R
857
06
K42
2006

MULTIRESOLUTION ANALYSIS FOR CLASSIFICATION AND COMPRESSION OF MEDICAL IMAGES

by

April Ellahe Khademi, B.Eng
Ryerson University, 2004.

A thesis
presented to Ryerson University
in partial fulfillment of the
requirement for the degree of
Master of Applied Science
in the Program of
Electrical and Computer Engineering.

Toronto, Ontario, Canada, 2006

© April Khademi, 2006

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

UMI Number: EC53503

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform EC53503
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature _____

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature _____

Instructions on Borrowers

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

Multiresolution Analysis for Classification and Compression of Medical Images

© April Khademi, 2006

Master of Applied Science
Department of Electrical and Computer Engineering
Ryerson University

This thesis contains the application of the DWT for classification and compression of biomedical images (mammogram, small bowel and retinal). The shift-invariant DWT and several textural descriptors were used to provide scale, translation and semi-rotational (RST) invariant features. The features were classified using LDA with the leave one out method to combat small database sizes. The small bowel images achieved a classification rate of 75% and is the first reported work in the area, the retinal images achieved 81% classification rate and the mammograms achieved a rate of 59%. The success of the system is due to the RST-invariant features which accounted for various sized masses, different camera angles and textural differences between pathologies. Any failures are a result of overlapping tissues which masked the pathologies. JPEG 2000 was the wavelet-based compressor used and it was compared to JPEG-LS, LJPEG, adaptive Huffman, arithmetic and LZW codes. For 12bpp mammograms, JPEG 2000 offered the best compression (CR of 9.319 and R of 1.288bpp), but suffered from slow compression speeds (501.3 Ksymbols/s). Compression was investigated solely for mammograms since they were the only images stored in raw formats.

Acknowledgments

I would like to acknowledge my parents and family for their continual support and encouragement, which has given me the ability to pursue my goals with confidence and full determination. Additionally, the knowledgeable guidance and moral support that Dr. Krishnan has patiently provided me with over the years has helped me to succeed beyond my wildest dreams. I am forever grateful to him for this. I would also like to thank my fellow colleagues from the Signal Analysis Research (SAR) group, especially Danoush Hosseinzadeh and Karthi Umapathy for the enlightening and informative discussions they had with me about my research.

I also would like to thank the Ontario Government and Ryerson University for awarding me the OGS and OGSST scholarships, which allowed me to pursue Master's level graduate research with full financial support. In addition, I am grateful to NSERC for awarding me the NSERC CGS-D3 scholarship for the pursuit of PhD studies in the area of biomedical image analysis.

Last but not least, I would like to thank everyone at Ryerson University who provided me with an enhanced educational experience, in which I will never forget.

“When the student is ready, the master appears”

- *Buddha*

Contents

1	Introduction	1
1.1	Biomedical Imagery	2
1.2	Biomedical Image Analysis	4
1.2.1	Nonstationary Behaviour	4
1.2.2	Nonstationary Image Analysis Tools	6
1.3	Experimental Images	15
1.3.1	Mammogram Images	16
1.3.2	Small Intestine Images	19
1.3.3	Retinal Images	20
1.4	Motivation	23
1.5	Outline of Thesis	23
2	Multiresolutional Analysis (MRA)	26
2.1	Introduction	26
2.2	Wavelets, Subspaces and Signal Decomposition	27
2.2.1	Scaling Function Subspaces	27
2.2.2	Wavelet Basis Functions	30
2.2.3	Wavelet Transforms	31
2.3	Discrete Wavelet Transform	32
2.3.1	2-D Extension for Images	33
2.3.2	5/3 Wavelet	35
2.4	Properties of the DWT	36
2.4.1	Scale Invariance	36
2.4.2	Shift-Variance	37
2.4.3	DWT for Image Analysis	43
3	Classification of Biomedical Images	48
3.1	Introduction	48
3.2	Feature Extraction Problem Formulation	52
3.2.1	Invariant Features	53
3.3	Shift-Invariant Discrete Wavelet Transform (SIDWT)	55
3.3.1	2-D SIDWT Algorithm	56

3.3.2	Experimental Proof	61
3.4	Feature Extraction for Biomedical Images	70
3.4.1	Image Texture	71
3.4.2	Texture Analysis	76
3.4.3	Multiresolutional Features	80
3.4.4	Spatial Features	83
3.5	Pattern Analysis/Classification	84
3.6	Results	86
3.6.1	Preprocessing	86
3.6.2	Block Diagram	88
3.6.3	Classification Results	88
4	Compression of Biomedical Images	92
4.1	Introduction	92
4.2	Image Model	94
4.3	Performance Measures	96
4.4	Source Encoder Requirements	96
4.5	Lossless Source Encoders for Mammograms	97
4.5.1	Adaptive Huffman	97
4.5.2	Arithmetic Codes	100
4.5.3	Lemple-Ziv Welch (LZW) Codes	105
4.5.4	Lossless-JPEG (LJPEG) Compression	108
4.5.5	JPEG-LS Compression	112
4.5.6	JPEG 2000 Compression	117
4.6	Compression Results	126
5	Conclusions	129
5.1	Discussion of Classification Results	129
5.1.1	Small Bowel Image Classification	129
5.1.2	Retinal Image Database	131
5.1.3	Mammogram Images	132
5.1.4	Overall Discussion on Classification Results	133
5.2	Discussion of Compression Results	134
5.3	Summary of Results	139
5.4	Future Work	139
A	Filters for DWT Decomposition	141
B	Discrete Wavelet Transform: Lifting Based Scheme	143
B.1	Split, Predict and Update	143
B.1.1	Splitting	144
B.1.2	Prediction	144
B.1.3	Updating	145

B.1.4 2-D Implementation	145
C Shift-Variance of the DWT	147
D Graphical User Interface	152
Publications	153
Bibliography	155

List of Figures

1.1	(a) CT image of a healthy heart [1], (b) image of small bowel with a polyp [2], (c) mammogram image of a breast containing a malignant, spiculated mass [3], (d) retinal image of an eye with central retinal artery occlusion and central retinal vein occlusion [4], (e) MRI image of a normal human head [5].	4
1.2	(a) Quasi-stationary image (256×256) containing a 40×40 square, (b) a nonstationary image.	6
1.3	Arbitrary time-frequency tiles in the time-frequency plane. Image is adapted from [6].	9
1.4	(a) The Fourier transform of the image $f(x, y)$ of Figure 1.2(a) revealing a two-dimensional $\text{sinc}(\cdot)$ function, (b) the Fourier transform of the Lena image $g(x, y)$ of Figure 1.2(b) which indicates there are directed edges within the image.	11
1.5	(a) Time domain representation of a 1-D sinusoidal basis function $\psi(t)$ with a frequency of f_1 , (b) frequency domain representation of $\psi(t)$ ($\Psi(f)$), (c) time-frequency tiles of a sinusoidal basis function set.	12
1.6	(a) Time domain representation of a delta dirac basis function $\psi(t)$ at the time instant t_1 , (b) frequency domain representation of $\psi(t)$ ($\Psi(f)$), (c) time-frequency tiles of a delta dirac basis function set.	12
1.7	(a) Time-frequency tiles of the 1-D STFT, (b) time-frequency tiles of a two-band 1-D wavelet transform.	14
1.8	(a) The discrete wavelet transform of the square image shown in Figure 1.2(a) using the (3,5) biorthogonal wavelet displaying the localization of horizontal, vertical and diagonal edge components and (b) the discrete wavelet transform of the Lena image shown in Figure 1.2(b) using the Meyer Wavelet where the directional edges associated with Lena are clearly isolated within the sub-bands. Some contrast enhancement was performed for visualization purposes.	16
1.9	Typical mass shapes. (Reproduced with the permission from H. Alto, R.M. Rangayyan, and J.E.L. Desautels, "Content-based retrieval and analysis of mammographic masses", <i>Journal of Electronic Imaging</i> , 2005, Vol. 14, No. 2. © SPIE.)	17
1.10	(a) Benign mass in a 1024×1024 mammogram, (b) malignant lesion in a 1024×1024 mammogram. Lesions are shown inside the square box.	18

1.11	Small bowel images (256×256) collected by the PillCam TM obtained from the Image Atlas of Given Imaging Ltd. See [2]. (a) Healthy small bowel, (b) normal jejunum, (c) normal colonic mucosa, (d) small bowel lymphoma, (e) jejunal carcinoma, (f) Kaposi's sarcoma.	21
1.12	Retinal images (700×605) supplied by the STare public database [4]. (a) Normal retinal image, (b) normal retinal image, (c) retinal image with background diabetic retinopathy, (d) retinal image with central retinal vein occlusion. . .	22
1.13	Thesis block diagram.	24
2.1	Nested wavelet and scaling signal spaces.	28
2.2	(a) Computing the 1-D wavelet and scaling coefficients using filtering and decimation with a 2-stage analysis filterbank, (b) corresponding decomposition tree showing the division of signal spaces.	33
2.3	Filterbank implementation of 2-D discrete wavelet transform (DWT).	34
2.4	Graphical depiction of wavelet coefficient placement for two levels of decomposition.	35
2.5	(a) DWT of $x(n)$, (b) DWT of $x(n-1)$. Image adapted from [7].	38
2.6	(a) Synthetic image $f(i, j)$ composed of a circle (radius = 60 pixels). Image has dimensions of 256×256	40
2.7	DWT decomposition of the image $f(i + \Delta i, j + \Delta j)$, where $(\Delta i, \Delta j) \in \{(0,0), (0,1), (1,0), (1,1)\}$	41
2.8	(a) Synthetic image $f(i, j)$ (256×256) containing an 80×80 square, (b) DWT of $f(i, j)$ in (a).	44
2.9	(a) Synthetic image $f(i, j)$ (256×256) containing a circle with a radius of 60 embedded in gaussian noise (mean = 0, variance = 1), (b) DWT of $f(i, j)$ in (a).	45
2.10	(a) Synthetic image $f(i, j)$ (256×256) containing a 50×100 rectangle, (b) DWT of $f(i, j)$ in (a).	46
3.1	Shift-invariant DWT decomposition tree for three decomposition levels. . . .	59
3.2	Best basis selection corresponding to the minimum cost path.	61
3.3	(a) Synthetic image $f(x, y)$ composed of a circle (radius = 20 pixels) and a square (width = 50 pixels). The image is 256×256	63
3.4	DWT decomposition of the image $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$. These are the four cosets.	64
3.5	SIDWT decomposition of the image $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$	65
3.6	The normalized histograms for the DWT coefficients computed for the images $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$	67
3.7	The cross-correlation sequence $\mathcal{R}_{(x_1, x_2)}$ between images $f(x + \Delta x, y + \Delta y)$, which are decomposed by the SIDWT for the circular translates $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$	70

