A Novel Approach for Automatic Detection of Abnormalities in Mammograms

B.N.Beena Ullala Mata[1] Dr. M. Meenakshi [2],
[1] Sl.Gr.Lecturer, Department Of Medical Electronics
B.M.S College of Engineering, Bangalore-560019, India
E-mail:bansbeena@gmail.com
[2] Professor, Department of Instrumentation Technology
Dr.Ambedker Institute of Technology, Bangalore-560056
E-mail:meenakshi_mbhat@yahoo.com

***Abstract*— This paper proposes a novel approach for the development of a computer aided decision system to automatically detect abnormalities in mammograms. In this method preprocessing of images is done by enhancing the contrast of the intensity image by transforming the values using contrast-limited adaptive histogram equalization (CLAHE). Then, Mathematical morphology is used for the extraction of abnormalities which are located on a non-uniform background. After performing the thresholding of the images by the extended maxima transformation, feature extraction is focused on the extraction of both statistical and textural features of the objects. Finally the extracted objects are classified using Naïve Bayes Classifier and abnormalities are detected. This forms a basic step in the automatic detection system of abnormality in breast images and thus increases in the sensitivity of breast cancer detecting algorithms. The accuracy of the method has been verified with the ground truth given in the data base and it is as high as 82.40%. The algorithm is validated by considering the mini – MIAS's data base, a benchmark data supplied by American Society of Radiology UK. With the help of this data base, the feasibility of the proposed method is demonstrated.**

***Index words*— Computer Vision, Mammography, Segmentation, Neive Bayes Classifier**

1. INTRODUCTION

Among the various types of human cancers, Breast cancer is the most common form, among the women. However, complete curing of the said disease is possible if it is detected in its early stage. Early detection of Breast cancer improves survival rate of women by 95% [1 - 2]. Mammography is currently the method of choice for early detection of breast cancer in women. However, the interpretation of mammograms is largely based on the radiologist's opinion. In this study, an attempt is made to develop a computer aided decision system using an Image processing algorithm for the detection of abnormalities [3-5].
This proposed method deals with a novel approach for the development of a computer aided decision system to automatically detect abnormalities in mammograms.

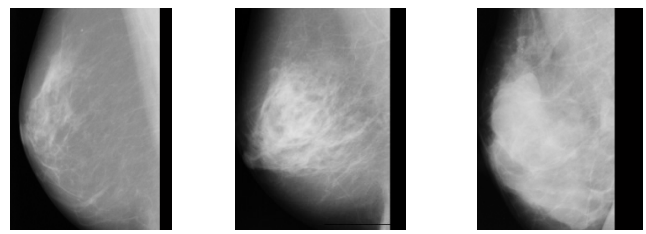


Figure. 1. Mammograms showing different breast densities (a) Fatty (b) Fatty Glandular (c) Dense Glandular

The method involves simple pre-processing techniques. It enhances the contrast of the intensity image by transforming the values using contrast-limited adaptive histogram equalization (CLAHE) [8]. After Morphological operations, every mammogram is threshold using the extended maxima transformation. The extended maxima transformation can be obtained by simply thresholding the h-dome image for values greater than zero. As a result, we get a binary image. A connected-component labeling operation [9] is performed, in order to evaluate the characteristics and the location of every object. Objects are selected based on equivalent diameter. The binary image is used only for the extraction of the exact location of every object. Using this, automatic Region of Interest is obtained and the original image is used for feature extraction, with the binary image as mask for every object.

Feature Extraction is focused on the extraction of both statistical and textural features of the objects. The highly textured regions of breast tissue in mammograms dictate the selection of such methods that are successful in dealing with texture regions. Finally the extracted objects are classified using NaïveBayes Classifier and abnormalities are detected [10-11].
The organization of this paper is as follows.
Section II explains the Principles of implementation of algorithm as architecture. The intermediate results of processing of the test data is also highlighted in this section II. Next, section III gives the final results and analysis of the system. Finally conclusions are drawn in Section IV.

II PRINCIPLES OF IMPLEMENTATION

A. Architecture

A block diagram describing the method used is shown in figure 2. The method consists of two stages namely, training and testing.

