Hyperspectral image classification using K-plane clustering and kernel principal component analysis

Sayeh Mirzaei[1]

Abstract
In this paper, we present a new approach for hyperspectral image classification. The pixels' spectra are grouped into clusters in an unsupervised manner using an improved version of plane based clustering. Since the pixels containing the same substances are linearly correlated, the proposed plane-based clustering can effectively group the data points. Plane-based clustering is a more appropriate choice than point based clustering schemes for grouping the datasets which are distributed around hyperplanes instead of hyperspheres. Then, Kernel Principal Component Analysis (KPCA) is applied to each cluster individually to obtain multiple kernel vectors for each data point. Applying non-linear kernels, can greatly increase the discrimination power of the acquired features. The feature vectors are extracted by a weighted linear combination of the kernel components obtained from each cluster. We compute optimal weights using the cluster hyperplane parameters. Since the whole procedure is performed in an unsupervised manner, the proposed approach can enhance the generalization power of the extracted features. Then, morphological attribute filters are applied to the feature maps to effectively utilize spatial relations. Hence, the acquired compact feature vectors include both spectral and spatial information. SVM is used for classification. The experiments performed on three well-known hyperspectral datasets reveal the effectiveness of the proposed feature extraction approach.

Abbreviations
AP   Attribute Profile
CNN   Convolutional Neural Network
EMAP   Extended Morphological Attribute Profile
GMM   Gaussian Mixture Model
HSI   Hyperspectral Image
KPC   K-Plane Clustering
KPCA   Kernel Principal Component Analysis

✉ Sayeh Mirzaei
    s.mirzaei@ut.ac.ir

1   School of Engineering Science, College of Engineering, University of Tehran, Tehran, Iran

LDA	Linear Discriminant Analysis
LKPPC	Local K-Proximal Plane Clustering
MAP	Morphological Attribute Profile
PCA	Principal Component Analysis
SVM	Support Vector Machine

1 Introduction

HYPERSPECTRAL image (HSI) analysis has found several applications in agriculture, health monitoring, mineral mapping, and many other remote sensing studies. The hyperspectral sensors can acquire images in hundreds of spectral bands which makes them very useful for recognizing spectrally different substances ([2, 24, 31]). Since the spectral sensors suffer from low spatial resolution, each pixel might contain multiple materials. Hyperspectral unmixing is the process of decomposing each pixel of the image to its constituent substances called endmembers and the abundance of the endmembers in constructing a pixel. Various spectral and spatial features have been extracted from the 3D hyperspectral data cube for the purpose of un-mixing the pixels ([16, 18, 28] and classification ([14, 15, 25, 26]. Deep neural network architectures have also been proposed for representation learning as well as providing well discriminant features for classification ([27, 29, 34]. In [30], a self-looping convolution neural network is proposed for efficient feature extraction. This model, obtains spate representations for different spatial levels through multiscale setting. Actually, deep learning models include many trainable parameters and they need many labeled samples to achieve optimal performance. However, large number of labeled data samples is not affordable for HIS classification tasks. A review of research works related to deep learning models for HSI classification with few labeled samples is presented in [17].

HSI data suffers from the curse of dimensionality. Many dimensionality reduction techniques have been applied to overcome this issue and eliminate redundant information. The most popular method for this purpose is Principal Component Analysis (PCA). PCA projects the data points into a lower-dimensional subspace with the objective to retain the variance and minimize the least square error. PCA does not use the class labels of the training samples and therefore is regarded as an unsupervised feature reduction technique. Linear Discriminant Analysis (LDA) is a supervised dimensionality reduction method that tries to find a lower-dimensional subspace so that the projected data points have maximized between-class scatter and minimized within-class scatter. However, for working efficiently, LDA requires many training samples and the performance is poor for small training data. Furthermore, LDA can maximally extract *k-1* features where k is the number of classes.

Since PCA is based on the second-order statistics, its performance is limited for high-dimensional HSI data. Kernel PCA (Kernel) first introduced in [32], can improve efficiency ([9, 11]. Kernel methods have been widely used in HSI classification algorithms ([1, 5, 13, 19]. The idea is to apply a non-linear transform with the aim to make the data points more separable in the transformed space. Single kernel learning is not necessarily capable of providing discriminatory features for performing classification. Multiple Kernel Learning (MKL) approaches have been already exploited for HSI classification [13]. Using these methods, effective multimodal information can be extracted from the HSI data. Moreover, they can efficiently compromise between the model accuracy and the power of generalization. In this paper, we aim to utilize multiple kernels obtained from different clusters of data points to effectively improve the discriminatory property of the extracted features.

Several clustering schemes have been extensively utilized in machine learning applications for capturing the structure of data in an unsupervised manner ([4, 20]. K-means, GMM, and hierarchical clustering are among the mostly used approaches. K-plane clustering (KPC) has been introduced in [3]. It is reasonable that plane based clustering is more suitable than point based clustering (e.g. k-means) for capturing the linear correlations of the data points. KPC is a more appropriate choice for grouping the datasets which are distributed around hyperplanes instead of hyperspheres. Therefore, we apply an improved version of k-plane clustering for grouping the HSI pixels' spectra.

We first apply k-plane clustering on the training data points.

It is the first time that this clustering scheme is applied to HSI data. We select a predefined value for the number of classes. Then, we employ the KPCA method to the pixels of each cluster individually. Hence, in contrast to the conventional approaches which estimate the covariance matrix of PCA using all of the data points and apply it for feature reduction, we obtain a separate covariance matrix corresponding to the data points of each cluster and then acquire a weighted combination of them for constructing the final discriminant features. This way, we have separate PCs corresponding to each cluster, and these PCs can be regarded as multiple kernels that are combined linearly. The weights used for this combination are obtained based on the distribution of the clusters' data points. Instead of using linear PCA, we propose to exploit kernel PCA to enhance the discriminatory property of the extracted components and improve the classification performance. Hence, we present a sort of multiple kernel learning approach in which the kernels are adaptively acquired from data in an unsupervised manner. Spatial information is implicitly taken into account since the feature vector of each pixel is obtained through combination of PCs extracted from different clusters. This is due to the fact that adjacent pixels containing the same materials are most likely aligned in the same clusters. However, we also apply morphological attribute filters to utilize the spatial structure of the pixels in a well-organized manner.

