Pattern Recognition 44 (2011) 1999–2007

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection

Giuseppe Papari *, Nicolai Petkov

*Johan Bernoulli Institute of Mathematics and Computing Science, University of Groningen, Groningen, Netherlands*

ARTICLE INFO

Available online 12 August 2010

*Keywords:*
Contour detection
Surround suppression
Steerable filters

ABSTRACT

Psychophysical and neurophysiological evidence about the human visual system shows the existence of a mechanism, called surround suppression, which inhibits the response of an edge in the presence of other similar edges in the surroundings. A simple computational model of this phenomenon has been previously proposed by us, by introducing an inhibition term that is supposed to be high on texture and low on isolated edges. While such an approach leads to better discrimination between object contours and texture edges w.r.t. methods based on the sole gradient magnitude, it has two drawbacks: first, a phenomenon called self-inhibition occurs, so that the inhibition term is quite high on isolated contours too; previous attempts to overcome self-inhibition result in slow and inelegant algorithms. Second, an input parameter called "inhibition level" needs to be introduced, whose value is left to heuristics. The contribution of this paper is two-fold: on one hand, we propose a new model for the inhibition term, based on the theory of steerable filters, to reduce self-inhibition. On the other hand, we introduce a simple method to combine the binary edge maps obtained by different inhibition levels, so that the inhibition level is no longer specified by the user. The proposed approach is validated by a broad range of experimental results.



## 1. Introduction

Edge detection is an important problem in computer vision and pattern recognition, with many applications in both scientific and practical problems, such as object recognition or shape analysis. Some existing approaches are based on differential [1], statistical [2], and local phase [3,4] analysis, machine learning [5,6], active contours [7,8], perceptual organization [6], graph theory [9,10], and multiresolution analysis [11]. Despite of the huge amount of work that has been done in this area, there is still room for improvement of the existing algorithms and the development of new, more effective ones.

We make use of the insights obtained in psychophysics and neurophysiology of the visual system (of primates) in order to improve edge detection algorithms. Specifically, we consider a neural mechanism called *surround suppression* that is observed in areas V1 and V2 of the (monkey) brain.[1] Its essence is that the response of an orientation selective neuron to a local oriented stimulus is inhibited by the presence of other similar stimuli in the immediate surroundings. In previous work [12], we proposed a simple computational model for surround suppression. It is based on the computation of an inhibition term [12], which is defined as the local average of the gradient magnitude on a ring around each pixel. The inhibition term is supposed to be high on texture and low on isolated edges. When the inhibition term (multiplied by a scaling factor called *inhibition level*) is subtracted from the gradient magnitude, the resulting quantity discriminates between object contours and texture edges better than the sole gradient magnitude.

This approach has two drawbacks. First, neighboring parts and the same contour will inhibit each other to a certain extent. Previous attempts to overcome this problem, called self-inhibition, result in slow inelegant algorithms [13,14]. Second, the weight with which the inhibition term, called the inhibition level, needs to be multiplied is an input parameter that must be specified by the user and its optimal value may vary from image to image (Fig. 1).

In this paper, we overcome the aforementioned limitations in two ways: (i) we develop a new operator for the computation of the inhibition term, based on the theory of steerable filters, which avoids self-inhibition and is much faster than previous methods. (ii) We propose a simple algorithm which combines the binary edge maps obtained for different values of the inhibition level. In this way, the inhibition level needs no longer be specified by the user.

The rest of this paper is organized as follows: after a short review of previous work in edge detection (Section 2), we present the proposed edge detector (Section 3), demonstrate its effectiveness and advantages over the existing surround suppression

* Corresponding author.
*E-mail address:* g.papari@rug.nl (G. Papari).

[1] The areas V1 and V2 of the brain in mammals, also called primary visual cortex and pre-striate visual cortex, respectively, are the first two areas in the visual cortex in which low level visual stimuli are processed.

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.08.013

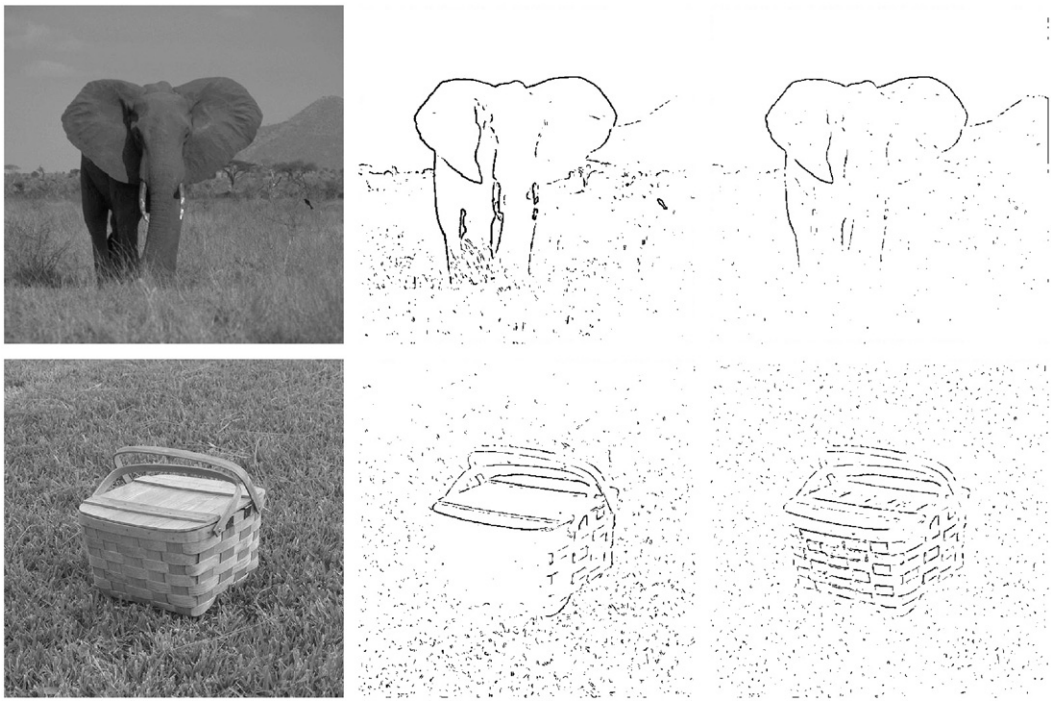

**Fig. 1.** From left to right: input image and output of the approach proposed in [12] with inhibition levels equal, respectively, to 1 and 3. For the first image, low inhibition levels give better results, while for the second image the best results are obtained for high inhibition levels.

algorithms (Section 4), and finally present a discussion and conclusions (Section 5).

