Modified Principal Component Analysis (MPCA) for Feature Selection of Hyperspectral Imagery

Cheng Wang, M.Menenti, Zhao-Liang Li
ENSPS-LSIIT/GRTR, University Louis Pasteur
ILLKIRCH 67400, France

Abstract- Principal Component Analysis (PCA) is a classical multivariate data analysis method that is useful in linear feature extraction and data compression. It can compress the most information in the original data space into a few features. Generally, remote sensing image contains (is composed of) many different objects such as land cover classes, but for a specific purpose of remote sensing application, only a few classes may be relevant. In this paper, a new method called Modified Principal Component Analysis (MPCA) is proposed and applied to a DAIS (Digital Airborne Imageing Spectrometer) image acquired in Venice, Italy. The results show that the features form MPCA is more effective in information compression, classes separablity and classification accuracy than those form PCA.

I. INTRODUCTION

Hyperspectral technology for earth observation makes it possible to collect large numbers of spectral bands. For example, the NASA/JPL airborne visible/infrared imaging spectrometer (AVIRIS) offers image data of 224 bands simultaneously. The high dimensionality of data carries detailed spectral information on the ground surface, but it increases complexity of data receiving, storing, transforming and processing [1].

Since data with high spectral resolution lead to correlated and redundant information for classification, one way to overcome the problem is to reduce the dimensionality of data space. Different feature extraction and selection techniques are proposed in the literature such as the branch-and-bound technique [2], floating search technique [3] and the discriminant analysis technique [4].

Principal component analysis (PCA) is an effective method for feature extraction, it involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. Many studies have been done with this method for information compression [5] [6].

Usually the whole image is used to compute the covariance matrix, so the transformed matrix used in PCA are also coming from the whole image. All possible cover classes in the research area are considered, possibly including some cover classes not relevant for a specific application. In this paper, a Modified Principal Component Analysis (MPCA) is proposed where the transformed matrix is computed from samples of selected classes only.

In the second section, the MPCA is described and compared with the tranditional PCA. By applying these two methods on the hyperspectral image of Venice, Italy, a few examples are given in the third section. In the fourth section, the transformed features obtained by the two methods are evaluated through the comparison of the discrimination of paired classes, the degree of information compression as well as the classification accuracy. Finally, the conclusion is given in the last section.

Π. Modified Prncipal Component Analysis (MPCA)

PCA is a classical multivariate data analysis method that is useful in linear feature extraction and data compression. The approach has been applied in many fields of information processing to extract useful important features for data compressing and classification due to its error minimizing and de-correlating properties.

Indicating the spectral data (original image) as the matrix: $X= [x_{ik}]_{m\times n}$, where m is the number of the origional spectral bands and n is the number of pixels in whole scene. So each line in this matrix stands for one band of the original bands.

In general, the linearly transform (PCA) can be expressed as following equation:

$$Y = TX \tag{1}$$

where T is the transform matrix, X is the origional vectors and Y is the transformed vectors. In order to solve the transform matrix T, the following equation:

$$(\lambda I - S)U = 0 \tag{2}$$

is used, where the matrices I , S, U and λ are the square matrix with unity along its diagonal, the covariance matrix of original images, the eigenvectors and the eigenvalues. U_j *and* $\lambda_j\ (j = 1,2,...,m)$ can be computed through the equation (2), with the eigenvalues ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. The eigenvectors U can be expressed as $U = [U_1, U_2, \cdots, U_m] = [u_{ij}]_{m\times m}$, where U is satisfies with the equation: $U^T U = UU^T = I$. The matrix T can be determined by inversing the matrix U.

Previous studies have demonstrated that PCA is effective in information compression for all classes within the imaged area. In most remote sensing applications we may be interested in a fewer classes and some classes present in the scene may be neglected. The PCA method cannot guarantee that the information related to the relevant classes is effectively compressed.

0-7803-7929-2/03/$17.00 (C) 2003 IEEE

In the new mothed MPCA, training samples, which are relevant for a given application, were selected from a scene, and the transformed matrix T' was obtained form these training samples. It can be expressed as the following aquation:

$$Y' = T' X \tag{3}$$

Comparing the two equations (2) and (3), the difference lies in the transform matrix, and essentially lies in the samples for calculating the covariance matrix, one is from training samples, the other is from the whole scene image.

III. APPLICATIONS

A. Data

The research area was selected at a wetland in the lagoon of Venice, Italy. This area is predominated by salt marsh vegetation and water body. Hyperspectral data of this region was aquired by the DAIS system (Digital Airborne Imageing Spectrometer) in July 1999, this sensor has 79 bands which covers the spectral range from the visible to the thermal infrared wavelengths (0.49-12.28 um) at variable spatial resolutions from 2m to 30 m. The system is integrated and operated by DLR (Institute of Opto-Electronics) since the beginning of 1994.