In this paper, a novel feature extraction approach based on the fusion of unsupervised k-plane clustering and KPCA is proposed. The objective is to find the best combination of kernel components as discriminant features for each pixel. Since the whole procedure is performed in an unsupervised manner, the proposed approach can enhance the generalization power of the extracted features. Morphological attribute filters have also been applied to the obtained feature maps to effectively exploit the spatial context of the image. This way, the extracted features include both spectral and spatial information. The other advantage of the proposed method over most conventional approaches is that it utilizes more compact feature vectors which contain joint spatial-spectral content. Many other previous methods extract spatial and spectral features separately and stack them as composite kernels while our suggested technique extracts feature vectors containing both spatial and spectral attributes in a compressed manner. SVM with the RBF kernel is used as classification method which is cheaper in terms of complexity and required computational resources for implementation than deep neural network architectures. Moreover, SVM performs well in a limited training dataset situation which is common for remote sensing applications. The experiments verify the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. In Section 2, the k-plane clustering scheme is described which is used to group the pixels in an unsupervised manner. Section 3, explains how the final feature vectors are extracted through weighted combination of kernel principal components acquired from each cluster and applying morphological attribute filters. Section 4 is dedicated to the experiments and the classification performance

evaluation is provided for well-known hyperspectral datasets. Conclusive remarks are ultimately presented in Section 5.

2 K-plane clustering

It is rational that the pixels containing the same substances be linearly correlated in the spectral domain. Therefore, plane-based clustering is regarded as a more effective and relevant unsupervised grouping method compared to point-wise clustering approaches. Central clustering methods such as k-means or fuzzy c-means assume that the data points are distributed around multiple centroids. However, this assumption is not valid for many applications. For instance, HSI pixel spectra most likely fall into clusters around center hyperplanes instead of center points.

K-plane clustering (KPC) [3] was proposed to address clustering the data with the mentioned structure. The algorithm starts with k random center hyperplanes. Then, the following two steps are repeated in a loop:

1. The data points are assigned to the nearest hyperplane.
2. The center hyperplanes are updated based on the points assigned to each cluster in the first step.

The issue with KPC is that the center hyperplane can extend infinitely. Local K-Proximal Plane Clustering (LKPPC) [33] is an improved version of KPC to solve this issue. It considers both within-cluster and between-cluster distances. Moreover, it enforces the data points to localize around some prototypes by incorporating k-means to the KPC problem. In summary, LKPPC tries to make each cluster data point close to both the center hyperplane and the prototype. At the same time, it makes the cluster points far from other cluster hyperplanes. A Laplace graph procedure is also suggested [33] for initialization which makes the algorithm more stable.

LKPPC groups the data points aligned in a matrix $\mathbf{A}_{m \times n}$ (m indicates the number of data samples and n is the number of features) into k clusters by optimizing the following objective function:

$$\min_{w_i, b_i, v_i} \left\| A_i w_i + b_i e_i \right\|_2^2 + c_1 \left\| A_i - e_i v_i^T \right\|_2^2 - c_2 \left\| B_i w_i + b_i \bar{e}_i \right\|_2^2$$
$$s.t. \left\| w_i \right\|_2^2 = 1 \tag{1}$$

where $\mathbf{e}_i$ and $\bar{\mathbf{e}}_i$ are vectors of ones of proper dimensions for $i = 1, 2, \ldots, k$. $\mathbf{w}_i^T \mathbf{x} + b_i = 0$ specifies the ith cluster hyperplane ($\mathbf{w}_i$ and b_i represent the hyperplane weight vector and bias respectively), $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$ shows the samples of the ith cluster, $\mathbf{B}_i \in \mathbb{R}^{(m-m_i) \times n}$ denotes the samples not belonging to the ith cluster, and $\mathbf{v}_i$ is the prototype of the ith cluster. Therefore, the first term in (1) enforces the closeness of the points to the ith cluster hyperplane. Parameter $c_1 \in (0, 1)$ restrains the extension of the ith hyperplane by penalizing the points far from the ith cluster prototype $\mathbf{v}_i$. So it controls localization of the ith hyperplane and performs similar to k-means. The parameter $c_2 > 0$ controls the distance of the other data points from the ith cluster hyperplane and makes them far away from it. This optimization problem has been solved with the Lagrangian multiplier method and the update relations for obtaining the cluster hyperplanes are given in [33]. Termination takes place based on monitoring the amount of stability of the acquired clusters or the number of iterations.

After finding the cluster hyperplanes using the training data samples, a new data point $\mathbf{x}$ is assigned to the cluster$y(\mathbf{x})$which minimizes the following criterion:

$$y(x) = \arg\min_i \left(\left\| w_i^T x + b_i \right\|_2^2 + c_1 \|x - v_i\|_2^2 \right), i = 1, 2, \ldots, k \tag{2}$$

In the current application, pixels' spectra construct the data matrix A, where the number of rows (m), denotes the total number of pixels used for training and the number of columns (n), indicates the number of spectral bands.

3 Feature extraction using KPCA and morphological filters

In the previous step, the pixels were grouped into k clusters using the LKPPC method where k is set to the actual number of classes. In the current stage, we apply KPCA to each group separately to obtain kernel components corresponding to each cluster. We take the number of principal components (PCs) equal to k. The polynomial kernel of degree 4 is used which has resulted in better performance empirically:

$$\mathbf{K}(\mathbf{x}, \mathbf{z}) = \left(\mathbf{x}^T \mathbf{z} + 1 \right)^4 \tag{3}$$

We examined other kernel types including the linear kernel (PCA), and RBF kernel, and concluded that the polynomial kernel lead to the superior performance for the HSI classification task. So we have multiple kernels extracted from multiple clusters. In order to acquire the feature vector for each pixel, we first obtain linear combination of the corresponding kernel Pcs with the weights$p_i$evaluated as follows for each pixel $\mathbf{x}$:

$$
\begin{aligned}
q_i &= \left\| w_i^T x + b_i \right\|_2^2 + c_1 \|x - v_i\|_2^2, \ i = 1, 2, \ldots, k \\
q &= \left[q_1 \, q_2 \ldots q_k \right]^T \\
p_i &= \exp\left(-2 \left(\frac{q_i - \min(q)}{\max(q)} \right) \right), \ i = 1, 2, \ldots, k
\end{aligned}
\tag{4}
$$

The weights p_i given by (4) are directly related to the membership probability of the pixel $\mathbf{x}$in the ith cluster. Hence, we give higher weight to the cluster PCs with higher probability in the combination. Consequently, if the kernel principal components corresponding to each cluster is denoted by $\mathbf{KPC}_i$, $i = 1, 2, \ldots, k$, the feature vector $\mathbf{f}$ is acquired by linearly combining these components as follows:

$$\mathbf{f} = \sum_{i=1}^{k} p_i \mathbf{KPC}_i \tag{5}$$

The proposed scheme provides some sort of multiple kernel features where the kernels are effectively acquired from different clusters to boost the discrimination power. We apply the morphological attribute filters to the features extracted through combination of kernel PCs to efficiently exploit the spatial relations. Morphological attribute profiles (MAP) have been already utilized to extract the spatial information of the image ([6, 7, 10, 22].

