Principal Component Analysis for Hyperspectral Image Classification

Craig Rodarmel and Jie Shan

ABSTRACT: The availability of hyperspectral images expands the capability of using image classification to study detailed characteristics of objects, but at a cost of having to deal with huge data sets. This work studies the use of the principal component analysis as a preprocessing technique for the classification of hyperspectral images. Two hyperspectral data sets, HYDICE and AVIRIS, were used for the study. A brief presentation of the principal component analysis approach is followed by an examination of the information contents of the principal component image bands, which revealed that only the first few bands contain significant information. The use of the first few principal component images can yield about 70 percent correct classification rate. This study suggests the benefit and efficiency of using the principal component analysis technique as a preprocessing step for the classification of hyperspectral images.

KEYWORDS: Hyperspectral images, image classification, land use, principal component analysis

Introduction

The spectral resolution of a sensor determines much of the capability and performance of a remote sensing system, which uses the detected spectral properties of the object for processing and analysis. The spectral resolution refers to the spectral width that a sensor can detect in one single image band. Several types of images with different spectral resolutions have been identified (Schowengerdt 1997). The common panchromatic image records the object in one band which covers the entire visible portion (ca. 300 nm wide) of the spectrum. It is therefore known as broadband image. Multispectral images, e.g., SPOT and Landsat images, have a relatively narrow spectral width of about 50-100 nm. Hyperspectral images are those having 5-10 nm spectral width, which can reach a nearly contiguous spectral record for the object. Though future development will lead to ultraspectral images at a spectral resolution of less than 5 nm, recent research activities are mostly focused on multispectral and leading increasingly toward hyperspectral images. Currently available airborne hyperspectral imaging systems are, among others, AVIRIS (Airborne Visible/Infrared Imaging Spectrometer (Porter and Enmark 1987)), HYDICE (Hyperspectral Digital Image Collection Experiment), DAIS (Digital Airborne Imaging Spectrometer (Lanzl and Mueller 1999)), HyMap (http://www.intspec.com, Australia), MAIS (Modular Airborne Imaging Spectrometer), and Push-broom Hyperspectral Imager (PHI) (Zhang et al. 2000). By the time of writing this article (January 2002), the space-borne hyperspectral sensor Hyperion with 220 bands on board of EO-1 satellite had successfully collected the first space-borne hyperspectral images [see http://eo1.gsfc.nasa.gov]. Other space-borne hyperspectral systems are being planned such as HyperSpectral Imager (NASA), NEMO (Naval EarthMap Observer) (Wilson and Davis 1999) and OrbiView-4 from ORBIMAGE.

The use of hyperspectral images brings in new capabilities along with some difficulties in their processing and analysis. Unlike the widely used multispectral images, hyperspectral images can be used not only to distinguish different categories of land cover, but also the defining components of each land cover category, such as minerals, and soil and vegetation type. On the other hand, there are also difficulties in processing so many bands. The large amount of data involved with hyperspectral imagery will, however, dramatically increase processing complexity and time. Effectively reducing the amount of data involved or selecting the relevant bands associated with a particular application from the entire data set becomes a unique, yet primary task for hyperspectral image analysis. Besides, as is described as the Hughes effect (Hughes 1968; Shahshahani and Landgrebe 1994), the classification quality may decrease if more image bands are

Craig Rodarmel, EarthData International of Maryland, 45 West Watkins Mill Road, Gaithersburg, MD 20878 , USA. **Jie Shan**, Geomatics Engineering, Purdue University, West Lafayette, IN 47907-1284, U.S.A. E-mail: <jsahn@ecn.purdue.edu>.

used for the reduction. Feature or subspace selection preprocessing therefore needs to be performed on the data (Campbell 1996). In this paper we use the principal component analysis (PCA) to select the best bands for classification, analyze their contents, and evaluate the correctness of classification obtained by using PCA images.

The principal component analysis has been used in remote sensing for different purposes. A mathematical derivation and historical review of PCA are presented in (Gonzalez and Woods 1993). Brief discussions may also be found in (Lillesand and Kiefer 2000; Campbell 1996). Most of the research explores ways of obtaining effective multispectral image classification, while the study of PCA performance and its improvement has been limited. Mather (1999) presented a comprehensive and detailed summary of different applications of PCA, including correlation analysis of Landsat TM images for effective feature recognition and identification of areas of change with multitemporal images. Carr and Matanawi (1999) introduced the correspondence analysis into PCA for SPOT multispectral image analysis and studied its possible application for image compression. As most existing work deals with multispectral imagery, it is open for discussion whether its conclusions apply to hyperspectral imagery as well.

The primary objective of this research is to determine the applicability of PCA in the classification of hyperspectral images. The contents of PCA bands for two common hyperspectral sensors (HYDICE and AVIRIS) were analyzed with a view of identifying the most informative bands. The selected PCA bands were then used for a supervised classification and the results were evaluated by comparing them to the classification results obtained using the original hyperspectral data.

