Combining bottom-up and top-down visual mechanisms for color constancy under varying illumination

Shao-Bing Gao, *Member*, *IEEE*, Yan-Ze Ren, Ming Zhang, Yong-Jie Li, *Senior Member*, *IEEE*

***Abstract*—Multi-illuminant based color constancy (MCC) is a quite challenging task. In this paper, we proposed a novel model motivated by the bottom-up and top-down mechanisms of human visual system (HVS) to estimate the spatially varying illumination in a scene. The motivation for bottom-up based estimation is from our finding that the bright and dark parts in a scene play different roles in encoding illuminants. However, the pure bottom-up processing is difficult to handle the color shift of large colorful objects. Thus, we further introduce a top-down constraint inspired by the findings in visual psychophysics, in which high level information (e.g., the prior of light source colors) plays a key role in visual color constancy. In order to implement the top-down hypothesis, we simply learn a color mapping between the illuminant distribution estimated by bottom-up processing and the ground truth maps provided by the dataset. We evaluated our model on four datasets and the results show that our method obtains very competitive performance compared to the state-of-the-art MCC algorithms. Moreover, the robustness of our model is more tangible considering that our results were obtained using the same parameters for all the datasets or the parameters of our model were learned from the inputs, that is, mimicking how HVS operates. We also show the color correction results on some real-world images taken from the web.**

***Index Terms*—color constancy, illuminant estimation, biologically inspired vision.**

I. Introduction

COLOR constancy (CC) aims at removing the color cast triggered by the illuminants in an image, which has quite general applications such as the white-balance in camera processing pipeline [1] and ensuring that the extracted color features of an image keep the invariance of illuminants [2]. Human Visual System (HVS) has a good CC ability under various illumination conditions [3]–[6], thus imitating the HVS may benefit some CC algorithms in improving efficiency [7]–[11]. Moreover, current HVS inspired models are limited to mimic the bottom-up mechanisms [5], [12], although the real HVS is a dynamical system combining both the bottom-up and top-down processing [13]–[15].

Fig. 1: Scenes with multiple light sources, taken from the web.

The majority of the existing CC models are generally implemented by estimating the color of the light source from a given image, and then discounting the light source color through a diagonal adaptation process [2], [16]. During the CC process, estimating illuminant is very challenging since it is ill-posed [2]. Thus, various hypotheses have been put forward in multiple models [10], [17]–[20]. Most existing CC algorithms assume that the illuminant is uniform across the scene [2], [3], [21]–[24]. However, this assumption could be easily violated in many real-world situations [25], [26], e.g., the indoor environments usually contain two or multiple illuminants (see Fig.1).

The multi-illuminant CC (MCC) problem has gained less attention in the past [2]. Land's Retinex algorithm is the first one to handle the MCC problem [10]. Barnard et al. [27] built a methodology that automatically detects non-uniform illuminants. Ebner [28] applied convolution technique to estimate the illuminants by hypothesizing that the extensively smoothed regions of images satisfy the gray-world assumption. Although this algorithm avoids explicitly segmenting the images, the regional grey-world assumption may result in an inaccurate estimation, especially for the images containing with large colorful objects [29].

Our proposed bottom-up processing utilizes convolution to locally infer the illuminants. But the differences with Ebner's method are twofold: (1) We use two different convolutional mechanisms inspired by HVS to separately handle the bright and dark areas in the images, since our analysis shows that the bright and dark regions in an image play quite different roles in locally encoding the illuminants. Then, the roughly estimated illuminant maps for bright and dark regions are further fused for constructing a bottom-up based illuminant map. (2) Our model further

Manuscript received Dec 1, 2017. This work was supported by Sichuan Province Science and Technology Support Project under Grant 2017JY0249 and 2017SZDZX0019, the National Natural Science Foundation of China under Grant 61806134, the Fundamental Research Funds for the central Universities under Grant YJ201751, and the China Aviation Science Fund under Grant ASFC-20171919002. (Corresponding author: Yong-Jie Li).

S.-B. Gao performed this work while studied as a PhD student at University of Electronic Science and Technology of China (UESTC). His current address is with the College of Computer Science, Sichuan University, Chengdu 610065, China (email:gaoshaobing@scu.edu.cn).

Y.-Z. Ren, M. Zhang, and Y.-J. Li are with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (email:{zezedi, zm_uestc}@163.com, liyj@uestc.edu.cn).

utilizes the top-down hypothesis to constrain the estimated bottom-up illuminant map, which can suppress the color bias introduced by locally convolving the colorful regions in the image.

Among others, a work that is relatively close to our top-down framework is the hybrid Retinex+SVR method [30], which first processes the images with Retinex, and then uses support vector regression (SVR) to predict a single illuminant that can diagonally balance the output of Retinex. In contrast, our model learns a full matrix transformation that accounts for the cross-channel effects in least mean square (LMS) sense, which is used for constraining the coarsely estimated maps by our bottom-up processing.

Recently, Kawakami et al. [31] formulated a physics-based method specifically designed to process the outdoor scenes with shadowed and unshadowed regions. Gijsenij et al. [25] proposed an algorithm for the scenes with two light sources. Beigpour et al. [32] formulated the multi-illuminant estimation as an energy minimization task within a conditional random field [17], [33]. Yang et al. [7] estimated the spatially varying illuminants by detecting the intrinsic grey pixels in the images. Gu et al. [34] proposed an unsupervised method that treats the multi-illuminant estimation as a segmentation problem. Similarly, Mutimbu and Robles-Kelly [35] further modeled the different image regions under multiple scales as a factor graph. Their methods need to first explicitly group the regions with similar illuminants.

Although these methods with pre-segmentation strategy reported promising results, the image segmentation itself is also a quite difficult task. Moreover, once the pre-segmentation is finished, grey world-based mechanisms are continually used to locally compute the illuminants. Thus, there is no high level constraint (supervised or unsupervised) to compensate the possible bias induced by the segmentation-based grey world estimates [35].

Most existing algorithms have paid attention to accurately detect or segment the image regions with similar light source colors [25], [32], [34]–[36]. However, most of them treat the role of luminance in the images equally without distinguishing the contribution of pixels with various luminances during illuminant estimation. One intuition is that the shadowed and unshadowed pixels in the images explicitly encode the visual cues of different illuminants. Moreover, luminance dependent processing would make sense if we further consider that in natural scenes, there always exist substantial spatial variations in both local contrast and local luminance [37].

In fact, HVS has strong luminance adaptation ability that allows us to use different mechanisms to deal with different scene regions with various luminance intensities [4], [37]. We have found that the relative luminance of pixels plays distinguishable role in estimating the illuminants through an eye tracking experiment [38].

Based on this physical motivation, we first segment the images into two parts with high and low luminance intensities. Note that such segmentation is simple and automatic without using any complex constraints as adopted by some recent pre-segmentation-based MCC methods [25], [32], [34], [35] that use the locally computed chromaticity by grey world for clustering. Then we use Difference-of-Gaussian (DoG) simulated ON-center and OFF-center receptive fields (RFs) to infer the local illuminants of the bright and dark areas of the images, respectively. A recent biologically inspired CC model [12] obtains very competitive performance using the dynamic kernel mechanisms conditioned on the local computation of luminance difference. Our proposed luminance-dependent filtering mechanisms follow the similar spirit.

Our experiments show that this step can provide a roughly accurate illuminant map for a given image. Then for top-down visual processing, we learn a color transformation that maps the estimated bottom-up illuminant map into the ground truth map. The advantage of our proposed model is that we avoid explicitly and accurately pre-segmenting a given image according to the light source colors since it is quite difficult and computationally intensive. In addition, with a learned global high level mapping that further constrains the rough illuminant map of bottom-up visual processing, the proposed method can achieve very competitive performance.

The rest of this paper is organized as follows. In section II, the motivation and hypothesis of our proposed model are described. Section III introduces the implementation of our model in details. In section IV, comprehensive experiments on four MCC datasets (one man-made dataset and three real-world datasets) are conducted to evaluate the proposed method and thoroughly compare against the state-of-the-art approaches. Finally, some discussions and concluding remarks are given in section V.

II. The motivation and hypothesis of our proposed model

The hypothesis behind our work could be simply stated as *the bottom-up visual processing provides a quick but inaccurate illuminant estimate, whereas the supervised top-down visual mechanism constrains the estimated bottom-up illuminant map to further refine the accuracy*.

A. The motivation of bottom-up hypothesis

HVS exploits various image statistics to infer the color appearance of scenes [4], [6], [39]. Our own behavioral experiments [38] using eye tracker also verified that (1) subjects are more likely to be attracted by the bright areas than dark areas in an image, and this coheres with the algorithms that use bright pixels for illuminant estimation [7], [40]–[42]; (2) in an image with unknown illuminants, neutral areas are more likely to be considered as achromatic surfaces (intrinsic grey) than colorful areas by subjects.

To further investigate the different roles of pixels with various intensities in helping the illuminant estimation, we conducted a statistical experiment to explore the relationship between the ratio of the bright and dark image regions and the ability of encoding the illuminants of the

TABLE I: The ability to encode the illuminant by the bright and dark regions under various proportions of the defined bright and dark areas of an image.

ratio of the bright area to the dark area	grey index of the bright area	grey index of the dark area
1:9	3.3	10.4
2:8	4.6	11.0
3:7	5.4	11.5
4:6	6.0	12.1
5:5	6.7	12.7

corresponding bright and dark image regions. As shown in many CC algorithms, the greyer the pixels are, the more accurate illuminant estimation is obtained with these pixels [7], [40]. By measuring the greyness of pixels with various intensities, we can indirectly check the roles of bright and dark parts of an image in contributing to encode the illuminants.

