CONTENT BASED IMAGE RETRIEVAL SYSTEM FOR MAMMOGRAMS: A REVIEW

Syed Jahangir Nazir* and **Puneet Kumar Kaushal****

Abstract: A CBIR methods are proposed for the retrieval of the images from the mammogram databases. The various types of technique are required to fetch the images from the database depending upon the requirements. The ultimate goal of the retrieval system is to show the results to the radiologists in the form of display. Breast cancer is the fifth most common disease leading to the cancer deaths around the world and the second common type of cancer leading to the death among the women. Nowadays, the technique called as mammography is one of the leading technique used for the breast cancer detection among the women. Mammography is based on the X-ray examination of the breast. The breast disease such as nipple discharge, pain or lumps among the women are discovered and also diagnosed by the mammography. CAD-e has been used for detecting the breast cancer of the women, which has shown to support screening mammography. CBIR systems fetches out the visual features of the "query" image which is feed into the system. The visual features of the query image such as texture, colour, or shape are educed and the comparability is checked between the query images and the images stored in database is done by applying the particular similarity measurements. The images which are nearly similar in the database are brought back to the user. The performance of the CBIR system is evaluated on the basis of precision and the recall value.

Key Words: Mammograms, Mammography, content based image retrieval (CBIR) Computer-aided detection (CADe), Content Based Image Retrieval (CBIR), Query Image, Visual Features, Similarity Measurements.

1. INTRODUCTION

The leading cause of the death is caused by the Breast cancer among the women in the developed countries. By 2030, the World Health Organization (WHO) has figured that 11 million deaths is to be caused by the cancer. Among the top five rank of the cancer, the breast cancer among the women In the Hesperian countries is the leading cause of the death. The mammography is the leading technic for spotting of the breast cancer among the women and is used to discover and diagnose the breast disease such as cancer, nipple discharge, pain or lumps benign tumours and vesicles (cysts) earlier before being spotted by palpation. Both the breasts of the patient are screened in two way direction i.e., mediolateral oblique (MLO) and cranio-caudal (CC) views using the X-ray imaging tools and the radiologists keep on inspecting the radiographs visually. Computer-aided detection (CADe) is used and it has shown to support for imaging mammography. The technology of Content-based image retrieval (CBIR) is proving in aiding the diagnosis of the cancer. On the basis of optical features such as shape, colour and texture, CBIR is used to fetch the

---

* Department of Computer Science and Engineering. Lovely Professional University Phagwara
  Email: s.jahan31@gmail.com Address: - Block 34, Lovely Professional University Phagwara
** Department of Computer Science Engineering Lovely Professional University Phagwara
  Email: - kkpuneet@gmail.com Address: Block 33, Lovely Professional University Phagwara

Mammographic images from the large mammogram database. Based on visual or optical patterns defined by attributes and distance functions, the query image is posed to CBIR system and the visually most similar images are brought back from the database of mammograms. The second step of CBIR-CAD is automatic processing of the image retrieved from the database to obtain the diagnostic suggestion by the Radiologists.