## 2. Background on edge detection

We now present a brief overview of the state of the art in the field of edge detection. We divide the existing algorithms into non-contextual and contextual methods. In the former, edges are detected by only looking at a small neighborhood of each pixels, while in the latter the information on a larger context is considered too.

### 2.1. Non-contextual methods

The existing non-contextual methods are based on (i) differential and (ii) statistical analysis, (iii) local energy and phase congruency, and (iv) combination of the aforementioned features by means of machine learning techniques. In the following, we shall briefly discuss these typologies of algorithms

*Differential methods*: Edges, identified as discontinuities of the input luminance profile $I(x,y)$, can be detected as points of high gradient magnitude. Specifically, edges are defined as the local maxima of the gradient magnitude. These points are given by the zero crossings of the second derivative $I_{vv}(x,y)$ of $I(x,y)$ in the direction $v$ of the gradient, for which the third derivative is negative. $I_{vv}(x,y)$ is often replaced by the Laplacian, which has a simpler analytical expression and, on low-curvature points, is a good approximation of $I_{vv}(x,y)$ [15]. Zero-crossings which do not satisfy the condition on the third derivative, known in the literature as *phantom edges* [16], are local minima of the gradient magnitude and do not correspond to edges.

As pointed out in [17], the computation of the derivatives of a digital image is an ill-posed problem. To regularize it, the input image must first be convolved with a low-pass pre-filtering. Canny proposed to optimize the template filter with respect to the following three criteria: good detection, good localization and low multiple responses [18]. It was found that the optimal filter for step edges is very close to the first derivatives of a Gaussian function. Discretized version of these criteria have been formulated in [19,20].

*Statistical approaches*: Differential methods are unable to detect boundaries defined by texture changes. To overcome this restriction, it has been proposed to analyze the local pattern on a neighborhood $\mathcal{N}(\mathbf{r})$ around each pixel $\mathbf{r}$ by means of statistical tools. The simplest technique consists in dividing $\mathcal{N}(\mathbf{r})$ into two equal parts along a given orientation and using a two-sample statistical test of independence to measure the dissimilarity between the two halves. High values of the dissimilarity indicate the presence of a region boundary. This analysis is repeated for several directions and the one which gives rise to the maximum dissimilarity is regarded as the local edge direction [2,21]. More recently, these ideas have been extended to color images [22] and integrated to texture models [5].

Other statistical approaches look at the local distribution of the gradient. The most common descriptive features are eigenvalues and eigenvectors of the local covariance matrix of the gradient [23,24], and local angular dispersion of the gradient [25,26]. In general, statistical approaches are more effective than differential methods in detecting color and texture transitions, but they are computationally more demanding.

*Phase congruency and local energy*: The human visual system responds strongly to points at which phase information is highly ordered [27]. In [3,4], a quantity called phase congruency is introduced, which is always between 0 and 1, being 1 on those points for which all Fourier components are in phase. Its local maxima correspond to salient visual events, such as step, peak and roof edges [4]. These maxima can be detected by analyzing a different quantity $L(t) \triangleq \sqrt{[(x \star f_e)(t)]^2 + [(x \star f_o)(t)]^2}$, called local energy. Here, $f_e(t)$ and $f_o(t)$ are a symmetric and an anti-symmetric low-pass or band-pass filter, such that $f_o(t)$ is the Hilbert transform of $f_e(t)$. Several pairs of functions $f_e(t)$ and $f_o(t)$, known in the

literature as quadrature pair filters, have been studied. Examples are the Gabor, log-Gabor, Gaussian derivatives, difference of Gaussians, and Cauchy functions [28]. The best known quadrature filter is the Gabor energy filter, which gives the optimal compromise between spatial and frequency indetermination [29], though it is not optimal for edge detection [28]. The success of local energy and phase congruency approaches is mainly due to their ability to detect different types of edges, such as step, ramp, roof, and line, thus making an unified framework. However, for practical applications, local energy methods perform similarly to the faster and conceptually simpler differential methods [5].

*Edge features combination*: The local features described above take into account the complementary aspects of the edge detection problem. Considerable effort has been made to combine the information carried out by different features. This is done by training classifiers that work in the concerned feature space by means of the semantic information contained in available ground truths. The classifier returns a quantity $P$ that we call *edge likelihood*, which can be, e.g., the posteriori probability or the Fisher discriminability [30], depending on the type of classifier that is deployed. Thresholding $P$ is equivalent to classifying an edge pixel with the optimal decision boundary in the concerned feature space. In [31], a Bayesian approach is followed to measure the performance improvement brought by combining the following local features: gradient magnitude, the output of a local energy analysis [32] and the Nitzberg edge strength [23]. A more exhaustive study is carried out in [5], where a larger set of local features is taken into account, including color and texture gradient. The dependence of the edge detection performance on the choice of the classifier is studied in [5,33].

### 2.2. Contextual methods

We now describe several attempts to improve local edge detection by using contextual information. The main existing approaches are based on mechanisms deployed by the visual system of the brain (surround suppression and facilitation), and on the graph theory (relaxation labeling).

*Biologically motivated approaches*: Various psychophysical and neurophysiological studies (see e.g. [34] and the references therein) show that the response of the human visual systems to an oriented stimulus is influenced by the presence of other similar stimuli in the surroundings. Two mechanisms have been identified: (i) facilitation from stimuli which are collinear with the central one [35,36] and (ii) inhibition by other stimuli. These findings inspired several authors to perform contour detection in two steps: first, local edge strength $E(\mathbf{r})$ is computed; second, $E(\mathbf{r})$ is inhibited or enhanced depending on the surrounding context. The result is a more informative quantity $G(\mathbf{r})$ that is called contour saliency.

Models of surround suppression will be discussed in the next sections. As to facilitation, one of the best known approaches is called *tensor voting* [37,38]. The idea is to associate a vector or a tensor field $\mathbf{v}_{\mathbf{r}}(\mathbf{x})$, called *extension field*, with each pixel $\mathbf{r}$ of the image. Each value of $\mathbf{v}_{\mathbf{r}}(\mathbf{x})$ on point $\mathbf{x}$ attempts to provide an *a priori* estimation of the edge strength and edge direction at point $\mathbf{x}$, only based on (i) the local edge strength and orientation in $\mathbf{r}$, and (ii) the fact that nearby edges are collinear or cocircular. Thus, at every point $\mathbf{x}$, many such estimation are collected for each edge pixel $\mathbf{r}$. Then, saliency is obtained from the statistical distribution of the extension fields of all pixels $\mathbf{r}$. In [37] the covariance matrix $K$ is considered and both saliency values and local edge orientation are obtained from the eigenvalues and eigenvectors of $K$.

