Improving Tone Prediction Accuracy in Calibration for Color Electrophotography Part I – Environmental and Consumable

Chao-Lung Yang[a], Yan-Fu Kuo[b], Yuehwern Yih[a], George T.-C. Chiu[b], Dennis A. Abramsohn[d], Gary R. Ashton[e], and Jan P. Allebach[c]; a: School of Industrial Engineering, b: School of Mechanical Engineering, c: School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, d: Hewlett-Packard Company, Boise, Idaho 83714, e: Consultant, Eagle, Idaho 83616

Abstract

Environmental and consumable operating conditions have been observed to have significant impact on the color electrophotographic (EP) process. This paper presents the results of utilizing operating information to improve the sensor mapping which predicts tone reproduction on printing media based on measurements on substitute media in off-line calibration. In this research, time-series sensor data and color measurements have been collected from off-the-shelf color EP platform printers under a variety of operating conditions. The data analysis shows that the sensor mapping has distinctive behaviors under different levels of relative humidity and cartridge toner consumption. In addition, the sensor mapping has been found to be sensitive to tone level. A new prediction model is proposed to compensate for environmental and consumable disturbances and to capture tone-level-dependent characteristics. The experimental results show the proposed model is able to improve the prediction accuracy by 30% on average.

Introduction

A color electrophotography (EP) printing system is physically a binary process in nature. It produces all output colors by the combinations of certain primary colors such as cyan, magenta, yellow, and black (CMYK). To reproduce a primary color with a desired tone, an EP printing system first translates the desired continuous tone image into a half-toned image labeled with a half-toned density. In order to represent colorimetric characterization of the EP printing system outputs, the colorimetric tone reproduction (CTR) measurement of the printed half-toned image is needed. Conventionally, the CTR can be determined by calculating the distance between the colorimetric measurement-of a printed color image and -of a reference color in a device-independent color space. For example, CIE L*a*b* is one of color spaces used for computing the CTR value.

Performing a calibration process periodically is a prevailing approach to maintain consistency of primary color by restoring EP parameter to a desired state. Conceptually, the calibration process can be defined as a feedback control system which first prints particular patterns on the desired media based on the pre-determined characterization function, and measure the print-outs to obtain CTR values. Then, the measured CTR values are applied as inputs of the inverted characterization function to obtain the control values required for calibration. The calibration described above is also categorized as "on-media" calibration.

Because of the costs of consuming media and user involvement on measuring printouts, the practicality and automaticity of on-media calibration is limited [1]. Therefore, the alternative "off-media" calibration illustrated in Figure 1 is commonly utilized in off-the-shelf devices. Basically, in off-media calibration, the on-board densitometer is installed in the device to measure color patches printed, not on media usually used for printing, but on the transfer belt during calibration process. The accuracy of off-media calibration highly relies on the correlation to determine how measured density on transfer belt (S_l) relates to actual printing density on media (CTR'_l). This sensor mapping function should be developed at factory during product development. In this research, we focused on improving the accuracy of the sensor mapping which predicts tone reproduction on printing media based on measurements on substitute media in off-media calibration

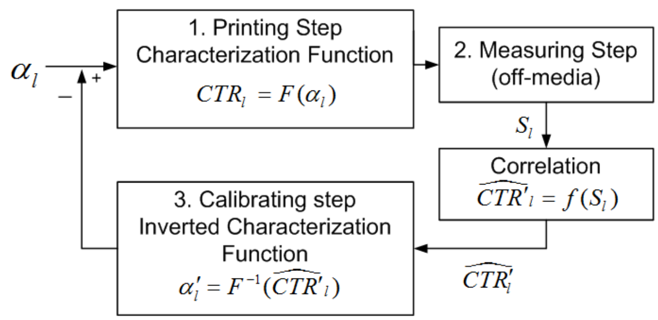

Figure 1. *Illustration of off-media calibration in an off-the-shelf device with on-board measurement.*

Data Collection

Since 2001, the data harvesting project was initiated at Purdue University to collect time-series sensor data and color measurements from several off-the-shelf color EP printers under a variety of operating conditions in the real-customer environment. The data harvesting procedure consists of three processes: 1) performing calibration, 2) retrieving data from printers, and 3) printing the test page after calibration. It took 30 minutes to finish a harvesting procedure. By the remote control, the procedure was triggered every six hours.

