Brain-inspired computer vision with applications to pattern recognition and computer-aided diagnosis of glaucoma

Guo, Jiapan

Publication date:
2017

Chapter 2

Inhibition-augmented Trainable COSFIRE Filters for Shape Detection and Object Recognition

Abstract

The shape and meaning of an object can radically change with the addition of one or more contour parts. For instance, a T-junction can become a crossover. We extend the COSFIRE trainable filter approach which uses a positive prototype pattern for configuration by adding a set of negative prototype patterns. The configured filter responds to patterns that are similar to the positive prototype but not to any of the negative prototypes. The configuration of such a filter comprises selecting given channels of a bank of Gabor filters that provide excitatory or inhibitory input and determining certain blur and shift parameters. We compute the response of such a filter as the excitatory input minus a fraction of the maximum of inhibitory inputs. We use three applications to demonstrate the effectiveness of inhibition: the exclusive detection of vascular bifurcations (i.e. without crossovers) in retinal fundus images (DRIVE data set), the recognition and localization of architectural and electrical symbols (GREC'11 data set) and the recognition of handwritten digits (MNIST data set).

---

[1]Subsection 'Localization of architectural and electrical symbols' in Section 2.3.2 is taken from this paper. The rest of this chapter is based on the first paper given above.

2.1   Introduction

Recently, a novel trainable filter for object recognition has been proposed in (Azzopardi and Petkov, 2013b). It is called combination of shifted filter responses or COSFIRE for brevity. A COSFIRE filter is configured to be selective for a given local pattern by extracting from that pattern characteristic properties of contour parts (such as orientation) and their geometrical arrangement. COSFIRE filters were demonstrated to be effective for detection of local patterns and recognition of objects and achieve very good performance in various applications (Azzopardi and Petkov, 2013a, 2014; Azzopardi et al., 2015; Guo, Shi, Azzopardi and Petkov, 2015a; Strisciuglio et al., 2015; Shi et al., 2015a; Strisciuglio, Vento, Azzopardi and Petkov, 2016). They were also used in a multi-layer hierarchical approach (Azzopardi and Petkov, 2014).

| (a) | (b) | (c) | (d) | (e) |
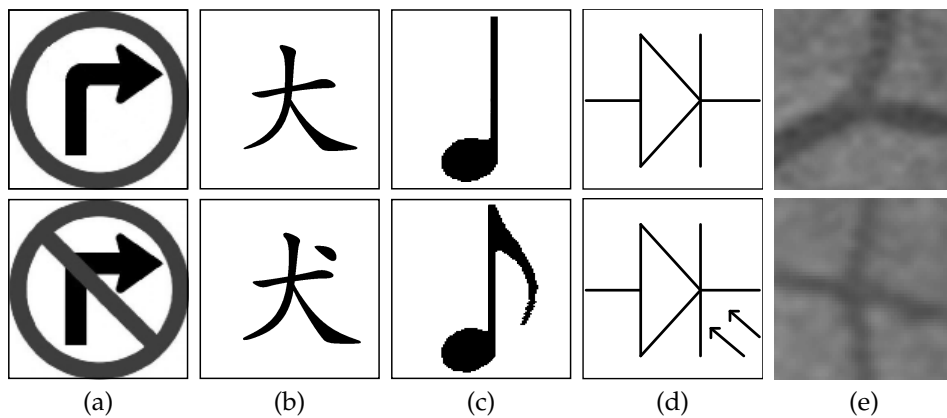
Figure 2.1: Examples of pairs of patterns that COSFIRE filters of the type proposed in (Azzopardi and Petkov, 2013b) are not able to distinguish. (a) Two traffic signs which have the opposite messages: permission and prohibition of turning right. (b) Two Chinese characters that are translated into English as "big" and "dog". (c) Two music notes: quarter and eighth. (d) Two electrical symbols: normal and photo-diodes. (e) Bifurcations and crossovers in retinal fundus images. A COSFIRE filter that is trained to detect the upper pattern in a pair of two, will give a strong response to the lower pattern too.

Fig. 2.1 shows some examples where a COSFIRE filter of the type proposed in (Azzopardi and Petkov, 2013b) may, however, not perform very well. COSFIRE filters that are configured to be selective for the patterns shown in the images in the top row of Fig. 2.1 also give strong responses to the images in the bottom row of Fig. 2.1. This is because all contour parts of a pattern in the top row are present, in the preferred arrangements, in the corresponding image shown in the bottom row

of Fig. 2.1. The presence of additional contour parts, such as the diagonal bar in Fig. 2.1a (bottom) or the extra stroke in Fig. 2.1b (bottom), do not have influence on the response of the filter.

The COSFIRE method (Azzopardi and Petkov, 2013b) was inspired by a specific type of shape-selective neuron in area V4 of visual cortex. This method, however, relies on contour parts that provide only excitatory inputs. This means that every involved contour part detector contributes to enhance the response of a COSFIRE filter.

There is neurophysiological evidence, however, that neurons in different layers of the visual cortex receive also inhibitory inputs (Hubel, 1988). For instance, neurons in the lateral geniculate nucleus (LGN) have center-surround receptive fields which have been modeled by difference-of-Gaussians (DoG) operators. A center-on DoG has an excitatory central region with an inhibitory surround. Similarly, simple cells in area V1, whose properties provided the inspiration for Gabor filters (Daugman, 1985; Jones and Palmer, 1987), derivative of Gaussians (Florack et al., 1994) and CORF (Azzopardi and Petkov, 2012; Azzopardi et al., 2014) filters, have receptive fields that consist of inhibitory and excitatory regions. Non-classical receptive field inhibition in orientation-selective visual neurons provided the inspiration for surround inhibition in orientation-selective filters (Petkov and Visser, 2005). It has been shown to improve contour detection by suppressing responses to textured regions. Moreover, shape-selective neurons of the type studied in (Brincat and Connor, 2004), located in the posterior inferotemporal cortex, respond to complex shapes that are formed by a number of convex and concave curvatures with a certain geometrical arrangement. The presence of some specific curvature elements can inhibit the response of such a neuron. Fig. 2.2 shows the response of a TEO neuron, studied in (Brincat and Connor, 2004), which is excited by the encircled curvatures A, B and C but is inhibited by the dashed encircled curvature D. The bar plots indicate the responses to the stimuli. Inhibition is also thought to increase the selectivity of neurons (Sami and Mriganka, 2014).

Inhibition is an important phenomenon in the brain. It facilitates sparseness in the representation of information that may result in an increase of the storage capacity and a higher number of patterns that can be discriminated (Rolls and Treves, 1990). End-stopped cells (Hubel and Wiesel, 1968; Bolz and Gilbert, 1986) in area V1 of visual cortex are another example.

In this work, we add inhibition to COSFIRE filters in order to increase their discrimination ability. The inhibition that we propose is learned in an automatic configuration process. We configure an inhibition-augmented COSFIRE filter by using two different types of prototype patterns, namely one positive pattern and one or more negative pattern(s), in order to extract excitatory and inhibitory contour parts,
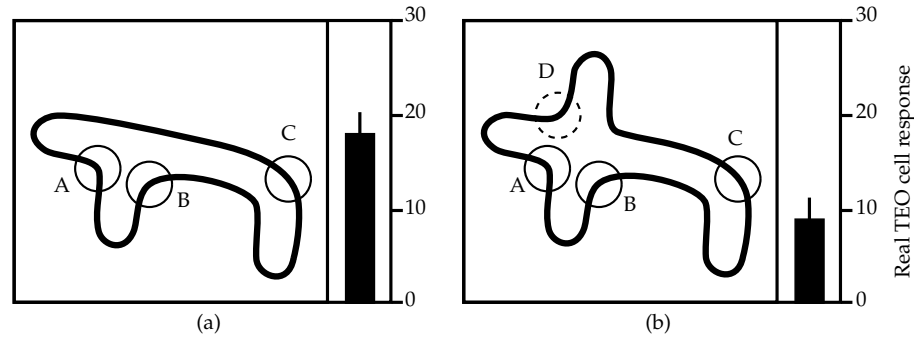
Figure 2.2: Selectivity of a shape-selective neuron in the posterior inferotemporal cortex (Brincat and Connor, 2004). (a) The curvatures marked with circles evoke excitation of the concerned cell, while (b) the curvature marked with a dashed circle inhibits the activation of the cell. The bars specify the strength of the response.

respectively. Such a filter can effectively detect patterns that are equivalent or similar to the positive prototype but does not respond to the negative prototype(s).

The proposed inhibition-augmented filters can be used in shape detection and object recognition. A large body of work has been done in these areas and many methods have been proposed (Beaudet, 1978; Julesz, 1981; Harris and Stephens, 1988; Lowe, 1999; Mikolajczyk and Schmid, 2001; Oliva and Torralba, 2001; Ojala et al., 2002; Lowe, 2004; Viola and Michael, 2004; Dalal and Triggs, 2005; Mikolajczyk and Schmid, 2005; Lazebnik et al., 2005; Zhang et al., 2007; H. Bay and Gool, 2008). The Hessian detector (Beaudet, 1978) and the Harris detector (Harris and Stephens, 1988), for instance, detect points of interest and are invariant to rotation but not so much to scale. Scaling-invariances of these two operators can be achieved by applying them in a Laplacian of Gaussian scale space (Lowe, 1999), resulting in the so-called Hessian-Laplace and Harris-Laplace detectors (Mikolajczyk and Schmid, 2001). A point of interest can be described by some local keypoint descriptors, such as the scale-invariant feature transform (SIFT) (Lowe, 2004), the histogram of oriented gradients (HOG) (Dalal and Triggs, 2005), the image descriptor GIST (Oliva and Torralba, 2001) and the gradient location and orientation histogram (GLOH) (Mikolajczyk and Schmid, 2005). Other keypoint descriptors include the speeded up robust features (SURF) (H. Bay and Gool, 2008), which is akin to SIFT but faster as it makes efficient use of integral images (Viola and Michael, 2004), the texture-based local binary patterns (LBP) (Ojala et al., 2002), textons (Julesz, 1981; Zhang et al., 2007), and the biologically inspired local descriptor (BILD) (Zhang et al., 2014), as well as the rotation invariant feature transform (RIFT) descriptor (Lazebnik et al., 2005). None of these methods employs inhibition.

Multiple keypoints can be used to represent bigger and more complex patterns, such as complete objects or scenes. In (Li and Perona, 2005), a bag-of-visual-words approach was proposed to describe an image or a region of interest with a histogram of prototypical keypoints. This method is improved by using spatial pyramids (Lazebnik et al., 2006) or a random sample consensus algorithm (Kalantidis et al., 2011). Other object recognition approaches use hierarchical representations of objects, which have been inspired by the visual processing in the brain. These include the HMAX model (Riesenhuber and Poggio, 1999), the object representation by parts proposed in (Fidler et al., 2006), neural networks (Krizhevsky et al., 2012), and the deep learning approach (LeCun et al., 2015).

These methods require many training examples to configure models of objects of interest. When such detectors and descriptors are trained, only positive examples are considered without the inclusion of inhibition mechanisms. The resulting detectors and descriptors can detect objects that are similar to positive examples but may also give strong responses to objects that contain additional contour parts. For instance, the detectors, which are trained by examples shown in the top row of Fig. 2.1, will give strong responses to objects that are equivalent or similar to the ones shown in the top row of Fig. 2.1. They will, however, also give strong responses to objects that are equivalent or similar to the ones in the bottom row of Fig. 2.1. Therefore, it is difficult for these methods to discriminate the pairs of patterns shown in Fig. 2.1(a-f).

The rest of the chapter is organized as follows. In Section 2.2, we explain how an inhibition-augmented filter is configured by given positive and negative prototype patterns. In Section 2.3, we demonstrate the effectiveness of the proposed approach in two applications. In Section 2.4 we discuss some aspects of the proposed method and finally we draw conclusions in Section 2.5.

