Band Subset Based Clustering and Fusion for Hyperspectral Imagery Classification[1]

Yongqiang Zhao, *Member, IEEE*, Lei Zhang, *Member, IEEE*, and Seong G. Kong, *Senior Member, IEEE*

Abstract—A band subset based clustering and fusion technique is proposed in this paper 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 pseudo-supervised hyperspectral classification scheme is developed by using evidence theory to fuse the information provided by each subset. The experimental results on real HYDICE hyperspectral imagery demonstrate that the proposed pseudo-supervised classification scheme can achieve higher accuracy than the spatially constrained fuzzy c-means clustering method, and it can achieve nearly the same accuracy as the supervised KNN classifier, and it also more robust to the noise.

Index Terms— Image segmentation, information fusion, hyperspectral, evidence theory

[1] Y. Zhao is with the College of Automation, Northwestern Polytechnical University, Shannxi 710072, China (e-mail: zhaoyq@nwpu.edu.cn).
L. Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: cslzhang@comp.polyu.edu.hk).
S.G. Kong is with the Department of Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA (e-mail: skong@temple.edu).

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, etc. [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. To improve the classification performance, it is necessary to investigate the spectra signature variation of hyperspectral data [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 dataset [1][2][3]. Actually, 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 utilize the spectral and spatial information jointly by using the Markov random field technique [18]. However, this method is computationally intensive because of the recursive or 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 sub-band cubes, each of which contains contiguous bands. These sub-bands are referred to as information sources. Channels within a sub-band cube are assumed to have similar noise

characteristics and discrimination ability. A non-parametric 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 uncertainty 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 wildly 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 allows representing the inaccuracy and uncertainty simultaneously using confidence, plausibility, and credibility functions [5][6]. Thus evidence theory could combine the items of evidence supporting certain hypotheses of a pattern from multiple information sources. In this paper, each sub-band 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 sub-band partition and confidence estimation. Section III presents a clustering fusion based classifier for hyperspectral imagery. Section IV describes experiments on HYDICE hyperspectral 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 methods. Section V concludes the paper.

II. Sub-band Partition in Hyperspectral Imagery

A. Sub-band 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: $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_l, ..., \mathbf{y}_L]^T$, where $\mathbf{y}_l$ is the $1 \times N$ row vector for the l^{th} 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. **Y** can be written as a linear mixture of target signals and background noise:

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

where $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_l, ..., \mathbf{x}_L]^T$ is the signal matrix and $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_l, ..., \mathbf{w}_L]^T$ 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. In actual application, however, 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 sub-band 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 i^{th} and the j^{th} 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 non-overlapped sub-band sets, which are viewed as nearly independent information sources. Denote by l_q the total number of bands in the q^{th} source, q=1,2,...,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 q^{th} source can be written as $\mathbf{Y_q} = \mathbf{X_q} + \mathbf{W_q}$. We assume that the noise level in the q^{th} 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 q^{th} source can be estimated by using eigenvalue analysis [21].

The spectral covariance matrix of the q^{th} source can be represented as:

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

where $\mathbf{\Lambda}_q = diag(\beta_1^2, \beta_2^2, \ldots, \beta_{l_q}^2)$ with $\beta_1^2 \geq \beta_2^2 \geq \ldots \geq \beta_{l_q}^2$ are the eigenvalues of noisy signal in the q^{th} source. The power of noiseless signal in the q^{th} 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 it by minimizing the Akaike information criterion (AIC) [16]:

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

where β_i , i=1,2,…, 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. Multi-Source Imagery Clustering Fusion

The hyperspectral imagery can be typically represented as a 2-dimensional 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-dimensional 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 mean shift as a feature extraction method to locate the cluster center by searching for the mode in the spatial-spectral domain.

After 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 pseudo-supervised hyperspectral classification method.