Attribute Profile (AP) is constructed by applying several attribute filters sequentially. Aps can be extracted for different attributes such as area, volume, etc. and stacked to make an Extended Multi-Attribute Profile (EMAP) [6]. The outputs of the filters are compared with predefined threshold values at each region of the image. If the attribute is smaller than the threshold, the region grayscale values are replaced with the neighboring region with

closer value. The operation is called thinning when the region is merged with a lower gray-scale value and it is called thickening when it is merged with larger grayscale value. Some useful attributes for HSI analysis include area, volume (sum of the intensities of the pixels belonging to each region),length of the diagonal of the box bounding each region, moment of inertia, shape factor, homogeneity, standard deviation, and entropy of the grayscale values of the pixels.

The length of the input feature vectors (**f**) to the morphological filters is equal to the number of PCs (*npcs*). Suppose that the length of the threshold set is equal to *T*. Then, the EMAP vector obtained for each pixel would be of length $(2 \times T + 1) \times npcs$. Factor 2 indicates the thinning and thickening operations corresponding to each threshold value. These EMAP vectors construct the discriminative inputs fed to the classifier. We employ SVM with RBF kernel for classification. SVM performs efficiently in limited training data size situations which is quite common for HSI datasets. Figure 1 demonstrates the feature extraction and classification process.

4 Experiments

Some widely studied HSI datasets were used to evaluate the performance of the proposed approach. The experiments are carried out on real datasets, Indiana Pines, Pavia University, and Salinas. The description of these HSI datasets is given in the following subsections. For each dataset, the number of clusters *k* is set to the actual number of classes and the number
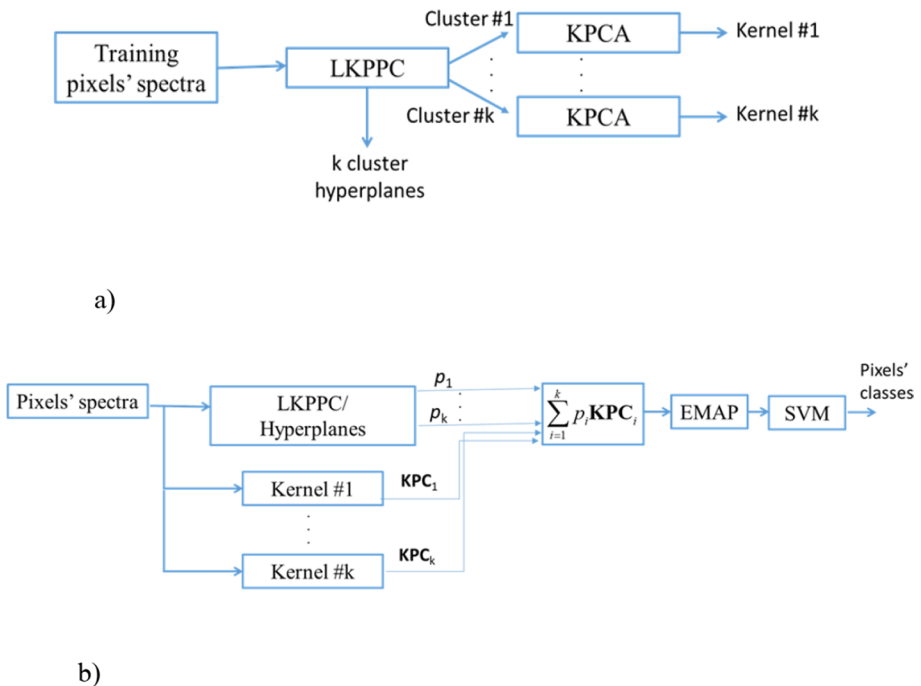

Fig. 1 Block diagram of the proposed feature extraction and classification schemes, (**a**) Training process, (**b**) Classification framework

of kernel PCs is taken equal to k. c_1 and c_2 parameters are both set to 0.9. Two morphological attributes are selected including the area and the length of the diagonal of the bounding box. The corresponding threshold values are taken as [10 15 20] and [50100500] respectively. Hence, the size of the EMAP vector is $k \times 13$. This vector is the input feature to the SVM classifier. 5-fold cross-validation is executed and the average results are reported. Therefore, at each run, 80% of the HSI pixels are used for training and the classification performance is evaluated with the remaining pixels. We have implemented the algorithms in MATLAB R2017b with Intel core i7 CPU 2.6GHz and 12GB RAM. The effectiveness of the proposed feature extraction method is demonstrated through comparison with two other approaches. In the first approach, KPCA is applied to the whole training data points ($npcs$ = Actual number of classes) and then the EMAP vector is obtained. In the second approach, the LKPPC scheme is replaced with k-means for clustering the pixels' spectra. Then, KPCA is applied to each cluster separately and the combination weights p_i for each pixel $\mathbf{x}$ are obtained similar to (4) by replacing the values q_i as follows:

$$q_i = \|x - \mu_i\|_2^2, \ i = 1, 2, \ldots, k$$
$$q = [q_1 \ q_2 \ldots q_k]^T$$
$$p_i = \exp\left(-2\left(\frac{q_i - \min(q)}{\max(q)}\right)\right), \ i = 1, 2, \ldots, k \tag{6}$$

$\mathbf{\mu}_i$, $i = 1, 2, \ldots, k$ denotes the ith cluster centroid given by k-means.

In the following reports, the first approach is called "KPCA-all" and the second approach is stated as "k-means".

4.1 Indian pines dataset

This scene was collected by AVIRIS sensor over the Indian Pines test site in North-western Indiana. It contains 145×145 pixels and 224 spectral bands in the wavelength range of 0.4–2.5 µm. The spatial resolution of this dataset is 20 m per pixel. The Image consists of two-thirds agriculture, and one-third forest or other natural perennial vegetation. The ground truth includes sixteen classes as demonstrated in Fig. 2. The number of bands has been reduced to 200 by removing those bands which contain the regions of water absorption: (104–108), (150–163), 220. Table 1 reports the classification accuracies acquired by the proposed method. In order to provide a useful comparative material, we have also evaluated the performance for k-means and KPCA-all approaches. The results manifest the outperformance of the proposed algorithm over the two other approaches in terms of individual, overall, and average accuracies. K-means perform better than KPCA-all. So acquiring PCs for individual clusters instead of the whole training data improves the performance. It can be associated with the well-discriminant spectral information extracted by clustering/KPCA combination. The results reveal that LKPPC is a more effective clustering scheme than k-means for grouping HSI pixels' spectra.