2.2 Method

2.2.1 Overview

Fig. 2.3a shows an input image containing a rectangle with a vertical line inside it. Let us consider the two local patterns encircled by a solid and a dashed line, which are shown enlarged in Fig. 2.3b and Fig. 2.3c, respectively. The two solid ellipses in Fig. 2.3b and Fig. 2.3c surround a line segment that is present in both patterns, while the dashed ellipse surrounds a line segment that is only present in Fig. 2.3c. We use these two patterns to configure an inhibition-augmented filter that will respond to the pattern shown in Fig. 2.3b, a line-ending, but not to the pattern shown in Fig. 2.3c, a continuous line.
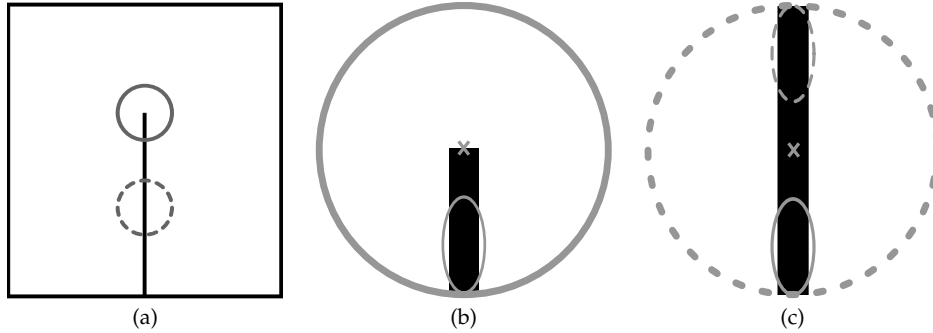
Figure 2.3: Example of a positive and a negative prototype that are used for the configuration of the inhibition-augmented COSFIRE filter. (a) Synthetic input image (of size 300×300 pixels). The solid circle indicates a positive prototype of interest (a line ending) while the dashed circle indicates a negative prototype of interest (a continuous line segment). The images in (b) and (c) show enlargements of the selected positive and negative prototype patterns, respectively. The gray crosses in (b) and (c) indicate the center positions of interest and the ellipses illustrate the orientation and location of the contour parts in the neighborhoods. The solid ellipses represent line segments that are present in both prototypes, while the dashed ellipse represents a line segment which is only present in the negative prototype.

We consider the line-ending and the continuous line shown in Fig. 2.3b and Fig. 2.3c as a positive and a negative prototype, respectively. A positive prototype is a local pattern to which the inhibition-augmented filter to be configured should respond, while a negative prototype is a local pattern to which it should not respond.

We use the positive and the negative prototypes to configure two COSFIRE filters with the method proposed in (Azzopardi and Petkov, 2013b). Next, we look for and identify pairs of contour parts with identical properties in the two filters. In Fig. 2.3 we use a solid ellipse to indicate that the corresponding contour part is an excitatory feature. We use a dashed ellipse to indicate the contour part that is only present in the negative prototype, and therefore we consider it as an inhibitory feature.

The response of the inhibition-augmented filter is the difference between the *excitatory input* and a fraction of the maximum of the *inhibitory inputs*. The resulting filter will only respond to the patterns that are identical with or similar to the positive prototype but will not respond to images similar to any of the negative prototypes. This design decision is inspired by the function of a type of shape-selective neuron in posterior inferotemporal cortex.

In the next sub-sections we elaborate further on the configuration steps mentioned above.
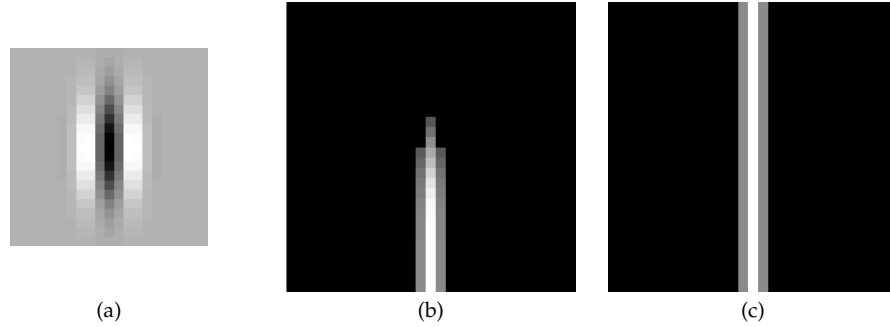
$$(a) \qquad\qquad (b) \qquad\qquad (c)$$

Figure 2.4: Gabor filter and its responses. (a) Intensity map (of size 21×21 pixels) of a symmetric Gabor function with wavelength $\lambda = 6$ and orientation $\theta = 0$. Light and dark regions correspond to positive and negative values of the Gabor function, respectively. (b-c) The thresholded (at $t_1 = 0.2$) Gabor response images (of size 30×30 pixels) to Fig. 2.3b and Fig. 2.3c, respectively.

2.2.2 Gabor Filters

The proposed inhibition-augmented filter uses as input the responses of 2D Gabor filters (Daugman, 1985) which are the generalization of 1D Gabor functions proposed by (Gabor, 1946). We denote by $g_{\lambda,\theta}(x, y)$ the response of a Gabor filter, which has a preferred wavelength λ and orientation θ, to a given input image at location (x, y). We threshold the responses of Gabor filters at a given fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response across all combinations of values (λ, θ) and all positions (x, y) in the image. We denote these thresholded response images by $|g_{\lambda,\theta}(x, y)|_{t_1}$. Fig. 2.4a shows the intensity map of a Gabor function with a wavelength $\lambda = 6$ and an orientation $\theta = 0$. Fig. 2.4b and Fig. 2.4c are the corresponding thresholded response images of this Gabor filter $|g_{6,0}(x, y)|_{t_1=0.2}$ to the input images in Fig. 2.3b and Fig. 2.3c, respectively. Such a filter has other parameters, including spatial aspect ratio, bandwidth and phase offset on which we do not elaborate further here. We refer the interested reader to (Petkov, 1997; Kruizinga, 1999; Azzopardi and Petkov, 2013b) for technical details and to an online implementation[2].

2.2.3 Configuration of an Inhibition-augmented COSFIRE Filter

The configuration of an inhibition-augmented filter involves two steps.

In the first step we configure two separate COSFIRE filters with the method proposed in (Azzopardi and Petkov, 2013b) to be selective for the specified positive and negative prototypes that are shown in Fig. 2.3b and Fig. 2.3c, respectively. Fig. 2.5a

---

[2]`http://matlabserver.cs.rug.nl`

and Fig. 2.5b show the corresponding superimposed thresholded responses of a bank of Gabor filters ($\theta \in \{0, \pi/8, \ldots 7\pi/8\}$ and $\lambda \in \{4, 4\sqrt{2}, 6, 6\sqrt{2}\}$) to the positive and negative prototypes. In this example, for the configuration of a COSFIRE filter with a given prototype, we consider the Gabor responses along two concentric circles with radii $\rho \in \{5, 14\}$ pixels around the specified point of interest. In Fig. 2.5c and Fig. 2.5d we illustrate the structures of the resulting selected filters. The size and orientation of an ellipse represent the preferred wavelength λ and orientation θ of a Gabor filter that provides input to the COSFIRE filter. The position of its center indicates the location at which we take the concerned Gabor filter response.

We specify a COSFIRE filter by a set of 4-tuples in which each 4-tuple represnts a Gabor filter and the positions at which its response has to be taken. We denote by P_f and N_f the two COSFIRE filters, configured with the patterns shown in Fig. 2.3b and Fig. 2.3c, respectively:

$$P_f = \left\{ \begin{array}{l} (\ \lambda_1 = 6,\ \theta_1 = 0,\ \rho_1 = 5,\quad \varphi_1 = 3\pi/2\ ), \\ (\ \lambda_2 = 6,\ \theta_2 = 0,\ \rho_2 = 14,\ \varphi_2 = 3\pi/2\ ) \end{array} \right\}$$

and

$$N_f = \left\{ \begin{array}{l} (\ \lambda_1 = 6,\ \theta_1 = 0,\ \rho_1 = 5,\quad \varphi_1 = \pi/2\quad), \\ (\ \lambda_2 = 6,\ \theta_2 = 0,\ \rho_2 = 5,\quad \varphi_2 = 3\pi/2\ ), \\ (\ \lambda_3 = 6,\ \theta_3 = 0,\ \rho_3 = 14,\ \varphi_3 = \pi/2\quad), \\ (\ \lambda_4 = 6,\ \theta_4 = 0,\ \rho_4 = 14,\ \varphi_4 = 3\pi/2\ ) \end{array} \right\}$$

In the second step we form a new set S_f by selecting tuples from the sets P_f and N_f as follows. We include all tuples from the set P_f in the new set S_f and add a new parameter $\delta = +1$ to indicate that the corresponding Gabor responses of such tuples provide excitatory input to the inhibition-augmented filter. We define a dissimilarity function, which we denote by $d(P_f^i, N_f^j)$, of the distance between the locations indicated by the i-th tuple in the set P_f and the j-th tuple in the set N_f:

$$d(P_f^i, N_f^j) = \left\{ \begin{array}{ll} 1, & \text{if } D > \zeta \\ 0, & \text{otherwise} \end{array} \right. \tag{2.1}$$

$$D = \sqrt{(\rho_i \cos \varphi_i - \rho_j \cos \varphi_j)^2 + (\rho_i \sin \varphi_i - \rho_j \sin \varphi_j)^2}$$

where D is the Euclidean distance between the polar coordinates (ρ_i, φ_i) of tuple i in the positive set P_f and the polar coordinates (ρ_j, φ_j) of tuple j in the negative set N_f. ζ is the threshold and we provide further details on the selection of its value in 2.3.

We compute the pairwise dissimilarity values between one tuple N_f^j from N_f
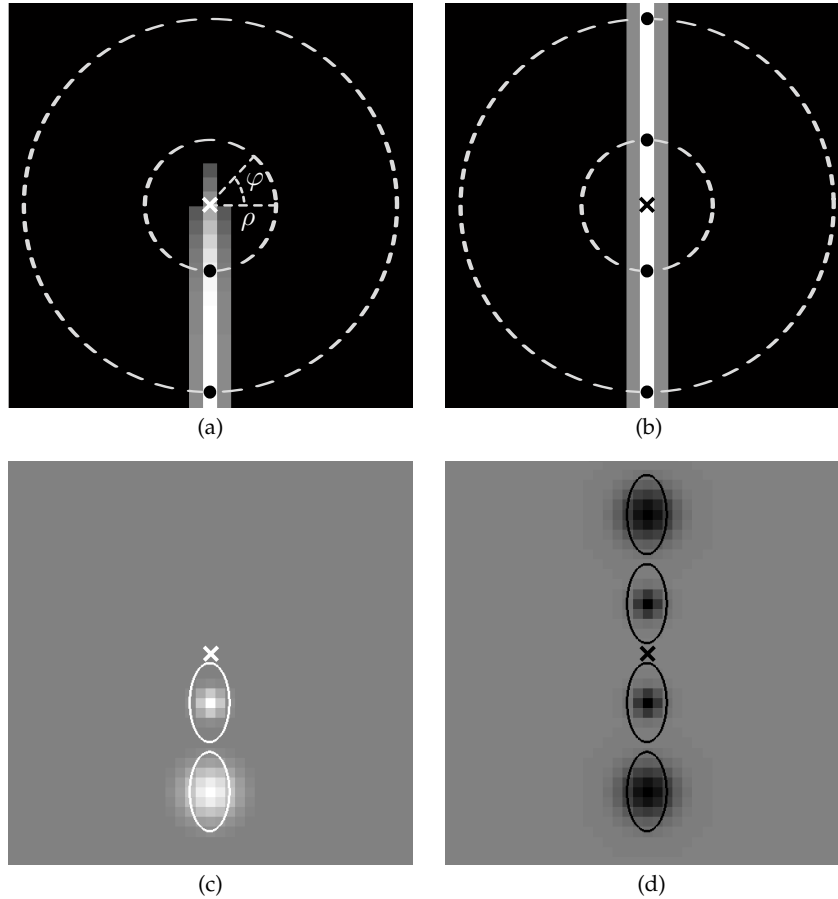
Figure 2.5: Configuration of two COSFIRE filters as proposed in (Azzopardi and Petkov, 2013b). Thresholded responses of a symmetric Gabor filter with $\lambda = 6$ and $\theta = 0$ to (a) the positive and (b) the negative prototypes shown in Fig. 2.3b and Fig. 2.3c, respectively. The cross markers indicate the centers of the filter. We consider the dashed circles of given radii (here $\rho \in \{5, 14\}$) around the center of the pattern of interest. The black dots indicate the positions of the local maxima of the Gabor responses along these circles. In each such point, we select the Gabor filter which gives this local maximum response. (c-d) Structure of the resulting two COSFIRE filters. The ellipses illustrate the wavelengths and orientations of the selected Gabor filters and their positions indicate the locations at which the responses of these Gabor filters are taken with respect to the center. The blobs represent the blurring functions that are used to provide some spatial tolerance to these positions.

and all tuples from P_f. If N_f^j is dissimilar to all tuples in P_f, we include it to the new set S_f and add a tag $\delta = -1$, which indicates that the corresponding Gabor re-

sponse provides an inhibitory input. We repeat the above procedure for each tuple in set N_f. With this process we ensure that a line segment that is present in both the positive and the negative prototypes in roughly the same position gives an excitatory input. On the other hand, a line segment that is only present in the negative prototype, i.e. it does not overlap with a line segment in the positive prototype, provides an inhibitory input.

