https://doi.org/10.2352/CIC.2025.33.1.15

Facial-Centric Color Constancy Dataset to Improve Scenario-Specific White Balance Algorithms

Yuyang Liu, Minchen Wei, The Hong Kong Polytechnic University, *minchen.wei@polyu.edu.hk*

Abstract

Eliminating the color cast of the illuminant is a critical step in modern image processing systems, which has been addressed with a great number of illuminant estimation algorithms. The algorithms are found not effective for some specific contexts and applications, which leads to the development of scenarios-specific algorithms leveraging domain-specific cues. This paper investigated how facial cues help illuminant estimation. A total of 1299 images were captured under various dual-illuminant conditions, including real-world environments and lab settings. Modifications were made on existing methods by considering the facial information, which resulted in better performance.

Introduction

Illuminant estimation is a core challenge in image processing, aiming to eliminate the color cast of the illumination achieve color constancy. This process simulates the chromatic adaptation mechanism of the human visual system. In modern digital cameras, it corresponds to the white balance adjustment in the image signal processing (ISP) pipeline, ensuring colors are rendered as if they were captured under a standard "neutral" light source.

The fundamental goal of white balance is to estimate the color of the illuminant based on a RAW image, which is indeed challenging due to the difficulties to separte the illuminant and object colors from the image. Traditional statistics-based methods address this by imposing some assumptions about the scene reflectance properties. For instance, the gray world method assumes that the average reflectance across the scene approximates a neutral gray, as in [1,2]. The gray edge method relies on the hypothesis that chromatic transitions (edges) are predominantly neutral, as proposed in [3–5]. The white patch approach assumes the presence of a white object in the scene, where the brightest pixel corresponds to the illuminant color, as detailed in [6–8]. And the achromatic pixel method identifies the gray pixels by detecting minimal color variation across the RGB channels in localized areas, as described in [9, 10]. All these methods work generally well if the assumptions hold. Many corner cases, however, cannot satisfy these assumptions, resulting in bad performance.

Instead of proposing more constraints or assumptions, great efforts have been made to develop learning-based methods, which rely on large-scale datasets. For instance, the FC4 method [11] employs a deep neural network to process RAW input images, using implicit semantic information to reduce the effects of metamerism. The CCC [12], FFCC [13], and C5 [14] methods convert images into log-uv histograms and formulate the task of illuminant estimation as a two-dimensional detection problem, which are also computationally efficient. The C4 method [15] builds upon the FC4 method by arranging three instances of its architecture in a cascaded configuration to iteratively enhance the estimation accuracy. The CLCC method [16] enhances FC4's robustness through a contrastive learning, guaranteeing that the intermediate outputs remain consistent and reliable across varying conditions. The OneNet method [17] assumes that the semantic information is irrelevant for illuminant estimation and instead employs CNNs with 1×1 kernels to achieve its objectives.

All these methods are based on the global illuminant assumption, with the performance validated on some color constancy datasets (e.g., INTEL-TAU [18] and NUS [5]). While improving the performance for covering a wider range of conditions remains an active research area, designing methods for specific cases can also be beneficial.

The facial skin appearance is always a critical element for image quality, and past studies (e.g., [19]) suggest that the human skin exhibits a limited color gamut and can be used as a stable reference for illuminant estimation. In this work, we collected a large-scale multi-illuminant color constancy dataset containing face-centric scenes and demonstrated its usefulness in developing illuminant estimation methods.

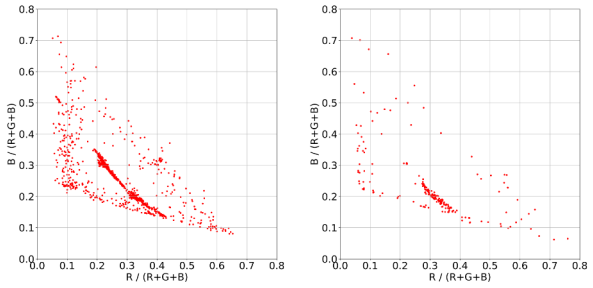
PolyU Face Color Constancy Dataset

Dataset description

The PolyU Face Color Constancy Dataset was collected for developing illuminant estimation methods for multi-illuminant scenes. The dataset includes 1299 facial images, with each image containing a facial portait and two illuminants, which were collected in field and lab settings. A total of eight human models were involved in the data collection. Such setups make the dataset useful for some applications, such as video streaming and selfies. The dataset includes a wide range of illuminants, as shown in Figure 1. Unlike the other multi-illuminant color constancy dataset (e.g., LSMI [20]) providing pixel-wise illuminant maps, this dataset only provides foreground and background illuminant conditions to simplify the problem formulation and data collection.

