Reducing The Number of Calibration Surfaces

Ali Alsam and Graham Finlayson*
The Norwegian Color Research Laboratory
Dept of Computer Science and Media Technology
Gjøvik University College
*****University of East Anglia UEA***
Dept of Computer Science

Abstract

Calibration charts are used in colour imaging to determine color correction transforms and for spectrally characterising imaging devices. Traditionally, quite complex charts have evolved as it was reasoned that the more reflectances in a chart the more the chart could represent all other reflectances. However, a chart with many reflectances is both expensive, difficult and tedious to use. The difficulty lies in assuming constant lighting conditions over the whole chart and the tedium appears when the chart must be measured using a spectrophotometer. To circumvent these problems researchers have sought methods to find smaller sets of reflectances which, in some sense, represent larger reflectance sets.

In this paper we develop an iterative selection procedure where we select individual reflectances from a colour chart. The first is chosen so it best accounts for the majority of the spectral variance. The next best accounts for the variance that is left. In general the ith selected chart reflectance best accounts for the variance among reflectances (given that $i-1$ reflectances are already selected). We show that this procedure is weakly optimal and as such compares with prior art which chooses reflectances using simple heuristics. The new method is also much faster than algorithms that are built on stronger optimality conditions. Experiments demonstrate that our new method represents a reasonable compromise between fast (and feasible) reflectance selection and the optimality of the chosen set.

Introduction

There are many applications in colour science involving algorithms that require the camera's response to a number of surfaces as their input. For example, calibrating a camera to determine a transform form the camera dependent *rgb* space to a camera independent space such as *xyz* [1], involves measuring its response to a number of surfaces with known spectral properties. The same measurements are required for spectral calibration [2, 3, 4]where the user is interested in estimating the sensor curves. To achieve accurate calibration results, the measurements are taken under laboratory controlled conditions making sure that the variation in the spectral data is minimised. Measuring the spectral data and the corresponding *rgb* responses is a tedious task and human error as well as error introduced by the variation in the lighting conditions are expected, especially, when the number of measurements taken is large. Furthermore, calibration results achievable are determined by the spectral properties of surfaces in a calibration target: a collection of predominately red surfaces would result in a calibration which is biased to correcting red colours but would not be so useful for calibrating blue colours. Hence we might wonder how many surfaces out of those available in a calibration chart are actually needed and which ones would result in a well behaved calibration. Finally, from a theoretical point of view, it is interesting to solve for a reduced set of surfaces that are sufficient to describe the spectral space of a larger set.

The search for a reduced set of surfaces to describe a larger dataset is not new: Hardeberg *et al* [5] introduced an algorithm to select a reduced set of reflectances to be used in scanner calibration. The algorithm [5] is a technique to select reflectances such that they are as different as possible to each other where the difference measure is determined by the condition number of the matrix encapsulating the reduced reflectance data, where the smaller the condition number [6] the more different the surfaces are. Unfortunately, the method is a greedy algorithm [7], which converges to a local rather than a global optimum. More recently, Alsam *et al* [8] proposed a surface-reduction algorithm which defines the most significant surfaces as the extreme points on the convex-hull of the spectral data. The results in [8] show that as many as 80 surfaces are needed to account, fully, for the data of the Munsell book of colors which is composed of 1269 surfaces. Further reduction is achieved by means of a volume projection technique where the optimum set of surfaces is taken as that which encloses the largest volume percentage. Unfortunately, due to the prohibitively large calculation complexity the algorithm is restricted to a maximum data dimensionality of five. Knowing that the dimensionality of some data sets extends to higher dimensional spaces, 3-11, [9]; it is believed that some significant information is likely to be lost when using this method.

