



Modified LDE for Dimensionality Reduction of Hyperspectral Image

Lei He, Hongwei Yang, and Lina Zhao^(✉)

The Beijing University of Chemical Technology, Beijing 10029, China
{2018200881, yanghw, zhao1n}@buct.edu.cn

Abstract. Hyperspectral image (HSI) has shown promising results in many fields because of its high spectral resolution. However, redundancy and noise in spectral dimension seriously affect the classification of HSI. For this reason, many popular dimensionality reduction (DR) methods are proposed to solve the problem. The local discriminant embedding (LDE) as an effective non-linear method for DR can be more discriminative by constructing two neighborhood graphs. However, HSI is very easy influenced by noise, and the LDE algorithm based on K nearest neighborhood is highly susceptible to interference from extreme point, which may lead to inaccurate graph construction and poor performance of classification. To overcome the problem and retain the advantages of LDE, a modified local discriminant embedding (MLDE) is firstly applied on HSI by constructing neighborhood graphs on a new spectral feature space instead of the original space. We use variance to characterize the pixels similarity of the same class and use covariance to characterize the separation of different classes of pixels. The combination of variance and covariance makes pixels in the same class to be closer and makes greater separation of pixels from different classes, which enhances classification performance of HSI. The way of representing data by using variance and covariance can attenuate the effects of noise. The Log-Euclidean metric is used to capture the similarity between spectral vectors, which can provide a more accurate similarity evaluation than euclidean distance. The experimental results of two hyperspectral datasets demonstrate the effectiveness of our proposed MLDE method.

Keywords: Classification of hyperspectral image (HSI) · Dimensionality reduction (DR) · Local discriminant embedding (LDE) · Log-Euclidean metric

1 Introduction

Hyperspectral image (HSI) usually consists of hundreds of spectral bands from the visible spectrum to the infrared spectrum [1]. Each pixel of HSI can be represented by a high dimensional spectral vector. It's because HSI's rich spectral

Supported by the National Natural Science Foundation of China, 11301021 and 11571031.

information that it has not only attracted the attention of the remote sensing community, but also aroused great interest in other fields, for instance, military [2], agriculture [3], urban planning, and environmental monitoring [4]. It is known that classification plays a crucial role in these fields. However, HSI generates a large amount of irrelevant or redundant data that causes a number of issues including significantly increased computation time, computational complexity and the classification performance especially when the training datasets are limited. A number of classical dimensionality reduction (DR) algorithms are explored to address these issues.

One of classic linear methods of DR is principle component analysis (PCA) [5]. But as an unsupervised methods, PCA doesn't take advantage of class label information. Another one of classic linear methods of DR is linear discriminant analysis (LDA) [6], as a supervised method, it often suffers from the small sample size problem. And the biggest disadvantage of these linear methods is the failure to discover the nonlinear structure inherent in HSI.

Since nonlinear techniques have the merit of preserving geometrical structure of data manifold, it can overcome the above-problem. Laplacian eigenmaps (LE) [7], local linear embedding (LLE) [8] and other manifold learning algorithms have been successfully applied to DR for HSI. Besides, as a linear version of LE, locality preserving projection (LPP) [9] has been introduced. In order to overcome the difficulty of LDA tending to produce undesirable results when the samples in a class is multimodal non-Gaussian class distributions [10], local Fisher's discriminant analysis (LFDA) [11] which having the advantages of LDA and LPP at the same time was introduced. After that, unlike LPP which uses only one graph to describe the geometry of the sample, local discriminant embedding (LDE) [12] method using two graphs to characterize the geometry structure of the sample was proposed. One as an intrinsic graph to characterize the compact nature of the sample, and the other as a penalty graph to describe the internal separation of the sample. Thus, LDE is more discriminative than LPP. The advantage of LDE is that it can make the data from the same class keep their intrinsic neighbor relations, and it also makes the data in different classes no longer close to each other. However, one thing in common among these above-mentioned methods is that the calculation of the affinity matrix is based on K nearest neighborhood, which is sensitive to outlier samples.

To overcome the above-problem, a graph embedding (GE) frame work [13] was proposed. In order to represent the sparse nature of the samples, a sparse graph embedding (SGE) [14] was developed. Later, a sparse graph-based discriminant analysis (SGDA) [15] model was developed by exploiting the class label information, resulting in a better performance than SGE. Above this, based on SGDA, sparse and low-rank graph discriminant analysis (SLGDA) [16] was proposed by increasing local information of samples. Recently, since considering curves changing description among spectral bands, a graph-based discriminant analysis with spectral similarity (GDA-SS) [17] method was proposed.