A. Feature Extraction by Mean Shift Clustering

The q^{th} sub-band set can be represented as N voxel points $\mathbf{Y}_n^q$, n=1,2,...,N in the l_q-dimensional space R^{l_q}, where l_q is the band number in the q^{th} sub-band 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 $K(\mathbf{y}) = k(\left\|\mathbf{y} - \mathbf{Y}_n^q\right\|^2)$. The PDF (probability density function) is estimated by the multivariate kernel density estimator:

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

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

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

Suppose that there exists a kernel $G: R^{l_q} \to R$ with $G(\mathbf{y}) = g(\left\|\mathbf{y} - \mathbf{Y}_n^q\right\|^2)$ 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:

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

Based on the above analysis, the feature extraction for the q^{th} sub-band set can be summarized as follows:

1) Initialize the n^{th} voxel as $\mathbf{y}_{n,1} = \mathbf{Y}_n^q$ and then update the voxel $\mathbf{y}_{n,j+1}$ according to (12) until $\left\|\mathbf{y}_{j+1} - \mathbf{y}_j\right\| < \varepsilon$, 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 $\{C_{p,q}\}_{p=1...m}$ by linking all $\mathbf{y}_{n,center}$ which are closer than a given threshold, where m is the number of clusters in the q^{th} sub-band set. The final feature extraction result is represented as a clustering map $C_q = \{C_{1,q},...,C_{m,q}\}$, and there are m homogeneous regions in this clustering map.

To utilize the spectral and spatial information jointly in the classification process, in the feature extraction process, a proper kernel should be chosen. It should be 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:

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

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 is used widely in classification, and 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, by controlling the size of the kernel, determines the resolution of the mode detection.

B. Pseudo-supervised Fusion

After performing feature extraction, there are Q clustering maps $C=\{C_1,C_2,...,C_Q\}$. Here we suppose 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 q^{th} sub-band set is named as the average spectrum and it is denoted by $\mathbf{S}_{r,q}$, where $r=1,...,R$ and $q=1,...,Q$. 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 pseudo-supervised fusion based classification method is proposed. The automatically labeled reference spectrum set is viewed as a training set $\mathbf{\Gamma}=\{(\mathbf{S}_{1,1},\omega^1),\ldots,(\mathbf{S}_{r,q},\omega^r),...,(\mathbf{S}_{R,Q},\omega^R)\}$ 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 $q=1,...,Q$. Suppose $n\in C_{p,q}$ is in the p^{th} homogeneous region of the q^{th} clustering map. To determine the class label of the i^{th} pixel, three situations are considered here:

1) If the Q homogeneous regions are the same, then the i^{th} pixel is assigned to the k^{th} class if the average spectrum belongs to the k^{th} group.
2) If the Q homogeneous regions are not the same but their average spectrums belong to the same group, say the k^{th} group, then the class label of the i^{th} 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 i^{th} pixel is determined by decision fusion, which is described as follows.

Suppose the average spectrum of homogeneous region in the q^{th} sub-band 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 i^{th} pixel can be regarded as the evidence that the i^{th} pixel belongs to class *k*. Based on this evidence and g_q, the confidence value of the q^{th} information source, the reliability function of the i^{th} pixel belonging to class *k* based on the q^{th} information source can be defined as:

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

where γ_q is a positive parameter associated to class ω_q. The reliability function α_k^q represents the reliability that the i^{th} location is labeled as ω_q based on the information provided by the q^{th} information source. While the reliability of the i^{th} 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 the Introduction section, 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 i^{th} pixel is label as ω_q can be calculated:

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

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

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

The final decision is made by assigning the i^{th} voxel to the group $\omega_{q\max}$ with maximum credibility:

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

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

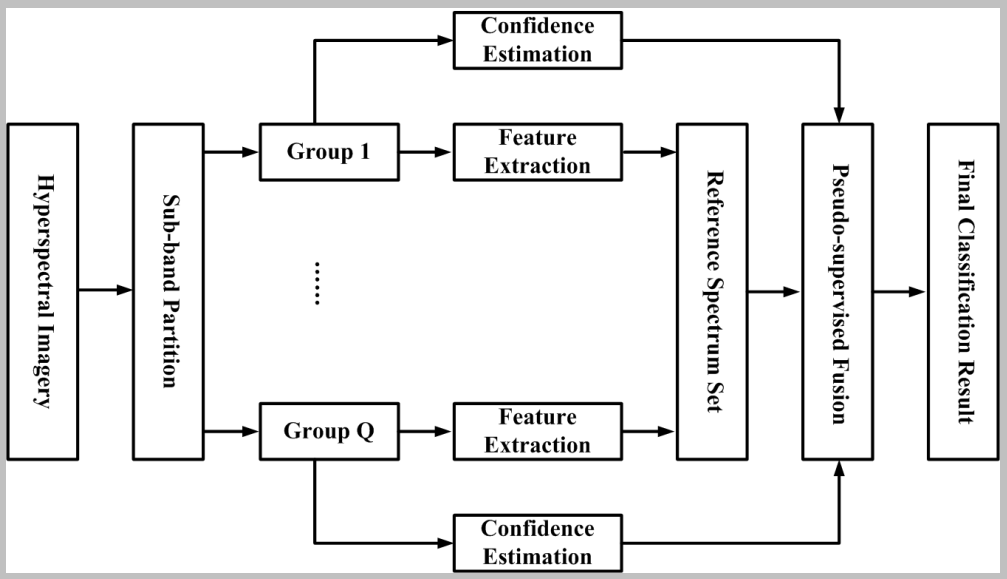
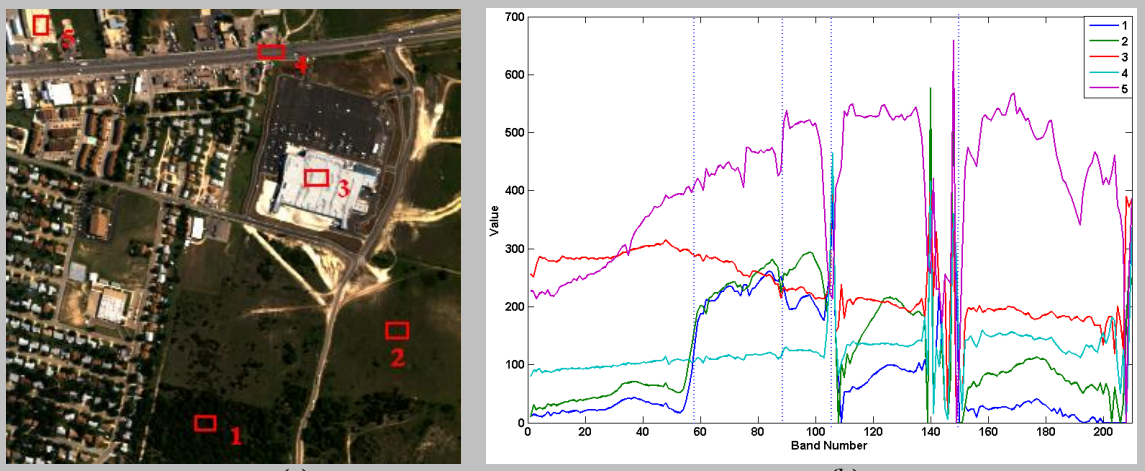


Figure 1. The flow chart of classification algorithm.

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 (b) shows the spectral characteristics of the five typical materials in the scene.

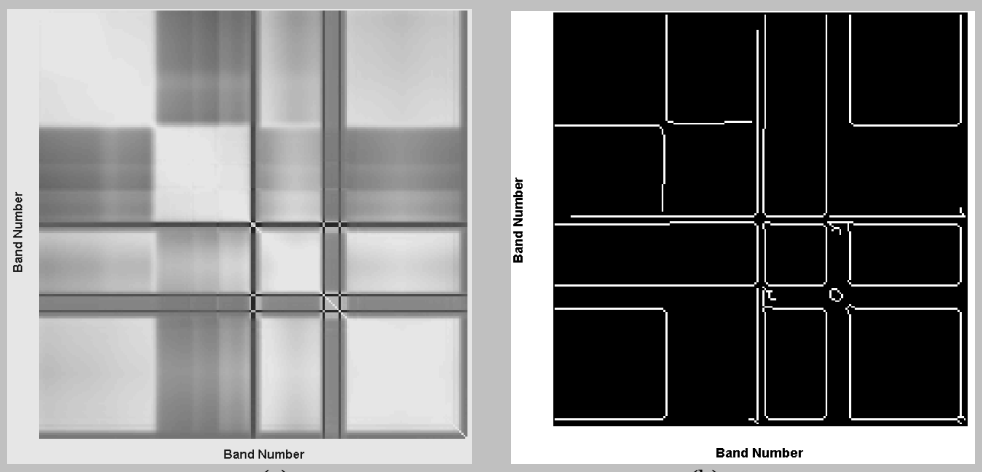


(a) (b)

Figure 2. The 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. Hyperspectral Sub-band Source Generation



(a) (b)