3.8	Some examples of (quasi-) periodic texture images. All images are obtained from the USC Signal and Image Processing Institute. (a) Brodatz - brick wall (D94 H.E), (b) Brodatz - plastic bubbles (D112 H.E), (c) Brodatz - Straw (D15).	72
3.9	Some examples of aperiodic texture images which contain random texture regions. All images are obtained from the USC Signal and Image Processing Institute. (a) Tree, (b) mandrill, (c) sailboat on a lake.	73
3.10	Small bowel images captured by the PillCam TM SB, which exhibit textural characteristics. (a) Healthy small bowel, (b) normal neocecal valve, (c) normal colonic mucosa, (d) normal small bowel, (e) normal jejunum, (f) small bowel polyp, (g) small bowel lymphoma, (h) GIST tumor, (i) polypoid mass, (j) small bowel polyp.	74
3.11	Retinal images which exhibit textural characteristics. (a)-(c) Normal, homogeneous retinal images, (d) background diabetic retinopathy (dense, homogeneous yellow clusters), (e) macular degeneration (large, radiolucent drusens with heterogeneous texture properties), (f) central retinal vein occlusion (oriented, radiating texture).	75
3.12	Manually segmented 128×128 regions of a mammogram. (a)-(c) Normal regions, (d)-(f) circumscribed benign masses, (g)-(i) spiculated malignant masses.	77
3.13	Histogram equalized version of Figure 3.12	78
3.14	(a) Test image, (b) DWT of (a).	81
3.15	System block diagram for the classification of the retinal and small bowel images.	89
3.16	System block diagram for the classification of mammogram images.	89
4.1	Zoomed in region (14×14) of a mammogram demonstrating pixel correlation.	95
4.2	Dynamic Huffman trees of (a) Algorithm FGK, (b) Algorithm Λ , both operating on sequence $\{fgh\}$	99
4.3	Components of a traditional arithmetic compressor.	101
4.4	Arithmetic coding: interval subdivision process for the encoding of <i>bacc</i> for a complete alphabet $\{a, b, c, d, e, f\}$. Initial probabilities and ranges are listed in Table 4.1.	104
4.5	LJPEG block diagram.	109
4.6	LJPEG context.	109
4.7	JPEG-LS encoder block diagram.	113
4.8	JPEG-LS: contextual template used for prediction, context and probability model generation.	114
4.9	The block diagram of JPEG 2000 for the compression of mammogram images.	117
4.10	Tiling of an image into 20 non-overlapping rectangles.	119
4.11	Placement of DWT coefficients for three levels of decomposition and the corresponding mammogram output.	121

4.12	Packet partition location (precinct) and codeblocks of a single tile or for an entire image.	122
4.13	The components of a binary adaptive arithmetic encoder.	123
4.14	Organization of packets, layers, codeblocks and code stream	125
4.15	Compression ratios CR and bitrates R for 100 mammograms. Note that the x axis corresponds to a particular coder: 1 - Adaptive Huffman, 2 - Arithmetic Order 0, 3 - Arithmetic Order 1, 4 - LZW 12bit, 5 - LZW 15bit, 6 - LJPEG, 7 - JPEG-LS, 8 - JPEG 2000.	128
4.16	Compression time/image and compression rate μ for 100 mammograms. Note that the x axis corresponds to a particular coder: 1 - Adaptive Huffman, 2 - Arithmetic Order 0, 3 - Arithmetic Order 1, 4 - LZW 12bit, 5 - LZW 15bit, 6 - LJPEG, 7 - JPEG-LS, 8 - JPEG 2000.	128
A.1	The frequency domain representation of a two level decomposition using the DWT.	142
B.1	Generalized 1-D lifting based implementation of the DWT.	144
B.2	Lifting-based implementation of the DWT for two dimensional signals. . . .	146
D.1	Graphical user interface created for the classification of biomedical imagery. .	152

List of Tables

2.1	Analysis and synthesis filter coefficients for the 5/3 wavelet.	36
2.2	Mean μ and variance σ^2 of the DWT coefficients of the LH band, for each DWT image shown in Figure 2.7.	42
3.1	Related works in retinal image feature extraction and classification.	50
3.2	Related works in mammogram image feature extraction and classification.	51
3.3	Translation parameters of the SIDWT for various translates $(\Delta x, \Delta y)$ of the input $f(x, y)$	66
3.4	Mean μ and variance σ^2 of the SIDWT output for translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$	68
3.5	% Energy ε per subband (1 level) of the SIDWT output for translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$	68
3.6	Correlation coefficient $\rho(X_1, X_2)$ *for each combination of two SIDWT outputs X_1 and X_2 . The SIDWT outputs are computed from the translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$. Note that X_1 and X_2 correspond to the SIDWT representations achieved by some translate, as dictated by $(\Delta x, \Delta y)$	69
3.7	The graylevel co-occurrence matrices (GCM) computed for each level of decomposition.	82
3.8	A list of features extracted from each subband for each decomposition level j	83
3.9	Characteristics of the biomedical images used for experimentation.	87
3.10	Confusion matrix containing the number of correctly classified small bowel images as either normal or abnormal.	90
3.11	Confusion matrix containing the number of correctly classified retinal images as either normal or abnormal.	90
3.12	Confusion matrix containing the number of correctly classified mammogram masses as either benign or malignant.	90
4.1	Symbol probabilities and initial encoding ranges assigned for the symbol alphabet $\{a, b, c, d, e, f\}$. The values of this table are used in conjunction with the arithmetic coding example shown in Figure 4.4.	103

4.2	The LZW dictionary which would be dynamically constructed for the encoding of <i>aabacaaaba</i> is on the left. Note that the original alphabet symbols $\{a, b, c\}$ would be initialized in the dictionary prior to any encoding. The right table contains The LZW coded output for <i>aabacaaaba</i> using the dictionary shown on the left side.	106
4.3	The seven predictor functions as specified by the LJPEG standard.	110
4.4	Average compression results for 100 mammograms.	127

List of Acronyms

1-D	1-Dimensional
2-D	2-Dimensional
CAD	Computer-Aided Diagnosis
CBIR	Content-Based Image Retrieval
CT	Computed Tomography
DCT	Discrete Cosine Transform
DSP	Discrete Signal Processing
DWT	Discrete Wavelet Transform
EBCOT	Embedded Block Coding with Optimized Truncation
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EZW	Embedded Zero Wavelet
GCM	Graylevel Co-occurrence Matrices
IS	International Standard
ISO	International Organization for Standardization
ITU	International Telecommunications Union
JPEG	Joint Photographic Experts Group
LDA	Linear Discriminant Analysis
LJPEG	Lossless JPEG
LOCO-I	Low Complexity Lossless Compression for Images
LOOM	Leave One Out Method
LZW	Lemple-Ziv Welch
MP	Matching Pursuit
MRA	Multiresolutional Analysis
MRI	Magnetic Resonance Imaging
OSGD	One-Sided Geometric Distribution
PDF	Probability Density Function
PET	Positron Emission Tomography
PSD	Power Spectral Density
RLE	Run-Length Encoder
ROI	Region of Interest
RST	Rotation, Scale and Translation
SIDWT	Shift-Invariant Discrete Wavelet Transform
SPECK	Set Partitioned Embedded Block Coder
SPIHT	Set Partitioning Into Hierarchical Trees
STFT	Short-Time Fourier Transform
TSGD	Two-Sided Geometric Distribution

Chapter 1

Introduction

DOCTORS and other healthcare professionals receive extensive training in order to be sufficiently qualified to detect and prevent diseases. Although the skills acquired by these medical facilitators are quite extensive, it is just as important for them to have access to an assortment of technologies to further improve their diagnostic capabilities. In fact, medical instruments provide healthcare workers with one of the primary sources of information needed for disease discovery.

Electrodes and signal acquisition technology can gather a variety of biomedical signals such as electrocardiograms (ECG), electroencephalograms (EEG) and electromyogram (EMG) signals, which record the electrical activity of the heart, brain and muscles, respectively. These type of one-dimensional (1-D) signals show the physiological behaviour of an organ or set of organs, based on a quantity which varies over time.

To get a visual representation of the patient's anatomy, different forms of acquisition technologies can capture radiological images of bones, vessels and soft tissue. This allows physicians to examine the physical state or function of internal structures for diagnostic purposes. Some common imaging modalities are x-ray, computer-assisted tomography (CT), nuclear medicine, medical resonance imaging (MRI), positron emission tomography (PET) and ultrasound. The radiological images generated from these modalities are stored in a 2-dimensional (2-D) representation, where the image's intensity is displayed as a function of its spatial dimensions.

As technology advances and the use of computing devices becomes common place in hospitals and clinics, it is only natural for the traditional imaging modalities to evolve to digital form. In hospital systems today, if digitized images are used, the medical images are either captured and directly stored in digital format, or the original filmed image is scanned and quantized to gain a digital representation [8]. In either method, the digital images must be high resolution to ensure that even the smallest irregularities can be represented. According to the American College of Radiology, a minimum of 8 bits per pixel is permissible [8].

