

RECURSIVE HIERARCHICAL CLUSTERING FOR HYPERSPECTRAL IMAGES

S. May

CNES DCT/SI
18, Av. Edouard Belin,
31401 Toulouse Cedex 9, France - stephane.may@cnes.fr

KEY WORDS: clustering, unsupervised, hierarchical, recursive, spectral angle, hyperspectral, classification, K-means, fuzzy C-means, self organizing map, Kohonen map

ABSTRACT:

Partition based clustering techniques are widely used in data mining and also to analyze hyperspectral images. Unsupervised clustering only depends on data, without any external knowledge. It creates a complete partition of the image with many classes. And so, sparse labeled samples may be used to label each cluster, and so simplify the supervised step. Each clustering algorithm has its own advantages, drawbacks (initialization, training complexity). We propose in this paper to use a recursive hierarchical clustering based on standard clustering strategies such as K-Means or Fuzzy-C-Means. The recursive hierarchical approach reduces the algorithm complexity, in order to process large amount of input pixels, and also to produce a clustering with a high number of clusters. Moreover, in hyperspectral images, a classical question is related to the high dimensionality and also to the distance that shall be used. Classical clustering algorithms usually use the Euclidean distance to compute distance between samples and centroids. We propose to implement the spectral angle distance instead and evaluate its performance. It better fits the pixel spectrums and is less sensitive to illumination change or spectrum variability inside a semantic class. Different scenes are processed with this method in order to demonstrate its potential.

1. INTRODUCTION

1.1 The problem of labeling

Hyperspectral images give us access to a wide range on information contained in the different spectral bands. The classification process suffers from several drawbacks : it reduces the wealth of the information (only a limited number of classes for example) and also requires of full set of learning samples (availability of ground truth). As we can see in many scientific papers, the availability of precise and validated ground truth is not always easy (Lange, 2018).

Clustering methods are able to exploit to whole amount of data into the hyperspectral cube. As they are unsupervised, they may extract some new classes into the data. The unsupervised clustering produces a list of clusters or centroids, a classification map, and a distance matrix to these centroids for each pixel.

In some cases of learning tasks, it is difficult or costly to build a ground truth especially for classification activities. Furthermore, creating a dedicated map for one or few classes is accessible, but creating a full ground truth over a geographical area is much more difficult. Clustering approach may be used to learn the structure of data, and using few labels to classify elements should be possible.

1.2 The problem of high dimension

In this paper, we propose to compare several clustering techniques such as K-Means, Fuzzy-C-Means, and Self Organizing Map (Kohonen Map). They are all using the Euclidean distance to compute the position updates of the centroids. One of our contribution consists in modifying the classical Euclidean distance or l_2 norm by a spectral angle and evaluate its impact. Indeed, in many analysis of high dimension hyperspectral vectors, the spectral angle is more robust to illumination change.

The Principal Component Analysis transform is a widely used method to reduce the data dimensionality. It may conduct to the loss of interesting information. For the main semantic classes, its use is recommended as it fastens the execution time.

1.3 About clustering algorithms

Among the different clustering techniques, K-Means is a widely used since it is fast, and quite robust (MacQueen, 1967). At each iteration, every sample is assigned to its nearest cluster. Then the new means (centroids) are computed with the assigned samples. The algorithm ends when the assignment no longer changes, and so the centroids do not move any more. This algorithm is quite simple but does not guarantee to find the optimum clusters. Given a set of observations $(X_1, X_2, X_3, \dots, X_N)$, the C sets of points are $S = \{S_1, S_2, \dots, S_C\}$. Distance between observations and clusters is computed with l_2 norm, which is :

$$D_{i,j} = \|X_i - S_j\| \quad (1)$$

Position is computed with the formula $S_j = 1/n_j \sum_{k \in S_j} X_k$

Its complexity is $O(N.D.C.I)$ with N number of samples, D dimensions of the data, C number of clusters and I number of iterations.

