Influence of the Size of the Training Set on Colour Measurements Performed using a Multispectral Imaging System

de Lasarte M., Pujol J., Arjona M., Vilaseca M; Centre for Sensors, Instruments and Systems Development (CD6), Department of Optics and Optometry, Technical University of Catalonia (UPC); Rambla Sant Nebridi 10, 08222 Terrassa, Barcelona (SPAIN)

Abstract

The accuracy of multispectral imaging systems for recovering reflectance spectra and/or measuring colour has been widely studied and a great deal of research work has been carried out to find the best set of filters and characterization methods. Less work has been carried out on the training set of colour patches used to characterize the system.

The main aim of this work is analyzing the influence of the size of the training set on colour measurements performed using an imaging system based on a CCD camera. In order to perform this study, training sets of different size were obtained using an easy selection method, based on differences in a and b* CIELAB coordinates between the selected colour samples from an initial training set.*

Two configuration of the imaging system have been used: a colorimetric configuration with 3 acquisition channels, and a multispectral configuration with 7 acquisition channels. Very similar results were obtained for both configurations, and it has been proved that the improvement in accuracy of colour measurements by increasing the size of the training set is limited, probably due to the fact that the expected improvement due to the increase of information is made up by the increase of errors accumulated in adjustment. Furthermore, a proper selection, uniform over the colour gamut available, of a low number of samples (40) allows to characterize the imaging system as well as a greater set of samples, for instance, the whole useful set of colour patches (166) of the GretagMacbeth ColorChecker DC chart.

Introduction

The accuracy of multispectral imaging systems for recovering reflectance spectra and /or measuring colour has been widely studied and a great deal of research work has been carried out to find the best set of filters and characterization methods. Less work has been carried out on the training set of colour patches used to characterize the system.

Some training set selection methods have been proposed, considered and compared by different authors [1,2,3]. For all of these methods the finally selected training set pretend to have relatively low numerosity and broad applicability. Hardeberg *et al.* [1] proposed a method for the selection of the most significant reflectance samples for the estimation of the system's spectral sensitivity. Cheung *et al.* [2] proposed two methods whose selection criteria was based on colorimetric metrics to determine how 'different' samples are from each other, rather than spectral metrics used by Hardeberg. Pellegri *et al.* [3] also proposed and compared two selection methods: one based on colorimetric considerations, and other one based on algebraic and geometrical facts.

The main aim of this work is analyzing the influence of the size of the training set on colour measurements performed using an imaging system based on a CCD camera. In order to perform this study, training sets of different size were obtained using an easy selection method, based on differences in a* and b* CIELAB coordinates (directly related to CIELAB colour differences) between the selected colour samples from an initial training set. It was previously proved that reliable and consistent results were obtained using this selection method.

In previous works, for both the colorimetric configuration (having 3 acquisition channels) and the multi-spectral configuration (having 7 acquisition channels) of the imaging system, slightly better results were obtained using the GretagMacbeth ColorChecker DC chart (CCDC, with 166 useful colour patches) as training set than those obtained using the GretagMacbeth ColorChecker Color Rendition chart (CCCR, with 24 colour patches) [4]. On the other hand, increasing the size of the CCDC chart training set by means of a set of 161 Munsell colour samples, selected in terms of their a* and b* CIELAB coordinates to complement the CCDC chart colour patches, did not improve the accuracy of colour measurements and very similar results were obtained for both training sets, without a lost of performance [5]. These results pointed out that there probably exists a limit in the improvement in accuracy of colour measurements with the size of the training set.

Material and Method

The imaging systems used in this work was constituted by a 12 bits cooled monochrome CCD camera (QImaging QICAM Fast1394 12 bit cooled), an objective lens (Nikon AF Nikkor 28 – 105 mm) and two sets of filters: an RGB liquid crystal tunable filter, and a set of seven interferential filters with a FWHM of approximately 40nm, covering the whole visible range of the spectrum, and fitted in a motorized filter wheel. Two different configurations of the imaging system were obtained by placing each set of filters between the CCD camera and the objective lens: a colorimetric configuration with 3 acquisition channels, by using the RGB liquid crystal tunable filter, and a multi-spectral configuration with 7 acquisition channels, by using the set of seven interferential filters fitted in a motorized filter wheel. Both configurations were assessed.

