Color Calibration Techniques for Print Quality Measurements

Wencheng Wu and René Rasmussen
Xerox Corporation, Webster, NY

Abstract

Accurate color calibration is crucial for obtaining consistent and meaningful print quality measurements. In this paper, we will investigate the effect of color calibration errors to print quality measurements, such as mottle, banding, micro-uniformity etc. Our objective is to identify the source of errors that affects the precision in measuring these quality metrics the most and propose color calibration procedures that will reduce these particular types of errors accordingly. By doing so, we provide calibration procedures that focus more on improving the measurement accuracy and consistency of these print quality metrics rather than strictly on the precision of color calibrations that is often quantified by the ΔE in CIE LAB.

1. Introduction

Measurements of print quality are essential in many applications such as in the development of printing systems and algorithms, or in the diagnosis of print defects. Conventionally, these measurements are made by using expensive microdensitometers. However, with the advent of high-precision and low-cost scanners, using scanners for color image quality measurements has become more and more promising[1]. However, in order to obtain consistent and meaningful measurements, it is important to calibrate the scanners. In fact, Lim and Mani[2] have shown that by performing a standard procedure for scanner calibration, the consistency of measuring mottle was greatly improved.

Although consistency is one of the main concerns, the accuracy of the measurements is even more important. Since most scanners are not colorimetric and/or noise-free, in practice scanner calibration is fine tuned to the prints to be evaluated instead (as discussed below). To distinguish from the standard scanner calibration model, we will refer to the calibration model that is tuned to a specific type of prints as print-specific scanner calibration. The difference between the two lies in the selection of training samples that are used to identify the calibration model. For the standard model, the training set consists of standard color samples that represent a large gamut in color space.

For a print-specific scanner model, the training samples are prepared using the same type of printer and substrate as the prints to be evaluated. By restricting the training samples in scanner calibration, color calibration errors for prints of interest are reduced. As a result, the accuracy of print quality measurements is improved due to more precise color calibration. This comes at the cost of calibration errors that increase dramatically for prints lying outside the color space formed by the training samples. Hence the print-specific model is less robust than the standard model.

If the evaluation of print quality are made by testing a series of patches with nearly the same hue with different densities, a full-color (FC) calibration can be replaced by monochromatic (MC) calibration.

In this paper, we focus on the impact of color calibration errors on measurements of print quality. In particular, we investigate the variation of mottle values that is due to the change in scanner calibration procedures (standard vs. print-specific, full-color vs. monochromatic). Our objective is to evaluate the significance of this effect, compare the results from different calibration models, and suggest a calibration procedure that focuses on the accuracy and consistency of the print quality measurement.

2. Scanner calibration

In this section, we will briefly describe the scanner calibration model that we applied in this study and its impact on some example print quality measurements. For simplicity, we will only focus on the estimation of lightness and those metrics that measure the non-uniformity in L^* of a print sample, such as mottle, banding, and uniformity.

Calibration model for luminance/lightness

One common technique[3,4] for color calibration is applying a two-step procedure that consists of gray-balance and matrix-transformation. First, the device outputs are gray-balanced so that the nonlinearity in the outputs is removed. Then a calibration matrix is designed and used to transform the linearized device outputs to a desired color space, such as CIE XYZ[5]. The design criterion that we chose for calibration matrix is a total least-square optimization over the training samples[4]. Note that in estimating lightness the calibration matrix degrades to a vector or even a scalar.

Since the nonlinearity of the scanner is independent of the prints to be scanned, only one set of gray-balance curves is needed for a given scanner. In our model, it is obtained by scanning six neutral patches in Macbeth ColorChecker and measuring their luminance. Once the nonlinearity is removed, a linear model is adequate for estimating the luminance.

$$
\begin{aligned}
Y &= \boldsymbol{a}^t\boldsymbol{x} = \begin{cases} a_r r + a_g g + a_b b & \text{for FC} \\ a_s(s+o_s) & \text{for MC} \end{cases} \\
L^* &= f(Y) = f(\boldsymbol{a}^t\boldsymbol{x}). \qquad (1)
\end{aligned}
$$

where, $s = r$, g, or b depending on which hue separation is tested, o_s is a luminance offset for the given separation, and

$$
f(Y) = \begin{cases} 116Y^{\frac{1}{3}} - 16 & Y > 0.008856 \\ 903.3Y & Y \le 0.008856 \end{cases} \qquad (2)
$$

Error analysis

Suppose that the calibration vector $\boldsymbol{a}$ is changed to $\boldsymbol{a}' = \boldsymbol{a} + \boldsymbol{\Delta a}$ due to a different selection of training samples in the scanner calibration.

