

Trainable COSFIRE Filters for Keypoint Detection and Pattern Recognition

George Azzopardi and Nicolai Petkov

Abstract—Background: Keypoint detection is important for many computer vision applications. Existing methods suffer from insufficient selectivity regarding the shape properties of features and are vulnerable to contrast variations and to the presence of noise or texture. **Methods:** We propose a trainable filter which we call Combination Of Shifted Filter REsponses (COSFIRE) and use for keypoint detection and pattern recognition. It is automatically configured to be selective for a local contour pattern specified by an example. The configuration comprises selecting given channels of a bank of Gabor filters and determining certain blur and shift parameters. A COSFIRE filter response is computed as the weighted geometric mean of the blurred and shifted responses of the selected Gabor filters. It shares similar properties with some shape-selective neurons in visual cortex, which provided inspiration for this work. **Results:** We demonstrate the effectiveness of the proposed filters in three applications: the detection of retinal vascular bifurcations (DRIVE dataset: 98.50 percent recall, 96.09 percent precision), the recognition of handwritten digits (MNIST dataset: 99.48 percent correct classification), and the detection and recognition of traffic signs in complex scenes (100 percent recall and precision). **Conclusions:** The proposed COSFIRE filters are conceptually simple and easy to implement. They are versatile keypoint detectors and are highly effective in practical computer vision applications.

Index Terms—Feature detection, feature representation, medical information systems, object recognition, optical character recognition, shape

1 INTRODUCTION

THE detection of perceptually salient features, often referred to as keypoints or landmarks, is an important task in many computer vision applications, such as image registration, stereo camera calibration, object tracking, and object recognition.

A substantial body of work has been done in this area and several methods have been proposed for the detection, description and matching of keypoints. These methods characterize a keypoint by a specific data structure derived from the image content in the surroundings of the concerned point. In this sense, the terms keypoint and landmark refer to a local pattern rather than a single point. The typical patterns of interest range from simple edges to corners and junctions, Fig. 1. The Harris detector [1], for instance, detects corner-like structures and achieves rotation invariance by using the eigenvalues of the Hessian matrix. This detector, which aroused much interest, was extended by including local gray-level invariants based on combinations of Gaussian derivatives [2]. Later, scale-invariant approaches were proposed by selecting keypoints as the maxima points in a Laplacian [3] or Difference-of-Gaussian (DoG) [4] scale space. The Laplacian-based scale selection and the Harris detector were also combined into the so-called Harris-Laplace operator [5].

• The authors are with the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands.
E-mail: {g.azzopardi, n.petkov}@rug.nl.

Manuscript received 7 Dec. 2011; revised 5 Apr. 2012; accepted 26 Apr. 2012; published online 8 May 2012.

Recommended for acceptance by T. Tuytelaars.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number
TPAMI-2011-12-0874.

Digital Object Identifier no. 10.1109/TPAMI.2012.106.

A salient feature or keypoint is often characterized by a local image descriptor, which may vary from a simple scalar value to a rich description, such as a feature vector, a bag of values, or some other data structure. An extensive survey of local descriptors can be found in [6]. It compares a number of descriptors, including derivatives of Gaussians [7], moment invariants [8], complex features [9], responses of steerable filters [10], phase-based local features [11], and shows that the best performance is achieved with the SIFT descriptor [12]. Various extensions of the SIFT descriptor have been proposed, including the PCA-SIFT [13] and the GLOH [6], which use principal component analysis for dimensionality reduction. Nevertheless, the original SIFT approach outperforms both mentioned variants and seems to be the most popular keypoint descriptor currently. Recently, another operator called SURF [14] has been introduced which is somewhat similar to SIFT and speeds up the efficiency of keypoint selection.

The detection of keypoints that are similar to some keypoint which is selected as a prototype is typically done by computing a similarity (or dissimilarity) measure that is usually based on the Euclidean (or some other) distance between the respective keypoint descriptors. These methods are not robust to contrast variations and as a result they suffer from insufficient selectivity to the shape properties of features. This issue is illustrated by Fig. 2. The pattern in Fig. 2a that is formed by two lines that make a right-angle vertex that is, as a shape, very different from a pattern that is formed by just one of the constituent lines, Fig. 2b. Approaches that are based on the dissimilarity between keypoint descriptors such as the ones mentioned above may find these two patterns similar to a considerable extent. On the other hand, such methods might produce lower similarity scores for patterns that are regarded as similar from the

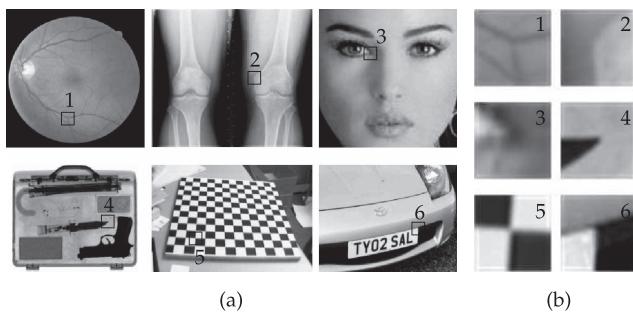


Fig. 1. Examples of corners and junction patterns marked in (a) photographic images and (b) their enlargements.

aspect of shape by a human observer, but show differences in contrast and/or contain texture, Figs. 2c, 2d.

In this paper, we are interested in the detection of **contour-based patterns**. We introduce trainable keypoint detection operators that are configured to be selective for given local patterns defined by the geometrical arrangement of contour segments. The proposed operators are inspired by the properties of a specific type of shape-selective neuron in area V4 of visual cortex which exhibit selectivity for parts of (curved) contours or for combinations of line segments [15], [16].

We call the proposed keypoint detector Combination Of Shifted Filter REsponses (COSFIRE) filter as the response of such a filter in a given point is computed as a function of the shifted responses of simpler (in this case orientation-selective) filters. Using shifted responses of simpler filters—Gabor filters in this study—corresponds to combining their respective supports at different locations to obtain a more sophisticated filter with a bigger support. The specific function that we use here to combine filter responses is weighted geometric mean, essentially multiplication, which has specific advantages regarding shape recognition and robustness to contrast variations. Such a model design decision is mainly motivated by the better results obtained using multiplication versus addition. It gets further support by psychophysical evidence [17] that curved contour parts are likely detected by a neural mechanism that multiplies the responses of afferent subunits (sensitive for different parts of the curve pattern). Due to the multiplicative character of the output function, a COSFIRE filter produces a response only when all constituent parts of a pattern of interest are present.

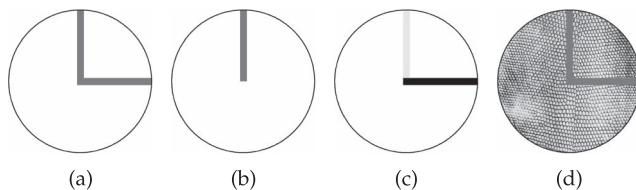


Fig. 2. (a) Prototype pattern. (b) Test pattern which has 50 percent similarity (computed by template matching) to the prototype. (c), (d) Test patterns that have only 30 percent similarity to the prototype due to (c) contrast differences and (d) presence of texture. From a shape detection point of view, the patterns in (c) and (d) are more similar to the prototype in (a) than the pattern in (b). This example shows the shortcomings of other models that are based on distance or dissimilarity of descriptors. The local image pattern is used as a descriptor in this example. Methods that compute local descriptors only shift the problem to a feature space.

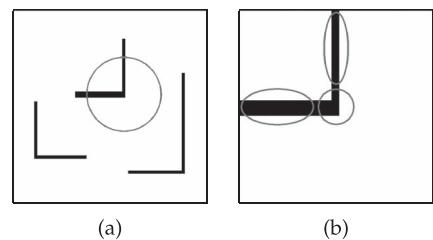


Fig. 3. (a) Synthetic input image (of size 256 × 256 pixels). The circle indicates a prototype feature of interest that is manually selected by a user. (b) Enlargement of the selected feature. The ellipses represent the support of line detectors that are identified as relevant for the concerned feature.

A COSFIRE filter is conceptually simple and straightforward to implement: It requires the application of selected Gabor filters, Gaussian blurring of their responses, shifting of the blurred responses by specific, different vectors, and multiplying the shifted responses. The questions of which Gabor filters to use, how much to blur their responses, and how to shift the blurred responses are answered in a COSFIRE filter configuration process in which a local pattern of interest that defines a keypoint is automatically analyzed. The configured COSFIRE filter can then successfully detect the same and similar patterns. We also show how the proposed COSFIRE filters can achieve invariance to rotation, scale, reflection, and contrast inversion.

The rest of the paper is organized as follows: In Section 2, we present the COSFIRE filter and demonstrate how it can be trained and used to detect local contour patterns. In Section 3, we demonstrate the effectiveness of the proposed trainable COSFIRE filters in three practical applications: the detection of vascular bifurcations in retinal fundus images, the recognition of handwritten digits, and the detection and recognition of traffic signs in complex scenes. Section 4 contains a discussion of some aspects of the proposed approach and highlights the differences that distinguish it from other approaches. Finally, we draw conclusions in Section 5.

2 METHOD

2.1 Overview

The following example illustrates the main idea of our method. Fig. 3a shows an input image containing three vertices. We consider the encircled vertex, which is shown enlarged in Fig. 3b, as a (prototype) pattern of interest and use it to automatically configure a COSFIRE filter that will respond to the same and similar patterns.

The two ellipses shown in Fig. 3b represent the dominant orientations in the neighborhood of the specified point of interest. We detect such lines by symmetric Gabor filters. The central circle represents the overlapping supports of a group of such filters. The response of the proposed COSFIRE detector is computed by combining the responses of these Gabor filters in the centers of the corresponding ellipses by multiplication. The preferred orientations of these filters and the locations at which we take their responses are determined by analyzing the local prototype pattern used for the configuration of the COSFIRE filter concerned. Consequently, the filter is selective for the presented local spatial arrangement of lines of specific orientations and widths. Taking the responses of Gabor

filters at different locations around a point can be implemented by shifting the responses of these Gabor filters by different vectors before using them for the pixel-wise evaluation of a multivariate function which gives the COSFIRE filter output.

In the next sections, we explain the automatic configuration process of a COSFIRE filter that will respond to a given prototype feature of interest and similar patterns. The configuration process determines which responses of which Gabor filters in which locations need to be multiplied in order to obtain the output of the filter.

2.2 Detection of Orientations by 2D Gabor Filters

We build the proposed COSFIRE filter using as input the responses of Gabor filters, which are known for their orientation selectivity.

We denote by $g_{\lambda,\theta}(x,y)$ the response of a Gabor filter of preferred wavelength λ and orientation θ to a given input image. Such a filter has other parameters, such as spatial aspect ratio, bandwidth, and phase offset, that we skip here for brevity. The responses of a symmetrical and an antisymmetrical filter can be combined in a Gabor energy filter. Surround suppression can also be applied to Gabor (energy) filter responses to reduce responses to texture and improve the detectability of object contours. For brevity of presentation, we do not consider all these aspects of Gabor filters here and we refer to [18], [19], [20], [21], [22], [23], [24] for technical details and to our online implementation.¹ We normalize² all Gabor functions that we use in such a way that all positive values of such a function sum up to 1 and all negative values sum up to -1.

