Multiple Classifier Ensembles with Band Clustering for Hyperspectral Image Classification

Hongjun Su\textsuperscript{1,2} and Peijun Du\textsuperscript{1*}

\textsuperscript{1}Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University, 163 Xianlin Road, 210046 - Nanjing, China
\textsuperscript{2}School of Earth Sciences and Engineering, Hohai University, 1 Xikang Road, 210098 - Nanjing, China
*Corresponding author, e-mail address: dupjrs@gmail.com

Abstract

Due to the high dimensionality of a hyperspectral image, classification accuracy of a single classifier may be limited when the size of the training set is small. A divide-and-conquer approach has been proposed, where a classifier is applied to each group of bands and the final output will be the fused result of multiple classifiers. Since the dimensionality in each band group is much lower, classification accuracy of the overall system can be improved even when training samples are limited. In this paper, we proposed a new multiple classifier ensembles which using SKMd-based band clustering features as input. We also investigate the impact of band partition for this approach. We find out that band partition based on spectral clustering (resulting in band groups composed of non-consecutive bands) can outperform the partition based on spectral correlation coefficient (resulting in band groups composed of consecutive bands only), in particular when the number of training samples is small.

Keywords: Hyperspectral classification, spectral clustering, multiple classifier ensembles.

Introduction

Supervised hyperspectral image classification requires sufficient training samples. Unfortunately, the numbers of training data are usually limited due to the expensiveness and difficulty in collecting ground-truth data. Small size training data leads to the curse of dimensionality [Hughes, 1968; Chang, 2013] which challenges the traditional single classifier algorithms.

In recent years, the divide-and-conquer approach has been developed [Kumar et al., 2001; Cheriyadat et al., 2003; Tsagaris et al., 2005; Prasad et al., 2012], where a classifier is applied to each group of bands and the final output will be the fused result of multiple classifiers. Since the dimensionality in each band group is much lower, the small training set may not be a problem any more. In addition, classification accuracy of the overall system can be improved with decision fusion.
In our research, we will focus on $k$-means-based band clustering for dimensionality reduction. 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. We have proposed a semi-supervised $k$-means clustering method (SKMd) that uses class signatures only (a class signature is the representative spectrum of a class) with a new initial technique and orthogonal projection divergence (OPD)-based bad band removal [Su et al., 2011]. Instead of using the band closest to the cluster centre, we use band cluster centers for the following data analysis (e.g. detection and classification).

In this paper, we will using SKMd-based band clustering for multiple classifier ensembles. The resulting band clusters include bands in different spectral ranges (i.e., not necessarily consecutive bands in the same spectral range), which may increase class separability. Support vector machine (SVM) is selected as the base classifier. Experimental results demonstrate the advantage of using SKMd-based band clustering over the traditional correlation coefficient-based band grouping, particularly when the number of training samples is limited. The main contribution of this paper is the application of non-consecutive bands to multiple classifier ensembles, which is a novel way different from using consecutive bands.

The remainder of this paper is organized as follows. In Section 2, the SKMd-based band clustering and multiple classifier ensemble algorithms are proposed. In Section 3, the experimental results are presented with three hyperspectral datasets. Finally, some discussion and conclusion remarks are made in Section 4.

Multiple classifier ensembles with band clustering

SKMd band clustering

Figure 1 is a typical spectral correlation coefficient matrix of an AVIRIS image. Its size is $L \times L$, where $L$ is the number of bands. It is a symmetric matrix. A white pixel at the $(i, j)$ location means a high correlation relation between the $i$-th and the $j$-th bands. Along the diagonal line, there are several bright squares, representing high spectral correlations among neighboring bands. Thus, it is straightforward to partition the bands into groups based on their correlation coefficients. It is worth mentioning that the resulting band partition is the same as that from using band mutual information [Pluim et al., 2003]. In this way, the partitioned bands in each group are consecutive bands within the same spectral range as reported in [Ceamanos et al., 2010].

After bands are partitioned into groups, a classifier (e.g., SVM), is applied to each group. The classification results from all the classifiers are fused to achieve the final classification. In this way, the number of training samples required is greatly reduced since the data dimensionality in each band group is much smaller than the dimensionality of the original data. Such a typical multi-classifier system is illustrated in Figure 2. It is denoted as band grouping (BG)-based multiple classifier ensembles.

