Semisupervised Band Clustering for Dimensionality Reduction of Hyperspectral Imagery

Hongjun Su, He Yang, *Student Member, IEEE*, Qian Du, *Senior Member, IEEE*, and Yehua Sheng

***Abstract*—Band clustering is applied to dimensionality reduction of hyperspectral imagery. Different from unsupervised clustering using all the pixels or supervised clustering requiring labeled pixels, the proposed semisupervised band clustering needs class spectral signatures only. After clustering, a cluster selection step is applied to select clusters to be used in the following data analysis. Initial conditions and distance metrics are also investigated to improve the clustering performance. The experimental results show that the proposed algorithm can outperform other existing methods with lower computational cost.**

***Index Terms*—Band clustering, dimensionality reduction, hyperspectral imagery, *k*-means clustering.**

I. Introduction

DIMENSIONALITY reduction is a commonly used preprocessing technique for hyperspectral image analysis [1]. It can be achieved by a transformation method [e.g., principal component analysis (PCA) and linear discriminant analysis (LDA)], where the original high-dimensional data are projected onto a low-dimensional space with a certain criterion. Dimensionality reduction can also be achieved by band selection, whose objective is to find a small subset of bands containing important data information.

Another approach is band grouping (BG) or band clustering. BG can be viewed as a special case of band extraction. For instance, adjacent bands can be grouped together, and a representative of each group can be selected to participate in the following data analysis. Intuitively, adjacent bands can be partitioned uniformly [denoted as BG(U)] or based on spectral correlation coefficient [denoted as BG(CC)]. Fig. 1 shows a 202×202 spectral correlation coefficient (CC) matrix, where a bright pixel at location (i, j) means high correlation between the ith and jth bands; if the pixel is in dark, then the correlation is low. The white blocks along the diagonal line indicate that adjacent bands usually have high correlation and should be grouped together. However, if examining Fig. 1 carefully, we can see that nonadjacent bands may also have high correlation; this is verified by the presence of white blocks in off-diagonal areas. Thus, nonadjacent bands should be allowed to be grouped together. In this letter, we use "band clustering" to represent the methods that can group nonadjacent bands and "BG" to represent the methods that group adjacent bands only.

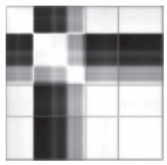

Fig. 1. Spectral CC matrix for a 202-band AVIRIS Indian Pines image scene.

Clustering algorithms have been applied to hyperspectral remote sensing data analysis. The typical implementation is to cluster pixels based on the spectral signatures so as to spatially segment an image scene into many subregions. For instance, a clustering-based anomaly detection approach was proposed in [2], where a clustering method, such as ISODATA, k-means, and self-organizing map, was adopted to model image pixels. In [3], a fuzzy-clustering algorithm that spatially exploited class membership relations was presented. A particle-swarm-optimization-based algorithm was proposed for pixel clustering to estimate class statistical parameters and the number of classes in image data in [4]. A modified k-means algorithm was developed to improve matched filter performance in [5].

Another type of implementation of a clustering algorithm is in the spatial domain; in other words, a spectral band is converted into a vector after column or row stacking, and then, these band vectors are clustered into several groups based on their similarity. In [6], two clustering methods, namely, Ward's linkage strategy using mutual information (WaLuMI) and Ward's linkage strategy using divergence (WaLuDi), were developed, and finalized clusters were further used for band selection.

In this letter, we will focus on k-means-based band clustering for dimensionality reduction [8]. One of its drawbacks is that it is sensitive to initial condition and may be trapped in local optima; different initial conditions may produce different clusters. In this letter, we will propose a new initial technique using band selection output. Because of unsupervised nature, k-means clustering may be time consuming when using all the pixels. In [7], k-means was extended to a supervised version, where training samples for each class were required for clustering. However, in practice, it may be difficult to obtain enough training samples; instead, it may be possible to have a spectral signature for each class. Therefore, in this letter, we

Manuscript received October 12, 2010; revised December 20, 2010, March 7, 2011, and April 8, 2011; accepted May 24, 2011. Date of publication June 23, 2011; date of current version October 28, 2011. The work of H. Su was supported in part by the China Scholarship Council under Grant 2009686004, in part by the National Natural Science Foundation of China under Program 40901200, and in part by the Outstanding Postgraduate Dissertation Cultivating Program of Nanjing Normal University under Grant 1243211601040.

