Journal of Electronic Imaging 6(4), 484–493 (October 1997).

Constrained least-squares regression in color spaces

Graham D. Finlayson
University of Derby
Colour Research Group
Derby, United Kingdom DE22 3BL

Mark S. Drew
Simon Fraser University
School of Computing Science
Vancouver, BC, V5A 1S6, Canada

Abstract. *To characterize color values measured by color devices (e.g., scanners, color copiers, and color cameras) in a device-independent fashion these values must be transformed to colorimetric tristimulus values. The measured RGB 3-vectors are not a linear transformation away from such colorimetric vectors, however, but still the best transformation between these two data sets, or between RGB values measured under different illuminants, can easily be determined. Two well-known methods for determining this transformation are the simple least-squares fit (LS) procedure and Vrhel's principal component method. The former approach makes no a priori statement about which colors will be mapped well and which will be mapped poorly. Depending on the data set a white reflectance may be mapped accurately or inaccurately. In contrast, the principal component method solves for the transform that exactly maps a particular set of basis surfaces between illuminants (where the basis is usually designed to capture the statistics of a set of spectral reflectance data) and hence some statement can be made about which colors will be mapped without error. Unfortunately, even if the basis set fits real reflectances well this does not guarantee good color correction. Here we propose a new, compromise, constrained regression method based on finding the mapping which maps a single (or possibly two) basis surface(s) without error and, subject to this constraint, also minimizes the sum of squared differences between the mapped RGB data and corresponding XYZ tristimuli values. The constrained regression is particularly useful either when it is crucial to map a particular color with great accuracy or when there is incomplete calibration data. For example, it is generally desirable that the device coordinates for a white reflectance should always map exactly to the XYZ tristimulus white. Surprisingly, we show that when no statistics about reflectances are known then a white-point preserving mapping affords much better correction performance compared with the naive least-squares method. Colorimetric results are improved further by guiding the regression using a training set of measured reflectances; a standard data set can be used to fix a white-point-preserving regression that does remarkably well on other data sets. Even when the reflectance statistics are known, we show that correctly mapping white does not incur a large colorimetric overhead; the errors resulting from white-point preserving least-squares fitting and straightforward least-squares are similar.* © *1997 SPIE and IS&T.* [S1017-9909(97)01204-X]

1 Introduction

Color sensors in scanners, color copiers, and color cameras are not colorimetric, in the sense that device RGB values are not a linear transformation away from the X, Y, Z tristimulus values[1] that would be produced by integrating the color signal impinging on the optical system with Commission Internationale de L'Éclairage (CIE) color-matching functions. The transformation from RGB to XYZ forms the first step in developing a device-independent description of color for these devices.[2] Given a particular set of targets or dyes one can map from RGB to XYZ using interpolation and lookup tables,[3–5] polynomial regression,[6] or spectral reconstruction from a set of basis functions.[7] More elegantly the device coordinates can be transformed to visual tristimuli using a single linear transform (3×3 matrix). This transform is often defined to be the best linear least-squares mapping for a particular calibration set of surface reflectances. Alternately, we could choose to map any three surfaces exactly and hope that this mapping would work for other reflectances as well. Vrhel[8] has demonstrated that if the mapping is defined relative to the first three principal components (those surfaces that optimally capture the variance) of the spectral surface reflectances of the calibration set then reasonable color correction results are often obtained for the entire set.

There are points for and against the least-squares regression and principal components methods. The least-squares regression optimally maps RGBs to XYZs so that the colorimetric error for a calibration data set is minimized. However, there is no mechanism for choosing the surfaces that will be mapped well and those that will be mapped poorly. For example, for one calibration set a white reflectance may be mapped exactly and for another it may be mapped with a high colorimetric error. Given the importance of white in color reproduction, we would rather not have this variable performance. In contrast, in the principal component method the surfaces that are mapped correctly are explicitly stated and this means that we can choose to map white by constraining one principal component to be a white reflectance. Unfortunately, the principal components method does not deliver the same mapping performance as simple least-squares regression. Indeed there is no reason that spectra which are statistically meaningful (e.g., principal

Paper 95-026 received Aug. 15, 1995; revised manuscript received Feb. 27, 1997; accepted for publication June 27, 1997.
1017-9909/97/$10.00 © 1997 SPIE and IS&T.

components) should supply a good basis for choosing the color correction transform. In the worst case, it is possible that a principal component spectrum might be orthogonal to (invoke a zero response in) the scanner (or other color input device) and as such color correction would be poor.

In this paper we develop a compromise constrained regression. Specifically we develop methods for determining the best least-squares transform that takes RGBs to XYZs subject to the constraint that the RGB responses induced by a single (or possibly two) surface reflectance(s) are mapped without error. Because it is important to map white correctly for color reproduction, we develop and examine in detail a white-point preserving least-squares (WPPLS) regression. Any constrained mapping is no longer optimal in the usual sense—it must produce larger errors than a least-squares regression for the particular calibration reflectance set used. However, it is interesting to ask how the transform derived for one calibration set fares when it is applied to other data. For example, it is fairly common to use a calibration set that contains all reflectances with equal likelihood (the maximum ignorance set). The results of simulation experiments demonstrate that the transform calculated using maximum ignorance WPPLS regression performs well when applied to real scanner or camera data (e.g., data that doesn't follow the maximum ignorance assumption). In contrast the maximum ignorance LS transform performs poorly on real scanner or camera data and white is mapped with a very large colorimetric error. Even better results can be obtained by altering the maximum-ignorance equations by applying covariance information from another, known, set of reflectances to the unknown set being imaged. The value of a constrained regression procedure is thus demonstrated.

2 Color Space Data Transforms

In this section we consider how scanner or camera RGBs can be mapped to XYZ tristimulus values. For concreteness, we shall focus the discussion on scanners although derived results apply equally well to color copiers and color cameras. In Section 6 calculations are repeated for a color camera in order to demonstrate that results apply across devices.

2.1 *Image Formation Model*

Consider the RGB scanner responses $\boldsymbol{\rho}$ for a given pixel. Color scanners are either one-pass or three-pass: in a one-pass device a fluorescent illumination source $E(\lambda)$ is directed at the surface, with spectral reflectance function $S(\lambda)$, and the resulting color signal, the product of $E(\lambda)$ and $S(\lambda)$, impinges on the imaging system. Here, three filters $\mathbf{Q}(\lambda)$ form the color three-vector ρ via

$$\boldsymbol{\rho}=\int E_s(\lambda)S(\lambda)\mathbf{Q}(\lambda)d\lambda \tag{1}$$

where $E_s(\lambda)$ denotes the illuminant in the scanner.