Each pixel of HSI is a high dimensional spectral vector that directly displays the spectral reflectance of the targets in different bands. Under an ideal

condition, the same targets should have the same spectral characteristics. Nevertheless, HSI is very easy influenced by environment change (i.e. atmosphere and illumination) and instrument problem (i.e. sensor) in the real world. And K nearest neighborhood based on euclidean distance is usually used to compute the similarity between two vectors, which is highly susceptible to interference from extreme point. These may lead to inaccurate graph construction and poor performance of classification. Inspired by the region covariance descriptor in [18] and the superiority of the second-order statistic representing data, a novel modified local discriminant embedding (MLDE) is proposed by constructing neighborhoods on a new spectral feature space instead of the original space. We use variance to characterize the pixel similarity of the same class and use covariance to characterize the separation of different classes of pixels. Considering the symmetric positive definite nature of covariance matrix lying on a Riemannian manifold, the Log-Euclidean metric is used to capture the similarity, which has a better effect than the euclidean distance. The main advantages in this paper are summarized as follows: (a) The combination of variance and covariance enables data points in the same class to be closer and enables greater separation of data points from different classes, which enhances classification performance of HSI. (b) The way of representing data by using variance and covariance can attenuate the effects of noise, which can better handle with noise in HSI. (c) The Log-Euclidean metric can provide a more accurate similarity evaluation than euclidean distance, which can better express the characteristics of spectral information.

2 Related Work

2.1 Local Discriminant Embedding (LDE)

Assume a hyperspectral dataset having N samples is denoted as $X = \{x_i\}_i^N$ existing in a $\mathbb{R}^{m \times 1}$ feature space, where m is the number of bands. And class labels $y_i \in 1, 2, \dots, C$, where C is the number of classes.

LDE which is defined for manifold learning and pattern classification tries to obtain an optimal projection matrix by considering the class label information of the data points and the local neighborhood information between data points. Specifically, the LDE algorithm can be described as follows.

Steps 1: Construct neighborhood graphs. An intrinsic graph G and a penalty graph G' can be constructed by K nodes of K nearest neighborhood (KNN) over all the data point.

Steps 2: Compute affinity weights. An affinity matrix W of the intrinsic graph G and an affinity matrix W' of the penalty graph G' can be computed as follows:

$$w_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t) & x_j \in O(K, x_i) \\ & \text{or } x_i \in O(K, x_j) \\ & \text{and } y_i = y_j; \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

and

$$w'_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2/t) & x_j \in O(K, x_i) \\ & \text{or } x_i \in O(K, x_j) \\ & \text{and } y_i \neq y_j; \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $O(K, x_i)$ represents the K nearest neighborhood of data x_i and the parameter t is a kernel width parameter.

The optimization problem of LDE is described as follows:

$$\begin{aligned} & \arg \min_P \sum_{i,j} \|P^T x_i - P^T x_j\|^2 w_{ij} \\ & \text{s.t. } \sum_{ij} \|P^T x_i - P^T x_j\|^2 w'_{ij} = 1 \end{aligned} \quad (3)$$

Steps 3: Complete the embedding. The projection matrix P can be obtained by solving the eigenvectors corresponding to the H smallest nonzero eigenvalues of the following generalized eigenvalue problem:

$$X(D - W)X^T P = \wedge X(D' - W')X^T P \quad (4)$$

where \wedge is a diagonal eigenvalue matrix. D and D' are diagonal matrices with $D_{ii} = \sum_{j=1}^N W_{i,j}$ and $D'_{ii} = \sum_{j=1}^N W'_{i,j}$.

2.2 Region Covariance Descriptor for HSI

As a robust and very novel data descriptor, region covariance descriptor has been successful and effectively applied to many computer vision problems [19,20]. Consider a HSI data $\mathbf{X} \in \mathbb{R}^{l \times w \times m}$ with m representing the number of bands and $l \times w$ representing the spatial structure. Consider a three order spatial-spectral tensor $x \in \mathbb{R}^{(2n-1) \times (2n-1) \times m}$ as a small patch of $\mathbf{X} \in \mathbb{R}^{l \times w \times m}$, the central of x is a pixel, the rest of the central of x is its local region neighborhood. Therefore, the pixels of a HSI data $\mathbf{X} \in \mathbb{R}^{l \times w \times m}$ can be denoted as $\{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^{(2n+1) \times (2n+1) \times m}$ denotes the i th pixel and N is the number of pixels [18]. And x_s ($s = 1, 2, \dots, (2n + 1) \times (2n + 1)$) is a spectral vector in the region of interest around the i th hyperspectral pixel. Then, a spectral region covariance descriptor C_i can be obtained by the Eq. (5).

$$\begin{aligned} C_i &= \frac{1}{S-1} \sum_{s=1}^S (x_s - \mu_i)(x_s - \mu_i)^T \\ \mu_i &= \frac{1}{S} \sum_{s=1}^S x_s \end{aligned} \quad (5)$$

where S is the number of spectral vectors in the region of interest, and μ_i is the mean vector. Meantime, C_i is considered to be the feature of \mathbf{X}_i .

