GUI Based Automatic Breast Cancer Mass and Calcification Detection in Mammogram Images using K-means and Fuzzy C-means Methods

Nalini Singh, Ambarish G Mohapatra, Biranchi Narayan Rath, and Guru Kalyan Kanungo

Abstract—**Mammogram breast cancer images have the ability to assist physicians in detecting brest cancer caused by cells abnormal growth.The first step of the cancer signs detection should be a segmentation procedure able to distinguish masses and micro calcifications from background tissue. This study is an attempt to reduce false alarm in Breast cancer detection. This paper presents a research on mammography images using K-means and Fuzzy C – means clustering for detecting cancer tumor mass and micro calcification. The proposed technique shows better results in less time (in Seconds) and user friendly as it is based on Graphical User Interfaces(GUI). The real time implementation of the proposed method can be implemented using data acquisition hardware and software interface with the mammography systems.**

Index Terms— **Clustering, Fuzzy C–means, GUI, K-means, Mammography, Segmentation.**

I. Introduction

A. Techniques Used for Breast Cancer Measurement

Breast image analysis can be performed using X-rays, magnetic resonance, nuclear medicine or ultrasound [1].

1) X-Ray Mammography

X-Ray Mammography is commonly used in clinical practice for diagnostic and screening purposes [2]. Mammography provides high sensitivity on fatty breast and excellent demonstration of micro calcifications; it is highly indicative of an early malignancy.

2) MRI Of The Breast

Magnetic Resonance Imaging is the most attractive alternative to Mammography for detecting some cancers which could be missed by mammography. In addition, MRI can help radiologists and other specialists determine how to treat breast cancer patients by identifying the stage of the disease [1]-[2].

B. Research Goal

The overall research of this project is to detect the breast cancer by using fuzzy k-means & fuzzy c-means algorithm [7]. Since screening mammography is currently the main test for early detection of breast cancer, a huge number of mammograms need to be examined by a limited number of radiologists, resulting misdiagnoses due to human errors by visual fatigue. In the previous work we were not able find out the total cancer affected area. It was only able to find out the masses of the tumor. Currently, there are several image processing methods proposed for the detection of tumors in mammograms. In this paper we have proposed a new technique for cancer mass detection of the mammogram image. Various technologies such as thresholding, intensity level slicing, contrast stretching, image negative, power transform, logarithm transform and segmentations have been designed for analysis of tumor and tumor like structures [2].

II. Literature Review

1. Jinshan Tang Rangayyan, R.M. Jun Xu El Naqa, and I. Yongyi Yang Dept. of Adv. Technol., Alcorn State Univ., Lorman, MS in their paper "*Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances*" described about an overview of recent advances in the development of CAD (Computer -aided diagnosis or detection) systems and related techniques for breast cancer detection and diagnosis.
2. Xia Xiao Kikkawa, T. Sch. of Electron. and Inf. Eng., Tianjin University, China, in their paper "*Early Breast Cancer Detection with Hemi-Elliptical Configuration by UWB Imaging*" described about an overview of Ultra-wideband (UWB) microwave imaging method for the early breast cancer detection physically based on the large contrast of electromagnetic parameters between the malignant tumor and the normal breast tissue.

III. Proposed Method for Breast Cancer Detection

We have proposed K-means clustering and Fuzzy C-Means clustering which are very similar in approaches [1]-[2].

A. K-Mean Clustering Algorithm

In statistics and machine learning, k-means clustering is a method of cluster analysis which aims to partition 'n' observations in to 'k' clusters in which each observation belongs to the cluster with the nearest mean [7]. For a given set of observation $(x_1, x_2, ...x_n)$, where each observation is a d-dimensional real vector, then k-means clustering aims to partition the 'n' observations in to 'k' sets (k<n), $\{s = s_1, s_2, ...s_n\}$ so as to minimize the within

Manuscript received November 18, 2011, revised February 3, 2012. This work was supported in part by Silicon Institute of Technology, India.