We threshold the responses of Gabor filters at a given fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response of $g_{\lambda,\theta}(x,y)$ across all combinations of values (λ, θ) used and all positions (x, y) in the image, and denote these thresholded responses by $|g_{\lambda,\theta}(x,y)|_{t_1}$. We comment on the choice of the value of t_1 in Sections 3 and 4.

2.3 Configuration of a COSFIRE Filter

A COSFIRE filter uses as input the responses of some Gabor filters, each characterized by parameter values (λ_i, θ_i) , around certain positions (ρ_i, ϕ_i) with respect to the center of the COSFIRE filter. A set of four parameter values $(\lambda_i, \theta_i, \rho_i, \phi_i)$ characterizes the properties of a contour part that is present in the specified area of interest: $\lambda_i/2$ represents the width, θ_i represents the orientation, and (ρ_i, ϕ_i) represents the location. In the following we explain how we obtain the parameter values of such contour parts around a given point of interest.

We consider the responses of a bank of Gabor filters along a circle³ of a given radius ρ around a selected point of interest, Fig. 4. In each position along that circle, we take the maximum of all responses across the possible values of (λ, θ) used in the filter bank. The positions that have values greater than the corresponding values of the neighboring positions along an arc of angle $\pi/8$ are chosen as the points

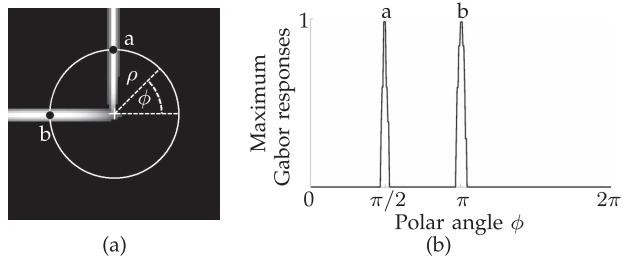


Fig. 4. Configuration of a COSFIRE filter. (a) The gray level of a pixel represents the maximum value superposition of the thresholded (at $t_1 = 0.2$) responses of a bank of Gabor filters (five wavelengths $\lambda \in \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$ and eight orientations $\theta \in \{\frac{\pi}{8}, i = 0 \dots 7\}$) at that position. The white cross indicates the location of the point of interest selected by a user and the bright circle of a given radius (here $\rho = 30$ pixels) indicates the locations considered around the point of interest. (b) Values of the maximum value superposition of thresholded Gabor filter responses along the concerned circle. The labeled black dots in (a) mark the positions (relative to the center of the filter) at which the respective strongest Gabor filter responses are taken. These two positions correspond to the two local maxima in the plot in (b).

that characterize the dominant orientations around the point of interest. We determine the polar coordinates (ρ_i, ϕ_i) for each such point with respect to the center of the filter. For such a location (ρ_i, ϕ_i) we then consider all combinations of (λ, θ) for which the corresponding responses $g_{\lambda,\theta}(x,y)$ are greater than a fraction $t_2 = 0.75$ of the maximum of $g_{\lambda,\theta}(x,y)$ across the different combinations of values (λ, θ) used. For each value θ that satisfies this condition, we consider a single value of λ , the one for which $g_{\lambda,\theta}(x,y)$ is the maximum of all responses across all values of λ . For each distinct pair of (λ, θ) and for location (ρ_i, ϕ_i) we obtain a tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$. Thus, multiple tuples can be formed for the same location (ρ_i, ϕ_i) . In Section 4, we provide further comment on the choice of the value of t_2 .

We denote by $S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i = 1 \dots n_f\}$ the set of parameter value combinations which fulfill the above conditions. The subscript f stands for the local prototype pattern around the selected point of interest. Every tuple in the set S_f specifies the parameters of some contour part in f .

For the point of interest shown in Fig. 4a, with two values of the parameter ρ ($\rho \in \{0, 30\}$), the selection method described above results in four contour parts with parameter values specified by the tuples in the following set:

$$\begin{aligned} S_f = \{ & \\ & (\lambda_1 = 8, \quad \theta_1 = 0, \quad \rho_1 = 0, \quad \phi_1 = 0), \\ & (\lambda_2 = 8, \quad \theta_2 = 0, \quad \rho_2 = 30, \quad \phi_2 = \pi/2), \\ & (\lambda_3 = 16, \quad \theta_3 = \pi/2, \quad \rho_3 = 0, \quad \phi_3 = 0), \\ & (\lambda_4 = 16, \quad \theta_4 = \pi/2, \quad \rho_4 = 30, \quad \phi_4 = \pi). \end{aligned}$$

The last tuple in S_f , $(\lambda_4 = 16, \theta_4 = \pi/2, \rho_4 = 30, \phi_4 = \pi)$, for instance, describes a contour part with a width of $(\lambda_4/2 = 8$ pixels and an orientation $\theta_4 = \pi/2$ that can be detected by a Gabor filter with preferred wavelength $\lambda_4 = 16$ and orientation $\theta_4 = \pi/2$, at a position of $\rho_4 = 30$ pixels to the left ($\phi_4 = \pi$) of the point of interest; this location is marked by the label "b" in Fig. 4. This selection is the result of the presence of a horizontal line to the left of the center of the feature that is used for the configuration of the filter.

1. <http://matlabserver.cs.rug.nl>.

2. This normalization ensures that the response to an image of constant intensity is 0. Without normalization, this is true only for antisymmetrical filters. It also ensures that the response to a line of width w will be largest for a symmetrical filter of preferred wavelength $\lambda = 2w$. We mention this explicitly because line detection is essential in one application that we present in Section 3.

3. For $\rho = 0$, we only consider the point of interest.

2.4 Blurring and Shifting Gabor Filter Responses

The above analysis of the considered local pattern of interest f indicates that this pattern produces four strong responses $g_{\lambda_i, \theta_i}(x, y)$ of Gabor filters with parameters $(\lambda_1 = 8, \theta_1 = 0)$, $(\lambda_2 = 8, \theta_2 = 0)$, $(\lambda_3 = 16, \theta_3 = \pi/2)$, and $(\lambda_4 = 16, \theta_4 = \pi/2)$ in the corresponding positions with polar coordinates (ρ_i, ϕ_i) with respect to the filter center. Next, we use these responses to compute the output of the COSFIRE filter. Since the concerned responses are in different positions (ρ_i, ϕ_i) with respect to the filter center, we first shift them appropriately so that they come together in the filter center. The COSFIRE filter output can then be evaluated as a pixel-wise multivariate function of the shifted Gabor filter responses.

Before these shift operations, we blur the Gabor filter responses in order to allow for some tolerance in the position of the respective contour parts. We define the blurring operation as the computation of maximum value of the weighted thresholded responses of a Gabor filter. For weighting, we use a Gaussian function $G_\sigma(x, y)$, the standard deviation σ of which is a linear function of the distance ρ from the center of the COSFIRE filter,

$$\sigma = \sigma_0 + \alpha\rho, \quad (1)$$

where σ_0 and α are constants. The choice of the linear function in (1) is explained in Section 4. The value of the parameter α determines the orientation tuning of the COSFIRE filter: The orientation bandwidth becomes broader with an increasing value of α .

Next, we shift the blurred responses of each selected Gabor filter (λ_i, θ_i) by a distance ρ_i in the direction opposite to ϕ_i . In polar coordinates, the shift vector is specified by $(\rho_i, \phi_i + \pi)$. In Cartesian coordinates, it 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 denote by $s_{\lambda_i, \theta_i, \rho_i, \phi_i}(x, y)$ the blurred and shifted response of the Gabor filter that is specified by the i th tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the set S_f :

$$s_{\lambda_i, \theta_i, \rho_i, \phi_i}(x, y) \stackrel{\text{def}}{=} \max_{x', y'} \{|g_{\lambda_i, \theta_i}(x - x' - \Delta x_i, y - y' - \Delta y_i)|_{t_1} G_\sigma(x', y')\}, \quad (2)$$

where $-3\sigma \leq x', y' \leq 3\sigma$.

Fig. 5 illustrates the blurring and shifting operations for this COSFIRE filter, applied to the image in Fig. 3a. For each of the four contour parts detected in the prototype feature pattern, we first compute the corresponding Gabor filter responses and then we blur and shift these responses accordingly.

In practice, the computation of one blurred response (for the same values of the parameters λ , θ , and ρ), for instance with $s_{\lambda, \theta, \rho, \phi=0}(x, y)$, is sufficient: The result of $s_{\lambda, \theta, \rho, \phi}(x, y)$ for any value of ϕ can be obtained from the result of the output of $s_{\lambda, \theta, \rho, \phi=0}(x, y)$ by appropriate shifting.

2.5 Response of a COSFIRE Filter

We define the response $r_{S_f}(x, y)$ of a COSFIRE filter as the weighted geometric mean of all the blurred and shifted thresholded Gabor filter responses $s_{\lambda_i, \theta_i, \rho_i, \phi_i}(x, y)$ that correspond to the properties of the contour parts described by S_f :

$$r_{S_f}(x, y) \stackrel{\text{def}}{=} \left| \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i, \theta_i, \rho_i, \phi_i}(x, y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_f|} \omega_i} \right|_{t_3} \quad (3)$$

$$\omega_i = \exp^{-\frac{\rho_i^2}{2\sigma'^2}}, \quad 0 \leq t_3 \leq 1,$$

where $|.|_{t_3}$ stands for thresholding the response at a fraction t_3 of its maximum across all image coordinates (x, y) . For $1/\sigma' = 0$, the computation of the COSFIRE filter becomes equivalent to the standard geometric mean, where the s -quantities have the same contribution. Otherwise, for $1/\sigma' > 0$, the input contribution of s -quantities decreases with an increasing value of the corresponding parameter ρ . In our experiments we use a value of the standard deviation σ' that is computed as a function of the maximum value of the given set of ρ values: $\sigma' = (-\rho_{\max}^2 / 2 \ln 0.5)^{1/2}$, where $\rho_{\max} = \max_{i \in \{1, \dots, |S_f|\}} \{\rho_i\}$. We make this choice in order to achieve a maximum value $\omega = 1$ of the weights in the center (for $\rho = 0$), and a minimum value $\omega = 0.5$ in the periphery (for $\rho = \rho_{\max}$).

Fig. 5 shows the output of a COSFIRE filter which is defined as the weighted geometric mean of four blurred and shifted images from the responses of two Gabor filters. Note that this filter responds at points where a pattern is present which is identical or similar to the prototype pattern f at and around the selected point of interest, which was used in the configuration of the filter. In this example, the COSFIRE filter reacts strongly in a given point to a local pattern that contains a horizontal line to the left of that point, a vertical line above it, together with a horizontal and a vertical line at the point.

