New Sample Election Method for Multispectral System Color Calibration

Miguel A. LÓPEZ-ÁLVAREZ*, Javier HERNÁNDEZ-ANDRÉS†, and Javier ROMERO†
* Hewlett-Packard Spain, Large Format Printing Division, Sant Cugat del Vallès, Barcelona
† Dpto. de Óptica, Universidad de Granada, Campus Fuentenueva s/n 18071, Granada
SPAIN

ABSTRACT

In this work we are dealing with one important issue in the optimization of multispectral systems [López-Álvarez 2005, Hardeberg 1999], which is the selection of a small but representative set of measurements -from a bigger available set- that permit us to appropriately train the system [Schettini 2002, Mohammadi 2004]. By training the multispectral system we relate the spectra belonging to the training set to their sensors' responses by some statistical mean, and this permit us then to obtain accurate spectral estimations from new data obtained by the multispectral camera and applying some spectral estimation method [López-Álvarez 2007]. We compare different previously proposed methods for the selection of training targets [Hardeberg 1999, Schettini 2002, Mohammadi 2004] with a new one proposed by the authors that has demonstrated to be fast and that gives good results. Finally, we also make a small study on the influence of the training set size on the results concerning spectral estimation.

1. INTRODUCTION

One of the critical points in multispectral image acquisition systems is the training process [López-Álvarez 2005]. In order to obtain accurate reconstructed (or estimated) spectral curves with a multispectral system it is essential to know in advance what are the sensors' responses for a certain training set of spectra [Hardeberg 1999]. Usually it is advisable to have a training set with a large number of spectra and with a high spectral gamut. However the computation time (proportional to the number of spectra) must be also taken into account. Therefore it is highly important to select the optimum training set of spectra (with a minimum number of spectra but with a maximum spectral gamut).

Different methods or algorithms have been proposed to select the most appropriate training set from a bigger dataset available. Harderberg [Hardeberg 1999] recommended an iterative method where a spectrum is added to the training set to minimize the ratio between the first and the last singular values (the condition number). Other authors [Schettini 2002] use the values of the cameras responses and then they select the spectra that maximize the average distance with the rest of camera responses values. There are other alternative methods [Schettini 2002, Mohammadi 2004, Lathi 1989] based on the maximum spectral differences between the training spectra using a certain metric (i.e. entropy [Lathi 1989], etc.) or on using Principal Component Analysis (PCA) to select the directions of maximum spectral variation.

2. METHOD

Here we compare some of the previous methods and we propose a new algorithm that is based on grouping [Mohammadi 2004] those spectra that are closer when we measure their respective distances using a colorimetric and spectral combined metric CSCM, that was proposed by the authors in recent papers [López-Álvarez 2005, 2007]. This combined metric is defined as

$$CSCM = Ln(1+1000(1-GFC)) + \Delta E^{*}{}_{ab} + IIE(\%) \qquad (1)$$

Where Ln is the natural logarithm, GFC stands for the goodness-fit-coefficient [Hernández-Andrés 2001] (a spectral metric with a value of 1 for perfect matches), $\Delta E^{*}{}_{ab}$ is the CIELAB distance and $IIE(\%)$ is the percentage of the integrated radiance error [Michalski 1985], a widely used metric in the field of solar radiation that takes into account differences in the total energy across the visible spectrum. The CSCM metric has proved to be a good candidate for evaluating mismatches between spectra, taking into account three different points of view - spectral, colorimetric and radiometric- at the same time. Hence, we will also use this metric to compare the quality of the spectral estimations obtained when using the training set from each of the studied methods.

Our grouping method takes one random spectrum from the global set of measurements, and measures its distance -in terms of the CSCM metric- to the rest of the spectra in the set. The nearest spectrum to the selected one is grouped to it and deleted from the training set. We then randomly choose another initial spectrum and repeat the grouping process (removing one spectrum at each step) until the desired number of spectra remains in the training set. The procedure is illustrated in figure 1. This algorithm is fast since we do not need to measure all the distances between the spectra since some of them are calculated in previous iterations and we can reuse them. We also noticed that, although the concrete set of training spectra selected each time that the algorithm runs is different (due to the random selection of the initial spectrum at each iteration), the quality of the training sets (measured by the quality of the spectral reconstructions obtained when using these sets) is almost the same in all the cases, hence proving the stability –in terms of spectral quality- of our training set selection algorithm.