However, it is well-observed that spectral correlation coefficient among non-consecutive bands may be higher as well; in some cases, it can be even larger than that between two neighboring bands. As shown in Figure 1, there are several bright squares in the off-diagonal areas. For instance, there is a bright square in the up-right corner, representing the high correlation between the optical bands and short-wave infrared bands. Thus, it seems to be reasonable to include non-consecutive bands in a band group. In this paper, we propose to conduct band clustering such that bands can be clustered into a same group, no matter they are consecutive or not.
Given a set of bands \( \{ B_1, \ldots, B_l, \ldots, B_L \} \), 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 \( C = \{ C_1, \ldots, C_m, \ldots, C_k \} \) (\( 1 \leq m \leq k \)) so as to minimize the following objective function:

\[
\arg \min_C \sum_{c=1}^{k} \sum_{B_j \in C_c} D(B_j, \mu_m) \quad [1]
\]

where \( \mu_m \) 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 it is assigned to. 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 \( S \) (\( S << N \)). This approach is denoted as semi-supervised \( k \)-means (SKM).

The SKM algorithm is initialized by using distinctive bands as cluster centroids. The idea of unsupervisedly selecting distinctive bands was presented [Du and Yang, 2008]. 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). The SKM algorithm deleting the worst cluster is denoted as SKMd [Su et al., 2011].
Multiple classifier ensembles

Figure 3 shows the proposed band clustering (BC)-based system diagram where band groups are replaced with band clusters to reflect this change. SKMd is adopted for band clustering [Su et al., 2011]. Traditionally, $k$-means clustering is applied to pixel vectors for unsupervised classification [Yang et al., 2013]. Here, it is applied to band images after a band image is rearranged into a vector. The proposed SKMd with SVMs can be detailed as below [Su et al., 2011].

1) Initialize the algorithm by using $k$ selected distinctive bands;

2) With the known class signatures, conduct $k$-means band clustering. The clustering is completed when no band is shuffled from one cluster to another. Compute band cluster centroids by averaging all the bands clustered;

3) OPD is employed to compute pair-wise cluster similarity. The cluster with the largest average OPD will be removed. The resulting $k-1$ clusters are the final band clustering result;

4) Calculate the OPD value between each band and its cluster centroid. A certain percentage of bands with large OPD values are removed. The $k-1$ cluster centroids are updated with the remaining bands, which are the final outputs;

5) After band clustering, SVM is applied to each cluster. The results from multiple SVM classifiers are simply fused with the rule of winner-take-all, although there are other fusion rules existing [Waske and Benediktsson, 2007].

It should be noted that several distance metrics can be used for clustering, including spectral angle (SA), and spectral correlation coefficient (CC). In addition to the correlation coefficient-based band grouping as mentioned before, the uniform (U) band grouping was also included in the experiments for comparison.
Experiment and Analysis

HYDICE DC Mall Experiment I

A subscene of a Hyperspectral Digital Imagery Collection Experiment (HYDICE) Washington DC Mall image data of size 304 × 301 pixels as shown in Figure 4 was used in the experiment. There are 210 bands in the 0.4 to 2.5µm spectral region, and its spatial resolution is approximately 2.8m. After some water absorption bands removed, 191 bands were left for analysis. In this scene, there are six classes present: roof, tree, grass, water, road, and trail. The total training and test set sizes of each class are shown in Table 1. The smallest training set had 46 samples only.

Figure 3 - The proposed diagram for the multi-classifier system.

Figure 4 - HYDICE Washington DC Mall image I and Test data map.
Table 1 - Training and test samples for DC Mall image I.

| Class Name | Training No. | Test No. |
|------------|--------------|---------|
| Road       | 55           | 892     |
| Grass      | 57           | 906     |
| Shadow     | 50           | 587     |
| Trail      | 46           | 578     |
| Tree       | 49           | 630     |
| Roof       | 69           | 1500    |

Based on spectral correlation coefficient, all the 191 bands were partitioned into 9 groups. For comparison purposes, they were also uniformly partitioned into 9 groups. When applying SKMd clustering, the number of clusters $K$ was also set to be 9.

