A CBIR System for Human Brain Magnetic Resonance Image Indexing

Mina Rafi Nazari
Electrical engineering department
Sharif University of Technology
International campus
Tehran, Iran

Emad Fatemizadeh
Electrical engineering department
Sharif University of Technology
Tehran-Ira

ABSTRACT

Content-based image retrieval (CBIR) is becoming an important field with the advance of multimedia and imaging technology everincreasingly. It makes use of image features, such as color, shape and texture, to index images with minimal human intervention. Among many retrieval features associated with CBIR, texture retrieval is one of the most powerful. Content-based image retrieval can also be utilized to locate medical images in large databases. In this research, we introduce a content-based approach to medical image retrieval. A case study, which describes the methodology of a CBIR system for retrieving digital human brain MRI database based on textural features retrieval, is then presented. This research intends to disseminate the knowledge of the CBIR approach to the applications of medical image management and to discrimination between the normal and abnormal medical images based on features. The main indices are finding Normal, Abnormal and clustering the abnormal images to detect two certain abnormalities: Multiple Sclerosis and Tumoral images to classify the database. A classification with a success of 95% has been obtained by the proposed method. This result indicates that the proposed method is robust and effective compared with other recently works.

General Terms

Content based image retrieval; Classification;

Keywords

Magnetic resonance image; Medical image; Feature extraction; Support vector machine; Co-occurrence matrices

1. INTRODUCTION

Content based image retrieval has been one of the most vivid research area in image processing field recently. On the other hand, due to the growing amounts of visual data specially in medical field and widespread requirement of internet, creating a method is needed that offer more than simple text based queries. In the medical field, digital images such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), nuclear medical imaging, endoscopy, microscopy, and scanning laser ophthalmoscopy (SLO), which used for diagnostics or therapy are produced in medical centers ever increasingly[1]. Magnetic resonance imaging (MRI) is often the medical imaging method of choice when soft tissue delineation is necessary. This is especially true for any attempt to classify brain tissues [2].

The fundamental content based image retrieval system which conceptually described in [3] consist of two major parts, feature extraction and classification. The main aspect of these systems is database management based on image retrieval using its content description. An image retrieval method combined color and texture features is proposed in [4]. The other method [5] is proposed an extensive analysis of the database structure has been carried out using fuzzy clustering algorithms to build the knowledge base. Mario Partio et al. [6] present an application of gray level co-occurrence matrix (GLCM) based on Haralick et al. Method [16] to texture based similarity evaluation of rock images. Also performance of the co-occurrence matrices was compared with Gabor wavelet features which shows the co-occurrence matrices performed better for the given rock image database.

Recent work has shown that classification of magnetic resonance (MR) images related to human brain have classified via supervised techniques such as artificial neural networks and support vector machine (SVM) [7], and unsupervised classification techniques unsupervised such as self organization map (SOM) [7] and fuzzy c-means combined with feature extraction techniques [8]. Other supervised classification techniques, such as k-nearest neighbors (k-NN) also group pixels based on their similarities in each feature image [2, 10, 11, 12] can be used to classify the MRI images. We used Gray Level Cooccurrence Matrix to extract texture features which has less computational complexity in comparison of other method like wavelet. Then Principal component analysis (PCA) has been applied to effectively reduce the dimensionality of data which directly reduces the computational cost. In classification part Support Vector Machine (SVM) as supervised machine technique is used to classify of MR images into three categories, normal, MS and Tumoral.

This paper is organized as follows: The proposed method and its relative theoretical background will present in section 2. Then a short description on the input dataset of MR images and also the simulation environment will explain. Also the methods for feature extraction and reduction as well for classification will present in this section. Section 3 contains results and discussion while conclusions and future work are presented in Section 4 and 5.

