

Comparative Study for Color Image Clustering based on Mahalanobis Distance Method

Fadhil Hanoon Abboud

Al Mustansiriyah University

Department of Computer Science, College of Education

Iraq

ABSTRACT

Color image has the potential to convey more information than monochrome or gray level images, RGB color model is used in many applications of image processing and image analysis such as Image Segmentation. The standard ways to deal with picture investigation and acknowledgment creatures by division of the picture into areas (articles) and figuring different properties and connections among these districts. Image segmentation algorithms, have been developed for extracting these regions. Important image segmentation is difficult to process due to the inherent noise and deterioration of the input indicators of the algorithm. The regions are not always specified, however, and it is often more fitting to consider them as fuzzy subjects in the image. In this work, an algorithm that is used to segment color images with clustering methods is defined. The lower conditions, results and results are illustrated. In the clustering algorithm, the results are compared using both Mahalanobis and Euclidean distances.

Key Words: Image Clustering, Fuzzy C- Mean, Mahalanobis, Euclidean Distances.

1. INTRODUCTION

For several image processing researchers, picture division was, is and will be a significant exploration subject. The reasons are straightforward and unlimited applications: most issues with PC vision and picture preparing include a division stage so as to distinguish antiquities or gap the picture into locales that can be viewed as homogeneous as indicated by what is used in biomedical fields, such as the identification of lung diseases, the automatic classification of white blood cells, the detection of cancerous cells. There are various applications for image segmentation. Clustering is the check in the feature space for distinct groups. These groups are supposed to have distinct structures and that can be obviously recognized. The assignment of grouping isolates the information into the quantity of parcels in the n-dimensional element space, which are volumes. Such parcels recognize a hard breaking point between the different gatherings and depend on the capacities used to show the appropriation of information. [1].

2. IMAGE SEGMENTATION

Picture isn't just an irregular arrangement of pixels for people; it is a significant course of action of locales and items. Various pictures actually exist: regular scenes, photos, and so forth. Regardless of the wide contrasts in these pictures, it isn't hard for people to decipher them. An amazing and robotized picture investigation would be welcome in our own photographic registries, given the broad WWW databases. Image segmentation is dividing of images into various regions, with certain properties in each region. It is a critical component of an image recognition process because segmentation errors will spread to even further processing steps, such as extraction and classification of features [2] [3]. The initial step of picture preparing and design acknowledgment is picture division. It is an essential and significant part of the cycle for picture investigation, one of the most requesting assignments in picture preparing, and decides the exactness of the last examination result. Picture division is the instrument by which a picture is isolated into different locales and that every district is homogeneous. Distinguishing locales of a picture that have comparative properties while isolating areas that are distinctive is one ordinary perspective on division. This is a difficult that is known as grouping. Traditional grouping relegates precisely one class to each protest, while the objects are assigned various degrees of membership to the various classes in fuzzy clustering [4]. Fuzzy clustering is an adaptive method to image segmentation using both cluster analysis and fuzzy logic principles. Methods of image segmentation can be classified as follows:

1. Thresholding histogram: accept that pictures comprise of areas with various scopes of dark (or shading), and partitions them into various pinnacles, each alluding to a district.
2. Edge-based methodologies: administrators, for example, Sobel and Laplacian, for instance, use edge recognition. The subsequent districts can't be associated, so it is critical to interface the edges.
3. Approaches dependent on area: in view of local picture information likeness. In this gathering, a portion of the more regularly utilized methodologies are: thresholding, bunching, district development, parting and consolidating.
4. Hybrid: think about the edges and areas of both.

The project is performed using a fuzzy Clustering Image Segmentation. It is based on the segmentation of color images using Mahalanobis distance. The Euclidean distance is also used to compare the consistency of the Mahalanobis-Euclidean distance segmentation.

2.1 Improved FCM Clustering

The clustering algorithm for Fuzzy C-means is based on the minimization of an objective function called functional C-means. In comparing the RGB value of each pixel to the value of the cluster core, C-Mean clustering follows the same rules as K-Means. The key difference is that it doles out an incentive somewhere in the range of 0 and 1 that determines "how much this pixel has a place with that bunch" for each group as opposed to settling on a hard choice on which group the pixel ought to have a place with. The Fuzzy principles determine that for all bunches, entirety of a pixel's participation esteem must be 1. The higher the estimation of the enrollment, the more probable the pixel is to have a place with that group. [4]. By minimizing an objective function shown in equation (1), the EFCM clustering is obtained: [5].

$$J = \sum_{i=1}^n \sum_{k=1}^c \mu_{ik}^m |p_i - v_k|^2 \dots\dots\dots(1)$$

Where:

