Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 4824 – 4838

doi:10.3906/elk-1412-29

Integration of spectral and spatial information via local covariance matrices for segmentation and classification of hyperspectral images

Uğur ERGÜL, Gökhan BİLGİN*
Department of Computer Engineering, Yıldız Technical University, İstanbul, Turkey

Received: 05.12.2014 • **Accepted/Published Online:** 17.09.2015 • **Final Version:** 06.12.2016

Abstract: In this work, a novel approach is presented for the feature extraction step in hyperspectral image processing to form more discriminative features between different pixel regions. The proposed method combines both spatial and spectral information, which is very important for segmentation and classification of hyperspectral images. For comparison, five different feature sets are formed using eigen decomposition of local covariance matrices of subcubes located around a pixel of interest in the scene. Subcubes of neighbor pixels are obtained by a windowed structure to expose pattern similarities. As a novel approach, local covariance matrices are computed in eigenspace and proposed feature sets are created after this stage. Before the formation of feature sets in eigenspace, the original input space is transferred to eigenspace by linear and nonlinear manner by principal component analysis (PCA) and its kernelized version (KPCA) and they are used in the experiments comparatively. In the simulations, one hyperspectral scene with ground-truth and one without ground-truth are used for the segmentation and classification tasks. Results of experiments are evaluated with four different unsupervised learning algorithms for data without ground-truth and three different supervised learning algorithms for data with ground-truth comparatively.

Key words: Hyperspectral imaging, remote sensing, segmentation, classification, covariance matrices, spectral spatial information

1. Introduction

Hyperspectral remote sensing technology has a wide range of application fields from geoscience, chemistry, forensic medicine, and military applications to space/planet research. Taking hundreds of contiguous and narrow spectral bands from the visible to the infrared bands in the electromagnetic spectrum into account is the main advantage of hyperspectral imaging [1]. Basically, hyperspectral imaging systems use measurements obtained from special hyperspectral sensors to determine the contribution of different materials appear in the imaging field with high spectral and spatial resolution.

Hyperspectral image pixels consist of brightness levels of multiple bands, which are also called spectral signatures. The spectral patterns of these signatures can be used in both classification and clustering tasks. Unsupervised classification or clustering/segmentation of hyperspectral images enables easier analysis of high-dimensional data [2]. However, it must be considered that spectral signatures are also spatially coherent. Making use of spatial information naturally improves the discrimination of pixels of interest.

Supervised classification algorithms need labeled samples (ground-truth, in remote sensing terminology) for building a model in the training stage. The number and quality of these samples play an important

*Correspondence: gbilgin@yildiz.edu.tr

role in getting good classification results. However, there are a limited number of labeled hyperspectral data sets available to all researchers that can be used for the comparison of supervised learning based studies in the literature. On the other hand, unsupervised classification algorithms do not need labeled samples. Basically, unsupervised classification algorithms use the similarity of samples to form distinct clusters as much as possible. Unsupervised and supervised classification of hyperspectral images is studied in the remote sensing literature with increasing attention considering its high dimensional nature. Both supervised and unsupervised classification algorithms suffer from high dimensionality of hyperspectral data. In the application of classification and clustering algorithms, the challenge of dimensionality must be considered for accuracy and complexity [3].

The task of clustering is especially important in the absence of ground-truth information. Collection of ground-truth information is expensive and time consuming. Hence, the realization of clustering/segmentation tasks without label information brings important advantages in the analysis of hyperspectral images. A modified k-means reclustering algorithm has been introduced for unsupervised classification of hyperspectral images [4]. In another work, a novel fuzzy c-means method using weighted cluster centers for unsupervised hyperspectral image classification was used [5]. A neuro-fuzzy approach based on weighted incremental neural networks for unsupervised hyperspectral image segmentation has been introduced in [6]. Unsupervised segmentation of hyperspectral images using region growing type segmentation with modified phase correlation of spectral signatures has been proposed in [7]. A multiobjective particle swarm optimization based clustering approach has been proposed with different types of optimization criteria [8].

Supervised algorithms analyze the training data using ground-truth information to build a learning model and this model can be used for mapping new samples in the scene. Kernel based classification algorithms such as support vector machines (SVMs) are very popular in classification tasks in hyperspectral images [9]. Radial basis functions, linear discriminant analysis, and SVMs have been combined for obtaining better classification results in [10]. A relevance vector machine (RVM) has been used to classify hyperspectral image data and the results have been compared with SVMs in [11]. Composite kernels have been studied for enhancing hyperspectral image classification in [12]. Kernel sparse representation based models [13], semisupervised graph-based approaches [14], semisupervised neural networks [15], and manifold learning-based studies [16] have been proposed for efficient hyperspectral image classification in recent years.

Feature extraction is an essential step for reducing computational complexity and improving the separability of samples in hyperspectral image processing. Several linear and nonlinear studies have been presented for this purpose in the literature. A wavelet and Fourier transformation based combined feature extraction has been proposed for enhancement of spectral signal quality and extraction of hidden features from hyperspectral images in [17]. The dyadic discrete wavelet transform has been introduced for feature extraction for dimension reduction [18]. A learning vector quantization based neural paradigm of generalized relevance learning vector quantization method has been developed for feature extraction to improve classification performance while keeping meaningful features [19]. Another proposed approach was a tensor extension of conventional supervised manifold-learning-based dimension reduction of hyperspectral images [20]. An innovative spectral feature extraction method called prototype space feature extraction (PSFE) was addressed with using only class spectral information in [21].

As well as spectral features, spatial information provides an important information source for classification and clustering tasks. There are many studies published about utilization of contextual or spatial information in the literature. In [22], integration of spectral and spatial information has been studied using morphological information and the original hyperspectral data together. In a similar work, spatial reclassification and

mathematical morphological approaches have been investigated for accurate and precise classification results in [23]. Results of a pixelwise SVM classification and the segmentation map, which were obtained by partitional clustering method, have been combined using a majority voting scheme in [24]. An agglomerative hierarchical clustering method was introduced in [25], which uses both spectral and spatial information for the aggregation decision. A Bayesian framework has been proposed using a multinomial logistic regression algorithm for spectral information and a multilevel logistic Markov–Gibbs random field for contextual information in [26]. Spectro-spatial features have been extracted using Gabor textural filters for a SVM based classification in [27].

