Proceedings of IOE Graduate Conference, 2017
Volume: 5 ISSN: 2350-8914 (Online), 2350-8906 (Print)

An Approach to Extract Features of Mammography Images for Early Detection of Breast Cancer

Ujjwal Thapa Magar [a], Dibakar Raj Pant [b]

[a, b] *Department of Electronics and Computer Engineering, Pulchowk Campus, IOE, TU, Nepal*
Corresponding Email: [a] ujjwalthapa2307@ioe.edu.np

Abstract

Lesions and its contours are prominent signatures to determine malignancy in mammograms. Detection of the masses and their spread in mammogram is important for radiologists. In this paper, Mammogram image is enhancement using homomorphic filtering and adaptive histogram equalization. The enhanced mammogram image is segmented using K means clustering and contour is extracted using morphological operations.The edges are detected by Sobel operator and extracted seven geometric features from the lesions. Lesions and its contours are prominent signatures to determine malignancy in mammograms.It is found that malignant lesions have speculated or ill-defined boundary and benign mass have smooth boundary. Classifications of malignant and benign are done by distance versus angle of signature.The average value of area in malignant and benign segmented mammogram images is 1073.6 and 316.7 square unit respectively. The value of area in malignant mammogram images is greater than benign images. The average range value of radius in malignant and benign mammogram images is 77.38 and 17.58 unit respectively. Signature value of range in malignant image is higher in comparison to benign image.

Keywords

Breast Cancer Detection – Mammography – K-Means Clustering – Boundary

1. Introduction

Medical images are rich in information that can be used for diagnosis and subsequent medical interventions. Information provide by medical image has become an indispensable part of today's patient care. Cancer is the unrepressed development of unusual cells in the body which account for the most dangerous and life threatening diseases in the world.Cancer is the second leading cause of death globally, and was responsible for 8.8 million deaths in 2015. Globally, nearly 1 in 6 deaths is due to cancer[1]. Mammography is specialized medical imaging that uses a low-dose x-ray system to see inside the breasts. A mammography exam, called a mammogram, aids in the early detection and diagnosis of breast diseases in women[2].

Work has been done on segmentation of mass in past to know the spread of spiculation in the breast tissue. Mean shift algorithm and Fuzzy C-means and active contour models are used in [4] for the detection of masses. Suspicious focal areas are found for testing

Figure 1: Mammograph [3]

morphologic concentric layer (MCL) criteria, to detect mass region in mammogram [5]. Gradient vector flow (GVF)snake and multi-scale analysis using Gaussian pyramid has been proposed in[6] to segment masses in mammogram. At first they applied gaussian pyramid to make the image coarse, so that GVF snake is able to converge to the mass contour easily and quickly with less computation. Shape features like elongatedness, eccentricity, Euler number, Max Radius, Min Radius were used to distinguish four different shapes round, oval, lobular, irregular of mass by using C5.0 decision tree algorithm in [7]. Gabor filter banks are used for extracting local spatial textural properties of masses at different orientations and scales[8]. Multilevel wavelet decomposition method is proposed to extract mean, variance, standard deviation, entropy and mean of absolute deviation from wavelet components [9]. Boundary extraction of this mass is also very important, so that radiologists can judge whether the mass is benign or cancer. Rangayan et al in [10] proposed a region-based measure of image edge profile acutance by polygonal approximation and measured shape features like compactness, Fourier descriptors, central invariant moments and chord-length statistics to distinguish between circumscribed and spiculated tumors.In this article, the method to detect breast cancer is presented using k-means clustering algorithm.

2. Methodology

The input mammogram image are taken from the National Cancer Hospital and from mammography database website[11]. Homomorphic filtering is used to remove multiplicative noise in the mammogram image. A common technique for contrast enhancement is the combined use of the top-hat and bottom-hat transforms.Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. Histogram equalization usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values.K-means is an algorithm to group objects into a K number of clusters based on features, where K is a positive integer number. The image segmentation of mass region is done by k-means algorithm and is one of the simplest unsupervised learning algorithms that classify a given data set through a certain number of clusters.Different features are extracted from tumor segemented image [12].