We have also provided classification maps obtained through different approaches in Fig. 3. They can give a better view of the superior classification performance of the proposed technique over the two other methods.

4.2 Pavia University dataset

This dataset was collected by the ROSIS sensor from Pavia, northern Italy. There are 610×340 pixels in the image and the number of spectral bands is 103. The spatial
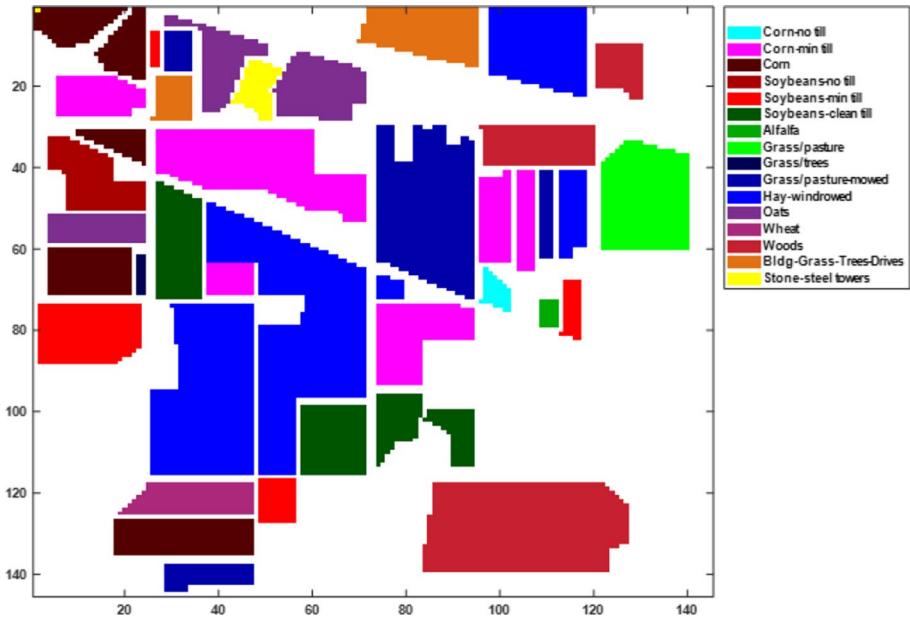
Fig. 2 Groundtruth map of Indiana Pines scene containing 16 classes

Table 1 The individual class accuracies (in percent) obtained for Indiana Pines dataset

Methods			
Classes	KPCA-all	k-means	Proposed method
Alfalfa	100	100	100
Corn-notill	80.77	94.76	94.76
Corn-mintill	63.25	87.95	100
Corn	77.08	87.50	100
Grass-pasture	91.75	95.88	100
Grass-trees	98.63	99.32	100
Grass-pasture-mowed	83.33	83.33	100
Hay-windrowed	100	100	100
Oats	25	50	100
Soybeans-notill	63.59	84.62	99.49
Soybeans-mintill	83.50	93.08	99.39
Soybeans-clean	88.24	96.64	100
Wheat	100	100	100
Woods	98.42	98.42	100
Buildings-Grass-Trees-Drives	64.10	88.46	100
Stone-Steel-Towers	84.21	89.47	94.74
Average Accuracy	81.37	90.59	**99.27**
Overall Accuracy	83.36	93.58	**99.03**

The bold items specify the maximum values for each dataset
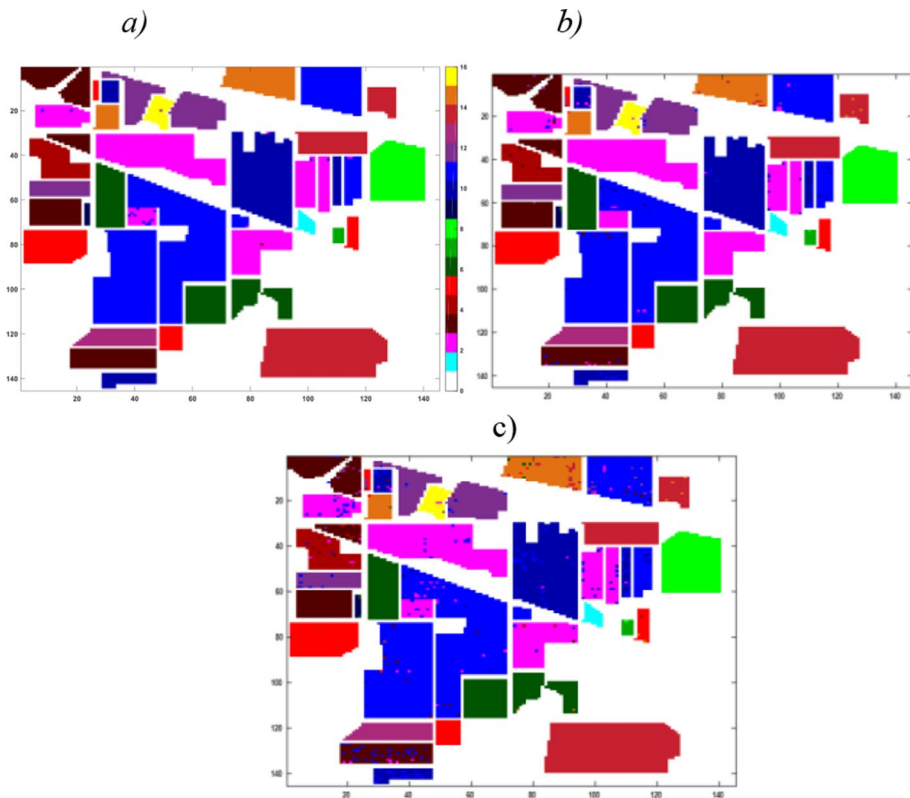
Fig. 3 Classification maps for Indiana dataset acquired by (**a**) The proposed method, (**b**) k-means, (**c**) KPCA-all

resolution is 1.3 m. The ground-truth data includes 9 classes as depicted in Fig. 4. The classification evaluation metrics are listed in Table 2. Performance improvement is notice-able using the proposed algorithm compared with k-means or KPCA-all methods. Again, k-means outperforms KPCA-all which indicates the advantage of applying clustering schemes before feature reduction through KPCA.

We have also compared classification maps obtained through different methods in Fig. 5 to visualize the measures reported in Table 2. Figure 5 exhibits the near-perfect classifica-tion achieved by the proposed approach.

