International Journal of Innovative Computing, Information and Control
Volume **13**, Number **5**, October **2017**
ICIC International ©2017 ISSN 1349-4198
pp. 1509–1525

AN IMPROVED *K*-MEANS CLUSTERING ALGORITHM BASED ON AN ADAPTIVE INITIAL PARAMETER ESTIMATION PROCEDURE FOR IMAGE SEGMENTATION

Zubair Khan[1], Jianjun Ni[1,2,*], Xinnan Fan[1,2] and Pengfei Shi[1]

[1]College of IOT Engineering
[2]Jiangsu Universities and Colleges Key Laboratory of Special Robot Technology
Hohai University
No. 200, Jinling North Road, Changzhou 213022, P. R. China
engr.zubairkhan13@gmail.com; fanxn519@163.com; flyshn@hotmail.com
*Corresponding author: njjhhuc@gmail.com

Received March 2017; revised July 2017

Abstract. *Image segmentation is of great importance in the field of image processing. K-means clustering algorithm is widely used in image segmentation because of its computational simplicity. However, the clustering results obtained from K-means heavily depend upon the initial parameters. Mostly, these initial parameters are selected through hit and trial rule, which leads to inconsistency in the image segmentation results. In this paper, an improved K-means clustering algorithm is proposed for image segmentation, employing an adaptive histogram based initial parameter estimation procedure. Being a histogram based technique, the proposed approach is computationally inexpensive and is able to deal with the large size images effectively, which suits the real-time image segmentation scenario very well. Additionally, the proposed algorithm requires less user interaction to determine K-means initialization parameters. Some experiments are conducted based on various grey images to test the proposed approach. The experiment results show that the proposed approach can improve the K-means based image segmentation results. Finally, image segmentation results obtained through the proposed method are objectively compared with other membership based clustering algorithms such as fuzzy C-means and K-medoids. The results prove that the overall performance of the proposed K-means method is better than other membership based clustering algorithms for image segmentation.*

Keywords: K-means clustering, Image segmentation, Fuzzy C-means, Image processing

1. **Introduction.** Image segmentation is applied to clustering pixels into salient image regions, such as regions corresponding to objects of the image. These regions can be considered homogeneous according to some characteristics, such as grey value, motion, texture, and shape [1, 2, 3]. Image segmentation is an important initial step in image processing and is used in various applications like medical imaging, locating objects in satellite images, face recognition, traffic control systems, and machine vision [4, 5, 6, 7].

Image segmentation can be achieved by using a lot of techniques, which can be broadly classified into region [8, 9] and edge based techniques [10, 11]. Edge based techniques represent the pixels in the image having sharp intensity variations. Sometimes there can be broken incomplete edges and there is not a clear separation between the objects in the image. On the other hand, region based techniques are based upon the collection of pixels having similar features, which can be grey level intensity information or color image components.

Recently, more and more researchers focus on the region based techniques using clustering methods, because of its simplicity [12, 13]. Clustering is one of the important and typically used techniques and is widely used in engineering fields such as pattern recognition, image processing, system modeling, and data mining [14, 15, 16, 17]. In clustering techniques, the prior knowledge of the initialization parameter estimate, i.e., number of clusters and the respective initial cluster centers is required, which is mostly input by the user or randomly chosen from the input data [18]. The effect of these parameters upon the final image segmentation is enormous, and variation of these parameters produces fluctuations in the final results.

One of the most well-known and fastest clustering technique is the K-means technique [19, 20]. K-means is a hard clustering algorithm, in which each pixel only belongs to a single cluster center. K-means re-computes each of the new centers by averaging the pixel intensities assigned to the old cluster centers. In this paper, we focus on the region based image segmentation using K-means clustering methods because of its simplicity and high efficiency. However, there are some shortcomings of the K-means based methods which should be solved, for example, it is sensitive to outliers and also its greedy nature makes it sensitive to initialization parameters [21].

To reduce the effect of outliers upon the cluster center selection procedure, some improvements have been done in the clustering algorithm and applied in image segmentation. For example, Bezdek [22] proposed a fuzzy C-means (FCM), a fuzzy version of K-means clustering algorithm, which is a membership based soft clustering algorithm. Benaichouche et al. [1] introduced a region based image segmentation method based on an improved spatial fuzzy C-means clustering (FCM). Zhang et al. [23] employed a possibilistic C-means (PCM) clustering algorithm based image segmentation method, by replacing Euclidean distance with Mahalanobis distance. Even though the segmentation results obtained by those methods above are good enough, the disadvantage is that the final segmentation results of FCM based algorithms provide better results only when the initial parameters are close to the final solution. In short, results of fuzzy clustering depend highly upon the initialization parameters and also it is computationally expensive as compared to standard K-means.

To reduce the effect of initialization parameters upon the final clustering results, many new clustering algorithms have been proposed. For example, Zhang et al. [24] proposed K-harmonic mean (KHM) clustering algorithm, in which harmonic averages are used as the performance function instead of the mean, to select a cluster center. Gui et al. [25] used KHM clustering algorithm for spectral clustering, to perform color image segmentation. Frackiewicz and Palus [26] used KHM as a clustering technique, to perform endoscopic image segmentation. Even though KHM based methods provide good clustering results, KHM clustering technique is computationally expensive as compared to K-means and also in case of proper initialization parameters K-means converges faster than KHM.

To deal with the problem of proper initialization parameters for clustering algorithms, several techniques have been devised. For example, Khan and Ahmad [27] devised a technique to find the initial cluster centers based upon the membership variations of the data objects for various data set attributes. Park and Jun [28] proposed an efficient PAM-based algorithm, in which the parameter initialization procedure is based upon the distance calculation between each pair of objects and then selected the required number of objects as cluster centers having the least average distance. Dhanachandra et al. [29] used subtractive clustering approach to determine one initial parameter, i.e., initial cluster centers. After that, these initial estimates were provided to the K-means clustering algorithm to perform image segmentation. Even though these techniques determined the cluster centers, the determination of cluster quantity remained an issue. Moreover, those above mentioned algorithms are computationally complex and the experiments are based upon small data sets.

Even though a lot of clustering algorithms and techniques have been devised to eliminate the shortcomings of K-means, and to determine initial estimates for clustering algorithms, some other problems are introduced into the K-means based image segmentation. For example, Goryawala et al. [30] proposed an improved K-means algorithm for the liver segmentation which benefited from prior knowledge regarding the liver image. Kurumalla and Rao [31] proposed an improved K-means algorithm enhancing the processing speed of K-means algorithm, which is based on reduction in the size of data objects. In addition, the main issue still remains that is to devise an efficient procedure, both in terms of computation complexity and segmentation quality to obtain consistent results for clustering type region based image segmentation approaches. So, to make the region based image segmentation using clustering techniques reliable and accurate, this paper focuses upon the clustering algorithms initialization parameters for image segmentation and devises a procedure to automatically determine two important initialization parameters for the clustering algorithms: 1) number of clusters; 2) initial cluster centers. For the clustering of pixels, K-means is chosen in this study.

