Spectral and Spatial Discriminant Analysis for Fused Feature Extraction in Hyperspectral Images

M. Imani[1], H. Ghassemian[2]*

Received:2015/11/25 Accepted:2015/12/25

Abstract

The fusion of valuable spectral and spatial features can significantly improve the performance of high resolution hyperspectral images classification. In this paper, we propose a spectral and spatial feature extraction method based on discriminant analysis. To increase the class discrimination, we maximize the between-class scatters and minimize the within-class scatters. To include the spatial information in the feature extraction process, we estimate the spatial scatters in a spatial neighborhood window with multi-scale fashion. We compare our proposed method, which is called spectral-spatial discriminant analysis (SSDA), with some spatial-spectral feature extraction methods included original spectral bands plus Gabor filters, gray level co-occurance matrix (GLCM), and morphology profiles and also with some popular spectral feature extraction methods such as nonparametric weighted feature extraction (NWFE) and locality preserving projection (LPP). Moreover, we compare SSDA with some recently proposed spectral-spatial classification approaches. The experimental results on two real hyperspectral images show the good performance of SSDA compared to the competitor methods.

Keywords: Spatial features, spectral features, discriminant analysis, classification.

Introduction

Hyperspectral remote sensing images are popular due to their high spectral resolution and are widely used for land cover discrimination. The design of a competitive supervised classification algorithm, which assigns one class label to each pixel of image, after some training procedures, is one of the most important tasks in the analysis of hyperspectral images. The acquisition of spectral information in a contiguous fashion provides a high capacity for land cover class discrimination. However, in the context of supervised classification, the high dimensionality of hyperspectral data introduces challenges such as curse of dimensionality or Hughes phenomenon [1]. To tackle this problem, several approaches have been introduced. The use of semi-supervised learning [2]-[3], the kernel based methods [4], and feature reduction [5-12] are the main solutions. Feature reduction can be implemented in two general ways: feature selection and feature extraction. Linear discriminant analysis (LDA) [13], and generalized discriminant analysis (GDA) [14], the nonlinear version of LDA, use the Fisher criterion for class discrimination. Both of the LDA and GDA methods can extract maximum $c-1$ features where c is the number of classes. Moreover, LDA has singularity problem in a small sample size situation. The nonparametric weighted feature extraction (NWFE) [15], which is the weighted version of LDA, introduces the non-parametric versions of between-class and within-class scatter matrices to extract more than $c-1$ features and copes with the singularity problem using a regularization method. In addition to LDA-based methods, there are manifold learning based approaches such as neighborhood preserving embedding (NPE) [16] and locality preserving projection (LPP) [17] that the main goal of them is preservation of data structure. Unsupervised NPE and LPP represent the topological structure of data using an adjacency graph without considering the class label information while in the supervised NPE and LPP only samples within the same class are considered during the graph construction.

The thematic maps obtained by spectral features often introduce a salt and pepper noise due to the lack of consideration of contextual information. So, with incorporation of neighborhood information (spatial features) with spectral information, the classification quality can be significantly improved. The spatial features can be included in the classification in some different ways. The contextual information can be incorporated in the classification procedure using object extraction [18] or by segmentation-aided classification [19]. In [20], a classification framework has been developed that pursues the combination of multiple features to integrate different types of features extracted from both linear and nonlinear transformations. This classification framework is able to cope with the linear and nonlinear class boundaries present in data. This approach requires no regularization parameters to control the weights of considered features. As a result, the different types of features can be efficiently integrated in a flexible and collaborative way. In [20], the posterior class probabilities are modeled by a multinomial logistic regression (MLR) and the used input feature is obtained by integration of multiple features. The original spectral information, the extended multi-attribute profiles (EMAPs), and two kernel features constructed over two previously mentioned sources of information are the linear and

1. Postdoctoral researcher, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran, maryam.imani@modares.ac.ir.

2 . Professor, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran, ghassemi@modares.ac.ir.

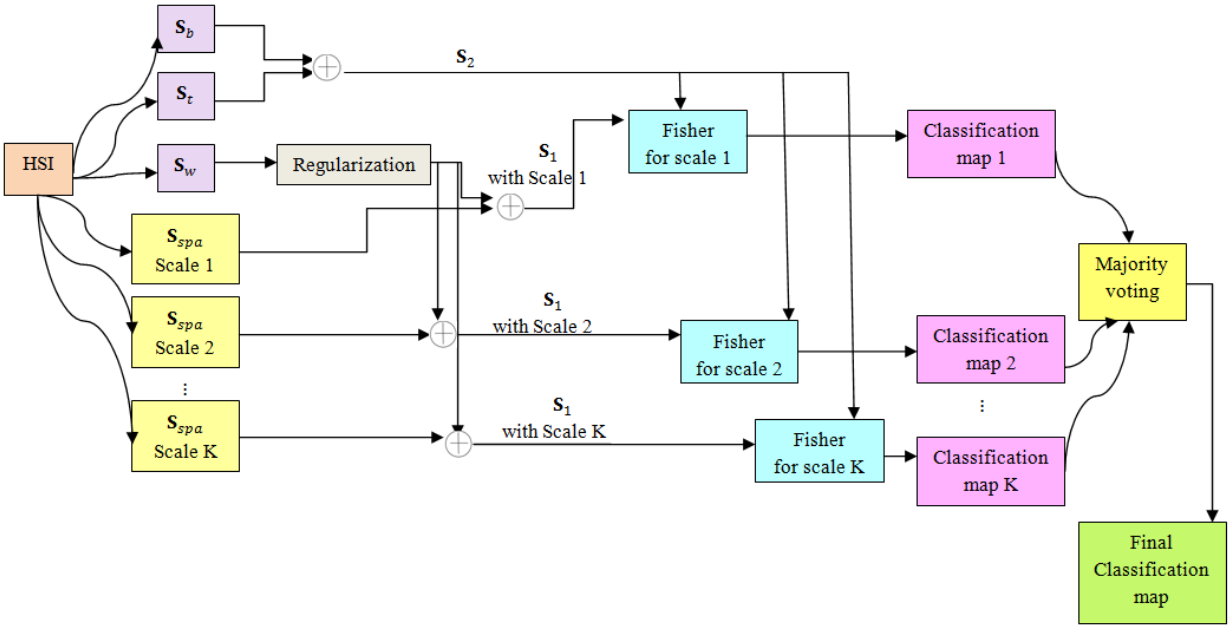

Fig. 1. The flowchart of proposed method.

