The Use of a Selective Database Technique in Order to Recover the Spectra of a Series of Acrylic Paints by the Principle Component Analysis

Ansari, Keivan
Department of Polymer and Color Engineering, Amirkabir University of Technology, Tehran, I.R. IRAN

***Amirshahi, Sayed Hossein* *+**
Department of Textile Engineering, Amirkabir University of Technology, Tehran, I.R. IRAN

Moradian, Siamak
Department of Polymer and Color Engineering, Amirkabir University of Technology, Tehran, I.R. IRAN

ABSTRACT: *A procedure for an efficient recovering of reflectance spectra of Acrylic paint samples from CIE tristimulus color values is described. By fixing a certain criteria based on color difference value, the proposed technique preliminarily selects a series of suitable samples from a main dataset containing the reflectance values of a series of different Acrylic paint samples, based on the color specifications of given samples. In this way, a series of different databases could be formed around a particular sample. The well-known principal-components linear model was used to recover the spectral data from their corresponding color coordinates by using only 3 basis functions.*

The surface spectra of a set of 2802 samples are collected for the recovery of the reflectance values of Acrylic paint samples whose tristimulus values were known. The role of the value of color difference for selecting suitable samples is discussed. The recovered spectra achieved by this method show considerable improvements in terms of root mean squarer (RMS) error and goodness-fitting coefficient as well as color difference values under different illuminants as compared to the recovery from the main database.

KEY WORDS: *Reflectance, Acrylic paint, Principal component analysis, Spectral estimation, Tristimulus values.*

INTRODUCTION

Principal component analysis, abbreviated PCA, has been used since 1964 as a mathematical tool in color technology [1]. Two purposes have been mainly followed by applying PCA, data reduction and reconstructing of

* *To whom correspondence should be addressed.*
+ *E-mail: hamirsha@aut.ac.ir*
1021-9986/06/2/39 7/$/2.70

spectral data. However, definition of principal directions of a set of data could be another aim of application of PCA [2-11]. In this aspect, the extraction of spectral reflectance of samples from their corresponding colorimetric values has been interesting subject to follow in certain aspects of optics such as color science as well as imaging technology. Linear models built by the approximation of spectral functions in the wavelength domain are normally used for this purpose [12].

The success of the method largely depends on the number of basis functions which have been chosen for approximating the spectral data. The numbers of principal components for acceptable recovery of reflectance or spectral compression varies between 3 to 9, based on the types of datasets applied. However, 7 basis vectors are generally considered to be adequate for acceptable approximation [13]. Clearly, the spectral information of a colored object is critical data in color manufacturing such as paints, textiles, ink and imaging industries. Providing of spectral data is more expensive in comparison with the colorimetric one and it is reasonable to prepare suitable data set for converting the later to the former with acceptable error. Formation of digital archive in image processing technique has been seriously pursued in recent years.

The colorimetric measurement of surface colors under a given set of CIE standard illuminant/observer combination yields the corresponding tristimulus X, Y and Z values. These values could be converted to any other color coordinates such as the L*a*b* coordinates. Determination of color values from their corresponding spectral data such as the reflectance or the transmittance behaviors of samples is a common type of psychophysical dimensionality reduction technique.

This procedure converts the multi dimensional spectrum into a 3 dimensional color space, based on the weighting of the spectral return of surface received by the standard observer, as shown in equation (1) [11, 14].

$$Q_j = \int_{400}^{700} S(\lambda).R(\lambda).q_j(\lambda).d\lambda \quad (1)$$

Where Q denotes the tristimulus values of X, Y and Z, and j varies from 1 to 3 over X, Y and Z. S, R and q respectively represent the spectral power distribution of a standard light source, the reflectance behavior of a surface color and the color matching functions of a standard observer. Several papers have been published [9, 11, 12, 15] attempting to extract the reflectance data, R, from the corresponding tristimulus values Q_j. In the present paper, a new technique is presented as an attempt to recover the spectral data from the color coordinates of a sample, using a progressive dataset selection technique which benefits from limited dimensionally adapted to the color specifications of a proposed color surface.

