Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification

Xiuping Jia, Member, IEEE, and John A. Richards, Fellow, IEEE

Abstract—In this paper, a segmented, and possibly multi-stage, principal components transformation (PCT) is proposed for efficient hyperspectral remote-sensing image classification and display. The scheme requires, initially, partitioning the complete set of bands into several highly correlated subgroups. After separate transformation of each subgroup, the single-band separabilities are used as a guide to carry out feature selection. The selected features can then be transformed again to achieve a satisfactory data reduction ratio and generate the three most significant components for color display. The scheme reduces the computational load significantly for feature extraction, compared with the conventional PCT. A reduced number of features will also accelerate the maximum likelihood classification process significantly, and the process will not suffer the limitations encountered by trying to use the full set of hyperspectral data when training samples are limited. Encouraging results have been obtained in terms of classification accuracy, speed, and quality of color image display using two airborne visible/infrared imaging spectrometers (AVIRIS) data sets.

Index Terms—Feature extraction, hyperspectral, image display and classification, principal components transformation (PCT).

I. INTRODUCTION

WILE the high spectral resolution available with imaging spectrometers, such as the airborne visible/infrared imaging spectrometer (AVIRIS) [1], with 224 channels, provides rich information on ground cover types, the enormous data volumes generated present particular problems for image handling, analysis, and display. Feature reduction, i.e., band reduction, therefore, has become a more significant part of the hyperspectral image interpretation process. For the conventional multispectral data with low dimensionality, feature reduction can be relatively easily achieved by using feature selection techniques, e.g., via finding the subset of features that provide the highest class separability, or by feature extraction techniques, e.g., via linear transformation to enhance class separability with fewer features [2], [3]. However, feature selection will be a time-consuming task for hyperspectral data since the best subset of features cannot be found until an exhaustive search of all the feature subset combinations is carried out. Since the combination of features to be examined increases exponentially, this is an infeasible task. When feature extraction techniques are applied to hyperspectral data, the principal components transform (PCT) outperforms those feature extraction techniques that are based on class statistics [4]. For example, a canonical analysis maps original data to a new feature space which maximizes the ratio of between-class variance to within-class variance. The difficulty is that the class covariance matrices are hardly ever estimated reliably since the number of training pixels is limited and its ratio to the number of dimensions is low for hyperspectral image data. The advantage of using a PCT is that global statistics are used to determine the transformation parameters. Therefore, there is no need to estimate class statistics before feature reduction. However, implementing this technique with a high-dimensional data set requires high computing load.

Let \( \mathbf{x} \) be the brightness vector for an image pixel with \( N \) spectral bands. The PCT is defined as

\[
\mathbf{z} = A^T \mathbf{x}
\]

where \( A \) is the matrix of normalized eigenvectors of the image covariance matrix \( \Sigma_x \) (\( T \) denotes the transpose operation).

PCT provides a new feature space \( \mathbf{z} \), and its first \( k \) components are the best choice of \( k \) linear functions for reconstructing the original data in the sense of minimizing mean square error of the residual [5]. The transformed data set has two main properties that are significant to the application here. The variance in the original data set has been rearranged and reordered so that the first few components contain almost all of the variance in the original data, and the components in the new feature space are uncorrelated [3].

Implementing the transformation consists of two tasks: an eigenanalysis to generate the transformation matrix \( A \) and pixel-by-pixel linear transformation given by (1). The former requires an insignificant amount of work since it is, in principle, independent of the size of the image. However, the latter is a time-consuming process that requires \( N \times N \) multiplications and \( N \times (N - 1) \) additions per pixel. The computation load is, therefore, a major consideration in the case of hyperspectral data transformation; i.e., it is inefficient to transform the complete data set. Moreover, the process can be biased to the high variance bands. For example, the data recorded by AVIRIS are affected in shape by the solar spectrum. This indicates that a spectral weighting is imposed. As a result, the variances of the spectral bands in the short wavelength region are much higher than the remaining bands.
if the data are not calibrated. Therefore, the conventional PCT will be dominated by the visible and near-infrared bands. The proposed scheme will overcome this problem as well as reducing computational time.

