Clustering based Band Selection for Hyperspectral Images

Aloke Datta
Center for
Soft Computing Research
Indian Statistical Institute
Kolkata, India
Email: daloked@gmail.com

Susmita Ghosh
Department of
Computer Science and Engineering
Jadavpur University
Kolkata, India
Email: susmitaghoshju@gmail.com

Ashish Ghosh
Center for
Soft Computing Research
Indian Statistical Institute
Kolkata, India
Email: ash@isical.ac.in

***Abstract*—An unsupervised band selection method for hyperspectral images is proposed in this article. Three steps are followed to carry out the algorithm. In the first step, characteristics (attributes) of the bands are generated. Next, redundancy among the bands is removed by using clustering. DBSCAN algorithm is used for clustering the bands. One representative band is selected from each cluster. Finally, the bands are ranked based on their discriminating capabilities for classification. To demonstrate the effectiveness of the proposed method, results are compared with a ranking based and two clustering based methods in terms of classification accuracy and *Kappa* coefficient. Results for the proposed methodology are found to be encouraging.**

Keywords:- Unsupervised band selection, hyperspectral imagery, clustering, feature ranking.

I. Introduction

Hyperspectral imagery [1], [2] can be viewed as an image cube where the first two dimensions indicate the size of the image and the third one specifies the band number of the image. Thus, each pixel may be treated as a pattern whose number of attributes is equal to the number of bands associated with it. In hyperspectral images, a strong correlation exists among the bands, i.e., increasing the number of bands may not always increase the discriminating capability for classification. On the contrary, it becomes a complex procedure to perform any classification task with this high dimension. As a result, reducing the dimensionality is considered as an important step where the aim is to discard the redundant bands and make it less time consuming for classification. Band selection and band extraction [1], [3]-[11] are two main approaches for dimensionality reduction. Depending on the availability of labeled patterns, band selection techniques can be categorized as supervised and unsupervised. If no labeled pattern is available, unsupervised band selection is used for dimensionality reduction.

Existing techniques of unsupervised band selection for hyperspectral images can be broadly classified into two categories: ranking based [7]-[9] and clustering based [10], [11]. The main idea of ranking based methods is to find the most distinctive and informative bands. Various ranking based methods for unsupervised band selection are present in the literature, like information divergences (ID) based method [7], maximum variance based principal component analysis (MVPCA) [8], similarity based band selection method [9]. Correlation among bands is not considered in these methods. If information content or discriminating ability of a band is considered for its selection, redundant bands may be selected. On the other hand, clustering based methods [10], [11] perform clustering on bands (treated as patterns) to group them according to their correlation, and select one band from each cluster representing the whole group. A few clustering based band selection techniques for hyperspectral images exist in the literature e.g., Ward's linkage strategy using divergence (WaLuDi) [10], or using mutual information (WaLuMI) [10], and band selection using affinity propagation (AP) [11].

In the present work, an unsupervised band selection technique is proposed by determining the characteristics of bands, and integrating the merits of both rank based and clustering based dimensionality reduction techniques in a single framework. The proposed work consists of three steps. In the first step, band characteristics are determined to generate the attributes of each band by estimating the region types present in the imagery and choosing a representative pixel from each region type. In the next step, clustering is performed over bands and an intermediate set of bands is formed which consists of one representative band from each cluster, along with the remaining uncorrelated (isolated) bands. DBSCAN [12] algorithm is used as a clustering technique for giving equal importance to both correlated bands and isolated bands. In the third step, bands of the intermediate set are ranked depending on their discriminating capabilities. To evaluate the effectiveness of the proposed method, overall classification accuracy (OA), and *Kappa* coefficient (*Kappa*) [13] are calculated for the selected (set of) bands. Performance of the proposed algorithm is compared with those of two other state of the art clustering based unsupervised band selection methods [10], [11] and one ranking based method [7]. The results show that the proposed method gives promising results compared to others in terms of OA and *Kappa*.

II. The Proposed Method

In band selection, d number of bands are selected from the original set of D $(> d)$ bands. The three steps of the proposed unsupervised band selection method for hyperspectral

978-1-4673-4700-6/12/$31.00 ©2012 IEEE

imagery are: determination of band characteristics, removal of redundant bands, and ranking of bands. Figure 1 shows the diagrammatic representation of the proposed method while block diagrams of Steps 1 and 2 are shown in Figures 2 and 3, respectively. For determination of band characteristics, a hyperspectral image cube is treated as input. Then it approximates the region types of the imagery, selects a representative from each region type and represents them in terms of the set of bands along with their characteristics. In the next step, clustering is performed to group the bands, and an intermediate set of bands is formed considering the representative from each cluster and the remaining isolated (uncorrelated) bands. Lastly, bands present in this intermediate set are ranked depending on their contents of information. The desired number of bands are selected from the set. The detail activity of these three steps are discussed below.

