ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 4, September 2014**

# Moving Fuzzy K-Means Conventional Clustering Algorithms for Microscopic Images

B.Swetha[1], D.Bulla Rao[2], P. Nageswara Rao[3]

[1,2,3] Department of CSE, Swetha Institute of Technology and Science, Tirupathi, India

**ABSTRACT:** Clustering algorithms have successfully been applied as a digital image segmentation technique in various fields and applications. However, those clustering algorithms are only applicable for specific images such as medical images, microscopic images etc. In this paper, we present a new clustering algorithm called Adaptive Fuzzy-K-means (AFKM) clustering for image segmentation which could be applied on general images and/or specific images (i.e., medical and microscopic images), captured using different consumer electronic products namely, for example, the common digital cameras and CCD cameras. The algorithm employs the concepts of fuzziness and belongingness to provide a better and more adaptive clustering process as compared to several conventional clustering algorithms. Both qualitative and quantitative analyses favour the proposed AFKM algorithm in terms of providing a better segmentation performance for various types of images and various number of segmented regions. Based on the results obtained, the proposed algorithm gives better visual quality as compared to several other clustering methods.[1]

**KEYWORDS**: Adaptive Fuzzy-K-means Clustering (AFKM), clustering, image segmentation, digital image processing.

## I. INTRODUCTION

Clustering is a process of grouping a set of objects into classes of similar characteristics. It has been extensively used in many areas, including in the statistics, machine learning, pattern recognition, data mining, and image processing. In digital image processing, segmentation is essential for image description and classification. The technique is commonly used by many consumer electronic products (i.e., conventional digital image) or in a specific application field such as the medical digital image. The algorithms are normally based on similarity and particularity, which can be divided into different categories; thresholding, template matching, region growing, edge detection, and clustering.

Clustering algorithm has been applied as a digital image segmentation technique in various fields such as engineering, computer, and mathematics. Recently, the application of clustering algorithms has been further applied to the medical field, specifically in the biomedical image analysis wherein images are produced by medical imaging devices. Previous studies proved that clustering algorithms are capable in segmenting and determining certain regions of interest in medical images. In a biomedical image segmentation task, clustering algorithm is often deemed suitable since the number of cluster for the structure of interest is usually known from its anatomical information. Among the clustering formulations based on minimizing formal objective functions, the most widely used and studied is the K-means (KM) clustering. KM is an exclusive clustering algorithm, (i.e., data which belongs to a definite cluster could not be included in another cluster). Although it is the most favourable technique, it does have some weaknesses:

1. It is dependent on initialization.
2. It is sensitive to outliers and skewed distributions.
3. It may converge to a local minimum.
4. It may miss a small cluster.

As a result, it may lead to poor or wrong representation of data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 4, September 2014

## II. THE PROPOSED CLUSTERING ALGORITHM

For details on the implementation of the proposed AFKM, consider a digital image with R×S pixels (i.e., R represents number of columns and S represents number of rows) to be clustered into nc regions or clusters. Let p(x,y) be the considered pixel and cj as the j-th centre, where x = 1, 2, ...., R,y=1, 2, ...., S and j = 1, 2, …., nc. Based on the aforementioned consideration, for the conventional KM and MKM, all data will be assigned to the nearest centre based on Euclidean distance. The new position for each centre is calculated using:

$$c_j = \frac{1}{n_{c_j}} \sum_{x \in c_j} \sum_{y \in c_j} p(x,y) \qquad (1)$$

For the conventional FCM, the process of allocating each data member to be assigned simultaneously to more than one class is based on the following membership function:

$$M^m jp(x,y) = \frac{1}{\sum_{k=1}^{n_1} \left( \frac{d_{jp(x,y)}}{d_{kp(x,y)}} \right)^{2/(m-1)}}; if\ d_{kp(x,y)} > 0, \forall j, p(x,y) \qquad (2)$$

$$\left.\begin{aligned} &M^m kp(x,y) = 1 \\ &M^m jp(x,y) = 0;\ for\ p(x,y) \neq k \end{aligned}\right\} if\ d_{kp(x,y)} = 0; \qquad (3)$$

where $d_{jp(x,y)}$ is distance from point $(x,y)$ to the current cluster centre $j$, $d_{kp(x,y)}$ is distance from point $(x,y)$ to other cluster centres $k$, $n_c$ is number of centres and m is an integer, $m > 1$ which determine the degree of fuzziness.

