

# Hyperspectral Image Classification Using Unsupervised Learning Algorithms

Iqra Ejaz

Department of Electrical Engineering  
Institute of Space Technology  
Islamabad, Pakistan  
46iqramalik@gmail.com

Wafa Fatima

Department of Electrical Engineering  
Institute of Space Technology  
Islamabad, Pakistan  
wafafatima23@gmail.com

**Abstract**—Hyperspectral image (HSI) classification is a mechanism of analyzing differentiated land cover in remotely sensed hyperspectral images. In last two decades, a number of different types of classification algorithms have been proposed for classifying hyperspectral data. These algorithms include supervised as well as unsupervised methods. Each of these algorithms has its own limitations. In this research, three different types of unsupervised classification methods are used to classify different datasets i-e Pavia Center, Pavia University, Cuprite, Moffett Field. The main objective is to assess the performance of all three classifiers K-Means, Spectral Matching, and Abundance Mapping, and observing their applicability on different datasets. This research also includes spectral feature extraction for hyperspectral datasets.

**Index Terms**—Hyperspectral Images, Classification, K means, Spectral Matching, Abundance Estimation

## I. INTRODUCTION

Recently, people have begun to obtain remotely sensed hyperspectral images with high spectral and spatial resolution easily. Because of the high resolving power of hyperspectral images for fine spectra, it is very useful to identify unique materials and to differentiate among different materials based on these spectrums. Each pixel in hyperspectral images is represented by a vector whose values correspond to the spectrum of reflected light. Using hyperspectral remote sensing technology, the classification is much more accurate than using any other remote sensing technology. Hyperspectral images have a variety of applications [1-4].

- To detect a new change in the hyperspectral imaging of the same geographical regions.
- For noninvasive cancer detection.
- For military object detection using multiple information extracted from hyperspectral imagery.
- Forensic document examination.

There are several techniques including computer vision, machine learning, pattern recognition, and statistical analysis that are used for the processing of hyperspectral images. Hyperspectral images are difficult to view because they contain a vast range of information on spectrums in narrow bands. For displaying HSI Usually three bands are used which can cause a lot of information loss making it not a very feasible method. Several supervised and unsupervised methods based on band selection can be used for the visualization of HSI, each of

which has its limitations. Minimum Estimated Abundance Covariance (MEAC) is a supervised method that depends on the spectral signatures of the data [7]. An unsupervised method like Principle Component Analysis (PCA), in which three bands are obtained from the linear combination of data [15] can also be used to view HSI. Another method, Saliency Based Visualization of Hyperspectral Images [16] has been recently proposed that uses spectral and spatial information of all the bands. Optimized class-separability in HIS [17], is another method that uses a fusion of three methods (Principle Component Analysis, Maximum Noise Fraction, Independent Component Analysis). In recent years several methods for the classification of hyperspectral images have been proposed. Based on the availability of training samples these methods can be either parametric or non-parametric and supervised or unsupervised, each of these methods have their limitations. In the supervised classification method, data is classified for each class using a set of representative data called training. The minimum distance method classifies the unknown image data, such that mean vectors for each class are used and the euclidean distance of each unknown pixel from the mean vector is calculated and the pixels are classified to the nearest class [7]. In the maximum likelihood method, it is assumed that the probability density function for each class is normal [8]. For the classification of data where a small number of training samples are available, the Support Vector Machine (SVM) method is used [7]. Integration of important spectral and spatial information is also of great interest. A supervised spatial enhanced density peaks-based algorithm was proposed for this purpose [9]. In this technique, first centers for classes are determined using a training dataset and then these centers are assigned to test samples. Unsupervised classification divides the hyperspectral dataset into several categories based on their similarities [5,6]. Change in the spectrum shape often corresponds to changes in land cover. Therefore, fine changes in land can be detected by spectral matching [1]. K-means and Spectral Clustering (SC) methods use an objective function, which is typically the sum of distances to a set of assumed cluster centers and is optimized until the best cluster centers are obtained [10,11]. Hyperspectral image classification always faces a variety of problems, which includes high dimensionality, spectral variability, and mixing pixels.

