2012-5-10

Adaptive Affinity Propagation with Spectral Angle Mapper for Semi-Supervised Hyperspectral Band Selection

Hongjun Su, *et al.*

Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection

Hongjun Su,[1,2] Yehua Sheng,[3] Peijun Du,[2,*] and Kui Liu[4]

[1]School of Earth Sciences and Engineering, Hohai University, 1 Xikang Road, Nanjing 210098, China

[2]Department of Geographical Information Science, Nanjing University, 22 Hankou Road, Nanjing 210093, China

[3]Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, 1, Wenyuan Road, Nanjing, Jiangsu 210046, China

[4]Department of Electrical Engineering, the University of Texas at Dallas, 800 West Campbell Road, Richardson, Texas 75080-3021, USA

*Corresponding author: dupjrs@126.com

Received 1 July 2011; revised 9 March 2012; accepted 12 March 2012; posted 14 March 2012 (Doc. ID 150380); published 10 May 2012

Band selection is a commonly used approach for dimensionality reduction in hyperspectral imagery. Affinity propagation (AP), a new clustering algorithm, is addressed in many fields, and it can be used for hyperspectral band selection. However, this algorithm cannot get a fixed number of exemplars during the message-passing procedure, which limits its uses to a great extent. This paper proposes an adaptive AP (AAP) algorithm for semi-supervised hyperspectral band selection and investigates the effectiveness of distance metrics for improving band selection. Specifically, the exemplar number determination algorithm and bisection method are addressed to improve AP procedure, and the relations between selected exemplar numbers and preferences are established. Experiments are conducted to evaluate the proposed AAP-based band selection algorithm, and the results demonstrate that the proposed method outperforms other popular methods, with lower computational cost and robust results.

OCIS codes: 100.2960, 100.5010, 300.6170.

1. Introduction

Hyperspectral image contains hundreds of spectral bands with very fine spectral resolution, providing a powerful data source for many applications. But its high dimensionality and difficulty in *a priori* information collection bring problems in data transmission and supervised application, which also presents a challenge to many traditional image analysis algorithms. To mitigate these problems, dimensionality reduction technology, which eliminates band numbers without decreasing critical information, has been widely used in hyperspectral image analysis [1]. This reduction in the hyperspectral data could be done by feature extraction or feature (band) selection [2–3]. In feature extraction, the original high-dimensional data is projected into a low-dimensional space with a certain criterion, but in band selection, its objective is to find a small subset of bands containing important data information.

For feature extraction, since the data are transformed, some critical information may be compromised and distorted, and the number of dimensionality is difficult to estimate. Band selection is preferable for hyperspectral image analysis because the critical information is preserved from the original data. In this paper, we focus on the band selection algorithm. However, band selection also suffers from two similar issues encountered in the feature

1559-128X/12/142656-08$15.00/0

extraction. One is the number of bands that should be selected in order to preserve necessary information. Another is the criterion to be used for band selection. For the first issue, the new concept of virtual dimensionality has been proposed and used to estimate the number of spectrally distinct signatures in the hyperspectral image [4], and it has achieved success in many hyperspectral applications [5]. For the band selection criterion, because of the unavailability of labeled information, many unsupervised band selection algorithms have been addressed in recent years. For example, liner prediction-based band selection [6], maximum-variance principal components analysis (MVPCA) and maximum-signal-to-noise principal components analysis (MSNRPCA) [7], linearly constrained minimum variance (LCMV)-based band correlation minimization (BCM) [LCMV (BCM)] and linearly constrained minimum variance (LCMV)-based band correlation constraint (BCC) [LCMV(BCC)] [8], particle swarm optimization [9], and some distance metrics-based methods [10].

From the other perspective, band selection algorithm can be viewed as a process of data clustering, which partitions the dataset into groups (clusters) and identifies the center of each cluster [11–12]. If the cluster centers are the data that come from the original space, they can be selected as the representative bands. All the representative bands are more informative and independent than the other bands in the original dataset by the clustering-based process [12–13]. Hence, certain clustering algorithms can be used for band selection under this assumption. Recently, Frey and Dueck proposed a new clustering algorithm, affinity propagation (AP) [14–15]. AP is initiated by simultaneously considering all data points as potential exemplars (cluster centers that can be viewed as the interesting bands) and exchanging real-valued messages between data points, at last the clusters are formed by assigning each data point to its most similar exemplar. Such message-passing methods have been shown remarkably efficient in hyperspectral unmixing and band selection [16–17]. But for the AP algorithm, an important problem to be addressed is that the number of exemplars is influenced by the value of preferences, that is, low preferences lead to a small number of exemplars, while high preferences lead to a large number of exemplars. So how to get a fixed number of clusters (exemplars) is still an open problem.

