CLODD BASED BAND GROUP SELECTION

Muhammad A. Islam, Derek T. Anderson, John E. Ball, and Nicolas H. Younan

Mississippi State University
Department of Electrical and Computer Engineering
Mississippi State, MS 39762

ABSTRACT

Herein, we explore both a new supervised and unsupervised technique for dimensionality reduction or multispectral sensor design via band group selection in hyperspectral imaging. Specifically, we investigate two algorithms, one based on the improved visual assessment of clustering tendency (iVAT) and the other based on the automatic extraction of "block-like" structure in a dissimilarity matrix (CLODD algorithm). In particular, the iVAT algorithm allows for identification of non-contiguous band groups. Experiments are conducted on a benchmark data set and results are compared to existing algorithms based on hierarchical and c-means clustering. Our results demonstrate the effectiveness of the proposed method.

Index Terms— band grouping, dimensionality reduction, hyperspectral, iVAT, CLODD

1. INTRODUCTION

Hyperspectral imaging is a demonstrated technology for numerous earth and space-borne applications involving tasks such as target detection, invasive species monitoring and precision agriculture. However, hyperspectral imaging suffers from the "curse of dimensionality". Of particular interest is new theory for dimensionality reduction or identification of fewer spectral bands for multispectral versus hyperspectral imaging, typically relative to some specific task, which aids efficient computation, improves classification and lowers system cost. Most techniques can be divided into two broad categories—projection or clustering. Projection techniques require all bands initially (versus feature selection) and they are focused on reducing dimensionality. Approaches include *principal component analysis* (PCA), *Fishers linear discriminant analysis* (FLDA) and *generalized discriminant analysis* (GDA), *random projections* (RP), and kernel extensions. Some methods are unsupervised, e.g., PCA and RP, while others are supervised, e.g., FLDA and GDA. Clustering is unsupervised learning and it can be applied to hyperspectral imagery in a number of ways. While it does not automatically do dimensionality reduction, it helps to identify structure and one can take that information and use it for dimensionality reduction or band group selection. For example, in [1] Martinez et al. used an information measure to compute dissimilarity between bands and they used hierarchical clustering with Ward's single linkage to produce a minimum variance partitioning of the bands. In [2], Imani and Ghassemain used (hard) c-means for supervised band grouping. Martinez's method suffers from the limitations of vanilla hierarchical clustering, e.g., how to pick clusters from the dendogram. Imani and Ghassemain's approach suffers from the limitations of the c-means clustering algorithm, e.g., initialization, selection of c, and lack of ability compared to "soft" clustering (probabilistic, fuzzy or possibilistic).

Herein, we explore a new band grouping approach based on the *improved visual assessment of clustering tendency* (iVAT) [3]. This approach is well-grounded theoretically, and it produces visual results that an expert or additional clustering algorithm, e.g., *clustering on ordered dissimilarity data* (CLODD) [4], can exploit. Our goal was to identify an algorithm that could reproduce the structure that an expert currently finds and also be useful in the context of classification, which might demand different structure than an expert "sees". A common practice is to use a proximity metric like correlation to measure the similarity between bands. Often, contiguous bands are highly similar and this structure "shows up" if one produces an image of the similarity matrix. The CLODD algorithm analyzes a dissimilarity matrix, e.g., distances between vectors in a data set or bands in hyperspectral imaging, and it automatically finds "block-like" structure. Structure is often found in a proximity matrix according to squares of high-contrast along the matrix diagonal. The CLODD algorithm exploits two properties, "edginess" and "contrast". CLODD obtains contiguous band groups. However, we can automatically identify non-contiguous clusters (band groups) if we re-order the bands according to a method like iVAT. Herein, we explore both contiguous and

This effort was partially sponsored by the Engineering Research and Development Center under Cooperative Agreement number W912HZ-15-2-0004. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Engineering Research and Development Center or the U.S. Government. Material presented in this paper is also a product of the CREATE-GV Element of the Computational Research and Engineering Acquisition Tools and Environments (CREATE) Program sponsored by the U.S. Department of Defense HPC Modernization Program Office.

non-contiguous band groups and compare their relative performances. On one hand, contiguous is useful if we wish to identify a simpler sensor, however it could very likely be the case that non-contiguous bands share similarity and should be grouped and lead to more of a dimensionality reduction approach. Overall, the "answer" to this question is very task specific. While CLODD and iVAT are naturally unsupervised techniques, we also explore a supervised CLODD and iVAT approach based on the construction of a dissimilarity matrix using the data labels. The following sections describe the proposed approach and results.

