Spectral Representation of Object Colours

Yi-Fan Chou[1,2], Vien Cheung[1], M Ronnier Luo[1] and San-Liang Lee[2]
[1]School of Design, University of Leeds, Leeds, UK
[2]Department of Electronic Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

Abstract

Reflectance functions can be represented by low-dimensional linear models with weighted sum of principal components (or, often, referred to as basis functions). Such method to obtain a low-dimensional linear model is based on principal component analysis (PCA). The specific requirement for a low-dimensional model is to accumulate fraction of variance of the basis functions. The more basis functions included the more fraction of variance accumulates. The investigation of how many basis functions required so as to represent reflectance functions accurately has been extensively studied over the last two decades [1, 2] since Cohen fitted a linear model to spectral reflectance functions of Munsell color chips in 1964 [3]. In this paper, a comprehensive dataset of 97593 including six types of materials has been accumulated. These materials are paint, graphic, plastic, textile, skin and natural samples. Principal component analysis for each material has been studied. The effective dimension of reflectance functions representations for these materials has examined. It was found that a single set of basis functions can essentially be applied to represent all spectra in the world.

Introduction

The smoothness of spectral reflectance can be represented by a linear combination of few basis functions. These functions can represent the physical property of surface colours. The weighted sum of basis functions is able to approximate any reflectance curve to a specified degree of accuracy. The method called Principal Component Analysis (PCA) is often applied to derive the basis functions from a dataset. The basis functions are constrained to be orthogonal to each other and allow large datasets to be accurately represented by a small number of linear combinations of basis functions with a corresponding set of coefficients. Although spectral representation using basis functions have been extensively studied over the last 20 years, the results have been not conclusive regarding to the basis functions needed to represent the spectral reflectance and the dimension of spectral reflectances [3-7].

The aim of this study is to derive a small number of basis functions from a set of spectra, representing the world of colours. A database, known as Leeds dataset, including 97593 spectral data items was accumulated. It comprises various materials such as paint, graphic, plastic, textile, skin and natural datasets. The similarities between the spectral characteristics of different materials were compared. The number of basis functions required to have accurately spectral reconstruction for each material was examined. A set of basis functions was derived to represent all spectral data with different materials.

Spectral Datasets

A comprehensive dataset having spectral reflectances of both artificial and natural materials was accumulated. It also included some spectra from the Standard Object Colour Spectra Database (SOCS) which is a collection of typical and difference sets of existing object colour reflectance spectra [8]. Table 1 tabulates all datasets and number of samples (NOS) in each dataset. According to their physical materials, they are divided into six groups: paint, graphic, plastic, textile, skin and natural. The spectra in different datasets originally had different wavelength ranges and sampling intervals. They are all interpolated into 10 nm intervals from 400 to 700 nm.

The paint group includes 672, 1562, 720, 1794, 981 and 505 samples from Dupont SpectraMaster®, Munsell Book of Color – Glossy collection [9], Munsell Limit Color Cascade [10], NCS [11], DIN [12] and SOCS-Paint, respectively. The graphic group has the sample printed by inks, much finer pigment particle than paint ones, including 26782, 30624, 1152, 7856 samples from Packaging Printing Ink [13], SOCS-Graphic, SOCS-Photo, and SOCS-Printer, respectively. The plastic group has one dataset including 5337 samples provided by a plastic company via private communication. The textile group has colours on cotton, polyester, wool and silk including 4028, 1925, 1063 and 2832 samples from Industry Cotton set, Pantone® Cotton, PCC [14] and SOCS-Textile, respectively. The skin group includes 357, 340, 8213 samples from Oulu Skin [15], RIT Skin [16, 17] and SOCS-Skin respectively. The skin colours were collected from different positions on face and different continents. The natural database consists of 494 and 404 samples collected by Cheung's [18] and Westland *et al*.'s [6] natural sets, respectively. The samples include leaves, petals, grass and barks. There are totally 97593 spectra in the collection.

