Comparison of Segmentation and Detection of Brain Tumor in MRI Images

K.G. Revathi and R. Varatharajan

Department of ECE, DMI College of Engineering, Chennai, India
Department of ECE, Sri Lakshmi Ammal College of Engineering, Chennai, India

Abstract: Tumor is an uncontrolled growth of tissues in any part of the body. As it is known, brain tumor is inherently serious and life-threatening because of its character in the limited space of the intracranial cavity. Most Research in developed countries show that the number of people who have brain tumors died due to the fact of inaccurate detection. Generally, Computed Tomography (CT) scan or Magnetic Resonance Imaging (MRI) that is directed into intracranial cavity produces a complete image of brain. This image is examined by the physician visually for detection and diagnosis of brain tumor. However, this detection method resists the accurate determination of stage and size of tumor. Tumor segmentation from MRI data is an important but time-consuming manual task by medical experts' performance. Automating this process is challenging due to high diversity in appearance of tumor tissue among different patients and, there is a similarity with normal tissue in many cases. There are two types of tumor viz., mass tumor and malignant tumor. In this paper, brain tumor such as mass tumor and malignant tumor is segmented and detected by using K-Means algorithm, Fuzzy C Means algorithm, texture based segmentation using Gray Level Co-occurrence matrix features and combination of Gray Level Co-occurrence matrix features and Gray level Run Length matrix features and all these methods are compared.

Key words: Brain Tumor • k-Means • GLCM • RLF

INTRODUCTION

Brain is the kernel part of the body which has a very complex structure. Brain can be affected by a problem which causes change in its normal structure and its normal behavior. This problem is known as brain tumor. Brain tumor causes the abnormal growth of the cells in the brain. The cells supplying the brain are tightly bound together in the arteries thereby routine laboratory test are inadequate to analyze the chemistry of brain. Computed tomography and magnetic resonance imaging are two imaging modalities that allow the doctors and researchers to study the brain and tumors that are present.

Magnetic resonance imaging (MRI) is an imaging technique used primarily in medical settings to produce high-quality images of the inside of the human body. MRI scan is that they are far superior for collecting images for areas of the body apart from bone. This procedure exposes the body to a strong magnetic field, which affects the hydrogen molecules of water in the body in such a way that a detailed picture can be produced. It does not affect the human body because it does not use any radiation as that of CT. MRI provides rich information about human soft tissues anatomy. MRI helps for diagnosis of the brain tumor.

Images obtained by the MRI are used for analyzing and studying the behavior of the brain. The segmentation of anatomic structures in the brain plays a crucial role in neuro imaging analyses. The complexity of human brain structure mandates the use of computerized approaches derived from computer vision, applied mathematics and image analysis field to extract brain data. However, extracting brain tumors in MR images is often highly challenging due to the convoluted shape, blurred boundaries, inhomogeneous intensity distribution, background noise and intensity contrast between adjacent brain tissues. MRI image acquisition parameters can be adjusted for generating high contrast image with different gray level for various cases of neuropathology. So, MRI image segmentation stands in the upcoming research limelight in medical imaging arena. Segmentation by the medical experts manually from the magnetic resonance images of the brain tumor is very much time-consuming task, tiresome, susceptible to error. Successful
numerical algorithms in segmenting anatomic structures in neuro images can help researchers, radiologists and neurosurgeons to investigate and diagnose the structure and function of the brain in both health and disease.

Work Flow of the Project:

Clusterizing Techniques: Clustering can be considered the most important unsupervised learning problem, so, it deals with finding a structure in a collection of unlabeled data. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. Cluster analysis is an effective method of analyzing and discovering useful information from numerous data. Cluster algorithm groups the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Dissimilarities are assessed based on the attribute values describing the objects. Often, distance measures are used. As a branch of statistics and an example of unsupervised learning, clustering provides us an exact and subtle analysis tool from the mathematic view. K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as centers of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function.

