Optica Applicata, Vol. XLII, No. 4, 2012
DOI: 10.5277/oa120422

Optimal physical primaries of spectral color information

SYAMAK FARAJIKHAH, FARAZ MADANCHI, SEYED HOSSEIN AMIRSHAHI*

Department of Textile Engineering, Amirkabir University of Technology, Tehran, Iran

*Corresponding author: hamirsha@aut.ac.ir

A new method for selection of optimal physical primaries is introduced. The reflectance spectra of 1269 matt Munsell color chips are used as a universal physical dataset and the most independent samples are extracted and used as actual primaries. The efficiency of selected primaries is compared with those obtained from principal component analysis (PCA), non-negative matrix factorization (NNMF) and non-linear principal component analysis (NLPCA) techniques. The performances of chosen primaries are evaluated by calculation of root mean square (RMS) errors, the goodness of fit coefficient (GFC) and the color difference (ΔE) values between the original and the reconstructed spectra.

Keywords: physical optimal primaries, data compression, data reconstruction.

1. Introduction

Reflectance spectra of non-fluorescent surfaces provide comprehensive information about the visual properties of objects in various viewing conditions. While the colorimetric data are expressed in three-dimensional (3D) spaces such as CIE XYZ color space, the spectral data are composed of the reflectance values of sample in the visible range of electromagnetic spectrum in predefined band length, *i.e.*, 400 to 700 nm at 10 nm intervals. Consequently, the size of spectral data is not comparable with colorimetric information and accordingly complete surface specification becomes accessible through the spectral communication. In fact, the spectral data is unique identification of object while, objects with different reflectance spectra can provide the same color under a given set of viewing condition that are known as metamer samples. There is no doubt that the spectral data suffer from huge sizes and some techniques have been presented to extract the main features of information to reduce the size [1, 2].

Several compression methods were tried to present the spectral data in reduced sizes. In fact, while the tristimulus values do not provide adequate information in some circumstances, the 31 dimensional behaviors of the reflectance spectra could contain some useless and/or minor information about the surfaces. Hence, some mathematical

techniques were utilized to reduce the dimensions of spectral reflectances by extracting the principal directions. One of the most applicable methods is the principal component analysis technique that is denoted by PCA [3–7].

TZENG and BERNS reviewed the PCA and its applications in the fields of colorant estimation, spectral data reduction, and defining multidimensional confidence regions for colorimetric scatter data [8]. FAIRMAN and BRILL [4] explained the application of classical PCA method for the compression of spectral reflectance as well as the reconstruction of spectra from the corresponding CIE XYZ tristimulus values. Different methods were also introduced to improve the efficiency of compression and the reconstruction techniques by choosing suitable sets in the learning step and/or weighing the samples prior to extraction of principal directions [5]. Recently, the application of non-linear version of PCA (NLPCA) was reported [9]. Opposed to classic PCA, the features extracted by NLPCA are not limited to the orthonormal vectors.

While the principal directions extracted by PCA were called "the statistical primaries" and contained both negative and positive values [10], the non-negative matrix factorization (NNMF) technique factorizes the data matrix into non-negative spectra and computes the more realistic, *i.e.*, all positive features of dataset. The method was employed by BUCHSBAUM and BLOCH in the field of color characterization to obtain the color names coincident with the established color naming categories [11]. AMIRSHAHI and AMIRSHAHI recently introduced the adaptive version of NNMF technique in reconstruction of spectral data from the corresponding CIE XYZ colorimetric information and compared the results with the classical and weighted versions of PCA [12]. They employed the spectral information of Munsell color chips and reported the priority of adaptive NNMF method over the classical one. While the efficiency of both types of NNMF methods was less than that of the corresponding PCA techniques, *i.e.*, the classical and weighted PCA, they claimed that the extracted primaries are closer to actual primaries in comparison to PCA due to their non-negative spectral behaviors, *i.e.*, physical availability [12].

In this study, we introduce a novel method for computing the more realistic primaries for compression and reconstruction of spectral data and compare the results to those obtained from PCA, NLPCA and NNMF methods. Compression process is performed by using different number of primaries, *i.e.*, 3 to 9 bases, and the effects of selected sizes on the reconstructed spectra are investigated. Two different spectral metrics, *i.e.*, root mean square (RMS) error and the goodness of fit coefficient (GFC) between the original and the reconstructed spectra are used to evaluate the performances of different methods. The CIELAB color difference values (ΔE), between the actual and synthesized spectra are also reported under D65 illuminant and 1964 standard observer.