3 Our Work

3.1 Variance and Covariance for HSI

Inspired by the region covariance descriptor in [18], we want to introduce the variance and covariance instead of the region covariance descriptor to attenuate the effects of noise, because in this paper the hyperspectral dataset is used as input in the form of a vector, not a tensor. Consider a hyperspectral dataset denoted as $X = \{x_i | x_{i1}, x_{i2}, \dots, x_{im}\}_{i=1}^N$ existing in a $\mathbb{R}^{m \times 1}$ feature space, where m is the number of bands. Then, a spectral variance C_i ($i = 1, 2, \dots, N$) and a covariance C_{ij} ($i, j = 1, 2, \dots, N$) can be obtained by the Eq. (6).

$$\begin{aligned} C_i &= \frac{1}{m-1} \sum_{k=1}^m (x_{ik} - \mu_i)(x_{ik} - \mu_i)^T \\ \mu_i &= \frac{1}{m} \sum_{k=1}^m x_{ik} \\ C_{ij} &= \frac{1}{m-1} \sum_{k=1}^m (x_{ik} - \mu_i)(x_{jk} - \mu_j)^T \end{aligned} \quad (6)$$

where μ_i is the spectral mean value. Meantime, the variance C_i is considered to be the feature of x_i , and the covariance C_{ij} is considered to be the feature of between x_i and x_j .

3.2 Modified Local Discriminant Embedding (MLDE)

Suffered by the euclidean distance which is sensitive for noise and the data which contain inevitable noise created by environment change (i.e. atmosphere and illumination) and instrument problem (i.e. sensor), the LDE algorithm may lead to inaccurate graph construction and a poor performance of classification. In this section, we propose an MLDE algorithm to overcome the problem.

Like LDE, the intrinsic graph G and the penalty graph G' should be constructed firstly. Nevertheless, in MLDE, the difference is that we use the variance features $\{C_i\}_{i=1}^N$ and the covariance features $\{C_{ij}\}_{i,j=1}^N$ obtained by Eq. (6) to construct the intrinsic graph and the penalty graph denoted as G_{var} and G'_{cov} , respectively. Due to the variance features and the covariance features lying on a Riemannian manifold, the Log-Euclidean metric is a good choice to compute the affinity.

$$D_{LE}(C_i, C_j) = |\log(C_i) - \log(C_j)| \quad (7)$$

Then, the affinity matrix W_{var} of the intrinsic graph G_{var} and the affinity matrix W_{cov} of the penalty graph G'_{cov} can be computed as follows:

$$w_{var \ ij} = \begin{cases} \exp(-D_{LE}(C_i, C_j)^2/t) & C_j \in O(K, C_i) \\ & \text{or } C_i \in O(K, C_j) \\ & \text{and } y_i = y_j; \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

and

$$w'_{cov\ ij} = \begin{cases} \exp(-|\log(C_{ij})|^2/t) & C_{ij} \in O(K, C_{ii}) \\ & \text{or } C_{ii} \in O(K, C_{ij}) \\ & \text{and } y_i \neq y_j; \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where $O(K, C_i)$ represents the K nearest neighborhood of covariance feature C_i and the parameter t is a kernel width parameter.

The optimization problem of MLDE is described as follows:

$$\begin{aligned} J(P) &= \arg \min_P \sum_{i,j} \|P^T x_i - P^T x_j\|^2 w_{var\ ij} \\ \text{s.t. } & \sum_{i,j} \|P^T x_i - P^T x_j\|^2 w'_{cov\ ij} = 1 \end{aligned} \quad (10)$$

Similarity to LDE, the optimization problem (10) can be rewritten as (11) by the nature of trace.

$$\begin{aligned} J(P) &= \arg \min_P \sum_{i,j} \|P^T x_i - P^T x_j\|^2 w_{var\ ij} \\ &= \arg \min_P \sum_{i,j} \text{tr}\{(P^T x_i - P^T x_j)(P^T x_i - P^T x_j)^T\} w_{var\ ij} \\ &= \arg \min_P \sum_{i,j} \text{tr}\{P^T (x_i - x_j)(x_i - x_j)^T P\} w_{var\ ij} \end{aligned} \quad (11)$$

By $w_{var\ ij}$ is a scalar and the operation of trace is linear, the Eq. (11) can be rewritten as (12):

$$\begin{aligned} J(P) &= \arg \min_P \text{tr}\{P^T \sum_{i,j} ((x_i - x_j)w_{var\ ij}(x_i - x_j)^T)P\} \\ &= \arg \min_P \text{tr}\{P^T (2X D_{var} X^T - 2X W_{var} X^T)P\} \\ &= \arg \min_P 2\text{tr}\{P^T X (D_{var} - W_{var}) X^T P\} \end{aligned} \quad (12)$$

where D_{var} is a diagonal matrix with $D_{var\ ii} = \sum_{j=1}^N W_{var\ ij}$. Then, the optimization problem (10) can be rewritten as (13):

$$\begin{aligned} J(P) &= \arg \min_P 2\text{tr}\{P^T X (D_{var} - W_{var}) X^T P\} \\ \text{s.t. } & 2\text{tr}\{P^T X (D_{cov} - W_{cov}) X^T P\} = 1 \end{aligned} \quad (13)$$

The projection matrix P can be obtained by solving the eigenvectors corresponding to the H smallest nonzero eigenvalues of the following generalized eigenvalue problem:

$$X(D_{var} - W_{var})X^T P = \wedge X(D_{cov} - W_{cov})X^T P \quad (14)$$

Thus, MLDE for hyperspectral image classification is carried out following the steps in Algorithm 1.

