

# LOCALITY-PRESERVING DISCRIMINANT ANALYSIS FOR HYPERSPECTRAL IMAGE CLASSIFICATION USING LOCAL SPATIAL INFORMATION

Wei Li<sup>1</sup>, Saurabh Prasad<sup>2</sup>, Zhen Ye<sup>3</sup>, James E. Fowler<sup>1</sup>, Minshan Cui<sup>2</sup>

<sup>1</sup> Mississippi State University, Mississippi State, MS 39762, USA

<sup>2</sup> University of Houston, Houston, TX 77004, USA

<sup>3</sup> Northwestern Polytechnical University, 127 Youyi Xilu, Xi'an 710072, China

## ABSTRACT

Locality-preserving projection as well as local Fisher discriminant analysis is applied for dimensionality reduction of hyperspectral imagery based on both spatial and spectral information. These techniques preserve the local geometric structure of hyperspectral data into a low-dimensional subspace wherein a Gaussian-mixture-model classifier is then considered. In the proposed classification system, local spatial information—which is expected to be more multimodal than strictly spectral features—is used. Results with experimental hyperspectral data demonstrate that this system outperforms traditional classification approaches.

**Index Terms**—Dimensionality reduction, linear discriminant analysis, hyperspectral data, pattern classification.

## 1. INTRODUCTION

Hyperspectral imagery (HSI) obtained by remote-sensing systems is a 3D cube covering a wide range of the electromagnetic spectrum [1]. Typically, hyperspectral data provides hundreds of narrow contiguous bands which include significant information about the spectral features of the materials in the scene. Over the last two decades, various hyperspectral image classification algorithms have been developed [2] for a variety of applications requiring robust ground-cover classification of remotely-sensed HSI.

Dimensionality reduction is an important preprocessing step for hyperspectral image classification. Since hyperspectral images possess a wealth of spectral information for each pixel in the image, an effective dimensionality reduction can exploit useful features specific for the classification task at hand and result in improved classification performance.

Principal component analysis (PCA) [3] and linear discriminant analysis (LDA) [3] are popular dimensionality-reduction techniques for hyperspectral image classification. Locality preserving projection (LPP) [4–6] is a recently proposed dimensionality-reduction algorithm which is designed to preserve the distance between samples while projecting data into a lower-dimensional subspace. In other words, LPP

can preserve the local structure of neighboring samples in the input space. This is distinctly different from PCA and LDA. Nevertheless, LPP is an unsupervised method. In [7], we have studied the supervised local Fisher discriminant analysis (LFDA), which can be viewed as LPP plus supervised LDA. This property of LFDA guarantees good between-class separation in the projection while preserving the within-class local structure.

In our previous work [7], LFDA is presented as a dimensionality reduction using only spectral features. Many recent studies have demonstrated that spatial information is helpful for hyperspectral image classification since spatial texture is useful for hyperspectral images, especially at high resolution. In this work, we investigate LFDA as well as LPP coupled with a Gaussian-mixture-model (GMM) [8] classifier to improve the classification performance based on spatial-spectral information. The combination of LPP and the GMM classifier can be effective for hyperspectral image classification.

The remainder of the paper is organized as follows. Sec. 2 briefly reviews some dimensionality-reduction techniques. Sec. 3 describes the classifiers we have used as well as our proposed method. Sec. 4 provides a description of the experimental hyperspectral dataset and presents experimental results. Finally, Sec. 5 makes some concluding remarks.

