

Two-Dimensional Clustering Algorithms for Image Segmentation

INTAN AIDHA YUSOFF¹, NOR ASHIDI MAT ISA²
 Imaging and Intelligent System Research Team (ISRT)
 Universiti Sains Malaysia Engineering Campus
 Nibong Tebal, Penang
 MALAYSIA
 intan_aidha@yahoo.com¹, ashidi@eng.usm.my²

Abstract : - This paper introduces modified versions of the K-Means (KM) and Moving K-Means (MKM) clustering algorithms, called the Two-Dimensional K-Means (2D-KM) and Two-Dimensional Moving K-Means (2D-MKM) algorithms respectively. The performances of these two proposed algorithms are compared with three of the commonly used conventional clustering algorithms, namely K-Means (KM), Fuzzy C-Means (FCM), and Moving K-Means (MKM). The new algorithms incorporate the median value of considered pixel intensity with its neighboring pixel; together with the pixel's own intensity for the assigning process of the pixel to the nearest cluster. From the observed qualitative and quantitative results, it is proven that 2D-KM and 2D-MKM perform better than KM, FCM, and MKM in terms of producing more homogeneous segmentation results, while taking shorter time in executing the process as compared to FCM.

Key-Words: - Two-Dimensional K-Means (2D-KM), Two-Dimensional Moving K-Means (2D-MKM), Image Segmentation, Clustering.

1 Introduction

Along with the fast development of consumer products in digital imaging and photography, there are numerous applications of segmentation process, especially in machine vision. Image segmentation is an important part in understanding many computer vision-based systems [1]. There are more than one approach in segmentation process, including region growing [2],[3], clustering [1],[4],[5] edge detection [6],[7], template matching [8],[9], and thresholding [10],[11].

Clustering has been implemented widely in the diverse scientific field, such as pattern recognition [12]-[14], machine learning [15],[16], spectral clustering [17], and medical image processing [1],[18]-[21]. In the medical image segmentation, most applications involve automatic extraction of features from the image which is then used for a variety of classification tasks, such as distinguishing normal tissues from abnormal tissues [20], or in the segmentation of soft tissues [21].

As many clustering algorithms have been developed over the years, with improvements proposed over time, the segmentation ability of each clustering algorithm is steadily improved through time. Some of the most widely used and studied clustering algorithms are K-Means (KM), Fuzzy C-Means (FCM), and Moving K-Means (MKM).

K-Means algorithm was originally proposed by Forgy and MacQueen in 1967 [22]. In image processing, KM clustering algorithm assigns a pixel to its nearest cluster centre using the Euclidean distance based on the pixel's intensity value. Later in 1973, Dunn had developed the FCM clustering, which was later further improved by Bezdek in 1981 [22]. This algorithm allows a data to be a member of more than one cluster with a certain level of membership.

Reference [23] has proposed the MKM clustering algorithm to overcome limitations of KM which are [4],[24]:

- Its dependency on initialization.
- It is sensitive to outliers and skewed distributions.
- It may converge to local minimum.
- It may miss a small cluster.

In addition, the MKM algorithm also minimizes dead centres and centre redundancy problems while indirectly reducing centres to be trapped at local minima [25].

All of the aforementioned algorithms perform the clustering process based on a single local parameter, namely the intensity value of a pixel. As in numerous image processing techniques (i.e. such as filtering, contrast enhancement etc) the feature of a pixel is commonly correlated to the effect of its

neighbouring pixels. By discarding this correlative behaviour, certain amount of image information is lost during the process. Hence, often in segmenting an image by using conventional one-dimensional clustering, the following limitations are observed:

- Noise pixels are considered as an independent feature, wrongly assigned to clusters, and stay visible after the segmentation process.
- By discarding the correlative effects of spatial parameters on a pixel, there are probabilities of information in an image being lost.
- The performance may degrade rapidly as the spatial interaction between pixels becomes more dominant than the gray level values [26].

In 1989, a study was carried out to utilize more information in an image, by using two-dimensional entropies (intensity/local average intensity) histogram into segmentation [26]. Since then, researchers have gone into utilizing spatial characteristics into image thresholding [27]-[30] and clustering [31]-[34]. These approaches have been proven to reduce information lost and noisy pixel interference in segmented images. Amongst the proposed methods, most threshold approaches use (intensity/local average intensity) of a pixel as spatial parameters, while most clustering approaches are adapting non-local spatial parameters to the Fuzzy C-Means clustering algorithm with modification on its calculation on membership. Whilst average intensity of local neighbouring pixels have always been an important spatial information of a pixel, the median value of neighbouring pixels may serve just as well, with insensitivity towards the skewness of intensity histogram as an advantage.

Thus, in this study we have chosen to incorporate local median as spatial information into KM and MKM clustering algorithms during the segmentation process in order to minimize information loss and produce a more homogeneous segmented image with less noise in the segmented regions.

The rest of this paper is organized as such: in Section 2 the proposed clustering algorithms are explained. Section 3 explains the methods of data analysis being used in this study. Section 4 analyses the results obtained from the proposed algorithm and evaluate its performances as well as comparison made with several selected conventional clustering algorithms by using both qualitative and quantitative analyses. Finally, Section 5 concludes the work of this paper.

2 Proposed Approach

As mentioned in Section I, the conventional KM and MKM clustering algorithms employ the nearest Euclidean distance concept in assigning pixels to their respective cluster, with pixels' intensity values as a sole parameter in this particular approach. We focus on the modification and enhancement of both algorithms by incorporating a new local spatial parameter in determining the nearest Euclidean distance, which is the value of the intensity median of the considered pixel and its 3×3 neighboring pixels. The proposed algorithms are known as 2D-KM and 2D-MKM. For the implementation of the proposed clustering algorithms, consider N as the number of data to be clustered into n_c regions or clusters. Let v_t be the t -th data where $t=1,2,\dots,N$ and c_k is the k -th centre.

