Research Article   **Open Access**

Image Processing of SEM Image Nano Silver using K-means MATLAB Technique

Elham Jasim Mohammad*, Aseel Mustafa Abdul Majeed, Faten Sajet Mater, Sanar Gasid Hassan, Ishraq Ahmed Shakir, Farouk Khalid Shaker

Department of Physics, Collage of Science, Mustansiriyah University, IRAQ
Correspondent author: elham.jasim@yahoo.com

Abstract

Nanotechnology is one of the non-exhaustive applications in which image processing is used. For optimal nanoparticle visualization and characterization, the high-resolution Scanning Electron Microscope (SEM) and the Atomic Force Microscope (AFM) are used. Image segmentation is one of the critical steps in nanoscale processing. There are also different ways to reach retail, including statistical approximations. In this study, we used the K-means method to determine the optimal threshold using statistical approximation. This technique is thoroughly studied for the SEM nanostructure silver image. Note that, the image obtained by SEM is good enough to analyze more recently images. The analysis is being used in the field of nanotechnology. The K-means algorithm classifies the data set given to k groups based on certain measurements of certain distances. K-means technology is the most widely used among all clustering algorithms. It is one of the common techniques used in statistical data analysis, image analysis, neural networks, classification analysis and biometric information. K-means is one of the fastest collection algorithms and can be easily used in image segmentation.

The results showed that K-means is highly sensitive to small data sets and performance can degrade at any time. When exposed to a huge data set such as 100.000, the performance increases significantly. The algorithm also works well when the number of clusters is small. This technology has helped to provide a good performance algorithm for the state of the image being tested.

Keywords: K-means, Nano Image, Threshold, Image Processing.

الخلاصة

تعد تقنية النانو واحدة من بين قائمة التطبيقات غير الشاملة التي تستخدم فيها معالجة الصور. ولتصور وتوصيف البنية النانوية بشكل مثالي، يتم استخدام الماسح الضوئي ذو الدقة العالية (SEM) ومجهر القوة الذرية (AFM). وتعتبر عملية تجزئة الصورة واحدة من الخطوات الحاسمة في معالجة الصور النانوية. كما إن هنالك طرق مختلفة للوصول إلى التجزئة ومنها طرق التقريب الإحصائية. في هذه الدراسة، استخدمنا طريقةK-meansلتحديد العتبة المثلى باستخدام التقريب الإحصائي. تمت دراسة هذه التقنية لصورة الماسح الضوئي ذو الدقة العالية للفضة النانوية. علما إن الصورة التي تم الحصول عليها عن طريق SEM جيدة بما يكفي لتحليلها. وفي الآونة الأخيرة يستخدم تحليل الصور في مجال تكنولوجيا النانو.إن خوارزميةK-means تقوم بتصنيف مجموعة البيانات المعطاة إلى مجموعات k استناداً إلى قياسات معينة لبعض المسافات. ان تقنية K-means هي الأكثر استخداما على نطاق واسع بين جميع خوارزميات التجميع. وهي واحدة من التقنيات الشائعة المستخدمة في تحليل البيانات الإحصائية، تحليل الصور، الشبكات العصبية، تحليل التصنيف والمعلوماتية الحيوية. حيث يعتبرK-means واحدًا من أسرع خوارزميات التجميع ويمكن استخدامه بسهولة في تجزئة الصور.

وأظهرت النتائج أن K-means حساس للغاية لمجموعات البيانات الصغيرة ويمكن أن يتدهور الأداء في أي وقت. وعندما يتعرض إلى مجموعة بيانات ضخمة مثل ١٠٠٠٠٠، يزداد الأداء بسرعة ملحوظة. وتعمل الخوارزمية أيضًا بشكل جيد عندما يكون عدد التجمعات صغيراً. لقد ساعدت هذه التقنية في تقديم الأداء الجيد للخوارزمية لحالة الصورة التي تم اختبارها.

Introduction

Nanotechnology is a field of applied science, which aims to produce and use engineered materials and structures close to the atomic or molecular scale. These structures are referred to as nanoparticles, possessing at least one dimen-

sion of less than 100 nanometers. In order to visualize and characterize this type of nanostructured material, the high-resolution Scanning Electron Microscope (SEM) and Atomic Force Microscopy (AFM) are ideally used [1]. The images are widely analyzed and processed to obtain quantitative and quantitative information on many physical properties including size, length, density, surface morphology, texture, and roughness. Although the images obtained from SEM and AFM are good enough to analyze the properties above, it cannot be analyzed in a raw image format as the abstract images. Thus, the raw image must undergo a number of pre-processing techniques for the image to facilitate the researcher's search process and also to obtain the correct information [2-4]. Pre-processing steps aim at improving contrast, background subtraction, restoration, and compression.

