IJICTR
International Journal of Information & Communication Technology Research

Volume 10- Number 1 – Winter 2018 (1-12)

Spectral-Spatial Feature Transformations With Controlling Contextual Information Through Smoothing Filtering and Morphological Analysis

Maryam Imani*
Faculty of Electrical and Computer Engineering
Tarbiat Modares University
Tehran, Iran
maryam.imani@modares.ac.ir

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

Received: 20 September, 2017 - Accepted: 25 December, 2017

Abstract— **A fusion method for spectral-spatial classification of hyperspectral images is proposed in this paper. In the proposed framework, at first, the dimension of hyperspectral image is reduced by several state-of-the-art spectral feature extraction methods, i.e., Binary Coding Based Feature Extraction (BCFE), Clustering Based Feature Extraction (CBFE), Feature Extraction Based on Ridge Regression (FERR), Feature Extraction Using Attraction Points (FEUAP), Feature Extraction using Weighted Training samples (FEWT), and Feature Space Discriminant Analysis (FSDA). Then, the spatial features are calculated from the spectral features extracted from each spectral feature extraction method individually using the proposed smoothing filters and morphological operators. Finally, majority voting decision rule is used to obtain the final classification map. The proposed framework, in addition to removing the useless spatial information such as noise and distortions, adds useful spatial information such as shape and size of objects presented in scene image. The use of complement information obtained from six spectral feature extraction methods with different ideas for class discrimination, significantly improves the classification results. The proposed framework provides in average 6.64%, 7.07%, 8.23%, 7.52% and 20.52% improvement in classification results of three real hyperspectral images compared to generalized composite kernel (GCK), multiple feature learning (MFL), weighted joint collaborative representation (WJCR), original hyperspectral bands stacked on extended morphological profile (HS+EMP) and original hyperspectral bands (HS), respectively in terms of overall accuracy.**

Keywords- ***spectral-spatial features; feature transformation; classification; majority voting; hyperspectral data.***

I. INTRODUCTION

High spectral dimensionality of hyperspectral images allows accurate classification of different land cover types. Supervised classifiers such as neural networks [1], Bayesian [2] and kernel-based methods [3]-[4] have provided good performance in terms of classification accuracy. By increasing the data

* Corresponding Author

dimensionality with a fixed number of training samples, the classification accuracy is first increased to a point and after that with increasing the data dimensionality, the classification accuracy is decreased. This is known as Hughes phenomenon [5]. Due to high dimension of hyperspectral images and because of limited number of available training samples, dealing with Hughes phenomenon is one of the main challenges of hyperspectral image classification. The high computational burden is another difficulty of using high spectral dimensionality. To solve these difficulties, feature reduction methods have been suggested [6]-[11].

In addition to spectral information, spatial characteristics have also been shown to be very useful for hyperspectral image classification [12]-[13]. Kernel-based methods such as support vector machines (SVMs) have been widely utilized because of their insensitivity to the curse of dimensionality. Composite kernels with integrating the spectral and spatial information provide significant improvement in hyperspectral image classification [14]. Standard composite kernels and also multiple kernel learning methods generally need convex combinations of kernels [15]. Moreover, their parameters optimization is difficult. The generalized composite kernel (GCK) method has been proposed to overcome these limitations [16]. Multiple kernels can be linearly combined without any restriction of convexity in GCK. Combination of the spectral and spatial information is done by GCK without any weight parameter.

Different spectral-spatial features have been used for hyperspectral image classification. These features are divided into two main groups: linear features and nonlinear ones. On the one hand, some methodologies such as maximum noise fraction [17] and independent component analysis [18] use the linear features extracted from the original spectral information. On the other hand, nonlinear features can be more effective for class discrimination in some cases. Some examples of the nonlinear transformations based techniques for modelling the inherent nonlinearity of data are kernel methods, manifold regularization ones [19], and extended multi-attribute profile [20]. Since generally both linear and nonlinear class boundaries exist in the scene, multiple feature learning (MFL) [21] has been introduced to integrate multiple linear and nonlinear features. The multinomial logistic regression (MLR) [22], with its flexibility in construction of nonlinear kernels, is exploited in both GCK and MFL methods.

The nearest subspace classification is coupled with a distance-weighted Tikhonov regularization in the nearest regularized subspace (NRS) classifier [23]. In NRS, each testing sample is represented via a linear combination of training pixels within each class. The label of class that best approximates the test pixel is assigned to it. However, NRS just exploits the spectral features and ignores the spatial characteristics at neighboring locations. The joint collaborative representation (JCR) method has been introduced in [24] to overcome the indigenous disadvantage of the NRS classifier. JCR uses a joint collaborative model of training samples. Therefore, JCR involves the contextual information in classification. The weighted JCR (WJCR), which is an improved version of JCR [25] uses more efficient collaborative representation with considering the similarity between the center pixel and its neighbors.

Morphological profiles (MPs) are efficient and popular tools for spatial feature extraction [26]-[27]. MP produces a multi-scale decomposition from a single band using opening and closing operators. A MP concatenates a closing profile and an opening profile. The extended morphological profile (EMP) is the generalization of MP for hyperspectral data [28].

A spectral-spatial classification method is proposed in this paper that is an extended version of work presented in IST 2016 symposium [40]. The proposed method removes the useless spatial characteristics and adds useful spatial ones to improve the class discrimination. In [40], the full hyperspectral cube is divided to some sub-cubes containing the adjacent bands. But, in this work, different feature extraction methods containing complementary information are used instead of them. At first, the high dimension of hyperspectral image is reduced using some state-of-the-art spectral feature extraction methods: Binary Coding Based Feature Extraction (BCFE) [6], Clustering Based Feature Extraction (CBFE) [29], Feature Extraction Based on Ridge Regression (FERR) [30], Feature Extraction Using Attraction Points (FEUAP) [31], Feature Extraction using Weighted Training samples (FEWT) [8], and Feature Space Discriminant Analysis (FSDA) [11]. These feature extraction methods are different approaches which use different ideas to produce appropriate spectral features for classification aims. This step reduces the data dimensionality and degrades the Hughes phenomenon. In each subgroup, we apply the proposed smoothing filter, which removes the spectral-spatial distortions such as noise from data. The morphological filters are applied to the smoothed images to add useful contextual information to them. Then, the spectral-spatial features in each subgroup are given to an appropriate classifier. The final classification map is obtained by majority voting (MV) rule from decisions made from all subgroups.

The extracted spectral features from the BCFE, CBFE, FERR, FEUAP, FEWT, and FSDA methods, contain complementary information and have

minimum overlap with respect to each other. So, the aggregation of decisions obtained from each group of features through the decision fusion using MV rule achieves high classification accuracy. The experimental results on three real hyperspectral images show the good performance of the proposed method compared to some state-of-the-art spectral-spatial classification methods such as GCK, MFL, WJCR, and integration of hyperspectral (HS) with EMP, which is denoted by HS+EMP.

The reminder of this paper is continued as follows. The feature extraction methods are represented in section II. The proposed method is introduced in section III. The experimental results are discussed in section IV. Finally, section V concludes the paper.

