Segmentation of Hyperspectral Images via Subtractive Clustering and Cluster Validation Using One-Class Support Vector Machines

Gökhan Bilgin, Sarp Ertürk, *Member, IEEE*, and Tülay Yıldırım, *Member, IEEE*

***Abstract*—This paper presents an unsupervised hyperspectral image segmentation with a new subtractive-clustering-based similarity segmentation and a novel cluster validation method using one-class support vector (SV) machine (OC-SVM). An estimation of the correct number of clusters is an important task in hyperspectral image segmentation. The proposed cluster validity measure is based on the power of spectral discrimination (PWSD) measure and utilizes the advantage of the inherited cluster contour definition feature of OC-SVM. Hence, this novel cluster validity method is referred to as SV-PWSD. SVs found by OC-SVM are located at the minimum distance to the hyperplane in the feature space and at the arbitrarily shaped cluster contours in the input space. SV-PWSD guides the segmentation/clustering process to find the optimal number of clusters in hyperspectral data. Because of the high computational load of subtractive clustering and OC-SVM, a subset of the image (only ground-truth data) is initially used in the clustering and validation phases. Then, it is proposed to use K-nearest neighbor classification, with the already clustered subset being used as training data, to project the initial clustering results onto the entire data set.**

***Index Terms*—Hyperspectral images, one-class support vector (SV) machines (OC-SVMs), phase correlation, segmentation, subtractive clustering, unsupervised classification.**

I. Introduction

HYPERSPECTRAL imaging attracts increasing attention among remote sensing technologies due to the high amount of information that can be acquired. Hyperspectral sensors gather information from hundreds of contiguous narrow spectral bands from the visible to the infrared bands in the electromagnetic spectrum [1]. This paper is targeted at unsupervised classification or clustering/segmentation of hyperspectral data which enables easier analysis of the high-dimensional data, and a novel approach is presented for cluster validity of segmented hyperspectral images.

Manuscript received July 1, 2010; revised November 15, 2010 and January 10, 2011; accepted January 16, 2011. Date of publication March 28, 2011; date of current version July 22, 2011.

G. Bilgin is with the Department of Computer Engineering, Yildiz Technical University, Istanbul 34349, Turkey. He is also with the Department of Computer and Information Science, Indiana University–Purdue University Indianapolis, Indianapolis, IN 46202 USA (e-mail: gokhanb@ce.yildiz.edu.tr).

S. Ertürk is with the Department of Electronics and Telecommunication Engineering, Kocaeli University, Kocaeli 41380, Turkey (e-mail: sarp@ieee.org).

T. Yıldırım is with the Department of Electronics and Telecommunication Engineering, Yildiz Technical University, Istanbul 34349, Turkey (e-mail: tulay@yildiz.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2011.2113186

Subtractive clustering with Euclidean distance as well as spectral angle measure (SAM) [2], spectral information divergence (SID) [3], and phase-correlation-based [4] similarity measures have been utilized in this paper for clustering/segmentation purposes. In addition, k-means (KM) and fuzzy c-means (FCM) algorithms have been evaluated for comparison.

Unsupervised classification (or clustering) of pixels in hyperspectral images typically relies on the spectral data of each pixel. In the literature, an automated clustering approach for hyperspectral images based on histogram thresholding has been studied in [5], and another thresholding technique based on histograms of principal components has been introduced in [6]. A hyperspectral image segmentation approach based on Gaussian mixture models has been presented in [7], and clustering of hyperspectral images using non-Gaussian mixture models has been investigated in [8]. Hyperspectral image segmentation using a multicomponent Markov chain model has been presented in [9]. Unsupervised hyperspectral image segmentation using a neuro–fuzzy approach based on weighted incremental neural networks has been introduced in [10]. An algorithm referred to as independent component analysis mixture model (ICAMM) has been developed in [11], where class distributions are modeled with non-Gaussian densities using ICAMM that employs higher order statistics for unsupervised classification. A KM reclustering algorithm has been presented in [12] for unsupervised segmentation of hyperspectral images.

One of the main problems that can be encountered in the processing of hyperspectral data is the high data dimensionality. Therefore, usually dimension reduction is applied in order to remove redundant information and accelerate processing time. In this paper, an unsupervised band selection algorithm, namely, singular value decomposition (SVD) band selection (SVDBS) [13], [14] is used before segmentation to reduce the computational load. SVDBS is utilized to reduce data dimensionality for efficient processing because it can be applied to data that do not have any labeled training samples available. Furthermore, the original spectral information is kept in SVDBS by selecting only the most informative bands and removing the others, without any transformation.

Finding an optimal number of clusters in a hyperspectral image using an unsupervised algorithm is one of the important problems in cluster/segmentation analysis. A novel cluster validity measure referred to as support vector power of spectral discrimination (SV-PWSD) has been developed in this paper for this purpose. The proposed method is based on one-class

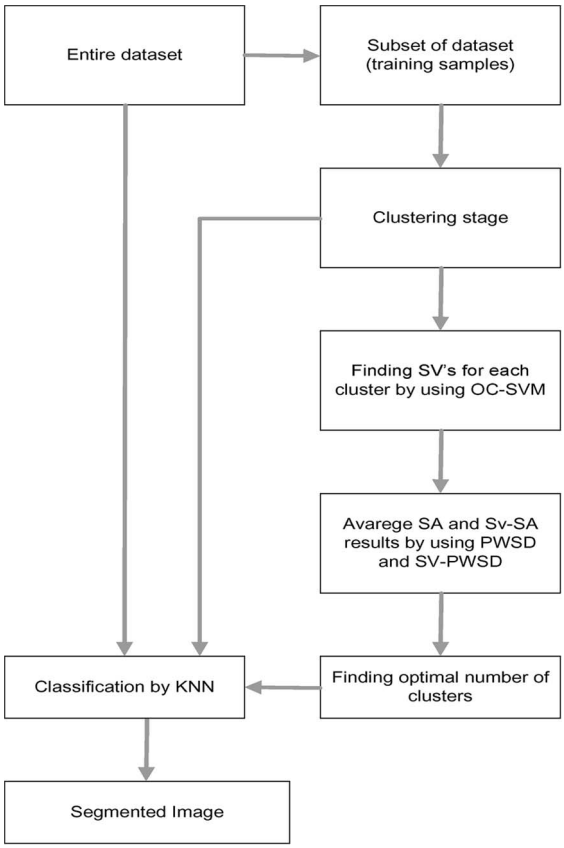

Fig. 1. Combined block diagram of cluster validation and segmentation stages.

SV machines (OC-SVMs) and gives a cluster validity measure by using SVs placed in the border of each cluster.