MATHEMATIC BACKGROUND

Equation (1) can be rewritten as equation (2) in the matrix form [12].

$$Q = A^T.R \quad (2)$$

Where A is the weight product sets of standard illuminant/observer combination. Suppose that the spectral data were presented in 10 nm intervals then, the spectral reflectance of sample with known tristimulus values Q, could be determined by equation (3).

$$R = \left(A^T\right)^+.Q \quad (3)$$

Where, $(AT)^+$ is the pseudo-inverse of A^T. Equation (3) would be underdetermined and the estimation of R, represented by $\widetilde{R}$, would be very poor. In fact, using a pseudo-inverse could be an immediate solution for recovery of spectral reflectance however; it does not lead to a realistic result. In this case, the distance between R and $\widetilde{R}$, called $d(R,\widetilde{R})$, is generally too high and the extracted spectrum is absolutely spiky, completely different from the smooth behavior of the surface reflectances.

A reasonable approach to solve this problem is based upon the fact that the reflectance spectra of natural and man-made surfaces are smooth functions of wavelength and can be represented by a linear model, as shown in equation (4) [12].

$$\widetilde{R}(\lambda) = V_0(\lambda) + V(\lambda).C \quad (4)$$

$V_0(\lambda)$ is the mean of the spectral reflectance values of the dataset and $V(\lambda)$ is the selected basis vectors. C is a column vector of k elements which contains the principal component coordinates. By substitution of equation (4) into equation (3), we have:

$$\left(A^T\right)^{-1}\cdot Q = V_0(\lambda) + V(\lambda)\cdot C \quad (5)$$

Then, the tristimulus values can be determined by:

$$Q = A^T.V_0 + A^T.V.C \quad (6)$$

Where, $A^T.V_0$ is the tristimulus values of mean vector and can be demonstrated by Q_{V_0} and the $A^T.V$ is a $3\times k$ matrix containing the tristimulus values of the first k principal components, shown by:

$$\begin{bmatrix} X_1 & X_2 & \dots & \dots & X_k \\ Y_1 & Y_2 & \dots & \dots & Y_k \\ Z_1 & Z_2 & \dots & \dots & Z_k \end{bmatrix}$$

If T is defined as:

$$T = A^T.V \quad (7)$$

Then, equation (6) can be written as:

$$Q = Q_{V_0} + T.C \quad (8)$$

If a set of target tristimulus values is known, it is possible to calculate the column vector C for that set of values.

$$C = T^+(Q - Q_{V_0}) \quad (9)$$

The pseudo-inverse, T^+, can be replaced by the common inverse, T^{-1}, if the dimension of k becomes equal to 3. Since, the selected basis vectors are not adequate in this case, the $d(R, \tilde{R})$ can not be good enough, generally. The residual between R and $\tilde{R}$ could be abandoned if the similarity between the reflectance values of samples in the dataset increases [8, 10] could be recovered by using equation (4).

METHOD

The reflectance data of 2802 matt Acrylic paint specimens obtained from Garcia-Beltran et al. work were implemented [8]. They have generated these samples set using Acrylic paint within 24 basic commercial Acrylic colors for artists on paper, with creating a total of 2802 different samples, each 4×4 cm in area, with a homogeneous texture. Beginning from these 24 basic colors (each one defined from a determined type of pigment) they obtained samples by mixing 2 or 3 of these colors within different ratios. The wavelength intervals of spectrum for samples were fixed to 20 nm within 400 t0 700 nm. The tristimulus values of samples were determined under equal energy illuminant and the CIE 1964 standard observer. Two dimensional a* and b* coordinates as well as the three dimensional L*a*b* representation of samples are shown in Fig. 1.

