Supplemental Materials of Optimizing Illuminant Estimation in Dual-Exposure HDR Imaging

Mahmoud Afifi, Zhenhua Hu, and Liang Liang

Google
{mafifi, zhenhuahu, annyliang}@google.com

1 Analogy to Chromagenic Color Constancy

In the main paper, we draw an analogy to the chromagenic color constancy theory [8–10]. Our argument is grounded in the empirical findings from [1, 10], indicating that even when the chromagenic filter constraints are not satisfied, color mapping matrices computed to map between the colors of the main camera and a filtered/second camera still exhibit a certain degree of correlation with the scene illuminant. Practically speaking, such mapping matrices capture the color differences (or "distortion") between the main camera and the filtered/second camera.

Our analogy is based on the observation that, in a dual-exposure setup, the long-exposure image, I_l, and the short-exposure image, I_s, exhibit varying levels of chromatic differences and distortions based on the scene irradiance per color channel. We illustrated in the main paper the variations in the red, green, blue ratios between long and short exposure images and showed that chromatic histograms exhibit differences in similarity between the two images. We also demonstrated that these differences can vary based on the scene lighting condition.

In Fig. 1, we present a similar study conducted on the two-camera dataset [1], which includes two cameras from a Samsung smartphone device. Comparing Fig. 1 with the corresponding figure (Fig. 2) in the main paper, we observe that both cases share a similar level of differences based on the lighting condition, albeit with less extent in the case of dual-exposure imaging. Thus, we draw our analogy by employing a 3×3 color matrix that maps the rgb-chromaticity of I_l and I_s along with the covariance matrix of the ratio between each color channel in I_s and I_l to build our dual-exposure feature (DEF).

It is important to note that analogizing I_s and I_l to images with and without a color filter does not imply their identity. While both the chromagenic color constancy theory and our method rely on color differences between paired images of the same scene, the use of two different camera response functions in chromagenic color constancy introduces a high level of color differences. In contrast, the dual-exposure case exhibits a lower level of color differences but remains valuable for our task. The disparities between I_s and I_l in dual-exposure scenarios arise from multiple factors, including varying photon counts captured by the camera

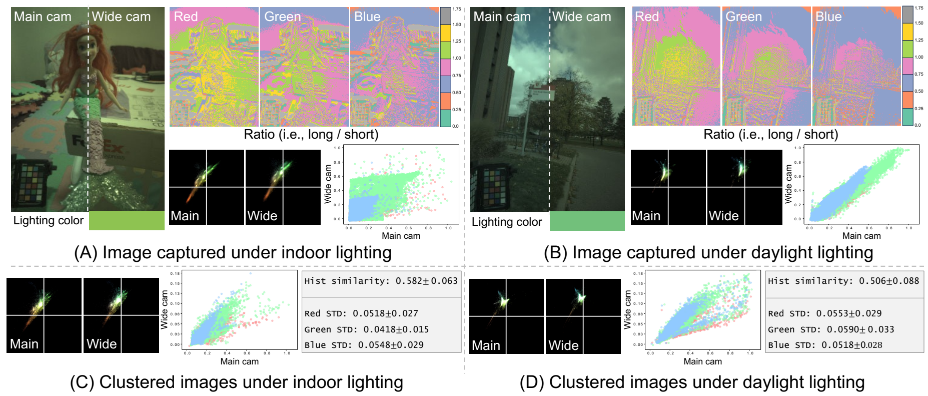

Fig. 1: In the main paper, we drew an analogy to chromagenic color constancy, demonstrating that both cases (i.e., two cameras and dual exposure capturing) result in variations per color channel, and these differences are linked to scene irradiance and camera response function. Here, we present images captured by two cameras from [1]. It can be observed that similar variations to the corresponding figure in the main paper in each channel occur due to the camera response function per channel. Moreover, spatial variations are noticeable based on scene irradiance and camera response function. (A) and (B) show scenes captured under indoor and outdoor lighting, respectively. In (C) and (D), we present the average rg-chromaticity histogram and aggregated red, green, and blue pixel values from 25 images sharing similar lighting conditions in (A) and (B), respectively.

and the non-linear nature of the camera response function (where each color channel is formulated by a different non-linear response function). As a result, these differences lead to variations in chromaticity, noise levels, and saturation between I_s and I_l. Collectively, these factors provide valuable clues derived from the correlation between dual-exposure images, I_s and I_l, aiding in the accurate estimation of the illuminant in a scene.