The main contributions of this paper are summarized as follows: 1) An adaptive histogram based approach to determine the initial parameters for K-means is proposed; 2) A 2-step initial parameter estimation procedure to choose proper amount of clusters and optimal initial cluster centers is presented; 3) A computationally economical initial parameter estimation procedure, being computationally independent of the size of image to be segmented is given out; 4) Comparison of initial parameter estimation procedure of the proposed approach with a histogram based method and density based subtractive clustering approach is carried out; and the subjective and objective evaluation of final image segmentation results with histogram based approach and subtractive clustering based K-means approach is also given out; 5) The final image segmentation results obtained by our improved K-means (IKM) algorithm and from other membership based soft clustering algorithms such as FCM and K-medoids are discussed.

This paper is organized as follows. Section 2 presents the proposed improved K-means clustering algorithm based image segmentation approach. The image segmentation experiments for various image data sets are given in Section 3. Section 4 discusses the performance of the proposed approach. Finally, the conclusions are given in Section 5.

2. **The Proposed Approach.**

In this paper, the problem of initialization parameters for K-means clustering algorithm focuses upon to benefit from the low computational cost of K-means, and to obtain reliable image segmentation results. K-means clustering, being an unsupervised image segmentation technique, has no prior knowledge about the type of image to be segmented. So, arbitrary selection of these initialization parameters affects the accuracy of the final image segmentation results based on the clustering methods. To deal with this problem, a histogram based adaptive initialization parameter estimation procedure is proposed in this study. Being a histogram based approach, it has fast processing speed and suits the image segmentation scenario very well. The overall computation time of the K-means based image segmentation approach consists of two sections: 1) the computation time required for the initialization parameters determination procedure; 2) the computation time required for K-means to produce clustering results. The purpose of the proposed algorithm is to obtain an enhanced image segmentation quality while keeping the computation cost minimum.

Histogram of a grey image is the graphical representation of (1-D) pixel intensities in the image. A sample grey image and its histogram are shown in Figure 1. In the grey

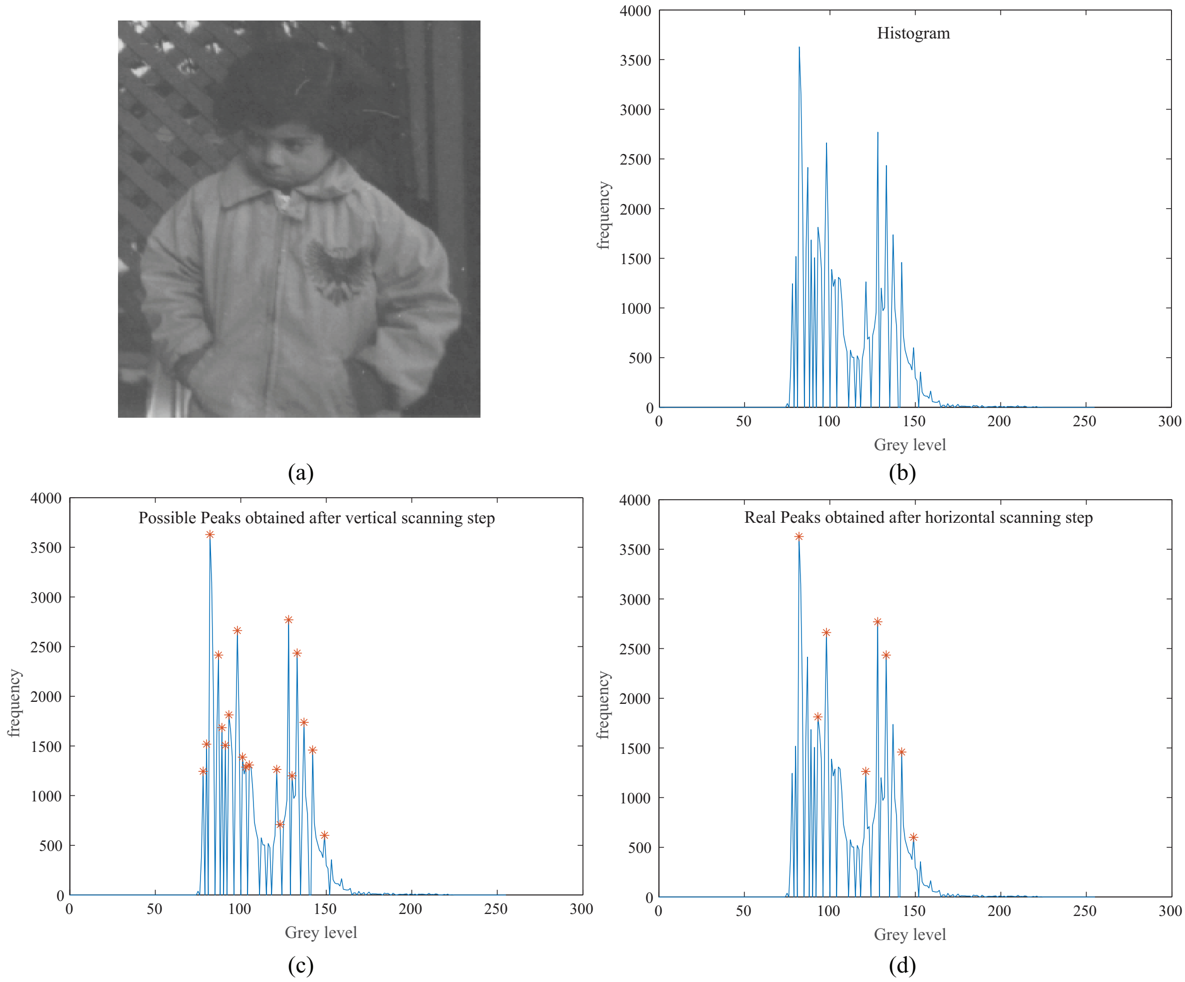

FIGURE 1. A sample grey image and its histogram: (a) the grey image; (b) the histogram of the grey image; (c) possible peaks obtained after the vertical scanning of histogram; (d) real peaks obtained after the horizontal scanning of histogram

image histogram (see Figure 1(b)), the x-axis represents grey levels and y-axis represents the frequency of each grey level occurrence in the image. The purpose of the proposed algorithm is to make use of the continuous and smoothed representation of the histogram of an image to automatically find the number of clusters and their respective cluster centers. In the proposed approach, the initialization parameters obtained from image histogram will be imputed to the K-means algorithm, which assigns each pixel to a nearest cluster center minimizing the sum of squared error (SSE). So all the intensity levels of the pixels lying within the cluster will be replaced by the intensity level of the respective cluster center. Then the final segmented image will only contain the grey levels of the cluster centers instead of each pixel having its own distinct intensity value. The work flow of the initial parameter estimation procedure for clustering algorithms is shown in Figure 2, and will be introduced as follows in detail.