H. Su and Y. Sheng are with the Key Laboratory of Virtual Geographic Environment of the Chinese Ministry of Education, School of Geography Science, Nanjing Normal University, Nanjing 210046, China (e-mail: hjsu1@163.com; shengyehua@njnu.edu.cn).

H. Yang and Q. Du are with the Department of Electrical and Computer Engineering and the Geosystem Research Institute in High Performance Computing Collaboratory, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: hy58@msstate.edu; du@ece.msstate.edu).

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/LGRS.2011.2158185

propose a semisupervised k-means (SKM) clustering method that uses class signatures only. Here, a class signature is the representative spectrum of a class.

Note that, in [8], k-means was applied for band selection using class signatures. Our method is different because of the following reasons.

1) We use band cluster centers for the following data analysis (e.g., detection and classification), while the band closest to a cluster center is selected in [8]. We will show that using cluster centers is better than using selected bands.
2) We conduct cluster selection, while in [8], all the clusters are used. We will show that deleting the worst cluster will provide better performance.

In the clustering process itself, our algorithm is also different because of the following reasons.

1) We show that different distance metrics will yield different clustering results. In our experiments, CC is better than others, such as Euclidean distance in [8].
2) The initial condition is critical to the clustering performance, and our band selection result in [9] can be used as initials, providing better performance than random initials in [8].

II. Proposed Method

A. SKM Band Clustering

Given a set of bands $(\mathbf{B}_1, \ldots, \mathbf{B}_l, \ldots, \mathbf{B}_L)$, where each band is arranged into N-dimensional vector, where N is the number of pixels. k-means band clustering aims to partition the L bands into k clusters $\mathbf{C} = \{c_1, \ldots, c_m, \ldots, c_k\}$ $(1 \leq m \leq k)$ so as to minimize the following objective function:

$$\arg\min_{\mathbf{C}} \sum_{m=1}^{k} \sum_{\mathbf{B}_l \in c_m} D(\mathbf{B}_l, \boldsymbol{\mu}_m) \tag{1}$$

where $\boldsymbol{\mu}_m$ is the cluster center of c_m and $D(\bullet, \bullet)$ is a distance metric gauging the similarity between a band and the center of the cluster that it is assigned to. This original algorithm with random initials is denoted as RKM. Its computational complexity is linearly proportional to the number of pixels N. In order to reduce the complexity, we use class signatures as algorithm input; then, the complexity becomes linearly proportional to the number of signatures S $(S \ll N)$. This approach is denoted as SKM. Several distance metrics can be adopted for k-means, including Euclidean distance (L_2), city-block distance (L_1), cosine (spectral angle), and spectral CC.

The SKM algorithm is initialized by using distinctive bands as cluster centroids. The idea of unsupervisedly selecting distinctive bands was presented in [9]. The band selection algorithm is initialized by choosing a pair of bands B_1 and B_2, leading to a band subset $\Phi = \{B_1, B_2\}$; it then finds a third band B_3 that is the most dissimilar to all the bands in the current Φ by using a certain criterion, resulting in an updated subset $\Phi = \Phi \cup B_3$; and the selection step is repeated until the number of bands in Φ is large enough. Here, a linear prediction (LP) error (i.e., the difference between an original band and its linear predicted version using bands in Φ) is employed as the similarity metric. A band with the maximum LP error is the most dissimilar band from those in Φ and should be selected. Note that the RKM can be initialized with this unsupervised band selection method as well, which is denoted as UKM.

