Colour Printer Characterization by Regression with Greyspace Constraint

***Mark S. Drew*⋆, *Graham D. Finlayson*†, *M. Ronnier Luo*†, & *Ján Morovič*†**

⋆School of Computing Science ***†The Colour & Imaging Institute***
Simon Fraser University ***The University of Derby***
Vancouver, B.C. V5A 1S6 ***Derby DE22 3BL***
Canada ***United Kingdom***
mark@cs.sfu.ca ***{graham,m.r.luo,j.morovic}@colour.derby.ac.uk***

Abstract

The White-Point Preserving Least Squares (WPPLS) algorithm is a method for colour correction that constrains the white point to be exactly mapped into its correct XYZ equivalent. For printers, however, the mapping is from device coordinates to colorimetric densities: the device white is thus mapped into a zero vector and the WPPLS method cannot go forward. Here we use a polynomial regression model and specify that both white (the zero vector) and an average grey be exactly mapped. Moreover we extend the method to accurately but approximately map a subspace of the entire achromatic curve, thus reproducing the neutral tones with far greater accuracy.

1. Introduction

An important consideration in colour printer calibration is finding an effective model for colour correction, i.e., mapping device RGB's or CMYK's to XYZ's. Of course, a lookup table provides the most straightforward method for characterizing colour printers; however one is still left with the problem of choosing an interpolation method, as well as possible storage problems. A simple printer model also has the advantage of capturing salient features of printer behaviour in model parameters. Understanding the structure of the model amounts to a better understanding of the device itself.

Luo et al.[1] showed that for a Cromalin proofing system, a Mitsubishi wax thermal-transfer printer, and an Iris inkjet printer a third-order masking model performed best for the colour correction task. In this paper, we show that a similar model can be adapted to the paradigm of constrained regression applied to printer models. We show the following: (1) Although one cannot apply the White-Point Preserving Least Squares (WPPLS) method[2, 3] directly to models based on colorimetric densities rather than tristimulus values, the method can be changed to preserve both the white point and a grey point as well, resulting in a Grey Point Preserving Least Squares (GPPLS) algorithm. The utility of this constrained regression is that errors along the curve of grey patches are reduced to zero at two locations, and diminished along the remainder of the curve. (2) As well, for this inherently nonlinear polynomial regression model one can in fact almost exactly map the entire achromatic curve in a constrained regression. Then, with little effect on the overall performance of such a regression, neutrals are reproduced almost exactly. We denote this new method as a Greyspace Preserving Least Squares (GSPLS) transform.

2. Polynomial Regression

For concreteness, throughout we study the behaviour of a Hewlett-Packard DeskJet 850C inkjet printer, characterizing the device separately for plain and glossy paper. For this printer, one cannot directly set CMYK values, but instead one can supply a file of device coordinates consisting of RGB values. Thus for the printer white point, we supply values RGB=(1,1,1) (normalizing to scale $0..1$) and the printer responds by depositing no colour for that pixel. We find that in fact this printer displays a reasonably linear relationship between RGB and the colorimetric densities $\log(\vec{X}_0/\vec{X})$, where $\vec{X}$ is the set of measured XYZ values and $\vec{X}_0$ are those for the printer white (see Fig.1). We start by finding a mapping from RGB to XYZ, since then we can use a perceptual ΔE measure to evaluate the mapping. The more practical mapping from desired XYZ to device RGB will be accurate if the reverse direction is accurate.

The WPPLS transform is based on performing a least squares regression from RGB's to XYZ's, but constraining the white point to be mapped exactly. It is based on a simple Least Squares (LS) regression.

2.1. Least Squares

Suppose we calibrate with a $5\times5\times5$ colour chart, and collect all $\xi \equiv \log(\vec{X}_0/\vec{X})$ measurements into an $n\times3$ matrix $\boldsymbol{H}$ where $n = 125$, and collect all RGB values into a similar matrix $\boldsymbol{Q}$. Here we wish to carry out a polynomial regression from $f(RGB)$ to XYZ. We form an 18-vector from each RGB triple, consisting of values R, G, B, R^2, G^2, B^2, RG, RB, GB, R^3, G^3, B^3, R^2G, R^2B, RG^2, G^2B, RB^2, GB^2. Thus matrix $\boldsymbol{Q}$ is $n \times 18$. Then the best least squares matrix $\boldsymbol{M}$ is that for which [1]

$$\boldsymbol{H} \simeq \boldsymbol{Q}\,\boldsymbol{M} \tag{1}$$

where $\boldsymbol{M}$ is 18×3.