2 Additional Details

2.1 Mapping matrices

Our DEF employs a 3×3 matrix that maps between the rgb-chromaticity values of images I_s and I_l. In the main paper, we presented ablation studies that utilized different mapping matrices between the chromaticity values of I_s and I_l. Specifically, we evaluated using the geometric affine transformation instead of the linear mapping matrix. Here, we use I_s^ν and I_l^ν to refer to the rgb-chromaticity of the long and short-exposure images. The affine transformation matrix, denoted as M, between I_s^ν and I_l^ν, after appending an additional constant 1 to the rgb-chroma triples, can be computed as follows:

$$M = \begin{bmatrix} \alpha R_{\text{aff.}} & T_{\text{aff.}} \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

$$T_{\text{aff.}} = \text{centroid}(I_s^{\nu}) - 2\ \text{centroid}(I_l^{\nu}) \tag{2}$$

$$\alpha = \left\| I_l^{\nu'} \right\| / \left\| I_s^{\nu'} \right\|, \tag{3}$$

$$R_{\text{aff.}} = UV^T, \tag{4}$$

where $I_s^{\nu'}$ and $I_l^{\nu'}$ refer to centered values of I_s^{ν} and I_l^{ν} obtained by subtracting the centroids of I_s^{ν} and I_l^{ν}, respectively. $\mathbf{0} \in \mathbb{R}^3$ is a zero vector, U and V are 3×3 matrices, and S is a 3×3 diagonal matrix. U, S, and V can be obtained via singular value decomposition of the matrix multiplication of $I_s^{\nu'}$ and ${I_l^{\nu'}}^T$. Since the last row of M is fixed, we excluded it from the color matrix, C_c used in our DEF.

We also explored the use of a 3×3 homography matrix as an alternative to the 3×3 linear mapping matrix discussed in the main design of our method. Homography mapping has demonstrated its utility in various color applications [6, 7]. The homography matrix is computed to map between $[r, g, 1]^T$ rg-chromaticity values of long and short-exposure images. Based on our results, the linear transformation outperforms both geometric transformation and homography mapping.

2.2 Exposure-Based Convolutional Color Constancy

In the main paper, we discussed a modification to the existing convolutional color constancy (CCC) framework by incorporating our DEF. The DEF is processed by a lightweight multilayer perceptron (MLP) that produces weighting factors to linearly interpolate between a set of learnable biases, generating DEF-based biases for use in the CCC. We referred to this modified version of CCC as exposured-based CCC, or ECCC for short. In our experiments we used a 64×64 histogram (also we presented an ablation study on using ECCC with 32×32 histograms) for ECCC and other CCC methods [2, 4]. The histogram, H, is computed as described in the following equation:

$$H(u, v) = \sum_{t=1}^{k} \left\| I^{(t)} \right\| \left[|u_t - u| \le \varepsilon \wedge |v_t - v| \le \varepsilon \right], \tag{5}$$

where k refers to the total number of pixels in the image, $\varepsilon = (b_{\max} - b_{\min})/h$, with $b_{\max} = 2.85$ and $b_{\min} = -2.85$ as the histogram boundary values. In ECCC, in contrast to FFCC [4] and C5 [2], only colors from long-exposure and short-exposure images (I_l and I_s) are used to create histograms, excluding edge color histograms for simplicity. Specifically, we utilized two histograms, H_l and H_s, which represent the uv chroma values of I_l and I_s, respectively, and thus, two