At each pixel of the scan, the value of the estimated L^* using Eq. (1) will vary from its nominal value due to this change. The amount of variation can be computed by

$$
\begin{aligned}
\Delta L^* &= f((\boldsymbol{a}+\boldsymbol{\Delta a})^t\boldsymbol{x}) - f(\boldsymbol{a}^t\boldsymbol{x}) \\
&\approx f'(\boldsymbol{a}^t\boldsymbol{x})\boldsymbol{\Delta a}^t\boldsymbol{x} = f'(Y)\boldsymbol{\Delta a}^t\boldsymbol{x}. \qquad (3)
\end{aligned}
$$

This equation accounts for the variation in L^* by applying different calibration vectors in the model. For example, using standard method rather than print-specific method or using vector obtained from printA to calibrate printB, etc.

Impact of lightness variation on measurements of print quality metric: mottle

The impact of lightness variation on the measurements of print quality depends on the definition of metrics. For illustration, we will discuss mottle mostly.

Since the value of mottle is used to describe the medium frequency lightness variation, we define the mottle M as the standard deviation of the lightness distribution over a filtered version of a given print area. The filtered version of the test patch is obtained by spatially averaging the luminance over an area of $1.2\,mm^2$, i.e. by filtering the scan image using a smoothing kernel with entries one and physical size $1.1mm \times 1.1mm$, and then decimating it by a factor of $1.1mm$ in both horizontal and vertical directions. The mottle is then computed as

$$
M = \sqrt{\frac{1}{N-1}\sum_{k\in S}(L_k^* - \overline{L^*})^2}. \qquad (4)
$$

where, S is the lowpass-filtered scan of the test sample.

Let $M_2 = M^2$ and M_2' be the square of the new mottle value due to lightness variation occurred in the calibration. Then,

$$
\begin{aligned}
M_2 &= \frac{1}{N-1}\sum_{k\in S}(L_k^* - \overline{L^*})^2; \\
M_2' &= \frac{1}{N-1}\sum_{k\in S}((L_k^* + \Delta L_k^*) - (\overline{L^*} + \overline{\Delta L^*}))^2 \qquad (5)
\end{aligned}
$$

Note that if $\Delta L_k^* = \overline{\Delta L^*}$ for all k, then $M_2' = M_2$. Therefore, the measurement of mottle is invariant to constant bias in the lightness calibration.

Index	Method	Model	Training samples
1	Standard FC	Full-color in Eq (1)	Macbeth ColorChecker
2	Print-specific FC	Full-color in Eq (1)	206 color squares in each print
3	Standard MC	Monochromatic in Eq (1)	CMYK colors in all 14 prints
4	Print-specific MC	Monochromatic in Eq (1)	CMYK colors in each print

Table 1: The characteristics of the calibration methods used in mottle measurements

Now let $\Delta M_2 = M_2' - M_2$ be the variation of the square of the mottle value due to different scanner calibration. From the appendix, it can be shown that

$$
|\Delta M_2| \le (f'(\overline{Y}))^2\|\boldsymbol{a}' + \boldsymbol{a}\|\|\boldsymbol{C_x}\|\|\Delta\boldsymbol{a}\| \qquad (6)
$$

Similar analysis can be obtained on banding and uniformity. The demerit of banding or uniformity is often defined as a *visually* weighted sum of the lightness deviation from the average lightness in a given printed area; and the derivation will be similar but easier (no square-root).

The key observation from Eq. (6) is: for print quality metrics that measure the non-uniformity of lightness, it is critical to minimize the size of $\Delta\boldsymbol{a}$ or to ensure that $\Delta\boldsymbol{a}$ is orthogonal to the correlation matrix $\boldsymbol{C_x}$. Bias in lightness estimation is irrelevent for this type of quality measure.