2. Short theoretical background

Linear models are mostly used for expression of spectral data in limited spaces. The models uncover the underlying linear pattern and extract the main features of

employed dataset in different manners and provide the possibility to redefine the samples in the compressed spaces. The spectral data mostly appear in matrix form, *i.e.*, the columns of the matrix are the reflectance spectra of different samples and the rows form the measurement points. Different mathematical methods like PCA, NLPCA and NNMF approaches with diverse potentiality were suggested to extract the basic functions of datasets based on different strategies [13].

2.1. Principal component analysis

The PCA is undoubtedly the most conventional method for reducing spectral datasets to lower dimensions. The method finds the eigenvectors of the correlation or covariance matrices of dataset and orders them according to the corresponding eigenvalues from the largest to the lowest ones. Then, proportional to the expected accuracy, the suitable dimensions are chosen and the arranged eigenvectors, *i.e.*, $\mathbf{V}_i(\lambda)$ ($i = 1, 2, ..., k$) are considered as basic functions of desired dataset. The values of ε_i are then obtained by the orthogonal projection of spectral data over the extracted directions. Hence, the spectral reflectance of a given sample in dataset can be expressed by a simple linear model such that $R(\lambda) = \sum_{i=1}^{k} \varepsilon_i \mathbf{V}_i(\lambda)$. It was proved that the reflectance spectra of natural surface colors could be described by about 7 to 9 eigenvectors, while depending on the type of dataset, they could be reduced in the smaller dimensional spaces [14]. If covariance matrix is used instead of correlation matrix, the subtracted mean vector of dataset should be added during the recovery of spectral data, hence, $R(\lambda) = \sum_{i=1}^{k} \varepsilon_i \mathbf{V}_i(\lambda) + \mathbf{V}_0(\lambda)$.

2.2. Non-linear principal component analysis

Similar to classical PCA, the non-linear principal component analysis recognizes and removes correlation among variables for dimensional reduction, while opposed to PCA it is not limited to linear correlation among the variables [15].

Let R be an $m \times n$ matrix where m and n represent the number of observations and the number of variables, respectively. By using a non-linear vector function, the non-linear principal components are obtained as illustrated in the following equation:

$$T = \mathbf{G}(\mathbf{R}) \tag{1}$$

where $\mathbf{G}$ is the non-linear vector function composed of f non-linear functions $\mathbf{G} = \{G_1, G_2, \ldots, G_f\}$ and T stands for the non-linear principal components. For reconstruction process, the second non-linear vector function is employed and data is reconstructed as shown in the following equation:

$$\hat{R} = \mathbf{H}(\mathbf{T}) \tag{2}$$

where $\mathbf{H}$ is the non-linear vector function composed of m non-linear functions $\mathbf{H} = \{H_1, H_2, \ldots, H_m\}$ and $\hat{R}$ is the reconstructed spectral reflectance. By using an artificial neural network the functions G and H are selected in such a manner as to minimize the $\|R - \hat{R}\|$ value [15].

2.3. Non-negative matrix factorization

For a non-negative matrix like $R_{m\times n}$ spectral data, using non-negative matrix factorization method gets factorization into two non-negative factors basic vectors $\mathbf{W}_{m\times u}$, and coefficient vectors, $\mathbf{H}_{u\times n}$ (u which is called the rank of factorization is smaller than m and n, hence provides a reduced space for describing a dataset). In the spectral domain, the reflectance spectra could be reconstructed by the following simple linear model

$$\mathbf{R}_{m\times n}(\lambda) = \mathbf{W}_{m\times u}(\lambda)\,\mathbf{H}_{u\times n}(\lambda) \tag{3}$$

Since the basic functions of any linear models could resemble the primaries in a color reproducing system, the non-negative behaviors of them in spectral domain make them closer to real primaries hence provide more sensible primaries. In other words, non-negativity is very important issue in real world and many of datasets like spectral data could not be imagined with any negative member. Therefore, opposed the PCA, the non-negative matrix factorization technique does not extract the basis according to the statistical correlation of members of dataset and finds the non-negative basic functions which are extracted through an optimization effort [16].

