

## Research Article

### De-Noising and Segmentation of Brain MR images by Spatial Information and K-Means Clustering

<sup>1,2</sup>Arshad Javed, <sup>1</sup>Wang Yin Chai and <sup>1</sup>Narayanan Kulathuramaiyer

<sup>1</sup>Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia

<sup>2</sup>Faculty of Computer Sciences and Information, Al Jouf University, Saudi Arabia

**Abstract:** Image Segmentation is the process of partitioning a digital image into non-overlapping distinct regions, so that significant information about the image could be retrieved and various analysis could be performed on that segmented image. The aim of this study is to reduce the noise, enhance the image quality by considering the spatial information without losing any important information about the images and perform the segmentation process in noise free environment. K-Means clustering technique is used for the purpose of segmentation of brain tissue classes which is considered more efficient and effective for the segmentation of an image. We tested the proposed technique on different types of brain MR images which generates good results and proved robust against noise. Conclusion had been concluded at the end of this study.

**Keywords:** Cluster validity index, image segmentation, k-means, MRI, spatial information

## INTRODUCTION

The process of image segmentation means dividing a digital image into numerous sets of pixels or segments. The main objective of segmentation is to simplify and/or modify the illustration of an image into something that is more meaningful and easier to analyze (Linda and George, 2001). The segmentation of image is typically used to detect boundaries and objects in images. In particular, image segmentation is the method of allotting a tag to every pixel in an image such that pixels having the same label share certain visual features. There are three basic tissue classes in the human brain, Gray Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF). The separation and segmentation of these tissue classes with accuracy while maintaining the image quality is a very challenging task. If noise in the image exist, there is high probability to lose the meaningful information which can cause wrong analysis.

Magnetic Resonance Imaging (MRI) is an imaging technique which is used in the field of radiology for the purpose of visualization of internal structure of the body. MRI contains or provides a wide and sufficient information about the human soft tissues anatomy. MRI uses a magnetic field and pulses of radio wave energy to construct pictures of structures and organs inside the body (Web Source, <http://www.webmd.com/a-to-z-guides/magnetic-resonance-imaging-mri>). For the

acquisition of MRI test, the body part being studied is placed inside a special machine that surrounds a strong magnetic equipment. MRI scanners acquire digital images that can be stored and saved on a computer for the purpose to more study. Images from MRI are acquired for numerous reasons. The images then can be used to find problems like tumors, injuries, blood vessel diseases, or infections etc. Contrast stretching techniques or materials may be used during the process of MRI to express abnormal tissue more distinctively. Echo Planer Imaging or EPI is the most frequent technique for acquiring the fast MRI images. These fast or single-shots acquisitions make scans very sensitive to magnetic susceptibility differences at air/tissues interfaces in the brain. The problems like blurring effects and geometric distortions in EPI in spiral imaging caused by field inhomogeneity and severely affect predictions about the brain of subject (Hien *et al.*, 2009).

In the last decade, many researchers proposed a family of algorithms or techniques to perform the segmentation process using MR images. All these algorithms or techniques perform segmentation with different manners to solve the segmentation problems.

**K-Means clustering:** K-Means clustering algorithm (Ayman *et al.*, 2003; Jobin Christ and Parvathi, 2011) is unsupervised method of clustering under the category

**Corresponding Author:** Arshad Javed, Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Malaysia

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

of Squared Error-Based. It is widely, successfully and effectively applied in fields such as agriculture, astronomy, computer vision, image segmentation, classifier designs and feature analysis etc. K-Means algorithm is dependent on a cost function which is minimized iteratively. The minimization of cost function is dependent on the distance of the object to a center of cluster in the feature domain. K-means algorithm starts with an initial guess for the each cluster center and allots pixels to each group by using memberships. After each assignments of the pixels to the nearest cluster, the clusters centers are then recomputed by taking the mean of clusters points values. Convergence of the k-means can be perceived by evaluating the changes in the membership function or the cluster center at two successive iteration steps.