In this paper, we present a new iterative method which is, like Principal component analysis, based on calculating the Eigenvectors and Eigenvalues of a data matrix. PCA can be formulated as finding the direction (spectrum) of maximum variance and then calculating the data orthogonal to this direction. Note that the best spectrum need not be in the original dataset. This strategy is iterated (we find the max variance direction of the new orthogonally projected) data and project orthogonal to this direction and so on until we account for all the variance in the data (or reach a suitable stopping criterion). The PCA method is known to be optimal (in terms of accounting for the largest % of variance in the data given a fixed number of k basis vectors). Here, we propose a similar approach, where the first vector, which is drawn from the set of spectra at hand, is the one that, considered as a basis

vector, accounts for greates variance in the data set. Like PCA we now iterate: the data is projected orthogonal the first basis, and then we search for the new max variance direction in the spectra that are left and continue this matrix deflation [10] until we reach a stopping criterion. In this way each basis vector we select is a reflectance spectrum in the original dataset.

This PCA-like algorithm can be directly applied to any spetral data set. Alternatively, we can transform the spectral data to a conical space as proposed by Reiner Lenz [11] and then apply the technique. This second technique has the advantage that t is known that the transformation of the spectra into a conical space results in a separation between intensity and hue. Thus, when reducing the data in a space we tend to recover get spectra that represent the dominant hues: red, green and blue etc.

Background

The problem which we are trying to solve is reducing the number of surfaces, c, contained in an $n \times m$ data matrix C where n is the dimensionality of the data governed by the sampling rate, normally 31, and m is the number of surfaces.

The Most Significant Reflectances

In [5] Hardeberg *et al* introduced an algorithm to select a reduced set of reflectances to be used in scanner calibration. The motivation behind the authors' work on selecting a subset of reflectances was the need for a very large set of spectrally different reflectances to facilitate an adequate sensor recovery in conjunction with the truncated singular value decomposition method [2]. More formally,

1. Starting from the full set of colour signals C we choose the one with maximum norm:

$$\|c_1\| \geq \|c_i\| \qquad for\ i = 1...m \tag{1}$$

   where c_1 is the first colour signal vector in the reduced set. Typically, this condition returns a white surface.
2. The second reflectance is selected by evaluating the condition number of the two column matrix, with the first column being the first chosen vector and the second vector is any other vector from the set. This is an iterative procedure which enables us to find the vector which will result in the minimum condition number for the matrix. We seek to find c_2 such that:

$$\frac{\sigma_{\max}\left([c_1 c_2]\right)}{\sigma_{\min}\left([c_1 c_2]\right)} \leq \frac{\sigma_{\max}\left([c_1 c_i]\right)}{\sigma_{\min}\left([c_1 c_i]\right)} \qquad for\ i = 1...m \tag{2}$$

   where σ_{max} denotes the maximum singular value of the matrix $c_1 c_i$ and σ_{min} denotes the minimum nonzero singular value.
3. the next colour signal is chosen in exactly the same fashion:

$$\frac{\sigma_{\max}\left([c_1 c_2 ... c_i]\right)}{\sigma_{\min}\left([c_1 c_2 ... c_i]\right)} \leq \frac{\sigma_{\max}\left([c_1 c_2 ... c_{i-1} c_i]\right)}{\sigma_{\min}\left([c_1 c_2 ... c_{i-1} c_i]\right)}$$
$$for\ i = 1...m, i \notin \{c_1, c_2 ... c_{i-1}\} \tag{3}$$

The basic problem with the Hardeberg method is that it is a greedy algorithm (it chooses the best spectrum using a simple criterion given the choices already made and the data that is left). Unfortunately, unlike some greedy procedures, the Hardeberg method does not converge the globally best answer.

However, finding the best global selection of surfaces is not easy. Let us assume that the number of reflectances needed to perform the calibration is 7 out of the 24 patches of the Macbeth colour checker. Then the number of possible combinations is:

$$N_z^y = \frac{y!}{z!(y-z)!} = \frac{24!}{7!(24-7)!} = 346104 \tag{4}$$

where z is the number of surfaces in the subset while y is the total number. Hardeberg's algorithm clearly does not explore all these combinations.

Figure (1) shows the first three reflectances achieved by Hardeberg's algorithm (the spectral data of the Macbeth Color Checker).

