Optimal Color Estimation by Combining Multiple Spectrodensitometer Measurements

Lei He; Library of Congress; Washington, DC, USA

Abstract

Cultural heritage digitization centers worldwide rely on quality management systems to insure the accuracy and consistency of their digitization efforts. Universal to all digitization quality management is the accurate measurement of color test targets. Testing at the Library of Congress has shown that significant variability exists between manufacturers and models of spectrodensitometers reading standard test targets. This variability has been measured commonly as ΔE2000 3 and as large as ΔE2000 4, leading to improper calibration of imaging systems and less than ideal color accuracy. In this paper we present two robust statistics-based algorithms that estimate the true color from multiple spectrodensitometer measurements, effectively reducing measurement error. The estimated colors can then be regarded as the ground truth for imaging device calibration and performance evaluation. Besides the true color estimation, the second algorithm in particular characterizes the performance of the devices by estimating their bias and variances.

1. Introduction

Color accuracy [1] is one of the critical factors in imaging quality analysis, which is extensively used in digital preservation and archiving applications. Color accuracy assessment is generally implemented by scanning or taking a picture of a standard target board, for example, Macbeth ColorChecker SG chart[1] or GoldenThread™ target[2] (see Figure 1), then comparing the derived image values with the pre-measured color patch values provided by the manufacturers. The metric to measure the color difference between two measurements is generally computed using the CIE ΔE2000 [2, 3]. Imaging devices with large ΔE2000 values should be recalibrated and profiled to achieve satisfactory results before production.

In practice, target board color values deviate from previously measured values for many reasons. To overcome this problem, users may measure the targets on their own with devices such as colorimeters or spectrodensitometers. However, this process assumes that these instruments produce accurate measurements when properly calibrated, which is not true due to the inter- and intra-variations of the devices. Such variations affect the true measured color values and produce inaccurate assessment results. To characterize such variations, color scientists have constructed regression models [4, 5, 6, 7] to depict different types of device error, for example, the photometric, wavelength, and bandwidth error. Both linear and nonlinear terms are constructed in the regression model to approximate these errors. To estimate the model parameters, a rather large set of standard tiles (e.g., > 400 tiles in [4]) of different materials with different gray scales and colors should be used as the training samples to fit the model. In addition, a reference device (e.g., Zeiss DMC-26 in [5] and Spectraflash 500 in [4]) is employed to produce the "ground truth" for the performance assessment of other devices. Such approaches have high complexity and experimental cost on sample collection. Moreover, these approaches are infeasible for general users who do not have access to those resources.

***Figure 1**. Examples of targets for device color accuracy assessment. Left: ColorChecker SG chart; Right: GoldenThread™ target.*

In this paper, we present two algorithms to estimate the true color values by optimizing the proposed objective functions that employ the robust statistics from multiple spectrodensitometer measurements. For each target patch, we conduct multiple measurements of the color with each device, and the output is the estimated true color for that patch. The first approach minimizes the total variance of the measurements. The total variance is constructed as the weighted sum of the individual device variances. This is a convex optimization problem with two constraints: (1) the valid range of each weight is restricted to [0, 1]; (2) the sum of the weights is 1. Linear programming is applied to compute the weights, and the true color is estimated as the weighted sum of the individual means of the multiple measurements of each device.

Our second algorithm is inspired by medical image segmentation applications, which combines multiple expert manual labels to derive the optimal estimation of the true object boundary [8, 9]. The objective function is the complete likelihood for the observed measurements and hidden data (true color values), given the estimated model parameters (bias and variance of each

---

1 ColorChecker Digital SG. http://xritephoto.com/ph_product_overview.aspx?ID=938

2 Image Science Associates. http://www.imagescienceassociates.com/

device). The expectation-maximization (EM) algorithm is employed to iteratively estimate the true color and the device parameters. This method produces not only the optimal color estimation, but also the bias and variance of each device, which provides a way to characterize the device performance.

