# Efficient Color Constancy with Local Surface Reflectance Statistics

Shaobing Gao[1], Wangwang Han[1], Kaifu Yang[1], Chaoyi Li[1,2], and Yongjie Li[1]

[1] University of Electronic Science and Technology of China
[2] Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences
{gao_shaobing,yang_kf}@163.com, cyli@sibs.ac.cn, liyj@uestc.edu.cn

**Abstract.** The aim of computational color constancy is to estimate the actual surface color in an acquired scene disregarding its illuminant. Many solutions try to first estimate the illuminant and then correct the image with the illuminant estimate. Based on the linear image formation model, we propose in this work a new strategy to estimate the illuminant. Inspired by the feedback modulation from horizontal cells to the cones in the retina, we first normalize each local patch with its local maximum to obtain the so-called locally normalized reflectance estimate (LNRE). Then, we experimentally found that the ratio of the global summation of true surface reflectance to the global summation of LNRE in a scene is approximately achromatic for both indoor and outdoor scenes. Based on this substantial observation, we estimate the illuminant by computing the ratio of the global summation of the intensities to the global summation of the locally normalized intensities of the color-biased image. The proposed model has only one free parameter and requires no explicit training with learning-based approach. Experimental results on four commonly used datasets show that our model can produce competitive or even better results compared to the state-of-the-art approaches with low computational cost.

**Keywords:** color constancy, illuminant estimation, reflectance, retina.

## 1 Introduction

To some extent, our visual system can constantly perceive the actual color of an object in a scene disregarding the large differences in illumination, which is called the ability of color constancy [17]. In contrast, the physical color of scenes captured with regular digital cameras or videos may be shifted by the varying external illuminant. Thus, for a robust color-based computer vision, removing the effect of light source color from the color-biased image is of very importance for many applications [1]. To solve this problem, one of the general ways is to estimate the scene illuminant and then utilize it to map the color-biased image to the so-called canonical image under white light source. According to the two steps mentioned above, many computational color constancy algorithms have been proposed (see [17,25,27] for recent overviews). For example, the classical

D. Fleet et al. (Eds.): ECCV 2014, Part II, LNCS 8690, pp. 158–173, 2014.

gamut mapping theory assumes that the distribution of RGB color values of an image captured under a canonical illuminant is a limited set, which is called canonical gamut [16]. So, for a color-biased image taken under an unknown illuminant will produce another gamut too, which is called observed gamut. The aim of gamut mapping is to compute the transformation that maps the observed gamut to the canonical gamut, and finally, the estimated illuminant could be derived from this transformation. Lately, the derivative structure of image was brought into the gamut mapping theory and achieved more robust color constancy performance than the standard gamut mapping algorithm [24].

Although the gamut mapping based methods have the elegant underlying theory to solve the problem of computational color constancy, the inherent drawbacks of these methods include their complication in implementation and the requirement of appropriate pre-processing [25]. Other typical approaches include learning-based methods [7,12,19], the bayesian color constancy [5,21], exemplar-based method [36], biologically-inspired methods [20,33], high level information-based methods [4,38], and physics-based methods [26,29,34].

Like most computer vision tasks, the problem of color constancy is ill-posed, and in general, most of the existing approaches mentioned above introduce specific assumptions based on, for example, the color distribution of per-pixel [16], filter-based structure of image [8,37], empirical prior distribution of light source color [5,13], etc, to estimate the illuminant. Based on the linear model $\boldsymbol{f}(x) = \int_{\omega} \boldsymbol{S}(\lambda) I(\lambda) C(x, \lambda) d\lambda$ [25], most existing methods put more emphasis on the relationship between the information of illuminant $I(\lambda)$ and the structure of image $\boldsymbol{f}(x)$. In addition, though almost all the existing models employ various sort of a prior on distribution of reflectance in natural scenes, they generally suffer from certain assumption.

For example, the grey-world theory based models assume that the average surface reflectance (or the edge) of natural scenes is statistically achromatic [6,15,37]. This assumption is too strong to well match the statistic diversity of surface reflectance distribution in indoor and outdoor scenes of natural world [10]. As another typical model, white patch [28] assumes that there are points with perfect reflection in natural scenes and those points can be used to estimate the illuminant of environment. Similarly to the gray-world assumption, this hypothesis also can't satisfy the statistic diversity of surface reflectance. For example, it is not always easy to retrieve points with perfect reflection in scenes [25]. The bayesian color constancy usually assumes that the distribution of surface reflectance is independent identically distributed and is normally imitated using Gaussian models. Although the Gaussian distribution is simple for calculating, it does not always accord with the actual situations [21,30,35].

Different from the traditional models mentioned above that directly estimate the illuminant by probing into the relationship between the illuminant and the derivative structure of images (e.g., the edges), or the prior information of illuminant combining with pixel color distribution (e.g., the gamut), we introduce an efficient and simple method for color constancy based on the retinal mechanism and the suitable modeling of the distribution of surface reflectance. Inspired by

the feedback modulation from horizontal cells to the cones in the retina [31], we first normalize each local patch of the image with its local maximum to obtain the so-called locally normalized reflectance estimate (LNRE). We then experimentally find that the global statistics of surface reflectances in an image could be well approached by the global statistics of LNRE. Based on this observation, we do not need to make any assumption about the distribution of the surface reflectances or the property of illuminant. Instead, the illuminant can be directly derived from the ratio between the global summation of per-pixels and the statistic summation of estimated surface reflectance in local regions according to the linear model of image formation [25]. We will also demonstrate that both the grey-world and white patch algorithms can be integrated into the proposed color constancy framework.

