Asia-Pacific Journal of Science and Technology: Volume: 22. Issue: 01. Article ID.: APST-22-01-02. Research Article

APST

Asia-Pacific Journal of Science and Technology

https://www.tci-thaijo.org/index.php/APST/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Feature reduction using minimum noise fraction and principal component analysis transforms for improving the classification of hyperspectral image

Murinto[1,*], Nur Rochmah Dyah PA[1]

[1] Department of Informatics, Faculty of Industrial Engineering, Universitas Ahmad Dahlan Yogyakarta, Indonesia
*Corresponding author: murintokusno@tif.uad.ac.id

Abstract

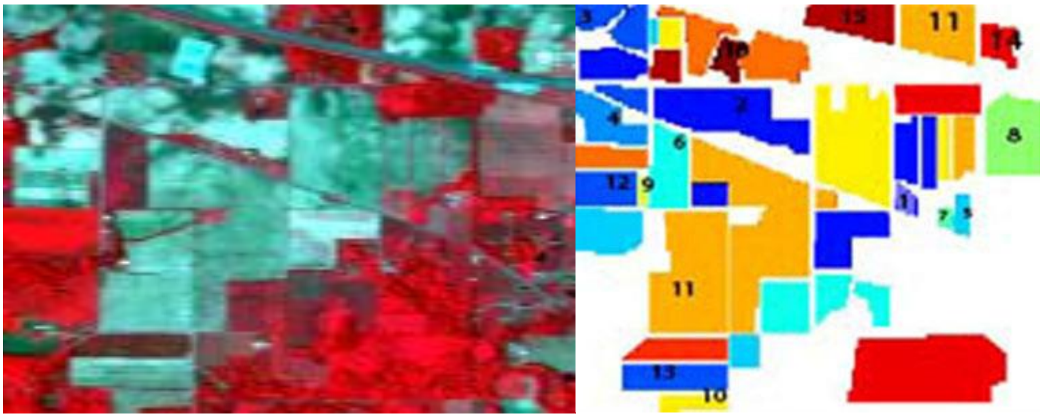
Dimensionality reduction is an important milestone in the preliminary process of high-dimensional data analysis. Most of the research on hyperspectral image fields deal with data extraction techniques. Each feature extraction technique is unique and has its advantages and disadvantages. However, using certain techniques may result in significant data loss. To avoid such problems, this research employs a combination of reduction techniques. In this paper, dimensionality reduction was conducted using principal component analysis (PCA) and minimum noise fraction (MNF). A combined principle component analysis and minimum noise fraction (PCA-MNF) method is proposed. Image classification using a minimum distance (MC) method was performed subsequent to the dimensionality reduction technique. We found that our proposed method increases the accuracy of image classification to 80.77% outperforming both PCA and MNF, which yield 40.37% and 77.21% accuracy, respectively.

Keywords: Classification, Feature Reduction, Hyperspectral image, MNF, PCA.

1. Introduction

The spectral characteristics of image pixels are commonly used as an information source in the classification process. In the context of supervised classification, the real challenge lies in the fact that the adaptation of high-dimensional data volume is needed. In other words, due to the limited number of training samples available, as well as the high number of features involved in remote sensing, it is difficult to achieve reliability with regard to class parameter statistical estimation. As a result, when using a limited number of training samples, the classification accuracy tends to decrease when the number of features increases. This phenomenon is known as the Hughes phenomenon [1].

Data dimensionality reduction is an important step in the preliminary process of high-dimensional data analysis. With the addition of spatial features into hyperspectral remote sensing images, the problems involved in high-dimensional data become more pronounced. For example the data model becomes more complex, and there is increase in the number of training samples required in supervised models. There are two kinds of feature reduction methods used for hyperspectral remote sensing images. The first is unsupervised. The most popular method in this group is Principal Component Analysis (PCA) [2]. The second type of feature reduction method is supervised, which requires training samples to learn the best data extraction structure for the classification process. One common method in this group is Fisher Discriminant Analysis (FDA) [3]. Other supervised methods are Local Fisher Discriminant Analysis (LFDA) [4] and Minimum Noise Fraction (MNF)[5]. The purpose of this study is to use the supervised method, MNF, for dimension reduction data. The MNF method was combined with PCA to obtain better classification accuracy.



