HYPER SPECTRAL IMAGE CLASSIFICATION BASED ON SPECTRA DERIVATIVE FEATURES AND LOCALITY PRESERVING ANALYSIS

Zhen Ye 1 
yzh525@gmail.com
1 Shaanxi Key Lab. of Information Acquisition and Processing, School of Electronics and Information Northwestern Polytechnical University, Xi’an, 710129, China

Mingyi He 1
myhe@nwpu.edu.cn
1 Northwestern Polytechnical University, Xi’an, 710129, China

James E. Fowler 2
fowler@ece.msstate.edu
2 Geosystems Research Institute, Mississippi State University, Starkville, MS, USA

Qian Du 3
du@ece.msstate.edu
3 Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA

ABSTRACT

High spectral resolution and correlation hinders the application of traditional hyperspectral classification methods in the spectral domain. To address this problem, derivative information is studied in an effort to capture salient features of different land-cover classes. Two locality-preserving dimensionality-reduction methods—specifically, locality-preserving nonnegative matrix factorization and local Fisher discriminant analysis—are incorporated to preserve the local structure of neighboring samples. Since the statistical distribution of classes in hyperspectral imagery is often a complicated multimodal structure, classifiers based on a Gaussian mixture model are employed after feature extraction and dimension reduction. Finally, the classification results in the spectral as well as derivative domains are fused by a logarithmic-opinion-pool rule. Experimental results demonstrate that the proposed algorithms improve classification accuracy even in a small training-sample-size situation.

Index Terms—Spectral derivative, locality-preserving analysis, hyperspectral image classification

1. INTRODUCTION

Hyperspectral imagery (HSI) is acquired and collected in hundreds of contiguous spectral bands. Due to such highly correlated and enormous data, the generalization capability of statistical classifiers may be reduced [1]. Dimensionality-reduction algorithms are typically employed to address this problem. Traditional approaches include unsupervised methods such as principal component analysis (PCA), as well as supervised methods, such as Fisher’s linear discriminant analysis (LDA) [2]. PCA is designed to minimize mean square error between the original and reduced spaces while LDA is designed to find a projection maximizing the class separation in a lower-dimension space [3].

PCA and LDA assume that the class-conditional distributions are Gaussian. However, real-life observational data are often not Gaussian and, in extreme cases, may be multimodal. In this work, two locality-preserving dimensionality-reduction methods based on locality-preserving non-negative matrix factorization (LPNMF) [4] and local Fisher’s discriminant analysis (LFDA) [5] are studied to exploit the rich statistical structure of HSI data. In [6], Li et al. argued that LFDA-based dimensionality reduction followed by Gaussian-mixture-model (GMM) [7] classifiers effectively captures the underlying statistical structure accurately and obtains superior classification performance. Although GMM combined with LFDA (LFDA-GMM) has proven to be an effective approach for classification, it still suffers from the fact that the classifier exploits only spectral information. In real applications, spectral reflectance is not exactly identical to laboratory measurements due to, e.g., illumination and atmospheric effects [8].

This paper proposes two classification methods called D-LPNMF-GMM-Fusion and D-LFDA-GMM-Fusion which are based upon spectral-derivative features and locality-preserving analysis. In this work, derivative information of each pixel is extracted under different orders. LPNMF and LFDA are used to reduce the dimension in the spectral and derivative domains, respectively. Following this, GMM classifiers are applied for obtaining local labels. Finally, the classification results from the spectral and derivative domains are fused by logarithmic-opinion-pool (LOGP) [9] decision-fusion rule. The experimental results demonstrate that the proposed system outperforms traditional algorithms, including LDA coupled with maximum likelihood estimation (LDA-MLE), LPNMF combined with GMM (LPNMF-GMM), LFDA-GMM [6], and a support vector machine (SVM) [10].

The remainder of this paper is organized as follows. Sec. II describes the two locality-preserving dimensionality-reduction techniques—LPNMF and LFDA. In Sec. III, we describe the formulation of the proposed classification system. A detailed discussion of the experimental results is presented in Sec. IV. Finally, Sec. V concludes the paper.