1.1 Biomedical Imagery

Medical imaging is important for numerous reasons. Radiological images can give radiologists insight into the functionality or condition of the patient's anatomy. For instance, MRI has been shown to be successful in the detection, evaluation and assessment of tumours of the musculoskeletal system [9]. A mammogram, or a x-ray image of the breast, can show any abnormal growths or changes in the breast tissue, even before it is palpable in a breast examination [10]. Retinal imaging uses an optical camera to show the optic nerve, fovea, surrounding vessels, and the retinal layer in order to diagnose diseases of the human eye [4]. All these images give the physician a tool which helps them to differentiate between normal and abnormal conditions. Furthermore, this diagnostic tool is non-invasive. The physician does not have to perform invasive procedures to gain a preliminary understanding of the internal structure of the patient's anatomy.

A particular branch of medicine is known as interventional radiology and its success is very dependant on radiological imaging technologies. Interventional radiology diagnoses and treats disease by inserting small needles, guidewires, and catheters into the skin and is guided by x-ray, ultrasound, or another form of radiologic imaging [11]. Any image-guided, interventional procedure is minimally invasive and performed by a highly skilled and trained doctor.

An example of an interventional procedure is called angioplasty. A patient undergoes angioplasty when he or she has a blocked or narrowed blood vessel which constricts the blood

flow. Without having to undergo painful and invasive surgery, the interventional radiologist can insert a small balloon into the artery using a catheter. Once the balloon is in the right position using x-ray guidance, the balloon can be inflated to open up the artery, allowing an increase in blood flow.

Many biomedical images collected in hospitals and clinics are being acquired in or converted to a digital format. Consequently, it is possible to apply digital signal processing (DSP) algorithms to these images to arrive at an entirely new series of beneficial applications. For instance, a compression scheme which uses a transformation to exploit the correlated statistical properties of a biomedical image can reduce the file size enough to warrant efficient transmission in a telemedicine-based application or efficient storage in a biomedical imaging archival system. Moreover, automated feature extraction algorithms can: (1) supply diagnostic information to the physician, ultimately providing a second-opinion of the patient's health and (2) be used to archive and retrieve the images based on their content. This type of Computer-Aided Diagnosis (CAD) system or Content-Based Image Retrieval (CBIR) scheme may use texture analysis to extract meaningful features.

A recent study indicates that digital mammograms aided in the detection of disease more readily than screen-film mammography, when tested on patients with dense breasts, women under 50 years of age, and pre/perimenopausal women [12]. This demonstrates the benefits of using digital mammography versus the traditional method of filming. Other benefits include digital zooming and image contrast enhancement capabilities, computer-aided diagnosis systems, telemammography, efficient database storage and the ability to immediately view the image after collection, thereby reducing the number of repeats [13]. Another important advantage comes from the digital acquisition technology itself; patients are exposed to smaller doses of radiation during the mammogram acquisition process [14]. Although these advantages are explicitly described for digital mammography, many other imaging modalities also benefit from the same applications as a result of DSP [8].

Many research efforts are dedicated to finding new and improved biomedical imaging applications since these images have moved to digital form. These applications allow radi-

ologists to have increased diagnostic capability, easy image management, as well as many other features and advantages. Consequently, it is necessary to investigate theories, signal processing methodologies and image analysis techniques for biomedical images, which will give way to beneficial applications. Some sample medical images are included in Figure 1.1.

Figure 1.1: (a) CT image of a healthy heart [1], (b) image of small bowel with a polyp [2], (c) mammogram image of a breast containing a malignant, spiculated mass [3], (d) retinal image of an eye with central retinal artery occlusion and central retinal vein occlusion [4], (e) MRI image of a normal human head [5].

1.2 Biomedical Image Analysis

To realize the medical imaging applications previously discussed, signal processing algorithms are needed to analyze and understand the images. Image analysis techniques can exploit the intrinsic properties of the raw image data, which allows a specific function to be performed on the images. Therefore, an analysis tool which is tuned to the image characteristics should be selected, in order to bring out the features of interest.

1.2.1 Nonstationary Behaviour

All signals and images may be categorized into one of two categories: 1) deterministic or 2) non-deterministic (random). Deterministic signals allow for advanced prediction of signal quantities, since the signal may be described by a mathematical function. In contrast, instantaneous values of non-deterministic signals are unpredictable due to their random nature and must be represented using a probabilistic model [15]. This stochastic model gives way to statistical quantities which describe the inherent behaviour of the signal or image in

question.

Random signals (both 1-D and 2-D) may be further classified into two groups: 1) stationary or 2) nonstationary. A stationary signal (1-D) is a signal which has a constant probability distribution for all time instants. As a consequence, first order statistics such as the mean and second order statistics such as variance must also remain constant. In contrast, a nonstationary signal has a time-varying probability distribution which causes other quantities which rely on the probability density function (PDF) to also be time-varying. For instance, the mean, variance and autocorrelation function of a nonstationary signal would also change with time, since they are computed from the PDF of a signal. Since the Fourier transform of the autocorrelation function is equal to the power spectral density (PSD) of a signal (which is related to the spectral content), the PSD of a nonstationary signal is also time-varying. Consequently, a nonstationary signal has time-varying spectral content.

For two dimensional signals (images) which possess nonstationary characteristics, the frequency changes with respect to its vertical and horizontal spatial dimensions. Therefore, the word “*nonstationary*” will strictly be associated to an image with spatially-varying frequency components. An example of a quasi-stationary image and a nonstationary image is shown in Figure 1.2. The image shown in Figure 1.2(a) may be considered quasi-stationary since large regions of the image have spatially-constant frequency content (for example, the black background). However, for all purposes, Figure 1.2(a) will be considered nonstationary since it contains spatially-changing frequency content when viewed as a whole.

It is important to see if biomedical images can be classified as nonstationary, since a specialized set of image analysis tools may be selected to investigate the nonstationary behaviour. Each individual radiological image in Figure 1.1 has regions which contain both coarse or fine texture patterns (high frequency), as well as regions with slowly varying or constant pixel values (low frequency). Since the frequency content is varying from region to region, biomedical images can be considered nonstationary, therefore requiring nonstationary image processing tools for analysis. In fact, most natural images are nonstationary.

Figure 1.2: (a) Quasi-stationary image (256×256) containing a 40×40 square, (b) a nonstationary image.

1.2.2 Nonstationary Image Analysis Tools

Biomedical images are nonstationary since they have spatially-varying frequency components. Many medical imaging applications which use DSP specifically rely on techniques and operations which identify these space-frequency structures. Therefore, an optimal image analysis tool should be identified and used to localize the features of interest.

Signal decomposition techniques can be used to transform the images into a different representation to highlight features of interest. A decomposition technique linearly expands a signal or image using a set of mathematical functions. This type of transformation is the primary interest of this thesis and the concept is explained below.

Using a set of real-valued, expansion coefficients a_k , and a series of 1-D mathematical functions $\psi_k(t)$ known as an expansion set ($\psi_k(t) = \psi(t - k)$ for all integer values of k), a 1-D signal $f(t)$ may be expressed as a weighted linear combination of these functions

$$f(t) = \sum_k a_k \cdot \psi_k(t), \quad k \in \mathbf{Z}. \quad (1.1)$$

If the members of the expansion set $\psi_k(t)$ are orthogonal to one another:

$$\langle \psi_k(t), \psi_l(t) \rangle = 0, \quad k \neq l, \quad (1.2)$$

then it is possible to find the expansion coefficients a_k

$$a_k = \langle f(t), \psi_k(t) \rangle, \quad (1.3)$$

where the inner product $\langle \cdot \rangle$ of two signals $x(t)$ and $y(t)$ is defined by

$$\langle x(t), y(t) \rangle = \int_t x^*(t) \cdot y(t) dt. \quad (1.4)$$

The definition of an expansion set depends on various properties. For instance, if there is a signal $f(t)$ which belongs to a subspace S ($f(t) \in S$), then $\psi_k(t)$ will *only* be called an expansion set for S if $f(t)$ can be expressed with linear combinations of $\psi_k(t)$. The expansion set forms a *basis* if the representation it provides is unique [16]. Similarly, a basis set may be defined first and then the space S *spans* all functions $f(t)$ which can be expressed by $f(t) = \sum_k a_k \cdot \psi_k(t)$.

For images, the basis functions may be a dependant on both the horizontal and vertical spatial variables (x, y) . This leads to 2-D basis functions $\psi_{m,n}(x, y)$, where $\psi_{m,n}(x, y) = \psi(x - n, y - m)$, for all $(m, n) \in \mathbf{Z}$. Therefore, a 2-D function $f(x, y)$ which belongs to the same space as the basis functions may be rewritten as a linear expansion using the following formula

$$f(x, y) = \sum_m \sum_n a_{m,n} \cdot \psi_{m,n}(x, y), \quad (1.5)$$

where $a_{m,n}$ are the 2-D expansion coefficients found by

$$a_{m,n} = \langle \psi_{m,n}(x, y), f(x, y) \rangle. \quad (1.6)$$

Using decomposition techniques, a new representation is generated (as in Equation 1.1 for 1-D signals and Equation 1.5 for 2-D images). To compute the coefficients a_k or $a_{n,m}$, it is necessary to choose basis functions appropriate for the application. For a biomedical imaging application that requires information about the nonstationary structures of the

images, choosing the appropriate basis functions is key in extracting relevant and useful information. The information obtained by such a transformation could be used in many different beneficial applications, including those discussed in Section 1.1.

Time-Frequency Tiles

In order to understand the optimality of a particular basis set for an application, time-frequency tiles may be used. For 1-D functions, time-frequency tiles provide a visual representation of a basis function's dual time-frequency characteristics. According to these properties, the tiles are placed in the time-frequency plane. To visualize the spatial-frequency relationship of the basis functions for 2-D images it is necessary to examine a high-dimensional plot, since any tiling representation would need 2 variables to represent the spatial dimensions and 2 variables to represent the frequency dimensions. To ease reading, explanation and visualization, only the 1-D case is considered here. Additionally in image processing applications, many 2-D transforms are realized by applying 1-D operations separately to rows and columns. Therefore, the 1-D behaviour of the basis functions can be used to analyze the effects they would have when acting solely on a row or column.

For a time-frequency tile, the time width (Δt) describes the time resolution of the basis function, whereas the frequency width ($\Delta \omega$) is related to the frequency resolution. The time and frequency resolutions determine how well the transformation will express the localized time-frequency structures. When considering basis functions and their tile widths, high resolution corresponds to a small tile width since resolution and tile width are inversely proportional.