The LS solution is, of course,

$$\boldsymbol{M} = (\boldsymbol{Q}^T\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\boldsymbol{H} \tag{2}$$

Since we note that eq.(2) is linear, it will map zero values into zero. However, colorimetric densities $\boldsymbol{H}$ are zero at the white point, whereas the elements of $\boldsymbol{Q}$ are zero at black. Therefore we alter the meaning of matrix $\boldsymbol{Q}$ so as to mean the 18-vector composed from values $(1 - RGB)$. Then white is zero, and is always exactly mapped using the LS transform. Note, however, that greys are not mapped exactly.

The first two rows of Table 1 show results for a LS transform for this printer, expressed as CIELAB ΔE error values for recovered XYZ tristimulus values, with white given by the paper white, for plain paper. Regression was performed on a $5 \times 5 \times 5$ color chart, and the resulting matrix was also applied to a separately measured achromatic scale consisting of 14 patches.

Algorithm	Min	Median	Mean	Max
LS (5x5x5)	0	6.20	6.98	24.77
LS (greys)	0	7.49	8.01	17.49
GPPLS (5x5x5)	0	6.83	7.54	24.64
GPPLS (greys)	0	4.43	6.11	13.61
GSPLS (5x5x5)	0	7.98	8.59	27.27
GSPLS (greys)	0	2.20	2.76	5.02

Table 1: *Plain paper. Statistics for CIELAB ΔE^* values comparing Least Squares (LS), Grey Point Preserving Least Squares (GPPLS), and Greyspace Preserving Least Squares (GSPLS) methods for 125 samples and 14 achromatic patches.*

Errors are not insubstantial, and Fig.2(a) shows the histogram for these errors. Errors for the 14 achromatic patches shown as vertical lines — errors for the grey scale are widely distributed.

2.2. Preserving a Grey Point: GPPLS

Although white is preserved automatically above, since it is represented as zero for both dependent and independent variables, we would like to apply a WPPLS [2, 3] approach to such printer models. In [3] a method for constraining the regression for higher dimensional models is presented. Matrix $\boldsymbol{M}$ is broken into two pieces, one denoted $\boldsymbol{D}$ that takes a constraint RGB point $\vec{\rho}_C$ into the correct tristimulus vector, and a second denoted $\boldsymbol{E}$ that preserves the constraint that $\vec{\rho}_C$ is mapped exactly:

$$\boldsymbol{M} \equiv \boldsymbol{D} + \boldsymbol{E} \tag{3}$$

[1]Errors in this printer model might be amenable to further reduction by the technique of finding error vectors for the achromatic scale and then remapping based on these corrections with gradually less influence as a function of chroma (cf. [4]); however in this work only the uncorrected results from regression are shown.

Now, here we are mapping an 18-vector $\vec{\rho}_C$ to a colorimetric density 3-vector $\vec{\xi}_C$. The problem is that we are representing white with both vectors equal to zero, and the WPPLS method cannot go forward. Subtracting each from unity will not be correct either, since then the model will have an offset. Instead, a simple approach is to insist that a distinguished point, the *average grey*, be mapped exactly. In this way both the white point as well as one grey point will be exactly mapped.

Matrix $\boldsymbol{D}$ is a higher-dimensional extension of a diagonal matrix relating $\vec{\rho}_C$ to $\vec{\xi}_C$; it consists of the pseudoinverse of $\vec{\rho}_C$ operating on $\vec{\xi}_C$:

$$\boldsymbol{D} = \vec{\rho}_C[(\vec{\rho}_C)^T\vec{\rho}_C]^{-1}(\vec{\xi}_C)^T \quad ,$$

$$(\vec{\rho}_C)^T\boldsymbol{D} = \vec{\xi}_C \tag{4}$$

We can further break matrix $\boldsymbol{E}$ into a part $\boldsymbol{Z}$ that automatically preserves the distinguished point and an arbitrary part $\boldsymbol{N}$,

$$\boldsymbol{E} = \boldsymbol{Z}\,\boldsymbol{N} \tag{5}$$

Then the job of the regression is to establish the best least squares $\boldsymbol{N}$.