The paper is organized as follows. Following this introductory section is a brief introduction of the PCA approach. The next section introduces the HYDICE and AVIRIS data sets used for this study and presents the test and analysis methods. The contents of the PCA bands are then examined, followed by a comparison of the classification results obtained using the PCA bands and the original hyperspectral data set. Graphic and numeric measures are given to illustrate the results and analyses in this section. The last-but-one section compares the efficiency of the PCA transformation in terms of computational time. The final section present experience gained and future work.

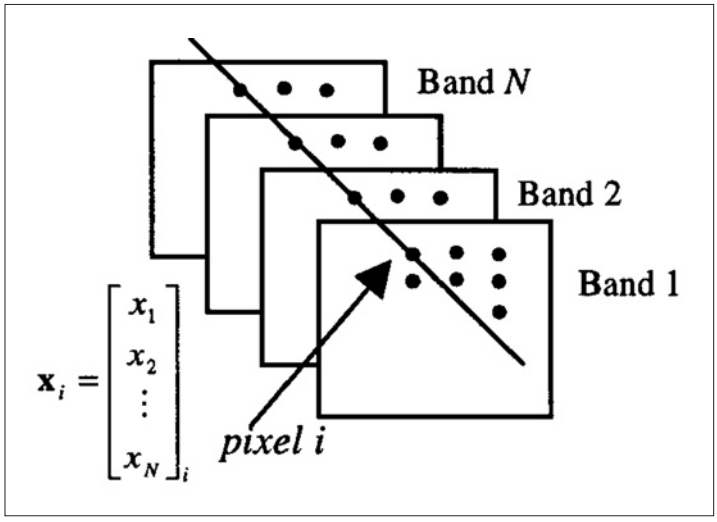


Figure 1. Pixel vector in principal component analysis [adapted from Gonzales and Woods (1993)].

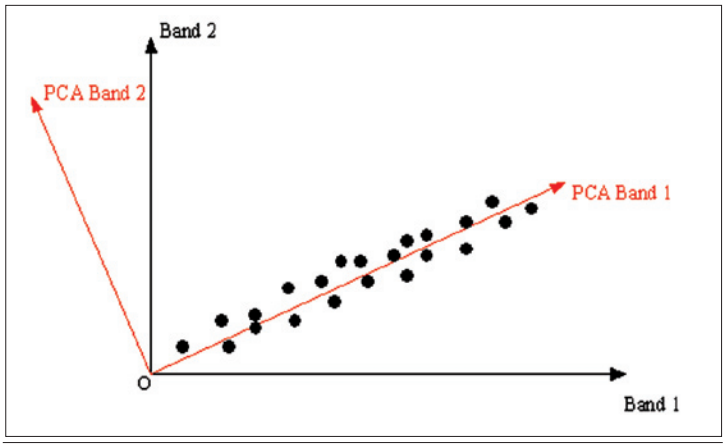
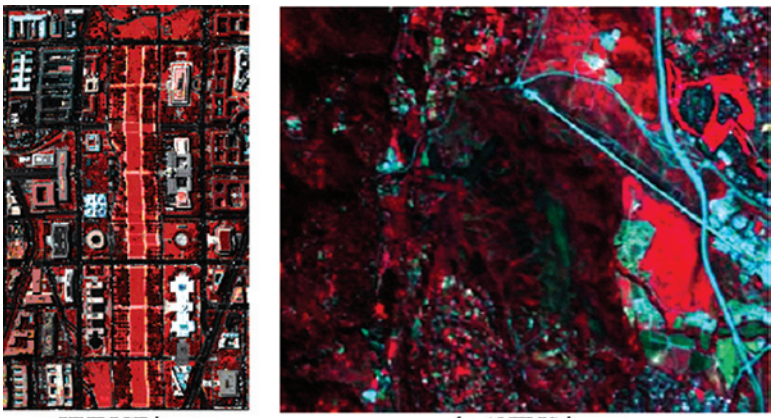


Figure 2. Geometry of principal component analysis and PCA bands.