Table 1: Description of each reflectance dataset

Dataset	NOS	Dataset	NOS
Paint Group	6189	Textile Group	9848
Dupont SpectraMaster®	672	Industry Cotton	4028
Munsell I	1562	Pantone Cotton	1925
Munsell II	720	PCC	1063
NCS	1749	SOCS-Textile	2832
DIN	981	Graphic Group	66414
SOCS-Paint	505	Packaging Ink	26782
Skin Group	8910	SOCS-Graphic	30624
Oulu Skin	357	SOCS-Photo	1152
RIT Skin	340	SOCS-Printer	7856
SOCS-Face	8213	Natural Group	898
Plastic Group	5337	Natural I	494
Industry Plastic	5337	Natural II	404

Linear Models for Reflectance Representation

For a given set of reflectance values with N sampling points for K samples, denoted by a $N \times K$ matrix $\boldsymbol{P}$, the basis functions $\boldsymbol{U} = [\vec{\boldsymbol{u}}_1\ \vec{\boldsymbol{u}}_2\ \vec{\boldsymbol{u}}_3 ... \vec{\boldsymbol{u}}_n ... \vec{\boldsymbol{u}}_N]$ and the associated singular values $\gamma_1 > \gamma_2 > ... > \gamma_n > ... > \gamma_N \geq 0$ can be found using the principal component analysis (PCA). The representation of a spectral curve is weighted sum of a set of basis functions. For a reflectance vector $\vec{\boldsymbol{p}}_k$ associated to the k_{th} sample can be reconstructed by the linear combination of the basis functions:

$$\vec{\boldsymbol{p}}_k = \boldsymbol{U} \cdot \vec{\boldsymbol{\Gamma}}_k = (\vec{\boldsymbol{u}}_1 \cdot \beta_1 + \vec{\boldsymbol{u}}_2 \cdot \beta_2 + ... + \vec{\boldsymbol{u}}_n \cdot \beta_n + ... + \vec{\boldsymbol{u}}_N \cdot \beta_N) \qquad \text{Eq. 1}$$

where β_n is a weighted coefficient of the vector $\vec{\boldsymbol{p}}_k$ corresponding to the basis function $\vec{\boldsymbol{u}}_n$; the $\vec{\boldsymbol{\Gamma}}_k$ is a $N \times 1$ column vector whose components are the weighted coefficients of the vector $\vec{\boldsymbol{p}}_k$ corresponding to the basis functions from the matrix $\boldsymbol{U}$. Since vectors of $\boldsymbol{U}$ are orthonormal, the coefficient can be obtain by

$$\beta_n = \vec{\boldsymbol{p}}_k \cdot \vec{\boldsymbol{u}}_n \text{ where } n = 1,2,...,N \qquad (2)$$

In accordance with previous studies [19-21], there is a strong concentration of variance or energy in the first few singular values. The first basis function maximally represents the variance in the data, and subsequent basis functions maximally represent the remaining variance. This indicates that the reflectance functions can be represented using subspace with $\widetilde{N} < N$ components to reduce the dimension of the vector space. The number of the basis function $\widetilde{N}$ can be chosen with $\widetilde{N}$ less than N according to the eigenvalues. The first $\widetilde{N}$ basis functions are denoted by an N by $\widetilde{N}$ matrix $\boldsymbol{U}^{(\widetilde{N})}$. The basis functions related to small singular values can be removed and Eq. 1 becomes:

$$\vec{\boldsymbol{p}}_k \approx \boldsymbol{U}^{(\widetilde{N})} \cdot \vec{\boldsymbol{\Gamma}}_k^{(\widetilde{N})} = (\vec{\boldsymbol{u}}_1 \cdot \beta_1 + \vec{\boldsymbol{u}}_2 \cdot \beta_2 + ... + \vec{\boldsymbol{u}}_{\widetilde{N}} \cdot \beta_{\widetilde{N}}), 1 \leq \widetilde{N} \leq N \qquad (3)$$

where $\vec{\boldsymbol{\Gamma}}_k^{(\widetilde{N})}$ is the coefficient vector corresponding to the basis functions $\boldsymbol{U}^{(\widetilde{N})}$. This reduces the number of vector spaces to represent a reflectance functions. To determine an estimation of the effective dimension of the space, the accumulated fraction of variance denoted as $E_a(\widetilde{N})$ is a useful notion:

$$E_a(\widetilde{N}) = \frac{\sum_{n=1}^{\widetilde{N}} \gamma_n}{\sum_{n=1}^{N} \gamma_n} \qquad (4)$$

