

Detection of microcalcification with top-hat transform and the Gibbs random fields

Akshay S. Bharadwaj and Mehmet Celenk
School of Electrical Engineering and Computer Science
Ohio University
Athens, OH 45701 USA
{ab101814, celenk}@ohio.edu

Abstract—Breast cancer is one of the most common causes of death in women aged 40 and above. Early detection of breast cancer has been one of the prime topics of research in biomedical engineering area. Micro-calcifications (MCs) are the indicators of early stages of breast cancer, and the detection of these MCs will, in turn, lead to diagnosis and treatment of breast cancer at its earliest stages. This paper proposes a new method to detect MCs in a digital mammogram. The approach starts with the segmentation of the digital mammogram to isolate the breast region, using fuzzy C means clustering algorithm. The segmented image is then further segmented using top-hat transform to localize the region of interest. A watershed transform is used to isolate the region of interest from rest of the image. The Gibbs random fields are employed to analyze the pixels in conjunction with the devised clique patterns and detect MCs in the image. A thresholding is performed on the processed image where the MCs are detected. The proposed algorithm is highly effective in reducing the region of interest to the region which has a high probability of finding a calcification or MC. It has an overall detection rate of 94.4% and accuracy of 88.2% with a false negative detection rate of 5.6%, respectively.

Keywords— *Microcalcifications, fuzzy C means clustering, top-hat transform, Gibbs random fields, clique patterns*

I. INTRODUCTION

Breast cancer is the most common type of cancer among women with a high mortality rate. According to the research by American Cancer Society (ACS) [1], there has been an estimate of 232,670 new cases in the United States of America alone in the year 2014, out of which the death count is more than 40,000. There are several causes of breast cancer and the most effective way to improve the success rate of treatment is its early detection. The vastly used diagnosis modality of breast cancer is through mammography, which has a high success rate in diagnosing radiography based cancer detection. The MCs are detected on the mammogram as highly small bright spots due to its high attenuation compared with the surrounding breast tissue [2]. The MCs are the initial indicators of malignant lesions in the breast tissue, which can be diagnosed by the pattern in which they occur [3]. The diameter of an MC can be as small as 0.1mm occurring among the soft tissue. The region of occurrence has high local luminance mean because of the direct reflection of the X-rays and decreased local contrast resulting from the attenuated rays falling on the X-ray film

(Fig. 1). This makes it hard to detect the MCs. Although significant research effort has been devoted to detection and computer aided diagnosis (CAD) of the breast cancer, there still have been numerous complex problems, among which the classification of the detected MCs as malignant or benign is perhaps one of the most challenging tasks.

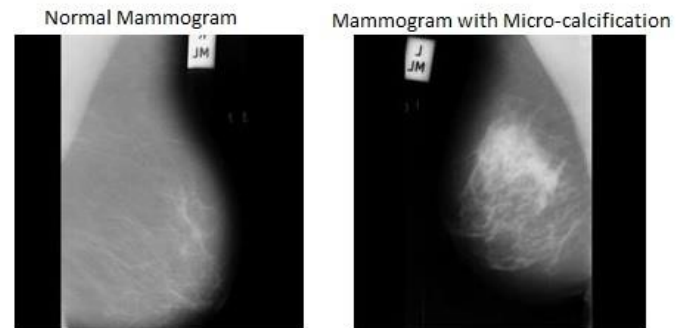


Figure 1. Comparison of a normal mammogram and a mammogram with calcifications.

Several researchers have proposed various approaches to detect MCs in the digital mammograms. For example, Wei Ping et al. [4] has presented a technique for detection of MCs using wavelet transforms with the Daubechies' 10 wavelets and a decomposition level of 4. The Region of Interest (ROI) is enhanced via the Difference of Gaussian (DOG) filter and adaptive thresholding to locate the MCs. The drawback of this method is that it gives a high rate of false positive and does not produce satisfactory results for mammograms with fatty tissue. Hajar Moradmand et al. [5] has presented a comparative study of segmentation methods involving adaptive thresholding and watershed segmentation where the image is processed by morphological operations and Otsu's algorithm. It is then denoised using 4x4 median filter. Contrast enhancement is carried out by unsharp masking and 2D wavelet transform. The results are promising with watershed segmentation compared to adaptive thresholding with the expense of extensive time consuming computations. Po-Tsun Liu et al. [6] proposed a method of detection of MCs and classification of the clusters based on the Gabor wavelet and fractal dimension. The mammograms are preprocessed using a Gaussian smoothing filter and image projection method to isolate the breast region. The back propagation technique is then employed to classify the clusters as malignant or benign. Their method is more