Figure 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

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 quite bright off diagonal blocks, which means the corresponding bands are correlated. But the correlation among bands in each diagonal block is 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 the correlation coefficient is scaled to 1~255 then Canny edge detector is applied to rescaled image. The detected edge map is shown in Fig. 3(b). Here we use the edge to determine the number of sub-band 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 it is too small, the influence of noise can not be removed efficiently and some sub-band sources will contain little useful classification information. If it is too big, the original correlation of bands will be changed. Here the threshold is set as 5. Based on this principle, the hyperspectral data

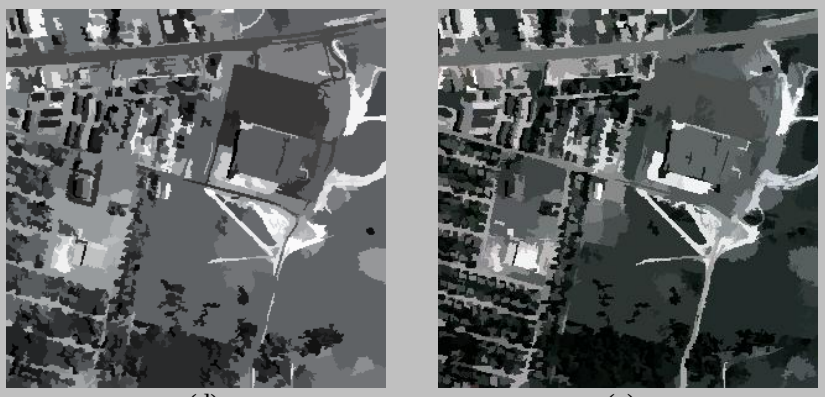
can be grouped into 5 sub-band sources, as listed in Table I. The numbers of bands in these sub-band sources are 57, 45, 34, 11, and 59, respectively. To reduce the volume of data before clustering and to obtain the same number of features for each q^{th} subset, a principal component transformation [4] is performed on each sub-band source.

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

$$\mathbf{\Omega} = \frac{1}{N}\bar{Y}_q\bar{Y}_q^T \tag{19}$$

Let and r_i and $\vec{e}_i$ be the i^{th} eigenvalue and the associated eigenvector of the covariance matrix $\mathbf{\Omega}$ and $r_1 \geq r_2 \geq \ldots \geq r_{lq}$. It is necessary to reduce the band number in every sub-band source, for example the band number in one sub-band source is reduced into three bands, we can perform the following steps. We choose the first three eigenvectors $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ that correspond to the first three most significant eigenvalues r_1, r_2 and r_3. Then we project $\bar{Y}_q$ onto $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ to obtain the first three most significant components as $\bar{Y}_{q,\lambda_1} = \bar{Y}_q \cdot \vec{e}_1$, $\bar{Y}_{q,\lambda_2} = \bar{Y}_q \cdot \vec{e}_2$ and $\bar{Y}_{q,\lambda_3} = \bar{Y}_q \cdot \vec{e}_3$. Reformatting the three vectors into 2D image format results in three images $\bar{Y}_{q,\lambda_1}$, $\bar{Y}_{q,\lambda_2}$ and $\bar{Y}_{q,\lambda_3}$. For every sub-band source, the rationale is set as 0.99 to decide the number of bands. The largest number of bands in these subsets is used to select the bands by PCA.

(a) (b) (c)



(d) (e)

Figure 4. Mean shift clustering results of different groups. (a) ~ (e) are group 1 ~ group 5.

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 resolution of the mode detection. As in [12], features with large spatial support will be represented when h_s increase, and feature with large spectral difference will be represented when h_p increase. Mean shift 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 bandwidth values associated with the data points, most of which use a pilot density estimate. The simplest way to obtain the pilot density estimate is by nearest neighbors[29]. Let $x_{i,k}$ be the k-nearest neighbor of the point x_i in spatial domain, then $h_s = \left\| x_i - x_{i,k} \right\|$. The number of neighbors k should be chosen large enough to assure that there is an increase in density within the support of most kernels having bandwidths h_s [28]. The same way can be used to determine the bandwidth h_p. In the simulation to make the classification accurate, every homogeneous region should contain the material's number as few 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 minimum region has 20 pixels. Fig.4. (a)-(e) show the mean shift clustering results of the different groups.

