Nonparametric Generic Substrate ICC Profile

Chunghui Kuo, Hwai-Tzuu Tai, Yee Ng, and Eric Zeise; Eastman Kodak Company, Graphic Communication Group; 2600 Manitou Road, Rochester, NY 14653, USA

Abstract

Digital printing has begun to play a more significant role in the current commercial printing market, and the demand for accurate color reproduction has also increased. Nonetheless, achieving wide selection of printable substrates imposes a limiting factor on color accuracy for a printing press manufacturer. The compromised solution is to only provide and maintain certain generic substrate ICC profiles based on the physical features of substrates. The main concern for this approach is that it is very difficult to establish a correlation between each substrate characteristics and the chosen imaging process. We propose a nonparametric approach by first sifting through a representative set of substrate-specific ICC profiles. Design a feature dimension reduction algorithm based on the independent component analysis, and we will be able to identify a small set of substrates achieving the optimal generic ICC profile performance against investigated substrates. At last, we will identify the important physical characteristics of a substrate affecting the performance of the ICC profile with respect to the selected imaging process.

Introduction

Digital printing has begun to play a significant role in the current commercial printing market. Aided by the advances in machine control and color management technology, the demand for accurate color reproduction has also increased. The most obvious approach to address this concern is to create an ICC output profile of the intended substrate on the specific digital press. Nonetheless, because of the wide variety of substrates, it is highly desirable for a printing press to be able to image on as broad substrate selections as possible. Hence, providing a substrate-specific ICC profile for each intended substrate would impose a significant burden on maintaining a digital press. Moreover, because the performance of each substrate-specific ICC profile is intimately tied with colorant selection, individual tone reproduction curve, color-mixing characteristics, etc., any modification in the imaging process might render the entire collection of substrate-specific ICC profiles less accurate; on the other hand, continuing updating each substrate-specific ICC profile is a very costly operation for a press manufacturer. The compromised solution is to only provide and maintain certain generic substrate ICC profiles based on the physical features of substrates, such as surface coating, weight and thickness, substrate brightness, etc., denoted as the parametric generic substrate ICC profile approach. The main concern for the parametric approach is that it is very difficult to establish correlation between each substrate parametric feature and the chosen imaging process due to complicated substrate/colorant interaction and the vast varieties of available substrates.

In this paper, we propose a nonparametric approach to bypass the initial physical characteristics analysis by first sifting through a representative set of N substrate-specific ICC profiles. Each substrate-specific ICC profile can be characterized by a substrate matrix, $M_i^{icc} \in \Re^{836\times3}, \ \ i = 1 \dots N$, with size, where 836 represents the number of unique color patches inside the IT8.7/3 standardized press characterization target, and the dimension of the *CIELAB* color space is 3. This means that each substrate-specific ICC profile resides in a hyperspace with dimensionality being 2508. The objective is to identify clusters within the collected data set, and each clustered substrate-specific ICC profiles can be optimally represented by their centroid, $C_j^{icc}, \ \ j = 1 \dots N_c$, where N_c is the number of clusters. No prior knowledge is assumed with respect of the physical properties of each substrate. In general, two algorithms are applicable to address this problem: multidimensional scaling with predefined dissimilarity metric for two substrate matrices and unsupervised learning. Since defining a suitable dissimilarity metric is equally difficult as the original problem, we choose to adopt the unsupervised learning approach. We first design a feature dimension reduction algorithm based on the independent component analysis [1, 2], and perform data clustering in the feature space with much reduced dimensionality. Furthermore, we will be able to identify a small set of substrates from each cluster substrate set, denoted as *generic substrates*, that achieves the optimal generic ICC profile performance against investigated substrates under the constraint of maximal number of generic ICC profiles. At last, we will identify the importance physical characteristics of a substrate affecting the performance of the ICC profile with respect to the selected imaging process.

Nonparametric Generic Profile

Physical characteristics of substrates, colorant, and the current state of the printing process all affect the color rendition achievable on the selected substrate. Thus, achieving accurate color reproduction not only requires a substrate-specific ICC profile but also precise control of the printing press, colorant, and substrate manufacturing process. Since it is unrealistic to continuously maintain a valid set of substrate-specific ICC profiles in a digital press, the digital printing industry is moving toward a pragmatic solution by providing a small set of generic ICC profiles based on preselected physical characteristics of substrate, such as surface coating, weight, surface color and roughness, etc., and in-line color re-calibration to achieve accurate color reproduction [3]. Nonetheless, a set of generic ICC profiles optimized in terms of color reproduction accuracy among the qualified substrates for a digital press will reduce or even eliminate the need for color modification during the printing process, which, in turns, increases the productivity.