The rest of this paper is organized as follows. In section 2, the proposed model is described in details. Section 3 provides extensive performance evaluations of the proposed algorithm on four commonly used datasets. Finally, some discussion and concluding remarks are given in Section 4. The source codes of the work are available at `http://www.neuro.uestc.edu.cn/vccl/home.html`

## 2 Color Constancy with Local Surface Reflectance Estimation

Suppose that the scene is illuminated by a single light source, e.g., the outdoor daylight illuminant. Based on the common form of the linear imaging equation, The captured image values $f(x) = [f_R(x), f_G(x), f_B(x)]^T$ depend on the color of the light source $I(\lambda)$, the surface reflectance $C(x, \lambda)$ and the camera sensitivity function $S(\lambda) = [S_R(\lambda), S_G(\lambda), S_B(\lambda)]^T$, where $x$ is the spatial coordinate and $\lambda$ is the wavelength of the light (e.g., [25]):

$$f_c(x) = \int_\omega S_c(\lambda) I(\lambda) C(x, \lambda) d\lambda \tag{1}$$

where the integral is taken over the visible spectrum $\omega$ and $c \in \{R, G, B\}$ are sensor channels.

Among various color constancy algorithms, the color transform of image induced by the illuminant can be well approximated by a diagonal transform [11,14,25,27]. Thus, based on the diagonal transform assumption in color constancy, Eq (1) can be simplified as

$$f_c(x) = I_c C_c(x) \tag{2}$$

The aim of the color constancy method proposed in this work is to estimate the color of the light source $I_c$, $c \in \{R, G, B\}$. Given the color-biased image values of $f_c(x)$ and the illuminant estimate $I_c$, color constancy can be achieved by a transformation of $f_c(x)$ to the one appearing to be taken under a canonical (often white) light source. In general, $I_c$ and $C_c(x)$ in Eq (2) are unknown and hence, given only the image values of $f_c(x)$, the estimation of $I_c$ is an ill-posed inverse problem that cannot be solved without further assumptions [25].

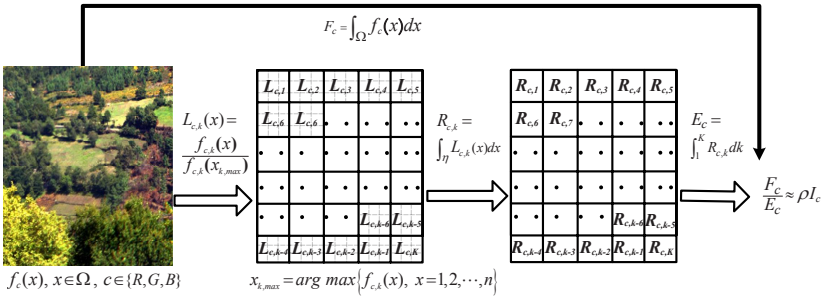


**Fig. 1.** The flowchart explaining the computational steps of the proposed algorithm

According to the linear multiplication relationship between the illuminant $I_c$ and surface reflectance $C_c(x)$ in Eq (2), if we can first obtain the rough estimate of $C_c(x)$, the illuminant $I_c$ could be simply derived by dividing $f_c(x)$ by the reflectance estimate $C_c(x)$ [25]. In this work, we are not to estimate the illuminant directly as did in most of the linear model based color constancy algorithms; in contrast, our philosophy is that we first estimate the reflectance $C_c(x)$ roughly based on certain appropriate assumption.

Fig. 1 shows the general flowchart of the proposed method, and the details will be described in the following sections.

### 2.1 Surface Reflectance Estimation in Local Region

The reflected lights entering into the eyes are first processed by the photoreceptors like cones (for color) and rods (for luminance). Then, at each spatial location, a horizontal cell (with large receptive field) pools a population of cones outputs within a relatively large region and then modulates the cones sensitivities via feedback [31]. In the following we simplify this retinal processing mechanism to modulate each local region by dividing by its local maximum.

We divide the full two dimensional (2D) area of the image $f_c(x)$ into $K$ non-overlapped patches with equal size. Let

$$L_{c,k}(x) = \frac{f_{c,k}(x)}{f_{c,k}(x_{k,max})} \tag{3}$$

where $x_{k,max}$ is the spatial coordinate of the maximum intensity pixel within the k-th local region:

$$x_{k,max} = arg\ max\{f_{c,k}(x), x = 1, 2, \cdots n\} \tag{4}$$

Clearly, $L_{c,k}(x)$ is the intensity value of the pixel at $x$ normalized by the local maximum of the intensity values within the k-th local region. Considering

$$L_{c,k}(x) = \frac{I_c C_{c,k}(x)}{I_c C_{c,k}(x_{k,max})} = \frac{C_{c,k}(x)}{C_{c,k}(x_{k,max})} \tag{5}$$

we call $L_{c,k}(x)$ the locally normalized reflectance estimate (LNRE) of the pixel at $x$ in channel $c \in \{R, G, B\}$. Note that Eq (3) will be used to compute $L_{c,k}(x)$ for the next step of illuminant estimation, since the true values of $C_c(x)$ in Eq (5) are unknown.