Accumulated fraction of variance is defined as a ratio of the energy represented by the first $\widetilde{N}$ singular values to the total energy. The amount of required accumulated fraction of variance does not have a clear criterion. A typical value is 99% used by several researchers [2-4, 22]. In this study, the spectral representation quality was analysed with colour perception error and spectral reconstruction error. Two measures were used by the CIE DE2000 colour difference and Goodness-of-Fit Coefficient (GFC). A value of 1 ΔE_{00} unit was used as a criterion which is generally considered to be approximately one just noticeable difference (JND) in colour. The calculations were performed for CIE 1964 10 degree standard observer under a set of standard illuminants including D65, A and F11. The GFC [23, 24] is defined by:

$$GFC = \frac{\left| \sum_{i=a}^{b} P_m(\lambda_i) P_e(\lambda_i) \right|}{\sqrt{\sum_{i=a}^{b} [P_m(\lambda_i)]^2} \sqrt{\sum_{i=a}^{b} [P_e(\lambda_i)]^2}} \qquad (5)$$

where $P_m(\lambda_i)$ and $P_e(\lambda_i)$ are the measured and the estimated spectral data at wavelength λ_i; a and b are the upper and lower bounds of the wavelength range.

Spectral Representation

The PCA was applied to all the datasets described above to derive basis functions. Figure 1 shows a set of subplots. The subplots, from top to bottom, indicate the basis function derived from each material group in the sequence of paint, graphic, plastic, textile, skin and natural samples. Figure 1 shows, from left to right, the first four basis functions.

In Figure 1, the first three basis functions from the subsets in the same group were closely overlapped to each other, especially for the paint group. All the paint subsets can be represented by basis functions with similar profile. The basis function representations of artificial material groups including the paint, graphic, plastic and textile are very similar. The two natural subsets had different characteristics and this can be explained by the differences in their colour gamuts. The fourth basis functions tend to oscillate more rapidly. The difference between subsets became clear. The present results showed that in each material group seven basis functions were sufficient to have spectral representation with average colorimetric accuracy less than 1.0 ΔE_{00} unit under illuminant D65, A and F11.Regardless of basis functions derived from which subset, they all gave good predictions to their material group. This inspired the idea to develop a universal set of basis functions for all world materials.

According to the materials, six material groups are obtained and denoted as “PAINT”, “GRAPHIC”, “PLASTIC”, “TEXTILE”, “SKIN”, and “NATURAL”. Their basis functions are shown in Figure 1. Two more universal sets named “ARTIFICIAL” and “MIX” are also generated. The former set combined the four artificial groups to derive a set of basis function. The latter set consists of 400 spectra randomly chosen from each of the six generic sets, totally 2400 spectra. This set is designed to avoid the bias of basis function caused by different numbers of samples in different groups. A set of basis functions was also derived from all data of the 97593 reflectances denoted as “ALL”. Figure 2 shows the first six basis functions. It can be seen that the “ALL” set had the same curves of basis functions as the “ARTIFICIAL” set since the artificial samples was the majority of the full set. Thus, it is not necessary to use the basis functions of “ALL” set.