2.1. **Initialization parameters estimation procedure.** In the grey image, a single random variable known as grey level, is used to assign intensity values to the pixels. In case of grey image, the grey level l_i is defined as $\{l_i \in \mathrm{Z}_+ \cup 0 | 0 \leq l_i \leq 255\}$. Each grey level has its own occurrence level in the grey image which is actually the histogram function value or the density of that grey level. Total number of required bins k can be computed

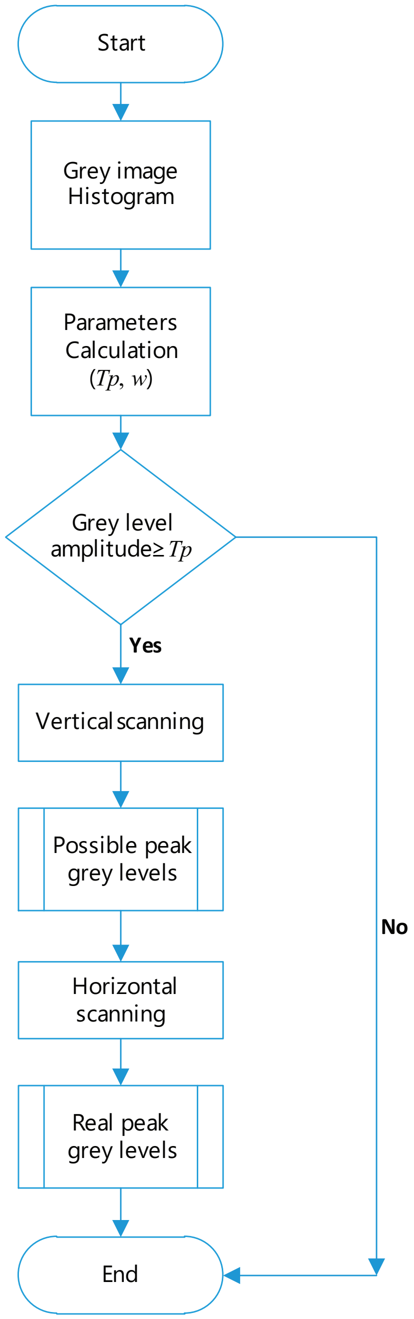

FIGURE 2. Flow diagram of the proposed approach to estimate initialization parameters

as

$$k = \left\lceil \frac{\max greylevel - \min greylevel}{h} \right\rceil \tag{1}$$

where $\lceil \cdot \rceil$ indicates the ceiling function; h denotes the bin width. In case of image histogram, the histogram of an 8-bit grey image contains 256 bins, indexed from 0 to 255. Grey image contains intensity levels having a minimum and maximum intensity value of 0 and 255 respectively. So the bin width h is selected to be 1 and maximum value of grey level is 255, and then the total number of bins k becomes 256.

In the proposed approach, there are two steps of histogram scanning. The first step is the vertical scanning step, where all the grey levels having an amplitude a_{l_i} greater than the threshold amplitude T_p are processed to obtain possible peak grey levels l_p. Possible peak grey levels l_p have an amplitude a_{l_p}, which is greater than their immediate neighboring grey levels, i.e., $a_{l_{i+1}}$ and $a_{l_{i-1}}$. The next step is the horizontal scanning step, where only the possible peak grey levels selected in the vertical scanning are processed to eliminate the local maxima peaks in the grey levels range w, obtaining real peak grey levels. The quantity and the grey level value of real peaks serve as the number of clusters and the initial cluster centers respectively for the K-means method.

Before proceeding to the histogram scanning procedure, some parameters are determined using the image histogram. Amplitude threshold T_p for an image histogram distribution is calculated, by counting the number of grey levels having an amplitude greater than one, i.e., $a_{l_i} \geq 1$, and excluding the grey levels having zero amplitude, i.e., $a_{l_i} = 0$, representing no information. The expressions to determine the threshold amplitude T_p are given below.

$$l_0 = \left(a_{l_i} \geq 1\right)_{i=0}^{255} \tag{2}$$

$$S_t = \sum_{i \in l_0} a_{l_i} \tag{3}$$

$$T_p = \left(\frac{S_t}{l_0}\right) \tag{4}$$

where l_0 denotes the number of levels having amplitude greater than or equal to 1; S_t denotes the summation of the occurrences of grey levels l_0.

An adaptive window width w applied in horizontal scanning step will be automatically determined for each grey image using the image histogram information. Adaptive window width w is determined as follows

$$l_t = (a_{l_i} \geq T_p)_{i=0}^{255} \tag{5}$$

$$w = \left\lceil \frac{l_t}{\alpha} \right\rceil \tag{6}$$

where l_t represents the total number of grey levels in the histogram having an amplitude greater than or equal to the threshold amplitude T_p; α is a constant $\{\alpha \in \mathrm{Z}_+ | 1 \leq \alpha \leq l_t\}$ whose value can be selected within the range $[1, l_t]$; in the proposed approach α is set to a value 10 for all the grey images simulation experiments.

The pseudo code of the proposed algorithm for automatic parameter estimation procedure is shown in Figure 3, and the two steps of histogram scanning will be introduced as follows.

2.2. **Vertical scanning of histogram.** In the vertical scanning step, only l_t grey levels will be processed for the selection of a possible peak, having an amplitude greater than threshold amplitude T_p. The rest of grey levels, i.e., $a_{l_i} < T_p$ will be discarded, decreasing the maximum number of grey levels from $255 \rightarrow l_t$. For a grey level l_i to be chosen as a possible peak grey level l_p, it must obey the condition

$$a_{l_{t-1}} \leq a_{l_t} \geq a_{l_{t+1}} \tag{7}$$

where a_{l_t} denotes the amplitude at grey level l_t, $\forall \{l_t | l_t > T_p\}$.

Possible peak amplitude obtained after the vertical scanning step is denoted as a_{l_p}. After the selection of the possible peak grey levels l_p, it is quite possible that some of the possible peaks selected are having grey levels, very close to each other, i.e., $(l_{p_i} \approx l_{p_{i+1}})$. To solve this issue a histogram horizontal scanning step is devised. The possible peaks and the respective grey levels obtained after vertical scanning step for the sample image in Figure 1(a) are shown in Figure 1(c).

Algorithm 1: Automatic initial parameter estimation for clustering algorithms

//Get the histogram of Image

Image = readImg ( );

Hist = histogram (Image);

// Calculation of Histogram Scanning Parameters

T_p = Calculate (*Hist*);

// Determine threshold Amplitude T_p;

$w = \left\lceil \frac{l_t}{\alpha} \right\rceil$;

// Determine Window size w*, where* α *is 10;*

// Vertical scanning step

For $l_i > l_t$

If $a_{l_{i-1}} \leq a_{l_i} \geq a_{l_{i+1}}$

Possible peak amplitude = $\{a_{l_p}\}$;

Possible peak grey level = $\{l_p\}$;

// Mark that particular grey levels as a possible peak grey levels.

End If

End For

// Horizontal Scanning step

For $l_i \in l_p$

If $a_{l_p} \geq a_{l_i}, \quad \forall a_{l_i} \in r$

Real peak grey level $(l_r) = l_p$;

End If

End For

// Calculate the initial parameter

Cluster centers $(C_i) = l_r$;

Number of clusters (N) *= length* (l_r).