Classification accuracy is directly dependent on the number of training samples. If the number of training samples is less than features, it greatly reduces the performance of the classifier. This problem is popularly referred to as the “Hughes phenomenon” [12]. To tackle this difficulty, feature extraction methods are used to reduce dimensionality by selecting prominent features. Unsupervised classification methods do not require any representative samples to train the data. The popular unsupervised methods are Principle Component Analysis (PCA) [13] and Independent Component Analysis (ICA) [14]. The proposed study deals with different types of unsupervised classifying algorithms. These algorithms include K-means, Spectral Matching, and Abundance Mapping. These classifying methods are applied separately on three types of datasets, and two different programming languages (Python, MATLAB) are used to get the desired results.

## II. METHODOLOGY

Unlike Normal imagery captured by standard cameras, hyperspectral imagery holds detailed spectrum information of each pixel [18]. This detailed information per pixel opens the realms of many new image classification techniques. The spectrum information helps us in segmenting the image based upon the reflective properties of materials and objects present inside the Image [19]. In this work, four separate hyperspectral image datasets i.e Cuprite, Moffett field, PaviaU, and Pavia Center have been segmented according to three different unsupervised classification methods to judge the merits of every technique independently.

- K means
- Spectral Matching through SAM
- Abundance estimation

DataSet	Sensor	Bands	Wavelength
Pavia Center	ROSIS	102	430-860nm
Pavia University	ROSIS	103	430-860nm
Cuprite	AVIRIS	224	400-2500nm
Moffett	AVIRIS	224	400-2500nm

### A. Through K means

K means clustering algorithms have been in use for quite a while now for unsupervised classification problems. In this algorithm hyperspectral data is organized in such a way that pixels having similar spectral characteristics are clustered together thus no need for the labeled data [20]. Due to the large spectral variability present in hyperspectral data, classification with the K Means clustering algorithm becomes a challenging task. K-means optimizes the error of each group as min by calculating the centers of k number of groups.

$$\sum_{j=1}^k \sum_{i=1}^{nk} \|x_i^j - c_j\|^2$$

where

$$\|x_i^j - c_j\|^2$$

is the euclidean distance between a data point  $x_j$  of the cluster  $j$  ( $nk$  is the observations within each cluster) and the

cluster center  $c_j$ , it is the point that minimizes the equation also known as centroid of cluster  $j$  [21]. For K means algorithm to work effectively, it's important to initialize it properly therefore a little inside about the data is important. Applying this method for the classification of hyperspectral images, satisfactory results can be obtained but the biggest drawback of this method is the computational complexity.

### B. Through Spectral Angle Mapper(SAM)

Hyperspectral images contain a pure and distinct signature of objects and materials present inside the data set known as endmembers [22]. To analyze and classify the hyperspectral data effectively, endmembers are of great interest therefore, a lot of detail spectral libraries are created such as ECOSTRESS, USGS by measuring and analyzing the spectral components of natural as well as man-made objects. Every single object possesses a distinct spectral signature and is recorded and stored inside the library. Spectral matching is a technique employed to determine various regions as well as to classify objects present inside the hyperspectral dataset by extracting the endmember information from it and as the name suggests by finding the similarity index between the endmember signatures and the spectrally distinct signatures present inside the library [23]. In this algorithm, spectra are considered as the vectors in space with dimensionality equals to the number of bands ( $nb$ ), and then the angle between two spectra is calculated to determine the spectral similarity between them. since the

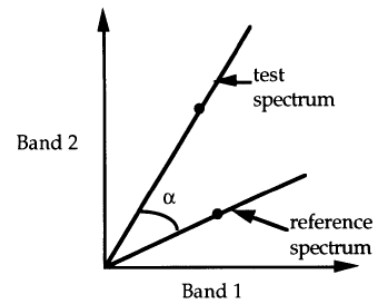


Fig. 1: Plot of a reference spectrum and test spectrum of a two-band image

similarity index is calculated through the angle alpha therefore this algorithm is often known as spectral angle mapper(SAM).

$$\arccos\left(\frac{\sum_{t=1}^{nb} (t_1 r_1)}{((\sum_{i=1}^{nb} (t_i)^2)^{1/2})((\sum_{i=1}^{nb} (r_i)^2)^{1/2})}\right)$$

where  $nb$  = total number of bands

$r$  = reference spectrum

$t$  = spectrum to be matched

As it can be observed from the equation above, the length of the vectors and the angle are not dependent upon each other thus completely insensitive to the gain factors therefore direct comparison can be made between laboratory spectra and remotely sensed spectral reflectances with unknown gain factor [23]. A new data cube is created after using spectral

angle mapping with the total number of bands being equal to the number of reference spectra present inside the reference library used in the mapping. Mainly this technique is employed in creating a ground truth file for completely unseen hyperspectral data and can also be used to identify the regions present in the hyperspectral image.