Segmentation of images is one of the most important techniques in image processing. It is a pre-processing step in image analysis, computer vision, and pattern recognition [5]. The process of splitting a digital image into multiple regions is called segmentation of images. Segmentation of images is commonly used to identify objects and borders in images. The result of a segmentation of image is a set of segments that includes the entire image, or a set of contours extracted from the image. All pixels are connected to an area with respect to certain characteristics or calculated properties, such as color, density, or texture [6]. Adjacent areas differ significantly in relation to the same characteristics [7].

Image clustering is a good technique used to segment images. After extracting features, these features are put together in clusters separate groups well based on each category of image. The clustering algorithm aims to develop partitioning decisions based on the initial set of groups that are updated after iteration [8]. This paper focused on the K-mean clustering and its procedures. The performance of the algorithms is tested with real images of 20 nm SEM image of nano silver.

Materials and Methods
K-means
The term k-means was first used by James Mac Queen in 1967 [9], although the idea dates back to Hugo Steinhaus in 1957 [10]. The standard algorithm was first proposed by Stuart Lloyd in 1957 as a method of pulse code formation, although it was not published outside Bell Labs until 1982 [11]. In 1965, E. W. Forgy published the same method, which is why it is sometimes referred to as Lloyd-Forgy [12]. K-means clustering is an algorithm that classifies the given data set to k clusters groups based on certain measurements for some specific distance. K-means is the most widely used and studied among all clustering algorithms. K-means is one of the simplest unsupervised learning algorithms for solving a well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters fixed a priori. The main idea is to determine k centroids, one for each cluster set. These centroids must be placed in a cunning manner because of the different location causing a different result. So, the best option is to put them as far as possible far away from each other. The next step is to take each point that belongs to a particular data set and associate it with the nearest centroid point. When there is no pending point, the first step is completed, and an early age is implemented for the group. At this stage, we need to recalculate k new centroids as bar centers of clusters resulting from the previous step. After we get these k new middle points, The new binding must be made between the same data set points and the nearest new central point. A loop has been created. As a result of this loop, we may observe that k centroids change their position step by step until no changes are made.

The clustering method is currently used to segmenting large scale images. Clustering is one of the unsupervised learning mode in which a set of essentials is separated into standardized groups. There are different types of clustering: hierarchical clustering, Fuzzy C-mean clustering, K-means clustering. K-means is one of the most generally techniques used clustering in various applications [13].

K-means clustering is a partition-based cluster analysis method. K-means is a clusteringtechnology that has been widely applied to solve low-level image segmentation tasks. The choice of primary cluster centers is very important because this prevents the clustering algorithm from producing incorrect decisions. The most common initialization procedure chooses the initial cluster center selects randomly from the input data [8].

K-means is the most important flatclustering algorithm. Its goal is to reduce the average square euclidean space of documents from their cluster centers where a cluster center is defined as the mean or centroid point $\vec{\mu}$ of central documents in the ω group [14]:

$$\vec{\mu}(\omega) = \frac{1}{(\omega)} \sum_{\vec{x} \in \omega} \vec{x} \qquad (1)$$

A measure of the extent to which the middle coasts (centroids) represent the members of their cluesresidual sum of lers is the residual sum of squares (RSS), and the square distance of each center-bound square collected on all vectors [14]:

$$RSS_k = \sum_{\vec{x} \in \omega} |\overrightarrow{x - \vec{\mu}(\omega_k)}|^2 \qquad (2)$$

$$RSS = \sum_{k-1}^{K} RSS_k \qquad (3)$$

RSS is the objective function in K-means and our goal is to reduce it. Since N is fixed, reducing RSS is equivalent to reducing the mean square distance, a measure of how centroids are represented in their documents. The first step of K-means is to choose, as a cluster K randomly selected documents the seeds. The algorithm then moves seeds the cluster centers around in space in order to reduce RSS [14].

