Classification of Medical Images using Fast Hilbert Transform and Decision Tree Algorithms

T.Baranidharan(Author)

Asst prof,Department of Electronics and Instrumentation
K.S.Rangasamy College of Technology
Erode, Tamilnadu state, INDIA

Dr.D.K.Ghosh (Supervisor)
Professor, Department of Mathematics
V.S.B Engineering College
Karur, Tamilnadu state, INDIA

Abstract— with generation of huge volume of patient data and advent of modern medical devices, large quantities of medical images are being generated. This has led to development of systems which can automatically compare, classify and index these medical images for future use. Non availability of efficient medical image classification systems have been a primary cause for under utilization of available medical image repositories. The current study investigates the efficiency of medical image transformation to frequency domain using the Fast Hilbert Transform and subsequently classification of the same using decision tree algorithms

Keywords-component; Fast Hilbert Transform, Image Classification, Medical Images

I. INTRODUCTION

Medical devices today extensively generate large quantities of images. Major challenges involved in the management of these images are indexing and classifying the images for future use. In a typical large hospitals terabytes of data reproduced each year. 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 [1]. 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. 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.

The tasks of image retrieval system is to retrieve relevant medical images from an image database based on the similarity of visual content of the given query image[2]. Histogram, texture analysis and color are some of the low level features in an image which has been effectively used for the image retrieval problem. However low level image features are prone to give a larger classification error due to the semantic gap[3]. Medical image retrieval differs from conventional image retrieval with respect to the image signatures which tend to be global in regular image retrieval but in the case of medical image retrieval due to the inherent localization of the pathological condition, global signatures fail. Features which are extracted from the whole images are called global features [5], while local features cover parts of region of interest. Three main techniques have been proposed namely classification, clustering and relevance feedback. Once features are extracted the next step involves similarity between the extracted features need to be compared. Distance based similarity measurement is a popular method to evaluate the closeness between two images. However distance based metrics have limitations in the classification accuracy. Learning based systems are increasingly being used for the image retrieval problem.

Scott et al, have proposed an content based image retrieval system for ophthalmology [6] to retrieve retinal images. Investigation was carried out on Stargardt's Disease and Drusen Detection for a given query image. Avi

Kak et al investigated the medical image retrieval problem in large databases [7] by using over 2871 images. Low feature extractions have been reported in [8, 9].

In this paper we propose to convert the image to frequency domain using Fast Hilbert transform, extract global features using zig – zag sampling and investigate the classification efficiency using various Decision tree algorithms. This paper is organized into the following sections. Section II briefly introduces to the Fast Hilbert transform, section III describes the classification algorithms used, section IV explains the experimental setup along with results obtained and section V discusses the result obtained.

II. FAST HILBERT TRANSFORM

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

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

The above equation can be extended upto 3 dimensions.

III. CLASSIFICATION ALGORITHM

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.

IV. EXPERIMENTAL SETUP

In this paper we chose 180 medical images and computed the features using Fast Hilbert transform. The obtained values were pre processed using Zig Zag scan to select the features. Some of the images used in the experimental setup is shown in figure I. Three types of images were selected with varying degree of noise. The obtained values were assigned class labels and a 10 fold cross validation was used to train the classification algorithm. The results obtained are shown in table I.

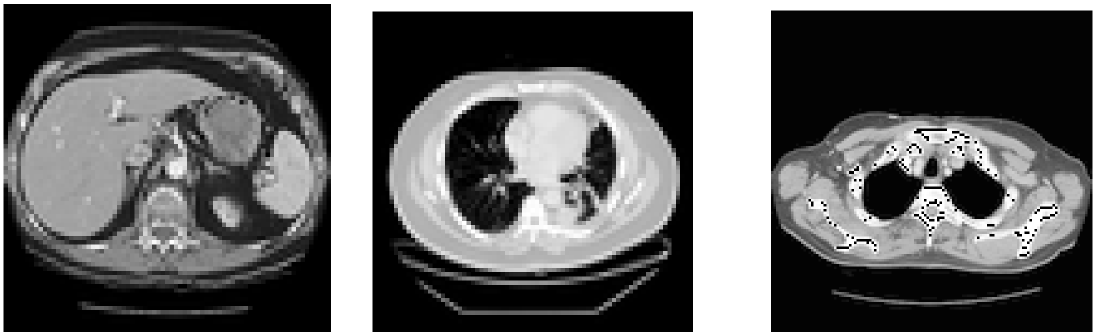
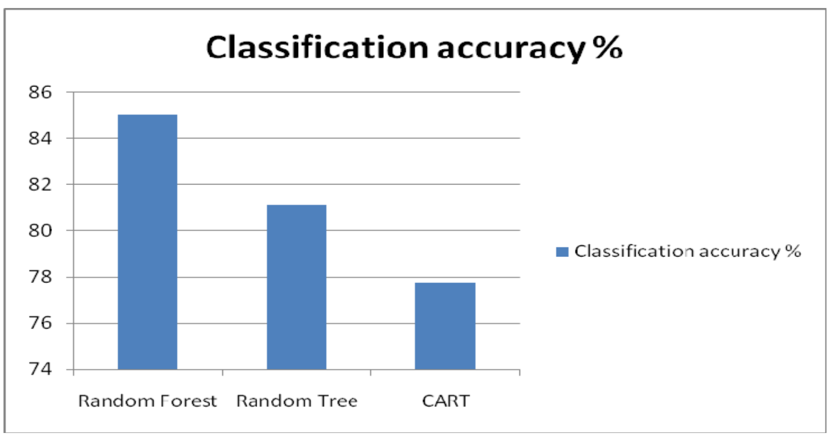


Figure I: Images used in our work.

We used the global signature of the images as we were interested in a broad spectrum image retrieval system. From table I we can observe that Random forest being an ensemble classifier has given the best results.

Table I: Classification accuracy by various Decision tree algorithms



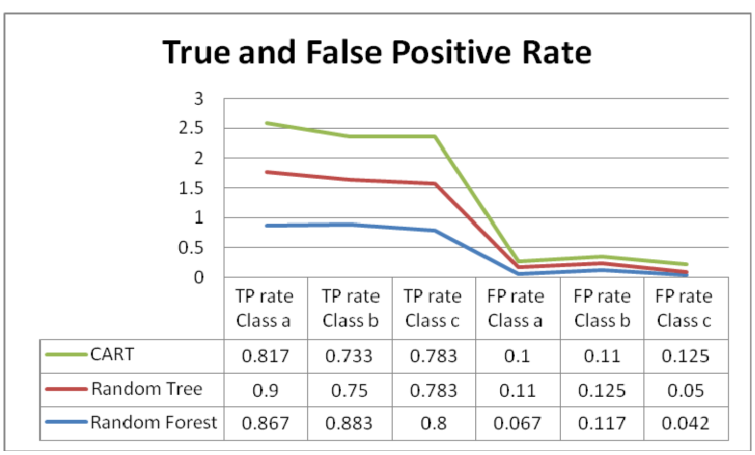


Figure II: The True positive and False Positive for the three class labels.

The True positive and False positive of all the three classification algorithms is shown in figure II.

V. CONCLUSION

In this paper we proposed to extract global features from medical images using the Hartley Transform. A subset of the extracted features was produced using the Zig Zag method. 180 images were used in the experimental setup and classified using Random forest, Random Tree and CART. Results obtained in classification accuracy were greater than 75% in all the three cases. Further work needs to be done in the area of preprocessing to improve the classification accuracy.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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

[6] 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,

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

[8] 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.

[9] 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.