The projector

$$\boldsymbol{P} = \vec{\rho}_C[(\vec{\rho}_C)^T\vec{\rho}_C]^{-1}(\vec{\rho}_C)^T \tag{6}$$

is the 18×18 matrix projecting onto the 1-dimensional subspace spanned by 18-vector $\vec{\rho}_C$. We need an 18×17 matrix $\boldsymbol{Z}$ orthogonal to $\vec{\rho}_C$, and for this we can take the set of eigenvectors of $\boldsymbol{P}$ spanning the complementary subspace. $\boldsymbol{P}$ has one eigenvalue equal to 1, and the rest 0. The eigenvectors for eigenvalues 0 make up matrix $\boldsymbol{Z}$. Then the solution for the best 17×3 matrix $\boldsymbol{N}$ that minimizes least squares under the constraint is [3]

$$\boldsymbol{N} = [\boldsymbol{Z}^T\boldsymbol{Q}^T\boldsymbol{Q}\,\boldsymbol{Z}]^{-1}[\boldsymbol{Z}^T\boldsymbol{Q}^T][\boldsymbol{H} - \boldsymbol{Q}\,\boldsymbol{D}] \tag{7}$$

Preserving grey, then, in this GPPLS model, takes the average (1-RGB) colour specifier, made into an 18-vector, into the average colorimetric density for the achromatic patches. [2] The 3rd and 4th rows of Table 1 show the results for this GPPLS method. As expected, overall GPPLS results are slightly worse (after all, LS necessarily produces *the* least squared error). However, errors for the achromatic patches are substantially reduced, as can also be seen in Fig.2(b).

3. Preserving Greyspace: GSPLS

Suppose we denote by $\boldsymbol{G}$ the set of 18-vectors formed from (1-RGB) values for a set of n achromatic patches. Then $\boldsymbol{G}$ is an $n \times 18$ matrix. We can ask the question: Is it possible to map all such grey values to their correct corresponding 3-vectors $\vec{\xi}$? Of course, the answer must be

[2]Once the average grey is calculated for any particular printer, units can be remapped so that that grey has units $(1, 1, 1)$ (cf.[2]); then matrix Z is fixed and need not be recalculated for each printer.

negative since there are not enough dimensions in the regression to allow for so many constraints. And in fact the more constraints the poorer overall should we expect the regression results to be, since we are pinning down more and more of matrix $\boldsymbol{M}$. Nevertheless we can expect to be able to approximately map a small subset of the grey patches if we are not too ambitious with respect to dimension n of preserved greys.

A reasonable size for our greyspace might be 5, since we here started with a $5 \times 5 \times 5$ RGB cube; let us take for the greyspace set $\boldsymbol{G}$ five evenly-spaced grey patches not including white. (The reason for excluding white is that white is represented in matrix $\boldsymbol{Q}$ and in matrix $\boldsymbol{H}$ as all zeros.) Thus we try using $n = 5$ and a 5×18 matrix $\boldsymbol{G}$.

However, we cannot expect $\boldsymbol{G}$ to have full rank, and in fact it turns out to be rank $r = 3$. [3] Denote by $\boldsymbol{L}$ the set of colorimetric densities $\vec{\xi}$ corresponding to these grey patches. Then $\boldsymbol{L}$ is an $n \times 3$ matrix.

If $\boldsymbol{G}$ is rank r, a Singular Value Decomposition (SVD) of $\boldsymbol{G}$ can be written

$$\boldsymbol{G} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{V}^T \tag{8}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices; $\boldsymbol{U}$ is $n \times r$ and $\boldsymbol{V}$ is $r \times 18$. Matrix $\boldsymbol{\Lambda}$ is diagonal and $r \times r$. The Moore-Penrose pseudoinverse of $\boldsymbol{G}$ is denoted $\boldsymbol{G}^+$, and is given by

$$\boldsymbol{G}^+ = \boldsymbol{V} \boldsymbol{\Lambda}^{-1} \boldsymbol{U}^T \tag{9}$$

Then we have that

$$\boldsymbol{G} \boldsymbol{G}^+ = \boldsymbol{U} \boldsymbol{U}^T \tag{10}$$

Now note that although we must have that the $r \times r$ matrix $(\boldsymbol{U}^T \boldsymbol{U})$ equals the $r \times r$ identity matrix $\boldsymbol{I}_r$ (which is $\boldsymbol{I}_3$, here), matrix $(\boldsymbol{U} \boldsymbol{U}^T)$ is *not* the identity matrix $\boldsymbol{I}_n$ (which is $\boldsymbol{I}_5$, here), but is only close to the identity. How close depends on how rank-reduced matrix $\boldsymbol{G}$ is.

