This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

Color Constancy Using Double-Opponency

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

Abstract—The double-opponent (DO) color-sensitive cells in the primary visual cortex (V1) of the human visual system (HVS) have long been recognized as the physiological basis of color constancy. In this work we propose a new color constancy model by imitating the functional properties of the HVS from the single-opponent (SO) cells in the retina to the DO cells in V1 and the possible neurons in the higher visual cortexes. The idea behind the proposed double-opponency based color constancy (DOCC) model originates from the substantial observation that the color distribution of the responses of DO cells to the color-biased images coincides well with the vector denoting the light source color. Then the illuminant color is easily estimated by pooling the responses of DO cells in separate channels in LMS space with the pooling mechanism of sum or max. Extensive evaluations on three commonly used datasets, including the test with the dataset dependent optimal parameters, as well as the intra- and inter-dataset cross validation, show that our physiologically inspired DOCC model can produce quite competitive results in comparison to the state-of-the-art approaches, but with a relative simple implementation and without requiring fine-tuning of the method for each different dataset.

Index Terms—color constancy, double opponency, human visual system, pooling mechanism

1 Introduction

COLOR constancy is one of the amazing abilities of perceptual constancy of the human visual system (HVS), which enables the perceived color of objects largely constant as the light source color changes [1]. In contrast, captured with regular digital cameras or videos, the physical color of scenes may be shifted by the varying external illuminant. As an example, Fig. 1(a) and (b) show the obvious shift of color distribution between the color-biased and canonical images of the same scene. One of the fundamental requirements in computer vision, especially for the robust color-based systems (e.g., color-based object recognition and tracking), is to extract reliable color cues that are invariant to the changes in external lighting. A common solution to such challenging task is to first estimate the scene illuminant, which is then used to correct the color-biased images to get the so-called canonical images under a white light source [2].

Based on the two steps mentioned above, a large number of elegant solutions have been proposed for computational color constancy, from the simple max or mean based algorithms to the sophisticated statistical and machine learning based algorithms. According to the underlying biological plausibility, the existing methods can be divided into two distinct groups: biologically inspired models [3], [4], [5], [6], [7], [8], [9] and non-biologically inspired models [10], [11], [12], [13], [14], [15], [16]; more examples before the year of 2011, see [2], [17], [18], [19] for good reviews.

1.1 Non-biologically based models

There is a large population of non-biologically based models that attempt to estimate the scene illuminant using the image statistics of low-level [5], [20], median- or high-level features [10], [11], [21]. For example, the well-known grey-world theory [20] estimates the illuminant components simply by computing the mean in each color channel (i.e., R, G, B) of the input image, based on the assumption that the average reflectance of a scene is normally achromatic. Some other grey-world based methods include white patch or max-RGB [5], shade of grey [10], grey-edge [11], etc. In general, with fixed (or static [2]) parameter setting for a given dataset, these methods have the advantage of simple implementation and fast computation. However, the diverse reflectance distribution of scenes does not always perfectly satisfy the grey-world assumption [2], [17].

In order to make more practical the simple assumption about the reflectance distribution made by grey-world based models, two types of promising solutions have been proposed: (1) more complex statistics and priori information were introduced about the surface reflectance and more sophisticated statistical computation was employed to estimate the illuminant. Representative examples include gamut mapping and its various extensions [12], [22], [23], [24], [25], Bayesian approaches [26], [27], spatio-spectral statistics based

- *S.-B. Gao and K.-F. Yang are with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (email:{gao_shaobing,yang_kf}@163.com)*
- *C.-Y. Li is School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China, and the Center for Life Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China (email:cyli@sibs.ac.cn)*
- *Y.-J. Li is with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (corresponding author, phone: +86-28-83208238; fax: +86-28-83208238; email:liyj@uestc.edu.cn)*

Manuscript received August 6, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

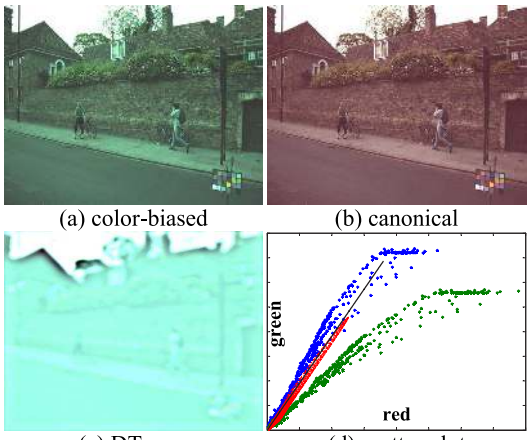


Fig. 1: The blue and green crosses in the scatter plot (d) denote the color distributions of the color-biased (a) and canonical (b) images [33], respectively. The red crosses in (d) show the color distribution of the responses of the double-opponent cells in our model (i.e., the *DT map* (c) shown in RGB space for display). The true illuminant direction is shown as a black solid line.

[28], neural network based [13], regression based [29], and corrected-moment based [30]. (2) The content of the individual images was used to automatically select and tune the most appropriate (normally simple assumption based) algorithm (or combination of algorithms) for each input image [14], [31], [15], [16], [32], [25]. In general, both of these two types of solutions are based on learning.

For example, by assuming that the distribution of RGB color values of an image captured under a canonical illuminant is a limited set, called the canonical gamut, the classical gamut mapping model proposed by Forsyth [22] needs a training phase to generate the canonical gamut containing possible colors that can be observed under a canonical illumination. In general, the gamut mapping based methods [12], [22], [23], [24], [25] possess with the elegant underlying theory and can achieve competitive performance; however, the weakness of these methods include their complication in implementation and the requirement of appropriate pre-processing [2].

Another group of typical models employing more complex statistics about the surface reflectance are the Bayesian approaches [26], [27]. These methods derive the estimate of the light source color from the posterior distribution conditioned on the image intensity data, based on the priori distribution learned from training images [27], [34]. Bayesian approaches are relatively simple to implement, but just like the grey world assumption, they do not always accord with the actual situations.

Different from assuming more sophisticated assumption, another learning strategy tries to estimate the illuminant by combining multiple methods that are normally based on simple assumption, considering the fact that none of the available methods seems universal for all images [16]. These models try to combine the output of different models using linearly or nonlinearly weighted average [35], [36], [21], or select the most appropriate methods for individual images or a certain class of scenes (e.g., indoor or outdoor) based on the intrinsic properties of input images [14], [31], [15], [16], [32], [25]. Most of such methods need to train multiple models using different machine learning techniques and feature descriptors in order to obtain the optimal combination weights or capture the relationship between the extracted features of images and the specific algorithms.

Basically, most of the above models are based on the simple Lambertian model of image formation. Differently, another group of non-biologically inspired algorithms, not mentioned too much in this paper, are based on the more comprehensive dichromatic reflection model of image formation [37], [38], [39]. These so-called physics-based methods [2], [40], [41] try to utilize the information about the physical process of reflected light by considering that both Lambertian and specular reflections occur in the scenes. Dichromatic-based methods have the promising merit of successfully dealing with the scenes containing only a very few surface colors; however, they generally suffer from the difficulty of accurately retrieving the specular reflections of a given scene [2].

1.2 Biologically inspired models

It is widely accepted that computer vision applications have always been inspired and challenged by the high efficiency of human vision [1], [19], [42], [43], [44]. There are several color constancy algorithms that take inspiration from biological systems [3], [4], [5], [6], [7], [8], [9], [18], [45], [46], among which only a few attempts seem to be inspired (to various degrees) by the biological vision at the levels of individual neurons [45], neural networks [3], [4], [7], [8], [9], [18] or psycho-physical behaviors (e.g., retinex-based models [5], [6], [47], [48]).

For example, Dufort and Lumsden [3] proposed a neural network based on DO cells. In their model, color constancy is achieved by using the output from the DO cells as input to a neural network of four hypothesized neurons in V4. Courtney et al. [8] proposed a multistage neural network incorporated with several visual mechanisms including cone adaptation in retina, center-surround spectral opponency in lateral geniculate nucleus (LGN), and spectrally-specific long-range inhibition in V4. It should be pointed out that both the models of Dufort and Lumsden [3] and Courtney et al. [8] did not show any color constancy results with real images, so it is unclear whether they work and how they perform on various real images.

Spitzer and Semo [4] proposed a biologically inspired color constancy model based on local and remote dynamical adaptation mechanisms in color-coded ganglion cells in the retina. Local adaptation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

occurs in the difference-of-Gaussians shaped classical receptive fields (CRFs) of the single-opponent ganglion cells, and the remote adaptation occurs in the regions peripheral to these CRFs. Lau and Shi [45] used the specifically normalized responses of DO neurons in V1 to obtain an illuminant invariant representation of color edges, and then the illuminant was estimated by applying the grey-world (GW) rule over the image where the color edges were weighted by the normalized DO responses [45]. It seems a question whether the normalization operation required by their DO cells, i.e., dividing DO response by the sum of all four low pass filtered responses, exists or not.