Consider the basis function $\psi_k(t)$ which has an energy distribution that is concentrated near the time instant k and is spread out over the interval Δt [6]. This basis function $\psi_k(t)$ can identify time-localized features (at k) with a resolution of $\frac{1}{\Delta t}$. Similarly, in the frequency domain, the Fourier representation $\Psi_\xi(\omega)$ is concentrated in energy near the frequency ξ and spread over the interval of $\Delta \omega$, which captures frequency-localized features (at ξ) with a resolution of $\frac{1}{\Delta \omega}$. A few arbitrary time-frequency tiles are shown in Figure 1.3.

Ideally, basis functions with infinitely small time and frequency tile widths would provide

Figure 1.3: Arbitrary time-frequency tiles in the time-frequency plane. Image is adapted from [6].

the best representation since time-frequency structures would be represented with infinite time and frequency resolution. However, there is a direct trade off between time resolution and frequency resolution of basis functions, which is governed by the Heisenburg uncertainty principal [16] [6]. The Heisenburg uncertainty principal states that resolution in time and resolution in frequency both cannot be made arbitrarily small at the same time and the product of the corresponding tile widths is lower bounded by

$$\Delta\omega \cdot \Delta t \geq 1/2. \quad (1.7)$$

Sinusoidal and Delta Dirac Basis Sets

In this section, basic functions will be investigated to gain insight into the localization properties of basis functions. The two functions which will be discussed are: sinusoidal and delta dirac functions.

The 1-D Fourier transform uses sinusoidal basis functions to decompose the original signal

$$f(t) = \int_{-\infty}^{+\infty} F(\omega) e^{j\omega t} dt, \quad (1.8)$$

where the Fourier representation of $f(t)$ may be found by

$$F(\omega) = \langle f(t), e^{j\omega t} \rangle = \int_{-\infty}^{+\infty} f(t) \cdot e^{-j\omega t} dt, \quad (1.9)$$

and ω is the angular frequency of the sinusoid. Each sinusoidal basis function is defined for all time (no time resolution) and a particular frequency ω (infinite frequency resolution). Due to the non-local nature (or lack of resolution in time of sinusoids), any signal expansion using these basis functions will spread time-local information (like abrupt amplitude changes, or transients) over the whole frequency axis of $F(\omega)$ [17]. Therefore, to identify time-local information within a signal, this transformation will not be sufficient.

In 2-D, the Fourier transform uses sinusoids which are a function of two frequency variables u and v . A 2-D image $f(x, y)$ may be written as the expansion of the 2-D basis functions, as shown below:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{j2\pi(ux+vy)} du dv, \quad (1.10)$$

where $F(u, v)$ is the 2-D Fourier transform of $f(x, y)$ and is defined as

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy. \quad (1.11)$$

Similar to the 1-D case, these 2-D sinusoids do not offer any insight into the spatially localized frequency structures. This is shown in Figure 1.4, where the 2-D discrete Fourier transform was computed for the nonstationary images of Figure 1.2. Figure 1.4(a) is a two-dimensional *sinc*($\cdot$) function $F(u, v)$, and Figure 1.4(b) contains $G(u, v)$ which has directional frequency components. As can be seen by these images, the frequency domain representation shows what frequency content is present in the original image, but does not indicate *where* in the images these frequency events occur.

Although 2-D sinusoidal basis functions cannot localize structures in space, any frequency component which is present within the image can be expressed using this basis set,

Fourier Transform of Square Image – $F(u,v)$

(a)

Fourier Transform of Lena Image – $G(u,v)$

(b)

Figure 1.4: (a) The Fourier transform of the image $f(x,y)$ of Figure 1.2(a) revealing a two-dimensional $\text{sinc}(\cdot)$ function, (b) the Fourier transform of the Lena image $g(x,y)$ of Figure 1.2(b) which indicates there are directed edges within the image.

due to the infinite frequency resolution of the sinusoids. For demonstrative purposes, the time, frequency and time-frequency representations for the 1-D sinusoidal basis and 1-D delta basis functions are shown in Figure 1.5 and Figure 1.6, respectively.

In contrast to sinusoids, delta dirac basis functions do not offer any frequency localization, but could reveal useful information about the localized spatial structures, due to their infinite resolution in space. Using the delta functions to decompose a 2-D signal is equivalent to sampling a continuous image, and dependant on the sampling rate, this representation could be equal to a digital representation, such as the images shown in Figure 1.2. The spatial representation of the digital images offer all information and features which are localized purely in space and does not give insight into the frequency properties.

The delta dirac functions can only offer good space localization and the sinusoidal basis sets can only offer good frequency localization properties. As a result, it is necessary to investigate alternative basis sets which offer good resolution in both spatial and frequency

Figure 1.5: (a) Time domain representation of a 1-D sinusoidal basis function $\psi(t)$ with a frequency of f_1 , (b) frequency domain representation of $\psi(t)$ ($\Psi(f)$), (c) time-frequency tiles of a sinusoidal basis function set.

Figure 1.6: (a) Time domain representation of a delta dirac basis function $\psi(t)$ at the time instant t_1 , (b) frequency domain representation of $\psi(t)$ ($\Psi(f)$), (c) time-frequency tiles of a delta dirac basis function set.

domains simultaneously so that qualities of the biomedical images may be examined and used in important medical applications.

Short-Time Fourier Transform (STFT)

The 1-D short-time Fourier transform (STFT) is an alternative signal decomposition technique [6] which windows the sinusoidal basis functions in an effort to add time localization. As was shown, the sinusoidal basis functions could only describe frequency information with high resolution and therefore, a windowed sinusoid may be a viable solution to this problem.

The 1-D STFT is computed by expanding a 1-D signal with sinusoidal basis functions

which are multiplied with a window function $w(t)$ which is nonzero only over a certain time interval (Δt of the time-frequency tile). The basis functions of the STFT may be generated by the following equation

$$w_{k,\xi}(t) = w(t - k) \cdot e^{-j2\pi\xi t}, \quad (1.12)$$

where $w(t)$ is an arbitrary window function, ξ is the frequency of the sinusoid, and k is the time shift. The window $w(t)$ determines both the time and frequency resolution of the basis functions which can be examined from the time and Fourier domain representations, respectively. To get the desired time-frequency resolution, one should choose the appropriate window (which is limited by the Heisenberg uncertainty principal). If $w_{k,\xi}(t)$ forms an orthogonal basis set, a function $f(t)$ may be uniquely represented by

$$f(t) = \sum_k \sum_{\xi} \hat{F}(k, \xi) \cdot w_{k,\xi}(t), \quad (1.13)$$

where $\hat{F}(k, \xi)$ is the STFT of $f(t)$ and is computed by

$$\hat{F}(k, \xi) = \langle w_{k,\xi}(t), f(t) \rangle. \quad (1.14)$$

Although the flexibility of the STFT allows a specific time-frequency resolution for decomposition, it has a fixed resolution for the entire decomposition process. This is shown in Figure 1.7(a), where the time-frequency tilings for a narrow window $w(t)$ are shown and only one resolution is offered for the entire time-frequency plane. Although there is flexibility in the choice of window $w(t)$, if a narrow window is selected, good time resolution is achieved but poor frequency resolution results. Similarly, if a wide window is used, the frequency resolution is good but the time resolution is poor and temporal events such as transients, cannot be represented using these functions. Therefore, the STFT would have limited use in many applications since the basis functions are unable to capture both time and frequency events with good resolution. In fact, this is a common and well understood problem of the STFT [6] [16].

For 2-D images $f(x, y)$, the STFT $\hat{F}(\tau_1, \tau_2, u, v)$ is a four-dimensional function, and may

be computed by

$$\hat{F}(\tau_1, \tau_2, u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \cdot w^*(x - \tau_1, y - \tau_2) \cdot e^{-j(ux+vy)} dx dy, \quad (1.15)$$

where τ_1 and τ_2 are spatial shifts, u and v are frequency variables and $w(x, y)$ is a 2-D windowing function. The 2-D STFT for images is not only difficult to visualize, but is also complex to compute. For instance, for each combination of τ_1 and τ_2 , there is a corresponding two-dimensional plot with frequency variables u and v as its axis'. This plot would correspond to the localized frequency information in the region of τ_1 and τ_2 . For an image with $N \times N$ dimensions, a full decomposition would require $N \times N$ two-dimensional frequency plots. This would require significant amounts of memory bandwidth for large biomedical images, be computationally complex to compute and it would be cumbersome to address and extract information from this representation.

The STFT basis functions are able to offer some space-frequency localization proper-

Figure 1.7: (a) Time-frequency tiles of the 1-D STFT, (b) time-frequency tiles of a two-band 1-D wavelet transform.

ties but: (1) offer a fixed resolution for the entire decomposition process, (2) does not offer an easy method to access and manage the coefficients and (3) creates a drastic increase in memory consumption and computational resources. To offer a solution to these problems,

the discrete wavelet transform (DWT) can be used [18] [17] [19] [6]. The 2-D DWT offers a multiresolutional representation (decomposes the image using various scale-frequency resolutions), which is achieved by dyadically changing the size of the window. This gives way to basis functions that are tuned to events which have high frequency components in a small analysis window (scale) or low frequency events with a large scale [16]. In biomedical images, it is likely to find all types of localized events which could be efficiently represented using a set of multiresolutional basis functions. Additionally, the discrete wavelet transform utilizes critical subsampling along rows and columns and uses these subsampled subbands as the input to the next decomposition level. For a 2-D image, this reduces the number of input samples by a factor of four for each level of decomposition. This representation may be stored back on to the original image for minimum memory usage and it also permits for an organized, computationally efficient manner to access these subbands and extract meaningful features. For demonstrative purposes, the time-frequency tiling offered by a two-band wavelet transform for 1-D signals is shown in Figure 1.7(b). As is illustrated in this figure, there isn't a fixed time-frequency resolution for the entire decomposition process, which allows for both good time and good frequency resolutions at different points in the decomposition. The 2-D DWT of the nonstationary images of Figure 1.2 are shown in Figure 1.8. Note how the DWT arranges the subbands into an organized, compact representation and also how the wavelet basis functions are able to isolate the directed edge components, indicating good space-frequency localization.

