2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia

ROI Segmentation using Local Binary Image

Shubhi Sharma
Computer Science & Engineering
PDPM IIITDM Jabalpur, India
shubhi.sharma@iiitdmj.ac.in

Pritee Khanna
Computer Science & Engineering
PDPM IIITDM Jabalpur, India
pkhanna@iiitdmj.ac.in

***Abstract*—Segmentation of ROI is an important and challenging task in the development of CAD system for the detection of breast cancer. This work proposes a Local Binary Image (LBI) to segment the ROI from the mammogram patches. The key idea is to use textural properties of mammogram patches for representing salient micro-patterns of the masses and preserving the spatial information at the same time. Corresponding to the patch, LBI is the binary image where the value 1 represents the presence of texture in the patch. Using LBI the threshold value is identified which is used to extract the mask image. Once the mask image is generated boundary is plotted to trace suspicious area in the patch. The efficiency of the proposed method is tested on a dataset of 819 suspicious patches from the IRMA reference database. The experimental results achieved that the proposed LBI method has successfully attained the value 0.934 for Quality measure.**

***Index Terms*— Histogram equalization, DDSM, Local Binary Image (LBI), Segmentation.**

I. INTRODUCTION

Breast Cancer has become one of the significant and frequent forms of cancer for women all over the world. In India, a death rate of one in eight women has been reported due to breast cancer [12]. Presently, risk factors of breast cancer cannot be avoided and the survival rate of the patient is only related to early detection. Mammography is one of the most reliable techniques for detecting early stage breast cancer. Image processing can prove the odds of mammograms in detecting breast cancer early. Mammography test has shown highest accuracy among all the existing techniques, but like most medical tests, it's not perfect. On an average, mammography detection rate of the breast cancer is 80-90% [7]. A mammogram is an X-ray image of the breast tissue which allows better visualization of internal structure of the breast. Fig. 1 shows an image of Digital Mammography.

High resolution mammographic images can be used to detect the signs of breast cancer such as micro-calcifications and masses [5, 6]. Lesions often occur in dense breast tissue areas in a number of different shapes such as circumscribed, speculate, lobulated or ill-defined. Micro-calcifications are deposited tiny calcium which accumulates in breast tissue [18]. It is shown that 90% of impalpable in situ ductal carcinomas and 70% of impalpable minimal carcinomas were visible as micro-calcification alone [8]. Accordingly detecting impalpable malignant calcifications within the breast can improve survival rate of breast cancer patients. Mammography screening method has been proved an effective technique for early breast cancer detection. As most of the countries are emphasizing on screening programs, number of mammograms to be analyzed by the radiologists is enormous. Manual reading of the mammograms is time consuming and labor intensive. Radiologist may miss some minor abnormalities due to large number of normal patients in screening programs. The problems associated with traditional diagnosis method motivated researchers to develop CAD systems to assist radiologists for detecting and diagnosing the breast cancer at very early stage. To develop an efficient CAD system, the most important step is the preprocessing of original image (mammogram) followed by the segmentation of the suspicious region of mammogram i.e. ROI (region of interest). The reduction of false positives/false negatives depends on how accurately the ROI is segmented.

The proposed work focuses on developing a new approach for efficient ROI segmentation. The texture information is, obtained through a Local Binary Image (LBI), which is used to generate the mask image to obtain ROI. The remaining text is organized in four sections; section 2 presents the efforts done in this direction, section 3 explains the proposed methodology, the database used for experimentation is discussed in section 4, and experimental results are discussed in section 5. Finally, section 6 concludes the work along with future direction.

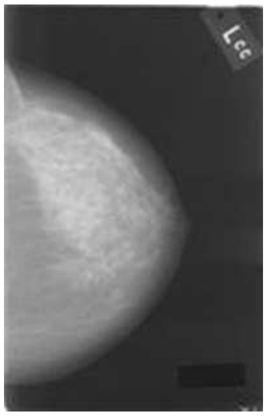

Fig. 1. A mammogram image from Digital Database of Screening Mammography of class-5 and case 1 C_0001_1.LEFT_CC.LJPEG [13, 19]

II. LITERATURE SURVEY

Image segmentation is the process of decomposing the image into its constituents regions or objects. Proper segmentation process isolates the ROI from the original image. Various approaches have been proposed to segment the breast profile region in mammograms. A few of them are presented here.