nonlinear features which used as input in the MLR classifier. In [21], a region kernel-based support vector machine (RKSVM) based on spatial homogeneous regions and kernel similarity metric has been proposed to classify the local homogeneous region of each pixel of hyperspectral image. A key point in the region classification is to measure the distance (or similarity) between different regions. In detail, for each pixel $\boldsymbol{x}$, a corresponding local region $\boldsymbol{R}$ containing $\boldsymbol{x}$ and its spatial neighbors is generated by using a distance similarity neighborhood or an area-filtering-based morphological neighborhood. Authors in [22] proposed a spectral–spatial classifier for hyperspectral image classification that addresses the issue of mixed pixel characterization. In [22], the spectral information is characterized both locally and globally for probabilistic classification. A subspace-based MLR (MLR*sub*) method is used to learn the posterior probabilities and a pixel-based probabilistic support vector machine (SVM) is used as an indicator to locally determine the number of mixed components, which participate in each pixel. Then, the information provided by local and global probabilities is fused and interpreted to characterize the mixed pixels. Finally, by using a Markov random field (MRF) regularizer, the spatial information is included. The resulting method is called SVM-MLR*sub*-MRF.

Moreover, the spatial features such as texture, shape, and size can be extracted using popular methods such as Gabor filters [23]-[24], gray level co-occurance matrix (GLCM) [25], and morphology profile (MP) [26]-[29]. Then, the extracted spatial features can be fused with spectral features to improve the classification results. GLCM is a square matrix and can reveal certain properties about the spatial distribution of gray levels in the texture image by considering the relationship between two neighbouring pixels in the image. A set of filter bank consisting of Gabor filters with various scales and directions can acquire localization properties in both of the spatial and frequency domains. The MPs model the spatial information by analyzing an interaction of a set of structure elements which have different shapes and sizes.

Due to this fact that adjacent pixels in a spatial window have the same materials with a high probability, we estimate the spatial scatter matrix to combine it with within-class and between-class scatter matrices using a discriminant analysis approach. In this way, the spectral and spatial information are combined and the extracted features have more ability for class discrimination. The proposed method, called spectral-spatial discriminant analysis (SSDA), is compared with NWFE, LPP, and some spatial feature extraction methods such as Gabor filters, GLCM, and MP combined with spectral bands.

Spectral-spatial discriminant analysis

The flowchart of proposed method, SSDA, is shown in Fig. 1. The within-class and between-class scatter matrices are calculated by spectral features using a non-parametric form. The total scatter matrix is also estimated. To include the spatial information in the feature extraction procedure, we calculate the spatial scatter matrix in a neighborhood window with multiple scales. The spectral and spatial discriminant

information is then incorporated using Fisher Criterion. Finally, using a majority voting rule, the classification results of individual scales are combined to provide the final classification map. In the following, each step of SSDA is described in detail.

To extract the spectral information for class discrimination, we calculate the within-class and between-class scatter matrices, which are denoted by $\mathbf{S}_w$ and $\mathbf{S}_b$ respectively, as follows:

$$\mathbf{S}_w = \sum_{j=1}^{n} \sum_{\substack{i=1 \\ l_i=l_j}}^{n} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \quad (1)$$

$$\mathbf{S}_b = \sum_{j=1}^{n} \sum_{\substack{i=1 \\ l_i\neq l_j}}^{n} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \quad (2)$$

where $\boldsymbol{x}_i \in \mathcal{R}^d$ is the ith training sample and $l_i \in \{1, 2, \dots, c\}$ denotes the class label of $\boldsymbol{x}_i$. n and c are the number of total training samples and the number of classes respectively. We use the above non-parametric forms for calculation of scatter matrices for two reasons:1- to cope with the singularity of within-class scatter matrix in a small sample size situation, 2-the rank of between-class scatter matrix is not limited to the number of classes in the non-parametric form, and so, we can extract more than $c-1$ features. In general, the neighboring pixels in a spatial homogeneous region have the same materials and belong to the same class. In SSDA, we want to preserve the spatial local neighborhood structure. In other words, the neighbor pixels in the original spatial space remain neighbor in the SSDA projected space and vice versa. Let (p_i, q_i) be the pixel coordinate of sample $\boldsymbol{x}_i$. The local pixel neighborhood centered at $\boldsymbol{x}_i$ is defined as:

$$N(\boldsymbol{x}_i) = \{\boldsymbol{x} \triangleq (p,q) \in [p_i - s, p_i + s], q \in qi\text{–}s, qi\text{+}s \quad (3)$$

where $L = 2s + 1$ is the width of the neighborhood window and is also called scale. The number of scales is denoted by K in Fig. 1. The neighbors of $\boldsymbol{x}_i$ in the spatial neighborhood,$N(\boldsymbol{x}_i)$, are denoted by $\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots, \boldsymbol{x}_{ip}$ where $p = L^2 - 1$ is the number of neighbors. The spatial scatter matrix $(\mathbf{S}_{spa})$ is calculated as follows:

$$\mathbf{S}_{spa} = \sum_{i=1}^{n} \sum_{j=1}^{p} \beta_{ij} (\boldsymbol{x}_i - \boldsymbol{x}_{ij})(\boldsymbol{x}_i - \boldsymbol{x}_{ij})^T \quad (4)$$

where $\boldsymbol{x}_{ij} \in N(\boldsymbol{x}_i)$ is the jth spatial neighbor pixel of $\boldsymbol{x}_i$ and β_{ij} is defined as:

$$\beta_{ij} = \frac{d^{-1}(\boldsymbol{x}_i, \boldsymbol{x}_{ij})}{\sum_{k=1}^{p} d^{-1}(\boldsymbol{x}_i, \boldsymbol{x}_{ik})} \quad (5)$$