In this paper, we mainly propose a combined approach for using both spatial and spectral information together. In line with this objective, local covariance matrices in eigenspace have been used to improve segmentation and classification tasks of hyperspectral images. In our first preliminary study [28], we introduced the use of local covariance matrices for integration of spectral and spatial information on the original input/sample space. Then we found that the local covariance matrices yield better results in eigenspace when transformed by principal component analysis (PCA) and kernel principal component analysis (KPCA) rather than original spectral signatures in input space. With adding the 'computation of local covariances in eigenspace' idea in the second preliminary study [29], segmentation of hyperspectral images has been addressed by using several clustering methods. In this manuscript, the previous studies are extended to prove the capability of the proposed approach using both supervised and unsupervised learning (classification and clustering) results in experiments with different window sizes, which are used for the handling of spatial relationships.

In accordance with this purpose, two hyperspectral images are utilized for the evaluation of the proposed feature extraction approaches in both the segmentation and classification tasks together. The first hyperspectral scene is a part of an AVIRIS image from NASA, which was captured at a cuprite mining site in Nevada, USA, and this scene is used in the evaluation of unsupervised segmentation experiments. The second scene is from a reflective optics system imaging spectrometer (ROSIS) hyperspectral image from the University of Pavia, Italy, and the scene is used in the evaluation of supervised classification results.

It must be noted that PCA and KPCA are exploited for two reasons: a) for transferring the original input space to eigenspace in linear (with PCA) and nonlinear (with KPCA) manner, and they are used in the experiments comparatively; b) for reducing computational complexity 'in the computation of local covariances' using different dimensions of transformed eigenspace (from dimension reduction perspective using 10, 20, and 30 bands/dimensions). Furthermore, different band sizes (10, 20, and 30 bands) and several windows sizes, such as 3×3, 5×5, 7×7, and 9×9 are utilized to investigate correlations of surrounding neighbor pixels within these windows. In addition, in the unsupervised segmentation part of this study, a more robust and more stable segmentation validity measure, namely power of spectral discrimination (PWSD), is utilized [30]. Furthermore, a well-known Haralick feature (HF) descriptor [31] method is adopted to integrate spectral and spatial features for comparison [32]. HFs are also extracted in the same window structure and size similar to the proposed method and compared in tables. In the unsupervised segmentation stage, K-means (KM), fuzzy C-means (FCM), Gustafson–Kessel clustering (GKC), and expectation maximization (EM) algorithms are used with unlabeled data. In the supervised classification of labeled data, SVMs, kernel discriminant analysis (KDA), and random forest (RF) methods are utilized for comparison.

This paper is organized as follows. Section 2 provides a brief introduction to the computation local covariance matrices in eigenspace and explains the proposed feature extraction approach in this paper. Experimental results are presented in Section 3. Finally, in Section 4, we conclude this paper with final remarks and observations on this study.

2. Spectro-spatial feature extraction in eigenspace

Integration of spectral and spatial information is a meaningful and useful strategy for accuracy increment in the classification and clustering of hyperspectral images. In this manuscript, the use of local covariance matrices for integration of spectral and spatial information on the eigenspace is introduced. Our previous preliminary studies [28,29] proved that the local covariance matrices yield better results in eigenspace when transformed by PCA and KPCA rather than original spectral signatures in input space.

2.1. Transferring hyperspectral data to eigenspace

As local covariance matrices yield better results in eigenspace, PCA and KPCA are mainly exploited for transferring original input space to eigenspace in linear (PCA) and nonlinear (KPCA) manner and they are used in the experiments comparatively. As well as the transferring issue, the advantage of dimension reduction possibilities with PCA [33] and KPCA [34] methods are taken into account for reducing computational complexity in the computation of local covariances. Because of the high dimensional nature of hyperspectral images, which contains quite rich information in bands, the computational cost also increases depending on the complexity of the machine learning algorithm in use. From the dimension reduction perspective, different dimensions of transformed eigenspace are assessed for comparison (i.e., 10, 20, and 30 bands/dimensions in the experiments).

2.2. Feature extraction with local covariance matrices

The proposed method with local covariance matrices enables the use of both spectral and spatial information in a pixel neighborhood. Furthermore, it is intended to utilize covariance matrices for feature extraction in eigenspace as a novel approach. Revealing the similarities between different image regions with using regional covariance matrices (RCMs) was first proposed by Tuzel at al. [35]. RCMs were generally used for the purpose of object recognition and texture classification. In previous works, RCMs were constructed using spatial location information, color values in each plane of RBG images, and first and second derivatives in the x and y directions of the images to find similar regions.

Hyperspectral images can be defined in two different ways: a) as a d-dimensional hypercube (height $\times$ width $\times d$), where d is the total number of bands in the hyperspectral image; b) if each pixel in a scene forms a vector as a line of a matrix, a two-dimensional matrix can be obtained ($\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}, \mathbf{x}_n \in \mathbf{R}^d$), where n is total number of pixels in the scene.

Unlike the original study, in our work each pixel (not only regions) is evaluated with its neighbor pixels using original spectral signature values. In this way, attempts are made to expose pattern similarities in pixel localization. For this purpose, $k \times k$ size (k denotes pixel length of one side of square window structure) windowed regions (i.e. B) are defined and $\mathbf{C}_B$ covariance matrices are calculated for each pixel $\left(\mathbf{x}_i \in \mathbf{R}^d\right)$, taking m neighbor pixels (xy) into account $(m = k \times k)$ as in (7).

$$\mathbf{C}_B = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}_i - \mu_B)(\mathbf{x}_i - \mu_B)^T, \tag{1}$$

where μ_B represents mean spectral signature calculated from pixels in the region B. In (8), $\mathbf{v}_B$ represents eigenvectors, while λ_B represents eigenvalues that indicate the importance of eigenvectors with respect to variance.

$$\mathbf{C}_B \mathbf{v}_B = \lambda_B \mathbf{v}_B \tag{2}$$

The obtained $d \times d$ covariance matrices provide discriminative features that can separate the pixel distribution in B region from other pixel regions. By using this information, proposed feature sets are extracted for each pixel to form new transformed spectral signatures. Eigenvalues and eigenvectors obtained from local covariance matrices for each pixel are sorted in descending order and used for extraction of different feature sets. Local covariance based feature extraction process steps are shown in Figure 1.