2. METHODS

First, the hyperspectral data cube (image) is re-arranged to form a 2D data set (so spatial context is lost) where each row represents a pixel in the image and each column is a band. Let the data set be $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\} \in \mathbb{R}^{n \times b}$, where n is the number of pixels in the image and b is the number of bands. The label for each pixel, $\mathbf{x}_i$ is $y_i \in \{1, 2, \cdots, L\}$, where L is the number of classes. Figure 1 shows the major steps in the proposed approach.

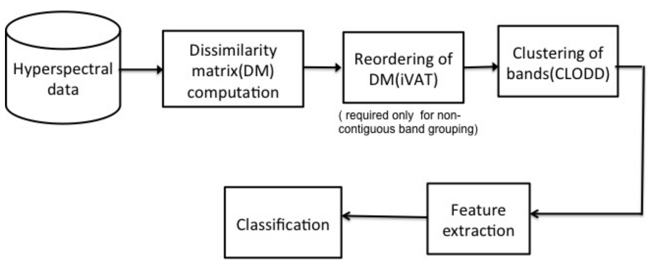
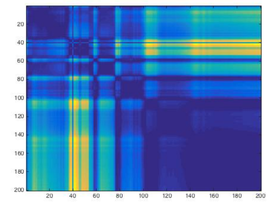


Fig. 1. Block diagram for the proposed method

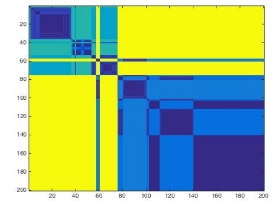
2.1. Calculation of dissimilarity matrix

The computation of the *dissimilarity matrix* (DM) differs for unsupervised and supervised band grouping. Note, there are numerous proximity measures and their aggregation that can, and have, been used for each, e.g., correlation, Bhattacharyya distance, Kullback-Liebler divergence, etc [5].

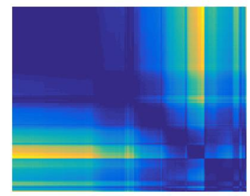
Herein, for supervised band grouping we compute the mean of the training samples in each class and a matrix, M, is formed such that ith row is the mean vector for the ith class. The square of the Euclidean distance between two bands, i and j is computed according to $d_1(i,j) = ||M_i - M_j||^2$, where M_i (and M_j) is the ith (jth respectively) mean vector. For unsupervised band grouping, we compute the square of the l_2 norm of the differences between pixel values for those two bands, $d_2(i,j) = ||X_{.,i} - X_{.,j}||^2$, where $X_{.,i}$ is a column vector of all pixel values for band i. The resultant pairwise dissimilarity matrix is sized $b \times b$. Figure 2 is an example for the Indian Pines data set.



(a)



(b)



(c)

(c)

Fig. 2. Supervised DM for Indian Pines data set; (a) "raw" DM, (c) VAT re-ordered, (b) iVAT enhanced minus the re-ordering step, and (d) iVAT enhanced with re-ordering.

2.2. Reordering of DM using iVAT

Figure 2(a) shows that some non-contiguous bands are similar. We can group those similar bands together in the matrix if we re-arrange the indicies (bands) using VAT [6]. VAT re-orders bands (data points in standard clustering) based on Prim's modified minimal single linkage. In [3], Havens et al. proposed an *improved VAT* (iVAT) that uses the graph theoretic distance to transform VAT to enhance our visualization and the effectiveness of the VAT algorithm. Figure 2(b) is the iVAT enhancement step on (a) without re-ordering and (c) is the enhanced iVAT on the re-ordered DM (b).

2.3. Clustering of a DM

CLODD, a "visual" clustering algorithm, which is more of an image processing technique than standard feature space clustering, exploits the "blockiness" in the raw DM or a reordered DM. Initially, VAT was created as a tool to help a user 'see" if there is any potential structure in the data. CLODD goes the next step and clusters the data. Its goal is to find a hard partitioning (aka clusters) via dark blocks along the matrix diagonal. While searching for the partition boundaries, it considers contrast between the on-diagonal dark block and off-diagonal lighter blocks known as "squareness" and visually apparent edges between the blocks, termed as "edginess".