For the above example, we include the two tuples in set P_f, which are illustrated by the two ellipses in Fig. 2.5c, in the new set S_f. We add to each of these two tuples a tag $\delta = +1$ to indicate that they provide excitatory input to the inhibition-augmented filter. These two tuples are also present in set N_f. Then we include in S_f the other two tuples from N_f indicated by the two ellipses at the top of Fig. 2.5d with a tag $\delta = -1$ as we do not find any matches in P_f. For the above example this method results in the following set S_f:

$$
S_f = \left\{
\begin{array}{l}
(\lambda_1 = 6,\ \theta_1 = 0,\ \rho_1 = 5,\quad \varphi_1 = 3\pi/2,\ \delta_1 = +1\ ), \\
(\lambda_2 = 6,\ \theta_2 = 0,\ \rho_2 = 14,\ \varphi_2 = 3\pi/2,\ \delta_2 = +1\ ), \\
(\lambda_3 = 6,\ \theta_3 = 0,\ \rho_3 = 5,\quad \varphi_3 = \pi/2,\quad \delta_3 = -1\ ), \\
(\lambda_4 = 6,\ \theta_4 = 0,\ \rho_4 = 14,\ \varphi_4 = \pi/2,\quad \delta_4 = -1\ )
\end{array}
\right\}
$$

Fig. 2.6 shows the structure of the resulting inhibition-augmented filter that is represented by the set S_f. The red ellipses indicate Gabor filters that provide excitatory input and the blue ellipses indicate Gabor filters that provide inhibitory input to the inhibition-augmented filter at hand.

For example, the second tuple in S_f ($\lambda_2 = 6, \theta_2 = 0, \rho_2 = 14, \varphi_2 = 3\pi/2, \delta_2 = +1$) corresponds to the ellipse in the bottommost of Fig. 2.6. It describes a line segment with a width of ($\lambda_2/2 =$) 3 pixels in a vertical ($\theta_2 = 0$) orientation at a position of ($\rho_2 =$) 14 pixels to the bottom ($\varphi_2 = 3\pi/2$) of the point of interest. This tuple provides excitatory ($\delta_2 = +1$) input to the inhibition-augmented filter. On the other hand, the last tuple in S_f ($\lambda_4 = 6, \theta_4 = 0, \rho_4 = 14, \varphi_4 = \pi/2, \delta_4 = -1$) corresponds to the topmost ellipse in Fig. 2.6. It describes a similar line segment at a position of ($\rho_2 =$) 14 pixels to the top ($\varphi_4 = \pi/2$) of the point of interest and provides inhibitory ($\delta_4 = -1$) input to the filter.

2.2.4   Configuration with Multiple Negative Prototypes

In the above example, we configured an inhibition-augmented filter to be selective for line-endings by using one positive and one negative prototype pattern. In practice, however, a positive pattern may be contained within multiple other patterns, and thus we may need multiple negative examples.
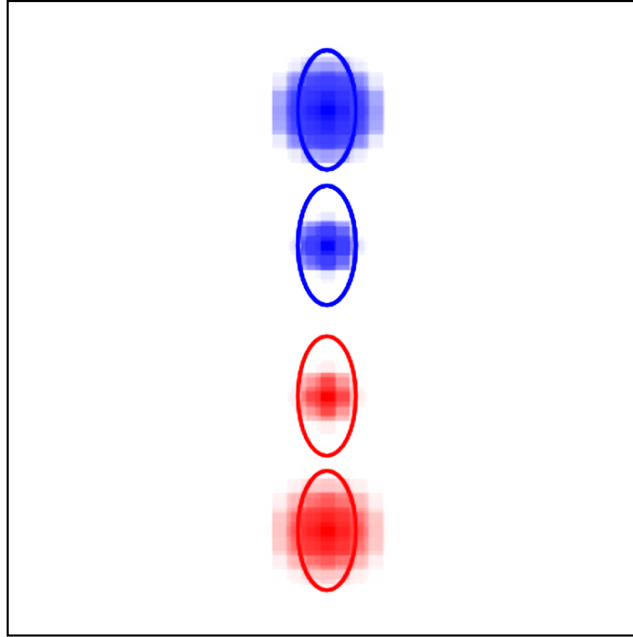
Figure 2.6: The structure of an inhibition-augmented filter. The four ellipses indicate the responses of four Gabor filters with the parameter values specified by the set S_f. The two red ellipses at the bottom represent the excitatory input to this inhibition-augmented filter, while the two blue ellipses at the top represent the inhibitory input.

Fig. 2.7(a-c) shows an example of three similar Chinese letters that have completely different meanings and are translated into English as "big", "dog" and "extremely", respectively. The letter in Fig. 2.7a is also present in Fig. 2.7b and in Fig. 2.7c but accompanied with additional strokes. Next we demonstrate how we configure an inhibition-augmented filter with more than one negative prototype pattern. Here, we use the letter image in Fig. 2.7a as our positive pattern of interest from which we extract contour parts that provide excitatory input to the resulting filter. The letter images in Fig. 2.7b and Fig. 2.7c are used as negative prototype patterns from which we determine inhibitory contour parts.

First we configure a filter P_f for the positive prototype pattern in Fig. 2.7a as proposed in (Azzopardi and Petkov, 2013b) that results in only excitatory inputs. For this example, we consider three values of the radius ρ ($\rho = \{0, 15, 33\}$) and we apply a bank of Gabor filters with four wavelengths ($\lambda \in \{8, 8\sqrt{2}, 16\}$) and eight orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0...7\}$). Then we use the procedure proposed in (Azzopardi and Petkov, 2013b) to apply the filter P_f to both the negative prototype patterns in
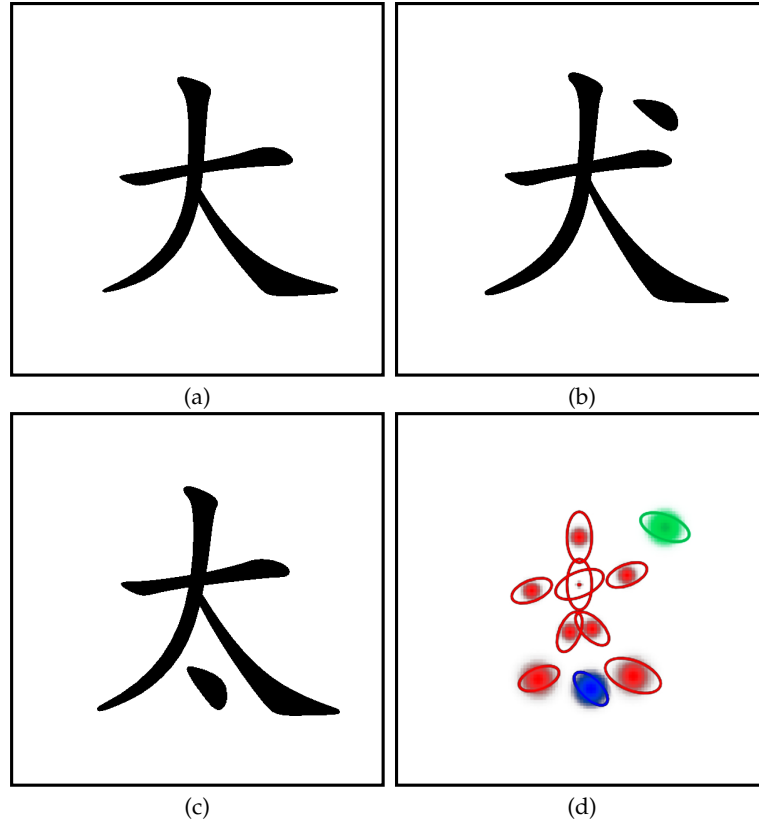
Figure 2.7: The inhibition-augmented COSFIRE filter configured with multiple negative prototypes. (a-c) Images of three similar Chinese letters (of size 120×120 pixels). (d) Structure of the resulting filter S_{big} that is configured by using (a) the positive prorotype and (b-c) the two negative prototypes. The red ellipses represent the preferred wavelengths and orientations of the Gabor filter responses that provide excitatory input to the concerned filter S_{big}, while the blue and green ellipses represent the Gabor filter responses provide inhibitory input.

Fig. 2.7(b-c). For each negative pattern we determine the location at which the maximum response is achieved by the filter P_f. We take the patterns from Fig. 2.7(b-c) that surround these locations and use them to configure two COSFIRE filters, which we denote by N_{f_1} and N_{f_2}, respectively. Finally, we form a new set S_{big} by selecting appropriate tuples from P_f, N_{f_1} and N_{f_2} as follows. We include all tuples from set P_f in the new set S_{big} with a tag $\delta = +1$ and compute the dissimilarity values between the locations of the tuples in N_{f_i} (here $i = 1, 2$) and those in set P_f by the method described in Section 2.3. The tuples in N_{f_1} and N_{f_2} that are not similar to

any of the tuples in P_f are added to S_{big} and marked as inhibitory parts with tags $\delta = -1$ and $\delta = -2$, respectively. These two different negative tags indicate that inhibitory contour parts are extracted from two separate negative patterns.

$$S_{\text{big}} = \left\{ \begin{array}{l} (\lambda_1 = 10, \theta_1 = \frac{\pi}{4}, \rho_1 = 15, \varphi_1 = \frac{25\pi}{16}, \delta_1 = +1), \\ \cdots \\ (\lambda_9 = 14, \theta_9 = \frac{3\pi}{8}, \rho_9 = 33, \varphi_9 = \frac{27\pi}{16}, \delta_9 = +1), \\ (\lambda_{10} = 12, \theta_{10} = \frac{3\pi}{8}, \rho_{10} = 33, \varphi_{10} = \frac{3\pi}{16}, \delta_{10} = -1), \\ (\lambda_{11} = 10, \theta_{11} = \frac{\pi}{4}, \rho_{11} = 33, \varphi_{11} = \frac{3\pi}{2}, \delta_{11} = -2) \end{array} \right\}$$

Fig. 2.7d shows the resulting structure of the inhibition-augmented filter S_{big}, in which the red ellipses indicate the tuples of the filter that provide excitatory input to the inhibition-augmented filter, while the blue and green ellipses indicate the tuples that provide inhibitory input.

2.2.5 Implementation

In the following we first explain how we blur and shift the responses of the involved Gabor filters, and then we describe the functions that we use to compute the collective excitatory input, the various collections of inhibitory inputs, and the ultimate filter output.

Blurring and Shifting Gabor Filter Responses

We blur the Gabor filter responses in order to allow for some tolerance in the positions at which their responses are taken. We define the blurring operation as the weighted maximum of local Gabor filter responses. 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:

$$\sigma = \sigma_0 + \alpha\rho \tag{2.2}$$

where σ_0 and α are constants. The choice of the linear function in Eq. 2.2 is advocated for more detail in (Azzopardi and Petkov, 2013b). For $\alpha > 0$, the tolerance to the positions of the considered contour parts increases with an increasing distance ρ from the center of the concerned COSFIRE filter. We use values of α between 0 and 2, depending on the aplication.