Algorithm for K-means:

Step 1: Take image as a dataset as \( X = \{x_1, x_2, x_3, \ldots, x_n\} \)

Step 2: Give the number of cluster value as k

Step 3: Randomly choose k cluster centers as \( Y = \{y_1, y_2, y_3, \ldots, y_n\} \)

Step 4: Initialize centre value for each cluster

Step 5: Calculate distance between the dataset and the cluster center

In this algorithm Euclidean distance measure is used between dataset values and the cluster center

\[
D(X, Y) = (\Sigma (x_i - y_i)^2)^{1/2}
\]

Step 6: If the distance is near to the center then move to that cluster otherwise move to next cluster

Step 7: The process continues for the whole dataset for k clusters

Step 8: The clustered image is formed

Initialization for Centroid Value

Step 1: The number of cluster chosen as k

Step 2: Calculate range by subtracting minimum intensity from maximum intensity value from the image
Step 3: The incremental step is calculated by dividing the range by number of cluster chosen

Step 4: Centroid is calculated from the incremental step

The aim of k-means algorithm is to partition objects into several classes and to make the distances between objects in the same class closer than the distances between objects in different classes. So if certain centroids in which each centroid represents a group of similar objects can be obtained, we will find out the centroids consistent with the distribution of data.

Fuzzy C Means Algorithm: FCM is a method of clustering which allows 1 piece of data to belong to 2 or more clusters. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. Fuzzy C-Means (FCM) is a data clustering technique in which a dataset is grouped into n clusters with every data point in the dataset belonging to every cluster to a certain degree in fuzzy clustering, a value between zero and one is assigned to each pattern by a membership function

Algorithm for Fuzzy C Means: The Fuzzy C-Means clustering algorithm is an iterative clustering method that produces an optimal c partition by minimizing the weighted within group sum of squared error objective function \( J_{km} \).

\[
J_{km} = \sum_{k=1}^{c} \sum_{i=1}^{n} (u_{ik})^q \cdot d^2(x_i, v_k)
\]  

(2)

where,

- \( n \) is the number of data items in the dataset
- \( c \) is the number of clusters with \( 2 \leq c < n \)
- \( u_{ik} \) is the degree of membership of object \( x_i \) in the \( i^{\text{th}} \) cluster
- \( q \) is the weighting exponent on each fuzzy membership used to adjust the fuzzy degree
- \( v_k \) is the prototype of center of cluster
- \( d^2(x_i, v_k) \) is a distance matrix measure between object \( x_i \) and cluster center
- \( \varepsilon \) is the minimum amount of improvement for termination process

A solution of the object function \( J_{km} \) can be obtained via an iterative process, which is carried out as follows:

Step 1: Set values for \( c, q \) and \( \varepsilon \).

Step 2: Initialize the fuzzy partition matrix \( U = [u_{ik}] \).

Step 3: Set the loop counter \( b = 0 \).

Step 4: Calculate the \( c \) cluster centers

\[
v_k = \frac{\sum_{i=1}^{n} (u_{ik})^q \cdot x_i}{\sum_{i=1}^{n} (u_{ik})^q}
\]

(3)

Step 5: Calculate the membership \( U(b+1) \).

\[
u_{ik} = \left( \frac{1}{\sum_{j=1}^{c} \left( \frac{1}{d^2(x_i, v_j)} \right)^{\frac{1}{q-1}}} \right)^{\frac{1}{q-1}}
\]

(4)

Step 6: If \( U(b) - U(b+1) \leq \varepsilon \), stop; otherwise, set \( b = b + 1 \) and go to step 4.

Formation of Fuzzy Cluster: The function parts for the fuzzy clustering analyzer are separated to 5 parts as follows. The first one is initialization, which generates initial fuzzy partition matrix for clustering. The second one is \( u' \), which gives the matrix after exponential modification. The third part is E-step is used to get the new center matrix, as described in formula. The forth part is distance calculation, which calculates the distance of the cluster center with input feature data. In general case, the Euclidean distance is used. The inverse distance can also be used. The fifth part is M-step, which get new fuzzy partition matrix and cost function value. Among them, the cost function value is used to control the iterations in the implementation. The final fuzzy partition matrix \( U \) is the result that we want, which includes the information of cluster and with dimension \((K \times n)\). \( K \) is the cluster number and \( n \) is the number of the feature data.