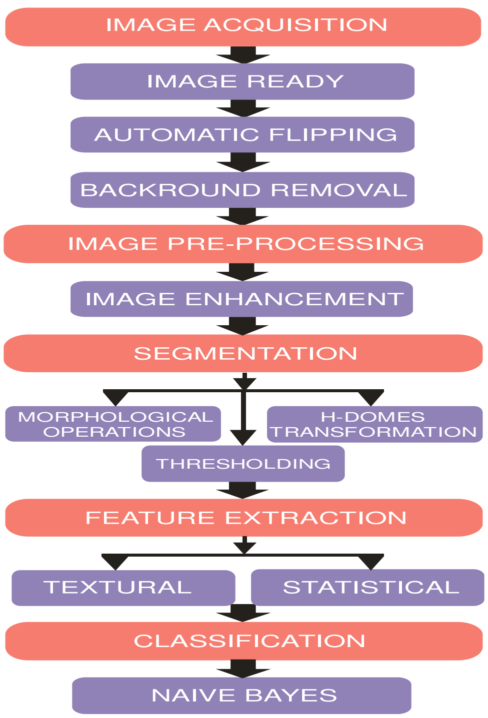


Figure. 2; Overview of method used

B. Pre-processing

The MIAS (Mammography Image Analysis Society) [6-7] database of mammograms which have been mentioned in the ground truth information given by radiologist is shown as [Figure. 1]. The database contains left and right breast images for 161 patients, and is available on a DAT-DDS tape. Its quantity consists of 322 images, which belong to three types such as Normal, benign and malignant. There are 208 normal, 63 benign, and 51 malignant (abnormal) images. It also includes radiologist's 'truth'-markings on the locations of any abnormalities that may be present. For each film, experienced radiologists have given the type, location's scale, and other useful information about them. Selected images in .pgm format are taken from the MIAS Database.

Each Image is automatically flipped to the left side in order to remove the background objects as shown in figure 3(a) and 3(b). Background objects are removed automatically as given in figure 3(c).

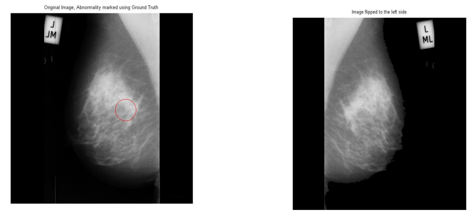
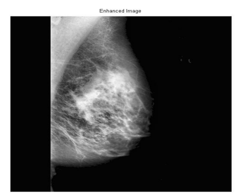


Figure 3; (a) Original image (b) Flipped image



(c) Background removed and enhanced.

Images are pre-processed for further analysis. Images are then enhanced. The contrast of the intensity image is enhanced by transforming the values using contrast-limited adaptive histogram equalization (CLAHE). An efficient spot-adaptive segmentation technique was developed by suitably combining in a cascade mode with the benefits of image enhancement .CLAHE was applied on each individual image-cell, initial SRG-seed was set at the image-cell's centre, and threshold was estimated from the image-cell Background. CLAHE operates on small regions in the image, called tiles, rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the 'Distribution' parameter. The neighbouring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.

C. Segmentation

One of the most recurrent prerequisites of an image processing system is the ability to analyze images and detect regions that have specific characteristics. Detection means extracting the clustered abnormalities from the local background breast tissue. This process is known as segmentation. The main principle behind thresholding is that image pixels, falling into a predefined range of intensity values, are assigned one single intensity value, and the rest of the pixels are assigned a different intensity value. A thresholding function can be formally defined as,

$$g(x,y) = \begin{cases} i & if \quad f(x,y) \in [L_l, L_u] \\ j & if \quad f(x,y) \notin [L_l, L_u] \end{cases}$$

Where g (.) is the threshold version of image f (.), i and j are the two intensities used to differentiate between two groups of pixels, and Ll and Lu are lower and upper limits of the intensity range used to define the two groups. The basic principle in thresholding makes it highly suitable for segmenting ROIs that will always have distinct intensities from the rest of the image.

Thresholding, a simple segmentation method is used to create binary images. Morphological operations in digital image processing are a way of extracting image components that can be used to express details about a regions shape, its boundaries, and its area and so on. Mathematical morphology can be defined as a theory for the analysis of spatial structures. It is called morphology because it aims at analyzing the shape and form of objects.