where $d(\boldsymbol{a}, \boldsymbol{b})$ is the Euclidean distance between $\boldsymbol{a}$ and $\boldsymbol{b}$. So, β_{ij} is a normalized similarity measure which measures the spectral distance of central pixel to neighboring pixels. If there is less distance between the central pixel $(\boldsymbol{x}_i)$ and jth spatial neighbor of $\boldsymbol{x}_i$, $(\boldsymbol{x}_{ij})$, the central pixel and the neighbor pixel have more similarity with together, and so, they consist of the same materials and belong to the same class with a high probability. Therefore, the neighbor pixels with more β_{ij} have more role to calculate the spatial scatter matrix and to preserve the spatial neighborhood structure of image. The total scatter matrix is estimated as:

$$\mathbf{S}_t = \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{M})(\boldsymbol{x}_i - \boldsymbol{M})^T \quad (6)$$

where $\boldsymbol{M} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$ is the mean of training samples. To cope with the singularity problem, when the number of training samples is limited, in addition to considering a non-parametric form for calculation of $\mathbf{S}_w$, we regularize it as follows:

$$\mathbf{S}_w^r = 0.5\mathbf{S}_w + 0.5 diag(\mathbf{S}_w) \quad (7)$$

After estimation of $\mathbf{S}_b$, $\mathbf{S}_w^r$, $\mathbf{S}_t$, and $\mathbf{S}_{spa}$, we combine the scatter matrices as follows to fuse the spectral and spatial information:

$$\mathbf{S}_1 = \mathbf{S}_w^r + \mathbf{S}_{spa} \quad (8)$$

$$\mathbf{S}_2 = \mathbf{S}_b + \mathbf{S}_t \quad (9)$$

To increase the class discrimination, we minimize the within-class scatters and maximize the between-class scatters and to preserve the local spatial neighborhood, we minimize the spatial scatters. In other words, for extraction of m features, SSDA seeks a linear projection matrix $\boldsymbol{A} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m]$ which can be obtained by solving the following generalized eigenvalue problem:

$$\mathbf{S}_2 \boldsymbol{v} = \lambda \mathbf{S}_1 \boldsymbol{v} \quad (10)$$

So, the projected feature vector is obtained as: $\boldsymbol{y}_{m\times 1} = \boldsymbol{A}_{m\times d} \boldsymbol{x}_{d\times 1}$.

Experimental results

The first dataset in our experiment (University of Pavia) was provided by the Reflective Optics System Imaging Spectrometer (ROSIS). It has a spatial resolution of 1.3 m per pixel. There is 115 spectral bands in the original recorded image with a spectral range from 0.43 to 0.86 μm, which after removal of noisy channels, 103 bands are selected. This urban image contains nine classes of interest and 610×340 pixels. The second dataset used in our experiments was collected over the valley of Salinas, Southern California, in 1998 by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The Salinas image consists of 224 spectral bands from 0.4 to 2.5 μm, with nominal spectral resolution of 10 nm and with a pixel size of 3.7 m which after removing 20

absorption channels, 204 bands are selected. This image contains 512 ×217 pixels and 16 classes.

We use the SVM classifier for classification of projected feature vectors. The polynomial kernel with default parameters defined in LIBSVM is used in our experiments [30]. The training samples are selected randomly and each experiment is repeated 10 times and the average of results are reported.

Some different criteria are used for evaluation of feature extraction methods from a classification accuracy point of view: average accuracy (AA), average reliability (AR), overall accuracy (OA), and kappa coefficient [31]. The accuracy (Acc.) and reliability (Rel.) for each class of data are calculated by:

$$Acc. = \alpha / A \quad (11)$$

$$Rel. = \alpha / B \quad (12)$$

where α denotes the number of correctly classified testing samples, A and B are the total testing samples of class and the total samples which are labeled as that class respectively. The number of testing samples whose labels are correctly determined divided to the total number of testing samples (in percentage) gives the overall accuracy. Moreover, for assessment of statistical significance of differences in the classification results, we use the McNemars test [32]. In this test, the sign of parameter Z_{12} indicates whether classifier 1 outperforms classifier 2 ($Z_{12} > 0$) or vice versa ($Z_{12} < 0$), and the difference between classification accuracy of two methods is statistically significant if $|Z_{12}| > 1.96$. In the SSDA method, we use the multi-scale spatial neighborhood windows with $L = 3 \times 3, 5 \times 5, \dots, 21 \times 21$. For implementation of Gabor filters, GLCM, and MP, at first, the principal component analysis (PCA) transform is applied to hyperspectral image, and then, the spatial features are extracted from the first principal component (PC1). The fast version of GLCM, with $d = 1, \theta = 0$, introduced in [33] with a 7×7 square neighbourhood window is used. 16 features introduced in [25] are extracted from the GLCM matrices.

We choose the number of scales and directions in the Gabor filters equal to 6 and 3 respectively. With applying the opening and closing operators by reconstruction on a single band image, I, a MP with $N_{mp} = 2n + 1$ features is constructed as follows:

$$MP_n(I) = \{\varphi_1(I), \dots, \varphi_n(I), I, \gamma_1(I), \dots, \gamma_n(I)\} \quad (13)$$

where $\varphi_i(I)$ and $\gamma_i(I)$; $(i = 1,2, \dots, n)$ are closing and opening operators by reconstruction respectively. A disk shape structure element with the radius $R \in \{1,2, \dots, n\}$ is used to extract a MP with N_{mp}=53 features. The 6×3=18 Gabor, 16 GLCM, and 53 MP features are extracted and combined with d original full band hyperspectral image (HS) into a vector using stacking. The feature vectors of Gabor-HS, GLCM-HS, and MP-HS have 18+d, 16+d, 53+d spectral-spatial features respectively. For extraction of m features from these long feature vectors, the PCA transform is used.

