Spectral data-driven model for overprint estimation

Pau Soler; Hewlett-Packard Large Format Printing; Barcelona, Spain
Ján Morovič; Hewlett-Packard Large Format Printing; Barcelona, Spain
Howard Doumaux; Hewlett-Packard Printing Technology Platforms; San Diego, CA/USA

Abstract

In this paper, we present a novel method to estimate color overprints from primary color spectra only. The method consists in building a mapping between primary color spectra and their combinations, based on a training set. The mapping is data-driven, without relying on assumptions based on classic theoretical models. By using this method, the number of prints and measurements for printer characterization is greatly reduced, which is of particular importance if there are many ink candidates to be combined. We tested the method on recently released Latex-based ink-jet ink, for different pigment and latex loads. The mean simulation accuracy error was below 2 dE, and gamut estimation error below 2% error. For pigment kinds, mean accuracy was about 5 dE, and gamut estimation error about 6%. In all cases, this new method outperforms other overprint estimation methods, such as Kubelka-Munk.

Introduction

To find the ink formulation that meets a demanding list of requirements, numerous variables and their resulting attributes need to be screened. Tests typically involve physical preparation of ink candidates, printing with prototypes and measuring and evaluating the results. Completing all of this process requires time and manpower, which increases exponentially with the number of ink candidates, as their interaction needs also to be characterized. In this paper, we will focus on how to simulate color gamut from a reduced number of tests.

We aim at a model that allows to predict the color of ink combinations (also known as overprints, or secondary colors), from primary ink colors. Reasons are manifold. This model helps reducing testing resources and eases screening, as it might be challenging to obtain color charts in a reliable way for a large number of candidates (e.g. around 15). Also, since many initial tests are performed with primary colors, ink developers are interested in obtaining color gamut information from this already available data. In addition, having a model to predict color gamut from primary inks can be reversed to optimize critical parameters, such as pigment load. Finally, such a tool would also help in the creation of color profiles for multi-ink systems.

There are several techniques in the literature to estimate overprint colors, although they are rarely specific to ink jet printing. The Kubelka-Munk model [Haa92] has been widely used to predict the combination of inks, especially for paints, plastics and textiles. Van de Capelle and Meireson [Van97] proposed a method for film-based proofing systems, which consisted in measuring the spectra of each primary and only a secondary for each primary, in particular, each color on top of black ink at two different shades (e.g. 50% and 100%). A variation of the Hoffmann-Schmelzer model is proposed by Hasler et al. [Has04]. This method permits to estimate the color output of a printer if an ink or the substrate changes, without the need to measure color charts again. The model is adapted to the packaging industry (e.g. offset, flexo, gravure) but, according to the authors, not to ink-jet. The Neugebauer model and its numerous extensions [Rol93] are very popular to estimate the color of halftone prints, but those models take the overprints as inputs and do not provide means to estimate them.

We propose a model build exclusively from measurement data. The study and application of models found in the literature with our data showed some limitations, most particularly that apparently the same law does not hold for the different overprints. For example, the relationship between magenta and cyan to form blue does not hold between cyan and yellow to form green. Thus, theory-based methods need in any case some sort of calibration to take into consideration these deviations. We followed this rationale further and built the model solely by generalizing the primary–overprint relationships present in a training set. As it will be further clarified, our method makes an assumption on the invariability of the relationship as a function of wavelength.

This paper is organized as follows. In Section “Proposed Solution”, we describe the technical problem that is tackled and the proposed method. In Section “Results”, we present the results of this method with Latex-based inks, and compare them to other techniques proposed in the literature. Finally, conclusions and next steps are discussed in last section.

Proposed Solution

Objective

We want to predict the color resulting from printing two inks – one on top of the other – on paper. We are looking for a relationship between the reflectance spectra of primary inks and their respective combinations, which we call overprints.

Similar problems have been tackled in the literature, especially in the areas of paints and paint films [Haa92]. In the case of ink mixtures, the resulting color spectra corresponds to approximately a multiplicative relationship. Notably the Kubelka-Munk theory, developed for ink films, accurately estimates reflective spectra of ink films and its application is common in the paint industry. Kubelka-Munk has shown also good results for predicting ink mixtures on paper [Wyb00]. However, modeling ink as delivered by ink jet on paper is more challenging, as the number of variables increases, such as: mechanical dot gain, optical dot gain (internal light scattering), pigment interaction, heating, humidity – to name a few. In addition, the actual way in which a halftoning algorithm places dots on the media also affects the final color. In short, the greater the number of variables (including ones depending on the physics of the print), the lesser the accuracy of theoretical models.

