On Finding Gray Pixels

Yanlin Qian[1,3], Joni-Kristian Kämäräinen[1], Jarno Nikkanen[2], Jiri Matas[1,3]
[1]Computing Sciences, Tampere University   [2]Intel Finland
[3]Center for Machine Perception, Czech Technical University in Prague

Abstract

We propose a novel grayness index for finding gray pixels and demonstrate its effectiveness and efficiency in illumination estimation. The grayness index, GI in short, is derived using the Dichromatic Reflection Model and is learning-free. GI allows to estimate one or multiple illumination sources in color-biased images. On standard single-illumination and multiple-illumination estimation benchmarks, GI outperforms state-of-the-art statistical methods and many recent deep methods. GI is simple and fast, written in a few dozen lines of code, processing a 1080p image in ~ 0.4 seconds with a non-optimized Matlab code.

1. Introduction

The human eye has the ability to adapt to changes in imaging conditions and illumination of scenes. The well-established computer vision problem of *color constancy*, CC in short, is trying to endow consumer digital cameras with the same ability. With "perfect" color constancy, finding a gray pixel is not a problem at all – just checking whether the RGB values are equal. However, given a color-biased image, detecting gray pixels, i.e. pixels observing an achromatic surface, is a hard and ill-posed problem – imagine a white piece of paper illuminated with a cyan light source; or is it a cyan paper under white light? On the other hand, "perfect" gray pixels in an image indicate that color constancy is satisfied. Thus, from this point onward, we treat *finding gray pixels* and *color constancy* as equivalent problems (see also Fig. 1). Color constancy problem arises in many computer vision and image processing applications, such as computational photography, intrinsic image decomposition, semantic segmentation, scene rendering, object tracking, *etc*. [18].

For decades, *learning-free methods*, the classical approach to color constancy, have relied on the assumption that the illumination color is constant over the whole scene and can therefore be estimated by global processing [6, 2, 38, 17, 19, 41, 12]. This approach has the advantage of being independent to the acquisition device, since
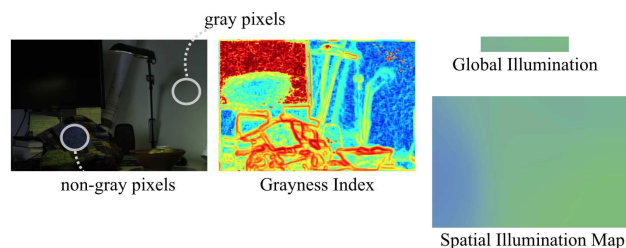

Figure 1: Gray and non-gray image pixels (left). The Grayness Index (GI) map (middle, blue denotes high grayness value). The global (top right) and spatially-variant illumination color (right) estimated from the GI map.

the illumination properties are estimated on a per-image basis. Recently, state-of-the-art *learning-based methods*, including convolutional neural networks (CNNs), have consistently outperformed statistical methods when validated on specific datasets [9, 25, 22, 24, 29]. We argue that learning-based methods depend on the assumption that the statistical distribution of the illumination and/or scene content is similar in training and test images. In other words, learning-based methods assume that imaging and illumination conditions of a given image can be inferred from previous training examples, thus becoming heavily dependent on the training data [21].

In this paper, we focus on the learning-free approach. For a practical example, consider the case when a user retrieves a linear-RGB (gamma corrected) image from the web and wants to correct its colors. In this scenario, in which the used CC method has never seen images from that camera, illumination estimation and color correction must be performed without strong assumptions on the imaging device or the captured scene. We experimentally show that in this setting, learning-free methods show more promising and robust results as compared to learning-based methods. As a result, there is a great need for learning-free approaches that are insensitive to parameters such as the camera and imaging process of captured images.

In most camera sensors, gray pixels are rendered gray in linear-RGB image under standard neutral illumination, making *grayness* a potential measure to estimate the color

of incident illumination. We adopt Shafer's Dichromatic Reflection Model (DRM) [33] to develop a novel grayness index (GI), which allow ranking all image pixels according to their grayness. The appealing points are: (i) GI is simple and fast to compute; (ii) it has a clear physical meaning; (iii) it can handle specular highlights to some extend (from qualitative comparison); (iv) it allows pixel-level illumination estimation; (v) it provides consistent prediction across different cameras. Comprehensive results on single-illumination and multi-illumination color constancy datasets show that GI outperforms the state-of-the-art learning-free methods and achieves state-of-the-art in the cross-dataset setting.

2. Related Work

Consider image I captured using a linear digital camera sensor, with black level corrected and no saturation. In the dichromatic reflection model, the pixel value at (x, y) under one global illumination source can be modeled as [33]:

$$I_i^{(x,y)} = \gamma_b^{(x,y)} \int F_i(\lambda) L(\lambda) R_b^{(x,y)}(\lambda) d\lambda$$
$$+ \gamma_s^{(x,y)} \int F_i(\lambda) L(\lambda) R_s^{(x,y)}(\lambda) d\lambda, \qquad (1)$$

where $I_i^{(x,y)}$ is the pixel value at (x, y), $L(\lambda)$ the global light spectral distribution, $F_i(\lambda)$ the sensor sensitivity, $i = \{R, G, B\}$ for trichromatic cameras, and λ the wavelength. The chromatic terms $R_b(\lambda)$ and $R_s(\lambda)$ account for body and surface reflection, respectively, while the achromatic terms γ_b and γ_s are the intensities of the above two types of reflection.