2.4. Extraction of realistic optimal primaries of dataset

Employing the PCA, NLPCA and NNMF methods leads to primaries that could be physically unavailable. In fact, finding the primaries with negative reflectance spectra like those obtained from PCA and its non-linear version would be impossible and the synthesizing of optimized non-negative primaries could not be practically feasible. Hence, selection of the most efficient primaries from the available members of spectral dataset could be more practical. The most important advantage of such primaries is their physical accessibility in comparison to those obtained from other techniques.

In any color reproducing effort, the independency of primaries and the size of gamut that could be provided by them are the most important criteria. In this work, the independent primaries were selected from samples of dataset through the computation of condition numbers of all combinations of different members of dataset. The set that provided the minimum condition number benefited from the most independent members.

Assume that $\mathbf{R}$ is an $m \times n$ matrix where m and n represent the number of observations and the number of variables, respectively. To find the first pair of primaries with maximum independency, all binary combinations of samples were formed and the set with the minimum condition number (that was introduced as a criterion for independency of members of reflectance dataset [17]), was selected as the first pair of primaries. The following equation shows the idea employed

$$\min\left(\frac{\omega_{\max}\left([R_i \quad R_j]\right)}{\omega_{\min}\left([R_i \quad R_j]\right)}\right), \qquad i, j = 1, 2, \ldots, n, \qquad i \neq j \tag{4}$$

where $\omega_{\max}(R)$ and $\omega_{\min}(R)$ are the maximum and minimum singular values of matrix **R**; R_i and R_j are the i-th and j-th columns of matrix **R**.

Assume that $[R_i \quad R_j]$ are the first pair of primaries denoted by P. Consequently, as Eq. (5) shows, for selecting the third primary, all remaining members of dataset were individually checked in combination with the selected pair and again the arrangement which provided the minimum value was selected and the procedure was continued for selection of optimal primaries

$$\min\left(\frac{\omega_{\max}\left([P \quad R_t]\right)}{\omega_{\min}\left([P \quad R_t]\right)}\right), \qquad t = 1, 2, \ldots, n, \qquad t \neq i, j \tag{5}$$

By finding the k-th ($k < m$) primary, matrix V, consisting of the selected primaries was formed and considered as basic functions. It is clear that while the extracted primaries benefit from maximum independency they are not orthonormal as those are extracted by PCA. However, they are physically available and as Eq. (6) shows, simply could be used as projection space for reducing the dimensional sizes of spectral data

$$C = V^T R \tag{6}$$

where C demonstrates the specification of spectra in the reduced space, *i.e.*, the amount of primaries for spectral reproduction of samples, V shows the matrix including the selected primaries, R is the reflectance spectra and T indicates the matrix transpose.

Similar to linear models, the reconstruction of samples could be readily possible through the following equation:

$$\hat{R} \approx \sum_{j=1}^{k} C_j V_j \tag{7}$$

where $\hat{R}$ is the reconstructed data.

3. Experiment

In this study, an ensemble of 1269 reflectance spectra of color chips in the Munsell Book of Color–Matt Finish Collection were borrowed [18]. The spectral reflectances of samples were measured with Perkin Elmer Lambda 9 spectrophotometer and the wavelength range was from 380 to 800 nm with 1 nm interval. In this work, the reflectance data were fixed between 400 to 700 nm at 10 nm intervals [19].

To increase the efficiency of the linear model employed a kernel space, *i.e.*, the Kubelka–Munk function of reflectance spectra, was used. The following equation shows the transitional space employed

$$\frac{K}{S} = \frac{(1-R)^2}{2R} \tag{8}$$

Then, the most independent of desired database, *i.e.*, the optimal primaries, were extracted by employing the suggested approach. Different numbers of primaries were chosen and used for compression and reconstruction purposes in desired limited spaces. Finally, the recovered K/S spectra were converted to reflectance spectra by

$$\hat{R} = 1 + \frac{K}{S} - \left[\left(1 + \frac{K}{S}\right)^2 - 1\right]^{0.5} \tag{9}$$

where $\hat{R}$ is the reconstructed data.

The classical routines were used to extract the basis functions of desired methods and the extracted features were employed in definition of samples in reduced spaces and consequent reconstruction of spectral data.

