

Contour detection model with multi-scale integration based on non-classical receptive field

Hui Wei*, Bo Lang, Qingsong Zuo

Cognitive Model and Algorithm Laboratory, Department of Computer Science and Technology, Fudan University, Shanghai, China

ARTICLE INFO

Article history:

Received 17 January 2012

Received in revised form

21 June 2012

Accepted 9 September 2012

Communicated by Pingkun Yan

Available online 26 October 2012

Keywords:

Non-classical receptive field

Contour detection

Neural modeling

ABSTRACT

The broad region outside the classical receptive field (CRF) of a vision neuron, known as the non-classical receptive field (nCRF), exerts a robust modulatory effect on the responses to visual stimuli presented within the CRF, and plays an important role in visual information processing. One possible role for the nCRF is the extract object contours from disorderly background textures. In this study, a multi-scale integration based contour extraction model, inspired by the inhibitory and disinhibitory interactions between the CRF and the nCRF is presented. Unlike previous models, our model not only includes both the simple and complex cell mechanisms but also introduces pre-processing of the external information by the retinal ganglion cells at an early stage. The multi-scale representation of a physical scene acquired through such pre-processing was filtered through Gabor filters, and then inhibited or disinhibited at different spatial locations on different scales until a final response was obtained. Our results show that by introducing this kind of mechanism into the model, numbers of non-meaningful texture elements can be removed significantly, while at the same time, the object contours can be detected effectively. In addition to the superior contour detection performance in comparison to other contour detection models, our model provides a better understanding of the role of the nCRF and a novel approach for computer vision and pattern recognition.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

A contour is a prominent structural feature and an important element for target recognition based on shape. In low-level vision, contour integration is considered to be a special process that integrates the sensory information via a lateral interaction in the primary visual cortex. The basic elements of any image can be represented by point-shaped data. An object corresponding to these point-shaped data is described using features such as shape, size, brightness, color and geometrical structure, which, when combined, characterize the object. Establishing the correspondence between the point-shaped data from the input and the integrated features of the object requires a process to integrate the data. Compared to edge detection, contour detection needs deeper levels of image information of, and the key is the integration of local information into coherent global features based on the context. The human visual system is able to quickly and effectively extract topographic features from an image through low-level processing, and can integrate local characteristic information at low levels into high-level entities. This capability helps

us to make appropriate decisions and reduce the post-processing complexity when individuals are faced with specific visual scenes. This vision characteristic provides a good model for a contour integration algorithm. The purpose of contour detection is to isolate a target object from a noisy background, and to determine which of the edges belong to the object and which belong to the background [1]. Over the past 20 years, considerable amount of research has been conducted in this field [2–12]. A natural scene image contains exceedingly rich information, and it is impossible for the visual system to afford the same level of attention to every point in the scene space. The human visual system is capable of acquiring relative integration and structure from scene images without any prior knowledge of the objects in the scene; this ability is called perception. Witkin has shown that structures that occupy a relatively large region of the visual space are especially conspicuous to vision systems [13]. The brain does not directly obtain the projected images on the retina presented by the external stimuli. Instead, it recognizes objects from visual information that has been aggregated and specialized. The main function of the visual cortex is to perform the extraction and calculation of perceptual signals, in a way that dramatically reduces the data volume but retains the useful structural information of the objects. In a complex natural scene, the components that have abundant structural information and specific attributes are more prominent, and are thus easier to isolate from the

* Corresponding author.

E-mail addresses: weihui@fudan.edu.cn, 09110240025@fudan.edu.cn, 09210240055@fudan.edu.cn (H. Wei).

background. This provides a physiologically plausible basis to develop computer algorithms for contour detection.

Neurophysiological studies have shown that in the visual pathway, from the ganglion cell (GC) projection first to the lateral geniculate nucleus (LGN) and then to the visual cortex, the information being processed along this pathway becomes more complex at higher levels. The receptive field (RF) (i.e., the region of space in which a visual stimulus can elicit a neuronal response) of cells along the pathway also becomes larger at the higher levels. This is the well-known “receptive field hierarchy” hypothesis. The areas surrounding the classical receptive field (CRF) are called the non-classical receptive field (nCRF). Although stimuli presented in the nCRF alone could not elicit neuronal responses, they could modulate the neural activities in response to stimuli in the CRF. This modulation allows the neurons to integrate information and participate in complex perceptual tasks [14–16]. Also, the CRF and nCRF of most visual neurons in the visual cortex are orientation selective [17,18]. The orientation selectivity of the nCRF may vary, and may modulate the CRF response differentially for different stimuli. For each single neuron, the output depends on the interaction between its nCRF and its CRF. In most cases, the modulation of the nCRF is inhibitory, and the horizontal connections between the neurons play an important role here [19]. About 80% of the orientation-selective neurons in the visual cortex show such inhibitory effects; in about 30% of these orientation-selective neurons, the inhibitory effect is weakest when the orientation of the stimuli in the surrounding area is orthogonal to that in the center [14,17,20]. This suggests that the nCRF is not just a simple expansion of the CRF. Instead, when combined with the CRF, they form a feature detector. This interaction between the CRF and the nCRF makes the fast, effortless detection of object contours possible in vision neurons.

Grigorescu et al. proposed two models for RF inhibition: the isotropic and anisotropic RFs [21,22]. With the use of techniques such as multi-scale method and denoising, the algorithm was further developed and became more robust [23,24]. The introduction of a multi-layer attention mechanism in addition to lateral inhibition allowed the model to obtain image contours with different resolutions [25,26]. In the version of the RF model with a butterfly-shaped inhibition surround, different inhibitory effects on the two ends and the two sides combined to eliminate texture information [27,28]. An iterative detection algorithm was also proposed in [29]. All of the above models have the following characteristics. (1) A Gabor function is used to represent the CRF, and a circular region is used to represent the nCRF. Compared to the Canny algorithm [6], these models can more effectively isolate the contours of target objects from a complex texture background. However, the circular shape of the RF can produce negative effects from auto-inhibition, which often leads to weak contour information being ignored, and in turn cause the “broken contour” phenomenon. This is particularly true when the target object is embedded in a noisy background [24,29]. (2) The algorithms of these models are all based on the RF properties of the orientation-selective simple and complex cells in the visual cortex. However, from a visual processing perspective, the external information reaches the visual cortex after it first arrives at the retina and then at the LGN. The GC is the first stage of information processing, and the RFs of GCs are relatively sensitive to changes in the spatial luminance, as shown in Fig. 1. The combined output from the output of GCs and their RF locations can therefore be used to detect the existence of edges. (3) Some of these algorithms do not take the multi-scale information of the image into account. In fact, multi-scale analysis can play an important role in contour detection models [30,25,31–33,24,34]. In the algorithm proposed in [27], two types of scale information, the coarse scale and the fine scale, were used. The Gabor energy

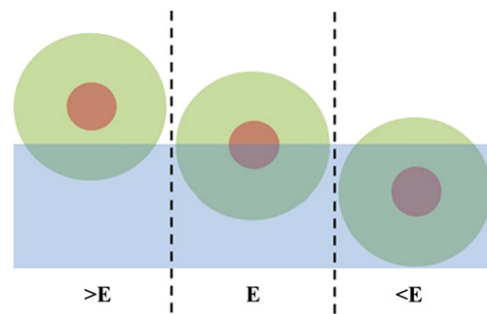


Fig. 1. Neuronal response to a luminance-change signal in the RF of a GC. When the luminance edge passes through the center of RF, the GC output is E; when the luminance edge is located elsewhere, the GC output is either higher or lower than E.

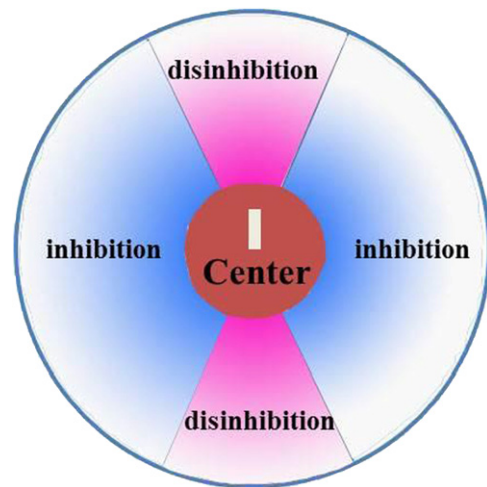


Fig. 2. The RF structure [37]. The regions on the sides along the preferred orientation axis are the disinhibitory regions (shown in pink); the regions on the sides orthogonal to the preferred orientation axis are the inhibitory regions (shown in blue). The shades of the colors indicate the magnitude of inhibitory/disinhibitory effects; the closer the area is to the RF center, then the stronger the effect becomes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was used to represent the coarse scale information, while the inhibitory information acts on a fine scale. This choice of scale was based on the assumption that the contour information of the target object was all on the coarse scale, whereas the texture information was all on the fine scale. However, division of the information contained in any natural image into two scales is too simplistic an approximation and would lead to information loss. Selection of the proper scale can maximally retain the object's contour information while removing texture information.

One study divided the RF of the neurons in the V1 area into three parts: the center, the surround and the extended surround. The surround area is inhibitory, whereas the extended surround area is disinhibitory [35]. When the cell responds to an external stimulus in its RF, it is not only inhibited by stimulation in the surround area, but is also influenced by the disinhibitory effect of the extended surround area. With the introduction of this disinhibitory mechanism, weak contours are strengthened, which helps to retain the completeness of the contour and increase the effectiveness of target detection. It has been proposed that the RF inhibitory regions are located at one of the two sides orthogonal to the axis of preferred orientation. Stimuli in the inhibitory regions will inhibit the neural response to similar center stimuli. In contrast, stimuli in the regions along the side of the preferred orientation will enhance the response to similar center stimuli,

and these regions are considered to be the disinhibitory area [15]. With regard to the spatial effects, the magnitude of the influence of the surround (including the inhibitory area and the disinhibitory area) on the response to center stimuli is inversely proportional to the distance between the surround and the center [15,14,36]. Based on the physiological evidence above, we propose the RF structure shown in Fig. 2.

The integrated multi-scale contour detection model proposed here is based on the nCRF mechanism. In contrast to previous contour detection models, our model adopts the strategy of combining the inhibitory and disinhibitory mechanisms. The inhibitory regions are on the two sides along the preferred orientation of the RF, while the disinhibitory regions are on the two sides orthogonal to the preferred orientation. More specifically, in accordance with the characteristics of the human visual

pathway, the external information is first processed by the GCs and the scale information is acquired in this step. In general, large-scale shape or contour information is detected, and the majority of the texture information (the high-frequency components in the image) is filtered out. The involvement of the scale information greatly reduces the influence of the irrelevant elements in the environment. Then, with the effects of distance, relative position and angular differences from the preferred orientation taken into full consideration, accurate contour information is obtained via the processing of the visual cortical neurons.