Fig. 6a shows a set of elementary features that are angles of different acuteness and orientations. For the illustration in Fig. 6, we configure a COSFIRE filter using the enframed local pattern in Fig. 6a where the point of interest is positioned on the corner of the vertex. The structure of the filter is determined by using three values of ρ ($\rho \in \{0, 12, 30\}$). Fig. 6b shows the responses of this COSFIRE filter where the strength of the maximum filter response to a given feature is rendered as a gray-level shading of that feature. The maximum response is reached at or near the corner. In this case, the COSFIRE filter achieves the strongest response to the local prototype pattern that was used to configure it, but it also reacts, with less than the maximum response, to angles that differ slightly in acuteness and/or orientation. This example illustrates the selectivity and the generalization ability of the proposed filter.

2.6 Achieving Invariance

In the following, we explain how we achieve invariance to rotation, scale, reflection, and contrast inversion.

2.6.1 Rotation Invariance

Using the set S_f that defines the concerned filter, we form a new set $\mathcal{R}_\psi(S_f)$ that defines a new filter, which is selective for a version of the prototype feature f that is rotated by an angle ψ :

$$\mathcal{R}_\psi(S_f) \stackrel{\text{def}}{=} \{(\lambda_i, \theta_i + \psi, \rho_i, \phi_i + \psi) | \forall (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\}. \quad (4)$$

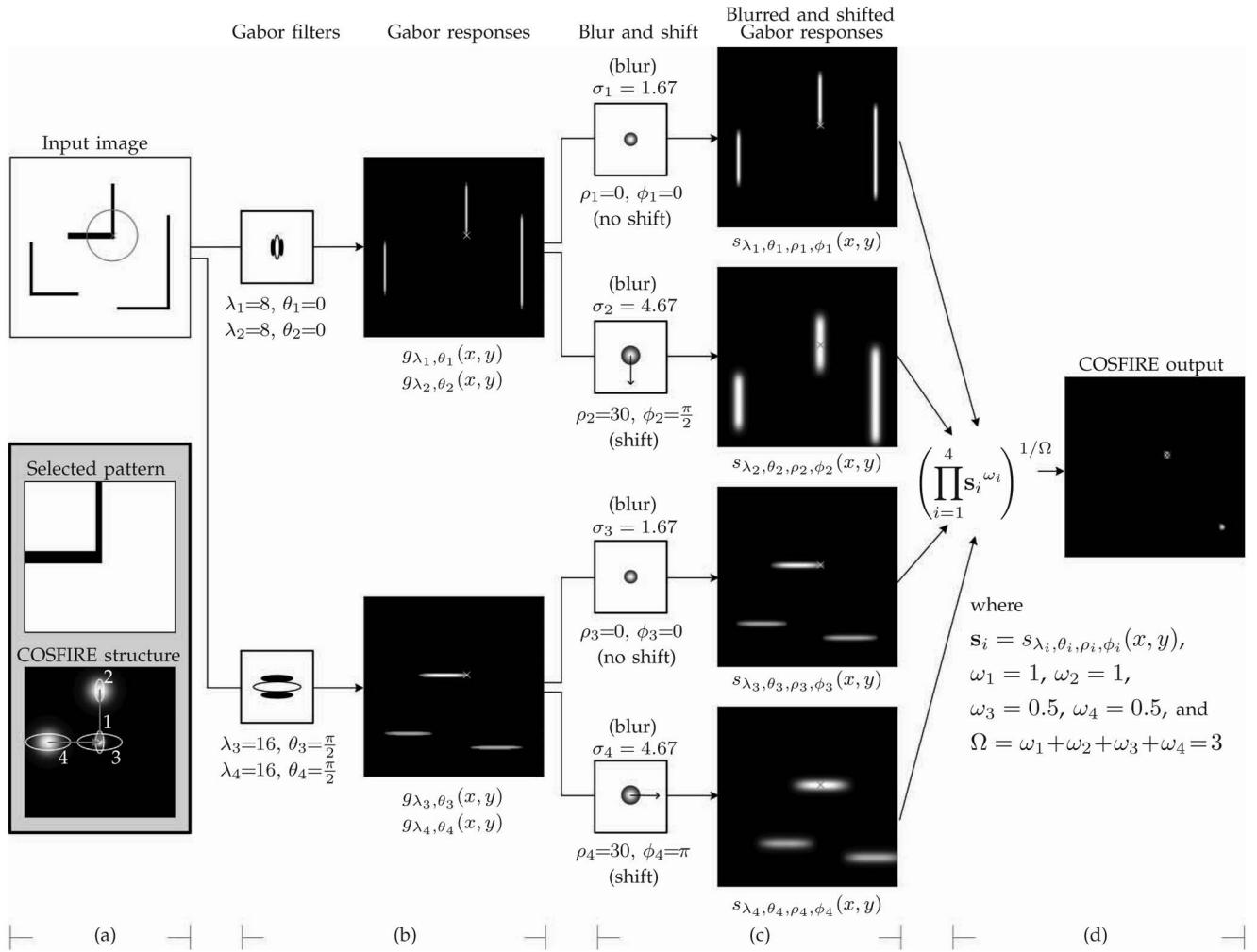


Fig. 5. (a) Input image (of size 256×256 pixels). The enframed inlay images show (top) the enlarged prototype feature of interest, which is the vertex encircled in the input image and (bottom) the structure of the COSFIRE filter that is configured for this feature. This filter is trained to detect the spatial local arrangement of four contour parts. The ellipses illustrate the wavelengths and orientations of the Gabor filters, and the bright blobs are intensity maps for Gaussian functions that are used to blur the responses of the corresponding Gabor filters. The blurred responses are then shifted by the corresponding vectors. (b) Each contour part of the input pattern is detected by a Gabor filter with a given preferred wavelength λ_i and orientation θ_i . Two of these parts ($i = \{1, 2\}$) are detected by the same Gabor filter and the other two parts ($i = \{3, 4\}$) are detected by another Gabor filter; therefore, only two distinct Gabor filters are selected from the filter bank. (c) We then blur the thresholded (here at $t_1 = 0.2$) response $|g_{\lambda_i, \theta_i}(x, y)|_{t_1}$ of each concerned Gabor filter and subsequently shift the resulting blurred response images by corresponding polar coordinate vectors $(\rho_i, \phi_i + \pi)$. (d) Finally, we obtain the output of the COSFIRE filter by computing the weighted geometric mean (here $\sigma' = 25.48$) of all the blurred and shifted thresholded Gabor filter responses. The \times marker indicates the location of the specified point of interest. The two local maxima in the output of the COSFIRE filter correspond to the two similar vertices in the input image.

For each tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the original filter S_f that describes a certain local contour part, we provide a counterpart tuple $(\lambda_i, \theta_i + \psi, \rho_i, \phi_i + \psi)$ in the new set $\mathcal{R}_\psi(S_f)$. The orientation of the concerned contour part and its polar angle position with respect to the center of the filter are offset by an angle ψ relative to the values of the corresponding parameters of the original part.

Fig. 6c shows the responses $r_{\mathcal{R}_\psi(S_f)}$ of the COSFIRE filter that correspond to $\mathcal{R}_\psi(S_f)$ to the set of elementary features shown in Fig. 6a. This filter responds selectively to a version of the original prototype feature f rotated counterclockwise at an angle of ($\psi = \pi/2$). It is, however, configured by manipulating the set of parameter value combinations, rather than by computing them from the responses to a rotated version of the original prototype pattern f .

A rotation-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter ψ :

$$\hat{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi} \{r_{\mathcal{R}_\psi(S_f)}(x, y)\}, \quad (5)$$

where Ψ is a set of n_ψ equidistant orientations defined as $\Psi = \{\frac{2\pi}{n_\psi} i \mid 0 \leq i < n_\psi\}$. Fig. 6d shows the maximum superposition $\hat{r}_{S_f}(x, y)$ for $n_\psi = 16$. The filter according to (5) produces the same response to local patterns that are versions of each other, obtained by rotation at discrete angles $\psi \in \Psi$.

As to the response of the filter to patterns that are rotated at angles of intermediate values between those in Ψ , it depends on the orientation selectivity of the filter S_f that is influenced by the orientation bandwidth of the involved Gabor filters and by the value of the parameter α in (1). Fig. 7 illustrates the orientation selectivity of the COSFIRE filter, which is configured with the enframed local prototype pattern in Fig. 6a using $\alpha = 0.1$. A maximum response is obtained for the local prototype pattern that was used to

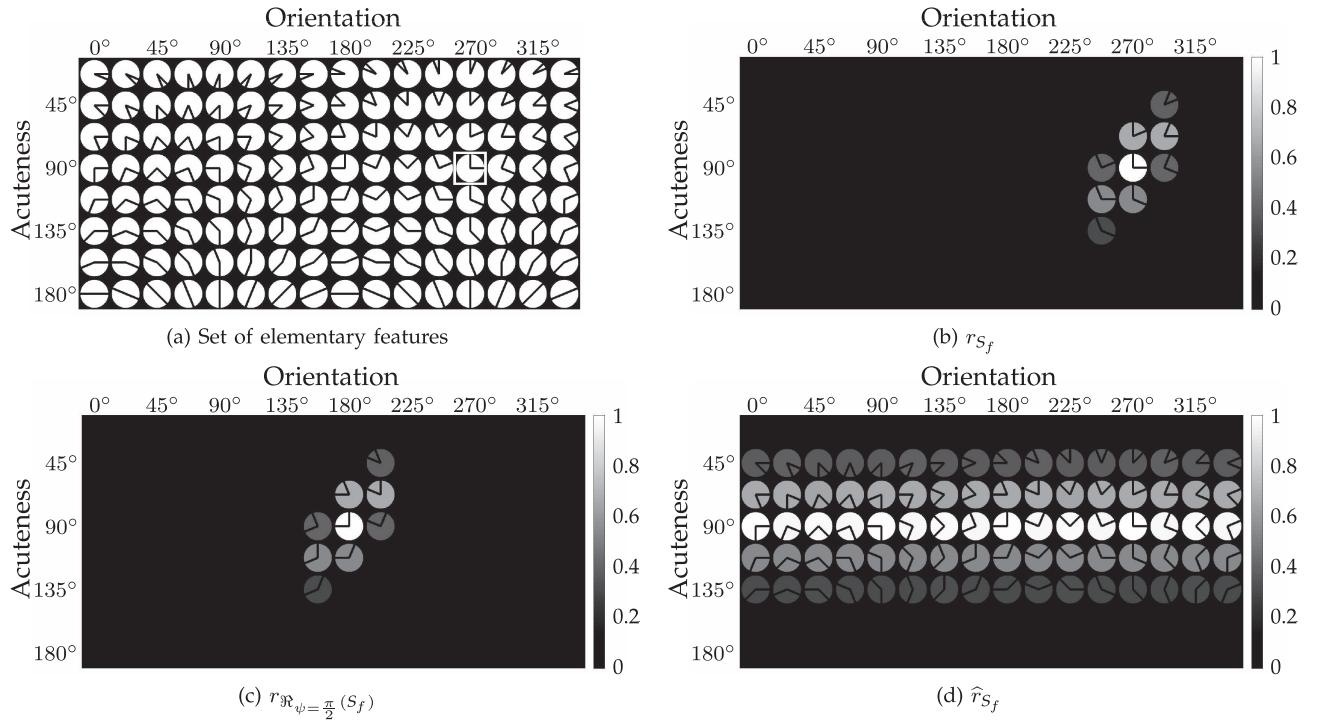


Fig. 6. (a) A set of elementary features. The enframed feature is used as a prototype for configuring a COSFIRE filter. (b) Responses of the configured filter rendered by shading of the features. (c) Responses of a rotated version ($\psi = \frac{\pi}{2}$) of the filter obtained by manipulation of the filter parameters. (d) Rotation-invariant responses for 16 discrete orientations.

configure this filter. The response declines with the deviation of the orientation of the local input pattern from the optimal one and practically disappears when this deviation is greater than $\pi/8$. When the deviation of the orientation is $\pi/16$, the response of the filter is approximately half of the maximum response. This means that the half-response bandwidth of this COSFIRE filter is $\pi/8$. Thus, $n_\psi = 16$ distinct preferred orientations (in intervals of $\pi/8$) ensure sufficient response for any orientation of the feature used to configure the filter.