II. FEATURE EXTRACTION METHODS

A) BCFE

The BCFE method extracts m features from d original spectral bands by decomposition of the whole spectral signature of each pixel to m segments. Then, it calculates the weighted mean of spectral bands in each segment and considers it as a new extracted feature. Two characterizations of binary coded class means is used to obtain a weight for each band. The first characteristic is the binary values of class means in each band and the second characteristic is the positive and negative edges of class means in each band. The BCFE calculates just a simple weighted mean. So, it is faster than supervised feature extraction methods such as linear discriminant analysis (LDA) [32], nonparametric weighted feature extraction (NWFE) [33], generalized discriminant analysis (GDA) [34], and median-mean line based discriminant analysis (MMLDA) [35] that need to calculate the scatter matrices. Moreover, BCFE achieves more classification accuracy than other feature extraction methods using limited training samples because other mentioned methods need to estimate the second-order statistics (scatter matrices) while BCFE just needs to estimate the first-order statistics, i.e., class means. Let $\boldsymbol{x} = [x_1 \quad x_2 \cdots \quad x_d]^T$ represents a pixel of hyperspectral image where d is the number of spectral bands and $\boldsymbol{y} = [y_1 \quad y_2 \cdots \quad y_m]^T$ denotes the extracted feature vector. For extraction of m features from d spectral bands, the spectral signature of each pixel is partitioned to m segments containing K bands where $K = \left\lfloor \frac{d}{m} \right\rfloor$. In each part, the weighted mean of bands is considered as a new extracted feature as follows:

$$y_j = \sum_{i=(j-1)K+1}^{jK} w_i x_i \ , \ 1 \le j \le m-1 \qquad (1)$$

$$y_m = \begin{cases} \sum_{i=(m-1)K+1}^{d} w_i x_i \ ; \ mK < d \ \left(m < \frac{d}{K}\right) \\ \sum_{i=(m-1)K+1}^{mK} w_i x_i \ ; \ mK = d \ \left(m = \frac{d}{K}\right) \end{cases} \qquad (2)$$

The weight in each band is calculated by:

$$w_i = \alpha(w_1)_i + (1-\alpha)(w_2)_i \ ; \ (1 \le i \le d) \qquad (3)$$

where $0 \le \alpha \le 1$ is a positive real-valued free parameter, which is tuned in the training process, and constitutes a tradeoff between the information contained in the values of bands (related to w_1) and the information contained in the edges of bands (related to w_2). To understand how to calculate weights w_1 and w_2, the authors refer the readers to [6].

B) CBFE

The CBFE method considers a vector associated with each spectral band which contains the mean values of training samples of all classes in that band. Then, a clustering algorithm such as k-means groups the vectors in some clusters. The mean of spectral bands whose associated vectors are located in the same cluster is considered as an extracted feature. CBFE just calculates the first order statistics of training samples, i.e., mean vectors and thus works well in the small sample size situations. In the CBFE method, the mean vector of training samples is calculated in each class and the matrix $\mathbf{A}$ is composed using the mean values as follows:

$$\mathbf{A} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1c} \\ m_{21} & m_{22} & \cdots & m_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ m_{d1} & m_{d2} & \cdots & m_{dc} \end{bmatrix} \qquad (4)$$

where m_{ij} is the mean of training samples of jth class in the ith band. The matrix $\mathbf{A}$ can be rewritten using the row vectors $\mathbf{a}_i$ $(i = 1, 2, \ldots, d)$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_d \end{bmatrix} \qquad (5)$$

where $\mathbf{a}_i$ $(i = 1,2,..,d)$ is the vector corresponding to ith dimension and contains the mean values of training samples for c classes. If two vector $\mathbf{a}_i$ and $\mathbf{a}_j$ $(i \ne j)$ become similar, then bands i and j are highly correlated. So, one of them can be removed. Based on this idea, the vectors $\mathbf{a}_i$ $(i = 1,2,..,d)$ are grouped into some clusters using a clustering algorithm such as k-means. The mean of spectral bands whose associated vectors are located in a cluster is considered as an extracted feature. The number of clusters determines the number of extracted features (12 features are extracted using CBFE in this work).

C) FERR

In the FERR method, a feature vector is defined for each dimension, and then, modeled as a linear combination of its farthest neighbors. By solving the ridge regression model, the representation coefficients are calculated and considers as the entries of the projection matrix. FERR can extract each number of features and has good efficiency in the small sample size situations. By Assuming $\boldsymbol{x}_{d\times 1}$ as the original feature vector of each pixel of hyperspectral image, the aim is extraction of m features from it using a linear transformation such as $\boldsymbol{y}_{m\times 1} = \mathbf{A}_{m\times d}\boldsymbol{x}_{d\times 1}$ $(m \ll d)$. The matrix of class mean is represented as follows where c and d are the number of classes and the number of spectral bands, respectively:

$$\begin{bmatrix} m_{11}\, m_{12}\, m_{13} \cdots m_{1c} \\ m_{21}\, m_{22}\, m_{23} \cdots m_{2c} \\ m_{31}\, m_{32}\, m_{33} \cdots m_{3c} \\ \vdots \quad \vdots \quad \vdots \quad \cdots \quad \vdots \\ m_{d1} m_{d2}\, m_{d3} \cdots m_{dc} \end{bmatrix} \tag{6}$$

where m_{ij} $(i = 1, 2, \dots, d; j = 1, 2, \dots, c)$ is the mean of class j in ith band. Feature vectors in d bands are defined as follows:

$$\boldsymbol{h}_i = [m_{i1} \quad m_{i2} \quad \cdots \quad m_{ic}]^T ; \quad i = 1, 2, \dots, d \tag{7}$$

By considering $\boldsymbol{h}_i$ as a central feature vector: $\boldsymbol{q}_i = \boldsymbol{h}_i$, m farthest neighbors of $\boldsymbol{q}_i$ among $\boldsymbol{h}_j (j = 1,2,..,d\,; j \neq i)$ are obtained and denoted by $\boldsymbol{h}_{i1}, \boldsymbol{h}_{i2}, \dots, \boldsymbol{h}_{im}$. Then, $\boldsymbol{q}_i$ is represented using a linear combination of m farthest neighbors:

$$\boldsymbol{q}_i = w_{i1}\boldsymbol{h}_{i1} + w_{i2}\boldsymbol{h}_{i2} + \cdots + w_{im}\boldsymbol{h}_{im} + \boldsymbol{\varepsilon} \;\; ; i = 1,2, \dots, d \tag{8}$$

where $w_{i1}, w_{i2}, \dots, w_{im}$ are the representation coefficients. The other form of (8) is given by:

$$(\boldsymbol{q}_i)_{c\times 1} = (\mathbf{H}_i)_{c\times m}(\boldsymbol{w}_i)_{m\times 1} + \boldsymbol{\varepsilon}_{c\times 1} \tag{9}$$

where

$$\mathbf{H}_i = [\boldsymbol{h}_{i1}\, \boldsymbol{h}_{i2} \dots \boldsymbol{h}_{im}] \;\; ; i = 1,2, \dots, d \tag{10}$$

$$\boldsymbol{w}_i = [w_{i1}\, w_{i2} \dots w_{im}]^T \;\; ; i = 1,2, \dots, d \tag{11}$$

To avoid the singularity problem, the ridge regression is used. The coefficient vector $\boldsymbol{w}_i$ is obtained by ordinary least square solution as follows:

$$\widehat{\boldsymbol{w}}_i = \left(\mathbf{H}_i{}^T\mathbf{H}_i + \delta\mathbf{I}\right)^{-1}\mathbf{H}_i{}^T\boldsymbol{q}_i \;\; ; i = 1,2, \dots, d \tag{12}$$

Note that the vector $\widehat{\boldsymbol{w}}_i$ describes the relationship between the ith feature vector $(\boldsymbol{q}_i = \boldsymbol{h}_i)$ and m feature vectors, $\boldsymbol{h}_{i1}, \boldsymbol{h}_{i2}, \dots, \boldsymbol{h}_{im}$, which have the largest distances from it. To obtain the projection matrix, the representation coefficients of each band is obtained in a similar way. These coefficients vector compose the projection matrix after normalization.