2. EASE OF USE

2.1 General Architecture

The Pre-processing is performed to alter the resolution and grey level of the images into the optimum value being estimated from the size of a database for the purpose to extraction features. Then the various features are drawn from the query image and the mammograms stored in database. The images which include the masses are simply cropped from the original digitized form into a simple square form image. For the creation of the features KL expansion method is used. Now in the next step similarity is checked amongst the image fed in the system (query image) and the other images which are present within the database.
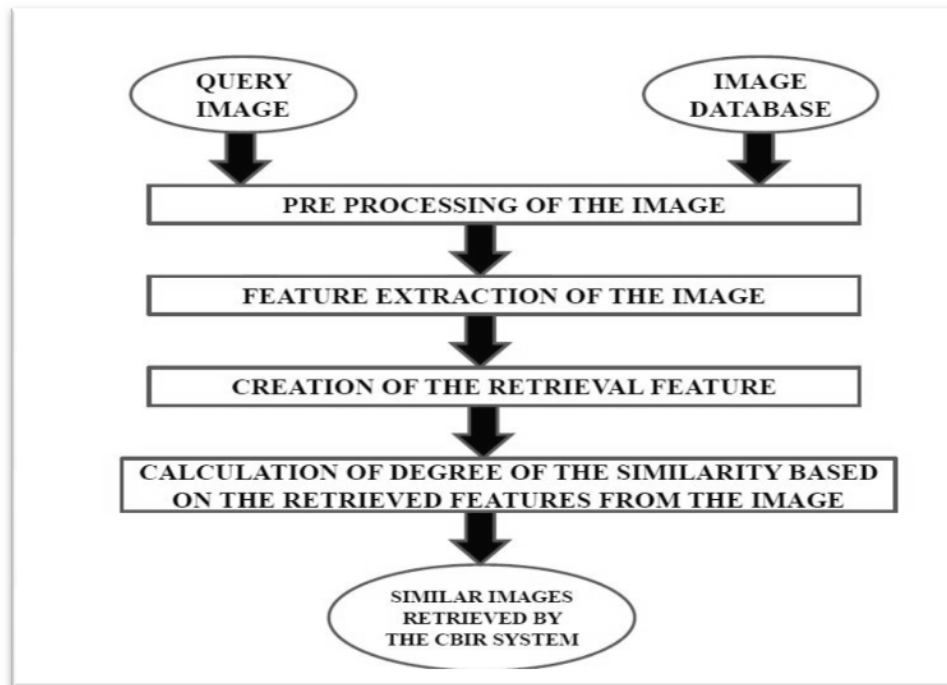
Figure 1.

3. LITERATURE REVIEW

Toshiaki Nakagawa (et al) aimed the concept of content-based image retrieval and showed the possible utility in the field of mammography. The method contained a local-pattern matching which is founded on Nth-order autocorrelation features with KL expansion (principal components analysis) to fetch out the same type of mass darkness on digital Mammogram images. [1]

T.J.Bose and P.Mythili proposed the hybrid approach for the content based image retrieval (CBIR) from the database. The method proposed fetches out the image from the database in two steps. In the initial step, the Self Organizing Map (SOM) also termed as the neural network is used for the purpose of bunching the images together based on their features. In the second step, the

search on the various subsets of the images which are having same basic features of the query image fed in the system by using the genetic algorithm (GA) based search. [2]

Thomas M. Deserno (at al) directs to move towards the computer-aided diagnostics for the detection of breast cancer among the women. The technique used for the content based image retrieval and feature extraction is two-dimensional (2D) principal component analysis with four eigenvalues and a support vector machine with Gaussian. The exactness of diagnosis of the mammograms is over 80% and may be used as alternative in screening the mammography. [3]

Menglin Jiang (et al) proposed the technique for the retrieval of images from the mammogram database and diagnosis based on the images retrieved from the database. The method used is for the detection of the region of interest (ROI) of the mammogram and SIFT descriptors are drawn out from the mammograms and are looked up in a vocabulary tree. All the quantized descriptors of the earlier diagnosed mammographic ROIs are stored in the vocabulary tree. Now the retrieved ROIs from the Vocabulary tree are used to check whether the query ROI are containing a mass or not. [4]

Gwenole Quellec (et al) in this article proposed the Content-Based Image Retrieval (CBIR) method for the diagnosis of various ailments of the human body. Images in the given technic, are indexed in a generic fashion such that domain-specific features are not extracted. From the wavelet transform of each image, a signature is created for every image. For the comparison of the two image signatures the distance measure is defined and thus the similar images are fetched out from the database when the physician feeds in the a query image into the system. [5]

Lazaros Tsochatzidis (et al) proposed a Content Based Image retrieval (CBIR) method called as support vector machine (SVM) that employs the extracted features as the input. The involvement value of each support vector machine (SVM) consists of the end features which are used for the Content Based Image Retrieval. [6]