In addition, under the the assumption of narrow spectral response $F_i(\lambda)$, Eq. 1 is further simplified to [3]:

$$I^{(x,y)} = W^{(x,y)} \circ L + V^{(x,y)} \circ L, \qquad (2)$$

where $\circ$ denotes Hadamard Product and,

$$W^{(x,y)} = [\gamma_b^{(x,y)} R_{b,R}^{(x,y)}, \gamma_b^{(x,y)} R_{b,G}^{(x,y)}, \gamma_b^{(x,y)} R_{b,B}^{(x,y)}]^T,$$
$$V^{(x,y)} = [\gamma_s^{(x,y)} R_{s,R}^{(x,y)}, \gamma_s^{(x,y)} R_{s,G}^{(x,y)}, \gamma_s^{(x,y)} R_{s,B}^{(x,y)}]^T,$$
$$L = [F_R L_R, F_G L_G, F_B L_B]^T, \qquad (3)$$

where the {R,G,B} subscripts represent the corresponding parts of the spectrum that intersect with F_i. Eq. 2 shows the formation of a pixel value in image I corresponding to a location in the scene exhibiting body W and surface reflection V, under a camera-captured global light L.

The goal of CC is to estimate L in order to recover W, given I. Based on the strategy used for solving this problem, we divide color constancy methods into two categories: learning-based, learning-free methods.

Learning-based Methods [9, 25, 22, 24, 29, 31, 32] aim at building a model that relates the captured image I and the sought illumination L from extensive training data. Among the best-performing state-of-the-art approaches, the CCC method [3] discriminatively learns convolutional filters in a 2D log-chroma space. This framework was subsequently accelerated using the Fast Fourier Transform on a chroma torus [4]. Chakrabarti *et al.* [8] leverage the normalized luminance for illumination prediction by learning a conditional chroma distribution. DS-Net [35] and FC⁴ Net [28] are two deep learning methods, where the former chooses an estimate from multiple illumination guesses using a two-branch CNN architecture and the later addresses local estimation ambiguities of patches using a segmentation-like framework. Learning-based methods achieve great success in predicting pre-recorded "ground-truth" illumination color fairly accurately, but heavily depending on the same cameras and/or scenes being in both training and test images (see Sec. 3 and Sec. 4.2). The Corrected-Moment method [14] can also be considered as a learning-based method as it needs to train a corrected matrix for each dataset.

Learning-free Methods estimate the illumination by making prior assumptions about the local or global regularity of the illumination and reflectance. The simplest such method is *Gray World* [7] that assumes that the global average of reflectance is achromatic. The generalization of this assumption by restricting it to local patches and higher-order gradients has led to more powerful statistics-based methods, such as White Patch [6], General Gray World [2], Gray Edge [38], Shades-of-Gray [17] and LSRS [19], among others [12].

Physics-based Methods [37, 15, 16], estimate illumination from the understanding of the physical process of image formation (*e.g.* the Dichromatic Model), thus being able to model highlights and inter-reflections. Most physics-based methods estimate illumination based on intersection of multiple dichromatic lines, making them work well on toy images and images with only a few surfaces but often failing on natural images [16]. The latest physics-based method relies on the longest dichromatic line segment assuming that the Phong reflection model holds and an ambient light exists [39]. Although our method is based on the Dichromatic Model, we classify our approach as *statistical* since the core of the method is finding gray pixels based on some observed image statistics. We refer readers to [26] for more details about physics-based methods.

The Closest Methods to GI are Xiong *et al.* [40] and Gray Pixel by Yang *et al.* [41]. Xiong *et al.* [40] method searches for gray surfaces based on a special LIS space, but it is camera-dependent. Gray Pixel [41] is closest to our work and is therefore outlined in details in Sec. 3.

3. Grayness Index

We first review the previous Gray Pixel [41] (derived from the Lambertian model) in the context of dichromatic reflection model (DRM).

3.1. Gray Pixel in [41]

Yang *et al.* [41] claims that gray pixels can be sought by a set of constraints. However, their formulation often identifies gray pixels that clearly are color pixels. This phenomenon has been noticed, but not properly analyzed. Herein we analyze GP using DRM and point out the potential failure cases of the original formulation.

Assuming narrow band sensor, Eq. 1 simplifies to:

$$I_i^{(x,y)} = \gamma_b^{(x,y)} F_i L_i R_{b,i}^{(x,y)} + \gamma_s^{(x,y)} F_i L_i R_{s,i}^{(x,y)},$$
$$i \in \{R,G,B\}. \quad (4)$$

Then, following Yang *et al.* [41], we apply $log(\cdot)$ and a local contrast operator $C\{\cdot\}$ (Laplacian of Gaussian, see Sec. 4 for more details) on the both sides, and obtain

$$C\{\log(I_i^{(x,y)})\} = C\{\log(F_i L_i R_{b,i}^{(x,y)})\}$$
$$+ C\left\{\log\left(\gamma_b^{(x,y)} + \gamma_s^{(x,y)} \frac{R_{s,i}^{(x,y)}}{R_{b,i}^{(x,y)}}\right)\right\}. \quad (5)$$

If $\gamma_s = 0$ (means no surface reflection), we obtain:

$$C\{\log(I_i^{(x,y)})\} = C\{\log(\gamma_b^{(x,y)} R_{b,i}^{(x,y)})\}. \quad (6)$$

If $\gamma_s \neq 0$, due to the interaction between γ_b and $\gamma_s \frac{R_{s,i}}{R_{b,i}}$ in Eq. 5, those colored pixels can be wrongly identified as gray pixels. Central to GP is that, when $\gamma_s = 0$, a non-uniform intensity casting on a homogeneous gray surface can induce the same amount of "contrast" in each channel. Varying intensity of light may result from the geometry between surface and illumination (shading) and that among different surfaces (occlusion). In order to resolve this problem we adopt the Dichromatic Reflection Model, exploring another path to identify gray pixels in a more complex environment.

