Joint estimation of spectral illuminant and reflectance from an RGB image

Hongyun Gao and Rafał K. Mantiuk

University of Cambridge, UK

Abstract. This paper addresses the challenge of accurately estimating both the full illuminant spectral power distribution (SPD) and the per-pixel spectral reflectance from an RGB image captured with a known camera. Full spectral information allows us to perform more accurate white balance, or render the scene under another illuminant. Traditional color constancy methods focus on predicting the illuminant color as a 3-dimensional projection of the infinite-dimensional illuminant SPD onto the camera spectral sensitivity functions (SSFs) space. Because different illuminants can have the same projection in the 3-dimensional SSFs space, those traditional methods cannot differentiate between such illuminants (metamers in the camera response space) and hence remove the color cast resulting from the illumination. We reconstruct the spectral information using a neural network with two interconnected branches: one branch predicts the illuminant SPD, and the other reconstructs the per-pixel spectral reflectance by integrating the predicted SPD within its intermediate layers. Experimental results demonstrate that our framework achieves superior performance in estimating both the illuminant SPD and the per-pixel spectral reflectance compared to the previous approach.

1 Introduction

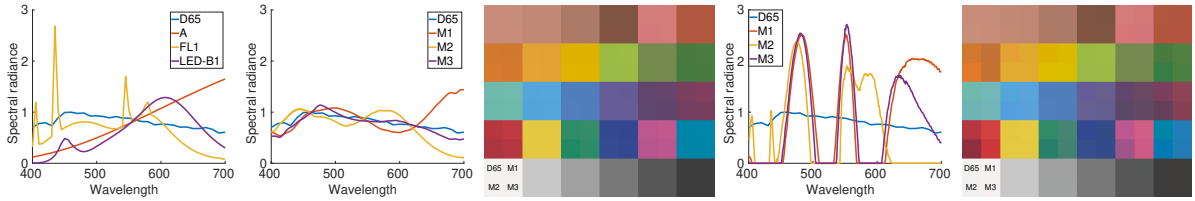
Spectral information is lost during camera image formation, as the spectral scene radiance—resulting from the product of the illuminant spectral power distribution (SPD) and object spectral reflectance—is projected onto the camera's spectral sensitivity functions (SSFs), yielding a 3-channel RGB value v_c

$$v_c = \int_\lambda l(\lambda)\, s(\lambda)\, q_c(\lambda)\, d\lambda, \quad c \in \{R, G, B\}. \tag{1}$$

where λ denotes the wavelength in the visible spectrum; $l(\lambda)$ is the illuminant SPD; $s(\lambda)$ denotes the spectral reflectance; $q_c(\lambda)$ is the SSF for red, green and blue. To remove the color cast caused by varying illumination and to recover object colors under a canonical light (typically D65), color constancy methods typically follow a two-stage pipeline. In the first stage, the illuminant color (e_r, e_g, e_b) defined as

$$e_c = \int_\lambda l(\lambda)\, q_c(\lambda)\, d\lambda, \quad c \in \{R, G, B\}\,. \tag{2}$$

is estimated from the RGB image. In the second stage, following von Kries–Ives adaptation [9], the white balanced image value $\bar{v}_c$ is calculated by dividing the illuminant color from the raw value $\frac{v_c}{e_c}$ in each color channel. While this procedure yields accurate color reconstruction in typical scenarios, it fails when distinct illuminant SPDs result in the same projection as defined in Eq. (2). We refer to such a set of illuminants sharing identical (e_r, e_g, e_b) values as camera metamers, adopting the term from colorimetry, where metamers denote different spectral colors that appear identical to the human eye.



(a) 4 CIE illuminants (b) SPD of metamers (c) ColorChart of (b) (d) SPDs for max ΔE (e) ColorChart of (d)

Figure 1: Four illuminants plotted in (b) result in the same illuminant color captured by our Sony A7R3 camera. The three metamers, M1 to M3, were optimized using the corresponding CIE illuminants shown in (a) as initialization. A color chart with 30 patches was synthesized in (c) under the four illuminants from (b), assuming CIE XYZ 1931 color matching functions. Each patch was subdivided into 2×2 parts corresponding to D65 and three metamers. Under the metamer constraint, one can also optimize illuminants to maximize the mean CIE Delta E 2000 on the color chart. The SPDs and color chart are shown in (d) and (e), illustrating the limitations of current methods under camera metamerism.