This paper is organized as follows: Section 2 briefly introduces the background of color accuracy and statistics technologies. Section 3 presents our proposed algorithms for optimal color estimation. Experiment results on a ColorChecker SG chart and a GoldenThread™ target are presented in Section 4. We draw conclusions in Section 5.

2. Background

Imaging quality assessment is usually implemented by examining different reference regions with desired features on target boards. For example, slanted edges or periodic patterns of different frequencies [10] are always used to evaluate the imaging resolution and sharpness; series of gray scale step wedges with increased density for OECF derivation and noise estimation; and arrays of color patches for color accuracy assessment. In this work we focus on color accuracy assessment, and we utilize the Macbeth ColorChecker SG chart (140 patches) and GoldenThread™ Target (30 patches) for the assessment. As introduced in Section 1, color accuracy is generally measured by computing the CIE ΔE2000 value between the imaging results and the ground truth. There are both commercial and free tools available for the assessment task, for example, the websites of delt.ae[3] and Bruce Lindbloom[4], and DICE™ software.

Colorimeters, spectrodensitometers, or spectrophotometers are the devices generally used for color value measurements following the ISO standards [1, 11, 12]. For example, we may choose the illuminants (e.g., D50 or D65), data format (spectral response or tristimulus colors), and output color space (e.g., XYZ or LAB) with the device user interfaces. In practice, different devices produce different measurements at different times on the same target. A single device will often produce different consecutive measurements on the same target. This introduces both intra- and inter-variations on the measurements, which confuses users on the identification of accurate and robust devices.

Statistics have been used to construct regression models [4, 5] that characterize the performance of spectrodensitometers, based on which the true color values can be estimated. For example, seven types of systematic errors are characterized in [5] to construct the regression model, including photometric zero error ($R_t(\lambda) = R_m(\lambda) + B_0$), photometric linear scale error ($R_t(\lambda) = R_m(\lambda) + B_1R_m(\lambda)$), photometric nonlinear scale error ($R_t(\lambda) = R_m(\lambda) + B_2[1-R_m(\lambda)]R_m(\lambda)$), wavelength linear scale error ($R_t(\lambda) = R_m(\lambda) + B_3\mathrm{d}R_m/\mathrm{d}\lambda$), wavelength nonlinear scale errors ($R_t(\lambda) = R_m(\lambda) + B_4w_1(\lambda)\mathrm{d}R_m/\mathrm{d}\lambda$, and $R_t(\lambda) = R_m(\lambda) + B_5w_2(\lambda)\mathrm{d}R_m/\mathrm{d}\lambda$), and bandwidth error ($R_t(\lambda) = R_m(\lambda) + B_6\mathrm{d}^2R_m/\mathrm{d}\lambda^2$). Here $R_t(\lambda)$ and $R_m(\lambda)$ correspond to the true color and measured values, respectively. $w_1(\lambda)$ and $w_2(\lambda)$ are wavelength weighting functions, and B_i (i = 0, …, 6) are the model parameters to be fitted. As indicated in Section 1, a large number of sample colors are generally needed for a robust parameter fitting. Therefore those approaches are not feasible for general users due to the very high cost of resources. Furthermore, all of these approaches require a high-end device to produce the ground truth as the reference for the regression model construction, which may not be accessible to general users. In [13], the reference values are provided by the manufacturer of the 14 BCRA tiles, which are employed to characterize 9 different spectrodensitometers by estimating their accuracy (bias) and precision (variance).

Robust statistics have been extensively employed in numerous signal and image processing applications. For example, image total variation minimization [14] is a major approach for image denoising, which estimates the true signal from noisy observations. Image total variation is regarded as an approximation of the image noise, thus the minimum corresponds to the noise free image. This is a convex optimization problem, and gradient descent-based algorithms may be applied to obtain the solution. When expert labels of the object boundary are available, the complete data likelihood maximization [8] is applied to estimate the true image segmentation results from multiple expert inputs. The complete data likelihood is constructed as the joint probability of both data (i.e., observed labels or measurements) and model parameters (expert or device characterization parameters). Because the model parameters are unknown, the conditional expectation given the observable data using the current parameter estimation is obtained instead. Computing the conditional expectation of the original complete data likelihood is implemented by the E-step of the EM algorithm, and identifying the model parameters that maximize this function is referred to as the M-step. Motivated by those applications, we developed two new algorithms for the true signal (color) estimation from multiple observations (spectrodensitometer measurements).