1.3 Experimental Images

This thesis investigates multiresolutional analysis tools for biomedical images since the DWT seems to offer many features which could be beneficial for an image analysis application. Mammogram images, retinal images and gastrointestinal images are the particular types of images which are used for experimentation. A brief description of these modalities and the challenges associated with each type of imagery, is included below.

Discrete Wavelet Transform for Square Image

(a)

Discrete Wavelet Transform for Lena Image

(b)

Figure 1.8: (a) The discrete wavelet transform of the square image shown in Figure 1.2(a) using the (3,5) biorthogonal wavelet displaying the localization of horizontal, vertical and diagonal edge components and (b) the discrete wavelet transform of the Lena image shown in Figure 1.2(b) using the Meyer Wavelet where the directional edges associated with Lena are clearly isolated within the subbands. Some contrast enhancement was performed for visualization purposes.

1.3.1 Mammogram Images

Mammography is an imaging technology which acquires an x-ray image of the breast [20]. Mammogram images are currently the most effective method for early detection of breast cancers [21] [22]. According to [23], for women over the age of 50, breast cancer mortality is reduced 33% for women who undergo regular screening. Additionally, for women less than 50 years of age, the reduction in breast cancer mortality is approximately 15% when screened with mammography [23].

The trained physician screens each of the mammograms looking for microcalcifications and masses [20], architectural distortion, developing density, asymmetry and isolated dilated ducts [24]. Microcalcifications and masses are used very regularly for disease identification [21]. However, although both pathologies are important, masses are harder to detect

since their features may be obscured or similar to normal breast parenchyma [21]. Furthermore, graylevel variances at lesion boundaries are usually ill-defined when compared to the boundaries of microcalcifications [22]. Therefore, a difficult and important problem faced by radiologists is the identification of masses within the breast tissue.

Some typical mass shapes are shown in Figure 1.9. Most often, higher density, ra-

Figure 1.9: Typical mass shapes. (Reproduced with the permission from H. Alto, R.M. Rangayyan, and J.E.L. Desautels, "Content-based retrieval and analysis of mammographic masses", *Journal of Electronic Imaging*, 2005, Vol. 14, No. 2. © SPIE.)

diopaque, irregular shaped masses (spiculated and microlobulated) are associated with malignancy and smoother (circumscribed and macrolobulated), radiolucent masses are benign [25]. Although the characteristics of the average benign and malignant mass are known, detection and differentiation between these two mass types is still difficult since normal tissue masks them and other features such as breast parenchyma are much more prominent than the lesion itself [20]. To reduce the number of negative to positive biopsies, which are about 5:1 and are as high as 11:1 in some clinics [26], it is necessary to have a good way to detect the benign and malignant masses. Such a method may be achieved by a computer-aided detection algorithm, which exploits the textural qualities of the masses using feature extraction. The features which are extracted in such a system may also be used in a content-based image retrieval scheme where the mammograms are retrieved based on the content, not by text annotations. Since text annotations are time consuming to create and error prone, implementation of such a system would be very beneficial.

Two mammograms are shown in Figure 1.10, where the one on the left contains a benign mass and the one on the right contains a malignant mass. As shown by these images, it is

quite difficult to visually differentiate between the masses since overlapping structures are obscuring each lesion. As a result of mass obstruction, discriminating between these types of lesions using an automated feature extraction and classification scheme could also be problematic. Overcoming this difficulty could pose a significant challenge.

As a result of the benefits associated with mammography, regular screening is recom-

Figure 1.10: (a) Benign mass in a 1024×1024 mammogram, (b) malignant lesion in a 1024×1024 mammogram. Lesions are shown inside the square box.

mended by leading health care societies for women between the ages 50 to 69 as often as two times a year [26]. These screening programs have created a large influx of digital mammograms and database management is difficult for hospitals and clinics. To combat this, a high performance source encoding scheme strictly dedicated mammograms must be utilized. This will not only aid in database management, but it will also allow for efficient bandwidth usage in a telemammography system. Such a compression scheme would employ a transformation which exploits the correlated or other properties of the digital mammogram to arrive at the maximum file size reduction. Identification of such a transform is key for medical imaging applications.

1.3.2 Small Intestine Images

Traditional endoscopes are operated by highly skilled doctors and are inserted through the mouth of the patient. Since the endoscopy's tube is rather stiff, it can cause the patient quite a bit of discomfort as the doctor navigates it through the gastrointestinal tract [27]. Furthermore, the endoscopic camera cannot reach the small intestine [27] and therefore, doctors have had substantial difficulties diagnosing diseases of the small bowel [28]. To combat this, Given Imaging Ltd. has developed a capsule endoscopy known as the PillCamTM SB video capsule, which was formally marketed as M2ATM.

The PillCamTM is a tiny capsule (10mm $\times$ 27mm [27]), which is ingested from the mouth with water. As natural peristalsis moves the capsule through the gastrointestinal tract (including the small bowel) it captures video and wirelessly transmits it to a data recorder the patient is wearing around his or her waist [29]. This video provides visualization of the 21 foot long small bowel, which was originally seen as a "black box" to doctors [30]. Video is recorded for approximately 8 hours and then the capsule is excreted naturally with a bowel movement [29]. Clinical results for the PillCamTM show that it is a superior diagnostic method for diseases of the small intestine [28].

In the small intestine, there are four main types of cancers, which are named after the cell they originate from. The names and descriptions are listed below [30]:

1. **Adenocarcinoma:** This type of small bowel cancer originates in the epithelial lining of the mucosa and is mostly found in the duodenum.
2. **Sarcoma:** This cancer originates in the muscle wall of the small bowel (most often it is found specifically in the ileum).
3. **Carcinoid:** These types of malignancies arise in the specialized neuroendocrine cells and appear in the small intestine, the ileum (most common) and sometimes within the appendix.
4. **Lymphoma:** The lymphoid tissue of the small bowel is where lymphoma tumours originate from. Majority of these lymphomas are known as non-Hodgkin's lymphoma,

and most commonly appear in the jejunum or ileum.

The PillCamTM offers gastroenterologists a new method of detecting disease of the small bowel and offers a live, video representation which was never available with the traditional endoscopy. The downfall of this technology comes from the large amount of data which is collected while the PillCamTM is on its journey through the gastrointestinal tract. The doctor has to watch and diagnose approximately 8 hours of footage! This is quite a labourious task for the physicians and it could cause them to miss important clues due to fatigue, boredom or due to the repetitive nature of the task. Therefore, a computer algorithm can be used as a secondary opinion, which views all the video images and points out suspicious areas to the doctor. Additionally, the extracted features may be used for a fast retrieval scheme, since trying to locate the instants at which the suspicious region(s) are visible in the video may be cumbersome since the footage is 8 hours long. If such a scheme was developed, it would allow text annotations to be avoided altogether, which are time consuming to create or access, and may be prone to human error on entry. Since the camera angle (and hence the images) can be expected to be different from patient to patient and suspicious regions may occur in several different places along the gastrointestinal tract, it may be difficult for a feature extraction algorithm to account for all these scenarios. Sample images taken by the PillCamTM are included in Figure 1.11, which includes three normal homogeneous images and three abnormal heterogenous, textured images.

1.3.3 Retinal Images

Ophthalmologists use digital fundus cameras to acquire and collect images of the human eye [31]. These cameras capture images of the optic nerve, fovea, surrounding vessels and the retinal layer [4]. Often, this type of imaging technique is referred to as *retinal imaging*.

The retinal images which are acquired by the medical professional are later interpreted by an ophthalmologist. The ophthalmologist may search the images for signs of various diseases, including diabetic retinopathy. This type of sight-threatening disease is common for diabetics [32]. In 2000, 16 million Americans were reported as having either Type I (juvenile onset)

Figure 1.11: Small bowel images (256×256) collected by the PillCamTM obtained from the Image Atlas of Given Imaging Ltd. See [2]. (a) Healthy small bowel, (b) normal jejunum, (c) normal colonic mucosa, (d) small bowel lymphoma, (e) jejunal carcinoma, (f) Kaposi's sarcoma.

or Type II (adult onset) diabetes [32]. Although screening with retinal imaging reduces the risk of blindness caused by diabetic retinopathy by 50% [31], it also creates a large number of images which the doctors need to interpret [33]. This is expensive, time consuming and may be prone to human error. To aid the doctors with this diagnostic task, an automated system could act as a second reader, offering a second opinion of the images. Furthermore, the extracted features may also be used as an efficient indexing scheme for a content-based image retrieval (CBIR) application. Text annotations and other database indexing schemes require human interaction to properly tag and log each image. A CBIR scheme would also allow for fast database searching according to a particular diagnosis. Four retinal images (two normal and two abnormal) are shown in Figure 1.12, exhibiting various textured regions.

Although diabetic retinopathy is one of the most common diseases sought with retinal

(a)

(b)

(c)

(d)

Figure 1.12: Retinal images (700×605) supplied by the STare public database [4]. (a) Normal retinal image, (b) normal retinal image, (c) retinal image with background diabetic retinopathy, (d) retinal image with central retinal vein occlusion.

imaging technologies, other diseases and medical conditions exist. For instance: choroidal neovascularization, central retinal vein occlusion, myelinated nerve fibers and age related macular degeneration are only a few examples [4]. Typical diagnostic techniques look for changes of the eye, which may include: increased number of blood vessels [34], presence of exudates/lesions (whitish/yellow patches of varying size, shape and location) [32] or drusens (cellular waste that form beneath the retina) [33]. These textural properties may be used for feature extraction and classification, as a way to numerically differentiate a particular pathology from the normal retinal images. A more useful application would discriminate

all abnormal retinal images from the normal ones. However, the development of such a scheme would be challenging since a single, unified framework would have to account for all pathology types.

1.4 Motivation

This thesis aims to evaluate the usefulness of multiresolutional image processing techniques for biomedical image applications via the discrete wavelet transform. In particular, two important applications will be addressed: classification and compression.