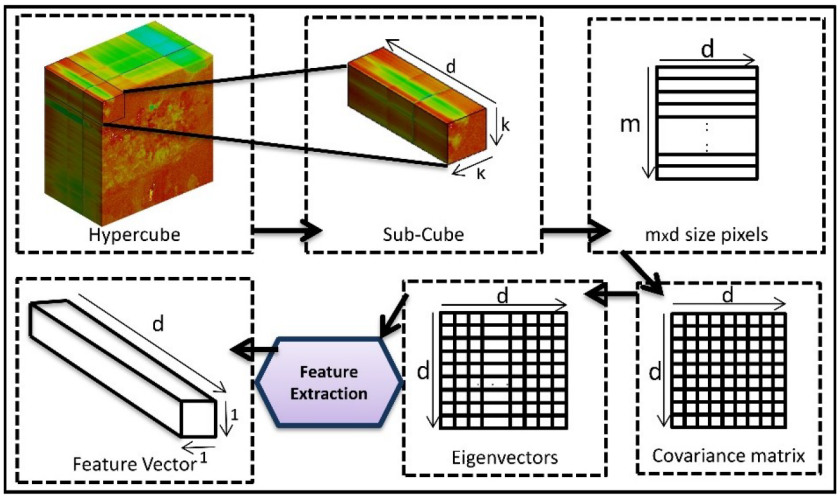


Figure 1. Local covariance based feature extraction process steps.

The proposed feature sets in this study can be given as follows:

2.2.1. Feature set (FS-1)

Only the first eigenvector corresponding to the biggest eigenvalue (i.e. $\mathbf{v}_B$) is chosen as new features for the pixel of interest (i.e. region B).

$$\mathbf{feat}_{(x,y)}=\mathbf{v}_B; for \lambda_{max} \tag{3}$$

2.2.2. Feature set (FS-2)

A new feature vector is formed for the pixel of interest by summing of all eigenvectors ($d^{'} < d$) at least 0.90% confidence of corresponding eigenvalues and proportional to eigenvalues in the confidence range.

$$\mathbf{feat}_{(x,y)}=\sum_{i=1}^{d'}\left(\frac{\lambda_B(i)}{\sum\lambda_B}\cdot\mathbf{v}_B(i)\right); \frac{\sum_{i=1}^{d'}\lambda_B(i)}{\sum\lambda_B}\geq 0.90 \tag{4}$$

2.2.3. Feature set (FS-3)

This feature set is formed similar to the previous feature set (i.e. FS-2) but it comprises at least 0.95% confidence of corresponding eigenvalues.

$$\mathbf{feat}_{(x,y)}=\sum_{i=1}^{d'}\left(\frac{\lambda_B(i)}{\sum\lambda_B}\cdot\mathbf{v}_B(i)\right); \frac{\sum_{i=1}^{d'}\lambda_B(i)}{\sum\lambda_B}\geq 0.95 \tag{5}$$

2.2.4. Feature set (FS-4)

Vectors of new feature space are formed by summing all eigenvectors proportional with respect to the corresponding eigenvalue.

$$\mathbf{feat}_{(x,y)}=\sum_{i=1}^{d}\left(\left(\lambda_B(i)/\sum_{j=1}^{d}\lambda_B\left(j\right)\right)\cdot\mathbf{v}_B(i)\right) \tag{6}$$

2.2.5. Feature set (FS-5)

All eigenvalues λ_B obtained in a surrounding windowed structure of a pixel constitute new features of the pixel of interest.

$$\mathbf{feat}_{(x,y)}=\lambda_B \tag{7}$$

Feature vectors $\mathbf{feat}_{(x,y)}$, which are obtained using proposed five feature extraction methods, are computed for each pixel in the scene. Covariance matrices combine multiple correlated features in a natural way. Noisy samples are also filtered out with the averaging operation by computation of covariance matrices [35]. It must be noted that, in small window sizes, the local covariance matrix may not be full rank. Hence, the number of eigenvectors corresponding to nonzero eigenvalues cannot be more than the number of pixels in the windowed area. Nevertheless, there is a low probability to come across an almost identical spectral pixel signature in a hyperspectral scene because of the large pixel coverage area in low resolutional hyperspectral sensors.

3. Experiments and results

In experiments, a part of the AVIRIS Cuprite-Nevada hyperspectral data, which will be called AVIRIS S4, and the hyperspectral data of the ROSIS University area with provided ground-truth are used. The AVIRIS Cuprite-Nevada hyperspectral data are provided by NASA and contain 224 spectral bands originally. After discarding noisy bands caused by atmospheric effects, the band number is reduced to 198. The scene consists of 256 rows/scene and 256 pixels/row and it does not contain any ground-truth information. Experimental results are evaluated for 9, 10, 11, and 12 clusters by utilizing a spectral signature library, provided by the US Geological Survey [36]. In Figure 2a, the 50th band of AVIRIS S4 hyperspectral image is shown.

(a) (b)

Figure 2. (a) 50th Band of the AVIRIS Cuprite S4 hyperspectral cube and (b) 75th rotated band of the ROSIS Pavia University hyperspectral cube used in the experiments.

The PAVIA University hyperspectral image is obtained over the University of Pavia, Italy, by the Deutschen Zentrum für Luft- und Raumfahrt (DLR, the German Aerospace Agency). It has 115 bands with

a spectral coverage ranging from 430 to 860 nm. The ROSIS Pavia University hyperspectral scene has 610 rows/scene and 340 pixels/row. The original band number is reduced to 103 bands after removing water absorption affected and excessively noisy bands. Nine classes are defined for the ROSIS Pavia University scene: tree, asphalt, bitumen, gravel, metal sheet, shadow, bricks, meadow, and soil. There are 3921 training and 42,776 testing pixels available for this hyperspectral scene [37]. In Figure 2b, the 75th band of the ROSIS Pavia University hyperspectral image is shown as an example.

It is necessary to determine an objective and statistical criterion for evaluating the performance of unsupervised segmentation results and for comparing accuracies of different segmentation maps. For this purpose, power of spectral discrimination (PWSD) is utilized, which was proposed in [30]. PWSD provides an assessment criterion depending on two reference cluster centers for a particular pixel. PWSD is shown in (14), where $\mathbf{x}$ is the vector of a pixel in the scene and $\mathbf{s}_i$ and $\mathbf{s}_j$ are cluster representative vectors obtained by averaging for each cluster.

$$\Omega\left(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}\right) = max\left\{\frac{SAM(\mathbf{s}_i, \mathbf{x})}{SAM(\mathbf{s}_j, \mathbf{x})}, \frac{SAM(\mathbf{s}_j, \mathbf{x})}{SAM(\mathbf{s}_i, \mathbf{x})}\right\} \tag{8}$$

SAM (spectral angular measure) represents a similarity measurement criterion.

$$SAM\left(\mathbf{x}, \mathbf{y}\right) = 1 - cos^{-1}\left[\frac{\sum_{i=i}^{n} \mathbf{x}_i \mathbf{y}_i}{\sqrt{\sum_{i=1}^{n} \mathbf{x}_i^2}\sqrt{\sum_{i=1}^{n} \mathbf{y}_i^2}}\right] \tag{9}$$

For every pixel $\mathbf{x}_i$ in the *ith* cluster, the PWSD is computed using the cluster representative $\mathbf{s}_i$ that the pixel belongs to and the representative signatures of other clusters $\mathbf{s}_j$, $(i \neq j)$. Segmentation accuracy (SA) is the mean of Ω values for each pixel as shown in (16), where $c$is total number of clusters.

$$SA\left(\mathbf{x}_i\right) = mean\left\{\Omega\left(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}\right) | i,j = 1, \ldots, c, i \neq j\right\} \tag{10}$$

By definition, the PWSD value is always bigger than one. It is assumed that the discrimination capability of the segmentation technique increases with the increasing value of the SA criterion.