concentrated towards the classification of the MC clusters sacrificing the accuracy of detection to 89.55%. D'Elia et al. [7] use the tree structured Markov Random Fields to segment the mammogram in order to identify the ROI. The textural and geometrical features are then used to classify the segments as MCs using support vector machine technique. Although they achieve creditable results, the algorithm does not perform well with true positives and are highly dependent on the higher resolution and size of the image. Cheng et al. [8] makes a comparative study among the various techniques followed at each stage of processing involved in computer aided detection of MCs in mammograms. Karahaliou et al. [9] have proposed an compelling work of analysis of the tissue surrounding the micro calcification. The proposed approach uses a wavelet based spatially adaptive procedure to enhance the contrast in the image. Further, a window of size 128x128 pixels is selected as the approximation of the surrounding tissue and the respective ROI. The ROI's grey level texture features and wavelet coefficient texture attributes are extracted, and based on the extracted characteristics the MCs are classified as benign or malignant. Although the results are impressive, the major drawback of this approach is the requirement of an input image with identified MC regions.

This paper intends to approach this problem with a perspective, where we concentrate to localize the ROI to the region with a high probability of finding MCs. A search is then conducted for the patterns in the image which are the characteristic attributes of MCs. This is followed by a progressive thresholding as needed.

II. PROPOSED ALGORITHM

The proposed technique consists of three important phases as explained in subsequent sections. The overall approach is also given in a block diagram as shown in Fig. 2.

A. Image Segmentation

In the mammograms, we can notice two distinct regions, the area of interest; i.e., the breast region and the background dark area. Considering only the ROI or the breast region with high probability of finding MCs enables us to reduce the computational cost involved in processing high resolution X-ray mammogram. Hence, we first segment the image into the breast region and the rest of the mammogram. The segmentation of a mammogram is challenging due to the low contrast between the background and the breast region. This is carried out by the Fuzzy C Means (FCM) clustering to isolate the breast region [10]. FCM clustering is a robust segmentation algorithm that allows the data to belong to multiple clusters with different degrees of memberships with each cluster. FCM considers all the pixels of the mammograms as a dataset. All the data points are divided into c clusters. An appropriate level of cluster fuzziness ' m ' is considered as a real value greater than 1. For the sake of simplicity, we set $m = 2$ and $c = 5$. This allows the C-means algorithm to classify the pixels into five different regions or clusters based on the intensity profile of the image. The first region is the unwanted dark background area. Progressively the pixels are classified from one to five with label five for the pixels with the higher intensities. The clusters labeled

four and five are considered for further analysis as they contain image areas necessary for further processing. The FCM segmented image still contains unwanted regions such as the region surrounding the dense soft tissue and the X-ray label which would interfere with the image analysis operations. To eliminate the unwanted areas, the image is further segmented using the white top-hat transform which is defined as difference of the image and its openings [11], as given in Eq. 1. $TH [I (n_1, n_2)]$ is the top-hat transform of the image, $I(n_1, n_2)$ is the input image, and $\gamma (I (n_1, n_2))$ is the opening of the image.

$$TH [I (n_1, n_2)] = I (n_1, n_2) - \gamma (I (n_1, n_2)) \quad (1)$$

The structuring element used is a disk structure with a radius of 18 pixels. The top-hat transform detects the areas with lower intensity and enhances it, whereas areas with high intensity and low contrast are attenuated. This difference allows to perform a histogram based threshold between these two areas to result in an image with reduced region of interest. The isolation of the ROI can be performed in several ways. Here, a marker controlled watershed segmentation is used [13] to capture only the ROI (Figure 5(b)). The ROI is further processed to detect the MCs.