The rest of this paper is organized as follows. The proposed biologically motivated model is explained in Section 2. The experimental results and performance evaluation are then presented in Section 3. Finally, in Section 4, we explain why our model can produce superior contour detection results.

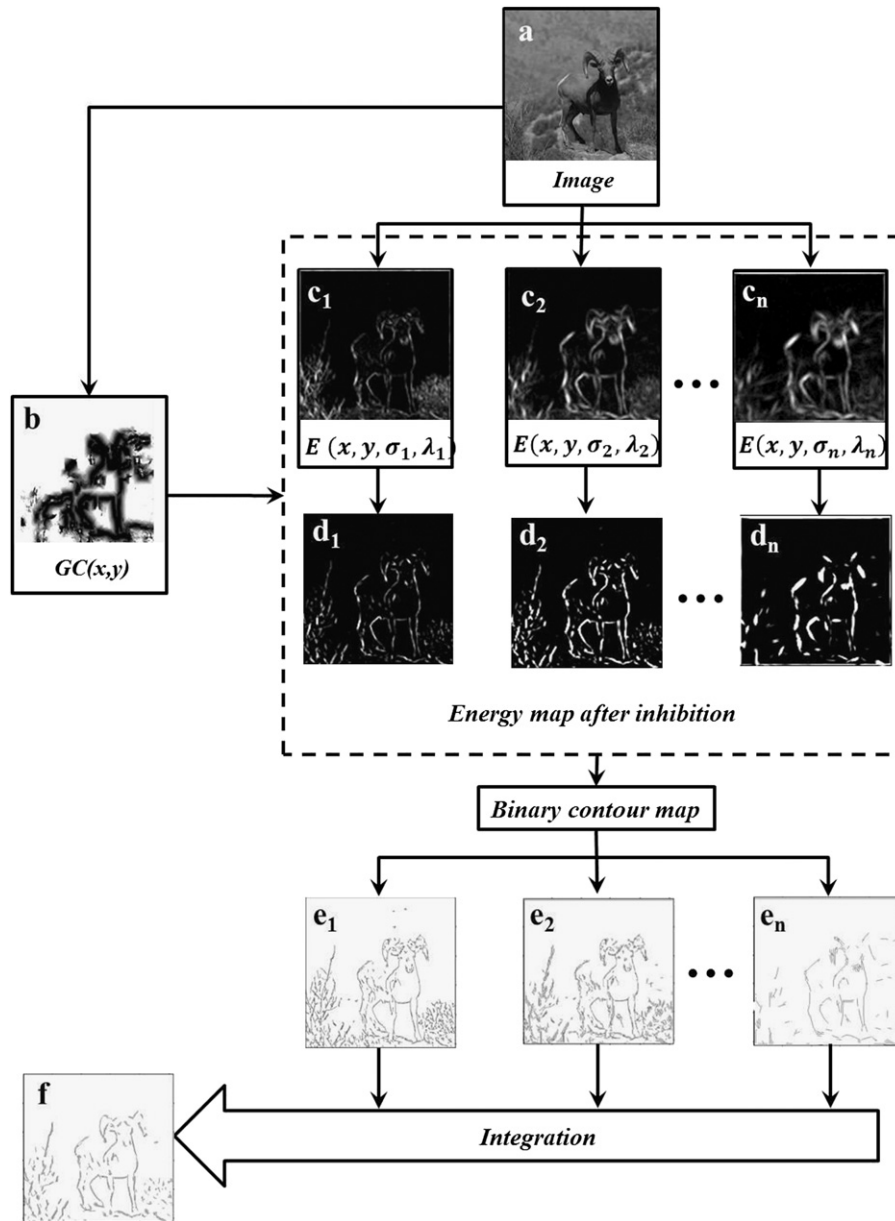


Fig. 3. (a) Original image. (b) $GC(x, y)$ is the result of pre-processing by retinal ganglion cells. (c_1 – c_n) Maximum Gabor energy maps on different scales. (d_1 – d_n) Energy maps after inhibition and disinhibition. (e_1 – e_n) Binary contour maps constructed by binarization based on the energy maps after inhibition. Note that the white part of the image is represented by the maximum RF size (coarse scale), i.e. the area that is most likely to contain textures; the dark part of the image is represented by the minimal RF size (fine scale). (f) Final contour map after multi-scale integration.

2. Design of the algorithm model

A good contour detection model should produce less inhibition in the regions containing contour information and greater inhibition in the regions containing texture information. To achieve this goal, two problems must be addressed. The first is how to distinguish the contour information of the target object from its texture information. The second is how to effectively control the magnitude of the inhibition to make the contour information prominent when it is present. In practice, it is contradictory in itself to increase texture information inhibition while decreasing contour information inhibition within the same image. To better solve this problem, we proposed a method that used multi-scale integration to determine the final contours of the target object. Unlike previous models, our model not only included both simple and complex cell mechanisms, but also introduced pre-processing of the external information by the GCs at an early stage. The multi-scale representation of the physical scene acquired from this pre-processing was filtered using Gabor filters and was then inhibited or disinhibited at different spatial locations on different scales until the final response was obtained. Because the contour information acquired on different scales is different, production of the optimal smooth and continuous contour information required all of the contour information that was acquired on different scales to be superimposed and integrated. The output of the model was the result of the final integration. A detailed flowchart of the algorithm used is shown in Fig. 3. To allow for easier comparison with other algorithm experiments, we performed binarization processing of the data using the non-maximum suppression and lag threshold methods of the Canny algorithm [6]. For comparison purposes, the two thresholds required in this process, t_l and t_h , were set as follows: $t_l = 0.5 \times t_h$.

2.1. Algorithm design

Let $E(x,y)$ be the intensity-response function of RF:

$$E(x,y) = \iint E_{\text{center}}(x,y) dx dy - \iint W_{\text{in}}(x,y) * E_{\text{in}}(x,y) dx dy + \iint W_{\text{disin}}(x,y) * E_{\text{disin}}(x,y) dx dy \quad (1)$$

W_{in} and W_{disin} are the weight functions of the inhibitory and disinhibitory regions, respectively. The specific definitions of these functions will be discussed later in the paper. Eq. (1) shows that the final response intensity of the RF is determined by the response intensity to the stimuli in the CRF together with the inhibitory and disinhibitory intensities of the nCRF.

Fig. 3 shows the flowchart of the algorithm. The information different obtained on scales was different, so the resulting contours were also different. Most of the information from the smallest scale was retained (see Fig. 3e₁), but most of its information was not related to the contours. On the largest scale, the main contour of the target object was retained (see Fig. 3e_n), but most of its details were lost, leading to the “broken contour” phenomenon. In practice, it is extremely difficult to choose an appropriate scale to maximally retain the contour information and simultaneously remove the texture information. To solve this problem, our algorithm adopted a multi-scale integration strategy in the final step. Fig. 3f shows the final result after multi-scale integration.

2.2. Multi-scale information obtained from changes in the retina ganglion cell RF size

There are over 100 million photoreceptor cells (RC) but only about one million GC on the retina. On average, every GC receives

inputs from 100 to 150 RCs. The magnitude of the electrical pulse generated by each RC is proportional to the light intensity that it receives; the RF of each GC projects to a certain region of the physical scene, and the radius of this region is generally proportional to the distance from the object to the retina. The image information that reaches our brain therefore depends not only on the light intensity distribution of the physical scene, but also on its distance from our eyes. The resolution of this image information is inversely proportional to the radii of the RFs. For GCs with RFs of the same size to acquire multi-scale representation of the physical scene, the distance from the scene to the eye must change constantly. However, such distance changes are neither realistic nor possible in a biological sense. From daily life, we know that the shape of a tree can be recognized on a coarse scale, while the leaves must be observed on a fine scale. Objects with sizes that are smaller than trees and larger than leaves are recognizable on an intermediate scale. The most important issue of scale-space theory is how to effectively establish the corresponding projections for the same physical scene on different scales. For this purpose, the brain has developed GCs with RFs of different sizes. Because the physical scene images acquired by RFs of different sizes correspond to the images acquired by a single RF at different distances, the image information that the brain obtains is a multi-scale representation of the physical scene. Outside the CRFs of these GCs, there are also the nCRFs [38,39]. These nCRFs can compensate for the loss of the low-frequency components caused by the antagonistic center-surround interaction in the CRFs [40]. Based on previous studies [40–42], the response profile of a GC can be stimulated by three Gaussian function as follows [43]:

$$GC(x,y) = \sum_{x \in \sigma_1} \sum_{y \in \sigma_1} W_{\text{excitatory}} RC(x,y) - \sum_{x \in \sigma_2} \sum_{y \in \sigma_2} W_{\text{inhibition}} RC(x,y) + \log_b \left(1 + \sum_{x \in \sigma_3} \sum_{y \in \sigma_3} W_{\text{disinhibition}} RC(x,y) \right) \quad (b > 1) \quad (2)$$

$GC(x,y)$ is the response of the GC; $RC(x,y)$ is the stimulus from the input image. $W_{\text{excitatory}}$, $W_{\text{inhibition}}$ and $W_{\text{disinhibition}}$ are the weight function of center (excitatory), surround (inhibitory/suppression) and extended surround (disinhibitory) areas of the RF, respectively, which can be defined as follows:

$$W_{\text{excitatory}} = \frac{A_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma_1^2}} \quad (3)$$

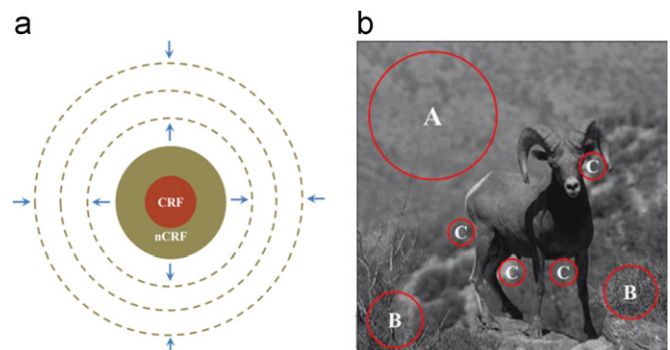


Fig. 4. (a) Diagram of the changes in the retinal GC nCRF size. Neurobiological studies have shown that the GC size can change in a range that is 2–5 times that of the CRF. (b) The red circles show the RF of GC. The image can be represented on three scales, i.e. A, B and C. A mostly represents the background information, B mostly represents the texture information, and C mostly represents the edge information between the main target and the background. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$W_{inhibition} = \frac{A_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma_2^2}}, \quad W_{disinhibition} = \frac{A_3}{\sqrt{2\pi}\sigma_3} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma_3^2}} \quad (4)$$

A_1 , A_2 and A_3 are the response amplitudes of the center, surround, and extended surround areas of the RF, respectively; σ_1 , σ_2 and σ_3 are the parameters representing scales; x and y are the position coordinates of the photoreceptor cells; x_0 and y_0 are the coordinates of the RF center. Based on the physiological data reported in previous studies [37,44], we set $\sigma_3 = 4\sigma_2$, $\sigma_2 = 5\sigma_1$, $A_1 = 1$, $A_2 = 0.18$, and $A_3 = 0.05$.

