

AUTOMATIC TUMOR SEGMENTATION USING OPTIMAL TEXTURE FEATURES

M.Sasikala¹, N.Kumaravel² and L.Subhashini²

¹Department of Instrumentation Engineering, Madras Institute of Technology, Anna University, Chromepet, Chennai – 600 044, India.

²Department of Electronics and Communication Engineering, College of Engineering, Guindy, Anna University, Chennai – 600 025, India.

sasi_yugesh@yahoo.com, kumaravel_n@annauniv.edu

Abstract: This paper presents an automatic segmentation of malignant tumor in magnetic resonance images (MRI's) of brain using optimal texture features. Texture features are extracted from normal and tumor regions (ROI) in the brain images under study using spatial gray level dependence method and wavelet transform. The normal and tumor regions are classified using an artificial neural network. A very difficult problem in classification techniques is the choice of features to distinguish between classes. In the proposed method, the optimal texture features that distinguish between the brain tissue and malignant tumor tissue is found using genetic algorithm (GA). The optimal features are used to segment the tumor. The proposed feature based segmentation technique is compared with few existing techniques. The performance of the algorithm is evaluated on a series of brain tumor images. The results illustrate that the proposed method outperforms the existing methods.

Keywords: Texture features, spatial gray level dependence method, wavelet transform, artificial neural network, genetic algorithm.

INTRODUCTION

Practical pattern classification and knowledge discovery problems require selection of a subset of attributes or features to represent the patterns to be classified. Exhaustive evaluation of possible feature subsets is usually infeasible in practice because of the large amount of computational effort required. Genetic algorithms (GA), which belong to a set of randomized heuristic search techniques, offer an attractive approach to find near-optimal solutions to such optimization problems. This paper presents an approach to feature subset selection using GA. The proposed image analysis process is outlined in fig. 1.

Computer based methods of texture analysis were originally developed for use in satellite applications, geological surveys, remote sensing and other related applications. Studies of intracranial tumors have demonstrated that MR image texture may be used to determine the tumor type [1].

Image → Selection → Feature → Dimensionality → Classifier
of ROI Extraction reduction

Fig. 1: The Image Analysis process

MR Image texture analysis is proved to be useful in the detection of Alzheimer's disease [2]. The main objective of this paper is the design of an automated system for the segmentation of glioblastoma-multiforme tumor in the digital images of brain. Glioblastoma-multiforme represents the highest grade of malignant astrocytoma and is also the most common primary brain neoplasm. It has a nonhomogeneous structure. The normal and tumor regions are classified using texture features from spatial gray level dependence method (SGLDM) and wavelet features. The optimal texture features are found using GA. The optimal features are used to segment the glioblastoma-multiforme tumors in MR images of brain.

TEXTURE FEATURE EXTRACTION

A set of ten texture features are extracted from the normal and the tumor regions. The methods for extracting these features are explained below.

Gray – Tone spatial dependence matrix

A gray tone spatial dependence matrix approach, introduced by Haralick [3] which is a well known statistical method for extracting second order texture information from images, is used for this study. This method is based on the estimation of the second order joint conditional probability density function $C(i, j / d, \theta)$ where $\theta = 0, 45, 90$ and 135 degrees. Each $C(i, j / d, \theta)$ is the probability of going from gray level i to gray level j , given that the inter-sample spacing is d and the direction is given by the angle θ . This is also referred to as cooccurrence matrix. The cooccurrence matrix is calculated for the normal and tumor regions (ROI) in the brain images for $\theta = 0$ degrees and distance $d=1$. fig 2 shows the region of interest.

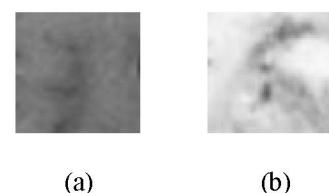


Fig 2: Region of Interest (a) Normal region (b) Tumor region

Eight texture features are calculated from the cooccurrence matrix. These features are named as follows: Angular Second Moment (ASM), Contrast (CON), Inverse Difference Moment (IDM), Sum Variance

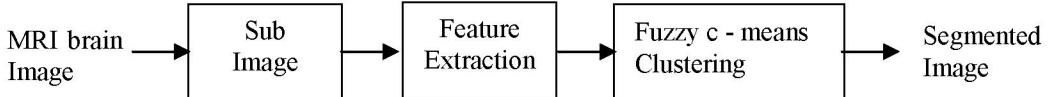


Fig.3 Block diagram of feature-based segmentation

(SVAR), Sum Entropy (SENT), Entropy (ENT), Difference Entropy (DENT) and Information Measure of Correlation (IMC).