We used the Color Checker dataset [43] that consists of 568 high dynamic range linear images to evaluate our hypothesis. We first corrected the original color-biased images with the provided ground truth illuminants so that the corrected images were fully rendered under a white light source. Then, we divided each corrected image into bright and dark areas according to the pixel luminance. We computed the angular error [2] between the luminance-segmented pixel and the standard grey pixel (i.e., R=G=B) as the index to evaluate the greyness of this pixel.

The angular error ε is defined as

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

Where $\vec{E}_e$ and $\vec{E}_t$ indicate respectively a RGB vector of a pixel in the image and the standard grey pixel (e.g., [1,1,1]). The smaller the index ε is, the greyer this bright or dark pixel is, and thus the stronger ability this pixel has to unambiguously encode the illuminant. From Table I, we get two conclusions. (1) The pixels with higher luminance tend to be greyer than the pixels with lower luminance. (2) The greyness of a bright or dark region is closely linked to the relative dynamic range of luminance in an image.

Thus, once we segment the images into bright and dark parts according to their luminance, we may exploit their various strengths in encoding the illuminants for multi-illuminant estimation. Our motivation is quite straightforward since the bright and dark areas in the scenes are usually rendered under different illuminants (e.g., a scene with shadowed and unshadowed regions). In particular, we segment the images into the high and low luminance regions adaptively according to the dynamic range of luminance. Then, two different filters based on the receptive field (RF) mechanisms of the visual system are used to convolute the segmented bright and dark regions to get the estimated illuminant map for each part. Finally, the two estimated maps are fused into one bottom-up illuminant map.

Fig.2 shows the estimated illuminant maps of four natural images rendered under the spatially varying illumination. We observed that this simple bottom-up processing works

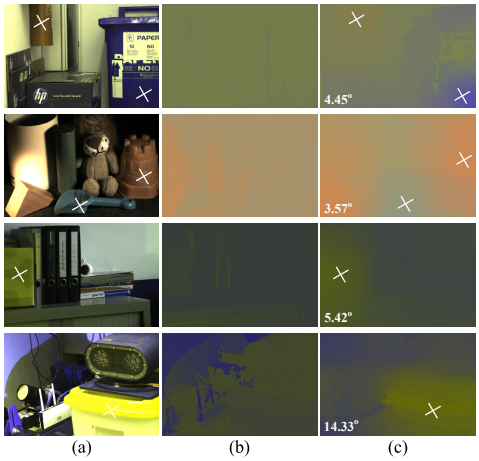


Fig. 2: The influence of large colorful objects on the illuminant estimation by our bottom-up processing. (a) Original images under the spatially varying illuminants [32]; (b) Ground truth illuminants; (c) Illuminant maps estimated by the pure bottom-up processing, in which the number indicates the averaged angular error between the ground truth and the estimated map. The white crosses indicate the locations of large colorful objects.

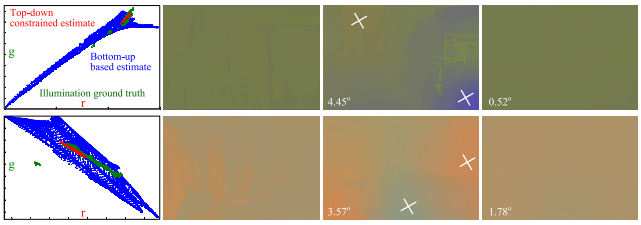


Fig. 3: The scatter plots (the first column) of the ground truth illuminant maps (the second column), the bottom-up illuminant maps (the third column), and the corrected color maps after the top-down processing (the fourth column). The images are from Fig.2.

quite well for preliminarily estimating the local illuminant maps. However, the pure bottom-up processing cannot well handle the local areas containing large colorful objects (e.g., the areas indicated by the white crosses in Fig.2), since the local convolution cannot discriminate the colors of the illuminants from the intrinsic colors of large objects. As a consequence, the estimated illuminant components in these locations are biased by the intrinsic colors of the objects.

B. The motivation of top-down hypothesis

Prior (e.g., the memory of object colors) plays an important role in visual color constancy [13], [14], [44], [45]. Our visual system can detect the change of the color appearance of a scene under the spatially varying illumination very quickly [3], [46]. However, the RF mechanism based processing may effectively encode the information of the scene illuminants [5], [12], [19], it is difficult to discriminate the surface colors of an object from the spatially varying light source colors.

Moreover, how the memory or other high level mechanisms constrain the bottom-up processing is still an open question [13], [14], [44], [45]. Here, for MCC we hypothesize that our visual system may exploit some prior knowl-

edge (e.g., the color distribution of the usually observed natural light sources [13], [47]) to further correct the bias of the roughly estimated illuminants by bottom-up processing. In this work, we use a transformation mechanism that learns a color mapping between the bottom-up illuminant maps and the ground truth illuminant maps provided by the dataset, for further constraining the roughly estimated bottom-up maps.

Fig.3 indicates the results by using the learned transformation. The blue crosses indicate the scatter plot of the estimated illuminant map purely based on the bottom-up mechanism, which is quite dispersively distributed around the real illuminant (green crosses). By exerting the learned color mapping, the final estimated illuminant map (red crosses) is more converged to the ground truth illuminant map (green crosses), and thus the accuracy of illuminant estimation is improved mainly by reducing the color bias introduced by the intrinsic colors of large objects (e.g., the local areas indicated by the white crosses).

III. THE PROPOSED MODEL

It should be stressed that we are not aiming at building biological circuits in details but providing a possible MCC framework integrating both the bottom-up and top-down mechanisms of HVS. Some works that are closely related to the biological implementation of the single illuminant estimation can be found elsewhere [5], [9], [11], [12], [48].

A. The bottom-up processing

The flowchart of our proposed MCC algorithm is illustrated in Fig.4. Our bottom-up based processing could be disintegrated into three steps.

1) Segmenting the bright and dark areas: We simply define the luminance of a pixel as the summation of R, G, B components of the pixel. Based on the information of pixel luminance, we utilize the simple clustering method of k-means to automatically group the pixels of the original color-biased image into the bright and dark parts. As for the implementation of this two-class clustering by k-means, the cluster number is set as 2; the maximum number of iterations is set as 100; the seeds are randomly initialized as a 2D matrix using the minimum and maximum luminance values of the input image; the number of times to repeat the clustering is set as 1; the empty cluster treatment is set to create a new cluster consisting of the one observation furthest from its centroid; the sum of absolute differences is used as the distance measure that k-means should minimize with respect to. Fig.5 shows an example of the grouped bright and dark areas for an image. This is different from the method in [25] which needs to locally estimate the illuminants using some low level methods (e.g., grey-world based algorithms) and then cluster the locally estimated illuminants into several groups using k-means with a preset parameter (i.e., the number of illuminants in the scene). In the following parts, the grouped bright and dark parts of an image are labelled by FB (bright area) and FD (dark area), respectively.

2) Illuminant estimation for bright and dark areas: In physiology, both On-center and Off-center RFs are simply modeled as a DoG (Difference of Gaussian) function [49], [50], which is quite accurate at depicting the linear response of a simple cell in the primary visual cortex. The mathematic formulation of DoG is described as

$$DoG(x,y) = A_1 exp(-\frac{x^2+y^2}{2\sigma_1^2}) - A_2 exp(-\frac{x^2+y^2}{2\sigma_2^2}) \tag{2}$$

where (x, y) denotes a location within the RF. σ_1 and σ_2 are the scales of the central and surrounding Gaussian kernels, respectively. A_1 and A_2 are the weights of the two kernels, which control the relative contributions from the two kernels, thus intrinsically encode the various image components (e.g., local image patches or local edges). When $A_1 > A_2$ and $\sigma_1 < \sigma_2$, the formula represents the ON-center RF. In contrast, the formula describes the OFF-center RF when $A_1 < A_2$ and $\sigma_1 > \sigma_2$.

In our implementation, the illuminant map for the bright part is estimated by convolving the bright part of image with a DoG of On-center type. Similarly, the illuminant map for the dark part is estimated by convolving the dark part of image with a DoG of Off-center type. Based on the DoG function given by Eq.(2), the function of On-center type $R_1(x, y; \sigma_{on}, k)$ is written as

$$\begin{aligned} R_1(x,y;\sigma_{on},k) = \frac{1}{2\pi\sigma_{on}^2} exp(-\frac{x^2+y^2}{2\sigma_{on}^2}) - \\ \frac{k}{2\pi(\lambda\sigma_{on})^2} exp(-\frac{x^2+y^2}{2(\lambda\sigma_{on})^2}) \end{aligned} \tag{3}$$

where the weight k is within the range of $(0, 1)$, which is very flexible in extracting various orders of image features [5]. For example, the DoG function with $k \neq 1$ primarily responds jointly to the luminance contrast and luminance regions, and with a higher k, the edges are emphasized much more than the regions (i.e., k=0.9 vs. k=0.3). On the other hand, with the unbalanced weights, the DoG plays a role of frequency analyzer that shows both of the low-pass and band-pass tuning properties, which may be quite important for extracting the globally changed illuminant in the scenes [5], [8], [19], [51].