Fig. 4 illustrates the working principles of the GCs. The sizes of their nCRFs can be dynamically modulated by the properties of the stimulus. For instance, in areas where the color changes continuously, the nCRFs of the GCs continuously expand. In areas such as edges, where the color changes abruptly, the nCRFs shrink. We treated each pixel of the image as a data point; the set of data points formed an image in space. Areas with similar colors were considered to contain data points with matching properties. Each such area was considered as a single class, represented by one large RF. If the characteristics of the data points in an area were all different, each was treated as a separate class and represented by a separate small RF. The visual system's representation of physical scenes involves multiple scales, where each scale corresponds to a group of neurons of a certain RF size. Therefore, object structures that can be observed on a large scale evoke responses from more neurons with more RF sizes. Different responses from neurons with different RF sizes reflect the outputs on the different scales. It is very useful to introduce this type of multi-scale information to the contour detection process. When an appropriate threshold value is chosen, information analysis of the image variations treats texture as a region with the same or similar properties throughout. These regions can be represented by large RFs. In contrast, there are usually abrupt color changes in the regions between the contour of the target object and the background. These regions can be represented by small RFs. Also, by adjusting the corresponding threshold values based on the actual requirements, information on different scales can be acquired by using RFs of different sizes. In this way, through pre-processing by the RF of each GC, the preliminary information on the target object and the texture can be obtained. This forms the basis for accurate contour detection carried out by downstream visual cortical neurons.

2.3. Response of the CRF

As a type of spatial filter, the Gabor function can effectively describe the RFs of the simple cells in the visual cortex of mammals. Response models composed of simple RF filters in parity are often used to simulate the detection of motion or feature location by the visual system, because they can capture the basic properties of typical complex cells. These complex cells can be viewed as local orientation energy operators, and their maximum levels can be used to accurately locate the edges and lines of the figures. The formula for a two-dimensional Gabor function is as follows:

$$g(x,y;\lambda,\theta,\sigma,\varphi,\gamma) = \exp\left(-\frac{\tilde{x}^2 + \gamma^2 \tilde{y}^2}{2\sigma^2} \cos\left(2\pi\frac{\tilde{x}}{\lambda} + \varphi\right)\right) \quad (5)$$

here, $\tilde{x} = x \cos\theta + y \sin\theta$, and $\tilde{y} = -x \sin\theta + y \cos\theta$, γ is a constant representing the ratio of the major axis to the minor axes for an oval RF; λ is the wavelength; $1/\lambda$ is the spatial frequency of the cosine function; σ/λ is the bandwidth of the spatial frequency; and φ is the phase parameter. R_c

The response of a simple cell R_s to an input image I can be represented by the convolution of the RF function and I .

$$R_s(x,y;\lambda,\theta,\sigma,\varphi) = I * g(x,y;\lambda,\theta,\sigma,\varphi,\gamma) \quad (6)$$

The response of the complex cell R_c can be defined as the combined response of a pair of simple cells in parity with a phase difference of $\pi/2$.

$$R_c(x,y;\lambda,\theta,\sigma,\varphi) = \sqrt{R_s(x,y;\lambda,\theta,\sigma,0)^2 + R_s(x,y;\lambda,\theta,\sigma,\frac{\pi}{2})^2} \quad (7)$$

Gabor filters show good orientation selectivity. The optimal response of a complex cell is the energy value of the Gabor function along the optimal orientation. The corresponding optimal response energy and optimal orientation are given as follows:

$$R_c(x,y;\lambda,\theta_k,\sigma) = \max\{R_c(x,y;\lambda,\theta_k,\sigma) | k = 1, 2, \dots, N_\theta\} \quad (8)$$

$$\hat{\theta}(x,y;\lambda,\sigma) = \operatorname{argmax}\{R_c(x,y;\lambda,\theta_k,\sigma) | k = 1, 2, \dots, N_\theta\} \quad (9)$$

here, $N_\theta = 12$. The energy value of the area m represented by the RFs then equals the sum of the optimal responses of all of the complex cells.

$$E_m(x,y) = \sum_{x,y \in m} \widehat{R}_c(x,y;\lambda,\hat{\theta},\sigma) \quad (10)$$

2.4. Mathematical simulation of inhibitory and disinhibitory regions

The inhibitory characteristics of nCRFs can be used as a basic biological method for boundary detection. They can detect the region boundaries and isolated contours, and do not respond strongly to the edge information in the texture. Physiological studies have shown that such inhibition exhibits dynamic feedback properties, and the input of the feedback is relayed to the cortical neurons via inhibitory interneurons [45–48]. To simplify the calculation process in these studies, the interneurons were abstracted, and the entire inhibition was represented by a long distance connection, which was mainly used to divide the inhomogeneous regions and emphasize the boundaries of the texture region.

In our study, we modified the inhibitory weight function model established by Grigorescu et al. [21] such that it only functions in specific areas (the areas on the two sides of the RF). First, a centered difference of Gaussian (DoG) function was constructed, where the distances were weighted for normalization:

$$\text{DoG}(x,y;\sigma,k) = \frac{1}{\sqrt{2\pi}(k\sigma)^2} \exp\left(-\frac{x^2 + y^2}{2(k\sigma)^2}\right) - \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (11)$$

$$W_d(x,y;\sigma) = \frac{H(\text{DoG}(x,y;\sigma,k))}{H(\text{DoG}(x,y;\sigma,k))_1} \quad (12)$$

$$H(\text{DoG}(x,y)) = \begin{cases} \text{DoG}(x,y) > 0, (x,y) \in A_{inhi} \\ 0 \end{cases} \quad (13)$$

here $\|\cdot\|_1$ denotes the L_1 normal form; $W_d(x,y;\sigma)$ denotes the distance weight; $H(\text{DoG}(x,y))$ was used to ensure that the operator only functions in the inhibitory area A_{inhi} . The diameter of the nCRF is typically 2–5 times that of the CRF [15,37]. For this reason, we set $k=4$.

Reports in the literature have shown that the inhibitory effect of the nCRF decreases as its distance to the center of the RF increases and as the angle between the nCRF orientation and the optimal RF orientation increases [49,37,14,15,20,36,16]. In this way, the inhibition intensity in the inhibitory area is determined by two factors: distance and orientation. The weight functions used to define the inhibitory area in our simulation were as follows:

$$W_{in}(x,y) = W_d(x,y;\sigma,A_{in})W_o(x,y) \quad (14)$$

$$W_o(x,y) = \begin{cases} 1 & (\beta \leq \theta) \\ 0 & \end{cases} \quad (15)$$

$W_o(x,y)$ defined the spatial dimensions of the inhibitory regions on the two sides; β was the angle between the input stimulus location

and the optimal orientation of the CRF. θ Defined the field angles of the inhibitory regions on the two sides; finally, A_{in} denoted the inhibitory regions. No previous studies have reported on the size relationships of the different sub-regions of the nCRF inhibitory area. To improve contour detection, we assumed that on the two



Fig. 5. Pre-processing of the images by the GC. (a) Input images. (b) Projection of the images to the RF of the GC. The red circles show the RFs. Based on the characteristics of the represented areas, the nCRF sizes change accordingly. (c) Results after pre-processing of the images by the nCRF of the GC. (The natural images employed from the first to third row were downloaded from the website: <http://www.cs.rug.nl/~imageing/>. The last three rows were downloaded from the website: <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/>.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sides of the CRF, the inhibitory sub-regions were larger than the disinhibitory sub-regions. We set $\theta = \pi/3$. We let K be the number of sampling orientations of the Gabor function. The inhibitory effect evoked by stimuli of different orientations in the nCRF was then as follows:

$$\text{Inhi}(x, y; \beta) = \sum_{i=1}^K W_{in}(x, y) * E(x, y; \beta_i) \quad (16)$$

The disinhibitory regions are usually located on the two sides of the RF center along the optimal orientation. They affect the CRF center response in a manner that is exactly opposite to the way that they affect the inhibitory regions. Stimulation of the disinhibitory regions enhances the cell's response and can prevent certain weak contour data being disregarded. Like the inhibitory regions, the weight functions of the disinhibitory regions are also influenced by the distance to the center of the RF and the angle between their orientation and the optimal RF orientation. We define the weight functions of the disinhibitory regions as follows:

$$W_{dis}(x, y) = W_d(x, y; \sigma, A_{dis})(1 - W_o(x, y)) \quad (17)$$

$(1 - W_o(x, y))$ represent the spatial sizes of the disinhibitory regions on the two sides, and A_{dis} denotes the disinhibitory regions. The disinhibitory effect evoked by stimuli with different orientations in the nCRF was as follows:

$$\text{Dis}(x, y; \beta) = \sum_{i=1}^K W_{dis}(x, y) * E(x, y; \beta_i) \quad (18)$$

based on Eq. (1), the output response to stimulation in a region m was represented as follows:

$$E(x, y) = \sum_{x, y \in m} \left(R_c(x, y; \lambda, \hat{\theta}, \sigma) - \mu \sum_{i=1}^K W_{in}(x, y) * E(x, y; \beta_i) + \sum_{i=1}^K W_{dis}(x, y) * E(x, y; \beta_i) \right) \quad (19)$$

here, μ was used to control the inhibitory magnitude of the nCRF.