The success of methods was evaluated by computing the RMS, GFC and ΔE values between the actual and the reconstructed spectra. The color difference values were computed under D65 standard illuminant and the CIE1964 standard observer and the GFC was calculated by using Eq. (10) and the results were evaluated as acceptable (GFC ≥ 0.9950), good (GFC ≥ 0.9990) and excellent (GFC ≥ 0.9999) [5]

$$\text{GFC} = \frac{\sum R(\lambda)\hat{R}(\lambda)}{\sqrt{\sum\left[R(\lambda)\right]^2}\sqrt{\sum\left[\hat{R}(\lambda)\right]^2}} \tag{10}$$

4. Results and discussion

Figure 1 shows the nine extracted primaries by PCA, NNMF and the proposed method. As plots in the figure show, while dipole positive–negative primaries were extracted by the PCA, the optimized primaries are all positive by NNMF and the methods suggested. In fact, PCA extracted the main features of dataset based on the statistical properties of dataset, while negative spectral reflectance would not be available in real-world tasks.

Figures 2 through 5 respectively show the results of spectral reconstruction by PCA, NLPCA, NNMF methods and the proposed technique for four randomly selected samples while 3 to 9 primaries have been used.

The variations of the mean values of RMSs against the numbers of employed primaries are also demonstrated in Fig. 6 for different methods.

As shown in Fig. 6, the mean value of RMS consistently decreases by increasing the number of primaries for PCA and suggested methods, while some types of fluctuations are evident for NNMF and NLPCA techniques. The figure also shows that the suggested method performs better than NNMF when the number of primaries is increased to 5 and still leads to better results than NLPCA when 7 primaries are employed.

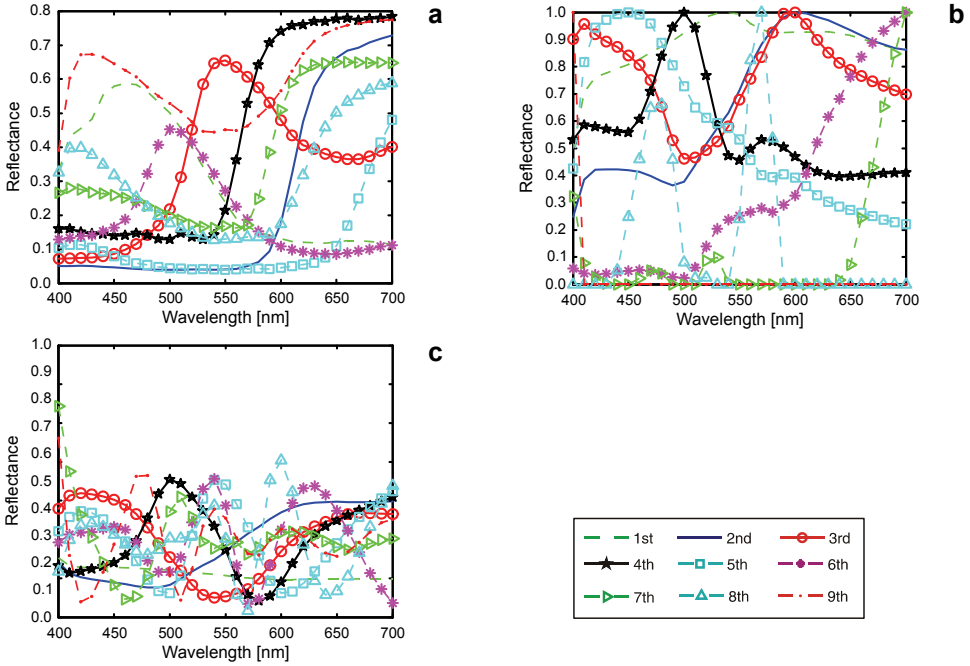

Fig. 1. Extracted primaries; proposed method (**a**), NNMF (**b**) and PCA (**c**).