Principal Component Analysis

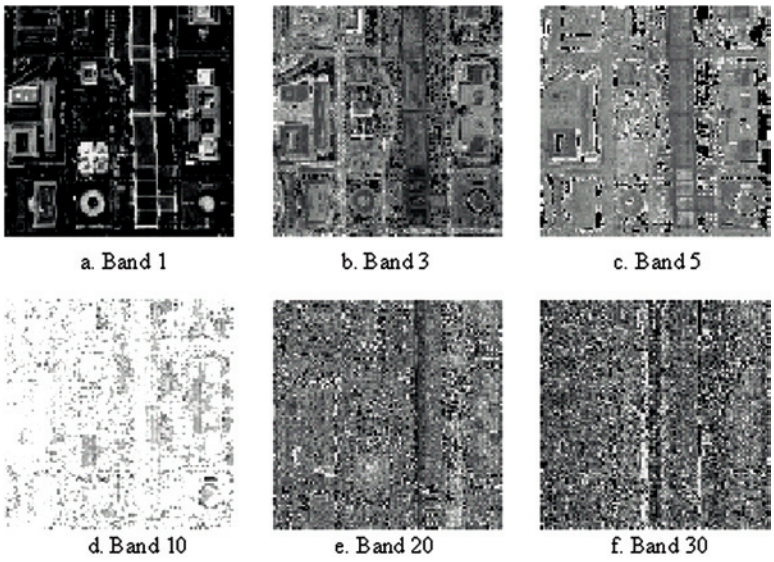
The principal component analysis is based on the fact that neighboring bands of hyperspectral images are highly correlated and often convey almost the same information about the object. The analysis is used to transform the original data so to remove the correlation among the bands. In the process, the optimum linear combination of the original bands accounting for the variation of pixel values in an image is identified.

The PCA employs the statistic properties of hyperspectral bands to examine band dependency or correlation. Though, one may find many synonyms for PCA, such as the Hotellling transforma-



a. HYDICE image b. AVIRIS image

Figure 3. Original hyperspectral images.



a. Band 1 b. Band 3 c. Band 5

d. Band 10 e. Band 20 f. Band 30

Figure 4. Sample PCA bands of the HYDICE image of the Washington, D.C., Mall area.

tion or Karhunen-Loeve transformation (Gonzalez and Woods 1993), all these transformations are based on the same mathematical principle known as eigenvalue decomposition of the covariance matrix of the hyperspectral image bands to be analyzed. Below is a brief formulation of the principle. Detailed discussions may be found in Gonzalez and Woods (1993) and Schowengerdt (1997).

An image pixel vector is calculated as:

$$\mathbf{x}_i = [x_1, x_2, \ldots\ldots x_N]_i^T \quad (1)$$

with all pixel values x_1, x_2, ... z_N at one corresponding pixel location of the hyperspectral image data. The dimension of that image vector is equal to the number of hyperspectral bands N. For a hyperspectral image with m rows and n columns there will be M=m*n such vectors, namely i=1,…,M. The mean vector of all image vectors is denoted and calculated as:

$$\mathbf{m} = \frac{1}{M}\sum_{i=1}^{M}[x_1 \quad x_2 \quad .. \quad .. \quad x_N]_i^T \quad (2)$$

The covariance matrix of **x** is defined as:

$$\mathbf{Cov(x) = E\{(x-E(x))(x-E(x))^T} \quad (3)$$

where:
E = expectation operator;
T superscript = transpose operation; and
Cov = notation for covariance matrix.

The covariance matrix is approximated via the following calculation:

$$\mathbf{C_x} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (4)$$

The PCA is based on the eigenvalue decomposition of the covariance matrix, which takes the form of:

$$\mathbf{C_x = ADA^T} \quad (5)$$

where:

$$\mathbf{D} = \mathrm{diag}(\lambda_1, \lambda_2 \ldots \lambda_N) \quad (6)$$

is the diagonal matrix composed of the eigenvalues λ_1, $\lambda_2 \ldots \lambda_N$ of the covariance matrix $\mathbf{C_x}$, and **A** is the orthonormal matrix composed of the corresponding *N* dimension eigenvectors $\mathbf{a}_k$ *(k=1,2,…, N)* of $\mathbf{C_x}$ as follows:

$$\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_N) \quad (7)$$

The linear transformation defined by:

$$\mathbf{y}_i = \mathbf{A}^T\mathbf{x}_i \quad (I = 1,2,\ldots, M) \quad (8)$$

is the PCA pixel vector, and all these pixel vectors form the PCA (transformed) bands of the original images.

Let the eigenvalues and eigenvectors be arranged in descending order so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, thus the first *K* (K ≤ N, usually *K*<<*N*) rows of the matrix $\mathbf{A}^T$, namely the first *K* eigenvectors $a_j^T (j = 1,2,\cdots K)$, can be used to calculate an approximation of the original images in the following way:

$$\mathbf{z}_i = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{bmatrix}_i = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2K} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} & \cdots & a_{KN} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \\ \vdots \\ x_N \end{bmatrix}_i \quad (i = 1,2,\ldots M) \quad (9)$$

where pixel vector $\mathbf{z}_i$ will form the first K bands of the PCA images.