In a three-pass system, the surface is imaged under three differently colored lights in turn. In this case, the triple of illuminants $\mathbf{E}_s(\lambda)$ and the single imaging system response $Q(\lambda)$, that includes the mirror and CCD, are combined to form the response vector:

$$\rho=\int \mathbf{E}_s(\lambda)S(\lambda)Q(\lambda)d\lambda. \tag{2}$$

In practice we do not use continuous functions for spectra. Sampling the visible spectrum, 400 to 700 nm, at intervals of 10 nm amounts to replacing all spectra by vectors with 31 elements. Let us denote column vectors and matrices by boldface lower case and upper case, respectively. Matrix $\mathbf{Q}$ is 31×3.

In terms of vectors, then, Eq. (1) for a one-pass scanner reads

$$\boldsymbol{\rho}^T=\mathbf{s}^T\mathbf{L}(\mathbf{e}_s)\mathbf{Q} \tag{3}$$

where ^{T} means transpose; here, $\mathbf{L}(\mathbf{e}_s)$ is a diagonal 31×31 matrix formed from the illuminant vector $\mathbf{e}_s$ sampled from $E_s(\lambda)$, and similarly $\mathbf{s}$ is the sampled surface reflectance vector.

Rewriting the three-pass scanner Eq. (2) in terms of vectors, we have

$$\boldsymbol{\rho}^T=\mathbf{s}^T\mathbf{L}(\mathbf{q})\mathbf{E}_s \tag{4}$$

where $\mathbf{E}_s$ is a 31×3 matrix of the three scanner illuminants, and $\mathbf{L}(\mathbf{q})$ is a diagonal matrix formed from the sampled optical system response.

Colorimetric tristimulus values X,Y,Z (which we simply call an XYZ vector and denote by $\mathbf{x}$) are defined by a similar equation:

$$\mathbf{x}^T=\int E_v(\lambda)S(\lambda)\mathbf{X}(\lambda)d\lambda,$$

or

$$\mathbf{x}^T=\mathbf{s}^T\mathbf{L}(\mathbf{e}_v)\mathbf{X}. \tag{5}$$

Here, the matrix $\mathbf{X}$ sampled from the set of color matching functions $\mathbf{X}(\lambda)$ is linearly related to the responses of human observers to color; the columns of $\mathbf{X}(\lambda)$ are usually denoted $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$, but here we collectively denote them by $\mathbf{X}(\lambda)$ for consistency with Eq. (1). The illuminant $E_v(\lambda)$, with sampled vector $\mathbf{e}_v$, is the illuminant for which we want tristimulus values: $\mathbf{x}$ is tied to a particular viewing illuminant. Typically, values of $\mathbf{x}$ under standard illuminant D65 are required.[1]

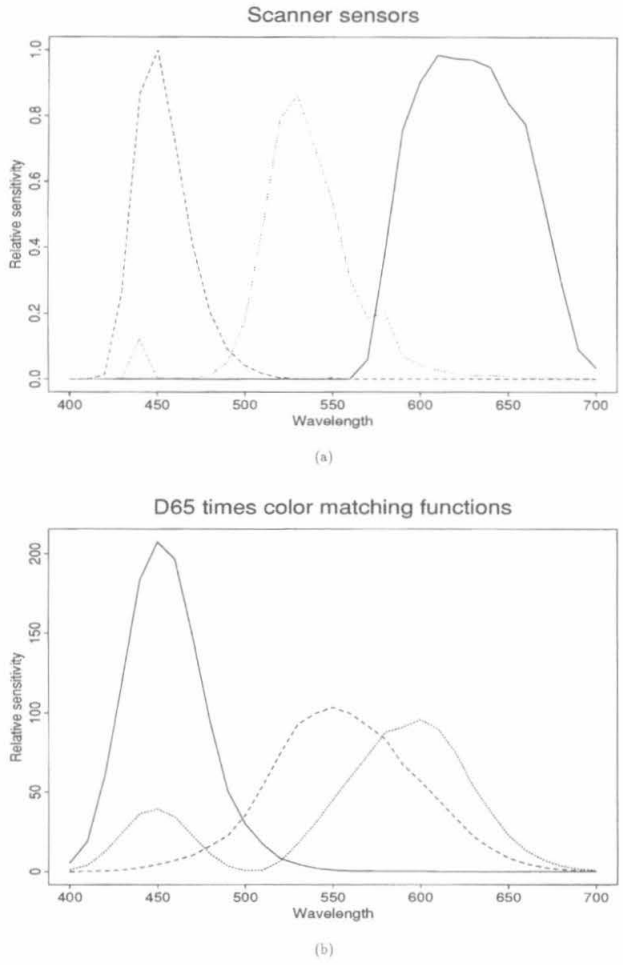


Fig. 1 (a) Response functions $Q_E(\lambda)$ for Sharp JX450 scanner. (b) $X_E(\lambda)$: Color matching functions times D65.

As a sensor set, we will use the measured sensor curves for the Sharp JX450 scanner.[2] This is a three-pass device.* Figure 1(a) shows the three functions $\mathbf{E}_s(\lambda)Q(\lambda)$. In matrix form we denote the three illuminants multiplied by the single sensor as $\mathbf{Q}_E$. The 31×3 matrix of color matching curves multiplied by a D65 illuminant is denoted as $\mathbf{X}_E$; we plot $\mathbf{X}_E$ in Fig. 1(b).

2.2 *Least-Squares Regression*

Below, we give a formal derivation of the least-squares regression method. While this technique is adequately explained elsewhere we include it here since it both motivates and simplifies the explanation of the constrained least-squares regression that follows this section. Under the least-squares regression a set of $\boldsymbol{\rho}$ vectors are mapped to the corresponding set of $\mathbf{x}$ vectors.[9] First let us collect all measurements $\boldsymbol{\rho}_I$ for $I=1...N$ surfaces into an N by 3 matrix $\mathbf{R}$; i.e., we stack all instances of $\boldsymbol{\rho}^T$. Similarly, we collect the corresponding $\mathbf{x}_I$ vectors into an N by 3 array $\mathbf{H}$.

*The scanner data for the Sharp JX450 are due to Brian Wandell and Joyce Farrell.

We wish to map scanner responses to their colorimetric counterparts. Assuming a linear mapping, the following objective function minimizes the sum of squared residuals for such a mapping:

$$\mathcal{J}=\sum_{I=1}^{N}\|\mathbf{x}_I^T-\boldsymbol{\rho}_I^T\mathbf{M}\|^2, \tag{6}$$

where $\mathbf{M}$ is the unknown transformation matrix sought. Taking partial derivatives with respect to the elements of $\mathbf{M}$, we arrive at the Euler equations

$$\mathbf{R}^T(\mathbf{H}-\mathbf{RM})=0, \tag{7}$$

so that a least-squares solution is

$$\mathbf{M}=(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{H}. \tag{8}$$

Let us adopt as an exemplar test data set the 462 surface spectral reflectance functions measured by Newhall *et al.*[10] These consisted of paint chips from the Munsell Book of Color[11] and, for the purposes of our simulation experiments, constitute typical reflectances that might be scanned. Hence the number of surfaces N used in our data set is 462. While the sum of squared residuals is a reasonable error measure it does not directly correlate with human visual performance. To remedy this we express the residual difference in terms of the CIELAB, or $L*a*b*$, error measure. A CIELAB residual, or ΔE^*_{ab}, corresponds roughly to human judgements of perceptual difference. CIELAB errors of 2 or 3 represent just noticeable color differences detectable by humans.[12]

The colorimetric performance of a least-squares fit for a linear map

$$(\mathbf{H}\simeq\mathbf{RM}) \tag{9}$$

is given, using the CIELAB measure, in the first row of Table 1. Small mean and median CIELAB errors result—2.1 and 1.3—both of which are not perceptually noticeable. Least-squares performance serves as a control for the other results presented in this paper.

