Abstract

SRIVASTAVA, VAIBHAV. Performance of microcalcification detection algorithms. (Under the direction of Dr Wesley Snyder)

Breast cancer is the most common malignancy in women and is three times more common than all gynecologic malignancies put together. The incidence of breast cancer has been increasing steadily from an incidence of 1:20 in 1960 to 1:8 women today. Seventy percent of all breast cancers are found through breast self-exams. However not all lumps are detectable by touch and mammography is a low-dose X-ray examination that can detect breast cancer up to two years before it is large enough to be felt. Some patterns of microcalcifications (small white deposits of calcium) give an early indication of cancer. Their small size makes their detection difficult for the radiologist. This brings in the role of CAD (Computer Aided Diagnosis) which serve as an assistant to the radiologist. In this thesis we have investigated the performance of three state of the art CAD techniques for the detection of microcalcifications in mammograms.

First, is a wavelet based technique which applies an adaptive wavelet filter to the input mammogram. Then it calculates HOS (Higher Order Statistics) values for maxima locations that are determined from the input image by an empirical method. This is followed by determination of a threshold using a cross entropy based thresholding algorithm. The thresholded image gives the locations of microcalcifications.

Second, is a technique that pre-processes the input mammogram with a tophat morphological filter which only preserves objects that are smaller than the size of the filter used in pre-processing. This is again followed by the determination of a threshold using the same thresholding algorithm as in the first technique. The thresholded image indicates the positions of microcalcifications. We have also done an investigation of co-occurrence matrix based entropy thresholding schemes. We found that two dimensional matrix based algorithms perform better than three dimensional based algorithms. Although both fail in case of images with high dynamic range which make them unsuitable for medical imaging. However the cross entropy based method performed better than co-occurrence matrix based techniques for both low as well as high resolution images.

Third, is a technique that makes use of a high pass filter for pre-processing. Classification of a location as a microcalcification is done by a SVM (Support Vector Machine) classifier using a scheme called SEL (Successive Enhancement Learning). These techniques have been compared by the use of ROC curves and we found out that the SVM based technique gives the lowest false positives for high detection rate. However cross entropy method does gives lower false positives for detection rates lower than 65%.

PERFORMANCE OF MICROCALCIFICATION DETECTION ALGORITHMS

by

Vaibhav Srivastava

A thesis submitted to the Graduate Faculty of
Department of Electrical and Computer Engineering
in partial satisfaction of the
requirements for the Degree of
Master of Science in Electrical Engineering

Department of Electrical And Computer Engineering

Raleigh

2005

Approved by:

_____

Dr Wesley Snyder

Chair of Advisory Committee

_____ _____

Dr Edward Grant Dr David Lalush

_____

Dr Huaiyu Dai

Biography

Vaibhav Srivastava was born in the year 1979 in the holy city of Varanasi, India. He graduated from IIT (Indian Institute of Technology), Kharagpur in 2001, with a Bachelor of Technology (Honors) degree in Electrical Engineering. During the next one and half years he worked with Ittiam Systems in the garden city of Bangalore, India where he gained extensive experience in video and image compression technologies. He then moved to Max Planck Institute of Biological Cybernetics, Tuebingen, Germany for a research internship where he gained valuable experience in machine learning and computer vision. In the fall of 2003 he began his Master of Science degree program in the Department of Electrical and Computer Engineering at North Carolina State University, Raleigh, NC and currently works with Dr Wesley Snyder. His research focuses on the application of computer vision and pattern recognition methods for automatic detection of breast cancer.

Acknowledgement

In the first place I would like to thank my father, Sri Satish Chandra Srivastava, my mother Srimati Anita and sister Disha Srivastava for their constant encouragement and values that allowed me to choose this career path. Next I would like to thank my adviser Dr Wesley Snyder for his most able and compassionate guidance. It was his constant encouragement and support that enabled me to complete this work. His guidance allied with most affable personality makes it a dream to work with him. I would also like to thank the imaging group at North Carolina State University headed by Dr Wesley Snyder for their help most notably, Karthik Krish, Geoff Chang and Pedro Lee. I would also like to thank North Carolina State University administration for its excellent institutional support throughout my studies.

I would also like to say thanks to my friends who helped me grow emotionally as well as technically. The list is long: Ujjwal Sinha, Amit Gupta, Aswini Kumar Jasrotia, Prabhat Kumar, Prashant Mudgal, Vivek Goel, Sachin Talwar, Shashank Shekhar. I am also an active member of HSC (Hindu Students Council) and I would like to thank HSC for giving me cherished friends whom I would not have been able to know without it. The cultural activities and intense discussions have enhanced my understanding and personality. I would like to thank Yogananda Saripalli, Arin Chaudhuri, Manav Tandon, Mrinmoyee Sanyal, Sridevi Chalasani, Anindita Banerjee for the same.

Contents

List of Figures **vi**

List of Tables **viii**

1 Introduction **1**

1.1 What are micro-calcifications 2

1.1.1 Types of calcifications 3

1.2 Selection of algorithms for comparison 4

1.3 Synthesis of calcifications on mammograms 5

1.3.1 Organization of the thesis 6

2 Adaptive Wavelet Filtering **9**

2.1 Wavelet based representation of images 10

2.1.1 Normal subband decomposition 10

2.1.2 Adaptive subband decomposition 11

2.1.3 HOS test 12

2.1.4 Algorithm 13

2.1.5 Experimental results 14

3 Entropy thresholding **18**

3.1 Algorithms 20

3.1.1 Minimum cross entropy thresholding 20

3.1.2 Two dimensional co-occurrence matrix based 21

3.1.3 Three dimensional co-occurrence matrix based 24

3.1.4 Experiments 26

4 Support Vector Machines **34**

4.1 Review of support vector machine classifiers 37

4.1.1 Linear SVMs 37

4.1.2 Nonlinear support vector machines 38

4.2 Successive enhancement learning algorithm 40

4.3 Experiments 41

4.3.1 Training phase 41

4.3.2 Testing phase 42

5 Conclusion **48**

Bibliography **50**

List of Figures

1.1 Typical mammogram 2
1.2 Calcifications in the breast 3
1.3 Low pass filter used in synthesis of microcalcifications 6
1.4 Simulation of microcalcifications on normal mammogram region 1 7
1.5 Simulation of microcalcifications on normal mammograms region 2 8

2.1 Two dimensional subband decomposition using wavelet filters 11
2.2 Adaptive subband decomposition structure 12
2.3 8 search directions around the center pixel 13
2.4 Best output with wavelet based approach. False positives = 1, False negatives = 1 (blue - true MC, red - detected MC) 16
2.5 Worst output with wavelet based approach. False positives = 285, False negatives = 6 (blue - true MC, red - detected MC) 17

3.1 Quadrants of co-occurrence matrix 22
3.2 Input image for comparison of thresholding algorithms 27
3.3 Thresholding the image in Figure 3.2 using algorithms in [41] 28
3.4 Thresholding the image in Figure 3.2 using algorithms in [41] 28
3.5 Thresholding the image in Figure 3.2 with scaling between 0-255 29
3.6 Thresholding the image in Figure 3.2 without scaling 29
3.7 Output with cross entropy based approach for image which gave worst performance with the wavelet approach. False positives = 3, False negatives = 4 (blue - true MC, red - detected MC) 30
3.8 Output when cross entropy based thresholding does not work well. False positives = 8, False negatives = 41(blue - true MC, red - detected MC) 31
3.9 ROC curve for wavelet and cross entropy methods 32

4.1 Linear support vector classifier with separable data. Support vectors are circled. (From [7]) . . 36
4.2 Linear support vector classifier with non-separable data. Support vectors are circled. (From [7]) 38
4.3 Effect of high pass filtering (length = 41, cut off frequency = 0.125) on mammogram 41
4.4 Selection of positive samples from the neighbourhood of a MC. Center pixel is shown black and extra positive locations selected are shown as gray 42
4.5 Variation of error with the number of iterations for SEL algorithm 43
4.6 Overall ROC curve with real microcalcifications 44
4.7 Overall ROC curve with simulated microcalcifications 45
4.8 Result with SVM classifier for the image which gave worst output with wavelet approach . . . 46

List of Tables

2.1 Results with Wavelet based approach for varying μ and pre-threshold $'t'$ 15

2.2 Results obtained for wavelet based algorithm with simulated microcalcifications 15

3.1 Results with cross entropy based approach with variation of pre-threshold $'t'$ 27

3.2 Results obtained for cross entropy based algorithm with simulated microcalcifications 33

4.1 Results obtained with SVM classifier on training set, test set 43

4.2 Overall results obtained with SVM classifier 47

4.3 Results obtained with SVM classifier for simulated microcalcifications 47

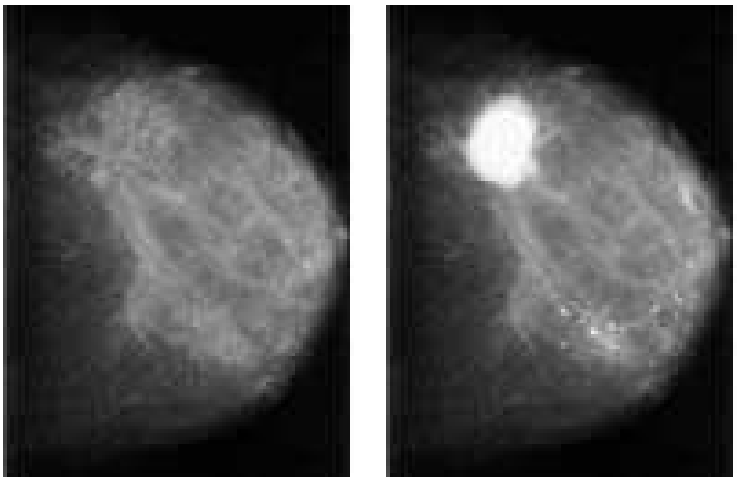
Chapter 1

Introduction

Breast cancer is the most frequent neoplasm in women in western countries and is the leading cause of cancer deaths among women age forty to forty-nine. It is second only to lung cancer as a cause of cancer death among women. Breast cancer accounts for one of every three cancer diagnoses in women. Over two million women living in the United States have been diagnosed with and treated for breast cancer. Breast cancer strikes Caucasian women most, followed by African American, Asian American and Pacific Islander, Hispanic of Latina and Native American women [52].

However most people tend to think of breast cancer as a woman's disease. But men get breast cancer too. According to the American Cancer Society, about 1200 new cases of breast cancer are diagnosed in American men each year (compared to about 200,000 cases of breast cancer in U.S. women). Research has shown that women with a family history of breast cancer have a higher risk for developing the disease. That is true whether the family history is on the mother's side or the father's [52].

Mammography is acknowledged as the single most effective method of screening for breast cancer. Mammography-based screening programs are carried out in many countries, and their effectiveness has had a great impact on diagnosis. Mammography is a technique that uses X-rays to provide an image of the breast. These images, called mammograms, are used to find potential signs of breast cancer: like tumors, small clusters of calcium (micro-calcifications) and abnormal changes in the skin. Overall, mammography is the best screening tool available today for breast cancer. It can find cancers at an early stage, when they are small and most responsive to treatment. A number of new methods are also currently being developed to detect breast cancer in early stages, including: ultrasound, magnetic resonance imaging (MRI), thermal imaging and positron emission tomography (PET). It is not yet clear if these methods will play an effective role in breast cancer screening. Ultrasound uses sound waves to make images of tissue, and it is commonly used as a follow-up test to mam-



(a) Normal breast (b) Cancerous breast

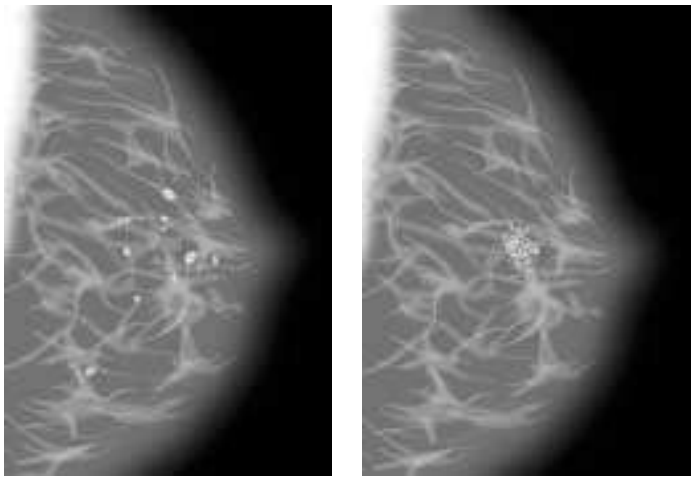
Figure 1.1: Typical mammogram

mography and clinical breast exam. However, ultrasound is not used alone for breast cancer screening. It does not produce an accurate image of the entire breast. MRI uses nuclear resonance to examine the breast, and there is promising evidence that it could eventually become a good tool for breast cancer screening [40]. Positron emission tomography detects how much sugar is being consumed by cells. Cancer cells tend to consume more sugar than normal tissue, which can help in identifying tumors. PET cannot provide a detailed enough image to find breast cancers in very early stages of development [13]. Therefore, it is currently not a promising option for breast cancer screening.

Computer Aided Detection (CAD) in mammography is a relatively new technology. The Food and Drug Administration (FDA) approved the first CAD device for film mammography in 1998 and expanded its approval to include the use of CAD for digital mammography detection in 2001.The technology is designed to provide an increased level of assistance to a radiologist when a mammography film is viewed. The computer software is designed to indicate areas in the mammogram that may be cancerous. However, the use of CAD as a second opinion, serves only to complement the interpreting radiologist, who must screen out the false indications of malignancies while accepting true lesions that might have been overlooked without CAD [20].

1.1 What are micro-calcifications

One of the hallmark features of cancer cells is that they do not stay within clear borders separate from the surrounding tissue; in fact, they invade the surrounding tissue. When cancer cells have started this type of tissue invasion, the core cancer cells will appear on a mammogram as a light patch, and the cells that are



(a) Benign macrocalcifications are larger and randomly spread throughout the breast. No follow-up care is usually needed.

(b) Microcalcifications are small, appear clustered and have irregular shapes. These may be a sign of cancer. Follow-up with more mammograms and perhaps a biopsy are needed.

Figure 1.2: Calcifications in the breast

creeping out into the normal tissue will create either a fuzzy or spiky appearance along the outer edge (called "spiculated").

Another potential sign of cancer are small, bright white spots that are tiny calcium deposits which form in the breast as a woman ages. They are common and can result from a number of different things such as trauma to the breast and inflammation. There is no known relationship between the amount of calcium in a womans diet and breast calcifications. Calcifications are too small to be felt however a mammographic image is characterized by a high spatial resolution which is adequate to detect these subtle fine-scale signs. Most calcifications are benign and are actually quite common, with about half of all mammograms showing some. However, certain patterns of calcifications are suspicious so they must be looked at carefully as can be seen in Figure 1.2 . Tight clusters or lines of tiny calcifications (micro-calcifications) can be indications of cancer or precursors to cancer [52]. Individual MCs are sometimes difficult to detect because of the surrounding breast tissue, their variation in shape, orientation, brightness and size (typically, 0.05-1mm). This makes detecting MCs a complex and cumbersome task which requires highly specialized radiologists.

1.1.1 Types of calcifications

- Macro-calcifications are almost always benign (not cancer). They appear large and round on the mammogram films and require no further follow-up.

- Micro-calcifications are smaller and more numerous than the larger macro-calcifications. They are usually benign, but can sometimes be a sign of cancer. The radiologist will look at the size, shape and distribution of the micro-calcifications to see if they are suspicious. More mammograms, an ultrasound study and a biopsy may be necessary. Suspicious micro-calcifications turn out to be cancer about 20 to 25 percent of the time.

