Investigation of impacting factors on camera calibration for spectral sensitivity estimation

Jueqin. Qiu and Haisong. Xu*

State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China

ABSTRACT

The acquisition of camera spectral sensitivity has become a hot issue in recent years because of its fundamental role in image processing and color reproduction. Although the estimation method based on camera response formation model has many advantages, the implementation procedures for spectral characterization are still tedious and time-consuming. In order for the improvement of calibration efficiency, it was investigated in this study that how the changes of the training samples selection would influence the estimation results and how many color samples at least should be included to achieve a high-fidelity color reproduction. The choice of the illuminants was also discussed to test whether the accuracy of the estimated spectra would degrade at the wavelength ranges with relatively low SPD of illuminant. Differing from the previous methods using relative errors in camera RGB space as cost function, we combine the colorimetric characterization with the spectral sensitivity estimation so as to minimize the estimation errors in device-independent color space like CIE XYZ. The detailed comparisons indicate that this modification could produce better perceived performance.

Keywords: Spectral sensitivity estimation, camera calibration, sample selection, impacting factor

1. INTRODUCTION

Determining the relation between camera output responses and the corresponding scene's radiometric quantities is important for photographic calibration and photometric measurements. The camera response is mainly determined by the spectral radiance of the scene being imaged and the spectral sensitivity of the sensor. As the spectral sensitivity of the camera plays a significant role in images formation, considerable attention has been paid to its capturing, which is also known as the spectral characterization.

The traditional method to acquire the spectral sensitivity of an imaging device is to use monochromator to generate near-monochromatic lights in sequence spanning the visible range and simultaneously record the responses of the camera[1]. Besides the monochromator method, some researchers focus on estimating camera spectral sensitivity by using imaging models and camera responses. The prevailing methods for estimating camera spectral sensitivity can be grouped into two categories, i.e., the algebraic method specified with constrains (denoted as AC hereafter) and the basis function method (denoted as BF hereafter). The AC method generally uses Moore–Penrose pseudoinverse to find the solution minimizing the cost function, being subject to some constrains of, for example, positivity, Tikhonov Regularization[2], Wiener filter[3], and first/second order derivatives[4], in order to satisfy some conditions like nonnegativity, smoothness, unimodality, and band-limitedness. The BF method represents the camera spectral sensitivity curves with a linear combination of some basis functions, and the methods of this category vary depending on the choice of the basis functions.

To obtain more generalized results, the AC method was selected in this study to acquire the camera spectral sensitivity. To our knowledge, the discussions about spectral characterization have been limited to minimizing the algebraically relative errors, rather than the visual perceived errors, between the actual camera responses and the predicted ones reconstructed from the estimated camera spectral sensitivity. For most applications in photography and photometry, since the responses in device-dependent space would have to be transformed to device-independent color space, the minimization of the mean CIEDE2000 (ΔE_{00}) color difference between the actual and the predicted camera responses was set as the cost function.

2. METHODS

2.1 Response Model

A Nikon D3x DSLR camera was employed in the experiments, mounted with a Nikkor AF-S 24-120 f/4 ED VR zoom lens. The lens was set to the focal length of 24mm and f-number of 4 throughout this study.

* Corresponding author: Haisong Xu, chsxu@zju.edu.cn

Whenever the camera responses are mentioned in this article, it refers to the triplet RGB values reordered from a quaternion in the sensor, to be specific, RGGB Bayer pattern in D3x. The response of the green channel in a triplet is the average of two green pixels in a quaternion.