**Noise modeling:** Image noise makes up undesirable information which drops image quality. Noise is defined as a process  $n$  which affects the acquired image  $f$  and is not part of the scene (initial signal- $s$ ). By using the model for additive noise, this process can be expressed as:

$$f(i, j) = s(i, j) + n(i, j) \quad (1)$$

The noise in digital images may occur from numerous sources. The noise in digital images can occur during the image acquisition process or during image conversion time

Noise ( $n$ ) may be modeled either by a histogram or a probability density function (Gonzales and Woods, 2002) which is superimposed on the probability density function of the original image ( $s$ ). There are many types of noises such as salt & pepper noise, Gaussian noise, Speckle noise, Rician noise etc. The most common and frequently occurring type of noise in MR images is salt and pepper noise.

In the model of salt and pepper noise, only two possible values are possible,  $x$  and  $y$  and the probability of obtaining each of them is less than 0.1 (otherwise, the noise would vastly dominate the image). For 8 bit grayscale image, the typical intensity value for salt noise is close to 255 and that for pepper is close to 0.

$$PDF_{\text{salt\&pepper}} = \begin{cases} A & \text{for } g = x \text{ ("pepper")} \\ B & \text{for } g = y \text{ ("salt")} \end{cases} \quad (2)$$

This type of noise is generally caused by synchronization errors in the image digitizing or transmission or malfunctioning of camera's sensor cells or by memory cell failure.

**Cluster validity index:** The procedure of evaluating quantitatively the clustering algorithm's results refers to the Cluster Validity. By using k-means clustering

technique, we can acquire the partition of the given data set. Since k-means technique needs pre-defined number of clusters from the user and different values of clusters corresponds to different partitions, so here we need some mechanism to validate the clusters.

Recently many researchers proposed many different methods to validate the cluster that are appropriate for k-means. One of them is proposed (Dunn, 1974) as the Dunn's index cluster validation technique.

The main objective of the study is to develop a mechanism to reduce the noise, enhance the image quality and perform the segmentation process in an automatic and robust way.

## LITERATURE REVIEW

A method (Tong *et al.*, 2008) proposed for the segmentation of CT brain images by using the k-means and expectation maximization clustering techniques. It showed some better results however the method is not robust against noise and also in this method, the clusters in the techniques are fixed manually. Another method (Singh *et al.*, 1996) proposed for the segmentation of MRI by k-means clustering. This method does not show satisfactory results because no mechanism is adopted for noise reduction from the image as well as selecting the optimal clusters. Another method (Ahmed and Mohamad, 2008) proposed for the brain MR image segmentation by the combination of k-means clustering and Perona Malik Anisotropic diffusion filter. This method used Anisotropic diffusion filter and performed some morphological operations for the image enhancement. This method generated very good results but there no mechanism is adopted to select the optimal number of clusters.

One of the main disadvantage of the k-means clustering technique is that it needs the number of clusters as an input to the algorithm manually to initiate. As designed, the algorithm is not able to determine the appropriate number of clusters and depends upon the user to identify this in advance. The other main disadvantage of the k-means clustering algorithm is that, it does not perform the segmentation process with perfection in the presence of noise in the images which reduces the efficiency and performance of this Algorithm or in other words, this is insensitive against noise. We proposed in this study the solution of these problems and developed a system which is capable to reduce the noise and made the k-means algorithm to perform the process of segmentation of brain MR images in an automatic and unsupervised way with great performance.

## METHODOLOGY

To achieve the good segmentation, we first performed the preprocessing step to enhance the image quality because the presence of noise in the images

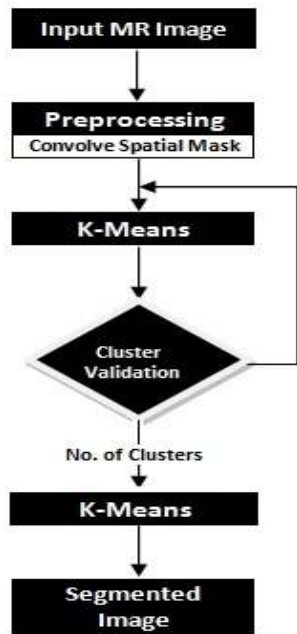


Fig. 1: System flow process

badly affects the segmentation results and there can be high chances of loss of regions partitioning information. After noise removal k-means technique is first executed for getting the optimal number of clusters and then for segmentation.

