An improved model for surround suppression by steerable filters and multilevel inhibition with application to contour detection
Papari, Giuseppe; Petkov, Nicolai

Published in:
Pattern recognition

DOI:
10.1016/j.patcog.2010.08.013

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
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

A R T I C L E   I N F O
Available online 12 August 2010
Keywords:
Contour detection
Surround suppression
Steerable filters

A B S T R A C T
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.

© 2010 Elsevier Ltd. All rights reserved.

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. 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
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 closer 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 \( \Lambda(r) \) around each pixel \( r \) by means of statistical tools. The simplest technique consists in dividing \( \Lambda(r) \) 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 \( L(t) \equiv \sqrt{(x \cdot f_{\phi}(t))^2 + (x \cdot f_{\phi}(t))^2} \), called local energy. Here, \( f_{\phi}(t) \) and \( f_{\phi}(t) \) are a symmetric and an anti-symmetric low-pass or band-pass filter, such that \( f_{\phi}(t) \) is the Hilbert transform of \( f_{\phi}(t) \). Several pairs of functions \( f_{\phi}(t) \) and \( f_{\phi}(t) \), known in the

![Image](https://example.com/image.png)

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.
both saliency values and local edge orientation are obtained from
based on (i) the local edge strength and orientation in
the performance improvement brought by combining the follow-
ing 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
adventages. Such indicators are high in presence of collinear
adjacent edges and low elsewhere. The main task of
results. 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 inhibition term $j_j$ the Gaussian gradient magnitude
of the two Gaussians, whose optimal value has been found to be
4 | [41], and the symbol $| \cdot |^+$ is defined as
$$|u|^+ \triangleq \begin{cases} u, & u > 0 \\ 0, & u \leq 0 \end{cases}$$
(3)
The inhibition term $T(r)$ is computed as the convolution of the
Gaussian gradient magnitude $|\nabla_g(r)|$ with the inhibition
filter $w_i(r)$:
$$T(r) \triangleq |\nabla_g|^+ w_i(r)$$
(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
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 $I(r)$ be an input graylevel image, with $r = (x,y) \in \mathbb{R}^2$, and
let $\nabla_g I(r)$ be its Gaussian gradient, which is defined as the
convolution between $I(r)$ and the gradient of a Gaussian function $g_0(r)$:
$$\nabla_g I(r) \triangleq (\nabla g_r)(r), \quad g_0(x,y) \triangleq \frac{1}{2\pi\sigma^2} e^{-\left((x^2+y^2)/2\sigma^2\right)}$$
(1)
As well known, this is equivalent to computing the gradient of the
convolution of $I(r)$ with $g_0(r)$. In [41], the inhibition term is
computed as a weighted local average of the Gaussian gradient
magnitude $|\nabla_g I(r)|$ over a ring around each pixel. Specifically, the
following weighting function $w_i(r)$, defined as a normalized
difference of Gaussians, is considered:
$$w_i(x,y) = \frac{1}{A} \left[ \frac{1}{2\pi(k^2 \sigma^2)} e^{-\left((x^2+y^2)/2(k^2 \sigma^2)\right)} - \frac{1}{2\pi\sigma^2} e^{-\left((x^2+y^2)/2\sigma^2\right)} \right]^+$$
(2)
where $A$ is a normalization factor defined such that
$$\int w_i(x,y) \, dx \, dy = 1, \quad k \text{ is the ratio between the scale parameters of the two Gaussians, whose optimal value has been found to be}$$
$$k = 4 \quad [41], \quad \text{and the symbol } | \cdot |^+ \text{ is defined as}$$
$$|u|^+ \triangleq \begin{cases} u, & u > 0 \\ 0, & u \leq 0 \end{cases}$$
(3)
on the texture at the bottom, therefore \( |\nabla r| \) 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]:

\[
c_{\lambda}(r) = s(r) - \lambda T(r)
\]

where the coefficient \( \lambda \) is an input parameter called the 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 \( T(r) \) 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 \( w_{\pi}(r) \). 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 \( T(r) \), 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 \( K_{\theta}(r) \), which continuously depend on the rotation angle \( \theta \)

\[
K_{\theta}(r) = H(R_{\theta}r), \quad H(x,y) = x^2e^{-(x^2+y^2)/2\sigma^2}
\]

where \( R_{\theta} \) is the rotation matrix with angle \( \theta \). This kernel is shown in Fig. 3 for different values of \( \theta \) (\( \theta = 0, \pi/8, 3\pi/8, \pi/2 \)). 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 do not contribute to the proposed inhibition term, thus avoiding self-inhibition.