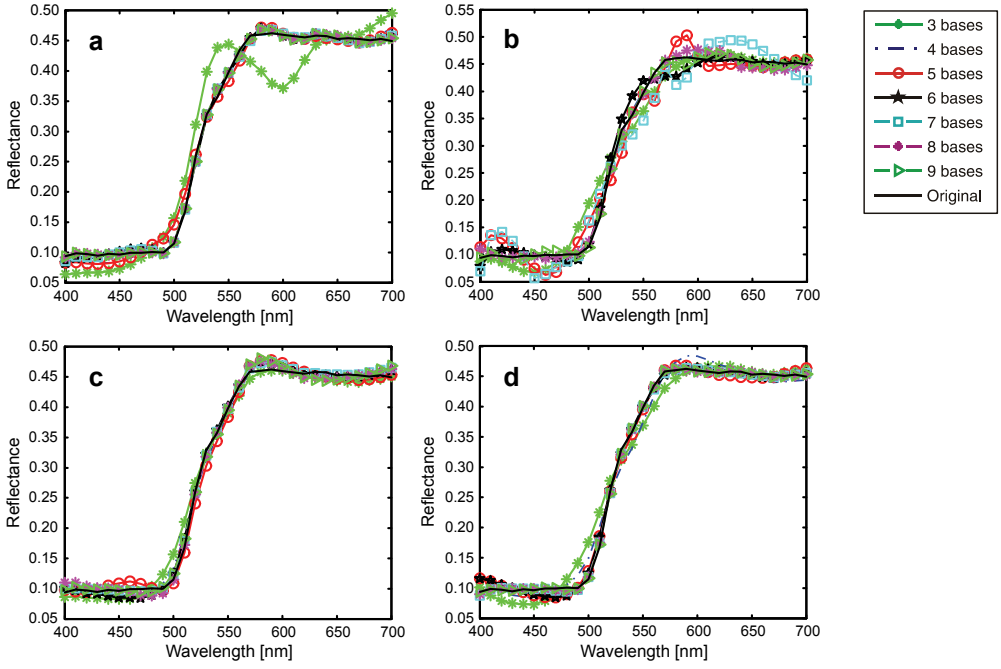

Fig. 2. Results of spectral reconstruction of sample #285 of Munsell chips; proposed method (**a**), NNMF (**b**), NLPCA (**c**) and PCA (**d**).

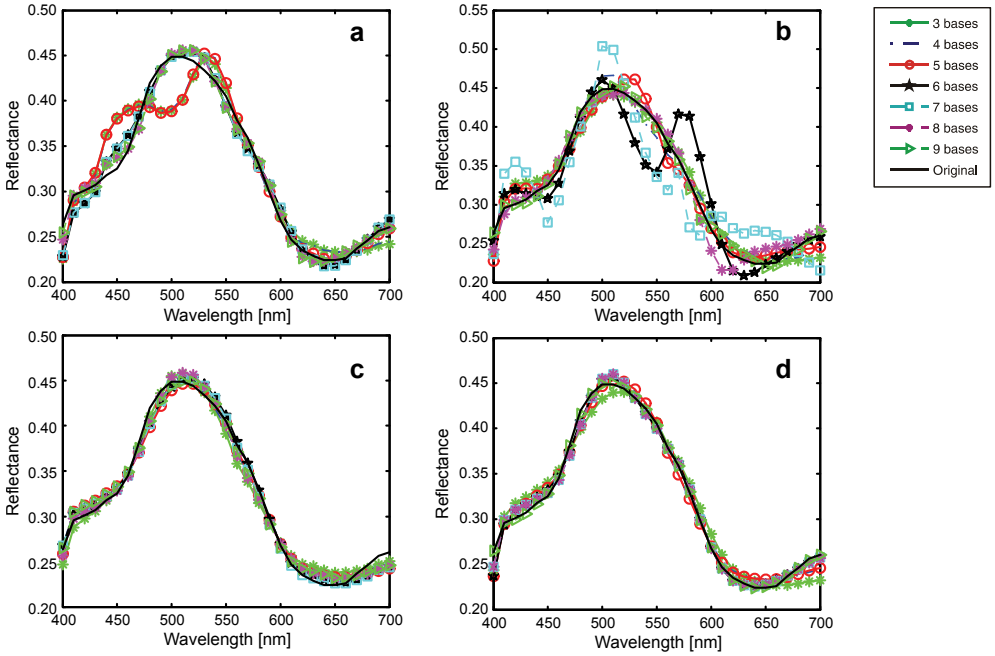

Fig. 3. Results of spectral reconstruction of sample #633 of Munsell chips; proposed method (**a**), NNMF (**b**), NLPCA (**c**) and PCA (**d**).