*Relaxation labeling*: A powerful probabilistic graph-based framework for the computation of saliency is *relaxation labeling* [9,10], which provides a rigorous framework to develop the intuitive ideas of tensor voting. The idea is to construct a graph whose nodes are edge pixels and whose arcs define the neighborhood on which the contextual feedback is based. Each node of such a graph is associated with a label $p$, which is a real number that measures the likeliness that the concerned edge belongs to an object contour. Arcs of this graph are weighted by real numbers which represent the compatibility between two adjacent edges. Such indicators are high in presence of collinear adjacent edge pixels and low elsewhere. The main task of relaxation labeling is to find a node labeling that maximizes a global objective function, called average consistency, which is high when long chains of collinear edges result in high values of $p$ and all other pixels are labeled with low values of $p$. This cost function offers a nice interpretation in terms of Markov models [39]. The simplest way to optimize the objective function is initialize the labels with the values of a local edge strength and to proceed iteratively with descent-gradient methods. In this way, at each iteration edge strength is reinforced on the underlying object contours and undesired responses are suppressed.

The main drawback of relaxation labeling is that, despite of a conspicuous amount of research [40], most of the existing algorithms do not converge to the global minimum of the cost function, but they get stuck in a local minima.

## 3. Proposed method

In this section, we first review surround suppression and illustrate the problem of self-inhibition (Section 3.1), then we describe the proposed inhibition term to avoid self-inhibition (Sections 3.2 and 3.3), and finally we present the proposed multilevel inhibition algorithm (Section 3.4).

### 3.1. Surround suppression and self-inhibition

Let $I(\mathbf{r})$ be an input graylevel image, with $\mathbf{r}=(x,y)\in\mathbb{R}^2$, and let $\nabla_\sigma I(\mathbf{r})$ be its Gaussian gradient, which is defined as the convolution between $I(\mathbf{r})$ and the gradient of a Gaussian function $g_\sigma(\mathbf{r})$:

$$\nabla_\sigma I(\mathbf{r}) \triangleq \{I \star \nabla g_\sigma\}(\mathbf{r}), \quad g_\sigma(x,y) \triangleq \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{1}$$

As well known, this is equivalent to computing the gradient of the convolution of $I(\mathbf{r})$ with $g_\sigma(\mathbf{r})$. In [41], the inhibition term is computed as a weighted local average of the Gaussian gradient magnitude $|\nabla_\sigma I(\mathbf{r})|$ over a ring around each pixel. Specifically, the following weighting function $w_\sigma(\mathbf{r})$, defined as a normalized difference of Gaussians, is considered:

$$w_\sigma(x,y) = \frac{1}{A}\left|\frac{1}{2\pi k^2\sigma^2} e^{-(x^2+y^2)/2k^2\sigma^2} - \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}\right|^+ \tag{2}$$

where $A$ is a normalization factor defined such that $\iint w_\sigma(x,y)\,dx\,dy = 1$, $k$ is the ratio between the scale parameters of the two Gaussians, whose optimal value has been found to be $k = 4$ [41], and the symbol $|\cdot|^+$ is defined as

$$|u|^+ \triangleq \begin{cases} u, & u > 0 \\ 0, & u \le 0 \end{cases} \tag{3}$$

The inhibition term $T(\mathbf{r})$ is computed as the convolution of the Gaussian gradient magnitude $|\nabla_\sigma(\mathbf{r})|$ with the inhibition filter $w_\sigma(\mathbf{r})$:

$$T(\mathbf{r}) \triangleq \{|\nabla_\sigma I| \star w_\sigma\}(\mathbf{r}) \tag{4}$$

The behavior of the inhibition term is illustrated in Fig. 2 for a synthetic input image. As we see, the gradient magnitude is strong both on the isolated edge in the top part of the image and

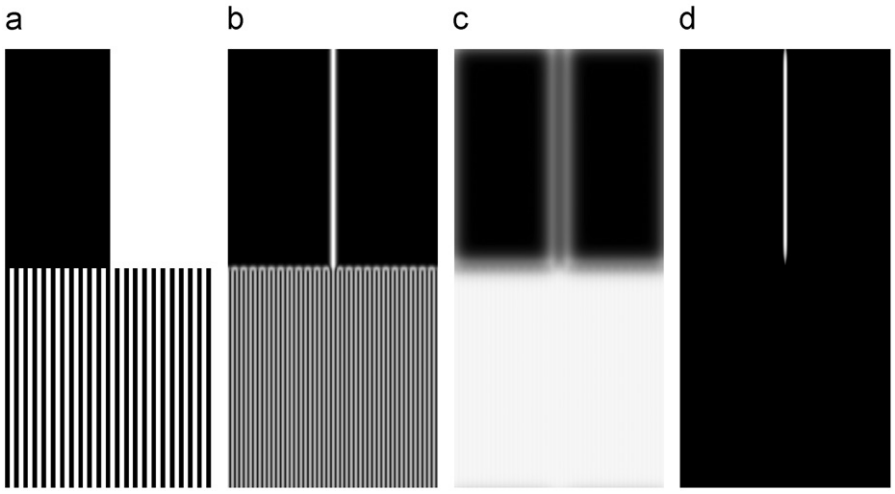


**Fig. 2.** Computational model of surround suppression proposed in [41]. (a) Synthetic input image, (b) gradient magnitude, (c) inhibition term $T(\mathbf{r})$, and (d) inhibited edge strength $c_\sigma^{(\lambda)}(\mathbf{r})$ for $\lambda=3$.