2. ALGORITHM OF PROPOSED METHOD

The proposed method is based on the following techniques: Gray Level Co-occurrence Matrix, the principle components analysis PCA and Support Vector Machine. It consists of three stages: (1) feature extraction stage, (2) feature reduction stage, and (3) classification stage. The proposed technique for MRI image classification is illustrated in Fig. 2.1. In the following sections, we illustrate how the proposed method is applied. A review of basic fundamentals of SVM, principal component analysis, and Gray Level Cooccurrence Matrix are introduced.

2.1 Feature extraction block

Extraction and collection of features are a basic part of any content based image retrieval system. Features of an image are the properties that describe the content of an image which consist of visual features extracted from the image.

Co-occurrence matrices are calculated for all the images in the normalized database. GLCM is build by incrementing locations, where certain gray levels i and j occur at a distance **d** apart from each other [13].

Features Energy, Entropy, Contrast, Inverse Difference Moment, Variance, Sum Average, Sum Entropy, Sum Variance, Difference Variance, Difference Entropy and Information measure of correlation are calculated from each co-occurrence matrix and their values are saved in the feature vector of the corresponding image.
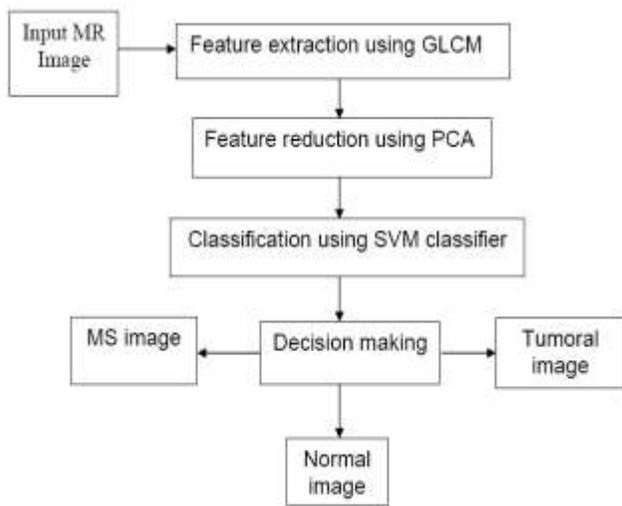

Figure 2.1: The proposed method algorithm

2.2 Feature reduction

One of the most common forms of dimensionality reduction is principal components analysis. Given a set of data, PCA finds the linear lower-dimensional representation of the data such that the variance of the reconstructed data is preserved [14, 15].

Using a system of feature reduction based on a combined principle component analysis on the feature vectors that are calculated from the GLCM, the feature vectors to the component selected by the PCA should lead to an efficient classification algorithm using supervised method. In fact it computes the linear lower-dimensional representation of the input matrix while the variance of output data remains steady [16]

So, to reduce the complexity of the system and dimensionality of the features matrix, PCA was used for feature reduction. Limiting the feature vectors to the component selected by the PCA leads to an increase in accuracy rates and decreases the time complexity.

In this experiment, MRI dataset that have healthy, tumoral and MS brain image are classified by the proposed classifiers. The eigenvalues which were calculated during PCA algorithm are shown in Table 2.1. We picked up 5 output features that are more valuable. Actually, the 4 primer eigenvalues are the bigger

ones so we try to consider the whole important numbers and use 5 of them.

Table 2.1: Relevance eigenvalues for a sample picture

4.20e+29	
1.07e+25	2.66e+05
7.42e+16	1.07e+25
3.28e+15	7.42e+16
4.46e+09	3.28e+15
5.31e+08	4.46e+09
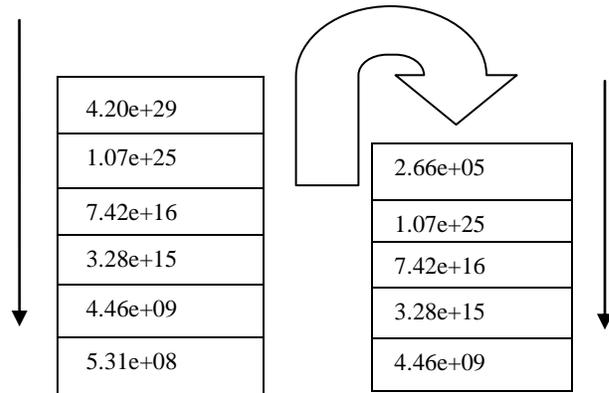
2.3 Classification