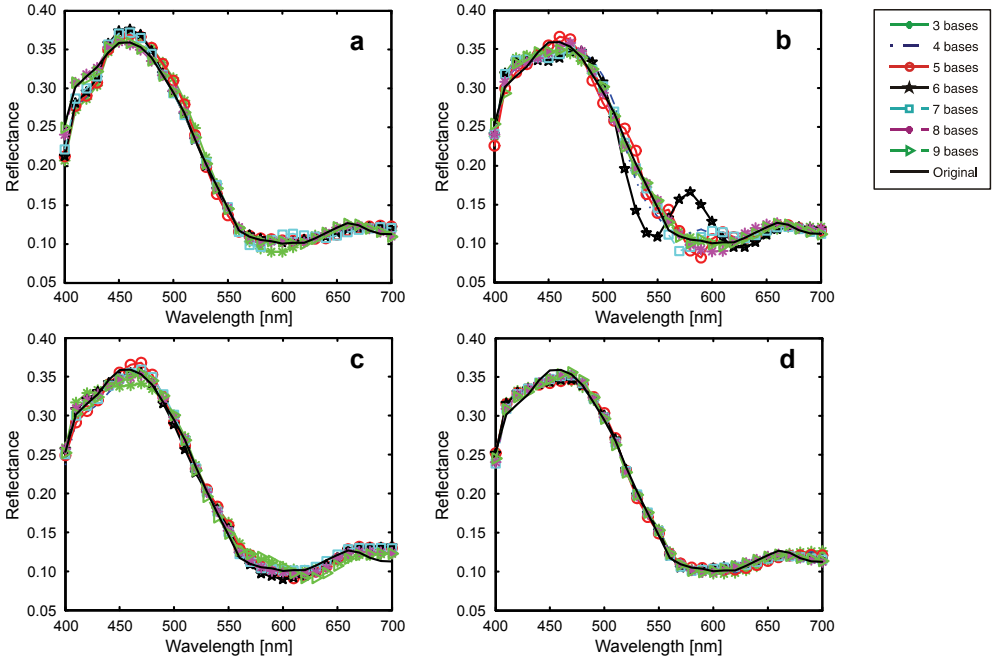

Fig. 4. Results of spectral reconstruction of sample #888 of Munsell chips; proposed method (**a**), NNMF (**b**), NLPCA (**c**) and PCA (**d**).

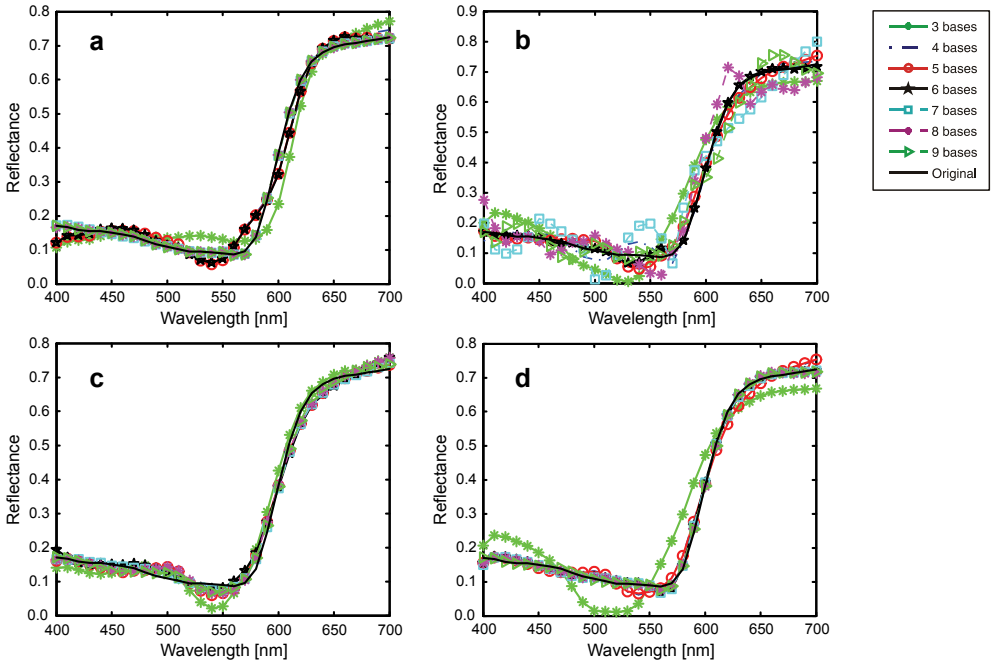

Fig. 5. Results of spectral reconstruction of sample #1268 of Munsell chips; proposed method (**a**), NNMF (**b**), NLPCA (**c**) and PCA (**d**).

