Image Processing

Ph.D. THESIS

CLASSIFICATION OF MAMMOGRAMS AND DWT BASED DETECTION OF MICROCALCIFICATION

Submitted to the
Cochin University of Science And Technology
in partial fulfillment of the requirements for the award of the degree of
Doctor of Philosophy
In the Faculty of Technology

by
MINI M.G.

Under the guidance of
Dr. Tessamma Thomas

DEPARTMENT OF ELECTRONICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN, KERALA, INDIA 682022

JULY 2004

Dedicated to.........

My parents, husband & children

DEPARTMENT OF ELECTRONICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN-22

CERTIFICATE

This is to certify that this Thesis entitled ***Classification of Mammograms and DWT based Detection of Microcalcification*** ***is*** *a bonafide record of the research work carried out by* ***Smt. Mini M.G.*** *under my supervision in the Department of Electronics, Cochin University of Science And Technology. The results presented in this thesis or parts of it have not been presented for the* ***award*** *of any other degree.*

Tessamma Thomas
Dr. Tessamma Thomas
(Supervising guide)
Reader
Department of Electronics
Cochin University of Science And Technology

Cochin-22
14-07-2004

DECLARATION

I hereby declare that this Thesis entitled ***Classification of Mammograms and DWT based Detection of Microcalcification*** *is based on the original research work carried out by* me *under the supervision of* ***Dr. Tessamma Thomas*** *in the Department of Electronics, Cochin University of Science And Technology. The results presented in this thesis or parts of it have not been presented for the award of any other degree.*

Mini MG

Mini M.G.

Cochin-22
14-07-2004

Acknowledgement

I would like to express my heartfelt gratitude to my research guide Dr. Tessamma Thomas, Reader, Department of Electronics, for her guidance and support. With her constant enquiries and suggestions she has been a great source of inspiration for me.

I sincerely thank the Director, IHRD for giving me an opportunity to carry out this research work at Cochin University of Science And Technology.

Let me express my sincere gratitude to Prof. K.Vasudevan, Head of the Department of Electronics, Cochin University of Science And Technology, Prof. K.G.Balakrishnan and Prof. P.R.S. Pillai, former Heads of the Department, for extending the facilities in the department for my research work.

My sincere thanks are due to all the faculty members of the Department of Electronics, particularly, Prof. K.T.Mathew, Mr. D.Raja Veerappa, Dr. Deepu Rajan and Mr. James Kurian, for their support and help.

I am greatly indebted to Prof. K.S.M. Panicker, Principal, N.S.S. College of Engineering, Palakkad for his continuous motivation and encouragement.

Dr.M.N.N. Nambooodiri, Department of Mathematics, Cochin University of Science And Technology has helped me a lot in understanding the complex mathematical concepts associated with wavelets. Let me express my heartfelt gratitude to him.

I sincerely thank all teaching and non-teaching faculty of Model Engineering College, Thrikkakkara, especially, Prof. Jyothi John, Principal, Prof. T.K.Mani and Prof. Jacob

Thomas, Asst.Professors, Mrs.Remadevi, Lecturer in Mathematics for their timely help and support.

I would like to express my gratitude to Dr. C.S.Sreedhar, Dr.A.Unnikrishnan, NPOL and Dr. J.C. Goswami for their very much encouraging words about my research work.

Dr.Joe Jacob, Department of Physics, Newman College, Thodupuzha and Prof. V. P. Devassia, Principal, College of Engineering, Chengannur has been constant sources of motivation and encouragement through out my research work. With a great sense of gratitude I remember their valuable suggestions and selfless assistance.

I thank all the research scholars of the department, especially Mr. Dinesh Kumar V.P., Mrs. Mridula S., Mrs. Binu Paul, Mr. Vinu Thomas and Mr. Anil Lonappan for their friendly and supportive attitude.

I thank the non-teaching, library and administrative staff of the department for their cooperation and support.

I also take this opportunity to thank all the M.Tech and M.Sc students, especially Mrs. Deepa J, Lecturer, College of Engineering, Chengannur and Mr.Benoy Jose, V.S.S.C who have collaborated with me.

It is beyond words to express my gratitude to my husband, Kannan and Ammu for their help and encouragement. Without their help and sacrifice, I am sure I could not have accomplished this task. I also thank my father, mother and mother-in-law for their support and understanding.

Mini M.G.

Contents

1 Introduction 1

1.1 Digital Image Processing 2

1.1.1 Image enhancement 2

1.1.2 Image Restoration 3

1.1.3 Image compression 3

1.1.4 Image segmentation 3

1.1.5 Image description and Representation 4

1.2 Medical Image Processing 4

1.3 Tools for image processing 5

1.3.1 The Wavelet Transform 6

1.3.1.1 History of Wavelets 7

1.3.1.2 The Continuous Wavelet Transform (CWT) 7

1.3.1.3 The Discrete Wavelet Transform (DWT) 8

1.3.1.4 The Multiplexed Wavelet Transform (MWT) 8

1.3.1.5 WT in Two Dimensions 9

1.3.1.6 Computation of DWT 9

1.3.1.6.1 Sectioned computation 10

1.3.1.7 WT in Biomedical Image Processing 10

1.3.1.7.1 Computer Assisted Mammography 11

1.3.1.7.2 Computer Assisted Tomography (CAT) 12

1.3.1.7.3 Magnetic Resonance Imaging (MRI) 12

1.3.1.7.4 Functional Image analysis 12

1.3.2 Neural Network 13

1.3.2.1 Target detection 15

1.4 Objective of the work 15

1.5 Layout of the Thesis 17

2 Literature Review 19

2.1 CAD in mammography 20

2.1.1 Classification of microcalcifications into benign and malignant 25

2.1.2 Normal mammogram characterization 27

2.2 DWT Computation 29

3 Breast Cancer – A Medical Perspective 33

3.1 Anatomy of the female breast 34

3.2 Malignancy in the breast 35

3.2.1 Symptoms & Diagnosis 36

3.3 Mammography 38

3.3.1 The Mammography Machine 39

3.3.2 Breast Composition Determination 40

3.4 Normal Mammograms 41

3.5 Mammographic Abnormalities 43

3.5.1 Microcalcifications 43

3.5.1 .1 Calcifications Distribution Modifiers 47

3.5.2 Circumscribed Masses 48

3.5.2.1 Architectural Distortion 51

3.5.2.2 Asymmetric Breast Tissue 52

3.5.3 Spiculated Lesions 52

4 Review of Basic Theory 55

4.1 The wavelets 55

4.2 The CWT 57

4.3 The DWT 58

4.4 Wavelets and Time- Frequency Representation - Concept of MRA 60

4.5 DWT computation 63

4.5.1 Basic Multirate Operations 63

4.5.1.1 Decimation 63
4.5.1.2 Interpolation 64
4.5.1.3 Sampling rate conversion by a rational factor L/M 64
4.5.2 The pyramidal algorithm 66
4.6 Computation of 2-D DWT 68
4.6.1 2-D Wavelets 68
4.6.2 2-D Wavelet Transform 69
4.7 The MWT 71
4.8 Edge detection 74
4.8.1 Edge detector using wavelets 74
4.9 Artificial Neural Networks Technology 76
4.9.1 Artificial Neurons 78
4.9.2 Teaching an ANN 82
4.9.2.1 Supervised Learning 82
4.9.2.2 Unsupervised Learning 83
4.9.2.3 Learning Rates 84
4.9.2.4 Learning Laws 84
4.10 Feature Extraction for classification 86
4.10.1 Statistical descriptors 86
4.10.2 Textural features 88
4.10.2.1 SGLD Matrix 88
4.11 Networks for classification 91
4.11.1 Back Propagation Neural Network (BPNN) 91
4.11.2 Competitive network 93
4.11.3 Radial Basis Function Network (RBFN) 94
4.11.4 PNN 95

5 Development of Block DWT Computation Algorithm 99

5.1 Block –wise Computation of 1-D DWT 101

5.1.1 Truncation of transform coefficients 103

5.1.2 BDWT by Overlap Save Method 104

5.1.3 Block IDWT (BIDWT) by Overlap Add Method 106

5.2 Block –wise Computation of 2-D DWT 107

5.2.1 2-D BDWT by overlap save method 109

5.2.2 2-D BIDWT by overlap add method 112

5.3 Computational Complexity 113

5.3.1 Estimation of computational burden for standard algorithm 113

5.3.2 Estimation of computational burden for BDWT algorithm 114

5.4 Results and Discussion 116

5.5 Conclusion 125

6 Neural Network Based Classification of Mammograms 127

6.1 Normal Mammogram Characterization 128

6.2 Features of Normal Mammograms 129

6.3 Residual Image generation 130

6.3.1 Removal of normal background 130

6.3.2 Detection and removal of linear markings 132

6.4 Neural Network Training and Testing Methodology 135

6.4.1 Training and Testing Data Sets 136

6.4.2 Detection criteria 137

6.4.3 *n*-fold Cross Validation 138

6.5 Normal/abnormal classification based on statistical features 138

6.5.1 Selection of Neural Network structure for classification 138

6.5.2 Feature selection 140

6.5.3 Classification Results 142

6.6 Normal/abnormal classification based on textural features 145
6.6.1 Selection of Neural Network structure for classification 145
6.6.2 Feature selection 146
6.6.3 Classification Results 148
6.7 Normal/abnormal classification based on both statistical and textural features 150
6.8 Conclusion 151

7 Multiplexed Wavelet Transform Technique for Detection of Microcalcification 153
7.1 Detection of Microcalcification as an edge detection operation 154
7.2 Edge Detection using MWT 154
7.3 MWT based Microcalcification Detection 155
7.3.1 Microcalcification Detection after Classification 160
7.4 Data and Detection Criteria 161
7.5 Results And Discussion 162
7.6 Conclusions 166

8 Summary And Conclusions 167
8.1 Summary of the work and important conclusions 167
8.2 Scope for further investigations 169

Appendix

A Line Detection Algorithm 171
A.1 Introduction 171
A.2 Detection Algorithm 172
Bibliography 175
List of publications 199
Index 203

List of Figures

3.1 Schematic Diagram of the Female Breast 34
3.2 Normal Mammogram - Dense-glandular type 42
3.3 Normal Mammogram - Fatty type 43
3.4 Normal Mammogram -Fatty-glandular type 44
3.5 Snippets of mammograms containing Microcalcifications 45
3.6 Basic types of malignant microcalcifications. 45
3.7 Different forms of benign calcifications. 46
3.8 Snippets of mammograms with circumscribed masses 49
3.9 Benign masses. a) Halo b) cyst c) capsule 49
3.10 Malignant masses. (a) High density radiopaque. (b) Solid tumor with random Orientation 50
3.11 Snippets of mammograms with ill-defined masses 50
3.12 Snippets of mammograms with architectural distortion 51
3.13 Snippets of mammograms having asymmetric breast tissue. 52
3.14 Snippets of mammograms with spiculated masses 53
4.1 Schematic of MRA decomposition 62
4.2 Fractional sampling rate conversion by multirate technique 65
4.3 Pyramid structure for 2-level DWT computation 66
4.4 Frequency bands for the analysis tree of the pyramid 67
4.5 Time-frequency tiling 68
4.6 Coefficient layout of a 3-level DWT of an image 71
4.7 Pyramidal Structure for 2-level 2-D DWT computation 72
4.8 A simple neuron 78
4.9 A Basic Artificial Neuron 79

4.10 Sigmoid Transfer Function. 80
4.11 A Simple Neural Network Diagram. 80
4.12 Simple Network with Feedback and Competition. 81
4.13 Two distributions having same variance and skew, but different kurtosis a) Leptokurtic distribution b) Platykurtic distribution 87
4.14 A general multilayer feed-forward network 92
4.15 General competitive network architecture 93
4.16 RBF Network 94
4.17 A Probabilistic Neural Network 96
5.1 Schematic of truncation for the computation of 2-level DWT 105
5.2 One block of WT coefficients in 2-level block DWT. 106
5.3 Block DWT by overlap save method 106
5.4 Block IDWT by overlap add method 107
5.5 Effect of truncation in DWT computation on the image 'coin'. 108
5.6 Partitioning of an image into 9 overlapping blocks. 109
5.7 2- level 2-D BDWT coefficients of a single block of data. 110
5.8 Distribution of interleaved transform coefficients of BDWT 111
5.9 Overlap add reconstruction of BDWT coefficients. 112
5.10 2- level decomposition and reconstruction of a music sample using BDWT technique. 117
5.11 Comparison of transform coefficients (2- level decomposition). 118
5.12 Verification of BDWT algorithm using an ECG signal. 119
5.13 Verification of BDWT algorithm using a guitar note. 120
5.14 2-level BDWT decomposition and reconstruction of the image 'camera man' using db2. 121
5.15 Comparison of pyramidal and BDWT algorithm for 2-D DWT computation 123
5.16 Effect of change in processing frame size 124
5.17 Normalized processing delay for multiprocessor computation of BDWT for various data sizes 125

6.1 Residual image generation (background removed)
Original normal mammogram b) Residual image 131

6.2 Residual image generation (background removed)
(a) Original image containing microcalcifications b) Residual image 131

6.3 Residual image generation (background removed)
(a) Original image containing a circumscribed mass b) Residual image 132

6.4 Residual image generation (background removed)
(a) Original image containing a spiculated mass b) Residual image 132

6.5 Residual image generation (background removed)
(a) Original image containing an asymmetry b) Residual image 133

6.6 Residual image generation (background removed)
(a) Original image containing an architectural distortion b) Residual image 133

6.7 Residual image generation (background removed)
(a) Original image containing an ill-defined mass b) Residual image 134

6.8 Residual image generation (normal linear markings removed)
(a) Original normal mammogram b) Residual image 135

6.9 Residual image generation (normal linear markings removed)
(a) Original image containing a microcalcification cluster b) Residual image 135

6.10 Residual image generation (normal linear markings removed)
(a) Original image containing a circumscribed mass b) Residual image 136

6.11 Residual image generation (normal linear markings removed)
Original image containing a spiculated lesion b) Residual image 136

7.1 Image reconstruction using MWT and 2-D DWT. 156

7.2 Steps for detection and segmentation of microcalcifications 159

7.3 Comparison of detection of microcalcifications from various mammograms
using Canny and M-H detectors on MIAS database 164

7.4 Comparison of detection of microcalcifications from various mammograms
using M-H and Canny detectors on the local database 165

A.1 Block diagram of the line detector 174

List of Tables

3.1 Positioning on performing mammograms 38

5.1 Ratio of computational complexity (in terms of real multiplications) of the BDWT to that of conventional method. 124

6.1 Training and testing data sets 137

6.2 Comparison of performance of different network architectures on normal / abnormal classification of mammographic data using statistical features. 139

6.3 Sensitivity for different features derived from the original image 141

6.4 Sensitivity for different features derived from the residual image (background removed) 141

6.5 Sensitivity for different features derived from the residual image (normal lines removed) 142

6.6 Classification results for 3 sets of feature vectors 143

6.7 Results of Statistical feature based classification 144

6.8 Detailed Results of Statistical feature based classification 144

6.9 Comparison of performance of different network architectectures on normal / abnormal classification of mammographic data data using textural features. 145

6.10 Selection of the single best feature using SFS 147

6.11 Selection of the best feature set of size two using SFS 147

6.12 Selection of the best feature set of size three using SFS 147

6.13 Selection of the best feature set of size 4 using SFS 148

6.14 Selection of the best feature set of size 5 using SFS 148

6.15 Classification result using textural features 149

6.16 Classification result using textural features for different orientations 149

6.17 Detailed result of Classification for an orientation of 0^0 150

6.18 Classification result using combined set of features. 150

6.19 Detailed results of Classification using combined set of features. 151
7. 1 Details of mammograms used for validation of the algorithm 161
7. 2 Comparison of detection capability of various edge detection algorithms on microcalcification detection 162
7.3 Comparison of sensitivity and specificity of microcalcification detection using Canny and M-H detectors for different values of k. 163
7.4 Detection Sensitivity for the two databases. 163

Abbreviations

1-D - One-Dimensional
2-D - Two-Dimensional
ACR- BIRADS- American College of Radiology - Breast Imaging Reporting and Data System
ACS - American Cancer Society
ANN - Artificial Neural Network
BDWT - Block DWT
BIDWT - Block IDWT
BPNN - Back Propagation Neural Network
CAD - Computer Aided Diagnosis
CAT - Computer Assisted Tomography
CC - Cranio-Caudal
CNN - Convolution Neural Network
CV - cross-validation
CWT - Continuous Wavelet Transform
DCT - Discrete Cosine Transform
DMRA - Discrete Multi Resolution Analysis
DSP - Digital Signal Processing
DWT - Discrete Wavelet Transform
EZW - Embedded Zero-tree Wavelet
FDG - 2-fluoro-2-dioxy-D- Glucose
FN - False Negative
FP - False Positive
FWT - Fast Wavelet Transform
I/O - Input / Output

IDWT	-	Inverse Discrete Wavelet Transform
IMWT	-	Inverse MWT
KvP	-	Kilo-voltage Peak
LMS	-	Least Mean Square
LVQ	-	Learning Vector Quantization
mAs	-	milli-Ampere-seconds
MCPCNN	-	Multiple Circular Path Convolution Neural Network
M-H	-	Marr-Hildreth
MIAS	-	Mammographic Image Analysis Society
MLO	-	Medio-Lateral Oblique
MLP	-	Multilayer perceptron
MRA	-	Multi Resolution Analysis
MRI	-	Magnetic Resonance Imaging
MWT	-	Multiplexed Wavelet Transform
PET	-	Positron Emission Tomography
PMS	-	Parallel Multiple Subsequence
PNN	-	Probabilistic Neural Network
QMF	-	Quadrature Mirror Filterbanks
RBF	-	Radial-Basis-Function
RBFN	-	Radial Basis Function Network
RBST	-	Rubber Band Straightening Transform
RGI	-	Radial Gradient Index
ROI	-	Region Of Interest
RPA	-	Recursive Pyramidal Algorithm
SFFS	-	Sequential Forward Floating Search
SFS	-	Sequential Forward Search
SGLD	-	Spatial Gray-Level Dependence
SPIHT	-	Set Partitioning In Hierarchical Trees
SSWT	-	Spatially Segmented Wavelet Transform

STFT	-	Short-Time Fourier Transform
TP	-	True Positive
WT	-	Wavelet Transform

Chapter 1

Introduction

Breast cancer is the second most common malignancy that affects women worldwide and is the leading cause among non-preventable cancer death [1]. The American Cancer Society (ACS) estimates that on an average, in every 15 minutes five women are diagnosed with breast cancer. It is also estimated that one in eight women will be diagnosed with this disease in her lifetime, and 1 in 30 will die from it [2]. Breast cancer is the second most prevalent cancer among Indian women, the first being cervical cancer [3]. In the age group of 30-70 years, one in fifty eight women are affected by this disease and the occurrence is mainly seen in the urban areas.

Mammography is the best technique for reliable detection of early, non-palpable, potentially curable breast cancer [4]. As a result of the increasing utilization of mammographic screening, the mortality rate due to this disease was observed to decrease for the first time in 1995 [5]. Since the interpretation of mammograms is a repetitive task that requires much attention to minute details, the opinion of radiologists may vary. To overcome this difficulty, during the past decade, the use of image processing techniques [6], [7], [8], [9], [10] for Computer Aided Diagnosis (CAD) in digital mammograms has been initiated. This has increased diagnostic accuracy as well as the reproducibility of mammographic interpretation.

1.1 Digital Image Processing

Digital image processing is a rapidly evolving field with growing applications in the fields of science and Engineering. Interest in Digital Image Processing stems from two principal application areas: improvement of pictorial information for human interpretation and processing of scenic data for machine perception. It finds application in a wide range of areas like image transmission and storage for remote sensing via satellites, automated inspection of industrial parts, industrial machine vision for product assembly, automatic character recognition, automatic processing of finger prints, RADAR, SONAR and acoustic image processing, Medical image processing etc.

Images have their information encoded in the spatial domain. In other words, features in images are represented by edges, not by sinusoids. Hence, the spacing and number of pixels are determined by how small a feature need to be seen, rather than by the formal constraints of the sampling theorem. A digital image can be considered as a matrix whose row and column indices identify a point in the image and the corresponding matrix element value identifies the gray level at that point.

Processing of digital images involve procedures that are usually expressed in algorithmic form. Thus with the exception of image acquisition and display, most image processing functions can be implemented in software. Transforms are the fundamental tools that are used in most of the image processing applications. The wavelet based multiresolution analysis is found to be one of the best tools for this.

The various realms of image processing are briefly described below [11], [12]:

1.1.1 Image enhancement

The principal objective of image enhancement techniques is to process a given image to make it more suitable than the original for some specific application. These techniques do not increase the inherent information content in the data but emphasize certain image characteristics. Enhancement is useful for feature extraction, image analysis and display of visual information. The enhancement techniques fall into two broad categories:

frequency domain methods and spatial domain methods. The former is based on the modification of the Fourier Transform of an image and the latter refers to the direct manipulation of pixels in an image. Image enhancement operations include contrast and edge enhancement, pseudo coloring, sharpening, magnifying and noise filtering.

1.1.2 Image Restoration

Image restoration is the process that reconstructs or recovers a degraded image, using some *apriori* knowledge of the degrading phenomenon. The ultimate goal of restoration is to improve a given image in some sense, as in image enhancement. The difference between enhancement and restoration is that the former is concerned with accentuation and extraction of image features while the latter restores degradations.

1.1.3 Image compression

Digital representations of images usually require a very large number of bits. In many applications it is important to consider techniques for representing an image or the information contained in it using fewer number of bits. Image compression addresses this problem. Image data compression methods fall into two categories: Predictive coding and Transform coding. In predictive coding compression is achieved by exploiting the redundancy of the data. Techniques such as delta modulation, differential pulse code modulation etc. fall into this category. In transform coding the given image is transformed into another domain such that a large amount of information is packed into a small number of samples. The compression process inevitably results in some distortion due to the removal of relatively insignificant information.

1.1.4 Image segmentation

Image segmentation is an essential preliminary step in most automatic pictorial pattern recognition and scene analysis problems. It is the process that subdivides an image into its constituent parts or objects. The concept of segmenting an image is generally based on the similarity or discontinuity of the gray level values of its pixels and can be applied to both static and dynamic images.

1.1.5 Image description and representation

Representation and description of objects or regions of interest, that have been segmented out of an image are the initial steps in the operation of most automated image analysis systems. After segmentation, the resulting aggregates of pixels are represented and described in a form suitable for further processing. Generally, an external representation is chosen when the primary focus is on morphological features. When one is interested in reflectivity properties such as color and texture, an internal representation is selected. The choice is dictated by the problem under consideration, so as to capture the essential differences between objects or class of objects, maintaining as much independence as possible to changes in factors such as location, size and orientation.

1.2 Medical Image Processing

The advent of medical imaging is one of the milestones in the progress of medical science. It serves as a beneficial tool for the medical practitioners during diagnosis of ailments. The application of image processing techniques to medical imaging has made the results accurate and reliable. In many cases it is possible to eliminate the necessity for invasive surgery, thus avoiding trauma to the patient as well as an inevitable element of risk.

One of the early applications of image processing in the medical field is the enhancement of conventional radiograms. When converted to digital form, it is possible to remove noise elements from X-ray images, thereby enhancing their contrast. This aids interpretation and removes blurring caused by unwanted movement of the patient. This form of representation also enables the physicians to measure the extent of tumors and other significant features accurately.

The basic image processing operations on medical images are conveniently placed in four categories: filtering, shape modeling, segmentation and classification [13]. Filtering includes linear and non-linear enhancement, deblurring and edge detection techniques using local operators or classification techniques. Shape modeling includes three-dimensional representation and graphics manipulation such as three-dimensional contours of the spinal column, coronary artery or shaded images. Clustering, object detection, and boundary detection are the main operations that come under segmentation. Simple histogram or thresholding techniques are used to segment objects of interest. When adequate prior information is available matched filters can be used effectively. Heuristic techniques are useful for tracing contours in the presence of highly structured background such as chest radiographs. Feature selection, texture characterization and pattern recognition are the major operations in classification. [14], [15].

Another application of digital image processing in medical imaging is 'tomography', the generation of images of a slice through the body [16] involving the reconstruction of two-dimensional images.

1.3 Tools for image processing

The first step after obtaining the image in any digital image processing system is preprocessing that image. The key function of this is to improve the image in ways that increase the chances of success of other processes. Wavelet Transform (WT) techniques are found to be a very effective processing tool for this purpose.

Neural Networks are found to be efficient tools for classification applications. They are rough models of human mental processes with powerful learning, memorization, and associative recall capabilities of pattern formatted information.

A brief introduction to these two image processing tools are provided in the sections below:

1.3.1 The Wavelet Transform

Perhaps the most prominent signal analysis technique is Fourier analysis, which breaks down a signal into its constituent sinusoids of different frequencies or transforms our view of the signal from a time-based one to a frequency-based one. But, This has the serious drawback of loss of time information while transforming into the frequency domain. This is not very prominent for stationary signals. However, Fourier analysis become inadequate when the local frequency contents of the signal are of interest or when it contains non-stationary or transitory characteristics like drift, trends, abrupt changes, etc.

In an effort to correct this, Dennis Gabor [17] adapted the Fourier transform to analyze only a small section of the signal at a time — a technique called windowing the signal. Gabor's adaptation, called the Short-Time Fourier Transform (STFT), maps a signal into a two-Dimensional (2-D) function of time and frequency. While the STFT's compromise between time and frequency information can be useful, the drawback is that once a particular size is chosen for the time window, it remains the same for all frequencies.

Wavelet analysis, a windowing technique with variable-sized regions, represents the next logical step. It allows the use of long time intervals where more precise low frequency information is needed and shorter intervals where high frequency information is needed. One major advantage offered by wavelets is the ability to analyze a localized area of a larger signal. Further, because it offers a different view of data than those presented by traditional techniques, wavelet analysis can often compress or de-noise a signal without appreciable degradation. Indeed, in their brief history within the signal processing field, wavelets have already proven themselves to be an indispensable addition to the analyst's collection of tools and continue to enjoy a burgeoning popularity today.

Wavelets are oscillatory functions that exist for a few cycles only and satisfy certain properties. Most of the wavelets are associated with a scaling function. There are various kinds of wavelets like compactly supported wavelets, symmetric and non-symmetric wavelets, orthogonal and biorthogonal wavelets and smooth wavelets.

1.3.1.1 History of Wavelets

From a historical point of view, wavelet analysis is a new method, though its mathematical underpinnings date back to the work of Joseph Fourier in the nineteenth century [18]. Fourier laid the foundations of frequency analysis with his theories, which proved to be enormously important and influential. When it became clear that an approach measuring average fluctuations at different scales might prove less sensitive to noise, the attention of researchers gradually turned from frequency-based analysis to scale-based analysis. The first recorded mention of the term "wavelet" was in 1909, in a thesis by Alfred Haar [19]. Morlet and the team working under Alex Grossmann at the Marseille Theoretical Physics Center in France first proposed the concept of wavelets in its present theoretical form [20]. The main algorithm for WT computation dates back to the work of S. Mallat in 1988 [21]. Since then, research on wavelets has become international and is particularly active in the United States, spearheaded by veteran scientists Ingrid Daubechies, Ronald Coifman, and Victor Wickerhauser [22].

1.3.1.2 The Continuous Wavelet Transform (CWT)

The WT of a signal represents the signal as a linear combination of scaled and shifted versions of the wavelets and scaling functions. When the scale and shift parameters are continuous, the transform under consideration is called a CWT. In the CWT a function ψ, which in practice looks like a little wave, is used to create a family of wavelets $\psi(at + b)$ where a and b are real numbers, a dilating (compressing or stretching) the function ψ and b translating or displacing it. The word continuous refers to the transform, not to the wavelet. The CWT turns a signal $f(t)$ into a function $W_\psi f$ of two variables, scale and time as:

$$W_\psi f(a,b) = |a|^{-1/2} \int_{-\infty}^{\infty} f(t)\psi^*(at+b)dt \qquad (1.1)$$

where ψ^* is the complex conjugate of ψ. This transformation in theory is infinitely redundant, but it can be useful in recognizing certain characteristics of a signal.

1.3.1.3 The Discrete Wavelet Transform (DWT)

The CWT maps a signal of one independent variable t into a function of two independent variables a and b. The highly redundant nature of this transform makes it inefficient from a computational point of view. One way to eliminate the problem of redundancy is to sample the CWT on a 2-D dyadic grid. That is, use wavelets only of the form $\psi(2^k t + l)$ with k and l being whole numbers. The resulting WT is called DWT. DWT is still the transform of a continuous time signal, with discretization performed in the a and b variables only. Hence it is analogous to the Fourier series, and also referred to as a continuous time wavelet series [23], [24].

1.3.1.4 The Multiplexed Wavelet Transform (MWT)

MWT is an alternate method for the time-scale representation of pseudo periodic signals with constant period, first proposed by Evangelista [25]. This transform simplifies the analysis of a pseudo periodic signal by decomposing it into a regular asymptotically periodic signal and a number of fluctuations over this.

Images can be treated as oscillatory signals, although they are not periodic in a strict mathematical sense. Contrary to the case of one-Dimensional (1-D) signals, no period detection is required in the case of images. When treated as quasi-periodic signals, the periods along the horizontal and vertical directions respectively are the width and length of the image segment. Hence, the DWT of the rows of the image gives the MWT of the image taken as a 1-D signal along the vertical direction and that of the columns corresponds to the MWT of the image taken as a 1-D signal along the horizontal direction.

1.3.1.5 WT in Two Dimensions

When the input signal is 2-D, it is necessary to represent the signal components by 2-D wavelets and 2-D approximation function. Often this is done by using separable products of 1-D wavelets and scaling functions which make it possible to use the Fast Wavelet Transform (FWT) algorithms.

For any scaling function and its corresponding wavelet function, we can construct three different 2-D wavelets and one 2-D approximation function using the tensor product approach. Each new wavelet measures the variations along a different direction ; vertical, horizontal and diagonal. As a result the 2-D extension of the wavelet transform is achieved by applying the 1-D algorithm along the rows and columns of the image. That is, the image is decomposed row wise first, for every row and then this is repeated column wise for every column.

1.3.1.6 Computation of DWT

The DWT of a signal is determined by finding the *detail* in the signal at each *level of resolution*; that is, for each successive value of the dilation variable. In essence, this is done by convolving the input signal with the appropriately dilated wavelet function at each translation. As the dilation increases, the number of translation points for which values must be determined drops; at the highest resolution the wavelet is being used to measure the difference between successive samples while at the lowest resolution the wavelet is comparing the first half of the signal with the second half. When the wavelet family is orthogonal, adding the detail at all levels of resolution yields the original signal.

Stephane Mallat [21] has shown how a scaling function and a wavelet function can be used in a recursive algorithm to compute the orthogonal forward and inverse WT of a signal in O (n $\log$ n) time. This is considered as the standard algorithm for WT computation. The scaling and wavelet functions are in effect low and high pass filters; at each level of recursion wavelet function is used to extract the details at that level of resolution and scaling function is used to construct a coarser version of the signal for analysis at the next level. The process is repeated on successively coarser representations of the signal, until only the steady-state (average) value of the signal remains.

1.3.1.6.1 Sectioned computation

Generally, the sequences involved in real time implementations are quasi-infinite and processing of such data is done after segmenting it to smaller blocks or frames. The DWT and Inverse Discrete Wavelet Transform (IDWT) are recursive-filtering processes. Hence, WT is not a block transform and due to the lack of data beyond block boundaries, edge artifacts will be produced on block boundaries in the reconstructed signals. For correct computation near the data boundaries each processor would need to access data allocated to other processors. This demands frequent data exchange between processors or requires large buffer storage for intermediate transform coefficients.

1.3.1.7 WT in Biomedical Image Processing

In the past few years, researchers in applied mathematics and signal processing have developed powerful wavelet methods for the multiscale representation and analysis of signals [23], [26]. These new tools differ from the traditional Fourier techniques by the way in which they localize the information in the time-frequency plane. They are capable of trading one type of resolution for the other, which makes them suitable for non-stationary signal analysis. One important area where these properties are found relevant is biomedical engineering.

The main difficulty in dealing with biomedical signals is their extreme variability and the necessity to operate on a case-by-case basis. Often there is no *apriori* knowledge about the pertinent information and /or at which scale it is located. Frequently, the deviation of some signal feature from the normal is the most relevant information for diagnosis. Another important aspect of biomedical signals is that the information of interest is often a combination of features that are well localized spatially or temporally (e.g. microcalcifications in mammograms) and others that are more diffuse (e.g. texture). This requires the use of sufficiently versatile analysis methods, to handle events that can be at opposite extremes in terms of their time-frequency localization.

The applications of wavelets in biomedical field include performing image processing tasks like noise reduction, enhancement, detection and reconstruction, acquisition techniques for X-ray tomography and MRI and statistical methods for localizing patterns of activity in the brain using functional imaging.

1.3.1.7.1 Computer Assisted Mammography

Image enhancement is especially relevant in mammography where the contrast between the soft tissues of the breast is inherently small and a relatively small change in the mammary structure can signify the presence of a malignant breast tumor. Because of the current interest in mammographic screening, wavelet based enhancement methods have been recently designed with that application in mind [27], [28], [29]. All these approaches invariably use reversible redundant or non-redundant wavelet decomposition and perform the enhancement by selective modification of WT coefficients. These enhancement techniques are not fundamentally different from the noise reduction techniques, since in the former case certain features of interest are amplified while in the latter some unwanted features are suppressed.

One of the key issues in computer-assisted mammography is the detection of clusters of fine granular microcalcifications, which are one of the primary signs of breast cancer. Individual calcifications typically range from 0.05-1mm in diameter. The detection of microcalcifications is closely related to the enhancement task described earlier, except that detection is typically performed by thresholding in the wavelet domain. The detection results so far reported suggest that wavelet techniques perform better than the best available single scale methods [30], [31], [32], [33].

1.3.1.7.2 Computer Assisted Tomography (CAT)

In 2-D computerized X-ray tomography image of an object is reconstructed from the measured values of its angular projections. These measurements are described by the Radon transform. The primary motivation for using wavelets for tomography is that the wavelet reconstruction formulas tend to be localized spatially and can be applied to obtain partial reconstructions when only a portion of the Radon transform is available (limited angle tomography). The WT also appears to have some merits for noise reduction in tomography.

1.3.1.7.3 Magnetic Resonance Imaging (MRI)

One of the major applications of the WT in medical imaging is the noise reduction in MR images. One approach proposed is to compute an orthogonal wavelet decomposition of the image after applying a soft thresholding rule on the coefficients [34]. A more sophisticated approach is an over complete wavelet decomposition followed by a reconstruction from the retained significant WT maxima by exploiting the correlation between adjacent scales [35], [36]. When applied to MR images this method compared favorably with the optimal Wiener filter and produced images with much sharper edges and did not induce any ringing artifacts [36], [37].

1.3.1.7.4 Functional Image Analysis

Functional neuro-imaging is a fast developing area aimed at investigating the neuronal activity of the brain in vivo. Positron Emission Tomography (PET) and *f*MRI are the two modalities that are used to obtain functional images. PET measures the spatial distribution of certain function specific radiotracers injected into the blood stream prior to imaging. A typical example is the measurement of cerebral glucose utilization with the tracer 2-fluoro-2-dioxy-D- Glucose (FDG). *f*MRI allows for a visualization of local changes in blood oxygenation induced by neuronal activation. It is substantially faster than PET and also offers better spatial resolution.

The functional images obtained with these two modalities are extremely noisy and variable and their interpretation requires the use of statistical analysis methods. The first step in this analysis is the registration of various images, which compensates for intersubject anatomical variability or intrasubject movement in the scanner. Efficient multiresolution solutions to this problem have been proposed resulting in much faster and robust algorithms compared to single scale counterparts [38].

The second step is the computation of difference between the aligned group averages and performing the statistical analysis. Direct testing in the image domain is difficult because of the amount of residual noise and the necessity to use a very conservative significance level to compensate for multiple testing. Testing in the wavelet domain has the advantage that the discriminative information, which is smooth and well localized spatially, becomes concentrated into a relatively small number of coefficients while the noise remains evenly distributed among all coefficients.

1.3.2 Neural Network

Traditional DSP is based on algorithms, changing data from one form to another through step-by-step procedures. Most of these techniques also need parameters to operate. For example, recursive filters using recursion coefficients, feature detection implemented by correlation and thresholds, image display depending on the brightness and contrast settings, etc. Algorithms describe what is to be done while parameters provide a benchmark to judge the data. The proper selection of parameters is often more important than the algorithm itself. Neural networks take this idea to the extreme by using very simple algorithms, but many highly optimized parameters. They replace the traditional problem-solving strategies with trial and error pragmatic solutions, and a "this works better than that" methodology.

A neural network structure can be defined as a collection of parallel processors connected together in the form of a directed graph, organized such that the network structure tends itself to the problem being considered [39]. It is radically different from the notions of ordinary serial computing strategy and forms a powerful tool for applications where the processing is to be done in parallel. They offer the following advantages:

i) Adaptive learning: This is learning to perform specific tasks by undergoing training with illustrated examples. This feature eliminates the need of elaborating *apriori* models or specifying probability distribution functions.

ii) Self-organization: Neural networks use self-organizing capabilities to create representations of distinct features in the presented data, which leads to the generalization of features.

iii) Fault tolerance: Networks can learn to recognize noisy and incomplete data and also exhibit graceful degradation when part of the network itself is destroyed.

iv) Real-time operation: Due to its parallel distributive structure most networks operate in the real time environment and the only time consuming operation is training the network.

Neural networks have been applied in many fields, some of which are mentioned below. In Aerospace applications it is used for high performance aircraft autopilot, flight path simulation, aircraft control systems, autopilot enhancements, and aircraft component simulation and fault detection. In automotive industry it is used for automobile automatic guidance system and warranty activity analysis. It is used in banking sector for cheque and other document reading and credit application evaluation. In the field of communication, neural network finds extensive applications in image and data compression, automated information services, real-time translation of spoken language and customer payment processing systems. In medical field neural networks are employed for breast cancer cell analysis, EEG and ECG analysis, prosthesis design, optimization of transplant times, hospital expense reduction and hospital quality improvement. It is also used in the fields of defense, entertainment, finance, manufacturing, oil and gas exploration, robotics, transportation etc [40], [41].

1.3.2.1 Target detection

Scientists and engineers often need to know if a particular object or condition is present. For instance, geophysicists explore the earth for oil, physicians examine patients for disease, astronomers search the universe for extraterrestrial intelligence, etc. These problems usually involve the comparison of the acquired data against a threshold and if the threshold is exceeded, the target is deemed present. The conventional approach to target detection (sometimes called pattern recognition) is a two-step process. The first step is called feature extraction, which uses algorithms to reduce the raw data to a few parameters, such as diameter, brightness, edge sharpness, etc. These parameters are often called features or classifiers. Feature extraction is needed to reduce the amount of data and to distill the information into a more concentrated and manageable form.

In the second step, an evaluation is made of the classifiers to determine if the target is present or not. This is quite straightforward for one and two-parameter spaces; the known data points are plotted on a graph and the regions separated by eye. As the number of parameters increases this cannot be done by the human brain and dedicated networks are required to carry out this .The neural network is the best solution for this type of problems. Some of the important neural classifiers include Perceptrons, Backpropagation network, Self-organizing map, Competitive networks, Learning Vector Quantization (LVQ) and Probabilistic Neural Network (PNN).

1.4 Objective of the work

Cancer is not preventable, but early detection leads to a much higher chance of recovery and lowers the mortality rate. Considering the incidence of breast cancer and the favorable prognosis associated with early detection, it is surprising to note that only 15 to 30% of eligible women have ever had a mammogram [42] and even fewer are involved in a regular screening program. Reasons for this are high cost, skepticism about reliability and the physical discomfort of the process.

The high cost of a mammography-screening program can be partly attributed to the fact that the mammographic images are difficult to interpret even for skilled radiologists with years of experience. One reason for this is that a mammographic image is a highly textured 3-D structure, which has been projected onto a 2-D plane. Additionally the images are often of low contrast, in order to maintain low radiation dose to the patients. It can be assumed that less than 10 percent of the mammograms from a screening population contain some type of abnormality. The visual fatigue of reading numerous mammograms, most of which are negative, and the existence of a wide variation of breast tissue structures lead to inconsistent readings between radiologists, and even by a single radiologist at different times.

CAD and automated pre-screening by computer makes it easy to interpret the multitudes of mammographic readings. Even if there is no large screening program computerized mammogram image analysis could be used to improve the quality of conventional mammography. In a CAD scenario, computerized image analysis is used to suggest possible suspicious regions in the image so that a radiologist can then examine these regions more carefully. Evidence is mounting that prompting the radiologist with computer detection results of mammographic images leads to an increased sensitivity without affecting specificity [5], [43], [44].

Cancer treatment is most effective when it is detected early and the progress in treatment will be closely related to the ability to reduce the proportion of misses in the cancer detection task. The effectiveness of algorithms for detecting cancers can be greatly increased if these algorithms work synergistically with those for characterizing normal mammograms. This research work combines computerized image analysis techniques and neural networks to separate out some fraction of the normal mammograms with extremely high reliability, based on normal tissue identification and removal.

The presence of clustered microcalcifications is one of the most important and sometimes the only sign of cancer on a mammogram. 60% to 70% of non-palpable breast carcinoma demonstrates microcalcifications on mammograms [44], [45], [46].

WT based techniques are applied on the remaining mammograms, those are obviously abnormal, to detect possible microcalcifications. The goal of this work is to improve the detection performance and throughput of screening-mammography, thus providing a 'second opinion ' to the radiologists.

The state-of- the- art DWT computation algorithms are not suitable for practical applications with memory and delay constraints, as it is not a block transform. Hence in this work, the development of a Block DWT (BDWT) computational structure having low processing memory requirement has also been taken up.

1.5 Layout of the Thesis

The thesis is organized in the following way:

A brief review of the previous research works in the field of computer-aided breast cancer detection is presented in chapter 2. Special stress is given to microcalcification detection and neural network based classification of normal / abnormal tissue in mammograms. Different methods of computation of both 1-D and 2-D WT are also reviewed in this section.

Chapter 3 summarizes the features of different types of breast lesions in digital mammograms, namely, microcalcifications, circumscribed lesions, and spiculated lesions.

Chapter 4 describes the basic theory for classification using neural networks and detection of microcalcifications using WT. An overview of neural networks for classification purposes and multiresolution representations of signals using wavelets are provided. One- dimensional wavelet analysis is discussed; including the orthogonal and biorthogonal wavelet representations and is extended to 2-D. Different types of WTs are also considered in this chapter.

Chapter 5 describes the algorithms developed for block-wise computation of both 1-D and 2-D DWT. The conventional method and its computational complexity are described in detail. The computational complexity of the BDWT algorithm is evaluated and compared against the standard methods.

Chapter 6 presents the classification of mammograms into normal and abnormal classes using neural networks. First the features of normal mammograms are explained followed by the derivation of different features for classification purpose. Finally results and conclusion are presented.

The new MWT based algorithms for automatic detection of microcalcifications is presented in chapter 7. The microcalcification detection problem is represented as an edge detection operation and different WT based edge detection methods are discussed in detail. Experimental results on mammographic data and discussions are also provided.

Chapter 8 is the concluding chapter, wherein the observations and inferences already brought out in the previous chapters are summarized. The suggestions for further work are also given.

This thesis includes one appendix, which describes a line detector that is capable of extracting linear mammographic features. This line detector is used to find out and remove normal linear markings from mammograms.

Chapter 2

Literature Review

Studies have shown that mammography can be used to detect breast cancer two years before it is palpable and can reduce the overall mortality due to this by up to 30% [47]. When detected early, localized cancers can be removed without resorting to breast removal (*masectomy*). However, radiologists' interpretation of the same mammogram may differ substantially [48] since mammograms are generally low in contrast and high in noise, while breast structures are small and complex. The false negative rate in current clinical mammography is reported to vary from 4% to 20% [2], [49], [50], [51], [52], [53]. Also, in the cases where positive mammograms have been reported and sent for biopsy, only 15 to 34% actually have been found to have cancer [54], [55]. Therefore, in the past decade tremendous research has been done on CAD techniques in mammography, so as to increase diagnostic accuracy of mammographic interpretation. However, the current state in computerized mammography techniques is not sufficient for large scale screening programs.

2.1 CAD in Mammography

Many researchers have attempted automated breast cancer detection by employing image processing techniques for detection of masses, lesions and microcalcifications. Other work in the field of digital mammography has been directed towards the enhancement of digital mammograms either to improve radiologist's reading or as a preprocessing step for some computerized process [56].

For circumscribed mass detection, a combination of criteria including shape, brightness, contrast, and uniform density of tumor areas was employed by Lai *et al* [57] and thresholding and fuzzy pyramid linking was used by Brzakovic *et al* [58]. Bilateral subtraction technique based on the alignment of corresponding right and left mammograms was tried by Yin *et al* [59] and Mendez *et al* [60]. Li *et al* [61], Comer *et al* [10] and Zheng *et al* [62] used Markov random fields to classify a mammogram into different texture regions, thereby singling out cancerous masses. A statistical method based on fitting broken regression lines to local intensity plots is proposed by Hastie *et al* [63]. Petrick *et al* used an adaptive density-weighted contrast enhancement filter in conjunction with Laplacian-Gaussian edge detection to detect suspicious mass regions in mammograms [64]. Wei *et al* proposed the use of local texture features in combination with global multiresolution texture features for the detection of masses from normal breast tissue [65]. Kupinski and Giger [66] developed two lesion segmentation techniques: one based on a single feature called the radial gradient index (RGI) and the other based on simple probabilistic models. Bovis and Singh employed a texture feature based mass detection technique [67]. A WT technique in conjunction with a novel Kalman-filtering neural network is proposed by Qian *et al* [68]. A multiple circular path convolution neural network (MCPCNN) architecture specifically designed for the analysis of tumor and tumor-like structures has been constructed by Lo *et al* [69].

For spiculated lesions, Kegelmeyer *et al* [5], [70] extracted a five-dimensional feature vector for each pixel, which included the standard deviation of the edge orientation histogram and the output of four spatial filters. Each feature vector was then classified using a binary decision tree. Huo *et al* [71] developed a technique that involves lesion extraction using region growing and feature extraction using radial edge-gradient analysis. Karssemeijer and Brake [72] investigated a method based on statistical analysis of a map of edge orientations. Kobatake and Yoshinaga [73] proposed the use of line skeletons and a modified Hough transform to characterize spiculated patterns. Liu *et al* designed a multiresolution scheme using a binary tree classifier for the detection of spiculated and stellate lesions [74], [75]. Qi & Snyder proposed a lesion-detection-and-characterization technique using Bezier histograms [76].

H.P.Chan *et al* [77], [78] investigated the application of computer-based methods for the detection of microcalcifications. Their system was based on an image subtracting technique in which a signal-suppressed image was subtracted from a signal-enhanced image to remove the background. Signal extraction techniques adapted to the known physical properties of the microcalcifications were used to isolate them from the remaining noise background. They have obtained a true positive cluster detection rate of approximately 80% at a false positive detection rate of 1 cluster per image on 20 mammograms, all of which containing clustered microcalcifications.

Davies and Dance [79] report a 96% true positive rate for clusters with an average of 0.18 false clusters per image for 50 mammograms, half of which were normal, using segmentation and local area thresholding. Karssemeijer [8] reports an algorithm using which he had obtained zero false negatives with about 2 false positive clusters per image on 40 mammograms containing microcalcifications.

Nishikawa *et al* [80] used a difference image technique to enhance microcalcifications first and then extracted potential microcalcifications with a series of three techniques: a global thresholding, an erosion operator, and a local adaptive thresholding. Finally, some false positives are eliminated by a texture analysis technique and remaining detections are grouped by a non-linear clustering algorithm. Employing a WT technique for enhancing the microcalcifications and combining this with the difference image technique mentioned above, Yoshida *et al.* [81] obtained an overall detection sensitivity of approximately 95%, with a false positive rate of 1.5 clusters per image on a database consisting of 39 mammograms with 41 clusters.

Chan *et al* [82] investigated a Convolution Neural Network (CNN) based approach and showed its effectiveness in reducing false positive detections. Strickland and Hahn [32] designed multiscale matched filters using WT for enhancing and detecting calcifications. On the Nijmegen database containing 40 mammograms, they had obtained a detection rate of 55% true positives at the cost of 0.7 false positives per image. Based on matching pursuit with optimally weighted wavelet packets, Yoshida [83], [84] achieved a sensitivity of 93% with a specificity of 80% in classifying 297 ROIs as containing microcalcifications or belonging to the background.

Gurcan *et al* [85] described a statistical method using skewness and kurtosis to detect microcalcifications. Ibrahim *et al* [86] employed a triple ring filter to extract the specific features of the pattern of the microcalcifications from contrast corrected mammograms. They have obtained a sensitivity of 95.8% with a false positive rate of 1.8 clusters per image on 43 mammograms from the Mammographic Image Analysis Society (MIAS) database.

Cheng *et al* [87] proposed a five-step approach based on fuzzy logic technique, which includes image fuzzification, enhancement, irrelevant structure removal, segmentation, and reconstruction. Nagel *et al* [88] examined three feature analysis methods, namely, rule based, Artificial Neural Network (ANN) and a combined method and concluded that the combined method performs best because each of the methods eliminates different types of false positives. A WT based technique where the detection is directly accomplished into the wavelet domain is presented in [89]. Texture-analysis methods can be applied to detect clustered microcalcifications in digitized mammograms. The surrounding region-dependence method of texture analysis is shown to be superior to the conventional texture-analysis methods with respect to classification accuracy and computational complexity [90].

Schmidt *et al* [91], [92] developed a fully automatic computer system for the identification and interpretation of clustered microcalcifications in mammograms with the ability to differentiate most benign lesions from malignant ones in an automatically selected subset of cases. From a total of 272 films of 100 patients, they have found 247 clusters of microcalcifications containing 5349 single microcalcifications with sensitivities of 0.90, 0.98 and 1.0 at the respective false positive alarm rates of 1.3, 5.3 and 7.4 groups per image.

