Retrieval from a Spectral Image Database by Reconstructed Spectral Images and the Combinations of Inner Product Images

Oili Kohonen and Markku Hauta-Kasari Department of Computer Science and Statistics, P.O.Box 111, FIN-80101 University of Joensuu, Finland

Abstract

A possibility to reduce the dimensionality of the training data in retrieval from a spectral image database by representing spectral images as the combinations of inner product images is examined. Moreover, the number of eigenvectors needed to reconstruct spectral images without significant percentage of classification error is studied. The experiments are performed by using a real spectral image database.

Introduction

The development of spectral imaging systems increases the amount of available spectral images, and the conceivable problems with different data formats will be tackled by the ongoing process on spectral image standardization [1]. Furthermore, these two aspects lead to the increase in the number and the size of available spectral image databases [2-4].

A technique for querying a spectral image database using both color and texture features was proposed in ref. [5]. There color spectra were classified into classes whose representatives were generated by selforganizing map (SOM) [6] algorithm and a color histogram over these classes was generated. Texture features for spectral images were calculated from the first inner product images generated by principal component analysis (PCA) [7]. The querying method was proved to be useful but the memory consumption during the training phase of SOM restricted the amount of the used training data. Training data should describe the spectral properties of the database properly and therefore the reasonable amount of spectra from each spectral image in a database should be included. However, when the amount of spectral images in a database rises high enough the amount of chosen spectra must be strongly restricted or the dimensionality of spectra must be reduced. If the number of spectra chosen from each image in a database is very small, the effect of randomness will be emphasized. Thus, the straightforward reduction of the amount of chosen spectra cannot be any farreaching solution.

In this paper, a possibility to reduce the dimensionality of the training data and perform successful queries by representing the spectral images as the combinations of inner product images is examined. Furthermore, the amount of eigenvectors needed to reconstruct spectral images without significant percentage of misclassification of spectra is studied.

Spectral image reconstruction

A dimensionality-reduced image S consisting of inner product images is defined as

$$S(i, j, k) = I_k(i, j) \tag{1}$$

in which I_k is the k^{th} inner product image of the image I, i and j correspond the location in spatial

domain and k is the location in spectral domain. The k^{th} inner product image is got by multiplying a spectral image by its k^{th} eigenvector which is calculated from a spectral image by PCA.

An example of the first three eigenvectors and the corresponding inner product images calculated from a spectral image are shown in Fig. 1.



Figure 1. The first three eigenvectors and the corresponding inner product images of a spectral image.

Another method for spectral image reconstruction is defined as

$$\tilde{I}(i, j, :) = \sum_{n=1}^{m} v_n^T I(i, j, k) v_n$$
 (2)

in which v_n is n^{th} eigenvector of the original spectral image I. When m equals to the number of spectral dimension in the original spectral image I, there is no difference between the original spectral image and the reconstructed spectral image.

Retrieval from a spectral image database

In the existing retrieval method the spectral images are represented by feature histograms which consist of both color and texture features. Feature histogram is generated for each spectral image in a database and the retrieval is based on dissimilarity calculation between histograms corresponding to a query image and the images in a database. Color histogram for a spectral image is generated by classifying the spectra of spectral image into the classes whose representatives are the weight vectors of a selforganizing map trained by the data collected from a spectral image database. The classification is performed in respect to the Euclidean distance between class representatives and a spectrum to be classified. After classifying each spectrum of a spectral image, a histogram over the classes will be established and normalized by the number of spectra in the image. Texture features have been calculated in ref. [5] from the first inner product images by the methods based on both co-occurrence matrix [8] and local binary pattern (LBP) [9], and there the use of LBP-based method gave better query results. The spectral reconstruction by using Eq. 1 has no effect on texture features because the first channel of reconstructed spectral image is represented by the first inner product image, from which the texture features can still be calculated. A diagram of retrieval from a spectral image database is shown in Fig. 2.



Figure 2. Retrieval from a spectral image database by using color features.