Here, we could in fact simply select 3 grey patches out of matrix $\boldsymbol{G}$ such that the subset produces a new rank-3 matrix $\boldsymbol{G}$. Then we would be mapping those three greys *exactly* in the GSPLS transform developed in the next three equations. However, that would leave part of the grey curve uncontrolled. Therefore here we keep all five patches in matrix $\boldsymbol{G}$ and consequently map those patches not exactly but only approximately.

To carry out a GSPLS, we again form matrix $\boldsymbol{D}$ akin to that in eq.(4), but now form it by application of the pseudoinverse of the grey space, $\boldsymbol{G}^+$, to the colorimetric densities for the greys, which form matrix $\boldsymbol{L}$:

$$\boldsymbol{D} = \boldsymbol{G}^+ \boldsymbol{L} \tag{11}$$

so that

$$\boldsymbol{G} \boldsymbol{D} \simeq \boldsymbol{L} \tag{12}$$

[3]Matrix rank can be obtained using a QR decomposition[5]. Any collection of 3 or more 18-vectors made from distinct patches with $R = G = B$ will have rank 3, barring noise.

Again, we need the 18×18 projector $\boldsymbol{P}$, which is here the projector onto the greyspace:

$$\boldsymbol{P} = \boldsymbol{G}^+ \boldsymbol{G} = \boldsymbol{V} \boldsymbol{V}^T \tag{13}$$

Then matrix $\boldsymbol{Z}$ is the set of $(18 - r) = 15$ eigenvectors of $\boldsymbol{P}$ which are orthogonal to the grey subspace. The solution for matrix $\boldsymbol{N}$ is again given by the same equation (7), where now $\boldsymbol{Z}$ is 18×15 and $\boldsymbol{N}$ is 15×3. They combine as in eqs.(3, 5) to form an 18×3 matrix $\boldsymbol{M}$.

The last two rows of Table 1 show the results for this GSPLS method, for plain paper. Again, the overall performance is slightly degraded, as expected, since this transform is not the 'least' possible. Nevertheless the performance for this GSPLS method is not greatly changed from that for the GPPLS one, which preserves a single grey point. However, there is a great improvement in the accuracy of mapping the achromatic patches, as can be seen in Fig.2(c).

4. Glossy Paper

Results for glossy paper are given in Table 2. As can be seen, these results are not much different than those for plain paper except that the error reduction for achromatic patches is somewhat better than for plain paper.

Algorithm	Min	Median	Mean	Max
LS (5x5x5)	0	8.36	9.08	27.00
LS (greys)	0	7.83	8.70	16.40
GPPLS (5x5x5)	0	9.33	10.50	27.42
GPPLS (greys)	0	3.83	3.78	7.19
GSPLS (5x5x5)	0	8.92	10.54	27.69
GSPLS (greys)	0	2.13	2.22	4.79

Table 2: Glossy paper. CIELAB ΔE^* values comparing LS, GPPLS, and GSPLS methods for 125 samples and 14 achromatic patches.

5. Conclusion

We have shown that it is possible to adapt the constrained regression method, based on mapping colour values to tristimulus values, to printer models for mapping device coordinates to colorimetric densities. Moreover, we have shown how to generalize the constrained regression method so as to accurately map part or all of the achromatic curve for a printer.

Of course, one could instead attempt to simply carry out a standard grey balancing of RGB values. But this is too coarse a coordinate change, even having a nonlinear polynomial regression at one's disposal following the grey balance: here a grey patch has $R = G = B$ so there is only one tone curve. For plain paper, we found that grey balancing followed by regression produced a maximum ΔE^* value of 48, and the minimum ΔE^* even for the neutrals was 5.2: these values are much better for the GSPLS method, as seen in Table 1.

The present method involves many tradeoffs, and it should be investigated to what degree one can increase the rank of the accurately mapped greyspace without too greatly decreasing overall accuracy. Also, the third-order model employed may not be the best to use in conjunction with the constrained regression method. Finally, the choice of just which grey patches to use should be further explored — we have found that this choice does make a difference, and possibly a substantial one. Clearly, one should choose patches that broadly span the printer's tone range, but just how to do so is an open question. E.g., it may be best to choose greys that lie on the most linear portion of the achromatic curve. The problem is something like choosing knots in a spline curve fit: if knots are too far from some data points unexpected variations in the curve can result. What is required is a compromise striking the best balance and choices made in this report are likely not yet optimal. Nevertheless, the method as presented does what it sets out to do, which is to more accurately map the greys.