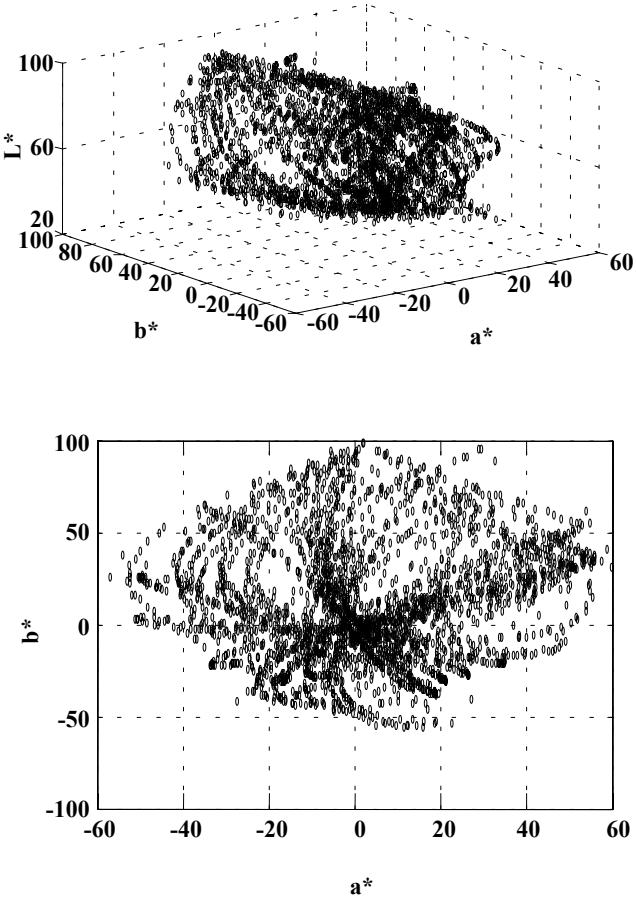

Fig. 1: Two and three dimensional distribution of Acrylic paints in L*a*b* color space.

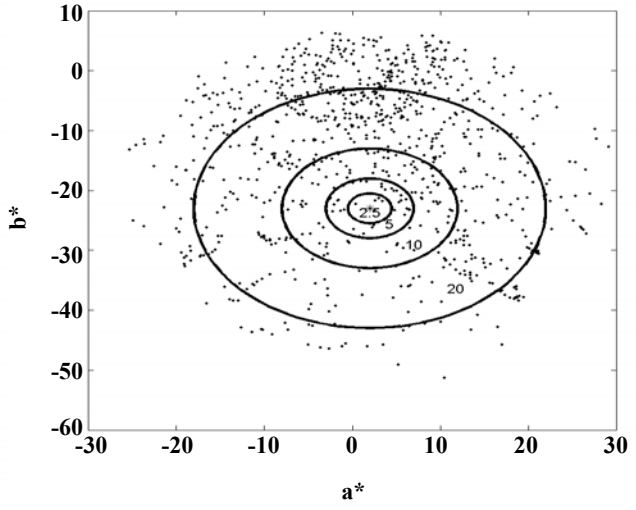

Fig. 2: The algorithm of finding of samples which are suitable for construction of dataset for recovery of spectral behavior of target sample, showed by *. Target sample is selected randomly. The contours show different ΔE ***values.***

In order to reduce the dimension of dataset, the reflectance vectors of suitable surfaces were extracted from the main data matrix by using a defined quantitative criterion depending on the color specification of the proposed sample. Two different criteria were adjusted simultaneously for data selection, based on the color specifications of target samples and the number of samples remained in the dataset. The criteria utilized for data selection were the conformance of a predetermined number of suitable samples (at least 3) giving a known color difference value. The tolerable color differences were then increased in order to minimize the number of samples should the algorithm not find at least three corresponding samples.