FIGURE 3. The pseudo code of the proposed approach

2.3. **Horizontal scanning of histogram.** Possible peak grey levels l_p selected after the histogram vertical scanning step can contain grey level that are close to each other, causing the problem of over segmentation in the final segmented image. So, in the histogram horizontal scanning step, actual peak grey levels are determined by the scrutiny of the possible peak grey levels, obtained after the histogram vertical scanning step. In the histogram horizontal scanning, scrutiny is done by comparing possible peak grey level amplitude a_{l_p} with each grey level amplitude a_{l_i} lying in range defined by w around the possible peak grey level l_p. In case of having no higher amplitude grey level peak, the possible peak grey level l_p is selected as the real peak grey level l_r. Real peak grey levels l_r chosen in this step will serve as the initialization parameters for K-means clustering algorithm.

For a possible peak grey level l_p to be selected as a real peak grey level l_r, it must satisfy the following condition:

$$l_r = (a_{l_p} \geq a_{l_i}), \quad \forall a_{l_i} \in r \tag{8}$$

where a_{l_i} is the amplitude at a grey level which belongs to the range r of grey levels, defined by window w.

The real peak grey levels l_r obtained after the horizontal scanning step for the sample image in Figure 1(a) are shown in Figure 1(d). After processing the histogram through two subsequent histogram scanning procedures, the grey levels l_r having the highest peaks and containing abundant information in the grey image provide the initial parameters. These parameters namely cluster quantity and initial cluster centers are provided to the K-means algorithm to output region based image segmentation results.

In short, the proposed algorithm efficiently retrieves the initialization parameters namely, number of clusters and initial cluster centers, maintaining the manual user interaction to a least level. These initialization parameters are regraded as highly critical to the performance of K-means method. The flow diagram of the K-means algorithm having the initialization parameters determined by the proposed histogram based method is shown in Figure 4.

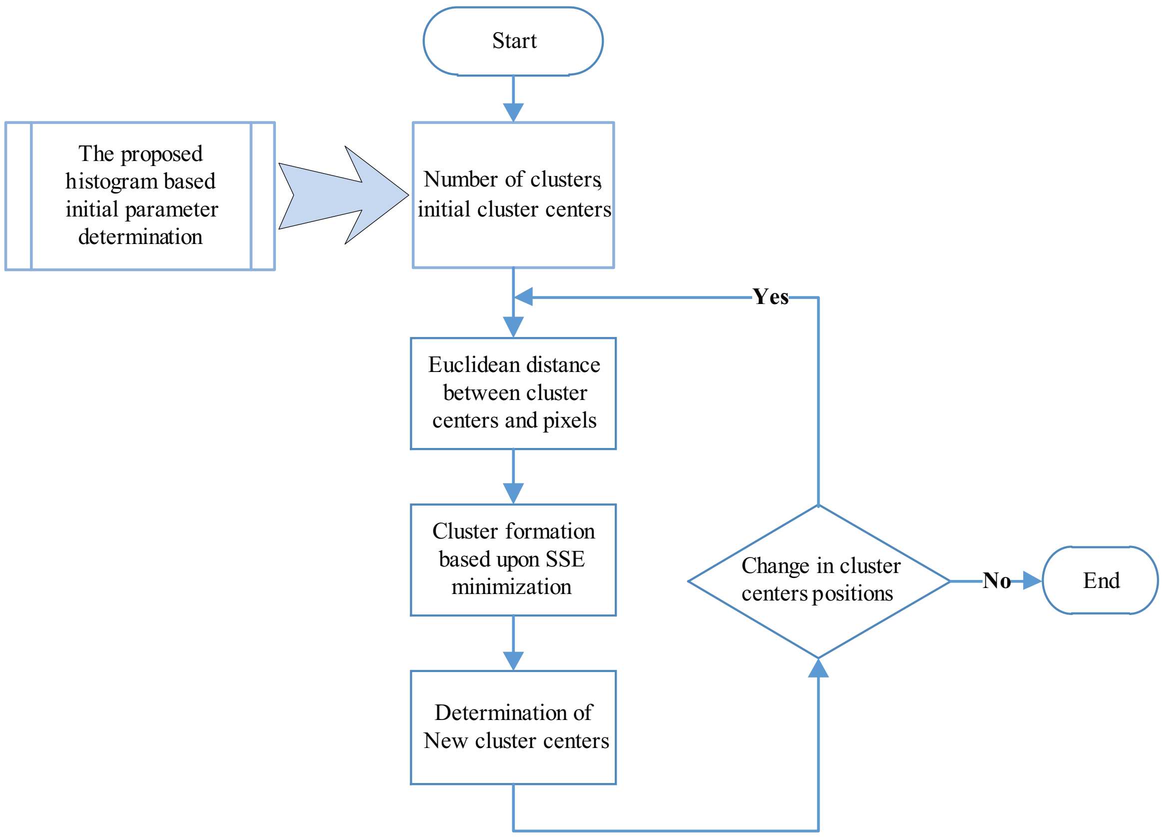

FIGURE 4. The flow diagram of K-means algorithm having the initialization parameters determined by the proposed histogram based method

In the next step to produce segmented image, K-means methods assign each pixel to a nearest cluster center minimizing the SSE (sum of squared error). In the subsequent step, new cluster centers determination is done by computing the mean of all the pixels within a cluster. After that, all the pixels belonging to a cluster are replaced by the intensity level of the cluster centers. This procedure continues until the threshold, i.e., the number of iterations is reached or the difference between the cluster center updates becomes less than the threshold value. So, in the final segmented image, individual pixel intensities are replaced by the cluster center intensity value of the respective cluster. The resulting image only contains regions instead of pixels and also retains the abundant information present in the image, which is used by other high level image processing applications, i.e., computer vision.

3. **Experiments.** To test the performance of the proposed approach, some experiments are carried out on grey images, which are retrieved from MATLAB media or downloaded from Berkeley image segmentation database [32], and used in the following sequence {(mri.tif denoted as MR), (bag.png denoted as BA), (cameraman.tif denoted as CA), (coins.png denoted as CO), (moon.tif denoted as MO), (pout.tif denoted as PO), (glass.png denoted as GL), (AT3_1m4_01.tif denoted as AT)} (see Figure 5(a) and Figure 6(a)), where the sizes of each image is adjusted to make the figure more regular and the actual sizes of these eight images are listed in Table 1.

In this section, experimental results of the proposed improved K-means (IKM) are described and compared with another histogram based method (HBA) [33] and the subtractive clustering method (SC) [29]. All the experiments were implemented in MATLAB version 8.5 and run on a 2.60 GHz Intel(R) Quad core CPU, under Microsoft Windows 10 Operating system. The parameter α is set to a value 10 for all the experiments in this paper.

To objectively compare the segmented image quality, Q value criterion function is used in this study [34]. Q value automatically estimates the image segmentation quality taking consideration of both the small and large regions in the final segmented images. The Q value evaluation function used in this paper is given below:

$$Q(I) = \frac{1}{10000(N \times M)} \sqrt{R} \times \sum_{i=1}^{R} \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{R(A_i)}{A_i} \right)^2 \right] \tag{9}$$

where N and M are the rows and columns in the image, R is the total number of regions in the segmented image, e_i is the color difference between original and segmented image, A_i is the area of the region i and $R(A)$ is the number of regions having the same area. e_i is determined by calculating the Euclidean distance between each pixel in original image and segmented image. The area of each region A_i is determined by calculating the total number of pixels constituting that region. Smaller values of Q represent good segmentation results whereas higher values indicate higher color error (under segmentation) or over segmentation and thus point toward inferior segmentation results.

