Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 3100–3109
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Extraction of salient contours from cluttered scenes

Qiling Tang, Nong Sang*, Tianxu Zhang

Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, PR China

Received 16 June 2006; received in revised form 7 February 2007; accepted 19 February 2007

Abstract

The responses of neurons in the primary visual cortex (V1) to stimulus inside the receptive field (RF) can be markedly modulated by stimuli outside the classical receptive field. The modulation, relying on contextual configurations, yields excitatory and inhibitory activities. The V1 neurons compose a functional network by lateral interactions and accomplish specific visual tasks in a dynamic and flexible fashion. Well-organized structures and conspicuous image locations are more salient and thus can pop out perceptually from the background. The excitatory and inhibitory activities give different visual physiological interpretations to the two kinds of saliencies.

A model of contour extraction, inspired by visual cortical mechanisms of perceptual grouping, is presented. We unify the dual processes of spatial facilitation and surround inhibition to extract salient contours from complex scenes, and in this way coherent spatial configurations and region boundaries could stand out from their surround. The proposed method can selectively retain object contours, and meanwhile can dramatically reduce non-meaningful elements resulting from a texture background. This work gives a clear understanding for the roles of the inhibition and facilitation in grouping, and provides a biologically motivated computational strategy for contour extraction in computer vision.

Keywords: Contour extraction; Contextual interactions; Surround inhibition; Spatial facilitation; Saliency

1. Introduction

Contour defining object shape is one of the most important features in object recognition tasks. However, it is extremely difficult to extract contours from cluttered scenes automatically. To do that, two major problems need to be solved: (I) eliminating non-meaningful edges engendering from texture fields rather than object boundaries, (II) grouping local elements into meaningful global features according to context information. Human visual system, which possesses a remarkable computational accomplishment, can rapidly and effectively perceive well-organized structure from abundant low-level features in complex scenes, and this provides a biologically plausible strategy for overcoming the difficulties mentioned above.

Psychophysical and neurophysiological findings [1–4] have shown that information about the stimulus of one cell can be transferred to that of its surrounding cells through the horizontal connections in V1, thus the cortical cell can be taken as a part of an interactional network rather than an isolated element. The interactions of neurons in area V1 play an important role in the process of visuospatial integration and have been interpreted as the neural substrate of a variety of psychophysical phenomena, such as contour integration, figure-ground segregation and surface perception. The contextual interactions depend on the relative orientation and spatial position of stimulus within the RF and stimuli falling outside the RF (surrounding regions of the RF) [5,6]. Stimuli presented in surrounding regions can suppress the response of cortical neuron to stimulus in the RF when they are tuned to the same orientation as that of the center of the RF, and this is referred to as surround inhibition; and on the other hand, they can also enhance the response of V1 neuron when they are aligned with the center to form collinear contextual stimuli, and this is called spatial facilitation. Inhibitory interactions are supposed to play a more important role in the segmentation of surfaces and textures [7], while excitatory contextual interactions are deemed to be more significant in contour integration and figure-ground segregation [8].

A number of different visual computational models have been proposed to extract salient contours of the object based on

* Corresponding author. Tel.: +86 27 8755 3671; fax: +86 27 8754 3594.
E-mail addresses: tqlinn@sohu.com (Q. Tang), nsang@hust.edu.cn (N. Sang).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.02.009

V1 characteristics, from which we benefited much. Field et al. [9] assumed that there exists an "associated field" linking neurons with similar orientation preference and proximally spatial location, thus grouping local edges with collinear alignment into a smooth and "elastic" contour. Li [10] proceeded cumulative propagation according to the orientations of stimuli inside and outside the RF and located different region boundaries by detecting the breakdown of spatial homogeneity in features. This model was achieved by using very complex neurons at a much high computational cost. Yen and Finkel [11] suggested that perceptual saliency should be directly related to the degree of the temporal synchronization of neurons, and they extracted salient contour by temporal binding. Ursino et al. [12] suppressed non-optimally oriented stimuli and guaranteed contrast invariance through the combination of feedforward and feedback mechanisms. These visual models of contour perception, primarily designed to interpret how human visual system groups local elements into global contours up to now, have only been of theoretical interest, and they have been mainly applied to artificial images.

