Band-Subset-Based Clustering and Fusion for Hyperspectral Imagery Classification

Yong-Qiang Zhao, *Member, IEEE*, Lei Zhang, *Member, IEEE*, and Seong G. Kong, *Senior Member, IEEE*

Abstract—This paper proposes a band-subset-based clustering and fusion technique to improve the classification performance in hyperspectral imagery. The proposed method can account for the varying data qualities and discrimination capabilities across spectral bands, and utilize the spectral and spatial information simultaneously. First, the hyperspectral data cube is partitioned into several nearly uncorrelated subsets, and an eigenvalue-based approach is proposed to evaluate the confidence of each subset. Then, a nonparametric technique is used to extract the arbitrarily-shaped clusters in spatial-spectral domain. Each cluster offers a reference spectral, based on which a pseudosupervised hyperspectral classification scheme is developed by using evidence theory to fuse the information provided by each subset. The experimental results on real Hyperspectral Digital Imagery Collection Experiment (HYDICE) demonstrate that the proposed pseudosupervised classification scheme can achieve higher accuracy than the spatially constrained fuzzy c-means clustering method. It can achieve nearly the same accuracy as the supervised K-Nearest Neighbor (KNN) classifier but is more robust to noise.

Index Terms—Evidence theory, hyperspectral, image segmentation, information fusion.

I. INTRODUCTION

CLASSIFICATION is a challenging but important task for hyperspectral remote sensing applications, including land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance [1], [2]. Various hyperspectral imagery classification methods have been proposed, such as statistical method [1], [3], [24], soft computing-based methods [2], and information fusion methods [2], [4]. Most of these methods apply the classifier to the complete data set, neglecting the varying data quality and discrimination ability across bands. It is necessary to investigate the spectra signature variation of hyperspectral data to improve the classification performance [1], [2].

There are many factors affecting the hyperspectral data quality, ranging from the external factors such as atmospheric conditions to the internal factors such as sensor noise, sensor transfer characteristics, and material spectrum. For a specific object or material, the noise-dominated bands will certainly deteriorate the discrimination capability, and hence degrade the classification performance. On the other hand, the spectral difference among materials also varies across bands. In this paper, we propose a divide-and-conquer approach that employs information fusion to classify the hyperspectral data. It partitions the hyperspectral data into contiguous subsets that have similar characteristics so that the discrimination information within each subset can be maximized [6].

Traditionally, only spectral information was employed to classify the hyperspectral data set [1]–[3]. The current hyperspectral imagery has fine spatial resolution, and therefore not only the spectral information but also the spatial information can be used to classify the scene. By integrating the spatial information into hyperspectral classification process, higher classification accuracy can be expected [17]. Rand and Keenan [18] proposed a hyperspectral segmentation method to jointly utilize the spectral and spatial information by using the Markov random field technique. However, this method is computationally intensive because of the recursive and global optimization procedures.

In this paper, we propose a new algorithm to exploit the spectral and spatial characteristics of hyperspectral imagery. We first partition the complete hyperspectral data cube into several nearly uncorrelated subband cubes, each of which contains contiguous bands. These subbands are referred to as information sources. Channels within a subband cube are assumed to have similar noise characteristics and discrimination ability. A nonparametric clustering method is then used to extract the joint spatial-spectral features of the hyperspectral data. The clustering result from each information source is called a clustering map. Due to the variation of data quality and discrimination ability across bands, there exist uncertainties and errors in clustering maps. The next key issue is how to effectively fuse the features extracted from these information sources to improve the final classification performance.

Fuzzy models and evidence theory are widely used in dealing with uncertainty and inaccurate information [5], [13], [22], [23]. Fuzzy theory usually represents the inaccurate information in terms of explicit functions of membership, while evidence theory represents the inaccuracy and uncertainty simultaneously using confidence, plausibility, and credibility functions [5], [6]. Thus evidence theory can combine the items of evidence supporting certain hypotheses of a pattern from multiple information sources. In this paper, each subband set provides an evidence of the pixel's label, and the final classification is made by fusing the evidences using the Dempster's rule.

The rest of the paper is organized as follows. Section II describes the model of hyperspectral imagery, the correlation-based subband partition and confidence estimation. Section III presents a clustering fusion-based classifier for hyperspectral imagery. Section IV presents experiments on Hyperspectral Digital Imagery Collection Experiment (HYDICE) data in comparisons with the spatially constrained fuzzy c-means clustering method, which exploits the spatial information in spectral clustering, and the K-Nearest Neighbor (KNN) classification method. Section V concludes the paper.

II. SUBBAND PARTITION IN HYPERSPECTRAL IMAGERY

A. Subband Partition

Hyperspectral imagery is represented as a 3-D data cube obtained through both spatial and spectral sampling. The data cube can be written as an $L \times N$ matrix: $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_l, \ldots, \mathbf{y}_L]^T$, where $\mathbf{y}_l$ is the $1 \times N$ row vector for the lth band $\mathbf{y}_l = [y_l^1, \ldots, y_l^n, \ldots, y_l^N]$ and N is the total number of spatial pixels in each band. $\mathbf{Y}$ can be written as a linear mixture of target signals and background noise

$$Y = X + W \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_l, \ldots, \mathbf{x}_L]^T$ is the signal matrix and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_l, \ldots, \mathbf{w}_L]^T$ is the noise matrix.

Traditional hyperspectral imagery classification methods [1], [3] work directly on all bands without considering the different qualities and discrimination capabilities across bands. However, in actual application, both the internal sensing factors and the external environment factors will affect the data quality and lead to large quality variations among different bands. It has been demonstrated that the statistical behavior of hyperspectral imagery varies across bands [1]. Therefore, we partition the whole data cube into several subband sets, in which the bands exhibit similar characteristic and can be modeled using the same model.

A widely used tool to measure the similarity between different bands is the correlation coefficients between them. The correlation coefficient between the ith and the jth bands is given by

$$\mathbf{R}_c(i, j) = \Sigma(i, j) / \sqrt{\Sigma(i, i)\Sigma(j, j)} \tag{2}$$

where Σ is the $L \times L$ covariance matrix of $\mathbf{Y}$

$$\Sigma(i, j) = (\mathbf{y}_i - \mathbf{E}(\mathbf{y}_i)) (\mathbf{y}_j - \mathbf{E}(\mathbf{y}_j))^T. \tag{3}$$

An example of the correlation coefficient matrix of HYDICE data is illustrated in Fig. 3(a), where the brightness represents the magnitude of the coefficient. Based on the correlation coefficient matrix $\mathbf{R}_c$, the data cube can then be partitioned into Q nonoverlapped subband sets, which are viewed as nearly independent information sources. Denote by l_q the total number of bands in the qth source, $q = 1, 2, \ldots, Q$.