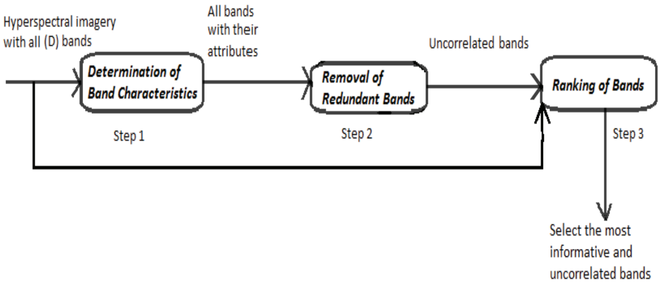
Hyperspectral imagery with all (D) bands
All bands with their attributes
Uncorrelated bands
Determination of Band Characteristics
Removal of Redundant Bands
Ranking of Bands
Step 1
Step 2
Step 3
Select the most informative and uncorrelated bands

Fig. 1. Block diagram of the proposed band selection method

Step 1: Determination of Band Characteristics:

Attributes of a band can be characterized by measuring the reflectance of different materials with respect to the said band (particular wavelength). More precisely, the number of attributes of a band with respect to a given hyperspectral imagery, is equal to the number of different materials present in the imagery. Attribute values are the responses of these materials with respect to the band under consideration. In this context, different materials mean different land cover types present in the image. To generate the attributes of bands, initially we have to find out the region types present in the given hyperspectral imagery. One of the ways to do this is to perform clustering operation over the pixels to group them into homogeneous regions.

Let D be the total number of bands present in the imagery. Then, each pixel can be treated as a pattern with D attributes (features). Although various types of clustering techniques can be used to find out the region types, DBSCAN is applied in the proposed method. This is due to the fact that, DBSCAN algorithm does not require any prior information about the number of region types. Moreover, if any pattern is far away from all the clusters then the DBSCAN technique treats it as an isolated point (outlier), rather than including it in any cluster. Let C be the number of clusters obtained from clustering, which gives us an approximation about the number of land cover types/ groups present in the imagery. Since, all the pixels of a cluster possess similar characteristics, a representative point is selected from each group. The mean of a cluster is treated as the representative point of the said cluster. As a result, there are C representative points each having D number of attributes. Let $A_{C,D}$ be a matrix containing the representatives of all the C clusters and it is of the form:

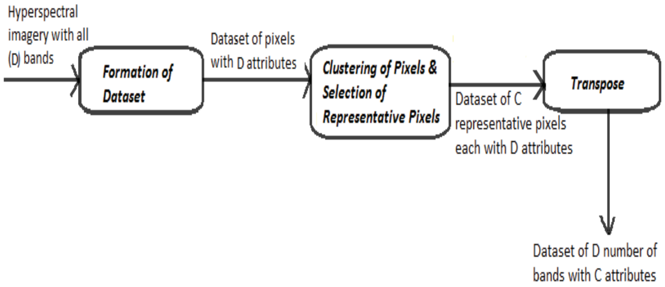
Hyperspectral imagery with all (D) bands
Formation of Dataset
Dataset of pixels with D attributes
Clustering of Pixels & Selection of Representative Pixels
Dataset of C representative pixels each with D attributes
Transpose
Dataset of D number of bands with C attributes

Fig. 2. Block diagram of the determination of band characteristics (Step 1)

$$A_{C,D} = \begin{array}{c} \\ c_1 \\ \vdots \\ c_i \\ \vdots \\ c_C \end{array} \begin{array}{c} \begin{array}{ccccc} b_1 & \cdots & b_j & \cdots & b_D \end{array} \\ \begin{pmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,D} \\ \vdots & \ddots & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,D} \\ \vdots & & \vdots & \ddots & \vdots \\ a_{C,1} & \cdots & a_{C,j} & \cdots & a_{C,D} \end{pmatrix} \end{array}.$$

Here, $c_i = [a_{i,1}, ..., a_{i,j}, ..., a_{i,D}]$ is the representative of cluster c_i. Alternatively, it can be said that $b_j = [a_{1,j}, ..., a_{i,j}, ..., a_{C,j}]^T$ is the characteristic vector or attribute vector. Hence, one can say that, there are D number of bands, each having C number of attributes.