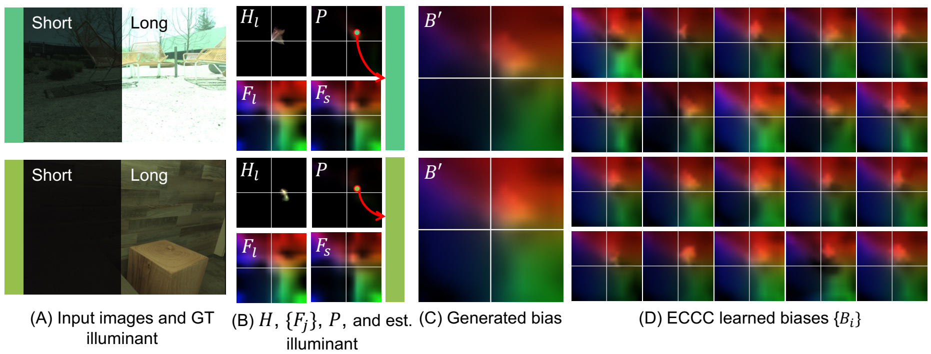

Fig. 2: This figure shows the generated bias map alongside the learned filters of the ECCC. In (A), we show a pair of input raw images captured with long and short exposure times, along with the ground-truth illuminant color. In (B), we show the histogram of image taken with long exposure (noting that our design employs the histograms of both short and long exposure images), the learned global filters F_j ($j \in l, s$), the probability map P, and the estimated illuminant color based on P. In (C), the generated bias is shown. (D) demonstrates the ECCC learned bias filters that are linearly interpolated based on the produced weights of the MLP using the input DEF associated with each pair of images.

convolutional filters, F_l and F_s, were learned in ECCC. Similar to FFCC and C5, FFTs are employed when convolving F_l and F_s over H_l and H_s, respectively.

To train ECCC, we used additional smoothness loss terms to encourage smoothness in the learned filters and biases. These smoothness terms can be described as follows:

$$S_B\left(B\right) = \lambda_B\left(\left\|B'_{\mathrm{up}} * \delta_u\right\|^2 + \left\|B'_{\mathrm{up}} * \delta_v\right\|^2\right), \tag{6}$$

$$S_F\left(\{F_j\}\right) = \lambda_F \sum_j \left(\left\|\uparrow\left(F_j\right) * \delta_u\right\|^2 + \left\|\uparrow\left(F_j\right) * \delta_v\right\|^2\right), \tag{7}$$

where δ_u and δ_v are 3×3 horizontal and vertical Sobel filters, respectively, and $\lambda_B = 0.01$ and $\lambda_F = 0.02$ are hyperparameters to control the strength of smoothness loss terms.

Figure 2 shows two examples of generated biases alongside the learned n biases (with $n = 20$). The figure also displays the learned convolutional filters for both histograms of images captured with long and short exposures (I_l and I_s). The convolutional filters (F_l and F_s) remain fixed in the model, while the bias dynamically changes based on the input DEF.

2.3 Dataset

As discussed in the main paper, we collected a dataset of multi-exposure raw images with ground-truth illuminant colors for training and evaluating our method.

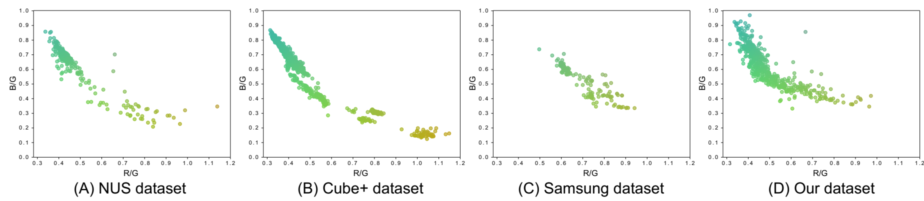

Fig. 3: Ground truth illuminant colors of the dataset used in our paper and other datasets. (A) NUS dataset [5]. (B) Cube+ dataset [3]. (C) Samsung dataset [1]. (D) Our dataset. For the NUS and Samsung datasets, we display the ground truth from a single camera: Canon EOS-1Ds for NUS and the main camera for Samsung.

