

Remote Sensing by using Unsupervised Algorithm

Afshan Salim

Department of Electrical Engineering
Institute of Space Technology
Islamabad, Pakistan.
Email: afshansaleem2015@gmail.com

Abstract—Hyper-spectral images contain wide range of bands or wavelength due to which they are rich in information. These images are taken by specialized sensors and then investigated through various supervised or unsupervised learning algorithms. Data that is acquired by hyper-spectral image contain plenty of information hence it can be used in applications where materials can be analyzed keenly, even smallest difference can be detected on the basis of spectral signature i.e. remote sensing applications. In order to retrieve information of concerned area, image has to be grouped in different segments and can be analysed conveniently. In this way only concerned portions of the image can be studied that have relevant information and the rest that do not have any information can be discarded. Image segmentation can be done to assort all pixels in groups. Many methods can be used for this purpose but in this paper we discussed k mean clustering to assort data in AVIRIS cuprite, AVIRIS Muffet and Rosis Pavia in order to calculate number of regions in each image and retrieved information of 1st, 10th and 100th band. Clustering has been done easily and efficiently as k mean algorithm is the easiest approach to retrieve information.

I. INTRODUCTION

Earth is under observation by thousands of satellites that helps us to get multiple types of information. Most of these information uses images of earth that are taken from specialised sensors and imagers. These images are then processed to retrieve useful data. So taking these images is one of the basic concern. Now a day hyper-spectral imaging technology has attracted many researchers as it gives treasure of information. Hyper-spectral sensors are designed to capture large number wavelength bands [1]. Hence hyper-spectral image contains wide and narrow range of wavelength which makes these images a source of rich information [2]. A Hyper-spectral image forms a 3-D (x,y,z) data cube in which x,y are spatial dimension and z is spectral resolution as shown in figure 1. These images are characterized by spatial and spectral resolution [3]. Spatial resolution is the smallest distinct detail in an image [4] and spectral resolution is the ability of sensor to distinguish the range of different wavelengths present in image, these wavelength intervals is also known as bands.

Different sensors have different spatial and spectral values. The images we are working on are taken from AVIRIS and ROSIS. Their spectral properties can be seen in table I [1].

Due to the rich details provided by hyper-spectral images, hyper-spectral imaging is used for various applications like agriculture and water resource management[5] [6], military defence, art preservation and archaeology [7] [8], medical diagnosis [9] [10], crime scene analysis[11][12], document

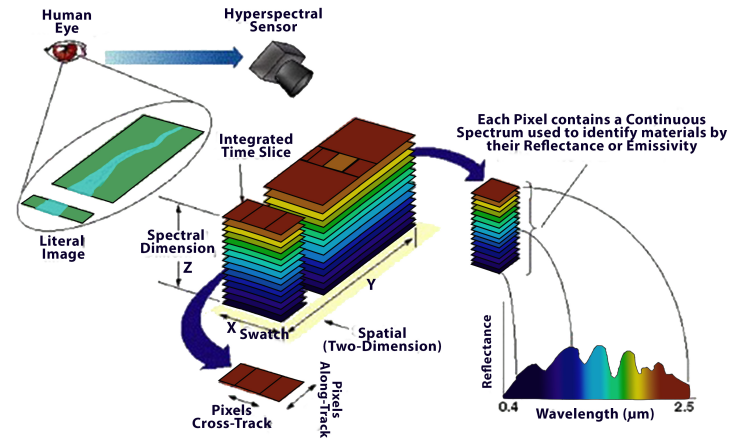


Fig. 1: Hyper-spectral imaging

Sensor	Spectral Band	Spectral Range	Spectral Resolution
AVIRIS	224	0.38-2.5	4-20
ROSIS	115	0.42-0.873	2

TABLE I: Spectral band, range and resolution of AVIRIS and ROSIS sensors

imaging [13], and forensics [14].