The goal of using SVMs is to find optimal hyper plane by minimizing an upper bound of the generalization error through maximizing the distance, margin, between the separating hyper plane and the data. SVMs uses the preprocessing strategy in learning by mapping input space, X to a high dimensional feature space, F. By this mapping, more flexible classifications are obtained. A separating hyper plane is found which maximizes the margin between itself and the nearest training points [19].

The output data vector from PCA comes to the classifier as an input with size 5. SVM classifies the MRI database into three certain classes consisting of normal images, tumoral images and MS images.

2.4 Database

In this section, the proposed techniques have been implemented on a real human brain MRI dataset. All the input dataset (total images is 120 which 43 images are MS, 36 tumoral and 41 normal) used for classification consists of axial, T2-weighted, 256 -256 pixel MR brain images. These images were collected from the Harvard Medical School website [16]. Figure 2.2 shows some samples from the used data for normal and pathological brain: a- normal, b- MS, c-Tumoral images.

Figure 2.2: sample MR images from the database (in order 1: tumoral,2: normal,3: MS)

2.5 Simulation environment

The algorithm described in this paper is developed locally and successfully trained in MATLAB version 7.5 using a combination of the Image Processing Toolbox of MATLAB. We performed all the computations of GLCM+PCA+SVM

classification on a personal computer with CPU 2.2 GHz, Core 2Duo processor and 2 GB of memory (RAM), running under Windows-Vista operating system and it takes less than half a minute to run. The programs can be run and tested on many different computer platforms where MATLAB is available.

3. CLASSIFICATION

The experimental results of the proposed classifiers are compared in Table 3.1, which shows the percentage of classification for the three different image classes.

Table 3.1: percentage of classification based on three normal, MS and tumoral images

MS class	Tumoral class	Normal class	Input image / Classification (percentage)
0%	0%	95%	Normal class
0%	84%	0%	Tumoral class
100%	16%	5%	MS class

Table 3.2: percentage of classification based on normal and abnormal classes

95%	Number of normal image which classify as normal class
100%	Number of abnormal image which classify as abnormal

3.1 Discusion

Finally, the classifier categorized three special classes consisting of MS images, tumoral images and normal images. Considering the analysis of the experimental results, which are shown in Tables 3.1, 3.2 the classification accuracy of detecting the normal image as a normal class is 100% and the percentage of finding abnormal images is 95%. This results are acceptable with comparison of previous works based on dividing into two classes that are mostly implemented [7, 8] on the same MRI database. We also detect the specific kinds of abnormality in the proposed method. Table 5.5 shows the average results based on a different number of training data which are given as an input to the SVM classifier. We collect the result three times with 65%, 70% and 75% of the whole MRI database as a training set of SVM classifier input and the average results of experiments is 95% normal images classification, 100% MS image classification and 84% tumoral images detection.

Table 3.3 gives the classification accuracies of our method and the recent results. The hybrid technique [9] gave 95.7% rate for DWT+PCA+FP-ANN classifier and 98.6% rate for DWT+PCA+ k-NN classifier. The same works[7, 8] with DWT+SOM find 94% rate and DWT+ one class SVM produced 96.5% and 98% for linear kernel SVM and with radial basis function based

kernel SVM respectively which shows the same results compared with our method. On the other hand, we try to detect the kinds of abnormality which means the implemented algorithm classify the MR images related to two kinds of abnormality consisting of MS and tumoral images plus classifying normal MR images with using multiclass SVM that has never been tried until now. This comparison shows that our system has high classification accuracy and less computation due to the feature reduction based on the PCA. Also it shows high classification in case of detecting just normal and abnormal classes in comparison with related work on the same database.