2. LOCALITY-PRESERVING ANALYSIS

2.1. LPNMF

LPNMF combines the advantages of non-negative matrix factorization (NMF) and locality-preserving projection (LPP) [11], which leads to a parts-based representation using only
additive operations wherein the intrinsic geometric structure is also preserved in the embedded space [12]. The objective function of LPNMF is

$$O = \sum_{i=1}^{d} \sum_{j=1}^{g} x_{ij} \log \frac{x_{ij}}{y_{ij}} + \lambda \Pi$$  \hspace{1cm} (1)

where $Y = \{y_{ij}\} = UV^T$, and $\lambda$ is the regularization parameter. The first part of (1) is the usual objective function of NMF which uses Kullback-Leibler divergence between $X$ and $Y$. On the other hand, $\mathcal{R}$ is used to force a geometric locality constraint among points in the reduced-dimensional subspace $V$ (t < d); i.e.,

$$\mathcal{R} = \frac{1}{2} \sum_{j=1}^{d} \sum_{k=1}^{g} (v_{ij} \log \frac{v_{ij}}{v_{jk}} - v_{ij} + v_{ik} \log \frac{v_{ik}}{v_{jk}}) W_{jk}$$  \hspace{1cm} (2)

where $W$ is the edge-weight matrix measuring the distance between points in the original space $X$. This matrix, following the theory of LPP, is used to preserve the intrinsic geometry of the data distribution. The following multiplicative rules are used to minimize the function $O$ and estimate the matrices $U$ and $V$,

$$u_{iq} \leftarrow \frac{\sum_j (x_{ij} v_{ij}) / \sum_j (u_{ij} v_{ij})}{\sum_j v_{ij}}$$  \hspace{1cm} (3)

$$v_{ij} \leftarrow \left( \sum_k u_{ik} I + \lambda L \right)^{-1} \left[ v_{ij} - \sum_k u_{ik} u_{ij} \right]$$  \hspace{1cm} (4)

where $v_{ij}$ is the $i^{th}$ column of $V$, $I$ is an $n \times n$ identity matrix, and $L$ is the graph Laplacian of $W$.

2.2. LFDA

LFDA has been recently proposed as supervised dimensionality reduction in order to handle multimodal, non-Gaussian distributions. LFDA combines the properties of LDA and LPP, exploiting the advantages of each. As in LPP, LFDA preserves the neighborhood relationships within the embedding by employing an “affinity” matrix that is described below.

Consider a dataset with training samples $\{x_i\}_{i=1}^{N}$ and class labels $y_i \in \{1, 2, \cdots, c\}$, where $c$ is the number of classes, and $n$ is the total number of training samples. Let $n_k$ be the number of available training samples for class $k$, and $\sum_{k} n_k = n$. Define $A_{ij} \in [0, 1]$ as the “affinity” between $x_i$ and $x_j$,

$$A_{ij} = \exp(-\|x_i - x_j\|^2 / (\gamma_i \gamma_j))$$  \hspace{1cm} (5)

where $\gamma_i = \|x_i - x_{i}^{(m)}\|$ denotes the local scaling of data $x_i$, and $x_{i}^{(m)}$ is the $m^{th}$ nearest neighbor of $x_i$. $A_{ij}$ is a symmetric matrix (referred to as the affinity matrix) of size $n \times n$ which measures the distance among data samples. Note that there are clearly many different ways to define an affinity matrix; in this work, the heat kernel as defined in (5) is employed. In LFDA, the “local” between-class $S^{(b)}$ and within-class $S^{(w)}$ scatter matrices are scaled via the affinity matrix. The solution that maximizes the Fisher’s ratio within the context of the local scatter matrices is given by

$$S^{(b)} \tilde{\phi} = \lambda S^{(w)} \phi$$  \hspace{1cm} (6)

where $\lambda$ is the diagonal eigenvalue matrix, and $\phi$ is the eigenvector matrix. The modified Fisher’s ratio in LFDA employs these local scatter matrices to estimate the dimensionality-reduction projection and obtains good between-class separation while preserving the within-class local structure. It is hence expected that LFDA will surpass LDA and LPP when the dataset has a multi-modal distribution.

3. THE PROPOSED CLASSIFICATION SYSTEM

3.1. Derivative Information

For HSI analysis, the spectral derivative is often employed to eliminate systematic error in spectral data and to reduce effects from atmospheric radiation and absorption. As for classification, derivative information can be used to capture salient features of different land-cover classes. The first-order spectral derivative are defined as follows,

$$\frac{\partial y_{ij,k}}{\partial \lambda_k} = \frac{y_{ij,k+1} - y_{ij,k-1}}{2 \Delta \lambda}$$  \hspace{1cm} (7)

where $y_{ij,k}$ and $y_{ij,k+1}$ are the spectral reflectance values corresponding to the wavelengths $\lambda_k$ and $\lambda_{k+1}$, respectively, and $N$ represents the number of spectral bands, $k = 1, 2, \cdots, N$. If $\Delta \lambda = \lambda_k - \lambda_{k+1} > 0$, the wavelength windows are fixed. In this paper, we set $\Delta \lambda = 1$.