- Sometimes it is hard to tell if micro-calcifications are cancer or not. These micro-calcifications are called indeterminate.When this happens the radiologist may take more X-rays to help decide if the micro-calcifications are benign, probably benign, suspicious, or malignant. If they are probably benign, then there is a 98 percent chance that they are not cancer. However, if they are suspicious, more follow-up is needed.

1.2 Selection of algorithms for comparison

In the literature there are many methods that have explored the challenge of detection of micro-calcifications(MCs) in mammograms and here we seek to evaluate three of the best state of the art algorithms. In our study we will concentrate only on the problem of the detection of individual MCs rather than clusters. We shall also not investigate the problem of classifying them as benign/malignant.

Lee et al. [30] have described a prototype of a CAD system for detection of cancerous calcifications in mammograms. It employs fractal dimension to determine the roughness of an area which is different for calcification and normal background. To calculate the fractal dimension they used the Blanket Method developed by Mandelbrot and extended by Peleg for estimating the fractal dimension of a surface area. Bagci et al. [2] have employed the wavelet transform and adaptive filtering to get the error in predicting high pass components. MCs are then obtained by thresholding the wavelet coefficients based on their HOS(Higher Order Statistics) values . HOS test is a robust indicator of the Gaussianity of an area [19]. It is based on sample estimates of the first four moments of the prediction errors and the observation that wavelet transform has a Gaussian distribution of values except around the MCs. Wan Mimi et al. [54] have compared these two methods [19, 30] and reported the percentage of clusters detected to be 92% for [30] and 99% for [2] . Hence we chose Bagci's approach [2] as the first method to be investigated.

El-Naqa et al. [16] have mentioned the use of Support Vector Machine(SVM) for detection of clusters of MCs in mammograms. SVM is trained after preprocessing the input with a high pass filter. They compared their method with four other techniques 1) image difference technique(IDT) [39] 2) DoG method [14] 3) wavelet decomposition(WD) [51] 4) two-stage multilayer neural network (TMNN) method [56]. SVM outperformed

all the other methods as shown in the FROC curves in [16]. Hence we chose this as the second method for our comparison.

Entropy thresholding forms another class of algorithms widely used in thresholding. It has also been used in detection of MCs in mammograms. Moti Melloul and Melloul et al. [38] have for the first time used a three dimensional co-occurrence matrix for threshold calculation. They obtained mean detection rates of 93.75% of true positives, 6.25% of false positives, and 2.0% of false negatives. It was reported to be better than three other state of the art algorithms that they used for comparison. Hence we chose this as the third algorithm for our investigation. Melloul et al. [38] have tested on the MIAS database (8 bit images of size 1024x1024), however we have tested it on the DSDM database (12 bit images). We have also compared its performance with three other entropy thresholding methods.

1.3 Synthesis of calcifications on mammograms

One of the fundamental problems in evaluating the performance of algorithms for detection of microcalcifications is that one is never certain if a false positive is not an actual calcification. This is because the radiologist is prone to human error. While most of the research has concentrated on detection methods little attention has been devoted to defining a gold standard for an objective evaluation of computerized detection accuracy. Some authors compare their results to the locations indicated by an experienced radiologist while others confirm the presence of microcalcification by biopsy or magnification. Biopsy and magnification are able to confirm the presence of microcalcifications and can accurately evaluate the rate of false positives when the aim is to detect clusters of microcalcifications on the whole mammogram. However these tools are not able to determine the number of microcalcifications and to localize them individually. Interpretation by human beings is highly subjective and the problem of reproducibility is seldom mentioned. It is mentioned in [31] that a study performed in their laboratory demonstrated that the evaluation based exclusively on the indications of the radiologist is not reliable for the detection of individual microcalcifications. Lado MJ et al. [29] have built upon the technique for simulation of clustered microcalcifications mentioned in [31]. It involves creation of a bank of real microcalcifications extracted from mammograms and simulation is based on the distribution of features such as shape, size, no of microcalcifications in a cluster etc. While in [31] simulated ones are placed on a normal background, in [29] they are placed on cancerous mammograms along with the real ones. The results of the observer testing of simulated versus real clustered microcalcifications in [29] were as follows: 47.74% of the simulated were called simulated and 44.14% were called real, the rest (8.12%) being classified as uncertain. Of the real clusters, 47.75% were called simulated and 45.94% were called real, the rest (6.31%) being classi-

1	2	1
2	4	2
1	2	1

Figure 1.3: Low pass filter used in synthesis of microcalcifications

fied as uncertain. Further review of literature on simulated of microcalcifications can be found in [29]. Some authors have also tried to artificially construct the whole mammogram. These can be found in [48, 44, 46, 45].

Our effort is similar to that mentioned in [29] where we have a bank of real microcalcifications as marked by an experienced radiologist. Microcalcifications randomly chosen from the bank are placed at random locations over a normal mammogram. Low pass filtering around the border of the microcalcification blends it with the background. Figure 1.3 shows the filter used during this process. However this method is not fool proof since even normal mammograms can have microcalcifications. We have not attempted to simulate microcalcification clusters since we only deal with the problem of detection of individual microcalcifications rather than clusters. In Figure 1.4 and Figure 1.5 we can see the result of this procedure on two mammogram regions.

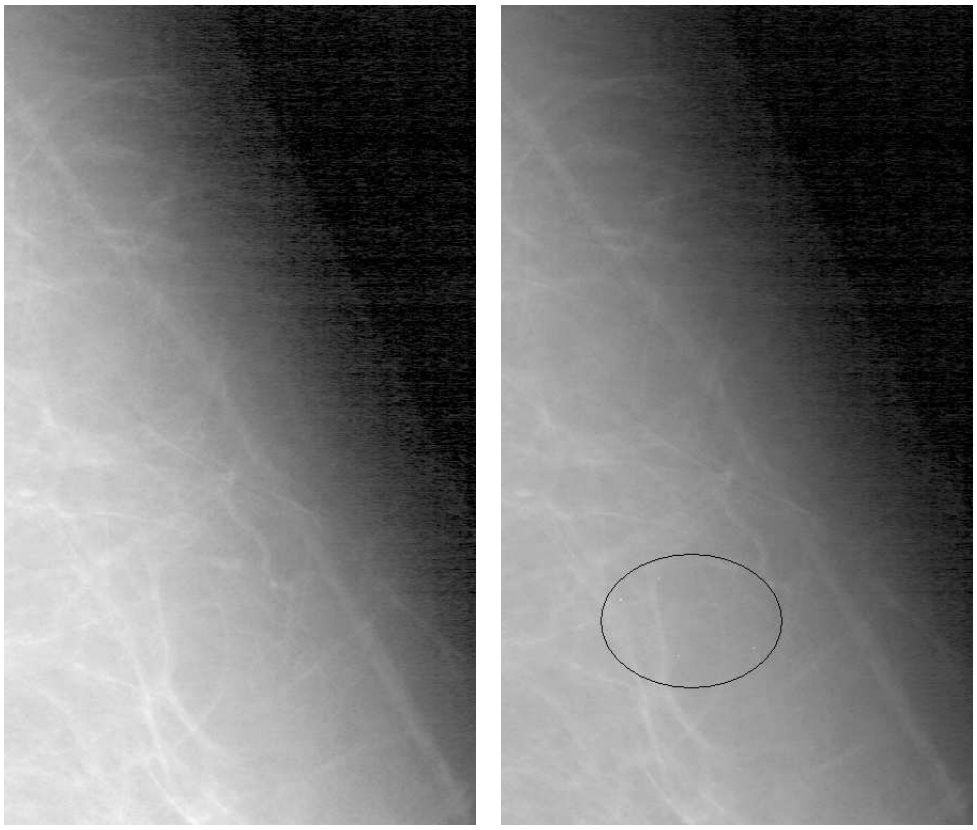
1.3.1 Organization of the thesis

Chapter 1 gives an introduction to CAD in breast cancer detection. It also gives a brief literature review and shows how were the three algorithms selected for comparison. It also describes a method for the synthesis of microcalcifications on normal mammograms.

Chapter 2 describes the wavelet filtering based algorithm mentioned in [2] and a brief literature review related to use of wavelets in detection of MCs. In [2] authors rank the locations by its HOS value but do not provide any statistics such as false positive or true positive detection rate. We have extended their approach by selection of threshold using minimum cross entropy thresholding as explained in the section 3.1.1.

Chapter 3 describes entropy thresholding based methods to detect microcalcifications. We compare the three dimensional co-occurrence matrix based algorithm mentioned in [38] with three other entropy thresholding algorithms and select the best performing algorithm. The selected algorithm is then used for the detection of microcalcifications. It also gives a brief literature review about the application of similar information theoretic measures in thresholding.

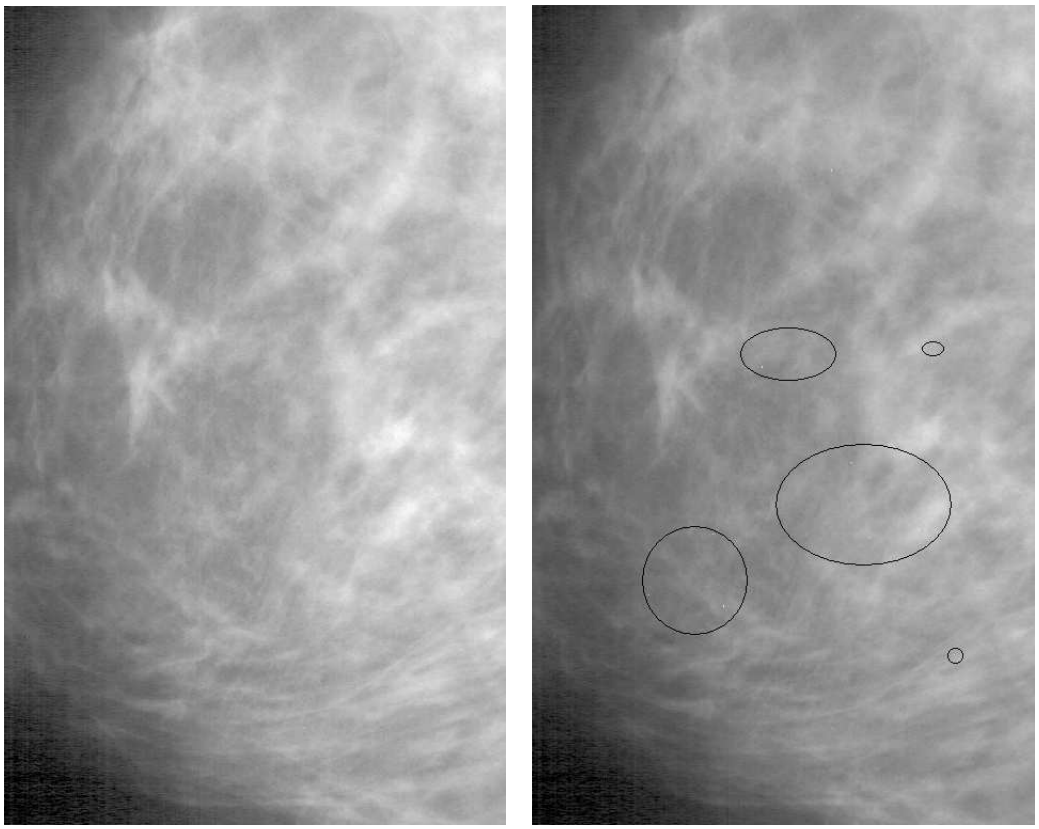
Chapter 4 describes the SVM algorithm mentioned in [16] and a brief literature review related to use of SVM classifiers in detection of MCs. Here the input mammogram is pre-processed with a high pass filter and detection is done by classifying each pixel using a SVM classifier.



(a) Normal mammogram without any microcalcification

(b) After addition of microcalfications

Figure 1.4: Simulation of microcalcifications on normal mammogram region 1



(a) Normal mammogram without any microcalcification (b) After addition of microcalfications

Figure 1.5: Simulation of microcalcifications on normal mammograms region 2

Chapter 2

Adaptive Wavelet Filtering

The theory of wavelets provides a common framework for numerous techniques developed independently for various signal and image processing applications. For example, multiresolution image processing, used in computer vision, subband coding, developed for speech and image compression, and wavelet series expansions, developed in applied mathematics, have been recently recognized as different views of a single theory. The classical approach for the analysis of non-stationary signals is the short-time Fourier transform (STFT) or Gabor transform. With the advent of wavelet transform (WT), short windows at high frequencies and long windows at low frequencies can be used to provide better signal resolution than the STFT. The Wavelet transform can also be seen from a signal decomposition view point. In this case, a signal is decomposed onto a set of basis functions which are called wavelets and are the core of wavelet analysis. These basis functions are obtained from a single mother wavelet by dilations and contractions (scalings), as well as translations or shifts. An excellent review of these techniques and concepts can be found in [55] where it is also shown that wavelet decomposition is closely related to multirate signal processing techniques. A particular wavelet decomposition relates to filter banks and can be the basis for subband coding schemes used in speech and image compression. Strickland et al. [51] have developed a two stage method for detecting MCs. The first stage is based on an undecimated wavelet transform where they do away with down-sampling thus producing wavelet images of the same size as the input image. Detection takes place in the HH and LH+HL subbands. By appropriate selection of the wavelet basis the detection of MCs in the relevant size range can be obtained. In fact the filters used are closely related to prewhitening matched filters for detecting Gaussian objects in Markov noise [51]. Mata, R. et al. [36] proposed a multi-resolution analysis by decomposing the image through a band-pass filter bank, so that in each subband image only the details at the given scale become visible. Thereafter all the images are combined in a final one to obtain an image that contains all the interest details at the scale where micro-

calcifications tend to appear. Clusters of MCs are then detected by performing a third order HOS Gaussianity test [23]. Wang and Karayiannis [55] have given an excellent review of wavelet theory and its application to images. According to their approach MCs are extracted from mammograms by suppressing the subband of the wavelet-decomposed image that carries the lowest frequencies and contains smooth (background) information. Only the high frequency components are used to reconstruct the image. The visibility of MCs is improved by using a nonlinear thresholding method to enhance the histogram of the resulting mammograms. Wan Mimi Diyana et al. have mentioned in [15] a wavelet based approach which used Higher Order Statistics(HOS). HOS test has been first mentioned in the literature in [53] and has been beautifully described in [34]. Below we outline normal subband decomposition and adaptive subband decomposition which has been explored in this work.

2.1 Wavelet based representation of images

2.1.1 Normal subband decomposition

Wavelet-based image decomposition can be interpreted as an image filtering process. Normal wavelet subband decomposition has been illustrated in Figure 2.1. For a given image I of size 2^nx 2^n wavelet-based subband decomposition can be performed as follows: the wavelet filters $h_l(n)$ and $h_h(n)$ are applied to the rows of the image I. The filter $h_l(n)$ is a low-pass filter with frequency response $H_l(\omega)$ and $h_h(n)$ is a high-pass filter with frequency response $H_h(\omega)$. By filtering the image I with $H_l(\omega)$, we obtain low-frequency information(background). By filtering the image with $H_h(\omega)$, we obtain the high-frequency information(edges). After downsampling by a factor of two, we obtain two subbands: x_l and x_h. Since we downsample by a factor of two in the horizontal direction of each subband, the size of these two downsampled subbands is 2^n x 2^{n-1} (see Figure2.1). The filters are then applied to the columns of the subbands x_l and x_h, and the following four subbands are obtained , x_{ll}, x_{lh}, x_{hl} and x_{hh}. Since we now downsample by a factor of two in the vertical direction of each subband, the four subbands have gone through downsampling by a factor of two in both directions and the final size of each subband is 2^{n-1} x 2^{n-1} (see Figure 2.1).

