International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 11 (2015) pp. 29825-29834

http://www.ripublication.com

Change Detection In Mammogram Images Using Fuzzy C- Means Clustering

[1]Nazneen P Kotwal, [2]M. Kirubaa, [3]R. Lavanya,[4]Devi Vijayan
1,2,3,4. Department of ECE, Biomedical Systems and Instrumentation, Amrita Vishwa Vidyapeetham University, Coimbatore 641 112,India.
Email:[1]nazneenpkotwal@gmail.com,[2]m.kirubaa@gmail.com,
[3]r_lavanya@cb.amrita.edu,[4]v_devi@cb.amrita.edu

Abstract

Experts have estimated that breast cancer is diagnosed in about one out of every eight women. At present mammography is the most efficient tool for the screening of breast cancer and studies show that misinterpretation is an important cause of missing breast cancer. In this paper we propose a computer aided detection system to identify changes in temporal mammographic images which would aid radiologists in the early and accurate detection of mammographic lesions. This system involves pre-processing, registration, generation of difference image and the analysis of difference image to obtain the changed and unchanged regions of the lesion. The novelty of this research work is to effectively find changes in mammogram images obtained from consecutive screening rounds using fuzzy c-means (FCM) clustering. The efficiency of FCM is compared with K-means clustering using overall error (OE) and kappa coefficient (KC). Experimental results show that the proposed method is a better alternative to the K-means clustering method. These techniques have been tested on mammogram images obtained from a private hospital.

Keywords: Temporal, Mammograms, Change Detection, Fuzzy c-Means, k-Means, Difference Image.

1.Introduction

Breast cancer is the most general cause of cancer in women all over the world and premature detection and treatment of breast cancer is the most effective method of reducing morbidity and mortality. At present mammography is the most efficient tool for the screening of breast cancer. Studies have shown that radiologists can fail to recognize a substantial proportion of abnormalities and hence a large number of computer aided detection (CAD) techniques are required to improve the accuracy of interpretation [1]. Temporal mammograms refer to mammograms of a single patient from two consecutive screening rounds. The most recent mammogram is called the current mammogram and the mammogram from an earlier screening round is called the prior mammogram. The use of temporal information can help improve the detection performance of a CAD system by bringing to attention subtle signs of malignancy which can be overlooked if the previous mammogram is not available for comparison. Computerized image change detection is especially useful in cases where the changes in mammogram images are difficult to notice by radiologists. In the technique proposed, automatic detection of lesions is carried out using adaptive thresholding and the regions where change has occurred is determined by means of fuzzy c-means clustering.

2. Literature Survey

A number of studies have been carried out in the recent years to evaluate and develop computer aided detection and diagnosis programs in order to assist radiologists in the detection of breast cancer. Warren Burhenne et al. [2] conducted a survey with 1000 screening mammograms and studied the benefit of CAD system in detection of initially missed cancer. Karssemeijer et al. [3] estimated the potential contribution of computer aided detection systems using screening mammograms of 500 cases. Helvie et al. [4] performed a study with 13 radiologists to evaluate the additional effect of using CAD on a dataset consisting of mammograms from 2,389 patients. These studies indicate a potential for CAD to help the breast radiology community while detecting breast cancers.

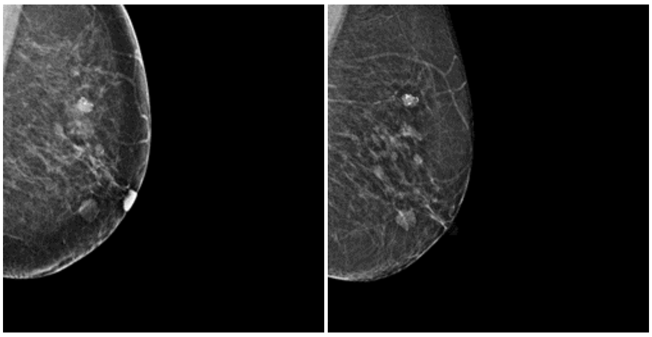
CAD systems that use a single image to characterize mass lesions have been considered by Chan et al. [5] and Huo et al. [6].Studies have shown that the detection of mass lesions by radiologists improve when comparing the current view with prior views [7], [8]. The purpose of the proposed paper is the analysis of two co-registered mammographic images that are taken at two different times. The analysis aims to discriminate the two classes - changed and unchanged regions in the images [9]. In this paper, the procedure for change detection in mammographic images involve three main steps namely 1) Image pre-processing on both the current and prior mammogram images. 2) Generating the difference image (DI) from temporal mammogram images using differencing technique. 3) Analyzing the difference image using fuzzy c-means clustering. The proposed CAD system processes mammograms that are obtained from consecutive screening rounds.

