Classification of MRI Brain Image using SVM Classifier

Mrs. M.C. Hingane, Mr. Satish B. Matkar, Mr. Ashish B. Mane, Mr. Ambadas M. Shirsat

[1]PG Student,[2,3,4]UG Student

PDEA's COEM, Pune

Abstract: **In this paper we describe the design and development of content-based image retrieval (CBIR) system. This system enables both multi-image   query and slide –level image retrieval in order to protect the semantic consistency among the retrieved images. Content of image retrieval is the process of finding relevant image from large collection of image database using visual queries. Medical images have led to growth in large image collection. To enhance the medical image retrieval for diagnostics, research and teaching purposes is done by CBIR. The system performance is improved by the multiple image queries instead of single image query. Pre-processing of the query image is done by median filter to remove the noise. Then the filtered image is given as input to the feature extraction technique which is a transformation of input image into set of features such as texture and shape. Feature extraction is done by the Gray level co-occurrence matrix algorithm that contains information about the position of pixels having similar gray level values. SVM (Support Vector machine) classifier is to group items that have similar feature values into three categories such as normal, benign and malignant. Then SVM classifier is followed by KNN (K-nearest neighbour) which search the corresponding database index will be computed by similarity feature matching. The query image is classified by the classifier to a particular class and the relevant images are retrieved from the database**.

Keywords: Feature Extraction, GLCM, Image Retrieval, MRI, SVM Classifier

1. Introduction

Today among the applications of computer science to the field of medicine, the processing of medical image data is playing an increasingly important role. With medical imaging techniques such as X-Ray, computer tomography, magnetic resonance imaging, and ultrasound, the amount of digital images that are produced in hospitals is increasing incredibly fast. Thus, the tasks of efficiently storing, processing and retrieving medical image data have become important research topics. Many hospitals use picture archiving and communication systems (PACS), which are basically computer networks that are used for storage, retrieval, and distribution of medical image data. In such a network, a central server provides access to an image database from which clients such as medical staff can retrieve images by using metadata like the name of the patient, the date, the imaging method, the body part, etc.

The applications of computer vision techniques present an image retrieval problem which is explicated as the problem of searching for digital images in large databases. An image retrieval system is a computer system for searching and retrieving images from a large database of digital images. Color, Shape and texture are important cue in extracting information from images; these histograms are widely used in content based image retrieval (Serge Belongie et al, 1998). Color and texture contain important information for two images with similar color histograms that can represent very different things. Therefore the use of shape-describing features is essential in an efficient content-based image retrieval system. Although shape description has been intensively researched, there exists no direct answer as to what kind of shape features is incorporated into such a system

Metadata-based retrieval is done via standard database tasks that are relatively easy to implement. If the metadata for an image is not sufficient to formulate a precise enough query, a textual query can be given. This query can for example be a set of keywords or a full textual description of the desired images. Then, the PACS searches for database images with similar descriptions. This retrieval task is more complicated and involves techniques from the field of text retrieval. If for example a doctor wants to compare X-ray images of his current patient with images from similar cases, he could also use these images as queries and let the PACS find the most similar entries in the database. This kind of searching for images is called content-based image retrieval (CBIR) and is currently part of the research of many computer science groups, who are trying to find models for the similarity of digital images. Several content based image retrieval systems are currently being developed.

2. Relevance

Brain tumors are a heterogeneous group of central nervous system neoplasm that arise within or adjacent to the brain. Some are curable by surgical resection, but many cannot be eradicated by current treatments, and when they are, disabling neurological injury, often ensues. Moreover, the location of the tumor within the brain has a profound effect on the patient's symptoms, surgical therapeutic options, and the likelihood of obtaining a definitive diagnosis. The location of the tumor in the brain also markedly alters the risk of neurological toxicities that alter the patient's quality of life. At present, brain tumors are detected by imaging only after the onset of neurological symptoms. No early detection strategies are in use, even in individuals known to be at risk for specific types of brain tumors by virtue of their genetic makeup. Current histopathological classification systems, which are based on the tumors presumed cell of origin, have been in place for nearly a century and were updated by the World Health Organization in 1999. Although satisfactory in many respects, they do not allow accurate prediction of tumor behavior in the individual patient, nor do they guide therapeutic decision-making as precisely as patients and physicians would hope and need. Current imaging techniques provide meticulous anatomical delineation and are the principal tools for establishing that neurological symptoms are

the consequence of a brain tumor. The current histopathological approach to the diagnosis and classification of brain tumors is satisfactory in many respects. This would be especially welcome to patients, who perceive that the evaluation process is slow, inefficient, and imprecise. Moreover, early identification of effective therapies quickly resets the clinical research agenda to include quality of life as well as efficiency.