The subband x_{ll} contains the smooth information and the background intensity of the image and the subbands x_{ll}, x_{lh} and x_{hl} contain the detail information of the image. The subband x_{ll} corresponds to the lowest frequencies, x_{lh} gives the horizontal high frequencies (vertical edges), x_{hl} gives the vertical high frequencies (horizontal edges), and x_{hh} the high frequencies in both directions (corners and diagonal edges).

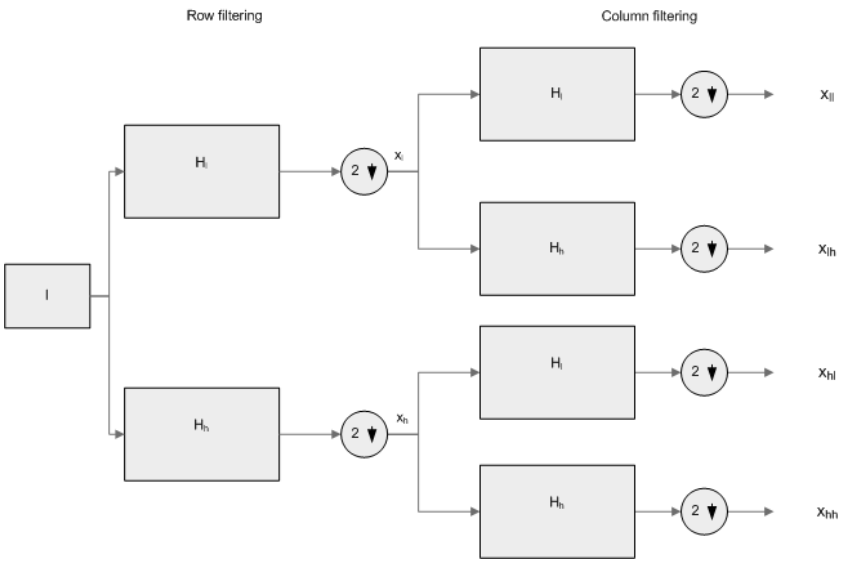


Figure 2.1: Two dimensional subband decomposition using wavelet filters

2.1.2 Adaptive subband decomposition

Bagci et al. [2] have used the adaptive wavelet structure proposed by Ömer Nezih Gerek et al. in [18] for image compression. They have shown that it works better than normal subband decomposition especially at the regions that are not as bright or significant as others. In this structure (Figure 2.2) the low band signal x_l is obtained by down-sampling the original signal, x and consists of only even samples. The high pass component x_h is obtained through an adaptive filter structure. The sequence x_2 is a shifted and downsampled version of the original signal x and consists of only odd samples. Next $\hat{x_2}$ the estimate of x_2 is obtained using x_l and filter $'w'$ as shown in equation 2.1. The adaptive FIR estimator used is obtained by predicting the odd samples x_2 from the even samples x_l as follows:

$$\hat{x_2} = \sum_{k=-N}^{k=N} w_{n,k} x_l(n-k) = \sum_{k=-N}^{k=N} w_{n,k} x(2n-2k) \tag{2.1}$$

The filter coefficients $w'_{n,k}s$ are updated using an LMS type algorithm as follows:

$$\hat{w}(n+1) = \hat{w}(n) + \mu \frac{\tilde{\nu}_n e(n)}{\| \tilde{\nu}_n \|^2} \tag{2.2}$$

where $\hat{w}(n) = [w_{n,-N}, \ldots, w_n, \ldots, w_{n,N}]$ is the weight vector and $\tilde{\nu}_n = [x_l(n-N), x_l(n-N+1), \ldots, x_l(n+N-1), x_l(n+N)]^T$ is the data vector at the time instant n. Estimation error $x_h (= e(n))$ is obtained by $x_h = x_2 - \hat{x_2}$. Here we have used Euclidean norm as in [2]. This is simple adaptive wavelet filter structure as applied to

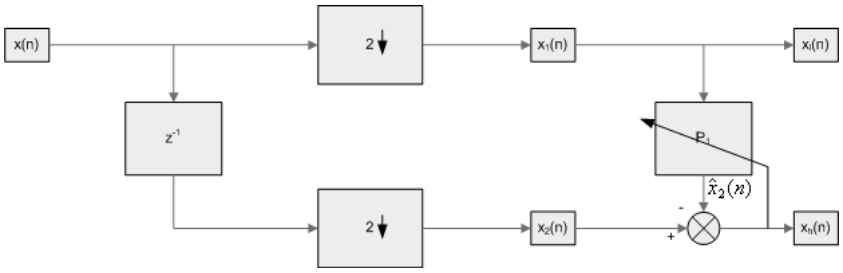


Figure 2.2: Adaptive subband decomposition structure

uni-dimensional signal. The extension of this adaptive filter bank structure to two dimensions is done in the same way as normal subband decomposition in section 2.1.1. Initially rows are filtered and the obtained images are column filtered thus again giving four subimages x_{ll}, x_{lh},x_{hl} and x_{hl}. Low-low image x_{ll} can be ignored in analysis as it gives background information and during the experiments it was observed that x_{hh} did not contain any information relevant to the current problem of detection of microcalcifications. Hence x_{lh} and x_{hl} have been used as in [2].

2.1.3 HOS test

Bagci et al. [2] have mentioned that MCs occur as isolated bright spots on the mammogram hence they correspond to local maxima on the image. Fourth order statistics or Kurtosis which we shall use in identification of MCs can be represented by

$$H = \gamma_4 = \frac{E\{[x - E(x)]^4\}}{(E\{[x - E(x)]^4\})^2} - 3 \tag{2.3}$$

It has been shown in [53] that H is significantly different in the regions which contain MCs as opposed to those that don't have them. However during our experiments we found out that it was not true for all the cases. HOS values were also high in the regions that did not have any tissue material which was surprising, and those portions were filtered out by rejecting pixels with mean in their neighbourhood less than a threshold. Average HOS value for negative positions was less than zero but average maximum for negative positions was 67.3 for a window size of 16x16 and 37 for 9x9. Hence we chose a window size of 9x9. We also experimented with

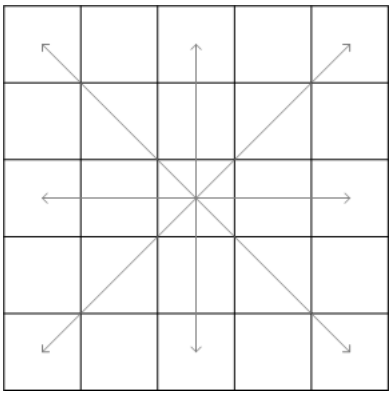


Figure 2.3: 8 search directions around the center pixel

calculating the HOS values on the actual image as mentioned in [2]. However average maximum for positive positions i.e. microcalcifications was greater for wavelet image, hence it was decided to work with the wavelet image. Some of the true MCs had HOS values less than non-MC which makes it difficult to decide upon a threshold.

2.1.4 Algorithm

Below we outline the steps in this algorithm:

- Bagci et al. [2] have outlined a pre-processing scheme to determine the location of maxima. These constitute a subset which dictates the positions where fourth order HOS test is carried out. For this purpose we extract a 5x5 window around each pixel and if the maximum is 1.1 times the minimum value in the window then that pixel is recorded as a maxima. If it does not satisfy this criteria but pixel value is 10% greater than the minimum value and greater than or equal to the maximum value of the neighbouring pixels then we check the following condition: check the two pixels in each 8 directions around the pixel as shown in the Figure 2.3. If the value of the center pixel is 1.05 times greater than at least one of the two pixels in each direction then the point is marked as possible microcalcification region.

- Obtain the wavelet transform of the input image using the adaptive subband structure explained in section 2.1.2. Thus we get the quarter sized wavelet image $e[m,n] = (\mid x_{lh} \mid + \mid x_{hl} \mid)[m,n]$. It is also pre-thresholded with a small threshold $'t'$, which avoids peaks due to the background which can affect the entropy thresholding process.

- Estimate the fourth order statistic, H, in a MxM window around each maximum location in the image at the corresponding position in the quarter size wavelet image. If the maximum is at the location (*i*,*j*) in the

image then we search at the location m = round(i/2), n = round(j/2) in the image $e[m,n]$. H is estimated as shown below

$$H(I_1, I_2, I_4) = I_4 + 2I_1^4 - 3I_2^2 \tag{2.4}$$

$$I_k = \frac{1}{M+N} \sum_{m=1}^{M} \sum_{n=1}^{N} e^k[m,n], k = 1,2,3,4 \tag{2.5}$$

- In the method mentioned in [2] first the local maximas of the image are detected as above and ranked according to a high order statistical (HOS) test performed over the subbands obtained by the adaptive wavelet transform. No statistical results have been provided in [2] terms of false positives or ROC curve against which one can evaluate the efficacy of their algorithm. However we differ from their method as we calculate a threshold T, using the entropy thresholding method mentioned in [33]. Presence of a microcalcification is determined if H is greater than or equal to this threshold. Thus the output of our procedure shows locations of MCs which is then compared with the truth. This enables us to obtain a ROC curve and compare it with other algorithms mentioned in this work.

2.1.5 Experimental results

We have used 5 mammograms from the DSDM database provided freely on the internet by the University of Florida. These contain 219 microcalcifications identified by an experienced radiologist. We have experimented with varying step size μ in equation 2.2 and pre-threshold $'t'$. In Table 2.1 we can see the results obtained with different values of these parameters. It can be seen that pre-threshold $'t'$ does not bear much upon the performance. Best performance for this algorithm is a detection rate of 33% with 35 false positives per image which is very high. In Figure 2.4 we can see its best output. Blue dots indicate positions indicated by the radiologist and red dots indicate positions identified by this algorithm. In Figure 2.5 we can see its performance when at its worst . Here it selects locations which actually belong to the background. Its performance can perhaps be improved if we supply more prior information. It could also be improved if we select a region of interest(ROI) and then apply it only on that portion; however we have been consistent in this respect while implementing the three algorithms being compared in this work. In Table 2.2 we can see its performance over mammograms with simulated microcalcifications.

Parameters	Fraction of true positives detected	False positives per image
$\mu = 2, t = 2$	0.19	147.91
$\mu = 1, t = 2$	0.31	49.41
$\mu = 0.1, t = 2$	0.32	36.77
$\mu = 0.2, t = -20$	0.33	35.18
$\mu = 0.2, t = 0$	0.33	35.18
$\mu = 0.2, t = 0.5$	0.33	35.18
$\mu = 0.2, t = 2$	0.33	35.18
$\mu = 0.2, t = 10$	0.33	35.18
$\mu = 0.2, t = 20$	0.33	35.18
$\mu = 0.5, t = 2$	0.32	33.86

Table 2.1: Results with Wavelet based approach for varying μ and pre-threshold $'t'$

Parameters	Fraction of true positives detected	False positives per image
$\mu = 2, t = 2$	0.29	86.4
$\mu = 1, t = 2$	0.82	25.7
$\mu = 0.1, t = 2$	0.82	34.7
$\mu = 0.2, t = -20$	0.84	32.1
$\mu = 0.2, t = 0$	0.84	32.1
$\mu = 0.2, t = 0.5$	0.84	32.1
$\mu = 0.2, t = 2$	0.84	32.1
$\mu = 0.2, t = 10$	0.84	32.1
$\mu = 0.2, t = 20$	0.84	32.1
$\mu = 0.5, t = 2$	0.76	26.3

Table 2.2: Results obtained for wavelet based algorithm with simulated microcalcifications



Figure 2.4: Best output with wavelet based approach. False positives = 1, False negatives = 1 (blue - true MC, red - detected MC)

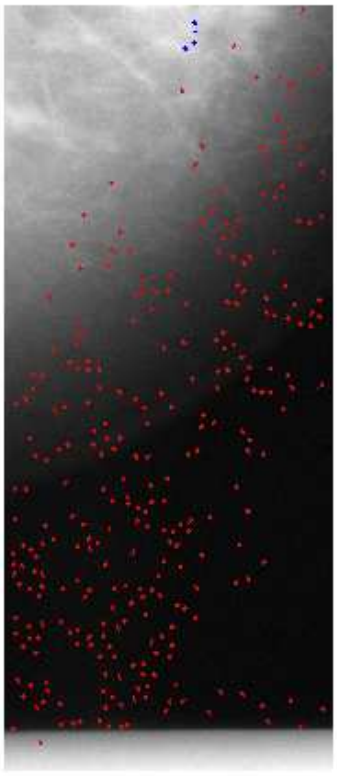


Figure 2.5: Worst output with wavelet based approach. False positives = 285, False negatives = 6 (blue - true MC, red - detected MC)

Chapter 3

Entropy thresholding

Shannon [49] defined the entropy of a system as a function of the probability of occurrence of different states of the system. If a system has n different states with probability of occurrence $p_i, i = 1,2,\ldots,n, \sum_{i=1}^{n} p_i = 1$, then the gain in information from the occurrence of the event i is defined as $\triangle I = -\log_2 p_i$. The expected value of such a gain in information is defined as the entropy of the system. Thus the entropy H of the system is: $H = -\sum_{i=1}^{n} p_i$. Let $F = [f(x,y)]_{PxQ}$ where $f(x,y)$ is the gray value at (x,y); $f(x,y) = \in G_L = \{0,1,\ldots,L-1\}$, the set of gray levels. Let N_i be the frequency of the gray level i. Then $\sum_{i=1}^{L-1} N_i = PxQ = N$. Following Shannon's definition of entropy, entropy of the image histogram can be defined as

$$H = -\sum_{i=0}^{L-1} p_i \log_2 p_i; \qquad p_i = N_i/N \tag{3.1}$$

for the image segmentation problem. Thus different images with identical histograms will result in same entropic value in spite of their different spatial distributions of gray levels. N.R. Pal et al. [41] have given a different formulation of entropy for images. We know that in an image pixel intensities are not independent of each other. This dependency of pixel intensities can be incorporated by considering sequences of pixels to estimate the entropy. In order to arrive at the expression of entropy of an image the following theorem due to Shannon can be stated:

Theorem: Let $p(s_i)$ be the probability of a sequence s_i of gray levels of length q, where a sequence s_i of length q is defined as a permutation of q gray levels. Let us define

$$H^{(q)} = -\frac{1}{q}\sum_{i} p(s_i)\log_2 p(s_i), \tag{3.2}$$

where the summation is taken over all gray level sequences of length q. Then $H^{(q)}$ *is a monotonic decreasing function of (q) and* $\lim_{q\to\infty} H^{(q)} = H$*, the entropy of the image.*

For different values of q we get different orders of entropy.

Case 1: q = 1 , i.e. sequence of length one. If q = 1 we get $H^{(1)} = -\sum_{i=0}^{L-1} p_i \log_2 p_i$, where p_i is the probability of occurrence of the gray level i. Such an entropy is a function of the histogram only and it may be called the "global entropy" of the image. Therefore different images with identical histograms would have same $H^{(1)}$ value irrespective of their contents.

Case2: q = 2, i.e. sequences of length two. Hence, $H^{(2)} = -\sum_{i=0}^{L-1} p_i \log_2 p_i$, where s_i is a sequence of gray level of length two, $= \sum_i \sum_j p_{ij} \log_2 p_{ij}$, where p_{ij} is the probability of co-occurrence of the gray levels i and j. Therefore $H^{(2)}$ can be obtained from the co-occurrence matrix, defined in section 3.1.2. $H^{(2)}$ takes into account the spatial distribution of gray levels. Therefore, two images with identical histograms but different spatial distributions will result in different entropy, $H^{(2)}$ values. Expressions for higher order entropies(q > 2) can also be deduced in a similar manner. $H^{(i)}$, $i \geq 2$, may be called the "local entropy of the image". The power of the gray level co-occurence approach is that it characterizes the spatial interrelationships of the gray levels in a textural pattern and can do so in a way that is invariant under monotonic gray level transformations. Its weakness is that it does not capture the shape aspects of the gray level primitives. Hence it is not supposed to work well for textures composed of large area primitives. Also it cannot capture the spatial relationships between primitives that are regions larger than a pixel. In section 3.1.2 more details about the approach followed by Pal et al. [41] have been given.