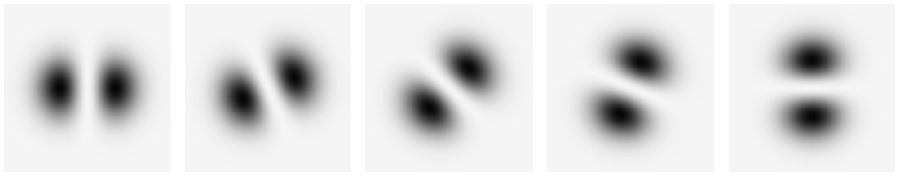

**Fig. 3.** From left to right: different rotated version $K_\theta(\mathbf{r})$ of the proposed inhibition kernel, in the spatial domain, for different values of $\theta$ ($\theta=0,\pi/8,\pi/4,3\pi/8,\pi/2$).

on the texture at the bottom, therefore $|\nabla_\sigma I(\mathbf{r})|$ is not sufficient to discriminate between the two patterns. In contrast, the inhibition term is much higher on texture than on isolated edges. Therefore, when the inhibition term is subtracted from the gradient magnitude, the resulting quantity has a strong response on isolated edges only. Specifically, the following quantity is considered in [41]:

$$c_\sigma^{(\lambda)}(\mathbf{r}) \triangleq g_\sigma(\mathbf{r}) - \lambda T(\mathbf{r}) \tag{5}$$

where the coefficient $\lambda$ is an input parameter called *inhibition level* which must be specified by the user. The higher the value of $\lambda$ is, the more edge responses are suppressed.

The main limitation of this approach is that, on isolated edges, the inhibition term computed as in (4) is not zero as it should be, but its value is about one-third of the value that $T(\mathbf{r})$ has on texture. This means that the suppression process will not only affect texture, but a considerable amount of isolated edges as well. This phenomenon is called *self-inhibition* [13] and it is due to the fact that part of the edge falls in the annular region defined by $w_\sigma(\mathbf{r})$. An example in which self-inhibition reduces the quality of the detected contour is shown in Fig. 1, top row, image on the left. As we see, some contour fragments of the elephant are not correctly detected. In the next subsections we present a different approach for the computation of the inhibition term $T(\mathbf{r})$, based on the theory of steerable filters, which does not suffer this limitation.

### 3.2. New inhibition term

To avoid self-inhibition, we introduce an orientation-selective inhibition kernel which completely excludes the central edge from the inhibition area. Specifically, we consider the following family of kernels $K_\theta(\mathbf{r})$, which continuously depend on the rotation angle $\theta$

$$K_\theta(\mathbf{r}) = H(R_\theta \mathbf{r}), \quad H(x,y) \triangleq x^2 e^{-(x^2+y^2)/2\sigma^2} \tag{6}$$

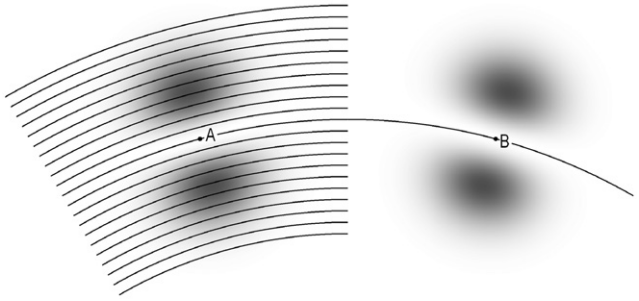


**Fig. 4.** Computation of the inhibition term for, respectively, the cases in which the central edge is surrounded by many other edges (point A) and isolated (point B). The area in which the inhibition kernel is significantly higher than zero is marked by the two lobes on the two sides of the central edge. In both cases, pixels of the central edge does not contribute to the proposed inhibition term, thus avoiding self-inhibition.

where $R_\theta$ is the rotation matrix with angle $\theta$. This kernel is shown in Fig. 3 for different values of the angle $\theta$. Also, let $t_\theta(\mathbf{r})$ be the result of the convolution of the Gaussian gradient magnitude $|\nabla_\sigma I(\mathbf{r})|$ of the input image with $K_\theta(\mathbf{r})$:

$$t_\theta(\mathbf{r}) = \{|\nabla_\sigma I| \star K_\theta\}(\mathbf{r}) \tag{7}$$

This quantity is a weighted local average of the gradient magnitude of the input image on a region which excluded the central edge when it is oriented orthogonally to $\theta$ and, consequently, avoids self-inhibition. Therefore, we define the new inhibition term $t(\mathbf{r})$ as the value of $t_\theta(\mathbf{r})$ in which $\theta$ is equal, for each pixel, to the local orientation $\theta_\nabla(\mathbf{r})$ of the gradient of the input image

$$t(\mathbf{r}) = t_{\theta=\theta_\nabla(\mathbf{r})}(\mathbf{r}) \tag{8}$$

As shown in Fig. 4, the edge on the central pixel is completely outside the inhibition area of the proposed filter, which would not happen

with the solution proposed in [41], in which the inhibition kernel is defined on an annular support.

### 3.3. Efficient implementation with steerable filters

Unlike the inhibition term $T(\mathbf{r})$ defined in Section 3.1, the quantity $t_\theta(\mathbf{r})$ cannot be computed with a single convolution. We present a simple implementation of $t_\theta(\mathbf{r})$, based on the theory of steerable filters, which requires two convolutions only.

#### 3.3.1. Background on steerable filtering

A family of filters $K_\theta(\mathbf{r}) = H(R_\theta \mathbf{r})$, which continuously depends on the parameter $\theta$, is *steerable* with order $N$ if it can be expressed as a linear combination of $N$ fixed basis $V_n(\mathbf{r})$, with coefficients $a_n(\theta)$ which depend on $\theta$:

$$K_\theta(\mathbf{r}) = \sum_n a_n(\theta) V_n(\mathbf{r}) \tag{9}$$

If $K_\theta(\mathbf{r})$ is steerable, its convolution $t_\theta(\mathbf{r})$ with an image $I(\mathbf{r})$ can be evaluated exactly for every $\theta$ by convolving $I(\mathbf{r})$ with the steering bases $V_n(\mathbf{r})$. Specifically, we have

$$\{I \star K_\theta\}(\mathbf{r}) = \sum_n a_n(\theta) \{I \star V_n\}(\mathbf{r}) \tag{10}$$

Let us express $\mathbf{r}$ in polar coordinates, $\mathbf{r} = (\rho, \phi)$, which implies $R_\theta \mathbf{r} = (\rho, \phi + \theta)$. It is easy to show that every family of filters expressed in the form

$$H(\mathbf{r}) = H(\rho, \phi) = \sum_n A_n(\rho) e^{i\omega_n \phi} \tag{11}$$

is steerable, where $A_n(\rho)$ are generic functions and $\omega_n$ are real numbers. In fact, we have

$$H(R_\theta \mathbf{r}) = H(\rho, \phi + \theta) = \sum_n e^{i\omega_n \theta} A_n(\rho) e^{i\omega_n \phi} \tag{12}$$

By comparing (9) and (12), we see that each steering base $V_n(\mathbf{r})$ is expressed as the product of a radial term $A_n(\rho)$ with a complex exponential $e^{i\omega_n \phi}$, and the coefficients $a_n(\theta)$ are complex exponential too, $a_n(\theta) = e^{i\omega_n \theta}$.