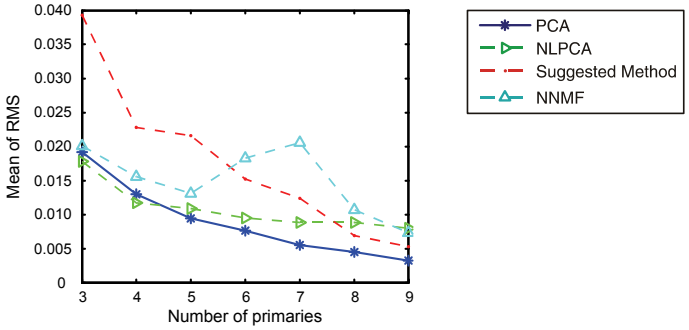

Fig. 6. The fluctuations of the mean value of RMSs against the number of chosen primaries.

Table 1 shows the mean of RMS values, the frequencies of different grades of GFC and the mean of color difference values for different numbers of primaries obtained by employing PCA, NLPCA, NNMF methods as well as those obtained by the technique proposed.

Again, by considering Tab. 1, it can be deduced that in both PCA and proposed method, the number of unacceptable synthesized samples constantly decreased and reached zero by increasing the number of employed primaries in contrast with two other methods in which the number of unacceptable synthesized samples show fluctuations by an increase in the number of primaries. On the other hand, the number

T a b l e 1. The spectral and colorimetric accuracies of spectral reconstruction by PCA, NLPCA, NNMF techniques and proposed method in different dimensional spaces.

		Errors						
		Colorimetric			Spectrophotometric			
	Number of primaries	ΔE			GFC			
		Mean	Max	RMS	Unacceptable	Acceptable	Good	Excellent
PCA	3	3.41	27.32	0.0192	327	600	337	5
	4	1.39	11.66	0.0130	146	421	661	41
	5	0.82	5.17	0.0094	53	329	806	81
	6	0.80	5.21	0.0076	33	225	888	123
	7	0.18	2.93	0.0055	5	148	699	417
	8	0.14	1.52	0.0045	2	119	559	589
	9	0.14	1.24	0.0032	1	20	438	810
NLPCA	3	3.65	39.03	0.0178	292	565	407	5
	4	2.02	15.59	0.0117	105	443	692	29
	5	1.98	24.38	0.0109	53	463	692	61
	6	1.51	20.38	0.0095	37	327	811	94
	7	1.25	13.57	0.0088	41	306	779	143
	8	0.94	6.06	0.0088	34	275	827	133
	9	1.22	17.07	0.0080	23	252	801	193
NNMF	3	3.31	27.50	0.0201	250	740	279	0
	4	1.70	12.54	0.0156	168	619	467	15
	5	0.82	7.37	0.0131	46	696	525	2
	6	2.81	10.95	0.0183	286	541	423	19
	7	1.54	12.49	0.0206	271	629	344	25
	8	1.15	17.39	0.0107	68	292	813	96
	9	0.33	2.54	0.0073	24	225	671	349
Proposed method	3	9.84	30.47	0.0393	891	363	11	4
	4	2.71	15.40	0.0228	365	741	154	9
	5	2.72	15.54	0.0216	338	758	162	11
	6	1.12	3.31	0.0152	107	711	435	16
	7	0.87	4.53	0.0124	43	605	566	55
	8	0.21	2.00	0.0069	4	57	1052	156
	9	0.17	1.65	0.0053	3	39	923	304

of excellent synthesized samples shows a consistent rising trend in PCA as well as the proposed method by an increase in the number of primaries, whereas such trend in NLPCA and NNMF is not evident and widely fluctuated. As shown in Tab. 1, similar results were totally obtained from colorimetric evaluations. In fact, the color difference values are higher for the method proposed when a limited number of primaries, *e.g.*, 3 or 4, were used but the results are completely different for more than 5 primaries.