Procedure
K-means is one of the fastest clustering algorithms and can be easily used in image segmentation. We will process the K-means algorithm as the most accurate of the other clustering algorithms, where K-means is one of the clustering techniques used to classify characteristics by giving a set of values and trying to divide them into K clusters groups based on the basis of some analogues. We need to enter the number of expected elements initially. This algorithm is based on making the intraset distance as small as possible and making the intreset distance as large as possible. The intraset distance and the intreset distance mean that the intraset distance represents the extent of the divergence between the patterns of one class, and the accuracy of the classification depends on the least intraset distance within the group. Where the lowest intraset distance within the class gives the accuracy classification, while the internal distance means the extent of the distance between each category of the other and the most accurate classification depends on the far distance. The algorithm is defined0 as follows:

1. Read SEM image of nano Silver and convert the three-dimensional image to two-dimensional.
2. Assume that the cluster numbers are (2, 4, 6, 8, 10 and 12).
3. Take the value of K-means iteration = 10.
4. Assume the locations of cluster centers randomly.
5. Define the distances and labels in relation to features and cluster group numbers.
6. Find the centers of each cluster group, thecluster points and the new cluster centers.
7. Show resulting images with histograms.

Results and Discussion
K-means clustering is one of the common algorithms in clustering and segmentation technique. The K-means segmentation treats each pixel image (with RGB values) as a feature point that has a location in space. Thus, the basic K-means algorithm then arbitrarily locates, and this number of cluster group centers is in a multidimensional measurement space. Each point is then assigned to the cluster that has the nearest arbitrary vector. The procedure continues until there is no significant change in the intermediate vectors location of the class

between successive iterations of the algorithms. In this theoretical study, the performance of K-means algorithms is tested with the real images of SEM nano Silver 20 nm shown below in Figure 1.
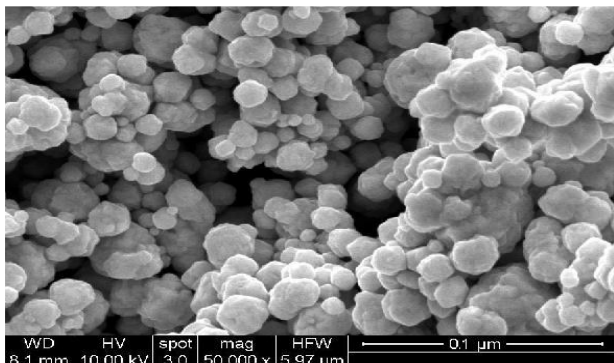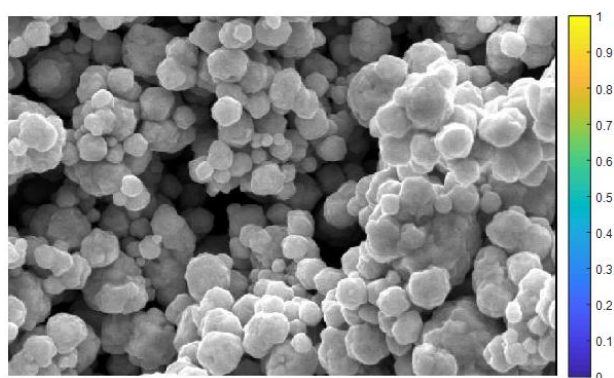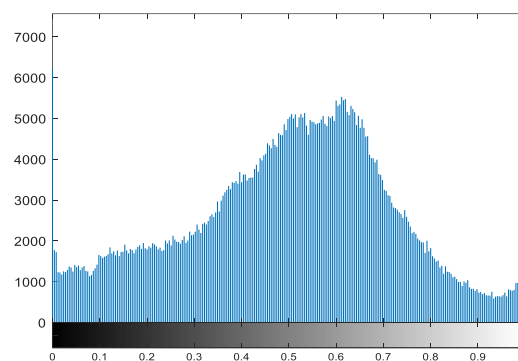

Figure 1: Real SEM Image of Nano Silver 20 nm.

The characteristics of the materials used in this study are Silver nano particles, AG 20nm, Spherical, 99.99%, metal basis 1. CAS No.:7440-22-4 2. Appearance: grey black solid powder 3. Apparent Density: 0.97g/ml 4. Tap Density: 2.16g/ml 5. COA of 20 nm sized silver nano particles.The original SEM image that was tested from the nanoSilver appears in Figure 2. All results are tested with the high-performance language of technical computing, MATLAB R2018a. Figure 2 show the original tested SEM image of nano Silver with its histogram.

(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 2: Original tested SEM image of nano Silver with its histogram.