### 2.2 Illuminant Estimation

In this section, we will show that under certain assumption, the estimate of illuminant $I_c$ could be accurately derived by computing the ratio of $\int_\Omega f_c(x)dx$ to $\int_1^K \int_\eta L_{c,k}(x)dxdk$, where $\Omega$ denotes the full area of the image, $K$ is the total number of the local regions within the image. $\eta$ represents the area of the k-th local region. Note that in our model, $K$ is the only one free parameter, since for a given image with certain size, the value of $\eta$ is determined by $K$. Let $FR_c$ represent the ratio given by

$$FR_c = \frac{\int_\Omega f_c(x)dx}{\int_1^K \int_\eta L_{c,k}(x)dxdk} \tag{6}$$

For a given color-biased image $f_c(x)$, we can easily obtain $FR_c, c \in \{R, G, B\}$. To exploit what information could be derived from $FR_c$, let us substitute Eq (2) into Eq (6), and we obtain:

$$FR_c = I_c \frac{\int_\Omega C_c(x)dx}{\int_1^K \int_\eta L_{c,k}(x)dxdk} = I_c \frac{T_c}{E_c} \tag{7}$$

where

$$T_c = \int_\Omega C_c(x)dx \tag{8}$$

$$E_c = \int_1^K \int_\eta L_{c,k}(x)dxdk \tag{9}$$

where $T_c$ represents the global summation of the true surface reflectance, and $E_c$ denotes the global summation of locally normalized reflectance estimate (LNRE) computed with Eq (3).

To exploit the relationship of these two global summation measures, we used the Gehler-Shi dataset containing 568 linear images [32], the SFU indoor dataset including 321 linear images [2], the SFU HDR dataset contaning 105 linear images [12,18], and the SFU grey ball dataset contaning 11346 nonlinear images [9] for quantitative analysis. The known illuminant of each color-biased image provided in the datasets was used to correct the color-biased image to obtain the ground truth image, and the values of the ground truth image are equivalent to $C_c(x)$, which were then used to compute $T_c$.

The first column of Fig. 2 plots in three separate channels the 568 ratios of these two global summation measures, i.e., $\frac{T_c}{E_c}$ with $c \in \{R, G, B\}$, and each scatter point corresponds to one image in Gehler-Shi dataset. Similarly, the second column plots the 321 ratios for SFU indoor dataset, the third column

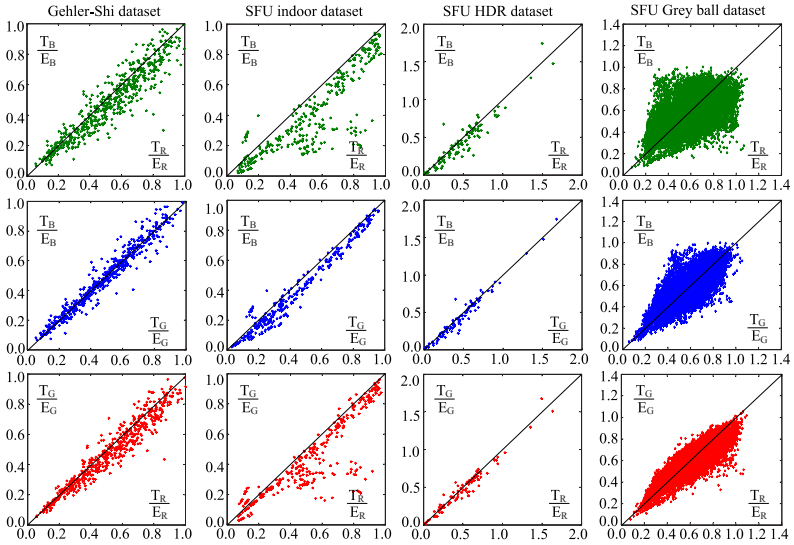


**Fig. 2.** Each point in the scatter plots represents the ratio of the statistic summation of real surface reflectance of an image (e.g. $T_B$) to the statistic summation of estimated surface reflectance in local regions of the image (e.g. $E_B$) in three separate color channels. These plots indicate that most of these scatter points in the three channels are closely aligned along the diagonal lines (the solid line).

plots the 105 ratios for SFU HDR dataset, and the fourth column plots the 11346 ratios for SFU grey ball dataset. Quite interestingly, from Fig. 2 we can clearly find that statistically, these scatter points are linearly correlated between any two channels of $\frac{T_c}{E_c}$, $c \in \{R, G, B\}$. More importantly, we can also obviously see that most of these scatter points in the three channels are closely aligned along the diagonal lines. These indicate that for most of the images, we have

$$\frac{T_R}{E_R} \approx \frac{T_G}{E_G} \approx \frac{T_B}{E_B} \approx \rho \tag{10}$$

where $\rho$ is a constant, which may be different for different images.

Equation (10) indicates that by choosing appropriate value of free parameter $K$ involved in $E_c, c \in \{R, G, B\}$, we have great chance to get an accurate estimate of the direction of vector $[T_R, T_G, T_B]$ denoting the global statistics (i.e., summation) of true surface reflectance of a scene from the easily computed $[E_R, E_G, E_B]$ denoting the summation of surface reflectance estimate in local regions, with a scaling factor (i.e., $\rho$) between the magnitudes of the two vectors.

Substituting Eq (10) into (7), we now have:

$$FR_c \approx \rho I_c \tag{11}$$

Based on this equation, we can estimate the color of the illuminant as

$$I_c \approx \frac{1}{\rho} FR_c \tag{12}$$

Given a color-biased image $f_c(x), c \in \{R, G, B\}$ as input, $FR_c$ can be easily computed using Eq (6). $\rho$ acts as a scaling factor that depends on the scene viewed. Considering that $\rho$ is identical for three color channels $c \in \{R, G, B\}$ according to Eq (10), we do not need to know the exact value of $\rho$, since $\rho$ will be cancelled by taking the normalized form of $I_c$ as the final estimate of the illuminant.