The basic tools of mathematical morphology are the morphological operations. A morphological operation P transforms an image A by means of a structuring element B (which can be chosen by the user) into a new image P (A; B). The basic morphological operations are dilation and erosion. Using the basic morphological operators we can design powerful morphological filters. The basic morphological filters are the morphological opening and the morphological closing. Two primitive and most widely used morphological operations are explained:

Dilation: Dilation is usually used to smooth boundaries of regions or bridge very small gaps between neighboring regions According to Gonzalez and Woods [12], the formal definition of dilation of a set A by another set B is denoted A © B and defined as,

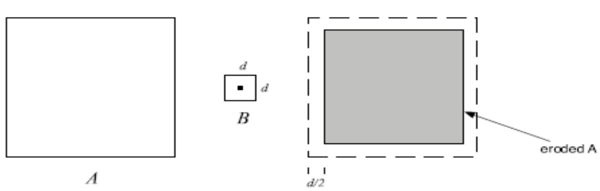
$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\}$$

Where B^ is the reflection of B. This definition means that dilation of A by B will is done by reflection B and then shifting B over A by z.

Erosion: Erosion has an opposite effect of dilation. Following the same notation a formal definition of erosion is,

$$A \ominus B = \{z | (B)_z \subseteq A\}$$

In other words, erosion of A by B is set of all points traversed by center of B such that B is totally contained within A at all times.



Erosion can be used for removing small unwanted components, such as thread like structures, from an image by using a structuring element that is bigger than the unwanted regions. The processes of dilation and erosion can be combined in different ways to make some interesting changes in an image. Morphological opening and closing are two such operations that are defined by specific combinations of dilation and erosion. Opening is generally used to smooth region contours in an image, and it also removes thin protrusions. Closing also adds smoothness to image contours; however it generally fuses two large regions separated by narrow breaks. This effect is opposite to that caused by morphological opening that breaks the narrow links between two large regions.

In morphological image analysis every gray tone image is considered as a topographic relief, where each pixel is associated with an elevation proportional to its intensity. Therefore, many morphological terms stem from geomorphology. So, the dark and light structures of the image correspond to the valleys and the domes of this relief. The plateau located at the top of the domes constitutes regional maxima (denotes maximum). Abnormalities appear on digitized mammograms as bright spots and it is as shown in figure 5 (a). These spots are small regions with higher intensity values than their surroundings. Each Abnormality constitutes regional maxima. Figure 4 a) shows a cluster of abnormalities and figure 4 b) shows the topographic representation of this region, where abnormalities appear as domes with higher intensity values than the surrounding tissue.

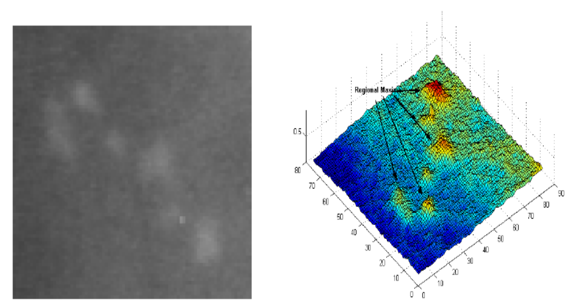


Figure 4 a) A cluster of Abnormalities   Figure 4 b) Topographic relief

A simple algorithm to extract all regional maximum in mammogram is based on the H-domes transform [Vincent (1993)]. The h-dome transformation is illustrated on Fig. 5(b). The h-dome transformation extracts light structures without involving any size or shape criterion.

The h-domes transformation extracts image regions fulfilling the following criteria:

- Every pixel in the dome has a gray value greater than any of the pixels surrounding it.

- The maximum gray level difference between two pixels in the dome is smaller than or equal to h.

The h-domes transformation can be defined by:

$$M_h(I) = I - \rho_I(I - h)$$

Where I is the original image that represents the result of subtracting a constant value h to the original image and M_h (I) is the morphological reconstruction of the original image. The choice of h turns out not to be a critical operation, since a wide range of values yields correct results.

