Content Based Image Retrieval System for BI-RADS Standard Mammogram Images Using SVM Classifier

Dr. Vijay Kumar Joshi[1] and Shallu[2]
[12,]*Deptt.of CSE, KCCEIT College, NawanShahar Punjab INDIA*

Abstract- A long-term learning algorithm is required for a real-time system in order to accumulate the user's search information and utilize it to shorten the retrieval time and the relevance feedback process during future query sessions. In addition to the retrieval effectiveness, the usability of the human–computer interaction should be evaluated for the retrieval system. Usability attributes evaluated for the user interface include efficiency, memorability, errors and satisfaction. When the same query image is presented to the system more than once, there are even chances that the results may differ in terms of number of retrieved images or some new images may retrieve or some image may be missed. This relevance feedback i.e. images of close similarity, is to be analysed and additional or missed images are to be carefully examined so that either they may be removed from all results or made part of the result in each query for the same image. If relevance feedback is not taken care-off, the precision and recall indices will deteriorate. Presented work consists of Feature Set generation using Mass and Calcification Spot detection as Query Contents. Further, Classification using SVM and image retrieval based on similarity Index is discussed followed by Performance Evaluation using Recall and Precision.

I. INTRODUCTION

"Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. While image libraries are growing at a rapid rate (personal image collections may contain thousands, commercial image repositories millions of images, most images remain un-annotated, preventing the application of a typical text-based search. Content-based image retrieval (CBIR) does not require any extra data, as it extracts image features directly from the image data and uses these, coupled with a similarity measure, to query image collections.

Traditional methods for retrieving images is not very satisfactory or may not meet user demand

- E.g. In Google image typing 'Apple' returns the Apple products as well as the apple fruit.
- Main reason is the ambiguity in the language. Several other limitations.

CBIR systems compensates such issues by analyzing the actual 'content' of the image hence yielding a more effective feature for describing the image rather than user defined meta-data [9].

Doctors use a standard system to describe mammogram findings and results. This system (called the Breast Imaging Reporting and Data System or BI-RADS) sorts the results into categories numbered 0 through 6. By sorting the results into these categories, doctors all over the country can describe what they find on a mammogram using the same words and terms. This makes accurately communicating about these test results and following up after the tests much easier.

II. LITERATURE SURVEY

Content Based Image Retrieval is a method by which user can draw out an image from the database in which large number of images are stored, using queries. While designing content based image retrieval system one must take care while selecting the image features that best describes the image contents in a database. [1]

Now a days the requirement for content based image retrieval and video retrieval from media archives has attracted the attention of researchers. There are methods developed from research that provides a means of approaching to image and video data. The origin of these methods is from pattern recognition. [2]

Today E-lecturing is liked by many people, due to which there is a large increase in the number of video lectures available on the World Wide Web (WWW). For retrieving video information on WWW a most efficient method is required. [3]

Video retrieval system is in immense demand with the growth of multimedia data type and available bandwidth, as users adapt content based retrieval system from text based retrieval system. In video retrieval based on content choice of drawn out contents is very important irrespective of attributes of video examined. [4]

Video retrieval based on content simplifies the browsing and searching of huge image database over World Wide Web. Low level visual properties withdrawn from video frame based video analysis is organized in this method. [5]

In CBVIR system, based on ongoing position of research, we determine the approaches of video retrieval from spatial and temporal analysis. Later on, we show the video classification approaches from multidimensional distributed Hidden Markov Models. Ultimately, we are provided with abstract of future trends and open problems of video modeling, retrieval and classification based on content. [6]

Now days, video retrieval is a recent field which has its genealogy based on artificial intelligence, statistics, digital signal processing, natural language understanding, databases, psychology, pattern recognition, computer vision. [7]

In various fields like imaging of medicinal, Geographic Information systems, find something in space, from big archives image retrieval based on content is essentiality. Though, the method of extracting the appropriate images is generally occurring in advance by retrieving few critical characteristics that can perfect explain the archive images. [8]

Now days, videos are an effective and informative media that can acquire and show data. Development of digital video data at such a high speed has stimulated progress of recent technical knowledge for efficient browsing, explaining, and extraction of video data. [9]

Video indexing and retrieval stimulates the attraction of researchers globally, have a broad spectrum of favorable applications. [10]

Image pattern features are discussed in texture, color and radial domain for pattern identification and classification. The feature base can be used in BIR system for object extraction. [11, 12]

III. ALGORITHM

The data based used for the proposed work is MIAS data base for mammogram images in PGM format. The images are converted to jpeg format for enhancing of the mammogram. The image enhancement is performed using histogram equalization. The spot or affected area is extracted using the R-, G-, B- color based thresholds. The extracted spot is then brought under the feature extraction process.