Cited models aim at finding the relationship between primaries and overprints through an analytical relationship. Actually, from a practical point of view, the ideal method would be to estimate the gamut uniquely from primary color data, without printing any overprints. However, initial results showed poor performance with such techniques (>20dE_{76}). In some methods, certain parameters are fitted based on a training set. The study of actual data shows that this aimed relationship is not necessarily constant across different inks, reinforcing the fact that overprint phenomena are difficult to model precisely.

Method

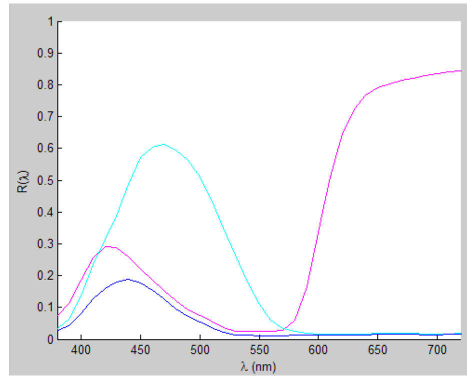
We propose to build the relationship function entirely from experimental data. A training set provides sparse 'ground truth' relationships, and interpolation is used to generate a continuous function from those. The main implicit assumption is that if the primaries have a certain pair of reflectances in a given wavelength band then the secondary will have the same reflectance regardless of what the wavelength band is. This is the key idea of our method.

In the evaluation of different ink candidates, for example, one set of candidates could be used as the training set – these need to be physically printed and measured – and the others estimated using this method. Similarly, when the different candidates correspond to a variation in one of the components of the ink (e.g. pigment loading), the training set can be composed of the most extreme values in the variation, and intermediate points can be estimated using this method.

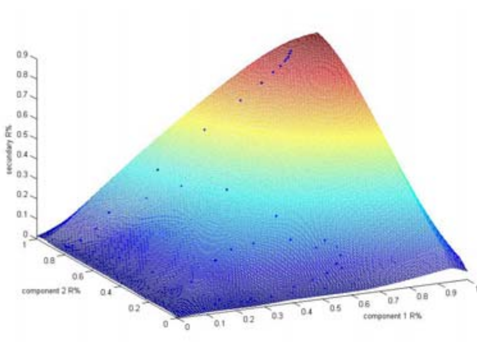
The problem of interpolating from scattered data points has been well documented in the literature. A general framework is that of Radial Basis Functions (RBF), and a particular case is bi-harmonic interpolation [San87]. This is one of the simplest RBFs implementations and provides a minimal-curvature interpolation by minimizing the bi-harmonic operator [Weiss]. Note also, that it is easily available in mathematical software packages (e.g. Matlab). To guarantee the convergence of the surface, it usually helps to apply a smoothing kernel before fitting the curve. The size of the kernel depends on the amount of data in the training set and its disparity.

The built relationship is a 3D mapping f_n: $[0,1]\text{x}[0,1] \rightarrow [0,1]$, for each overprint. The input intervals correspond to the normalized reflectance of the primary inks, and the output corresponds to the normalized reflectance of the overprint. To compute the overprint of a particular cyan and a particular magenta candidate, we would map though f_{blue} each of the values of their reflectance spectra. To give a specific example, assume the reflectance of cyan at 500 nm is 0.45, and magenta's is 0.12. Following our mapping, the overprint's reflectance would be f_{blue} (0.45, 0.12). This process would be repeated for all wavelengths in order to obtain the complete reflectance curve.

Figure 1 shows such a mapping for pigmented inks, specifically for the combination of cyan and magenta (blue). On the surface displayed on the right, the blue dots correspond to the samples of the training set, and the colored surface is the estimated response. The input axes (labeled as 'component 1' and 'component 2') correspond to the normalized reflectance of the primary inks, and the output axis (labeled as 'secondary') corresponds to the estimated spectral reflectance of the overprint. This surface has been obtained from the training data set shown at Figure 1.a.