In Figure 3, after using the statistical approach we get image segmentation. Where, we fit the background and background histogram for the Gaussian curves then the parameters are found.
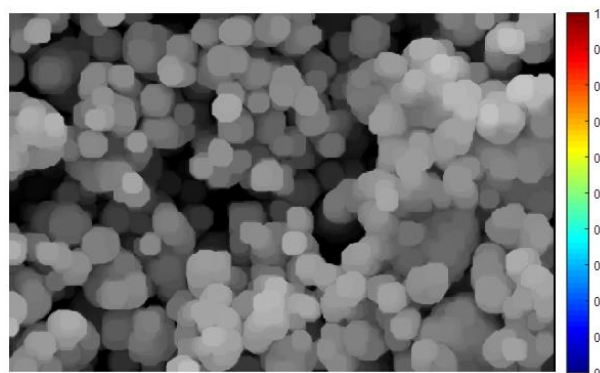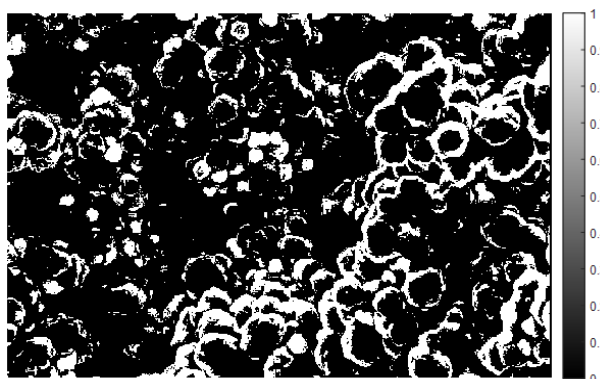

(a) Tested SEM image of nano Silver.
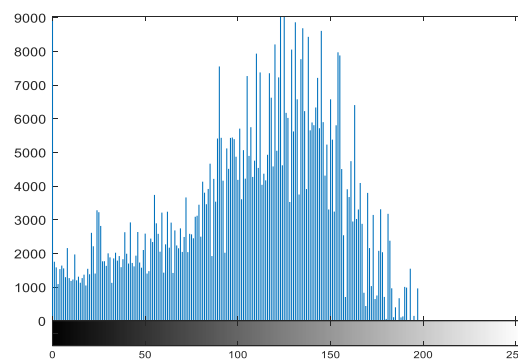
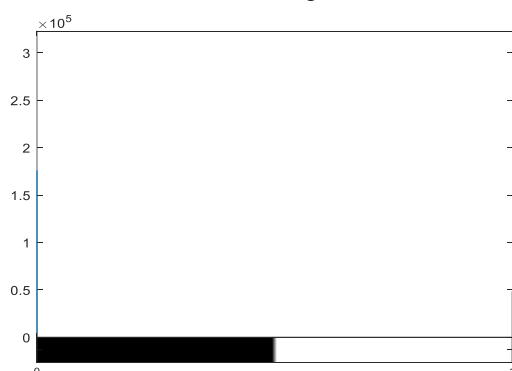(b) Histogram of tested SEM image of nano Silver.

Figure 3: Background tested SEM image of nano Silver with its histogram.

Figure 4, optimization was found by solving a set of equations using several parameters of a mixture of two Gaussian distribution to deter-

mine the optimum threshold using the statistical approach.
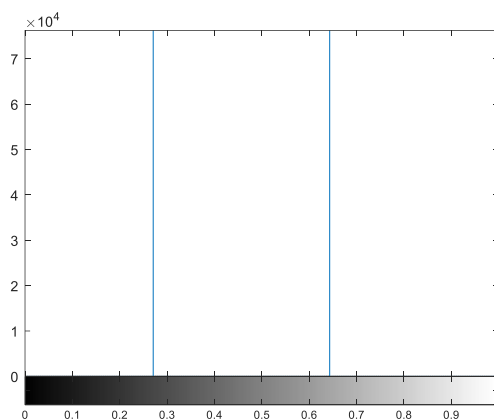

(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 4: Segmentation by thresholding. The gray threshed of tested SEM image of nano Silver with its histogram.