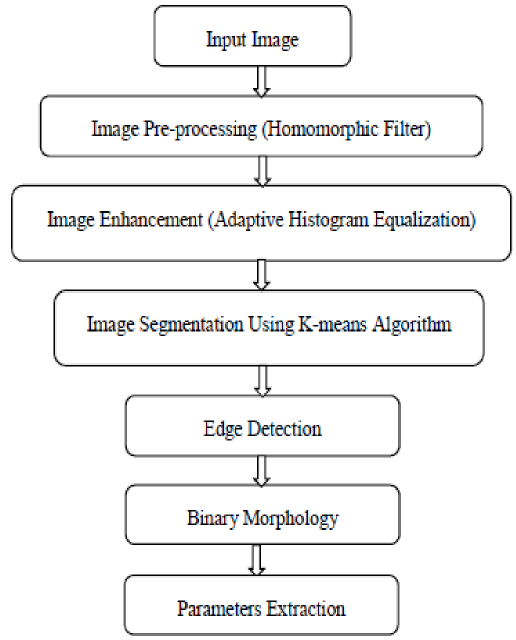


Figure 2: methodoogy

2.1 Homo-morphic Filtering

Homo-morphic filtering is a process of image pre-processing tehnique and is most commonly used for correcting non-uniform illumination as well as remove the multiplicative noise in images [13].

$$I(x,y) = L(x,y)R(x,y)$$

$$ln(I(x,y)) = ln(L(x,y)) + ln(R(x,y)) \tag{1}$$

where I is the image, L is scene illumination, and R is the scene reflectance.

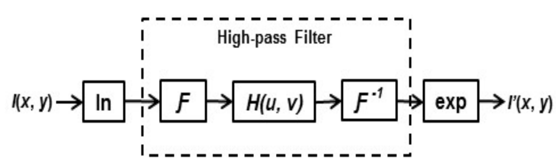


Figure 3: Homo-morphic Filtering

2.1.1 Fast Fourier Transform

Fast Fourier Transform is applied to convert an image from the image (spatial) domain to the frequency domain. Applying filters to images in frequency domain is computationally faster than to do the same in the image domain[14].The mathematical formula of Fourier transform and Inverse Fourier Transform of performing the 2D transform in the frequency space can be expressed :

$$F(x,y) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)e^{-j2\pi(x\frac{m}{M}+y\frac{n}{N})} \tag{2}$$

$$f(m,n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} F(x,y)e^{j2\pi(x\frac{m}{M}+y\frac{n}{N})} \tag{3}$$

Where f(m,n) is the pixel at coordinates (m, n), F(x,y) is the value of the image in the frequency domain corresponding to the coordinates x and y. M and N are the dimensions of the image.

2.2 Image pre-processing and Image Enhancement

Following are the steps in Image pre-processing and Image Enhancement process:

Step1:Apply homomorphic filter to compress brightness range and enhance contrast of image as shown in Figure 3. It removes non-uniform illumination without any loss. Let the output be G. f(x, y) is an input image. Z(x, y) is the output after log transformation. Z (u, v) is the output of Fourier transform. H (u, v) is a transfer function of frequency domain filter. H' (u ,v) is output of the Z(u, v) filtered with H(u,v).

Step2: Tophat transform is applied to G using disk of radius 15 as a structuring element. Shape and size of structuring element is selected based on the shape and size of the masses. It can be used to separate the objects .Let the output be thf

Step 3: Dilation operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels . It is applied to smooth the borders of tophat transformed image. Let the output be thf1.

Step4: Bothat transform is applied to the original image to smooth the objects in original image. Let the output be bhf.

Step 5: These images are combined using Image arthimetic addition and subtraction.

$$Enhancedimage = (G+thf1)–(bhf) \tag{4}$$

Step6: Adaptive histogram equalization technique is applied to improve local contrast. Adaptive method computes several histograms on small tiles of image and improves local contrast giving more details.