4.3 Salinas dataset

This scene was collected by the AVIRIS sensor over Salinas Valley, California. The spatial resolution is 3.7 m. The image consists of 512×217 pixels and 224 spectral bands. 20 water absorption bands, (108–112), (154–167), 224 are discarded. Salinas groundtruth contains 16 classes including bare soils, vegetables, and vineyard fields (see Fig. 6). The outcomes of the different classification approaches are reported in Table 3. Again, the best results are achieved by the proposed method. The same pattern as the other two datasets appears in the classification results which reveals the effectiveness of
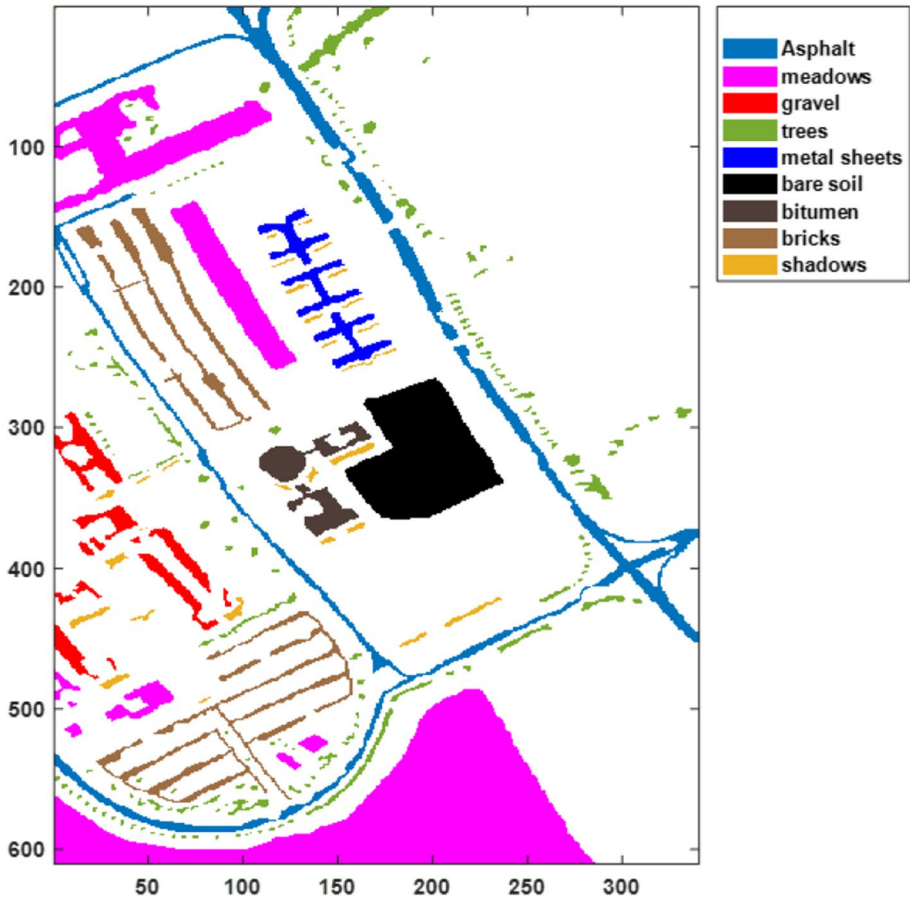
Fig. 4 Groundtruth map of Pavia University scene containing 9 classes

Table 2 The individual class accuracies (in percent) obtained for Pavia University dataset

Methods			
Classes	KPCA-all	k-means	Proposed method
Asphalt	90.13	97.89	99.85
Meadows	94.99	98.74	99.92
Gravel	66.67	83.57	98.81
Trees	88.58	95.27	99.51
Painted metal sheets	100	100	99.63
Bare Soil	77.63	91.25	100
Bitumen	57.52	97.37	100
Self-Blocking Bricks	86.30	94.03	99.73
Shadows	99.47	99.47	100
Average Accuracy	84.59	95.29	**99.72**
Overall Accuracy	88.69	96.34	**99.81**

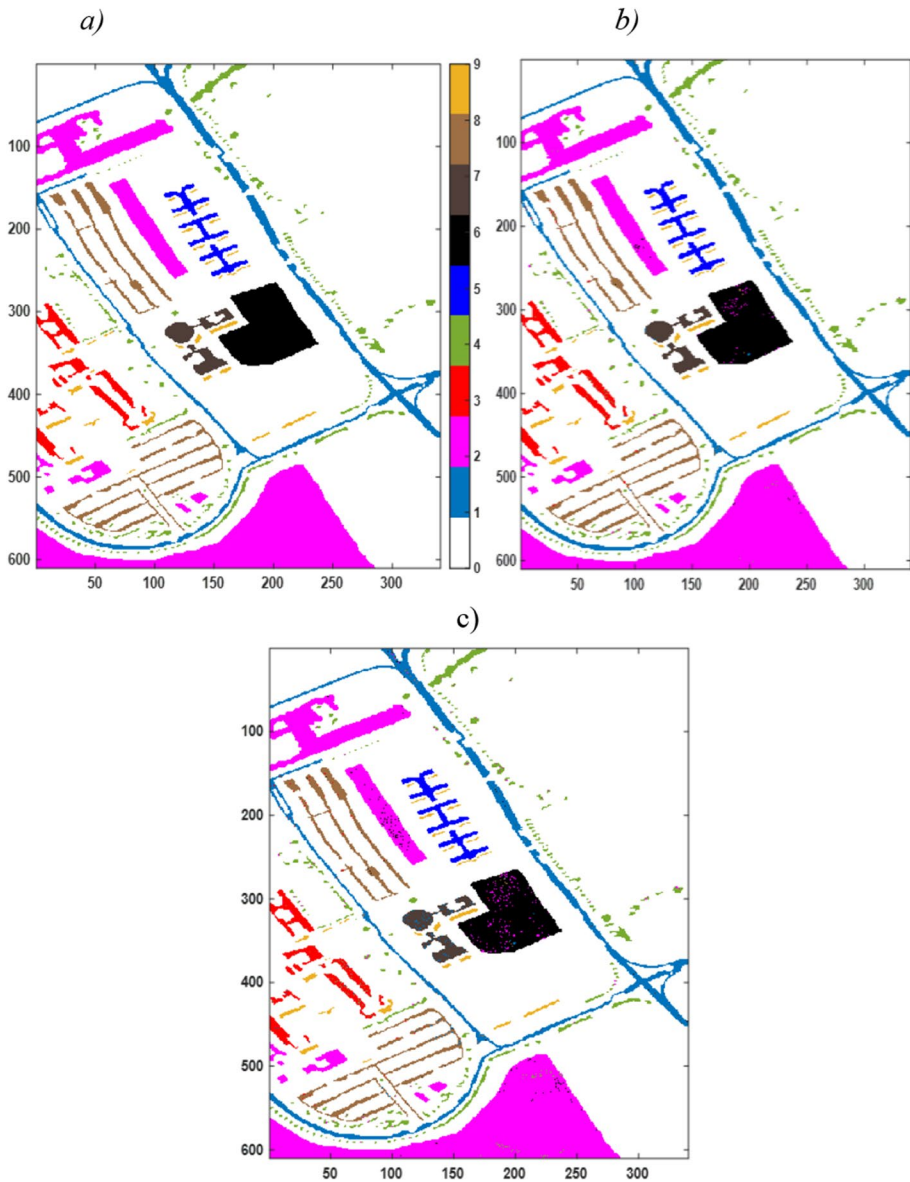The bold items specify the maximum values for each dataset

Fig. 5 Classification maps for Pavia dataset acquired by (**a**) The proposed method, (**b**) k-means, (**c**) KPCA-all

the proposed plane clustering approach. Figure 7 represents the classification maps corresponding to different approaches which indicates the superior performance attained by the proposed method.