As an example shown in Figure 1b, we optimize camera metamers M1, M2, and M3 for Sony A7R3 camera given the corresponding initial illuminants, i.e. incandescent illuminant A, fluorescent lamp FL1, and LED-B1, plotted in Figure 1a. When we while-balance a color chart using Kries–Ives adaptation model, we end up with consistent colors for gray values, but varying colors for color patches (see red, pink, and purple patches), as shown in Figure 1b. This is because all four camera metamers result in the same RGB illuminant color, which cannot account for the differences between the four illuminants. These inaccuracies become more pronounced when the optimization objective is modified to maximize CIE Delta E 2000 color differences under the metamer constraint. As illustrated in Figure 1e, significant color variations emerge across many patches, highlighting the necessity of estimating the full illuminant SPD rather than relying solely on the 3-dimensional illuminant color for accurate color reconstruction.

Therefore, this paper addresses the joint estimation of spectral illuminant and surface reflectance from a single RGB image captured by a known camera. We propose a two-branch neural network architecture: the first branch predicts the illuminant SPD, while the second branch estimates per-pixel spectral reflectance using the predicted SPD as a condition.

2 Related Work

Computational color constancy aims to estimate the illuminant color (e_r, e_g, e_b, see Eq. (2)), which represents the projection of the illuminant SPD onto the camera's SSFs subspace. Many statistics-based color constancy methods have been proposed to improve the accuracy of the estimation, including the classical Gray-World [5], Max-RGB [16], Shades of Gray (SoG) [8], Gray-Edge [23], optimal method selection from Weibull parameterization and mixture of Gaussians (MoG) classifiers [11], frequency-based features using difference of Gaussians and discrete cosine transforms [6], first principal component from top and bottom pixel projection distances [7]. Bayesian approaches have also been explored [4, 21, 10], framing illuminant estimation as a posterior inference problem conditioned on image intensities. More recently, CNN-based methods, including [3, 2, 22, 13, 1], learn color constancy models from large-scale images and their corresponding ground-truth illumination color.

In contrast to a 3-dimensional RGB illuminant color, the SPD preserves full spectral information and is essential in dealing with illumination metamerism. Khan et al. [14] extend the RGB statistics-based methods into multispectral images and recover the SPD of the input multispectral images. To handle illuminant estimation for both RGB and multispectral images, Robles-Kelly and Wei [20] exploit image patches of three different scales and estimate the illumination spectrum using a 5-layer CNN.

Several works attempted to estimate the spectral illumination and reflectance from RGB images. Nguyen et al. [19] propose a training-based method to reconstruct both illuminant SPD and spectral reflectance from RGB images. The SoG method [8] is first used to produce white-balanced RGB images, and the trained radial basis function network is then used to recover the spectral reflectance. The illuminant SPD is then estimated using the PCA bases, which are calculated from their collected illuminant SPDs. In [24], multispectral illumination and reflectance estimation are modeled as a constrained matrix factorization problem, where the low-rank property of spectral reflectance is exploited to constrain the separation of illumination and reflectance. Li et al. [17] propose a deep unrolling network to solve the constrained matrix factorization using alternating direction method of multipliers(ADMM) optimization.