The test page was designed by PostScript commands. Thirteen color patches with various tone level ranging from 0 to 160, which are also used for printing on the transfer belt in off-media calibration, are determined to represents the colorimetric tone reproduction. All test pages which bypass the calibration effect were printed to investigate the sensor mapping issue. The spectrophotometer (Xrite DTP70®) is used to measure color patches and obtain the CTR values (ΔE76 from paper white).

The data retrieval process also collected sensor data and printer operating conditions from the I/O interface of printers. The collected dataset contains five groups of information: 1) printer setting information, 2) densitometer readings (S_l), 3) environmental conditions such as temperature and humidity, 4) consumable conditions such as cartridge toner consumption (CTC), and 5) other EP related parameters.

Data Analysis

Sensor Mapping Model

A simple regression model is conventionally used to construct the sensor mapping between S_l and CTR_l' in off-media calibration. Figure 2 displays an empirical example of the linear correlation between S_l and CTR_l' (ΔE76 values from paper white). The data corresponding to a certain tone level are plotted as a data cluster. Several data clusters in X-Y dimension show a high correlation between S_l and CTR_l'.

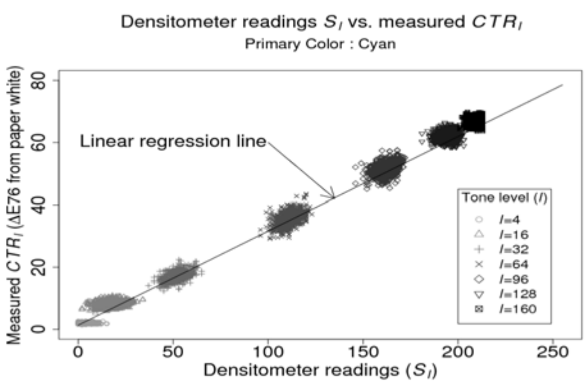

Figure 2. *An example of sensor mapping between densitometer sensor readings and measured CTR values (ΔE76 values from paper white).*

Although the linear relationship between S_l and CTR_l' is obvious, the prediction error is perceptible due to sensor mapping variation. Based on data analysis, the prediction error (root mean squared error, RMSE) of the existing method is greater than 2.5 ΔE76 on average in CMYK. The error distinguishable by human eye, in fact, degrades the calibration accuracy and need to be controlled properly.

Based on previous research works [2][3], the environmental and consumable conditions such as temperature, humidity, and cartridge toner consumption have been proved to impact the EP process. In the following sections, the impacts of the relative humidity and cartridge toner consumption on sensor mapping are reviewed. Variety of sensor mapping behaviors under different tone level is also investigated.

Environmental Factors – Relative Humidity

Figure 3 displays an example of the measured CTR values (ΔE76 from paper white) on media under different relative humidity (RH) levels (1871 data points in total). Each bar in the figure shows average CTR values of the collected data in certain RH level. The line segments on the top of each bar represent the confidence intervals surrounding the mean based on 95% confidence level. Obviously, we can observe the average CTR value is lower while RH is higher in this example. ANOVA analysis was performed and confirmed that RH is a significant factor to explain the measured CTR variation. This observation is consistent with results of the previous research works in [2].

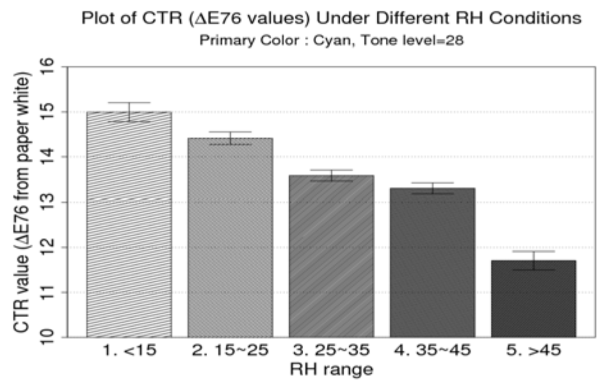

Figure 3. *Average CTR values (ΔE76 from paper white) under different RH conditions. The bars on the plot indicate mean values under different RH range. The mean CTR value under a certain RH range is significantly different from one under another RH range.*

Consumable Factors – Cartridge Toner Consumption (CTC)

Cartridge toner consumption (CTC) is a measurement to represent the usage of the cartridge toner. It is defined as a percentage value. For example, 0% CTC specifies that the cartridge is brand new; 100% CTC represents that the toner in the cartridge is running out.