3. Experiment on mottle measurements

Three flatbed scanners (Epson Expression 836XL, Scitex EverSmart Supreme, and Umax PowerLook III), four calibration methods, and fourteen prints were used in this experiment. The characteristics of the calibration methods are listed in Table 1. We selected prints from various types of printers (Dye-sublimation, Electrophotography, Lithography, and Ink-jet) and papers (coated and plain) to avoid systematic errors caused by the choice of training sample sets in this experiment.

Each print consists of 206 color squares. Among them, 32 samples (8 samples each in C, M, Y, or K with different densities) were used for mottle measurements and MC calibration. A total of 448 test samples were thus used for this experiment. While measuring the mottle values, the scans of these samples were first cropped to a square image with physical size of $30\,mm^2$. The cropped image element was then calibrated to its luminance value pixel by pixel following the procedure described in Sec. 2. After luminance averaging (lowpass filtering) and conversion from luminance to lightness, the mottle value of this element was then computed using Eq. (4). Since we used three scanners and four possible ways to calibrate the scanned image, there are twelve possible mottle values for a given test sample. Ideally, all 12 numbers should be the same. The discussion of the experimental results will be based on this fact.

For clarity, these mottle values are denoted as

$$
M_i^{jk} \quad i = 1,\ldots,448 \quad j = 1,2,3 \quad k = 1,2,3,4. \qquad (7)
$$

Calibration	$\overline{\sigma}$	$\sigma_{\max}$
FC	3.0 %	42.0 %
MC	8.2 %	61.4 %

Table 2: The results of robustness test on full-color and monochromatic calibration methods. The percentage mottle variation σ_{ij} for FC is computed by $\sigma(M_i^{j1}, M_i^{j2})$ using Eq. (8). That is, for a fixed i and j the comparison is made between two computed mottle values, where the scanned image is calibrated to L^* using standard or print-specific full-color calibration. The $\overline{\sigma}$ for FC is the average value of σ_{ij} averaging over i (the print sample) and j (the scanner); and $\sigma_{\max}$ is the maximum over all i and j. Similarly, σ_{ij} for MC equals to $\sigma(M_i^{j3}, M_i^{j4})$ using Eq. (8). $\overline{\sigma}$ and $\sigma_{\max}$ are the average and maximum over all i, j.

where, index i, j, and k are associate with the print sample index, scanner index, and calibration method index, respectively. Since the significance of the variation in mottle values are relative rather than absolute, a quantity σ is defined as

$$\sigma(M', M) = 100\frac{|M' - M|}{M}\ \% \tag{8}$$

for the purpose of comparison between M' and M.

Robustness test

The purpose of this test is to evaluate the robustness of the full-color and monochromatic calibration methods. The concern is the impact on mottle measurement when the print used to identify the calibration vector differs from the print that is about to be evaluated. The results of this test are listed in Table 2.

Accuracy test

We evaluated the accuracy of mottle measurements from two aspects: the consistency across scanners and the accuracy in estimating lightness. The consistency itself has already played an important role in the measurement. When a calibration method is good in both aspects, it can be inferred that this method has accurate mottle measurements. The results of this test are listed in Table 3. Example scatter plots are shown in Fig. 1. Note that the consistency is perfect if all points fall on the $45°$ line.

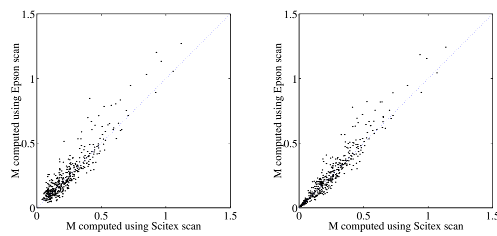
In Fig. 1b the scattering around the $45°$ line gets wider when M gets larger (resembling a cone centers at this line), while in Fig. 1a the scattering is about the same at most M levels. The trend in Fig. 1b is preferable since it implies that the relative variation is the same at all mottle levels. The trend in Fig. 1a on the other hand has significant percentage of variation in the range of low/moderate mottle values. This is undesirable since this is the region that demands the most precise measurement of mottle values.

Effect in image cropping

The location of the cropped region will affect the mottle measurements, if there exist streaks or banding in the test patches. Although mottle only measures low frequency variation, there are still some impacts due to the limitation of crop size. We tested this effect by randomly re-cropping the scanned images twice, computing the mottle values for new cropped images, and evaluating the percentage variation σ compared to the measurements of the original crop. The results are listed in Table 4. Figure 2 shows example scatter plots of cropping effect. Note that the impacts are about the same on all 4 calibration methods. All scattering plots are cone-shaped as well.