The retinex algorithms [47], [48] have been considered as the first attempts at designing computational model for human color constancy [19], [49]. Although the retinex theory offers an explanation of our ability for color constancy, and various retinex based models are still widely used with their parameters optimized for a range of color dependent applications [50], [51], [52], it is yet difficult to well reconcile with human constancy performance and its underlying biological mechanisms [19], [26], [27].

1.3 General description of the proposed model

This work aims to develop a physiologically based color constancy model by simulating the double-opponent (DO) mechanism in V1. The idea behind our so-called double-opponency based color constancy (DOCC) model originates from the interesting observation from computational experiments that the responses of DO cells to the color-biased images provide clear information about the scene illuminant. From Fig. 1 we can see that the color distribution of the responses of DO cells to a color-biased image coincides well with the vector direction of light source color. Based on this substantial observation on various scenes of multiple datasets, we directly use the standard pooling mechanisms like max or sum to extract the true illuminant from the responses of a population of DO cells, under the common assumption that the spectrum of the illumination is uniform across a scene.

This paper is based on our previous work reported orally at an international conference [53], which is substantially extended here as follows. (1) Considering that the spectral sensitivities of digital cameras used to record the images are different from the cone spectral sensitivities of HVS, we transform the input images from RGB space to LMS cone inputs for more physiological plausibility. (2) Experimental tests are carried out on more datasets, as well as more quantitative analysis including both the metric of angular error and the Wilcoxon Signed-Rank Test (WST) [4] for performance evaluation, and the inter- and intra-dataset tests for robustness evaluation. (3) Two kinds of pooling techniques (max and sum) are introduced and evaluated when estimating the light source color from the DO responses. (4) The spatial frequency tuning properties are analyzed in order to explain why the proposed model works well. The source codes is available at http://www.neuro.uestc.edu.cn/vccl/projcc.html.

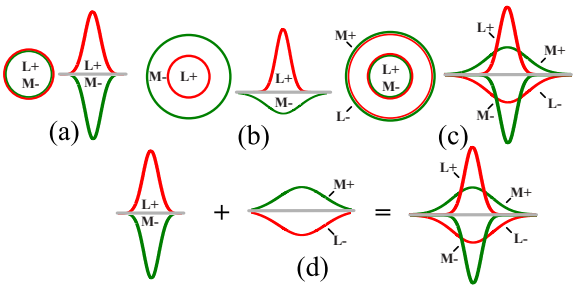


Fig. 2: The receptive field (RF) of red-green single-opponent Type II cells in LGN with color-opponent centre-only RF (a) and Type I cells with color-opponent centre-surround RF (b). The RF of red-green double-opponent cells in the primary visual cortex (V1) (c) can be computationally constructed using two single-opponent Type II cells with different RF scales and opposite signs (d). In the expression of "A+" or "B-", the sign "+" and "-" denote the excitation and inhibition, respectively. Adapted from [1].

The rest of this paper is organized as follows. In Section 2, we begin with an overview of the color processing mechanisms in HVS. We describe in details our DOCC model in Section 3. In Section 4 our model was experimentally verified on three image datasets. We then discuss some attributes of our model in Section 5. Finally, some concluding remarks and future directions are presented in Section 6.

2 Color Processing in Visual System

Though we are still very far from fully understanding the neural mechanisms underlying the transformation of color signals from retina to color perception in the higher cortexes [54], the neural basis for color information processing in the early visual stages are relatively better characterized at the level of individual neurons. In short, color processing progresses through a series of hierarchical stages [1]: after the light absorption by cone photoreceptors, cone activities are compared by cone-opponent retinal ganglion cells; these color signals are transmitted via LGN to V1 and then higher areas (e.g., V2 and V4).

Cone photoreceptor layer of retina: There are two types of photoreceptors: rods and cones, and cones are responsible for color vision. Based on the spectral sensitivities, cones can be classified into short-wavelength cone (S-cone), medium-wavelength cone (M-cone) and long-wavelength cone (L-cone), which respond preferably to the blue (B), green (G) and red (R) colors, respectively. The color information into the eyes is first coded in a trichromatic way via L-, M- and S-cones in the retina, and then propagated in the way of color opponency via single-opponent and double-opponent neurons at the levels of retinal ganglion

layer, LGN and V1 (and the higher cortical areas).

Single-opponent (SO) cells: Most color-sensitive cells in ganglion layer and LGN are SO cells that code the color information within their receptive fields (RFs) in the way of red-green, blue-yellow, and black-white opponency. Fig. 2(a) and (b) show the RF structures of the red-on/green-off SO cells of type-II and -I, respectively. Type-II cells (Fig. 2(a)) have center-only color opponent RF and respond well to uniformly colored areas [55]. In contrast, type-I cells (Fig. 2(b)) have center-surround RF structure with color opponency and respond well to color areas and luminance contrast [55], [56], [57].

Double-opponent (DO) cells: Many experiments [56], [57], [58] have revealed that DO cells exist widely in V1. The most important feature that emerges with DO color cells in V1 is a spatially transformed receptive field capable of detecting local color contrast [59], which shows band-pass spatial frequency tuning to colored gratings. It has been suggested that neurons computing local color contrast between center and surround are the basis to discount the illuminant changes of environment within receptive field, and thus helpful to color constancy [1], [55], [57].

Based on the spatial structure of RF, DO cells could be classified into two types: DO cells with concentric RFs (Fig. 2(c)) and with oriented RFs [60] (not shown here). In particular, the RFs of the first type (Fig. 2(c)) have concentrically organized center-surround structure and are both spectrally and spatially opponent. These properties make such DO cells the reasonable building blocks of color constancy [1], [58]. It has been found that the majority of DO cells receive unbalanced cone inputs [55]. In this study we use the DO cells with unbalanced concentric RFs to build color constancy model.

Color sensitive cells in higher visual cortexes: The color selective cells of thin stripes in V2 are involved in the elaboration of hues [61], [62], which probably suggests that the color stream originally coded in color opponency space (red-green, blue-yellow, and black-white) in early visual cortex is now ready to transform to trichromacy space (red, green, and blue) in higher visual cortexes like V2. Subsequently, the color stream coded in trichromacy way further propagates to glob region in posterior inferior temporal cortex (PIT, a loosely defined area including V4 and its neighboring regions), where a population of cells in PIT of macaque monkey were recently identified to contain an explicit representation of primary hues (red, green, blue) [63], [64].

Neurons in V4 with extensive RFs are able to pool information from large fields and show constant responses when the illuminant composition of wavelength lighting the surfaces is changing [65], which indicates that V4 neurons are important for defining global color constancy [66].

3 Double-opponency based color constancy (DOCC)

Fig. 3 shows the flowchart of our DOCC model. The different layers in this hierarchical framework mainly correspond to the color processing strategies involved in the HVS from the retina to V1 and the possible higher areas like V4. The key idea is that some V4 cells serve to estimate the illuminant from the responses of DO cells in V1 with the canonical neural pooling mechanism of max or sum [67].

Cone layer: The image is first transformed from RGB space to cone's LMS space according to [68]

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.3192 & 0.6098 & 0.0447 \\ 0.1647 & 0.7638 & 0.0870 \\ 0.0202 & 0.1296 & 0.9391 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1)$$

Then, the three components of the input image in LMS space, denoted by $l(x,y)$, $m(x,y)$, and $s(x,y)$, are sent into the L-, M-, and S-cones, respectively. Also, an extra yellow (Y) component, given by $y(x,y) = m(x,y)+s(x,y)$, is constructed for the computation of blue-yellow (S-Y) opponency. In addition, a luminance channel, denoted by $b(x,y)=l(x,y)+m(x,y)+s(x,y)$, is constructed for the computation of black-white opponency of luminance.

