Segmentation of Digital Mammograms Using Clustering Algorithm and Evolutionary Algorithm

[1]Athira krishnan and [2]R.Sivakumar
[1]M.E (Computer Science & Engineering) and [2]Professor and Head,
[2]Department of IT, [1,2]Tamilnadu College of Engineering
Karumathampatti, Coimbatore, India.

Abstract: Breast cancer is one of the major causes for the increase in mortality among women, especially in developed countries. Micro calcifications in breast tissue is one of the most incident signs considered by radiologist for an early diagnosis of breast cancer, which is one of the most common forms of cancer among women. In this paper the thresholding algorithm is applied for the breast boundary identification and a tracking algorithm is introduced for pectoral muscle determination in Mammograms. Fuzzy C-means Clustering algorithm (FCM) and Artificial Bee colony algorithm is frequently used for image segmentation purpose. It has the advantage of giving good modeling results in many cases, although, it is not capable of specifying the number of clusters by itself.

Keywords–Mammography, CAD, modified tracking algorithm, Fuzzy C-means clustering.

I. INTRODUCTION

Computer technology has a tremendous impact on medical imaging. The interpretation of medical images, however, is still almost exclusively the work of humans. In the next decades, the use of computers in image interpretation is expected to increase vastly. Breast cancer is a malignant tumor that starts in the cells of the breast. A malignant tumor is a group of cancer cells that can grow into (invade) surrounding tissues or spread (metastasize) to distant areas of the body. The disease occurs almost entirely in women, but men can get it, too. Mammography is widely used as a principal breast cancer screening method, however mass screening is generating large number of images.

A mammogram is a low-dose x-ray exam of the breasts to look for changes that are not normal. The results are recorded on x-ray film or directly into a computer for a doctor called a radiologist to examine. Digital mammography has the potential to offer several advantages over traditional film mammography, including: faster image acquisition, shorter exams, easier image storage, easy transmission of images to other physicians and computer processing of breast images for more accurate detection of breast cancer [1, 2].

II. PREPROCESSING AND ENHANCEMENT

A. Image Preprocessing:

Image processing modifies pictures to improve them (enhancement, restoration), extract information (analysis, recognition), and change their structure (composition, image editing). Images can be processed by optical, photographic, and electronic means, but image processing using digital computers is the most common method because digital methods are fast, flexible, and precise. An image can be synthesized from a micrograph of various cell organelles by assigning a light intensity value to each cell organelle. The sensor signal is "digitized" converted to an array of numerical values, each value representing the light intensity of a small area of the cell.

Digitized values are called picture elements, or "pixels," and are stored in computer memory as a digital image. A typical size for a digital image is an array of 512 by 512 pixels, where each pixel has value in the range of 0 to 255. The digital image is processed by a computer to achieve the desired result [3, 4, 5].

B. Image Enhancement:

Image enhancement improves the quality (clarity) of images for human viewing. Removing blurring and noise, increasing contrast, and revealing details are examples of enhancement operations. For example, an image might be taken of an endothelial cell, which might be of low contrast and somewhat blurred. Reducing the noise and blurring and increasing the contrast range could enhance the image. The original image might have areas of very high and very low intensity, which mask details. An adaptive enhancement algorithm reveals these details. Adaptive algorithms adjust their operation based on the image information (pixels) being processed. In this case the mean intensity, contrast, and sharpness (amount of blur removal) could be adjusted based on the pixel intensity statistics in various areas of the image [6. 7].

C. Database (Image Acquisition):

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database (ftp://peipa.essex.ac.uk). The data used in these experiments was taken from the MIAS. The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Programme and digitized with a Joyce-Lobel scanning microdensitometer. The database contains left and right breast images for 161 patients, is used. Its quantity consists of 322 images, which belong to three types such as Normal, benign and malignant. There are 208 normal, 63 benign and 51 malignant (abnormal) images. Figure 1 shows the input image.

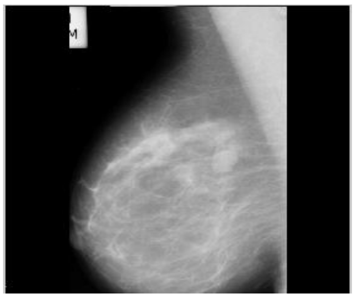

Figure 1: The input image.

D. Thresholding Algorithm-Label Removal:

First step of the algorithm is to identify breast boarder. In many mammograms back ground objects with high intensity values make breast boundaries identification a challenging task, especially for the scanned ones where the original film has some artifacts [8, 9]. The following algorithm is used to exclude back ground objects (noise) and identify breast boundaries.

-Using the histogram curve, the threshold point is determined for the input image (10% from the bottom of the image).

