

# CLUSTERING HYPERSPECTRAL DATA

Arwa Alturki and Ouiem Bchir

Department of Computer Science,  
King Saud University, Riyadh, Saudi Arabia

## **ABSTRACT**

*Spectroscopy or hyperspectral imaging consists in the acquisition, analysis, and extraction of the spectral information measured on a specific region or object using an airborne or satellite device. Hyperspectral imaging has become an active field of research recently. One way of analysing such data is through clustering. However, due to the high dimensionality of the data and the small distance between the different material signatures, clustering such a data is a challenging task. In this paper, we empirically compared five clustering techniques in different hyperspectral data sets. The considered clustering techniques are K-means, K-medoids, fuzzy C-means, hierarchical, and density-based spatial clustering of applications with noise. Four data sets are used to achieve this purpose which is Botswana, Kennedy space centre, Pavia, and Pavia University. Beside the accuracy, we adopted four more similarity measures: Rand statistics, Jaccard coefficient, Fowlkes-Mallows index, and Hubert index. According to accuracy, we found that fuzzy C-means clustering is doing better on Botswana and Pavia data sets, K-means and K-medoids are giving better results on Kennedy space centre data set, and for Pavia University the hierarchical clustering is better.*

## **KEYWORDS**

*Image Processing, Hyperspectral Imaging, Imaging Spectroscopy, Clustering, FCM, K-means, K-medoids, hierarchical, DBSCAN*

## **1. INTRODUCTION**

Hyperspectral imagery (HSI) is a main tool of remote sensing applications. As a signal is transmitted, reflected, and scattered again by interacting with the various components of the atmosphere and surface; the reflectance spectra analysis may allow recognition or quantification of the materials. It can be used in several applications such as environmental, geographic, and military applications [1, 2].

HSI is a three-dimensional data cube that containing the values of the radiation that has been collected over an area in a wide range of wavelengths. It is perceived as a high resolution image with hundreds of wavelength (spectral) bands[1, 3].

HSI provides more accurate and detailed information than normal colored image. When HSI captures hundreds or thousands of spectral channels, the normal colored image captures only three: red, blue, and green [3]. In fact, each pixel in the HSI is a spectrum with continuous values to precisely represent an object in a natural scene. Thus, each pixel in a hyperspectral image is a high dimensional vector that represents the radiance of a specific material. It can be considered as its signature.

Clustering is a powerful machine learning tools that allows analysing the content of HSI. Although the signature of each material is unique, two adjacent materials can have very similar signatures. Due to the characteristics of the hyperspectral data, clustering this data is challenging. In this work, we aim to compare empirically several existing clustering techniques in four hyperspectral data sets.

## 2. CLUSTERING TECHNIQUES

Number of clustering techniques have been applied on HIS [4]. Depending on the spectral content information, clustering methods such as k-means[5, 6], K-medoids [7, 8], fuzzy C-Means (FCM) [9, 10], hierarchical [11, 12], and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [13] are highly exploited [14]. To further enhance the clustering performance, the spatial context information is incorporated with the spectral information to enhance the clustering performance [15]. In the following we describe the considered clustering algorithms in some details.

### 2.1. K-means Clustering

K-means is one of the simplest and well known clustering techniques that aims to partition  $N$  observations into  $K$  clusters [5]. The algorithm starts by randomly choosing  $K$  centroids, where one centre for each cluster. Let  $X = \{x_1, x_2, \dots, x_N\}$  be the set of observations that we want to clusters, and  $C = \{c_1, c_2, \dots, c_K\}$  be the initial centroids set, the following two steps are repeatedly visited [6]:

- Assignment, and this can be accomplished by: (1) calculate the mean (distance) from the observation and each cluster centre, (2) assign the observation to the cluster with minimum mean among all clusters. The algorithm tries to minimize the mean (total squared error) depending on Equation (1):

$$I = \sum_{i=1}^K \sum_{j=1}^N \|x_j - c_i\|^2 \quad (1)$$

- Update, by recalculating the clustercentres according to Equation (2):

$$c_i = \left(\frac{1}{m_i}\right) \sum_{j=1}^{m_i} x_j \quad (2)$$

Where  $m_i$  is the number of observations in the  $i^{\text{th}}$  cluster. The algorithm stops when the results converge and the observations are no longer reassigned to other clusters. K-means clustering is a fast one with time consumption equals to  $\theta(NKdt)$ , where  $d$  is the dimension of each observation and  $t$  is the number of iteration needed to converge [5]. It is important to note that, K-means gives the best results when the data is well separated.