Retinal ganglion/LGN layer: The retinal ganglion cells (RGCs) locate in the output layer of retina, which receive the outputs of the cones through different retinal layers like bipolar cells, and then send signals to LGN. Note that we did not explicitly simulate the cells in the sub-layers between cones and RGCs, since these sub-layers function mainly to form the RF properties of RGCs as shown in Fig. 2(a) and (b). RGCs and LGN cells have similar RF properties. Here we implement the processing of RGC and LGN within a single step for simplicity. In this study we only consider the type-II SO cells (Fig. 2(a)), which are used to construct the RFs of DO cells (Fig. 2(c) and (d)). The signals in this layer are first transformed from the LMS space to the SO space according to [18]

$$\begin{pmatrix} O_{lm} \\ O_{ys} \\ O_{b_+} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} l \\ m \\ s \end{pmatrix} ; \begin{pmatrix} O_{ml} \\ O_{sy} \\ O_{b_-} \end{pmatrix} = -\begin{pmatrix} O_{lm} \\ O_{ys} \\ O_{b_+} \end{pmatrix} \quad (2)$$

The RF spatial structure of each component of a type-II SO cell could be described using a two-dimensional (2D) Gaussian function [1], [55] written as

$$RF(x,y;\sigma) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (3)$$

where the standard deviation σ controls the scale (i.e., the size) of RF. Taking a SO cell of type-II (Fig. 2(a)) with red-on/green-off (L+M-) opponency as example, its response is given by

$$SO_{l+m-}(x,y;\sigma) = O_{lm}(x,y) * RF(x,y;\sigma) \quad (4)$$

where $*$ denotes the convolution. Similarly, we can compute $SO_{m+l-}(x,y)$ for M+L- SO cells, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

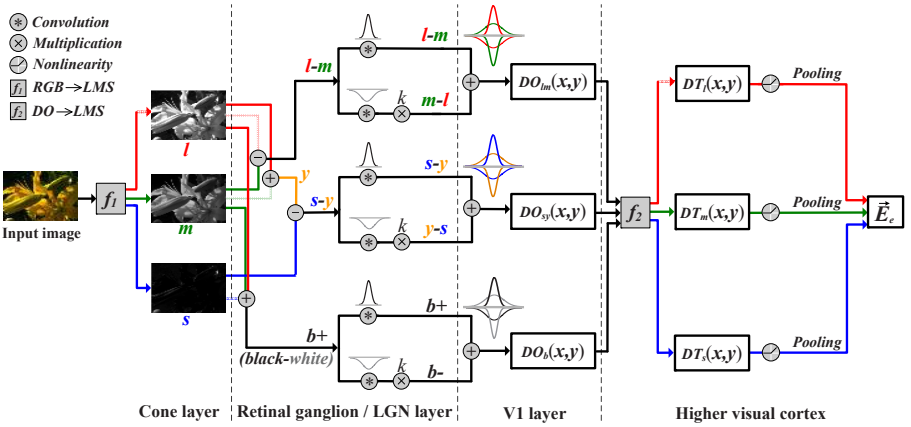


Fig. 3: The flowchart of our DO based color constancy model. The symbol f_1 denotes the transformation from RGB to LMS space. f_2 denotes the transformation from DO to LMS space. l, m, s, y, and b denote respectively the components of the red, green, blue, yellow, and luminance channels. $Pooling$ denotes the mechanism of max or sum.

$SO_{s+y-}(x,y)$ and $SO_{y+s-}(x,y)$ for the responses of blue-yellow SO cells, and $SO_{b+}(x,y)$ and $SO_{b-}(x,y)$ for the response of brightness-sensitive cells. In the expression of "A+B-", the sign "+" and "-" denote the excitation and inhibition, respectively.

V1 layer: Based on the viewpoint of physiological experiments [56], we construct the RF of a V1 DO cell of L+M-/M+L- shown in Fig. 2(c) using the outputs from two LGN SO cells of type-II with different scales: one red-on/green-off (L+M-) SO cell with smaller RF scale and another green-on/red-off (M+L-) SO cell with larger RF scale (Fig. 2(d)). Thus, the response of a DO cell can be computed as

$$\begin{aligned} DO_{lm}(x,y) &= SO_{l+m-}(x,y;\sigma) + k \cdot SO_{m+l-}(x,y;\lambda\sigma) \\ DO_{sy}(x,y) &= SO_{s+y-}(x,y;\sigma) + k \cdot SO_{y+s-}(x,y;\lambda\sigma) \\ DO_{b}(x,y) &= SO_{b+}(x,y;\sigma) + k \cdot SO_{b-}(x,y;\lambda\sigma) \end{aligned} \tag{5}$$

where σ and $\lambda\sigma$ define respectively the scales of the RF center and its surround of a DO cell. We set $\lambda = 3$ based on the finding that the size of RF surround is roughly 3 times (in diameter) larger than that of RF center [53]. k is a relative cone weight that controls the contribution of RF surround. $k \neq 1$ implies that DO cells receive unbalanced cone inputs [55].

Higher visual cortex: It is unknown yet at which stage the human visual system finally realizes color constancy. What is known, however, is that color constant cells have been found at the level of V4 [1], [58]. V4 cells normally have very large receptive fields [66], which may endow the V4 cells with the ability to extract light source color based on global statistics. Along this line, we first transform the output of DO cells from DO to LMS space according to [18]

$$\begin{pmatrix} DT_l(x,y) \\ DT_m(x,y) \\ DT_s(x,y) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}^{-1} \begin{pmatrix} DO_{lm}(x,y) \\ DO_{sy}(x,y) \\ DO_{b}(x,y) \end{pmatrix} \tag{6}$$

Note that transformation of DO to LMS space by Eq (6) is based on some physiological evidences suggesting that transforming signals from DO to LMS space may happen in the cortex [62], [63], [64]. In the following, we will call the output of DO cells given by Eq (6) the *DT map* in LMS space.

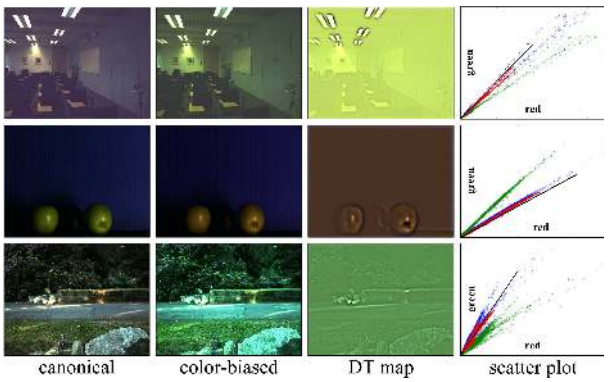


Fig. 4: More examples illustrating the close fitting of the color distribution (red dots) of DO responses (i.e., *DT map*) in RGB space to the true illuminant (black line). From top to bottom: images are from the Gehler-Shi dataset [33], SFU lab dataset [69], and SFU HDR dataset [70], respectively.

We also need to transform the information (e.g., *DT map*) from LMS to RGB space for the purpose of display. The transformation is written as [68]

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 5.3341 & -4.2829 & 0.1428 \\ -1.1556 & 2.2581 & -0.1542 \\ 0.0448 & -0.2195 & 1.0831 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{7}$$

Taking the color-biased image as shown in Fig. 1 as input, we compute the output of DO cells using Eqs (1)∼(6). We can clearly find from the scatter plot shown in Fig. 1 that the color distribution of the DO responses (i.e., the *DT map*) represented in RGB space fits closely to the true illuminant (provided in RGB space by the datasets), which holds true for the most of color-biased images in the three datasets [33], [69], [70] used in this study. Fig. 4 shows more examples on the images selected from the three datasets.

Based on this substantial observation, we speculate that one of the functional roles of the specific type of

DO cells in V1 with concentric RFs (Fig. 2(c)) is to code the illuminant information of scenes. Then, the color sensitive cells with large RFs in higher visual cortexes, e.g., V4, might adopt certain pooling mechanism to compute an accurate illuminant estimate of the scene using the color distribution of the DO cells' output in the LMS space. Considering that neuronal responses are non-negative, a simple non-linear operation is used to set the negative DO responses in LMS space to be zero before pooling them.

As done in most existing models, we assume that the scene is illuminated by one single light source with spatially uniform color across the scene. Then, the color of the illuminant $\vec{E}_e = (e_l, e_m, e_s)$ is estimated using certain pooling mechanism according to

$$\begin{aligned} e_i &= \underset{(x,y)}{pooling}\big(DT_i(x,y)\big)/coef, \quad i \in \{l,m,s\} \\ coef &= \sum_{i\in\{l,m,s\}} \underset{(x,y)}{pooling}\big(DT_i(x,y)\big) \end{aligned} \tag{8}$$

where $pooling(\cdot)$ denotes the canonical neural computation of max or sum [67] pooling the DO responses in separate channels over the whole image. With these two different pooling schemes, our model proposed in this paper has two variations: $DOCC\text{-}LMS(max)$ and $DOCC\text{-}LMS(sum)$. In contrast, we named our previous model in [53] as $DOCC\text{-}RGB(max)$.

It should be mentioned that from a signal-processing point of view, the operations performed in the proposed DOCC system (Fig. 3) are all linear, except for the non-linear operation dealing with the negative DO values. It is clear that when ignoring the non-linear operation, the transfer from LMS to SO (as well as the transfer from DO to LMS) is computationally redundant and can be skipped, and our model is equivalently to directly apply the linear DO filters in LMS space.

4 Experimental results

The proposed model was compared with multiple methods on three typical datasets, i.e., Gehler-Shi dataset [33], SFU lab dataset [69], and a more recent SFU HDR (high dynamic range) dataset [30], [70].