- Convert the given input image into binary image using the threshold value from the histogram curve.
- Label is extracted and the binary image is obtained.
- Since breast object is the biggest object, then the object with maximum number of pixels is identified to be the breast. Source mammogram is ANDed with mask image and the resulting image is the denoised using median filtering process. The proposed algorithm aims to reduce the complexity for abnormality detection by identifying breast object and excluding pectoral muscle using modified tracking algorithm for the mammogram images. Figure 2 shows the binary image and label removed image.

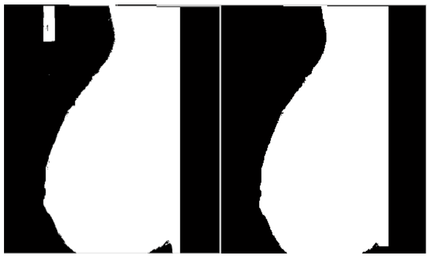

Figure 2: Binary image and label removed image.

E. Adaptive Filter:

An adaptive filter is a filter that self-adjusts its transfer function according to an optimization algorithm driven by an error signal. Because of the complexity of the optimization algorithms, most adaptive filters are digital filters. By way of contrast, a non-adaptive filter has a static transfer function. Adaptive filters are required for some applications because some parameters of the desired processing operation (for instance, the locations of reflective surfaces in a reverberant space) are not known in advance. Considering an example, a hospital is recording a heart beat (an ECG), which is being corrupted by a 50 Hz noise (the frequency coming from the power supply in many countries). However, due to slight variations in the power supply to the hospital, the noise signal may contain harmonics of the noise and the exact frequency of the noise may vary. One way to remove the noise is to filter the signal with a notch filter at 50 Hz. Such a static filter would need to remove all the frequencies in the vicinity of 50 Hz, which could excessively degrade the quality of the ECG since the heart beat would also likely have frequency components in the rejected range.

To circumvent this potential loss of information, an adaptive filter could be used. The adaptive filter would take input both from the patient and from the power supply directly and would thus be able to track the actual frequency of the noise as it fluctuates. Such an adaptive technique generally allows for a filter with a smaller rejection range, which means, in our case, that the quality of the output signal is more accurate for medical diagnoses.

The Characteristics of adaptive filters are as follows:

– Automatically adjust (adapt) in changing system.

– Can perform specific filtering or decision-making.

– Have adaption algorithm for adjusting parameters.

F. Normalization:

Normalization is a process that changes the range of pixel intensity values. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching. In more general fields of data processing, such as digital signal processing, it is referred to as dynamic range expansion. The purpose of dynamic range expansion in the various applications is usually to bring the image, or other type of signal, into a range that is more familiar or normal to the senses, hence the term normalization [10, 11].

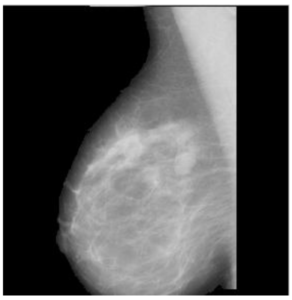

Figure 3: Normalized Image.

III. PROPOSED ALGORITHM

A. Pectoral Muscle Removal Algorithm:

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms, and can affect the results of image processing methods. Intensity-based methods, for example, can present poor performance when applied to differentiate dense structures such as the fibro-glandular disc or small suspicious masses, since the pectoral muscle appears at approximately the same density as the dense tissues of interest in the image. And to increase the reliability of boundary matching, the pectoral muscle can be removed from the breast region. Another important need to identify the pectoral muscle lies in the possibility that the local information of its edge, along with an internal analysis of its region, may be used to identify the presence of abnormal axillary lymph nodes, which may be the only manifestation of occult breast carcinoma.

Algorithm: Removal of Pectoral Muscle Region

S Read the mammogram image

[m,n] Size of the image

th thresholding algorithm used to convert the gray scale image to binary image**.**

Procedure:

- Consider, the histogram curve the threshold point is determined for the input image (15% from the bottom of the image).
- Convert the given input image into binary image using the threshold value from the histogram curve.
- Label is extracted and the binary image is obtained.
- Since breast object is the biggest object, then the object with maximum number of pixels is identified to be the breast.

Figure 4 shows the thresholded image and Removal of Pectoral Muscle Region in the mammogram image using modified tracking algorithm.

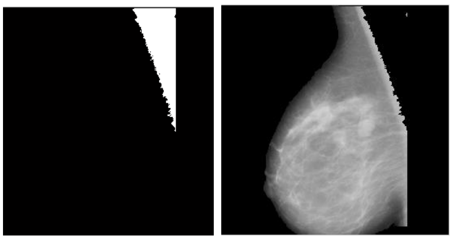

Figure 4: Threshold image and Pectoral Muscle Region removed from the mammogram image.

B. Fuzzy C-Means Clustering Algorithm:

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. Among the fuzzy clustering method, the fuzzy c-means (FCM) algorithm is the most well known method because it has the advantage of robustness for ambiguity and maintains much more information than any hard clustering methods. The algorithm is an extension of the classical and the crisp k-means clustering method in fuzzy set domain. It is widely studied and applied in pattern recognition, image segmentation and image clustering, data mining, wireless sensor network and so on.

