Unsupervised Classification on Hyperspectral Imagery

S. Avusali, S. O'Day, T. Lagaunne, A. Shafiekhani, and S. Sruthikesh
University of Missouri
Columbia, MO, 65211

***Abstract*—In this project, several methods were used to classify pixels in a hyperspectral data set taken from Indian Pines. In general, the data was first decomposed using one of the dimensionality reduction techniques discussed and a classification algorithm was applied to the remaining dimensions. Several different approaches to dimensionality reduction and classification were used and the results of these experiments were computed using the Rand Index on the classified image compared with the ground truth provided. First, a Principal Component Analysis (PCA) was performed on the entire image preserving the 8 most varied bands. Next, a Local PCA was performed using local regions in the image. After this, PCA was similarly performed using local bands in the image. Finally, these two methods of Local and Split Band PCA were combined. Several classification algorithms were then performed on each of these datasets. In the end, it was found that the best result came from combining GMM, K-Means, and LDA methods with voting to achieve a Rand Index of 88%. There were also a number of failed experiments for dimensionality reduction, using KPCA and Laplacian Eigenmaps, that may be worthy of further study.**

I. Introduction

The precise classification of remote sensing images has numerous real time applications such as environmental monitoring, plant and mineral exploration. The emergence of high resolution sensors and supercomputing devices has compelled the use of hyperspectral images for image analysis and classification. Hyperspectral imagery (HSI) captures a dense spectral sampling of reflectance values over a wide range of spectrum. This rich spectral information in every spatial location increases the capability to distinguish different physical structures, leading to the potential of a more precise image classification. However, they suffer from the dimensionality of the data. Hence, we use a two stage approach, dimensionality reduction and unsupervised classification.

The rest of the paper will be divided into sections discussing the implementation, experiments and corresponding results. The dataset used for these experiments is collected by AVIRIS from the Indian Pines site in Northwest Indiana. It consists of 224 bands from visible to near infrared (400 - 2500 nm). A subset of this data was then created by removing the bands covering water absorption, leaving 200 bands covering 145 x 145 pixels [1].

.

II. Implementation

The implementation covers two stages. First, the data is subjected to a dimensionality reduction. Next, this reduced dataset is processed through a classification algorithm. These methods are detailed below.

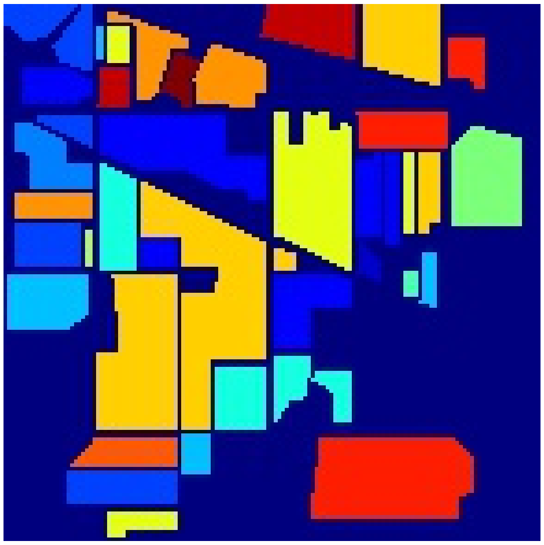

Figure 1. Image of the ground truth for the Indian Pines dataset

A. Dimensionality Reduction

1) Global PCA: Global PCA is the conventional dimensionality reduction. Principle Component Analysis is performed on all bands across the entire image. This should find highly uncorrelated bands across the entire image. Global PCA for our experiments was set to return 8 bands.

2) Localized PCA: A PCA implementation was used to determine bands that had a high covariance in specific regions of the image. Theoretically, this should preserve the bands that correspond to differences in local sections of the image. Localized PCA uses the spatial information to first divide the image into sections. PCA is then performed on each section, preserving 4 bands for each section. These bands are then combined for clustering.

3) Split Band PCA: The working theory behind the split band theory of PCA is that bands close to one another will be similar. With this in mind, there should be a minimal loss of information if we perform a component analysis on local clusters of bands. So, Split Band PCA takes the bands of the image and divides them up into bandwidths to perform PCA upon. In the event the number of dimensions is not divisible by the bandwidth, the extra bands are distributed evenly to the first chunks of classes one by one. PCA is then performed on each chunk of bands, transforming the data to preserve only the band with the highest covariance. For this experiment, the bandwidth used was of size 50, giving a total of 4 bands to use for clustering.