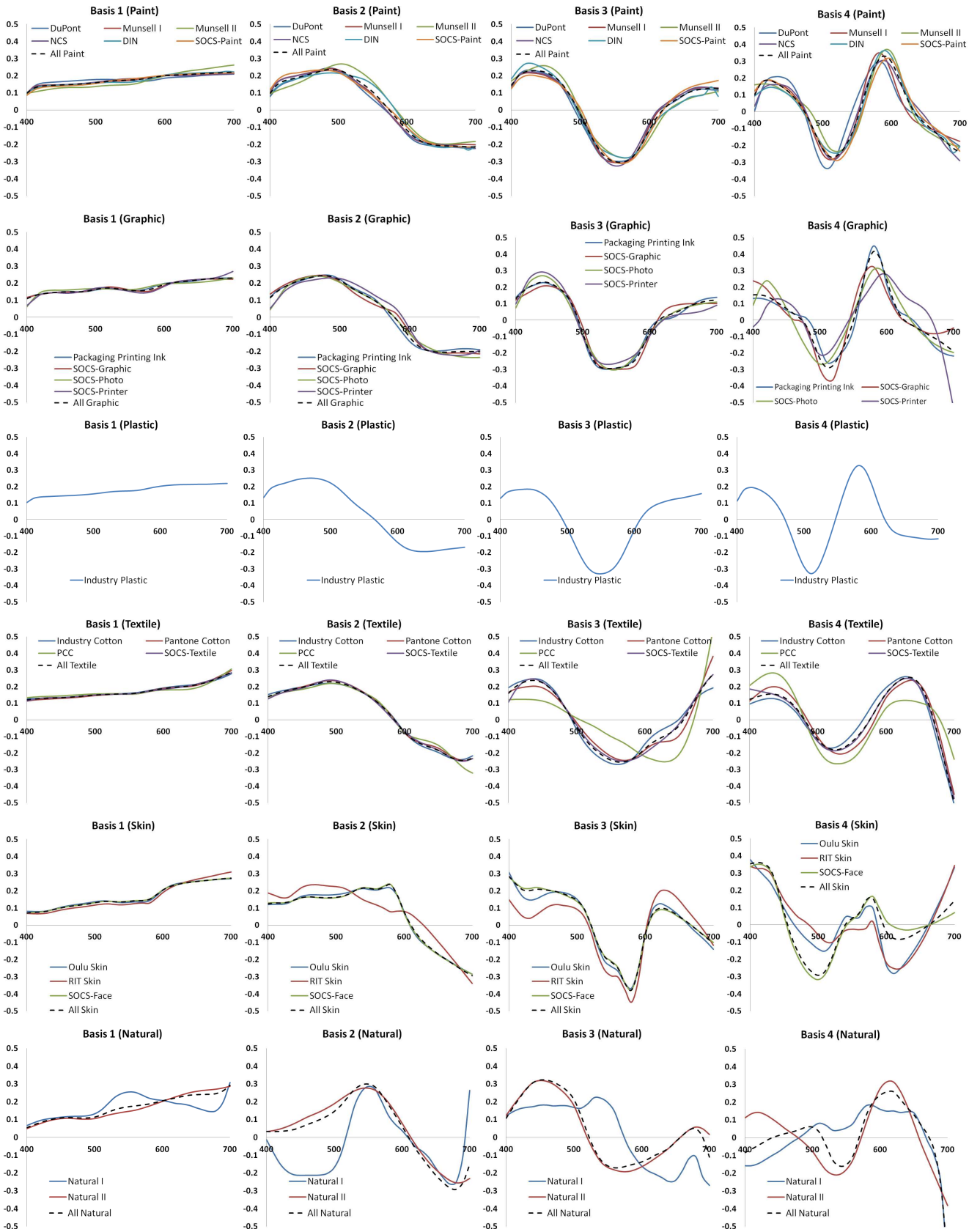

***Figure 1**: The first four basis functions derived from each dataset*

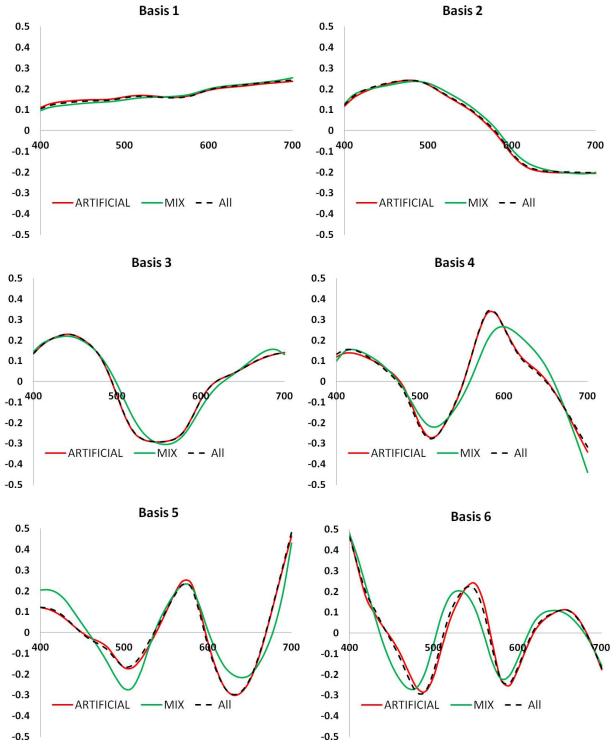

***Figure 2**: The six basis function derived from the universal sets*

A set of basis functions was derived from each material group and universal set. The reconstruction errors tested by the training group itself in average ΔE_{00} [25], average GFC [24] and accumulated fraction of variance [2] against number of basis functions are presented in Figure 3. For a perfect agreement, the former measure should be zero and the latter two measures should be one. The performance presented is expressed by the mean of the performance under D65, A and F11. In most cases, the errors were largely improved when the number of basis increased from three to four. The "SKIN" group was an exception that only three basis functions were enough to reconstruct the spectra with the mean ΔE_{00} less than one, average GFC higher than 0.999, and accumulated fraction of variance higher than 