Algorithm 1 MLDE-Based HSI Classification Algorithm

Input: Training set X_{train} , the class labels of training set y_{train} , testing set X_{test} , the class labels of testing set y_{test} , where $X_{train} = \{x_i | x_i \in R^{m \times 1}, i = 1, \dots, N_{train}\}$, $X_{test} = \{x_i | x_i \in R^{m \times 1}, i = 1, \dots, N_{test}\}$, $y_{train} \in \{1, \dots, C\}$, and $y_{test} \in \{1, \dots, C\}$.

Output: The class labels of testing set.

Initialize: $K = k, H = h$.

- 1: Structure the intrinsic graph G_{var} and the penalty graph G_{cov} .
- 2: Compute the affinity matrices W_{var} and W_{cov} by Eq. (8) and Eq. (9), respectively;
- 3: Compute the projection matrix P by Eq. (14).
- 4: Obtain the training set $B_{train} = \{b_i | b_i \in R^{h \times 1}, i = 1, 2, \dots, N_{train}\}$ in a low dimensional space, where $B_{train} = P^T X_{train}$.
- 5: Obtain the testing set $B_{test} = \{b_i | b_i \in R^{h \times 1}, i = 1, 2, \dots, N_{test}\}$ in a low dimensional space, where $B_{test} = P^T X_{test}$.
- 6: Perform classification on the testing set B_{test} by support vector machine.

4 Experimental Results and Discussions

In this section, we will apply MLDE on two hyperspectral datasets. Firstly, we introduce the experimental datasets. Secondly, how to choose the best experimental parameters would be given. Finally, The classification accuracy and classification maps on compared algorithms and MLDE algorithm would be shown. The MLDE algorithm is implemented by matlab. The results are generated on a personal computer equipped with an Intel Core i7-3370 with 3.40 GHz. The personal computer's memory is 4GB.

Table 1. Number of training and testing samples for the University of Pavia dataset

Class	Name	Training	Testing
1	Asphalt	530	6101
2	Meadows	1492	17157
3	Gravel	168	1931
4	Trees	245	2819
5	Painted Metal Sheets	108	1237
6	Bare Soil	402	4627
7	Bitumen	106	1224
8	Self-Blocking Bricks	295	3387
9	Shadows	76	871
Total		3422	39354

4.1 Experimental Dataset

The first experimental dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the University of Pavia in Italy. The image includes 610×340 pixels and 115 spectral bands in the wavelength

Table 2. Number of training and testing samples for the Salinas dataset

Class	Name	Training	Testing
1	Brocoli-green-weeds-1	100	1909
2	Brocoli-green-weeds-1	186	3540
3	Fallow	99	1877
4	Fallow-rough-plow	70	1324
5	Fallow-smooth	134	2544
6	Stubble	198	3761
7	Celery	179	3400
8	Grapes-untrained	564	10707
9	Soil-vinyard-develop	310	5893
10	Corn-senesced-green-weeds	164	3114
11	Lettuce-romaine-4wk	53	1015
12	Lettuce-romaine-5wk	96	1831
13	Lettuce-romaine-6wk	46	870
14	Lettuce-romaine-7wk	54	1016
15	Vinyard-untrained	363	6905
16	Vinyard-vertical-trellis	90	1717
Total		2706	51423

range $0.43 - 0.86 - \mu\text{m}$. In our experiments, 12 spectral bands covering noisy are removing. Then, a total of 103 bands is used. Thus, the image contains 9 different classes and a total of 42776 ground-truth samples (Table 1).

The second experimental dataset was acquired by the National Aeronautics and Space Administration’s Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS) sensor over Salinas Valley in California. The image includes 512×127 pixels and 204 bands after 20 water-absorption bands are removed. Thus, the image contains 16 different classes and a total of 54129 ground-truth samples.

8% and 5% samples in each class are randomly selected as training samples in the University of Pavia dataset and the Salinas dataset, respectively. And the rest are chosen as the testing samples. More detailed information of the number of training and testing samples is summarized in Tables 1 and 2.

4.2 Experiment Parameters

The SVM is used to verify the proposed MLDE algorithm. The SVM classifier is implemented by libsvm (the kernel is rbf, the penalty parameter is 1000 and the sigma is searched in $\{0.01, 0.05, 0.5, 1, 5, 10, 50, 100, 500, 1000\}$). And to demonstrate the benefits of MLDE algorithm, the experimental results would be compared with nine other classical algorithm of DR, i.e., PCA, LDA, LPP, LDE, LFDA, LGDA, SGDA, SLGDA, GDA-SS.