Calibration	$\overline{\sigma}$	$\sigma_{\max}$	$\overline{dL}$	$dL_{\max}$
Standard FC	17.9%	93.4 %	1.6	10.1
Print-specific FC	16.9%	86.7%	0.9	8.3
Standard MC	11.1%	50.7%	1.3	9.8
Print-specific MC	10.6%	52.2%	0.7	4.6

Table 3: The results of accuracy test: consistency across scanners and accuracy in L^* estimation. The notation is the same as in Table 2. The percentage mottle variation σ_{ij} in the kth row of the table is computed by $\sigma(M_i^{jk}, (M_i^{1k} + M_i^{2k} + M_i^{3k})/3)$ using Eq. (8). Therefore, the average value $\overline{\sigma}$ over all i, j quantify the consistency of mottle measurement across scanners for method k. The lightness estimation error dL_{ij} in the kth row of the table is the absolute difference between estimated L^* of the ith sample using scanner j, calibration k and the true L^* value measured by spectrophotometer. $\overline{dL}$ and $dL_{\max}$ are the average and maximum of dL_{ij} over all i, j, respectively. The value of $\overline{dL}$ indicates the accuracy in L^* estimation for a give calibration method k.



(a) (b)

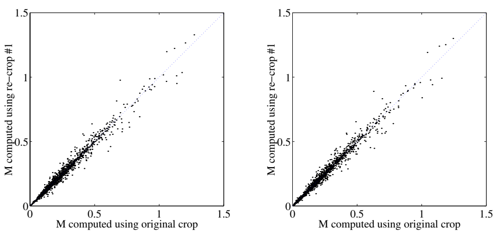
Figure 1: Example scatter plots for accuracy test. Here the computed mottle values using scanned images from Epson scanner are plotted against those using scanned images from Scitex scanner. The scanned images from both scanners were calibrated to L^* using (a) standard FC calibration and (b) standard MC calibration.

Discussion

From Table 2, it is observed that the full-color calibration is more robust than monochromatic calibration. The reason is that in mottle measurements the test samples contain only a fairly small set of hues (C, M, Y, or K). Even though these hues vary from one print to another, the variation is still limited. Since the training samples used in standard FC calibration represent a gamut that contains all sort of hues including those CMYK test samples, the standard model is suitable for all prints in calibrating the test samples. Thus for each print, the variation of mottle values using standard or print-specific model is insignificant (robust). For monochromatic calibration, it is assumed that each separation has one hue only. The calibration scale factor and the offset is highly dependent of the print. Due to

Calibration	$\overline{\sigma}_1$	$\sigma_{1,\max}$	$\overline{\sigma}_2$	$\sigma_{2,\max}$
Standard FC	7.0%	48.6%	7.0%	45.5%
Print-specific FC	7.0%	47.8%	7.1%	44.8%
Standard MC	7.8%	76.4%	8.1%	66.5%
Print-specific MC	7.8%	76.4%	8.1%	66.5%

Table 4: The effect of image cropping on the calibration methods. The σ_{ij} for crop #m in the kth row of the table is computed by $\sigma(M_i^{jk}$ from crop #m,M_i^{jk} from original crop) using Eq. (8). $\overline{\sigma}_m$ and $\sigma_{m,\max}$, $m = 1, 2$ are used to represent the comparison of the mottle values between crop #m and original crop.



(a) (b)

Figure 2: Example scatter plots for cropping effect #1. Here the computed mottle values using cropped images #1 are plotted against those using original cropped images. Mottle values M were computed after the cropped images were calibrated to L^* using (a) standard FC calibration for all scanners and (b) standard MC calibration for all scanners.

the hue variation in CMKY among prints, the standard MC model that uses all 448 CMYK samples describes only the MC calibrations for *average* CMYK separations. When the CMYK test samples from a given print differ from the *average* CMKY separations, the one hue per separation assumption is viloated; therefore the variation of mottle values using standard or print-specific model will be large for this print (less robust).