Nalini Singh, Ambarish G Mohapatra, Biranchi Narayan Rath and Guru Kalyan Kanungo are with the Silicon Institute of Technology, Silicon Hills, Patia, Bhubaneswar-751024, India (email: nalini.singh@silicon.ac.in; ambarish.mohapatra@silicon.ac.in; biranchi.rath@silicon.ac.in; gurukalyan@rocketmail.com).

cluster sum of squares (WCSS) in equation (1).

$$\arg\min_s \sum_{i=1}^{k} \sum_{x_j \in s_i} \left\| x_j - \mu_i \right\|^2 \tag{1}$$

where, μ_i is the mean of s_i .The number of cluster k is assumed to be fixed in k-means clustering.

Standard algorithm

Given an initial set of k-means in which may be specified randomly or by some heuristic, the algorithm produces by alternating between two steps.

1) Assignment Step

Assign each observation to the cluster with the closest mean (i.e. partition the observation according to the voronoi diagram generated by the means) in equation (2).

$$s_i = \{x_j : \| x_j - m_i^2 \| \le \| x_j - m^2{}_{i^*} \| \forall i^* = 1......k\} \tag{2}$$

2) Update Step

Calculate the new means to be the centroid of the observations in the cluster in equation (3)

$$m_i^{t+1} = \frac{1}{|s_i^t|} \sum_{x_j \in s_i^t} x_j \tag{3}$$

The algorithm is usually very fast, it is common to run it multiple times with different starting conditions. Theoretically it has seen that there exit certain point sets on which k-means takes super-polynomial time, but practically it is not so far.

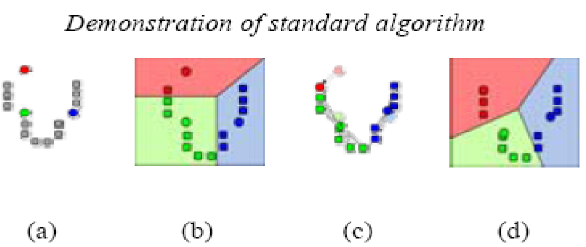


Fig. 1. Demonstration of Standard Algorithm

The above Fig. 1 shows the demonstration of standard algorithm.

(a) *k* initial "means" (in this case *k*=3) are randomly selected from the data set (shown in color)

(b) *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

(c) The centroid becomes the new means

(d) Steps (b) and (c) are repeated until convergence has been reached.

B. FUZZY C-Mean Algorithm

The Fuzzy C-means algorithm, also known as fuzzy ISODATA, is one of the most frequently used methods in pattern recognition. Fuzzy C-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [7]. It is based on the minimization of objective function to achieve a good classification. 'J' is a squared error clustering criterion, and solutions of minimization are least squared error stationary point of 'J' in equation (4).

$$j_m = \sum_{i=1}^{k} \sum_{j=1}^{c} u_{ij} \left\| x_i - c_j \right\|^2 \tag{4}$$

where, $1 \le m \le \infty$

Where 'm' is any real number greater than 1, is the degree of membership of in the cluster 'j', is the dimensional measured data, is the dimension center of the cluster and is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of member ship u_{ij} in equation (5) and the cluster centers c_j by equation (6)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right]^{\frac{2}{m-1}}} \tag{5}$$

$$c_j = \frac{\sum_{i=1}^{k} u_{ij} \cdot x_i}{\sum_{i=1}^{k} u_{ij}} \tag{6}$$

The iteration will stop when

$$\max_{ij} = \{ \left| u_{ij}^{k+1} - u_{ij}^{k} \right| \} < \in \tag{7}$$

where $\in$ is the termination criterion between 0 & 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of j_m .

The fuzzy c means algorithm composed of following steps.

1. Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$
2. At k-step calculate the center vectors $C^{(k)} = [C_j]$ with $U^{(k)}$.
3. Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right]^{\frac{2}{m-1}}}$$

4. If $\left\| u_{ij}^{k+1} - u_{ij}^{k} \right\| < \in$ then STOP, otherwise return to step 2.

C. Performance Evaluation of K-means & Fuzzy C-Means

The main difference is that, in FCM, each point has a weighted associated with a particular cluster. So, a point doesn't in a cluster as much as has a weak or strong association to cluster, which is determined by the inverse distance to the center of the cluster. FCM will tend to run slower than K-means, since it is actually doing more work [7]. Each point is evaluated with each cluster, and more operations are involved in each evaluation. K-means just needs to do a distance calculation, whereas FCM needs to do a full inverse-distance weighting. In this proposed work we make these differences or weakness our strong point for full detection of breast cancer. From this we were able to find out the masses as well as the cancerous area i.e. how far the cancer has affected the breast [5].

IV. Results and Discussions

This section details the detection of breast cancer mass and calcification in mammograms using image processing functions [3], K-means clustering [7], and fuzzy C-means algorithm [8].

A. Gray Level Transformations (Negative Image)

The negative of an image with gray levels in the range [0, L-1] is obtained by using the negative transformation as given below in equation (8)

$$\mathbf{S=L-1-r} \tag{8}$$

Reversing the intensity levels of an image in this manner produces the equivalent of a photographic negative. This type of processing is particularly suited for enhancing white or gray detail embedded in dark region of an image especially when the dark areas are dominant in size [3]. The Fig. 2 shows the negative image of an original image. The image negative transformation response also plotted using mat lab [Fig. 3.].

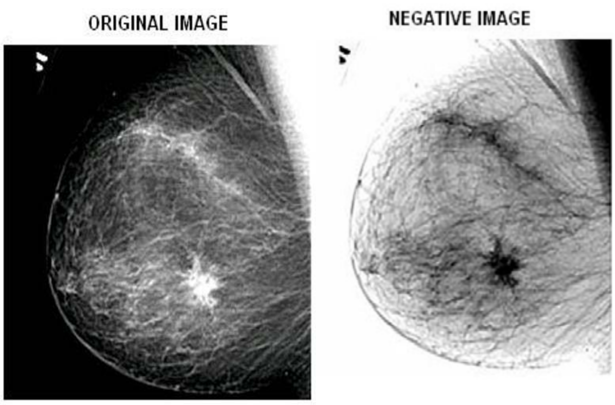


Fig. 2. Shows the negative of an original image

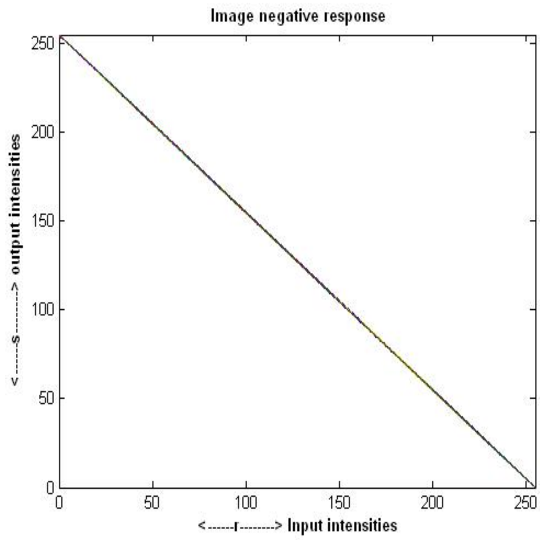


Fig. 3. Shows the image negative transformation response

B. Log Transformation

The general form of log transformation is in equation (9)

$$\mathbf{S=c\ log\ (1+r)} \tag{9}$$

where, c is a constant, and it is assumed that r ≥ 0**.** Log transformation maps a narrow range of low gray level values in the input image into a wider range of output levels. The opposite is true of higher values of input levels. We would use a transformation of this type to expand the values of dark pixel in an image while compressing the higher level values [8]. The opposite is true of the inverse log transformation. The Fig. 4 shows result of log transformation applied to original image .The log transformation response of original image is plotted using mat lab [Fig. 5.].