The fuzzy C-Means (FCM) is a method of clustering which allows one pixel sample to belong to several clusters (Dunn, 1973) (Bezdek, 1980). It is based on the minimization of the objective function :

$$J_m = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^m \|S_i - X_j\|^2$$

m is the fuzzyfication coefficient with $m \geq 1$

$$u_{ik} = \frac{1}{\sum_{j=1}^C \left(\frac{\|S_i - X_k\|}{\|S_j - X_k\|} \right)^{2/(m-1)}} \quad (2)$$

The new position of the updated centroid is computed with the

$$\text{formula : } S_i = \frac{\sum_{k=1}^N u_{ik}^m X_k}{\sum_{k=1}^N u_{ik}^m} \quad (3)$$

The Self Organizing Map differs from the K-Means algorithm (Kohonen, 1982). For each pixel, we assign a winning node with the distance $\mathcal{D}_{i,j} = \|X_i - S_j\|$ (4)

Considering an external topology describing the spatial organization of nodes, spatial neighbors of the winning node are also updated. Consequently, spatial neighbors nodes into the topology will become spectral neighbors.

The updated centroid is computed with the formula :

$$S_{k,t+1} = S_{k,t} + \beta_t * w_{dist} * (X_i - S_{k,t}) \quad (5)$$

The parameter β_t is the learning rate and w_{dist} is the function of distance into the topological representation of the network (may be simply equal to 1 if neuron is updated, or 0 if not).

There are many other interesting evolutions relative to clustering algorithms around K-Means (Xu, 2016) or (Nasser, 2006) for example. But our study is limited for the moment to the three algorithms presented in the paragraph.

1.4 About hierarchical or recursive approaches

To solve the problem of large dataset, many attempts have been done with hierarchical analysis or recursive analysis (Gowda, 2017) (De Silva, 2018) (Cardot, 2011). Divide and conquer strategy is applied : learn a K-Means clustering once, apply it on the whole dataset, and divide it in several sub-groups. Then apply a new K-Means recursively on each sub-groups till the partitions meet a predefined criterion : last level reached, min number of samples in the last group, statistical homogeneity of vectors, etc.

2. PROPOSED METHOD

First of all, considering the 3 clustering algorithms described previously, we propose to replace the Euclidean distance function by a spectral angle distance. It is defined as :

$$\mathcal{D}_{i,j} = \text{acos} \left(\frac{X_i \cdot S_j}{\|X_i\| \cdot \|S_j\|} \right) \quad (6)$$

The spectral angle (SA) distance replaces the l_2 norm into formulas (1) (2) and (4). The Spectral Angle Mapper algorithm is widely used for remote sensing, for classification purpose or to fit a database spectrum with pixel spectrums (May, 2013). The aim is to analyse its impact on the results.

The proposed hierarchical clustering is a top-down approach. It is defined by a number of levels L , and number of clusters C for each clustering. We consider here that the total number of clusters is not a question : the final numbers of clusters will be large relatively to the number of labeled classes that we are looking for. Many methods try to find an optimal partition of the input data.

A first clustering is applied on random samples (N pixels) inside the image. It produces C clusters, with C groups of image pixels. Then, at the second (hierarchical) stage, for each group of image pixels, we apply (recursively) the same clustering algorithm on random samples (N pixels), producing C clusters, and C groups of image pixels. At this level, we obtain C^2 clusters, C^2 groups of pixels. The same algorithm is applied recursively to each group of pixels, producing each time C clusters and C

groups of pixels. At the final level L , we obtain C^L clusters.

We perform $\sum_{k=0}^{L-1} C^k$ separate clusterings.

Each individual clustering algorithm may be a K-Means, Fuzzy-C-Means, Kohonen map or any other clustering approach.