Table 3.3: Classification performance (P) comparisons for the proposed technique and the recent works for the same MR images datasets

Technique	P (%)
Our technique GLCM+ PCA+ SVM	95
DWT+ PCA+ ANN[9]	95.7
DWT+ PCA+ k-NN[9]	98.6
DWT+SOM[17]	94
DWT+SVM with linear kernel[18]	96.15
DWT+SVM with radial basis function based kernel[18]	98

We should also consider some images which the simulated algorithm could not classify in the correct category. Figure 3.1 shows some kind of incorrect classified images. In spite of the fact that these images have a fundamental difference with the other expressed images in database, the problem is related to the amount of background in these data to the main part of the image. Also for distinction of tumoral class, the variety of size, type and the distributed region of tumor bring some misclassification about. As we can see in Table 3.1 maximum error happens in classification of this category with 16%.
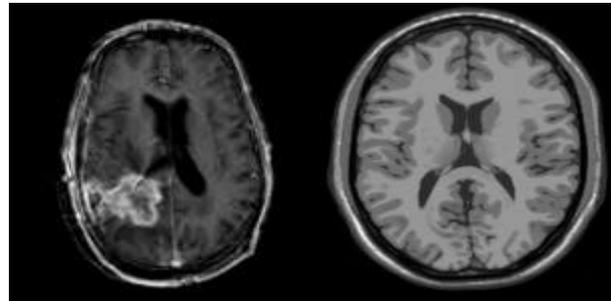

Figure 3.1: some kinds of unclassified sample images in the correct category

There is also some kind of image which is always truly detected as we have shown in Figure 3.2. We can recognize some kind of homogeneous image in this part in comparison with previous Figures (Figure 3.1) about the occupation of background and the main image.
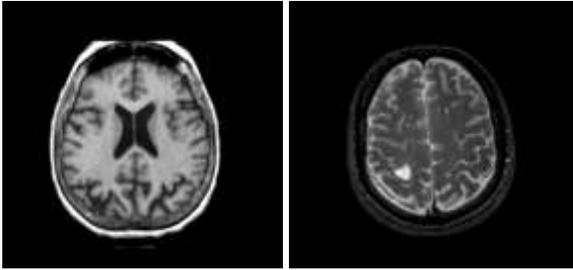
Figure 3.2: some classified image in relative class.

According to Table 3.3 from relevant works on the same MR database as we can see, most previous algorithms try to extract features by using DWT which brings about a huge computational cost. It also takes a lot of time for running the simulated program. Our proposed method using Co-occurrence matrices classify the database into three certain categories and distinguish the type of abnormality in a short time and present acceptable results.

4. CONCLUSION

In this study, we have developed a medical decision support system with normal and finding two certain abnormalities. The medical decision making system has been designed by the gray level co-occurrence matrices (GLCM), principal component analysis (PCA), and support vector machine as a supervised learning method (SVM) which help we have got very promising results in classifying the normal images, images with tumor and image of multiple sclerosis. The benefit of the system is to assist the physician to make the final decision without hesitation.

According to the experimental results, the proposed method is efficient for classification of the human brain MR database to normal, MS and Tumoral classes. Our work produced 100 % classification for MS images, 95% for normal images and84% for Tumoral images. According to our results the number of normal images which are classified as Tumoral or MS classes is zero. Therefore, it is a powerful method for distinction of normal images with 100% accuracy. On the other hand, the accuracy of finding Tumoral or MS images as the two abnormal classes is 95%. Hence in comparison with the previous method on the same data this is a powerful method. The hybrid technique [9] gave 95.7% rate for DWT+PCA+FP-ANN classifier and 98.6% rate for DWT+PCA+ k-NN classifier. The same works with DWT+SOM and DWT+SVM [7, 8] produced similar results.

Our results have been compared with the results reported very recently based on the same T2-wieghted MRI database. Our method can be employed for all types of MR images. This research developed a technique, GLCM+PCA+SVM to classify the human brain MR images. The stated results show that the proposed method can make an accurate and robust classifier. The classification performances of this study show the advantages of this technique: it is easy to operate, noninvasive and inexpensive.

