

# A Study of Brain Tumor Segmentation from MR Images Using Fuzzy C Means Clustering

Chasen Campbell, University of North Georgia, Gainesville, USA

September 2019

## 1 Introduction

The use of clustering algorithms for image segmentation shows great potential amongst a variety of scientific and engineering fields because image segmentation takes an image of interest and extracts portions of the image for ease of analysis for some particular application[3]. Specifically, the medical industry has taken great interest in using clustering algorithms for image segmentation to extract the pixels of a brain tumor from MR images of cancer patients. The use of computer vision functions can then be used to count the number of pixels in the tumor with excellent accuracy[1]. By doing this, the area of the tumor can be determined and health care professionals can use the processed images to aid in diagnosis, treatment, and study of brain tumors [2].

In this paper, we'll examine the use of fuzzy c-means clustering (FCM) to achieve image segmentation of a brain tumor from MR images in the Matlab environment. An introduction to FCM, methods to extract clustered pixels, and solutions to problems with FCM implementation will be discussed. Over-segmentation and determination of the optimal number of clusters prove to be the main obstacles in this paper. The elbow method, an optimal cluster number method, is implemented. However ambiguous and unreliable the elbow method may be in several studies, it will prove to be a useful method of cluster approximation for this particular case study of the two MR images[4]. The lack of reliability of the elbow method and the design of the algorithm unfortunately results in poor code reuse and algorithm run times that are not ideal. The literature shows that FCM is not ideal for noise corrupted images, as its membership function does not consider the spatial relationship among data [2]. Pre-processing methods such as removal of unnecessary tissue and noise reduction have been proven to enhance FCM accuracy[3].

## 2 Fuzzy C-Means Clustering

FCM is a clustering method that assigns data points to multiple clusters, such that each data point has an assigned degree of membership to each cluster. For

instance, consider an input vector of data  $\vec{x}$  that has  $n$  many data points during a FCM calculation with  $j$  many clusters, so that  $C = j$ . Matlab's built in FCM function will output a matrix  $U$  that contains the membership information for all of the data.

$$U = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{j1} & m_{j2} & \cdots & m_{jn} \end{bmatrix} \quad (1)$$

Where  $0 \leq m_{jn} \leq 1$  is measures the degree of similarity the  $n$ th data point has with the  $j$ th cluster. It is also true that for any column in the matrix, say column 1, the following holds:

$$\sum_{k=1}^j m_{k1} = 1 \quad (2)$$

Example: If you had 2 clusters and 2 data points, then a potential case could be that  $x_1$  is 30 percent similar to  $C_1$  and 70 percent similar to  $C_2$ , while  $x_2$  might be 23.4 percent similar to  $C_1$  and 76.6 percent similar to the  $C_2$ . FCM would then assign  $x_1$  to  $C_2$  and  $x_2$  to  $C_2$ . FCM in Matlab also yields a  $c$  by  $dim$  matrix called centers and contains the center points of each cluster, where  $c$  is the number of clusters and  $dim$  is the dimension of the data [1].

## 2.1 FCM Algorithm

FCM seeks to minimize the objective function [1]

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2 \quad (3)$$

where

- $D$  is the number of data points.
- $N$  is the number of clusters.
- $m$  is a matrix exponent that controls the fuzzy overlap and  $m > 1$ . This controls just how "fuzzy" the membership of each data point is or how many data points have significant membership values in more than one cluster.
- $x_i$  is the  $i$ th data point
- $c_i$  is the center of the  $j$ th cluster
- $\mu_{ij}$  is the degree of membership of  $x_i$  in the  $j$ th cluster. As mentioned earlier, for a  $j$ th data point the sum of the member ship values is one.

- The operator  $||.||$ , called the norm, is a generalization of the absolute value for vector spaces.

First, the membership values  $\mu_{ij}$  are randomly initialized. Second, the cluster centers are calculated as such

$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m} \quad (4)$$

Third,  $\mu_{ij}$  is updated accordingly

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (5)$$

The fourth step is to recalculate  $J_m$  with the updated  $\mu_{ij}^m$  value. Steps 2-4 are repeated until the objective function converges[1].

### 3 Implementation

Unfortunately, reuse of this function is not currently possible and an already segmented tumor image from this function must be provided for image2 and image3 implementation.

During implementation, the file that will locate and segment the tumor is called findTumor.m. The function will take a p by q image file and its corresponding provided file like so

$$findTumor('image.filetype','image6Clusters.filetype')$$

where p and q are the numbers of the vertical and horizontal pixels of the image respectively. This would make p\*q the resolution of the image.

