

Decision fusion for hyperspectral image classification based on multiple features and locality-preserving analysis

Zhen Ye, Lin Bai & Yongjian Nian

To cite this article: Zhen Ye, Lin Bai & Yongjian Nian (2017) Decision fusion for hyperspectral image classification based on multiple features and locality-preserving analysis, European Journal of Remote Sensing, 50:1, 166-178, DOI: [10.1080/22797254.2017.1299556](https://doi.org/10.1080/22797254.2017.1299556)

To link to this article: <https://doi.org/10.1080/22797254.2017.1299556>



© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 Mar 2017.



Submit your article to this journal [↗](#)



Article views: 558



View Crossmark data [↗](#)

Decision fusion for hyperspectral image classification based on multiple features and locality-preserving analysis

Zhen Ye^a, Lin Bai^a and Yongjian Nian^b

^aSchool of Electronics and Control Engineering, Chang'an University, Xi'an, China; ^bSchool of Biomedical Engineering, Third Military Medical University, Chongqing, China

ABSTRACT

A novel fusion-classification system is proposed for hyperspectral image classification. Firstly, spectral derivatives are used to capture salient spectral features for different land-cover classes and a Gabor filter is applied to extract useful spatial features at neighbouring locations. Then, two locality-preserving dimensionality reduction methods are employed to reduce the dimensionality of data and preserve the local structure of neighbouring samples in the original image, derivative-feature and Gabor-feature domains. Finally, the classification results from Gaussian-mixture-model classifiers are fused by a decision-fusion approach. We have compared the proposed system to several traditional and state-of-the-art methods on two benchmark classification data sets. In both cases, our system achieved improved accuracy than the current best performing methods. Especially in the case of classification for the Indian Pines data set, the proposed DG-locality-preserving nonnegative matrix factorization (LPNMF) has 7% higher accuracy than support vector machine-Markov random field and LPNMF-G-MRF, and also has about 4% higher accuracy than Gabor-LPNMF. The proposed method has practical relevance as it shows potential to create smooth classification maps even for complex, spatially heterogeneous land-cover classes.

ARTICLE HISTORY

Received 27 January 2017
Revised 21 February 2017
Accepted 22 February 2017

KEYWORDS

Hyperspectral image classification; spectral derivative; Gabor filter; dimensionality reduction; locality-preserving; decision fusion

Introduction

Hyperspectral imagery (HSI) records hundreds of spectral bands for each pixel, each of which contains values that correspond to the reflected light in a defined range of the electromagnetic spectrum (Ye, Prasad, Li, Fowler, & He, 2014). Due to the detailed spectral information, HSI can be used to efficiently distinguish land-cover types (Kang, Li, & Benediktsson, 2014). However, the HSI often has a much greater dimensionality than the number of available training samples. The lack of training samples and the high computational burden caused by high-dimensional data processing are inevitable obstacles for designing HSI classifiers (Li, Zhang, & Zhang, 2015). To avoid this “curse of dimensionality” problem, dimensionality reduction is often employed to project the HSI data onto a lower-dimensional feature space. Principal component analysis (PCA) and Fisher’s linear discriminant analysis (LDA) are traditional methods used for HSI dimensionality reduction (Martinez & Kak, 2001). PCA aims to find projections with a minimal reconstruction error for the overall information content in the data, whereas LDA seeks to find projections that preserve the information necessary to separate target classes (Prasad & Bruce, 2008a). A popular method, which employed PCA-based dimensionality reduction as a

preprocessing to LDA, is named SLDA (Li, Prasad, & Fowler, 2013). Due to the influence of various factors, such as geometric features of material surfaces, the spectral signatures of pixels within the same land-cover class often have complicated distributions. The mentioned three dimensionality reduction methods all assume that the distributions of spectral signatures within one class of the HSI are Gaussian. However, the distributions of classes in HSI data are often not Gaussian and, in extreme cases, may have complicated multimodal structures.

To address this problem, locality-preserving dimensionality reduction methods are studied to exploit the complicated statistical structure of HSI data. These methods include, for example, locality-preserving projection (LPP) (He & Niyogi, 2004), locality-preserving non-negative matrix factorization (LPNMF) (Cai, He, Wang, Bao, & Han, 2009) and local Fisher’s discriminant analysis (LFDA) (Sugiyama, 2007). LPP is an unsupervised dimensionality reduction method which can preserve the local structure of each class. LPNMF combines the advantages of LPP and the nonnegative matrix factorization (NMF). The NMF decomposes the data into two non-negative matrices and extracts features for HSI classification, unmixing of HSI and other applications (Li, Prasad, Fowler, & Cui, 2012). LFDA combines

the properties of LDA and LPP, exploiting the advantages of each and preserving the neighbourhood relationships within the embedding by employing an “affinity” matrix. In Li et al. (2012), it is argued that LFDA-based dimensionality reduction followed by Gaussian-mixture-model (GMM) (Berge & Solberg, 2006) classifiers (named LFDA-GMM) could capture the underlying statistical structure accurately and obtain higher classification accuracies as compared to LDA-GMM.

Although LFDA-GMM was shown to outperform traditional dimensionality reduction methods (i.e. LDA-GMM) for HSI classification, it still suffers from the fact that the classifier exploits only original spectral signatures. During the last decade, many researchers have worked with other spectral features. In Bao, Chi, and Benediktsson (2013), spectral derivatives have been proven to be an effective measure to capture salient features of different land-cover classes. In our previous work (Ye et al., 2014), spectral derivatives were used for HSI classification, and two fusion-classification algorithms, called D-LPNMF and D-LFDA, were proposed. In the two algorithms, the locality-preserving dimensionality reduction method (LPNMF or LFDA) is used to reduce the dimensionality of the original image and derivative features. Following this, GMM classifiers are applied to classify the image, and a decision-fusion approach is used to fuse the classification results. It has been proved that derivative features are helpful to improve the classification performance in this fusion-classification system. However, there is further space for improvements. Since there is a high probability that two adjacent pixels belong to the same HSI class, spatial information can be helpful to create accurate classification maps. In the aforementioned spectral-feature methods, the spatial information at neighbouring locations is not considered. Recently, several strategies have been proposed for spatial feature extraction. In particular, one popular approach adopts a Bayesian maximum a posteriori formulation partitioned into class-conditional and class-prior probability terms, in which the Markov random field (MRF) (Xu & Li, 2014) was viewed as a post-processing stage for HSI classification. In Tarabalka, Fauvel, Chanussot, and Benediktsson (2010), SVM-MRF as a popular method was proposed for HSI classification. For this method, the support vector machine (SVM) (Camps-Valls & Bruzzone, 2005) and MRF were employed for class-conditional probability and class-prior probability of the classes, respectively. In Li, Prasad, and Fowler (2014), LFDA-G-MRF and LPNMF-G-MRF strategies have been proven to outperform SVM-MRF. In LFDA-G-MRF and LPNMF-G-MRF strategies, LFDA and LPNMF were employed for dimensionality reduction followed by GMM classifiers to preserve

multimodal structures of HSI, and then an MRF model as post-processing integrates the classification results to exploit spatial-context information. Recently, Gabor filters have also been successfully used for spatial feature extraction. In Bau, Sarkar, and Healey (2010), Shen and Jia (2011) and Jia, Zhu, Shen, and Li (2014), three-dimensional Gabor feature extraction approaches were found to represent useful spatial and spectral information. In Huo and Tang (2011), Zhang, Zhang, Tao, and Huang (2012) and Li and Du (2015), two-dimensional Gabor features extracted in PCA-projected subspaces have also been proven to be efficient for HSI classification.

The aforementioned algorithms, such as the MRF-based and the Gabor-based, can incorporate the spatial-contextual information into the classification process. However, only single features have been considered in these algorithms, which may lead to the loss of useful information and unstable classification accuracy. Specifically, spectral features cannot differentiate between objects made of the same material, while spatial features alone may fail to discriminate between objects that are made of different materials but resemble in shape and size (Liao, Dalla Mura, Chanussot, & Pizurica, 2016). In this work, we develop a new framework for the classification of hyperspectral scenes that pursues the combination of multiple features.

