World Applied Sciences Journal 3 (5): 785-796, 2008
ISSN 1818-4952

Projection Pursuit and Lowpass Filtering for Preprocessing of Hypespectral Images

Bahram Salehi, Mohammad Javad Valadan Zoej and Masood Varshosaz

K.N. Toosi University of Technology, Mirdamad Cr., Vali-Asr St., Tehran, Iran

Abstract: Hyperspectral data potentially contain more information than multispectral data because of their higher spectral resolution. However, the stochastic data analysis approaches, successfully applied to classification of multispectral data, are not as effective as those for hyperspectral data. Various investigations indicate that the key problem causing poor performance in the stochastic approaches to hyperspectral data classification is inaccurate class parameters estimation. It has been found that the conventional approaches can be retained if a preprocessing stage is established before feature extraction stage in the classification process. This paper, presents a combined preprocessing algorithm which includes dimensionality reduction followed by class separability improvement. For the dimensionality reduction, the Sequential Parametric Projection Pursuit was used because of its special characteristics. For class separability improvement, a Lowpass filter was used. Experimental results showed that applying such a combination, improves the classification accuracy as compared with the case where either a dimensionally reduction or a class separability improvement algorithm is used individually.

Key words: Hyperspectral . projection pursuit . lowpass filter . preprocessing . feature extraction . classification

INTRODUCTION

Hyperspectral data potentially contain more information than multispectral data because of their higher spectral resolution. However, the stochastic approaches applied to hyperspectral data analysis do not provide as accurate results as when applied to multispectral data analysis. This is because stochastic approaches lead to inaccurate classification performance as the dimensionality (i.e. the number of spectral bands) increases [1]. In the stochastic approaches, the characteristics of a class are modelled with a set of training samples [2]. Hughes [3] showed that if the number of training samples is finite and fixed, the accuracy of statistical parameter estimation decreases as the dimensionality increases, leading to a decline in the classification accuracy. Although increasing the number of spectral bands potentially provides more information about class separability, this positive effect is diluted by inaccurate parameter estimation [1]. As a result, if the number of training samples is finite and remains constant, the classification accuracy first grows and then declines as the number of spectral bands increases (Fig. 1). This is often referred to as the Hughes phenomenon [3].

In order to increase the accuracy of statistical parameter estimation, Jimenez *et al*. [4] and Hsieh *et al*. [1] suggested a preprocessing stage before the feature extraction stage in classification of hyperspectral image data. Fig. 2 shows the steps of classifying hyperspectral data.

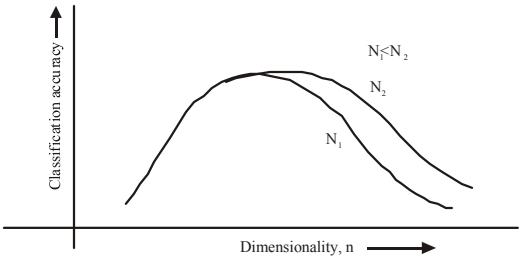


Fig. 1: Conceptual presentation of the hughes phenomenon (N is the number of samples) [1]

The preprocessing stage should be designed in a way that the classification accuracy increases. In general, the classification accuracy depends on four factors which are class separability, dimensionality, training sample size and classifier type [5].

To date, various investigations have been carried out to increase the classification accuracy using the above factors. Hsieh *et al*. [1] showed that applying a lowpass filter increases the separability of image classes, which in turn leads to an increment in the classification accuracy. Jimenez *et al*. [4] used a statistical algorithm, so called the Projection Pursuit (PP), to reduce the dimensionality of hyperspectral

Corresponding Author: Dr. Bahram Salehi, K.N. Toosi University of Technology, Mirdamad Cr., Vali-Asr St., Tehran, Iran

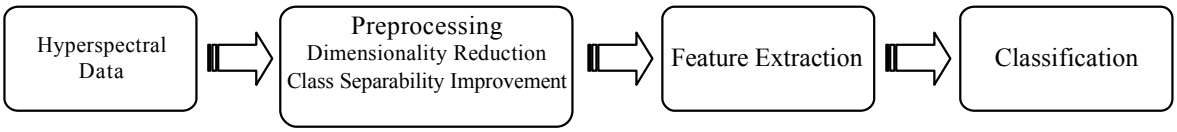


Fig. 2: Sequence of analysis step for hyperspectral data classification

image data. Also Vaiphasa *et al*. [6] used a hyperpectral band selector that falls into the category of dimensionality. For the factor "training sample size", Shahshahani *et al*. [7] showed that by adding un-labeled samples to the classifier design process the accuracy of the classification can also increase. Finally, Friedman [8], Hoffbeck *et al*. [9] and Bandyopadhyay [10] argued that the type of classifier has a reasonable effect on the accuracy of the classification.

The objective of this paper is to propose the combination of the first two factors, i.e. dimensionality reduction and increasing class separability, as the two steps of the preprocessing stage (Fig. 2). For the first step a Projection Pursuit (PP) and for the second step a LowPass Filter (LPF) algorithm was used. To demonstrate the applicability of such a combination, the effect of each individual factor on the classification accuracy was evaluated first. Then, the accuracy of combining them on the classification accuracy was studied.

The Hughes phenomenon indicates that by reducing the dimensionality, the class separability decreases, leading to a decrease in the classification accuracy. To mitigate this negative effect, the second step in the preprocessing stage in Fig. 2 is to increase the class separability in a new lower dimensional space resulting from the first step.

The rapid increase with dimensionality in training sample size required for density estimation has been termed the "curse of dimensionality" by Bellman [11] which leads to the Hughes phenomenon in a classifier design. The curse of dimensionality for the high dimensional data analysis using statistical methods has been known for decades. Many feature extraction algorithms have been developed and implemented for solving this problem. Some well known and widely used algorithms include principal component analysis (PCA) [12, 13] discriminant analysis feature extraction (DAFE) [14] and decision boundary feature extraction (DBFE) [15]. A number of problems associated with these algorithms are discussed in Jimenez *et al*. [4] and Landgrebe [2], the most important of which are the computational performance in full dimensionality and the Hughes phenomenon. Therefore, in order for a feature extraction algorithm to have accurate results, the dimensionality of the hyperspectral image has to be reduced. For this, the data from the high dimensional space (original image bands) are to be transferred to a lower dimensional space with fewer number of bands, where the feature extraction process is, then, carried out.