The LS transform makes no *a priori* assumption about the surfaces that will be mapped with low error and those that will be mapped with high error. In particular, we do not expect the white point to be preserved even although we might like this to be the case given the importance of white for color reproduction. If we take "white" to be the scanner output for a uniformly reflecting surface, we find a CIELAB ΔE^*_{ab} error of 1.406 units for the white point for the regression described above. That this is small is a property of the Munsell data set we used; there is no reason why a larger error might not have occurred. Indeed, before performing the regression we simply have no idea what the error in white will be; it just so happens that in this case we were fortunate. To ensure the correct mapping of white (or any other surface) we need to use a constrained least-squares regression.

Table 1 Scanner. Top two rows: statistics for CIELAB ΔE^*_{ab} values comparing LS and WPPLS methods for 462 Munsell samples (see text in Section 3.1). Bottom two rows: statistics for CIELAB ΔE^*_{ab} values comparing polynomial LS and polynomial WPPLS methods (see text in Section 3.3).

	Wt.Pt.ΔE^*_{ab}	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell LS	1.41	0.03	1.29	2.10	11.50	77%
Munsell WPPLS	0	0.073	1.58	2.37	11.17	84%
Munsell PLS	1.39	0.04	1.012	1.58	8.34	85%
Munsell PWPPLS	0	0.05	1.074	1.59	8.42	85%

3 Constrained Data Transforms

We wish to develop a method for mapping RGBs to XYZs such that the sum of least-squares residuals is minimized subject to the constraint that a particular surface reflectance (e.g., white) is mapped without error. To do so, we apply a constraint to the optimization for matrix **M**.

3.1 *Constrained Least-Squares Regression*

Let $\boldsymbol{\rho}^C$ denote the scanner RGB vector for a constraint surface (a surface we wish to map without error). Let $\mathbf{x}^C$ denote the corresponding constraint surface XYZ tristimulus vector. We can ensure that the constraint surface is mapped without error by augmenting Eq. (6) with a Lagrange multiplier term:

$$\mathcal{J}=\sum_{I=1}^{N}\{\|\mathbf{x}_I^T-\boldsymbol{\rho}_I^T\mathbf{M}\|^2\}+\Lambda\sum_{k=1}^{3}\left(x_k^C-\sum_{j=1}^{3}\rho_j^C M_{jk}\right)^2. \quad (10)$$

Here, Λ is a Lagrange multiplier that captures the idea that the constraint surface in RGB coordinates $\boldsymbol{\rho}^C$ must go over to the correct XYZ vector $\mathbf{x}^C$. Note that metamers for the scanner are unlikely to be the same as for the human eye. It follows that the constrained least-squares regression correctly maps RGB to XYZ for the unique constraint surface but will incorrectly map other scanner metamers (of the constraint surface). Taking partial derivatives with respect to the Lagrange multiplier yields the constraint condition

$$\|(\mathbf{x}^C)^T-(\boldsymbol{\rho}^C)^T\mathbf{M}\|=0 \quad (11)$$

and taking derivatives with respect to the elements M_{jk} of **M**, we obtain

$$\sum_{I=1}^{N}\left\{\rho_{Ij}\left(x_{Ik}-\sum_{h=1}^{3}\rho_{Ih}M_{hk}\right)\right\}+\Lambda\sum_{k=1}^{3}\rho_j^C\left(x_k^C-\sum_{h=1}^{3}\rho_h^C M_{hk}\right)=0 \quad (12)$$

where ρ_{Ij} is the j'th element of set I. Clearly, the solution of Eqs. (11) and (12) is that the constraint surface is correctly mapped,

$$(\mathbf{x}^C)^T=(\boldsymbol{\rho}^C)^T\mathbf{M}, \quad (13)$$

and, once again, **M** satisfies Eq. (7).

However, now **M** must be of a form that obeys the constraint (13). Any such matrix **M** must be of the form

$$\mathbf{M}=\mathbf{D}+\mathbf{E} \quad (14)$$

with **D** that diagonal matrix formed from the ratios of $\mathbf{x}^C$ and $\boldsymbol{\rho}^C$, and **E** any matrix that satisfies $(\boldsymbol{\rho}^C)^T\mathbf{E}=0$. Every such matrix **E** can be written

$$\mathbf{E}=\mathbf{ZN}, \quad (15)$$

where **Z** is a 3×2 matrix composed of any two vectors $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ orthogonal to $\boldsymbol{\rho}^C$, and **N** is an arbitrary 2×3 matrix.

Substituting Eqs. (14) and (15) into Eq. (7), we can solve for **N** by premultiplying by $\mathbf{Z}^T$; then we have

$$\mathbf{Z}^T\mathbf{R}^T\mathbf{H}-\mathbf{Z}^T\mathbf{R}^T\mathbf{RD}=[\mathbf{Z}^T\mathbf{R}^T\mathbf{RZ}]\mathbf{N} \quad (16)$$

in terms of unknown **N**. This is precisely the same Euler equation as one arrives at starting from the minimization (6) with **M** in the special form of Eqs. (14) and (15). The matrix **M** is comprised of the diagonal **D** augmented by a rank-2 matrix that does not affect the mapping of the constraint surface.

Solving, we have

$$\mathbf{N}=[\mathbf{Z}^T\mathbf{R}^T\mathbf{RZ}]^{-1}[\mathbf{Z}^T\mathbf{R}^T\mathbf{H}-\mathbf{Z}^T\mathbf{R}^T\mathbf{RD}], \quad (17)$$

where in the above we are inverting a 2×2 matrix.

When the constraint surface equals a white reflectance we call the constrained least-squares regression white point preserving least-squares (or WPPLS). We expect that this WPPLS regression will not do as well as the LS method since we have constrained matrix **M**. This expectation is borne out in Table 1, which shows that for our exemplar data set the WPPLS results are slightly worse than for LS. These figures could be improved for surfaces that best obey Eq. (9) by carrying out a robust regression instead of a least-squares regression,[13] but at the expense of increasing errors for surfaces that obey Eq. (9) less well. However, independent of the measure of error we choose (simple least-squares or robust counterparts such as least median of squares) we are always confident that the white point (or constrained surface in general) is mapped without error.