In this paper, we present an adaptive AP (AAP) clustering algorithm with spectral angle mapper (SAM) for hyperspectral band selection, in which an AAP algorithm which provides a new way to get the fix number of exemplars is addressed.

The rest of this paper is organized as follows. Section 2 presents the proposed AAP algorithm for band selection in detail and introduces some approaches for performance evaluation. We present experimental results in Section 3 with three hyperspectral data, and conclusions are drawn in Section 4.

2. Proposed Band Selection Method

A. AP Clustering

In order to use AP as the band selection algorithm, we begin with a brief review of this model proposed by Frey and Dueck for clustering data points [14]. Suppose that there is a dataset of centered random vector $x \in \mathbf{R}^n$ with N observations $x_i, i \in [1, \ldots, N]$ For affinity propagation, the real-valued similarity collection $\{s(i,k)\}$ between data points was taken as inputs, where the similarity $s(i,k)$ indicates how well the data point with index k is suited to be the exemplar for data point i. Same as other clustering algorithms, AP aims at finding a set of data points $C = \{c_1, c_2, \ldots, c_k\}$ in the original space $X = \{x_1, x_2, \ldots, x_N\}$ for each data point i as its exemplar. A value of $c_i = k$ for $i \neq k$ indicates that data point i is assigned to a cluster with point k as its exemplar, and $c_i = k$ indicates that data point k serves as a cluster exemplar. When the goal is to minimize squared error, the common negative Euclidean distance can be computed by

$$s(i,k) = -\|x_i - x_k\|^2. \tag{1}$$

In $s(k,k)$, for each data point k, which with larger values of $s(k,k)$ are more likely to be chosen as exemplars, these values are referred to as "preferences," which play important roles in determining the number of exemplars; thus, low preferences lead to a small number of exemplars, and high preferences lead to a large number of exemplars. In most cases, the value could be the median of the input similarities or their minimum. Also, AP can be viewed as searching through valid configurations of the labels $C = \{c_1, c_2, \ldots, c_n\}$ to minimize the energy:

$$E(c) = -\sum_{i=1}^{N} s(i, c_i). \tag{2}$$

On the other hand, AP is a message-passing algorithm. There are two kinds of messages: "responsibility" and "availability," which are exchanged between data points. In the algorithm, messages can be combined at any stage to decide which points are exemplars and, for every other point, which exemplar it belongs to. The responsibility $r(i,k)$ is the message sent from data point i to candidate exemplar point k, reflecting the accumulated evidence for how well-suited point k is as the exemplar for point i. The availability $a(k,k)$ is the message sent from candidate exemplar k to data point i, indicating how appropriate it is that point i would choose candidate k as its exemplar.

In the first place, all the availability initializes from zero. Then, the responsibility and availability will update by the following rules [14]:

$$r(i,k) \leftarrow s(i,k) - \max_{k' s.t. k' \neq k} \{a(i,k') + s(i,k')\}, \tag{3}$$

$$a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i's.t.i' \in \{i,k\}} \max\{0, r(i',k)\}\right\}. \quad (4)$$

It should be noted that, for $k = i$, the self-responsibility $r(k,k)$ is set to the input preference at the point k is chosen as an exemplar, $s(k,k)$, minus the largest of the similarities between point i and all other candidate exemplars. The self-availability $a(k,k)$ is updated differently:

$$a(k,k) \leftarrow \sum_{i's.t.i' \neq k} \max\{0, r(r',k)\}. \quad (5)$$

The above updating rules require only simple, local computation and can be easily implemented with Eq. (4), and messages need only be exchanged between pairs of points with known similarities. The exemplar of each point i is determined by

$$c(i) = \max_k (a(i,k) + r(i,k)). \quad (6)$$

The message-passing procedure may be terminated by an iteration procedure, with changes in the messages falling below a threshold. However, in practice, the updating rules in Eq. (3) and Eq. (4) often lead to oscillations caused by "overshooting" the solution, so the responsibility and availability messages are always "dumped" as follows:

$$M = \lambda M_{\text{old}} + (1 - \lambda) M_{\text{new}}, \quad (7)$$

where M_{new} and M_{old} are the message values from the previous and current iterations respectively;λ is the damping factor ranging from 0 to 1. Frey and Dueck suggest that the damping factor $\lambda \geq 0.5$ is the best value for experiments, which is considered as the convergence of both speed and stability.