With respect to colour measurements, the predicted XYZ tristimulus values were calculated from the output digital levels of the imaging system (for the colorimetric and the multispectral configurations) applying a direct transformation that relates both sets of values: the XYZ values were directly related to digital levels by a transformation matrix, which was calculated applying the Moore-Penrose pseudo-inverse technique (PSE).

Colour samples were imaged and measured, placed into a special light booth (63cm x 64cm x 52cm) with six incandescent lamps (MAZDA 22c 40W 230V Softone), which provided a uniform illumination field over them. A big window on the opposite side of the booth allowed the measurement of the patches with both the CCD camera, that provided the digital signals, and a spectra colorimeter (PhotoResearch PR-650 with the MS-75 objective lens) that provided the measured XYZ tristimulus values and the radiance spectra. Accuracy of colour measurements was assessed in terms of mean, minimum and maximum CIELAB colour differences between the calculated and the measured XYZ values for the test set.

Training sets of different size were selected among the colour samples of the GretagMacbeth ColorChecker DC chart (CCDC, with 166 useful colour patches). For each training set, the GretagMacbeth ColorChecker Color Rendition chart (CCCR, with 24 colour patches) was used as test set. The selection method applied was based on differences in a* and b* CIELAB coordinates between each pair of colour samples in the selected set. Starting from a first selected sample, samples were selected one by one, and each pair of selected samples must satisfy that $\Delta a^* \geq inca$ and $\Delta b^* \geq incb$. The variables *inca* and *incb* were chosen so that *inca* = *incb*. Different values for these variables allow to fix the number of colour samples selected. In order to determine the dependence of the system's performance on the colour samples of the training set, two selections of the same number of samples were carried out, starting from two different first selected samples. In the first selection, the first useful sample of the CCDC (B2) was selected as the fixed initial sample. In the second selection, the initial sample was selected randomly, and 5 training sets of each size were considered in order to analyze the influence of the colour samples of the training set on the system's performance, depending on the size of the training set.

Results

The size of the training sets selected from the CCDC chart for both selections performed and the corresponding values for the variables *inca* and *incb* are shown in Table 1.

Table 1. Values for the inca and incb variables used in selection of training sets, and number of colour samples (size) of each selected training set

inca = incb	*# colour samples*	*inca = incb*	*# colour samples*
17.00	10	2.355	90
9.20	20	2.15	100
6.98	30	1.69	109
5.35	40	1.45	120
4.50	50	1.22	130
4.00	60	0.80	140
3.12	70	0.58	150
2.68	80	0.20	161
		0.05	166

Colorimetric Configuration

For the colorimetric configuration and the first selection of the training set carried out (fixed initial sample), an outstanding improvement in accuracy of colour measurements is initially observed when the size of the training set is increased up to approximately 40 samples. From 40 samples on the improvement in accuracy when increasing the size of the training set stops being noticeable and accuracy in colour measurements is kept approximately constant (Figure 1.).

Considering the results for the second selection (first sample selected randomly and 5 training set selections of each size), it can be seen that the dependence of the system's performance on the colour samples of the training set (fluctuation of results depending on the training set used) decreases outstandingly when the size of the training set increases (Figure 2., Table 2.).

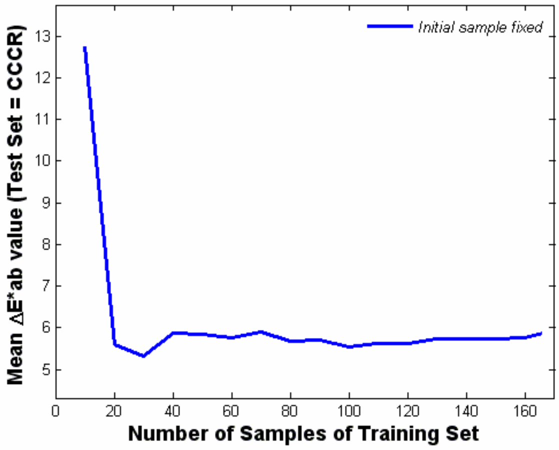

Figure 1. *Mean CIELAB colour difference ΔE_{ab}^* over the samples of the CCCR chart used as test set, versus the number of samples (size) of the training set for the colorimetric configuration.*

