International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.3 (2016), pp.37-44
http://dx.doi.org/10.14257/ijmue.2016.11.3.04

Tumor Diagnosis Based on the GMM Feature Decision Classification of Brain MR Images

Yang Li[1], Ye Mingquan[1] and Zhang Hao[1]

1. Department of Basic Public, WanNan Medical College, Wuhu Anhui 241002, China

Abstract

The MR image has provided lots of information used for medical examination. Accurate and robust brain MR image segmentation, feature extraction and feature classification are very important for clinical tumor diagnosis. A new tumor diagnosis method based on brain MR images is hereby put forward. Firstly, detect the deformed area of the images through multi-threshold segmentation morphology, and then, extract the GMM feature used for the classification, and finally, classify the types of tumor images by using decision tree classifier. The whole classification consists of two stages, during training stage extract the different features of tumor images and non-tumor images, and during testing stage conduct the classification of tumor and non-tumor based on knowledge databank. The computing method is appraised according to the three performance index including accuracy, false alarm rate and loss detecting rate, the experiment results show that the computing function is excellent and is helpful for better brain tumor diagnosis.

Keywords*: Brain Tumor; MR Images; Segmentation; GMM Feature; Decision Tree*

1. Introduction

The MRI plays a vital role in the field of medical imaging, provides qualitative, quantitative and accurate medical diagnosis information. It has a superior advantage than other modal medical imaging technologies in many application fields, such as angiocarpy, nerve and muscle, especially in the field of brain imaging. However, the most of the MR image processing problems arise from the strength change which is caused by the unevenness of B1 and B0 wave field. The unevenness, even a single tissue may mislead lots of image analysis algorithms, especially in the image segmentation stage.

In the past decade, the malignant tumor diagnosis based on MRI has aroused the wide concern. Cancer is one of the most common diseases that influence the human health, and its feature is the control-free cell proliferation, especially the primary malignant brain tumors. Unluckily, lots of the seeking work for new treatment methods failed. Based on this, many recent research work focus on the development of a stronger MR image computer-aided detecting and analyzing system. The robust MR image analysis includes several sequent steps: the system must detect and extract the deformed area from the surrounding medium through multi-threshold segmentation technology and morphological image processing. This step needs to process the high-resolution gray MR images by selecting a suitable segmentation technology. The segmentation is often the first step of image processing. Several segmentation technologies based on MR image have been put forward, besides, there are also some methods removing the noise through filtering. But these technologies are not applicable to MR images in general because the key feature of brain tumor may be removed due to carelessness. The first segmentation is based on the threshold or multi-threshold technology. The threshold segmentation method adopts multi threshold to divide the image into some areas. When the deformed area of an image is detected, the system will extract some feature reference and textural features in order to

ISSN: 1975-0080 IJMUE

identify the type of brain tumor. The texture analysis decides the pixel points surrounding the texture elements based on the gray-level co-occurrence matrix.

In this article, we have put forward to study the brain tumor's features and those possible measuring features. For this, we focus on the analysis on GMM features obtained from weighted T1, T2 and the FLAIR MR images. The GMM method based methods have been mentioned in the literatures related to face identification, the literatures have a method that can provide the best compromise among complication, robustness and distinguishing degree, in addition, the literature conducts identification of speakers through feature and score normalization technology. Different from others, this article uses GMM in order to extract the sub-feature of each GMM. The decision tree is used for the appraisal of the classifying technology of GMM feature's function, during our study, it is reflected on the brain tumor and the normal brain MR image's identifying ability. This article conducts appraisal on the algorithm's function by distinguishing the tumor images from non-tumor MR images.

2. This Article's Methods

The brain MR image tumor diagnosis method flow chart put forward in this article is shown as per the Figure 1.