Combining difference-image technique, gaussianity, statistical properties and multiresolution properties of WT, Bazzani *et al* [93] yielded a sensitivity of 91.4% with 0.4 false positive clusters per image on the 40 images of the Nijmegen database. Yu and Guan developed a method that segments potential microcalcification pixels in the mammograms by using wavelet features and gray level statistical features [94]. 90% mean true positive detection rate is achieved at the cost of 0.5 false positive per image by applying this to the 40 mammograms of Nijmegen database containing 105 clusters of microcalcifications. By exploiting information gained through evaluation of Renyi's entropy at the different decomposition levels of the wavelet space, microcalcifications are separated from background tissue. Gulsrud and Husoy [95] proposed a scheme for texture feature extraction based on the use of a single optimal filter for microcalcification detection achieving approximately 89% true positive detection rate with only one false positive cluster per image on the MIAS database.

A method is presented by Diekmanna *et al* [96] for visualizing microcalcifications by full-field mammography using wavelet frames, an enhancement operator, and a suitable reconstruction technique. In all cases, microcalcifications were depicted with a markedly higher contrast for 24 digital mammographies (Senographe 2000D,GE Medical Systems) containing microcalcifications. Serrano *et al* [97] detected microcalcifications in mammograms based on region growing with pre-filtering and a seed selection procedure based on 2-D linear prediction error. They have achieved a detection capability of 86% over all of the existing microcalcifications in three test mammograms.

Melloul and Joskowicz [98] described a fully automatic, parameter-free algorithm for microcalcification segmentation in mammograms. Removing the background tissue with a multiscale morphological operation and then applying thresholding based on a 3-dimensional co-occurrence matrix, they have obtained detection rates of 93.75% of true positives, 6.25% of false positives, and 2% of false negatives on the MIAS database. A hybrid intelligent system is presented by Papadopoulosa *et al* [99] for the identification of microcalcification clusters in digital mammograms based on a three-step procedure of preprocessing and segmentation, Regions Of Interest (ROI) specification, and feature extraction and classification. The proposed methodology produced a detection accuracy of 0.91 and 0.92 for the Nijmegen and MIAS mammographic databases respectively with 1.80 and 1.15 false positive clusters per image, at a sensitivity level higher than 0.90.

Mammographic image enhancement methods are typically aimed at either the improvement of the overall visibility of features or the enhancement of a specific sign of malignancy. Laine and Schuler [27], [100] introduced a method of contrast enhancement on digital mammograms based on the hexagonal wavelet transform that improved the visualization of breast pathology without excessive noise amplification. Gagnon *et al* [101] proposed a multiscale sharpening enhancement algorithm using complex symmetric Daubechies wavelets, useful for low contrast digitized mammograms. Li *et al* [102] showed that general mammographic parenchymal and ductal patterns could be well modeled by a set of parameters of *affine* transformations and hence, microcalcifications can be enhanced by taking the difference between the original image and the modeled image. They also demonstrated that the fractal modeling method is an effective way to enhance microcalcifications so that the detection and classification of them in a computer-aided diagnosis system can be improved.

Kim *et al* [103] proposed an adaptive image enhancement method for enhancement of microcalcifications in mammographic images, which is based on the first derivative and the local statistics. Local statistics of the image is utilized for adaptive realization of the enhancement whereby image details can be enhanced and image noises can be suppressed. Koren *et al* [104] used a redundant B-spline WT decomposition followed by image fusion for enhancement of microcalcifications, circumscribed masses and stellate lesions. Sersic and Loncaric [105] introduced an approach based on redundant DWT to enhance digital mammography images for more accurate detection of microcalcification clusters. In [106] filter banks derived from the CWT, called integrated wavelets, were employed for the enhancement of microcalcifications in mammograms.

2.1.1 Classification of microcalcifications into benign and malignant

In the United States, the positive predictive value, i.e., the ratio of the number of breast cancers found to the total number of biopsies, of mammography is typically between 15 and 34% [54], [107]. An improvement in the positive predictive value would reduce health care costs and eliminate the anxiety and morbidity of patients who would have to undergo unnecessary biopsy otherwise. One of the potential approaches to improving the specificity of mammography is the use of computerized feature extraction techniques to extract information that may not be readily perceived by human readers. These features may complement the visual characteristics of the mammographic abnormalities and provide additional information to the radiologists in distinguishing malignant and benign lesions. The computer-extracted features, alone or in combination with human-perceived features, may also be input to a trained classifier to estimate the likelihood of malignancy of a mammographic lesion, thereby assisting radiologists in making diagnostic decisions.

Shen *et al* [108] developed a set of shape factors, including measures of compactness, moments and Fourier descriptors, to measure the roughness of contours of calcifications for using in their classification as malignant or benign. Dhawan *et al* [109] presented a feature analysis approach using features from first and second order gray-level histograms.

Using a back propagation ANN and texture features, Chan *et al* [110] have got 39% specificity without missing any malignant cases (100% sensitivity) for patients who had undergone biopsy. The test was conducted on a data set of 54 cases (26 benign and 28 malignant). The ANN was trained and tested with a leave-one-case-out method to recognize the malignant or benign microcalcification clusters.

Salfity *et al* [111] identified clusters by a detection scheme based on morphological filtering and classified the data in benign and malignant cases using K-nearest neighbor method and a naive Bayesian classifier. On a database containing 23 malignant and 59 benign microcalcification clusters, they have attained true positive fractions of 87% and 78% respectively at a false positive fraction of 10%.

Christoyianni *et al* [112] investigated the efficiency of Radial-Basis-Function (RBF) and Multilayer perceptron (MLP) classifiers in recognizing cancer regions of suspicion on mammograms. Similar recognition scores were obtained on the MIAS database, for two types of texture features: statistical descriptors based on high-order statistics and the Spatial Gray-Level Dependence (SGLD) matrix.

Varma and Zakos [113] presented an intelligent computer-aided diagnosis system which distinguishes a benign microcalcification pattern from a malignant one using a fuzzy technique in conjunction with combination of 3 features; entropy, standard deviation and number of pixels. A total of 40 microcalcification areas (20 benign, 20 malignant) from Nijmegen database were used as training samples and 10 images (5 benign and 5 malignant) were used for testing. The algorithm detected most microcalcification areas with detection rates of 83.3% for benign case and 77.8% for malignant case.

Buchbinder *et al* [114] investigated whether the size of mammographically detected microcalcifications is predictive of malignancy and came to the conclusion that the average length and area of the calcifications in benign clusters were significantly smaller than those in malignant clusters.

Others have explored methods for classifying breast lesions as benign or malignant. Kilday *et al* [115] studied the use of tumor boundary roughness, circularity and other shape factors for classification of mammographic lesions as fibroadenomas, cysts and cancers. Pohlman *et al* [116] developed a technique based on shape morphology for classifying breast lesions as benign or malignant. Rangayyan *et al* [117] investigated the potential of acutance in quantifying the sharpness of the boundaries of tumors and distinguishing between benign and malignant mammographic tumors. Sahiner *et al* [118] introduced a Rubber Band Straightening Transform (RBST) for characterization of mammographic masses. Features extracted from the RBST images were found to be significantly more effective for tumor classification than those extracted from the original images.

Guliato *et al* [119] proposed two segmentation methods, one based on region growing and enhancing the ROI using fuzzy sets and the other based on fuzzy region growing method that takes into account the uncertainty around the boundary of a mass. The methods are successful not only in identifying a mass, but also classifying it into benign or malignant.

2.1.2 Normal mammogram characterization

The effectiveness of algorithms for detecting cancers can be greatly increased if these algorithms work synergistically with algorithms for characterizing normal mammograms. However, little work has been done on understanding normal mammograms [120]. Sahiner *et al* [121] developed a fast and stable implementation of the CNN to classify ROIs in mammograms as either mass or normal tissue. On a data set consisting of 168 biopsy proven ROIs with masses and 504 ROIs having normal breast tissue, they have got a true positive detection rate of 90% at a false positive fraction of 31%.

Heine, *et al* [122] used a statistical method based on wavelet expansion to separate normal regions from potentially abnormal regions containing isolated calcifications. This is the initial phase of the development of a general method for the automatic recognition of normal mammograms. On a dataset of 30 mammograms containing 17 clinically abnormal ones and the rest normal, they were able to achieve 92% specificity and 89% sensitivity at the cost of 0.12 false positives per image. Kalman, *et al* [123] studied the feasibility of combining WT and ANN to screen normal mammograms from those containing masses. They have obtained a sensitivity of 75% and specificity of 56% using a dataset of 350 mammograms with 221 cases containing masses.

Liu *et al* [120], [124] presented the development of algorithms for recognizing normal mammograms by eliminating the effects produced by normal glandular tissues and connective tissues in mammograms. Bovis and Singh [125] investigated a new approach to the classification of mammographic images based on the underlying texture contained within the breast tissue. This study demonstrated a high sensitivity in the classification of breast types justifying the use of this prior knowledge for the detection of lesions in a proposed CAD system.

Zaýane *et al* [126] proposed a classification method based on association rule mining to classify digital mammograms into three categories: normal, benign and malignant. The experimental results show that the method performs well reaching over 80% in accuracy for the MIAS database. Lo *et al* [127] developed an ensemble classifier that identifies whether suspicious calcification clusters are benign or malignant based on features extracted by automated image processing algorithms as well as features manually interpreted by radiologists.

With the advent of digital mammography systems, the implementation of CAD systems in everyday clinical applications looks highly promising [128] and is obviously revolutionizing the practice of radiology.

2.2 DWT Computation

In recent years, there have been considerable research activities centered on building efficient systems for computing the DWT. This is certainly because the WT is a powerful tool for multiscale time-frequency signal decomposition and analysis which has found applications in many areas such as signal processing, digital communications, numerical analysis, and computer graphics [129]. Moreover, practical system design is itself a very challenging problem because of the stringent constraints, such as buffer size, delay, power, chip area and control complexity, imposed by specific DWT applications [130], [131], [132].

The most popular DWT algorithm is the recursive filtering approach using the corresponding wavelet filter bank, the so-called standard algorithm [21], whose computational complexity is $O(L)$ per output coefficient (L is the filter length). The FFT-based DWT algorithm proposed by Rioul *et al* [24] reduced the complexity from $O(L)$ to $O(log\ L)$ for large filter lengths. For short filters, a "fast running FIR filtering" technique was presented [133], which has achieved 30% saving in computations. Using a lattice structure, Vaidyanathan *et al* [134], [135] have shown that the complexity can be reduced by a factor of 50% for orthogonal wavelet filter banks. The ladder structure by Marshall [136] and the lifting algorithm by Daubechies and Sweldens [137] further show that, for large filter lengths, 50% savings in computations can be achieved for any FIR wavelet filter bank.

In many cases, a sequential architecture is used where the DWT is computed by splitting the input into blocks, with the processor operating on one block at a time [138], [139], [140]. One reason for such a choice is that only a limited amount of memory is available for the transform computation. Example scenarios include image compression/decompression systems using a DSP/ASIC chip in consumer electronics products (e.g., digital cameras) or space-borne instruments [21], [133]. In these applications, reducing the memory buffer size helps not only to maintain low costs but also to reduce the chip design area and thus the volume of the final product.

As an alternative a parallel architecture would split the input among several processors to speed up the transform computation [131], [141], [142], [143]. This is typical for applications such as the seismic data processing [143] or illumination computations in computer graphics [129] where a large volume of data has to be processed in a reasonably short time. Obviously, fast DWT computation to meet stringent delay constraints is critical to the success of any wavelet-based techniques.

In sequential architecture designs, most approaches adopt the standard FFT-based filtering techniques [144], overlap-add or overlap-save. These include the recursive pyramid algorithm (RPA) by Vishwanath [138], the Spatially Segmented Wavelet Transform (SSWT) by Kossentini [145], and the reduced line-based compression system by Chrysafis *et al* [146]. Since the SSWT overlaps data only once before the start of the transform, the overlap buffer size increases exponentially with the increase of decomposition levels. Alternatives are implemented in [138] and [146] where data is overlapped at each level of decomposition and the buffer size is reduced.

In parallel architecture designs, most approaches proposed require communication of the boundary data at each level of decomposition [131]. To reduce the overhead caused by frequent inter-processor communication, Yang at el. [147] proposed to use boundary extensions in their DWT system configured from a cluster of SGI workstations. However, this computes incorrect wavelet coefficients near data boundaries, which causes performance degradation in some applications, for example low-bit rate image coding.

Various image codecs using hierarchical subband decomposition emerged during the last decade. One of the first algorithms of this family is due to Shapiro [148] and is known as the Embedded Zero-tree Wavelet algorithm (EZW). Said and Pearlman [149] successfully improved the EZW algorithm by providing symbols for combination of parallel zero-trees. The implementation is based on a set-partitioning sorting algorithm called Set-Partitioning in Hierarchical Trees (SPIHT).

These wavelet based image codecs can produce smoother and more perceptually pleasant reconstructed images than conventional Discrete Cosine Transform (DCT) based image codecs. However, implementing these codecs for the entire image is cumbersome for large images, since it requires a large amount of memory. To reduce the memory requirement, the most direct approach is to divide the image into non-overlapping blocks and process each block independently. However, due to the lack of pixels beyond block boundaries, edge artifacts will be produced on block boundaries in the reconstructed image. Williams and Amaratunga [150] removed edge artifacts by extrapolating the missing pixels using polynomials, but their method introduces a large computational overhead. Eom and Kim [151] proposed a different method of removing the edge artifacts by using samples of adjacent blocks. Their scheme can perform the forward WT independently for each block, but independence is not achieved for the inverse transform.

The edge artifacts are removed in SSWT algorithm by making use of the overlapping between blocks. This algorithm was evaluated by M. Adams [145], [152]. Within the JPEG2000 baseline system and found suitable. Chrysafis and Ortega [146] proposed a complete low-memory image compression system using a line-based approach. The memory requirement of this method was reduced to be proportional to the width of image, which when very large, still require large amount of memory for both transform and coding. Also, the line-based codec is not easily parallelized.

Recently Nealand *et al* [153] proposed an overlap save convolution method for wavelet analysis, which eliminated boundary distortion with the exception of first and last frames. But they have not considered reconstruction. Devassia *et al* [154], [155] introduced a Parallel Multiple Subsequence (PMS) structure which involves much less computation than state-of-the-art algorithms up to 6 levels of processing for Haar wavelet and 3 levels for others, for any data length. For higher decomposition levels and for real-time applications, this algorithm can be used effectively by optimal selection of processing frame size.

Chapter 3

Breast Cancer – A Medical Perspective

Breast cancer is one among the most common and deadly cancers, affecting nearly one in ten women. Standard Film Mammography is an important type of medical imaging used to uniquely screen healthy women for small curable breast cancers. Currently, it is the "gold standard " in breast cancer detection. Controlled medical studies have shown that mammography can lead to decreases in death due to breast cancer sufficient to measurably lengthen life. The reported cancer free 5 year survival for cancer detected by mammography is 92%, with 96% overall 5 year survival [1], [2].

To construct a system for automatic detection of abnormalities in mammographic images, it is important to have some basic medical knowledge in the area. It is also important to investigate how an expert radiologist works. Hence a medical point of view of mammography is presented in this section.

3.1 Anatomy of the female breast

The breast is a complex organ consisting of different types of tissue (see figure 3.1). The female breast contains milk-producing glands called lobes or lobules. These glands vary in size during the menstrual cycle due to the difference in hormone levels, and regress after menopause [156]. The lobules are linked together by tiny tubes called ducts through which the milk is transported from the lobules to the nipple. The surrounding fatty tissue and ligaments, called parenchyma, supports the lobules and ducts. The structure of the breasts of different women varies much, but usually the two breasts of the same woman are much alike.

There are also blood vessels and lymphatics, small thin channels that collect and carry tissue fluids, present in the breast. Breast tissue fluid drains through the lymphatics into axillary lymph nodes, located in the underarm. They filter the lymph fluid and serve as a barrier to the further spread of cancerous cells or bacteria that may

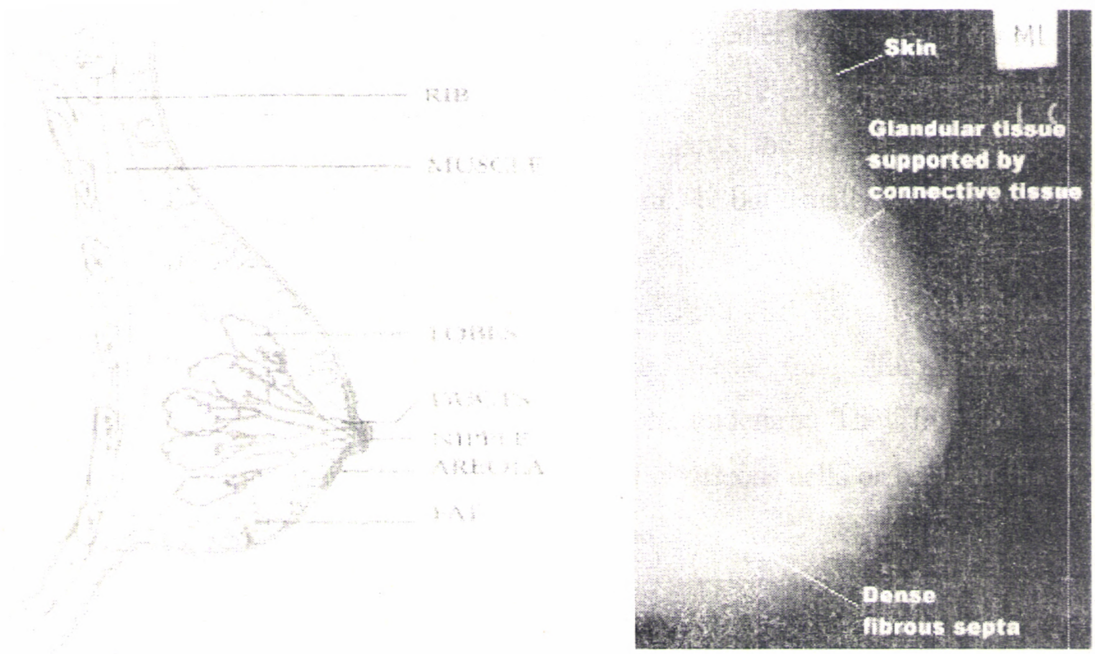


Figure 3.1: Schematic Diagram of the Female Breast.

have entered the lymph fluid. Lymph nodes are not completely effective in filtering out cancerous cells and despite their presence, once cancerous cells have gained access either to lymph channels or blood stream they have the potential to spread to any part of the body, particularly bone, lungs, liver and brain [157].

3.2 Malignancy in the breast

Cancer is a disease in which abnormal, mutated cells in some organs grow out of control [158]. Healthy cells reproduce themselves continuously throughout life, growing new tissue and replacing old or damaged ones, which is a normal, controlled and orderly process. However, sometimes this orderly process is disturbed and cells begin to reproduce in an abnormal way building a tumor.

A tumor may refer to both benign and malignant growth. Benign tumors remain similar to the tissue of their origin and generally do not invade surrounding tissues or produce *metastasis* [156] and their growth is usually slow. Malignant cells appear in many different forms. Some remain similar to the surrounding tissue and are referred to as *well-differentiated.* Cells bearing very little similarity to surrounding tissue are referred to as *undifferentiated* or *anaplastic.* These are usually more aggressive in their growth and behaviour than well-differentiated malignancies [156]. Metastasis occur when cancerous cells break away from the primary tumor, and travel through the body via blood or lymphatic channels to other organs where they grow and form new tumors.

A cancerous cell has characteristics that differentiate it from normal tissue cells with respect to the cell outline, shape, structure of nucleus and most importantly, its ability to metastasize and infiltrate. When this happens in the breast, it is commonly termed as ‘Breast Cancer’. Cancer is confirmed after a biopsy (surgically extracting tissue samples) and pathological evaluation.

The majority of breast cancers begin in either lobules or ducts [158]. Breast cancer is classified as either *invasive* (infiltrating) or *non-invasive* (*in-situ*). Invasive cancer has the ability to spread to other parts of the body, whereas in-situ cancers does not spread to other parts, but may develop and become invasive and should therefore be removed.

Since breast cancer grows close to lymphatic channels, detection, diagnosis and treatment of the cancer in an early stage is important. According to Greshon [157] a tumor may be considered early, not because it is small or because it is believed to have existed for a short time, but because it has not metastasized. Few palpable tumors can be thought of as early [157]. The most important tool for early detection is the use of mammography in mass screening programs [156], [157], [159]. Clinical examination and self-examination are the other methods of finding breast cancer.

The exact cause of breast cancer is not known. However, studies show that a woman's chance of getting breast cancer increases with age, personal and family history and certain genetic alterations. Also, breast cancer occurs more often in white women than in African or Asian women.

3.2.1 Symptoms & Diagnosis

Early breast cancer usually does not cause pain. In fact when breast cancer first develops, there may be no symptom at all. As the cancer grows it can cause changes that women should watch for. Some of them are:

1) A lump or thickening in or near the breast or in the underarm area.
2) A change in the size or shape of the breast.
3) Nipple discharge or tenderness or the nipple pulled back into the breast
4) Ridges or pitting of the breast
5) A change in the way the skin of the breast, aerola or nipple looks or feels like warm, swollen, red or scaly.

To find the cause of any sign or symptom, a doctor does a careful physical examination and analyses family and personal medical history. In addition the doctor may do one or more of the following examinations:

Clinical Breast Examination: The doctor can tell a lot about a lump by carefully feeling it and the tissue around it. Benign lumps often feel different from cancerous ones. The doctor can examine the size and texture of the lump and determine whether the lump moves easily.

Mammography: X-rays of the breast can give the doctor important information about a breast lump.

Ultrasonography: Using high frequency sound waves, ultrasonography can often show whether a lump is a fluid-filled cyst (not cancerous) or a solid mass (which may or may not be cancerous). This examination may be used along with mammography.

Based on these tests the doctor may decide whether further tests are needed or treatment is necessary. In such cases, the doctor may need to check the woman regularly to watch for any changes.

Biopsy

Often fluid or tissue must be removed from the breast so that the doctor can make a diagnosis. This is done by:

Fine needle aspiration: A thin needle is used to remove fluid and /or cells from a breast lump. If the fluid is clear, it need not be checked in a lab.

Needle biopsy: Using special techniques, tissue can be removed with a needle from an area that looks suspicious in a mammogram but cannot be felt. Tissue removed in a needle biopsy goes to a lab to be checked by pathologists for cancerous cells.

Surgical biopsy: In an incisional biopsy, the surgeon cuts out a sample of a lump or suspicious area. In an excisional biopsy, the surgeon removes all of a lump or suspicious area and an area of healthy tissue around the edges. A pathologist then examines the tissue under a microscope for cancerous cells.

3.3 Mammography

A mammogram is a specialized X-ray examination of the breast. It is an effective non-invasive means of examining the breast, searching for breast cancer. Two types of mammogram studies are commonly performed: screening mammography and diagnostic mammography.

A Screening mammogram is performed on women who have no current symptoms or breast problems while a diagnostic mammogram is performed specifically to evaluate a breast problem or revisit a previous abnormal finding. Mammograms are done using two different positions for each breast, the details of which are given in table 3.1. This allows more thorough evaluation of breast tissue. A compression paddle is used to spread out the breast tissue and obtain more uniform thickness. This greatly improves detail and image quality, making it possible to see very small abnormalities. At the same time the amount of X-rays needed for the examination are significantly reduced.

Abbreviation	**Projection/Position**	**Direction of the X-Ray**
CC	Cranio-Caudal	Direction from head (cranium) to the feet (caudal)
MLO	Medio-Lateral Oblique	X-ray direction is from medial (inner) to lateral (outer) aspect and the orientation of the breast is at an angle (Oblique)

Table 3.1: Positioning on performing mammograms

When a mammogram of the breast is taken, the different forms of tissue appear as different shades of gray depending on the level of absorbed radiation. Skin and fat tissue absorb very little radiation and does not usually show. Glandular tissues normally appear in medium or light gray shades. The pectoralis muscle located behind the glandular tissue covering the ribs is visible as a white area in a normal mammogram.

X-ray images depend on differences in x-ray stopping power (attenuation) of separate tissues. In general, a clear separation between normal functioning tissue and abnormal cancerous tissue is not possible since their attenuations are very similar. However, both functional tissue and cancerous ones can be separated from fatty storage tissue, which normally surround active breast tissue, even in lean persons. This is due to a substantially lower attenuation caused by fat.

In older women, the functional glandular tissue diminishes leaving only thin supporting tissues clearly outlined by fatty tissue. Mammography in these "mature" breasts is very effective since even small cancers are well outlined by fat. In addition, many cancers develop calcium deposits that strongly stop X-rays and are easily seen on mammograms.

Since mammography cannot separate normal gland tissue from tumors, it is much more effective when gland tissue diminishes with age. Many women retain glandular tissue as they "mature" and it camouflages tumors until they are large. The young women's breast normally contains more active tissue that again interferes with detection of small cancers.

3.3.1 The Mammography Machine

The first dedicated mammography machine was developed in 1966 [160]. Until then, mammographic images had been produced using standard X-ray machines. They have a reciprocating grid to reduce scatter radiation, thus avoiding fog and blurry images. A 0.03 mm molybdenum filter is generally used to make the beam hard and more penetrable.

The second-generation design introduced in 1980's reduced the exposure time significantly. The machine also provided increased resolution and accuracy and used more advanced type of film to provide better detail. Film processing is done under specific conditions. Depending on the type of film used, standard processing or extended processing techniques are employed to develop an exposed film [161]. The Films used for mammography are single emulsion fast films, which enhance image sharpness by eliminating geometric distortion. Films commonly used are Kodak Min-

RE, Agfa, Fuji, Dupont, and Konica. The screens consist of a rare earth phosphor called terbium activated gadolinium oxysulfide. Screens have to be compatible with the film. The newest film-screen combination is responsible for dose reduction by 30 - 50 %.

The technique used for mammography is low Kilo-voltage Peak (KvP) of about 24 to 30. The milli-Ampere-seconds (mAs) vary depending on breast tissue density. When the photo timer cells are used, it provides the optimum mAs for the tissue to be imaged. This technique results in mammograms with a high film contrast, making it easier for the radiologist to read.

The design of mammography equipment has progressed rapidly over the last four decades. In developed countries, dedicated mammography units are used. A whole range of manufacturers, GE, Bennett, Lorad, Siemens, Fischer, Phillips, etc. make these machines. In 2000 GE introduced the first full field digital mammography system the Senographe 2000D [160]. A digital mammography system uses essentially the same system as conventional mammography, but it is equipped with a digital receptor and a computer instead of a film cassette.

3.3.2 Breast Composition Determination

Because mammographic screening procedures are applied to all persons at risk, millions of mammograms and hundreds of thousands of biopsies must be preformed. To minimize the natural anxiety and inconvenience, mammographic procedures must be quick and accessible and every effort must be made to inform, counsel and support women undergoing the procedure.

Breast care involves many people, *viz*, personal physicians who order mammograms, radiologists who interpret mammograms and do needle biopsies, surgeons who perform incisional biopsies and curative cancer operations, pathologists who interpret biopsies, and radiotherapists who deliver radiation treatment. Patients often must function under great stress in evaluating several physician recommendations to make important treatment decisions. Hence, clear and accurate communication is almost as important as technical competence. To facilitate communication, the American College of Radiology (ACR) has created a standardized system for reporting the results of mammography, called ACR - Breast Imaging Reporting and Data System (BIRADS). It includes standard interpretation and reporting formats, a standard dictionary of terms, and standard disease classifications used in checking program effectiveness.

The ACR-BIRADS recognizes the background composition of the breast in the following categories:

1. Almost Entirely Fatty: Mammography very effective, sensitive to even small tumors.
2. Scattered Fibroglandular tissue: Minor decrease in sensitivity.
3. Heterogeneously Dense tissue present: moderate decrease in sensitivity.
4. Extremely dense tissue present: marked decrease in sensitivity.

Mammography does retain some value even in dense breasts by detecting calcium deposits (which are so dense, surrounding tissue does not interfere), but is not reliable in detecting small non-calcified cancers. In general, women with "dense" breasts remain so from year to year and it is possible to let a women know when she cannot depend on mammography. In dense breasts, more emphasis on self-examination may be appropriate, particularly if there is a family history of breast cancer.

3.4 Normal Mammograms

Unfortunately, there is no "normal" appearance on a mammogram that can be memorized. What constitutes "normal" varies within a wide spectrum. In addition, the appearance of the breast differs during pregnancy and in the postpartum period. This spectrum is due to the differences in breast composition. A breast with a high composition of adipose tissue will appear darker on a mammogram than a breast with a high composition of connective tissue stroma (lighter). From a population perspective, the mammogram will appear radiographically denser in a higher percentage of younger women than in older women. Even for an individual female, the image of breast density may vary over the years. Such changes are often gradual and the trend is generally towards a less dense (higher percentage of fat) breast tissue, but the reverse trend may happen as in the case of weight loss or hormone replacement therapy. Examples of different, entirely normal mammograms are shown in Figures 3.2 to 3.4.

Other normal variations of breast tissue include asymmetric patterns and asymmetric size. Although the breasts usually develop symmetrically, differences in the symmetry of breast tissue patterns or breast size are not necessarily abnormal. Without other indices of an abnormal process, such asymmetry may simply be a developmental phenomenon.

Figure 3.2: Normal Mammogram
Dense-glandular type

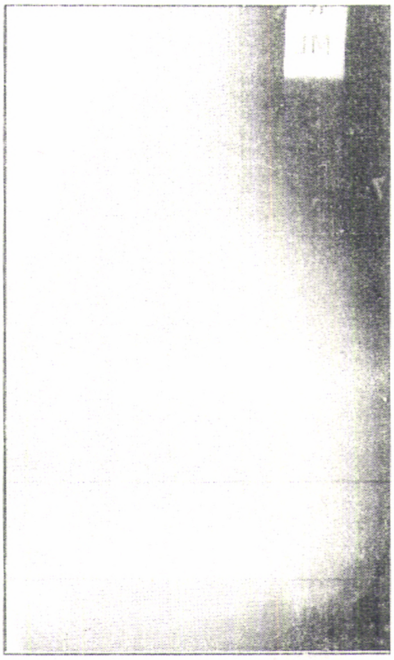


Figure 3.3: Normal Mammogram
Fatty type

3.5 Mammographic Abnormalities

3.5.1 Microcalcifications

A microcalcification is a tiny calcium deposit that has accumulated in the tissue in the breast and it appears as a small bright spot in the mammogram. A cluster is typically defined to be at least 3 to 5 microcalcifications within a 1 square centimeter region [1], [31]. Up to 50% of malignant masses demonstrate clustered microcalcifications and in a number of cases the clusters are the only sign of malignancy [1].

Suspicious calcifications occur in about one-third of breast cancers, and may develop prior to the invasive phase of tumor growth (*in situ* cancer) when cancers are most curable. Calcium deposits are easy to be seen in X-rays because they are much denser (have higher X-ray stopping power) than all types of soft tissues in the breast. Calcifications associated with cancer are usually very small. However calcifications

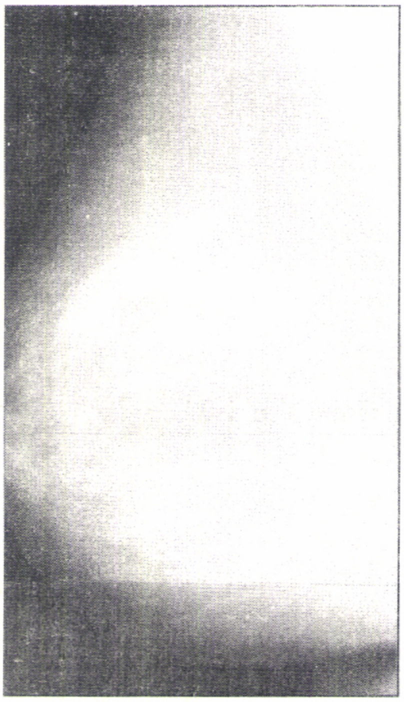
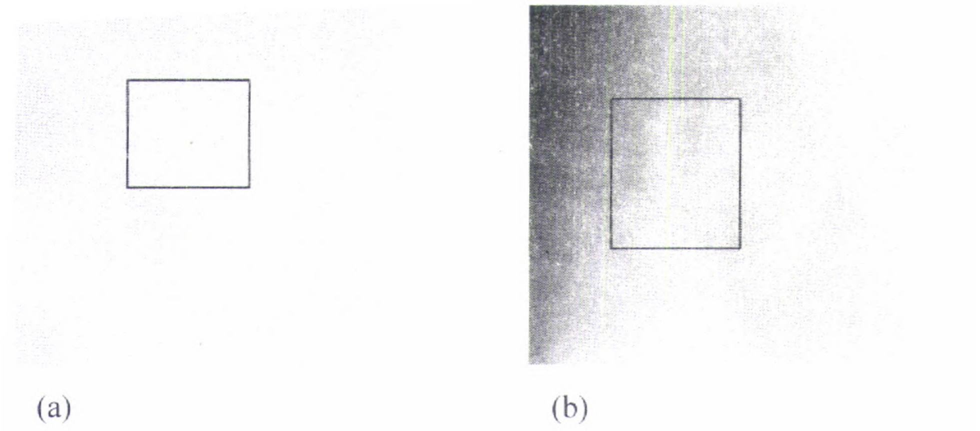


Figure 3.4: Normal Mammogram
Fatty-glandular type

commonly occur in benign breast processes, where they may be confused with cancer. Figure 3.5 shows snippets of mammograms containing malignant and benign microcalcifications.

Malignant microcalcifications vary extremely in form, size, density and number. They are usually clustered within one area of the breast, often within one lobe. They are of two types: granular and casting.

Pleomorphic or heterogeneous calcifications (Granular): These are tiny calcifications with dot-like or elongated shape and innumerable. They are having varying size, usually less than 0.5 mm..

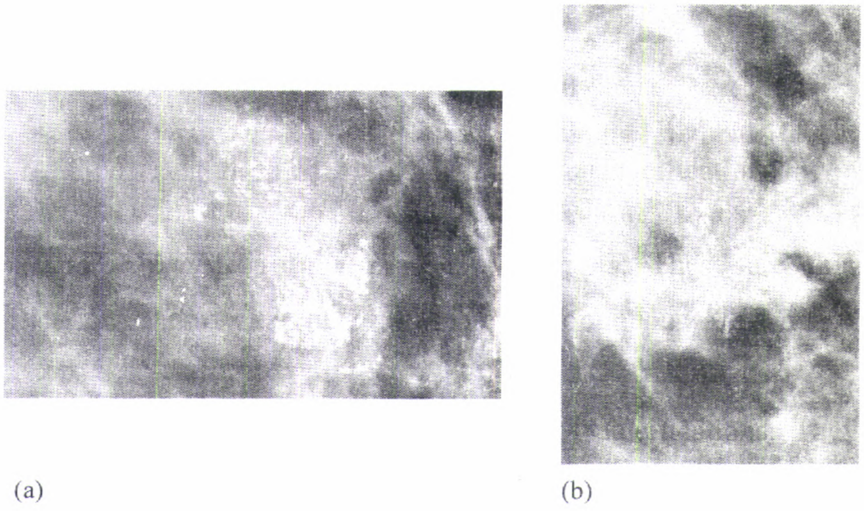


(a) (b)

Figure 3.5: Snippets of mammograms containing microcalcifications
a) Benign calcifications b) Malignant calcifications

Fine and/or branching (casting) calcifications: These are thin, irregular calcifications that appear linear, but are discontinuous and under 0.5 mm in width.

Figure 3.6 shows examples of these types of calcifications. They are often associated with cancer and clearly merit immediate biopsy.

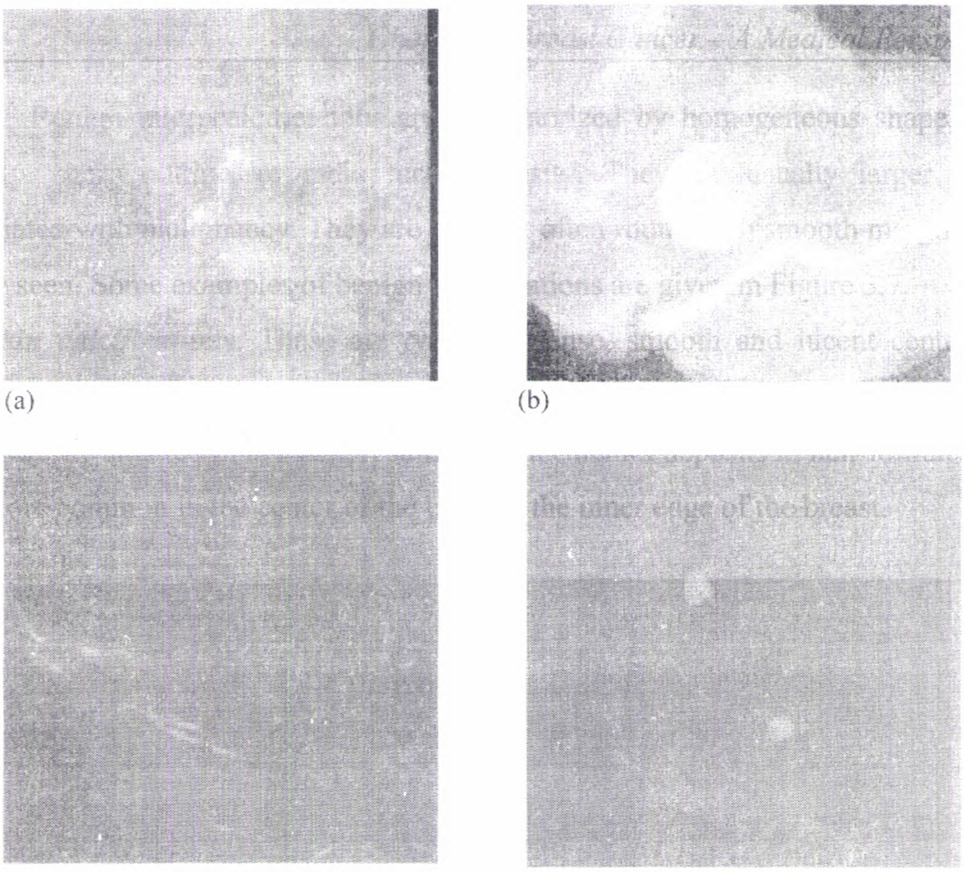


(a) (b)

Figure 3.6: Basic types of malignant microcalcifications
(a) Granular type (b) Casting type

Benign microcalcifications are characterized by homogeneous shape, uniform density, sharp outline, or radio lucent density. They are usually larger than that associated with malignancy. They are coarser, often round with smooth margins and are easily seen. Some examples of benign calcifications are given in Figure 3.7.

Skin calcifications: These are typically dense, smooth and lucent centered (less dense in center than margin) that are *pathognomonic* (appearance is always benign). They are situated in the skin, resulting from calcium deposits in hair follicles and are more common in the center of the chest at the inner edge of the breast.



(a) (b)

(c) (d)

Figure 3.7: Different forms of benign calcifications.
a) Skin calcifications b) Vascular calcifications
c) Rod shaped calcification d) Round calcifications.

Vascular calcifications: These are parallel paired tracks or linear tubular calcifications that are clearly associated with small arteries.

Coarse or popcorn like calcification: Rounded groups of coarse calcifications develop in an involuting *fibro adenoma*. When completely developed the appearance is reliable, but during early phases of development calcifications in fibro adenomas may be suspicious.

Large rod shaped calcification: These are benign calcifications forming continuous rods that may occasionally branch. They are usually more than 1 mm in diameter and may have lucent center, if calcium surrounds rather than fills an enlarged duct. These kinds of calcifications are found in *secreatory disease*, *plasma cell mastitis*, and *duct ectasia*.

Round Calcifications: They are smooth, dense and round calcifications with size less than 1 mm.

Spherical or lucent centered calcifications: These are benign calcifications that range from under 1 mm to over a centimeter. These deposits have smooth surfaces, are round or oval and tend to have a lucent center. They arise from areas of *fat necrosis*, *calcified duct debris* and occasional *fibro adenoma* of a duct involved irregularly by breast cancer.

3.5.1.1 Calcification Distribution Modifiers

Breast cancer frequently spreads locally in characteristic patterns. These patterns are used as modifiers of the basic morphologic description and describe the arrangement of the calcifications. The significance of groups of calcifications is influenced by the pattern of distribution. Multiple similar groups may be indicated when there is more than one group that show similar morphology and distribution.

Grouped or Clustered: The term is used when multiple small calcifications occupy a small volume of tissue (less than two cubic centimeters).

Linear: Calcifications arrayed in a line that may have branch points.

Segmental: These are worrisome in that their distribution suggests deposits in a duct and its branches raising the possibility of multi-focal breast cancer in a lobe or segment of the breast. Although benign causes of segmental calcifications exist such as *secreatory disease*, this distribution is of greater concern when the morphology of the calcifications is not specifically benign.

Regional: These are calcifications scattered in a large volume of breast tissue not necessarily conforming to a duct distribution that are likely benign, but are not everywhere in the breast and do not fit the other more suspicious categories.

Diffuse/Scattered: These are calcifications that are distributed randomly throughout the breast.

Multiple groups: Multiple groups may be indicated when there is more than one group of calcifications that are similar in morphology and distribution.

3.5.2 Circumscribed Masses

The presence of a localized collection of tissue represents a mass. By ACR-BIRAD definition, a mass is a space-occupying lesion seen in 2 different projections (X-ray points of view). When an apparent collection is seen in only one view, it is referred to as a mammographic "density". Although the density may be a mass, perhaps obscured by overlying glandular tissue on other views, it may be nothing more than several overlapping normal areas. When a density is seen on only one view, additional views must be done to confirm or exclude the presence of a mass.

Circumscribed Masses have a distinct border and are typically circular in shape. High-density radio opaque and random oriented masses are most likely to be malignant whereas radio lucent and radio lucent / radio opaque combined masses are almost always benign [1], [31]. Examples of benign and malignant circumscribed masses are shown in Figure 3.8. Halo and capsules are characteristics of benign masses with rare exceptions (see fig.3.9). A halo is a narrow radiolucent ring or a segment of a ring around the periphery of a tumor. A capsule is a thin, curved, radiopaque line that is seen

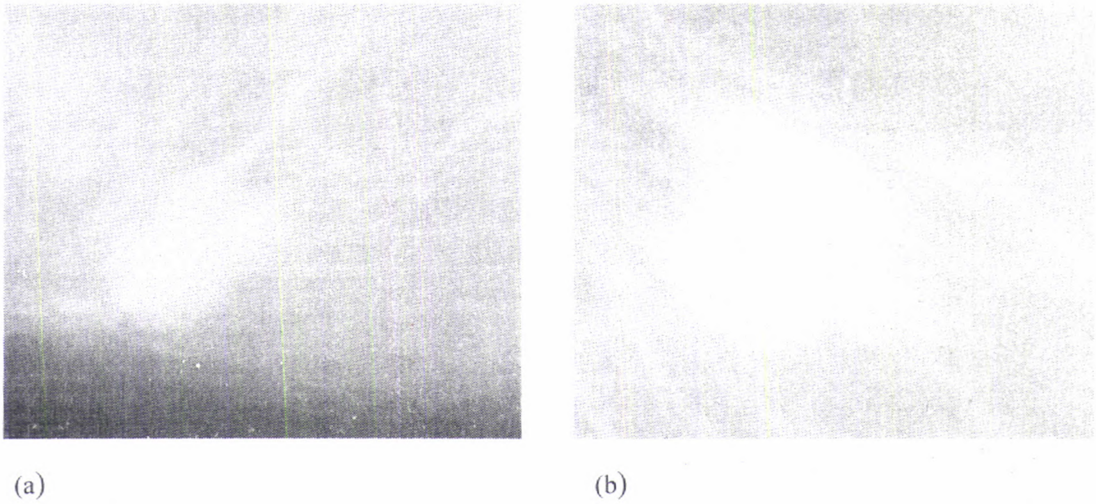


Figure 3.8: Snippets of mammograms with circumscribed masses
a) Benign mass b) Malignant mass

only when it surrounds tumors containing fat. A cyst with smooth borders and orienting in the direction of the nipple following the trabecular structure of the breast also indicates a benign lesion. Figure 3.10 shows some typical examples of different types of malignant masses. Contour, density, shape, orientation and size of the mass are important factors to be considered when analyzing a visible mass.

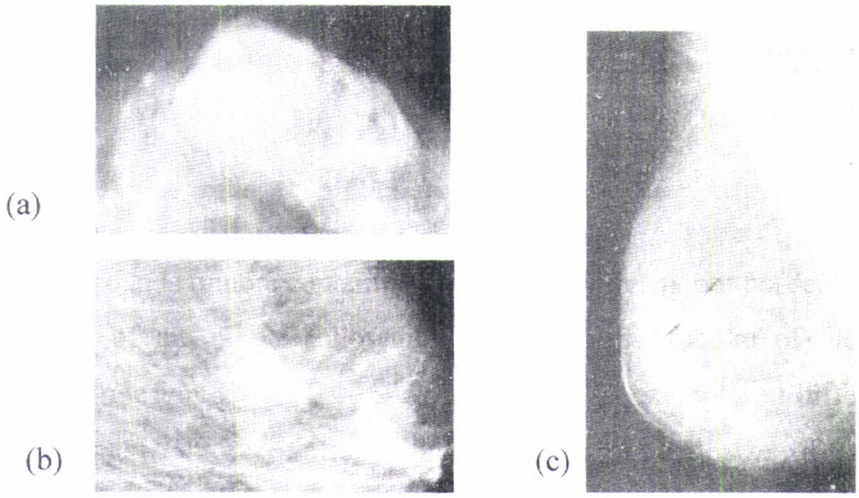


Figure 3.9:Benign masses.
a)halo b)cyst c) capsule

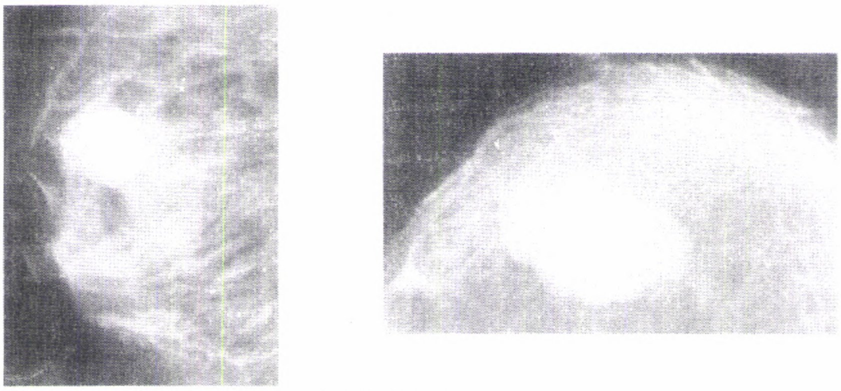
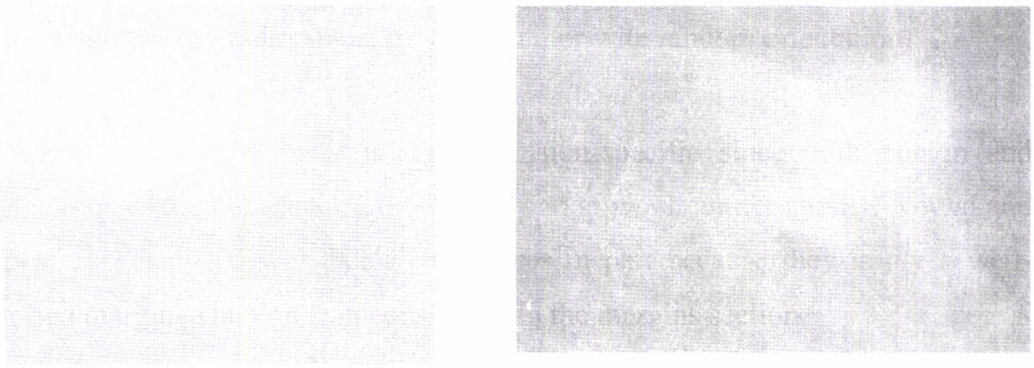


Figure 3.10: Malignant masses.
(a) High density radiopaque. (b) Solid tumor with random orientation.

General shape of a mass is relatively non-specific since both benign and malignant processes tend to arise from one spot and grow circumferentially. Round and oval shapes are associated with benign processes in part because they imply a well-circumscribed margin, a benign sign considered in the margins section.

Round, oval, and lobular shape: Masses in these categories imply a well-defined smooth edge and is often benign. If their margin is not smooth, their shape alone does not tend to exclude malignancy.



(a) (b)

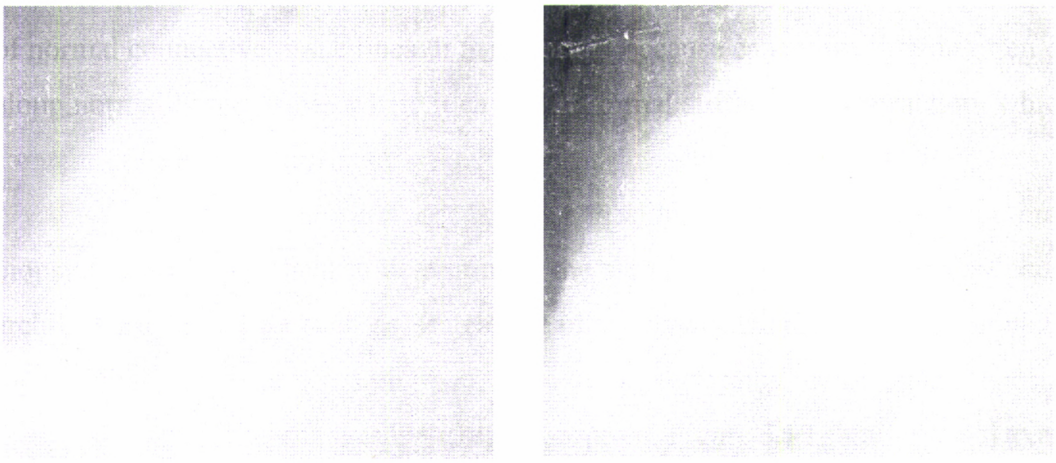
Fig.3.11: Snippets of mammograms with ill-defined masses
a) Benign mass b) Malignant mass

Irregular shape: Irregular shapes are more concerning, in part because they imply indistinct margins and are more often malignant (tumor infiltrating edges). Figure 3.11 shows typical benign and malignant ill-defined masses.