Here we give a review of some of the recent works in this field. Earlier works related to entropy thresholding can be found in Pal et al. [41]. Sahoo, P.K et al. [47] have given a very good review of entropy thresholding techniques. Zuo HongQing [25] used entropy thresholding to segment to blood vessels in retinal images. He employed several matched filters to enhance the contrast of blood vessels. Subsequently entropies are derived from gray level- normalized gradient co-occurrence matrices instead of a gray level transition matrix. Chanwimaluang, T. et al. [12] use matched filters and most common gray level transition matrix to segment blood vessels from retinal images by maximizing the sum of entropies of object and the background. Li X. et al. [34] have compared three distinct thresholding algorithms, mixture modeling, fuzzy sets and entropy thresholding when applied for the segmentation of aurora boundary in satellite images. Jianwei Wang et al. [54] introduce two new relative entropy based thresholding techniques which are local relative entropy(LRE) and joint relative entropy (JRE). Since relative entropy thresholding is generally sensitive to sparse histograms, a histogram compression and translation (HCT) is proposed to improve and enhance their performance. In order to evaluate image quality objectively, they also propose two measures of uniformity and shape and show

by experiment that visual inspection was consistent with uniformity and shape measures. Brink & Pendock [6] use cross-entropy to determine the appropriate threshold. They also compare it with metric measures such as cross correlation, χ^2 measure of discrepancy and show that use of a correlation measure for thresholding is valid (in terms of maximum entropy theory) only if the distributions are normal and the first moment is preserved. [12, 25, 54] use two dimensional co-occurrence matrix for entropy thresholding. [58] have reported the superiority of two dimensional histogram instead of co-occurrence for entropy thresholding. Kurani et al. [28] investigate a new approach to the co-occurrence matrix currently used to extract textural features: co-occurrence matrices for volumetric data. It is used to calculate volumetric texture descriptors that can be used for segmentation and classification of soft tissues in CT studies. C.H. Li & P.K.S. Tam [33] derive a fast iterative method for minimum cross entropy thresholding which accurately locates the threshold while significantly reducing the computational cost.

3.1 Algorithms

We have compared four entropy thresholding techniques and applied the best one for the detection of micro-calcifications on mammograms. Below we describe each of the four techniques.

3.1.1 Minimum cross entropy thresholding

This algorithm has been described by Li et al. in [33]. For a histogram h defined on the gray level range [0,L-1] , the zeroth and the first moments of the foreground and background portions of the thresholded histogram are respectively,

$$m_{0a}(t) = \sum_{i=0}^{t-1} h(i) \qquad m_{ob} = \sum_{i=t}^{L-1} h(i) \tag{3.3}$$

$$m_{1a}(t) = \sum_{i=0}^{t-1} ih(i) \qquad m_{ob} = \sum_{i=t}^{L-1} ih(i) \tag{3.4}$$

The portions' means are defined as

$$\mu_a(t) = \frac{m_{1a}(t)}{m_{0a}(t)} \qquad \mu_b(t) = \frac{m_{1b}(t)}{m_0(t)} \tag{3.5}$$

The minimum cross entropy method selects the threshold which minimizes the cross entropy of the image and its segmented version. The criterion function [33] is found to be

$$\eta(t) = -m_{1a}(t)\log(\mu_a(t)) - m_{1b}(t)\log(\mu_b(t)) \tag{3.6}$$

Thus the optimal threshold is given by the minimizer of equation 3.6, $t_{op} = \arg\min_t \eta(t)$.

3.1.2 Two dimensional co-occurrence matrix based

N.R. Pal, S.K. Pal [41] is one of the earliest works in entropy thresholding. It is based on maximizing the conditional entropy of object and background. Suppose an image has two distinct portions, the object X and the background Y. Suppose the object consists of the gray levels $\{x_i\}$ and the background contains the gray levels $\{y_i\}$. The conditional entropy of the object X given the background Y, i.e. the average amount of information that may be obtained from X given that one has viewed the background Y, can be defined as

$$H(X/Y) = -\sum_{x_i \in X}\sum_{y_j \in Y} p(x_i/y_j)\log_2 p(x_i/y_j) \tag{3.7}$$

Similarly the conditional entropy of the background Y given the object X is defined as

$$H(X/Y) = -\sum_{x_i \in X}\sum_{y_j \in Y} p(y_j/x_i)\log_2 p(y_j/x_i) \tag{3.8}$$

An additional constraint is imposed that y_j and x_i have to be adjacent.

Algorithm 1

Based upon the above formulation the co-occurrence matrix of an image I of size PxQ and L gray levels is an LxL dimensional matrix $T = [t_{ij}]_{LxL}$ that gives an idea about the transition between adjacent pixels. Each entry in the matrix t_{ij} gives the number of times the gray level j follows the gray level i in some pattern. Depending upon different patterns different definitions of co-occurrence matrix are possible. It has been reported in [41] that consideration of both horizontal and vertical transitions allows all the edges to participate in the threshold selection. Thus t_{ij} is defined as follows:

$$t_{ij} = \sum_{l=1}^{P}\sum_{k=1}^{Q}\delta \tag{3.9}$$

where

$$\delta = 1, \text{ if } \begin{cases} f(l,k) = i \quad \text{and} \quad f(l,k+1) = j \\ \text{or} \\ f(l,k) = i \quad \text{and} \quad f(l+1,k) = j; \end{cases} \tag{3.10}$$

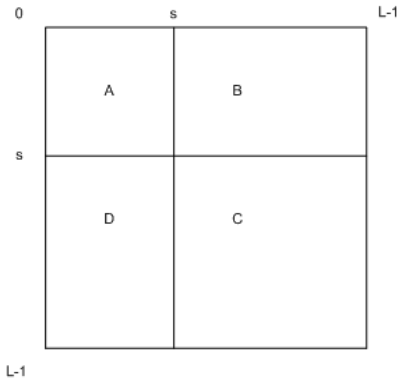


Figure 3.1: Quadrants of co-occurrence matrix

The probability of co-occurrence p_{ij} of gray levels i and j therefore can be written as

$$p_{ij} = t_{ij} \Big/ \sum_i \sum_j t_{ij} \tag{3.11}$$

If s, $0 \leq s \leq L-1$ is a threshold, then 's' partitions the co-occurrence matrix into four quadrants, namely A,B,C and D as seen in Figure 3.1. Let us define the following quantities

$$\begin{aligned} P_{A=} \sum_{i=0}^{s} \sum_{j=0}^{s} p_{ij}, \quad & P_{B=} \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_{ij} \\ P_{C=} \sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} p_{ij}, \quad & P_{D=} \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij} \end{aligned} \tag{3.12}$$

Normalizing the probabilities within each individual quadrant, such that the sum of the probabilities of each quadrant equals one, we get the cell probabilities for different quadrants:

$$\begin{aligned} p_{ij}^{A} &= \frac{p_{ij}}{P_A} = \frac{t_{ij} \Big/ \left(\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij} \right)}{\sum_{i=0}^{s} \sum_{j=0}^{s} \left\{ t_{ij} \Big/ \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij} \right\}} \\ &= \frac{t_{ij}}{\sum_{i=0}^{s} \sum_{j=0}^{s} t_{ij}} \\ &\text{for } 0 \leq \text{i} \leq \text{s}, \; 0 \leq \text{j} \leq \text{s} \end{aligned} \tag{3.13}$$

Similarly,

$$p_{ij}^{B} = \frac{p_{ij}}{P_B} = \frac{t_{ij}}{\sum_{i=0}^{s}\sum_{j=s+1}^{L-1} t_{ij}} \quad \text{for s+1}\leq\text{i}\leq\text{L-1, s+1}\leq\text{j}\leq\text{L-1} \tag{3.14}$$

and

$$p_{ij}^{C} = \frac{p_{ij}}{P_C} = \frac{t_{ij}}{\sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1} t_{ij}} \quad \text{for s+1} \leq \text{i} \leq \text{L-1, s+1} \leq \text{j} \leq \text{L-1} \tag{3.15}$$

$$p_{ij}^{D} = \frac{p_{ij}}{P_D} = \frac{t_{ij}}{\sum_{i=s+1}^{L-1}\sum_{j=0}^{s} t_{ij}} \quad \text{for s+1} \leq \text{i} \leq \text{L-1, s+1} \leq \text{j} \leq \text{L-1} \tag{3.16}$$

Now the second order local entropy of the background can be defined as

$$H_A(s) = -\frac{1}{2}\sum_{i=0}^{s}\sum_{j=0}^{s} p_{ij}^{A}\log_2 p_{ij}^{A}. \tag{3.17}$$

Similarly the second order entropy of the object can be written as

$$H_C(s) = -\frac{1}{2}\sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1} p_{ij}^{C}\log_2 p_{ij}^{C}. \tag{3.18}$$

Hence the total second order local entropy of the object and the background can be written as

$$H_T(s) = H_A(s) + H_C(s). \tag{3.19}$$

The gray level corresponding to the maximum of $H_T(s)$ gives the threshold for object-background classification.

Algorithm 2

This algorithm is based on the concept of conditional entropy. Suppose *'s'* is an assumed threshold then pixels with gray level values ranging from 0 to s correspond to the background while the remaining pixels with gray levels between s+1 to L-1 constitute the object. Let t_{ij} be an entry of the quadrant B then t_{ij} gives the number of transitions such that i belongs to the background and j belongs to the object, and i and j are adjacent. Therefore,

$$\text{H(object/background)} = H(O/B) = -\sum_{i=s+1}^{L-1}\sum_{j=0}^{s} p_{ij}^{D}\log_2 p_{ij}^{D} \tag{3.20}$$

and

$$\text{H(background/object)} = H(B/O) = -\sum_{i=0}^{s}\sum_{j=s+1}^{L-1} p_{ij}^{B}\log_2 p_{ij}^{B} \tag{3.21}$$

Now the conditional entropy of the image is

$$H_T(s) = (H(O/B) + H(B/O))/2. \tag{3.22}$$

In order to get the threshold for object-background classification H_T is maximized with respect to *'s'* and that maximum value of *'s'* is the selected threshold.

3.1.3 Three dimensional co-occurrence matrix based

Melloul et al. [38] have described an algorithm similar to that in [41] with the difference based in the use of three dimensional co-occurrence matrix for thresholding and tophat morphological filter as a pre-processor. Tophat filtering removes objects which are larger than the kernel which in this case is of size 5x5. It involves morphological gray scale opening of the input image. The opened image is then subtracted from the input image. The effect of this pre-processing can be seen in the Figure 3.2. A three dimensional co-occurrence matrix of an image is an LxLxL dimensional matrix, $T = [t_{ijk}]_{LxLxL}$ which contains information regarding spatial dependency of the gray levels in image I and information about the number of transitions between three adjacent pixels, horizontally and vertically below. The information about the background and microcalcifications and appear in two distinct regions. The background region is a cube in the upper left corner of the three dimensional matrix. The region of the microcalcifications is the remaining volume. The size of these regions is dynamically adjusted by changing the position of the boundary between the two regions in the three dimensional matrix. A region in the three dimensional co-occurrence matrix defines a distribution of the gray level transitions. The entropy of each region is computed according to the boundary position. The boundary

separating background and microcalcifications is the one which gives maximum sum of the entropies. An entry t_{ijk} of the 3D co-occurrence matrix is defined by

$$t_{ijk} = \sum_{m=1}^{M} \sum_{n=1}^{N} \delta(m,n) \tag{3.23}$$

where

$$\delta(m,n) = \begin{cases} 1 & \text{if f(m,n)=i and f(m,n+1)=j and f(m+1,n)=k} \\ 0 & \text{otherwise} \end{cases} \tag{3.24}$$

The two volume objects (O) and background (B) can be separated in the 3D co-occurrence matrix defined by the equation 3.23. Let $t \in G$ be the threshold of the two groups the foreground and background in the image

$$B = \{(i,j,k) \mid 0 < i < t, 0 < j < t, 0 < k < t\}$$

$$O = \{(i,j,k) \mid t+1 < i < L-1, t+1 < j < L-1, t+1 < k < L-1\}$$

Microcalcifications are associated with the foreground and are pixels with gray levels greater than or equal to the threshold. Background is associated with the pixels which are below the threshold. The probabilities associated with each volume are defined by

$$P_B(t) = \sum_{i=0}^{t} \sum_{j=0}^{t} \sum_{k=0}^{t} p_{ijk} \tag{3.25}$$

$$P_O(t) = \sum_{i=t+1}^{L-1} \sum_{j=t+1}^{L-1} \sum_{k=t+1}^{L-1} p_{ijk} \tag{3.26}$$

The probabilities in each volume are defined by normalization

$$p_{ijk}^{B} = \frac{p_{ijk}}{P_B} = \frac{t_{ijk}}{\sum_{i=0}^{t} \sum_{j=0}^{t} \sum_{k=0}^{t} t_{ijk}} \tag{3.27}$$

$$p_{ijk}^{O} = \frac{p_{ijk}}{P_O} = \frac{t_{ijk}}{\sum_{i=t+1}^{L-1} \sum_{j=t+1}^{L-1} \sum_{k=t+1}^{L-1} t_{ijk}}$$

The background entropy $H_B(t)$ and the object entropy $H_O(t)$ are computed on the volumes B and O which reflect the background to background transitions and object to object transitions respectively. The entropies of

each are defined by

$$H_B(t) = -\frac{1}{2}\sum_{(i,j)\in B} p_{ij}^B \log_2 p_{ij}^B \tag{3.28}$$

$$H_O(t) = -\frac{1}{2}\sum_{(i,j)\in O} p_{ij}^O \log_2 p_{ij}^O \tag{3.29}$$

Note that H_B and H_O are determined by the threshold, thus they are a function of t. By summing up the entropies of the object and the background, the image entropy can be obtained by

$$H(t) = H_B(t) + H_O(t). \tag{3.30}$$

This algorithm selects a threshold which maximizes H over t. The optimal threshold is value of t which maximizes the image entropy.

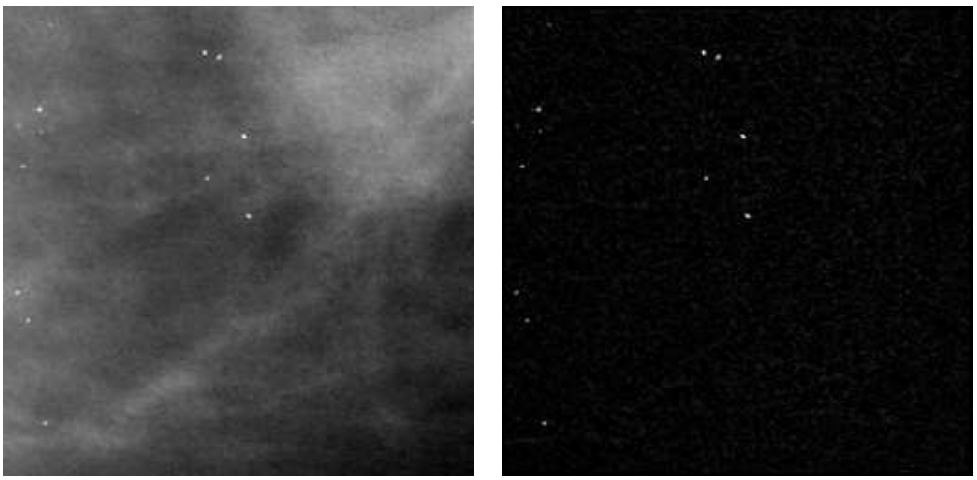
3.1.4 Experiments

Experimental comparison of entropy thresholding algorithms

Here we have compared three-dimensional entropy thresholding in [38], two dimensional algorithms in [41] and minimum cross entropy based thresholding in [33]. Figure 3.2 shows the 12-bit test input image before and after filtering with the morphological tophat filter mentioned in [38]. Both images are gray scale and microcalcifications are easily visible in the tophat filtered image, thus it appears to be an easy job to segment microcalcifications from the background. For all the algorithms the tophat filtered image is pre-thresholded with the same threshold (= 75) and first few values in the histogram/co-occurrence matrix are ignored. This is because background occupies a major portion of the mammogram and it gives rise to peaks in the early part of the histogram which clouds peaks due to breast anatomy.