3.2. Grayness Index using Dichromatic Reflection Model

For simplicity, in the sequel we will drop the superscripts (x,y), as all operations are applied in a local neighborhood centered at (x,y). We first calculate the residual of the red channel and luminance in log space and then apply local contrast operator $C\{\cdot\}$ to Eq. 5 as:

$$C\{\log(I_R) - \log(|I|)\} = C\{\log(F_R L_R) + \log(\gamma_b R_{b,R} + \gamma_s R_{s,R})\}$$
$$- C\{\log(F_R L_R(\gamma_b R_{b,R} + \gamma_s R_{s,R}) + F_G L_G(\gamma_b R_{b,G} + \gamma_s R_{s,G})$$
$$+ F_B L_B(\gamma_b R_{b,B} + \gamma_s R_{s,B}))\}, \quad (7)$$
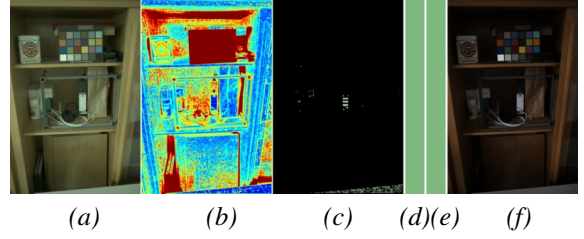

(a)   *(b)*   *(c)*   *(d)(e)*   *(f)*

Figure 2: Finding gray pixels. (a) input image. (b) computed grayness index GI. darker blue indicates higher degree of grayness. (c) the $N\%$ most gray pixels rendered using the corresponding pixel color (greenish) in (a). (d) estimated illumination color. (e) ground truth color. (f) corrected image using (d).

where $|I|$ denotes the luminance magnitude $(I_R + I_G + I_B)$.

In this case, the neutral interface reflection (NIR) assumption establishes that, for gray pixels, we have that $R_{j,R} = R_{j,G} = R_{j,B} = \bar{R}_j$ with $j \in \{s,b\}$ [30]. In this case, Eq. 7 simplifies to:

$$C\{\log(I_R) - \log(|I|)\} = C\{\log(F_R L_R) + \log(\gamma_b \bar{R}_b + \gamma_s \bar{R}_s)\}$$
$$- C\{\log((F_R L_R + F_G L_G + F_B L_B)(\gamma_b \bar{R}_b + \gamma_c \bar{R}_s))\}. \quad (8)$$

In a small local neighborhood, the casting illumination and sensor response can be assumed constant [41], such that $C\{\log(F_R L_R)\} = 0$ and $C\{\log((F_R L_R + F_G L_G + F_B L_B)\} = 0$, leading to:

$$C\{\log(I_R) - \log(|I|)\} = C\left\{\log \frac{\gamma_b \bar{R}_b + \gamma_c \bar{R}_s}{\gamma_b \bar{R}_b + \gamma_c \bar{R}_s}\right\} \overset{\text{gray}}{=} 0. \quad (9)$$

Eq. (9) is a necessary yet not a sufficient condition for gray pixels. A more restrictive requirement for the detection of gray pixels is given by extending Eq. 9 to one more color channel (using all channels in redundant, the spectral response of R and B rarely overlap in sensors) as:

$$C\{\log(I_R) - \log(|I|)\} = C\{\log(I_B) - \log(|I|)\} = 0. \quad (10)$$

From Eq. (7), we define the grayness index w.r.t. $I(x,y)$ as:

$$GI(x,y) = \|[C\{\log(I_R) - \log(|I|)\},$$
$$C\{\log(I_B) - \log(|I|)\}]\|, \quad (11)$$

where $\|\cdot\|$ refers to the $\ell 2$ norm. The smaller the GI is, the more likely the corresponding pixel is gray.

In addition, we impose a restriction on the local contrast to ensure that a "small" GI value comes from grey pixels in varying intensity of light, not a flatten color patch (no spatial cues), written as:

$$C\{I_i\} > \epsilon, \forall i \in \{R,G,B\}, \quad (12)$$

where ϵ is a small contrast threshold.

The process of computing GI is in two steps:

1. Compute a preliminary GI map using Eq. 11.

2. Discard pixels in GI with no spatial cues using Eq. 12. To weaken the effect of isolated gray pixels mainly due to camera noise, GI map is averaged in 7×7 window.

For illustration, Fig. 2 shows a flowchart of computing GI and its predicted illumination.

The proposed GI differs from GP in two important aspects. At first, it utilizes a novel mechanism to detect gray pixels based on a more complete image formation model that leads to different formulation. Secondly, the proposed GI works without selectively enhancing bright and dark pixels according to their luminance. In other words, the proposed GI does not weaken the influence of dark pixels.

3.3. GI Application in Color Constancy

Color Constancy is a direct application of gray pixels. Here we describe two pipelines to compute illumination color from gray pixels: *single illumination* and *multi-illumination* pipelines.

When a scene contains only one global illumination, the pipeline is straightforward. As shown in Fig. 2, after ranking all image pixels according to their GI, the global illumination is computed as the average of top $N\%$ pixels.

Given a scene cast by more than one light source, the desired output is a pixel-wise illumination map. Similar to [41], the GI map is first computed and then followed by a K-means clustering of the top $N\%$ pixels into preset number of M clusters. Now, the averaging is applied on cluster basis, giving a illumination vector L_m for the cluster m. The final spatial illumination map is computed using:

$$L_i(x, y) = \sum_{m=1}^{M} \omega_m L_m^i, i \in \{R, G, B\} \qquad (13)$$

where ω_m controls the connection between the pixel $I(x, y)$ to the cluster m, written as:

$$\omega_m = e^{-\frac{D_m}{2\sigma^2}} / \sum_{n=1}^{M} e^{-\frac{D_n}{2\sigma^2}}, \qquad (14)$$

where D_m is the Euclidean distance from the pixel to the centroid of cluster m. Eq. 14 encourages nearby pixels to share a similar illumination.

4. Evaluation

We evaluated GI in two color constancy settings: (1) single-illumination estimation, where the illumination of the whole captured scene is described by a single chroma vector for the red, green and blue channels; and (2) multi-illumination estimation, where in each scene there are two or more effective illuminants. Moreover, we conducted experiments in the cross-dataset setting which is very challenging for the learning-based methods.