For the proposed AFKM algorithm, all centres are firstly initialized to a certain value. In order to ensure a better clustering process, (1) is no longer employed to update the centre, hence, fuzziness and belongingness concepts are introduced in the proposed AFKM algorithm. The concept of fuzzy partitioning concept is applied to allow each data to be assigned to more than one class simultaneously by different degrees of membership, $M^m jp(x,y)$. The membership function $M^m jp(x,y)$ is determined by using (2) and (3). In AFKM, the membership function measurement is performed within the data members without any influence from outsiders (non-members). Usually, the closer the connection of the data, the bigger the significance of impact on the clustering results. Thus, this provides a significant impact on the degree of membership, which is then used in determining the new position of the centre.

For a good clustering process, the modification concept introduced in the AFKM algorithm suggests that each cluster should have a significant value of belongingness which measures the relationship strength between the centre and its members. Therefore in the proposed AFKM, after specifying the membership for each data, the degree of belongingness, $B_j$ for each cluster is calculated. This relationship provides a significant impact on the clustering results through the degree of belongingness. Higher degree of belongingness shows a stronger relationship between the centre and its members, which will ensure a better data clustering. The degree of belongingness, $B_j$ is calculated based on:

$$B_j = \frac{c_j}{M^m jp(x,y)} \qquad (4)$$

In order to improve the clustering process, it is necessary for the degree of membership is needed to always be updated. The degree of membership is optimized based on the degree of belongingness to ensure that the procedure of the reassigning member places the data to its appropriate centre or cluster. By introducing the concept of belongingness, the AFKM suggests that each centre or cluster should have members with a strong relationship formed between them and the difference of belongingness among the clusters should be minimized. This criterion could be achieved using the followingprocesses.

$M^m{}_{jp(x,y)}$ is updated in the iteration according to :

$$\left(M^m{}_{jp(x,y)}\right)' = M^m{}_{jp(x,y) + \Delta M^m{}_{jp(x,y)}} \quad (5)$$

Where $\left(M^m{}_{jp(x,y)}\right)'$ is the new membership. $\Delta M^m{}_{jp(x,y)}$ is defined as :

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 4, September 2014**

$$\Delta M^{m}{}_{jp(x,y)} = \alpha\left(c_j\right)\left(e_j\right) \qquad (6)$$

where α is a designed constant with value between 0 and 1 and is typically set to 0.1. Subsequently, the value of $e_j$ is calculated according to:

$$e_j = \mathrm{B}_j - \hat{\mathrm{B}}_j \qquad (7)$$

Where $\hat{\mathrm{B}}_j$ is the normalized value for degree of belongingness.

Finally the new centre positions of the entire existing clusters are calculated based on the new (optimized) membership function according to :

$$C_j = \frac{\sum_{x\in c_j}\sum_{y\in C_j}\left(M^{m}{}_{jp(x,y)}\right)' p\left(x,y\right)}{\sum_{x\in c_j}\sum_{x\in c_j}\left(M^{m}{}_{jp(x,y)}\right)^{1}} \qquad (8)$$

## III. DATA SAMPLES AND ANALYSIS

The MKM clustering is chosen because it has been proven to produce better segmentation performance as compared to the conventional KM clustering algorithm. For all clustering algorithms, centre initializations are set according to:

$$C_j = \min{}_{p(x,y)} + \left(2j+1\right)\times\left(\frac{\max{}_{p(x,y)-\min_{p(x,y)}}}{2n_c}\right) \qquad (9)$$