Figure 3.3 shows output with two algorithms from [41] when the input image is scaled to 8-bit. As already mentioned in their paper algorithm-2 (section 3.1.2) outperforms algorithm-1 (section 3.1.2). For comparison Figure 3.4 shows the output for both algorithm-1 and algorithm-2 when images are not scaled. This goes to show that both the algorithms are highly sensitive to scaling. This could be because the co-occurrence matrix is highly sparse in the unscaled case.

In Figure 3.6 we can see a comparison between the three-dimensional dimensional co-occurrence matrix based algorithm in [38] and cross entropy based method in [33] for unscaled 12-bit images. In Figure 3.5 we can see the results with test input image scaled to 8-bit. It is clearly apparent that the algorithm in [33] outperforms both [41] and [38] for scaled as well as unscaled case. This shows a limitation of co-occurrence



(a) Input image for comparison (b) Input image after tophat filtering

Figure 3.2: Input image for comparison of thresholding algorithms

Parameters	Fraction of true positives detected	False positives per image
$t = 25$	1	13120.5
$t = 50$	0.98	8180.32
$t = 75$	0.87	2742
$t = 100$	0.73	373
$t = 125$	0.48	11.73
$t = 150$	0.4	5.91
$t = 175$	0.37	3.45
$t = 200$	0.35	2.45

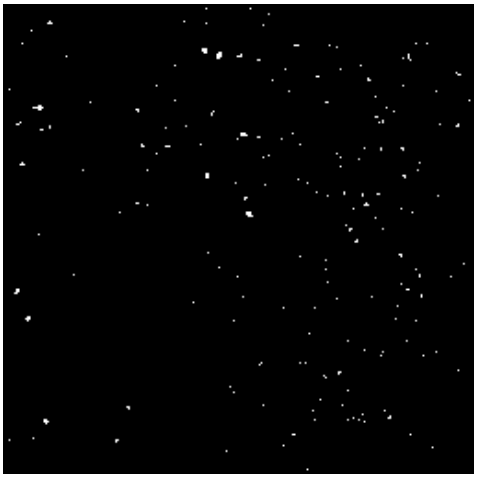
Table 3.1: Results with cross entropy based approach with variation of pre-threshold $'t'$

based methods which fail in case of images with greater brightness resolution. This can probably be attributed to the sparsity of the co-occurrence matrix with increasing grayscale resolution. Increased resolution will also result in an almost flat histogram.

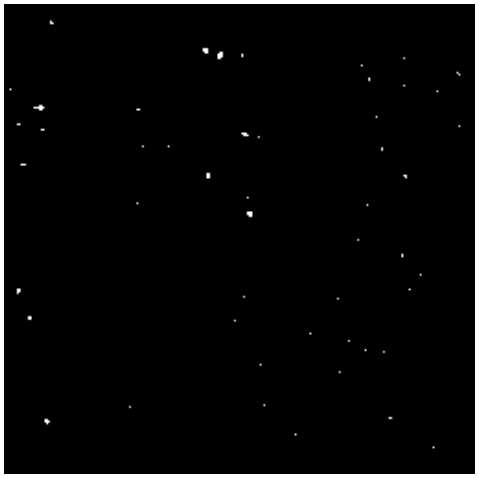
Experiments with the cross entropy based thresholding algorithm

We chose the cross entropy based thresholding algorithm to be compared with other approaches mentioned in this work since it worked best as shown in section 3.1.4. Here too we pre-threshold the image after pre-processing (in this case a tophat morphological filter) with a pre-threshold $'t'$. In section 2.1.5 also we threshold the image after filtering with the adaptive wavelet filter. This threshold is varied to obtain various performance points. Table 3.1 shows its performance as we vary the pre-threshold $'t'$. This is again done to avoid peaks due to the background which could be peaks due to the breast anatomy.

If we compare its performance with the wavelet filtering based algorithm in Table 2.1 then we can see

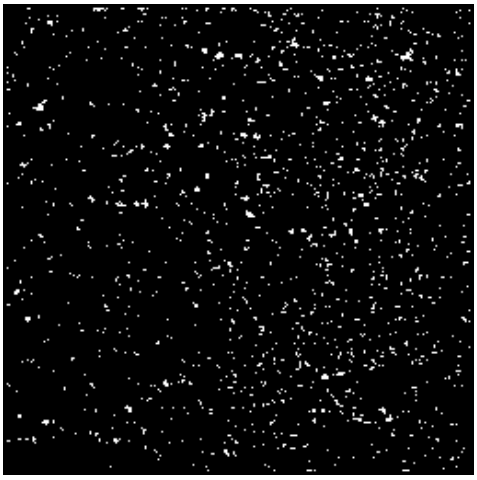


(a) Output with two dimensional co-occurrence matrix based thresholding for algorithm 1



(b) Output with two dimensional co-occurrence matrix based thresholding for algorithm 2

Figure 3.3: Thresholding the image in Figure 3.2 using algorithms in [41]

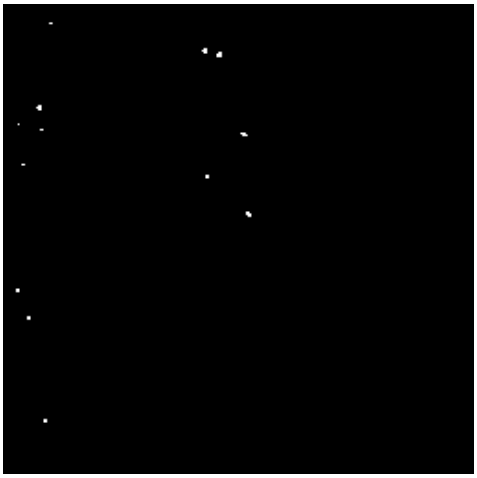


(a) Output with two dimensional co-occurrence matrix based thresholding for algorithm 1 without scaling

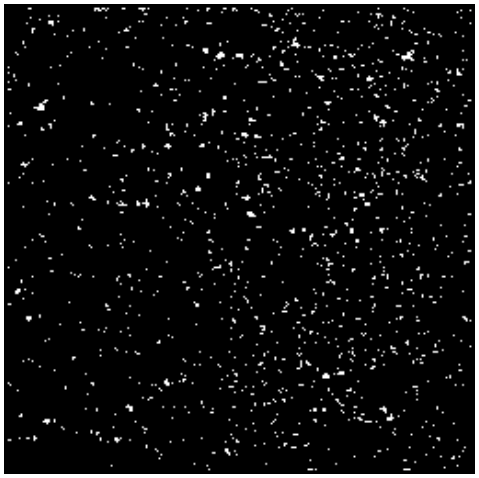


(b) Output with two dimensional co-occurrence matrix based thresholding for algorithm 2 without scaling

Figure 3.4: Thresholding the image in Figure 3.2 using algorithms in [41]

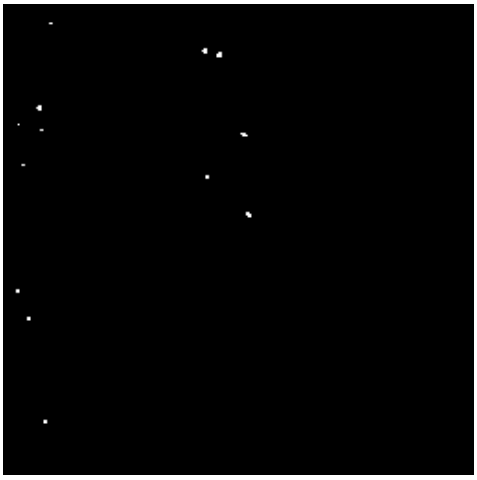


(a) Output with cross entropy based thresholding

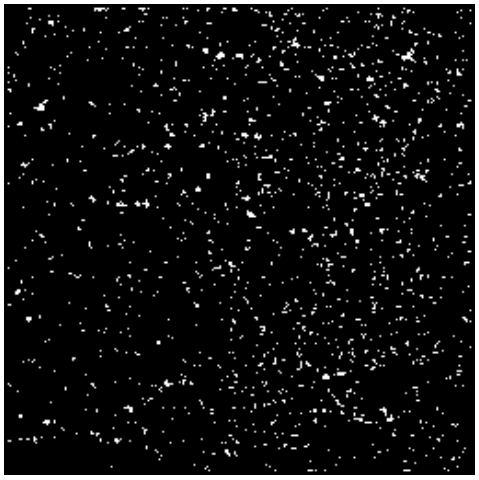


(b) Output with three dimensional co-occurrence matrix based thresholding

Figure 3.5: Thresholding the image in Figure 3.2 with scaling between 0-255



(a) Output with cross entropy based thresholding without scaling



(b) Output with three dimensional co-occurrence matrix based thresholding without scaling

Figure 3.6: Thresholding the image in Figure 3.2 without scaling



Figure 3.7: Output with cross entropy based approach for image which gave worst performance with the wavelet approach. False positives = 3, False negatives = 4 (blue - true MC, red - detected MC)

that whereas wavelet filtering achieves a maximum detection percentage of 33% with 35.18 false positives per image, cross entropy based algorithm detects 48% of MCs with only 11.73 false positives per image. Also in Figure 3.7 we can see its output for the same image which gave worst performance for the wavelet based approach (Figure 2.5). It is apparent that it outperforms the wavelet approach since it has only 3 false positives and 4 false negatives as against 285 false positives and 6 false negatives with the wavelet approach. In Figure 3.8 we can see an instance when it does not work well.

This algorithm also offers the advantage of a handle in terms of pre-threshold $'t'$ which can be tuned to achieve higher detection rates, though at the cost of more false positives per image. In Figure 3.9 we can see the ROC curve for wavelet and cross entropy based methods on absolute as well as logarithmic scale for real images. ROC curve is a plot of percentage/fraction of true positives detected versus false positives per image.

We have also tested its performance on normal mammograms with simulated microcalcifications. Table

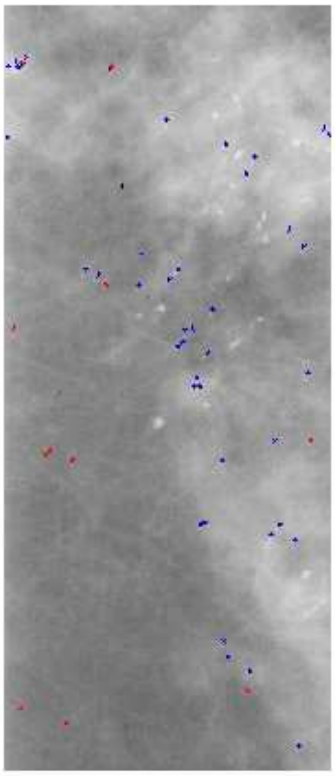


Figure 3.8: Output when cross entropy based thresholding does not work well. False positives = 8, False negatives = 41(blue - true MC, red - detected MC)

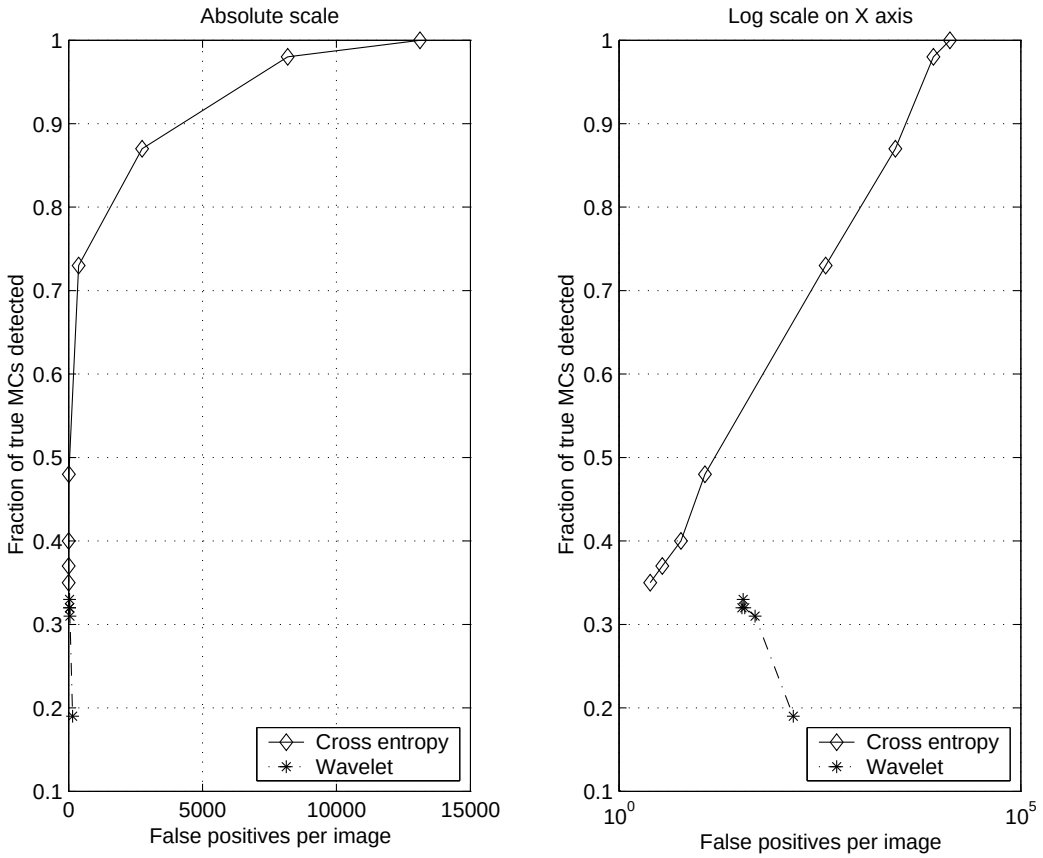


Figure 3.9: ROC curve for wavelet and cross entropy methods

Parameters	Fraction of true positives detected	False positives per image
$t = 25$	0.98	7800.5
$t = 50$	0.88	1628.4
$t = 75$	0.77	12.6
$t = 100$	0.68	6.1
$t = 125$	0.66	4.7
$t = 150$	0.62	3.7
$t = 175$	0.56	3
$t = 200$	0.5	2.7

Table 3.2: Results obtained for cross entropy based algorithm with simulated microcalcifications

3.2 shows the results obtained with these images. If we compare it with wavelet based approach in Table 2.2 then we can see that it detects 77% of microcalcifications with 12.6 false positives per image while former achieves similar performance with 26.3 false positives per image. However wavelet based method gives a detection rate of 0.82 with 32.1 false positives, cross entropy method requires 1628.4 false positives for a similar performance. Hence it will not be easy to choose between the two on the basis of their performance over simulated microcalcifications.