An ideal dimensionality reduction scheme should be able to consider the redundancy between the spectral bands and to avoid the computations in the high dimensional space in order to reduce the required number of training samples and to take advantage of the high dimensionality of hyperspectral data [4]. For this purpose, in this paper parametric projection pursuit [4, 16] was used and implemented because of its special characteristics. Projection Pursuit (PP) projects the total number of bands of a hyperspectral image into several subspaces, accomplishing the computations of parameter estimation in these subspaces [4]. As a result, the Hughes phenomenon is expected to be reduced and, thus, the classification accuracy to increase. The method used to increase the class separability is a lowpass filter in the lower space resulting from the previous step.

In the following sections, the PP and the LPF algorithms used and implemented in this research are discussed further. The results of tests carried out to evaluate the performance of each of these algorithms alone, along with the combined technique proposed in this paper, are discussed and conclusions are made.

MATERIALS AND METHODS

In this section, the PP algorithm which is used here for reducing the dimensionality of hyperspectral image data is described. As will be shown, the PP algorithm has different approaches, the best of which is defined. In addition, the LPF and its effect on the class separability is discussed as well.

Projection pursuit and dimensionality reduction: Friedman and Tukey [17] introduced the term projection pursuit for a technique for exploratory analysis of multivariate data sets. The method seeks out "interesting" linear projection of the multivariate data onto a lower dimensional subspace [16]. For the first time, projection pursuit was used by Jimenez *et al*. [4] in order to reduce the dimensionality of hyperspectral data and is briefly described in the following lines adapted mainly from Jimenez *et al*. [4] and Lin *et al*. [18].

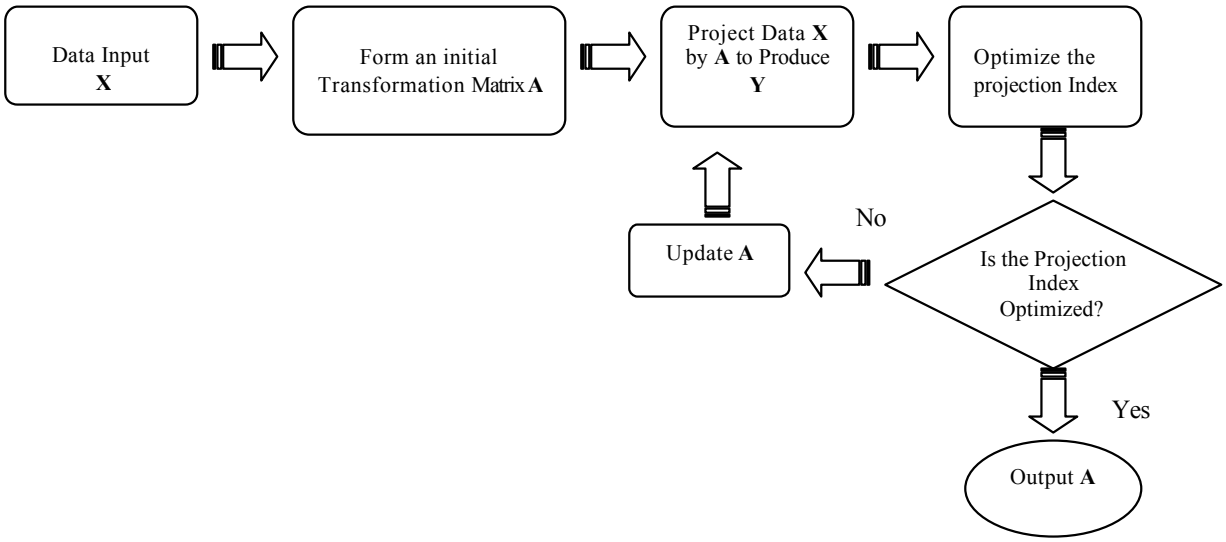


Fig. 3: Block diagram for projection pursuit algorithm

The hyperspectral data set for each class forms a $D \times N$ matrix X, where D denotes the spectral dimensionality; N is the number of pixels and X represents the data set in the original space. If A is a transform matrix of dimension $D \times M$, where M is the number of projections then the transformation procedure is done by matrix A as:

$$Y = A^T X \quad (1)$$

where Y is an $M \times N$ matrix, which is the orthogonal projection of the hyperspectral data onto a new M-dimensional coordinate system. The axes of the new coordinate system lie in the direction of a linear combination of the original coordinate, D. In this case; the columns of A are required to be mutually orthogonal. If H is a function measuring the efficiency of a sample in the projected subspace, then the function $H(A^T X)$ is referred to as the projection index; and projection pursuit attempts to find the transformation matrix, A, which produces a local optima of $H(A^T X)$ by numerical optimization. Once the optimized transform matrix A is obtained, the original data can be projected onto a lower dimensional space. The transform matrix, A, reduces the dimensionality of the original data from D to M. As a result, the transformation reduces the required number of training samples and thus mitigates the Hughes phenomenon for an accurate classification to overcome the curse of dimensionality. The steps and details of the projection pursuit algorithm are shown in Fig. 3.

As can be seen in Fig. 3, the optimization of projection index is one of the steps in the projection pursuit dimensionality reduction algorithm. The main issue in the projection pursuit algorithm is the definition of the projection index [4]. Such a projection index needs to be defined such that by its optimization within the PP algorithm, the "interesting" projections can be selected [4, 16]. What "interesting" means depends on what function or projection index one uses.

In remote sensing data analysis, "interesting" would certainly be a projection which separates data into different meaningful clusters which are exhaustive, separable and of information value [4, 19].

To date, a number of non-parametric projection indices have been proposed. Among these are the Friedman-Tukey index [16], negative Shannon entropy [12] and the Moment index suggested by [20]. These nonparametric projection indices require a lot of training samples in order to estimate the statistical parameters in their algorithms [4]. Therefore, to solve this problem, researchers have proposed some parametric indices. Parametric indices use *a prior* information in the form of training samples to carry out their supervised calculations [4]. Among them are the Divergence distance [19], Fisher criterion and Bhattacharyya distance [4].

A major problem with Divergence and Fisher indices is that they do not have a linear, one to one relationship with classification accuracy [2]. This is the main reason why these indices were not used as projection indices in this work; instead the Bhattacharyya distance was used as the projection index because of its special characteristics.