The algorithm steps are listed below :

1. Start at first hierarchical level ($l = 1$)
2. Extract N samples X_n from group of pixels
3. Learn the C clusters $S_k \quad \forall k \in [1; C]$
4. For all pixels of the image, compute the distance to the nearest cluster, and assign the corresponding cluster id
5. Next hierarchical level ($l + 1$) : loop on C
6. For a given cluster id (let's consider i_0), extract N samples X_n from the data
7. Learn the C clusters $S_{i_0,k} \quad \forall k \in [1; C]$
8. For all pixels of the image assigned to i_0 , compute the distance to the nearest cluster, and assign the corresponding cluster ids
9. Stop condition (example : max number of hierarchical layer level L reached).
10. Else, recursive call to 5)

At each stage, the initial number of pixels is divided into C groups of pixels. So the complexity of the clustering is reduced for the lower levels. Each clustering algorithm conducts to add C nodes into a hierarchical tree of clusters.

For a target number of clusters C' , we define a couple of values (C, L) as parameters of the recursive hierarchical clustering, to build C^L clusters. Examples:

$$C = 2, L = 8 \Rightarrow C^L = 256$$

$$C = 10, L = 3 \Rightarrow C^L = 1000$$

Each learning and clustering achieved at a given level is done on its own group of pixels, so is completely independent from the other parallel learning.

Finally, each cluster must be linked with a labeled class. In this work, we consider that the number of available labeled samples is sparse. Each cluster is labeled related to its nearest class sample. If a class label has no representative cluster, it is linked to its nearest cluster.

The main advantages of this method are :

1. Scalability : as at each level, we only learn a reduced number of nodes, there is no need to take a very high number of samples. It is able to manage large amount of data
2. It produces a high number of clusters
3. Easier convergence : the number of clusters (C) is limited at each step so the convergence is easier
4. Recursive algorithm : the same method is applied on the son clusters
5. Parallelism : each pixel subgroup may be process independently

The main drawback is the loss of homogeneous distribution of clusters. Two clusters in two separate branches of the hierarchical tree may be spectrally close. The number of clusters is chosen voluntary high, and at the end, distance between spectrums in the image and clusters should be reduced.

3. EXPERIMENTS AND RESULTS

3.1 Pavia University

The ROSIS sensor acquired a scene over University of Pavia during a flight campaign. The image contains 103 spectral bands. It is a 610*610 pixels image. The geometric resolution is 1.3 meters. The image ground truths has 9 classes : gravel, painted metal sheet, trees, asphalt, self-blocking bricks, bitumen, shadows, meadows, bare soil.

A very limited set of manual labels have been done with the input image and the ground truth. 30 small squares or rectangles with size around 3x3 pixels have been edited. It means that each class is represented by ~3 patches. They are used for supervised classification (for comparison) and for the clusters labeling. We indicate that in the theoretical ground truth, some areas labeled as meadows contains meadows and bare soil. Also a large area with bare soils contains several classes.

The classification result is compared to a supervised Random Forest (RF) method (Breiman, 2001) : number of estimator = 40, max depth = 8, max features = 3.

The clustering algorithms K-Means, Fuzzy C-Means and Self Organizing Map are trained with 25 classes, 50k samples inside the image. For each method, an evaluation is done without Principal Component Analysis (PCA) and with a PCA with the 10 first components.

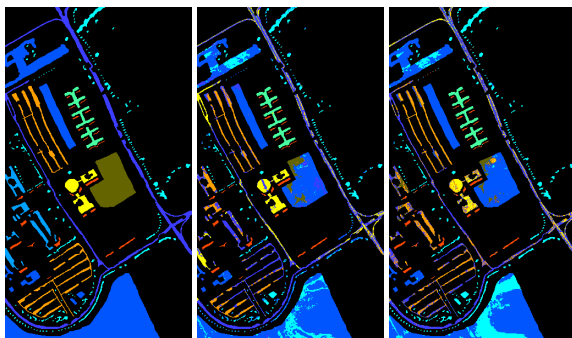


Figure 1. Classification map : 1: official ground truth 2: K-Means clustering with l_2 norm 3: K-Means with SA