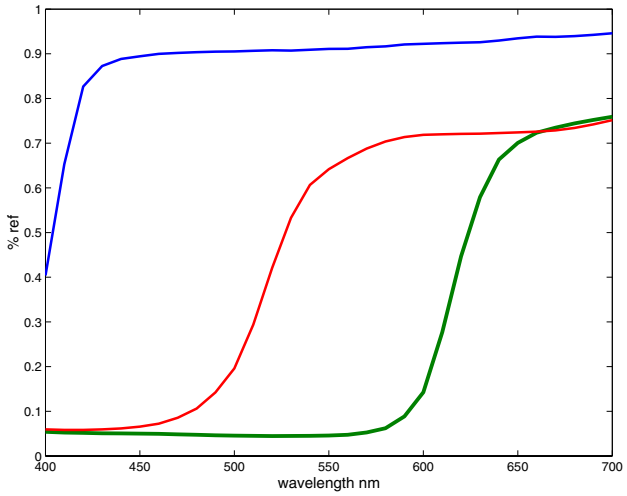

Figure 1. *The first three reflectances achieved by Hardeberg's algorithm based on the spectral data of the Macbeth Color Checker.*

Convex Reduction

In [8] Alsam *et al* proposed a geometric solution to the surface reduction problem. The fundamental idea is that a set of reflectances occupy a subspace in an n-dimensional hyper-cube. Moreover, the space occupied by the reflectances is defined by a number of extreme points, which are themselves natural surfaces. Hence Alsam proposed that it is sufficient to describe an arbitrarily large set of reflectances using those surfaces which lie on the convexhull of the spectral gamut. Further, the authors argue that the most significant surfaces are those which enclose the maximum possible volume in the spectral space. Choosing the volume as the optimality criterion is in keeping with [12], where the prime colors are defined as those, which enclose the maximum volume in the chromaticity space. Moreover, the volume defined by a set of points is directly associated with determinant of the resultant matrix; and it is well established that the stability of matrix inversion is directly related to the determinant of the matrix: where the

greater the determinant the more stable the inverse. Though, maximising this determinat is a computationally laborious: we arrive at a more rigorous algorithm for spectral selection (in comparison with the Hardeberg method) but at a much higher computational cost.

The main criterion which a reduced set needs to satisfy is that any other reflectance in the global set should be a convex combination of the reduced set, i.e. a reflectance c

$$c = \sum_{i=1}^{k} \lambda_i c_i^e \ : \ \sum_{i=1}^{k} \lambda_i = 1, \text{ and } \lambda_i \geq 0 \ \forall i \tag{5}$$

where c_i^e is the ith extreme point in C.

To satisfy Equation (5) ,exactly, for any reflectance c the reflectances c_i^e in the reduced set have to be the extreme points defining the convexhull of the spectral data C in the spectral space. Unfortunately, the complexity of calculating the convexhull of a 31 dimensional data is forbiddingly, high [13]. Thus to proceed the authors make use of linear models where, the spectral data in C is approximated using 3-5 basis vectors.

In Figure (2), we plotted the first three most significant surfaces arrived at by using Alsam's algorithm, here the calculations were based on the 24 surfaces of the Macbeth Color Checker. As we see the most significant surfaces include a white and a black.

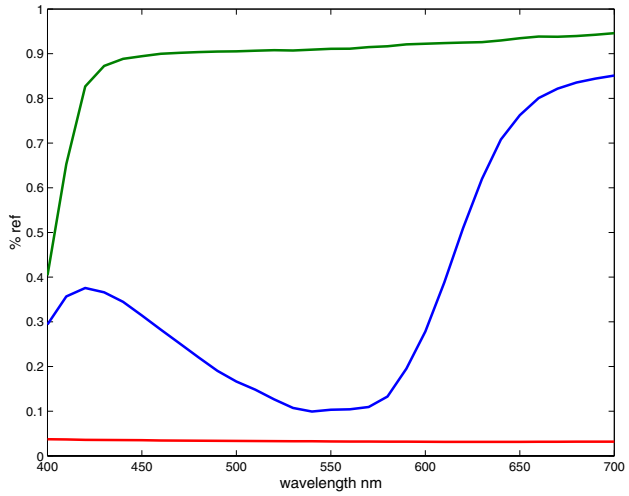

Figure 2. *The most significant surfaces arrived at by using Alsam's algorithm, here the calculations were based on the 24 surfaces of the Macbeth Color Checker.*