In the computer vision field, the challenging work of extracting contours has also been approached in many different ways. A recent model proposed by Grigorescu et al. [13,14] utilized the non-classical receptive field inhibition to reduce texture edges and to retain isolated contours in real images. However, the object contours embedded in a cluttered background would be subjected to a certain degree of destruction due to the absence of excitatory interactions. Deng and Clausi [15] developed a simple MRF model to locate texture boundaries, in which a variable weighting parameter is used to combine the region labeling component and the feature-modeling component, but this approach for texture boundaries cannot detect brightness edges effectively. On the other hand, the properties of proximity and collinearity are utilized to group contours. For example, Sha'shua and Ullman [16] introduced a saliency network through several low-level features—contour length and curvature. Guy and Medioni's model [17] let each image element cast a 'vote' on the position and salience of every other image location. Farag and Delp [18] linked the goal nodes with a maximum likelihood metric by sequential search. Zhu et al. [19] detected underlying boundaries by minimizing the directional potential function. These approaches work well on contour grouping in noisy images, but they fail to deal with cluttered and textured regions. More recently, Wang et al. [20] presented a ratio-contour method for finding closed contours as maximum likelihood cycles in a graph which encodes geometric relationships between neighboring edges. Our model presented in this paper does not utilize the effect of closure and our detected results contain non-closed curves.

The present model, inspired by visual cortical mechanisms of perceptual grouping, unifies the dual processes of spatial facilitation and surround inhibition to extract salient contours from the cluttered background. For complex natural scenes, we considered two different cases: (I) coherent configurations, as shown in Fig. 1a, embedded in a field of randomly oriented distractor elements, where excitatory interactions enhance the neuron's response to elements with well-organized structure, enabling them to pop out perceptually from their environment; (II) region boundaries, the regions with uniform textures (similar orientation) are shown in Fig. 1b, where homogeneities in inputs break down at texture boundaries where there is a change in orientation, thus boundaries suffer less iso-orientation surround inhibition than other part, leading to relatively higher neural activity near the texture boundaries, making them stand out from their surround. Both cases provide psychophysical and physiological data for extraction of salient contours from cluttered scenes.

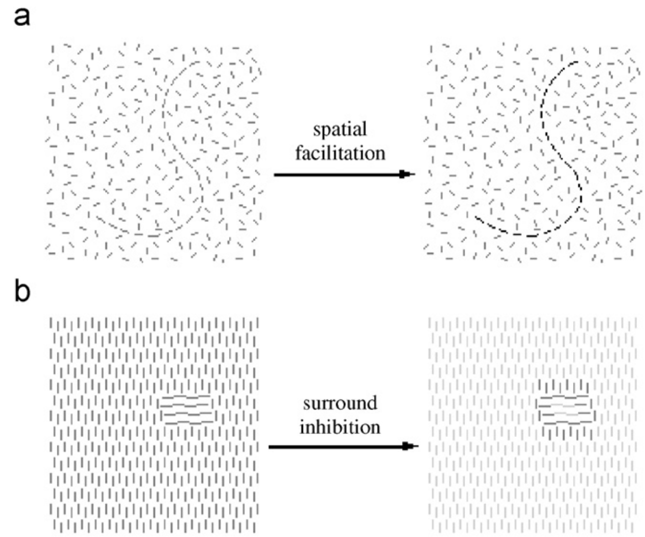

Fig. 1. Role of facilitatory and inhibitory interactions in contour extraction, (a) spatial facilitation, standing out coherent configurations; (b) iso-orientation surround inhibition, standing out boundaries of texture regions, adapted from Ref. [21].

The rest of this paper is organized as follows. Section 2 details how to construct a model of contour extraction based on the characteristics of the primary visual cortex. Section 3 discusses some practical aspects in the application of the method and then tests the performance of this algorithm through various synthetic and natural images. Finally, Section 4 summarizes the main conclusions of this work and presents plans for future work. A preliminary version of the present paper can be found in a proceeding volume [22].

2. Contour extraction

Neurons in V1 are selective for the orientation, spatial frequency and direction of motion [23,24]. The visual inputs are first decomposed by channels tuned to specific properties, and interactions between these channels make elements with coherent spatial configurations and elements with distinctive property pop out perceptually from their surround. Sha'ashua and Ullman [16] referred to the two kinds of saliencies as "structural saliency" and "local saliency", respectively. Spatial facilitation and surround inhibition give the two saliencies different biological interpretations—collinear excitation enhances well-organized structures and homogeneous suppression makes elements distinguishing from uniform surround more salient.