2.3 K-Means Algorithm

K-Means algorithm is to group objects into a K number of clusters based on features, where K is a positive integer number. We consider the input as image pixels and their features are their grey-level values. The algorithm aims at minimizing sum of any pixel to cluster centroid distances, we have chosen Euclidean distances as distance measure.This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function:

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n} |x_i^j - c_j|^2 \tag{5}$$

where $|x_i^j - c_j|^2$ is a chosen distance measure between a data point x_i^j and the cluster center c_j , is an indicator of the distance of the n data points from their respective cluster centers.

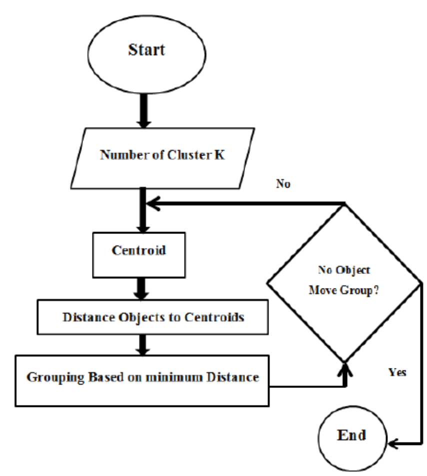


Figure 4: Flowchart of K-Means Algorithm

2.4 Cancer cell Features Extraction

The figure 5 are of the normal cell and cancer cell. In figure 5 you can clearly see that normal cells have large cytoplasm, single nucleus, single nucleolus, fine chromatin and smooth cell border whereas cancer cells have small cytoplasm, multiple nuclei, multiple and large nucleoli, coarse chromatin and irregular cell border.

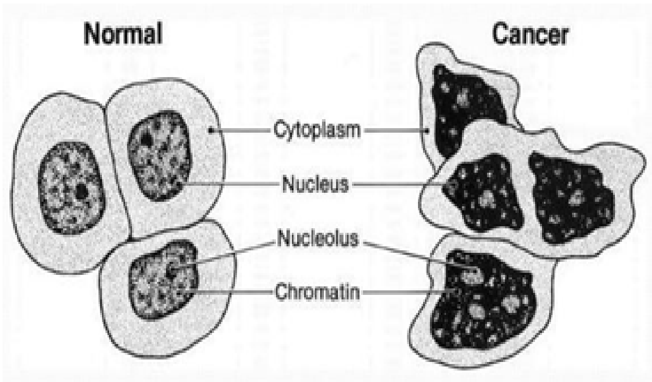


Figure 5: Cancer cell and Normal cell [15]

2.5 Binary Morphology

Combination of Dilation, Erosion and Image subtraction gives morphological gradient.Let, f is an input image.

$$M(x,y) = Dilation(f) - Erosion(f) \tag{6}$$

Where, M(x, y) is the morphological gradient.

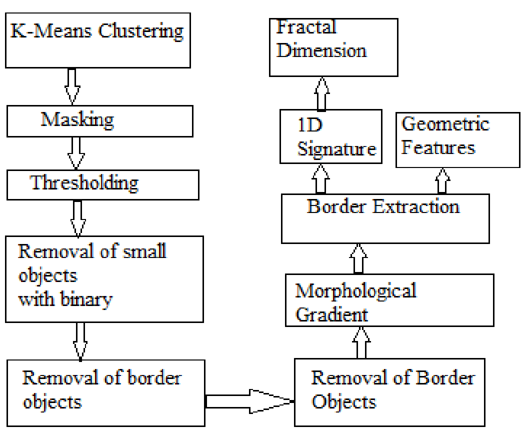


Figure 6: Steps Diagram for Extraction of Border and Features

3. Result and Discussion

Algorithms have been implemented on 25 mammograms of which nine mammograms are from national cancer hospital, jawlakhel and 16 mammograms are from DDSM database [11]. From this 10 mammograms have benign lesions and 15 mammograms have malignant lesions. We implemented image enhancement algorithm on 9 mammograms that are taken from national cancer hospital, jawlakhel, as they are high quality digital mammograms compared to DDSM mammograms.