B. AAP as Band Selection

Suppose that a hyperspectral dataset consists of L matrix X_i, $1 \leq i \leq L$, where L is the number of wavebands, and each vector has $M \times N$ dimensions. In order to use the AP algorithm, the similarity matrix $S \in R^{L \times L}$ of hyperspectral image data should be provided, in which the element $s(i,k)(i \neq k)$ measures how well the band k can represent the band i. As mentioned in the AP algorithm, different from other clustering algorithms such as k-means, which requires the number of clusters be prespecified, AP takes a real number $s(k,k)$ as input for each data point k so that the preferences with larger values of $s(k,k)$ are more likely to be chosen as exemplars. We can see that the value of preferences can determine the number of clusters. So in order to get the identify band numbers, we established relations between the selected band number k and preferences:

$$\text{pref} = h - 10^i(h - l). \quad (8)$$

First, we set $i = -4$; h and l are the highest and lowest value in similarity matrix, respectively. Then we execute the AP algorithm many times (i is increased by 1 for each iteration, so the value of preference is changed in each step) until the number of exemplars is lower than k.

Second, we use the bisection method to find fixed number k. The bisection method is a root-finding algorithm that recursively bisects an interval, then for further processing, selects a subinterval in which a root must lie. This processing step is relatively simple and robust. In the bisection method, the AP algorithm was repeatedly executed by changed preferences using the following expression until we get the exactly k clusters:

$$\text{pref} = 0.5h - 0.5l. \quad (9)$$

Also, we set the damping factor as $\lambda = 0.9$. In the experiments, the fixed numbers of exemplars from 5 to 25 were produced based on Eq. (8) and Eq. (9). So, a novel band selection method that combines adaptive k selection approach and the AP algorithm was put forward as the following steps: AAP. The input, output, and steps of the AAP are summarized as follows.

Input:

$s(i,k)$: Similarity of point i to point k.
k: How many bands are selected?
$p(j)$: Preferences array indicating the preference that data point j is chosen as a cluster center.

Output:

idx(j): Index of the cluster center for data point j.
dpsim: Sum of the similarities of the data points to their cluster centers.
expref: Sum of the preferences of the identified cluster centers.
netsim: Net similarity (sum of the data point similarities and preferences).
pref: Final value of preferences for k bands.

Steps:

Step 1: Negative distance was used to compute the similarity matrix S.
Step 2: All preferences are set to be the same value. At first, the value of preferences is set by Eq. (8).
Step 3: Initialize the availabilities $a(i,k)$ to zero.
Step 4: Update the responsibilities using the rule in Eq. (3).
Step 5: Update the availability using the rule in Eq. (4)
Step 6: After changes in the messages fall below a threshold for some number of iterations, terminate the iterations.
Step 7: Go to step 2, and change the preferences by Eq. (8) and Eq. (9) to obtain more bands.
Step 8: Get the exemplars, i.e., the representative bands of hyperspectral imagery.

AAP takes a collection of real-valued similarities between different bands as inputs, and in the AP, Euclidean distance (EUD) is used as the initial input. In the information field, the many metrics or information distance can be used as similarity measures, and all the similarities that are real and symmetric distance can be used for the AAP algorithm. In fact, there are significant differences when using different distances for similarity matrix. Other similar metrics, such as negative SAM, mutual information (MI), Jeffreys-Matusita distance (JM), and spectral information divergence (SID), are experimented and compared in this paper. Because the similarity metric is based on class signatures only, this AAP-based hyperspectral band selection is a semi-supervised method.

C. Computational Complexity and Methods for Comparison

Table 1 lists the computational complexity of different methods during the band selection process. For the AAP, it is $O(Kt+N)$, compared to ($O(N^2+Kt+N)$ in the AP, $O(Kt+N^2)$ in the MVPCA and MSNRPCA, where K is the number of bands and t is the number of iterations. Obviously, the computational complexity of the AAP is smaller than MVPCA and MSNRPCA, and it is also much lower than LCMV(BCM) and LCMV(BCC), since $N \ll LN^2$.