Equation (10) describes the key idea underlying the method proposed in this work, which will be discussed in details in Section 4. Note that basically, $E_c$ is equivalent to the white-patch estimate of reflectance in a local region. However, our model is far more than a robust version of white patch, since we just use it as a rough reflectance estimate to further compute illuminant color using our empirical observation based rule.

## 3 Experimental Results

The proposed model was compared with various methods on foure typical datasets [2,9,18,32]. The methods considered for comparison include: Do Nothing (DN), inverse-intensity chromaticity space (IICS) [34], Grey World (GW) [6], White Patch (WP) [28], 1st-Grey Edge (GE1) and 2nd-Grey Edge (GE2) [37], Shades of Grey (SG) [15], general Grey World (GG), Bayesian (Bayes) [21], Regression (SVR) [19], automatic color constancy algorithm selection (AAS) [3], using natural image statistics (NIS) [23], spatial correlations (SC)[25], spatio-spectral statistics (SS (with reg.)) [8], pixel-based gamut mapping (GM(pixel)) [16], edge-based gamut mapping (GM(edge)) [24,25], Exemplar-based [36], Corrected-moment based (CM) [12]. Recently, the standard survey paper [25] reported the results for most of the methods mentioned above, here we directly use the result data from [22,25] for analysis and comparison except that results of SS and CM are from Ref [8] and [12], respectively.

The generally employed angular error is chosen as error metric [25].

$$\varepsilon = cos^{-1}\Big(\big(I_e \cdot I_g\big)/\big(\|I_e\| \cdot \|I_g\|\big)\Big) \tag{13}$$

Where $I_e \cdot I_g$ is the dot product of the estimated illuminant $I_e$ and the ground truth illuminant $I_g$, $\|\cdot\|$ is the Euclidean norm of a vector. Besides the commonly used measure of median angular error, we also reported the measure of mean, trimean, best-25%, and worst-25% for more comprehensive comparison.

### 3.1 Real-World Image Set

Gehler-Shi dataset [21,32] 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.

**Table 1.** Performance statistics of various methods on the Gehler-Shi dataset [32]

| Methods | | Median | Mean | Trimean | Best-25% | Worst-25% |
|---|---|---|---|---|---|---|
| | DN | 13.6° | 13.7° | 13.5° | 10.4° | 17.2° |
| Physics-based | IICS | 13.6° | 13.6° | 13.4° | 9.5° | 18.0° |
| | GW | 6.3° | 6.4° | 6.3° | 2.3° | 10.6° |
| (Static) | WP | 5.7° | 7.5° | 6.7° | 1.5° | 16.1° |
| low-level | GE2 | 4.5° | 5.3° | 4.9° | 1.9° | 10.0° |
| statistics | SG | 4.0° | 4.9° | 4.4° | 1.1° | 10.2° |
| -based | GG | 3.5° | 4.7° | 4.0° | 1.0° | 10.2° |
| | SVR | 6.7° | 8.1° | 7.4° | 3.3° | 14.9° |
| | SC | 5.1° | 5.9° | 5.5° | 2.4° | 10.8° |
| | Bayes | 3.5° | 4.8° | 4.1° | 1.3° | 10.5° |
| | AAS | 3.3° | 4.5° | 3.7° | 0.9° | 10.1° |
| Learning | NIS | 3.1° | 4.2° | 3.5° | 1.0° | 9.2° |
| -based | SS(reg.) | 3.0° | 3.6° | 3.2° | 0.9° | 7.4° |
| | GM(edge) | 5.6° | 6.7° | 6.0° | 2.0° | 13.5° |
| | GM(pixel) | 2.4° | 4.2° | 3.3° | 0.5° | 11.2° |
| | Exemplar | 2.3° | 2.9° | 2.5° | 0.8° | 6.0° |
| | CM(19 Edge-Moments) | 2.0° | 2.8° | – | – | – |
| Proposed | | 2.6° | 3.4° | 2.9° | 0.8° | 7.2° |

**Table 2.** Mean computation time taken to compute the illuminant for per image, which was averaged on 100 test images for one dataset with repeated 20 times using MATLAB codes. Here, we list the mean computation time of image from two typical datasets: one dataset with small image size of $360 * 240$ pixels [9] and one dataset with large image size of $2041 * 1359$ pixels [32]. Note that the time taken by GM(pixel) is only for the process of test, without including the process of training. Computer used here is Intel Core2, 2.53GHZ with 2.0G RAM.

| Dataset | GM(pixel) | GE2 | Proposed |
|---|---|---|---|
| SFU grey ball dataset [9] | 1.44(s) | 0.27(s) | 0.22(s) |
| Gehler-Shi dataset [32] | 9.21(s) | 12.90(s) | 1.36(s) |

The results for various algorithms on this database are listed in Table 1. Fig. 3 shows examples of indoor and outdoor images corrected by different algorithms.

It can be seen from Table 1 that the performance of the proposed method almost arrives at (in terms of median angular error and best-25%) or beyond (in terms of other measures) the best performance of the state-of-the-art learning-based algorithms, e.g., the GM (pixel), GM (edge), SS(reg.), NIS, Exemplar, and CM. Note that the methods of GM (pixel), Exemplar, and CM perform better in terms of median angular error of 2.4°, 2.3°, and 2.0° respectively. In addition, Exemplar also performs better in terms of other measures.