The existing methods considered are classified into three types according to [2], including: (1) dichromatic reflection model based: inverse-intensity chromaticity space (IICS) [38]; (2) low-level statistics based: Grey World (GW) [20], White Patch (WP) [5], 1st-Grey Edge (GE1), 2nd-Grey Edge (GE2) [11], Shades of Grey (SG) [10], General Grey World (GG); (3) learning based: Bayesian [27], Regression (SVR) [2], [29], Automatic Color Constancy Algorithm Selection (CART-AAS) [31], using natural image statistics (CCNIS) [2], [16], Spatio-spectral Statistics (SS) [2], [28], Pixel-based Gamut Mapping (GM(pixel)) [2], [22], Edge-based Gamut Mapping (GM(edge)) [2], [24], Exemplar-based (Exemplar) [15] and Corrected-Moment (CM) based [30] models. Note that roughly, the proposed model could be classified into the low-level statistics based ones, since it utilizes the statistics of DO responses emerging at the early stages of the visual system.

In this paper we directly cited the results of Gehler-Shi dataset and SFU lab dataset from [2], [71] for comparison. As for the third SFU HDR dataset, since it was only recently available [30], [70] and few methods have reported results on this dataset, we compared multiple algorithms on this dataset by running their Matlab source codes available from [71] with optimal parameter settings.

4.1 Performance Measure

We choose the frequently used angular error ε as the error metric, which is computed as [2]

$$\varepsilon = cos^{-1}\Big(\big(\vec{E}_e \cdot \vec{E}_t\big)/\big(\|\vec{E}_e\| \cdot \|\vec{E}_t\|\big)\Big) \tag{9}$$

where $\vec{E}_e$ and $\vec{E}_t$ are respectively the estimated and the true light source colors, and $\vec{E}_e \cdot \vec{E}_t$ is their dot product. $\|\cdot\|$ denotes the Euclidean norm. Due that the true light source colors provided in the three datasets are all in RGB space, we first transformed the estimated illuminant (e_l, e_m, e_s) from LMS to RGB space with Eq (7) before calculating the angular error.

We also tested our model using both the intra-dataset evaluation with cross validation and the inter-dataset evaluation on the three datasets. Similar as done in [2], we adopted the form of threefold cross validation for intra-dataset evaluation, which is as follows. For the SFU lab dataset [69] and SFU HDR dataset [70], we first randomly divided the whole dataset into three parts. Next, applying the model on any two parts, the optimal parameters (σ and k) were obtained with an exhaustive search. Then, the model was tested on the third part of the data with this optimal parameter setting. For a complete procedure of threefold cross validation, the steps mentioned above were repeated 3 times to ensure that each image occurs in the test set only once and all images in the whole dataset is either in the training set or in the test set at the same time. In this work, we repeated 100 times of threefold cross validation procedure for both of the SFU lab and SFU HDR datasets, and hence, obtained 300 angular error measures for each dataset, which were averaged as the intra-dataset evaluation based performance. For the Gehler-Shi dataset [33], we adopted a simpler threefold cross validation since the threefolds are provided by dataset. With the provided three parts of the data, we tested the model on one part with the optimal parameter setting obtained by exhaustive search on other two parts, which was repeated 3 times. We averaged the 3 angular error measures as the cross validation based performance for this dataset. Note that we only conducted the cross validation for $DOCC\text{-}LMS(max)$. In the following,

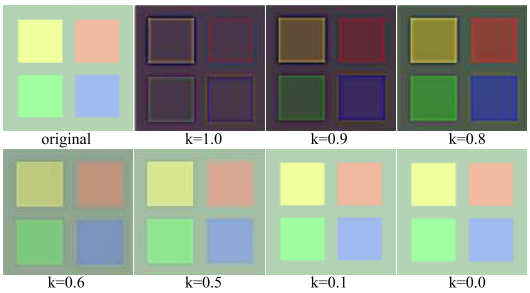


Fig. 5: The responses of model DO cells to a synthetic image rendered under greenish illuminant. With the flexible cone weights (k), the output of DO cells can encode both of the color edges and color regions.

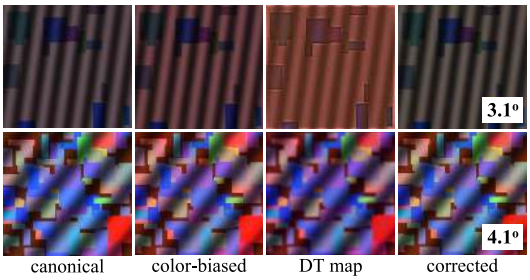


Fig. 6: Mondrian images with less (first row) or more (second row) color blocks. The *DT map* and the corrected images are also shown (the angular error is shown on the lower right corner). $k = 0.9$ and 0.3 are used for the *DT map* computation of the two images. The mondrian images were generated using the surface reflectance spectra combined with illuminant spectra from [69].

we will name $DOCC\text{-}LMS(max)$ with intra-dataset cross validation as $DOCC\text{-}LMS(max\text{-}cro)$.

In order to further determine whether the difference of angular errors between two algorithms is statistically significant or not, the Wilcoxon Signed-Rank Test (WST) has been recommended as another valuable tool for performance evaluation [2], [14], [31], [25], [72]. In this study, given any two different algorithms, the WST was run on their angular error distributions on the whole dataset, and its result was used to conclude that at a specific (e.g., 95%) confidence level, the angular error of one algorithm is often lower or higher than that of another algorithm, or there is no significant difference between them.

4.2 Parameter Setting

Our model has two free parameters, i.e., the scale of RF (σ) (in Eqs (3)∼(5)) and the cone weight (k) (in Eq (5)). Fig. 5 indicates that DO cells with balanced cone weights (i.e., $k = 1$) only respond to the color contrast (i.e., color edges), since the double-opponent operation makes the different cone responses to homogeneous surfaces always cancel each other. In contrast, DO cells receiving unbalanced cone inputs (i.e., $k \neq 1$) can respond to both the color edges and color patches, which is consistent with the physiological observations [1]. DO cells with higher k respond much more strongly to the color edges than to the color patches. Fig. 6 further suggests that if only fewer objects and hence fewer edges are available in an image (the top row of Fig. 6), emphasizing on enhancing edges (with higher k) is much better than on enhancing color regions (with lower k) in terms of benefiting the accuracy of illuminant estimation. More discussion about the influence of cone weights on the properties of DO cells will be presented in section "5 DISCUSSION".

Fig. 7 shows the influences of different parameter values on the performance of the proposed $DOCC\text{-}LMS(max)$ in terms of median angular error on three image datasets. According to which we roughly set $\sigma = 3.0, 3.0, 3.5$, $k = 0.1, 0.6, 0.7$ for the Gehler-Shi dataset, SFU lab dataset, and SFU HDR dataset, respectively. We also exploited the parameter space of $DOCC\text{-}LMS(sum)$ (the figure is not shown for space limitation), and obtained the optimal parameter setting of $DOCC\text{-}LMS(sum)$ $\sigma = 0.5, 1.5, 8.0$, $k = 1.0, 1.0, 1.0$ for the three datasets. Similar to other algorithms, the optimal parameter setting of our models varies across different datasets, but fixed for all the images of the same dataset. Note that though our models have different optimal parameter settings on different datasets, it is interesting to find that from Fig. 7, the median measures on three datasets are consistently lower than 5.0^o by setting arbitrary values among $1.5 \leq \sigma \leq 8.0$, and $0.1 \leq k \leq 1.0$, which are quite wide ranges for parameter setting that can produce the acceptable performance on all of the datasets tested here. This indicates that our models have promising potential to easily obtain acceptable results by loosely tuning the only two parameters.

4.3 Gehler-Shi Dataset of Real-World Images

Gehler-Shi dataset [27], [33] contains 568 high dynamic range linear images, including a variety of indoor and outdoor scenes, captured using a high-quality digital SLR camera in RAW format and therefore free of any color correction. In this study, the color-checker patch in each image used for computing ground truth illuminant was masked out in order to fully evaluate the performance of a specific model. Moreover, we down-sampled the original large images to half of its original size for computational reason.

The results of multiple methods are listed in Table 1. All methods were executed with optimal parameter settings. From Table 1, we can find that the performances of $DOCC\text{-}LMS(max)$ and $DOCC\text{-}LMS(max\text{-}cro)$ almost arrive at or beyond the best performance (for all of error measures) of the state-of-the-art learning-based algorithms on the indoor, outdoor and entire datasets. In addition, we can see that with a slightly higher median error than GG (3.53^o vs. 3.45^o), $DOCC\text{-}LMS(sum)$ ranks first among other static algorithms on the whole dataset in all measures.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

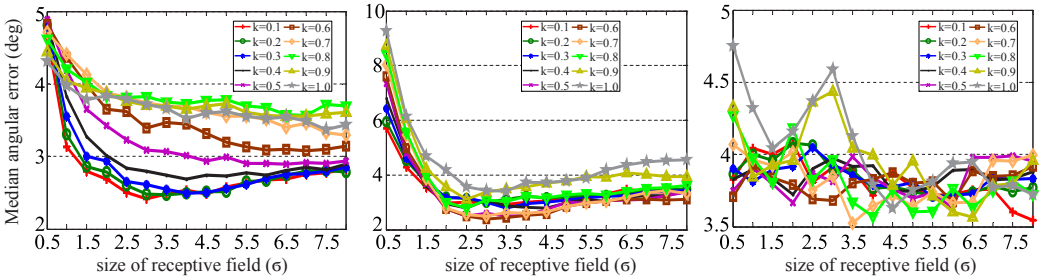


Fig. 7: The influence of receptive field size (σ) and the cone weight (k) of our model ($DOCC\text{-}LMS(max)$) on the measure of median angular error. Left: Gehler-Shi dataset, middle: SFU lab dataset, right: SFU HDR dataset. For experimental evaluation, in this study we respectively set $k = 0.1, 0.6, 0.7$ and $\sigma = 3.0, 3.0, 3.5$ for Gehler-Shi dataset, SFU lab dataset, and SFU HDR dataset.