Texture Based Segmentation: Texture can be defined as a regular repetition of an element or pattern on a surface. Texture quantifies local contrast (gray level differences) and local spatial structure. Statistical methods analyze the spatial distribution of gray values, by computing local features at each point in the image, and deriving a set of statistics from the distributions of the local features which is one of the defining qualities of texture. Depending on the number of pixels defining the local feature, statistical methods can be further classified into first-order (one...
pixel), second-order (two pixels) and higher-order (three or more pixels) statistics. The basic difference is that first-order statistics estimate properties (e.g. average and variance) of individual pixel values, ignoring the spatial interaction between image pixels, whereas second-order and higher-order statistics estimate properties of two or more pixel values occurring at specific locations relative to each other.

**Gray Level Co-Occurrence Matrix Based Features:** The most popular second-order statistical features for texture analysis are derived from the so-called co-occurrence matrix. It gives effective texture discrimination in biomedical images. Histogram based features are local in nature. This feature does not consider spatial information into consideration. So for this purpose gray-level spatial co-occurrence matrix \( h(i,j) \) based features are defined which are known as second order histogram based features.

Second order histogram based features are based on the joint probability distribution of pairs of pixels. Distance \( d \) and angle \( \theta \) within a given neighborhood are used for calculation of joint probability distribution between pixels. Normally \( d=1, 2 \) and \( \theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ \) are used for calculation. Instead of using angular features directly, because there are two instances where two horizontally adjacent pixels have the values 1 and 2, functions viz., energy and entropy’s average and range can also be used as inputs. Texture features can be described using this co-occurrence matrix. The features which defines this matrix are:

- Angular Second Moment (Energy)
- Correlation
- Inertia
- Absolute value
- Inverse Difference
- Entropy
- Maximum Probability

In this paper, Energy and Entropy have been concentrated for segmentation of brain tumor from MRI image.

**Steps for Segmentation Based on Texture Features**

**Step 1:** The input image is converted to gray scale image.

**Step 2:** Co-occurrence matrix is formed based on joint probability distribution of pairs of pixels.

Co-occurrence matrix calculation is calculated by how often a pixel with gray-level (grayscale intensity) value \( i \) occurs horizontally adjacent to a pixel with the value \( j \). Each element \((i,j)\) in this matrix specifies the number of times that the pixel with value \( i \) occurred horizontally adjacent to a pixel with value \( j \) from the image. Gray level Co-occurrence matrix is the size of the image. Here spotting the tumor is based on neighborhood pixel analysis. It is illustrated in Fig. as follows.

In the output GLCM, element \((1,1)\) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. GLCM \((1,2)\) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element \((1,3)\) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. It is named as \( p(i,j) \).

In the output GLCM, element \((1,1)\) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. GLCM \((1,2)\) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element \((1,3)\) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values 1 and 3. It is named as \( p(i,j) \).

GLCM show how often each gray level occurs at a pixel located at a fixed geometric position relative to each other pixel, as a function of the gray level. The \((1,2)\) entry in a matrix for right neighbors, for example, would show the frequency or probability of finding gray level 2 immediately to the right of pixel with gray level 1. The number of threshold levels is \( 8 \). The 2 in the co-occurrence matrix indicates that there are two occurrences of a pixel with gray level 2 immediately to the right of pixel with gray level 1. The size of co-occurrence matrix will be the number of threshold levels. When we consider neighboring pixels, the distance between the pair of pixels is 1. However, each different relative position between the two pixels to be compared creates a different co-occurrence matrix.
Step 3: Determine Energy from Gray Level Co-occurrence matrix to define texture feature by using this equation,

\[ E = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(i,j)]^2 \]  

(5)

Step 4: Determine Entropy from Gray Level Co-occurrence matrix to define texture feature by using this equation,

\[ H = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \log_2[p(i)] \]  

(6)

where

\[ P(i) = \frac{h(i)}{NM}, I=0,1,\ldots,G-1 \]  

(7)

Step 5: To determine Threshold point coordinates \((t_1, t_2)\),

Step 5.1: Find maximum and minimum of Energy from the Energy matrix as \(E_{\text{max}}\) and \(E_{\text{min}}\) respectively.

Step 5.2: Find maximum and minimum of Entropy from the determined. From Co-occurrence probability features, from energy and entropy features are determined. From these Co-occurrence probability and run length features common seed point is determined and proceeded to spot the tumor based on neighbourhood pixel analysis.