D) FEUAP

The basic idea of FEUAP is represented as follows: if we consider an appropriate attraction point for each class, the samples of each class can move toward the attraction point of their class using a proper transformation. If attraction points of different classes are chosen away enough from each other, different classes become separable by aggregation of samples of each class around the attraction point of the same class. The FEUAP method is done in two basic phases: 1- obtaining appropriate attraction points, 2- achieving the proper transformation to move toward attraction points. Two approaches are proposed in [31] to obtain the attraction points (selection based on distance measure and selection based on dense measure). For more details, authors refer the readers to [31]. FEUAP has no need to estimate the statistical moments (mean vector or scatter matrix) and so works well using limited training samples. In FEUAP, the samples in the reduced feature space are in such a way that: 1- Each sample has minimum distance from the attraction point of its class (attraction), 2- Each sample has maximum distance from the attraction points of other classes (repulsion). Based on this idea, two functions are defined, attraction function (ψ_1) and repulsion function (ψ_2):

$$\psi_1 = \sum_{c=1}^{n_c}\sum_{i=1}^{n_{tc}}\|\boldsymbol{y}_{ic} - \boldsymbol{y}^{ac}\|^2 \tag{13}$$

$$\psi_2 = -\sum_{c=1}^{n_c}\sum_{i=1}^{n_{tc}}\sum_{\substack{k=1 \\ k\neq c}}^{n_c}\|\boldsymbol{y}_{ic} - \boldsymbol{y}^{ak}\|^2 \tag{14}$$

where $\boldsymbol{x}^{ac}$ is the attraction point of cth class and $\boldsymbol{y}^{ac} = \mathbf{A}\boldsymbol{x}^{ac}$ is the attraction point of cth class in the new feature space, i.e., $\boldsymbol{y}_{ic} = \mathbf{A}\boldsymbol{x}_{ic}$. For finding the transformation matrix $(\mathbf{A})$, the following optimization problem is solved:

$$\min_{\mathbf{A}}(\psi = \psi_1 + \psi_2) \tag{15}$$

E) FEWT

In popular feature extraction methods, all spectral bands of each training sample have the same role in the feature extraction process. But, different spectral bands have different abilities in identification of classes. FEWT considers the relative importance of each feature (spectral band) in predicting the class label of sample as a weight for that feature. Each arbitrary feature extraction approach can use these weighted training samples. In [8], the weighted training samples are used in the supervised locality preserving projection (LPP).

Let $\boldsymbol{X}_{n\times d}=[\boldsymbol{x}_1,\boldsymbol{x}_2,\dots,\boldsymbol{x}_n]^T$ be the training samples where n is the number of training samples and $\boldsymbol{L}\in\mathcal{R}^{n\times c}$ be the class label matrix. The entries of matrix $\boldsymbol{L}$ are zero or one. If ith sample belongs to kth, then, in the ith row of $\boldsymbol{L}$ just column k is one and other columns are zero. The relationship between the training samples and the class labels can be modeled as follows:

$$\boldsymbol{L}_{n\times c}=\boldsymbol{X}_{n\times d}\boldsymbol{W}_{d\times c}+\boldsymbol{1}_n\boldsymbol{b}^T \quad (16)$$

where $\boldsymbol{b}\in\mathcal{R}^{c\times 1}$ is the bias term and $\boldsymbol{1}_n$ is a $n\times 1$ constant vector whose elements are all one. The entries of the weight matrix $\boldsymbol{W}$ are $w_{jk}(j=1,..,d\,;k=1,\dots,c)$ where w_{jk} measures the relative importance of jth feature (band) in predicting the class k. The least square with ridge regularization is used to calculate the weight matrix $\boldsymbol{W}$ as follows:

$$\min_{\boldsymbol{W},\boldsymbol{b}}\psi(\boldsymbol{W},\boldsymbol{b})=\|\boldsymbol{X}\boldsymbol{W}+\boldsymbol{1}_n\boldsymbol{b}^T-\boldsymbol{L}\|_F^2+\gamma\|\boldsymbol{W}\|_F^2 \quad (17)$$

$\gamma>0$ is a tradeoff parameter and $\|\cdot\|_F$ denotes the Frobenius norm. The matrix $\boldsymbol{W}$ is obtained using the matrix theory:

$$\boldsymbol{W}=(\boldsymbol{X}^T\boldsymbol{H}\boldsymbol{X}+\gamma\boldsymbol{I}_d)^{-1}\boldsymbol{X}^T\boldsymbol{H}\boldsymbol{L} \quad (18)$$

where $\boldsymbol{H}=\boldsymbol{I}_n-(1/n)\boldsymbol{1}_n\boldsymbol{1}_n{}^T$ and $\boldsymbol{I}_n$ is an $n\times n$ identity matrix. The matrix $\boldsymbol{W}$ is used for weighting the training samples. If $\boldsymbol{x}_i$ belongs to class k, then, the bands of $\boldsymbol{x}_i$ are weighted as follows:

$$z_{ij}=x_{ij}w_{jk}\,;1\le j\le d \quad (19)$$

where w_{jk} is the weight of jth feature in class k, x_{ij} is the jth feature of ith sample and z_{ij} is jth feature of ith weighted sample.

