Boundary Detection Using Double-Opponency and Spatial Sparseness Constraint

Kai-Fu Yang, Shao-Bing Gao, Ce-Feng Guo, Chao-Yi Li, and Yong-Jie Li, *Member, IEEE*

Abstract—Brightness and color are two basic visual features integrated by the human visual system (HVS) to gain a better understanding of color natural scenes. Aiming to combine these two cues to maximize the reliability of boundary detection in natural scenes, we propose a new framework based on the color-opponent mechanisms of a certain type of color-sensitive double-opponent (DO) cells in the primary visual cortex (V1) of HVS. This type of DO cells has oriented receptive field with both chromatically and spatially opponent structure. The proposed framework is a feedforward hierarchical model, which has direct counterpart to the color-opponent mechanisms involved in from the retina to V1. In addition, we employ the spatial sparseness constraint (SSC) of neural responses to further suppress the unwanted edges of texture elements. Experimental results show that the DO cells we modeled can flexibly capture both the structured chromatic and achromatic boundaries of salient objects in complex scenes when the cone inputs to DO cells are unbalanced. Meanwhile, the SSC operator further improves the performance by suppressing redundant texture edges. With competitive contour detection accuracy, the proposed model has the additional advantage of quite simple implementation with low computational cost.

Index Terms—Boundary, color opponent, receptive field, visual system, texture suppression.

I. Introduction

OBJECT boundaries represent important cues for visual perception such as scene understanding and object recognition [1]. Boundary detection is also a fundamental building block for a large variety of computer vision applications, such as image segmentation [2] and object detection [2], [3]. Among numerous computational boundary detection, typical methods include zero crossing [4], phase congruency [5], Canny detector [6], level set [7]–[9], and
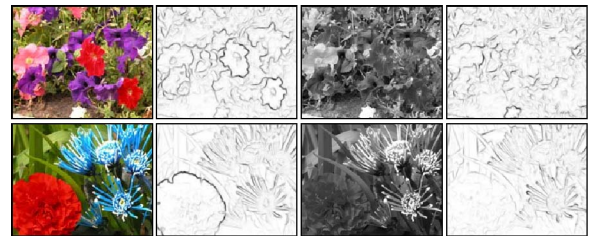
Fig. 1. Examples showing that color boundaries are lost when ignoring color information. Color images and their boundary maps (the first and second column) provide more object information than gray-level images and their boundary maps (the third and last column).

oriented energy approaches [10]–[12] (see [3] for reviews). However, most traditional edge detection methods usually extract edges by computing the abrupt change of local luminance, and normally are not capable of distinguishing boundaries from abundant of textured edges. To endow the boundary detection models with human-level performance, many groups have been making extensive efforts, for example, to combine more visual features extracted from the scenes (see [2], [13]).

As a basic feature of external world, color information plays an important role in human visual perception such as shape and object recognition [14]. From the perspective of engineering, color is also necessary for various image processing tasks, such as edge detection [13], [15], image segmentation [16], junction/corner detection [15], [17]. Fig. 1 shows typical examples illustrating that some important contours of objects (e.g., flowers in the first column) in color images are lost in the gray-scale space, especially for those boundaries with only color contrast in regions of iso-luminance.

A. Related Work

In order to detect boundaries from color images, many early studies focused on extending the standard edge detectors, such as Canny [6], to color space. These methods are inherently difficult to discriminate salient object boundaries and texture edges due that they respond to all the discontinuities in image intensity, color or texture. In recent decades, many new approaches have been developed for boundary detection in complex scenes. Typically, in the famous *Pb* method, Martin *et al.* [13] took into account multiple local cues (i.e., color, brightness, and texture) and combined these cues with certain learning technique to detect and localize the boundaries. Other learning-based methods tried to take multiple scales [18], more local features [19] or global

information [2], [17] for better results. Recent methods also improve contour detection by learning to classify the various SIFT features [20] or sparse code features [21], [22] extracted at multiple scales. However, the performances of most learning-based methods mentioned above are dependent on the appropriate selection of training sets, which makes the methods inflexible for individual images. Furthermore, the high computational cost resulted from training needs to be carefully dealt with.

Another important issue is to make the salient contours pop out in cluttered scenes. There are mainly two classes of methods including contour grouping and texture suppression. Contour grouping methods usually integrate low-level elements produced by basic edge detectors into mid-level features. For example, Zhu *et al.* [23] proposed a contour grouping method with the topological formulation called Untangling Cycles. Ren *et al.* [24] presented a model to enforce the curvilinear continuity with Conditional Random Fields framework. By utilizing the Gestalt rules (i.e., good-continuation, proximity, contour-closure, etc.), existing methods introduced the local interactions between contour segments [25], [26] and global effect [27] to extract perceptually salient contours. Salient contours were also extracted by solving the min-cover problem [28] or building Ultrametric Contour Maps [2], [29].

Texture analysis methods have also been used to suppress the undesired textured edges while extracting boundaries. For example, texton-based approaches have been developed for texture analysis and image segmentation [33], [34]. Martin *et al.* further developed the Texture-Gradient (TG) for texture boundary detection [13]. These detectors respond well to texture-defined boundaries and are insensitive to unwanted edge segments within homogeneous textured regions. However, texton-based methods usually take high computational cost on multiple convolution operations and high-dimensional analysis. Recently, some more time-saving texture boundary detection algorithms have been proposed. For example, biologically inspired surround inhibition methods make texture boundaries pop out by suppressing the unwanted short edges surrounded by similar textured patterns [35]–[38]. Hidayat *et al.* detected texture boundaries almost in real-time by extracting ridges in the standard deviation space [39].

It is exciting to see that several more recent models are capable of obtaining a quite well balance between boundary detection accuracy and computational cost, especially for the multi-scale and learning-based frameworks. For example, Leordeanu *et al.* [40] proposed a generalized boundary detection model (termed *Gb*), which estimates a measure of boundary strength for each location in a closed-form computation by utilizing multiple layers of image representations of various features at both low- and mid-levels. Then a logistic model learned from training set is used to obtain the probability of boundary. This *Gb* model achieves competitive results at a quite lower computational cost. More recently, Lim *et al.* [41] proposed a quite fast and accurate edge detector that learns and detects mid-level features, called *Sketch Tokens*, to capture local edge structure. They learn the features of sketch tokens using supervised mid-level information from human labeled edges in natural images, and then classify edge patches of new images into sketch tokens using random forest classifiers. In order to discard the requirement of pre-defined classes of edge patches in sketch tokens based model, Dollár and Zitnick [42] proposed a novel framework of structured random forests to learn more subtle variations in edge structure by taking advantage of the inherent structure in edge patches. This model can achieve state-of-the-art results almost in realtime.

Contrary to the recent trends employing higher-order or higher-level information for boundary detection as mentioned above, Isola *et al.* [43] introduced a very local variance termed pointwise mutual information (PMI) as the statistical association between pixels to predict whether or not two pixels belong to the same object. Then a spectral clustering was used to get segments and boundaries. The state-of-the-art results obtained by this model suggest that with the help of certain grouping operation, it is possible to achieve excellent boundary detection performance by only using very local information and low-dimensional feature spaces [2], [43].

Along another line, it has a long history that to employ early visual mechanisms for image analysis, such as texture discrimination in gray-scale images [44], [45]. Recently, the increasing success of biologically based methods for edge detection in gray-scale images [35]–[38], [46], [47] motivated us to build a biologically inspired framework for color boundary detection in natural images in an effective way. Several methods based on the color-opponent channels, i.e., Red-Green (R-G) and Blue-Yellow (B-Y) channels found in the visual system, have exhibited promising performance on color boundary detection. Martin *et al.* [13] computed the color gradients in these two single-opponent channels for color boundary detection. Zhou and Mel [48] applied custom "pairwise difference" oriented edge detector on the smoothed R-G and B-Y. More recently, Zhang *et al.* [49] proposed a new color descriptor based on color-opponent mechanisms, which improves the performances of several classical object recognition and boundary detection systems by extending them from grayscale to color space. However, one of the key limitations of these color-opponency based approaches is that the color-opponent channels in them are blind to the luminance-defined boundaries when no luminance channel is explicitly included [13], [49].

B. Color Mechanisms in Early Visual System

Color processing in the human visual system (HVS) progresses through a series of hierarchical steps [30], [50], [51]: after the light absorption by cone photoreceptors, cone activities are transmitted via horizontal cells, bipolar cells, etc, and then compared by cone-opponent retinal ganglion cells (RGCs); these color signals are transmitted via the Lateral Geniculate Nucleus (LGN) to the primary visual cortex (V1) and then higher cortical areas. One of the amazing properties in the early stages of HVS, i.e., retina $\rightarrow$ LGN $\rightarrow$ V1, is on the color coding, which can be summarized as follows:

- *Trichromacy*. The first stage of color processing of HVS takes place in the photoreceptor layer of the retina. There are two types of photoreceptors: rods and cones, and
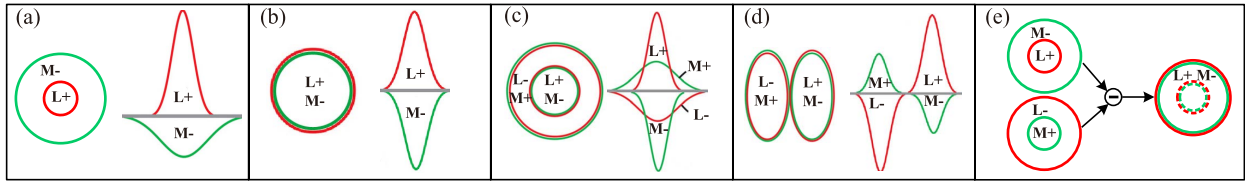
Fig. 2. The receptive fields of single-opponent cells of Type I (a) and Type II (b) in RGC and LGN levels, and double-opponent cells in V1 with concentric RF(c) and oriented double-opponent cells in V1 with side-by-side spatially antagonistic regions with unbalanced cone weights (d). In the expression of "A+" or "B-", the sign "+" and "-" denote the role of excitation and inhibition, respectively. Adapted from [30] and [31]. (e) An illustration to explain that the center-only RF of Type II in LGN is constructed by differencing two center-surround ganglion cells. Adapted from Fig. 38 in [32].

cones are responsible for color vision. There exist three kinds of cone photoreceptors, namely L, M, and S cones, which preferably absorb long, middle, and short wavelengths in a local spatial space, respectively. This is well known as trichromacy. As described in many other literatures, in the following the L-, M- and S-cones will be also referred to as red (R), green (G), and blue (B) cones, though each cone class does not specifically code the perception of a single color of R, G or B.