The final results of the image segmentation based on the SC, the HBA and the proposed IKM method are shown in Figure 5 and Figure 6. The comparison of number of clusters, computation time and the Q value of the SC, the HBA and the proposed IKM method is listed in Table 1.

The results in Table 1 show that because SC method took account of each and every pixel for the determination of initial estimates for K-means clustering algorithm, its computational complexity grows in a linear way with the increasing image size. Even though the segmentation results produced by SC based K-means are good enough subjectively and objectively, the estimate of cluster number is needed to be imputed by the user.

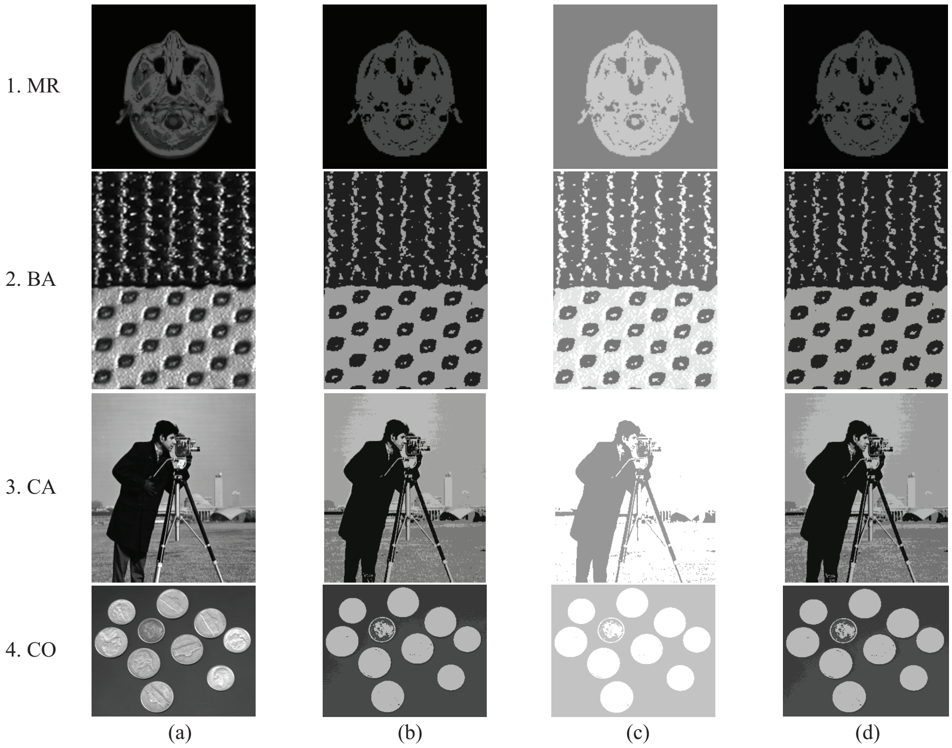

FIGURE 5. The final image segmentation results for image #1-#4: a) the original image; b) based on the SC method; c) based on the HBA method; d) based on the IKM method

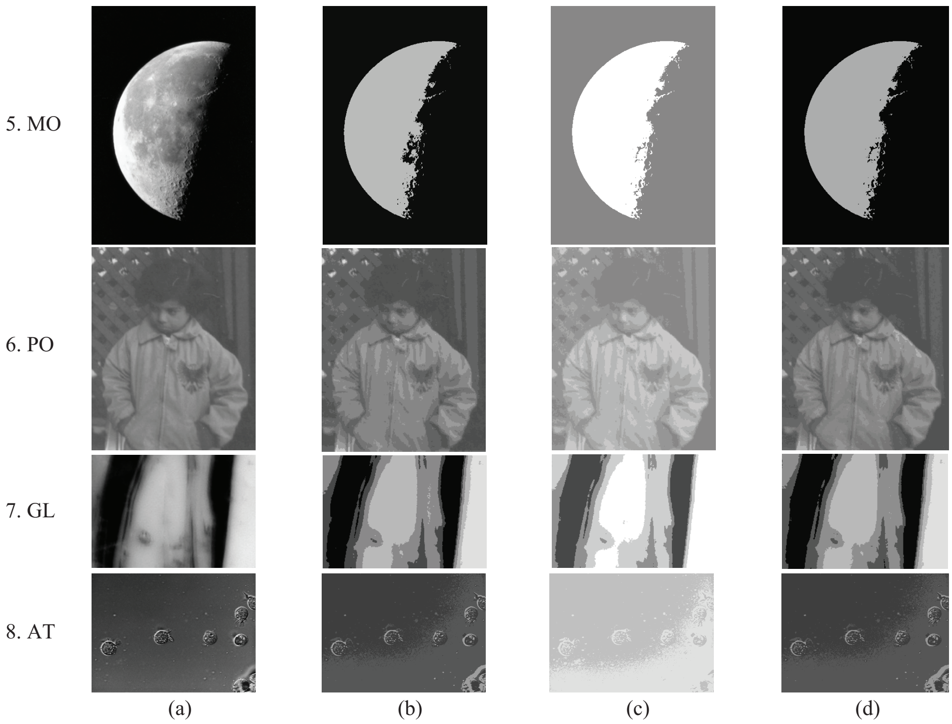

FIGURE 6. The final image segmentation results for image #5-#8: a) the original image; b) based on the SC method; c) based on the HBA method; d) based on the IKM method

Table 1. The image segmentation results based on three parameter initialization methods

Samples		Number of clusters			Q-value of segmentation			Computational time (s)		
Name	Size	SC	HBA	IKM	SC	HBA	IKM	SC	HBA	IKM
1. MR	128*128	2	2	2	0.001	0.34	0.001	2.54	0.23	0.19
2. BA	250*189	2	3	2	0.02	0.43	0.02	17.36	0.62	0.46
3. CA	256*256	4	2	4	0.005	1.23	0.004	31.98	0.68	0.81
4. CO	246*300	3	2	3	0.006	1.31	0.006	39.84	0.70	0.81
5. MO	537*358	2	2	2	0.06	3.67	0.06	358.78	1.79	1.66
6. PO	291*240	8	5	8	0.0009	0.12	0.0003	37.44	1.14	1.29
7. GL	181*282	5	4	5	0.004	0.20	0.005	20.53	0.69	0.76
8. AT	480*640	3	2	3	0.01	4.47	0.01	1181.10	2.77	16.30

Generally, the number of clusters used for SC method is the same as determined by the proposed algorithm as can be seen in Table 1. However, the high computation cost of SC makes it less feasible for real-time image processing scenario. On the other hand, HBA based method, even though being computationally inexpensive as compared to SC, encountered the problem of determining less number of clusters which resulted in loss of important information, hence reducing the image segmentation quality as can be observed in the case of sample image #8 (AT3_1m4_01.tif) in Figure 6. The final segmented image results in Figure 5, Figure 6 and the objective evaluation in Table 1 indicate that the performance of our improved K-means is better as compared to other two approaches.