### C. Through Abundance Estimation

Linear unmixing is one of the most commonly used techniques for the determination of content obtained via hyperspectral image. The main concept behind it is that each captured pixel, a combination of  $N_b$  spectral bands, is represented in the form of a linear combination [24]. The linear combination comprises of a group of endmembers (spectrally pure constituent spectra) and they are weighted using an abundance factor. This factor helps in the establishment of the endmember proportion in each pixel that is being inspected. The linear model is based on the assumption of neglecting the scattering effects and secondary reflections. Therefore the resultant spectra under consideration is a linear combination of pure spectral signatures of the different materials present inside the mixed pixel [25]. The Linear unmixing technique comprises of three steps:

- Estimating endmembers population
- Extracting the endmembers
- Calculating the abundances

Other than these, the following operations are used for enhancement of unmixing results:

- Filtering of noise
- Dimensional reduction of data set

For every step involved in linear unmixing, a large number of algorithms have been proposed over time. For example, HySIME (hyperspectral subspace identification by minimum error) algorithm and the virtual dimensionality algorithm are used for estimation of the endmembers population [26]. With the help of these algorithms, we can find the intrinsic dimensionality of the hyperspectral data sets. VCA (vertex component analysis) and N- finder algorithms are vastly used by researchers for the extraction of endmembers. The endmembers are extracted as image pixels having the largest volume in the N-finder algorithm. This method required more computational powers since it requires the solution of  $p \times p$  determinants. Contrary to this, the endmembers are extracted as pixels having the largest orthogonal projection in the VCA algorithm [26]. The algorithm is based on the  $p$  iterations which involve computation of the orthogonal projection of hyperspectral image pixel making it computational more complex but faster as compared to the N-finder algorithm [27]. Lastly, the FCLSU (fully constraint least squares linear spectral unmixing) algorithm is used for the calculation of the abundances [28]. All the algorithms mentioned fulfill their purpose and need efficiently.

## III. RESULTS AND DISCUSSIONS

### A. Definition of Problem

To identify the regions present inside the Cuprite, Pavia, PaviaU, and Moffett Field data sets through k means, spectral matching and abundance estimation models.

### B. Results

For the classification of the Pavia Centre dataset, the K-Means clustering algorithm was used. This algorithm works well for small datasets, in which a number of regions can be predicted. Less number of bands makes it easy to visualize and analyze this dataset in order to find the optimized value for  $k$ . Results given in figure(3) and figure(4) were obtained by initializing

$$k = 9$$

The processing time was slower due to higher dimensional data, making it computationally complex.

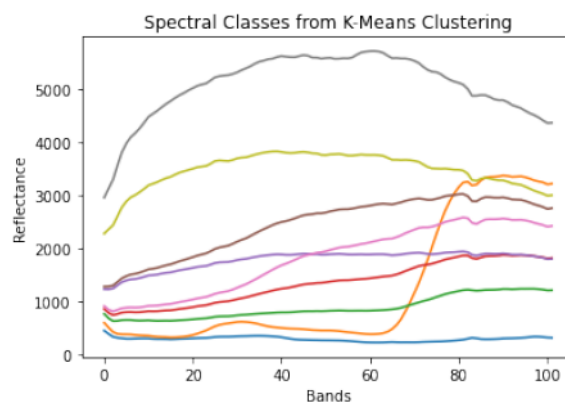


Fig. 2: Spectral Classes obtained from K means Algorithm

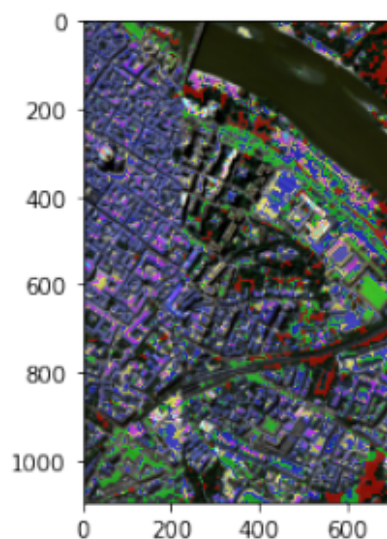


Fig. 3: Pavia Center Colour Segmented Image



Pavia University dataset was segmented through spectral matching, the ECOSTRESS library available in Matlab was used as a reference for the endmembers obtained from the dataset at hand. N findr algorithm was used for the calculation of endmembers. The wavelengths of calculated endmembers were matched with the natural and man-made objects' wavelengths from the library. This method works well for the unsupervised classification of data where we don't have prior knowledge and the dataset is not easy to visualize. Therefore, by estimating the total endmembers and comparing their spectral information with the reference, we can roughly segment our data and get more insight as to what kind of materials are present inside the image. Figure (5) and figure (6) are the results obtained by using the ECOSTRESS library as a reference. The segmented image is compared to the estimated RGB Image in figure(4) whereas, the pure spectral signatures available in the ECOSTRESS library are shown in figure(5).