Chapter 4

Support Vector Machines

Burges [7] has provided an excellent tutorial for Support Vector Machines and explains the key concepts in a very lucid language. The description that follows has been taken from the same. SVM's have provided a new approach to the problem of pattern recognition with clear connections to the underlying statistical learning theory. They differ radically from comparable approaches such as neural networks: SVM training always finds a global minimum and has simple geometric interpretation. The problem which drove the initial development of SVMs occurs in several guises- the bias variance trade-off, capacity control, overfitting - but the basic idea is the same. For a given learning task with a given amount of training data, the best generalization performance will be achieved if the right balance is struck between the accuracy attained on that particular training set and the "capacity" of the machine. A machine with too much capacity is like a botanist with a photographic memory who when presented with a new tree concludes that it is not a tree because it has a different number of leaves from anything he has seen before; a machine with too little capacity is like his lazy friend who says that if it is green then it must be a tree. Neither can generalize well. The formulation and exploration of these concepts led to a path breaking development in the theory of statistical learning theory. This theory is the principle of structural risk minimization (SRM) (Vapnik, 1979). It grew out of considerations of under what circumstances, and how quickly the mean of some empirical quantity converges uniformly to the true mean (that which would be calculated from an infinite amount of data), as the number of data points increases. We will briefly try to cover this principle since this forms the backbone of SVMs.

Suppose we have a machine whose task is to learn the mapping $x_i \mapsto y_i$. The machine is actually defined by a set of possible mappings $x \mapsto f(x,\alpha)$, where the functions $f(x,\alpha)$ themselves are labeled by the adjustable

parameter α. The expectation of the test error for a trained machine is therefore:

$$R(\alpha) = \int \frac{1}{2} \mid y - f(x,\alpha) \mid dP(x,y) \tag{4.1}$$

The quantity $R(\alpha)$ is called the expected risk or just the risk and this is the quantity that we are ultimately interested in. The "empirical risk" $R_{emp}(\alpha)$ is defined to be just the measured error rate on the training set (for a fixed, finite number of observations):

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^{l} \mid y_i - f(x_i,\alpha) \mid \tag{4.2}$$

where l is the number of observations. The quantity $\frac{1}{2} \mid y_i - f(x_i,\alpha) \mid$ is called the loss. Suppose we define our problem such that the loss can only be either 0 or 1. Now choose some η such that $0 \leq \eta \leq 1$. Then for losses taking these values with probability $1-\eta$, the following bound holds (Vapnik, 1995):

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\left(\frac{h(\log 2l/h) + 1) - \log(\eta/4)}{l} \right)} \tag{4.3}$$

where h is a non-negative integer called the Vapnik Chervonenkis (VC) dimension and is a measure of the notion of capacity above. The second term on the right in is called the "VC confidence". It is a monotonically increasing function of h for any value of l. Given some selection of learning machines whose empirical risk is zero, one wants to choose that machine whose associated set of functions has minimal VC dimension.

We can now summarize the principle of SRM. The VC confidence term in equation (4.3) depends on the chosen class of functions whereas the empirical risk and actual risk depend on the one particular function chosen by the training procedure. We would like to find that subset of the chosen functions, such that the risk bound for that subset is minimized. One cannot arrange things so that the VC dimension varies smoothly, since it is an integer. Instead introduce a "structure" by dividing the entire class of functions into nested subsets. For each subset we must be able to compute h, or to get a bound on h itself. SRM then consists of finding that subset of functions which minimizes the bound on the actual risk. This can be done by simply training a series of machines one for each subset where for a given subset the goal of training is simply to minimize the empirical risk. One then takes that trained machine in the series whose sum of empirical risk and VC confidence is minimal.

Bazzani et al. [3] have investigated the feasibility of using an SVM (support vector machine) classifier in their automatic system for the detection of clustered microcalcifications in mammograms. They compared the SVM classifier with an MLP (multi-layer perceptron) in the false-positive reduction phase of their detection

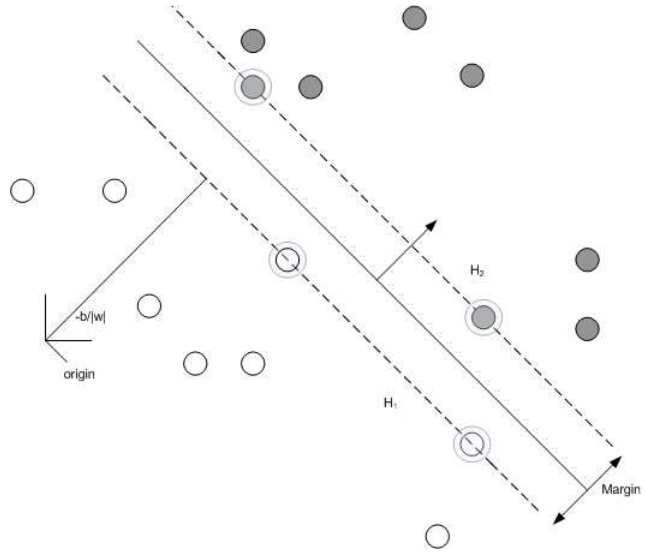


Figure 4.1: Linear support vector classifier with separable data. Support vectors are circled. (From [7])

scheme where a detected signal is considered either microcalcification or false signal, according to the value of a set of its features. Here SVM performs slightly better than MLP and even better in case of reduced training set. The drawback of this scheme is that it relies on five features of MCs viz area, average pixel value, edge gradient, degree of linearity and average local gradient in the false positive reduction step. Campaninni et al. [8] use wavelet representation followed by SVM to detect masses in mammograms. A multi-resolution overcomplete wavelet representation is performed, in order to codify the image with redundancy of information. A cascade of two SVM classifiers is used in the false positive reduction step. To further reduce false positives they use an ensemble of experts (classifiers) and final decision is made by majority vote. El-Naqa et al. [16] have for the first time reported the use of SVM for detection of MCs with minimal pre-processing. They filter the image with a high pass filter in order to reduce the effect of slowly varying background. SVM classifier is then trained by taking samples from the input image. Their proposed SVM framework has been shown to outperform four other schemes: 1) image difference technique(IDT) [39] 2) DoG method [14] 3) wavelet decomposition(WD) [51] 4) two-stage multilayer neural network (TMNN) method [56]. They achieved a sensitivity of 94% by the SVM method at an error rate of one false-positive cluster per image. Hence we decided to use the algorithm mentioned in [16] in our experiments.

4.1 Review of support vector machine classifiers

4.1.1 Linear SVMs

SVMs are best understood in the linear case. We assume that the training data is linearly separable i.e. there exists a hyperplane of the form $w \bullet x + b = 0$,where w is normal to the hyperplane, $\mid b \mid / \parallel w \parallel$ is the perpendicular distance from the hyperplane to the origin, $\parallel w \parallel$ is the Euclidean norm of w and $y_i \in [+1, -1]$ is class label or identifier. Since this hyperplane separates the data we can say:

$$w \bullet x_i + b \geq 1 \text{ for } y_i = +1 \tag{4.4}$$

$$w \bullet x_i + b \geq 1 \text{ for } y_i = -1 \tag{4.5}$$

These can be combined into one set of inequalities:

$$y_i(w \bullet x_i) + b \geq 0 \quad \forall i \tag{4.6}$$

Now consider the points for which the equality in Eq. (4.4) holds. These points lie on the hyperplane $H_1 : x_i \bullet w + b = 1$ with normal w and perpendicular distance from the origin $\mid 1 - b \mid / \parallel w \parallel$. Similarly, the points for which the equality in Eq. (4.5) holds lie on the hyperplane $H_2 : x_i \bullet w + b = -1$ with again the normal being w and perpendicular distance from the origin $\mid -1 - b \mid / \parallel w \parallel$. Hence the margin, distance between the positive and negative hyperplanes is $2/ \parallel w \parallel$. Note that H_1 and H_2 are parallel (they have the same normal) and that no training points fall between them. Thus one can find the pair of hyperplanes which gives the maximum margin by minimizing $\parallel w \parallel^2$, subject to constraints (Eq. 4.6). For a given training set, while there may exist many hyperplanes that separate the two classes, the SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes (Figure 4.1) . This hyperplane can be found by minimizing the margin given by

$$J(w) = \frac{1}{2} w^T w = \parallel w \parallel^2 \tag{4.7}$$

The description above assumes that the data is separable however it does not hold when applied to non-separable data. We can apply the same formulation to this situation by the introduction of positive slack

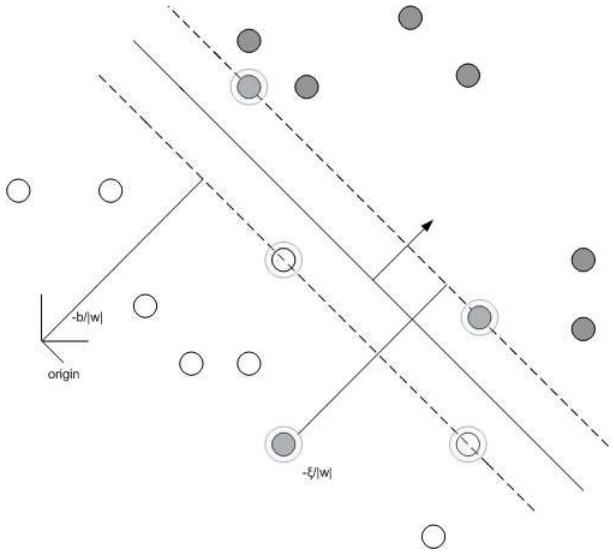


Figure 4.2: Linear support vector classifier with non-separable data. Support vectors are circled. (From [7])

variables $\xi_i, i = 1, \ldots, l$ in the constraints. Now the equations (4.5) and (4.4) look like:

$$\begin{aligned} x_i \bullet w + b &\geq +1 - \xi_i \qquad \text{for } y_i = +1 \\ x_i \bullet w + b &\leq -1 - \xi_i \qquad \text{for } y_i = -1 \\ \xi_i &\geq 0 \forall i \end{aligned} \tag{4.8}$$

Thus for an error to occur the corresponding ξ_i must exceed unity, so $\sum_i \xi_i$ is an upper bound on the number of training errors. Hence the objective function is modified to $\| w \|^2 /2 + C(\sum_i \xi_i)^k$ where C is a parameter to be chosen by the user, a larger C corresponding to assigning a higher penalty to errors. This situation can be visualized in Figure 4.2.

4.1.2 Nonlinear support vector machines

In section 4.1.1 we assume that the decision function is a linear function of the data. So how do we deal with the case when it is not so obvious (Figure 4.2)? This is accomplished by mapping the data to some other (possibly infinite dimensional) Euclidean space using a mapping

$$\Phi : R^d \mapsto \mathcal{H}$$

In the solution of Eq. (4.7) (refer [7] for details) with constraints (4.4) & (4.5); the data x_i only appears in the form of dot products $x_i \bullet x_j$; in a similar manner now it will appear in the form $\Phi(x_i) \bullet \Phi(x_j)$. Now if there is a

kernel function K such that $K(x_i, x_j) = \Phi(x_i) \bullet \Phi(x_j)$, we will only need to use K in the training algorithm and would never need to explicitly even know what Φ is. For example the radial basis kernel:

$$K(x_i, x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2}.$$

can be shown to satisfy the required properties for such a kernel. In this particular example $\mathcal{H}$ is infinite dimensional so it would not be very easy to work with Φ explicitly. However if one replaces $x_i \bullet x_j$ by $K(x_i, x_j)$ everywhere in the training algorithm, it will easily produce a support vector machine which lives in an infinite dimensional space. For example for kernels of the form $K = (x_i \bullet x_j)^p$, a dot product in $\mathcal{H}$ would require computations of order $\binom{d_L + p - 1}{p}$ operations, whereas the computation of $K = (x_i \bullet x_j)$ requires only $O(d_L)$ operations (d_L being the dimension of the data). It is this fact that allows us to construct hyperplanes in these very high dimensional spaces and yet be left with a tractable computation. Thus SVMs circumvent both forms of the "curse of dimensionality": the proliferation of parameters causing intractable complexity, and the proliferation of parameters causing overfitting.

Some kernels used in Nonlinear SVMs

An SVM is largely characterized by the choice of its kernel, thus SVM links the problems they are designed for with a large body of existing work on kernel methods. But perhaps this kernel is also its biggest limitation. Once the kernel is fixed, SVM classifiers have only one user-chosen parameter (the error penalty) but the best choice of kernel for a particular problem is still a research issue. A second limitation is speed and size, both in training and testing. Training for very large datasets (millions of support vectors) is an unsolved problem. Below are examples of kernels used for pattern recognition problems:

$$K(x, y) = (x \bullet y + 1)^p \tag{4.9}$$

$$K(x, y) = e^{-\|x - y\|^2 / 2\sigma^2} \tag{4.10}$$

$$K(x, y) = \tanh(\kappa x \bullet y - \delta) \tag{4.11}$$

Eq. (4.9) results in a classifier that is a polynomial of degree p in the data, Eq. (4.10) gives a Gaussian radial basis function classifier, and Eq. (4.11) gives a particular kind of two-layer sigmoidal neural network. Polynomial kernel function (4.9) assumes its maximum when x and y are aligned in the same direction (with their respective lengths fixed); while the radial basis function (4.10) assumes its maximum when x and y are

identical. The parameters such as σ are determined during the training phase.

Now the question arises that how does one say if there exist a pair $\{\mathcal{H}, \Phi\}$, with the properties mentioned above and for which there does not? The answer is given by the Mercer's condition: There exists a mapping Φ and an expansion

$$K(x,y) = \sum_i \Phi(x)_i \Phi(y)_i \tag{4.12}$$

if and only if, for any $g(x)$ such that $\int g(x)^2 dx$ is finite then

$$\int K(x,y)g(x)g(y)dxdy \geq 0. \tag{4.13}$$

A simple way to express the same condition is that any kernel which can be expressed as $K(x,y) = \sum_{p=0}^{\infty} c_p (x \bullet y)^p$, where c_p are positive real coefficients and the series is uniformly convergent, satisfies Mercer's condition.

4.2 Successive enhancement learning algorithm

Training a SVM classifier that operates on the whole mammogram is difficult due to the overwhelming number of negatives and very few positives (microcalcifications). El Naqa et al. [16] have described a SEL (Successive Enhancement Learning) algorithm to this purpose which selects only the most representative negative samples while keeping the training set within tolerable limits. It has been outlined below:

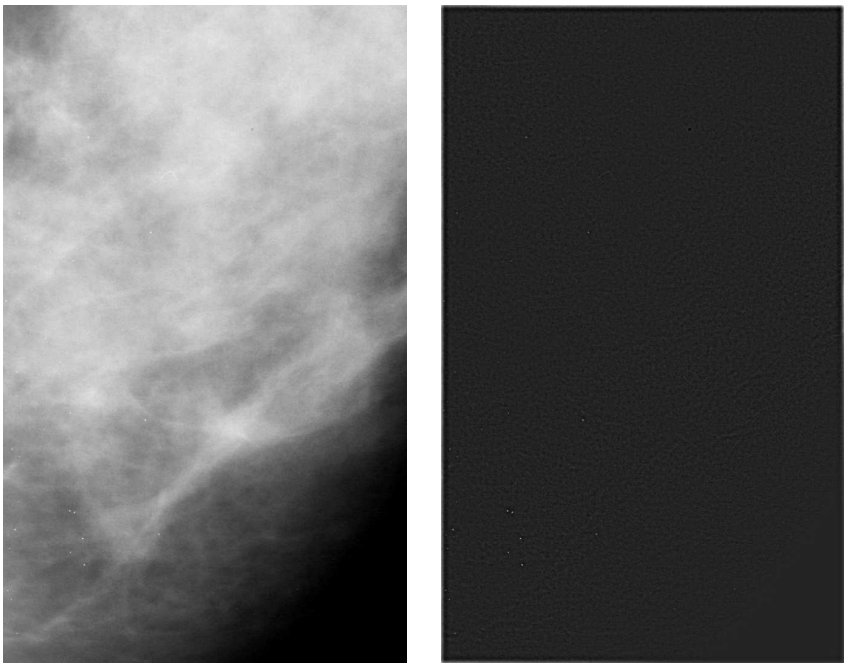
1. Extract an initial set of training examples from the available training images. Let $Z = \{(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)\}$ denote the resulting set of training samples.
2. Train a SVM classifier with Z.
3. Apply the resulting classifier to all the mammogram regions (except those in Z) in the available training images and record the "MC absent" locations that have been misclassified as "MC present".
4. Gather N new input examples from the misclassified "MC absent" locations; update the set by replacing "MC absent" examples that have been classified correctly by the classifier with the newly collected "MC absent" examples.
5. Re-train the SVM classifier with the updated set Z.
6. Repeat steps 3-5 until convergence is achieved.