Dataset acquisition

The images were captured using the main camera of a commercially available smartphone, with the ISO set to a low value (i.e., 200) to reduce the noise. A standard ColorChecker was used to capture the lighting condition. Each scene contains two illuminants, with a background illuminant behind the portrait and a foreground illuminant shining on the portrait. For the lab settings, we used to two LED light sources to provide the background and foreground illumination L_b and L_f; for the field settings, the background illumination L_b was from the ambient light (a single ambient illuminant) and the foreground L_f was provided by our light source. The LED light sources were tunable, so that we were



(a) (b)

Figure 1: Distribution of the illuminants in the dataset in the camera color space. (a) Foreground illuminant; (b) background illuminants.

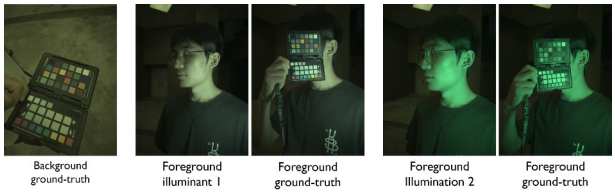


Figure 2: Examples of data collection.

able to create a wide range of lighting conditions. The intensity of the foreground illumination was carefully adjusted, so that it would not affect the background illumination. During the data collection, we used the ColorChecker to capture the background illumination L_b, and let the model to hold the ColorChecker when capturing the foreground illumination L_f. Figures 2 and 3 give examples of how we captured the background and foreground illumination during the data collection. Thus, each setting had three RAW images, with two images for capturing the ground-truth illuminants of the foreground and background illumination and one image for the portrait. During the data collection, around five to ten different foreground illumination share the same background illumination, by varying the color of the light source.

All the RAW images were 10-bit files and were converted to 16-bit PNG files with three steps:

1. Black-level subtraction and scaling: images were normalized to the range $[0,1]$ using the black and white level values from the EXIF metadata.
2. Demosaicing: the RAW images were in an RGGB Bayer pattern. Demosaicing was performed using a simple method, where the two green channels in each 2×2 pixel block were averaged to derive the value G.
3. Resolution adjustment: the RAW images were downsampled to a resolution of 1024×768.

For the ground-truth illuminant color, the 20th patch on the ColorChecker was used, with the average value of a small window at the center of the patch used to reduce the effect of noise.

Methodology

Image model

For a pre-procssed RAW image I, typical methods assuming a single illuminant define the image model as

$$I = W \times L \tag{1}$$

where W is the white-balanced image, and $L = \{R,G,B\}$ represents the global illuminant color. In such a case, the white balance algorithms aim to predict L from I.

For the algorithms deriving pxiel-wise ground-truth under two illuminants (e.g., [20]), the problem is formulated as:

$$I(x) = W(x) \times \alpha(x) L_a + (1-\alpha(x)) L_b \tag{2}$$

where x represents the pixel location, L_a and L_b are the first and second illuminants respectively, and α is the coefficient map for the first illuminant. For such a formulation, the algorithms must predict either L_a, L_b, and α, or the illuminant map derived from L_a, L_b, and α directly.

In contrast, We simplify such a problem by removing the coefficient map α, which is indeed difficult to obtain in real-world conditions. In our formulation, for a RAW image I, white balance algorithms are expected to predict two illuminants L_a and L_b in the scene.

Multi-illuminant white balance algorithms

Most previous algorithms predict the primary illuminant in a scene, and some of them can be extended to multi-illuminant conditions with some minor modifications. In this work, we adapt several state-of-the-art algorithms to fit such a formulation.