The classification application will employ a feature extraction scheme in the wavelet and spatial domains for all three images (retinal, small bowel and mammogram), where the numerical descriptors are selected to differentiate the images based on their textural characteristics. Extraction of optimal features will be one of the main challenges faced since there are many overlapping structures and other textured regions in the images. Additionally, since the discrete wavelet transform is shift-variant [35] [36], it will also be necessary to investigate techniques which permit for the extraction of a consistent feature set. The extracted features can either serve as the input to a computer-aided diagnosis (CAD) system, or they can be used as indexes for a CBIR scheme.

Wavelet-based compression techniques will be investigated for the mammograms only, since the other images are stored in an already compressed format. The performance of a wavelet-based coder will be compared to several other lossless coders for digital mammograms. An optimal compressor will warrant a file size reduction that permits efficient storage and transmission in teleradiological applications. The block diagram of the proposed works is included in Figure 1.13.

1.5 Outline of Thesis

As previously stated, this thesis investigates the benefits of the DWT for classification and compression of biomedical images. Finding a good image analysis tool will permit localization

Figure 1.13: Thesis block diagram.

of nonstationary events, which may be beneficial for both these applications.

Chapter 2: Multiresolutional Analysis

This chapter discusses multiresolutional analysis by the way of the discrete wavelet transform (DWT). Multiresolutional analysis is looked at from a signal decomposition view, and the discrete implementation is briefly discussed. Properties of the DWT which effect both compression and feature extraction are investigated (such as: scale-invariance and shift-variance).

Chapter 3: Classification of Biomedical Images

This chapter looks at multiresolutional feature extraction for mammogram, retinal and small bowel images. Features extracted from the retinal and small bowel images are computed from the compressed domain. Invariant features are discussed as well as the shift-invariant discrete wavelet transform, which is used to arrive at translation-invariant features. Image texture is described and the extracted features are shown. A brief discussion on the classification scheme used is also included, along side experimental results.

Chapter 4: Compression of Biomedical Images

An important application for mammogram images is compression. Chapter 4 looks at JPEG 2000, which uses the discrete wavelet transform for decomposition, and compares it to various

other lossless compression schemes for mammogram images. The other coders reviewed include JPEG-LS, LJPEG, arithmetic, LZW and adaptive Huffman. A comparative and comprehensive overview is given for these source encoders, and the overall performance and advantages for a mammogram image compression application are discussed. Experimental results are also given.

Chapter 5: Conclusions

Chapter 5 will contain all experimental discussions, a summary of the results and directions for future work.

Chapter 2

Multiresolutional Analysis (MRA)

2.1 Introduction

IN the previous chapter, biomedical images were shown to be nonstationary and a non-stationary analysis tool must be identified to extract useful information from the images. The basis functions of the STFT provide a space-frequency resolution that is inadequate for capturing the wide variety of localized structures which are common in biomedical imagery. Additionally, this transformation is very computationally complex and memory intensive when computed for images. Consequently, for biomedical image analysis, a multiresolutional approach may be a good alternative to the STFT. Multiresolutional analysis (MRA) techniques can represent low and high frequency structures with different spatial resolutions (and in one basis set), in a computationally efficient manner.

Wavelet analysis techniques were independently researched and developed in the fields of mathematics, quantum physics, electrical engineering and seismic geology [37]. Consequently, a series of applications which utilize these techniques have been created. Such applications include FBI fingerprint compression, computer and human vision schemes, data denoising, detecting self-similar behaviour in a time-series and many more [37].

In this thesis, MRA techniques will be used on biomedical images for two applications: (1) feature extraction and (2) compression. The particular MRA technique which will be applied is known as the Discrete Wavelet Transform (DWT). To give the reader some background on wavelets, some multiresolutional theory is included in this Chapter. For simplicity,

wavelet subspace and basis function theories will be discussed for only 1-D signals, and then a 2-D extension will be given. The properties of the DWT will also be described, including the benefits and limitations this transform has to offer for both compression and feature extraction.

2.2 Wavelets, Subspaces and Signal Decomposition

Prior to applying any transformation, it is important to understand the properties of the underlying basis functions and subspaces which these function sets occupy.

2.2.1 Scaling Function Subspaces

Consider a set of basis functions $\{\phi_k(t)\}$ which may be created by translating the prototype scaling function $\phi(t)$

$$\phi_k(t) = \phi(t - k), \quad k \in \mathbf{Z}, \quad (2.1)$$

where $\phi_k(t)$ spans the space $\mathcal{V}_o$. This may be written as

$$\mathcal{V}_o = \overline{\text{Span}_k\{\phi_k(t)\}}, \quad (2.2)$$

where the overbar denotes closure [16]. If a set of basis functions span a signal space $\mathcal{V}_o$, then any function $f(t)$ which also belongs to that space can be completely represented using those basis functions as in: $f(t) = \sum_k a_k \cdot \phi_k(t)$ (for any $f(t) \in \mathcal{V}_o$).

For added flexibility, the time and frequency resolution of these scaling functions may be adjusted by including an additional scale parameter j in the characteristic basis function expression

$$\phi_{j,k}(t) = 2^{j/2} \cdot \phi(2^j t - k), \quad j, k \in \mathbf{Z}, \quad (2.3)$$

where the scalar multiple $2^{j/2}$ is included to ensure orthonormality [18]. Therefore, an entire series of basis functions can be created by simply dilating (changing the j value) or translating (changing the k value) the prototype scaling function $\phi(t)$. These basis functions span the subspace $\mathcal{V}_j$

$$\mathcal{V}_j = \overline{\text{Span}_k\{\phi_k(2^j t)\}},$$

subset of those with smaller temporal resolution (can represent finer detail) [6]. This can be formally expressed by

$$\mathcal{V}_j \subset \mathcal{V}_{j+1}, \quad \forall j \in \mathbf{Z}. \quad (2.7)$$

Equation 2.7 indicates that signals which reside inside $\mathcal{V}_j$ must also reside inside $\mathcal{V}_{j+1}$. Consequently, spaces which contain high resolution signals must also contain those of lower resolution as well. This allows coarse detail that is isolated by the transformation at scale j to be constructed from the finer details at another scale $j + 1$. Therefore, if $\mathcal{V}_j$ is spanned by $\phi_k(2^j t)$ and is used to represent $f(t)$, then $\mathcal{V}_{j+1}$ is spanned by the scaling functions $\phi_k(2^{j+1} t)$ and is able to represent all functions $f(2t)$ [16]. For the spaces $\mathcal{V}_0$ and $\mathcal{V}_1$: if $\mathcal{V}_0$ is spanned by $\phi(t)$, $\mathcal{V}_1$ is spanned by $\phi(2t)$ and must also contain the scaling functions $\phi(t)$, since $\mathcal{V}_0 \subset \mathcal{V}_1$ (recall $\mathcal{V}_j \subset \mathcal{V}_{j+1}$). Consequently, $\phi(t)$ can be expressed in terms of weighted, linear combinations of time shifted versions of $\phi(2t)$. This creates the multiresolutional, recursive equation as defined by

$$\phi(t) = \sum_n h_o(n) \cdot \sqrt{2} \phi(2t - n), \quad n \in \mathbf{Z}, \quad (2.8)$$

where $h_o(n)$ are the scaling functions' coefficients.

One of the last requirements for multiresolutional decomposition comes from the definition of the subspaces. In order for $\{\mathcal{V}_j\}_{j \in \mathbf{Z}}$ to be a multiresolution approximation of $L^2(\mathbf{R})$, along with the properties previously discussed, the following must also be true

$$\mathcal{V}_{-\infty} = \{0\}, \quad (2.9)$$

which tells us that the basis functions at this scale are infinitely wide (in time), and

$$\mathcal{V}_{\infty} = L^2, \quad (2.10)$$

which indicates that these basis functions are infinitely narrow. In fact, Equation 2.9 and Equation 2.10 are the essence of multiresolutional analysis since within one family of functions, it is possible to arrive at a variety of basis functions, in which the time-frequency properties are changing.

2.2.2 Wavelet Basis Functions

Although the scaling functions give way to a multiresolution representation, it is also necessary to investigate the spaces which span the *differences* of the spaces spanned by the scaling functions. These types of basis functions are known as wavelets $\psi(t)$ and their corresponding signal spaces are denoted as $\mathcal{W}$. Similar to scaling functions, a series of wavelet basis functions can be generated by dilating and translating the mother wavelet $\psi(t)$

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z}. \quad (2.11)$$

In order to find the mother wavelet $\psi(t)$, it is necessary to find the relationship between the mother wavelet $\psi(t)$ and the generating scaling function $\phi(t)$.

