Pattern Recognition Letters 30 (2009) 985–993

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

A neural contextual model for detecting perceptually salient contours

Wentao Huang *, Licheng Jiao, Jianhua Jia, Hang Yu

Institute of Intelligent Information, Processing and Key Laboratory of Intelligent, Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an 710071, PR China

ARTICLE INFO

Article history:
Received 6 September 2008
Received in revised form 29 March 2009
Available online 18 May 2009

Communicated by J.A. Robinson

Keywords:
Contour detection
Computational model
Surround suppression
Collinear facilitation

ABSTRACT

A computational model, inspired by visual cortical mechanisms of contextual modulation, is presented in this paper, and is applied to detect perceptually salient contours. The presented model incorporates two mechanisms of contextual modulation, surround suppression and collinear facilitation. An oriented filterbank generated by Gaussian derivatives and their Hilbert transform is proposed for pre-processing. The operators of surround suppression and collinear facilitation are applied to the orientation energy resulting from the outputs of oriented filterbank. To avoid augmenting the noise when the facilitation operator enhances the saliency parts, we employ a contrast enhancement transformation for the facilitation operator. For drawing the binary contours, we present an automatic thresholding approach for post-processing. The performance of our model is tested by artificial images with heavy noise and nature images with texture background. Results show that the model has a good performance on extracting the salient contours from images.

1. Introduction

Salient contour detection is one of the major topics in perceptual organization exploring the true meanings for contours. Many works can be found in this area (Canny, 1986; Perona and Malik, 1990; Mallat and Hwang, 1992; Grigorescu et al., 2003, 2004; Tang et al., 2007). It is extremely difficult to extract contours in noisy environments automatically, because it is difficult to make a distinction between contours of objects and edges originating from textured and noise regions.

In the primary visual cortex (V1), neurons are selective to the orientation of the object presented in their receptive fields (RFs). When an appropriate stimulus, such as an oriented bar or a patch of texture, is presented in the RF of a visual neuron, it may drive the cell to evoke action potential responses (classical receptive field, CRF) (Hubel and Wiesel, 1959, 1965). In particular, it was observed that when multiple objects or natural scenes were shown, stimuli placed outside the CRF can modulate the activity evoked by the stimulus placed within the CRF (Maffei and Fiorentini, 1976; Nelson and Frost, 1978; Allman et al., 1985; Carandini and Heeger, 1994; Ferster and Miller, 2000; Albright and Stoner, 2002; Series et al., 2003; Meese et al., 2007). We call this general phenomenon extra-RF (ERF) contextual modulation.

Psychophysical and neurophysiological findings have shown that contextual modulation in the primary visual cortex (V1) plays an important role in the visual information processing (Albright and Stoner, 2002), such as contour integration (Field et al., 1993; Huang et al., 2006), figure-ground segregation (Olzak and Laurinen, 2005), saliency map (Li, 1998; Yen and Finkel, 1998; Grigorescu et al., 2003, 2004; Tang et al., 2007) and so on.

Contextual modulation is a universal phenomenon in the primary visual cortex (V1). It is often allocated to the two categories, suppression and collinear facilitation, which are either weakened or strengthened by the contextual stimuli, respectively. Using the suppression and facilitation, some authors (Li, 1998; Yen and Finkel, 1998; Ursino and La Cara, 2004) proposed dynamic models to extract salient contours of the object. Yen and Finkel (1998) suggested that perceptual saliency should be directly related to the degree of the temporal synchronization of neurons, and proposed a model to extract the salient contour by oscillatory neurons. Li (1998) also use oscillatory neurons to represent contours, but use a different disposition for intracortical synapses. Li assumes that cells with orthogonal preferred orientation are connected by inhibitory synapses, whereas Yen and Finkel assume a facilitatory connection, which they call 'trans-axial', between orthogonal cells. Ursino and La Cara (2004) presented a model of contour extraction and perceptual grouping in the primary visual cortex, which suppressed non-optimally oriented stimuli and guaranteed contrast invariance of the response in combination with feedforward and feedback mechanisms. Using the surround suppression, Grigorescu et al. (2003, 2004) proposed a biologically motivated computational step to improve contour detection in machine vision. Guy and Medioni (1996) proposed a voting process to infer the global salient contours, and in the following work, a unified tensor voting framework was proposed by Medioni and collaborator (Medioni