### 2.2. K-medoids Clustering

A clustering algorithm that is very similar to K-means, but with small differences. Unlike K-means, K-medoids tries to minimize the total dissimilarities between the observations in a cluster and its medoid, which make K-medoids more robust to noise [7]. A medoid of cluster is an observation that have the minimal dissimilarity average to all other observations [8]. Similar to K-means, the initial medoids are randomly selected then the assignment and updating step are processed repeatedly.  $X = \{x_1, x_2, \dots, x_N\}$  and  $M = \{m_1, m_2, \dots, m_K\}$  be the set of observations and

medoids respectively, the difference between K-means and K-medoids appears in the update step as the following:

- Swap between medoid  $m_i$  and observation  $x_j$  in the  $i^{\text{th}}$  cluster.
- Recalculate the dissimilarity average between  $x_j$  and other observation in the cluster.
- If the resulted dissimilarity is larger than before swapping, return to the previous medoid.

### 2.3. Fuzzy C-Means (FCM)

Unlike other clustering techniques, an observation can belong to more than one cluster with FCM clustering [9, 10]. At the end of the clustering process, each observation is given a degree of membership to each cluster, where the degree is computed based on specific similarity measure [16]. Let  $X = \{x_1, x_2, \dots, x_N\}$  be the observations that we want to be partitioned into  $K$  cluster and  $m$  is the clusters overlapping scaler, the algorithm assigns a random degree for each observation with each cluster. After that, the following steps are repeatedly processed:

- For each cluster  $i$ , find its centroid by calculating the mean of all observations weighted by the degree of membership  $U$  to the  $i^{\text{th}}$  cluster. This can be summarized in Equation (3):

$$c_i = \frac{\sum_{j=1}^N u_{ji}^m \cdot x_j}{\sum_{j=1}^N u_{ji}^m} \quad (3)$$

- Re-compute the degree of membership for all observations of all clusters as in Equation (4):

$$u_{ji} = \frac{1}{\sum_{k=1}^K \left( \frac{\|x_j - c_i\|}{\|x_j - c_k\|} \right)^{\frac{2}{m-1}}} \quad (4)$$

The above two steps are repeated until the algorithm is converging, and the change of membership degrees  $U$  in two consecutive iterations is no more than a predefined threshold. FCM clustering is a good choice when the data set is overlapped and not well separated.

### 2.4. Hierarchical Clustering

A clustering algorithm that builds a hierarchy of clusters based on two approaches: agglomerative and divisive [11, 12]. Agglomerative approach (also known as bottom-up) treats each observation as a cluster and recursively merge (agglomerate) pairs of clusters until all of them are merged into one cluster. On the other hand, divisive hierarchical clustering (also known as top-down) starts by putting all observations in one cluster and recursively split (divide) the clusters until each observation represent one cluster. The hierarchical clustering results usually presented in a dendrogram graph.

In order to determine which clusters to merge (or split); we need to measure the dissimilarity between observations by calculating the distance between them. There are several metrics of distance, which are: Euclidean, squared Euclidean, Manhattan, maximum, and Mahalanobis. Then, the distance metric itself is used by the linkage criteria to measure the dissimilarity between two clusters. Here are some of the well-known linkage criteria: maximum (complete) linkage, minimum (single) linkage, and average (mean) linkage.

## 2.5. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN discovers the clusters (of arbitrary shapes) and noise in spatial database (points) based on the density [13]. It requires determining two parameters, which are: the maximum distance between two points to be neighbours (*Eps*) and the minimum number of points to form a cluster (*MinPts*). Then, DBSCAN starts with arbitrary un-visited point  $P$  by retrieving its neighbours, i.e. points where the distance between  $P$  and them is less than or equals to *Eps*. Then, if the number of neighbour is equal to or larger than *MinPts*, a new cluster is created; otherwise  $P$  labelled as a noise point. The process repeated until all points are visited.