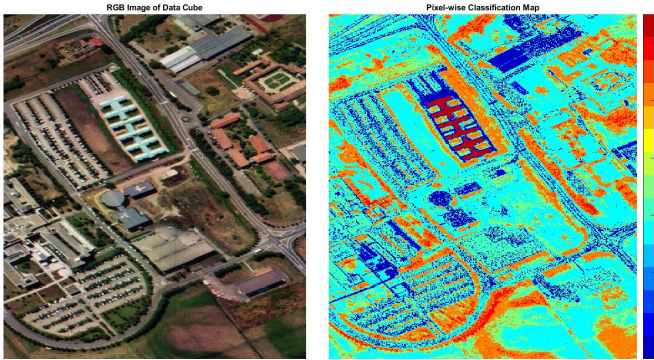


Fig. 4: Pavia University segmented image vs RGB estimated Image



Fig. 6: Estimated RGB Image

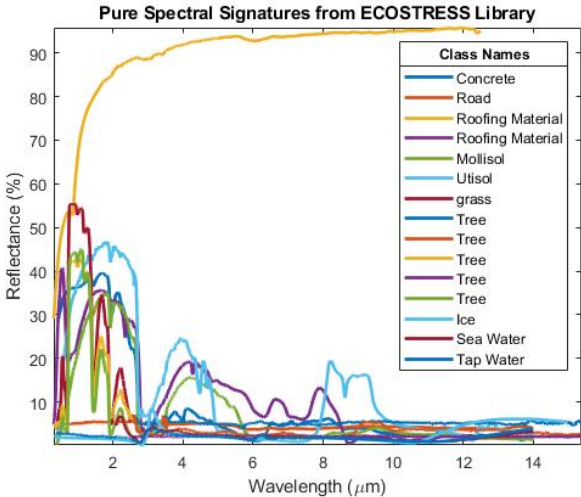


Fig. 5: Library EndMember spectrum

Cuprite and Moffett's fields are large datasets. Hence Abundance mapping algorithm was used for the segmentation of these datasets. The results obtained for Cuprite are shown in

figure(6), figure(7), and figure(8). The Cuprite data is a challenge as it contains a lot of distinct spectral regions. The use of spectral matching for this dataset resulted in an exponential increase of processing time because 20 endmembers have to be matched with spectral data of thousands of various types of minerals reference wavelengths present inside the UGSC library. Using abundance mapping reduces the processing time by a considerable factor. The results of applying K-Means on Cuprite will not be much reliable, because it requires estimation of k just from visualizing it. Moffett field dataset is also very large and tricky to handle so it was reduced to our region of interest as shown in figure (9). Abundance mapping was applied for its segmentation and the results obtained in figure(10) were satisfactory.