* Corresponding author. Tel./fax: +86 29 88207354.
E-mail address: wthuang@mail.xidian.edu.cn (W. Huang).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.05.006

et al., 2000). And in the recent work, Loss et al. (2008) presented an iterative multi-scale tensor voting approach for perceptual grouping of oriented segments in highly cluttered images. A recent model proposed by Tang et al. (2007) utilized the surround suppression and co-circularity facilitation to extract contours from images.

A computational model, inspired by visual cortical mechanisms of contextual modulation, is presented in this paper, and is applied to detect salient contours. The present model incorporates two mechanisms of contextual modulation, surround suppression and collinear facilitation. Firstly, an oriented filterbank generated by Gaussian derivatives and their Hilbert transform is presented to express the output response of V1. Then, we give the computation models of surround suppression and collinear facilitation. The performance of models has then been tested by artificial images with heavy noise and nature images with texture background. Results show that the model has a good performance on extracting the salient contours from images. The rest of this paper is organized as follows. In Section 2, we give the detailed description for models. In the following, Section 3 tests the performance of models, which are applied in detecting salient contours from various synthetic and natural images. Finally, we summarize the main conclusions of this work and presents plans for future work in Section 4.

2. Models description

2.1. Orientation energy filters

In area V1, the outputs of neurons are the responses of directional characteristics for the inputs of images. Contour information of images can be described as orientation energies (Adelson and Bergen, 1985; Morrone and Owens, 1987; Morrone and Burr, 1988). Usually, a family of two-dimensional Gabor functions can effectively model the receptive field profiles of simple cells in V1 (Daugman, 1985), and the Gabor energy (Adelson and Bergen, 1985; Morrone and Burr, 1988) is related to a model of complex cells. Here, we adopt an alternate approach, which uses Gaussian derivatives for modeling receptive fields of simple cells and oriented energy for modeling receptive fields of complex cells. Let $G_0^1(x,y)$ be the second derivative of an elongated Gaussian kernel and $G_0^2(x,y)$ be the Hilbert transform of $G_0^1(x,y)$:

$$\begin{cases} G_0^1(x,y) = \frac{d^2}{dy^2}\left(\frac{1}{C}\exp\left(-\frac{x^2}{2(l\sigma)^2}-\frac{y^2}{2\sigma^2}\right)\right), \\ G_0^2(x,y) = \mathrm{Hilbert}(G_0^1(x,y)), \end{cases} \tag{1}$$

where C is a normalization constant and the scale σ determine the size of the filter. The spatial aspect ratio l determines the elongation of the filter. The orientation energy at angle 0^0 is defined as:

$$E_0(x,y) = \sqrt{(I * G_0^1)^2 + (I * G_0^2)^2}, \tag{2}$$

where the symbol $*$ denotes the convolution operator, $I(x,y)$ is an input image. $E_0(x,y)$ has maximum response for horizontal direction contours. Rotated copies of the two filter kernels will be able to pick up edge contrast at various orientations θ, we define them as $G_\theta^1(x,y)$, $G_\theta^2(x,y)$ and $E_\theta(x,y)$, respectively. For a number of M sampling orientations θ_m:

$$\theta_m = \frac{(m-1)\pi}{M}, \quad m = 1,2,\ldots,M, \tag{3}$$

$\theta_m \in [0,\pi)$, we can get M different energy responses $E_{\theta_m}(x,y)$ ($m = 1,2,\ldots,M$), and define the orientation energy and the orientation at each pixel as:

$$\begin{cases} r(x,y) = \max\{E_{\theta_m}(x,y) | m = 1,2,\ldots,M\}, \\ \theta(x,y) = \arg\max_{\theta_m}\{E_{\theta_m}(x,y) | m = 1,2,\ldots,M\}, \end{cases} \tag{4}$$

which can be characterized as a vector:

$$\boldsymbol{R}(x,y) = r(x,y)(\cos\theta(x,y), \sin\theta(x,y)). \tag{5}$$

The number of sampling orientation M, trade-off computational cost and performance, is set to 12 in this paper. Fig. 1 shows these filter kernels defined in Eq. (1).

