

Breast cancer detection using modified Hough transform.

Osama R. Shahin^{1,2*}, Meshrif Alruily³, Mansi Alsmarah¹, Musharraf Alruwaill¹

¹Computer Science and Information Department, Faculty of Science & Arts, Jouf University, Saudi Arabia

²Physics and Mathematics Department, Faculty of Engineering, Helwan University, Egypt

³College of Computer and Information Sciences, Jouf University, Saudi Arabia

Abstract

Breast cancer is a result of the uncontrolled cells division. These cells may spread from the breast to other parts of the body through blood vessels. Many symptoms were associated to the breast cancer, such as change in color, skin, and shape. In this work, the breast tumor will detect according to its geometric, texture, and gradient features. Firstly, region of interest (ROI) will detected by using the Hough transform. After the location of breast tumor was determined, these locations will cropping from the original mammogram image to detect the actual tumor, then twenty two features will be computed for the ROIs using predefined equations. These features are used to classify the infected tissues from the normal one. A comparison was made between system results of the breast features and reading of an expert's radiologist to ensure the feasibility and algorithm efficiency.

Keywords: Mammograms, Breast cancer, Hough transform.

Accepted on August 06, 2018

Introduction

Breast cancer is where anomalous cells develop in an uncontrolled manner. It is the most widely recognized reason for death among moderately aged ladies. The World Health Organization's International Agency for Research on Cancer in Lyon, France, has evaluated that in excess of million ladies overall kicked the bucket because of breast cancer every year [1].

Mammography helps in early location and it assumes a critical part in cancer treatment and permits a quicker recuperation for a large portion of the patients [2]. There are two sorts of mammography, film and computerized. Computerized mammography is superior to film mammography since radiation beams can be lessened up to half can even now identify breast cancer, while in film mammography the standard radiation beams can't be diminished [3,4]. Keeping in mind the end goal to enhance the biopsy yield proportion, masses must be named amiable or threatening.

Masses with sporadic in shapes are normally threatening, and those joined with radiolucent shapes are favourable [5,6]. In this work the digital mammogram images that were provided from online mammogram database (MIAS database) will be used [7]. Firstly, for each image the location for each suspicious areas will be decided and secondly a number of significant features-geometric, texture, and gradient features-will be computed [6,8-10].

Tumor Features

There are a lot of features types that describe tumors in image. The function of features is to translate the medical statistics into decisions. Firstly, geometric features describe the geometric properties of the ROI. It is represented as a set of pixels in image [11,12]. Secondly, texture is really very fuzzy concept, these features of attempt to describe grayscale items between adjacent pixels in the image [13]. Finally, gradient is a subsidiary of the image among the neighbourhood respects. The Sobel operator is used to record gradient image as well [14].

Methodology

The suspicious area (ROI) can be easily detected in a picture if the area contains enough background variation. The steps of the proposed can be summarized as follows:

Reading the mammogram image

The images that are used in the algorithm from MIAS database [7], are firstly converted into gray level.

Create gradient image

ROI differs significantly in contrast from the background. Changes in contrast can be detected by different operators such as Sobel, Prewitt, Roberts, Canny, etc. In this work, Sobel detector is used, due to its design simplicity.

Subtracting the original image from the image after filling its gaps

After some of pre-processing process such as image dilates, fill interior gaps; the subtraction of the background from the filling images will be preceded. This process is very essential, to enhance the periphery of the ROI to make it easy to fit it in a circle that made in the following stage using Hough transform.