## 2. DIMENSIONALITY REDUCTION

Dimensionality reduction is commonly used as a preprocessing step to reduce the dimensionality of the data while ensuring a well-conditioned representation of class statistics. Popular dimensionality-reduction methods include PCA and LDA [3]. PCA [9] seeks to find a linear transformation which projects high-dimensional data into a low-dimensional space in terms of maximizing the variance of the data in projected space. LDA [3] seeks to find a linear transformation which maximizes the between-class scatter matrix while minimizing the within-class scatter matrix. In general, for classification tasks, PCA-based dimensionality reduction is suboptimal compared to LDA-based methods [2] due to the fact that

PCA-based methods aim to find projections with minimal reconstruction error, whereas LDA-based methods seek to find projections that preserve the discriminant information. However, a limitation of LDA is that the technique assumes that the class-conditional distributions are single Gaussian [10]. In this paper, we introduce dimensionality reduction designed to exploit the rich statistical structure of hyperspectral data. These techniques do not make the single-Gaussian assumption but rather assume that the data is highly non-Gaussian and multimodal, which is suitable for hyperspectral data.

## 2.1. LPP

Unlike PCA and LDA, LPP seeks to find a linear map  $\mathbf{W}$  that preserves the local structure of neighboring samples from the input space. Fig. 1 illustrates the difference between PCA and LPP (both are unsupervised dimensionality-reduction methods) using synthetic data.

LPP preserves neighborhood relationships and forces neighboring points in the input space to remain close in the projected space. Consider a dataset with training samples  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$  in  $\mathbb{R}^d$  ( $d$ -dimensional feature space) and class labels  $y_i \in \{1, 2, \dots, C\}$ , where  $C$  is the number of classes, and  $n$  is the total number of training samples. Let  $n_l$  be the number of available training samples for the  $l^{\text{th}}$  class,  $\sum_{l=1}^C n_l = n$ . Define  $A_{i,j} \in [0, 1]$  as the *affinity* between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ,

$$A_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\gamma_i \gamma_j}\right), \quad (1)$$

where  $\gamma_i = \|\mathbf{x}_i - \mathbf{x}_i^{(k_{\text{nn}})}\|$  denotes the local-scaling of data samples in the neighborhood of  $\mathbf{x}_i$ , and  $\mathbf{x}_i^{(k_{\text{nn}})}$  is the  $k_{\text{nn}}$ -nearest neighbor of  $\mathbf{x}_i$ .  $\mathbf{A}$  is then a symmetric matrix (referred to as the affinity matrix) of size  $n \times n$  which measures the distance among data samples.

The eigenmap is calculated via computing the eigenvectors and eigenvalues of the generalized eigenvalue problem,

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{W} = \mathbf{\Lambda}\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{W}, \quad (2)$$

where  $\mathbf{\Lambda}$  is the diagonal eigenvalue matrix;  $\mathbf{W}$  is the eigenvector matrix; and  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  is the Laplacian matrix such that  $\mathbf{D}$  is a diagonal matrix with the  $i^{\text{th}}$  diagonal element being  $D_{ii} = \sum_{j=1}^n A_{i,j}$ . Further details can be found in [4].

## 2.2. LFDA

LFDA is a supervised dimensionality-reduction technique which is designed to handle multimodal, non-Gaussian distributions. In essence, LFDA combines the properties of LDA and LPP. In LFDA, the *local* between-class  $S^{(\text{lb})}$  and

within-class  $S^{(\text{lw})}$  scatter matrices are defined as,

$$S^{(\text{lb})} = \frac{1}{2} \sum_{i,j=1}^n W_{i,j}^{(\text{lb})} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top, \quad (3)$$

$$S^{(\text{lw})} = \frac{1}{2} \sum_{i,j=1}^n W_{i,j}^{(\text{lw})} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top, \quad (4)$$

where  $W^{(\text{lb})}$  and  $W^{(\text{lw})}$  are  $n \times n$  matrices defined as,

$$W_{i,j}^{(\text{lb})} = \begin{cases} A_{i,j}(1/n - 1/n_l), & \text{if } y_i = y_j = l, \\ 1/n, & \text{if } y_i \neq y_j, \end{cases} \quad (5)$$

$$W_{i,j}^{(\text{lw})} = \begin{cases} A_{i,j}/n_l, & \text{if } y_i = y_j = l, \\ 0, & \text{if } y_i \neq y_j. \end{cases} \quad (6)$$