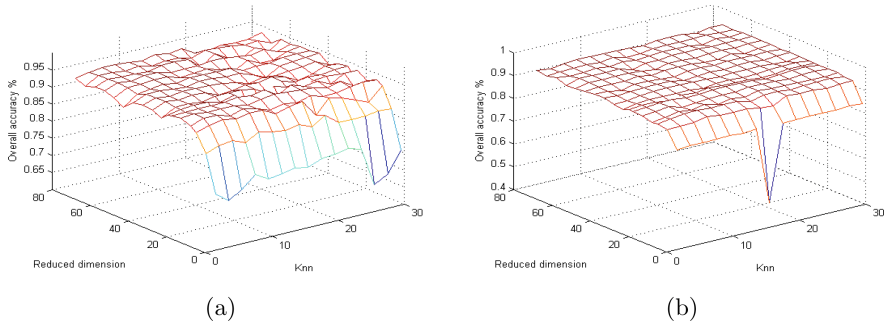


Fig. 1. The overall accuracy corresponding to different reduced dimensionality and different K for MLDE on two hyperspectral datasets

It is very easy to note that the reduced dimensionality and the value of the K nearest neighborhood are two important parameters, which have a significant influence on the performance of the classification.

If the K is too small, it may reduce classification accuracy. And if the K is too large, it would increase computational complexity, increase the noise and reduce the classification effect. To find a good value of K , the even numbers are chosen from 2 to 60, and the reduced dimensionality is searched in the range of $\{2, 7, 12, 15, 20, 25, 27, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75\}$. To have a better presentation, we only show the range of 2–30 of the value of K in Fig. 1.

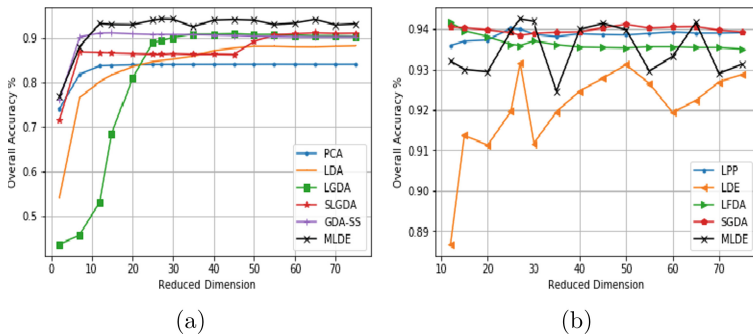


Fig. 2. The overall accuracy corresponding to different reduced dimensionality for MLDE on the University of Pavia dataset

Figure 1 shows the classification performances of MLDE in different K for two hyperspectral datasets. It can be seen from Fig. 1 that the overall accuracy would increase as K increasing when K is at a relatively small value, while the overall accuracy would decline as K increasing when K is at a relatively big value. It’s noticed that the overall accuracy will be stable and less affected by k when the spectral number is in a high position. From Fig. 1, the highest value of

overall accuracy are 94.28% and 93.30% in the University of Pavia dataset and the Salinas dataset, at the same time, K are 12 and 22, respectively.

Thus, the K is respectively fixed as 12 and 22 according to Fig. 1. Next, a good value of the reduced dimensionality would be searched in the above spectral range, the way in which other algorithms do, e.g. LFDA, SGDA, SLGDA.

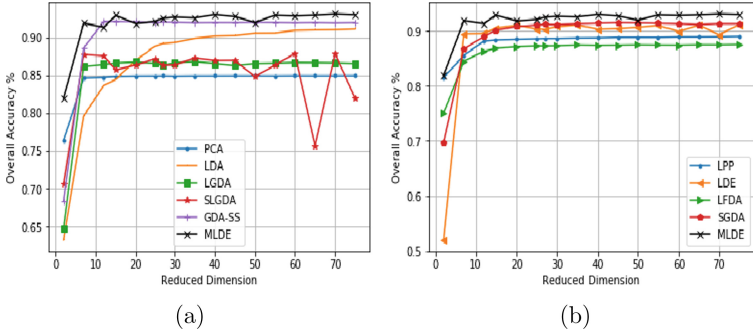


Fig. 3. The overall accuracy corresponding to different reduced dimensionality for MLDE on the Salinas dataset

Figure 2 illustrates the overall accuracy corresponding to the reduced dimensionality H for all the algorithms mentioned in the University of Pavia dataset. The performance is poor when the reduced dimensionality is low, and it would increase and stabilize as the reduced dimensionality increasing. From Fig. 2(a), PCA, LDA, LGDA, SLGDA, and GDA-SS apparently don't have a better classification performance than MLDE. Although the curves of LPP, LFDA, SGDA and MLDE alternatively rise, the highest point 94.28% can be found on the MLDE curve in Fig. 2(b). So, the reduced dimensionality being set as 27 can be considered a good choice.

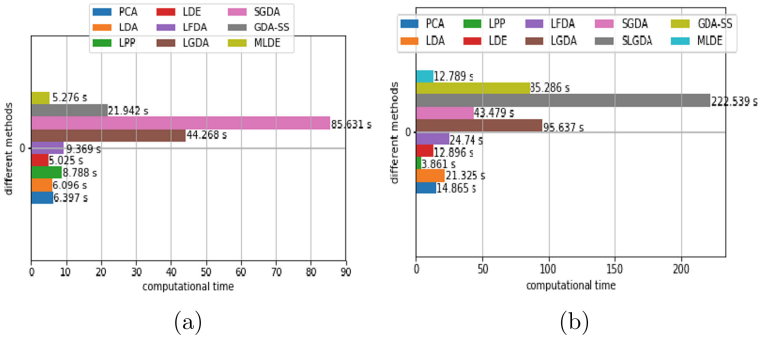


Fig. 4. The computational time of different methods in two hyperspectral datasets: (a) the University of Pavia dataset, (b) the Salinas dataset

