ISSN:0975-9646

T.Baranidharan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (6) , 2011, 2773-2775

Medical Image Classification Using Information Gain for Global Feature Reduction

T.Baranidharan[#], Dr.D.K.Ghosh[*]

[#]Dept. Of EIE, K.S.Rangasamy College of Technology
Tiruchengode, Tamil Nadu, INDIA
[*]V.S.B Engineering College
Karur, Tamil Nadu, INDIA

***Abstract*— Medical image classification and retrieval systems have been finding extensive use in the areas of image classification according to imaging modalities, body part and diseases. One of the major challenges in the medical classification is the large size images leading to a large number of extracted features which is a burden for the classification algorithm and the resources. In this paper, it is proposed to investigate the efficacy of information gain of the extracted energy with respect to the class. Results obtained from the proposed method indicate the classification accuracy is not affected by our proposed data reduction method.**

***Keywords*— Fast Hilbert Transform, Image Classification, Medical Images, Information gain.**

I. INTRODUCTION

Medical images are generated using computerized tomography (CT), magnetic resonance image (MRI), positron emission tomography (PET), single photon emission computed tomography (SPECT), ultrasound from radiological departments. Manually classifying and archiving of medical images is a time consuming and complex process. Medical image retrieval systems are using image processing and information retrieval techniques for automated classification and retrieval [1, 2]. The medical images retrieval techniques from a large source of image database include the histrogram, color, texture analysis with reference images [3].

In medical image retrieval system histogram is not dependent upon the feature such as image size & orientation. It represents the intensity of 3 colors i.e. Red, Green, and Blue [4]. Texture is used to measure similarity between images based on the human perception of the image.Gobar filter are normally used to extract texture features in medical image retrieval [5,6].

Most picture archiving and retrieval systems use textual information to retrieve these images which are ineffective with most of the time the required images not being retrieved. An emerging area of research is content based image retrieval where the query parameter for retrieving an image is a image [7,8]. Based on the query image similar images are retrieved from the database. In medical imaging, automatic classification and retrieval is useful to insert the new radiographs into existing archive without interaction, searching for a specific diagnoses based on an image input[9,10]. Image retrieval system reduces the cost in medical care significantly as the clinical decision process by a physician can be faster as anatomical features or pathologic appearance can be compared in the image database[11,12].

Sadek, et al., 2008 [13] proposed a new architecture for CBIR using Splines Neural Network based image retrieval (SNNIR). The efficiency of traditional neural networks is limited due to the assumption of linear relationship among features and the difficulty of representing high level concepts in low level features. The proposed neural network is based on an adaptive model called splines neural network. The splines neural network facilitates the system to determine nonlinear relationship between different features in images which betters the comparison accuracy. Results of the proposed system show that it is more effective and efficient to retrieve visual-similar images for a set of images with same conception can be retrieved.

Dimitris K. Iakovidis, et al., 2009[14] proposed a novel scheme for content based image retrieval (CBIR) for medical images. The proposed method was formalized according to the patterns for next generation database systems (PANDA) framework for pattern representation and management. The low level features extracted from the medical images were clustered in the feature space to form higher level, semantically meaningful patterns. Expectation-maximization algorithm was used to automatically determine the number of clusters. The similarity between the clusters was estimated as a function of the similarity of both their structures and the measure of components. Experiments showed that the proposed scheme can be efficiently applied for medical image retrieval from large database.

II. METHODOLOGY

The Hilbert transform of a function $f(x)$ is given by

$$F(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{f(x)}{t-x}d(x)$$

Since the integral is evaluated using the Cauchy principal value the above equation can be written as the convolution

$$F(t) = \frac{1}{\pi t}.f(t)$$

Using the convolution theorem of Fourier transform we can evaluate the above equation as the product of $f(x)$ with $-i \times \operatorname{sgn}(x)$
where

$$\operatorname{sgn}(x) = \begin{cases} -1 & for \quad x < 0 \\ 0 & for \quad x = 0 \\ 1 & for \quad x = 1 \end{cases}$$

Recent trend in medical image retrieval proposes to use learning algorithms. Popular classification algorithms include SVM, ANN, Naive Bayes and Decision tree algorithms. In this paper we propose to evaluate Random tree, Random forest and CART. Random forest is an aggregation of random trees. The input vector for an object to be classified is predicted using each of the trees in the forest. The forest chooses the best answer based on the number of votes obtained in each forest. The mode of the random tree results gives the class label.