From Table 3, it can be seen that monochromatic calibration obtain better accuracy in mottle measurement than full-color calibration. Print-specific method performs better than standard method. In addition, we see that monochromatic methods are more suitable in mottle measurements especially for the print samples with moderate mottle values (see trends in Fig. 1). Even though the standard mono-chromatic calibration does not do well in estimating L^*, the consistency $\overline{\sigma}$ and its trend in percentage mottle variation out-perform both full-color methods.

The effect of image cropping has a noticeable impact on all four calibration methods. It is not surprising that the extend of impact is the same among them since it depends on the quality of the print samples rather than the calibration methods. This effect can be reduced by increasing the crop size in computing mottle values. Other factors, such as the MTF of a scanner, the noise in scanner RGB values, and the color to color registration errors, could affect the precision of print quality measurements as well. We will not discuss their impacts on mottle since the effects of the above three factors are limited due to the averaging in mottle computation. Furthermore, the color to color registration errors can be eliminated by applying MC calibration.

4. Conclusion

In this work, we investigated the effect of calibration errors on print quality measurement. In particular, those errors that occurred when different scanner calibrations are applied. From our analysis, we found that it is important to identify calibration vectors according to the print to be evaluated. Bias in lightness estimation does not affect the measurement of mottle. This observation can be used to improve or simplify the calibration method for the purpose of print quality measurement. For example, using total least-square optimization to identify the calibration vector can be modified with an offset term. Though the accuracy of L^* estimation might suffer, the accuracy of mottle values can be improved since the estimation of L^* only need to be correct within a constant offset for this measurement.

From our experiment, we found that monochromatic methods are more suitable in mottle measurement. In fact, the print-specific monochromatic calibration performs the best among these four methods. However, the robustness issue could play an important role in practice when the evaluation of print quality are done separately from scanner calibration.

References

1. D. R. Rasmussen, B. Mishra, and M. Mongeon, "Using drum and flatbed scanners for color image quality measurements," *Proceedings of IS&T's PICS2000*, Portland, OR, 26-29 March 2000, pp 108 – 113.
2. W. Lim and S. Mani, "Application of digital imaging to measure print quality," *Proceedings of IS&T's NIP14*, Toronto, Ontario, Canada, 18-23 October, 1998, pp 611 – 614.
3. H. R. Kang, "Color scanner calibration," *Journal of IS&T*, vol. 36, pp. 162 – 170, 1992.
4. W. Wu, J. P. Allebach, and M. Analoui, "Using a digital camera for colorimetry of human teeth," *Proceedings of IS&T's PICS1998*, Portland, OR, 17-20 May 1998, pp 37 – 42.
5. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, Inc, 2 ed., 1982.

Appendix

Ideally the lowpass-filtered scan area S where mottle measurement is applied should have a constant lightness. Hence a first-order approximation is adequate for prints that have *moderate* quality on mottle. That is, for all k in S, we have $f'(Y_k) \approx f'(\overline{Y})$, $L_k^* - \overline{L^*} \approx f'(\overline{Y})\boldsymbol{a}^t(\boldsymbol{x}_k - \overline{\boldsymbol{x}})$, and $\Delta L_k^* - \overline{\Delta L^*} \approx f'(\overline{Y})\Delta\boldsymbol{a}^t(\boldsymbol{x}_k - \overline{\boldsymbol{x}})$. Using these approximations and Eq. (5), it can be shown that

$$\begin{aligned}\Delta M_2 &\approx (f'(\overline{Y}))^2(2\boldsymbol{a} + \Delta\boldsymbol{a})^t\boldsymbol{C_x}\Delta\boldsymbol{a} \\ &= (f'(\overline{Y}))^2(\boldsymbol{a}' + \boldsymbol{a})^t\boldsymbol{C_x}(\boldsymbol{a}' - \boldsymbol{a}) \quad \text{(A-1)} \\ &\leq (f'(\overline{Y}))^2\|\boldsymbol{a}' + \boldsymbol{a}\|\|\boldsymbol{C_x}\|\|\Delta\boldsymbol{a}\|. \quad \text{(A-2)}\end{aligned}$$

where, $\boldsymbol{C_x} = \frac{1}{N-1}\sum_{k\in S}(\boldsymbol{x}_k - \overline{\boldsymbol{x}})(\boldsymbol{x}_k - \overline{\boldsymbol{x}})^t$.