## 3. EXPERIMENTS

In this paper, we investigated the behaviour of clustering techniques that explained in Section 1 on hyperspectral data (see Section 2). The followed approach starts by applying two pre-processing steps which are (1) noise reduction by removing zero labelled data, and (2) normalization to reduce the variance between data and exclude the outliers. After that, different clustering techniques are applied on the data. Finally, for each clustering result, five performance measures are computed and will be explained in Section 3.2. The flowchart in Figure 1 summarize the proposed approach on this paper.



Figure 1. Proposed approach

### 3.1. Data Description

Four different data sets are used for experiments: Botswana, KSC, Pavia, and PaviaU. Here are brief description and information about each data set:

#### 3.1.1. Botswana

Obtained by NASA-EO1 satellite over Okavango Delta, Botswana in 2001-2004. The data is acquired by EO1 Hyperion sensor at 30m pixel resolution from an altitude of 7.7km in 242 bands of 400-2500nm wavelength and 10nm width [17]. Feature selection is conducted in order to remove the noise of water absorption to result in 145 bands. The data is used for analysis purposes such as classification, where there are 14 classes representing various land cover types in seasonal and occasional swamps and drier woodlands in Delta.

#### 3.1.2. Kennedy Space Centre (KSC)

Obtained by NASA AVIRIS over Kennedy Space Centre, Florida in 1996. The data is acquired at 18m pixel resolution from an altitude of 20km in 224 bands of 400-2500nm wavelength and 10nm width [17]. Feature selection is applied to remove the water absorption and SNR bands and the resulted data have 176 bands. For classification purposes, there are 13 classes representing various land cover types that defined and usually occurred on the environment site.

#### 3.1.3. Pavia Centre and University

Obtained by the ROSIS sensor during a flight campaign over Pavia, Italy in 2002. The data is acquired at 1.3m per pixel spectral resolution with 102 bands and 103 bands for Pavia centre and Pavia university (PaviaU) respectively [18]. Pavia centre data set have 9 ground truth classes which are: water, trees, asphalt, self-blocking bricks, tiles, bitumen, shadows, meadows, and bare soil. Pavia university also have the same 9 classes but with gravel instead of tiles and painted metal sheets instead of self-blocking bricks.

### 3.2. Experiments Parameters

The following settings have been adopted in our experiments:

- Using cosine distance in K-means, K-medoids, and hierarchical clustering.
- Average function is used as a linkage strategy in hierarchical clustering.
- FCM is applied with an overlapping scaler equals to 1.1 to make it fuzzier.
- Due to huge memory requirements of hierarchical and DBSCAN clustering; data selection is applied. The process is achieved by selecting  $p\%$  of each ground-truth class in the data set to be  $p\%$  of the whole data set. The chosen percentages are **25%** and **10%** for hierarchical and DBSCAN clustering respectively.

In order to assess the performance of the proposed approach, five performance measurements are computed. The first one is clustering accuracy that calculated by assigning label for each cluster based on its elements true labels majority. Then the weighted accuracy of each cluster is computed as the clustering correct rate multiplied by the number of elements in each cluster. The final clustering accuracy is computed as the average of all clusters weighted accuracies.

To compute the other measurements, four parameters should be defined for each cluster  $C_i$ , which are: (1) number of correctly clustered elements (TP), (2) number of correctly rejected elements (TN), (3) number of incorrectly clustered elements (FP), and (4) number of incorrectly rejected data elements (FN).