The Bhattacharyya distance is a special case of Chernoff distance [2, 14]defined as:

$$B = -\ln \int_x \sqrt{P(x|\omega_i) P(x|\omega_j)}\, dx \quad (2)$$

Where $P(x|\omega_i)$ and $P(x|\omega_j)$ are the values of ith and jth class probability distribution at the position x. When

the distribution function of classes is normal, this distance is computed by [2, 14]:

$$B_{ij} = \frac{1}{8}(M_i - M_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2}\right]^{-1} (M_i - M_j) + \frac{1}{2} Ln \frac{\left|\frac{1}{2}[\Sigma_i + \Sigma_j]\right|}{\sqrt{|\Sigma_i||\Sigma_j|}} \quad (3)$$

where M_i and M_j are the means of classes i and j and Σ_i and Σ_j are the covariance matrices of those classes. B_{ij} is the Bhattacharyya distance between the two classes. In case the number of classes is more than two, the minimum Bhattacharyya distance among the classes could be used.

The Bhattacharyya distance fulfills the requirements of a projection index to be "interesting" due to some of its inherent properties. Unlike the Fisher and Divergence indices, Bhattacharyya distance has a nearly linear, nearly one-to-one relationship with classification accuracy [2]. Also it estimates the parameters of each pair of classes separately; thus it is class specific. The other advantage of Bhattacharyya distance is that it is constructed from two terms, due to the mean and covariance of classes. This illustrates, for example, that two classes can have the same mean value and still be quite separable, in which case the first term is zero [21]. On the basis of these arguments and in view of empirical results [4], in the projection pursuit algorithm the Bhattacharyya distance is preferred to the Fisher criterion and Divergence index as a projection index in this research. Because of the use of parametric index (Bhattacharyya distance) in PP algorithm, it is called parametric projection pursuit (PPP).

In the PPP algorithm, the matrix A is used to project the data set X, from the original dimension onto a lower dimensional subspace, Y as $Y = A^T X$. In the matrix A, the number of columns corresponds to the number of groups into which the entire features in the original space are projected. The columns of matrix *A* are orthogonal to each other because the PPP seeks to find an orthogonal basis onto which the data are projected. Therefore, the entries of each column are zero except the position of the corresponding group of adjacent bands. The form of matrix A is shown in equation (4) [18].

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \\ a_{1n_1} & 0 & & & \\ 0 & a_{21} & & 0 & \vdots \\ & \vdots & & a_{2n_2} & \\ & a_{2n_2} & & \vdots & \\ \vdots & 0 & & a_{k1} & 0 \\ & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & a_{kn_k} & a_{kn} \end{bmatrix} \quad (4)$$

Each group of adjacent bands in equation (4) will be initialized with a bank of estimated guesses for the a_i. If the initial number of groups is assumed to be given, then the construction of matrix A is to choose one estimated guess a_i from the corresponding vector bank. There are four initial guesses in each bank. The first and second are based on the assumption that the mean and covariance difference between class 1 and class 2 is respectively dominant in the Bhattacharyya distance. The third initial guess is a vector that averages all the features in each group of adjacent bands in the original space. The fourth initial guess for the a_i is a vector that selects only one feature in the *i*th group.

As mentioned above the main goal of PP is the optimization of the projection index in order to find the "interesting" projections. In this respect, Jimenez *et al*. [4] proposed two techniques for the optimization of the projection index which are parallel parametric projection pursuit (PPPP) and sequential parametric projection pursuit (SPPP). In PPPP, for each group of adjacent bands, the projection index is computed separately, by optimization of which a feature is extracted for each group. Jimenez at al. [4] demonstrated that the accuracy of PPPP is lower than that of SPPP, as the projection indices are computed independently in each group without incorporating other groups. Therefore, they proposed SPPP, in which the linear combinations of adjacent bands are calculated in a way that optimizes the global projection index in the projected subspace [22]. The global projection index is the Bhattacharyya distance of the entire projected data set instead of the Bhattacharyya distance of each group of adjacent bands [18].

In order to estimate the number of groups and the number of adjacent bands in each group, a top-down binary decision tree algorithm developed by Jimenez *et al*. [4] was used in this research. This algorithm, starts by considering the total number of initial bands as a group. It then continues by using a series of binary decisions and ends when it reaches the maximum number of features established by the analyst. Alternatively, the process stops when the increasing rate of Bhattacharyya distance reaches a certain threshold.

The whole process of the SPPP algorithm can be summarized as follows. First, the total number of spectral bands in the original data space is divided into *G* groups of adjacent bands; *G* is the initial number of groups which depends on the minimum number of training samples in each class. Then the transformation matrix A, with characteristics described earlier, with *G* columns is created. Now the optimization of A is started with the first group of adjacent bands. That is, to replace non-zero elements in the first column of A with