Starting with an initial resolution of $j = 0$, the nested subspaces may be written as

$$\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset L^2. \quad (2.12)$$

The corresponding spaces spanned by the wavelet basis functions are shown in Figure 2.1, which illustrates how each $\mathcal{W}$ subspace spans the difference of two subspaces. As shown in Figure 2.1, the signal spaces $\mathcal{V}_1$ and $\mathcal{V}_2$ may be expressed as

$$\mathcal{V}_1 = \mathcal{V}_0 \oplus \mathcal{W}_0, \quad (2.13)$$

and

$$\mathcal{V}_2 = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1, \quad (2.14)$$

where $\oplus$ is a direct sum. If $\mathcal{V}_j$ is the space spanned by the scaling functions $\phi_{j,k}(t)$ and $\mathcal{V}_{j+1}$ is the space spanned by the functions $\phi_{j+1,k}(t)$, then $\mathcal{W}_j$ is the disjoint difference or the orthogonal compliments of $\mathcal{V}_j$ and $\mathcal{V}_{j+1}$ spanned by the wavelet basis functions $\psi_{j,k}(t)$. This may be shown by

$$\mathcal{V}_{j+1} = \mathcal{V}_j \oplus \mathcal{W}_j, \quad \forall j \in \mathbb{Z}. \quad (2.15)$$

Using Equation 2.13, Equation 2.14 and Figure 2.1, a general expression for the L^2 subspace may be developed:

$$L^2 = \mathcal{V}_0 \oplus \mathcal{W}_0 \oplus \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \cdots, \quad (2.16)$$

which demonstrates that these subspaces are orthogonal to one another

$$\mathcal{V}_o \perp \mathcal{W}_o \perp \mathcal{W}_1 \perp \mathcal{W}_2 \perp \mathcal{W}_3 \cdots, \quad (2.17)$$

indicating that the corresponding basis functions which span these spaces are also orthogonal

$$\langle \phi_{j,k}(t), \psi_{j,k}(t) \rangle = \int \phi_{j,k}(t) \cdot \psi_{j,k}(t) dt = 0. \quad (2.18)$$

Furthermore, wavelet spaces at a scale j , are a subset of the scale spaces at the next scale $j + 1$

$$\mathcal{W}_j \subset \mathcal{V}_{j+1}. \quad (2.19)$$

Consequently, wavelets reside in the space spanned by the next narrower scaling function and can be expressed as a weighted sum of shifted scaling functions, $\phi(2t)$

$$\psi(t) = \sum_n h_1(n) \cdot \sqrt{2} \cdot \phi(2t - n), \quad n \in \mathbb{Z}, \quad (2.20)$$

where $h_1(n)$ are the wavelets' coefficients. Equation 2.20 shows that the generating wavelet $\psi(t)$ can be produced from the prototype scaling function $\phi(t)$ by choosing the appropriate $h_1(n)$. In order to ensure orthogonality, the scaling and wavelet coefficients must be related by [16]

$$h_1(n) = (-1)^n h_o(1 - n). \quad (2.21)$$

2.2.3 Wavelet Transforms

In the last section, it was shown that the L^2 subspace is made up of a series of orthogonal subspaces

$$L^2 = \mathcal{V}_o \oplus \mathcal{W}_o \oplus \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \cdots,$$

which includes both a scaling subspace as well as several wavelet subspaces. Consequently, starting with the initial scale $j = 0$, any function $f(t)$ which belongs to $L^2(\mathbb{R})$ may be decomposed using both wavelets and scaling functions, as shown below:

$$f(t) = \sum_{k=-\infty}^{k=\infty} c(k) \cdot \phi_k(t) + \sum_{j=0}^{j=\infty} \sum_{k=-\infty}^{k=\infty} d(j, k) \cdot \psi_{j,k}(t), \quad (2.22)$$

where $c(k)$ are the scaling or averaging coefficients defined by

$$c(k) = c_0(k) = \langle f(t), \phi_k(t) \rangle = \int f(t) \phi_k(t) dt, \quad (2.23)$$

and $d_j(k)$ are the detail wavelet coefficients defined by

$$d_j(k) = d(j, k) = \langle f(t), \psi_{j,k}(t) \rangle = \int f(t) \psi_{j,k}(t) dt. \quad (2.24)$$

The representation achieved by Equation 2.22 is known as the Wavelet Transform.

2.3 Discrete Wavelet Transform

In the last section, the wavelet transform was shown and both the scaling and wavelet basis functions were used for decomposition. In order to perform the wavelet transform for discrete signals and images, implementation of the DWT using filterbanks will be discussed. All discussions for one dimension will be later extended to two dimensions for images.

After performing a series of simplifications and change of variables [16] [6] [19], Equation 2.8 and Equation 2.20 may be rewritten as

$$c_j(k) = \sum_m h_o(m - 2k) c_{j+1}(m), \quad (2.25)$$

$$d_j(k) = \sum_m h_1(m - 2k) c_{j+1}(m). \quad (2.26)$$

This illustrates that $c_j(k)$ and $d_j(k)$ can be found by filtering $c_{j+1}(k)$ with $h_o(-n)$ and $h_1(-n)$, respectively, followed by a decimation by a factor of 2. To compute the DWT coefficients, a two stage analysis filterbank for 1-D signals may be used as shown in Figure 2.2(a), where c_{j+1} is decomposed over two stages. After one level of decomposition, the coefficients c_j and d_j are produced. The scaling coefficients $c_j(k)$ are further decomposed using the filters $h_o(-n)$ and $h_1(-n)$ to produce the coefficients $c_{j-1}(k)$ and $d_{j-1}(k)$. In Figure 2.2(a), the resultant signal spaces are listed on top of the filter output and the corresponding decomposition tree is shown in Figure 2.2(b). Note that the initial scale here is $j + 1$, and therefore c_{j+1} would represent the original input signal. A description of the type of filters used for the DWT is included in Appendix A. The most important things to know about

Figure 2.2: (a) Computing the 1-D wavelet and scaling coefficients using filtering and decimation with a 2-stage analysis filterbank, (b) corresponding decomposition tree showing the division of signal spaces.

the filters is that $h_0(n)$ and $h_1(n)$ are half-band lowpass and highpass filters, respectively. Consequently, the lowpass filter $h_0(n)$ produces lowpassed or averaged coefficients $c_j(k)$ and the highpass filter $h_1(n)$ creates highpassed or detail coefficients $d_j(k)$.

2.3.1 2-D Extension for Images

Instead of having a wavelet or filter which is a function of the two spatial dimensions of an image, the filter can be separable, which allows a particular 1-D filter to be applied to the rows and columns of an image separately to gain the desired overall 2-D response [38]. A separable filter for two dimensions may be denoted by:

$$H(z_1, z_2) = H(z_1) \cdot H(z_2), \quad (2.27)$$

where z_1 and z_2 relate to the spatial dimensions of an image. Therefore, the filters defined for the 1-D DWT may be applied separably to gain a 2-D DWT representation for images. The 2-D DWT filterbank scheme for an $N \times N$ image $x(m, n)$ is shown in Figure 2.3. Initially, the filters $H_0(z)$ and $H_1(z)$ are applied to the rows of image $x(m, n)$, creating two images which respectively contain the low and high frequency content of the image in question. After this, both frequency bands are subsampled by a factor of 2, and are sent to the next set of filters for filtering along the columns. After these bands have been filtered, decimation by a

factor of 2 is again performed, but this time along columns. At the output of one level of decomposition, as shown in Figure 2.3, there are four subband images of size $\frac{N}{2} \times \frac{N}{2}$ labeled LL , LH , HL and HH . Using the separability concept, at scale j , these subbands may be computed by

$$LL^j(x, y) = \sum_m \sum_n h_o(m - 2x) h_o(n - 2y) \cdot LL^{j+1}(m, n), \quad (2.28)$$

$$HL^j(x, y) = \sum_m \sum_n h_o(m - 2x) h_1(n - 2y) \cdot LL^{j+1}(m, n), \quad (2.29)$$

$$LH^j(x, y) = \sum_m \sum_n h_1(m - 2x) h_o(n - 2y) \cdot LL^{j+1}(m, n), \quad (2.30)$$

$$HH^j(x, y) = \sum_m \sum_n h_1(m - 2x) h_1(n - 2y) \cdot LL^{j+1}(m, n). \quad (2.31)$$

The first letter of the subimages indicates the operation that was performed on the columns (i.e. L is for lowpass filtering with $H_o(z)$ and H is for highpass filtering with $H_1(z)$) whereas the last letter indicates which operation was performed on the rows. If more levels of decom-

Figure 2.3: Filterbank implementation of 2-D discrete wavelet transform (DWT).

position are required, the LL band may be recursively reapplied to the analysis filterbank of Figure 2.3. For two levels of decomposition, the placement of the coefficients back onto the

image is shown in Figure 2.4.

Figure 2.4: Graphical depiction of wavelet coefficient placement for two levels of decomposition.

2.3.2 5/3 Wavelet

The integer wavelet which will be used for decomposition is part of the Odd-Length Analysis/Synthesis Filter (OLASF) family, where the number of filter taps in the FIR filter (for the filterbank implementation) are odd [39]. Additionally, biomedical images are high resolution images, which results in large image sizes. Consequently, for these large-sized images, a wavelet with fewer taps is desired so that the overall computational load may be reduced. The 5/3 Le Gull wavelet will be used since the filter lengths are small (5 and 3 taps for the analysis low and highpass filters) and can warrant an efficient implementation [40] [41]. The 5/3 filter coefficients are listed in Table 2.1.

Using the 5/3 integer wavelet, the highpass details $d_j(n)$ can be computed using a lifting based DWT (see Appendix B for more details on lifting):

$$d_j(n) = c_{j+1}(2n+1) - \left\lfloor \frac{c_{j+1}(2n) + c_{j+1}(2n+2)}{2} \right\rfloor, \quad (2.32)$$

where $\lfloor X \rfloor$ is the greatest integer less than or equal to X. The low frequency, average coefficients $c_j(n)$ may be found using an update function

$$c_j(n) = c_{j+1}(2n) + \left\lfloor \frac{d_j(n) + d_j(n-1) + 2}{4} \right\rfloor. \quad (2.33)$$

Table 2.1: Analysis and synthesis filter coefficients for the 5/3 wavelet.

	Analysis Coefficients		Synthesis Coefficients	
i	$h_o(i)$	$h_1(i)$	$h_o(i)$	$h_1(i)$
0	$+\frac{6}{8}$	+1	+1	$+\frac{6}{8}$
± 1	$+\frac{2}{8}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{2}{8}$
± 2	$-\frac{1}{8}$			$+\frac{1}{8}$

For reconstruction, the reverse DWT can be found by reversing the arithmetic operations of the forward transform. This is shown below:

$$c_{j+1}(2n) = c_j(n) + \left\lfloor \frac{d_j(n) + d_j(n-1) + 2}{4} \right\rfloor, \quad (2.34)$$

$$c_{j+1}(2n+1) = d_j(n) - \left\lfloor \frac{c_{j+1}(2n) + c_{j+1}(2n+2)}{2} \right\rfloor. \quad (2.35)$$

These equations may be applied separably to the images in order to gain a 2-D DWT representation. More information can be found in Appendix B.

2.4 Properties of the DWT

The DWT has been discussed as a signal decomposition technique as well as from an implementation point of view. The utility of such a transform for both compression and feature extraction of biomedical images requires an investigation into the important properties of the DWT. These will be discussed here.