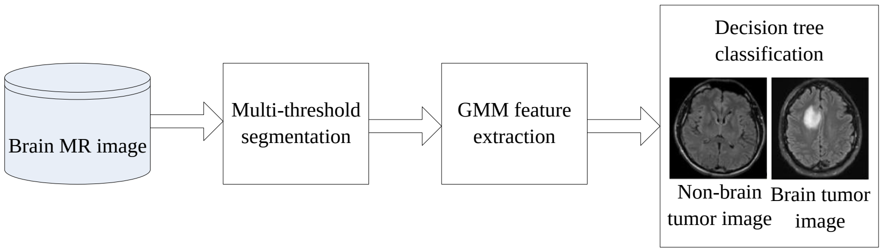


Figure 1. Flow Chart

2.1. Multi-Threshold Segmentation

The thresholding segmentation is a traditional image segmentation method, it is especially applicable to the images of which the object and background occupy different gray levels because it is easy to be realized, with small computing quantity and stable function. It can not only greatly compress data size, but also greatly simplify the analyzing and processing step and it is the necessary image pre-processing procedure before image analysis, feature extraction and mode identification. Its aim is to, according to the gray level, divide the pixels, and each obtained sub-pixels segmentation forms an area corresponding to the real scene, and the various internal attributes accord with each other, and the neighboring area distribution has such according attribute. Such segmentation can be realized by starting from the gray level and selecting one or more thresholds.

The histogram analysis is often used for the statistics of estimation, for example, the average distributed value and variance. This article uses the three modes T1, T2 and FLAIR of the MR images to detect the brain tumor. Based on the single- threshold segmentation, the many-point area is common in the brain tumor image. This article adopts multi-threshold segmentation to solve this problem. According to this, we carry out preliminary estimation by adopting the threshold-based judging method put forward in the literature. Its main idea is to seek for the threshold which can make the interclass variance max and the intraclass variance minimum. When giving a 2-D image with gray levels from 1 to L, it should contain N pixels. When the gray level is i, the pixel point volume form is shown as f_i and the probability of the gray level i is expressed as:

$$P_i = f_i / N \tag{1}$$

The two thresholds of an image divide pixel points into two types, the C_1 gray level [1, …, t], the C_2 gray level [t+1 , …, L]. In such condition, the two types of probabilities are expressed as follows:

$$C_1 = \frac{p_1}{w_1(t)}, \frac{p_2}{w_1(t)}, \text{n} \frac{p_t}{w_1(t)} \tag{2}$$

$$C_2 = \frac{p_{t+1}}{w_2(t)}, \frac{p_{t+2}}{w_2(t)}, \text{n} \frac{p_L}{w_2(t)} \tag{3}$$

Of which, the two $w_1(t) = \sum_{i=1}^{t} P_i$, $w_2(t) = \sum_{i=t+1}^{L} P_i$, $C_1(\mu_1(t))$ and $C_2(\mu_2(t))$'s average expression is as follows respectively：

$$\mu_1(t) = \sum_{i=1}^{t} i \cdot \frac{P_i}{w_1(t)} \tag{4}$$

$$\mu_2(t) = \sum_{i=t+1}^{L} i \cdot \frac{P_i}{w_2(t)} \tag{5}$$

The whole image's average strength (μ_T) is expressed as follows:

$$\mu_T = w_1\mu_1 + w_2\mu_2 \tag{6}$$

Of which $w_1 + w_2 = 1$. According to literature description, the threshold image's interclass variance (σ^2) is expressed as follows:

$$\sigma^2 = w_1(\mu_1 - \mu_T)^2 + w_2(\mu_2 - \mu_t)^2 \tag{7}$$

According to formula (7), the best threshold (t) selects value $1 \le t < L$ that can make the variance (σ^2) max. In order to find and detect the brain tumor in an image with strong strength, we can expand the image into multi-threshold according to the above formula. The image 2 shows a brain MR image and its single-threshold and double-threshold effect segmentation effect image.