IV. Mass Computation

Shape:This work uses the Zernike moments for describing mammographic masses as the Zernike basis function satisfies the orthogonal property image under test is unique and independent. Further, calculation of Zernike moments do not require knowledge of the precise boundary of an object, i.e., Zernike moments are insensitive to mass segmentation results. This makes Zernike moments suitable for representing masses with obscure boundaries.

Margin: Sobel operators is used to detect the margin of a segmented mass and then the resultant edge map showing the variation in gray levels is obtained to measure its sharpness degree, which is used as a feature to describe margin characteristics.

Density: As mammograms are presented in gray level, the density degree is represented by brightness variation of a mass.

V. CALCIFICATION SPOT EXTRACTION

Four features are extracted to describe calcification types as

Spot Size: All calcified spots are first found from the segmentation results, and then the average size of all the calcification spots is computed.

Spot Shape: Individual spot shapes are described by the first 15 Zernike moments.

Brightness: This is the average pixel value of all pixels inside the calcified spots.

Contrast: This is the ratio between the average pixel value of calcified spots and that of their surrounding regions, formed by using the morphological ‘dilation’ operation with a circular structuring element.

VI. SIMILARITY MEASURE

Similarity measure is performed using the Weighting Distance function and hierchichal arrangement of mammogram features. The feature arrangement in the similarity measurement hierarchy is designed to meet the following requirements:

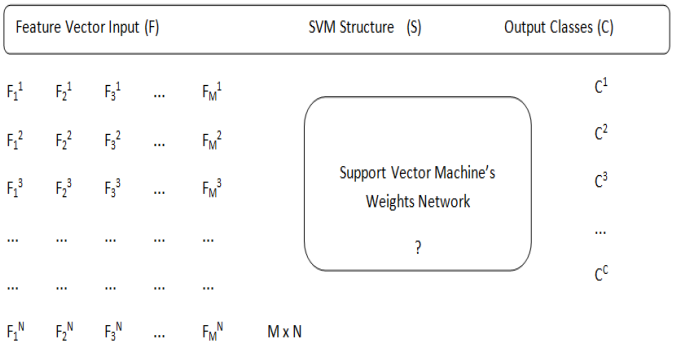
(1) The measure calculated at a lower layer determines the weighting factor of the next layer, excluding those dissimilar images from similarity calculation in the higher layer;

(2) The weighting should be proportional to similarity ranking.

The proposed distance weighting function is utilized for the similarity measure in the hierarchical structure. When the distance between the descriptor of the query image and a database image descriptor at the current feature layer is large, the weighting function will assign a smaller weight factor for the next feature layer, and vice versa.

VII. SVM TRAINING

SVM is trained using the known samples of leaf patterns in order to compute the weights of the SVM structure. In the training phase, input leaf feature vector and their respective class is know and weights network is unknown. Therefore, by training, the weight network is computed.



VIII. RESULTS AND DISCUSSION

Data Base Used: Online available MIAS data base for mammograms in PGM format

No. of Images= 210

No. of Test Images = 50

No. of Class = 5

Class-1 → Round

Class-2 →Lobular

Class-3 → Irregular

Class-4 → Oval

Class-5 → Micro-Lobular

Table 1-Proposed Algorithm Results for QBE for Mass Lesions

S. No.	**No. of Images**	Precision Lobular (%)	Precision Round (%)	Precision Oval (%)	Precision Irregular (%)	Precision Micro-Lobular (%)
1	15	90	95	91	95	97
2	20	91	93	94	90	95
3	25	89	94	88	88	96
4	30	89	92	85	88	96
5	35	87	90	80	87	97