Figure 3 also illustrates the overall accuracy corresponding to the reduced dimensionality H for all the algorithms mentioned in the Salinas dataset. The performance is poor when the reduced dimensionality is low, and it would increase and stabilize as the reduced dimensionality increasing. From Fig. 3 (b), LPP, LDE, LFDA, SGDA apparently don't have a better classification performance than MLDE. Except the curver of GDA-SS has some intersection with the curver of MLDE, the other methods don't have a better value than MLDE in Fig. 3(a), and the highest overall accuracy 93.12% will be found in the curve of MLDE. So, the reduced dimensionality being set as 70 can be considered a good choice.

From Fig. 4(a), the computational time of MLDE is 5.276 s and ranked second with a small difference of 0.251 s of the first place. Because the computational time of SLGDA is 1564.8 s is very big will cause the figure don't have a good presentation, it don't be shown in Fig. 4(a). And from Fig. 4(b), the computational time of MLDE is 12.879 s and ranked third.

4.3 Experimental Results

Through our experiments, for the University of Pavia dataset, the value of K would be set as 12, the reduced dimensionality would be set as 27, for the salinas dataset, the value of K would be set as 22, the reduced dimensionality would be set as 70.

The each class's accuracy, overall accuracy (OA), average accuracy (AA) and kappa coefficient of two hyperspectral datasets are listed in Tables 3 and 4.

From Table 3, the MLDE achieves the best classification performance in the class 3, the class 7, and the class 8, respectively. And the classification accuracy of OA, AA, and κ are all better than other compared methods. On details, the OA of MLDE increases from 0.44% to 10.89%, the AA of MLDE increases from 1% to 17.08%, and the κ of MLDE increases from 0.59% to 15.27%, when compared with other methods. Especially, the classification performance of the class 7 is 83.83% when the accuracy of other methods is basically no more than 80%, and the classification performance of the class 8 is 91.53% when the accuracy of other methods is basically no more than 90%. Meaawhile, when other methods achieve the best results in a certain class, the results of MLDE are not inferior, for instance, the class1, the class2, the class 5, and the class 9.

From Table 4, although the MLDE only achieves the best classification performance in the class 16, the classification performance of the other classes has a good performance, for example, the classification accuracy of the class 1 and the class 12 are also good. And the classification accuracy of OA is better than other compared methods. On details, the OA of MLDE increases from 3.64% to 7.21%.

Figure 5 illustrates the classification maps resulting from the classification of those methods in the University of Pavia dataset. In Fig. 5, the number of misclassified points in the class 3 (Gravel), the class 8 (Self-Blocking Bricks) of MLDE is significantly less than other methods, which further illustrates that the results in Table 3 are indeed believable.

Table 3. Classification accuracy (%) for the University of Pavia dataset

*	PCA	LDA	LPP	LDE	LFDA	LGDA	SGDA	SLGDA	GDA-SS	MLDE
1	89.78	89.93	87.42	94.03	95.32	92.02	95.81	89.61	93.55	94.81
2	94.69	95.55	98.24	96.60	97.21	96.21	97.62	95.60	96.16	97.61
3	23.96	43.83	54.12	72.61	74.75	62.32	70.13	33.40	66.89	76.56
4	77.97	85.87	78.49	88.25	96.83	91.68	95.82	88.97	92.69	94.32
5	98.66	99.78	99.26	99.63	97.17	99.85	99.85	99.78	99.78	99.55
6	59.00	72.06	32.95	83.69	92.05	78.29	90.59	75.66	80.41	90.91
7	68.05	30.45	62.63	74.29	72.48	64.06	82.63	38.65	72.26	83.83
8	89.33	76.91	88.10	90.55	87.64	81.10	87.72	84.27	85.55	91.53
9	100.00	80.89	99.79	99.79	97.47	99.47	99.79	99.47	99.79	99.89
OA	84.01	84.87	83.39	91.86	93.59	89.35	93.84	86.27	90.75	94.28
AA	77.94	75.03	77.89	88.83	90.10	85.00	91.11	78.38	87.45	92.11
κ	78.38	79.64	77.14	89.15	91.51	85.78	91.82	81.65	87.70	92.41