Mapping the constraint surface without error amounts to allowing matrix **N** to account for a best least-squares fit in a plane orthogonal to $\boldsymbol{\rho}^C$. We could, if desired, extend the constrained regression idea and fix two distinguished directions and thus preserve two different color points (and hence two constraint surfaces); then **N** would be restricted to accounting for the best least-squares fit in a single direction orthogonal to the plane spanned by those directions.

Preserving no color points (zero constraint surfaces) yields the simple least-squares regression. Correctly mapping three color points, where these are defined by the principal components of the reflectances under consideration, yields Vrhel's color correction method.[8]

3.2 *Constrained Regression for Higher Dimensions*

The above analysis can be extended to the case of more than three scanner sensors. In the case that there are p sensors the matrix $\mathbf{M}$ has dimensions $p \times 3$, so that

$$\mathbf{H} \approx \mathbf{RM} \tag{18}$$

with an $N \times p$ ensemble of scanner measurements $\mathbf{R}$ and an $N \times 3$ ensemble $\mathbf{H}$.

In Section 3.1 we had broken $\mathbf{M}$ into a diagonal part $\mathbf{D}$ and additional part $\mathbf{E}$, with $\mathbf{D}$ defined by

$$(\mathbf{x}^C)^T = (\boldsymbol{\rho}^C)^T \mathbf{D}. \tag{19}$$

In fact, $\mathbf{D}$ above is composed of two parts. The first part is simply the pseudoinverse of $(\boldsymbol{\rho}^C)^T$ operating on $(\mathbf{x}^C)^T$:

$$\mathbf{D} = (\boldsymbol{\rho}^C)[(\boldsymbol{\rho}^C)^T(\boldsymbol{\rho}^C)]^{-1}(\mathbf{x}^C)^T. \tag{20}$$

The second part is constraint surface preserving; that is, its columns are orthogonal to $(\boldsymbol{\rho}^C)^T$. Therefore, in the 3×3 case we can use Eq. (20) for $\mathbf{D}$ instead of a diagonal matrix (since $\mathbf{D}$ defined in Eq. (20) will not in general be a diagonal matrix). This reformulation is especially useful in higher dimensions; in that case $\mathbf{D}$ is $p \times 3$ because it is formed as the outer product of a p-vector with a three-vector.

Matrix $\mathbf{E}$ is the constraint-surface preserving part of matrix $\mathbf{M}$; it is composed of a $p \times (p-1)$ factor premultiplying a $(p-1) \times 3$ one. The first factor is a matrix whose columns span the subspace complementary to the one-dimensional subspace spanned by $\boldsymbol{\rho}^C$.

A simple way to find a basis for this subspace is to first form the projector $\mathbf{P}$ onto the one-dimensional subspace:

$$\mathbf{P} = \boldsymbol{\rho}^C[(\boldsymbol{\rho}^C)^T\boldsymbol{\rho}^C]^{-1}(\boldsymbol{\rho}^C)^T, \tag{21}$$

a $p \times p$ symmetric, idempotent matrix.

One eigenvalue of P is 1, and the rest are zero. Eigenvectors corresponding to the zero eigenvalues span the complementary subspace. In the case of p sensors, this submatrix is $p \times (p-1)$. Denote by $\mathbf{Z}$ the $p \times (p-1)$ matrix orthogonal to $\boldsymbol{\rho}^C$.

Thus the constraint-surface preserving part of $\mathbf{M}$ is $\mathbf{E} = \mathbf{ZN}$, with $\mathbf{N}$ an unknown $(p-1) \times 3$ matrix to be solved for by regression. The equation for the unknown $\mathbf{N}$ is again given by Eq. (17), but with higher dimension matrices.

Below, in Section 4.2, we give results for an example using six sensors.

3.3 *Constrained Multiple Polynomial Regression*

Berns and Shyu[7] and Vrhel[14] have shown that color correction performance can be improved using a polynomial expansion of the calibration data set. Importantly, we show here that a similar polynomial expansion can also improve the colorimetric performance that is delivered by the constrained regression method.

As before, let $\mathbf{R}$ denote an $N \times 3$ matrix of RGB measurements and $\mathbf{H}$ the corresponding matrix of XYZ tristimuli. As a first toward presenting the polynomial method, it is useful to expand the matrices $\mathbf{R}$ and $\mathbf{H}$ so that their column vectors are made explicit:

$$\mathbf{R} = [\mathbf{rgb}], \quad \mathbf{H} = [\mathbf{xyz}]. \tag{22}$$

Here, $\mathbf{r}$, $\mathbf{g}$, and $\mathbf{b}$ are $N \times 1$ vectors corresponding to the R, G, and B device measurements. Similarly, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are $N \times 1$ vectors of X, Y and Z tristimulus coordinates.

The basic idea underpinning linear color correction is that each column of $\mathbf{H}$ can be written as a linear combination of the columns of $\mathbf{R}$. For example, $\mathbf{x} = \alpha\mathbf{r} + \beta\mathbf{g} + \gamma\mathbf{b}$ (where α, β, and γ are scalars). In second-order polynomial color correction each column of $\mathbf{H}$ is represented as a linear combination of not just the columns of $\mathbf{R}$, but also the columns of $\mathbf{R}$ squared $(\mathbf{r}^2, \mathbf{g}^2, \mathbf{b}^2)$. Moreover, cross-column terms, for example, $\mathbf{rg}$, and a translational term are also added into the mix. The $N \times 3$ matrix $\mathbf{R}$ is expanded to the $N \times 10$ matrix $\mathbf{R}'$:

$$\mathbf{R}' = [\mathbf{r}\ \mathbf{g}\ \mathbf{b}\ \mathbf{r}^2\ \mathbf{g}^2\ \mathbf{b}^2\ \mathbf{rg}\ \mathbf{rb}\ \mathbf{gb}\ \mathbf{1}] \tag{23}$$

where $\mathbf{1}$ denotes the $N \times 1$ vector with all N components equal to 1 (and accounts for translations). The color correction problem (18) is then rewritten as

$$\mathbf{H} \approx \mathbf{R}'\mathbf{M}' \tag{24}$$

where $\mathbf{M}'$ is now a 10×3 matrix.

In Section 3.2 we showed how a constrained regression could be carried out when p measurements were made of an ensemble set of reflectances. Specifically, we developed methods to deal with $N \times p$ response matrices. Clearly these methods are equally applicable to the polynomial expansion matrix $\mathbf{R}'$ constructed above.