The parameters σ_{on} and $\lambda\sigma_{on}$ respectively define the extent of center and surround RFs. We set $\lambda = 3$ based on the physiological finding that the size of surround RF is roughly three times larger than that of the center RF [50], [52]. Moreover, the DoG of Off-type $R_2(x, y; \sigma_{off}, k)$ is defined in the similar way to Eq.(3). The extent of the DoG shaped RFs will vary depending on the stimulus contrast according to the neurophysiology [49], [53]. In summary, there are three parameters (i.e., σ_{on}, σ_{off}, k) for controlling the property of DoG function. Finally, the two estimated illuminant maps for the bright and dark areas are further fused together

$$\begin{aligned} IM(x,y;c) = FB(x,y;c) \otimes R_1(x,y;\sigma_{on},k) + \\ FD(x,y;c) \otimes R_2(x,y;\sigma_{off},k) \end{aligned} \tag{4}$$

where $\otimes$ denotes the convolution operator and $c \in \{R, G, B\}$.

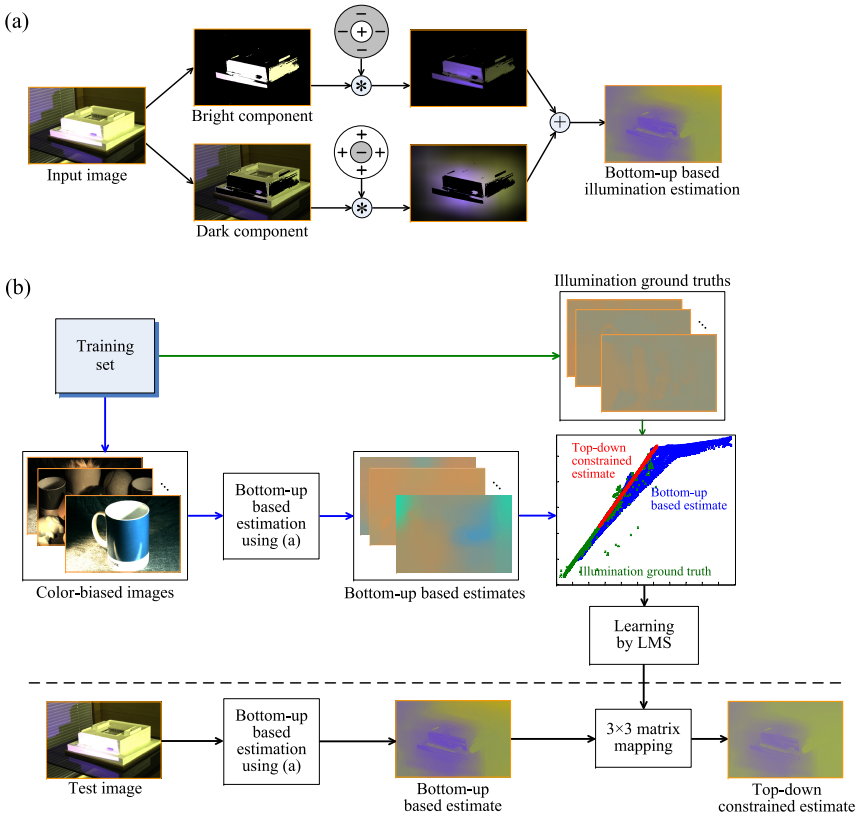


Fig. 4: The framework of our model for multi-illuminant estimation. (a) The bottom-up processing. (b) The top-down processing.

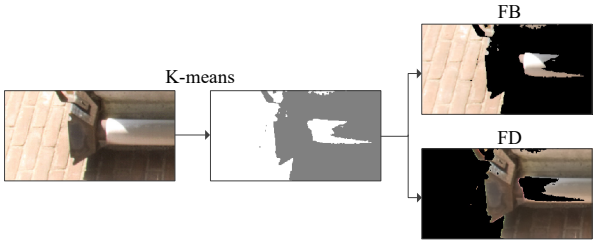


Fig. 5: An example of segmenting an original image into the bright (FB) and dark (FD) parts by k-means clustering.

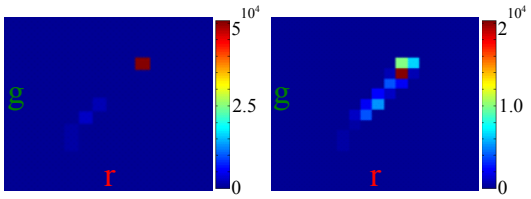


Fig. 6: Left panel indicates the ground truth illuminant map, while the right panel represents the estimated illuminant map by our bottom-up processing. The various blocks mean the different colors of the multiple illuminants and the frequencies of the corresponding illuminant colors are indicated in the attached colorbar.

3) The post-processing: In MCC, we don't know how many light sources exist in a given scene, and thus we just assume that each pixel corresponds to an independent illuminant. This hypothesis avoids the difficulty of image segmentation according to the similarity of illuminant colors as many MCC models did [25], [32], [34], [35] (e.g., some methods need to preset the number of light source colors before grouping), though this may ignore the fact that the usually observed light sources have quite restricted color gamut [32]. Fig.6 shows the distinction of the color gamut between the ground truth and the estimated illuminant map by our proposed bottom-up processing. We observed that the gamut of the ground truth mainly concentrates on a local distribution, while the gamut of the estimated map distributes quite dispersively. One of the reasons is that the color bias introduced by the intrinsic colors of the objects results in the further extension of the gamut. Thus, we process the estimated multi-illuminant map to make its color gamut more concentrated. Fig.7 shows this process. We first sort all the estimated illuminant colors by their frequencies and select the top 80% as the valid illuminant colors (♯1). Then, the estimated illuminant colors are smoothed using a fixed Gaussian kernel (♯2), and the remaining 20% of the invalid illuminant colors are replaced by the smoothed color values of the corresponding locations (♯3, ♯4). We can see that the unwanted dispersive colors induced by the intrinsic colors of the objects (labelled in the red box) can be suppressed through this step.

B. The top-down processing

With respect to the top-down implementation, we assume that HVS has evolved to build a kind of mapping mechanism, which helps refine the bottom-up illuminant maps with some constraints of usually observed natural

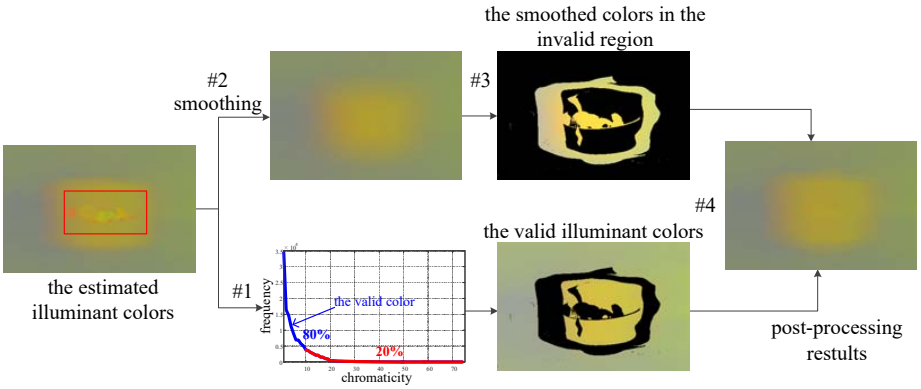


Fig. 7: The implementation of post-processing for the estimated illuminant colors by smoothing operation.

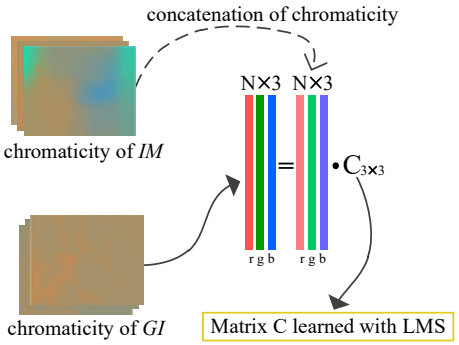


Fig. 8: The flowchart of learning the matrix $C_{3\times3}$ that maps the estimated illuminant map (IM) into the illuminant ground truth (GI).

Algorithm 1: MCC algorithm

Input: data and $\{\sigma_{on}$=2, σ_{off}=1, k=0.3$\}$
1. Bottom-up (BU) processing
1.1 segmenting $FB(x,y,c)$ and $FD(x,y,c)$ using Fig.5
1.2 computing $IM(x,y,c)$ using Eq.(4)
1.3 post-processing using Fig.7 on $IM(x,y,c)$
2. Top-down (TD) processing
2.1 learning the matrix $C_{3\times3}$ using Fig.8
2.2 applying $C_{3\times3}$ on the results of BU processing
Output: the estimated multi-illuminant map

illuminants. For HVS, it makes sense to learn the mapping using the general color distribution of natural illuminants as the constraint [15], [47], [54]. Here, we learn a chromatic mapping between the bottom-up estimated maps and the illuminant ground truth provided by the datasets [25], [32], [36].

$$GI(x,y;c) = IM(x,y;c)\cdot C_{3\times3} \tag{5}$$

where GI and IM indicate respectively the illuminant ground truth provided by the dataset and the roughly estimated illuminant map by our bottom-up processing, and $C_{3\times3}$ is a 3×3 matrix denoting the learned mapping.