Stylianos Tzikopoulos proposed the two computer-aided diagnosis (CAD) systems. These systems are implemented and their evaluation performance is measured. In the first system, machine-driven classification and segmentation of the mammograms is done on the basis of breast density idea and detection of asymmetry in the mammograms. Image pre-processing and segmentation methods are applied first. Secondly, the extraction of the features for breast density categorization is done and Support Vector Machines (SVMs) are applied for classification and hence attaining the accuracy measure up to the level of 85.7%. [7]

K. Vaidehi and T.S. Subashini proposed the method to fetch out same mammographic images based on the type of breast tissue density of the given query image. The mean of extracted features are fed in as input to the SVM classifier for the purpose of classification of the tissue density into any of the three classes namely dense, glandular and fatty. After classification the mammogram images are stored into three separate databases namely dense, glandular and fatty along with its feature vector. Separation of each database into two clusters is done by using K-means clustering algorithm. For retrieval of the mammograms based on content of the given query image of the system, firstly classification of the query image is done into any of the three dense, glandular and fatty tissue class. Then the comparison of the feature vector of the query image is done with the two cluster centroids of the matching class, in order to limit the search within the closest cluster. The highest performance rate ranging between 98 and 99% is achieved. [8]

Issam El-Naqa (et al) describes the content based retrieval method of medical images from a Database. The neural networks and support vector machines are used to tell in advance the user's belief of similarity. The similarity is learned from the training examples which are provided by human observers. For the optimization of retrieval effectiveness and efficiency, a cascade of a binary classifier and a regression module is used. [9]

J´ulia E. E. de Oliveira (et al) proposed content-based image retrieval (CBIR) system called as MammoSVD. This content-based image retrieval CBIR system is formulated on the basis of breast density i.e. Fatty or dense. To perform the retrieval operation on the database the Support-vector machine (SVM) is used. The performance of the system is evaluated primarily on the 10% of the retrieved images from the database. The accuracy of the system is achieved up to 90%. [10]

K.Vennila (et al) has aimed the detection of the masses of the Digital Mammograms where the mass is detected by using Otsu thresholding in which digital mammogram is subjected to morphological pre-processing, artifacts removal, and pectoral muscle segmentation by using seeded region growing algorithm. The Mammographic Image Analysis Society (MIAS) database is used for the test of the method. [11]

S. Mangijao Singh (et al) proposed a method for content-based image retrieval (CBIR) in which features are combined such as texture and colour. The discriminating power of colour indexing techniques is improved by encoding minimum number of spatial information in the colour index. As the colour features of an image, the image is fractioned into same non-intersecting three areas horizontally. The first three moments of the colour distribution are extracted from each of the region of the image, from every colour channel and are stored in the index i.e., for a HSV colour space and then 27 floating point numbers per image are stored. Gabor texture descriptors are took over for the texture features. The values are assigned to each feature and Canberra distance is used to compute the resemblance with combined features of texture and colour. [12]

Karin Kailing (et al) proposed the two models for image representation which are Image Representation as Containment tree and Image Representation as Segmentation Graphs which Incorporates the structural features and content features in a tree and a graph data structure respectively. The combination of these models enhances the retrieval accuracy of the images from the database. [13]

Pranali Prakash Lokhande (et al) Proposed a productive picture recovery procedure or the image retrieval method which utilizes dynamic dominant colour, texture and shape elements of a image. By applying this method, in an initial step a picture can be consistently isolated or divided into coarse allotments or partitions. [14]

Ms. Pragati Ashok Deole (et al) proposed the content based image retrieval strategy which is applied to fetch the query or inquiry image from the vast image database utilizing three components which are shape, colour and surface or texture. The primary aim is to order or the classification of image by using K-nearest neighbours Algorithm (KNN). [15]

