ISSN No. 0976-5697

Volume 7, No. 5, September-October 2016

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

FEATURES EXTRACTION FOR MASSES IN DIGITAL MAMMOGRAMS USING STATISTICAL AND TEXTURAL MEASURES

P.Shanmugavadivu
Department of Computer Science and Applications
Gandhigram Rural Institute - Deemed University,
Gandhigram, Tamil Nadu,
India

A.Krishnaveni
Department of Computer Science and Applications
Gandhigram Rural Institute - Deemed University,
Gandhigram, Tamil Nadu,
India

S.G.Lakshmi Narayanan
Department of Computer Science and Applications
Gandhigram Rural Institute - Deemed University, Gandhigram, Tamil Nadu,
India

Abstract: Breast Cancer is reported to be one of the major causes of mortality among the women across the world. Digital mammogram is considered to be a reliable breast screening tool due to its low-dose of energy conservation. This paper aims at the detection and segmentation of breast masses using watershed segmentation. Further, this research work explores the statistical and texture features of the segmented masses. These features form the foundation for the classification and detection of breast cancer. The proposed method named as Feature Extraction using Statistical and Texture Measure (FESTM) is experimented on the Mammographic Image Analysis Society (MIAS) test datasets, for validation and verification. It is confirmed that the segmentation results align with the ground truth data of test images.

Keywords: Digital Mammogram, Threshold, Watershed Segmentation, Feature Extraction , Cancer Detection, Statistical and Textural features

I. INTRODUCTION

A two-dimensional function f(x,y) is termed as a digital image, where x and y are the spatial co-ordinates and f designates the intensity at (x,y). A digital image is a composition of finite pixels, each with unique spatial coordinates. Digital Image Processing helps to process such images using digital computer [1,2].

Medical image processing captures the characteristics of the internal parts of the human body for the clinical intervention that leads to the analysis of vital organs/tissues [3,4]. Breast cancer has became a predominant threat, leading to women mortality. Early detection is the only remedy to contain the victims of this disease. The tumors/masses present in the breast may either be benign or malignant [5,6].

Digital mammography, obtained using low-range x-rays is the best and reliable source for the early detection of breast cancer. Computer Aided Diagnosis (CAD) tools are widely used for breast cancer analysis that comprises of a sequential processing of digital mammograms namely pre-processing, segmentation, feature extraction and classification [7,8].

MIAS is commonly used as a benchmark for testing new proposals dealing with processing and analysis of mammograms for breast cancer detection. Each case in this database is annotated by expert radiologists and the complete information is provided as an overlay file.

In the MIAS description the first column denotes the feature of background tissue as F – Fatty, G - Fatty-glandular and D - Dense-glandular. The second column specifies the class of abnormalities present in the image as CALC – Calcification, CIRC - Well-defined /circumscribed masses, SPIC - Spiculated masses, MISC - Other, ill-defined masses, ARCH - Architectural distortion, ASYM – Asymmetry and NORM – Normal. The third column specifies the severity of the abnormality as B – Benign and M – Malignant. The fourth and fifth column presents co-ordinate pair of centre of abnormalities (x,y) and the sixth column specifies about the approximate radius (in pixel) of a circle enclosing the abnormality.

Image segmentation is an element of digital image processing, which helps to sub-divide a given image into its constituent objects/regions [9,10]. It is the process of separating the masses in the RoI from the mammogram, for the purpose of cancer diagnosis / detection [11,12].

Generally, the intensity of the masses are confirmed to differ from the surrounding tissues.

Thresholding techniques help to bifurcate the images into its foreground and background regions, based on the histogram of the image [13,14]. Watershed segmentation is a technique that distributes and groups the pixels of the image based on their intensity levels. Features extraction phase aims at collecting the intensity details of the masses / tumors [15].

In this research article, Section II describes the methodology of the newly devised method for image segmentation and feature extraction technique. The results and discussion are given in section III and the conclusion is presented in section IV.

II. METHODOLOGY OF FESTM

The devised method Feature Extraction using Statistical and Texture Measures (FESTM), has two distinct phases: segmentation and feature extraction. The segmentation is performed using fixed point watershed algorithm, whereas the latter uses statistical and textural measures.