Then we shift all blurred Gabor filter responses so that they meet at the support center of the inhibition-augmented filter. This is achieved by shifting the blurred responses of a 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 $(\rho_i, \varphi_i + \pi)$. In Cartesian coordinates, it is $(\Delta x_i, \Delta y_i)$ where $\Delta x_i = -\rho_i \cos \varphi_i$, and $\Delta y_i = -\rho_i \sin \varphi_i$. We denote by $s_{\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i}(x, y)$ the blurred and shifted thresholded response of a Gabor filter in position (x, y) that is specified by the i-th tuple $(\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i)$ in the set S_f:

$$s_{\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i}(x, y) \overset{\text{def}}{=}$$
$$\mathbf{max}_{x', y'} \{ |g_{\lambda_i, \theta_i}(x - x' - \Delta x_i, y - y' - \Delta y_i)|_{t_1} G_\sigma(x', y') \} \tag{2.3}$$

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

In order to prevent interference of inhibitory and excitatory parts of the filter, we restrict ζ (in Eq. 1) to be three times the maximum standard deviation of any pair of tuples in P_f and N_f.

Response of an inhibition-augmented COSFIRE filter

We denote by $r_{S_f}(x, y)$ the response of an inhibition-augmented COSFIRE filter which we define as the difference between excitatory response $r_{S_f^+}(x, y)$ and a fraction of the maximum of the inhibitory responses $r_{S_f^{-j}}(x, y)$.

$$r_{S_f}(x, y) \overset{\text{def}}{=} |r_{S_f^+}(x, y) - \eta \, \mathbf{max}_{j=1}^{n} \{r_{S_f^{-j}}(x, y)\}|_{t_3} \tag{2.4}$$

where $S_f^+ = \{(\lambda_i, \theta_i, \rho_i, \varphi_i) \mid \forall (\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i) \in S_f, \delta_i = +1\}$, $S_f^{-j} = \{(\lambda_i, \theta_i, \rho_i, \varphi_i) \mid \forall (\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i) \in S_f, \delta_i = -j\}$, $n = \mathbf{max} |\delta_i|$, η is a coefficient that we call inhibition factor and $|.|_{t_3}$ stands for thresholding the response at a fraction t_3 of its maximum across all image coordinates (x, y).

We denote by $r_{S_f^+}$ and $r_{S_f^-}$, the weighted geometric means of all the blurred and shifted responses of the Gabor filters $s_{\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i}(x, y)$ that correspond to the contour parts described by S_f^+ and S_f^{-j}:

$$r_{S_f^\delta}(x, y) \overset{\text{def}}{=} \left| \left(\prod_{i=1}^{|S_f^\delta|} (s_{\lambda_i, \theta_i, \rho_i, \varphi_i, \delta_i}(x, y))^{\omega_i} \right)^{1/\Sigma_{i=1}^{|S_f^\delta|} \omega_i} \right|_{t_2} \tag{2.5}$$

$$\omega_i = \exp^{-\frac{\rho_i^2}{2\sigma'^2}}, 0 \leq t_2 \leq 1$$

where $|\cdot|_{t_2}$ stands for thresholding the response at a fraction t_2 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. We refer the interested
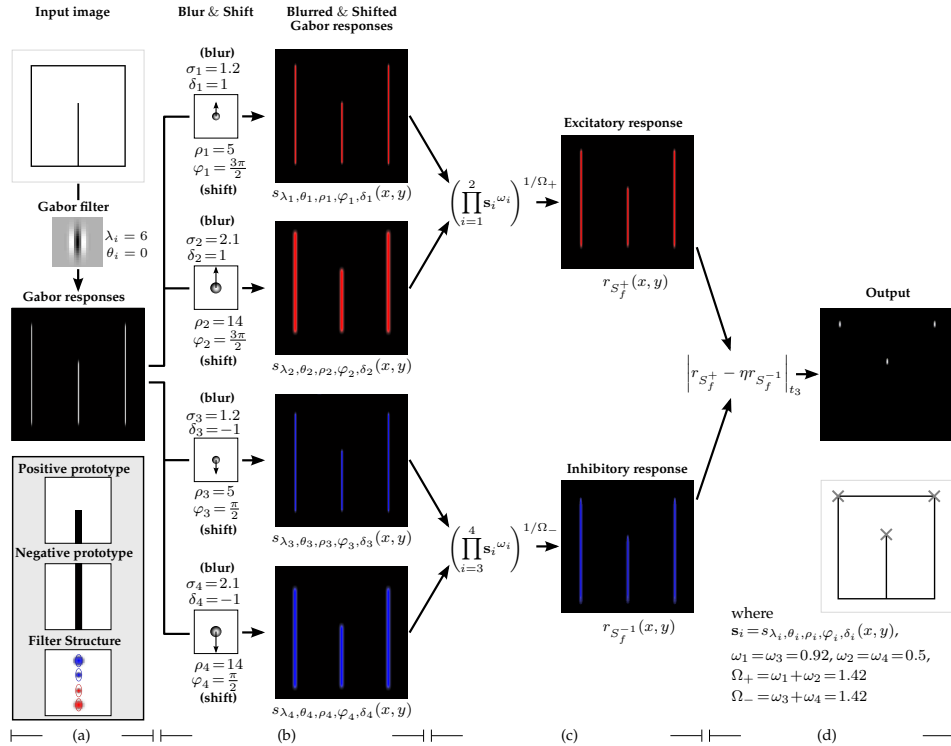
Figure 2.8: Illustration of the intermediate computations performed in an inhibition-augmented filter that is selective to vertical line endings pointing upwards. (a) We first convolve the input image (of size 300×300 pixels) with a Gabor filter which has a wavelength $\lambda = 6$ and an orientation $\theta = 0$. The three enframed inlay images illustrate (top two) the enlarged positive and negative prototype patterns and (bottom) the structure of the resulting inhibition-augmented filter. The red ellipses represent the preferred wavelengths and orientations of the Gabor filters that provide excitatory input to the concerned filter S_f, while the blue ellipses represent the channels of the Gabor filters that provide inhibitory input. Then, (b) we blur and shift all thresholded (here $t_1 = 0.2$) Gabor responses appropriately. Next, (c) we compute two weighted geometric means (here $\sigma' = 11.89$), one for the excitatory blurred and shifted Gabor responses and the other one for the inhibitory input. Finally, (d) we calculate the thresholded (here $t_3 = 0.4$) output of the inhibition-augmented filter by subtracting a fraction η (here $\eta = 1$) of the inhibitory response from the excitatory response. The local maxima responses correspond correctly to the three vertical line endings in the input images, which are indicated by the cross markers.

reader to (Azzopardi and Petkov, 2013b) for more details.

Fig. 2.8 shows an illustration of the application of an inhibition-augmented filter that is selective for vertical line endings pointing upwards. Fig. 2.8d shows the
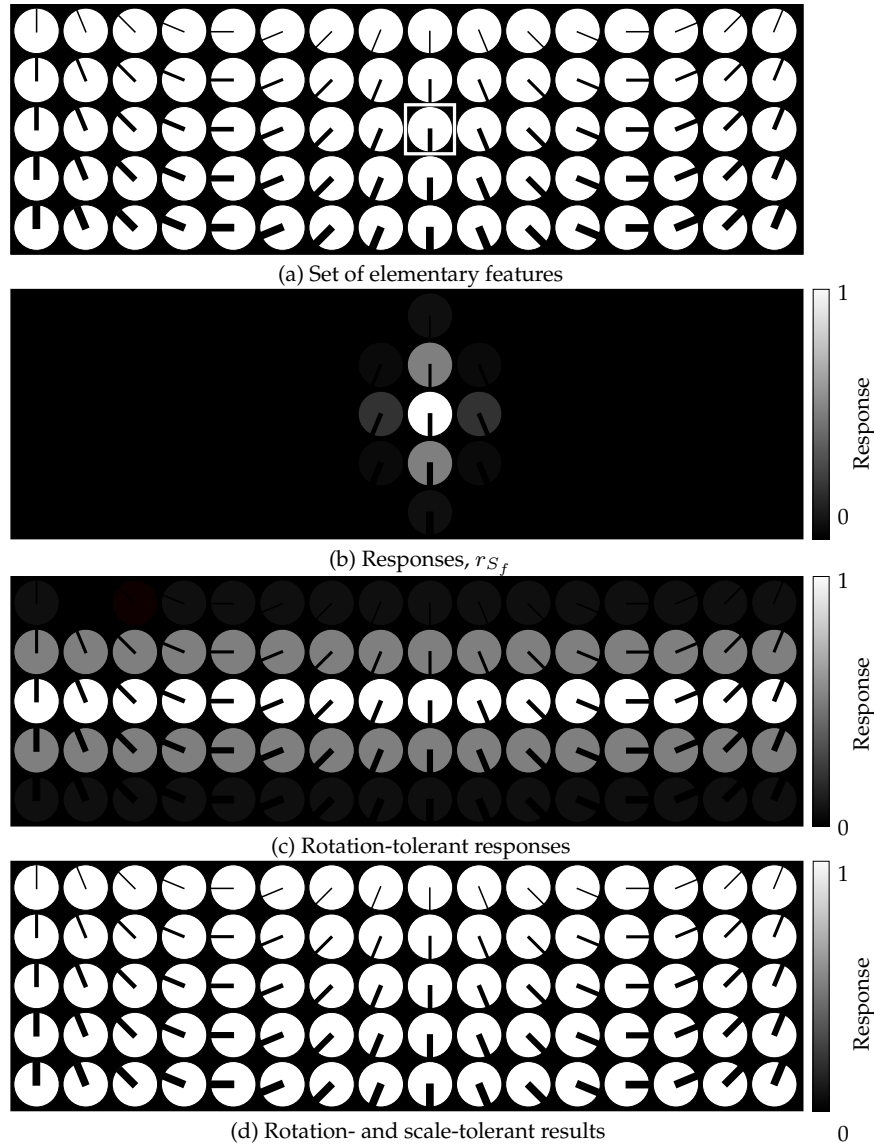
(a) Set of elementary features

(b) Responses, r_{S_f}

(c) Rotation-tolerant responses

(d) Rotation- and scale-tolerant results

Figure 2.9: Illustration of the rotation- and scale- tolerances of the inhibition-augmented COS-FIRE filter. (a) A systematically designed data set of line-endings that vary in orientation (in intervals of $\pi/8$) as well as in scale (the line width ranges from 1 pixel to 5 pixels). The en-framed feature is the same one shown in Fig. 2.3b which is used as a positive prototype for configuring an inhibition-augmented filter. The resulting filter is applied to all features in the data set and (b) the responses are rendered by shading the features in gray. (c) Rotation-tolerant responses. (d) Rotation- and scale-tolerant results.
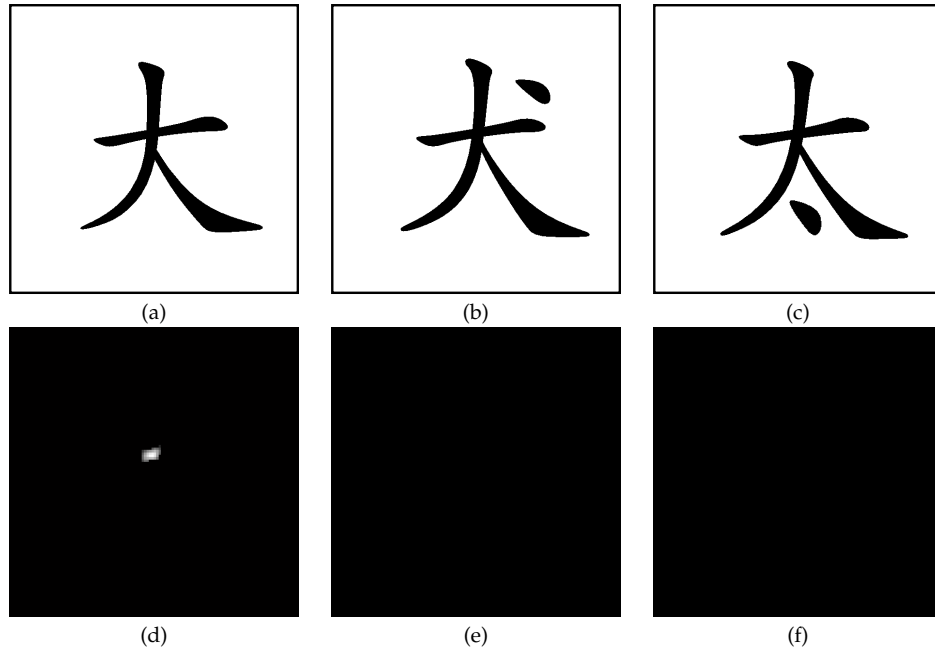
Figure 2.10: The responses of the configured filters to the prototype patterns. (a-c) Images of three similar Chinese letters (of size 120×120 pixels). (d-f) The response images of the resulting filter S_{big} to the corresponding images in (a-c). The filter responds to the letter shown in (a) but does not respond to the letters shown in (b) and (c).

output of this filter and the positions of the strongest local output are marked by crosses in the input image. In this example, this filter only responds strongly at the locations where the positive pattern is present.