The logic for random forest is given below

- Specify the number of training set J
- Specify the number of attributes K
- Use bootstrapping to select the training set
- Randomly choose k (k<K) on which the decisions are to be made
- Compute the best split for making a decision
- Grow each tree without pruning

Classification and Regression Tree (CART) functions depending on the type of target variable. The variable can either be continuous or categorical. For a categorical class label, use the values of the predictor variables to span through the tree till a leaf node is encountered. The class label assigned will be the value of the leaf node. Gini index is used to decide the attribute split criteria.

III. EXPERIMENTAL SETUP

In this work 180 medical images were used for investigating the hypothesis. Energies where extracted using fast Hilbert transform and coefficients were selected using zig zag method. The information gain was computed for all the attributes with respect to the class and ranked. All attributes with the ranking value of greater than 0.099 was selected. Figure 1 shows some of the images used in this work. The down sized data set was used to find the classification accuracy of random forest , random tree & CART. The results obtained are shown in fig 2.

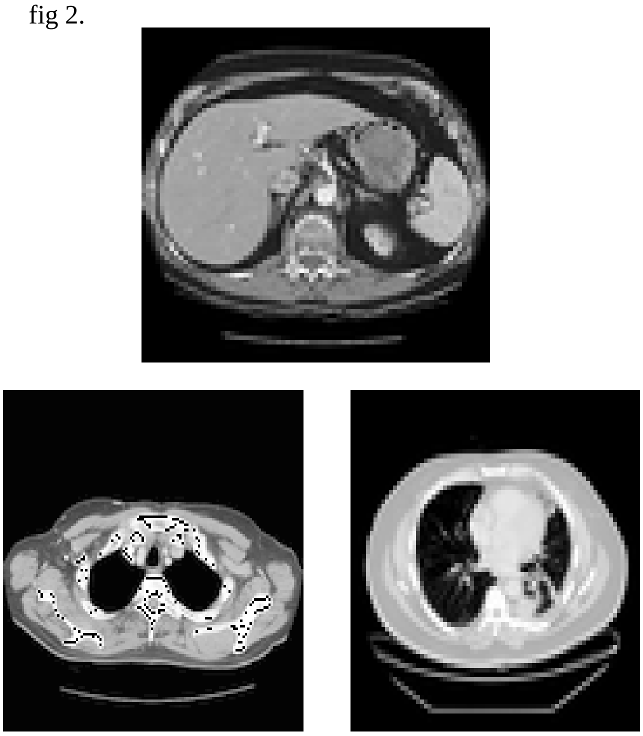

Figure 1. Medical images used

From figure 1 it is observed that images with very high noise content are also being used in this work to check the efficacy of the proposed method.

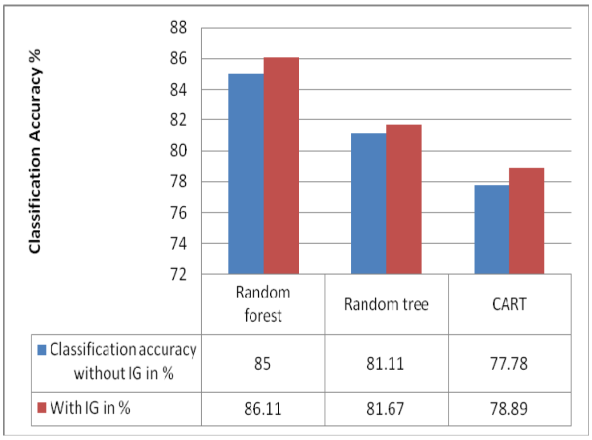

Figure 2. The classification accuracy of the proposed method.