After k-means clustering, k clusters with their centroids are ready for further analysis. However, it does not mean that all of them should be used. Some clusters may not be helpful for object classification, and they may even bring about confusion. Thus, we propose to remove a cluster by exhaustively searching for the worst one (when it is removed, the remaining clusters provided the most similar classification maps to those from using all the original bands). It is observed that deleting one cluster usually results in improvement, but deleting more than one cluster may not necessarily provide further improvement. Thus, only one cluster is removed hereafter. The SKM algorithm deleting the worst cluster is denoted as SKMd.

SKMd-based band clustering can be detailed as follows.

1) Initialize the algorithm by using k selected distinctive bands.
2) Using the known class signatures, conduct band clustering based on CC. The k-means clustering is completed when no band is shuffled from one cluster to another. The center of each cluster is used as the representative.
3) Conduct classification when each cluster center is removed in turn. If the removal of a specific cluster center yields the largest accuracy, then this cluster will be removed permanently. The resulting $k-1$ clusters are the final output.

B. Automatic Removal of the Worst Cluster

In addition to exhaustive searching (ES), as mentioned in Section II-A, a cluster may be deleted based on a criterion, such as orthogonal projection divergence (OPD) [10]. Let $\mathbf{c}_i$ and $\mathbf{c}_j$ denote the ith and jth cluster centroids, respectively. Their OPD value is defined as

$$\mathrm{OPD}(\mathbf{c}_i, \mathbf{c}_j) = \left(\mathbf{c}_i^{\mathrm{T}} \mathbf{P}_{\mathbf{c}_j}^{\perp} \mathbf{c}_i + \mathbf{c}_j^{\mathrm{T}} \mathbf{P}_{\mathbf{c}_i}^{\perp} \mathbf{c}_j\right)^{1/2} \tag{2}$$

where $\mathbf{P}_{\mathbf{c}_m}^{\perp} = \mathbf{I} - \mathbf{c}_m(\mathbf{c}_m^{\mathrm{T}}\mathbf{c}_m)^{-1}\mathbf{c}_m^{\mathrm{T}}$ for $m = i$, j and $\mathbf{I}$ is an identity matrix. $\mathbf{P}_{\mathbf{c}_j}^{\perp}$ is the orthogonal subspace of $\mathbf{c}_j$, and $\mathbf{c}_i^{\mathrm{T}}\mathbf{P}_{\mathbf{c}_j}^{\perp}\mathbf{c}_i$ is the squared norm of the projection of $\mathbf{c}_i$ onto $\mathbf{P}_{\mathbf{c}_j}^{\perp}$. Similarly, $\mathbf{c}_j^{\mathrm{T}}\mathbf{P}_{\mathbf{c}_i}^{\perp}\mathbf{c}_j$ is the squared norm of the projection of $\mathbf{c}_j$ onto $\mathbf{P}_{\mathbf{c}_i}^{\perp}$. A larger OPD value means that $\mathbf{c}_i$ and $\mathbf{c}_j$ are more different.

For k cluster centroids, each pair of OPD value is computed. A cluster will be removed if the average OPD to other $k-1$ cluster is the largest. Based on our experience, this cluster usually has lower image quality. In other words, bands included in such a cluster generally have low signal-to-noise ratio (SNR).

C. Computational Complexity and Methods for Comparison

Table I lists the computational complexity of different methods during band clustering process. For the SKM, it is $O(LSkt)$, compared to $O(LNkt)$ in the RKM, where t is the number of iterations. The complexity of SKM is also much lower than those of WaLuMI and WaLuDi. In WaLuMI and WaLuDi, mutual information or Kullback–Leibler divergence is used as similarity metric, hierarchical clustering is conducted, and a representative band from each final cluster is chosen.