**Fig. 1. The original images of (a) Air Force (b) Peppers (c) Elaine**

The performance analysis is carried out based on qualitative and quantitative analyses. The qualitative analysis refers to the visual interpretation that is observed via unaided human visual perception. In this case, it determines the capability of the segmentation algorithms in detecting and distinguishing the regions of interest from unwanted backgrounds. In our studies, the qualitative analysis is a judged by a panel of experts from Universiti Sains Malaysia and Universiti Teknologi MARA Malaysia. Nonetheless, the quantitative analysis enables the performance of the segmentation results produced by the proposed algorithm and the other algorithms to be evaluated statistically. Here, three evaluation functions are used as the quantitative benchmarks. The three functions are; *F(I)* proposed by Liu and Yang,

$$F(I) = \sqrt{R}\sum_{i=1}^{R}\frac{e_i^{2}}{\sqrt{A_i}} \qquad (10)$$

*F'(I)* proposed by Borsotti et al. [36],

$$F(I) = \frac{1}{1000\left(N\times M\right)}\sqrt{\sum_{A=1}^{Max}\left[R(A)\right]^{1+1/A}}\sum_{i=1}^{R}\frac{e_i^{2}}{\sqrt{A_i}} \qquad (11)$$

And *Q(I)* the improved version from *F(I)* proposed by Borsotti et al. [36],

$$Q(I) = \frac{1}{1000\left(N\times M\right)}\sqrt{R}\sum_{i=1}^{R}\left[\frac{e_i^{2}}{1+\log A_i} + \left(\frac{R(A_i)}{A_i}\right)^{2}\right] \qquad (12)$$

For the above three formulae, $I$ is the resultant image to be evaluated, $N \times M$ is the image size, $R$ is the number of regions found, Ai is the size of the $i$-th region, and $R(A_i)$ is the number of regions having area $A_i$. $e_i$, is defined as the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 4, September 2014**

sum of Euclidean distances between the features of pixels of region I and the corresponding region in the segmented image. With reference to, R*(A)* is the number of regions having exactly area A, and Max denotes the area of the largest region in the segmented image. These functions allow the segmentation to be evaluated without labelling the image and without the requirement of user-set parameters. Moreover, these functions correspond more closely to visual judgement. Smaller values of *F(I), F'(I),* and *Q(I),* show better segmentation results.

## IV.**RESULT AND DISCUSSION**

**Qualitative Analysis**

We have tested the AFKM on various images with different numbers of clusters, some of which are illustrated in Figs. 2 to 4. In this study, the main criterion used to evaluate the segmentation performance of the proposed algorithms is based on the capability to outline the desired important regions in the image.

First, the proposed AFKM, the FCM and the MKM clustering algorithms are tested on the aforementioned ten standard images with the number of clusters being set to 3. The resultant images are illustrated in Fig. 2. Then we compare the capability of those clustering algorithms in providing good image segmentation performance as the number of clusters increases. Figs. 3 and 4 illustrate the image segmentation performance as the number of clusters is set to 4 and 5 respectively. Significant changes on the images are highlighted with arrows and/or circles.

Let us consider the segmentation results on the three standard images (i.e., Air Force, Elaine, and Golden Gate) when the number of cluster is set to 3. From the first row of Fig. 2, (i.e., the Air Force image) it can be seen that the conventional algorithms and the proposed AFKM algorithm produce almost very similar results; but as compared with the FCM, the AFKM shows abetter shading result particularly on the mountain area. Moreover, the label "U.S Air Force" on the body of plane appears clearer and sharper in the resultant image from the proposed AFKM as compared to other images produced by the conventional clustering algorithms. The picture becomes clearer as we increase the number of cluster to 4 and then 5 (as depicted in Figs 3, and 4, respectively), with the "U.S Air Force" label and the mountain area become sharper and crispier. In Fig. 5 (i.e., the cropped image of Air Force on its tail area) it can be seen that the details on the tail of the plane are remain unaffected by the proposed AFKM as compared to the FCM, and the MKM. The serial number, highlighted with a circle and arrow is sharper and clearer for the image after the application of the proposed AFKM, while in the other resultant images produced by the FCM, and the MKM, the serial number is corrupted with noise (i.e., occurrence of white spots on the serial numbers). Overall, the proposed AFKM algorithm has outperformed other clustering algorithms. The segmented images produced by the proposed AFKM are sharper and crispier with less noisy pixels.