Such formed PCA bands have the highest contrast or variance in the first band and the lowest contrast or variance in the last band. Therefore, the first K PCA bands often contain the majority of information residing in the original hyperspectral images and can be used for more effective and accurate analyses because the number of image bands and the amount of image noise involved are reduced. According to Gonzalez and Woods (1993), the PCA bands are mutually independent or orthogonal and their covariance matrix takes the form of:

$$\mathbf{C}_z = \text{diag}(\lambda_1, \lambda_2 ... \lambda_K) \qquad (10)$$

Detailed discussions on the properties and applications of PCA in multispectral images may be found in (Mather 1999). The geometry of the PCA concept is illustrated in Figure 2, where the original data consist of two bands, band 1 and band 2. There is considerable correlation between the two bands: a move in band 1 creates an almost linear change in band 2. Once the PCA takes place, however, the correlation between the PCA band 1 and 2 vanishes. Another aspect of PCA analysis that can be seen in this illustration pertains to the variability within bands. Once the transformation has taken place, PCA band 1accounts for the maximum amount of variability or contrast possible in the image and PCA band 2 accounts for the second largest amount. This trend is likely to continue in the first few PCA bands, with the remainder containing less and less useful information.

Hyperspectral Images for the Study

Two hyperspectral data sets were selected and used for the analyses. The first was a HYDICE image of the Washington, D.C., Mall area (Figure 3a). This data set was acquired by an aircraft-borne sensor and originally contained 191 bands. The second was an AVIRIS image for Jasper Ridge, California, which originally contained 224 bands (Figure 3b). The HYDICE image shows in urban area, while the AVRIS image covers a predominantly rural area. Other information on these two hyperspectral images is summarized in Table 1.

Sensor	Developer	Band range (μm)	Number of bands	GSD	Swath pixels	Sensor type
HYDICE	U.S Navy	0.4-2.5	210	1m	320	Pushbroom
AVIRIS	JPL	0.4-2.5	224	20m	614	Whiskbroom

GSD = Ground sample distance. JPL = Jet Propulsion Laboratory. AVIRIS = Airborne Visible/ Infrared Imaging Spectrometer. HYDICE = Hyperspectral Digital Image Collection Experiment.

Table 1. Hyperspectral image data used.

Following the collection of image data, ground references for the two sites were collected. For the Washington, D.C., Mall area, previous classifications performed at Purdue University were used as reference (http://dynamo.ecn.purdue.edu/~landgreb/Hyperspectral.Ex.html), along with general knowledge about the site. For the Jasper Ridge, California, image, a GIS map created at the Center for Conservation Biology and JRBP (Jasper Ridge Biological Preserve) at Stanford University was used as the reference (http://jasper1.stanford.edu/).

The ENVI image processing package (Research Systems, Inc 1999) and ArcView GIS from ESRI, Inc. were used as the major processing and analysis tools in this study. ENVI hyperspectral image analysis functions were used to perform PCA on both hyperspectral data sets. To limit the data to be handled to a reasonable size, only subsets of the original bands were used for the analysis, and the PCA was performed on these subsets. In fact, for one particular application, usually only a small portion (subspace) of the entire band of hyperspectral data is used, as is discussed in Landgrebe (2000). In our analysis, the first 1-50 bands of the HYDICE image were selected and the first 1-60 bands were used for the AVIRIS image.

Content Analysis of PCA Bands

As indicated, the information content of PCA bands decreases with an increasing number of PCA bands, and most of the information may only be contained in the first few PCA bands. This fact was confirmed by the following analysis performed for hyperspectral imagery. The HYDICE image was used to calculate the PCA bands. Figure 4, which contains 128*128 pixel samples of the bands, clearly shows that the majority of the variability is accounted for in the first few PCA bands and that the remaining bands quickly become noise. In the Washington, D.C. urban area we studied, the first five PCA bands contained virtually all the useful information and, therefore, were used to substitute for the entire original data set for processing and analysis. In PCA bands 5 through to 10, only the most contrasting areas of the scene were visually distinguishable. After band 10, the PCA bands appeared to be made entirely of noise.

Similar studies collected in Mather (1999) give quantitative descriptions of the contents of

Sensor	Bands Classified	Correct Rate (%)	Classification Time (sec)	Total Time[1] (sec)
HYDICE	Original 1-50	100	46	46
	PCA 1-50	99.99	46	66
	PCA 1-25	92.05	16	36
	PCA 1-10	83.47	7	27
	PCA 1-5	77.24	5	25
AVIRIS	Original 1-60	100	65	65
	PCA 1-10	73.89	7	20
	PCA 1-5	69.63	4	17

[1] The total time is the sum of classification time and PCA calculation time (except for the original imagery). The PCA calculation time for HYDICE and AVIRIS are respectively 20 and 13 seconds on Pentium II.

Table 2. Correct classification rate and calculation time.

PCA bands by using the ratios of the eigenvalue of each PCA band to the sum of all eigenvalues. Suprisingly, both the theoretical and experimental studies indicate that PCA bands with smaller eigenvalues (or low-order PCA bands) may contain apparently visible information that is useful and can contribute to the image classification As a matter of fact, Mather (1999) suggested that inference cannot be made solely based on the magnitudes of the eigenvalues, and a visual check of the obtained PCA bands is necessary and important. Given this conclusion, the next section will conduct image classifications by using both PCA bands and original images, and then analyze the effect of PCA bands on image classification. In this way, it is expected that a reliable examination would be conducted in this study.