Table 4. Classification accuracy (%) for the salinas dataset

*	PCA	LDA	LPP	LDE	LFDA	LGDA	SGDA	SLGDA	GDA-SS	MLDE
1	97.31	99.15	97.90	96.91	99.30	98.80	99.60	99.65	98.70	97.71
2	96.91	99.00	99.70	97.42	94.60	90.31	99.06	99.35	99.06	96.70
3	86.13	96.00	91.29	84.91	99.39	69.88	84.15	90.73	75.35	95.29
4	98.78	98.49	99.06	98.70	99.56	99.42	98.78	96.84	93.68	97.13
5	92.15	91.97	94.24	94.36	95.07	92.49	98.69	96.82	98.65	96.00
6	98.91	99.44	99.19	98.40	99.11	98.56	98.68	98.86	99.36	97.17
7	99.13	99.46	99.10	99.49	98.91	99.16	99.63	99.38	99.74	97.59
8	87.33	86.09	89.82	88.53	84.38	85.53	83.09	78.14	77.51	83.10
9	96.16	99.01	99.06	97.64	98.85	95.38	99.90	99.50	99.90	95.87
10	79.01	89.38	85.96	79.22	80.81	82.45	64.88	63.51	93.77	86.12
11	72.00	93.25	82.86	88.01	94.66	89.88	94.10	91.10	90.91	87.27
12	96.83	77.58	98.02	97.66	61.13	84.12	100.00	99.63	97.71	99.01
13	98.58	97.27	98.03	98.25	99.01	99.12	99.01	99.45	99.45	93.77
14	88.69	90.56	89.71	90.46	86.72	81.58	91.21	82.61	88.41	84.01
15	40.05	55.71	44.12	347.8	63.34	63.01	63.47	70.58	73.34	61.64
16	88.43	95.84	83.50	97.39	86.22	91.09	98.50	97.67	98.61	99.89
OA	84.93	88.34	87.38	87.15	87.49	86.63	88.21	87.86	89.50	92.14
AA	88.53	91.77	90.73	90.95	90.07	88.80	92.05	91.49	92.77	91.77
κ	83.13	86.99	85.88	85.63	86.06	85.10	86.86	86.48	88.32	86.84

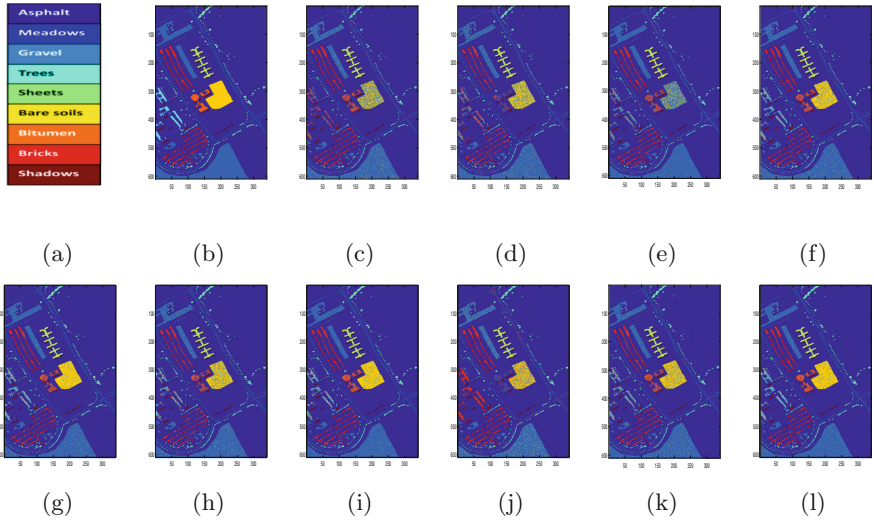


Fig. 5. Classification maps of different methods for the University of Pavia dataset: (a) legend (b) ground truth; (c) PCA: 84.01%; (d) LDA: 84.87%; (e) LPP: 83.39%; (f) LDE: 91.86%; (g) LFDA: 93.59%; (h) LGDA: 89.35%; (i) SGDA: 93.84%; (j) SLGDA: 86.27%; (k) GDA-SS: 90.75% (l) MLDE: 94.28%

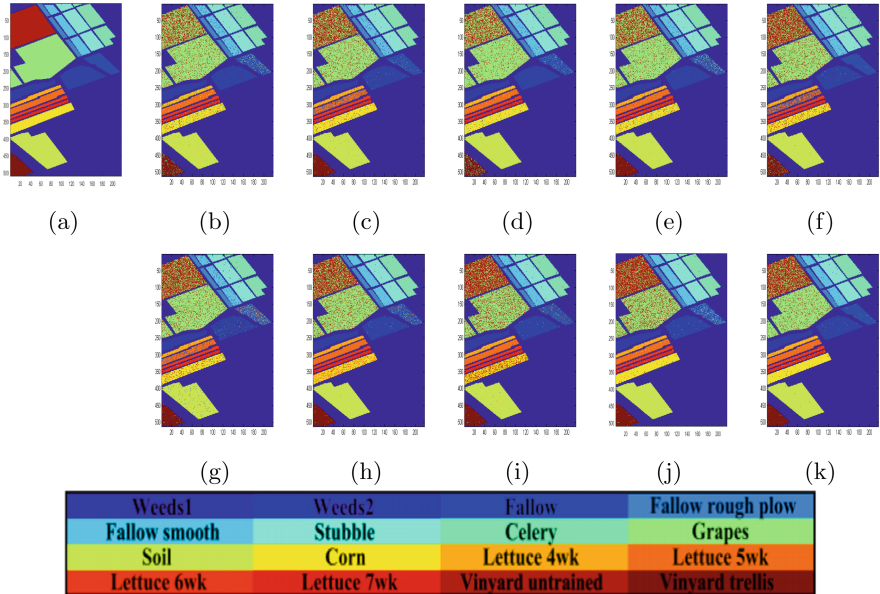


Fig. 6. Classification maps of different methods for the salinas dataset: (a) ground truth; (b) PCA: 84.93%; (c) LDA: 88.34%; (d) LPP: 87.38%; (e) LDE: 87.15%; (f) LFDA: 87.49%; (g) LGDA: 86.63%; (h) SGDA: 88.21%; (i) SLGDA: 87.86%; (k) GDA-SS: 89.50% (l) MLDE: 92.14%

Figure 6 illustrates the classification maps resulting from the classification of those methods in the salinas dataset. In Fig. 6, the number of misclassified points in the class 16 (Vinyard-vertical-trellis) is significantly less than other methods.