However, compared to the complex implementation of the learning-based GM(pixel), Exemplar, and CM, our model is quite simpler, which endows our method with a remarkable advantage in saving computational cost. Table 2 shows

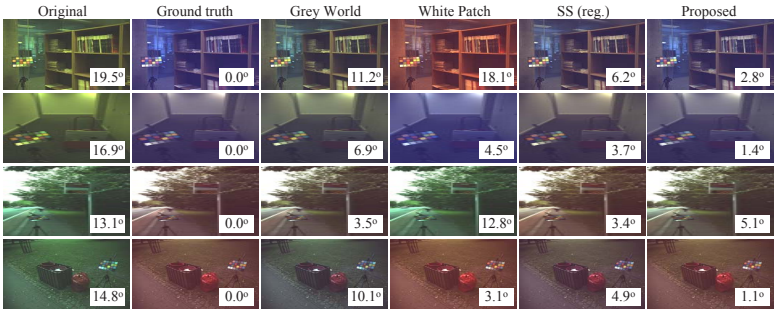


**Fig. 3.** Results of several algorithms on Gehler-Shi dataset, the angular error is given on the lower right corner of image

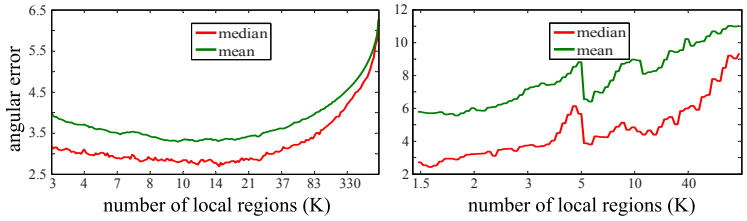


**Fig. 4.** The influence of parameter K of the proposed algorithm on the measure of median and mean angular error for two datasets. **Left**: real world Gehler-Shi dataset [32], **Right**: SFU indoor image dataset [2]. According to the two figures, for a real world image dataset, the suitable setting of parameter is $8 \leq K \leq 37$. For an indoor image dataset, the suitable setting of parameter is $1.5 \leq K \leq 4$. More discussion about parameter setting is given in the last section. Note the patch number $K = image\_size/patch\_size$, which may result in non-integer $K$. The curves were obtained with a step of $5 * 5$ pixels for patch size and the patches have no overlap.

that our model just takes about a seventh of the time needed by GM (pixel). Note that without the source codes of Exemplar and CM, we did not list the computation time of per image taken by them.

By visually comparing the performance of the proposed method with other algorithms on some examples shown in Fig. 3, we can find that for the indoor or outdoor scenes with various distributions of surface reflectance, it is difficult for a certain assumption (e.g., grey-world or white-patch) based model to perform well on all the scenes. For example, for the scene of grassland shown in the last row of Fig. 3, the distribution of surface reflectance is obviously not achromatic, which does not meet the grey-world assumption. In contrast, the proposed approach achieves the more robust performance with a suitable value for the only one parameter (i.e. $K$ in Eq(6)) on the whole dataset.

Fig. 4 demonstrates the influence of the only one parameter $K$ on the performance of the proposed method on two datasets: a real world Gehler-Shi dataset [32] and SFU indoor image dataset [2].

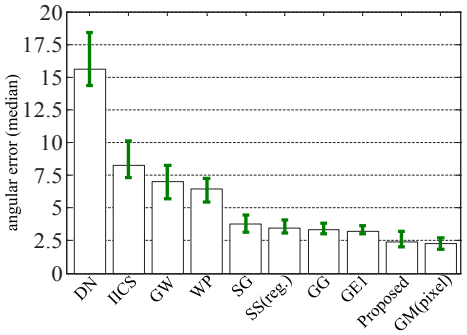


**Fig. 5.** Angular error (median) for different algorithms on SFU indoor dataset plotted with a 95% confidence interval

### 3.2 An Indoor Image Dataset in Laboratory

SFU lab dataset [2] contains 321 available images of 31 different objects captured with calibrated camera under 11 different lights in laboratory. Table 3 reports the results on this dataset for various algorithms (all methods are with optimal parameters). Similar to the results on the first dataset, we find that in terms of median angular error, our model performs better than both static- and learning-based algorithms on this dataset, but slightly worse than the method of GM(pixel) and CM. However, there is almost no significant difference between our model and GM(pixel) in terms of the measure of median angular error (Fig. 5). Note that though CM (9 Edge-moments) obtains the best performance, it required a special preprocessing on this dataset [12]. Fig. 6 presents some examples of images from SFU indoor dataset. As analyzed previously, no certain assumption can perform best on all of the scenes with various distribution of surface reflectance, even just in a simple indoor environment. However, the proposed approach still achieved competitive performance on those images comparing with the more complicated algorithm (Gamut (pixel)). We questioned

**Table 3.** Performance of various methods on the SFU indoor dataset

| Methods | Median | Mean | Trimean | Best-25% | Worst-25% |
|---|---|---|---|---|---|
| DN | 15.6° | 17.3° | 16.9° | 3.6° | 32.4° |
| IICS | 8.2° | 15.5° | 12.0° | 2.2° | 40.0° |
| GW | 7.0° | 9.8° | 8.1° | 0.9° | 23.3° |
| WP | 6.5° | 9.1° | 7.6° | 1.8° | 20.9° |
| SG | 3.7° | 6.4° | 5.0° | 0.6° | 16.4° |
| SS(with reg.) | 3.5° | 5.6° | 4.8° | 1.2° | 12.8° |
| GG | 3.3° | 5.4° | 4.1° | 0.5° | 13.7° |
| GE1 | 3.2° | 5.6° | 4.2° | 1.0° | 14.0° |
| GM(pixel) | 2.3° | 3.7° | 2.7° | 0.5° | 9.3° |
| CM(9 Edge-Moments) | 2.0° | 2.6° | – | – | – |
| Proposed | 2.4° | 5.7° | 4.1° | 0.5° | 15.0° |