Remark 3.1. *The results in Table 1 show that the computation time of the proposed method may be longer than the HBA based method for some certain image segmentation tasks; however, the proposed method definitely produces far better segmented images, as represented by the Q value. The performance to balance the computation time and the image segmentation results is very important for the image processing approach.*

4. **Discussions.** The results of the image segmentation experiments in Section 3 show that the proposed approach can perform grey image segmentation task efficiently while automatically determining both the initialization parameters properly. Some performance aspects of the proposed approach are discussed in this section.

To discuss the performance of the proposed approach in image segmentation for other grey images, some further experiments are conducted on some commonly used images ({(Lena.png denoted as LE), (Valley.jpg denoted as VA), (Airplane.jpg denoted as AI), (Mountain.jpg denoted as MT)}). The final image segmentation results based on the proposed IKM method are shown in Figure 7 and Table 2. The results obtained are satisfactory as the number of regions determined by our automatic initialization parameter estimation procedure, segmented the image capturing the majority of information in the image and also preserving the boundaries between different regions in the final segmented image.

The initialization parameters obtained by the proposed approach in these four images are compared with those of the HBA and SC methods (see Figure 8 and Figure 9). In each approach cluster centers are presented in the order of retrieval. SC method determined the initial cluster centers, in the order of the grey levels having the least average Euclidean distance to other pixel intensities in the grey image. HBA based approach determined cluster centers in the ascending order of grey levels. For the initialization parameter determination more priority is given to grey level peaks having a large span whereas, grey

FIGURE 7. The final image segmentation results: a) the original image; b) the image segmentation results based on the IKM method

TABLE 2. The image segmentation results based on three methods

Samples		Number of clusters			Q-value of segmentation			Computational time (s)		
Name	Size	SC	HBA	IKM	SC	HBA	IKM	SC	HBA	IKM
9. LE	512*512	4	6	4	0.02	0.40	0.01	731.18	3.58	3.06
10. VA	321*481	4	3	4	0.03	1.16	0.02	229.85	1.80	1.85
11. AI	321*481	2	2	2	0.03	3.69	0.03	230.24	7.37	6.68
12. MT	321*481	4	4	4	0.01	0.50	0.02	225.83	1.92	1.86

level peaks having more occurrence in the grey image are ignored. This procedure sometimes results in the determination of less number of clusters, which causes loss of salient information in the image (see the case of sample image #10 (Valley.jpg) in Figure 8(b)). Also, it presented the problem of selecting cluster centers very close to each other, as can be seen in case of sample image #9 (Lena.jpg) in Figure 8(b), whereas our proposed algorithm determined proper number of clusters which accurately segmented the image. Moreover, the grey levels determined as initial cluster centers by our proposed algorithm have higher occurrence in the grey image. It can be figured out that the initialization parameters determined by our proposed approach are better, as the proposed IKM technique provides proper number of clusters and better initial cluster centers, representing abundant information in the image.

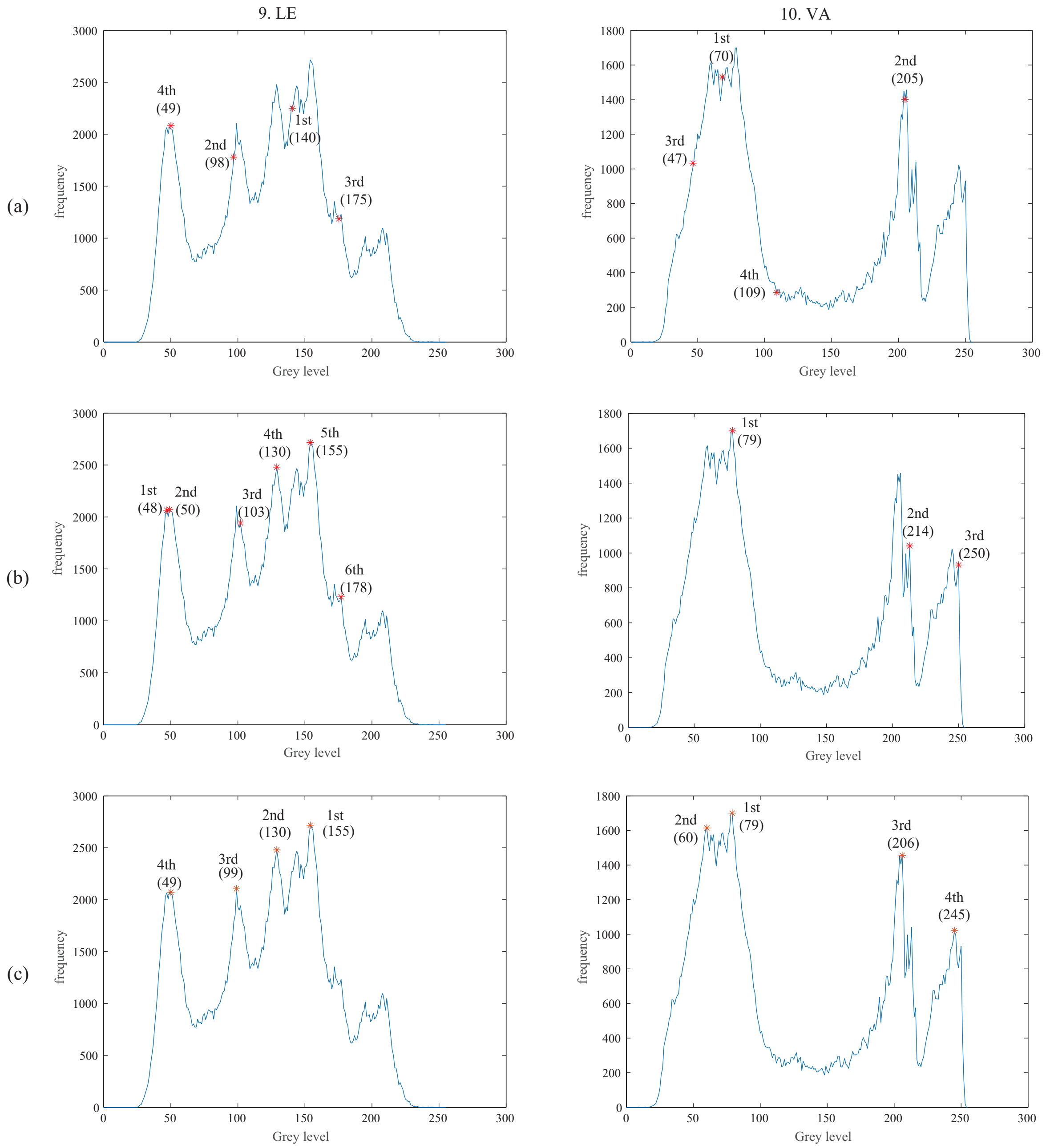

FIGURE 8. The initial cluster centers and cluster quantity for image #9-#10: (a) obtained by SC method; (b) obtained by HBA method; (c) obtained by the IKM method

To further discuss the performance of the proposed improved K-means method with other membership based soft clustering algorithms, i.e., FCM [35] and K-medoids [36], some experiments are conducted. All the 12 images used in this paper are employed to perform this experiment and the results are listed in Table 3. For the proper performance evaluation, initialization parameters provided to each clustering algorithm are the same which are determined by our proposed IKM method. K-medoids algorithm has shown the highest overall computation cost, even though in some cases it shows better segmentation quality. From the data analysis, it can be observed that the computation cost of K-medoids becomes expensive when an image is segmented into less number of clusters. This aspect can be observed in the case of sample image #5 (moon.tif), which