- *Color opponency*. Many researchers have reported that color information is processed in the visual pathway in an opponent way, i.e., red versus green (red-green, or R-G) and blue versus yellow (blue-yellow, or B-Y) channels [50]. Responses to the two colors of an opponent channel are antagonistic to each other, which makes opposite opponent colors never be perceived together. For example, there is no color like "greenish red" or "yellowish blue".

- *Color opponent cells*. The RGCs or LGN cells have been found to have single-opponent (SO) receptive field (RF) and some kinds of cells in V1 have double-opponent (DO) RF [30], [51]. There are mainly two types of single-opponent cells: Type I cells have center-surround opponent RF (Fig. 2(a)), and in contrast, Type II cells have center-only opponent RF (Fig. 2(b)). It has been found that these single-opponent cells come in four varieties, i.e., L-on/M-off (or L+M-), M-on/L-off (or M+L-), S-on/(L+M)-off (or S+(L+M)-), and (L+M)-on/S-off (or (L+M)+S-), where "on" and "off" correspond to the RF center and surround, respectively. In V1, the RF of DO neurons shows more complex properties. Their RFs are both chromatically and spatially opponent [52], which was considered as an important role in color scene understanding [49]. Especially, it has been reported that some DO neurons in V1 have concentric center-surround RF (Fig. 2(c)) or side-by-side spatially oriented RFs (Fig. 2(d)). The DO cells with concentric RF shown in Fig. 2(c) were thought to be physiological building blocks of color constancy and color contrast [30], [51], and our previous model has shown that such DO cells contribute to color constancy by coding the external light source color [53], [54]. Differently, the DO cells with oriented RF shown in Fig. 2(d), which will be specifically modeled in this study, have been assumed to play crucial role in boundary detection in (color) natural scenes [55]. Physiological experiments indicated that the DO cells shown in Fig. 2(d) can respond well to
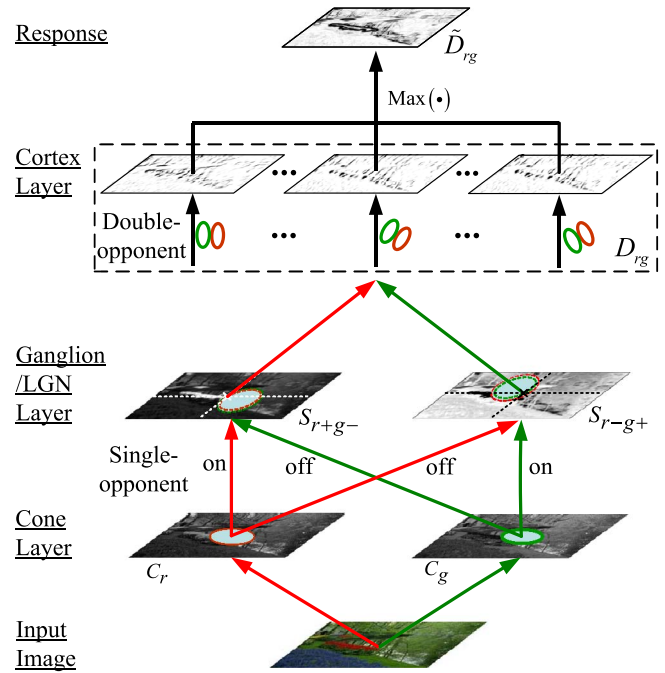

Fig. 3. The flowchart of color-opponent processing in the R-G channel. The similar computational steps are used in the other channels.

color-defined boundaries when the cone weights are well balanced [31], and can respond sensitively to both achromatic and iso-luminant gratings when the cone weights are unbalanced [30], [51]. It has been physiologically hypothesized that the RF structure shown in Fig. 2(d) was composed of two side-by-side subregions, each of which resembles the receptive field of a Type II cell shown in Fig. 2(b) [31].

C. Contributions of the Proposed Work

To summarize the above, most of the existing models utilize the two basic cues of brightness and color normally by first extracting edges from each cue, and then combining them based on certain strategy, e.g., supervised learning [13]. This work aims to utilize these two basic cues in a more efficient way for boundary detection. We describe our new contributions of this paper as follows.

First, this paper proposes a novel contour detection model based on the color-opponent mechanisms of the biological visual system by specifically simulating the DO cells with oriented RF shown in Fig. 2(d). The new model includes three layers simulating the levels of retina, LGN, and V1 (Fig. 3).

Particularly, in the last layer (Cortex layer), a pool of oriented DO cells with different preferred orientations is used at each location to extract boundaries by receiving the responses of SO LGN cells, followed by a max operator across all orientations to combine responses to boundaries in separate DO channels. Finally, we compute the maximum to combine the boundaries across all DO channels. To our knowledge, this work is the first attempt to introduce the DO mechanism of color-sensitive V1 cells with oriented RF, a very important group of cells in V1, for detecting boundaries.

Second, this work also develops a new texture suppression method with spatial sparseness constraint (SSC). We suppress the neuronal responses to the edges in the local regions with low sparseness measure. This operator works well because the local regions containing (unwanted) regularly distributed textures tend to exhibit lower local sparseness measure, while the regions covering salient boundaries usually have high spatial sparseness in response.

As briefly mentioned above, in this work, we simulate the biological mechanisms of color information processing along the Retina-LGN-Cortex visual pathway and propose a feedforward hierarchical system for boundary detection in real natural scenes by using only low-level local information. The results on a commonly used dataset will show that our model has the capacity of jointly detecting the color- and luminance-defined boundaries and efficiently suppressing textural edges. In addition, our model has obvious advantage in saving computational cost. This paper is based on our previous work reported in a conference proceeding [56], which is substantially extended here mainly by introducing the SSC-based texture suppression and more performance analysis. The source code is available at http://www.neuro.uestc.edu.cn/vccl/.

The remaining of this paper is organized as follows. In Section II, we describe the details of the proposed boundary detection system, including color opponent system and texture suppression with SSC. In Section III, we evaluate the performance of the proposed model on BSDS300/500 datasets. Finally, we draw conclusions in Section IV.

II. BOUNDARY DETECTION SYSTEM

A. From Single- to Double-Opponent Processing

Our framework shown in Fig. 3 is a feedforward hierarchical model including three layers, which correspond to the levels of retina, LGN, and V1 of the visual system, respectively. Based on the physiological hypothesis that the two subregions of the RF of oriented DO cells Fig. 2(d) resemble the RF of a Type II cell shown in Fig. 2(b) [31], we model that each DO V1 cell receives the neuronal responses of two single-opponent LGN cells of Type II. In Fig. 3, we just show the computational steps in the R-G channel, and information processing along another channel of B-Y shares the similar computational steps.

1) Cone Layer: At the layer of cone photoreceptors, the input color image is first separated into three components: $I_r(x, y)$ for red (R), $I_g(x, y)$ for green (G), and $I_b(x, y)$ for blue (B), which are respectively sent into L, M, and S cones.

In addition, when the information from the cones is passed forward via horizontal cells, bipolar cells, etc., to the retinal ganglion cells, the output layer of the retina, a yellow (Y) component is constructed by a kind of bipolar cells that receive both R and G cone signals, i.e., $I_y(x, y) = 0.5(I_r(x, y) + I_g(x, y))$, which will be then sent to the single-opponent ganglion cells of B-Y type.

2) Ganglion/LGN Layer: The majority of ganglion cells in retina have center-surround RFs, which send information to LGN, a place that is widely regarded as a relay center between the retina and V1 [31]. Many physiological findings reveal that the ganglion and LGN cells have similar RF properties (e.g., single-opponent), and the main difference is that LGN cells have relatively larger RFs [31]. Meanwhile, physiological studies have also reported that Type II cells with center-only RFs do exist in the dorsal layers of LGN [57], though they are in the minority. It has been suggested that the RF of a Type II LGN cell could be constructed by differencing two center-surround SO ganglion cells [32], [58], as shown in Fig. 2(e). Based on this idea, we unify the ganglion and LGN layers into a single processing by center-only LGN cells.

We first define

$$C_k(x, y; \sigma) = I_k(x, y) * gf(x, y; \sigma); k \in \{r, g, b, y\} \quad (1)$$

$$gf(x, y; \sigma) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

where $*$ is a convolution operator.

Then, the response of the SO cells of R-G type (including R-on/G-off and G-on/R-off) is computed as

$$S_{rg}(x, y) = w_1 \cdot C_r(x, y; \sigma) + w_2 \cdot C_g(x, y; \sigma) \quad (3)$$

$$where \begin{cases} w_1 \cdot w_2 \leq 0 \\ |w_1|, |w_2| \in [0, 1] \end{cases} \quad (4)$$

where w_1 and w_2 are the connection weights from cones to RGCs. w_1 and w_2 always have opposite signs. With $w_1 > 0$ and $w_2 < 0$, we obtain the responses of R-on/G-off (or R+/G-) cells, and with $w_1 < 0$ and $w_2 > 0$, we get the responses of G-on/R-off(or G+/R-) cells. Note that though the RF scales of ganglion cells vary (e.g., from fovea to periphery), here we model all ganglion cells with same RF scales (σ), which implicitly indicates that we only care about the stimuli within the fovea, which are processed and transmitted to the ganglion cells with almost same (fine) RF scales.