0.95. Generally, five basis functions were required to achieve mean ΔE_{00} less than one unit except the "TEXTILE" and "NATURAL" groups which needed 6 basis functions. The spectral errors needed seven basis functions to achieve average GFC of 0.999 except the two universal sets and the "GRAPHIC" group which needed eight basis functions. To have the accumulated fraction of variance higher than 0.95, seven to eight basis functions were required.

Eight sets of basis functions were derived from each of the following group: PAINT, GRAPHIC, PLASTIC, TEXTILE, SKIN, NATURAL, ARTIFICIAL and MIX. These sets of basis functions were used to reconstruct each of the material group. With the criterion in mean ΔE_{00} of 1 unit for the average colorimetric errors under the three illuminants, the number of needed basis functions is shown in Table 2.

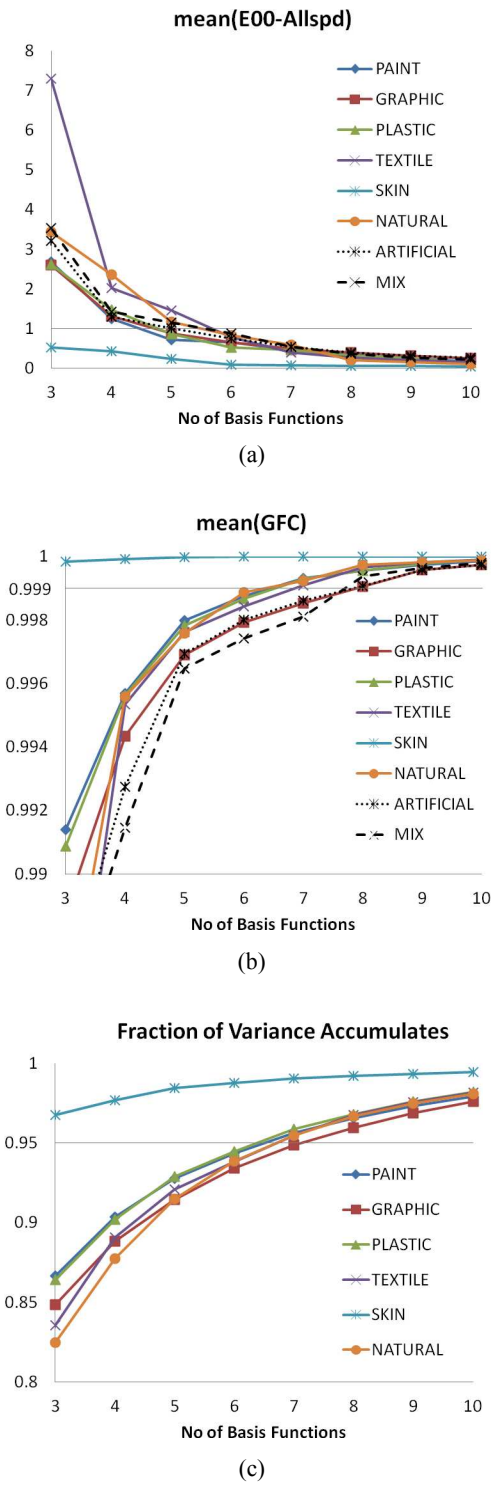

***Figure 3**: Reconstruction error in (a) average ΔE_{00} (b) average GFC (c) Accumulated Fraction of variance against number of basis functions*