Table 2 lists the overall accuracy (OA), average accuracy (AA), and Kappa coefficients when using 10, 20, 30, and all training samples. The proposed band clustering-based partition with two metrics was compared with band grouping-based partition and without partition (a single SVM for classifying the entire data). As we can see, band clustering with spectral angle as similarity metric provided the best results when using 10, 20, and 30 training samples for each class. However, when using all the training samples, band grouping with spectral correlation coefficient (CC) as similarity metric provided the best result.

Table 2 - Classification accuracy for DC Mall image I.

|                | 10 training samples | 20 training samples | 30 training samples | all training samples |
|----------------|---------------------|---------------------|---------------------|---------------------|
|                | OA      | AA      | Kappa   | OA      | AA      | Kappa   | OA      | AA      | Kappa   |
| 1 SVM          | 0.868   | 0.878   | 0.837   | 0.847   | 0.876   | 0.814   |
| BC (CC)        | 0.919   | 0.923   | 0.900   | 0.907   | 0.921   | 0.886   |
| BC (SA)        | 0.937   | **0.944** | **0.922** | **0.961** | **0.963** | **0.952** |
| BG (CC)        | 0.911   | 0.915   | 0.890   | 0.933   | 0.936   | 0.917   |
| BG (U)         | 0.835   | 0.861   | 0.799   | 0.884   | 0.909   | 0.858   |
|                | OA      | AA      | Kappa   | OA      | AA      | Kappa   |
| 1 SVM          | 0.816   | 0.849   | **0.775** | 0.910   | 0.921   | 0.889   |
| BC (CC)        | 0.902   | 0.918   | 0.879   | 0.953   | 0.954   | 0.942   |
| BC (SA)        | **0.901** | **0.924** | **0.878** | 0.922   | 0.936   | 0.904   |
| BG (CC)        | 0.888   | 0.913   | 0.862   | **0.981** | **0.980** | **0.977** |
| BG (U)         | 0.877   | 0.906   | 0.850   | 0.911   | 0.921   | 0.890   |

Note: 1 SVM means SVM classifier on the original spectral features after removal of water absorption bands; BC(CC) means band clustering(SKMd) using CC; BC(SA) band clustering using SA; BG (CC) band group using CC; and BG (U) band group using uniform way.
HYDICE DC Mall Experiment II

Another subimage with 266×304 pixels of the DC Mall data as shown in Figure 5 was also used in the experiment. In this subscene, there are seven classes: roof, tree, grass, water, road, trail, and shadow. The training and test set sizes of each class are shown in Table 3. The smallest training set included 63 samples. As in the previous experiment, the number of band groups/clusters was fixed to be 9.

![HYDICE image](image1.jpg) ![Test data map](image2.jpg)

Figure 5 - HYDICE Washington DC Mall image II and Test data map.

Table 3 - Training and test samples for DC Mall image II.

| Class Name | Training No. | Test No. |
|------------|--------------|----------|
| Road       | 63           | 1074     |
| Grass      | 124          | 2142     |
| Water      | 159          | 1347     |
| Trail      | 236          | 1668     |
| Tree       | 300          | 3465     |
| Shadow     | 63           | 1074     |
| Roof       | 124          | 2142     |

Table 4 lists the values of OA, AA, and Kappa coefficients when using the first 10, 25, 50, and all training samples. As we can see, band clustering with spectral angle as similarity metric still provided the overall best results when using 10, 25, and 50 training samples for each class. However, when using all the training samples, band grouping with spectral correlation coefficient (CC) as similarity metric provided the best result. However, in this experiment, when using all the training samples, band clustering performed similarly as band grouping using spectral correlation coefficient.
Table 4 - Classification accuracy for DC Mall image II.

|                  | 10 training samples | 25 training samples |      |
|------------------|----------------------|----------------------|------|
|                  | OA       | AA       | Kappa | OA       | AA       | Kappa |
| 1 SVM            | 0.874    | 0.903    | 0.849  | 0.850    | 0.880    | 0.820  |
| BC (CC)          | 0.901    | 0.930    | 0.881  | 0.919    | 0.938    | 0.902  |
| BC (SA)          | **0.920** | **0.947** | **0.938** | **0.949** | **0.959** | **0.939** |
| BG (CC)          | 0.908    | 0.922    | 0.889  | 0.907    | 0.883    | 0.887  |
| BG (U)           | 0.910    | 0.934    | 0.892  | 0.889    | 0.918    | 0.867  |