In order to evaluate the performance of the AAP algorithm, four widely used band selection algorithms, MVPCA, MSNRPCA [7], LCMV(BCM) and LCMV(BCC) [8], are used for the comparison. MVPCA and MSNRPCA are two unsupervised band selection methods derived from principal components analysis (PCA). In these methods, the band prioritization is based on an eigenanalysis, and a matrix is decomposed into an eigenform matrix, from which a loading-factors matrix could be constructed and used to prioritize bands. After all bands are ranked in accordance with their associated priorities, determined by the loading factors, a divergence-based band decorrelation, which used the divergence measure to remove redundant or insignificant bands, are conducted for band prioritization. In LCMV (BCM) and LCMV(BCC), they linearly constrain a desired target signature while minimizing interfering effects caused by other unknown signatures. BCM and BCC, which use LCMV to linearly constrain a band image, also minimized band correlation or dependence provided by other band images, by interpreting a band image as a desired target signature vector while considering other band images as unknown signature vectors.

Table 1. Computational Complexity of Band Selection Methods[a]

Methods	Number of Multiplications
AAP	$O(Kt+N)$
AP	$O(N^2+Kt+N)$
MVPCA	$O(KL+N^2)$
MSNRPCA	$O(KL+N^2)$
LCMV(BCM)	$O(L+LN^2)$
LCMV(BCC)	$O(L+L^2N+LN^2)$

[a] N is the number of pixels, L is the number of bands, K is the number of bands, and t is the number of iterations.

Fig. 1. (Color online) Washington, DC, Mall data.

D. Performance Evaluation

The selected bands are used for classification, and the quality can be evaluated with classification accuracy. Because training and test samples are available, the support vector machine (SVM) is applied, and the overall accuracy is used as the evaluation criterion. The SVM classifier is selected due to its excellent performance when dealing with small-sized training data. SVM discriminates two classes by fitting an optional separating hyperplane to the training samples of two classes in a high-dimensional feature space. In linearly inseparable cases, the input space is mapped into a high-dimensional feature space using a kernel function. Details of SVM are available in [18].

3. Experiments and Discussion

A. Hyperspectral Digital Imagery Collection Experiment (HYDICE) Washington, DC, Mall Experiments

A subscene of the Washington, DC, Mall image data, Fig. (1), is used for a real data experiment. There are 210 bands from 0.4 to 2.4 μm in the visible and infrared wave range, and its spatial resolution is approximately 2.8 m. After removing some water absorption bands, only 191 bands are used in the experiments. Of the data, there are seven classes: roof, tree, grass, water, road, trail, and shadow. Training and test samples are available for this scene.

Figure 2 shows the overall classification accuracy using SVM classifier from the DC Mall data. As we can see, the proposed AAP-based band selection algorithm significantly outperforms other methods. All bands means classification accuracy using the original data (before band selection), and its performance is lower than that of the AAP (after band selection). In this case, the performance of MVPCA and

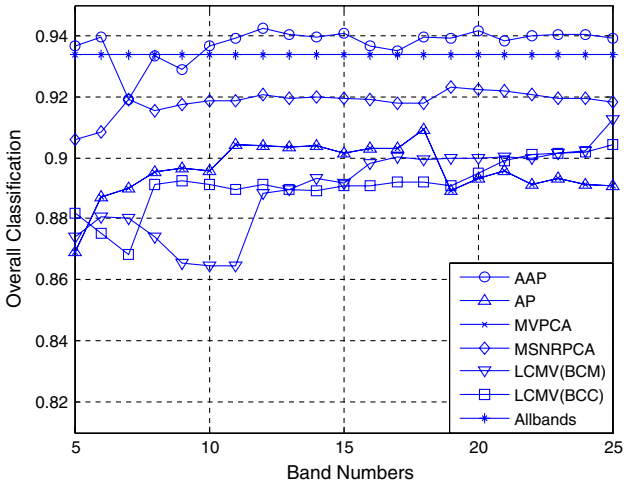

Fig. 2. (Color online) Classification comparison with different band selection methods for DC Mall data.

MSNRPCA are higher than that of LCMV(BCM) and LCMV(BCC).

Figure 3 presents classification maps when using twelve bands. There was a substantial amount of misclassification between trail (in yellow) and roof (in orange) using all bands of original data. The AP further enlarged the orange (roof) areas but not significantly. The AAP using twelve bands could significantly reduce the yellow (trail) areas.