Classification of PCA bands

Hyperspectral images are usually used for detailed analysis of the properties of a given object. In theory, PCA transformation affects the classification of hyperspectral images. We will now evaluate the classification results of PCA bands by comparing them with the results obtained from the original images. Both HYDICE and AVIRIS images were used for this analysis. After obtaining PCA bands, the first step in the classification is to select training sites (see Figure 5). In this study, the training sites were areas of known land cover. The radiometric responses of the pixels in the images were statistically compared with the responses in the training sites by using the maximum likelihood classifier (Schowengerdt 1997; Landgrebe 2000). Each pixel in the images was then categorized into the class with which it was most closely associated.

After the selection of training sites, image classification was conducted first for the original images, then for the entire PCA bands, and finally for subsets of the PCA bands. The results of these classifications are shown in Figure 6 (for HYDICE) and Figure 7 (for AVIRIS). A visual comparison immediately shows that classifications from both original and entire PCA images reveal the same class patterns or distributions over the image. A further detail analysis is therefore necessary based on a pixel-by-pixel comparison to study the effects of the PCA bands on the image classification.

When comparing the two classification results obtained, respectively, from the PCA-transformed and the original images, the latter were considered as the reference. These two classification results are subtracted pixel-by-pixel to detect misclassified pixels in the classification results of the PCA bands. The differential images obtained showed the correctly classified and the misclassified pixels along with the statistical data about the percentage of the correct classification. Samples of these differential images are shown in Figure 8 (for HYDICE) and Figure 9 (for AVIRIS). The percentages of pixels that were correctly classified relative to the total number of pixels in the image are listed in Table 2. Figure 10 is a graphical representation of the percentage of correctly classified pixels. An examination of Figure 8 and 9, Table 2 and Figure 10 will reveal the effects of PCA bands on image classification.

The pattern of the classified image did not deteriorate as much as one would have expected based on the percentage of the misclassified pixels. In fact, the class pattern was depicted very well, even when only a few (~10 percent of the total images) PCA bands were involved in the classification. This fact was valid for both urban and rural areas. It might be the biggest benefit in using the PCA approach for hyperspectral image classification, since the overall appearance of the class patterns is not affected by the misclassification while the amount of data involved is significantly reduced to a few bands.

The correct classification rate increases slowly in a nearly linear manner as the number of the PCA bands used for the classification increases. For both data sets, the use of the first 10 percent (~5) PCA bands can obtain a correct classification rate of about 70 percent; the first 20 percent (~10) PCA bands leads to about 75-80 percent

correct classification rate. The use of more PCA band images will only slightly increase the correct classification rate. With the use of 50 percent or more PCA bands the correct classification rate may reach up to about 90 percent. Therefore, the PCA approach can effectively ensure a practically acceptable and accurate classification result by handling only a small data set (5-10 percent) derived from the original large amount of image data.

The misclassification mainly occurs at feature boarders or edges. The boarder effect is due to the loss of information or contrast in the process of transformation, such that the boarder becomes "smoothed" or less contrasted in the PCA band images. This misclassification will decrease the geometric accuracy of the classified image and, hence, the accuracy of derived thematic maps. However, as discussed earlier on, this misclassification does not change the general class patterns and, therefore, the dominating classification results still remain correct.

A comparison of the results of classified HYDICE and AVIRIS images shows that the rural area (AVIRIS image), which is mostly covered by forest and vegetation, is slightly more affected by the PCA images, while the urban area (HYDICE), with its roads, buildings, and trees, is less affected. This fact indicates that urban land-

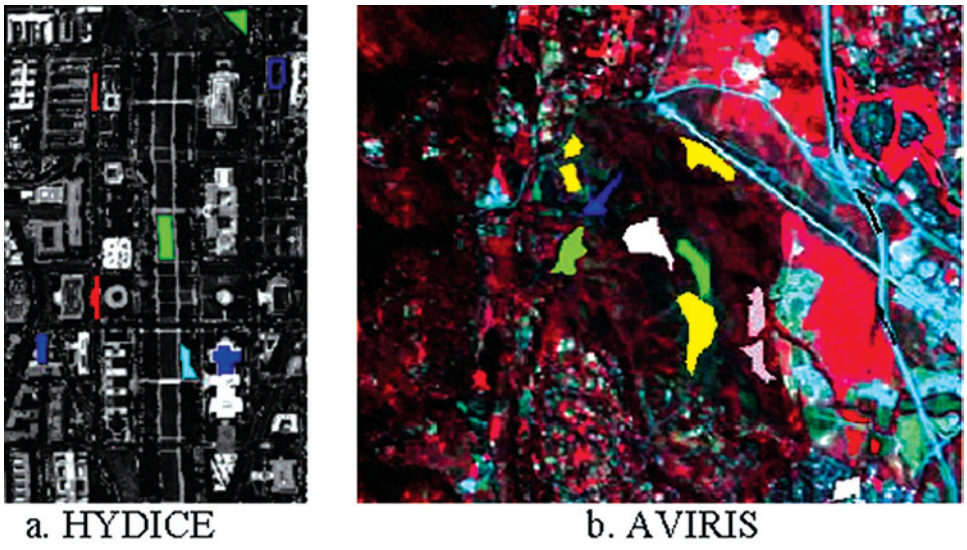


Figure 5. Study sites where the images were taken.