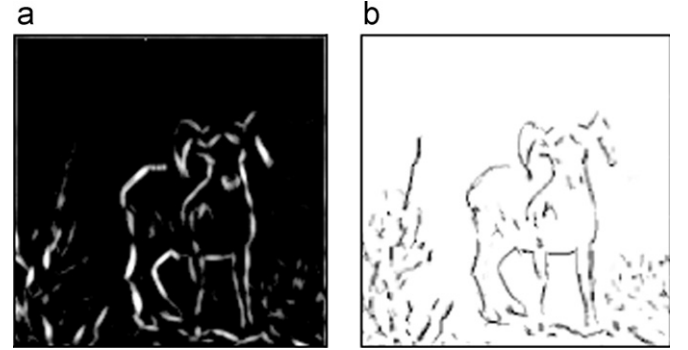


Fig. 8. (a) Contour map after integrated multi-scale Gabor filtering. (b) Contour map after binarization.

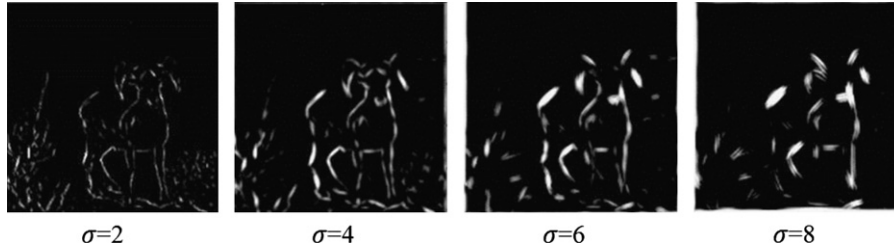


Fig. 6. Gabor energy maps on different scales.

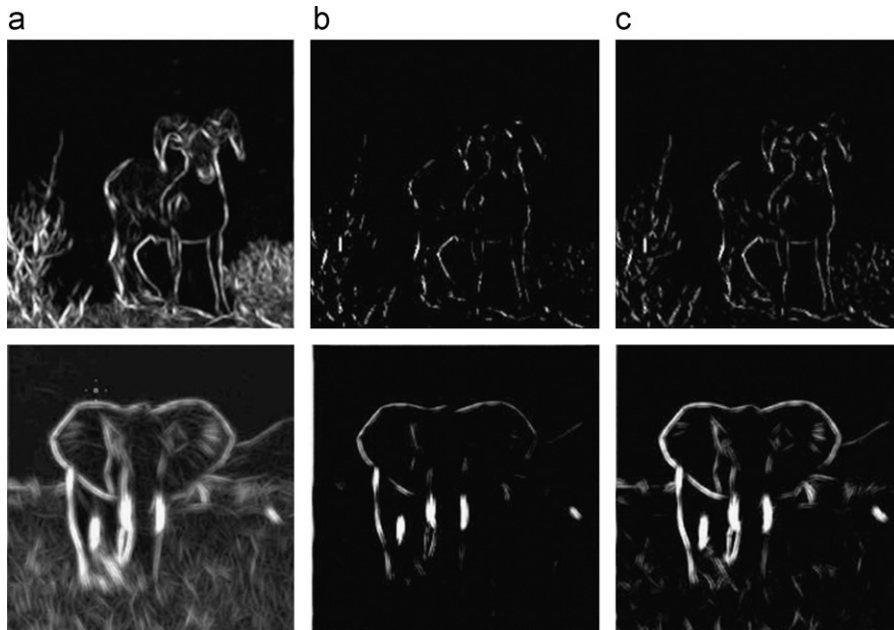


Fig. 7. (a) Gabor energy map at $\sigma \text{ QUOTE} = 2$. (b) Gabor energy map after inhibition, represented by Eq. (14). (c) Gabor energy map after disinhibition, represented by Eq. (15). Under inhibitory modulation, most of the texture information was removed, but part of the target object contour information was also sacrificed. Disinhibition partially recovered the lost contour information and helped to provide more continuous contours.

Table 1
Summary of parameter meanings and setting.

Parameters	Meaning	Setting value	Physiological grounds
k	Ratio of the two standard deviations in DOG	4.0	Refs. [44,14,15]
A_1,A_2,A_3	The response amplitudes of the center, surround and extended surround areas of the RF	1,0.18,0.05	Refs. [44,37]
$\sigma_1,\sigma_2,\sigma_3$	Scale information(we use radius of the RF representing scales in our experiments)	$\sigma_3 = 4\sigma_2, \sigma_2 = 5\sigma_1$	Refs. [44,37]
σ/λ	Spatial frequency bandwidth, where λ is the wavelength of Gabor filter	0.56	Refs. [57–59]
γ	Spatial aspect ratio of the Gaussian envelop of Gabor filter	0.5	Refs. [57]
N_θ	Number of orientations of Gabor filter	12	Refs. [57]
μ	Suppression factor	1.0	It was used to control the inhibitory magnitude of the nCRF

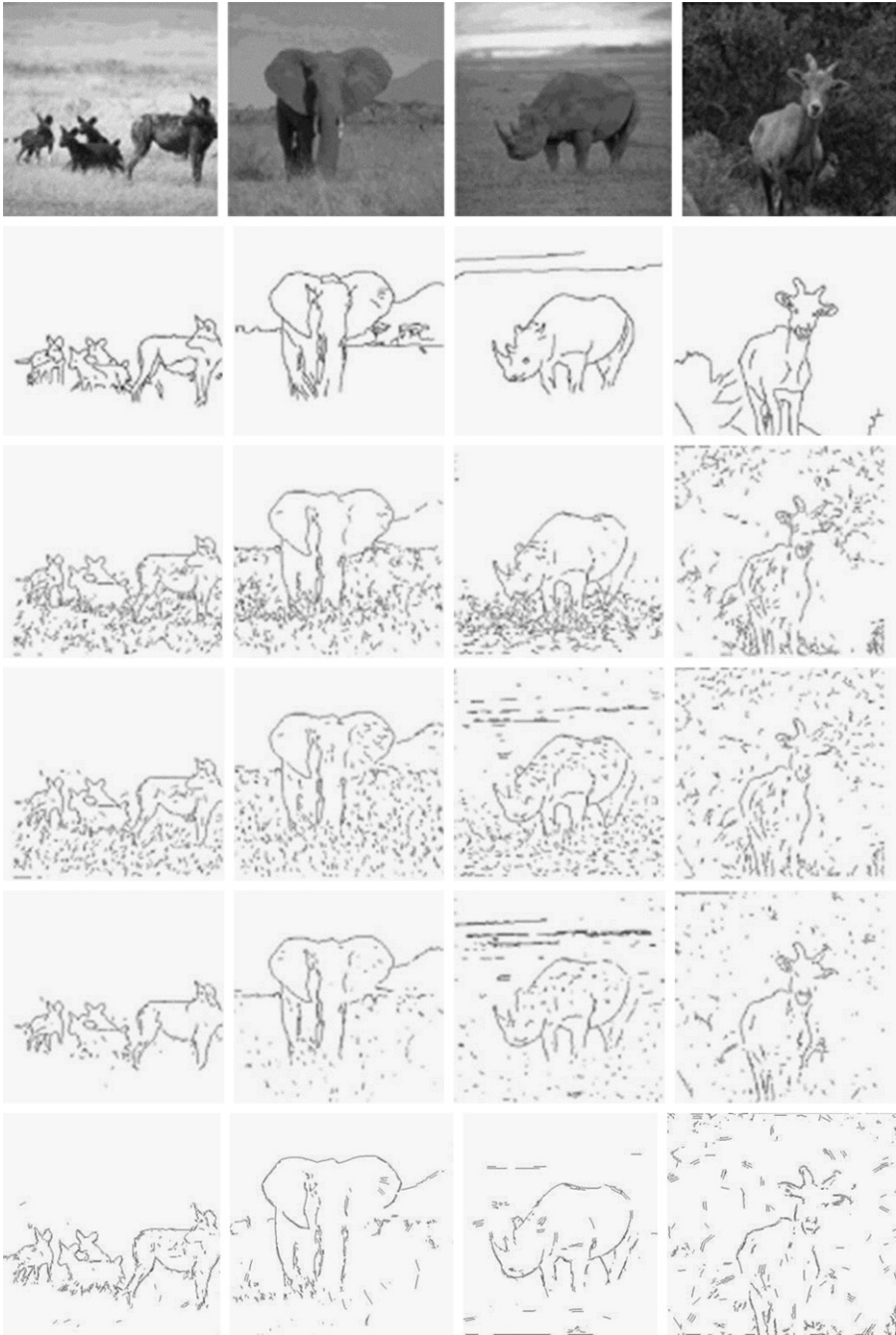


Fig. 9. Results of contour detection on four natural images selected from the Rug database. The first row and the second row represent the original images and ground-truth contours, respectively. The third row and the fourth row represent the output of anisotropic and isotropic inhibition detectors, respectively. The fifth row shows the results of adaptive inhibition as proposed by Zeng et al. [28]. The last row shows the output of our computation model.

3. Experimental results

3.1. Pre-processing results of the retinal GC RF

For comparison purposed, all of the testing images came from the Rug dataset from University of Groningen and BSD dataset from UC Berkeley. The images in this database have been used for testing and evaluation contour detection algorithms [21,22,27,28,23,30,50–52]. In Section 2.2, we explained the role of the RF of the GC in the acquisition of multi-scale information in detail. Pre-processing of the RF of the GC can approximately deduce both the target object information and the texture information, mainly through a multi-scale representation of the physical scene and using the fact that the nCRFs of GCs can dynamically vary their sizes [53–56]. Image information on different scales can be obtained, depending on the nCRF size variation. Fig. 5 shows the images after the pre-processing of the nCRF of the GC. Fig. 5b illustrates the effects of changes in the size of the RF of the GC. The RF size was largest in the background

area and the texture areas. It became smaller in areas with abundant color changes, such as the object boundaries. To simplify the calculations, we assumed that the RF projection was even. Let $r_{CRF} = 2$, and $r_{nCRF} = 10$, with the increases in size occurring in steps of 2. The spatial scale information was described by $\sigma = \{2, 4, 6, 8, 10\}$. Fig. 5c shows the images when they were pre-processed by the nCRF. Most of the texture information was filtered out, whereas the contour information of each target object was largely retained.

The Gabor energy map based on Eq. (8) is shown in Fig. 6, which illustrates the results of Gabor filtering on different scales with an inhibitory factor of $\mu = 1.5$. It can be seen that different information was obtained from the image on different scales.