The paper is organized in the following manner. We first describe the initial CAD program and the segmentation process in section 3 and 4 respectively. Then we explain the generation and analysis of difference image in section 5. Experimental results and conclusion are provided in section 6 and 7 respectively.

3. Pre Processing

Image pre-processing is necessary in order to find the orientation of the mammogram, for the purpose of noise removal and to enhance the quality of the images. Fig. 1

shows a pair of temporal mammograms with the corresponding current and prior views. Prior to the application of any mammogram image analysis algorithm, a number of pre-processing steps are crucial to limit the search for abnormalities without excessive impact from background of the mammogram images.



(a) (b)

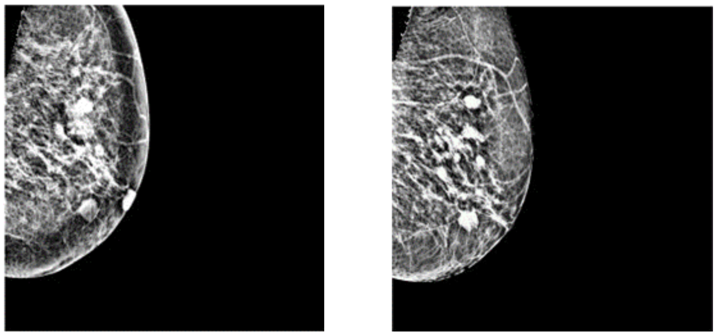
Figure 1: Pairs of Temporal Images. Left And Right Images Correspond To Current and Prior Views

Pre-processing is applied to all the prior and current mammogram images. We start with global registration of the current and prior mammograms by applying the centre of mass alignment technique [10]. This registration process is used to correct the horizontal and vertical translations present between the current and prior images that may have occurred during the acquisition process. We next enhance each image separately. Further pre-processing steps involve the following: a) removal of noise b) suppression of radiopaque artifact c) removal of pectoral muscle.

The process of the image formation, transmission, receiving and processing involves various types of unavoidable disturbances. The kinds of noise that corrupt the mammogram images are quantum noise, additive gaussian noise and impulse noise. Quantum noise occurs in the mammogram images during acquisition due to low-count X-ray photons. In order to reduce the effect of such noise contrast enhancement is applied on the image. Impulse noise forms the main cause of error as this noise is present owing to bit errors during transmission.

The two kinds of impulse noise are random valued noise and salt and pepper noise. In mammograms salt and pepper noise largely corrupts the images where the degraded pixels take either the maximum or minimum gray level value which then leads to severedegradation of image quality and the loss of fine details. The objective of noise suppression in such corrupted images is to filter the specks of salt and pepper so that the noise free image is fully resorted with minimum signal distortion. In the proposed paper the salt and pepper noise is eliminated using median filtering [11]. Median filtering is particularly suitable since it is able to smooth the non-repulsive noise from images without blurring of the edges and thus preserves the details in the image.

In this paper we further use histogram equalization and morphological operations to remove radiopaque artifacts, eliminate digitization noise and to make the hidden features of the image more visible. Seeded region growing (SRG) [12] based segmentation is used to remove the pectoral muscle in order to highlight the region of breast profile for use in computer aided detection algorithm. Fig. 2 shows the result of the pre-processing steps followed by SRG algorithm performed on the current and prior images.