The other four measurements are: Rand statistics ( $Q_{Rand}$ ), Jaccard coefficient ( $Q_{Jaccard}$ ), Fowlkes-Mallows index ( $Q_{FM}$ ), and Hubert index ( $Q_{Hubert}$ ) [19]. These measurements are defined in Equation (5), (6), (7) and (8) :

$$Q_{Rand} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Q_{Jaccard} = \frac{TP}{TP + FP + FN} \quad (6)$$

$$Q_{FM} = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}} \quad (7)$$

$$Q_{Hubert} = \frac{TP (TP + TN + FP + FN) - ((TP + FP)(TP + FN))}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

### 3.3. Results and Analysis

For all considered data sets, FCM [9, 10] is giving better clustering results with respect to accuracy measure. However, since the data sets are unbalanced, we investigate the other considered measures.

For Botswana data set[17] (Figure 2), it is confirmed that FCM is outperforming the other considered clustering algorithms with respect to all performance measures. Similarly, for Pavia data set[18] (Figure 4), FCM is either the best clustering results or slightly behind hierarchical clustering algorithm [11, 12]. For KSC data set[17] (Figure 3), K-means [5, 6] and K-medoids [7, 8] are giving the best results. Concerning PaviaU data set[18] (Figure 5), FCM is giving better

results according to the accuracy and  $Q_{\text{Rand}}$  measures. However, according to  $Q_{\text{Jaccard}}$ ,  $Q_{\text{FM}}$ , and  $Q_{\text{Hubert}}$ , the hierarchical clustering is giving better results.

In summary, we noticed that FCM is doing better than the other considered clustering algorithms. However, the accuracy is between 0.5 and 0.6. Moreover, FCM result depends on the value of the overlapping scaler  $m$  that needs to be tuned in order to find the optimal result.