3.2. Realization of inhibitory and disinhibitory activities

After the processing of the GC, most of the texture information was filtered out and the approximate contour information of the



Fig. 10. Results of contour detection on four natural images selected from the BSD database. The first row and the second row represent the original images and ground-truth contours [60], respectively. The third row represents the output of standard Canny edge detector [61], and the fourth row represents the output of like-CARTOON multiscale edge detector [62]. The fifth row shows the results of single scale detector with surround inhibition [21]. The last row shows the output of model.

target object was acquired. However, this is still far from the desired refined and accurate contour representation of the object. In previous studies, artificial images were used to test the inhibitory functions [21,29]. As even texture and high contrast between the foreground and the background are typical characteristics of artificial images, the inhibitory functions usually achieve prominent results in contour detection. To determine how the inhibitory and disinhibitory activities affect the contour detection under realistic circumstances, we chose to use natural images. The inhibitory/disinhibitory model discussed in Section 2.4 was used for further processing. Fig. 7 shows the Gabor energy output on a scale of $\sigma=2$. Fig. 7a illustrates the processing results: without the participation of the inhibitory components, all of the information, including the texture and the target object information, was detected. Fig. 7b shows the contour map after inhibitory modulation: the texture information was substantially reduced, but the target object contour information was also partially inhibited. This led to “broken contours”. Fig. 7c shows the contour map after the introduction of the disinhibitory modulation, which partially recovered the lost contour information. For reason of space, we only show the results on one scale in this paper. In the actual test, the scales used included $\sigma = \{2, 4, 6, 8\}$. The parameters used for the Gabor filter were the same as those used by Grigorescu et al. [21]; $\sigma/\lambda = 0.56$, $\gamma = 0.5$, and $N_\theta = 12$. The variance of the DoG function was $\sigma = 4$ and $\sigma_{\text{dis}} = 5\sigma$. The sizes of all testing images were 512 512 pixels (Rug dataset) and 481 321 pixels (BSD from UC

Berkeley); and the noise edges were defined as edges that were smaller than 20 pixels.

3.3. Multi-scale integration σ

Section 3.2 shows the effects of the inhibitory and disinhibitory modulation of the nCRF. Different image information was contained on the different scales. On the largest scale, this mainly involved structural information, whereas on the smallest scale, it mainly represented information on the image details. If the computation is only performed on one scale, it becomes extremely difficult to guarantee that the information obtained is optimal. If we want to remove the texture information, we must increase inhibition, which will also cause the loss of contour information. A similar situation occurs with disinhibition. This problem can be solved nicely by using multi-scale integration, i.e. by integrating all four scales. (From an alternative perspective, this integration can be viewed as another step of the inhibition and disinhibition process.) Fig. 8 shows the contour map after integration of the four scales.

In order to find the best results and conduct a statistical analysis for our model within a specific parameter range. Table 1 summarizes the meanings and values of the parameter sets involved in our model. For those parameters with direct physiological grounds and the corresponding references are listed in the last column of Table 1.

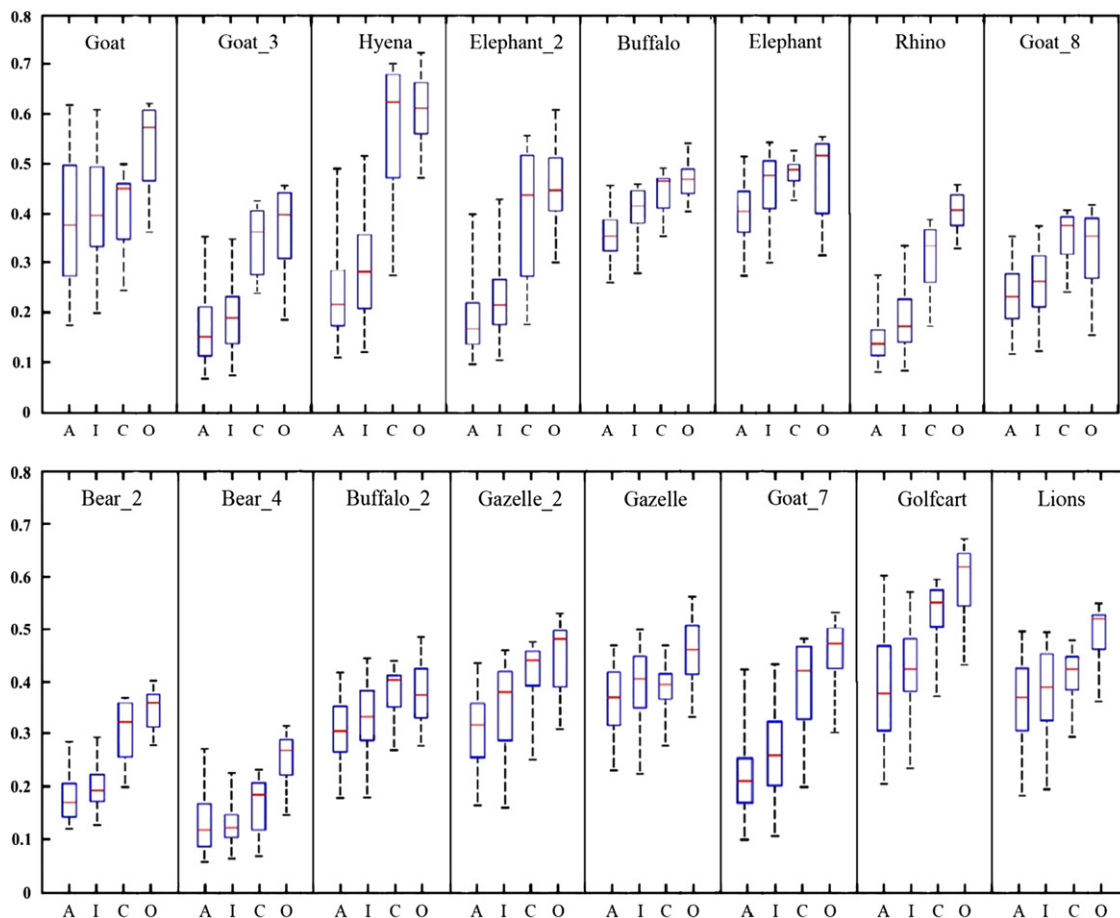






Fig. 11. Contour detection performance of the anisotropic model (denoted by A), the isotropic model (denoted by I) from literature [21], the adaptive inhibition model (denoted by C) from literature [27,28] and of our model (denoted by O) on sixteen images. Each lattice represents the performance plots based on a similar image. In each plot, the top end of each whisker represents the optimal performance among 50 different results for each image, and the horizontal red line in the box shows the median value of the performance. The top and bottom lines of each box denote the upper and lower quartiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Parameter settings and performance evaluation.

Images	Detectors	μ	σ	E_{FP}	E_{FN}	P
	Isotropic	1.2	2.4	0.98	0.27	0.42
	Anisotropic	1.2	2.4	1.17	0.27	0.39
	Butterfly-shaped inhibition model	1.0	2.0+12.0	0.36	0.27	0.57
	Our model	1.0	2.0+6.0	0.28	0.25	0.61
	Isotropic	1.0	2.4	1.36	0.33	0.35
	Anisotropic	1.2	2.2	1.47	0.26	0.36
	Butterfly-shaped inhibition model	1.0	2.0+10.0	0.91	0.31	0.42
	Our model	1.0	2.0+6.0	0.45	0.51	0.46
	Isotropic	1.0	2.0	0.73	0.19	0.51
	Anisotropic	1.2	2.4	0.80	0.20	0.49
	Butterfly-shaped inhibition model	1.0	2.0+10.0	0.23	0.17	0.70
	Our model	1.0	2.0+6.0	0.21	0.18	0.72
	Isotropic	1.0	2.4	1.65	0.28	0.33
	Anisotropic	1.0	2.4	2.13	0.37	0.27
	Butterfly-shaped inhibition model	1.0	2.0+10.0	0.54	0.32	0.39
	Our model	1.0	2.0+6.0	0.34	0.25	0.47

To make the comparison more convenient, we selected the same four test images that were used by Grigorescu [21]. Fig. 9 illustrates the results. The first row shows the raw images. The second row shows the corresponding ground-truth contour images. The third and fourth rows show the isotropic and anisotropic contour maps from the Grigorescu et al. study [21], respectively. The fifth row shows the detection results using the adaptive inhibitory mechanism described by Zeng et al. [28]. The last row shows the experimental results using our model. By comparison, it can be seen that our model is superior to a certain extent in terms of both inhibiting the texture information and retaining the continuity of the contours.

Also we have selected four images randomly from BSD 500 [60]. Fig. 10 presents the contour detection result using different models. The first and second rows list the original images and corresponding ground-truth image, respectively. The rows from third to fifth show the result using standard Canny edge detector [61], the result using like-CARTOON multiscale edge detector [62] and the result using Grigorescu et al. model [21]. The last row shows the result of our contour model. The results clearly show that our computational model much more effectively than the other three models, e.g. suppresses the edge originating from textural regions like grass and preserve the real contour information to some extent.

3.4. Performance evaluation

The evaluation method used was the same as that used by Grigorescu [21]. We let E_{GT} and B_{GT} be the contour pixel set and the background pixel set of the ground truth contour image, respectively. E_D and B_D were the contour pixel set and the background pixel set in the final detection result, respectively. The correctly detected contour pixel was set $E = E_D \cap E_{GT}$; a missed contour was $E_{FN} = E_{GT} \cap B_D$; and a falsely detected contour was $E_{FP} = E_D \cap B_{GT}$.

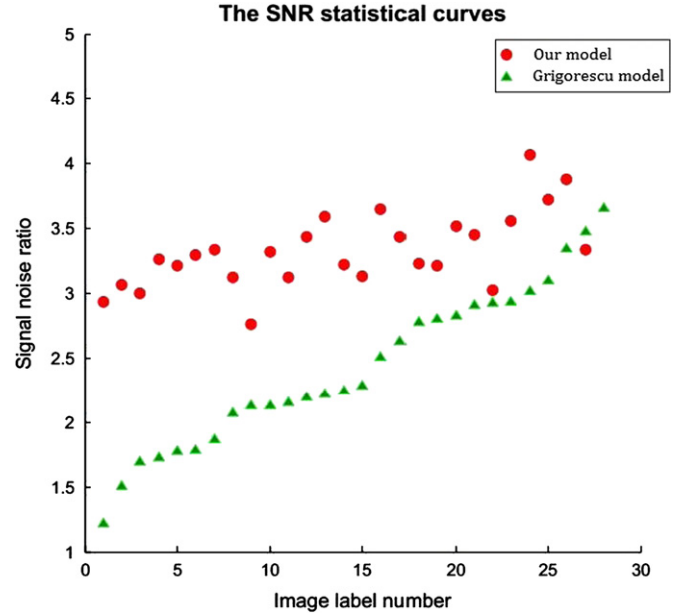


Fig. 12. Comparison of signal–noise–ratio testing on Rug Dataset. The red dot and green triangle represent the SNR value of our model and Grigorecu model, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The performance evaluation index for the contour detection operator was as follows:

$$P = \frac{\text{card}(E)}{\text{card}(E) + \text{card}(E_{FP}) + \text{card}(E_{FN})}$$

here, $\text{card}(X)$ represented the number of set members in set X , $P \in [0, 1]$. If all of the true contour pixels were correctly detected and no background pixels were falsely classified as contour pixels, then $P = 1$. In all other cases, as the value of P decreased, then the number of false or missed detections increased. We selected 16 images from the Rug database for the comparison test. Table 2 shows the comparison of the performance of our model with those of the isotropic and anisotropic contour detectors proposed by Grigorescu et al. [21] and butterfly-shaped inhibition model proposed by Zeng et al. [28]. From Table 2, we can see that our results were better than those of Grigorescu et al. and Zeng et al. The E_{FP} and E_{FN} values indicated that our model was superior with respect to inhibition of the texture information and in rendering

the object contours more prominently. Also, all P values calculated from our model were higher than those of the isotropic and anisotropic models. This was especially true for the elephant_2, hyena, and rhino images.