B. HYMAP Purdue Campus Experiments

The dataset used in this experiment is a hyperspectral remote sensed image of the Purdue University, West Lafayette campus. The data was collected on September 30, 1999, with the airborne mapper hyperspectral (HYMAP) system, providing image data in the visible and infrared regions from 0.4 to 2.4 μm. It includes six classes: road, grass, shadow, soil, tree, and roof. The original image has 128 bands and about 3.5 m spatial resolution, and 126 bands participated in band selection after bad-band removal. An image of the scene is shown in Fig. 4.

Figure 5 illustrates the overall accuracy of classification based on SVM classifiers. From the results, the proposed band selection algorithm performs better than the others. MVPCA also performs better when the feature number is larger, but the accuracy of it is still lower than the proposed AAP method. LCMV(BCM) and LCMV(BCC) generate the same results, except in several band numbers; this is because the band selection results are almost the same, and they get the worst results. In this experiment, increasing the value of band numbers did not improve the accuracy. Instead, the highest accuracy appears when band number equals 12.

In order to evaluate the performance discrepancy when using different similarity metrics for the AAP algorithm, five distance metrics, such as SAM, EUD, MI, JM, and SID, are conducted for the AAP-based band selections. The experiment results are demonstrated in Fig. 6; it is obvious that SAM provided the best performance among all the metrics, such as Euclidean distance, MI, JM, and SID, and JM metric gives the worst results.

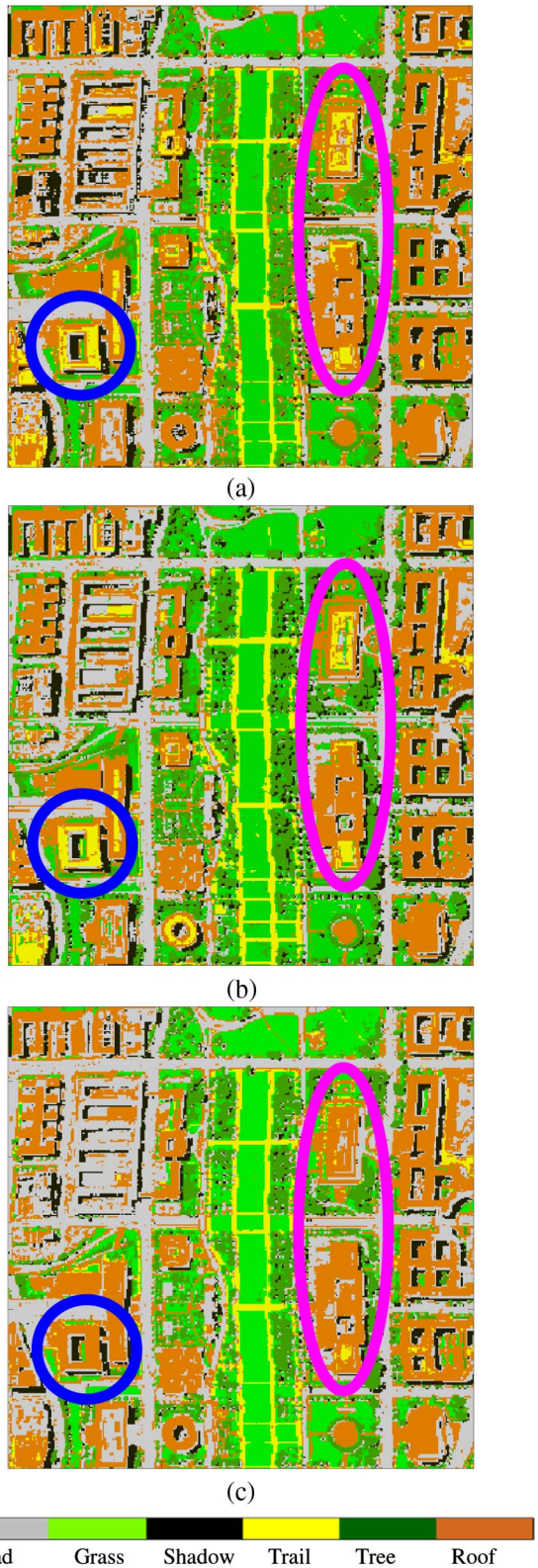

Fig. 3. (Color online) Classification maps of DC Mall data (with twelve bands). (a) All bands (OA = 0.9340), (b) AP (OA = 0.8331), and (c) AAP (OA = 0.9426)

Fig. 4. (Color online) Purdue campus data.