In this research, the original DAIS data includes 190 rows and 400 columns. In order to demonstrate the method detailly and reduce the time-cost, only the first 10 bands in visible range is used as the original bands, Table 1 shows the the central wavelength of each band.

Table1 Central wavelength of each band in original bands (um)

Band	1	2	3	4	5
Wavelength	0,490	0,512	0,530	0,548	0,564
Band	6	7	8	9	10
Wavelength	0,585	0,602	0,635	0,654	0,672

B. Procedure

Before applying the MPCA, we should analyse the original bands. Correlation coefficients among all bands were calculated and shown in table 2. From this table, a very high correlation can be found among the neighbouring bands. On the basis of the value of the correlation coefficient (above 0.8), there existes three sub-regions, the first region is

Table 2 Correlation coefficient among the original bands

	1	2	3	4	5	6	7	8	9	10
1	1,0	0,9	0,9	0,9	0,8	0,5	0,2	0,1	-,1	-,3
2	0,9	1,0	0,9	0,9	0,8	0,6	0,2	0,1	-,1	-,3
3	0,9	0,9	1,0	0,9	0,8	0,6	0,3	0,1	0,1	-,2
4	0,9	0,9	0,9	1,0	0,9	0,7	0,4	0,2	0,1	-,1
5	0,8	0,8	0,8	0,9	1,0	0,9	0,7	0,6	0,4	0,2
6	0,5	0,6	0,6	0,7	0,9	1,0	0,9	0,8	0,7	0,5
7	0,2	0,2	0,3	0,4	0,7	0,9	1,0	0,9	0,9	0,8
8	0,1	0,1	0,1	0,2	0,6	0,8	0,9	1,0	0,9	0,9
9	-,1	-,1	0,0	0,1	0,4	0,7	0,9	0,9	1,0	0,9
10	-,3	-,3	-,2	-,1	0,2	0,5	0,8	0,9	0,9	1,0

from the first band to the fifth band, the second region covers from the sixth band to the seventh band, the last region ranges from the eighth band to the tenth band.

The first step is to select the trainning samples from the real data for MPCA processing. Since the ecosystem of wet land in Venice is the objective of this research, six cover classes (sea, river, pond, shrub, marsh and crop were selected in the real data scene and the total number of selected samples is 1430 pixels. Table 3 displays the number of the training samples for each class.

Table 3 Description of six interested classes in this experiment

Class	1	2	3	4	5	6
Surface type	Sea	River	Pond	Shrub	Marsh	Crop
Samples	225	183	289	269	186	278

For MPCA, transform matrix T' is first computed from the training samples, then applying this transform matrix to each pixel in the image, all the pixels in the image can be therefore projected, using (3), into another space in which the first one or two principal components (PC1, PC2) contain most information (>95%).

The method of MPCA is applied to the three sub-regions respectively using the transform matrix T' resulting from the training samples. In order to compare with PCA, then we also applied the traditional PCA where the transform matrix T comes from the entire image for the same bands.

Comparing the principal components obtained respectively using the two methods with the the first sub-region channels (Table 4), it is found that the information of PCs using MPCA is more concentrated than PCA. The information contained in PC1 using MPCA is about 97% of the whole information, it is improved nearly 2% with respect to the PCA which is only about 95%. This shows that the MPCA is relatively better than the PCA on the information compression.

After these two transforms, the correlation coefficient among all the PCs in each method is produced (Table 5 and Table 6). For PCA, the transformed space is an orthogonal space because of the transform matrix coming from the whole image, it can be seen from the Table 5 that the correlation coefficient between two different PCs is 0. But the MPCA, based on the transform matrix of the training samples, is not the same result as the standard PCA, the values of correlation coefficients in Table 6 are greater than the corresponding values in table 5 excluded the diagonal line, especially the correlation coefficient between the PC1 and PC2 is near to 0.5.

Combining the above results, though MPCA destroys the orthogonal principle of the transformed space, it is better than PCA in imformation compression because the first principal component (PC1) of MPCA can provide more information for the specific purpose (only six classes in our research).