Let D be the DM, U is a c partitioning and b_i is the number of (contiguous) bands in cluster i. Squareness is

$$E_{sq}(U;D) = \frac{\sum_{i=1}^{c}\sum_{s\in i, t\notin i} d_{st}}{\sum_{i=1}^{c}(b-b_i)b_i} - \frac{\sum_{i=1}^{c}\sum_{s,t\in i, s\neq t} d_{st}}{\sum_{i=1}^{c}(b_i^2-b_i)}.$$

Table 1. Classification accuracy (percentages) for unsupervised band grouping.

Bandgrouping method	Feature extraction method	corn (notill)	corn (min)	grass (pasture)	grass (trees)	hay (windrowed)	soybeans (notill)	soybeans (min)	soybeans (clean)	woods	Overall of 9 classes
CLODD(contiguous)	Mean	**71.84**	**55.62**	89.92	95.98	**98.72**	68.09	**81.56**	**52.14**	**98.74**	**79.30**
CLODD(contiguous)	Weight	66.43	47.98	87.41	94.97	**98.72**	64.86	78.98	43.18	98.65	75.95
CLODD(non-contiguous)	Mean	71.05	53.37	88.16	95.81	98.47	66.41	80.34	39.51	98.84	77.55
CLODD(non-contiguous)	Weight	66.78	47.98	87.66	**96.31**	**98.72**	65.50	78.93	41.96	98.74	76.12
Hierarchical	Mean	66.87	49.48	87.91	93.13	97.70	67.05	78.88	31.57	98.45	75.39
Hierarchical	Weight	65.30	51.42	**90.68**	95.14	97.95	**68.60**	78.52	47.66	98.55	76.78

Table 2. Classification accuracy (percentages) for supervised band grouping.

Bandgrouping method	Feature extraction method	corn (notill)	corn (min)	grass (pasture)	grass (trees)	hay (windrowed)	soybeans (notill)	soybeans (min)	soybeans (clean)	woods	Overall of 9 classes
CLODD(contiguous)	Mean	**71.40**	**55.62**	89.67	95.64	**98.72**	**70.41**	**82.02**	**52.14**	98.65	**79.54**
CLODD(contiguous)	Weight	65.30	48.73	89.67	95.48	**98.72**	63.57	77.86	43.38	98.65	75.59
CLODD(non-contiguous)	Mean	69.40	52.32	**88.92**	**96.15**	98.47	68.99	80.04	43.79	98.65	77.71
CLODD(non-contiguous)	Weight	65.74	47.23	**88.92**	95.81	98.47	64.60	78.42	43.58	**98.84**	75.79
Hierarchical	Mean	67.31	49.03	88.16	92.63	97.70	64.21	78.37	31.16	98.45	74.94
Hierarchical	Weight	65.48	50.67	89.92	95.14	97.95	68.73	79.28	47.45	98.55	76.90
c-means	Mean	66.78	53.97	86.15	93.13	97.19	67.96	77.66	33.20	97.78	75.45
c-means	Weight	63.73	49.63	83.88	93.13	97.19	68.60	77.10	44.81	95.65	74.86

The first part is the average between dark and non-dark regions. The second is just for dark regions. Edginess is

$$E_{edge}(U;D) = \frac{1}{c-1}\left(\sum_{j=1}^{c-1} \frac{\sum_{i=m_j-1}^{m_j} |d_{i,m_j} - d_{i,m_j+1}|}{b_j + b_{j+1}} + \frac{\sum_{i=m_j+1}^{m_{j+1}} |d_{i,m_j} - d_{i,m_j+1}|}{b_j + b_{j+1}}\right),$$

where $m_j = \sum_{k=1}^{j} b_k$ and $m_0 = 1$.

The objective function has two controlling parameters: mixing coefficient, α to trade-off between squareness and edginess; and γ to impose minimum cluster size,

$$E(U,D) = s(\min_{1\le i\le c} b_i, \gamma b)(\alpha E_{sq}(U,D) + (1-\alpha)E_{edge}(U,D)),$$

where s(.) is a spline function and is maximized with respect to U to obtain the optimum partition, U^*.