B. Confidence Estimation

The confidence level of an information source could reflect the reliability of the classification result by using that source. We assume that the noise in the hyperspectral data is independently and identically distributed (i.i.d) Gaussian white noise [19]. With (1), the data cube for the qth source can be written as $\mathbf{Y_q} = \mathbf{X_q} + \mathbf{W_q}$. We assume that the noise level in the qth source is approximately constant across the l_q spectral bands in that source, while different sources may have different noise levels. Then, the signal in the qth source can be estimated by using eigenvalue analysis [21].

The spectral covariance matrix of the qth source can be represented as

$$\Sigma_{\mathbf{Y}_q} = \mathbf{U}_q \mathbf{\Lambda}_q \mathbf{U}_q^T \tag{4}$$

where $\mathbf{\Lambda}_q = \text{diag}(\beta_1^2, \beta_2^2, \ldots, \beta_{l_q}^2)$ with $\beta_1^2 \geq \beta_2^2 \geq \cdots \geq \beta_{l_q}^2$ are the eigenvalues of noisy signal in the qth source. The power of noiseless signal in the qth source is estimated by computing the average of the R_q largest singular values of covariance matrix $\mathbf{R}_{\mathbf{Y}_q}$ [21]

$$\hat{\sigma}(\mathbf{X}_q)^2 = 1/R_q \sum_{i=1}^{R_q} \beta_i^2. \tag{5}$$

The noise power $\hat{\sigma}(\mathbf{W}_q)^2$ is estimated by computing the average of the $l_q - R_q$ smallest singular values of covariance matrix $\mathbf{R}_{\mathbf{Y}_q}$

$$\hat{\sigma}(\mathbf{W}_q)^2 = 1/(l_q - R_q) \sum_{i=R_q+1}^{l_q} \beta_i^2. \tag{6}$$

Considering that the performance of the classifier is affected by the signal-to-noise ratio (SNR), we define the confidence g_q of information source q as

$$g_q = \frac{\hat{\sigma}(\mathbf{X}_q)^2}{\hat{\sigma}(\mathbf{W}_q)^2} \tag{7}$$

which is actually the ratio between the estimated signal power to the noise power.

Now only the parameter R_q is left to estimate. We estimate this parameter by minimizing the Akaike information criterion (AIC) [16]

$$\text{AIC}(r_q) = -2M_q \sum_{i=r_q+1}^{l_q} \ln \beta_i + M_q(l_q - r_q)$$

$$\times \ln \left(\frac{1}{l_q - r_q} \sum_{i=k_q+1}^{l_q} \ln \beta_i \right) + 2r_q(l_q - r_q) \tag{8}$$

where $\beta_i, i = 1, 2, \ldots, l_q$, are the l_q singular values of $\mathbf{R}_{\mathbf{Y}_q}$ and M_q is the number of columns of Y_q. The estimated rank R_q is set to be the r_q which minimizes AIC

$$R_q = \arg\min_{r_q} AIC(r_q). \tag{9}$$

III. MULTISOURCE IMAGERY CLUSTERING FUSION

The hyperspectral imagery can be typically represented as a 2-D lattice of L-dimensional spectral vectors. The space of the lattice is known as the spatial domain while the spectral information is represented in the spectral domain. The spectral vectors can be concatenated locally to obtain a spatial-spectral domain. Hyperspectral imagery classification is performed by assigning a label to each voxel in the 2-D lattice. Kernel-based methods are widely used in hyperspectral imagery classification. One way to implement the kernel-based method is to find a kernel density estimate in the data space and then search for the model of the estimated density. The mean shift is a simple but effective technique that can be used to find the mode of a kernel density estimate [10]–[12], [14], [15]. Once the location of a mode is determined, the cluster associated with it is delineated based on the local structure of the feature space. Here, we use the mean shift as a feature extraction method to locate the cluster center by searching for the mode in the spatial-spectral domain.

After the mean shift-based feature extraction, each cluster provides a map of spatial-spectral homogeneous regions. We will use the average spectral vectors in homogeneous regions to form a set of reference spectrums. Then, the whole image can be classified based on these reference spectrums by using evidence theory. We call this classification method a pseudosupervised hyperspectral classification method.

A. Feature Extraction by Mean Shift Clustering

The qth subband set can be represented as N voxel points $\mathbf{Y}_n^q$, $n = 1, 2, \ldots, N$ in the l_q-dimensional space R^{l_q}, where l_q is the band number in the qth subband set. The mean shift procedure [7] is a simple gradient based technique to find the modes of voxel's probability density, which is acquired by kernel density estimate. Therefore, the first step of feature extraction is to estimate the probability density of $\mathbf{Y}_n^q$. Let $K : R^{l_q} \to R$ be a kernel with $K(\mathbf{y}) = k(\|\mathbf{y} - \mathbf{Y}_n^q\|^2)$. The probability density function is estimated by the multivariate kernel density estimator

$$\hat{f}_K(\mathbf{y}) = \sum_{n=1}^{N} k\left(\|\mathbf{y} - \mathbf{Y}_n^q\|^2\right) w\left(\mathbf{Y}_n^q\right) \quad (10)$$

where $w(\cdot)$ is a weight function. The gradient of the estimated density is

$$\nabla \hat{f}_K(\mathbf{y}) = 2 \sum_{n=1}^{n} \left(\mathbf{y} - \mathbf{Y}_n^q\right) k'\left(\|\mathbf{y} - \mathbf{Y}_n^q\|^2\right) w\left(\mathbf{Y}_n^q\right). \quad (11)$$

Suppose that there exists a kernel $G : R^{l_q} \to R$ with $G(\mathbf{y}) = g(\|\mathbf{y} - \mathbf{Y}_n^q\|^2)$ such that $k'(\mathbf{y}) = cg(\mathbf{y})$, where c is a constant. Substituting $k'(\mathbf{y}) = cg(\mathbf{y})$ into formula (11) and letting the gradient estimator $\nabla \hat{f}_K(\mathbf{y})$ be zero, we can derive a mode estimate as

$$\mathbf{y} = \frac{\sum_{n=1}^{N} g\left(\|\mathbf{y} - \mathbf{Y}_n^q\|^2\right) w\left(\mathbf{Y}_n^q\right) \mathbf{Y}_n^q}{\sum_{n=1}^{N} g\left(\|\mathbf{y} - \mathbf{Y}_n^q\|^2\right) w\left(\mathbf{Y}_n^q\right)}. \quad (12)$$

Based on the above analysis, the feature extraction for the qth subband set can be summarized as follows.