The new contribution of this research is an innovative framework to combine three processing steps for efficient classification of HSI. In the first step, spectral derivative and Gabor filters are combined to extract the spectral and spatial features of HSI and thereby capturing the salient features of different land-cover classes and the orientation and scale properties of HSI. In the second step, the locality-preserving dimensionality reduction is studied to avoid the distortion of the information contained in multimodal distribution and preserve the local structure of each class in both derivative-feature and Gabor-feature domains. Furthermore, GMM classifiers are employed after applying locality-preserving dimensionality reduction for preserving the multimodal distribution of classes for HSI. In the third step, the pipeline outputs of multiple features are combined effectively by decision fusion in the form of logarithmic-opinion-pool (LOGP) (Prasad & Bruce, 2008b). Merging the classification results of the original signatures, the derivative features and the Gabor features into a final decision, is an effective way to exploit useful information from different sources. Thus, we assume that the proposed framework can significantly increase the classification accuracy and result in smoother classification maps with fewer erroneous outliers, compared to some traditional and state-of-the-art algorithms.

Locality-preserving analysis

Dimensionality reduction, which is an essential preprocessing step in high-dimensional data spaces, can decrease the computational cost and may also improve the classification accuracy. Here, locality-preserving analysis is studied for dimensionality reduction while preserving the local structure of each class. LPP, as an unsupervised dimensionality reduction method, seeks to find a linear map, \mathbf{W} , that preserves the local statistical structure of neighbouring samples from the input data space. We consider a hyperspectral image data with training samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$, where d is the dimensionality of the data (the number of bands) and n is the total number of pixels per band. Class labels are given as $\omega = \{\omega_i\}_{i=1}^n \in \mathbb{R}\{1, 2, \dots, c\}$, where c is the number of classes. The “affinity” matrix between \mathbf{x}_i and \mathbf{x}_j are defined as $A_{i,j} \in [0, 1]$, where

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

where $\gamma_i = \|\mathbf{x}_i - \mathbf{x}_i^{(m)}\|$ denotes the local scaling of the data \mathbf{x}_i , and $\mathbf{x}_i^{(m)}$ is the m th nearest neighbour to \mathbf{x}_i . $A_{i,j}$ is an $n \times n$ symmetric matrix which measures the distance among data samples.

The eigenmap is calculated via computing the eigenvectors and eigenvalues of the following 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, $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian matrix and \mathbf{D} is a diagonal matrix with the i th diagonal element.

An example of dimensionality reduction for a two-dimensional multimodal synthetic data set with two classes is shown in Figure 1. To demonstrate the advantages of the locality-preserving approaches, the data set is projected by LPP and traditional dimensionality reduction (i.e. PCA). Although the data set

is projected onto a one-dimensional subspace, the example given in Figure 1(b) is in two-dimension for demonstration purposes. As seen in Figure 1, LPP not only can force far-apart data pairs of the same class to be close for preserving the within-class local structure, but also can obtain good between-class separation while PCA obviously fails for such a data set.

LPNMF

As a popular method to process high-dimensional data, the NMF focuses on the analysis of data matrices whose elements are nonnegative. The NMF seeks to find two nonnegative matrices $\mathbf{U} = [u_{iq}] \in \mathbb{R}^{d \times t}$ and $\mathbf{V} = [v_{jq}] \in \mathbb{R}^{n \times t}$, which minimize the following objective function:

$$\mathcal{O} = \sum_{i=1}^d \sum_{j=1}^n \left(x_{i,j} \log \frac{x_{i,j}}{y_{i,j}} - x_{i,j} + y_{i,j} \right) \quad (3)$$

where $\mathbf{Y} = \{y_{i,j}\} = \mathbf{U}\mathbf{V}^T$. Since the objective function is not convex in both the \mathbf{U} and \mathbf{V} variables simultaneously, it is difficult to study an algorithm to determine the global minimum of \mathcal{O} . Lee and Seung (1999) proposed an iterative algorithm to solve the NMF problem.

Although the NMF has been frequently applied in hyperspectral image processing tasks, it fails to consider the intrinsic geometric structure. As a consequence, LPNMF, which combines the advantages of LPP and NMF, was found to have more discriminating power for high-dimensional data (Cai et al., 2009). Compared to the traditional NMF, the objective function of the LPNMF can be defined as

$$\mathcal{O}' = \sum_{i=1}^d \sum_{j=1}^n \left(x_{i,j} \log \frac{x_{i,j}}{y_{i,j}} - x_{i,j} + y_{i,j} \right) + \lambda \mathfrak{R} \quad (4)$$

where λ is the regularization parameter. The first part of Equation (4) is the usual objective function of the NMF,

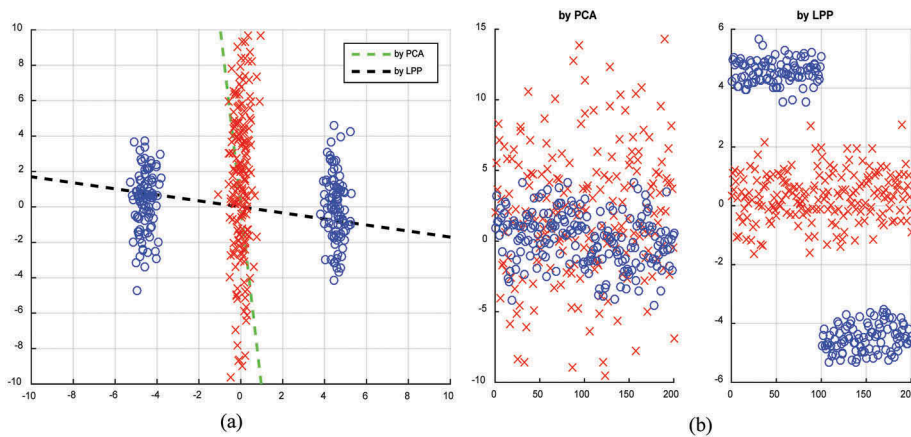


Figure 1. Example of dimensionality reduction. (a) Synthetic 2-D multimodal data and projection direction. (b) The data distribution after projection.

which uses Kullback–Leibler divergence between \mathbf{X} and \mathbf{Y} . In the second part, \mathfrak{R} is used to force a geometric locality constraint among points in the reduced-dimensional subspace $\mathbf{V}(t < d)$, expressed as follows:

$$\mathfrak{R} = \frac{1}{2} \sum_{j,s=1}^n \sum_{q=1}^t \left(v_{j,q} \log \frac{v_{j,q}}{v_{s,q}} - v_{s,q} + v_{s,q} \log \frac{v_{s,q}}{v_{j,q}} \right) \mathbf{W}_{j,s} \quad (5)$$

Since the local geometric structure can be effectively modelled through a nearest neighbour graph on a scatter of data points, a graph can be considered with N vertices where each vertex corresponds to a document in the corpus. The edge-weight matrix \mathbf{W} can be defined as

$$\mathbf{W}_{j,s} = \begin{cases} 1, & \mathbf{x}_j \in N_p(\mathbf{x}_s) \text{ or } \mathbf{x}_s \in N_p(\mathbf{x}_j) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $N_p(\mathbf{x}_s)$ denotes the set of p nearest neighbours of \mathbf{x}_s to measure the distance between points in the original space, \mathbf{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 \mathbf{U} and \mathbf{V} ,

$$u_{i,q} \leftarrow u_{i,q} \frac{\sum_j (x_{i,j} v_{j,q}) / \sum_q (u_{i,q} v_{j,q})}{\sum_j v_{j,q}} \quad (7)$$

$$\mathbf{v}_q \leftarrow \left[\sum_i u_{i,q} \mathbf{I} + \lambda \mathbf{L} \right]^{-1} \begin{bmatrix} v_{1,q} \sum_i \frac{x_{i,1} u_{i,q}}{\sum_q (u_{i,q} v_{1,q})} \\ v_{2,q} \sum_i \frac{x_{i,2} u_{i,q}}{\sum_q (u_{i,q} v_{2,q})} \\ \vdots \\ v_{n,q} \sum_i \frac{x_{i,n} u_{i,q}}{\sum_q (u_{i,q} v_{n,q})} \end{bmatrix} \quad (8)$$

where \mathbf{v}_q is the q th column of \mathbf{V} , \mathbf{I} is the $n \times n$ identity matrix and \mathbf{L} is the Laplacian graph of \mathbf{W} . Due to the limited spatial resolution of many HSI, the spectrum of a pixel may be a mixture of several endmembers. The LPNMF can be used to extract an endmember-based feature and preserve the intrinsic geometric structure of the HSI.