3. Proposed Work

The fundamental content based image retrieval system consists 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. The proposed method is based on three stages:

• Feature extraction stage, using GLCM (Gray level co-occurrence matrix)

• Feature reduction stage, using PCA (Principal Components Analysis)
Classification stage using SVM (Support Vector Machine). The proposed technique for MRI image classification is illustrated in Fig. 2.

3.1 Feature extraction block

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.
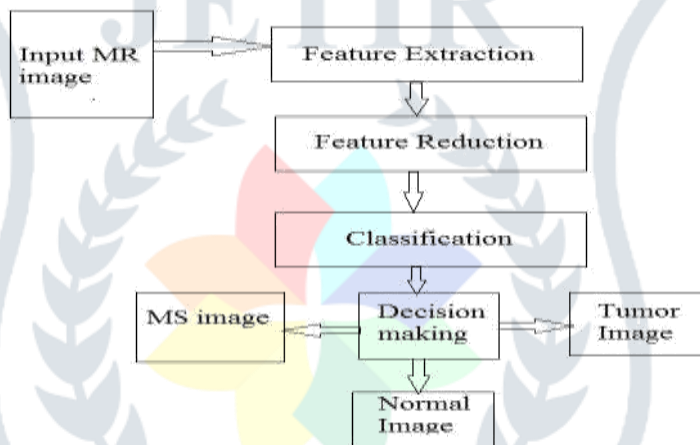

Fig 1: The proposed method algorithm.

• GLCM (Gray level co-occurrence matrix)
• DWT (Discrete Wavelet Transform)
• Direct variance, etc

A performance of the co-occurrence matrices when compared with wavelet features shows that co occurrence matrix performed better for the given rock image database. 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. Features such as energy, entropy, variance, correlation are calculated. Here we calculate correlation between i and j defined by

$$R_{ij} \quad (i \quad i).(j \quad j)\, P_{ij} \quad i \quad j \qquad (1)$$

Where

i – Gray level of reference pixel j- Gray level of neighboring pixel
μi- expected or mean value of gray level of reference pixel
μj- expected or mean value of gray level of neighboring pixel
Pij – joint probability of i and j σi - standard deviation of i
σy- standard deviation of j

3.2 Feature Reduction

One of the most common forms of dimensionality reduction is PCA (principal components analysis).Principal component analysis (PCA) is appropriate when we have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables (called principal components) that will account for most of the variance in the observed variables. The principal components may then be used as predictor or criterion variables in subsequent analyses. The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. 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. Using a system of feature reduction based on a combined principle component analysis feature vectors are calculated

from the GLCM. In fact it computes the linear lower-dimensional representation of the input matrix. PCA uses an orthogonal transformation which converts a set of observations of possible correlated variables into a set of values of uncorrelated variables called principal components. Number of principal components is less than or equal to the number of original variables. Limiting the feature vectors to the component selected by the PCA leads to an increase in accuracy rates and decreases time complexity. The Eigen values are calculated using PCA.

Let X be the original data set, where each column is a single sample of our data and Y is a re-representation of that data set such that

$$PX = Y$$

Where P is a matrix that transforms X into Y. The rows of P, [p1. . . pm], are a set of new basis vectors for expressing the columns of X.

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad P = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \quad (2)$$

$$Y = PX = \begin{bmatrix} p_1 x_1 & \cdots & p_1 x_n \\ \vdots & & \vdots \\ p_m x_1 & \cdots & p_m x_n \end{bmatrix}$$

3.3 Classfication

The Support Vector Machine (SVM) is a state-of-the-art classification method. The SVM classifier is widely used in bioinformatics (and other disciplines) due to its high accuracy, ability to deal with high-dimensional data such as gene expression, and flexibility in modeling diverse sources of data. SVMs belong to the general category of kernel methods. A kernel method is an algorithm that depends on the data only through dot-products. When this is the case, the dot product can be replaced by a kernel function which computes a dot product in some possibly high dimensional feature space. This has two advantages: First, the ability to generate non-linear decision boundaries using methods designed for linear classifiers. Second, the use of kernel functions allows the user to apply a classifier to data that have no obvious fixed-dimensional vector space representation. 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[8]. The output data vector from PCA comes to SVM classifier as an input with size 5. SVM classifier classifies the MRI database into three certain classes consisting of normal images, tumoral images and MS images.