In the pattern recognition literature, there are some spatial feature extraction methods that aim to reveal spatial similarities/dependencies between distant or local neighbor regions. In this study, the well-known Haralick feature (HF) [31] descriptor method is adopted to integrate spectral and spatial features for comparison. The Haralick method uses a gray-level co-occurrence matrix (GLCM) for feature extraction. As a part of this study, GLCM matrices are formed with degrees of rotational direction of 0, 45, 90, and 135 for each band of the hyperspectral scene in a local window structure. In this way, a compatible comparison can be realized with the proposed local covariance feature extraction method. After obtaining GLCM matrices, a Haralick feature descriptor is applied to compute contrast, entropy, standard deviation end energy values as features. HFs are extracted from four different window sizes, which are defined as $k \times k(k = 3, 5, 7, 9)$ surrounding the neighborhood of the center pixel. Hence, the Haralick based local feature extraction method is compared with the proposed approach in terms of the same local spatial relations. HFs are extracted in the same window size similar to the proposed method and the results are presented in the tables.

In order to get comparative results, KM, FCM, GKC, and EM algorithms are used for unsupervised segmentation of hyperspectral data with no ground-truth. SVM, KDA, and RF methods are selected as

supervised classification methods for hyperspectral data with ground-truth. In all unsupervised algorithms, the error criterion is set to 0.0001 for termination condition and the weighting exponent of (m_f) fuzzy clustering algorithms is set to 2. The radial basis kernel is chosen with SVM and KDA. The best parameter optimization of kernel classifiers is accomplished by applying a 5-fold cross validation method. RF is an ensemble supervised classifier method that consists of multidecision trees. The maximum number of trees is set to 100 in the experiments.

Without applying any local covariance based feature extraction, evaluations of experimental results are obtained after clustering and classification processes as shown in Table 1. Original and dimension reduced bands with PCA and KPCA are presented for comparison. SA, namely mean of PWSD, measure is used for the unlabeled AVIRIS S4 hyperspectral scene and overall accuracy (OA) measure is used for the labeled ROSIS Pavia University hyperspectral scene to find segmentation and classification accuracies.

Table 1. AVIRIS Cuprite S4 hyperspectral scene SA results and ROSIS Pavia University hyperspectral scene OA results for original bands and 10, 20, 30, and 40 band sizes obtained by PCA and KPCA.

		AVIRIS Cuprite S4 SA results				ROSIS Pavia University OA results (%)		
		KM	FCM	GKC	EM	SVM	KDA	RF
With original bands		1.064	1.059	1.036	1.059	78.857	77.560	71.524
10 Band	PCA	1.025	1.033	1.015	1.056	73.992	72.017	74.041
	KPCA	1.066	1.057	1.083	1.064	72.660	74.348	66.480
20 Band	PCA	1.017	1.023	1.010	1.058	73.768	72.849	75.058
	KPCA	1.072	1.034	1.074	1.058	77.161	78.539	70.675
30 Band	PCA	1.013	1.011	1.010	1.054	73.908	73.050	74.840
	KPCA	1.073	1.028	1.070	1.052	80.538	80.938	72.936
40 Band	PCA	1.013	1.011	1.009	1.045	73.180	72.852	73.410
	KPCA	1.065	1.028	1.069	1.051	79.285	79.380	72.460

PCA and KPCA are mainly utilized for dimension reduction before the proposed novel feature extraction method with local covariance matrices. Experiments are performed on the hyperspectral scenes after dimension reduction with 10, 20, and 30 bands. After passing from original input space to the eigenspace, local covariance matrices are computed in a windowed structure. Several windows sizes, such as 3×3, 5×5, 7×7, and 9×9, are defined for considering different correlations of surrounding neighbor pixels within these windows. Afterwards, local covariance matrices are computed by taking all pixels within these windows and for every pixel in the scene. Then eigen decompositions of obtained covariance matrices are made. Several feature sets (FS-1 to FS-5) are composed using eigenvalues and eigenvectors as explained in Section 2.2.

The best SA results for the AVIRIS S4 hyperspectral scene are obtained by using 3×3 window size out of all other window sizes (i.e. 5×5, 7×7, 9×9) in the experiments. Thus, all results for 3×3 window size are presented in Table 2 and the best SA result is 1.417 for 10 bands using FS-1 in this table. In all reduced bands, KPCA gives the best SA results for KM, FCM, and GKC with FS-1. EM also produces higher results with KPCA but it produces the best results for different feature sets (e.g., FS-3 for 20 bands and HF for 30 bands). Furthermore, decreasing of SA values can be observed from Table 2 as the number of reduced bands increases for all segmentation methods. On the other hand, the best OA results for the ROSIS Pavia University hyperspectral scene are obtained with 5×5 window size out of all other window sizes (i.e. 3×3, 7×7, 9×9) in the experiments. Hence, all results for 5×5 window size are presented in Table 3 and the best OA is 90.73% for 30 bands using FS-1 in the table. It can be seen from the Table 3 that KPCA gives the best OA results for all classifiers but the best results for different reduced bands change arbitrarily with the feature set.

Table 2. AVIRIS Cuprite S4 hyperspectral scene SA results with 3 × 3 window size and 10, 20, and 30 bands using five different feature sets (FS-1 to FS-5) and Haralick features (HFs).