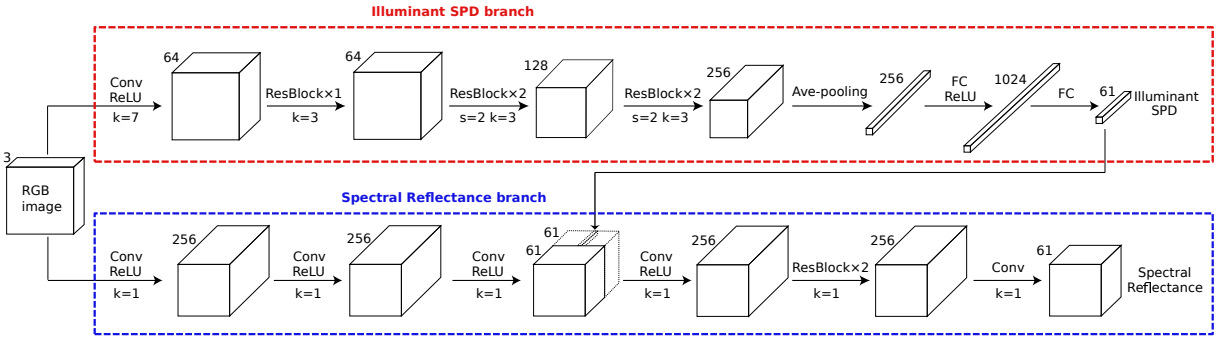


Figure 2: Our proposed joint estimation network consists of an illuminant SPD branch (red) and a spectral reflectance branch (blue). The channel dimension is marked next to each feature map. The applied operators between neighboring feature maps are listed on top of each arrow and the respective convolution stride (s) and kernel size (k) are listed below the arrow.

3 Our method

3.1 The joint framework

As illustrated in Figure 2, the proposed network consists of an illuminant SPD branch and a spectral reflectance branch. The SPD branch first extracts shallow features using a convolutional layer followed by a residual block (ResBlock) [12]. The feature maps are then passed through 2 concatenated ResBlocks in which the feature map is spatially downsampled by 2 and channel dimension extended by 2. Another two ResBlocks are applied to capture nonlinear transformations tailored to illuminant estimation. The resulting feature maps are average pooled to remove spatial dimensions, producing a 1D vector, which is further transformed into the estimated illuminant SPD using two fully connected layers.

In addition to the SPD branch, the spectral reflectance branch estimates per-pixel reflectance spectra. The first three convolutional layers extract reflectance-relevant features, from which the predicted SPD is concatenated to condition the spectral reflectance prediction. This formulates an inverse problem—recovering spectral reflectance from the RGB input and predicted SPD. The concatenated features are processed through a merging point-wise convolutional layer, two ResBlocks, and another channel-reduction point-wise convolutional layer to generate the final per-pixel reflectance, where point-wise convolutions are used to model the inversion process for each pixel, which is independent for each pixel.

For training the network, we apply Mean Squared Error (MSE) loss to both the outputs of SPD and spectral reflectance branches as follows:

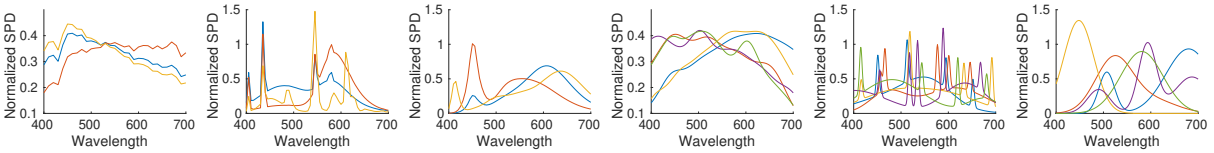
$$\begin{aligned} \mathcal{L}_{spd} &= \|l(\lambda) - \hat{l}(\lambda)\|^2, \\ \mathcal{L}_{ref} &= \frac{1}{M}\sum_{p} \|s_p(\lambda) - \hat{s}_p(\lambda)\|^2. \end{aligned} \tag{3}$$

where $l(\lambda)$ and $\hat{l}(\lambda)$ denote the ground-truth and predicted SPDs and $s_p(\lambda)$ and $\hat{s}_p(\lambda)$ represent the ground-truth and predicted spectral reflectance at pixel position p. M is the total number of pixels in the predicted spectral reflectance.

3.2 Illuminant SPD generation

The small size of the available illuminant SPD datasets restricts the effective use of deep neural networks in modeling the relationship between illuminants and their corresponding RGB values. When using existing datasets, such as CIE standard illuminants, the network learns to predict one of the SPDs from the dataset, and it cannot generalize to unseen SPDs. This motivates the need to generate a rich and diverse set of illuminant SPDs, which could be used for training.