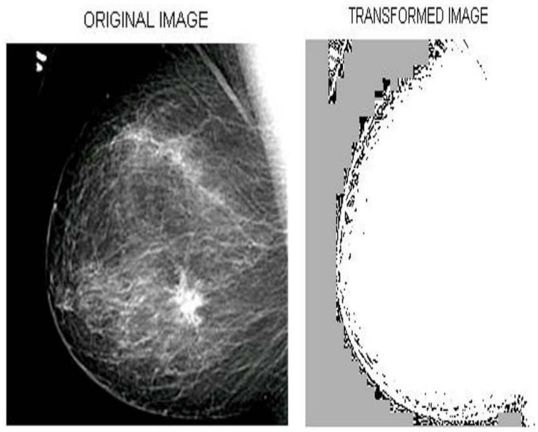


Fig. 4. result of log transformation applied to original image

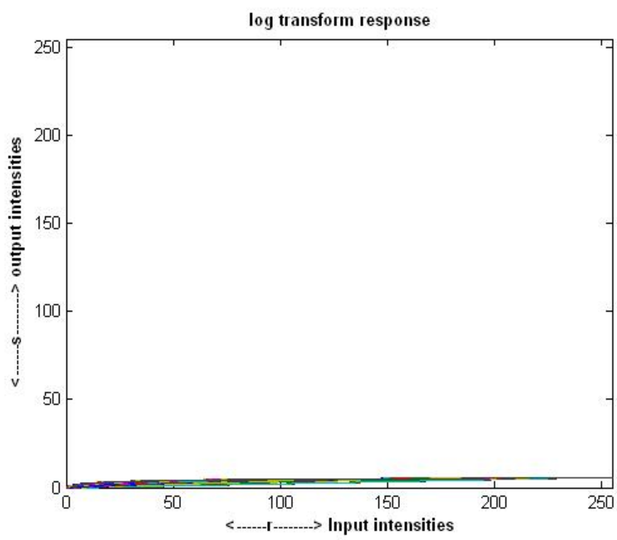


Fig. 5. Shows log transformation response of the original image

C. Power Law Transformation

Power law transformation has a basic form as in equation (10)

$$s = cr^{\gamma} \tag{10}$$

where, 'c' and 'γ' are positive constant. Sometimes it can be written as $s = c(r+\varepsilon)^{\gamma}$ to account for an offset. However, offsets typically are an issue of display calibration and as a result they are normally ignored. Power law curves with fractional values of γ map a narrow range of dark input values into wider range of output values, with opposite is true for higher input values [4]. The figure 6 shows the power law transformation of RGB image. The power law transformation response is also plotted using mat lab [Fig. 7.]

Fig. 6. Displays the power law transformation of a RGB image

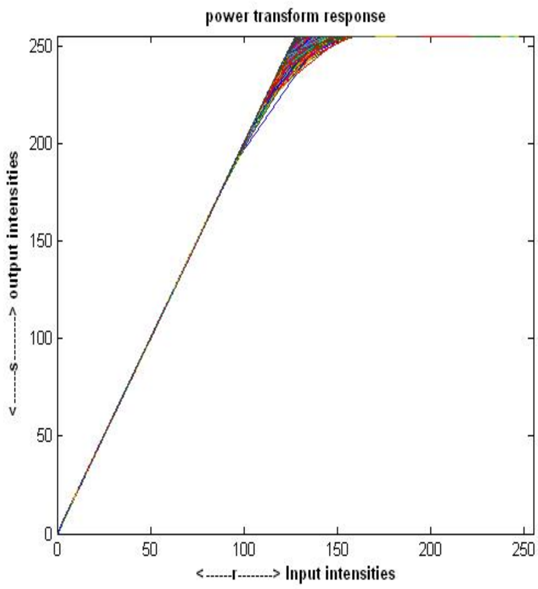


Fig. 7. plot of $s = cr^{\gamma}$ for values of c=2 and γ =0.