Fig.8 shows the training process. Specifically, the estimated illuminant map IM is first normalized so that we get the chromaticity for each map, then we concatenate the chromaticity of each map to get a $N\times3$ matrix, where N indicates the number of chromaticity of the training dataset. Similar steps are applied to the illuminant ground truth GI to get another $N\times3$ matrix. Finally, the two $N\times3$ matrices are fitted in the least mean square (LMS) sense to get the learned mapping matrix $C_{3\times3}$.

It is worth noting that the simple mapping mechanism has been applied in a single illuminant hypothesis based CC method [55], where the edge moments of image (e.g., the 3rd order edge moments of image is consisted of a $m\times19$ matrix, where m indicates the number of images) is used to learn the mapping matrix (e.g., $C_{19\times3}$). However, our method doesn't adopt the edge moments as proposed by [55], but directly concatenates the normalized RGB chromaticity of the illuminant map to learn the mapping.

We find that this simple mapping mechanism works quite well on the MCC problem with quite effective computation. In practical implementation, we adopt the three-fold cross validation to evaluate the proposed color mapping on specific dataset [2]. For example, we randomly divide the dataset into three parts and then learn the matrix $C_{3\times3}$ on two parts, and test the learned mapping on the remaining part. These steps are repeated three times to ensure that each image occurs in the test set only once and all the images in the whole dataset is either in the training set or test set at the same time. Algorithm 1 gives the implementation of our proposed BU+TD MCC model in details.

The underlying assumption of our top-down proposal is that the learned transformation is especially usable within a database that has some common properties in terms of illuminant. The motivation behind this assumption is the fact that the ideal illumination space contains a rich ensemble of illuminants, and thus its average is likely to be similar to the average of many scenes [56]. The new prior information of our top-down proposal is that if the images under evaluation are part of a coherent image database, we will illustrate that by assuming the average of illumination of an image to be equal to the average illumination of the database, the results of the bottom-up method can be improved significantly. To the best of our knowledge, this is the first work introducing such prior to constrain the multi-illuminant distribution of a test image.

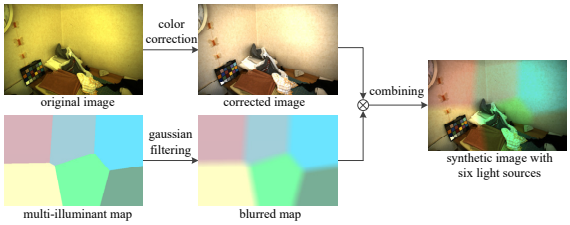


Fig. 9: The flowchart of generating a synthetic image with the spatially varying illuminants.

Fig. 10: (a) a synthetic outdoor image of color-checker dataset [43], [57] and its corresponding illuminant map; (b) an indoor image and a real-world image of Gijsenij et al. [25]; (c) a real-world image and a laboratory image from the MIMO dataset [32]; (d) two indoor images from Bleier et al. [36].

IV. EXPERIMENTS

We evaluated the proposed method against the state-of-the-art MCC approaches on one synthetic dataset and three real datasets with multiple illuminants. To generate the synthetic dataset, we first corrected the color-biased images in the color-checker dataset [43], [57], which is commonly used for single illuminant estimation, using the illuminant ground-truth to obtain the corrected images as rendered under the white light source. Then, we integrated the illuminant spectra and the camera sepectral sensitivity curves provided by [58] into (R,G,B)-values. In practical computation, all the spectrum curves are sampled and represented as vectors. Both vectors of the illuminant spectra and the camera sepectral sensitivity curves are further multiplied to produce an illuminant set with various chromaticities. Then, several (at least four) illuminants with different chromaticities were randomly selected to constitute the spatially varying illuminant maps, which were further smoothed using a Gaussian filter to get a multi-illuminant map. Finally, two images (i.e., one corrected color-checker image and one generated multi-illuminant map) were fused to construct one synthetic image with the spatially varying illumination (Fig.9).

Then, we conducted experiments on three publicly available real datasets that have been recently established to benchmark MCC [25], [32], [36]. The first dataset used here contains 59 indoor and 9 outdoor images with varying illumination [25]. The second dataset is the multi-illuminant multi-object (MIMO) dataset presented in [32]. This dataset comprises 78 images of scenes (58 laboratory images and 20 real-world images) lit with multiple illuminants. The third dataset is the multi-illuminant image dataset of Bleier et al. [36]. It consists of four scenes acquired in two-illuminant lighting setup, where each lamp can bear

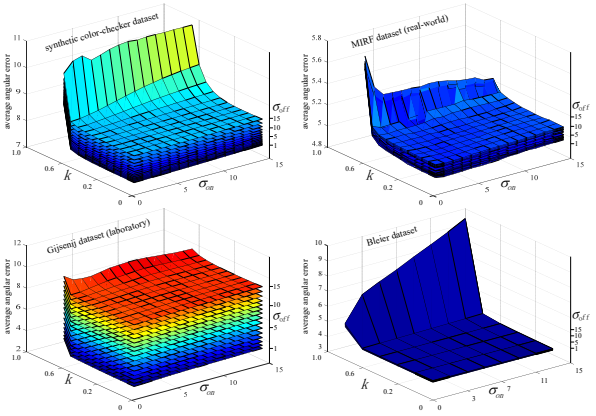


Fig. 11: The impact of DoG parameters involved in Eq (4) of our bottom-up method on the average angular error on each dataset.

several colour filters so that this dataset totally contains 36 high quality images under the varying illumination. Fig.10 (b), (c), and (d) respectively show the examples of these datasets.

A. Performance measure and parameter setting

We used the angular error (i.e., Eq.(1)) as the measure to evaluate the performance of MCC algorithms. As the scenes are lit by varying illumination, the angular error was computed pixel by pixel over the whole image. Then, the average angular error across the scene was considered as the measure of this image [25], [32], [36].

Fig.11 shows the impact of DoG parameters on the average angular error for each dataset. Basically, the performance is not sensitive to the parameters of σ_{on} and k when k<1. The possible reason for the poor performance with k=1 is that the DoG filter with k=1 has balanced RF center and surround, which makes it to primarily extract the high frequency components (e.g, edges) that may not be very informative for locally inferring the light source colors composed mainly by low frequency components in the images [12], [59]. For another parameter σ_{off}, we have observed that our method can always achieve good results for all the datasets when $\sigma_{off} < 5$. According to these observations, we always set the fixed parameters as $\{\sigma_{on} = 2,\ \sigma_{off} = 1,\ k = 0.3\}$ on all the datasets to report the performance of our bottom-up method (BU-MCC) in this work. Note that such fixed set of model parameters may not provide a clear guarantee of optimum. However, similar approaches [2], [12] have shown that this strategy is effective at evaluating the CC algorithms.

We always used the same parameters for BU-MCC to evaluate the performance of the top-down method on the four datasets. In that sense, our TD-MCC is parameter free, since we don't need to set the specific parameters for each dataset. This is in contrast to most of the learning-based methods, which need to fine tune their parameters in order to obtain good results for each dataset [2].

In order to validate the fixed parameter setting described above, we further built an automatic way to determine the optimal parameters depending on the training images.

For each dataset, we first randomly divided the whole dataset into three parts. Then, our TD-MCC model was applied on any two parts, and the optimal parameters (k, σ_{on}, and σ_{off}) were obtained with an exhaustive search and then the mapping of top-down was learned on the same training images using the optimal parameters. Finally, the optimal parameters and the mapping learned from the training parts were applied on the third part of the data to test the performance of model. For a complete procedure of threefold cross validation, the steps mentioned above were repeated 3 times. We repeated 20 times of threefold cross validation procedure, and hence, obtained 20 angular error measures (median and mean) for each dataset, which were averaged to report the performance. Note that in such a situation, our proposed model was fully automatic, since all the involved parameters were learned from the data. We report the performance of our method in this way as TD-MCC (fully automatic).

Here, we compared with a number of MCC methods, including the Retinex [10], the method of local space average color (LSAV) [28], the Grey Pixel (GP) [7], and the recent retinal model proposed by Zhang [48]. For the Retinex, we used the version implemented by Funt [60] and implemented our own version of LSAV since there is no source code publicly available. For the methods of GP and Zhang, we directly implemented the source code downloaded from their website [7], [48]. All of these algorithms share the common property that they don't need to explicitly segment the image before estimating the pixel wised illuminant map.

In addition, we also compared our method with the state-of-the-art MCC alternatives, all of which need to segment the images before the illuminant map recovery process. These are the method of Gijsenij et al. [25], the multi-illuminant recovery method of Gu et al. [34], the multi-illuminant random field (MIRF) algorithm presented in [32], and the most recent multi-illuminant method based on factor graphs (FG) [35]. Note that these algorithms employ a variety of existing grey world methods as an integral part of the illuminant recovery process. Thus, the results produced by these methods are more or less dependent upon the grey world algorithm used. Moreover, since there is no publicly available source code for these algorithms, we directly cite the results reported by [25], [35].

Finally, two deep-learning methods [61], [62] designed primarily for single illuminant estimation were compared, one of which [61] also roughly employs the merits of luminance dependent filtering mechanisms. In [61], Shi et al. built a Deep Specialized Network with two interacting branches that were trained to handle the regions of different appearances (e.g., the bright and dark regions) for illuminant estimation.