In the literature, there are a few studies dealing with SVs for cluster validity, but it has not been applied to hyperspectral images so far. A cluster validity measure with outlier detection and cluster merging algorithms for SV clustering has been introduced in [15]. The determination of the number of clusters in a data set via SV machines has been studied in [16]. Another cluster validity measure that evaluates the separation in clustering results using SVs has been presented in [17].

In brief, the presented approach consists of two stages. The first is the clustering and validation for finding the correct number of clusters from a subset of hyperspectral image. In the second stage, after obtaining the number of clusters for a scene, segmentation of the full image is accomplished using K-nearest neighbor (KNN) classification. In KNN, the samples obtained in the first stage with the appropriate number of clusters are used for training. Consequently, the test set of KNN is the entire data set to be segmented. Fig. 1 shows the cluster validation and segmentation stages in a combined block diagram.

This paper is organized as follows. Section II provides a brief introduction to SVDBS that is used as preprocessing step in this paper. An overview of subtractive-clustering-based segmentation is presented in Section III. In Section IV, OC-SVM is introduced, and the proposed cluster validity method referred to as SV-PWSD is widely explained. Experimental results are presented in Section V. Finally, in Section VI, we conclude this paper with final remarks and observations on this subject.

II. Band Selection for Dimension Reduction

Because of the high number of captured bands, hyperspectral images can contain redundant information [18]. Therefore, dimension reduction without losing important information can be possible. In contrast to transformation-based dimension reduction methods, original spectral data values are kept in band selection techniques by selecting only the most informative bands and removing the others. At the end of the band selection process, an m-dimensional original hyperspectral cube is reduced to d-dimensional space ($d < m$). In this paper, unsupervised SVDBS is utilized for band selection.

Hyperspectral images can be represented in a matrix form $\mathbf{X^T} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where n shows the number of pixels in the scene and $\mathbf{x} \in \mathbf{R}^d$ represents the spectral data. In the normalization phase, z-scores of spectral data are obtained. The z-score transform makes the data distribution zero mean with unit standard deviation using (1). Hence, the normalized spectral data become more comparable

$$\mathbf{z} = \frac{(\mathbf{x} - \mu_i)}{\sigma_i}. \tag{1}$$

Here, $\mathbf{x}$ shows a pixel in hyperspectral image; μ_i and σ_i show the mean and standard deviation of the ith $(i = 1, 2, \ldots, d)$ band, respectively, as in [13]. Then, the SVD of the $\mathbf{Z^T} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n]$ matrix is computed, and the number of selected bands can be determined as its rank or it can be set *a priori*. To obtain d selected bands, the rank revealing SVD decomposition or QR factorization can be used. QR factorization is a linear algebra operation that factors a matrix into orthogonal components. QR factorization is carried out with a pivoting matrix $(\mathbf{P})$, in the form of $\mathbf{V}^{\mathrm{T}}\mathbf{P} = \mathbf{QR}$, where $\mathbf{V}$ has the first d eigenvectors of $\mathbf{X}$. The matrices $\mathbf{Q}$ and $\mathbf{R}$ are the QR factors. After finding the pivoting matrix, $\mathbf{Y} = \mathbf{P}^{\mathrm{T}}\mathbf{X}$ is computed. The selected number of bands is obtained from the first d rows of the matrix $\mathbf{Y}$. Note that the transformations used in SVDBS are only utilized to select independent bands by selecting the most uncorrelated variables, and the bands themselves are not transformed. More information about this band selection technique can be found in [13] and [14].

III. Subtractive-Clustering-Based Segmentation

Subtractive clustering is an improved version of the mountain clustering method [19]. In mountain clustering, the data space is divided into small regions using a grid function, a potential function referred to as the mountain function is computed at each grid point, and grid points with higher values of this potential function are accepted as cluster centers. The mountain function is calculated as

$$M(\mathbf{v}_i) = \sum_{j=1}^{n} e^{-\alpha\|\mathbf{v}_i - \mathbf{x}_j\|^2} \tag{2}$$

where $\mathbf{v}_i$ denotes a grid point in the data space and α is a positive distance factor (referred to as radii) that determines the neighborhood of a cluster center. If a grid point has the highest mountain function value, this means that there are more data

points around this grid point than the other grid points, as normally dense regions in the data space can be accepted as cluster centers. The spatial resolution of the grid function affects the accuracy in the definition of the exact cluster centers. However, if the spatial resolution is increased to obtain finer grids, the cost of computation will increase exponentially, depending on the dimension of the data.

In subtractive clustering [20], it has been proposed to label each individual data point as a potential cluster center instead of using grid points. The potential function of each data point can be calculated similar to the mountain method as

$$M(\mathbf{x}_i) = \sum_{j=1}^{n} e^{-\alpha\|\mathbf{x}_i - \mathbf{x}_j\|^2}. \tag{3}$$

By using this potential function, the maximum value for all data points in the data space can be obtained in the form of

$$M_1^{\max} = \max_i \left(M(\mathbf{x}_i)\right). \tag{4}$$

Here, $M_1^{\max}$ shows the potential value of the first cluster center for the data point $\mathbf{x}_i^{\max}$. The density of data points surrounding $\mathbf{x}_i^{\max}$ is high, and these points have also high potential values. Hence, in the next step, one of these surrounding points could be obtained as the next cluster center. In order to eliminate the effect of high-density areas, all data points in the vicinity of the obtained cluster center must be removed, so that the next cluster center can be found in the next high-density region of the data space. The next cluster center $\hat{M}^j(\mathbf{x}_i)$ can be obtained in the form of

$$\hat{M}^j(\mathbf{x}_i) = \hat{M}^{j-1}(\mathbf{x}_i) - M_{j-1}^{\max} \sum_{j=1}^{n} e^{-\beta\left\|\mathbf{x}_i - \mathbf{x}_{j-1}^{\max}\right\|^2} \tag{5}$$

where $\mathbf{x}_{j-1}^{\max}$ is the previous data point obtained as the cluster center and $M_{j-1}^{\max}$ is its potential value. A positive constant β is used as a quash distance factor which is used to multiply the radii value. Similar to radii factor, it quashes the potential for surrounding points to be regarded as part of the cluster. This iteration continues to a predefined number of clusters [21].