D. Thresholding

General form of thresholding function is in equation (11)

$$S=T(r) \tag{11}$$

where s & r are the input and output images respectively. The effect of this transformation would be to produce an image of higher contrast than the original by darkening the levels below and brightening the levels above around particular value which is known as threshold value [4]. The figure 8 shows the thresholding image of original image. Gray level transformation function for thresholding is plotted using mat lab [Fig. 9.]

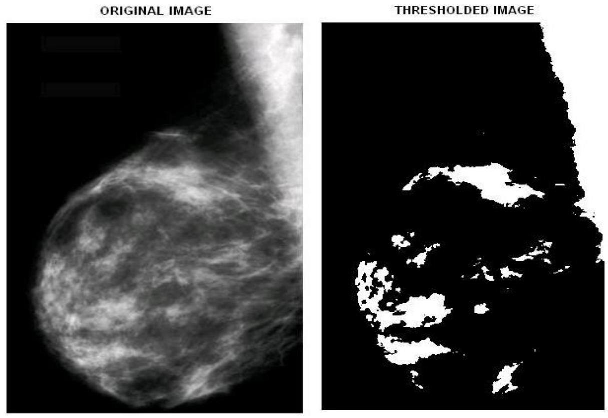


Fig. 8. Shows the thresholding image of the original image

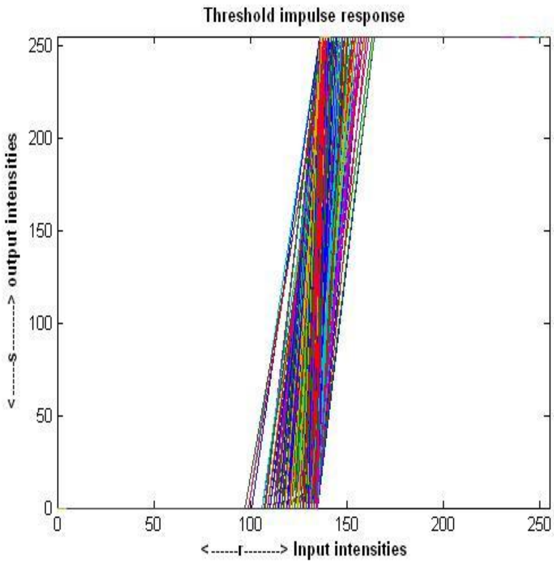
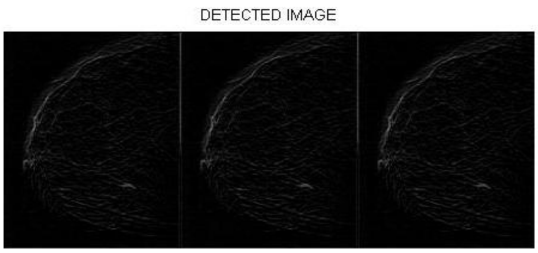


Fig. 9. Gray level transformation function for thresholding

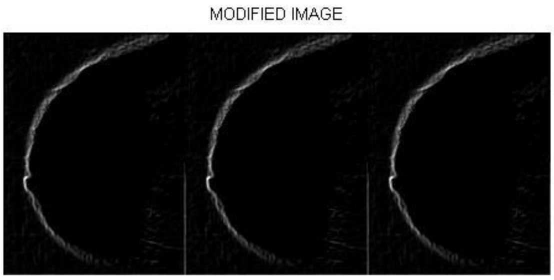
E. Detection Techniques

We have applied different edge detection techniques based on Robert edge detection, prewitt edge detection, sobel edge detection on the raw mammogram image [Fig. 10(a)]. From the processed image the medical practitioner can easily compare different edge detected mammogram images [Fig. 10 (b), (c), (d)] [10].