3.1 Quality measures of image enhancement

3.1.1 Entropy

Image entropy is a quantity which measures the information of an image. It is represented by H (I)

$$H(I) = -\sum_{i=0}^{n-1} P_i ln P_i \tag{7}$$

Where Pi is probability of ith gray level intensity value n is a gray level number in the image. If the entropy is greater, the image is more clear .We observed E2 is more than E1.

3.1.2 Standard Deviation (SD)

It is a value on the gray level axis, showing the average distance of all pixels to the mean. SD of the histogram tells us about the average contrast of the image. Greater the Standard deviation, greater is the contrast of the image, std2 is greater than std1.

3.1.3 Edge based contrast measure (EBCM)

EBCM measures the intensity of edge pixels in small windows of the image.

3.2 Image pre-processing and Image Enhancement

The proposed algorithm based on hybrid approach combination of both frequency domain homomorphic filtering and spatial domain morphology and adaptive histogram equalization technique to the output of hybrid approach. Homomorphic filtering was applied to the input image to improve the contrast of image and morphological operations were applied to remove the noise and to smooth the edges of the image.

Table 1: Quality Measures of Enhanced Images in Cancer

Patient No.	E1	E2	EBCM1	EBCM2	SD1	SD2
C1(Cancer)	3.3023	4.0535	58.4866	110.0120	0.0215	0.0648
C2	3.5350	4.4489	95.6216	99.0381	0.0132	0.0363
C3	2.9220	3.3419	85.7870	95.1428	0.0130	0.0338
C4	3.5648	4.3746	111.5529	81.9789	0.0120	0.1237
C5	2.6656	3.0532	93.0772	90.4355	0.0118	0.0443
C6	3.4680	4.2274	90.3118	88.2642	0.0121	0.0369
C7	2.4541	2.8501	60.1479	92.1521	0.0148	0.0549
C8	2.7730	3.0913	97.9758	90.1452	0.0074	0.0252
C9	1.7965	2.1380	45.2919	87.588	0.0168	0.0570
C10	2.7979	3.2509	120.8388	86.9146	0.0076	0.0269
C11	2.6002	3.1353	66.6428	90.4510	0.0141	0.0554
C12	2.0944	2.5337	55.0490	88.4676	0.161	0.0580
C13	2.7222	3.1991	70.5318	92.8534	0.0139	0.0478
C14	2.9045	3.5197	74.0608	90.8344	0.0175	0.0550
C15	2.2406	2.6668	55.8464	87.9221	0.0158	0.0578

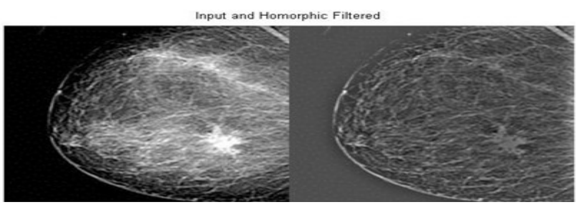


Figure 7: (a) Input (b) Homorphic Filtered Image

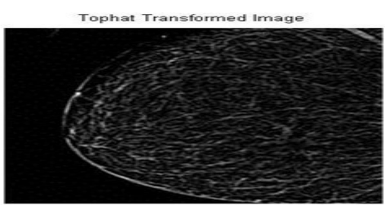


Figure 8: Tophat Transform Image

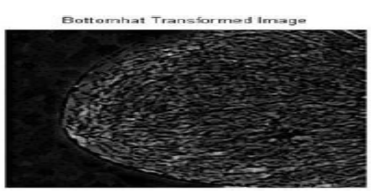


Figure 9: Bottomhat Transform Image

3.3 Image Segmentation

Thresholding is a process of binarization of an image. Binary thresholding converts the gray scale image I to a binary image. The output image replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels

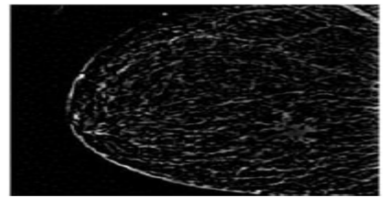


Figure 10: Enhanced Image

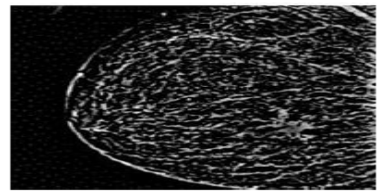


Figure 11: Enhanced Image using Adaptive Histrogram equalization

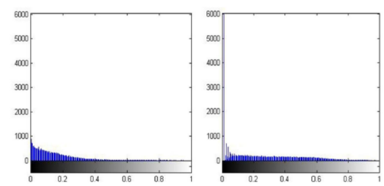


Figure 12: Histrogram equalization

with the value 0 (black), specify level in the range [0, 1]. Therefore, a level value of 0.5 is midway between black and white.

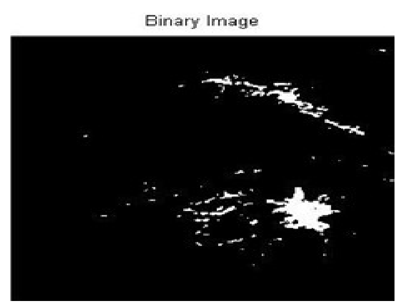


Figure 13: Binary Image Segmentation using threshold

3.4 K-Means clustering

K-Means clustering algorithm and morphological operators were used to segment mass and extract the border. The procedure of image segmentation consists of :

Step1: K-means Clustering
Step2: Morphological operations

Step 3: Morphological gradient

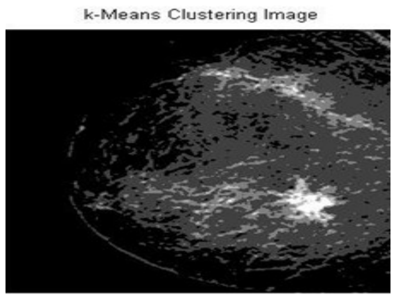


Figure 14: Image Segmentation by K-Means Clustering Algorithm

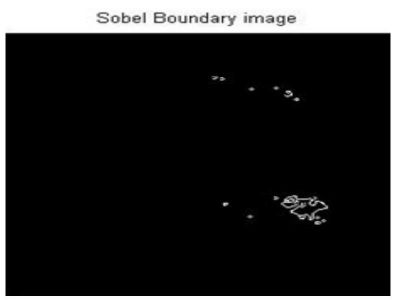


Figure 15: Edge Detection using Sobel Method

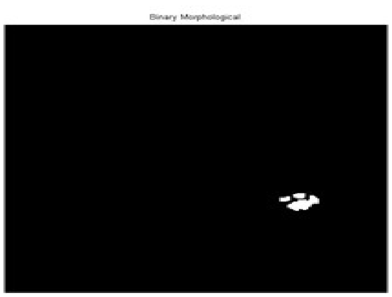


Figure 16: Binary Morphological Image

3.5 Geometric Features Extraction

Geometric feature learning is a technique combining machine learning and computer vision to solve visual tasks. The main goal of this method was to find a set of representative features of geometric form to represent an object by collecting geometric features from images and learning them using efficient machine learning methods. Feature plays a very important role in the area of image processing.

a) Area: Number of pixels contained in the lesion. Greater the value of area, it is more likely the lesion is malignant.

b) Perimeter: The distance around the boundary of the region. Regionprops computes the perimeter by calculating the distance between each adjoining pair of pixels around the border of the region. Perimeter is the circumference of Lesion.

c) PA-ratio : It is the ratio of perimeter to area of the lesion.

d) L: S Ratio: It is the length ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the lesion. If L: S ratio is more, it is likely the lesion is malignant.

e) ENC (Elliptical normalized circumference): Anfractuosity is common morphological feature for malignant contour. ENC is circumference ratio of the lesion and its equivalent ellipse. Anfractuosity of a lesion contour is characterized by ENC.