A corollary of this theorem is that every polynomial in $x$ and $y$, $P(x,y) = \sum_{j,k} a_{j,k} x^j y^k$, is steerable. In fact, by writing $x = \rho\cos\phi$ and $y = \rho\sin\phi$, and expressing $\cos\phi$ and $\sin\phi$ in terms of complex exponentials of $\phi$, it is easy to reduce $P(x,y)$ to the form (11). Moreover, the functions $A_n(\rho)$ are polynomials in $\rho$ of degree $n$.

#### 3.3.2. Steered inhibition filter

With basic algebra, it is easy to show that the filter $H(x,y)$ defined in (6) is steerable with order three. In fact, $H(x, y)$ is the product of the Gaussian term $e^{-(x^2+y^2)/2\sigma^2}$ with a polynomial. Since the Gaussian term does not depend on the angular coordinate, and the polynomial is steerable, we conclude that $H(x,y)$ is steerable too. Specifically, the steering bases $V_n(\mathbf{r})$, $(n = -2, 0, 2)$ are the following:

$$V_0(\rho,\phi) = \frac{\rho^2}{2}, \quad V_{\pm 2}(\rho,\phi) = \frac{\rho^2}{2} e^{\pm 2i\phi} \tag{13}$$

We notice that $V_{-2}(\rho,\phi) = \overline{V_2}(\rho,\phi)$, where $\overline{z}$ denotes the complex conjugate of $z$, which allows to simplify the expression of the steered inhibition filter as follows:

$$t(\mathbf{r}) = \{V_0 \star |\nabla I|\}(\mathbf{r}) + \mathrm{re}\{e^{2i\theta_\nabla(\mathbf{r})} [(V_2 \star |\nabla I|)(\mathbf{r})]\} \tag{14}$$

### 3.4. Multilevel inhibition

We now present the proposed approach for edge detection, which is depicted in Fig. 5. Unlike previous techniques based on surround suppression, the user does not need to specify the value of the inhibition level $\lambda$ defined in (5).

For an input image of $N$ pixels, let $b(p,\lambda)$ be the set of the $Np$ pixels which have the highest values of $c_\sigma^{(\lambda)}(\mathbf{r})$, where $p$ is the fraction of pixels of $b(p,\lambda)$ with respect to the total number of pixels of the input image. In Fig. 6, the sets $b(p,\lambda)$ are shown for different values of $\lambda$ by keeping $p$ constant, for the two input images of Fig. 1. The following facts can be observed:

- Undesired edges have small overlap across different values of $\lambda$, thus suggesting that the *intersection* of the $b_k \triangleq b(p,\lambda_k)$ will

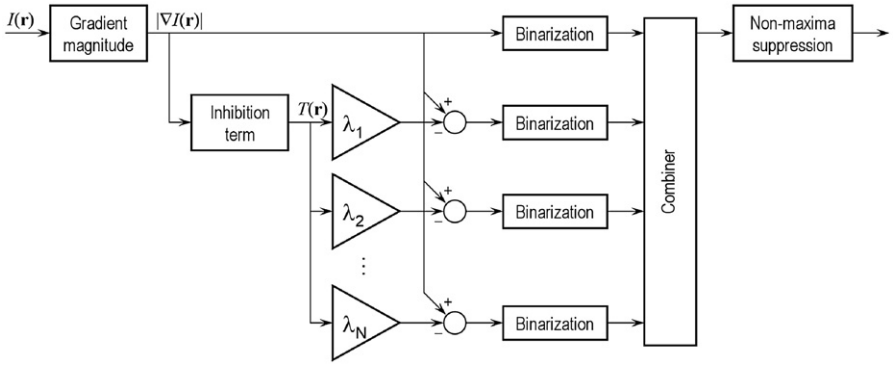


**Fig. 5.** Proposed approach for edge detection.

**Fig. 6.** Sets $b(p,\lambda_k)$ for $p = 0.1$ and for $\lambda_k = 0,1,2,3,4$.

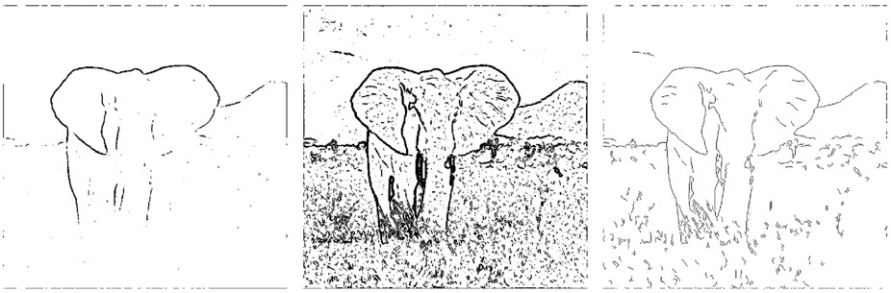

**Fig. 7.** From left to right: intersection and union of the sets $b(p,\lambda_k)$, and output of the proposed operator.

contain a very little amount of undesired responses. However, since some object contours will be missing for some $\lambda_k$, they will be missing in the intersection too (Fig. 7, left).

- Some object contours or parts of them which are missing for some values of $\lambda$ are well detected for other values of $\lambda$, thus suggesting that the union of the *union* of the $b_k$ will contain all object contours. However, much undesired response will be presented too (Fig. 7, center).

This suggests to look for some combination of different binary maps whose output is in between an intersection and a union. We propose a strategy based on the following observations:

- A considerable amount of undesired response can still be removed by intersecting a few binary maps $b_k$ instead of all of them. In this way, a smaller amount of good contours will be removed by the intersections.
- The binary maps $b_k$ associated to different inhibition levels will contain different parts of interesting contours. Therefore, after removing the majority of undesired response by means of intersections, their union will contain a larger amount of good responses w.r.t. each $b_k$.

In force of the above considerations, we combine the advantages of intersections and unions by the following output, which is implemented by the block "combiner" of Fig. 5:

$$B(p) = \bigcup_{k=0}^{N_\Lambda - 1} b(p,\lambda_k) \cap b(p,\lambda_{k+1}) \tag{15}$$

where the $N_\Lambda$ values of $\lambda_k$ are equally spaced in the range $[0,\lambda_{\max}]$.