As demonstrated by Fig. 6d, when the concerned filter is applied in rotation-invariant mode ($n_\psi = 16$), it responds selectively to the prototype pattern, a right angle, independently of the orientation of the angle.

2.6.2 Scale Invariance

?twb=0pc2.58>Scale invariance is achieved in a similar way. Using the set S_f that defines the concerned filter, we form a new set $T_v(S_f)$ that defines a new filter which is selective for a version of the prototype feature f that is scaled in size by a factor v :

$$T_v(S_f) \stackrel{\text{def}}{=} \{(v\lambda_i, \theta_i, v\rho_i, \phi_i) | \forall (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\}. \quad (6)$$

For each tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the original filter S_f that describes a certain local contour part, we provide a counterpart tuple $(v\lambda_i, \theta_i, v\rho_i, \phi_i)$ in the new set $T_v(S_f)$. The width of the concerned contour part and its distance to the center of the filter are scaled by the factor v relative to the values of the corresponding parameters of the original part.

A scale-invariant response is achieved by taking the maximum value of the responses of filters that are obtained with different values of the parameter v :

$$\tilde{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{v \in \Upsilon} \{r_{T_v(S_f)}(x, y)\}, \quad (7)$$

where Υ is a set of v values equidistant on a logarithmic scale defined as $\Upsilon = \{2^{\frac{i}{n}} | i \in \mathbb{Z}\}$.

2.6.3 Reflection Invariance

As to reflection invariance, we first form a new set \dot{S}_f from the set S_f as follows:

$$\dot{S}_f \stackrel{\text{def}}{=} \{(\lambda_i, \pi - \theta_i, \rho_i, \pi - \phi_i) | \forall (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\}, \quad (8)$$

The new filter which is defined by the set \dot{S}_f is selective for a reflected version of the prototype feature f about the y -axis. A reflection-invariant response is achieved by taking the maximum value of the responses of the filters S_f and \dot{S}_f :

$$\dot{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max\{r_{S_f}(x, y), r_{\dot{S}_f}(x, y)\}. \quad (9)$$

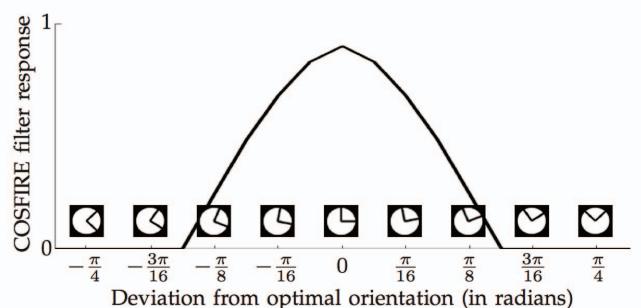


Fig. 7. Orientation selectivity of a COSFIRE filter that is configured with a right-angle vertex.

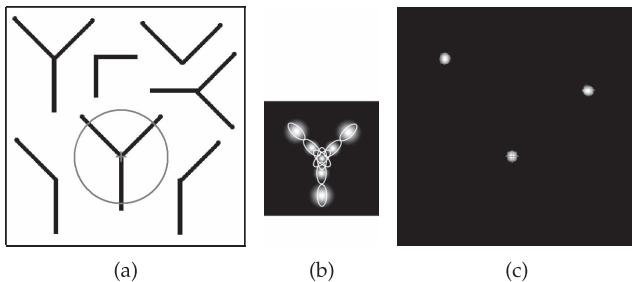


Fig. 8. (a) Synthetic input image (of size 256×256 pixels). (b) The structure of a COSFIRE filter that is configured using the encircled pattern in (a) with three values of ρ ($\rho \in \{0, 12, 30\}$) and $\sigma_0 = 2.5$. (c) Rotation-invariant response \hat{r}_{S_f} of the COSFIRE filter (here $\sigma' = 25.48$).

2.6.4 Combined Invariance to Rotation, Scale, and Reflection

A combined rotation-, scale-, and reflection-invariant response is achieved by taking the maximum value of the rotation- and scale-invariant responses of the filters S_f and \dot{S}_f that are obtained with different values of the parameters ψ and v :

$$\bar{r}_{S_f}(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi, v \in \Upsilon} \{\hat{r}_{\mathcal{R}_\psi(T_v(S_f))}(x, y), \hat{r}_{\mathcal{R}_\psi(T_v(\dot{S}_f))}(x, y)\}. \quad (10)$$

2.6.5 Invariance to Contrast Inversion

Next to the above geometric invariances, we can achieve invariance to contrast inversion by using Gabor filters with inverse polarity.

We do not elaborate further on this possibility because we do not use it in the applications presented below.

2.7 Detection of More Complex Patterns

The filter considered above is selective for a local pattern that consists of two lines forming an angle. However, in the configuration of the COSFIRE filter we made no assumptions about the specific type of local pattern it should detect. The configuration result is determined by the local prototype pattern presented.

Next, we illustrate the configuration of a filter that can detect a bifurcation pattern formed by three lines of different orientations, Figs. 8a, 8b. In Fig. 8c, we show the rotation-invariant response of the concerned COSFIRE filter for the input image in Fig. 8a. Besides the original local prototype

pattern that was used to configure this filter, it correctly detects the presence of another two similar features: one in a vertical orientation and the other pattern in a horizontal orientation.

3 APPLICATIONS

In the following, we demonstrate the effectiveness of the proposed COSFIRE filters by applying them in three practical applications: the detection of vascular bifurcations in retinal fundus images, the recognition of handwritten digits, and the detection and recognition of traffic signs in complex scenes.

3.1 Detection of Retinal Vascular Bifurcations

Retinal fundus images give a unique possibility to take a noninvasive look at the state of the vascular system of a person. The vascular geometrical structure in the retina is known to conform to structural principles which are related to certain physical properties [25], [26], [27], [28]. The analysis of the geometrical structure is important as deviations from the optimal principles may indicate (increased risk of) some cardiovascular diseases, such as hypertension [29] and atherosclerosis [30]; a comprehensive analysis is given in [31]. The identification of vascular bifurcations is one of the basic steps in this analysis. There are no state-of-the-art automatic techniques yet, and hence a time-consuming manual process is usually adopted [30]. Automating the identification of vascular bifurcations is an essential step in the description of the vascular tree that is needed for further analysis.

In the following, we show how trainable COSFIRE filters of the type introduced above can be configured to detect vascular bifurcations in retinal fundus images.

Figs. 9a, 9b show a retinal fundus image and its segmentation in blood vessels and background, which are both taken from the DRIVE dataset [32]. The latter image contains 107 blood vessel features, shown encircled, which present Y- or T-form bifurcations or cross overs.

We apply to the binary segmentation image a bank of symmetric Gabor filters with eight equidistant orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0 \dots 7\}$) and five wavelengths equidistantly spaced on a logarithmic scale ($\lambda \in \{4(2^i) \mid i = 0 \dots 4\}$) and threshold the results at $t_1 = 0.2$ of the maximum possible response. This threshold value is sufficient to preserve all

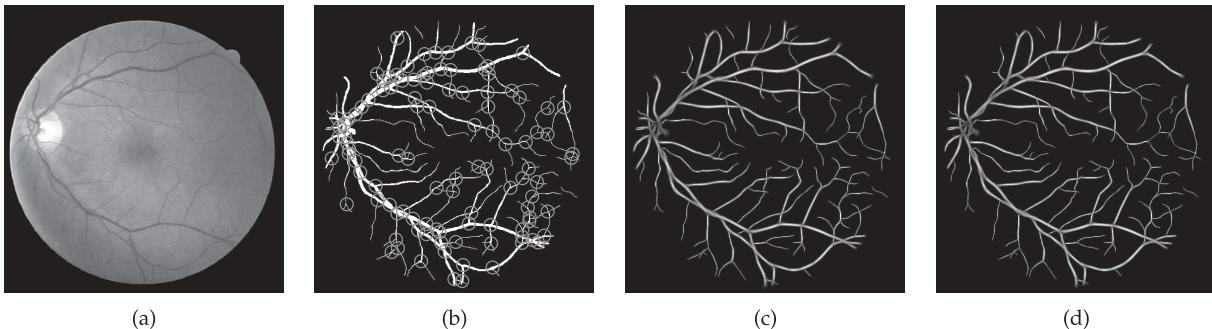


Fig. 9. Example of a retinal fundus image from the DRIVE dataset. (a) Original image (of size 564×584 pixels) with filename 21_training.tif. (b) Binary segmentation of vessels and background (also from DRIVE). The typical widths of blood vessels vary between 1 and 7 pixels. This range of width values determines our choice of the values of the wavelength λ used in the bank of Gabor filters. The circles surround Y- and T-formed vessel bifurcations and crossings. (c), (d) Superposition of the responses of a bank of symmetric Gabor filters with a threshold (c) $t_1 = 0$ and (d) $t_1 = 0.2$.

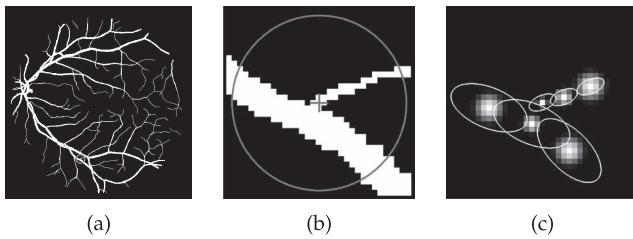


Fig. 10. Configuration of a COSFIRE filter. (a) The circle indicates a bifurcation feature f_1 selected for the configuration of the filter. (b) Enlargement of the selected feature. (c) Structure of the COSFIRE filter S_{f_1} configured for the specified bifurcation. The ellipses illustrate the involved Gabor filters and the positions in which their responses are taken.

junction regions and suppress the undesirable responses of Gabor filters, Fig. 9d.