Fig. 2.9a shows a data set of line-endings with different line widths and orientations. We applied the same configured inhibition-augmented filter to the stimuli in this data set and the responses of this filter are rendered by a gray level shading of the features (Fig. 2.9b). The maximum response is reached for the feature that was used as a positive prototype in the configuration process while it also reacts, with less than the maximum response, to line-endings that differ slightly in scale and orientation. This example illustrates the selectivity and the generalization ability of the proposed filter. The rotation-invariance in Fig. 2.9c indicates the circularity of the extracted features, which is defined as the invariance of probability distribution of variables that are rotated by any angles (Mandic and Goh, 2009).

Moreover, in Fig. 2.10(d-f) we show the response images of the filter S_{big}, which we configured in Section 2.2.4, to the corresponding patterns in Fig. 2.10(a-c). The

configured inhibition-augmented filter correctly responds only to the pattern shown in Fig. 2.10a but not to the ones in Fig. 2.10(b-c).

2.2.6 Tolerance to Geometric Transformations

The proposed inhibition-augmented filter can achieve tolerance to scale, rotation and reflection by similar parameter manipulation as proposed for the original COS-FIRE filters (Azzopardi and Petkov, 2013b). Fig. 2.9 (c-d) show the rotation- and scaling- tolerant responses of the inhibition-augmented filter that correspond to the set of elementary features shown in Fig. 2.9a. We do not elaborate on these aspects here and we refer the reader to (Azzopardi and Petkov, 2013b) for a thorough explanation.

2.3 Applications

In the following we demonstrate the effectiveness of the proposed inhibition-augmented filter in two practical applications: the detection of vascular bifurcations in retinal fundus images and the recognition of handwritten digits.

2.3.1 Detection of Retinal Vascular Bifurcations

The retina contains cues of the health status of a person. For instance, its vascular geometrical structure can reflect the risk of some cardiovascular diseases such as hypertension (Tso and Jampol, 1982) and atherosclerosis (Chapman et al., 2002). The identification of vascular bifurcations is one of the basic steps in such analysis. For a thorough review on retinal fundus image analysis we refer to (Patton et al., 2006) and (M. Abramoff and Sonka, 2010).

Fig. 2.11 shows an example of a retinal fundus image and its segmentation in blood vessels and background, both of which are taken from the DRIVE data set (Staal et al., 2004). It contains 109 blood vessel features (81 bifurcations marked by red circles and 28 crossovers marked by blue squares). A bifurcation-selective filter configured by the basic COSFIRE approach (Azzopardi and Petkov, 2013b) gives a response also to crossovers and therefore cannot be used to exclusively detect bifurcations. The existing methods that are used to distinguish bifurcations from crossovers preprocess the binary retinal fundus images by morphological operators, such as thinning. Then they typically apply template matching or connected component labeling, which do not work very well for complicated situations, e.g. two bifurcations that are close to each other can be detected as a crossover. An overview of these methods can be found in (Tsai et al., 2004; Bhuiyan et al., 2007; Azzopardi
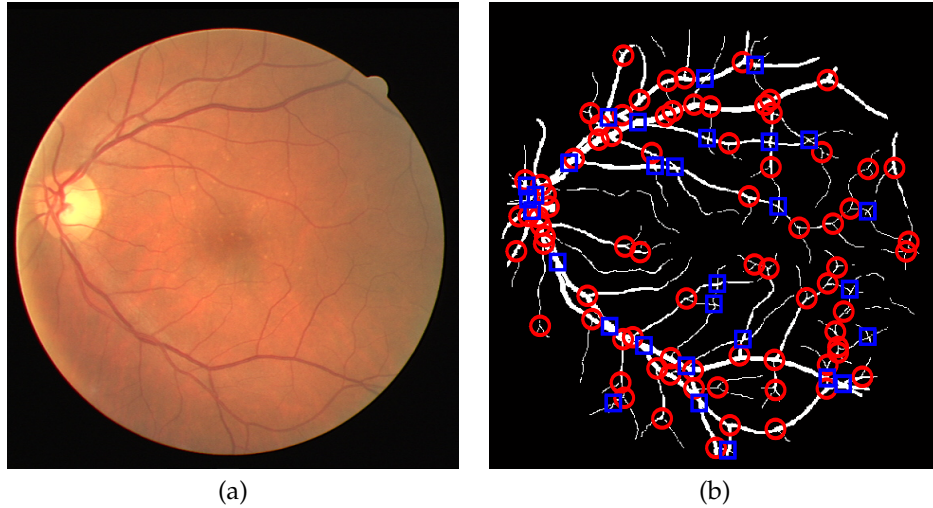
(a) (b)

Figure 2.11: Example of a retinal fundus image from the DRIVE data set (Staal et al., 2004). (a) Original image (of size $564{\times}584$ pixels) with filename 21_training.tif. (b) Binary segmentation of vessels and background (also from DRIVE). The red circles surround vessel bifurcations and blue squares surround crossovers and this labeling is part of the current work.

and Petkov, 2011). In the following, we illustrate how inhibition-augmented filters that we propose can be configured to detect only vascular bifurcations in retinal fundus images.

First, we select a bifurcation prototype from a given retinal fundus image and use it as a positive example to configure a COSFIRE filter P_{f_1} that is composed of excitatory vessel segments. For the configuration of this filter, we use three values of the distance ρ ($\rho = \{0, 5, 10\}$), threshold value $t_1 = 0.2$, $t_2 = 0.45$, a bank of symmetric Gabor filters with eight orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0...7\}$) and five wavelengths ($\lambda \in \{4(2^{\frac{i}{2}}) \mid i = 0 \ldots 4\}$). Fig. 2.12b and Fig. 2.12e show an enlarged prototype and the corresponding filter structure, respectively. Then we apply the configured filter P_{f_1} to all 20 training retinal fundus images (with filenames from 21_manual1.gif to 40_manual1.gif) without tolerance to rotation, scale and reflection transformations. We consider the points that characterize crossover patterns and evoke sufficiently strong responses (which is more than a fraction ε of the maximum response to the positive pattern, here $\varepsilon = 0.2$). We then use these patterns as negative prototypes. Fig. 2.12a and Fig. 2.12c show two of the negative prototypes and the structures of the resulting COSFIRE filters are shown in Fig. 2.12d and Fig. 2.12f. We generate an inhibition-augmented filter S_{f_1} by the method proposed in Section 2.4. Fig. 2.12(d-i) show how two groups of inhibitory line segments are automatically

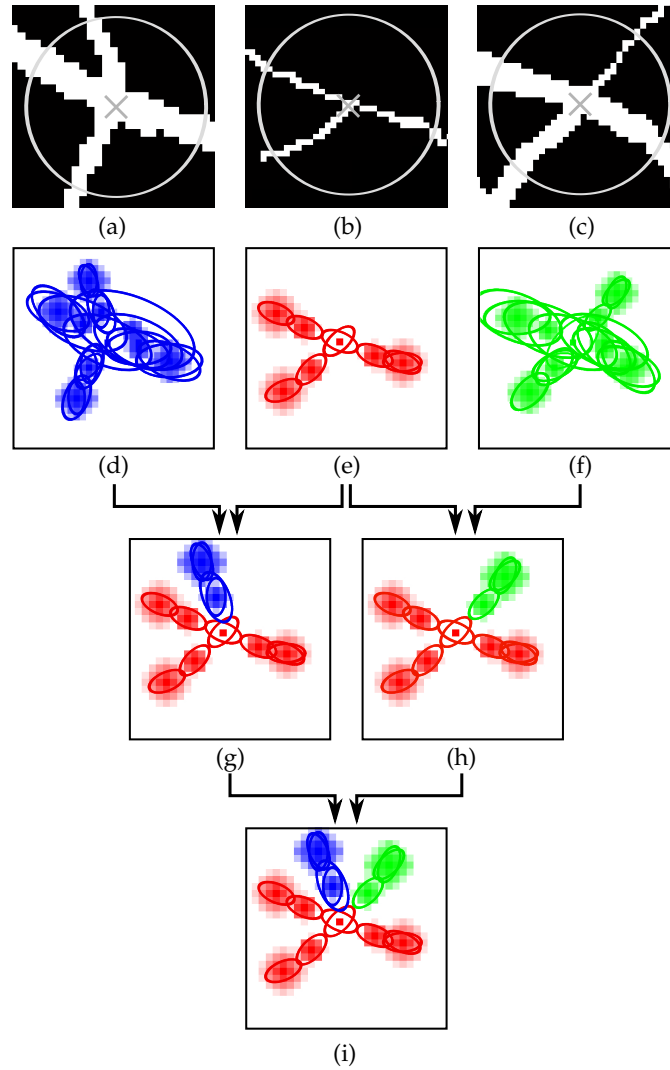selected by the proposed configuration procedure.

Figure 2.12: Examples of positive and negative prototype patterns. (b) A positive prototype pattern, which is the feature of interest. (a and c) Negative prototype patterns. (d-f) The structure of the filters that are selective for the features in (a-c). (g-h) Two inhibition-augmented filters configured by one positive and one negative prototype. (i) An inhibition-augmented filter configured by one positive and two negative prototypes. The tuples that are indicated by the red ellipses come from the positive pattern in (b) and the tuples that are indicated by the blue and green ellipses come from the negative patterns in (a) and (c), respectively.
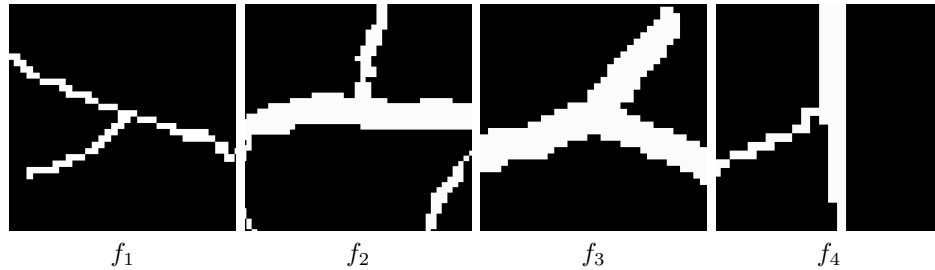
Figure 2.13: A set of 4 bifurcations ($f_1 \ldots f_4$) taken from the DRIVE data set. These four bifurcations are extracted from the binary retinal fundus image shown in Fig. 2.11b with filename 21_manual1.gif.

We repeat the above procedure by applying the filter P_{f_1} in reflection- and rotation-tolerant mode in order to find more negative patterns. Finally, the filter S_{f_1} contains 19 groups of inhibitory tuples.

The values of the inhibition factor η and the threshold t_3 are determined as follows. We apply the filter S_{f_1} to the 20 training retinal fundus images and perform a grid search to estimate the best pair of parameters η and t_3. For η we consider the range of values [0,5] and for t_3 we consider the range [0,1], both of which are in intervals of 0.01. For each combination of these two parameters, we calculate the precision P and recall R. The corresponding harmonic mean ($2PR/(P+R)$) reaches a maximum at an inhibition factor $\eta = 2$ and threshold $t_3 = 0.29$ when the precision P is at least 90 percent. Here, the filter S_{f_1} detects 30 bifurcations and achieves 100 percent precision.

For the remaining bifurcations that are not detected by S_{f_1}, we perform the following steps. We randomly select one of the undetected bifurcations and use it as a new positive prototype. Then, we use the same procedure as described above to find the inhibitory parts of the new filter S_{f_2} as well as the corresponding inhibition factor η and threshold value t_3. The prototype pattern f_2 is shown in Fig.2.13. By applying the filters S_{f_1} and S_{f_2} ($\eta(S_{f_2}) = 1.80$, $t_3(S_{f_2}) = 0.37$) together, we correctly detect 42 correctly detected bifurcations and no crossovers. We continue increasing the number of filters by using vascular features that are not detected by the previously configured filters. For this given retinal fundus image, we achieve 95 percent recall and 100 percent precision with only four filters, Fig.2.13. Table 2.1 reports the values of the parameters η and t_3 that were determined with the grid search method described above.