Experiments

Experiments were performed on a spectral image database consisting of 147 spectral images. The images in the database were transformed into the same format, each image containing 59 components from 410 nm to 700 nm with the interval of 5 nm. The images have been measured in the University of Joensuu (Finland) [10], Chiba University (Japan) [11], Marine Biological Laboratory of Maryland (United States) [12] and University of East Anglia (United Kingdom) [13]. In addition to real spectral images there exist also 24 synthetic images created by virtual coloring technique [14]. The spectral image database is shown in an RGB-format in Fig. 2.



In the first part of experiments the spectral images were reconstructed by Eq. 1 and represented as the combinations of inner product images. The retrieval experiments were performed for one by seven inner product images and using one to four inner product images was noticed to be reasonable. However, in most cases the best results were achieved by using three inner product images. The inner product images generated by using eigenvectors higher than five were generally noisy. Therefore, an increase in the number of used inner product images will weaken the retrieval results after a certain limit. An example of retrieved images by using both the original spectral image and spectral images represented by the combinations of inner product images is shown in Fig. 4. If the query image is included in the database it will always be the first of the retrieved images. However, to save the space in horizontal plane we have removed the query image from the retrieved images in this case. The query image is shown on the top of the image and then the retrieved images by using original and reconstructed images are shown row by row. The number of retrieved images is restricted to seven, the number not including the query image. Furthermore, the lowest dissimilarity is between the query image and the leftmost retrieved image.

The retrieval results shown in Fig. 4 are based on the color features only, whereas in Fig. 5 both color and LBP-based texture features are applied. The effect of noisy inner product images appears when the number of inner product images increases. This can be seen very clearly in Fig. 4 in the case of seven used inner product images. When LBP-texture features are combined with color features, the effect is noticeable but not that powerful. According to subjective evaluation, the most semantically meaningful outputs for retrieval in the case of reconstructed images are got by using the first three inner product images.

Figure 3. The used spectral image database in an RGB-format



Figure 4.. Retrieved images by using original spectral image (the first row) and the reconstructed spectral images. The used features are color features.

More examples of color-based retrieval using both the original and reconstructed spectral images consisting of three inner product images are shown in Fig. 6. In these cases the retrieval is performed by using color features only.

In the second part of the experiments spectral images were reconstructed by using Eq. 2 and the percentage of misclassified spectra in the case of reconstructed images was examined. Both the eigenvectors and percentage of misclassified spectra were calculated separately for each image. The maximum and average percentages of misclassification over the images for the first one to ten eigenvectors are shown in Fig. 7. One needs to notice that the scaling in the left diagram differs from the one in the right. The higher the number of used eigenvectors, the lower the percentage of misclassified spectra. The limit of one percent misclassification is achieved by seven and four eigenvectors when considering maximum and average percentages, respectively. Moreover, the effect of spectral misclassification on spectral image retrieval was examined as follows.



Figure 5. Retrieved images by using original spectral image (the first row) and the reconstructed spectral images. The features are combined color and LBP-texture features.

Image retrieval was first performed by using original spectral image and the number of ordered retrieved images was restricted to ten. This output was then used as a reference for the outputs achieved by using reconstructed images. Next, the spectral image reconstruction was performed by using the first eigenvector. The reconstructed image was used as a query image and the ordered output was compared to

the reference output. If the two outputs were not the same the number of eigenvectors in a reconstruction was increased by one, the image retrieval was performed again and the achieved output was compared to the reference output again. This process was continued until the two outputs were identical. The described process was performed for each image in the database. A histogram over the number of necessary eigenvectors is shown in Fig. 8, the average and the median being 3.4 and 3, respectively.



Figure 6.. Retrieved images by using original spectral image and the first three inner product images. The used features are color features.