Table 2: Using different sets of basis functions to represent each material group required to achieve mean colour difference less than 1 ΔE_{00} unit

Training \ Testing	PAINT	GRAPHIC	PLASTIC	TEXTILE	SKIN	NATURAL
PAINT	**5**	5	5	8	5	8
GRAPHIC	5	**5**	5	7	4	9
PLASTIC	5	5	**5**	8	4	9
TEXTILE	6	6	6	**6**	5	8
SKIN	7	8	7	9	**3**	>10
NATURAL	7	7	7	8	7	**6**
ARTIFICIAL	5	5	5	7	4	9
MIX	5	6	6	7	4	7
SUMMARY	5~7	5~8	5~7	6~9	3~7	6~10

For the PAINT, GRAPHIC, and PLASTIC groups, they all required five basis functions for representing each other due to their high similarity in the spectral characteristics. The five to six basis functions of the TEXTILE group were sufficient to represent all of these groups except the NATURAL set. For reconstructing skin spectra, fewer basis functions were needed, but the basis functions derived from the SKIN group were not suitable to represent the other sets. The NATURAL generally needed more basis functions to reconstruct because the spectral characteristics were different from the others.

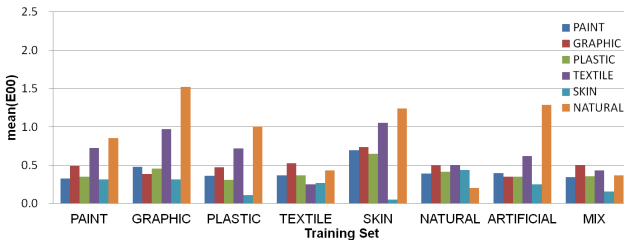

Figure 4: *The colorimetric errors in terms of mean ΔE_{00} with eight basis functions of each set tested by the six material groups.*

Figure 4 show the colorimetric errors in terms of mean ΔE_{00} with eight basis functions of each set by testing the six material groups. The spectral reconstructions of the TEXTILE and NATURAL were worse than other groups with exception of using basis functions of themselves. Even the ARTIFICIAL set did not give better reconstruction for the TEXTILE group than using the material itself. By using the eight basis functions of the MIX group, the errors in the TEXTILE and NATURAL groups were reduced. The basis functions derived from the TEXTILE and MIX groups gave 0.37 and 0.36 mean ΔE_{00} for all groups, respectively. This was better than using the other sets to derive basis functions.

Both of the GFC and ΔE_{00} were used in this study, and they gave the similar results so that only ΔE_{00} results are given above. In terms of spectral errors, the basis functions derived from the MIX group gave 26.66% reconstructed spectra with the GFC values higher than 0.9999 where the TEXTILE set gave 14.50%. Both two sets only had less than 1% reconstructed spectra with the GFC values less than 0.99.

Conclusions

A comprehensive dataset of 97593 spectra was accumulated in this study. They included six materials: paint, graphic, plastic, textile, skin and natural objects. Each of them had several subsets. The number of basis functions required for representing each set has been investigated. The results concerning the needed basis functions are based on our calculations with 10 degree standard observer under illuminant D65, A and F11. The curves of basis functions for the paint, graphic and plastic were very similar. They gave a good reconstruction to each other. In the natural or skin database, the spectra can be reconstructed well by using their own basis functions. However, even though the basis functions derived from different kind of subsets had different spectra characteristics, they still reconstructed the spectra well but more basis functions were needed to get equivalent accuracy. Generally, the first basis function was the smoothest and the higher order principal components tend to oscillate more rapidly. This is consistent with many previous researches [5, 26].

A set of basis functions was also developed from a mixture set. The mixture set consisted of 400 spectra randomly selected from each of the six materials, and 2400 spectra in total. The result shows the eight basis functions derived from this MIX set provide good spectral reconstruction. The mean colorimetric error was 0.36 ΔE_{00}, and more than 99% reconstructed spectra had GFC value higher than 0.999.

References

[1] Dannemiller, J.L. Spectral reflectance of natural objects - how many basis functions are necessary. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 1992, **9**(4), pp.507-515.