To evaluate the colorimetric performance of the polynomial regression method, we repeated the simulation experiment of Section 3.1. We generated the RGBs of 462 Munsell reflectances and from these the corresponding polynomial expansion. We then found the best least-squares mapping taking the expanded terms to corresponding XYZs. A second, white-point preserving mapping was also calculated. In Table 1 (bottom), we denote by PLS the results for polynomial least-squares regression, and by PW-PPLS the corresponding white-point preserving results. As shown in Table 1, the mean CIELAB error of the least-squares regression is reduced to 1.58 (reduced from 2.10 for linear correction) and for the least-squares regression, when white is preserved, to 1.59 (from 2.37). It is apparent that the white-point preserving regression now delivers comparable correction performance. Notice, however, that the error in mapping white is still relatively large for the unconstrained regression: 1.39 CIELAB units (compared with 1.41 for linear correction).

Table 2 Scanner. Statistics for CIELAB ΔE^*_{ab} values comparing matrix transforms derived from maximum-ignorance LS and maximum-ignorance WPPLS for three data sets: the set of Munsells, the Macbeth ColorChecker, and the Object data set. For three-sensor LS, the white point moves 10.53 units; for WPPLS the white point error is zero. For six-sensor LS, the white point moves 3.97 units.

	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell MaxIgLS	2.70	7.92	8.08	17.36	0%
Munsell MaxIgWPPLS	0.16	4.26	5.49	30.12	35%
Macbeth MaxIgLS	2.84	7.77	8.46	18.30	4%
Macbeth MaxIgWPPLS	0.07	6.08	7.08	23.95	33%
Neutrals MaxIgLS	3.56	7.28	6.77	10.17	0%
Neutrals MaxIgWPPLS	0.07	0.21	0.31	0.99	100%
Object MaxIgLS	2.73	7.68	7.91	23.50	0%
Object MaxIgWPPLS	0.22	4.32	6.02	40.13	28%
Munsell 6-sensor MaxIgLS	1.02	2.91	2.86	6.31	53%
Munsell 6-sensor MaxIgWPPLS	0.04	1.80	2.52	15.97	71%

Since the improvement offered by the polynomial method is quite small we will not consider it further in this paper.

4 Maximum Ignorance Calibration Set

So far we have investigated a matrix transformation derived from a particular data set. However, we may not have knowledge of what colors are being scanned. Therefore it is of value to consider what can be determined by applying knowledge of **M** derived from the sensor curves in Fig. 1 without any reference to the data (cf. Ref. 15). If we simply assume that reflectance data is uniformly distributed in spectrum in the world—a "maximum ignorance" assumption—then the equations derived above still apply except that instead of using data matrices **R** and **H** we substitute the sensor curves $\mathbf{Q}_E$ and $\mathbf{X}_E$ themselves.[9]

4.1 *Maximum Ignorance Least-Squares Regression*

In this case the matrix derived from the sensor set $\mathbf{Q}_E$ alone parallels the LS solution Eq. (8):

$$\mathbf{M}=(\mathbf{Q}_E{}^T\mathbf{Q}_E)^{-1}\mathbf{Q}_E{}^T\mathbf{X}_E\,. \tag{25}$$

Table 2 shows how well this maximum ignorance least-squares method performs when applied to our exemplar data set of Munsell reflectances: not as well as a data-driven method, of course. Notice that the white point shifts by a ΔE^*_{ab} of 10.5 units.

4.2 *Maximum Ignorance Constrained Regression*

The constrained-surface preserving solution Eq. (17) in this maximum ignorance case becomes

$$\mathbf{N}=[\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{Q}_E\mathbf{Z}]^{-1}[\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{X}_E-\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{Q}_E\mathbf{D}]. \tag{26}$$

Table 2 shows that when applied to the Munsell data set, in terms of the CIELAB error when the constraint surface is white, the WPPLS (white point preserving least-squares) method outperforms the LS method. This shows that, for this particular data set, the solution embodied in Eq. (26) is closer to the best solution (17). The mean CIELAB error in this case is 5.49, in the range identified as acceptable by Meyer.[16]

We can picture these results by showing errors in a chromaticity diagram.[1] Three-dimensional points are shown in a plane by forming chromaticity coordinates $x=X/(X+Y+Z)$ and $y=Y/(X+Y+Z)$. Figure 2 shows the 'horseshoe' shaped spectrum locus as well as the convex set of chromaticities for the set of Munsell colors. Lengths of errors are not well indicated as such a diagram—the size of just-noticeable differences under human perception must be displayed using ellipses of sizes that vary throughout the chromaticity diagram, the MacAdam ellipses.[1] The white point is shown as an '*O*' in this figure. Errors are less for colors that are more desaturated ("whiter") and the data set contains many such colors. This effect is just what is needed for producing an image that appears as it should for desaturated colors in terms of subjective color.[17] In particular, consider the commercially important case of "whites." Consider the set of desaturated Munsell colors that are within the first octile of distance on

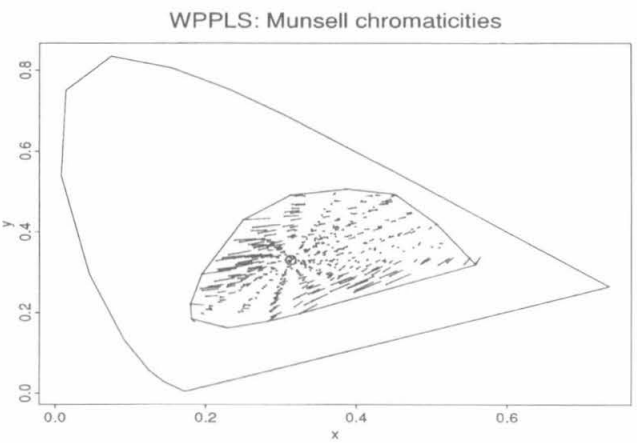


Fig. 2 Scanner. WPPLS regression: errors plotted on chromaticity diagram. The white point is shown by 'O'.

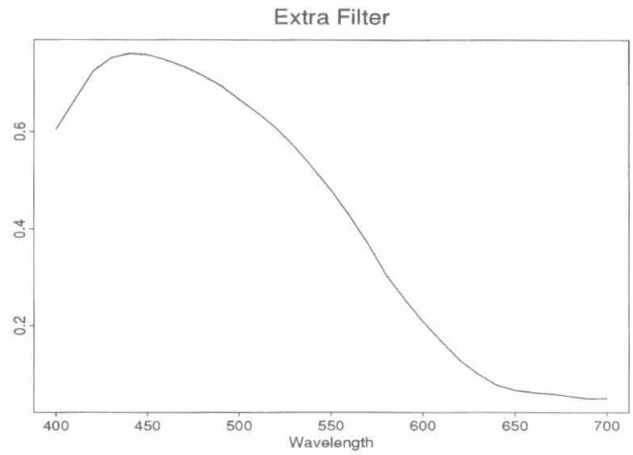


Fig. 3 Extra filter for forming a six-sensor scanner.

the chromaticity diagram from the white point. Hunt (Ref. 17, p. 133) gives a measure of whiteness for colors that are "called white commercially." For the white point itself the index equals 100, and for the desaturated Munsell colors the index is in the range −13 to 123. For the LS fit, the error in whiteness is 35.6 whereas for the WPPLS fit the error is only 3.4. Therefore, as expected, the WPPLS method does much better in reproducing the important class of desaturated colors.