- J is the objective function.
- n is the number of pixels in the entire image.
- c is the number of clusters.
- μ is the fuzzy membership value.
- m is a fuzziness factor (a value > 1).
- p_i is the i 'th pixel in E
- v_k is the centroid of the k 'th cluster
- $|p_i - v_k|$ is the Euclidean distance between p_i and v_k defined by equation (2):

$$|p_i - v_k| = \sqrt{\sum_{i=1}^n (p_i - v_k)^2} \dots\dots\dots(2)$$

The calculation of the centroid of the k th cluster is achieved using equation (3):

$$v_k = \frac{\sum_{i=1}^n \mu_{ik}^m p_i}{\sum_{i=1}^n \mu_{ik}^m} \dots\dots\dots(3)$$

The fuzzy membership is calculated using the original equation (4):

$$\mu_{ik} = \frac{1}{\sum_{i=1}^c \left(\frac{|p_i - v_k|}{|p_i - v_i|} \right)^{\frac{2}{m-1}}} \dots\dots\dots(4)$$

2.2 Improved FCM Algorithm Processes (proposed method)

Depending on new improved method In this proposed method, the number of cluster centers is calculated randomly and then the distance between the pixels and the cluster center will be calculated through the mahalanobis method of distributing pixels on clusters that are determined at the beginning of the open method in a way where the center will be determined for each cluster and then the distance between the center of the cluster and each pixel will be calculated to determine the pixel affiliation of any cluster for all RGB space to gather.

- process 1: Set the number of clusters, the fuzzy (constant > 1) parameter, and the stop condition. [6]
- process 2: Initialize the matrix of fuzzy partitions
- process 3: Set the $k = 0$ loop counter
- process 4: Determine the centroids of clusters randomly, calculate the objective value J
- process 5: Compute the membership values in the matrix for each pixel, for each cluster, using mahalanobis distance
- process 6: If the value of J is less than the stop condition between consecutive iterations, then stop; otherwise, set $k = k+1$ and go to step 4.
- process7: segmentation using the clustering

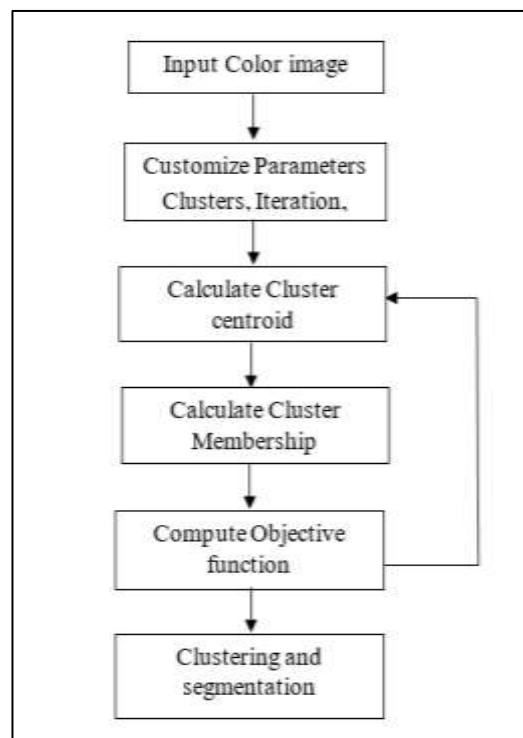


Figure 1: shows Block Diagram to Proposed Method Improve FCM

3. HOW TO MEASURE THE DISTANCE FROM MAHALANOBIS AND EUCLIDEAN

The Euclidean method used to measure the distance in classical C-Mean clustering. Both the distances from Mahalanobis and

Euclidean are clearly defined below in this work.

3.1 Mahalanobis Distance

A set of properties can be described as follows:

- Mahalanobis Distance is a very convenient way to evaluate the "similarities" of a "unknown" set of values: sample to a set of values determined from a "known" collection of samples.
- Superior to Euclidean distance as it takes into account the distribution of points (correlations)
- Traditionally, observations can be divided into various categories.
- It not only takes into account the average value, but also the variance and covariance of the measured variables.
- It compensates for interactions between variables (covariance).
- It is dimensionless.

The method used to compute Mahalanobis distance is shown below [7]:

$$Dt(x) = (x - Ci) * Inverse(S) * (x - Ci) \dots\dots\dots (5)$$

Where:

The X is a data point in the 3-D RGB space, Ci is the center of a cluster.

S is the covariance matrix of the data points in the 3-D RGB space Inverse(S) is the inverse of covariance matrix S.