It must be noted that only cluster centers ($\mathbf{cnt}_k$, $k = 1, \ldots, c$, where c shows the number of clusters) are obtained by subtractive clustering. After obtaining the cluster centers, several similarity-based methods have been utilized for segmentation in this paper. Hence, cluster centers and pixel values are compared and clustered according to a similarity measure for segmentation. The first similarity measure evaluated for this purpose is the well-known Euclidean-distance-based similarity measure that can be formulated as

$$EUC(\mathbf{x}, \mathbf{cnt}_k) = \sqrt{\sum_{i=1}^{d} (\mathbf{x}_i - \mathbf{cnt}_{k,i})^2} \tag{6}$$

where d shows the dimension of the spectral data, $\mathbf{x}$ shows the spectral data of the current pixel being processed, and $\mathbf{cnt}$ shows the spectral data of a cluster center.

The second similarity measure used is the normalized SAM [2], which can be formulated as

$$SAM(\mathbf{x}, \mathbf{cnt}_k) = 1 - \cos^{-1}\left(\sum_{i=1}^{d} \mathbf{x}_i \mathbf{cnt}_{k,i} \Big/ \sqrt{\sum_{i=1}^{d} \mathbf{x}_i^2} \sqrt{\sum_{i=1}^{d} \mathbf{cnt}_{k,i}^2}\right). \tag{7}$$

The third one, called SID, measures the information divergence between the spectral signatures. SID is based on the concept of divergence arising in information theory, and it is used to describe the statistics of a spectrum [3]

$$\begin{aligned} SID(\mathbf{x}, \mathbf{cnt}_k) &= 1 - [D(\mathbf{x}\|\mathbf{cnt}_k) - D(\mathbf{cnt}_k\|\mathbf{x})] \\ D(\mathbf{x}\|\mathbf{cnt}_k) &= \sum_{n=1}^{n} p_n(\mathbf{x}) \log\left(p_n(\mathbf{x})/p_n(\mathbf{cnt}_k)\right) \\ p_n(\mathbf{x}) &= \mathbf{x}_n / \sum_{i=1}^{n} \mathbf{x}_i. \end{aligned} \tag{8}$$

The fourth similarity measure evaluated is the phase-correlation-based similarity measure [4]. If X and CNT show the corresponding discrete Fourier transforms (DFTs) of the spectral data of the current pixel and the spectral data of a cluster center, the phase correlation of these two spectral signals is defined as

$$Ph.Cor(\mathbf{x}, \mathbf{cnt}_k) = F^{-1}\left[\frac{X \cdot CNT_k^*}{|X \cdot CNT_k^*|}\right] \tag{9}$$

where F^{-1} represents the inverse DFT. The phase correlation result of two identical signals has a peak value of unity. The peak of the phase correlation can be detected more accurately than classical cross-correlation, because phase correlation gives a distinct sharp peak in the case of high similarity between two spectral signatures [22].

In order to perform subtractive-clustering-based segmentation, pixels are grouped into cluster whose cluster center data have the highest similarity.

IV. OC-SVM and SV-PWSD Cluster Validity

OC-SVM was introduced by Schölkopf *et al.* [26]. OC-SVM was proposed for estimating the support of a high-dimensional distribution, and it is currently a popular method for outlier detection or anomaly detection. In summary, OC-SVM maps the data points into a high-dimensional feature space using an appropriate kernel function. After mapping to high-dimensional space, OC-SVM tries to find a hyperplane that separates the high-dimensional feature vectors from the origin of the transformed space with maximum margin. OC-SVM handles the origin of the transformed feature space as the second class, and the feature vectors in that class are regarded as outliers or anomalies. In the quadratic programming formulation of OC-SVM, a tradeoff parameter is utilized to control the distance of the hyperplane from the origin and to keep most of the samples which are regarded as training data [23]. Another version of one-class SVM, introduced by Tax and Duin, is

called SV domain description (SVDD) [24]. SVDD estimates a sphere as a classification decision function that encloses most of the training points in the high-dimensional feature space. It distinguishes a set of target points from the outlier points while keeping most of the target points inside the sphere.

In this paper, Schölkopf's version of OC-SVM is considered, which is therefore explained briefly in this section. The hyperspectral data set can be shown as $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, where n shows the number of samples in the data set and $\mathbf{x} \in \mathbf{R}^d$. The main aim is to estimate a decision function that takes the value $+1$ in a small region containing most of the data points and -1 elsewhere. Finding the hyperplane with maximal margin can be defined as a convex problem

$$\begin{aligned} \min_{\mathbf{w},\xi_1,\ldots,\xi_n,\rho} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \\ \text{subject to :} \quad & \langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle \geq \rho - \xi_i \quad \text{and} \quad \xi_i \geq 0, \\ & i = 1, \ldots, n \end{aligned} \tag{10}$$

where $\xi_i (i = 1, 2, \ldots, n)$ shows the slack variables and $\nu \in (0, 1)$ is the tradeoff parameter that controls the number of outliers and model complexity. Here, ρ denotes the margin. Any data point in the data set is projected to a feature space by a nonlinear transformation function $(\phi : X \to F)$. For the classification of a data point, a decision function that assigns a label to this point can be obtained with

$$f(\mathbf{x}) = \operatorname{sgn}\left(\langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle - \rho\right). \tag{11}$$

Introducing the Lagrange multiplier theorem and the Karush–Kuhn–Tucker complementary, the following equation can then be derived using the Lagrange multipliers (α_i)

$$\mathbf{w} = \sum_i \alpha_i \phi(\mathbf{x}_i) \tag{12}$$

where only a subset of points $\mathbf{x}_i$ that have minimum distances to the hyperplane has nonzero values of α_i. These points are called SVs. Instead of solving this primal optimization problem, the Wolfe dual of this problem can be represented as follows:

$$\begin{aligned} \min_{\alpha_1,\ldots,\alpha_n} \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j K(x_i, x_j) \\ \text{subject to :} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu n}, \\ & \sum_i \alpha_i = 1, \quad i = 1, \ldots, n \end{aligned} \tag{13}$$

where $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$ shows a Mercer kernel. Using kernel functions, the decision function can be transformed into a high-dimensional feature space for the nonlinearly separable case. In this way, the hyperplane can be represented in the feature space [25], [26]. Thus, the decision function in the feature space can be formulated as

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) - \rho\right). \tag{14}$$

An objective statistical criterion needs to be defined for the validation of clustering methods. For this purpose, a modified version of the PWSD, which is given in [2], is utilized. Basically, the proposed cluster validity measure is based on the modified PWSD measure and utilizes the advantage of the inherited cluster contour definition feature of OC-SVM, as stated in [27] and [28], by using Gaussian kernel. SVs obtained by OC-SVM are located at the minimum distance to the hyperplane in the feature space and at the arbitrarily shaped cluster boundary of the input space.

In the proposed cluster validation measure, referred to as SV-PWSD, the correct number of clusters is determined using SVs of each cluster obtained by OC-SVM. In the input space, SVs lie on the cluster boundaries, and instead of using all spectral data, only SV spectral data that represent the shapes of the clusters will be utilized.