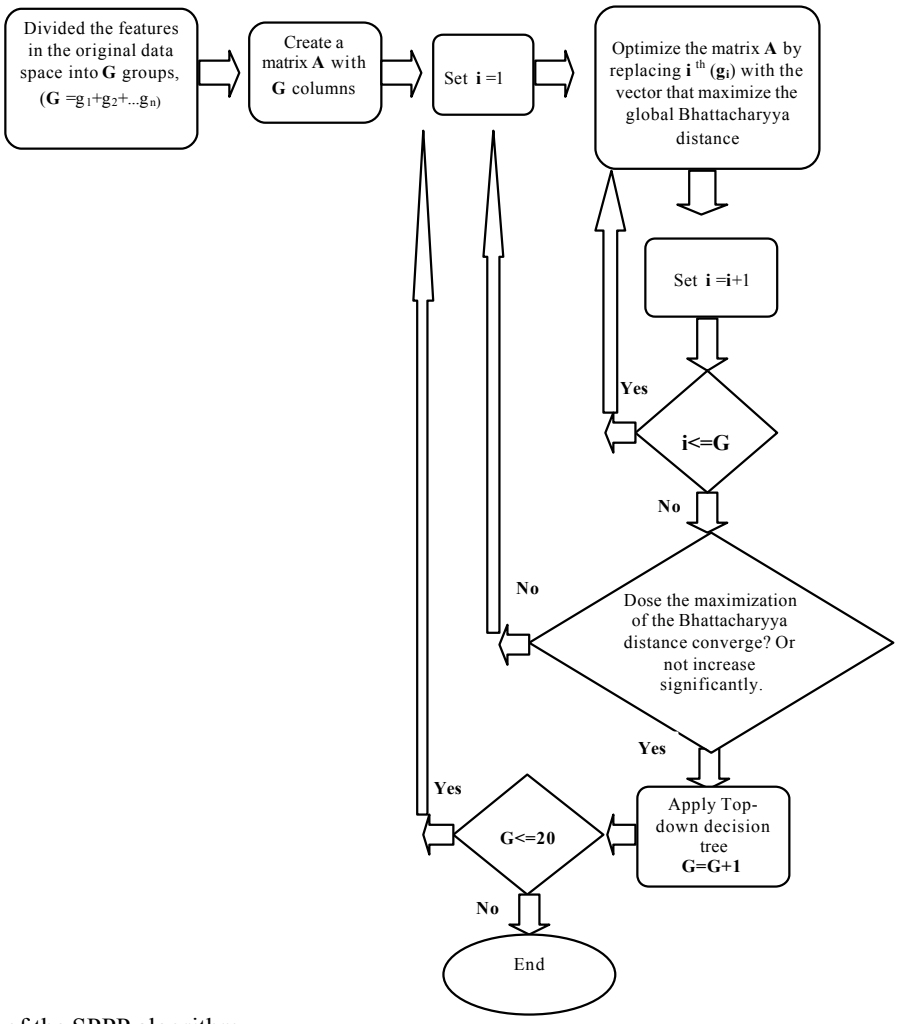


Fig. 4: Processes of the SPPP algorithm

the vector that maximizes the Bhattacharyya distance, while holding the remaining columns of A unchanged. The choices for the vector are the four groups of vectors described earlier. This procedure is repeated for each of the remaining columns of A. Again, the optimization of A is repeated until the increase of Bhattacharyya distance is either below a preset threshold or no longer increases. In this step, by applying the top-down decision tree, the number of groups (*G*) increases by one. The procedure is repeated, this time by new created matrix A with *G+1* columns. These steps are repeated until the number of columns of A becomes equal to 20. Figure 4 shows the procedure of SPPP algorithm in brief.

Lowpass filter and class separability: Class separability is invariant under any nonsingular transformation [14]. However, any singular transformation maps X onto a lower dimensional Y (e.g. projection pursuit), losing some of the classification information [14]. Moreover, the class separability decreases when the dimensionality is reduced by PP. As a result a technique to mitigate this negative dimensionality reduction effect is required. Increasing the class separability compensates for the loss of classification information caused by the PP dimensionality reduction algorithm.

In general, the lowpass filter is used to smooth the image in the field of image processing. However, in this paper, as proposed by Hsieh *et al*. [1], this filter is used to increase class separability. This filter is further described based mainly on the text adapted from Hsieh *et al*. [1].

The lowpass filter is a spatial averaging operator [23]. Assuming $X(i,j)$ is the value of a pixel whose spatial coordinates are i and j, the new value after applying a lowpass filter with a window size of w at spatial location (i,j), is Y where [25]:

$$Y(i,j) = \sum_{(k,l) \in W} c_{kl} X(i+k, j+l) \quad (5)$$

In this equation, c_{kl} is the average of the weights within the window. To simplify the implementation process, c_{kl} is considered to be equal for all weights [1]; therefore $c_{kl} = 1/w$.

If the samples $X(i-k, j-l)$ with $(k,l) \in W$ within the window are independent and identically distributed random vectors with the normal density $N(M_X, \Sigma_X)$, then the lowpass filtered sample Y obtained from equation (5) would possess a normal density $N(M_Y, \Sigma_Y)$ where [25] :

$$M_Y = M_X \qquad \sum_Y = \frac{1}{w}\sum_X \tag{6}$$

As pointed out earlier, the Bhattacharyya distance is used as a measure of class separability because of its special characteristics. From equation (3) it is observed that the Bhattacharyya distance between two classes, 1 and 2, has been constructed from two terms which are [1]:

$$B_X = B_{1X} + B_{2X} \tag{7}$$

$$B_{1X} = \frac{1}{8}(M_{1X} - M_{2X})^T\left[\frac{\sum_{1X} + \sum_{2X}}{2}\right]^{-1}(M_{1X} - M_{2X})$$

$$B_{2X} = \frac{1}{2}\ln\frac{\left|\frac{1}{2}\left[\sum_{1X} + \sum_{2X}\right]\right|}{\sqrt{\left|\sum_{1X}\right|\left|\sum_{2X}\right|}} \tag{8}$$

The first and the second terms in equation (7) represent the class separability due to the mean and covariance differences respectively. By applying a lowpass filter to the image, the Bhattacharyya distance from equation (8) becomes [1]:

$$B_Y = B_{1Y} + B_{2Y} = wB_{1X} + B_{2X} \tag{9}$$

where the first term, $B_{1Y} = wB_{1X}$, indicates that the class separability increases by w times if $M_{1X} \neq M_{2X}$ due to the mean difference. The second term $B_{2Y} = B_{2X}$ shows that the class separability is not affected by the lowpass filter due to the covariance difference.

In this paper, in order to show the effect of lowpass filter on class separability a TM sensor image (Fig. 5) from the Environment for Visualizing Images (ENVI 4.2) software samples was used. For this, the two bands of the feature space of the image were plotted before and after applying the lowpass filter with the window size 3 (Fig. 6).

As can be seen in Fig. 6, after applying the lowpass filter on the image, the variation of each class is reduced; consequently, the gap between classes in the

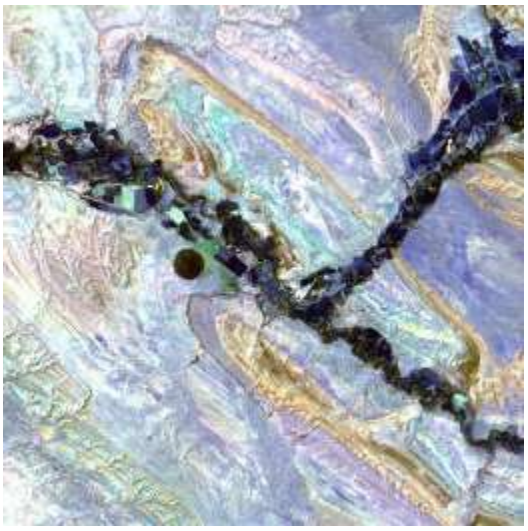
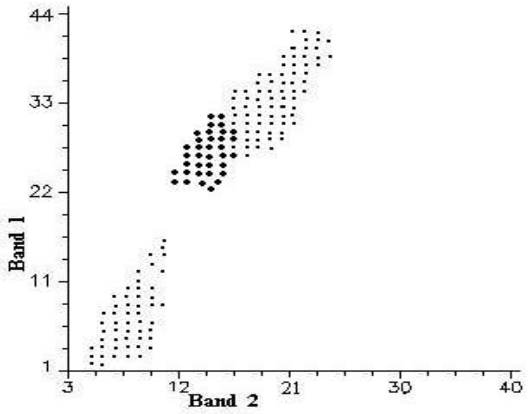
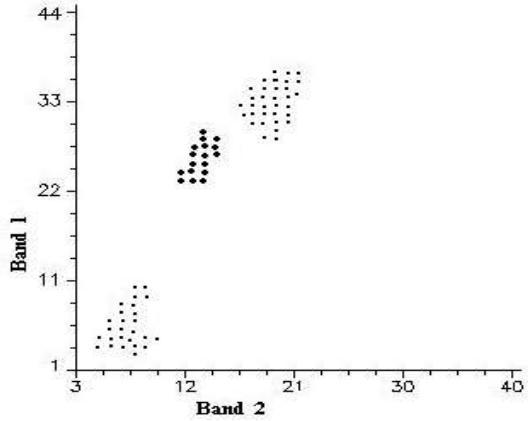