Complete understanding of the physics behind the interaction between the selected substrate and colorant in a specific printing process will undoubtably solve the generic profile optimization problem; however, the high complexity of physical interaction between colorant and substrate poses a significant challenge to researchers [4]. We propose to tackle this problem with a nonparametric approach. Let $\bar{M}_o^{icc}$ be the average press characteristics across all qualified substrates of the selected printing process, and there exists N_f physical properties, $\{\theta_j | j = 1 \dots N_f\}$, that drive the final press characteristics M_i^{icc} away from $\bar{M}_o^{icc}$. Furthermore, two assumptions are proposed to simplify the original problem:

A1: The amount of deviation from $\bar{M}_o^{icc}$ is the superposition of the effect from each individual physical feature θ_j.
A2: The contribution from each physical feature θ_j is linear.
A3: The contributions from $\{\theta_j | j = 1 \dots N_f\}$ are mutually independent.

As a result, we can formulate **A1** and **A2** into following equation:

$$\begin{aligned} X_i^{icc} = \quad & M_i^{icc} - \bar{M}_o^{icc} \quad && = f_i(\theta_j | j = 1 \dots N_f) \\ \Longrightarrow \quad & \Sigma_{j=1}^{N_j} f_{ij}(\theta_j) \quad && \Longrightarrow \sum_{j=1}^{N_j} h_{ij} S_j^{icc} \end{aligned} \tag{1}$$

where f_i represents the overall contribution from all physical properties, $\{\theta_j | j = 1 \dots N_f\}$, on substrate i, $f_{ij}(\theta_j)$ contains only the contribution from property θ_j, and a constant coefficient h_{ij} and a basis matrix $S_j^{icc} \in \Re^{836 \times 3}$ are used to approximate $f_{ij}(\theta_j)$.

Reorder X_i^{icc} and S_j^{icc} into vector forms to be x_i and s_j with the range space being $\Re^{2508 \times 1}$, and Equation (1) can be rewritten as follows:

$$x_i = \begin{bmatrix} s_1 & s_2 & \dots & s_{N_f} \end{bmatrix} \begin{bmatrix} h_{i1} & h_{i2} & \dots & h_{iN_f} \end{bmatrix}^t. \tag{2}$$

Denote

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix} \tag{3}$$

$$S = \begin{bmatrix} s_1 & s_2 & \dots & s_{N_f} \end{bmatrix} \tag{4}$$

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1N_f} \\ h_{21} & h_{22} & \dots & h_{2N_f} \\ \vdots & \vdots & \vdots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{NN_f} \end{bmatrix}^t \tag{5}$$

where N is the total number of substrates being analyzed. The original nonparametric generic substrate ICC profile problem can be recapitulated as the following matrix form:

$$X = SH + \Lambda \tag{6}$$

where $X \in \Re^{2508 \times N}$, $S \in \Re^{2508 \times N_f}$, $H \in \Re^{N \times N_f}$, and Λ is the additive white noise. S, H, and Λ are unknown.

Independent Component Analysis

We can treat X as the observed mixed signal matrix, S as the source signal matrix (latent variables), and H as the mixing matrix. Because Equation (6) is ill-posed with more unknown variables than number of equations, it is essential to first impose other constraints. For example, the singular value decomposition, *SVD*, has been widely adopted by researchers to solve the matrix decomposition, dimensionality reduction, and latent variable analysis by imposing orthogonality constraint [5, 6]. $X = U\Sigma V^t$, where U and V are orthogonal matrices and Σ is a diagonal matrix. Thus, we can define $S = U$ and $H = \Sigma V^t$. The imposed orthogonality constraint also means that there exists no correlation between two vector components, s_α and s_β, in S; however, this might be too restrictive to only allow θ_α and θ_β to drive x_i in perpendicular directions. Based on **A3**, we propose to adopt the independent component decomposition to estimate a separation matrix W such that

$$Y = XW = S(HW) \rightarrow S. \tag{7}$$

That is, the separation matrix W is the pseudo-inverse of the mixing matrix H.