The PWSD provides a measure of discrimination for a certain pixel with respect to two reference cluster centers. Representative cluster signatures are calculated by averaging the pixel data in each cluster. If $\mathbf{x}$ is used to show the spectral data of a pixel and $\mathbf{s_i}$ and $\mathbf{s_j}$ are two cluster representatives, the PWSD measure is defined in this paper by using phase correlation as

$$\Omega(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}) = \max\left\{\frac{Ph.Cor(\mathbf{s}_i, \mathbf{x})}{Ph.Cor(\mathbf{s}_j, \mathbf{x})}, \frac{Ph.Cor(\mathbf{s}_j, \mathbf{x})}{Ph.Cor(\mathbf{s}_i, \mathbf{x})}\right\}. \tag{15}$$

For each pixel $\mathbf{x_i}$ in the ith cluster, the PWSD is computed using the representative signature of the cluster to which the pixel belongs to $(\mathbf{s_i})$ and the representative signatures of all other clusters $\mathbf{s_j}$, $(i \neq j)$. The segmentation accuracy (SA) for one pixel (the pixel $\mathbf{x_i}$) is then obtained as the average of the PWSD measures and can be formulated as

$$SA(\mathbf{x}_i) = mean\left\{\Omega\left(\mathbf{s}_i, \mathbf{s}_j, \mathbf{x}_i\right) | i, j = 1, \ldots, c, i \neq j\right\} \tag{16}$$

where c shows the total number of clusters.

The PWSD value is by definition always larger than one, and the discrimination capability is regarded to increase with increased PWSD value. The segmentation performance of a certain approach can be obtained by averaging the SA values computed for all pixels in the scene.

In SV-PWSD, SV representatives $(\mathbf{s_{SV}}_i)$ are calculated by separately averaging the SVs of each cluster. For each SV $(\mathbf{x_{SV}}_i)$ in the ith cluster, the SV-PWSD is computed by using the SV representative of the cluster $(\mathbf{s_{SV}}_i)$ to which $\mathbf{x_{SV}}_i$ belongs to and the SV representatives of other clusters $\mathbf{s_{SV}}_j$, $(i \neq j)$ in the form of

$$\begin{aligned} &\Omega_{SV}(\mathbf{s_{SV}}_i, \mathbf{s_{SV}}_j, \mathbf{x_{SV}}_i) \\ &\quad = \max\left\{\frac{Ph.Cor(\mathbf{s_{SV}}_i, \mathbf{x_{SV}}_i)}{Ph.Cor(\mathbf{s_{SV}}_j, \mathbf{x_{SV}}_i)}, \frac{Ph.Cor(\mathbf{s_{SV}}_j, \mathbf{x_{SV}}_i)}{Ph.Cor(\mathbf{s_{SV}}_i, \mathbf{x_{SV}}_i)}\right\}. \end{aligned} \tag{17}$$

The SV-SA for one SV (the SV $\mathbf{x_{SV_i}}$) is then obtained as the average of the SV-PWSD measures and can be formulated as

$$\begin{aligned} SV\text{-}SA(\mathbf{x_{SV}}) = mean\{&\Omega_{SV}(\mathbf{s_{SV}}_i, \mathbf{s_{SV}}_j, \mathbf{x_{SV}}) \\ &| i, j = 1, \ldots, c, i \neq j\}. \end{aligned} \tag{18}$$

Fig. 2. Part of sample scene in DC Mall data (50th band), which is used in the experiments.

Fig. 3. Part of the sample scene in ROSIS 03 University data (75th band), which is used in the experiments.

The SV-PWSD value is also by definition always larger than one, and the discrimination capability is regarded to increase with increased SV-PWSD value. The segmentation performance of a clustering method can be obtained by averaging all SV-SA values obtained for only SV pixels in the scene. It must be noted that a clustering method is regarded to have good segmentation capability if the corresponding SV-SA measure is much larger than unity. Furthermore, a good segmentation capability is highly correlated with the cluster validity measure, because a cluster validity measure indicates the quality of clustering results and, by implication, the capability of good segmentation. Hence, the cluster validity of utilized clustering methods can be evaluated by the distance of their SV-SA measure from one and assessed comparatively with each other.

V. Experimental Results

In the experiments, initially, the HYDICE hyperspectral image of Washington DC Mall is used [29]. This hyperspectral data set consists of bands between 0.4 to 2.4 μm in the electromagnetic spectrum. The scene contains 1280 lines/scene and 307 pixels/line. A rotated sample image is shown in Fig. 2. The band number is first reduced from 210 to 191 spectral bands by removing water absorption bands affected by the atmosphere.

In the second experiment, the Reflection Optics System Imaging Spectrometer (ROSIS 03) hyperspectral data collected over the University of Pavia, Pavia, Italy, are used [30]. A rotated sample image is shown in Fig. 3. This scene has 610 lines/scene and 340 pixels/line and contains bands between 0.43 to 0.86 μm in the electromagnetic spectrum. Initially, water absorption bands affected by the atmosphere are removed, and the band number is reduced from 115 to 103 spectral bands.

The ground truth of ROSIS University includes nine classes, i.e., asphalt, meadows, gravel, trees, metal, soil, bitumen, bricks, and shadow. After removing some of the training samples which have zero features, the full training set contains 3921 samples [30]. Ground truth is also available for the DC Mall data set, and there are 8079 labeled pixels for a total of seven classes, including roofs, roads, grass, trees, trail, water, and shadow [29].

In the dimension reduction step, SVDBS has been applied. The implementation of this algorithm is available through the Hyperspectral Image Analysis Toolbox platform [31]. The numbers of selected bands are set to 20, 30, 40, and 50 for both experiments. In the DC Mall and University data sets, only supplied training samples (ground truths) are utilized in the clustering and validation phases to enable comparison with the available ground truth.

As stated before in the introduction section, the presented approach consists of two stages. The first stage has initial clustering and validation phases. To overcome the high computational load of subtractive clustering and OC-SVM, particularly for a high volume of hyperspectral images, it is initially proposed to use a subset of the image (training samples) in this paper. In the experiments, only the ground-truth data of each hyperspectral image have been selected as an initial subset. In the second stage, after assigning the labels for each pixel in this subset, KNN-based classification is used to project the initial clustering results onto the entire data set. Hence, the entire data set is assigned to corresponding clusters according to the KNN classification results.