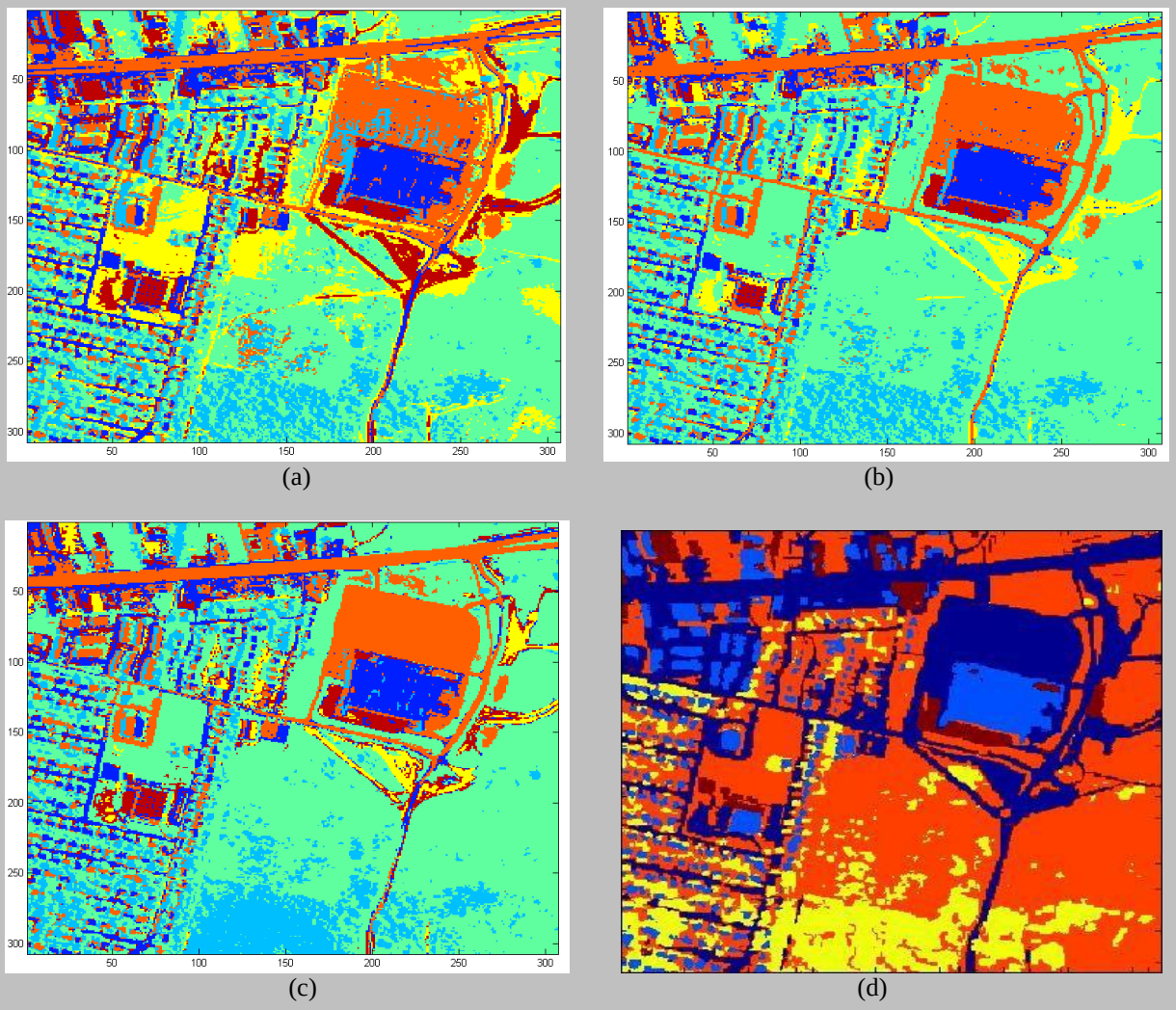


Figure 5. Classification results by using different methods: (a) FCM_S method; (b) KNN method; (c) the proposed method; (d) the ground truth.