In general, the superiority of the suggested scheme over KPCA-all and k-means shows the effectiveness of applying unsupervised clustering for obtaining multiple kernel PCs and the advantage of using k-plane clustering respectively.
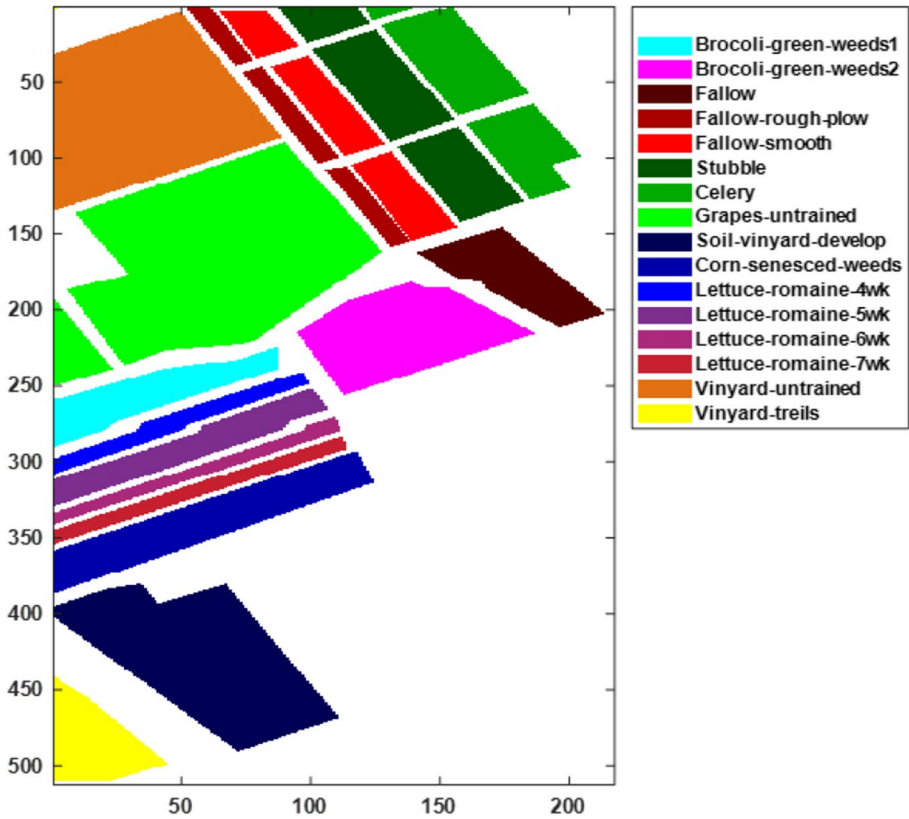
Fig. 6 Groundtruth map of Salinas scene containing 16 classes

4.4 Performance comparison with the state-of-the-art methods

To verify the effectiveness of the proposed approach, we provide the results of the comparison with other recent methods ([12, 21, 23] on Pavia and Salinas datasets. A non-linear multiple kernel learning approach is proposed in [12] in which the kernels are obtained based on morphological attribute profiles. In [21], a Convolutional Neural Network (CNN) architecture called contextual deep CNN is introduced which uses local spatio-spectral relationships of neighboring pixels through applying multi-scale convolutional filter bank. An automatic clustering-based two-branch convolutional neural network is proposed in [23]; First, to reduce the intraclass spectral variation, the HSI pixels are automatically subdivided into smaller classes by clustering; second, in order to suppress the interference of spectral amplitude variation, the SincNet is introduced to capture the spectral pattern by giving more weight to the spectral shape; third, the DS-CNN with double directional strip convolution kernel is designed to extract spatial feature. The resulted overall accuracies obtained versus different number of training samples per class are reported in Table 4. Similar to the reports provided by the above references, we perform the random train/test splitting 20 times and compute the mean and standard deviation of overall classification accuracy. Table 5

Table 3 The individual class accuracies (in percent) obtained for Salinas dataset

Methods			
Classes	KPCA-all	k-means	Proposed method
Brocoli_green_weeds_1	99.75	100	100
Brocoli_green_weeds_2	100	100	100
Fallow	99.49	100	100
Fallow_rough_plow	100	99.64	10
Fallow_smooth	99.63	99.63	99.81
Stubble	99.75	100	100
Celery	99.86	99.86	99.86
Grapes_untrained	89.14	91.84	99.87
Soil_vinyard_develop	99.84	99.92	100
Corn_senesced_green_weeds	98.48	99.09	99.70
Lettuce_romaine_4wk	99.53	100	100
Lettuce_romaine_5wk	100	100	100
Lettuce_romaine_6wk	100	100	100
Lettuce_romaine_7wk	96.73	98.13	99.07
Vinyard_untrained	65.27	78.47	99.79
Vinyard_vertical_trellis	100	100	100
Average Accuracy	96.72	97.91	**99.88**
Overall Accuracy	92.82	95.27	**99.89**

The bold items specify the maximum values for each dataset

exhibits the significant performance improvement achieved by the proposed method for all different numbers of training samples. This improvement is attained in spite of the fact that the computational burden of the proposed method is noticeably less than the other methods; Particularly compared with Deep CNN approaches, our suggested scheme requires much fewer computational resources.

To provide more comparison material verifying the effectiveness of the proposed approach, we compare our method with two other recent studies as well. In [35], a deformable CNN structure is proposed (DHCNet) in which the size and shape of the convolutional sampling locations can be adaptively adjusted. Experimental results are reported for Pavia University dataset. The training set consists of 45, 55, and 65 samples, respectively, randomly selected per class. The comparison between DHC-Net and our proposed approach can be observed in Table 5. Different classification performance metrics including Overall Accuracy (OA), Average Accuracy (AA), and Kappa are reported for 44, 55, and 65 training samples per class. It is evident that the proposed scheme outperforms the DHCNet approach in terms of all performance measures.