Wavelet feature extraction

The ROI is decomposed using 2D wavelet transform into four subbands: low-low (LL) band, low-high (LH) band, high-low (HL) band and high-high (HH) band. Daubechies wavelet filter of order two is used. The LL-band texture is a smoothed version of the input texture, while the LH-band, the HL-band and HH-band textures are high pass filtered textures in the x direction, y direction, and both x and y directions respectively. The energy of the LL subband is calculated [4]. If $x(m,n)$ is the LL subband with $1 \leq m \leq M$ and $1 \leq n \leq N$ the energy is

$$\text{Energy} = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |x(m,n)|^2. \quad (1)$$

The energy feature is calculated for the normal and tumor regions in the brain images.

FEATURE SELECTION

The main purpose of feature selection is to reduce the number of features used in classification while maintaining acceptable classification accuracy.

The Genetic Algorithm Approach

For selecting the best set of features, GA based global search method is used. The algorithm is based on Darwin's fittest principle, which states that an initial population of individuals evolves through natural selection in such a way that the fittest individuals have a higher chance of survival. In the proposed method each feature subset is encoded in a binary string called chromosome with as many binary-coded genes as the feature space dimension. Each bit in the binary vector is associated with a feature. If the i^{th} bit of this vector equals 1 then the i^{th} feature is allowed to participate in classification. If the bit is a 0, then the corresponding feature does not participate. Each resulting subset of features is evaluated according to its classification accuracy on a set of test data [5]. Fitness value is assigned based on a performance measure in classification. e.g. the classification error.

SEGMENTATION METHODS

Optimal Feature based segmentation

The block diagram of the proposed method for segmentation of malignant tumor is given in fig.3. The glioblastoma multiforme tumor image is divided into overlapping sub-blocks of size 3×3 . For each of these sub-blocks, gray level co-occurrence matrix is obtained. The optimal texture features such as Inverse Difference Moment, Entropy and Energy are extracted from each of the sub blocks.

The optimal features are given as input to fuzzy c-means clustering algorithm and the number of classes are chosen to be two. One is normal brain tissue, and the other is malignant tumor. But due to block processing, the boundary of segmented image suffers from stair case effect. This is overcome by using post-processing techniques.

Segmentation of tumor using region growing

Region growing is a procedure that groups pixels or sub regions into larger regions. In this method, the brain image is converted to binary image using a threshold. The threshold is computed using Otsu's method [6], which chooses the threshold to minimize the intraclass variance of the thresholded black and white pixels. In the binary image, a pixel (seed) representing tumor region is taken as starting point. The seed pixel is grown along the eight connected neighbors. In the region growing process, only the homogeneous region in tumor grows leaving behind the heterogeneous regions. This is a drawback of region growing.

Segmentation of tumor using Fuzzy c-means algorithm

The pixel intensities of the image is given as input to the fuzzy c-means clustering (FCM) algorithm. The FCM clustering algorithm assigns a fuzzy membership value to each data point based on its proximity to the cluster centroids in the feature space [7]. Formally FCM takes a finite dataset X as an input, each $x_i \in X$ is a feature vector; $x_i = [x_{i1}, x_{i2}, \dots, x_{is}]$, where x_{ij} is the j^{th} feature of the i^{th} subject and s is the dimensionality of x_i . Functions $u_k: X \rightarrow [0,1]$ are defined; they assign to each x_i in X its grade of membership in the fuzzy set u_k . A function u_k describes a fuzzy subset of X . The goal is to partition X by

means of fuzzy sets. Together, the functions u_k yield a fuzzy c -partition U , which can be represented as a $c \times n$ matrix. The FCM algorithm consists of an iterative optimization of an objective function J_m . The objective function that is minimized is given by

$$J_m(U, v) = \sum_{i=1}^n \sum_{j=1}^c (u_{ji})^m (d_{ji})^2. \quad (2)$$

where $v = (v_1, v_2 \dots v_c)$ with v_j being the cluster center of class j , $1 \leq j \leq c$, and $d_{ji}^2 = ||x_i - v_j||^2$. The parameter m determines the fuzziness of the result. Given a partition, the cluster centers are calculated using

$$v_j = \frac{\sum_{i=1}^n u_{ji}^m x_i}{\sum_{i=1}^n u_{ji}^m}, \quad 1 \leq j \leq c \quad (3)$$