TABLE 1: Performance of various methods on the Gehler-Shi dataset.

Methods		All images (568)			Indoor (246)			Outdoor (322)		
		Median	Mean	Worst-25%	Median	Mean	Worst-25%	Median	Mean	Worst-25%
Do nothing (DN)		13.55°	13.65°	17.15°	13.37°	13.85°	19.02°	13.61°	13.50°	15.54°
Physics-based	IICS	13.56°	13.61°	17.95°	13.09°	13.35°	19.45°	13.68°	13.81°	16.68°
(Static) low-level statistics-based	GW	6.28°	6.35°	10.58°	5.67°	5.85°	10.42°	6.60°	6.73°	10.63°
	WP	5.68°	7.54°	16.12°	8.00°	8.78°	17.05°	4.22°	6.60°	15.25°
	GE2	4.50°	5.12°	9.25°	4.64°	5.07°	8.96°	4.20°	5.16°	9.41°
	SG	4.00°	4.92°	10.19°	4.59°	5.39°	10.69°	3.52°	4.56°	9.70°
	GG	3.45°	4.66°	10.18°	3.97°	5.11°	10.70°	3.14°	4.32°	9.62°
Learning-based	SVR	6.72°	8.08°	14.89°	7.84°	9.26°	16.79°	6.08°	7.18°	12.55°
	Bayesian	3.46°	4.82°	10.48°	5.93°	6.52°	11.91°	2.44°	3.52°	7.81°
	CART-AAS	3.34°	4.49°	10.10°	4.26°	5.23°	10.86°	2.73°	3.92°	9.15°
	CCNIS	3.13°	4.19°	9.21°	3.86°	4.83°	9.60°	2.76°	3.70°	8.50°
	SS	2.96°	3.59°	7.43°	3.64°	4.15°	7.88°	2.37°	3.17°	6.86°
	GM(pixel)	2.44°	4.20°	11.15°	4.43°	5.61°	12.57°	1.61°	3.12°	8.87°
	GM(edge)	5.60°	6.71°	13.46°	7.34°	7.92°	14.45°	4.34°	5.79°	11.90°
	Exemplar	2.27°	2.89°	5.97°	3.07°	3.42°	6.39°	1.93°	2.49°	5.26°
	CM(19 edge)	2.00°	2.80°	–	–	–	–	–	–	–
DOCC-RGB (max) [53]		2.60°	4.03°	9.35°	3.95°	5.01°	10.86°	2.06°	3.13°	7.46°
DOCC-LMS (max)		2.42°	4.09°	10.37°	4.51°	5.54°	12.33°	1.82°	2.98°	7.57°
DOCC-LMS (max-cro)		2.43°	3.98°	9.08°	3.79°	4.77°	15.62°	1.83°	3.05°	11.31°
DOCC-LMS (sum)		3.53°	4.10°	8.07°	3.75°	4.10°	7.61°	3.30°	4.09°	8.35°

Interestingly, Table 1 also shows that almost all algorithms yield poorer performance on the indoor subset than on the outdoor images. This observation is consistent with the results of the same dataset reported in [28], in which it was speculated that indoor scenes are more likely to contain some outlying regions, e.g., the surfaces with significant interreflections. We further speculate that the indoor scenes may contain more shadow regions, which are also likely to act as outliers, since the illuminant may not light and influence these regions. This observation also indicates that the most of learning-based models achieve better performance on the entire dataset mainly by improving its performance on outdoor set.

We also notice that $DOCC\text{-}LMS(sum)$ outperforms $DOCC\text{-}LMS(max)$ on the indoor set. In addition, $DOCC\text{-}LMS(sum)$ shows almost equivalent performance on both of the indoor and outdoor subsets, which indicates that pooling with sum scheme may enhance its generalizability across diverse scenes, instead of biasing to a certain kind of images (e.g., outdoor or indoor).

Fig. 8 reports the results of the statistical significance test with WST. A sign (1) at location (i, j) indicates that the median angular error of method i is significantly lower than that of method j at the 95% confidence level, and a sign (-1) at (i, j) indicates

	1) DN	2) IICS	3) SVR	4) GW	5) WP	6) GM(edge)	7) GE2	8) SG	9) Bayesian	10) GG	11) CART-AAS	12) CCNIS	13) SS	14) GM (pixel)	15) Exemplar	16) Proposed (*sum*)	17) Proposed (*max*)	Score
1)	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0
2)	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0
3)	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	2
4)	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3
5)	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3
6)	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3
7)	1	1	1	1	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	6
8)	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	7
9)	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	-1	-1	0	-1	7
10)	1	1	1	1	1	1	1	0	0	0	0	0	-1	-1	-1	0	-1	7
11)	1	1	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	0	-1	9
12)	1	1	1	1	1	1	1	1	1	0	0	0	-1	-1	-1	0	-1	9
13)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	-1	1	0	13
14)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	13
15)	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	14
16)	1	1	1	1	1	1	1	1	0	0	0	0	-1	-1	-1	0	-1	8
17)	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	13

Fig. 8: WST test on the Gehler-Shi dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 2: Performance on the SFU lab dataset.

Methods	Median	Mean	Worst-25%
DN	15.60°	17.27°	32.42°
IICS	8.23°	15.52°	40.01°
GW	7.00°	9.78°	23.33°
WP	6.47°	9.08°	20.88°
SG	3.74°	6.38°	16.40°
SS	3.45°	5.63°	12.84°
GG	3.32°	5.41°	13.66°
GE1	3.18°	5.58°	13.96°
GM (pixel)	2.26°	3.69°	9.26°
SVR	2.17°	–	–
CM (3 edge)	3.60°	4.10°	–
CM (9 edge)	2.00°	2.60°	–
DOCC-RGB (max) [53]	2.38°	4.82°	12.73°
DOCC-LMS (max)	2.21°	5.71°	16.08°
DOCC-LMS (max-cro)	2.41°	4.93°	12.80°
DOCC-LMS (sum)	4.93°	6.25°	12.43°

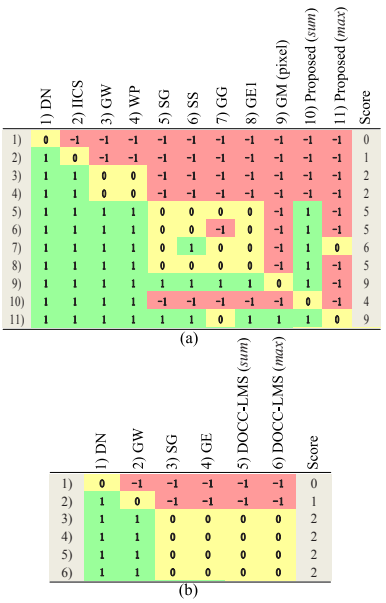


Fig. 10: WST test on the SFU lab dataset (a) and SFU HDR dataset (b).

the opposite situation. A sign (0) means that the median errors of the two methods have no significant difference. In addition, the score in the last column reports the times that a method is significantly better than others. Note that on this dataset, as well as the datasets in the following experiments, the WST analysis of CM [30] was omitted, since no error distributions is currently available for CM.

Fig. 8 indicates that the performance of our $DOCC$-$LMS(sum)$ exhibits significant improvement over various non-learning based algorithms (e.g., IICS, GW, WP, GE2, SG, with the only exception of GG). In comparison to learning-based models, the WST confirms that $DOCC$-$LMS(sum)$ significantly outperforms SVR and GM(edge), there is no distinguishable difference between the performance of $DOCC$-$LMS(sum)$ and Bayesian, CART-AAS, and CCNIS.

Fig. 8 also shows that among all the models, $DOCC$-$LMS(max)$ exhibits an indistinguishable behavior with the best performed learning-based models like SS, GM(pixel), and Exemplar. In addition, the WST indexes show that $DOCC$-$LMS(max)$ exhibits statistically better performance than the remaining models, either static or learning-based.

Fig. 9 show some examples of both indoor and outdoor images corrected with the illuminant estimates of various methods.

It should be pointed out that the Gehler-Shi dataset is provided in RAW color space, and the conversion from RGB to LMS color space by Eq (1) was applied here as if the images in this dataset were in a standard RGB color space, thus what we obtained is not a true LMS encoding. This is different from the other two datasets tested below.