(a) (b)

Figure 1 Indian Pines dataset (a) *false*-composite(band 57,27,17) (b) available data references*:* (1)*alfalfa*, (2) *corn-notill, (3)corn-mintill,* (4) corn (5)*grass-pasture,* (6) *grass-trees,* (7) *grass-pasture-mowed*, (8) *hay-windrowed*,(9) oat, (10) *soybean-notill,* (11) *soybean-mintill,* (12) *soybean-clean, (13) wheat,* (14) *woods,* (15) *buildings-grass-trees-drives*, (16) *stone-steel-towers.*

2. Materials and methods

2.1. Hyperspectral dataset

This study uses the Aviris Indian Pines hyperspectral imagery dataset, obtained from Aviris sensors over northwest Indiana in 1992. The original image size was 145 x 145 pixels. The imagery consists of 220 spectral bands (after removal of spectral bands that contained noise and water absorption). The three-band false-color composite image (57, 27, 17 for red, green, and blue) from Aviris Indian pines and the ground truth data are sequentially shown in Figure 1. (ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.tif.zip)[6].

2.2. Experimental method

Data dimensionality reduction was performed using the following transformation techniques: MNF, PCA, and combined MNF-PCA using ENVI 4.5. The MNF method was performed first. Then the results obtained were used for PCA data dimension reduction. There were 25 bands of principal components selected for each technique. The PCA data were obtained from data variance, whilst MNF data were obtained from SNR values. The minimum distance (MD) method was employed for classification.

3. Results and discussion

3.1. Data dimensionality reduction

In this study, the intrinsic model of data dimension [7-8] was used to determine the effectiveness of the principal components of the hyperspectral data used. Moreover, 25 first principal components were taken from the Aviris Indian Pines image data. These 25 principal components were obtained by first calculating the correlation matrix and then calculating the eigenvalues of the image data. Table 1 shows the statistical cumulative variance percentage values and eigenvalues of the 25 first principal components of the image data.

PC, MNF and combined MNF-PC transformation were performed on the image data obtained from the AVIRIS sensor. A covariance matrix was used to calculate the eigenvector and eigenvalue. Two principal components of Aviris Indian Pines utilize the PCA method, two utilize MNF, and two utilize combined the MNF-PCA method, based on the scree test criteria. In applying the cumulative variance method, a minimum total variance of 90% is required to retain PC (Table 1), showing that two PC components of the Aviris Indian Pines images use the PCA method, 15 PC components use MNF and 15 PC components use combined MNF-PCA. The next step of the analysis process is to observe the results of image reduction directly, to determine a better estimation of what components contain useful information for image classification. Figure 2 shows 25 first-principal components of the Aviris Indian Pines image data using the combined MNF-PCA method. In this study, a combined MNF-PCA method was performed as a data dimensionality reduction technique on the hyperspectral image classification process. This technique was performed by first applying an MNF data dimensionality reduction technique, followed by a PCA data dimensionality reduction technique.

