Project 1
Image Segmentation via Clustering

- CAP6610 - Machine Learning - Spring 2018 -

Project Report

Group 9

Garg, Ashutosh Mahesh
Charles, Thelijjagoda Subodha
Chauhan, Karan
Singh, Devanshu

University of Florida

1 MEMBER ROLES

Table 1. Roles of group members

	Cluster Alg	Eval Alg	Cluster Test	Eval Test	Cluster Write	Eval Write	Cluster Edit	Eval Edit
Garg	1			1	1			1
Charles	1			1	1			1
Chauhan		1	1			1	1	
Singh		1	1			1	1	

2 EXPERIMENTS

2.1 RGB Experiments and Tabular Results

2.2 Hyperspectral Experiments

3 OBSERVATIONS

3.1 K-means advantages and disadvantages

The K-means clustering algorithm is by far the most popular and one of the simplest unsupervised learning algorithms.

Advantages:

- Simple, easy to implement
- Easy to interpret the clustering results;
- For small K values, K-Means is computationally faster than hierarchical clustering.

Disadvantages:

- Initial seeds have a strong impact on the final results.
- Uniform effect: often produce clusters with relatively uniform size even if the input data have different cluster size;
- Sensitive to outliers

References: [1–4]

3.2 FCM advantages and disadvantages

An extension of Hard C-Mean clustering method, it can be viewed as an advanced version of K-Means. Unlike K-Means which just calculates distance, FCM does a full inverse-distance weighting [5].

Advantages:

- Unsupervised
- Always converges
- High detection rate and lower false positive rates

Disadvantages:

- Long computational time
- Sensitivity to the initial guess (speed, local minima)
- Sensitivity to noise
- Low (or even no) membership degree for outliers (noisy points)

3.3 Spectral Clustering advantages and disadvantages

Most often than not, spectral clustering outperforms traditional clustering methods such as K-Means.

Advantages:

- Does not make strong assumptions on the statistics of the clusters
- Easy to implement.
- Good clustering results.
- Reasonably fast for sparse data sets of several thousand element

Disadvantages:

- Sensitive to choice of parameters
- Computationaly expensive for large datasets

References: [6–9]

3.4 Gaussian Mixture Models advantages and disadvantages

Gaussian Mixture Models (GMM) can be considered as a more generalized version of K-Means. Unlike K-Means, it is more flexible in terms of cluster covariance. It also allows mixed membership. i.e., a point can belong to a several clusters at different degrees.

Advantages:

- More flexibility clustering
- Allows mixed membership of points

Disadvantages:

- All eigenvectors of the local covariance matrix need to be extracted
- Can fail to work if the dimensionality of the problem is too high

References: [10]

3.5 Relative performance of algorithms

Table 2. Algorithm Ranking Table

Ranks	RGB Rankings	Hyperspectral Rankings
1		
2		
3		
4		

REFERENCES

[1] Hui Zhang, Jason E Fritts, and Sally A Goldman. Image segmentation evaluation: A survey of unsupervised methods. *computer vision and image understanding*, 110(2):260–280, 2008.

[2] King-Sun Fu and JK Mui. A survey on image segmentation. *Pattern recognition*, 13(1):3–16, 1981.

[3] Yu Jin Zhang. A survey on evaluation methods for image segmentation. *Pattern recognition*, 29(8):1335–1346, 1996.

[4] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[5] Abhinay B Angadi, Akshata B Angadi, and Karuna C Gull. International journal of advanced research in computer science and software engineering. *International Journal*, 3(6), 2013.

[6] Mathworks: File Exchange - Fast and efficient spectral clustering. https://www.mathworks.com/matlabcentral/fileexchange/34412-fast-and-efficient-spectral-clustering. [Online].

[7] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

[8] Pietro Perona and William Freeman. A factorization approach to grouping. In *European Conference on Computer Vision*, pages 655–670. Springer, 1998.

[9] Guy L Scott and Hugh Christopher Longuet-Higgins. Feature grouping by'relocalisation'of eigenvectors of the proximity matrix. In *BMVC*, pages 1–6, 1990.

[10] Stephen Adams and Peter A Beling. A survey of feature selection methods for gaussian mixture models and hidden markov models. *Artificial Intelligence Review*, pages 1–41, 2017.