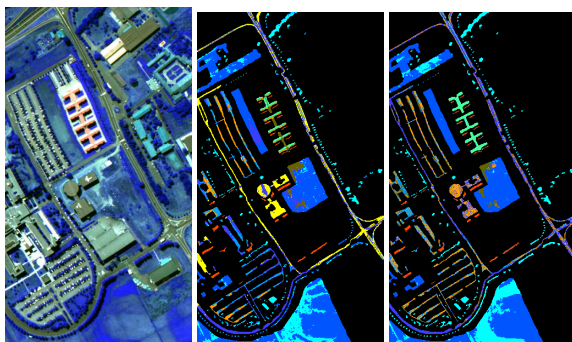


Figure 2. Classification map : 1: RGB composite 2: FCM with l_2 norm 3: FCM with SA

The Recursive Hierarchical Clustering is computed with the input image, with 2 clusters per node ($C=2$), a number of levels L varying between 3 to 8 (producing between 8 clusters to 256 clusters). At each node, a K-Means clustering method is

used with a maximum of 5000 samples to compute the C clusters. Comparison with classical clustering such as K-Means, Fuzzy C-Means, and the different distances, is useful to evaluate which base clustering shall be used into the Recursive Hierarchical Clustering.

For each cluster id, a maximum of 5000 samples are used to compute the C child clusters. Between each levels, all pixels of the image are assigned to the corresponding cluster ids. At each level, and for each cluster id, a new set with a maximum of 5000 samples is used to compute the recursive clustering. Consequently, with this method, if the number of levels is high, all pixels of the image are taken into account to compute the clusters.

Method	No PCA		PCA 10c	
	0.72			
Random Forest	SAM	Euc	SAM	Euc
K-Means	0.63	0.65	0.52	0.61
Fuzzy C-Means	0.46	0.58	0.43	0.63
SOM	0.63	0.62	0.39	0.65
RHC K-Means 8	0.38	0.56	0.41	0.58
RHC K-Means 16	0.53	0.65	0.45	0.53
RHC K-Means 32	0.55	0.65	0.48	0.60
RHC K-Means 64	0.63	0.60	0.46	0.65
RHC K-Means 128	0.57	0.65	0.45	0.61
RHC K-Means 256	0.64	0.65	0.52	0.64

Table 1. Overall Accuracy

As we don't try in this paper to use a large amount of labeled samples, the performance is lower than in other scientific papers. But we focus on classification with low number of labeled samples. The Random Forest is used as a reference method for supervised classification for the configuration with PCA and without PCA.

The clustered image reveals the real classes in the image, with their heterogeneity. As seen in table 1. The comparison of Spectral angle to Euclidean distance gives an advantages to the Euclidean distance. Legend of classes is given in figure 3. From a qualitative point of view, both clustering with spectral angle are able to separate the very green meadows. In the final classification map, K-Means (figure 1) and Fuzzy C-Means (figure 2) are able to separate several classes in the two large areas at the bottom right of the image. On the road, as there are several types of asphalt, the method would require more samples to label the clusters. It is important to note that the saturation of the overall accuracy at 0.65 is due to the problem of cluster labeling. Some labels are badly represented by the clusters due to the combined heterogeneity of classes into this image, and the low number of used patches. Study has not been done with a higher number of labeled samples.



Figure 3. Legend of Pavia University classes

3.2 Pavia Centre

The ROSIS sensor acquired a scene over Pavia centre during a flight campaign, in northern Italy. The number of spectral bands is 102. It is a 1096*1096 pixels image. The geometric resolution is 1.3 meters. The image ground truth differentiates 9 classes : water, trees, asphalt, self-blocking bricks, bitumen, tiles, shadows, meadows, bare soil. Pavia scenes were provided by Professor Paolo Gamba from the Telecommunications and Remote Sensing Laboratory, Pavia university (Italy).

For the evaluation of the method, 22 small squares or rectangles with size around 3x3 pixels have been edited. It means that each class is represented by ~2 or 3 patches. They are used for supervised classification (for comparison) and for the clusters labeling.