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 3, the 550 average spectrums can be extracted from these homogeneous regions, and then clustered into 5 groups to form a training set $\mathbf{\Gamma} = \{(\mathbf{S}_{1,1}, \omega^1), \ldots, (\mathbf{S}_{r,q}, \omega^r), ..., (\mathbf{S}_{110,5}, \omega^{110})\}$. Here the mean shift clustering algorithm is used. As the training set is obtained from the clustering results, some times it can not match the real class, this is the main source of classification 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:

$$d_{i,k}^{q} = (\mathbf{S} - \mathbf{S}_{i,q})^{T} \boldsymbol{\Sigma}_{k} (\mathbf{S} - \mathbf{S}_{i,q}) \tag{18}$$

where $\boldsymbol{\Sigma}_{k}$ is the covariance matrix of the samples in group k, i=1,…,110 and q=1,...,5. The hyperspectral imagery is classified by using the pseudo-supervised fusion classification proposed in Section 3. The final classification result is shown in Fig. 5 (c).

C. Evaluation Measures

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 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

The propose method is an unsupervised classification method, it combines the spectral and spatial information in the classification process, it also utilizes the idea of supervised classifier. In order to illustrate the effectiveness of our method the unsupervised classification method, which utilizes the spectral and spatial information should be compared, and the supervised classification method should be also compared. We compare it 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]. PCA is used as a pre-processor before FCM_S and KNN to reduce dimension and noise, the rational is set as 0.99 to decide the number of bands. Both the global measure and local measure as outlined 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 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.

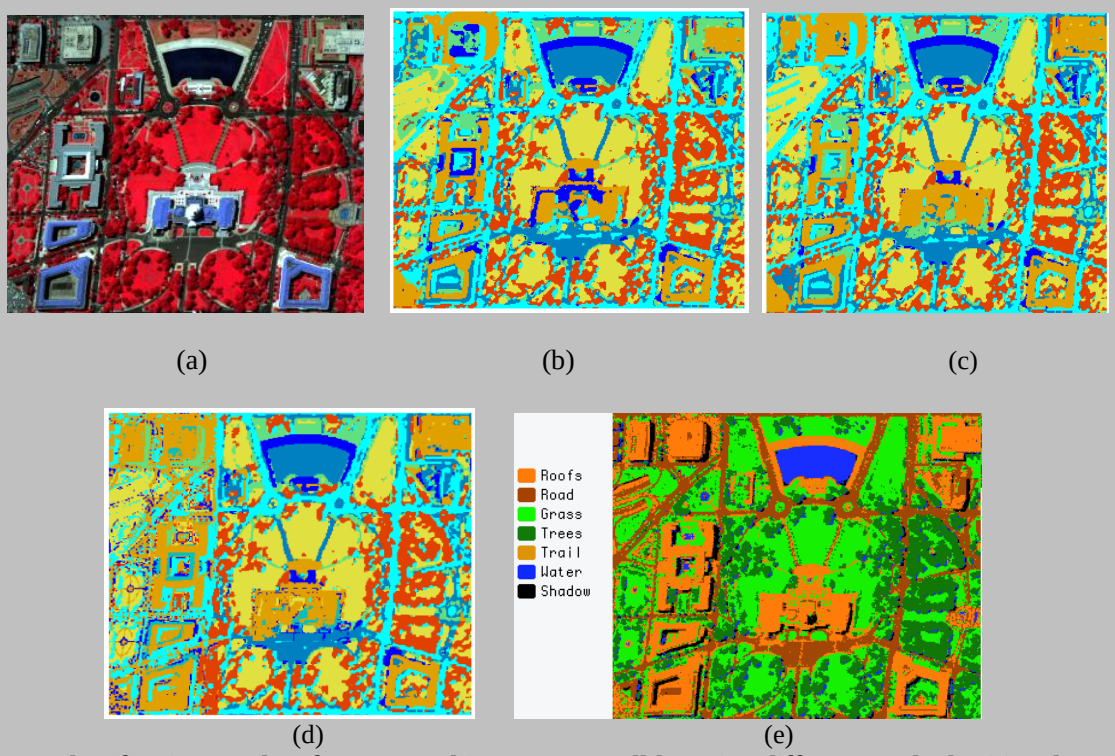
D. Results

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. It can be seen that the proposed algorithm achieves very good result in 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 which is taken over Washington DC Mall. The data consist of 260(1001:1260) × 307 pixels with 210 bands, was 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 Ⅳ. By comparing Fig. 6(b) and (c) with (d) and results in Table Ⅳ, it can be seen that the proposed algorithm also achieves good results in either in classification accuracy or in homogeneity.



(a) (b) (c)

(d) (e)

Figure 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) the ground truth.