2.1 Two-Dimensional K-Means Clustering Algorithm

Generally, the conventional KM clustering algorithm will minimize the following objective function of partitioning a dataset $\{v_t\}_{t=1}^N$ into k -th centre, c_k [28]:

$$J = \sum_{k=1}^{n_c} \sum_{t=1}^N \|v_t - c_k\|^2 \quad (1)$$

where $\|\cdot\|$ stands for a distance measure that is normally taken to be the Euclidean norm. In segmenting an 8-bit gray scale digital image with 256 gray levels in the interval $[0, 255]$ by the conventional KM, $v_t = p(x,y)$ where $p(x,y)$ is the pixel at location (x,y) with the intensity p (where $x=1,2,3,\dots,R$ and $y=1,2,3,\dots,S$, with R and S are number of columns and rows of the image respectively). With predetermined initial values for all clusters, all data will be first assigned to the nearest centre based on the Euclidean distance. Then, the new position for each centre is calculated using:

$$c_k = \frac{1}{n_{c_k}} \sum_{t \in c_k} v_t \quad (2)$$

The process is repeated until the value of all centres no longer change. In order to include the effect of the local spatial information of an image (i.e. median intensity value of 3×3 neighboring pixels) as in the proposed 2D-KM, the v_t and c_k are modified and represented by (3) and (4) respectively:

$$v_t = (v_{tINT}, v_{tMED}) \quad (3)$$

$$c_k = \frac{1}{n_{c_k}} \left[\left(\sum_{t \in c_k} v_{tINT} \right), \left(\sum_{t \in c_k} v_{tMED} \right) \right] \quad (4)$$

where v_{tINT} is the intensity vector of t -th data, v_{tMED} is the median vector of t -th data, and n_{c_k} is number of pixels assigned to k -th centre.

2.2 Two-Dimensional Moving K-Means Clustering Algorithm

For the proposed 2D-MKM clustering algorithm, it uses the similar concept of the conventional MKM proposed by [35]. Concept of fitness is introduced to ensure that each cluster should have a significant number of members and final fitness values before the new position of cluster is calculated. The fitness for each cluster is calculated using:

$$f(c_k) = \sum_{t \in c_k} (\|v_t - c_k\|)^2 \quad (5)$$

where in the proposed 2D-MKM, v_t and c_k are represented by (3) and (4) respectively. From (5), C_s and C_l the centre with the smallest and the largest fitness values respectively, are determined. Based on the MKM algorithm the relationship between C_s and C_l should satisfy the following condition:

$$f(C_s) \geq \alpha_a f(C_l) \quad (6)$$

where α_a is a small constant value, initially set to be equal to α_o . α_o is a designated constant with value in range $0 < \alpha_o < 1/3$. If (6) is not fulfilled, the members of C_l which are larger than C_l are assigned as members of C_s while the rest are maintained as the members of C_l . Then, the positions of C_s and C_l are recalculated according to (7) and (8) respectively:

$$C_s = \frac{1}{n_{C_s}} \left[\left(\sum_{t \in C_s} v_{tINT} \right), \left(\sum_{t \in C_s} v_{tMED} \right) \right] \quad (7)$$

$$C_l = \frac{1}{n_{C_l}} \left[\left(\sum_{t \in C_l} v_{tINT} \right), \left(\sum_{t \in C_l} v_{tMED} \right) \right] \quad (8)$$

The value of α_a is then updated according to:

$$\alpha_a = \alpha_a - \alpha_a / n_c \quad (9)$$

The above processes are repeated until (6) is fulfilled. After the (6) is fulfilled the following condition is observed:

$$f(C_s) \geq \alpha_b f(C_l) \quad (10)$$

If it is not fulfilled, all processes are repeated. In each iteration, the value of α_b is updated according to:

$$\alpha_b = \alpha_b - \alpha_b / n_c \quad (11)$$

While the value of α_a is reset to α_o .

3 Data Analysis

In order to analyze the segmented performance for processing images, a total of 73 gray-scale standard images have been tested using the conventional and proposed clustering algorithms. In addition, for evaluation on real world applications, all clustering algorithms were applied on medical pathology image of cervical cells.

Each image is tested using KM, FCM, MKM, 2D-KM, and 2D-MKM clustering algorithms with three different number of clusters; three, four, and five. Ten standard images and five Thin-Prep cell images are elaborated qualitatively. The rest of the tested images' results will be used for average quantitative performance analysis. In evaluating a clustering process, there are no predefined classes and examples that show what kind of desirable relations should be valid amongst data [36]. However there are analyses which have been proposed to evaluate the quality of segmentation of clustering algorithms [37],[38]. In this study, four types of quantitative analyses are used, namely $F(I)$, $F'(I)$, $Q(I)$, and processing time.

In image and signal processing applications, short processing time is one of the most desired capabilities and has always been one of the most important benchmark in determining field performance. In image processing, it also denotes the simplicity of an algorithm. Thus, we have taken this parameter into consideration. In addition, a good segmentation should incorporate the following criteria [39]:

- The segmented regions must be uniform and homogeneous.
- The region's interiors must be simple, without too many small holes.
- Adjacent regions must present significantly different values for uniform characteristics.