As we see in Fig. 7, right, this procedure gives better results than both the intersection and the union of the $b(p,\lambda_k)$, as well as the best binary map associated to a single value of $\lambda_k$.

## 4. Experimental results

We have tested the proposed edge detector on a dataset of 40 natural images. In this section, we compare the output of the proposed algorithms with the outputs of the previous inhibition term and the standard Canny edge detector. Some experimental results are shown in Fig. 8, where all images have been generated with the same values of the input parameter. A larger set of examples is available online.[2] As we see, the proposed method outperforms all the others in terms of larger suppression of undesired responses and better preservation of low contrast contours.

To quantify the achieved performance improvement, we have measured the similarity between detected contours and hand-drawn ground truths. Ground truths are obtained by simply asking human observers to draw the set of lines that, according to them, should be regarded as contours in the concerned image. Several datasets of ground truth are available, we used the Rug dataset [41]. Despite a certain degree of subjectivity that is introduced, this procedure is widely deployed due to the large agreement among different observers [5]. We measure similarity with a ground truth in terms of the well established Pratt's figure of merit, which is defined as

$$F = \frac{1}{\max(\mathrm{card}\{DC\},\mathrm{card}\{GT\})} \sum_{\mathbf{x}\in DC} \frac{1}{1+\left[\frac{d_{GT}(\mathbf{x})}{d_0}\right]^2} \tag{16}$$

where $DC$ indicates the set of detected contours, $GT$ indicates the set of ground truth pixels, $\mathrm{card}\{X\}$ indicates the number of pixels of set $X$, $d_{GT}$ is the distance transform of $GT$, and $d_0$ a scale parameter. $F$ takes values in $(0,1]$, being equal to 1 iff $DC$ coincides with $GT$. The scale parameter $d_0$ controls the sensitivity of $F$ to differences between $GT$ and $DC$: for small values of $d_0$, $F$ is close to 1 only if $DC$ is very similar to $GT$, while for large values of $d_0$ larger differences between $GT$ and $DC$ can be tolerated. A certain tolerance is needed because of possible errors in tracing contours when ground truths are drawn. We used the standard value $d_0 = 2$, which is based on the reasonable assumption that humans can draw contours with a position accuracy of 2 pixels.

The values of $F$, averaged over 40 images, are plotted in Figs. 9 and 10. Specifically, in Fig. 9, the average values of $F$ are plotted for single level inhibition versus the inhibition level $\lambda$, both for the isotropic [41] and new inhibition terms. As we see, the inhibition term proposed here outperforms the previous one for all values of $\lambda$. We also see that, for the isotropic inhibition term, performance collapses for $\lambda > 2$. These are the values of $\lambda$ for which self-inhibition becomes more serious. In contrast, for the new inhibition term performance decreases less dramatic and self-inhibition is serious only for $\lambda > 2.5$.

In Fig. 10, the values of $F$ are plotted versus the fraction $p$ of pixels which survive thresholding, both for single and multilevel inhibition. This allows us to measure, for each method, the optimal value of $p$. As we see, the proposed method outperforms all others in terms of $F$. This plot relates to the new inhibition term; a similar plot could be obtained for the isotropic inhibition term, but with lower values of $F$.

In conclusion, both the new inhibition term and the multilevel inhibition scheme contribute to the improvement.

## 5. Discussion, summary and conclusions

Surround suppression is a mechanism in the human visual systems which significantly contributes to distinguish texture edges from object contours. Existing mathematical models suffer two drawbacks: self-inhibition and an input parameter called inhibition level. A previous attempt to solve self-inhibition was

[2] http://www.cs.rug.nl/~imaging/PR

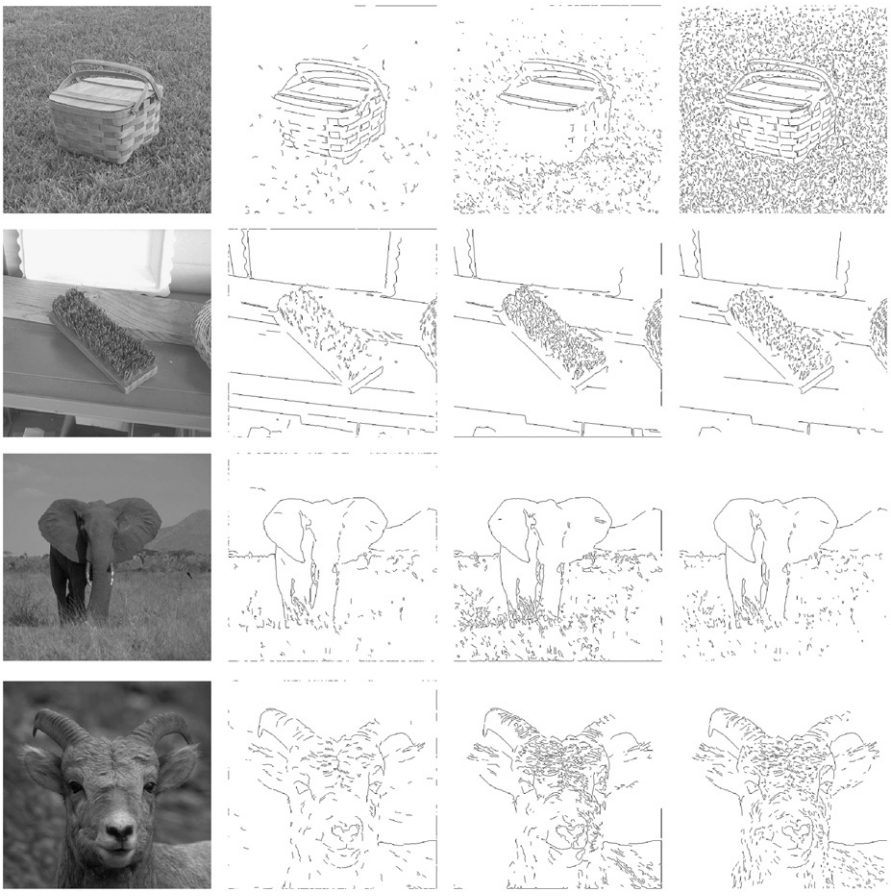

**Fig. 8.** From left to right: input images, output of the proposed operator, the Canny edge detector [18], and the single level inhibition approach proposed in [41].