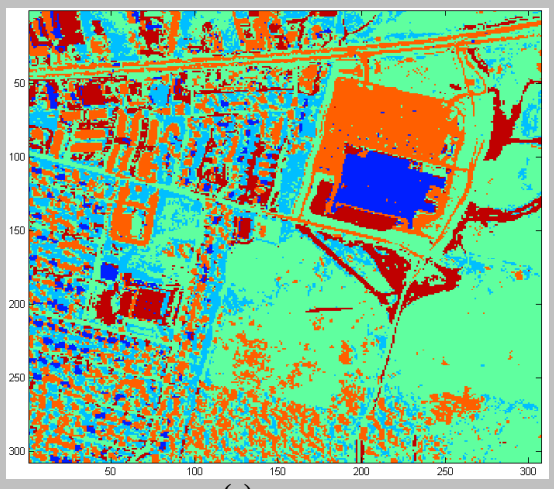
TABLE Ⅳ CLASSIFICATION RESULTS BY DIFFERENT METHODS ON WASHINGTON DC MALL DATA

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

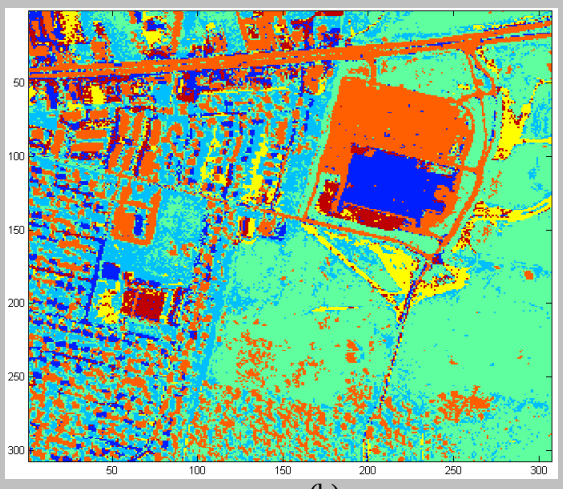
E. Robustness to Noise

The hyperspectral imagery was corrupted by noise in the data acquisition process, as the additive Gaussian white noise models the instrumental photonic or electronic noise. To simplify the noise analysis, only additive Gaussian white noise will be considered in the hyperspectral noise

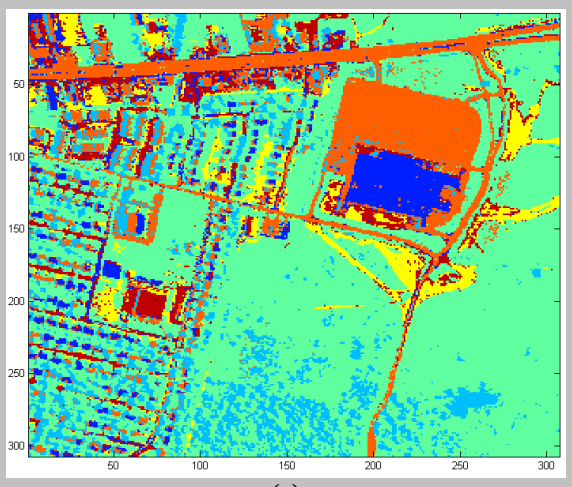
analysis process [21]. To test the robustness to noise by different classification methods, an additive Gaussian white noise is added to the hyperspectral image with peak signal-to-noise ratios of 10db, 16db, and 20.13 dB. The classification results with peak signal-to-noise ratio = 20.13 dB are shown 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 by the proposed method gives almost the same segmentation as the original hyperspectral data. The quantitative results are listed in Table III. We see that the proposed method achieves much better performance than the FCM_S and KNN classifiers. FCM_S and KNN employ a pixel-by-pixel classification scheme, while the proposed method is based on the region-based classification. The pixel-by-pixel classification scheme is sensitive to noise, while region-based classification can reduce the errors introduced by noise through averaging. On the other hand, by comparing the Fig. 7(a) and Fig. 7(b) with Fig. 7(c), it can be conclude 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 can be proved by the quantitative results in Table II and Table III. As discussed, the proposed method is robust to the noise.



(a)



(b)



(c)

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

V. Conclusions

In this paper, we presented 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 pseudo-supervised fusion method. The proposed method was verified on the real HYDICE hyperspectral 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.

Acknowledgments

This work is supported by the Natural Science Foundation of China under grants No.60602056 and 60634030, and the State Key Laboratory of Remote Sensing Opening Fund under grant No.SK050013.