4) Local-Split: The Local-Split method simply combines the 4 bands from Local and 4 bands from Split and uses the 8 total bands for clustering.

B. Classification

1) K-Means & C-Means: The K-Means algorithm uses the Euclidean distance between the points and a centroid to determine the class that each point belongs to. Each centroid is initialized randomly and pixels are assigned to the closest centroid. The centers are then recalculated by averaging the points that have been assigned to it. The process is then repeated until the centroids converge to stable points [2].

The C-Means algorithm is similar, but instead of giving each pixel a specific cluster membership, it assigns each point a membership to all clusters based on a slightly modified objective function [3]. The assignment of clusters is then the same as for K-Means except the membership values is used as a weight for all points in the average. The maximum membership is then used as the class assignment.

2) ISODATA: The ISODATA algorithm is an extension of the K-Means algorithm. It uses the same assignment of clusters and update of centers, however it will discard clusters that become to small or split clusters that become to large. As such, it does not need to be explicitly given a number of clusters to create [4]. For purposes of this implementation, the initial clusters was set to 58 with a minimum cluster size of 20 pixels.

3) Latent Dirichlet Allocation (LDA): In Latent Dirichlet Allocation a document that has collection of words is assumed to be a mixture of topics. Specifically, LDA will find the topic mixture in each document and find the topic assignment for each words. The mixture of topics in each documents is assumed to have Dirichlet prior. For image segmentation problems, document can be some superpixel of images and the words are the label of pixels that might be constructed with some other algorithm (ex. K-Means or other clustering algorithms). The documents also can be constructed with some other algorithm like normalized cut or K-Means . Using K-Means, the number of documents used was set to 17 with 100 words.

4) Gaussian Mixture Model (GMM): Gaussian Mixture Models can be used for clustering. They assign each point to a multivariate Gaussian distribution such that they maximize the probability of belonging to that distribution.

5) Two Stage Classification: In this section, a two-stage classification using two different classifiers has been proposed. The first stage includes an unsupervised classification methods such as K-means or C-means which provides initialized labels for next stage. In the second stage, using an algorithm to eliminate outliers and given labels calculated in the first stage, SVM algorithm has been used to train for the inliers. Once we trained using inliers, it can be applied to entire data set and obtain different classes.

Two algorithms have been considered to eliminate outliers, the first one uses searching window to find data points that have similar neighbor labels and by doing that data points that are more consistence and far from class boarder have been selected. The alternative method which can be only used for C-mean classification in first stage uses weights of labels to threshold outliers, that is data point that are more likely to represent class characteristics will be selected and used to train SVM.

6) Voting: The final classification algorithm explored was a voting method on the results of the other methods. Because the labels from each result image do not necessarily correspond to the labels in one another, it is necessary to use an algorithm that can handle such situations. For this, we implemented the Hungarian method on these results [5]. The Hungarian method optimizes the combinations and relabels each image so that the labels are now the same as one another. The resultant images then vote to produce the final result. As implemented, the voting method only votes on images from the GMM, K-Means, and LDA algorithms.

III. Experiments

For each dimensionality technique, each clustering algorithm was applied and the results can be seen below. K-Means, C-Means and LDA algorithms require a random seed to initialize the centers. For purposes of these experiments, these random seeds are set to 100 for K-Means and C-Means; and 1 for LDA.

For the first experiment, each clustering algorithm was used on the Global PCA dataset. The results can be seen in Table I

Global PCA	8 dimensions
Clustering Algorithm	**Rand Index (%)**
K-Means	83.4
C-Means	55.0
ISODATA	82.5
GMM	85.0
LDA	85.2
Two Stage	72.7
Voting	83.9

Table I
Results for different cluster algorithm for Global PCA implementation

Next, each classification algorithm was applied to the Local PCA result, as seen in Table II.

After performing calssification using Local PCA, the Split Band approach was attempted. These results can be seen in Table III.

Finally, the two methods of Local and Split Band PCA were combined together and the bands were used for classification. These results can be seen in Table IV. For the voting in this method only the results from K-Means, GMM, and LDA were combined to create the best Rand Index score.