Datasets
- The Gehler-Shi Dataset [34, 22]: single illumination, 568 high dynamic linear images, 2 cameras [1].
- The NUS 8-Camera Dataset [12]: single illumination, 1,736 high dynamic linear images, 8 cameras (see Table 2 for the camera list).
- MIMO Dataset [5]: multi-illumination, 78 linear images, 58 laboratory images and 20 harder wild images.

Single-illumination Experiment Settings
- The local contrast operator in Eq. 11 is the Laplacian of Gaussian filter of the size 5 pixels.
- The proportion of the best gray pixels used for color estimation is set to $N = 0.1\%$.
- The contrast threshold is set to $\epsilon = 1e-4$

These parameters were selected based on preliminary grid search (see Section 4.3) and remained fixed for all experiments with the both datasets.

Multi-illumination Experiment Settings
- The local contrast operator and the contrast threshold are the same as in the single-illuminant experiment.
- The proportion of chosen pixels is set to $N = 10.0\%$ as more illuminants are involved.
- The tested number of clusters M were 2,4 and 6.

Dataset Bias of Learning-based Methods When trained with images from a single data that is divided to training and testing sets, the state-of-the-art learning-based methods (*e.g.* [4]) outperform the best learning-free methods by a clear margin. However, it is important to know how these values are biased since images in the training and test sets often share the same camera(s) and same scenes. It can happen that a learning-based method overfits to the camera and scene features that are not available in the real case. To investigate the dataset bias, we evaluated several top performing learning-based methods in the cross-dataset setting, where the methods were trained on one dataset (*e.g.*, the Gehler-Shi) and tested with another. This allows evaluating the performance of learning-based algorithms for unseen cameras and scenes.

Performance Metric As the standard tool in color constancy papers we adopted the angular error $\arccos(\frac{L^T \hat{L}}{\|L\| \|\hat{L}\|})$ between the estimated illumination $\hat{L}$ and ground-truth L as the performance metric. Obtained results are summarized in Table 1 and discussed in Sections 4.1 and 4.2.

4.1. Single-dataset Setting

Single-dataset setting is the most common setting in related works, allowing extensive pre-training using k-fold

---

[1]cameras: Canon 1D, Canon 5D

Table 1: Quantitative Evaluation of CC methods. All values correspond to angular error in degrees. We report the results of the related work in the following order: 1) the cited paper, 2) Table [1] and Table [2] from Barron *et al.* [4, 3] considered to be up-to-date and comprehensive, 3) the color constancy benchmarking website [23]. We left dash on unreported results. In (a) results of learning-based methods worse than ours are marked in gray. The training time and testing time are reported in seconds, averagely per image, if reported in the original paper.

(a) single-dataset setting

	Gehler-Shi					NUS 8-camera				
	Mean	Median	Trimean	Best 25%	Worst 25%	Mean	Median	Trimean	Best 25%	Worst 25%
Learning-based Methods (camera-known setting)										
Edge-based Gamut [25]	6.52	5.04	5.43	1.90	13.58	4.40	3.30	3.45	0.99	9.83
Pixel-based Gamut [25]	4.20	2.33	2.91	0.50	10.72	5.27	4.26	4.45	1.28	11.16
Bayesian [22]	4.82	3.46	3.88	1.26	10.49	3.50	2.36	2.57	0.78	8.02
Natural Image Statistics [24]	4.19	3.13	3.45	1.00	9.22	3.45	2.88	2.95	0.83	7.18
Spatio-spectral (GenPrior) [9]	3.59	2.96	3.10	0.95	7.61	3.06	2.58	2.74	0.87	6.17
Corrected-Moment[1](19 Edge) [14]	3.12	2.38	2.59	0.90	6.46	3.03	2.11	2.25	0.68	7.08
Corrected-Moment[1](19 Color) [14]	2.96	2.15	2.37	0.64	6.69	3.05	1.90	2.13	0.65	7.41
Exemplar-based [29]*	2.89	2.27	2.42	0.82	5.97	–	–	–	–	–
Chakrabarti *et al.* 2015 [8]	2.56	1.67	1.89	0.52	6.07	–	–	–	–	–
Cheng *et al.* 2015 [13]	2.42	1.65	1.75	0.38	5.87	2.18	1.48	1.64	0.46	5.03
DS-Net (HypNet+SelNet) [35]	1.90	1.12	1.33	0.31	4.84	2.24	1.46	1.68	0.48	6.08
CCC (dist+ext) [3]	1.95	1.22	1.38	0.35	4.76	2.38	1.48	1.69	0.45	5.85
FC[4] (AlexNet) [28]	**1.77**	1.11	1.29	0.34	**4.29**	2.12	1.53	1.67	0.48	4.78
FFCC [4]	1.78	**0.96**	**1.14**	**0.29**	4.62	**1.99**	**1.31**	**1.43**	**0.35**	**4.75**
GI	3.07	1.87	2.16	0.43	7.62	2.91	1.97	2.13	0.56	6.67

[1] For Correct-Moment [14] we report reproduced and more detailed results by [3], which slightly differs with the original results: mean: 3.5, median: 2.6 for 19 colors and mean: 2.8, median: 2.0 for 19 edges on Gehler-Shi Dataset.
* We mark Exemplar-based method with asterisk as it is trained and tested on a uncorrected-blacklevel dataset.