To test the robustness of our proposed model and that of the other models, we implemented a statistical analysis of the performances of the different models using the aforementioned combinations of the different parameters. Fig. 11 shows comparative box-and-whisker plots for sixteen of the images used in our experiments. The top end of a whisker represents the best P value for a specific image, and the horizontal red line in the box of each whisker shows the median P value. It is clear from these plots that for all of the test images, the best P values attained by our model

Table 3
Comparison between gPb and our model.

	gPb	Our model
Multiscale	mPb represents the fine scale sPb represents the coarse scale	The small size RF represents the fine scale The large size RF represents the coarse scale
σ	The gPb detector considers brightness, color, and texture gradients at three scales. $[(\sigma/2), \sigma, 2\sigma]$	We use radius of the nCRF representing scales in our experiments. $\sigma_3 = 4\sigma_2$, $\sigma_2 = 5\sigma_1$
Filter	Second-order Savitzky–Golay smoothing filter	Gabor filter
θ	8 Orientations	12 Orientations
Combination	sPb ⊕ gPb	The large size RFs ⊕ the small size RFs

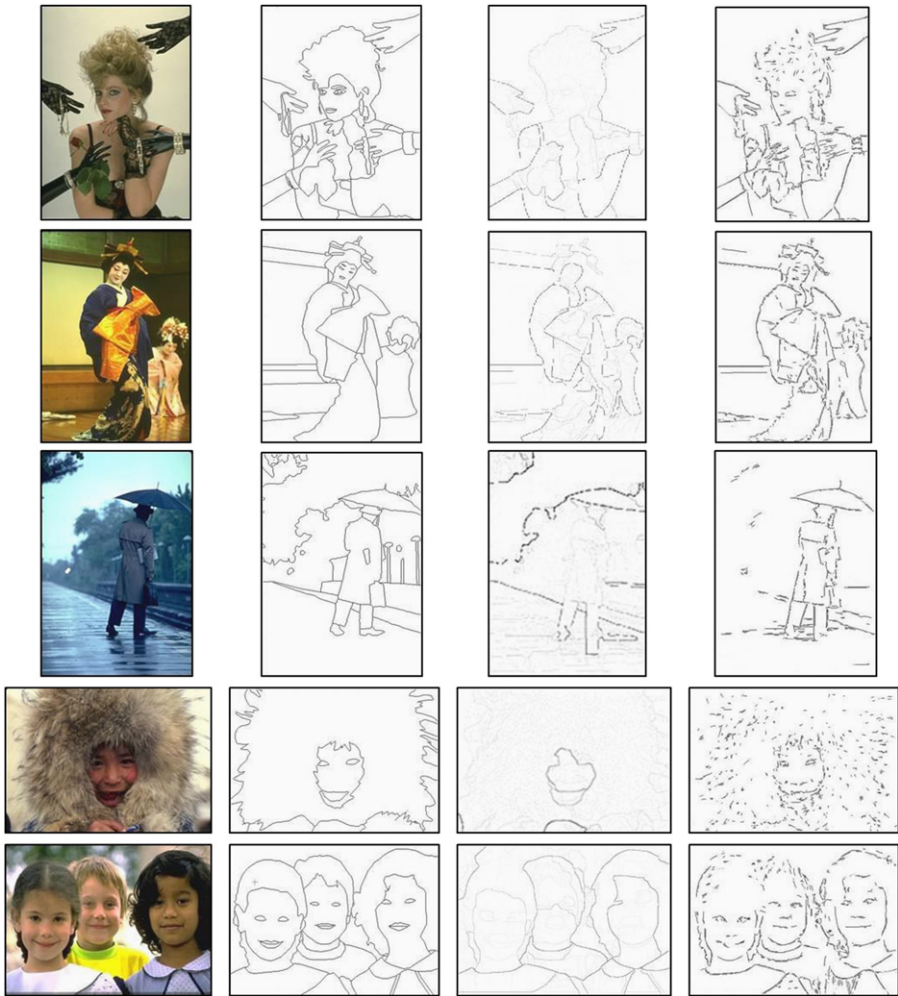


Fig. 13. Comparison results between gPb and our model in color images. The columns from left to right represent the original color images, ground-truth contour, output of gPb and our model, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were noticeably higher than those of the other models. Our model also showed a significantly higher median P value in almost all cases, and showed a noticeably smaller interquartile range in most cases. This indicated that our model performed better than the standard inhibition model [21] and butterfly-shaped inhibition model [27,28].

Note that the presented results of standard surround inhibition model (anisotropic and isotropic) and butterfly-shaped inhibition model shown in Table 2 are not exactly identical to those listed in the work of Grigorescu et al. [21] and Zeng et al. [28]. This difference is mainly due to the difference in implementation details (e.g. parameters or initialization values setting) between these works. However, this difference does not affect the performance comparison and the deduced conclusion, because different models in this study remain identical values for those common parameters.

We also introduce the statistical signal–noise-ratio curve to evaluate the efficiency of our model. We take the ground-truth as the standard benchmark. Similar to the method was proposed in [63], the signal-to-noise ratio (SNR) defined as:

$$\text{SNR}_{DB} = 10 \log_{10} \left(\frac{\sigma_{\text{image}}}{\sigma_{\text{noise}}} \right) \quad (20)$$

The final contour information obtained from our model and Grigorescu model [21] along with those of the ground-truth image was tested using Eq. (20). The final testing result shown in Fig. 12 and clearly indicates that our model has better performance in

terms of inhibiting the texture information (i.e. noise) while retaining contour information (i.e. signal).

3.5. Comparison with gPb detector

As an important point for contour detection, we consider the work of Arbelaez et al. [64], who define two functions $\text{mPb}(x,y,\theta)$ and $\text{sPb}(x,y,\theta)$ to convey different information, as the former fires at all the edges (i.e. fine scale), while the latter extracts only on the most salient curves in the image (i.e. coarse scale). The final globalized probability of boundary (gPb) is a weighted sum of two. The gPb detector combines the orientated gradient signals obtained from transforming an input image into four separate feature channels and processing each channel independently. The first three correspond to the channels of the CIE Lab colorspace, which we refer to as the brightness (L), color (a), and color (b) channels. For gray-scale images, the brightness channel is the image itself and no color channels are used. Table 3 summarizes the difference between gPb and our model.

The experimental results demonstrate that the gPb shows excellent performance for majority color images, but it shows signs of performance degradation for gray images. Especially, we also found that the gPb detector cannot extract contour information for some images, while our model presents preferable result in same images. Figs. 13 and 14 show the contour detection results in color images and gray images, respectively.



Fig. 14. Comparison results between gPb and our model in gray images. The columns from left to right represent the original gray images, ground-truth contour, output of gPb and our model, respectively.

Contour detection performance is evaluated in the precision-recall framework of [65]. The precision-recall curve captures the trade-off between accuracy and noise as the detector threshold is varied. Precision is the fraction of detections which are true positives, while recall is the fraction of positives that are detected. These are computed using a distance tolerance of two pixels to allow for small localization errors in both the machine and human contour maps. A particular application will define a relative cost α between these quantities, which focuses attention at a specific point on the precision-recall curve. The F -measure, defined as $F = PR/(\alpha R + (1-\alpha)P)$, captures this trade-off [65]. The location of the maximum F -measure along the curve provides the optimal threshold given $\alpha=0.5$. Fig. 15 shows the performance comparison between gPb and our model, it provides a clearly quantitative justification for performance evaluation.

3.6. Facilitation to shape-based recognition

Depending on contour, a direct application of contour-detection is to recognize object. Undoubtedly, a clean output with fewer disturbances from noise will facilitate recognition. The contour output of our model seems to be cleaner than other ones. Here, we introduce famous shape context algorithm to do recognition [66]. This algorithm is based on shape, thus if as much contour information as possible is retained then this algorithm will be more applicable to map a testing image to a proper ground-truth benchmark. We test the contours obtained from our model and from Grigorescu's model along with the ground-truth image using the shape context algorithm. As shown in Fig. 16c, there are many points in background being matched with points on elephant body. And our contour result is cleaner than that of Grigorescu's model, the point-to-point correspondence in Fig. 16d is more accurate than Fig. 16c.