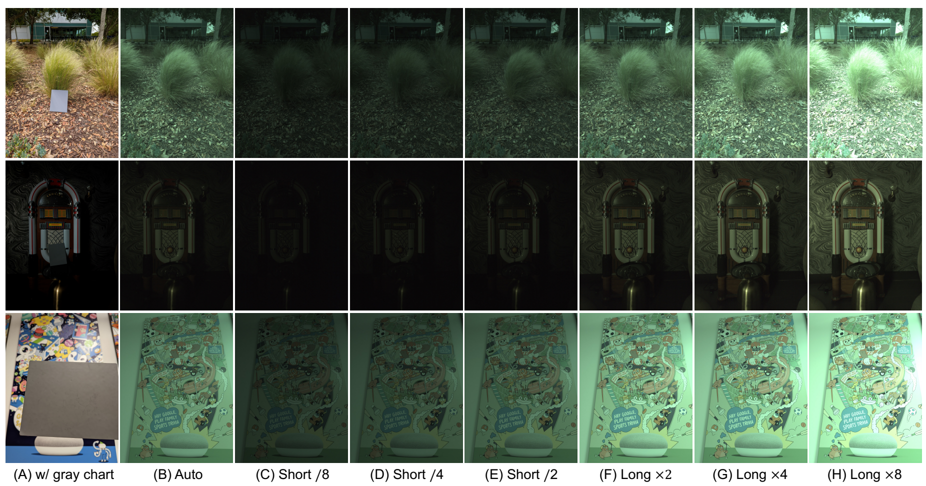

Fig. 4: Additional examples from the dataset used in this work. For each scene, we captured the scene with a gray calibration object placed in the scene to obtain the ground-truth illuminant (A) and captured the scene using different exposure settings without the gray object (B-H). The terms 'short $/e$' (C-E) and 'long $\times e$' (F-H) refer to multiplying and dividing auto exposure time by a factor e, respectively. The first image in (A) is displayed in sRGB, while the rest are shown in raw RGB space.

Figure 3 illustrates the distribution of R/G and B/G values for the ground-truth illuminant colors in the collected dataset. We also present illuminant distributions from other datasets (NUS [5], Cube+ [3], and Samsung [1]). Our dataset exhibits reasonable diversity, sometimes better, as observed when comparing with the Samsung dataset [1]. Notably, our dataset does not lack examples for certain regions in the Planckian-like curve, unlike the NUS and Cube+ datasets [3, 5]. Additional example images from our dataset are shown in Fig. 4.

References

1. Abdelhamed, A., Punnappurath, A., Brown, M.S.: Leveraging the availability of two cameras for illuminant estimation. In: CVPR (2021)
2. Afifi, M., Barron, J.T., LeGendre, C., Tsai, Y.T., Bleibel, F.: Cross-camera convolutional color constancy. In: ICCV (2021)
3. Banić, N., Koščević, K., Lončarić, S.: Unsupervised learning for color constancy. arXiv preprint arXiv:1712.00436 (2017)
4. Barron, J.T., Tsai, Y.T.: Fast Fourier color constancy. In: CVPR (2017)
5. Cheng, D., Prasad, D.K., Brown, M.S.: 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. Finlayson, G., Gong, H., Fisher, R.B.: Color homography: Theory and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(1), 20–33 (2017)
7. Finlayson, G.D., Gong, H., Fisher, R.B.: Color homography color correction. arXiv preprint arXiv:1607.05947 (2016)
8. Finlayson, G.D., Hordley, S.D.: The chromagenic colour camera and illuminant estimation. In: Color Imaging Conference (2005)
9. Finlayson, G.D., Hordley, S.D., Morovic, P.: Chromagenic colour constancy. In: Congress of the International Colour Association (2005)
10. Finlayson, G.D., Hordley, S.D., Morovic, P.: Colour constancy using the chromagenic constraint. In: CVPR (2005)