The purpose of steps 3 and 4 is to identify those difficult "MC absent" examples in the training mammograms that were not included in the initial training set Z. In step 4 we used $N(= 50)$ samples which were selected

randomly from the set of misclassified “MC absent” locations. The convergence in step 6 was monitored by its error over another set of “MC absent” positions selected randomly from the training set such that they did not overlap with the existing “MC absent” samples.

4.3 Experiments

4.3.1 Training phase



(a) Input image before high pass filtering (b) Input image after high pass filtering

Figure 4.3: Effect of high pass filtering (length = 41, cut off frequency = 0.125) on mammogram

In this work as in [16] MC detection has been treated as a binary pattern recognition problem. A trained classifier is applied to each location of the mammogram and tested for presence or absence of a MC. The input image is pre-processed with a high pass filter with length 41 and cut-off frequency $w = 0.125$. An example of this operation can be seen in the Figure 4.3. The database consisted of 25 mammogram regions selected from 5 mammograms and it was divided into two sets - training and testing. The training set consisted of 13 mammogram regions with 138 MCs and test set was composed of 12 mammogram regions with 81 MCs. The test set was not used in any step of training procedure. Positive samples (pixels identified as MC present, $y_i = +1$) are collected using the truth data supplied by an experienced radiologist. In order to account for human

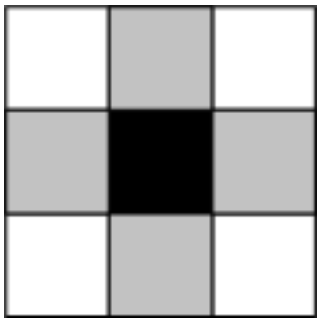


Figure 4.4: Selection of positive samples from the neighbourhood of a MC. Center pixel is shown black and extra positive locations selected are shown as gray

error we collect windows of size MxM (M = 9), centered on a 3x3 neighbourhood as shown in the Figure 4.4. Negative samples (pixels identified as MC absent, $y_i = -1$) number twice that of positive samples and were selected randomly such that they don't overlap. So for a total of 677 positive samples we collect a total of 1354 negative samples. This constitutes the initial data set Z in step 1 in the SEL algorithm (section 4.2). A SVM classifier was trained on this data with radial basis function kernel using 'Spider', a Matlab package provided by the Max Planck Institute of Biological Cybernetics, Tuebingen, Germany. Parameters obtained after training were $\sigma = 1$ and $C = 1000$ (as in Eq. 4.10). This was followed by repeating steps 3-6 in the SEL algorithm till the generalization error (measured over an exclusive set of negative samples as explained in section 4.2) was sufficiently low. Figure 4.5 shows the variation of generalization error against number of iterations with negative sample size twice of that of the positive sample size. After nine iterations it achieves an error rate of 0.61% with 467 support vectors.

We also experimented with negative sample size being ten times the size of positive samples, however it gave no significant improvement in performance along with an increase in number of support vectors to 608. Hence we decided against taking this big a sample size since with an increase in number of support vectors the testing time also increased.

4.3.2 Testing phase

We have introduced a pre-threshold $'t'$ and we only test the trained SVM classifier on pixels with gray level value greater than t. Thus by varying t we get a ROC curve for this algorithm which can then be compared with those obtained with the other algorithms. In Table 4.1 we can see the performance obtained on training and test set separately and in Table 4.2 on the overall database. A comparison of all the three methods implemented in this work can be seen in the Figure 4.6 where SVM outperforms the other methods in terms of detection rate

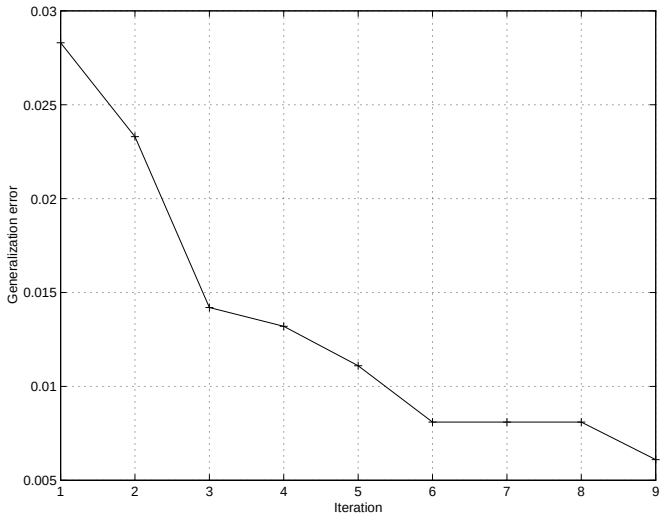


Figure 4.5: Variation of error with the number of iterations for SEL algorithm

Initial threshold	Training set		Testing set	
	TP detected	FP per image	TP detected	FP per image
$t = 3000$	0.12	21.38	0.11	30.89
$t = 2500$	0.46	89.46	0.46	109.89
$t = 2000$	0.72	138.23	0.56	158.22
$t = 1500$	0.78	162	0.6	178.78
$t = 1000$	0.81	169.38	0.68	186.33
$t = 700$	0.81	173.23	0.7	189.11

Table 4.1: Results obtained with SVM classifier on training set, test set

and false positives. However it is considerably slower than other methods. In order to compare with results mentioned previously for the image that performed worst for the wavelet method (2.5), output for that image has been shown in the Figure 4.8.

Table 4.3 shows results obtained with the SVM classifier for simulated microcalcifications. Figure 4.7 shows a comparison of ROC plots for the three algorithms for simulated microcalcifications. Here wavelet method performs the best and SVM the worst. Though cross entropy method does get quite close to wavelet's performance.

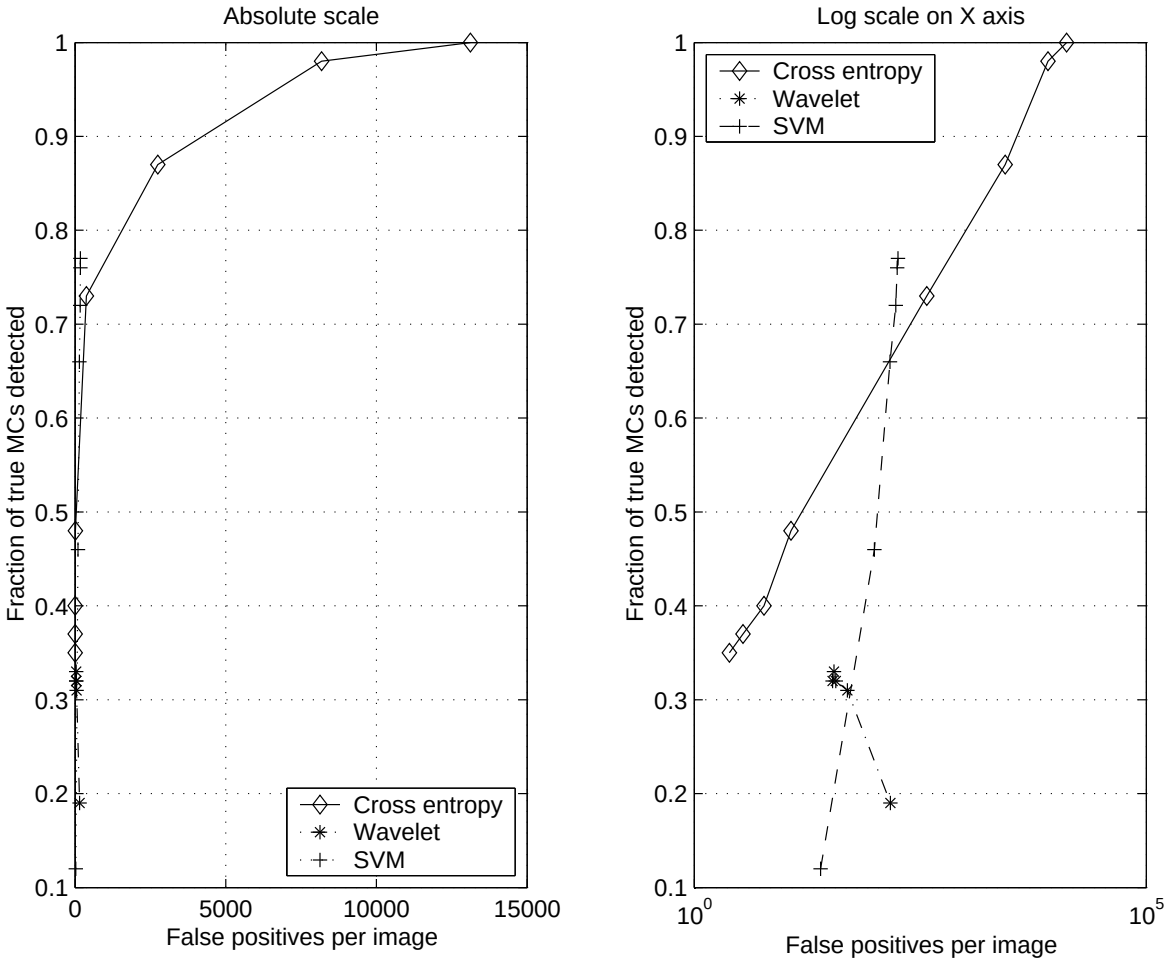


Figure 4.6: Overall ROC curve with real microcalcifications

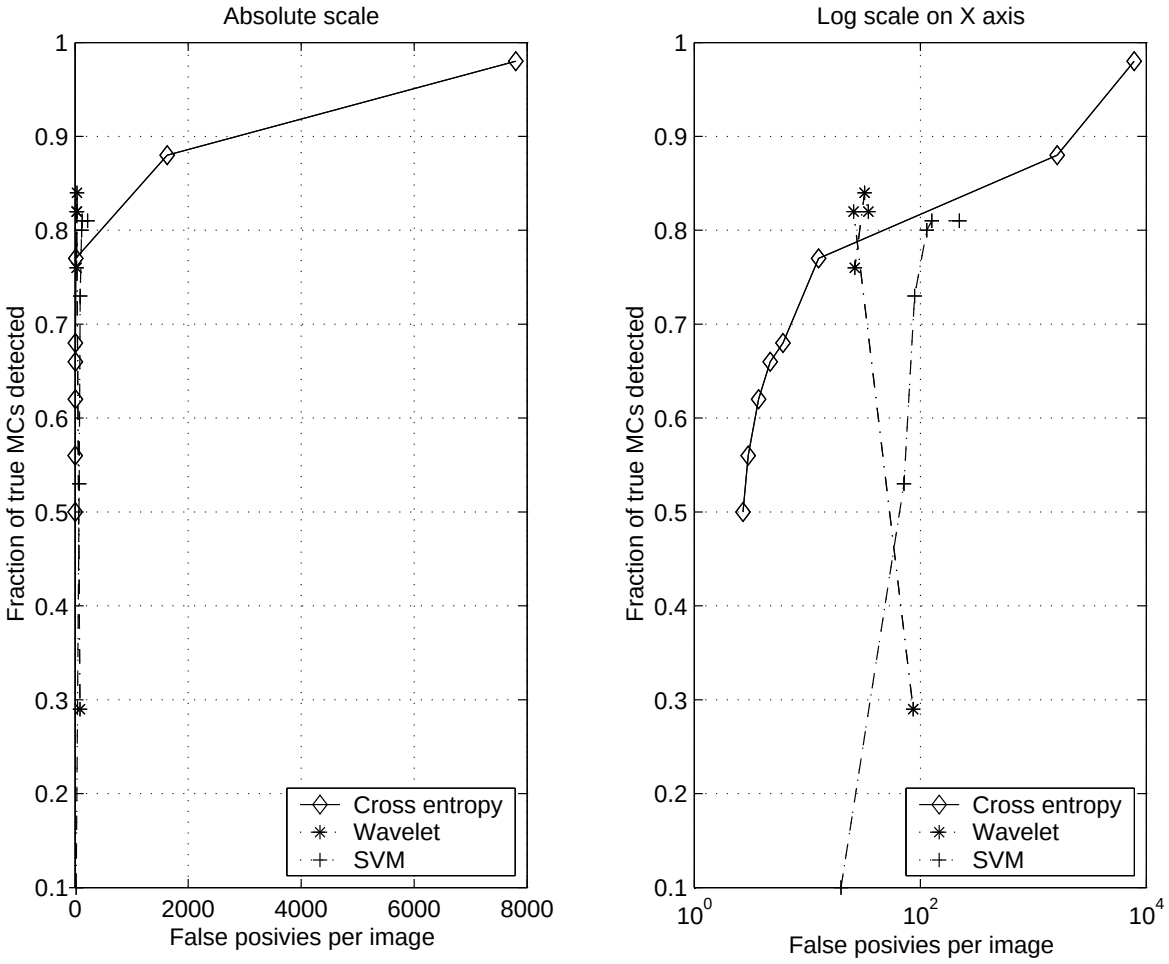


Figure 4.7: Overall ROC curve with simulated microcalcifications

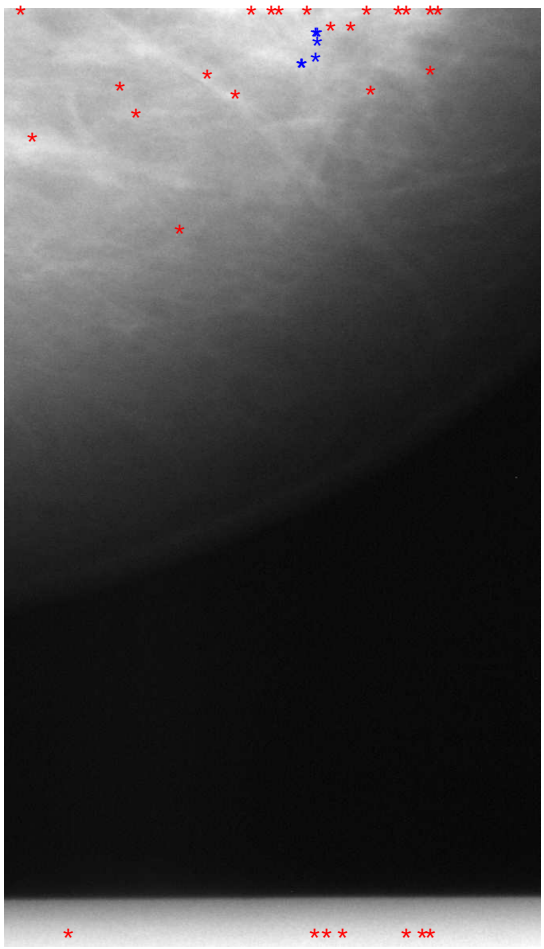


Figure 4.8: Result with SVM classifier for the image which gave worst output with wavelet approach

Initial threshold	Overall - train and test set	
	TP detected	FP per image
$t = 3000$	0.12	25.27
$t = 2500$	0.46	97.82
$t = 2000$	0.66	146.41
$t = 1500$	0.72	169.86
$t = 1000$	0.76	176.32
$t = 700$	0.77	179.73

Table 4.2: Overall results obtained with SVM classifier

Initial threshold	Simulated microcalcifications	
	TP detected	FP per image
$t = 3000$	19.88	0.1
$t = 2500$	71.63	0.53
$t = 2000$	89	0.73
$t = 1500$	114.25	0.8
$t = 1000$	126.63	0.81
$t = 700$	222	0.81

Table 4.3: Results obtained with SVM classifier for simulated microcalcifications

Chapter 5

Conclusion

In this thesis we have done an unbiased comparison of three state of the art methods for detection of microcalcifications in mammograms which are early signs of breast cancer. Only clustered presence of these calcifications indicate cancer however our objective was to only detect their presence and not to classify the mammogram as benign or cancerous.