Fig. 5: TM sensor image in bands 1 as red, 3 as green and 5 as blue



a) Before applying the lowpass filter



(b) After applying the lowpass filter

Fig. 6: Scatter plot of three classes before and after applying the lowpass filter for a TM sensor image

Table 1: Number of training and test samples for each class

Class name	Training samples No. of pixels	Test samples No. of pixels
Grass/pasture-mowed Wood	373	2468
Soybeans no-till	239	1294
Corn no-till	197	968
Corn	208	1434
Soybeans-clean	197	714
Corn-min	205	614
Grass/pasture	220	836
Grass/trees	197	497
Hay-windrowed	217	747
	207	489
Total	2260	10061

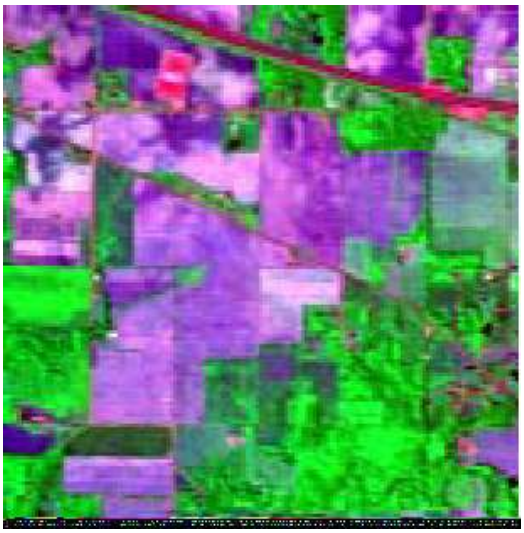


Fig. 7: AVIRIS Hyperspectral test data

feature space is widened, leading to higher class separability.

A drawback of applying a lowpass filter on the image is its spectral mixing effect. In other words, by applying the lowpass filter the small classes become mixed as the spatial resolution is lost. Therefore, the lowpass filter is to be used when the image consists mainly of homogenous large classes.

Hyperspectral test data: The hyperspectral data set, used in this research is the one used by Landgrebe and his group [1, 2, 4] for several years and it was available on the internet. The data had been obtained in June 1992 and was a segment of Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data of north-west Indiana's Indian Pine test site. Figure 7 shows this portion of AVIRIS image in bands 10 as red, 60 as green and 135 as blue. The data set consists of 145× 145 pixels of the AVIRIS image in 195 spectral bands at 10 nm interval in the spectral region from 0.40 to 2.45 μm at a spatial resolution of 20 m and covering an agricultural area. The AVIRIS sensor collects data in 224 spectral bands. Four of the 224 AVIRIS bands contain no data or zero values. From the 220 spectral bands, 195 were used, discarding the water absorption bands.

As mentioned above, the image was obtained in June, by which time most of the crops in the agricultural portion of the test site had not reached their maximum ground cover. This in turn, leads to inaccurate classification result as the image includes not only crops but also variations in the soil type, soil moisture and previous crop residuals [2].

In the experiment, ten classes were defined: corn, corn no-till, soybean-clean, soybean no-till, wood, grass/pasture-mowed, corn-min, grass/pasture, grass/trees and hay-windrowed. No-till, min and clean are three levels of tillage indicating a great, moderate and small amount of residue of previous year's crops, respectively. The total number of training samples was 2260. Table 1 shows the number of training samples and test samples for each class.

RESULTS AND DISCUSSION

The aim of the experiments discussed in this section was to evaluate how the preprocessing stage (Fig. 2) increases the classification accuracy. As the first step of preprocessing (i.e. the dimensionality reduction) the data were projected from the original dimension to a lower dimensional subspace by the SPPP algorithm. Then a lowpass filter with a window size 3 was applied to the image in order to increase the class separability. To evaluate the class separability the Bhattacharyya distance was used. Having performed the preprocessing, two different feature extraction techniques were used to compare the results. These are DAFE [14] and DBFE [15]. Because many researchers [1, 2, 4] have used these techniques as supervised feature extraction in their research, they were adopted here in order for the results of this research to be comparable to those performed by the others. Once the features were extracted, the data were classified by the maximum likelihood (ML) [14] classifier. The ML classifier was used because it is the most conventional parametric classifier which uses first and second order statistics in its algorithm. Finally, a field test was done to evaluate the classification accuracy of different methods; the overall accuracy vs., the number of features in each method were plotted. In the following, the results of the mentioned tests are presented and discussed.

Table 2: The partition group of adjacent bands for different numbers of features and their corresponding Minimum and incremental percentage of Bhattacharyya distance in SPPP algorithm

Number of features	Numbers of adjacent bands per feature and listed in increasing order of wavelength	Minimum Bhattacharyya Distance (MBD)	Increment percentage of MBD (ΔBI_i)
1	195	0.0188	
5	12 12 24 49 98	0.5069	45%
10	12 12 6 6 6 6 49 49 24 25	1.0290	6%
15	12 6 6 6 6 6 6 49 24 25 12 12 6 6 13	1.3876	8%
20	12 6 6 6 6 6 3 3 24 25 12 12 6 6 13 12 12 6 6 13	1.7644	3%