B. Pattern Recognition using the Gibbs Random Fields (GRFs) and Thresholding

Since the MCs do not have a specific shape and occurrence pattern, they can appear in random shapes and orientations. To detect the patterns created by the MCs, we use the relation of the Markov Random Fields (MRFs) and the Gibbs distribution, called the GRFs [14]. The Gibbs conditional probability is given by

$$P(\omega_{ij} | N_{ij}) = \frac{1}{Z} \exp(-\frac{1}{T} \sum_k F_k(C_k(i, j))) \quad (2)$$

where Z is the normalizing factor, T is a parameter, $F_k()$ is the function of states of the pixels in the cliques, $C_k(i, j)$ are cliques in the image, N_{ij} is the defined neighborhood around ω_{ij} , and ω_{ij} is the point at which the observation window is centered. Cliques are the patterns among a pixel and its neighbour that can be observed in a given neighbourhood. A clique can contain a single pixel or multiple pixels. The search for finding the cliques can be carried out using four connectivity or eight connectivity. In this work, eight connectivity is employed. All the possible cliques that can appear in an image are given in Fig. 3. To illustrate the concept, consider a 5x5 local window in Fig. 4, selected from the test image. It can be seen that the pixels in the center has a significantly higher intensity values compared to the background and hence represents an MC pattern. Observing the texture pattern that the center point forms with the neighborhood, the pattern can be represented by two 'L' shaped cliques with different orientations from Fig. 3(c). Similarly, all the MCs in the image can be characterized as a combination of one or more such cliques in Fig. 3.

In order to detect the MCs, the average intensity of the pixels in a clique (3x3 pixel window) is calculated and is compared to the average intensity of the pixels in the neighborhood of 5x5 window size.

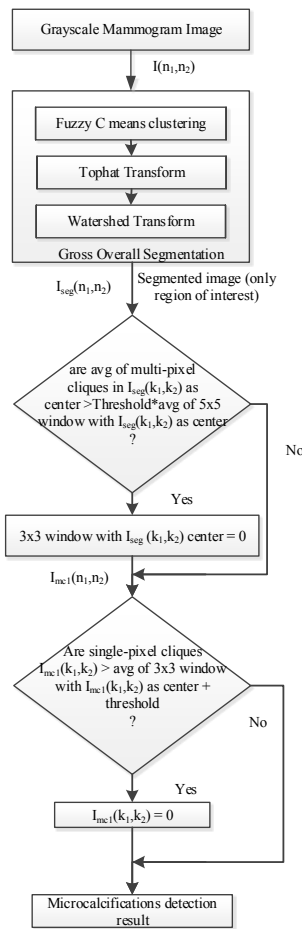


Figure 2. System for the modification of local contrast as a function of luminance mean

If the average intensity of the cliques is greater than a threshold value times the average of the neighborhood, a thresholding is performed on the cliques to indicate the detection of MCs pattern. Furthermore, the small MCs which are in the high intensity areas are missed when searching for MCs using multi-pixel cliques. Thus, single pixel cliques are adopted and by examination of such regions, it is found that the intensity of the pixel in the clique is greater than the product of the average intensities of pixels in the neighborhood of 3x3 window and threshold value.

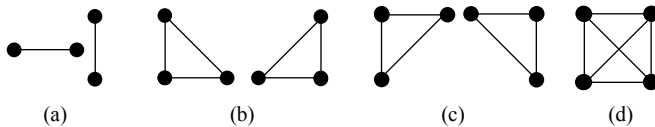


Figure 3. (a) Two pixel cliques, (b) and (c) three pixel cliques, (d) four pixel clique

201	203	204	202	199
202	205	211	203	199
199	211	218	203	196
198	208	217	203	198
196	198	201	197	195

Figure 4. A 5x5 image window from the test image which contains an MC

A thresholding is performed in the selected areas where the above condition is satisfied.

III. EXPERIMENTAL RESULTS

For this research, fifty one mammograms were selected out of 322 images from the digital mammogram database of the Mammographic Image analysis society [15], out of which 25 images have calcifications and MCs, whereas rest were normal.