Feature	Equation	
Area meas- ure	$Area = \sum_i \sum_j (A_{i,j}, X_{ROI}(Area) = i, Y_{ROI}(Area) = j)$	(1)
Perimeter	$Perimeter = \sum_i \sum_j (P_{i,j}, X_{edge}(Perimeter) = i, Y_{edge}(Perimeter) = j)$	(2)
Compact- ness	$Compactness = \frac{p^2}{A}$	(3)
Mean	$Mean = \sum_{i=0}^{Area} Input\ Image[X_{ROI}(i), Y_{ROI}(i)]$	(4)
Mean Global Area	$Mean\ Global = \sum_i \sum_j Input\ Image[i, j]$	(5)
Mean Local Area	$Mean\ Local = \sum_i \sum_j Input\ Image[i, j]$	(6)
Uniformity	$Uniformity = \sum_{i=0}^{Area} NH^2[Input\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(7)
Standard Deviation	$SD = \sum_{i=0}^{Area} ([Input\ Image[X_{ROI}(i), Y_{ROI}(i)]] - Mean)^2$	(8)
Smoothness	$Smoothness = \frac{1}{1 + SD^2}$	(9)
Skewness	$Skewness = \sum_{i=0}^{Area} ([Input\ Image[X_{ROI}(i), Y_{ROI}(i)]] - Mean)$	(10)
Entropy	$Entropy = \sum_{i=0}^{Area} NH[Input\ Image[X_{ROI}(i), Y_{ROI}(i)]] * \log NH[Input\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(11)
Correlation	$Correlation = \sum_{i=0}^{Area} NH[Input\ Image[X_{ROI}(i), Y_{ROI}(i)]] * Input\ Image[X_{ROI}(i), Y_{ROI}(i)]$	(12)
Inverse	$Inverse = \sum_{i=0}^{Area} \left(\frac{1}{1 + (X_{ROI}(i) - Y_{ROI}(i))^2} \right) * NH[Input\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(13)
Sobal-Mean	$Sobal\ Mean = \sum_{i=0}^{Area} Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]$	(14)
Sobal-Mean Global Area	$Sobal\ Mean\ Global = \sum_i \sum_j Gradient\ Image[i, j]$	(15)
Sobal- Uniformity	$Sobal\ Uniformity = \sum_{i=0}^{Area} GNH^2[Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(16)
Sobal- Standard Deviation:	$Sobal\ SD = \sum_{i=0}^{Area} ([Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]] - Mean)^2$	(17)
Sobal- Smoothness:	$Sobal\ Smoothness = \frac{1}{1 + Sobal\ SD^2}$	(18)
Sobal- Skewness	$Sobal\ Skewness = \sum_{i=0}^{Area} ([Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]] - Mean)$	(19)
Sobal- Entropy	$\sum_{i=0}^{Area} GNH[Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]] * \log PG[Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(20)
Sobal- Correlation	$Sobal\ Correlation = \sum_{i=0}^{Area} GNH[Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]] * Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]$	(21)
Sobal- Inverse	$Sobal\ Inverse = \sum_{i=0}^{Area} \left(\frac{1}{1 + (X_{ROI}(i) - Y_{ROI}(i))^2} \right) * GNH[Gradient\ Image[X_{ROI}(i), Y_{ROI}(i)]]$	(22)

Figure 1. List of tumor features.

Detect the circular shape using the Hough transform

Hough transform is used here to surround the tumor area by a circle that closes to the surface dimension of the tumor [15]. The pseudo code of the algorithm can be summarized in Table 1.

Table 1. Pseudo code for modified Hough transform.

```

Load image
Find the edges in the image
For each edge point (xo, yo)
For (xo=xo_min; xo ≤ xo_max; xo++)
    
```

```

For (yo=yo_min; yo ≤ yo_max; yo++)
r=sqrt ((x-xo)^2+(y-yo)^2);
Accumulator (r) (xo) (yo)++; //Voting
Find the average values in accumulator for (r) (xo) (yo)
    
```

Determine the features for the ROI

Here, the geometric, texture, and gradient features of ROI will be calculated using equations that predefined in [6,12,13]. Equations that needed to detect tumor features were listed in Figure 1.

Experimental Results

This part introduces the experimental results of the proposed algorithm on the dataset consisting of 322 breast images from MIAS database that covering all cases of masses (malignant and benign) by different shapes, and size. The goal is to test and validate the proposed system presented in this paper in order to select the most suitable system. The total number of malignant samples that used here were 52, while the number for benign was 66 samples; finally normal samples were equal to 204. Table 2 lists the description of all images in the MIAS database and supplies the following significant information [16]. A portion of the tissues are darkened by a splendid area in the focal point of the mammogram. This brilliant locale makes the image be unevenly lit up.

Table 2. MIAS database description.