LFDA

As a supervised dimensionality reduction technique, the LFDA, which combines LDA and LPP, has been proposed to handle multimodal, non-Gaussian distributions. The LFDA preserves the neighbourhood relationships within the embedding by employing an “affinity” matrix, as performed in the LPP. In this part, the heat kernel of the LFDA as described in Equation (1) is employed. In the LFDA, the “local” between-class $\mathbf{S}^{(lb)}$ and within-class $\mathbf{S}^{(lw)}$ scatter matrices are scaled via the affinity matrix and defined as

$$\mathbf{S}^{(lb)} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{i,j}^{(lb)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (9)$$

$$\mathbf{S}^{(lw)} = \frac{1}{2} \sum_{i,j=1}^n \mathbf{W}_{i,j}^{(lw)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (10)$$

where $\mathbf{W}^{(lb)}$ and $\mathbf{W}^{(lw)}$ are $n \times n$ matrices

$$\mathbf{W}_{i,j}^{(lb)} = \begin{cases} A_{i,j}(1/n - 1/n_l), & \text{if } y_i = y_j = l \\ 1/n, & \text{if } y_i \neq y_j \end{cases} \quad (11)$$

$$\mathbf{W}_{i,j}^{(lw)} = \begin{cases} A_{i,j}/n_l, & \text{if } y_i = y_j = l \\ 0, & \text{if } y_i \neq y_j \end{cases} \quad (12)$$

The values of the in-class pairs are weighted by the affinity in $\mathbf{S}^{(lb)}$ and $\mathbf{S}^{(lw)}$, which are weakly dependent on the far-part, in-class pairs. The solution that maximizes the Fisher’s ratio within the context of the local scatter matrices is given by

$$\mathbf{S}^{(lb)} \boldsymbol{\phi} = \boldsymbol{\Lambda} \mathbf{S}^{(lw)} \boldsymbol{\phi} \quad (13)$$

where $\boldsymbol{\Lambda}$ is the diagonal eigenvalue matrix and $\boldsymbol{\phi}$ is the eigenvector matrix. Maximizing Fisher’s ratio as defined using the local scatter matrices, the solution \mathbf{T}_{LFDA} can be calculated by

$$\mathbf{T}_{\text{LFDA}} = \arg \max_{\mathbf{T} \in \mathbb{R}^{d \times r}} \text{tr}[(\mathbf{T}^T \mathbf{S}^{(lw)} \mathbf{T})^{-1} \mathbf{T}^T \mathbf{S}^{(lb)} \mathbf{T}] \quad (14)$$

\mathbf{T} is determined so that data pairs of different classes are apart and nearby data pairs of the same class are close. Since the far-apart data pairs of the same class are not assumed to be close, the LFDA can be viewed as a localized LDA. The modified Fisher’s ratio in the LFDA treats samples of a class within each mode independently so that it can estimate the dimensionality-reduction projection by these local scatter matrices and obtain good between-class separation while preserving local neighbourhoods. It is hence expected that the LFDA will be superior to the LDA when the distributions of input data are complex and even multimodal.

In conventional approaches, locality-preserving analysis is presented for dimensionality reduction using only the spectral features, which may be sub-optimal. In this article, these two advanced locality-preserving dimensionality-reduction methods – LPNMF and LFDA – are utilized to reduce the dimensionality of both local spatial features and spectral features. The LPNMF and LFDA need to optimize the parameter r , which is the reduced dimensionality of the LPNMF projection and LFDA projection. In this work, we varied r over a wide range of values and chose the optimal r as those that maximized the training accuracy for the Indian Pines data set and the Pavia University data set.

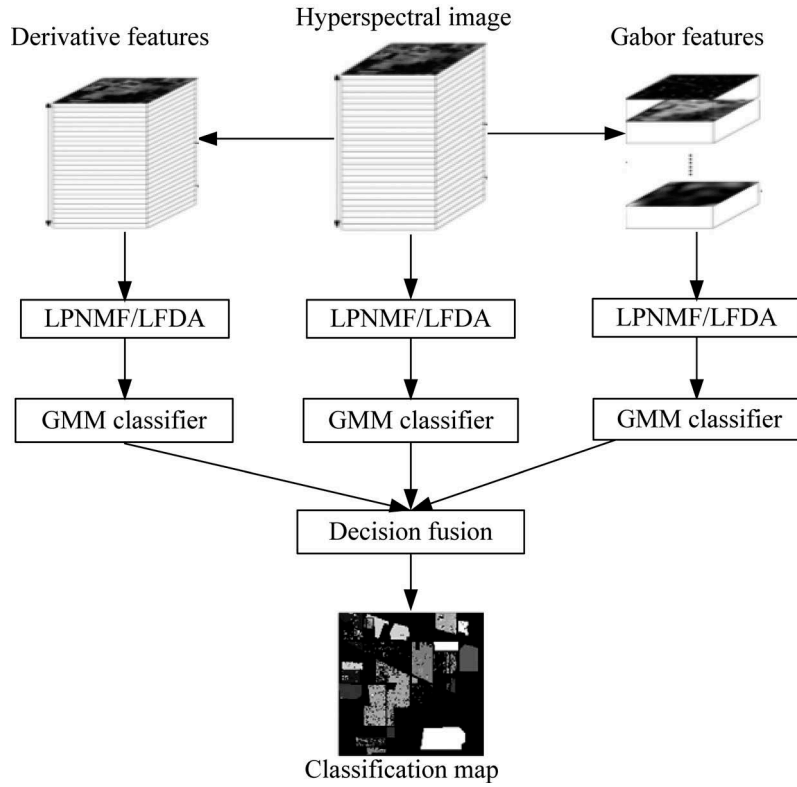


Figure 2. Flow chart of the proposed framework.

Proposed fusion-classification system

Framework

Spectral derivatives have been proven to be an effective measure to capture salient features of different land-cover classes (Bao et al., 2013). In our previous work, D-LFDA and D-LPNMF were studied by fusing original spectral signatures and derivative features (Ye et al., 2014). It also has been proven that Gabor filters can effectively capture the orientation and scale of the physical structures for the HSI spatial feature extraction (Zhang et al., 2012; Li et al. 2015). However, little work has been conducted to study the combination of derivative features and Gabor features for HSI classification. In this work, we focus on a fusion-classification framework, in which the locality-preserving dimensionality reduction followed by GMM classifiers is introduced into both derivative-feature and Gabor-feature domains. The objective here is to preserve the local structure of each class, obtain the output labels by GMM classifiers and then combine the results by using a decision-fusion approach for a final decision.

As shown in Figure 2, the proposed fusion-classification framework consists of four steps. First, the derivative features are extracted in the spectral direction and the Gabor features are extracted in the spatial direction. Then, either the LPNMF or LFDA is used to reduce the dimensionalities of the original image, the derivative features and the Gabor features. Afterwards, GMM classifiers are

subsequently applied to classify the image. Finally, the classification results are merged by the LOGP decision-fusion approach. A detailed description of the proposed algorithms (named DG-LPNMF and DG-LFDA) is described as Strategy 1.

Strategy 1 The proposed DG-LPNMF/LFDA

Input: Training data $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, class labels ω_i , testing samples $\mathbf{y} \in \mathbb{R}^{d \times n}$

The spectral derivative is conducted on \mathbf{y} , giving the first-order derivative features via Equation (15);

A Gabor-filter is applied to \mathbf{y} , giving Gabor features via Equation (17);

Either the LPNMF or LFDA followed by the GMM classifier is implemented to the derivative features, Gabor features and \mathbf{y} ;

Classification results are fused using Equation (23);

Output: class (y)

Multiple feature analysis

Derivative features

The spectral derivatives, estimated by obtaining the slope information from the reflectance curve over the available wavelengths in the spectrum, can be used to alleviate spectral distortions and to eliminate systematic errors in the original HSI data (Kalluri, Prasad, & Bruce, 2010; Bao et al., 2013). Traditionally, the spectral derivatives are evaluated in different step-lengths and different orders for

hyperspectral image classification. The first-order spectral derivative is defined as

$$\frac{\partial y_{i,j,\lambda}}{\partial \lambda_l} = \frac{y_{i,j,\lambda_l} - y_{i,j,\lambda_k}}{\lambda_l - \lambda_k} \quad (15)$$

where y_{i,j,λ_k} and y_{i,j,λ_l} are the spectral reflectance values corresponding to the wavelengths λ_k and λ_l , respectively, where k and l represent the different spectral bands. The difference between k and l is considered to be the step-length. If the step-length $\Delta\lambda = \lambda_l - \lambda_k > 0$, the wavelength windows are fixed. In this article, we set $\Delta\lambda = 1$.

