Content Based Image Retrieval Approach to Tumor Detection in Human Brain Using Magnetic Resonance Image

Aditi P. Killedar   Veena P. Patil   Megha S. Borse

Abstract— with the advancement in multimedia and imaging technology, Content Based Image Retrieval (C.B.I.R.) is becoming very important and emerging field. CBIR is the part of image processing that extracts features of image to index images with minimal human interventions. There are many feature extraction techniques such as color, shape or texture retrieval among which texture retrieval is the most powerful and optimal technique .C.B.I.R. finds better future scope in medical field. A huge database of medical images can be created using C.B.I.R. In   this   paper, we introduce a content-based approach to retrieve information from digital human brain   Magnetic Resonance Image (M.R.I.) based on texture retrieval. The paper describes methodology of C.B.I.R. applied to images of medical field to differentiate them between normal and abnormal medical images based on feature extraction a n d clustering the abnormal images to detect two certain abnormalities: Multiple Sclerosis(M.S.) and Tumoral images to classify database.

Key words — Co-occurrence matrices, Feature extraction, Magnetic resonance image, Support vector machine.

I. INTRODUCTION

Content-Based Image Retrieval (C.B.I.R.) technology has seen proposed to benefit not only the management of increasingly large image collections, but also to aid clinical medicine, research, and education relying on visual content in the data. CBIR can be briefly defined as a set of methods that attempts to index an image based on the characteristics of its visual content, and to retrieve the images by similarity to queries that express some combination of these characteristics. These characteristics may include intensity, color, texture, shape, size, or location, or their combination. Imaging plays a central role in the diagnosis of brain tumors. There are various techniques to detect the brain tumor, which are CT scan, MRI, Biopsy, and Ultrasound etc [1]. MRI makes use of the property of nuclear magnetic resonance (NMR) to image nuclei of atoms inside the body. MRI provides good contrast between the different soft tissues of the body, which makes it especially useful in imaging the brain, muscles, the heart, and cancers compared with other medical imaging techniques such as computed tomography (CT) or X-rays[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]. 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 described in Mario Partio et al. [6]. Recent work has shown that classification of magnetic resonance (MR) images related to human brain have classified via

- Supervised techniques
    - a. Artificial Neural Networks (ANN)
    - b. Support Vector Machine (SVM)
    - c. k-Nearest Neighbors (k-NN)
- Unsupervised techniques
    - a. Self Organization Map (SOM)

We use Support Vector Machine (SVM) as supervised machine technique to classify of magnetic resonance (MR) images into three categories, normal, MS and Tumoral. One of the most common forms of dimensionality reduction is principal components analysis.

II. 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.

III. 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.

A. 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
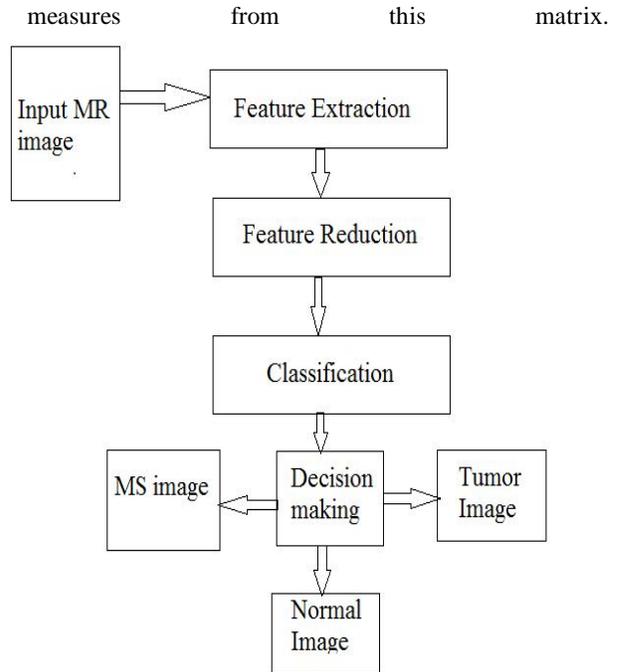
measures from this matrix.

Fig 1: The proposed method algorithm.

The feature extraction techniques are
1) GLCM (Gray level co-occurrence matrix)
2) DWT (Discrete Wavelet Transform)
3) 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

$$Rij = \sum \sum (i - \mu i).(j - \mu j) Pij \div \sigma i \sigma j \quad (1)$$

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

B. 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, $[p1 \ldots pm]$, are a set of new basis vectors for expressing the columns of X.

$$PX = \begin{bmatrix} p1 \\ pm \end{bmatrix} \begin{bmatrix} x1 & xn \end{bmatrix} \qquad (2)$$

$$Y = \begin{bmatrix} p1x1 & p1xn \\ pmx1 & pmxn \end{bmatrix}$$

C. Classification

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.

IV. 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.

V. SOFTWARE AND HARDWARE REQUIREMENT

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.

VI. 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.

VII. 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.
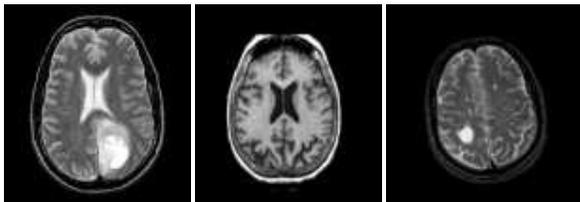

Fig. 2: sample MR images from the database (in order (a) Tumoral, (b) normal, (c) Multiple Sclerosis MS)

ACKNOWLEDGMENT

REFERENCES

[1] 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.
[2] Fletcher-Heath, L. M., Hall,D, L. O., Goldgof, B., F. Murtagh, R.,2001. Automatic segmentation of non enhancing brain tumors in magnetic resonance images, Artificial Intelligence in Medicine 21, pp. 43-63.
[3] 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 MachineIntelligence 22, pp. 1349–1380.
[4] Dahshan, A. E., Salem, and Younis, T. H., 2009. A hybrid technique for automatic MRI brain images classification,
[5] Haralick, R. M., 1979. Statistical and structural approach to texture, proceeding of the IEEE, vol. 67, pp. 786-804. BABES–BOLYAI, Informatica, Volume LIV.
[6] Partio, M., Cramariuc, B., Gabbouj, M., and Visa, A., 2001. Rock texture retrieval using Gray Level Co-occurrence Matrix.
[7] Chaplot, S., Patnaik, L.M., and Jagannathan, .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] 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
[9] Harvard Medical School, Web: data available at http://med.harvard.edu/ AANLIB/.

AUTHOR'S PROFILE

Aditi P. Killedar

She has completed her graduation in electronis and telecommunication engineering from KKWCOE,Nasik.Currently persuing her post graduation in digital signal processing from Cummins COE,Pune.She is working as a lecturer in Matoshri COE ,Nasik.She has2.5 years of teaching experience.

Veena P. Patil

She has completed her graduation in electronis engineering from SSVPSCOE,Dhule.Currently persuing her post graduation in Electronics from TKIET,Warananagar.Dist Kolhapur.She is working as a lecturer in Matoshri COE ,Nasik.She has 6.5 years of teaching experience.

Megha S. Borse

She has completed her graduation in electronis and telecommunication engineering and post graduation in digital signal processing She is working as a assistant professor in Cummins COE, Pune She has 20 years of teaching experience.