In order to evaluate the performance of proposed approach, we apply the four inhibition-augmented filters to the 20 test retinal fundus images in the DRIVE data set. We perform two experiments with the four filters; the first one using the fine-
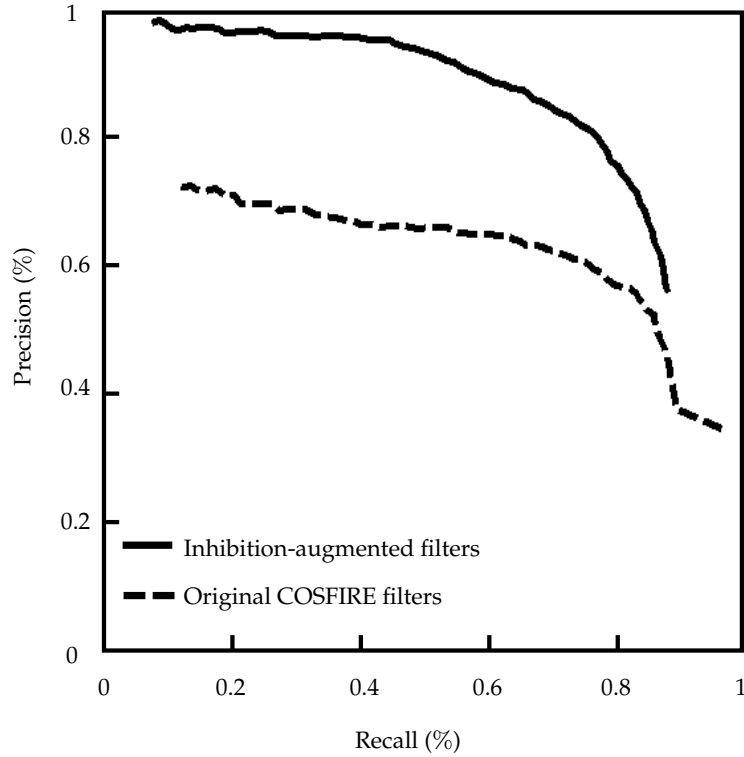
Figure 2.14: Precision-recall plots of the inhibition-augmented method and original COSFIRE method, indicated by the dashed and solid line, respectively.

tuned inhibition factors η and the other one with $\eta = 0$. We change the value of the threshold parameter $t_3(S_{f_i})$ to compute the precision P and recall R. For each filter, we alter the threshold value $t_3(S_{f_i})$ by the same offset value (ranging between -0.2 and 0.2 in intervals of 0.01) which results in the P-R plots shown in Fig.2.14. For the same value of recall, the precision of the inhibition-augmented method is substantially higher than that of the method without inhibition.

Table 2.1: The optimal values of η and t_3.

	S_{f_1}	S_{f_2}	S_{f_3}	S_{f_4}
η	2.00	1.80	2.00	1.60
t_3	0.29	0.22	0.29	0.14

2.3.2 Detection of Architectural and Electrical Symbols

Recognition of Architectural and Electrical Symbols

Recognition of hand-drawn or scanned architectural and electrical symbols is an important application for the automatic conversion to a digital representation which can then be stored efficiently or processed by auto CAD systems. Examples of applications are the recognition of architectural and electrical symbols, optical music notes, document analysis, logo and mathematical expressions (Zanibbi et al., 2002; Valveny et al., 2007; Rebelo et al., 2012; Tang et al., 2013). In such applications, it is common to find that a symbol is contained within another symbol that has a different meaning. Fig. 2.1(a-d) show some pairs of patterns taken from such applications.

Existing symbol recognition algorithms can be categorized into statistical and structural-based approaches. The former methods extract hand-crafted features from symbols and use them to form feature vectors and train classification models (Zhang et al., 2006). While such methods may be effective, they require large numbers of training examples. Moreover, the selection of features is specific to the application at hand and typically requires domain knowledge. The structural-based approaches, which usually describe symbols by the geometrical relation between their constituent parts (Lin et al., 2004), are not suitable to distinguish symbols with similar shapes (Yajie et al., 2007). In the following, we illustrate how the inhibition-augmented filter that we propose is effective for such an application.

We evaluate the proposed approach on the Graphics Recognition Contest (GREC'11) data set (Valveny et al., 2011). The GREC'11 data set contains 150 different symbol classes, in which the images are of size 256×256 pixels. Fig. 2.15 shows some examples of the symbol classes from the data set. This data set consists of three different sets of images, namely SetA, SetB and SetC. SetA contains 2500 images from 50 symbol classes, SetB comprises 5000 images from 100 classes, and SetC consists of 7500 images from 150 classes. The three data sets contain examples with different scale, rotation and various levels of noise degradation. Fig. 2.16a shows a model symbol and Fig. 2.16(c-d) show three symbols of the same class with different levels of noise.

In the following, we explain how the proposed inhibition-augmented filters are configured to be exclusively selective for specific symbol classes. Fig.2.17 shows two such examples of symbol images from the GREC'11 data set. All contour parts of the symbol in Fig.2.17a are contained in the symbol in Fig.2.17b.

For configuration, we do the following steps. First, we consider a model symbol, such as the one in Fig.2.17a, as a positive prototype pattern to configure a COSFIRE filter without inhibition. Fig.2.18a shows the structure of the resulting filter. Then, we apply the configured filter in rotation- and scale- tolerant mode to all the other
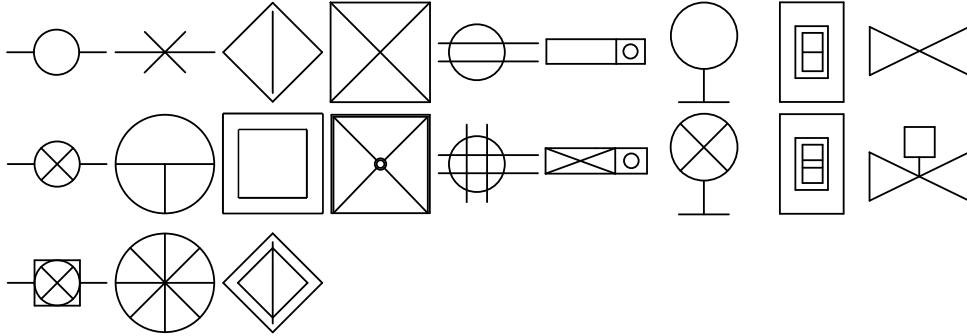
Figure 2.15: Examples of symbol classes from the GREC 2011 data set (Valveny et al., 2011). The symbol in the top left corner is contained within the symbol below it, which in turn is contained within the symbol in the bottom left corner. The top two symbols in the second column are both contained within the symbol in the third row of the same column. The top two symbols in the third column are both contained within the symbol in the third row of the same column. The one in the second row is contained in the third one in a rotation- and scale- mode. The symbols in the first row of the remaining columns are contained within the corresponding symbols of the second row.
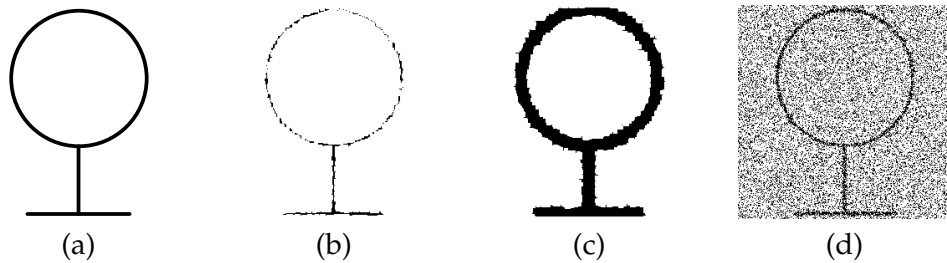

Figure 2.16: Examples from the GREC 2011 data set (Valveny et al., 2011). (a) A model symbol image. (b-d) Degraded symbols of the same class from the data sets of noisy images with different levels of degradations.

149 model images. We threshold the responses at a given fraction ε ($\varepsilon = 0.3$) of the maximal filter response to the positive pattern used for configuration. The other symbol images which evoke strong responses to the filter are considered as negative prototype patterns. For instance, the symbol shown in Fig.2.17b is one negative prototype for the pattern in Fig.2.17a. The COSFIRE filter structure that corresponds to the pattern in Fig.2.17b is shown in Fig.2.18b. Next we compare the structures shown in Fig.2.18a and Fig.2.18b to identify contour parts to be used for inhibition. In Fig.2.18c, we show the structure of the resulting inhibition-augmented filter, in which red and blue ellipses and blobs indicate Gabor responses that provide respec-
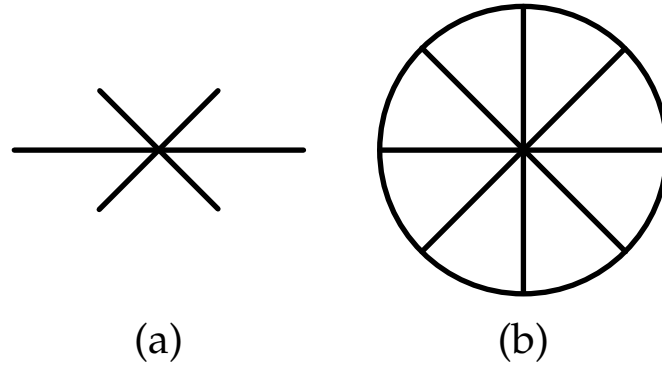
Figure 2.17: Example of (a) a symbol that is contained within (b) another symbol.
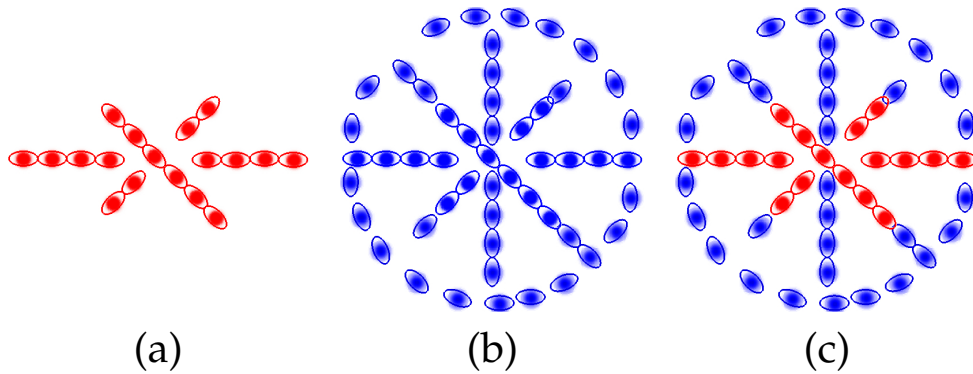

Figure 2.18: Structures of the configured COSFIRE filters. (a-b) The structures of COSFIRE filters configured with the positive and negative prototype in Fig. 2.17a and Fig. 2.17b, respectively. (c) The structure of the resulting inhibition-augmented filter obtained. The ellipses illustrate the wavelengths and orientations of the selected Gabor filters and their positions indicate the locations at which their responses are used as input to the concerned COSFIRE filter. The blobs within the ellipses represent blurring functions that are used to provide some tolerance regarding the preferred positions. Red and blue ellipses and blobs indicate Gabor responses that provide respectively positive and negative inputs to the inhibition-augmented filter.

tively positive and negative inputs to the filter. In this implementation, we consider a bank of Gabor filters with eight orientations ($\theta \in \{ \frac{\pi i}{8} \mid i = 0...7\}$) and two wavelengths ($\lambda \in \{10, 18\}$). We use the empirically-determined threshold values $t_1 = 0.2$ and $t_2 = 0.5$. For the blurring function, we use a fixed standard deviation $\sigma = 4$. In order to make sure that we extract information from all the line segments of a given prototype, we first use a large set of ρ values, and then we remove redundant
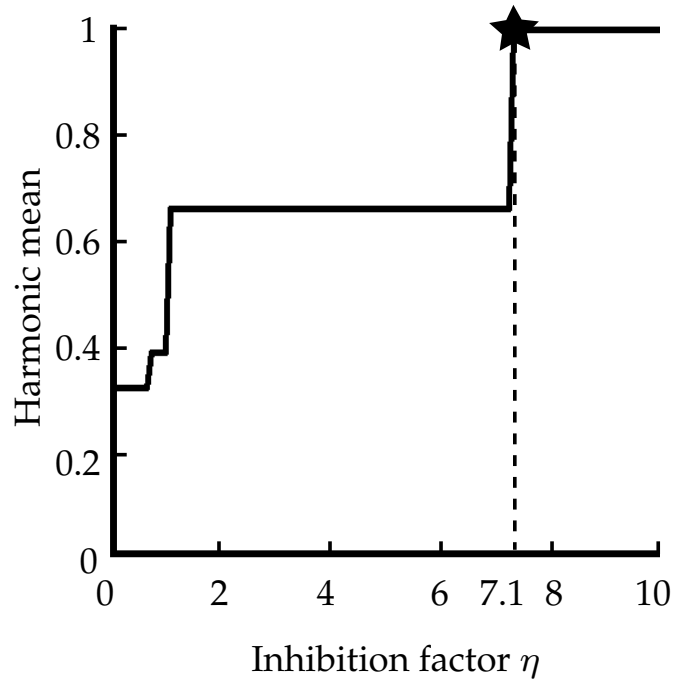
Figure 2.19: The harmonic mean of the precision and recall for the filter shown in Fig.2.18c with different values of inhibition factor η. The star indicates the minimum inhibition factor ($\eta = 7.1$) that achieves the maximum harmonic mean.

tuples from the filters as follows. We compute the pairwise dissimilarity proposed in Section 2.2.3 with parameter ζ equal to 3 times the maximum standard deviation of any pair of tuples, and delete one tuple from the pair whose dissimilarity value is 0. In this way, the corresponding blurring maps of tuples do not overlap each other.