(a) Original Image (b) Single-Single-Threshold (c) Multi-Threshold Segmentation

Figure 2. Segmentation Effect Diagram

2.2. GMM Feature Extraction

GMM is an extension of single Gaussian probability-density function, it can be expressed as the weighted sum of k Gaussian density and it is detailed as follows:

$$p(x|\lambda) = \sum_{i=1}^{k} w_i g(x|\mu_i, \Sigma_i) \tag{8}$$

Of which, x is the data vector of N-dimension sequent digit, w is hybrid weights and g is sub-item Gaussian density. Each sub-item density is controlled by the N Gaussian function shown in the formula (9).

$$g(x|\mu_i,\Sigma_i)=\frac{1}{\sqrt{(2\pi)^N|\Sigma_i|}}\exp\left\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right\} \tag{9}$$

Of which, μ_i is mean vector, Σ_i is covariance matrix. The complete GMM is obtained by the parameterization of mean vector, covariance matrix and all the sub-item strength hybrid weights. These parameters are uniformly expressed as:

$$\lambda=\{w_i,\mu_i,\Sigma_i\} \tag{10}$$

The diagonal of covariance matrix Σ_i stands for the component variance v_i . We adopt 5 components which form the GMM, the feature vector R is expressed as:

$$R=\{w_1\dots w_5,\mu_1\dots\mu_5,v_1\dots v_5\} \tag{11}$$

Each divided brain tumor area is expressed by the 1-dimension feature vector R_Y which contains 15 elements, its matrix and divided normal area is similar in size. As for the n brain tumor areas, the brain tumor feature matrix Rt_Y is n*15 in size. The brain normal area matrix Rt_N is similar to it. The classified operation is prepared and ready by calculating the matrix Rt_Y and Rt_N .

2.3. Decision Tree Classifier

The decision tree can be regarded as a tree-shape prediction model and it is the hierarchical structure composed of nodes and directed edges. The tree includes 3 kind of nodes: root nodes, internal nodes and leaf nodes. The decision tree only has one root node, which is the collection of all the training data. Every internal node in the tree is a fission issue; the certain type of test is designated, and it will meet the sample requirement and be divided according to a certain attribute. In addition, every subsequent branch of this node is corresponding to a possible value of the attribute. Every leaf node is the collection of data with classification label, that is, the classification of the example.

The decision tree has lots of algorithms, for example: ID3, C4.5, CART and others. All these algorithms adopt greedy algorithm from top to bottom, each internal node selects the attribute with the best classification effect, and it can be divided into two or more sub-nodes, and this procedure will be continued until the decision tree can classify all the training data accurately, or all the attributes are used. All the literatures have put forward the optimal decision tree, of which the latter two include two conception periods: growth and trim. Other researchers all focus on the growth period only. This article uses one method that automatically construct a decision tree according to the training collection given. Our goal is to find the optimal decision tree with the minimized and generalized errors. One of the most important aspects of the decision tree induced strategy is the split criterion, which is a method of attribute-selection test and decides the distribution of the subset of training object, and the sub-tree arises hereby. This operation uses a built-in Matlab function to execute regression and classification.

Because the available sample data are limited, this article adopts cross validation. Partial samples of the data (including n-1 brain tumor images, n-1 normal images) are used to conduct algorithm training, the rest 1 brain tumor image and 1 normal image are used for appraisal of the classification accuracy of algorithm. We repeat the procedure for every brain tumor and normal image group. We measure the accuracy of the classifier through three performance index, separately are accuracy, false alarm rate and loss detecting rate, which can be expressed by the following formula:

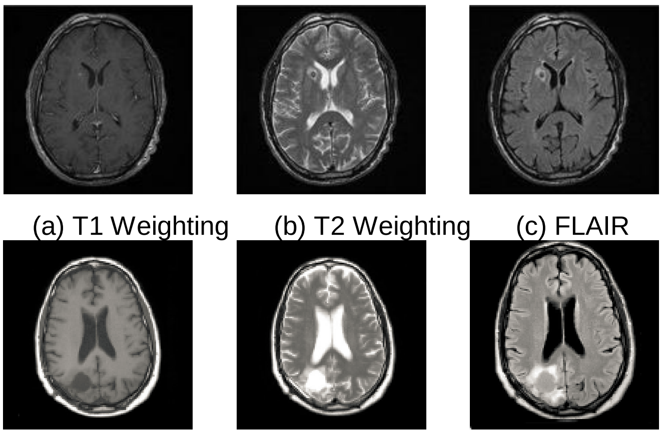
$$\text{False alarm rate}=\frac{\text{wrongly classified number of normal sample}}{\text{total sample number}} \tag{12}$$

$$\text{Loss detecting rate} = \frac{\text{wrongly classified number of brain tumor image samples}}{\text{total sample numbers}} \tag{13}$$

$$\text{Accuracy} = \frac{\text{normally classified sample number}}{\text{total sample number}} \tag{14}$$

3. Experiment Result and Analysis

This article has made experimental analysis on 34 images containing brain tumor and 34 normal brain T1 weighting, T2 weighting and FLAIR MR images. The Figure3(a)(b)(c) are respectively the normal brain T1 weighting, T2 weighting and FLAIR MR image samples, Figure 3(d) (e)(f) are the brain MR T1 weighting, T2 weighting and FLAIR MR images samples which contain brain tumor. All the experiments are conducted in the software Matlab 2013a. The brain tumor detection uses the multi-threshold segmentation based on Otsu technology and morphological operation, so that the accurate deformed area can be obtained. The Figure 4 shows the several steps of brain tumor segmentation. Figure 4(a) is the T1-MR image of a sample, obviously, it contains some MR images with similar gray level and color scope and it seems to be a tough task for the segmentation, just as shown in the Figure 4(b). The adoption of morphological operation or the filtering of marginal noise is necessary, as shown in Figure 4(c)(d). And then, finish the detection and positioning of brain tumor (Figure 4 (e)(f)). Here, we set the components based on the empirical measurement. This article adopts 5 components to conduct Gaussian Fitting for the brain tumor detection. The Figure 5 shows a histogram of brain tumor sample image.



(a) T1 Weighting (b) T2 Weighting (c) FLAIR

(d) T1 Weighting (e) T2 Weighting (f) FLAIR

Figure 3. Normal and Brain MR Image Sample

As for the training data, in order to cover the different brain areas in an image, we select normal areas from all areas of brain images. Use the GMM feature extracted from the brain tumor area and normal area, we adopt the decision tree classifier to distinguish the brain tumor from normal image samples. Table 1 shows the comparative study results of the three modes of brain MR images based on the classifier's accuracy, false alarm rate and loss detecting rate. It can be seen from the Table that the algorithm has obtained that the classification accuracy is 82.35%-94.11%, the false alarm rate and loss detecting rate are within 2.94%-8.82%.

Table 1. Simulation Results of Performance Index Based on GMM Feature

MR mode	Accuracy	False Alarm Rate	Loss Detecting Rate
T1 weighting	94.11	2.94	2.94
T2 weighting	82.35	8.82	8.82
FLAIR	91.18	4.41	4.41

Because the data dimensionality is large, we adopt PCA to decrease the data dimensionality and obtain the principal component of GMM feature extracted. The Table 2 shows the classifier's accuracy and performance index based on PCA. It is obvious that the three modes of MR image classifying accuracy has improved greatly, the classification accuracy has increased to 91.18%-94.11% and the false alarm rate is within 2.94%-4.41%. GMM feature has provided the super high classification accuracy, especially the T1 weighting and T2 weighting MR images have improved to 94.11% in accuracy.