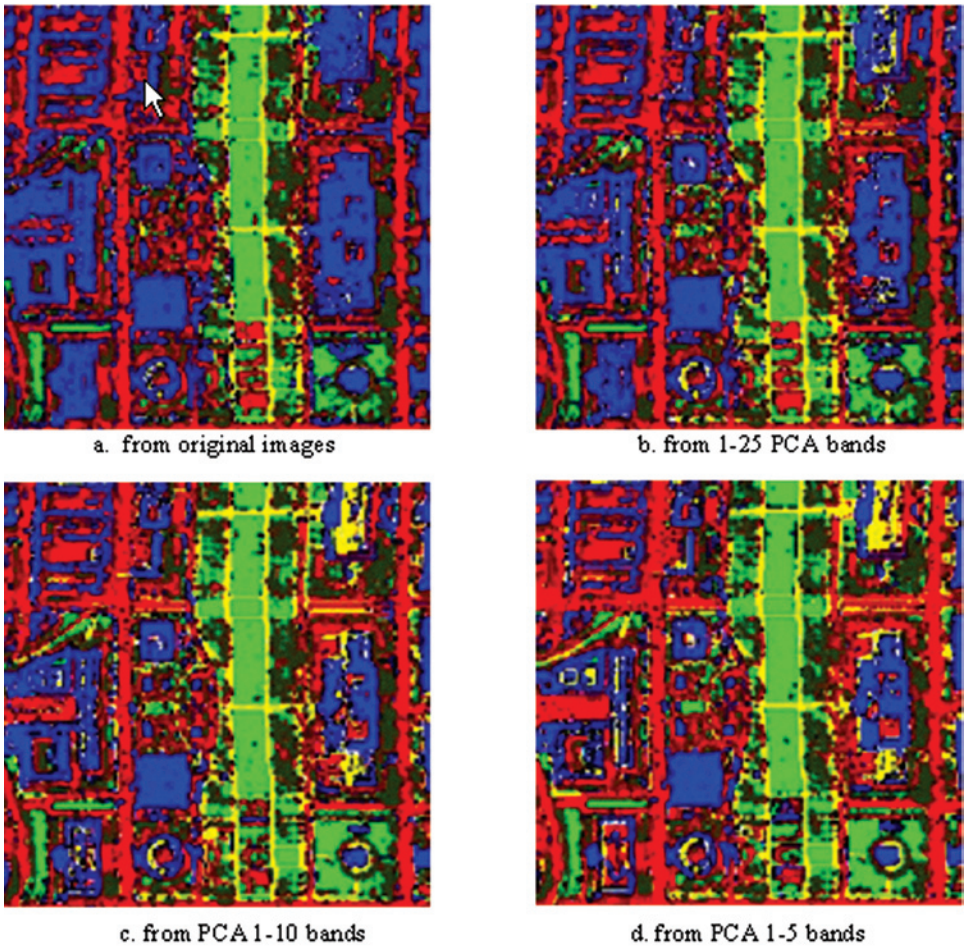


Figure 6. Classification results for the HYDICE images.