The first algorithm is a wavelet based approach proposed by Bagci et al. [2]. The location of maxima/microcalcifications is pre-determined using an empirical method which serves as a bank of candidate locations. It uses adaptive wavelet filter first used by Ömer Nezih Gerek et al. [18] (for image compression) to filter the input image. It is observed that this wavelet image has Gaussian distribution on places other than the microcalcifications whose presence skews the distribution. HOS (Higher Order Statistics) test in [53] has been shown to be an effective measure of this skewness. So we calculate HOS values (only for locations determined by the empirical method in a previous step as explained in section 2.1.4) from the wavelet image and threshold it using the cross entropy method proposed by Li et al. in [33]. The thresholded image gives the presence of microcalcifications. Bagci et al. [2] have given the ranking of true microcalcifications based on their HOS values instead of a ROC curve or detection rate which makes it difficult to compare it with any other algorithm. Hence we introduced thresholding in the final step.

Entropy thresholding was employed as the second method. However we had to investigate a variety of algorithms before selecting any one. We started with the three dimensional co-occurrence matrix based algorithm by Melloul et al. [38] which was the first time anyone had used a 3D co-occurrence matrix. We compared it with two other 2D co-occurrence matrix by N.R. Pal, S.K. Pal [41] and a cross entropy based method by Li et al. in [33]. This investigation (section 3.1.4) has shown that 2D co-occurrence matrix based methods outperformed 3D co-occurrence matrix based method for both scaled as well as unscaled images. However the cross

entropy method outperformed co-occurrence matrix based algorithms for both scaled as well as unscaled case. Next in order to detect microcalcifications we used a graylevel tophat morphological filter used by Melloul et al. [38], which only retains objects lesser than the kernel size and then thresholded the resulting gray scale image with the cross entropy method in [33].

Support Vector Machine algorithm proposed by El-Naqa et al. [16] was used as the third algorithm for our comparison. Training a SVM classifier on mammograms to detect microcalcifications is difficult due to overwhelming number of negative samples. [16] proposed a SEL (Successive Enhancement Learning) scheme that strives to select the most representative samples without bloating the training size.

These algorithms were tested on a database of 25 mammogram regions selected from 5 mammograms containing a total of 219 MCs. A comparison of their performance can be seen in the ROC curve in Figure 4.6. SVM gave the best performance for higher detection rates while for lower detection rates cross entropy method performed the best. However the SVM method is slowest of the three and also needs to be well trained but the cross entropy and wavelet based methods do not depend on training. Wavelet based method did not perform as well as the other methods. In order to reduce the testing time for SVM it can be combined with a suitable pre-processor. One of the methods could be to use a threshold lesser than one given by the cross entropy method and then use SVM to reduce the false positives. That would significantly reduce the training set since now it does not need to classify all the pixels in the mammogram but only select candidate locations.

Bibliography

[1] Altunbasak, Y. (June 2000) “A statistical approach to threshold selection in temporal video segmentation algorithms,” *ICASSP ’00. Proceedings. IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 6 , 5-9, Pages:2421 - 2424*

[2] Bagci, A.M., Yardimci, Y., Cetin, A.E. (May 2002) “Detection of microcalcification clusters in mammogram images using local maxima and adaptive wavelet transform analysis,” *Proceedings. (ICASSP ’02). IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 4, Pages:IV-3856 - IV-3859*

[3] Bazzani A., Bevilacqua A., Bollini D., Brancaccio R., Campanini R., Lanconelli N., Riccardi A., Romani D. (2001) “An SVM classifier to separate false signals from microcalcifications in digital mammograms,” *Physics in Medicine and. Biology 46 1651-1663*

[4] Bazzani et al. (2002) “Automatic detection of clustered microcalcifications in digital mammograms using an SVM classifier,” *ESSAN’ 2000*

[5] Bocchi L, Coppini G, Nori J, Valli G. (2004) “Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks,” *Medical Engineering & Physics 26, 303312*

[6] Brink A. D. and Pendock N. E. (1994) “Minimum cross-entropy threshold selection,” *Pattern Recognition, Vol. 29, No. 1, pp. 179-188*

[7] Burges C. J. C. “A tutorial on Support Vector Machines for Pattern Recognition,”

[8] Campanini R., Dongiovanni D., Lampieri E., Lanconelli N.,Masotti M., Palermo G., Riccardi A., Roffilli M. (2004) "A novel featureless approach to mass detection in digital mammograms based on support vector machines," *Phys. Med. Biol. (49), 961-975*

[9] Caputo B., La Torre E., Gigante G. E. (2002) “Microcalcification detection using a kernel Bayes classifier,” *LNCS, ISDMA02, http://www.nada.kth.se/~caputo/publik/CaLaTorGig-isdma02.pdf*

[10] Chang C., Chen K., Wang J., Althouse M. L. G. (1994) “A relative entropy based approach to Image thresholding,” *Pattern Recognition. Vol. 27, no. 9, pp. 1275-1289.*

[11] Chang, Y., Zheng B., Good W. F., Gur D. (Oct. 1998) “Identification of Clustered Microcalcifications on Digitized Mammograms Using Morphology and Topography-Based Computer-Aided Detection Schemes: A Preliminary Experiment,” *Investigative Radiology 33(10) 746-751*

[12] Chanwimaluang T., Guoliang F. (May 2003) “An efficient blood vessel detection algorithm for retinal images using local entropy thresholding,” *ISCAS ’03. Proceedings of the 2003 International Symposium on Circuits and Systems, Volume: 5 Pages:V-21 V-24*

[13] Delbeke D., Martin WH. (2001) "Positron emission tomography imaging in oncology," *Radiol Clin North Am. 39: 883-917*

[14] Dengler J., Behrens S., Desaga J. F. (Dec. 1993) “Segmentation of microcalcifications in mammograms,” *IEEE Trans. Med. Imag., vol. 12, pp.634-642*

[15] Diyana W. M., Larcher J., Besar R. (April 2003) “A comparison of clustered microcalcifications automated detection methods in digital mammogram,” *Proceedings. (ICASSP ’03). IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 2 , 6-10 Pages:II - 385-8*

[16] El-Naqa, I., Yongyi Y., Wernick, M.N., Galatsanos N.P., Nishikawa R. (2002) "A support vector machine approach for detection of microcalcification in mammograms," *IEEE Transactions on Medical Imaging , Vol 21(12), 1552-1563*

[17] Fuhr H., Treiber O., Wanninger F. “Cluster-oriented Detection of Microcalcifications in Simulated Low-Dose Mammography,”

[18] Gerek Ö. N., Çetin A. E. (Oct. 2000) “Adaptive Polyphase Subband Decomposition Structures for Image Compression,” *IEEE Transactions on Image Processing, Vol. 9, No. 10*

[19] Gonzalez J.G., Griffith D.W. Arce, G.R. (1997) “Zero-order statistics: a signal processing framework for very impulsive processes, Higher-Order Statistics,” *Proceedings of the IEEE Signal Processing Workshop on , 21-23 July 1997 Pages:254-258*

[20] Gregory V. S., Superintendent of insurance & Antonia C. N., Commissioner of health (March 2003) "A report by the superintendent of insurance in consultation with the commissioner of health to the Governor George E. Pataki and the legislature on the advisability of a law that would expand the definition of mammography screening to include the review of X-Ray examinations by the use of a computer aided detection device,"

[21] Gulsrud T. O. “Computer-Aided Diagnosis in Digital Mammography,”

[22] Gulsrud T.O., Husoy J.H. (2001) “Optimal filter-based detection of microcalcifications,” *IEEE Transactions on Biomedical Engineering, Volume: 48 , Issue: 11 , Nov. 2001,Pages:1272 - 1281*

[23] Gurcan, M. N., Yardimci Y., Cetin A.E. (Oct. 1999) “Influence function based Gaussianity tests for detection of microcalcifications in mammogram images,” *ICIP 99. Proceedings. International Conference on Image Processing, Volume: 3 , 24-28 Pages:407 - 411*

[24] Halkitis S., Mantas J., Botsis T. “Computer-aided detection of clustered microcalcifications in digital mammograms

[25] HongQing Z. (May 2004) “Segmentation of blood vessels in retinal images using 2D entropies of gray level-gradient co-occurrence matrix,” *Proceedings. (ICASSP ’04). IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 3 , 17-21 Pages:iii 509-12*

[26] Hosseini A. E. H., Saradjian M.R., Amini J., Lucas C. “Generalized Co-occurrence Matrix to Classify IRS-1D Images using Neural Network,“

[27] http://ww3.komotv.com/Global/story.asp?S=1230164

[28] Kurani Arati S.,Xu Dong-Hui, Furst Jacob Raicu Daniela Stan (2004) “Co-occurence matrices for volumetric data,” *http://www.cra.org/Activities/craw/ dmp/awards/2004/Xu/Papers/iasted_06_01.pdf*

[29] Lado M. J., Tahoces P. G., Souto M., Mendez A.J., Vidal J. J. (Sep 1997) “Real and simulated clustered microcalcifications in digital mammograms. ROC study of observer performance,” *Medical physics 24(9) , 1385-94*

[30] Lee S., Lo C.,Wang C., Chung P., Chang C., Yang C., Hsu P. (2000) “A computer-aided design mammography screening system for detection and classification of microcalcifications,” *International Journal of Medical Informatics, 60, 29-57*

[31] Lefebvre F., Benali H., Gilles R., Di Paola R. (Dec 1994) “A simulation model of clustered breast microcalcifications,” *Medical physics 21(12), 1865-74*

[32] Lemaur G., Drouiche, K., DeConinck J. (March 2003) “Highly regular wavelets for the detection of clustered microcalcifications in mammograms,” *IEEE Transactions on Medical Imaging , Volume: 22 , Issue: 3,Pages:393-401*

[33] Li C.H., Tam P.K.S. (2003) “An iterative algorithm for minimum cross entropy thresholding,” *www.microserf.org.uk/academic/archives/MIUA2003_CJR.pdf*

[34] Li X., Ramachandran R., He M., Movva S., Rushing J., Graves S., Lyatsky W., Tan A. (April 2004) “Comparing Different Thresholding Algorithms for Segmenting Auroras,” *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004 594- 601 Vol.2*

[35] Masek M., Chandrasekhar R., DeSilva C.J.S., Attikiouzel Y. (Nov. 2001) “Spatially based application of the minimum cross-entropy thresholding algorithm to segment the pectoral muscle in mammograms,” *Intelligent Information Systems Conference, The Seventh Australian and New Zealand 2001 , 18-21 Pages:101-106*

[36] Mata R., Nava E., Sendra F. (Sept. 2000) “Microcalcifications detection using multiresolution methods,” *Proceedings. 15th International Conference on Pattern Recognition,Pages:344 - 347 vol.4*

[37] McLoughlin, K.J.; Bones, P.J.; Karssemeijer, N. (March 2004) “Noise equalization for detection of microcalcification clusters in direct digital mammogram images,” *IEEE Transactions on Medical Imaging, Volume: 23(3) ,313 - 320*

[38] Melloul M., Joskowicz L. (2002) “Segmentation of microcalcification in X-ray mammograms using entropy thresholding,” *CARS 2002*

[39] Nishikawa R. M., Giger M. L., Doi K., Vyborny C. J., Schmidt R. A., (1995) “Computer aided detection of clustered microcalcifications in digital mammograms,” *Med. Biol. Eng. Compu., vol. 33, pp. 174-178*

[40] Orel S. G. (2000) “MR imaging of the breast,” *Radiol Clin North Am. 38: 899-913*

[41] Pal N.R., Pal S.K. (1989) “Entropic thresholding,” *Signal Process. 16, 97-108*

[42] Park J. M., Choi H. K., Bae S. J., Lee M.S., Ahn S.H., Gong G. (2000) “Clustering of microcalcifications revisited,” *Clinical Radiology, 55, 114-118*

[43] Rodriguez-Sanchez R., Garcia J.A., Fdez-Valdivia J., Fdez-Vidal X. (Sept. 2000) “How to define the notion of microcalcifications in digitized mammograms,” *Proceedings. 15th International Conference on Pattern Recognition, Volume: 1 , 3-7 Pages:494 - 499*

[44] Rollan J. P., Strickland R. N. (Dec 1997) "An approach to the synthesis of biological tissue," *Vol. 1, No. 13 Optics Express*

[45] Rose C. J. and Taylor C. J. (2003) “A Generative Statistical Model of Mammographic Appearance,” *www.microserf.org.uk/academic/archives/MIUA2003_CJR.pdf*

[46] Rose C. J. and Taylor C. J. (2004) “A Statistical Model of Texture for Medical Image Synthesis and Analysis,” *www.microserf.org.uk/academic/ archives/cjr-miua-2004.pdf*

[47] Sahoo P.K., Slaaf D.W., Albert T.A. (1997) “Threshold Selection Using a Minimal Histogram Entropy Difference,” *Optical Engineering, Vol.36, No.7, pp.1976- 1981*

[48] Sajda P., Spence C., Lucas P. (Aug. 2003) “A multi-scale probabilistic network model for detection, synthesis and compression in mammographic image analysis,” *Medical Image Analysis, 7, 187-204*

[49] Shannon C. E. (1948) “A Mathematical Theory of Communication,” *The Bell System Technical Journal, Vol. 27, pp. 379423, 623656, July, October*

[50] Snyder W. E., Hairong Q. (2004) “Machine Vision,” *Cambridge University Press*

[51] Strickland R.N., Hahn H. I. (Nov. 1994) “Wavelet transforms for detecting microcalcifications in mammography,” *Proceedings. ICIP-94., IEEE International Conference on Image Processing, Volume: 1 , 13-16 Pages:402-406*

[52] Susan G. Komen Breast Cancer Foundation - *http://www.komen.org*

[53] Tsatsanis M. K., Giannakis G. B. (July 1992) "Object and texture classification using higher order statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 14,pp. 773-750*

[54] Wang J., Du E. Y., Chang C., Thouin, P.D. (2002) “Relative entropy-based methods for image thresholding,” *ISCAS 2002. IEEE International Symposium on Circuits and Systems, Volume: 2 , 26-29 May 2002 Pages:II-265 - II-268*

[55] Wang T.C., Karayiannis N.B. (Aug. 1998) “Detection of microcalcifications in digital mammograms using wavelets,” *IEEE Transactions on Medical Imaging Volume 17, Issue 4, Page(s):498 - 509*

[56] Yu S., Guan L. (Feb. 2000) “A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films,” *IEEE Trans. Med. Imag., vol. 19, pp. 115-126*

[57] Yan Chang, Alan M. N. Fu, Hong Yan and Mansuo Zhao, “Comparison of Five Conditional Probabilities in 2-level Image Thresholding Based on Bayesian Formulation”

[58] Zhu H., Fu Z., Li Z. “A new image thresholding method based on relative entropy,”