4.4 SFU Laboratory Image Set

SFU lab dataset [69] contains 321 available images of 31 different objects captured with calibrated camera under 11 different lights in laboratory. Note that the results of CM method on this dataset are directly from [30], which has a special pre-processing step of raising each image to the power of 2 to give brighter colors greater weights. Table 2 reports the angular error statistics. From Table 2, $DOCC$-$LMS(max)$ achieves quite promising performance, which is better than all the existing models compared here including the learning-based models, i.e., GM(pixel) and CM(3 edge). The exceptions are CM(9 edge) and SVR, which have slighter lower median error of 2.00^o and 2.17^o. But considering the small difference (2.21^o vs. 2.17^o), it is hard to conclude that $DOCC$-$LMS(max)$ is significantly worse than SVR (note that no WST test was done between them due to the lack of error distributions of CM and SVR on this dataset).

From the WST analysis shown in Fig. 10(a), we see that although both of the GM(pixel) and $DOCC$-$LMS(max)$ obtain same best score and almost perform significantly better than other algorithms except for GG, the WST indexes also indicate that $DOCC$-$LMS(max)$ performs significantly better than GM(pixel) at the 95% confidence level. In addition, $DOCC$-$LMS(max$-$cro)$, applying random threefold cross validation with $DOCC$-$LMS(max)$, also exhibits quite competitive performance in comparison to all other models considered here.

$DOCC$-$LMS(sum)$ also achieves significantly better performance than the static methods of IICS, GW, and WP, as indicated by all of their measures in Table 2 and the WST indexes in Fig. 10(a). Furthermore, although the median error of $DOCC$-$LMS(sum)$ is poorer than that of SG, SS, GG, and GE1, the $DOCC$-$LMS(sum)$ performs better than SG, SS, GG, and GE1 in terms of angular error measure of worst-25%, which indicates better robust performance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

Fig. 9: Some examples of indoor and outdoor images from Gehler-Shi dataset corrected with multiple methods.

TABLE 3: Performance on the SFU HDR dataset.

Methods	Median	Mean	Worst-25%
DN	14.66°	15.10°	19.40°
GW	7.42°	8.05°	15.03°
Max-RGB (blur)	3.90°	6.30°	–
SG	3.89°	5.65°	12.72°
GE	3.78°	6.00°	13.60°
CM (3 edge)	3.20°	4.00°	–
CM (9 edge)	2.70°	3.50°	–
DOCC-LMS (max)	3.50°	6.19°	13.94°
DOCC-LMS (max-cro)	4.14°	6.04°	13.30°
DOCC-LMS (sum)	5.27°	6.42°	13.04°

4.5 SFU HDR dataset

The SFU HDR dataset, recently collected by Funt and Shi [70], includes 105 high dynamic range (HDR) linear images with indoor and outdoor scenes. A color checker was placed in the scenes for recording the light source color (again the color checker was masked during illuminant estimation). Considering that the HVS indeed directly deals with the real scenes with high dynamic range, testing our algorithm on HDR dataset is basically a logical option.

In Table 3, the results are shown for multiple algorithms. In this table, the results of GW, SG, and GE were obtained by running the Matlab codes from [2], [71] with optimal parameter settings. The results of Max-RGB (post-blur) and CM are directly from [30]. Table 3 indicates that $DOCC\text{-}LMS(max)$ performs better than all the static algorithms compared (i.e., GW, SG, GE and Max-RGB (post-blur)) except for the learning-based CM in terms of median measures. The WST analysis in Fig. 10(b) further indicates that all algorithms tested here show indistinguishable performance on this set at the 95% confidence level except better than GW.

Note that on this dataset our $DOCC\text{-}LMS(max)$ obtains a median angular error of $3.50°$, a performance that is competitive, but not quite outstanding. This may partially be explained by the last sub-figure in Fig. 7, in which the plot is much more spiky than the other two datasets. We have investigated and found that for some high dynamic range images in this dataset, DO cells show extremely strong responses to the quite high contrast boundaries between the regions with extremely high and low brightness. Such striking variations in DO responses could result from even a small change of receptive field size (σ), e.g., from 1.0 to 1.5. This may result in accidental fluctuation in angular errors because of the inaccurate estimate of illuminant components when using the max pooling. This behavior indicates that the proposed model is not so robust especially when emphasizing the edge information in HDR images with the pooling scheme of max, and suggests more complicated pooling schemes to improve the illuminant estimate from the responses of DO cells.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence

TABLE 4: Performance of inter-dataset test.

Training dataset	Test dataset					
	a.Gehler-Shi		b.SFU lab		c.SFU HDR	
–	med	mean	med	mean	med	mean
a.	–	–	2.99°	5.95°	3.91°	6.07°
b.	3.39°	5.03°	–	–	3.67°	6.20°
c.	3.69°	5.11°	2.61°	5.41°	–	–

TABLE 5: Performance of two non-linear operations.

DOCC-LMS	Gehler-Shi		SFU lab		SFU HDR	
–	med	mean	med	mean	med	mean
pos-max	2.42°	4.09°	2.21°	5.71°	3.50°	6.19°
pos-sum	3.53°	4.10°	4.93°	6.25°	5.27°	6.42°
abs-max	3.12°	4.50°	2.39°	5.27°	3.52°	5.78°
abs-sum	2.95°	3.87°	4.92°	6.24°	5.31°	6.46°

4.6 Inter-dataset evaluation

Inter-dataset evaluation was done by testing our model with optimal parameters learned from one dataset (Fig. 7) on all other two datasets. This is a more challenging task and almost has not been considered by both of statistics-based and learning-based illuminant estimation papers with the only exception of the recent work in [15].

The results of inter-dataset evaluation are listed in Table 4, which demonstrates the quite competitive inter-dataset accuracy for various cases mentioned. For example, we first look at the inter-dataset performance of $DOCC\text{-}LMS(max)$ on Gehler-Shi dataset. The median and mean errors of $DOCC\text{-}LMS(max)$ are respectively 3.39^o and 5.03^o with the optimal parameter setting from SFU lab dataset, and 3.69^o and 5.11^o with the optimal parameter setting from SFU HDR dataset, which indicates that no matter the parameter setting was based on SFU lab dataset or SFU HDR dataset, $DOCC\text{-}LMS(max)$ always produced consistently competitive or even better performance (in terms of both median and mean) in comparison with most of the static models and learning-based algorithms as listed in Table 1.

Similar consistency in acceptable performance holds for the SFU HDR dataset with the optimal parameters from Gehler-Shi dataset or SFU lab dataset. Analyzing from another point of view, we can also find that with the same optimal parameters on Gehler-Shi dataset, $DOCC\text{-}LMS(max)$ shows consistently and equivalently competitive performance on the SFU lab dataset (e.g., with the median error of 2.99^o and mean error of 5.95^o, respectively). Similarly, $DOCC\text{-}LMS(max)$ obtained competitive performance (2.61^o for median and 5.41^o for mean) for the inter-dataset evaluation on the SFU lab dataset when using the optimal parameters determined on SFU HDR dataset.

4.7 About the nonlinear operation

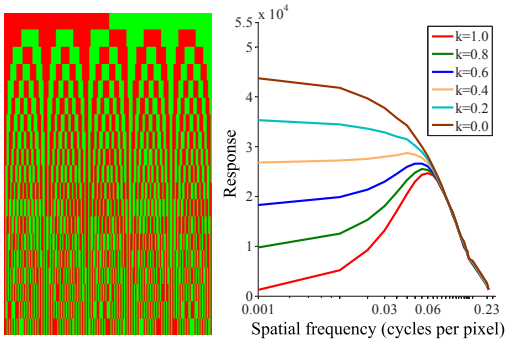
The only non-linear operation in our model is to set the negative DO responses in LMS space to be zero before pooling (*pos-max* and *pos-sum* in Table 5). In order to see the impact of non-linear operation, we also tested another standard approach of taking the absolute values (as in the grey-edge algorithms) to deal with the negative DO responses ($abs\text{-}max$ and $abs\text{-}sum$ in Table 5). Table 5 compares the results of two different non-linear operations. Note the results of $pos\text{-}max$ and $pos\text{-}sum$ are directly from Table 1∼3 for convenience of comparison. It is clear that the non-linear operation by taking the absolute values also provides competitive performance, but slightly worse than that by setting the negative to be zero.

5 Discussion

At the level of individual neurons, though DO cells have been modeled by several researchers, even for the purpose of color constancy [3], our model differs entirely from them in how to utilize the output of DO cells. For example, our model estimates the light source color by pooling (with sum or max) the information from the separate LMS channels, which are transformed from the DO responses obtained in the double-opponent space. In addition, our DO cells receive unbalanced cone inputs, which provides a flexible way to utilize the information of scenes. Furthermore, compared to many other existing models that depend only on the statistical distributions of individual pixels and ignore their spatial contexts, our model could capture the strong dependencies between nearby pixels using the concentrically organized center-surround structure with both spectral and spatial opponency of the DO cells.