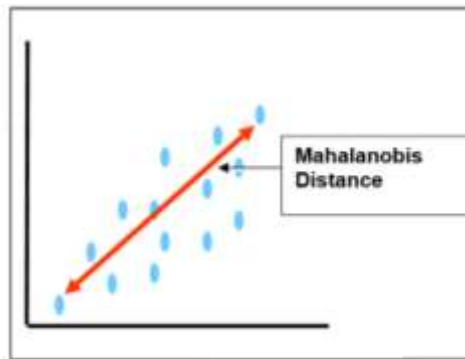


Figure 2: shows how the Mahalanobis distance is calculated as the variances of the data points in the RGB space.

3.2 Euclidean Distance

The straight-line distance between two pixels is the Euclidean distance

$$\text{Euclidean distance} = \sqrt{((x1 - x2)^2 + (y1 - y2)^2)} \dots\dots\dots (6)$$

where (x1,y1) & (x2,y2) are two pixel points or two data points.

In C-Mean clustering the distance has been calculated bases on equation 2:

$$|P_i - v_k| = \sqrt{\sum_{i=1}^n (P_i - v_k)^2}$$

The only difference between Mahalanobis and Euclidean distance is that Mahalanobis considers the Inverse of the covariance matrix of the set of data points in the 3-d space.

So,

$$\text{Mahalanobis distance} = (P_i - V_k) * INV (Cov (S)) * (P_i - V_k)' \dots\dots\dots (7)$$

$$\text{Euclidean distance} = (P_i - V_k) * (P_i - V_k)' \dots\dots\dots (8)$$

Here P_i is a data point and V_k is the center of a cluster.

S is a vector containing all the data points the 3-d color space.

5. RESULT AND DISCUSSION

Different color images were examined using the Mahalanobis distance when using proposed Improved FCMC method, as shown in the figures below (3-5) and compared with the results (segmented images) obtained by the FCMC using the Euclidean distance Different images were used containing a set of overlapping objects. The quality of work was measured by using a set of quality calculation criteria set out in spreadsheets below.

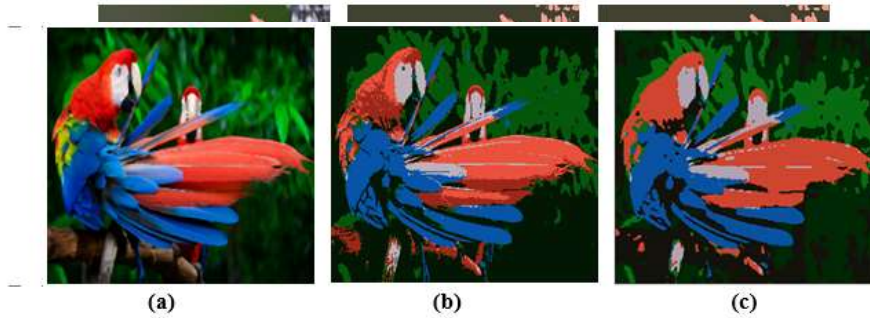


Figure 4: shown (a) Parrot Original Image (b) Segmented Image using EFCMC by Mahalanobis distance with 6-clusters and (c) Parrot Segmented Image using EFCMC by Euclidean distance with 6-clusters.

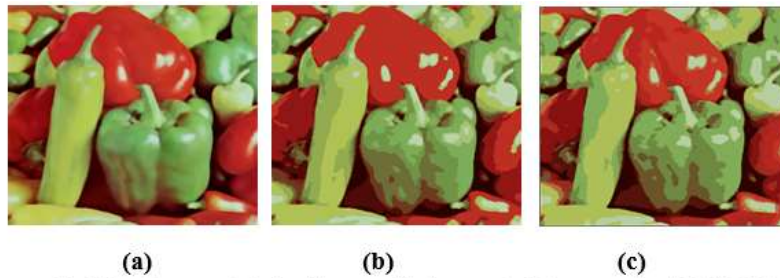


Figure 5: shown (a) Color Pepper Original Image (b) Segmented Image using EFCMC by Mahalanobis distance with 12-clusters and (c) Color Pepper Segmented Image using EFCMC by Euclidean distance with 12-clusters

Table 1: Shows the quality measures for color images after application Improved FCMC using Mahalanobis and Euclidean distance.

Image	Clusters	Improved FCMC (Mahalanobis)			Improved FCMC (Euclidean)		
		Iteration	PSNR	Normalized Absolute Error	Iteration	PSNR	Normalized Absolute Error
Canary	3	13	29.749	0.2394	15	30.419	0.1364
Parrot	3	10	30.866	0.2301	15	31.314	0.1386
Peppers color	3	12	29.560	0.2402	18	31.073	0.1374

Table 2: Shows the quality measures for color images after application Improved FCMC using Mahalanobis and Euclidean distance.