Table 2: Features Extraction of Cancer

Image	Area	Perimeter	PA Ratio	Major Axis	Minor Axis	LS Ratio	ENC
C1	245	17.6619	0.3474	29.122	119.801	1.4707	55.486
C2	1428	42.6402	0.2524	256.3406	69.6153	3.6822	133.95
C3	348	21.0496	0.3728	34.2217	24.8925	1.3748	66.129
C4	1741	47.082	0.0169	168.8417	119.563	1.4122	147.91
C5	409	22.8201	0.3760	51.2765	36.0112	1.4239	71.691
C6	1280	40.3701	0.2683	296.8708	94.355	3.1463	126.82
C7	1022	36.0729	0.2868	152.2419	74.3404	2.0479	113.32
C8	336	20.6835	0.4104	51.5467	34.4832	1.4948	64.979
C9	136	13.1590	0.4614	30.1860	11.8856	2.5397	41.340
C10	316	20.0585	0.4207	65.3548	41.1373	1.5445	63.015
C11	1420	42.5206	0.3140	149.9097	89.5525	1.6740	133.58
C12	1722	46.8243	0.3338	135.3552	81.9427	1.6518	147.10
C13	1054	36.6332	0.3131	116.4685	73.6592	1.5812	115.08
C14	2885	60.6077	0.2765	236.2648	77.1992	3.0605	190.40
C15	1762	47.3651	0.3196	143.0897	98.2015	1.4571	148.80

3.6 Signature

A signature is a 1D functional representation of a boundary. It is a plot of the distance from the centroid to the boundary as a function of angle. A signature is a 1-D representation of a boundary (which is a 2-D thing): it should be easier to describe. For example distance from the centroid vs. angle. Signatures are invariant to translation. Signatures are invariance to rotation and depends on the starting point .the starting point could be the one farthest from the centroid. Scaling varies the amplitude of the signature and invariance can be obtained by normalizing between 0 and 1.

Figure 20 and Figure 21 are the extracted boundary of

Table 3: Features Extraction of Benign

Image	Area	Perimeter	PA Ratio	Major Axis	Minor Axis	LS Ratio	ENC
B1	168	14.6255	0.3857	368.5450	17.2290	21.3909	45.9473
B2	548	26.4147	0.3587	78.0633	76.6882	1.0179	82.9842
B3	319	20.1535	0.4292	58.9635	50.9702	1.1568	63.3141
B4	608	27.8232	0.4027	118.0374	62.2037	1.8976	87.4091
B5	134	13.0619	0.4229	25.9273	15.9948	1.6210	41.0353
B6	52	8.1369	0.5318	11.2679	10.0193	1.1246	25.5627
B7	325	20.3421	0.3967	52.1927	34.8037	1.4996	63.9067
B8	319	20.1535	0.4292	58.9635	50.9702	1.1568	63.3141
B9	579	27.1515	0.0529	267.2050	55.5585	4.8094	85.2911
B10	115	12.1005	0.4449	21.9452	16.4682	1.3326	38.0149