The camera responses formation model is constructed in the form as follows:

$$\hat{\mathbf{p}} = \left[g \cdot f \cdot \Delta\lambda \cdot \mathbf{C} \cdot \mathbf{S}^{\mathrm{T}} \cdot \mathbf{l} + c_1 \right]^{\beta} + c_2 \tag{1}$$

where $g = \left(\pi \cos^4 \alpha / 4F^2\right)$ is a geometric constant only depending on the position of the pixels or regions $\mathbf{x}$ on the sensor to be investigated, f is the function of exposure time T and ISO sensitivity and can be simply expressed as $f = T \cdot (\text{ISO value}/100)$ here, $\mathbf{l}$ is a vector representing the spectral radiance on the location $\mathbf{X}$ in 3D space. $\mathbf{x}$ and $\mathbf{X}$ are related by the general projective camera model [5], $\Delta\lambda$ is the wavelength interval in accordance with the sampling interval of $l(\lambda)$. The camera spectral sensitivity $\mathbf{S} = [\mathbf{S}^R, \mathbf{S}^G, \mathbf{S}^B]$ is the element-wise product of the spectral transmissions and quantum efficiency, i.e., $\mathbf{S}^{(k)} = \mathbf{T}_o \mathbf{T}_c^{(k)} \mathbf{S}_Q$, where $T_o(\lambda)$ is the transmission of optical system including lens, low-pass filter, IR filters, etc., $T_c^{(k)}(\lambda)$ is the transmission of the color filter on $k=R,G,B$ channel. The quantum efficiency $S_Q(\lambda)$, in $\mathrm{m}^2 \cdot \mathrm{w}^{-1} \cdot \mathrm{s}^{-1}$, is informally defined here as the sensitivity of the sensor converting spectral irradiance to the amount of electrons.

The crosstalk matrix $\mathbf{C}$ is added into our model to take account of the crosstalk among the adjacent pixels:

$$\mathbf{C} = \begin{bmatrix} c_{RR} & c_{GR} & 0 \\ c_{RG} & c_{GG} & c_{BG} \\ 0 & c_{GB} & c_{BB} \end{bmatrix} \tag{2}$$

where c_{kk} weights how many photons arriving at the pixel k come from the color filter k over that pixel, and $c_{kk'}$ denotes the crosstalk coefficient from pixel k to k'. Since red pixels and blue pixels are always located diagonally, $c_{RB} = c_{BR} \equiv 0$.

2.2 Optimization Method

The interior-point method is selected as the algorithm to solve the constrained nonlinear optimization problem, as the response formation model, Eq. (1), is a power function and the gradient can be hardly provided during optimization. Similar to other nonlinear optimization algorithms, the interior-point method requires the initial guess to start the minimum search. The initial estimation of camera spectral sensitivity is obtained analogous to Barnard and Funt's work [4], but with the L-curve method to find the suitable regularization parameter [6]. During our optimization, the iteration would be programmingly stopped at the number of function evaluations exceeding 10000 times.

As suggested by Finlayson *et al.* [7], the root polynomial color correction (RPCC) regression is adopted to perform the colorimetric characterization that transforms the camera responses from RGB space into CIEXYZ and then CIELAB color space, so that the ΔE_{00} color difference is evaluated. Therefore, the cost function in this study is defined by

$$\mathcal{F} = \frac{\sum_{i=1}^{N} E_{00}\left(\mathbf{M}\boldsymbol{\rho}_i, \mathbf{M}\hat{\boldsymbol{\rho}}_i\right)}{N} + \lambda \max\left\{E_{00}\left(\mathbf{M}\boldsymbol{\rho}_i, \mathbf{M}\hat{\boldsymbol{\rho}}_i\right)\right\}. \tag{3}$$

where $\boldsymbol{\rho}_i$ and $\hat{\boldsymbol{\rho}}_i$ are the root-polynomial expansions of actual and predicted response of i^{th} sample, respectively. In our experiment, the root-polynomial expansions of 4^{th} degree were used in consideration of the accuracy.

3. RESULTS AND DISCUSSION

3.1 Sample Selection

The X-Rite ColorChecker Digital SG (DSG), illuminated by D65, was selected as the training set to extract the camera responses and their corresponding spectral radiances. From the 140 color patches on DSG color checker, 44 repeated neutral colors around the edge were excluded, thus a total of 96 color samples were adopted as the training and testing colors.