Let each source signal s_α be a random process with its probability distribution being $p(s_\alpha)$. The independent component analysis assumes that source signals are mutually independent. That is,

$$p(S) = \prod_{j=1}^{N_f} p(s_j). \tag{8}$$

Various loss functions have been proposed to measure the degree of independence such as Kullback-Leibler divergence and high-order statistics (*Kurtosis*) [1, 7]. In a linear system as Equation (6) where x_i is a linear combination of independent variables $\{s_j,\ j = 1 \cdots N_f\}$, the probability distribution of x_i, $p(x_i)$, will be closer to a Gaussian distribution than any of $p(s_j)$ based on the *Central Limit Theorem*. As a result, the measures of nongaussianity, such as *Kurtosis*, can serve as an indirect metric for the level of mutual independence [7]. In this paper, we adopt the *Fast ICA* fixed-point algorithm with:

$$\bar{M}_o^{icc} = \frac{1}{N} \sum_{i=1}^{N} M_i^{icc} \tag{9}$$

which includes various physical features, such as surface coating, substrate weight/thickness, whiteness, surface roughness, material, etc. [2].

Generic Substrates

The first step before decomposing X is to estimate its rank, N_r, via *SVD*, where $N_r \ll N$. Let $\hat{X}$ be the projection of X onto the subspace spanned by the first N_r singular vectors, and the *fast ICA* algorithm results in the following decomposition:

$$\hat{X} = \hat{S}^\dagger \hat{H}^\dagger = \hat{S}^\dagger \begin{bmatrix} \bar{h}_1 & \bar{h}_2 & \cdots & \bar{h}_N \end{bmatrix} \tag{10}$$

where $\hat{S}^\dagger \in \Re^{2508 \times N_r}$, $\hat{H}^\dagger \in \Re^{N_r \times N}$, and $\bar{h}_i \in \Re^{N_r \times 1}$. Hence, in the feature space spanned by $\hat{S}^\dagger$, substrate i can be numerically represented by $\bar{h}_i$, of which dimension N_r is significantly reduced from that of the original substrate-specific ICC profile space, i.e., 2508. The difference between substrate α and β, $\delta_{\alpha\beta}$, is defined as follows:

$$\delta_{\alpha\beta} = \|\bar{h}_\alpha - \bar{h}_\beta\|_2 \tag{11}$$

where $\|\cdot\|_2$ is *2-norm*. As a result, we can begin to group neighboring substrates into clusters in the feature space sequentially [5]. Each cluster is denoted as one type of *generic substrate*. Let G_A and G_B be two generic substrates with n_a and n_b number of substrates at clustering stage k, and the difference between G_A and G_B, Δ_{AB}, is defined as follows:

$$\Delta_{AB} = \frac{1}{n_a n_b} \sum_{i=1}^{n_a} \sum_{j=1}^{n_b} \delta_{ab}^{ij}. \tag{12}$$

Two generic substrates with the smallest difference are merged to form a new generic substrate at stage $k+1$. The ICC profile for generic substrate A at each stage can be computed as follows:

$$\bar{M}_A^{icc} = \frac{1}{n_a} \sum_{A_i=1}^{n_a} M_{A_i}^{icc} \quad (13)$$

where A_i contains the substrate indices assigned to generic substrate A. Thus, all substrates are optimally represented by a group of generic substrate ICC profiles at each stage. A clustering tree is formed after completing the merging process. The final clustering result will be obtained by pruning the clustering tree. Let $\Delta_s = \{\Delta_1 \ \ \Delta_2 \ \ \cdots \ \ \Delta_{N-1}\}$ be the merged distance during the clustering process, and we can treat Δ_s as a stochastic process with two hypotheses:

H_0**:** Δ_i is measured from the same type of substrate.
H_1**:** Δ_i is measured from different types of substrates.

We propose to adopt the *Gamma distribution*, $P_g(x|\mu,\nu)$, to model the null hypothesis, H_0, and formulate the tree-pruning problem as a one-sided interval estimation problem [8]. The threshold η is set so that $\int_0^\eta P_g(x|\mu,\nu)dx = 0.95$.

At last, a short list of substrates are identified in each group of generic substrate, denoted as *Generic Substrate Prototypes*, of which color characteristics are closest to the corresponding $\bar{M}_A^{icc}$.