(b) cross-dataset setting

Training set Testing set	NUS 8-Camera Gehler-Shi					Gehler-Shi NUS 8-Camera					Average runtime (s)	
	Mean	Median	Trimean	Best 25%	Worst 25%	Mean	Median	Trimean	Best 25%	Worst 25%	Train	Test
Learning-based Methods (agnostic-camera setting), Our rerun												
Bayesian [22]	4.75	3.11	3.50	1.04	11.28	3.65	3.08	3.16	1.03	7.33	764	97
Chakrabarti *et al.* 2015 [8] Empirical	3.49	2.87	2.95	0.94	**7.24**	3.87	3.25	3.37	1.34	7.50	–	0.30
Chakrabarti *et al.* 2015 [8] End2End	3.52	2.71	2.80	0.86	7.72	3.89	3.10	3.26	1.17	7.95	–	0.30
Cheng *et al.* 2015 [10]	5.52	4.52	4.79	1.96	12.10	4.86	4.40	4.43	1.72	8.87	245	0.25
FFCC [4]	3.91	3.15	3.34	1.22	7.94	3.19	2.33	2.52	0.84	7.01	98	0.029
Physics-based Methods												
IIC [36]	13.62	13.56	13.45	9.46	17.98	–	–	–	–	–	–	–
Woo *et al.* 2018 [39]	4.30	2.86	3.31	0.71	10.14	–	–	–	–	–	–	–
Biological Methods												
Double-Opponency [20]	4.00	2.60	–	–	–	–	–	–	–	–	–	–
ASM 2017 [1]	3.80	2.40	2.70	–	–	–	–	–	–	–	–	–
Learning-free Methods												
White Patch [6]	7.55	5.68	6.35	1.45	16.12	9.91	7.44	8.78	1.44	21.27	–	0.16
Grey World [7]	6.36	6.28	6.28	2.33	10.58	4.59	3.46	3.81	1.16	9.85	–	0.15
General GW [2]	4.66	3.48	3.81	1.00	10.09	3.20	2.56	2.68	0.85	6.68	–	0.91
2st-order grey-Edge [38]	5.13	4.44	4.62	2.11	9.26	3.36	2.70	2.80	0.89	7.14	–	1.30
1st-order grey-Edge [38]	5.33	4.52	4.73	1.86	10.43	3.35	2.58	2.76	0.79	7.18	–	1.10
Shades-of-grey [17]	4.93	4.01	4.23	1.14	10.20	3.67	2.94	3.03	0.99	7.75	–	0.47
Grey Pixel (edge) [41]	4.60	3.10	–	–	–	3.15	2.20	–	–	–	–	0.88
LSRS [19]	3.31	2.80	2.87	1.14	6.39	3.45	2.51	2.70	0.98	7.32	–	2.60
Cheng *et al.* 2014 [12]	3.52	2.14	2.47	0.50	8.74	2.93	2.33	2.42	0.78	**6.13**	–	0.24
GI	**3.07**	**1.87**	**2.16**	**0.43**	7.62	**2.91**	**1.97**	**2.13**	**0.56**	6.67	–	0.40

cross-validation for learning-based methods. The results for this setting are summarized in Table 1a. Among all the compared methods, up to the date of submission of this paper, FFCC [4] achieves the best overall performance

Table 2: Each-camera evaluation on the NUS 8-Camera Dataset. *Std* in the last column refers to the standard deviation of statistics (*e.g.* mean angular error) on 8 cameras.

					NUS 8-camera Dataset				
	Canon 1DS Mark3	Canon 600D	Fujifilm X-M1	Nikon D5200	Olympus E-PL6	Panasonic DMC-GX1	Samsung NX2000	Sony SLT-A57	Std
Cheng et al. 2014 [12]									
Mean	2.93	2.81	3.15	2.90	2.76	2.96	2.91	2.93	0.1152
Median	2.01	1.89	2.15	2.08	1.87	2.02	2.03	2.33	0.1465
Tri	2.22	2.12	2.41	2.19	2.05	2.31	2.22	2.42	0.1309
Best-25%	0.59	0.55	0.65	0.56	0.55	0.67	0.66	0.78	0.0798
Worst-25%	6.82	6.50	7.30	6.73	6.31	6.66	6.48	6.13	0.3558
Chakrabarti et al. *[8] (best), trained on Gehler-Shi, tested here*									
Mean	3.00	3.26	3.12	3.26	3.31	3.30	3.30	3.32	0.1056
Median	2.17	2.48	2.45	2.48	2.50	2.49	2.48	2.56	0.1171
Tri	2.31	2.64	2.60	2.64	2.72	2.69	2.68	2.75	0.1365
Best-25%	0.74	0.83	0.83	0.83	0.85	0.84	0.83	0.86	0.0390
Worst-25%	6.77	7.04	6.89	7.04	7.11	7.12	7.16	7.12	**0.1312**
GI									
Mean	3.02	2.85	2.89	2.85	2.84	2.86	2.86	2.75	**0.0753**
Median	1.87	1.96	1.98	1.96	1.97	1.97	1.97	1.89	**0.0420**
Tri	2.16	2.12	2.15	2.12	2.15	2.17	2.13	2.07	**0.0321**
Best-25%	0.54	0.55	0.55	0.55	0.56	0.56	0.55	0.53	**0.0114**
Worst-25%	7.29	6.79	6.86	6.79	6.70	6.75	6.81	6.51	0.2198

Table 3: Quantitative Evaluation on the MIMO dataset.

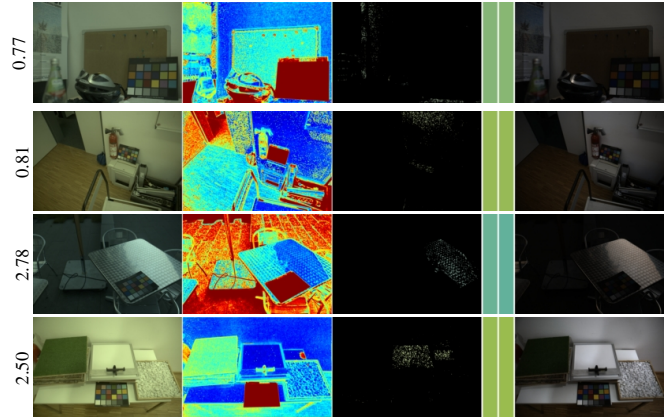| Method	Laboratory(58)		Real-world(20)	
	Median	Mean	Median	Mean
Doing Nothing	10.5	10.6	8.8	8.9
Gijsenij *et al.* [27]	4.2	4.8	3.8	4.2
CRF [5]	2.6	2.6	3.3	4.1
GP (best) [41]	2.20	2.88	3.51	5.68
GI (M=2)	2.09	2.66	**3.32**	**3.79**
GI (M=4)	2.09	2.65	3.47	3.96
GI (M=6)	**2.07**	**2.60**	3.49	3.94