Next, findTumor.m will read the input image and separate the red, blue, and green intensities of each pixel and store this information in a p by q by 3 array.

The three dimensional array is then rearranged into a p\*q by 3 double valued matrix. Each row of the matrix represents a pixel and the columns will contain the red, green, and blue intensity values of each pixel (the (R,G,B) coordinates for each pixel in RGB space).

Ideally, the function would determine the optimal number of clusters automatically, but the cluster numbers for each image have been predetermined for this paper and will be addressed in the next section.

Once the number of clusters is determined, the function will run the FCM algorithm and create a vector of maximum values from each column of matrix U eqn(1). Using both matrix and logical indexing, each pixel is assigned to

the cluster for which it has the highest membership value. The implementation requires a twice nested for loop for each cluster and unfortunately results in an expected run time of at least  $O(cpq)$  where  $p$  and  $q$  are the dimensions of the input image and  $c$  is the number of clusters.

The function will output the original MR image,  $C$  many segments of the image, plus an optimally segmented tumor by means of texture filter from the cluster that contained the tumor.

### 3.1 Issues With Implementations and Proposed Solutions

It is evident that MRI pre-processing methods such as noise reduction and removal of the skull and eye regions can help improve clustering [2][3]. FCM in fact does not consider spatial relationship between data points while assigning membership values, which can result in loss of accuracy in segmentation if there is noise corruption in the image [2].

The issue of determining the appropriate number of clusters was addressed by mostly trial and error. One of the issues in the implementation is that each image must be taken case by case. A proposed solution would be to use the elbow method to gain an idea of where to start with the clustering for each image[4]. The elbow method compares how much of the variance of the data can be explained by each number of clusters[5]. This is achieved by calculating the total within-cluster sum of square WSS and plotting it against each number of clusters[4]. The idea is that there is a particular value for  $C$  such that when you pass it, you won't learn much more information about the data. But it is merely a heuristic to help provide a starting point for implementation and can't currently be unambiguously defined[7]. However, it is found that the appropriate number of clusters for analysis in RGB space did in fact agree with elbow plot.

Before the use of texture filter to refine segmentation, FCM alone provided under-segmentation at 5 clusters and a slight over-segmentation at 6 clusters. This could be possibly be explained by the use of RGB color space in addition to the fact that the membership function does not consider the spatial relationship between data points, which could result in slight ambiguity of membership values[2].

## 4 Results

The results seem reasonable and texture filtering was used to reduce the over-segmentation of the FCM clustering. However, it seems that image2 has slight under-segmentation and image3 has slight over-segmentation after texture filters. The tumor pixel count for image2 is 97165 and the pixel count for image3 is 1396083. It seems unlikely that image3 would have a higher pixel count, but this is most likely explained by the over-segmentation. 6 clusters were used for each image and conveniently agreed with the elbow method.

In figure 1, the image segment in the middle-far right is the segmented image that contains the tumor. The image on the top left was the segmented image refined with a text filter and is expected to contain most of the tumor pixels. 6 clusters were used. Image2 shows less non cancerous tissue in the final segmented picture than image3. Figures 3 and 4 show a comparison between using 5 clusters and 6 clusters respectively. It seems by inspection that 6 clusters is more accurate.

Figure 5 shows the results for image3, where the the segment with the bright pink ball is the segment with the tumor. The image in the top left is the segmented tumor after texture filter.