In 1994, Liu and Yang designed a function which caters for evaluating a segmentation performance based on all the aforementioned criteria [37]:

$$F(I) = \frac{1}{1000(N \times M)} \sqrt{R} \sum_{i=1}^R \frac{e_i}{\sqrt{A_i}} \quad (12)$$

where I is the segmented image, $N \times M$ is the image size, R is the number of regions in the clustered image, A_i is the area, and e_i is the Euclidean distance between the gray level color vectors of the pixels of i -th region and the color vector attributed to region i in the segmented image. In 1998, Borsotti et al revised the $F(I)$ function and came up with $F'(I)$ and $Q(I)$ evaluation functions[38]:

$$F'(I) = \frac{1}{10000(N \times M)} \sqrt{\sum_{A=1}^{Max} [R(A)]^{1+1/A}} \times \sum_{i=1}^R \frac{e_i^2}{\sqrt{A_i}} \quad (13)$$

$$Q(I) = \frac{1}{10000(N \times M)} \sqrt{R} \times \sum_{i=1}^R \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{R(A_i)}{A_i} \right)^2 \right] \quad (14)$$

where for (13), $R(A)$ is the number of region having exactly area A , and Max is the area of the largest region in the segmented image. In (14), $R(A_i)$ is the number of regions having an area equal to A_i . As observable from the functions, bigger number of regions, and a smaller region size will yield a larger result of $F(I)$, $F'(I)$ and $Q(I)$. Thus lower values of all three functions are desired as it proves that the segmentation done produces a smoother and more homogeneous segmentation where the number of noisy pixels presented in the segmented image is minimized.

4 Result and Discussion

4.1 Qualitative Analysis

In image processing and computer vision, image segmentation is a process of partitioning an image into multiple regions that are homogeneous with respect to one or more characteristics [40]. By using both standard images and medical images, we will

visually study the ability of proposed algorithms in segmenting images for general applications and also for professional needs, such as in the medical field.

4.1.1 Standard Image



Fig. 1. Original standard images. From top left: *Man*, *House*, *Flower*, *Lady*, *Nature*, *Elaine*, *Air Force*, *Tree*, *Peppers*, *Bird*.

For standard images as shown in Fig. 1, ten images namely *Man*, *House*, *Flower*, *Lady*, *Nature*, *Elaine*, *Air Force*, *Tree*, *Peppers*, and *Bird* have been chosen as test images for qualitative evaluation. The resultant images after applying the KM, FCM, MKM, 2D-KM, and 2D-MKM for number of clusters equal to 3, 4, and 5 are shown in Figs. 2 to 4 respectively. In all images, arrows are used to indicate the differences between these resultant images.

As seen in Fig. 2, when the number of clusters is set at 3 clusters, for the image *Lady*, the hand and face areas are more homogeneously segmented by the 2D-KM and 2D-MKM compared to the conventional clustering algorithms. Both proposed algorithms managed to give a cleaner segmented area of hand without any noise pixels which can be seen in results of conventional clustering algorithms. The 2D-MKM algorithm removed all small isolated regions in the face area, which could be seen in resultant image of the KM, FCM, MKM, and 2D-KM. Furthermore, although all conventional algorithms have difficulties in segmenting homogeneous background area, the proposed 2D-KM and 2D-MKM discard most of the untamed hair strands of the lady, making the background area more homogeneous. The segmented result of *Air Force* shows similar observations. The KM, FCM, and MKM algorithms have managed to segment the background into a single region, but with the presence of small regions inside it, this contributes to a less homogeneous clustering result. These small insignificant regions are successfully reduced by the

proposed 2D-KM and 2D-MKM algorithms. Furthermore, for the image titled *Tree*, a more homogeneous segmentation result could be seen in the leaves and tree regions when segmented using the proposed 2D-KM and 2D-MKM clustering algorithms.



Fig. 2. Segmented image with number of clusters equals to 3. First column: Image processed with KM. Second column: Image processed with FCM. Third column: Image processed with MKM. Fourth column: Image processed with 2D-KM. Fifth column: Image processed with 2D-MKM.

As we increase the number of clusters to 4 (as shown in Fig. 3), the segmented images of *Lady* still show more homogeneous face, hand, and

background areas for the proposed 2D-KM and 2D-MKM as compared to the conventional KM, FCM, and MKM clustering algorithms. For the image labeled *Air Force*, the results clearly show that both the 2D-KM and 2D-MKM give much better results in segmenting this image, by successfully segmenting the background (i.e. land) area into a homogeneous single region while the conventional algorithms segmented it into two different regions.



Fig. 3. Segmented image with number of clusters equals to 4. First column: Image processed with KM. Second column: Image processed with FCM. Third column: Image processed with MKM. Fourth column: Image processed with 2D-KM. Fifth column: Image processed with 2D-MKM.

For the *Tree* image, comparing all tested clustering algorithms, more homogeneous segmentation results can be seen in the leaves and tree areas when segmented using the 2D-KM and 2D-MKM clustering algorithms as compared to those using the KM, FCM, and MKM algorithms.



Fig. 4. Segmented image with number of clusters equals to 5. First column: Image processed with KM. Second column: Image processed with FCM. Third column: Image processed with MKM. Fourth column: Image processed with 2D-KM. Fifth column: Image processed with 2D-MKM.

For 5 clusters segmentation results (as shown in Fig. 4), it is still observed that image *Lady* has more homogeneous face and hand regions when

processed using the 2D-KM and 2D-MKM. The conventional clustering methods have poorly produced small isolated regions inside these two areas. For image *Air Force*, a single-clustered background is achieved only by using the proposed clustering methods.