Figure 3: Qualitative results on the single-illumination Gehler-Shi. From left to right: angular error, input image, GI, top 1% pixels chosen as gray pixel, estimated illumination color, the ground truth color and corrected image using the predicted illumination. Macbeth Color Checker is always masked as GI finds perfect gray patch as gray pixels.

with the both datasets. It is important to remark that cross-validation makes no difference to the performance of statistical methods. Therefore, in order to avoid repetition, the performance of competing non-learning methods are shown only once in Table 1b. For visualization purposes, results of learning-based methods that are outperformed by the proposed GI are highlighted in gray. Remarkably, it is clear that, even in the setting which is friendly to learning-based method, GI outperforms several popular learning-based methods (from Gamut [25] to the industry-standard Corrected-Moment [14]) without the need of extensive training and parameter tuning. Visual examples of GI are shown in Fig. 3.

Comparing to the best learning-based methods (*e.g.* [8]),

GI has a noticeable heavy tail in its angular error distribution (*e.g.* amont the worst 25% cases), which suggests that GI would be more optimal if gray pixels would be i.i.d over the whole datasets (*e.g.* natural images). Learning-based methods perform well on these "rare" cases using 3-fold cross-validation, and can further improve "rarity case" per-
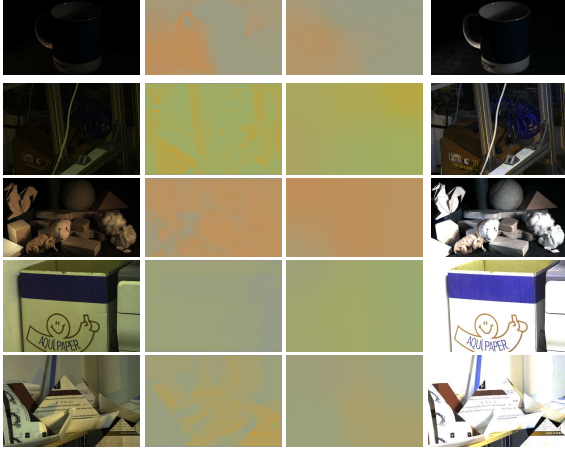
Figure 4: Qualitative results on (multi-illumination) MIMO dataset. From left to right, color-biased input, groundtruth spatial illumination, our spatial estimation using GI, our corrected image.

formance by including more training data (*e.g.* via 10-fold cross-validation) [8].

4.2. Cross-Dataset Setting

We were able to re-run the Bayesian method [22], Chakrabarti *et al.*[8], FFCC [4], and the method by Cheng *et al.* 2015 [13], using the codes provided by the original authors. Note that this list of methods includes FFCC, which showed the best overall performance in the camera-known setting. From the provided code we found different approaches to correct the black level and saturated pixels. For consistency, we used a uniform correction process (given in supplement), which was applied to GI as well.

When we trained on one dataset and tested with another, we made sure that the datasets share no common cameras. For the results reported in this section, we used the best or final setting for each method: Bayes (GT) for Bayesian; Empirical and End-to-End training for Chakrabarti *et al.* [8]; 30 regression trees for Cheng *et al.*; full image resolution and 2 channels for FFCC. Obtained results are summarized in Table 1b. From this table, it is clear that GI outperforms **all** learning-based and statistical methods.

All selected learning-based methods perform worse in this setting, as compared to some statistical methods (*e.g.* LSRS [19], Cheng *et al.* 2014 [12]). It is not surprising that the performance of learning-based methods degrades in this scenario. For example, in [4] it is visualized that FFCC models two varying camera sensitivity for Gehler-Shi in preconditioning filter (two wrap-around line segments), which in cross-dataset setting will be improperly used to evaluate performance on the NUS 8-camera Dataset.

A special feature of the NUS 8-Camera Benchmark is that it includes 8 cameras that share the same scenes. We leveraged this feature to evaluate the robustness of the
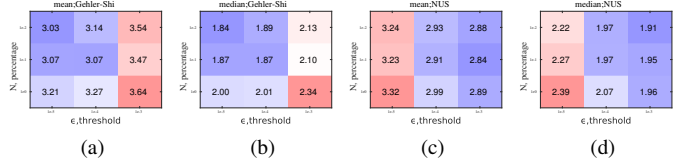

Figure 5: The colormaps of mean and median angular errors corresponding to various N and ϵ (see the text) for (a,b) Gehler-Shi; (c,d) NUS 8-camera.

well-performing learning-free and learning-based methods. These results are summarized in Table 2, where GI achieves much more stable results (standard variance is smaller) across 8 cameras. Due to space limitations, we refer readers to [12] for more results on individual cameras with other methods, including but not limited to [2, 38, 25, 22, 9]. Among all methods in Table 2 of [12], GI is less sensitive to camera hardware.

4.3. Grid Search on Parameters

The only two parameters in GI are: the percentage $N\%$ of pixels chosen as gray for illumination estimation, and the threshold ϵ of Eq. 12 used to remove regions without spatial cues. The former restricts the domain range where illumination norm is measured, analogous to the *receptive field* in deep learning, while the later one passes only noticeable activation, like the *ReLU* activation. Figure 5 summarizes the obtained median and mean angular errors corresponding to a grid search of the parameters, with $N \in \{10^{-2}, 10^{-1}, 1\}$ and $\epsilon \in \{10^{-5}, 10^{-4}, 10^{-3}\}$, on Gehler-Shi Dataset and NUS 8-camera Dataset. The setting ($N = 1e{-}1$ and $\epsilon = 1e{-}4$) results in a good trade-off between mean and median error on both datasets. The shown parameter grid seems loose, but on the contrary, this shows that our method is robust to parameter tuning across orders of magnitude.