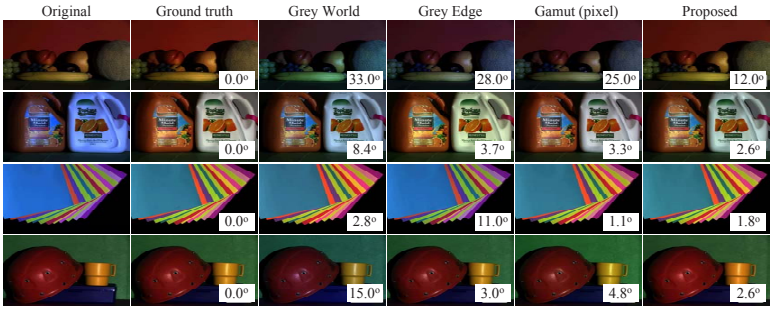


**Fig. 6.** Results of several algorithms on SFU indoor dataset, the angular error is given on the lower right corner of image

**Table 4.** Performance statistics of various methods on the SFU grey ball image dataset

| Methods | Median | Mean | Worst-25% |
|---|---|---|---|
| GW | 7.0° | 7.9° | 15.2° |
| DN | 6.7° | 8.3° | 18.7° |
| GM(pixel) | 5.8° | 7.1° | 14.7° |
| IICS | 5.6° | 6.6° | 13.3° |
| WP | 5.5° | 6.8° | 14.7° |
| GE2 | 4.9° | 6.1° | 13.2° |
| Proposed | 5.1° | 6.0° | 11.9° |

whether the high worst-25% of our model (15°) on this set was caused by black background of some images. To answer it, we tried to ignore the black background by using different levels of thresholding, and the performance varied as: median: 2.4° → 2.3°, mean: 5.7° → 5.5°, worst-25%: still 15.0°. This indicates that the high worst-25% might not be mainly caused by black background.

### 3.3 SFU Grey Ball Image Datasets

SFU grey-ball image dataset [9] contains 11346 nonlinear images. The images in this dataset have been processed with gamma-correction, automatic white balance, and other unknown post-processing in the camera (the image is more post-processed by camera, the more unknown factors affect the image and thus, the harder the illuminant estimation becomes [12]). So, the images in this datasets are no longer to meet the linear model of image formation described by Eq (1) [25]. Before this experiment, the grey sphere in the images, which were originally used to compute the ground truth, were masked in order to get a fair and full evaluation to the performance of a specific model.

Table 4 reports the measures on the entire dataset for several available static algorithms and gamut mapping algorithm (again all algorithms are with optimal parameters). From Table 4 we can see that among the multiple methods considered, the proposed algorithm produces the as best performance as that of GE2 on this dataset, in terms of all the three measures.

**Table 5.** Performance statistics for the SFU HDR image dataset. Here dash in table means that the result for that dataset was not reported by their authors.

| Methods | Median | Mean | Worst-25% |
|---|---|---|---|
| DN | 14.7° | 15.1° | 19.5° |
| GW | 7.4° | 8.0° | 15.1° |
| MaxRGB (post-blur) | 3.9° | 6.3° | – |
| SG | 3.9° | 5.7° | 12.9° |
| GE | 3.8° | 6.0° | 13.8° |
| CM(9 Edge-Moments) | 2.7° | 3.5° | – |
| Proposed | 2.9° | 4.7° | 10.8° |

However, we also notice that none of the methods can provide a very low measure of median angular error (comparing with other three linear image datasets). One of the main possible reasons is that the unknown non-linear effects embedded in this dataset may seriously degrade the performances of most linear image formation based methods. For example, from Table 4 we observed that the median angular error of the grey world (GW) algorithm on this dataset is even higher than that of DN (i.e., do nothing on the original color biased images). For that reason, it could be expected that the linear image formation based methods would produce better performance when the camera dependent images are first transformed to the device independent raw images [25]. Note that on this dataset, the learning-based GM(pixel) also provides relatively poor performance.

### 3.4 SFU HDR Dataset

The SFU HDR dataset has been recently collected by [18] and includes 105 high dynamic range linear images with indoor and outdoor scenes. Also, a color checker were placed in the scenes for recording the overall light source color (again the color checker was masked during testing the illuminant estimation algorithms).

For comprehensive comparison, we also reported the results of GW, SG, and GE on this dataset by running the matlab codes downloaded from [22,37] with optimal parameters. The results of MaxRGB (post-blur) and Corrected-Moment (CM) are directly from paper [12]. Table 5 reports the performance statistics for various algorithms on this dataset.

On this dataset, our model also performed well comparing with other static-based models in terms of various measures. Although the performance of CM is slightly better than our method (in median angular error), CM needs to iteratively train a corrected matrix for every dataset with ground truth.