TABLE I
COMPUTATIONAL COMPLEXITY OF BAND CLUSTERING

Method	Number of Multiplications
SKM	$O(LSkt)$
RKM	$O(LNkt)$
WaLuMI	$O(L^2N)+O(L^3)$
WaLuDi	$O(LN+L^2G)+O(L^3)$

G: the number of gray scale levels

TABLE II
LIST OF METHODS FOR COMPARISON

Method	Description
Band Clustering	
RKM	k-means using all data with random initials
UKM	k-means using all data with unsupervisedly selected bands as initials
SKMd	**semi-supervised k-means using class signatures with selected bands as initials; delete the worst cluster for classification**
SKM	semi-supervised k-means using class signatures with selected bands as initials
SKMd (CC)	semi-supervised k-means(correlation) based on signatures with cc-grouped bands as initials; delete the worst cluster for classification
SKM (CC)	semi-supervised k-means(correlation) based on signatures with cc-grouped bands as initials
Band Grouping	
BG (U)	Band grouping using uniform spacing bands
BG (CC)	Band grouping based on correlation coefficients
Band Selection	
SKM (BS)	Same as SKM except that the bands closest to cluster centers are selected.
SKMd (BS)	Same as SKMd except that the bands closest to cluster centers are selected.
WaLuMI	Hierarchical clustering using all data
WaLuDi	Hierarchical clustering using all data
Linear Transformation	
PCA	PCA for all data
LDA	LDA for all data
Others	
All bands	All bands for classification

In addition, BG(U) that simply groups the same number of adjacent bands and BG(CC) that groups adjacent bands based on spectral CC are used for comparsion. PCA and LDA are implemented for comparison when training samples are available. A list of methods for comparison, including the variants of k-means, is presented in Table II.

III. EXPERIMENTS

Three real-data experiments were conducted. Clustering quality can be evaluated with classification accuracy. When training and test samples are available, support vector machine (SVM) can be applied. If only class signatures are available, then a method that does not require the training process, such as orthogonal subspace projection (OSP) [11], can be used. The classification maps are compared with those using all the original bands, and the similarity is assessed with spatial CC; a larger value of average spatial CC means better performance.

A. AVIRIS Lunar Lake Experiment

The same 200 × 200 AVIRIS Lunar Lake image scene as in [11] was used in this experiment. After the water absorption and low-SNR bands were removed, 158 bands were left. Five classes are present: cinders, playa, rhyolite, shade, and vegetation. Thus, five signatures were used for k-means clustering.

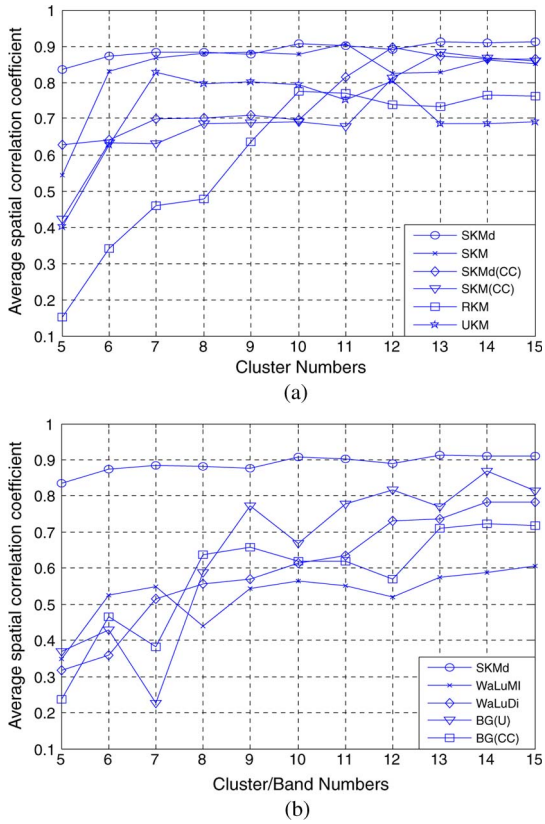

Fig. 2. Comparison with different methods in Lunar Lake experiment. (a) Compared with k-means related methods. (b) Compared with other methods.