4.4. Multi-illumination Setting

As a side product of grayness index, we evaluate the proposed method on a multi-illumination dataset. Table 3 indicates that despite the fact that GI is not designed to deal with spatial illumination changes, it still outperforms well-performing methods [5, 41] with a clear margin. From the mean value over real-world images, it is obvious that GI can better handle multi-illumination situations. Increasing the number of clusters M from 2 to 6 further improved our results on indoor images, but not for wild ones. Figure 4 shows the spatial estimation predicted using GI. Due to Euclidean distance used by the K-means, GI predictions are not sharp in some scenes with complex geometry but still obtain the best overall error rate and plausible visual color correction.

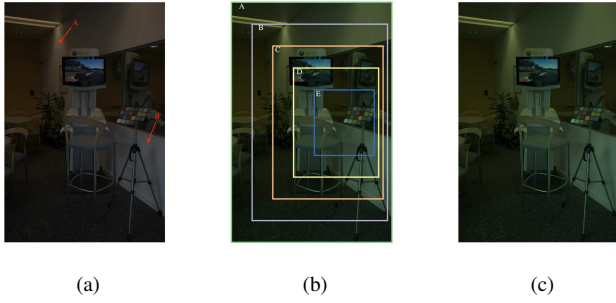| | (a) | (b) | (c) |

Figure 6: (a) Example images from the Gehler-Shi corrected using groundtruth, where two different illuminations (red arrow A and B) exist. (b) We test CC methods in decreasing box sizes (from A to E). (c) Color-biased (a).

Table 4: Testing GI, FFCC on varying-size cropped images from Gehler-Shi, given illumination split from [11].

(a) Double-illumination Setting					(b) Single-illumination Setting				
Gehler-Shi: 66 two-illumination Images					Gehler-Shi: 502 single-illumination Images				
Mean	Median	Trimean	Best-25%	Worst-25%	Mean	Median	Trimean	Best-25%	Worst-25%
GI					*GI*				
A 6.12	4.54	5.24	0.70	13.72	A **2.78**	**1.79**	**2.03**	**0.41**	**6.75**
B 6.06	3.88	4.90	0.92	14.08	B 2.95	1.86	2.12	0.41	7.28
C 6.02	3.63	5.04	0.92	14.55	C 3.32	2.30	2.49	0.50	7.96
D 5.46	3.46	4.13	0.77	13.69	D 3.93	2.97	3.14	0.70	8.90
E **4.96**	**2.94**	**3.45**	**0.53**	**12.42**	E 4.81	3.79	3.94	0.82	10.74
FFCC [4]					*FFCC [4]*				
A **3.11**	**1.67**	**2.25**	**0.44**	**8.00**	A **1.68**	**0.94**	**1.16**	**0.27**	**4.22**
B 3.44	1.84	2.39	0.42	8.69	B 1.72	1.01	1.20	0.27	4.30
C 4.01	2.47	2.92	0.56	10.03	C 1.84	1.11	1.29	0.29	4.58
D 4.64	3.13	3.53	0.62	11.38	D 2.13	1.29	1.43	0.36	5.45
E 4.99	3.29	3.72	0.60	11.92	E 2.39	1.39	1.58	0.38	6.17

5. Problems with the "Ground-truth"

We investigated those cases where GI made erratic predictions (see the supplement for erratic cases) and have observed that, in some images, there exists gray pixels casted by two illumination sources. A similar problem was noticed by Cheng *et al.* [11], who claimed that in the Gehler-Shi [34], there are 66 two-illumination images. An example of this problem is illustrated in Fig. 6. In Fig. 6(a), where pixels near arrows A and B share the same surface (white wall) but have different illuminations, the color of pixel in the neighborhood of B is close to the Macbeth Color Checker (MCC). In such case, our GI does a good job in identifying gray pixels by following the designed rules and finding gray pixels lying in two illuminants, but this comes at the cost of a large angular error. As a first impression, we suppose this is due to the MCC being more dominated by one of the illuminants.

We designed a simple experiment to investigate our observation. For the list of 66 two-illumination images (given in [11]) in the Gehler-Shi and the remaining 502 single-illumination images, we test GI and FFCC [4] (full resolution, 2 channels, pretrained on whole Gehler-Shi) on images cropped by boxes of decreasing sizes centered at the MCC (from box A to box E in Fig. 6b. Specifically, the boxes are generated by halving the width and height of the previous box.

The results summarized in Tables 4a and 4b show a crucial fact: in the single-illumination subset, GI yields larger angular errors as the testing box gets smaller (from box A to E). In contrast, in the double-illumination subset, this tendency is reversed. It makes sense that the performance of GI decreases as the testing box shrinks since less reference points are available. A reasonable explanation to the abnormal tendency in the two-illumination subset is that the MCC is placed mainly in one illumination, reflecting a biased "ground-truth". This problem restricts the upper limit of the performance of GI and possibly also other statistical color constancy methods. Learning-based methods (especially CNN-based method) suffer less from this problem, as they can learn to reason about some structural information, *e.g.* whole-image chroma histogram, the physical geometry of the scene, the location where MCC is placed. As expected, FFCC performs worse on smaller boxes. Bearing these results in mind, we argue that learning-based methods and statistical methods should be compared by considering their corresponding advantages and limitations in both single-dataset and cross-dataset scenarios.

6. Conclusions

We derived a method to compute grayness in a novel way – Grayness Index. It relies on the Dichromatic Reflection Model and can detect gray pixels accurately. Experiments performed on the tasks of single-illumination estimation and multi-illumination estimation verified the effectiveness and efficiency of GI. On standard benchmarks, GI estimates illumination more accurately than state-of-the-art learning-free methods in about 0.4 seconds. GI has a clear physical interpretation, which we believe can be used for other vision tasks, *e.g.* intrinsic image decomposition.