Analysis of CIE standard illuminants, including natural daylights, fluorescent lamps, and LEDs, reveals distinct spectral characteristics. Natural daylights (e.g., D65 and D50) exhibit balanced spectra, while fluorescent lamps and LEDs present sharp spectral peaks. Both bell-shaped (LED) and spike-shaped (fluorescent) SPDs can be approximated using Gaussian functions with varying standard deviations, and the spectra of natural daylights can be modeled with a series of Gaussian functions at different mean values. These findings motivate the approximation of illuminant SPDs using a mixture of Gaussian basis



(a) GT natural (b) GT FL (c) GT LED (d) Gen. natural (e) Gen. FL (f) Gen. LED

Figure 3: The comparison between CIE standard illuminants (left 3) and generated SPDs (right 3) for natural daylight, fluorescents and LEDs.

functions with adaptive location, width, and intensity. Our SPD generation is defined as

$$t(\lambda; \boldsymbol{w}, \boldsymbol{c}, \boldsymbol{\sigma}) = \sum_{i=1}^{N} w_i \, e^{-\frac{(\lambda - c_i)^2}{2\sigma_i^2}} \tag{4}$$

where $t(\lambda; \boldsymbol{w}, \boldsymbol{c}, \boldsymbol{\sigma})$ denotes the approximated SPD and N is the number of basis functions. c_i, σ_i, and w_i represents the location, width, and intensity for the i-th Gaussian basis. The vector $\boldsymbol{c} = [c_1, c_2, \ldots, c_N]$, $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_N]$ and $\boldsymbol{w} = [w_1, w_2, \ldots, w_N]$ are generation parameters. To validate the precision of Eq. (4) in representing CIE standard illuminants, we can optimize $\boldsymbol{c}$, $\boldsymbol{\sigma}$, and $\boldsymbol{w}$ for each CIE illuminant. The optimization objective is formulated as

$$\begin{aligned} \arg\min_{\boldsymbol{w},\boldsymbol{c},\boldsymbol{\sigma}} & \int_{\lambda} (l(\lambda) - t(\lambda; \boldsymbol{w}, \boldsymbol{c}, \boldsymbol{\sigma}))^2 + \alpha \, \Delta E(l(\lambda), t(\lambda; \boldsymbol{w}, \boldsymbol{c}, \boldsymbol{\sigma})) \\ & \text{s.t.} \quad \lambda_{min} \leq \boldsymbol{c} \leq \lambda_{max}, \quad \boldsymbol{w} \geq 0, \quad N \leq \tau \end{aligned} \tag{5}$$

where $l(\lambda)$ denotes the ground-truth SPD and ΔE is CIE Delta E 2000 [18] color difference metric. The wavelength range is bounded by λ_{min} and λ_{max}, and α balances the trade-off between L^2 error and color accuracy. Theoretically, we could approximate the target with any precision by adding adequate basis functions, but allowing for an extremely large number of Gaussian basis functions to approximate a target SPD is not practical for our realistic SPDs generation task. To ensure practical and efficient SPD generation, we impose a constraint $N \leq \tau$ where τ is a small integer.

4 Experiment

4.1 Implementation and dataset

All experiments were conducted on the visible spectrum range from 400 to 700 nm. As CIE standard illuminants are measured at 5 nm intervals, all generated illuminant SPDs and spectral reflectance are 61-dimensional vectors. We used the KAUST multispectral image dataset [17], which contains 409 reflectance images of size 256×256 with 33 spectral bands from 400 to 720 nm at 10 nm intervals. To match our spectral range, we extracted the data between 400 and 700 nm and resampled it at 5 nm intervals using linear interpolation. The dataset was randomly split into 300 training and 109 testing images.

During training, SPDs representing natural daylight, fluorescent, and LED lighting were dynamically generated with equal probability and normalized to a vector sum of 100. To reduce GPU memory usage and accelerate training, the input RGB images for the reflectance branch were downsampled by a factor of four. The network was trained for 100 epochs with a batch size of 16 using the Adam optimizer [15] with default parameters. During testing, RGB images were synthesized from 109 test reflectance images under 48 CIE illuminants, including 5 natural daylights, 1 incandescent, 1 equal-energy, 27 fluorescents, 9 LEDs, and 5 high-pressure lamps.