The proposed segmented PCT scheme is presented in the next section, followed by a fast feature selection method from the eigenfeatures obtained. Experimental results are presented in Section III.

II. TECHNIQUES

A. Segmented PCT

The segmented PCT scheme is developed based on the following considerations.

Let \( x_l \) be a pixel vector, \( l = 1, 2, \ldots, L \), whose components are the individual spectral responses in each band. The unbiased estimate of the mean vector \( \mathbf{m} \) and covariance matrix \( \Sigma \) are given by [3]

\[
\mathbf{m} = L^{-1} \sum_{l=1}^{L} x_l
\]

and

\[
\Sigma = (L - 1)^{-1} \sum_{l=1}^{L} (x_l - \mathbf{m})(x_l - \mathbf{m})^T.
\]

The correlation matrix \( R \) is related to the covariance matrix, and its elements \( Q_{ij} \) are determined by

\[
Q_{ij} = \frac{\nu_{ij}}{\sqrt{\nu_{ii} \nu_{jj}}}
\]

where \( \nu_{ij} \) are elements of the covariance matrix and \( \nu_{ii} \) and \( \nu_{jj} \) are the variances of the \( i \)th and \( j \)th bands of data. \( Q_{ij} \) describes the correlation between band \( i \) and band \( j \).

It has been observed previously that when the original bands are highly correlated, PCT works efficiently. However, for poorly correlated data, there may be little change after application of the PCT [6]. We have also observed for hyperspectral data the property in which the correlations between neighboring spectral bands are generally higher than for bands further apart, with high correlations appearing in blocks [6], [7]. An example of the correlation matrix of AVIRIS data in image form is shown in Fig. 1.

If the conventional PCT is modified so that the transform is carried out by avoiding the low correlations between the highly correlated blocks, the efficiency of the PCT will be improved. This idea leads to the proposed segmented PCT scheme discussed below.

Fig. 2 shows the multistage segmented PCT process schematically. The complete data set is first partitioned into \( K \) subgroups. Highly correlated bands are selected as subgroups. We denote by \( n_1, n_2, \ldots, n_K \), the number of bands in subgroups 1, 2, \ldots, \( K \), respectively. The PCT is conducted separately on each subgroup of data. Feature selection from the new data set obtained after the first stage transformation is then carried out either by making use of variance information in each component for simplicity or by pursuing single-band separability due to the orthogonal property of the transformed data. The features selected can be regrouped and transformed again to compress the data further. Generally, the steps can be repeated until the required data reduction ratio is achieved for classification or storage purposes, while important information is essentially preserved. The most informative three features will be used for color composite display.

The major advantage of segmenting the PCT is saving computation time. Eigenanalysis time is reduced, but more significantly, the actual transformation time is reduced. For the conventional PCT, the transformation given by (1) requires \( N \times N \) multiplications for each pixel vector, as noted earlier. For the segmented PCT, it requires \( n_k \times n_k \) multiplications in each subgroup and, thus, the total multiplications are \( \sum_{k=1}^{K} n_k^2 \) for each pixel vector. When addition operations are ignored, the saving factor (SF) can be expressed in terms of the number of subgroups \( K \) and the size of each subgroup \( n_k \), as given in (2), shown at the bottom of the page, where \( \sum_{k=1}^{K} n_k = N \).

For example, 2/3 of the total time is saved when three subgroups of uniform size are used (i.e., \( K = 3 \) and \( n_1 = n_2 = n_3 \)).

Highly correlated bands are selected as the subgroups in the proposed scheme so that the PCT works efficiently since it depends on redundancy reduction. Moreover, the difference between the transformed data obtained by conventional PCT and that from the new scheme will be minimized when the new bands corresponding to the high eigenvalues in all subgroups

---

\[
SF = 1 - \frac{\text{no. of multiplications required by segmented PCT}}{\text{no. of multiplications required by conventional PCT}} = 1 - \frac{\sum_{k=1}^{K} n_k \times n_k}{N \times N} = 1 - \frac{1}{N^2} \sum_{k=1}^{K} n_k^2 = 1 - \left( \frac{n_k}{N} \right)^2
\]
are kept. The solar spectrum weighting imposed on each band within a subgroup, corresponding to a narrow region of wavelength, tends to be relatively uniform. Therefore, the bands in a subgroup with similar variance will not suffer from the bias problem that occurs with a conventional PCT.