Other conclusions also emerged from the research: learning-based methods generally perform worse in the cross-dataset setting; When testing on an image with color checker masked by zeros, learning-based methods can still exploit the location of the color checker and overfit to scene and camera specific features.

References

[1] A. Akbarinia and C. A. Parraga. Colour constancy beyond the classical receptive field. *TPAMI*, 2017. 5

[2] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *TIP*, 11(9):972–984, 2002. 1, 2, 5, 7

[3] J. T. Barron. Convolutional color constancy. In *ICCV*, 2015. 2, 5

[4] J. T. Barron and Y.-T. Tsai. Fast fourier color constancy. In *CVPR*, 2017. 2, 4, 5, 7, 8

[5] S. Beigpour, C. Riess, J. Van De Weijer, and E. Angelopoulou. Multi-illuminant estimation with conditional random fields. *IEEE Transactions on Image Processing*, 23(1):83–96, 2014. 4, 6, 7

[6] D. H. Brainard and B. A. Wandell. Analysis of the retinex theory of color vision. *JOSA A*, 3(10):1651–1661, 1986. 1, 2, 5

[7] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, 1980. 2, 5

[8] A. Chakrabarti. Color constancy by learning to predict chromaticity from luminance. In *NIPS*, 2015. 2, 5, 6, 7

[9] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *TPAMI*, 34(8):1509–1519, 2012. 1, 2, 5, 7

[10] X. Chen and C. Zitnick. Minds eye: A recurrent visual representation for image caption generation. In *CVPR*, 2015. 5

[11] D. Cheng, A. Kamel, B. Price, S. Cohen, and M. S. Brown. Two illuminant estimation and user correction preference. In *CVPR*, 2016. 8

[12] 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*, 31(5):1049–1058, May 2014. 1, 2, 4, 5, 6, 7

[13] D. Cheng, B. Price, S. Cohen, and M. S. Brown. Effective learning-based illuminant estimation using simple features. In *CVPR*, 2015. 5, 7

[14] G. D. Finlayson. Corrected-moment illuminant estimation. In *ICCV*, pages 1904–1911, 2013. 2, 5, 6

[15] G. D. Finlayson and G. Schaefer. Convex and non-convex illuminant constraints for dichromatic colour constancy. In *CVPR*, volume 1, pages I–I. IEEE, 2001. 2

[16] G. D. Finlayson and G. Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *IJCV*, 42(3):127–144, 2001. 2

[17] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color Imaging Conference (CIC)*, 2004. 1, 2, 5

[18] D. H. Foster. Color constancy. *Vision research*, 51(7):674–700, 2011. 1

[19] S. Gao, W. Han, K. Yang, C. Li, and Y. Li. Fefficient color constancy with local surface reflectance statistics. In *ECCV*, 2014. 1, 2, 5, 7

[20] S.-B. Gao, K.-F. Yang, C.-Y. Li, and Y.-J. Li. Color constancy using double-opponency. *TPAMI*, 37(10):1973–1985, 2015. 5

[21] 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*, 34(8):1448–1462, 2017. 1

[22] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *CVPR*, 2008. 1, 2, 4, 5, 7

[23] A. Gijsenij. Color constancy research website: http://colorconstancy.com. In *http://colorconstancy.com*, 2019. 5

[24] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *TPAMI*, 33(4):687–698, 2011. 1, 2, 5

[25] A. Gijsenij, T. Gevers, and J. Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *IJCV*, 86(2-3):127–139, 2010. 1, 2, 5, 6, 7

[26] A. Gijsenij, T. Gevers, and J. Van De Weijer. Computational color constancy: Survey and experiments. *TIP*, 20(9):2475–2489, 2011. 2

[27] A. Gijsenij, T. Gevers, and J. Van De Weijer. Improving color constancy by photometric edge weighting. *TPAMI*, 34(5):918–929, 2012. 6

[28] Y. Hu, B. Wang, and S. Lin. Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017. 2, 5

[29] H. R. V. Joze and M. S. Drew. Exemplar-based color constancy and multiple illumination. *TPAMI*, 36(5):860–873, 2014. 1, 2, 5

[30] H.-C. Lee, E. J. Breneman, and C. P. Schulte. Modeling light reflection for computer color vision. *TPAMI*, 12(4):402–409, 1990. 3

[31] Y. Qian, K. Chen, J. Kämäräinen, J. Nikkanen, and J. Matas. Deep structured-output regression learning for computational color constancy. In *ICPR*, 2016. 2

[32] Y. Qian, K. Chen, J. Kämäräinen, J. Nikkanen, and J. Matas. Recurrent color constancy. In *ICCV*, 2017. 2

[33] S. A. Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985. 2

[34] 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/*, 2010. 4, 8

[35] W. Shi, C. C. Loy, and X. Tang. Deep specialized network for illumination estimation. In *ECCV*, 2016. 2, 5

[36] R. T. Tan, K. Ikeuchi, and K. Nishino. Color constancy through inverse-intensity chromaticity space. In *Digitally Archiving Cultural Objects*, pages 323–351. Springer, 2008. 5

[37] S. Tominaga. Multichannel vision system for estimating surface and illumination functions. *JOSA A*, 13(11):2163–2173, 1996. 2

[38] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *TIP*, 16(9):2207–2214, 2007. 1, 2, 5, 7

[39] S.-M. Woo, S.-h. Lee, J.-S. Yoo, and J.-O. Kim. Improving color constancy in an ambient light environment using the phong reflection model. *TIP*, 27(4):1862–1877, 2018. 2, 5

[40] W. Xiong, B. Funt, L. Shi, S.-S. Kim, B.-H. Kang, S.-D. Lee, and C.-Y. Kim. Automatic white balancing via gray surface identification. In *Color and Imaging Conference (CIC)*, 2007. 2

[41] K.-F. Yang, S.-B. Gao, and Y.-J. Li. Efficient illuminant estimation for color constancy using grey pixels. In *CVPR*, 2015. 1, 2, 3, 4, 5, 6, 7