The double-opponent properties have been physiologically suggested as the ideal neural substrate for color constancy [1], [55]. However, it is still unclear how the DO cells contribute to color constancy. In this study we have demonstrated in Fig. 5 that DO cells with unbalanced cone inputs (i.e., $k \neq 1$) can respond jointly to the color contrast and color regions, and with a higher k, the color edges are emphasized much more than the color regions (e.g., $k = 0.9$ vs. $k = 0.6$). Actually, the property of the DO cells with diverse inhibitory cone weights could also be interpreted by the spatial frequency tuning of DO cells shown in Fig. 11. This figure indicates that the red(L)-green(M) DO model cells show diverse spatial frequency tuning properties with various cone weights k. When $k = 1.0$, indicating the balanced cone inputs to the RF center and surround, the DO cells show band-pass tuning property in double-opponency channels. In contrast, when k is relatively low, e.g., $k = 0.0 \sim 0.2$, indicating the quite low contribution from surround, the DO cells show typical low-pass tuning property. For other k values, e.g., $k = 0.4 \sim 0.8$, the DO cells receive typically unbalanced cone inputs in RF center and its surround, which endows the DO cells with a band-pass and partially low-pass spatial frequency tuning property. These tuning properties of our DO model cells are quite consistent with the physiological recording [55], [59].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2015.2396053, IEEE Transactions on Pattern Analysis and Machine Intelligence



(a) (b)

Fig. 11: The spatial frequency responses of model L-M DO cells to the input equluminance red-green color gratings (a) The input equluminance red-green color gratings with various spatial frequency. (b) The responses of model DO cells to the input color gratings.

Roughly, the band-pass spatial frequency information mainly signals the color edges in an image (e.g., Fig. 5, $k = 1$); in contrast, the information of low spatial frequency mainly characterizes the smoothed regions (e.g., Fig. 5, $k = 0.1$). Because of the band-pass and partially low-pass tuning property resulting from the unbalanced cone inputs, these DO cells would respond best to the edges defined by the chromatic and luminance differences [1], [58], and meanwhile, transfer partially the smoothed low-frequency global (possibly the light source) color information while smoothing and suppressing the color of local surfaces. In fact, as revealed by various models mentioned in the Introduction section, both the local edges and global information in natural scenes are important cues for estimating illuminant. For example, the mean chromaticity of an image is mainly contributed by the color regions, which is a significant statistic indicator for illuminant [73] and often be used for illuminant estimation [10], [20]. Similarly, the mean of luminance edges has also been used to estimate the illuminant [11]. Differently, the DO cells modeled here could efficiently code these two kinds of spatial frequency information of scenes simultaneously by its unbalanced center-surround structure. We speculate that such attribute of DO cells in V1 makes them be perfect physiological building blocks for encoding illuminant information of scenes.

It has been speculated that human color constancy might be (partially) achieved by the distinct characteristic of double opponency of DO cells, which could cancel the spectral bias of a light source since DO cells seems strongly responsive to color patterns but weakly responsive to full-field color stimuli or color stimuli of low spatial frequency [1], [19], [54], [58]. Conversely, our results indicate that DO cells in V1 can carry information about the light source color, which might be further utilized by the higher visual cortexes, e.g., V4, to achieve the color constancy by estimating and removing the light source color from the scenes. This testable prediction, of course, needs further validation by physiological experiments.

Another point that deserves a brief comment from a mathematical point of view is that considering the resemblance of Laplace-of-Gaussian (LoG) with difference-of-Gaussian (DoG) function [74], the linear DO filters implemented in this study can be approximately constructed by subtracting two LoGs. For example, the L+M-/M+L- DO filter shown in Fig. 2(c) can be approximated by subtracting a LoG applied in M channel from another LoG with a different scale applied in L channel. This makes the proposed DO filter seem close to the second order grey edge (GE2) filter. However, the better performance of our model mainly results from the unbalanced cone inputs (which allows to also employ zero-order information) and the non-linear operation to remove the negative DO values. Roughly speaking, our model works as a kind of GE2 filter accompanied with zero-order component and followed by non-linear suppression of negative values in LMS space.

6 Conclusion and Future Work

We proposed a physiologically inspired color constancy model based on the physiological findings on color information processing in the human visual system (HVS). We found that the responses of double-opponent (DO) cells in V1 to the color-biased images contain the exact information about the scene illuminant. Inspired by the physiological findings that the color sensitive cells in V4 have large receptive field and are suggested to be important for defining global color constancy, we utilized two possible pooling mechanisms (i.e., sum or max) at the level of V4 to estimate the illuminant from the responses of DO cells. Systematical evaluation on three typical datasets validated the efficiency of the proposed model, especially with the pooling mechanism of max. In addition, the results show that transforming the cone inputs from RGB to LMS space provides a clear improvement in illuminant estimation.

As an important future direction, it is interesting to incorporate top-down modulation. High-level feedback is known to exist abundantly in visual cortexes and play important role in stable visual functions including color constancy. An evidence supporting this idea comes from the psychophysical experiments by Hansen et al. [75] indicating that high-level visual memory of colors could be used as valuable hints by HVS for estimating the scene illuminant. Computational models have also demonstrated the advantage of memory colors of grass and sky [76] and faces [25] in improving color constancy performance.

Another challenging direction is to extend our model to work well under the situations of multiple illuminants or multiple light sources with different colors. Recently, increasing attention has been drawn

to consider the presence of multiple light sources when building color constancy models [15], [77]. Basically, the DO cells in our model process the local information involved in their receptive fields, which provides us reasonable options in the future to easily estimate the multiple illuminants.

ACKNOWLEDGMENTS

We thank all anonymous reviewers for their thoughtful comments. This work was supported by the Major State Basic Research Program under Grant 2013CB329401, the Natural Science Foundations of China under Grant 61375115, 91420105, 91120013, and the Doctoral Support Program of UESTC. The work was also supported by the 111 Project (B12027).

REFERENCES

[1] B. R. Conway, S. Chatterjee, G. D. Field, G. D. Horwitz, E. N. Johnson, K. Koida, and K. Mancuso, "Advances in color science: from retina to behavior," *J Neurosci*, vol. 30, no. 45, pp. 14 955–14 963, 2010.

[2] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Computational color constancy: Survey and experiments," *IEEE TIP*, vol. 20, no. 9, pp. 2475–2489, 2011.

[3] P. A. Dufort and C. J. Lumsden, "Color categorization and color constancy in a neural network model of v4," *Biological Cybernetics*, vol. 65, no. 4, pp. 293–303, 1991.

[4] H. Spitzer and S. Semo, "Color constancy: a biological model and its application for still and video images," *Pattern Recognition*, vol. 35, no. 8, pp. 1645–1659, 2002.

[5] E. H. Land, *The retinex theory of color vision*. Scientific America., 1977.

[6] ——, "Recent advances in retinex theory," *Vision research*, vol. 26, no. 1, pp. 7–21, 1986.

[7] A. Moore, J. Allman, and R. M. Goodman, "A real-time neural system for color constancy," *Neural Networks, IEEE Transactions on*, vol. 2, no. 2, pp. 237–247, 1991.

[8] S. M. Courtney, L. H. Finkel, and G. Buchsbaum, "A multistage neural network for color constancy and color induction," *Neural Networks, IEEE Transactions on*, vol. 6, no. 4, pp. 972–985, 1995.

[9] C.-H. Huang and C.-T. Lin, "Bio-inspired computer fovea model based on hexagonal-type cellular neural network," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 1, pp. 35–47, 2007.

[10] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *CIC*, vol. 2004, no. 1, 2004, pp. 37–41.

[11] J. Van De Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE TIP*, vol. 16, no. 9, pp. 2207–2214, 2007.

[12] G. D. Finlayson, "Color in perspective," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 10, pp. 1034–1038, 1996.

[13] V. C. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity by using a neural network," *JOSA A*, vol. 19, no. 12, pp. 2374–2386, 2002.

[14] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Improving color constancy using indoor–outdoor image classification," *Image Processing, IEEE Transactions on*, vol. 17, no. 12, pp. 2381–2392, 2008.

[15] J. H. Vaezi and M. Drew, "Exemplar-based colour constancy and multiple illumination." *IEEE transactions on pattern analysis and machine intelligence*, pp. 860–873, 2013.

[16] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE PAMI*, vol. 33, no. 4, pp. 687–698, 2011.

[17] S. D. Hordley, "Scene illuminant estimation: past, present, and future," *Color Research & Application*, vol. 31, no. 4, pp. 303–314, 2006.

[18] M. Ebner, *Color constancy*. John Wiley & Sons, 2007, vol. 6.

[19] D. H. Foster, "Color constancy," *Vision research*, vol. 51, no. 7, pp. 674–700, 2011.

[20] G. Buchsbaum, "A spatial processor model for object colour perception," *J Franklin Inst*, vol. 310, no. 1, pp. 1–26, 1980.

[21] J. Vazquez-Corral, M. Vanrell, R. Baldrich, and F. Tous, "Color constancy by category correlation," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1997–2007, 2012.