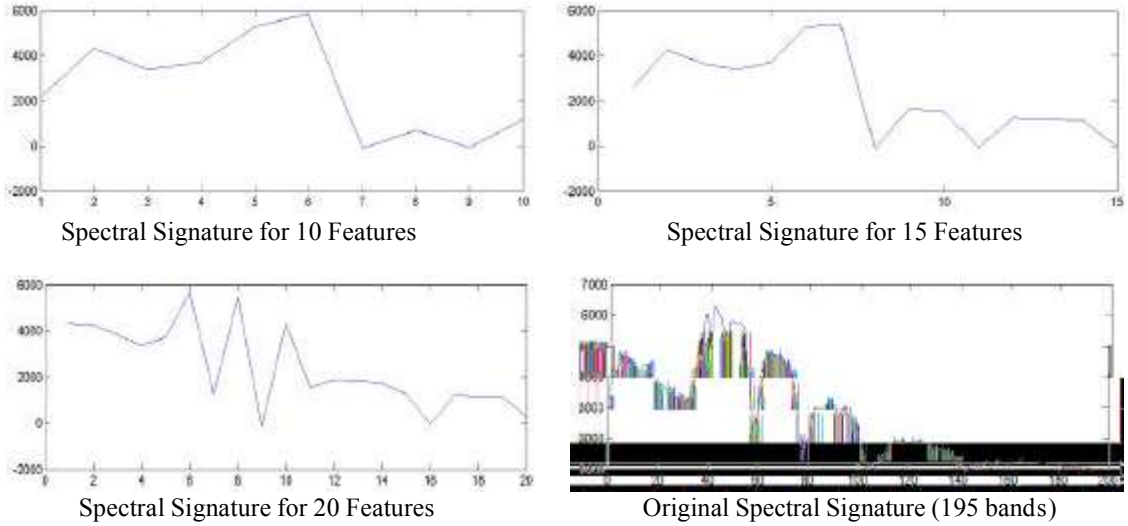


Fig. 8: Spectral signature in high dimension (195 bands) and in low dimension (10, 15and 20 features selected by decision tree in SPPP algorithm) for Corn class

SPPP for dimensionality reduction: In this section, the results of applying the SPPP algorithm on the hyperspectral test data to reduce its dimensionality are presented. The algorithm projects every group of adjacent bands into one feature and the final number of features is the dimensionality of the projected subspace. As discussed earlier, the top-down decision tree was used to estimate the number of adjacent bands in each group. Table 2 shows the partition groups of adjacent bands combined to create features and their corresponding minimum and incremental percentage of Bhattacharyya distance for different set of features (i.e.1, 5, 10, 15 and 20). In Table 2, the incremental percentage of Minimum Bhattacharyya Distance (MBD) for each set of features is calculated with respect to its previous set. For example the 45%, is the incremental percentage of MBD for the case of 5 features compared to that of 4 features. As can be seen, in this table when the number of features increases, the corresponding MBD increases as well. The maximum value of MBD for up to 20 features is 1.7644 and it occurs in the 20 features case. Also, the incremental percentage of MBD decreases with a rise in the number of features. This increment is about 45% and 3% for 5 and 20 features, respectively. This means the increment is very low when the number of features is close to 20. Thus the class separability does not increase significantly for more than 20 features. This is the reason for the algorithm stopping when the number of features is 20 (Table 2).

An example of actual signature of one class (Corn) for 195 bands of the image and the different set of features selected by the top-down decision tree in the SPPP algorithm is shown in Fig. 8. As the number of features increases, the shape of the spectral signature becomes more similar to that of the original one. It could be explained that, by increasing the number of features, the number of adjacent bands which are combined to create one feature is reduced; consequently more important information of the class is maintained. Therefore, the shape of the spectral signature in the projected subspace becomes more similar to that of the original one within the original space.

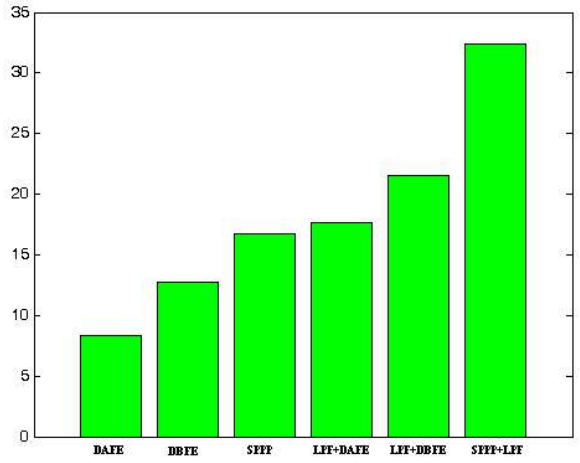


Fig. 9: Average Bhattacharyya distance among the classes for DAFE, DBFE, SPPP, LPF+DAFE, LPF+DBFE and SPPP+LPF

Impact of SPPP and LPF on the class separability: Figure 9 shows the average Bhattacharyya distance as a measure of separability between the 10 classes defined for the test image for different algorithms in the 20 features space. The Bhattacharyya distance was the average of 45 Bhattacharyya distances where 45 refers to the number of 2 class combinations within the 10 classes. In Fig. 9, the DAFE and DBFE are the results of projection from 195 bands to 20 features by discriminant analysis and decision boundary feature extraction respectively. Likewise, LPF+DAFE and LPF+DBFE are the results of the lowpass filter applied to the image followed by the DAFE and DBFE algorithms, to extract 20 features from 195 original bands. SPPP is the result of reducing the 195 bands to 20 features by SPPP algorithm while SPPP+LPF is the result of dimensionality reduction using SPPP followed by applying the lowpass filter on the image. The window size of the lowpass filter is 3.

As can be seen in Fig. 9, applying the lowpass filter to the image increases the average Bhattacharyya distance in the case of DAFE from 8.3 (B_X in Eq.7) to 17.7 (B_Y in Eq.9) and in the case of DBFE from 17.7 to 21.5. The increase is due to the application of the lowpass filter on the image making the first term in Bhattacharyya distance increase (Eq.9). Also, the SPPP algorithm preserves more information, in terms of average Bhattacharyya distance, than DAFE and DBFE. This is because DAFE and DBFE do their computations at the original dimensionality (i.e.195 bands) where the Hughes phenomenon takes place. Moreover, in order to calculate the features, DAFE maximizes the Fisher criterion rather than the Bhattacharyya distance. Figure 9 indicates that the maximum average Bhattacharyya distance (32.4) is achieved when the combination of the SPPP and LPF is used.

Impact of preprocessing stage on classification accuracy: In this part of the experiment, the impact of SPPP and LPF as the two steps of preprocessing block (Fig. 2) on the classification accuracy is evaluated. For this, two experiments were accomplished: one by using DAFE and the other by DBFE. In each experiment, four data sets were used to do the preprocessing and feature extraction processes. The data sets were, then, classified by the ML classifier. The first data set was prepared by direct application of DAFE and DBFE on the 195 original bands to reach 20 features (DAFE in Fig. 10 and DBFE in Fig. 11). For preparing the second data set, first the lowpass filter was applied on the image and then DAFE and DBFE algorithms were used to extract 20 features (LPF +DAFE in Fig. 10 and LPF+DBFE in Fig. 11). The second data set was used to evaluate the effect of the lowpass filter on the classification accuracy. For preparing the third data set, first the data were projected from the original 195 bands to 20 features by SPPP method; then DAFE and DBFE were used to extract different number of features among these 20 features (SPPP +DAFE in Fig. 10 and SPPP+DBFE in Fig. 11). In fact, the third data set was used to evaluate the impact of the first step of the preprocessing block (SPPP) on the classification accuracy.