As the first step, the image segmentation is achieved by applying the FCM to the original image. The unwanted background and the required region of interest are classified separately. The White top-hat transform is performed on the segmented image (Fig. 5(b)). The resulting top-hat transformed image is depicted in Fig. 6(a), the histogram of the top-hat transformed image is shown in Fig. 6(b). This segmentation operation isolates the ROI from the rest of the image as shown in Fig. 7(a). Watershed segmentation is performed to extract the ROI from the rest of the image (Fig. 7(b)). Further processing is conveyed on the extracted ROI (Fig. 7(c)). This saves computational time and hence makes the system computationally less demanding. In the search of MCs, thresholding is performed on the 3x3 windows containing the cliques. The result of finding multi-pixel cliques in the ROI is given in Fig. 8(a). The threshold used to detect the MCs in a multi-pixel clique is 1.047, which is empirically determined in this case. Since the patterns searched are for multi-pixel cliques, several small sized MCs are missed. Hence, a search for single pixel cliques is conducted. For this purpose, a 3x3 window is considered with the center pixel of the window as a single clique, if the intensity of the clique is greater than the product of average of the 3x3 window and the determined threshold value. The threshold value is empirically determined to be 1.054. The results of searching for 1 pixel cliques are shown in Fig. 8 (b). The results of the search demonstrate high detection rate and low false negative rate. One of the drawbacks of the algorithm is that it misclassifies the overlapping of the soft tissues which appear bright, similar to MC but has a smooth appearance. Hence, this method has a slightly higher rate of false positive than desired.

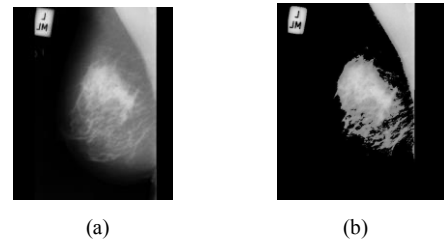


Figure 5. (a) Shows the original mammogram with calcification. (b) Segmented mammogram with isolated breast region.

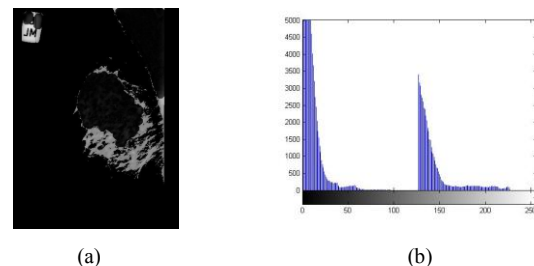


Figure 6. (a) Top-hat transform of the segmented image, (b) Histogram of the top-hat transformed image

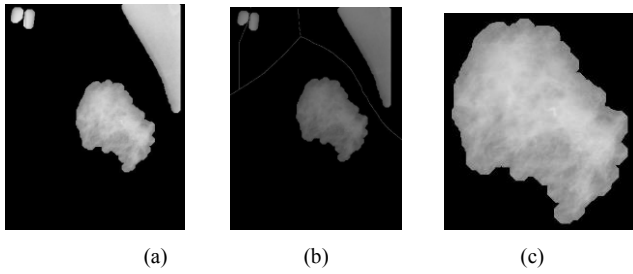


Figure 7. (a) Resultant image of top-hat transform after histogram thresholding, (b) Watershed segmentation of the top-hat transformed image, (c) ROI extracted after watershed segmentation.

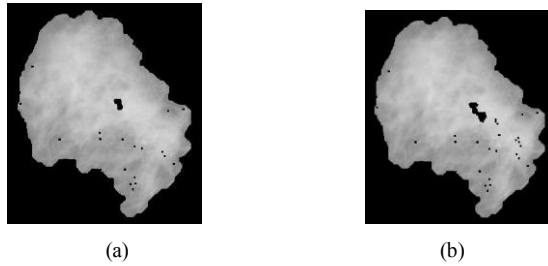


Figure 8. (a) Result of searching for multi-pixel cliques. (b) Result of detection of MCs after the search for multi-pixel cliques and single pixel cliques.