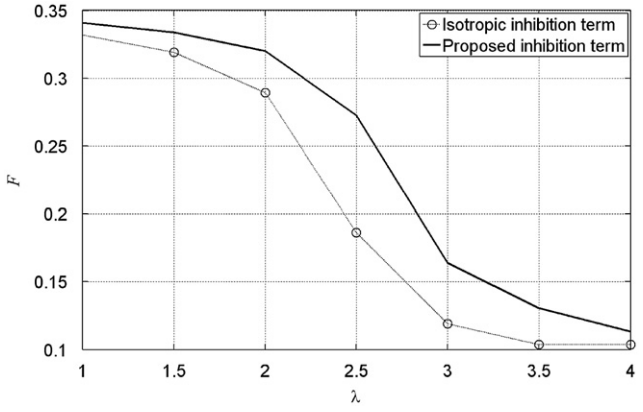


**Fig. 9.** Performance of the old (isotropic) and the new inhibition term.

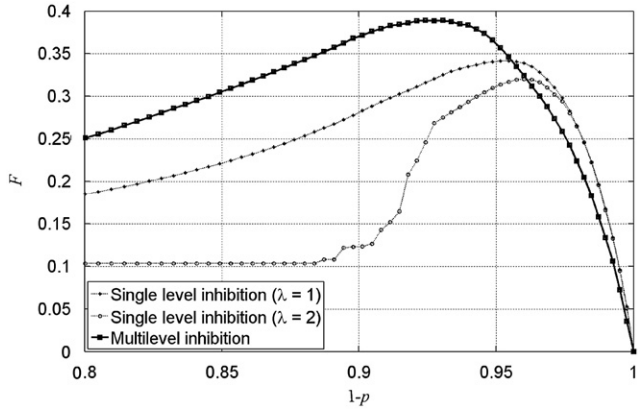


**Fig. 10.** Performance of single and multilevel surround suppression, for the proposed inhibition term.

proposed in [13]. However, it is affected by several limitations: first of all, a large number of convolutions is required, thus making the method computationally demanding; also, a discretization error on the local edge orientation is introduced; finally, it is based on the assumption that the value of the inhibition term is equal to the minimum between the two amounts of edges on the left and on the right side of the concerned edge, which is not supported by biological evidence. In contrast, we propose a method based on the theory of steerable filters which only requires two convolutions, it does not introduce discretization errors on the local edge orientation, and does not need special assumptions about the human visual system. Moreover, the inhibition kernel presented here introduces much less input parameters with respect to the technique developed in [13].

We also introduce a simple technique to combine the binary maps obtained with different values of the inhibition level, by taking advantage of both intersections and unions of different point sets. The benefits of this method are twofold: first, edge detection performance improves of about 15% with respect to single level inhibition; furthermore the user does not need to specify the inhibition level, thus making the method more unsupervised. The importance of being unsupervised with respect to the inhibition level is testified by the fact that the performance of single level approaches is strongly influenced by the value of $\lambda$ (see, e.g., Fig. 1).

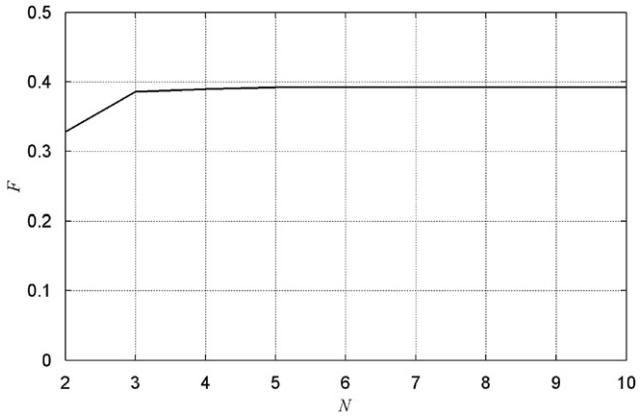


**Fig. 11.** Performance of the proposed method for different discretizations of $\lambda$.

One limitation of such a strategy is that it relies on heuristics rather than on general principles. For example, it is unclear why one should intersect only pairs of consecutive binary maps $b_k$ instead of triples or, more generally, $n$-ples. Our choice is mainly motivated by conceptual simplicity, while leaving a theoretical study of optimal solutions for future research.

The performance improvement of the proposed algorithm w.r.t. the isotropic single level surround suppression proposed in [41] is brought by (i) lessening self-inhibition by means of a new inhibition term and (ii) combining the advantages of both high and low inhibition levels through a new multilevel inhibition scheme. Quantitative performance evaluation shows that both techniques contribute significantly to the overall improvement w.r.t. previous methods (Figs. 9 and 10).

In the proposed multilevel inhibition scheme, $N_\Lambda$ evenly spaced values of the inhibition level $\lambda$ in the range $[0,\lambda_{max}]$ are considered. We now briefly discuss the influence of $N_\Lambda$ and $\lambda_{max}$. As for $\lambda_{max}$, we observe that for sufficiently high inhibition levels ($\lambda > 4$) practically all edges are inhibited, thus the corresponding binary maps are empty and they do not contribute to the union defined in (15). Therefore, choosing $\lambda_{max} \geq 4$ is equivalent to set $\lambda_{max} \to \infty$ and there is no particular preferred range for $\lambda$. Regarding the discretization of $\lambda$, we measured the performance of the proposed algorithm for different values of $N_\Lambda$ between 2 and 10 (Fig. 11). As we see, for $N_\Lambda \geq 4$ the performance stabilize to a constant value, thus this parameter is not significantly influent either.

In conclusion, the proposed approach is validated by a broad range of experimental results. Both qualitative and quantitative comparison show the superiority of the proposed method with respect to both the Canny edge detector and single level surround inhibition, in terms of both texture suppression and low contrast contour preservation.