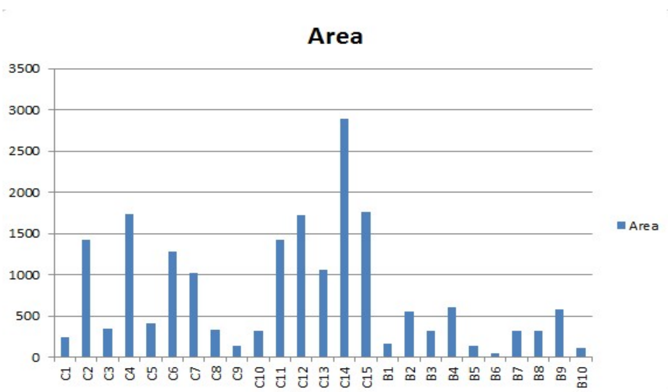


Figure 17: Comparative Analysis of Area of Cancer and Benign

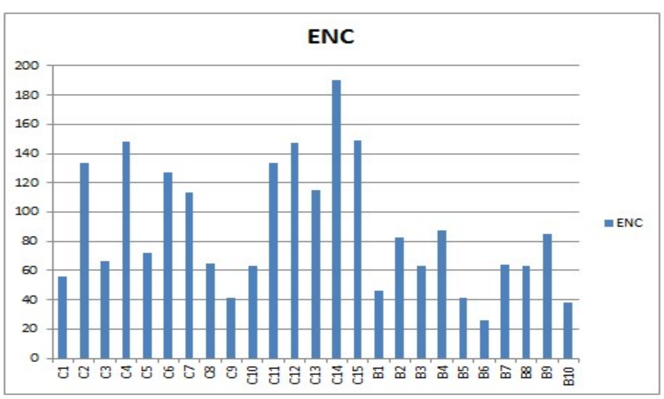


Figure 18: Comparative Analysis of ENC of Cancer and Benign

cancer and benign. Figure 20 shows that the extracted boundary of the cancer is speculated or ill-defined

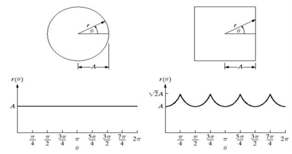


Figure 19: Distance versus Angle Signature

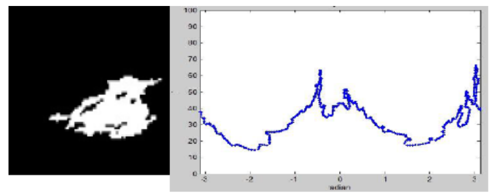


Figure 20: Extracted Boundary of Cancer and its signature

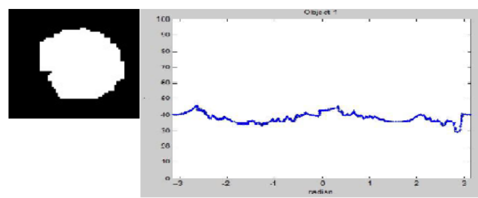
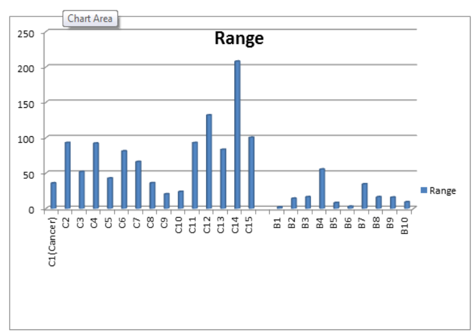


Figure 21: Extracted Boundary of Benign and its signature



Comparative Analysis of Signature Radius (R) in Benign and Cancer

Figure 22: Comparative Analysis of range of Radius (R) in Benign and Cancer

boundary whereas Figure 21 which is the extracted boundary of the benign is smooth boundary. From 22, it is shown that the range value of R in cancer mammogram have higher value in comparison to benign. Figure 20 and 21 are the distance versus angle in signature, shown that the variation of R is high in terms of angle. The shape of the contour or boundary to delineate malignant and benign lesions as malignant lesions have speculated or ill-defined boundary and benign mass have smooth boundary.

4. Conclusion

Automatic detection of boundary helps the doctors in analyzing the lesion in less time and prevents unnecessary biopsies. The shape of the contour or boundary to delineate malignant and benign lesions as malignant lesions have speculated or ill-defined boundary and benign mass have smooth boundary. In this paper Mammogram image is enhanced using homomorphic filtering and adaptive histogram equalization. The enhanced mammogram image is segmented using K means clustering and extracted geometric features from the lesions. Geometric features of the border are also calculated. Geometric features of Lesion boundary are characterized as malignant or benign lesion. Seven morphologic features are extracted from each lesion to describe features such as shape, contour, and size. Classifications of malignant and benign are done by distance versus angle of signature. Image enhancement and segmentation methods are implemented to extract the border and distance versus angle of signatures is calculated. Signature value of range in malignant image is higher in comparison to benign image.