A study about abnormalities and their imaging properties showed that region offset average i.e. the difference between average intensity values of every calcification and their surrounding tissue, were similar for all calcifications and only few statistically significant differences were found between benign and malignant offsets. Indeed, choosing a threshold greater than 20 intensity values, all abnormalities that were used at the testing phase of the algorithm, were extracted.

The new image contains all the abnormalities and many more elevations, which doesn't correspond to calcifications, since mammograms are complex images with great number of regional maxima. In order to suppress noise and to reduce the number of the extracted domes, the image is opened using a disk shaped structuring element of radius of two.

D Feature extraction

Then in step, the image is used for the feature extraction Feature Extraction is focused on the extraction of both statistical and textural features of the objects. The highly textured regions of breast tissue in mammograms dictate the selection of such methods that are successful in dealing with texture regions.

Textural features:

Texture features have been widely used as classification of masses in digital mammogram. Texture features can be divided into three classes based on what they are derived; from gray level co-occurrence matrices (GLCM) - based features, gray level difference statistics (GLDS)-based features. The texture of the region surrounding the mass can yield important features for its classification, thus some study have used GLDS or RBST (Rubber Band Straightening Transform) that maps a band of pixels surrounding the mass onto the Cartesian plane.

Several previous studies have calculated RLS and used that to extract the texture information in the region of interest. Besides, most researchers used GLCM or SGLD (spatial gray level dependence) matrices to measure the texture-context information in the mammogram.

Even though, most of the researchers used GLCM to extract the features, different set of features have been used for classification of masses. This is extracted using GLCM for further classification of masses. The four features used in this work are contrast, mean, standard deviation, entropy and correlation. They have been computed in the four directions, which are 0º, 45º, 90º and 135º, thus resulting in a total of 16 texture features for each ROI.

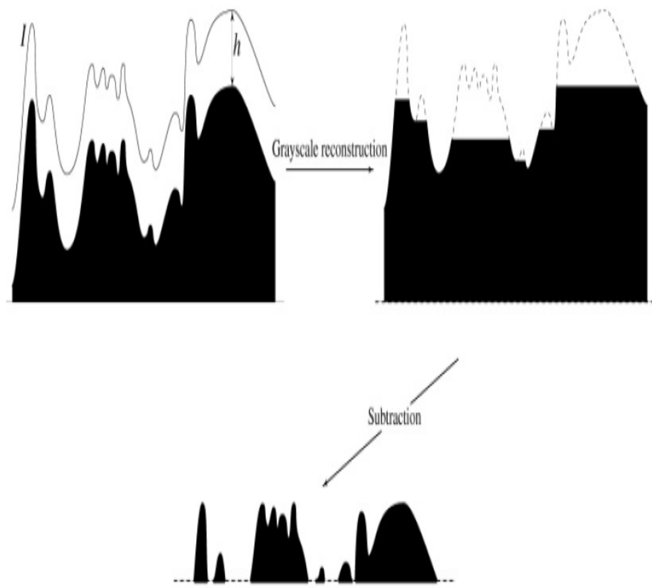


Figure 3: H-domes transformation

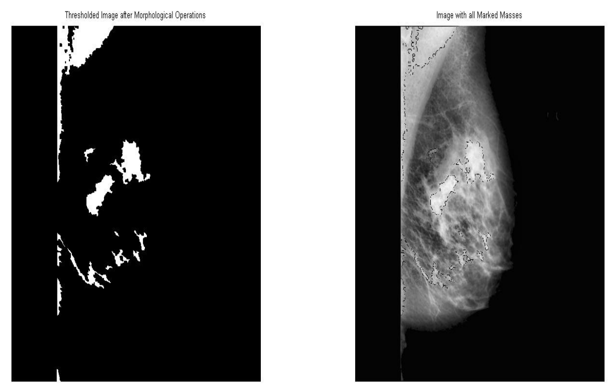


Figure 5 (a); H-dome transformation
Figure 5 (b); Threshold and overlaid images

Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}.$$

Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of p_x and p_y.

Entropy:

$$f_9 = -\sum_i \sum_j p(i,j) \log (p(i,j)).$$

.

Statistical features:

The most famous statistical approach is the co-occurrence matrix. This was the result of the first approach to describe, and then classify, image texture. It remains popular today, by virtue of good performance. The co-occurrence matrix contains elements that are counts of the number of pixel pairs for specific brightness levels, when separated by some distance and at some relative inclination.