In [8], a novel squeeze multibias network (SMBN) is suggested for HSI classification. The multibias module adaptively selects meaningful CNN patches for classification. The squeeze convolution module can greatly reduce the number of parameters in the network. We compare the performance of our method with SMBN technique for Indiana Pines dataset with 10% training. Individual class accuracies along with the statistical metrics are reported in Table 6. The proposed method results in better AA and OA measures compared with SMBN approach.
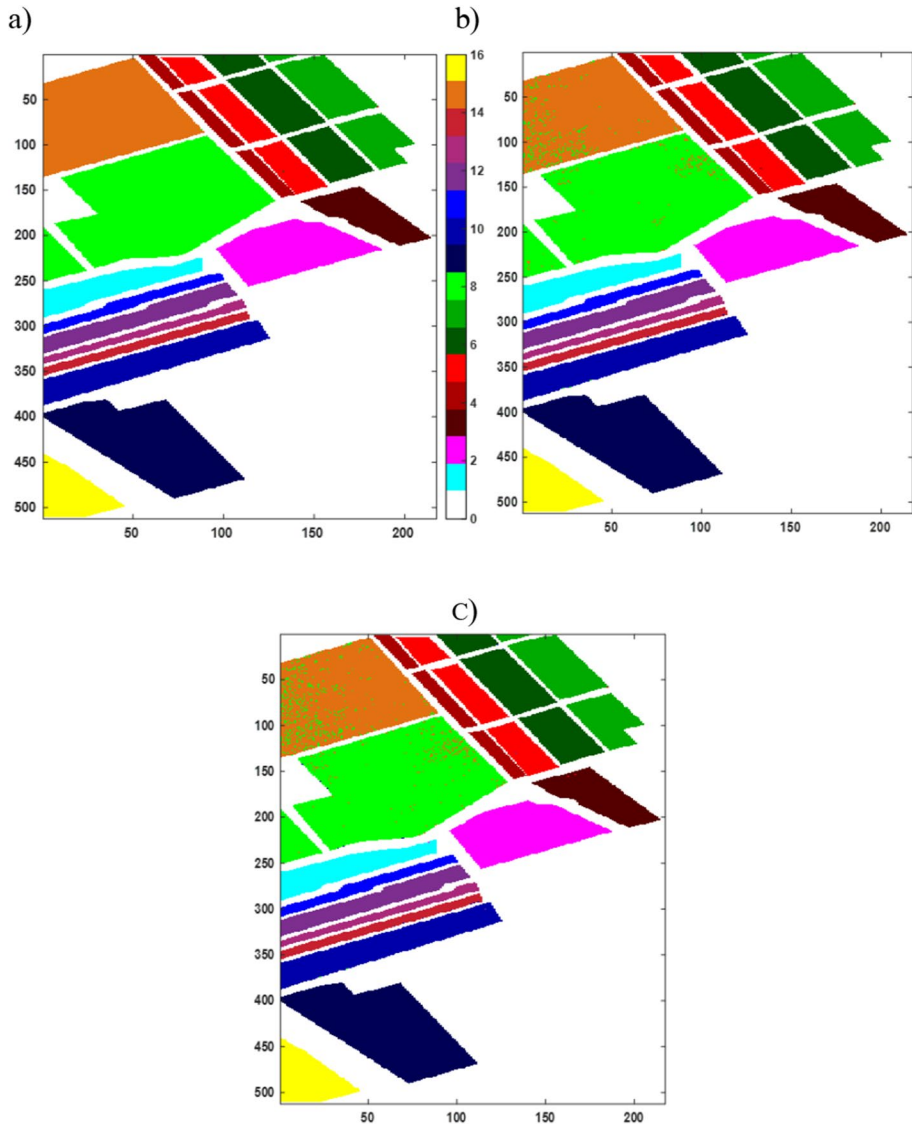
Fig. 7 Classification maps for Salinas dataset acquired by (**a**) The proposed method, (**b**) k-means, (**c**) KPCA-all

5 Conclusion

We propose a novel approach for HSI classification. We use a plane based clustering scheme to group the pixels' spectra without supervision. Then, KPCA is applied to each cluster to obtain kernel components of the clusters separately. Weighted combination of these kernel components is acquired for each pixel to construct the feature map. Hence, we present a multiple kernel learning approach in which the kernels are adaptively acquired

Table 4 Overall accuracies (in percent) obtained for different number of training samples

Dataset	Method	50	100	200	400	800
Pavia University	NMKL[12]	91.52±0.98	92.72±0.33	–	–	–
	Deep CNN [21]	91.39±0.8	93.10±0.45	95.97±0.46	96.81±0.25	97.31±0.26
	DS-CNN and SincNet [23]	–	**–**	–	–	99.68±0.31
	Proposed method	**94.49±0.42**	**96.24±0.51**	**98.24±0.32**	**99.36±0.11**	**99.72±0.23**
Salinas	NMKL[12]	89.33±0.44	90.60±0.43	–	–	–
	Deep CNN [21]	91.36±1.11	93.15±0.43	95.07±0.23	96.55±0.29	97.14±0.53
	Proposed method	**95.94±0.14**	**97.15±0.35**	**99.12±0.21**	**99.39±0.32**	**99.59±0.19**

The bold items specify the maximum values for each dataset

Table 5 Performance comparison for Pavia University dataset

Training samples per class	Tr = 45			Tr = 55			Tr = 65		
	OA	AA	Kappa	OA	AA	Kappa	OA	AA	Kappa
DHCNET [35]	97.37	97.70	96.52	97.95	98.03	97.28	98.44	98.54	97.93
Proposed method	**98.26**	**98.48**	**96.75**	**98.92**	**98.97**	**97.81**	**99.42**	**99**	**98.22**

The bold items specify the maximum values for each dataset

Table 6 Individual class accuracies obtained for Indiana pines dataset with 10% training

Method Class	SMBN [8]	Proposed method
1	92.68	100
2	98.83	98.52
3	99.87	100
4	100	96.23
5	97.47	99.40
6	100	99.54
7	100	100
8	100	99.68
9	96.23	93.35
10	98.52	97.22
11	100	98.94
12	99.36	97.51
13	99.58	99.68
14	100	99.15
15	99.68	100
16	92.31	97.52
OA	98.26	98.30
AA	97.49	98.55
KAPPA	98.96	98.10

from data in an unsupervised manner. Multiple morphological attribute filters are applied to these feature maps to exploit spatial information. Therefore, we extract joint spectral-spatial features in a compact way instead of using multiple kernels corresponding to each modality and stacking them to make a large feature vector. Furthermore, SVM classifier is utilized which leads to accurate and stable results for HSI data. This reduces the computational burden significantly compared to deep neural network-based classification frameworks.