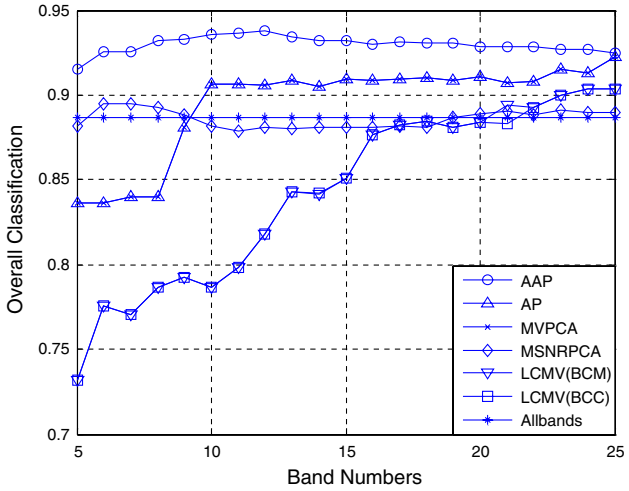

Fig. 5. (Color online) Classification comparison with different band selection methods for Purdue campus data.

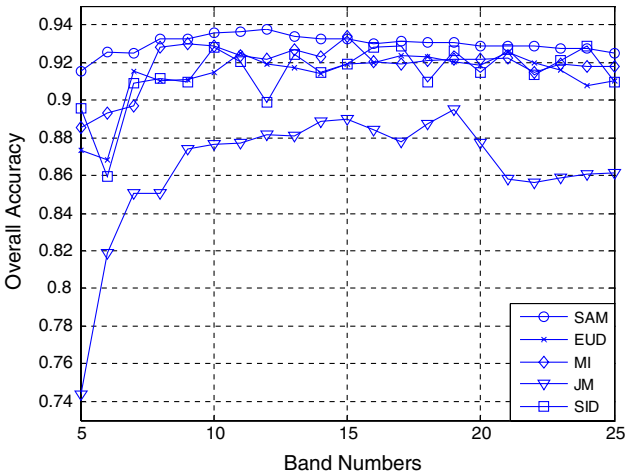

Fig. 6. (Color online) Classification comparison with different metrics in the AAP for Purdue campus data.

Fig. 7. (Color online) Classification maps of Purdue campus data (with twelve bands). (a) All bands (OA = 0.8870), (b) AP (OA = 0.8821), and (c) AAP (OA = 0.9379).

Figure 7 presents classification maps when using twelve bands. The improvement in the vegetation area (highlighted in the circles in pink) was obvious. In the roof areas circled in blue, the AP could slightly reduce the gray areas (for road) that were misclassified, but the AAP using twelve bands could more significantly reduce the gray (and black) areas.

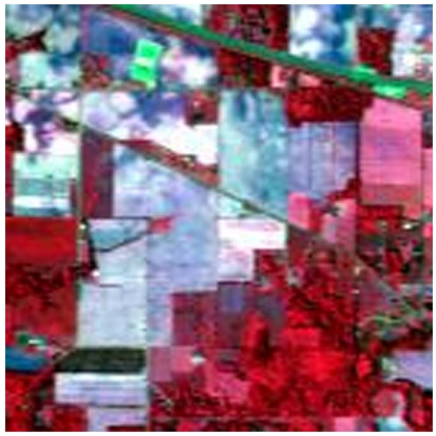

Fig. 8. (Color online) Indian Pines data.

C. Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines Experiments

The AVIRIS subimage taken over northwest Indiana's Indian Pines, with 145×145 pixels and 202 bands (Fig. 8). In this dataset, sixteen different land-cover classes were presented based on the ground truth. Since training and test samples are known, overall classification accuracy was computed for evaluation using SVM outputs.

The purpose of this experiment is to demonstrate the performance discrepancy on information changes and computational complexity when using different band selection methods. Figure 9 shows the information preserved by the selected band using different band selection approaches; the information was measured by orthogonal projection divergence [19]. The AAP contains the most information among all the algorithms, and LCMV(BCM) takes second place. PCA and LCMV variants performed quite similarly, while LCMV(BCC) yielded the worst result when the band number was larger. Once again, the performance of the AAP could be better than MVPCA, MSNRPCA, LCMV(BCM), and LCMV(BCC).