IV. CONCLUSION

In this paper a novel method for image retrieval was proposed. Energies were extracted from images using fast Hilbert transform and global energy coefficient using ZigZag method was computed. It was proposed to use information gain for attribute selection and reduction . The proposed method's classification accuracy was compared against the classification accuracy of data not preprocessed with information gain. It is observed the classification accuracy improves for random forest &CART by 1% . From the above it can be inferred that the information can be used for reducing the extracted features in medical image classification.

. REFERENCES

[1] Md. Mahmudur Rahman, Tongyuan Wang,Bipin C. Desai Medical Image Retrievaland Registration: Towards ComputerAssisted Diagnostic Approach Proceedings of the IDEAS Workshop on MedicalInformation Systems: The Digital Hospital(Ideas- Dh'04) Issue , 1-3 Sept. 2004Page(s): 78 – 89

[2] Lilian H. Y. Tang, R. Hanka, Ringo W. K.Lam, Kent K T Cheung, Horace H. S. Ip,"Automatic Semantic Analysis for Histological Images", Medical Imaging Understanding and Analysis 99 (MIUA'99), pp. 53-56, Oxford, UK, 19-20 July, 1999.

[3] G. D. Guo, A. K. Jain, W. Y. Ma, and H. J. Zhang,"Learning similarity measure for natural image retrieval with relevance feedback", *IEEE Trans. Neural Networks*., vol. 13, no. 4, pp. 811–820, Jul. 2002.

[4] B.S.Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan, and Akio Yamada, "Color and Texture Descriptors". In: IEEE Transactions on Circuits and Systems for Video chnology, Vol. 11, No. 6, June 2001,pp. 70-715.

[5] B.S.Manjunath,W.Y.Ma,"Texture Features For Browsing and Retrieval of Image Data",IEEE Transcation on pattern Analysis and Machine Intelligence, 18(8), pp.837-842 1996.

[6] Y. Chen and J. Z. Wang, "A region-based fuzzy featurematching approach to content-based image retrieval", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp.1252–1267, Sep. 2002.

[7] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: Current techniques, promising directions, and open issues", J. Vis. Commun. Image Represen., vol. 10, pp 39–62, Mar. 1999.

[8] P. Kelly, T. Cannon, and D. Hush. Query by image example:The CANDID approach. In Storage and Retrieval for Image and Video Databases III, pages 238–248. SPIE Vol. 2420,1995.

[9] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Tools for content-basedmanipulation of image databases. In Storage and Retrieval for Image and Video Databases, pages 34–47. SPIE, 1994

[10] Scott T. Acton, Peter Soliz,Stephen Russell, Marios S. Pattichis "CONTENT BASED IMAGE RETRIEVAL: THE FOUNDATION FOR FUTURE CASE-BASED AND EVIDENCE-BASED OPHTHALMOLOGY" ICME,

[11] Avi Kak and Christina Pavlopoulou "Content-Based Image Retrieval from Large Medical Databases" First International Symposium on 3D Data Processing Visualization and Transmission, 2002)

[12] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. H. bs B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content The QBIC system. *IEEE Computer*, pages 23–32, 1995.

[13] Samy Sadek, Ayoub Al-Hamadi, Bernd Michaelis , and Usama Sayed; “Image retrieval using Cubic Splines Neural Networks.” International Journal of Video & Image Processing and Network Security IJVIPNS 2008 Vol: 9 No: 10, pp:17-23.

[14] Dimitrios K. Iakovidis, Nikos Pelekis, Evangelos E. Kotsifakos, Ioannis Kopanakis, Haralampos Karanikas, Yannis Theodoridis: A Pattern Similarity Scheme for Medical Image Retrieval. IEEE Transactions on Information Technology in Biomedicine 13(4): 442-450 (2009)