Because no training samples are available, OSP was used for classification, and spatial CC with the classification maps from using all the original bands was considered as accuracy. As shown in Fig. 2(a), SKMd provided the best result, which was better than SKM using all the resulting clusters for classification. With CC-grouped cluster centroids as initials, SKMd(CC) and SKM(CC) were worse than their counterparts SKMd and SKM, respectively, which means that the band selection result was a better choice as the initial condition. RKM was the average result from 50 runs of the k-means with random initials; it could not outperform UKM with the selected bands as the initial. As shown in Fig. 2(b), SKMd was compared with WaLuMI, WaLuDi, BG(U), and BG(CC), where it was significantly better. The simplest BG(U) offers a better result than WaLuMI, WaLuDi, and BG(CC).

B. HYDICE DC Mall Experiment

The HYDICE subimage scene with 304 × 301 pixels over the Washington, DC Mall area was used as in [12]. After bad-band removal, 191 bands were used in the experiment. In these data, there are six classes: roof, tree, grass, water, road, and trail. These six class centers were used for band clustering. The overall accuracy from SVM was computed with the available training and test samples [12].

As shown in Fig. 3(a), SKMd provided the best result among all the k-means variants, and its performance was always better

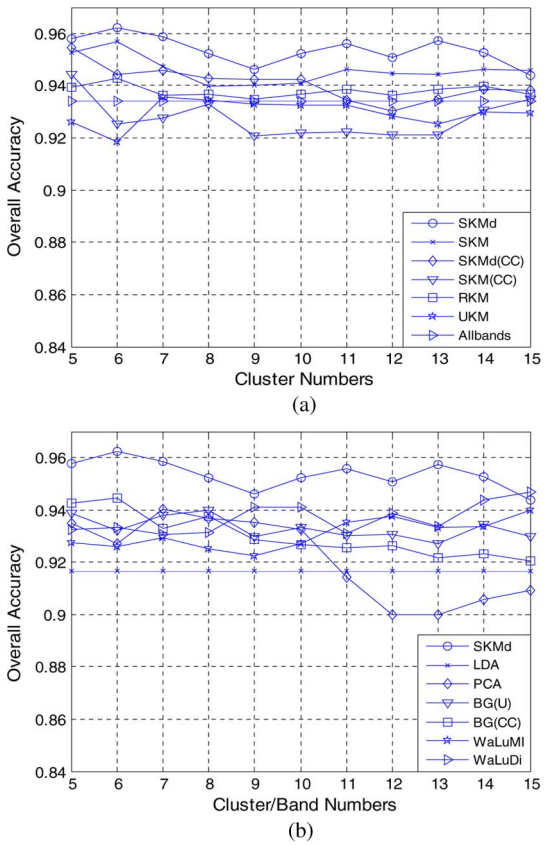

Fig. 3. Comparison with different methods in DC mall experiment. (a) Compared with k-means related methods. (b) Compared with other methods.

than that using all the bands regardless of the number of clusters. All other k-means variants performed quite similarly, while SKM(CC) yielded the worst result. Fig. 3(b) shows the performance of other BG and clustering approaches, and SKMd outperformed BG(U), BG(CC), WaLuMI, WaLUDi, PCA, and LDA. Once again, the performance of PCA could be better than LDA. In this experiment, increasing the value of k did not improve the accuracy; instead, the highest accuracy appeared when $k = 6$.

C. AVIRIS Indian Pines Experiment

The AVIRIS subimage scene taken over northwest Indiana's Indian Pines with 145 $\times$ 145 pixels and 202 bands was used, and 16 different land-cover classes were present based on the ground truth [13]. Since class samples are available, the overall classification accuracy was computed for evaluation using SVM outputs.

The purpose of this experiment is to demonstrate the performance discrepancy when using different similarity metrics for clustering and when using cluster centers or the bands closest to the cluster centers in [8]. As shown in Figs. 4 and 5, CC provided better performance than other metrics such as Euclidean distance (L_2), cosine (spectral angle), and city block (L_1) in both SKM and SKMd. Fig. 6 shows that using cluster centers is better than using selected bands.