1) Initialize the nth voxel as $\mathbf{y}_{n,1} = \mathbf{Y}_n^q$ and then update the voxel $\mathbf{y}_{n,j+1}$ according to (12) until $\|\mathbf{y}_{j+1} - \mathbf{y}_j\| < \varepsilon$, where ε is a small positive number.
2) Let $\mathbf{y}_{n,center} = \mathbf{y}_{n,j}$, where n is the spatial location of the voxel and j is the number of iteration.
3) Identifying clusters $\{C_{p,q}\}_{p=1,\ldots,m}$ by linking all $\mathbf{y}_{n,center}$ which are closer than a given threshold, where m is the number of clusters in the qth subband set. The final feature extraction result is represented as a clustering map $C_q = \{C_{1,q}, \ldots, C_{m,q}\}$, and there are m homogeneous regions in this clustering map.

To utilize the spectral and spatial information jointly in the classification process, a proper kernel should be chosen in the feature extraction process. It should be a multivariate kernel, which can take several factors into consideration. For more information, please refer to [10]–[12]. To integrate the spatial and spectral information in the clustering process, the multivariate kernel for the joint spatial-spectral domain is used [12]. It is the product of two radially symmetric kernels and allows a single bandwidth parameter for each kind of information

$$G_{h_s, h_p}(x) = \frac{T}{h_s^2 h_p^3} g\left(\left\|\frac{x^s}{h_s}\right\|^2\right) g\left(\left\|\frac{x^p}{h_p}\right\|^2\right) \quad (13)$$

where x^s is the spatial part and x^p is the spectral part of a feature vector, $g(x)$ is the common profile used in joint spatial-spectral domain, h_s and h_p are the employed bandwidths, and T is the corresponding normalization constant. In the simulation, the Epanechnikov kernel is used. The Epanechnikov kernel is optimal when considering approximation accuracy and it is widely used in classification. The bandwidth is chosen according to the texture characteristics of the scene [12]. The mean shift feature extraction algorithm requires the selection of the bandwidth (h_s, h_p), which determines the resolution of the mode detection by controlling the size of the kernel.

B. Pseudosupervised Fusion

After performing feature extraction, there are Q clustering maps $C = \{C_1, C_2, \ldots, C_Q\}$. Here, we assume that there are R different homogeneous regions in all Q clustering maps, and the spectral voxels in the same homogeneous region are similar. For a homogeneous region r, the average value of all voxels in the qth subband set is named as the average spectrum and it is denoted by $\mathbf{S}_{r,q}$, where $r = 1, \ldots, R$ and $q = 1, \ldots, Q$. The average spectrums of R different homogeneous regions form a reference spectrum set $\mathbf{S} = \{\mathbf{S}_{1,1}, \ldots, \mathbf{S}_{r,q}, \ldots, \mathbf{S}_{R,Q}\}$, and it will be clustered into K groups, denoted by $\Omega = \{\omega_1, \ldots, \omega_K\}$.

With these reference spectrums and their corresponding class labels, a pseudosupervised fusion based classification method is proposed. The automatically labeled reference spectrum set is viewed as a training set $\mathbf{\Gamma} = \{(\mathbf{S}_{1,1}, \omega^1), \ldots, (\mathbf{S}_{r,q}, \omega^r), \ldots, (\mathbf{S}_{R,Q}, \omega^R)\}$ of $R \times Q$ reference spectrum $\mathbf{S}_{r,q}$ and their corresponding group label ω^r, $\forall \omega^r \in \Omega, r = 1, \ldots, R$ and $q = 1, \ldots, Q$. Suppose $n \in C_{p,q}$ is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

in the pth homogeneous region of the qth clustering map. To determine the class label of the ith pixel, three situations are considered here.

1) If the Q homogeneous regions are the same, then the ith pixel is assigned to the kth class if the average spectrum belongs to the kth group.
2) If the Q homogeneous regions are not the same but their average spectrums belong to the same group, say the kth group, then the class label of the ith pixel is assigned as k.
3) If the Q homogeneous regions are not the same and their average spectrums belong to different groups, the class label of the ith pixel is determined by decision fusion, which is described as follows.

Suppose the average spectrum of homogeneous region in the qth subband set, denoted as $\mathbf{AS}_q$, belongs to class k. The spectral distance $d_{i,k}^q$ between $\mathbf{AS}_q$ and the spectral vector at the ith pixel can be regarded as an evidence that the ith pixel belongs to class k. Based on this evidence and g_q, the confidence value of the qth information source, the reliability function of the ith pixel belonging to class k based on the qth information source can be defined as

$$\alpha_k^q = g_q \exp\left(-\gamma_q \left(d_{i,k}^q\right)^2\right) \tag{14}$$

where γ_q is a positive parameter associated to class ω_q. The reliability function α_k^q represents the reliability that the ith location is labeled as ω_q based on the information provided by the qth information source. While the reliability of the ith pixel is label as α_k^q, the reliability of the rest of Ω is defined as $1 - \alpha_k^q$.

If the average spectrums of Q regions belong to different groups, it implies that there are some conflicts and disagreements among the decisions made by the Q information sources. As mentioned in Section I, evidence theory [5], [8] is a proper tool to deal with these conflicts and disagreements. Here, we use the evidence theory to fuse these decisions. Through fusion, the reliability that the ith pixel is label as ω_q can be calculated

$$m\left(\{\omega_k\}\right) = \frac{1}{\Upsilon}\left(1 - \prod_{i \in \Theta_q}\left(1 - \alpha_{i,k}^q\right)\right)\prod_{r \neq q}\prod_{i \in \Theta_q}\left(1 - \alpha_{i,k}^q\right) \tag{15}$$

where $\mathbf{\Theta}_Q = \{\theta_1, \ldots, \theta_Q\}$ contains the indices provided by the Q information sources, $\mathbf{\Theta}_q$ is the subset of $\mathbf{\Theta}_Q$ corresponding to those reference spectrums belong to class ω_q, and Υ is a normalization factor

$$\Upsilon = \sum_{q=1}^{Q} \alpha_k^q \prod_{r \neq q}\left(1 - \alpha_k^r\right) + \prod_{q=1}^{Q}\left(1 - \alpha_k^q\right). \tag{16}$$

The final decision is made by assigning the ith voxel to the group $\omega_{q\,\max}$ with maximum credibility

$$\omega_{k\,\max} = \arg\max_k m\left(\{\omega_k\}\right). \tag{17}$$

By fusing the decisions made from the Q information sources, the different discrimination information provided by different information sources can be effectively exploited, and
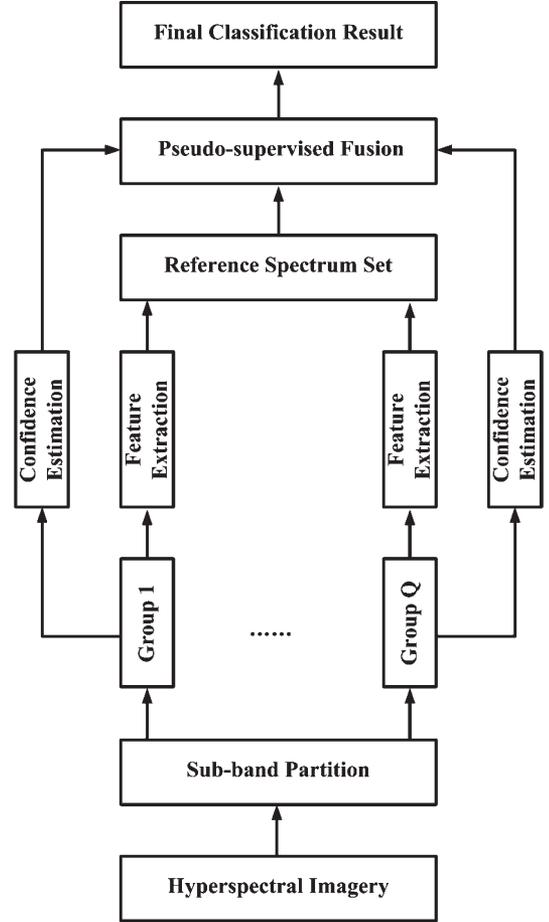