Similar to the first-order derivative, the n th-order spectral derivative is defined as

$$\frac{\partial^n y_{i,j,\lambda}}{\partial \lambda^n} = \frac{\partial}{\partial \lambda} \left[\frac{\partial^{n-1} y_{i,j,\lambda}}{\partial \lambda^{n-1}} \right] \quad (16)$$

In our previous work (Ye et al., 2014), two fusion-classification algorithms were studied to fully exploit the spectral information of HSI. D-LFDA indicates the fusion-classification algorithm-based LFDA-GMM using original spectral signatures and derivative features, and D-LPNMF indicates the fusion-classification-algorithm-based LPNMF-GMM using original spectral signatures and derivative features. In previous work, we conducted experiments to study the classification performance of D-LFDA and D-LPNMF with fusing different order derivatives. The results demonstrated that the performance of both approaches improves slowly with an increasing number of derivative features and the growth of the classification accuracy diminishes for more than eight derivative features. As in this work, the application of Gabor features further increased the complexity of the workflow; we followed a trade-off between classification accuracy and system complexity by employing only the first-order derivative features.

Gabor features

The Gabor filter, which is a sinusoidal function modulated by a Gaussian envelope, can be viewed as an orientation dependent bandpass filter and can effectively capture the orientation and scale of the physical structures for the HSI spatial feature extraction (Li et al., 2015). However, the HSI contains a wealth of spectral information over a wide range of the spectrum and the wavelength interval is relatively small, which adds to the processing complexity and statistical ill-conditioning in the classification tasks. To improve the efficiency, we considered dimensionality reduction before the Gabor feature extraction. Since Gabor features extracted in PCA-projected feature spaces have been proven to deliver very good results (Huo & Tang, 2011), (Zhang et al., 2012) and (Li & Du, 2015), we employed this strategy in our feature-extraction pipeline.

In a two-dimensional (a, b) coordinate system, the Gabor filter, including both a real and imaginary component, can be represented as

$$g(a, b; \delta, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{a'^2 + \gamma^2 b'^2}{2\sigma^2}\right) \times \exp(j(2\pi \frac{a'}{\delta} + \psi)) \quad (17)$$

with

$$a' = a \cos \theta + b \sin \theta \quad (18)$$

$$b' = -a \sin \theta + b \cos \theta \quad (19)$$

where δ represents the wavelength of the sinusoidal factor, ψ is the phase offset and γ is the spatial aspect ratio (the default value was set to 0.5 in (Daugman, 1985) and (Clausen & Jernigan, 2000) specifying the ellipticity of the support of the Gabor function. $\psi = 0$ and $\psi = \pi/2$ return the real and imaginary parts of the Gabor filter, respectively (Chen, Li, Su, & Liu, 2014). In Equations (18) and (19), θ represents the orientation separation angle between the Gabor kernels. The parameter σ is the standard derivation of the Gaussian envelope and is determined by δ and the spatial frequency bandwidth, bw , as

$$\sigma = \frac{\delta}{\pi} \sqrt{\frac{\ln 2}{2} \frac{2^{bw} + 1}{2^{bw} - 1}} \quad (20)$$

In our proposed fusion-classification system, all eight orientations $(0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8})$ for each band pass are initially considered and system parameters (δ and bw) are tuned for optimal classification results. For example, we discuss the overall classification accuracy (OA) of the proposed algorithms for different δ and bw for the Indian Pines data set in Figure 3. It can be observed that the classification performance of the proposed algorithms is affected by the parameter δ but insensitive to the parameter bw .

GMM

GMM can be viewed as a combination of two or more normal Gaussian distributions. In typical GMM representation, a probability density function for $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n \in \mathbb{R}^{d \times n}$ as the sum of K Gaussian components (the default value is 5 in (Li et al., 2012)) can be estimated as

$$P(\mathbf{X}) = \sum_{k=1}^K \alpha_k N(\mathbf{X}, \mu_k, \Sigma_k) \quad (21)$$

where

$$N(\mathbf{X}, \mu_k, \Sigma_k) = \frac{1}{2\pi^{d/2} |\Sigma_k|^{1/2}} \times \exp\left[-\frac{1}{2} (\mathbf{X} - \mu_k)^T \Sigma_k^{-1} (\mathbf{X} - \mu_k)\right] \quad (22)$$

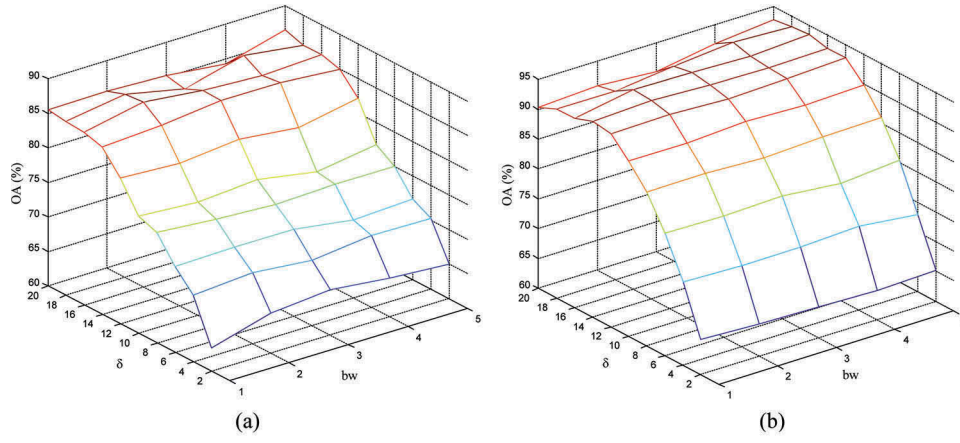


Figure 3. Overall classification accuracy (%) versus varying δ and bw for the Indian Pines data set using 50 labelled samples and 50 testing samples per class. (a) DG-LFDA. (b) DG-LPNMF.

represents the k th Gaussian component of the mixture. The variables α_k , μ_k and Σ_k are the mixing weights, mean vector and covariance matrix, respectively, which are expressed by the parameter vector $\Theta(\alpha_k, \mu_k, \Sigma_k)$. Once the optimal number of components has been determined, the parameters for the mixture model can be estimated by using an iterative optimization strategy – expectation maximization (Dundar & Landgrebe, 2002). In recent years, GMM classifiers have been shown to be well-suited for remotely sensed data. In Li et al. (2012) a LFDA-GMM strategy is proposed which achieved higher performances for hyperspectral image classification as compared to traditional LDA-GMM. We also investigated the LPNMF combined with GMM (named LPNMF-GMM) and observed good classification results under the proposed framework. In this work, we focus on discussing multimodal distribution of classes for HSI and preserving the within-class local structure. GMM classifiers are hence a natural fit for this topic.

Decision fusion

In this article, the LPNMF-GMM and LFDA-GMM are employed as the basic classification strategies for multiple features, such as the derivative features and Gabor features. We propose a fusion-classification system based on the LOGP decision fusion rule to combine the distinct decisions from different feature spaces into a final decision. The LOGP, which is a popular soft-decision fusion rule (Prasad & Bruce, 2008a), entails the use of posterior probabilities or some class-membership function from every classifier for making the final decision. In the LOGP, the distinct decisions from different classifiers are independent, and the output is a weighted product of different probability distributions (Li et al., 2015).

The global class-membership function for the LOGP can be defined as

$$P(w_j|\mathbf{X}) = \prod_{i=1}^n p_i(w_j|\mathbf{X})^{\alpha_i} \quad (23)$$

$$\Rightarrow \log P(w_j|\mathbf{X}) = \sum_{i=1}^n \alpha_i \log p_i(w_j|\mathbf{X})$$

where α_i is the confidence score for the i th classifier, w is the class label, i is the classifier index, n is the number of classifiers and class j is detected in the bank of classifiers.

Experiments

Data description and experimental set-up

All experiments of the study were conducted using MATLAB. We applied two hyperspectral data sets. The first data set is the Indian Pines collected by the National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer sensor. The data set represents a vegetation-classification scenario with 145×145 pixels with a pixel size of 20 m and 220 spectral bands. The original Indian Pines data set consists of 16 land-cover classes and some of them contain a small number of samples. Since the probability distribution obtained from such a small number of training samples cannot well represent their statistical characteristics, we selected a subset of eight classes (see Table 1). For this data set, an average of 50 training samples per class and 8624 testing samples (the ratio of testing and training is approximately 21.6:1) are set up for experiments.