Table 1 Statistical values of PCA, MNF, and MNF PCA in the Indian Pines Image

PC#	PCA		MNF		MNF PCA	
	eigenvalue	Variance cumulative	eigenvalue	Variance cumulative	eigenvalue	Variance cumulative
1	25383112	72.58	46.54	27.30	46.54	27.30
2	8629688	97.25	19.34	38.64	19.34	38.64
3	606429	98.98	14.81	47.33	14.81	47.33
4	128538	99.35	12.75	54.81	12.75	54.81
5	69276	99.55	11.92	61.80	11.92	61.80
6	53271	99.70	8.96	67.05	8.96	67.05
7	25984	99.78	8.12	71.82	8.12	71.82
8	13618	99.82	6.51	75.64	6.51	75.64
9	12671	99.85	5.84	79.06	5.84	79.06
10	8336	99.88	4.79	81.87	4.79	81.87
11	6824	99.89	3.35	83.83	3.35	83.83
12	5838	99.91	3.09	85.65	3.09	85.65
13	5055	99.93	2.48	87.10	2.48	87.10
14	3165	99.93	2.14	88.35	2.14	88.35
15	3037	99.94	2.01	89.53	2.01	89.53
16	2716	99.95	1.93	90.66	1.93	90.66
17	2575	99.96	1.86	91.75	1.86	91.75
18	2268	99.97	1.84	92.83	1.84	92.83
19	2085	99.97	1.81	93.89	1.81	93.89
20	1868	99.98	1.76	94.93	1.76	94.93
21	1790	99.98	1.75	95.95	1.75	95.95
22	1742	99.99	1.74	96.97	1.74	96.97
23	1600	99.99	1.73	97.99	1.73	97.99
24	1539	99.99	1.72	98.99	1.72	98.99
25	1514	99.99	1.71	99.00	1.71	99.00