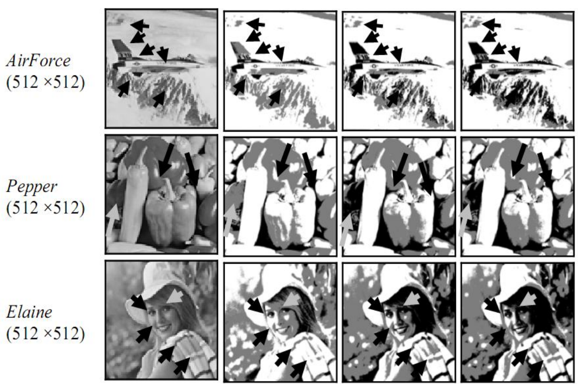


**Fig.2. Resultant segmented images with 3 clusters after applying: First Column : Original images. Second column: FCM. : Original Third column: MKM. Fourth column : AFKM. Third column:**


**Fig.3. Resultant segmented images clusters after applying : First column images. Second column : FCM. MKM. Forth column : AFKM.**

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 4, September 2014**

For the Elaine image as shown in the third row of Fig. 2, as we cluster the image into three regions, the folded line on the Elaine's towel is seen to be unclear for the resultant images produced by the FCM and the MKM. Nonetheless, the proposed algorithm has lined the folded towel much better than the conventional techniques. Elaine's hair and eyebrow are segmented clearly by the proposed AFKM algorithm as compared to the FCM, and the MKM, which segmented the aforementioned areas rather poorly. The segmentation regions of hair and eyebrow of Elaine, and the folded line on the Elaine's towel produced by the proposed AFKM are much clearer as we increase the number of cluster from 3 to 4 and 5. The resultant images for number of clusters set to 4 and 5 are shown in Figs. 3, and 4, respectively. A more significant comparison between the clustering algorithms can be observed from the cropped image of Elaine on her towel as shown in Fig. 6.


**Fig. 4. Resultant segmented images with 5 clusters after applying: First column: Original images. Second column: FCM. Third column: MKM. Forth column: AFKM.**

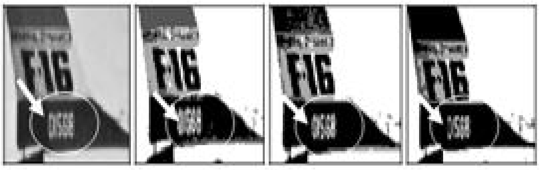

**Fig. 5. The segmentation results (3 clusters) for cropped image of Air Force. First column: Original images, Second column: FCM. Third column: MKM. Forth column: AFKM.**

**Fig. 6. The segmentation results (4 clusters) for cropped image of Elaine. First column: Original images. Second column: FCM.**

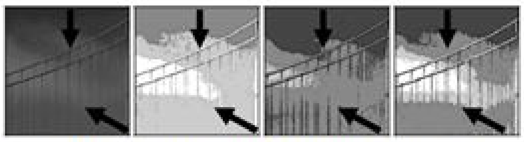

**Fig. 7. The segmentation results (5 clusters) for cropped image of Golden Gate. First column: Original images. Second column: FCM. Third column: MKM. Forth column: AFKM.**

With reference to the seventh row of Fig.2, (i.e., the Golden Gate image) it can be seen that again, the proposed AFKM algorithm has successfully produced a better segmentation performance. The main structure of the bridge image namely the beams and the pillars are significantly detected and segmented while the conventional FCM clustering algorithm have evidently produced poor segmentation performance. The same observation can also be seen when the number of cluster is set to 4 and 5 as shown in Figs. 3 and 4 respectively. The aforementioned advantages can be clearly seen from the cropped Golden Gate image as shown in Fig. 7.

For the remaining seven images, the results obtained in Figs. 2 to 4 further favour the proposed AFKM as the best segmentation technique compared to the FCM and the MKM clustering algorithms. Significant differences between those algorithms are highlighted with arrows. Based on these results, the proposed algorithm can be suggested to be used as post processing tools for digital images; an added-value to existing digital image processing techniques in recent consumer electronics products (e.g., CCD camera).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(An ISO 3297: 2007 Certified Organization)*

