Principal Component Analysis of Spectral Images Based on the Independence of the Color Matching Function Vectors

Nobuhito Matsushiro*, Francisco Hideki Imai and Noboru Ohta
Munsell Color Science Laboratory, Center for Imaging Science
Rochester Institute of Technology, 54 Lomb Memorial Drive
Rochester, New York 14623-5604.
*Visiting Scientist from Oki Data Corporation, Japan.

Abstract

We propose a new method for data compression of images generated by multi-channel visible spectra imaging (MVSI) using principal component analysis based on the independence of the color matching function vectors. In the proposed method, resamplings of spectral values are performed considering the independence of the color matching function vectors. Our experiments showed that the method is effective in reducing the redundancy of MVSI images.

1 Introduction

Image compression for multispectral images have been explored in the remote-sensing field [1] but it was just recently been also an issue in the multi-channel visible spectra imaging (MVSI). For simplification we are going to refer MVSI images as mutispectral images. Multispectral images even for the visible spectra involves imaging with more than the traditional three channels and problems arise with respect to processing time, transmission and storage due to the huge amounts of data generated. Therefore, image compression methods need to be considered for MVSI images in practical applications. MVSI image compression has been emerging as a relevant field of research [2-5]. In these researches, principal component analysis (PCA) has been used successfully for data compression method for spectral images. Research at Chiba University, Japan, presents a method using PCA and discrete cosine transformation for spectral images followed by an analysis of the image quality of the compressed images [1]. A group at University of Joensuu, Finland has been used JPEG and MPEG-types subsampling for image compression based on PCA [3,5] as well as independent component analysis [5]. At Munsell Color Science Laboratory we started a preliminary study on multispactral image compression using arithmetic coding [4].

All methods mentioned before performs image compression without considering the response of human visual system. In this article we propose a new data compression using PCA based on the independence of the color matching function vectors. The assumption of the proposed method is that spectral values are weighted by the color matching functions at the final human vision stage. In the proposed method, resamplings of spectral values are performed considering the independence and dependence of the color matching function vectors, and redundancy is reduced in the data compression method.

2 Proposed Method

The independent color matching function vectors,

$$VI = \{VI(i) = (\bar{x}I(i), \bar{y}I(i), \bar{z}I(i)) : i = 1,2,\Lambda, sn\}, \quad (1)$$

are defined as the smallest number (sn) of vectors which represent all of the color matching function vectors $\{(\bar{x}(j), \bar{y}(j), \bar{z}(j)) : j = 1,2,\Lambda, n\}$ by linear combinations of vectors of VI. Where n is the wavelength number of samples, i, j indicate wavelength indices from the smallest wavelength to the largest wavelength. It is assumed that the coefficients of the linear combinations are only positive. The assumption has been derived because spectral values do not take negative values. The definition of the independent vectors is with this assumption, though without the assumption, the maximum number of independent vectors are three.

The dependent color matching function vectors,

$$VD(j_k), \quad (k = 1,2,\ldots,kn), \tag{2}$$

are represented using independent vectors,

$$VI(i_h), \quad \left(h = 1,2,\ldots,sn,\ i_h \neq j_k,\ i_h \in \{j : j = 1,2,\ldots,n\}\right), \tag{3}$$

as follows:

$$VD(j_k) = \sum_h \alpha(k,i_h) \cdot VI(i_h), \quad (k = 1,2,\ldots,kn), \tag{4}$$

where

j_k : indices for kn numbers of dependent vectors as shown in the diagram of figure 1,

i_h : indices for sn numbers of independent vectors as shown in the diagram of figure 2,

$\alpha(k,i_h)$: positive weight for i_h independent vector.

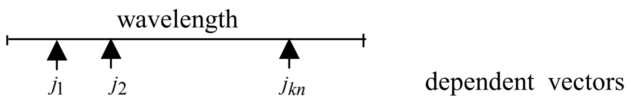

Figure 1

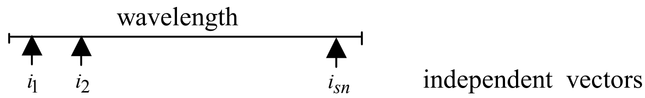

Figure 2

The spectral values $S(j_k)$, $(k = 1,2,\ldots,kn)$ with same wavelength as the dependent vectors $VD(j_k)$, $(k = 1,2,\ldots,kn)$ can be taken into the coefficients on the independent vectors of eq.(3). The spectral values $S(j_k)$, $(k = 1,2,\ldots,kn)$ have influence on all independent vectors of eq.(3), and using the relation of eq.(4), this is formulated as follows:

$$S^*(i_h) = \sum_k \alpha(k,i_h) S(j_k) + S(i_h), \quad (h = 1,2,\ldots,sn) \tag{5}$$

where $S^*(i_h)$, $(h = 1,2,\ldots,sn)$ indicate resampled spectral values. The coefficients correspond to the total effect of dependent vectors on the independent vectors of eq.(3).