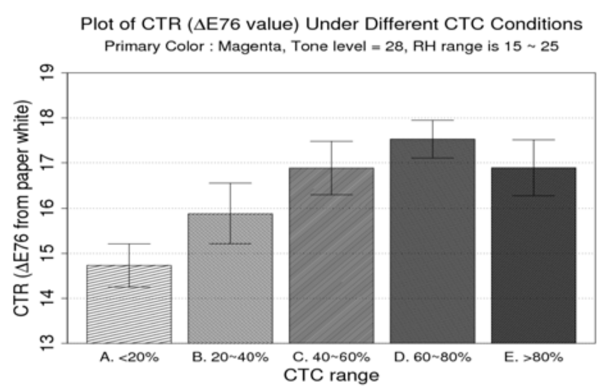

Figure 4. *Average CTR values under different CTC conditions (magenta, tone level = 28). The bars on the plot indicate mean CTR values under different CTC range. It shows the average CTR value in CTC < 20% group is relatively lower than it of other CTC groups.*

Following the same analysis method in the previous section, the collected data is categorized to different CTC groups, from low to high (or from new to old cartridge). In order to emphasize the impact of CTC, here, only data under a certain RH level (15 ~ 25) is showed in Figure 4. We can see that the mean of CTR values in low CTC group is significantly lower than other CTC groups. This interesting pattern of CTR values can also be observed under different levels of CTC, and also different RH range. ANOVA analysis is conducted to show not only CTC but also the interaction between RH and CTC are significant factors on the measured CTR values, based on 95% confidence level.

Tone-level-dependent Characteristics

In order to investigate tone-level-dependent characteristics, we plot the data of the selected four different tone levels (4, 16, 28, and 64) to check the sensor mapping correlation (see Figure 5). Similar to Figure 2, the solid straight line indicates the regression line of the existing sensor mapping method. The dotted lines present the regression line of each data cluster. Obviously, the linear relationship between S_l and CTR'_l, in terms of regression slope, are not consistent among four tone levels.

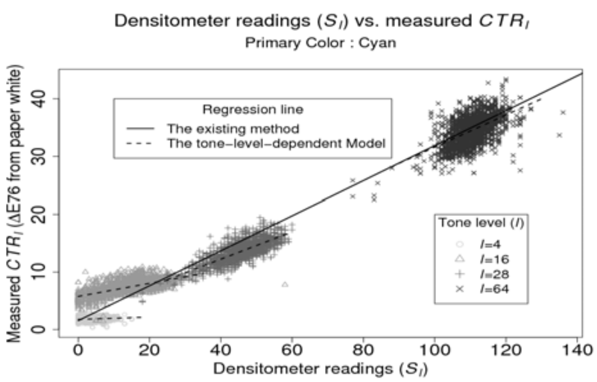

Figure 5. *An example of sensor mapping between densitometer sensor readings and measured CTR values (ΔE76 values from paper white). The data with 4 tone levels are plotted. The dotted lines representing the regression lines of each data cluster are compared to the regression line by using all data.*

Figure 6 further compares regression coefficients of the selected 12 data clusters (12 different tone levels) with the regression coefficient applied by the existing method. Again, the straight line indicates the regression coefficient of the existing method. We can observe that the coefficient is higher (steeper) in middle tone range and smaller (flatter) in low and high tones. All of them are different from the constant regression coefficient of the existing method. Therefore, the existence of different tone-level-dependent characteristic in the sensor mapping can be concluded.