3. Proposed Approaches

We propose two approaches to estimate the optimal color values from the measurements of multiple spectrodensitometers. Five spectrodensitometers are used in our experiments: X-Rite® 528, two X-Rite® i1Pro, Barbieri® Spectro LFP, and Konica Minolta® FD7. All of these devices are either new or recertified by the manufacturers for our experiments. Thus their measurements are regarded close to the true colors.

Our first approach is motivated by the total variation-based image denoising [14]. Given the measurements of the same target patch from a set of I (I = 5 in our application) spectrodensitometers, we define the true color of the target patch as a weighted sum of the average measurements (μ_i, i = 1, …, 5) of the spectrodensitometers, i.e.,

$$\sum_{i=1}^{I=5} \alpha_i \mu_i \ , \tag{1}$$

where α_i are the weighting parameters. We derive these parameters by minimizing the total variance,

$$\sum_{i=1}^{I=5} \alpha_i^2 \sigma_i^2 \ , \tag{2}$$

[3] http://delt.ae

[4] http://brucelindbloom.com/index.html?Eqn_DeltaE_CIE2000.html

where σ_i^2 are the variances of the devices. For each spectrodensitometer, we measure the target patch multiple times and compute the mean of the patch color (μ_i) and the corresponding variance (σ_i^2). To minimize Eq. (2), we further add the constraints of the parameters:

$$\sum_{i=1}^{I=5} \alpha_i = 1 \text{ and } 0 \le \alpha_i \le 1 .$$

This is a convex optimization problem that the optimization is guaranteed to converge to the global minimum. The linear programming technique is applied to optimize Eq. (2), then the derived parameters are used in Eq. (1) to obtain the estimated true color.

Our second approach employs the same principle in [8, 9] that estimate the true object boundary from a set of expert manual segmentation results. The objective is to maximize the complete data likelihood using the EM algorithm. Because the complete data likelihood consists of the unknown device parameters (i.e., accuracy and precision in [13]), we instead compute the expected value of a conditional probability density function for the true color given the measurements and previous estimates of the model parameters. Here we adopt the statistical terminologies to use variance (σ^2) and bias (β) to characterize the device performance. The objective function is

$$\theta^{(t)} = \arg\max_\theta E\left[\log \Pr(s, \tau \mid \sigma^2, \beta) \mid s, \theta^{(t-1)}\right] \quad (3)$$

where s and τ represent the measurements and ground truth, respectively. This conditional expectation is evaluated with respect to $p(\tau \mid s, \theta^{(t-1)})$. With the EM algorithm, we iteratively estimate the expectation and the device parameters. In the E-step, we can derive the joint variance of the devices and individual mean of each patch as:

$$\frac{1}{\sigma^2} = \sum_{i=1}^{I=5} \frac{1}{\sigma_i^2} \quad (4)$$

$$\mu_j = \frac{\sum_{i=1}^{I=5} \left(s_{ij} - \beta_i\right) / \sigma_i^2}{1/\sigma^2} \quad (5)$$

where σ_i^2 and β_i (i = 1, …, I) are the variance and bias of the spectrodensitometers. μ_j (j = 1, …, J) are the average color for the target patches, and s_{ij} are the measurement of the j-th patch by the i-th spectrodensitometer. It can be seen that the total variance is the harmonic mean of the individual device variance, and the mean color for the patch is an inverse of total variance-weighted sum of the difference between the measurement and the bias. In the M-step, we estimate the device parameters that maximize the conditional expectation of the complete data likelihood.