**Vol.2, Special Issue 4, September 2014**

## IV. CONCLUSION

This paper presents a new clustering algorithm named the AFKM algorithm for segmentation purposes. It employs an adaptive and iterative fuzzy and belongingness concept to obtain the optimum value of clusters centre for a better segmentation process. The proposed clustering algorithm is applicable to numerous outdoor and indoor images whereby ten images have been presented as case studies. The conclusion of this paper sees the proposed algorithm outperforming the conventional FCM and MKM algorithms by successfully producing better segmented images. The proposed AFKM also successfully preserves important features on digital images. Both qualitative and quantitative analyses have justified the conclusion that the proposed approach has been able to illustrate good segmentation results efficiently. Thus, it is recommendable for this algorithm to be applied in the post image processing in consumer electronic products such as the digital camera for general applications and the CCD camera which is extensively used with the microscope in capturing microscopic images, especially in segmenting medical images.

| Number of Cluster | Algorithms | Quantitative evaluation functions | | |
|---|---|---|---|---|
| | | F(I) ($*1.0e+^{12}$) | F'(I) ($*1.0e+^{5}$) | Q(I) ($*1.0e+^{6}$) |
| 3 | FCM | 7.8030 | 0.3269 | **0.8374** |
| | MKM | 10.9354 | 0.4441 | 1.0336 |
| | AFKM | **6.1938** | **0.2609** | 0.8495 |
| 4 | FCM | 13.6391 | 0.5766 | 1.4149 |
| | MKM | 6.7291 | 0.2902 | 1.4543 |
| | AFKM | **5.6972** | **0.2448** | **1.2936** |
| 5 | FCM | 46.4756 | 1.8677 | 2.2439 |
| | MKM | 9.5417 | 0.4205 | 2.1243 |
| | AFKM | **8.7510** | **0.3744** | **1.9115** |

**Table -1 : Average Segmentation evaluation functions on 64 images**

| Images | Algorithms and Cluster Number | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | FCM | | | MKM | | | AFKM | | |
| | 3 | 4 | 5 | 3 | 4 | 5 | 3 | 4 | 5 |
| *AirForce* | 4.9 | 4.65 | 4.40 | 2.46 | 2.76 | 2.55 | 4.74 | 5.47 | 6.36 |
| *Pepper* | 5.14 | 5.24 | 5.33 | 2.46 | 2.46 | 2.57 | 3.77 | 3.80 | 4.52 |
| *Elaine* | 3.27 | 4.20 | 4.55 | 2.46 | 2.64 | 2.59 | 4.03 | 5.05 | 5.40 |
| *Smarties1* | 1.47 | 1.68 | 1.89 | 0.97 | 1.35 | 1.31 | 1.37 | 1.93 | 1.97 |
| *Pirate* | 2.95 | 4.71 | 5.71 | 2.43 | 2.47 | 2.73 | 4.64 | 6.06 | 7.44 |
| *Smarties2* | 1.68 | 1.67 | 1.66 | 1.14 | 1.01 | 1.25 | 1.46 | 1.62 | 2.25 |
| *GoldenGate* | 3.16 | 5.12 | 4.45 | 2.39 | 2.55 | 2.51 | 4.50 | 4.93 | 6.08 |
| *Butterfly* | 3.20 | 4.52 | 5.13 | 2.56 | 2.61 | 2.49 | 3.39 | 3.58 | 4.71 |
| *River Bridge* | 3.46 | 4.46 | 6.34 | 2.42 | 2.44 | 2.60 | 3.51 | 3.65 | 3.87 |
| *Boat* | 4.73 | 4.72 | 5.68 | 2.56 | 2.71 | 2.54 | 3.83 | 4.33 | 4.73 |

**Table – 2 : Execution Time (in Seconds)**

## REFERENCES

[1] D. Auber, and M. Delest, "A clustering algorithm for huge trees," Advances in Applied Mathematics, vol. 31, no. 1 pp. 46-60 2003.

[2] A. S. Mahani, A. E. Carlsson, and R. Wessel, "Motion repulsion arises from stimulus statistics when analyzed with a clustering algorithm," Biological Cybernetics, vol. 92, no. 4, pp. 288-291, 2005.

[3] T. Abeel, Y. V. d. Peer, and Y. Saeys, "Java-ML: A machine learning library," Journal of Machine Learning Research, vol. 10, pp. 931-934, 2009.