B. Results on the synthetic images

To verify the reliability of our proposed hypothesis, we evaluated our model on a set of 100 synthetic images with multiple illuminants (Fig.10 (a)). These images were randomly selected from the Color Checker dataset [43], [57] and each one was re-rendered under at least four different illuminants. Table II shows the performance of our BU-MCC and other single illuminant hypothesis based CC algorithms. Note that in all the tables of this paper, "DN (do nothing)" means that no color correction was applied on the color biased images when computing the measures. In addition, we set the optimal parameters for both algorithms of Shade of Grey [18] and Grey Edge [19] to report their best results on this dataset. We can observe that the single illuminant based algorithms perform poorly in estimating of multiple illuminants (e.g., these methods obtain quite similar performance as DN), while the proposed BU-MCC significantly boosts the accuracy of multi-illuminant estimation.

TABLE II: The performance of multiple CC methods on the synthetic multi-illuminant dataset.

Method	Median	Mean
DN	10.4°	10.5°
Grey World (GW)	10.7°	10.9°
White Patch (WP)	10.2°	10.2°
Shade of Grey (SG)	9.9°	10.0°
Grey Edge (GE)	10.9°	10.9°
BU-MCC	$\mathbf{6.5}^\circ$	$\mathbf{7.2}^\circ$

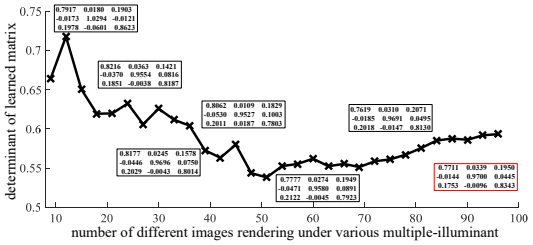


Fig. 12: Visualization of the learned matrix with various input images and illuminants for training.

We conducted a new experiment to visualize how the learned 3x3 matrix returned by top-down processing varies with different input images and illuminants. Geometrically, the determinant of a square matrix can be viewed as the scaling factor of the linear transformation described by the matrix. Fig.12 shows the determinant of the learned matrix on the synthetic color checker dataset with multiple illuminants, which clearly indicates that the learned matrix of top-down proposal is quite stable under various input images and illuminants. Fig.13 further verifies that the learned matrix (Proposed-TD) within a database could always reduce the angular error (Proposed-BU) under various input images and illuminants. We can see that as the number of synthetic images continues to increase (i.e., the numbers of both the training and test sets are increased), the performance (mean and median angular errors) of our top-down based MCC (TD-MCC) almost keeps stable. This result further grounds that the top-down proposal with a learned matrix could be universal to capture the mean illumination color distribution of a group of more complicated scenes within a database, hence can improve the results of the bottom-up method. It should be noted that in comparison to the varying illumination with sharp boundaries in our synthetic images,

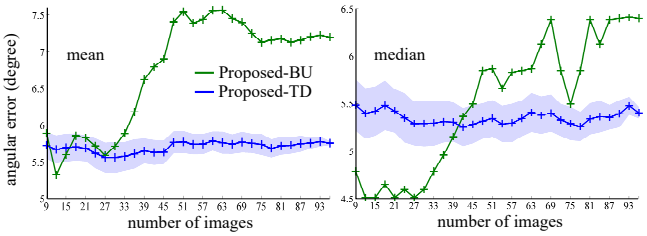


Fig. 13: The influence of the learned matrix on the results of our methods evaluated on the synthetic color checker dataset with the three-fold cross validation.

TABLE III: Results of multiple MCC methods on the Gijsenij dataset [25]. All the results of our TD-MCC in this table and other tables are based on the three-fold cross validation and reported as the average of 20 repetitions.

Gijsenij dataset Method		Laboratory (59)		Outdoor (9)	
		Med	Mean	Med	Mean
DN		18.7^o	20.3^o	3.6^o	4.4^o
Single-Illuminant	GW	14.6^o	14.5^o	8.9^o	8.3^o
	GE2	14.6^o	15.0^o	5.1^o	5.7^o
LSAC [28]	-	12.9^o	12.8^o	7.3^o	7.7^o
LSAC (mask)	-	5.4^o	5.6^o	-	-
Retinex [60]	-	13.2^o	13.2^o	6.6^o	7.3^o
Retinex (mask)	-	3.6^o	3.4^o	-	-
Zhang [48]	-	14.6^o	14.4^o	8.5^o	8.2^o
Zhang (mask)	-	11.5^o	12.3^o	-	-
Gijsenij et al. [25]	GW	11.7^o	-	6.4^o	-
	GE2	12.4^o	-	5.1^o	-
Gu et al. [34]	GE1	-	-	3.3^o	3.3^o
	WP	-	-	3.0^o	3.2^o
MIRF [32]	GW	-	-	10.0^o	10.0^o
	GE1	-	-	4.7^o	7.1^o
FG [35]	-	-	-	2.7^o	3.1^o
GP [7]	M=2	10.9^o	12.2^o	6.8^o	7.1^o
	M=6	10.4^o	12.0^o	7.1^o	6.8^o
	M=10	10.2^o	11.7^o	7.1^o	6.8^o
DS-Net [61]	-	-	-	4.6^o	4.8^o
BU-MCC	-	3.3^o	3.4^o	7.1^o	7.5^o
TD-MCC	-	$\mathbf{2.9^o}$	$\mathbf{3.2^o}$	$\mathbf{2.4^o}$	$\mathbf{3.1^o}$

the multi-illuminant color distribution is relatively simple in the real world [25], thus the synthetic dataset tested here has already represented the most extreme multi-illuminant situation that we can suffer in the real world.

C. Results on three real-world datasets

We further illustrate the discrimination ability of the illuminant map estimated by our bottom-up hypothesis (BU-MCC) and the top-down hypothesis (TD-MCC) on three real-world datasets [25], [32], [36] (Fig.14). As indicated in the previous section, we can observe that the simple segmentation mechanism just based on the pixel luminance actually works quite well on roughly locating the illuminants of various chromaticities. An extreme case is the shadowed image in Fig.14(b), where the illuminant map with both shadowed and unshadowed regions are correctly detected by our bottom-up processing. The reason is that this shadowed image contains quite high luminance dynamic range and thus satisfies the motivation of our method of BU-MCC (i.e., the image areas in the scene with different luminances probably have different illuminants).

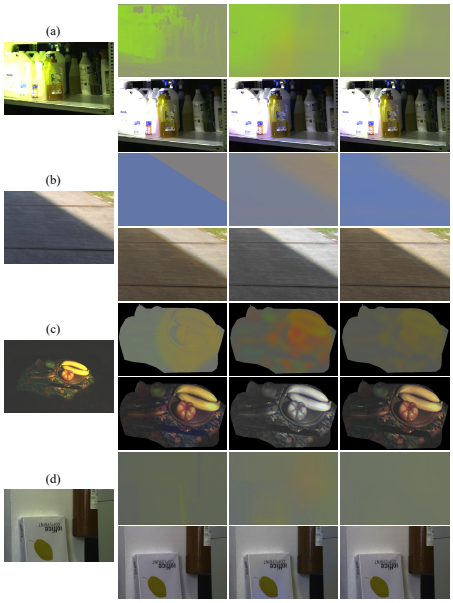


Fig. 14: Multiple results on various real datasets. (a) and (d) Two real-world images from [32]; (b) An outdoor image from [25]; (c) An indoor image from [36]. For each of (a)∼(d), the first column shows the original images, the second to fourth columns list the illuminant maps (top) and the corrected images (bottom) by the ground truth, our BU-MCC and TD-MCC, respectively.

TABLE IV: Performance comparison of our BU-MCC with and without the luminance-dependent operation.

MIRF dataset	Laboratory		Real-world	
	Med	Mean	Med	Mean
lum-dependent	3.45^o	3.65^o	4.04^o	5.06^o
lum-independent	3.59^o	3.96^o	4.15^o	5.07^o
Gijsenij dataset	Laboratory		Outdoor	
	Med	Mean	Med	Mean
lum-dependent	3.34^o	3.43^o	7.08^o	7.51^o
lum-independent	3.47^o	3.55^o	7.67^o	7.75^o

To verify the necessity of applying different DoG filters on the segmented bright and dark regions when estimating the spatially varying illumination, we conducted experiments on the MIRF dataset [25] and the Gijsenij dataset [32] using our BU-MCC without the luminance-dependent filtering. In details, we convolved the images respectively with the On-center and Off-center based DoG filters, and then the two intermediate estimates were directly summed up to form the final illuminant maps. We can see from Table IV that the luminance-dependent operation always provides clear or slight improvement over the processing without the luminance-dependent operation. In particular, the luminance-dependent operation provides quite clear improvement on the real-world subset of Gijsenij dataset (7.08^o vs 7.67^o) which includes the images with many shadows.