Reference

[1] D. Landgrebe, “Hyperspectral image data analysis,” *IEEE Signal Processing Magazine*, Vol. 19, no. 1, pp. 17-28, 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 Transactions on Geoscience and Remote Sensing*, Vol. 37, no. 3, pp.538-542. 1999.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-Based Methods for Hyperspectral Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, no. 6, pp.1351-1362, 2005.

[4] V. Tsagaris, V. Anastassopoulos, and G. A. Lampropoulos, "Fusion of hyperspectral data using segmented PCT for color representation and classification," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, no. 10, pp.2365-2375, 2005.

[5] M. Zribi, A. Rekik and M. Benjelloun, "Utilization of Dempster-Shafer Theory of Evidence in Unsupervised Image Segmentation," *Journal of ASTM International*, Vol. 4, no. 4, pp.1-13, 2007.

[6] S. Prasad and L.M. Bruce, "Decision Fusion With Confidence-Based Weight Assignment for Hyperspectral Target Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 46, no. 5, pp.1448-1456, 2008.

[7] Y. Cheng, "Mean shift, mode seeking and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, no. 8, pp.790-799, 1995.

[8] T. Denceux, "A k-nearst neighbor classification rule based on Dempster-Shafer theory," *IEEE Transactions on Systems, Man. And Cybernetics*, Vol. 25, no.5, pp.804-813, 1995.

[9] H. Wang and D. Bell, "Extend k-nearest neighbours based on evidence theory," *The Computer Journal*, Vol. 47, no.6, pp.662-672, 2004.

[10] U. Ozertem, D. Erdogmus and R. Jenssen, "Mean shift spectral clustering," *Pattern Recognition*, Vol.41, no.5, pp.1924-1938, 2008.

[11] Kuo-L. Wu and Miin-S. Yang, "Mean shift-based clustering," *Pattern Recognition*, Vol.40, no.5, pp.3035-3052, 2007.

[12] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, no. 5, pp.603-619, 2002.

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

[14] W. Tao, H. Jin and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts", *IEEE transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol.37, no. 5, pp.1382-1389, 2007.

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

[16] C. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, no. 3, pp.608-619, 2004.

[17] F. D. Acqua, P. Gamba, A. Ferrari, *et al.*, "Exploiting Spectral and Spatial Information in Hyperspectral Urban Data With High Resolution," *IEEE Transactions on Geoscience and Remote Sensing Letters*, vol. 1, no. 4, pp. 322–326, 2004.

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

[19] J. M. Bioucas-Dias, and J. M. P. Nascimento, "Hyperspectral Subspace Identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.

[20] G. M. Food, "Status of Land Cover Classification Accuracy Assessment," *Remote Sensing of Environment*, Vol.80, no.1, pp. 185-201, 2002.

[21] D. Letexier and S. Bourennance, "Noise removal from hyperspectral images by Multidimensional filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 7, pp. 2061–2069, 2008.

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

[23] K. Chuang, H. Tzeng, S. Chen, *e. al*, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, Vol. 30, no. 1, pp. 9-15, 2006.

[24] T. Bandos , L. Bruzzone , G. Camps-Valls. "Classification of Hyperspectral Images With Regularized Linear Discriminant Analysis", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 862 - 873, 2009.

[25] H. Xin and Z. Liangpei "An Adaptive Mean-Shift Analysis Approach for Object Extraction and Classification From Urban Hyperspectral Imagery", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 12, pp. 4173 - 4185, 2008.

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

[27] X. Jia and J.A. Richards. "Efficient maximum likelihood classification for imaging spectrometer data sets", IEEE Trans. on Geoscience and Remote Sensing, vol. 32, no,2, pp.274-281, 1994.

[28] B. Georgescu, I. Shimshoni, P. Meer. “Mean shift based clustering in high dimensions: a texture classification example”, Ninth IEEE International Conference on In Computer Vision, vol. 1, pp. 456-463, 2003.

[29] B.W.Silverman. “Density Estimation for Statistics and Data Analysis”. Chapman & Hall, 1986

[30] E. Christophea, D. Légera and C. Mailhesb. “Comparison and evaluation of quality criteria for hyperspectral imagery”, SPIE Proceedings on Image quality and system performance II, vol. 5668, 2005:204-213