F) FSDA

Popular feature extraction methods such as LDA-based methods just use the class discrimination for feature extraction. In addition to separability between classes, FSDA considers the difference between spectral bands in the transformed feature space. FSDA extracts features such a way that: the extracted features are as different from each other as possible, and, separability between classes is increased. For extraction of features with minimum redundant information, FSDA at first estimates the between-spectral band scatter matrix as follows:

$$\boldsymbol{S}_{\mathrm{f}}=\sum_{j=1}^{n_t}\sum_{i=1}^{d}\left(\boldsymbol{h}_{ij}-\bar{\boldsymbol{h}}_j\right)\left(\boldsymbol{h}_{ij}-\bar{\boldsymbol{h}}_j\right)^T \quad (20)$$

where $\boldsymbol{h}_{ij}$ and $\bar{\boldsymbol{h}}_j$ are considered as follows:

$$\boldsymbol{h}_{ij}=[x_{i1j}\quad x_{i2j}\quad\cdots\quad x_{icj}]^T\,;\quad i=1,2,\dots,d,\ j=1,2,\dots,n_t \quad (21)$$

$$\bar{\boldsymbol{h}}_j=\frac{1}{d}\sum_{i=1}^{d}\boldsymbol{h}_{ij} \quad (22)$$

where the jth training sample of class k in ith feature is denoted by x_{ikj} $(i=1,2,\dots,d;\ k=1,2,\dots,c;\ j=1,2,\dots,n_t\ )$ and n_t is the number of training samples per class. With maximizing $tr(\boldsymbol{S}_{\mathrm{f}})$, the projection matrix $\boldsymbol{W}$ is calculated and then, the new feature space is obtained by:

$$\left(\boldsymbol{g}_{ij}\right)_{c\times 1}=\boldsymbol{W}_{c\times c}\left(\boldsymbol{h}_{ij}\right)_{c\times 1}\,;\quad i=1,2,\dots,d,\ j=1,2,\dots,n_t. \quad (23)$$

According to $\boldsymbol{W}$ transformation, training samples x_{ikj} are transformed to r_{ikj} where r_{ikj} is the jth training sample of class k in ith dimension of the transformed feature space. Then, the between-class scatter matrix $(\boldsymbol{S}_{\mathrm{b}})$ and within-class scatter matrix $(\boldsymbol{S}_{\mathrm{w}})$ are defined as follows:

$$\boldsymbol{S}_{\mathrm{b}}=\sum_{j=1}^{n_t}\sum_{k=1}^{c}\left(\boldsymbol{R}_{kj}-\bar{\boldsymbol{R}}\right)\left(\boldsymbol{R}_{kj}-\bar{\boldsymbol{R}}\right)^T \quad (24)$$

$$\boldsymbol{S}_{\mathrm{w}}=\sum_{k=1}^{c}\sum_{j=1}^{n_t}\sum_{i=1}^{n_t}\left(\boldsymbol{R}_{ki}-\boldsymbol{R}_{kj}\right)\left(\boldsymbol{R}_{ki}-\boldsymbol{R}_{kj}\right)^T \quad (25)$$

where

$$\boldsymbol{R}_{kj}=[r_{1kj}\quad r_{2kj}\quad\cdots\quad r_{dkj}]^T\,;\quad k=1,2,\dots,c,\ j=1,2,\dots,n_t \quad (26)$$

and

$$\bar{\boldsymbol{R}}=\frac{1}{c\times n_t}\sum_{j=1}^{n_t}\sum_{k=1}^{c}\boldsymbol{R}_{kj} \quad (27)$$

To maximize the class discrimination, $tr\left(\boldsymbol{S}_{\mathrm{w}}{}^{-1}\boldsymbol{S}_{\mathrm{b}}\right)$ is maximized. For more information, the authors refer the readers to [11].

II. PROPOSED METHOD

The proposed method is a spectral-spatial classification framework that improves the classification accuracy by controlling spatial information involved in the classification process. The flowchart of the proposed method is shown in Fig. 1. At first, the huge hyperspectral cube is given to six different feature extraction blocks for dimensionality reduction. The BCFE, CBFE, FERR, FEUAP, FEWT, and FSDA methods are used for feature extraction. These feature extraction methods use different approaches and ideas for feature transformation and so the feature spaces obtained by them contain complement information. Using this step of the

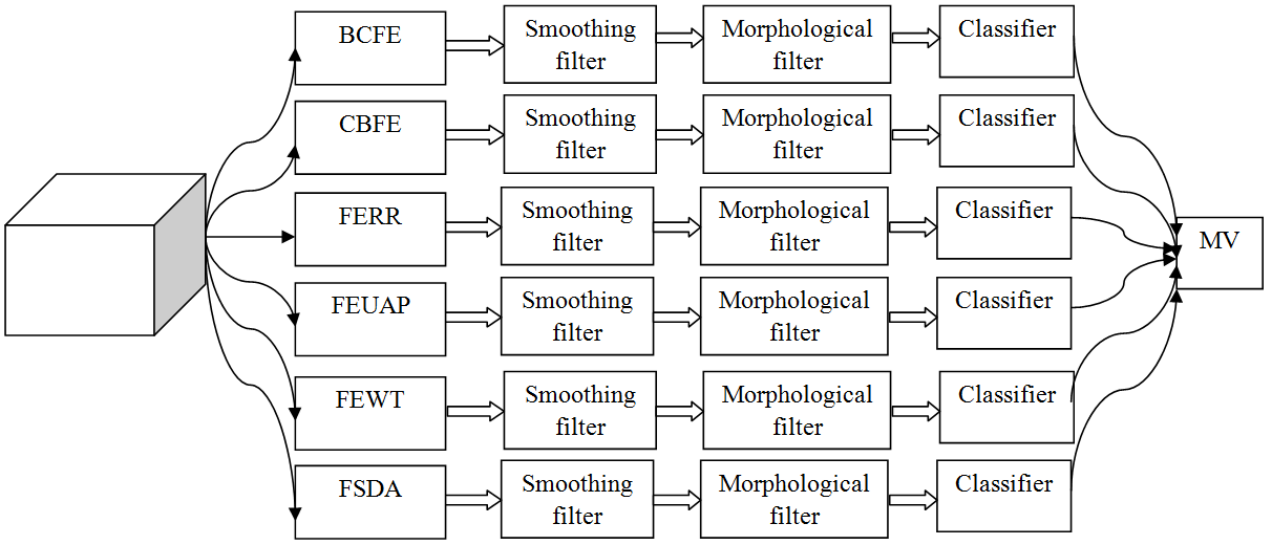

Fig.1. Block diagram of the proposed spectral-spatial classification framework.

proposed method, the dimensionality of data is reduced and the Hughes phenomenon is degraded. So, the proposed method can work well in small sample size situations. Moreover, the complement information extracted by different feature extraction methods, after spatial processing, are fused using decision rule.

Most of the spectral-spatial classification methods just add contextual features to data. But, two main contributions are used in this work. At first, the useless contextual information such as noise which causes spectral-spatial distortions is removed. Then, the useful spatial information is added to data. The proposed smoothing filters and then morphological filters are applied to each subgroup of extracted features to remove useless spatial information and add useful spatial information, respectively.