Table 4 Information of each principal component

Method	PC1	PC2	PC3	PC4	PC5
MPCA	0,968	0,029	0,001	0,001	0,001
PCA	0,950	0,045	0,002	0,002	0,001

Table 5 Correlatioon coefficient between PCs of PCA

	PC1	PC2	PC3	PC4	PC5
PC1	1,00	0,00	0,00	0,00	0,00
PC2	0,00	1,00	0,00	0,00	0,00
PC3	0,00	0,00	1,00	0,00	0,00
PC4	0,00	0,00	0,00	1,00	0,00
PC5	0,00	0,00	0,00	0,00	1,00

Table 6 Correlatioon coefficient between PCs of MPCA

	PC1	PC2	PC3	PC4	PC5
PC1	1,00	0,44	-0,02	0,02	0,05
PC2	0,44	1,00	-0,01	-0,06	-0,01
PC3	-0,02	-0,01	1,00	0,03	0,14
PC4	0,02	-0,06	0,03	1,00	0,11
PC5	0,05	-0,01	0,14	0,11	1,00

Finding a subset of features is the key point in classification. In many methods of feature selection, the J-M distance is an efficient method which can be conceptualized as a measure of the average difference between two class density functions, it has a range from 0, in the case of two classes that completely overlap, to 2, for two classes being completely separable. A J-M distance is calculated for each pair of interclass comparisons and then averaged to give an overall measure of class separability [7]. Table 7 shows the J-M distance of the first sub-region bands in the original space (the first row) and the five PCs in two transformed space respectively (the second and third row). Comparing with the distances obtained by the three different feature spaces, the values in the first row (in the original space) are very similar, indicating that the five original bands present the similar ability to discriminate the six classes, whereas , the PCs of the two transformed spaces have different abilities to discriminate the six classes. In the table 7, J-M distances of PC1 in MPCA are greater than these of PCA, implying that the PC1 in MPCA is more efficient in the discriminatation of the six classes.

Table7 J-Mdistance of each feature in three feature spaces

Method	1	2	3	4	5
Original	1,5395	1,5593	1,6099	1,6663	1,6509
PCA	1,7266	1,5690	0,0500	0,1226	0,0748
MPCA	1,7493	1,2121	0,0787	0,0739	0,1017

Using the same procedures, results of the other two sub-regions show the similar conclusions as these given above from the first sub-region.

In this research, maximum likelihood classification (MLC) and four different feature combinations (Table 8) from three different feature spaces are used to classify the image.

Table 8 Description of the input features for classification

	Features for classification	Description
1	Original space: band 8, 4, 9	the frist three largest J-M distance's bands in original space
2	Original space: band 4, 6, 8	Band 4 has the largest J-M distance in the first sub-region, Band 6 and 8 are the other two sub-region respectively ,
3	three PC1s in PCA space	Each sub-region has a PC1 in PCA space
4	three PC1s in MPCA space	Each sub-region has a PC1 in MPCA space

Table 9 Classification accuracy of four feature combinations

Combination	1	2	3	4
Accurancy	0.985	0.985	0.989	0.991

The classification accuracy resulting from the confusion matrice shows that the features of MPCA space provide the highest accuracy of the four band combinations, as it can be seen from table 9.

IV. SUMMARY AND CONCLUSIONS

The fundamental objective of this research is to develope an effective and pratical feature extraction method for classification. The new method MPCA was developed on the basis of the traditional PCA and the difference between these two methods lies in the different correlation matrix, the former uses only the training samples for calculation of correlation matrix, whereas the latter uses the entire image.

In this paper, a hyperspectral image (DIAS) taken in Velice, Italy was selected as the original data. MPCA and PCA were applied to the same image respectively, and results show that MPCA is more powerful in information compression, class discrimination and classification for specific purpose.

MPCA has great advantage in feature extraction for classification because it has a specific purpose in a certain application. For one image, there are different purposes of application to different researchers, so the relevant information, which the researchers want to extract, is not the same.

In addition, since the training simples for calculating the covariance of MPCA is only based on the training samples, it cost less time for computing than PCA.

REFERENCES

[1] Fukanaga, K., Introduction to Statistical Pattern Recognition, Academic Press, San Diego, CA, 1990.

[2] Narendra,P., Fukanaga, K, "Abranch and bound algorithm for feature subset selection." *IEEE Trans. Comput*. V26 (9), 1977, pp. 917-922..

[3] Pudil, P., Novovicova, J., Kittler, J., "Floating search methods in feature selection. " Pattern Recognition Lett. 15, 1994, pp. 1119-1125.

[4] Lee, C., Landgrebe, D., "Feature extraction based on decision boundaries." IEEE Trans. Pattern Anal. Machine Intell. 15 (4), 1993, pp. 388-400.

[5] Mark W. Patterson and Stephen R. Yool, "Mapping Fire-Induced. Vegetation Mortality Using Landsat Thematic Mapper Data: A Comparison of Linear Transformation Techniques," *Remote Sensing of Environment*, 65, 1998, pp. 132–142.

[6] Yuji Hirosawa, Stuart E. Marsh and Douglas H. Kliman, "Application of Standardized Principal Component Analysis to Land-Cover Characterization Using Multitemporal AVHRR Data," *Remote Sensing of Environment*, 1996, pp. 267-281.

[7] Swain, P. H., and Davis, S. M., *Remote sensing:The quantitative Approach*. McGraw Hill, New York, pp. 369, 1978.