Fig. 1.   Flow chart of the proposed classification algorithm.

a more accurate classification result can be expected. On the other hand, as the classification is performed locally based on the joint spatial-spectral information, the classification map will be smoother than that by using only the spectral information. Fig. 1 shows the flow chart of the whole classification method. The following experiments validate the performance of the proposed algorithm.

IV. EXPERIMENTAL RESULTS

The data used in this experiment were recorded by the HYDICE sensor (16-bit BIL, 307 rows by 307 columns by 210 bands). There are five different materials in the scene, including asphalt, concrete, grass, trees, and soil. Fig. 2(a) is the false color image of the scene, and Fig. 2(b) shows the spectral characteristics of the five typical materials in the scene.

A. Hyperspectral Subband Source Generation

Fig. 3(a) shows the correlation coefficient matrix of the HYDICE image, based on which the hyperspectral bands can be partitioned. The correlation matrix shown in Fig. 3(a) contains bright off-diagonal blocks, which means that the corresponding bands are correlated. The correlation among bands in each diagonal block is the highest. The bright off-diagonal blocks also correspond to the intersection between two different diagonal blocks. We apply the Canny edge detector to Fig. 3(a), where

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

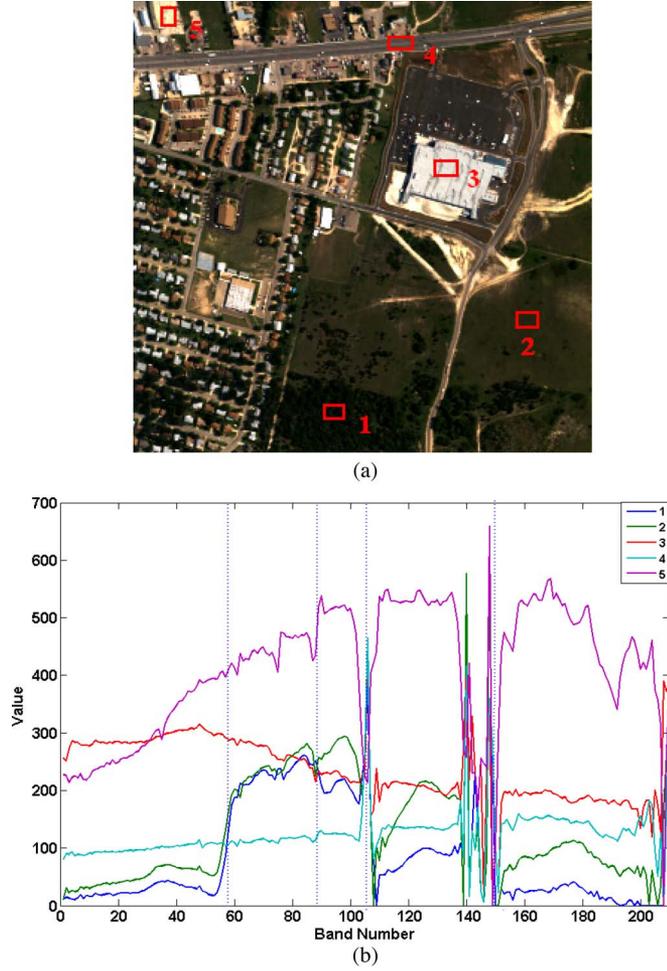ZHAO *et al.*: BAND-SUBSET-BASED CLUSTERING AND FUSION

5

(a)

(b)

Fig. 2. Scene of interest. (a) False color image of the scene (R: band 49, G: band 35, B: band 18). (b) The spectral characteristics of the five materials in the scene.
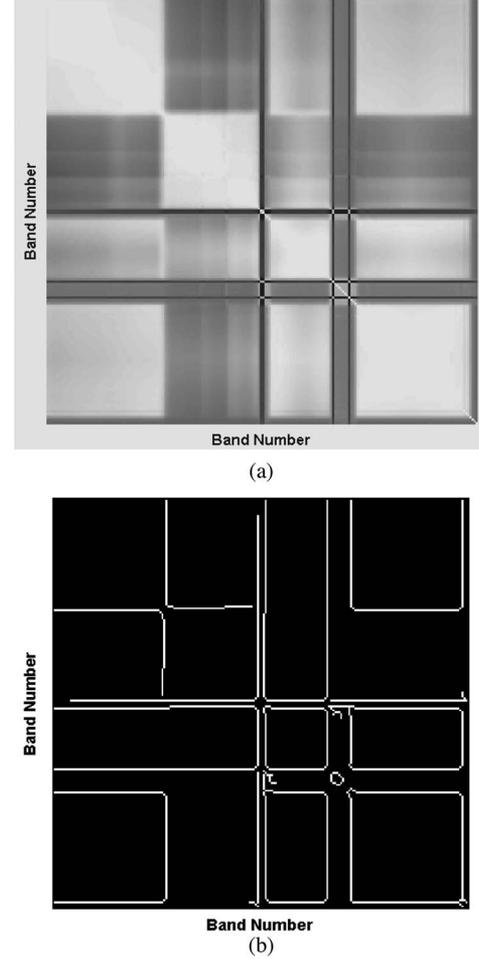

(a)

(b)

Fig. 3. (a) Correlation coefficient matrix of the HYDICE image. The brightness corresponds to the magnitude of the matrix elements. (b) The Canny edge detection result of the correlation coefficient matrix.

TABLE I
SEGMENTATION OF THE BANDS

Source 1	Source 2	Source 3	Source 4	Source 5
1~57	58~103	104~138	139~150	151~210

the correlation coefficient is scaled to $1 \sim 255$. The detected edge map is shown in Fig. 3(b). We use the edge to determine the number of subband source [26], [27]. If two edges are closer than a given threshold, these two edges will be merged. The threshold is set to preserve the original correlation of bands and remove the influence of noise. If the threshold is too small, the influence of noise cannot be removed efficiently and some subband sources will contain little useful classification information. If the threshold is too big, the original correlation of bands will be changed. Here, we set the threshold as 5 based on experience. Based on this principle, the hyperspectral data can be grouped into 5 subband sources, as listed in Table I. The numbers of bands in these subband sources are 57, 45, 34, 11, and 59, respectively. To reduce the volume of data before clustering and to obtain the same number of features for all the subsets, the principal component analysis (PCA) [4] is applied to each subband source.