In the subtractive-clustering-based similarity segmentation method, subtractive clustering is initially employed to determine high-density areas as candidate cluster centers. Subtractive clustering is applied to obtain from 5 to 10, and also 12 and 14 cluster centers for the DC Mall data and from 7 to 12 and also 14 and 16 candidate cluster centers for the University data in the presented results. Then, clustering/segmentation is performed using these cluster centers and an appropriate similarity measure. In order to perform subtractive-clustering-based segmentation, first of all, the densest cluster centers are found by adjusting the radii value. The quash distance factor β value is fixed experimentally to 1.1. Spectral data that are most similar to a cluster center obtained by subtractive clustering are grouped into the same cluster by using the Euclidean distance (EUC)-, SAM-, SID-, and phase-correlation-based similarity measures. Furthermore, for comparison purposes, KM and FCM algorithms are also utilized, with a predefined number of clusters. In KM and FCM, the error tolerance is set to 0.0001 experimentally.

After clustering of the subsets, OC-SVM is utilized for obtaining SVs that define the shape of the cluster boundary in the input space of each cluster. Note that SVs are used for validation purposes but not for any segmentation purposes. In SV-PWSD, SVs are considered to represent the border of a cluster instead of each sample in the cluster. This approach can also reduce the computational load of validation. In OC-SVM data, samples are mapped from input space to high-dimensional feature space by a Gaussian kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2\right). \tag{19}$$

The Gaussian kernel permits one to get cluster contour representations in the input space [28]. There are two important user-specified parameters in OC-SVM: the width of the Gaussian kernel (σ) in (19) and the tradeoff term (ν) in (10). These parameters have been set after evaluation in a range to get the best results. For Gaussian kernel (σ), results are obtained in

TABLE I
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR DC MALL SCENE (20 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
5	1.3862	1.4028	1.3837	1.4008	1.3612	1.3791	1.3835	1.4007	1.3364	1.3609	1.3647	1.3817
6	1.3736	1.3872	1.3694	1.3923	1.3558	1.3848	1.3747	1.3927	1.3638	1.3560	1.3865	1.3975
7	**1.4039**	1.3946	1.3764	1.3899	1.3723	1.3870	1.3796	1.3900	**1.3674**	1.3880	**1.3915**	1.4015
8	1.3956	1.4016	**1.3890**	1.4018	1.3800	1.3999	1.3868	1.4020	1.3633	1.3863	1.3663	1.3809
9	1.3824	1.3877	1.3760	1.3901	1.3701	1.3867	1.3772	1.3903	1.3482	1.3693	1.3564	1.3775
10	1.3984	1.4000	1.3874	1.4003	**1.3805**	1.3947	**1.3917**	1.4002	1.3398	1.3540	1.3546	1.3699
12	1.3755	1.3830	1.3845	1.3843	1.3642	1.3762	1.3802	1.3841	1.3363	1.3463	1.3567	1.3704
14	1.3773	1.3869	1.3781	1.3866	1.3692	1.3781	1.3862	1.3864	1.3327	1.3486	1.3318	1.3487

TABLE II
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR DC MALL SCENE (30 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
5	1.4802	1.5104	1.4630	1.4997	1.4350	1.4841	1.4604	1.4995	1.4518	1.4936	1.4312	1.4844
6	1.4879	1.5271	1.4794	1.5170	**1.4843**	1.5087	1.4909	1.5167	1.4687	1.5089	1.4581	1.5063
7	**1.4901**	1.5080	**1.4879**	1.5068	1.4758	1.4953	1.4889	1.5062	**1.4839**	1.5084	**1.4692**	1.5044
8	1.4757	1.5014	1.4695	1.4975	1.4680	1.4889	1.4693	1.4977	1.4648	1.4892	1.4670	1.4948
9	1.4829	1.4859	1.4646	1.4876	1.4530	1.4752	1.4663	1.4870	1.4778	1.5084	14535	1.4855
10	1.4615	14678	1.4587	1.4683	1.4450	1.4603	1.4611	1.4674	1.4664	1.4904	1.4330	1.4750
12	1.4822	1.4976	1.4828	1.4987	1.4481	1.4811	**1.4940**	1.4985	1.4466	1.4799	1.4365	1.4770
14	1.4612	1.4823	1.4542	1.4779	1.4371	1.4633	1.4640	1.4779	1.4445	1.4779	1.4253	1.4588

TABLE III
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR DC MALL SCENE (40 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
5	1.4858	1.5403	1.4608	1.5328	1.4594	1.5257	1.4582	1.5328	1.4690	1.5286	1.4573	1.5257
6	1.5247	1.5578	1.4998	1.5306	1.4862	1.5199	1.4944	1.5297	1.4931	1.5208	1.4404	1.4986
7	**1.5555**	1.5634	**1.5443**	1.5450	1.4862	1.5221	**1.5358**	1.5446	**1.5109**	1.5171	**1.4904**	1.5292
8	1.5192	1.5261	1.5371	1.5301	**1.4935**	1.5142	1.5169	1.5223	1.4877	1.5183	1.4828	1.5420
9	1.5181	1.5299	1.4993	1.5151	1.4673	1.4990	1.5150	1.5140	1.4868	1.5014	1.4629	1.5172
10	1.5152	1.5108	1.4916	1.4950	1.4710	1.4846	1.4945	1.4935	1.4610	1.4978	1.4788	1.5223
12	1.5173	1.5271	1.4995	1.5255	1.4831	1.5143	1.5159	1.5254	1.4581	1.4882	1.4785	1.5235
14	1.4908	1.4980	1.4860	1.4910	1.4673	1.4869	1.4793	1.4913	1.4221	1.4822	1.4744	1.4994

TABLE IV
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR DC MALL SCENE (50 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
5	1.2429	1.2840	1.2317	1.2818	1.2448	1.2814	1.2304	1.2814	1.2443	1.2790	1.2348	1.2777
6	1.2759	1.3010	1.2544	1.2979	**1.2763**	1.2958	1.2778	1.2973	1.2762	1.2957	1.2280	1.2617
7	**1.2796**	1.2921	**1.2799**	1.2899	1.2624	1.2822	1.2782	1.2883	**1.3002**	1.3220	**1.2815**	1.3006
8	1.2778	1.2840	1.2672	1.2813	1.2541	1.2732	1.2559	1.2806	1.2854	1.2963	1.2752	1.2905
9	1.2578	1.2795	1.2777	1.2820	1.2670	1.2758	**1.2835**	1.2812	1.2779	1.2860	1.2581	1.2781
10	1.2686	1.2661	1.2676	1.2673	1.2560	1.2620	1.2744	1.2666	1.2704	1.2747	1.2544	1.2682
12	1.2482	1.2506	1.2544	1.2534	1.2371	1.2471	1.2538	1.2529	1.2696	1.2747	1.2705	1.2948
14	1.2709	1.2730	1.2725	1.2719	1.2587	1.2688	1.2647	1.2723	1.2277	1.2359	1.2551	1.2600

the range of between 0.1 and 2 with 0.1 incremental steps. The tradeoff term (ν) is changed between 0.1 and 0.5 with 0.1 incremental steps for all data subsets to obtain the best results.