Ms. K. Arthi (et al) proposed an effective image retrieval algorithm which is dependent on colour co-occurrence matrix (CCM). The CCM for every pixel of an image is discovered utilizing (HSV) Hue Saturation value of the pixel and after that it is compared with CCM of the images in the database and the images are recalled or fetched out from the database. [16]

K. Haridas (et al) experiments with the different methods which are used for content based image retrieval system. These strategies are applied and tried in the light of three parameters like exactness worth, recall value and Precision rate or accuracy. The experimentation results demonstrate that FCTH (Fuzzy Colour and Texture Histogram) technique is more productive when comparing with other different methods. The exactness indicated by the (FCTH) for image retrieval system is 93.21 %. [17]

N.Puviarasan (et al) Proposed methodology which is a combination of features like shape and texture. For the portioning of the images, the clustering methods are the K-means clustering and Fuzzy C-means clustering. After that 5 Haralick texture features, 7 features of shape are drawn out. The features which are drawn out or extracted are combined together. [18]

Aruna Verma (et al) proposed a technique for aggregating all the three features i.e. texture, colour and shape data and accomplish greater image retrieval effectiveness. Content based image retrieval (CBIR) technique is proposed by adventure the wavelets, which represent the visual component. The image extraction and similarity match is achieved by using Haar wavelet, which breaks down the colour images into multilevel scale and wavelet coefficients. The Two images are then viewed and are compared with each other by checking if their feature vectors lay close in the feature space. [19]

Prof. C. S. Gode(et al) proposed the model which combines all the three image features such as texture, shape, colour data and the higher image retrieval effectiveness is accomplished. The 18 partitions of the image and its compliment is done into non-intersecting thin rectangular slabs of equivalent size. The features serve as nearby or local descriptors of colour, shape and texture which are drawn out from the conditional co-occurrence histograms amongst the image slabs and corresponding compliment slabs in RGB colour space. [20]

4. CONCLUSION

The various Image processing techniques are used in the content based image retrieval systems to extract or fetch out visual features such as colour, shape and from images. This study discuss about the retrieval of the mammogram Region of interest (ROI) from the mammogram images database based on the query image. This system is very helpful to the radiologist for making effective diagnosis and making decisions based on the retrieved Region of interest (ROI) from the mammograms stored in the database. The performance of the Content Based Image Retrieval System (CBIR) is studied by calculating the recall value and the precision value. These two values are used for the purpose of improving the retrieval of the images from the database.

Precision value is defined as the ratio of retrieved images which are similar to the input or query image feed into the system. The precision value takes all the retrieved images into account but it can also be evaluated at a given cut-off rank , conceiving only the topmost results retrieved by the Content Based Image Retrieval System CBIR system.

$$\text{Precision} = \frac{\text{total no. relevant retrieved images}}{\text{total no. of images retrieved}}$$

Recall value is defined as the ratio of the image which are similar to the query image that is successfully retrieved from the database. In the binary classification, the recall is called as sensitive. So it can be seen as the probability that a relevant image is retrieved by the query image. It is trivial to attain recall of 100% by retrieving all the images in response to the query image. Therefore recall is not sufficient but it is also needed it to measure the number of non-relevant images retrieved from the database.

$$\text{Recall} = \frac{\text{total no. of relevant retrieved images}}{\text{total no. of relevant images}}$$

References

[1]   Toshiaki Nakagawa, Takeshi Hara, Hiroshi Fujita, Takuji Iwase and Tokiko Endo.   (2004) Local Pattern Matching method for image retrieval and its application to digital mammograms. International Conference Visualization, Imaging, and image processing September 6- 8,2004,Marbella, Spain.