The Acrylic paint samples were used for determining the basis vectors and the proper data matrices extracted for each of 2802 Acrylic paint samples and the recovery process was continued by using a variety of smaller data sets with different mean vectors in comparison with the original data. By this method, the dimension of spectral dataset was initially reduced before determining the eigenvectors and in this case, it was supposed that 3 basis functions could approximate the reflectance behaviors of such samples.

This would means that, different datasets could be created for determining of the basis vectors, depending on the color coordinates of the target sample and the adjusted color difference value. In order to compare the effect of the preliminarily preprocessing on the results, the mean and median values of cumulative variance of initial database for different values of ΔE were calculated and are shown in table 1. The table shows that, the cumulative information increases by the proposed method, considerably.

Since, the value of cumulative variance was adequate for three of the most important eigenvectors using the suggested method, the first three of basis components were selected for the recovery process and the goal was restricted to finding the tristimulus-constrained PC coordinates, denoted by C in the equations in this paper. Hence, a fully defined equation was used in the computational steps.

It is clear that the number of approved samples for any given specimen would depend on the applied color difference value. Figs. 2 and 3 show the distribution of the selected samples for different ΔE values in the a*b*

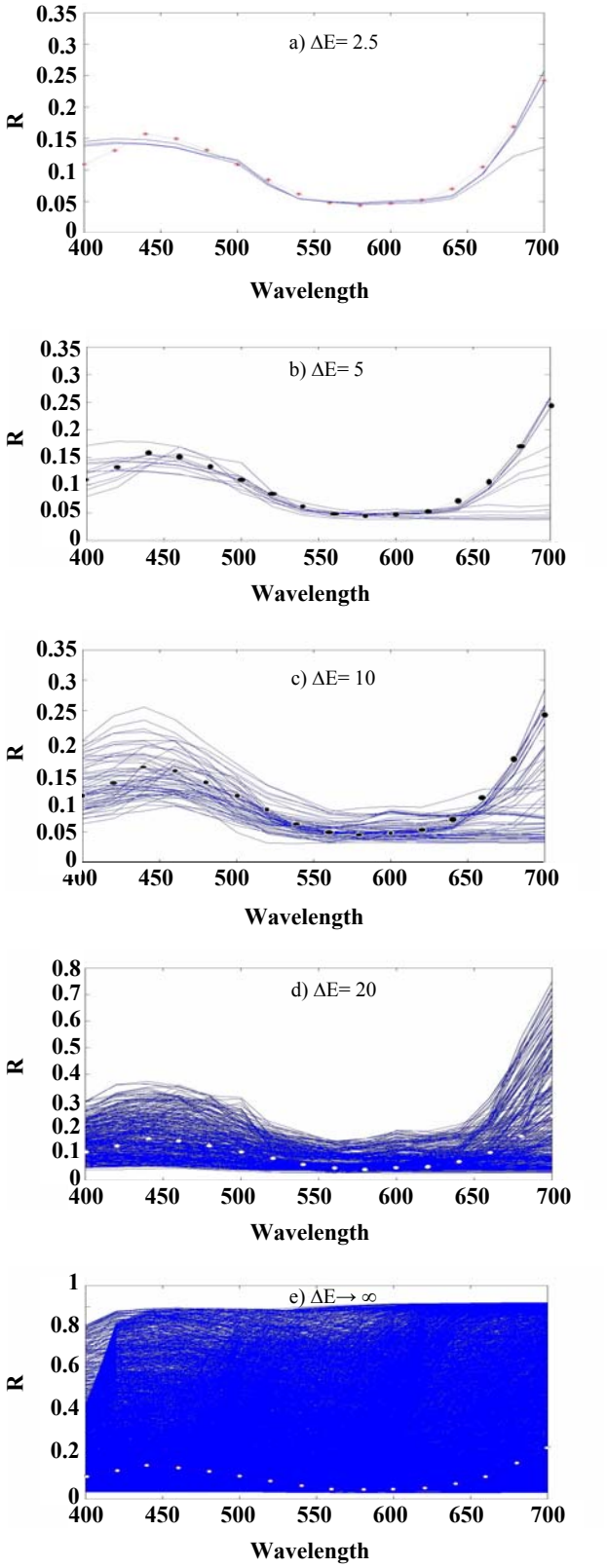

Fig. 3: The reflectance values of conformed samples for a randomly selected target in spectral domain as a function of selected ΔE value.