The Recursive Hierarchical Clustering is computed with the same varying parameters than for the previous scenario : 2 clusters per node (C=2), number of levels L varying between 3 to 8 (producing between 8 clusters to 256 clusters). At each node, a K-Means clustering method is used with a maximum of 5000 samples to compute the C clusters.

Method	No PCA		PCA 10c	
	0.83		0.92	
Random Forest	SAM	Euc	SAM	Euc
K-Means	0.82	0.87	0.88	0.88
Fuzzy C-Means	0.79	0.89	0.88	0.86
SOM	0.87	0.74	0.73	0.83
RHC K-Means 8	0.86	0.80	0.43	0.68
RHC K-Means 16	0.82	0.66	0.86	0.87
RHC K-Means 32	0.80	0.87	0.89	0.88
RHC K-Means 64	0.82	0.88	0.86	0.89
RHC K-Means 128	0.84	0.88	0.87	0.90
RHC K-Means 256	0.86	0.90	0.89	0.90

Table 2. Overall Accuracy

Several conclusions can be done with the table 2 of results. For this scenario, the use of PCA is an advantage on the clustering process. The Random Forest reference is better with the PCA, which is due to the small number of labeled samples. The K-Means in this case directly produces interesting results. It is able to produce comparable or better results than Fuzzy C-Means and Self Organizing Map. Performance and low computation complexity (with low value of clusters) explain why K-Means is proposed as base clustering algorithm for the Recursive Hierarchical Clustering method. Difference between Euclidean and Spectral angle is not very important but gives a small advantage to the Euclidean distance. When the number of clusters increases for the Recursive Hierarchical Clustering method, results become really close to the Random Forest reference. Beyond 32 clusters, the Overall Accuracy is very stable. Because each local clustering produces only C=2 clusters with N=5000 samples, the method is quite stable. The proposed method to label each cluster is quite simple (class label of the nearest samples), but seems to be robust if the number of clusters is high.

The figure 4 has been labeled with a very low number of samples. Some errors remains between water and shadows or dark roads. Errors may be reduced by simply adding new samples.

The table 3 presents all the learning times. Between each level, the Recursive Hierarchical Clustering method assign the pixels to the corresponding cluster ids, and extract new samples. So it processes a higher volume of pixels with file access, which