After this vectorial process, the PCA is applied to the resampled spectral data S^*. $S^*(i_h)$, $(h = 1,2,\ldots,sn)$ has different scales in comparison with the original sampled spectral values, and scale factors ω_h, $(h = 1,2,\ldots,sn)$ are applied to $S^*(i_h)$, $(h = 1,2,\ldots,sn)$. The scaling operations are performed from the view point of the optimization of the PCA. The scaling factors are optimized minimizing the distributions of the first principal component. In the decoding process of spectral data from the principal components, the inverse scaling factors $1/\omega_h$, $(h = 1,2,\ldots,sn)$ are applied. These scaling factors are shared by both the coding processor and the decoding processor, and will be included in a coding format. The proposed method is visually lossless coding in the meaning that all of spectral information is taken into the wavelength of the independent color matching function vectors. The overhead of the scaling factors is relatively decreased with the increase of the number of data.

3 Experiments

Let consider the case that the wavelength range is 400nm to 700nm, and sampled in 10nm intervals (31 sampling points), and the color matching functions are the CIE1931.

Based on analysis of the color matching function vectors, the wavelength of the independent color matching function vectors are the following:
400nm, 410nm, 420nm, 430nm, 450nm, 460nm, 470nm, 480nm, 490nm, 500nm, 510nm, 520nm, 530nm, 540nm, 550nm, 560nm, 570nm, 630nm, 700nm.
And the wavelength of the dependent color matching function vectors are the following:
440nm, 580nm, 590nm, 600nm, 610nm, 620nm, 640nm, 650nm, 660nm, 670nm, 680nm, 690nm.

The spectral data employed in the experiments are data 1 (poster color ink), data 2 (various objects reflectance). The data 1 consists of 105 reflectances, and the data 2 consists of 170 reflectances. The data 1 has been created in Munsell Color Science Laboratory, and the data 2 has been retrieved from North Carolina State University ftp site [6].

Tables 1.1, 1.2, 2.1, 2.2 show the experimental results of contribution ratio. Table 1.1, Table 1.2 are for the data 1, and Table 2.1, Table 2.2 are for the data 2. Table 1.1, Table 2.1 are the results using the conventional PCA, and Table 1.2, 2.2 are the results using the new proposed PCA. In these tables, NPC indicates the number of principal components (in order of contribution to the reconstruction), and CR indicates the contribution ratio.

Table 1.1 Results of conventional PCA for data 1.

NPC	1	2	3	4	5	6
CR	0.7341	0.9249	0.9544	0.9843	0.9909	0.9963

Table 1.2 Results of new PCA for data 1.

NPC	1	2	3	4	5	6
CR	0.7638	0.9388	0.9771	0.9939	0.9971	0.9990

Table 2.1 Results of conventional PCA for data 2.

NPC	1	2	3	4	5	6
CR	0.8031	0.9191	0.9555	0.9768	0.9891	0.9932

Table 2.2 Results of new PCA for data 2.

NPC	1	2	3	4	5	6
CR	0.8489	0.9377	0.9628	0.9894	0.9951	0.9982

These results show that in the new PCA method, one or two principal components can be reduced keeping the contribution ratio almost the same, in comparison with the conventional PCA. Consequently, 16.6% to 33.2% amount of data is reduced in the new PCA data compression with visually lossless in comparison with the conventional PCA data compression. These results show the validity of the new method. The performance of the new PCA method can be explained by the fact that when we reduce the samplings from 31 to 19 also reduce the dimension of spectral variation that is not relevant for human visual system.

4 Conclusions

A new PCA based data compression method for spectral images has been described. The assumption of the proposed method is that spectral values are weighted by the color matching functions at the final human vision stage. In the new method, resampling of spectral values are performed considering the independence of the color matching function vectors.

The experimental results have shown that one or two components are reduced keeping the contribution ratio almost the same, in comparison with the conventional PCA. Consequently, 16.6% to 33.2% amount of data is reduced in the new PCA data compression with visually lossless, in comparison with the conventional PCA data compression method. These results have shown the validity of the new method. We are going to perform further analysis to evaluate the possibilities of this new method.

References

1 A. Kaarna, P. Zemcik, H. Kalvainen and Jussi Parkkinen, Compression of multispectral remote sensing images using clustering and spectral reduction, IEEE Transactions on Geoscience and Remote Sensing 38, 1037-1082 (2000).

2 W. Kondou, K. Miyata, H. Haneishi and Y. Miyake, An evaluation of image quality for compressed multi-spectral image, Proc. of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, Chiba University, Japan, 143-146, 1999.

3 J. Parkkinen, M. Hauta-Kasari, A, Kaarna, J. Lehtonen, P. Koponen and T. Jaaskelainen, Multispectral Image Compression, Proc. of the Second International Symposium on Multispectral Imaging and High Accuate Color Reproduction, 2000, pp.49-58.

4 F. H. Imai, M. R. Rosen, N. Ohta and N. Matsushiro, Preliminaly study on spectral image compression, Proc. of Color Forum Japan 2000, pp.67-70.

5 M. Hauta-Kasari, J. Lehtonen, J. Parkkinen and T. Jaaskelainen, Spectral image compression for data communications, Proc. of SPIE Vol.4300, 2001, pp.42-49.

6 ftp at ftp.eos.ncsu.edu (/pub/spectra) at North Carolina State University.