However, as indicated by Fig. 2, our BU-MCC method suffers the bias triggered by the large colorful objects since the convolution cannot discriminate the colors of the illuminant from the intrinsic colors of large objects. It should be noted that at present, there is no existing algorithm that can distinguish this ambiguity. Many algorithms [25], [32],

TABLE V: The performance of multiple MCC methods on the MIRF dataset [32]

MIRF dataset Method		Laboratory (58)		Real-world (20)	
		Med	Mean	Med	Mean
DN		10.5^o	10.6^o	8.9^o	9.0^o
Single-Illuminant	GE1	2.8^o	3.2^o	3.9^o	5.3^o
LSAC [28]	-	3.4^o	3.7^o	4.1^o	5.0^o
Retinex [60]	-	4.9^o	5.4^o	4.7^o	5.8^o
Zhang [48]	-	2.7^o	3.2^o	4.4^o	5.2^o
Gijsenij et al. [25]	WP	4.2^o	5.1^o	3.8^o	4.2^o
	GE1	4.2^o	4.8^o	9.2^o	9.1^o
MIRF [32]	WP	2.8^o	3.0^o	3.3^o	4.1^o
	GW	2.6^o	2.6^o	4.5^o	4.9^o
Gu et al. [34]	GE1	3.2^o (Med)		3.6^o (Mean)	
	GW	3.9^o (Med)		4.4^o (Mean)	
FG [35]	-	3.0^o (Med)		3.5^o (Mean)	
GP [7]	M=2	2.5^o	3.1^o	3.3^o	5.7^o
	M=6	2.2^o	2.9^o	3.5^o	5.7^o
	M=10	2.2^o	2.9^o	3.5^o	5.7^o
CNN [62]	-	$\mathbf{2.0^o}$	$\mathbf{2.2^o}$	3.0^o	$\mathbf{3.1^o}$
BU-MCC	-	3.4^o	3.7^o	4.0^o	5.0^o
TD-MCC	-	$\mathbf{2.5^o}$	$\mathbf{2.8^o}$	$\mathbf{2.9^o}$	$\mathbf{3.8^o}$

TABLE VI: The performance of multiple methods on the Bleier dataset [36]

Bleier dataset Method		Laboratory (36)	
		Med	Mean
DN		10.5^o	10.1^o
Single-Illuminant	GE1	14.3^o	13.8^o
LSAC [28]	-	3.0^o	3.3^o
Retinex [60]	-	2.7^o	3.4^o
Zhang [48]	-	4.0^o	4.5^o
Gijsenij et al. [25]	GW	4.7^o	4.9^o
	GE1	14.9^o	14.5^o
MIRF [32]	GW	6.2^o	6.5^o
	WP	7.7^o	7.9^o
Gu et al. [34]	GW	$\mathbf{1.2^o}$	$\mathbf{1.2^o}$
	GE1	3.4^o	3.3^o
FG [35]	-	2.9^o	3.0^o
GP [7]	M=2	5.5^o	5.2^o
	M=6	5.4^o	5.5^o
	M=10	5.4^o	5.5^o
BU-MCC	-	2.9^o	3.4^o
TD-MCC	-	$\mathbf{2.7^o}$	$\mathbf{2.8^o}$

TABLE VII: Inter-dataset based evaluation on three real-world datasets using a matrix learned on a synthetic multi-illuminant dataset as a general top-down constraint.

Gijsenij dataset	Laboratory		Real-world	
	Med	Mean	Med	Mean
DN	18.7^o	20.3^o	3.6^o	4.4^o
GW	14.6^o	14.5^o	8.9^o	8.3^o
GE2	14.6^o	15.0^o	5.1^o	5.7^o
TD-MCC (general prior)	$\mathbf{7.3^o}$	$\mathbf{7.8^o}$	$\mathbf{4.3^o}$	$\mathbf{3.6^o}$
MIRF dataset	Laboratory		Real-world	
	Med	Mean	Med	Mean
DN	10.5^o	10.6^o	8.9^o	9.0^o
GE1	2.8^o	3.2^o	3.9^o	5.3^o
TD-MCC (general prior)	$\mathbf{7.5^o}$	$\mathbf{7.5^o}$	$\mathbf{5.5^o}$	$\mathbf{6.2^o}$
Bleier dataset	Med		Mean	
DN	10.5^o		10.1^o	
GE1	14.3^o		13.8^o	
TD-MCC (general prior)	$\mathbf{6.1^o}$		$\mathbf{6.5^o}$	

[34], [35] using the sophisticated segmentation principle before the illuminant recovery process also cannot avoid such problem. These methods use the results of grey world-based algorithms as the initialization, which will also surfer the ambiguity problem once a local image region doesn't meet the assumption of grey world [35].

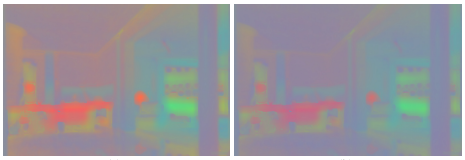
With further constraint by learning a color mapping, our top-down hypothesis (TD-MCC) could suppress the color bias introduced by the intrinsic properties of objects, thus further improve the accuracy of the estimated illuminant map. This observation is further confirmed by the error statistics reported in Table III, V and VI. These results show that the average of the pixel-wised mean and median angular errors between the illuminant map delivered by our BU-MCC approach and the ground truth map are further decreased when the learned color mapping is introduced.

In Table III and V, our approach (TD-MCC) outperforms all the alternatives (including a deep-learning based method [61]) when applied on both the Gijsenij et al. [25] dataset and the MIRF dataset [32]. The only exception is that the CNN-based method proposed by Bianco et al. [62] performs better than our approach on the part of laboratory images of MIRF dataset and obtains lower average error than our approach on the part of real-world images of MIRF dataset (i.e., 3.8^o vs 3.1^o). However, our approach (TD-MCC) slightly outperforms the CNN-based approach on the part of real-world images of MIRF dataset with lower median error (i.e., 2.9^o vs 3.0^o). In Table VI, our algorithm's performance on the Bleier et al. dataset [36] is preceded only by the method proposed by Gu et al. [34]. However, their method appears sensitive to the grey-world methods used (e.g., GW vs GE1).

Table VIII shows that our TD-MCC (fully automatic) also obtains very good performance on both Gijsenij dataset and MIRF dataset. The results show that the use of the fixed set of parameters allows to obtain comparable results that are close to those adopting the optimal parameters by an exhaustive search, which further proves the generalizability of our proposed method for real-world data.

In addition, it should be noted that our method of BU-MCC without the learned mapping also performs quite competitive to other methods. Moreover, we also observed that the methods of Retinex [60], LSAC [28], Zhang [48] and GP [7] without the sophisticated segmentation also obtain acceptable results in comparison to those computationally intensive methods [32], [35]. This means that for MCC, sophisticated segmentation rule adopted by many state-of-the-art algorithms [25], [32], [34], [35] may not be necessary since the spatially varying nature of these methods make them computationally intensive (e.g., the FG needs 125 seconds to process one image [35]). Actually, one recent MCC work [63] has shown that an user guided global correction without explicit segmentation can also work quite well for the problem of two-illuminant estimation.

In Table III, many methods obtain lower performance even than DN on the outdoor images of Gijsenij dataset, which may be caused by several factors: (1) In most circumstances, CC algorithms implicitly assume that there is serious color bias in the images, however, the colors of the two light sources deviate only marginally from white in this dataset. (2) This dataset consists of only



(a) (b)

Fig. 15: The role of general prior in top-down processing. The top-down processing with the general prior (b) could help further suppress the salient colors in the estimated illumination map by the bottom-up processing (a).

TABLE VIII: The results of our TD-MCC (fully automatic) on both Gijsenij dataset and MIRF dataset.

MIRF dataset	Laboratory		Real-world	
	Med	Mean	Med	Mean
TD-MCC (fully automatic)	2.54^o	2.72^o	2.92^o	3.57^o
Gijsenij dataset	Laboratory		Outdoor	
	Med	Mean	Med	Mean
TD-MCC (fully automatic)	2.73^o	2.77^o	2.07^o	1.98^o

nine outdoor images with quite low resolution. (3) The dataset simply classifies the light sources into the shadowed and unshadowed parts, and supposes that the light of unshadowed part to be rendered under standard white light source, which may be different from the real situations [25]. For the indoor images of this dataset, since the ground truth values in the central parts of some images are not labeled accurately [32], [34], [35], we evaluated the MCC algorithms on the masked ground truth maps. Thus, the results evaluated here for some algorithms (e.g., Retinex and LSAC) may be different from the results reported by Gijsenij et al. [25]. Note that these segmentation based methods didn't report their performance [32], [34], [35] on the indoor part of Gijsenij dataset. For the Bleier dataset and MIRF dataset, the methods of Retinex [60], LSAC [28], Zhang [48], and GP [7] actually obtain quite competitive performance in comparison to those image segmentation based algorithms when further considering the simplicity of Retinex, LSAC, Zhang and GP methods.

D. Inter-dataset based cross validation

As a limitation of our top-down stage, the learned transformation matrix is dependent on the property of the training set, which would make the estimated illuminant color distribution of the test image bias towards the mean color of the illumination in the training set.

The underlying prior for the top-down constraint can be described as follows. If the test image is part of a coherent image database, by assuming that the average of illumination of the test image is close to the average illumination of the database, the results of the bottom-up method can be improved by the transformation matrix learned on the training part of this database. It is easily understood that many existing single-illuminant based CC methods cannot well handle the image areas containing large colorful objects, due that the estimated illuminant components in such areas are biased by the intrinsic colors of the objects. In contrast, the proposed database-compensated top-down correction provides chance to resolve this problem.

To well resolve the limitation of our top-down processing mentioned above, we have been working to build a good default top-down mapping that could be generalized to any database. This is the so-called inter-dataset based cross validation and many CC methods suffer this problem [64], [65].

As a first attempt to generalize the top-down mapping, we investigated whether there exists a good default top-down color mapping that could be generalized to estimate the multiple illuminants of a new image. As indicated in Fig.12, we have trained the color mapping matrix on a synthetic multi-illuminant dataset without knowing any database contexts. We selected the learned matrix $C_{3\times3}$ marked by a red rectangle in Fig.12 as an example of default color mapping. Then, this matrix $C_{3\times3}$ was used to correct the result of our bottom-up method for any new images from other datasets. In this way, we treat the learned matrix $C_{3\times3}$ as a general prior of illumination and see whether it could reduce the color-bias of a new image with multiple illuminants. Table VII shows the results on the new images of three real-world datasets using the matrix $C_{3\times3}$ learned on the synthetic dataset.