The nonparametric McNemar test was employed to evaluate the statistical significance in accuracy improvement with the proposed methods [14]. It is based on the standardized normal

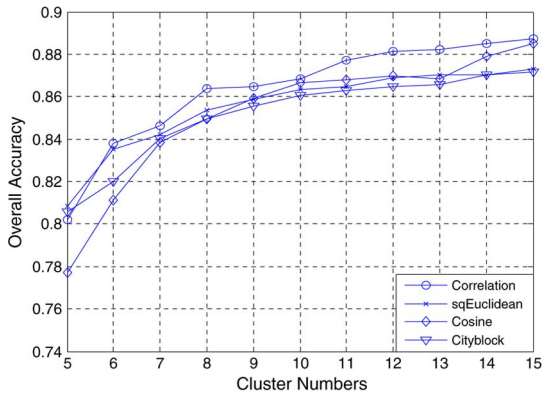

Fig. 4. Classification accuracy using different similarity metrics for SKMd in Indian Pines experiment.

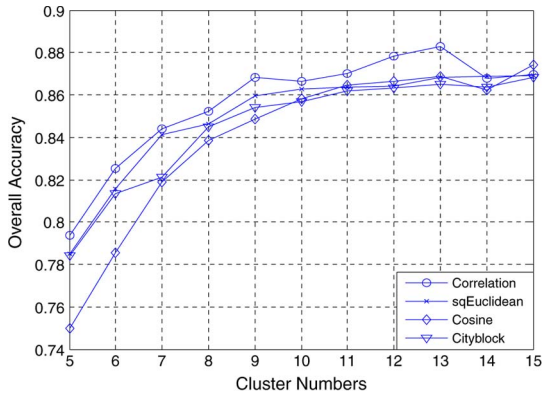

Fig. 5. Classification accuracy using different similarity metrics for SKM in Indian Pines experiment.

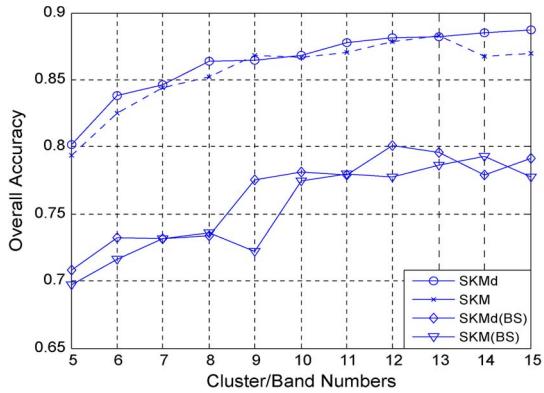

Fig. 6. Classification accuracy using cluster centers or selected bands in Indian Pines experiment.

test statistic. For two methods to be compared, let f_{11} denote the number of samples that both methods can correctly classify, f_{22} the number of samples that both cannot, f_{12} the number of samples misclassified by method 1 but not method 2, and f_{21} the number of samples misclassified by method 2 but not method 1. Then, the McNemar's test statistic for these two methods can be defined as

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}. \tag{3}$$

For 5% level of significance, the corresponding $|z|$ value is 1.96; a $|z|$ value greater than this quantity means that two

TABLE III
Z VALUES IN THE MCNEMAR'S TEST FOR AVIRIS INDIANA PINES DATA (THE 5% LEVEL OF SIGNIFICANCE IS SELECTED)

	SKMd	
	$\|z\|$	Significant?
RKM	22.18	Yes
UKM	17.69	Yes
SKM	2.52	Yes
SKMd (CC)	5.82	Yes
SKM (CC)	4.32	Yes
SKMd (BS)	20.35	Yes
SKM (BS)	21.98	Yes
WaLuMI	22.90	Yes
WaLuDi	12.73	Yes
BG (U)	4.54	Yes
BG (CC)	4.07	Yes

TABLE IV
CLASSIFICATION ACCURACY WITH BAD CLUSTER REMOVAL

	k	5	7	9	11	13	15
Lunar Lake	ES	0.84	0.88	0.88	0.90	0.91	0.91
	OPD	0.84	0.88	0.88	0.82	0.86	0.89
DC Mall	ES	0.96	0.96	0.95	0.96	0.95	0.94
	OPD	0.95	0.96	0.95	0.96	0.95	0.94
Pines	ES	0.80	0.85	0.86	0.88	0.88	0.89
	OPD	0.76	0.83	0.86	0.87	0.88	0.89

methods have significant performance discrepancy. Table III tabulates the average $|z|$ values when SKMd was compared against other methods, with k being changed from 5 to 15. Obviously, the performance of the proposed SKMd is statistically different from others, but the discrepancy between SKMd and SKM is less than other pairs.