$$\beta_i^{(t)} = \frac{1}{J} \sum_{j=1}^{J} \left(s_{ij} - \mu_j^{(t-1)}\right) \quad (6)$$

$$(\sigma_i^2)^{(t)} = \frac{1}{J} \sum_{j=1}^{J} \left(s_{ij} - \beta_i^{(t)} - \mu_j^{(t-1)}\right)^2 + (\sigma^2)^{(t-1)} \quad (7)$$

The device bias is the average difference between the measurement and the previously estimated true color over all patches, and the device variance is a sum of the term of current variance estimation and the total variance of Eq. (4). In practice, we compute the initial $\mu_j^{(0)}$ and the $(\sigma_i^2)^{(0)}$ with the measurements, i.e., $\mu_j^{(0)}$ is the average of all the I spectrodensitometer measurements on the j-th patch, and $(\sigma_i^2)^{(0)}$ is similarly computed as the variance of the measurements of each device over all the patches, i.e., the average variance on all the target patches. Thus we start with Eq. (6) to compute Eq. (4) and (5), based on which we compute the Eq. (6) and (7) to continue the iteration until the final convergence.

4. Experiments

Our experiments include tests of the two algorithms on a ColorChecker SG chart and a GoldenThread™ Target. In these experiments we measure each target patch five times for the computation of μ_i and σ_i^2 in Eq. (1) and (2), as well as for the initialization of $\mu_j^{(0)}$ and the $(\sigma_i^2)^{(0)}$ in Eq. (4) and (6). We choose D50 as the illuminant for all the devices, and CIE LAB space as the output color values due to its device independence. After the true color estimation (i.e., Eq. (1) and Eq. (5)), we further conduct the hypothesis tests to compare the results of our two algorithms. Besides the true color estimation, the second algorithm directly characterizes the device performance by computing their variance and bias. In this presentation, in order to avoid endorsing a specific product, we instead use the letters *A*, *B*, *C*, *D*, *E* to represent the devices.

Our first experiment is to measure the color values on a GoldenThread™ Target, which consist of 18 color and 12 gray patches. For the first algorithm, after computing the μ_i and σ_i^2, we use a Matlab® optimization package to minimize Eq. (2), and derive the weighting parameters α_i. We then compute the bias β_i for each spectrodensitometer by comparing the measurements with the estimated true color (LAB), as shown in Table 1. They are the average values over all patches. It can be seen that the device *C* obtains the minimum variance and the device *D* has the smallest bias, see the bolded numbers. In Table 2, we present mean and variance of ΔE2000 over all patches by comparing the measurements with the estimated true color. Overall, the device *C* obtains the optimal measurements, i.e., closest to the true color.

For the second algorithm, Table 3 presents the directly estimated σ_i^2 and β_i for each spectrodensitometer using Eq. (6) and (7). We obtain the same results as that of the first algorithm, i.e., the device *C* has the smallest variance and the device *D* obtains the smallest bias compared with the true color. Similar to Table 2, we compute the mean and variance of ΔE2000 for the second algorithm results in Table 4. Again the device *C* obtains the smallest measurement error.

	A	*B*	*C*	*D*	*E*
L Var	0.0098	0.0048	**0.0005**	0.003	0.0071
A Var	0.0014	0.0203	**0.0002**	0.0015	0.0013
B Var	0.0031	0.05	**0.0008**	0.0036	0.0024
L Bias	0.7116	-0.2838	0.2069	**0.0289**	-0.2244
A Bias	-0.0347	0.2814	-0.0804	**-0.0186**	-0.058
B Bias	0.3988	0.5364	-0.0844	**-0.0776**	0.0818

Table 1. Spectrodensitometer bias and variance estimated by the total variance minimization (GoldenThread™ Target)

	ΔE2000 Mean	ΔE2000 Variance
A	0.7984	0.1382
B	1.3749	0.8221
C	**0.4984**	**0.1238**
D	0.6294	0.2101
E	0.5261	0.1229

