[[1]](#footnote-1)

**Unsupervised Hyperspectral Image Segmentation of Indian Pines (April 2017)**

Jason Bonvie, Michael Lowe, Ruoyu Wang

University of Florida Department of Electrical and Computer Engineering

*Abstract*—Currently hyperspectral imaging is used to obtain a spectrum for each given pixel in an image by a remote sensing device. This large multi-dimensional information allows researchers to analyze an image to detect, identify, and classify objects in application. Experimentation reflects on the testing, abstraction, and implementation of the K-Means Algorithm, Principle Component Analysis, Neighboring Bias and the Gaussian Mixture Model to accurately generate a classified hyperspectral image.

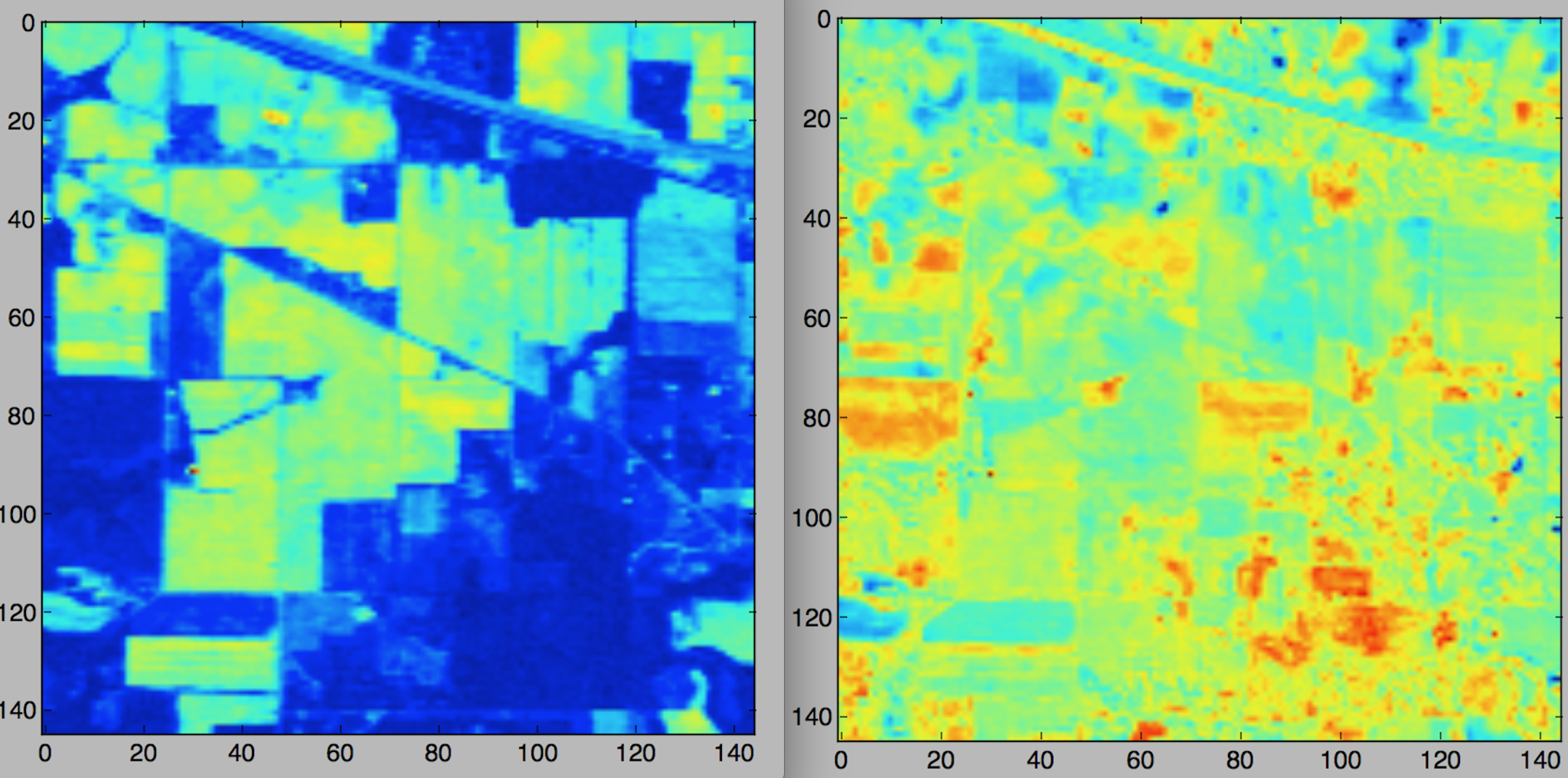
INTRODUCTION

Hyperspectral imaging (HSI) is performed using remote sensing devices. These devices obtain information about a specific object area in observance. By analyzing electromagnetic radiation transmitted through, reflected from, or absorbed by a local observed medium; i.e. Land; imaging information can be gathered and applied to obtain specific physical information about said medium [1].