4. Discussion

According to the literature, most contour detection models based on nCRFs rely on the inhibition mechanism of the V1 neurons [21,23,30,50,51,67,24]. The model proposed in this paper

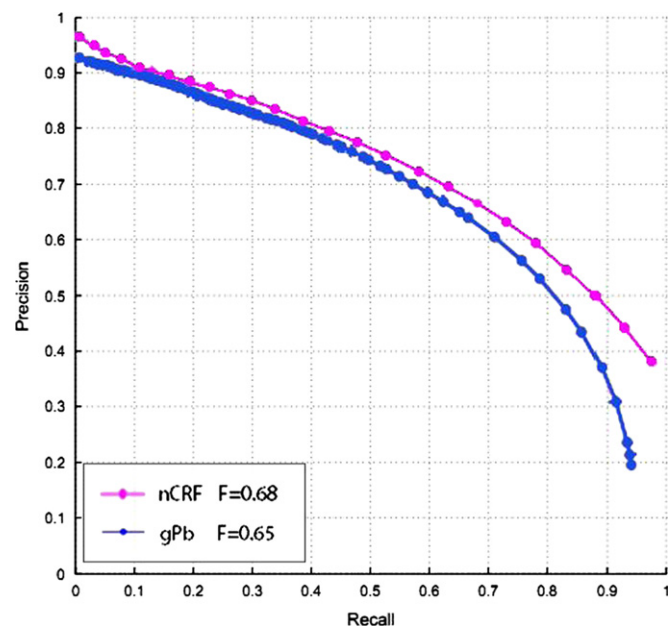


Fig. 15. Evaluation of contour detection on the BSD 300 gray Benchmark. The contour detection approaches are ranked according to their maximum F -measure with respect to human ground-truth boundaries. Obviously, our model performs better than gPb detector.

uses a multi-scale integration algorithm based on nCRFs (retinal GCs and visual cortical neurons). In our model, preprocessing (by the retinal GCs), later-stage processing (by the visual cortical neurons), and the final representation (multi-scale integration) form a complete processing chain that is consistent with the human visual pathway. With regard to the working mechanism of the nCRF, both disinhibition and inhibition are included. In this way, some of the inhibited contour information can be restored to a certain extent. The completeness of the contour obtained is thus guaranteed. The most prominent differences between our model and previous models are that our model introduces GC preprocessing at an early stage and uses multi-scale integration for the output at a later stage. The rationale for adopting this strategy includes the following reasons:

- (i) The nCRFs can adjust their filtering properties according to changes in the spatial frequency and can compensate for the loss of low-spatial-frequency components to a certain extent. This renders them capable of transferring brightness information that occurs over a large area or that change slowly over time. The role of the nCRF in visual processing has been discussed extensively [68–70]. Evidence has been presented in the literature that shows that non-local, long-range connections are involved in the nCRFs of V2 neurons [45]. These connections allow these neurons to receive feedback input from neurons at higher levels and to form feedback pathways in visual processing [46–48]. This provides anatomical evidence that can be used for the design of algorithm models for nCRFs. It has also been shown that, under the influence of acetylcholine, the size of the RFs can change dynamically based on differences in the brightness, stimuli, background images, and movement speed [71–74].
- (ii) In recent years, with the development of neurophysiology and computer-aided anatomical studies, some few fairly accurate preliminary visual system models have been proposed. These models have been built on different levels of different regions of the visual system, and the scale-space theory is one of these models [75,76]. In this theory, every scale corresponds to a group of neurons of a certain RF size. GCs that have different RF sizes represent the different properties of the data using different scales.
- (iii) A good contour detection model should provide truthful, continuous contour information that describes the target object while removing the texture information. Typically, most contour detection models based on the nCRF only take inhibition into account, even though increasing numbers of studies have shown that excitation also plays an important role in contour detection [77,15,78,79]. On this basis, we included the disinhibitory effect of the nCRF in our model, with the aim of partly restoring the inhibited contour information. The main disadvantage of this method is that it may produce some “fake” contours. For this reason, at the final output stage, a multi-scale integration method is used to produce the ideal contour detection results. The experimental results shown in Section 3.3 indicate that our method can indeed improve the quality and efficiency of contour detection.

Our ultimate goal is to establish an optimal contour detection model, which requires accurate controls for the magnitude of the inhibition and disinhibition based on the external information, along with an optimal balance between retention of the target object information and removal of the texture information. However, it is extremely difficult to realize such accurate controls when relying on the response of only one type of neuron, especially in complex scenes. The fact that the human visual

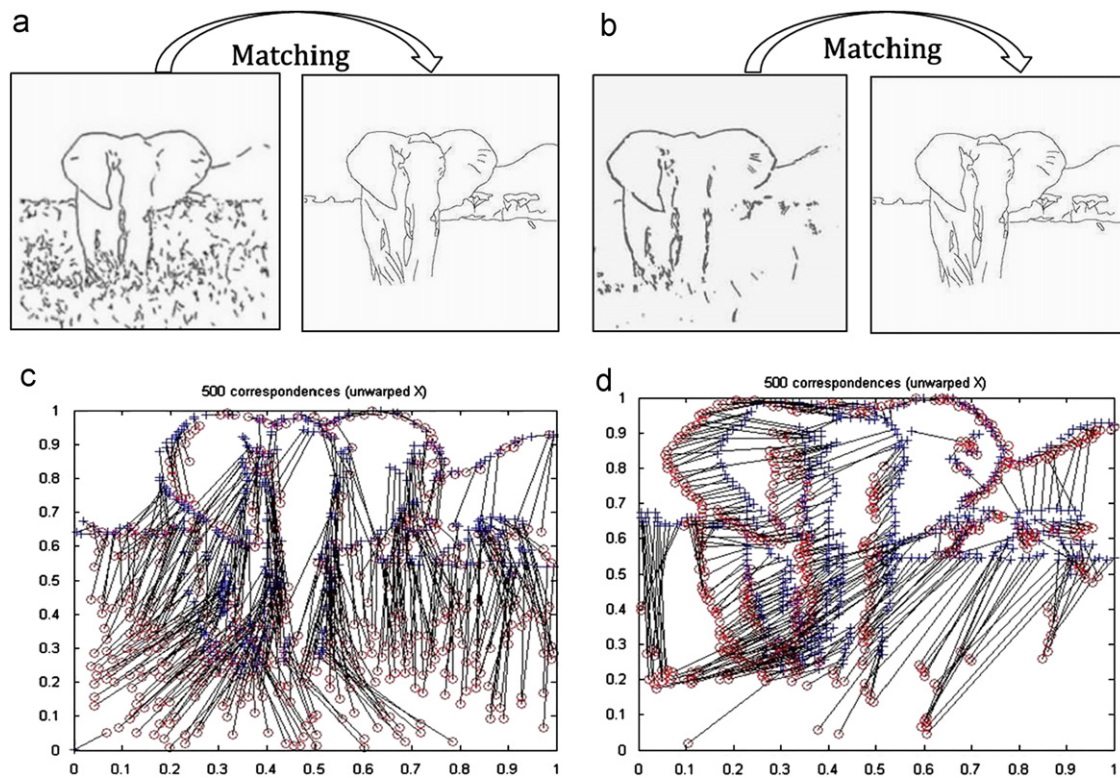


Fig. 16. (a) The result of Grigorescu's model and ground-truth image. (b) The result of our model and ground-truth image. (c) Is point-to-point matching between the ground-truth (blue crosses) and the contour result of Grigorescu's model (red circle) produced by shape context algorithm. (d) Is point-to-point matching result between the ground-truth (blue crosses) and the contour result of our model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

system can achieve this effortlessly and efficiently is also because of the interaction of many neural mechanisms. Improvement in this area will be the subject of further studies.

Acknowledgment

This work was supported by the Major State Basic Research Program of China (2010CB327900) and the Natural Science Foundations of China (30990263). The authors would like to thank Grigorescu, Petkov and Arbelaez for their natural image database with associated ground-truth contour maps. We also thank the anonymous reviewers for their valuable comments and suggestions on the manuscript.

References

- [1] D.A. Forsyth, J. Ponce, *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
- [2] S. Ando, Image field categorization and edge/corner detection from gradient covariance, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2) (2000) 179–190.
- [3] F. Bergholm, Edge focusing, *IEEE Trans. Pattern Anal. Mach. Intell.* 6 (1987) 726–741.
- [4] J.C. Bezdek, R. Chandrasekhar, Y. Attikouzel, A geometric approach to edge detection, *IEEE Trans. Fuzzy Syst.* 6 (1) (1998) 52–75.
- [5] K. Bowyer, C. Kranenburg, S. Dougherty, Edge detector evaluation using empirical ROC curves, in: *Proceedings of the IEEE*, 1999.
- [6] J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (6) (1986) 679–698.
- [7] L.A. Iverson, S.W. Zucker, Logical/linear operators for image curves, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (10) (1995) 982–996.
- [8] P. Meer, B. Georgescu, Edge detection with embedded confidence, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (12) (2001) 1351–1365.
- [9] C.A. Rothwell, J. Mundy, W. Hoffman, V.D. Nguyen, Driving vision by topology, in: *Proceedings of the IEEE*, 1995, pp. 395–400.
- [10] S. Sarkar, K. Boyer, Optimal infinite impulse response zero crossing based edge detectors, *CVGIP: Image Understanding* 54 (2) (1991) 224–243.
- [11] S.M. Smith, J.M. Brady, SUSAN—a new approach to low level image processing, *Int. J. Comput. Vision* 23 (1) (1997) 45–78.
- [12] M. Tabb, N. Ahuja, Multiscale image segmentation by integrated edge and region detection, *IEEE Trans. Image Process.* 6 (5) (1997) 642–655.
- [13] A.P. Witkin, Scale-space filtering, in: *Proceeding of the International joint conference on Artificial Intelligence (IJCAI'83)* (2) (1983) 1019–1022.
- [14] H. Jones, K. Grieve, W. Wang, A. Sillito, Surround suppression in primate V1, *J. Neurophysiol.* 86 (4) (2001) 2011.
- [15] M.K. Kapadia, G. Westheimer, C.D. Gilbert, Spatial distribution of contextual interactions in primary visual cortex and in visual perception, *J. Neurophysiol.* 84 (4) (2000) 2048–2062.
- [16] G.A. Walker, I. Ohzawa, R.D. Freeman, Suppression outside the classical cortical receptive field, *Visual Neurosci.* 17 (3) (2000) 369–379.
- [17] J.J. Knierim, D.C. van Essen, Neuronal responses to static texture patterns in area V1 of the alert macaque monkey, *J. Neurophysiol.* 67 (4) (1992) 961–980.
- [18] J.B. Levitt, J.S. Lund, Contrast dependence of contextual effects in primate visual cortex, *Nature* 387 (6628) (1997) 73–76.
- [19] J. R. Cavanaugh, W. Bair, and J. A. Movshon, Orientation-selective setting of contrast gain by the surrounds of macaque striate cortex neurons, *Neurosci. Abstr.* 23 (1997) 227.2.
- [20] H.C. Nothdurft, J.L. Gallant, D.C. van Essen, Response modulation by texture surround in primate area V1: correlates of popout under anesthesia, *Visual Neurosci.* 16 (01) (1999) 15–34.
- [21] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour detection based on nonclassical receptive field inhibition, *IEEE Trans. Image Process.* 12 (7) (2003) 729–739.
- [22] N. Petkov, M.A. Westenberg, Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition, *Biol. Cybern.* 88 (3) (2003) 236–246.
- [23] C. Grigorescu, N. Petkov, M.A. Westenberg, Contour and boundary detection improved by surround suppression of texture edges, *Image Vision Comput.* 22 (8) (2004) 609–622.
- [24] G. Papari, P. Campisi, N. Petkov, A. Neri, A biologically motivated multi-resolution approach to contour detection, *EURASIP J. Appl. Signal Process.* 1 (1) (2007) 119.
- [25] G.E. La Cara, M. Ursino, A model of contour extraction including multiple scales, flexible inhibition and attention, *Neural Networks* 21 (5) (2008) 759–773.
- [26] M. Ursino, G.E. La Cara, A model of contextual interactions and contour detection in primary visual cortex, *Neural Networks* 17 (5–6) (2004) 719–735.
- [27] C. Zeng, Y. Li, C. Li, Center-surround interaction with adaptive inhibition: a computational model for contour detection, *Neuroimage* 1 (55) (2010) 49–66.
- [28] C. Zeng, Y. Li, K. Yang, C. Li, Contour detection based on a non-classical receptive field model with butterfly-shaped inhibition subregions, *Neurocomputing* 74 (2011) (2011) 1527–1534.