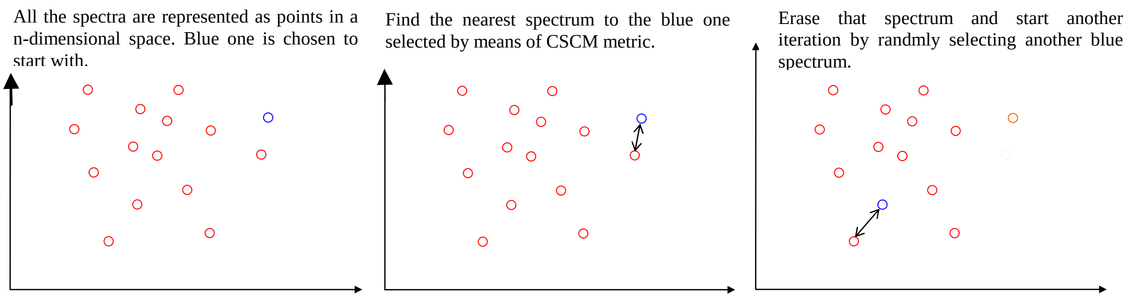

Figure 1. Steps of the sample election method.

As we said before, we have tested three methods proposed by other authors and our new method with 900 spectra of skylight measured in Granada (Spain) during six months, for different weather conditions, different view elevations angles and different solar elevation angles. Simultaneously with the spectroradiometric measurements a trichromatic image was obtained from the same area of the sky using a digital CCD 12-bit camera, relating skylight spectra with camera's responses. The different selection methods were tested to obtain a training set of just 40 spectra, a number that some authors have shown to be an adequate number to obtain accurate spectral reconstructions using the Linear Pseudoinverse method [López-Álvarez 2007]. This method (see eq. (2)) is based on the pseudoinversion [Hardeberg 1999] of the matrix containing the sensors' responses of the training set, ρ_{ts}, and multiplying it by the matrix containing the set of training spectra E_{ts}, in order to obtain a matrix W_L that permits us later to obtain spectral estimations, E_R, from new measurements sensors' responses ρ (see eq. (3)).

$$W_L = E_{ts}\rho_{ts}^{+} \quad (2)$$

$$E_R = W_L\rho \quad (3)$$

3. RESULTS

The quality of each training set obtained was evaluated using the CSCM metric when the complete set of 900 spectra was reconstructed using eq. (3) with a W_L matrix calculated from the training set obtained with each of the training selection methods. Table 1 shows the average values and standard deviations of these reconstructions. Clearly our new method improves the other three methods studied here and could be used in the future in order to obtain the optimum set of training spectra from a large dataset. The set found by using this method tries to maximize the spectral and colorimetric differences between the spectra chosen, assuring a high quality of the selection.

	Condition number	**CSCM maximization**	**PCA**	**Entropy**
CSCM	38.20±32.97	12.22±18.33	19.13±20.89	85.83±95.21

Table 1. Average (± standard deviation) of CSCM metric when the 900 skylight spectra are reconstructed using the Linear Pseudoinverse method [López-Álvarez 2007] and using as training set 40 spectra selected using tour different methods.

In figure 2 we show the CIE31 chromaticity diagram with the chromaticity coordinates of each of the 900 spectra of the complete set in red. In green we show the 40 training spectra selected with our method. We can see how the training spectra selected are placed among the whole spectral range covered by the original complete set, hence proving that the chromatic variance of the training set is as big as that of the original set.