[2]   T.J.Bose and P.Mythili (2009) Neural Network and Genetic Algorithm based Hybrid model for content based mammogram image retrieval. Journal of applied Sciences 9(19):3531-3538, 2009

[3]   Thomas M. Deserno, Michael Soiron, J´ulia E.E. de Oliveira and Arnaldo de A. Ara´ujo (N\A) Towards Computer-Aided Diagnostics of screening mammography using content-based image retrieval

[4] Menglin Jiang, Shaoting Zhang, Jingjing Liu, Tian Shen, Dimitris N. Metaxas. (N\A) Computer-Aided Diagnosis of mammographic masses using vocabulary tree-based image retrieval.

[5] Gwenole Quellec, Mathieu Lamard Guy Cazuguel, Beatrice Cochener, Christian Roux.(2010) Wavelet Optimization for content-based image retrieval in medical databases. Med image Anal. 2010 Apr;14(2):227-41

[6] Lazaros Tsochatzidis, Konstantinos Zagoris, Michalis Savelonas, and Ioannis Pratikakis. (N\A) SVM-based cbir of breast masses on mammograms.

[7] Stylianos Tzikopoulos (N/A) Analysis and Retrieval of mammographic images.

[8] K. Vaidehi and T.S. Subashini (2014) Content based mammogram retrieval based on breast tissue characterization using statistical features. Research Journal of applied sciences, Engineering and technology 8(7):871-878, 2014

[9] Issam El-Naqa, Yongyi Yang, Nikolas P. Galatsanos, Robert M. Nishikawa and Miles N.Wernick. (2004) A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 23, NO. 10, OCTOBER 2004

[10] J´ulia E. E. de Oliveira, Ana Paula B. Lopes, Guillermo C´amara-Chavez and Arnaldo de A. Ara´ujo. (N\A) MAMMOSVD: A Content-based image retrieval system using a reference database of mammographies.

[11] K.Vennila,K.Sivakami, R.Padmapriya.(2014) Detection of Mass in Digital Mammograms. *International Journal of Computer Applications (0975 − 8887) Volume 104 − No.5, October 2014*

[12] S. Mangijao Singh , K. Hemachandran (2012) Content-Based Image Retrieval using Color Moment and Gabor Texture Feature. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012 ISSN (Online): 1694-0814

[13] Karin Kailing, Hans-Peter Kriegel, and Stefan Schonauer (2004) Content-Based Image Retrieval Using Multiple Representations. Proc. 8th int. conf. on knowledge based intelligent information and engineering systems (KES'04), Wellington, New Zealand, LNAI 3214,pp. 982- 988,2004

[14] Pranali Prakash Lokhande, P. A. Tijare (2012) Feature Extraction Approach for Content Based Image Retrieval. Volume 2, Issue 2, February 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

[15] Ms. Pragati Ashok Deole, Prof. Rushi Longadge.(2014) Content Based Image Retrieval using Color Feature Extraction with KNN Classification. International Journal of Computer Science and Mobile Computing, Vol.3 Issue.5, May- 2014, pg. 1274-1280

[16] Ms. K. Arthi Mr. J. Vijayaraghavan.(2013) Content Based Image Retrieval Algorithm Using Colour Models. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2013

[17] K. Haridas, Dr. Antony SelvadossThanamani, (2014)Efficient Content Based Image Retrieval System in Visual Words, Color and Edge Directive Descriptors and Fuzzy Color and Texture Histogram. International Journal of Innovative Research in Computer and Communication Engineering *(An ISO 3297: 2007 Certified Organization)* Vol. 2, Issue 10, October 2014

[18] N.Puviarasan, Dr.R.Bhavani, A.Vasanthi.(2014) Image Retrieval Using Combination of Texture and Shape Features. International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 3, March 2014

[19] Aruna Verma, Deepti Sharma. (2014) Content Based Image Retrieval Using Color, Texture and Shape Features. ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering

[20] Prof. C. S. Gode, Ms. A. N. Ganar. (2014) Image Retrieval by Using Colour, Texture and Shape Features. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering *(An ISO 3297: 2007 Certified Organization)* Vol. 3, Issue 4, April 2014