Maximizing the Fisher's ratio as defined using the local scatter matrices, we have that

$$\begin{aligned} \Phi_{\text{LFDA}} = \\ \arg \max_{\Phi_{\text{LFDA}}} \text{tr} \left[ \left( \Phi_{\text{LFDA}}^\top S^{(\text{lw})} \Phi_{\text{LFDA}} \right)^{-1} \Phi_{\text{LFDA}}^\top S^{(\text{lb})} \Phi_{\text{LFDA}} \right] \end{aligned} \quad (7)$$

is given by  $S^{(\text{lb})} \Phi_{\text{LFDA}} = \mathbf{\Lambda} S^{(\text{lw})} \Phi_{\text{LFDA}}$ , where  $\mathbf{\Lambda}$  is the diagonal eigenvalue matrix, and  $\Phi_{\text{LFDA}} \in \mathbb{R}^d$  is the transformation matrix. LFDA obtains good between-class separation in the projection while preserving the within-class local structure at the same time. A more complete description of LFDA can be found in [7, 11].

## 3. CLASSIFICATION

The maximum-likelihood-estimation (MLE) [12] classifier is one of the most popular methods of classification of remote-sensing data. However, a fundamental limitation of MLE classifiers is the assumption that class-conditional likelihoods are Gaussian—for a variety of classification tasks, this assumption can be inaccurate, resulting in suboptimal classification performance.

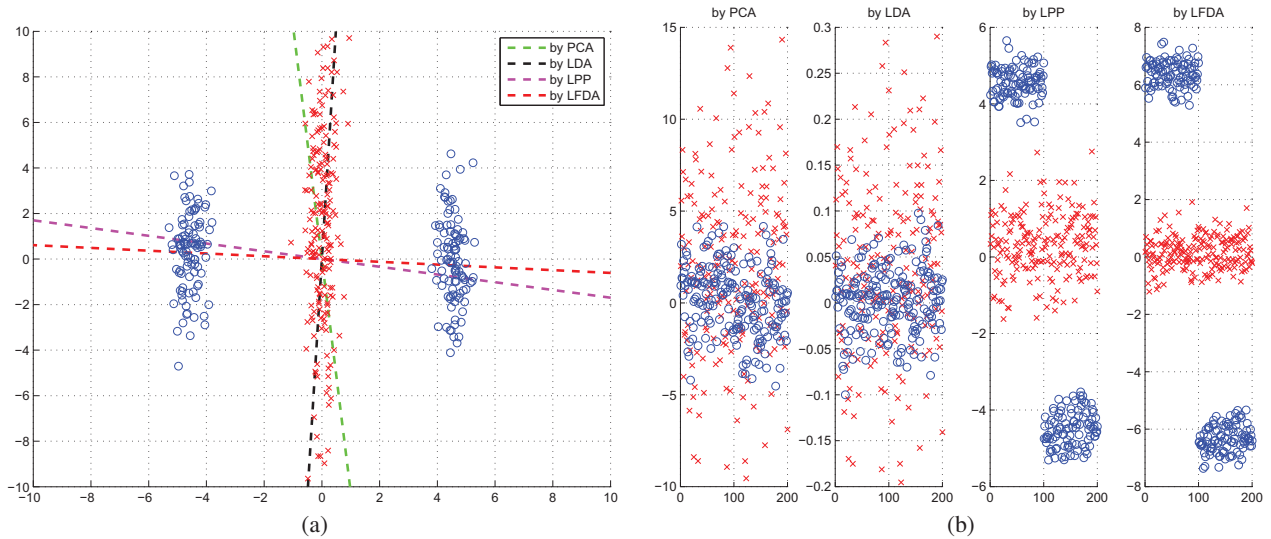
### 3.1. GMM

A GMM [8] can be viewed as a combination of two or more normal Gaussian distributions. In a typical GMM representation, a probability density function for  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$  in  $\mathbb{R}^d$  is written as the sum of  $K$  Gaussian components (modes); i.e.,

$$p(\mathbf{x}) = \sum_{k=1}^K \alpha_k \mathcal{N}(\mathbf{x}, \mu_k, \Sigma_k), \quad (8)$$

where

$$\begin{aligned} \mathcal{N}(\mathbf{x}, \mu_k, \Sigma_k) = \\ \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right]. \end{aligned} \quad (9)$$



**Fig. 1.** (a) A synthetic 2D multimodal data and the directions of PCA, LDA, LPP and LFDA; (b) the data distribution after projection into a 1D subspace.