For the data cube of the qth source Y_q, there are l_q images in this source, and then the data cube Y_q can be represented in a matrix form: $\overline{Y}_q = [\overline{Y}_{q,1}, \ldots, \overline{Y}_{q,l_q}]$. The covariance matrix of $\overline{Y}_q$ is

$$\Omega = \frac{1}{N}\overline{Y}_q \overline{Y}_q^T. \tag{18}$$

Let r_i and $\vec{e}_i$ be the ith eigenvalue and the associated eigenvector of the covariance matrix Ω and $r_1 \geq r_2 \geq \cdots \geq r_{lq}$. The projection $Y_q \cdot \vec{e}_i$ is called the ith principal component (PC) of Y_q. Since Y_q is redundant, in general by using only the several most significant PCs (i.e., those PCs corresponding to the most significant eigenvalues) of Y_q, most of the energy of Y_q can be preserved. Therefore, the number of bands in each source can be reduced by PCA transform. For the convenience of discussion, suppose that by using only three PCs, 99% of the energy of each source can be preserved. Then we project $\overline{Y}_q$ onto $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ to obtain the first three most significant PCs as $\overline{Y}_{q,\lambda_1} = \overline{Y}_q \cdot \vec{e}_1$, $\overline{Y}_{q,\lambda_2} = \overline{Y}_q \cdot \vec{e}_2$ and $\overline{Y}_{q,\lambda_3} = \overline{Y}_q \cdot \vec{e}_3$. Reformatting the three vectors into 2-D image format results in three images $\overline{Y}_{q,\lambda_1}$, $\overline{Y}_{q,\lambda_2}$ and $\overline{Y}_{q,\lambda_3}$.

B. Feature Extraction and Classification

In the feature extraction process, the bandwidth (h_s, h_p) should be selected carefully as they will determine the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
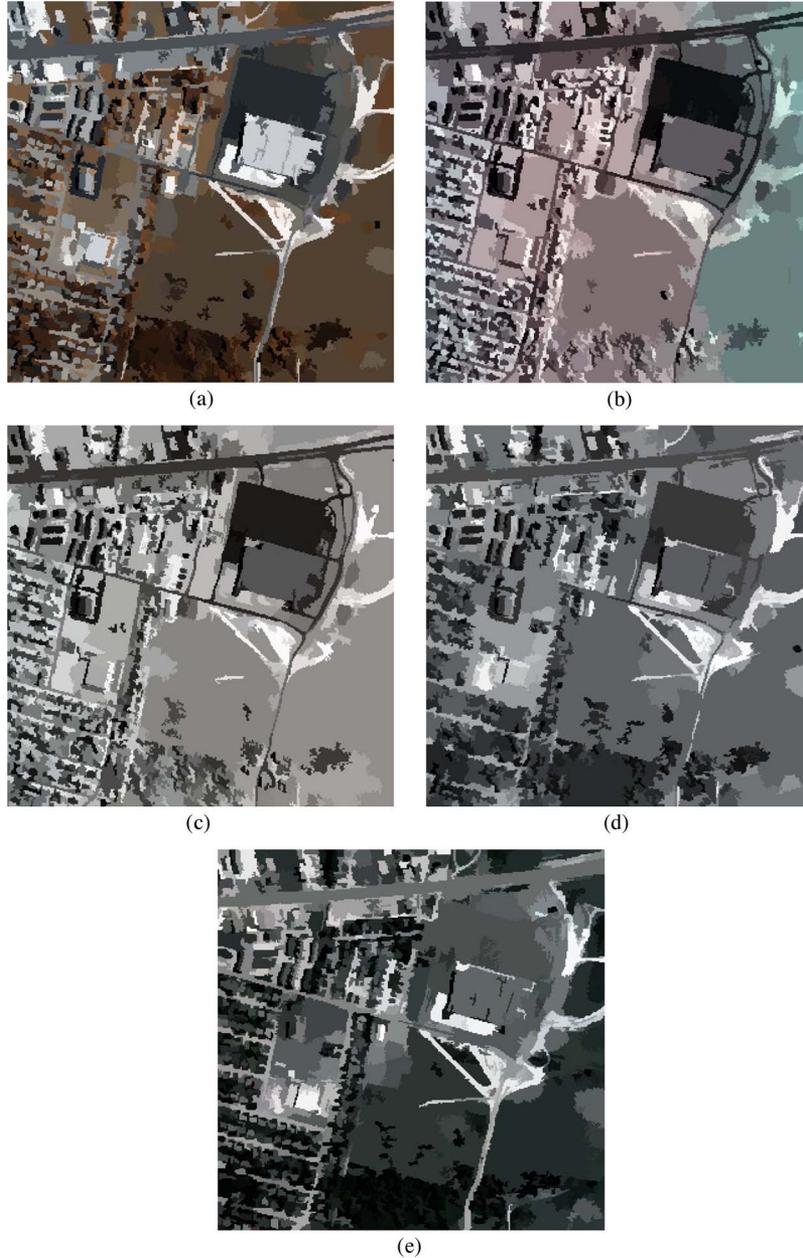

Fig. 4.   Mean shift clustering results of different subband groups. (a)–(e) are group 1 to group 5.

resolution of the mode detection. As in [12], features with large spatial support will be represented when h_s increases, and features with large spectral difference will be represented when h_p increases. Mean shift clustering is used to divide the scene into a series of homogeneous regions, and the average spectrum in every homogeneous region is regarded as training data. The setting of bandwidth depends on the data points and is decided by using pilot density estimation. The simplest way to obtain the pilot density estimate is the nearest neighbor method [29]. Let $x_{i,k}$ be the kth nearest neighbor of point x_i in spatial domain, then $h_s = \|x_i - x_{i,k}\|$. The number of neighbors should be chosen large enough to assure that there is an increase in density within the support of kernel having bandwidths h_s [28]. The bandwidth h_p is determined similarly. To make the classification accurate, the number of different materials in each

homogeneous region should be as small as possible. Based on this principle and to capture the small variation of texture and spectrum, we choose the parameters $h_s = 7$ and $h_p = 6.5$. The smallest region has 20 pixels. Fig. 4(a)–(e) shows the mean shift clustering results of the different groups.