SO cells in the ganglion/LGN layer are important for separating color and achromatic information. As indicated by (3) and Fig. 2(b), when the weights of different cone inputs are balanced (with opposite signs), i.e., $|w_1| = |w_2|$, the SO ganglion/LGN cells are blind to achromatic information, because achromatic information (i.e., $R = G = B$) provides equal but antagonistic components in a SO channel. In contrast, when the cone weights are unbalanced (i.e., $|w_1| \neq |w_2|$), the SO model cell can respond to both color and luminance information, due that the two cone inputs with opposite signs can not cancel each other out. Such properties of SO cells are crucial for the functional role of the DO cells shown in Fig. 2(d), which will be modeled as a combination of two such SO cells, as described below.
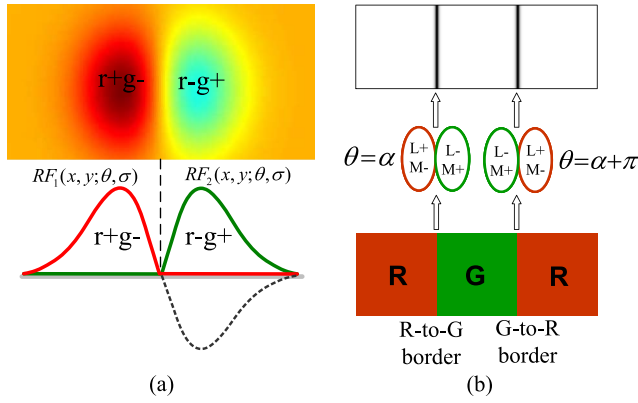
Fig. 4. An illustration of the Gaussian first derivative (GFD) based receptive field (RF) model of the oriented double-opponent (DO) neurons in V1 with vertical orientation. (a) The RF with two side-by-side subregions could be modeled as the combination of the left part of GFD curve (the red curve) and the reverse of the right part of GFD (the green curve), which receive respectively opposite single-opponent signals from LGN, e.g., r+g- and r-g+. (b) Two DO neurons with 180 degree difference between their preferred orientations ($\theta = \alpha$ and $\theta = \alpha + \pi$) detect the color boundaries with the same orientation but different polarities (i.e., R-to-G and G-to-R).

3) Cortex Layer: In the cortex layer of V1, the RFs of most color- and color-luminance-sensitive neurons are both chromatically and spatially opponent. Mathematically, the DO RF with two side-by-side spatially antagonistic regions shown in Fig. 2(d) can be modeled using the first-order (partial) derivative of a two-dimensional (2D) Gaussian given by

$$V(x, y; \theta, \sigma) = \frac{\partial f(x, y; \theta, \sigma)}{\partial \widetilde{x}} \tag{5}$$

$$f(x, y; \theta, \sigma) = \frac{1}{2\pi (k\sigma)^2} exp\left(-\frac{\widetilde{x}^2 + \gamma^2 \widetilde{y}^2}{2(k\sigma)^2}\right) \tag{6}$$

$$\begin{bmatrix} \widetilde{x} \\ \widetilde{y} \end{bmatrix} = \begin{bmatrix} x\cos(\theta) + y\sin(\theta) \\ -x\sin(\theta) + y\cos(\theta) \end{bmatrix} \tag{7}$$

where γ is the spatial aspect ratio of Gaussian that controls the ellipticity of RF. In this study we set $\gamma = 0.5$ based on the physiological finding [30], [51]. θ is the preferred orientation of a given cell. $k\sigma$ determines the RF size of V1 neurons, where σ is same as the scale of Gaussian filters used in the ganglion/LGN layer. We always set $k = 2$ for different scenes, which means that the RF of V1 neurons is 2 times larger (in diameter) than that of ganglion/LGN cells [59].

Fig. 4(a) shows an example of Gaussian derivative based RF model of an oriented DO neuron with vertical preferred orientation. The vertical axis separates the RF into two side-by-side spatially antagonistic subregions: the left part receives the R+/G- response and the right receives the G+/R- response of the SO ganglion/LGN cells. The vertical axis indicates the orientation of the boundaries that can be extracted by this DO neuron.

Computationally, the response of oriented DO cells can be calculated by directly convoluting the response map of SO cells in LGN given by (3) with the oriented RF filter. That is

$$D_{rg}(x, y; \theta_i, \sigma) = S_{rg}(x, y) * RF_1(x, y; \theta_i, \sigma)$$
$$- S_{rg}(x, y) * RF_2(x, y; \theta_i, \sigma) \tag{8}$$

$$RF_1(x, y; \theta_i, \sigma) = \mathbb{H}(V(x, y; \theta_i, \sigma)) \tag{9}$$

$$RF_2(x, y; \theta_i, \sigma) = \mathbb{H}(-V(x, y; \theta_i, \sigma)) \tag{10}$$

$$\mathbb{H}(s) = \begin{cases} s, & s > 0 \\ 0, & s \leq 0 \end{cases} \tag{11}$$

where $*$ denotes the convolution operator. $\theta_i \in [0, 2\pi)$ are N_θ different orientations computed at $\theta_i = (i-1)2\pi/N_\theta$, $i = 1, 2, \cdots, N_\theta$. We set $N_\theta = 16$ in this work. Fig. 4(b) shows that the neuron with $\theta = \alpha$ responds preferably to the R-to-G boundaries, and in contrast, the neuron with $\theta = \alpha + \pi$ can most suitably detect the G-to-R boundaries. Taken together, a set of cells with various preferred orientations $\theta_i \in [0, 2\pi)$ can be used to detect the boundaries defined by the same color pairs but with different polarities and orientations.

Then, in order to detect edges with any orientation, a max-pooling operator (a canonical neural computation [60]) is applied across a pool of DO cells with various orientations at each location to obtain the final DO response to the boundary in R-G opponent channel according to

$$\widetilde{D}_{rg}(x, y; \sigma) = max\{D_{rg}(x, y; \theta_i, \sigma)|i = 1, 2, \cdots, N_\theta\} \tag{12}$$

The final responses in other three DO channels like $\widetilde{D}_{gr}$, $\widetilde{D}_{by}$, and $\widetilde{D}_{yb}$, could be computed with the similar steps described by (3)$\sim$(12). To make the four DO channels comparable, the output of each DO channel is separately normalized linearly to [0, 1]. Note that $\widetilde{D}_{rg}$ and $\widetilde{D}_{gr}$ are not equivalent due to the unbalanced cone weights as shown in Fig. 2(d).

B. Texture Suppression With Spatial Sparseness Constraint

The outputs obtained by (12) usually include undesired responses to textured edges. Here we propose a new method to suppress the responses to unwanted textures by introducing the spatial sparseness constraint (SSC). It has been well recognized that our visual system represents the natural scenes with a quite efficient style, such as sparse coding [61], [62]. Sparse coding models have been used to account for the responsive properties of V1 neurons [62], [63]. From the perspective of engineering, sparse features have been trained to discriminate contours of objects [22]. In specific, Alpert *et al.* [64] employed the sparseness measure proposed in [65] to distinguish the textured and non-textured regions based on the assumption that the non-textured regions are fairly sparse. Alpert *et al.* computed the sparseness measure as [64]

$$sparseness(x, y; \vec{h}) = \frac{1}{\sqrt{n} - 1}\left(\sqrt{n} - \frac{\|\vec{h}(x, y)\|_1}{\|\vec{h}(x, y)\|_2}\right) \tag{13}$$

where $\vec{h}(x, y)$ denotes the gradient's magnitude histogram of a local region centered at (x, y), and n is the dimensionality of $\vec{h}(x, y)$. $\|\vec{h}(x, y)\|_p$ denotes the ℓ_p norm of $\vec{h}(x, y)$, i.e., $\|\vec{h}(x, y)\|_1 = \sum |h_i|$, $\|\vec{h}(x, y)\|_2 = \sqrt{\sum |h_i|^2}$. Considering that information about the spatial distribution of pixels is lost in the histogram-based analysis, it would be inaccurate to evaluate whether or not a local region is characterized by texture (Fig. 6(b)).
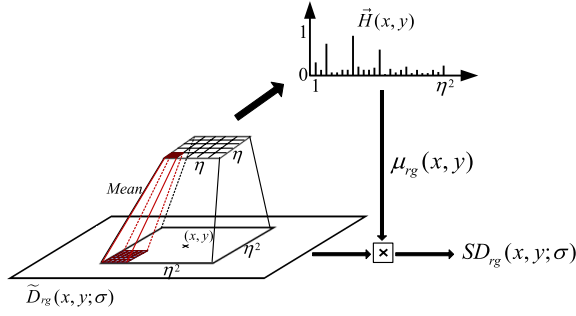
Fig. 5. The computation of texture suppression with spatial sparseness constraints (SSC).
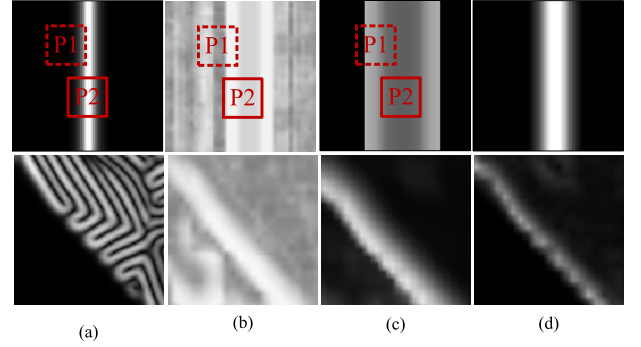

Fig. 6. Modified sparseness measure improves the accuracy of locating boundaries. (a) The responses of double-opponent cells contain boundaries (top) and textured edges (bottom). (b) The original sparseness measure (in (13)). (c) The modified sparseness measure without λ-correction (i.e., $\lambda = 1$ in (14)). (d) The λ-modified sparseness measure.