5. Limitations and Future Enhancement

In future we would like to develop algorithms for classification of cancer and noncancer patients by considering different types of abnormalities like Micro calcifications, Architectural distortion, Lesions, Bilateral Asymmetry in mammogram. A pattern classification step, based on fractal analysis, support vector machine and Bayes linear classifier can be implemented to have accuracy of the lesions malignancy assessment procedure. In future, the mass obtained from the mammogram will be realized in 3D and suitable modification will be carried out with the proposed shape and margin properties. Limitations that are present in this paper can be removed by increasing parameter and region of interest.

6. Acknowledgement

We would like to express our deepest appreciation to all those who provided us the possibility to complete this paper work. Furthermore, We would like to acknowledge with much appreciation the crucial role of, National Cancer Hospital. We express our gratitude for providing us with all the mammogram images, valuable information for the thesis. Finally, We would like thank our families and friends who always encouraged and supported us.

References

[1] Cancer data and its factsheets. http://www.who.int/mediacentre/factsheets/fs297/en/. Accessed: 2016-06-15.

[2] Mammography. https://www.radiologyinfo.org/en/info.cfm?pg=mammo. Accessed: 2016-06-15.

[3] Mammography machine. https://www.acrin.org/patients/aboutimagingexamsandagents/aboutmammographyandtomosynthesis.aspx. Accessed: 2016-06-15.

[4] Arianna Mencattini, Marcello Salmeri, Paola Casti, Grazia Raguso, Samuela L'Abbate, Loredana Chieppa, Antonietta Ancona, Fabio Mangieri, and Maria Luisa Pepe. Automatic breast masses boundary extraction in digital mammography using spatial fuzzy c-means clustering and active contour models. In *Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on*, pages 632–637. IEEE, 2011.

[5] Nevine H Eltonsy, Georgia D Tourassi, and Adel S Elmaghraby. A concentric morphology model for the detection of masses in mammography. *IEEE transactions on medical imaging*, 26(6):880–889, 2007.

[6] Hongwei Yu, Lihua Li, Weidong Xu, Wei Liu, Juan Zhang, and Guoliang Shao. Gaussian pyramid based multi-scale gvf snake for mass segmentation in digitized mammograms. In *Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on*, pages 1–4. IEEE, 2009.

[7] A Vadivel and B Surendiran. A fuzzy rule-based approach for characterization of mammogram masses into bi-rads shape categories. *Computers in biology and medicine*, 43(4):259–267, 2013.

[8] Muhammad Hussain, Salabat Khan, Ghulam Muhammad, Mohamed Berbar, and George Bebis. Mass detection in digital mammograms using gabor filter bank. 2012.

[9] M Vidhya, N Sangeetha, MN Vimalkumar, and K Helenprabha. Early stage detection of cancer

in mammogram using statistical feature extraction. In *Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEeCE), 2011 International Conference on*, pages 401–404. IEEE, 2011.

[10] Rangaraj M Rangayyan, Nema M El-Faramawy, JE Leo Desautels, and Onsy Abdel Alim. Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on medical imaging*, 16(6):799–810, 1997.

[11] Usf digital mammography home page. http://marathon.csee.usf.edu/Mammography/Database.html. Accessed: 2016-06-15.

[12] Spandana Paramkusham, KMM Rao, and BVVSN Prabhakar Rao. Automatic detection of breast lesion contour and analysis using fractals through spectral methods. In *Advances in Computer Science, AETACS*, 2013.

[13] Chaofu Zhang, Li-ni Ma, and Lu-na Jing. Mixed frequency domain and spatial of enhancement algorithm for infrared image. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on*, pages 2706–2710. IEEE, 2012.

[14] Rahil Garnavi, Mohammad Aldeen, and James Bailey. Computer-aided diagnosis of melanoma using border- and wavelet-based texture analysis. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1239–1252, 2012.

[15] Normal and cancer cell. https://visualsonline.cancer.gov/details.cfm?imageid=2512. Accessed: 2016-06-15.