AVIRIS Cuprite S4		Segmentation methods							
		KM		FCM		GKC		EM	
		PCA	KPCA	PCA	KPCA	PCA	KPCA	PCA	KPCA
10 Band	FS-1	1.104	1.417	1.179	1.325	1.089	1.380	1.058	1.371
	FS-2	1.101	1.224	1.177	1.179	1.040	1.215	1.050	1.208
	FS-3	1.054	1.247	1.062	1.232	1.046	1.344	1.056	1.410
	FS-4	1.100	1.217	1.178	1.177	1.086	1.207	1.049	1.165
	FS-5	1.003	1.096	1.001	1.081	1.052	1.064	1.042	1.051
	HFs	1.032	1.118	1.027	1.068	1.052	1.185	1.056	1.213
20 Band	FS-1	1.101	1.280	1.67	1.218	1.040	1.253	1.061	1.236
	FS-2	1.101	1.200	1.166	1.160	1.038	1.188	1.058	1.190
	FS-3	1.063	1.145	1.070	1.150	1.041	1.202	1.055	1.264
	FS-4	1.091	1.189	1.166	1.167	1.096	1.198	1.057	1.159
	FS-5	1.002	1.026	1.001	1.050	1.047	1.057	1.040	1.048
	HFs	1.032	1.110	1.026	1.036	1.026	1.120	1.091	1.250
30 Band	FS-1	1.095	1.213	1.091	1.174	1.041	1.199	1.055	1.185
	FS-2	1.092	1.189	1.099	1.155	1.039	1.180	1.037	1.157
	FS-3	1.048	1.153	1.062	1.126	1.030	1.155	1.056	1.203
	FS-4	1.081	1.197	1.099	1.167	1.038	1.191	1.045	1.158
	FS-5	1.002	1.017	1.001	1.034	1.039	1.052	1.040	1.400
	HFs	1.033	1.071	1.028	1.049	1.011	1.103	1.095	1.221

Table 3. ROSIS Pavia University hyperspectral scene OA results in percentage with 5 × 5 window size and 10, 20, and 30 bands using five different feature sets (FS-1 to FS-5) and Haralick features (HFs).

ROSIS Pavia University		Classification methods					
		SVM		KDA		RF	
		PCA	KPCA	PCA	KPCA	PCA	KPCA
10 Band	FS-1	65.53	67.05	62.38	62.54	64.95	64.62
	FS-2	64.33	75.76	73.17	75.10	64.13	76.39
	FS-3	63.86	76.52	64.34	77.21	56.09	72.49
	FS-4	64.23	76.69	73.62	75.00	64.02	77.23
	FS-5	64.53	67.68	47.17	48.41	57.99	73.05
	HFs	54.40	59.07	56.49	60.52	41.26	58.04
20 Band	FS-1	70.08	79.22	72.81	75.15	73.09	73.94
	FS-2	74.46	79.21	72.91	77.54	71.70	77.15
	FS-3	73.49	77.20	71.87	77.02	67.35	75.55
	FS-4	71.92	77.17	73.06	77.18	72.59	76.56
	FS-5	60.18	61.21	51.62	53.40	51.54	71.82
	HFs	54.04	60.30	54.46	68.13	41.26	52.97
30 Band	FS-1	67.53	**90.73**	70.93	87.91	72.39	78.80
	FS-2	71.82	88.13	71.94	84.91	72.10	82.16
	FS-3	73.04	84.86	72.90	81.27	65.81	82.86
	FS-4	69.77	85.76	71.49	83.82	71.70	82.80
	FS-5	41.68	59.38	22.93	34.30	49.50	65.83
	HFs	54.40	67.11	55.45	68.64	41.26	71.81

In Tables 2 and 3, extracted features with the proposed method produce better accuracies if compared with Table 1. Note that FS-1 gives the best result in both segmentation and classification tasks. FS-1 is

formed by only the first eigenvectors corresponding to the biggest eigenvalue. It can be inferred that in FS-1 eigenvectors that correspond to axes with lower variance do not contribute to the composition of the feature set as in FS-2, FS-3, and FS-4.

Segmentation and classification maps could enable better visual assessment of obtained accuracies for all methods. The results without using the proposed feature sets with original data and also using the HF method are provided in Figures 3a–3d for visual comparison. Segmentation and classification maps of the two hyperspectral scenes using the proposed approach are shown in Figures 4a–4d and Figures 5a–5c, respectively. The best maps are generated with FS-1 for all segmentation and classification methods for comparison.

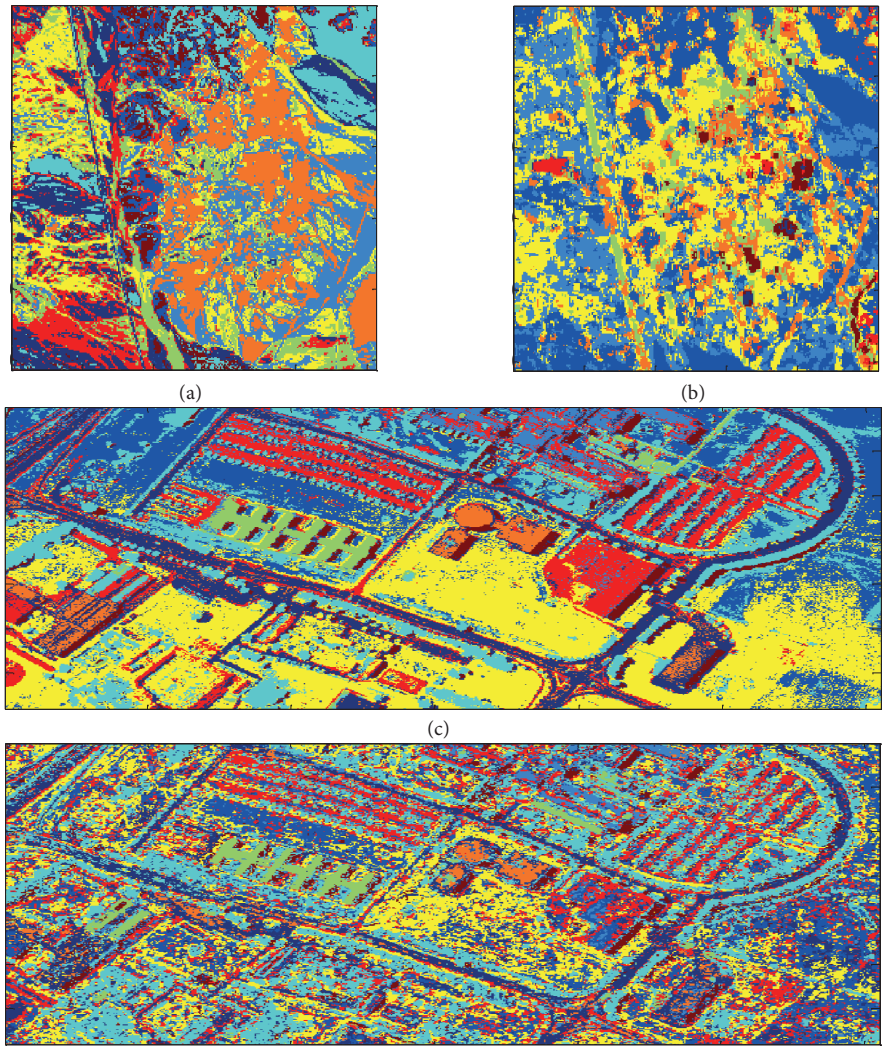
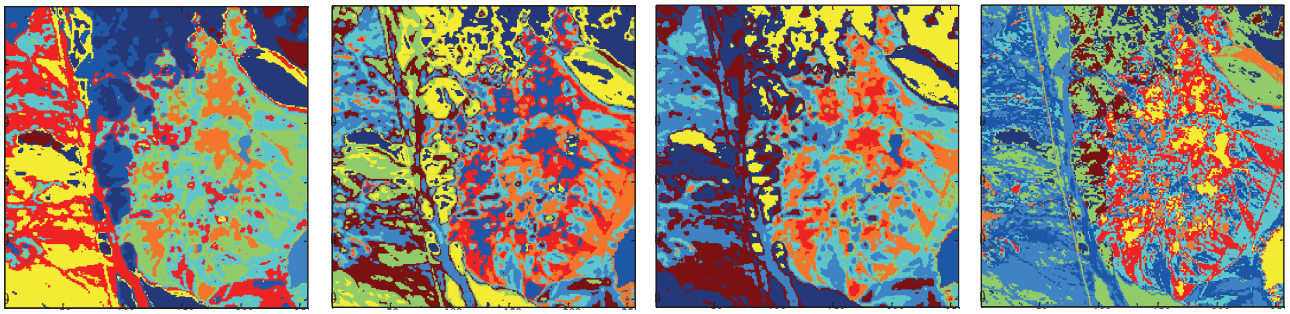