Table 2. Error between the spectrodensitometer measurements and the estimated true color by the total variance minimization (GoldenThread™ Target)

	A	*B*	*C*	*D*	*E*
L Var	0.0695	0.4408	**0.0005**	0.606	0.2928
A Var	0.0481	2.5081	**0.0023**	0.7131	0.513
B Var	0.1171	1.7077	**0.0006**	0.4225	0.4286
L Bias	0.6238	-0.3716	0.1191	**-0.0589**	-0.3123
A Bias	-0.0526	0.2634	-0.0983	**-0.0366**	-0.0759
B Bias	0.2278	0.3654	-0.2554	-0.2486	**-0.0892**

Table 3. Spectrodensitometer bias and variance estimated by the EM algorithm (GoldenThread™ Target)

	ΔE2000 Mean	ΔE2000 Variance
A	0.5641	0.0398
B	1.2193	0.5213
C	**0.2256**	**0.0045**
D	0.6885	0.26
E	0.6443	0.0811

Table 4. Error between the spectrodensitometer measurements and the estimated true color by the EM algorithm (GoldenThread™ Target)

After the color estimation, we compare the estimated color values obtained by our two approaches. First we conduct the hypothesis tests (paired-sample *t*-test) that the two estimated colors are the same. Statistically, the two approaches produce the same results for all 30 patches, i.e., we cannot reject the null hypothesis that the difference between the two estimations is 0. We further compute the ΔE2000 between the two estimated true colors for each pair of the 30 patches. The mean and variance of ΔE2000 are 0.47 and 0.1, respectively. The average difference is less than 1 (i.e., the empirically used "just noticeable difference" of color), so we determine that the two approaches produce the same results.

Our second experiment is conducted on the ColorChecker SG chart, which consists of 140 color and gray patches. As in the first experiment, we estimate the true colors using our two approaches and characterize the spectrodensitometers by their variance and bias. Table 5-8 show the results similar to those on Table 1-4. With this experiment, it is not obvious to identify the best device with their variance and bias (Table 5 and 7). From Table 6 and 8, we can see that the device *C* obtains the best performance on color accuracy, i.e., smallest ΔE2000 compared to the true colors. In the end, the hypothesis tests are conducted to compare the two approaches, which produce statistically the same results on 134 out of the 140 patches. Computing the ΔE2000 over the 140 patches for the two estimated colors, the mean and standard variation of are 0.57 and 0.21. Again, they obtain the same results with ΔE2000 less than the just-noticeable difference (JND).

	A	*B*	*C*	*D*	*E*
L Var	0.0058	0.0146	0.0016	**0.0009**	0.0018
A Var	0.0021	0.0083	**0.0004**	0.0008	0.0022
B Var	**0.0027**	0.0334	0.0033	0.0036	0.0028
L Bias	0.5321	-0.3916	**-0.2085**	0.4505	0.3987
A Bias	-0.1332	**-0.0082**	-0.2018	-0.0834	-0.1333
B Bias	**0.0662**	0.4766	-0.2396	-0.1846	-0.1548

Table 5. Spectrodensitometer bias and variance estimated by the total variance minimization (ColorChecker SG Target)

	ΔE2000 Mean	ΔE2000 Variance
A	0.7633	0.261
B	1.3652	0.7811
C	**0.6357**	**0.2064**
D	0.7929	0.3803
E	0.8056	0.4411

Table 6. Error between the spectrodensitometer measurements and the estimated true color by the total variance minimization (ColorChecker SG Target)

	A	*B*	*C*	*D*	*E*
L Var	0.1998	0.6069	**0.0451**	0.2028	0.1837
A Var	**0.0067**	3.1996	0.0286	0.4133	0.5021
B Var	**0.0958**	2.585	0.1222	0.4773	0.4655
L Bias	0.3758	-0.5478	-0.3648	0.2942	**0.2425**
A Bias	**-0.0245**	0.1169	-0.0931	0.0253	-0.0246
B Bias	**0.0734**	0.4838	-0.2323	-0.1773	-0.1476