B. Feature Selection on PCT Data

For the original data, reliable feature selection is based on pairwise separability measures, such as the Bhattacharyya distance [8]

$$BD = \frac{1}{8}(m_i - m_j)^T \left\{ \frac{\Sigma_i + \Sigma_j}{2} \right\}^{-1} (m_i - m_j)$$

$$+ \frac{1}{2} \ln \left( \frac{(\Sigma_i + \Sigma_j)^{1/2}}{\Sigma_i^{1/2} \Sigma_j^{1/2}} \right)$$

where $m_i$, $\Sigma_i$, and $m_j$ are the mean vectors and covariance matrices, respectively, for classes $i$ and $j$. If $\Sigma_i$ and $\Sigma_j$ are diagonal matrices (i.e., the data of classes $i$ and $j$ are uncorrelated), the above equation becomes

$$BD = \sum_{n=1}^{N} \left[ \frac{1}{4}(m_i(n) - m_j(n))^2 (\sigma_i^2(n) + \sigma_j^2(n))^{-1} + \frac{1}{2} \ln (\sigma_i^2(n)/2 + \sigma_j^2(n)/2)(\sigma_i^2(n)\sigma_j^2(n))^{-1/2} \right]$$

III. EXPERIMENTS AND RESULTS

The data sets used in this work were recorded by AVIRIS over the standard scenes of Jasper Ridge (1989) and Moffett Field (1989). The data sets have been reduced from the original 224 bands to 196 bands in Jasper Ridge and 187 bands in Moffett Field, after overlapping bands, water absorption bands, and the bands that are totally zero or have very small means (i.e., <2) were removed.

A. Comparison of the Segmented and Conventional Transformation

A portion of image data, i.e., every fifth column and fifth row in the 256 x 256 Jasper Ridge image was used in the following tests. The conventional PCT, which transforms the complete set of 196 bands, was run first as a reference. This is referred to as T196. A segmented PCT was examined and compared with the conventional PCT. The segmentation was based on the results obtained by, first, considering only correlations whose absolute value exceeds a threshold of 0.5 and, then, searching for edges in the “image” of the correlation matrix (Fig. 1) [6]. As a result, the complete set of bands is divided into three subgroups, as given in Table I. The size of each subgroup is 35, 58, and 103 bands, respectively. Fig. 3 shows that the average correlations within these three subgroups (three diagonal blocks) are much higher than those of the off-diagonal segments. This segmentation is referred to as T35/58/103.

The covariance matrix was calculated for the complete set of bands, and an eigenanalysis was made on the full covariance matrix and the corresponding blocks for T35/58/103, respectively. In order to examine the similarity of the segmented PCT to the conventional PCT, the eigenvalues obtained from each group were sorted by descending order. The cumulative
TABLE I  
SEGMENTATION OF THE COMPLETE SET OF BANDS

<table>
<thead>
<tr>
<th>Images</th>
<th>Subgroup 1</th>
<th>Subgroup 2</th>
<th>Subgroup 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jasper Ridge</td>
<td>Bands 1 - 35</td>
<td>Bands 36 - 93</td>
<td>Bands 94 - 196</td>
</tr>
<tr>
<td>Moffett Field</td>
<td>Bands 1 - 33</td>
<td>Bands 34 - 91</td>
<td>Bands 92 - 187</td>
</tr>
</tbody>
</table>

![Fig. 3. Average correlations within diagonal blocks and within selected off-diagonal segments of Fig. 1.](image)

The cumulative eigenvalues for the new order are plotted in Fig. 4. That graph shows that, while cumulative eigenvalues from the segmented PCT are slightly lower than the conventional PCT for the first ten PC’s, they become very close after that. When the first 36 PC’s are considered, the difference in cumulative eigenvalues obtained from T35/58/103 and T196 is only 0.1%. The computation time, however, is reduced by 60% using the segmented PCT (T35/58/103).