In addition to spectral-spatial methods, i.e., Gabor-HS, GLCM-HS, and MP-HS, we compare our proposed method with two popular spectral feature extraction methods, NWFE and supervised LPP. Figs. 2 and 3 show the average accuracy versus the number of extracted features using a) 15 training samples and b) 30 training samples for Pavia and Salinas datasets respectively. The accuracy and reliability for each class, average accuracy, average reliability, overall accuracy, and kappa coefficient obtained by 5 and 8 extracted features for Pavia and Salinas datasets are represented in tables 1 and 2 respectively. For University of Pavia using 15 training samples, the use of HS provides the following results: 75.53% average accuracy, 68.54% average reliability, 60.02% overall accuracy, and 0.52 kappa coefficient. Also, for Salinas dataset using 15 training samples, the following results are achieved by HS: 88.83% average accuracy, 85.25% average reliability, 80.48% overall accuracy, and 0.78 kappa coefficient. The ground truth map (GTM), and the classification maps obtained by 15 training samples are shown in Figs. 4 and 5 for Pavia and Salinas datasets respectively.

Tables 3 and 4 represent the classification results using 30 training samples for Pavia (with 9 extracted features) and Salinas (with 7 extracted features) respectively. The McNemars test results are reported in table 5. As we can see from the results, SSDA outperforms other methods in the most cases. This is expected because SSDA extract spectral and spatial features simultaneously while in Gabor-HS, GLCM-HS, and MP-HS, at first the spatial features are extracted, and then combined with original spectral bands. After that, a transformation such as PCA is needed to reduce the dimensionality of stacked feature vector in Gabor-HS, GLCM-HS, and MP-HS. Also, NWFE and LPP feature extraction methods just use the spectral features and do not consider the correlation of adjacent pixels in an image and so, a salt and pepper noise is presented in the obtained classification maps.

The highest average accuracies among the first 12 features for University of Pavia and Salinas datasets are represented in table 6. Each number in the parentheses is the optimal number of extracted features. The highest values in each row are shown in bold. As can be seen, the highest AA is obtained by SSDA in both datasets. In addition, we compared our proposed method with some classification approaches [20-[22], which use the both of spectral and spatial information. The classification results for University of Pavia are reported. Authors in [20] use a linear feature $\boldsymbol{h}_{linear}$ (the original spectral information), a nonlinear feature $\boldsymbol{h}_{EMAP}$ (Extended multi-attribute profiles, i.e., EMAP), and also two kernel features constructed over two previously mentioned sources of information (spectral and spatial,

respectively) using the Gaussian radial-basis-function kernel. The spectral kernel $\boldsymbol{K}_{linear}$ and the spatial kernel $\boldsymbol{K}_{EMAP}$ are built by using the original spectral data and the EMAP respectively. All nonlinear features considered in $\boldsymbol{h}_{all} = [\boldsymbol{h}_{linear}, \boldsymbol{h}_{EMAP}, \boldsymbol{K}_{linear}, \boldsymbol{K}_{EMAP}]$. A suitable subset of features, including only $\boldsymbol{h}_{linear}$ and $\boldsymbol{h}_{EMAP}$, i.e., $\boldsymbol{h}_{subset} = [\boldsymbol{h}_{linear}, \boldsymbol{h}_{EMAP}]$ can be obtained to have a comparable solution with very competitive computational cost. The comparison results of proposed method (SSDA) with types of features considered in [20] are represented in table 7. The number of features extracted by SSDA, which achieved the highest OA, is shown in the parentheses. In this table, totally 3921 samples are used for training and 42776 samples are used for testing. The OA and computation time are reported in this table. SSDA provides reasonable overall classification accuracy compared to classification approach introduced in [20]. SSDA is implemented faster than $\boldsymbol{K}_{linear}$, $\boldsymbol{K}_{EMAP}$, and $\boldsymbol{h}_{all}$.

In [21], RKSVM using a squared neighborhood (SN) and a morphological neighborhood (MN) is called RKSVM-SN and RKSVM-MN, respectively. K_S^r and K_M^r denote RKSVM-SN and RKSVM-MN with the single region kernel respectively. μK_S^{r+w} and μK_M^{r+w} represent RKSVM-SN and RKSVM-MN with the weighted summation region kernel respectively, where the combination coefficient μ is set as 0.8. $K_S^{[r,\omega]}$ and $K_M^{[r,\omega]}$ refer to RKSVM-SN and RKSVM-MN with the stack region kernel respectively. All definition of single region kernel, weighted summation region kernel, and stack region kernel can be found in [21]. The comparison results between SSDA and proposed methods in [21] using 5 and 20 training samples are reported in table 8. When there is a very small training set (5 samples per class), SSDA outperforms K_M^r and μK_M^{r+w}. But, using 20 training samples, which is approximately a small training set, RKSVM provides better performance than SSDA.

The OA results obtained by SSDA and SVM-MLR*sub*-MRF (proposed in [22]) are given in table 9 for different number of training samples. In the last column of this table, 3921samples are used for training, and 42776 samples are used for testing. Using a large training set, the scatter matrices in SSDA can be estimated with a high accuracy. The classification results show the better performance of SSDA compared to SVM-MLR*sub*-MRF using a large training set.

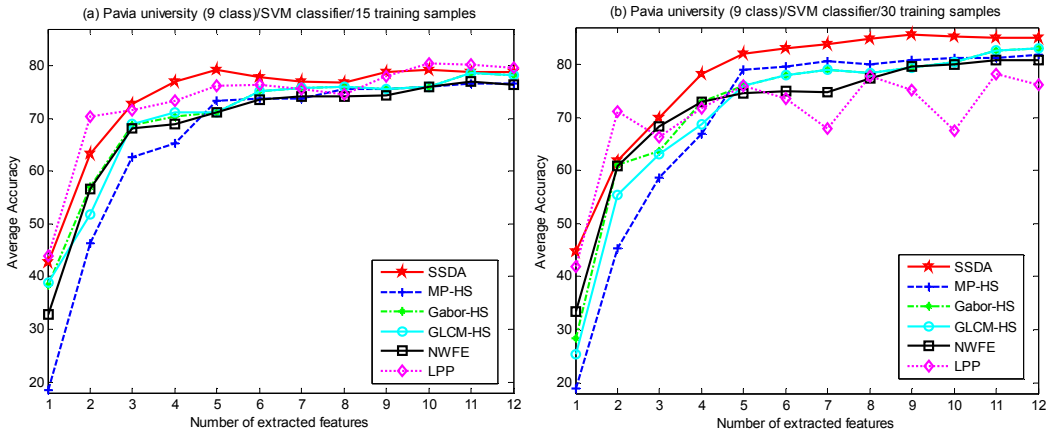

Fig. 2. The average classification accuracy versus the number of extracted features for Pavia dataset using a) 15 training samples, and b) 30 training samples.