Table 1: The means and medians of the cumulative variances obtained from different values of sorting criteria, using three basis vectors.

Tolerable ΔE_{EE}	Cumulative variance					
	1		2		3	
	Mean	Median	Mean	Median	Mean	Median
2.5	99.01	99.79	99.97	99.99	99.99	100
5	98.26	99.40	99.85	99.93	99.96	99.99
10	96.49	98.43	99.40	99.55	99.75	99.82
13	95.89	97.94	99.14	99.33	99.62	99.71
No selection ($\Delta E \rightarrow \infty$)	86.13		95.19		98.37	

Table 2: Number of nonconforming and obtained samples as a function of CIELAB color difference value under equal energy illuminant and 10^o standard observer and the results of recovery of spectral data. A database consisted of 2802 was used to recover the spectral behavior of Acrylic samples.

Tolerable ΔE_{EE}	Nonconforming samples in targets	Mean of obtained samples from source	Recovery results						
			RMS	GFC	%GFC		ΔE_{10^o} (median)*		
			median*	median*	>0.95	>0.99	D_{65}	A	TL_{84}
2.5	1855	4	0	1	89.9	78.3	0	0	0
5.0	206	8	0.0029	0.9883	71.9	48.0	0.08	0.30	0.39
10.0	1	34	0.0106	0.9490	49.9	19.1	0.26	0.98	1.23
13.0	**0**	**63**	**0.0127**	**0.9318**	**42.8**	**14.7**	**0.31**	**1.22**	**1.41**
20	0	167	0.0155	0.9151	37.9	8.6	0.41	1.66	1.88
30	0	406	0.0180	0.9091	33.5	6.9	0.47	1.90	1.98
40	0	744	0.0157	0.9192	34.1	6.7	0.44	1.77	1.94
50	0	1160	0.0145	0.9228	35.0	7.7	0.42	1.75	1.74
No selection ($\Delta E \rightarrow \infty$)	0	2801	0.0154	0.9262	34.6	7.5	0.55	2.20	2.01

**Since the ΔE, RMS and GFC values are not normally distributed the median values are reported.*

and the reflectance domains for one randomly selected colored paint, respectively. As expected, while the number of selected samples increase with increasing ΔE value, the similarity of collected specimens decrease with increasing ΔE value which would lead to a rise in the actual dimensions of chosen dataset.

RESULTS AND DISCUSSION

Fig. 4 shows the results of the recovery of spectral reflectance for 8 randomly selected colored Acrylic paints obtained from the common and the suggested methods.

As table 2 shows the recovery procedure improved by using progressive dataset selection in accordance, to the proposed color and the adjusted ΔE value. The performance of suggested algorithm was evaluated by RMS and GFC (goodness-fitting coefficient). The best results were usually achieved when the ΔE value was small and the numbers of selected samples in the dataset were also adequate. However, as the table shows, in such circumstance, it is not possible to satisfy the condition of minimum number of samples for most of the specimens. Columns 2 and 3 of table 2 show the number of nonconforming samples and the mean of selected samples, respectively. As the results show, the number of