Table 2 also shows how the WPPLS regression performs compared to a LS regression for two additional data sets. The first is the set of paint chips that form the 24 patches on the Macbeth ColorChecker chart;[18] there are six neutral paint chips in this set, drawn from the Munsell set, and the formulation of all patches is similar to that for Munsell chips. One can see that the WPPLS method substantially improves accuracy for the neutrals, as one would expect. Another data set examined in Table 2 is the set of 170 natural objects measured by Vrhel *et al.*[19]

As well, Table 2 shows how error is affected by carrying out a WPPLS regression on XYZ estimates constructed from a six-sensor scanner. Here, we follow Ref. 20 and, after a first image is taken, we overlay our scanner with another filter and produce a second image, so as to effectively produce a six-sensor output. The extra filter used is Wratten filter 38 (Ref. 21, p. 70), shown in Fig. 3. Overall, the error is substantially reduced by applying the WPPLS estimate instead of the LS one: the median error is 62% of that for the LS regression, and there are many more colors with errors under the just-noticeable level. As well, while the WPPLS white point error is zero, that for the LS regression is 3.97 CIELAB units even for this six-sensor data.

In general, we expect the performance of a maximum ignorance assumption to be good when the data set is close to being uniform in expectation value of spectral power distribution, i.e., having covariance matrix equal to the identity. Below, we show how to exploit this relationship between covariance matrix and estimates of matrix transforms.

5 Non-Maximum Ignorance

In Ref. 9, it was pointed out that the LS solution (8) can be written in terms of the non-mean subtracted covariance matrix—the 31×31 matrix of products $\mathbf{K}$—of the set of reflectance data. If $\mathbf{S}$ is an $n\times31$ matrix, where each row represents a given reflectance spectrum, then $\mathbf{K}$ is defined as $\mathbf{S}^T\mathbf{S}$. It is straightforward to show that a least-squares regression between scanner responses and tristimulus values depends only on $\mathbf{K}$. To see this let us rewrite Eq. (8) making the role of the reflectances $\mathbf{S}$ and the effective sensors $\mathbf{Q}_E$ and $\mathbf{X}_E$ explicit:

$$(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{H}=(\mathbf{Q}_E{}^T\mathbf{S}^T\mathbf{S}\mathbf{Q}_E)^{-1}\mathbf{Q}_E{}^T\mathbf{S}^T\mathbf{S}\mathbf{X}_E\,. \tag{27}$$

Substituting $\mathbf{K}$ for $\mathbf{S}^T\mathbf{S}$ we see that

$$(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{H}=(\mathbf{Q}_E{}^T\mathbf{K}\mathbf{Q}_E)^{-1}\mathbf{Q}_E{}^T\mathbf{K}\mathbf{X}_E\,. \tag{28}$$

The importance of the above is that one need only know the matrix $\mathbf{K}$ in order to determine the best LS transform. Turning this idea around, we can say that for any particular LS transform $\mathbf{M}$, any data set will share this same $\mathbf{M}$ provided it has the same product matrix $\mathbf{K}$.

The validity of this idea turns out to persist in the WPPLS calculation as well. For rewriting Eq. (17), we have

$$\mathbf{N}=[\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{K}\mathbf{Q}_E\mathbf{Z}]^{-1}[\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{K}\mathbf{X}_E-\mathbf{Z}^T\mathbf{Q}_E{}^T\mathbf{K}\mathbf{Q}_E\mathbf{D}]. \tag{29}$$

Using this equation, suppose that we did not wish to simply use the information contained in the sensor sets $\mathbf{Q}_E$ and $\mathbf{X}_E$ alone, but wanted to add additional information about the type of reflectance sets likely to arise in imaging situations. Then substituting the matrix $\mathbf{K}$ for such a guiding or training set of reflectances, we arrive at a matrix transform $\mathbf{M}$ that can be applied to the particular image at hand. If the training set is like the actual image, then the matrix transform derived from Eq. (29) should perform well.

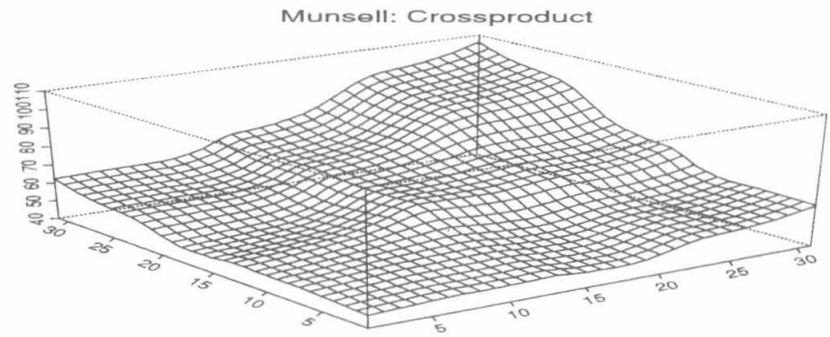
Consider the standard set of 24 reflectances that make up the Macbeth ColorChecker chart.[18] The six neutral patches in this chart are drawn from the Munsell Book of Color; the other patches are new paints. Therefore we expect matrix $\mathbf{K}$ to be similar to but not identical with that for our exemplar data set. This turns out to be the case, as evidenced by the results in Table 3. Here, results of applying the matrix $\mathbf{M}$ derived from the ColorChecker chart to all 462 Munsell cases gives excellent results.

Figure 4 shows why this should be the case: here the product matrix $\mathbf{K}$ for the Munsell set is plotted in Fig. 4(a), while Fig. 4(b) shows that for the ColorChecker chart. Since matrix $\mathbf{K}$ appears in both matrices in Eq. (29), the absolute numbers in $\mathbf{K}$ are not important: just the shape of the plot changes the results. In contrast, the maximum-ignorance approach amounts to using a matrix $\mathbf{K}$ equal to a diagonal matrix plus a uniform plane.