The second data set is located in an urban environment and provided by the University of Pavia. The imagery was collected by the Reflective Optics System Imaging Spectrometer sensor. The data set has 103 spectral bands with 610×340 pixels and a pixel size of 1.3 m. The data set has nine classes, with 50 training samples per class and 24,776 testing samples

Table 1. Class labels and train-test distribution of samples for the Indian Pines data set.

	Class	Train	Test
1	Corn-no till	50	1460
2	Corn-min till	50	834
3	Grass/pasture	50	497
4	Hay-windowed	50	489
5	Soybean-no	50	968
6	Soybean-min	50	2468
7	Soybean-clean	50	614
8	Woods	50	1294
	Total	400	8624

Table 2. Class labels and train-test distribution of samples for the Pavia University data set.

	Class	Train	Test
1	Asphalt	50	6631
2	Meadows	50	18,649
3	Gravel	50	2099
4	Trees	50	3064
5	Painted sheets	50	1345
6	Bare soil	50	5029
7	Bitumen	50	1330
8	Bricks	50	3682
9	Shadows	50	947
	Total	450	42,776

(the ratio of testing and training is approximately 55:1). For this data set, the number of training and test samples for each class is summarized in Table 2.

Parameter selection

Existing classification algorithms, such as the SVM-based and the dimensionality-reduction-based, need to optimize the system parameters for any HSI classification task. For the SVM-based, the kernel parameter, σ , is an important factor that impacts the generalization ability in the kernel-induced space. For the LFDA-based and the LPNMF-based, we need to find an optimal parameter, r , which is the reduced dimensionality of the LFDA projection and the LPNMF projection. We varied the parameter σ from 0.1 to 1.0 and r over a wide range from 5 to 40. The optimal parameters were selected by maximizing the training accuracy.

The proposed two fusion-classification algorithms are named DG-LFDA and DG-LPNMF. Both employ Gabor filters to extract spatial information of HSI. For the Gabor-feature-based algorithms, all eight orientations for each pass band are initially considered and system parameters are tuned for optimal classification results. First, we present the parameters of the Gabor filter with different δ and bw for the Indian Pines data set (see Figure 3). For this data set, the available training samples and testing samples are both 50 per class. Figure 3(a,b) illustrate OA of the proposed DG-LFDA and DG-LPNMF versus the varying δ and bw . We observe that the classification performance of the proposed two fusion-classification algorithms is affected by the

parameter δ but is insensitive to the parameter bw . We selected $\delta = 18$ and $bw = 4$ as the optimal parameters in the following experiments. As a result of the Gabor features extracted in the PCA-projected subspaces, the number of principle components (PCs) should be viewed as a parameter for the proposed system. For the LFDA and LPNMF, the r for dimensionality reduction is expected to vary with the experimental data set. Figure 4(a,b) shows the OA of the proposed DG-LFDA and DG-LPNMF versus varying the PC and r . For the DG-LFDA, the optimal values of PC and r are 20 and 7, respectively. For the DG-LPNMF, the optimal values of PC and r are 16 and 33, respectively. Similar experiments using the Pavia University data set are performed for finding the appropriate parameters. Accordingly, we selected $\delta = 20$ and $bw = 5$ as the optimal parameters of Gabor filter for the Pavia University data set. For the DG-LFDA, the optimal values of PC and r are 16 and 7, respectively. For the DG-LPNMF, the optimal values of PC and r are 20 and 29, respectively.

Classification results

To quantify the efficacy of the proposed system, we compare the proposed fusion-classification algorithms (DG-LFDA and DG-LPNMF) with SVM (Camps-Valls & Bruzzone, 2005), SVM-MRF (Tarabalka et al., 2010), IFRF (Kang et al., 2014) and several corresponding methods (LDA-GMM, LPP-GMM, LFDA-GMM, LFDA-GMM (Li et al., 2012), D-LFDA (Ye et al., 2014), D-LPNMF (Ye et al., 2014), LFDA-G-MRF (Li et al., 2014) and LPNMF-G-MRF (Li et al., 2014). Table 3 shows the classification results for the two data sets. Two mathematical criteria are derived from the confusion matrix: the OA and the Kappa (κ) coefficient. To avoid any bias, the experiments are repeated 20 times, reporting the average classification accuracy and the standard deviation in Table 3.

We investigated the performance of three SVM-classifier methods, including traditional SVM, SVM-MRF and IFRF. Since IFRF is also designed to fuse multiple features and published recently, we added IFRF as comparison to the experiments. The code of IFRF is available on the web.¹ From Table 3, it can be seen that, although IFRF perform much better than traditional SVM, the OA and κ of IFRF is unsatisfactory when compared with SVM-MRF for the Indian Pines data set. The proposed DG-LFDA and DG-LPNMF always outperform SVM and SVM-MRF for the Indian Pines data set and the Pavia University data set. Compared with IFRF, DG-LPNMF shows comparable performance and DG-LFDA has a slight advantage for the Pavia University data set, while the improvements from both DG-LPNMF and DG-LFDA are very good for the Indian Pines data set.

¹<http://xudongkang.weebly.com>.

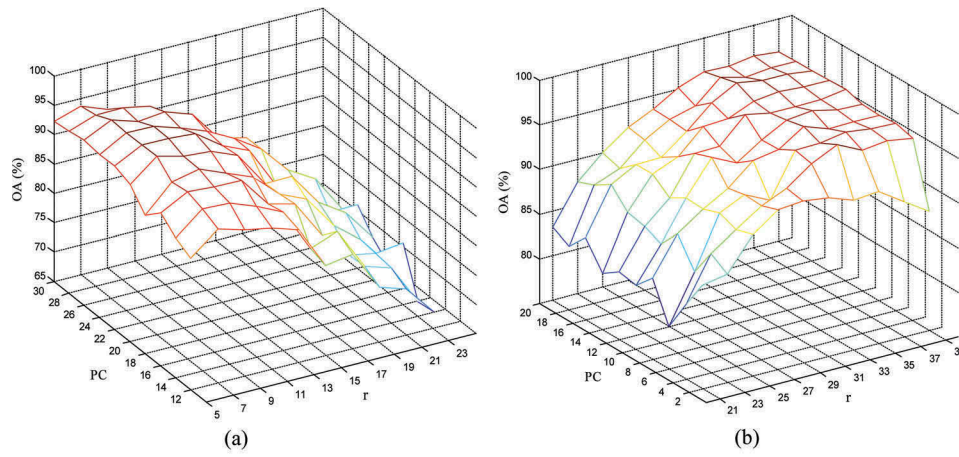


Figure 4. Overall classification accuracy (%) versus varying PC and r for the Indian Pines data set using 50 labelled samples and 50 testing samples per class. (a) DG-LFDA. (b) DG-LPNMF.

Table 3. Classification results for the Indian Pines data set and the Pavia University data set.

Algorithms	Indian Pines		Pavia University	
	OA (%)	κ coefficient	OA (%)	κ coefficient
SVM	77.46 (± 1.71)	0.7337 (± 0.0195)	84.43 (± 1.28)	0.7991 (± 0.0155)
SVM-MRF	88.04 (± 1.14)	0.8578 (± 0.0135)	91.09 (± 2.70)	0.8846 (± 0.0336)
IFRF	84.62 (± 1.62)	0.8168 (± 0.0189)	94.24 (± 1.42)	0.9250 (± 0.0184)
LDA-GMM	62.79 (± 1.29)	0.5629 (± 0.0135)	74.94 (± 1.83)	0.6800 (± 0.0197)
LFDA-GMM	64.17 (± 2.04)	0.5767 (± 0.0218)	81.15 (± 1.94)	0.7585 (± 0.0229)
D-LFDA	67.85 (± 1.94)	0.6222 (± 0.0199)	86.96 (± 2.09)	0.8301 (± 0.0252)
Gabor-LFDA	87.08 (± 1.71)	0.8439 (± 0.0210)	89.39 (± 2.23)	0.8714 (± 0.0271)
DG-LFDA	93.23 (± 1.15)	0.9192 (± 0.0170)	95.27 (± 1.46)	0.9382 (± 0.0187)
LFDA-G-MRF	82.09 (± 2.89)	0.7840 (± 0.0351)	90.06 (± 2.05)	0.8732 (± 0.0248)
LPP-GMM	58.34 (± 1.97)	0.5136 (± 0.0205)	77.54 (± 1.89)	0.7148 (± 0.0218)
LPNMF-GMM	73.01 (± 1.67)	0.6806 (± 0.0194)	84.57 (± 1.78)	0.8006 (± 0.0211)
D-LPNMF	74.92 (± 1.39)	0.7004 (± 0.0149)	87.64 (± 2.61)	0.8398 (± 0.0312)
Gabor-LPNMF	91.88 (± 1.15)	0.9026 (± 0.0136)	90.12 (± 2.32)	0.8856 (± 0.0279)
DG-LPNMF	95.60 (± 0.74)	0.9473 (± 0.0089)	94.36 (± 1.50)	0.9259 (± 0.0189)
LPNMF-G-MRF	88.17 (± 2.59)	0.8491 (± 0.0303)	93.95 (± 1.65)	0.9217 (± 0.0197)