Figures (5-10), we used the K-means method (cluster numbers= 2, 4, 6, 8, 10 and 12)to determine the optimum threshold using the statistical approach.
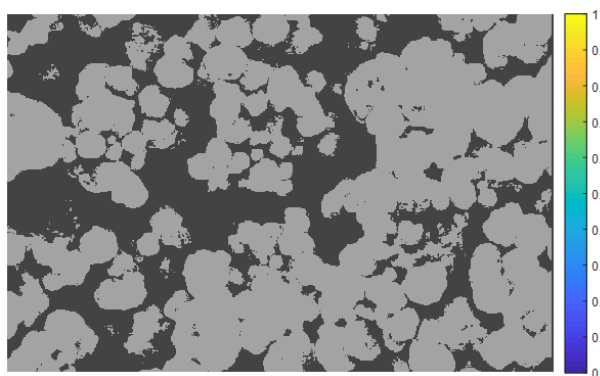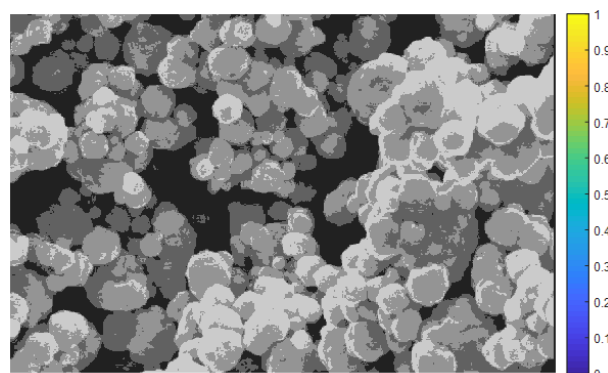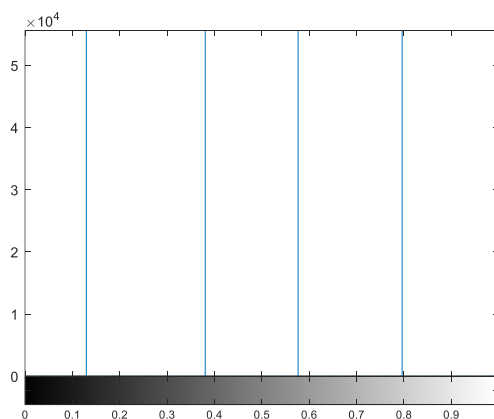

(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 5: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 2 with its histogram.
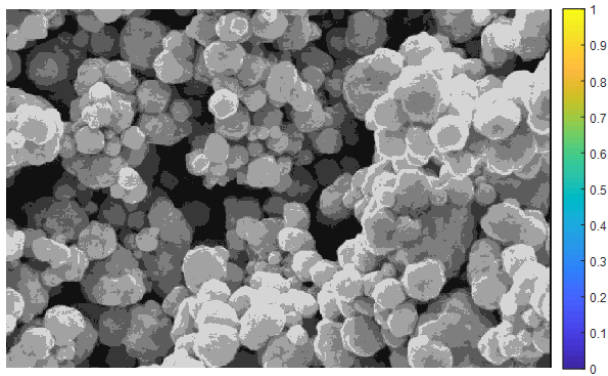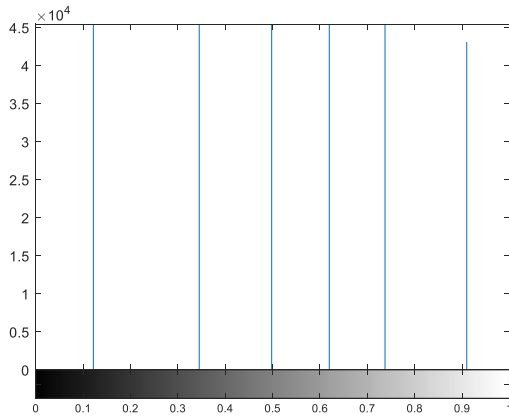

(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 6: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 4 with its histogram.
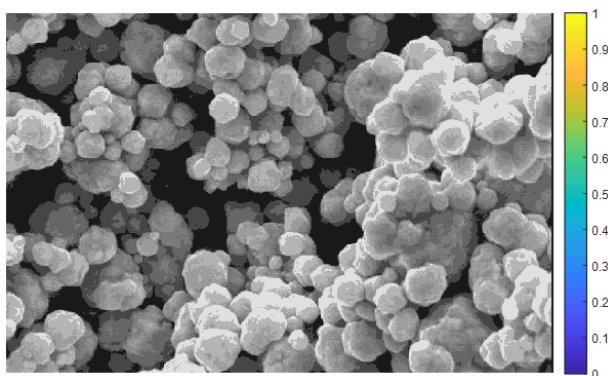
(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 8: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 8 with its histogram.