Complete system design of the proposed method is given in Fig. 1, while details about the each major components are discussed in the following subsections one by one.

**Spatial convolution mask:** One of the significant characteristics of the image is that, the neighboring pixels have great chances of correlation. A spatial convolution mask  $K_{mask}$  is used to utilize the spatial information where the neighborhood constitutes a squared window which is centered at the pixel  $x_j$  in spatial domain of the image. A square window of size  $5 \times 5$  was used throughout this study. In this way the each pixels in the image is corrected by taking the average of its neighbors. The spatial convolution mask (square window of size  $5 \times 5$ ) is designed for this purpose and is given below:

$$K_{mask} = \frac{1}{25} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

We convolved this mask with the original image and resulting image is obtained by the following function:

$$g(x, y) = K_{mask} * f(x, y) \tag{3}$$

where,  $f(x, y)$  represents the original image and  $g(x, y)$  resulting or noise reduced image.

**Segmentation:** The Segmentation process is performed to extract the abnormal (malignant) portion of the brain from the brain MR images. This segmentation process segments the brain image into two portions. One segment part contains the normal brain tissues while the other brain segment portion contains the abnormal (malignant) cells. The segment portion containing the abnormal tissues is the desired region which is known as tumorous region.

After the noise reduction from the brain MR image, K-Means clustering algorithm was applied to segment the image. This procedure exemplifies the image using the pixel intensity feature only. First K-Means algorithm is iterated for a range of hypothesized numbers of clusters to find the optimal number of clusters. Optimal number of clusters are selected based on cluster validity measurement criteria. The Dunn's index for cluster validity method (Jobin Christ and Parvathi, 2011) generated amazingly some good results for some of the tested images.

**K-Means clustering:** K-means clustering (Proposed by MacQueen, 1967) is one of the simplest unsupervised method of clustering (Ayman *et al.*, 2003; Jobin Christ and Parvathi, 2011) that solve the well known clustering and many practical problems. The algorithm follows a simple and easy way to classify a given data set  $n$  through a certain number of clusters  $k$  which are fixed a priori. The main objective of this algorithm is to define  $k$  centroids or centers, one for each cluster. The next step of this algorithm is to take each data point belonging to a given data set and associate it to the nearest cluster centroid. If no point is in pending, then the first step is completed and an early grouping of data set is done. At this stage the algorithm re-calculate  $k$  new centroids as centers of the clusters resulting from the preceding step. After that, we have  $k$  new centroids, a new allocation has to be done between the nearest new centroid and the same data set points. This is an iterative method and as a result of this we may notice that the  $k$  centroids change their location step by step until no more changes are done or centroids do not move any more.

The k-means algorithm aims at minimizing an objective function. The objective function of this algorithm is given as:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x^{(j)}_i - c_j\|^2 \tag{4}$$

where,  $\|x^{(j)}_i - c_j\|^2$  represents an Euclidean distance measure between a data point  $x^{(j)}_i$  and the cluster centre  $c_j$ .

**Dunn's cluster validity index:** Dunn's index (Dunn, 1974) for validation of clusters is a technique based on the idea of identifying the cluster sets that are compact and well-separated. For any given partition of clusters, Dunn's index is defined as:

$$DI = \frac{\min_{1 \leq i \leq n_c} \left\{ \min_{1 \leq i \leq n_c, j \neq i} \{dist(X_i, X_j)\} \right\}}{\max_{1 \leq i \leq n_c} \{diam(X_k)\}} \quad (5)$$

where,  $X_i$ ,  $n_c$ ,  $dist(X_i, X_j)$  and  $diam(X_k)$  represent the  $i$ th-cluster partition, cluster number, inter-cluster distance and intra-cluster distances respectively. The maximal value of  $DI$  means that given sample/data is well-clustered.