In (8),  $K$  is the number of mixture components, while  $\alpha_k$ ,  $\mu_k$ , and  $\Sigma_k$  are the mixing weight, mean, and covariance matrix, respectively, of the  $k^{\text{th}}$  component. These last three quantities can be expressed by the parameter vector  $\Theta = \{\alpha_k, \mu_k, \Sigma_k\}$ . Once the optimal number of components,  $K$ , per GMM has been determined, the parameters for the mixture model can be estimated by the expectation-maximization (EM) algorithm [13], an iterative optimization strategy.

The spectral response of hyperspectral data can be affected by many factors, such as differences in illumination conditions, geometric features of material surfaces, and atmospheric effects. It is hence reasonable to expect that the statistical distribution of classes in hyperspectral imagery has a complicated multimodal structure. Classifiers such as those based on GMM are hence a natural fit for hyperspectral data. The visualization in Fig. 1 suggests that the combination of LPP and a GMM classifier can be effective for hyperspectral image classification.

In this work, we exploit LPP as well as LFDA for dimensionality reduction based on the spatial-spectral information for hyperspectral image classification. PCA first reduces the dimensionality of original image bands from  $d$  to a certain value  $d'$  ( $d' \ll d$ ). For each pixel, we extract a spatial window of size  $B \times B$  surrounding the current pixel; the resulting  $B \times B \times d'$  cube from this window is then “rasterized” into a vector of dimension  $B^2 \cdot d'$  which is considered to be the new feature vector (including spectral information as well as spatial information) for the center sample (pixel) in the current spatial window. Following this procedure, LFDA or LPP is employed on this new space to extract features. In doing so, the local spatial information of hyperspectral image is utilized, and it is expected that the spatial texture tends to be

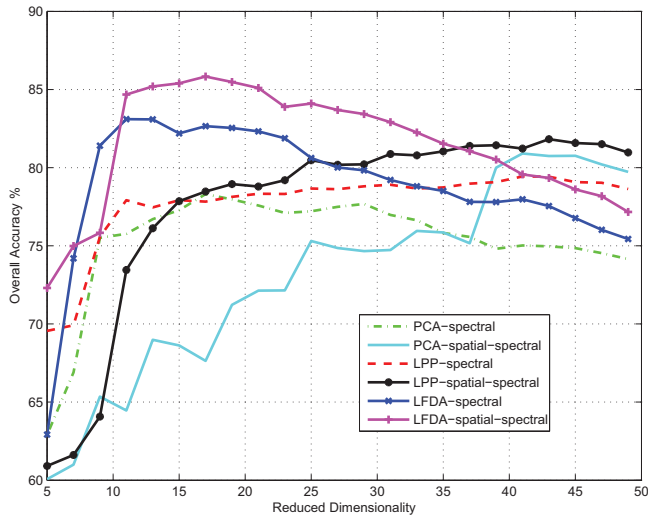
more multimodal than features based solely on spectral information. Thus, LPP/LFDA coupled with the GMM classifier is expected to improve classification performance.

#### 4. EXPERIMENTS AND ANALYSIS

In this section, we evaluate the classification performance of the proposed algorithm for hyperspectral data classification. The HSI dataset employed was acquired using NASA’s Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor and was collected over northwest Indiana’s Indian Pine test site in June 1992<sup>1</sup>. The image represents a vegetation-classification scenario with  $145 \times 145$  pixels and 220 spectral bands with a spatial resolution of 20 m. Eight classes are used in this study. There are about 188 training samples per class and a total of 7,120 testing samples.