The leave-one-out cross-validation (LOOCV) was used to investigate how the changes of the training samples selection would influence the estimation results and how many color samples at least should be included to achieve a high-fidelity color reproduction. LOOCV selected 95 patches from the totally 96 patches as training samples to obtain the optimized parameters in Eq. (1), then used the remaining one patch as the testing sample to evaluate the performance of the spectral characterization. This process was repeated 96 times by only altering the choice of the testing sample, generating 96 subsets of patches, each of which consisted of 95 training samples. Different subsets of training samples were ranked according to their generalization abilities, i.e., the color difference of the predicted test sample. Poorer prediction result of one testing sample means that this sample varies significantly from others and should be excluded from the estimation of the camera spectral sensitivity to avoid additional noise.

Given the anticipant number of samples, *k*, being enough to achieve good color reproduction, the worst (96-*k*) samples that produced poorest color difference results were excluded from training phase, then the optimal *k* samples were utilized to reestimate the parameters in the camera response formation model. Finally, the generalization error was evaluated based on the mean ΔE_{00} color difference of the excluded (96-*k*) samples as illustrated in Figure 1, along with the standard deviations, as the function of the number *k* of training samples. It can be fairly deduced that as long as the training samples are appropriately selected, only a few color patches (less than 10) are necessary to obtain a good performance of the spectral characterization. And the spectral radiances of the optimal 24 samples with the highest rank during the LOOCV are plotted in Figure 2.

According to the testing results, the response formation model, with the crosstalk matrix and nonlinearity function added, has been verified to be of good performance to predict camera responses, given the spectral radiance of the target surface. When the initial guess of the camera spectral sensitivity is elaborately provided, a fast convergence can be achieved for the color difference between the actual and the predicted responses. During our optimization, the derivative of the cost function converged at below 1% after approximately 40 iterations.

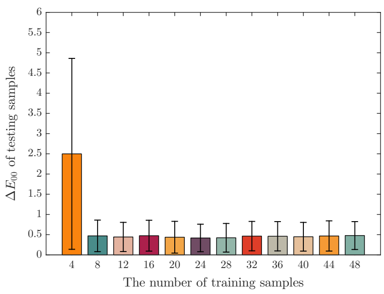

Figure 1. The mean color difference, along with the standard deviation, of the (96-*k*) testing samples.

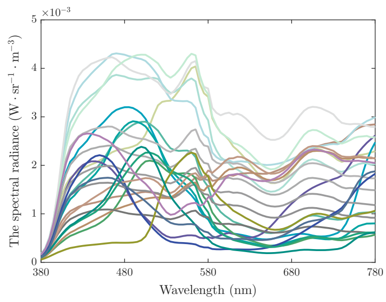

Figure 2. The spectral radiance of the optimal 24 samples obtained from LOOCV.

3.2 Illuminant Selection

To investigate how the spectral radiances of training samples would influence the performance of the spectral characterization, we replaced D65 in the aforementioned experiment with illuminant A and repeated all the captures and measurement processes. As mentioned above, a small amount of training samples could generate a satisfactory result, the X-Rite ColorChecker Classic was adopted for saving the experimental time, from the totally 24 patches of which the color differences of testing samples, as the function of *k* according to the LOOCV, are shown in Figure 3. Apparently the performance of the spectral characterization using illuminant A is poorer than that using D65. And the estimated camera spectral sensitivity for D3x, using the optimal 16 training samples under D65 are demonstrated in Figure 4.

Considering that the training and testing samples in the previous experiments might be of high correlation due to their being illuminated by the identical illuminant, the cross-validation was carried out through the color samples lit by two different illuminants. Figure 5 illustrates the color differences between the actual and the predicted responses of the 24 color samples under illuminant A, in which the predicted responses were constructed using the estimated camera spectral sensitivity, crosstalk matrix and the nonlinearity parameters obtained from the optimal 12 samples under D65. And the corresponding results by exchanging the training and testing illuminants are shown in Figure 6.

