HYPERSPECTRAL IMAGE (HSI) PROCESSING BASED ON LOCAL DATA DISTRIBUTION ANALYSIS IN LFDA ALGORITHM

Lina Yang
Faculty of Science and Technology
University of Macau, Macau
Av. Padre Tomás Pereira Taipa, Macau, P.R. China
e-mail: yb27411@umac.mo

Huiwu Luo and Yuan Yan Tang
Faculty of Science and Technology
University of Macau, Macau
Av. Padre Tomás Pereira Taipa, Macau, P.R. China
e-mail: yb17409@umac.mo, yytang@umac.mo

ABSTRACT
A pixel in Hyperspectral Image (HSI) can be regarded as a signal, which is represented by a vector with high dimension. In this paper, an intelligent signal processing method is used to treat HSI. Dimension reduction, as a preprocessing step, plays a significant role in the procedure of HSI classification. Based on the typical behavior of Local Fisher Discriminant Analysis (LFDA), an improved LFDA which is called ILFDA is proposed in this paper. According to the affinity matrix in the local linear discriminant analysis, the local variance and prior probability are adopted to describe the local data distribution. This method considers not only the sample data distribution characteristics, but also the discriminant analysis in HSI classification. We carry out experiments on a real HSI database, and the results of the overall accuracy and kappa coefficient indicate that the proposed method is more effective than conventional dimension reduction methods.

KEY WORDS
Intelligent Signal Processing, Applications, Local Linear discriminant analysis, Dimension reduction, Hyperspectral image (HSI).

1 Introduction
Each pixel in Hyperspectral image (HSI) is a signal, which is represented by a high dimension vector. Thus, HSI results in highly dimension redundancy. This property of HSI obstructs the performance of HRI classification. Hence, an effective feature selection is necessary, and dimension reduction is an effective tool to select the useful low-dimension features. A numerous of dimension reduction algorithms have been proposed in the past few decades. These algorithms can be categorized into linear and nonlinear approaches. The most significate linear approaches are Principal Component Analysis (PCA)[10], Random Projection (RP)[2], Linear Discriminant Analysis (LDA)[5], Joint Global and Local Discriminant Analysis (JGLDA)[4], and Locality Preserving Projection (LPP)[6]. While the typical nonlinear ones[7] are Isomap embedding (Isomap), Diffusion Maps (DMaps), Laplacian Eigenmap (LE) and Local Linear Embedding (LLE).

The linear discriminant analysis method is in a better position in the dimension reduction of the hyperspectral image (HSI) [3]. However, the conventional linear discriminant analysis methods, such as LDA, PCA and LPP assume that the Gaussian or Gaussian mixture distribution is used to predict the data distribution, which is impossible for the real HRI data[8]. Due to the fact that the distribution of HSI data is often unknown, and many difficulties are faced in estimating the concrete distribution of HSI data, which is to accommodate the model in reality. A single Gaussian or Gaussian mixture distribution is unable to complete all feature expression under multi-modal data distribution. Hence, the conventional methods lost the priority. To overcome such weakness, a number of extension-LDA are proposed. e.g. M. Sugiyama proposed Local Fisher Linear Discriminant Analysis (LFDA) [13], which incorporates the conventional LDA and PCA. According to different forms of distribution, the proposed LFDA not only overcomes the small sample size problem but also gets rid of the assumption of Gaussian distribution. W. Li adopted it to HRI data [9], and applied the Maximum Likelihood Estimate (MLE), Support Vector Machine (SVM) and Gaussian Mixture Model (GMM) after dimension reduction. From the literature reports, LFDA consumes less time, and obtains prior classification accuracy.

However, conventional LFDA fails to involve the projected data distribution of HRI. In this research, an improved LFDA called ILFDA is proposed. ILFDA applies the prior probability of the class, so that the dimension-reduced HRI has better classification results. ILFDA replaces ‘longest distance’ with ‘local variance’ to describe the data characteristics, which well match the ‘local focus’ property of HRI.