Methods and Results

Methods

Multiple linear regression method was applied to construct a new sensor mapping model which not only includes S_l but also considers RH, CTC and their interaction factors. Instead of developing a single model for all tone levels (the existing method), an individual model is developed for each pre-defined tone level to catch the unique tone-level-dependent characteristic. In order to evaluate the effectiveness of applying individual sensor mapping in each tone level, the tone-level-dependent model as Equation (1) is first compared with the existing method. Notice that this tone-level-dependent model does not include RH, CTC, and their interaction factor in the model.

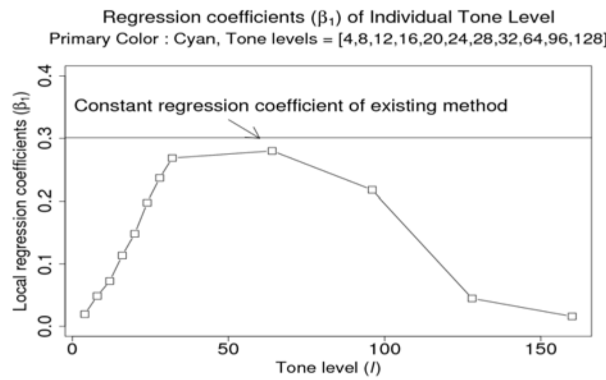

Figure 6. *12 regression coefficients of the selected 12 data clusters (12 different tone levels) are compared with the regression coefficient applied by the existing method.*

The second model includes RH, CTC, and their interaction factor as Equation (2). By comparing the first and second models, the effectiveness of including environmental and consumable conditions can be evaluated. In this research, only first-order regression is considered due to the memory constraint of the printer firmware.

$$CTR_{l,i} = \beta_o + \beta_1 \times S_{l,i} + \varepsilon_i \quad (1)$$

$$CTR_{l,i} = \beta_o + \beta_1 \times S_{l,i} + \beta_2 \times RH_i + \beta_3 \times CTC_i + \beta_4 (RH_i \times CTC_i) + \varepsilon_i \quad (2)$$

β_k is the regression coefficient for the relation between CTR_l and the associated dependent variables

ε_i is independent $N(0, \sigma^2)$

$i = 1,...,n;$ n is total number of dataset

By utilizing the collected data, the existing method and proposed models were compared with each other by the model performance measure. The root mean square error on cross validation (CVRMSE) defined as Equation (3) is a conventional performance measure to compare the prediction models. Essentially, the whole dataset is divided by k folds. In each of k iterations, one fold of data is used as validation set and rest of data is the training set. The RMSE of all predictions in each testing set is computed [4]. In this research, the 10-fold cross validation which is commonly used in literatures was considered.

$$\text{CVRMSE} = \sqrt{\frac{1}{m \times n} \sum_{l=1}^{m} \sum_{i=1}^{n} (CTR_{l,i} - \hat{f}^{-k(i)}(x_i))^2} \quad (3)$$

m is number of tone levels

x_i is a vector of predictive varaibles (S_l, RH, CTC)

$\hat{f}^{-k(i)}(x_i)$ is the predicted value by k-fold cross validation

$i = 1,...,n,\ n$ is number of sample size

Results

Totally, 1871 data points are used to evaluate the existing method and proposed models. Table 1 shows the average CVRMSE across all tone levels of the existing method and proposed models. The parentheses denote the percentage improvement of the proposed models compared with the existing method. Obviously, the existing method has largest CVRMSE representing the worst case. The experimental result of the 1st tone-level-dependent model shows significant improvement on reducing the CVRMSE. The paired T-test comparing the existing method and the tone-level-dependent model on each tone level also confirm this improvement (95% confidence level). It implies that the individual model of each tone level is able to catch the tone-level-dependent characteristic and improve the sensor mapping.