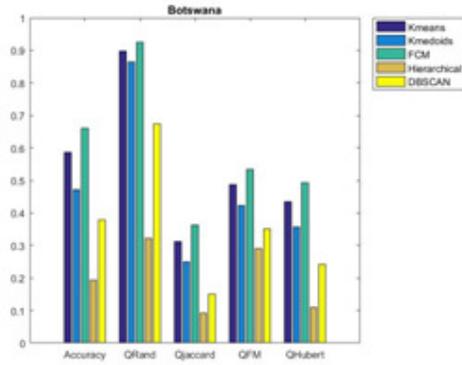


Figure 2. Botswana data set clustering results

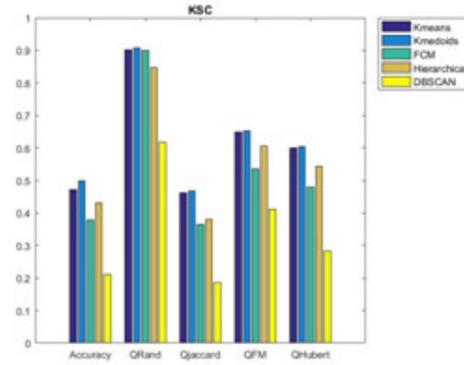


Figure 3. KSC data set clustering results

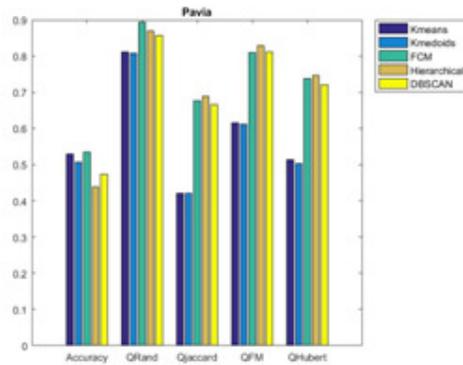


Figure 4. Pavia data set clustering results

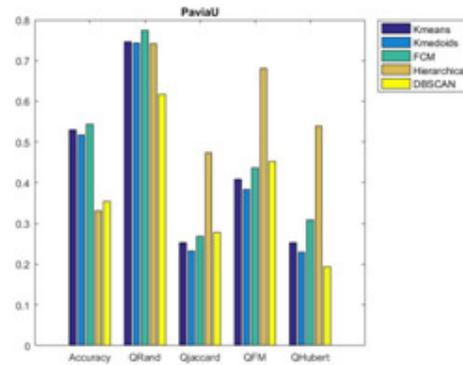


Figure 5. PaviaU data set clustering results

## 4. CONCLUSION

Recently, hyperspectral imaging (HSI) attracted researchers' interests. Classification and clustering algorithms are conducted to analyse and study their results on HSI. In this paper we investigated the behaviour of known clustering techniques on HSI, which are: k-means[5, 6], K-medoids [7, 8], fuzzy C-Means (FCM)[9, 10], hierarchical[11, 12], and DBSCAN[13]. Four different HSI data sets are chosen in our experiments, which are: Botswana[17], KSC[17], Pavia[18], and PaviaU[18]. We found that FCM is doing better on Botswana and Pavia data sets, K-means and K-medoids are giving better results on KSC data set, and for PaviaU the hierarchical clustering is better.

In order to enhance further the results, we plan to use fusion techniques on the considered clustering algorithms. We will also consider semi-supervised clustering techniques [19, 20] where some of the known information is used to guide the clustering processes to a better partition. In addition, we intend to conduct advanced noise reduction process by applying some filtering techniques[21, 22]. As an example of filtering techniques that we may consider are: maximum likelihood adaptive filter for partially observed Boolean dynamical systems[23], optimal state

estimation for Boolean dynamical systems using a Boolean Kalman smoother[24], and automated lane detection by k-means clustering [25].

## REFERENCES

- [1] H. Grahn and P. Geladi, Techniques and applications of hyperspectral image analysis. John Wiley & Sons, 2007.
- [2] G. Vane and A. F. Goetz, "Terrestrial imaging spectroscopy," *Remote Sensing of Environment*, vol. 24, no. 1, pp. 1-29, 1988.
- [3] C.-I. Chang, *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007.
- [4] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003.
- [5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, vol. 1, no. 14, pp. 281-297: Oakland, CA, USA.
- [6] 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.
- [7] Y. Dodge, *Statistical data analysis based on the L1-norm and related methods*. Birkhäuser, 2012.
- [8] A. Struyf, M. Hubert, and P. Rousseeuw, "Clustering in an object-oriented environment," *Journal of Statistical Software*, vol. 1, no. 4, pp. 1-30, 1997.
- [9] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," 1973.
- [10] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
- [11] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval* (no. 1). Cambridge university press Cambridge, 2008.
- [12] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241-254, 1967.
- [13] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, 1996, vol. 96, no. 34, pp. 226-231.
- [14] S. Li, B. Zhang, A. Li, X. Jia, L. Gao, and M. Peng, "Hyperspectral imagery clustering with neighborhood constraints," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 588-592, 2013.
- [15] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote sensing of environment*, vol. 113, pp. S110-S122, 2009.
- [16] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern recognition*, vol. 40, no. 3, pp. 825-838, 2007.
- [17] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to knowledge transfer for hyperspectral data analysis," in *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on, 2006*, pp. 541-544: IEEE.

- [18] Y. Bazi, L. Bruzzone, and F. Melgani, "An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 874-887, 2005.
- [19] O. Bchir, "Unsupervised and Semi-supervised Clustering with Learnable Cluster Dependent Kernels," University of Louisville, 2011.
- [20] E. Bair, "Semi-supervised clustering methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 349-361, 2013.
- [21] R. Chandel and G. Gupta, "Image Filtering Algorithms and Techniques: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 10, 2013.
- [22] S. Tania and R. Rowaida, "A comparative study of various image filtering techniques for removing various noisy pixels in aerial image," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9, no. 3, pp. 113-124, 2016.
- [23] M. Imani and U. M. Braga-Neto, "Maximum-Likelihood Adaptive Filter for Partially Observed Boolean Dynamical Systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 359-371, 2017.
- [24] M. Imani and U. Braga-Neto, "Optimal state estimation for boolean dynamical systems using a boolean kalman smoother," in *Signal and Information Processing (GlobalSIP), 2015 IEEE Global Conference on*, 2015, pp. 972-976: IEEE.
- [25] A. Gupta and P. S. Merchant, "Automated Lane Detection by K-means Clustering: A Machine Learning Approach," *Electronic Imaging*, vol. 2016, no. 14, pp. 1-6, 2016.