Fig. 2 illustrates the overall accuracy versus reduced dimensionality for the Indian Pines dataset (8 classes). It is worth mentioning that the reduced dimensionality of LFDA is not restricted to  $C - 1$  where  $C$  is the number of classes. For traditional LDA, the classification accuracy using only spectral information for this dataset is 76.42% while the accuracy obtained using spatial-spectral information is 79.75%, which are both lower than the techniques shown in Fig. 2—PCA-spectral, PCA-spatial-spectral, LPP-spectral, LPP-spatial-spectral, LFDA-spectral and LFDA-spatial-spectral. We also show the overall classification accuracy versus varying reduced dimensionality for the Indian Pines dataset. From the experimental results, it can be seen that the overall accuracy of LPP is always higher than that of PCA, both for spatial as well as spectral classification. Furthermore, when

<sup>1</sup><ftp://ftp.ecn.purdue.edu/biehl/MultiSpec>



**Fig. 2.** Indian Pines dataset: overall classification accuracy versus reduced dimensionality.

the dimensionality is 17, the proposed LFDA-spatial-spectral classifier provides the best performance, approximately 3% higher accuracy than the LFDA-spectral technique introduced in [7, 14]. For this dataset, in Fig. 2, the reduced dimensionality indicates the final dimensionality of the space in which the classifiers are applied; spatial window size  $B$  is 5; and parameter  $d'$  is 24, 10, and 22 for LDA, LPP, and LFDA, respectively. Finally, we note that we report results from the proposed spatial-spectral classification following an optimization of all system parameters.

## 5. CONCLUSIONS

In this paper, we investigated LPP and LFDA as dimensionality reduction for hyperspectral image classification. Unlike traditional PCA and LDA, LPP and LFDA are designed to handle non-Gaussian class distributions and preserve the underlying structure of such distributions in the projected space. We exploited these techniques for hyperspectral data based on spatial and spectral information, and GMM was employed as the classifier to predict class labels. The hyperspectral dataset—representing vegetation classification—was used to evaluate the proposed strategy and compare it to traditional classification methods. Experimental results demonstrated that the proposed system provides better classification performance compared to traditional approaches.

## 6. REFERENCES

[1] D. A. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 17–28, January 2002.

[2] S. Prasad and L. M. Bruce, "Limitations of principal component analysis for hyperspectral target recognition," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 625–629, October 2008.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[4] X. He and P. Niyogi, "Locality preserving projections," in *Advances in Neural Information Processing System*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2004.

[5] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 894–898, September 2011.

[6] Z. Wang and B. He, "Locality preserving projections algorithms for hyperspectral image dimensionality reduction," in *Proceedings of the International Conference on Geoinformatics*, Shanghai, China, June 2011, pp. 1–4.

[7] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 4, pp. 1185–1198, April 2012.

[8] A. Berge and A. H. S. Solberg, "Structured Gaussian components for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3386–3396, November 2006.

[9] K. Liu, Q. Du, H. Yang, and B. Ma, "Optical flow and principal component analysis-based motion detection in outdoor videos," *EURASIP Journal in Advances in Signal Processing*, vol. 2010, July 2010, Article ID 680623.

[10] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, February 2001.

[11] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proceedings of the International Conference on Machine Learning*, Pittsburgh, PA, June 2006, pp. 905–912.

[12] S. Di Zeno, R. Bernstein, S. D. Degloria, and H. C. Kolsky, "Gaussian maximum likelihood and contextual classification algorithms for multicrop classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 25, no. 6, pp. 805–814, November 1987.

[13] N. Vlassis and A. Likas, "A greedy EM algorithm for Gaussian mixture learning," *Neural Processing Letters*, vol. 15, no. 1, pp. 77–87, February 2002.

[14] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "A multi-modal pattern classification framework for hyperspectral image analysis," in *Proceedings of the Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Lisbon, Portugal, June 2011.