For implementation of the proposed smoothing filter, a spatial neighborhood window is considered around each central pixel where its spectral feature vector is modeled by its surrounding pixels. If $\boldsymbol{x}_i^k$ indicates i th pixel in k th subgroup, it can be represented by:

$$\boldsymbol{x}_i^k = \sum_{j\in w_i} \alpha_{ij}^k \boldsymbol{x}_j^k \qquad (28)$$

where w_i is a local window containing $(2a+1)\times(2a+1)$ pixels around pixel $\boldsymbol{x}_i^k$ where a indicates the radius of the window. $\boldsymbol{x}_j^k$ denotes jth neighbor of $\boldsymbol{x}_i^k$ in the local window and α_{ij}^k measures the similarity between central pixel $(\boldsymbol{x}_i^k)$ and its neighbor, $(\boldsymbol{x}_j^k)$, in subgroup k. The weight α_{ij}^k is defined by:

$$\alpha_{ij}^k = \frac{1}{1+dist\left(\boldsymbol{x}_i^k,\boldsymbol{x}_j^k\right)} \qquad (29)$$

where

$$dist\left(\boldsymbol{x}_i^k,\boldsymbol{x}_j^k\right) = \left(\boldsymbol{x}_i^k-\boldsymbol{x}_j^k\right)^T\left(\boldsymbol{x}_i^k-\boldsymbol{x}_j^k\right) \qquad (30)$$

is the Euclidean distance. Then, the morphological filters are applied to the smoothed images. In other words, the spatial features are added to images using EMP. The morphological filters are efficient tools for spatial feature extraction. The degree of processing of input image is determined by the geometrical characteristics of the structure element (SE). The opening profile (Π_γ) and closing profile (Π_φ) are concatenated to provide a MP:

$$\Pi_\gamma(z) = \left\{\Pi_{\gamma\lambda}: \Pi_{\gamma\lambda} = \gamma_\lambda^*(z), \forall\lambda\in[0,n]\right\} \qquad (31)$$

$$\Pi_\varphi(z) = \left\{\Pi_{\varphi\lambda}: \Pi_{\varphi\lambda} = \varphi_\lambda^*(z), \forall\lambda\in[0,n]\right\} \qquad (32)$$

where z denotes a pixel of single band image I, $\gamma_\lambda^*(z)$ and $\varphi_\lambda^*(z)$ indicate the morphological opening and closing operators by reconstruction using SE with the size of λ, respectively. A MP consists of $2n+1$ bands is provided from the single band image I by applying n opening operators and n closing operators by reconstruction:

$$MP_n(I) = \{\varphi_1^*(I),\dots,\varphi_n^*(I), I, \gamma_1^*(I),\dots,\gamma_n^*(I)\} \qquad (33)$$

To handle the hyperspectral images, EMP is used. The principal component analysis (PCA) transformation [36] as a dimension reduction method is used to reduce the hyperspectral image dimensionality. In the proposed method in this work, the PCA transformation is applied to each subgroup of extracted features to reduce the dimensionality from d_i to $(p<d_i)$. The p principal components (PCs) of data corresponding to the p largest eigenvalues of covariance matrix of each

Table I. The individual results for each of BCFE, CBFE, FERR, FEUAP, FEWT, and FSDA branches in the proposed block diagram compared to the final fused result obtained by the MV rule.

Dataset	BCFE	CBFE	FERR	FEUAP	FEWT	FSDA	MV
Indian	84.89	83.54	79.51	73.38	76.02	78.45	94.05
Pavia	95.33	94.42	92.97	89.99	90.52	89.90	98.95
Salinas	93.57	93.57	93.33	93.04	92.84	93.94	95.74

subgroup are chosen and the remained components are discarded. EMP in each subgroup is formalized as follows:

$$EMP = \{MP_n(PC_1), MP_n(PC_2), \dots MP_n(PC_p)\} \quad (34)$$

The spectral channels are integrated with the spatial features extracted by EMP in each subgroup and the spectral-spatial features are fed to an efficient classifier such as SVM. Finally the classification maps provided by different subgroups of extracted features are contributed to achieve the final classification map using the MV rule.

IV. EXPERIMENTAL RESULTS

The performance of the proposed method is evaluated in comparison with some spectral-spatial classification methods such as GCK [16], MFL [21], WJCR [25], HS+EMP (an integration of hyperspectral data with EMP), and original hyperspectral (HS). In each class, 20 training samples, which is a small training set relative to the high dimension of used datasets, are randomly selected for doing experiments and evaluation of the proposed method in small sample size situation. Three hyperspectral datasets are used in the experiments: Indian pines with agriculture/forest context and low spatial resolution of 20 m by pixel, University of Pavia with an urban context and high spatial resolution of 1.3 m per pixel, and Salinas with a spatial resolution of 3.7 m. The Indian image collected from Northwestern Indiana by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) has 145×145 pixels containing 16 classes, and 224 spectral channels where the number of channels is reduced to 200 by removing water absorption bands. Ten classes of Indian are selected for doing experiments. University of Pavia image acquired by the ROSIS instrument over the city of Pavia, Italy has 610×340 pixels with 115 spectral bands and 9 classes where 103 spectral channels are remained after discarding noisy and water absorption bands. The Salinas hyperspectral image collected over the valley of Salinas, Southern California by AVIRIS contains 512 × 217 pixels, 16 classes, and 224 spectral channels where 204 channels are remained after removing absorption bands. Several measurements are used for evaluation of classification accuracy: classification accuracy (Acc.), classification reliability (Rel.), average accuracy, average reliability, overall accuracy as the percentage of correctly classified samples and kappa coefficient [37]. In the formulas of the accuracy $Acc = N/A$ and reliability $Rel = N/B$ N denotes the number of testing samples that are correctly classified, A indicates the total testing samples of class and B denotes the total samples that are labeled as the class. Moreover, the McNemars test is used [38] to assess the statistical significance of differences in the classification results. The sign of McNemars test parameter Z_{12} indicates whether classifier 1 works more accurate than classifier 2 ($Z_{12} > 0$) or vice versa ($Z_{12} < 0$). If $|Z_{12}| > 1.96$, the difference in the classification accuracies of two classifiers is statistically significant. SVM implemented by LIBSVM [39] is used as classifier in the proposed method, HS+EMP, and HS. The polynomial kernel with default parameters in the LIBSVM is used. MLR is used as classifier in GCK and MFL (according to their definitions in [16] and [21]) and the spectral-spatial version of the nearest regularized subspace method is used as classifier in WJCR (according to its definition in [25]).

The individual results for each of BCFE, CBFE, FERR, FEUAP, FEWT, and FSDA branches in the proposed block diagram are obtained and compared with the final fused result obtained by the MV rule. The overall accuracies are reported in Table I. As seen from the results, it can be found that the fused results with MV rule are the best. It is expected because the features extracted by different ideas are fused together which provide complementary information for image classification. The classification results for Indian, Pavia, and Salinas datasets are represented in Tables II-IV respectively. The associated ground truth map (GTM) and the classification maps are also shown in Figs. 2-4. The McNemars test results are reported in Table V. The following conclusions can be seen from the obtained results:

1- The proposed classification framework achieves the highest classification accuracy.
2- In Indian dataset, MFL is superior to GCK, but in Pavia dataset, GCK is better than MFL.