Hence to attain full information of the hyper-spectral each region of the image is analysed along with its full spectrum signature. As each image contains thousands of small regions (pixel) each pixel has its own spectral signature which defines the property of material. For this purpose both supervised and unsupervised learning algorithm can be used.

The primary goal of satellite imagery study is to accurately assess terrain features and extract relevant data [14]. The two main types of algorithms are unsupervised (clustering) and supervised (classification) algorithms. Every classification and clustering algorithm gives different performance on different type of datasets. The performance depends upon the type of data used, scale of data, training sample size and the type of classification method used[15]. In this paper we used k-means clustering algorithm that is an unsupervised learning technique. Figure 2[16] shows unsupervised clustering chain,

with K-Means algorithm as the clustering algorithm.

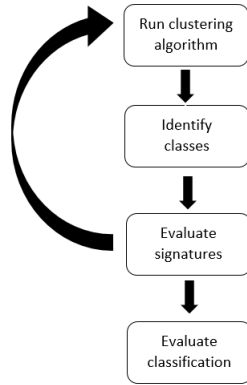


Fig. 2: Unsupervised learning algorithm

II. METHODOLOGY

Each pixel of hyper-spectral image contains a spectral signature that depends upon the reflectance of the specific material. If we can retrieve that spectral information we can tell the specific characteristic of material in that pixel. In this paper we worked on three data sets, information of them are given below.

A. Sample data

Three data sets that are used for analysis purpose are Aviris Cuprite, Aviris Muffet and RosiS Pavia. Cuprite and Muffet are retrieved from NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) that has a spatial resolution of 20 meters when flown at its typical altitude of 20 km, but a 4-meter resolution when flown at an altitude of 4 km. The number of spectral band of cuprite and Muffet is 224 with wavelength ranged from 400nm-2500nm. Pavia university scene has been retrieved by the ROSIS sensor during a flight campaign over Pavia, Italy. The number of spectral bands of Pavia university is 113 with wavelength ranged from 430nm to 860nm, pixels are 610*610, but some of the samples in this image contain no information and have to be discarded before the analysis. The geometric resolution is 1.3 meters. As hyper-spectral image consists of number of spectral bands and human eye can only detect 3 bands (RGB), hence its display for the human eye to visualize is an important task [17]. An efficient band reduction algorithm is required to visualize hyper-spectral image[18]. In figure 3 data sets has been shown.

B. K-Means Clustering algorithm

The K-means algorithm is a simple method for calculating the mean of a group with k number of sets. The algorithm is designed to reduce cluster inconsistency [16][19][20] by organizing the hyper-spectral data in such a way that pixels that contains similar spectral characteristics are clustered together[20]. K-means introduces centroids c_k among the k number of clusters and keeps on calculating the distance of each cluster from it using euclidean distance and segregating

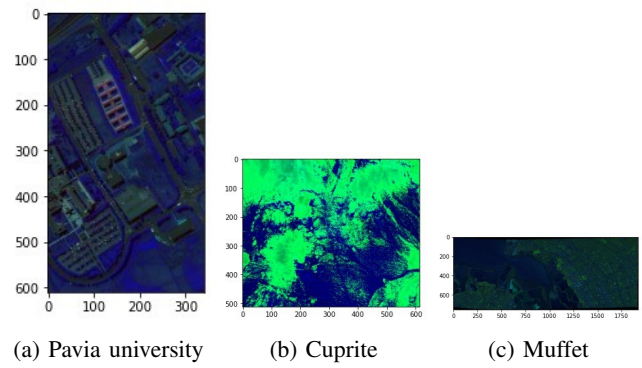


Fig. 3: Data sets

these cluster on the basis of minimum distance m from it. The algorithm keeps on calculating the distance until there is no change in distance m. Hence new set of groups are created with similar properties. Euclidean distance of centroid to each cluster can be found by using following equation.

$$J = \sum_{j=1}^k \sum_{i=1}^n ||x_i^j - c_j||^2 \quad (1)$$

$$\sum_{j=1}^k = \text{sum of clusters from 1 to k}$$

$$\sum_{i=1}^n = \text{sum of data points from 1 to n}$$

where, n = number of data points.