G. Rabottino, A. Mencattini, M. Salmeri, F. Caselli and R. Lojacono proposed mass contour extraction in mammographic image for breast cancer detection [1]. The proposed algorithm for massive lesions segmentation is based on the region growing technique and is tested on mammographic images taken from Digital Database for Screening Mammography (DDSM). Region growing algorithm is one of the most acceptable techniques used for the segmentation of medical images. But it is not always feasible as it needs extraction of ROI by radiologist which is a time consuming and labor intensive job.

A. Mencattini, G. Rabottino, S. Salicone and M. Salmeri introduced uncertainty propagation for the assessment of tumoral masses segmentation [2]. The tumoral mass segmentation and characterization algorithms are evaluated by implementing the uncertainty propagation through blocks. Monte Carlo method owing to the iterative and very complex structure of the algorithms is used. The basis of segmentation is the region growing segmentation algorithm and the performance is tested on the dataset taken from DDSM.

Marker-Controlled Watershed algorithm is one of the robust and acceptable segmentation techniques in many areas. Xu, Shengzhou, Hong Liu, and Enmin Song applied marker controlled watershed for lesion segmentation in mammograms [3]. The method is based on the traditional watershed transformation to obtain a boundary in the belt between the internal and external markers. The dataset of 363 ROIs from DDSM is used to evaluate the performance of the method. To automatically determine the internal and external markers, the rough region of the lesion is identified by template matching and thresholding.

Oliver, Arnau, Xavier Lladó, Jordi Freixenet, and J oan Martí [4] proposed a method to reduce false positive rate in mammographic mass detection using local binary patterns. They develop a new approach in the field of mammographic mass detection. Their proposal is based on Local Binary Patterns (LBP) for representing salient micro-patterns and preserving the spatial structure of the masses too. Once the descriptors are extracted, support vector machines (SVM) are used for classifying the detected masses. The results are generated on a dataset of 1792 suspicious regions of interest extracted from the DDSM database.

It is well-known that texture is one of the commonly used features for the analysis and interpretation of images. It plays an important role in detecting the disease pattern in breast from the mammogram image [14]. A Local Binary Image (LBI) is proposed to segment the ROI. The experimental results are generated on a dataset of 819 suspicious regions of interest extracted from IRMA reference database belonging to DDSM repository [13].

III. METHODOLOGY

The most challenging step in the development of a Computer Aided Diagnosis (CAD) system for the detection of breast cancer is the lesion segmentation in ROI. Texture analysis plays an important role in segmentation. It also helps in maintaining the spatial information of the mass ROI. Considering these facts a 'Local Binary Image (LBI)' is developed for segmentation of the ROI and generation of mask image.

Fig. 2 shows the flowchart of the proposed method. First, mammogram gray-scale patch image is enhanced to make pixels of mass area more visible then gray scale erosion is applied two times. LBI is a binary image which works in a 3x3 block of pixels. It assigns a value 1 to the pixels where texture information is present otherwise 0 value is assigned. The LBI is used for calculating the threshold parameter to generate a binary mask image for the original gray image. This mask image is used to plot the boundary of the suspicious area in the original patch termed as ROI here. This ROI can be tested further to determine malignancy. The following sub sections explain the steps shown in flowchart.