From the results of GMM-classifier methods, it inferred that the performance of LFDA-GMM and LPNMF-GMM are better than LDA-GMM and LPP-GMM, while D-LFDA and D-LPNMF are superior to LFDA-GMM and LPNMF-GMM, respectively. As spatial-spectral classification algorithms, LFDA-G-MRF and LPNMF-G-MRF achieve much higher OA and κ than aforementioned spectral classification algorithms, wherein the performance of LPNMF-G-MRF is excellent for both the Indian Pines data set and the Pavia University data set. Moreover, we investigated the performance of two Gabor-feature-based algorithms (named Gabor-LFDA and Gabor-LPNMF). From the results of Table 3, these two Gabor-feature-based algorithms can achieve improved OA and κ compared to the derivative-feature-based algorithms (such as D-LFDA and D-LPNMF). Especially for the Indian Pines data set, Gabor-LFDA shows over 19% higher OA than D-LFDA, and Gabor-LPNMF yields about 17% higher OA than D-LPNMF. However, the performance of the proposed DG-LFDA and DG-LPNMF exceeds those of the traditional and the state-of-art algorithms based on locality-preserving dimensionality reduction. For the Indian Pines data set, the proposed

DG-LFDA has an OA that is 25% and 11% higher than the D-LFDA and LFDA-G-MRF respectively, and the proposed DG-LPNMF has an OA 20% and 7% higher than the D-LPNMF and LPNMF-G-MRF. The performance of the proposed DG-LPNMF and DG-LFDA is also better than all aforementioned GMM-classifier methods for the Pavia University data set described in Table 3.

Figure 5 shows the classification maps resulting from all approaches for the Indian Pines data set with the same labelled samples as listed in Table 1. It can be observed that four spectral-spatial algorithms (including LFDA-G-MRF, LPNMF-G-MRF, DG-LFDA and DG-LPNMF) are superior to other spectral algorithms, tending to result in smoother classification maps with fewer erroneous outliers. MRF-based methods, such as SVM-MRF, LFDA-G-MRF and LPNMF-G-MRF, can smooth classification maps by capturing spatial contextual information as compared with SVM, LFDA-GMM and LPNMF-GMM correspondingly. But upon closer inspection for LFDA-G-MRF and LPNMF-G-MRF, the mislabelling of large contiguous regions can be observed for the classes *Corn-no till* with *Corn-min till*.

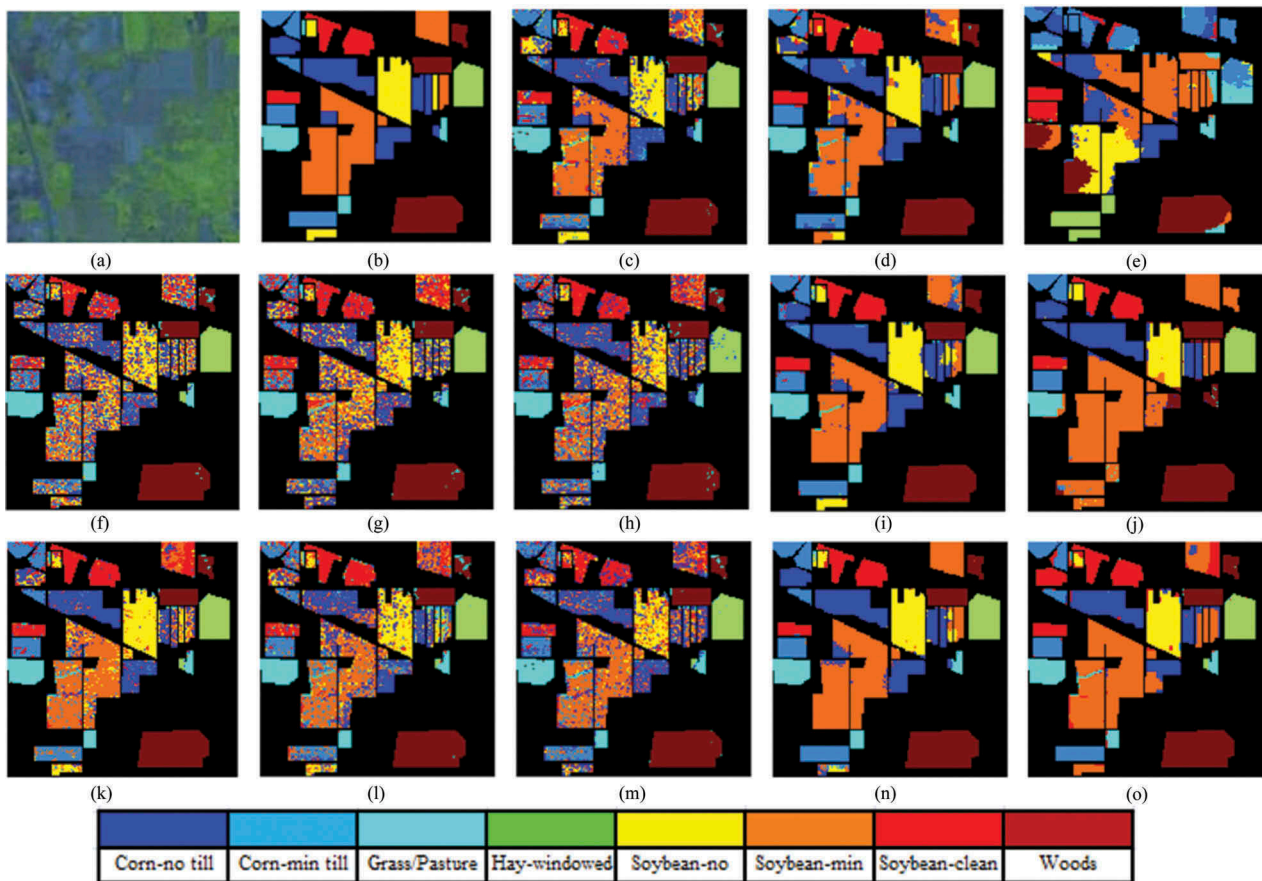


Figure 5. Classification maps for the Indian Pine data set. (a) False colour image. (b) Ground-truth. (c) SVM. (d) SVM-MRF. (e) IFRF. (f) LDA-GMM. (g) LFDA-GMM. (h) D-LFDA. (i) DG-LFDA. (j) LFDA-G-MRF. (k) LPP-GMM. (l) LPNMF-GMM. (m) D-LPNMF. (n) DG-LPNMF. (o) LPNMF-G-MRF.

Additionally, we observe that D-LFDA and D-LPNMF cannot clearly improve classification maps as compared with LFDA-GMM and LPNMF-GMM, while the proposed DG-LFDA and DG-LPNMF can clear and smooth the classification maps significantly due to the ability of Gabor features to capture additional spatial information. From what has been discussed earlier, we may conclude that the proposed system is superior to current MRF-based methods and Gabor features have a bigger contribution than derivative features in the system for improved classification maps. Figure 6 illustrates that the classification maps for the Pavia University data set is consistent with Table 2. It is clear that the classification maps of the proposed DG-LFDA and DG-LPNMF have less noise and are more accurate. Furthermore, the execution time (in seconds) of the proposed approaches using the Indian Pine data set and the Pavia University data set is shown as follows. The experiments are carried out using MATLAB on a 2.8 GHz machine with 8 GB of RAM. For the Indian Pine data set, the execution times are 74.94 for DG-LFDA and 74.57 for DG-LPNMF; for the Pavia University data set, the execution times are 381.21 for DG-LFDA and 421.13 for DG-LPNMF.