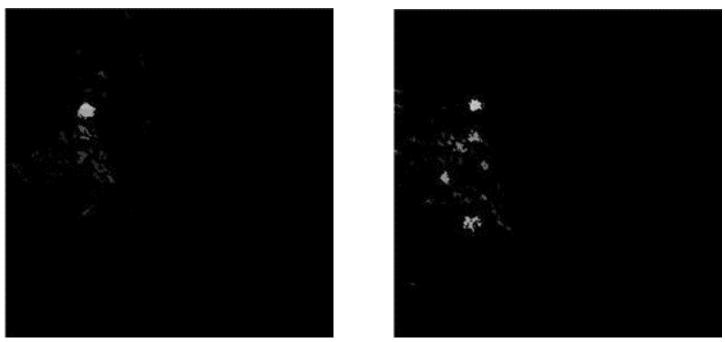


(a) (b)

Figure 2: Pectoral muscle segmentation using SRG. (a) Current image after SRG. (b) Prior image after SRG.

4. Segmentation

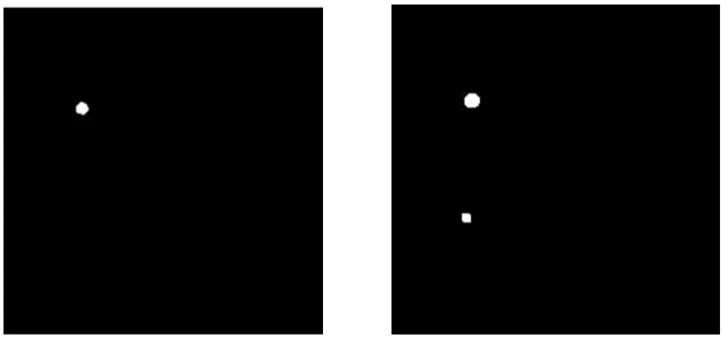
Detection of mass lesion is performed on multiresolution images using the technique of adaptive thresholding [13]. In this technique we combine the coarse segmentation thresholding process and the fine segmentation thresholding process. Coarse segmentation technique is used to obtain a rough local position of suspicious lesions present in the mammogram images. Fig. 3 shows the coarse segmented images for the current and prior images. In this paper we use tophat filtering technique to perform coarse segmentation. Tophat filtering is performed twice on the same image using a disc shaped structuring elements having a different radius each time. The radius of the structuring element is decide



(a) (b)

Figure 3: Images after coarse segmentation. (a) Segmented current image. (b) Segmented prior image

taking into consideration the size of the lesions present. The coarse segmented image is obtained by differencing the resultant images. After performing the coarse segmentation, we implement a window-based adaptive thresholding to perform fine segmentation. In fine segmentation thresholding technique the threshold is calculated for each pixel separately and hence it is a form of local segmentation. Using fine segmentation we obtain a more precise segmentation result of suspicious lesions. Fig. 4 shows the fine segmented results for the current and prior images.



(a) (b)

Figure 4: Images after fine segmentation. (a) Segmented current image. (b) Segmented prior image.

5. Generation of Difference Image

The process of change detection includes the generation of the DI and the analysis of it. The DI is generated by conventional simple differencing. The next step is analysis of DI. The purpose of the DI analysis is to distinguish the changed class from the unchanged one, and the process actually belongs to the domain of image segmentation. Fuzzy c-means clustering is a well-known and popular technique for performing image segmentation.

6. Fuzzy C-Means Clustering

The Fuzzy C-means algorithm is widely used in pattern recognition, clustering and image analysis. It is a process of clustering which allows data elements to group into two or more clusters and assigns membership levels associated with each element. The main aim of FCM is to cluster the data sets such that the similarity of data elements is maximized within each cluster and is minimized in different clusters. Its main objective is to minimize the following objective function ‘J’ in equation (1).

(1)

$$J_m = \sum_{i=1}^{k} \sum_{j=1}^{l} u_{ij} \left\| x_i - c_j \right\|^2 \quad (1)$$

where, m is the fuzziness exponent and u_{ij} represents the degree of membership of the d- dimensional measured data x_i in the j^{th}cluster. $c_j$is the cluster's dimension center and ‖*‖ is a norm that expresses the similarity between the center and a data value.