There are 110 different homogeneous regions in five mean shift clustering subsets. Every homogeneous region can be labeled as one of the five different materials: asphalt, concrete, grass, trees, and soil. As described in Section III, the 550 average spectrums can be extracted from these homogeneous regions, and then clustered into five groups to form a training set $\mathbf{\Gamma} = \{(\mathbf{S}_{1,1}, \omega^1), \ldots, (\mathbf{S}_{r,q}, \omega^r), \ldots, (\mathbf{S}_{110,5}, \omega^{110})\}$. The mean shift clustering algorithm is used. As the training set is obtained from the clustering results, sometimes it may not match the real class, and this is the main source of classification

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHAO *et al.*: BAND-SUBSET-BASED CLUSTERING AND FUSION   7
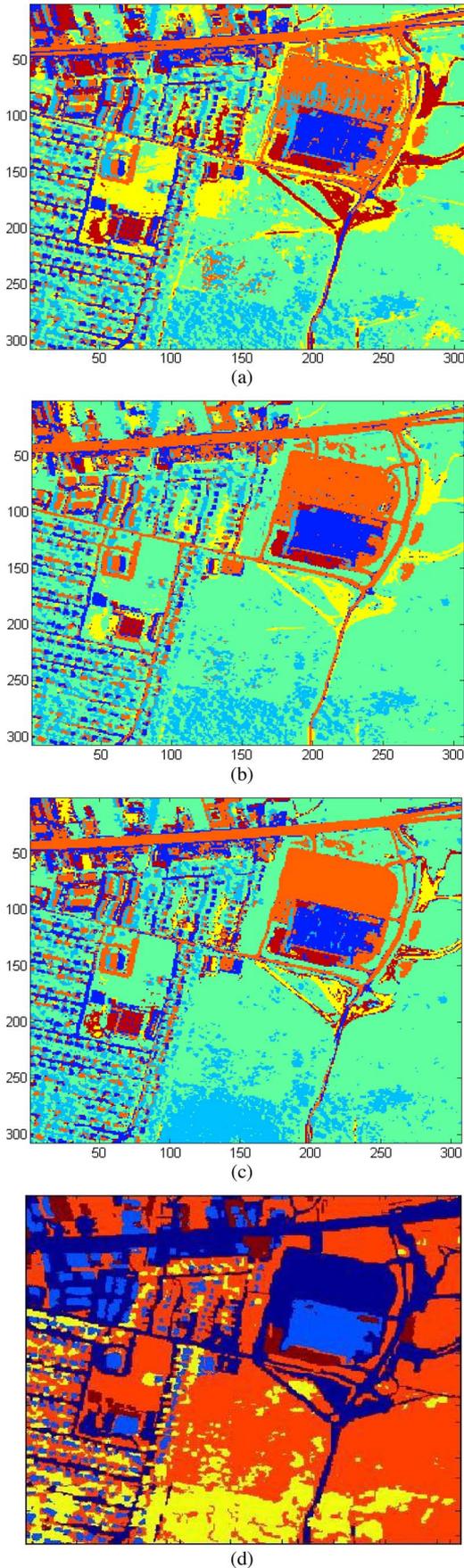
Fig. 5.   Classification results by using different methods. (a) FCM_S method. (b) KNN method. (c) Proposed method. (d) The ground truth.

TABLE II
CLASSIFICATION RESULTS BY DIFFERENT METHODS
ON THE HYDICE DATA

Classification Algorithm	OA	K-tree	K-asphalt	K-grass	$K_{clutter}$
FCM_S	79.6%	76.1%	82.7%	73.3%	0.20
KNN	90.7%	78.2%	87.1%	87.0%	0.18
Proposed Algorithm	88.5%	87.6%	87.8%	86.2%	0.10

TABLE III
CLASSIFICATION RESULTS BY DIFFERENT METHODS ON
WASHINGTON DC MALL DATA

Classification Algorithm	FCM_S	KNN	Proposed Algorithm
OA	**81.7%**	**89.6%**	**89.0%**
$K_{clutter}$	**0.17**	**0.15**	**0.14**

errors. On the other hand, as the method is unsupervised classification, the final class label is assigned based on the training set. The Mahalanobis distance is used to measure the spectral difference $d_{i,k}^q$ between the testing sample S and the training sample S_i

$$d_{i,k}^q = (\mathbf{S} - \mathbf{S}_{i,q})^T \boldsymbol{\Sigma}_k (\mathbf{S} - \mathbf{S}_{i,q}) \qquad (19)$$

where $\boldsymbol{\Sigma}_k$ is the covariance matrix of the samples in group k, $i = 1, \ldots, 110$ and $q = 1, \ldots, 5$. The hyperspectral imagery is classified by using the pseudosupervised fusion classification proposed in Section III. The final classification result is shown in Fig. 5(c).

C. Evaluation Measures

The proposed method is an unsupervised classification method that combines the spectral and spatial information in the classification process, and it also utilizes the idea of a supervised classifier. In order to illustrate the effectiveness of our method, the unsupervised classification method, which utilizes the spectral and spatial information, and the supervised classification method are used for comparison. Specifically, we compare the proposed method with the unsupervised fuzzy c-means clustering method with spatial constraints (FCM_S for short), which incorporates spatial information into the membership function [13], [22], [23], and the supervised KNN method [8]. Similar to that in the proposed method, PCA is used to reduce the dimension before performing FCM_S and KNN. Both the global measure and local measure as outline below are used to evaluate the performance.

1) *Global Measure* is used to characterize the classification accuracy [20]. The overall accuracy (OA) is employed to measure the labeling accuracy of the whole scene, and the kappa coefficient is employed to measure the correspondence of the labeling with three categories—trees, grass, and asphalt. The kappa coefficients of the three categories are represented as K-tree, K-grass, and K-asphalt.