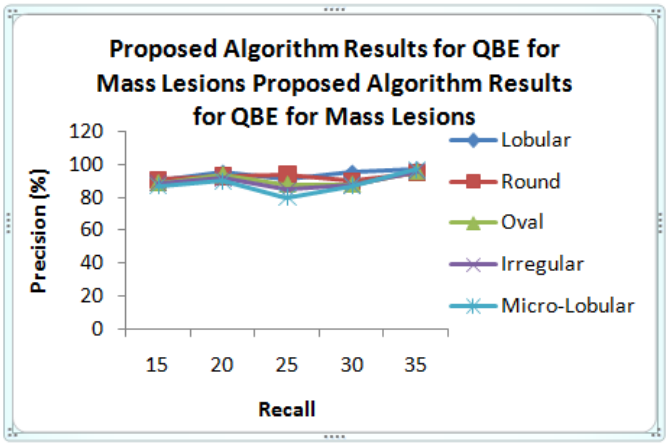


Fig.1: Proposed Algorithm Results

IX. CONCLUSION

The proposed work is targeted to be implemented on mammogram images as obtained from the hospital as well as on online image data base. The speed of the algorithm primarily depends upon the image size and therefore is expected to vary from image to image. The accuracy of retrieved image to that of the query image depends primarily upon the quality of query and data base image. Mammograms are usually at high noises. Therefore, it is recommended to use denoising algorithm before retrieval. The precision, recall and accuracy are discussed in result table and show a fine accuracy in retrieving the mammograms based on query mammograms.

X. REFERENCES

[1]. Kommineni Jenni1, Satria Mandala, Mohd Shahrizal Sunar," Content Based Image Retrieval Using Colour Strings Comparison", Procedia Computer Science 50 (2015) 374 – 379

[2]. SameerAntania,RangacharKasturia; RameshJainb," A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video", Pattern Recognition 35 (2002) 945–965

[3]. Bhagyashri Babhale,Kaumudinee Ibitkar,Priyanka Sonawane,Renuka Puntambekar ," Content Based Lecture Video Retrieval using Video Text Information", International Journal of Advance Foundation and Research in Computer (IJAFRC) Volume 2, Special Issue (NCRTIT 2015), January 2015. ISSN 2348 – 4853

[4]. B V Patel and B B Meshram ," CONTENT BASED VIDEO RETRIEVAL SYSTEMS", International Journal of UbiComp (IJU), Vol.3, No.2, April 2012

[5]. B. V. Patel and B. B. Meshram," CONTENT BASED VIDEO RETRIEVAL", The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.5, October 2012

[6]. Xiang Ma, Xu Chen, Ashfaq Khokhar and Dan Schonfeld," Content based Video Retrieval, Classification and Summarization: The State-of-the-Art and the Future",1-35

[7]. Madhav Gitte, Harshal Bawaskar, Sourabh Sethi, Ajinkya Shinde," CONTENT BASED VIDEO RETRIEVAL SYSTEM", Volume: 03 Issue: 06 | Jun-2014, Available @ http://www.ijret.org

[8]. Ammar Huneiti, Maisa Daoud," Content-Based Image Retrieval Using SOM and DWT",52-61

[9]. Dipika H Patel," Content based Video Retrieval: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 109 – No. 13, January 2015

[10]. Weiming Hu, Senior Member, IEEE, Nianhua Xie, Li Li, Xianglin Zeng, and Stephen Maybank," A Survey on Visual Content-Based Video Indexing and Retrieval", VOL. 41, NO. 6, NOVEMBER 2011

[11]. Goel V., Sohi B.S. and Singh H., "Radial Vectored Features for Pattern Enveloping and Retrieval," Journal of Optoelectronics and Advanced Materials, Vol. 18, No. 9-10, September - October 2016, pp. no. 878 – 883

[12]. Goel, V., Sohi, B.S. and Singh, H., 2017, February. "Statistical optimization of radial vectors for pattern reconstruction/retrieval" In Computational Intelligence & Communication Technology (CICT), 2017 3rd International Conference on (pp. 1-5). IEEE.