Using the same strategy as the first-order derivative, the $n^{th}$-order spectral derivative is defined by

$$\frac{\partial^n y_{ij,k}}{\partial \lambda^n} = \partial^n \left[ \frac{\partial^{n-1} y_{ij,k}}{\partial \lambda^{n-1}} \right]$$  \hspace{1cm} (8)

3.2. Decision Fusion

LOGP is a popular soft-decision fusion rule. LOGP entails the use of posterior probabilities, or, more generally, some class-membership function from every classifier, for making the final decision. In LOGP, the global class-membership function is modified to be a weighted product of the posterior probabilities of all classifiers, instead of a weighted sum:
\[
C(w_j | x) = \prod_{i=1}^{n} p_i(w_j | x)^{a_i} \tag{9}
\]

\[ \Rightarrow \log C(w_j | x) = \sum_{i=1}^{n} a_i \log p_i(w_j | x) \]

where \(a_i\) is the confidence score for the \(i\)-th classifier, \(w\) is the class label from one of the \(C\) possible classes for the test pixel, \(i\) is the classifier index, \(n\) is the number of classifiers, and class \(j\) is detected in the bank of classifiers.

3.3. Proposed System

In this paper, we study two classification strategies to fully exploit the spectral information of HSI via the spectral derivative. The first strategy employs LPNMF to preserve local information. The second one involves LFDA-based dimensionality reduction. Fig. 1 illustrates the flowchart of the proposed algorithms. Firstly, derivative information of each pixel is extracted in the spectrum direction. Then, LPNMF or LFDA is used to reduce the dimension in both the spectral and derivative domains. GMM classifiers are subsequently applied for obtaining local labels. Finally, the classification results from the spectral domain and all the derivative domains are fused by the LOGP decision-fusion rule.

![Flowchart of the proposed system.](image)

**4. EXPERIMENTS AND ANALYSIS**

The dataset used is the Pavia University data, collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. The dataset has 103 spectral bands with a spatial size of 610×340 pixels. Approximately 8100 labeled pixels are employed to train and validate/quantify the efficacy of the proposed system. This dataset is partitioned into approximately 1350 training samples and 6750 testing samples (the ratio of testing and training is 5:1).

We have conducted experiments to study the classification performance of the proposed system under different numbers of derivatives. For example, when a single derivative is used, the classification result is obtained by merging decisions from the spectral domain and the first-order derivative domain. The parameter \(r\) which represents the dimensionality in LPNMF or LFDA, is selected from the range 5 to 30. To avoid any bias, the experiments are repeated 20 times, reporting the average classification accuracy and the standard deviation in Table I. It is clear that the classification performance improves with increasing number of derivatives. However, the growth of classification accuracy diminishes for more than 8 derivatives.

<table>
<thead>
<tr>
<th>Derivative Number</th>
<th>D-LPNMF-GMM-Fusion Accuracy</th>
<th>Deviation</th>
<th>D-LFDA-GMM-Fusion Accuracy</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93.04% (r=15) 0.0036</td>
<td>93.23% (r=9) 0.0039</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>93.26% (r=15) 0.0044</td>
<td>93.56% (r=9) 0.0037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>93.78% (r=15) 0.0040</td>
<td>94.19% (r=9) 0.0044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>94.08% (r=15) 0.0040</td>
<td>94.60% (r=9) 0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>94.38% (r=15) 0.0047</td>
<td>94.95% (r=9) 0.0036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>94.65% (r=15) 0.0037</td>
<td>95.24% (r=9) 0.0027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>94.79% (r=11) 0.0033</td>
<td>95.38% (r=9) 0.0031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>95.21% (r=13) 0.0034</td>
<td>95.74% (r=7) 0.0032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>95.52% (r=13) 0.0031</td>
<td>95.71% (r=7) 0.0036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>95.32% (r=13) 0.0030</td>
<td>95.83% (r=7) 0.0028</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following experiments are conducted using both the original spectral-domain and the first-order-derivative features. The proposed D-LPNMF-GMM-Fusion and D-LFDA-GMM-Fusion are compared to classical algorithms, including LDA-MLE, LPNMF-GMM, LFDA-GMM [6], and SVM [10]. A radial-basis-function (RBF) kernel is used in the SVM implementation. All the free parameters used in the above algorithms, such as \(r\) for dimensionality reduction in LPNMF/LFDA, and \(\sigma\) for the RBF kernel used in SVM, are empirically tuned over wide ranges to get the optimal classification accuracy. The proposed algorithms, which are conducted by fusing the classification results both with original spectral features only as well as with the first-order-derivative features only, outperform traditional dimensionality-reduction and classification algorithms. When the number of training samples is 150 and the number of test samples is 750 per class, the classification accuracy of proposed algorithms is significantly higher than that of other algorithms. We also compare the computational cost of the classification methods. All experiments are carried out using MATLAB 7.6.0 on a 2.8GHz machine with 7.89GB of RAM, and the resulting execution times are shown in Table II.