4.2 Accuracy of SPD generation

We evaluated the effectiveness of the proposed SPD generation method in Eq. (4) by approximating CIE standard illuminants through the optimization defined in Eq. (5). To initialize the optimization, we identify all local maxima and convex points in the target SPD, assigning their individual wavelengths to c_i. The corresponding w_i is initialized as $l(c_i)$, and σ_i is set inversely proportional to the local curvature. For practical SPD generation, we set τ to 10 and perform the optimization using MATLAB's interior-point method via the `fmincon` function. The average CIE Delta E 2000 of our method on all 48 standard illuminants is 0.28, indicating a relatively accurate approximation. We generate diverse SPDs during

Table 1: Quantitative comparisons between our method and TBSR [19]. MSE-l, MSE-s and MSE-v refers to the MSE of illuminant SPD, spectral refletance, and reconstructed RGB image, respectively. Angle is the angular error between the predicted and ground-truth illuminant color.

	MSE-l	MSE-s	MSE-v	Angle
Ours	1.0674	0.0047	0.3004	2.808
TBSR	1.3358	0.0052	0.5979	9.445

training by randomly sampling the parameters c_i, σ_i and w_i. Distinct generation strategies are employed for natural daylight, fluorescent lamps, and LEDs. Some random examples of the synthesized SPDs are plotted in Figure 3 along with CIE standard illuminants of the corresponding category.

4.3 Our results

We compared our method with TBSR [19], a leading approach for joint illuminant SPD and spectral reflectance estimation. As their model was trained on their collected SPDs, we re-trained their radial basis networks on our synthesized RGB images from the training reflectance images and 48 CIE illuminants. Although this comparison favors their method due to the inclusion of the same illuminants in both training and testing, our approach still achieves superior performance in the quantitative metrics reported in Table 1. The superior accuracy of our SPD prediction stems from the ability to learn RGB images under generated illuminant SPDs, whereas TBSR relies only on a 3-dimensional PCA-based representation, limiting its capacity to handle diverse SPDs. TBSR estimates spectral reflectance from white-balanced RGB images, hence inevitably introduces errors for the reflectance estimation. In contrast, our reflectance branch leverages both the predicted SPD and the original RGB image. Visual comparisons for 3 illuminants are presented in Figure 4. While TBSR achieves comparable results under natural daylight D75, its accuracy degrades under fluorescent FL2 and LED-V1. In contrast, our approach is more robust across different illuminants and gives much smaller errors.

5 Conclusions and limitations

We propose a network that jointly estimates the illuminant SPD and per-pixel spectral reflectance. Our SPD generation method supports diverse SPD synthesis, enhancing the network's capacity to generalize across varied illumination conditions. Despite its effectiveness, our framework has several limitations. First, it is trained and evaluated on synthesized RGB images, excluding real-world factors such as sensor noise and inaccuracies in camera SSF measurements. Second, the current model assumes uniform illumination throughout the image, which is rarely true. Addressing pixel-wise or region-based illuminant SPD estimation under non-uniform lighting remains a more general yet challenging extension.

Acknowledgements

We would like to thank Joseph Meehan and Ge Song for their feedback on this project.

References

[1] M. Afifi and M. S. Brown. Deep white-balance editing. In *CVPR*, pages 1397–1406, 2020.

[2] J. T. Barron. Convolutional color constancy. In *ICCV*, pages 379–387, 2015.

[3] S. Bianco, C. Cusano, and R. Schettini. Color constancy using cnns. In *CVPRW*, pages 81–89, 2015.

[4] D. H. Brainard and W. T. Freeman. Bayesian color constancy. *JOSA A*, 14(7):1393–1411, 1997.

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

[6] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *IEEE TPAMI*, 34(8):1509–1519, 2011.

[7] 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, 2014.

[8] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 12, pages 37–41, 2004.

[9] K. R. Gegenfurtner and L. T. Sharpe. *Color vision: From genes to perception*. Cambridge University Press, 2001.