Reducing Calibration Surfaces by Matrix Projections

The algorithm proposed in this paper is similar to the standard algorithm used to calculate the eigenvectors of a data matrix. We start by normalizing each colour signal, c', by its norm, i,e,:

$$c = \frac{c'}{||c'||} \tag{6}$$

where $||c'||$ is the second norm of c'. The result, for all colour signals, is an $n \times m$ matrix C whose columns have norm 1. Having done that we calculate, for each vector, c a score value s_i defined as:

$$s_i^{(it)} = ||[c_i^{(it)}]^T C^{(it)}|| \text{ for } i = 1...m \tag{7}$$

where T is the vector transpose operator. We use the superscript (it) to denote the itth iteration (in terms of the narrative thus far $it = 1$). Based on the score value $s_i^{(it)}$ obtained for each vector, we chose the first vector in the set as the one corresponding to the highest score.

$$k = \max_i s_i^{(it)}$$
$$v^{(it)} = c_k^{(it)}$$

where $it = 1$ when the 1st vector is chosen. The motivation behind this choice is that the vector which results in the highest score is the one which is most correlated, i.e. representative of the data set.

To find the second most representative vector we define a projection operator onto the null space of the first vector, $v^{(it)}$:

$$P^{(it)} = I - v^{(it)}[v^{(it)}]^T \tag{8}$$

Using the projection matrix $P^{(1)}$, we project the data in C onto the null space of $v^{(1)}$:

$$C^{(it+1)\perp} = P^{(it)} C^{(it)} \tag{9}$$

Clearly, starting with $it = 1$, we can calculate the first reflectance spectrum $v^{(1)}$. We then increment it to equal 2 then 3 then 4 and so on and use the procedure (7) through (9) above to calculate the basis functions $v^{(2)}$, $v^{(3)}$, $v^{(4)}$ and so on.

Interestingly, this simple procedure can be shown to equal Principal component analysis given a dataset that comprises a closed convex set of spectra i.e. if we allow all convex combinations of our input spectra then PCA and the above procedure (operating on these combinations) will return the same basis vectors. In this respect we say that our selection procedure is *weakly optimal*.

The Conical Intensity and Chromaticity

As mentioned in the introduction section, in this paper, we tested the influence of transforming the spectral data into the conical space proposed by Reiner Lenz [11]. Thus in this section, we offer a brief description of this transformation. We start with a general linear model where any spectrum is written as a linear sum of a set of basis vectors:

$$c \approx \sum_{i=1}^{n} \sigma_i b_i \tag{10}$$

It is known that: when the basis functions b_i form an orthonormal system, then the expansion coefficients σ_i are given by the scalar products $\langle c, b_i \rangle$ in the Hilbert space. Furthermore, in Principle Component Analysis PCA, the basis vectors b_i are as the eigenvectors of the $n \times n$ correlation matrix encompassing the spectral data, i.e. CC^T.

To transform the spectral data into a conical space of intensity and chromaticity, Lenz noted that the first eigenvector b_i is equal to the mean and positive everywhere. He then showed that the first eigenvalue σ_1 is a measure of the intensity of the spectral

distribution. Thus to transform the spectra to an intensity independent space he defined a non linear transformation where all the eigenvalues where normalized by the first, i.e.:

$$\left(\frac{\sigma_2}{\sigma_1},...,\frac{\sigma_n}{\sigma_1}\right) \tag{11}$$

Given this transformation and Equation (10) a colour signal in the chromaticity space can be defined as:

$$\tilde{c} = \frac{\sigma_2}{\sigma_1}b_2 + \frac{\sigma_3}{\sigma_1}b_3 + ... + \frac{\sigma_n}{\sigma_1}b_n \tag{12}$$

In this paper we are interested in using this chromaticity spectral representation as input to our new selection procedure.

Experiments and Results

To test the new algorithm and the influence of the conical transformation on surface selection, we indexed the 24 reflectances of the Macbeth Color checker in order of significance. To compare the results we performed the selection using the new algorithm first in the spectral space and then in the conical space. We also indexed the surfaces using Hardeberg's algorithm described in the background section.