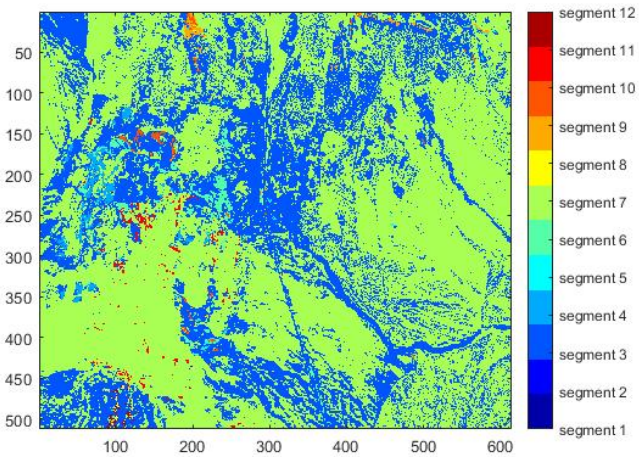


Fig. 7: Segmented Cuprite Image

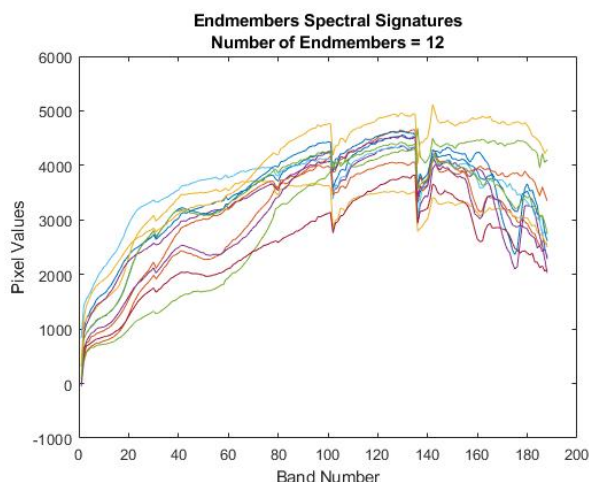


Fig. 8: EndMembers spectrum

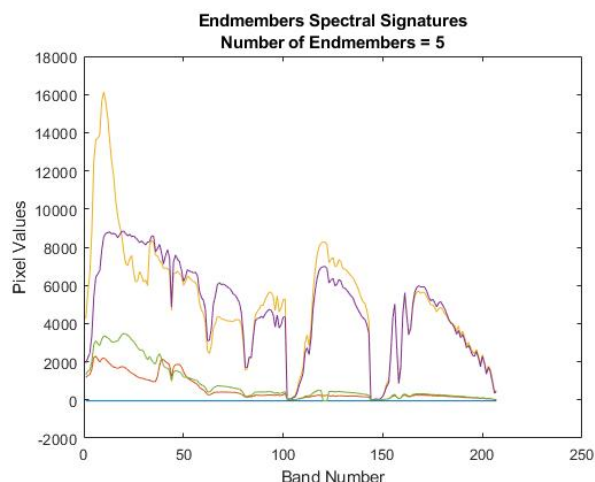


Fig. 11: EndMembers spectrum

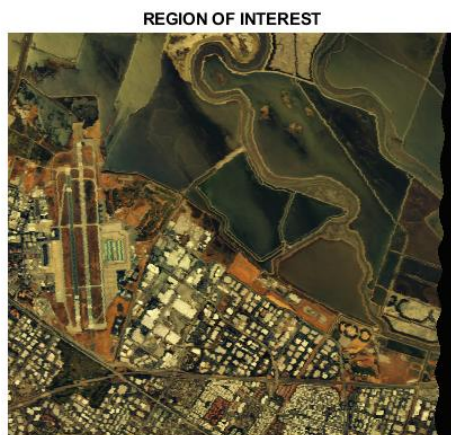


Fig. 9: Moffett Field Region of interest

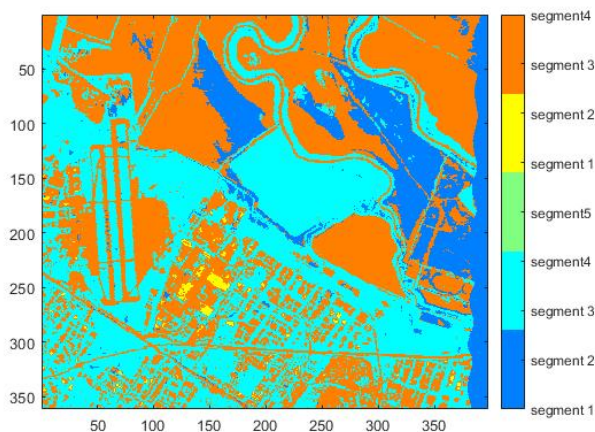


Fig. 10: Segmented Moffett Field Image

#### IV. CONCLUSION

The applicability of different classification techniques on different datasets was analyzed. The pure spectral signatures for each dataset were extracted and the image was then classified according to the endmembers obtained. K means clustering was found to work better with low dimensional data, whereas spectral matching was found to work better for datasets with a lesser number of endmembers and datasets with basic objects present in them. Abundance estimation mapping worked better on higher dimensional as well as on lower-dimensional data.