## References

[1] B.M. ter Haar Romeny, Front-end Vision and Multi-scale Image Analysis, Kluwer Academic Publishers, 2002.
[2] D.H. Lim, S.J. Jang, Comparison of two-sample tests for edge detection in noisy images, Journal of the Royal Statistical Society. Series D (The Statistician) 51 (1) (2002) 21–30.
[3] M. Morrone, R. Owens, Feature detection from local energy, Pattern Recognition Letters 6 (1987) 303–313.
[4] P. Kovesi, Image features from phase congruency, Videre: Journal of Computer Vision Research 1 (3) (1999) 26.
[5] D.R. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5) (2004) 530–549.
[6] S. Sarkar, P. Soundararajan, Supervised learning of large perceptual organization: graph spectral partitioning and learning automata, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (5) (2000) 504–525.
[7] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, International Journal of Computer Vision 1 (4) (1987) 321–331.
[8] H. Li, A.J. Yezzi, Local or global minima: flexible dual-front active contours, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 1–14.
[9] R.A. Hummel, S.W. Zucker, On the foundation of relaxation labeling process, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (3) (1983) 267–286.
[10] P. Parent, S.W. Zucker, Trace inference, curvature consistency, and curve detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 11(8) (1989) 823–839.
[11] F. Bergholm, Edge focusing, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (6) (1987) 726–741.
[12] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour detection based on nonclassical receptive field inhibition, IEEE Transactions on Image Processing 12 (7) (2003) 729–739.
[13] G. Papari, P. Campisi, N. Petkov, A. Neri, A biologically motivated multiresolution approach to contour detection, EURASIP Journal on Advances in Signal Processing, Special issue on Human Perception 2007 (2007) 28.
[14] G. Papari, P. Campisi, N. Petkov, Contour detection by multiresolution surround inhibition, in: IEEE International Conference on Image Processing, 2006, pp. 749–752.
[15] J. Vriendt, Accuracy of the zero crossings of the second directional derivative as an edge detector, Multidimensional Systems and Signal Processing 4 (3) (1993) 227–251.
[16] J.J. Clark, Singularity theory and phantom edges in scale space, IEEE Transactions on Pattern Analysis and Machine Intelligence 10 (5) (1988) 720–727.
[17] M. Bertero, T.A. Poggio, V. Torre, Ill-posed problems in early vision, Proceedings of IEEE 76 (8) (1988) 869–889.
[18] J.F. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6) (1986) 679–698.
[19] S. Sarkar, K.L. Boyer, On optimal infinite impulse response edge detection filters, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (11) (1991) 1154–1171.
[20] D.R. Demigny, T. Kamlé, A discrete expression of canny's criteria for step edge detector performances evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (11) (1997) 1199–1211.
[21] E. Chuang, D. Sher, $\chi^2$ test for feature detection, Pattern Recognition 26 (11) (1993) 1671–1681.
[22] M.A. Ruzon, C. Tomasi, Edge, junction, and corner detection using color distributions, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1281–1295.
[23] M. Nitzberg, D. Mumford, T. Shiota, Filtering Segmentation and Depth, Lecture Notes in Computer science, vol. 662, Springer, 1993.
[24] S. Ando, Image field categorization and edge/corner detection from gradient covariance, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2) (2000) 179–190.
[25] P.H. Gregson, Using angular dispersion of gradient direction for detecting edge ribbons, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (7) (1993) 682–696.
[26] J.B. Martens, Local orientation analysis in images by means of the hermite transform, IEEE Transactions on Image Processing 6 (8) (1997) 1103–1116.
[27] M.C. Morrone, D. Burr, J. Ross, R. Owens, Mach bands are phase dependent, Nature (1986) 250–253.
[28] D. Boukerroui, J.A. Noble, M. Brady, On the choice of band-pass quadrature filters, Journal of Mathematical Imaging and Vision 21 (1) (2004) 53–80.
[29] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America A 2 (1985) 1160–1169.
[30] R. Duda, P. Hart, D. Stork, Pattern Classification, John Wiley and Sons, 2001.
[31] S. Konishi, A.L. Yuille, J.M. Coughlan, S.C. Zhu, Statistical edge detection: learning and evaluating edge cues, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003) 57–74.
[32] P. Perona, J. Malik, Detecting and localizing edges composed of steps, in: International Conference on Computer Vision, 1990, pp. 52–57.
[33] D.R. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using brightness and texture, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Neural Information Processing Systems, MIT Press, 2002, pp. 1255–1262.
[34] N. Petkov, M.A. Westenberg, Suppression of contour perception by band-limited noise and its relation to non-classical receptive field inhibition, Biological Cybernetics 88 (2003) 236–246.
[35] S.C. Yen, L.H. Finkel, Extraction of perceptually salient contours by striate cortical networks, Vision Research 38 (5) (1998) 719–741.
[36] Z. Li, A neural model of contour integration in the primary visual cortex, Neural Computation 10 (4) (1998) 903–940.
[37] G. Guy, G. Medioni, Inferring global perceptual contours from local features, International Journal of Computer Vision 20 (1) (1996) 113–133.
[38] M.S. Lee, G. Medioni, Grouping …into regions, curves, and junctions, Computer Vision and Image Understanding 76 (1) (1999) 54–69.
[39] S.Z. Li, H. Wang, M. Petrou, Relaxation labeling of Markov random fields, Proceedings of IEEE International Conference on Pattern Recognition, vol. 1, 1994, pp. 488–492.

[40] T. Werner, A linear programming approach to max-sum problem: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (7) (2007) 1165–1179.

[41] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour and boundary detection improved by surround suppression of texture edges, Image and Vision Computing 22 (8) (2004) 609–622.

**Giuseppe Papari** was born in Rome in 1979. He got his Master degree in Electrical Engineering in 2003 from "Universita' degli Studi di Roma Tre" (University of Rome III) and his Ph.D. degree in Computing Science in 2009 from the University of Groningen, both of them with cum laude distinction. Since 2009 he is a postdoctoral researcher at the Department of Mathematics and Computing Science of the University of Groningen.

He has published several papers about biologically motivated contour detection, perceptual grouping, multiresolution analysis, and artistic imaging, one of which was invited to CODEC 2006. His main research interests are nonlinear filtering, contour detection, image segmentation, multiresolution analysis, clustering, pattern recognition, artistic imaging, and morphological analysis.

Dr. Giuseppe Papari received one of the four IBM best student paper awards at the IEEE International Conference on Image Processing in 2006.

**Nicolai Petkov** is professor of computer science and head of the Research Institute of Mathematics and computing Science of the University of Groningen. He got the Dr.sc.techn. degree in Computer Engineering (Informationstechnik) from Dresden University of Technology. Prior to joining the University of Groningen in 1991, he held research positions at the University of Wuppertal, the University of Erlangen-Nürnberg, the Academy of Sciences in Berlin, and Dresden University of Technology. In 1989 he was awarded an Alexander von Humboldt scholarship of the Federal Republic of Germany.

Nicolai Petkov is the author of two monographs and (co-)author of another book, four patents and over 100 scientific papers. He is member of the editorial boards of several journals. His current research focusses on computer simulations and understanding of the visual system of the brain and using the obtained insights for the development of effective computer vision algorithms. He is also interested in using computers for artistic expression.