Table 2. Simulation Results of Performance Index Based on GMM-PCA Feature

MR Mode	Accuracy	False Alarm Rate	Loss Detecting Rate
T1 weighting	94.11	2.94	2.94
T2 weighting	94.11	2.94	2.94
FLAIR	91.18	4.41	4.41

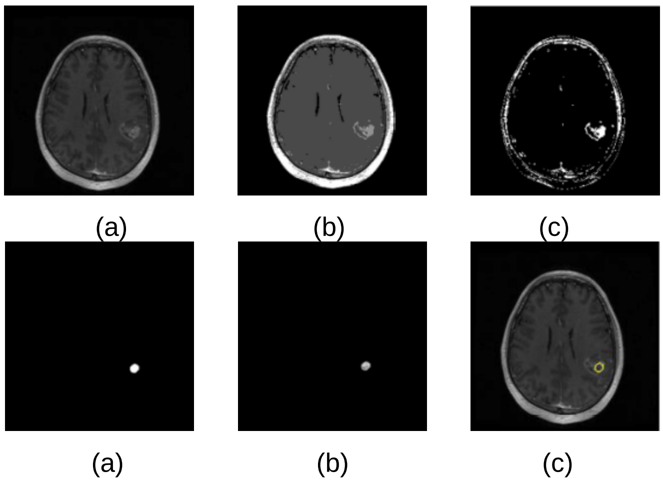


Figure 4. Brain Tumor Detection by Segmentation and Morphological Operation

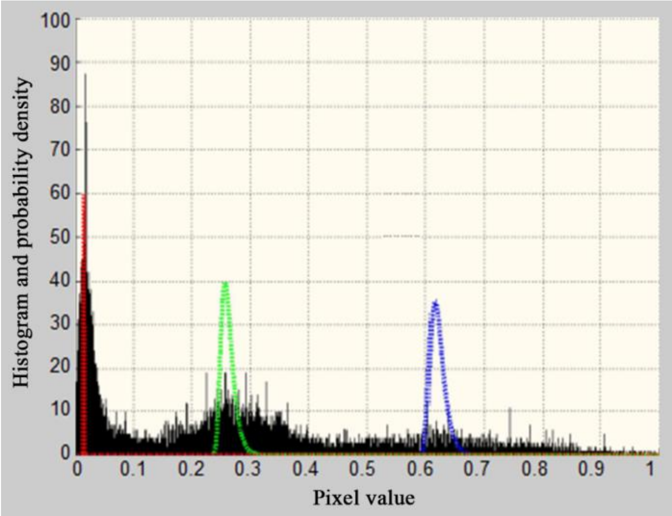


Figure 5. Histogram of Brain Tumor

4. Conclusion

This article has put forward and realized the extraction of brain tumor's GMM feature by using MR images. The brain tumor identification or diagnosis has used multi-threshold segmentation, GMM feature calculation and extraction of brain tumor, as well as the distinguishing technology of brain tumor and normal area based on decision tree classifier and other technologies. The simulation results of MR images based brain tumor prove that the use of the technology extracted from GMM feature has obtained the expected result. The preliminary experiment results show that the method has provided very good performance in the cancer cell and brain tumor detection. The ability of the method lies on its high accuracy performance that it is superior to other algorithms when detecting the brain tumor. This article uses the image samples normalized to 256*256 in size of pixel, which is helpful when carrying out automatic segmentation of the different modes of MR images with T1 weighting, T2 weighting and FLAIR. The experiment results show that the algorithm provided by this article has excellent performance and is helpful for better brain tumor diagnosis. It is the work goals of the next step to integrate the algorithm and equipment, use the GMM feature or other types of features for processing, and realize the automatic diagnosis and identification.

Acknowledgement

This work was supported by the Key Project of Natural Science Research in Anhui Province (KJ2014A266).

References

[1] Y. Wang, Y. Su and G. Agrawal, "A Novel Approach for Approximate Aggregations Over Arrays", In Proceedings of the 27th international conference on scientific and statistical database management, ACM, **(2015)**.
[2] J. Hu and Z. Gao, "Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity", Journal of Applied Mathematics, **(2012)**.
[3] X. Li, Z. Lv, J. Hu, L. Yin, B. Zhang and S. Feng, "Virtual Reality GIS Based Traffic Analysis and Visualization System", Advances in Engineering Software, **(2015)**.
[4] Z. Lv, C. Esteve, J. Chirivella and P. Gagliardo, "Clinical Feedback and Technology Selection of Game Based Dysphonic Rehabilitation Tool", 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2015), IEEE, **(2015)**.