Tables I–IV show the average SA values obtained by PWSD and SV-SA values obtained for the proposed SV-PWSD cluster validation methods for the DC Mall. This data set has been evaluated with different numbers of bands selected by SVDBS (20,

TABLE V
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR ROSIS UNIVERSITY SCENE (20 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
7	1.1771	1.3529	**1.4129**	1.3477	1.3345	1.3045	**1.4110**	1.3419	1.2498	1.3338	**1.3239**	1.3294
8	1.2398	1.3828	1.3039	1.3158	1.3015	1.2777	1.3046	1.3134	1.2961	1.3109	1.2935	1.3092
9	**1.4389**	1.4976	1.3325	1.3669	1.3651	1.3834	1.3282	1.3430	**1.3005**	1.2921	1.2898	1.2971
10	1.3866	1.4842	1.2926	1.2983	1.3618	1.3676	1.2909	1.2948	1.2589	1.2776	1.2588	1.2706
11	1.3250	1.4100	1.2829	1.2874	**1.4709**	1.3945	1.2970	1.2827	1.2456	1.2618	1.2492	1.2640
12	1.3396	1.4169	1.3048	1.3296	1.3990	1.4326	1.3000	1.3301	1.2748	1.2632	1.2380	1.2517
14	1.3841	1.4658	1.3523	1.3463	1.4237	1.4075	1.3315	1.3453	1.2393	1.2513	1.2494	1.2454
16	1.3824	1.4508	1.3961	1.3853	1.4158	1.4005	1.3826	1.3867	1.2386	1.2391	1.2766	1.2366

TABLE VI
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR ROSIS UNIVERSITY SCENE (30 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
7	1.3026	1.4029	**1.4480**	1.4322	1.3785	1.3939	**1.4160**	1.4206	1.2943	1.3821	1.2971	1.3596
8	1.3262	1.3936	1.3827	1.3851	1.3487	1.3669	1.3905	1.3773	1.2652	1.3326	1.2642	1.3304
9	**1.3349**	1.4149	1.4260	1.4304	1.3840	1.3346	1.4697	1.4281	**1.4026**	1.4218	**1.2995**	1.3069
10	1.2866	1.3522	1.3946	1.4055	1.3365	1.3155	1.4145	1.4082	1.2648	1.2975	1.2536	1.2835
11	1.2640	1.3693	1.2936	1.2911	1.4252	1.3653	1.2854	1.2833	1.2421	1.2910	1.2477	1.2695
12	1.2897	1.3484	1.3254	1.3169	1.4079	1.3622	1.3259	1.3093	1.2381	1.2606	1.2399	1.2657
14	1.3166	1.3710	1.2856	1.2911	1.3500	1.2987	1.2774	1.2842	1.2609	1.2518	1.2481	1.2500
16	1.3046	1.3804	1.3315	1.3403	**1.4374**	1.3789	1.3817	1.3341	1.2244	1.2370	1.2464	1.2271

TABLE VII
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR ROSIS UNIVERSITY SCENE (40 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
7	1.5525	1.4464	**1.5166**	1.5584	1.3423	1.3432	**1.5578**	1.5498	1.4256	1.4744	**1.3962**	1.4709
8	1.4907	1.4000	1.4743	1.5101	1.3119	1.3118	1.5024	1.5014	1.4066	1.4773	1.3586	1.4260
9	**1.5632**	1.5547	1.4804	1.4969	1.3156	1.3030	1.5485	1.5485	**1.5186**	1.5401	1.3631	1.3944
10	1.4505	1.3972	1.4196	1.4592	1.2934	1.2774	1.4650	1.4650	1.3783	1.4383	1.3407	1.3636
11	1.5091	1.5006	1.4387	1.4438	1.2900	1.2673	1.4851	1.4851	1.3580	1.4123	1.3235	1.3414
12	1.5256	1.5227	1.3814	1.4238	1.2991	1.2618	1.4036	1.4036	1.3126	1.3266	1.3360	1.3854
14	1.4817	1.4581	1.4154	1.3945	1.3464	1.2667	1.4196	1.4196	1.2912	1.2982	1.3220	1.3544
16	1.5295	1.4890	1.5093	1.4769	**1.4742**	1.3354	1.5087	1.5087	1.3084	1.3284	1.3065	1.3286

TABLE VIII
AVERAGE SV-SA AND SA RESULTS FOR DIFFERENT UNSUPERVISED CLUSTERING METHODS FOR ROSIS UNIVERSITY SCENE (50 BANDS)

Number of Clusters	Sub.Clus+Ph.Corr.		Sub.Clus+SAM		Sub.Clus+Euc.		Sub.Clus+SID		K-Means		FCM	
	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA	SV-SA	SA
7	1.6310	1.5525	**1.6353**	1.6062	1.3385	1.3719	**1.6645**	1.5995	1.5448	1.5353	**1.5267**	1.5243
8	1.6549	1.5976	1.5908	1.4826	1.3088	1.3374	1.5370	1.4444	1.4954	1.4788	1.4852	1.4748
9	**1.6715**	1.6333	1.5589	1.4417	1.3118	1.3231	1.4841	1.4099	**1.6533**	1.7300	1.4404	1.4343
10	1.5965	1.5786	1.5652	1.4723	1.2800	1.2920	1.5271	1.4597	1.4274	1.4054	1.4780	1.4023
11	1.6110	1.5928	1.5574	1.4561	1.3140	1.2820	1.5274	1.4474	1.3916	1.3813	1.3890	1.3766
12	1.6110	1.5870	1.4684	1.4906	1.3660	1.2977	1.4793	1.4795	1.3657	1.4481	1.3669	1.3570
14	1.5767	1.5545	1.4425	1.4993	1.3801	1.3013	1.4593	1.4923	1.3567	1.3340	1.3289	1.4020
16	1.6093	1.5448	1.6258	1.5693	**1.4801**	1.3634	1.6462	1.5768	1.3283	1.3094	1.3257	1.3035

30, 40, and 50 bands, respectively). In Tables I–III, the highest SV-SA values are obtained by subtractive clustering with phase correlation method. The only exception is in Table IV; in this case, the highest result is achieved using KM. In Tables I–IV, the highest SV-SA values for subtractive clustering with phase correlation, KM, and FCM are always obtained for seven clusters, which is the actual number of clusters that is known to be included in the scene. Moreover, according to the results of the DC Mall, it can be said that SV-PWSD gives more accurate and consistent information about the number of

clusters compared with PWSD. If all tables related with DC Mall scene (Tables I–IV) are examined, it can be seen that all of the methods except subtractive clustering with Euclidean distance clearly indicate that the number of clusters should be taken as seven for the DC Mall data (which is concordant with the number of labels supplied with the ground truth). In Table I, subtractive clustering with SAM is the only exception. Hence, there is actually a consensus between the methods.