In order to determine the optimal value of the inhibition factor η for such an inhibition-augmented filter, we perform the following steps. First, we apply the filter to all 150 model symbol images with different values of inhibition factor η in a range between 0 and 10 in interval of 0.1. Then for each inhibiton factor, we calculate the harmonic mean of the precision[3] and recall[4] of this filter. Fig.2.19 shows the harmonic mean of the concerned filter with different values of inhibition factor. The optimal inhibition factor ($\eta = 7.1$) is the minimum value of η that achieves the highest harmonic mean. In Fig.2.19 we indicate this point by a star marker.

---

[3]We compute precision as the number of images to which the filter correctly responds divided by the total number of images to which the filter responds.

[4]We compute recall as the number of images to which the filter correctly responds divided by the total number of images to which the filter should respond.
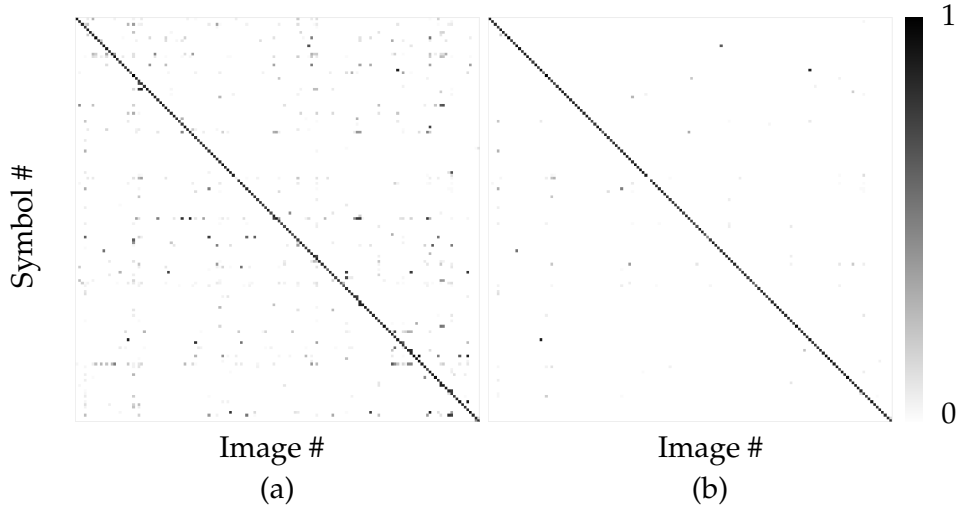
Figure 2.20: Matrices (of size 150×150) obtained by (a) original COSFIRE filters and (b) inhibition-augmented COSFIRE filters with the optimal inhibition factors. The columns represent images and the rows represent filters which are configured from symbols. The elements of the matrices are the maximum responses of the configured filters to the images, which are rendered as grey levels.

We perform the same procedure on the remaining 149 symbols. We apply the resulting 150 inhibition-augmented filters to the 150 symbol images. Fig.2.20(a-b) show matrices (of size 150×150) obtained using the COSFIRE filters without inhibition ($\eta = 0$) and the inhibition-augmented COSFIRE filters, respectively. The value of the element (i, j) in each matrix is the maximum response of the filter configured by symbol i to symbol image j. For each filter we compute the precision and recall. The average precision achieved by the COSFIRE filters without inhibiton is 48.0% while the one for the inhibiton-based filters is 81.7%. The recall for both methods is 100%. Compared to the results of the original COSFIRE fillters, the matrix obtained by the inhibition-augmented filters is much sparser and the precision is significantly improved.

Before applying the configured filters to the test images in SetA, SetB and SetC, we pre-process each image as follows. We compute the mean value of the intensities of all the pixels in an image. For the images that have a mean intensity value of at least 90% of the maximum, we apply the morphological operations proposed in (Guo, Shi, Azzopardi and Petkov, 2015a). First, we dilate the images by six line-shaped structuring elements of 6-pixel length with different orientations ($\{0, \frac{\pi}{6}, \frac{\pi}{3}, \dots, \frac{5\pi}{6}\}$). Then, we perform a thinning operation followed by six dilations using line-shaped structuring elements of 4-pixel length with equidistant ori-

Table 2.2: Recognition rates (%) for the three data sets in the GREC'11 set (Valveny et al., 2011)

	SetA	SetB	SetC	Average
Our proposed method	**98.80**	**97.64**	**98.11**	**98.18**
COSFIRE filters (Azzopardi and Petkov, 2013b)	96.80	94.50	91.81	94.37
Spectra of shape context (Yang, 2014)	92.96	90.70	89.79	90.62
Geometric matching (Valveny et al., 2011)	94.76	91.98	85.88	89.39

entations. Finally, we apply opening and thinning followed by a dilation operation using a series of line-shaped structuring elements of 4-pixel length in six orientations. We do not pre-process the images that have a mean value less than 90% of the maximum since most of them do not lose part of their contour segments.

We apply the 150 inhibition-augmented filters to each pre-processed image by using the proposed method in rotation- and scaling-tolerant mode with parameters $\psi = \left\{ \frac{\pi i}{32} \mid i = 0, 1, \ldots, 31 \right\}$ and $v = \{0.5, 0.6, \ldots, 2.5\}$. A given image is classified to the class of the positive prototype symbol by which the inhibition-augmented filter that achieves the maximum response was configured. In Table 2.2, we compare the results that we achieve with the existing methods on the three data sets. The proposed approach achieves the best results in all subsets of the data sets.

Localization of Architectural and Electrical Symbols

In this subsection, we evaluate the performance of the inhibition-augmented COSFIRE filters for localization of symbols in circuit diagrams and floor plan images.

We experiment on the Graphics Recognition (GREC2011) localization data set (Valveny et al., 2011). The data set contains 16 architectural and 21 electrical symbols. It consists of a training and a test set. In the training set, there are 40 drawing images, 20 of which are floor plans consisting of architectural symbols and the rest are electrical diagrams. In each domain, there are four subsets, namely Ideal, Level 1, 2 and 3, of five images each. The latter three subsets have different levels of noise degradation. The test set has 160 drawing images, 80 images of which are floor plans and the rest are electrical diagrams. Similar to the training set, every domain has four subsets (Ideal, Level 1, 2 and 3) of 20 images each. In total, the test set of GREC2011 data set for localization contains 3463 symbols in all drawing images. The lines in the ideal symbols have a thickness of 5, 9 or 12 pixels. Fig. 2.21 shows an example of an electrical diagram taken from the Ideal subset.

First, we configure 16 inhibition-augmented COSFIRE filters to be selective for the 16 architectural symbols given in the training set. The way we configure the filters is as follows. We configure an excitatory-only COSFIRE filter as proposed in (Azzopardi and Petkov, 2013b) for a given symbol prototype and apply the re-
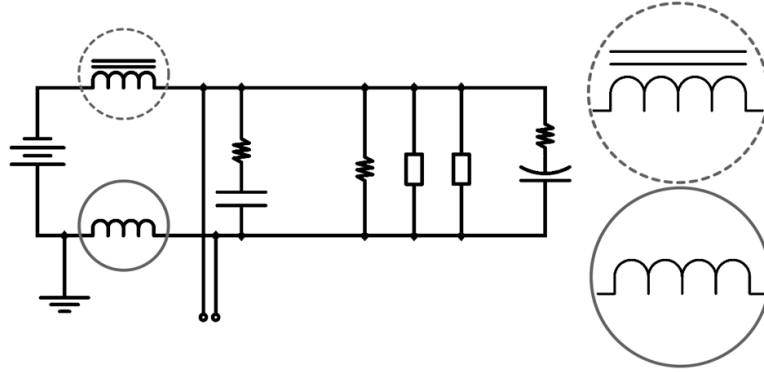
Figure 2.21: An example image (of size 2981×1463 pixels) of an electrical circuit. The symbol encircled by a solid line, which indicates a core-air inductor, is considered here as a positive prototype pattern of interest, while the symbol marked by a dashed circle, which indicates a core-iron inductor, is considered as a negative prototype. The two inset images (of size 810×294 pixels) on the right are enlargements of the positive and the negative prototypes.

sulting filter to the remaining 15 symbols. After that, we threshold each response map at a fraction ϵ of the maximal response that the resulting filter achieves when applied to the given prototype. We take the symbol images that elicit responses greater than $\epsilon(= 0.1)$ as negative prototype patterns and determine the inhibitory line segments. Then the inhibition-augmented COSFIRE filter is configured to be selective for the given pattern. We perform the same procedure for each of the 16 architectural symbols. For the symbols from the electrical domain, we apply the same configuration procedure as above to configure 21 inhibition-augmented COSFIRE filters. In this experiment, we use a bank of Gabor filters with eight orientations ($\theta \in \{\frac{\pi i}{8} \mid i = 0...7\}$) and three wavelengths ($\lambda \in \{10, 18, 24\}$). The wavelength of a symmetric Gabor filter is roughly twice the thickness of the preferred line. The choice of the three wavelength values is motivated from the empirical determination of line thicknesses in the images of the training set. We threshold the Gabor responses with $t_1 = 0.2$ and the excitatory and inhibitory inputs with $t_2 = 0.5$. For the blurring function, we set σ to 4 pixels. We use a value of 12 pixels for ζ, which is three times the standard deviation σ, in order to prevent the interference of inhibitory and excitatory parts of the filter. Next, we apply the above-configured inhibition-augmented COSFIRE filters to all training symbol images in each domain. We investigate the inhibition factor by varying the value of parameter η between 0 and 2 in intervals of 0.1. For $\eta = 1$, the filters give responses only to the preferred positive prototype pattern. Then in order to determine the optimal value of t_3 for each configured filter, we apply each filter in rotation- and scaling- tolerant mode to

Table 2.3: Comparison of results (in terms of precision, recall and F-score) between the proposed method, excitatory-only COSFIRE and the Geometric matching (Valveny et al., 2011) approach. (The results in bold font represent the maximum values among the three methods.