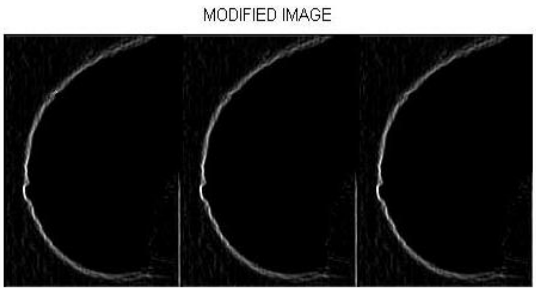
(a)



(b)



(c)

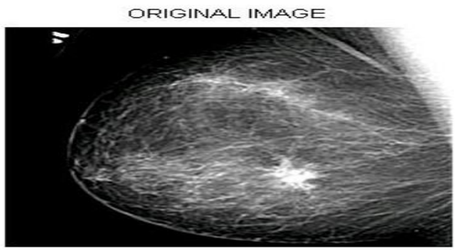


(d)

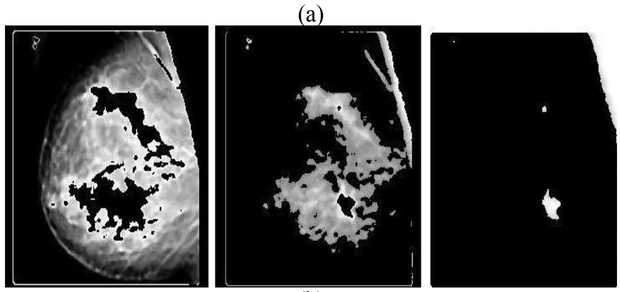
Fig. 10. The comparison of the edge detections for the example image. (a) The original image (b) using Robert edge detection (c) using Prewitt edge detection (d) using Sobel edge detection

In this analysis the first procedure is determining the seed regions. When dealing with mammograms, it is known that the pixels of tumor regions tend to have maximum allowable digital value. Based on this information, image processing functions such as transformation and segmentation are used to detect the possible clusters which contains masses and calcifications [8]. The image features are then extracted to remove those clusters that belong to background or normal tissue as a first cut. The K-means clustering algorithm and fuzzy C-means algorithm is applied as segmentation strategy to function as a better classifier and aims to class data in to separate groups according to their characteristics [7]. The clustering method for both K-means algorithm and fuzzy C-means algorithm is same, but in K-means algorithm when it clusters, it takes the mean of the weighted clusters so as easy to identify masses or the origin point of cancer/ tumor. Similarly in FCM, it considers that each point has weighted value associated with cluster. This is directly implies that, the cluster that have a small value, that will also taken in to calculation. Doing this we were able to find out how much the cancer has spread out [Fig. 11.]. This helped us to find out the stages of breast cancer [10].

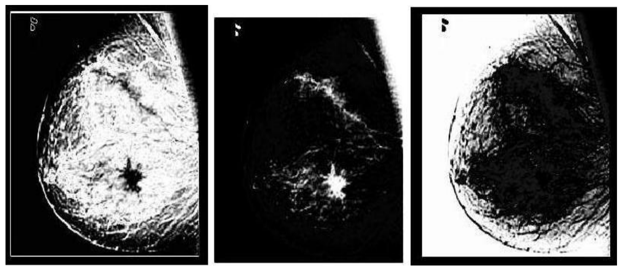
K-means and FCM are helpful in early stage of clustering in medical diagnosis [7]. The cancerous mode can easily be separated from a fatty breast region as well as from dense region. As the number of cluster increases more and more information is obtained about the tissue which can't be identified by the pathologist [11].



(a)



(b)



(c)

Fig. 11. Results from three clusters with k-means & FCM. (a) The original image (b) results obtained from three clusters with k-means algorithm. (c) The results obtained from three clusters with FCM.