Next, we select a vascular bifurcation that we use to configure a COSFIRE filter. In practice, the selection is done by specifying a region of appropriate size centered at the concerned feature. Fig. 10a illustrates the selection of one such region that is shown enlarged in Fig. 10b. In the following, we denote this prototype feature by f_1 . Fig. 10c shows the structure of a COSFIRE filter S_{f_1} that is configured for the specified feature. For the configuration of this filter, we use three values of the radius ρ ($\rho = \{0, 4, 10\}$).

Fig. 11 shows the results that are obtained by the application of filter S_{f_1} ($\sigma' = 8.49$) in different modes to the binary retinal fundus image shown in Fig. 10a. For this filter, we use a threshold value of $t_3 = 0.21$ as it produces the largest number of correctly detected bifurcations and no falsely detected features. The encircled regions⁴ are centered on the local maxima of the filter response and if two such regions overlap by 75 percent, only the one with the stronger response is shown.

When no invariance is used (Fig. 11a), the filter S_{f_1} detects four vascular bifurcations, one of which is the prototype pattern that was used to configure this filter. When the filter is applied in a rotation-invariant mode ($\psi \in \{\frac{\pi}{8} | i = 0 \dots 7\}$) it detects 24 features. With the addition of scale invariance ($\nu \in \{2^{-\frac{1}{2}}, 1, 2^{\frac{1}{2}}\}$) the filter detects 34 features, and with the inclusion of reflection invariance the COSFIRE filter S_{f_1} detects 67 bifurcations. These results illustrate how invariance to such geometric transformations can be used to boost the performance of a COSFIRE filter. It also shows the strong generalization capability of this approach because 62.62 percent (67 out of 107) of the features of interest are detected by one filter.

As to the remaining features that are not detected by the filter corresponding to feature f_1 , we proceed as follows: We take one of these features that we denote by f_2 (Fig. 12) and train a second COSFIRE filter S_{f_2} using it. With this second filter we detect 50 features of interest of which 35 overlap with features detected by the filter S_{f_1} and 15 are newly detected features ($t_3(S_{f_2}) = 0.25$). Applying the two filters together results in the detection of 82 distinct features. We continue adding filters that are configured using features that have not been detected by the previously trained filters. By configuring another two COSFIRE filters,

4. The radius of the circle is the sum of the maximum value of the radial parameter ρ and blur radius used at this value of ρ .

S_{f_3} and S_{f_4} (Fig. 12), and using them together with the other two filters we achieve 100 percent recall and 100 percent precision for the concerned image. This means that all 107 features shown in Fig. 9b are correctly detected and that there are no false responses of the filters.

We use an individual threshold value $t_3(S_{f_i})$ for each COSFIRE filter S_{f_i} by setting it to the smallest number for which the precision is still 100 percent for the training image.

We apply the same four COSFIRE filters on a dataset (DRIVE) of 40 binary retinal images⁵ and evaluate the obtained results with the ground truth data⁶ that was defined by the authors of this paper. The recall R and the precision P that we achieve depend on the values of the threshold parameters $t_3(S_{f_i})$: P increases and R decreases with increasing values of $t_3(S_{f_i})$. For each COSFIRE filter we add to (or subtract from) the corresponding learned threshold value $t_3(S_{f_i})$ the same offset value. With the referred four COSFIRE filters, the harmonic mean ($2PR/(P + R)$) of the precision and recall reaches a maximum at a recall R of 95.58 percent and a precision P of 95.25 percent when each $t_3(S_{f_i})$ is offset by the same amount of +0.05 from the corresponding learned threshold value. We extend our experiments by configuring up to eight COSFIRE filters of which the new four filters are configured for four prototype features taken from the same retinal image with filename 04_manual1.gif. We achieve the best results for six filters (Fig. 12) and show them together with the results for four filters in Fig. 13. With six filters the maximum harmonic mean is reached at a recall R of 98.50 percent and a precision P of 96.09 percent when the corresponding learned $t_3(S_{f_i})$ values are offset by the same amount of +0.07. We made this application available on the Internet.⁷

The optimal results for four and six COSFIRE filters are reached for a very small value of the offset: +0.05 and +0.07, respectively. This shows that the learned threshold values that are determined individually for each filter give results near to the optimal that may be expected.

In principle, all vascular bifurcations can be detected if a sufficient number of filters are configured and used. Furthermore, the precision can be improved by performing additional morphological analysis of the features that are detected by the filters. Even without these possible improvements, our results are better than those achieved in [33] where a recall of 95.82 percent was reported on a small dataset of five retinal images only.

3.2 Recognition of Handwritten Digits

Handwritten digit recognition is a challenging task in the community of pattern recognition which has various commercial applications, such as bank check processing and postal mail sorting. It has been used as a benchmark for comparing shape recognition methods. Feature extraction plays a significant role in the effectiveness of such systems. A detailed review of the state-of-the-art methods is given in [34].

In the following, we show how the proposed trainable COSFIRE filters can be configured to detect specific parts of

5. Named in DRIVE 01_manual1.gif, ..., 40_manual1.gif.

6. The ground truth data (coordinates of bifurcations and cross overs) can be downloaded from http://www.cs.rug.nl/~imaging/databases/retina_database.

7. <http://matlabserver.cs.rug.nl/RetinalVascularBifurcations>.

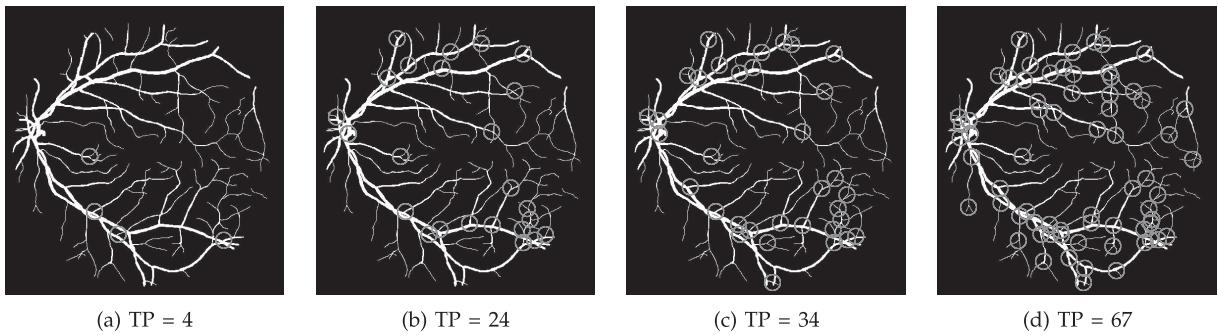


Fig. 11. Results of using the filter S_{f_1} in different modes: (a) noninvariant, (b) rotation-invariant, (c) rotation- and scale-invariant, and (d) rotation-, scale-, and reflection-invariant. The number of correctly detected features (TP—true positives) increases as the filter achieves invariance to such geometric transformations.

handwritten digits. Consequently, the collective responses of multiple such filters can be used as a shape descriptor of a given handwritten digit. We use the well-known modified NIST (MNIST) dataset [35] to evaluate the performance of this approach. This dataset comprises 60,000 training and 10,000 test digits⁸ where each digit is given as a grayscale image of size 28×28 pixels, Fig. 14.

In the configuration step, we choose a random subset of digit images from each digit class. For each such digit image we choose a random location in the image and use the local stroke pattern around that location to configure a COSFIRE filter. We use a given randomly selected location for the configuration of a COSFIRE filter only if that filter consists of at least four tuples; otherwise, we choose a different location. We impose this restriction in order to avoid the selection of small digit fragments as prototype patterns, which may consequently result in filters with low discriminative power. We provide further comments on the discriminative abilities of these COSFIRE filters in Section 4. For this application, we use three values of ρ ($\rho \in \{0, 3, 8\}$), $t_2 = 0.75$, $\sigma_0 = 0.83$, $\alpha = 0.1$, and a bank of antisymmetric Gabor filters with 16 equidistant orientations ($\theta \in \{\frac{\pi}{8}|i = 0 \dots 15\}$) and one wavelength of ($\lambda = 2\sqrt{2}$). Fig. 15 illustrates the configuration of four such COSFIRE filters using local prototype patterns (parts of digits) that are randomly selected from four handwritten digits.

We perform a number of experiments with different values of the threshold parameter t_1 ($t_1 \in \{0, 0.05, 0.1, 0.15\}$). The values of the other parameters mentioned above are kept fixed for all experiments. For each value of t_1 , we run an experiment by configuring up to 500 COSFIRE filters per digit class. We repeat such an experiment five times and report the average recognition rate. Repetition of experiments is necessary in order to compensate for the random selection of training digit images and the random selection of locations within these images that are used to configure the concerned filters.

After the configuration of a certain number of COSFIRE filters, every digit to which the set of these filters is applied can be described by a vector where each element corresponds to the maximum response of a COSFIRE filter across all locations in the input image. For instance, with 500 filters per digit class and 10 digit classes, a digit image to which this set of 5,000 COSFIRE filters is applied is described by a vector of 5,000 elements. For this application, the responses of the

concerned Gabor filters provide equal contribution ($1/\sigma' = 0$) to the output of the corresponding COSFIRE filter.

The feature vectors obtained for the digit images of the training set are then used to train an all-pairs multiclass (with majority vote) support vector machine (SVM) classifier with a linear kernel. In Fig. 16a, we plot the recognition rates that we achieve for different values of the threshold t_1 and for different numbers of COSFIRE filters used. We achieve a maximum recognition rate of 99.40 percent with 4,500 COSFIRE filters, where the filters are used in a noninvariant mode, i.e., without compensation for possible pattern

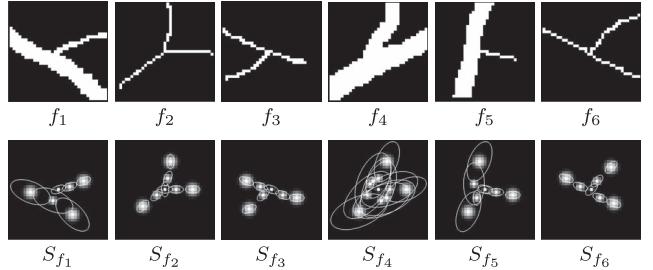


Fig. 12. (Top row) A set of six bifurcations and (bottom row) the structures of the corresponding six COSFIRE filters. The first four bifurcations are taken from the binary retinal image shown in Fig. 10a with filename 21_manual1.gif and the last two bifurcations are extracted from the retinal image with filename 04_manual1.gif. The following are the learned threshold values: $t_3(S_{f_1}) = 0.21$, $t_3(S_{f_2}) = 0.25$, $t_3(S_{f_3}) = 0.36$, $t_3(S_{f_4}) = 0.29$, $t_3(S_{f_5}) = 0.17$, and $t_3(S_{f_6}) = 0.25$.