The overall objective in applying hyperspectral image segmentation is to correctly cluster and classify the data provided to our system in comparison to the actual overlay of the geographic landscape.

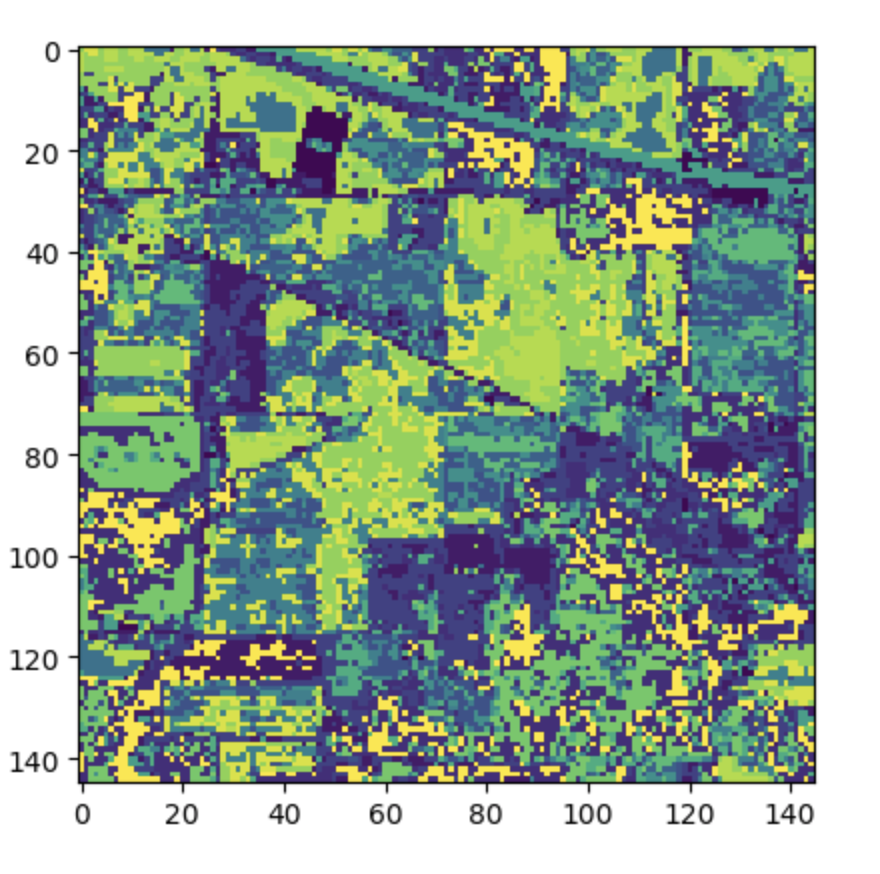
# PCA Experimentation

Principle Component Analysis(PCA) is a procedure used to reduce datasets by identifying a small group of uncorrelated data with the largest variance—effectively reducing the observed feature space. This method allows for the given Indian Pine data to be reduced from a feature space of 200 components to a 20—Essentially providing 97 percent of the variance in from the original dataset with 10 percent of the total components.