Data availability The datasets analysed during the current study are available in the following repository: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_ScenesReferences

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Binol, H (2018) Ensemble learning based multiple kernel principal component analysis for dimensionality reduction and classification of hyperspectral imagery. Math Probl Eng https://doi.org/10.1155/2018/9632569
2. Bioucas-Dias, JM, Plaza, A, Dobigeon, N, Parente, M, Du, Q, Gader, P, Chanussot, J (2012) Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. IEEE J Select Top Appl Earth Observ Remote Sens https://doi.org/10.1109/JSTARS.2012.2194696
3. Bradley, PS, Mangasarian, OL (2000) K-Plane Clustering. J Glob Optim https://doi.org/10.1023/A:1008324625522
4. Cai, W, Chen, S, Zhang, D (2007) Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern Recogn https://doi.org/10.1016/j.patcog.2006.07.011
5. Camps-Valls, G, Bruzzone, L (2005) Kernel-based methods for hyperspectral image classification. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2005.846154
6. Dalla Mura, M, Benediktsson, JA, Waske, B, Bruzzone, L (2010a) Extended profiles with morphological attribute filters for the analysis of hyperspectral data. Int J Remote Sens https://doi.org/10.1080/01431161.2010.512425
7. Dalla Mura, M, Benediktsson, JA, Waske, B, Bruzzone, L (2010b) Morphological attribute profiles for the analysis of very high resolution images. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2010.2048116
8. Fang L, Liu G, Li S, Ghamisi P, Benediktsson JA (2018) Hyperspectral image classification with squeeze multibias network. IEEE Trans Geosci Remote Sens 57(3):1291–1301
9. Fauvel, M, Chanussot, J, Benediktsson, JA (2006) Kernel principal component analysis for feature reduction in hyperspectrale images analysis. Proceedings of the 7th Nordic signal processing symposium, NORSIG 2006. https://doi.org/10.1109/NORSIG.2006.275232
10. Fauvel, M, Benediktsson, JA, Chanussot, J, Sveinsson, JR (2008) Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2008.922034
11. Fauvel, M, Chanussot, J, Benediktsson, JA (2009). Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. Eurasip J Adv Signal Process https://doi.org/10.1155/2009/783194
12. Gu, Y, Liu, T, Jia, X, Benediktsson, JA, Chanussot, J (2016) Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2015.2514161
13. Gu, Y, Chanussot, J, Jia, X, Benediktsson, JA (2017) Multiple kernel learning for hyperspectral image classification: a review. In IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2017.2729882
14. He, Z, Liu, L, Deng, R, Shen, Y (2016) Low-rank group inspired dictionary learning for hyperspectral image classification. Signal Process https://doi.org/10.1016/j.sigpro.2015.09.004

15. He, Z, Hu, J, Wang, Y (2018) Low-rank tensor learning for classification of hyperspectral image with limited labeled samples. Signal Process https://doi.org/10.1016/j.sigpro.2017.11.007

16. Iordache, MD, Bioucas-Dias, JM, Plaza, A (2011) Sparse unmixing of hyperspectral data. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2010.2098413

17. Jia S, Jiang S, Lin Z, Li N, Xu M, Yu S (2021) A survey: deep learning for hyperspectral image classification with few labeled samples. Neurocomputing 448:179–204

18. Keshava, N, Mustard, JF (2002) Spectral unmixing. IEEE Signal Process Mag https://doi.org/10.1109/79.974727

19. Kuo, BC, Ho, HH, Li, CH, Hung, CC, Taur, JS (2014) A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. IEEE J Select Top Appl Earth Observ Remote Sens https://doi.org/10.1109/JSTARS.2013.2262926

20. Leahy, R (1993) An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. IEEE Trans Pattern Anal Mach Intell https://doi.org/10.1109/34.244673

21. Lee, Hyungtae, and Heesung Kwon (2017) Going deeper with contextual CNN for hyperspectral image classification. IEEE Transactions on Image Processing 26(10):4843–4855

22. Li, J, Marpu, PR, Plaza, A, Bioucas-Dias, JM, Benediktsson, JA (2013) Generalized composite kernel framework for hyperspectral image classification. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2012.2230268

23. Li, Yuan, Qizhi Xu, Wei Li, and Jinyan Nie (2020) Automatic clustering-based two-branch CNN for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 59(9):7803–7816

24. Manolakis, D, Lockwood, R, Cooley, T (2016) Hyperspectral imaging remote Sensing_Physics, sensors, and algorithms. Hyperspectral Imaging Remote Sens

25. Melgani, F, Bruzzone, L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2004.831865

26. Mirzaei, S (2019) Hyperspectral image classification using non-negative tensor factorization and multinomial logistic regression. J Appl Remote Sens https://doi.org/10.1117/1.jrs.13.026501

27. Mirzaei, S, Van Hamme, H, Khosravani, S (2019) Hyperspectral image classification using non-negative tensor factorization and 3D convolutional neural networks. Signal Process Image Commun https://doi.org/10.1016/j.image.2019.05.004

28. Nascimento, JMP, Dias, JMB (2005) Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2005.844293

29. Pan, B, Shi, Z, Xu, X (2018) MugNet: deep learning for hyperspectral image classification using limited samples. ISPRS J Photogramm Remote Sens https://doi.org/10.1016/j.isprsjprs.2017.11.003

30. Pande S, Banerjee B (2022) HyperLoopNet: hyperspectral image classification using multiscale self-looping convolutional networks. ISPRS J Photogramm Remote Sens 183:422–438

31. Richards, JA, Jia, X (1999) Remote Sensing Digital Image Analysis. Remote Sens Digit Image Anal https://doi.org/10.1007/978-3-662-03978-6

32. Schölkopf, B, Smola, A, Müller, KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput https://doi.org/10.1162/089976698300017467

33. Yang, Zhi-Min, Yan-Ru Guo, Chun-Na Li, and Yuan-Hai Shao (2015) Local k-proximal plane clustering. Neural Computing and Applications 26:199–211

34. Zhong, Z, Li, J, Luo, Z, Chapman, M (2018) Spectral-spatial residual network for hyperspectral image classification: a 3-D deep learning framework. IEEE Trans Geosci Remote Sens https://doi.org/10.1109/TGRS.2017.2755542

35. Zhu J, Fang L, Ghamisi P (2018) Deformable convolutional neural networks for hyperspectral image classification. IEEE Geosci Remote Sens Lett 15(8):1254–1258