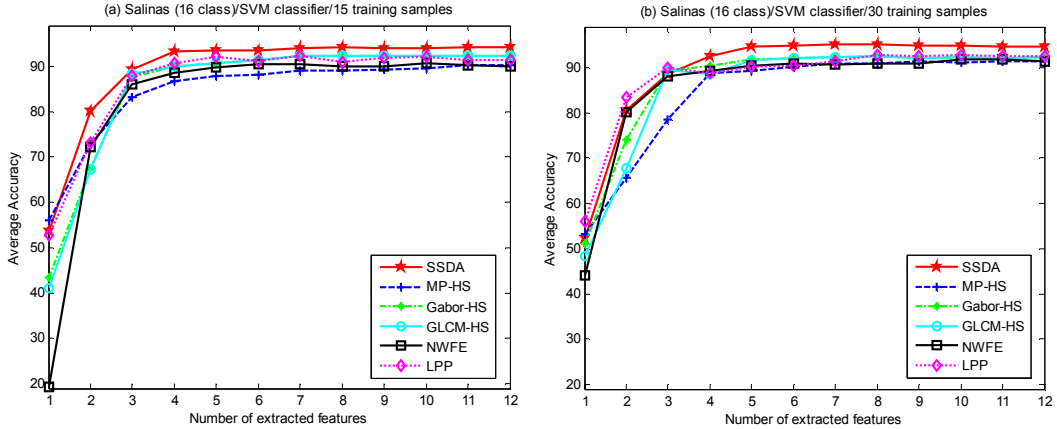

Fig. 3. The average classification accuracy versus the number of extracted features for Salinas dataset using a) 15 training samples, and b) 30 training samples.

Table 1. The classification results obtained by 15 training samples and 5 extracted feature for Pavia dataset.

class			SSDA		MP-HS		Gabor-HS		GLCM-HS		NWFE		LPP	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Asphalt	6631	78.54	78.90	58.66	67.07	60.46	82.29	60.44	82.35	60.40	79.04	64.95	85.76
2	Meadows	18649	72.93	92.97	74.68	95.15	64.78	85.58	64.78	85.58	62.19	84.54	53.88	86.37
3	Gravel	2099	75.56	48.91	53.36	29.27	35.64	25.21	35.97	25.43	54.69	28.43	62.46	39.88
4	Trees	3064	82.54	76.40	77.42	59.20	75.46	85.00	75.46	85.00	74.15	93.85	82.87	75.88
5	Painted metal sheets	1345	99.18	99.93	95.61	93.12	98.66	98.01	98.66	98.01	98.81	97.01	99.33	96.53
6	Bare Soil	5029	79.04	49.53	54.03	56.88	63.89	35.92	63.87	35.94	60.27	30.52	65.68	31.05
7	Bitumen	1330	75.34	48.90	86.54	35.76	87.97	36.76	87.97	36.76	77.89	41.86	86.24	44.72
8	Self-Blocking Bricks	3682	50.05	69.84	62.17	54.45	53.61	53.70	53.69	53.61	52.93	69.48	70.15	65.33
9	Shadows	947	99.58	99.79	97.57	99.89	99.68	100.00	99.68	100.00	99.79	99.47	99.89	100.00
AA and AR			**79.20**	**73.91**	**73.34**	**65.64**	**71.13**	**66.94**	**71.17**	**66.96**	**71.24**	**69.36**	**76.16**	**69.50**
OA			**74.84**		**69.36**		**64.92**		**64.94**		**63.83**		**64.30**	
Kappa coefficient			**0.68**		**0.61**		**0.56**		**0.56**		**0.55**		**0.56**	

Table 2. The classification results obtained by15 training samples and 8 extracted feature for Salinas dataset.

class			SSDA		MP-HS		Gabor-HS		GLCM-HS		NWFE		LPP	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Brocoli_green_weeds_1	2009	99.70	99.45	100.00	99.41	99.35	99.75	99.35	99.75	98.26	89.89	99.60	99.11
2	Brocoli_green_weeds_2	3726	99.60	99.97	92.51	97.65	99.41	99.52	99.41	99.52	92.40	98.85	99.28	99.73
3	Fallow	1976	99.54	98.20	99.80	93.33	98.13	82.51	98.13	82.51	95.45	89.30	99.54	90.48
4	Fallow_rough_plow	1394	99.00	98.57	98.85	97.59	98.42	97.44	98.42	97.44	99.43	96.12	99.07	95.50
5	Fallow_smooth	2678	98.58	98.73	94.77	99.22	95.03	98.91	95.03	98.91	92.83	98.77	95.52	99.15
6	Stubble	3959	99.12	100.00	94.62	93.42	99.12	99.14	99.12	99.14	99.62	99.60	99.37	99.34
7	Celery	3579	99.69	99.97	94.52	92.10	99.13	99.33	99.13	99.33	98.91	95.16	99.30	98.56
8	Grapes_untrained	11271	66.40	82.40	53.15	81.72	59.45	82.54	59.47	82.55	59.28	73.26	37.64	62.56
9	Soil_vineyard_develop	6203	99.92	97.33	99.74	94.85	97.68	96.51	97.68	96.51	99.15	97.31	98.69	97.70
10	Corn_senesced_green_weeds	3278	83.89	91.64	72.15	65.59	74.92	88.60	74.92	88.60	73.76	87.36	80.60	95.83
11	Lettuce_romaine_4weeks	1068	92.60	94.37	96.72	91.82	94.01	81.83	94.01	81.83	89.61	89.61	95.60	83.28
12	Lettuce_romaine_5 weeks	1927	99.90	88.38	99.27	97.90	99.22	92.77	99.22	92.77	98.03	82.17	98.91	98.50
13	Lettuce_romaine_6 weeks	916	98.14	95.23	95.09	57.38	99.02	84.06	99.02	84.06	98.03	83.23	98.91	90.42
14	Lettuce_romaine_7 weeks	1070	91.68	96.46	94.11	67.40	89.35	95.41	89.35	95.41	90.65	88.10	90.09	91.03
15	Vineyard_untrained	7268	79.64	61.64	68.73	61.83	80.94	58.46	80.94	58.47	65.64	51.75	67.83	41.77
16	Vineyard_vertical_trellis	1807	99.83	100.00	71.56	40.62	93.86	87.20	93.86	87.20	89.37	93.73	95.96	96.01
AA and AR			**94.20**	**93.90**	**89.10**	**83.24**	**92.32**	**90.25**	**92.32**	**90.25**	**90.03**	**88.39**	**90.99**	**89.94**
OA			**88.67**		**81.49**		**86.07**		**86.07**		**83.12**		**80.43**	
Kappa coefficient			**0.87**		**0.80**		**0.85**		**0.85**		**0.81**		**0.78**	