3.5.2.1 Architectural Distortion

In this class, the normal outline of tissues is distorted, sometimes with no definable mass. It includes spiculations (lines radiating from a center) and retraction (puckering) of normal connective tissue lines. It is important because cancer infiltration often occurs along normal tissue planes where it causes abnormal stiffness or contraction, which can sometimes be seen before an actual mass.

Architectural distortion occurs with healing after injury including previous biopsy and so it is critical to determine if the area has been injured. Benign causes of architectural distortion such as scarring tend to remain unchanged or improve. So in cases where previous mammograms have been performed, it is most important to compare to observe if changes have really occurred. Figure 3.12 shows typical examples of benign and malignant architectural distortions.



(a) (b)

Figure 3.12: Snippets of mammograms with architectural distortion
a) Benign distortion b) Malignant distortion

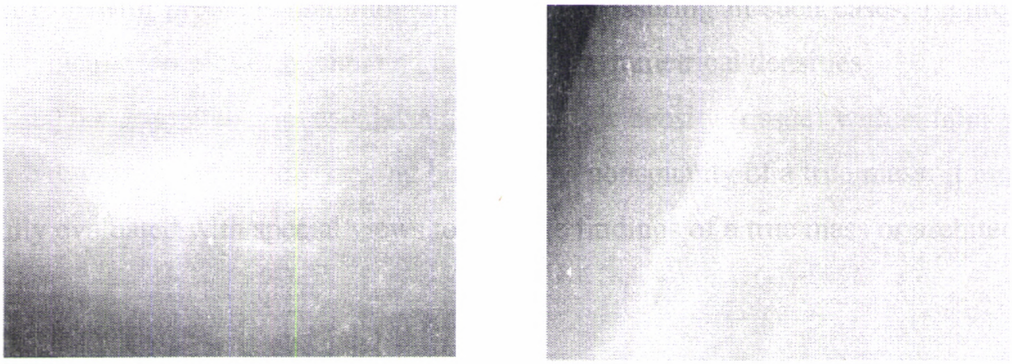
3.5.2.2 Asymmetric Breast Tissue

Breast tissue is usually very similar from one side to the other. When a greater volume or density of tissue is present on one side, concern arises even if no mass is seen. Although asymmetry does occur as an occasional normal finding, it is important to obtain detailed views, usually with small "focal compression" devices to spread tissues out and exclude a mass. Even if no mass is identified, accelerated follow-up mammogram in 6 months time is usual. If the area can be palpated (felt), biopsy must be considered. Once asymmetry is identified, it tends to remain constant over years and comparison with previous mammograms can be reassuring in such cases. Figure 3.13 shows typical examples of benign and malignant asymmetrical densities.

Focal Asymmetric Density: This is an area of breast density (tissue) with similar shape on two views, but completely lacking borders and conspicuity of a true mass. It must be carefully evaluated with special views to exclude findings of a true mass or architectural distortion.

3.5.3 Spiculated Lesions

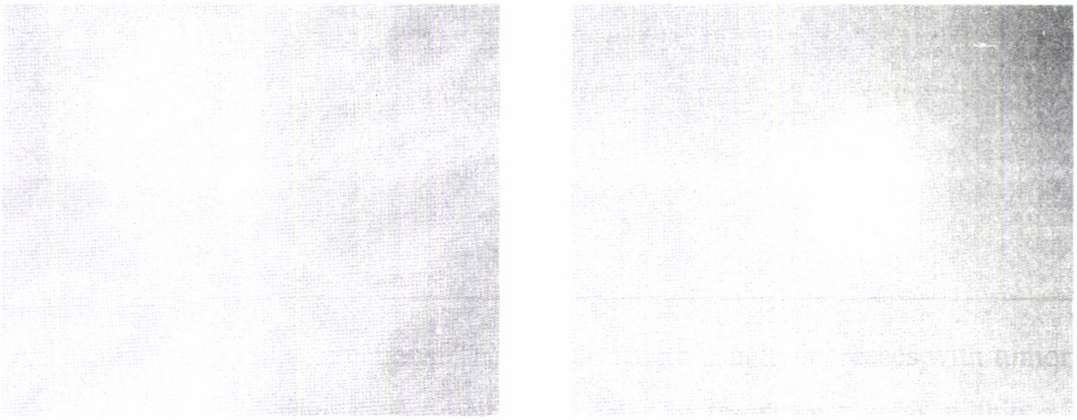
Spiculated lesions appear star shaped with blurred borders (see figure 3.14). They are almost all malignant. Generally the lesion has a distinct central tumor mass with dense



(a) (b)

Figure 3.13: Snippets of mammograms having asymmetric breast tissue.
a) Benign b) Malignant

spicules radiating in all directions. The spicule length usually increases with tumor size. Occasionally, translucent, oval or circular center or translucent areas within a loose structure and low-density spicules characterize benign spiculated lesions.



(a) (b)

Figure 3.14: Snippets of mammograms with spiculated masses
a) Benign mass b) Malignant mass

Chapter 4

Review of Basic Theory

4.1 The Wavelets

The concept of wavelet analysis has been developed in the late 1980's. However, its idea can be traced back to the Littlewood-Paley technique and Calderón-Zygmund theory [162] in harmonic analysis. Wavelet analysis is a powerful tool for time-frequency analysis. In many applied areas like Digital Signal Processing (DSP), time-frequency analysis is critical [163]. To perform a WT we need a wavelet, which is a localized wave that satisfies certain mathematical criteria [164]. These are:

1. A wavelet must have finite energy:

$$E = \int_{-\infty}^{\infty} \left|\psi(t)\right|^{2} dt < \infty \tag{4.1a}$$

2. It should have zero mean, i.e. it should have no zero frequency components

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \quad \text{or} \qquad \hat{\psi}(0) = 0 \tag{4.1b}$$

where $\hat{\psi}(f)$ is the Fourier Transform of $\psi(t)$.

3. The following condition known as admissibility condition must hold good

$$0 < \int_{-\infty}^{\infty} \frac{|\hat{\psi}(f)|^2}{|f|} df < \infty \tag{4.1c}$$

An additional criterion that must hold for complex wavelets is that the Fourier transform must both be real and vanish for negative frequencies.

In 1985, Yves Meyer discovered that one could obtain orthonormal bases for $L^2(R)$ and the expression for decomposing a function into these orthonormal wavelets converged in many function spaces. An orthonormal wavelet system satisfy the following orthogonal relationships [165]:

$$\int \phi_{J,k}(t)\phi_{J,k'}(t)dt = \delta_{k,k'} \tag{4.2a}$$

$$\int \psi_{j,k}(t)\phi_{J,k'}(t)dt = 0 \tag{4.2b}$$

$$\int \psi_{j,k}(t)\psi_{j',k'}(t)dt = \delta_{j,j'}\delta_{k,k'} \tag{4.2c}$$

where $\phi_{i,j}(t)$ and $\psi_{i,j}(t)$ are the scaling and wavelet functions at the i^{th} scale and j^{th} shift respectively.

A particular example of an orthonormal wavelet system was introduced by Alfred Haar [19]. However, the Haar wavelets are discontinuous and therefore poorly localized in frequency.

The conditions of orthogonality are quite stringent. When using the WT these conditions are not necessarily needed. For many of the image processing applications, all that is needed is a reversible transform and the inverse transform should be simple enough. One way of relaxing the orthogonality, while keeping the invertible property is to impose biorthogonality.

Let $\phi(t)$ and $\tilde{\phi}(t)$ be two scaling functions that satisfy the following relations:

$$\phi(t) = 2\sum_{n} g(n)\phi(2t-n) \tag{4.3a}$$

$$\tilde{\phi}(t) = 2\sum_{n} \tilde{g}(n)\tilde{\phi}(2t-n) \tag{4.3b}$$

$$\int \phi_{J,k}(t)\tilde{\phi}_{J,k'}(t)dt = \delta_k \tag{4.3c}$$

Also let $\psi(t)$ and $\tilde{\psi}(t)$ be two functions that integrates to zero with the additional properties that

$$\int \psi_{J,k}(t)\tilde{\psi}_{J,k'}(t)dt = \delta_k \tag{4.4a}$$

$$\int \psi_{J,k}(t)\tilde{\phi}_{J,k'}(t)dt = 0 \tag{4.4b}$$

$$\int \tilde{\psi}_{J,k'}(t)\phi_{J,k}(t)dt = 0 \tag{4.4c}$$

$$\psi(t) = 2\sum_n h(n)\phi(2t-n) \tag{4.4d}$$

$$\tilde{\psi}(t) = 2\sum_n \tilde{h}(n)\tilde{\phi}(2t-n) \tag{4.4e}$$

These 4 functions $\phi(t)$, $\tilde{\phi}(t)$, $\psi(t)$ and $\tilde{\psi}(t)$ form a biorthogonal wavelet system [164].

There are in fact a large number of wavelets to choose from for analyzing our data. The best one for a particular application depends on both the nature of the signal and what physical phenomena or process we are looking to interrogate or how we are manipulating the signal.

4.2 The CWT

The CWT of any square integrable function $f(t)$ is given by equation 1.1. It is a reversible transform and $f(t)$ can be recovered from it using the relation:

$$f(t) = \frac{1}{C_\psi}\int_0^\infty \int_R a^{-2} W_\psi f(a,b)\psi(\frac{t-b}{a})dadb \tag{4.5}$$

where $\psi(t)$ is the mother wavelet function, $W_\psi f(a,b)$ are the CWT coefficients for dilation a and translation b and

$$C_\psi = 2\pi \int_{-\infty}^{\infty} |w|^{-1} \left|\hat{\psi}^*(w)\right|^2 dw < \infty \tag{4.6}$$

$\hat{\psi}(w)$ is the Fourier transform of ψ and $\hat{\psi}^*$ is the complex conjugate of $\hat{\psi}$.

CWT offers time and frequency selectivity. The segment of $f(t)$ that influences the value of $W_\psi f(a,b)$ for any (a,b) is that stretch of $f(t)$ that coincides with the interval over which $\psi_{a,b}$ has the bulk of its energy. This windowing effect results in the time selectivity of the CWT.

The Quality factor is invariant with respect to wavelet dilation because the Fourier Transform $F[\psi(\frac{t}{a})] = |a|\hat{\psi}(aw)$. The center frequency and the 3 dB bandwidth of the dilated wavelet are $\frac{1}{|a|}$ times that of the corresponding values of the mother wavelet, yielding the same value for the quality factor. Thus the continnum of filters to which we alluded is a set of constant Q band pass filters. It is this band pass nature that gives the frequency selectivity of the CWT.

4.3 The DWT

CWT provides a redundant representation of the signal, in the sense that the entire support of $W_\psi(f)(a,b)$ need not be used to recover $f(t)$. The following relation gives a non-redundant wavelet representation:

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d(k,l) 2^{-\frac{k}{2}} \psi(2^{-k}t - l) \tag{4.7}$$

Here, the dilation and translation take values of the form $a = 2^k$ and $b = 2^k l$; k, l being integers. This is called dyadic sampling because consecutive values of the discrete scales as well as the corresponding sampling intervals differ by a factor of 2.

$$d(k,l) = W_\psi f(a,b) \tag{4.8}$$

at $a=2^k$ and $b=2^k l$. $d(k,l)$ the DWT, is still the transform of a continuous time signal. The discretization is only in the dilation and translation variables. The time scale parameters being discrete it has been recognized as a natural WT for discrete time signals [166], [167].

The DWT of a discrete time signal $x(n)\ \varepsilon\ l^2(z)$ with respect to a set of analysis wavelets $(\xi_{n,m})_{n=1,2,\ldots;m\Lambda\ \mathrm{int\,eger}}$ is defined as the set of coefficients

$$X_{n,m} = \sum_k x(k)\xi_{n,m}^{*}(k) \tag{4.9}$$

$\xi_{n,m}(k)$ is obtained by translating $\xi_{n,0}(k)$ by a scale dependent amount that monotonically increases with m at each fixed scale.

$$\xi_{n,m}(k) = \xi_{n,0}(k-2^n m) \tag{4.10}$$

The signal can be recovered using the following expansion,

$$x(k) = \sum_n \sum_m X_{n,m}\psi_{n,m}(k) \tag{4.11}$$

if one can determine a dual set of synthesis wavelet sequences $(\psi_{n,m})_{n=1,2,\ldots;m\Lambda\ \mathrm{int\,eger}}$,on the same time-scale grid, that satisfy the completeness relationship

$$\sum_n \sum_m \xi_{n,m}^{*}(s)\psi_{n,m}(p) = \delta_{s,p} \tag{4.12}$$

where $\delta_{s,p}$ is the Kronecker delta defined as $\delta_{s,p}\ =\ 1\ ;\ s=p$

$= 0$ *otherwise.*

For band-limited signals, a finite-level expansion will be quite sufficient. Hence such signals can be defined by the following finite sum.

$$x(k) = \sum_{n=1}^{N} \sum_m X_{n,m}\psi_{n,m}(k) + \gamma_N(k) \tag{4.13a}$$

where

$$\gamma_N(k) = \sum_m \beta_{N,m}\phi_{N,m}(k) \tag{4.13b}$$

and

$$\beta_{N,m} = \sum_k x(k)\lambda_{N,m}^{*}(k) \tag{4.13c}$$

Here, the sequence $\gamma_N(k)$ represents the residue of the expansion over the finite wavelet set that converges in norm to zero for any arbitrary finite energy signal as $N \to \infty$. $\lambda_{N,m}(k)$ is the analysis scaling sequence at the N^{th} level and $\phi_{N,m}(k)$ is the corresponding synthesis dual.

4.4 Wavelets and Time- Frequency Representation -Concept of MRA

Multi Resolution Analysis (MRA), formulated in 1986 by Mallat and Meyer, provided the frame work for understanding the logic behind the wavelet basis and, subsequently, became the tool for constructing new ones. The concept of MRA is based on the analysis of a function at various levels of resolution. The wavelet decomposition obtained using an MRA is a successive approximation method, which adds more and more projections into detail spaces spanned by the wavelets and their shifts at different scales.

To achieve an MRA of a function x (t), we must have a finite energy function $\phi(t) \in L^2(R)$, called a scaling function, that generates a nested sequence { Aj }, namely $\{0\} \leftarrow \mathrm{K} \subset A_{-1} \subset A_0 \subset A_1 \subset \mathrm{K} \to L^2$ and satisfies a dilation equation $\phi(t) = \sum_k g_0(k)\phi(at-k)$ for some $a>0$ and coefficients $\{g_0[k] \in l^2\}$. The space A_0 is generated by $\{\phi(.-k) : k \in Z\}$ and in general A_s, by $\{\phi_{k,s} : k,s \in Z\}$. Consequently we have the following results.

$$x(t) \in A_s \Leftrightarrow x(2t) \in A_{s+1} \tag{4.14a}$$

$$x(t) \in A_s \Leftrightarrow x(t + 2^{-s}) \in A_s \tag{4.14b}$$

These dilation equations are unique to MRA.

For each s, since A_s is a proper subspace of A_{s+1}, there is some space left in A_{s+1}, called W_s, which when combined with A_s gives A_{s+1}. This space $\{W_s\}$ is called the wavelet sub-space and is complementary to A_s in A_{s+1}, meaning that

$$A_s \, I \;\; W_s = \{0\} \quad , \qquad\qquad s \in Z \tag{4.15a}$$

$$A_s \oplus W_s = A_{s+1} \tag{4.15b}$$

Subspace $\{\ W_s\ \}$ are generated by $\psi(t) \in L^2(R)$, called the wavelet, in the same way as A_s is generated by $\phi(t)$. i.e. for any $x_s(t) \in A_s$

$$x_s(t) = \sum_k a_{k,s}\phi(2^s t - k), \tag{4.16}$$

and any function $y_s(t) \in W_s$ can be written as

$$y_s(t) = \sum_k w_{k,s}\psi(2^s t - k) \tag{4.17}$$

for same coefficients $a_{k,s}, w_{k,s} \in l^2$.

Since we have

$$\begin{aligned} A_{s+1} &= W_s \oplus A_s \\ &= W_s \oplus W_{s-1} \oplus A_{s-1} \\ &= W_s \oplus W_{s-1} \oplus W_{s-2} \oplus \mathrm{K}\ , \end{aligned} \tag{4.18}$$

we have

$$A_s = \bigoplus_{l=-\infty}^{s-1} W_l \tag{4.19}$$

Observe that the $\{A_s\ \}$ are nested while the $\{W_s\ \}$ are mutually orthogonal. Consequently, we have

$$A_l \, I \;\; A_m = A_l, \qquad\qquad m > l \tag{4.20a}$$

$$W_l \, I \;\; W_m = \{0\}, \qquad\qquad l = m \tag{4.20b}$$

$$A_l \, I \;\; W_m = \{0\}, \qquad\qquad l \le m \tag{4.20c}$$

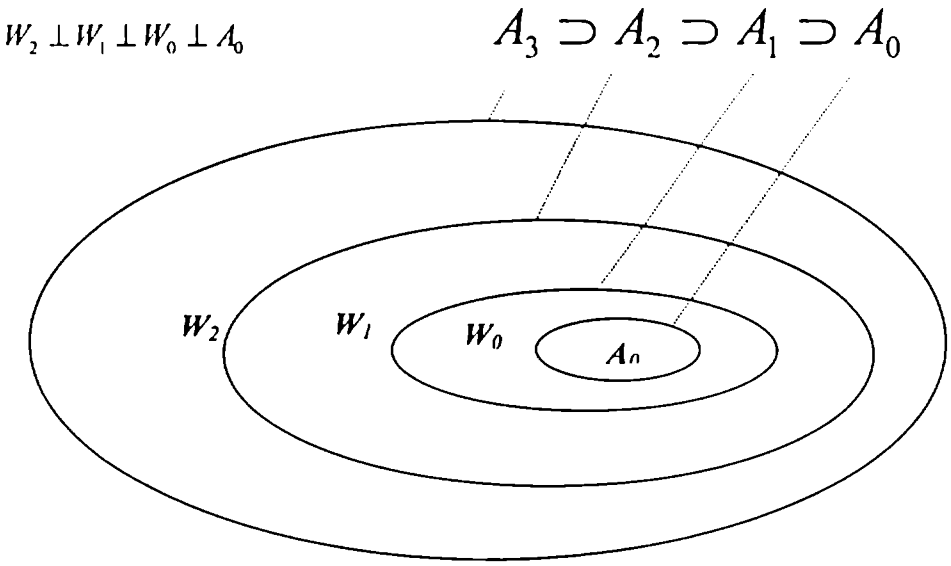


Figure 4.1: Schematic of MRA decomposition

The hierarchical nature of A_s and W_s can be shown by the schematic representation in figure 4.1. In the case of an orthogonal decomposition, in addition to the wavelet space W_s being complementary to A_s, they are mutually orthogonal also, such that $W_s \perp A_s$.

Similar to MRA in $L_2(R)$, a Discrete Multi Resolution Analysis (DMRA) for $l_2(Z)$ has been developed [168], [169], [170], [171]. Accordingly any discrete time sequence $f(n)$ of finite energy can be expressed in terms of the discrete time basis functions $\psi_{j,k}(n)$ as:

$$f(n)=\sum_{j,k} d_j(k)\psi(2^j n-k) \tag{4.21}$$

If the expansion basis functions form a tight frame, the expansion coefficients form an inner product by

$$d_j(k)=\left\langle f(n),\psi(2^j n-k)\right\rangle=\sum_n f(n)\psi(2^j n-k) \tag{4.22}$$

4.5 DWT Computation

Stéphane Mallat made a decisive step in the theory of wavelets when he proposed a fast algorithm for the computation of wavelet coefficients [21], [172]. He developed the pyramidal scheme, relating the wavelet functions to a set of Quadrature Mirror Filterbanks (QMF)[21], [168], [173] that decompose signals into subbands. This involves successive filtering and multirate operations.

4.5.1 Basic Multirate Operations

The basic operations involved in sampling rate conversion of digital signals is decimation, which reduces the sampling rate by an integer M and interpolation, which increases the sampling rate by an integer factor L.

4.5.1.1 Decimation

Decimation or downsampling a sequence $x(n)$ by a factor of M is achieved simply by selecting every M^{th} sample of $x(n)$.

$$Y_D(n)=x(Mn) \tag{4.23}$$

The downsampling operation results in a time-variant system. In general, it may not be possible to recover $x(n)$ from $Y_D(n)$,as it is an aliased version of $x(n)$ with a folding frequency of $F_S / 2M$,where F_S is the sampling frequency of $x(n)$.The Aliasing can be avoided if $x(n)$ is band limited to $F_S / 2M$. Hence before decimation, usually, a low pass filtering is performed, which is characterized by a response $H_D(w)$, where

$$\begin{aligned} H_D(w) &= 1 \qquad |w| \leq \pi/M \\ &= 0 \qquad \textit{otherwise.} \end{aligned} \tag{4.24}$$

The frequency domain characteristics of $Y_D(n)$ is found to be

$$Y_D(e^{jw}) = \frac{1}{M}\sum_{k=0}^{M-1} X(e^{j(w-2\pi k)/M}) \qquad (4.25)$$

4.5.1.2 Interpolation

Interpolation or upsampling by a factor of L can be accomplished by inserting L-1 zeroes between successive samples of the signal followed by a stage of filtering to remove the image frequencies. The upsampled sequence can be expressed as

$$\begin{aligned} Y_U(n) &= x(n/L) && n=0, \pm L, \pm 2L,\ldots && (4.26)\\ &= 0 && \textit{otherwise} \end{aligned}$$

In frequency domain it can be expressed as

$$Y_U(e^{jw}) = X(e^{jw\,L}) \qquad (4.27)$$

whose spectrum is an L-fold periodic repetition of the input signal spectrum. The new sampling rate will be LF_S. Since only the frequency components of x (n) in the range $0 \le w_s = 2\pi F_S \le \pi/L$ are unique, the images of $X(w)$ in $X_U(w)$ beyond this range should be rejected by passing the sequence $x_U(n)$through a filter with frequency response

$$\begin{aligned} H_U(w) &= C && |w| \le \pi/L\\ &= 0 && \textit{otherwise} && (4.28) \end{aligned}$$

where C is a scale factor required to normalize the output sequence.

4.5.1.3 Sampling rate conversion by a rational factor *L/M*

Sampling rate conversion by a factor of L/M can be achieved by first performing interpolation by a factor of L followed by decimation by a factor M. This is shown as a cascaded operation in fig 4.2(a) that can be combined to single filter as in figure 4.2(b). The single filter will be having an ideal frequency response

$$\begin{aligned} H(w) &= L && |w| \le \min(\pi/\mathrm{D}, \pi/L)\\ &= 0 && \textit{otherwise} && (4.29) \end{aligned}$$

'↑' and '↓' indicates upsampling and downsampling operations respectively.

In the time domain, the output of the upsampler is the sequence

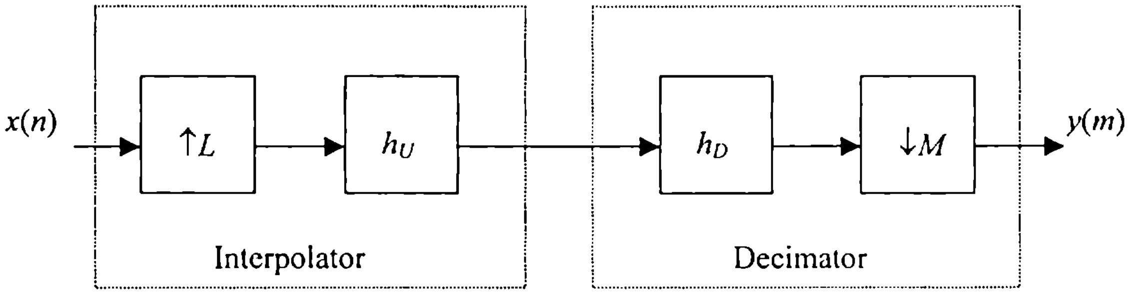
$$v(k) = x(k/L) \qquad k=0, \pm L, \pm 2L, \ldots \tag{4.30}$$
$$= 0 \qquad otherwise$$

and the output of the linear time-invariant filter is

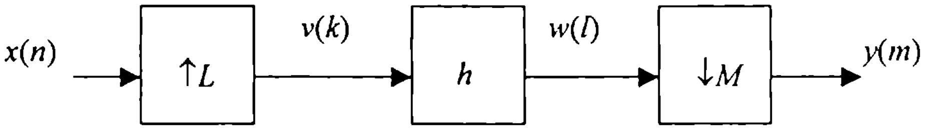
$$w(l) = \sum_{k=-\infty}^{\infty} h(l-k)v(k)$$
$$= \sum_{k=-\infty}^{\infty} h(l-kL)x(k) \tag{4.31}$$

Finally, the output of the sampling rate converter is the sequence $\{y(m)\}$,which is obtained by downsampling the sequence $\{w(l)\}$ by a factor M. Thus

$$y(m) = \sum_{k=-\infty}^{\infty} h(mM-kL)x(k) \tag{4.32}$$



(a)



(b)

Figure 4.2: Fractional sampling rate conversion by multirate technique
(a) Cascaded interpolation and decimation (b) Combined filtering

4.5.2 The pyramidal algorithm

An attractive feature of the wavelet series expansion is that the underlying multiresolution structure leads to an efficient discrete-time algorithm based on a filter bank implementation. This is shown in figure 4.3 for a 2-level DWT computation using the pyramidal tree structure for a discrete-time sequence $x(n)$. The filters h, g, $\tilde{h}$ and $\tilde{g}$ characterize the wavelet system.

The transform coefficients at the j^{th} level are computed from the scaling output of the previous stage of the analysis structure as

$$X_j(k) = (\beta_{j-1}(r) * h(r)) \downarrow 2 \qquad (4.33a)$$

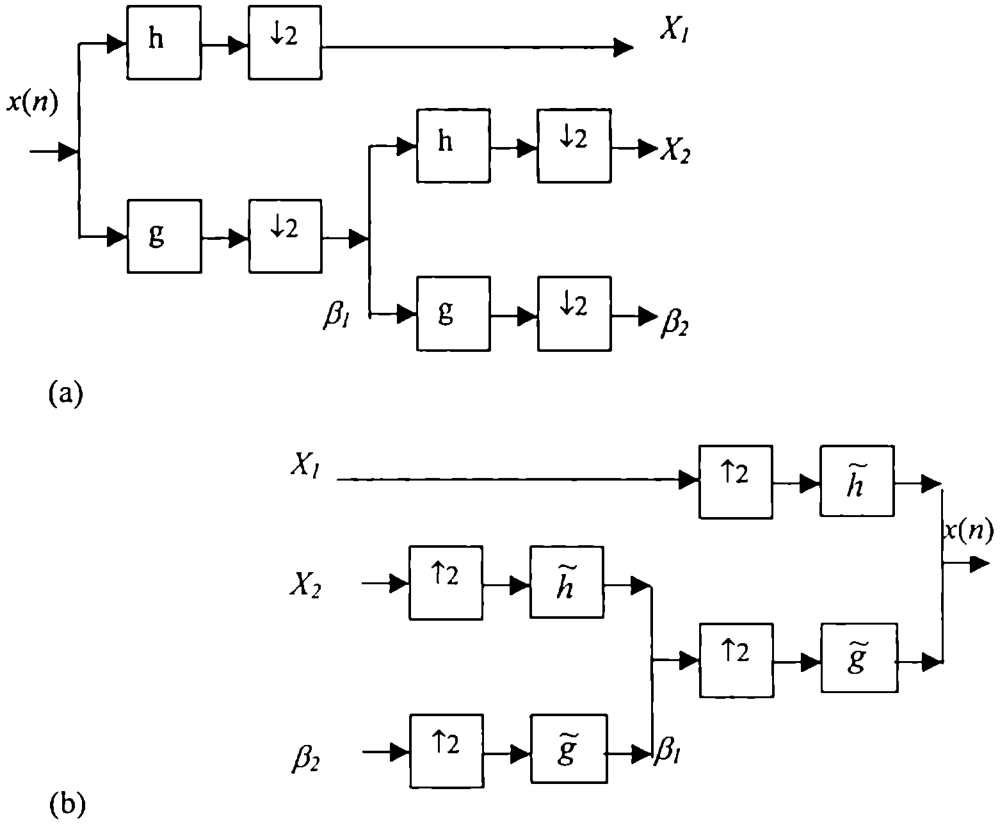


Figure 4.3: Pyramid structure for 2-level DWT computation
(a) Decomposition (b) Reconstruction

$$\beta_j(k) = (\beta_{j-1}(r) * g(r)) \downarrow 2 \tag{4.33b}$$

where $\beta_0(k) = x(k)$, $h(k)= \psi_{0,0}(k)$, $g(k) = \phi_{0,0}(k)$. Similarly, employing the scaling and wavelet sequences of the upper levels, the scaling coefficients at each of the immediate lower levels are computed using the multirate filtering operations.

The first stage of the two banks in figure 4.3(a) divides the spectrum of the input signal into a low pass and high pass band, resulting in the scaling coefficients and wavelet coefficients at a lower scale. The second stage then divides the low pass band into another lower low pass band and a band pass band. The first stage divides the spectrum into two equal parts. The second stage divides the lower band into quarters and so on. This results in a logarithmic set of bandwidths as illustrated in figure 4.4. Correspondingly the time-frequency tiling will appear as in figure 4.5. Each horizontal strip in the tiling corresponds to each channel, which in turn corresponds to a scale j. The span covered by each of the translations is marked on the horizontal axis. The residue of the finite scale representation is a low pass signal that reproduces the trend of the input signal. The remaining terms in the expansion represent the fluctuations of the signal over this trend at various scales.

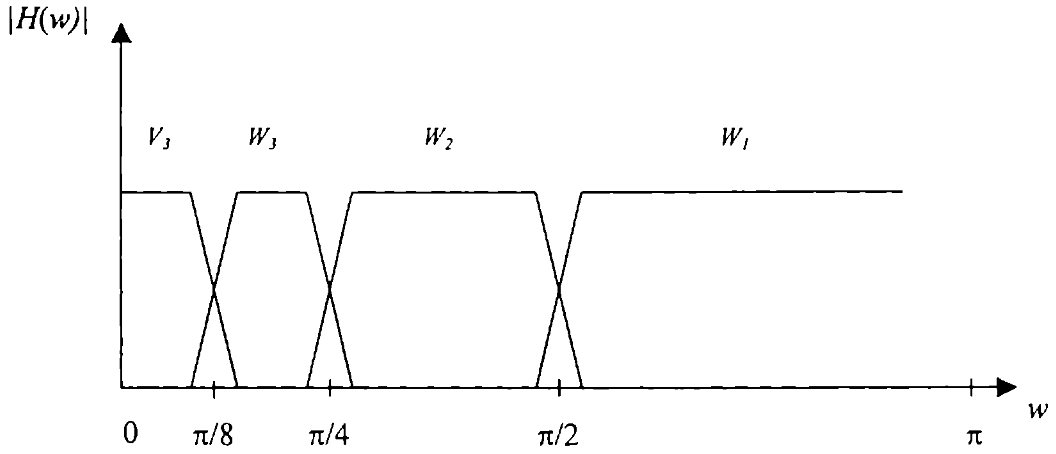


Figure 4.4: Frequency bands for the analysis tree of the pyramid

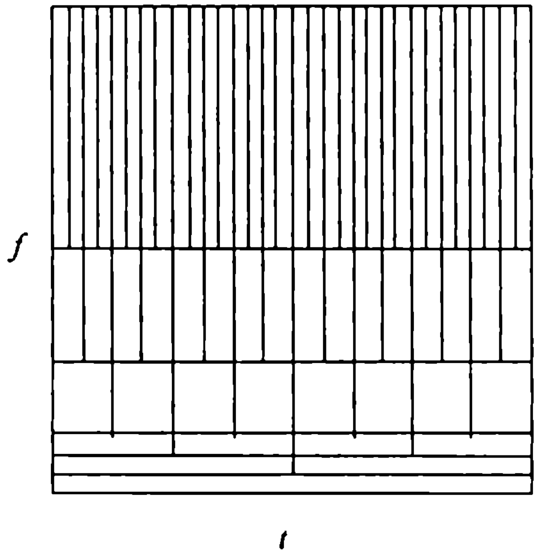


Figure 4.5: Time-frequency tiling

4.6 Computation of 2-D DWT

4.6.1 2-D Wavelets

Wavelets are basis functions in a vector space comprising of a scaling function ϕ with its associated wavelet function ψ and their dual functions $\tilde{\phi}$ and $\tilde{\psi}$. The basis functions at scale j and translation or shift k in the case of 1-D may be denoted by $\{\phi_{j,k}\}$ and $\{\psi_{j,k}\}, j,k \in Z$, where

$$\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k) \qquad (4.34a)$$

and

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) \qquad (4.34b)$$

The dual functions are defined in a similar way. In order to apply wavelet decomposition to images, 2-D extension of wavelets are required. This can be achieved by the use of separable or non-separable wavelets. Here, only separable wavelets are considered.

From the 1-D basis, one can construct a 2-D separable wavelet basis with four basis functions; one scaling function $\phi^0_{j,k,l}(x,y)$ and three wavelet functions $\psi^\tau_{j,k,l}(x,y)$, $\tau\in\{1,2,3\}$ given by

$$\phi^0_{j,k,l}(x,y)=\phi_{j,k}(x)\phi_{j,l}(y) \tag{4.35a}$$

$$\psi^1_{j,k,l}(x,y)=\phi_{j,k}(x)\psi_{j,l}(y) \tag{4.35b}$$

$$\psi^2_{j,k,l}(x,y)=\psi_{j,k}(x)\phi_{j,l}(y) \tag{4.35c}$$

$$\psi^3_{j,k,l}(x,y)=\psi_{j,k}(x)\psi_{j,l}(y) \tag{4.35d}$$

These basis functions span the four j-level linear vector spaces rather than just two as in the 1-D case. An analogous definition holds for the dual scaling function $\tilde{\phi}^0_{j,k,l}(x,y)$ and wavelet function $\tilde{\psi}^\tau_{j,k,l}(x,y)$.

4.6.2 2-D Wavelet Transform

The M-level wavelet representation of a 2-D function f is given by

$$f(x,y)=\sum_{k,l}c^M_{k,l}\phi^0_{M,k,l}(x,y)+\sum_{j=1}^{M}\sum_{\tau\in\tau}\sum_{k,l}d^{j,\tau}_{k,l}\psi^\tau_{j,k,l}(x,y) \tag{4.36}$$

The approximation and the detail coefficients in the above expression are $c^M_{k,l}=\left\langle f,\tilde{\phi}^0_{M,k,l}\right\rangle$ and $d^{j,\tau}_{k,l}=\left\langle f,\tilde{\psi}^\tau_{j,k,l}\right\rangle$ respectively, where $\langle .,.\rangle$ denotes the inner product in the $l^2(Z^2)$ space.

For fast DWT computation using the Mallat's pyramid algorithm, a sub band filtering scheme is used, where ϕ and ψ are represented by the corresponding discrete filters $g(n)$ and $h(n)$, $n\in Z$, respectively, called the decomposition or analysis filters. Furthermore, there exist the reconstruction or synthesis filters as the dual filters $\tilde{g}\,(n)$ and $\tilde{h}\,(n)$.

The 2-D basis given by equation (4.35) may be represented by the four possible tensor products gg, gh, hg and hh of the 1-D filters g and h. Let c^j and $d^{j,\tau}$ denote the

2-D sequences $c^j_{k,l}$ and $d^{j,\tau}_{k,l}$; $k,l \in Z$ and $\tau=\{1,2,3\}$. The scaling and wavelet transform coefficients at a courser level j+1 are computed from c^j by convolution followed by downsampling as follows

$$c^{j+1} = (\downarrow\downarrow 2)(gg * c^j) \tag{4.37a}$$

$$d^{j+1,1} = (\downarrow\downarrow 2)(gh * c^j) \tag{4.37b}$$

$$d^{j+1,2} = (\downarrow\downarrow 2)(hg * c^j) \tag{4.37c}$$

$$d^{j+1,3} = (\downarrow\downarrow 2)(hh * c^j) \tag{4.37d}$$

for $j = 0,1,..,M$-1. Here $*$ denotes 2-D convolution and ($\downarrow\downarrow$2) denotes downsampling by a factor of 2 in both x and y directions. The given image is treated as c^0.

To perform the 2-D DWT computation as above, instead of using the 2-D filters, one can employ a separable extension of the 1-D decomposition algorithm [174]. The rows of the data are convolved with the first 1-D filter and every other column is retained. The resulting data is then convolved column-wise using the other 1-D filter. Further stages of 2-D decomposition are obtained by recursively applying the procedure to the low-pass filtered output of the previous stage.

Following the common terminology found in the literature, the coefficients given by equation (4.37) may be denoted as LL, LH, HL and HH partials respectively as shown in the layout of figure 4.6.

The wavelet reconstruction is performed recursively, starting at level M by upsampling (denoted by $\uparrow\uparrow$2) followed by convolution using the dual filters. The signal reconstructed at the j^{th} level from coefficients at the j+1th level may be expressed as:

$$c^j = \tilde{g}\tilde{g} * ((\uparrow\uparrow 2)c^{j+1}) + \tilde{g}\tilde{h} * ((\uparrow\uparrow 2)d^{j+1,1}) + \tilde{h}\tilde{g} * ((\uparrow\uparrow 2)d^{j+1,2}) + \tilde{h}\tilde{h} * ((\uparrow\uparrow 2)d^{j+1,3}) \tag{4.38}$$

Figure 4.7 shows the block diagram for a 2-level decomposition and reconstruction of an image using the pyramidal algorithm.

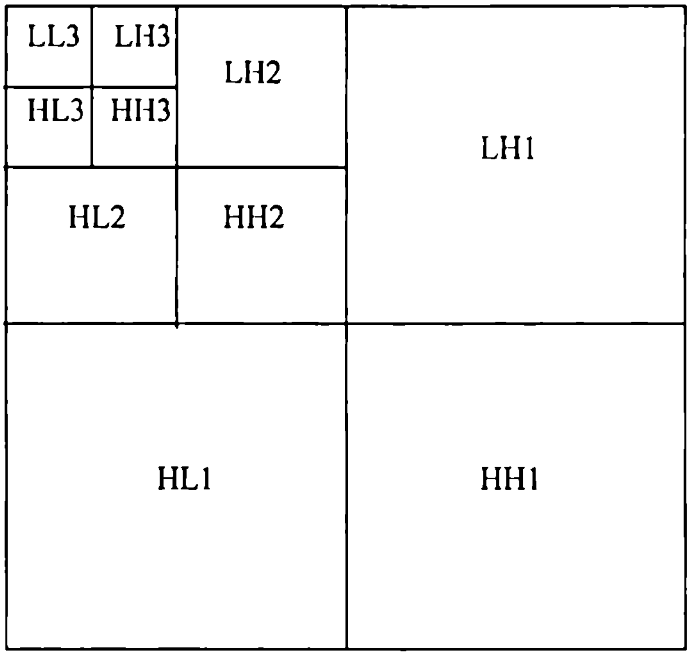


Figure 4.6: Coefficient layout of a 3-level DWT of an image

4.7 The MWT

The concept of MWT was first proposed by Evangelista [25], as a class of transforms for the representation of pseudo periodic signals with constant period. This transform simplifies the analysis of the pseudo periodic signals by decomposing them into a regular asymptotically periodic signal and a number of fluctuations over this signal.

The MWT of a signal $x(n)$, of period M, is defined as the set of coefficients

$$X_{j,k,q} = \sum_{n} x(n)\varsigma_{j,k,q}(n) \tag{4.39}$$

where $j = 1,2\ldots$; k is an integer; q=0,1,2...M-1. $\varsigma_{j,k,q}(n)$ is the multiplexed wavelets, defined as

$$\varsigma_{j,k,q}(n) = \sum_{s} \psi_{j,k}(s)\delta(n - sM - q) \tag{4.40}$$

given a complete and orthonormal set of ordinary wavelets $\psi_{j,k}(n)$. The Inverse MWT (IMWT) is given by

$$x(n)=\sum_{j}\sum_{k}\sum_{q=0}^{M-1}X_{j,k,q}\varsigma_{j,k,q}(n) \tag{4.41}$$

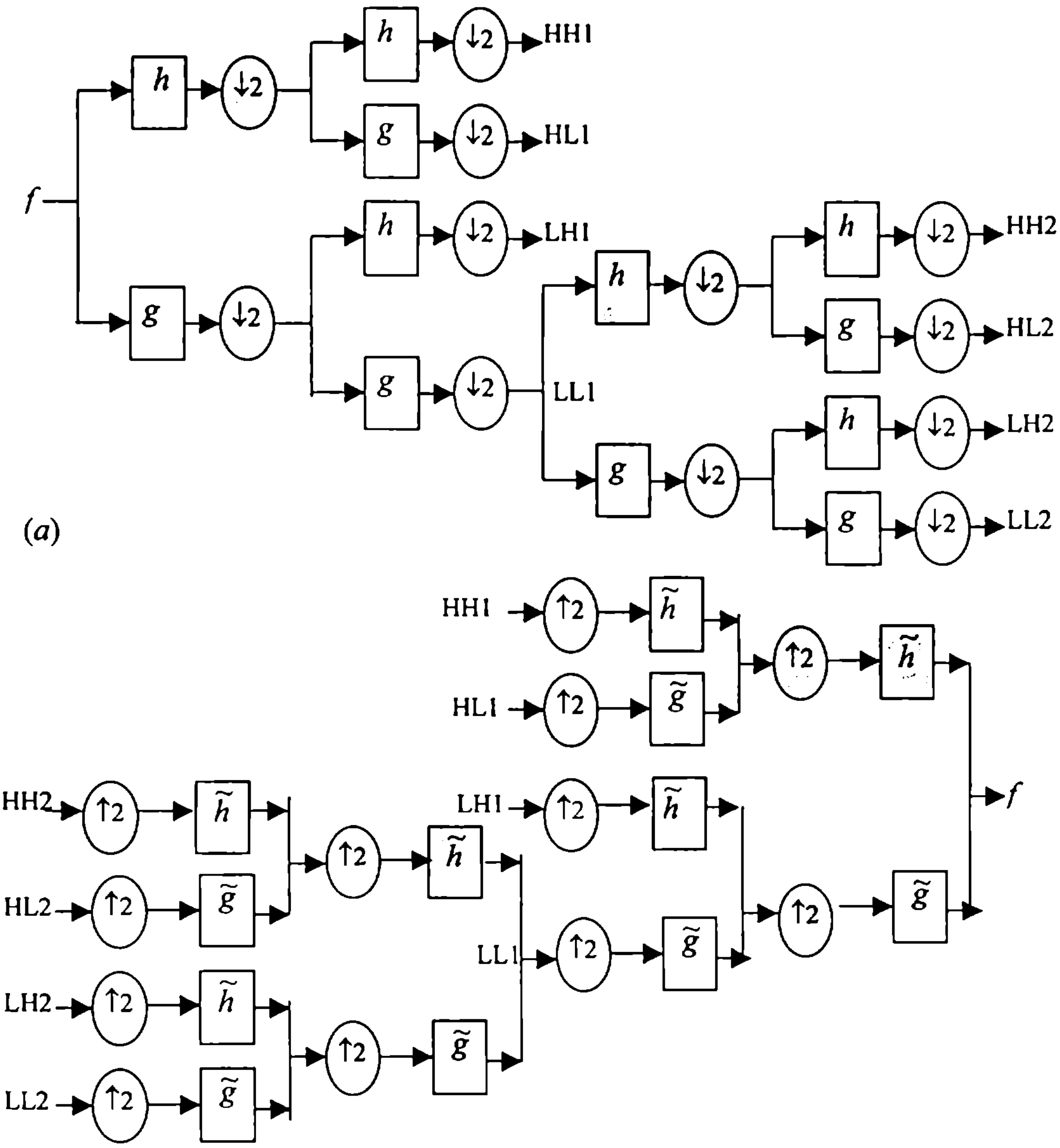


Figure 4.7: Pyramidal Structure for 2-level 2-D DWT computation. (a) Decomposition. (*b*) Reconstruction. (Shaded blocks represent row wise operation, the rest being column-wise)

In this work, images are treated as oscillatory signals, although they are not periodic in a strict mathematical sense. Taking the periods along the horizontal and vertical directions to be the width and length of the image segment respectively, any pixel in an image of size LxW can be represented as $V_q(r)$ or $V_r(q)$ where r =1,2...L-1 and q =1,2...W-1.

Considering its periodicity along the vertical direction alone, the J-level MWT of the image may be expressed as

$$\hat{V}_{j,k,q} = \sum_r V_q(r)\psi_{j,k}(r) \; ; \tag{4.42a}$$

and

$$\hat{\sigma}_{J,k,q} = \sum_r V_q(r)\phi_{J,k}(r) \tag{4.42b}$$

where j = 1,2...J ;$k \in Z$ and $\phi_{j,k}$, the scaling function associated with the wavelet function $\psi_{j,k}$. $\hat{V}_{j,k,q}$ are the MWT coefficients of the signal $V_q(r)$ at j^{th} scale and k^{th} shift and $\hat{\sigma}_{J,k,q}$ are the multiplexed scaling transform coefficients of the same section of the signal at J^{th} scale and k^{th} shift (residue after J-levels of MWT decomposition). Similar expressions are obtained by considering the image as a 1-D signal along the horizontal direction also.

From equation (4.42) we can see that in the MWT the WT is taken over samples that are spaced one period apart. Hence, the inter-period fluctuations of the signal are better sieved out in the wavelet partials, whereas the residue holds the asymptotically periodic information. It is quite different for the DWT, where the oscillatory part as well as the fluctuations gets filtered into different wavelet and scaling partials altogether, depending on the frequency content of the signal. As the edges correspond to inter-period fluctuations, accurate reconstruction of the edges can be achieved from the MWT partials. Hence, MWT is more suited to edge detection applications than DWT.

4.8 Edge detection

An edge in an image is a contour across which the brightness of the image changes abruptly. It is often interpreted in image processing as one class of singularities. In a function, singularities can be characterized easily as discontinuities where the gradient approaches infinity. However, image data is discrete and edges in an image often are defined as the local maxima of the gradient.

Edge detection is an important tool in pattern recognition, image segmentation, and scene analysis. An edge detector is basically a highpass filter that can be applied to extract the edge points in an image. Many classical edge detectors have been developed, based on the principle of matching local image segments with specific edge patterns. The popular operators belonging to this class are Roberts, Sobel, Prewitt, FreiChen and Laplacian operators [175], [176], [177], [178]. They are realized by convolving the image with a set of directional derivative masks defined on a 3 by 3 pattern grid [175]. Hence, they are efficient and easy to apply.

However, classical edge detectors usually fail to handle images with strong noise. To reduce the influence of noise, two techniques were developed from 1979 to 1984; filtering the images with the Gaussian before edge detection [179], [180], [181] and approximating the image with a smooth function [178], [182]. The weakness of the above approaches is that the optimal result may not be obtained by using a fixed operator. A computational approach to edge detection was developed in 1991, in which an optimal detector was approximated by the first derivative of a Gaussian [183].

4.8.1 Edge detector using wavelets

Edges in images can be mathematically defined as local singularities. Until recently, the Fourier transforms was the main mathematical tool for analyzing singularities. However, the Fourier transform is global and not well adapted to local singularities. Wavelet analysis is a local analysis and is especially suitable for time-frequency analysis [184], which is essential for singularity detection. With the growth of wavelet theory, the WT has been found to be a remarkable mathematical tool to analyze the singularities including the edges, and further, to detect them effectively. This idea is similar to that of approximating an optimal detector by the first derivative of a Gaussian [183]. It has been proved that [35], [185] the maxima of the WT modulus can detect the location of the irregular structures. Further, a numerical procedure to calculate their Lipschitz exponents has been provided. 1-D and 2-D signals can be reconstructed, with a good approximation, from the local maxima of their WT modulus.

The WT characterizes the local regularity of signals by decomposing signals into elementary building blocks that are well localized both in space and frequency. This not only explains the underlying mechanism of classical edge detectors, but also indicates a way of constructing optimal edge detectors under specific working conditions.

Most multiscale edge detectors smooth the signal at various scales and detect sharp variation points from their first or second derivatives. The extrema of the first derivative corresponds to the zero-crossings of the second derivative and to the inflation points of the smoothed signal. It is proved that if a wavelet is the second derivative of a smoothing function, the zero crossings of the WT indicate the location of the sharper signal variations [35].

Any function $\theta(x)$ whose integral is equal to 1 and that converges to 0 at infinity can be considered as a smoothing function. An example is a Gaussian function. Let $\theta(x)$ be a function twice differentiable and let $\psi^a(x) = \frac{d\theta(x)}{dx}$ and $\psi^b(x) = \frac{d^2\theta(x)}{dx^2}$. By definition $\psi^a(x)$ and $\psi^b(x)$ can be considered to be wavelets because their integral is equal to zero. The WT of a function $f(x)$ at scale s and position x computed with respect to $\psi^a(x)$ and $\psi^b(x)$ is defined by:

$$W_s^a f(x) = f * \psi_s^a(x) \quad ; \tag{4.43a}$$

and

$$W_s^b f(x) = f * \psi_s^b(x) \quad ; \tag{4.43b}$$

$$W_s^a f(x) = s\frac{d}{dx}(f * \theta_s)(x); \quad (4.43c)$$

and

$$W_s^b f(x) = s^2\frac{d^2}{dx^2}(f * \theta_s)(x); \quad (4.43d)$$

The local extrema of W_s^a thus correspond to the zero crossings of W_s^b and to the inflation points of $f * \theta_s(x)$. In the particular case $\theta(x)$ is a gaussian, the zero-crossing detection is equivalent to a Marr-Hildreth (M-H) edge detection [186] and extrema detection corresponds to a Canny edge detection [183].