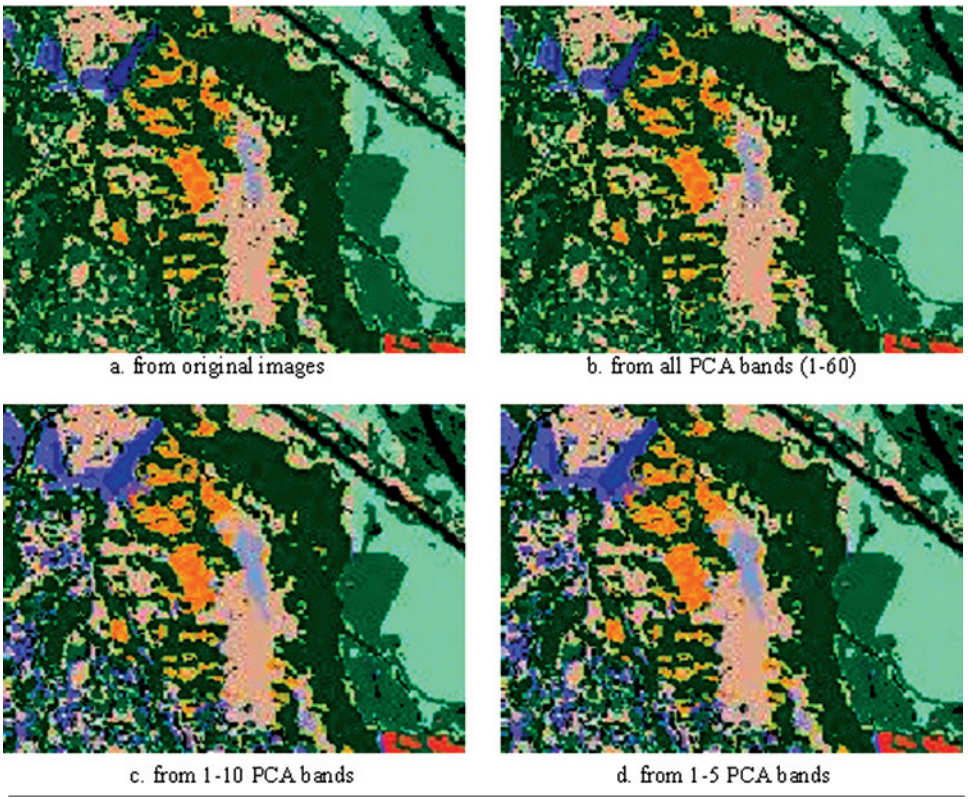


Figure 7. Classification results for the AVIRIS images.

use classification is less sensitive to information loss due to the mathematical transformation involved in the PCA approach than are rural areas and, consequently, requires fewer PCA bands for data reduction. The selection of the number of PCA bands for the classification may thus vary depending on the type of study areas.

Comparison of Computational Time

Table 2 also lists the computational time for the PCA analysis. The total time is the sum of the classification time and the PCA transformation time. The PCA transformation for the HYDICE and AVIRIS images took, respectively, 20 and 13 seconds on a Pentium II CPU. If the classification is applied to the original images, no PCA transformation is needed and the total time will only be the classification time. Because PCA transformation only needs to be carried out once, the total computational time essentially depends on the number of PCA bands used in the classification. As can be seen in Table 2, the classification time is decreased when fewer bands are used in the classification. When the first five PCA bands are used, the classification time is seven and four seconds, respectively, for the HYDICE and AVIRIS images, which is about 20.0 percent (5/25, HYDICE) and 23.5 percent (4/17, AVIRIS) of the total time. This result suggests that the PCA transformation will dominate the total processing time in practice, which in our studies is about 43.5 percent (20/46, HYDICE) and 20.0 percent (13/65, AVIRIS) of the time when all original images are used for the classification.

Conclusions

This study reveals that the PCA approach is a useful preprocessing technique for hyperspectral image classification. Among all the obtained PCA bands, the first few (about 5) bands may contain most of the information contained in the entire hyperspectral image data. After the first 10 PCA bands, virtually all other bands contain only noise. Classifications using the most significant PCA bands yield the same class patterns as when entire hyperspectral data sets are used. The correct classification rate increases slowly in a nearly linear manner as more PCA bands are involved in the classification. The use of the most significant 5 (~10 percent) and 10 (~20 percent) PCA bands can lead to correct classification rates of about 70 percent and 80 percent or higher. Misclassifications caused by PCA-induced information loss mainly occur at feature class boarders in the image and are more sensitive for rural than for urban areas. The CPU time for PCA transformation will dominate the entire processing time if the most significant PCA bands are used. All these findings suggest that the use of the PCA approach for hyperspectral image classification is beneficial and effective. It significantly reduces the amount of data to be handled and achieves practically acceptable and accurate classification results that are comparable with those obtained using the entire hyperspectral image data. Future research efforts will include other alternative preprocessing approaches, such as canonical component analysis

and quantitative measures, and the image classification quality based on theses transformations.

ACKNOWLEDGMENT

The AVIRIS image data used in this study was obtained from the issue of ENVI software as a sample data set.

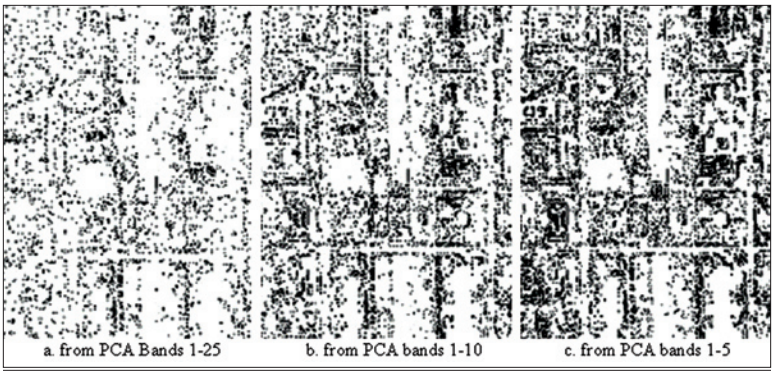


Figure 8. HYDICE misclassified pixels.