The rest of this paper is organized as follows. A brief review of conventional linear approaches is introduced in section 2. The proposed ILFDA algorithm is presented in section 3. The experimental results are provided in section 4. We conclude our work in section 5.

2 Review of Conventional Literatures
The purpose of linear approaches is to find an optimal projected direction where the information of embedding features is preserved as much as possible. To formulate the problem, let $x_i$ be the $p$-dimensional feature in
the original space, and \( \{x_1, x_2, \ldots, x_N \} \) be the \( N \) samples. For supervised learning, let \( l_i \) be label of \( x_i \), then the label set of all samples can be represented by notation \( \{l_1, l_2, \ldots, l_N \} \). Suppose there are \( C \) classes, and the number of samples in the \( c \)th class is \( N_c \), that fulfills the condition \( N = \sum_{c=1}^CN_c \). That is, the number of all samples are the total sum of each class. Let \( x_i^{(c)} \) be the \( i \)th sample of the \( c \)th class, then the corresponding sample mean becomes \( m_c = (1/N_c)^T \sum_{i=1}^{N_c} x_i^{(c)} \), yet the data center of all samples is denoted as \( m = (1/N)^T \sum_{i=1}^{N} x_i \). Suppose that the data set \( X \) in \( p \)-dimensional hyperspace is distributed to a low \( q \)-dimensional subspace. A general problem of linear discriminant is to find a transformation \( T \in \mathbb{R}^{p \times q} \) that maps the \( p \)-dimensional data into a low \( q \)-dimensional subspace data by \( Y = T^TX \), each \( y_i \) represents \( x_i \) without losing any useful information. The transformed matrix \( T \) is obtained by different methods and different objective functions, which results different algorithms.

2.1 Fisher’s linear discriminant analysis (LDA)

LDA introduces the within-scatter matrix \( S_w \) and between-scatter matrix \( S_b \) to describe the distribution of data samples, where the calculation can be formulated as:

\[
S_w = \sum_{c=1}^{C} \sum_{i=1}^{N_c} (x_i^{(c)} - m_c)(x_i^{(c)} - m_c)^T \tag{1}
\]

\[
S_b = \sum_{c=1}^{C} N_c (m_c - m)(m_c - m)^T \tag{2}
\]

respectively. Fisher’s constraint criterion seeks a transformation \( T \) that maximized the scatter between the classes, while minimized the scatter within the class. This can be achieved by optimizing the following objective function:

\[
T_{LDA} = \text{arg max}_{T \in \mathbb{R}^{p \times q}} \frac{\text{tr}\{T^T S_b T\}}{\text{tr}\{T^T S_w T\}} \tag{3}
\]

It is implicitly assumed that \( T^T S_w T \) is full rank. Thus, the problem can then be attributed to the generalized eigenvectors \( \{\varphi_1, \varphi_2, \ldots, \varphi_d\} \) by solving:

\[
S_b \varphi = \lambda S_w \varphi \tag{4}
\]

Finally, the solution of \( T_{FDA} \) is given by \( T_{FDA} = \{\varphi_1, \varphi_2, \ldots, \varphi_q\} \) which are associated with the first \( q \) largest eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \lambda_q \). Since the max number of ranks of matrix \( S_b \) is \( C - 1 \), which indicates that there are \( C - 1 \) meaningful features in conventional LDA. For the small sample issue, LDA is unacceptable due to the irreversibility of within-scatter matrix \( S_w \). To deal with this issue, a regularization procedure is essential in practice.

2.2 Local Fisher discriminant analysis (LFDA)

Local Fisher discriminant analysis (LFDA) [13] measures the “weights” of two data points by the corresponding distance, the affinity matrix is calculated by this weights. Noted that the “pairwise” representation of within-scatter matrix and between-scatter matrix is momentous for LFDA. Following simple algebra steps, the within-scatter matrix of LDA can be transformed into the following forms:

\[
S_w = \sum_{c=1}^{C} \sum_{i=1}^{N_c} (x_i^{(c)} - m_c)(x_i^{(c)} - m_c)^T = \frac{1}{2} \sum_{i,j=1}^{N} P_w(i,j)(x_i - x_j)(x_i - x_j)^T \tag{5}
\]

where

\[
P_w(i,j) = \begin{cases} 
1/N_c, & \text{if } l_i = l_j \\
1/N, & \text{if } l_i \neq l_j
\end{cases} \tag{6}
\]