This can be easily extended to the 2-D case. Here, a 2-D smoothing function $\theta\,(x,y)$ whose integral over x and y equal to 1 and converges to 0 at infinity is used. Edges are defined as points (x_0,y_0) where the modulus of the gradient vector is the maximum in the direction towards which the gradient vector points in the image plane $f * \theta_s(x,y)$. Relating this to a 2-D WT, one can locate the edge points from the two components $W_s^a f(x,y)$ and $W_s^b f(x,y)$, as discussed in the 1-D case.

In addition to the WT techniques, another major area linked with the present work is the ANNs. The basic theory of the relevant ANNs is briefed in the next section.

4.9 Artificial Neural Networks Technology

Artificial Neural Networks are relatively crude electronic models based on the neural structure of the brain. An ANN structure is composed of a number of interconnected units called artificial neurons. Each of these units has an Input / Output (I/O) characteristics and implements a local computation or function. The output of any unit is determined by its I/O characteristics, its interconnection to other units, and possibly external units. And the network develops an overall functionality through one or more forms of training [187].

The biologically inspired methods of computing are thought to be the next major advancement in the computing industry. The brain modeling promises a less technical way to develop machine solutions and provides a more graceful degradation during system overload than its traditional counterparts. Now, advances in biological research promises an initial understanding of the natural thinking mechanism and shows that the brain store information as patterns. Some of these patterns are very complicated and allow us to recognize individual faces from many different angles. This process of storing information as patterns and utilizing those patterns for solving problems encompasses a new field in computing. This field, as mentioned before, does not utilize traditional programming but involves the creation of massively parallel networks and the training of them to solve specific problems. This field also utilizes words much different from traditional computing like " behave, react, self-organize, learn, generalize, and forget".

The exact workings of the human brain are still a mystery. Yet, some aspects of this amazing processor are known. In particular, the most basic element of the human brain is a specific type of cell, which unlike the rest of the body, does not appear to degenerate. These cells, known as neurons, are the only part of the body that is not slowly replaced. Hence, it is assumed that they provide us with our abilities to remember, think, and apply previous experiences to our every action. Each of these neurons can connect up to 200,000 other neurons, although 1,000 to 10,000 connections are typical. The power of the human mind comes from the sheer numbers of these basic components and the multiple connections between them. It also comes from genetic programming and learning.

The individual neurons are complicated. They have a myriad of parts, sub-systems, and control mechanisms. They convey information via a host of electrochemical pathways. There are over one hundred different classes of neurons, depending on the classification method used. Together these neurons and their connections form a process, which is not binary, stable and synchronous. In short, it is nothing like the currently available electronic computers or even ANNs. ANNs try to replicate only the most basic elements of this complicated, versatile and powerful organism. They do it in a primitive way. But for the software engineer who is trying to solve problems, neural computing was never about replicating human brains. It is about machines and a new way to solve problems.

4.9.1 Artificial Neurons

The fundamental processing element of a neural network is a neuron. Within humans there are many variations on this basic type of neuron, which further complicates man's attempts at electrically replicating the process of thinking. Yet, all natural neurons have the same four basic components as shown in figure 4.8. These components are known by their biological names - dendrites, soma, axon, and synapses. Dendrites are hair-like extensions of the soma, which act like input channels. These input channels receive their input through the synapses of other neurons. The soma then processes these incoming signals over time and turns the processed value into an output, which is sent out to other neurons through the axon and the synapses.

The basic unit of neural networks, the artificial neurons, simulates the four basic functions of natural neurons. Figure 4.9 shows a fundamental representation of an artificial neuron. Each of the various inputs to the network $x(n)$ is multiplied by a connection weight represented by $w(n)$. In the simplest case, these products are summed, fed through a transfer function to generate a result and then output.

Some applications like the recognition of text, the identification of speech, and

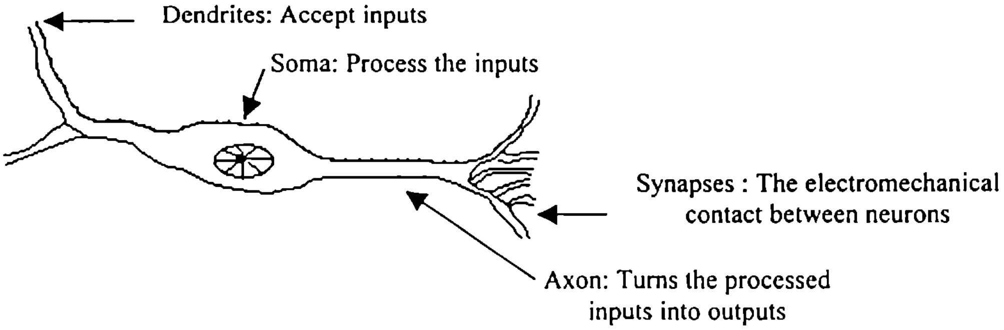


Figure 4.8: A simple neuron

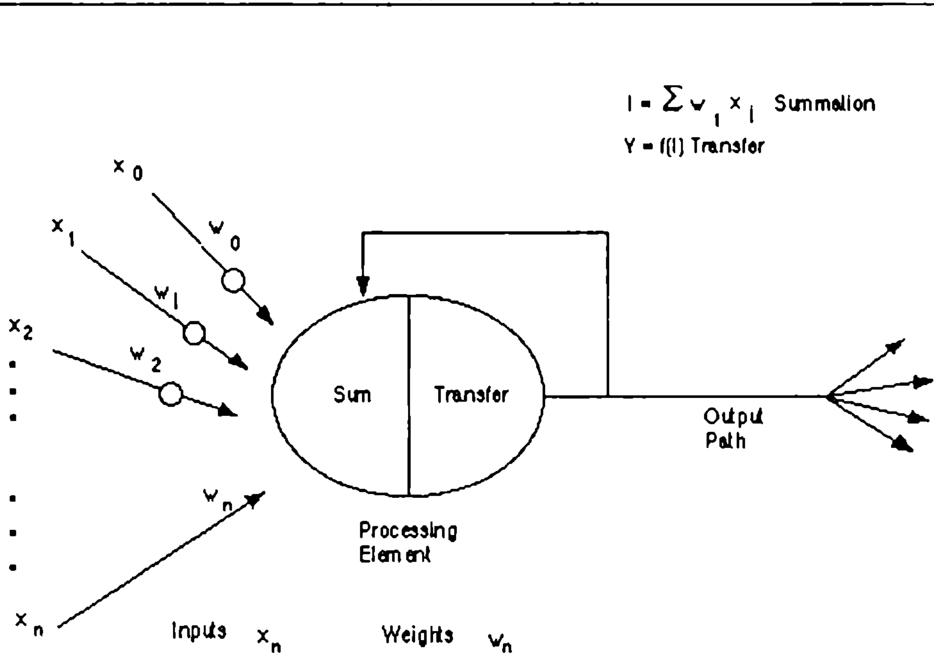


Figure 4.9: A Basic Artificial Neuron.

the image deciphering of scenes require binary answers. Because of this limitation of output options, these applications do not always utilize networks composed of neurons that simply sum up and thereby smooth inputs. These networks may utilize the binary properties of ORing and ANDing of inputs. These functions, and many others, can be built into the summation and transfer functions of a network. Other applications might simply sum and compare to a threshold producing one of two possible outputs, a zero or a one or scale the outputs to match the application. Some functions even integrate the input data over time, creating time-dependent networks.

The output of the summing function is then sent to a transfer function, which turns this number into a real output via some algorithm. The transfer functions that are commonly supported are sigmoid, sine, hyperbolic tangent, etc. Transfer functions can also scale the output or control its value via thresholds. The result of the transfer function is usually the direct output of the processing element. Figure 4.10 shows a sigmoid transfer function.

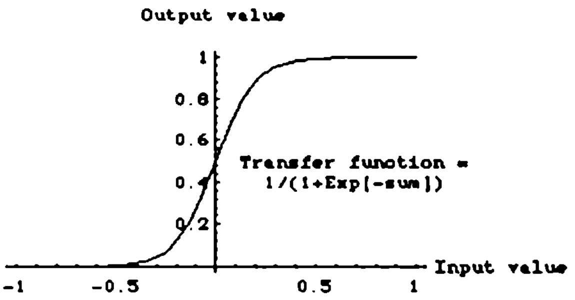


Figure 4.10: Sigmoid Transfer function.

Basically, all ANNs have a similar structure or topology as shown in Figure 4.11. Here, some of the neurons interface to the real world to receive its inputs and some others provide the real world with the network's outputs. All the rest of the neurons are hidden from view. Although there are useful networks, which contain only one layer or even one element, most applications require networks that contain at least three normal types of layers - input, hidden, and output. The layer of input neurons receives the data either from input files or directly from electronic sensors in real-time applications. The output layer sends information directly to the outside world, to a

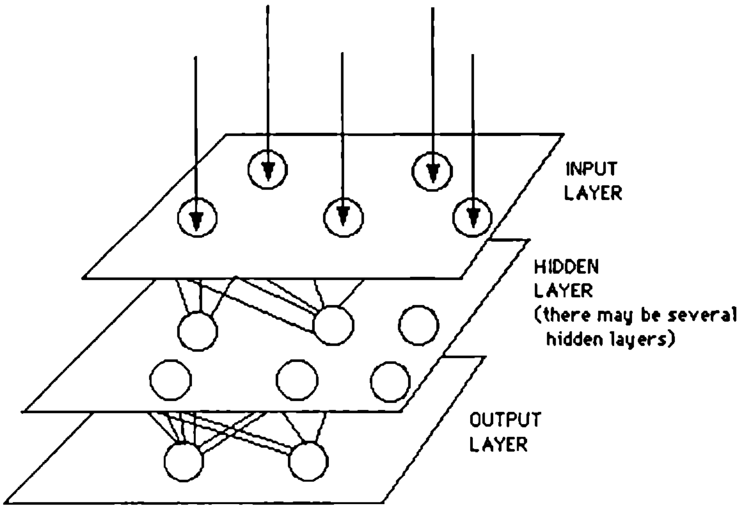


Figure 4.11: A Simple Neural Network Diagram.

secondary computer process, or to other devices such as a mechanical control system. Between these two layers there can be many hidden layers. These internal layers contain many of the neurons in various interconnected structures. The inputs and outputs of each of these hidden neurons go to other neurons.

In most networks each neuron in a hidden layer receives the signals from all of the neurons in a layer above it, typically an input layer. After a neuron performs its function it passes its output to all of the neurons in the layer below it, providing a feed forward path to the output. Some networks want a neuron to inhibit the other neurons in the same layer. This is called lateral inhibition. The most common use of this is in the output layer. For example in text recognition if the probability of a character being a "P" is 0.85 and the probability of the character being an "F" is 0.65, the network needs to choose the highest probability and inhibit all the others. It can do that with lateral inhibition. This concept is also called competition. Another type of connection is feedback, where the output of one-layer routes back to a previous layer. An example of competition and feedback is shown in Figure 4.12.

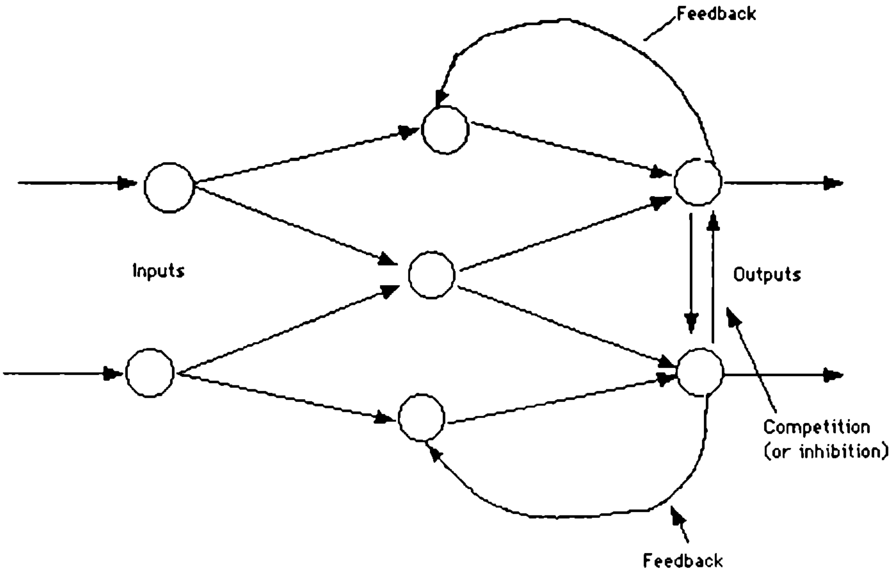


Figure 4.12: Simple Network with Feedback and Competition.

4.9.2 Teaching an ANN

4.9.2.1 Supervised Learning

The vast majority of ANN solutions have been trained with supervision. In this mode, the actual output of a neural network is compared to the desired output. The network then adjusts weights, which are usually randomly set to begin with, so that the next iteration will produce a closer match between the desired and the actual output. The learning method tries to minimize the current errors of all processing elements. This global error reduction is created over time by continuously modifying the input weights until acceptable network accuracy is reached.

With supervised learning, the ANN must be trained before it becomes useful. Training consists of presenting input and output data, which is often referred to as the training set, to the network. The training phase can consume a lot of time. Training is considered complete when the neural network reaches a user defined performance level. This level signifies that the network has achieved the desired statistical accuracy as it produces the required outputs for a given sequence of inputs. When no further learning is necessary, the weights are typically frozen for the application. Some network types allow continual training, at a much slower rate, while in operation. This helps a network to adapt to gradually changing conditions.

Training sets need to be fairly large to contain all the needed information for the network to learn the features and relationships that are important. In addition, the training sessions must include a wide variety of data. If the network is trained just one example at a time, all the weights set so meticulously for one fact could be drastically altered in learning the next fact. The previous facts could be forgotten in learning something new. As a result, the system has to learn everything together, finding the best weight settings for the total set of facts.

How the input and output data is represented or encoded is important in successfully instructing a network. Artificial networks only deal with numeric input data. Therefore, the raw data must often be converted from the external environment.

Additionally, it is usually necessary to scale the data or normalize it to the network's paradigm.

After a supervised network performs well on the training data, it is important to see what it can do with data it has not seen before. If a system does not give reasonable outputs for this test set, the training period is not over. Indeed, this testing is critical to ensure that the network has not simply memorized a given set of data but has learned the general patterns involved within an application.

4.9.2.2 Unsupervised Learning

Unsupervised learning is a great promise for the future. It shouts that computers can someday learn on their own in a true robotic sense. Currently, this learning is limited to networks known as self-organizing maps [188]. It has been proven to be more effective than many algorithmic techniques for numerical aerodynamic flow calculations.

This promising field of unsupervised learning is sometimes called self-supervised learning. These networks internally monitor their performance. They look for regularities or trends in the input signals and make adaptations according to the function of the network. Even without being taught whether it is right or wrong, the network still must have some information about how to organize itself. This information is built into the network topology and learning rules.

An unsupervised learning algorithm might emphasize cooperation among clusters of processing elements. If some external input activated any node in the cluster, its activity as a whole could be increased. Likewise, if external input to nodes in the cluster is decreased, that could have an inhibitory effect on the entire cluster.

Competition between processing elements could also form a basis for learning. Training of competitive clusters could amplify the responses of specific groups to specific stimuli. Normally, when competition for learning is in effect, only the weights belonging to the winning processing element will be updated.

4.9.2.3 Learning Rates

The rate at which ANNs learn depends upon several controllable factors. Obviously, a slower rate means a lot more time is spent in accomplishing the off-line learning to produce an adequately trained system. With the faster learning rates, however, the network may not be able to make the fine discriminations possible with a system that learns more slowly.

Generally, several factors such as network complexity, size, paradigm selection, architecture, type of learning rules employed and desired accuracy, besides time have to be considered when discussing the off-line training task. Changing any one of these factors may either extend the training time to an unreasonable length or even result in an unacceptable accuracy.

Most learning functions have some provision for a learning rate or learning constant. Usually this term is positive and the value is between zero and one. If the learning rate is greater than one, it is easy for the learning algorithm to overshoot in correcting the weights, and the network will oscillate. Small values for the learning rate will not correct the current error quickly, but if small steps are taken in correcting errors, there is a good chance of arriving at the best minimum convergence.

4.9.2.4 Learning Laws

Many learning laws are in common use. Most of these laws are some sorts of variation of the best-known and oldest learning law, Hebb's Rule. A few of the major laws are presented below as examples.

Hebb's Rule: The first, and undoubtedly the best known, learning rule was introduced by Donald Hebb. His basic rule is that if a neuron receives an input from another neuron and if both are highly active (mathematically have the same sign), the weight between the neurons should be strengthened.

Hopfield Law: It is similar to Hebb's rule with the exception that it specifies the magnitude of the strengthening or weakening. It states that if the desired output and the input are both active or inactive, increment the connection weight by the learning rate, otherwise decrement the weight by the learning rate.

The Delta Rule: This is one of the most commonly used rules. It is based on the simple idea of continuously modifying the strengths of the input connections to reduce the difference (the delta) between the desired output value and the actual output of a processing element. This rule changes the synaptic weights in the way that minimizes the mean squared error of the network. This rule is also referred to as the Widrow-Hoff Learning Rule and the Least Mean Square (LMS) Learning Rule.

When the Delta Rule is implemented the delta error in the output layer is transformed by the derivative of the transfer function and is then used in the previous neural layer to adjust input connection weights. In other words, this error is back propagated into previous layers, one layer at a time, until the first layer is reached. When using the delta rule, it is important to ensure that the input data set is well randomized. Well-ordered or structured presentation of the training set can lead to a network, which cannot converge to the desired accuracy. When that happens, the network is incapable of learning the problem.

The Gradient Descent Rule: This rule is similar to the Delta Rule in that the derivative of the transfer function is still used to modify the delta error before it is applied to the connection weights. Here, however, an additional proportional constant tied to the learning rate is appended to the final modifying factor acting upon the weight. This rule is commonly used even though it converges to a point of stability very slowly.

Kohonen's Learning Law: This procedure, developed by T. Kohonen, was inspired by learning in biological systems [188]. Here, the processing elements compete for the opportunity to learn or update their weights. The processing element with the largest output is declared the winner and has the capability of inhibiting its competitors as well as exciting its neighbors. Only the winner is permitted an output and only the winner and its neighbors are allowed to adjust their connection weights.

Further, the size of the neighborhood can vary during the training period. The usual paradigm is to start with a larger definition of the neighborhood and narrow in as the training process proceeds. Because the winning element is defined as the one that has the closest match to the input pattern, Kohonen networks model the distribution of the inputs. This is good for statistical or topological modeling of the data and is sometimes referred to as self-organizing maps or self-organizing topologies.

4.10 Feature Extraction for classification

Neural networks are frequently employed to classify patterns based on learning from examples. While different neural network paradigms employ different learning rules, all these paradigms determine pattern statistics from a set of training samples and then classify new patterns on the basis of these statistics.

The texture is found to be the main descriptor for all kinds of mammograms. In this work, we concentrate on statistical descriptors that include averages, standard deviations and higher-order statistics of intensity values and also on the SGLD Matrix for texture description.

4.10.1 Statistical descriptors

The four gray level sensitive histogram moments, mean (μ), variance (σ^2), skewness (μ_3) and kurtosis (μ_4) are described in this section.

Mean: It represents the (probability-weighted) average value for the random variable. Mean of a distribution can be thought as being analogous to a center of mass.

$$\mu = \sum_{k=1}^{N} f_k n_k / n \tag{4.44}$$

where N denotes the number of gray levels in the mammogram, f_k is the k^{th} gray level n_k is the number of pixels with f_k gray-level and n is the total number of pixels in the region considered.

Variance: Variance is a parameter that measures how dispersed a random variable's probability distribution is. The positive square root of variance is called standard deviation.

$$\sigma^2 = \sum_{k=1}^{N} (f_k - \mu)^2 n_k / n \tag{4.45}$$

Skewness: Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness of the normal distribution or any perfectly symmetric distribution is zero.

$$\mu_3 = \frac{1}{\sigma^2} \sum_{k=1}^{N} (f_k - \mu)^3 n_k / n \tag{4.46}$$

Kurtosis: Kurtosis is the degree of peakedness of a distribution, defined as a normalized form of the fourth central moment of a distribution. It is based on the size of a distribution's tails. Kurtosis is the classical measure of nongaussianity. It can be positive or negative. Distributions with relatively large tails have a negative kurtosis and are called subgaussian or "leptokurtic". Those with small tails and positive kurtosis are called supergaussian or "platykurtic." A distribution with the same kurtosis as the normal distribution is called "mesokurtic." The kurtosis of a normal distribution is 0.

$$\mu_4 = \frac{1}{4} \sum_{k=1}^{N} (f_k - \mu)^4 n_k / n - 3 \tag{4.47}$$

The two distributions shown in figure 4.13 have the same variance; approximately the same skew, but differ markedly in kurtosis.

Kurtosis, or rather its absolute value, has been widely used as a measure of

Figure 4.13: Two distributions having same variance and skew, but different kurtosis a) Leptokurtic distribution b) Platykurtic distribution

nongaussianity in ICA and related fields. The main reason is its computational and theoretical simplicity. However, kurtosis has also some drawbacks in practice when its value has to be estimated from a measured sample. The main problem is that kurtosis can be very sensitive to outliers [189]. Its value may depend on only a few observations in the tails of the distribution, which may be erroneous or irrelevant. In other words, kurtosis is not a robust measure of nongaussianity.

4.10.2 Textural features

Texture feature extraction has long been recognized to be a fundamental task in image analysis, segmentation and classification. In particular, the employment of texture features in medical imaging has proved to be valuable. Two basic methods for texture description exist: statistical and structural [190]. Statistical methods employ features extracted from the image, which measure coarseness, contrast, directionality and other textural characteristics whereas structural methods describe texture by means of primitive descriptions and primitive placement rules.

4.10.2.1 SGLD Matrix

The SGLD method is very powerful for statistical texture description. According to [191], all known visually distinct texture pairs can be discriminated using the above method. In medical imaging, its performance has been shown to be one of the best, especially in Ultrasonic, MR and CT image analysis [192], [193]. SGLD method is based on second order statistics (the spatial relationships of pairs of gray levels). Texture is described by extracting a number of textural features from the image. The texture feature extraction method introduced by Haralick [194] has been proven to be one of the best in overall performance [193].

The SGLD method is based on the estimation of the second order joint conditional probability density functions,

$$f(i,j \mid d,\theta), \theta = 0^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0, 315^0 .$$

Each $f(i,j \mid d,\theta)$ is the probability of going from gray level i to gray level j. It is constructed by counting the number of occurrences of pixel pairs at a given displacement d for a given direction θ.

The co-occurence matrix or SGLD matrix [190] , [194] is used as the underlying structure for storing the textural content of an image that is employed by the SGLD method for the computation of the features characterizing the texture of the image. The co-occurrence matrix is computed by specifying a distance between the pixels in each pair (displacement vector) and whether the direction of the vector is important or not. In particular, to compute the SGLD matrix for an image I (i,j), a displacement vector $d = (x,y)$ is defined. The $(i,j)^{th}$ element of the SGLD matrix:

$$s(i,j \mid d,\theta) = \frac{N(i,j \mid d,\theta)}{\sum_{i,j} N(i,j \mid d,\theta)} \tag{4.48}$$

$N(i,j \mid d,\theta)$ is the number of occurrences of grey levels i and j at a distance d in I (i,j). The matrix is then normalized so that it can be treated as a probability density function.

If a texture is coarse and d is small compared to the sizes of the texture elements, the pairs of points at the intersample distance should usually have similar gray levels. This means that the probability distribution in the matrix is concentrated on or near its diagonal. On the other hand, for a fine texture, the gray levels of the points separated by the distance should be quite different so that the probability distribution is distributed away from its diagonal.

The most significant disadvantage of the co-occurence matrix is its dependency on the number of gray levels in the entire image [193], [195], [196]. Since texture is usually measured in a small region, a large number of entries are zero contributing nothing to the texture description of the region. The computational time for the texture feature extraction operations include the time for processing these also. The above problems become more serious when the examined images are composed of a large number of gray levels.

In texture classification, individual elements of the SGLD matrix are rarely used. Instead, features are derived from the matrix. A large number of textural features have been proposed starting with the original thirteen features described by Haralick [194] namely energy, entropy, correlation, local homogeneity, inertia, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy, information measure of correlation 1, and information measure of correlation 2. However, only some of these are in wide use.

Wezka *et al*[197] used four of Haralick's fourteen features, Conners and Harlow [198] used 5 features and Conners *et al* [199] introduced two new features which address a deficiency in the Conners and Harlow set.

In this work, ten texture features are evaluated from the SGLD matrix for fixed d and θ. They are:

1.Angular second moment or Energy : $\sum_{i,j} s(i,j)^2$ (4.49)

2.Entropy : $-\sum_{i,j} s(i,j)\ \log s(i,j)$ (4.50)

3.Correlation $\sum_{i,j} \frac{s(i,j)\ ij - \mu_x \mu_y}{\sigma_x \sigma_y}$ (4.51)

4.Local Homogenity : $\sum_{i,j} \frac{1}{1+(i-j)^2} s(i,j)$ (4.52)

5.Contrast : $\sum_{n} n^2 \sum_{i,j:n=|i-j|} s(i,j)$ (4.53)

6.Sum of squares variance : $\sum_{i,j} (i-\mu)^2 s(i,j)$ (4.54)

7.Sum average $\sum_{i=2}^{2N_g} i s_{x+y}(i)$ (4.55)

8.Sum variance : $\sum_{i=2}^{2N_g} i + \sum_{j=2}^{2N_g} s_{x+y}(j) \log(s_{x+y}(j))$ (4.56)

9.Sum entropy $\sum_{i=2}^{2N_g} s_{x+y}(i) \log(s_{x+y}(i))$ (4.57)

10.Information measure of correlation: $$\frac{HXY - HXY1}{\max\{HX, HY\}} \tag{4.58}$$

where μ_x, μ_y, σ_x, σ_y are the means and the standard deviations of $s_x(i)$, $s_y(j)$, and s_{x+y} is the mean of the density function $s(i,j)$.

$$s_x(i) = \sum_k s(i,k) \tag{4.59a}$$

$$s_y(j) = \sum_k s(k,j) \tag{4.59b}$$

$$s_{x+y}(i) = \sum_{j,k;j+k=i} s(j,k) \tag{4.59c}$$

HX and HY are the entropies of s_x and s_y and

$$HXY1 = -\sum_{i,j} s(i,j) \log s_x(i) s_y(j) \tag{4.59d}$$

4.11 Networks for classification

4.11.1 Back Propagation Neural Networks (BPNN)

BPNN are multilayer feed-forward networks with differentiable transfer functions, which can perform function approximation, pattern association and pattern classification. The term back propagation refers to the process by which derivatives of network error with respect to network weights and biases are computed. This process can be used with a number of different optimization strategies. The architecture of a multilayer network is not completely constrained by the problem to be solved. The number of inputs to the network is constrained by the problem and the number of neurons in the output layer is constrained by the number of outputs required by the problem. However, the number of layers between network inputs and the output layer and the sizes of the layers are up to the designer to decide.

A general multilayer feed-forward network is illustrated in figure 4.14. This is a feed-forward fully connected hierarchical network consisting of an input layer, one or more hidden layers and an output layer. Real-valued n-dimensional input feature vectors

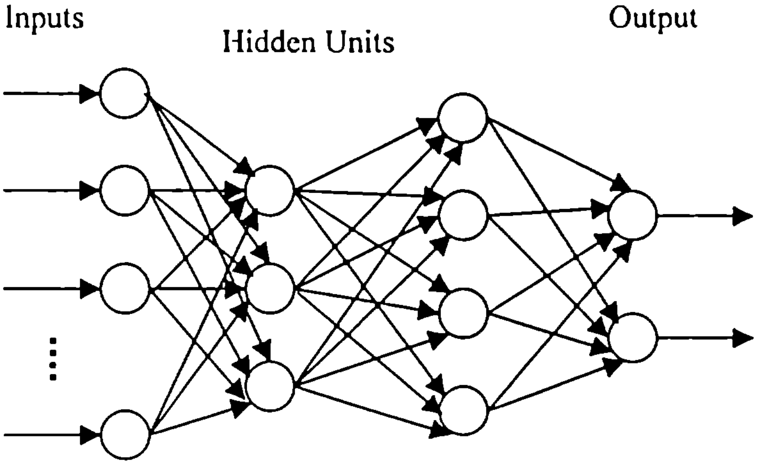


Figure 4.14: A general multilayer feed-forward network

x are presented to each of the first hidden layer units through weights w_{ji}. Hidden layer unit j receives input i through the synoptic weight w_{ji}, $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,h$. It computes a function of the input signal x and the weight vectors w_{ji} and passes its output forward to all of the units in the next successive layer. The second hidden layer also compute a function of their inputs and weight vectors and pass their output to the next layer. This process is repeated until the final computation is produced by the output layer.

There are several different back propagation training algorithms. They have a variety of different computation and storage requirements and no one algorithm is best suited for all purposes. The drawbacks of this method are the local minima, the long training time and the stability of the learned patterns. This algorithm tries to adjust the weights to yield a minimum error with the desired outputs. But the network can get trapped in a local minimum when there is a much deeper minimum nearby, which yields a less accurate solution. This algorithm also requires lots of supervised training and there is no guarantee that the system will converge.

4.11.2 Competitive networks

A competitive network learns to categorize the input vectors presented to it. If a neural network only needs to learn to categorize its input vectors, then a competitive network is the best choice. Competitive networks also learn the distribution of inputs by dedicating more neurons to classifying parts of the input space with higher densities of input. The basic element of a competitive learning algorithm is a mechanism that allows

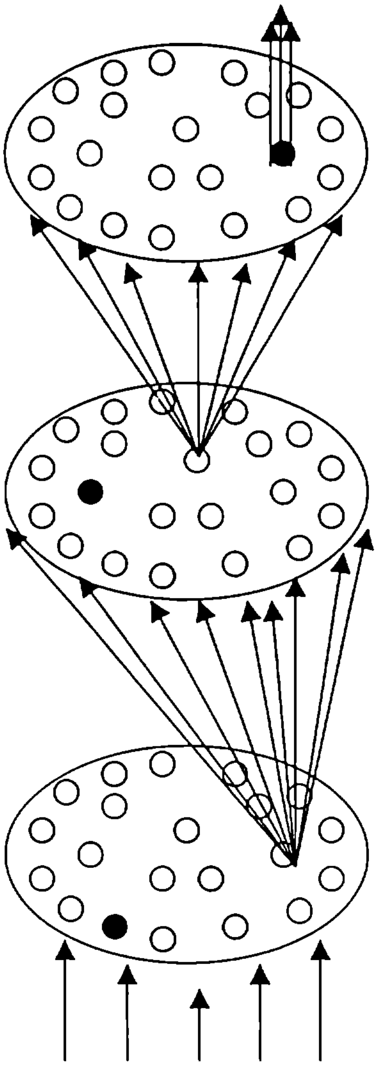


Figure 4.15: General competitive network architecture

units to competitively respond to a given stimulus. Usually for a specific input, only one unit or group of neurons is denoted as the winner. These are treated specially during the training phase. A basis for competitive behavior in biological systems may be the orientation- and location – sensitive arrangements of neural cells in the receptive fields. A three-layer competitive network is shown in figure 4.15.

4.11.3 Radial Basis Function Networks (RBFN)

In the nervous system of biological organisms there is evidence, for the presence of neurons whose response characteristics are tuned to some region of the input space. An example is the orientation-sensitive cells of the visual cortex. The RBF emulates the behavior of such biological networks. RBF is a feed forward structure with a modified hidden layer and training algorithm. RBF networks may require more neurons than standard feed-forward back-propagation networks, but often they can be designed in a fraction of the time it takes to train standard feed-forward networks. They work best when many training vectors are available.

Figure 4.16 shows the RBFN structure. This consists of two layers, a locally

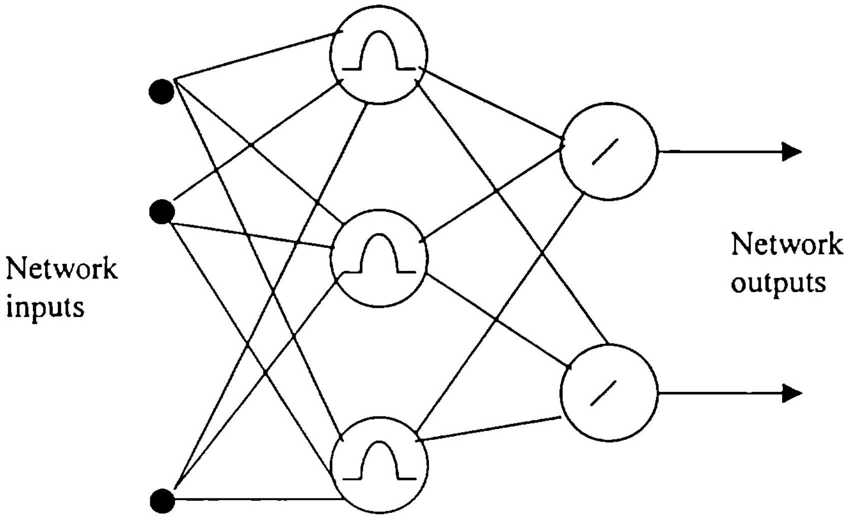


Figure 4.16: RBF Network

tuned or locally sensitive hidden layer and an output layer consisting of linear units. In hidden layer units, the response is localized and decreases as a function of the distance of inputs from the unit's receptive field center.

Methods described above often use heuristic approaches to discover underlying class statistics [200]. These approaches usually involve many incremental modifications to networks parameters in order to improve system performance gradually. They require long computation times for training. Furthermore, the incremental adaptation approaches can be susceptible to converging to local minima of an error function. Statistical classifiers like Bayesian classifiers can be used to eliminate the disadvantages of heuristic approaches. The PNN is an example of a Bayesian classifier.

4.11.4 PNN

The PNN was developed by Donald Specht [201], [202]. This network provides a general solution to pattern classification problems by following an approach developed in statistics called Bayesian classifiers. Bayes theory, introduced in the 1950's, takes into account the relative likelihood of events and uses *apriori* information to improve prediction. The network paradigm also uses Parzen Estimators, which were developed to construct the probability density functions required by Bayes theory.

The PNN uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category or class. The learned patterns can also be combined or weighted with the *apriori* probability, also called the relative frequency, of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class.

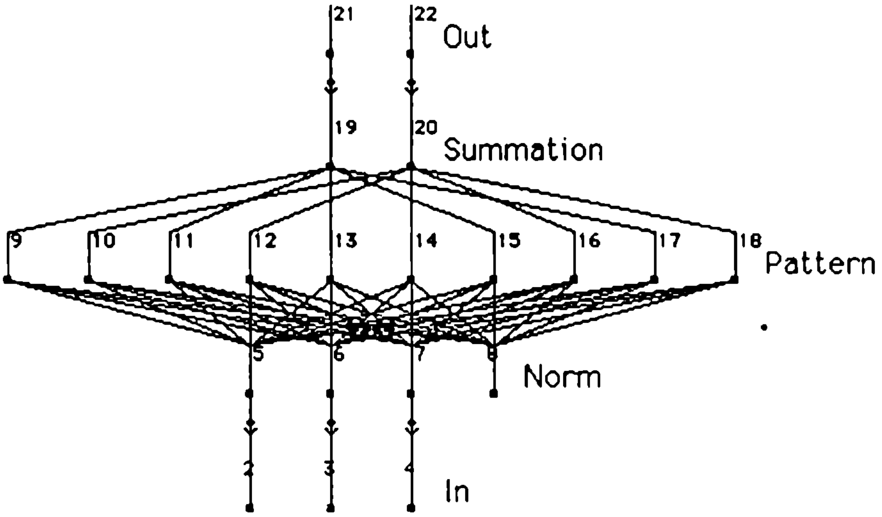


Figure 4.17: A Probablistic Neural Network

An example of a PNN is shown in Figure 4.17. This network has three layers; an input layer, a pattern layer and an output or summation layer. Input layer has as many elements as there are separable parameters needed to describe the objects to be classified. Pattern layer organizes the training set such that each input vector is represented by an individual processing element. The summation layer has as many processing elements as there are classes to be recognized. Each element in this layer combines via processing elements within the pattern layer, which relate to the same class and prepares that category for output. Sometimes a fourth layer is added to normalize the input vector, if the inputs are not already normalized before they enter the network. As with the counter-propagation network, the input vector must be normalized to provide proper object separation in the pattern layer.

As mentioned earlier, the pattern layer represents a neural implementation of a version of a Bayes classifier where the class dependent probability density functions are approximated using a Parzen estimator. This approach provides an optimum pattern classifier that minimizes the expected risk of wrongly classifying an object. With the estimator, the approach gets closer to the true underlying class density functions as the number of training samples increases, so long as the training set is an adequate representation of the class distinctions.

In the pattern layer, there is a processing element for each input vector in the training set. Normally, there are equal amounts of processing elements for each output class. Otherwise, one or more classes may be skewed incorrectly and the network will generate poor results. Each processing element in the pattern layer is trained once. An element is trained to generate a high output value when an input vector matches the training vector. The training function may include a global smoothing factor to better generalize classification results. In any case, the training vectors do not have to be in any special order in the training set since the category of a particular vector is specified by the desired output of the input. The learning function simply selects the first untrained processing element in the correct output class and modifies its weights to match the training vector.

The pattern layer operates competitively where only the highest match to an input vector wins and generates an output. In this way, only one classification category is generated for any given input vector. If the input does not relate well to any patterns programmed into the pattern layer, no output is generated.

The Parzen estimation can be added to the pattern layer to fine tune the classification of objects. This is done by adding the frequency of occurrence for each training pattern built into a processing element. Basically, the probability distribution of occurrence for each example in a class is multiplied into its respective training node. In this way, a more accurate expectation of an object is added to the features that make it recognizable as a class member.

Training of the PNN is much simpler than that of back-propagation. However, the pattern layer can be quite huge if the distinction between categories is varied and at the same time quite similar in special areas. The network structure of PNN is similar to back-propagation; the primary difference is that the sigmoid activation function is replaced by one of a class of functions, which includes the exponential in particular.

Key advantages of PNN are that training requires only a single pass and that the decision surfaces are guaranteed to approach the Bayes optimal decision boundaries as the number of training samples grows. Furthermore the shape of decision curve can be made as complex as necessary or as simple as desired, by choosing the appropriate value for the smoothing parameter.

The key disadvantage of PNN is that all training samples must be stored and used in classifying new patterns. However, because memory is inexpensive and dense, storage will not be a problem. For many classification problems, the speed of a software simulation running on a microprocessor is entirely satisfactory. This type of simulation is several magnitudes faster than back propagation [202], [203]. Furthermore, special-purpose processors and custom chips will allow the PNN paradigm to achieve even higher speeds through the use of parallel hardware.

Chapter 5

Development of Block DWT Computation Algorithm

The advances in technology in the last few decades have made the use of digital images very common in everyday life. While the usefulness of digital images in communication is unquestionable, the overheads associated with storing, processing and transmitting images is very large.

Recently, a great deal of attention has been dedicated to the design of multidimensional filter banks not only for image compression [204], [205], [206] but also for tasks including directional decomposition, image interpolation, linear feature detection and enhancement [207], [208], [209]. Memory is an important constraint in many image processing applications especially for mass-market consumer products such as printers and digital camera where it is highly imperative to maintain low costs. Even if sufficient memory is available, inefficient memory utilization may limit scalability and hinder overall performance. Processing of digitized pathology images is an example. These images have very high resolution making it difficult to display in their entirety on the computer screen and inefficient to transmit over the network for educational purposes. For the computation of DWT of such images the processing memory required is prohibitively high. While wavelet or sub band coding has been proved to be superior to more traditional transforms, the important issue of low memory implementation of the WT has not been given due consideration. Existing DCT is very efficient in its memory utilization because, if needed, it can operate on individual image blocks and the memory requirement is low indeed.

Many of the state-of-the-art algorithms for image processing assume that the WT for the whole image has been computed so that all the corresponding coefficients are available [148], [149]. Global image information, i.e., information that can be obtained only after the whole image has been transformed, is used for classification, selection of specific decomposition levels etc [210]. Algorithms that provide progressive transmission also require the complete set of WT coefficients [148]. All these algorithms typically require buffering the whole image so that the memory usage increases proportional to the image size necessitating prohibitively large on chip memory for filtering operations.

Many algorithms are available for efficient computation of DWT, the most popular one being the Mallat algorithm based on the filter bank tree structure [21]. In real-time applications with memory and delay constraints, this standard algorithm has the following disadvantages:

(1) The requirement of a buffer having the same size as the input to store intermediate results for recursive filtering.

(2) Large latency since all the outputs of one sub band are to be generated before the output of the next sub band.

For efficient multilevel decompositions where memory and delay constraints have to be strictly adhered to, the most direct approach of transform computation is to divide the data into non-overlapping blocks and process each block separately, either sequentially or in parallel. Compression/ decompression applications using DSP/ ASIC chips, space borne instruments etc. require a sequential architecture, because of the availability of a limited amount of memory for transform computation. The parallel architecture will split the input among several processors to speed up the computation. This is used for applications like videoconferencing, digital broadcasting etc., where real time encoding/decoding is involved. However, in both the above cases, due to the non-availability of pixels beyond block boundaries, edge artifacts will be produced near the boundaries in the reconstructed images.

To enable parallel processing without interprocessor data exchange and to reduce memory requirements, a novel algorithm that performs block partitioning on the original data is presented in this chapter. The overlap region between the blocks is used to make the sub band decomposition of each block the same as the corresponding spatial blocks of the full size decomposition. The computational overhead due to overlapping can be reduced to an acceptable level by carefully selecting the block size. During reconstruction, non-overlapping blocks are processed to reduce memory requirements.

By combining block convolution techniques with Mallat algorithm, the proposed scheme converts DWT into a block transform. Parallel processing capability and low memory requirement are the salient features of the algorithm presented here.

5.1 Block –wise Computation of 1-D DWT

The wavelet and scaling transform coefficients X_j and β_j of a signal $x(n)$ at any level j (see fig.4.3) are given by [24]:

$$X_j[n] = \sum_k \beta_{j-1}[k]h[n-2k] \tag{5.1a}$$

and

$$\beta_j[n] = \sum_k \beta_{j-1}[k]g[n-2k] \tag{5.1b}$$

where, $j = 1,2 \ldots J$; $n, k \in Z$, $\beta_0[n]=x[n]$, the input signal and $h[n]$ and $g[n]$ are the analysis high pass and low pass filters respectively. The synthesis equation is:

$$\gamma_j[n] = \gamma_{j+1}[n/2] * \tilde{g}[n] + X_{j+1}[n/2] * \tilde{h}[n] \quad \text{for } j < J \tag{5.2}$$

where, $\gamma_J[n] = \beta_J[n]$ and $\gamma_0[n] = y[n]$, the reconstructed signal which is the same as the input signal x[n]. $\tilde{h}\,[n]$ and $\tilde{g}\,[n]$ are the synthesis high pass and low pass filters respectively. From the above equations, it can be seen that the DWT and IDWT are recursive-filtering processes. This recursiveness poses extra challenges in the implementation of DWT, as it is not a block transform.

Generally, the sequences involved in real time implementations are quasi-infinite. While theoretically we can store the entire data and compute the DWT as a single block for a large number of points, it is too large to compute practically. Also, no processed samples can be obtained until all the input points have been collected. To avoid such delays, the data to be processed must be segmented into smaller blocks or frames.

Consider for example, the above system implemented using two processors with each one allocated with half the input data. For correct computation near the data boundaries each processor would need to access data allocated to the other processor. In this case, either each processor should exchange data before each level of computation or each should be given sufficient overlap of data. The first approach demands frequent data exchange between processors, which will affect the system performance, particularly in the case of slow communication links. The overlap needed in the second approach can be very large due to the recursiveness of the algorithm as the number of levels of decomposition increases.

Direct computation of BDWT using the standard filter bank algorithm results in a sequence longer than the original input sequence. Since the approximation/details at any level are obtained by convolution between the input and filter coefficients, the length of these at the j^{th} level is $l_x/2^j + (1\text{-}2^{-j})\,(l_w\text{-}1)$ where l_x is the length of input sequence and l_w is the length of filter. The total length of transform coefficients at level J is given by:

$$\sum_{j=1}^{J}\{l_x/2^j + (l_w - 1)(1 - 2^{-j})\} + l_x/2^J + (l_w - 1)(1 - 2^{-J})$$

which gets simplified to $l_x+J(l_w-1)$. This clearly indicates that in-place computation is not possible and a separate memory block considerably larger than the input block is necessary for the storage of transform coefficients. For example, consider a speech signal sampled at 8KHz, to be decomposed by a wavelet system having a length of 8, to a depth of 6 levels. If we take blocks of 10ms duration, each of the processed blocks will have a length of 122 samples and 1sec. of processed data will need 12200 locations of storage, whereas the input size is only 8000.

A method for finding the DWT of large sequences on a block-by-block basis is described below. The method is intrinsically similar to the popular block convolution technique of FFT computation [144] except for the problems created by the recursiveness of DWT computation. To meet the challenges made by this, the proposed algorithm is divided into three parts: decomposition using a technique similar to the overlap save method of block convolution followed by truncation to achieve storage of transform coefficients in the same place as the original image and reconstruction using a technique similar to the overlap add method.

5.1.1 Truncation of transform coefficients

For storing the transform coefficients in the same place as the original image, the size of the transform coefficients is to be truncated to the original size after each level of processing. In order to understand how the truncation affects the coefficients, a detailed analysis of the DWT/IDWT pair for 1-D data was performed.

Truncation can be done either symmetrically with respect to the centre point or from one end. During reconstruction also this is needed to maintain the length of the signal. Truncation introduces error in a few samples of the reconstructed signal. If the coefficients are cut symmetrically with respect to the central sample after each level of decomposition and reconstruction, erroneous samples are found at both the ends symmetrically spaced with respect to the central sample. Unsymmetrical truncation can be carried out in two ways:

(1) By removing the additional coefficients from the tail/front end after each level of decomposition and from the front/tail end after each level of reconstruction. In doing so, erroneous terms appear at the tail/front end only. The number of erroneous terms is a function of the length of wavelet l_w and level of decomposition J, given by the relation

$$M=(l_w-2)(2^J-1) \tag{5.3}$$

for any data length, l_x.

(2) By removing the additional coefficients from the tail/front end after each level of decomposition and reconstruction. Number of erroneous terms is more in this case and they appear at both ends.

In the discussions to follow, the methodology adopted is to truncate the WT coefficients from the tail end during decomposition and from the front end during reconstruction, which makes the last M points of the reconstructed signal erroneous [211]. A schematic representation of the above, for a two level DWT computation is shown in figure 5.1.

5.1.2 BDWT by Overlap Save Method

To perform the block-wise decomposition, the principle of overlap save method is incorporated into DWT computation, so as to form an overall transform coefficients sequence, which is identical to that obtained when the signal is processed via conventional DWT. The input data is segmented into blocks of length $N > M$. Each data block overlaps the preceding block by M data points. A J level decomposition of each data block consists of J sets of *details* of length $N/2^j$ where j=1,2,...J, and one approximation component of length $N/2^J$. $M/2^j$ points of the details at the j^{th} level and $M/2^J$ points of the approximation of each transformed block, will be in error due to aliasing, and must be discarded. On removing these erroneous terms, N-M correct transform coefficients are obtained from each input block, except for the first block where there is no aliasing and hence no erroneous terms. This is illustrated in Figure

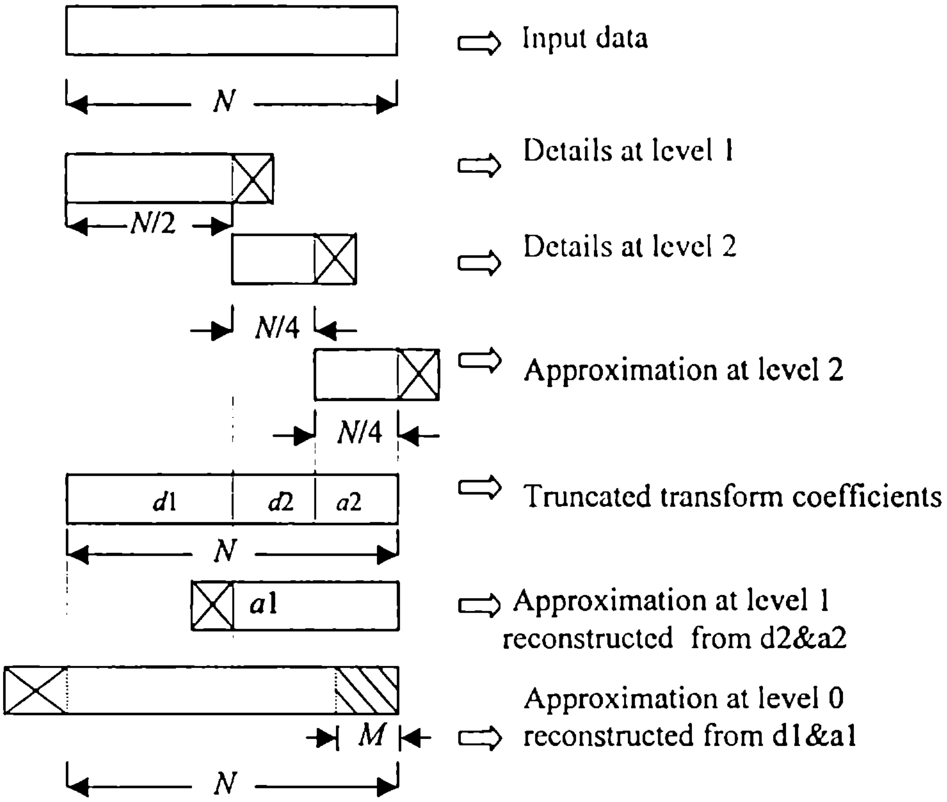


Figure.5.1: Schematic of truncation for the computation of 2-level DWT d1, d2, a2 - details at level 1& 2 and approximation at level 2 respectively after truncation. ('X'indicates truncated portions of transform coefficients and hatched region indicate erroneous portions in the synthesized signal).