$||x_i^j - c_j||^2$ = Euclidean distance between data point and centroid

This equation will calculate Euclidean distance between data points and centroid for each cluster.

1) *Mean initialisation*: Mean initialisation is the major step in clustering algorithm in order to determine the accuracy of classification. Forgy and Random Partition methods are commonly used mean initializing methods. Forgy partition method randomly chooses value of k i.e k number of observations. Then this value of k initialise as means. While random Partition method initialises by assigning cluster to each observation. The Forgy method is a preferable method of mean initialization.

III. RESULTS AND DISCUSSION

K-means clustering algorithm is applied on Pavia University, Cuprite and Muffet to classify all these data sets. Algorithm divides image pixels into group on the basis of the similarities between their spectral reflectance values. In such way regions that have similar reflectance spectrum can be categorized as a group hence vegetation, rocks, water is differentiated. For k-mean algorithm predicting number of cluster is an important task. Value of k used for Pavia university is 9 (k=9) and for Cuprite and Muffet is 8 (k=8). Algorithm runs 30 times for PAVIA and Cuprite data set for 5

times for Muffet dataset, results can be seen in figure 4. Each cluster is labeled with different color as shown in figure 5. The process time is slow because of high computation. Cuprite data set took larger computational time as data set is complicated as compared to other two data sets. K-means clustering algorithm is implemented on all 3 data sets using Python.

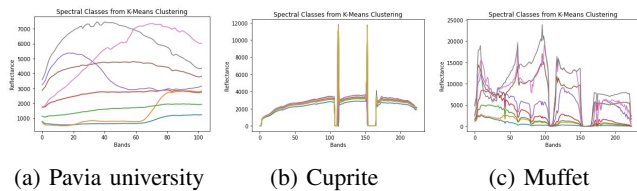


Fig. 4: Spectral classes from k mean clustering

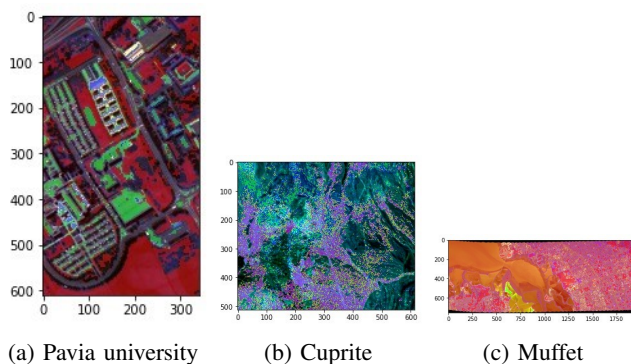


Fig. 5: Color labeling

IV. CONCLUSION

In satellite remote inspection, Hyperspectral imaging is a promising technique for evaluation of different materials. For different material detection like vegetation, water, soil etc, we employed K-means clustering. This method happened to give great results in detecting different regions through clustering. It is simple to implement, faster in terms of computing, and ensures convergence. One of the technique's drawbacks is that it requires the user to manually select the cluster number, which in this case is "k."

V. FUTURE WORK

In future implementation of different classification and clustering techniques can be used and their results can be compared. This will eventually give us best clustering or classification algorithm that could take lesser time. Like ISODTA is also an unsupervised learning technique but gives better and efficient results than K mean clustering[20].

REFERENCES

- [1] F. Masood, I. H. Qazi and K. Khurram "Saliency Based Visualization of Hyperspectral Satellite Images using Hierarchical Fusion", in Journal of Applied Remote Sensing, vol. 12, 2018, doi: 10.1117/1.JRS.12.046011.
- [2] Editor(s): E.P. Xanthoula, M.Dimitrios, B.Dionysis, "Intelligent Data Mining and Fusion Systems in Agriculture", Academic Press, pp.103-173, 2020, ISBN: 9780128143919.