The fourth data set is similar to the third one. The difference is that in the former, the lowpass filter was applied on the image after the application of SPPP. Then, DAFE and DBFE were used to extract different numbers of features (SPPP+LPF+DAFE in Fig. 10 and SPPP+LPF+DBFE in Fig. 11). In fact the fourth data set was used to evaluate the impact of the first as well as the second step of the preprocessing block (SPPP and LPF).

Figure 10 and 11 show the test field ML classification accuracy versus the number of features (1 to 20) for the four data sets in each experiment. Figure 10 shows that the classification accuracy is saturated after 9 features in DAFE. The reason is that discriminant analysis feature extraction algorithm extracts only (L -1) features where L (9 in this case) is the number of classes [14, 24]. Therefore, the maximum classification accuracy occured in 9 features and is 76.9%. The rest of features in DAFE method were selected randomly. This means, the classification accuracy does not increase after 9 features.

By comparing Fig.10 and 11 it can be seen that the classification accuracy is greater in DAFE than that in DBFE in all subsets of features. Since the DBFE

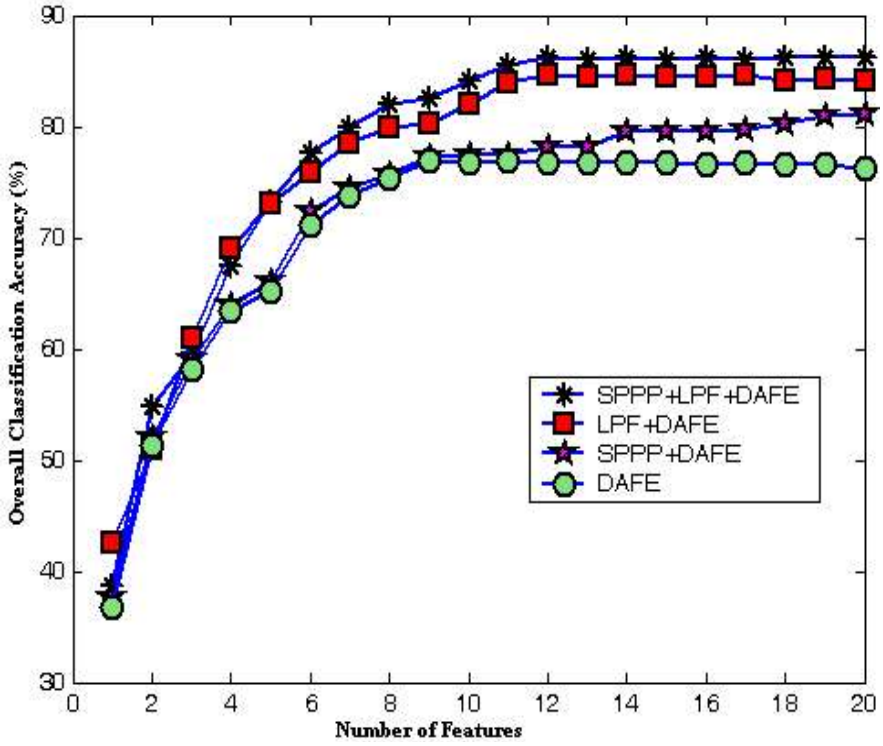


Fig. 10: Test field classification accuracy for different subset of features and for different methods: DAFE, SPPP+DAFE, LPF+DAFE and SPPP+LPF+DAFE

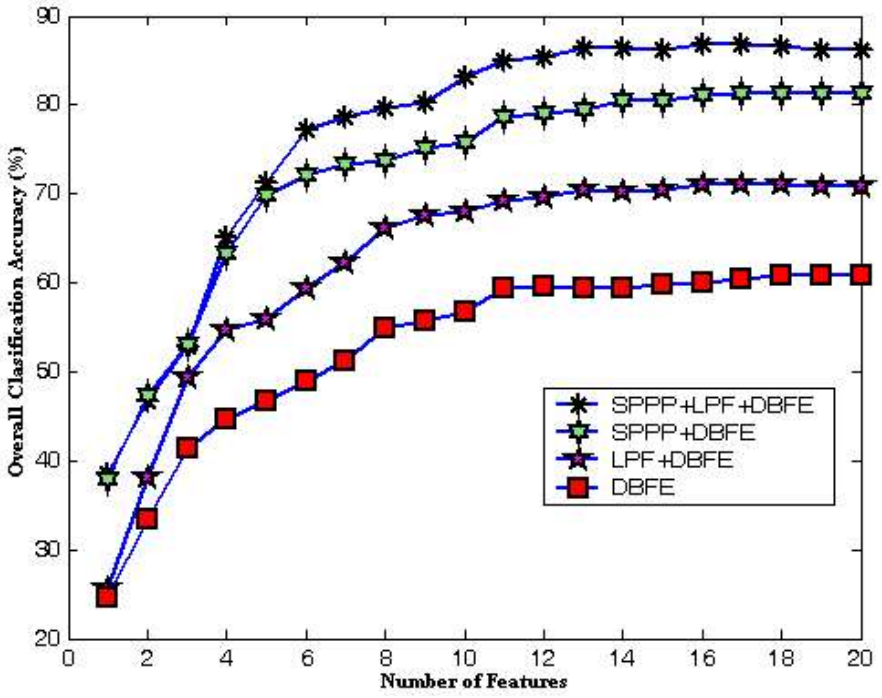


Fig. 11: Test field classification accuracy for different subset of features and for different methods: DBFE, LPF+DBFE, SPPP+DBFE and SPPP+LPF+DBFE

algorithm depends on how well the training samples approximate the decision boundaries, the required number of training samples could be much more for the high dimensional data. Here, only 2260 samples were used (Table 1) and, as expected, the classification accuracy after DBFE is weaker than that of DAFE. It is also observed, that the overall classification accuracy is increased when the SPPP is applied to the data. This increment happens for almost all subsets of features (especially for more than 11 features) in Fig. 10 and all subset of features in Fig. 11. The reason is that unlike the DAFE (or DBFE) that does the computations at full dimensionality (195 bands), the SPPP algorithm does the computations at the lower dimensional space. This allows the approach to deal better with Hughes phenomenon and high dimensional space characteristics, preserving more information. The increment of classification accuracy for 20 features is around 5% for SPPP+DAFE as compared with DAFE and 33% for SPPP+DBFE as compared with DBFE data sets. Also, the classification accuracy improves when the lowpass filter is applied on the image before using the feature extraction algorithms. In the case where the number of features is 20, the overall classification accuracy is 76.3% before and 84.1% after applying LPF for DAFE whereas it was 60.9% before and 70.9% after applying LPF for DBFE. As pointed out earlier, the lowpass filter makes the Bhattacharyya distance increase, leading to an improvement to the classification accuracy.