5.2. (The approach can be generalized by adding *M* zeroes at the beginning of the input sequence and removing *M* terms corresponding to this from the transform coefficients). The forward WT is implemented without using extra memory for storage of intermediate results. When the transform is complete, the resultant coefficients are the same as those obtained using the pyramidal algorithm, but the coefficients of different sub bands appear in an interleaved manner. The segmentation of the input data stream into smaller blocks and the fitting of the decomposed blocks together to form the overall transform coefficients are graphically illustrated in figure 5.3.

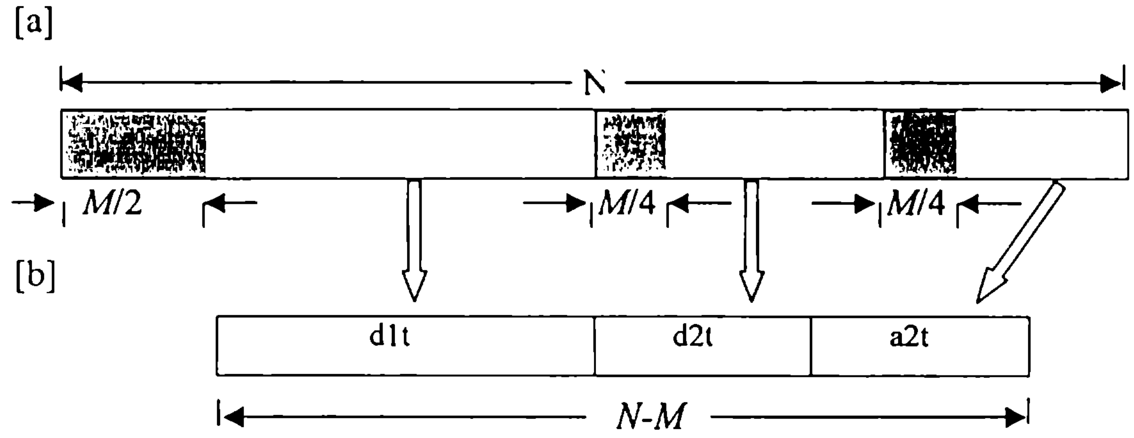


Figure. 5.2 :One block of WT coefficients in 2-level BDWT.
a) Showing erroneous terms due to aliasing as shaded regions b) After removing the erroneous terms d1t, d2t, a2t - details at level 1 & 2 and approximation at level 2 respectively excluding erroneous terms.

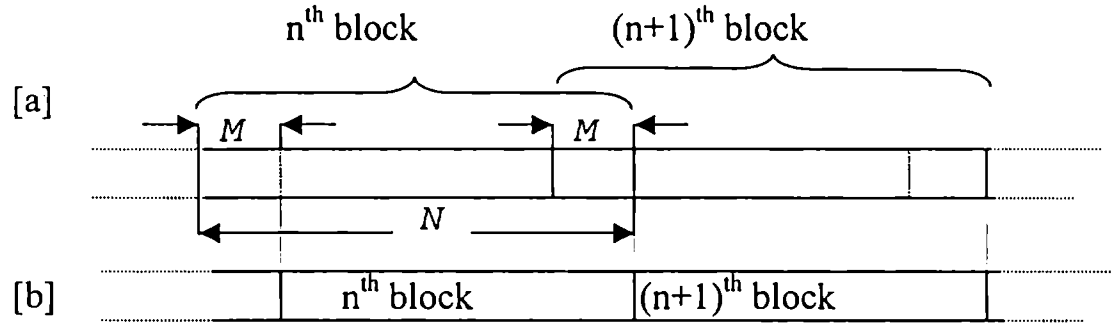


Figure. 5.3: BDWT by overlap save method
a) Input data segmented into overlapping blocks
b) In-place storage of transform coefficients .

5.1.3 Block IDWT (BIDWT) by Overlap Add Method

The reconstruction is also done block-wise. In order to avoid the use of extra buffer memory a method resembling the overlap add method of block convolution is used. Hence, input block for reconstruction is non-overlapping, having a length $N'=N-M$ consisting of J sets of details of length $N'/2^j$, where $j=1,2\ldots J$ and one approximation component of length $N'/2^J$. On reconstruction, blocks of size N with each synthesized block overlapping its succeeding block by M terms are obtained. These overlapping

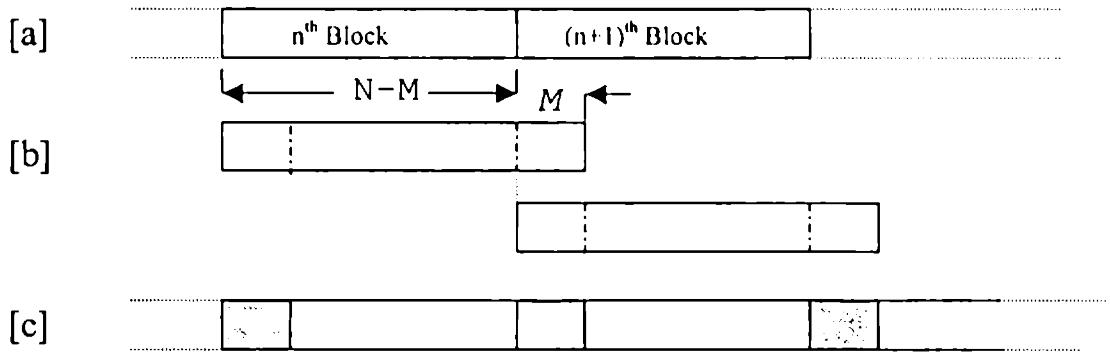


Figure.5.4: BIDWT by overlap add method a) BDWT coefficients. b) Block-wise reconstructed signal c) Final reconstructed signal obtained by adding the overlapping portions of contiguous blocks

terms are added to get a perfect reconstruction of the original signal as shown in figure 5.4.

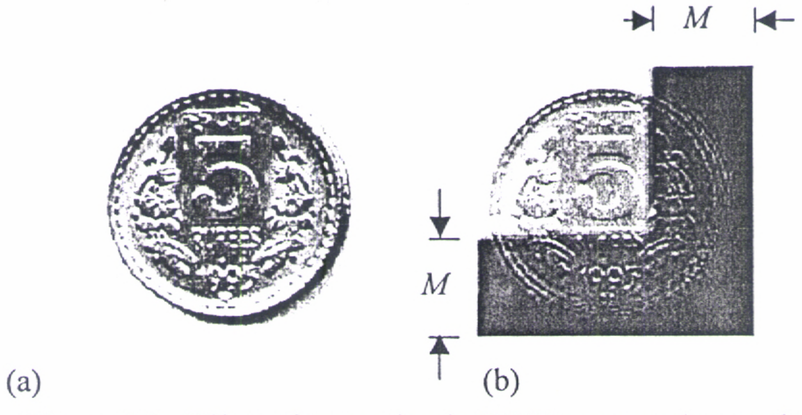
5.2 Block -wise Computation of 2-D DWT

Since the 2-D DWT is separable, it can be computed by cascaded 1-D DWTs i.e. by computing 1-D DWT on rows followed by columns. For computing the DWT of very large images, the images are segmented into smaller frames as in the case of 1-D signals.

As the DWT computation is a convolution operation, the size of the processed blocks will be considerably larger than the size of the input data block. Consider for example, the decomposition of a 256 × 256 pixel image into 4 uniform sub bands using a separable, uniform, two-band filter-bank with 16–tap filters. The overall number of samples in the processed data will increase roughly by 15%[204]. This effect is more pronounced for small data sizes.

One approach to eliminate data expansion is to use circular instead of linear convolution for the implementation of analysis and synthesis filters. However, since circular convolution is equivalent to first periodically extending the finite length sequence and then filtering, artificial high frequency artifacts will be introduced at the block boundaries, making perfect reconstruction impossible [212]. This problem can be corrected using symmetric extension of the signal at the boundaries as long as the filter kernel is symmetric. But for symmetric extension biorthogonal filters rather than orthogonal ones are to be used. The only orthogonal filter that is also symmetric is the Haar filter [173]. By extending the 1-D BDWT technique mentioned above, to the 2-D domain, a new 2-D BDWT algorithm has been developed here which can be applied efficiently with both orthogonal and biorthogonal filters.

Extending the results obtained in the previous section to the 2-D case, it can be seen that, if the transform coefficients are truncated to store them at the same location as the original image, the last M elements of each row and column will be in error in the reconstructed image. This is shown in figure 5.5 for a typical image, which is decomposed into 3 levels using a 16-tap wavelet and then reconstructed after removing the additional coefficients due to convolution.



(a) (b)

Figure.5.5 : Effect of truncation in DWT computation on the image 'coin'.a) Original image. b) Reconstructed image after removing the additional coefficients due to convolution

5.2.1 2-D BDWT by overlap save method

The input data is segmented into overlapping blocks of size $N \times N$, where $N > M$. Each block overlaps with the top neighboring block by M rows and the left neighboring block by M columns. See fig. 5.6 for details. (By adding M rows of zeroes at the top and M columns of zeroes to the left of the image, this approach can be generalized).

A J-level decomposition of each data block consists of J sets of horizontal, vertical and diagonal details each of size $N/2^j \times N/2^j$, where $j=1,2,\ldots J$, and one approximation component of size $N/2^J \times N/2^J$. The first $M/2^j$ rows and columns of the details at the j^{th} level of each transformed block will be in error due to aliasing. Similarly the first $M/2^J$ rows and columns of the approximation will also be in error and all these erroneous terms must be discarded. On removing the above-mentioned erroneous terms, $(N\text{-}M) \times (N\text{-}M)$ correct transform coefficients are obtained from each input block. This can be stored in the same place as the input data, thus making the buffer requirement a minimum. This method of DWT computation resembles the

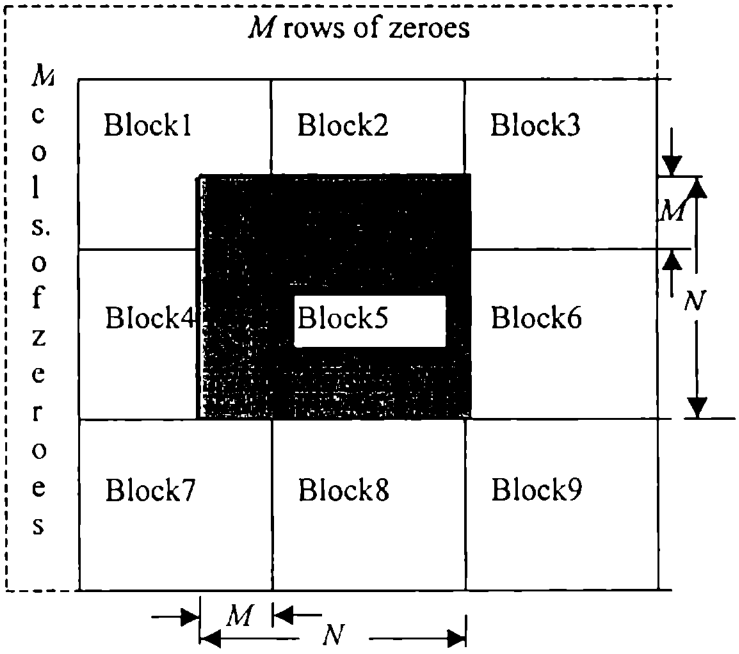


Figure 5.6: Partitioning of an image into 9 overlapping blocks. (The overlap is marked for block 5 only).

overlap save method of block convolution. The distribution of erroneous coefficients for a 2-level decomposition of one data block is illustrated in figure 5.7. The transform coefficients are obviously the same as those obtained for conventional DWT, but the coefficients corresponding to various sub bands appear in an interleaved manner in different blocks.

Figure 5.8 shows the interleaved transform coefficients corresponding to the 2-level block decomposition of a data segmented into 4 blocks. The different levels of detail and approximation coefficients corresponding to the conventional DWT of taking the data as a single block is obtained by concatenating the respective interleaved

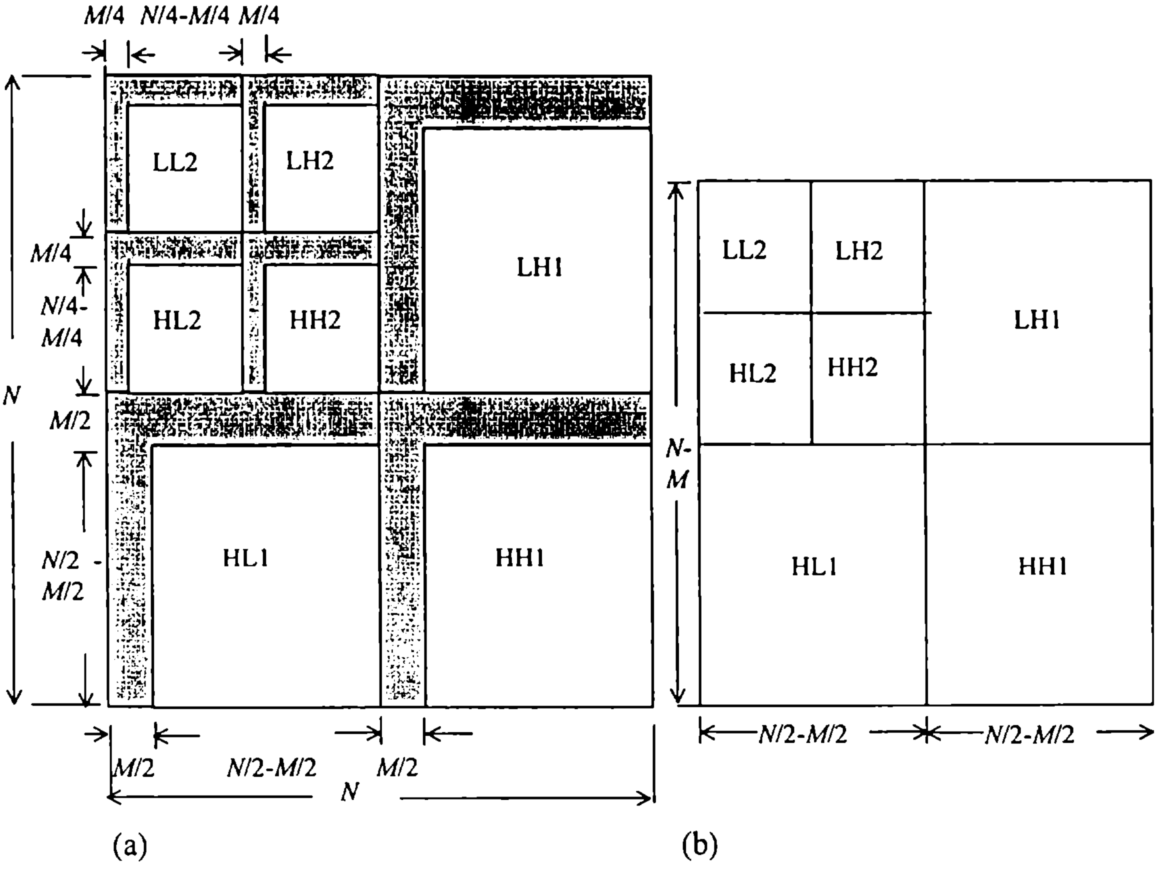


Figure. 5.7: 2- level 2-D BDWT coefficients of a single block of data.
a) Showing erroneous terms due to aliasing as shaded regions.
b) Resultant block of size (*N*-*M*) x (*N*-*M*) after removing the erroneous terms.

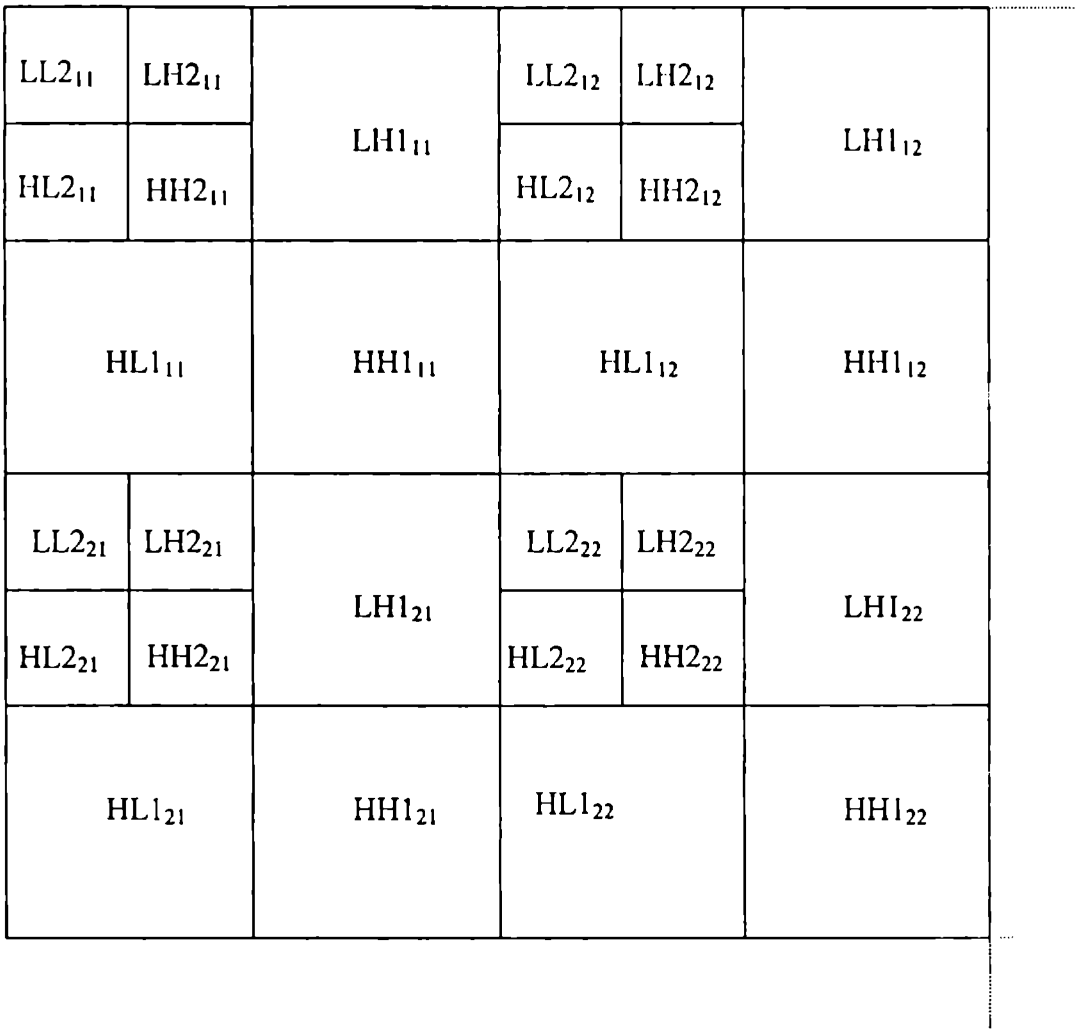


Figure5.8: Distribution of interleaved transform coefficients of BDWT

BDWT coefficients. The second level approximation coefficients LL2of the conventional DWT is obtained by concatenating LL2$_{11}$, LL2$_{12}$, LL2$_{21}$, LL2$_{22}$ i.e. LL2 = { LL2$_{11}$ LL2$_{12}$; LL2$_{21}$ LL2$_{22}$}.Similarly, The second level detail coefficients LH2, HL2, HH2 and the first level detail coefficients LH1, HL1, HH1 of the conventional DWT can be obtained from the BDWT coefficients as LH2 ={LH2$_{11}$ LH2$_{12}$; LH2$_{21}$ LH2$_{22}$}and so on.

5.2.2 2-D BIDWT by overlap add method

To obtain perfect reconstruction of the original image from the transform coefficients, a technique similar to the overlap add method of block convolution is carried out. Processing is done on non-overlapping blocks of transform coefficients of size $N' \times N'$, where $N'=N-M$. Each input block for IDWT consists of J sets of horizontal, vertical and diagonal details, each of size $N'/2^j \times N'/2^j$, $j = 1,2,\ldots J$ and one approximation component of size $N'/2^J \times N'/2^J$. Each synthesized block is of size $N \times N$. The last M rows and columns of each synthesized blocks overlaps with the first M rows of its bottom neighboring block and the first M columns of its right neighboring block as shown in figure 5.9(a). All these overlapping terms are added to get a perfect reconstruction of the original input image. Figure 5.9(b) shows the overlap add operation for one reconstructed block.

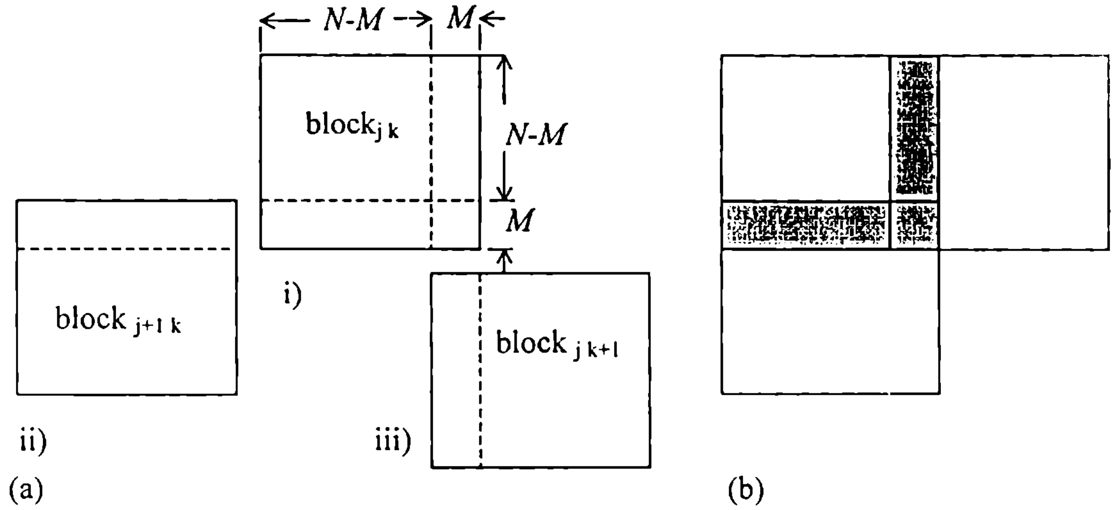


Figure.5.9: Overlap add reconstruction of BDWT coefficients.
a. i)A reconstructed block, block $_{j\,k}$. ii) Its bottom neighboring block, block $_{j+1\,k}$ with top M rows overlapping with block $_{j\,k}$. iii) Its right neighboring block, block $_{j\,k+1}$ with left M columns overlapping with block $_{j\,k}$
b) Resultant image after addition of overlapping terms in the above blocks.

5.3 Computational Complexity

In this section, the computational complexity of the BDWT algorithm is compared with that of the conventional 1-D and 2-D algorithms. The number of real multiplications and additions has been considered as a measure of the computational complexity.

5.3.1 Estimation of computational burden for standard algorithm

For J level decomposition of a signal of length l_x using a wavelet of length l_w, the total computational burden for the pyramidal algorithm is N_{mult} multiplications and N_{add} additions where [213] ,

$$N_{mult} = l_w[4(1-2^{-J})(l_x-l_w+1) + 2J(l_w-1)] \tag{5.4a}$$

$$N_{add} = 2(l_w-1)[J(l_w-2)+2(1-2^{-J})(l_x-l_w+1)] \tag{5.4b}$$

It is shown [24] that for IDWT, the number of arithmetic operations required is exactly the same as that of DWT.

For 2-D DWT computation employing separable filters, each level of computation involves two sets of convolutions as evident from Figure 4.7, one being row-wise on all rows and the other column-wise along all columns. The length of input sequence to each level gets successively modified due to increase in length resulting from convolution at the previous level, decrease in length resulting from downsampling operation and row-column transposition between each horizontal/ vertical convolution.

Consider an image of size $X \times Y$, decomposed to J levels using a wavelet of length l_w. At the j^{th} level of decomposition, the input to each block in which convolution is carried out row-wise, will be of size $X_j^{row} \times Y_j^{row}$, where,

$$X_j^{row} = [X + (2^{j-1} - 1)(l_w - 1)]/2^{j-1} \tag{5.5a}$$

$$Y_j^{row} = [Y + (2^{j-1} - 1)(l_w - 1)]/2^{j-1}, \tag{5.5b}$$

for $j = 1,2,3,\ldots J$. Similarly, for column-wise convolution the input data size at the j^{th} level is $X_j^{col} \times Y_j^{col}$,where ,

$$X_j^{col} = X + (2^{j-1} - 1)(l_w - 1)/2^{j-1} \qquad (5.6a)$$

$$Y_j^{col} = Y + (2^j - 1)(l_w - 1)/2^{j-1} \qquad (5.6b)$$

The decomposition of each level consists of two row-wise convolutions and four column-wise convolutions. Considering all these, the computational burden for a J-level 2-D DWT computation is found to be N_{mult2D} multiplications and N_{add2D} additions where,

$$N_{mult2D} = \sum_{j=1}^{J} (2X_j^{row} Y_j^{row} + 4X_j^{col} Y_j^{col}) l_w \qquad (5.7a)$$

and

$$N_{add2D} = \sum_{j=1}^{J} (2X_j^{row}(Y_j^{row} - 1) + 4(X_j^{col} - 1)Y_j^{col})(l_w - 1) \qquad (5.7b)$$

5.3.2 Estimation of computational burden for BDWT algorithm

Here, since an in-place computation is performed, after each stage of convolution and down sampling, the length of the resulting sequence is adjusted to exactly half the length of the input signal to that stage. Hence, for J-level 1-D BDWT computation using a wavelet of length l_w, N_{mult} and N_{add} given by equations (5.4*a*) & (5.4*b*) gets modified to $l_w(4L(1-2^{-J})-J(l_w-1))$multiplications and $4L(l_w-1)(1-2^{-J})-J(3\ l_w-4)$ additions, for an input block of length L.

Considering the entire data, the total computational burden will be the product of that for a single block and the number of blocks to be computed. Since overlapping blocks are taken during decomposition, the number of blocks is equal to $l_x/(L-M)$, where M is the block overlap length. The minimum value required for M is given by equation (5.3). The overall computational burden is $N_{block\text{-}mult}$ multiplications and $N_{block\text{-}add}$ additions, where

$$N_{block\text{-}mult} = l_w[4\ L\ (1-2^{-J})-J(\ l_w-1)]\ l_x/(L-M) \qquad (5.8a)$$

$$N_{block\text{-}add} = [4\ L(\ l_w-1)\ (1-2^{-J})-J(3\ l_w-4)]\ l_x/(L-M) \qquad (5.8b)$$

For BIDWT the expressions of computational complexity for a block of length L are the same as that of the conventional DWT as we are adopting the overlap add method for reconstruction. This is slightly greater than that for decomposition for a single block of same length. But since the block length for reconstruction is only *L-M*, the computational complexity for a single block is the same for both decomposition and reconstruction. Reconstruction requires an additional number of additions proportional to the overlap length and number of blocks as the overlapping portions of the succeeding reconstructed blocks of the signal are to be added. This makes its total computational complexity slightly greater than that for decomposition.

For 2-D BDWT computation also, after each stage of convolution and down sampling, the size of the resulting 2-D sequence is adjusted to half the size of the input sequence at that stage. Hence, considering a block of initial size $P \times Q$, at the j^{th} level, the input for row-wise convolution will be of size $P_j^{row} \times Q_j^{row}$, where,

$$P_j^{row} = P/2^{j-1} \tag{5.9a}$$

and

$$Q_j^{row} = Q/2^{j-1} \tag{5.9b}$$

$j=1,2,\dots J$.Similarly, for column-wise convolution, the data size is $P_j^{col} \times Q_j^{col}$, where,

$$P_j^{col} = P/2^{j-1} \tag{5.10a}$$

$$Q_j^{col} = Q/2^{j} \tag{5.10b}$$

Hence, the total number of multiplications required for a *J*-level 2-D BDWT decomposition will be $N_{block\text{-}mult2D}$ where,

$$N_{block-mult2D} = \sum_{j=1}^{J} (2(P_j^{row} Q_j^{row} - l_w(l_w - 1)) + 4(P_j^{col} Q_j^{col} - l_w(l_w - 1)))l_w \tag{5.11a}$$

Proceeding in a similar way number of additions required is estimated to be $N_{block\text{-}add2D}$, where,

$$N_{block-add2D} = \sum_{j=1}^{J} (l_w - 1)(2P_j^{row}(Q_j^{row} - 1) + 4Q_j^{col}(P_j^{col} - 1)) - (l_w - 2)(l_w - 3)(P_j^{row} + 2Q_j^{row}) \tag{5.11b}$$

The computational burden for the entire data is obtained by multiplying that of a single block with the number of blocks.

For 2-D BIDWT, the expression for the number of multiplications required for a block of size $P \times Q$ is the same as that for decomposition. Since the overlapping portions of the succeeding blocks are to be added, the number of additions required for reconstruction is increased by a factor proportional to the overlap length and number of blocks.

5.4 Results and Discussion

The BDWT algorithm has been verified for real-time processing of 1-D and 2-D signals. Figure 5.10 shows the illustration of the above algorithm using a segment of a music note. The signal is decomposed to 2 levels using '*db5*' wavelet. The frame length for processing depends on the processing delay tolerable for the particular application and the block overlap length, *M*. Here, a frame length of 256 samples was selected. Figure 5.10(a) shows the input signal under consideration, segmented into three overlapping blocks. Since the signal under consideration contains only 500 samples, the last block is appended with zeroes to make its length 256.The block overlap length as given by equation (5.3) is 24 samples.

The transform coefficients of each of the three blocks are shown in figures 5.10(b), 5.10(c) and 5.10(d). (The coefficients are drawn with level 1 details coming first, followed by level 2 details and level 2 approximation). The shaded regions indicate the erroneous terms due to aliasing in each of the details and approximation. After removing the erroneous terms the length of the transform coefficients is 232 (except for the first block, which is 256 itself as there is no aliasing) and hence they can be stored in the same place as the input.

The overlap add reconstruction of the signal from the BDWT coefficients is shown in figures 5.10(e)-(h). Figure 5.10(e) shows the segmentation of the transform coefficients into non-overlapping blocks, for reconstruction. The 24 samples at the tail end of each of the reconstructed block, overlap with the succeeding block as shown in

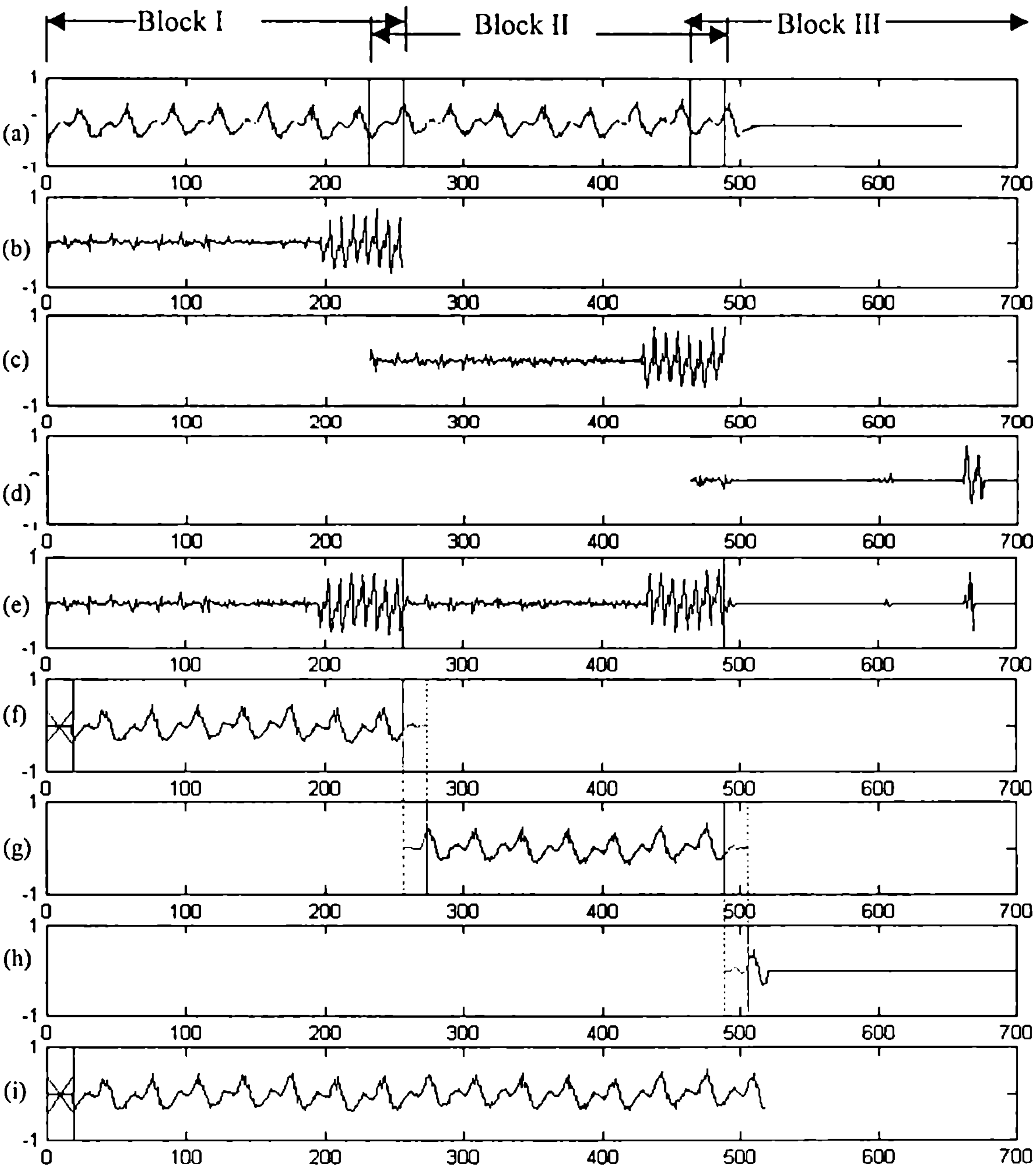


Figure 5.10: 2- level decomposition and reconstruction of a music sample using BDWT technique.(a) Segmentation of input signal (500 samples of a musical note 'D$_4$'in the natural scale by a female voice sampled at 8 kHz, 8-bit resolution. (b),(c),(d) Transform coefficients of individual blocks. (e) Entire transform coefficients obtained by overlap save BDWT. (f),(g),(h) Reconstructed signals from individual blocks. (i) Final reconstructed signal obtained by overlap add BIDWT ('X' indicate additional samples to be removed)

figures 5.10(f)-(h). The final reconstructed signal is obtained by adding all these individual blocks, which is shown in fig.5.10 (i).

Figure 5.11 shows a comparison between the transform coefficients of a speech signal segment computed using the block-wise approach and conventional approach. The signal under consideration is 500 samples taken from a speech signal sampled at 8 kHz with 8-bit resolution (shown in 5.11(a)). A 2-level decomposition is performed on this using '*db2*' wavelet for block lengths of 256. It is seen that the BDWT coefficients in Figure 5.11(b) is the same as the conventional DWT coefficients (shown in figure 5.11(d)) with data interleaving as already mentioned. Also, perfect reconstruction is achieved from the BDWT coefficient as shown in 5.11(c).

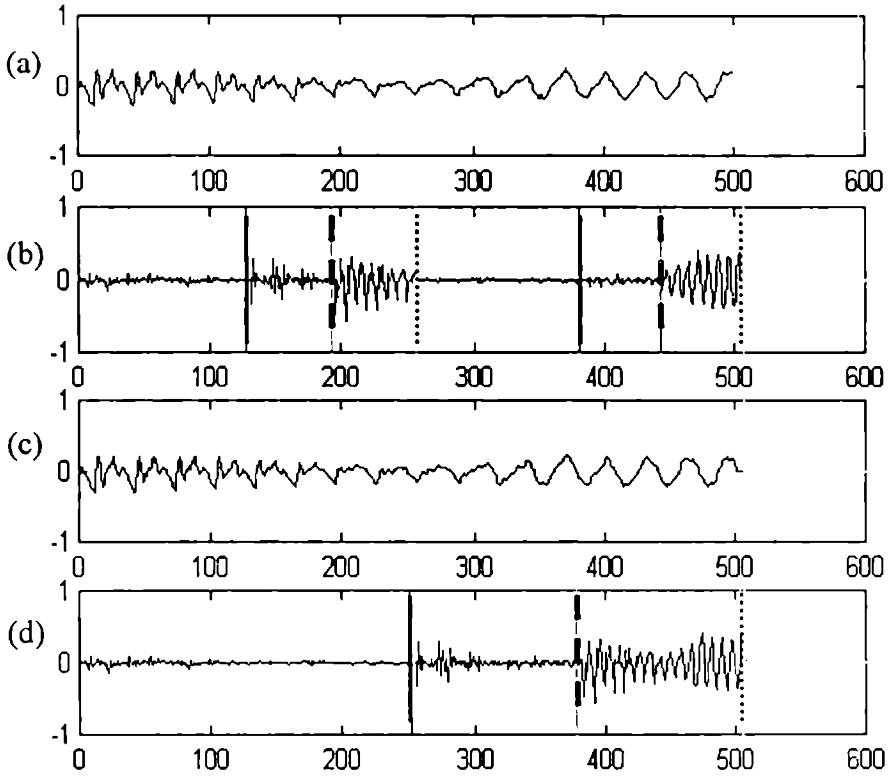


Figure 5.11: Comparison of transform coefficients (2-level decomposition). (a) Original speech signal of 500 samples. (b) Interleaved transform coefficients obtained by BDWT. (c) Reconstructed signal from BDWT coefficients.
(d). Transform coefficients obtained by conventional method.(── lines indicate the end of first level details, ▬ ▬ second level details and ······· approximations)

The algorithm has also been verified using various orthogonal and biorthogonal wavelets, for different levels of decomposition. Figures 5.12 and 5.13 illustrate the verification of BDWT algorithm using an ECG segment and a guitar note .The ECG segment shown in fig 5.12(a) consists of 2000 samples taken from an ECG signal having a sampling rate of 256 Hz with 8-bit resolution. This is decomposed to 4 levels using a biorthogonal wavelet '*bior 4.4*'. A comparatively larger block size of 512 is selected for processing .The interleaved transform coefficients obtained are shown in fig. 5.12(b) and reconstructed signal in fig.5.12(c).

The guitar note shown in fig 5.13(a) contains 600 samples, sampled at 22 kHz. A 3 level decomposition is performed on this using '*db2*' wavelet, selecting small blocks of length 40. Figures 5.13(b) and (c) shows the interleaved BDWT coefficients

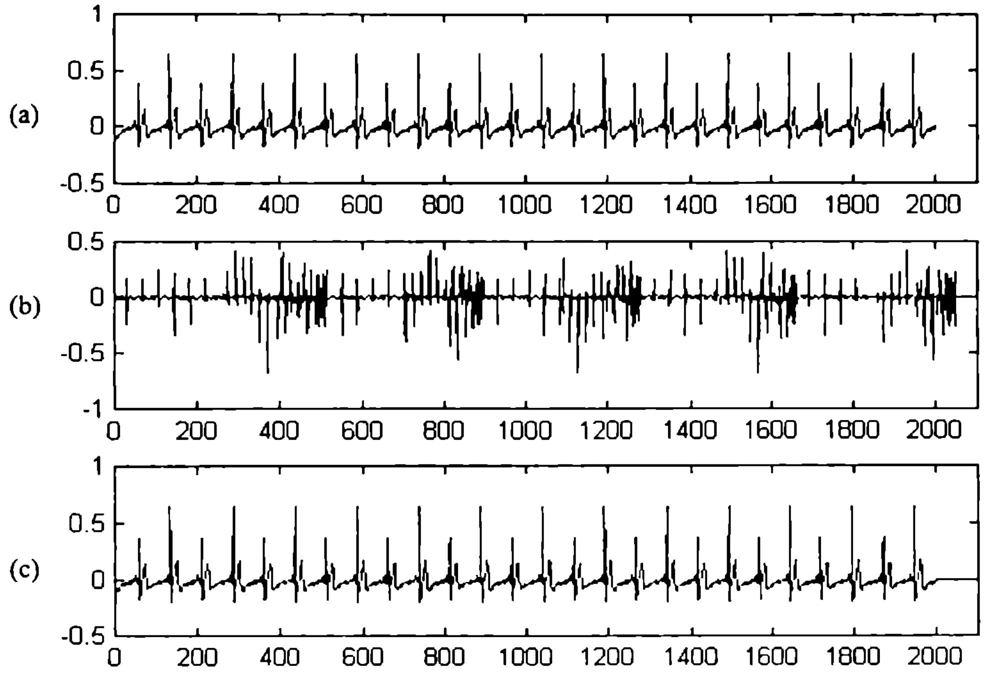


Figure 5.12: Verification of BDWT algorithm using an ECG signal.
a) ECG segment of 2000 samples. (Sampling rate 256Hz).
b) Interleaved BDWT coefficients, block size = 512, 4 levels of decomposition wavelet- '*bior 4.4*'. (c) Reconstructed signal

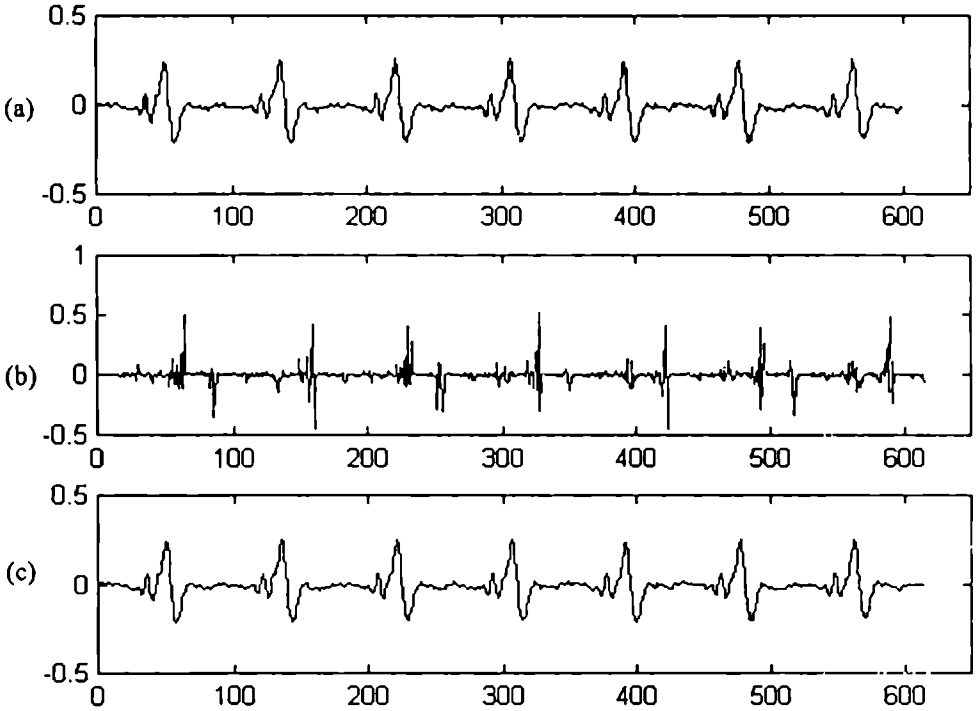
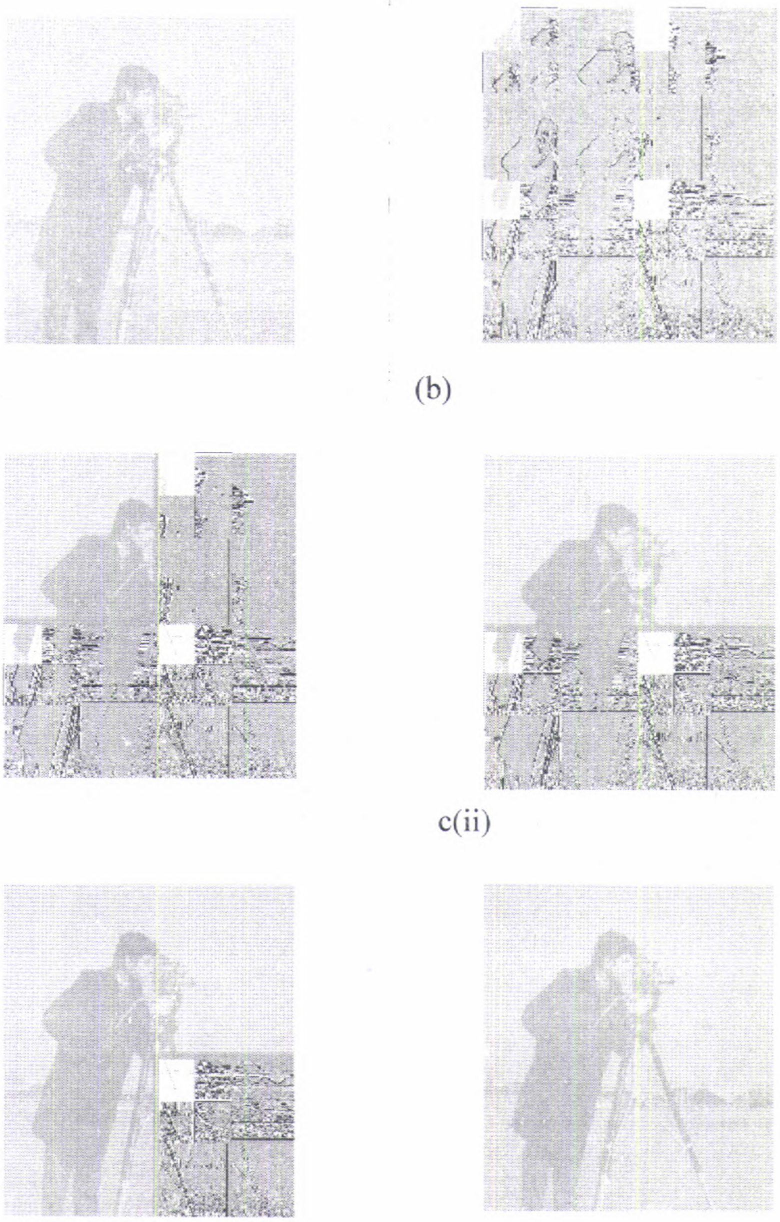


Figure 5.13: Verification of BDWT algorithm using a guitar note.
a) Segment of guitar note sampled at 22 kHz. b) Interleaved BDWT coefficients, block size = 40, wavelet- 'db2' 3 levels of decomposition.
(c) Reconstructed signal

and the signal reconstructed from these coefficients.

Irrespective of the wavelet, level of decomposition, type and size of the signal and block length used, perfect reconstruction of the original signal from the BDWT coefficients has been achieved. The only constraint is that the block length selected should be greater than or equal to the minimum overlap length *M*.

The 2-D BDWT algorithm has been verified using various monochrome still images. Figure 5.14 illustrates this technique on the standard image 'camera man' having 256 x 256 pixels (Fig. 5.14 (a)). Here, for simplicity of illustration, a block size of 140 x 140 is selected, resulting in 4 overlapping blocks with each block overlapping its predecessor on the top and left by 8 rows and 8 columns respectively. On 2-level decomposition employing the '*db2*' wavelet and removal of erroneous terms as



(a) (b)

c(i) c(ii)

c(iii) c(iv)

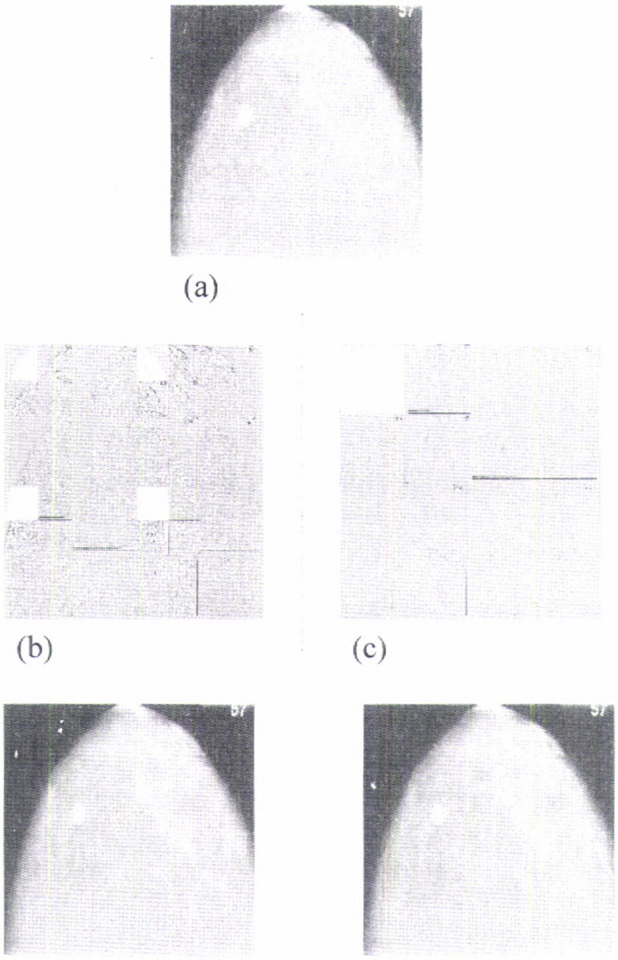
Figure 5.14: 2-level BDWT decomposition and reconstruction of the image ‘camera man’using *db2*. a) Original image. b) Interleaved BDWT coefficients of 4 blocks with each block containing 2nd level approximation of size 32 x 32 & details of size 3 x 32 x32 and 1st level details of size 3 x 64 x 64.
c) (i)-(iv) Block-wise Reconstruction .

described in section 3.3, each of these 4 overlapping blocks produced second level approximation of size 33 x 33, second level details of size 3 x 33 x 33 and first level details of size 3 x 66 x 66.These are stored in an interleaved manner (see fig. 5.14 (b)) in the same place as the original image. The block size for processing depends on the delay tolerable for the particular application, available processing memory and the blockoverlap length *M*. At present, a block of size, which is an integral multiple of 2^J only has been considered.