Let \( S_t \) be the total mixed scatter matrix of LDA, then:

\[
S_b = S_t - S_w = \frac{1}{2} \sum_{i,j=1}^{N} P_b(i,j)(x_i - x_j)(x_i - x_j)^T \tag{7}
\]

where

\[
P_b(i,j) = \begin{cases} 
1/N - 1/N_c, & \text{if } l_i = l_j \\
1/N, & \text{if } l_i \neq l_j
\end{cases} \tag{8}
\]

LFDA is achieved by weighting the pairwise data points:

\[
\hat{S}_w = \frac{1}{2} \sum_{i,j=1}^{N} \hat{P}_w(i,j)(x_i - x_j)(x_i - x_j)^T \\
\hat{S}_b = \frac{1}{2} \sum_{i,j=1}^{N} \hat{P}_b(i,j)(x_i - x_j)(x_i - x_j)^T
\]

where \( \hat{P}_w(i,j) \) and \( \hat{P}_b(i,j) \) denote the weight matrix of different pairwise points for within class samples and between class samples respectively:

\[
\hat{P}_w(i,j) = \begin{cases} 
W(i,j)/N_c, & \text{if } l_i = l_j, \\
0, & \text{if } l_i \neq l_j
\end{cases} \tag{9}
\]

\[
\hat{P}_b(i,j) = \begin{cases} 
W(i,j)(1/N - 1/N_c), & \text{if } l_i = l_j, \\
1/N, & \text{if } l_i \neq l_j
\end{cases} \tag{10}
\]

where \( W \) indicates the affinity matrix. The construction of \( W \) is critical for the performance of classification accuracy. The details will be presented in the following section.

3 Proposed Method

The calculation of (9) and (10) are very important to the performance of LFDA. There are many methods to compute the affinity matrix \( W \) in (9) and (10). The simplest way is to suppose \( W \) is equivalent to a constant, i.e., \( W(i,j) = a \), where \( a \) is a real nonnegative number.

Another construction adopts the heat kernel derived from a more adaptive LPP version [1].
Furthermore, the density of HSI data points may vary according to different patches. Hence, a local scaling technique is proposed to cope with this issue in ILFDA, where the sophisticated computation is given by:

\[
W(i, j) = \begin{cases} 
\exp(-\frac{||x_i - x_j||^2}{\rho_i \rho_j}), & \text{if } x_i \in \text{KNN}(x_j, K), \\
0, & \text{or } x_j \in \text{KNN}(x_i, K), \\
\end{cases}
\]

(11)

where \( \rho_i \) denotes the local data around the corresponding sample \( x_i \), with the definition \( \rho_i = ||x_i - x_i^{K}|| \), of which \( x_i^{K} \) represents the \( K^{th} \) nearest neighbor of \( x_i \), and \( ||*||^2 \) denotes the square Euclidean distance, \( K \) is a self-tuning parameter predefined. A recommended value of \( K = 7 \) is studied in [15].

Due to the approximation of adjacent HSI pixels, spectrum of neighboring landmarks behave great similarly. The calculation of variance by using the diversity of its \( K^{th} \) nearest neighborhoods is not fully correct for a local point.

Based on the discussion above, a more sophisticated construction of affinity matrix is proposed in the following calculation:

\[
W(i, j) = \begin{cases} 
p(l_i) \exp(-\frac{||x_i - x_j||^2}{\rho_i \rho_j}) \left(1 + \exp(-\frac{||x_i - x_j||^2}{\rho_i \rho_j})\right), & \text{if } l_i = l_j; \\
0, & \text{if } l_i \neq l_j.
\end{cases}
\]

(12)

where \( \rho_* \) denotes local scaling around \( x_* \), with the definition

\[
\hat{\rho}_m = \frac{1}{K} \sum_{i=1}^{K} \frac{||x_i^{(m)}||^2}{\sqrt{\sum_{j=1}^{K} ||x_j^{(m)}||^2}}
\]

(13)

of which \( x_i^{(m)} \) represents the \( i^{th} \) nearest neighbor of \( x_m \), \( ||*||^2 \) is the squared Euclidean distance, and \( K \) is a self-tuning parameter predefined. A recommended value of \( K = 7 \) is proposed in this paper. Compared \( \rho_i = ||x_i - x_i^{K}|| \) with (13), it is noticeable that \( \rho_m \leq \rho_i \) holds.