FC4 [11]: FC4 takes a whole image as the input and uses a relatively large convolutional neural network as a backbone to extract the image feature maps and the corresponding confidence maps. The final estimation results are derived as the weighted sum of the feature maps based on the confidence map. In the original design, the backbone outputs a tensor with a shape of $h\times w\times 4$, where the first three channels represent the image features, and the last channel is the confidence map. For our purpose, this architecture can be slightly modified to output a tensor by doubling the number of channels.

C5 [14]: C5 is a hyper-net that outputs parameters for the FFCC algorithm [13] for each image, then applies the parameters to produce the estimation result. Instead of using the whole image as the input, it transforms the image into the log-uv color space and uses the histogram for prediction. C5 can also be adapted to our formulation by outputting another set of parameters, which share the same regularization constraints as the original ones.

OneNet [17]: OneNet assumes that the semantic information is irrelevant to white balance. It employs a simple convolutional neural network for prediction, where all the convolutional operators are implemented as 1×1 kernels to remove semantic clues. In its original design, it directly predicts the global illuminant. We, however, modify it to predict two illuminants.

Experiment and results

All the algorithms were implemented and evaluated in Python. Learning-based methods were re-implemented in PyTorch [21] to ensure compatibility with our codebase when required.

For training and testing, all the images were resized to a resolution of 256×256. In the training phase, data augmentation techniques such as random cropping and random illuminant scaling were applied to improve model generalization.

The learning-based models were optimized using the AdamW optimizer [22] with a weight decay of 0.01. A linear

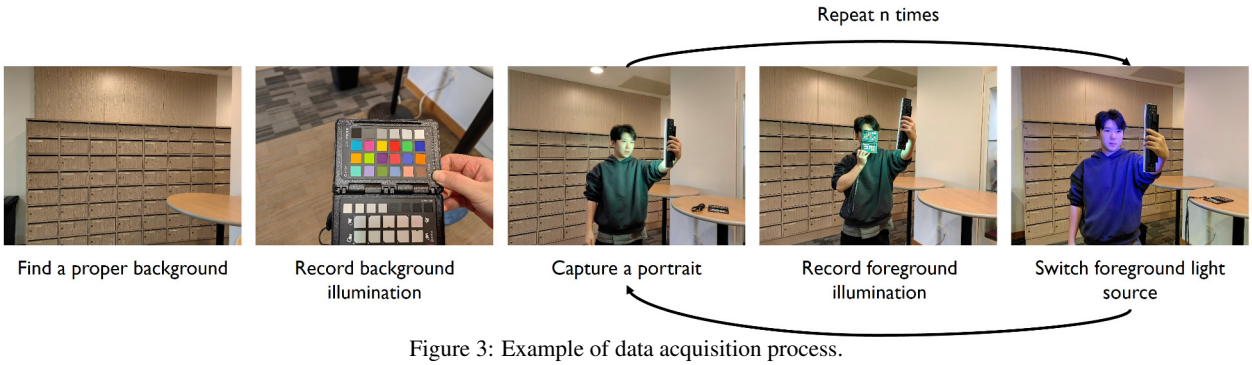


Figure 3: Example of data acquisition process.

learning rate schedule was employed, starting from 0.002 and decaying to 0.0002 over 500 epochs. A fixed batch size of 32 was utilized throughout the training. To mitigate the effects of random initialization and sampling bias, a five-fold cross-validation strategy was adopted.

A range of statistics- and learning-based methods were evaluated using standard metrics based on the angular error (Equation 3), such as: the average error, the median error, the tri-mean of all errors, the average of the lowest 25% of errors, and the average of the highest 25% of errors. For the statistics-based methods, no modifications were applied to enable them to output multiple estimations, with the estimated illuminant considered as both the first and second illuminant in the scene.

$$l_{ae} = \frac{180}{\pi} arccos\left(\frac{\hat{L} \cdot L}{||\hat{L}|| \cdot ||L||}\right) \tag{3}$$

Since there are two illuminants in each scene, the average of the angular error values was calculated for the final comparison, with the results summarized in Table 1. It can be found that all the statistics-based methods have much worse performance, suggesting the failure for multi-illuminant conditions. Among the learning-based methods, FC4 [23] had the best performance, while both C5 [14] and OneNet [17] had worse performance, with FC4 leverages full-image information and C5 and OneNet do not rely on semantic information.