|                  | 50 training samples | all training samples |      |
|------------------|----------------------|-----------------------|------|
|                  | OA       | AA       | Kappa | OA       | AA       | Kappa |
| 1 SVM            | 0.947    | 0.932    | 0.936  | 0.949    | 0.934    | 0.938  |
| BC (CC)          | 0.959    | 0.962    | 0.950  | 0.961    | 0.961    | 0.953  |
| BC (SA)          | **0.958** | **0.964** | **0.950** | 0.959    | 0.961    | 0.947  |
| BG (CC)          | 0.946    | 0.949    | 0.935  | **0.965** | **0.965** | **0.958** |
| BG (U)           | 0.912    | 0.935    | 0.894  | 0.898    | 0.923    | 0.877  |

Note: 1 SVM means SVM classifier on the original spectral features after removal of water absorption bands; BC(CC) means band clustering(SKMd) using CC; BC(SA) band clustering using SA; BG (CC) band group using CC; and BG (U) band group using uniform way.

**HyMap Purdue Experiment**
The last dataset is collected by the airborne Hyperspectral Mapper (HyMap) about a residential area near the campus of Purdue University in 1999. The HyMap system providing image data in 128 spectral bands (with 0.45-2.48 spectral coverage). In this experiment, 126 bands are used excluding the atmospheric water vapor bands. The image size is 377×512. The spatial resolution is about 5 m and spectral bandwidth is about 16 nm.

(a) HyMap image  (b) Test data map

Figure 6 - HyMap Purdue image and Test data map.
The image scene includes six classes: {road, grass, shadow, soil, tree, roof}. As listed in Table 5, 404 training samples and 5463 testing samples were available. Compared to the HYDICE image, roof class in this image was more spectrally homogeneous. However, the road class had within-class spectral variation, particularly in the upper right subdivision.

| Class Name | Training No. | Test No. |
|------------|--------------|---------|
| Road       | 73           | 1230    |
| Grass      | 72           | 1072    |
| shadow     | 49           | 213     |
| soil       | 69           | 371     |
| Tree       | 67           | 1321    |
| roof       | 74           | 1236    |

The classification performance of OA, AA, and Kappa coefficients were listed in Table 6 when using 10, 25, 40, and all training samples. From the table, band clustering with spectral angle as similarity metric still provided the best results.

|                           | 10 training samples | 25 training samples | 40 training samples | all training samples |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
|                           | OA                  | AA                  | Kappa               | OA                  | AA                  | Kappa               | OA                  | AA                  | Kappa               |
| 1 SVM                     | 0.872               | 0.865               | 0.839               | 0.870               | 0.864               | 0.837               |
| BC (CC)                   | 0.895               | 0.882               | 0.866               | 0.923               | 0.910               | 0.895               |
| BC (SA)                   | 0.890               | 0.885               | 0.877               | 0.928               | 0.915               | 0.903               |
| BG (CC)                   | 0.891               | 0.884               | 0.861               | 0.912               | 0.899               | 0.885               |
| BG (U)                    | 0.732               | 0.707               | 0.660               | 0.826               | 0.807               | 0.782               |
| 1 SVM                     | 0.931               | 0.924               | 0.909               | 0.953               | 0.942               | 0.918               |
| BC (CC)                   | 0.937               | 0.925               | 0.914               | 0.959               | 0.942               | 0.926               |
| BC (SA)                   | 0.922               | 0.914               | 0.895               | 0.955               | 0.940               | 0.929               |
| BG (CC)                   | 0.826               | 0.807               | 0.782               | 0.842               | 0.824               | 0.801               |

Note: 1 SVM means SVM classifier on the original spectral features after removal of water absorption bands; BC(CC) means band clustering(SKMd) using CC; BC(SA) band clustering using SA; BG (CC) band group using CC; and BG (U) band group using uniform way.
Discussions and Conclusions

From the HYDICE DC Mall experiment, we can find that the performance of SVM is sensitive to training samples. For instance, the roofing material is different from others in the upper left area Figure 4. If the selected training samples do not locate in this area (30 selected training samples in the experiment), it will lead to more misclassification. This is the reason for which 30 training samples perform worse than 20 training samples in Table 2. As long as training samples have good representation, SVM is a truly powerful classifier [Yang et al., 2012].