Discussion

A method for achieving a reduction of dimensionality in the case of retrieval from a spectral image database is proposed. The proposed method is based on the idea of representing the spectral images as the combination



Figure 7. The maximum and average misclassification



Figure 8. A histogram over the number of eigenvectors needed to achieve the output similar to the reference output in the case of the first ten ordered output images.

of inner product images generated by principal component analysis. The method is tested by performing queries in a real spectral image database by using both original spectral images and their reconstructions. The results are compared and evaluated by subjective evaluation. The sufficient results are achieved by using the first one to four inner product images, and generally the use of the first three inner product images gives the best output. The results are promising. According to performance of the experiments, using the first three inner product images might be enough to achieve a rough approximation on image properties for certain image retrieval and browsing purposes such as the one discussed in ref. [15]. Furthermore, in these cases the techniques developed for retrieving RGB-images might be applied.

In addition, the percentage of misclassified spectra and the effect of misclassification on retrieval from a spectral image database in the cases of reconstructed spectral images are examined. The reconstruction is based on eigenvectors instead of inner product images. The limit of one percentage misclassification is achieved for all images by using seven eigenvectors. An identical retrieval output containing ten images was accomplished for the both original and reconstructed images by using twelve eigenvectors. However, the use of one to four eigenvectors was enough for 83% of images in a database.

Acknowledgements

The authors thank associate professor Yoshitsugu Manabe for suggesting to examine the use of the spectral images reconstructed by inner product images.

References

- T. Jetsu, P. Hertzog, T. Jaaskelainen and J. Parkkinen, Standardization of Spectral Image Formats, Pattern Recognition and Image Analysis, 15(3), 618-620, (2005).
- M. Rosen. Lippmann2000: A spectral image database under construction, International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, Chiba, Japan, 117-122, (1999).
- S. Hordley, G. Finlayson and P. Morovic, A Multi-Spectral Image Database and Its Application to Image Rendering across Illumination, Third International Conference on Image and Graphics (ICIG), Hong Kong, 394-397, (2004).
- H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi and M. Abidi, An Indoor and Outdoor Multimodal, Multispectral and Multi-Illuminant Database for Face Recognition, Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), New York, USA, (2006).
- O. Kohonen and M. Hauta-Kasari, Importance of the Texture Features in a Query from a Spectral Image Database, Eight International

Symposium on Multispectral Color Science, Electronic Imaging 2006, San Jose, California, USA, (2006).

- T. Kohonen, Self-Organizing Maps, Springer, Berlin Heidelberg, (2001).
- I. T. Jolliffe, Principal Component Analysis, Springer, New York, (1986).
- R. M. Haralick, K. Shanmugam and I. Dinstein, Textural Features for Image Classification, IEEE Trans Systems, Man, and Cybernetics, 3, 610-621, (1973).
- T. Ojala, M. Pietikäinen and D.Harwood, A comparative study of texture measures with classification based on feature distributions, Pattern Recognition, 29, 51-59, (1996).
- 10. Joensuu spectral image database http://www.spectral.joensuu.fi
- T. Fujimaki, K. Ishii, N. Tsumura and Y. Miyake, Proposals of Standard Spectral Image and Its Application to Designing of CCD camera, Fifth International Symposium on Multispectral Color Science, Digital Photography Conference (PICS) Rochester, USA, 496-499, (2003).
- C.-C. Chiao, T. W Cronin, and D. Osorio, Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants, J. Opt. Soc. Am. A, 2/17, 218-224, (2000).
- G. D. Finlayson, S. D. Hordley, P. Morovic, Using the SpectraCube to Build a Multispectral Image Database, Second European Conference on Colour in Graphics, Imaging and Vision (CGIV 2004), Aachen, Germany, 268-273, (2004).
- V. Botchko, S. Nakauchi, J. Parkkinen, H. Kälviäinen, Multispectral texture derivation in virtual coloring, in M. Pietikäinen (Ed.) Texture Analysis in Machine Vision, World Scientific Publishing Co., Singapore, (2000).
- M. Hauta-Kasari, J. Lehtonen and J. Parkkinen, Image Format for Spectral Image Browsing, Journal of Imaging Science and Technology, 50(6), 572-582, (2006).