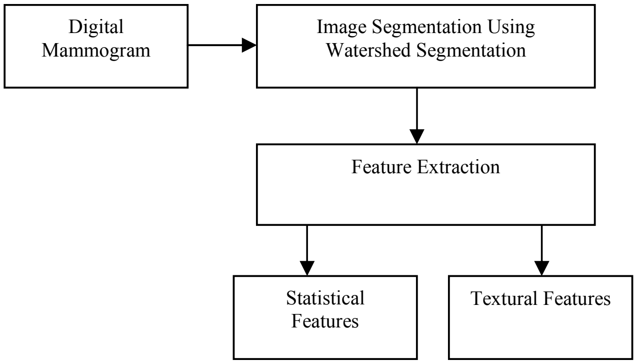

Fig.1: Block Diagram of FESTM

A. Segmentation

Image segmentation plays a major role in medical sciences. The purpose of image segmentation is to partition the images which have different characteristic tissues into semantically interpretable regions. Based on the characteristics of each region, objects of interest are extracted [16,17].

For extracting the specific region, threshold and watershed transformation are used. The main objective of using threshold is to classify each image pixel into two categories either as foreground or background. Thresholding is the process of picking up a fixed gray scale value. Moreover, it is used to locate the pixel with distinct intensity variations [18-20].

Image segmentation by mathematical morphology is a methodology based on the watershed transformation. It is a powerful tool for image segmentation, which objectively targets the watershed lines in topographic surface. The region that the watershed separates is called catchment basin [21,22]. The algorithmic description of the fixed point watershed technique is described in the following section.

B. Computational Procedure of FESTM

Input: 2D Digital Mammogram
Output: Segmented masses and the feature of masses
Step 1: Read input digital mammogram.
Step 2: Compute Histogram, max and min of the input mammogram.
Step 3: Compute ω and μ value.
/* ω: Cumulative sum of peak values and
μ: Product of ω and number of bins in the histogram */
Step 4: Calculate σ value by using Full Objective criteria matrix,

$$\sigma = (\mu_t * \omega - \mu)^2 / \omega * (1-\omega)$$

Step 5: Find maximum of the σ

$$Max_{value} = \max(\sigma)$$

Step 6: Compute

$$idx = find(\sigma == Max_{value})$$

Step 7: Find threshold (t) value for segmentation

$$t = mean(idx) - 1$$

Step 8: Repeat the process for entire *t* value
Initialize the $quantized_{image}$
SI = ones(size($Gray_{image}$))
SI=(SI+($Gray_{image}$>t))
where,i=1,2,….max(t)
(SI - Segmented Image)
Step 9: Stop

C. Feature Extraction

Feature extraction is essentially inevitable for mammogram analysis and better selection of feature gives higher accuracy. It basically separates the visual information from the image and stores them in the form of feature vectors in a feature database. These feature values are called as feature vectors of image which help to find the image information from the feature extraction. These feature vectors are used to compare the query image with the images stored in the database [23].

The extraction of texture feature plays a vital role in detecting abnormalities of mammograms. Texture features are proved to be useful in differentiating masses and normal breast tissues [24]. The proposed method uses five statistical features and four texture features, which can further be examined for mammogram classification and analysis. The principle of the statistical and texture features chosen for this study are described in the next section.

A. Statistical Features

The statistical measures and texture features used in this article for the feature extraction of the mammogram images are listed and described in the following section.

In the presented work, Statistical features and Texture features are extracted for the segmented mass from the given input mammogram image. The features used for this research article and the computation details are explained herein under.

i. Mean

The mean gives the average value of all pixels that represents the average intensity of the image, is given by

$$\mu = \frac{1}{MN} \sum_{j=1}^{N} p(i,j) \tag{1}$$

where M and N are numbers of rows and columns in the input image.

ii. Standard Deviation

The standard deviation gives the mean square deviation the pixel value p(i,j) from its mean value μ, and is given as

$$\sigma = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (p(i,j) - \mu)^2 \tag{2}$$

where M and N are numbers of rows and columns in the input image.

iii. Variance

The variance is measured as the squared root of the standard deviation σ which is given by

$$\mathrm{var} = \sqrt{\sigma} \tag{3}$$

iv. Skewness

Skewness gives the asymmetry of the gray levels with respect to the sample mean. If skewness is negative, the data are around the left of the mean than to the right. If skewness is positive, the data are out to the right. The skewness S is given by,

$$S = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{p(i,j)-\mu}{\sigma}\right)^3 \quad \textbf{(4)}$$

v. Kurtosis

Kurtosis explains the shape of the tail of the histogram and is given by

$$K = \left\{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{p(i,j)-\mu}{\sigma}\right)^4\right\} - 3 \quad \textbf{(5)}$$

where p(i,j) are pixels of the image, M and N are the number of rows and columns in the input image, μ is the mean and σ is the standard deviation of the input image.