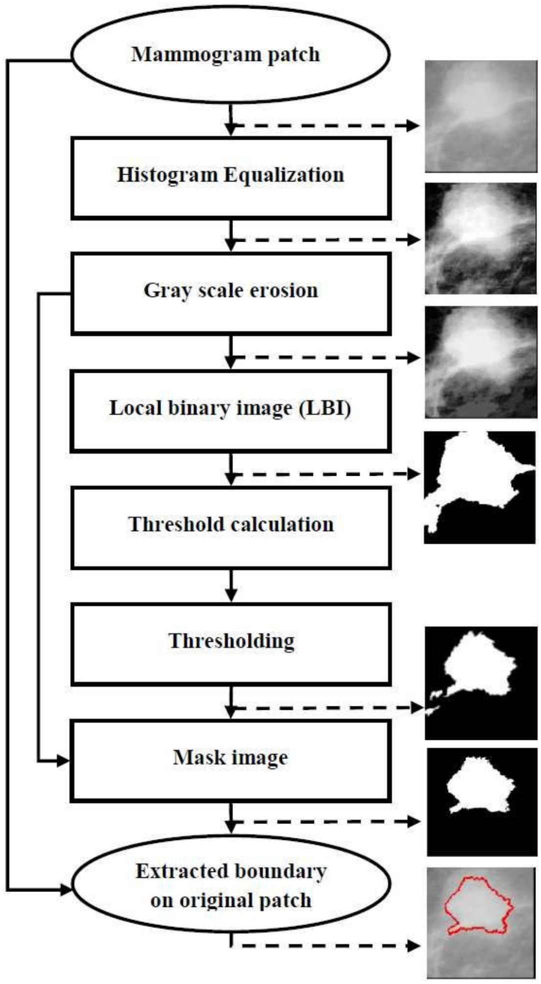

Fig. 2. Flowchart of the proposed method.

A. Histogram Equalization

Image enhancement is used to generate a visually desirable image. In this work, the contrast of the ROI is improved by applying a histogram equalization method which is a widely acceptable technique for medical images enhancement. By distributing the pixel values uniformly, it enhances the contrast of the image. The required transformation function is given in Eq. 1 [15].

$$s = T(r) \quad 0 \le r \le 1. \tag{1}$$

This transformation will normalize an image 'r' to the interval [0, 1]. A value 0 and 1 stands for black and white respectively. The transformation function, $T(r)$, is single valued and monotonically increasing in the interval $0 \le r \le 1$ and satisfies the condition in Eq. 2.

$$0 \le T(r) \le 1 \text{ for } 0 \le r \le 1. \tag{2}$$

The result of histogram equalization applied on original ROI is shown in Fig. 3.

B. Gray Scale Erosion

Erosion is one of the two basic operators in the area of mathematical morphology that is used to erode away the boundaries of the region of foreground pixels [20]. It makes the area of foreground pixels to shrink in size and enlarges the holes within that area. Here, gray scale erosion operation is applied on the enhanced image so that the pixels of the malignant mass become more visible and get separated from the background. To get efficient output of LBI, the gray scale erosion is applied twice. In general, gray scale erosion of image $A(x, y)$ by $B(x, y)$ is defined as given in Eq. 3.

$$(A\Theta B)(x, y) = \min\{A(x + x\text{^}') - B(x\text{^}'y\text{^}') \mid (x\text{^}'y\text{^}') \in D\_B\} \tag{3}$$

where D_B is the domain of structuring element B and $A(x, y)$ is assumed to be $+\infty$ outside the domain of image [11]. In mammograms, lesions are circular in shape, so the structuring element needs to be flat and disk-shaped. This work uses a structuring element with non-zero height and a radius of eleven pixels.

C. Local Binary Image (LBI)

LBI is proposed to identify the textural information present in the mammogram patch. In the proposed method central pixel value in 3x3 block is thresholded on the basis of gray level values in its 8-neighborhood. The central pixel will be assigned 1 if any of the pixel in the 8-neighborhood does not equals the intensity of the central pixel. Otherwise intensity of central pixel is set to 0. The resulting image thus obtained is termed as LBI wherein the patterns of 1's reflect texture information.

The LBI is generated using Eq. 4.

$$LBI(x, y) = \begin{cases} 0, & if\,|I(x, y) - P_i(x, y)| = 0 \forall i,\ i = 0,1,... \\ 1, & otherwise \end{cases} \tag{4}$$

where, I is the double eroded image and P_i is one of the 8-neighbors of pixel (x, y) considered as central pixel in a block. Using these pixel coordinates threshold value is calculated by