The optimal solution of improved scheme can be achieved to maximize the following criterion:

\[
T_{ILFDA} \equiv \arg \max_{T \in \mathbb{R}^{p \times q}} \left\{ \frac{\text{tr}\{T^T \hat{S}_b T\}}{\text{tr}\{(T^T \hat{S}_w T\)} \right\} 
\]

(14)

Eq.(14) is similar with Eq.(3), which indicates that the transformation \( T \) can be simply archived by solving the generalized eigenvalue decomposition of \( \hat{S}_w^{-1} \hat{S}_b \). Moreover, Let \( G \in \mathbb{R}^{q \times q} \) be a q-dimensional invertible square matrix. It is clear that \( T_{ILFDA} G \) is also an optimal solution of Eq.(14). This property indicates that under the arbitrary arithmetic transformation of \( T_{ILFDA} G \), the optimal solution is not uniquely determined. Let \( \hat{\phi}_i \) be the eigenvector of \( \hat{S}_w^{-1} \hat{S}_b \) corresponding to eigenvalue \( \lambda_i \), i.e., \( \hat{S}_b \hat{\phi}_i = \lambda_i \hat{S}_w \hat{\phi}_i \). To cope with this issue, a re-scaled procedure is adopted [12]. Each eigenvector \( \{\hat{\phi}_i\}^q_{i=1} \) is re-scaled to satisfy the following constraint:

\[
\hat{\phi}_i \hat{S}_w \hat{\phi}_j = \begin{cases} 
1, & \text{if } i = j; \\
0, & \text{if } i \neq j.
\end{cases}
\]

(15)

Then each eigenvector is weighted by the square root of its associated eigenvalue. The transformed matrix \( T_{ILFDA} \) of improved scheme is finally given by:

\[
T_{ILFDA} = \{\sqrt{\lambda_1 \hat{\phi}_1}, \sqrt{\lambda_2 \hat{\phi}_2}, \cdots, \sqrt{\lambda_q \hat{\phi}_q}\} \in \mathbb{R}^{p \times q}
\]

(16)

with descending order: \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \).

For new testing points \( \hat{x} \), the projected points in the new feature space can be captured by \( \hat{y} = T_{ILFDA}^T \hat{x} \). Thus, it can be further analyzed in the transformed space.

To summarize the discussion aforementioned, the detailed calculation steps of improved scheme are summarized as follows:

1. Construct the affinity matrix \( W \) via Eq.(12), Eq.(9), and Eq.(10), according to the available training samples.

2. To pursue the optimal solution under the framework of graph embedding [14, 11], construct Laplacian matrix \( L \) as follows:

\[
L = D - W
\]

(17)

where \( D \) denotes the row sum (or column sum) of \( W \), i.e., \( D_i = \sum_j W_{ij} \) (or \( D_i = \sum_i W_{ij} \)).

3. Solve the general eigen problem of \( [T^T \hat{S}_b T/(T^T \hat{S}_w T)] \) according to Eq.(14) and Eq.(15). The projected matrix \( T_{ILFDA} \) is achieved via the \( q \) eigenvector that corresponding the \( q \) leading eigenvalue: \( T_{ILFDA} = \{\sqrt{\lambda_1 \hat{\phi}_1}, \sqrt{\lambda_2 \hat{\phi}_2}, \cdots, \sqrt{\lambda_q \hat{\phi}_q}\} \in \mathbb{R}^{p \times q} \).