Step 2: Removal of Redundant Bands:

One main property of hyperspectral imagery is that high correlation exists among the neighboring bands and there exist a few more bands which are totally uncorrelated with all other bands. Bands which are correlated provide similar information. Thus, it is reasonable to remove some of the correlated bands without affecting the performance. In our investigation, highly correlated bands form clusters, and only one representative band is selected from each cluster while discarding the other bands. Uncorrelated bands which do not belong to any cluster are treated as isolated ones and should be considered with equal importance. Here, again DBSCAN is used since it treats the uncorrelated and isolated bands with equal importance. It can also automatically recognize the number of clusters.

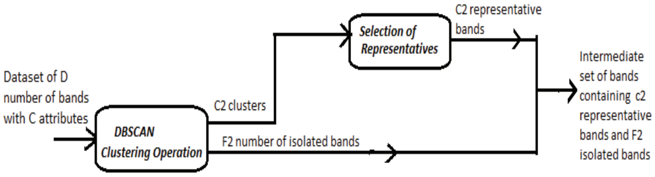
Dataset of D number of bands with C attributes
DBSCAN Clustering Operation
C2 clusters
F2 number of isolated bands
Selection of Representatives
C2 representative bands
Intermediate set of bands containing c2 representative bands and F2 isolated bands

Fig. 3. Block diagram of removal of redundant bands (Step 2)

Minimal requirement of domain knowledge for adjusting its parameter values, and discovery of clusters of arbitrary shapes are advantages of the said clustering technique.

Let us consider that the clustering operation provides $C2$ number of clusters and $F2$ number of isolated bands. One representative band from each of the $C2$ clusters and $F2$ number of isolated bands form an intermediate set of bands. Thus, the intermediate set will have $C2 + F2$ bands. It is expected that the bands of the intermediate set are highly uncorrelated.

Let c_i $(i = 1, 2, ..., C2)$ be the i^{th} cluster and n_i $(i = 1, 2, ..., C2)$ be the number of patterns present in it. Now, each pattern p_j (where p_j belongs to the cluster c_i) is considered as the cluster center and the distances of all other patterns $\{x_k\}$ (x_k belongs to the cluster c_i) are calculated from the center. The distance, Δ_{ij}, calculated for all j (where, $j = 1, 2, ..., n_i$) is as follows:

$$\Delta_{ij} = \frac{\sum_{k=1}^{n_i} d(x_k, p_j)}{n_i}, \tag{1}$$

where $d(x_k, p_j)$, the distance between the patterns x_k and p_j, is mathematically defined as:

$$d(x_k, p_j) = \left(\sum_{m=1}^{C} (x_{k,m} - p_{j,m})^2 \right)^{1/2}, \tag{2}$$

where C is the number of attributes of each pattern. Now a pattern, p_j is found out for which Δ_{ij} is the minimum. This pattern is treated as the representative of cluster c_i (denoted as, CR_i). Thus,

$$CR_i = \underset{p_j}{\operatorname{argmin}}(\Delta_{ij}). \tag{3}$$

Step 3: Ranking of Bands:

The task, ranking of bands, is performed over the set of bands present in the intermediate set. As mentioned, intermediate set contains uncorrelated bands and its size is much smaller than the original number of bands present in the hyperspectral imagery. The uncorrelated bands of the intermediate set are sorted in descending order depending on their discriminating capabilities. Discriminating capability of a band is measured by considering the deviation of normalized probability distribution of a band image from the corresponding Gaussian probability distribution with the same mean and variance. (This is called the non-Gaussianity of that band.) This deviation is quantitatively measured by information divergence [14] of that band image with its corresponding Gaussian image. The probability distribution of the k-th band image, q_k, is estimated by calculating the normalized histogram of the said image. Let, g_k be the Gaussian distribution whose mean and variance are the same as that of the image. The information divergence of this two probability distributions, denoted as, $D(q_k, g_k)$, is defined as:

$$\mathrm{D}(q_k, g_k) = \sum_i q_{ki} \log(q_{ki}/g_{ki}) + \sum_i g_{ki} \log(g_{ki}/q_{ki}). \tag{4}$$

Higher value of information divergence indicates more deviation of q_k from the Gaussian distribution g_k. This measure will help us to select the desired d number of bands from the top of the list (of intermediate set).