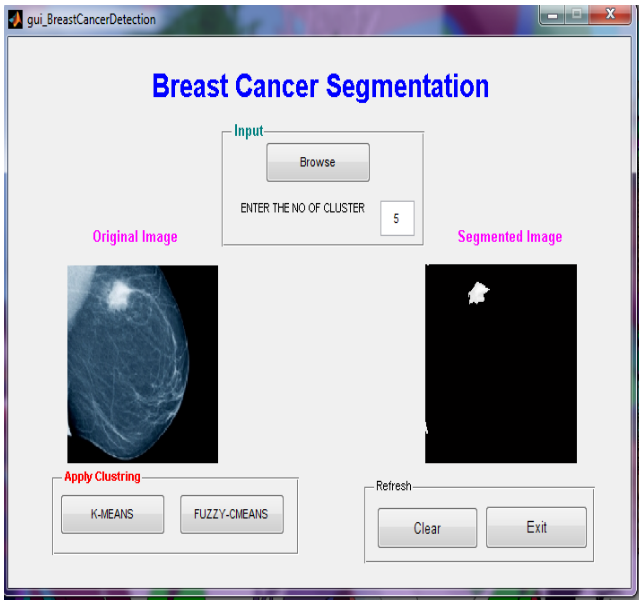


Fig. 12. Shows GUI based Breast Cancer Detection using K-means with cluster size 5.

V. Conclusion

Breast cancer is one of the major causes of death among women. So early diagnosis through regular screening and timely treatment has been shown to prevent cancer. In this paper we have presented a novel approach to identify the presence of breast cancer mass and calcification in mammograms using image processing functions, K-means and Fuzzy C-Means clustering for clear identification of clusters. Combining these we have successfully detected the breast cancer area in raw mammograms images. The results indicate that this system can facilitate the radiologist to detect the breast cancer in the early stage of diagnosis as well as classify the total cancer affected area. This will help doctor to take or analyze in which stage of cancer the patient have and according to which he/she can take necessary and appropriate treatment steps. This proposed method is low cost as it can be implemented in general computer .This paper is based on visual detection method of the processed mammogram images. A real-time system can be implemented using suitable data acquisition software and hardware interface with digital mammography systems.

ACKNOWLEDGMENT

The authors thank to Dr. R. N. Pal, retired Professor IIT Kharagpur, India and his team for their help in conducting experiments with different raw mammogram images. The authors also thank to Dr. P. C. Mahapatra, former vice-president FOGSI (2004), associate Professor, Dept. of Obstetrics and Gynecology, SCB Medical College, Cuttack, India for their assistance with digital mammography systems.

REFERENCES

[1] Mohiy Hadhou, Mohamed Amin, and Walid Dabbour, “Detection of Breast Cancer Tumor Algorithm using Mathematical Morphology and Wavelet Analysis”, in *Proc. GVIP conference, CICC, Cairo, Egypt*,2005, pp. 208-213.

[2] Kanaga, K.C., Anandan, S., Chin, M.Y., and Laila, S.E., “Characterization of micro-calcifications and mass on female breast using processing in full field digital mammography (FFDM)”, in *Proc. Simposium Sains Kesihatan Kebangsaan ke 7, hotel legend, Kuala Lumpur*,2008, pp. 83-87.

[3] Bhattacharya, M. and Das, A., Indian Institute of Information Technology & Management, Gwalior, “Fuzzy Logic Based Segmentation of Microcalcification in Breast Using Digital Mammograms Considering Multiresolution”, in *Proc. IEEE International conference on Machine Vision and Image Processing (IMVIP)*,2007, pp.98 – 105.

[4] M.J. Bottema, G.N.Lee, and S.Lu, “Automatic image feature extraction for diagnosis and prognosis of breast cancer”, in *Proc.*International conference on Artificial intelligence techniques in breast cancer diagnosis and prognosis, Series in machine perception and artificial intelligence, vol.39,2000, pp.17-54.