Finally, the leaves and shadow regions of the tree in *Tree* image processed using the conventional KM, FCM, and MKM clustering algorithms are less homogeneous, unlike the ones processed using the 2D-KM and 2D-MKM. In general, regardless of the number of clusters, the 2D-KM and 2D-MKM algorithms continue producing more homogeneous segmented images as compared to the KM, FCM, and MKM algorithms.

4.1.2 Case Study – Medical Images Segmentation

For evaluation on case study (i.e. medical image segmentation), we have purposefully selected cervical cell images. The main objective of medical image segmentation is to extract and characterize anatomical structures with respect to important features for expert interpretation [18]. In such application, issues such as limited spatial resolution, poor contrast, noise, and non-uniform intensity variations make accurate segmentation a difficult task [41].

For the segmentation of cervical cell image, the number of clusters is set to 3 in order to segment the images into background, nucleus, and cytoplasm regions. A good clustering algorithm should not only be able to cluster these images into background, cytoplasm, and nucleus regions, but it also needs to preserve dimensional criteria of the cell such as the size of nucleus and cytoplasm. These criteria are important to pathologists in screening for cell abnormalities.

Fig. 5 shows 5 cervical cell images used as test images, while Fig. 6 presents the resultant images of test images segmented using the conventional KM, FCM, MKM, and the proposed 2D-KM, and 2D-MKM algorithms. Noted from Fig. 6, the 2D-KM clustering algorithm is able to cluster all cell images into background, cytoplasm, and nucleus regions with less ‘holes’ in the nucleus and cytoplasm areas, two important features which are crucial for features’ extraction of dimensions (i.e. size, area) of a cell. Smoother cytoplasm areas are produced as compared to those produced by the KM, FCM, MKM, and 2D-MKM algorithms. Thus these findings prove that 2D-KM has better potential in

the application of segmenting pathological-standard images as compared to the conventional clustering algorithms.

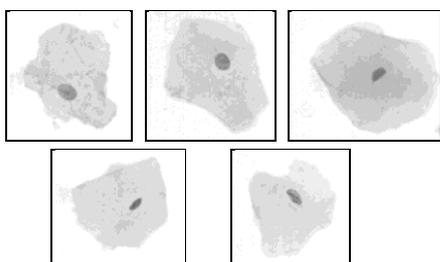


Fig. 5. Original Image of (from top left) *Cell1*, *Cell2*, *Cell3*, *Cell4*, *Cell5*.

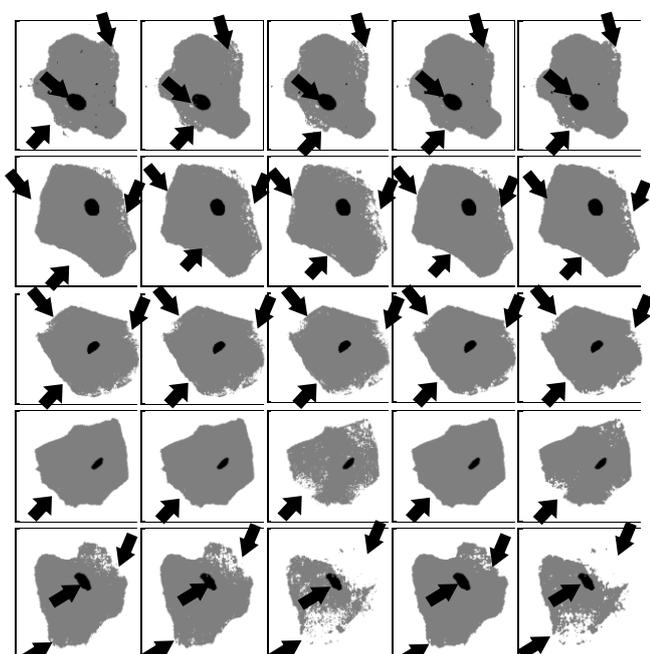


Fig. 6. Segmented cervical cell image with number of clusters equal to 3. From top down: *Cell1*, *Cell2*, *Cell3*, *Cell4*, *Cell5*. First column: Image processed with KM. Second column: Image processed with FCM. Third column: Image processed with MKM. Fourth column: Image processed with 2D-KM. Fifth column: Image processed with 2D-MKM.

4.2 Quantitative Analysis

Tables 1 to 5 show the results of the quantitative analysis for standard images. The best results obtained for all analyses are made bold. As seen in Tables 1 to 3, when clustering the tested images into 3 clusters, the $F(I)$, $F'(I)$, and $Q(I)$ values for the 2D-KM and 2D-MKM are smaller as compared to those stemming from the KM, FCM, and MKM algorithms. This proves better segmentation qualities obtainable from both algorithms.

As the number of clusters increases to 4 and 5 clusters, the proposed 2D-KM and 2D-MKM algorithms still produce better results as compared to the conventional methods. As these three functions are designed to penalize images with too many regions, ‘holes’, and noise, thus the results support the qualitative analysis where the resultant images segmented using the 2D-KM and 2D-MKM consist of more homogeneous and smoother regions. In developing clustering algorithms, one of the most important features is the simplicity and less time-consumption of an application. Thus, processing time analysis aims to favor an algorithm which takes less time to execute. From Table 4, it can be observed that even though the proposed algorithms does not execute in the shortest time, the readings are still in small variance from the conventional algorithms and is still comparable. In almost all of the images, the 2D-KM and 2D-MKM algorithms execute faster than the conventional FCM algorithm.