In order to narrow the unexpected spreading of high sparseness measure around the true boundaries, here we improve the computation of sparseness by introducing spatial information. Similar as Alpert *et al*. [64], we assume that the textured regions are far from sparse, with a low probability of containing wanted boundaries, considering the observation that the textured regions are often characterized by strong responses to edges at various orientations and scales [44]. Our strategy is as follows.

For each location (x, y) of the boundary response map obtained with (12), we compute the spatial sparseness measure of boundary response based on the information within a local window centered at (x, y). Fig. 5 illustrates the steps of sparseness computation. In this work, we define a square window with a size of $\eta^2 \times \eta^2$ pixels for the computation of spatial sparseness at each location. This window is further combined into $\eta \times \eta$ sub-patches with the size of $\eta \times \eta$ pixels, and the mean response magnitude is obtained for each sub-patch. Finally, the totally η^2 mean magnitude values are arranged into a one-dimensional vector denoted by $\vec{H}(x, y)$. Then, $\mu_{rg}(x, y)$, the new sparseness measure of location (x, y), is obtained by

$$\mu_{rg}(x, y) = \lambda(x, y) \cdot sparseness(x, y; \vec{H}) \qquad (14)$$

$$\lambda(x, y) = min\left(1, \frac{\widetilde{D}_{rg}(x, y; \sigma)}{mean\{\vec{H}(x, y)\}}\right) \qquad (15)$$

where $mean\{\vec{H}(x, y)\}$ denotes the mean value of the elements of $\vec{H}(x, y)$, which is equivalent to the mean response within the local window. Hence, $\lambda(x, y)$ is a factor that acts to weight the sparseness measure of certain location based on the distribution of neuronal responses within its local region. The operator $min(\cdot)$ makes $\lambda(x, y) \le 1$, indicating that $\lambda(x, y)$ acts only to reduce the sparseness measure if necessary.

Fig. 6 displays examples illustrating how the modified sparseness measure given by (14) works. Fig. 6(a) shows the edge responses, including boundaries (top row) and textured edges (bottom row). Fig. 6(b) shows that the original sparseness measure produced by (13) is incapable of distinguishing boundaries from textures. Fig. 6(c) indicates that our strategy without λ-correction (given by (14) with $\lambda = 1$) improves the accuracy of sparseness measure, but still results in the unexpected spreading of high sparseness measure from the true boundaries to its unwanted neighboring textures. From the images in the top row, the points (e.g., $P1$) around the

real boundaries (e.g., $P2$) exhibit incorrect high sparseness, because some unexpected locations with high responses are involved in the window (used for sparseness computation) of $P1$, which noise the histogram and hence the ℓ_1 and ℓ_2 norms in (13). Conversely, the final λ-modified sparseness measure defined in (14) can easily weaken the incorrect evaluation of sparseness, because the response of $P1$ is much weaker than the mean response of the local window, which results in a quite smaller $\lambda(x, y)$. In addition, boundary point $P2$ always maintains high sparseness (Fig. 6(d)).

Finally, the textured edges can be suppressed by multiplying the response of DO cells with the modified sparseness measure, which is written as

$$SD_{rg}(x, y; \sigma) = \mu_{rg}(x, y) \cdot \widetilde{D}_{rg}(x, y; \sigma) \qquad (16)$$

C. Obtain the Final Boundary Response

For the convenience of computation, we set one of w_1 and w_2 as 1, and another one as $w \in [-1, 0]$ to meet (4). The neuronal responses in four SO channels, i.e., $S_{rg} = C_r + wC_g$, $S_{gr} = C_g + wC_r$, $S_{by} = C_b + wC_y$, and $S_{yb} = C_y + wC_b$, are first computed with (3), and then, the responses of DO cells with texture suppression, i.e., SD_{rg}, SD_{gr}, SD_{by}, and SD_{yb}, are computed according to (16) for four DO channels. The final output $O(x, y; \sigma)$ is obtained by taking the max response over four DO channels at (x, y) according to

$$O(x, y; \sigma) = max\{SD_c(x, y; \sigma) | c \in \{rg, gr, by, yb\}\} \quad (17)$$

The full computational flowchart is summarized in Fig. 7. Note that for convenience of description, in the following we will use CO (Color-Opponency) to denote our model without SSC-based texture suppression (i.e., $\mu_c(x, y) = 1$, $c \in \{rg, gr, by, yb\}$ in (16)), and use SCO to denote the final model with SSC-based texture suppression.

III. EXPERIMENTS

A. Basic Properties of the Proposed Approach

To begin with, we evaluated the effect of the cone weights (i.e., w in Fig. 7) on the performance of the proposed model.
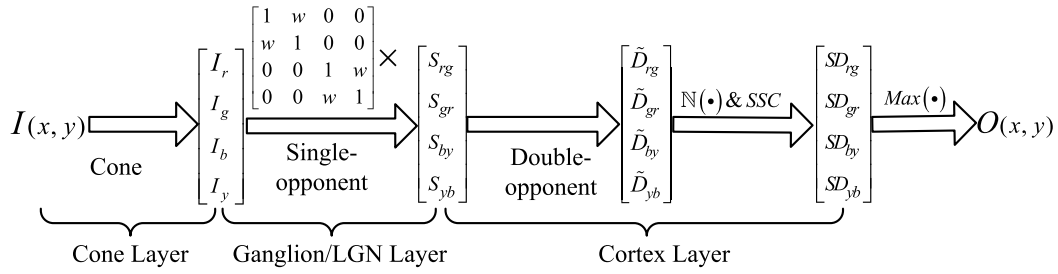
Fig. 7. The full steps of the proposed system for boundary detection. $\mathbb{N}(\cdot)$ denotes the linear normalization.
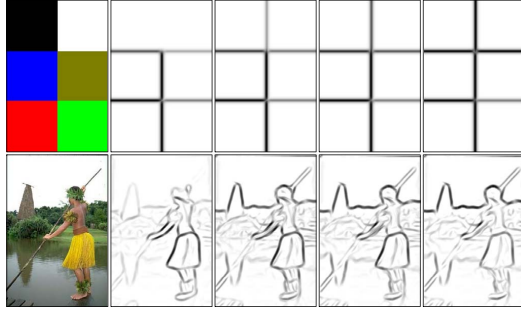

Fig. 8. The responses of CO to color and achromatic boundaries with various cone weights. **From left to right**: Original images and the responses of CO with different cone weights (w): -1.0, -0.6, -0.4 and 0.0. The four color blocks in the image of the top row are of equi-luminance.

Fig. 9. The capacity of CO to reconstruct broken contours while suppressing noise. **From left to right**: synthetic images and the responses of our model with different filter scales (σ) in the Ganglion/LGN Layer. The top and bottom rows are the images added with low- and high-level random noises.

We first analyzed the basic properties of CO (without texture suppression). Fig. 8 illustrates the responses of CO to pure color and achromatic boundaries with various w. It is clear that when $w = -1.0$ (i.e., $|w_1| = |w_2|$ in (3)), our CO model can only respond to the color-defined edges (e.g., a pure red-to-green edge). The reason that CO with balanced cone weights is blind to luminance-defined boundaries could be explained as follows. The double-opponent RF shown in Fig. 2(d) is mathematically composed of two side-by-side single-opponent RFs. With balanced cone weights, single-opponent RF is blind to achromatic patterns with any brightness, since equal cone inputs with opposite signs cancel each other out (as analyzed in Section II), which eventually makes the double-opponent RF irresponsive to any luminance-defined edge.

When w varies from -1.0 to 0.0, besides the obvious responses to color-defined edges, the responses to luminance-defined boundaries increase, since neither the achromatic nor chromatic cone inputs can completely cancel each other out when the double-opponent RF covers the edge. Taken together, when $w \neq -1.0$ (i.e., $|w_1| \neq |w_2|$), our model has the exciting ability of jointly extracting the color- and luminance-defined boundaries with an opponent way. For the pure color and brightness boundaries (top row in Fig. 8), the model shows equivalent responses to the two types of boundaries when $w = 0$. In contrast, both the color- and luminance-defined boundaries in the natural image (bottom row in Fig. 8) are well responded when $w = -0.6$. This is perhaps because that in natural images, the absolute values of color difference are normally weaker than that of the brightness difference [66].

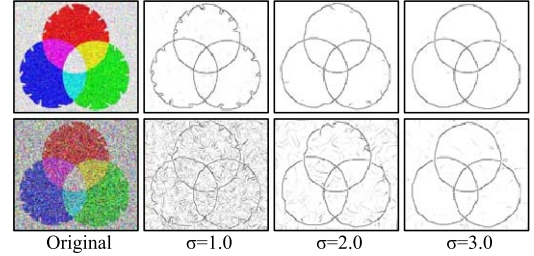We also designed synthetic images with broken circles at different noise levels to verify the capacity of our CO to extract smooth contours in noised images. Fig. 9 clearly shows that our model has wonderful ability of suppressing noises while reconstructing the broken portions of the contours. When the scale σ is increased to proper values, the smooth circles can pop-out clearly from the noised background.

B. Results on BSDS

We further evaluated the performance of our model using a publicly available dataset, i.e., Berkeley Segmentation Data Set (BSDS300 and BSDS500) [2], [67]. Each image in the dataset has multiple human-labeled segmentations as the ground truth data. For the convenience of comparison, we represented the pixel values of the output image (i.e., the boundaries) of all detectors as the probability of being true boundaries after an operation of non-maxima suppression [6], [13], [35]. We also computed the so-called *F-measure* [13]

$$F = 2PR/(P + R) \tag{18}$$

where P and R represent respectively the precision and recall, which have been widely used for evaluating the performance of edge detectors [13]. In short, precision reflects the probability that the detected edge is valid, and recall reflects the probability that the ground truth edge was detected.