For brightness levels b1 and b2 the co- occurrence matrix C

$$C_{b1,b2} = \sum_{x=1}^{N} \sum_{y=1}^{N} (P_{x,y} = b1) \wedge (P_{x',y'} = b2)$$

is,

Where the x co-ordinate x′ is the offset given by the specified distance d and inclination θ

by,

$$x' = x + d\cos(\theta) \quad \forall \left(d \in 1, \max(d)\right) \wedge \left(\theta \in 0, 2\pi\right)$$

and the y co-ordinate y′ is,

$$y' = y + d\sin(\theta) \quad \forall \left(d \in 1, \max(d)\right) \wedge \left(\theta \in 0, 2\pi\right)$$

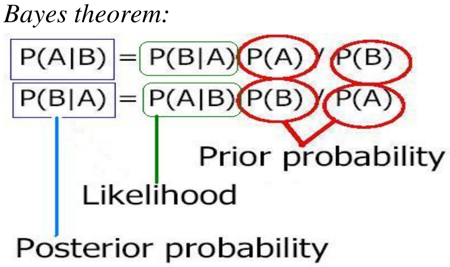
2.4 Classification using Naïve Bayes Classifier

Naive Bayes is one of the simplest density estimation methods from which we can form one of the standard classification methods in machine learning.

Its fame is partly due to the following properties

- Very easy to program and intuitive.
- Fast to train and to use as a classifier.
- Very easy to deal with missing attribute.

Bayes theorem:



E. Performance Evaluation and Comparison

One approach to assess the performance of the detection scheme is by constructing a receiver operating characteristic (ROC) curve which represents the false positive (FP) rates against the true positive (TP) rates obtained by varying a predefined control parameter such as K. Furthermore, other studies have used other metrics such as specificity and sensitivity parameters. Specificity and sensitivity parameters are defined in terms of TP, TN, and FP as follows:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Where a true positive (TP) rate represents the probability of classifying a malignant tissue as target object while false positive (FP) refers to the probability of classifying healthy tissue as target one. Also, true negative (TN) rate is defined as the probability of classifying a healthy tissue as none-malignant and a false negative (FN) represents the case when a malignant tissue being classified as a healthy one.

III. RESULTS

Finally, results of the classification indicating TP, FP, TN and FN for Sensitivity and specificity parameter calculations are as shown in Table-1. The performance of the classification is evaluated by Receiver operating curve which is as shown in figure 6. The original image marked abnormality masses using ground truth and final image indicating detected abnormality masses after Classification results are shown in figure 7(a) and 7(b) respectively.

Table-1; Results of Classification with TP, FP, TN and FN calculations.

		CLASSIFICATION RESULTS				
TESTING IMAGE	TP	FP	TN	FN	SENSITIVITY(%)	SPECIFICITY(%)
IMAGE 2	1	1	4	0	100	80
IMAGE 3	1	1	14	0	100	93.3333
IMAGE 4	0	2	13	1	0	86.6667
IMAGE 6	1	1	12	0	100	92.3077
IMAGE 11	1	1	7	0	100	87.5
IMAGE 13	0	0	8	0	NaN	100
IMAGE 21	1	3	9	0	100	75
IMAGE 22	0	3	7	1	0	70
IMAGE 23	1	4	10	0	100	71.4286
IMAGE 24	1	1	4	0	100	80
IMAGE 25	1	3	7	0	100	70
TOTALS	8	20	95	2	80	82.38511818

True Positive TP = 8
False Positive FP = 20
True Negative TN = 95
False Negative FN = 2
Sensitivity Percentage SNP= 80
Specificity Percentage SPP = 82.6087
Percentage Correct Classification=82.400000%
Percentage Incorrect Classification=17.600000%