Table 2. Computing Time of Different Algorithms for Indian Pines Data (in Seconds)

K	5	10	15	20	25
AAP	44.37	32.70	50.57	36.50	47.30
AP	161.29	95.90	68.15	70.57	82.70
MVPCA	37.64	36.38	56.28	57.70	55.45
MSNRPCA	63.15	34.53	44.59	48.07	54.63
LCMV(BCM)	43.62	48.06	46.47	46.16	46.81
LCMV(BCC)	47.22	51.60	71.32	77.69	71.42

Table 3. Average Classification Accuracies Over 5–25 Selected Bands with Different Techniques

	DC Mall	Purdue	Pines
AAP	0.9375	0.9299	0.8125
AP	0.8240	0.8620	0.5330
MVPCA	0.8958	0.8948	0.6467
MSNRPCA	0.9182	0.8858	0.6590
LCMV(BCM)	0.8896	0.8427	0.7595
LCMV(BCC)	0.8914	0.8422	0.6185
All bands	0.9340	0.8870	0.7655

Table 2 tabulates the computing time when the algorithms run in a personal computer with 2.26 GHz CPU and 4.0 GB memory. We can see that the AAP can save a significant amount of time, compared to the traditional MVPCA, MSNRPCA LCMV(BCM), and LCMV(BCC). Note that the running time of the AAP does not include the time for similarity matrix calculation, so it can approximately represent the running time of the AAP algorithm; using different similarity matrices based on signatures does not have much impact on the convergence speed of the AAP algorithms in these experiments. But the similarity matrix based on original data will consume much more time (AP).

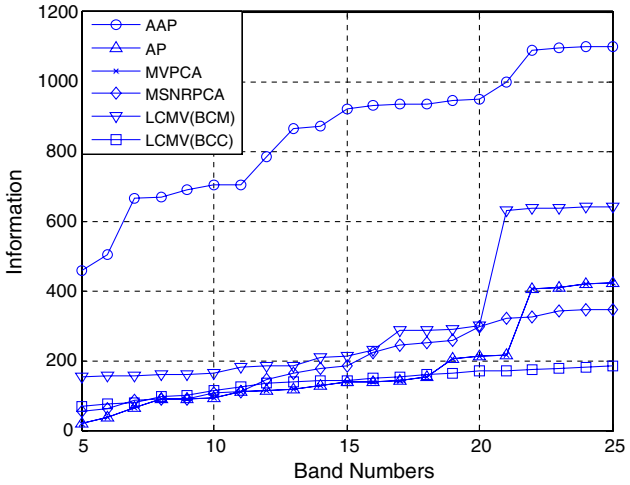

Fig. 9. (Color online) Amount of information with different band selection for Pines data.

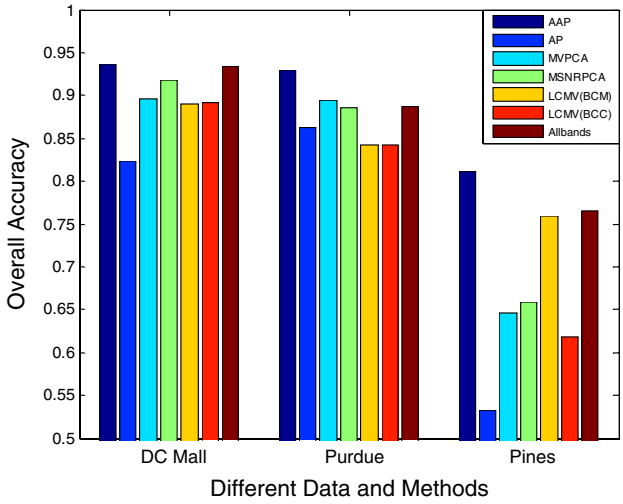

Fig. 10. (Color online) Classification comparison with different band selection methods for the Indian Pines data.

D. Analysis and Discussion

For more details, Table 3 shows the average classification accuracy over 5–25 selected bands with different techniques for all datasets. Figure 10 represents the average classification accuracy comparison of different band selection algorithms for all datasets.

From Table 3, for DC data, the AAP-based band selection algorithm obtains the highest classification accuracy (93.75% for SVM classifier), and LCMV is about 5% lower than the proposed AAP algorithm; MVPCA is 4.17% lower than the AAP. In the Purdue campus data, the accuracy of the AAP is higher than the LCMV by 8.77%, and higher than PCA variants by 3.51% and 4.41%, respectively. In the Indian Pines data, the AAP is higher than PCA variants by 16.58% and 15.35%, respectively; and LCMV variants are lower than our method by 5.30% and 19.40%. We can see that the original AP yielded the worst result.