**Figure 1.0** Useful hyperspectral band(right) vs Useless hyperspectral band(left).

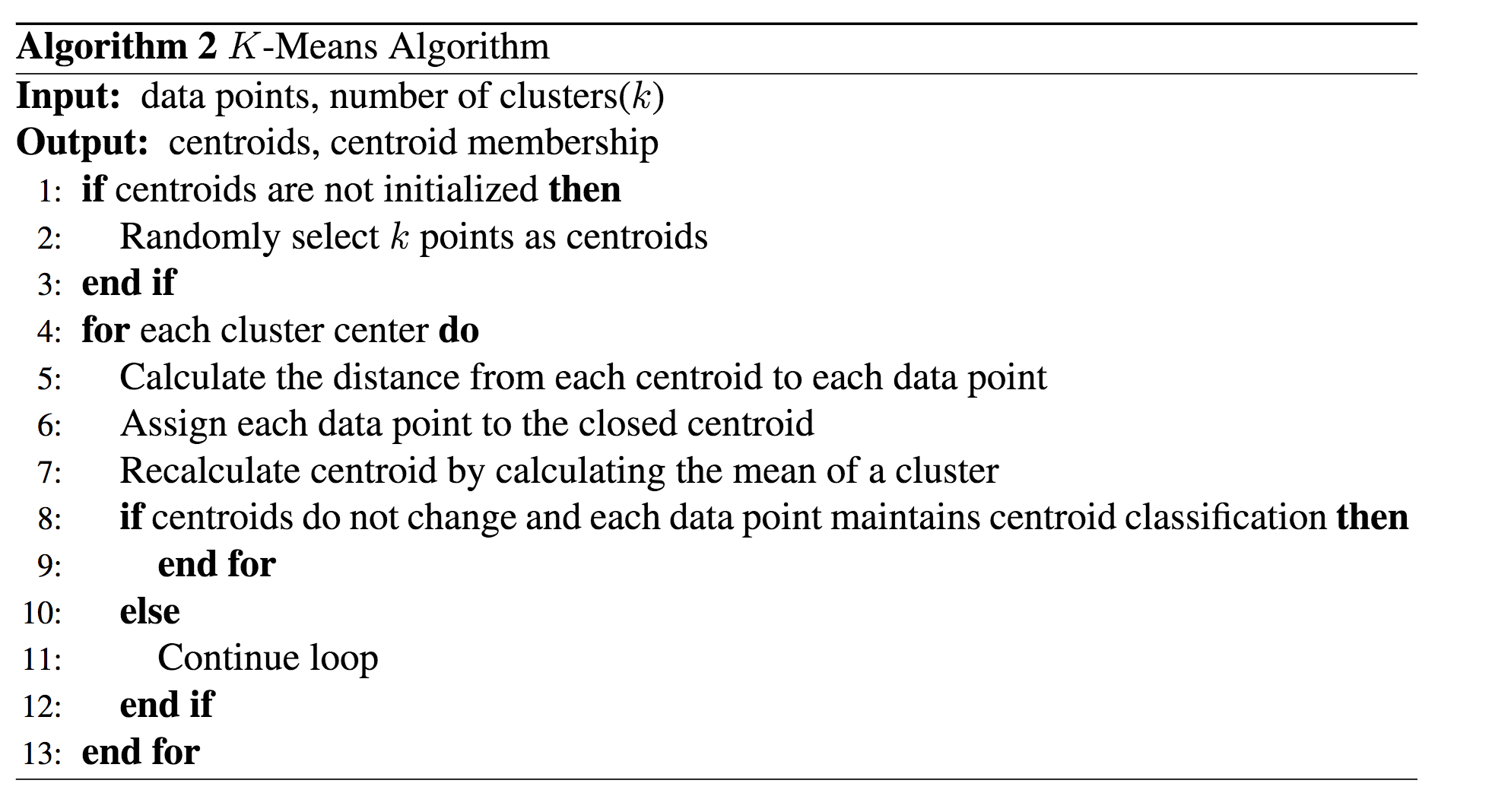
Because there were several bands in the dataset that provided little to no information, PCA allows one to narrow the feature set for a smaller set of data with higher variance. The motivation behind doing so is to weed out what would be 23 percent of components that provided useless features and unneeded noise to the dataset (Bands 30-100). The following figure illustrate PCA as performed on the 200-component and after reduction.