Figure 7: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 6 with its histogram.

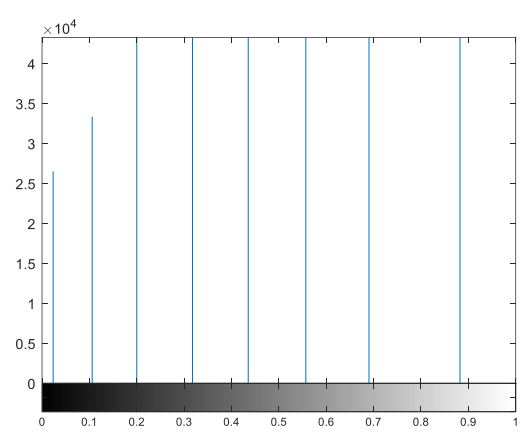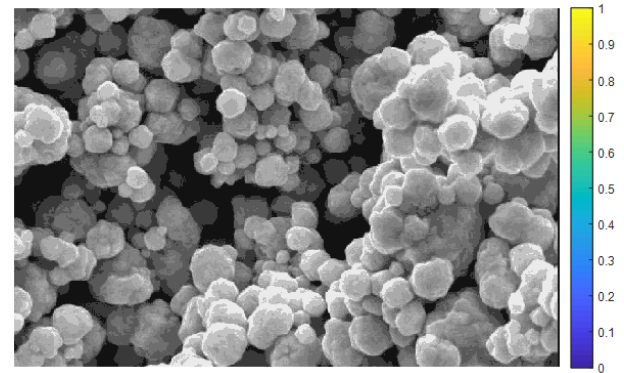(a) Tested SEM image of nano Silver.

(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 9: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 10 with its histogram.
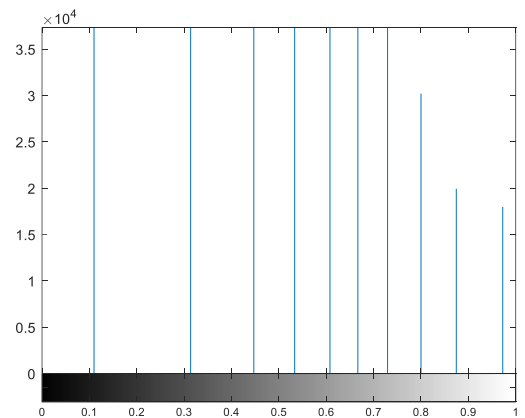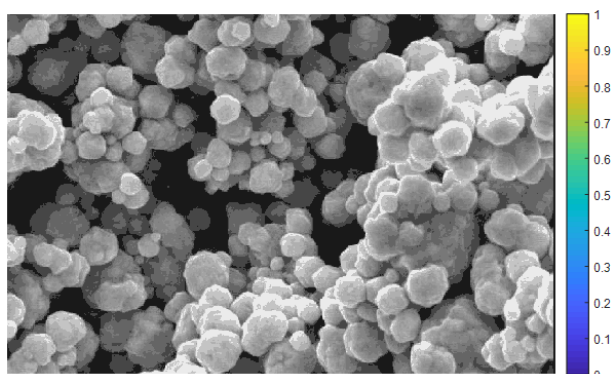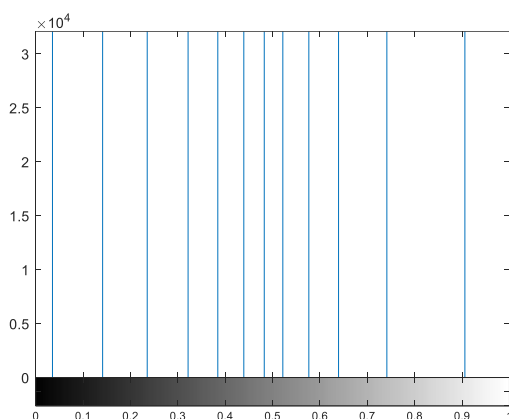
(a) Tested SEM image of nano Silver.

(b) Histogram of tested SEM image of nano Silver.

Figure 10: Segmentation by K-means for tested SEM image of nano Silver using cluster numbers= 12 with its histogram.

Conclusion

1. One of the most important treatments on images is image cutting, K-means and more algorithms. Image classification accuracy is the classification algorithm, which gives high accuracy because it depends on the correction centers in each implementation cycle.