[22] D. A. Forsyth, "A novel algorithm for color constancy," *IJCV*, vol. 5, no. 1, pp. 5–35, 1990.

[23] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1209–1221, 2001.

[24] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *IJCV*, vol. 86, no. 2-3, pp. 127–139, 2010.

[25] S. Bianco and R. Schettini, "Adaptive color constancy using faces," *TPAMI*, pp. 1505–1518, 2014.

[26] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *JOSA A*, vol. 14, no. 7, pp. 1393–1411, 1997.

[27] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *CVPR*, 2008, pp. 1–8.

[28] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE PAMI*, vol. 34, no. 8, pp. 1509–1519, 2012.

[29] W. Xiong and B. Funt, "Estimating illumination chromaticity via support vector regression," *Journal of Imaging Science and Technology*, vol. 50, no. 4, pp. 341–348, 2006.

[30] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1904–1911.

[31] S. Bianco, G. Ciocca, C. Cusano, and R. Schettini, "Automatic color constancy algorithm selection and combination," *Pattern recognition*, vol. 43, no. 3, pp. 695–705, 2010.

[32] J. Van De Weijer, C. Schmid, and J. Verbeek, "Using high-level visual information for color constancy," in *ICCV*, 2007, pp. 1–8.

[33] L. Shi and B.Funt, "Re-processed version of the gehler color constancy dataset of 568 images," accessed from http://www.cs.sfu.ca/ colour/data/.

[34] C. Rosenberg, A. Ladsariya, and T. Minka, "Bayesian color constancy with non-gaussian models," in *Advances in neural information processing systems*, 2003, pp. 1595–1602.

[35] V. C. Cardei and B. Funt, "Committee-based color constancy," in *Color and Imaging Conference*, vol. 1999, no. 1. Society for Imaging Science and Technology, 1999, pp. 311–313.

[36] G. Schaefer, S. Hordley, and G. Finlayson, "A combined physical and statistical approach to colour constancy," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 148–153.

[37] H.-C. Lee, "Method for computing the scene-illuminant chromaticity from specular highlights," *JOSA A*, vol. 3, no. 10, pp. 1694–1699, 1986.

[38] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *JOSA A*, vol. 21, no. 3, pp. 321–334, 2004.

[39] M. S. Drew, H. R. V. Joze, and G. D. Finlayson, "Specularity, the zeta-image, and information-theoretic illuminant estimation," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer, 2012, pp. 411–420.

[40] G. D. Finlayson and G. Schaefer, "Solving for colour constancy using a constrained dichromatic reflection model," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 127–144, 2001.

[41] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Improving color constancy by photometric edge weighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 918–929, 2012.

[42] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 3, pp. 411–426, 2007.

[43] L. Itti, C. Koch, E. Niebur *et al.*, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[44] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *Image Processing, IEEE Transactions on*, vol. 6, no. 3, pp. 451–462, 1997.
[45] H. Y. Lau, "Neural inspired color constancy model based on double opponent neurons," 2008.
[46] M. Ebner, "Evolving color constancy," *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1220–1229, 2006.
[47] E. H. Land and J. McCann, "Lightness and retinex theory," *JOSA*, vol. 61, no. 1, pp. 1–11, 1971.
[48] E. H. Land, "Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, no. 16, p. 5163, 1983.
[49] D. H. Brainard and B. A. Wandell, "Analysis of the retinex theory of color vision," *JOSA A*, vol. 3, no. 10, pp. 1651–1661, 1986.
[50] B. Funt, J. McCann, and F. Ciurea, "Retinex in matlab?" *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 48–57, 2004.
[51] E. Provenzi, L. De Carli, A. Rizzi, and D. Marini, "Mathematical definition and analysis of the retinex algorithm," *JOSA A*, vol. 22, no. 12, pp. 2613–2621, 2005.
[52] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *Image Processing, IEEE Transactions on*, vol. 6, no. 7, pp. 965–976, 1997.
[53] S. Gao, K. Yang, C. Li, and Y. Li, "A color constancy model with double-opponency mechanisms," in *ICCV*, 2013, pp. 929–936.
[54] B. R. Conway, "Color vision, cones, and color-coding in the cortex," *The neuroscientist*, vol. 15, no. 3, pp. 274–290, 2009.
[55] R. Shapley and M. J. Hawken, "Color in the cortex: single- and double-opponent cells," *Vision research*, vol. 51, no. 7, pp. 701–717, 2011.
[56] M. S. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in the primate visual cortex," *J Neurosci*, vol. 4, no. 1, pp. 309–356, 1984.
[57] S. G. Solomon and P. Lennie, "The machinery of colour vision," *NAT REV NEUROSCI*, vol. 8, no. 4, pp. 276–286, 2007.
[58] K. R. Gegenfurtner, "Cortical mechanisms of colour vision," *NAT REV NEUROSCI*, vol. 4, no. 7, pp. 563–572, 2003.
[59] E. N. Johnson, M. J. Hawken, and R. Shapley, "The spatial transformation of color in the primary visual cortex of the macaque monkey," *Nature neuroscience*, vol. 4, no. 4, pp. 409–416, 2001.
[60] K. Yang, S. Gao, C. Li, and Y. Li, "Efficient color boundary detection with color-opponent mechanisms," in *CVPR*, 2013, pp. 2810–2817.
[61] Y. Xiao, Y. Wang, and D. J. Felleman, "A spatially organized representation of colour in macaque cortical area v2," *Nature*, vol. 421, no. 6922, pp. 535–539, 2003.
[62] D. C. Kiper, S. B. Fenstemaker, and K. R. Gegenfurtner, "Chromatic properties of neurons in macaque area v2," *Visual neuroscience*, vol. 14, no. 06, pp. 1061–1072, 1997.
[63] B. R. Conway, S. Moeller, and D. Y. Tsao, "Specialized color modules in macaque extrastriate cortex," *Neuron*, vol. 56, no. 3, pp. 560–573, 2007.
[64] C. M. Stoughton and B. R. Conway, "Neural basis for unique hues," *Current Biology*, vol. 18, no. 16, pp. R698–R699, 2008.
[65] S. Zeki, "Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours," *Neuroscience*, vol. 9, no. 4, pp. 741–765, 1983.
[66] A. W. Roe, L. Chelazzi, C. E. Connor, B. R. Conway, I. Fujita, J. L. Gallant, H. Lu, and W. Vanduffel, "Toward a unified theory of visual area v4," *Neuron*, vol. 74, no. 1, pp. 12–29, 2012.
[67] M. Carandini and D. J. Heeger, "Normalization as a canonical neural computation," *NAT REV NEUROSCI*, vol. 13, no. 1, pp. 51–62, 2011.
[68] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.
[69] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Research & Application*, vol. 27, no. 3, pp. 147–151, 2002.
[70] B. Funt and L. Shi, "The rehabilitation of maxrgb," in *Color and Imaging Conference*, vol. 2010, no. 1. Society for Imaging Science and Technology, 2010, pp. 256–259.
[71] A. Gijsenij, "Color constancy: research website on illuminant estimation," accessed from http://colorconstancy.com/.
[72] S. D. Hordley and G. D. Finlayson, "Reevaluation of color constancy algorithm performance," *JOSA A*, vol. 23, no. 5, pp. 1008–1020, 2006.
[73] J. Golz and D. I. MacLeod, "Influence of scene statistics on colour constancy," *Nature*, vol. 415, no. 6872, pp. 637–640, 2002.
[74] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
[75] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner, "Memory modulates color appearance," *Nature neuroscience*, vol. 9, no. 11, pp. 1367–1368, 2006.
[76] E. Rahtu, J. Nikkanen, J. Kannala, L. Lepistö, and J. Heikkilä, "Applying visual object categorization and memory colors for automatic color constancy," in *Image Analysis and Processing–ICIAP*. Springer, 2009, pp. 873–882.
[77] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 697–707, 2012.

Shao-Bing Gao received M.Sc. degree in Biomedical engineering from University of Electronic Science and Technology of China (UESTC) in 2013, where he is currently pursuing his Ph.D. degree in Biomedical engineering. His research interests include visual mechanism modeling and image processing.

Kai-Fu Yang received a B.Sc. and M.Sc. degree in Biomedical engineering from UESTC in 2009 and 2012, where he is currently pursuing his Ph.D. degree in Biomedical engineering. His research interests include visual mechanism modeling and image processing.

Chao-Yi Li graduated from the Chinese Medical University in 1956 and Fudan University in 1961. He became an academician of the Chinese Academy of Sciences in 1999. He is currently a professor of UESTC, and professor of Shanghai Institutes for Biological Sciences. His research interest is visual neurophysiology.

Yong-Jie Li received his Ph.D. degree in biomedical engineering from UESTC in 2004. He is now a professor of the Key Lab for Neuroinformation of Ministry of Education, School of Life Science and Technolgy, UESTC, China. His research interests include visual mechanism modeling and image processing. He is a member of the IEEE.