In Fuzzy partitioning membership u_{ij} by (2) and the center of the clusters $c_j$by (3) are updated after every iteration.

$$u_{ij} = \frac{1}{\sum_{k=1}^{l}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{(m-l)}}} \tag{2}$$

$$c_j = \frac{\sum_{i=1}^{k} u_{ij}\, x_i}{\sum_{i=1}^{k} u_{ij}} \tag{3}$$

The iteration will stop when,

$$max_{ij} = \{|u_{ij}^{k+1} - u_{ij}^{k}|\} < \varepsilon \tag{4}$$

Algorithm

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

$$u_{ij} = \frac{1}{\sum_{k=1}^{l}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{(m-l)}}} \tag{5}$$

4. Stop if, $\|u_{ij}(k+1) - u_{ij}(k)\| < \varepsilon$ otherwise go back to step 2.

7. K-Means Clustering

K-means clustering is a simple partitioning method which partitions 'n' observations/data into 'k' mutually exclusive clusters in which it assigns each observation to a particular cluster with the nearest centroid. For a particular observation set denoted by ($x_1, x_2, \ldots, x_n$), where each of the given observation is a d-dimensional real vector, the K--means clustering algorithm partitions 'n' given observations in to 'k' sets where the value of k is less than the value of n, {S = S_1, S2,…,S_n} to minimize the equation in (6)

$$\arg min_S \sum_{i=1}^{k} \sum_{x \in S} \|x - u_i\|^2 \quad (6)$$

Where, u_i is the mean value of S_i. The number of cluster, k is assumed to be fixed in K-means clustering.

Algorithm

Given the cluster number k, the algorithm yields by alternating between the assignment step and the update step.

1. Step for Assignment: Each data point is assigned to the cluster with the nearest centroid using equation (7).

$$s_i = \{x_i : \|x_j - m_i^t\| \leq \|x_j - m_i^t\| \forall_i = 1 \ldots k \quad (7)$$

2. II. Update Step: Calculating the new centroid of the cluster using equation (8)

$$m_i^{t+1} = \left(\frac{1}{|s_i|}\right) \sum_{x_j \in s_i} x_j \quad (8)$$

1. Select k points (k=2 in this case) randomly from the data set as the points denoting initial centroids.
2. Obtain k clusters by associating every point with the closest centroid.
3. Now compute the centroid of each cluster.
4. Repeat steps (2) and (3) until it converges.

8. Experimentation Result

The results of the experiments are shown by providing some values of criteria to evaluate the algorithm [9]. The total number of the pixels N present in the ground truth provided by the radiologist is first calculated. Then, using the ground truth, the number of pixels that actually belong to the class – unchanged, Nu, and the class – changed, Nc is determined. The ground truth is also compared with the image generated from the above approach pixel by pixel. From the above comparison the number of False Positive (FP) pixels(pixels that belong to the unchanged class but are incorrectly classified as those belonging to the changed class) and the number of False Negative (FN) pixels (pixels that belong to the changed class but are incorrectly classified as those belonging to the unchanged class) are determined. Another two values that represent the number of correctly detected unchanged and changed pixels are written as True Negative, TN and True Positive TP, respectively and are defined as (9).

$$TP = Nc - FN$$
$$TN = Nu - FP \tag{9}$$

In order to evaluate the results obtained further, the overall error (OE) is determined. The OE is defined as (10).

$$OE = FP + FN \tag{10}$$

Table 1: Evaluation Criteria Using Fuzzy C Means

Criteria	Value
TP	316
TN	65214
OE	6
KC	0.9905

Table 2: Evaluation Criteria Using K Means

Criteria	Value
TP	108
TN	65410
OE	18
KC	0.9229

Moreover, the Kappa coefficient (KC) is used as an overall evaluation criterion. KC is calculated as (11):

$$KC = \frac{PCC - PRE}{1 - PRE} \tag{11}$$

Where

$$PRE = \frac{(TP + FP) \cdot Nc + (FN + TN) \cdot Nu}{N^2} \tag{12}$$

$$PCC = \frac{TP + TN}{N} \tag{13}$$

The usual range for the value of KC is found to be 0 to 1. Here PRE is a measure of chance agreement and PCC is the percentage correct classification measure. From the above equation we note that KC is influenced by the dependent values of TP and TN, whereas PCC is influenced by the summation of TP and TN values. The main criteria values of evaluation on application of fuzzy c means are organized in tabular form in Table 1. The analysis of the difference image is repeated using k-means clustering for the purpose of comparison and the result is organized in Table 2.