Although Fig. 9 shows that the performance of DBFE is better than that of DAFE in terms of average Bhattacharyya distance, Fig.10 and 11 indicate that DAFE performs better than DBFE in terms of classification accuracy. The reason is that, the values plotted in Fig. 9 are the averages of 45 numbers of Bhattacharyya distances among the 10 classes and do not show the class separability between each individual class pair. Another reason is that, the DBFE is more sensitive to the number of training samples rather than the separability between the classes.

Also from Fig. 9 it is obvious that LPF contribute significantly over SPPP for average Bhattacharyya distance while as shown in Fig.11 the most contribution comes from SPPP for increasing the classification accuracy. This could be explained by the fact that DBFE needs many more training samples to approximate the decision boundaries in its algorithm for high dimensional data. In addition, as mentioned earlier, it is more sensitive to the number of training samples (2260 training samples i.e. 10% of total samples in this work) than the separability between the classes.

Fig. 10 and 11 show that the maximum classification accuracy happens with 20 features and is 86.3% for both feature extraction algorithm, i.e. SPPP+LPF+DAFE and SPPP+LPF+DBFE. All these experiments suggest that the preprocessing of hyperspectral image data by the combination of SPPP and LPF, as proposed in this paper, gives best results in terms of classification accuracy.

CONCLUSION

This paper reported on the development of a preprocessing stage used before feature extraction stage in the classification of hyperspectral images data. The preprocessing stage includes a dimensionality reduction by SPPP and then applying a LPF on hyperspectral images data to increase the class separability.

The SPPP algorithm reduces the dimensionality by projecting the total number of bands into several subsets of features, doing the computations for each subset separately. This allows the approach to deal better with the Hughes phenomenon and the hyperspectral data characteristics, preserving more information. However, after the application of SPPP, the class separability decreases. This negative effect can be mitigated by application of LPF which increases the Bhattacharyya distance as a measure of class separability. As a consequence, the classification accuracy increases. Although class separability increases by increasing the size of the lowpass filter, small classes might be lost, because the spatial resolution is lost. Therefore, the size of the window must be selected with respect to the smallest class in the image.

The results of this research indicate that, although the application of SPPP and LPF before feature extraction stage increases the classification accuracy individually, the combination of SPPP and LPF, as the two steps of preprocessing stage results a higher accuracy in the classification.

REFERENCES

1. Hsieh, P.F. and D.A. Landgrebe, 1998.Classification of High Dimensional Data. School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana.
2. Landgrebe, D.A., 2003. Signal Theory Methods in Multispectral Remote Sensing. John Wiley and Sons. West Lafayette, Indiana.
3. Hughes, G.F.,1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Transactions on Information Theory, 14 (1): 55-63.

4. Jimenez, L. and D.A. Landgrebe, 1995. High Dimensional Feature Reduction Via Projection Pursuit. School of Electrical and Computer Engineering. Purdue University, West Lafayette, Ph.D. Thesis, Report No. TR-ECE, pp: 96-95.
5. Raudys, S. and V. Pikelis, 1980. On dimensionality, sample size, classification error and complexity of classification algorithm in pattern recognition. IEEE Trans. Pattern Anal. Machine Intell., 2 (3): 242-252.
6. Vaiphasa, C., A.K. Skidmore, W.F. Boer and T. Vaiphasa, 2007. A hyperspectral band selector for plant species discrimination. ISPRS Journal of Photogrammetry and Remote Sensing, 62: 225-235.
7. Shahshahani, B.M. and D.A. Landgrebe, 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Transaction on Geoscience and Remote Sensing, 32 (5): 1087-1095.
8. Friedman, J.H., 1989. Regularized discriminant analysis. Journal of the American Statistical Association, 84: 165-175.
9. Hoffbeck, J.P. and D.A. Landgrebe, 1995. Classification of high dimensional multispectral data. TR-EE 95-14, Purdue University, West Lafayette, Indiana.
10. Bandyopadhyay, S., 2005. Satellite image classification using genetically guided fuzzy clustering with spatial information. International Journal of Remote Sensing, 26 (3): 579-593.
11. Bellman, R., 1961. Adaptive control processes: A Guided Tours. Princeton University Press.
12. Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice and Visualization, John Wiley and Sons. New York.
13. Webb, A., 1999. Statistical Pattern Recognition, Arnold, Great British.
14. Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, San Diego.
15. Lee, C. and D.A. Landgrebe, 1993. Feature extraction based on decision boundaries. IEEE Transaction of Pattern Analysis and Machine Intelligence, 15 (3): 388-400.
16. Jones, M.C. and R. Sibson, 1987. What is Projection Pursuit? J.R. Statistics Soc., A Part 1, pp: 1- 36.
17. Friedman, J.H. and J.W. Tukey, 1974. A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Trans. Computers, 23: 881–889.
18. Lin, H.D. and L.M. Burce, 2004. Parametric Projection Pursuits for Dimensionality Reduction of Hyperspectral Signals in Target Recognition Applications, M.Sc. Thesis, Mississippi State University, USA.
19. Swain, P.H. and S.M. Davis, 1978. Remote Sensing: The Quantitative Approach. McGraw-Hill, New York.
20. Huber, P.J., 1985. Projection Pursuit. The Annals of Statistics, 13 (2): 435-475.
21. Chen, C.H., 1999. Information Processing for Remote Sensing. World Scientific Publishing Co., USA.
22. Jimenez, L. and D.A. Landgrebe, 1999. Hyperspectral Data Analysis and Supervised Feature Reduction via Projection Pursuit. IEEE Transaction on Geoscience and Remote Sensing, Vol: 37 (6).
23. Jain, A.K., 1989. Fundamentals of Digital Image Processing, Prentice-Hall.
24. Gonzalez, R.C. and R.E. Woods, 2003. Digital Image Processing. Prentice-Hall, New Jersey.
25. Kuo, B.C. and D.A. Landgrebe, 2001. Improved Statistics Estimation and Feature Extraction for Hyperspectral Data Classification. School of Electrical Engineering, Purdue University, West Lafayette, Ph.D Thesis.