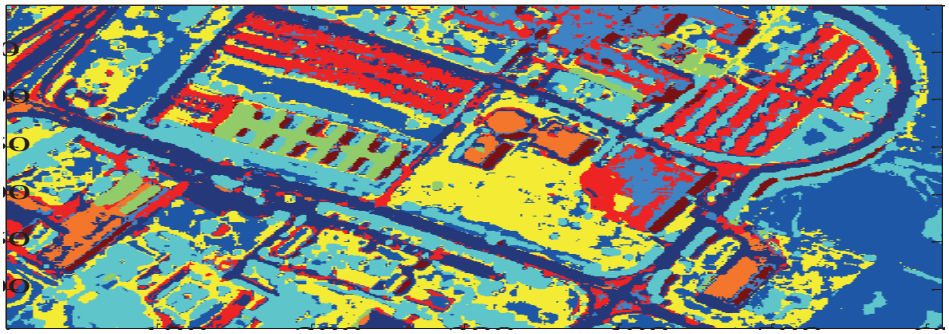
Figure 3. AVIRIS Cuprite S4 segmentation maps obtained by (a) original data and (b) HF with using KM. PAVIA Rosis University's rotated classification maps obtained by (c) original data and (d) HF with using SVM.

For comparison, AVIRIS Cuprite S4 hyperspectral scene SA results are provided with respect to window sizes for all bands using FS-1 with KM, FCM, GKC, and EM clustering methods in Figures 6a–6d and OA results of the ROSIS Pavia University hyperspectral scene are plotted with respect to 3×3, 5×5, 7×7, and 9×9 window sizes for all bands using FS-1 with SVM, KDE, and RF classification methods in Figures 6e–6g.

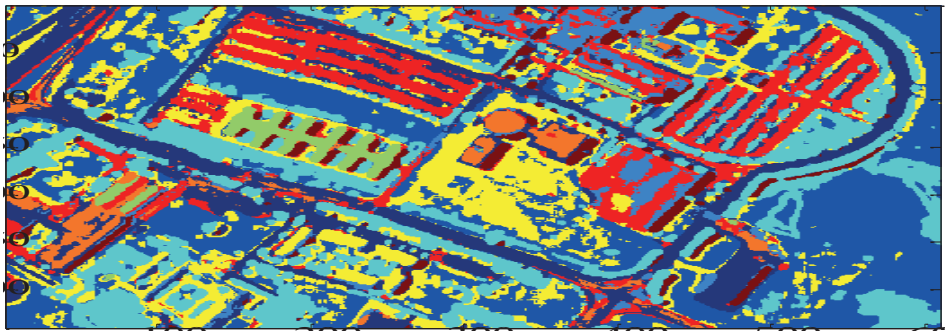


(a) (b) (c) (d)

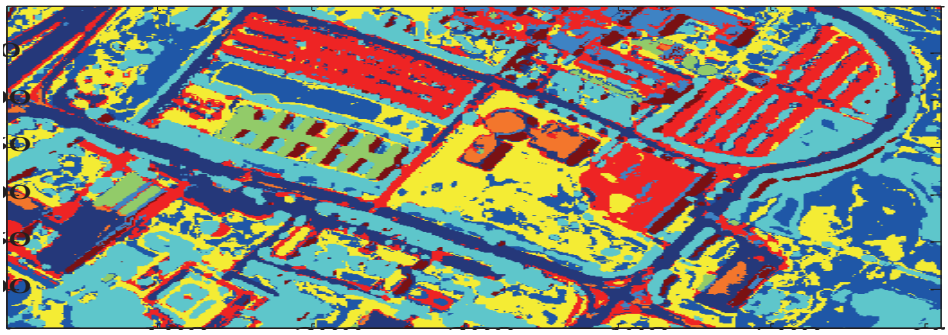
Figure 4. Segmentation maps obtained by 3 × 3 window size with FS-1 for (a) KM, (b) FCM, (c) GKC, (d) EM clustering methods.



(a)



(b)



(c)

Figure 5. Rotated classification maps obtained by 5 × 5 window size with FS-1 for (a) SVM, (b) KDA, (c) RF classification methods.