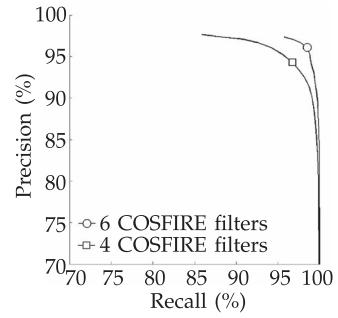


Fig. 13. Precision-recall plots obtained with four and six COSFIRE filters. For each plot the threshold parameter t_3 of each filter is varied by adding the same offset (ranging between -0.1 and 0.1) to the corresponding learned threshold value. The precision P increases and the recall R decreases with an increasing offset value. The harmonic mean (often used as a single measure of performance) of R and P reaches a maximum at $R = 98.50$ percent and $P = 96.09$ percent with six filters and at $R = 95.58$ percent and $P = 95.25$ percent for four filters. These points are marked by circle and square markers, respectively.

⁸ The MNIST dataset is available online: <http://yann.lecun.com/exdb/mnist>.

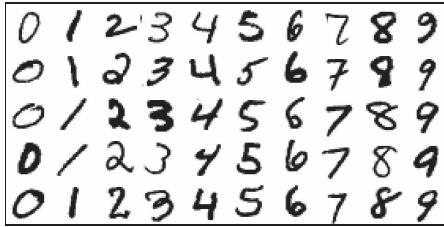


Fig. 14. Examples of handwritten digits from the MNIST dataset.

reflection, rotation, and scaling ($\psi = 0, v = 1$). In Fig. 14, one can observe, however, that some of the handwritten digits given in the MNIST dataset differ slightly in orientation. We consider this fact and repeat the five experiments for the threshold $t_1 = 0$ (that contributed to the best performance so far), but this time applying the same COSFIRE filters in a partially rotation-invariant mode with five values of the rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The plots in Fig. 16b show that the performance that is achieved with the partially rotation-invariant filters is improved to a maximum recognition rate of 99.48 percent with 4,000 filters. This means that the error rate is decreased by 13.33 percent and that 500 less filters are required.

The recognition rate of 99.48 percent that we achieve is comparable to the best results obtained with other approaches⁹ applied on the MNIST dataset. In particular, our method outperforms the shape context approach (99.37 percent [36]) and three other approaches (94.2 percent [37], 97.62 percent [38], and 98.73 percent [39]) that use biologically inspired features combined with a multilayer perceptron (MLP) [37] and a linear SVM classifier [38], [39]. The highest recognition rate achieved to date is 99.61 percent [40]. The approach used to achieve that result extends the original training dataset by elastically distorting the training images.

Notable is the fact that we achieve the above result without any optimization regarding the COSFIRE filters and the parameters used. Moreover, we do not perform any preprocessing and/or postprocessing operations, neither do we use an extended training dataset with elastic distortion. The fact that we achieve a result that is very close to the best result ever achieved is remarkable because our method is versatile and has not been developed with the specific handwritten digit recognition application in mind, while the best methods are results of long-lasting research effort in which elaborate application-specific techniques have been developed.

3.3 Detection and Recognition of Traffic Signs

The detection and recognition of specific objects in complex scenes is one of the most challenging tasks in computer vision. Here, we show how the proposed COSFIRE filters can be used for the detection of traffic signs in complex scenes.

We use a public dataset¹⁰ of 48 color images (of size 360×270 pixels) that was originally published in [41]. Each of these images contains (at least) one of three possible traffic signs illustrated in Figs. 17a, 17b, 17c.

For this application, we configure filters for patterns that are more complex than the ones involved in the previous two

9. A list of results obtained by state-of-the-art approaches is maintained at <http://yann.lecun.com/exdb/mnist/>.

10. Traffic sign dataset is online: http://www.cs.rug.nl/~imaging/databases/traffic_sign_database/traffic_sign_database.html.

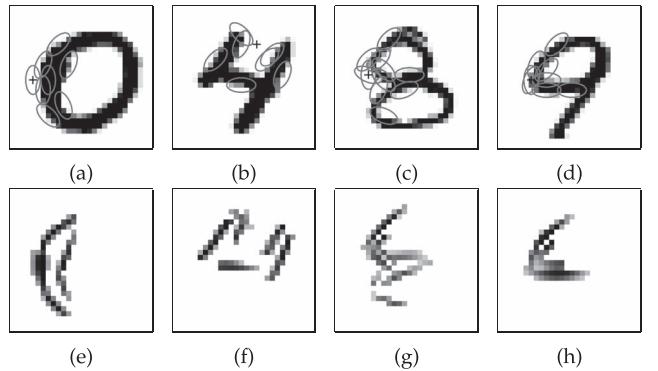


Fig. 15. Example of the configuration of four COSFIRE filters, one for each of the handwritten digits 0, 4, 8, and 9. (a)-(d) The “+” markers show the randomly selected locations. The ellipses around the marked locations represent the support of the Gabor filters that are determined in the configuration of the concerned COSFIRE filters. (e)-(h) The corresponding reconstructions of the local patterns that are illustrated as a superposition of the thresholded ($t_1 = 0.1$) Gabor filter (inverted) responses, which contribute to the responses of the respective COSFIRE filters.

applications. We configure one COSFIRE filter for each of the three traffic signs, Figs. 17d, 17e, 17f. We use a bank of antisymmetric Gabor filters with one wavelength ($\lambda = 4$) and 16 equidistant orientations ($\theta \in \{\frac{\pi i}{8} | i = 0 \dots 15\}$), and threshold their responses with $t_1 = 0.1$. The reference traffic signs that are used to configure the filters and the signs embedded in the complex scenes have approximately the same viewpoint and their sizes differ only by at most 10 percent. For such rigid objects, it is more appropriate to configure COSFIRE filters that achieve high selectivity. With this requirement in mind, we choose to configure the filters with a large number of ρ values ($\rho \in \{0, 2, 4, 7, 10, 13, 16, 20, 25\}$) and a small value of the parameter α ($\alpha = 0.04$) that allows little tolerance in the position of the involved edges.

We use the three COSFIRE filters to detect and recognize the corresponding traffic signs in the entire dataset of

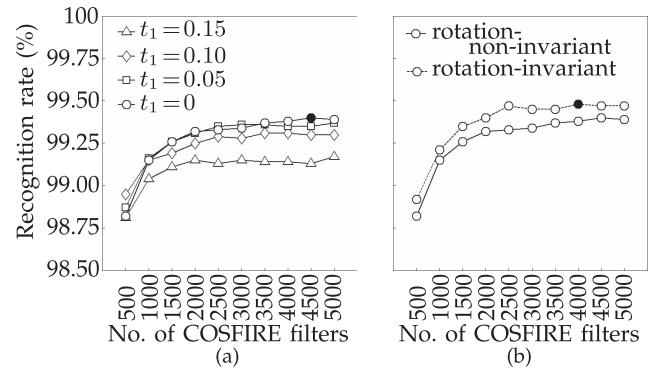


Fig. 16. Experimental results achieved on the MNIST dataset. (a) The four plots show the recognition rates achieved for different values of the threshold t_1 as a function of the number of COSFIRE filters used. The filled-in circle represents the maximum recognition rate of 99.40 percent achieved for $t_1 = 0$ with 4,500 filters. In these experiments, the COSFIRE filters are used in rotation-noninvariant mode. (b) Performance comparison between the same set of COSFIRE filters and $t_1 = 0$ that are first applied in rotation-noninvariant mode and then in a partially rotation-invariant mode. Here partial rotation-invariance is based on five values of the rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The performance improves with partial rotation-invariant filters that achieve a maximum recognition rate of 99.48 percent (shown as a filled-in circle) with 4,000 filters.

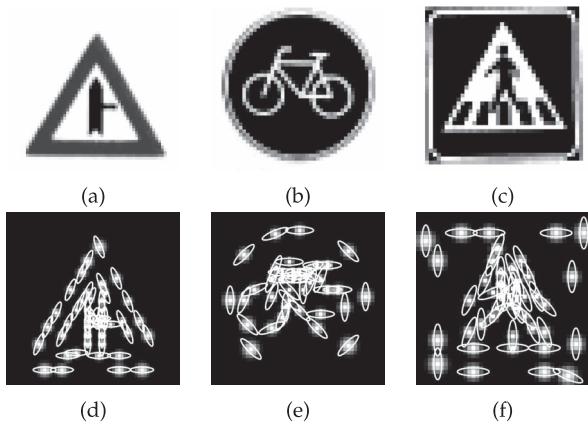


Fig. 17. Three reference traffic signs: (a) an intersection, (b) compulsory giveway for bikes, and (c) a pedestrian crossing. (d)-(f) The structures of the corresponding COSFIRE filters determined by the following parameter values: $\rho \in \{0, 2, 4, 7, 10, 13, 16, 20, 25\}$, $\sigma_0 = 0.67$, $\alpha = 0.04$, $\lambda = 4$, and $\theta \in \{\frac{\pi}{8}|i = 1 \dots 15\}$.

48 images. For each color image, we first convert it to grayscale and subsequently apply the filters. The antisymmetric Gabor filters that we use to provide inputs to the COSFIRE filters are applied with isotropic surround suppression [42] (using an inhibition factor of 2) in order to reduce responses to the presence of texture in these complex scenes. Rather than using the parameter t_3 to threshold the filter responses at a given fraction of the maximum filter response, we choose to threshold the responses at a given absolute value. Moreover, we also threshold responses that are smaller than a fraction of the maximum value of all the responses produced by the three filters. We call this threshold validity ratio. For an absolute threshold of 0.04 and a validity ratio of 0.5, we obtain perfect detection and recognition performance for all the 48 traffic scenes. This means that we detect all the traffic signs in the given images with no false positives and correctly recognize every detected sign. Fig. 18 illustrates the detection and recognition of two different traffic signs, shown encircled, in one of the input images. For this application, we apply the COSFIRE filters in a noninvariant mode ($\psi = 0$, $v = 1$) and compute their output by a weighted geometric mean of the concerned Gabor filter responses ($\sigma' = 21.23$).

4 DISCUSSION

When presenting the method in Section 2, we indicated that a prototype feature used for the configuration of a COSFIRE filter is selected by a user. The detection of vascular bifurcations and the detection and recognition of traffic signs presented in Sections 3.1 and 3.3, respectively, are examples of such applications. The method is, however, not restricted by this aspect: There exists the possibility that a system “discovers” patterns to be used for configuration and Section 3.2 provides an example of such an application.

We use Gabor filters for the detection of lines and edges. Gabor filters, however, are not intrinsic to the proposed method and other orientation-selective filters can also be used.