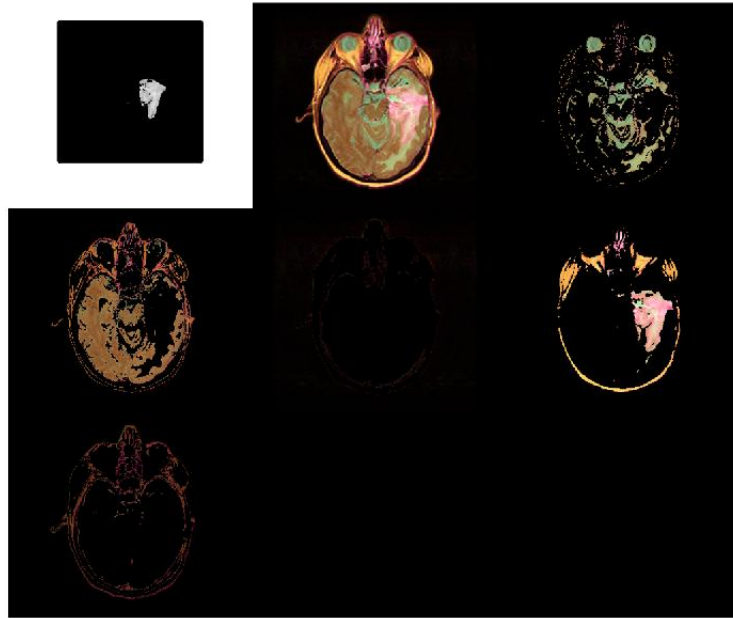


Figure 1: Results for image2

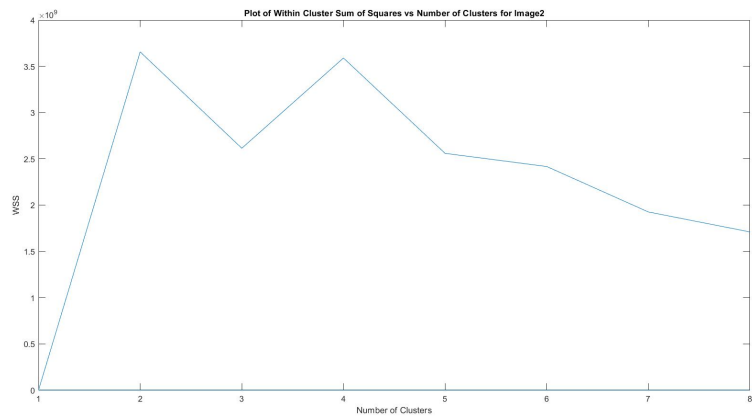


Figure 2: WSS vs number of clusters for image2



Figure 3: 5 cluster segmentation (image2)

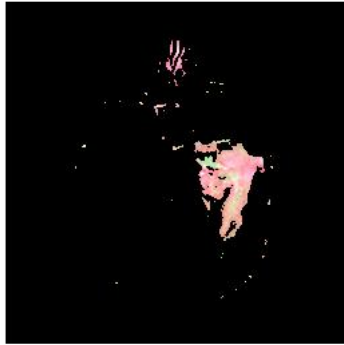


Figure 4: 6 cluster segmentation (image2)