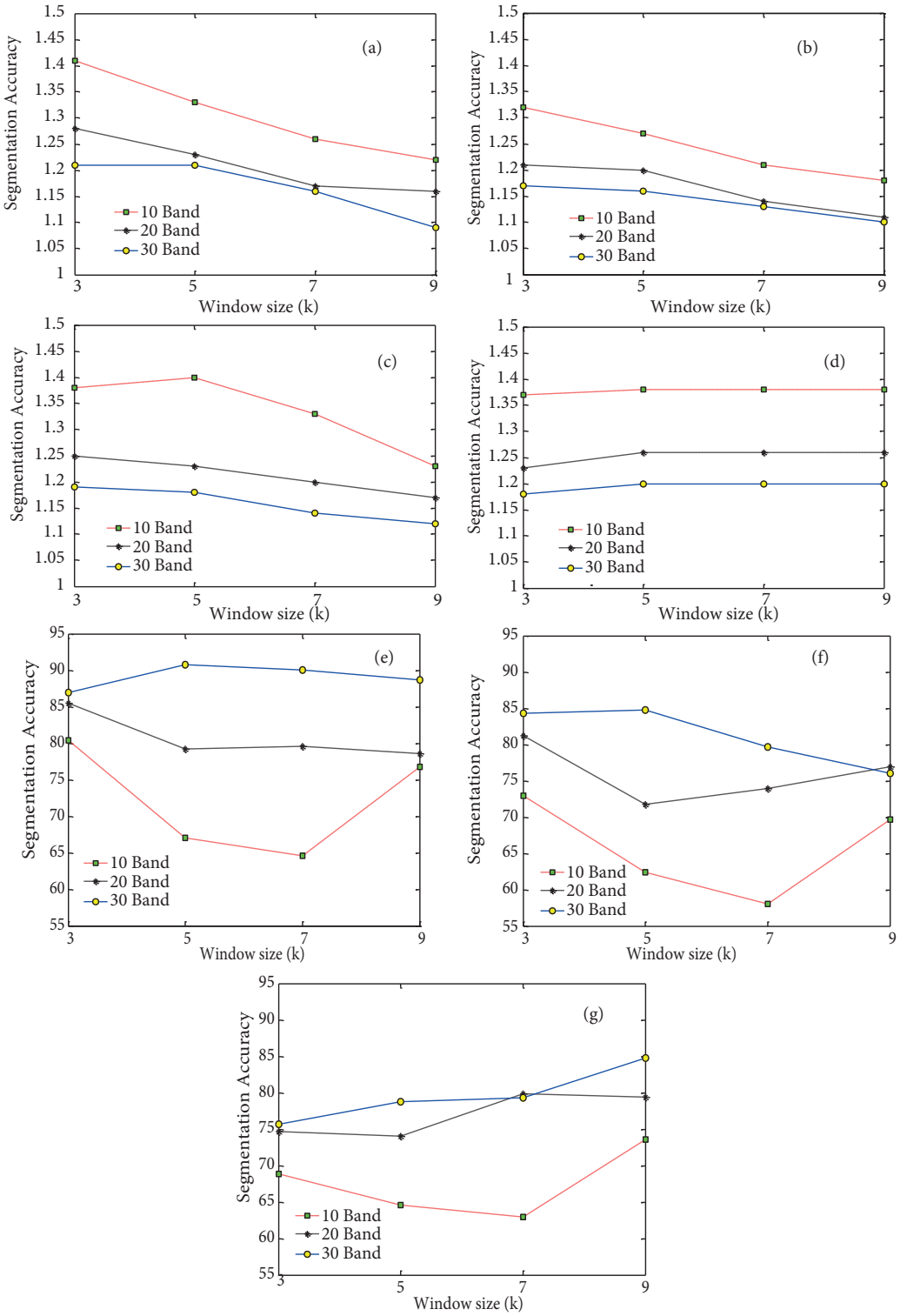


Figure 6. AVIRIS Cuprite S4 hyperspectral scene segmentation accuracy results (in terms of SA) with respect to 3 × 3, 5 × 5, 7 × 7, and 9 × 9 window sizes for 10, 20, and 30 bands with (a) KM, (b) FCM, (c) GKC, and (d) EM clustering methods using FS-1. ROSIS PAVIA University hyperspectral scene classification accuracy results (in terms of OA) with respect to 3 × 3, 5 × 5, 7 × 7, and 9 × 9 window sizes for 10, 20, and 30 bands with (e) SVM, (f) KDE, and (g) RF classification methods using FS-1.

The samples in a windowed area (which constitute a small subcube) have typically similar statistical distributions. Therefore, one variance value is generally calculated as the largest only for one dimension as a result of the computation of local covariance. In this case, variances of other dimensions will be relatively small or close to zero. This situation reveals that the data can be highly distinctive in only one dimensional space in most cases. Thus, only one eigenvector corresponding to the largest variance has high capacity of distinctiveness. Other eigenvectors that corresponds to relatively small or zero variances have low information on the local data distribution. In most cases, it can be inferred that these eigenvectors totally consist of noise and redundant information. Hence, taking these eigenvectors into account can reduce the separability of extracted features. The samples projected by a nonlinear transform like KPCA are expected to be more separable, when samples have nonlinear data distributions in the same windowed area particularly. In this case, the distribution of samples in a windowed area in feature space produces relatively few larger variances compared to linearly transformed samples.

4. Conclusion

In this paper, a novel feature extraction approach is proposed for utilization of both spectral and spatial information of hyperspectral images using local covariance matrices. Spatial information is an important source to achieve better segmentation and classification accuracies as well as spectral information. For this reason, local covariance matrices are computed with neighbor pixels in a windowed structure to expose pattern similarities in eigenspace. Experimental results show that integration of spectro-spatial features based on the proposed approach led to the improvement of accuracies in all unsupervised and supervised classification algorithms.

Acknowledgment

This research has been supported by Yıldız Technical University, Scientific Research Projects Coordination Department, Project Number: 2014–04-01–KAP01.

References

[1] Landgrebe DA. Signal Theory Methods in Multispectral Remote Sensing. Hoboken, NJ, USA: Wiley, 2003.

[2] Varshney PK, Arora MK. Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data. Berlin, Germany: Springer, 2004.

[3] Plaza A, Benediktsson JA, Boardman JW, Brazile J, Bruzzone L, Camps-Valls G, Chanussot J, Fauvel M, Gamba P, Gualtieri A et al. Recent advances in techniques for hyperspectral image processing. Remote Sens Environ 2012; 113: 110-122.

[4] Meyer AW, Paglieroni DW, Astaneh C. K-means re-clustering: algorithmic options with quantifiable performance comparisons. In: International Society for Optics and Photonics High-Power Lasers and Applications Conference; 25 Jan 2003; San Jose, CA, USA. Bellingham, WA, USA: SPIE. pp. 84-92.