The overlap add reconstruction of the image from the block DWT coefficients is shown in Figures 5.14 (c (i)-(iv)). Each block of transform coefficients gets reconstructed into an image segment of size 140 x 140, with each segment overlapping with its successor on the right and bottom by 8 columns and 8 rows respectively. By adding these overlapping terms the final reconstructed image is obtained (Figure 5.14 (civ)), which is an exact replica of the original image shown in figure 5.14(a). Figures clearly show that the reconstructed blocks can be stored in the same place as the transform coefficients, reducing the buffer requirement to a mininmum.

Figure 5.15 illustrates the comparison of BDWT and pyramidal algorithms for DWT computation of images. The image under consideration is a biomedical image. The BDWT coefficients shown in 4.15(b) are formed by segmenting the image into 4 blocks. Hence the different subbands appear in an interleaved manner in these 4 blocks. 4.15(c) shows the WT coefficients obtained using the pyramidal algorithm. Figures 4.15 (d) and (e) clearly indicate that perfect reconstruction is obtained using both BDWT and pyramidal algorithms.

Table 5.1 gives the relationship between the ratio of computational complexity of the BDWT algorithm to that of the conventional algorithm for various values of wavelet length, level of decomposition, block size and data size, based on the number of real multiplications. Similar results have been obtained for addition operation also. For reconstruction, the trend remains the same as for decomposition even though the numerical values change slightly for addition operation. From the table it can be noted



(a)

(b) (c)

(d) (e)

Figure 5.15: Comparison of pyramidal and BDWT algorithm for 2-D DWT computation. a) Original image b) BDWT coefficients c) DWT coefficients obtained using pyramidal algorithm d) Image reconstructed from (b) e) Image reconstructed from (c)

that the ratio remains more or less the same for different data sizes. It increases with wavelet length and level of decomposition and decreases with increase in block size. The processing memory requirement of the BDWT algorithm is to the extent of the block size only, whereas the conventional algorithm needs a processing memory of the data size itself. Hence this algorithm is well suited for data –intensive applications with limited processing memory.

Data size	Length of wavelet	Level of Decom-position	Block size			
			500	750	1000	5000
2000	4	2	1.007	1.005	1.003	-
		4	1.05	1.03	1.02	-
		5	1.13	1.08	1.06	-
	10	2	1.036	1.025	1.017	-
		4	1.28	1.18	1.12	
	→	5	1.925	1.45	1.3	-
	16	2	1.06	1.04	1.03	-
		4	1.65	1.34	1.24	-
		5	7.2	2.3	1.7	-
10000	4	2	1.007	1.005	1.004	1
		4	1.05	1.03	1.02	1
		5	1.13	1.08	1.06	1.01
	10	2	1.03	1.025	1.018	1
		4	1.29	1.18	1.12	1.02
		5	1.93	1.45	1.31	1.05
	16	2	1.06	1.04	1.03	1
		4	1.66	1.36	1.24	1.04
		5	7.3	2.3	1.73	1.08

Table 5.1: Ratio of computational complexity (in terms of real multiplications) of the BDWT to that of conventional method.

Figure 5.16 shows the effect of frame size change on computational burden for a data size of 512 x 512 subjected to 3-level decomposition using an 8-tap wavelet. The computational burden is expressed as the ratio of the total number of multiplications involved in processing the data block-wise to that of the entire data as a single block. From the figure we note that, the increase in computational burden is quite negligible for frames of size greater than 32 x 32.

Figure 5.17 shows the reduction in processing delay (normalized with respect to a single processor) that can be achieved in a parallel processing environment. The figure shows that the processing delay for a 3 level decomposition of a 512 x 512 image can be reduced by a factor of 0.11 using 4 processors. The improvement in speed is more as the data size increases.

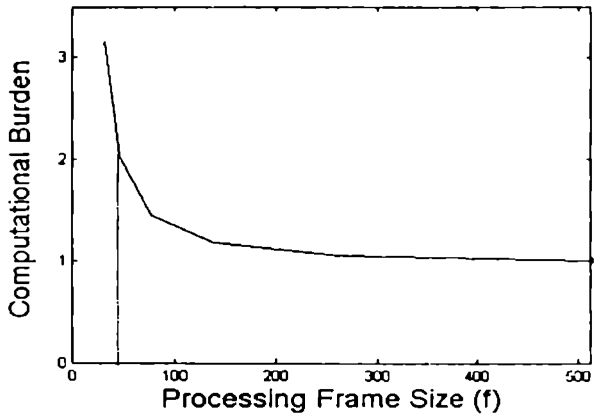


Figure 5.16 : Effect of change in processing frame size (square Frames of size f x f assumed) on Computational Burden (normalized with respect to maximum frame size)

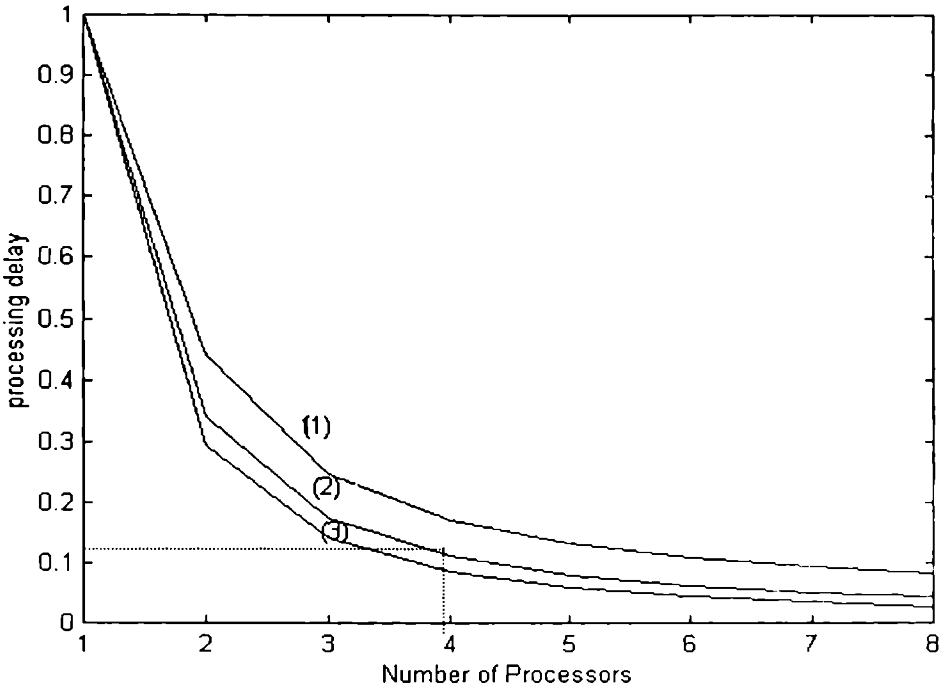


Figure 5.17: Normalized processing delay for multiprocessor computation of BDWT for various data sizes (1) 256x256. (2) 512x512(3) 1024x1024 (3-level decomposition using 8-tap wavelet is carried out).

5.5 Conclusion

Novel efficient algorithms, suitable for real time in-place computation of DWT/IDWT pair of 1-D and 2-D signals, employing the principle of block convolution have been presented. Edge artifacts were removed using the overlap save method for DWT computation. Application of overlap save method for DWT and overlap add method for IDWT computation, reduced the buffer requirement to the order of the size of a block from the size of image required for the present-day algorithms. The minimum block size and hence the buffer requirement is determined by the length of the wavelet and the level of decomposition needed. For any application, appropriate frame sizes can be selected, depending on the available processing memory and tolerable processing delay thereby optimizing the computational burden as well as the processing memory.

The need for interprocessor communication is eliminated since the overlap is taken only once at the input. Hence, this algorithm is apt for both sequential and parallel implementation of real time systems. Since the algorithm is a modification of the popular pyramidal filter bank algorithm, the techniques used for improving the efficiency of computation of the filter bank structure like the lattice structures, FFT techniques for large filters, Running FIR technique for short filters, etc. can be applied here also.

Because of the low memory requirement this algorithm is particularly suitable for processing of images of very large size like digitized pathology images. Hence, this is used in the following sections for WT computation of digitized mammograms.

Chapter 6

Neural Network Based Classification of Mammograms

A great deal of effort has been devoted to CAD in digital mammography to increase diagnostic accuracy as well as the reproducibility of mammographic interpretation. Majority of the work done in this area aims at detecting one or more of the three abnormal structures in mammograms [10]: microcalcifications [32], [77], [90], [95], [99] circumscribed masses [57], [59], [69] and spiculated lesions [73], [76] which often characterize early breast cancer. Others have explored classifying breast lesions as benign or malignant [110], [112]. There has also been work on distinguishing normal regions from regions containing calcifications or masses. However, little work has been done on specifically characterizing normal mammograms [120].

Treatment of cancer is most effective when it is detected early. The ability to reduce the proportion of “misses" in the cancer detection task enhances the chances of survival. Prescreening mammograms to identify the relatively large number of normal mammograms as well as areas of normal tissue in potentially abnormal mammograms will substantially reduce the workload of radiologists and increase the accuracy of their diagnosis in subtle cases.

In this chapter a neural network based classifier is developed for separating normal regions from potentially abnormal regions in mammograms. Removing normal background structures and normal linear markings from the mammograms under consideration, enhances the abnormal features. Statistical as well as structural features are extracted from these and the mammographic regions are classified into normal / abnormal classes with the help of a PNN classifier.

6.1 Normal Mammogram Characterization

The problem of detecting normal mammograms is different from the detection of lesions and is not simply "1 - the detection of abnormal mammograms." Let *P*(*cancer*) be the probability of presence of cancer; *P*(*normal*) be the probability of absence of cancer; *P*(*image*) be the probability of obtaining a specific mammogram; *P*(*cancer/image*) be the probability of cancer detection given a specific mammogram; *P*(*image/cancer*) be the probability of obtaining a specific mammogram given there is cancer and *P*(*image/normal*) be the probability of obtaining a specific mammogram given there is no cancer. In the context of a decision model using Bayes's rule in probability theory we have

$$P(cancer\,/\,image) = \frac{P(image\,/\,cancer)P(cancer)}{P(image)} \qquad (6.1)$$

where

$$P(image) = P(image\,/\,cancer)P(cancer) + P(image\,/\,normal)P(normal) \qquad (6.2)$$

$$P(cancer\,/\,image) = \frac{P(image\,/\,cancer)P(cancer)}{P(image\,/\,cancer)P(cancer) + P(image\,/\,normal)P(normal)} \qquad (6.3)$$

Detection is based on the posterior probability *P* (*cancer/image*), but the decision maybe highly sub-optimal if *P* (*image/normal*) is not known. For example, even if *P* (*image/cancer*) is small, the posterior probability can be large if *P*(*image/normal*) is close to zero. This means that misses can be avoided more easily, if the detection algorithms work synergistically with algorithms for characterizing normal mammograms.

Our approach to the normal mammogram recognition problem is based on normal tissue identification and removal, which is independent of the type of abnormalities that may exist in the mammogram. This approach also facilitates the classification of abnormalities, since suppressing normal background structures enhances the contrast and obviousness of abnormal structures. In addition, the normal tissue characterization problem is fundamentally simpler and easier for computers to solve than the tumor detection problem, because the properties of images of normal tissue are much simpler than the properties of images of abnormalities of various types, sizes, and stages of development. The classification is achieved by presenting the features from the residual image to a neural network classifier.

6.2 Features of Normal Mammograms

Breast tissue composition varies with age and hormone levels in a woman. Generally, a young woman has denser or fibro-glandular breasts, which appear very white or "cloudy" in a mammogram (Figure 3.2). Middle-aged women have a mixture of fibrous and glandular tissues (Figure 3.3). Their mammograms look black and white. In a mature breast, most of the fibrous tissue is replaced with fatty tissue. The mammograms tend to look black or gray (Figure 3.4).

Completely normal mammograms may have entirely different appearances and hence a clear definition of normal mammograms is not easy [1], [2]. As there are no spikes corresponding to microcalcifications and no large bright areas corresponding to masses they have a lower overall density than abnormal ones. Normal regions have linear markings, which are shadows of ducts and connective tissue elements. These are distinct from spiculated or stellate lesions, in which linear markings radiate locally in all directions [120], [75](see fig 3.14). Normal linear markings in mammograms can be considered as straight-line segments of dimensions 1 to 2 mm or greater in length and

0.1 to 1mm in width. Removal of the normal background and linear markings enhances the contrast and obviousness of abnormal structures making their detection easier.

6.3 Residual Image Generation

Normal mammogram characterization is achieved by identifying and removing normal tissue structures. The mammograms obtained after removing the normal structures are called residual images. Accordingly, the residual image of a normal mammogram would be uniformly dark and featureless whereas the residual image of an abnormal mammogram would show microcalcifications or masses against a featureless background.

In this work two types of residual images are generated; one by removing the normal background structures and the other by removing the normal linear markings present in mammograms.

6.3.1 Removal of normal background

Removal of normal background region helps in enhancing any abnormalities present in the ROI under consideration. Here, a WT based average subtraction technique is employed for background removal, as the WT is capable of separating small objects such as microcalcifications from large objects such as large background structures. It was found [214] that the resolution level 1 of the WT showed mainly the high frequency noise included in the mammogram, whereas levels 2 and 3 enhanced microcalcifications effectively. Levels greater than 3 showed a large correlation with the non-uniform background.

To enhance the abnormalities properly, the coarser background should be suppressed, without suppressing the finer abnormal features. This can be done using a smooth mother wavelet. As the wavelet becomes smoother, higher degree of background non-uniformity can be corrected but at the cost of localization property of the wavelet [26]. In this study, we have selected a mother wavelet, which has intermediate length and a high degree of smoothness, to extract subtle and small abnormalities and to suppress background structures effectively.

Figures 6.1 to 6.7 show residual images generated by background-subtraction corresponding to a normal mammogram, and mammograms containing various types of abnormalities like microcalcification, masses, asymmetry and architectural distortion. The images are decomposed to 4 levels using '*bior 6.5*' wavelets and reconstructed after discarding the fourth level approximations. This helps to remove the normal background structures and projects abnormalities, if any.

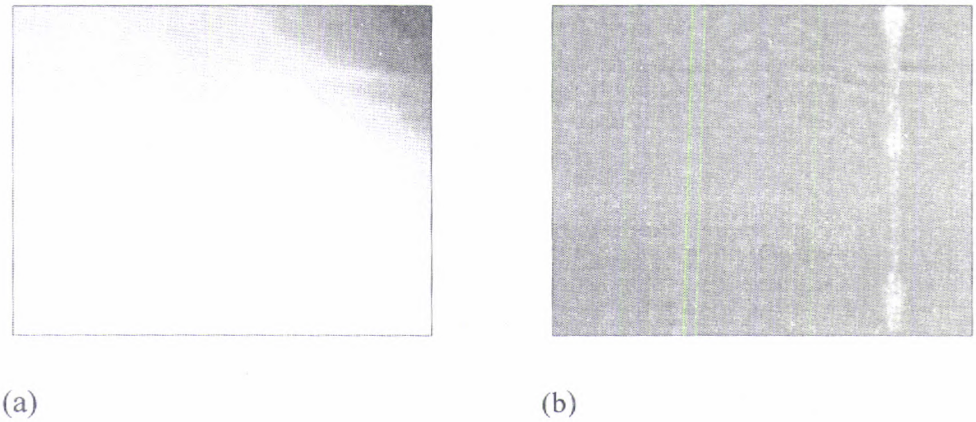

(a) (b)

Figure 6.1: Residual image generation (background removed)
(a) Original normal mammogram b) Residual image

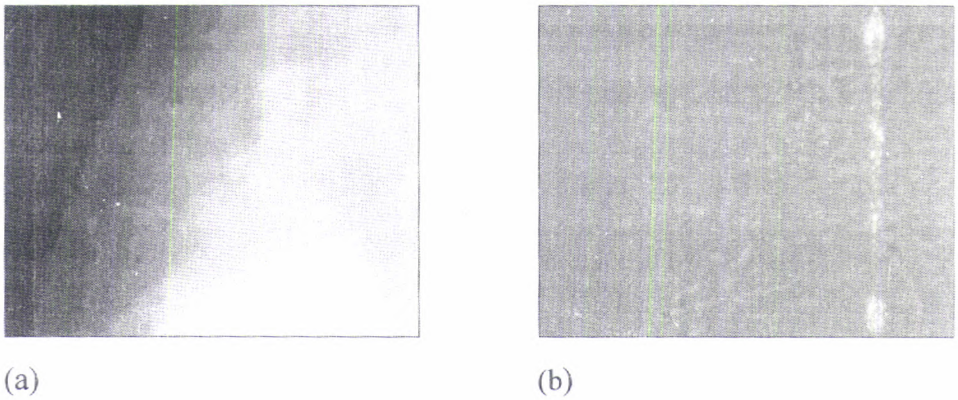
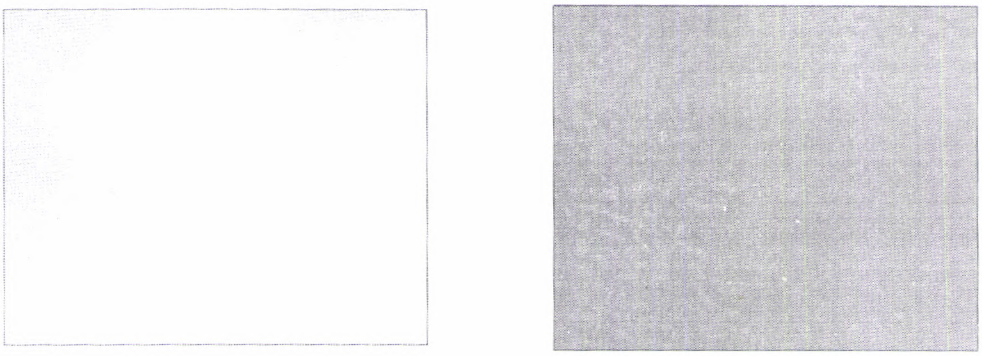
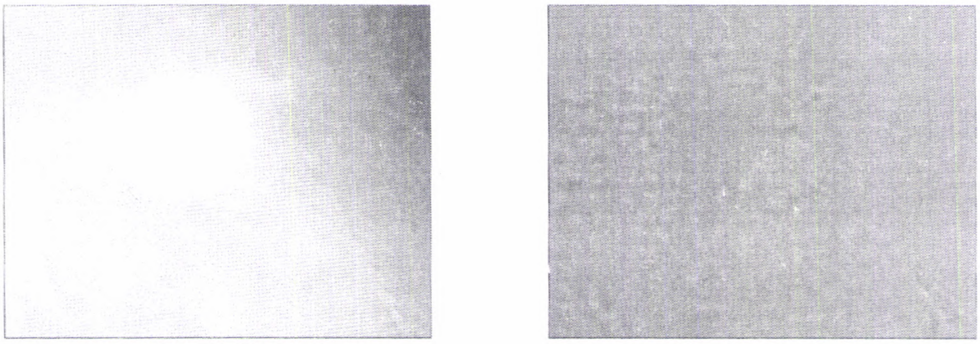

(a) (b)

Figure 6.2: Residual image generation (background removed)
(a) Original image containing microcalcifications b) Residual image



(a) (b)

Figure 6.3: Residual image generation (background removed)
(a) Original image containing a circumscribed mass b) Residual image

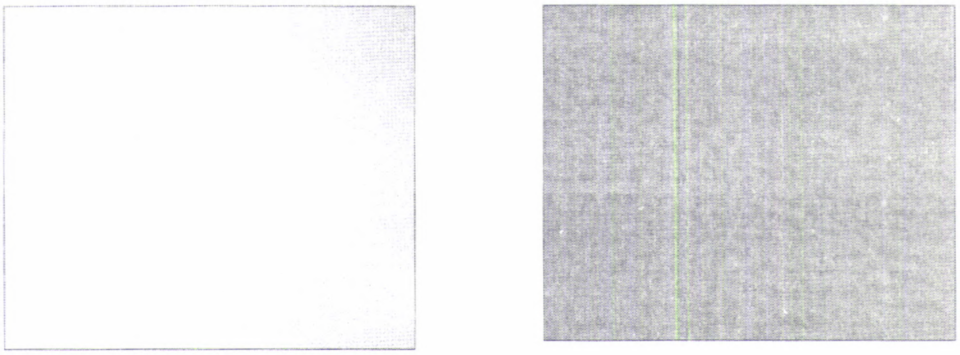


(a) (b)

Figure 6.4: Residual image generation (background removed)
(a) Original image containing a spiculated mass b) Residual image

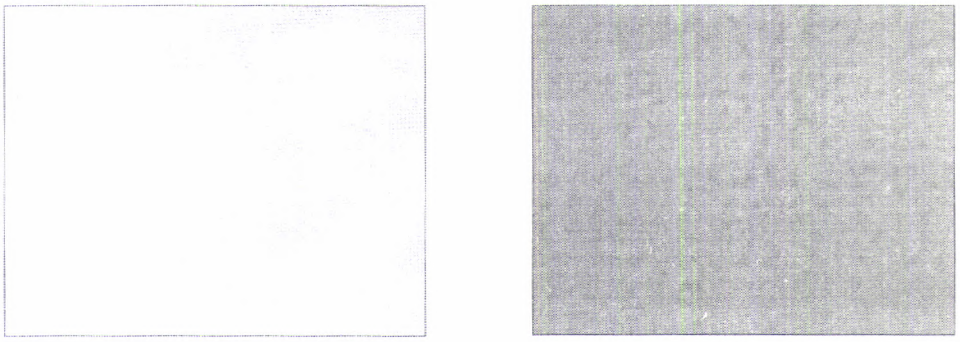
6.3.2 Detection and removal of linear markings

Linear structure detection is a very basic, yet important problem in image processing and computer vision. It is also often the preprocessing step in other applications such as feature extraction, pattern recognition, and image enhancement. Since 1960s, the Hough transform [11], [215], [216] has been widely used for detecting lines in binary images. The Hough transform is fairly robust to noisy or missing data [216], [217] and it can be



(a) (b)

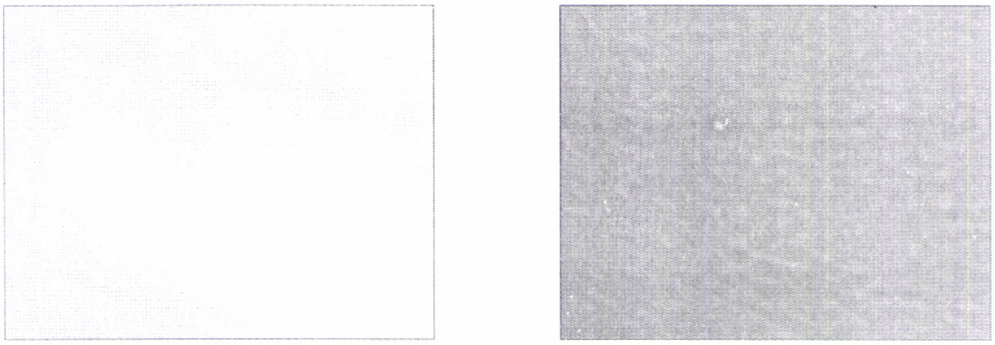
Figure 6.5: Residual image generation (background removed)
(a) Original image containing an asymmetry b) Residual image



(a) (b)

Figure 6.6: Residual image generation (background removed)
(a) Original image containing an architectural distortion b) Residual image

easily extended to detect shapes other than straight lines, such as ellipse [218], [219] and circular objects [220]. However, this requires increased storage and computational complexity [221]. Another drawback of the Hough transform is that it is not suitable for direct use in grayscale images. It requires some preprocessing steps, such as edge detection and thresholding, to make an input grayscale image a binary pattern. In addition, the Hough transform does not provide the actual position of the line in the (x,y) plane [222], but only the angle of the line θ and its distance from the origin ρ. This is not sufficient for many applications.

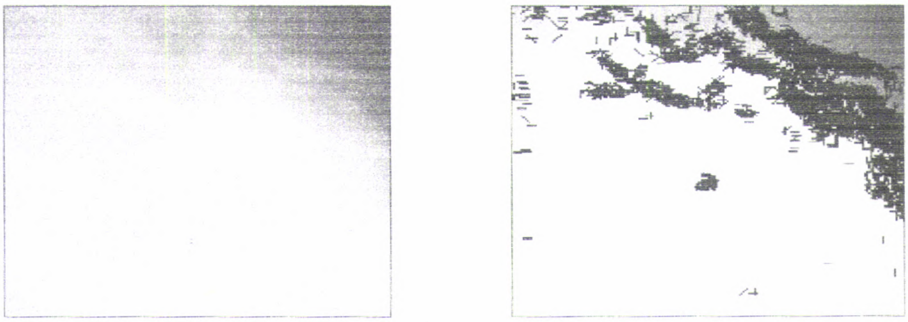


(a) (b)

Figure 6.7: Residual image generation (background removed)
(a) Original image containing an ill-defined mass b) Residual image

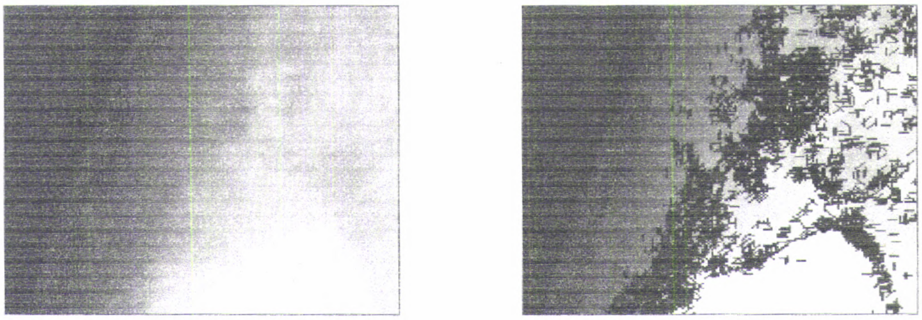
Lines are commonly viewed as extended or contiguous edges. Consequently, many line detection algorithms extract local edges first and then group them into more globally defined lines based on certain criteria [223], [224], [225], [226], [227], [228], [229]. However, local edge operators usually enhance the noise and tend to generate dense edge maps due to their small spatial extent, which makes subsequent processing difficult [230].

Normal linear markings in digitized mammograms vary from 0.1 to 1mm in width. Hence, it is assumed that the basic characteristic of a line in a mammogram, regardless of its thickness, is that pixels on it have similar gray levels. To detect such lines, line detector should be capable of extracting lines with very different and irregular width, as well as curves. Hence the line detector proposed by Liu et al [331] is adopted in this work, which is described in Appendix A. Removal of normal linear markings reduces the possibility of wrong classification of normal mammograms into abnormal class. Residual images generated by removing normal linear markings using the above algorithm for different mammograms are shown in figures 6.8 to 6.11.



(a) (b)

Figure 6.8: Residual image generation (normal linear markings removed)
(a) Original normal mammogram b) Residual image

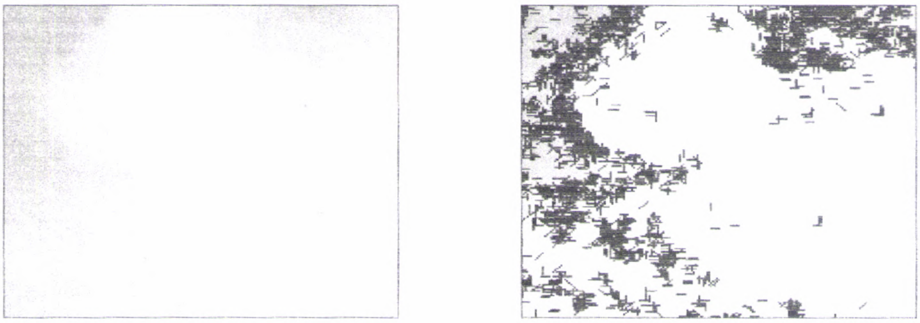


(a) (b)

Figure 6.9: Residual image generation (normal linear markings removed)
(a) Original image containing a microcalcification cluster b) Residual image

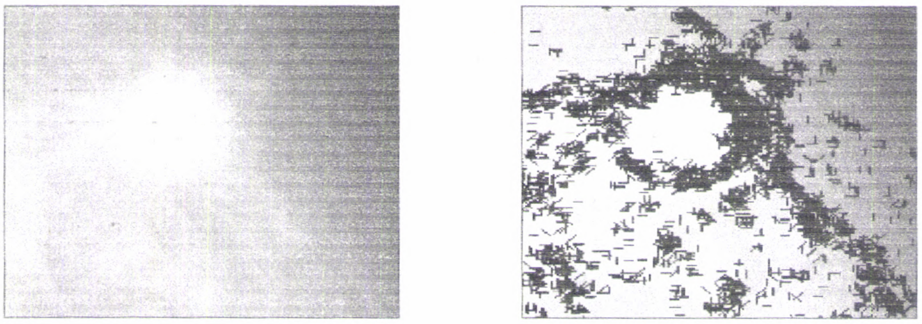
6.4 Neural Network Training and Testing Methodology

Before a neural network model can be used as a pattern classifier, its structure has to be designed and trained. In this section the selection of training and testing data sets and training techniques are discussed.



(a) (b)

Figure 6.10: Residual image generation (normal linear markings removed) (a) Original image containing a circumscribed mass b) Residual image



(a) (b)

Figure 6.11: Residual image generation (normal linear markings removed) (a) Original image containing a speculated lesion b) Residual image

6.4.1 Training and Testing Data Sets

The datasets used were the digitized mammographic data obtained from the freely available database provided by the MIAS [232]. The images in the database are digitized at 50-micron pixel edge, which are then reduced to 200-micron pixel edge and clipped or padded so that every image is having 1024 x 1024 pixels with the image portion centered in the matrix. The origin of the coordinate system is the bottom-left corner. The accompanied 'Ground Truth' contains details regarding the character of the background tissue, class and severity of the abnormality and *x*, *y* coordinate of its center and radii. The database contains 207 normal mammograms and 115 mammograms containing various abnormalities such as calcification, circumscribed masses, spiculated masses, other ill-defined masses, architectural distortion and asymmetry. A detailed description of the data set is given in table 6.1.

Type	Total number	Approximate range of radius (in pixels) of a circle enclosing the abnormality.
Normal	207	-
Calcification	25	3 to 87
Circumscribed masses	23	18 to 197
Spiculated masses	19	17 to 174
Ill-defined masses	14	20 to 123
Architectural distortion	19	23 to 117
Asymmetry	15	23 to 131
Total	322	-

Table 6.1: Training and testing data sets.

The mammograms in this database are arranged in pairs of films, where each pair represents the left (even filename numbers) and right mammograms (odd filename numbers) of a single patient. When calcifications are present, centre locations and radii apply to clusters rather than individual calcifications. In some cases calcifications are widely distributed throughout the image rather than concentrated at a single site. In these cases centre locations and radii are inappropriate and have been omitted.

6.4.2 Detection criteria

The results of classification are expressed in terms of three parameters, True Positive (TP), False Positive (FP) and False Negative (FN). A TP is obtained when a normal/abnormal ROI is correctly classified into normal/abnormal class. When a normal ROI is incorrectly classified as abnormal, it is defined as a FP. A FN is obtained when an abnormal ROI is incorrectly classified into normal class.

6.4.3 *n*-fold Cross Validation

For the purpose of generalization, the cross-validation (CV) technique [233], [234] is used. The cross validation method is a leave-one-out algorithm. For large data sets, this algorithm makes very heavy demands on computing resources. For instance, a Sun Ultra 1 workstation running a PNN simulator required three weeks continual processing to complete a cross validation exercise for the size/structure classification based on a database of around 5000 examples [233]. This is clearly impractical. A compromise solution is to use an *n*-fold cross validation.

The database is divided, with random selection of examples, into *n* partitions (known sometimes as folds) of varying sizes. The literature on *n*-fold cross validation shows little discussion on the selection of *n*, with many researchers using values between 5 and 20 and the majority use $n=10$ [233]. In this study, the partition sizes are varied from 25:75 training/validation to 50:50 training/validation with $n=10$.

6.5 Normal/abnormal classification based on statistical features

6.5.1 Selection of Neural Network structure for classification

The performance of four types of network architectures available in MATLAB, which are popular for classification applications, is compared here. The architectures selected for this study were the BPNN, the RBFN, the PNN and the Competitive network. The four statistical features described in section 4.11.1 *viz.* mean (μ), variance (σ^2), skewness (μ_3) and kurtosis (μ_4) were fed to the input of the networks and the results are tabulated in table 6.2.

Type of network	No.of ROIs	Original image			Residual image with background removed			Residual image with normal lines removed		
		TP	FP	FN	TP	FP	FN	TP	FP	FN
BPNN	Normal 165	17	148	-	45	120	-	91	74	-
	Abnormal 69	41	-	28	37	-	32	10	-	59
	Total 234	58	148	28	82	120	32	101	74	59
RBFN	Normal 165	49	116	-	66	99	-	68	97	-
	Abnormal 69	23	-	46	33	-	36	40	-	29
	Total 234	72	116	46	99	99	36	108	97	29
PNN	Normal 165	**87**	**78**	-	**95**	**70**	-	**86**	**79**	-
	Abnormal 69	**38**	-	**31**	**52**	-	**17**	**52**	-	**17**
	Total 234	**125**	**78**	**31**	**147**	**70**	**17**	**138**	**79**	**17**
Competitive layer	Normal 165	2	163	-	66	99	-	113	52	-
	Abnormal 69	45	-	24	12	-	57	14	-	55
	Total 234	47	163	24	78	99	57	127	52	55

Table 6.2:Comparison of performance of different network architectures on normal/abnormal classification of mammographic data using statistical features.

From the MIAS database 345 ROIs of size 256 x 256 were selected for this study. The selected ROIs include 221 normal ROIs and 124 ROIs containing various abnormalities. 111 ROIs (55 abnormal ones and 56 normal ones) from the above are used for training the networks and the performance of the networks are tested using the remaining 234 ROIs (165 normal and 69 abnormal). The table clearly shows that best results are obtained for the PNN architecture. Hence, PNN is selected for the classification of the mammograms into normal and abnormal groups.

6.5.2 Feature selection

A good feature must be sufficiently discriminating. However, in order to keep the classification problem tractable, the total number of features selected must be limited. Neural network technology offers techniques for selecting, developing, clustering and compressing features into a useful set.

In this thesis, the problem of feature selection is attacked using a "feature wrapper" approach. The guiding principle of this approach is that the features that can best be used for classification should be chosen. A consequence of this principle is that one must know exactly how the samples will be classified before feature selection can be done. A major advantage of the feature wrapper approach is accuracy, because the feature selection is "tuned" for the classification method. Another advantage is that the approach provides some protection against over fitting because of the internal cross validation employed by the jackknife approach. One drawback of feature wrapper method is that the method can be computationally intensive.

The process of feature selection is described below [235]:

(1) A candidate set of features is considered.
    a. The mammograms are divided into training and test sets.
        i. The classifier is trained on the training set of samples.
        ii. The classifier is used on the test set of samples.
    b. Step 1(a) is repeated with alternative divisions into training and test sets.
    c. The candidate feature set is evaluated using all classifications from 1(a) (i)-(ii).

(2) Step 1 is repeated with another candidate feature set.

The result of the feature wrapper approach for the statistical features μ, σ^2, μ_3 and μ_4 derived from the original ROIs using PNN is tabulated in table 6.3. The PNN is trained and tested using the training and test data described in section 6.5.1. The table shows that very low FP values are obtained for the feature kurtosis alone, but at the cost of a FN rate of 100%. This classification is intended as a pre-processing step for automated detection of breast cancer. Even though a slightly larger FP rate can be tolerated, the FN rate should be as small as possible. Hence, best results are obtained for feature sets mean and variance, mean, skewness and kurtosis and all 4 together. But the classification result obtained here is far from the requirement. So the procedure is repeated on residual images obtained by subtracting the background and removing the normal linear markings. The results are tabulated in table 6.4 and 6.5.

	Selected feature														
	μ	σ^2	μ_3	μ_4	μ & σ^2	μ & μ_3	μ & μ_4	σ^2 & μ_3	σ^2 & μ_4	μ_3 & μ_4	μ, σ^2 & μ_3	μ, σ^2 & μ_4	μ, μ_3 & μ_4	σ^2 μ_3 & μ_4	μ, σ^2 μ_3& μ_4
FP	95	76	102	2	106	78	95	120	2	103	93	95	78	103	78
FN	38	68	49	69	36	31	36	30	69	30	30	36	31	30	31

Table 6.3: Sensitivity for different features derived from the original image

	Selected feature														
	μ	σ^2	μ_3	μ_4	μ & σ^2	μ & μ_3	μ & μ_4	σ^2 & μ_3	σ^2 & μ_4	μ_3 & μ_4	μ, σ^2 & μ_3	μ, σ^2 & μ_4	μ, μ_3 & μ_4	σ^2 μ_3 & μ_4	μ, σ^2 μ_3& μ_4
FP	121	74	60	4	153	69	123	113	2	64	106	92	71	62	70
FN	45	42	63	69	17	36	28	25	69	35	11	28	17	35	17

Table 6.4: Sensitivity for different features derived from the residual image (background removed)

For residual images generated by background subtraction, lowest FN rate is produced by feature set mean, variance and skewness. But the FP rate of this feature set is unacceptably high. The feature sets mean, skewness and kurtosis and all 4 together provide acceptable FP and FN rates. Since lesser number of features make the computation easier, it is better to select the features mean, skewness and kurtosis for classification purpose, even though its FP rate is slightly higher compared to the result obtained by using all 4 features. The results obtained here are better than those obtained on features extracted from the original image. This is because of the enhancement occurred to the abnormalities due to the removal of normal background tissue from the images. Residual images formed by removing normal linear markings provide best results when the parameters mean and skewness are used (see table 6.5). Based on these observations three features mean, skewness and kurtosis are selected for classification of mammograms into normal and abnormal classes.

	Selected feature														
	μ	σ^2	μ_3	μ_4	μ & σ^2	μ & μ_3	μ& μ_4	σ^2 & μ_3	σ^2 & μ_4	μ_3 & μ_4	μ, σ^2 & μ_3	μ, σ^2 &μ_4	μ, μ_3& μ_4	σ^2, μ_3 &μ_4	μ, σ^2, μ_3& μ_4
FP	80	66	92	1	85	79	80	99	1	92	84	80	79	92	79
FN	26	68	47	69	25	17	25	32	69	32	17	25	17	32	17

Table 6.5: Sensitivity for different features derived from the residual image (normal lines removed)

6.5.3 Classification results

3 sets of the feature vectors containing three elements each, the mean, skewness and kurtosis are extracted from each snippet of mammogram of size 256 x 256 and presented to 3 PNNs. The first set of feature vectors is extracted from the original mammograms, second from residual images obtained after the removal of normal background and third from residual images obtained after detection and removal of normal linear markings. The input layer of the PNN handles the features extracted from each ROI. As mentioned earlier, two output units denote the presence or absence of an abnormal tissue.

The results for the three sets of features are tabulated in table 6.6. 100% detection accuracy was obtained for the training set. Using the features from the original image a sensitivity of 52.7% for normal mammograms and 55% for abnormal mammograms were obtained for the test data. A sensitivity of 57% for normal mammograms was obtained for the background-removed case and 52% for residual images with linear markings removed. Both types of residual images produced a sensitivity of 75.4 % for the abnormal case. But none of these is sufficient for the first stage of an automated breast cancer detection system. Hence a two-step classification, as detailed below, was developed for increasing the sensitivity and specificity of detection.

	No.of ROIs	Original image			Residual image with background removed			Residual image with normal lines removed		
		TP	FP	FN	TP	FP	FN	TP	FP	FN
Normal	165	87	78	-	94	71	-	86	79	-
Abnormal	69	38	-	31	52	-	17	52	-	17
Total	234	125	78	31	146	71	17	131	86	17

Table 6.6: Classification results for the 3 sets of feature vectors

Step1: Classify the given mammogram into normal or abnormal groups using the features derived from the original mammograms.

Step2: If the ROI is found to be normal, project any abnormality that can be present in it by removing the background and again classify.

Else if it is abnormal, remove the normal linear markings that may be misunderstood as abnormalities and again classify.

The result of this two-step classification is given in table 6.7. The detection sensitivity of the abnormal cases is 91% and that for the normal cases is only 58%. The low sensitivity for the normal cases will not be much of a problem as this is intended as a pre-processing stage only. A detailed classification result on the MIAS database is given in table 6.8.

	No.of ROIs	TP	FP	FN	%Detection
Normal	167	97	70	-	58.08
Abnormal	69	63	-	6	91.3
Total	236	160	70	6	67.8

Table 6.7. Results of Statistical feature based classification

			No.of ROIs	TP	FP	FN	%Detection
Normal		Training set	56	56	-	-	100
		Test set	167	97	70	-	58.08
Abnormal	Training set	Calcification	14	14	-	-	100
		circumscribed masses	12	12	-	-	100
		Spiculated masses	13	13	-	-	100
		ill-defined masses	10	10	-	-	100
		Architectural distortion	3	3	-	-	100
		Asymmetry	4	4	-	-	100
	Test set	Calcification	20	20	-	-	100
		circumscribed masses	11	10		1	90.1
		Spiculated masses	6	6	-	-	100
		ill-defined masses	4	3	-	1	75
		Architectural distortion	16	14	-	2	87.5
		Asymmetry	11	9	-	2	81.8
Total			345	269	70	6	78

Table 6.8. Detailed Results of Statistical feature based classification

6.6 Normal/abnormal classification based on textural features

6.6.1 Selection of Neural Network structure for classification

The performance of the same four types of networks as in section 6.5.1 is estimated here to select the best network for classification based on textural features. Various texture features described in section 4.11.2 were used for the evaluation and the results are tabulated in table 6.9. From the MIAS database, 332 ROIs of size 256 x 256 including 222 normal ones and 110 ROIs containing various abnormalities like microcalcifications, masses and architectural distortions are selected for this study.

Type of network	No.of ROIs	Original image			Residual image with background removed			Residual image with normal lines removed		
		TP	FP	FN	TP	FP	FN	TP	FP	FN
BPNN	Normal 168	19	149	-	47	121	-	94	74	-
	Abnormal 55	7	-	48	13	-	42	9	-	46
	Total 223	26	149	48	60	121	42	103	74	46
RBFN	Normal 168	49	119	-	66	102	-	71	97	-
	Abnormal 55	9	-	46	19	-	36	26	-	29
	Total 223	58	119	46	85	102	36	97	97	29
PNN	**Normal 168**	**151**	**17**	-	**92**	**76**	-	**100**	**68**	-
	Abnormal 55	**6**	-	**49**	**25**	-	**30**	**52**	-	**3**
	Total 223	**157**	**17**	**49**	**117**	**76**	**30**	**152**	**68**	**3**
Compet-itive layer	Normal 168	110	58	-	87	81	-	129	39	-
	Abnormal 55	35	-	20	10	-	45	12	-	43
	Total 223	145	58	20	97	81	45	141	39	43

Table 6.9: Comparison of performance of different network architectures on normal/abnormal classification of mammographic data using textural features.

109 (55 abnormal ones and the rest normal) ROIs from these were used to train the networks. Mammograms containing the abnormality, asymmetry is not included in this study.

6.6.2 Feature selection

Ideally, the feature wrapper approach requires all possible candidate feature sets to be considered. But this is difficult even for modest number of candidate features. Hence, clever strategies are required to search through the space of feature sets. Xiong et al. evaluated two relatively simple search procedures, Sequential Forward Search (SFS) and Sequential Forward Floating Search (SFFS) [236]. The sequential forward search procedure is adopted here. Different steps involved in this are:

(1) Choose the single best feature
(2) Choose the best feature set of size two that includes the feature from (1)
(3) Choose the best feature set of size three that includes the feature set from (2)
(4) And so on.

The textural features used for classification purpose are the Haralick's texture features described in section 4.11.2. The results of various steps of SFS for feature selection on the test data set of 168 normal mammograms and 55 abnormal ones described in the previous section are tabulated below. Using this approach a feature set of 4 elements, which can best be used for classification is selected from the 10 features. The selected features are angular second moment or energy, entropy, correlation, and contrast.

Table 6.10 shows that the best single feature for classification is correlation. The features that were not able to classify even the elements of the training set into their correct group were eliminated at this stage. The best result for a feature set size of 2 is obtained for the features correlation and entropy. Hence these two were selected as the best feature set of size 2 (see table 6.11). From table 6.12 the best feature set of size 3 is obtained by adding energy to that obtained in step 2. Following this procedure it can be seen that the best set of features of size 4 is correlation, entropy, energy and contrast (see table 6.13) and that for a 5 vector feature set includes local homogeneity in addition to the above 4. But when the 5-vector feature set is used, there is no improvement in performance over the 4-feature set. Hence, it is better to use the 4- feature set to reduce computational burden.

	energy		entropy		**correlation**		contrast		Sum of squares variance		Local Homogenity	
	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN
Original	87	40	78	39	**61**	**40**	0	55	8	55	56	55
Back-ground subtracted	87	48	106	31	**94**	**43**	0	55	4	55	44	55
Normal lines removed	165	19	113	34	**88**	**33**	0	55	105	35	45	55

Table 6.10 Selection of the single best feature using SFS

	correlation & energy		**correlation & entropy**		Correlation & contrast		Correlation & Local Homogenity		Correlation & Sum of squares variance	
	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN
Original	28	55	**34**	**52**	61	40	102	31	66	38
Background subtracted	90	40	**77**	**34**	106	27	125	23	107	27
Normal lines removed	85	32	**72**	**24**	88	27	107	30	139	11

Table 6.11 Selection of the best feature set of size two using SFS

	correlation entropy &energy		correlation entropy & contrast		correlation entropy & Local Homogenity		Correlation Entropy & Sum of squares variance	
	FP	FN	FP	FN	FP	FN	FP	FN
Original	**17**	**52**	34	50	39	42	36	56
Background subtracted	**75**	**38**	97	34	101	27	80	38
Normal lines removed	**69**	**6**	92	7	95	7	134	10

Table 6.12.Selection of the best feature set of size three using SFS

	correlation entropy energy & contrast		correlation entropy energy & Local Homogenity		correlation entropy energy & Sum of squares variance	
	FP	FN	FP	FN	FP	FN
Original	17	49	68	50	19	55
Background subtracted	76	30	99	31	78	38
Normal lines removed	68	3	93	4	130	7

Table 6.13: Selection of the best feature set of size 4 using SFS

	correlation entropy energy contrast & Local Homogenity		correlation entropy energy contrast & Sum of squares variance	
	FP	FN	FP	FN
Original	17	49	17	49
Background subtracted	76	30	76	30
Normal lines removed	68	3	68	3

Table 6. 14: Selection of the best feature set of size 5 using SFS

6.6.3 Classification results

Since the image matrix is discrete, the displacement vector used in the feature calculation was chosen to have the following phase and displacement values: $(0^0,1)$, $(45^0, 1)$, $(90^0,1)$, $(135^0 ,1)$. The input layer of the PNN handles the four features extracted from each ROI.

The three sets of features extracted from each ROI corresponding to the original as well as the two residual images were tested using the PNN and the results are tabulated in table 6.15. Though the false positive rate using the features from the original mammograms is less than 10%, the false negative rate is very high (94%), which cannot be tolerated. Best detection accuracy is got with the features obtained after removing normal lines. A recognition accuracy of 60% was obtained for the normal mammograms and 94.2% for the abnormal cases. For the entire data set an overall TP recognition score of 68.2% was obtained. The recognition score with features from the background-subtracted images is less. This is due to the fact that the texture information is lost when the normal background is subtracted from the images.

	No of mammograms	Original			With background subtracted			With normal lines removed		
		TP	FP	FN	TP	FP	FN	TP	FP	FN
Normal	168	151	17	-	92	76	-	100	68	-
Abnormal	55	6	-	49	25	-	30	52	-	3
Total	223	157	17	49	117	76	30	152	68	3

Table 6.15 Classification result using textural features

Table 6.16 compares the performance of the algorithm on the features obtained after removal of normal lines from the ROIs for different orientations. Best performance is attained for an orientation of 0^0. A detailed classification result is provided in table 6.17.

	No of mammogram	$\theta=0^0$			$\theta=45^0$			$\theta=90^0$			$\theta=135^0$		
		TP	FP	FN	TP	FP	FN	TP	FP	FN	TP	FP	FN
Normal	168	100	68	-	92	76	-	90	78	-	91	77	-
Abnormal	55	52	-	3	42		13	45	-	10	43	-	12
Total	223	152	68	3	134	76	13	135	78	10	134	77	

Table 6.16 Classification result using textural features for different orientations

			No.of ROIs	TP	FP	FN	%Detection
Normal		Training set	54	54	-	-	100
		Test set	168	100	68	-	59.5
Abnormal	Training set	Calcification	23	23	-	-	100
		circumscribed masses	9	9	-	-	100
		Spiculated masses	9	9	-	-	100
		ill-defined masses	6	6	-	-	100
		Architectural distortion	8	8	-	-	100
	Test set	Calcification	12	12	-	-	100
		circumscribed masses	14	14	-	-	100
		Spiculated masses	10	8	-	2	80
		ill-defined masses	8	-	-	-	100
		Architectural distortion	11	10	-	1	90.9
Total			332	261	68	3	78.6

Table 6.17:Detailed result of Classification for an orientation of 0^0

6.7 Normal/abnormal classification based on both statistical and textural features

The classification task is repeated using a combined feature set from the above sections. Residual images are formed by removing the normal linear markings. The 3 statistical features *viz.* mean, skewness and kurtosis and the four texture features *viz.* correlation entropy, energy and contrast described in previous two sections, are derived from the residual image and fed to the PNN for classification. 111 ROIs (55 abnormal ones and 56 normal ones) from the MIAS database are used for training the PNN and its performance is tested using 236 ROIs (167 normal and 69 abnormal). The results are tabulated below in tables 6.18 and 6.19.