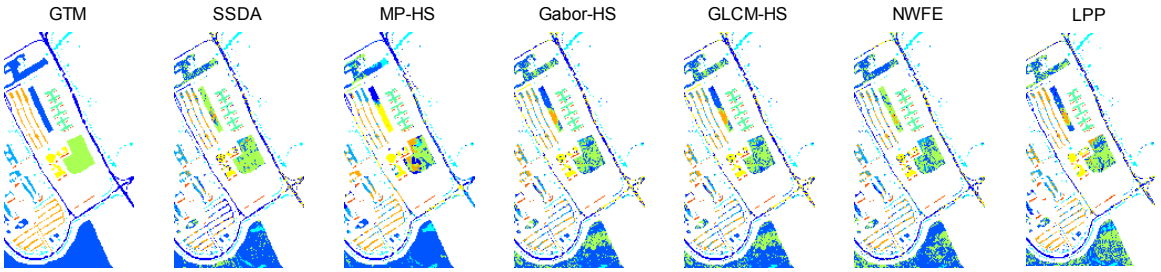

Fig. 4. GTM and classification maps obtained by 15 training samples and 5 extracted features for Pavia dataset.

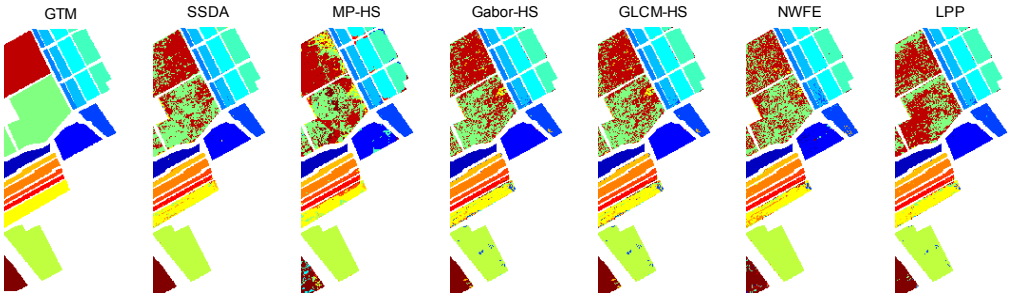

Fig. 5. GTM and classification maps obtained by 15 training samples and 8 extracted features for Salinas dataset.

Table 3. The classification results obtained by 30 training samples and 9 extracted feature for Pavia dataset.

class			SSDA		MP-HS		Gabor-HS		GLCM-HS		NWFE		LPP	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Asphalt	6631	73.96	88.12	81.66	90.07	82.05	87.28	82.05	87.28	81.45	84.76	73.67	87.29
2	Meadows	18649	87.54	97.00	68.80	86.88	65.89	85.83	65.89	85.83	61.72	85.93	48.45	80.76
3	Gravel	2099	87.61	50.91	74.89	61.41	74.32	57.31	74.32	57.31	71.22	61.50	73.03	52.59
4	Trees	3064	78.88	74.55	78.49	80.33	79.90	73.78	79.86	73.77	77.68	79.23	75.59	63.84
5	Painted metal sheets	1345	97.17	91.53	99.85	99.85	99.63	83.65	99.63	83.65	99.78	96.34	99.70	88.51
6	Bare Soil	5029	78.25	71.56	68.68	39.69	66.18	37.56	66.18	37.56	67.99	34.10	63.39	27.40
7	Bitumen	1330	92.71	64.79	87.52	58.46	84.29	59.56	84.29	59.56	88.27	59.47	82.33	46.92
8	Self-Blocking Bricks	3682	76.02	74.96	67.52	71.89	63.88	81.13	63.88	81.13	68.44	77.90	60.67	74.00
9	Shadows	947	99.16	97.61	99.89	100.00	99.68	100.00	99.68	100.00	99.47	99.47	99.89	100.00
AA and AR			**85.70**	**79.00**	**80.81**	**76.51**	**79.53**	**74.01**	**79.53**	**74.01**	**79.56**	**75.41**	**75.19**	**69.03**
OA			**83.44**		**73.86**		**72.01**		**72.00**		**70.50**		**62.04**	
Kappa coefficient			**0.79**		**0.67**		**0.65**		**0.65**		**0.63**		**0.54**	

Table 4. The classification results obtained by 30 training samples and 7 extracted feature for Salinas dataset.