B. Texture Features

Texture features indicate the surface characteristics and appearance of an object represented by the size, shape, density, arrangement and proportion of its elementary parts.

i. Coarseness

Coarseness helps to locate the largest size at which a texture exists. Computation uses averages at every point over neighborhoods, the linear size of which are powers of 2. The average over the neighborhood of size $2k \times 2k$ at the point (x, y) is given by

$$Ak(x,y) = \sum_{i=x-2k}^{x+2k}\sum_{j=2k-1}^{y+2k-1}\frac{f(i,j)}{2^{2k}} \quad \textbf{(6)}$$

ii. Contrast

The contrast is used to capture the range of gray levels in an image, together with the polarisation of the distribution of black and white using the standard deviation of gray levels and the kurtosis. The contrast measure is defined as

$$F_{con} = \frac{\mu}{(a_4)^n} \text{ where } a_4 = \frac{\mu^4}{\sigma^4} \quad \textbf{(7)}$$

where μ_4 is the fourth moment about the mean and σ^2 is the variance.

iii. Directionality

Directionality is computed using the frequency distribution of local angles. This feature measures the total degree of directionality.

$$\left[\text{Directionality} = 1 - rn_{peaks}\sum_{p=1}^{npeaks}\sum adwp(2-2p)^{2m}\,\text{Directionality}(a)\right] \quad \textbf{(8)}$$

where npeaks number of peaks, ap is the position of the peak, wp is the range of the angles attribute to the p^{th} peak, r denotes the normalization factor related to the quantizing levels of the angles and a indicates quantized directional angle, H Directionality is the histogram of quantized direction values, a is found by counting the number of edge pixels with the corresponding directional angles.

iv. Roughness

Roughness is measured as the summation of contrast and coarseness of an image.

$$\text{Roughness} = \text{Contrast} + \text{Corseness} \quad \textbf{(9)}$$

III. Results and Discussion

The proposed method has been tested by using MIAS test dataset in which the mammograms are in the size of 1024x1024 pixels, with the resolution of 200 microns. This database is composed of 322 mammograms of right and left breast, from 161 patients, where 54 are diagnosed as malignant, 69 benign and 207 normal.

For illustrative purpose, the segmentation results are obtained for five mammograms among the test dataset, mdb001, mdb003, mdb025, mdb049 and mdb054 are given in fig.2.

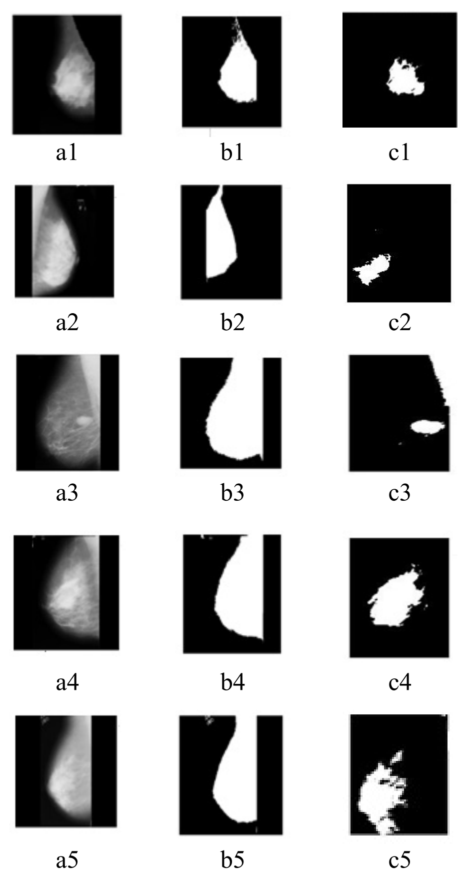

Fig.2:
(a1) - (a5): Input Digital Mammogram; (a1): mdb001; (a2): mdb003; (a3): m db025; (a4): mdb049 and (a5): mdb054.
(b1) - (b5): Thresholded binary Images of (a1) - (a5).
(c1) - (c5): Segmented image of (a1) - (a5).