No of Clusters	Image	$F(I)$ For Clustering Algorithms (*1.0e+3)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	<i>Man</i>	1.0528	2.0400	1.7018	0.8447	1.3301
	<i>House</i>	1.5793	3.8532	1.5757	1.1209	1.0761
	<i>Flower</i>	1.0388	2.8542	1.0838	0.8475	0.8408
	<i>Lady</i>	2.4413	3.3044	2.4413	1.9761	2.0348
	<i>Nature</i>	1.6340	3.2572	3.5218	1.2741	1.9674
	<i>Elaine</i>	1.0452	1.1108	1.1450	0.8213	0.8125
	<i>Peppers</i>	1.4755	2.6298	2.4174	1.2745	1.3683
	<i>Air Force</i>	0.6032	0.5650	0.6032	0.4375	0.4371
	<i>Bird</i>	2.1114	5.6279	2.2753	1.8893	1.9238
	<i>Tree</i>	2.0412	7.7388	2.2159	1.8408	2.4957
4	<i>Man</i>	0.5675	1.0935	0.7427	0.4780	0.4891
	<i>House</i>	0.8131	1.5459	0.8413	0.6325	0.7368
	<i>Flower</i>	0.7888	1.0612	0.9338	0.7738	1.2307
	<i>Lady</i>	2.2426	1.3853	2.2417	1.6901	1.7540
	<i>Nature</i>	0.5786	1.1919	1.7883	0.5317	0.9795
	<i>Elaine</i>	0.3077	0.3304	0.7466	0.2384	0.9518
	<i>Peppers</i>	1.3236	1.3073	1.6919	0.6039	1.0185
	<i>Air Force</i>	0.2417	0.4236	0.4142	0.8439	0.6167
	<i>Bird</i>	1.0792	1.8317	1.3749	0.7246	1.6643
	<i>Tree</i>	0.8303	1.6700	0.8721	1.0123	0.9993
5	<i>Man</i>	0.3593	0.5211	0.5865	0.3524	0.4226
	<i>House</i>	0.5038	0.7566	0.6273	0.4238	0.5054
	<i>Flower</i>	0.4184	0.6968	0.6592	0.3910	0.3967
	<i>Lady</i>	0.4187	1.0647	0.6428	0.4860	0.6213
	<i>Nature</i>	0.4335	0.5567	0.9825	0.3853	0.6262
	<i>Elaine</i>	0.1714	0.3209	0.3392	0.1577	0.2756
	<i>Peppers</i>	0.5889	0.6065	0.9115	0.5801	0.6464
	<i>Air Force</i>	0.2902	0.2446	0.2899	1.0412	0.6888
	<i>Bird</i>	0.6412	0.7015	0.7268	0.5243	0.6548
	<i>Tree</i>	0.6170	0.7213	0.7053	0.7625	0.9135

Table 1 : Quantitative evaluation $F(I)$ on segmented standard images

No of Clusters	Image	F'(I) For Clustering Algorithms (*1.0e+2)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Man	1.1007	2.1436	1.7802	0.9006	1.4145
	House	1.6194	3.9564	1.6200	1.1703	1.1242
	Flower	2.0556	3.5704	2.0549	1.6291	1.4928
	Lady	2.5197	3.4114	2.5197	2.0754	2.1370
	Nature	1.6817	3.3416	3.6445	1.3374	2.0769
	Elaine	1.0835	1.1523	1.1897	0.8678	0.8602
	Peppers	1.4986	2.6667	2.4550	1.3066	1.4060
	Air Force	0.6263	0.5891	0.6263	0.4617	0.4627
	Bird	2.2254	5.8615	2.3992	2.0276	2.0582
	Tree	2.1140	7.9366	2.2923	1.9384	2.6081
4	Man	0.5884	1.1369	0.7691	0.5064	0.5157
	House	0.8335	1.5846	0.8643	0.6585	0.7657
	Flower	0.8179	1.1022	0.9606	0.8102	1.2762
	Lady	2.3030	1.4247	2.3020	1.7607	1.8289
	Nature	0.5944	1.2206	1.8377	0.5549	1.0213
	Elaine	0.3162	0.3400	0.7736	0.2500	0.9995
	Peppers	1.3448	1.3255	1.7182	0.6169	1.0409
	Air Force	0.2467	0.4318	0.4215	0.8769	0.6416
	Bird	1.1226	1.9011	1.4383	0.7608	1.7658
	Tree	0.8485	1.7167	0.8912	1.0464	1.0324
5	Man	0.3703	0.5389	0.6078	0.3689	0.4438
	House	0.5155	0.7750	0.6421	0.4405	0.5263
	Flower	0.4284	0.7230	0.6789	0.4065	0.4118
	Lady	0.4289	1.0914	0.6595	0.5046	0.6454
	Nature	0.4430	0.5718	1.0054	0.3992	0.6513
	Elaine	0.1756	0.3297	0.3488	0.1645	0.2895
	Peppers	0.5961	0.6135	0.9221	0.5924	0.6601
	Air Force	0.2960	0.2498	0.2957	1.0746	0.7121
	Bird	0.6612	0.7213	0.7510	0.5473	0.6832
	Tree	0.6274	0.7355	0.7177	0.7836	0.9426