5. Conclusions

A novel method for selection of physical optimal primaries was introduced. The efficiency of extracted primaries was compared with those obtained from the most applicable types of linear models, *i.e.*, classical PCA and non-negative matrix factorization methods as well as non-linear PCA approach. To compare the performance of the method proposed with those of PCA, NLPCA and NNMF techniques, the colorimetric differences, *i.e.*, ΔE, and the spectrophotometric errors, *i.e.*, RMS and GFC values were calculated between the original and synthesized spectra. Based on the results, it can be deduced that the proposed method generally performed better than NLPCA and NNMF techniques and the physical and realistic aspects of primaries provided the outstanding advantage among all extracted bases computed by other techniques.

References

[1] HAWKYARD C.J., *Synthetic reflectance curves by subtractive colour mixing*, Journal of the Society of Dyers and Colourists **109**(7–8), 1993, pp. 246–251.

[2] ESLAHI N., AMIRSHAHI S.H., AGAHIAN F., *Recovery of spectral data using weighted canonical correlation regression*, Optical Review **16**(3), 2009, pp. 296–303.

[3] COHEN J.B., *Dependency of the spectral reflectance curves of the Munsell color chips*, Psychonomic Science **1**, 1964, pp. 369–370.

[4] FAIRMAN H.S., BRILL M.H., *The principal components of reflectances*, Color Research and Application **29**(2), 2004, pp. 104–110.

[5] GARCÍA-BELTRÁN A., NIEVES J.L., HERNÁNDEZ-ANDRÉS J., ROMERO J., *Linear bases for spectral reflectance functions of acrylic paints*, Color Research and Application **23**(1), 1998, pp. 39–45.

[6] MALONEY L.T., *Evaluation of linear models of surface spectral reflectance with small numbers of parameters*, Journal of the Optical Society of America A **3**(10), 1986, pp. 1673–1683.

[7] AGAHIAN F., AMIRSHAHI S.A., AMIRSHAHI S.H., *Reconstruction of reflectance spectra using weighted principal component analysis*, Color Research and Application **33**(5), 2008, pp. 360–371.

[8] DI-YUAN TZENG, BERNS R.S., *A review of principal component analysis and its applications to color technology*, Color Research and Application **30**(2), 2005, pp. 84–98.

[9] DEL FRATE F., SCHIAVON G., *Nonlinear principal component analysis for the radiometric inversion of atmospheric profile by using neural networks*, IEEE Transactions on Geoscience and Remote Sensing **37**, 1999, p. 2335.

[10] BERNS R.S., *A generic approach to color modeling*, Color Research and Application **22**(5), 1997, pp. 318–325.

[11] BUCHSBAUM G., BLOCH O., *Color categories revealed by non-negative matrix factorization of Munsell color spectra*, Vision Research **42**(5), 2002, pp. 559–563.

[12] AMIRSHAHI S.H., AMIRSHAHI S.A., *Adaptive non-negative bases for reconstruction of spectral data from colorimetric information*, Optical Review **17**(6), 2010, pp. 562–569.

[13] WESTLAND S., RIPAMONTI C., *Computational Color Science Using MATLAB*, Wiley, West Sussex, 2004.

[14] MALONEY L.T., WANDELL B.A., *Color constancy: a method for recovering surface spectral reflectance*, Journal of the Optical Society of America A **3**(1), 1986, pp. 29–33.

[15] KRAMER M.A., *Nonlinear principal component analysis using autoassociative neural networks*, AIChE Journal **37**(2), 1991, pp. 233–243.
[16] CICHOCKI A., ZDUNEK R., ANH HUY PHAN, SHUN-ICHI AMARI, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separations*, Wiley, New York, 2009.
[17] HARDEBERG J.Y., *Acquisition and reproduction of colour images: colorimetric and multispectral approaches*, PhD Thesis, École Nationale Supérieure des Télécommunications, 1999.
[18] Spectral Database, University of Joensuu Color Group, available at: http://spectral.joensuu.fi/
[19] KOHONEN O., PARKKINEN J., JÄÄSKELÄINEN T., *Databases for spectral color science*, Color Research and Application **31**(5), 2006, pp. 381–390.

Received July 24, 2011