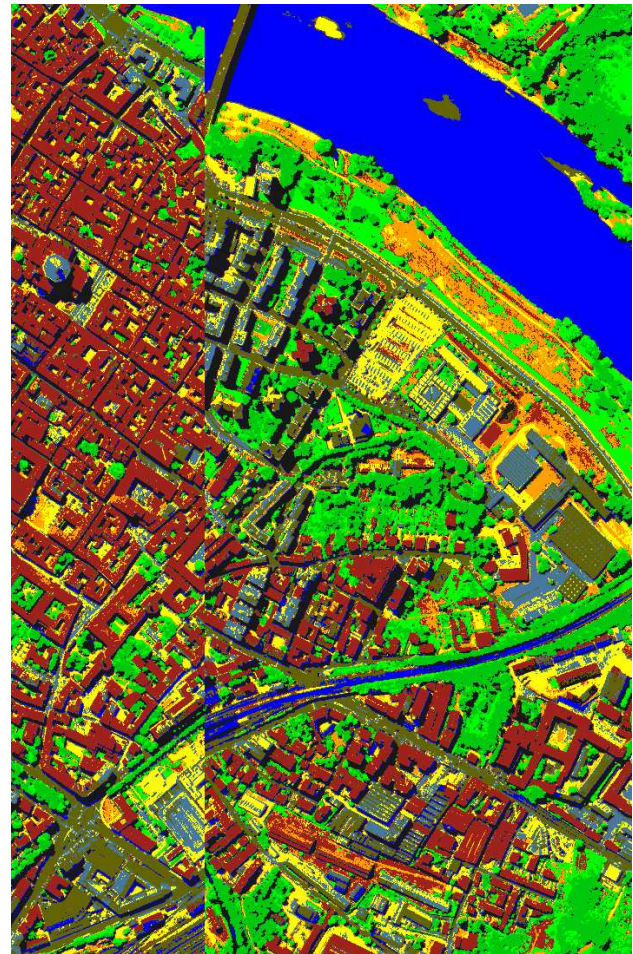


Figure 4. Classification map RHC K-Means SAM PCA

Method	No PCA		PCA 10c	
	0.8		0.8	
Random Forest	SAM	Euc	SAM	Euc
K-Means 25	33	50	6.3	4.8
Fuzzy C-Means 25	80	80	18.3	42.0
SOM 25	283	275	366.3	351.0
RHC K-Means 8	50	47	6.3	5.8
RHC K-Means 16	64	68	8.4	7.9
RHC K-Means 32	86	85	10.6	10.1
RHC K-Means 64	111	115	13.6	12.8
RHC K-Means 128	134	151	17.6	16.0
RHC K-Means 256	188	202	22.6	20.3

Table 3. Learning time (s)

Number of clusters	K-Means	RCH
8	-	5.8
16	3.0	7.9
25	4.8	-
32	6.0	10.1
64	11.2	12.8
128	21.0	16.0
256	41.2	20.3

Table 4. Learning time (s) and number of clusters, PCA 10c, Euclidean distance

explains the higher values of processing time. The table 4 illustrates the K-Means complexity which is $\mathcal{O}(N.D.C.I)$. As expected, the computation times of RHC is faster than K-Means for a high number of clusters. Furthermore, a high number of

clusters can be easily computed without initialization or convergence problems thanks to the recursive approach. For the last scenario with 256 clusters on Pavia, K-Means overall accuracy is equal to 0.91 which is very near the reference value obtained with the supervised classifier. But the use of K-Means with such a high number of clusters is not always possible when we consider large images and no dimension reduction.

Xu, J., H. X. N., 2016. Robust and sparse fuzzy k-means clustering. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence.

4. CONCLUSIONS AND PERSPECTIVES

This paper presents the current work related to unsupervised clustering algorithms applied to hyperspectral images. The clustered images directly outline some areas of interest. Quantitative results with the spectral angle relative to the Euclidean distance are lower. Nevertheless, the classification map got with the spectral angle method better reveals some particular regions.

The proposed method with Recursive Hierarchical Clustering method is able to outperform the standard clustering algorithms thanks to the high number of clusters with a reduced processing time. Unsupervised clustering techniques are promising for classification purpose but also for anomaly detections.

REFERENCES

- Bezdek, J., 1980. A convergence theorem for the fuzzy isodata clustering algorithms. *IEEE Transactions on Pattern Analysis Machine Intelligence*.
- Breiman, L., C. M., 2001. Random forests. *Machine Learning*, 45.
- Cardot, H., C. M., 2011. A fast and recursive algorithm for clustering large datasets with k-medians. *Computational Statistics Data Analysis*, 56(6).
- De Silva, P., F. W. F., 2018. Recursive Hierarchical Clustering Algorithm. *International Journal of Machine Learning and Computing*, 8(1).
- Dunn, J. C., 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3(3).
- Gowda, H., S. G. N. R., 2017. Semi-Supervised Text Categorization using Recursive K-means clustering. *Class Specific Feature Selection for Interval Valued Data Through Interval K-Means Clustering*.
- Kohonen, T., 1982. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 43(1).
- Lange, J., C. G. E. R., 2018. The influence of sampling methods on pixel-wise hyperspectral image classification with 3d convolutional neural networks. IGARSS.
- MacQueen, J. B., 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*.
- May, S., 2013. Classification of hyperspectral images with self organizing map. Whispers.
- Nasser, S., A. V., 2006. A Modified Fuzzy K-means Clustering using Expectation Maximization. *IEEE International Conference on Fuzzy Systems*.