Since labelling training samples for hyperspectral image classification is a complicated task, the number of training samples is often insufficient to reliably estimate the classifier models for each class. To match the challenging conditions, we investigate the influence of the different training sample sizes. Hence, we report the performance of the traditional methods (LFDA-GMM and LPNMF-GMM), the state-of-art methods (IFRF, LFDA-G-MRF and LPNMF-G-MRF) and the proposed methods (DG-LFDA and DG-LPNMF) as measured by the OA, along with 95% confidence intervals. For both the Indian Pine and the Pavia University data sets, the OA of the algorithms mentioned earlier is studied over an available range of training samples (from 50 to 100 per class).

Figure 7 demonstrates that the performance of the proposed DG-LFDA and DG-LPNMF obviously achieve much higher OA than all of the corresponding algorithms for the Indian Pine data set. For Pavia University data set as shown in Figure 8, the performance of DG-LFDA performs consistently better than other algorithms, while DG-LPNMF has a close performance to IFRF. When 100 samples per class are used for training, DG-LPNMF approaches a 98% OA for the Indian Pine data set, and DG-LFDA approaches an OA higher than 97% for the Pavia University data set.

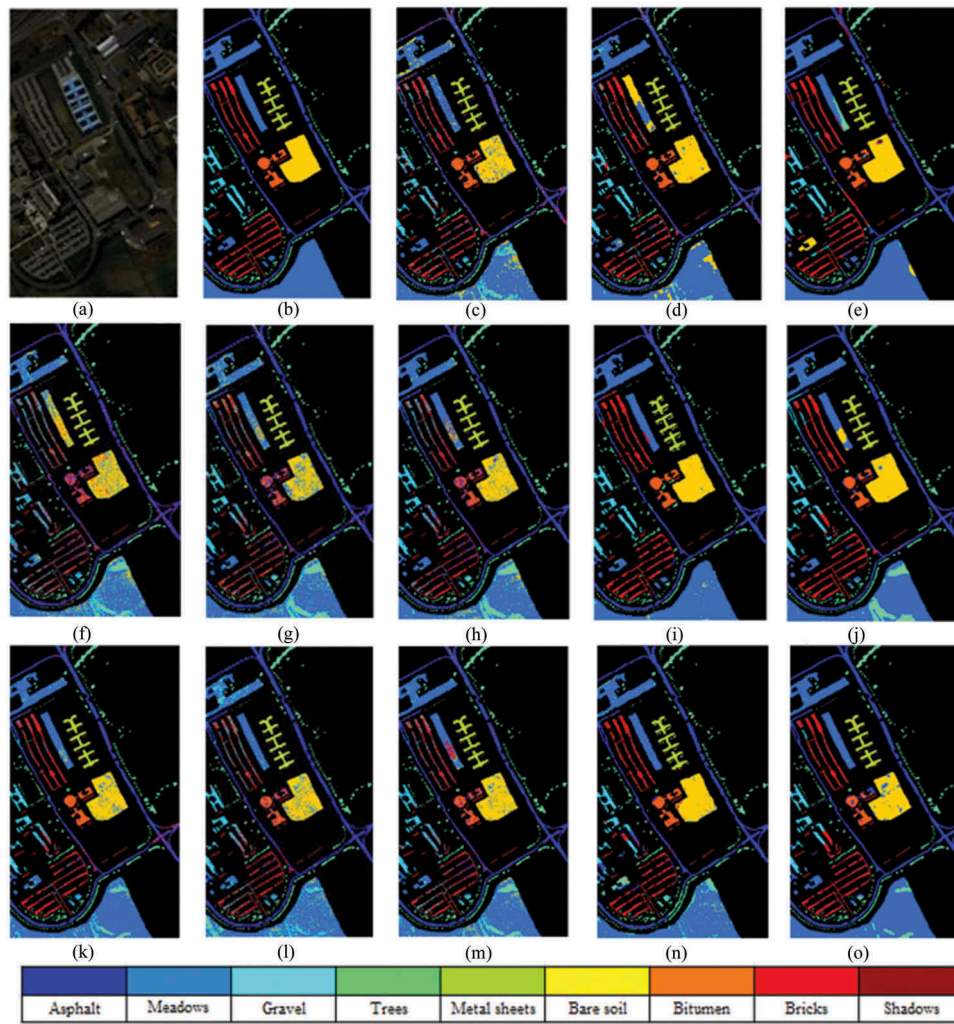


Figure 6. Classification maps for the Pavia University data set. False colour image. (b) Ground-truth. (c) SVM. (d) SVM-MRF. (e) IFRF. (f) LDA-GMM. (g) LFDA-GMM. (h) D-LFDA. (i) DG-LFDA. (j) LFDA-G-MRF. (k) LPP-GMM. (l) LPNMF-GMM. (m) D-LPNF. (n) DG-LPNMF. (o) LPNMF-G-MRF.

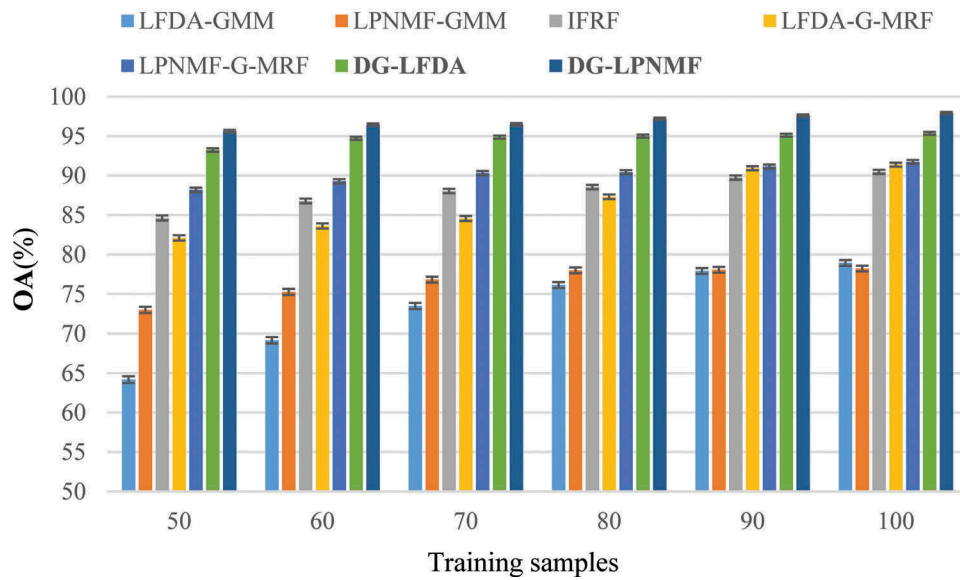


Figure 7. Overall classification accuracy (%) as a function of training samples for the Indian Pines data set.

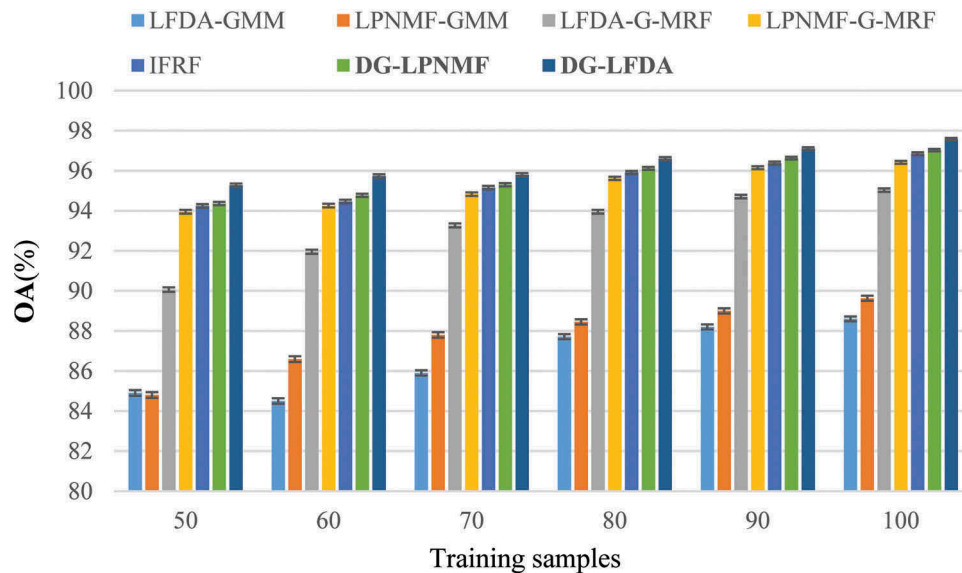


Figure 8. Overall classification accuracy (%) as a function of training samples for the Pavia University data set.