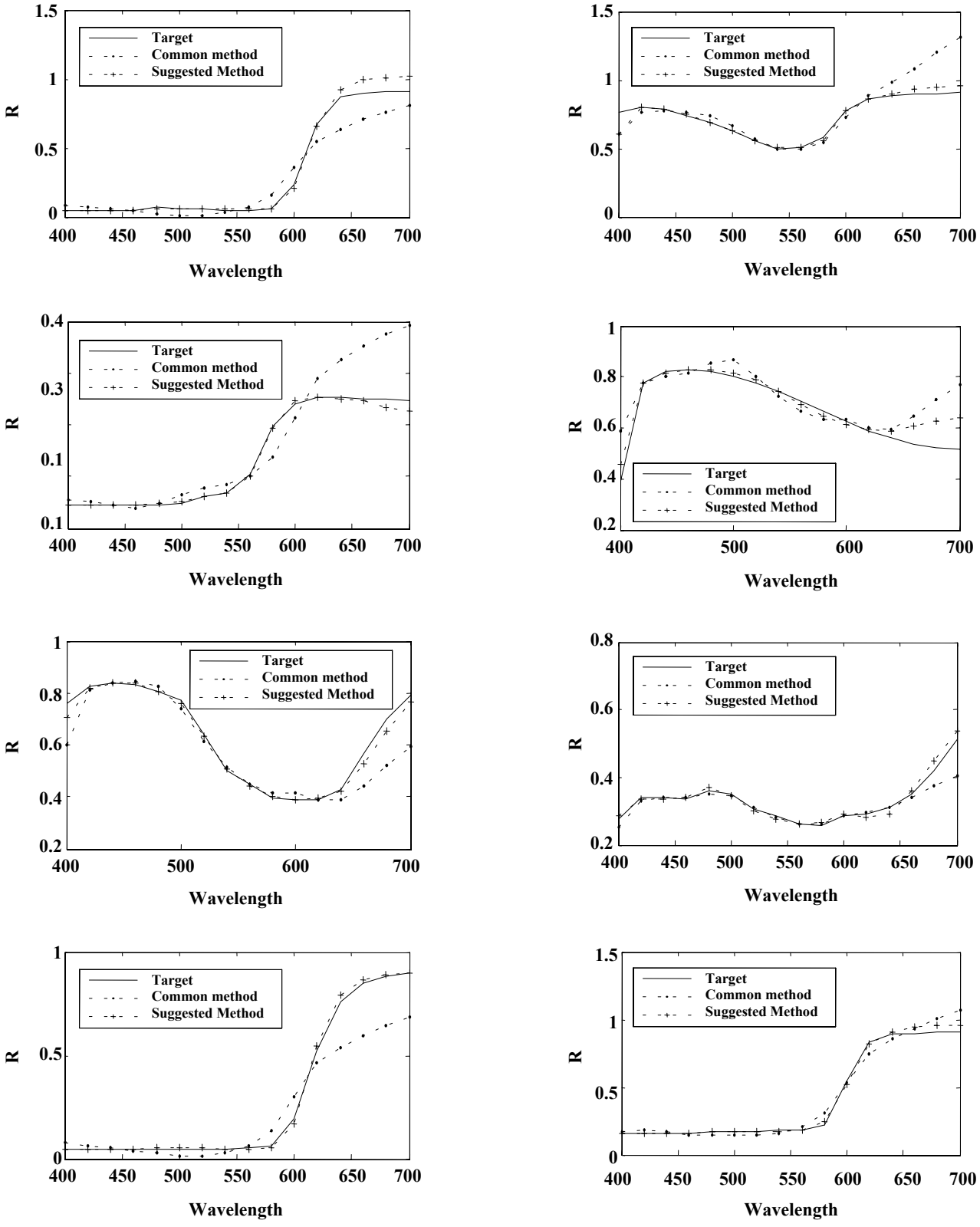

Fig. 4: Results of spectral recovery of some randomly selected Acrylic samples from different dataset by suggested and common method.

nonconforming samples decrease with increasing ΔE value, whilst the mean of conforming samples increasing rapidly. As table 2 shows, different ΔE values are needed for finding suitable datasets for all samples, demonstrated by bold fonts in the shadow row.