The results of the indexing experiment are shown in Figure (3) for the original reflectances and (4) for those obtained in the conical space. As we note from the two figures, transforming the spectra to the conical space as a pre-indexing step results in surfaces which correspond to the dominant colours, i.e. red, green, blue, purple and white.

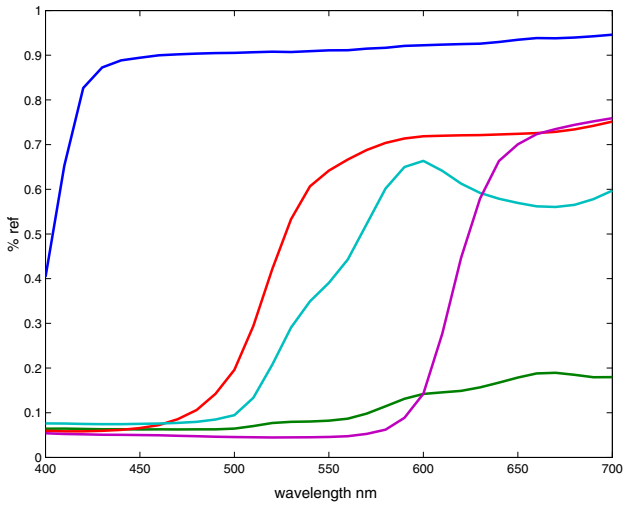

Figure 3. *The most significant 5 surfaces arrived at by using the new proposed algorithm, here the calculations were based on the 24 surfaces of the Macbeth Color Checker.*

The evaluation of the usefulness of the selection and the comparison of the reduced set to the global (that from the whole chart), is not trivial and depends on the application considered, i.e if the surfaces selected are to be used for spectral calibration, colorimetric characterisation or reflectance recovery. In this paper, we chose a simple experiment which involved solving for the optimal 3×3 transformation matrix, in the least squares sense,

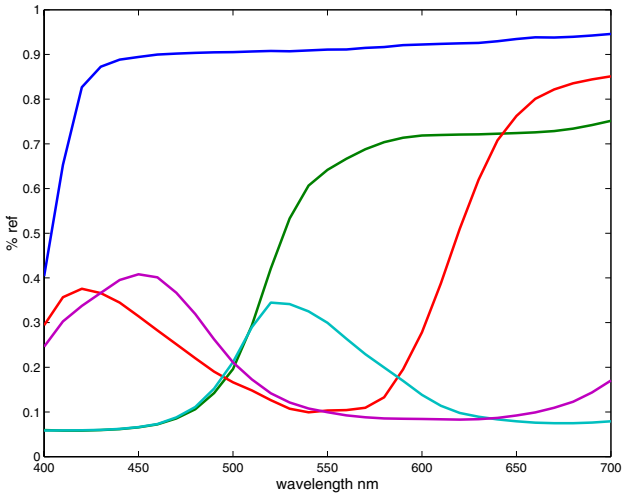

Figure 4. *The most significant 5 surfaces arrived at by using the new proposed algorithm in the conical space, here the calculations were based on the 24 surfaces of the Macbeth Color Checker.*

to transform the rgb data to CIEXYZ. Knowing that a minimum of three surfaces are needed, we used an increasing number of surfaces, as indexed by the algorithms, to solve for the transformation matrices. The number was increased from 3-24 which is the total number available from the chart. Using each matrix we transformed the rgb data for the whole set to the corresponding CIEXYZ values. Finally, we transformed both the measured and estimated XYZ values to the perceptually uniform LAB space and calculated the difference between the two sets. These calculations were based on the assumption of equi-energy illumination spectra; and were carried out for the Sony DX. The spectral sensitivities of this camera are shown in Figures (5). Finally, the results of fitting the data in the Lab space based on the rgb data are reported in Table (1). We report the minimum mean and median values of the Lab differences as well as the number of surfaces which were required to achieve those values. From Table (1), we note that the selection algorithm performed best when the indexing was carried out in the conical space where we find that as few as 4 and 3 surfaces were sufficient to result in the minimum mean and median differences. Further, the new algorithm resulted in the lowest median error, however, 9 surfaces were required. Here, we have to note that the relation between CIEXYZ and Lab is none linear and that the errors depend on the relation between the camera sensitivities and the color matching functions [14].