2) *Local Measure* is used to characterize the classification smoothness. Smoothness in classification labeling can be measured from the viewpoint of clutter $K_{clutter}$ [18]. In a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
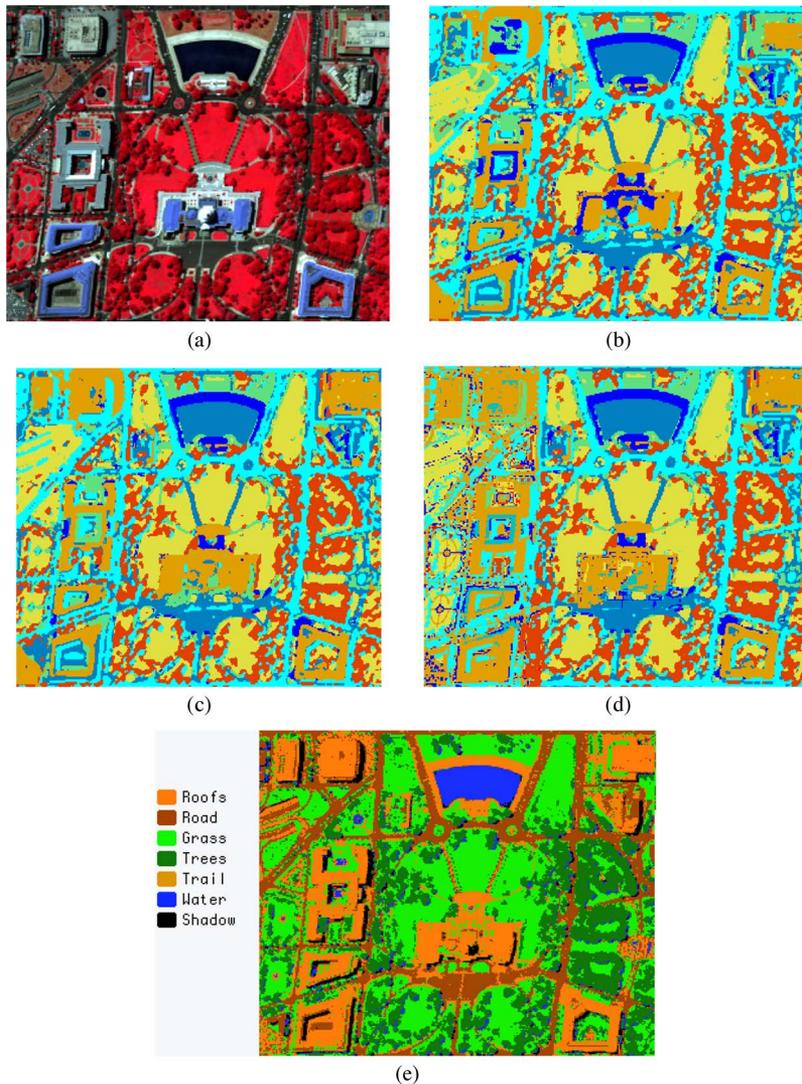

Fig. 6. Classification results of Scene Washington DC Mall by using different methods. (a) False color image of the scene (R: band 60, G: band 27, B: band 17). (b) FCM_S method. (c) KNN method. (d) The proposed method. (e) Ground truth.

homogenous region, all the pixels should be assigned to one label. Clutter can be defined as the number of pixels assigned to labels that are different from the predominant label in the region.

D. Results

Fig. 5(a) shows the results of FCM_S classification method. It can be seen that the classification result provides a poor partition of the scene. The concrete and soil are not well separated, the shade and asphalt are mixed, and there is a lot of clutter. Fig. 5(b) shows the results of KNN classification method (the training data is the average spectrum of 5×5 regions which is chosen based on ground truth). We see that the classification result is much better than the FCM_S classification. However, it is sensitive to noise in the forested areas and still has too much clutter. As shown in Fig. 5(c), the proposed classification method is more robust to the noise than previous classification methods, as it exploits the spatial and spectral information in the classification process.

Table II presents quantitative results of FCM_S clustering method, KNN classification method and the proposed algorithm. The proposed algorithm achieves good results in terms of classification accuracy. As an unsupervised scheme, it achieves almost the same global measures (OA, K-tree, K-grass, and K-asphalt) as the supervised KNN classification method. Table II also summarizes the measures of local smoothness $K_{clutter}$. There is clearly an improvement in homogeneity over the FCM_S and KNN algorithms by using the proposed clustering fusion algorithm.

The proposed algorithm is applied to a sample hyperspectral image that was taken over Washington DC Mall. The data consist of 260 (1001:1260) $\times$ 307 pixels with 210 bands, recorded with the HYDICE sensor. Fig. 6 shows the classification results using FCM_S, KNN, and the proposed classification method. The quantitative results are listed in Table III. By comparing Fig. 6(b) and (c) with Fig. 6(d) and results in Table III, we see that the proposed algorithm also achieved good results either in classification accuracy or in homogeneity.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHAO *et al.*: BAND-SUBSET-BASED CLUSTERING AND FUSION

9

TABLE IV
CLASSIFICATION RESULTS BY DIFFERENT METHODS
ON THE NOISY HYDICE DATA

Classification Algorithm	OA with different noise level			$K_{clutter}$ with different noise level		
	10 DB	**16 DB**	**20.13 DB**	**10 DB**	**16 DB**	**20.13 DB**
FCM_S	74.3%	70.0%	67.1%	0.26	0.33	0.41
KNN	86.9%	80.1%	73.4%	0.21	0.28	0.38
Proposed Algorithm	87.3%	86.8%	85.1%	0.11	0.12	0.15

E. Robustness to Noise

The hyperspectral imagery was corrupted by noise in the data acquisition process. The additive Gaussian white noise (AGWN) model is often used to model the instrumental photonic or electronic noise. To simplify the noise analysis, only AGWN is considered in the hyperspectral noise analysis process [21], [30]. To test the robustness to noise by different classification methods, AGWN is added to the hyperspectral image with peak signal-to-noise ratios (PSNRs) of 10, 16, and 20.13 dB, respectively. The classification results with PSNR of 20.13 dB are shown in Fig. 6. Both the FCM_S and KNN classifiers give poor partitioning of the scene in comparison to their segmentation on the original hyperspectral data. However, the classification results of the proposed method give almost the same segmentation as the original hyperspectral data. The quantitative results are listed in Table IV. We see that the proposed method achieves much better performance than the FCM_S and KNN classifiers. FCM_S and KNN perform classification at the pixel level, while the proposed method is a region-based classification scheme. The pixel-level classification is sensitive to noise, while region-based classification can reduce the errors introduced by noise through averaging. On the other hand, by comparing Fig. 7(a) and (b) with Fig 7(c), it can be concluded that even in a noisy condition, the proposed method can obtain better classification performance than FCM_S and KNN classifiers applied to the denoised sources. This conclusion can be validated by the quantitative results in Tables II and IV. As discussed, the proposed method is robust to noise.