Table 2 : Quantitative evaluation F'(I) on segmented standard images

No of Clusters	Image	Clustering Algorithms				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Man	1.66	2.08	1.67	2.08	1.60
	House	1.49	1.80	1.48	1.88	2.79
	Flower	1.50	2.28	1.41	1.57	1.60
	Lady	1.41	1.53	1.53	1.74	1.64
	Nature	2.62	1.74	1.48	1.77	1.55
	Elaine	1.53	1.89	1.52	1.53	2.81
	Peppers	1.48	2.82	2.57	1.65	2.78
	Air Force	1.52	1.82	1.43	1.69	1.62
	Bird	1.47	1.80	2.57	1.58	2.89
	Tree	1.55	3.09	1.44	1.65	2.76
4	Man	1.57	2.74	1.47	2.34	1.58
	House	1.51	3.88	2.63	1.65	1.64
	Flower	1.48	3.96	1.50	2.67	2.89
	Lady	1.52	2.45	1.49	2.34	2.89
	Nature	2.69	3.09	2.60	2.94	1.57
	Elaine	1.53	2.39	2.58	2.11	1.57
	Peppers	1.45	3.18	2.59	1.90	1.57
	Air Force	1.49	3.51	1.48	1.88	2.81
	Bird	1.55	2.35	1.40	1.74	1.60
	Tree	1.69	1.80	2.64	2.58	1.59
5	Man	2.96	6.46	1.52	3.99	1.63
	House	1.61	3.27	2.67	2.87	1.61
	Flower	2.79	2.98	1.48	2.69	1.64
	Lady	1.73	2.10	2.57	4.71	1.68
	Nature	1.59	8.55	1.45	2.55	1.62
	Elaine	1.53	2.57	2.55	2.05	1.63
	Peppers	3.16	5.90	1.51	3.15	1.64
	Air Force	2.61	1.86	1.46	2.58	1.71
	Bird	1.62	4.75	1.49	2.95	1.61
	Tree	1.63	3.22	1.45	2.21	1.65

Table 4 : Execution time (in seconds)

No of Clusters	Image	Q(I) For Clustering Algorithms (*1.0e+4)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Man	0.1845	0.4675	0.3632	0.1533	0.2795
	House	0.3196	1.0006	0.3138	0.2240	0.2083
	Flower	0.4562	0.8437	0.4561	0.3445	0.2849
	Lady	0.7216	0.8334	0.7216	0.5541	0.5719
	Nature	0.3038	0.6699	0.9542	0.2344	0.4523
	Elaine	0.3426	0.3833	0.3933	0.2792	0.2784
	Peppers	0.3251	0.6850	0.6365	0.2651	0.3003
	Air Force	0.2572	0.2391	0.2572	0.1888	0.1883
	Bird	0.5418	1.0476	0.5942	0.4817	0.5023
	Tree	0.3935	2.0958	0.4651	0.3250	0.4985
4	Man	0.0752	0.1978	0.1234	0.0648	0.0706
	House	0.1162	0.3073	0.1186	0.0895	0.1120
	Flower	0.1780	0.1893	0.1414	0.1608	0.2437
	Lady	0.5883	0.2565	0.5882	0.4044	0.4357
	Nature	0.0757	0.1740	0.3632	0.0637	0.1800
	Elaine	0.0562	0.0596	0.1843	0.0431	0.3041
	Peppers	0.2990	0.3420	0.4438	0.1023	0.1998
	Air Force	0.0748	0.1281	0.1374	0.3616	0.2552
	Bird	0.2394	0.3599	0.3280	0.1106	0.4111
	Tree	0.1230	0.3001	0.1322	0.1341	0.1311
5	Man	0.0434	0.0678	0.0844	0.0508	0.0502
	House	0.0725	0.1087	0.1012	0.0578	0.0785
	Flower	0.0708	0.1454	0.1041	0.0627	0.0592
	Lady	0.0701	0.2046	0.1379	0.0719	0.1146
	Nature	0.0701	0.0709	0.1896	0.0430	0.0886
	Elaine	0.0299	0.0589	0.0626	0.0249	0.0519
	Peppers	0.2075	0.2685	0.3072	0.1005	0.1186
	Air Force	0.0955	0.0764	0.0955	0.4470	0.2956
	Bird	0.0997	0.0920	0.1179	0.0758	0.1016
	Tree	0.1678	0.1206	0.1643	0.0954	0.1135

Table 3 : Quantitative evaluation Q(I) on segmented standard images

No of Cluster	Algorithm	Quantitative Functions		
		F(I) (*1.0e+4)	F'(I) (*1.0e+3)	Q(I) (*1.0e+4)
3	KM	0.1677	0.1729	0.4279
	FCM	0.3518	0.3623	0.9791
	MKM	0.1999	0.2063	0.5393
	2D-KM	0.1415	0.1478	0.3540
	2D-MKM	0.1675	0.1752	0.4413
4	KM	0.0908	0.0931	0.2131
	FCM	0.1426	0.1463	0.3461
	MKM	0.1147	0.1177	0.2783
	2D-KM	0.0898	0.0930	0.1898
5	2D-MKM	0.1150	0.1192	0.2656
	KM	0.0599	0.0610	0.1878
	FCM	0.0705	0.0722	0.1981
	MKM	0.0680	0.0696	0.1986
	2D-KM	0.0574	0.0592	0.1239
2D-MKM	0.0841	0.0869	0.2016	

Table 5 : Average quantitative evaluation functions on 73 standard images

To study the ability of proposed algorithms to perform on a wider scale of standard image applications, we have tested 73 standard images, and the average results are tabulated in Table 5. All quantitative analysis functions are relatively low for the proposed 2D-KM algorithm. Even though 2D-MKM does not give the best result when compared to the conventional algorithms, but it performs better than FCM and MKM. By incorporating two-dimensional Euclidean distance into KM by adding

a new parameter (i.e. spatial information of intensity median), we have managed to increase the performance of the conventional KM by 34% in average.