In this paper, we investigate the impact of band partition in the divide-and-conquer approach, where each classifier is applied to a band group/cluster and classifier fusion is conducted for final classification. We find out that using SKMd clustering-based band partition (BC (SA)) outperforms spectral correlation coefficient-based band grouping (BC (CC)) when the number of training samples is small. The former may include non-consecutive bands, while the latter consists of consecutive bands only. The spectral angle-based similarity metric yields the most robust performance over other similarity metric in clustering. As future work, we will compare with other more advanced band clustering methods and investigate the performance when using other decision fusion rules.

Acknowledgements

The authors would like to thank Dr. Landgrebe and Dr. Huang Xin for providing the data used in the experiments. This paper was partially supported by National Natural Science Foundation of China (No.41201341), China Postdoctoral Science Foundation (No.2012M521044), the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (No.12R02), and Key Laboratory of Advanced Engineering Surveying of National Administration of Surveying, Mapping and Geoinformation (No.TJES1301).

References

Ceamanos X., Waske B., Benediktsson J.A., Chanussot J., Fauvel M., Sveinsson J.R. (2010) - A Classifier Ensemble Based on Fusion of Support Vector Machines for Classifying Hyperspectral Data. International Journal of Image and Data Fusion, 1: 293-307. doi: http://dx.doi.org/10.1080/19479832.2010.485935.

Chang C-I. (2013) - Hyperspectral Data Processing: Algorithm Design and Analysis. New Jersey: Wiley-Interscience, 1-5.

Cheriyadat A., Bruce L.M., Mathur A. (2003) - Decision level fusion with best-bases for hyperspectral classification. In: Proceeding IEEE Workshop Advances in Techniques for Analysis of Remotely Sensed Data, 399-406. doi: http://dx.doi.org/10.1109/WARSD.2003.1295221.

Du Q., Yang H. (2008) - Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geoscience and Remote Sensing Letters, 5: 564-568. doi: http://dx.doi.org/10.1109/LGRS.2008.2000619.

Hughes G.F. (1968) - On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, IT-14: 55-63. doi: http://dx.doi.org/10.1109/TIT.1968.1054102.
Kumar S., Ghosh J., Crawford M.M. (2001) - *Best-bases feature extraction algorithms for classification of hyperspectral data*. IEEE Transactions on Geoscience and Remote Sensing, 39: 1368-1379. doi: http://dx.doi.org/10.1109/36.934070.

Pluim J.P.W., Maintz A., Viergever M.A. (2003) - *Mutual-information-based registration of medical images: A survey*. IEEE Transactions on Medical Imaging, 22: 986-1004. doi: http://dx.doi.org/10.1109/TMI.2003.815867.

Prasad S., Li W., Fowler J.E., Bruce L.M. (2012) - *Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification*. IEEE Geoscience and Remote Sensing Society, 50: 3474-3486. doi: http://dx.doi.org/10.1109/TGRS.2012.2185053.

Su H., Yang H., Du Q., Sheng Y. (2011) - *Semi-supervised band clustering for dimensionality reduction of hyperspectral imagery*. IEEE Geoscience and Remote Sensing Letters, 8: 1135-1139. doi: http://dx.doi.org/10.1109/LGRS.2011.2158185.

Tsagaris V., Anastassopoulos V., Lampropoulos G.A. (2005) - *Fusion of hyperspectral data using segmented PCT for color representation and classification*. IEEE Transactions on Geoscience and Remote Sensing, 43: 2365-2375. doi: http://dx.doi.org/10.1109/TGRS.2005.856104.

Waske B., Benediktsson J.A. (2007) - *Fusion of support vector machines for classification of multisensor Data*. IEEE Transactions on Geoscience and Remote Sensing, 45: 3858-3866. doi: http://dx.doi.org/10.1109/TGRS.2007.898446.

Yang H., Du Q., Chen G. (2012) - *Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5: 544-554. doi: http://dx.doi.org/10.1109/JSTARS.2012.2185822.

Yang H., Peng J., Xia B., Zhang D. (2013) - *Remote Sensing Classification Using Fuzzy C-means Clustering with Spatial Constraints Based on Markov Random Field*. European Journal of Remote Sensing, 46: 305 - 316. doi: http://dx.doi.org/10.5721/EuJRS20134617.