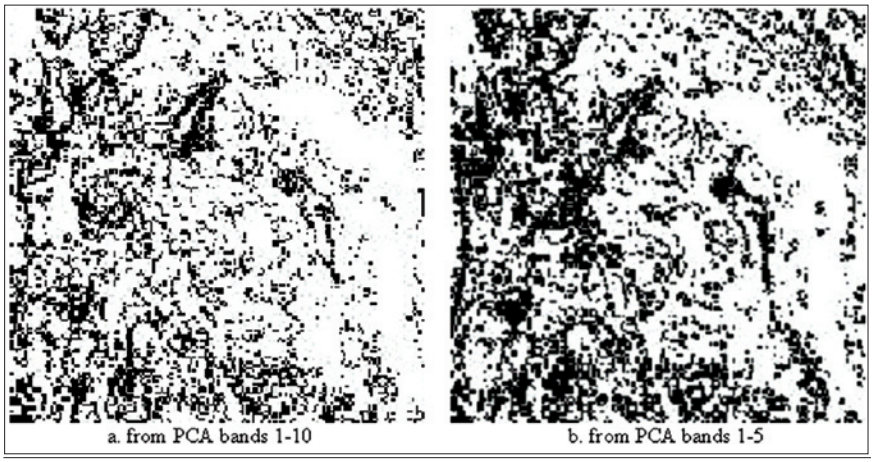
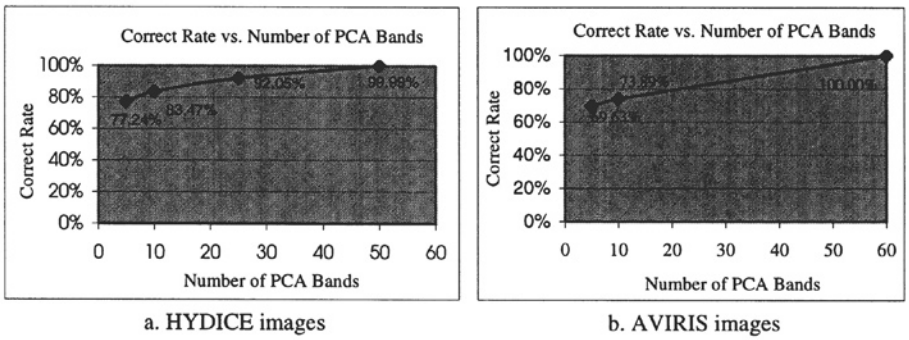


Figure 9. AVIRIS misclassified pixels.



a. HYDICE images b. AVIRIS images

Figure 10. Correct classification rate for PCA band images.

REFERENCES

Campbell, J. 1996. *Introduction to remote sensing*. New York, New York: The Guilford Press. pp.287 - 92.

Carr, J., and K. Matanawi. 1999. Correspondence analysis for principal components transformation of multispectral and hyperspectral digital images. *PE&RS* 65(8): 909-14.

Gonzalez, R., and R. Woods. 1993. Digital image processing. Reading, Massachusetts, Addison-Wesley Publishing Company. pp.148-56.

Hughes, G. 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transaction on Information Theory* IT-14(1): 55-63.

Landgrebe, D. 2000. *Information extraction principles and methods for multispectral and hyperspectral image data*. Chapter 1: Information Processing for Remote Sensing, edited by C. H. Chen. River Edge, New Jersey, World Scientific Publishing Co., Inc.

Lanzl, F., and A. Mueller. 1999. Imaging spectroscopy: Strategic activities at the DLR Institute of Optoelectronics. In: Proceedings of SPIE, *Imaging Spectrometry* V(3753): 25-34.

Lillesand, T., and R. Kiefer. 2000. *Remote sensing and image interpretation*. New York, New York: John Wiley and Sons. pp. 518-23.

Mather, P. 1999. *Computer processing of remotely-sensed images'An introduction*, 2nd ed. New York, New York: John Wiley & Sons. pp. 126-38.

Porter, W., and H. Enmark. 1987. A system overview of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). In: *Proceedings of SPIE* 834:114-26.

Research Systems, Inc., 1999. ENVI users guide and tutorials.

Schowengerdt, R. 1997. *Remote sensing*: *Models and methods for image processing*, 2nd ed. Academic Press. pp.1-11; 179-98.

Shahshahani, B., and D. Landgrebe. 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE *Transactions on Geoscience and Remote Sensing* 32(5): 1087-95.

Wilson, T., and C. Davis. 1999. Naval EarthMap Observer (NEMO) Satellite. In: Proceedings of SPIE, *Imaging Spectrometry* V (3753): 2-11.

Zhang, B., X. Wang, J. Liu, L. Zheng, and Q. Tong. 2000. Hyperspectral Image Processing and Analysis System (HIPAS) and its application. *Photogrammetric Engineering and Remote Sensing* 66(5): 605-9.