[5] W. Ke, "Next generation job management systems for extreme-scale ensemble computing", Proceedings of the 23rd international symposium on High-performance parallel and distributed computing. ACM, **(2014)**.

[6] L. Tonglin, "Distributed Key-Value Store on HPC and Cloud Systems", 2nd Greater Chicago Area System Research Workshop (GCASR), **(2013)**.

[7] T. Su, W. Wang, Z. Lv, W. Wu and X. Li, "Rapid Delaunay Triangulation for Random Distributed Point Cloud Data Using Adaptive Hilbert Curve", Computers & Graphics, **(2015)**.

[8] J. Hu, Z. Gao and W. Pan, "Multi-angle Social Network Recommendation Algorithms and Similarity Network Evaluation", Journal of Applied Mathematics, **(2013)**.

[9] Z. Lv, T. Yin, Y. Han, Y. Chen and G. Chen. "Web VR——web virtual reality engine based on P2P network", Journal of Networks, vol. 6, no. 7, **(2011)**, pp. 990-998.

[10] J. Yang, B. Chen, J. Zhou and Z. Lv, "A portable biomedical device for respiratory monitoring with a stable power source", Sensors, **(2015)**.

[11] Z. Dongfang, "Fusion FS: Toward supporting data-intensive scientific applications on extreme-scale high-performance computing systems", Big Data (Big Data), 2014 IEEE International Conference on. IEEE, **(2014)**.

[12] S. Dang, J. Ju, D. Matthews, X. Feng and C. Zuo, "Efficient solar power heating system based on lenticular condensation", Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference, **(2014)**.

[13] X. Zhang, Y. Han, D. Hao and Z. Lv, "ARPPS：Augmented Reality Pipeline Prospect System", 22th International Conference on Neural Information Processing (ICONIP), Istanbul, Turkey. In press, **(2015)**.

[14] J. Hu and Z. Gao, "Distinction immune genes of hepatitis-induced heptatocellular carcinoma", Bioinformatics, vol. 28, no. 24, **(2012)**, pp. 3191-3194.

[15] W. Ke, "Overcoming Hadoop Scaling Limitations through Distributed Task Execution".

[16] Z. Su, X. Zhang and X. Ou, "After we knew it: empirical study and modeling of cost-effectiveness of exploiting prevalent known vulnerabilities across IAAS cloud", Proceedings of the 9th ACM symposium on Information, computer and communications security. ACM, **(2014)**.

[17] G. Bao, L. Mi, Y. Geng and K. Pahlavan, "A computer vision based speed estimation technique for localizing the wireless capsule endoscope inside small intestine", 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), **(2014)**.

[18] W. Gu, Z. Lv and M. Hao, "Change detection method for remote sensing images based on an improved Markov random field", Multimedia Tools and Applications. **(2016)**.

[19] Z. Lu, C. Esteve, J. Chirivella and P. Gagliardo, "A Game Based Assistive Tool for Rehabilitation of Dysphonic Patients", 3rd International Workshop on Virtual and Augmented Assistive Technology (VAAT) at IEEE Virtual Reality 2015 (VR2015), Arles, France, IEEE, **(2015)**.

[20] Z. Chen, W. Huang and Z. Lv, "Uncorrelated Discriminant Sparse Preserving Projection Based Face Recognition Method", Multimedia Tools and Applications, **(2016)**.

Author

Yang Li, He received his M.S. degree in computer technology from Anhui University in Hefei, China. He is currently an Assistant in the Department of Basic Public at Wannan Medical College. His research interest is mainly in the area of Data mining, Medical Informatics. He has published several research papers in scholarly journals in the above research areas and has participated in several books.