Our model includes three parameters, i.e., the scale of Gaussian (σ), cone weight (w), and the size of window used in SSC (η). In order to ensure the integrity of evaluation, we only employed the training set (200 images) of BSDS300 dataset for parameter setting during the optimization phase, and then benchmarked our method on the test set (100 images). Our SCO achieves the optimal performance (F=0.68 on training set of BSDS300) with $\sigma = 1.1$, $w = -0.7$, and $\eta = 5$.
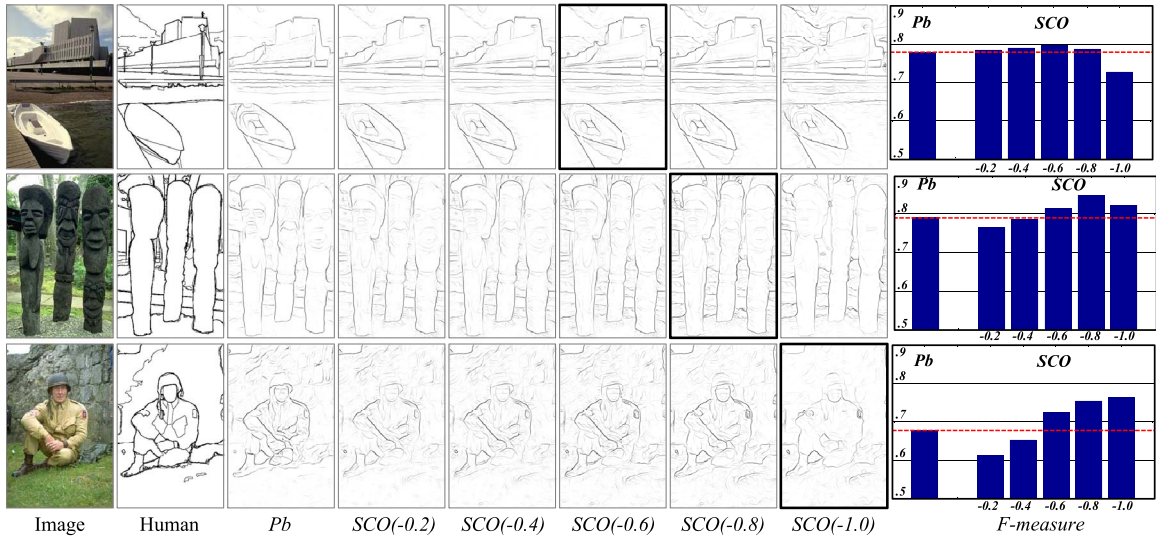
Fig. 10. Comparisons of $SCO(w)$ (with $\sigma = 1.1$, $\eta = 5$ and different cone weights w) with Pb detector [13]. The last column presents the *F-measure* of each boundary map listed in the third to eighth columns. The optimal results (marked by black bold rectangle) correspond to the maximum of *F-measure*.

In the following, we will fix this setting for different scenes when comparing with other models. In addition, we will use $CO(w)$ and $SCO(w)$ to respectively denote the model of CO and SCO with certain w value for the convenience of clear description. Meanwhile, other parameters are always fixed as $\sigma = 1.1$ and $\eta = 5$.

Fig. 10 shows some examples comparing SCO to the Pb detector proposed by Martin *et al.* [13]. The boundaries of each image with cone weight w varying from -0.2 to -1.0 are presented in Fig. 10 (the forth to the eighth column), and the optimal results are marked with bold rectangles. Note that $SCO(-1.0)$ corresponds to the response of double-opponent cells with balanced cone weights to the pure color edges. It is clear from Fig. 10 that our model exhibits more powerful ability of extracting the structured object boundaries with low color and brightness contrast. The last column in Fig. 10 shows the quantitative comparison of our SCO with Pb. The maximum *F-measure*, which was computed from the probability of boundary map at a particular threshold, captures the trade-off between precision and recall [13]. The red dotted horizontal line indicates the optimal performance of Pb. It is clear that compared to Pb, our SCO model provides higher *F-measure* at most w levels; particularly, the maximum of *F-measure* across various w (i.e., the optimal detection result) of our model is notably higher than that of Pb, at least for these specific images in Fig. 10. Such property of our SCO provides us a wide space to introduce new ways to improve our model in the future work by adaptively selecting optimal w for individual images, considering another remarkable advantage of SCO in saving computational cost (see details later).

To check whether our model can jointly extract sufficient color and luminance defined edges, we tried to add a separate channel of color or luminance to $SCO(-0.7)$ as follows. When adding the separate channel of luminance, we averaged the R, G, and B as the luminance channel, and then, we got the luminance defined edges according to the similar steps given by (8) and (12). These luminance edges were

combined with the edges given by $SCO(-0.7)$ using the max operation. We added the separate channel of color by using the max operation on the outputs of $SCO(-0.7)$ and $SCO(-1.0)$, where $SCO(-1.0)$ is the response to pure color edges. The results shown in Fig. 11(left) clearly indicate that there is no obvious improvement contributed by additional color or luminance channels, which indicates that by selecting an appropriate w value, our SCO model responds almost all color and luminance defined edges, though the two types of edges were not separately and explicitly processed.

Fig. 11(right) shows that extracting boundary in DO space produces better performance than that directly in RGB or CIELAB color spaces. In RGB space, color information is represented by the relative proportion among each channel, so individual RGB channels have not the ability of fully representing the color contrast (color boundaries). In fact, the boundaries detected in RGB space are equivalent to the results of our CO model with $w = 0$ (mainly luminance boundaries). In surprise, we find that CIELAB based detection also produces good results (F=0.63), only slightly worse than our $CO(-0.7)$ (without SSC, F=0.64). This is mainly because that CIELAB is also a kind of color-opponent space with dimension L^* for lightness (luminance) and a^* and b^* for the color-opponent dimensions. However, compared to our model with unbalanced cone weights, CIELAB is of less flexibility in selectively detecting color and luminance defined boundaries. Note that the boundaries in RGB or CIELAB space were detected in the way similar to (8)$\sim$(12), i.e., the edges were first obtained by oriented filtering at individual channels, and then combined with a max operator.

We also evaluated the influence of SSC. Fig. 12(left) shows that the overall performance is improved when modified sparseness measure in (14) is used in SSC. That is, our both versions of modified SSC outperform the original sparseness measure in (13), which has negligible contribution to performance (F=0.63); In addition, with λ-correction in modified SSC, the F-measure was further increased from
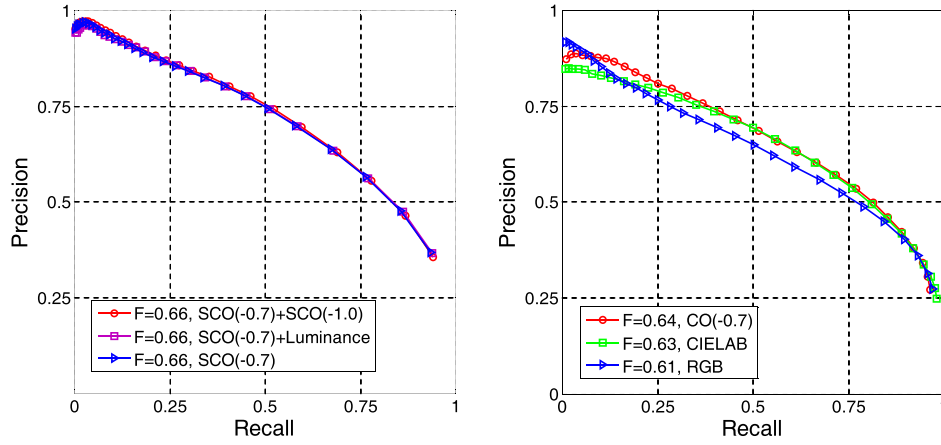
Fig. 11.   **Left**: Performance comparison after adding a separate luminance ((R+G+B)/3) or pure color ($SCO(-1.0)$) channel to $SCO(-0.7)$; **Right**: Comparison of the performance in double-opponent space with that in RGB and CIELAB color spaces on BSDS300.
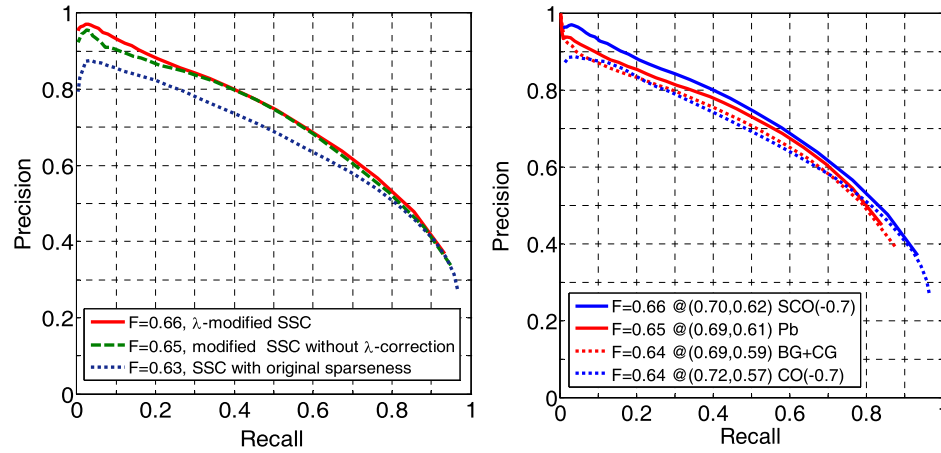

Fig. 12.   **Left**: *P-R curve* comparison of $SCO(-0.7)$ with modified and un-modified SSC versions on BSDS300. **Right**: Tested on BSDS300, the *P-R curves* show that SSC exhibits more efficient ability to suppress unwanted edges than Pb, which suppresses texture based on Texture Gradient (TG).