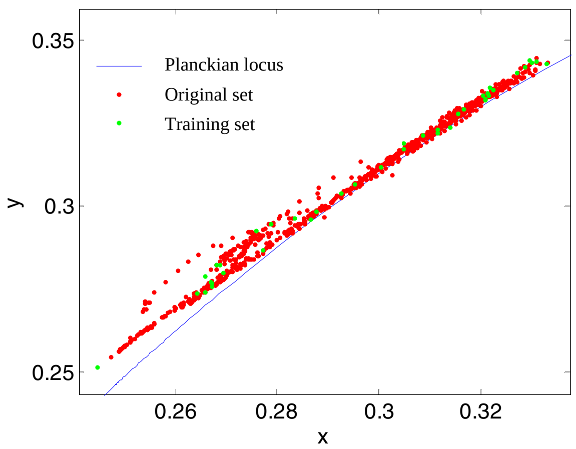

Figure 2. CIE31 chromaticity coordinates of the complete set (in red) and the selected training set (green) of skylight spectra. We also show the Planckian locus in blue line.

In a second round of experiments we tested the influence of the training set size, *m*, in our proposed method for the selection of samples. We used our algorithm to select training sets of various sizes, and calculated their quality by -again- recovering the complete set of 900 skylight spectra bu using the Linear Pseudoinverse method. In figure 3 we show how using more than 100 samples results in no benefit in spectral estimation. Hence, we proposed to use this number as an upper limit for training set sizes in skylight multispectral systems. For different systems regarding different kinds of spectra this value should be different, but must be studied as well.

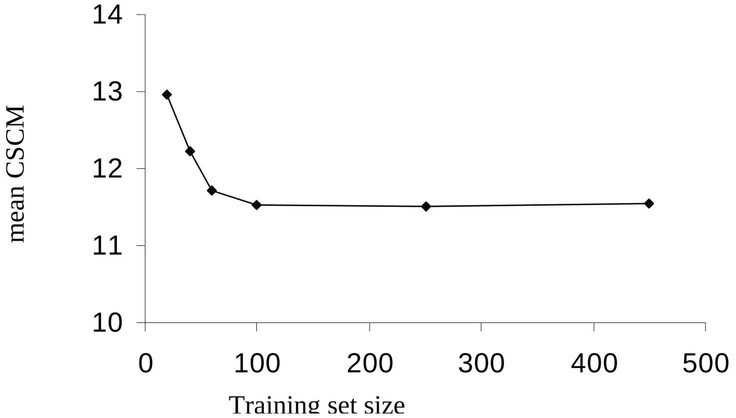

Figure 3. Mean CSCM across the compelte set of 900 spectra when recovered using the Linear Pseudoinverse method trained with training sets of different sizes m.

REFERENCES

Hardeberg J. Y., *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral approaches*, 2001, (Dissertation.com, Parkland, Fla.,). Revised second edition of Ph.D. dissertation, Ecole Nationale Supérieure des Télécommunications, Paris. 134-140.

Hernández-Andrés J., Romero J. y Lee R. L. Jr., "Colorimetric and spectrorradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain", J. Opt. Soc. Am. A, **18**, (2001) 412-420.

Lathi B. P., *Modern digital and analog communication systems*, Second ed. (Oxford University Press, 1989).

López-Álvarez M. A., Hernández-Andrés J., Romero J. y Lee Jr. R. L., "Designing a practical system for spectral recovery of skylight", Appl. Opt., 2005, **44**, 5688-5695.

López-Álvarez M. A., Hernández-Andrés J., Valero E. M. y Romero J., "Selecting algorithms, sensors, and linear bases for optimum spectral recovery of skylight", J. Opt. Soc. Am. A, 2007, **24**, 942-956.

Michalsky J. J., "Estimation of continuous solar spectral distributions from discrete filter measurements: II. A demonstration of practicability," Sol. Energy **34**, 439-445 (1985).

Mohammadi M., Nezamabadi M., Berns R. S., Taplin L. A., "Spectral imaging target development based on hierarchical cluster analysis", Proc. of 12th Color Imaging Conference: Color Science and Engineering, Systems, Technologies and Applications, IS&T, Springfield, 2004, pp. 59-64, (www.art-si.org).

Schettini R., Novati G. Y Pellegri P., "Training set and filters selection for the efficient use of multispectral acquisition systems", Proc. of the 2nd European Conference on Colour Graphics, Image and Vision (IS&T, Springfield, Va., 2002), 422-426.

emails: migangel(a)ugr.es, javierha(a)ugr.es or jromero(a)ugr.es