***1.a)**. training set example*



***1.b)**.Mapping built from the training set 1.a*

***Figure 1**. The relationship between two primaries and a secondary is built from a training set. For each wavelength, a pair of input points is related to an output point. Thus, the relationship does not depend on wavelength.*

As mentioned, the built mapping is specific for each secondary. Also, steps that take place in the printer pipeline, and especially the ink limiting, need to be considered. For example, if the actual values of 100% cyan and 100% magenta are clipped inside the pipeline to 70% and 80% respectively, then these latter ones are the spectra that should be used – both in training and evaluation sets.

Results

The method was tested in two scenarios: with varying loadings of pigment and latex and with different pigment kinds. In both cases the objective was to optimize the gamut volume while maintaining other attributes.

Different Pigment and Latex Loadings

For different loadings, three test points for latex and three test points for pigment were included for each primary color (CMYK). The total number of primaries was 9 (3 pigment x 3 latex loadings) times 4 (colors), that is, a total of 36 candidates. The total number of potential secondary colors was a 486 combinations, per media, a prohibitive figure even for evaluating this method.. The 36 candidates were printed and measured (and labeled as 'L', 'M' and 'H' for 'low', 'medium' and 'high' loadings), on two different media, Avery MPI3000 calenderer permanent glossy (Vinyl) and Normandy Pro PVC Scrim Banner (Banner). Building the mapping from the corner cases only ('LL', 'LH', 'HL', 'HH'), one mapping for each secondary color and media, secondary colors for all other conditions were predicted, with an average accuracy error of 1.90 dE76, and a 95 percentile of 3.72 dE76 — as shown in Table 1. This represents considerably better performance compared to the alternative methods, which typically are above 10 dE76. Please notice that even the combinations that are included in the training set are not zero, due to smoothing interpolator used and measurement errors. Also, these results do not conclude that the same accuracy would be obtained for all potential 486 overprints (as we could not measure them), but as these are within the range of combinations screened in this experiments, deviations should be minimal.

Table 1: Color estimation error (dE_{76}) for different pigment/latex loadings. Labels 'L', 'M' and 'H' stand for 'low, 'medium' and 'high' loadings.

Pigment - Latex loading		Red			Green			Blue		
		L	M	H	L	M	H	L	M	H
Vinyl	L	1.8	0.4	0.7	0.5	2.4	1.3	2.7	2.5	0.8
	M	1.9	1.1	2.4	0.7	1.2	1.5	0.6	3.0	3.3
	H	1.1	1.9	2.3	0.9	1.3	1.0	1.6	1.0	6.2
Banner	L	0.9	1.8	1.0	0.9	1.0	0.2	1.2	3.0	3.3
	M	4.7	1.7	3.6	3.6	3.0	1.8	1.3	2.2	2.7
	H	2.3	1.3	2.2	3.2	3.8	1.4	0.2	1.0	0.5
Average / 95 prc		1.88 / 3.82			1.70 / 3.68			2.12 / 3.81		
		1.90 / 3.72								

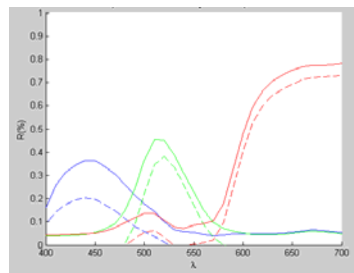
Different Pigment kinds

The test was conducted by printing color patches using the pigments from different kinds, as shown in Table 2, and comparing the measured and estimated gamut. Estimation of secondary colors had an average error of 5.1 dE_{76}, and a 95 percentile of 11.0 dE_{76}. Tests showed that color gamut was estimated with an error of 6%. Kubelka-Munk misestimated color gamut by 12% and secondary estimation was biased by an average of 12.4 dE_{76} and 95 percentile of 20.4 dE_{76}. Van de Capelle's method's error for secondary colors was much larger, and it varied with the specified ink order.

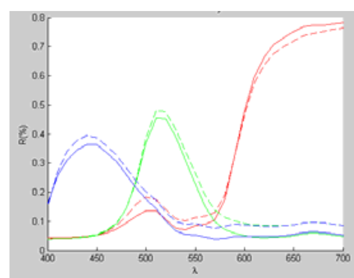
Table 2: Ink sets with different pigment kinds and loadings. Latex levels are constant.