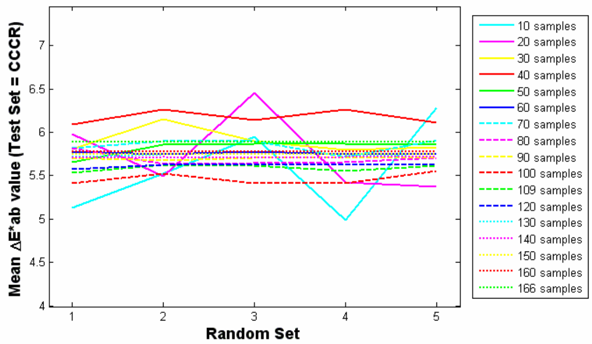

Figure 2. *Mean CIELAB colour difference ΔE_{ab}^* over the samples of the CCCR chart used as test set for the colorimetric configuration, for all numbers of samples (size) and for the 5 training sets selected of each size.*

Table 2. Percentage of fluctuation (eq. [1]) of mean CIELAB colour difference ΔEab* over the samples of the CCCR chart between the 5 training sets selected of each size in the second selection, for the colorimetric configuration.

# colour samples	*%fluctuation*	*# colour samples*	*%fluctuation*
10	9.75	90	0.27
20	8.06	100	1.37
30	2.51	109	0.80
40	1.30	120	0.43
50	1.58	130	0.38
60	0.26	140	0.10
70	1.43	150	0.00
80	1.27	161	0.00
		166	0.00

©2008 Society for Imaging Science and Technology

%fluctuac= 100·σ(mean)/mean [1]

The same behaviour in terms of the number of samples of the training set and for the 5 training sets of the second selection is observed in the minimum and maximum CIELAB colour difference values over the samples of the CCCR chart used as test set.

Multispectral Configuration

Results obtained for the multispectral configuration and the first selection of the training set carried out (fixed initial sample), are very similar to those for the colorimetric configuration, as it can be seen in Figure 3. Increasing the size of the training set improves accuracy of colour measurements outstandingly up to approximately 40 samples. From here on, accuracy slightly fluctuates around a constant value (Figure 3.).

Results obtained for the multispectral configuration and the second selection of the training set carried out (randomly selected initial sample, 5 training sets of each size), are also very similar to those for the colorimetric configuration (Figure 4., Table 3.).

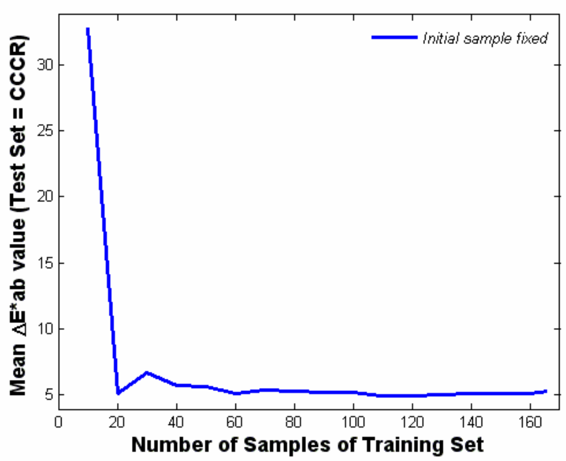

Figure 3. *Mean CIELAB colour difference ΔE_{ab}* over the samples of the CCCR chart used as test set, versus the number of samples (size) of the training set for the multispectral configuration.*

All the mean, minimum, and maximum values of the CIELAB colour differences over the samples of the CCCR chart, for the 5 training sets selected of each size, show a very similar behaviour depending on the number of samples of the training set. For a low number of samples, system's performance greatly depends on the colour samples composing the training set, as it can be seen in fluctuations of the mean CIELAB colour difference depending on the training set considered (Figure 4., Table 3.).

According to the improvement in accuracy of system's performance, increasing the size of the training set up to approximately 40 samples also reduces its dependence on the colour samples of the training set. From 40 samples on, dependence of system's performance on the training set and its accuracy remain approximately constant. Consequently, it can be stated that a training set of low numerosity (40 samples) properly selected over the colour gamut available, is enough to perform a complete training of the imaging system, independently of the concrete set of colour samples selected.