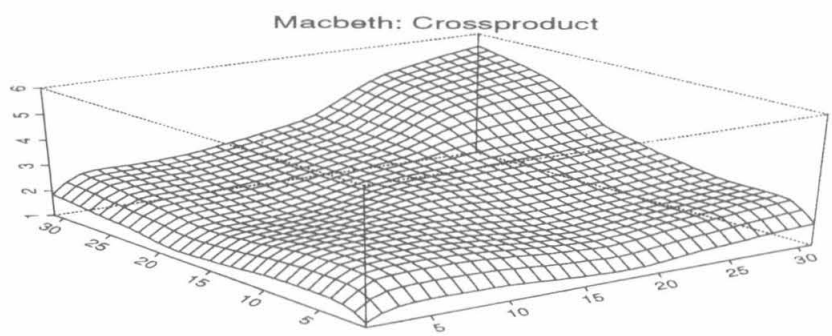
As a more stringent test of the non-maximal ignorance training set idea, consider the plot of Fig. 4(c). Here we show the product matrix $\mathbf{K}$ for the set of 170 natural objects measured by Vrhel *et al.*;[19] these objects were skins, rocks, plants, man-made objects, etc., and should be quite unre-

Table 3 Scanner. Statistics for CIELAB ΔE^*_{ab} values for 462 Munsell samples derived from a non-maximal ignorance WPPLS for two different training sets: the Macbeth ColorChecker and the Object data set.

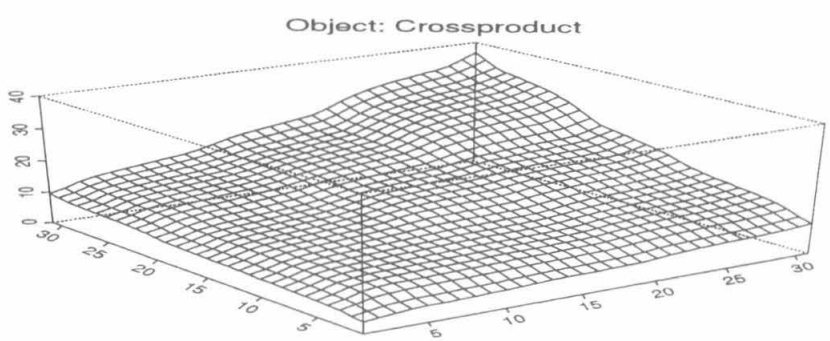
	Wt.Pt.ΔE^*_{ab}	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell from Macbeth WPPLS	0	0.12	1.74	2.57	12.66	68%
Munsell from Object WPPLS	0	0.13	2.42	3.02	13.95	61%



(a)



(b)



(c)

Fig. 4 (a) Crossproduct matrix for 426 Munsell patch reflectances. (b) Crossproduct matrix for 24 Macbeth ColorChecker patches. (c) Crossproduct matrix for 170 object reflectances.

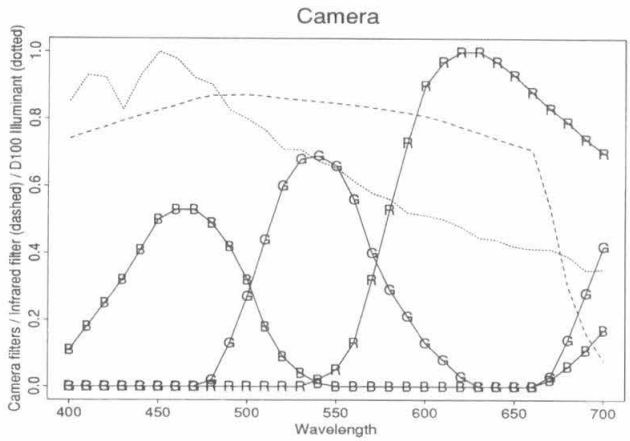


Fig. 5 Camera sensors. Sony DXC151 sensors are shown (solid lines), along with infrared filter (dashed line), and standard illuminant D100 (dotted line).

lated to the Munsell data set. Nevertheless it is surprising how closely Fig. 4(c) resembles Fig. 4(a). The results in Table 3 show that even in this case the non-maximally ignorant WPPLS method performs quite well.

6 Color Camera Experimental Results

In this section, we repeat the calculations above, but apply the WPPLS method not to a scanner but to the case of a color camera, in order to demonstrate that results apply across devices.

For concreteness, let us take a particular camera; we use the sensor curves for a Sony DXC151 camera, shown in Fig. 5. We also multiply by an infrared filter, and take as illuminant the standard daylight D100, simply as a typical example and also so as to not equal D65, in order to test the method. Figure 5 shows the infrared filter and the illuminant as well as the camera sensor curves.

The scanner experiments with results shown in Tables 1, 2, and 3 are repeated in Tables 4, 5, and 6.

As can be seen, the success of the WPPLS method when applied to a color scanner carries over to the case of a color camera. As well as improved color correction, the salient feature of the results is that neutrals are very much better corrected using the WPPLS method than the LS method.

Table 4 Camera. Statistics for CIELAB ΔE^*_{ab} values comparing LS and WPPLS methods for 462 Munsell samples.

	Wt.Pt. $\Delta Eab*$	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell LS	1.16	0.04	1.15	1.78	10.72	82%
Munsell WPPLS	0	0.07	1.40	2.00	10.87	80%

Table 5 Camera. Statistics for CIELAB ΔE^*_{ab} values comparing matrix transforms derived from maximum-ignorance LS and maximum-ignorance WPPLS for three data sets: the set of Munsells, the Macbeth ColorChecker, and the Object data set. For LS, the white point moves 2.92 units; for WPPLS the white point error is zero.

	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell MaxIgLS	0.44	3.40	3.92	13.18	40%
Munsell MaxIgWPPLS	0.15	1.99	2.65	12.35	50%
Macbeth MaxIgLS	2.74	1.03	4.29	12.90	42%
Macbeth MaxIgWPPLS	0.17	3.10	3.45	12.27	46%
Neutrals MaxIgLS	1.14	1.83	1.99	3.32	83%
Neutrals MaxIgWPPLS	0.12	0.35	0.43	1.08	100%
Object MaxIgLS	2.45	3.74	4.37	16.78	33.53%
Object MaxIgWPPLS	0.10	3.08	3.53	16.71	49%

7 Conclusions

This paper examines the problem of mapping scanner (or other device) RGBs to corresponding XYZ tristimulus values. In particular a new constrained least-squares regression method is formulated. The constrained least-squares regression represents a compromise between simple least-squares and Vrhel's principal component method. In particular the constrained least-squares regression allows the specification of k surfaces ($k=0,1,2,3$) for which the mapping from RGB to XYZ must be exact. Particular attention was paid to the case where a single surface ($k=1$) must be mapped without error. Because of the importance of white in color reproduction a white-point preserving least-squares (WPPLS) regression was examined. We showed that a WPPLS transform works well when the statistics of likely surfaces are known in advance (almost as well as least-squares) and performs reasonably even when they are not known. In the latter case performance is superior to simple least-squares. Additional information from the covariance for a known reflectance set can be applied to further improve results. The constrained regressions when $k=0$ and $k=3$ are equivalent to simple least-squares regression and Vrhel's principal component method, respectively. Thus, the current theory subsumes these two color correction methods.