Tables V–VIII show the average SA obtained by PWSD and the SV-SA values for the proposed SV-PWSD cluster validation algorithm for the ROSIS University data. Similar to the previous tables, Tables V–VIII show the results for different numbers of bands obtained by SVDBS (20, 30, 40, and 50 bands, respectively). In Tables V–VIII, the highest SV-SA values are always obtained by subtractive clustering with phase correlation or KM (in Table VI). For this data set, only subtractive clustering with phase correlation and KM give consistent results with the correct number of clusters. Hence, the number of clusters should be taken as nine for the ROSIS University data, and it is also concordant with the number of labels supplied with the ground truth.

According to the experimental results presented in Tables I–VIII, it can be concluded that SV-PWSD is more suitable than PWSD in defining the number of clusters using subtractive clustering with phase correlation and KM. However, in Tables I–VIII, it can be seen that subtractive clustering with phase correlation method generally gives higher SV-SA values than KM (in six out of eight tables). Note that SV-PWSD uses only SVs in a cluster representation for computing the cluster validity measure. Furthermore, it can be seen that, among all of the clustering methods in the Tables I–VIII, subtractive clustering with phase correlation method and KM give more accurate results for obtaining the exact number of available clusters. Thus, it can be said that subtractive clustering with phase correlation method and KM are more steady and reliable methods against other methods in the validation of cluster number.

Because of the high computational load of subtractive clustering and OC-SVM, these approaches cannot be applied to data sets with large number of samples directly, as observed in the experiments. Therefore, it is proposed to use KNN-based classification to project the initial clustering results obtained with a small subset onto the entire data set. KNN as a supervised classification technique has advantages of fast computation and good generalization performance. In the experiments K (number of nearest neighbors) was chosen as nine (after evaluations in the range of $K = 1, 3, 5, \ldots, 15$), which gives the maximum average SA value. Finally, the standard average SA values are computed using the regular PWSD measure to assess the segmentation performance (SV-PWSD is not used in this case because of computational constraints). These values are shown in Tables IX and X for the DC Mall and University data sets. From Tables IX and X, it can be seen that the best average SA values are obtained by subtractive clustering with phase correlation and KM, respectively. The highest values are obtained for 40 bands for the DC Mall and 50 bands for the University hyperspectral images. Figs. 4 and 5 show the segmentation maps of the scenes obtained by subtractive clustering with phase-correlation-based similarity measure and KM, respectively.

TABLE IX
AVERAGE SA RESULTS OF THE ENTIRE DC MALL SCENE

Number of Bands	SC+ EUC	SC+ SAM	SC+ SID	SC+ PC	KM	FCM
20	1.3642	1.3591	1.3593	1.3475	1.3527	**1.3717**
30	1.4051	1.4317	1.4308	1.4325	**1.4386**	1.4270
40	1.4122	1.4472	1.4475	**1.4494**	1.4321	1.3870
50	1.2205	1.2330	1.2322	1.2436	**1.2526**	1.2473

TABLE X
AVERAGE SA RESULTS OF ENTIRE UNIVERSITY SCENE

Number of Bands	SC+ EUC	SC+ SAM	SC+ SID	SC+ PC	KM	FCM
20	1.3442	**1.4148**	1.3795	1.3834	1.3026	1.3508
30	1.2981	1.3017	1.3713	**1.4428**	1.2293	1.2773
40	1.4520	1.4870	1.3276	**1.5080**	1.1963	1.3635
50	1.4847	1.4750	1.3687	1.5224	**1.6538**	1.2774

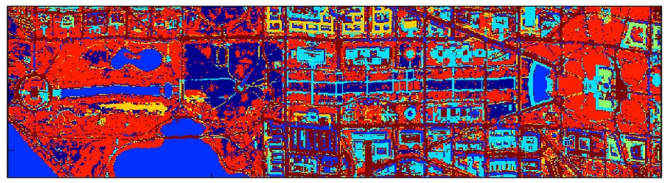

Fig. 4. Segmentation map of the DC Mall data with seven clusters obtained by subtractive clustering with phase correlation method for 40 bands.

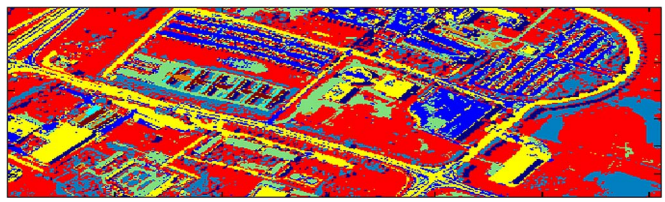

Fig. 5. Segmentation map of the ROSIS University data with nine clusters obtained by KM for 50 bands.

VI. CONCLUSION

Hyperspectral image segmentation based on subtractive clustering with phase correlation method and a novel approach for cluster validation based on OC-SVM and PWSD have been presented in this paper. In the proposed cluster validation technique (SV-PWSD), OC-SVM is utilized for obtaining SVs that identify the shape of the cluster boundary of each cluster in the input space. SVs found by OC-SVM are located at the minimum distance to the hyperplane in the feature space and at the arbitrarily shaped cluster contours in the input space. It is shown that the proposed cluster validity method finds accurate number of clusters by using subtractive clustering with phase correlation and KM. Thus, it can be said that the proposed cluster validity method can help to explain the unknown cluster structure in hyperspectral data sets.

REFERENCES

[1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ: Wiley, 2003, pp. 3–90.
[2] F. van der Meer, "The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 8, no. 1, pp. 3–17, 2006.