FCM Algorithm

- Let x_i be a vector of values for data point g_i.

1. Initialize membership $U^{(0)} = [\ u_{ij}\ ]$ for data point g_i of cluster cl_j by random
2. At the k-th step, compute the fuzzy centroid $C^{(k)} = [\ c_j\ ]$ for j = 1, .., nc, where nc is the number of clusters, using

$$c_j = \frac{\sum_{i=1}^{n} (u_{ij})^m x_i}{\sum_{i=1}^{n} (u_{ij})^m}$$

where m is the fuzzy parameter and n is the number of data points.

Figure: 5 show the segmented image using the fuzzy C means algorithm.

Figure 5: FCM segmented Image

Segmentation of Image With ABC:

The ABC algorithm is developed by inspecting the behaviors of the real bees. The real bees are finding food source, which is called the nectar, and sharing the information of food sources to another bees in the nest. In ABC model, the colony consists of three groups of bees: employed bees, onlookers and scouts [12]. It is assumed that there is only one artificial employed bee for each food source. In other words, the number of employed bees in the colony is equal to the number of food sources around the hive. Employed bees go to their food source and come back to hive and dance on this area. The employed bee whose food source has been abandoned becomes a scout and starts to search for finding a new food source. Onlookers watch the dances of employed bees and choose food sources depending on dances.

In the ABC algorithm, a food source position stands a possible solution of the problem to be optimized which is represented by a d-dimension real-valued vector. The nectar amount of a food source corresponds to the quality of the associated solution. The number of employed bees or the onlookers is equal to the number of the food sources in the population. In ABC, a population based algorithm, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. The number of the employed bees is equal to the number of solutions in the population.

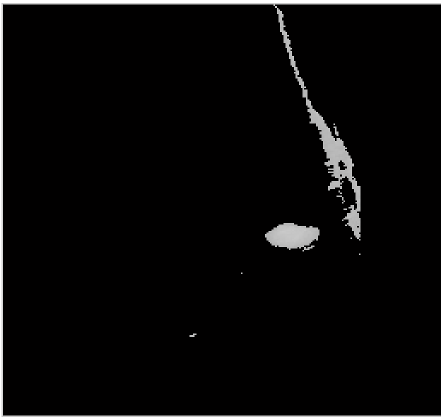

Figure: 6 ABC segmented image

CONCLUSION

In this paper a new pre-processing algorithm is introduced. The thresholding algorithm is used to identify the breast boarder and modified tracking algorithm is used to remove the pectoral muscles. Using the fuzzy C Means clustering algorithm and evolutionary algorithm the mammogram image is segmented effectively. This outcome will be used to further analysis of breast tissue more accurately than other algorithms.

REFERENCES

[1] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, A.F Frere, Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets, IEEE Trans. Med. Imag. 20 (9) (2001) 953-964.

[2] M. Karnan, K. Thangavel,"Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications"Computer Methods and Programs in bio medicine, Elsevier Ltd 8 7 (2 0 0 7) 12-20.

[3] Lobregt, S. & Viergever, M. A. (1995),"A discrete dynamic contour model", IEEE Trans. Med. Imaging 14(1), 12–24.

[4] A.J. Mendez, P.G. Tahocesb, M.J. Lado, M. Souto, J.L. Correa, J.J. Vidal, Automatic detection of breast border and nipple in digital mammograms, Comput. Meth. Prog. Biomed. 49 (1996)253-262.

[5] Griffiths & Tenenbaum,"From algorithmic to subjective randomness, In advances in neural information processing", 2004.

[6] H Sheshadri, A Kandaswamy,"Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms", Medical Imaging and Information Sciences vol. 23 2006.

[7] K. Thangavel, M. Karnan, R. Siva Kumar, A. Kaja Mohideen, Automatic detection of microcalcification in mammograms—a review, Int. J. Graph. Vision Image Process. 5 (5) (2005)31-61.

[8] K. Thangavel, M. Karnan, CAD system for preprocessing and enhancement of digital mammograms, Int. J. Graph. Vision Image Process. 9 (9) (2006) 69-74.

[9] Bezdek JC (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.

[10] Das S, Konar A, Chakraborty UK (2006). Automatic Fuzzy Segmentation of Images with Differential Evolution, In IEEE Congress on Evolutionary Computation, pp. 2026-2033.

[11] Jain A, Dubes R (1998). Algorithms for Clustering Data, Prentice-Hall, Englewood Cliffs, NJ.

[12] R.Sivakumar , M.Karnan, "A Novel Approach to Detect Microcalcification in Mammogram Image Using Evolutionary Algorithm", CiiT International Journal of Artificial Intelligent Systems and Machine Learning. Vol 4, No.3, pp 135-140, Mar 2012.