Here, we have emphasized transformations from RGB to XYZ values, with an eye to providing device-independent color descriptors. Nevertheless, the method can also be

Table 6 Camera. Statistics for CIELAB ΔE^*_{ab} values for 462 Munsell samples derived from a non-maximal ignorance WPPLS for two different training sets: the Macbeth ColorChecker and the Object data set.

	Wt.Pt. $\Delta Eab*$	Min	Median	Mean	Max	% Under 3 ΔE^*_{ab}
Munsell from Macbeth WPPLS	0	0.08	1.53	2.17	11.68	77%
Munsell from Object WPPLS	0	0.10	2.00	2.60	12.78	70%

used to transform, in a white-point preserving fashion, between color values imaged under different illuminants.[22] Thus the problem of color appearance can be addressed via the present method.

Future work involves incorporating the idea of spectral sharpening[23,24] into the present methodology. In Ref. 15 it was shown how matrix transforms that 'sharpen'—make more narrow-band—the sets of sensor curves can lead to improved results in a white-point preserving transform. We shall pursue the idea of including spectral sharpening transforms into the present method in order to simplify methods of color appearance correction.

References

1. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*, 2nd ed., Wiley, New York (1982).
2. B. A. Wandell and J. E. Farrell, "Water into wine: converting scanner RGB to tristimulus XYZ," in *Device-Independent Color Imaging and Imaging Systems Integration, Proc. SPIE* **1909**, 92–101 (1993).
3. S. Suzuki, T. Kusunoki, and M. Mori, "Color characteristic design for color scanners," *Applied Optics* **29**, 5187–5192 (1990).
4. P. C. Hung, "Color rendition using three-dimensional interpolation," in *Imaging Applications in the Work World, Proc. SPIE* **900**, 111–115 (1988).
5. P. C. Hung, "Colorimetric calibration for scanners and media," in *Proc. SPIE* **1448**, 164–174 (1991).
6. H. R. Kang, "Color scanner calibration," *J. Imaging Sci. and Tech.* **36**, 162–170 (1992).
7. R. S. Berns and M. J. Shyu, "Colorimetric characterization of a desktop drum scanner using a spectral model," *J. Electronic Imaging*, **4**, 360–372 (1995).
8. M. J. Vrhel and H. J. Trussel, "Color correction using principal components," *Color Research and Application* **17**, 328–338 (1992).
9. M. J. Vrhel and H. J. Trussel, "Optimal scanning filters using spectral reflectance information," in J. P. Allebach and B. E. Rogowitz, eds., *Human Vision, Visual Processing and Digital Display IV, Proc. SPIE* **1913**, 404–412 (1993).
10. S. M. Newhall, D. Nickerson, and D. B. Judd, "Final report of the OSA subcommittee on the spacing of the Munsell colors," *J. Opt. Soc. Am.* **33**, 385–418 (1943).
11. Munsell Color, *Munsell Book of Color*, Macbeth, a division of Kollmorgen Corp. (1979).
12. M. Stokes, M. D. Fairchild, and R. S. Berns, "Precision requirements for digital color reproduction," *ACM Transactions on Graphics* **11**(4), 406–422 (Oct. 1992).
13. G. D. Finlayson and M. S. Drew, "White-point preserving data transforms in color spaces," Technical Report CSS/LCCR TR 94-22, Simon Fraser University, School of Computing Science (1994). Available using ftp://fas.sfu.ca/pub/cs/techreports/1994/CSS-LCCR94-22.ps.Z.
14. M. J. Vrhel, "Mathematical methods of color correction," PhD thesis, North Carolina State University, Department of Electrical and Computer Engineering (1993).
15. M. S. Drew and G. D. Finlayson, "Device-independent color via spectral sharpening," in *2nd Color Imaging Conference: Color, Science, Systems and Applications*, Scottsdale, AZ, Nov. 15–18, 1994, pp. 121–126, Society for Imaging Science & Technology (IS&T)/Society for Information Display (SID) joint conference (1994). To be reprinted in R. Eschbach and K. Braun, Eds., *Recent Progress in Color Science*, IS&T, Springfield, VA (1997).
16. G. W. Meyer, "Wavelength selection for synthetic image generation," *Comp. Vision, Graphics, and Image Proc.* **41**, 57–79 (1988).
17. R. W. G. Hunt, *The Reproduction of Color*, 3rd ed., Wiley, New York (1987).
18. C. S. McCamy, H. Marcus, and J. G. Davidson, "A color-rendition chart," *J. App. Photog. Eng.* **2**, 95–99 (1976).
19. M. J. Vrhel, R. Gershon, and L. S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Research and Application* **19**, 4–9 (1994).
20. J. E. Farrell and B. Wandell, "Color characterization of a scanner," in *IS&T's Tenth International Congress on Advances in Non-Impact Printing Technologies*, IS&T, Springfield, VA (1994).
21. Eastman Kodak Co., *Kodak Filters for Scientific and Technical Uses*, 2nd ed. (1981).
22. M. J. Vrhel and H. J. Trussel, "Physical device illumination color correction," in *Device-Independent Color Imaging and Imaging Systems Integration, Proc. SPIE* **1909**, 84–91 (1993).
23. G. D. Finlayson, M. S. Drew, and B. V. Funt, "Spectral sharpening: sensor transformations for improved color constancy," *J. Opt. Soc. Am. A* **11**(5), 1553–1563 (May 1994).
24. G. D. Finlayson and B. V. Funt, "Coefficient channels for colour constancy," in *Proc. John Dalton Conf. Colour Vision*, C. M. Dickinson et al., Eds., Taylor and Francis (1997).

Graham D. Finlayson was awarded his doctorate at Simon Fraser University, Vancouver, Canada, in July 1995, for research in color constancy. In September of that year he joined the Department of Computer Science at the University of York. There he has worked on a diverse range of color problems, including color correction, color constancy, color object recognition, and color indexing. In October 1997, he joined the Colour Research Group at the University of Derby where he is a Reader in Colour Image Analysis. His research is sponsored by Hewlett-Packard Corporation, EPSRC, and the Nuffield Foundation. He is a member of the British Machine Vision Association and the Colour Group of Great Britain.

Mark S. Drew is a senior lecturer in the School of Computing Science at Simon Fraser University. His undergraduate education was in engineering science and his subsequent education was in theoretical physics. He received the BASc degree from the University of Toronto in 1970, and studied the foundations of quantum mechanics for the MSc degree in 1971 in the mathematics department there. He studied field theory at the University of British Columbia and received the PhD in 1976. Combining work on energy systems and computer applications, he held an Industrial Postdoctoral Fellowship and subsequently an Industrial Research Fellowship in industry until he joined Simon Fraser University in 1982. His interests are in the fields of multimedia, computer vision, computer graphics, computer algorithms for color reproduction, scientific applications of algebraic programming, and the application of Lie groups to the solution of systems of nonlinear partial differential equations. He is the holder of a US patent in digital color processing.