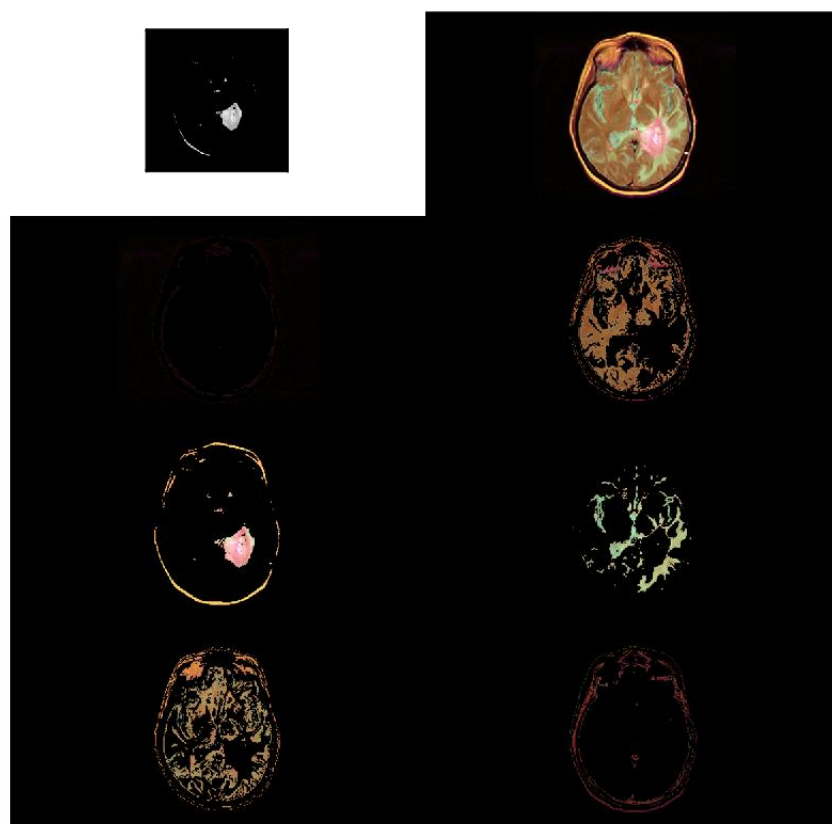


Figure 5: Results for image3



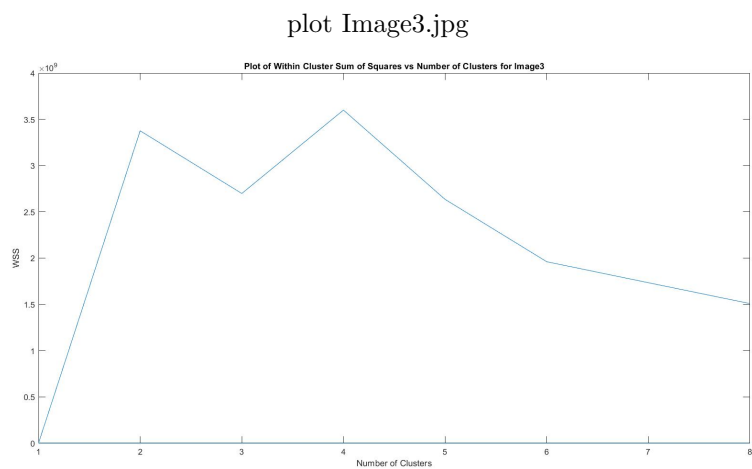


Figure 6: WSS vs number of clusters for image3

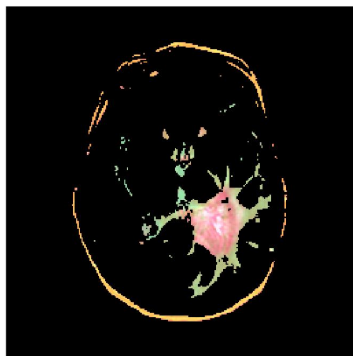


Figure 7: 5 cluster segmentation (image3)

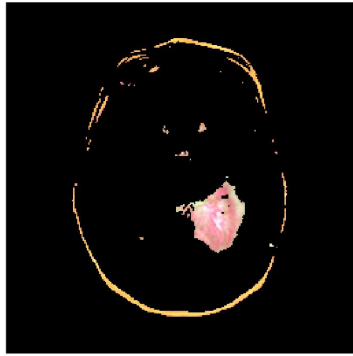


Figure 8: 6 cluster segmentation (image3)

## 5 conclusion

Overall, the results seem reasonable. The biggest issue is most likely from the utilization of FCM in RGB space. Not that it is impossible, but perhaps implementation in another space such as  $l^*a^*b$  would have been more agreeable with the FCM membership function.

A majority of the literature proposed alternative approaches to brain tumor segmentation such as, thresh-holding, region growing, morphological image processing, and rough set approximations[8]. In addition, several studies also modified the FCM algorithm itself to improve the segmentation process[1][2][3][8].

The implementation would have poor run time at substantial input sizes with a run time of at least  $O(cpq)$ , where  $c$  is the number of clusters,  $p$  is the horizontal dimension of the image and  $q$  is the vertical dimension of the image. There is poor reuse quality, but as far as this particular case study goes, the results seem reasonable. It is difficult to judge the results with confidence with the lack of a control group for comparison, however.

The topic is certainly an exciting application of clustering methods, and the biggest issues seem to be with optimizing the accuracy of the pixel count of the tumor area.

## 6 References

- [1] Bezdec, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
- [2] Praylin Selva Blessy, Helen Sulochana, Performance analysis of unsupervised optimal fuzzy clustering algorithm for MRI brain tumor segmentation, IOS Press, 2014.
- [3] M. Sucharita, K. Parimala Geetha, Brain Tissue Segmentation Using Fuzzy Clustering Techniques, IOS Press, 2015.
- [4] Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. "NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set." Journal of Statistical Software 61: 1–36. <http://www.jstatsoft.org/v61/i06/paper>.
- [5] Kaufman, Leonard, and Peter Rousseeuw. 1990. Finding Groups in Data: An Introduction to Cluster Analysis.
- [6] Ketchen, Jr, David J.; Shook, Christopher L. (1996). "The application of cluster analysis in Strategic Management Research: An analysis and critique". Strategic Management Journal. 17 (6): 441–458. doi:10.1002
- [7] Goutte, Cyril; Toft, Peter; Rostrup, Egill; Nielsen, Finn Årup; Hansen, Lars Kai (March 1999). "On Clustering fMRI Time Series". NeuroImage. 9 (3): 298–310. CiteSeerX 10.1.1.29.2679.
- [8] Bhide, A.S., Patil, Priyanka, Dhande, Shraddha. Brain Segmentation using Fuzzy C means clustering to detect tumour Region. International Journal of Advanced Research in Computer Science and Electronics engineering, Vol 1, issue 2, 2012.