Step 5.3: Threshold coordinates are determined by using following equations as

\[ t_1 = (E_{\text{min}} + \text{Entropy}_{\text{max}})/2 \]
\[ t_2 = (E_{\text{max}} + \text{Entropy}_{\text{min}})/2 \]

Step 6: After finding out the threshold coordinates, by keeping this point as seed point the steps are proceeded to spot the tumor based on neighbourhood pixel analysis given under V.

Gray Level Run Length Features: Gray level runs can be characterized by the gray tone of the run, the length of the run and the direction of the run. The length of the run is the number of picture points (pixels) in the run.

The major reason is that the length of the runs reflect the size of the texture elements. Using the run length matrix, five features can be computed viz.,

- Short run emphasis
- Long run emphasis
- Run length Non Uniformity
- Gray Level non-uniformity
- Run percentage

Among these features, the following two features give the good difference between homogenous and non-homogeneous regions are considered.

\[ \text{Long run emphasis} = \frac{\sum_{i=1}^{G} \sum_{j=1}^{R} j^2 P(i,j)}{\sum_{i=1}^{G} \sum_{j=1}^{R} P(i,j)} \]  

(8)

\[ \text{Run length Non Uniformity} = \frac{\sum_{i=1}^{G} \sum_{j=1}^{R} (P(i,j))^2}{\sum_{i=1}^{G} \sum_{j=1}^{R} P(i,j)} \]  

(9)

where

\( P(i,j) \) – run length matrix

\( G \) - Number of gray levels

\( R \) - Longest run

The long run emphasis is high for homogeneous region and the run length non-uniformity is low for non-homogeneous.

From these two features seed point co ordinates are determined. From Co-occurrence probability features, from energy and entropy features are determined. From these Co-occurrence probability and run length features common seed point is determined and proceeded to spot the tumor based on neighbourhood pixel analysis.

Steps for Tumor Detection Using Neighbourhood Pixel Analysis:

a) Convert the rgb image to grayscale image
b) Typecast the input image from uint8 to double class for detailed analysis
c) Set the threshold value to get the neighbour pixel
d) Seed = get the pixel intensity from abnormal region using coordinate.
e) Initialize the output image as zeros matrix and set value 1 at seed coordinate in output matrix.
f) Count and suit variables for counting the pixels at abnormal region and update the seed values.

g) Initialize the loop for growing process uptotumor detection

h) Initialize the ‘for’ loop for getting row and col coordinates from input matrix and fill ‘1’ in corresponding coordinate of output

i) From the coordinate of nonzero pixel value (at seed coordinate), create 3*3 matrix around it if row and col are lesser than size of input image.

j) After that, within 3*3 mask we have to find pixel intensity difference between initial seed pixel value and neighbourhood pixels and if the difference is less than threshold value then fix ‘1’ at corresponding coordinates of output matrix.

k) Count the pixels and sum the seed value for updating process finally.

l) This process will be continued for all neighbourhood pixels and loop will be terminated if the intensity difference is greater than the threshold value. Segmented tumor is obtained.

RESULTS AND DISCUSSIONS

CONCLUSION

In this paper, Brain tumor of both mass and malignant tumor in MRI images are segmented using K-Means algorithm, Fuzzy C Means algorithm, Gray Level Cooccurence matrix features, combination of both Gray Level Cooccurence matrix features and Gray Level Run Length matrix features of texture based segmentation. The results are then compared.

REFERENCES


4. Jie Wu1, Skip Poehlman1, Michael D. Noseworthy and Markad V. Kamath, 2008. Texture Feature based Automated Seeded Region Growing in Abdominal MRI Segmentation in International Conference on Biomedical Engineering and Informatics


7. Qurat-Ul-Ain, Ghazanfar Latif, Sidra Batool Kazmi, M. Arfan Jaffar and Anwar M. Mirza, XXXX. Classification and segmentation of Brain Tumor using Texture Analysis, in proc. 9thWSEAS Int. Conf. on Recent Advances in Artificial Intelligence, Knowledge Engineering and Databases, pp: 147-155.