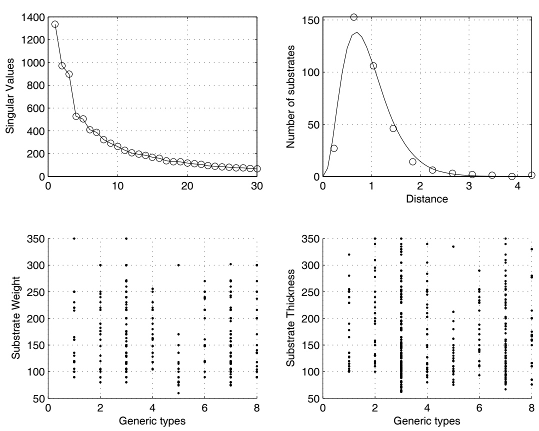

Figure 1. *Merging distance histogram and feature analysis.*

Experimental Results

The 469 substrates printed via the electrophotographic process are analyzed in our experiment, which include glossy-coated, matte-coated, cast-coated, uncoated, and texture substrate surface types. The weights of substrates range from 60 to 350 *gsm*. The measured media white points stored in each substrate-specific ICC profile indicate that the $L^* \subseteq [88 \ \ 98.1]$, $a^* \subseteq [-1 \ \ 4]$, $b^* \subseteq [-10 \ \ 10]$. Based on the extracted singular values as shown in Figure 1, it can be seen that the singular value decreases become slower beyond the eighth singular value. Hence, we set $N_r = 8$, and the first six independent components are shown in Figure 2, where coherent structures are observable. The clustering tree based on estimated $\hat{H}^\dagger$ is shown in Figure 3. Figure 1 also indicates that the *Gamma distribution* satisfactorily approximates the set of merging distance with estimated parameters $\mu = 3.66$ and $\nu = 0.26$. Hence, the result $\eta = 1.87$.

Table 1: Substrate Surface Classification

Class	1	2	3	4	5	6	7	8
Uncoated	24	26	6	0	28	11	33	7
Matte	1	1	53	2	2	9	69	26
Glossy	1	0	111	23	0	3	13	0
Texture	0	0	3	6	0	0	1	0
CastCoat	0	3	3	0	1	0	3	0

We can compare the nonparametric clustering result against various physical features such as substrate surface, weight, thickness, media whiteness, and resulted color gamut volume. Table 1 lists the substrate surface characteristics in each nonparametric generic substrate class. Although there exists crosstalk between surface types and generic classes, it appears that surface type/coating plays a significant role in the resulted color characteristics in the selected printing process. We can roughly classify class 1, 2, 5 and 6 to belong to the *"uncoated"*, class 7 and 8 as the *"matte coated"*, and class 3 and 4 as the *"glossy coated"*. How-

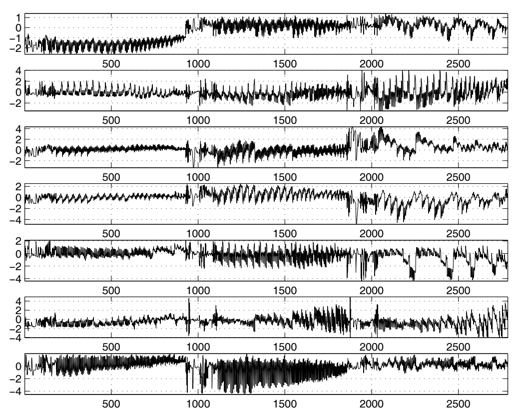

Figure 2. *The first six ICA bases.*

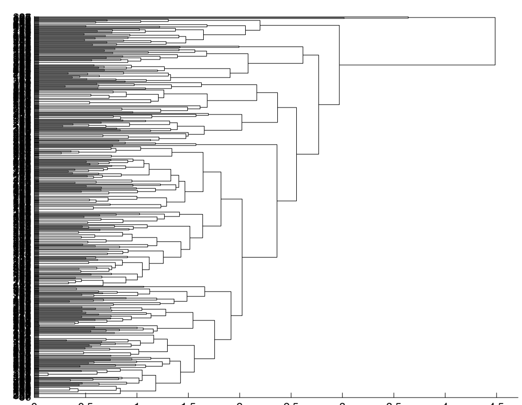

Figure 3. *Substrate clustering tree.*

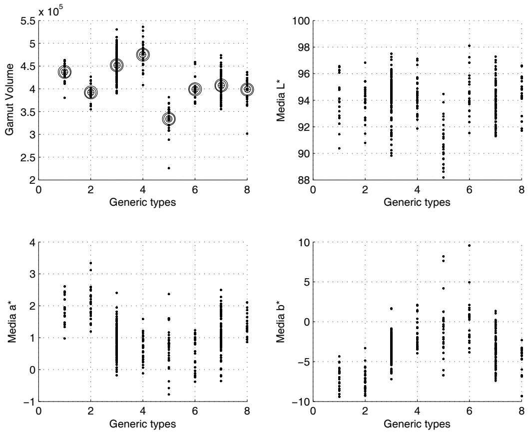

Figure 4. *Color gamut and media white point correlation.*

ever, Figure 1 suggests that the substrate weight and thickness have little impact on the color characteristics in the electrophotographic print process. Our explanation is that the impact from different substrate weight and thickness is mostly compensated by the fusing process with different fuser settings. Figure 4 shows the correlation between the nonparametric generic substrate class and the color gamut volume and media whiteness. On average, class 3 and 4 with more glossy coated substrates result in larger gamut volume than the rest, and class 2 and 5 generic substrate containing most uncoated substrates has the smallest color gamut volume. It is worthwhile noting that the class 1 generic substrate containing mainly uncoated substrates results in larger color gamut volume than that of class 7 and 8 with mostly matte-coated substrates. At last, while the substrate luminance L^* shows little correlation with the substrate class, the media color measured in $[a^* \;\; b^*]$ serves a very useful separating feature to classify generic substrates in the selected electrophotographic printing process.