Conclusion

Traditional algorithms for hyperspectral image classification exploit only spectral information or spatial information. In this article, spectral derivatives and Gabor filters were first combined to extract both spectral and spatial features. Afterwards, a locality-preserving dimensionality reduction followed by GMM classifiers was conducted in multiple-feature domains to avoid the distortion of the information contained in multimodal distribution and preserve the local structure of each class. Moreover, an LOGP-based decision fusion strategy was invoked to merge classification results into a final decision. The experimental results demonstrated that the proposed system can significantly increase the classification accuracy and result in smoother classification maps with fewer erroneous outliers, compared to the traditional and state-of-the-art algorithms.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China [41601344, 61601059], the Fundamental Research Funds for the Central Universities [310832171006] and Chongqing Research Program of Basic research and Frontier Technology [No. cstc2016jcyjA0539]. The authors would like to thank the Editor-in-Chief, the anonymous editor and the reviewers for their insightful comments and suggestions in improving the quality of this paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported in part by the National Natural Science Foundation of China [41601344,

61601059], the Fundamental Research Funds for the Central Universities [310832171006] and Chongqing Research Program of Basic research and Frontier Technology [No. cstc2016jcyjA0539].

References

- Bao, J., Chi, M., & Benediktsson, J.A. (2013). Spectral derivative features for classification of hyperspectral remote sensing images: Experimental evaluation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2), 594–601. doi:10.1109/JSTARS.2013.2237758
- Bau, T.C., Sarkar, S., & Healey, G. (2010). Hyperspectral region classification using a three-dimensional Gabor filterbank. *IEEE Transactions on Geoscience Remote Sensing*, 48(9), 3457–3464. doi:10.1109/TGRS.2010.2046494
- Berge, A., & Solberg, A.H.S. (2006). Structured Gaussian components for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 44(11), 3386–3396. doi:10.1109/TGRS.2006.880626
- Cai, D., He, X., Wang, X., Bao, H., & Han, J. (July, 2009). Locality preserving nonnegative matrix factorization. *Proceedings of the International Joint Conference on Artificial Intelligence*, Pasadena, California, USA.
- Camps-Valls, G., & Bruzzone, L. (2005). Kernel-based methods for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 43(6), 1351–1362. doi:10.1109/TGRS.2005.846154
- Chen, C., Li, W., Su, H., & Liu, K. (2014). Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sensing*, 6(6), 5795–5814. doi:10.3390/rs6065795
- Clausiand, D.A., & Jernigan, M.E. (2000). Designing Gabor filters for optimal texture separability. *Pattern Recognition*, 33(11), 1835–1849. doi:10.1016/S0031-3203(99)00181-8
- Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7), 1160–1169. doi:10.1364/JOSAA.2.001160

- Dundar, M.M., & Landgrebe, D. (2002). A model-based mixture-supervised classification approach in hyperspectral data analysis. *IEEE Transactions on Geoscience Remote Sensing*, 40 (12), 2692–2699. doi:10.1109/TGRS.2002.807010
- He, X., & Niyogi, P. (2004). *Locality preserving projections*. Advances in Neural Information Processing System, USA: MIT Press.
- Huo, L.Z., & Tang, P. (2011, July). Spectral and spatial classification of hyperspectral data using SVMs and Gabor texture. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Vancouver, BC, Canada.
- Jia, S., Zhu, Z., Shen, L., & Li, Q. (2014). A two-stage feature selection framework for hyperspectral image classification using few labeled samples. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4), 1023–1035. doi:10.1109/JSTARS.2013.2282161
- Kalluri, H.R., Prasad, S., & Bruce, M.L. (2010). Decision-level fusion of spectral reflectance and derivative information for robust hyperspectral land cover classification. *IEEE Transactions on Geoscience Remote Sensing*, 48(11), 4047–4058. doi:10.1109/TGRS.2010.2072787
- Kang, X., Li, S., & Benediktsson, J.A. (2014). Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Transactions on Geoscience Remote Sensing*, 52(6), 3742–3752. doi:10.1109/TGRS.2013.2275613
- Lee, D.D., & Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788–791. doi:10.1038/44565
- Li, J., Zhang, H., & Zhang, L. (2015). Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 53(10), 5338–5351. doi:10.1109/TGRS.2015.2421638
- Li, W., Chen, C., Su, H., & Du, Q. (2015). Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Transactions on Geoscience Remote Sensing*, 53(7), 3681–3693. doi:10.1109/TGRS.2014.2381602
- Li, W., & Du, Q. (2014). Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4), 1012–1022. doi:10.1109/JSTARS.2013.2295313
- Li, W., Prasad, S., & Fowler, J.E. (2013). Noise-adjusted subspace discriminant analysis for hyperspectral imagery classification. *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1374–1378. doi:10.1109/LGRS.2013.2242042
- Li, W., Prasad, S., & Fowler, J.E. (2014). Hyperspectral image classification using Gaussian mixture model and Markov random field. *IEEE Geoscience and Remote Sensing Letters*, 11(1), 153–157. doi:10.1109/LGRS.2013.2250905
- Li, W., Prasad, S., Fowler, J.E., & Bruce, L.M. (2012). Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience Remote Sensing*, 50(4), 1185–1198. doi:10.1109/TGRS.2011.2165957
- Li, W., Prasad, S., Fowler, J.E., & Cui, M. (2012, July). Locality-preserving nonnegative matrix factorization for hyperspectral image classification. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany.
- Liao, W., Dalla Mura, M., Chanussot, J., & Pizurica, A. (2016). Fusion of spectral and spatial information for classification of hyperspectral remote sensed imagery by local graph. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2), 583–594. doi:10.1109/JSTARS.2015.2498664
- Martinez, A.M., & Kak, A.C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. doi:10.1109/34.908974
- Prasad, S., & Bruce, L.M. (2008a). Limitations of principal components analysis for hyperspectral target recognition. *IEEE Geoscience and Remote Sensing Letters*, 5(4), 625–629. doi:10.1109/LGRS.2008.2001282
- Prasad, S., & Bruce, L.M. (2008b). Decision fusion with confidence-based weight assignment for hyperspectral target recognition. *IEEE Transactions on Geoscience Remote Sensing*, 46(5), 1448–1456. doi:10.1109/TGRS.2008.916207
- Shen, L., & Jia, S. (2011). Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification. *IEEE Transactions on Geoscience Remote Sensing*, 49(12), 5039–5046. doi:10.1109/TGRS.2011.2157166
- Sugiyama, M. (2007). Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of Machine Learning Research*, 8(5), 1027–1061. <http://www.cs.wvu.edu/~richas/ML-Papers/KLFDA.pdf>
- Tarabalka, Y., Fauvel, M., Chanussot, J., & Benediktsson, J. A. (2010). SVM- and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 7(4), 736–740. doi:10.1109/LGRS.2010.2047711
- Xu, L., & Li, J. (2014). Bayesian classification of hyperspectral imagery based on probabilistic sparse representation. *IEEE Geoscience and Remote Sensing Letters*, 11(4), 823–827. doi:10.1109/LGRS.2013.2279395
- Ye, Z., He, M.Y., Fowler, J.E., & Du, Q. (2014, July). Hyperspectral image classification based on spectra derivative features and locality preserving analysis. *Proceedings of IEEE China Summit and International Conference on Signal and Information Processing*, Xi'an, China.
- Ye, Z., Prasad, S., Li, W., Fowler, J.E., & He, M.Y. (2014). Classification based on 3D DWT and decision fusion for hyperspectral image analysis. *IEEE Geoscience and Remote Sensing Letters*, 11(1), 173–177. doi:10.1109/LGRS.2013.2251316
- Zhang, L.F., Zhang, L.P., Tao, D.C., & Huang, X. (2012). On combining multiple features for hyperspectral remote sensing image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 50(3), 879–893. doi:10.1109/TGRS.2011.2162339