Figure 10 illustrates classification comparison with different band selection methods. It can be obviously found that the proposed AAP algorithm outperforms other band selection methods in these three hyperspectral imagery data, especially in the Indian Pines data. It is proved that SAM distance can be an effective threshold to discriminate subset bands and distinctive bands from the point of classification tasks.

4. Conclusion

As one of the most important hyperspectral dimensionality reduction methods, band selection has received much attention for hyperspectral image analysis. This paper presents a novel AAP-based band selection algorithm for hyperspectral dimensionality reduction; it also provides a novel procedure for getting the fixed selected band numbers in the AP algorithm. Different from the conventional methods like k-mean and hierarchical clustering, the AAP uses a message-passing procedure to search the exemplars in the data points; it has several advantages over other band selection methods. Experiments have demonstrated that the proposed method can provide robust and promising results and perform better than other similar band selection algorithms. In the experiment, we also found that SAM metric outperforms other measures for hyperspectral imagery, and the computation complexity of the AAP algorithm is much lower than other methods.

The authors would like to thank Dr. Qian Du at Mississippi State University for providing the hyperspectral data used in the experiments. This work is partially supported by the Program of National Natural Science Foundation of China (nos. 40901200, 41171323, 41171321), the Priority Academic Program Development of Jiangsu Higher Education Institutions, and the Fundamental Research Funds for the Central Universities (no. 2012B01614).

References

1. G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," IEEE Trans. Inf. Theory **14**, 55–63 (1968).
2. C.-I. Chang, *Hyperspectral Imaging: Signal Processing Algorithm Design and Analysis* (Wiley, 2009).
3. D. Landgrebe, "Hyperspectral image data analysis," IEEE Signal Process. Mag. **19**, 17–28 (2002).
4. C.-I Chang and Q. Du, "Estimation of the number of spectrally distinct signal sources in hyperspectral imagery," IEEE Trans. Geosci. Remote Sens. **42**, 608–619 (2004).
5. J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," IEEE Trans. Geosci. Remote Sens. **46**, 2435–2445 (2008).
6. Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," IEEE Geosci. Remote Sens. Lett. **5**, 564–568 (2008).
7. C.-I Chang, Q. Du, T. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," IEEE Trans. Geosci. Remote Sens. **37**, 2631–2641 (1999).
8. C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," IEEE Trans. Geosci. Remote Sens. **44**, 1575–1585 (2006).
9. A. Paoli, F. Melgani, and E. Pasolli, "Clustering of hyperspectral images based on multiobjective particle swarm optimization," IEEE Trans. Geosci. Remote Sens. **47**, 4175–4188 (2009).
10. N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," IEEE Trans. Geosci. Remote Sens. **42**, 1552–1565 (2004).
11. A. Martínez-Usó, F. Pla, and J. M. Sotoca et al., "Clustering-based hyperspectral band selection using information measures," IEEE Trans. Geosci. Remote Sens. **45**, 4158–4171 (2007).
12. H. Su, H. Yang, Q. Du, and Y. Sheng, "Semi-supervised band clustering for dimensionality reduction of hyperspectral imagery," IEEE Geosci. Remote Sens. Lett. **8**, 1135–1139 (2011).
13. C. Cariou, K. Chehdi, and S. Le Moan, "BandClust: an unsupervised band reduction method for hyperspectral remote sensing," IEEE Geosci. Remote Sens. Lett. **8**, 565–569 (2011).
14. B. J. Frey and D. Dueck, "Clustering by passing messages between data points," Science **315**, 972–976 (2007).
15. B. J. Frey and D. Dueck, "Response to comment on 'Clustering by passing messages between data points,'" Science **319**, 726 (2008).
16. S. Jia, Y. T. Qian, and Z. Ji, "Band selection for hyperspectral imagery using affinity propagation," in *Digital Image Computing: Techniques and Applications Conference* (IEEE, 2008), pp. 137–141.
17. S. Jia, Z. Ji, and Y. T. Qian, "Band selection based hyperspectral unmixing," in *International Workshop on Imaging Systems and Techniques* (IEEE, 2009), pp. 303–306.
18. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining Knowledge Discovery **2**, 121–167 (1998).
19. C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer, 2003).