	1	2	3	4
C	Type A, Low	Type A, Med	Type B, Low	Type B, Med
M	Type A, Med	Type A, High	Type B, Med	Type B, High
Y	Type B, Low	Type B, High	Type B, Low	Type B, High
K	Type B, Low	Type B, Med	Type B, Low	Type B, Med

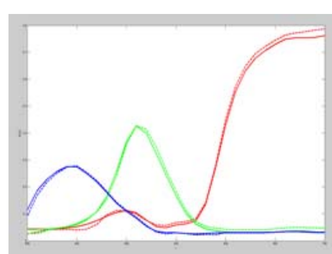
Figure 2 shows an example of the estimated and the predicted spectra by the three compared techniques.



***2.a)** Kubleka-Munk*



***.b)** Van de Capelle*



***2.c)** Proposed Method*

Figure2: Example of overprint estimation with a) Kubelka-Munk, b) Van de Capelle and 3) the proposed method.

As mentioned, an interesting outcome of this experiment is that mappings from measured data vary significantly form color to color. In other words, non data-driven models that do not take this into account will hardly perform on an acceptable level.

Conclusions

We presented a method to estimate overprints from primary reflectance spectral data. The goal was to be able to estimate color attributes as gamut from few samples, thus reducing the needed amount of experimentation. The method showed good results, estimating colors with little color differences, and in all cases outperforming classic color modeling methods such as Kubelka-Munk. Even within a limited accuracy, relative performance among ink candidates might be sufficient for candidate selection.

Acknowledgements

The authors would like to specially thank Karen Waggoner, Phil Cagle, Vicente Lázaro, Jesus García and Aleix Oriol for their collaboration and support.

References

[Haa92] C.S. Haase and G.W. Meyer, "Modeling pigmented materials for realistic image synthesis," ACM Transactions on Graphics, 11: 4, 305-335, 1992.

[Has04] D. Hasler, B. Zimmermann and P. Ehbets, "Efficient Characterisation of Printing Systems for the Packaging Industry," Proceedings of IS&T's NIP20: International Conference on Digital Printing Technologies, 20: 357-363, 2004

[Van97] J.P. Van De Capelle and B. Meireson, "A new method for characterizing output devices and its fit into ICC and HIFI color workflows," Proceedings of 5th CIC, 66-69, 1997.

[Rol93] R. Rolleston and R. Balasubramanian, "Accuracy of various types of Neugebauer model," Proceedings of First IS&T/SID Color Imaging Conference, Scottsdale, AZ, 2413, 32-37, 1993.

[Wyb00] D. R. Wyble and R. S. Berns, "Critical review of spectral models applied to binary color printing," Color Research and Application 25(1), pp. 4–19, 2000.

[Weiss] E. W. Weisstein, "Biharmonic Operator," From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/BiharmonicOperator.html

[San87] Sandwell, David T., "Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data," Geophysical Research Letters, 14: 2, 139-142, 1987.

Author Biography

Pau Soler obtained his MS in Telecommunication from the Universitat Politècnica de Catalunya (Barcelona, Spain) in 1999, and his PhD on image processing from École Nationale Supérieure des Télécomunications (Paris, France) in 2006. His Thesis work was done at Philips Research, under a Marie-Curie Fellowship. He is currently working as research scientist at Hewlett-Packard (Barcelona) on color image processing. His research interests include color management, color science, image processing and inverse problems. He is member of IS&T and SPIE.

Ján Morovič received his Ph.D. in color science from the Colour & Imaging Institute (CII) of the University of Derby (UK) in 1998, where the title of his thesis was To Develop a Universal Gamut Mapping Algorithm. After working as a lecturer in digital color reproduction at the CII he became senior color scientist at Hewlett-Packard in Barcelona, Spain. He was also the chairman of the CIE's Technical Committee 8–03 on Gamut Mapping.

Howard Doumaux received his B.S. in Chemical Engineering from Lehigh University in 1989, and his PhD in Chemical Engineering from the University of Minnesota in 1995. Howard has worked at Hewlett Packard as an ink chemist since 1997. While at Hewlett Packard Howard has worked with a wide range of ink chemistries utilizing both dye and pigment colorants, and has been awarded eight U.S. patents.