We can see from Table VII that the performance of top-down processing (i.e., TD-MCC (general prior)) indeed performs better than DN and some single-illuminant based algorithms (e.g., GE1 and GW on the Gijsenij and Bleier datasets). This indicates that the top-down approach could reduce the color bias of a new image with multiple illuminants to some extent, even without knowing any database contexts. The reason is that many single-illuminant based algorithms cannot well handle the image areas containing large colorful objects, due that the estimated illuminant components in such areas are biased by the intrinsic colors of the objects (Fig.15 (a)). In contrast, we found that the top-down processing with the general prior could help suppress these salient intrinsic colors in the estimated illumination map by the bottom-up processing (Fig.15 (b)).

E. Results on web images

The comparison among multiple algorithms on several real-world images taken from the web shown in Fig.16 were only conducted qualitatively since no ground truth map is available for quantitative comparison. Moreover, we adopted the same parameters as that used for the MIRF dataset to produce the corrected results of each algorithm [7], [28], [48], [60].

It is clear that each of the original images shown in Fig.16 (a) is influenced by at least two different illuminants. We can see that BU-MCC can successfully reduce the serious color cast induced by the varying light sources (Fig.16 (b)). Although the true color appearance of these images could be of debate (e.g., our method may overcompensate the color of the sofa in the image of the fourth row), it can be observed that the effect of the varying light sources is much less visible in the corrected images. We further list the results by the top-down processing with the general

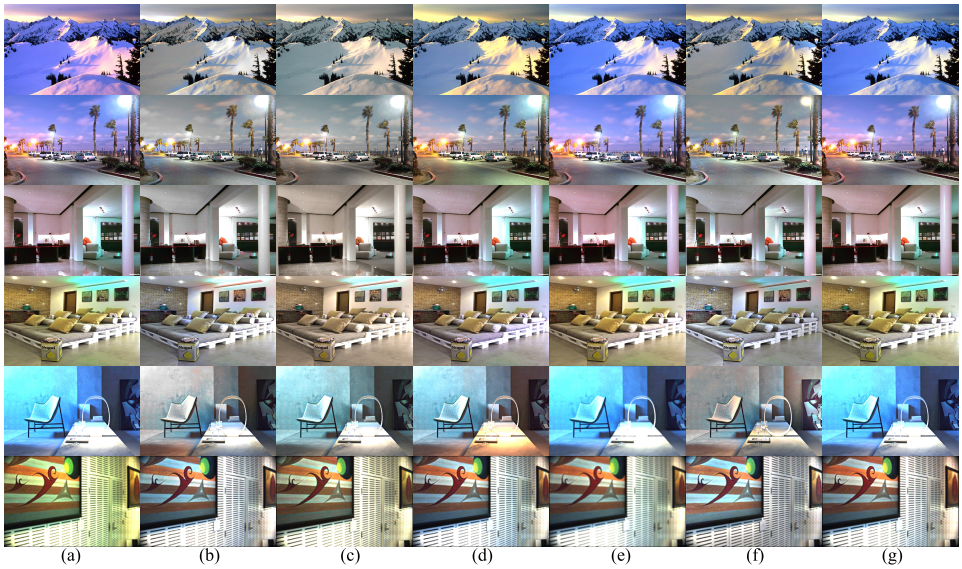


Fig. 16: Results of MCC on some real-world images taken from the web. (a) The real-world images with varying color cast, the results yielded by (b) our bottom-up (BU) method, (c) BU+TD, (d) GW [17], (e) Zhang [48], (f) LSAC [28], and (g) GP [7], respectively.

prior (Fig.16 (c)), which was learned from a synthetic multi-illuminant dataset mentioned above. We can see that the effect of varying light sources is also significantly reduced from the original input, even the corrected images by top-down processing may present some color difference compared to the appearance of bottom-up processing.

Fig.16 (d) shows the results by directly applying the grey world (GW). We can see that the global correction is inappropriate for these images, which even further degrades the color quality of these images. Fig.16 (e), (f), and (g) are the results produced by the MCC methods of Zhang [48], LSAC [28], and GP [7]. These algorithms exhibit various degrees of ability of removing the color cast in local regions. However, the global or local color cast is still quite obvious in their outputs. LSAC is the one that most approaches the performance of our method. This is within our expectation since our method (BU-MCC) performs algorithmically like a diffusion filter as did by LSAC.

F. Further Analysis of the limitations of our method

In Fig.2, we have demonstrated the limitation of our BU-MCC method resulted from the fact that the local convolution cannot well discriminate the colors of the illuminant from the intrinsic colors of large objects. To explicitly examine the performance of our method with shadowless images with stronger surface texture contrast, we conducted new experiments in Fig.17, in which the first row is from the real-world dataset [25] that shows an almost shadowless scene with quite uniform illumination and strong color contrast, and the second row shows a synthetic image with multiple illuminants and very strong surface texture contrast. Our visual system could easily percept that there is clear variation of different light source colors in these two scenes regardless of the stronger surface texture contrast of the second scene or the more uniform illumination of the first scene.

We observed that our BU-MCC using the luminance-dependent clustering tends to group the areas with strong surface texture contrast into distinct regions (i.e., the regions labeled by the red crosses), which are actually rendered under the same illuminant. This indicates that our BU-MCC may fail for the images with strong surface texture contrast, since our BU-MCC performs worse in locating the illuminant for these areas and hence results in high angular errors. However, as indicated in Fig.2, the high errors are also partly resulted from the bias introduced by convolving the intrinsic colors of large objects. Furthermore, our TD-MCC with the learned high level mapping can suppress these salient intrinsic colors in the illumination map estimated by the BU-MCC, which further reduces the errors to some extent.

V. Discussion and Conclusion

The promising performance of our method is primarily grounded on three sides: (1) The intrinsic luminance distinction between the bright and dark areas in an image gives quite useful cues to locate the illuminants with various chromaticities. (2) With the unbalanced DoG filters, our bottom-up hypothesis based method can simultaneously extract different orders of image components (e.g., local edges and local patches), which are proven to be very effective in locally encoding the illuminants and have been adopted by many single-illuminant CC methods [2], [5], [12], [19]. (3) The global color mapping can well suppress the color shift of the estimated illuminants induced by the intrinsic colors of large objects.

However, we also observed in Fig.17 that our BU-MCC method may perform poorly in correctly locating the illuminants for the images with strong surface texture contrast. Considering the fact that the difference in luminance between two locations of a natural scene remains strongly associated with the difference in color of the surfaces [66], [67], an useful future direction is to investigate the benefits

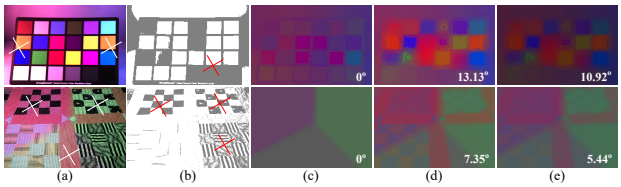


Fig. 17: Results of our BU-MCC and TD-MCC on two example images. The first row is from the real-world dataset [25] that shows an almost shadowless scene with quite uniform illumination and strong color contrast, and the second row shows a synthetic image with multiple illuminants and very strong surface texture contrast. (a) The original image; (b) The locations with different light source colors estimated by our BU-MCC; (c) The ground truth illuminant map; (d) The illuminant map estimated by our BU-MCC; (e) The illuminant map further corrected by our TD-MCC. The white crosses in (a) roughly point out the true locations with different light source colors. The red crosses in (b) roughly point out the areas with strong surface texture contrast.

of developing the filtering mechanisms that are conditioned on the information of lightness, which may be estimated using an intrinsic image decomposition method instead of luminance-dependent filtering. This may further improve the performance of the proposed BU-MCC for the scenes without substantial spatial variation in illumination.

REFERENCES

[1] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline," *Signal Processing Magazine, IEEE*, vol. 22, no. 1, pp. 34–43, 2005.

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

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

[4] J. M. Kraft and D. H. Brainard, "Mechanisms of color constancy under nearly natural viewing," *Proceedings of the National Academy of Sciences*, vol. 96, no. 1, pp. 307–312, 1999.

[5] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li, "Color constancy using double-opponency," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 1973–1985, 2015.

[6] J. Golz and D. I. MacLeod, "Influence of scene statistics on colour constancy," *Nature*, vol. 415, no. 6872, pp. 637–640, 2002.

[7] K.-F. Yang, S.-B. Gao, Y.-J. Li, and Y. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2254–2263.

[8] S. Gao, K. Yang, C. Li, and Y. Li, "A color constancy model with double-opponency mechanisms," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 929–936.

[9] 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.

[10] E. H. Land and J. McCann, "Lightness and retinex theory," *JOSA*, vol. 61, no. 1, pp. 1–11, 1971.

[11] S. M. Courtney, L. H. Finkel, and G. Buchsbaum, "Network simulations of retinal and cortical contributions to color constancy," *Vision Research*, vol. 35, no. 3, pp. 413–434, 1995.

[12] A. Akbarinia and C. A. Parraga, "Colour constancy beyond the classical receptive field," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 9, pp. 2081–2094, 2018.