Table II. The classification results for Indian dataset.

class			Proposed		GCK		MFL		WJCR		HS+EMP		HS	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Corn-no till	1434	94.21	88.94	71.97	86.50	73.57	72.76	91.21	79.13	80.96	89.58	62.27	55.88
2	Corn-min till	834	95.68	82.35	86.93	71.57	91.97	79.40	86.09	72.23	87.89	62.54	50.72	34.09
3	Grass/pasture	497	98.59	92.80	97.38	83.30	96.58	97.76	97.99	92.41	98.39	81.50	93.96	63.54
4	Grass/trees	747	98.13	96.19	97.59	95.92	97.46	95.54	88.62	99.85	93.44	97.21	86.88	89.52
5	Hay-windrowed	489	99.59	97.01	99.59	99.80	99.59	100.00	100.00	100.00	99.39	99.59	99.18	98.38
6	Soybeans-no till	968	89.88	94.46	85.74	65.05	86.47	73.94	80.99	90.01	84.92	73.52	58.16	57.16
7	Soybeans-min till	2468	88.21	98.28	73.26	88.11	73.50	88.19	63.86	91.05	73.18	89.94	33.39	57.78
8	Soybeans-clean till	614	98.05	92.33	92.35	76.93	90.55	84.63	99.84	57.24	78.50	73.59	75.57	47.54
9	Woods	1294	97.30	100.00	90.65	99.83	99.15	99.84	86.71	99.56	96.52	100.00	66.69	98.18
10	Bldg-Grass-Tree-Drives	380	99.74	95.23	98.16	82.89	96.84	84.02	96.84	61.44	99.74	90.02	81.32	46.33
Average Acc. and Average Rel.			95.94	93.76	89.36	84.99	90.57	87.61	89.21	84.29	89.29	85.75	70.81	64.84
Overall Acc.			94.05		84.40		86.12		83.57		85.40		61.08	
Kappa coefficient			0.93		0.82		0.84		0.81		0.83		0.56	

Table III. The classification results for Pavia dataset.

class			Proposed		GCK		MFL		WJCR		HS+EMP		HS	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Asphalt	6631	99.98	98.73	90.47	87.04	86.85	92.40	81.68	98.71	78.31	98.37	64.27	95.60
2	Meadows	18649	96.76	100.00	91.82	98.13	89.92	98.09	91.46	94.40	89.26	98.65	75.90	92.72
3	Gravel	2099	99.62	84.66	91.19	71.10	87.57	66.38	75.51	73.21	91.62	64.25	76.89	59.40
4	Trees	3064	91.85	93.09	91.45	86.06	83.52	76.96	97.23	66.27	96.90	88.73	92.62	67.88
5	Painted metal sheets	1345	100.00	100.00	99.41	84.57	85.35	100.00	100.00	99.12	98.14	84.83	99.63	95.85
6	Bare Soil	5029	100.00	100.00	90.50	92.16	97.41	84.79	80.77	88.63	95.63	76.22	81.13	56.16
7	Bitumen	1330	100.00	98.64	98.80	65.08	99.62	94.04	98.42	79.77	99.40	58.29	92.48	43.10
8	Self-Blocking Bricks	3682	86.83	100.00	75.20	96.35	93.54	81.65	88.67	80.54	77.24	88.74	76.15	75.95
9	Shadows	947	100.00	99.41	99.47	87.71	81.31	96.25	97.68	100.00	99.37	99.58	99.79	99.89
Average Acc. and Average Rel.			97.23	97.17	92.03	85.35	89.45	87.84	90.16	86.74	91.76	84.18	84.32	76.28
Overall Acc.			98.95		90.59		90.03		88.70		88.76		77.77	
Kappa coefficient			0.97		0.88		0.87		0.85		0.85		0.72	

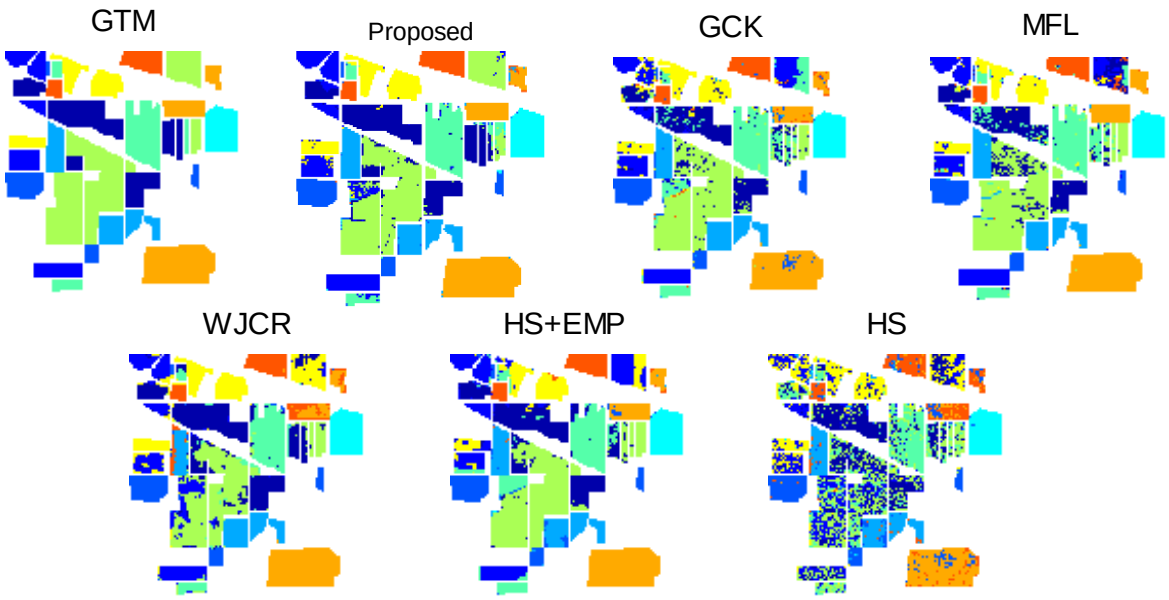

Fig. 2. The classification maps for Indian dataset.

Table III. The classification results for Salinas dataset.

class			Proposed		GCK		MFL		WJCR		HS+EMP		HS	
No	Name of class	# samples	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.	Acc.	Rel.
1	Brocoli_green_weeds_1	2009	100.00	90.50	100.00	100.00	100.00	100.00	43.06	74.19	100.00	98.38	97.11	99.90
2	Brocoli_green_weeds_2	3726	97.34	100.00	99.49	100.00	98.93	100.00	91.81	74.94	95.12	100.00	99.14	98.32
3	Fallow	1976	98.76	99.50	99.80	96.62	99.80	99.85	100.00	99.90	95.45	98.33	99.14	95.47
4	Fallow_rough_plow	1394	99.43	89.87	99.21	97.88	99.43	97.47	99.00	97.11	99.64	97.13	99.57	94.42
5	Fallow_smooth	2678	97.52	99.18	98.69	98.62	98.36	99.28	98.32	99.32	98.10	96.44	98.02	98.65
6	Stubble	3959	98.59	100.00	98.71	100.00	94.42	100.00	97.30	100.00	95.98	100.00	97.50	99.95
7	Celery	3579	99.02	98.84	99.75	100.00	99.44	99.50	99.86	99.97	99.55	93.74	99.69	96.88
8	Grapes_untrained	11271	96.14	96.57	78.74	94.91	69.56	92.81	89.14	86.64	74.49	97.80	70.19	81.58
9	Soil_vineyard_develop	6203	100.00	100.00	99.94	98.41	99.27	98.12	99.27	99.63	97.95	99.51	98.13	98.93
10	Corn_senesced_green_weeds	3278	95.35	100.00	93.11	97.32	95.30	99.55	98.96	93.19	91.76	82.77	89.48	76.28
11	Lettuce_romaine_4weeks	1068	99.85	100.00	98.31	90.28	98.78	96.97	98.88	99.81	99.81	89.43	97.57	85.06
12	Lettuce_romaine_5 weeks	1927	98.78	99.59	100.00	98.27	99.84	97.52	99.90	100.00	97.98	98.18	98.39	93.54
13	Lettuce_romaine_6 weeks	916	98.29	93.59	99.89	95.81	99.34	95.89	99.56	97.85	97.49	77.12	98.91	98.05
14	Lettuce_romaine_7 weeks	1070	98.45	96.71	98.32	96.60	97.76	95.18	97.48	98.12	92.43	77.81	97.01	89.95
15	Vineyard_untrained	7268	95.64	95.84	93.26	74.61	92.09	64.61	80.17	84.19	95.69	75.84	72.33	66.13
16	Vineyard_vertical_trellis	1807	95.75	100.00	97.51	100.00	96.29	100.00	97.62	100.00	95.79	94.59	94.36	99.30
Average Acc. and Average Rel.			98.06	97.51	97.17	96.21	96.16	96.05	93.15	94.05	95.45	92.32	94.16	92.03
Overall Acc.			95.74		93.85		91.40		91.78		92.04		88.34	
Kappa coefficient			0.95		0.93		0.90		0.91		0.91		0.87	