Along with these experiments, there was also an attempt to use non-linear dimensionality techniques, such as Kernel

Local PCA	4 dimensions
Clustering Algorithm	**Rand Index (%)**
K-Means	86.5
C-Means	86.8
ISODATA	86.5
GMM	86.0
LDA	86.4
Two Stage	85.3
Voting	87.0

Table II
RAND INDEX RESULTS FOR CLASSIFICATION ALGORITHMS ON LOCAL PCA REDUCTION

Split Band PCA	4 dimensions
Clustering Algorithm	**Rand Index (%)**
K-Means	85.0
C-Means	85.2
ISODATA	84.1
GMM	85.6
LDA	84.2
Two Stage	84.9
Voting	84.5

Table III
SPLIT BAND PCA RESULTS FOR DESCRIBED CLASSIFICATIONS

PCA and Laplacian Eigenmaps. These gave results much lower than random assignment (˜75%) and so were not pursued. It is possible however, that because these methods require several parameters to be set for calculation, they may not have been optimally set.

IV. CONCLUSION

Global PCA easily performed the worst out of all these methods, but it is still better than random chance for most clustering methods. The best results were obtained by using the Local-Split method of PCA with voting applied to the results of LDA, K-Means, and GMM. It is worth noting here that though we attempted to use non-linear KPCA and Eigenmap reduction techniques, we did not attempt at all to optimize the parameters. This was largely due to time of computation, and may be a subject to be further investigated.

REFERENCES

[1] Hyperspectral remote sensing scenes of indian pines data set. http://www.ehu.eus/ccwintco/index.php?$title = Hyperspectral\_Remote\_Sensing\_Scenes$.

[2] J. A. Hartigan and M. A. Wong, “Algorithm as 136: A k-means clustering algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[3] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, “A possibilistic fuzzy c-means clustering algorithm,” *Fuzzy Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 517–530, 2005.

[4] G. H. Ball and D. J. Hall, “Isodata, a novel method of data analysis and pattern classification,” DTIC Document, Tech. Rep., 1965.

[5] H. G. Ayad and M. S. Kamel, “Cluster-based cumulative ensembles,” in *Multiple Classifier Systems*. Springer, 2005, pp. 236–245.

Combined Local-Split PCA	4 dimensions
Clustering Algorithm	**Rand Index (%)**
K-Means	87.9
C-Means	87.2
ISODATA	86.5
GMM	87.2
LDA	87.4
Two Stage	86.3
Voting	88.0

Table IV
RESULTS FOR COMBINED LOCAL AND SPLIT BAND PCA

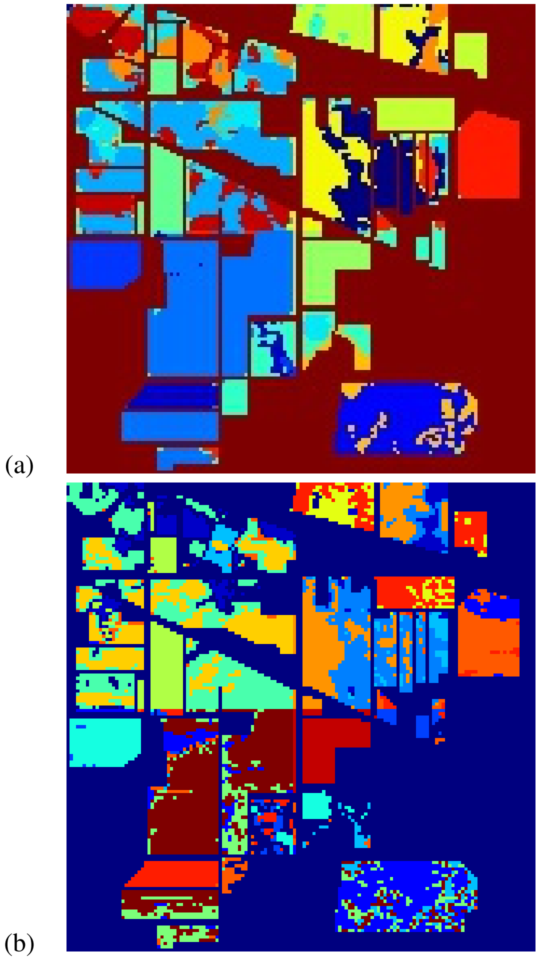

Figure 2. Image of clustering results for voting on (a) Local-Split PCA (b) LPCA with background removed.