9. Conclusion

At present mammography is the most efficient tool for the screening of breast cancer and studies show that misinterpretation is an important cause of missing breast cancer. The contribution of our work is to effectively find changes in mammogram images obtained from consecutive screening rounds which can prompt the radiologist for further diagnosis. Experimental results show that fuzzy c-means is a better technique in successfully classifying the difference image into changed and unchanged class in comparison with k-means technique. The overall error in identification of changed pixels is lower in case of fuzzy c-means clustering compared to k-means clustering and the value of KC is greater in case of FCM indicating better segmentation. As an extension of this work we will be developing an effective automated system capable of classifying the tumor into malignant and benign which would help the radiologist in early detection of breast cancer.

References

[1] Sheila Timp, Ceila Varela, and Nico Karssemeijer, Computer-Aided Diagnosis With Temporal Analysis to Improve Radiologists' Interpretation of Mammographic Mass Lesions, IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 3, pp.803-808, May 2010.

[2] Warren Burhenne, L.J., S.A. Wood, C.J. D'Oris, S.a. Feig, D.B. Kopans, K.F. O'Shaughnessy, E.A. Sickles, L. Tabar, C.J. Vyborny, and R.A. Castellino, Potential contribution of computer- aided detection to the sensitivity of screening mammography, Radiography, pp. 554-562, 2000.

[3] Karssemeijer, J.D. Otten, A.L. Verbeek, J.H. Groenewoud, H.J. De Koning, J.H. Hendriks, and R. Holland, Computer-aided detection versus independent double reading of masses on mammograms, Radiology, pp.192-200, 2003.

[4] Helvie, M.A., L. Hadjiiski, E. Makariou, H.P. Chan, N. Petrick, B. Sahiner, S.C. Lo, M. Freedman, D, Adler, J. Bailey, C.Blane, D. Hoff, K. Hunt, L. Joynt, K. Klein, C. Paramagul, S.K. Patterson, and M.A. Roubidoux, Sensitivity of non-commercial computer-aided detection system for mammographic breast cancer detection: pilot clinical trial, Radiology, pp. 208-214.

[5] H.P.Chan, B.Sahinder, M.Helvie, N.Petrick, M.Roubidoux, T. Wilson, D.Adler, C.Paramagul, J.Newman, and S.Sanjay-Gopal, Improvement of radiologists' characterisation of mammographic masses by using

computer-aided diagnosis: An ROC study. Radiology, vol. 212, no. 3, pp. 817-827, 1999.

[6] Z.Huo, M.Giger, C.Vyborny, and C.Metz, Breast cancer: Effectiveness of computer-aided diagnosis observer study with independence database of mammograms. Radiology, vol. 224, no. 2, pp. 560-568, 2002.

[7] C.Varela, N.Karssemeijer, J.Hendriks, and R.Holland, Use of prior mammograms in the classification of benign and malignant masses, Eur.J.Radiol., vol. 56, no. 2, pp. 248-255, 2005.

[8] L.Hadjiiski, B.Sahiner, H.-P.Chan, N.Petrick, M.Helvie, and M.Gurean, Analysis of temporal changes of mammographic features: Computer-aided classification of malignant and benign masses. Med. Phys., vol. 28, no. 11, pp. 2309-2317, 2001.

[9] Maoguo Gong, Linzhi Su, Meng Jia, and Weisheng Chen, Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar image, IEEE Trans. Fuzzy Syst., vol. 22, no. 1, pp. 98-109, Feb. 2014.

[10] Shelia Timp, Ceila Varela, and Nico Karessemeijer, Temporal change analysis for characterisation of mass lesion in mammography, IEEE Trans. Med. Imag., vol. 26, no. 7, pp. 945-953, Jul. 2007.

[11] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, and Syed Khaleel Ahmed, Automated breast profile segmentation for ROI detection using digital mammograms, 2010 IEEE EMBS conference on Biomedical Engineering & Sciences(IECBES 2010), Kuala Lumpur, Malaysia, pp. 87-92, Dec. 2010.

[12] Rolf Adams, and Leanne Bischof, Seeded region growing, IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 6, pp. 641-647, Jun. 1994.

[13] Kai Hu, Xieping Gao, and Fei Li, “Detection of Suspicious Lesion by Adaptive Thresholding Based on Multiresolution Analysis in Mammograms,” IEEE Trans. Instrum. Meas., vol. 60, no. 2, pp. 462-472, Feb. 2011.