	Fatty tissues	Fatty glandular tissues	Dense tissues	Total	
Calcification	B:2	B:5	B:5	B:12	25
	M:4	M:4	M:5	M:13	
Circumscribed masses	B:7	B:6	B:3	B:16	20
	M:2	M:2	M:0	M:4	
Spiculated masses	B:2	B:4	B:6	B:12	21
	M:4	M:3	M:2	M:9	
Other, ill-defined masses	B:4	B:2	B:1	B:8	15
	M:4	M:4	M:1	M:7	
Architectural distortion	B:4	B:3	B:4	B:10	20
	M:2	M:4	M:4	M:10	
Asymmetry	B:2	B:3	B:3	B:8	17
	M:3	M:2	M:4	M:9	
Normal	66	64	74	-	204
Total	106	104	112	-	322

B: Benign; M: Malignant.

Because the background of the mammogram is dark and the image of tumor masses are lighter than the background of the

image, the using of the top-hat transform is required to reduce uneven lighting. This transformation is defined as the difference between the original image and its opening. The opening of the image is the set of front parts of the image that fit a particular structural element. After the location of the tumor area from the entire breast locale the features for these district will be ascertained, as specified before; 22 features will be calculated and will be spared in database which contains 322 row, each row for one image and 22 features extraction that calculated based on equations (1-22). The terminology measurement, i.e. (performance criteria) of the work for system is shown in Table 3. After the location of breast cancer was determined by previous method, these locations will cropping

Table 3. Terminology measurement of the algorithm.

TN	TP	FP	FN	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)
164	136	16	6	95.77	90.9	93.17	89.47

Table 4. Feature extraction range values for cancer tumor.

Feature	Maximum	Minimum
Area	260	75
Perimeter	3580	270
Compactness	34.554	17.456
Mean	242.802	142.133
Mean global area	9.245	1.968
Mean local area	13.176	1.858
Uniformity	4.993	1.177
SD	32.632	12.866
Smoothness	0.006	0
Skewness	37.466	-8.203
Entropy	-27.676	-197.477
Correlation	569	48
Inverse	0.6446	0.001
Sobal mean	97.459	44.597
Sobal mean global	8.472	1.97
Sobal uniformity	3.583	0.005
Sobal SD	27.639	5.948
Sobal smoothness	0.0538	0.001
Sobal skewness	25.005	-5.538
Sobal entropy	-20.214	-250.459
Sobal correlation	378	8
Sobal inverse	5.846	0.001

Table 5. Performance comparison with the other reliable classifiers in terms of accuracy (Acc.).

from the original mammogram image to detect the actual tumor as depicted in Figure 2. As a result of breast tumor detection, a 22 features based upon the equations that explained in the methodology section will be calculate that shown in Table 4.

The results of the computed features by using the equations (2-23) appeared in Table 4. Based upon these features, the classification between the normal tissues and the other infected tissues can be determined by aiding of an expert radiologist.

The performance measures of the proposed algorithm with the other reliable classifiers are depicted in Table 5.

Author	Methodology	Acc.
Trbalić et al. [17]	ROIs are obtained using a proficient Otsu's N thresholding and then subjected to pre-processing stage. Then a collection of 32 Zernike moments with dissimilar iterations and orders have been extracted. These moments applied to the neural network classifier.	90%
Liu et al. [18]	The regions of interest (ROIs) are located by using multiple concentric layers (MCL) approach. Then, ROIs are further refined with narrow band based active contour (NBAC), which can enhance the segmentation process.	92%
Xie et al. [19]	A new CAD system depends on the maximum learning machine (ELM). The proposed system contains extraction feature, selection of parameters, and rating. After pre-treatment, areas containing tumor masses are detected.	95%~96%
Proposed	The regions of interest (ROIs) can be detected via modified Hough transform, the find the features of this region by means of predefined equations described in paper, and compare it with normal values with advice of expert radiologist.	96%

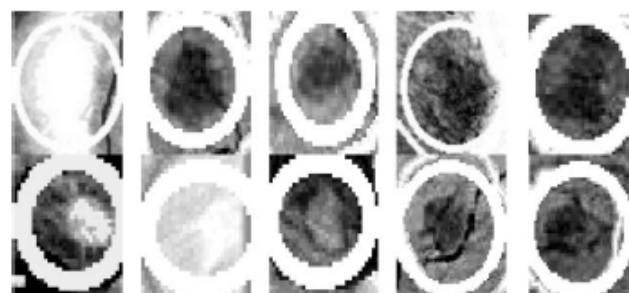


Figure 2. Tumor detection by Hough transform.