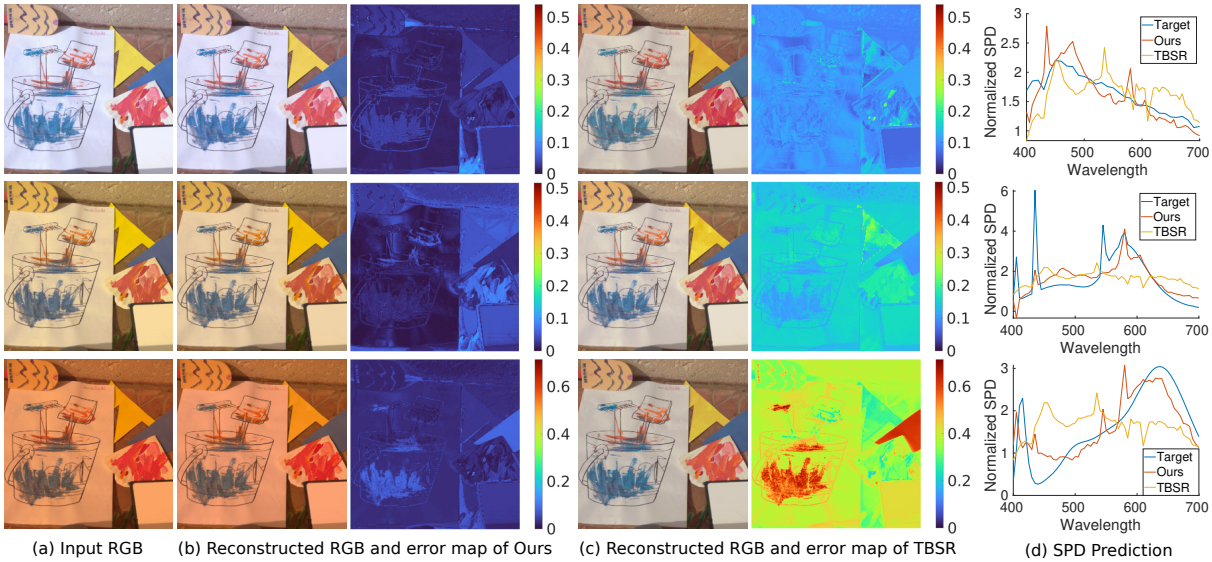


Figure 4: Comparison of our proposed method with TBSR [19] on three illuminants i.e., D75 (first row), FL2 (second row), and LED-V1(third row). The input RGB images are listed in (a). Reconstructed RGB images using the predicted SPD and spectral reflectance and error maps of our method and TBSR are shown in (b) and (c). The SPD predictions with the ground-truth are illustrated in (d).

[10] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *CVPR*, pages 1–8, 2008.

[11] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE TPAMI*, 33(4):687–698, 2010.

[12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

[13] Y. Hu, B. Wang, and S. Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, pages 4085–4094, 2017.

[14] H. A. Khan, J.-B. Thomas, J. Y. Hardeberg, and O. Laligant. Illuminant estimation in multispectral imaging. *JOSA A*, 34(7):1085–1098, 2017.

[15] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[16] E. H. Land and J. J. McCann. Lightness and retinex theory. *JOSA A*, 61(1):1–11, 1971.

[17] Y. Li, Q. Fu, and W. Heidrich. Multispectral illumination estimation using deep unrolling network. In *ICCV*, pages 2672–2681, 2021.

[18] M. R. Luo, G. Cui, and B. Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application*, 26(5):340–350, 2001.

[19] R. M. Nguyen, D. K. Prasad, and M. S. Brown. Training-based spectral reconstruction from a single rgb image. In *ECCV*, pages 186–201, 2014.

[20] A. Robles-Kelly and R. Wei. A convolutional neural network for pixelwise illuminant recovery in colour and spectral images. In *IEEE ICPR*, pages 109–114, 2018.

[21] C. Rosenberg, A. Ladsariya, and T. Minka. Bayesian color constancy with non-gaussian models. In *NeurIPS*, 2003.

[22] W. Shi, C. C. Loy, and X. Tang. Deep specialized network for illuminant estimation. In *ECCV*, pages 371–387, 2016.

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

[24] Y. Zheng, I. Sato, and Y. Sato. Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization. In *CVPR*, pages 1779–1787, 2015.