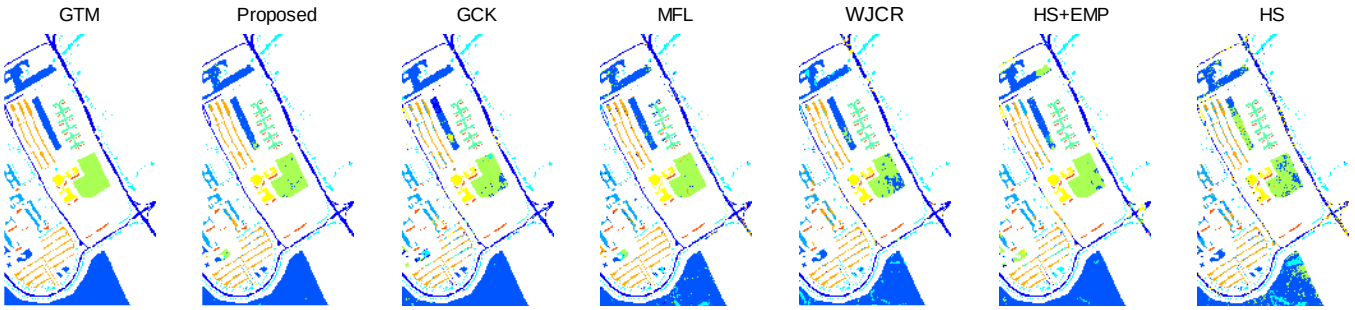

Fig. 3. The classification maps for Pavia dataset.

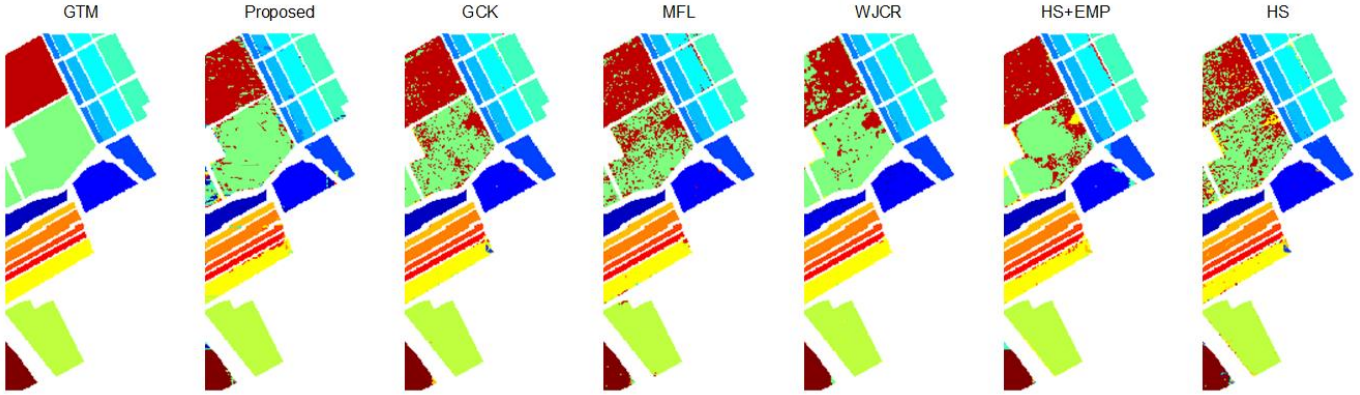

Fig. 4. The classification maps for Salinas dataset.

Table V. The McNemars test results.

Indian						
	Proposed	GCK	MFL	WJCR	HS+EMP	HS
Proposed	0	15.41	10.01	20.78	14.09	57.11
GCK	-15.41	0	-5.11	2.16	-2.99	42.57
MFL	-10.01	5.11	0	6.33	2.02	43.05
WJCR	-20.78	-2.16	-6.33	0	-4.99	39.65
HS+EMP	-14.09	2.99	-2.02	4.99	0	42.10
HS	-57.11	-42.57	-43.05	-39.65	-42.10	0

Pavia						
	Proposed	GCK	MFL	WJCR	HS+EMP	HS
Proposed	0	40.72	44.14	48.68	50.56	86.70
GCK	-40.72	0	3.04	9.84	11.29	53.58
MFL	-44.14	-3.04	0	6.69	6.70	50.67
WJCR	-48.68	-9.84	-6.69	0	-0.29	45.46
HS+EMP	-50.56	-11.29	-6.70	0.29	0	45.72
HS	-86.70	-53.58	-50.67	-45.46	-45.72	0

Salinas						
	Proposed	GCK	MFL	WJCR	HS+EMP	HS
Proposed	0	15.93	2.41	3.02	3.00	20.17
GCK	-15.93	0	22.84	14.79	14.28	43.81
MFL	-2.41	-22.84	0	-2.45	-4.56	21.34
WJCR	-3.02	-14.79	2.45	0	-1.79	21.75
HS+EMP	-3.00	-14.28	4.56	1.79	0	25.28
HS	-20.17	-43.81	-21.34	-21.75	-25.28	0

3- In Indian, HS+EMP is superior to WJCR. But, in Pavia, the performances of HS+EMP and WJCR are similar.
4- The worst classification results are related to the use of just spectral bands (HS).
5- The spatial neighborhood information significantly attenuates the salt and pepper noise in classification maps obtained by the proposed, GCK, MFL, WJCR, and HS+EMP methods.

The main advantages of the proposed framework can be represented as follows:
The use of different feature extraction methods with different ideas provides the useful features for classification. The features obtained by different feature extraction methods contain complement information and have minimum overlap with respect to each other. This is an important principle in the decision fusion rule.