[3] C.-I. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Hamburg, Germany, Jun. 28–Jul. 2, 1999, pp. 509–511.
[4] A. Ertürk and S. Ertürk, "Unsupervised segmentation of hyperspectral images using modified phase correlation," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 527–531, Oct. 2006.
[5] J. Silverman, C. E. Caefer, J. M. Mooney, M. M. Weeks, and P. Yip, "An automated clustering/segmentation of hyperspectral images based on histogram thresholding," in *Proc. SPIE*, 2002, vol. 4480, pp. 65–75.
[6] J. Silverman, S. R. Rotman, and C. E. Caefer, "Segmentation of hyperspectral images based on histograms of principal components," in *Proc. SPIE*, 2002, vol. 4816, pp. 270–277.
[7] N. Acito, G. Corsini, and M. Diani, "An unsupervised algorithm for hyperspectral image segmentation based on the Gaussian mixture model," in *Proc. IEEE IGARSS*, Toulouse, France, Jul. 2003, vol. 6, pp. 3745–3747.
[8] M. D. Farrell, Jr. and R. Mersereau, "Robust automatic clustering of hyperspectral imagery using non-Gaussian mixtures," in *Proc. SPIE*, 2004, vol. 5573, pp. 161–172.
[9] G. Mercier, S. Derrode, and M. Lennon, "Hyperspectral image segmentation with Markov chain model," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Toulouse, France, Jul. 2003, vol. 6, pp. 3766–3768.
[10] H. H. Muhammed, "Unsupervised hyperspectral image segmentation using a new class of neuro–fuzzy systems based on weighted incremental neural networks," in *Proc. IEEE 31st Appl. Image Pattern Recognit. Workshop*, Washington, DC, Oct. 2002, pp. 171–177.
[11] C. A. Shah, P. Watanachaturaporn, M. K. Arora, and P. K. Varshney, "Some recent results on hyperspectral image classification," in *Proc. IEEE Workshop Adv. Tech. Anal. Remotely Sensed Data*, Greenbelt, MD, 2003, vol. 19, pp. 346–353.
[12] A. Meyer, D. Paglieroni, and C. Astaneh, " K-means re-clustering: Algorithmic options with quantifiable performance comparisons," in *Proc. SPIE*, 2003, vol. 5001, pp. 84–92.
[13] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, Feb. 2007.
[14] E. Arzuaga-Cruz, L. O. Jimenez-Rodriguez, and M. Velez-Reyes, "Feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis," in *Proc. SPIE*, vol. 5093, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX*, Apr. 2003, pp. 462–473.
[15] J.-S. Wang and J.-C. Chiang, "A cluster validity measure with outlier detection for support vector clustering," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 78–89, Feb. 2008.
[16] J. M. Moguerza, A. Munoz, and M. M. Merino, "Detecting the number of clusters using a support vector machine approach," in *Proc. Int. Conf. Artif. Neural Netw.*, vol. 2415, *Lecture Notes in Computer Science*, 2002, pp. 763–768.
[17] V. Estivill-Castro and J. Yang, "Cluster validity using support vector machines," in *Proc. 5th Int. Conf. Data Warehousing Knowl. Discovery*, vol. 2737, *LNCS*, 2003, pp. 244–256.
[18] A. C. Zelinski and V. K. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Geosci. Remote Sens. Conf.*, Denver, CO, Aug. 2006, pp. 387–390.
[19] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 8, pp. 1279–1284, Aug. 1994.
[20] S. L. Chiu, "Fuzzy model identification based on cluster estimation," *J. Intell. Fuzzy Syst.*, vol. 2, no. 3, pp. 267–278, Sep. 1994.
[21] D.-W. Kim, K. Y. Lee, D. Lee, and K. H. Lee, "A kernel-based subtractive clustering method," *Pattern Recognit. Lett.*, vol. 26, no. 7, pp. 879–891, 2005.
[22] O. Urhan, M. K. Güllü, and S. Ertürk, "Modified phase-correlation based robust hard-cut detection with application to archive film," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 753–770, Jun. 2006.
[23] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.
[24] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognit. Lett.*, vol. 20, no. 11–13, pp. 1191–1199, Nov. 1999.
[25] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," in *Advances in Neural Information Processing Systems*, vol. 12, S. A. Solla, T. K. Leen, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2000, pp. 582–588.
[26] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.
[27] A. Ben-Hur, D. Horn, T. Siegelmann, and V. Vapnik, "A support vector clustering method," in *Proc. Int. Conf. Pattern Recognit.*, 2000, vol. 2, pp. 724–727.
[28] F. Camastra, "Kernel methods for clustering," in *WIRN/NAIS*, vol. 3931, *Lecture Notes in Computer Science*, 2006, pp. 1–9, Berlin, Germany: Springer-Verlag.
[29] F. Camastra, *Signal Theory Methods in Multispectral Remote Sensing*. DC Mall image and band specifications for the HYDICE Washington D.C. Mall image provided on the CD with. [Online]. Available: http:// www. lars. purdue. edu/ home/ image_data/ hydice_dc_wavelengths. html
[30] M. Chi, Q. Kun, J. A. Benediktsson, and R. Feng, "Ensemble classification algorithm for hyperspectral remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 762–766, Oct. 2009.
[31] E. Arzuaga-Cruz, L. O. Jimenez-Rodriguez, M. Velez-Reyes, D. Kaeli, E. Rodriguez-Diaz, H. T. Velazquez-Santana, A. Castrodad-Carrau, L. E. Santos-Campis, and C. Santiago, "A MATLAB toolbox for hyperspectral image analysis," in *Proc. IGARSS*, 2004, vol. 7, pp. 4839–4842.

Gökhan Bilgin received the B.Sc., M.Sc., and Ph.D. degrees in electronics and telecommunication engineering from Yildiz Technical University (YTU), Istanbul, Turkey, in 1999, 2003, and 2009, respectively.

He is currently appointed as an Assistant Professor with the Department of Computer Engineering, YTU, where he worked as a Research Assistant between 2001 and 2010. He is also a Postdoctoral Researcher with the Department of Computer and Information Science, Indiana University–Purdue University Indianapolis, Indianapolis. His research interests are in the areas of image and signal processing, machine learning, and pattern recognition with applications to biomedical engineering and remote sensing.

Sarp Ertürk (M'99) received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 1995 and the M.Sc. degree in telecommunication and information systems and the Ph.D. degree in electronic systems engineering from the University of Essex, Colchester, U.K., in 1996 and 1999, respectively.

From 1999 to 2001, he carried out his compulsory service at the Army Academy, Ankara. He is currently appointed as a full Professor with Kocaeli University, Kocaeli, Turkey, where he worked as an Assistant Professor between 2001 and 2002, and an Associate Professor between 2002 and 2007. His research interests are in the areas of digital signal and image processing, video coding, remote sensing, and digital communications.

Tülay Yıldırım (M'03) received the B.Sc. and M.Sc. degrees in electronics and communication engineering, Yildiz Technical University (YTU), Istanbul, Turkey, in 1990 and 1992, respectively, and the Ph.D. degree in electrical and electronics engineering, University of Liverpool, Liverpool, U.K., in 1997.

She is currently a Professor with YTU. Her research interests are analog and digital IC design, neural networks, and intelligent systems.