The iteration is completed by calculating the new partition:

$$u_{ji} = \left[\sum_{k=1}^c \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad (4)$$

$$1 \leq j \leq c : 1 \leq i \leq n$$

The steps described by equation (3) and equation (4) are repeated until the Euclidean distance between U matrices at successive iterations falls below a threshold ϵ .

$$\|U_{k+1} - U_k\| < \epsilon. \quad (5)$$

Clustering can result in a large number of misclassifications, particularly when pixel gray level alone is used as a feature.

RESULTS AND DISCUSSION

Nineteen scans containing glioblastoma multiforme and eighteen scans containing normal brain images are used for the study. A ROI is selected in the normal and tumor regions of the brain images. The texture features such as ASM, CON, IDM, SVAR, SENT, ENT, DENT, IMC are extracted from ROI using spatial gray level dependence method. The feature, ‘energy’ is extracted using wavelet transform. The mean gray value of the ROI is also included in the feature set.

The feature set containing ten texture features is used as the input of the neural network. The neural network is trained using back propagation algorithm. Due to the small size of the dataset, the neural network classifier is employed by means of the leave-one-out method. This procedure takes N training samples, trains the classifier on $N-1$ samples, then uses the remaining one sample to test. Classification is continued in this manner until all N samples have been used as the test sample. Final performance is reported as an average of results for all N trials. Thus the normal region and the glioblastoma multiforme are classified. The accuracy obtained using the entire feature vector before optimization is presented in table 1.

Feature set	Classifier accuracy
10 features	97.3%

Table 1: Classification performance of the entire feature set

The optimal feature set is found using GA. A population of twenty chromosomes is randomly generated. Each chromosome contains ten genes. One gene for each feature. The genetic operators, two point crossover and mutation are used. The crossover rate is 90% and mutation rate is 10%. Roulette wheel selection method is used to select the mating pool, which holds the parents that generate offsprings. The inverse of the classification error is used as fitness function. The algorithm did not converge to a single set of features but to several sets. The ten best chromosomes the algorithm found during the execution is presented in table 2.

The texture feature ASM is present in most of the feature vectors selected by GA. The features such as DENT and MEAN which are not selected are found to be less significant. The feature IDM, ENT and Energy classifies the normal and tumor regions with an accuracy of 97.3%. The classification accuracy of 97.3% is obtained with three of the available ten features using GA. Therefore it is possible to classify the normal and tumor images with minimum number of features. Thus the cost of classifier can be reduced. Hence the optimal features are used to segment glioblastoma-multiforme tumor in brain images.

Feature set	Classifier accuracy
IDM, ENT, ENERGY	97.3%
ASM, CON, IDM, SVAR	97.3%
ASM, IDM, SVAR, SENT	97.3%
CON, IDM, SVAR, IMC	97.3%
ASM, CON, SENT IMC	97.3%
ENT, IMC, ENERGY, MEAN	97.3%
ASM, CON, SVAR, IMC	97.3%
ASM, SVAR, IMC, ENERGY	97.3%
ASM, SVAR, SENT, DENT	97.3%
ASM, SVAR, ENT, IMC	97.3%

Table 2: Best chromosomes selected by GA

The glioblastoma-multiforme tumor image is divided into subblocks of size 3×3 . The optimal texture features IDM, ENT and Energy are extracted from each subblock. The extracted features are given as input to the FCM clustering algorithm and are classified into malignant tumor and normal brain tissue. The results obtained from this method are compared with region growing and FCM algorithm. The performance measures such as Sensitivity, Specificity and Accuracy are calculated for each of these methods and are presented in table 3.