The proposed segmentation works efficiently in isolating the ROI in all the tested images. The overall detection rate (DR) is given by Eq. 3, where FP is the false positive, TP is the true positive, FN is the false negative, and TN is the true negative. The detection rate of the proposed method is found to be 94.4%.

$$DR = \frac{FP + TP}{FP + TP + FN + TN} \quad (3)$$

The sensitivity (S) of the algorithm is measured by

$$S = \frac{TP}{TP + FN} \quad (4)$$

which is found to be 93.7% in this case. The rate of detection of false negatives in the image is 5.6%. The detection accuracy (DA) is given by

$$DA = \frac{TP}{TP + FP} \quad (5)$$

and it is determined to be 88.2%, respectively.

IV. CONCLUSION AND FUTURE WORK

The low contrast nature of the mammogram makes the detection of MCs difficult. Hence, we search for possible MC patterns in the extracted ROI. The search process is computationally efficient since it is implemented in relatively small size local windows as opposed to [9]. Further, using the method described in [9], an analysis can be carried out on the tissue regions close to the detected MCs in order to classify the MCs as malignant or benign. The method proposed in this paper is robust in the sense that the algorithm is not dependent on the orientation and size of the MCs, resolution and size of the image. The method uses

relatively simple image processing operations which is considerate for embedded system implementation.

REFERENCES

- [1] American Cancer Society, "Breast Cancer," 2007.
- [2] M. P. Morgan, et al., "Microcalcifications associated with breast cancer: An epiphenomenon or biologically significant feature of selected tumors?," *J. Mammary Gland Biol. Neoplasia*, vol. 10, no. 2, pp. 181–187, Apr. 2005.
- [3] R. Smithuis and R. H. Pijnappel, Ruud Radiology department, "The Radiology Assistant : Breast - Calcifications Differential Diagnosis," 2008.
- [4] W. Ping, et al., "A method of detection micro-calcifications in mammograms using wavelets and adaptive thresholds," in *2nd Int. Conf. on Bioinformatics and Biomedical Eng., iCBBE*, 2008, pp. 2361–2364.
- [5] H. Moradmand, et al., "Comparing Methods for segmentation of Microcalcification Clusters in Digitized Mammograms," *Int. J. Comput. Sci. Issues*, vol. 8, no. 6, pp. 104–108, 2011.
- [6] L. Po-Tsun, et al., "The Classification of Breast Calcification using Texture Feature of Fractal-based and Gabor Wavelet-based," *Proc. 3rd Int. Conf. Multimed. Technol.*, pp. 1830–1837, 2013.
- [7] C. D'Elia, et al., "Detection of microcalcifications clusters in mammograms through TS-MRF segmentation and SVM-based classification," *Proc. Int. Conf. Pattern Recognit.*, vol. 3, no. C, pp. 742–745, 2004.
- [8] H. D. Cheng, et al. "Approaches for automated detection and classification of masses in mammograms," *Pattern Recognit.*, vol. 39, no. 4, pp. 646–668, 2006.
- [9] A. N. Karahaliou, et al., "Breast cancer diagnosis: Analyzing texture of tissue surrounding microcalcifications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 6, pp. 731–738, 2008.
- [10] S. D. Gahukar and S. S. Salankar, "Segmentation of MRI Brain Image Using Fuzzy C Means For Brain Tumor Diagnosis," *Int. J. Eng. Res. Appl.*, vol. 4, no. 4, pp. 107–111, 2014.
- [11] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," *Optical Engineering*, vol. 34, 1992, pp. 433–433.
- [12] R. C. Gonzalez and R. E. Woods, "Morphological Image Processing," in *Digital Image Processing*, Second Ed., Prentice Hall, 2002, pp. 519 – 566.
- [13] "Marker-Controlled Watershed Segmentation - MATLAB & Simulink Example." [Online]. Available: <http://www.mathworks.com/help/images/examples/marker-controlled-watershed-segmentation.html>. [Accessed: 23-Mar-2015].
- [14] S. Theodoridis and K. Koutroumbas, "Context-Dependent Classification," in *Pattern Recognition*, Fourth Ed., Elsevier, 2009, pp. 554–556.
- [15] "The mini-MIAS database of mammograms." [Online]. Available: <http://peipa.essex.ac.uk/info/mias.html>. [Accessed: 26-Mar-2015].