	No.of ROIs	TP	FP	FN	%Detection
Normal	167	101	66	-	60.5
Abnormal	69	62	-	7	89.86
Total	236	163	66	7	69.07

Table 6.18 Classification result using combined set of features.

			No.of ROIs	TP	FP	FN	%Detection
Normal		Training set	56	56	-	-	100
		Test set	167	101	66	-	60.5
Abnormal	Training set	Calcification	14	14	-	-	100
		circumscribed masses	11	11	-	-	100
		Spiculated masses	13	13	-	-	100
		ill-defined masses	10	10	-	-	100
		Architectural distortion	3	3	-	-	100
		Asymmetry	4	4	-	-	100
	Test set	Calcification	20	19	-	1	95
		circumscribed masses	12	12		-	100
		Spiculated masses	6	6	-	-	100
		ill-defined masses	4	2	-	2	50
		Architectural distortion	16	14	-	2	87.5
		Asymmetry	11	9	-	2	81.8
Total			345	272	66	7	78.84

Table 6.19 Detailed results of Classification using combined set of features

6.8 Conclusion

PNN classifier for normal/abnormal classification of digitized mammogram has been implemented based on statistical features, textural features and a combination of these.

A TP identification rate of 58%, 59.5% and 60.5% were produced for the normal cases and 91.3%, 94.5% and 89.86% for the abnormal cases respectively. The texture-based classifier was unable to classify mammograms having the abnormality, asymmetry. The high TP rate of 94.5% for abnormal case using texture-based classifier is obtained excluding the mammograms containing asymmetry. As this is intended as the pre-processing stage of an automatic detection system, the major aim of this stage is to reduce the FN rate as far as possible. Hence, for further applications, the classifier based on statistical features is selected, even though it has slightly lower TP rate compared to the other two.

Chapter 7

Multiplexed Wavelet Transform Technique for Detection of Microcalcification

Today, breast cancer is one of the most frequent forms of cancer in women and also the leading cause of mortality. There is clear evidence, which reveals that early diagnosis and treatment can significantly increase the chance of survival [237], [238]. Among the different diagnostic methods currently available for detection of breast cancer, mammography is widely recognized as the chief modality for early detection in asymptomatic women [238].

One of the early symptoms of breast cancer is the appearance of microcalcification clusters, which have a higher X-ray attenuation than the normal breast tissue and appear as a group of small, localized, granular bright spots in mammograms. Popular methods for computer-aided detection of clustered microcalcifications include the difference image [44], multiscale processing based on fuzzy pyramidal linking [239], and spatial filtering techniques [8].

An efficient method for the detection of microcalcification must be capable of detecting objects with very small but varying sizes. Recently discovered multiorientation and multiresolution properties of the human visual system [240] has led to the idea of wavelet based multiresolution analysis for detection of microcalcification.

This approach has been used for detection and segmentation of microcalcifications and contrast enhancement [32], [241], [242].

7.1 Detection of Microcalcification - An Edge Detection Operation

Microcalcifications, small concentrations of calcium in the breast, is an early sign of possible cancerous growth. Individual calcifications are not worrisome; but when they appear in groups a potential tumor can be suspected. Five or more calcifications, measuring less than one millimeter, in a volume of one cubic centimeter define a 'cluster'. The possibility of malignancy increases as the size of the individual calcification decreases and also when the total number of calcifications per limit area increases. The risk increases when they are heterogeneous in size and shape.

Microcalcifications located in parenchymal structures show detectable edge features. The major constraint in their detection is the low contrast between normal and malignant tissues, especially in younger women. Their small size also contributes to a lower subject contrast. The assumptions made with regard to the nature of the microcalcifications are as follows:

1) They are of higher frequency than the surrounding breast tissue. Hence, they appear brighter.
2) They are usually 0.1-1mm in size.
3) The average calcification is roughly circular, and can be treated as a circular-symmetric Gaussian function.

7.2 Edge detection using MWT

Image sequences are found to be pseudo-periodic in both horizontal and vertical directions when treated as 1-D signals, with period same as the size of the image. Utilizing this pseudo-periodic nature, the MWT separates the periodic information and the inter-period fluctuations present in the sequence effectively into the scaling partials and wavelet partials respectively. In the case of the DWT, the asymptotically periodic part of the signal as well as the fluctuations gets filtered into different wavelet and scaling partials altogether. Hence MWT is found to be better for edge detection operations than DWT. A comparison between the MWT and 2-D DWT techniques for edge detection on a simple synthetic image using '*bior 6.8*', a biorthogonal wavelet basis, is shown in Figure 7.1. The figures clearly illustrate that the edge information gets precisely filtered into the wavelet partials of the MWT (see fig.7.1 (b), (c), (d)) leaving its scaling partials with only the edges blurred, whereas additional information is lost from the scaling partials of DWT (fig. 7.1(e)). Hence, accurate reconstruction of the edges can be achieved from the MWT wavelet partials.

In this work, we represent the microcalcification detection in mammograms as an edge detection operation, utilizing the pseudo-periodic nature of image sequences when treated as 1-D signals. Taking the MWT, i.e. DWT over samples that are spaced one period apart [25], pushes the edge information into the detail spaces more efficiently. Also, the edge detection methods using 2-D operators smear the edge information as a smoothing operation always precedes the edge detection operator. But smoothing is very important, as the differential operators are very much sensitive to noise [12], [243]. Since a 1-D processing technique is used here, the gray level transitions of an edge in the orthogonal direction will not get disturbed [244]. Moreover, the processing memory requirement is reduced to a size equal to the length / breadth of an image from that of the whole image in the 2-D DWT methods.

7.3 MWT based Microcalcification Detection

This section describes the novel WT based 1-D processing techniques developed for detecting and segmenting microcalcifications in digitized mammograms. The edge features formed by microcalcifications located in parenchymal structures are detected using the zero-crossings / local extrema of the MWT coefficients. The use of zero-

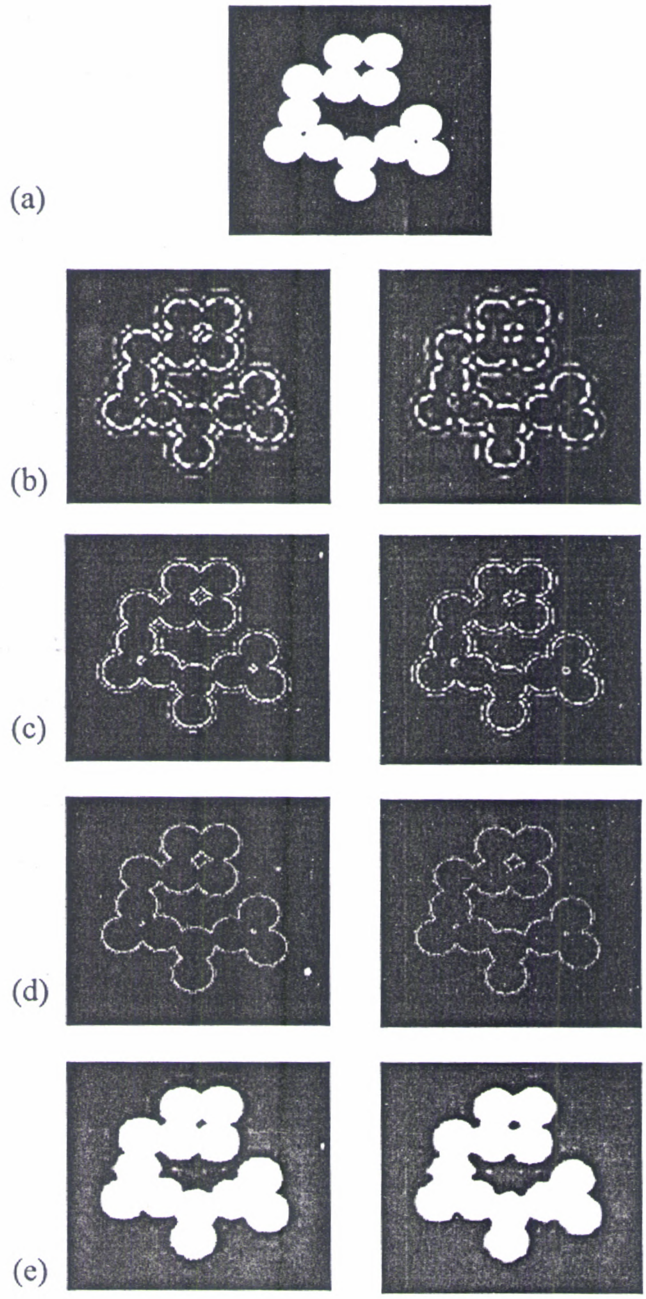


Figure 7.1: Image reconstruction using MWT (left) and 2DDWT (right) a) original image b) reconstructed from 3rd level WT coefficients alone. c) from 2nd level WT coefficients alone d) from 1st level WT coefficients alone. e) from 3rd level scaling transform coefficients alone

crossings results in a M-H edge detector [179] and local extrema results in a Canny edge detector [183]. Here, the Biorthogonal wavelet '*bior 1.3*' of support 6 is selected for edge detection. The level of processing required depends upon the resolution of image data. It was experimentally determined that a 3 level decomposition is sufficient for detecting the microcalcifications from the images in the databases used in this work.

The interperiod fluctuations corresponding to the intensity changes along the horizontal direction are determined by computing the MWT of the image up to the desired level, taking the image sequence to be periodic along the vertical direction. Similarly, considering the periodicity along the other direction, the MWT is computed to give the edge information along the vertical direction. Singular points are determined from the zero-crossings / local extrema of the MWT coefficients along both directions. The isolated single pixel intensity changes i.e. noise, are eliminated by retaining only those zero-crossings that hold a parent child relationship [148] from the coarsest to the finest levels of details. In this way the sensitivity of the differential operators to noise is taken care of without smearing the edges.

The retained points are boosted so as to enhance the edge features. The IMWT is applied on these to get the horizontal and vertical edge maps, which are then combined and scaled to get the complete microcalcification information. Global gray level thresholding based on image statistics is applied on the combined edge map to segment possible microcalcifications, with the rest of the pixel intensities being reduced to zero. The threshold T is selected to be proportional to the mean of the reconstructed edge map 'M'. i.e. $T=kM$; k was experimentally determined to be a real number such that $1<k<2$. The implementation consists of the following steps:

Step I :MWT computation -Computation of the MWT coefficients up to the desired level for a row/ column of the image.

Step II: Removal of noise -Discarding all zero-crossings/ local extrema that does not hold a parent-child relationship from coarser to finer levels.

Step III: Inverse transform- Scalar multiplication of the retained coefficients to boost them and taking inverse transform.

Steps I to III produces the edge map for a single row/column. These steps are repeated for all rows/ columns to get the complete horizontal/ vertical edge maps.

Step IV: Thresholding- Global gray level thresholding on the combined edge map to segment possible microcalcifications.

Figure 7.2 shows a summary of various operations involved in the detection and segmentation of microcalcifications employing the technique. The MWT based approach can enhance even subtle microcalcifications effectively, but it is also sensitive to mammographic noise such as quantum mottle and film graininess [245], [246]. It is difficult to suppress mammographic noise, while enhancing microcalcifications, because this noise is quite similar in size to that of microcalcifications. To reduce false alarms, a local thresholding depending on the power spectral density of the regions containing detected calcifications is carried out on the segmented calcifications as follows:

1. Determine the power spectral density from the original mammograms corresponding to all of the segmented regions obtained by the application of the algorithm in figure 7.2.
2. Determine the maximum value of power spectral density for a particular mammogram.
3. If the power spectral density is less than the minimum threshold set experimentally, discard those regions as false alarms.

   Else check whether the value is less than 0.6 times the maximum power spectral density.

      If so, discard it.

      Else consider it as a TP detection.

Incorporating the pre-processing stage developed in the previous chapter, the false alarm rate can be reduced further.

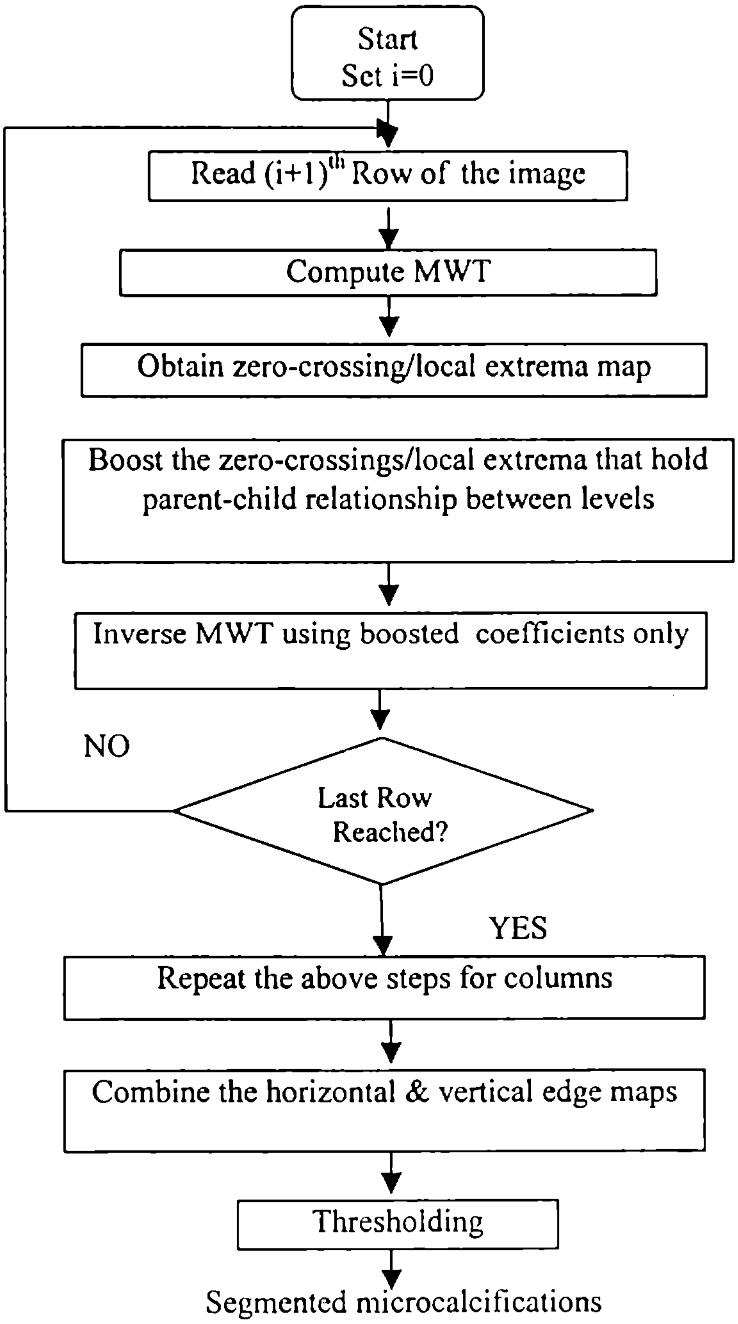


Figure 7.2: Steps for detection and segmentation of microcalcifications

7.3.1. Microcalcification Detection after Classification

The detection accuracy can be further increased by prescreening mammograms to identify the relatively large number of clearly normal mammograms, as well as large areas of clearly normal tissue in potentially abnormal mammograms. This will reduce unnecessary computations and will substantially increase the accuracy of diagnosis in subtle cases. The normal / abnormal classification algorithm based on statistical features, developed in the previous chapter is employed here as the pre-processing stage. This method has got high TP detection rate for the abnormal cases, and it has produced 100% detection rate for mammograms containing microcalcification, even though the detection sensitivity for the normal case is only 58%. The computational burden is also less for this method as only 3 features are to be extracted.

The distinct steps involved in the MWT based microcalcification detection system after PNN based normal / abnormal classification are listed below:

Step 1 Segmentation of image data: Overlapping blocks of image sections are used for processing in order to eliminate the possibility of misses near the block boundaries. The overlap depends on the length of wavelet used and level of decomposition. Since 3 levels of decomposition is performed using a wavelet of support 6 for microcalcification detection, the minimum overlap required is found to be 28. The block size selected is 256 x 256.

Step 2 Normal/abnormal classification: The features, mean, skewness and kurtosis are derived from the grey-level histogram and fed to the trained PNN to determine whether the given snippet is normal or abnormal.

Step 3 Microclcification detection: If a snippet of mammogram is found to be abnormal, the MWT based microcalcification detection algorithm is performed on that to see whether it contained any microcalcifications.

The abnormality need not be microcalcification alone. For detecting other abnormalities algorithms are to be developed.

7.4 Data and Detection Criteria

The algorithm has been validated on an ensemble of 40 mammograms, taken from the freely distributed digitized mammograms of the MIAS database of the University of Essex, England [232]. It has also been validated using a local database consisting of 30 mammograms. These were collected from restricted mammographic centers and scanned using UMAX Powerlook III scanner at a resolution of 2000 pixels/square inch. Abnormal regions in these mammograms were identified by an expert radiologist. Some of these were containing more than one type of abnormality. The details of mammograms used for evaluation purpose is given in table 7.1.

Database	No. of mammograms considered			No. of clusters
	Normal	Containing calcifications	Total	
MIAS	16	21 3	40	25 Distributed
Local	11	18 1	30	19 Distributed

Table 7. 1: Details of mammograms used for validation of the algorithm

Two closely related issues involved in any diagnostic investigation are sensitivity or 'true positive' and specificity or 'false positive'. Hence it is more appropriate to express the results in terms of these two. The detection criteria selected is that proposed by Karssemeijer [8], as there are no universally accepted detection criteria. Accordingly, for counting TP's a cluster is considered detected if two or more microcalcifications are found in the region of film identified by an expert radiologist. A FP is counted if two or more erroneous detections are made within an empty closed region of 0.5cm width.

7.5 Results and Discussion

As the detection of microcalcification in digitized mammograms is represented as an edge detection problem in this work, the detection capability of some of the popular edge detection techniques has been validated and the results are tabulated in table 7.2. The 24 mammograms containing microcalcifications from the MIAS database as detailed in table 7.1 are used for this study. The edge detection techniques selected for comparison are the conventional 2-D DWT based edge detection method and those available in MATLAB, Viz. the Sobel, Roberts, Prewitts, LOG, zero cross and Canny methods. The FP rate produced by the Canny detector is very high as it detects both strong and weak edges. Hence, it is found to be unsuitable for microcalcification detection and is not included in the table. The table shows the number of TPs, FPs and FNs for optimal threshold values experimentally selected for each of the algorithms mentioned above.

Method	TP	FP	FN
Sobel	10	51	18
Roberts	12	56	16
Prewitts	9	59	19
LOG	9	49	19
Zero-cross	9	42	19
2-D DWT	21	39	7

Table7. 2:Comparison of detection capability of various edge detection algorithms on microcalcification detection

The detection capability of the Canny and M-H methods of microcalcification detection on the MIAS database are tabulated in table 7.3 for different thresholds. The table shows the number of TPs, FPs and FNs for different values of *k*. It can be seen from the table that as the threshold is increased the FPs decrease, but the sensitivity also gets decreased. For screening mammography applications, where a high rate of FP is not tolerable, a high threshold is preferred. In diagnostic purposes, sensitivity is the important factor and hence smaller thresholds are to be used.

MIAS database 40 mammograms with 16 normal and 24 with 25 microcalcification clusters and 3 with distributed calcifications						
k	Canny			M-H		
	TP	FP	FN	TP	FP	FN
1.3	11	16	17	13	20	15
1.2	21	25	7	22	28	6
1.15	24	30	4	26	32	2

Table 7.3: Comparison of sensitivity and specificity of microcalcification detection using Canny and M-H detectors for different values of *k*.

It is seen that the application of our algorithm on the MIAS database resulted in a TP identification rate of 93% against 0.8 FP clusters per image for the M-H edge detector. (Rate of FP is computed considering all 40 images). With the locally obtained mammograms, a TP identification rate of 85% at the rate of 0.83 FP clusters per image was obtained. The Canny method produced 85.7% TP detection rate at the cost of a slightly lower FP rate per image, 0.75 for the MIAS database and 0.8 for the local database. Table 7.4 tabulates the detection sensitivity obtained for the two databases.

Database	Method	TP	FP	FN
MIAS database 40 mammograms-16 normal +21 mammograms with 25 microcalcification clusters and 3 with distributed calcifications	M-H	26	26	2
	Canny	24	22	4
Local 30 mammograms – 11 normal +18 mammograms with 20 microcalcification clusters and 1 with distributed calcifications	M-H	17	25	3
	Canny	17	24	3

Table 7.4: Detection Sensitivity for the two databases.

Typical examples of microcalcification detection using the two methods are illustrated in Figures 7.3 and 7.4. Figure 7.3 shows the detection of microcalcifications from three mammograms from the MIAS database, one containing a well defined cluster, second one having a hard to find cluster and the third, having distributed

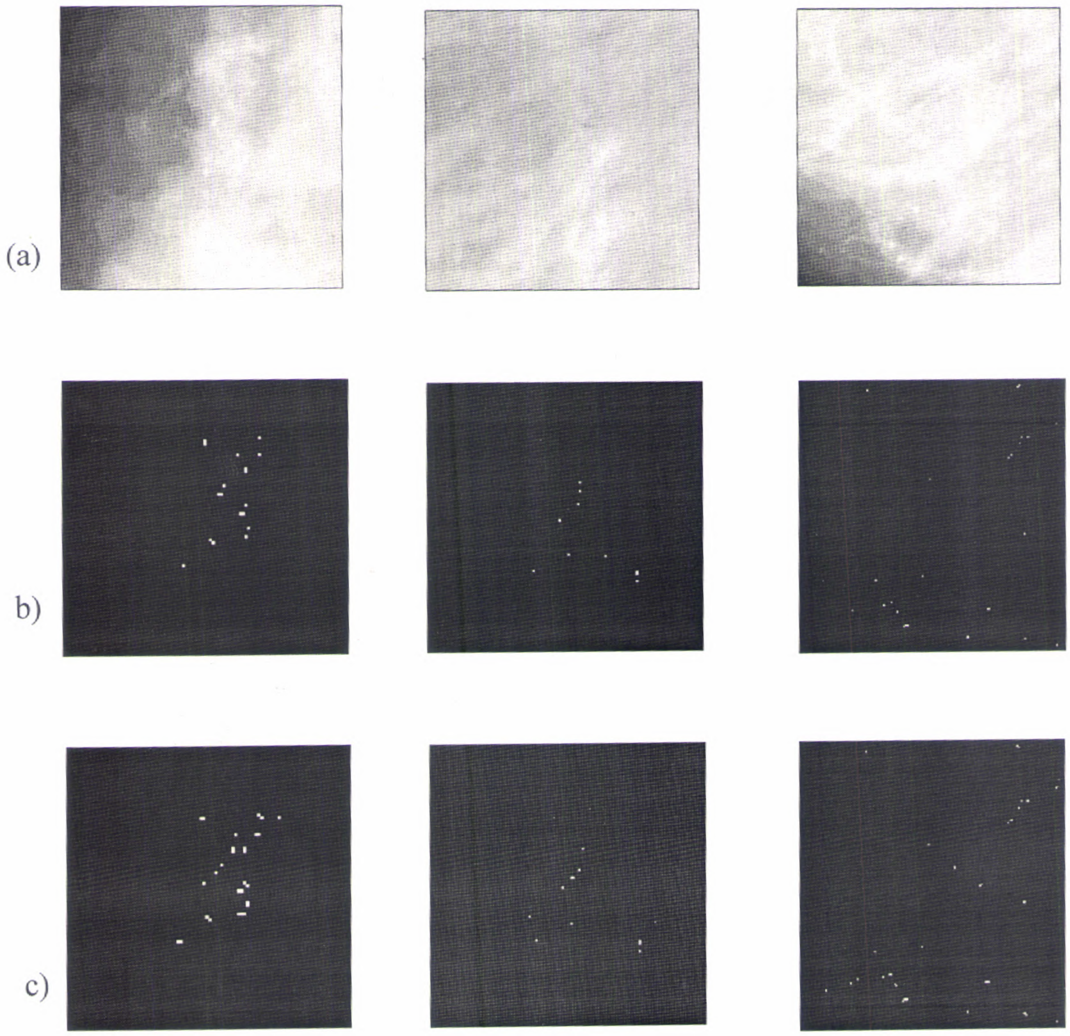


Figure 7.3: Comparison of detection of microcalcifications from various mammograms using Canny and M-H detectors on MIAS database (a)Sections of original mammograms containing a well-defined microcalcification cluster (left) hard to find cluster (middle) widely distributed calcifications (right). (b)& c) microcalcifications detected by Canny and M-H detectors respectively from the above mammograms.

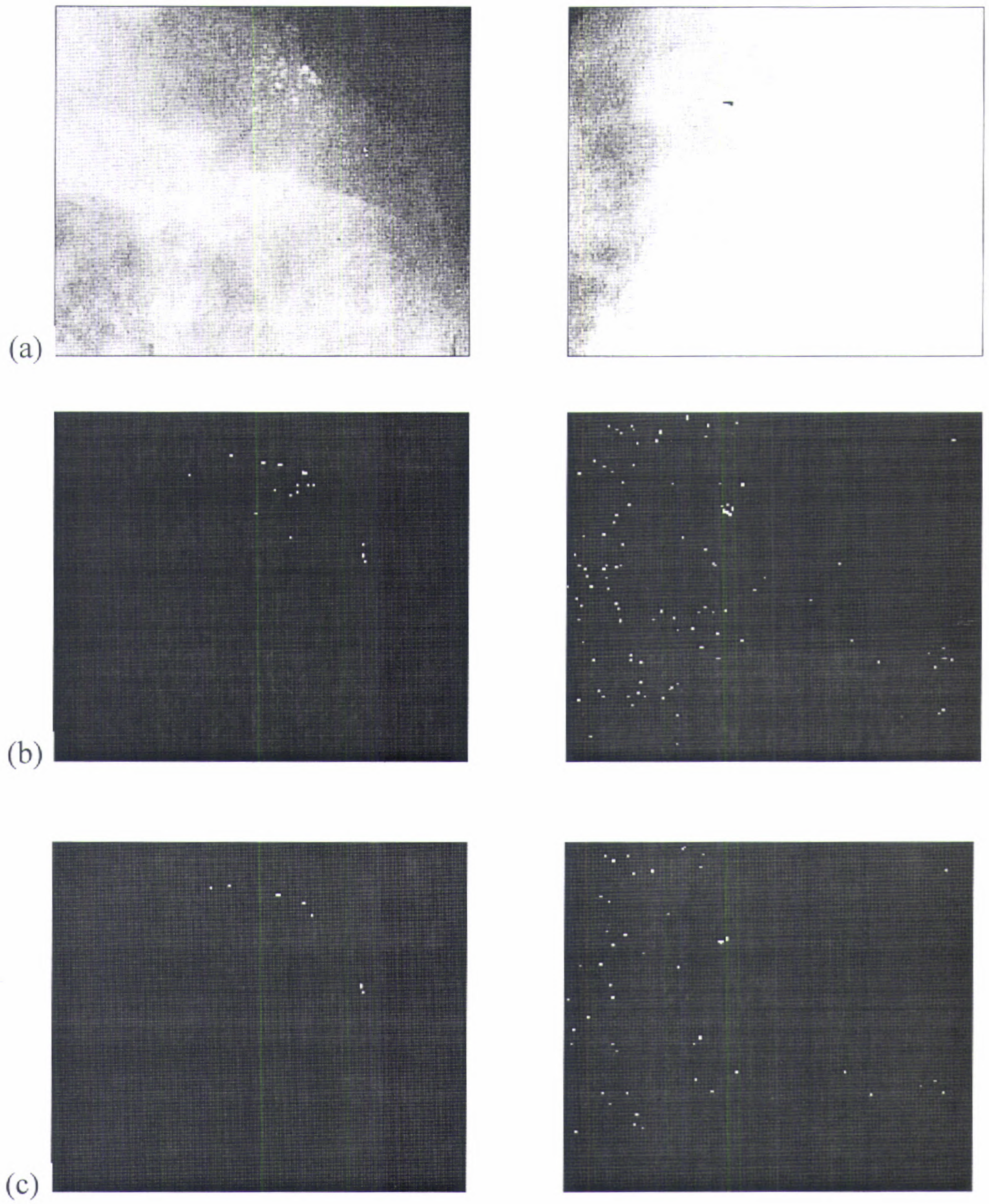


Figure 7.4: Comparison of detection of microcalcifications from various mammograms using M-H and Canny detectors on the local database (a) Sections of original mammograms containing a well-defined microcalcification cluster (left) and widely distributed calcifications (right). (b)& c) microcalcifications detected by M-H and Canny detectors respectively from the above mammograms.

calcifications. The detection of microcalcifications from two mammograms from the local database, having a well defined cluster and widely distributed calcifications are illustrated in fig 7.4. Figures clearly indicate that the shape of the cluster is preserved in the M-H method compared to Canny method. This is due to the symmetric nature and the ability to form closed contours of the M-H operator. The M-H edge detector provides better detection sensitivity for mammograms containing widely distributed calcifications.

Detection can be achieved using any of the biorthogonal wavelets. We have selected '*bior1.3*', being one of the smaller biorthogonal wavelets. It is found that the shape of the cluster is captured more accurately by the larger biorthogonal wavelets such as '*bior4.4*', '*bior 5.5*' and '*bior6.8*'. This is probably due to the fact that these filters approximate the response of the human visual system, in the sense that they are similar in form to the Laplacian of Gaussian described by Marr [17].

7.6 Conclusions

Microcalcification detection using two wavelet based edge detectors were performed and their performance was evaluated on two databases. A detection accuracy of 95% has been obtained for the MIAS database. It has been found that the M-H edge detector retains the shape information of the clusters, which is essential for classification. Also it detects distributed microcalcifications more efficiently. Both methods are found to be suitable for detecting subtle microcalcifications that could not be detected by other methods. The absence of edge smearing along the orthogonal direction due to the 1-D processing technique employed has helped in providing better detection efficiency. Moreover, the processing memory requirement is reduced to the size of one row or column, from that of the whole image, for a 2-D operator.

Chapter 8

Summary And Conclusions

8.1 Summary of the work and important conclusions

The work presented in this thesis describes the application of computer based image analysis techniques on digitized mammographic images, to classify them into normal / abnormal classes and detect clustered microcalcifications. The image analysis techniques are implemented using two image processing tools, *viz.* neural networks and WT. A BDWT technique without edge artifacts is also developed to process image blocks independently with reduced memory usage, thereby eliminating the need for the very large buffer requirement of DWT computation of digitized pathology images.

The first topic presented in this thesis is about the development of the BDWT computation algorithm. Employing the principle of block convolution, efficient algorithms having very low memory requirement and suitable for real time in-place computation of DWT/IDWT pair of 1-D and 2-D signals have been developed. Application of the overlap save method for DWT and the overlap add method for IDWT computation eliminates edge artifacts and reduces the buffer requirement to the order of the size of a block from that of the size of an image for the present-day algorithms. Also, the need for interprocessor communication is eliminated since the overlap is taken only once at the input.

The developed algorithms have been validated for various 1-D signals and monochrome images. The transform coefficients obtained using these methods were exactly the same as those obtained if the entire data is taken as a single block. But the coefficients corresponding to various sub bands appear in an interleaved manner in different blocks. Perfect reconstruction was achieved irrespective of the type and size of input signals, length of wavelet and level of decomposition. Both orthogonal and biorthogonal wavelets were employed in the study.

The block transform technique may result in a computation overhead, but this can be optimized for the available processing memory and tolerable processing delay by careful selection of frame sizes, wavelet filters and decomposition levels. The parallel processing ability and low-memory requirement help to speed up the algorithm.

Application of image analysis techniques for computerized detection of mammographic abnormalities is depicted in two parts. The first part deals with the classification of ROIs in mammograms into two groups; normal and abnormal, irrespective of the type, size and severity of the abnormalities. Large number of normal mammograms and large part of clearly normal tissue regions from potentially abnormal ones were separated out to ease the detection procedure. This also helps to increase the accuracy of interpretation. The second part is concerned with the development of MWT based algorithms for the detection of microcalcification clusters, which is the most important and sometimes the only sign of presence of cancer.

A PNN based approach was adopted for normal /abnormal classification of mammograms. Residual images were generated by removing normal background structures and normal linear markings from the ROIs. WT based background subtraction technique was used for removing the normal background regions. Lines were detected and removed using a line detector algorithm proposed by S.Liu. Statistical and textural features were derived from these ROIs using histogram and SGLD matrix approaches and fed to a PNN based classifier. The results have been validated using mammograms from the MIAS database.

The major aim of this stage is to segment all or most of the abnormalities present in a mammogram. A TP identification rate of 58%, for the normal cases and 91% for the abnormal cases has been achieved on the MIAS database.

The abnormal ROIs obtained from the above stage were checked for the presence of microcalcifications, using a 1-D processing technique based on DWT. The 1-D processing technique employed here reduces the processing memory requirement to a size equal to the length/ breadth of an image from its size of length x breadth in the 2-D DWT methods. This technique has the added advantage of not disturbing the gray level information along the orthogonal direction, as smoothing is performed in one direction only.

Depending on the zero-crossings/ local extrema of the MWT coefficients, two wavelet based edge detectors *viz.* the M-H and Canny detectors were developed for detection of microcalcifications. Their performances were evaluated on two databases, the MIAS database and a locally obtained database. It has been found that the M-H edge detector retains the shape information of the clusters, which is essential for classification. Also it detected distributed microcalcifications more efficiently. Both methods were found to be suitable for detecting subtle microcalcifications that could not be detected by other methods.

8.2 Scope for further investigations

Mammographic abnormalities can be broadly classified into two groups: microcalcifications and masses. Any system for automated detection of breast cancer will be incomplete if it does not take into consideration the detection of these two types of abnormalities. In this thesis, only the detection of microcalcification is taken care of. Efficient mass detection schemes are to be developed in order to make this a useful tool for the radiologists. Additionally, techniques are to be developed for considering multiple types of abnormalities in the same mammogram.

Classification of the detected abnormalities into benign and malignant groups is another important function required from an automated system for detection of breast cancer. Studies have shown that prior knowledge about the breast tissue type has increased the sensitivity of detection of lesions. Hence, classification of mammographic images based on the underlying texture contained within the breast tissue can also be attempted.

The generation of a database containing ample number of mammograms depicting various types of abnormalities as well as normal mammograms at various resolutions can be done as an extension to this work.

Appendix A

Line Detection Algorithm

A.1 Introduction

Linear structure detection is often a very important preprocessing step in image processing and computer vision, for applications such as feature extraction, pattern recognition, and image enhancement. Many line detection algorithms extract local edges first and then form lines based on various grouping criteria. However, they do not distinguish between edges resulting from lines and object boundaries [247]. Hough transform, the most widely used technique for detecting lines in binary images do not provide locations of lines and are not suitable for grayscale images. Hence a line detection algorithm based on a new line model is considered here for detection of normal linear markings in mammograms.

Normal linear markings in mammograms are the shadows of ducts and connective tissue elements that tend to radiate from the nipple toward the chest wall. They usually appear slightly curved, but over short segments, can be considered as straight line segments of dimensions 1 to 2 mm or greater and having varying width of 0.1 to 1mm. They may have very low contrast.

For detecting lines with varying width and low contrast, it is assumed that the most basic characteristic of a line, regardless of its thickness, is that pixels on it have similar gray levels. In other words, if a pixel belongs to a line, then there exists a string of pixels along the direction of the line with similar gray levels. A line can be perceived if its surrounding region has different gray levels from those of the pixels on the line. Also, it is assumed that the length of a line is always greater than its width.

A.2 Detection Algorithm

As described above, the basic characteristic of a line is that pixels on it have similar gray levels. A good measure of gray level similarity among pixels is the standard deviation. Let (i, j) be the spatial location in the image at row i and column j, $f(i, j)$ be the pixel gray level at (i, j), $L(\theta, l)$ be a string of pixels in direction θ and of length l and $N_{L(\theta,l)}$ be the number of pixels within $L(\theta, l)$. Then the standard deviation of pixel gray levels in $L(\theta, l)$ is

$$\sigma(\theta,l) = \sqrt{\frac{1}{N_{L(\theta,l)}-1}\sum_{m,n\in L(\theta,l)}(f(m,n)-\bar{f}_{L(\theta,l)})^2} \qquad \text{A(1)}$$

where (m, n) ε $L(\theta, l)$ means that the string $L(\theta, l)$ passes through the pixel (m, n) and $\bar{f}_{L(\theta,l)})$ is the average gray level of $L(\theta, l)$

$$\bar{f}_{L(\theta,l)} = \frac{1}{N_{L(\theta,l)}}\sum_{m,n\in L(\theta,l)} f(m,n) \qquad \text{A(2)}$$

Let $L_{i,j}(\theta, l) = L(\theta, 1)$ such that $\sigma_{i,j}(\theta,l) = \min_{(i,j)\in L_{(\theta,l)}} \sigma(\theta,l)$. If pixel (i, j) belongs to a line in the direction θ^* and of length greater than l, then $\sigma_{i,j}(\theta^*,l)$ is small. In the case of an ideal line where all pixels on it have the same gray levels, $\sigma_{i,j}(\theta^*,l) = 0$ does not lead to the conclusion that (i, j) belongs to a line. It may lie in a uniform region instead.

However, if (i,j) is in a uniform region, then $\sigma_{i,j}(\theta^*,l)$ is small for all θ. If we take the standard deviation of $\sigma_{i,j}(\theta^*,l)$ with regard to θ and denote

$$\sigma^2_{\sigma(i,j)}(l) = \int_0^{2\pi} \{(\sigma_{i,j}(\theta,l) - \bar{\sigma}_{i,j}(l))^2\} \frac{1}{2\pi} d\theta \qquad A(3)$$

where

$$\bar{\sigma}_{i,j}(l) = \int_0^{2\pi} \{\sigma_{i,j}(\theta,l) \frac{1}{2\pi}\} d\theta \qquad A(4)$$

Then $\sigma^2_{\sigma(i,j)}(l)$ is small for the case that (i, j) lies in a uniform region. Based on the above analysis, this line detector is capable of detecting lines of very different width from single pixel wide up to l and lines of any length greater than l. Detection of lines with varying width is possible if the changes are slower than l and that of curves if they can be approximated as lines of length greater than l over short segments.

In the actual implementation for any given θ and l, there are $N_{L(\theta,l)}$ number of strings passing through each pixel. Taking advantage of the fact that many $L(\theta, l)$ are the same for neighboring pixels, we only need to consider one of the $L(\theta, l)$s for each pixel (i,j). The steps to obtain $\sigma_{i,j}(\theta^*,l)$ can be implemented as follows:

1. Initiate $\sigma_{i,j}(\theta^*,l)$ to a large value for every pixel (i,j) in the image.
2. For each pixel (i,j), obtain $\sigma(\theta, l)$ for the string $L(\theta, l)$ that starts at (i, j).
3. Compare this $\sigma(\theta, l)$ to every $\sigma_{m,n}(\theta, l)$ for every $(m, n) \varepsilon L(\theta, l)$.
4. Replace $\sigma_{m,n}(\theta, l)$ with $\sigma(\theta, l)$ if the latter is smaller.

Due to the spatial redundancy in images, we can obtain a good estimation of $\sigma^2_{\sigma(i,j)}(l)$ using a number of equally spaced θ s. To make a binary decision as to whether or not a pixel (i, j) belongs to a line, we need to threshold $\sigma_{\sigma(i,j)}(l)$ and $\sigma_{i,j}(l)$. Figure A.1 illustrates the different steps involved in the implementation of the detection algorithm.

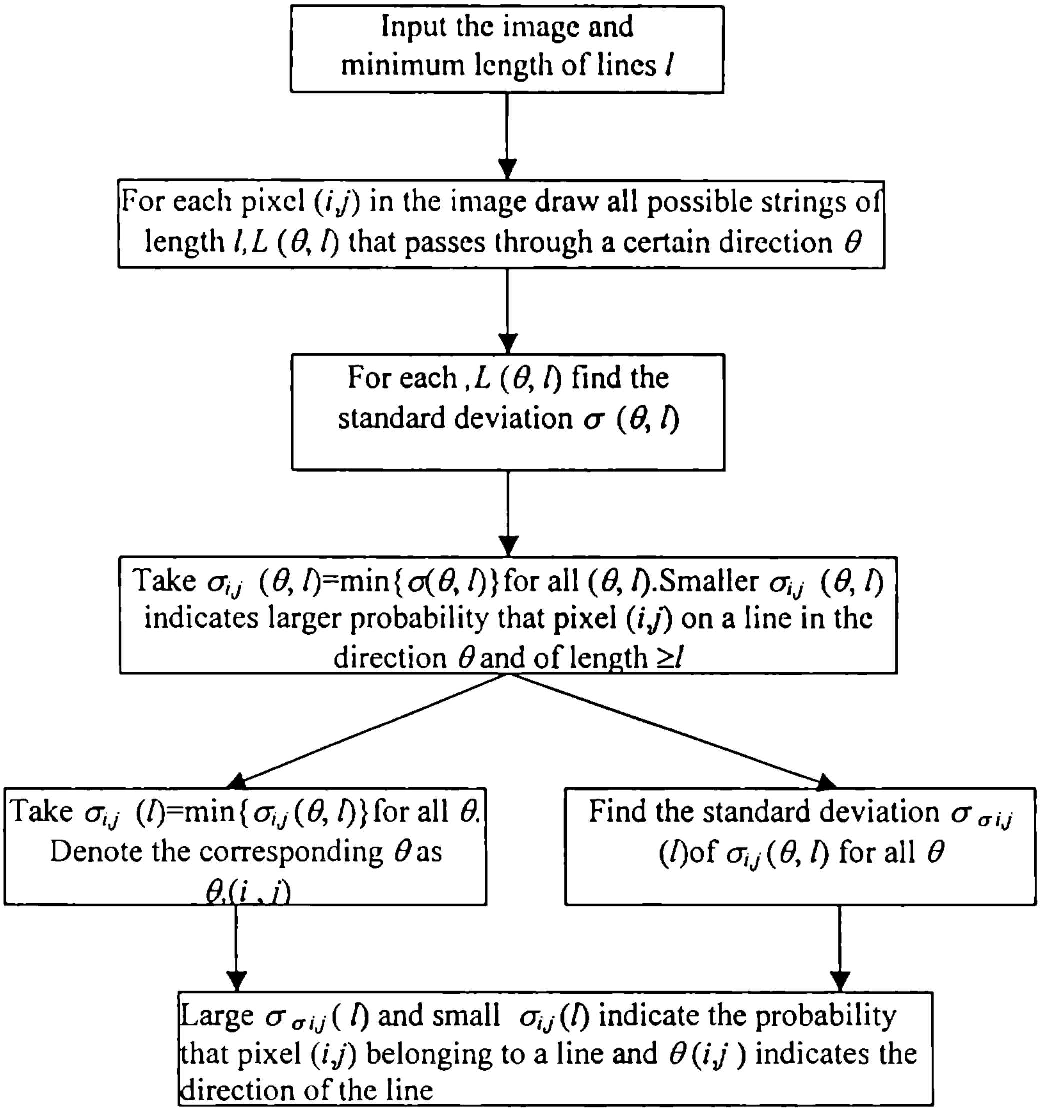


Figure A.1: Block diagram of the line detector

Bibliography

[1] D. B. Kopans, *Breast Imaging*, J. B. Lippincott Company, 1989.

[2] L. W. Bassett, V. P. Jackson, R. Jahan, Y. S. Fu, and R. H. Gold, *Diagnosis of Diseases of the Breast*, W. B. Saunders Company, 1997.

[3] Rao D.N, Ganesh B. "Estimate of cancer incidence in India in 1991", *Indian J Cancer*; vol 35, No.1, pp. 10-18, 1998.

[4] G. Cardenosa, "Mammography: An overview," *Proceedings of the 3rd International Workshop on Digital Mammography*, Chicago, Illinois, pp. 3-10, June 9-12, 1996.

[5] W. P. Kegelmeyer, Jr., J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M.L.Nipper, "Computer-aided mammographic screening for speculated lesions," *Radiology*, vol. 191, no. 2, pp. 331-336, May 1994.

[6] K. Woods and K. Bowyer, "A general view of detection algorithms," *Proceedings of the 3rd International Workshop on Digital Mammography*, Chicago, U.S.A., pp 385-390, June 9-12, 1996.

[7] M. L. Giger, F.F. Yin, K. Doi, C. E. Metz, R. A. Schmidt, and C. J. Vyborny, "Investigation of methods for the computerized detection and analysis of mammographic masses," *Proceedings of SPIE*, Bellingham, Washington, pp. 183-184, February 6 - 8 1990.

[8] N. Karssemeijer, "A stochastic model for automated detection of calcifications in digital mammograms," *Proceedings of 12th International Conference on Information Processing in Medical Imaging*, Wye, UK, pp. 227-238, 1991.

[9] C. J. Vybomy and M. L. Giger, "Computer vision and artificial intelligence in mammography," *American Journal of Roentgenology*, vol. 162, no. 3, pp. 699-708, March 1994.

[10] M. L. Comer, S. Liu, and E. J. Delp, "Statistical segmentation of mammograms," *Proceedings of the 3rd International Workshop on Digital Mammography*, Chicago, Illinois, pp. 475-478, June 9-12, 1996.

[11] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[12] R.C.Gonzales & P.Wintz , *Digital Image Processing*, Addison- Wesley, 1987.

[13] *Handbook of pattern recognition and Image processing*, Eds. T.Y.Young & K.S.Fu, Academic press Inc. Orlando, Florida, 1986.

[14] W.Pratt, *Digital image processing*, Wiley, Newyork, 1977.

[15] Rosenfield and Kak, *Digital image processing*, Academic Press, New York, 1980.

[16] G.J.Awcock and R Thomas, *Applied image processing*, Macmillan New Electronics, Hampshire, 1995.

[17] Gabor,D.,"Theory of Communication", *J. IEE* (London) , Vol. 93 ,Part III ,No. 26, pp. 429-457, November 1946.

[18] L.Prasad and S.S.Iyengar, *Wavelet Analysis with Applications to Image Processing*, CRC Press, Boca Raton, FL, 1997.

[19] A. Haar, "Zur theorie der orthogonalen Funktionensysteme," *Math. Ann.* 69: pp. 331-371, 1910.

[20] A.Grossman and J.Morlet, "Decompositionsof hardy functions into square integrable wavelets of constant shape", *SIAM Journal of Mathematical Analysis*, vol 15, pp 723-736, July 1984.

[21] S. Mallat, "A Theory for multi-resolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Recognition and Machine Intelligence*, vol.11, No.7, pp. 674-693, July1989.

[22] A.Abbate, C.M. DeCusatis, P.K.Das, *Wavelets and Subbands Fundamentals and Applications*, Birkhauser, Boston, 2002.

[23] M.Vetterli & J.Kovacevisc, *Wavelets and Sub band coding*, Englewoodcliff, NJ Prentice Hall, 1995.

[24] O.Rioul and P.Duhamel, "Fast algorithm for continuous and discrete wavelet transforms," *IEEE Trans. Inf. Theory*, vol.38, pp.569-586, March 1992.

[25] G. Evangelista, "Comb and Multiplexed Wavelet Transforms and their applications to signal processing," *IEEE Trans. S P*, vol.42, no.2, pp. 292-303, Feb. 1994.

[26] I.Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.

[27] A.F.Laine, S.Schuler, J.Fan, W.Huda, "Mammographic Feature Enhancement by multiscale analysis", *IEEE Transactions on Medical Imaging*, vol.13, No 4, pp 725-752, September 1994.

[28] A.F.Laine and S.Song , "Multiscale wavelet representations for mammographic feature analysis," *in Proc SPIE conf. Mathemat. Methods in Med. Imaging*. , vol. 1768, pp 306-316, 1992.

[29] W.B.Richardson Jr., " Nonlinear filtering and Multiscale texture discrimination for mammograms," *in Proc SPIE conf. Mathemat. Methods in Med. Imaging*. , vol. 1768, pp 293-305, 1992.

[30] L.P.Clarke et al. " Tree-structured nonlinear filter and wavelet transform for microcalcification segmentation in digital mammography", *Cancer letters*, vol.77, no 2-3, pp 173-181, 1994.

[31] W.Qian et al., "Digital mammography: m channel Quadrature mirror filters for microcalcification extraction", *Computerized medical imaging and Graph*ics, vol.18, No.5, pp 301-314, 1994.

[32] R. N. Strickland and H. I. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 2, pp. 218-229, April 1996.

[33] H.Yoshida, K.Doi, R.M.Nishikawa, M.L.Giger, R.A.Schmidt, "An improved CAD scheme using Wavelet Transform for detection of clustered Microcalcifications in digitized mammograms." *Academic Radiology*, No.3, pp. 621-627, 1996.

[34] J.B.Weaver, X.Yansun, D.M.Healy, Jr., and L.D.Cromwell, "Filtering noise from images with Wavelet Transforms," *Magnet. Reson. In Med.*, Vol 21, No.2, pp. 288-295, 1991.

[35] S.Mallat and S.Zhong, "Characterization of signals from multiscale edges", *IEEE Trans. Patt.Anal. Machine Intell.* ,Vol 14, pp. 710-732, July 1992.

[36] Y.Xu, J.B.Weaver, D.M.Healy, Jr.,and J.Lu, "Wavelet Transform domain filters: A spatially selective noise filtration technique" *IEEE Trans on Image Processing*, vol.3 , pp. 747-758, Dec 1994.

[37] M.Malfeit and D. Roose, "Biomedical Image denoising with wavelets and Bayesian geometrical constraints", *in. Proc. IEEE EMBS workshop on wavelets in Med and Biol.*, Baltimore, vol 2, pp 1228-1229, 1994.