class			SSDA		MP-HS		Gabor-HS		GLCM-HS		NWFE		LPP	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Brocoli_green_weeds_1	2009	99.50	100.00	100.00	99.95	99.30	99.90	99.30	99.90	98.31	95.04	99.30	99.01
2	Brocoli_green_weeds_2	3726	99.19	99.73	94.77	100.00	98.31	99.38	98.31	99.38	94.31	98.54	98.39	99.35
3	Fallow	1976	99.80	95.54	98.68	95.68	96.20	94.96	96.20	94.96	93.22	87.34	96.71	96.27
4	Fallow_rough_plow	1394	99.35	99.00	93.26	96.87	99.00	99.50	99.00	99.50	99.57	94.74	99.35	99.28
5	Fallow_smooth	2678	97.72	99.39	95.97	96.47	97.16	99.20	97.16	99.20	93.43	96.98	96.90	99.35
6	Stubble	3959	99.82	100.00	98.23	98.08	99.34	99.47	99.34	99.47	99.55	99.90	99.72	99.45
7	Celery	3579	99.69	99.69	96.93	92.85	98.80	97.33	98.80	97.33	99.22	95.51	99.02	99.11
8	Grapes_untrained	11271	74.70	80.79	74.49	78.07	58.03	77.32	58.02	77.30	61.80	76.73	47.70	68.91
9	Soil_vineyard_develop	6203	99.36	98.89	96.36	99.15	94.39	98.44	94.39	98.44	98.31	99.19	93.71	98.31
10	Corn_senesced_green_weeds	3278	92.34	87.11	85.91	85.23	88.53	86.09	88.53	86.09	79.84	86.34	89.08	89.43
11	Lettuce_romaine_4weeks	1068	98.78	97.24	97.85	81.39	94.29	70.27	94.29	70.27	93.35	78.07	94.38	63.44
12	Lettuce_romaine_5 weeks	1927	100.00	98.42	99.22	97.01	97.56	97.66	97.56	97.66	95.74	96.65	94.76	98.65
13	Lettuce_romaine_6 weeks	916	99.56	95.30	97.49	60.54	99.02	76.41	99.02	76.41	98.25	81.45	99.02	91.34
14	Lettuce_romaine_7 weeks	1070	93.27	88.71	88.04	78.30	89.81	84.00	89.81	84.00	85.14	82.89	90.09	82.04
15	Vineyard_untrained	7268	71.99	67.61	64.89	67.78	72.74	55.04	72.73	55.04	72.43	56.58	67.09	46.43
16	Vineyard_vertical_trellis	1807	98.23	98.34	72.55	70.26	96.07	97.75	96.07	97.75	86.88	91.33	97.29	97.88
AA and AR			**95.21**	**94.11**	**90.91**	**87.35**	**92.41**	**89.54**	**92.41**	**89.54**	**90.58**	**88.58**	**91.41**	**89.27**
OA			**89.92**		**86.28**		**85.07**		**85.07**		**84.71**		**82.11**	
Kappa coefficient			**0.89**		**0.85**		**0.84**		**0.84**		**0.83**		**0.80**	

Table 5. The McNemars test results for Pavia and Salinas datasets.

Pavia, 15 training samples, 5 extracted features						
	SSDA	MP-HS	Gabor-HS	GLCM-HS	NWFE	LPP
SSDA	0	21.96	36.31	36.26	42.44	38.97
MP-HS	-21.96	0	15.53	15.47	19.42	17.12
Gabor-HS	-36.31	-15.53	0	-1.02	4.32	2.39
GLCM-HS	-36.26	-15.47	1.02	0	4.39	2.45
NWFE	-42.44	-19.42	-4.32	-4.39	0	-1.82
LPP	-38.97	-17.12	-2.39	-2.45	1.82	0

Salinas, 15 training samples, 8 extracted features						
	SSDA	MP-HS	Gabor-HS	GLCM-HS	NWFE	LPP
SSDA	0	37.29	18.27	18.25	35.20	42.55
MP-HS	-37.29	0	-24.50	-24.52	-7.96	5.31
Gabor-HS	-18.27	24.50	0	-1.41	17.40	29.72
GLCM-HS	-18.25	24.52	1.41	0	17.42	29.74
NWFE	-35.20	7.96	-17.40	-17.42	0	13.96
LPP	-42.55	-5.31	-29.72	-29.74	-13.96	0

Pavia, 30 training samples, 9 extracted features						
	SSDA	MP-HS	Gabor-HS	GLCM-HS	NWFE	LPP
SSDA	0	15.91	9.15	9.16	17.40	43.80
MP-HS	-15.91	0	42.22	42.22	46.82	72.09
Gabor-HS	-9.15	-42.22	0	1.00	6.94	36.21
GLCM-HS	-9.16	-42.22	-1.00	0	6.93	36.20
NWFE	-17.40	-46.82	-6.94	-6.93	0	32.74
LPP	-43.80	-72.09	-36.21	-36.20	-32.74	0

Salinas, 30 training samples, 7 extracted features						
	SSDA	MP-HS	Gabor-HS	GLCM-HS	NWFE	LPP
SSDA	0	21.28	32.44	32.46	32.94	45.90
MP-HS	-21.28	0	6.87	6.89	8.39	22.86
Gabor-HS	-32.44	-6.87	0	0.37	2.30	18.91
GLCM-HS	-32.46	-6.89	-0.37	0	2.28	18.87
NWFE	-32.94	-8.39	-2.30	-2.28	0	13.69
LPP	-45.90	-22.86	-18.91	-18.87	-13.69	0

Table 6. The highest average accuracy among the first 12 features.

Dataset	No. of training	SSDA	MP-HS	Gabor-HS	GLCM-HS	NWFE	LPP
University of Pavia	15	**79.20 (5)**	76.65 (12)	78.63 (11)	78.62 (11)	76.94 (11)	80.43 (10)
	30	**85.70 (9)**	81.86 (12)	83.02 (12)	83.02 (12)	80.92 (11)	78.18 (11)
Salinas	15	**94.20 (8)**	90.31 (12)	92.43 (11)	92.44 (12)	90.45 (7)	92.28 (7)
	30	**95.21 (7)**	91.47 (11)	92.55 (8)	92.55 (8)	91.86 (11)	92.87 (8)

Table 7. The comparison results between SSDA and proposed method in [20] using totally 3921 training samples for University of Pavia dataset.

	SSDA	Proposed frame work in [20]				
		$\boldsymbol{h}_{EMAP}$	$\boldsymbol{K}_{linear}$	$\boldsymbol{K}_{EMAP}$	$\boldsymbol{h}_{all}$	$\boldsymbol{h}_{subset}$
OA	97.61 (22 features)	97.37	79.50	97.43	97.80	97.53
Time (seconds)	93.42	3.56	156.08	166.50	2082.3	5.00

Table 8. The comparison results between SSDA and proposed method in [21] using different number of training samples per class for University of Pavia dataset.