- [29] Q. Tang, N. Sang, T. Zhang, Contour detection based on contextual influences, *Image Vision Comput.* 25 (8) (2007) 1282–1290.
- [30] G. Papari, P. Campisi, N. Petkov, A. Neri, A multiscale approach to contour detection by texture suppression, in: *Proceedings of the SPIE*, 2006, pp. 107–118.
- [31] K.H. Liang, T. Tjahjadi, Y.H. Yang, Bounded diffusion for multiscale edge detection using regularized cubic B-spline fitting, *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.* 29 (2) (1999) 291–297.
- [32] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, in: *Proceedings of the IEEE*, 1996, pp. 465–470.
- [33] S.G. Mallat, Multifrequency channel decompositions of images and wavelet models, *IEEE Trans. Acoust., Speech Signal Process.* 37 (12) (1989) 2091–2110.
- [34] X. Ren, Multi-scale improves boundary detection in natural images, in: *Proceedings of European Conference on Computer Vision (ECCV'08)* (2008).
- [35] T. Shou, Y. Zhou, Orientation and direction sensitivity of cells in subcortical structures of the visual system, *Sheng li xue bao: Acta Physiol. Sin.* 48 (2) (1996) 105.
- [36] A.F. Rossi, R. Desimone, L.G. Ungerleider, Contextual modulation in primary visual cortex of macaques, *J. Neurosci.* 21 (5) (2001) 1698.
- [37] L. Chao-Yi, L. Wu, Extensive integration field beyond the classical receptive field of cat's striate cortical neurons—classification and tuning properties, *Vision Res.* 34 (18) (1994) 2337–2355.
- [38] H. Ikeda, M. Wright, The outer disinhibitory surround of the retinal ganglion cell receptive field, *J. Physiol.* 226 (2) (1972) 511.
- [39] J. Krüger, B. Fischer, Strong periphery effect in cat retinal ganglion cells. Excitatory responses in ON- and OFF-center neurones to single grid displacements, *Exp. Brain Res.* 18 (3) (1973) 316–318.
- [40] L. Chao-Yi, P. Xing, Z. Yi-Xiong, M. Hans-Christoph, Role of the extensive area outside the X-cell receptive field in brightness information transmission, *Vision Res.* 31 (9) (1991) 1529–1540.
- [41] K. Ghosh, S. Sarkar, K. Bhaumik, Low-level brightness–contrast illusions and non-classical receptive field of mammalian retina, in: *Proceedings of the IEEE International Conference on Intelligent Sensing and Information Processing*, 2005, pp. 529–534.
- [42] K. Ghosh, S. Sarkar, K. Bhaumik, A possible explanation of the low-level brightness–contrast illusions in the light of an extended classical receptive field model of retinal ganglion cells, *Biol. Cybern.* 94 (2) (2006) 89–96.
- [43] K. Ghosh, S. Sarkar, K. Bhaumik, Image enhancement by high-order Gaussian derivative filters simulating non-classical receptive fields in the human visual system, *Pattern Recognition Mach. Intell.* 3776 (2005) 453–458.
- [44] L. Chao-Yi, Integration field beyond the classical receptive field: organization and functional properties, *News Physiol. Sci.* 11 (1996) 181–186.
- [45] J. Jeffs, J.M. Ichida, F. Federer, A. Angelucci, Anatomical evidence for classical and extra-classical receptive field completion across the discontinuous horizontal meridian representation of primate area V2, *Cereb. Cortex* 19 (4) (2009) 963.
- [46] A. Angelucci, J. Bullier, Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J. Physiol.—Paris* 97 (2–3) (2003) 141–154.
- [47] L. Schwabe, K. Obermayer, A. Angelucci, P.C. Bressloff, The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model, *J. Neurosci.* 26 (36) (2006) 9117–9129.
- [48] P. Series, J. Lorenceau, Y. Frégnac, The silent surround of V1 receptive fields: theory and experiments, *J. Physiol.—Paris* 97 (4–6) (2003) 453–474.
- [49] G. Chen, Y. Dan, C.Y. Li, Stimulation of non-classical receptive field enhances orientation selectivity in the cat, *J. Physiol.* 564 (1) (2005) 233–243.
- [50] G. Papari, P. Campisi, N. Petkov, A. Neri, Contour detection by multiresolution surround inhibition, in: *Proceedings of the IEEE*, 2006, pp. 749–752.
- [51] Q. Tang, N. Sang, T. Zhang, Extraction of salient contours from cluttered scenes, *Pattern Recognition* 40 (11) (2007) 3100–3109.
- [52] G. Papari, N. Petkov, Edge and line oriented contour detection: state of the art, *Image Vision Comput.* 29 (2–3) (2011) 79–103.
- [53] C.D. Gilbert, T.N. Wiesel, Receptive field dynamics in adult primary visual cortex, *Nature* 356 (6365) (1992) 150–152.
- [54] M.W. Pettet, C.D. Gilbert, Dynamic changes in receptive-field size in cat primary visual cortex, *Proc. Natl. Acad. Sci.* 89 (17) (1992) 8366.
- [55] A. Das, C. Gilbert, Receptive field expansion in adult visual cortex is linked to dynamic changes in strength of cortical connections, *J. Neurophysiol.* 74 (2) (1995) 779–792.
- [56] T. Tsao, L. Kanak, A scene registration method based on a dynamical receptive field model of biological vision, *Pattern Recognition Lett.* 20 (11–13) (1999) 1423–1430.
- [57] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Opt. Soc. Am., J. A: Opt. Image Sci.* 2 (1985) 1160–1169.
- [58] R.L. De Valois, D.G. Albrecht, L.G. Thorell, Spatial frequency selectivity of cells in macaque visual cortex, *Vision Res.* 22 (5) (1982) 545–559.
- [59] P. Kruizinga, N. Petkov, Nonlinear operator for oriented texture, *IEEE Trans. Image Process.* 8 (10) (1999) 1395–1407.
- [60] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: *Proceedings of the Eighth IEEE International Conference Computer Vision*, vol. 412, 2001, pp. 416–423.
- [61] J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 6 (1986) 679–698.
- [62] W. Richards, H. Nishihara, B. Dawson, *CARTOON: A Biologically Motivated Edge Detection Algorithm*, MIT Press, Cambridge, 1982.
- [63] J.C. Russ, *The Image Processing Handbook*, sixth ed., CRC Press, Boca Raton, Florida, 2011, p. 376.
- [64] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (9) (2011) 1.
- [65] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (5) (2004) 530–549.
- [66] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4) (2002) 509–522.
- [67] A. Ghosh, N. Petkov, Robustness of shape descriptors to incomplete contour representations, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (11) (2005) 1793–1804.
- [68] W.E. Vinje, J.L. Gallant, Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1, *J. Neurosci.* 22 (7) (2002) 2904–2915.
- [69] Z.M. Shen, W.F. Xu, C.Y. Li, Cue-invariant detection of centre-surround discontinuity by V1 neurons in awake macaque monkey, *J. Physiol.* 583 (2) (2007) 581–592.
- [70] H. Wässle, Parallel processing in the mammalian retina, *Nat. Rev. Neurosci.* 5 (10) (2004) 747–757.
- [71] M.E. Hasselmo, Neuromodulation and cortical function: modeling the physiological basis of behavior, *Behav. Brain Res.* 67 (1) (1995) 1–27.
- [72] M.E. Hasselmo, J.M. Bower, Cholinergic suppression specific to intrinsic not afferent fiber synapses in rat piriform (olfactory) cortex, *J. Neurophysiol.* 67 (5) (1992) 1222–1229.
- [73] F. Kimura, Cholinergic modulation of cortical function: a hypothetical role in shifting the dynamics in cortical network, *Neurosci. Res.* 38 (1) (2000) 19–26.
- [74] F. Kimura, M. Fukuda, T. Tsumoto, Acetylcholine suppresses the spread of excitation in the visual cortex revealed by optical recording: possible differential effect depending on the source of input, *Eur. J. Neurosci.* 11 (10) (1999) 3597–3609.
- [75] J.J. Koenderink, The structure of images, *Biol. Cybern.* 50 (5) (1984) 363–370.
- [76] T. Lindeberg, *Scale-Space Theory in Computer Vision*, Springer, New York, 1993.
- [77] M.K. Kapadia, M. Ito, C.D. Gilbert, G. Westheimer, Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys, *Neuron* 15 (4) (1995) 843–856.
- [78] J. Nelson, B. Frost, Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex, *Exp. Brain Res.* 61 (1) (1985) 54–61.
- [79] U. Polat, K. Mizobe, M.W. Pettet, T. Kasamatsu, A.M. Norcia, Collinear stimuli regulate visual responses depending on cell's contrast threshold, *Nature* 391 (6667) (1998) 580–584.



Hui Wei has a Ph.D. at the Department of Computer Science at Beijing University of Aeronautics and Astronautics in July 1998. From September 1998 to November 2000, he was a Postdoctoral Fellow at the Department of Computer Science and Institute of Artificial Intelligence at Zhejiang University. In November 2000, he joined the Department of Computer Science and Engineering at Fudan University. His research interests include artificial intelligence and cognitive science.



Bo Lang received his M.Sc. degree in computer science and technology from the Nanchang University, China, in 2006. He is now pursuing his Ph.D. degree in Artificial Intelligence with the University of Fudan of China. His research interests include visual mechanism modeling and image processing.



Qing-song Zuo received his BS degree in software engineering from the Nanjing University of Science and Technology, China, in 2007. He has been a master candidate in the department of computer science and engineering, Fudan University, China. His research interests include artificial intelligence, cognitive model, computer vision and image processing, with emphasis on models inspired by human visual system.