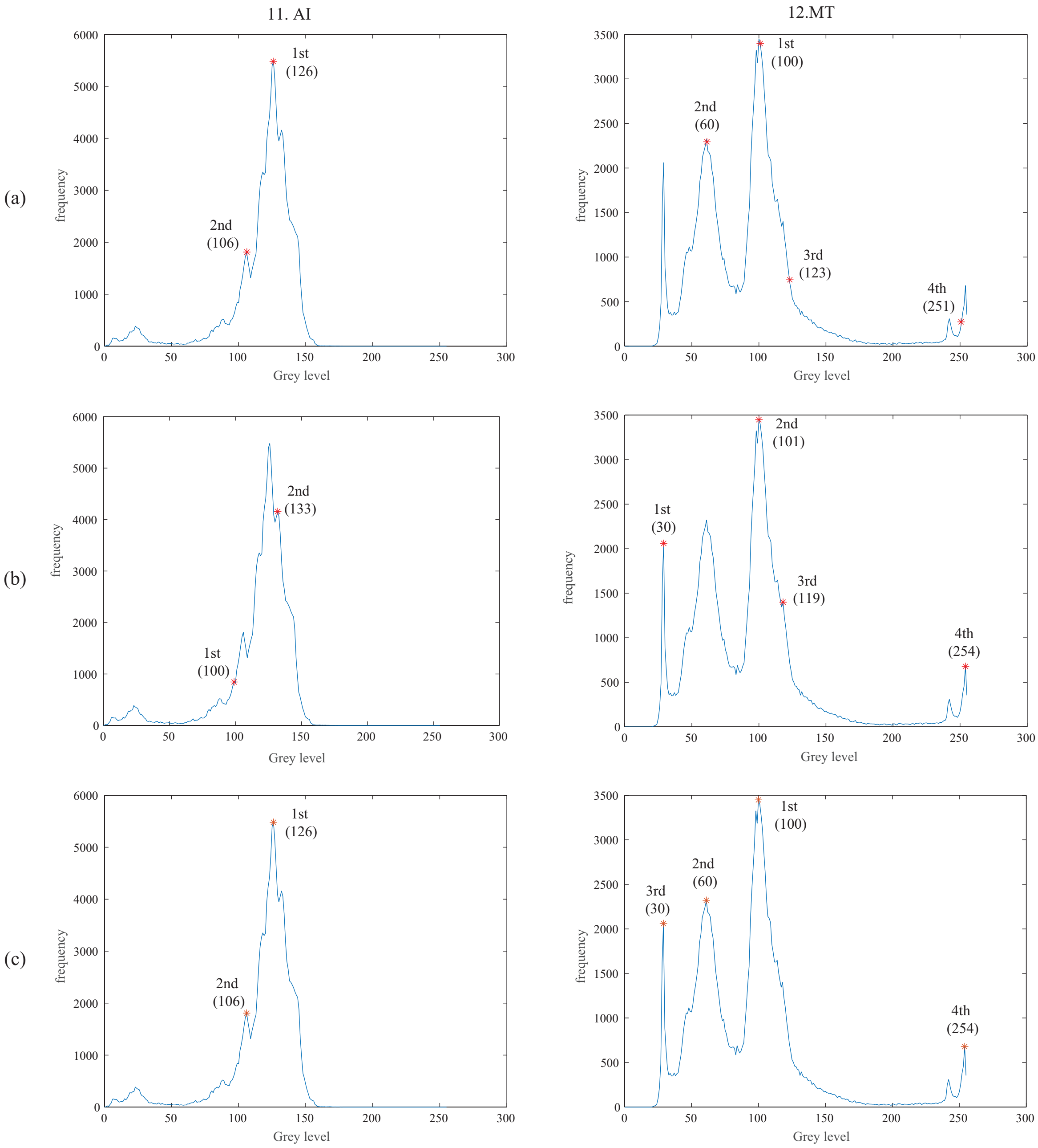

FIGURE 9. The initial cluster centers and cluster quantity for image #11-#12: (a) obtained by SC method; (b) obtained by HBA method; (c) obtained by the IKM method

has more computation cost than sample image #9 (Lena.png), even though Lena.png has larger image size as compared to moon.tif, the computation cost for moon.tif is more as it is segmented into two regions whereas Lena.png is segmented into four regions. In the experimental results presented in Table 3, FCM shows better performance as compared to K-medoids algorithm, yet large size images have higher increments in computation cost. K-means algorithm has better image segmentation quality in the majority of cases discussed and the computation time of K-means makes it more suitable for image segmentation. The results show that the overall performance of the proposed improved K-means is superior to other soft clustering algorithms in terms of computational complexity and segmented image quality.

Table 3. The image segmentation results based on three clustering methods

Name of samples	FCM method		K-medoids method		The proposed method	
	Time	Q-value	Time	Q-value	Time	Q-value
1. MR	0.29	0.001	2.35	0.0007	0.19	0.001
2. BA	0.76	0.02	11.83	0.02	0.46	0.02
3. CA	1.83	0.004	11.98	0.004	0.81	0.004
4. CO	1.61	0.006	17.83	0.006	0.81	0.006
5. MO	2.99	0.07	416.39	0.04	1.66	0.06
6. PO	3.71	0.0003	9.43	0.0003	1.29	0.0003
7. GL	1.80	0.006	6.99	0.006	0.76	0.005
8. AT	55	0.01	286.50	0.01	16.30	0.01
9. LE	7.19	0.04	160	0.04	3.06	0.01
10. VA	4.15	0.02	89.90	0.02	1.85	0.02
11. AI	16.92	0.02	1252	0.02	6.68	0.03
12. MT	4.35	0.02	97.30	0.02	1.86	0.02

5. **Conclusions.** In this paper, we presented an improved K-means algorithm, to overcome the drawbacks of K-means algorithm in image segmentation scenario. First of all, an automatic initialization parameter determination algorithm was devised for clustering algorithms, employing the image histogram information. The process was organized into two histogram scanning steps for better determination of initialization parameters. The initial parameters obtained were fed to K-means algorithm to output regions in the image, producing the segmented image. The objective and subjective evaluation of the proposed improved K-means with HBA and SC method was done, in terms of image segmentation quality and computational complexity. The results endorsed the performance of our proposed improved K-means algorithm. Moreover, the objective evaluation of the proposed improved K-means algorithm with other soft clustering algorithms also indicated better performance of the proposed improved K-means.

For the future research, a region merging technique can be devised to merge the disconnected pixels efficiently produced by the K-means method. Furthermore, the initialization problem of K-means can be extended for the color images and the task of grouping the individual peaks in the histograms, to represent the true color and the object boundaries in the image.

Acknowledgments. The authors would like to thank the National Natural Science Foundation of China (61203365, 61573128), National Key Research Program of China (2016YFC0401606), the Jiangsu Province Natural Science Foundation (BK2012149), and the Fundamental Research Funds for the Central Universities (2015B20114, 2015B25214) for their support of this paper.