D. Automatic Bad Cluster Removal

The OPD-based cluster removal in Section II-B was also implemented for SKMd. As listed in Table IV, the results were slightly degraded compared to ES. It is useful when the number of clusters k is larger. ES needs to repeat the process of classification and evaluation for K times.

E. Computing Time

To further compare the computational complexity in addition to Table I, the computing times when the algorithms run in a personal computer with 2.26-GHz CPU and 4.0-GB memory were recorded and listed in Table V. We can see that SKM can save significant amount of time compared to the traditional RKM, WaLuMI, and WaLuDi. Note that the running time spent by SKM does not include the time for band selection (as the initial condition), so it can approximately represent for the running time of SKM(CC); using different initials do not have much impact on the convergence speed of the SKM-based algorithms in these experiments.

IV. CONCLUSION

Band clustering has been investigated for hyperspectral dimensionality reduction. By allowing nonadjacent bands to be clustered together, its performance is better than those grouping adjacent bands only. Different from unsupervised clustering using all the pixels or supervised clustering requiring labeled pixels, the proposed semisupervised band clustering needs class

TABLE V
COMPUTING TIMES OF DIFFERENT ALGORITHMS (IN SECONDS)

	k	5	10	15
Lunar Lake	SKM	**0.75**	**0.75**	**0.86**
	RKM	119.59	188.48	210.54
	WaLuMI	182.06	183.50	185.72
	WaLuDi	177.77	179.90	188.87
DC Mall	SKM	**3.27**	**3.24**	**3.77**
	RKM	419.89	464.26	662.06
	WaLuMI	534.83	541.19	548.92
	WaLuDi	547.78	560.76	555.15
Pines	SKM	**20.16**	**18.30**	**24.08**
	RKM	998.85	2068.78	2366.58
	WaLuMI	209.73	219.28	232.48
	WaLuDi	284.83	260.57	262.17

spectral signatures only, thereby significantly reducing computational cost. After clustering, a cluster selection step can further improve the following data analysis performance; ES or the OPD-based automated cluster removal can be adopted. The experimental results also showed that preselected distinctive bands can be a good choice as algorithm initial, and the distance metric plays a role in the clustering performance.

However, the proposed algorithm does require class signatures. These can be obtained by prior information. In practice, these signatures can be extracted directly from the image scene using an endmember extraction algorithm.

REFERENCES

[1] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.
[2] O. Duran and M. Petrou, "A time-efficient method for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3894–3904, Dec. 2007.
[3] G. Bilgin, S. Ertürk, and T. Yildirim, "Unsupervised classification of hyperspectral image data using fuzzy approaches that spatially exploit membership relations," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 673–677, Oct. 2008.
[4] A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4175–4188, Dec. 2009.
[5] C. C. Funk, J. Theiler, D. A. Roberts, and C. C. Borel, "Clustering to improve matched filter detection of weak gas plumes in hyperspectral thermal imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1410–1420, Jul. 2001.
[6] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[7] S. H. Al-Harbi and V. J. Rayward-Smith, "Adapting k-means for supervised clustering," *Appl. Intell.*, vol. 24, no. 3, pp. 219–226, Jun. 2006.
[8] B. Mojaradi, H. Emami, M. Varshosaz, and S. Jamali, "A novel band selection method for hyperspectral data analysis," in *Proc. Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, 2008, pp. 447–451.
[9] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.
[10] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA: Kluwer, 2003.
[11] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
[12] H. Yang, Q. Du, and B. Ma, "Data fusion on supervised and unsupervised classifiers for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 875–879, Oct. 2010.
[13] [Online]. Available: http://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
[14] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.