Since an appropriate combination of the spectral radiances of the training samples is the most significant factor that determines the performance of the spectral characterization, the training samples and the illuminant should be carefully selected for the spectral calibration of the camera. As seen from

Figures 5-6, using illuminant A as light source performs poorly in comparison with D65, which is considered to be resulted from the low spectral radiance in the short wavelength region of illuminant A. Although the regularization was utilized to resist noise, the low SNR in the short wavelength region give rise to the fluctuation during iterations of the interior-point method, especially for the blue channel mainly located at the short wavelength side of the visible spectrum.

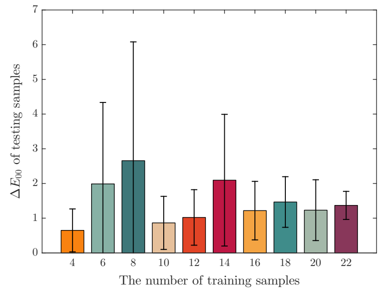

Figure 3. The mean color difference of the (24-k) testing samples.

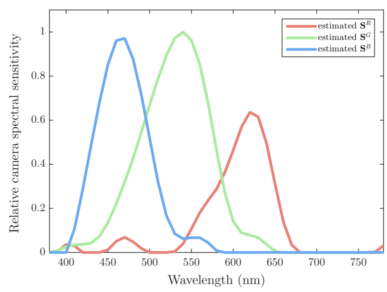

Figure 4. The estimated camera spectral sensitivity using 16 training samples under D65.

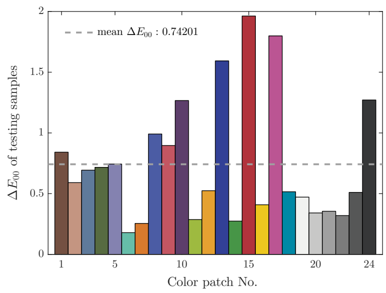

Figure 5. The color difference of the 24 testing samples under illuminant A with the training data obtained under D65.

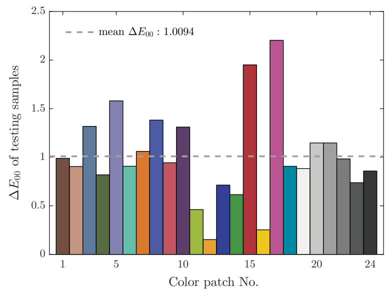

Figure 6. The color difference of the 24 testing samples under D65 with the training data under illuminant A.

4. CONCLUSION

Through the experiments, it has been proven that the procedure of the spectral characterization could be of convenience, i.e., using less than 10 training samples to achieve a high-fidelity color reproduction with the mean ΔE_{00} being below 1.0 for the testing samples. The selections of the training color samples as well as the illuminant are the critical factors for a reliable and accurate camera characterization. The SPD of the illuminant should be as flat as possible over the visible spectrum, so as to reduce perturbations for the calculation of the camera spectral sensitivity. If both the light source and training samples are elaborately selected, the spectral characterization of the digital camera could be carried out in a fairly straightforward way.

REFERENCES

1. M. M. Darrodi, G. Finlayson, T. Goodman, and M. Mackiewicz, "Reference data set for camera spectral sensitivity estimation," *JOSA A* **32**, 381–391 (2015).
2. B. Dyas, "Robust color sensor response characterization," in *Color and Imaging Conference* (Society for Imaging Science and Technology, 2000), Vol. 2000, pp. 144–148.
3. P. Urban, M. Desch, K. Happel, and D. Spiehl, "Recovering camera sensitivities using target-based reflectances captured under multiple led-illuminations," in *Proc. of Workshop on Color Image Processing* (2010), pp. 9–16.
4. K. Barnard and B. Funt, "Camera characterization for color research," *Col. Res. Appl.* **27**, 153–164 (2002).
5. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge university press, 2003.
6. P. C. Hansen, "Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms* **6**, 1–35 (1994).
7. G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Color Correction Using Root-Polynomial Regression," *IEEE Trans. on Image Process* **24**, 1460–1470 (2015).