TABLE I

MEAN COMPUTATION TIME TAKEN TO COMPUTE PER BOUNDARY MAP AVERAGED OVER THE 100 TEST IMAGES IN BSDS300 [67] WITH MATLAB CODES. COMPUTER USED HERE IS INTEL CORE2, 2.8GHz WITH 2.0G RAM

Method	Pb	CO	SCO
Time(s)	49.00	5.37	6.75

0.65 to 0.66. Meanwhile, a clearly higher performance in the low recall and high precision regime was achieved by $SCO(-0.7)$ with modified SSC. Fig. 12(right) clearly shows that SCO performs excellent in suppressing textural edges. Compared to *Texture Gradient (TG)* introduced in Pb [13], which improves the performance of Pb from 0.64 ($BG+CG$) to 0.65 ($BG+TG+CG$), the SSC-based suppression improves the performance of our model from 0.64 (CO) to 0.66 (SCO). In addition, SSC is computational saving and without explicit training processing. The comparison of computational cost is listed in Table I, which indicates that our CO model took only one tenth of the computation time needed by Pb. Introducing SSC-based texture suppression into CO (i.e., SCO) improves

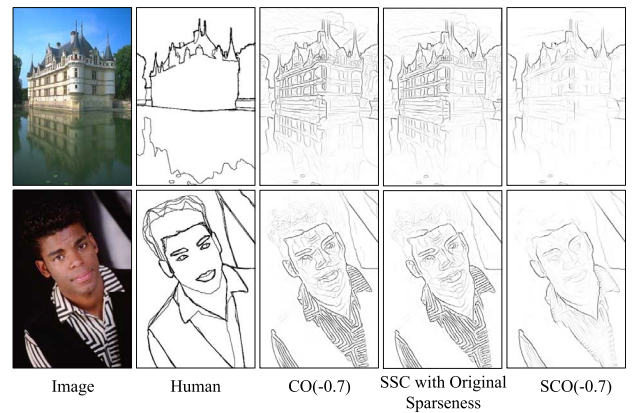Fig. 13.   Examples illustrating the role of texture suppression with SSC.

the performance still in a quite effective way, with a quite slight increase of computation time over CO.

Two examples in Fig. 13 shows that SSC efficiently suppresses the edges within textured regions. With SSC, the unwanted fine edges located in textured regions (e.g., the texture of the shirt in the bottom row) are largely weakened
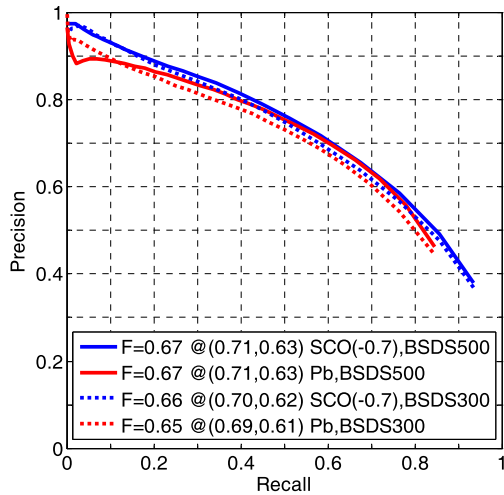
Fig. 14. The overall performance of *P-R curve* on BSDS300 and BSDS500.

TABLE II

COMPARISON WITH MORE COLOR BOUNDARY DETECTORS ON THE
TEST SET OF BSDS300/500 DATASET [2], [67]

Method on BSDS300	F-measure
SCO(-0.7)	**0.66**
CO(-0.7)	**0.64**
Pb (2004) [13]	0.65
CRF (2005) [24]	0.64
Min-Cover (2006) [28]	0.65
BEL (2006) [19]	0.66
Untangling-Cycles (2007) [23]	0.64
Sparse (2008) [21]	0.66
Multiscale-Pb (2008) [18]	0.68
gPb (2008) [2], [17]	0.70
SCG (2012) [22]	0.71
Gb(C) (2012) [40]	0.65
Gb(C+S) (2012) [40]	0.67

Method on BSDS500	F-measure
SCO(-0.7)	**0.67**
Pb (2004) [13]	0.67
gPb (2011) [2]	0.71
gPb-owt-ucm (2011) [2]	0.73
SCG (2012) [22]	0.74
Sketch-Tokens (2013) [41]	0.73
SE (2013) [42]	0.74
PMI (2014) [43]	0.74

since they normally have low spatial sparseness; in contrast, the wanted fine boundaries between two different regions could be largely reserved because they normally have high spatial sparseness. Note that if two textured regions have similar spatial sparseness measures, the fine boundaries between them would be also weakened to certain extent. To solve this problem, more efficient texture descriptors are expected.

In Fig. 14, we give *P-R curve* based comparison of our *SCO* with *Pb* on the datasets of both BSDS300 and BSDS500. The figure shows that the proposed model (*SCO*) outperforms the *Pb*, when employing the same information (i.e., color, luminance, and texture). Especially, our model achieves higher precision at high thresholds (top end of curve) and higher recall at low thresholds (bottom end of curve). Fig. 15 presents qualitative comparisons on some images.

Table II lists more comparisons with other state-of-the-art methods on BSDS300/500. Some state-of-the-art

algorithms achieve higher *F-measure* than the proposed model; however, these methods always employ more information, such as multiple scales (*Multiscale Pb*, [18]), global information (*gPb*, [2], [17]), or high dimensional representation (*SCG*, [22]), etc. Furthermore, considering the fact that many models (e.g., *Untangling Cycles*, *Multiscale Pb*, *gPb*, etc.) presented in Table II are various extended versions of *Pb*, they generally take more computation time (not shown here) than *Pb*. For example, compared to the F-measure of 0.65 by *Pb*, The method of *multiscale Pb* obtains an F-measure of 0.68 by combining the output of *Pb* at three scales; the state-of-the-art global *Pb (gPb)* obtains an F-measure of 0.70 by using multiscale information and soft-segmentations from *Normalized Cuts (Ncuts)*. When introducing soft-segmentations by a modified *Ncuts*, the F-measure of *Gb* was increased from 0.65 to 0.67.

In addition, if a supervised learning based model is training data dependant, it sometimes obtains poor results on specific images. Fig. 16 shows two examples, which clearly illuminate that many color-defined equiluminant edges were not well responded by *Pb*; conversely, our method responded strongly to most of the boundaries, whether they are chromatic, achromatic, or mixed. Obviously, it is a quite challenging task to detect the boundaries when the foreground and background of an image are difficult to be distinguished in BOTH the brightness and color, and we admit that our model may miss such boundaries more or less, since we do not introduce effective high-order features (like texture) to distinguish such foreground and background.

Note that in the above, we paid more attention to the comparison of our model with *Pb* due to the fact that they use the same information, i.e., luminance, color, and texture. Thought *Multiscale Pb* [18] and *gPb* [2], [17] provide better F-measure, they are substantial extensions of *Pb* by further introducing multiple scales and global information.

It should be pointed out that we also tried to extend our model into multi-scale space and combined the boundaries with a simple strategy of max operation across scales. However, from the results (not shown here for space considerations) we did not see the clear performance improvements. Considering the similar observation with the *Pb* model proposed in [13], we believe that multi-scale processing may significantly benefit the boundary detection in natural images only with appropriate choice of scales and combination rules [2], [18].

Another point worthy of noticing is that our model provides higher resolution to the boundaries that are very close to each other, which is illustrated by several typical examples with zoomed-in local regions shown in Fig. 17. As pointed out by Isola *et al.* [43], the standard metrics can not distinguish whether or not a detector is capturing detailed edges above a certain spatial frequency, since detected edges that are within several pixels away from a ground truth edge is always considered a correct detection by the benchmark metrics. The ability of our model to precisely localize the high-frequency boundaries comes from the use of simple and very local features (within the range of receptive field) to detect highly detailed salient edges and the use of local statistical
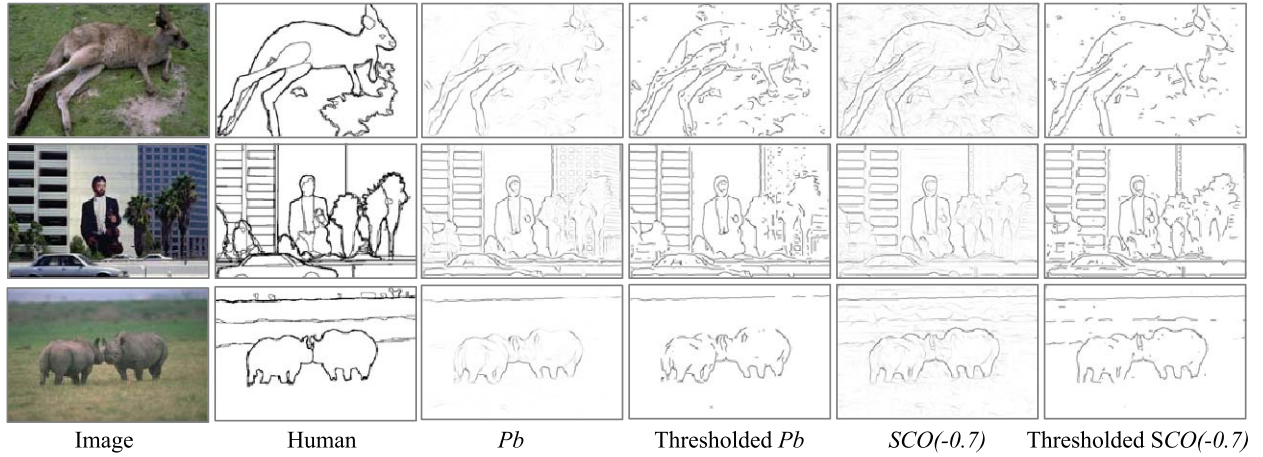
| Image | Human | *Pb* | Thresholded *Pb* | *SCO(-0.7)* | Thresholded S*CO(-0.7)* |