5 Conclusion

In this paper, we proposed a MLDE algorithm for HSI by constructing neighborhood graphs on a new spectral feature space instead of the original space. We use variance to characterize the pixels similarity of the same class and use covariance to characterize the separation of different classes of pixels. The combination of variance and covariance enables pixels in the same class to be closer and enables greater separation of pixels from different classes, which enhances classification performance of HSI. The way of representing data by using variance and covariance can attenuate the effects of noise, which can better handle with noise in HSI. Considering the symmetric positive definite nature of covariance lying on a Riemannian manifold, the MLDE algorithm using the Log-Euclidean metric to capture the similarity between spectral vectors, which can provide a more accurate similarity evaluation than euclidean distance and can better express the characteristics of spectral information. The experimental results of two hyperspectral datasets demonstrate the effectiveness of our proposed MLDE method.

Acknowledgement. We would like to thank Prof. Wei Li for sharing the codes of LFDA, LGDA, SGDA, and SLGDA. We would also like to thank Changming Jia for offering the code of GDA-SS.

References

1. Lianru, G., Bin, Y., Qian, D., et al.: Adjusted spectral matched filter for target detection in hyperspectral imagery. *Remote Sensing* **7**(6), 6611–6634 (2015)
2. Zhang, L., Zhang, L., Tao, D., et al.: Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE Trans. Geosci. Remote Sens.* **52**(8), 4955–4965 (2014)
3. Onoyama, H., Ryu, C., Suguri, M., et al.: Integrate growing temperature to estimate the nitrogen content of rice plants at the heading stage using hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* **7**(6), 2506–2515 (2014)
4. Cheng, G., Zhu, F., Xiang, S., et al.: Semisupervised hyperspectral image classification via discriminant analysis and robust regression. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* **9**(2), 595–608 (2017)
5. Jolliffe, I.T.: Principal component analysis. *J. Mark. Res.* **87**(100), 513 (2002)
6. Bandos, T.V., Bruzzone, L., Camps-Valls, G.: Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **47**(3), 862–873 (2009)
7. Zhang, X., Liang, Y., Cahill, N.: Using superpixels to improve the efficiency of Laplacian Eigenmap based methods for target detection in hyperspectral imagery. In: *Geoscience & Remote Sensing Symposium, IEEE* (2016)

8. Wang, M., Yu, J., Niu, L., et al.: Unsupervised feature extraction for hyperspectral images using combined low rank representation and locally linear embedding. In: IEEE ICASSP 2017–2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) - New Orleans, LA, USA, 5 March 2017–9 March 2017, pp. 1428–1431 (2017)
9. Zhang, M., Jia, P., Shen, Y., et al.: Hyperspectral image classification method based on orthogonal NMF and LPP. In: Instrumentation & Measurement Technology Conference, IEEE (2016)
10. Sugiyama, M.: Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **8**(1), 1027–1061 (2007)
11. Li, W., Prasad, S., Fowler, J.E., et al.: Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **50**(4), 1185–1198 (2012)
12. Chen, H.T., Chang, H.W., Liu, T.L.: Local discriminant embedding and its variants. In: IEEE Computer Society Conference on Computer Vision & Pattern Recognition (2005)
13. Yan, S., Xu, D., Zhang, B., et al.: Graph embedding: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(1), 40 (2007)
14. Cheng, B., Yang, J., Yan, S., et al.: Learning with 1-graph for image analysis. *IEEE Trans. Image Process.* **19**(4), 858–866 (2010)
15. Ly, N.H., Du, Q., Fowler, J.E.: Sparse graph-based discriminant analysis for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **52**(7), 3872–3884 (2014)
16. Li, W., Liu, J., Du, Q.: Sparse and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **54**(7), 1–12 (2016)
17. Fubiao, F., Wei, L., Qian, D., et al.: Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sensing* **9**(4), 323 (2017)
18. Deng, Y.J., Li, H.C., Pan, L., et al.: Modified tensor locality preserving projection for dimensionality reduction of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **15**(2), 277–281 (2018)
19. Yang, X., Tu, S., Bai, Y., et al.: Fusion of intensity/coherent information using region covariance features for unsupervised classification of SAR imagery. In: Geoscience & Remote Sensing Symposium, IEEE (2016)
20. Yang, J., Xing, C., Chen, Y.: Improving the ScSPM model with Log-Euclidean Covariance matrix for scene classification. In: 2016 International Conference on Computer on Information and Telecommunication Systems (CITS), IEEE (2016)