The reflectance curves in Fig. 4 conform to the same outcome. The figure shows the result of recovery for eight randomly selected Acrylic paint samples when a database containing 2802 Acrylic paints was used. The results of reconstruction of spectral data improve significantly by the suggested method.

The most important drawback of the spectral space is the availability of metamers [14]. In fact, separation of large psychophysical color metamers in spectral space could lead to some difficulties. Sorting samples according to their color differences could result the spectral reflectances which are not similar in the spectral domain. Generally, this is not a case in real metamers. The smooth spectral behaviors of object colors, such as paints, textile materials, prepare several crossovers which are distributed in the whole region of visual spectrum [15]. In the other words, the heavy metamerism is not a serious case in the paints and metameric pairs generally exhibit similar trends in the spectral domain and the proposed procedure does not direct to any difficulties in practice. In fact, the suggested technique pulls the suitable samples out of the large library which orders the samples at the beginning of the dimensional reduction step leading to a proper approximation with three basis functions.

CONCLUSIONS

An adaptive dataset of reflectance vectors was created in relation to the color specifications of given sample to extract the basis components. So, different datasets were formed for each target sample whose reflectance behavior was aimed after. Each dataset was consisted of the reflectance values of different samples. So, the dimensions of samples were preliminarily reduced before determining the basis functions. By this type of pre-selection, the mean of the cumulative variance increased for 3 basis functions considerably in comparison to the common method.

Finally, the spectral reflectances of colored samples were determined by using the standard routine. Results of recovery, obtained from the suggested and common method, were compared by GFC and RMS values of the reflectance curves as well as the color difference values under different illuminants. A significant improvement was observed by the new algorithm.

Acknowledgement

The authors thank Dr. Garcia-Beltran and his colleagues for supplying the spectral data of Acrylic paints.

Received : 13th July 2005 ; Accepted : 26th September 2005

REFERENCES

[1] Cohen, J., *Psychonomic Science*, **1**, 369 (1964).
[2] Maloney, L. T., *J. Opt. Soc. Am. A.*, **3**, 1673 (1986).
[3] Jaaskelainen, T., Parkkinen, J., and Toyooka, S., *J. Opt. Soc. Am. A.*, **7**, 725 (1990).
[4] Lenz, R., Österberg, M., Hiltunen, J., Jaaskelainen, T., and Parkkinen, J., *J. Opt. Soc. Am. A.*, **13**, 1315 (1996).
[5] Parkkinen, J., Hallikainen, J., and Jaaskelainen, T., *J. Opt. Soc. Am. A.*, **6,** 318 (1989).
[6] Dannemiller, J. L., *J. Opt. Soc. Am. A.*, **9**, 507 (1992).
[7] Vrhel, M. L., Gershon, R., and Iwan, L. S., *Col. Res. Appl.*, **19**, 4 (1994).
[8] Garcia-Beltran, A., Nieves, J. L., Hernandez-Andres, J., and Romero, J., *Col. Res Appl.*, **23**, 39 (1998).
[9] Maloney, L. T., and Wandell, B. A., *J. Opt. Soc. Am. A.*, **3,** 29 (1986).
[10] Romero, J., Garcia-Beltran, A., Jimenez-delBarco, L., and Garcia, J. A., *J. Optics.*, **23**, 41 (1992).
[11] Marimont, D. H., and Wandell, B. A., *J. Opt. Soc. Am. A.*, **9**, 1905-1913 (1992).
[12] Fairman, H. S., and Brill, M. H., *Col. Res. Appl.*, **29,** 104 (2004).
[13] Hardeberg, J. Y., 1st European Conference in Colour in Graphics, Image and Vision 480-485 (2002).
[14] Ramanath, R., Kuehni, R. G., Snyder, W. E., and Hinks, D., *Col. Res. Appl.*, **29,** 29 (2004).
[15] Choudhury, A., Roy, K., and Chatterjee, S. M., *Col. Res. Appl.*, **21,** 26 (1996).