Since labeling training samples for HSI classification is a complicated task, the number of training samples is often insufficient to reliably estimate the classifier models for each class. Hence, the classification accuracy of the proposed algorithms is studied over a range of total training samples from 425 to 1350. Fig. 2 shows that the proposed D-LPNMF-GMM-Fusion and D-LFDA-GMM-Fusion algorithms always produced the highest overall classification accuracies, even under small sample sizes. To compare the results on the whole test set, selecting 50 training samples per class randomly, the classification maps of our strategies
and the state-of-art algorithms are shown in Fig. 3. The proposed techniques significantly outperform other algorithms.

Table II. The classification performance for the Pavia University dataset.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Classification Algorithms</th>
<th>(σ = 0.6)</th>
<th>(r = 23)</th>
<th>(r = 9)</th>
<th>(r = 15)</th>
<th>(r = 9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asphalt</td>
<td>Train</td>
<td>77.07</td>
<td>78.67</td>
<td>91.20</td>
<td>88.27</td>
<td>89.08</td>
</tr>
<tr>
<td>Meadows</td>
<td>Test</td>
<td>83.87</td>
<td>91.07</td>
<td>93.87</td>
<td>89.47</td>
<td>96.17</td>
</tr>
<tr>
<td>Gravel</td>
<td></td>
<td>71.07</td>
<td>86.00</td>
<td>82.67</td>
<td>80.13</td>
<td>83.97</td>
</tr>
<tr>
<td>Trees</td>
<td></td>
<td>95.47</td>
<td>95.33</td>
<td>97.07</td>
<td>95.20</td>
<td>98.52</td>
</tr>
<tr>
<td>Metal sheets</td>
<td></td>
<td>100.00</td>
<td>99.87</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Bare soil</td>
<td></td>
<td>88.67</td>
<td>92.40</td>
<td>87.20</td>
<td>92.00</td>
<td>95.76</td>
</tr>
<tr>
<td>Bitumen</td>
<td></td>
<td>81.47</td>
<td>95.20</td>
<td>88.80</td>
<td>86.13</td>
<td>91.67</td>
</tr>
<tr>
<td>Bricks</td>
<td></td>
<td>71.73</td>
<td>84.00</td>
<td>75.47</td>
<td>86.67</td>
<td>84.00</td>
</tr>
<tr>
<td>Shadows</td>
<td></td>
<td>98.13</td>
<td>100.00</td>
<td>98.80</td>
<td>98.00</td>
<td>98.23</td>
</tr>
<tr>
<td>Overall Accuracy (%)</td>
<td></td>
<td>85.27</td>
<td>91.39</td>
<td>90.56</td>
<td>90.65</td>
<td>93.04</td>
</tr>
<tr>
<td>Kappa Coefficient</td>
<td></td>
<td>0.8395</td>
<td>0.9112</td>
<td>0.8917</td>
<td>0.9015</td>
<td>0.9227</td>
</tr>
<tr>
<td>Execution Time (s)</td>
<td></td>
<td>30.46</td>
<td>2.01</td>
<td>22.06</td>
<td>12.23</td>
<td>39.87</td>
</tr>
</tbody>
</table>

6. ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (project number: 61171154) and the Science and Technology Innovation Engineering Program for Shaanxi Provincial Key Laboratories.

Fig. 2. The classification performance of proposed algorithms versus training-data size.

5. CONCLUSION AND FUTURE WORK

In this paper, an HSI classification system based on spectral-derivative features and locality-preserving analysis was proposed. We can obtain the following conclusions: (1) the novel strategies capture the underlying statistical structure accurately and extract the useful derivative information sufficiently; (2) the classification performance of the D-LPNMF-GMM-Fusion and D-LFDA-GMM-Fusion approaches can be improved by increasing derivative orders; and (3) the proposed approaches are effective for addressing the problem of small training-sample sizes.

Although the proposed system provides effective classification results, we believe that there is room for further improvement. In future work, we plan to vary the derivative step-length (Δλ = λ−1 − λ−1) involved in (7) and to incorporate a spatial feature to improve the system’s classification performance.
7. REFERENCES