### IMPLEMENTATION

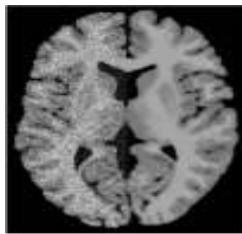
The proposed method is implemented using the Matlab R2009a environment. The entire procedure model is explained in Fig. 1.

First the image is given as input to the system. Since noise corrupts the image quality and this also affect the segmentation process, so the image is

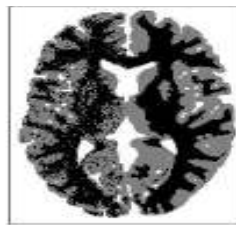
processed first for enhancing the image quality. For this purpose, a spatial convolution mask is designed and convolved with the image which corrects each pixel of the image by taking the average of its neighbors. A  $5 \times 5$  square window is used throughout this effort for incorporating the spatial information. In spatial mask all boundary pixels are kept equal to 1 while all remaining neighbors of each pixels equal to 2. After reducing noise from the image, the standard k-means algorithm is iterated for a number of times to obtain the optimal number of clusters. The best choice for optimal number of clusters is selected on the basis of cluster validity index, which measures the compactness and separation of the clusters. Dunn's index (Dunn, 1974) method is used to validate the clusters. This validity index generated good results on the tested images. After selecting the optimal number of clusters, k-means algorithm is applied on the image for getting the segmented image.

### RESULTS AND DISCUSSION

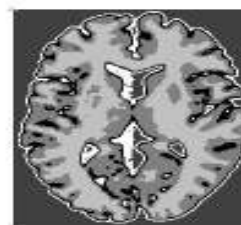
Though we have tested our proposed method over a large number of images with varying range of



(a) Image1



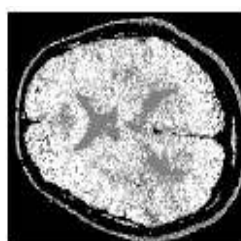
(a-1)



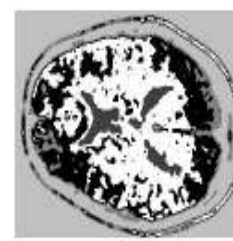
(a-2): Segmented by proposed method



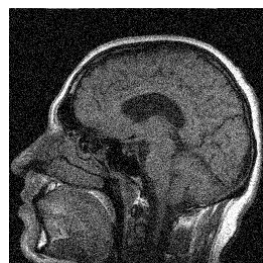
(b) Image2



(b-1)



(b-2): Segmented by proposed method



(c) Image3



(c-1)



(c-2): Segmented by proposed method

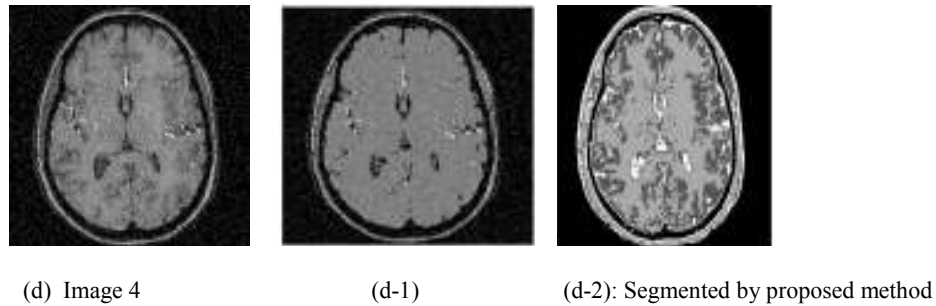


Fig. 2: (a, b, c, d) Input noisy images, (a-1, b-1, c-1, d-1) Segmented by standard k-means

Table 1: Dunn's index values (DI) calculated by Eq. (5)