2.4. Feature extraction

Herein, we explore two feature extraction methods, mean and weight. In the 'mean' method, the resultant feature in each band group is the mean value of all bands in that group. In 'weight', the weight of each band is determined as $W_i = \frac{1}{R}\sum_{j\in C_i, j\neq i} \frac{1}{\epsilon + d(i,j)^2}$. The band with the highest weight, i.e., minimum average distance from all other bands in that group, is selected as the representative [1]. While we could have performed more advanced feature-level fusion or dimensionality reduction methods, the (simple) mean and weight were used as they are more easily translated into a physical sensor when designing a multispectral sensor.

2.5. Classification

Herein, we use a soft margin *support vector machine* (SVM) with RBF kernel for classification [7]. While we could have used a more sophisticated classifier, e.g., multiple kernel learning [8], we desired to reduce the number of "free parameters" to study just the proposed band grouping technique relative to related work.

3. PRELIMINARY FINDINGS

The publicly available benchmark data set Indian Pines is used to validate our method. The image has 145×145 pixels with a spatial resolution of 20 meters and 220 spectral channels (bands). We removed 20 water absorption bands, $104-108$, $150-163$ and 220; we consider only those classes with more than 5% of the total samples— Corn-notill, Corn-mint, Grass-pasture, Grass-trees, Hay-windowed, Soybean-notill, Soybean-mintill, Soybean-clean and woods.

We used a random jack-knife partitioning of the data, where 20% are for training and the remainder is testing. We modified CLODD and instead of letting it pick c we varied $c \in \{3, 4, ..., 35\}$ and α. We keep γ fixed to minimum cluster size of 2. Whereas CLODD will pick the "best visual clustering", we wanted to generate multiple candidate partitions and pick the "winner" based on classification accuracy. Next, the training data is standardized for each feature to have zero mean and unit variance. The testing data is standardized with the mean and standard deviation of the training data. We varied the RBF and adopted a one-vs-all strategy for multi-class classification. The best classification accuracy for each scenario is reported and used for comparison.

Table 1 is the classification accuracy for unsupervised band grouping. Non-contiguous CLODD-mean has the best overall performance. For soybean clean, it shows great improvement of approximately 5% over the Hierarchical based method. Note that wood has almost the same accuracy for different methods because its characteristics are distinctly different from all other classes which makes it easily classifiable by any of the methods. On the other hand, corn(min) and corn(notill) have very similar characteristics and are difficult to differentiate. For these two classes, contiguous CLODD-mean is the best.

Supervised band grouping is reported in Table 2. Contiguous CLODD-mean is still the top performer and it has a slightly better overall accuracy than unsupervised. For this data set, we see that unsupervised or supervised (how to construct the DM) does not make a significant difference in performance. Another point is performance appears to greatly depends on the band group feature extraction method. For example, CLODD and c-means favor the mean whereas Hierarchical likes weight.

4. CONCLUSION AND FUTURE WORK

Herein, we explored a visual clustering algorithm, CLODD, and re-ordering technique, iVAT, for contiguous and non-contiguous band group selection in hyperspectral imaging. Previously, these clustering techniques were used for feature space clustering instead of band group selection. Experimental results indicate that contiguous CLODD is the top performer. However, in future work we will explore the proposed algorithms on additional data sets, compare to more band selection and band group selection algorithms, explore additional DM functions for supervised and unsupervised, feature extraction methods, fusion strategies and more sophisticated classifiers.

5. REFERENCES

[1] A. Martínez-Usó, F. Pla, J. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, 2007.

[2] M. Imani and H. Ghassemian, "Band clustering-based feature extraction for classification of hyperspectral images using limited training samples," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1325–1329, 2014.

[3] T. Havens and J. Bezdek, "An efficient formulation of the improved visual assessment of cluster tendency (ivat) algorithm," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 813–822, 2012.

[4] T. Havens, J. Bezdek, J. Keller, and M. Popescu, "Clustering in ordered dissimilarity data," *International Journal of Intelligent Systems*, vol. 24, no. 5, pp. 504–528, 2009.

[5] J. E. Ball, D. T. Anderson, and S. Samiappan, "Hyperspectral band selection using multiple proximity measures and its aggregation for automated target detection," in *SPIE Defense and Security*, 2014.

[6] J. Bezdek, R. Hathaway, J. Huband, and M. Jacalyn, "Visual assessment of clustering tendency for rectangular dissimilarity matrices," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 5, pp. 890–903, 2007.

[7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[8] A. Pinar, T. C. Havens, D. T. Anderson, and L. Hu, "Feature and decision level fusion using multiple kernel learning and fuzzy integrals," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Aug 2015, pp. 1–7.