The configuration of a COSFIRE filter is based on the spatial arrangement of contour parts that lies along concentric circles of given radii around a specified point

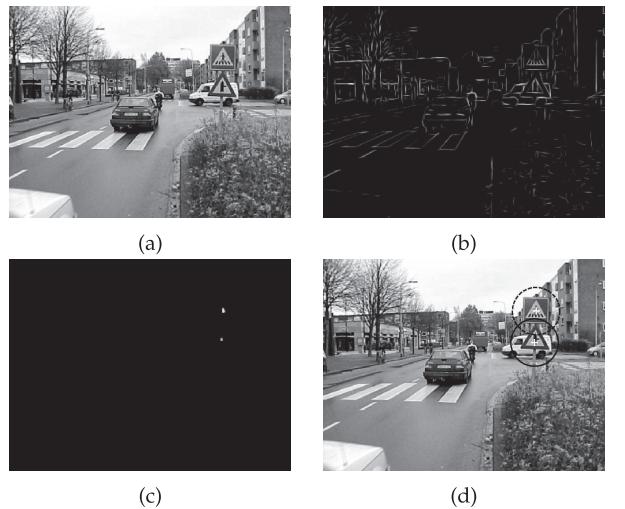


Fig. 18. (a) Input image with filename crossing_004.png. (b) Superposition of thresholded responses ($t_1 = 0.1$) of a bank of antisymmetric Gabor filters ($\lambda = 4$ and $\theta \in \{\frac{\pi}{8}|i = 0 \dots 15\}$) with isotropic surround suppression (inhibition factor is 2). (c) Superposition of the thresholded responses of the three COSFIRE filters. (d) Correct detection and recognition of two traffic signs. The cross markers indicate the locations of the two local maxima responses, each surrounded with a circle that represents the support of the corresponding COSFIRE filter (the continuous circle represents the intersection sign and the dashed circle represents the pedestrian crossing sign).

of interest. In the first two applications that we present we choose to configure COSFIRE filters with three values of the radius parameter ρ as they provide sufficient coverage of the corresponding features. However, for the third application we use nine values of the parameter ρ in order to configure COSFIRE filters that are selective for more complex patterns. The choice of the number of ρ values is related to the size and complexity of the local prototype pattern that is used to configure a filter. The number of ρ values used also controls the tradeoff between the selectivity and generalization ability of a filter: A COSFIRE filter becomes more selective and more discriminative with an increasing number of ρ values.

A COSFIRE filter uses three threshold parameters: t_1 , t_2 , and t_3 . The value of parameter t_1 depends on the contrast of the image material involved in a given application and the presence of noise. It controls the level at which the response of a Gabor filter is supposed to indicate the presence of a line or an edge at a given position. For the first application, which concerns binary input images, we achieved good results for $t_1 = 0.2$. Yet, for the second and third applications, which use grayscale input images, we obtained the best results for $t_1 = 0$ and $t_1 = 0.1$, respectively. The threshold parameter t_2 , which is used only in the configuration phase, is application-independent. It implements a condition that the selected responses are significant and comparable with the strongest possible response. We fix the value of this threshold to ($t_2 = 0.75$). The parameter t_3 is optional. It may be used to suppress the responses of the COSFIRE filter that are below a given fraction of the maximum response value across all locations of the input image. For instance, in the first application we evaluate the performance of the COSFIRE filters with different values of the parameter t_3 , while for the second application we do not threshold ($t_3 = 0$) the responses. In the third application

we threshold the responses at a given absolute value rather than use this threshold parameter.

The proposed COSFIRE filters can be applied in various modes. For the detection of vascular bifurcations in retinal images we applied COSFIRE filters in rotation-, scale-, and reflection-invariant mode, while for the recognition of handwritten digits we only made use of partial rotation invariance and for the detection and recognition of traffic signs in complex scenes we used noninvariant COSFIRE filters.

In the following, we highlight three main aspects in which the proposed COSFIRE filters can be distinguished from other keypoint detectors. First, a COSFIRE filter gives a response only when all parts of the filter-defining prototype feature are present. In contrast, dissimilarity-based approaches also give responses to parts of the prototype pattern. Second, while a COSFIRE filter combines the responses of Gabor filters at different scales, typical scale-invariant approaches, such as SIFT, use the same scale, the one at which the concerned keypoint is an extremum in a given scale space. Third, the area of support of a COSFIRE filter is adaptive. It is composed of the support of a number of orientation-selective filters whose geometrical arrangement around a point of interest is learned from a given local contour prototype pattern. On the contrary, the area of support of other operators is typically related to the appropriate scale rather than to the shape properties of the concerned pattern. To the best of our knowledge the proposed filters are the first ones which combine the responses of orientation-selective filters with their main area of support outside the point of interest. The presence of added noise around a pattern of interest has little or no effect on a COSFIRE filter response. For other operators, any added noise in the surroundings of a pattern of interest results in a descriptor that may differ substantially from the descriptor of the same but noiseless pattern.

The computational cost of the configuration of a COSFIRE filter is proportional to the maximum value of the given set of ρ values and to the size of the bank of Gabor filters used. In practice, for the parameter values that we used in the three applications, a COSFIRE filter is configured in less than half of a second for a Matlab implementation that runs on a 3 GHz processor. The computational cost of the application of a COSFIRE filter depends on the computations of the responses of a bank of Gabor filters and their blurring and shifting. In practice, in the first application a retinal fundus image of size 564×584 pixels is processed in less than 45 seconds on a standard 3 GHz processor by six rotation-, scale-, and reflection-invariant COSFIRE filters. For the second application, a handwritten digit of size 28×28 pixels is described by 5,000 rotation-noninvariant COSFIRE filters in less than 10 seconds on a computer cluster.¹¹ Finally, in the third application, a complex scene of size 360×270 pixels is processed in less than 10 seconds on the same standard 3 GHz processor by three noninvariant COSFIRE filters. For this application we achieve the same performance as

11. We executed the experiments for the MNIST dataset on a computer cluster of 255 multicore nodes (<http://www.rug.nl/cit/hpcv/faciliteiten/HPCCluster/>). We split the MNIST dataset of 70,000 images (60,000 training and 10,000 test digits) into 250 batches of 280 images each, and processed the 250 batches in parallel. In this way, the digit descriptors of one experiment using 5,000 rotation-noninvariant COSFIRE filters takes approximately $(9.5 \text{ seconds} \times 280 \text{ images}) = 45 \text{ minutes}$. An experiment with 5,000 partial rotation-invariant COSFIRE filters (five values of the parameter ψ) takes five times as much.

reported in [41] but with a much lower computational cost. We used Matlab implementation¹² for all the experiments.

The application of the proposed method to the recognition of handwritten digits contains an interesting aspect from a machine learning point of view. In traditional machine learning, the features to be used are fixed in advance and the machine learning aspect concerns the classification of observed feature vectors. If traditional machine learning is concerned with features at all, this is typically limited to the selection of predefined features or using them to derive "new" features as (linear) combinations of the original ones. Examples are principle component analysis and generalized matrix learning vector quantization [43]. Traditional machine learning is typically not concerned with the question of how the original features are defined. This aspect of the problem is, however, crucial for the success: Almost any machine learning method will perform well with good features. The interesting aspect we would like to point out is that in the proposed approach the appropriate prototype features are learned in the filter configuration process when a feature of interest is presented.

In our experiments, we do not analyze the discriminative ability of the individual COSFIRE filters because in this work we are not concerned with the optimization of the filters, but rather with showing their versatility. As a consequence, some of the configured filters that we used for the handwritten digit application might be redundant due to being selective for correlated patterns or for patterns with low distinctiveness. One way of dealing with such redundancy is to compute a dissimilarity measure between the prototype patterns used for the configuration of different COSFIRE filters. Moreover, a prototype feature selection method may also be incorporated in a machine learning algorithm, such as relevance learning vector quantization [44] or a support feature machine [45], to identify the most relevant COSFIRE filters.

The COSFIRE filters that we propose are inspired by the properties of one class of shape-selective neuron in area V4 of visual cortex [15], [16], [46], [47]. The selectivity that is exhibited by a COSFIRE filter which we configured in Section 2 to a dataset of elementary features (Fig. 6) is qualitatively similar to the selectivity of some V4 neurons studied in [15]. The way we determine the standard deviation of the blurring function in (1) is also motivated by neurophysiological evidence that the average diameter of receptive fields¹³ of V4 neurons increases with the eccentricity [48]. Since there is a considerable spread in the behavior across neurons of the concerned type, different computational models may be needed to adequately cover the diversity of functional properties in that empirical space. In this respect, the proposed COSFIRE filter can be considered as a computational model of shape-selective V4 neurons that is complementary to other models [49], [50], [51], [52], [53].

The specific type of function that we use to combine the responses of afferent (Gabor) filters for the considered applications is weighted geometric mean. This output

12. Matlab scripts for the configuration and application of COSFIRE filters can be downloaded from <http://matlabserver.cs.rug.nl/>.

13. In neurophysiology a receptive field refers to an area in the visual field which provides input to a given neuron. Its mathematical counterpart is the support of an operator.

function proved to give better results than various forms of addition. Furthermore, there is psychophysical evidence that human visual processing of shape is likely performed by multiplication [17]. In future work, we plan to experiment with functions other than (weighted) geometric mean.

The proposed COSFIRE filters are particularly useful due to their versatility and selectiveness, in that a COSFIRE filter can be configured by any given local feature and is built on top of other—here orientation-selective—simpler filters. Elsewhere, we have used other types of simple filters (Mexican hat operators) to build a contour operator, which we call Combination of Receptive Fields (CORF) [54]. We use the terms COSFIRE and CORF for the same design principle in an engineering and neuroscience context, respectively.

There are various directions for future research. One direction is to apply the proposed trainable COSFIRE filters in other computer vision tasks, such as geometric stereo calibration, image retrieval, the recognition of handwritten characters, architectural symbols, and pedestrians. Another direction is to enrich the properties of a COSFIRE filter by including information about the color and texture distribution in a given local prototype pattern. A third direction is to extend the proposed approach to 3D COSFIRE filters that can be applied, for instance, to tubular organ registration and bifurcation detection in X-ray computed tomography medical images or to video sequences.

5 CONCLUSIONS

We demonstrated that the proposed COSFIRE filters provide effective machine vision solutions in three practical applications: the detection of vascular bifurcations in retinal fundus images (98.50 percent recall and 96.09 percent precision), the recognition of handwritten digits (99.48 percent correct classification), and the detection and recognition of traffic signs in complex scenes (100 percent recall and precision). For the first application, the proposed COSFIRE filters outperform other methods previously reported in the literature. For the second, it is close to the performance of the best application-specific method. For the third, it gives the same performance as another method which has much higher computational complexity.

The novel COSFIRE filters are conceptually simple and easy to implement: The filter output is computed as the product of blurred and shifted Gabor filter responses. They are versatile detectors of contour related features as they can be trained with any given local contour pattern and are subsequently able to detect identical and similar patterns. The COSFIRE approach is not limited to the combination of Gabor filter responses: More generally, it can be applied to the responses of filters that provide information about texture, color, contours, and motion.