As for the pathology images, the results of execution time and all three functions as shown in Table 6 to Table 9 have further verified the good and comparable performance of the proposed algorithms. The 2D-KM algorithm yields the best results for all tested images, while the 2D-MKM gives better result than all conventional clustering algorithms; making both proposed algorithms surface with better overall performance. These findings suggest that the 2D-KM and 2D-MKM are able to offer better performance in segmenting pathological images for medical purposes.

No of Cluster	Image	$F(I)$ For Clustering Algorithms (*1.0e+3)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Cell1	0.0718	0.1011	0.1473	0.0480	0.0616
	Cell2	0.0581	0.0744	0.1768	0.0429	0.0551
	Cell3	0.2999	0.3061	0.4410	0.1868	0.2096
	Cell4	0.0561	0.0616	0.4806	0.0293	0.1263
	Cell5	0.0266	0.0285	0.9730	0.0134	0.3867

Table 6 : Quantitative evaluation $F(I)$ on segmented pathology image

No of Cluster	Image	$F'(I)$ For Clustering Algorithms (*1.0e+2)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Cell1	0.0779	0.1079	0.1589	0.0536	0.0686
	Cell2	0.0640	0.0805	0.1867	0.0459	0.0609
	Cell3	0.3129	0.3197	0.4576	0.2042	0.2256
	Cell4	0.0592	0.0650	0.4958	0.0322	0.1368
	Cell5	0.0287	0.0309	1.0124	0.0148	0.4155

Table 7 : Quantitative evaluation $F'(I)$ segmented pathology image

No of Cluster	Image	$Q(I)$ For Clustering Algorithms (*1.0e+3)				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Cell1	0.1904	0.2676	0.4060	0.1277	0.1661
	Cell2	0.1520	0.1932	0.4685	0.1120	0.1442
	Cell3	0.9203	0.9380	1.3846	0.5742	0.6527
	Cell4	0.1490	0.1602	1.4061	0.0779	0.3588
	Cell5	0.0664	0.0714	2.9020	0.0336	1.0854

Table 8 : Quantitative evaluation $Q(I)$ on segmented pathology image

No of Cluster	Image	Algorithms				
		KM	FCM	MKM	2D-KM	2D-MKM
3	Cell1	1.09	1.24	1.07	1.34	1.22
	Cell2	0.97	1.15	1.03	1.26	2.14
	Cell3	1.45	2.1	1.35	1.4	1.47
	Cell4	1.13	1.23	1.04	1.3	1.13
	Cell5	1.02	1.18	1.05	1.03	1.24

Table 9 : Execution time (in seconds)

5 Conclusion

In this paper, two modified versions of the conventional KM and MKM clustering algorithms have been introduced, namely the 2D-KM and 2D-MKM clustering algorithms. Both algorithms were tested against standard images and cervical cell images (i.e. as case study) qualitatively and quantitatively. From the results, it is observed that both 2D-KM and 2D-MKM perform better as compared to conventional KM, FCM, and MKM clustering algorithms. Qualitatively, the images produced by the proposed algorithms are more homogeneous and smoother. Quantitatively, the 2D-KM and 2D-MKM algorithms give lower readings of $F(I)$, $F'(I)$, and $Q(I)$, which are desired in image segmentation. Execution times of the proposed algorithms are also shorter than FCM in most cases, further adding to their advantages when compared against conventional clustering algorithms. As a conclusion, the new proposed 2D-KM and 2D-MKM clustering algorithms perform better than the conventional KM, FCM, and MKM clustering algorithms in terms of quality and which credibility further proven in their quantitative records.

References:

- [1] T.D. Pham, U. Eisenblatter, J. Golledge, B.T. Baune, K. Berger, Segmentation of medical images using geo-theoretic distance matrix in fuzzy clustering, *16th IEEE International Conference on Image Processing (ICIP)*, 2009.
- [2] Y. Seunghwan, and P. Rae-Hong, Red-eye detection and correction using inpainting in digital photographs, *IEEE Transactions on Consumer Electronics*, vol. 55, no. 3, pp. 1006-1014, 2009.
- [3] N. A. Mat-Isa, M. Y. Mashor, and N. H. Othman, Automatic seed based region growing for pap smear image segmentation, in *Kuala Lumpur International Conference on Biomedical Engineering*. 2002.
- [4] C. Ordenez, Clustering binary data streams with K-means, *8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery*, pp. 12-19, 2003.
- [5] Castro, R.M.; Coates, M.J.; Nowak, R.D.; Likelihood based hierarchical clustering, *IEEE Transactions on Signal Processing*, vol.52, no.8, pp. 2308- 2321, 2004.