Table 1: Statistical and Textures Features of Segmented Mammograms

Features	**mdb001**	**mdb003**	**mdb028**	**mdb049**	**mdb054**
Mean	12.0162	1.2331	16.8283	7.9786	16.2411
Standard Deviation	46.3805	16.3455	54.8789	40.8954	55.9209
variance	6.8103	4.0430	7.4080	6.3950	7.4780
Skewness	3.6247	13.1843	2.9541	4.9350	3.1615
Kurtosis	11.2358	1.9923	7.0992	22.3776	8.0284
Coarseness	0.2591	0.0754	0.3066	0.1951	0.2904
Contrast	23.8775	4.4948	30.7846	18.2206	30.6864
Directionality	3.1416	3.1416	3.1416	3.1416	3.1416
Roughness	24.1366	4.5702	31.0912	18.4157	30.9708

This database contains a list of the mammograms in the MIAS database and provides appropriate details, namely, the class of abnormality, image-coordinates of centre of abnormality and approximate radius (in pixels) of a circle enclosing the abnormality. The performance of the Watershed Segmentation is measured using accuracy, sensitivity and specificity for malignancy detection, where

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \qquad (12)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (13)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (14)$$

TP: the number of true positives.
(predicts breast images are correctly)
TN: the number of true negatives.
(predicts non-tumor images are non- tumor)
FN: the number of false negatives.
(predicts tumor images are wrongly non- tumor)
FP: the number of false positives.
(predicts wrongly breast images incorrectly)

The value of each point in the output image is based on a comparison of the corresponding point in the input image with its neighbours.

For this work, 60 mammograms are taken from the MIAS database, among which the proposed method has detected 26 as malignant, 12 as benign and 22 as normal images. The results are compared with the coordinates described in MIAS ground values.

From the total sample set of 60 data, the nature of severity is computed and compared with the MIAS descriptions in the form of confusion matrix, as recorded in Table 2.

Table 2: Segmentation Evaluation

Evaluation		
Positive		Negative
Test Positive	TP = 38	FP = 8
Test Negative	FN = 2	TN = 12
Accuracy	Sensitivity	Specificity
83.333%	95%	60%

Table 3 exhibits the accuracy of detection in comparison with the coordinates in the MIAS description.

The obtained results are validated with the ground reality descriptions provided in the MIAS dataset. The above results are very close to those specifications for all the test images. The initial and final co-ordinate points of the experimental results were obtained, by sampling the input images in Adobe Photoshop CS3 version.

The comparative description between MIAS specifications and the obtained results of the proposed method are furnished in Table 3.

Table 3: Comparative Description between MIAS and Obtained results

Image Referen ce	**MIAS Description**	**Co-ordinates of segmented region**	
		Initial position	***Final position***
mdb001	G CIRC B 535 425 197	546 432	670 454
mdb003	D NORM	-	328 512
mdb025	F CIRC B 674 443 79	628 439	743 472
mdb049	G NORM	-	421 602
mdb054	D NORM	-	354 534

It is visually evident that the segmentation technique has segmented the regions of varying intensities indicating varying disorders of different sizes in various locations. These segmented images may further be subjected to intensity classification using which the features of microcalcifications.

IV. CONCLUSION

In this paper, Fixed point Watershed Segmentation is used for mass detection and 5 Statistical Features and 4 Texture Features are extracted. The research outcomes of the proposed FESTM has further scope for the classification of the extracted mass textures either as normal (benign) or abnormal (malignant) leading to mammogram analysis.

V. REFERENCES

[1] Gonzalez. R. C, R.E. Woods, S.L. Eddins, "Digital Image Processing", Prentice Hall of India, II Edition, 2006.

[2] D.Pnraj, M.Jenifer, P.Poongodi and J. Manoharan, "A Survey on the Preprocessing Techniques of Mammograms for the Detection of Breast Cancer," Journal of Emerging Trends in computing and Information Sciences, Vol. 2, No. 12, PP.656-664, 2011.

[3] Gøtzsche PC, Nielsen M, "Screening for breast cancer with mammography", Cochrane Database Syst*Rev*(Online): Available. doi:10.1002/14651858.CD001877.pub4,2011.

[4] K. H. Ng and M. Muttarak, "Advances in mammography have improved early detection of breast cancer," J. Hong Kong College Radiol., vol. 6, no. 3, pp. 126131, 2003.

[5] P.Shanmugavadivu, V.Sivakumar, "Fractal Dimension-Bound Spatio-Temporal Analysis of Digital Mammograms", Eur. Phys. J. Special Topics © EDP Sciences, Springer-Verlag, DOI: 10.1140/epjst/e2016-02613-0, Springer, 2016.

[6] P.Shanmugavadivu, V.Sivakumar, "Segmentation of Masses in Digital Mammograms using Fractal-Bound Computing Technique for Breast Cancer Prognosis", International Journal of Applied Engineering Research, ISSN 0973-4562, pp. 23187 – 23192, Vol. 10 No.31, 2015.