REFERENCES

- [1] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.*, pp. 147-151, 1988,
- [2] C. Schmid and R. Mohr, "Local Grayvalue Invariants for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530-535, May 1997.
- [3] T. Lindeberg, "Feature Detection with Automatic Scale Selection," *Int'l J. Computer Vision*, vol. 30, no. 2, pp. 79-116, 1998.
- [4] D.G. Lowe, "Object Recognition from Local Scale-Invariant Features," *Proc. Seventh IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1150-1157, 1999.
- [5] K. Mikolajczyk and C. Schmid, "Indexing Based on Scale Invariant Interest Points," *Proc. Eighth IEEE Int'l Conf. Computer Vision*, vol. 1, pp. 525-531, 2001.
- [6] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.
- [7] L. Florack, B. ter Haar Romeny, J. Koenderink, and M. Viergever, "General Intensity Transformations and Differential Invariants," *J. Math. Imaging and Vision*, pp. 171-187, 1994.
- [8] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons, "Moment Invariants for Recognition under Changing Viewpoint and Illumination," *Computer Vision and Understanding*, vol. 94, nos. 1-3, pp. 3-27, 2004.
- [9] A. Baumberg, "Reliable Feature Matching Across Widely Separated Views," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 774-781, 2000.
- [10] W. Freeman and E. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, Sept. 1991.
- [11] G. Carneiro and A. Jepson, "Multi-Scale Phase-Based Local Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. I-736-I-743, 2003.
- [12] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l J. Computer Vision*, vol. 60, pp. 91-110, 2004.
- [13] Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. II-506-II-513, 2004.
- [14] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.
- [15] A. Pasupathy and C.E. Connor, "Responses to Contour Features in Macaque Area V4," *J. Neurophysiology*, vol. 82, no. 5, pp. 2490-2502, Nov. 1999.
- [16] A. Pasupathy and C.E. Connor, "Population Coding of Shape in Area V4," *Nature Neuroscience*, vol. 5, no. 12, pp. 1332-1338, 2002.
- [17] E. Gheorghiu and F.A.A. Kingdom, "Multiplication in Curvature Processing," *J. Vision*, vol. 9, no. 2, pp. 23:1-23:17, 2009.
- [18] N. Petkov, "Biologically Motivated Computationally Intensive Approaches to Image Pattern-Recognition," *Future Generation Computer Systems*, vol. 11, nos. 4/5, pp. 451-465, 1995.
- [19] N. Petkov and P. Kruizinga, "Computational Models of Visual Neurons Specialised in the Detection of Periodic and Aperiodic Oriented Visual Stimuli: Bar and Grating Cells," *Biological Cybernetics*, vol. 76, no. 2, pp. 83-96, 1997.
- [20] P. Kruizinga and N. Petkov, "Non-Linear Operator for Oriented Texture," *IEEE Trans. Image Processing*, vol. 8, no. 10, pp. 1395-1407, Oct. 1999.
- [21] S.E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of Texture Features Based on Gabor Filters," *IEEE Trans. Image Processing*, vol. 11, no. 10, pp. 1160-1167, Oct. 2002.
- [22] N. Petkov and M.A. Westenberg, "Suppression of Contour Perception by Band-Limited Noise and Its Relation to Non-Classical Receptive Field Inhibition," *Biological Cybernetics*, vol. 88, no. 10, pp. 236-246, 2003.
- [23] C. Grigorescu, N. Petkov, and M.A. Westenberg, "The Role of Non-CRF Inhibition in Contour Detection," *J. Computer Graphics, Visualization, and Computer Vision*, vol. 11, no. 2, pp. 197-204, 2003.
- [24] C. Grigorescu, N. Petkov, and M.A. Westenberg, "Contour Detection Based on Nonclassical Receptive Field Inhibition," *IEEE Trans. Image Processing*, vol. 12, no. 7, pp. 729-739, July 2003.
- [25] C.D. Murray, "The Physiological Principle of Minimum Work: I. The Vascular System and the Cost of Blood Volume," *Proc. Nat'l Academy of Sciences USA*, vol. 12, pp. 207-214, 1926.
- [26] C.D. Murray, "The Physiological Principle of Minimum Work Applied to the Angle of Branching of Arteries," *J. General Physiology*, vol. 9, pp. 835-841, 1926.
- [27] T. Sherman, "On Connecting Large Vessels to Small—the Meaning of Murray Law," *J. General Physiology*, vol. 78, no. 4, pp. 431-453, 1981.
- [28] M. Zamir, J. Medeiros, and T. Cunningham, "Arterial Bifurcations in the Human Retina," *J. General Physiology*, vol. 74, no. 4, pp. 537-548, 1979.

- [29] M. Tso and L. Jampol, "Path-Physiology of Hypertensive Retinopathy," *Ophthalmology*, vol. 89, no. 10, pp. 1132-1145, 1982.
- [30] N. Chapman, G. Dell'omo, M.S. Sartini, N. Witt, A. Hughes, S. Thom, and R. Pedrinelli, "Peripheral Vascular Disease Is Associated with Abnormal Arteriolar Diameter Relationships at Bifurcations in the Human Retina," *Clinical Science*, vol. 103, no. 2, pp. 111-116, Aug. 2002.
- [31] N. Patton, T.M. Aslam, T. MacGillivray, I.J. Deary, B. Dhillon, R.H. Eikelboom, K. Yugesan, and I.J. Constable, "Retinal Image Analysis: Concepts, Applications and Potential," *Progress in Retinal and Eye Research*, vol. 25, no. 1, pp. 99-127, Jan. 2006.
- [32] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-Based Vessel Segmentation in Color Images of the Retina," *IEEE Trans. Medical Imaging*, vol. 23, no. 4, pp. 501-509, Apr. 2004.
- [33] A. Bhuiyan, B. Nath, J. Chua, and K. Ramamohanarao, "Automatic Detection of Vascular Bifurcations and Crossovers from Color Retinal Fundus Images," *Proc. Third IEEE Int'l Conf. Signal-Image Technologies and Internet-Based System*, pp. 711-718, 2007,
- [34] C. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten Digit Recognition: Benchmarking of State-of-the-Art Techniques," *Pattern Recognition*, vol. 36, no. 10, pp. 2271-2285, 2003.
- [35] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
- [36] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, Apr. 2002.
- [37] D. Oberhoff and M. Kolesnik, "Unsupervised Shape Learning in a Neuromorphic Hierarchy," *Pattern Recognition and Image Analysis*, vol. 18, pp. 314-322, 2008.
- [38] A. Borji, M. Hamidi, and F. Mahmoudi, "Robust Handwritten Character Recognition with Features Inspired by Visual Ventral Stream," *Neural Processing Letters*, vol. 28, no. 2, pp. 97-111, 2008.
- [39] M. Hamidi and A. Borji, "Invariance Analysis of Modified C2 Features: Case Study-Handwritten Digit Recognition," *Machine Vision and Applications*, vol. 21, no. 6, pp. 969-979, 2010.
- [40] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient Learning of Sparse Representations with an Energy-Based Model," *Advances in Neural Information Processing Systems*, J. Platt et al., eds., MIT Press, 2006.
- [41] C. Grigorescu and N. Petkov, "Distance Sets for Shape Filters and Shape Recognition," *IEEE Trans. Image Processing*, vol. 12, no. 10, pp. 1274-1286, Oct. 2003.
- [42] C. Grigorescu, N. Petkov, and M.A. Westenberg, "Contour and Boundary Detection Improved by Surround Suppression of Texture Edges," *Image and Vision Computing*, vol. 22, no. 8, pp. 609-622, Aug. 2004.
- [43] K. Bunte, M. Biehl, M.F. Jonkman, and N. Petkov, "Learning Effective Color Features for Content Based Image Retrieval in Dermatology," *Pattern Recognition*, vol. 44, no. 9, pp. 1892-1902, 2011.
- [44] B. Hammer and T. Villmann, "Generalized Relevance Learning Vector Quantization," *Neural Networks*, vol. 15, nos. 8/9, pp. 1059-1068, 2002.
- [45] S. Klement and T. Martinetz, "The Support Feature Machine for Classifying with the Least Number of Features," *Proc. 20th Int'l Conf. Artificial Neural Networks: Part II*, pp. 88-93, 2010.
- [46] A. Pasupathy and C.E. Connor, "Shape Representation in Area V4: Position-Specific Tuning for Boundary Conformation," *J. Neurophysiology*, vol. 86, no. 5, pp. 2505-2519, Nov. 2001.
- [47] J. Hegde and D.C. Van Essen, "A Comparative Study of Shape Representation in Macaque Visual Areas V2 and V4," *Cerebral Cortex*, vol. 17, no. 5, pp. 1100-1116, 2007.
- [48] R. Gattass, A.P. Sousa, and C.G. Gross, "Visuotopic Organization and Extent of V3 and V4 of the Macaque," *J. Neuroscience*, vol. 8, no. 6, pp. 1831-1845, 1988.
- [49] M. Riesenhuber and T. Poggio, "Hierarchical Models of Object Recognition in Cortex," *Nature Neuroscience*, vol. 2, no. 11, pp. 1019-1025, Nov. 1999.
- [50] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, "A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex," AI Memo 2005-036/CBCL Memo 259, Massachusetts Inst. of Technology, 2005.
- [51] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust Object Recognition with Cortex-Like Mechanisms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411-426, Mar. 2007.
- [52] C. Cadieu, M. Kouh, A. Pasupathy, C.E. Connor, M. Riesenhuber, and T. Poggio, "A Model of V4 Shape Selectivity and Invariance," *J. Neurophysiology*, vol. 98, no. 3, pp. 1733-1750, Sept. 2007.
- [53] S. Fidler and A. Leonardis, "Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [54] G. Azzopardi and N. Petkov, "A CORF Computational Model of a Simple Cell That Relies on LGN Input Outperforms the Gabor Function Model," *Biological Cybernetics*, vol. 106, pp. 177-189, 2011, doi: 10.1007/s00422-012-0486-6.



George Azzopardi received the BSc degree with honours (first class) in computer science from Goldsmiths University of London in 2006 and was awarded an academic award. In 2007, he was awarded a government scholarship in order to pursue a master's degree in advanced methods of computer science at Queen Mary University of London, where he graduated with distinction (ranked first) in 2008. Currently, he is working toward the PhD degree at the Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands. His current research interests include brain-inspired machine vision, which includes computational models of the visual system with applications to contour detection, feature, and shape recognition.



Nicolai Petkov received the Dr.sc.techn. degree in computer engineering (Informationstechnik) from Dresden University of Technology, Germany. He is a professor of computer science and head of the Intelligent Systems Group of the Johann Bernoulli Institute of Mathematics and Computer Science of the University of Groningen, The Netherlands. He is the author of two monographs and coauthor of another book on parallel computing, holds four patents, and has authored more than 100 scientific papers. His current research is in image processing, computer vision, and pattern recognition, and includes computer simulations of the visual system of the brain, brain-inspired computing, computer applications in health care and life sciences, and creating computer programs for artistic expression. He is a member of the editorial boards of several journals.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.