4. For a new testing sample \( x_t \in \mathbb{R}^p \), the extracted feature is \( z_t = T_{ILFDA}^T x_t \), where the embedding space is \( z_t \in \mathbb{R}^q \).

The advantage of ILFDA is discussed below.

First, the rank between class scatter matrix \( S_b \) can be computed:

\[
\text{rank}(S_b) \leq \text{rank}([N_1(m_1 - m), N_2(m_2 - m), \ldots, N_L(m_L - m)]) \leq C - 1
\]

(18)

It is easy to deduce that the rank of matrix \( S_b \) is \( C - 1 \) at most. Thus, there are \( C - 1 \) sub-features can be extracted at most. Compared with LDA, the method of ILFDA benefit from the affinity matrix \( W \), which has a much higher than ILDA, can be employed into dimension reduction, and the subspace can be any sub-dimensional space. On the other hand, the classical local fisher’s linear discriminant analysis only weights the value of sample pairs in the same classes, while the proposed method takes into account the sample pairs in different classes. The objective function of proposed method is quite similar to LDA, hereby the optimal solution is almost the same as LDA, which indicates that it is also simple to implement and easy to revise.
Hence, the proposed method will be more flexible, and the results will be more adaptive.

The objective function of LDA and ILFDA can be rewritten individually as:

\[ T_{LDA} = \arg \max_{T \in \mathbb{R}^{p \times s}} \text{trace}\{T^TS_iT\} \quad \text{s. t. } T^TS_wT = I \]

(19)

\[ T_{ILFDA} = \arg \max_{T \in \mathbb{R}^{p \times s}} \text{trace}\{T^TS_bT\} \quad \text{s. t. } T^TS_wT = I \]

(20)

This implies that benefiting from the flexible designing of affinity matrix \( W \), ILFDA gains more freedom in Eq.(20). The separability of ILFDA will be more distinct, and the degree of freedom remains more than conventional LDA. Thus, the proposed method is expected to be more robust and preponderant.

For large scale data sets, we discuss a scheme that accelerate the computation procedure of within-scatter matrix \( S_w \). In our algorithm, we have put penalty on the affinity matrix for different class samples in constructing between scatter matrix. Hence, the accelerated procedure will remain for further discussion.

The within class scatter \( \hat{S}_w \) can be reformulated as:

\[
\hat{S}_w = \sum_{i=1}^{N} \left( \sum_{j=1}^{N} \hat{P}_w(i,j) \right) x_i x_i^T - \sum_{i,j=1}^{N} \hat{P}_w(i,j) x_i x_j^T
= X \hat{L}_w X^T
\]

(21)

Here,

\[
\hat{D}_w(i,i) = \sum_{j=1}^{N} \hat{P}_w(i,j), \quad \hat{L}_w = \hat{D}_w - \hat{P}_w
\]

(22)

\( \hat{P}_w \) can be block diagonal if all samples \( \{x_i\}_{i=1}^{N} \) are sorted according to their labels. This property implies that \( \hat{D}_w \) and \( \hat{L}_w \) can also be block diagonal matrix. Hence, if we compute \( \hat{S}_w \) through Eq.(21), then the procedure will be much more efficient. Similarly, \( \hat{S}_b \) can also formulated as:

\[
\hat{S}_b = X \hat{L}_b X^T = X (\hat{D}_b - \hat{P}_b) X^T
\]

(23)

Nevertheless, \( \hat{P}_b \) is dense and can not be further simplified. Moreover, the simplified computational procedure of \( \hat{P}_w \) save part of computation time. In this paper, we adopt above procedure to accelerate \( \hat{S}_w \) while pursues \( \hat{S}_b \) normally. In addition to locality structure, some papers show that marginal information should be preserved in reduced space. The theory of extended LDA and LPP algorithm is developed rapidly recently. S.Yan[14] summarized these algorithms in a graph embedding framework, and also proposed a marginal fisher analysis embedding (MFA) algorithm under this framework.