**Figure 1.1** Reduced 20 component HSI after PCA performed.

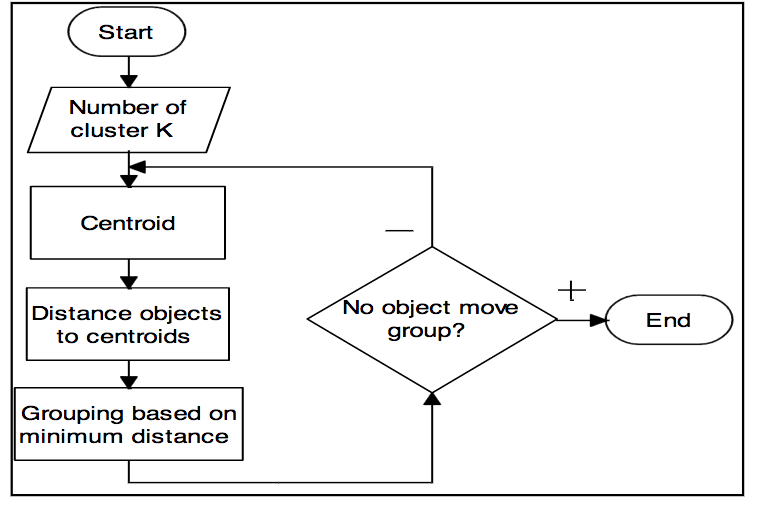
# K-Means Experimentation

The K-Means Algorithm (see figures 1.3 and 1.4) is an iterative clustering technique used to classify data points in an unsupervised environment. The Algorithm compares the distance of each data point to the center of a preset number N of select centroids. The centroid is then updated at the end of each traversal until the centroid values become static [2].



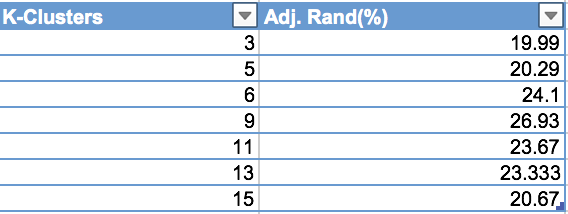
**Figure 1.3** K-means algorithm structure [2].

After applying PCA on the sample bands and reducing the feature space, an N-Clusters vs Adjusted Rand Score comparison can be made. The experimentation process can is mapped out by **figure 1.3** By adjusting the number of clusters used to classify the data points in the spectral bands, and updating the neighboring bias of each data point to its nearest centroid, an optimal Rand Score is found at 9 clusters.



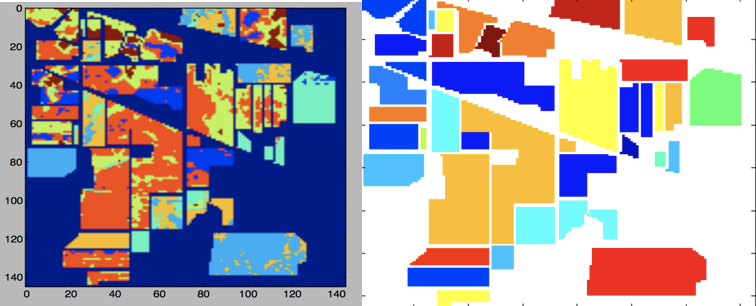
**Figure 1.**4 State Diagram for K-Means process [1]

Though the ground truth provided implies that the number of clusters used for classification cannot exceed 16, there was a significant drop off in classification accuracy the further the clusters were moved away from its local maximum value of 9(As observed in **figures 1.5** and **1.6**). However, the problem encountered from this inference is that the closer the number of clusters become to the ground truth, the less accurate the model becomes.



**Figure 1.5** Adjusted Clusters Tested by K-Means Algorithm.

**Figure 1.6** Plot of Adjusted Clusters Tested by K-Means Algorithm.



**Figure 1.7** Experimental Yield(left) vs Ground Truth[3](right)

# Conclusions and Drawbacks

To more effectively reduce the feature space used in analyzing the hyperspectral bands, KPCA was applied to the data; however, given that the KPCA algorithm has a time complexity of **O(p2n+p3)**, this process took too long to draw kind of inference simply because of dimensionality of the data, and local computing power.

Given the dimensionality of the data, and the ground truth supplied, the optimal clustering classification levels were surprisingly far away from the ground truth. Given that most of the data from all classes exists in 9-10 larger groups than the total 16 classes the classification algorithm used is not sensitive enough to pick up on the smaller classes or able to accurately separate them enough to generate a higher accuracy.

References:

[1] Sahar A.El\_Rahman, “Hyperspectral Image Classification Using Unsupervised Algorithms”, Benha University, Cairo, Egypt, 2016.

[2] Hao Sun, “MAP-GUIDED HYPERSPECTRAL IMAGE SUPERPIXEL SEGMENTATION

USING SEMI-SUPERVISED PARTIAL MEMBERSHIP

LATENT DIRICHLET ALLOCATION”, M.S. thesis, Graduate School at the University of Missouri, Columbia, Missouri, year 2016.

## [3] University of the Basque Country,

“Hyperspectral Remote Sensing Scenes”,

Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral\_Remote\_Sensing\_Scenes

1. [↑](#footnote-ref-1)