The orientation energy defined in above has some nice properties. The second derivative of the gaussian and the Hilbert transform of it are a quadrature pair. Quadrature pairs are phase independent and they will not be affected by exact localization of the edges. Filters are insensitive to linear variations of intensity,such as smooth shading. They do not respond well to high curvature contours. This property agrees with human perception that curvilinear grouping does not occur over high curvature contours. Moreover, they can respond well to contours under strong noise environment, which will be illustrated in Section 3.1.

2.2. Suppression effects

Electrophysiological studies have found that neurons in V1 show significant suppression when increasing the diameter of a central grating beyond the CRF, or adding to the central grating a large annular iso-oriented surround (Carandini and Heeger, 1994; Walker et al., 2000). Stimuli presented in surrounding regions can suppress the response of cortical neuron to stimulus in the RF when they are tuned to the cross-oriented or the same orientation (sometimes called co-oriented) as that of the center of the RF, and this is referred to as surround suppression (Morrone et al., 1982; Levitt and Lund, 1997; Meese et al., 2007). And suppression has also been found for cross-oriented masks in the surround although the effects are weaker than in the co-oriented case, and suppression degree depends on the orientation contrast of stimuli in the surround and the RF. Suppression is the strongest when the surround and center stimulus orientations are the same and is the weakest when they are orthogonal to each other (Knierim and van Essen, 1992; Levitt and Lund, 1997; Kapadia et al., 2000; Rossi et al., 2001; Bair et al., 2003).

In order to describe the suppression varying according to the orientation similarity and diameter, we give the following model. We define a normalized vectorial weighting function,

$$\boldsymbol{S}_\theta(x,y) = s_\theta(x,y)(\cos\theta, \sin\theta), \tag{6}$$

where $s_\theta(x,y)$ is of the dissociated Gaussians distance between the surround and the center, and satisfies:

$$\begin{cases} s_\theta(x,y) = \frac{1}{Z_s}\exp\left(-\frac{\varphi_\theta^2}{2\sigma_\varphi^2}\right)\left(\exp\left(-\frac{\left(\sqrt{x^2+y^2}-\sigma_s\right)^2}{2\sigma_s^2}\right)\right. \\ \qquad\qquad \left.+\exp\left(-\frac{\left(\sqrt{x^2+y^2}+\sigma_s\right)^2}{2\sigma_s^2}\right)\right), \\ \varphi_\theta = \mathrm{atan}(\frac{x'}{y'}), \\ x' = x\cos\theta + y\sin\theta, \\ y' = -x\sin\theta + y\cos\theta, \end{cases} \tag{7}$$

where Z_s is a normalizing constant, θ is the direction parameter and is a constant for a specific sampling orientation. The standard deviation σ_φ determines the size of angle of the surround suppression. Suppression effects decline with increasing angle φ_θ, and have the maximal weight at the vertical direction. σ_s is the standard devia-

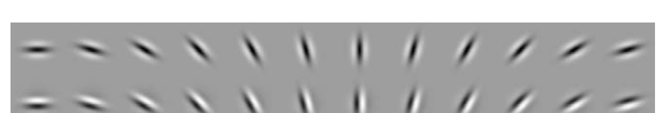

Fig. 1. The filter kernels definied in Eq. (1), $G_{\theta_m}^1(x,y)$ (top) and $G_{\theta_m}^2(x,y)$ (bottom), $\theta_m \in [0,\pi)$, $M = 12$, $\sigma = 2$, $l = 2.5$.