[5] Kuo BC, Huang WC, Liu HC, Tseng SC. A novel fuzzy c-means method for hyperspectral image classification. In: IEEE International Geoscience and Remote Sensing Symposium; 7–11 July 2008; Boston, MA, USA. Hoboken, NJ, USA: IEEE. pp. 1002-1005.

[6] Muhammed HH. Unsupervised hyperspectral image segmentation using a new class of neuro-fuzzy systems based on weighted incremental neural networks. In: The 31st Applied Imagery Pattern Recognition Workshop; 16–17 October 2002; Washington DC, USA. Piscataway, NJ, USA: IEEE. pp. 171-177.

[7] Erturk A, Erturk S. Unsupervised segmentation of hyperspectral images using modified phase correlation. IEEE Geosci Remote S 2006; 3: 527-531.

[8] Paoli A, Melgani F, Pasolli E. Clustering of hyperspectral images based on multiobjective particle swarm optimization. IEEE T Geosci Remote 2009; 47: 4175-4188.

[9] Mercier G, Lennon M. Support vector machines for hyperspectral image classification with spectral-based kernels. In: IEEE International Geoscience and Remote Sensing Symposium; 21–25 July 2003; Toulouse, France. Hoboken, NJ, USA: IEEE. pp. 288-290.

[10] Camps-Valls G, Bruzzone L. Kernel Methods for Remote Sensing Data Analysis. New York, NY, USA: Wiley Online Library, 2009.

[11] Demir B, Erturk S. Hyperspectral image classification using relevance vector machines. IEEE Geosci Remote S 2007; 4: 586-590.

[12] Camps-Valls G, Gomez-Chova L, Muñoz-Mari J, Vila-Frances J, Calpe-Maravilla J. Composite kernels for hyperspectral image classification. IEEE Geosci Remote S 2006; 3: 93-97.

[13] Chen Y, Nasrabadi NM, Tran TD. Hyperspectral image classification via kernel sparse representation. IEEE T Geosci Remote S 2013; 51: 217-231.

[14] Camps-Valls G, Marsheva TB, Zhou D. Semi-supervised graph-based hyperspectral image classification. IEEE T Geosci Remote S 2007; 45: 3044-3054.

[15] Ratle F, Camps-Valls G, Weston J. Semisupervised neural networks for efficient hyperspectral image classification. IEEE T Geosci Remote S 2010; 48: 2271-2282.

[16] Ma L, Crawford MM, Tian J. Local manifold learning-based-nearest-neighbor for hyperspectral image classification. IEEE T Geosci Remote S 2010; 48: 4099-4109.

[17] Wang C, Gao RX. Wavelet transform with spectral post-processing for enhanced feature extraction-machine condition monitoring. IEEE Instru Meas Mag 2003; 52: 1296-1301.

[18] Bruce LM, Koger CH, Li J. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. IEEE T Geosci Remote 2002; 40: 2331-2338.

[19] Mendenhall MJ, Merényi E. Relevance-based feature extraction for hyperspectral images. IEEE T Neural Networ 2008; 19: 658-672.

[20] Zhang L, Zhang L, Tao D, Huang X. Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction. IEEE T Geosci Remote 2013; 51: 242-256.

[21] Mojaradi B, Abrishami-Moghaddam H, Zoej MJV, Duin RP. Dimensionality reduction of hyperspectral data via spectral feature extraction. IEEE T Geosci Remote 2009; 47: 2091-2105.

[22] Fauvel M, Benediktsson JA, Chanussot J, Sveinsson JR. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE T Geosci Remote 2008; 46: 3804-3814.

[23] Dell'Acqua F, Gamba P, Ferrari A, Palmason JA, Benediktsson JA, Arnason KAAK. Exploiting spectral and spatial information in hyperspectral urban data with high resolution. IEEE Geosci Remote S 2004; 1: 322-326.

[24] Tarabalka Y, Benediktsson JA, Chanussot J. Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. IEEE T Geosci Remote 2009; 47: 2973-2987.

[25] Marçal ARS, Castro L. Hierarchical clustering of multispectral images using combined spectral and spatial criteria. IEEE Geosci Remote S 2005; 2: 59-63.

[26] Li J, Bioucas-Dias JM, Plaza A. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. IEEE T Geosci Remote 2012; 50: 809-823.

[27] Huo LZ, Tang P. Spectral and spatial classification of hyperspectral data using SVMs and Gabor textures. In: IEEE International Geoscience and Remote Sensing Symposium; 24–29 July 2011; Vancouver, Canada. Hoboken, NJ, USA: IEEE. pp. 1708-1711.

[28] Bilgin G, Uslu E. Segmentation of hyperspectral images using local covariance matrices. In: 20th Signal Processing and Communications Applications Conference; 18–20 April 2012; Fethiye, Turkey. Piscataway, NJ, USA: IEEE. pp. 1-4.

[29] Ergul U, Bilgin G. Segmentation of hyperspectral images using local covariance matrices in eigenspace. In: 21st Signal Processing and Communications Applications Conference; 24–26 April 2013; Girne, Cyprus. Piscataway, NJ, USA: IEEE. pp. 1-4.

[30] Van der Meer F. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. Int J Appl Earth Obs 2006; 8: 3-17.

[31] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE T Syst Man Cyb 1973; 3: 610-621.

[32] Huang X, Zhang L. A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, Northern Italy. Int J Remote Sens 2009; 30: 3205-3221.

[33] Jimenez-Rodriguez OL, Arzuaga-Cruz E, Velez-Reyes M. Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data. IEEE T Geosci Remote 2007; 45: 469-483.

[34] Mathieu F, Chanussot J, Benediktsson JA. Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas. EURASIP J Adv Sig Pr 2009; 11: 1-14.

[35] Tuzel O, Porikli F, Meer P. Region covariance: A fast descriptor for detection and classification. In: European Conference on Computer Vision; 7–13 May 2006; Graz, Austria. Berlin, Germany: Springer. pp. 589-600.

[36] Swayze G, Clark R, Kruse F, Sutley S, Gallagher A. Groundtruthing AVIRIS mineral mapping at Cuprite, Nevada. In: Summaries of the Third Annual JPL Airborne Geoscience Workshop; 1 Jun 1992; Denver, CO, USA. pp. 47-49.

[37] Chi M, Kun Q, Benediktsson JA, Feng R. Ensemble classification algorithm for hyperspectral remote sensing data. IEEE T Geosci Remote 2009; 6: 762-766.