In MFA, the criterion is characterized by intra-class compactness and inter-class marginal superability, which it is replaced the ‘within-class scatter’ with ‘between class scatter’, severally. The intra-class relationship is reflected by a intrinsic graph, which is constructed by \( K \)-nearest neighborhood sample data points in the same class. The inter-class superability is mirrored by a penalty graph computed for marginal points from different classes. Intra-class compactness is given as follows:

\[
S_i = \sum_{i,j: i \in N(k)(j) \text{ or } j \in N(k)(i)} \|T^T x_i - T^T x_j\|^2 = 2T^T X (D - W) X^T T
\]

(24)

where

\[
W(i,j) = \begin{cases} 1, & \text{if } i \in N(k)(j) \text{ or } j \in N(k)(i) \\ 0, & \text{otherwise.} \end{cases}
\]

(25)

Here, \( N(k)(j) \) represents the \( K \) nearest neighborhood index set of \( x_j \) from the same class, and \( D \) is the row sum (or column sum) of \( W \): \( D(i,i) = \sum_j W_{ij} \). Interclass separability is indicated by a penalty graph whose term is expressed as follows:

\[
S_c = 2T^T X (\hat{D} - \hat{W}) X^T T
\]

\[
\hat{W}(i,j) = \begin{cases} 1, & \text{if } (i,j) \in P(k)(l_j) \text{ or } (i,j) \in P(k)(l_i) \\ 0, & \text{otherwise.} \end{cases}
\]

Note that \( S_i \) and \( S_c \) are corresponding to “within scatter matrix” and “between scatter matrix” of traditional LDA, alternatively. The optimal solution of MFA can be achieved by solving the following minimization problem, i.e.,

\[
\hat{T} = \arg \min_{T} \frac{T^T X (D - W) X^T T}{T^T X (D - W) X^T T}
\]

(26)

Note that Eq.(26) is also a general eigenvalue decomposition problem. Let \( T_{PCA} \) indicates the transformation matrix from the original space to PCA subspace, then the final projection of MFA is described as \( T_{MFA} = T_{PCA} \hat{T} \).

It indicates that MFA constructs two weighted matrix: \( W \) and \( \hat{W} \) according to intra-class compactness and inter-class separability. In LFDA and ILFDA, only one affinity is constructed. The difference lies in that the “weight” in LFDA and ILFDA is in the range of \([0, 1] \) according to the level of difference. MFA distributes the same weight for its \( K \)-nearest neighborhoods. The optimal solution of MFA, LFDA and ILFDA can be attributed to a general eigenvalue decomposition problem. Hence, the idea of MFA, LFDA and ILFDA is approximately similar in a certain interpretation. Relationship with other methodologies can be analyzed in an analogous way.

4 Experimental Results

To illustrate the performance of ILFDA, the experiments on a real HRI data set—AVIRIS Indian Pine 1992, are conducted. It consists of \( 145 \times 145 \) pixels and 224 spectral reflectance bands ranging from 0.4\( \mu m \) to 2.45\( \mu m \) with a spatial resolution of 20m. The 20 water absorption bands were discarded.
The performance of different dimension reduction methods, i.e., PCA, LPP, LFDA, LDA, JGLDA and R-P are compared with ILFDA. Classification accuracy is reported via a concrete classifier. $K$ nearest neighborhood classifier (KNN) and support vector machine (SVM) classifier are selected to measure the performance of extracting features after dimension reduction. The value of $K$ is chosen from 1, 5, and 9 (1NN, 5NN and 9NN). SVM, as a robust and successful classifier, has been widely used to evaluate the performance.