- [6] J. K. Paik, Y. C. Park, and S. W. Park, An edge detection approach to digital image stabilization based on tri-state adaptive linear neurons, *IEEE Transactions on Consumer Electronics*, vol. 37, no. 3, pp. 521-530, 1991.
- [7] Y. Siyoung, K. Donghyung, and J. Jechang, Fine edge-preserving deinterlacing algorithm for progressive display, *IEEE Transactions on Consumer Electronics*, vol. 53, no. 3, pp. 1654-1662, 2009.
- [8] M. Lalonde, M. Beaulieu, and L. Gagnon, Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching, *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp.1193-1200, 2001.
- [9] S. K. Warfield, K. Michael, F. A. Jolesz, and K. Ron, Adaptive, template moderated, spatially varying statistical classification, *Medical Image Analysis*, vol. 4, no. 1, pp. 43-55, 2000.
- [10] M. Cheriet, J. N. Said, and C. Y. Suen, A recursive thresholding technique for image segmentation, *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 918-921, 1998.
- [11] S. Arora, J. Acharya, A. Verma, and P. K. Panigrahi, Multilevel thresholding for image segmentation through a fast statistical recursive algorithm, *Pattern Recognition Letters*, vol. 29, pp. 119-125, 2007.
- [12] P. C. Boutros, and A. B. Okey, Unsupervised pattern recognition: An introduction to the whys and wherefores of clustering microarray data, *Briefings In Bioinformatics*, vol. 6, no. 4, pp. 331-343, 2005.
- [13] J. P. Jesan, The neural approach to pattern recognition, *Ubiquity*, vol.5, no.7, 2004.
- [14] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, An efficient k-means clustering algorithm: analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881-892, 2002.
- [15] 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.
- [16] A. Y. Al-Omary, and M. S. Jamil, A new approach of clustering based machine-learning algorithm, *Knowledge-Based Systems*, vol.19, no.4, pp. 248-258, 2006.
- [17] X. Zhang; X. Qian; L. Jiao; G. Wang; , An immune spectral clustering algorithm, *International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS*, pp.296-299, 2007.
- [18] M. Tabakov, A fuzzy clustering technique for medical image segmentation, *International Symposium on Evolving Fuzzy System*, Nov. 2006.
- [19] M. H. Mohamed, and M. M. Abdel Samea, An efficient clustering based texture feature extraction for medical image, *11th International Conference on Computer and Information Technology ICCIT*, pp. 88, 2009.
- [20] K. Z. Mao, P. Zhao, and P.-H. Tan, Supervised learning-based cell image segmentation for p53 immunohistochemistry, *IEEE Transactions Biomedical Eng.*, vol. 53, no. 6, pp. 1153-1163, 2006.
- [21] M. Li, T. Huang, and G. Zhu, Improved fast fuzzy c-means algorithm for medical MR images segmentation, *Second International Conference on Genetic and Evolutionary Computing WGECC '08.* , vol., no., pp.285-288, 25-26, 2008.
- [22] N. K. Verma, A. Roy, and S. Vasikarla, Medical image segmentation using improved mountain clustering technique version-2, *7th International Conference on Information Technology: New Generations (ITNG)*, pp. 156-161, 2010.
- [23] M. Y. Mashor, Hybrid training algorithm for RBF network, *International Journal of The Computer, The Internet and Management*, vol. 8, no. 2, pp. 50-65, 2000.
- [24] P. K. Agarwal, and N. H. Mustafa, K-means projective clustering, *Proceedings of the twenty-third ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2004.
- [25] S. N. Sulaiman, and N. A. Mat Isa, Adaptive fuzzy k-means clustering algorithm for image segmentation, *IEEE Transactions on Consumer Electronics*, vol. 56, issue 4, pp. 2661, 2011.
- [26] A.S. Abutaleb, Automatic thresholding of grey-level pictures using two- dimensional entropy, *Computer Vision Graphics Image Processing* pp. 22-32, 1989.
- [27] D. Brink, Thresholding of digital images using two dimensional entropies, *Pattern Recognition* 25, 803-808 (1992).

- [28] L. Li, R. Gong, and W. Chen Gray level image thresholding based on fisher linear projection of two-dimensional histogram, *Pattern Recognition*, vol. 30, pp. 743-749, 1997.
- [29] S. Jianzhong, T. Jinwen, L. Jianguo, S. Zailong, Y. Zhiqiang, C. Xiaohua, Z. Renli, and Y. Chang. Fast algorithm for 2D entropic thresholding of image segmentation, *Processing SPIE.*, pp. 328- 333, 2001.
- [30] J. Xie, T. Liu, J. Wang, and Y. He, A fast two-dimensional entropic thresholding algorithm, *International Conference on Information and Automation, ICIA*, pp.275-278, 20-23, 2008.
- [31] L. P. Hu, H. W. Liu, K. Y. Yin, and S. J. Wu, Two-dimensional clustering-based discriminant analysis for SAR ATR, . *8th International Symposium on Antennas, Propagation and EM Theory, ISAPE*, pp.509-513, 2-5 Nov. 2008.
- [32] S. Chen, and D. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol.34, no.4, pp.1907-1916, 2004.
- [33] K.S. Chuang, H.L. Tzeng, S. Chen, J. Wu and T.J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, *Computer Medical Imaging Graphics* , pp. 9-15, 2006.
- [34] W.L. Cai, S.C. Chen and D.Q. Zhang, Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation, *Pattern Recognition*, pp. 825-838, 2007.
- [35] M. Y. Mashor, Improving the performance of k-means clustering algorithm to position the centres of RBF network, *International Journal of Computer, Internet and Management*. vol.6, no.2, 1998.
- [36] M. J. A. Berry, and G. Linoff, *Data Mining Techniques For marketing, Sales and Customer Support*, John Willey & Sons, Inc. 1996.
- [37] J. Liu, and Y. H. Yang, Multiresolution color image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.16, no.7, pp.689-700, Jul 1994.
- [38] M. Borsotti, P. Campadelli, and R. Schettini, Quantitative evaluation of color image segmentation results, *Pattern Recognition Letters*, vol. 19, no. 8, pp. 741-747, 1998.
- [39] R. H. Haralick, and L. G. Shapiro, Image segmentations techniques, *Computer Vision Graphics Image Processing* 29, pp. 100-132, 1985.
- [40] K. S. Tan, and N. A. Mat Isa, Color image segmentation using histogram thresholding – Fuzzy C-means hybrid approach, *Pattern Recognition*, vol.44, pp. 1-15, 2011.
- [41] A. W. Liew, and H. Yan, An adaptive fuzzy clustering algorithm for medical image segmentation, *International Workshop on Medical Imaging and Augmented Reality*, pp. 272, 2002.