III. Experiment and Evaluation

To evaluate the effectiveness of the proposed method, experiments are conducted on a set of hyperspectral image, but due to limitations of space, results on Indiana data set [5] are discussed here. Indiana image data was captured by AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) over an agricultural land of northwest Indiana's Indian pine test site in the early growing season of 1992. The data has been taken within the spectral range from 400nm-2500nm with spectral resolution of about 10nm and has 220 spectral bands. The size of the image is 145×145 and spatial resolution is 20m. Twenty water absorption bands and fifteen noisy bands were removed, resulting in a total of 185 bands. There are 16 classes in this image.

Settings of two parameters are required to execute DBSCAN clustering algorithm: (i) the radial distance with which the neighborhood is defined, denoted as, *Eps*, and (ii) the minimum number of points required to be present within the defined neighborhood, denoted by, *MinPts*. Ester et. al. [12] suggested to use *MinPts* equal to 4 and to calculate the value of *Eps*, they suggested a method by depicting a graph of the number of points with respect to their 4th nearest neighbor. The value of *Eps* is computed by noting the distance of the first valley of this graph. In the proposed method, both *MinPts* and *Eps* are calculated according to Ester et. al. [12].

As mentioned earlier, performance of the proposed unsupervised band selection technique has been compared with three other existing techniques. Two clustering based methods (namely, Ward's linkage strategy using divergence (WaLuDi) [10] and affinity propagation (AP) [11]) and a ranking based method which uses information divergence (ID) [7] are considered for comparison. Before conducting the experiments, normalization of data has been done. The desired number of bands to be selected is not known apriori and it varies from image to image. In the present investigation, experiments are carried out using different number of selected bands ranging from 4 to 30 with a step size of 2. At termination of unsupervised band selection algorithm, fuzzy kNN based classification is performed on the selected subset of bands using 10-fold cross validation (theoretically, any good classification algorithm can be used). There may be overlapping of information between neighboring pixels of the hyperspectral imagery. Fuzzy kNN, rather than other classification techniques, is used to take care of the fuzziness present in hyperspectral imagery. Overall classification accuracy (OA), and *Kappa* coefficient (*Kappa*) are calculated from the labeled information present in the data set. The higher the value of *Kappa*, the better is the classification.

The OA and *Kappa* values obtained with Indiana image using ID, WaLuDi, AP based algorithms and the proposed method are given in Table I. From this table, it is observed that the proposed methodology outperforms the other three methods used in our experiments in terms of both the performance indices, OA and *Kappa*. It is also noticed that clustering based methods (WaLuDi and AP) are better than ranking based method (ID). AP based method gives slightly better results than WaLuDi based one. It is noticed that the optimum performance is obtained (*w.r.t.* both the measures) when the number of selected bands is 18. This holds true for WaLuDi, AP based method and the proposed method. Figure 4 depicts the variation of OA (in percentage) with number of bands for all the methods used in the experiment. The upper horizontal dotted line in the graph shows the classification accuracy with all the bands present in the data set. From the graph, it is seen that the peak performance is obtained when the number of selected bands is 18 in case of WaLuDi based, AP based and the proposed techniques; whereas for ID based method, OA increases slowly and becomes stable when the number of selected bands is above 24. Moreover, from Fig. 4, it is noticed that the values of OA using the proposed method are higher than those of the other three methods used in our investigation. Thus, it can be concluded that the proposed algorithm gives better subset of bands for classification than the other methods.

TABLE I
OVERALL ACCURACY AND *Kappa* VALUES OBTAINED USING ID, WALUDI, AP AND THE PROPOSED METHOD WITH DIFFERENT NUMBER OF SELECTED BANDS FOR INDIANA DATA SET