Table 1: Summary of the performance of various methods, in terms of the angular error, evaluated on PolyU Face Color Constancy Dataset, with the best result of each metric is highlighted in red.

Method	Avg Mean	Avg Med.	Avg Tri.	Avg Best 25%	Avg. Worst 25%
White Patch [7]	14.98	15.55	14.81	6.21	23.98
Gray World [1]	9.70	8.19	8.50	4.16	18.13
Bright Pixels [8]	10.48	8.93	9.44	4.47	18.92
Gray Edge [4]	9.84	8.47	8.61	4.06	18.17
Shades of Gray [3]	12.80	11.75	12.07	5.84	21.64
PCA CC [5]	10.90	9.36	9.92	4.92	19.21
Gray Pixels [9]	11.43	9.89	10.27	4.28	20.93
Gray Index [10]	10.87	9.19	9.87	3.85	20.51
LSRS [24]	9.68	8.21	8.46	4.04	18.14
C5 (m=1) [14]	5.43	3.24	3.73	1.29	13.51
OneNet [17]	3.59	2.52	2.64	1.09	8.01
FC4 (SqueezeNet) [11]	1.12	0.92	0.94	0.45	2.16

Discussion

It remains a topic of debate whether semantic information is useful for illuminant estimation. Methods, such as FC4 [23], CLCC [16], and C4 [15], were found to effectively deal with the ambiguities between the illuminant and object colors, while other methods, such as CCC [12], FFCC [13], C5 [14], and OneNet [17], were found to achieve superior results without leveraging semantic information on the datasets (e.g., Cube+ [25], INTEL-TAU [18], and NUS-8 [5]). In our previous work [26], we also found that the semantic information does not enhance performance in pure color scenes. Our results presented here, however, suggest that the semantic information can significantly improve the performance for multi-illuminant scenes containing faces. This could be due to two reasons. Firstly, human faces or skin tones serve as a stable and reliable cue for illuminant estimation, as demonstrated in [19]; secondly, region separation is inherently required for multi-illuminant estimation. Notably, the learning-based methods that process full images are able to locate the regions of face and infer the dominant illuminant color from these areas, even without the ground-truth illumination maps. This can be verified by activating the weighting mechanism in FC4 [23] and visualizing the coefficient map, as illustrated in Figure 4 as an example.

(a) (b) (c)

Figure 4: Example of showing the effectiveness in isolating regions of interest (ROI). (a) RAW input image; (b) ROI of the first illuminant; (c) ROI of the second illuminant.

Pseudo pixel-wise white balance

Since the semantic-based models were found to be able to isolate the face region from the image, it enables pixel-wise adjustments in post-processing, as illustrated in Figure 5. In our experiment, the image encoder was based on the backbone of FC4 [11], though it can be replaced with other architectures that are capable of extracting semantic information. The post-processing converts the original attention maps A into weight maps W, which

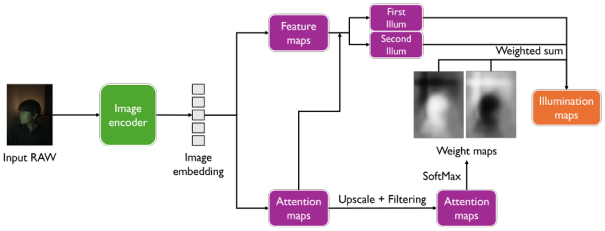


Figure 5: Illustration of post-processing for deriving a pseudo illumination map.

are up-scaled to the size of the input RAW image. Since the attention maps highlight the regions of interest, overlapping areas and ignored regions may exist at the same time. The Softmax function is applied along the illumination axis to generate the weight maps, assigning each pixel a weight corresponding to the two illumination. To control the blending strength of the two illumination, the attention maps are first scaled to a range of $[0,1]$ using their respective maximum values. This scaled value is then multiplied by a positive blending factor b before applying the Softmax function. This can be expressed as:

$$W(x) = Softmax(\frac{A(x)}{A_{max}}b) \tag{4}$$

A filter is then used to smooth the edges for a better appearance, with the final weight map showing the coefficients of the two illuminants for each pixel. Combined with the predicted illuminant colors, these maps are used to compute the final illumination maps. Figure 6 shows an example of applying the pseudo illumination map. In our design, the blending parameter b balances the visual artifacts and the correctness of the white-balanced results, a lower b value reducing the artifacts at the edges by using the average of the two illuminants and a higher b value introducing noticable artifacts by using one of the two illuminants, with Figure 7 showing an example.