Date set	Image size (pixels)	Dunn's index value and number of clusters					
		K = 2	K = 3	K = 4	K = 5	K = 6	K = 7
Image 1	202×250	0.049	0.054	0.054	0.054	0.054	0.054
Image 2	247×204	0.131	0.139	0.143	0.143	0.143	0.143
Image 3	256×256	0.079	0.084	0.087	0.087	0.087	0.087
Image 4	512×512	0.086	0.090	0.093	0.093	0.093	0.093

Table 2: Performance results comparison

Date set	Image size	No. of clusters	Time (seconds)	
			k-mean	Proposed method
Image 1	202×250	3	0.1079	0.1042
Image 2	247×204	4	0.1450	0.1363
Image 3	256×256	4	0.3381	0.1671
Image 4	512×512	4	0.4592	0.4248

complexity, but here we showed the experimental results of some of the selected images only. We tested our proposed method on the noisy images also which are taken some from internet (the real noisy images as given in Fig. 2a and b and the real data set of patients. We added first salt and pepper noise in the real data images as given in Fig. 2 (c) Image 3 and (d) Image 4. Our proposed method has showed the good results.

Our proposed method calculates the optimal number of clusters by the Dunn's index clusters validation method. The results generated by the Dunn's index are shown in Table 1. Segmentation performance of our proposed method is compared with the standard k-means. The performance parameters are given in Table 2. which shows that our method perform segmentation process fast as compared to standard k-means.

### CONCLUSION AND RECOMMENDATIONS

In this study we presented a method which performs the segmentation process in an efficient and automatic way. There is no need for the user involvement to set the number of classes in the system. The important characteristic of our method is that, it calculates the optimal number of clusters automatically by using the cluster validity function. Furthermore, the proposed method utilize the spatial information of each pixel in the image and enhances the quality of image which is important factor for the correct segmentation. The proposed method was tested on various images and

proved that the noise effect in segmentation is significantly less than the standard k-means algorithm.

Though the results of segmentation are very much satisfactory and the method is robust against noise as well but to make this study as a part of some automated system, many improvements can be done.

### REFERENCES

- Ahmed, M.M. and D.B. Mohamad, 2008. Segmentation of brain MR images for tumor extraction by combining kmeans clustering and perona-malik anisotropic diffusion model. *Intl. J. Image Process.*, 2(1): 27-34.
- Ayman, E.B., A.F. Aly, F. Robert and L.R. Renato, 2003. A unified approach for detection, visualization and identification of lung abnormalities in chest spiral CT scan. *Proceedings of the 17th International Congress and Exhibition on Computer Assisted Radiology and Surgery*, London, 1256: 998-1004.
- Dunn, J.C., 1974. Well separated clusters and optimal fuzzy partitions. *J. Cybern.*, 4: 95-104.
- Gonzales, R.C. and R.E. Woods, 2002. *Digital Image Processing*. 2nd Edn., Prentice Hall Inc.
- Hien, M.N., P.S. Bradley, L.M. Robert and N.D. Minh, 2009. Joint estimation and correction of geometric distortions for EPI functional MRI using harmonic retrieval. *IEEE T. Med. Imag.*, 28(3).
- Jobin Christ, M.C. and Dr. R.M.S. Parvathi, 2011. An adaptive mean-shift algorithm for MRI brain segmentation. *Int. J. Eng. Sci. Tech.*, 3(7): 5550-5554.
- Linda, G.S. and C.S. George, 2001. *Computer Vision*. Prentice-Hall, New Jersey, pp: 279-325, ISBN: 0-13-030796-3.

- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, 1: 281-297.
- Singh, M., P. Patel, D. Khosla and T. Kim, 1996. Segmentation of functional MRI by K-means clustering. IEEE T. Nucl. Sci., 43(3): 2030-2036.
- Tong, H.L., F.A.F. Mohammad and K. Ryoichi, 2008. Segmentation of CT brain images using K-means and EM clustering. Proceeding of the 5th International Conference on Computer Graphics, Imaging and Visualisation (CGIV '08), Penang, pp: 339-344.