[2] Hardeberg, J.Y. On the spectral dimensionality of object colours. *In: 1st European Conference on Colour in Graphics, Imaging, and Vision*, 2002, pp.480-485.

[3] Cohen, J. Dependency of the spectral reflectance curves of the Munsell color chips. *Psychonomic Science*, 1964, **1**(12), pp.369-370.

[4] Maloney, L.T. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 1986, **3**(10), pp.1673-1683.

[5] Jaaskelainen, T., J. Parkkinen and S. Toyooka. Vector-subspace model for color representation. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 1990, **7**(4), pp.725-730.

[6] Westland, S., J. Shaw and H. Owens. Colour statistics of natural and man-made surfaces. *Sensor Review*, 2000, **20**(1), pp.50-55.

[7] Kohonen, O., J. Parkkinen and T. Jaaskelainen. Databases for spectral color science. *Color Research and Application*, 2006, **31**(5), pp.381-390.

[8] Jsa. *Graphic Technology-Standard object color spectra database for colour reproduction evaluation*. Information Processing Divisional Council, Japanese Standards Association, 1998.

[9] Nickerson, D. History of Munsell Color System. *Color Engineering*, 1969, **7**(5), pp.42-&.

[10] Pointer, M.R. The gamut of real surface colors. *Color Research and Application*, 1980, **5**(3), pp.145-155.

[11] Tonnquist, G. Das Natu□rliche Farbsystem NCS. 1986, pp.462-469.

[12] Richter, M. The official German standard color chart. *Journal of the Optical Society of America*, 1955, **45**(3), pp.223-226.

[13] Li, X., C.J. Li, M.R. Luo, M. Pointer, M. Cho and J. Kinf. A new colour gamut for object colours. *In:* 2007, pp.283-287.

[14] Park, J. and K.M. Park. Professional colour communicator (PCC) - the definitive colour selector. *Book of Papers - Aatcc 1994 International Conference & Exhibition*, 1994, pp.294-296.

[15] Jones, M.J. and J.M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 2002, **46**(1), pp.81-96.

[16] Sun, Q. and M.D. Fairchild. Statistical characterization of face spectral reflectances and its application to human portraiture spectral estimation. *Journal of Imaging Science and Technology*, 2002, **46**(6), pp.498-506.

[17] Sun, Q. and M. Fairchild. A new procedure for capturing spectral images of human portraiture. *Proceedings of the 9th Congress of the International Colour Association, AIC Color 2001*, 2001.

[18] Cheung, V. *Colour statistics of natural surfaces*. MSc thesis, University of Derby, 2000.

[19] Hardeberg, J.Y., F. Schmitt and H. Brettel. Multispectral color image capture using a liquid crystal tunable filter. *Optical Engineering*, 2002, **41**(10), pp.2532-2548.

[20] Imai, F.H. and R.S. Berns. A comparative analysis of spectral reflectance reconstruction in various spaces using a trichromatic camera system. *In: IS&T/SID 7th Color Imaging Conference*, 1999, pp.21-25.

[21] Hardeberg, J.Y. ed. *Acquisition and reproduction of color images: Colorimetric and multispectral approaches*. Paris, France: Ecole Nationale Superieure des Telecommunications, 1999.

[22] Haneishi, H., T. Hasegawa, N. Tsumura and Y. Miyake. Design of color filters for recording artworks. *Proc. of IS&T's 50th Annual Conference*, 1997, pp.369-372.

[23] Mansouri, A., T. Sliwa, J.Y. Hardeberg and Y. Voisin. Representation and estimation of spectral reflectances using projection on PCA and wavelet bases. *Color Research and Application*, 2008, **33**(6), pp.485-493.

[24] Romero, J., A. García-Beltrán and J. Hernández-Andrés. Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 1997, **14**(5), pp.1007-1014.

[25] Luo, M.R. The CIE 2000 colour difference formula: CIEDE2000. *In: 9th Congress of the International Colour Association*, 2001, pp.554-559.

[26] Parkkinen, J.P.S., J. Hallikainen and T. Jaaskelainen. Characteristic spectra of Munsell colors. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 1989, **6**(2), pp.318-322.