1- It copes with the small sample size problem with dimensionality reduction of hyperspectral image and by providing some groups of useful extracted features and classification of each subgroup individually.
2- The useless spatial information such as noise and distortions is removed from the image by applying the smoothing filter on each subgroup of data.
3- The useful spatial information is added to data by applying morphological filters to the smoothed images.
4- For exploiting the complement information contained in different subgroups of extracted features, the MV rule is used as a decision fusion technique.

V. CONCLUSION

A spectral-spatial classification framework was proposed in this paper. The proposed method uses the complement groups of spectral features extracted from several different state-of-the-art feature extraction methods with minimum redundant information. The proposed framework utilizes the smoothing filters for removing the useless spatial information while utilizes the morphological filters for adding the useful spatial information to hyperspectral image. The proposed method copes with the curse of dimensionality by feature extraction of the high dimensional data in six different groups and data processing on each subspace individually. The experimental results on Indian hyperspectral image with low spatial resolution and Pavia and Salinas hyperspectral images with high spatial resolution demonstrate the superiority of the proposed framework compared to several spectral-spatial classification methods.

REFERENCES

[1] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 5, pp. 2271–2282, May 2010.

[2] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," IEEE Trans. Geosci. Remote Sens., vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[3] M. Borhani and H. Ghassemian, "Kernel Multivariate Spectral–Spatial Analysis of Hyperspectral Data," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 2418-2426, June 2015.

[4] F. Mirzapour and H. Ghassemian, "Multiscale Gaussian Derivative Functions for Hyperspectral Image Feature Extraction," IEEE Geosci. Remote Sens. Lett, vol. 13, no. 4, pp. 525-529, April 2016.

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

[6] M. Imani and H. Ghassemian, "Binary coding based feature extraction in remote sensing high dimensional data," Information Sciences, vol. 342, pp. 191-208, 2016.

[7] Q. Zhang, Y. Tian, Y. Yang, and C. Pan, "Automatic Spatial–Spectral Feature Selection for Hyperspectral Image via Discriminative Sparse Multimodal Learning," IEEE Trans. Geosci. Remote Sens., vol. 53, no. 1, pp. 261–279, Jan. 2015.

[8] M. Imani and H. Ghassemian, "Feature Extraction Using Weighted Training Samples," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 7, pp. 1387-1391, July 2015.

[9] X. Liu, L. Wang, J. Zhang, J. Yin, and H. Liu, "Global and Local Structure Preservation for Feature Selection," IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 6, pp. 1083–1095, June 2014.

[10] S. A. Hosseini and H. Ghassemian, "Rational Function Approximation for Feature Reduction in Hyperspectral Data," Remote Sens. Lett., vol. 7, no. 2, 101-110, 2016.

[11] M. Imani and H. Ghassemian, "Feature space discriminant analysis for hyperspectral data feature reduction," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 102, pp. 1-13, 2015.

[12] M. Golipour, H. Ghassemian, and F. Mirzapour, "Integrating Hierarchical Segmentation Maps With MRF Prior for Classification of Hyperspectral Images in a Bayesian Framework," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 2, pp. 805-816, Feb. 2016.

[13] A. Zehtabian and H. Ghassemian, "An Adaptive Pixon Extraction Technique for Multispectral/Hyperspectral Image Classification," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 4, April 2015.

[14] G. Camps-Valls, L. Gomez-Chova, J. Muoz-Mar, J. Vila-Francs, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," IEEE Geosci. Remote Sens. Lett., vol. 3, no. 1, pp. 93–97, Jan. 2006.

[15] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," J. Mach. Learn. Res., vol. 9, no. 11, pp. 2491–2521, Nov. 2008.

[16] J. Li, P. Marpu, A. Plaza, J. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens., vol. 51, no. 9, pp. 4816–4829, Sept. 2013.

[17] A. Green, M. Berman, P. Switzer, and M. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," IEEE Trans. Geosci. Remote Sens., vol. 26, no. 1, pp. 65–74, Jan. 1988.

[18] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," IEEE Trans. Geosci. Remote Sens., vol. 49, no. 12, pp. 4865–4876, Dec. 2011.

[19] B. Du, L. Zhang, L. Zhang, T. Chen, and K. Wu, "A discriminative manifold learning based dimension reduction method," Int. J. Fuzzy Syst., vol. 14, no. 2, pp. 272–277, Jun. 2012.

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

[21] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L.Zhang, and J. Atli Benediktsson, A. Plaza, "Multiple Feature Learning for Hyperspectral Image Classification," IEEE Trans. Geosci. Remote Sens., vol. 53, no. 3, pp. 1592-1606, March 2015.

[22] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," IEEE Trans. Geosci. Remote Sens., vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[23] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 477–489, Jan. 2014.

[24] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 7, no. 6, pp. 2200–2208, June 2014.

[25] M. Xiong, Q. Ran, W. Li, J. Zou, and Q. Du, "Hyperspectral Image Classification Using Weighted Joint Collaborative Representation," IEEE Geosci. Remote Sens. Lett., vol. 12, no. 6, pp. 1209–1213, June 2015.

[26] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 3, pp. 466–479, Mar. 2005.

[27] 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.

[28] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," IEEE Trans. Geosci. Remote Sens., vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[29] M. Imani and H. Ghassemian, "Band Clustering-Based Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples," IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 8, pp. 1325-1329, August 2014.

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

[31] M. Imani and H. Ghassemian, "Feature Extraction Using Attraction Points for Classification of Hyperspectral Images in a Small Sample Size Situation," IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 11, pp. 1986-1990, Nov. 2014.

[32] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, San Diego, CA, USA: Academic, 1990.

[33] 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.

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

[35] J. Xu, J.Yang, Z. Gu, N. Zhang, "Median–mean line based discriminant analysis," *Neurocomputing*, vol. 123, pp. 233–246, 2014.

[36] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," IEEE Geosci. Remote Sens. Lett., vol. 5, no. 4, pp. 625–629, Oct. 2008.

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

[38] 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, May 2004.

[39] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 27:1-27:27, April 2011.

[40] M. Imani and H. Ghassemian, "Hyperspectral Images Classification by Spectral-Spatial Processing," 8th International Symposium on Telecommunications (IST'2016), Tehran, Iran, 27-29 Sept. 2016.

Maryam Imani received the B.Sc. and M.Sc. degrees in electrical engineering from Shahed University, Tehran, Iran, and the Ph.D. degree in electrical engineering from Tarbiat Modares University, Tehran, Iran in 2009, 2011, and 2015 respectively. She continued her research in Tarbiat Modares University as a postdoc. She is an Assistant professor of the Faculty of Electrical and Computer Engineering at Tarbiat Modares University, Tehran, Iran. Her research interests include pattern recognition, signal and image processing, information analysis, and remote sensing.

Hassan Ghassemian received the B.S.E.E. degree from Tehran College of Telecommunication in 1980 and the M.S.E.E. and Ph.D. degree from Purdue University, West Lafayette, USA in 1984 and 1988 respectively. Since 1988, he has been with Tarbiat Modares University in Tehran, Iran, where he is a Professor of Computer and Electrical Engineering. His research interests focus on Multi-Source Signal/Image Processing and Information Analysis and Remote Sensing.