- [3] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid and A. Abbas, "Modern Trends in Hyperspectral Image Analysis: A Review", in IEEE Access, vol. 6, pp. 14118-14129, 2018, doi: 10.1109/ACCESS.2018.2812999.
- [4] F.Fei, "Hyperspectral imaging in medical applications", Data Handling in Science and Technology, Elsevier, Vol. 32, pp. 523-565, 2020, ISSN 0922-3487, ISBN 9780444639776.
- [5] M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," Water SA, vol. 33, no. 2, pp. 145-151, 2007.
- [6] E. Adam, O. Mutanga, and D. Rugege, "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review," Wetlands Ecol. Manage., vol. 18, no. 3, pp. 281-296, 2010.
- [7] C. Fischer and I. Kakoulli, "Multispectral and hyperspectral imaging technologies in conservation: Current research and potential applications," Stud. Conservation, vol. 51, pp. 3-16, Dec. 2006.
- [8] H. Liang, "Advances in multispectral and hyperspectral imaging for archaeology and art conservation," Appl. Phys. A, vol. 106, no. 2, pp. 309-323, 2012.
- [9] H. Akbari, Y. Kosugi, K. Kojima and N. Tanaka, "Blood vessel detection and artery-vein differentiation using hyperspectral imaging," 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 1461-1464, doi: 10.1109/IEMBS.2009.5332920.
- [10] J. Freeman, F. Downs, L. Marcucci, E. N. Lewis, B. Blume and J. Rish, "Multispectral and hyperspectral imaging: applications for medical and surgical diagnostics," Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No.97CH36136), 1997, pp. 700-701 vol.2, doi: 10.1109/IEMBS.1997.757727.
- [11] J. Kuula et al., "Using VIS/NIR and IR spectral cameras for detecting and separating crime scene details," Proc. SPIE, vol. 8359, p. 83590P, Jun. 2012.
- [12] R. L. Schuler, P. E. Kish, and C. A. Plese, "Preliminary observations on the ability of hyperspectral imaging to provide detection and visualization of bloodstain patterns on black fabrics," J. Forensic Sci., vol. 57, no. 6, pp. 1562-1569, 2012.
- [13] R. Padoan, T. A. Steemers, M. Klein, B. Aalderink, and G. De Bruin, "Quantitative hyperspectral imaging of historical documents: Technique and applications," in Proc. ART, 2008, pp. 25-30.
- [14] G. J. Edelman, E. Gaston, T. G. van Leeuwen, P. J. Cullen, and M. C. G. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," Forensic Sci. Int., vol. 223, nos. 1-3, pp. 28-39, 2012.
- [15] I.A.Huqqani, K.Khurshid, "Comparative Study of Supervised Classification of Urban Area Hyperspectral Satellite Imagery," in Journal of Space Technology, vol. 4, pp. 7-15, 2014.
- [16] Available: <http://disi.unitn.it/seminars/281>. [Accessed 22 September 2013].
- [17] H. A. Khan, M. M. Khan, K. Khurshid and J. Chanussot, "Saliency based visualization of hyper-spectral images," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 1096-1099, doi: 10.1109/IGARSS.2015.7325961.
- [18] S. Sattar, H. Ahmad Khan and K. Khurshid, "Optimized class-separability in hyperspectral images," 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 2711-2714, doi: 10.1109/IGARSS.2016.7729700.
- [19] Tan, "The K-Means algorithm," December 2006. [Online]. Available: <http://www.cs.uvm.edu/~xwu/kdd/Slides/Kmeans-ICDM06.pdf>. 2013.
- [20] N. Acito, G. Corsini and M. Diani, "An unsupervised algorithm for hyperspectral image segmentation based on the Gaussian mixture model," IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477), 2003, pp. 3745-3747 vol.6, doi: 10.1109/IGARSS.2003.1295256.