Conclusion and Future Works

A nonparametric approach is proposed to address the generic substrate classification problem, where *ICA* is first used to reduce the dimensionality of the feature space, and 469 substrates printed with an electrophotographic process are grouped into different classes of generic substrates in the derived feature space. Our analysis results in 8 classes of generic substrates, where discriminating physical features include substrate surface, color gamut volume, and media color. On the other hand, substrate weight and thickness shows no correlation with generic substrate classes. In the future, we plan to analyze the extracted independent components and correlate with specific substrate physical characteristics, and extend this algorithm to other printing processes such as inkjet and offset printing processes.

References

[1] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing, John Wiley, 2002.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, John Wiley, 2001.

[3] ISO12647-2:2004, Graphic technology-Process control for the production of half-tone colour separations, proof and production prints-Parts 2:Offset lithographic processess, TC130, International Organization for Standardization, 2004.

[4] J. Bloch, S. Roscoat, P. Verhnes, C. Mercier, B. Pineaux, and A. Blayo, Influence of Paper Structure on Printability: Characterisation Using X-Ray Synchroton Microtomography, IS&T NIP22, 2006.

[5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Data Mining, Inference, and Prediction, Springer, 2001.

[6] G. Golub and C. Van Loan, Matrix Computations, Third Edition, The Johns Hopkins University Press, 1996.

[7] P. Comon, Independent Component Analysis, a new concept? Signal Processing, Elsevier, 36(3):287-314, April 1994, Special issue on High-Order Statistics.

[8] G. Casella and R. Berger, Statistical Inference, Second Edition, Duxbury, 2002.

Author Biography

Chunghui Kuo is a Scientist at Eastman Kodak Company. He received his Ph.D. in Electrical and Computer Engineering from the University of Minnesota and joined NexPress division of Eastman Kodak Company in 2001. His research interest is in image processing, image quality, blind signal separation and classification, and neural network applied in signal processing. He is a senior member of IEEE signal processing society and a member of IS&T and SPIE.

Hwai-Tzuu Tai is a technical member of Digital Printing Solutions of Graphic Communications Group (GCG), Eastman Kodak Company. Before that, he worked in R&D of NexPress Solutions LLC&Office Imaging of GCG. His research interests are electrophotographic processes, color profiling, workflow, and image processing related to digital halftoning and color printing. He holds 30+ patents in the image processing field.

Yee S. Ng is a Manager Intellectual Property, GCG, at Eastman Kodak Company. Prior to that, he was Senior Research Associate and Chief Engineer with responsibility for the engine image chain and image quality at NexPress Solutions LLC. He is a Kodak Distinguished Inventor and holds 90 U.S. patents. He is a Fellow of IS&T, a Senior Member of IEEE, and a member of New York Academy of Sciences. He is Project Editor for ISO/IEC 19799 (Gloss Uniformity), and ISO/IEC 24734 (Printer Productivity) and Liaison officer from ISO/IEC JTC1 SC28 (Office Equipment) to ISO/IEC TC130 (Graphics). He was General Chairman of NIP19 and received the Chester Carlson Award from IS&T in 2000. He also received the INCITS Team Award for the his contribution in the INCITS/W1 Productivity technical committee in the International Committee for Information Technology Standards (INCITS) in 2008.

Eric K. Zeise is a Research Associate and Group Leader for Image Quality Evaluation in the GCG of Eastman Kodak Company in Rochester, NY. He received his BS in physics from the University of Massachusetts, and MS and PhD degrees in low-temperature physics from Cornell University. He holds 19 patents in the areas of systems modeling, printing systems architecture, and perceptual image quality evaluation. Eric is a member and contributor to IS&T, IEEE, and APS, and has been invited to give several tutorials in the areas of image quality, color printing, and color management. He is active in standards development as Project Editor for the ISO/IEC JTC-1 SC28 and INCITS W1.1 team developing appearance-based image quality standards for printers and is Convenor of the ISO/IEC JTC-1 SC28 Working Group 4 (Image Quality).