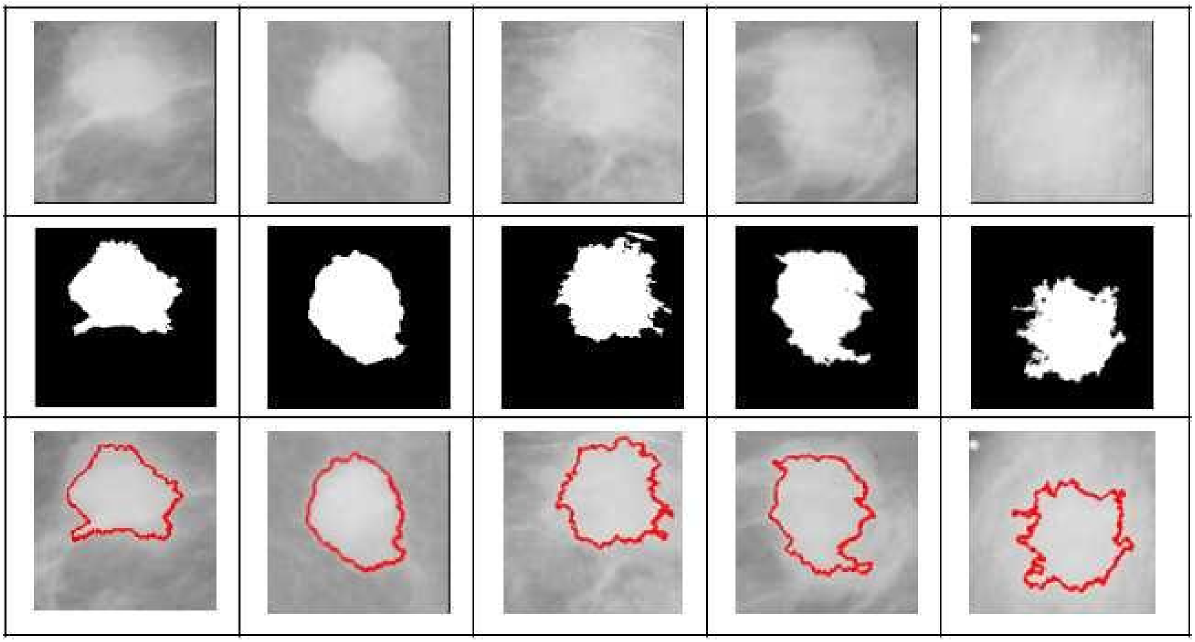

Fig.3. Five examples of segmentation results by our method. First row is the original mammogram patches, second row is the mask images obtained and the last row contains masses boundary extracted on original mammogram patch images.

taking mean of the gray values of the double eroded image for the pixels whose corresponding values in LBI is 1.

Given a $LBI(x, y)$ and a double eroded image $I(x, y)$, the threshold T can be calculated as given in Eq. 5.

$$Z = LBI(x, y) \times I(x, y)$$
$$M = \sum_x \sum_y z(x, y) \qquad (5)$$
$$T = \frac{M}{Total \quad numberof \quad non \quad zero\, pixels \quad in \quad Z}$$

In this way threshold T is calculated automatically for the targeted mammogram patch. To extract the mask image the double eroded image is converted into a binary image using the threshold calculated in the previous step. For the binary image obtained, connected components in the image are identified as regions and the region with maximum area is considered which gives final mask image. The extracted mask image is as shown in the second row of Fig. 3. Exterior boundary of the region is traced using mask image and is plotted on the original image using the same pixel coordinate values. The results obtained are shown in last row of Fig. 3.

IV. DATABASE

IRMA reference database is used for experiment. It contains 9,870 mammographic images from four different repositories [13, 19]. For experimental evaluations 819 mammographic patches belonging to class-5 of DDSM repository are selected randomly. Fig. 4 shows the six mammogram patches from the set of chosen images. Table 1 shows 9,870 mammographic images with relevant Breast Imaging-Reporting and Data System (BI-RADS) codes. It is a quality assurance tool originally designed to use with mammography. The system is a collaborative effort of many health groups but is published and trademarked by the American college of radiology (ACR) [16]. BI-RADS Assessment Categories and class definitions are as follows [17, 19]:

- 0: Incomplete
- 1: Negative
- 2: Benign finding (s)
- 3: Probably benign
- 4: Suspicious abnormality
- 5: Highly suggestive of malignancy
- 6: Known biopsy – proven malignancy.