[5] T. C. Wang and N. B. Karayiannis, "Detection of micro-calcifications in digital mammograms using wavelets", IEEE Trans. on Medical imaging, vol. 17, no. 4, pp. 498-509,August 1998.

[6] Aijuan Dong and Baoying Wang, “Feature selection and analysis on mammogram classification”, Communications, in *Proc. IEEE Pacific Rim Conference on Computers and Signal Processing*,2008, pp.23-26.

[7] K. Thangavel and A.K.Mohideen, Department of Computer Science, Periyar University, Salem, India, “Semi-supervised k-means clustering for outlier detection in mammogram classification”, in *Proc. IEEE international conference on Trendz in Information Sciences & Computing (TISC)*, 2010,pp. 68 – 72.

[8] T.C. Cahoon, M.A. Sutton, and J.C. Bezdek, Dept. of Computer Science, Univ. of West Florida, Pensacola, FL, “Breast cancer detection using image processing techniques”, in *Proc.IEEE Ninth International Conference on Fuzzy Systems (IEEE FUZZ)*, vol.2,2000, pp. 973 - 976.

[9] Jasmine, J.S.L. Govardhan, and A. Baskaran, S. Dept. of CSE, Velammal Eng. Coll., Chennai, India, “Microcalcification detection in digital mammograms based on wavelet analysis and neural networks”, in *Proc.IEEE International Conference on Control, Automation, Communication and Energy Conservation (INCACEC)*,2009, pp.1 – 6.

[10] Yu Guang Zhang, Wen Lu, Fu Yun Cheng, and Li Song, Taishan Med. Univ., Taian, China, “A Edge Detection Method for Microcalfication Clusters in Mammograms”, in *Proc. IEEE 2nd International Conference on Biomedical Engineering and Informatics (BMEI)*,2009, pp. 1 – 5.

[11] M. Sameti, R.K. Ward, J. Morgan-Parkes, B. Palcic, Braintech Inc., Vancouver, BC, “Image feature extraction in the last screening mammograms prior to detection of breast cancer”, *IEEE Journal of Signal Processing*, vol. 3, no.1, pp.46 – 52, Feb. 2009.

[12] M. Bajger, Ma. Fei, S. Williams, and M. Bottema, Sch. of Comput. Sci., Eng. & Math., Flinders Univ., Bedford Park, SA, Australia, “Mammographic Mass Detection with Statistical Region Merging”, in *Proc. IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA)*,2010, pp. 27 – 32.

Nalini Singh received the M.Tech. degree in Communication System Engineering from Kalinga Institute of Industrial Technology, KIIT University, Bhubaneswar, India, in 2008 and the Bachelor in Electronics and Instrumentation Engineering (E & I) from Ghanashyam Hemalata Institute of Technology & Management (GHITM), Puri, Inida, in 2004.

She is currently working as an Asst. Professor in Silicon Institute of Technology, Bhubaneswar, India. She has a number of national and international papers in the field of Biomedical Engineering, where her interests are sensors and transducers, Biomedical signal processing and Digital Image Processing.

Ambarish G. Mohapatra received the M.Tech. degree in Sensor System Technology from Vellore Institute of Technology, VIT University, Vellore, India, in 2008 and the Bachelor in Electronics and Communication Engineering (ECE) from National Institute of Science and Technology (NIST), Berhampur, Inida, in 2004.

He is currently pursuing the Doctor of Science in Technology degree. From 2008 to 2010, he was a Asst. Professor in Silicon Institute of Technology, Bhubaneswar, India. He has number of national and international papers in the field of sensing and control, where his interests are sensors and transducers, MEMS wireless sensor network and Biomedical signal processing.

Guru Kalyan Kanungo pursuing M.Tech degree in Electronics and Telecommunication Engineering from Silicon Institute of Technology, Bhubaneswar, India. Currently he is having many international conference papers and his research interests are biomedical image processing, digital systems and signals.