Fig. 15.   More examples compared with *Pb*. The thresholds used here correspond to the maximal *F-measure* for each image.
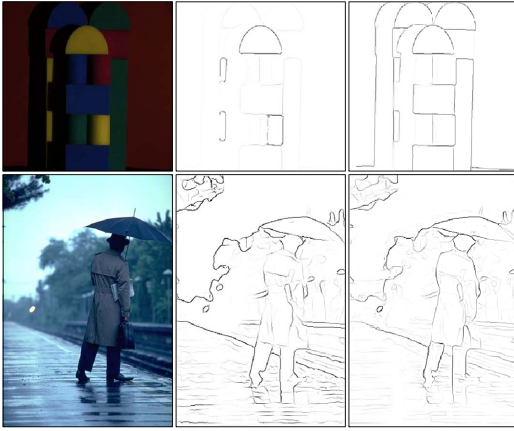

Fig. 16.   Examples on images with color-defined equiluminant edges. **From left to right**: original images, results of *Pb*, and results of *SCO*(−0.7).
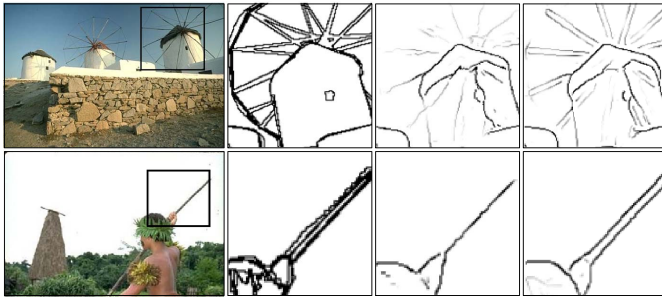

Fig. 17.   Zoomed in view of several test images. **From left to right**: original images with local regions of interest (black rectangle), human-marked boundaries, boundaries detected by *Pb* and our *SCO*(−0.7). Note that our model can preserve the high frequency contours.

regularities (i.e., sparseness constraint) to suppress textural regions.

IV. Discussion

In this paper we have proposed a novel way for the challenging task of detecting salient boundaries in complex color scenes, inspired by the information processing mechanisms emerging in the early visual stages. In specific, the SO ganglion cells function to enhance region information, and the oriented DO cells in V1 serve to detect the boundaries among regions. The results on synthetic images and large image datasets (BSDS300 and BSDS500) with human-labeled boundaries in natural images revealed the competitive performance of the proposed model with a simple framework.

The main novelty of the work is summarized as follows. (a) Our new boundary detection system is based on the double-opponent (DO) mechanism and has the amazing property of jointly extracting color- and luminance-defined edges, which is really different from the two-step way of some existing models that explicitly extract the color and luminance edges in separate channels and then combine them, e.g., with a supervised learning. (b) A new strategy of spatial sparseness constraint (SSC) was introduced to weight the edge responses of the proposed CO system, which provides a simple while efficient way for texture suppression. The main merits of the proposed *SCO* model include: (a) competitive performance for edge detection and texture suppression with only low-level local information; (b) flexibility in responding to color- and luminance-defined boundaries; (c) as few as only one free parameter (i.e., cone weight w) (the model is almost insensitive to another parameter, Gaussian scale σ); (d) quite low computational cost. Note that though our model also needs a fine tuning of the free parameter w, quite acceptable performance was always achieved when w is not close to -1.0. This is different from some of the current state-of-the-art methods (e.g., *Pb* and its several variants [2], [13], [17], [18]), which not only need supervised learning on specific training dataset, but also require to first optimize each involved parameter with respect to the ground truth dataset.

In recent years, bio-inspired methods have been proposed for various computer vision applications [53], [56], [63], [68], [69]. This is an interesting trend in the progress of computer vision. For the task of boundary detection studied in this paper, though recent research suggested that some state-of-the-art contour detection algorithms seem to perform as well as humans using low-level visual cues [70], there are still a lot of unclear issues about how the real visual system percept contours from complex scenes far efficiently

than computer vision. For example, how to integrate color and luminance information in V1 or higher level areas of the visual system is less known. In addition, combination of various visual cues (color, orientation, etc.) and the contextual influence between neurons have also aroused widely concern [38], [71]. In this paper we only tried to model the boundary-related subcortical and V1 mechanisms, and obviously, the proposed model is not a full V1 model, but a bottom-up V1 model for the specific task of boundary detection.

A quite recent physiological study suggested that the mixing of color and luminance signals by certain cells in V1 was optimized to process the diversity of edges existing in natural scenes [72]. Clearly, our results in this work support this physiological suggestion. It should be pointed out that the early stages of the human visual system are much more complicated than the simplistic system introduced in this work. For example, the non-linear color-luminance interactions have been found between the blue-yellow and luminance channels [73], which indicates that important aspects of the cortical representation of color information cannot be well described by classical linear analysis [30]. In addition, Feedback controls from higher cortical areas exert important influence on the V1 information processing, including contour detection [71]. All these complicated mechanisms in the biological visual systems suggest us future directions to improve our current model.

In fact, the more challenging task is to find salient object-level boundaries in complex natural scenes, which are considered as quite important elements for visual perception [74]. To this end, some mid-level perceptual grouping rules (e.g., the Gestalt laws of "good continuity", "closure", etc.) are thought to play critical role [2]. From the perspective of computer vision, some existing models like *Untangling Cycles* [23], *gPb* [2], [17], *Gb* [40], *Sketch Tokens* [41], *PMI* [43] etc., are excellent examples introducing the idea of mid-level perceptual grouping for contour detection. Using Gestalt rules to guide boundary detection is also one of our future directions. Though suitable supervised learning may, implicitly, find some mid-level rules like Gestalt (from a group of images), we hope to explicitly accomplish it (on individual images) by simulating the biological properties like the collinear facilitation in V1 [71] and the border-ownership selectivity produced by the dynamic interactions between V1, V2, and V4 [75].

To summarize, the new model proposed in this work offers the novel property of responding well to diversity of edges including both color- and luminance-defined ones. With competitive performance in comparison to the state-of-the-art approaches, we hope that the proposed double-opponent based detector will soon become a standard component of boundary detection approaches for color scenes.

REFERENCES

[1] D. B. Walther, B. Chai, E. Caddigan, D. M. Beck, and L. Fei-Fei, "Simple line drawings suffice for functional MRI decoding of natural scene categories," *Proc. Nat. Acad. Sci.*, vol. 108, no. 23, pp. 9661–9666, 2011.

[2] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[3] G. Papari and N. Petkov, "Edge and line oriented contour detection: State of the art," *Image Vis. Comput.*, vol. 29, nos. 2–3, pp. 79–103, 2011.

[4] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, 1980.

[5] P. Kovesi, "Image features from phase congruency," *Videre, J. Comput. Vis. Res.*, vol. 1, no. 3, pp. 1–26, 1999.

[6] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[7] X. Yang, X. Gao, D. Tao, and X. Li, "Improving level set method for fast auroral oval segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2854–2865, Jul. 2014.

[8] X. Yang, X. Gao, J. Li, and B. Han, "A shape-initialized and intensity-adaptive level set method for auroral oval segmentation," *Inf. Sci.*, vol. 277, pp. 794–807, Sep. 2014.

[9] X. Yang, X. Gao, D. Tao, X. Li, and J. Li, "An efficient MRF embedded level set method for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 9–21, Jan. 2015.

[10] P. Perona and J. Malik, "Detecting and localizing edges composed of steps, peaks and roofs," in *Proc. 3rd ICCV*, Dec. 1990, pp. 52–57.

[11] M. C. Morrone and R. A. Owens, "Feature detection from local energy," *Pattern Recognit. Lett.*, vol. 6, no. 5, pp. 303–313, 1987.

[12] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.

[13] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.

[14] S. K. Shevell and F. A. Kingdom, "Color in complex scenes," *Annu. Rev. Psychol.*, vol. 59, pp. 143–166, Jan. 2008.

[15] M. A. Ruzon and C. Tomasi, "Edge, junction, and corner detection using color distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1281–1295, Nov. 2001.

[16] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1454–1466, Oct. 2001.

[17] M. Maire, P. Arbeláez, C. Fowlkes, and J. Malik, "Using contours to detect and localize junctions in natural images," in *Proc. IEEE CVPR*, Jun. 2008, pp. 1–8.

[18] X. Ren, "Multi-scale improves boundary detection in natural images," in *Proc. ECCV*, 2008, pp. 533–545.

[19] P. Dollar, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," in *Proc. IEEE CVPR*, vol. 2. Jun. 2006, pp. 1964–1971.

[20] I. Kokkinos, "Highly accurate boundary detection and grouping," in *Proc. IEEE CVPR*, Jun. 2010, pp. 2520–2527.

[21] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce, "Discriminative sparse image models for class-specific edge detection and image interpretation," in *Proc. ECCV*, 2008, pp. 43–56.

[22] X. Ren and L. Bo, "Discriminatively trained sparse code gradients for contour detection," in *Proc. NIPS*, 2012, pp. 593–601.

[23] Q. Zhu, G. Song, and J. Shi, "Untangling cycles for contour grouping," in *Proc. IEEE 11th ICCV*, Oct. 2007, pp. 1–8.

[24] X. Ren, C. C. Fowlkes, and J. Malik, "Scale-invariant contour completion using conditional random fields," in *Proc. 10th IEEE ICCV*, vol. 2. Oct. 2005, pp. 1214–1221.