## 4 Discussion and Conclusion

We proposed in this work a new idea to estimate the illuminant of a scene, based on the experimental finding that the ratio of the global summation of

the true surface reflectance to the global summation of locally estimated surface reflectance in a scene is almost achromatic. This so-called achromatic-ratio-mean observation inspired us to estimate the illuminant just by computing the ratio of the global summation of the observed intensities to the global summation of the locally normalized intensities of the color-biased image. There is only one free parameter in the proposed model, without the need for pre-learning. Extensive experimental tests on four commonly used datasets (three datasets with linear images and one with nonlinear images) indicated that the proposed model can produce quite competitive or even better results compared to the state-of-the-art approaches. In addition, our model shows significant advantage in terms of computational efficiency.

The main condition under which our model works well is the achromatic-ratio-mean observation described by Fig. 2 and Eq (10). Based on the derivation of equations, it is clear that we would obtain an accuracy of 100% if the ratio $\rho$ defined by Eq (10) was always a constant, which is impossibly realized due to the complexity of scenes. Higher correlation between $T_c$ and $E_c$ makes $\rho$ closer to be constant. Extensive empirical evidence on multiple datasets with large sizes and various scenes has already supported the visible correlation. For a given dataset, this requirement of high correlation could be matched as much as possible by choosing a suitable value for the only one free parameter $K$. In general, the distributions of the real reflectances in outdoor scenes are more complicated than the reflectance distributions in indoor scenes, since most of the indoor scenes are generally composed of only several simple color regions. Thus, based on the experimental results (Fig. 4), we suggest the bigger $K$ values for outdoor images, because the estimated surface reflectance from a larger number of small local regions is able to better imitate the complex properties of surface reflectances in natural scenes. In contrast, for indoor images in laboratory, the model with smaller $K$ values could work well.

Computationally, our model novelly combines the ideas of max-RGB (or White-Patch, WP) and Grey-World (GW), i.e., estimate each pixels reflectance in each local patch by normalizing the intensities with the local max (the idea of WP), then sum (or average) the estimated reflectances across the whole scene (the idea of GW). Then we estimate the light source color by computing the ratio of summation of observed intensities to the summation of estimated reflectances. Such idea underlying our model is quite different from others. For example, though the shade-of-grey (SG) also seems to balance the advantages of WP and GW, the difference is, SG averages intensities (different from the estimated reflectances in our model) that have been processed by emphasizing pixels with larger intensities (different from the WP based normalizing in each patch in our model). In addition, our achromatic-ratio-mean assumption is basically different from the GW theory assuming that the average reflectance in a scene under a neutral light source is achromatic. However, we can mathematically demonstrate that both the GW and WP assumptions are the extreme cases of Eq (10) (see Appendix).

Besides the achromatic-ratio-mean observation that was substantially validated across different datasets, a deeper insight explaining the good performance of our model is as follows. By applying the idea of WP in local patches, the influences of the pixels unexpected for WP (e.g. bright non-specularity or noise) could be limited within local regions, and meanwhile, the robustness of reflectance estimates could be enhanced by utilizing more normal bright patches.

Actually, a suitable value of parameter $K$ may build a good balance between the gray world assumption [6] and the white patch assumption [28], which relaxes the strong limitation to the statistics distribution of surface reflectance in natural world assumed by these two theories, and hence, to better satisfy the diversity of the statistics distribution of surface reflectance in natural world.

## 5 Appendix

In the following, we mathematically derive how both grey world and white patch algorithms can be basically integrated into the proposed framework. When setting $K = 1$ in Eq (9), which is equivalent to set the whole input image as the only single local region, then Eq (9) is rewritten as

$$E_c = \int_1^K \int_\eta L_{c,k}(x)dxdk = \int_\Omega L_{c,k}(x)dx = \int_\Omega \frac{C_c(x)}{C_c(x_{max})}dx \tag{14}$$

$c \in \{R, G, B\}$, where $x_{max}$ is the spatial coordinate of the maximum of the intensity values within whole image. Thus, combining Eqs (8), (14), and (10) together, we have

$$C_R(x_{max}) \approx C_G(x_{max}) \approx C_B(x_{max}) \approx \rho \tag{15}$$

This is just equivalent to the white patch algorithm (e.g. max-RGB) [28], which assumes the maximum reflectance in a scene is achromatic and then taking the maximum value of every channel of image as the estimated illuminant.

Similarly, when setting $K = N$ in Eq (9), where $N$ is the pixel number of the full two dimensional area ($\Omega$) of the image. This is equivalent to shrink each local region as small as one pixel, then Eq (9) is equal to

$$E_c = \int_1^K \int_\eta L_{c,k}(x)dxdk = \int_\Omega dk \tag{16}$$

where $\int_\eta L_{c,k}(x)dx = 1$, since each local region contains only one pixel. Similarly, combining Eqs (8), (16), and (10) together, we get

$$\frac{\int_\Omega C_R(x)dx}{\int_\Omega dk} \approx \frac{\int_\Omega C_G(x)dx}{\int_\Omega dk} \approx \frac{\int_\Omega C_B(x)dx}{\int_\Omega dk} \quad \approx \rho \tag{17}$$

This is equivalent to the gray world algorithm [6], which hypothesizes that the average surface reflectance in a scene is achromatic and then computing the mean of every channel of image as the estimated illuminant.

**Acknowledgments.** We thank all anonymous reviewers for their thoughtful comments. This work was supported by the Major State Basic Research Program (#2013CB329401), the Natural Science Foundations of China (#61375115, #61075109, #91120013, #31300912), and the Doctoral Support Program of University of Electronic Science and Technology of China. The work was also supported by the 111 Project (#B12027) and PCSIRT (#IRT0910) of China.