Conclusion

From the comparison tests that for result shown in Table 5, we see that all features obtained by the proposed algorithm are matching with those features in the previous works. So, the proposed algorithm is able to detect the region of the tumors

and determine its locations. In addition to the last benefit it could detect more than one tumor in the same breast region which gives power to the proposed algorithm. As seen from the experimental results and from the performance comparison, the proposed features extraction shows a good result. Thus, it has an ability to implement such breast cancer detection system.

References

1. Nixon Mark S, Alberto S Aguado. Feature extraction & image processing for computer vision. Academic Press 2012.
2. Sameti M, Rabab Kreidieh W, Jacqueline M-P, Branko P. Image feature extraction in the last screening mammograms prior to detection of breast cancer. *IEEE J Selected Topics in Signal Processing* 2009; 3: 46-52.
3. Madabhushi A, George L. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Analysis* 2016; 33: 170-175.
4. Mustra M, Mislav G, Rangaraj MR. Re-view of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. *Med Biol Eng Computing* 2016; 54: 1003-1024.
5. Qiu Y, Shiju Y, Maxine T, Samuel C, Hong L, Bin Z. Computer-aided classification of mammographic masses using the deep learning technology: a preliminary study. In *Med Imaging 2016: Computer-Aided Diagnosis* 2016; 9785: 978520.
6. Shahin OR. Brain tumor detection using watershed transform. *Ann Clin Cytol Pathol* 2018; 4: 1096.
7. Astley S, Betal D, Cerneaz N, Dance DR, Kok SL, Parker J, Ricketts I, Savage J, Stamatakis E, Taylor P. The mini-MIAS database of mammograms. MAIS database 2018.
8. Syed L, Saima J, Manimala S. Telemammography: a novel approach for early detection of breast cancer through wavelets based image processing and machine learning techniques. In *Advances in Soft Computing and Machine Learning in Image Processing* 2018; 149-183.
9. Koundal D, Savita G, Sukhwinder S. Computer aided thyroid nodule detection system using medical ultrasound images. *Biomed Signal Processing and Control* 2018; 40: 117-130.
10. Yeo SY, Romero J, Loper M, Machann J, Black M. Shape estimation of subcutaneous adipose tissue using an articulated statistical shape model. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 2018; 6: 51-58.
11. d'Anjou M-A. Principles of computed tomography and magnetic resonance imaging. *Textbook of Veterinary Diagnostic Radiology-E-Book* 2017; 71.
12. Dhungel N, Gustavo C, Andrew PB. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med Image Analysis* 2017; 37: 114-128.
13. Mughal B, Nazeer M, Muhammad S, Tanzila S, Amjad R. Extraction of breast border and removal of pectoral muscle in wavelet domain. *Biomed Res* 2017; 28.
14. María Victoria CC, Patricia Rayon V. Circumscribed mass detection in digital mammograms. *IEEE* 2006.
15. Jain L, Devesh M. Application of Hough transform for finding parametric curves. *Int J Comput Appl Eng Sci* 2011; 1: 100-103.
16. Oliveira Júlia EE, Mark OG, Arnaldo de AA, Bastian O, Thomas MD. Toward a standard reference database for computer-aided mammography. In *Medical Imaging* 2008; 69151Y.
17. Šerifović-Trbalić AT, Demirović D, Prljača N, Cattin PC. Classification of benign and malignant masses in breast mammograms *MIPRO* 2014; 26-30.
18. Liu X, Zeng Z. A new automatic mass detection method for breast cancer with false positive reduction. *Neurocomputing* 2015; 152: 388-402.
19. Xie W, Yunsong L, Yide M. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* 2016; 173: 930-941.

*Correspondence to

Osama R. Shahin

Computer Science and Information Department

Faculty of Science & Arts

Jouf University

Saudi Arabia