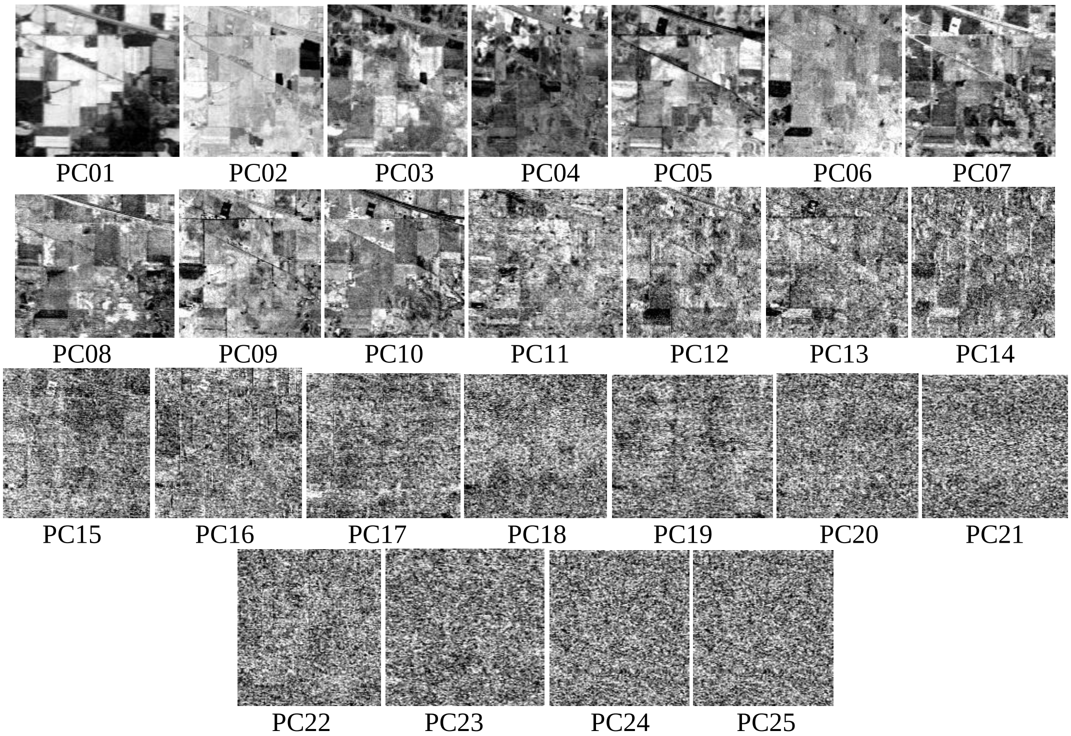


Figure 2 25 first PC of combined PCA and MNF of Aviris Indian Pines image data

Through observation of the PCs in both images, it is clear that there is additional information. In images that depict water, these components reflect variations in water depth. The high value of the second PC indicates sallow areas and the decrease in brightness indicates a decrease of water depth. The third PC indicates basically the same thing as the second PC, but more detailed information is revealed. Vegetation on the land is easier to differentiate, and water depth variation, hills and settlement areas are clearly visualized. Therefore, based on both the scree test and visual observation of the images, 15 PCs provide most efficient and most informative visualization and represent the intrinsic dimensions of the image. The cumulative variance at 15 PCs indicates that the rest of the PC band contains less correlated information. In general, the rest of the PCs do not reveal additional information about the image. However, there are several other findings that can be derived from this comparison. At bands 5 and 6 the image is not defined clearly. Unlike previous image components, the image cannot be differentiated properly. Bands 7 and 8 are full of noise, as are all bands beyond 10, due to small variations caused by noise in the original data. In this problem, it does not give image with degrading quality. Based on the scree test method, cumulative variance, direct observation of the reduction result, only the 10 PC band should be processed for image classification.

3.2. Classification using the minimum distance (MD) method

In this research, overall accuracy (OA) of the classification was improved by using the data dimensionality reduction model with a combined MNF-PCA method. Based on the abovementioned parameters, the 10 PC band was employed, using PCA, MNF, and combined PCA-MNF data dimensionality reduction methods. MD classification was then performed and the overall accuracy value was calculated. Table 2 shows that using MD classifiers and the combined MNF-PCA data dimensionality reduction method resulted in overall accuracy of 80.77% and 0.78 Kappa Index. This overall accuracy was higher than that of the PCA method, which resulted in overall accuracy of 40.37% and 0.351 Kappa Index. It was also higher than that of the MNF method, which resulted in overall accuracy of 77.21% and 0.74 Kappa Index. The result of 10 principal components classification using a MNF-PCA combined data dimensionality reduction method is shown in Figure 3. The overall accuracy and Kappa Index of Aviris Indian Pines image data are shown in Table 2.

Figure 3 Classification using MD of reducted data (a) Aviris Indian Pines (b) using PCA (c) using MNF (d) using combined MNF-PCA

Table 2 Overall accuracy (OA) and Kappa index of data dimensionality reduction using the minimum distance (MD) classifier of AVIRIS Indian Pines data

Types of features	Aviris Indian Pines		Processing time
	OA (%)	Kappa	
Original	40.35	0.35	7m
PCA	40.37	0.351	3m
MNF	77.21	0.74	2m
MNF+PCA	80.77	0.78	6m

5. Conclusions

Each feature extraction method has its own advantages and disadvantages. Employing certain techniques in data dimensionality reduction may result in significant information loss. The approach proposed in this study overcomes that problem. The results showed that the overall accuracy of classification of the proposed data dimensionality reduction method was 80.77%, outperforming both PCA and MNF. The classification accuracy obtained by using PCA is only 40.37% and that of MNF is only 77.21%. The processing time obtained using PCA is 3m, MNF is 2m, and MNF + PCA is 6m

6. References

[1] Hughes, G., 1968. On the accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14, 55-63.
[2] Bhisop, C.M., 2006. Pattern Recognition and Machien Learning. Springer; number 978-0387-31073-2.
[3] Fukunaga, K., 1990. Introduction to statistical pattern recognition, second ed. Academic Press.
[4] Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. Journal of machine learning research 8, 1027-1061.
[5] Rodarmel, C., Shan, J., 2002. Principal component analysis for hyperspectral image classification. journal of the American Congress on surveying and mapping 2.
[6] Aviris Dataset 2002. [WWW Document]. URL ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C .tif.zip. (accessed 1. 3. 16).
[7] Cattel, R.B., 1966. The cree test for number od factors. Multivariate behavioral research 1, 245-276.
[8] Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and psychological measurement.