V. RESULTS AND DISCUSSIONS

To test the performance of the method, all 819 ROIs were segmented by the proposed method. The accuracy of this technique was evaluated through quantitative measures derived through manual comparison of each segmented "mask i mage" with its corresponding original ROI. In the segmentation a positive case means correct detection or classification of breast glandular or dense tissue while a negative case means

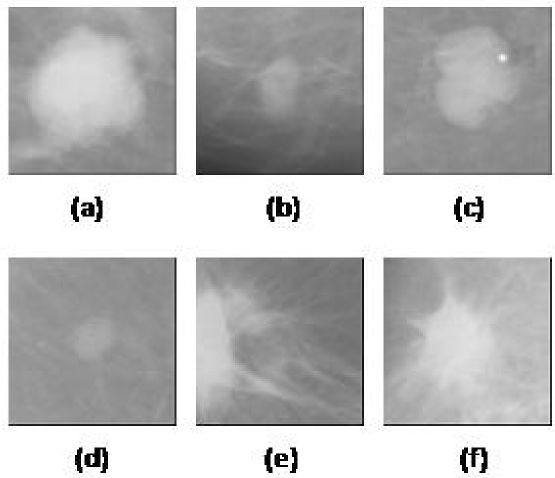

Fig. 4 (a-f) ROIs with malignant mass [19].

misclassification of other tissues as such a type. The quantitative measures are based on the following:

- ***True Positive (TP)*** means segmentation algorithm extracts the region (mask image) correctly.
- ***False Positive (FP)*** means segmentation algorithm has segmented extra pixels other than the suspicious area (mask image and more pixels) i.e. over segmentation.
- ***False Negative (FN)*** segmentation algorithm has segmented few pixels of the suspicious area (not all the pixels of mask image) i.e. under segmentation.

From these measures two popular metrics, Completeness (CM) and Correctness (CR), can be derived. Completeness defines the sensitivity i.e. true positive rate or recall rate and it is measured as shown in Eq.6. On the other hand, Correctness describes the specificity. It measures the proportion of negatives which are correctly identified and is calculated as given in Eq. 7.

$$Completeness = \frac{TP}{(TP + FN)} . \qquad (6)$$

$$Correctness = \frac{TP}{(TP + FP)} . \qquad (7)$$

TABLE. 1. DATA DISTRIBUTION IN THE IRMA MAMMOGRAPHY REFERENCE DATABASE [13].

BI-RADS		Assessment category			Sum
		1	2	5	
TISSUEC LASSES	I	2,518	691	591	3,785
	II	1,855	515	471	2,541
	III	1,295	383	263	1,940
	IV	834	237	233	1,304
SUM		6,501	1,811	1,558	9,870

The optimum value for both the measures, CR and CM, is 1. With this proposed approach the values for CR and CM is determined as 0.961 and 0.970 which are nearer to 1. The approach leads us towards efficient ROI extraction. In general, completeness and correctness are combined together to measure the performance of an algorithm in more efficient way. The combined single measure is termed as *Quality* and is given in equation 8. The optimum value of this measure is also 1 and using the proposed approach the obtained value for this measure is 0.934.

$$Quality = \frac{TP}{(TP + FN + FP)} . \tag{8}$$

The values of various quantitative measures for the dataset consisting of 819 ROIs of class-5 are shown in Table 2.

TABLE. 2. QUANTITATIVE MEASURES ON 819 ROIS OF CLASS-5 FROM IRMA REFERENCE DATABASE.

Mean Result		Values
Intersection	(TP)	765
Over segmentation	(FP)	31
Under Segmentation	(FN)	23
Correctness	(CR)	0.961
Completeness	(CM)	0.970
Quality		0.934

A. Comparative Analysis

G. Rabottino, A. Mencattini, M. Salmeri, F. Caselli and R. Lojacono [1] has also used region growing algorithm. The method chooses the seed manually and then region grows if the values of neighboring pixels are similar to the seed. Finally the region obtained is extracted to be processed further. For the similarity condition they considered two thresholds based on some condition. The results obtained for boundary extraction with this approach and the proposed approach is shown in Table 3. It is clear that the proposed approach is performing better.

TABLE. 3. COMPARISON OF PROPOSED METHOD WITH MASS CONTOUR EXTRACTION METHOD [1].