No. of training	SSDA	RKSVM-SN proposed in [21]			RKSVM-MN proposed in [21]		
		K_S^r	μK_S^{r+w}	$K_S^{[r,\omega]}$	K_M^r	μK_M^{r+w}	$K_M^{[r,\omega]}$
5	71.83 (3 features)	72.66	72.95	74.62	71.42	71.52	73.37
20	83.11 (8 features)	92.67	92.94	93.09	91.81	91.86	92.86

Table 9. The comparison results between SSDA and proposed method in [22], as a function of the number of training samples per class where the total number of training samples is given in parentheses, for University of Pavia dataset.

No. of training	20 (180)	40 (360)	60 (540)	80 (720)	All samples (3921)
SSDA	83.11 (8 features)	89.72 (12 features)	92.53 (13 features)	95.22 (17 features)	97.61 (22 features)
SVM-MLR*sub*-MRF [22]	85.88	90.97	92.10	94.57	95.56

Conclusion

The spectral-spatial discriminant analysis method was proposed in this paper for feature extraction of hyperspectral images. The SSDA method simultaneously maximizes the class discrimination and preserves the spatial local structure of data in the projected space using the Fisher criterion.

The experimental results showed the good performance of SSDA in comparison with Gabor-HS, GLCM-HS, MP-HS, NWFE, LPP, and some recently proposed spectral-spatial classification approaches.

References

[1] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[2] G. Camps-Valls, T. V. Bandos, and D. Zhou, "Semi-supervised Graph-based Hyperspectral Image Classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3044 - 3054, Oct. 2007.

[3] Z. Feng, S. Yang, S. Wang, and L. Jiao, "Discriminative Spectral Spatial Margin-Based Semisupervised Dimensionality Reduction of Hyperspectral Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 224–228, Feb. 2015.

[4] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative Multiple Kernel Learning for Classification in Hyperspectral Imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2852–2865, July 2012.

[5] M. Imani and H. Ghassemian, Two Dimensional Linear Discriminant Analysis for Hyperspectral Data, *Journal of Photogrammetric Engineering & Remote Sensing (PE&RS)*, vol. 81, no. 10, pp. 777-786, Oct. 2015.

[6] Y. Yuan, G. Zhu, and Q. Wang, "Hyperspectral Band Selection by Multitask Sparsity Pursuit," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 631–644, Feb. 2015.

[7] J. Xia, J. Chanussot, P. Du, and X. He, "Spectral–Spatial Classification for Hyperspectral Data Using Rotation Forests With Local Feature Extraction and Markov Random Fields," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2532–2546, May 2015.

[8] M. Imani, H. Ghassemian, Feature Extraction Using Median-Mean and Feature Line Embedding, International Journal of Remote Sensing, vol. 36, no. 17, pp. 4297-4314, 2015.

[9] J.Yan, N. Liu, S. Yan, Q. Yang, W. (P.) Fan, W. Wei, and Z. Chen, "Trace-Oriented Feature Analysis for Large Scale Text Data Dimension Reduction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 7, pp. 1103–1117, July 2011.

[10] J. Qian, J. Yang, and Y. Xu, "Local Structure-Based Image Decomposition for Feature Extraction With Applications to Face Recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3591–3603, Sept. 2013.

[11] M. Imani, H. Ghassemian, Ridge regression-based feature extraction for hyperspectral data, International Journal of Remote Sensing, vol. 36, no. 6, pp. 1728–1742, 2015.

[12] M. Imani, H. Ghassemian, Feature reduction of hyperspectral images: discriminant analysis and the first principal component, Journal of AI and Data Mining, vol. 3, no. 1, pp.1-9, 2015.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, San Diego: Academic Press Inc., 1990.

[14] G. Baudat and F. Anouar, ''Generalized discriminant analysis using a kernel approach,'' *Neural Comput*., vol. 12, pp. 2385–2404, 2000.

[15] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens*, vol. 42, no. 5, pp. 1096-1105, May 2004.

[16] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in Proc. 10th IEEE Int. Conf. Comput. Vis., vol. 2, pp. 1208–1213, 2005.

[17] X. F. He and P. Niyogi, "Locality preserving projections," in Proc. Adv. Neural Inf. Process. Syst., vol. 16, pp. 153–160, 2004.

[18] H. Ghassemian and D. A. Landgrebe, "Object-Oriented Feature Extraction Method for Image Data Compaction," *IEEE Control Systems Mogorine*, vol. 8, no. 3, pp. 42–48, 1988.

[19] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral image using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[20] J. Li, X. Huang, P. Gamba, J. M. B. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple Feature Learning for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3 pp. 1592-1606, March 2015.

[21] P. Jiangtao, Z. Yicong, and C. L. P. Chen, "Region-Kernel-Based Support Vector Machines for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 4810- 4824, Sept. 2015.

[22] M. Khodadadzadeh, J. Li, A. Plaza, H. Ghassemian, J. M. Bioucas-Dias, and X. Li, "Spectral-Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6298- 6314, Oct. 2014.

[23] T.P. Weldon, W.E. Higgins, and D.F. Dunn, "Efficient Gabor filter design for texture segmentation," PREPRINT of Pattern recognition, 1996.

[24] W. Li and Q. Du, "Gabor-Filtering-Based Nearest Regularized Subspace for Hyperspectral Image Classification," *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing*, vol. 7, no. 4, pp. 1012–1022, April 2014.

[25] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *IEEE*

Transactions on Systems, Man, and Cybernetics, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[26] F. Mirzapour and H. Ghassemian, "Improving hyperspectral image classification by combining spectral, texture, and shape features," *International Journal of Remote Sensing*, vol. 36, no. 4, pp. 1070-1096, April 2015.

[27] J. Atli Benediktsson, J. Aevar Palmason, and J. R. Sveinsson, "Classification of Hyperspectral Data From Urban Areas Based on Extended Morphological Profiles,"*IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, March 2005.

[28] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological Attribute Profiles for the Analysis of Very High Resolution Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[29] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, Nov. 2010.

[30] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp: 27:1–27:27, 2011.

[31] J. Cohen, A coefficient of agreement from nominal scales, *Edu. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46,1960.

[32] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.

[33] F. Mirzapour and H. Ghassemian, "Using GLCM and Gabor filters for classification of PAN images," 21st Iranian Conference on Electrical Engineering (ICEE 2013), Mashhad, Iran, May 2013.