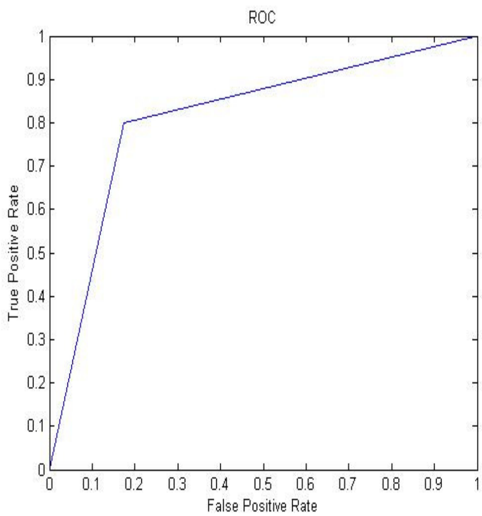


Figure 6; Performance Curve

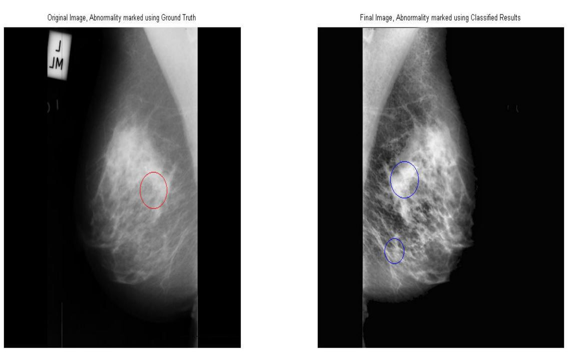


Figure7 (a); Mammogram Image Marked With Abnormality Masses Using Ground Truth
(b); Mammogram Image Marked With Abnormality Masses Using Classified Results

IV CONCLUSION

Breast cancer is the most common cancer and a significant cause of this disease still remains unknown. It is believed that the most promising way to decrease the number of patients suffering from the disease is by early detection. Early the cancer detection, the better the chances that treatment will work. The radiologist’s performance increases when they incorporate automatic image analysis in their decision making process for both the detection and diagnosis of cancer. Thus, a development of “Automatic Detection of Abnormalities in Mammograms” is highly desirable in order to assist radiologists in interpretation of specific abnormalities and to improve the diagnostic accuracy in making diagnostic decisions. It is well known that mammogram interpretation is a very difficult task even for experienced radiologists.

Mathematical morphology proves to be a useful tool for the detection of abnormalities in digital mammograms. We proposed a new algorithm for the detection of abnormalities on mammograms. Every suspicious object is marked using a binary image, which is used as a mask for object extraction from the original image. The features of the extracted objects are classified using Naïve Bayes classifier. Our feature extraction method and the selected Naïve Bays classifiers need to be tested using a larger database in order to perform reliably in clinical situations.

Thus, the main objective of this work to discuss the automatic detection of abnormalities in mammograms has been proposed, designed and developed.

REFERENCES

[1] R H. Zonderland: "BI-RADS - Introduction to the Breast Imaging Reporting and Data System", Available at: www.radiologyassistant.nl segmented Image
[2] Imaginis Corporation: www.imaginis.com
[3] E.D. Pisano, et al.: "Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening", New England Journal of Medicine, Vol. 353, No. 17, 2005, pp. 1773-1783
[4] R.M. Rangayyan: Biomedical Image Analysis, CRC Press LLC, Boca Raton, Florida, USA, 2005
[5] J.S. Suri, R.M. Rangayyan: Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer, Bellingham, Washington, USA, 2006
[6] The MIAS database of mammograms.
[7] American College of Radiology (ACR): ACR Breast Imaging Reporting and Data System, Breast Imaging Atlas, 4th Edition, Reston, VA. USA, 2003
[8] M. Rangayyan, F.J. Ayres, J.E.L. Desautels: "A Review of Computer-Aided Diagnosis of Breast Cancer: Toward the Detection of Subtle Signs", Journal of the Franklin Institute, Vol. 344, Issues 3-4, 2007, pp. 312-348
[9] A.C. Bovik: Handbook of Image and Video Processing, Elsevier Academic Press,
Amsterdam, 2005
[10] A. Oliver, J. Freixenet, R. Zwiggelaar: "Automatic Classification of Breast Density", Image Processing, ICIP 2005, Vol. 2, 2005, pp. 1258-1261
[11].E.S. de Paredes: Atlas of Mammography, 3rd Edition, Lippincott Williams & Wilkins, Philadelphia, USA, 2007
[12].Rafael C.Gonzalez and Richard E.Woods, Digital Image Processing, second edition, 2002, Prentice Hall India.