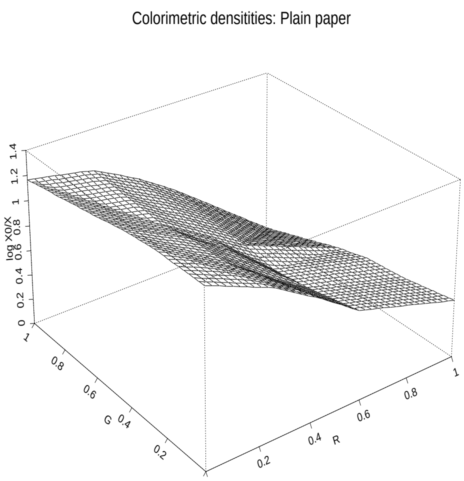

Figure 1: *Colorimetric densities as a function of input RGB file values (for fixed Blue), 850C printer, plain paper.*

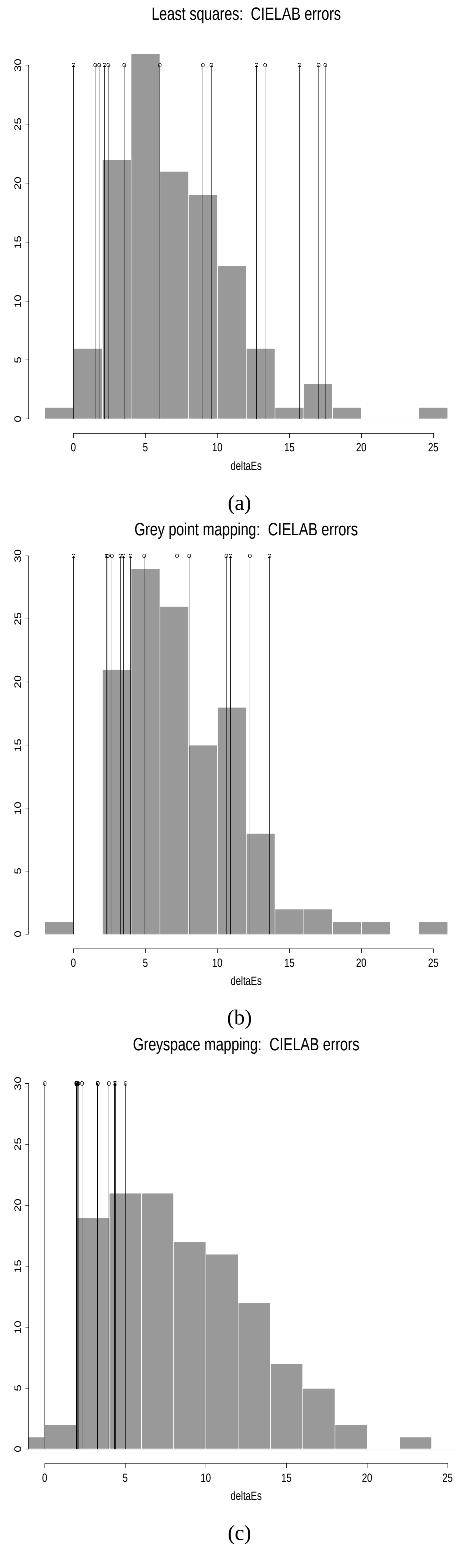

Figure 2: *Plain paper. Histogram of CIELAB errors for (a) LS, (b) GPPLS, and (c) GSPLS regressions. Results for regression matrices applied to a separately measured achromatic scale are shown as vertical lines.*

References

[1] M.R.Luo, A.J. Johnson, J.H. Xin, and S.A.R. Scrivener. Mathematical models for characterizing printing devices. *SID Digest*, 23:17–22, 1992.

[2] G.D. Finlayson and M.S. Drew. White–point preserving color correction. In *5th Color Imaging Conference*, pages 258–261, 1997.

[3] G.D. Finlayson and M. S. Drew. Constrained least–squares regression in color spaces. *Journal of Electronic Imaging*, 6:484–493, 1997.

[4] Ján Morovič and M. Ronnier Luo. Gamut mapping algorithms based on psychophysical experiment. In *5th Color Imaging Conference*, pages 44–49, 1997.

[5] G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins U. Press, 1983.