[7] J.S. Suri, R.M. Rangayyan: "Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer", SPIE, Bellingham, Washington, USA, 2006.

[8] Moumena Al-Bayati, Ali El-Zaart, "Mammogram Images Thresholding for Breast Cancer Detection Using Different Thresholding Methods", Advances in Breast Cancer Research, 2, 72-77, 2013.

[9] P.Shanmugavadivu, V.Sivakumar, "A Novel Technique for Mammogram Mass Segmentation Using Fractal Adaptive Thresholding", Lecture Notes in Electrical Engineering, Chapter 25, Vol. 291, pp. 213-220, ISSN No: 1876-1100, DOI 10.1007/978-981-4585-42-2_25, Springer, 2014.

[10] P.Shanmugavadivu, V.Sivakumar, "Intensity-Based Detection of Microcalcification Clusters in Digital Mammograms using Fractal Dimension", Advances in Intelligent Systems and Computing, Vol. 236, pp. 1293-1299, ISSN No: 2194-5357, DOI 10.1007/978-81-322-1602-5_135, Springer, 2014.

[11] P.DanielRatnaRaju, G.Neelima, "Image Segmentation by using Histogram Thresholding,", International Journal of Computer Science & Engineering Technology (IJCSET) Vol 2, Issue 1,776-779, January 2012.

[12] Sura Ramzi Shareef, "Breast Cancer Detection Based on Watershed Transformation", IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 1, No 1, January 2014.

[13] Shanmugavadivu P, Lakshmi Narayanan S G," Detection of Microcalcifications in Mammograms using Statistical Measures Based Region Growing", Proceedings of SPIE of the International Conference on Communication and Electronics System Design, 2013, vol 8760, pp. 8760 1M-1-6.

[14] Shanmugavadivu P, Lakshmi Narayanan S G," Segmentation of Microcalcifications in Digital Mammogram Images using Intensity Modified BlackTop Hat Transformation and Gauss Distribution", LNAI proceedings of the International Conference on Mining Intelligence and Knowledge Exploration – (MIKE)2013, pp. 234 – 240.

[15] Neelima Bagri and Punit Kumar Johari, "A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval", International Journal of Advanced Science and Technology Vol.80, pp.41-52, 2015.

[16] Shanmugavadivu P, Lakshmi Narayanan S G," Segmentation of Microcalcification Regions in Digital Mammograms using Self-Guided Region Growing", Recent Advancements on Science and Engineering Innovation, IEEE proceedings of the International Conference on Emerging Trends in Science Engineering and Technology INCOSET-2012, pp. 274 – 279.

[17] Shanmugavadivu P, Lakshmi Narayanan S G," Segmentation of Microcalcifications in Mammogram Images using Intensity-Directed Region Growing", IEEE proceedings of the International Conference on Computer Communication and Informatics (ICCCI – 2013).

[18] R.G.Selkar, M.N.Thakare, "Brain Tumor Detection and Segmentation by using Thresholding and Watershed Algorithm" , International Journal of Advanced Information and Communication Technology (IJAICT) Volume 1, Issue 3, July 2014.

[19] Salem Saleh Al-amri, N.V. Kalyankar and Khamitkar S.D, "Image Segmentation by Using Thershod Techniques", Journal of computing, volume 2, issue 5 , issn 2151-9617, may 2010.

[20] Amandeep Kaur, Aayushi , "Image Segmentation Using Watershed Transform", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4, Issue-1, March 2014.

[21] Khushboo Mantri, Shiv Kumar, "MRI Image Segmentation Using Gradient Based Watershed Transform In Level Set Method for a Medical Diagnosis System", Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 11(Version 1), pp.27-36, November 2014.

[22] Pradeep N, Girisha H, Sreepathi B and Karibasappa K., "Feature Extraction of Mammograms", International Journal of Bioinformatics Research ISSN: 0975–3087 & E-ISSN: 0975–9115 , Volume 4, Issue 1, pp.-241-244, 2012.

[23] Neelima Bagri and Punit Kumar Johari, "A Comparative Study on Feature Extraction using Texture and Shape for Content Based Image Retrieval", International Journal of Advanced Science and Technology Vol.80, pp.41-52, 2015.

[24] Tamura, H., Mori, S., Yamawaki, T., "Textural features corresponding to visual Perception", IEEE Trans on Systems,Man and Cybernetics 8, 460–472, 1978.