No of Bands	ID		WaLuDi		AP		Proposed method	
	OA	*Kappa*	OA	*Kappa*	OA	*Kappa*	OA	*Kappa*
4	51.80	0.4498	48.30	0.4060	67.48	0.6266	71.56	0.6733
6	62.54	0.5684	62.27	0.5645	72.51	0.6836	77.60	0.7413
8	64.95	0.5945	63.24	0.5760	76.61	0.7310	78.74	0.7552
10	65.17	0.5972	74.83	0.7100	78.79	0.7563	79.56	0.7628
12	67.68	0.6264	76.04	0.7238	78.71	0.7552	80.30	0.7736
14	69.36	0.6455	79.33	0.7652	80.74	0.7787	82.94	0.8059
16	70.02	0.6532	81.68	0.7924	81.73	0.7903	84.40	0.8207
18	70.98	0.6644	82.54	0.8023	83.38	0.8091	84.81	0.8255
20	72.30	0.6798	82.37	0.7972	82.18	0.7952	84.57	0.8227
22	72.68	0.6844	81.33	0.7852	81.90	0.7919	84.28	0.8194
24	73.64	0.6955	81.39	0.7859	81.85	0.7914	83.82	0.8141
26	74.31	0.7034	80.53	0.7758	82.12	0.7946	83.56	0.8111
28	73.94	0.6987	80.95	0.7839	80.92	0.7807	83.70	0.8128
30	73.83	0.6977	81.34	0.7852	79.83	0.7682	83.94	0.8152

IV. CONCLUSIONS

In this article, a new unsupervised band selection technique is proposed for hyperspectral images which involves determination of the attributes of bands, removal of redundant bands and providing ranks among the remaining bands. In the proposed method, DBSCAN clustering technique is used for obtaining band characteristics, and removing redundant bands. Measurement of non-Gaussianity of bands is used for prioritization of bands according to their discriminating abilities. A comparison of the proposed method with three other existing methods (two clustering based methods and one ranking based method) shows significant improvement in terms of overall accuracy and *Kappa* coefficient.

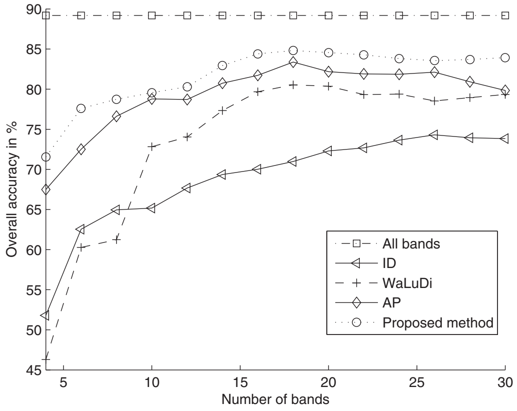

Fig. 4. Comparison of the performance of all the bands present in the dataset (denoted as,‘all bands’) along with those of ID, WaLuDi, AP and the proposed method in terms of overall accuracy for Indiana data set

REFERENCES

[1] P. K. Varshney and M. K. Arora, *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Berlin: Springer-Verlag, 2004.

[2] D. Landgrebe, “Hyperspectral image data analysis,” *IEEE Signal Processing Magazine*, vol. 19, pp. 17–28, 2002.

[3] J. M. Yang, P. T. Yu, and B. C. Kuo, “A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1279–1293, 2010.

[4] B. Mojaradi, H. A. Moghaddam, M. J. V. Zoej, and R. P. W. Duin, “Dimensionality reduction of hyperspectral data via spectral feature extraction,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2091–2105, 2009.

[5] L. O. Jimenez and D. A. Landgrebe, “Hyperspectral data analysis and supervised feature reduction via projection pursuit,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2653–2667, 1999.

[6] B. Guo, R. I. Damper, S. R. Gunn, and J. D. B. Nelson, “A fast separability-based feature-selection method for high-dimensional remotely sensed image classificaion,” *Pattern Recognition*, vol. 41, no. 5, pp. 1653–1662, 2008.

[7] C.-I. Chang and S. Wang, “Constrained band selection for hyperspectral imagery,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[8] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, “A joint band prioritization and band decorrelation approach to band selection for hyperspectral image classification,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, 1999.

[9] Q. Du and H. Yang, “Similarity-based unsupervised band selection for hyperspectral image analysis,” *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, 2008.

[10] A. Martinez-Uso, F. Pla, J. M. Sotoca, and P. Garcia-Sevilla, “Clustering-based hyperspectral band selection using information measures,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, 2007.

[11] Y. Qian and F. Yao, “Band selection for hyperspectral imagery using affinity propagation,” *IET Computer Vision*, vol. 3, no. 4, pp. 213–222, 2009.

[12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)*. Portland, Oregon: IEEE Computer Society, 1996, pp. 226–231.

[13] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data.*, 2nd ed. Boca Raton, London: CRC Press, 2009.

[14] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.