[13] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner, "Memory modulates color appearance," *Nature neuroscience*, vol. 9, no. 11, pp. 1367–1368, 2006.

[14] A. Vandenbroucke, J. Fahrenfort, J. Meuwese, H. Scholte, and V. Lamme, "Prior knowledge about objects determines neural color representation in human visual cortex," *Cerebral Cortex*, vol. 26, no. 4, pp. 1401–1408, 2016.

[15] K. R. Gegenfurtner, M. Bloj, and M. Toscani, "The many colours of the dress," *Current Biology*, vol. 25, no. 13, pp. R543–R544, 2015.

[16] G. D. Finlayson, M. S. Drew, and B. V. Funt, "Diagonal transforms suffice for color constancy," in *Computer Vision, 1993. Proceedings., Fourth International Conference on*. IEEE, 1993, pp. 164–171.

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

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

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

[20] S. Gao, W. Han, K. Yang, C. Li, and Y. Li, "Efficient color constancy with local surface reflectance statistics," in *European Conference on Computer Vision*. Springer, 2014, pp. 158–173.

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

[22] B. Li, W. Xiong, W. Hu, B. Funt, and J. Xing, "Multi-cue illumination estimation via a tree-structured group joint sparse representation," *International Journal of Computer Vision*, vol. 117, no. 1, pp. 21–47, 2016.

[23] Y. Hu, B. Wang, and S. Lin, "Fc 4: Fully convolutional color constancy with confidence-weighted pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4085–4094.

[24] J. T. Barron and Y.-T. Tsai, "Fast fourier color constancy," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[25] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 697–707, 2012.

[26] M. A. Hussain and A. S. Akbari, "Color constancy algorithm for mixed-illuminant scene images," *IEEE Access*, vol. 6, pp. 8964–8976, 2018.

[27] K. Barnard, G. Finlayson, and B. Funt, "Color constancy for scenes with varying illumination," *Computer vision and image understanding*, vol. 65, no. 2, pp. 311–321, 1997.

[28] M. Ebner, "Color constancy using local color shifts," in *European Conference on Computer Vision*. Springer, 2004, pp. 276–287.

[29] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, 2008, p. 70.

[30] W. Xiong and B. Funt, "Color constancy for multiple-illuminant scenes using retinex and svr," in *Color and Imaging Conference*, vol. 2006, no. 1. Society for Imaging Science and Technolog, 2006, pp. 304–308.

[31] R. Kawakami, K. Ikeuchi, and R. T. Tan, "Consistent surface color for texturing large objects in outdoor scenes," in *Proceedings of the IEEE international conference on computer vision*, vol. 2. IEEE, 2005, pp. 1200–1207.

[32] S. Beigpour, C. Riess, J. Van de Weijer, and E. Angelopoulou, "Multi-illuminant estimation with conditional random fields," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 83–96, 2014.

[33] J. Van De Weijer and T. Gevers, "Color constancy based on the grey-edge hypothesis," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–722.

[34] L. Gu, C. P. Huynh, and A. Robles-Kelly, "Segmentation and estimation of spatially varying illumination," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3478–3489, 2014.

[35] L. Mutimbu and A. Robles-Kelly, "Multiple illuminant color estimation via statistical inference on factor graphs," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5383–5396, 2016.

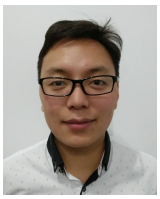
[36] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?" in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 774–781.

[37] V. Mante, R. A. Frazor, V. Bonin, W. S. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nature neuroscience*, vol. 8, no. 12, p. 1690, 2005.

[38] Y. Li, S. Gao, W. Han, R. Li, and C. Li, "Local regions with normal brightness contribute more to color constancy," in *I-PERCEPTION*, vol. 5, no. 4. PION LTD 207 BRONDESBURY PARK, LONDON NW2 5JN, ENGLAND, 2014, pp. 275–275.

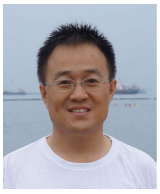
[39] M. G. Bloj, D. Kersten, and A. C. Hurlbert, "Perception of three-dimensional shape influences colour perception through mutual illumination," *Nature*, vol. 402, no. 6764, pp. 877–879, 1999.
[40] M. S. Drew, H. R. V. Joze, and G. D. Finlayson, "The zeta-image, illuminant estimation, and specularity manipulation," *Computer Vision and Image Understanding*, vol. 127, pp. 1–13, 2014.
[41] Y. Wang and Y. Luo, "Color constancy using bright-neutral pixels," *Journal of Electronic Imaging*, vol. 23, no. 2, pp. 023 011–023 011, 2014.
[42] C. Fredembach and G. Finlayson, "Bright chromagenic algorithm for illuminant estimation," *Journal of Imaging Science and Technology*, vol. 52, no. 4, pp. 40 906–1, 2008.
[43] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2008, pp. 1–8.
[44] M. Olkkonen, T. Hansen, and K. R. Gegenfurtner, "Color appearance of familiar objects: Effects of object shape, texture, and illumination changes," *Journal of Vision*, vol. 8, no. 5, pp. 13–13, 2008.
[45] C. Witzel, H. Valkova, T. Hansen, and K. R. Gegenfurtner, "Object knowledge modulates colour appearance," *i-Perception*, vol. 2, no. 1, pp. 13–49, 2011.
[46] D. H. Foster and S. M. Nascimento, "Relational colour constancy from invariant cone-excitation ratios," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 257, no. 1349, pp. 115–121, 1994.
[47] B. Pearce, S. Crichton, M. Mackiewicz, G. D. Finlayson, and A. Hurlbert, "Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations," *PloS one*, vol. 9, no. 2, p. e87989, 2014.
[48] X.-S. Zhang, S.-B. Gao, R.-X. Li, X.-Y. Du, C.-Y. Li, and Y.-J. Li, "A retinal mechanism inspired color constancy model," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1219–1232, 2016.
[49] K. Chen, X.-M. Song, and C.-Y. Li, "Contrast-dependent variations in the excitatory classical receptive field and suppressive nonclassical receptive field of cat primary visual cortex," *Cerebral cortex*, vol. 23, no. 2, pp. 283–292, 2013.
[50] G. C. DeAngelis, R. D. Freeman, and I. Ohzawa, "Length and width tuning of neurons in the cat's primary visual cortex," *Journal of Neurophysiology*, vol. 71, no. 1, pp. 347–374, 1994.
[51] 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.
[52] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vision research*, vol. 5, no. 12, pp. 583–601, 1965.
[53] J. R. Cavanaugh, W. Bair, and J. A. Movshon, "Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons," *Journal of neurophysiology*, vol. 88, no. 5, pp. 2530–2546, 2002.
[54] A. D. Winkler, L. Spillmann, J. S. Werner, and M. A. Webster, "Asymmetries in blue–yellow color perception and in the color of the dress," *Current Biology*, vol. 25, no. 13, pp. R547–R548, 2015.
[55] G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proceedings of the IEEE international conference on computer vision*. IEEE, 2013, pp. 1904–1911.
[56] R. Gershon, A. D. Jepson, and J. K. Tsotsos, "From [r, g, b] to surface reflectance: Computing color constant descriptors in images." in *IJCAI*, 1987, pp. 755–758.
[57] 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/.
[58] 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.
[59] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *JOSA A*, vol. 31, no. 5, pp. 1049–1058, 2014.
[60] B. Funt, J. McCann, and F. Ciurea, "Retinex in matlab?" *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 48–57, 2004.
[61] W. Shi, C. C. Loy, and X. Tang, "Deep specialized network for illuminant estimation," in *European Conference on Computer Vision*. Springer, 2016, pp. 371–387.
[62] S. Bianco, C. Cusano, and R. Schettini, "Single and multiple illuminant estimation using convolutional neural networks," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4347–4362, 2017.
[63] D. Cheng, A. Abdelhamed, B. Price, S. Cohen, and M. S. Brown, "Two illuminant estimation and user correction preference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 469–477.
[64] S.-B. Gao, M. Zhang, C.-Y. Li, and Y.-J. Li, "Improving color constancy by discounting the variation of camera spectral sensitivity," *JOSA A*, vol. 34, no. 8, pp. 1448–1462, 2017.
[65] Ç. Aytekin, J. Nikkanen, and M. Gabbouj, "A data set for camera-independent color constancy," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 530–544, 2018.
[66] I. Fine, D. I. MacLeod, and G. M. Boynton, "Surface segmentation based on the luminance and color statistics of natural scenes," *JOSA A*, vol. 20, no. 7, pp. 1283–1291, 2003.
[67] A. Chakrabarti, "Color constancy by learning to predict chromaticity from luminance," in *Advances in Neural Information Processing Systems*, 2015, pp. 163–171.

Shao-Bing Gao received his Ph.D. degree from UESTC, Chengdu, China, in 2017. He is currently a tenure-track associate professor in College of Computer Science, Sichuan University. His research interests include biologically inspired vision and image processing.



Yan-Ze Ren received his B.Sc. and Master degree in Biomedical engineering from UESTC in 2013 and 2016. His research interests include image processing.

Ming Zhang received his B.Sc. and Master degree in Biomedical engineering from UESTC in 2015 and 2018. His research interests include visual mechanism modeling and image processing.



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 Technology, UESTC, China. His research interests include visual mechanism modeling and image processing.