V. CONCLUSION

In this paper, we proposed a clustering and fusion method to improve the classification performance of hyperspectral imagery. Due to the various external environment factors and internal factors of imaging system, the quality of hyperspectral imagery data varies across the spectral bands. Therefore, the segmentation of hyperspectral data should account for the quality variation of different bands. On the other hand, the spatial redundancy should also be exploited to improve the hyperspectral imagery segmentation performance. The proposed algorithm combines the spatial and spectral information by using the mean shift clustering, and then classifies the hyperspectral data by using a pseudosupervised fusion method. The proposed method was verified on the real HYDICE hyper-
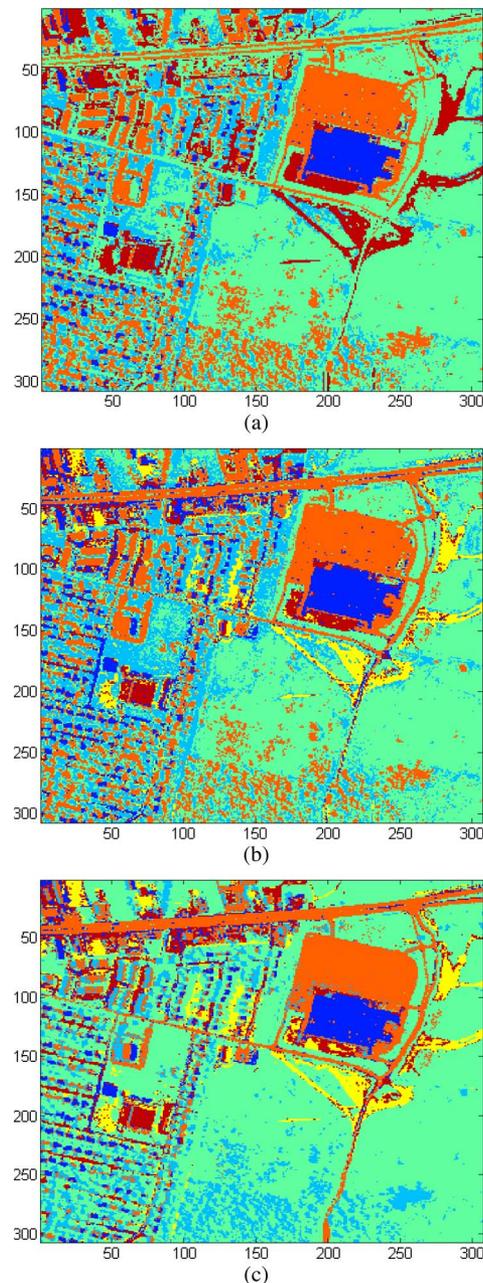

Fig. 7. Noisy hyperspectral data classification results by different methods. (a) The FCM_S method. (b) The KNN method. (c) The proposed method.

spectral imagery data. The experimental results demonstrated the robustness and the higher accuracy of the proposed classification method over the fuzzy c-means clustering with spatial constraints and KNN schemes.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] L. O. Jimenez, A. M. Morell, and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1360–1366, May 1999.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[4] V. Tsagaris, V. Anastassopoulos, and G. A. Lampropoulos, "Fusion of hyperspectral data using segmented PCT for color representation and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2365–2375, Oct. 2005.

[5] M. Zribi, A. Rekik, and M. Benjelloun, "Utilization of Dempster-Shafer theory of evidence in unsupervised image segmentation," *J. ASTM Int.*, vol. 4, no. 4, pp. 1–13, Apr. 2007.

[6] S. Prasad and L. M. Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, May 2008.

[7] Y. Cheng, "Mean shift, mode seeking and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.

[8] T. Denux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 5, pp. 804–813, May 1995.

[9] H. Wang and D. Bell, "Extend k-nearest neighbours based on evidence theory," *Comput. J.*, vol. 47, no. 6, pp. 662–672, 2004.

[10] U. Ozertem, D. Erdogmus, and R. Jenssen, "Mean shift spectral clustering," *Pattern Recognit.*, vol. 41, no. 6, pp. 1924–1938, Jun. 2008.

[11] K.-L. Wu and M.-S. Yang, "Mean shift-based clustering," *Pattern Recognit.*, vol. 40, no. 11, pp. 3035–3052, Nov. 2007.

[12] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[13] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825–838, Mar. 2007.

[14] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.

[15] A. D. Stocker, E. Ensafi, and C. Oliphint, "Application of eigenvalue distribution theory to hyperspectral processing," in *Proc. SPIE—Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imager IX*, 2003, vol. 5093, pp. 651–665.

[16] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[17] F. Dell'Acqua, P. Gamba, A. Ferrari, J. A. Palmason, J. A. Benediktsson, and K. Arnason, "Exploiting spectral and spatial information in hyperspectral urban data with high resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 322–326, Oct. 2004.

[18] R. S. Rand and D. M. Keenan, "Spatially smooth partitioning of hyperspectral imagery using spectral/spatial measures of disparity," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1479–1490, Jun. 2003.

[19] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[20] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, Apr. 2002.

[21] D. Letexier and S. Bourennance, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[22] J. Yang, S. Hao, and P. Chung, "Color image segmentation using fuzzy C-means and eigenspace projections," *Signal Process.*, vol. 82, no. 3, pp. 461–472, Mar. 2002.

[23] K. Chuang, H. Tzeng, S. Chen, J. Wu, and T. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imaging Graph.*, vol. 30, no. 1, pp. 9–15, Jan. 2006.

[24] T. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[25] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[26] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.

[27] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 274–281, Mar. 1994.

[28] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, vol. 1, pp. 456–463.

[29] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.

[30] E. Christophe, D. Léger, and C. Mailhes, "Comparison and evaluation of quality criteria for hyperspectral imagery," in *Proc. SPIE—Image Quality and System Performance II*, 2005, vol. 5668, pp. 204–213.

Yong-Qiang Zhao (M'05) received the B.S. degree in automation and the M.S. and Ph.D. degrees in control theory and control engineering from the Northwestern Polytechnic University, Xi'an, China, in 1998, 2001, and 2004, respectively.

From 2007 to 2009, he worked as a Postdoctoral Researcher at McMaster University, Hamilton, ON, Canada and Temple University, Philadelphia, PA, respectively. He is currently an Associate Professor with Northwestern Polytechnic University with research interests in polarization imaging analysis, imaging spectrometry, information fusion, and pattern recognition.

Lei Zhang (M'04) received the B.S. degree from the Shenyang Institute of Aeronautical Engineering, Shenyang, China, in 1995, and the M.S. and Ph.D. degrees in electrical and engineering from Northwestern Polytechnical University, Xi'an, China, respectively, in 1998 and 2001, respectively.

From 2001 to 2002, he was a Research Associate with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong. From January 2003 to January 2006, he was a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. Since January 2006, he has been an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University. His research interests include image and video processing, biometrics, pattern recognition, multisensor data fusion and optimal estimation theory, etc.

Seong G. Kong (SM'03) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1982 and 1987, respectively, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1991.

From 1992 to 2000, he was an Associate Professor of electrical engineering at Soongsil University, Seoul. He served as Chair of the department from 1998 to 2000. During 2000–2001, he was with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, as a visiting scholar. From 2002 to 2007, he was an Associate Professor in the Electrical and Computer Engineering Department, University of Tennessee, Knoxville. Since 2007, he has been an Associate Professor and Graduate Program Coordinator in the Electrical and Computer Engineering Department, Temple University, Philadelphia, PA.

Dr. Kong is a member of SPIE. He was the Editor-in-Chief of *Journal of Fuzzy Logic and Intelligent Systems* from 1996 to 1999 and an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS from 2003 to 2009. He received the Award for Academic Excellence from Korea Fuzzy Logic and Intelligent Systems Society (2000), Honorable Mention Paper Award from the American Society of Agricultural and Biological Engineers (2005), and the Most Cited Paper Award for the journal Computer Vision and Image Understanding (2007, 2008).