2. By increasing the dimensions of the image, the effect of deleting some elements on the images is reduced, despite the increase in the number of deleted items, which are not important properties in the images. This feature is useful in exploiting these elements in certain applications, such as data security applications, and sending them across networks.

3. If the sum of the absolute difference in the binary Gaussian distribution is less than the sum of the absolute difference, the K-means method of the initial Gaussian distribution, we apply the image segmentation by non-directional classification, the image segmentation technique.

4. The initial assembly will determine the algorithm dramatically. Therefore, the initial compilation should be based on some reasonable assumptions.

5. The number of clusters group k should be determined initially, and the initial conditions must be appropriate because the K-means algorithm is very sensitive to the initial conditions. Different initial conditions may produce different results that eventually affect the algorithm to a large extent.

6. In K-means technology we can get multiple segments according to our cluster group size. Increase performance according to cluster sizes. The more cluster size is the more accuracy ratio.

7. K-means clustering technology can be used for better performance in applications such as face recognition and video retrieval. In different ways, the performance of the traditional K-means clustering group can be improved. Perhaps, the performance of most of the algorithm depends primarily on selected cluster centers. By properly repairing the cluster centers by knowing the domain, or any other means, the time of convergence can be improved with accuracy. So here is the point of finding a better mechanism for reforming cluster centers through prior knowledge.

References

[1] G. Poletti, F. Orsini, C. Lenardi and E. Barborini, "A comparative study between AFM and SEM imaging on human scalp hair," Journal of Microscopy-Oxford, vol. 211, no. 7, 2003, pp. 249-255.

[2] A. P. Chuklanov, S. A. Ziganshina and A. A. Bukharaev, "Computer program for the grain analysis of AFM images of nanoparticles placed on a rough surface," Surface and Interface Analysis, vol. 38, no. 4, 2006, pp. 679-681.

[3] E. Ficarra, L. Benini, E. Macii and G. Zuccheri, "Automated DNA fragments recognition and sizing through AFM image processing," IEEE Transactions on Information Technology in Biomedicine, vol. 9, no. 4, 2005, pp. 508-517.

[4] T. S. Spisz, Y. Fang, R. H. Reeves, C. K. Seymour, I. N. Bankman and J. H. Hoh, "Automated sizing of DNA fragments in atomic force microscope images," Medical & Biological Engineering & Computing, vol. 36, no. 6, 1998, pp. 667-672.

[5] R. R. N. Senthilkumaran, "A Study on Rough Set Theory for Medical Image Segmentation," International Journal of Recent Trends in Engineering, vol. 2, november 2009.

[6] S. k. A. P. Prasad Dakhole, "Fabric Fault Detection Using Image Processing Matlab," International Journal For Emerging Trends in Engineering and Management Research (IJETEMR), vol. 2, no. 1, 21 january 2016.

[7] C. R. Tippana, "Homogeneous Regions for Image Segmentation Based on Fuzzy," international journal & magazine of engineering, technologymanagement and research.

[8] S. panda, "Color Image Segmentation Using K-means Clustering and Thresholding Technique," IJESC, march 2015.

[9] J. B. MacQueen, (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability.1.University of California Press. pp. 281–297. MR 0214227. Zbl 0214.46201. Retrieved 2009-04-07.

[10] H. Steinhaus,(1957). "Sur la division des corps matériels en parties". Bull. Acad. Polon. Sci. (in French). 4 (12): 801–804. MR 0090073. Zbl 0079.16403.

[11] S. P. Lloyd, (1957). "Least square quantization in PCM". Bell Telephone Laboratories Paper. Published in journal much later: Lloyd., S. P. (1982). "Least squares quantization in PCM" (PDF). IEEE Transactions on Information Theory. 28 (2): 129–137.doi:10.1109/TIT.1982.1056489. Retrieved 2009-04-15.

[12] E.W. Forgy (1965). "Cluster analysis of multivariate data: efficiency versus interpretability of classifications". Biometrics. 21: 768–769. JSTOR 2528559.

[13] L. H. a. J. Y. LihuaTian, "Research on Image Segmentation based on Clustering Algorithm," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 9, pp. 1-12, 2016.

[14] D. M. Christopher, P. Raghavan and H. Schütze, "An Introduction to Information Retrieval", Cambridge University Press, Cambridge, England, April 1, 2009.