References

[1] H. R. Kang. Color scanner calibration. *Journal of Imaging Science and Technology*, 36(2), 1992.

[2] G. Sharma and H. Trussell. Characterization of scanner sensitivity. *IS&T/SID Color Imaging Conference*, pages 103–107, 1993.

[3] G. Finlayson, S. Hordley, and P. Hubel. Recovering device sensitivities with quadratic programming. *IS&T/SID Sixth Color Imaging Conference: Color Science, Systems and Applications*, 1998.

[4] A. Alsam and G. D. Finlayson. Recovering spectral sensitiv-

The results of fitting the data in the Lab space based on the rgb data.

Camera	MegaVis	no
PA		
mean ΔE	1.855	6
median ΔE	1.381	9
PA in conical space		
mean ΔE	1.804	3
median ΔE	1.494	4
HardA		
mean ΔE	1.848	6
median ΔE	1.526	4

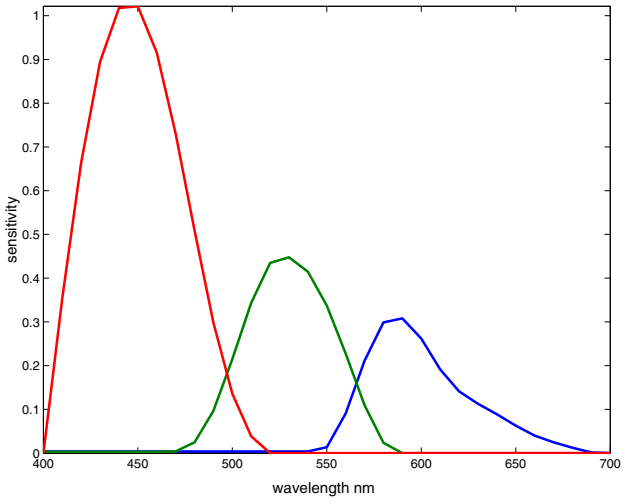

Figure 5. *The spectral sensitivities of a Sony DX camera.*

ities with uncertainty. *In The First European Conference on Color in Graphics, Imaging and Vision*, pages 22–26, 2002.

[5] Jon Yngve Hardeberg, Hans Brettel, and Francis Schmitt. Spectral characterisation of electronic cameras. In *Electronic Imaging*, volume 3409 of *SPIE Proceedings*, pages 100–109, 1998.

[6] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, 1989.

[7] T.H Cormen, C.E. Leiserson, and R.L. Rivest. *Introduction to Algorithms*. MIT Press, 1990.

[8] Ali Alsam and Jon Y. Hardeberg. Convex reduction of calibration charts. *Electronic Imaging Conference San Jose, California*, pages 28–46, 2005.

[9] J. Parkkinen, J. Hallikainen, and T. Jaaskelainen. Characteristic spectra of munsell colors. *Journal of Optical Society of America, 6(2)*, pages 318–322, 1989.

[10] Erwin Kreyszig. *Advanced Engineering Mathematics, 8th Edition*. Wiley, 1998.

[11] Reiner Lenz. Spaces of spectral distributions and their natural geometry. *In The First European Conference on Color in Graphics, Imaging and Vision*, pages 249–254, 2002.

[12] M. H. Brill, G. Finlayson, P. Hubel, and W. A. Thornton. Prime colors and color imaging. *Sixth SID/IS&T Color Imaging Conference, Scottsdale, AZ, 17-20 November*, pages 33–42, 1998.

[13] Joseph O'Rourke. *Computational Geometry in C*. Cambridge University Press, 2 edition, 1998.

[14] P. L. Vora and H. J. Trussell. Measure of goodness of a set of color scanning filters. *Journal of the Optical Society of America A*, 10(7):1499–1508, 1993.