Method	CM	CR
Mass contour extraction method [1]	0.8834	0.9338
Proposed Approach	0.970	0.961

VI. CONCLUSIONS

The proposed algorithm efficiently extracts ROIs from mammogram patch image. The results obtained over randomly selected patches from the IRMA database are promising. This resultant segmented region can be used further for the automated abnormalities detection in the human breast like calcification, circumscribed masses, speculated masses and other ill-defined masses, circumscribed lesions, asymmetry analysis etc. Future work aims to develop CAD system for breast cancer detection using this algorithm because of the simplicity and promising results shown by the algorithm.

REFERENCES

[1] G. Rabottino, A. Mencattini, M. Salmeri, F. Caselli and R. Lojacono, “Mass Contour Extraction in Mammographic Images for Breast Cancer Identification,”16th IMEKO TC4 Symposium, Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements, Florence, Italy, 2008.

[2] A. Mencattini, G. Rabottino, M. Salmeri and S. Salicone, “Uncertainty propagation for the assessment of tumo ral masses segmentation,” Advanced Methods for Uncertainty E stimation in Measurement, 2009. AMUEM 2009. IEEE International Workshop on, pp. 39-43. IEEE, 2009.

[3] Xu, Shengzhou, H. Liu, and E. Song. “Marker-control led watershed for lesion segmentation in mammograms,” J ournal of digital imaging 24.5, pp. 754-763, 2011.

[4] A. Oliver, X. Llad´o, J. Freixenet, and J. Mart, “F alse positive reduction in mammographic mass detection using local binary patterns,” Medical Image Computing and Computer-Assi sted Intervention–MICCAI 2007.,Springer Berlin Heidelberg, pp. 286-293, 2007.

[5] T. O. Gulsrud and J.H. Husoy, “Optimal filter based detection of microcalcifications,” Biomedical Engineering, IEEE Transactions on 48, no. 11, pp. 1272-1281, 2001.

[6] T. Netsch and H. Peitgen, “Scale space signatures f or the detection of clustered micro-calcifications in digital mammograms,” Medical Imaging, IEEE Transactions on 18, no. 9, pp. 774-786, 1999.

[7] I. K. Maitra, S. Nag, and S. Kumar Bandyopadhyay, “T echnique for preprocessing of digital mammogram”, Computer me thods and programs in biomedicine 107.2, pp. 175-188, 2012.

[8] Feig, B.M. Galkin, “Breast microcalcifications: early warning for cancer?” Diagn. Imag., pp. 132–138, 1990.

[9] Timp, Sheila, and N. Karssemeijer. "A new 2D segmentation method based on dynamic programming applied to computer aided detection in mammography." Medical Physics 31, 5, pp. 958-971, 2004.

[10] M. Pietikäinen , “ Computer vision using local bina ry patterns”. vol. 40. Springer, 2011.

[11] R. M. Haralick,., S. R. Sternberg, and X. Zhuang, “Im age analysis using mathematical morphology,” Pattern An alysis and Machine Intelligence, IEEE Transactions on 4, pp. 532-550, 1987.

[12] American Cancer Society Breast Cancer: facts and figures.

[13] T. Deserno, M. Soiron, J. Oliveira, and A. Araujo. “Towards computer-aided diagnostics of screening mammography using content-based image retrieval,” In Graphics, Patter ns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on, pp. 211-219. IEEE, 2011.

[14] Nanni, Loris, A. Lumini, and S. Brahnam, "Survey on LBP based texture descriptors for image classification," Expert Systems with Applications 39.3, pp. 3634-3641, 2012.

[15] R.C. Gonzalez and R.E. Woods “Digital Image processing ,” Pearson education, Inc, publishing as prentice Hall, 3rd edition, 2009.

[16] American College of Radiology (ACR) Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). Reston, Va : © American College of Radiology; 2003.

[17] http://en.wikipedia.org/wiki/BI-RADS last access on 31/07/2013.

[18] M. Rizzi, M. D'Aloia, and B. Castagnolo, "Computer aided detection of microcalcifications in digital mammograms adopting a wavelet decomposition," J. Integrated Co mputer-Aided Engineering 16.2, pp. 91-103, 2009.

[19] T.M. Deserno web page: http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php

[20] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digit al image processing using MATLAB," Knoxville: Gatesmark Publ ishing, vol. 2, 2009.