Table 7. Spectrodensitometer bias and variance estimated by the EM algorithm (ColorChecker SG Target)

	ΔE2000 Mean	ΔE2000 Variance
A	0.3894	0.0577
B	1.4011	0.5737
C	**0.3757**	**0.0281**
D	0.5989	0.0692
E	0.6041	0.0853

Table 8. Error between the spectrodensitometer measurements and the estimated true color by the EM algorithm (ColorChecker SG Target)

5. Conclusion

In this paper we present two algorithms to estimate the true colors from multiple measurements of different spectrodensitometers. Motivated by the image and signal processing applications, our approaches employ robust statistics to derive the optimum estimation. Hypothesis tests and the ΔE2000 comparison show that the two approaches produce the same results. Besides the true color estimation, we also characterize the device performance by their bias, variance and the error to the true color. In practice, in case users do not have multiple spectrodensitometers to derive the optimal color estimation, they may choose advanced devices, for example, the device *C* in our test, to measure the target as an estimation of the ground truth.

References

[1] ISO 13656:2000, Graphic Technology – Application of Reflection Densitometry and Colorimetry to Process Control or Evaluation of Prints and Proofs.

[2] A. R. Robertson, "Historical development of CIE recommended color difference equations", Color Research & Application, 15, 3, pg. 167–170. (1990).

[3] G. Sharma, W. Wu, E. N. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations", Color Research & Applications, 30, 1, pg. 21–30. (2005)

[4] D. Rich, D. Martin, "Improved model for improving inter-instrument agreement of spectrodensitometers", Analytical Chemica Acta, 380, pg. 263-276. (1999).

[5] R. S. Berns, K. H. Petersen, "Empirical modeling of systematic spectrodensitometric errors", Color Research and Application, 13, 4. pg. 243-256. (1988).

[6] R. S. Berns, L. Reniff, "An abridged technique to diagnose spectrodensitometric errors", Color Research and Application, 22, 1. pg. 51-60. (1997).

[7] R. S. Berns, M. J. Shyu, "Colorimetric characterization of a desktop drum scanner using a spectral model", J. Electronic Imaging, 4, 4. pg. 360-372. (1995).

[8] S. K. Warfield, K. H. Zou, W. M. Wells, "Simultaneous truth and performance level estimation (STAPE): an algorithm for the validation of image segmentation", IEEE Trans. Medical Imaging, 23, 7. pg. 903-921. (2004).

[9] S. K. Warfield, K. H. Zou, W. M. Wells, "Validation of image segmentation by estimating rater bias and variance", Philosophical Trans. A Mathematical, Physical & Engineering Sciences, 266, 1874. pg. 2361-2375. (2008).

[10] L. He, P. Michel, S. Puglia, D. Williams, autoSFR: new extensions to automated scanning resolution determination for photographic collections, Proc. IS&T Archiving, pg. 101-105. (2014).

[11] ISO 13655:2009, Graphic Technology – Spectral Measurements and Colorimetric Computation for Graphic Arts Images.

[12] ISO 3664:2009, Graphic Technology and Photography – Viewing Conditions.

[13] P. Nussbaum, A. Sole, and J. Y. Hardeberg. "Analysis of color measurement uncertainty in a color managed printing workflow", Colour Measurement and Print Quality Assessment in a Colour Managed Printing Workflow: 253.

[14] T. F. Chan, S. Osher, J. Shen, "The digital TV filter and nonlinear denoising", IEEE Trans. Image Processing, 10, 2. pg. 231-241. (2001).

Author Biography

Lei He received his PhD in Electrical Engineering from University of Cincinnati in 2003. Since then he has worked as a faculty member at Armstrong Atlantic State University for six years and was awarded the tenure in 2009 as an associate professor. Before joining the Library of Congress, he has visited the National Library of Medicine for two years. Currently he is digital imaging scientist at the LOC. His work focuses on imaging quality analysis for digital preservation at the LOC.