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

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 \( t(r) \) as the value of \( t_{\theta}(r) \) in which \( \theta \) is equal, for each pixel, to the local orientation \( \theta_v(r) \) of the gradient of the input image

\[
t(r) = t_{\theta_v(r)}(r)
\]

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...
3.3. Efficient implementation with steerable filters

Unlike the inhibition term \( T(r) \) defined in Section 3.1, the quantity \( t_0(r) \) cannot be computed with a single convolution. We present a simple implementation of \( t_0(r) \), based on the theory of steerable filters, which requires two convolutions only.

3.3.1. Background on steerable filtering

A family of filters \( K_0(r) = H(K,r) \), 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 \( V_0(r) \), with coefficients \( a_n(\theta) \) which depend on \( \theta \):

\[
K_0(r) = \sum_{n} a_n(\theta) V_n(r) \tag{9}
\]

If \( K_0(r) \) is steerable, its convolution \( t_0(r) \) with an image \( I(r) \) can be evaluated exactly for every \( \theta \) by convolving \( I(r) \) with the steering bases \( V_n(r) \). Specifically, we have

\[
[I \ast K_0](r) = \sum_{n} a_n(\theta) [I \ast V_n](r) \tag{10}
\]

Let us express \( r \) in polar coordinates, \( r = (\rho, \phi) \), which implies \( K_0(r) = (\rho, \phi, \theta) \). It is easy to show that every family of filters expressed in the form

\[
H(r) = H(\rho, \phi) = \sum A_n(\rho) e^{i\omega_n \phi} \tag{11}
\]

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

\[
H(K, r) = H(\rho, \phi + \theta) = \sum A_n(\rho) e^{i\omega_n \phi} \tag{12}
\]

By comparing (9) and (12), we see that each steering base \( V_n(r) \) 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 \( y \), \( P(x,y) = \sum_{n} a_n x^n y^n \), 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 \( e^{-\alpha^2 + y^2/2\sigma^2} \) 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 \( V_n(r) \), \( n = -2, 0, 2 \) 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 \( V_{\pm 2}(\rho, \phi) = V_{\mp 2}(\rho, \phi) \), where \( z \) denotes the complex conjugate of \( z \), which allows to simplify the expression of the steered inhibition filter as follows:

\[
t(r) = |V_0 \ast |V(I)||r| + r e^{2i \omega_n \phi} |\{V_{\pm 2} \ast |V(I)||r|\}| \tag{14}
\]

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 \( N \) pixels which have the highest values of \( c^p(\lambda) \), 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 \( b_k = b(p, \lambda_k) \) will
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:

\[
B(p) = \bigcup_{k=0}^{N\lambda - 1} b(p, \lambda_k) \cap b(p, \lambda_{k+1})
\]

(15)

where the \( N\lambda \) values of \( \lambda_k \) are equally spaced in the range \([0, \lambda_{\text{max}}] \).

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

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

(16)

where \( \text{DC} \) indicates the set of detected contours, \( \text{GT} \) indicates the set of ground truth pixels, \( \text{card}(\text{X}) \) indicates the number of pixels of set \( X \), \( d_{\text{GT}} \) is the distance transform of \( \text{GT} \), 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 \( \lambda > 2.5 \). 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
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. 8.](image1.png)

**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.](image2.png)

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

![Fig. 10.](image3.png)

**Fig. 10.** Performance of single and multilevel surround suppression, for the proposed inhibition term.
We now briefly discuss the influence of observe that for sufficiently high inhibition levels choosing and they do not contribute to the union defined in (15). Therefore, all edges are inhibited, thus the corresponding binary maps are empty particular preferred range for inhibition, in terms of both texture suppression and low contrast comparison show the superiority of the proposed method with range of experimental results. Both qualitative and quantitative 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.

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, $N_1$ evenly spaced values of the inhibition level $\lambda$ in the range $[0, \lambda_{\text{max}}]$ are considered. We now briefly discuss the influence of $N_1$ and $\lambda_{\text{max}}$. As for $\lambda_{\text{max}}$, 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 $\lambda_{\text{max}} \geq 4$ is equivalent to set $\lambda_{\text{max}} \rightarrow \infty$ 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 $N_1$ between 2 and 10 (Fig. 11). As we see, for $N_1 \geq 4$ the performance stabilize to a constant value, thus this parameter is not significantly influent either.

References

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.