## References

1. Barnard, K., Martin, L., Coath, A., Funt, B.: A comparison of computational color constancy algorithms. ii. experiments with image data. IEEE Transactions on Image Processing 11(9), 985–996 (2002)
2. Barnard, K., Martin, L., Funt, B., Coath, A.: A data set for color research. Color Research & Application 27(3), 147–151 (2002)
3. Bianco, S., Ciocca, G., Cusano, C., Schettini, R.: Automatic color constancy algorithm selection and combination. Pattern Recognition 43(3), 695–705 (2010)
4. Bianco, S., Schettini, R.: Color constancy using faces. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 65–72. IEEE (2012)
5. Brainard, D.H., Freeman, W.T.: Bayesian color constancy. JOSA A 14(7), 1393–1411 (1997)
6. Buchsbaum, G.: A spatial processor model for object colour perception. Journal of the Franklin Institute 310(1), 1–26 (1980)
7. Cardei, V.C., Funt, B., Barnard, K.: Estimating the scene illumination chromaticity by using a neural network. JOSA A 19(12), 2374–2386 (2002)
8. Chakrabarti, A., Hirakawa, K., Zickler, T.: Color constancy with spatio-spectral statistics. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(8), 1509–1519 (2012)
9. Ciurea, F., Funt, B.: A large image database for color constancy research. In: Color and Imaging Conference, vol. 2003, pp. 160–164. Society for Imaging Science and Technology (2003)
10. Ebner, M.: Color constancy, vol, vol. 6. Wiley. com (2007)
11. Ebner, M.: Color constancy based on local space average color. Machine Vision and Applications 20(5), 283–301 (2009)
12. Finlayson, G.: Corrected-moment illuminant estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1904–1911 (2013)
13. Finlayson, G.D.: Color in perspective. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(10), 1034–1038 (1996)
14. Finlayson, G.D., Hordley, S.D.: Color constancy at a pixel. JOSA A 18(2), 253–264 (2001)
15. Finlayson, G.D., Trezzi, E.: Shades of gray and colour constancy. In: Color and Imaging Conference, vol. 2004, pp. 37–41. Society for Imaging Science and Technology (2004)
16. Forsyth, D.A.: A novel algorithm for color constancy. International Journal of Computer Vision 5(1), 5–35 (1990)
17. Foster, D.H.: Color constancy. Vision Research 51(7), 674–700 (2011)
18. Funt, B., Shi, L.: The rehabilitation of maxrgb. In: Color and Imaging Conference, vol. 2010, pp. 256–259. Society for Imaging Science and Technology (2010)

19. Funt, B., Xiong, W.: Estimating illumination chromaticity via support vector regression. In: Color and Imaging Conference, vol. 2004, pp. 47–52. Society for Imaging Science and Technology (2004)
20. Gao, S., Yang, K., Li, C., Li, Y.: A color constancy model with double-opponency mechanisms. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), pp. 929–936 (2013)
21. Gehler, P.V., Rother, C., Blake, A., Minka, T., Sharp, T.: Bayesian color constancy revisited. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
22. Gijsenij, A.: Color constancy: research website on illuminant estimation, `http://colorconstancy.com/` (accessed from)
23. Gijsenij, A., Gevers, T.: Color constancy using natural image statistics and scene semantics. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(4), 687–698 (2011)
24. Gijsenij, A., Gevers, T., Van De Weijer, J.: Generalized gamut mapping using image derivative structures for color constancy. International Journal of Computer Vision 86(2-3), 127–139 (2010)
25. Gijsenij, A., Gevers, T., Van De Weijer, J.: Computational color constancy: Survey and experiments. IEEE Transactions on Image Processing 20(9), 2475–2489 (2011)
26. Gijsenij, A., Gevers, T., Van De Weijer, J.: Improving color constancy by photometric edge weighting. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5), 918–929 (2012)
27. Hordley, S.D.: Scene illuminant estimation: past, present, and future. Color Research & Application 31(4), 303–314 (2006)
28. Land, E.H., McCann, J.J., et al.: Lightness and retinex theory. Journal of the Optical society of America 61(1), 1–11 (1971)
29. Lee, H.C.: Method for computing the scene-illuminant chromaticity from specular highlights. JOSA A 3(10), 1694–1699 (1986)
30. Nascimento, S., Ferreira, F.P., Foster, D.H.: Statistics of spatial cone-excitation ratios in natural scenes. JOSA A 19(8), 1484–1490 (2002)
31. Schiller, P.H.: Parallel information processing channels created in the retina. Proceedings of the National Academy of Sciences 107(40), 17087–17094 (2010)
32. Shi, L., Funt, B.: Re-processed version of the gehler color constancy dataset of 568 images, `http://www.cs.sfu.ca/~colour/data/` (accessed from)
33. Spitzer, H., Semo, S.: Color constancy: a biological model and its application for still and video images. Pattern Recognition 35(8), 1645–1659 (2002)
34. Tan, R.T., Nishino, K., Ikeuchi, K.: Color constancy through inverse-intensity chromaticity space. JOSA A 21(3), 321–334 (2004)
35. Tsin, Y., Collins, R.T., Ramesh, V., Kanade, T.: Bayesian color constancy for outdoor object recognition. In: Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. I–1132 (2001)
36. Vaezi, J.H., Drew, M.: Exemplar-based colour constancy and multiple illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence (2013)
37. Van De Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. IEEE Transactions on Image Processing 16(9), 2207–2214 (2007)
38. Van De Weijer, J., Schmid, C., Verbeek, J.: Using high-level visual information for color constancy. In: International Conference on Computer Vision (ICCV), pp. 1–8 (2007)