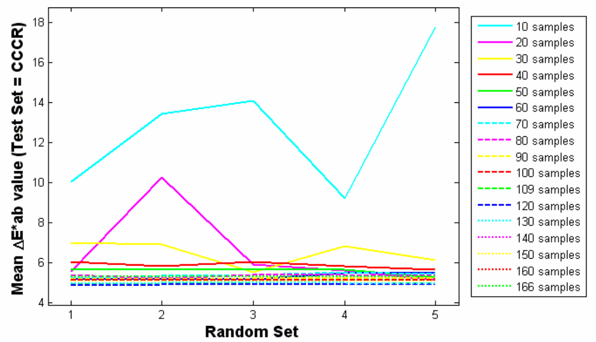

Figure 4. *Mean CIELAB colour difference ΔE_{ab}* over the samples of the CCCR chart used as test set for the multispectral configuration, for all numbers of samples (size) and for the 5 training sets selected of each size.*

Table 3. Percentage of fluctuation (eq. [1]) of mean CIELAB colour difference ΔEab* over the samples of the CCCR chart between the 5 training sets selected of each size in the second selection, for the multispectral configuration.

# colour samples	% fluctuation
10	26.55
20	32.66
30	9.76
40	2.89
50	3.03
60	3.13
70	1.86
80	1.50

# colour samples	% fluctuation
90	0.52
100	0.32
109	0.69
120	0.35
130	0.41
140	0.11
150	0.00
161	0.00
166	0.00

Conclusions

In this work, the influence of the size of the training set on colour measurements performed using an imaging system based on a CCD camera has been analyzed. Selection of training sets of different sizes has been performed using an easy selection method, based on differences in a* and b* CIELAB coordinates between the selected colour samples from an initial training set.

Two configuration of the imaging system have been used: a colorimetric configuration with 3 acquisition channels, and a multispectral configuration with 7 acquisition channels.

Very similar results were obtained for both configurations. Training sets having from 40 samples on, lead to very similar results, practically independent of the number of samples of the training set. It has been observed that as the size of the training set increases, fluctuations between results obtained using equal-sized sets selected from a first sample selected randomly, decrease and are greatly reduced from 40 samples on. System's performance depends on the colour samples of the training set for sets with less than 40 samples.

Taking into account these results, it can be concluded that improvement in accuracy of colour measurements by increasing the size of the training set is limited, probably due to the fact that the expected improvement due to the increase of information is made up by the increase of errors accumulated in adjustment. Furthermore, a proper selection, uniform over the colour gamut available, of a low number of samples (40) allows to characterize the imaging system as well as a greater set of samples, for instance, the whole useful set of colour patches (166) of the CCDC.

References

[1] J. Y. Hardeberg, H. Brettel, and F. Schnitt, Spectral characterisation of electronic cameras, Proc. SPIE 3409, p. 100 (1998).

[2] T. L. V. Cheung and S. Westland, Color Selections for Characterization Charts, Second European Conference on Color in Graphics, Imaging and Vision (CGIV'04), Proc. IS&T's, pg. 116. (2004).

[3] P. Pellegri, G. Novati, and R. Schettini, Training Set Selection for Multispectral Imaging Systems Characterization, J. of Imaging Sci. and Technol., 48(3), 203 (2004).

[4] de Lasarte M., Vilaseca M., Pujol J., and Arjona M., Color measurements with colorimetric and multi.spectral imaging systems, Spectral Imaging: Eighth International Symposium on Multispectral Color Science, Proc. SPIE-IS&T Electronic Imaging, SPIE Vol. 6062, Bellingham, WA, pg. 60620F1. (2006).

[5] Pujol J., de Lasarte M., Vilaseca M., Arjona M., High Dynamic Range Multispectral System for Wide Color Gamut Measurements, Third European Conference on Color in Graphics, Imaging and Vision (CGIV'06), Proc. IS&T's, pg. 404. (2006).

Author Biography

Marta de Lasarte completed her BSc Degree in Physics at the Autonomous University of Barcelona (Spain) in 2004 and received the 2004 year Extraordinary Award of Degree End from the Autonomous University of Barcelona. She is currently enrolled in the PhD program in Optical Engineering at the Technical University of Catalonia, having received a PhD grant from the Ministerio de Educación y Ciencia of Spain. Her work focuses on color imaging (device calibration and characterization, color management) and industrial colorimetry.

Acknowledgements

This research was supported by the Spanish Ministry for Education and Science (Ministerio de Educación y Ciencia) by means of the grant number DPI2005-08999-C02-01. M. de Lasarte would like to thank the Ministerio de Educación y Ciencia of Spain for the Ph.D. grant she has received.

 ©2008 Society for Imaging Science and Technology