[38] M.Unser , A.Aldroubi and C.R.Gerfen, "A multiresolution image registration procedure using spline pyramids", *in Proc. SPIE Conference Wavelet Applications in Signal and Image Processing* ,San Diego, CA, vol. 2034, pp 160-170, 1993.

[39] J.A.Freeman, D.M.Skapura, *Neural networks: Algorithms, Applications and Programming Techniques*, Addison-Wesley Publishing Co.,NewYork, 1992.

[40] Lippman, R. P., "An introduction to computing with neural nets,"*IEEE ASSP Magazine*, pp. 4-22, 1987.

[41] Caudill, M., *Neural Networks Primer*, San Francisco, CA: Miller Freeman Publications, 1989.

[42] R.A.Clark, P.S.King, C.E.Cox, J.Bromley and K. Mauer, "Screening mammography in the Tampa Bay area: Current status and Implications for the next decade", *The Journal of the Florida Medical Association*, pp. 449-453, May 1989.

[43] S.Astley, L.Hutt, S.Adamson, P.Miller, P.Rose, C.Boggis, C.J.Taylor, T.Valentine, J. Davies and J.Armstrong, "Automation in Mammography: Computer Vision and Human Peerception," *in Proc. SPIE/IS&T Syposium on Electronic Imaging Science and Technology,* vol.1905, San Jose, CA, pp.716-730, Jan 31-Feb4, 1993.

[44] Chan H P, Doi K, Vybomy C J, Schmidt R A, Metz C E, Lam K L, Ogura T, Wu Y, Mac Mahon H, " Improvements in radiologists' detection of clustered microcalcifications on mammograms: The potential of Computer-aided Diagnosis," *Investigative Radiology,* Vol. 25, pp. 1102-1110, 1990.

[45] A.L. Rosenberg, G.F. Schwarz, S.A. Feig, A.S. Patchefsky, "Clinically Occult Breast Lesions: Localization and significance," *Radiology*; vol.162, pp. 167-170, 1987.

[46] G. Hermann, C. Janus , S. Schwartz , A. Papatestas , D.G. Hermann, J.G. Rabinovitz, "Occult malignant Breast lesions in 114 patients: Relationship to age and the presence of microcalcifications," *Radiology* vol. 169, pp. 321-324, 1988.

[47] D. Brzakovic and M. Neskovic, "Mammogram screening using multiresolution based image segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 6, pp. 1437-1460, 1993.

[48] J. G. Elmore, C. K. Wells, C. H. Lee, D. H. Howard, and A. R. Feinstein, Variability in radiologists' interpretations of mammograms," *New England Journal of Medicine*, vol. 331, no. 22, pp. 1493-1499, December 1994.

[49] D. B. Spring and K. Kimbrell-Wilmot, "Evaluating the success of mammography at the local level: how to conduct an audit of your practice," *Radiologic Clinics of North America*, vol. 25, no. 5, pp. 983-992, September 1987.

[50] E. A. Sickles, "Quality assurance. How to audit your own mammography practice," *Radiologic Clinics of North America.*, vol. 30, no. 1, pp. 265-275, January, 1992.

[51] P. Y. Poon, P. Cairns, A. C. Lata, D. W. Marcuzzi, and G. M. Cooke, "Medical audit of mammography: A simplified alternative," *Canadian Association of Radiologists Journal*, vol. 43, no. 3, pp. 191-194, June 1992.

[52] L. Berlin and R. W. Hendrix, "Perceptual errors and negligence," *American Journal of Roentgenology*, vol. 170, no. 4, pp. 863-867, April 1998.

[53] P. J. Robinson, "Radiology's achilles' heel: Error and variation in the interpretation of the roentgen image," *British Journal of Radiology*, vol. 70, no 839, pp. 1085-1098, November 1997.

[54] D. B. Kopans, "The positive predictive value of mammography," *American Journal of Roentgenology*, vol. 158, no. 3, pp. 521-526, March 1993.

[55] A. M. Knutzen and J. J. Gisvold, "Likelihood of malignant disease for various categories of mammographically detected, nonpalpable breast lesions," *Mayo Clinic Proceedings*, vol. 68, no. 5, pp. 454-460, May 1993.

[56] M.Unser and A.Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, pp. 626-638, April 1996.

[57] S.M. Lai, X. Li, and W. F. Bischof, "On techniques for detecting circumscribed masses in mammograms," *IEEE Transactions on Medical Imaging*, vol. 8, no. 4, pp. 377-386, December 1989.

[58] D. Brzakovic, X. M. Luo, and P. Brzakovic, "An approach to automated detection of tumors in mammograms," *IEEE Transactions on Medical Imaging*, vol. 9,no. 3, pp. 233-241, September 1990.

[59] F.F. Yin, M. L. Giger, K. Doi, C. J. Vybomy, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Medical Physics*, vol. 21,no. 3, pp. 445-452, March 1994.

[60] A. J. M'endez, P. G. Tahoces, M. J. Lado, M. Souto, and J. J. Vidal, "Computer aided diagnosis: Automatic detection of malignant masses in digitized mammograms," *Medical Physics*, vol. 25, no. 6, pp. 957-964, June 1998.

[61] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Transactionson Medical Imaging*, vol. 14, no. 3, pp. 565-576, September 1995.

[62] L. Zheng, A. K. Chan, G. McCord, S. Wu, J. S. Liu, "Detection of Cancerous Masses for Screening Mammography using DWT based Multiresolution Markov Random Field," *Proceedings of the SCAR conference*, 1999.

[63] T. Hastie, D. Ikeda, and R.Tibshirani., "Computer-aided diagnosis of mammographic masses". Technical report, Departments of Preventive Medicine & Biostatistics, University of Toronto, June 1996.

[64] N. Petrick, H.P. Chan, B. Sahiner, and D. Wei, "An adaptive density-weighted Contrast enhancement filter for mammographic breast mass detection," *IEEE Transactions on Medical Imaging*, vol. 15, no. 1, pp. 59-67, February 1996.

[65] D. Wei, H.P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, and M. M.Goodsitt, "False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis," *Medical Physics*, vol. 24, no. 6, pp. 903-914, June 1997.

[66] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 17, no. 4, pp. 510-517, August 1998.

[67] K. Bovis and S. Singh, "Detection of Masses in Mammograms Using Texture Features",*Proceedings of the 2000 International Conference on Pattern Recognition.* Vol. II (2000) pp. 267-270, 2000.

[68] W. Qian, X. Sun, D. Song, R. A. Clark, "Digital Mammography: Wavelet Transform and Kalman-filtering Neural Network in Mass Segmentation and Detection",*Acad Radiol*, No.8, pp1074–1082, 2001.

[69] S. B. Lo, H. Li, Y. Wang, L. Kinnard, and M. T. Freedman, "A Multiple Circular Path Convolution Neural Network System for Detection of Mammographic Masses," *IEEE Transactions on Medical Imaging*, Vol. 21, No. 2,pp 150-158, February 2002.

[70] W. P. Kegelmeyer, Jr., "Computer detection of stellate lesions in mammograms,"*Proceedings of the SPIE Conference on Biomedical Image Processing and Three-Dimensional Microscopy*, February 10-13 1992, San Jose, California, pp. 446-454.1992

[71] Z. Huo, M. L. Giger, C. J. Vybomy, U. Bick, P. Lu, D. E. Wolverton, and R. A. Schmidt, "Analysis of spiculation in the computerized classification of mammographic masses," *Medical Physics*, vol. 22, no. 10, pp. 1569-1579, October 1995.

[72] N. Karssemeijer and G. M. Brake, "Detection of stellate distortions in mammograms," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 611-619, October 1996

[73] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 235-245, June 1996.

[74] S. Liu , C. F. Babbs , and E. J. Delp , "Multiresolution Detection of Spiculated Lesions in Digital Mammograms", *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 874-884, June 2001.

[75] S. Liu and E. J. Delp, "Multiresolution detection of stellate lesions in mammograms," *Proceedings of the IEEE International Conference on Image Processing*, October 26-29, Santa Barbara, California, pp. 109-112, 1997.

[76] H. Qi, W. E. Snyder, "Lesion detection and characterization in digital mammography by Bezier histograms", *SPIE Medical Imaging: Image Processing*, Pt.1-2, pp 1521-1526, February 1999.

[77] H.P. Chan, K. Doi, S.Galgotra, C.J. Vyborny , H.Macmohan and P.M.Jockih, "Image feature analysis and computer-aided diagnosis in digital radiography: automated detection of microcalcifications in mammography", *Medical Physics*, Vol 14, pp. 538-548, July/ August 1987.

[78] H.P. Chan and K. Doi and C.J. Vyborny and K.L. Lam and R.A. Schmidt, "Computer-aided detection of microcalcifications in mammograms: Methodology and preliminary clinical study", *Investigative Radiology*, Vol. 23, No. 9, pp. 664-671, Sept. 1988.

[79] D.H.Davies and D.R.Dance, "Automatic computer detection of clustered calcifications in digital mammograms," *Physics in Medicine and Biology*, vol. 35, no.8, pp 1111-1118, 1990.

[80] R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computer-aided detection of clustered microcalcifications on digital mammograms," *Medical and Biological Engineering and Computing*, vol. 33, no. 2,pp. 174-178, March 1995.

[81] H.Yoshida, K.Doi, R.M.Nishikawa, M.L.Giger, R.A.Schmidt, "An improved CAD scheme using Wavelet Transform for detection of clustered Microcalcifications in digitized mammograms." *Academic Radiology*, vol. 3 pp. 621-627, 1996.

[82] H.P. Chan, S.C. B. Lo, B. Sahiner, K. L. Lam, and M. A. Helvie, "Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network," *Medical Physics*, vol. 22, no. 10, pp. 1555-1567, October 1995.

[83] H.Yoshida, R.M.Nishikawa,M.L.Giger and K.Doi, "Signal/background separation by wavelet packets for detection of microcalcifications in mammograms ", *SPIE Proceedings Vol. 2825, Wavelet Applications in Signal and Image Processing IV*,6-9 Aug. 1996, Denver, CO, USA, 1996.

[84] H.Yoshida, "Matching pursuit with optimally weighted wavelet packets for extraction of microcalcifications in mammograms", *Applied Signal Processing*, vol. 5, no.3, pp 127-141, 1999.

[85] M. N. Gurcan, Y. Yardimci, A. E. Cetin, and R. Ansari, "Detection of microcalcifications in mammograms using higher order statistics," *IEEE Signal Processing Letters*, vol. 4, no. 8, pp. 213-216, August 1997.

[86] N. Ibrahim, H. Fujita, T. Hara and T. Endoz , "Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database", *Phys. Med. Biol.* Vol. 42 , pp. 2577–2589, 1997.

[87] H.D. Cheng, Y. M. Lui, and R. I. Freimanis, "A novel approach to microcalcification detection using fuzzy logic technique," *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 442-450, June 1998.

[88] R. H. Nagel, R. M. Nishikawa, J. Papaioannou, and K. Doi, "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms," *Medical Physics*, vol. 25, no. 8, pp. 1502-1506, August 1998.

[89] G. Boccignone, A. Chianese, A. Picariello, "Entropy-Based Detection of Microcalcifications in Wavelet Space", *Proceedings of ICASSP* 1998, Seattle, Washington, U.S.A, May 12-15, 1998.

[90] J. K. Kim and H. W. Park, "Statistical Textural Features for Detection of Microcalcifications in Digitized Mammograms", *IEEE Transactions On Medical Imaging*, Vol. 18, No. 3, pp. 231-238 , March 1999.

[91] F. Schmidt, E. Sorantin, C. Szepesv'ari, E. Graif, M. Becker, H. Mayer and K.Hartwagner, "An automatic method for the identification and interpretation of clustered microcalcifications in mammograms", *Phys. Med. Biol.* Vol.44 pp. 1231–1243, 1999.

[92] Sorantin E., Schmidt F., Mayer H., Becker M., Szepesvari C., Graif E., Winkler P., "Computer Aided Diagnosis of Clustered Microcalcifications Using Artificial Neural Nets", *Journal of Computing and Information Technology* - CIT 8, No. 2, pp.151-160, 2000.

[93] A.Bazzani, D.Bollini ,R.Brancaccio, R.Companini, N.Lanconelli and D.Romani, "System for automatic detection of clustered microcalcifications in mammograms ", *International Journal of Modern PhysicsC*, Vol.11, No.5, pp 1-12, 2000.

[94] S. Yu and L.Guan, "A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films", *IEEE Transactions on Medical Imaging*, Vol. 19, No. 2, pp 115-126, February 2000.

[95] T. O. Gulsrud and J. H. Husøy, "Optimal filter based detection of microcalcifications," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1272-1280, 2001.

[96] F.Diekmanna, P.Heinleinb, J.Drexlb, S.Grebea, A.Gosslerb, W. Schneiderb , B. Hamma ,"Visualization of microcalcifications by full-field digital mammography using a wavelet algorithm," *International Congress Series 1230* (2001) pp.526–530,2001.

[97] C. Serrano, J. Díaz-Trujillo, B. Acha and R. M. Rangayyan, "Use of 2d linear prediction error to detect microcalcifications in mammograms,"*II Congreso Latinoamericano de Ingeniería Biomédica*, La Habana, May 2001.

[98] M.Melloul and L. Joskowicz, "Segmentation of microcalcification in X-ray mammograms using entropy thresholding ," *CARS 2002* – H.U. Lemke, M.W. Vannier; K. Inamura, A.G. Farman, K. Doi & J.H.C. Reiber (Editors), 2002.

[99] A. Papadopoulosa, D.I. Fotiadisb, A. Likasb, " An automatic microcalcification detection system based on a hybrid neural network classifier", *Artificial Intelligence in Medicine*, vol.25, pp. 149–167, 2002.

[100] A. Laine and S. Schuler, "Hexagonal wavelet processing of digital mammography", *in Medical Imaging 1993 (Part of SPIE's Thematic Applied Science and Engineering Series)*, Newport Beach, California, Feb. 1993.

[101] L.Gagnon,J.M. Lina ,B. Goulard, "Sharpening enhancement of digitized mammograms with Complex Symmetric Daubechies wavelets," *Proceedings of the17th Annual International Conference of Engineering in Medicine and Biology Society (IEEE)*, Montréal, 1995.

[102] H. Li, K. J. Ray Liu, and S.B. Lo, "Fractal Modeling and Segmentation for the Enhancement of Microcalcifications in Digital Mammograms", *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp.785-798, December 1997.

[103] J. K. Kim, J. M. Park, K. S. Song, and H. W. Park, "Adaptive Mammographic Image Enhancement Using First Derivative and Local Statistics", *IEEE Transactions on Medical Imaging*, Vol. 16, No. 5, pp.495-502, October 1997.

[104] I.Koren, A.Laine ,F. Taylor ,"Enhancement via fusion of mammographic features ," *Proc. Int. Conf. Image Processing*, vol. 2, Chicago,IL, pp. 722–726, Oct. 1998.

[105] D. Sersic, S. Loncaric, "Enhancement of mammographic images for detection of microcalcifications", *Proceedings of the IX European signal processing conference* ,vol2, pp 693-696, Island of Rhodes, Greece, 1998.

[106] P. Heinlein, J. Drexl, and W. Schneider, "Integrated Wavelets for Enhancement of Microcalcifications in Digital Mammography," IEEE *Transactions on Medical Imaging*, Vol. 22, No. 3, pp.402-413, March 2003.

[107] D.D.Adler and M.A.Helvie, "Mammographic biopsy recommendations", *Current Opinion Radiol,* No. 4, pp. 123–129,1992.

[108] L. Shen, R. M. Rangayyan, and J. E. L. Desautels," Applications of shape analysis to mammographic calcifications," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 263-274, June 1994.

[109] A. P. Dhawan, Y. Chitre, C. Kaiser-Bonasso, and M. Moskowitz, "Analysis of mammographic microcalcifications using gray-level image structure features," *IEEE Transactions on Medical Imaging*, vol. 15, no. 3, pp. 246-259, June 1996.

[110] H-P. Chan, B. Sahiner, N. Petrick, M. A Helvie, K. L. Lam, D. D. Adler and M.M. Goodsitt, "Computerized classification of malignant and benign microcalcifications on mammograms: texture analysis using an artificial neural network," *Phys. Med. Biol.*, Vol .42 , pp 549–567, 1997.

[111] M. F. Salfity, G. H. Kaufmann, P. Granitto and H. A. Ceccatto, "Automated detection and classification of clustered microcalcifications using morphological filtering and statistical techniques," *29th International Conference of the Argentine Computer Science and Operational Research Society*, Tandil, Argentina, 2000.

[112] I. Christoyianni, E. Dermatas and G. Kokkinakis, "Neural Classification of Abnormal Tissue in Digital Mammography using Statistical Features of the Texture", *International Conference on Image Processing (ICIP'01)*, Thessaloniki, Greece, October 2001.

[113] B. Verma and J. Zakos , "A Computer-Aided Diagnosis System For Digital Mammograms Based On Fuzzy-Neural And Feature Extraction Techniques",*IEEE Transactions on Information Technology in Biomedicine*, vol. 5, No. 1, pp 46-54, March 2001.

[114] S. S. Buchbinder, I. S. Leichter, R. B. Lederman, B. Novak, P. N. Bamberger, H. Coopersmith, S. I. Fields, "Can the Size of Microcalcifications Predict Malignancy of Clusters at Mammography? ",*Acad Radiol* no. 9, pp.18–25, 2002.

[115] J. Kilday, F. Palmieri, and M. D. Fox, "Classifying mammographic lesions using computerized image analysis," *IEEE Transactions on Medical Imaging*, vol. 12, no. 4, pp. 664-669, December 1993.

[116] S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote, and S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms quantitative classification of breast tumors in digitized mammograms," *Medical Physics*, vol. 23, no. 8, pp. 1337-1345, August 1996.

[117] R. M. Rangayyan, N. M. El.Faramawy, J. E. L. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 799-810, December 1997.

[118] B. Sahiner, H.P. Chan, N. Petrick, M. A. Helvie, and M. M. Goodsitt, "Computerized characterization of masses on mammograms: The rubber band straightening transform and texture analysis," *Medical Physics*, vol. 25, no. 4, pp. 516-526, April 1998.

[119] D. Guliato, R.M. Rangayyan, W.A. Carnielli, J.A. Zuffo, and J.E.L. Desautels, "Segmentation of breast tumors in mammograms by fuzzy sets", *Journal of Electronic Imaging*. 2003 (In press).

[120] S. Liu, C. F. Babbs, and E. J. Delp, "Normal mammogram analysis and recognition,"*Proceedings of the IEEE International Conference on Image Processing*, October 4-7 1998, Chicago, Illinois, pp. 727-731, 1998.

[121] B.Sahiner, H-P. Chan, N.Petrick,D.Wei, M.A.Helvie, D.D.Adler, M.M.Goodsitt, "Classification of Mass and Normal Breast Tissue: A Convolution Neural Network Classifier with spatial Domain and Texture Images", *IEEE Transactions on Medical Imaging* , Vol 15, No. 5 , pp. 598-610,October 1996.

[122] J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke, Multiresolution statisitical analysis of high-resolution digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 16, no. 5, pp. 503-515, October 1997.

[123] B. L. Kalman, S. C. Kwasny, and W. R. Reinus, "Diagnostic screening of digital mammograms using wavelets and neural networks to extract structure",*Academic Radiology*, vol. 4, no. 6, pp. 405-414, June 1997.

[124] S. Liu, C. F. Babbs, Z. Pizlo, and E. J. Delp, "Characterization of Normal Mammograms", *Proc. IWDM* 2000, Canada, June 2002.

[125] K.Bovis and S. Singh, "Classification of Breast Density in Digital Mammograms", PANN Research, Department of Computer Science, Exeter University, 2003.

[126] O. R. Zaýane, M. Antonie, A. Coman , "Mammography Classification by an Association Rule-based Classifier," *MDM/KDD 2002: International Workshop on Multimedia Data Mining (with ACM SIGKDD)* 2002.

[127] J. Y. Lo, M. Gavrielides, M. K. Markey, J. L. Jesneck, "Computer-aided classification of breast microcalcification clusters:Merging of features from image processing and radiologists," *Medical Imaging 2003: Image Processing, Proceedings of the SPIE 5032*, pp. 882-889, 2003.

[128] C. J. Vyborny and M. L. Giger, "Computer vision and artificial intelligence in mammography," *American Journal of Roentgenology*, vol. 162, no. 3, pp. 699-708, March 1994.

[129] *Proc. of The IEEE: Special Issue on Wavelets*, No 4. 1996.

[130] C. Chakrabarti and M. Vishwanath, "Effcient realizations of the discrete and continuous wavelet transforms: From single chip implementations to mappings on SIMD array computers," *IEEE Trans. on Signal Proc.*, vol.43, no. 3, pp. 759-771, Mar. 1995.

[131] J. Fridman and E. S. Manolakos, "On the scalability of 2-D discrete wavelet transform alogrithms," *Multidimensional Systems and Signal Processing*, no. 8, pp. 185-217, 1997.

[132] K. K. Parhi and T. Nishitani, "VLSI architectures for discrete wavelet transforms," *IEEE Trans. on VLSI System*, vol. 1, no. 2, pp. 191-202, June 1993.

[133] M. Vetterli, "Running FIR and IIR filtering using Multirate Filter banks," *IEEE Trans. on Acoustic, Speech & Signal Processing*, vol. ASSP-36, pp. 730-738, May 1988.

[134] P. P. Vaidyanathan and P. -Q. Hoang, "Lattice structures for optimal design and robust implementation of two-channel perfect-reconstruction QMF banks," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 1, pp.81-94, Jan. 1988.

[135] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proceedings of the IEEE*, vol. 78, no.1, pp. 56-93, Jan. 1990.

[136] T. G. Marshall, "Zero-phase filter bank and wavelet coder matrices: Properties, triangular decompositions, and a fast algorithm," *Multidimensional Systems and Signal Processing*, vol. 8, pp. 71-88, 1997.

[137] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247-269, 1998.

[138] M. Vishwanath, "The recursive pyramid algorithm for the discrete wavelet transform," *IEEE Trans. on Signal Proc.*, vol. 42, no. 3, Mar. 1994.

[139] C. Chrysafis and A. Ortega, "Line based, reduced memory, wavelet image compression," *in Proc. of Data Compress. Conf.*, 1998.

[140] P. Cosman, T. Frajka, and K. Zeger, "Image compression for memory constrained printers," *in Proc. of ICIP*, Oct. 1998.

[141] H. Sava, A. C. Downtown, and A. F. Clark, "Parallel pipeline implementation of wavelet transform," *Proc. Inst. Elect. Eng., Vision Image Process.*,vol. 144, no. 6, pp. 355-359, Dec. 1997.

[142] F. Marino, V. Piuru, and E. Swartzlander Jr., "A parallel implementation of the 2D discrete wavelet transform without interprocessor communication," *IEEE Trans. Signal Proc.*, vol. 47, no. 11, pp. 3179-3184, Nov. 1999.

[143] L. Yang and M. Misra, "Coarse-grained parallel algorithms for multi-dimensional wavelet transforms," *The Journal of Supercomputing*, vol. 12, no. 1/2, pp. 9-118, 1998.

[144] A.V. Oppenheim, and R.W.Schafer, '*Discrete-Time Signal Processing*' Prentice Hall, Inc., N.J., 1989

[145] M. D. Adams and F. Kossentini, "Performance evaluation of the spatially segmented wavelet transform in the JPEG-2000 baseline system", *ISO/IEC JTC 1/SC 29/WG* 1 N 868, June 1998

[146] C. Chrysafis and A. Ortega, "Line based, reduced Memory, wavelet Image compression", *IEEE Transactions on Image Processing*, Vol. 9, pp. 378-389, March 2000.

[147] L. Yang and M. Misra, "Parallel wavelet transforms for image processing," *in SAE Workshop on Application of Parallel Processing for Image Processing*, Aug. 1997.

[148] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients", *IEEE Transaction on Signal Processing*, Vol. 41, pp. 3445-3462, December 1993.

[149] A. Said and W. A. Pearlman. "A new, fast and efficient image codec based on set partitioning in hierarchal trees", *IEEE Transaction on Circuits and Systems for Video Technology*, vol 6 , No.3, pp. 243-250, June 1996.

[150] J. R. Williams and K. Amaratunga, "A discrete wavelet transform without edge effects using wavelet extrapolation", *IESL Technical Report* No. 95-02, 28 January 1995.

[151] E. K. Eom, Y. S. Kim, and J. H. Kim, "A block wavelet transform for sub-image coding/decoding", *SPIE Electronic Imaging*, Vol. 2669 pp.169-177, January 1996.

[152] Motorola Australian Research Center, "A memory saving method of calculating the discrete wavelet transform", *Project report*, December 1997.

[153] J. H.Nealand, A. B.Bradley, M. Lech, "Overlap-Save Convolution Applied to Wavelet Analysis" *IEEE Signal Processing Letters* vol.10, No.2 pp. 47-49, Feb.2003.

[154] V.P. Devassia, M.G. Mini, Tessamma Thomas, "Novel Computational Structure for Real-Time Wavelet Analysis", Accepted in the *SPIE International Conference on Modeling, Signal Processing and Control*, SS03, San Diego, USA, June 24-27, 2002

[155] V.P. Devassia, M.G. Mini, Tessamma Thomas , "A Novel Parallel Structure for Computation of Discrete Wavelet Transform of Images," *AMSE J. Advances in Modelling & Analysis, Series D*: vol. 7, no. 3, pp. 25-40, 2002.

[156] L.Benjamin, K.T.Bruchak, M.Hampshire and E.L.Buhle, "The nature of cancer", July, 1997. http://cancer.med.upenn.edu/specialty/med_one/nature_cancer.html

[157] J.Gershon- *Cohen Atlas of mammography*, Springer- Verlag, Berlin, 1970.

[158] N.Bjurstam, *Breast cancer*, Zeneca AB, Sweden, 1996.

[159] L.Tabar, "Mammography and breast cancer: the new era", *International Journal of Gynecology & Obstetrics*, Vol. 82, Issue 3, pp 319 - 326

[160] http://gemedicalsystem.com.

[161] A.G.Haus, M.J.Yaffe, *Screen -film and Digital Mammography: Image Quality and radiation dose considerations* , Eds. Breast imaging – Radiologic clinics of North America, W.B.Saunders, Philadelphia, July 2000.

[162] A. Zygmund, *Trigonometric Series*, 2nd ed., Cambridge: Cambridge Univ.Press, 1968.

[163] A. Cohen, R. D. Ryan, *Wavelets and Multiscale Signal Pro*cessing, Chapman & Hall, 1995.

[164] R.M .Rao & A.S.Bopardikar, *Wavelet Transforms: Introduction to Theory and Applications*,India: Addison-Wesley, 2000.

[165] A.Bruce and H.Y.Gao, *Applied Wavelet Analysis with S-PLUS*. NY: Springer-Verlag,1996.

[166] G.Evangelista," Orthogonal Wavelet Transforms and Filter Banks", presented at *IEEE 23rd Asilomar Conf. Circuits, Syst., Computers*, November 1989.

[167] M.Vetterly and C.Herly, "Wavelets and Filterbanks: Relationships and new results," in Proc. *1990 IEEE Int. Conf. Acoust.,Speech,Signal Processing*,(Albuquerque,NM), pp 1723-1726, April 3-6, 1990.

[168] M.R.K.Khansari and A. leon-Garcia, "Subband decomposition of signals with generalized sampling",*IEEE Trans. Signal Processing*, Vol. 41,no.12, pp 3365-3376,1993.

[169] M.Vetterly and J.Kovacevic, *Wavelets and subband coding*, Englewood Cliffs, NJ, Prentice-Hall, 1995.

[170] O.Rioul,"Regular wavelets:a discrete time approach", *IEEE Trans. Signal Processing*, Vol. 41, pp 3572-3579, December 1993.

[171] O.Rioul,"A discrete time multiresolution theory", *IEEE Trans. Signal Processing*, Vol. 41, pp 2591-2606, August 1993.

[172] S. Mallat, *A Wavelet Tour of Signal Processing*, New York: Academic, 1998.

[173] G.Strang and T. Nguyen, *Wavelets and Filterbanks* , MA: Wellesly-Cambridge Press, 1996.

[174] M. A. Westenburg and J. B.T.M. Roerdink, "Frequency Domain Volume Rendering by X-ray Wavelet Transform", *IEEE Trans. Image Processing*, vol.9, no.7, pp. 1249-1261, July 2000.

[175] K. R. Castleman, *Digital Image Processing*, Englewood Cliffs, NJ: PrenticeHall, 1996.

[176] L. Mero, "A simplified and fast version of the Hueckel operator for finding optimal edges in pictures," *Pric. IJCAI*, pp. 650-655, 1975.

[177] R. Nevatia, "Evaluation of simplified Hueckel edgeline detector," *Comput.,Graph., Image Process.*, vol. 6, no. 6, pp. 582-588, 1977.

[178] M. Hueckel, "An operator which locates edges in digital pictures," *J. ACM*, vol.18, no. 1, pp. 113-125, 1971.

[179] D. Marr, E. C. Hildreth, "Theory of edge detection," *Proc. R. Soc.*, B 207, pp.187-217,London, U.K, 1980.

[180] D.Marr, S. Ullman, "Directional selectivity and its use in early visual processing," *Proc. R. Soc.* Lond. B, vol. 208, 1981.

[181] D. Marr, T. Poggio, "A theory of human stereo vision," *Proc. R. Soc.*, London,U.K., B 204, pp. 301-328, 1979.

[182] R. M. Haralick, "Digital step edges from zero crossing of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI 6, no. 1, pp. 58-68, 1984.

[183] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI8, pp. 679-698, 1986.

[184] J. C. Goswami, A. K. Chan, *Fundamentals of wavelets: theory, algorithms, and applications*," John Wiley & Sons, Inc, 1999.

[185] S. Mallat, W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol.38, no.2, pp. 617-643, 1992.

[186] Marr D, *Vision*, San Francisco: W.H.Freeman, 1982.

[187] R.J.Schalkoff, *Artificial Neural Networks*, Mc-Graw Hill Int. Editions, 1997.

[188] T.Kohonen, "The Self-Organizing Map", *Proceedings of the IEEE*, vol.78, no.9 pp 1464-1480, September 1990.

[189] P.J. Huber.,'Projection pursuit", *The Annals of Statistics*, 13(2):435-475, 1985.

[190] Haralick R. "Statistical and Structural approaches to Texture", *Proc. IEEE*, Vol. 67, No. 4, pp. 786-804, May 1979.

[191] R. W. Conners, M. M. Trivedi, and C. A. Harlow "Segmentation of a high resolution urban scene using texture operators", *CVGIP*, vol. 25, pp. 273-310, Mar. 1984.

[192] A. H. Mir, M. Hanmandlu and S. N. Tandon "Texture analysis of CT images", *IEEE Engineering in Medicine and Biology*, vol. 14, pp. 781-786, Nov./Dec. 1995 .

[193] C.M. Wu, Y.C. Chen and K.S. Hsieh, "Texture features for classification of ultrasonic images", *IEEE Trans. On Med. Imag.*, vol. 11, pp. 141-152, June 1992.

[194] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural features for image classification", *IEEE Trans. Syst. Man and Cybern.*, vol. SMC-3, pp.610-621,Nov. 1973.

[195] J. M. H. Du Buf, M. Kardan and M. Spann, "Texture feature performance for image segmentation", *Pattern Recognition*, vol. 23, pp. 291-309, 1990.

[196] F. Argenti, L. Alparone, and G. Benelli, "Fast algorithms for texture analysis using co-occurence matrices", *IEE Proc.*, vol. 137, pp. 443-448, Dec. 1990.

[197] J.S.Weszka, C.R.Dyer, and A. Rosenfield ,"A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man,Cybern.*, vol. SMC-6, pp. 269–285, Apr. 1976.

[198] R.W.Conners and C.A.Harlow, "A theoretical comparison of texture algorithms", IEEE *Transactions on Pattern Analysis and Machine Intelligence*, no.2, pp. 204-222,1980.

[199] R.W.Conners, M.M.Trivedi, and C.A.Harlow, "Segmentation of a high resolution urban scene using texture operators", *Computer vision, Graphics and Image Processing*, vol.25, pp.273-310, 1984.

[200] D.E.Rumalhart , J.L. McClelland, and the PDP Research Group, *Parallel Distributed Processing, vol.1 Foundations*, Cambridge MA and London England:M.I.T.Press, 1986, ch.8.

[201] D.F.Specht and P.D.Shapiro, "Generalization accuracy of probabilistic neural network compared with backpropogation network," in *Proc. JCNN'91*, pp. 458-461, july 1991.

[202] D. F. Specht , "Probabilistic Neural Networks," *Neural Networks*, vol. 3, pp.109–118, 1990.

[203] P.D.Wasserman, *Neural Computing: Theory and Practice* ,Van Nostrand Reinhold, NewYork, 1989.

[204] M.J.T.Smith &S.L.Eddins ,"Analysis/synthesis techniques for subband image coding," *IEEE Trans. on Acoust., Speech, and Signal Proc* ., vol.38, pp. 1446-1456,Aug. 1990.

[205] M. Vetterli, "Multi-dimensional subband coding: some theory and algorithms," *Signal Processing*, vol 6, pp. 97-112, Apr.1984.

[206] J.M. Woods and S.D. O'Neil, " Sub band coding of images", *IEEE Trans. On Acoust., Speech,and Signal Proc* ., vol.ASSP(34) ,pp. 1278-1288,Oct. 1986.

[207] R.H.Bamberger , "A method for image interpolation based on a novel multirate filter bank structure and properties of human visual system", i*n Proc.1992 SPIE/IS&T Symp. Electron Imaging sci. Technol*, pp 351-362, Feb 1992.

[208] R.H.Bamberger & M.J.T.Smith, "A filter bank for the directional decomposition of images: Theory and design", *IEEE Trans. on Signal Proc.*, vol.40, pp 882-893, Apr. 1992.

[209] R.H.Bamberger & M.J.T.Smith, "A multirate filter bank approach to the detection and enhancement of linear features in images", *in. Proc.IEEE ICASSP*, pp. 2557-2560, May 1991.

[210] R.L.Joshi, H.Jafarkhani, J.H.Cosner, T.R.Fischer, N.Farwardin, M.W,Marcellin & R.H.Bamberger , "Comparison of different methods of classification in subband Coding of images," *IEEE Trans. on Image Proc* ., vol.6, pp. 1473-1486, Nov 1997.

[211] M.G.Mini, V.P.Devassia, Tessamma Thomas, "Block Discrete Wavelet Transform for Real-Time Processing of Speech Signals", *in Proc.of the National Conference on Communications*, NCC-2002, Bombay, pp.374-378, Jan.2002.

[212] R.H.Bamberger,S.L.Eddins,and V. Nuri, "Generalized Symmetric Extension for Size-Limited Multirate Filter Banks ," *IEEE Trans. on Image Proc* .,vol.3, No.1, pp. 82-87, January 1994.

[213] V. P. Devassia, M. G. Mini, and T. Thomas, "An Efficient Algorithm for Time-Domain Computation of Discrete Wavelet Transform," *Knowledge-Based Intelligent Information Engineering & Allied Technologies,* Ed. N. Baba, L.C. Jain, and R. J. Howlett, IOS Press, Netherlands, pp.892-895, 2001.

[214] H.Yoshida, W.Zhang, W.Cai, K.Doi, R.M.Nishikawa, M.L.Giger, "Optimizing Wavelet Transform based Supervised learning for detection of micro calcification in digitized mammograms," ," *IEEE Proc. Int. Conf. On Image Processing,* ICIP-95, pp 152-155, 1995.

[215] R. O. Duda and P. E. Hart, *Pattern Recogntion and Scene Analysis*. New York, John Wiley, 1973.

[216] J. Illingworth and J. Kittler, "Survey of the Hough transform," *Computer Vision,Graphics, and Image Processing,* vol. 44, no. 1, pp. 87-116, October 1988.

[217] R. Lo and W. Tsai"Gray-scale Hough transform for thick line detection in gray-scale images," *Pattern Recognition*, vol. 28, no. 5, pp. 647-661, 1995.

[218] C. Daul, P. Graebling, and E. Hirsch, "From the Hough transform to a new approach for the detection and approximation of elliptical arcs," *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 215-236, December 1998.

[219] R. A. McLaughlin, "Randomized Hough transform: improved ellipse detection with comparison," *Pattern Recognition Letters*, vol. 19, no. 3-4, pp. 299-305, March 1998.

[220] J. Y. Goulermas and P. Liatsis, "Genetically fine-tuning the Hough transform feature space, for the detection of circular objects," *Image and Vision Computing*, vol. 16, no. 9-10, pp. 615-625, July 1998.

[221] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, New Jersey, Prentice-Hall, 1982.

[222] V. Kamat-Sadekar and S. Ganesan, "Complete description of multiple line segments using the hough transform," *Image and Vision Computing*, vol. 16, no. 9-10, pp. 597-613, July 1998.

[223] R. Nevatia and K. Babu, "Linear feature extraction and description," *Computer Graphics and Image Processing*, vol. 13, pp. 257-269, 1980.

[224] J. Basak, B. Chanda, and D. D. Majumder, "On edge and line linking with connectionist," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 3, pp. 413-428, March 1994.

[225] J.-W. Lee and I.-S. Kweon, "Extraction of line features in a noisy image," *Pattern Recognition*, vol. 30, no. 10, pp. 1651-1660, October 1997.

[226] S. Zucker, R. Hummel, and A. Rosenfield, "An application of relaxation labeling to line and curve enhancement," *IEEE Transactions on Computers*, vol. C-26, no. 4, April 1977.

[227] Y. Kanazawa and K. Kanatani, "Optimal line fitting and reliability evaluation," *IEICE Transactions on Information and Systems*, no. 9, pp. 1317-1322, September 1996.

[228] R. C. Nelson, "Finding line segments by stick growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 5, pp. 519-523, May 1994.

[229] A.-R. Mansouri, A. S. Malowany, and M. D. Levine, "Line detection in digital pictures: a hypothesis prediction/verification paradigm," *Computer Vision,Graphics, and Image Processing*, vol. 40, no. 1, pp. 95-114, October 1987.

[230] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, pp. 425-455, July 1986.

[231] S.Liu, "The analysis of digital mammograms: Spiculated tumour detection and normal mammogram characterization, " *Ph.D.Thesis* ,School of Electrical and Computer Engineering, Purdue University, May 1999.

[232] J. Suckling et al. "The Mammographic Image Analysis Society Digital Mammogram Database" *Exerpta Medica. International Congress Series 1069*, pp 375-378, 1994.

[233] Shao, J. , "Linear model selection by cross-validation," *J. of the American Statistical Association*, 88, pp.486-494. 1993

[234] Stone, M. , "Asymptotics for and against cross-validation," *Biometrika*, vol.64, pp.29-35, 1977.

[235] F.P Roth, "Bringing out the best features of expression data", *Genome Research (Insight/Outlook)*. Vol.11, no.11, pp. 1801-1802, 2001.

[236] Xiong, M., Fang, X., and Zhao, J. "Biomarker Identification by Feature Wrappers".*Genome Res.* No. 11, pp. 1878–1887, 2001.

[237]. Lester R G, " The contribution of radiology to the diagnosis, management, and cure of breast cancer", *Radiology*, 151: 1. 1984

[238]. Smith R A , "Epidemiology of breast cancer categorical course in physics, Tech. Aspects Breast Imaging", *Radiol. Soc. N. Amer.*, pp.21–33, 1993.

[239] Lo B C, Chan H P, Lin J S, Freedman M T, Mun S K , Artificial convolution neural network for medical image pattern recognition, *Neural Netw.* , No.8, pp. 1201-1214, 1995.

[240] Wiesel T N , "Post natal development of the visual cortex and the influence of environment", *Nature*, 229(5883), pp. 583-591,1982.

[241] Qian W, Chian W, Clarke L P, Kallergi M, Li H D, Velthuizen R L, Clarke R A, Silbigier M L , "Tree structured non-linear filter and wavelet transform for microcalcification segmentation in mammography", *in Biomedical Image Processing and Biomedical Visualization,* R.S.Acharya and D.B.Goldgof, Eds. *Proc. S*PIE, pp.509-520, 1993.

[242] Yoshida Y, Doi K, Nishikawa R.M, Ema T, and Zhang W , " Automated detection of clustered microcalcifications in digital mammograms using wavelet processing techniques" , *in Medical Imaging. Proc. SPIE 2167*, pp. 868-886, 1994.

[243] Shirai Y , *Three Dimensional Computer Vision*, Springer-Verlag, 1995.

[244]. Kiran Kumar P, Sukhendu Das, and Yegnanarayana B , " One Dimensional Processing for Edge detection using Hilbert Transform", *in Proc. of the Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2000)*, pp.25-31,2000.

[245] Nishikawa R.M. and Yaffe M.J, "Signal to noise properties of mammographic film- screen systems, *Med. Physics*, no.12, pp 32-39,1985.

[246] Bunch C, Huff K.E and Metter R.V., "Sources of noise in high-resolution screen-film radiography, P*roc. SPIE*, vol. 626, pp.63-71, 1986.

[247] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8,no. 4, pp. 425-455, July 1986.

List of Publications

International journal/Conference Papers

1. **M.G. Mini** and Tessamma Thomas, “A Neural Network Method for Normal / Abnormal Classification of Digitized Mammograms”' *Proceedings of IASTED Int.Conf on Neural Network and Computational Intelligence* (NCI 2004), Grindelwald, Switzerland ,Feb. 22 - 25, 2004.

2. **M.G. Mini**, V.P. Devassia and Tessamma Thomas, “Low-Memory Block Discrete Wavelet Transform for Image Processing,” *AMSE Journal on Signal Processing and Pattern Recognition.* (In Press)

3. **M.G. Mini**, V.P. Devassia and T. Thomas, “Block Discrete Wavelet Transform for Real-time Processing of One-Dimensional signals,” *AMSE Journal on Signal Processing and Pattern Recognition.* (In Press)

4. **M.G. Mini**, V.P. Devassia and T. Thomas, “Multiplexed Wavelet Transform Technique for Detection of Microcalcification in Digitized Mammograms”, *Journal of Digital Imaging.*(In Press)

5. **M.G. Mini** and Tessamma Thomas, “A Neural Network Method for Mammogram Analysis Based on Statistical Features” *Proc. of the IEEE TENCON-2003*, Bangalore, Oct.2003.

6. **M.G. Mini**, V.P. Devassia and Tessamma Thomas, “Detection of Microcalcification in Digitized Mammograms using wavelet Transform Local Extrema” *Journal of Digital Imaging*, Vol 16, Supplement1, pp 8-10, July 2003.

7. **M.G. Mini**, V.P. Devassia and Tessamma Thomas “Detection of Microcalcification in Digitized Mammograms using wavelet Transform Local Extrema”, *Proceedings of SCAR-2003*, Boston, June 2003.

8. **M.G. Mini**,V.P. Devassia and Tessamma Thomas, " One-Dimensional processing for edge detection in Biomedical Images using Pitch-Synchronous Wavelet Transform." Accepted in the *Int. Conf METMBS-02*, Las Vegas USA, June 24-27, 2002.

9. **M.G. Mini**,V.P. Devassia and Tessamma Thomas, "Block Discrete Wavelet Transform for Real-time Processing of One-Dimensional signals," Proceedings of the International *Conference on Business Decision &Technology, BDT-02*, Kuwait vol.1, pp 155-162, March 18-20, 2002.

10. Deepa J., **M.G. Mini**, Tessamma Thomas , 'A Robust Method for Blind Digital Image Watermarking using Wavelet Techniques' Accepted for presentation in the IEEE Int . Conference ITPC_03, Nepal, May 2003.

11. V.P. Devassia, **M.G. Mini** & Tessamma Thomas, 'Novel computational structure for real time wavelet analysis' *Proceedings of the SPIE 2003 Conference in Modeling, Signal Processing, and Control*, San Diego, March 2003.

12. V.P.Devassia, **M.G.Mini** , Tessamma Thomas, "A Novel Parallel Structure for Computation of Discrete Wavelet Transform of images" , *AMSE Journal on Computer Science and Statistics* vol. 7, no.3, pp 25-40, 2002.

13. V.P. Devassia, **M.G. Mini** & Tessamma Thomas, ' Efficient Parallel Structure for Wavelet Transform Based Processing of Biomedical Images', Proceedings of the *Int. Conf METMBS-02*, Las Vegas USA,June 2002.

14. V.P. Devassia, **M.G. Mini** & Tessamma Thomas, ' PSWT Based Predictive Coding and Compression of ECG Signals' Proceedings of *the Int. Conf METMBS-02*, Las Vegas USA,June 2002.

15. V.P. Devassia, **M.G. Mini** & Tessamma Thomas, ' Wavelet Transform Based Classification Of Vocal Music Into Voiced, Unvoiced, Silent And Transition Regions 'Proceedings of the *Int. Conf MS-02*,Spain, June 2002.

16. V.P.Devassia, **M.G.Mini**& Tessamma Thomas, "Feature Enhancement Using PSWT Techniques:A Study On Instrumental Music", *Proceedings of the International Workshop on Linear Algebra , Numerical Functional analysis, and Wavelet Analysis*, Eds. S.H.Kulkarni,M.N.N. Namboodiri, India, pp. 65-77, Aug.6-15, 2001.

17. V.P.Devassia, **M.G.Mini** & Tessamma Thomas, 'An Efficient Algorithm for Time-Domain Computation of Discrete Wavelet Transform,' Accepted for presentation in the *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Utah, May 7-11, 2001.

National Journal/ Conference Papers

18. **M.G. Mini**,V.P. Devassia and Tessamma Thomas, "A comparative study on the application of wavelet based techniques for microcalcification detection in mammograms', Proceedings of *Nat. Conf. On Communications (NCC-2003*), IIT, Chennai,pp 601-605, Feb.2003.

19. **M.G Mini**, V.P Devassia and Tessamma Thomas, "In-Place Computation of Block Discrete Wavelet Transform", *IETE journal of Education*, vol.43, No.2 pp 79-82 April-June 2002.

20. **M.G.Mini**, V.P.Devassia and Tessamma Thomas, 'Block Discrete Wavelet Transform for Real-Time Processing of Speech Signals,' Proc. of the *Nat. Conf. On Communications (NCC-2002)*, pp 374-378, Jan.2002.

21. V.P.Devassia, **M.G.Mini** & Tessamma Thomas, " Instrument –Dependent Noise Suppression Using PSWT", J. Acoust. Soc. of India, vol. 30, pp 164-167, 2002.

22. V.P.Devassia, **M.G.Mini** & Tessamma Thomas, "Application of Pitch-Synchronous Wavelet Transform In The Analysis Of Guitar Music," J. Acoust. Soc. of India, vol. 29, No 2-4, pp 296-302, 2001.

Index

Anatomy of the female breast, 34

Architectural Distortion, 51

Artificial Neural Networks Technology, 13, 76

- for classification, 91
- Learning Laws, 84
- Learning Rates, 84
- Selection of structure, 138, 145
- Teaching, 82
- Training and Testing , 136

Artificial Neurons, 78

Asymmetric Breast Tissue, 52

BPNN, 91

BDWT by Overlap Save Method

- 1-D, 106
- 2-D, 109

BIDWT by Overlap Add Method

- 1-D, 105
- 2-D, 112

Biopsy, 38

Block –wise Computation

- 1-D DWT, 101
- 2-D DWT, 107

Breast Composition Determination, 40

CAD in mammography, 20

Calcifications Distribution Modifiers, 47

Circumscribed Masses, 48

Classification of microcalcifications into benign and malignant, 25

Competitive network, 93

Computation

- DWT, 9,29, 63
- 2-D DWT, 68

Computational Complexity, 113

Computer Assisted Mammography, 11

Computer Assisted Tomography , 12

Continuous wavelet transform, 7,57

Decimation, 63

Delta Rule, 85

Detection and removal of linear markings, 132

Detection Criteria, 137, 161

Digital Image Processing, 2

Discrete wavelet transform 8, 58

Edge detection, 74

- Wavelets, 74

    MWT, 154
Estimation of computational burden
    BDWT algorithm, 114
    Standard algorithm, 113
Feature Extraction for classification, 86
Feature selection, 140, 146
Features of Normal Mammograms, 129
Functional Image analysis, 12
Gradient Descent Rule, 85
Hebb's Rule, 84
History of Wavelets, 7
Hopfield Law, 84
Image compression, 3
Image description and Representation, 4
Image enhancement, 2
Image Restoration, 3
Image segmentation, 3
Interpolation, 64
Kohonen's Learning Law, 85
Kurtosis, 87
Line Detection Algorithm, 171
Magnetic Resonance Imaging, 12
Malignancy in the breast, 35
Mammographic Abnormalities, 43
Mammography, 38
Mammography Machine, 39
Mean, 86
Medical Image Processing, 4
Microcalcifications, 43
Microcalcification Detection
    after Classification, 160
    as an edge detection operation, 154
    MWT based, 155
MRA, 60
Multirate Operations, 63
Multiplexed wavelet transform, 8,71
n-fold Cross Validation, 138
Normal Mammograms, 41
    Characterization, 27,128,
Normal/abnormal classification
    Statistical and textural features, 149
    Statistical features, 138
    Textural features, 145
PNN, 95
Pyramidal algorithm, 66
RBFN, 94
Removal of normal background, 130
Residual Image generation, 130
Sampling rate conversion, 64
Sectioned computation, 10

SGLD Matrix, 88
Skewnesss, 86
Spiculated Lesions, 52
Statistical descriptors, 86
Supervised Learning, 82
Symptoms & Diagnosis, 36
Target detection, 15
Textural features, 88
Training and [illegible]
Truncations of transform coefficients, 103
Unsupervised Learning, 83
Variance, 87
Wavelets, 55
2-D Wavelets, 68
Wavelet Transform, 6
  CWT, 7,57
  DWT 8, 58
  in Biomedical Image Processing, 10
  in Two Dimensions, 9
  MWT, 8,71
2-D Wavelet Transform, 69