REFERENCES

[1] A. Benaichouche, H. Oulhadj and P. Siarry, Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization, Mahalanobis distance and post-segmentation correction, *Digital Signal Processing*, vol.23, no.5, pp.1390-1400, 2013.

[2] M. J. Christ and R. Parvathi, Magnetic resonance brain image segmentation, *International Journal of VLSI Design & Communication Systems*, vol.3, no.4, p.119, 2012.

[3] T. Uemura, G. Koutaki and K. Uchimura, Image segmentation based on edge detection using boundary code, *International Journal of Innovative Computing, Information and Control*, vol.7, no.10, pp.6073-6083, 2011.

[4] M. Balafar, A. Ramli, M. Saripan and S. Mashohor, Review of brain MRI image segmentation methods, *Artificial Intelligence Review*, vol.33, no.3, pp.261-274, 2010.

[5] J. Yuan and A. M. Cheriyadat, Image feature based GPS trace filtering for road network generation and road segmentation, *Machine Vision and Applications*, vol.27, no.1, pp.1-12, 2016.

[6] M. Ferrara, A. Franco and D. Maio, A multi-classifier approach to face image segmentation for travel documents, *Expert Systems with Applications*, vol.39, no.9, pp.8452-8466, 2012.

[7] X. Ling, Y. Zhang, J. Xiong, X. Huang and Z. Chen, An image matching algorithm integrating global SRTM and image segmentation for multi-source satellite imagery, *Remote Sensing*, vol.8, no.8, 2016.

[8] V. Dey, Y. Zhang and M. Zhong, A review on image segmentation techniques with remote sensing perspective, *ISPRS Technical Commission VII Symposium on Advancing Remote Sensing Science*, Vienna, Austria, vol.38, pp.31-42, 2010.

[9] K. C. Ciesielski and J. K. Udupa, *Region-Based Segmentation: Fuzzy Connectedness, Graph Cut and Related Algorithms*, Springer, Berlin, Heidelberg, 2011.

[10] A. Nakib, H. Oulhadj and P. Siarry, A thresholding method based on two-dimensional fractional differentiation, *Image and Vision Computing*, vol.27, no.9, pp.1343-1357, 2009.

[11] C. Liu, W. Liu and W. Xing, An improved edge-based level set method combining local regional fitting information for noisy image segmentation, *Signal Processing*, vol.130, pp.12-21, 2017.

[12] R. Ma, Y. Zhu and J. Ding, Soft image segmentation based on center-free fuzzy clustering, *Transactions of Nanjing University of Aeronautics and Astronautics*, vol.30, no.1, pp.67-76, 2013.

[13] P. Sharma and J. Suji, A review on image segmentation with its clustering techniques, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol.9, no.5, pp.209-218, 2016.

[14] C. Bouveyron and C. Brunet-Saumard, Model-based clustering of high-dimensional data: A review, *Computational Statistics and Data Analysis*, vol.71, pp.52-78, 2014.

[15] S. Aghabozorgi, A. S. Shirkhorshidi and T. Y. Wah, Time-series clustering – A decade review, *Information Systems*, vol.53, pp.16-38, 2015.

[16] N. Ahmed, Recent review on image clustering, *IET Image Processing*, vol.9, no.11, pp.1020-1032, 2015.

[17] P. C. Besse, B. Guillouet, J.-M. Loubes and F. Royer, Review and perspective for distance-based clustering of vehicle trajectories, *IEEE Trans. Intelligent Transportation Systems*, vol.17, no.11, pp.3306-3317, 2016.

[18] B. B. Firouzi, M. S. Sadeghi and T. Niknam, A new hybrid algorithm based on PSO, SA, and K-means for cluster analysis, *International Journal of Innovative Computing, Information and Control*, vol.6, no.7, pp.3177-3192, 2010.

[19] P. Shi, X. Fan, J. Ni and G. Wang, A detection and classification approach for underwater dam cracks, *Structural Health Monitoring*, vol.15, no.5, pp.541-554, 2016.

[20] S. Khanmohammadi, N. Adibeig and S. Shanehbandy, An improved overlapping k-means clustering method for medical applications, *Expert Systems with Applications*, vol.67, pp.12-18, 2017.

[21] S. M. R. Zadegan, M. Mirzaie and F. Sadoughi, Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets, *Knowledge-Based Systems*, vol.39, pp.133-143, 2013.

[22] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[23] Y. Zhang, D. Huang, M. Ji and F. Xie, Image segmentation using PSO and PCM with Mahalanobis distance, *Expert Systems with Applications*, vol.38, no.7, pp.9036-9040, 2011.

[24] B. Zhang, M. Hsu and U. Dayal, K-harmonic means – A data clustering algorithm, *HP Laboratories Technical Report*, no.124, 1999.

[25] Y. Gui, X. Bai, Z. Li and Y. Yuan, Color image segmentation using mean shift and improved spectral clustering, *The 12th International Conference on Control, Automation, Robotics and Vision*, 2012.

[26] M. Frackiewicz and H. Palus, KHM clustering technique as a segmentation method for endoscopic colour images, *International Journal of Applied Mathematics and Computer Science*, vol.21, no.1, pp.203-209, 2011.

[27] S. S. Khan and A. Ahmad, Cluster center initialization algorithm for K-means clustering, *Pattern Recognition Letters*, vol.25, no.11, pp.1293-1302, 2004.

[28] H.-S. Park and C.-H. Jun, A simple and fast algorithm for K-medoids clustering, *Expert Systems with Applications*, vol.36, no.2, pp.3336-3341, 2009.

[29] N. Dhanachandra, K. Manglem and Y. J. Chanu, Image segmentation using k-means clustering algorithm and subtractive clustering algorithm, *Procedia Computer Science*, vol.54, pp.764-771, 2015.

[30] M. Goryawala, M. R. Guillen, S. Gulec, T. Barot, R. Suthar, R. Bhatt, A. Mcgron and M. Adjouadi, An accurate 3D liver segmentation method for selective internal radiation therapy using a modified k-means algorithm and parallel computing, *International Journal of Innovative Computing, Information and Control*, vol.8, no.10(A), pp.6515-6538, 2012.

[31] S. Kurumalla and P. S. Rao, An improved k-means clustering algorithm for image segmentation, *International Journal of Applied Engineering Research*, vol.10, no.13, pp.33143-33147, 2015.

[32] D. Martin, C. Fowlkes, D. Tal and J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *Proc. of the IEEE International Conference on Computer Vision*, Vancouver, BC, United States, vol.2, pp.416-423, 2001.

[33] E. Kucukkulahli, P. Erdomu and K. Polat, Histogram-based automatic segmentation of images, *Neural Computing and Applications*, vol.27, no.5, pp.1445-1450, 2016.

[34] 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.

[35] W. Tang, C. Zhang, X. Zhang and C. Liu, Medical image segmentation based on improved FCM, *Journal of Computational Information Systems*, vol.8, no.2, pp.887-894, 2012.

[36] A. Khatami, S. Mirghasemi, A. Khosravi and S. Nahavandi, A new color space based on k-medoids clustering for fire detection, *Proceedings – 2015 IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon Tong, Hong Kong, pp.2755-2760, 2015.