5. FUTURE WORKS

The presented algorithm in this thesis successfully demonstrated innovation and novelty. The results that have been obtained are highly promising and have reasonably demonstrated a proof of concept. However, there is substantial scope for further research and development to the system.

The limitation of this work is that it requires fresh training each time whenever there is an increase in image database. On the other hand, our proposed method can diagnose just two kinds of abnormality plus normal images so extension of the developed techniques for processing more abnormality on brain MRI databases is the topic of future research.

The other major research area is designing a preprocessing step to assimilate various databases or different type of images in a certain database to make the algorithm more practical.

6. REFERENCES

[1] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain,V.,2000. Content-Based Image Retrieval at the End of the Years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22, pp. 1349–1380.

[2] Fletcher-Heath, L. M., Hall,D, L. O., Goldgof, B., Murtagh, F. R.,2001. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images, Artificial Intelligence in Medicine 21, pp. 43-63.

[3] Partio, M., Cramariuc, B., Gabbouj, M., and Visa, A., 2001. Rock texture retrieval using Gray Level Co-occurrence Matrix.

[4] Dahshan, A. E., Salem, and Younis, T. H., 2009. A hybrid technique for automatic MRI brain images classification, BABES–BOLYAI, Informatica, Volume LIV.

[5] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R., 2000. Content based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380.

[6] Eakins, J., and Graham, M., 1999. Content-based image retrieval: A report to the jisc technology applications programme, Tech. Rep. JTAP-039, Institute for Image and Data Research, University of Northumbria at Newcastle.

[7] Chaplot, S., Patnaik, L.M., and Jagannathan, N.R., 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomedical Signal Processing and Control 1, pp. 86-92.

[8] Maitra, M., Chatterjee, A., 2007. Hybrid multiresolution Slantlet transform and fuzzy c- means clustering approach for normal-pathological brain MR image segregation, Med Eng Phys, doi:10.1016/j.medengphy.2007.06.009.

[9] Dahshan, A. E., Salem, A. B. M., and Younis, T. H., 2009. A hybrid technique for automatic MRI brain images classification, BABES–BOLYAI, Informatica, Volume LIV.

[10] Abdolmaleki, P., Mihara, F., Masuda, K., and Buadu, L. D., 1977. Neural networks analysis of astrocytic gliomas from MRI appearances, Cancer Letters 118, pp. 69-78.

[11] Rosenbaum, T., Engelbrecht, V., Krolls, W., and Dorstenc, F. A., Berlagec, M. H., and Lenard, H. G., 1999. MRI abnormalities in neurofibromatosis type 1 (NF1): a study of men and mice, Brain & Development 21, pp. 268-273.

[12] Cocosco, C., Zijdenbos, A. P., and Evans, A. C., 2003. A fully automatic and robust brain MRI tissue classification method, Medical Image Analysis 7, pp. 513-527.

[13] Haralick, R. M., 1979. Statistical and structural approach to texture, proceeding of the IEEE, vol. 67, pp. 786-804.

[14] Sengur, A., 2007. An expert system based on principal component analysis; artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases, Comp. Biol. Med.

[15] Duda, R. O., Hart, P. E., and Stork, D. G., 2001. Pattern Classification, New York: John Wiley and Sons.

[16] Sengur, A., 2007. An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases, Comp. Biol. Med, doi: 10.1016/j.compbiomed.2007.11.004.

[17] Harvard Medical School, Web: data available at http://med.harvard.edu/ AANLIB/.

[18] Chaplot, S., Patnaik, L.M., Jagannathan, N.R., 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomedical Signal Processing and Control 1.

[19] Shin, H. J., Eom, D. H., and Kim, S. S., 2005. One-class support vector machines-an application in machine fault detection and classification, Science Direct, volum 48, issue 2. Pages 395-408.