#### REFERENCES

- [1] Wen X, Yang X. A new change detection method for two remote sensing images based on spectral matching. In 2009 International Conference on Industrial Mechatronics and Automation 2009 May 15 (pp. 89-92). IEEE.
- [2] Pike, R., Lu, G., Wang, D., Chen, Z.G. and Fei, B., 2015. A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging. *IEEE Transactions on Biomedical Engineering*, 63(3), pp.653-663.
- [3] Ke, C., 2017, December. Military object detection using multiple information extracted from hyperspectral imagery. In 2017 International Conference on Progress in Informatics and Computing (PIC) (pp. 124-128). IEEE.
- [4] Khan, M.J., Khan, H.S., Yousaf, A., Khurshid, K. and Abbas, A., 2018. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, 6, pp.14118-14129.
- [5] Xu, R. and Wunsch, D., 2005. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3), pp.645-678.
- [6] Von Luxburg, U., 2007. A tutorial on spectral clustering. *Statistics and computing*, 17(4), pp.395-416.
- [7] Huqqani, I.A. and Khurshid, K., 2014. Comparative study of supervised classification of urban area hyperspectral satellite imagery. *Journal of Space Technology*, 4(1), pp.7-14.
- [8] Tou, J.T. and Gonzalez, R.C., 1974. Pattern recognition principles.



- [9] Li, T., Zhang, J. and Zhang, Y., 2016, August. A supervised density-peaks-based classification approach for hyperspectral images. In 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-4). IEEE.
- [10] Wagstaff, K.L., Shu, H.P., Mazzoni, D. and Castano, R., 2005. Semi-supervised data summarization: Using spectral libraries to improve hyperspectral clustering.
- [11] Sanderson, M., Christopher, D. and Manning, H., 2010. Introduction to information retrieval. Natural Language Engineering, 16(1), p.100.
- [12] Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE transactions on information theory, 14(1), pp.55-63.
- [13] Wold, S., Esbensen, K. and Geladi, P., 1987. Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), pp.37-52.
- [14] Hyvärinen, A. and Oja, E., 2000. Independent component analysis: algorithms and applications. Neural networks, 13(4-5), pp.411-430.
- [15] Jia, X. and Richards, J.A., 1999. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. IEEE transactions on Geoscience and Remote Sensing, 37(1), pp.538-542.
- [16] Khan, H.A., Khan, M.M., Khurshid, K. and Chanussot, J., 2015, July. Saliency based visualization of hyper-spectral images. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 1096-1099). IEEE.
- [17] Sattar, S., Khan, H.A. and Khurshid, K., 2016, July. Optimized class-separability in hyperspectral images. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 2711-2714). IEEE.
- [18] Ranjan, S., Nayak, D.R., Kumar, K.S., Dash, R. and Majhi, B., 2017, January. Hyperspectral image classification: A k-means clustering based approach. In 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 1-7). IEEE.
- [19] Villa, A., Chanussot, J., Benediktsson, J.A. and Jutten, C., 2010. Spectral unmixing for the classification of hyperspectral images at a finer spatial resolution. IEEE Journal of Selected Topics in Signal Processing, 5(3), pp.521-533.
- [20] Cariou, C. and Chehdi, K., 2015. Unsupervised nearest neighbors clustering with application to hyperspectral images. IEEE Journal of Selected Topics in Signal Processing, 9(6), pp.1105-1116.
- [21] Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A. and Vigo-Aguiar, J., 2017. Yinyang K-means clustering for hyperspectral image analysis. In Proc. 17th Int. Conf. Comput. Math. Methods Sci. Eng (pp. 1625-1636). Cadiz, Spain: Rota.
- [22] Keshava, N., 2003. A survey of spectral unmixing algorithms. Lincoln laboratory journal, 14(1), pp.55-78.
- [23] Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J. and Goetz, A.F.H., 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote sensing of environment, 44(2-3), pp.145-163.
- [24] Guerra, R., Santos, L., López, S. and Sarmiento, R., 2015. A new fast algorithm for linearly unmixing hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing, 53(12), pp.6752-6765.
- [25] Parra, L.C., Spence, C., Sajda, P., Ziehe, A. and Müller, K.R., 1999, November. Unmixing Hyperspectral Data. In NIPS (pp. 942-948).
- [26] Chang, C.-I., and Q. Du. "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery." IEEE Transactions on Geoscience and Remote Sensing 42, no. 3 (March 2004): 608–19. <https://doi.org/10.1109/TGRS.2003.819189>.
- [27] Winter, M.E., 1999, October. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Imaging Spectrometry V (Vol. 3753, pp. 266-275). International Society for Optics and Photonics.
- [28] Keshava, N. and Mustard, J.F., 2002. Spectral unmixing. IEEE signal processing magazine, 19(1), pp.44-57.