[25] R. Kennedy, J. Gallier, and J. Shi, "Contour cut: Identifying salient contours in images by solving a Hermitian eigenvalue problem," in *Proc. IEEE CVPR*, Jun. 2011, pp. 2065–2072.

[26] S. Mahamud, L. R. Williams, K. K. Thornber, and K. Xu, "Segmentation of multiple salient closed contours from real images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 4, pp. 433–444, Apr. 2003.

[27] Y. Ming, H. Li, and X. He, "Connected contours: A new contour completion model that respects the closure effect," in *Proc. IEEE CVPR*, Jun. 2012, pp. 829–836.

[28] P. Felzenszwalb and D. McAllester, "A min-cover approach for finding salient curves," in *Proc. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 185.

[29] P. Arbelaez, "Boundary extraction in natural images using ultra-metric contour maps," in *Proc. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 182.

[30] B. R. Conway *et al.*, "Advances in color science: From retina to behavior," *J. Neurosci.*, vol. 30, no. 45, pp. 14955–14963, 2010.

[31] S. G. Solomon and P. Lennie, "The machinery of colour vision," *Nature Rev. Neurosci.*, vol. 8, no. 4, pp. 276–286, 2007.

[32] A. Stockman and D. H. Brainard, "Color vision mechanisms," in *The Optical Society of America Handbook of Optics* (Vision and Vision Optics), vol. 3, M. Bass *et al.*, Eds., 3rd ed. New York, NY, USA: McGraw-Hill, 2010, pp. 11.11–11.104.

[33] J. Malik, S. Belongie, J. Shi, and T. Leung, "Textons, contours and regions: Cue integration in image segmentation," in *Proc. ICCV*, vol. 2. 1999, pp. 918–925.

[34] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, 2001.

[35] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 729–739, Jul. 2003.

[36] Q. Tang, N. Sang, and T. Zhang, "Extraction of salient contours from cluttered scenes," *Pattern Recognit.*, vol. 40, no. 11, pp. 3100–3109, 2007.

[37] C. Zeng, Y. Li, and C. Li, "Center–surround interaction with adaptive inhibition: A computational model for contour detection," *NeuroImage*, vol. 55, no. 1, pp. 49–66, 2011.

[38] K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Multifeature-based surround inhibition improves contour detection in natural images," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5020–5032, Dec. 2014.

[39] R. Hidayat and R. D. Green, "Real-time texture boundary detection from ridges in the standard deviation space," in *Proc. BMVC*, 2009, pp. 1–10.

[40] M. Leordeanu, R. Sukthankar, and C. Sminchisescu, "Efficient closed-form solution to generalized boundary detection," in *Proc. ECCV*, 2012, pp. 516–529.

[41] J. J. Lim, C. L. Zitnick, and P. Dollár, "Sketch tokens: A learned mid-level representation for contour and object detection," in *Proc. IEEE CVPR*, Jun. 2013, pp. 3158–3165.

[42] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE ICCV*, Dec. 2013, pp. 1841–1848.

[43] P. Isola, D. Zoran, D. Krishnan, and E. H. Adelson, "Crisp boundary detection using pointwise mutual information," in *Proc. ECCV*, 2014, pp. 799–814.

[44] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Opt. Soc. Amer. A*, vol. 7, no. 5, pp. 923–932, 1990.

[45] N. Petkov and P. Kruizinga, "Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells," *Biol. Cybern.*, vol. 76, no. 2, pp. 83–96, 1997.

[46] G. Papari, P. Campisi, N. Petkov, and A. Neri, "A biologically motivated multiresolution approach to contour detection," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, 2007, Art. ID 071828.

[47] N. Petkov and M. A. Westenberg, "Suppression of contour perception by band-limited noise and its relation to nonclassical receptive field inhibition," *Biol. Cybern.*, vol. 88, no. 3, pp. 236–246, 2003.

[48] C. Zhou and B. W. Mel, "Cue combination and color edge detection in natural scenes," *J. Vis.*, vol. 8, no. 4, pp. 1–25, Apr. 2008.

[49] J. Zhang, Y. Barhomi, and T. Serre, "A new biologically inspired color image descriptor," in *Proc. ECCV*, 2012, pp. 312–324.

[50] K. R. Gegenfurtner, "Cortical mechanisms of colour vision," *Nature Rev. Neurosci.*, vol. 4, no. 7, pp. 563–572, 2003.

[51] R. Shapley and M. J. Hawken, "Color in the cortex: Single- and double-opponent cells," *Vis. Res.*, vol. 51, no. 7, pp. 701–717, 2011.

[52] E. N. Johnson, M. J. Hawken, and R. Shapley, "The orientation selectivity of color-responsive neurons in macaque V1," *J. Neurosci.*, vol. 28, no. 32, pp. 8096–8106, 2008.

[53] S. Gao, K. Yang, C. Li, and Y. Li, "A color constancy model with double-opponency mechanisms," in *Proc. IEEE ICCV*, Dec. 2013, pp. 929–936.

[54] S. Gao, K. Yang, C. Li, and Y. Li, "Color constancy using double-opponency," *IEEE Trans. Pattern Anal. Mach. Intell.*, Jan. 2015.

[55] E. N. Johnson, M. J. Hawken, and R. Shapley, "The spatial transformation of color in the primary visual cortex of the macaque monkey," *Nature Neurosci.*, vol. 4, no. 4, pp. 409–416, 2001.

[56] K. Yang, S. Gao, C. Li, and Y. Li, "Efficient color boundary detection with color-opponent mechanisms," in *Proc. IEEE CVPR*, Jun. 2013, pp. 2810–2817.

[57] T. N. Wiesel and D. H. Hubel, "Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey," *J. Neurophysiol.*, vol. 29, no. 6, pp. 1115–1156, 1966.

[58] P. Lennie and M. D'Zmura, "Mechanisms of color vision," *Critical Rev. Neurobiol.*, vol. 3, no. 4, pp. 333–400, 1987.

[59] A. Angelucci and K. Sainsbury, "Contribution of feedforward thalamic afferents and corticogeniculate feedback to the spatial summation area of macaque V1 and LGN," *J. Comparative Neurol.*, vol. 498, no. 3, pp. 330–351, 2006.

[60] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *Nature Rev. Neurosci.*, vol. 13, no. 1, pp. 51–62, 2011.

[61] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[62] M. Zhu and C. J. Rozell, "Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system," *PLoS Comput. Biol.*, vol. 9, no. 8, p. e1003191, 2013.

[63] M. W. Spratling, "Image segmentation using a sparse coding model of cortical area V1," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1631–1643, Apr. 2013.

[64] S. Alpert, M. Galun, A. Brandt, and R. Basri, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 315–327, Feb. 2012.

[65] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.

[66] T. Hansen and K. R. Gegenfurtner, "Independence of color and luminance edges in natural scenes," *Vis. Neurosci.*, vol. 26, no. 1, pp. 35–49, 2009.

[67] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE ICCV*, vol. 2. 2001, pp. 416–423.

[68] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[69] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.

[70] C. L. Zitnick and D. Parikh, "The role of image understanding in contour detection," in *Proc. IEEE CVPR*, Jun. 2012, pp. 622–629.

[71] G. Loffler, "Perception of contours and shapes: Low and intermediate stage mechanisms," *Vis. Res.*, vol. 48, no. 20, pp. 2106–2127, 2008.

[72] X. Li *et al.*, "Mixing of chromatic and luminance retinal signals in primate area V1," *Cerebral Cortex*, 2014.

[73] G. D. Horwitz, E. J. Chichilnisky, and T. D. Albright, "Blue-yellow signals are enhanced by spatiotemporal luminance contrast in macaque V1," *J. Neurophysiol.*, vol. 93, no. 4, pp. 2263–2278, 2005.

[74] X. Hou, A. Yuille, and C. Koch, "Boundary detection benchmarking: Beyond F-measures," in *Proc. IEEE CVPR*, Jun. 2013, pp. 2123–2130.

[75] O. W. Layton, E. Mingolla, and A. Yazdanbakhsh, "Dynamic coding of border-ownership in visual cortex," *J. Vis.*, vol. 12, no. 13, pp. 1–21, 2012.

Kai-Fu Yang received the B.Sc. and M.Sc. degrees in biomedical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2009 and 2012, respectively, where he is currently pursuing the Ph.D. degree. His research interests include visual mechanism modeling and image processing.
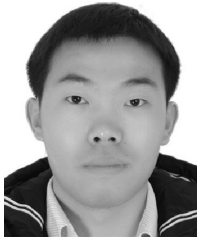
Shao-Bing Gao received the M.Sc. degree in biomedical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2013, where he is currently pursuing the Ph.D. degree in biomedical engineering. His research interests include visual mechanism modeling and image processing.

Chao-Yi Li received the degrees from Chinese Medical University, Shenyang, China, in 1956, and Fudan University, Shanghai, China, in 1961. He became an Academician with the Chinese Academy of Sciences, Beijing, China, in 1999. He is currently a Professor with the University of Electronic Science and Technology of China, Chengdu, China, and the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. He is also the Honorary Director of the Key Laboratory for Neuroinformation of Ministry of Education, China. His research interests include visual neurophysiology and brain-computer interface.

Ce-Feng Guo received the B.Sc. degree from the the Southwest University of Science and Technology, Mianyang, in 2013. He is currently pursuing the M.Sc. degree in biomedical engineering with the University of Electronic Science and Technology of China. His research interests include visual mechanism modeling and image processing.

Yong-Jie Li (M'14) received the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2004. He is currently a Professor with the Key Laboratory for Neuroinformation of Ministry of Education, School of Life Science and Technology, UESTC. His research interests include visual mechanism modeling, image processing, and intelligent computation.