Table 1: Root mean square error on cross validation (CVRMSE) comparison of sensor mapping models

Model	CVRMSE (ΔE76)			
	Cyan	Magenta	Yellow	Black
Existing Method	1.968	2.184	2.675	3.476
1st Tone-level-dependent Model	1.327 (32.6%)	1.826 (16.4%)	2.194 (18.0%)	2.155 (38.0%)
2nd Tone-level-dependent Model with RH and CTC Factors	1.180 (40.1%)	1.724 (21.1%)	2.105 (21.3%)	2.039 (41.3%)

The 2nd tone-level-dependent model including the environmental and consumable conditions (RH, CTC, and their interaction) is further compared to the existing method. The result shows that adding RH, CTC, and their interaction factors can help on further explaining the variations and improve the prediction, except yellow color. Again, the paired T-test was conducted to confirm this result based on 95% confidence level. On average, around 30% CVRMSE across CMYK can be reduced by the proposed new model comparing with the existing method using in off-media calibration.

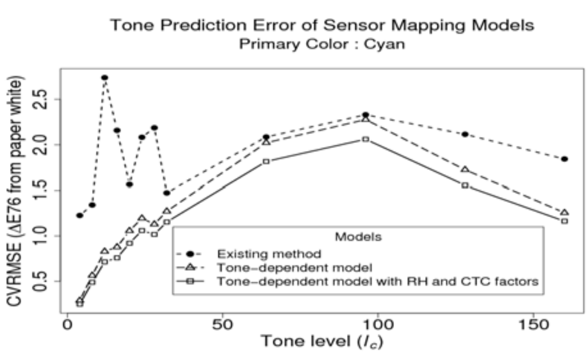

Figure 7. The root mean square error on cross validation (CVRMSE) of sensor mapping models are compared in each tone level.

Figure 7 illustrates the prediction improvement for cyan as an example. The solid circles denote the CVRMSE of the existing method across 12 tone levels. The triangles indicate the performance of the model which considers tone-level-dependent characteristics. The 1st tone-dependent model is able to reduce the sensor mapping error especially in low and high tone levels. The square symbols represent the results of the 2nd model which considers not only tone-level-dependent characteristics, but also RH and CTC factors. The prediction accuracy of the tone-level-dependent model with RH and CTC factors is improved by 40% compared with the existing model in this case.

Conclusion

A more accurate sensor mapping model was proposed to improve the prediction accuracy which is important for off-media calibration on the color EP printer. Based on the data analysis, the relative humidity (RH) and cartridge tone consumption (CTC) conditions have significant impacts on the sensor mapping between color tone reproduction (CTR) and on-board densitometer readings (S_l). Besides, the tone-level-dependent characteristic was observed to violate the assumption of the existing method assuming the constant mapping across tone levels. We proposed a new model to compensate for environmental and consumable disturbances and to capture tone-level-dependent characteristics. To verify the validity of the proposed approach, the prediction models are compared by using collected data. CVRMSE is used to investigate the model accuracy. The results show that the proposed model outperforms the existing method in terms of lower prediction error, and is able to reduce 30% of sensor mapping error on average.

Acknowledgements

The authors gratefully acknowledge the support from the Hewlett-Packard Company. They would like to specially thank Terry Nelson for his valuable comments and suggestions in this research.

References

[1] A. D. Johnson, Calibration of printing devices, U. S. Patent 6,982,812 (2006).

[2] K. Totsuka, G. T.-C. Chiu, D. Abramsohn, Tone Curve Stabilization for Color Electrophotography, Proc. IS&T's NIP21, 173-177, (2005).

[3] C. Staelin, R. Bergman, M. Fischer, M. Vans, D. Greig, G. Braverman, S. Harush, E. Shelef, "Dot Gain Table and Developer Voltage Prediction for the HP Indigo Press," J. Imaging Sci. Tech., 49, 620-628, (2005).

[4] T. Hastie , R. Tibshirani, J. Freidman, The Elements of Statistical Learning – Data Mining, Inference, and Prediction (Springer-Verlag, NY, 2001) pg. 253-268.

Author Biography

Chao-Lung Yang is currently a PhD candidate at School of Industrial Engineering, Purdue University. He received his BS in mechanical engineering and MS in automation control from National Taiwan University of Science and Technology, in 1996 and 1998, respectively. He also received MSIE in industrial engineering from Purdue University (2004). He was an intern engineer at Hewlett Packard in 2007 and 2008 to develop the data-mining tool for improving the product design