B. Feature Selection and Classification

Seven and nine cover types were selected visually from the Jasper Ridge and Moffett Field images, respectively, using MultiSpec (a Macintosh version of the LARSYS software package) for this classification exercise.

The complete set of bands was segmented into three highly correlated subgroups for each image, as given in Table I. Each subgroup’s global statistics, i.e., mean vector and covariance matrix, were found using the selected samples from the image (every fifth row and fifth column). An eigenanalysis was then conducted to generate the eigenvectors, which were used in the PCT in the first-stage processing.

The features corresponding to high average separability (BD > 1.0) are selected for classification. They are PC1, PC3, PC5 from subgroup 1, PC1, PC2, PC3 from subgroup 2, and PC1 from subgroup 3 for the Jasper Ridge image and PC1 from subgroup 1, PC1, PC2, PC3 from subgroup 2, and PC1, PC2 from subgroup 3 for the Moffett Field image. Consequently, seven PC’s for the Jasper Ridge image and six PC’s for the Moffett Field image are kept.

![Fig. 4. Comparison between cumulative eigenvalues.](image)

The substantial feature reduction after the first-stage block-based transformation, from 196 to seven for the Jasper Ridge image and from 187 to six for the Moffett Field image, discussed above, suggests that there is no need to carry out the second-stage PC transform for classification purposes. Maximum likelihood classification using the selected features will be significantly accelerated compared with the full data set, as the computational demand in this step is equivalent to processing just a thematic mapper (TM) data set. The numbers of training pixels are adequate even for the smallest class; “road” in the image of Jasper Ridge, in which only 108 training pixels are available. This would cause a problem in statistical estimates of the class signature if the original hyperspectral data set is used [9], [10].

Classification into the seven and nine classes defined for the two images, respectively, were conducted using the selected features. The overall classification accuracies are 98.6 and 97.0% for the Jasper Ridge and Moffett Field images, respectively. These results are encouraging since, with only seven (for the Jasper Ridge image) or six features (for the Moffett Field image), satisfactory classification performance is achieved.

C. Color Composite Image Display

Normally, wavebands corresponding to visible green, red, and near infrared are used to compose a color image. Those three bands contain, however, only a small percentage of the total variance of a hyperspectral data set. In order to compress the information into three channels for color display, the features selected after the first-stage PCT, which are regarded as maximally information-bearing data because of their relatively strong separabilities and high variances, are used again for transformation. The resulting first three components, which are uncorrelated, contain as much as 99% of the variance for both the images studied. The image displayed using these
three components for red, green, and blue is richly colored, as expected.

IV. DISCUSSIONS AND CONCLUSION

This scheme is a practical and efficient method for dealing with hyperspectral data. It makes use of the block structure of the correlation matrix so that the PCT is conducted on data of smaller dimensionality. Therefore, computational load is reduced significantly. Moreover, the solar spectrum weighting imposed on each band within a subgroup, corresponding to a narrow region of wavelength, tends to be relatively uniform so that all bands in a subgroup with similar variance will not experience the bias problem that occurs with a conventional PCT.

Single-feature separability may be used as a guide for feature selection. A reduced number of features will also accelerate the maximum likelihood classification process significantly.

High-quality color composite images can be formed by the three most significant and decorrelated components, which efficiently accommodates color image display, providing the maximum visual impact for photointerpretation and training field selection.

ACKNOWLEDGMENT

The authors wish to thank Dr. D. Landgrebe of the School of Electrical Engineering, Purdue University, for providing the MultiSpec software package.

REFERENCES


Xiuping Jia (M’93) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and the Ph.D. degree in electrical engineering from The University of New South Wales, Canberra, Australia, in 1996. She has been with the School of Electrical Engineering, University College, Australian Defence Force Academy, The University of New South Wales, since 1988. She is currently an Associate Lecturer with research interests in remote sensing and image spectrometry.


Dr. Richards is a Fellow of the Australian Academy of Technological Sciences and Engineering, a member of the Fulbright National Selection Committee for Australia, and on the Editorial Board of