(a) (b) (c) (d)

Figure 6: Example of psudo pixel-wise white balance. (a) White balance using the foreground illuminant only; (b) white balance using the background illuminant only; (c) pixel-wise white balance using a pseudo illumination map; (d) the corresponding pseudo weight map for deriving the psudo illumination map.

Conclusion

In this work, we present a color constancy dataset. The dataset includes 1299 portrait images under multi-illuminant conditions, which were specifically collected for enhancing performance of white balance algorithms under multi-illuminant conditions. It addresses a critical gap in the existing literature which often overlooks the significance of human facial features in color constancy. Through a systematic evaluation on the state-of-the-art methods, we propose a simple method to adapt the methods for performing white balance on portrait images.

(a) (b) (c) (d)

Figure 7: Example of how the blending parameter b affect the final image. (a) b=0.5; (b) b=3; (c) b=5; (d) b=7.

References

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

[2] Robert W Hunt. *The reproduction of colour*. John Wiley & Sons, 2005.

[3] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 12, pages 37–41. Society of Imaging Science and Technology, 2004.

[4] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007.

[5] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *Journal of the Optical Society of America A*, 31(5):1049–1058, 2014.

[6] Edwin H Land and John J McCann. Lightness and retinex theory. *Journal of the Optical Society of America A*, 61(1):1–11, 1971.

[7] David H Brainard and Brian A Wandell. Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, 3(10):1651–1661, 1986.

[8] Brian Funt and Lilong Shi. Maxrgb reconsidered. *Journal of Imaging Science and Technology*, 56(2):020501–1, 2012.

[9] Kai Fu Yang, Shao Bing Gao, and Yong Jie Li. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2254–2263, 2015.

[10] Yanlin Qian, Joni Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8062–8070, 2019.

[11] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017.

[12] Jonathan T Barron. Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015.

[13] Jonathan T Barron and Yun Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, pages 886–894, 2017.

[14] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1981–1990, 2021.

[15] Huanglin Yu, Ke Chen, Kaiqi Wang, Yanlin Qian, Zhaoxiang Zhang, and Kui Jia. Cascading convolutional color constancy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12725–12732, 2020.

[16] Yi Chen Lo, Chia Che Chang, Hsuan Chao Chiu, Yu Hao Huang, Chia Ping Chen, Yu Lin Chang, and Kevin Jou. Clcc: Contrastive learning for color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8053–8063, 2021.

[17] Ilija Domislović, Donik Vršnak, Marko Subašić, and Sven Lončarić. One-net: Convolutional color constancy simplified. *Pattern Recognition Letters*, 159:31–37, 2022.

[18] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-tau: A color constancy dataset. *IEEE access*, 9:39560–39567, 2021.

[19] Simone Bianco and Raimondo Schettini. Adaptive color constancy using faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1505–1518, 2014.

[20] Dongyoung Kim, Jinwoo Kim, Seonghyeon Nam, Dongwoo Lee, Yeonkyung Lee, Nahyup Kang, Hyong-Euk Lee, ByungIn Yoo, Jae-Joon Han, and Seon Joo Kim. Large scale multi-illuminant (lsmi) dataset for developing white balance algorithm under mixed illumination. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2410–2419, 2021.

[21] A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

[22] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[24] Shaobing Gao, Wangwang Han, Kaifu Yang, Chaoyi Li, and Yongjie Li. Efficient color constancy with local surface reflectance statistics. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pages 158–173. Springer, 2014.

[25] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017.

[26] Shuwei Yue and Minchen Wei. Color constancy from a pure color view. *Journal of the Optical Society of America A*, 40(3):602–610, 2023.

Author Biography

Yuyang Liu is a PhD student at The Hong Kong Polytechnic University. His research mainly focuses on color constancy and computational photography.

Minchen Wei is a professor at The Hong Kong Polytechnic University. His research mainly focuses on color science and imaging science.