Segmentation Methods	Sensitivity	Specificity	Accuracy
Region growing	54.58%	97.34%	94.29%
Fuzzy c means	80.76%	93.35%	94.65%
Feature based segmentation	91.99%	98.96%	98.45%

Table 3: Performance measures of different segmentation algorithms

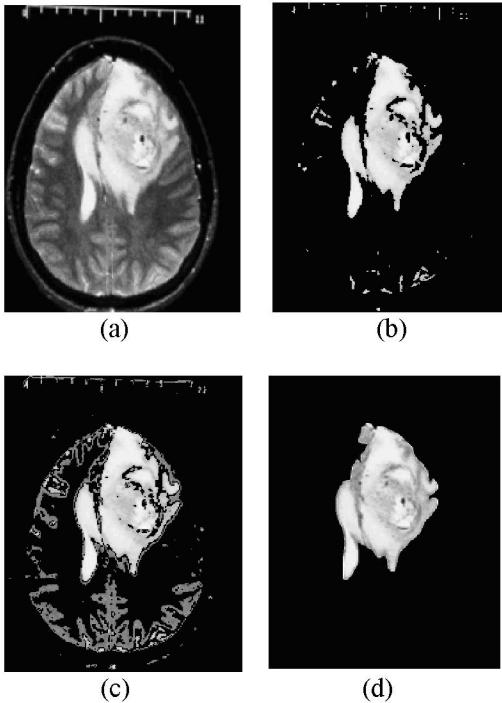


Fig.4: Comparison of segmentation results on a brain MRI with glioblastoma-multiforme tumor (a) Original image (b) using region growing (c) using FCM (d) using optimal feature based segmentation

The segmentation results of various methods are shown in fig.4. The results in Table.3 and Fig.4 shows that in the region growing method, the true positive (TP) value is low because it discards heterogeneous regions within the malignant tumor and grows only the homogeneous regions. Region growing technique cannot classify unconnected parts of an image as the same region. FCM algorithm overcomes this limitation by considering all pixels concurrently. However FCM algorithm results in a large number of misclassifications, since only the pixel intensities are used as input features to the classifier. But in feature based segmentation technique, the tumor region which is a combination of homogeneous and heterogeneous region is segmented with high accuracy using three optimal texture features extracted from each sub-block of the image. The optimal feature based segmentation technique produces better segmentation accuracy than other methods.

CONCLUSION

In this paper, an automatic method of segmentation of malignant tumor using optimal texture features is proposed. The optimal features that distinguish between the brain tissue and malignant tissue is found using GA. The optimal features are used to segment the glioblastoma multiforme tumor in MR images of human brain. This method segments the tumor, which has both homogeneous and heterogeneous characteristics, using optimal features. The proposed method is compared with few other existing segmentation techniques. The results show the high performance of the proposed method and its superiority to the existing methods.

REFERENCES

- [1] Zuna I., Schad L.R. and Bluml S. (1993): 'MR tissue characterization of intracranial tumors by means of texture analysis', *Magnetic Resonance Imaging*, 11, pp. 889-896.
- [2] Nick C. Fox and Peter A. Freeborough. (1998): 'MR Image Texture analysis Applied to the diagnosis and tracking of Alzheimer's disease', *IEEE Trans. on Medical Imaging*, 17(3), pp.475-479.
- [3] Dinstein, Haralick M. and Shanmugam. (1973): 'Textural Features for Image Classification', *IEEE Trans. on Systems, Man and Cybernetics*, 3(6), pp.610-621.
- [4] Jay Kuo, C.C, Tianhorng Chang. (1993): 'Texture Analysis and Classification with Tree Structured Wavelet Transform', *IEEE Transactions on Image processing*, 2(4), pp.429-441.
- [5] Goodman E.D., Kuhn L.A., Jain A., Raymer M.L. and Punch W.F. (2000): 'Dimensionality reduction using genetic algorithms', *IEEE Trans. On Evolutionary Computation*, 2, pp.164-171.,
- [6] Otsu N. (1979): 'A threshold selection method from gray level histograms', *IEEE Trans. Systems Man and Cybernetics*, 9, pp.62-66.,
- [7] Timothy J. Ross. (1995): 'Fuzzy logic with engineering applications', Tata McGraw-Hill.