<table>
<thead>
<tr>
<th>ID</th>
<th>Class Name</th>
<th>ID</th>
<th>Class Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alfalfa (46/12, 39.19%)</td>
<td>9</td>
<td>Oats (201/5, 79.90%)</td>
</tr>
<tr>
<td>2</td>
<td>Corn-notill (162/8, 9.52%)</td>
<td>10</td>
<td>Soybean-notill (972/86, 8.41%)</td>
</tr>
<tr>
<td>3</td>
<td>Corn-tilled (389/8, 9.48%)</td>
<td>11</td>
<td>Soybean-tilled (2452/214, 9.37%)</td>
</tr>
<tr>
<td>4</td>
<td>Corn (237/34, 14.56%)</td>
<td>12</td>
<td>Soybean-clean (945/54, 8.11%)</td>
</tr>
<tr>
<td>5</td>
<td>Grass-pasture (49/34, 11.18%)</td>
<td>13</td>
<td>Wheat (205/28, 13.66%)</td>
</tr>
<tr>
<td>6</td>
<td>Grass-trees (170/17, 9.77%)</td>
<td>14</td>
<td>Woods (126/102, 8.06%)</td>
</tr>
</tbody>
</table>
| 7  | Grass-pasture-mowed (28/17, 17.41%) | 15 | Buildings-Grass-Trees-Dirt 
| 8  | Hay-windrowed (478/58, 10.46%) | 16 | Stone-Steel-Towers (924, 25.81%) |
| Total | ~1029/1029, 10.04% | ~1029/1029, 10.04% |

- Numerical value in this table refers to No. samples, No. training samples and p.c., respectively.

In this experiment, totally 1,029 sample were selected for training, and the rest is used for testing. Noted that all the labeled samples in Table 1 distributed uneven, and the available labeled samples of each category differ dramatically. The fixed 15 samples are randomly selected for training samples, the absent samples are randomly selected from the rest samples. Under this strategy, the training samples and testing samples are listed in Table 1.

Figure 1. shows the overall accuracy of different dimension reduction methods on AVIRIS Indian 92AV3C data set. The neighborhood of KNN classifier is selected from 1, 5, to 9, respectively. The three derived SVM classifiers are also posed in this experiment, i.e., linear SVM, polynomial SVM and RBF-SVM.

It can be deduced from Figure 1. (a)~(c) that when the embedding space is greater than 5, the proposed ILFDA performs the best overall accuracy among 7 algorithms.

Moreover, it can be observed from (d) that, with the linear kernel SVM classifier, ILFDA outperforms the rest when the reduced feature space is less than 3. However, LDA demonstrates the highest overall accuracy when the reduced features vary from 4 to 9, while LFDA outperforms the rest when the number of reduced features is greater than 9. This phenomenon of Figure 1. (d) indicates the instability of linear SVM. Nevertheless, the situation reversed when the subspace is greater than 7 for polynomial SVM and RBF-SVM in Figure 1. (e) and Figure 1. (f), wherein the proposed ILFDA wins significant improvement. In summary, the proposed ILFDA algorithm achieved better performance in most cases.

Figure 2. displays the classified maps of proposed ILFDA in pseudo-color images. It is obviously that the best performance achieved by ILFDA is under 7NN classifier, with 83.79% in overall accuracy, and 89.91% in average accuracy. Moreover, the worst algorithm is JGLDA whose overall accuracy is 61.95% and the average accuracy is only 62.09%.

5 Conclusion

In this paper, the procedure of Local Fisher Discriminant Analysis (LFDA) is discussed in details. The ‘maximum distance’ is replaced with ‘local variance’ in the construction of weight matrix. Simultaneously, class prior probability is introduced into the computation of affinity, and the improved LFDA algorithm (ILFDA) is proposed. ILFDA has been evaluated on a real removing sensing AVIRIS Indian Pine 92AV3C data set. We compared the performance of PCA, LPP, LFDA, LDA, JGLDA, RP and ILFDA by numerical results. The KNN classifier and SVM classifier are imposed for generating class labels. We also found that the accuracy of HSI data relies on an intrinsic dimension [5]. Hereby, future work can be carried out on exploring the distribution of landmarks, spatial relationship, as well as intrinsic dimension estimation for classification performance improvement.

Acknowledgements

This work was supported by the University of Macau Grants: MYRG205 (Y1-L4)-FST11-TYY and No. MYRG187 (Y1-L3)-FST11-TYY and SRG010-FST11-TYY, and the National Natural Science Foundation of China :61273244, as well as the Science and Technology Development Fund (FDCT) of Macau :100-2012-A3.

References


Figure 1. Overall accuracy versus reduced subspace for AVIRIS Indian Pines.

Figure 2. Classified map produced by the proposed method.