	Our method Precision/Recall/F-score	Excitatory-only COSFIRE Precision/Recall/F-score	Geometric matching Precision/Recall/F-score
Architectural Ideal	**0.85** / 0.95 / **0.90**	0.81 / 0.95 / 0.88	0.62 / 0.99 / 0.76
Architectural Level 1	**0.98** / 0.85 / **0.91**	0.45 / 0.90 / 0.60	0.64 / **0.98** / 0.77
Architectural Level 2	**0.95** / 0.87 /**0.91**	0.64 / 0.88 / 0.74	0.62 / 0.93 / 0.74
Architectural Level 3	**0.95** / 0.88 / **0.91**	0.47 / 0.93 / 0.62	0.57 / **0.98** / 0.72
Electrical Ideal	**0.88** / 0.87 / 0.87	0.88 / **0.92** / **0.90**	0.37 / 0.56 / 0.45
Electrical Level 1	**0.97** /**0.97** / **0.97**	0.79 / 0.95 / 0.86	0.44 / 0.63 / 0.52
Electrical Level 2	**0.96** / 0.89 / **0.92**	0.73 / **0.90** / 0.81	0.40 / 0.61 / 0.48
Electrical Level 3	**0.95** / 0.89 / **0.92**	0.80 / **0.94** / 0.86	0.43 / 0.64 / 0.51
Average	**0.94** / 0.90 / **0.91**	0.70 /**0.92** / 0.78	0.51 / 0.79 / 0.62

the drawing images in the training set (subsets Ideal, Level 1, 2 and 3). For the images in the noisy sets, we do not apply any pre-processing method. We use different values of t_3 between 0 and 1 in intervals of 0.1 to threshold the response of such an inhibition-augmented COSFIRE filter to a given image. We compute the harmonic mean of the precision and recall for each value of the threshold. The optimal threshold is the minimum value that achieves the highest harmonic mean.

We apply the resulting 16 inhibition-augmented COSFIRE filters selective for architectural symbols to the images in the subsets Ideal, Level 1, 2 and 3 of architectural domain in rotation- and scaling- tolerant mode. Similarly, we apply the 21 filters selective for electrical symbols on the corresponding subsets of electrical domain. We compare our results with the ground truth provided in the dataset. In each test drawing image, the ground truth provides a rectangular region in which a symbol is located. If the distance between the center of the detected symbol and the center of the provided region is less than 0.25 of the maximum of the width and the height of the provided region, we consider it as a successful localization. We report the achieved precision, recall and F-score in Table 2.3, in which we highlight (in bold font) the best results achieved among different methods. We compare the results of the proposed filters with those obtained by the original excitatory-only COSFIRE filters by setting the inhibition factor $\eta = 0$. The method proposed in (Azzopardi and Petkov, 2013b) achieves the best recall in the architectural subsets while in average the excitatory-only COSFIRE filters have better recall than that from the other two methods. In average, the best precision and F-score are achieved by our proposed inhibition-augmented COSFIRE filters which have slightly lower recall (0.90) than that (0.92) achieved by excitatory-only COSFIRE filters.
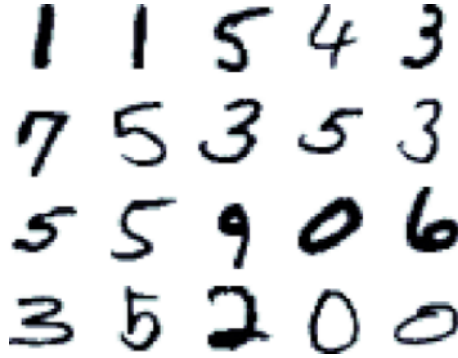
Figure 2.22: Examples of handwritten digits in the MNIST data set.

2.3.3 Recognition of Handwritten Digits

Handwritten digit recognition is an important application in optical character recognition (OCR) systems. It helps to accelerate the distribution of postal mails, processing of bank checks and so on. Various benchmark data sets and approaches have been proposed for such application, a thorough literature review of which is given in (Liu et al., 2003).

In this application, we use the MNIST data set (LeCun et al., 1998) to evaluate the performance of our approach. The data set contains 60000 training and 10000 test digit images in gray scale of size 28×28 pixels[5]. Fig. 2.22 shows some examples of the handwritten digits in the MNIST data set. We first configure inhibition-augmented COSFIRE filters to extract the local contour parts of digit images. Then we apply these filters to all the training images and generate feature vectors from the maximum responses of these configured filters. Finally, we use these feature vectors to train support vector machine classifier to recognize the digits in the test data sets.

For configuration, we randomly select 20 training images from each digit class. We select a random location as the point of interest to configure a COSFIRE filter in each image from the same digit class. The local pattern around such a point should provide at least four tuples to the resulting filter, otherwise we select another random location. Then we apply this filter to all the other 180 training images from different digit classes in order to identify negative prototypes. We use the method described in Section 2 to configure an inhibition-augmented filter. We repeat the above process for all the 200 training digit images and configure 200 inhibition-augmented filters. In this application, we use a bank of antisymmetric Gabor filters with 16 equidistant orientations ($\theta \in \{ \frac{\pi i}{8} \mid i = 0...7\}$), one wavelength ($\lambda = 2\sqrt{2}$)

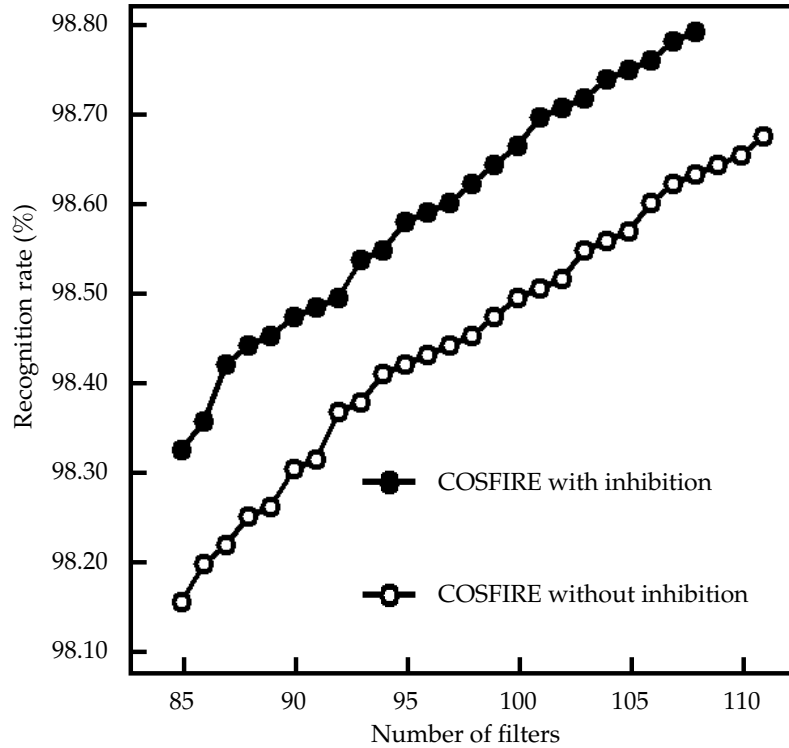[5]The MNIST set is available online: `http://yann.lecun.com/exdb/mnist`

Figure 2.23: The plots show the recognition rates on the MNIST data set with different numbers of filters by using the methods with and without inhibition.

and three values of ρ ($\rho \in \{0, 3, 8\}$), and threshold their responses with $t_1 = 0.2$, $t_2 = 0.99$.

Next, in the application process we apply these 200 inhibition-augmented filters to 60000 training images by using the proposed method. We take the maximum response of each filter to these digit images and generate a matrix of size 60000×200.

Then, we apply a wrapper method for feature selection using support vector machines (SVMs) with a linear kernel. We iteratively add the result of one filter, the one that best improves the 7-fold cross validation accuracy, and stop the process until no more improvement is achieved. This process results in 108 filters when the 200 inhibition-augmented COSFIRE filters are applied ($\eta = 1$) and 111 filters when the 200 original COSFIRE filters are applied ($\eta = 0$). Then we use the inhibition-augmented and non-inhibition-augmented training vectors with the selected features to train two multi-class SVMs.

The plots in Fig. 2.23 show the recognition rates as a function of increasing

number of selected filters. The method with inhibition achieves a recognition rate of 98.77% with 108 filters while the method without inhibition achieves 98.66% with 111 filters. The inhibition-augmented training vectors of 108 dimensions have 753019 (11.62%) zero elements, which is substantially greater than the 277641 (4.17%) zero elements in non-inhibition-augmented vectors of 111 dimensions.

2.4 Discussion

We proposed an inhibition-augmented COSFIRE approach which uses a positive prototype and a set of negative prototypes to configure a filter. The choice of negative prototypes can be either manually specified by a user or automatically discovered by the system. For instance, the negative prototype shown in Fig. 2.3, a complete line, is selected by the user. For more complex situations, such as the recognition of symbols and handwritten digits, it is more practical to use an automated process. To discover negative prototypes, we first apply the COSFIRE filter which is configured by a positive prototype pattern to all the other pattern images. The ones which evoke strong responses to the filter are negative prototype patterns.

We use Gabor filters, which are the model of orientation-selective cells in area V1 and V2 of the visual cortex. Gabor filters are, however, not intrinsic to the proposed model and other computational models of simple cells, for example the models proposed in (Azzopardi and Petkov, 2012) and (Azzopardi et al., 2014), can also be used.

The response of an inhibition-augmented filter is defined as the difference between the excitatory input and a fraction of the maximum of the inhibitory inputs. The inhibition factor can be adjusted by changing the value of the parameter η. In the detection of vascular bifurcations and the symbol recognition applications, we determine an optimal value of η for each filter as the one that contributes to the maximum harmonic mean on the training images. For the other application, we set the same η value for all filters so that none of them achieves a response to any of the negative patterns.

We use four parameters t_1, t_2, t_3 and η to control the output of an inhibition-augmented COSFIRE model. The parameter t_1 regulates the threshold at which the response of a Gabor filter indicates the presence of a contour part. The choice of the value of parameter t_1 depends on the contrast of the image and the presence of noise. The value of t_2 controls the minimum valid value of the response that provides the excitatory and inhibitory inputs to the model. The parameter t_3 depends on the noise of an input image and it is used to suppress the responses that are below a given fraction of the maximum response value across all locations of the

input image. We use the value of the parameter η to adjust the inhibition factor. The inhibition contour parts and the inhibition strength parameter that we propose are determined from the training set of a given application. The generalization ability of an inhibition-augmented COSFIRE model decreases with an increasing number of inhibitory contour parts and/or with an increasing value of the inhibition strength parameter η. We determine optimal values of t_1, t_2, t_3 and η for each model as the ones that contribute the best to the results on the training images.

In neurophysiology there is an ongoing debate about what kind of neural coding the brain uses to encode the representation of objects. The two extremes in the debate are the grandmother cell theory (i.e. only one specific cell fires for a given pattern) and population coding (i.e. a number of neurons fire for a given pattern with different rates). In the recognition of handwritten digits, the proposed inhibition-augmented COSFIRE filters are similar to the population coding. This application demonstrates that the inhibition mechanism facilitates sparseness in the representation of information. This is attractive for the brain because sparse coding increases storage capacity and allows the discrimination of more objects.

The computational cost of the configuration of a COSFIRE filter with inhibition depends on the number of negative prototype patterns and the bank of Gabor filters it uses. An inhibition-augmented filter is configured in less than one second for one positive and one negative prototype pattern with the size of 512×512 pixels and a bank of Gabor filters of eight orientations and five wavelengths. The computational cost of the application of an inhibition-augmented filter is proportional to the computations of the excitatory and inhibitatory responses and their blurring and shifting operations. For the detection of vascular bifurcations, a retinal fundus image of size 564×584 pixels is processed in less than 20 seconds by four rotation- and reflection-tolerant inhibition-augmented filters. For the recognition of architectural and electrical symbols, a symbol image of size 256×256 pixels is processed in less than 30s by 150 inhibition-augmented filters without any rotation or scaling tolerances. For the third application, a handwritten digit image of size 28×28 pixels is described by 200 inhibition-augmented COSFIRE filters without any rotation- or scaling- tolerances in less than 5 seconds. We used a sequential implementation in Matlab[6] for all experiments that runs on the same standard 3GHz processor.

There are various possible directions for future research. One direction is to apply the proposed inhibition-augmented filters in other objection localization and recognition tasks, as well as image classification. Another direction is to investigate a learning algorithm to determine the output function by assigning different weights to inhibitory and excitatory contour parts.

---

[6]Matlab scripts for the configuration and application of inhibition-augmented filter can be downloaded from `http://matlabserver.cs.rug.nl`.

2.5 Conclusions

The proposed inhibition-augmented filters are versatile trainable shape and object detectors as they can be trained with any given positive and negative prototype patterns. We demonstrated the effectiveness of the method with three applications: detection of vascular bifurcations (i.e. without crossovers) in retinal fundus images (DRIVE data set), recognition and localization of the architectural and electrical symbols (GREC2011 data set) and the recognition of handwritten digits (MNIST data set). The inclusion of the inhibition mechanism improves the discrimination properties and the performance of COSFIRE filters.