4. Database

The proposed techniques have been implemented on a real human brain MRI dataset. We are working on images collected from the Harvard Medical School website [9]. Fig. 2 shows three types of brain MRI a- normal, b- MS, c-Tumoral images.

5. Software and Hardware Requirements

The algorithm described in this paper can be developed locally and successfully trained in MATLAB version 7.5 using a combination of the Image Processing Toolbox of MATLAB. We are performing 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. The programs can be run and tested on many different computer platforms where MATLAB is available.

6. Discussion

At present there are some hybrids techniques for detection of brain tumor .Based on feature extraction, reduction and classification, methods are given such as

1)Discrete Wavelet Transform(DWT)+ principal components

analysis (PCA)+ Artificial Neural Networks(ANN)

2)   DiscreteWaveletTransform (DWT)+   principal

components analysis(PCA)+ k-Nearest Neighbors(k-NN) 3) Discrete Wavelet Transform(DWT)+ Self Organization

Map (SOM)

The proposed method in this project is GLCM+ PCA+ SVM

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.

7. Conclusion

In this study, we are developing 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 will help us to get 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. This system can also be well utilized for detecting tumors in the whole body i.e. not only concentrating on the brain but also the other organs. This system represents an innovative idea to implement an efficient system with powerful algorithm. Just like texture retrieval as one of the methods of CBIR to implement the system, the other feature retrieval techniques can also be considered for comparative study. 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.

Fig. 2: sample MR images from the database (in order

(a) Tumoral, (b) normal, (c) Multiple Sclerosis MS)

ACKNOWLEDGMENT:

Refrences:

[1] Cocosco, C., Zijdenbos, A. P., and Evans, A. C., 2003. A fully automatic and robust brain MRI tissueClassification method, Medical Image Analysis 7, pp. 513-527
[2] Fletcher-Heath, L. M., Hall,D, L. O., Goldgof, B., F. Murtagh, R.,2001. Automatic
[3] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, Content-Based Image Retrieval at the End of the Years, IEEE Transactions on Pattern Analysis and MachineIntelligence.
[4] Arati Kothari, Detection and classification of brain cancer using artificial neural network in MRI images, World Journal of Science and Technology, 2012, ISSN: 2231 – 2587.
[5] Ben George E, Karnan M , MRI Brain Image Enhancement Using Filtering Techniques, International Journal of Computer Science & Engineering Technology, Vol. 3 No. 9, 399- 403.
[6] Ben George E, Karnan M, Feature Extraction and Classification of Brain Tumor using Bacteria Foraging Optimization Algorithm and Back Propagation Neural Networks, European Journal of Scientific Research, 2012, 88 (3), 327-333.
[7] Ben George E, Karnan M, MR Brain Image Segmentation using Bacteria Foraging Optimization Algorithm, International Journal of Engineering and Technology 2012.
[8] Bhupal Singh and Jai Singh, Classification of Brain MRI in Wavelet Domain, International Journal of Electronics and Computer Science Engineering,2011,ISSN 2277-1956/V1N3-879.
[9] Daljit Singh, and Kamaljeet Kaur, Classification of Abnormalities in Brain MRI Images Using GLCM, PCA and SVM , International Journal of Engineering and Advanced Technology (IJEAT)2012,ISSN: 2249 – 8958, Volume-1, Issue-6.

Author's Profile:
[1]**Mrs. M.C. Hingane. :** She is currently appearing Masters in Computer Engineering and works on same project.
[2] **Mr.Ashish Mane :** He is currently appearing Bachelors of engineering in computer engineering. Works on project Classification of MRI Brain Image using SVM Classifier
[3] **Mr. Satish Matkar :** He is currently appearing Bachelors of engineering in computer engineering. Works on project Classification of MRI Brain Image using SVM Classifier
[4]**Mr. Ambadas Shirsat :** He is currently appearing Bachelors of engineering in computer engineering. Works on project Classification of MRI Brain Image using SVM Classifier