Image	Clusters	Improved FCMC (Mahalanobis)			Improved FCMC (Euclidean)		
		Iteration	PSNR	Normalized Absolute Error	Iteration	PSNR	Normalized Absolute Error
Canary	6	20	30.943	0.2338	13	31.052	0.1385
Parrot	6	18	28.816	0.2356	25	31.305	0.1410
Peppers color	6	12	26.215	0.2464	22	32.931	0.1356

Table 6: Shows the quality measures for color images after application Improved FCMC using Mahalanobis and Euclidean distance.

Image	Clusters	Improved FCMC (Mahalanobis)			Improved FCMC (Euclidean)		
		Iteration	PSNR	Normalized Absolute Error	Iteration	PSNR	Normalized Absolute Error
Canary	12	36	32.012	0.2235	42	34.082	0.1415
Parrot	12	28	30.688	0.2326	32	33.004	0.1423
Peppers color	12	34	32.785	0.2284	36	34.309	0.1426

6. CONCLUSION

1. The segmented color image is done using Improve FCMC clustering algorithm in RGB space, it workings perfectly fine with color images using both method Mahalanobis and Euclidean distance.
2. The segmented image precision using Euclidean distance is very good compared to the segmented image using Mahalanobis distance.
3. The precision of the segmented image based on clusters using Improve FCMC algorithm by Euclidean and Mahalanobis distance method.
4. All results are matched using Improve FCMC algorithm by Mahalanobis and Euclidean distance with some of quality measures.
5. The results obtained from the Improve FCMC algorithm using Euclidean distance more accurate than the results obtained from Improve FCMC algorithm using Mahalanobis distance. Where the Improve FCMC algorithm using Euclidean distance can extract the information and the features of colored images by adapting the value of precision of the image input to the algorithm
6. As one can see from the above image in the previous page that the image segmented using Improve FCMC with Euclidean distance did come better than using Mahalanobis distance when the image is segmented with 6 and 12 clusters

REFERENCES

1. Schmid, P.: Image segmentation by color clustering, <http://www.schmid-saugeon.ch/publications.html>, 2011
2. Neary D., "Fractal Methods in Image Analysis and Coding", M. Eng. thesis, Dublin City University, 2002.
3. Srikanteswara S., "Feature Identification in Wooden Boards Using Color Image Segmentation", M. Sc Thesis, State University, Electrical and Computer Engineering, 1997.
4. Young I.T., Gerbrands J.J., and . van Vliet L.J," Image Processing Fundamentals", Netherlands Organization for Scientific Research (NWO) Grant 900-538-040, 1998.
5. Baogang W., Dongming L., Yunhe P., and Wenhua X., "Interactive Image Segmentation Using Multiple Color Spaces and Its Application in Ancient Art Preservation", Artificial Intelligence Institute, Zhejiang University, Hangzhou, 2000, China P. R. 310027, www.rostock.zgdv.de.
6. Zhao B., "Color Space", Electrical Engineering, SUNY, NY, 2002, www.ece.sunysb.edu
7. Moore R., "Digital Image Processing ", Mathematics Department, Macquarie University, Sydney, 1999.
8. Scott .E. Umbaugh, "Computer Vision & Image Processing: A practical Approach Using CVIP tools ", Prentice Hall. Inc. 1998.
9. Gonzalez c. Rafael , Richard E. Woods" Digital Image processing " Addison-Wesley, 2002.

10. Rastislav Lukac , Konstantinos N. Plataniotis " Color Image Processing – Method and Application " University of Toronto
11. karbek W. ,and Koschan A., "Colour Image Segmentation :A Survey", Technical University of Berlin, 1994.
12. Wesolkowski S. B., "Color Image Edge Detection and Segmentation: A Comparison of the Vector Angle and the Euclidean Distance Color Similarity Measures ", M. Sc. Thesis, University of Waterloo, 1999.
13. Ilic S. , and Ulicny B., "Seeded Region Growing Method for Image Segmentation", the Swiss Federal Institute of Technology, 2000.
14. Sadiq A Mehdi, Khalid Kadhim Jabbar, Fadhil Hanoon Abbood "Image Encryption Based on The Nove15D Hyper-Chaotic System Via Improved AES Algorithm" International Journal of Civil Engineering and Technology, Publication date 2018.
15. Rana Saad Mohammed, Fadhil Hanoon Abbood, Intisar Abid Yousif "Image encryption technique using clustering and stochastic standard map" Conference :2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA) IEEE, 2016/5/9.
16. Fadhil Hanoon Abbood, Rana Saad Mohammed, Intisar Abid Yousif "Random Chaotic Number Generation based Clustered Image Encryption" International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2763 Issue 03, Volume 3 (March 2016)