2.4.1 Scale Invariance

As was discussed in Section 2.2, two nested subspaces $\mathcal{V}_j$ and $\mathcal{V}_{j+1}$ are spanned by two sets of basis functions which are scaled versions of one another $\phi(2^j t)$ and $\phi(2^{j+1} t)$, respectively. Signals which belong to these subspaces may be expanded using these basis sets. For example, at scale $j = 0$, if $f(t) \in \mathcal{V}_0$, then $f(t)$ may be expanded by $\phi(t)$: $f(t) = \sum_k a_{0,k} \cdot \phi(t - k)$. Similarly, at the next scale $j = 1$, signals $f(2t) \in \mathcal{V}_1$, may be expressed as linear combinations of $\phi(2t)$, as in: $f(2t) = \sum_k a_{1,k} \cdot \phi(2t - k)$. In the general case, it is possible to express any

signal $f(2^j t) \in \mathcal{V}_j$ by $f(2^j t) = \sum_k a_{j,k} \cdot \phi(2^j t - k)$. Therefore, a complete multiresolutional decomposition will contain all basis functions $\{\phi(2^j t)\}_{0 \leq j \leq J}$ necessary to decompose various scaled versions of the original signal $f(t)$. In other words, any scaled version of the input function $f(t)$ will be matched to a scaling function with the same scale in the basis family. For signals which are scaled by non-dyadic factors (powers of two), the basis function most similar to this scaled signal will be used to decompose it and the remaining parts of the signal will be diffused across the other bands. Generally, for any decomposition scheme, as long as scaled versions of the basis functions are present in the dictionary, it is possible to achieve a scale-invariant representation.

Why Scale-Invariance is Beneficial

Although scale-invariance was briefly described for 1-D signals, similar methodology also applies to 2-D images as the basis functions are 2-D extensions of those used for 1-D analysis. The reason why scale-invariance is beneficial and important for biomedical image analysis is because pathological features do not come in a predefined size and using a fixed scale (window size) for a feature extraction application would neglect to pick up all pathologies. For instance, within mammograms, spiculated masses range in size from a few millimeters to several centimeters [42]. If the scale used for analysis corresponds to a region of a few square millimeters within a mammogram image, all malignant masses larger than the window would be missed since only a portion of the mass would be captured. Other examples include various sized lesions within the gastrointestinal tract and different sized pathologies in the retinal images. To capture all these lesions, masses and tumours, which come in different sizes, an analysis tool which provides a scale-invariant representation is needed. Since the DWT transform is scale-invariant, various sized image features may be captured within one basis set, making it a viable option for robust feature extraction.

2.4.2 Shift-Variance

It is a well known fact that the DWT is shift-variant [6] [16]. Therefore, for different translations of the input image or signal, a different set of DWT coefficients would be generated.

This would pose difficulties in a feature extraction application since the same structures located at different positions within the image may cause different features to be extracted. The DWT will be shown to be shift-variant by examining the effect that the subsampling operators have on the transformation, first for the 1-D case and then a 2-D extension will be given followed by an illustrative example. A mathematical proof of the shift-variant property is included in Appendix C.

Consider a 1-D, discrete signal $x(n)$. To compute the 1-D DWT for $x(n)$, it must be low-pass and highpass filtered with $h_0(n)$ and $h_1(n)$, respectively, followed by a downsampling of two. This is illustrated in Figure 2.5(a), where the output after one level of decomposition is one set of scaling coefficients $c_1(n)$ and one set of wavelet coefficients $d_1(n)$. If the same

Figure 2.5: (a) DWT of $x(n)$, (b) DWT of $x(n-1)$. Image adapted from [7].

signal is delayed by one ($x(n - 1)$) and reapplied to the same filterbank, the coefficients $c_1(n)$ and $d_1(n)$ are completely different than those obtained from computing the DWT for $x(n)$. This is shown in Figure 2.5(b), where ‘ $\times$ ’ is marked on the discrete time axis to show the time-relation between the filtered versions of $x(n)$ and $x(n - 1)$. The reason that the output coefficients are different after one level of DWT decomposition for $x(n)$ and $x(n - 1)$ is because the subsampling operators are deleting different samples from the filtered outputs. Consequently, the DWT coefficients for these two signals $x(n)$ and $x(n - 1)$ are *not* time shifted versions of one another.

For all other even shifts of $x(n)$, these signals result in DWT coefficients which are time shifted versions of the coefficients generated by $x(n)$. Similarly, for all odd shifts of $x(n)$, the computed DWT coefficients at this level would be time shifted versions of the coefficients found by filtering and decimating $x(n - 1)$. This behaviour is a consequence of the periodicity of the rate change operator.

As a result of the two different sets of coefficients obtained by shifting the input by either even or odd shifts, the DWT coefficients have two unique representations per decomposition level. What is meant by two *representations* is that one distribution of amplitude coefficients corresponds to the even shifts of the input and another distribution of amplitude coefficients corresponds to the odd shifts. Since these coefficients are not-time shifted versions of one another, the DWT is *shift-variant*. These results can be easily extended to 2-D, for images, since the 1-D filtering and decimation operations are applied separably to rows and columns.

By separately decimating both the rows and columns by two, the 2-D DWT can be expected to have shift-variant properties in both directions. Therefore, as a result of the periodicity of the decimation operators in both the row and column directions, the 2-D DWT would produce four different distributions of coefficients, which are in response to shifts of the input by: $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, where the first index corresponds to the row shift and the second index corresponds to the column shift. The coefficients for all other shifts of the input can be obtained by circularly translating one of the sets of DWT coefficients created

by one of the four fundamental shifts of the input $((0, 0), (0, 1), (1, 0), (1, 1))$. For instance, all other translations of the input by $\{0, 2, 4, 6, \dots\}$ rows and $\{0, 2, 4, 6, \dots\}$ columns would result in DWT representations which are space translated versions of the DWT coefficients computed from the input when it is shifted by $(0, 0)$. Similar results apply for the other input shifts $(0, 1), (1, 0), (1, 1)$, see [43].

Why Shift-Variance is a Problem

Consider the 2-D synthetic image $f(i, j)$ shown in Figure 2.6, where the circle in the middle may be considered some type of circumscribed lesion or mass which needs to be identified by a feature extraction application. If the the image $f(i, j)$ is translated by the fundamental shifts

Figure 2.6: (a) Synthetic image $f(i, j)$ composed of a circle (radius = 60 pixels). Image has dimensions of 256×256 .

$((0, 0), (0, 1), (1, 0), (1, 1))$, let these new images be considered as four separate biomedical images which have the same mass located at different spatial locations. The corresponding DWT for each of these four images is computed and shown in Figure 2.7(b). Visually the DWT coefficients may look similar, although subtle differences can be noticed with careful inspection. However, to prove this concretely, a numerical example will be given.

DWT of $f(i,j)$

(a)

DWT of $f(i-1,j)$

(b)

DWT of $f(i,j-1)$

(c)

DWT of $f(i-1,j-1)$

(d)

Figure 2.7: DWT decomposition of the image $f(i + \Delta i, j + \Delta j)$, where $(\Delta i, \Delta j) \in \{(0,0), (0,1), (1,0), (1,1)\}$.

Table 2.2: Mean μ and variance σ^2 of the DWT coefficients of the LH band, for each DWT image shown in Figure 2.7.

Input shift $(\Delta x, \Delta y)$	Mean μ	Variance σ^2
(0,0)	-0.050537	97.017
(0,1)	-0.051025	100.42
(1,0)	0.057861	96.82
(1,1)	0.058350	98.383

In a feature extraction application, each of the masses contained within the four images ($f(x, y)$, $f(x - 1, y)$, $f(x, y - 1)$, $f(x - 1, y - 1)$) can be characterized by extracting numerical descriptors from the DWT domain. Highly discriminatory descriptors would uniquely describe the lesions and may be sent to a classifier for a CAD system or they may be used as indexes in a CBIR scheme. In order for either of these applications to be robust, the same features must be extracted for each of the four images shown in Figure 2.7 since the *image content* is the same (i.e. they contain the exact same mass, only located at different positions).

To evaluate the effect of the shift-variance property of the DWT for a feature extraction application, an arbitrary feature vector $F = \{F_1, F_2\}$ may be defined. These two features may be any desired descriptor which numerically characterizes the masses, but for demonstrative purposes, let F_1 be the mean μ and F_2 the variance σ^2 of the wavelet coefficients contained in the LH band. Note that features are only extracted from the LH band for simplicity and they also may be extracted from the other bands to arrive at similar results. The features $\{F_1, F_2\}$ may be computed using Equation 2.36 and Equation 2.37:

$$\mu_X = E(X) = \sum_i i \cdot p_X(i), \quad (2.36)$$

$$\sigma_X^2 = E(X^2) - E^2(X) = \sum_i i^2 \cdot p_X(i) - \mu_X^2, \quad (2.37)$$

where $p_X(\cdot)$ is the normalized histogram of X . The results are tabulated in Table 2.2.

As shown by the extracted features listed in Table 2.2, for each of the input translates,

the means μ of each image and variances σ^2 of each image are not equal to one another, even though the same ‘mass’ is present in all images. This would be highly undesirable for a computer-aided diagnosis system since these masses could be classified to be pathologically different, even though they are the same. This would result in a wrong diagnosis and could cause an unnecessary biopsy or a missed malignancy. Furthermore, since the extracted features are different, a query for these mass types would not retrieve all these images in a content-based image retrieval scheme, even though the lesions have the *exact* same characteristics. Therefore, the inability of the DWT to provide a consistent set of coefficients will cause problems for feature extraction applications and a solution to this must be sought.

2.4.3 DWT for Image Analysis

In this thesis, the DWT will be investigated for two very different applications for biomedical images: compression and feature extraction. For compression, the DWT must provide a compact representation in order to achieve high coding gains. Conversely, for feature extraction, the DWT coefficients are required to provide discriminatory information in order to arrive at a sound automatic classification result. Since these two applications have two very different requirements of the DWT, this section will examine how the DWT can realize these two almost ‘contradictory’ applications simultaneously.

DWT for Feature Extraction

To extract meaningful features from the biomedical images, the wavelet representation must somehow isolate events which are relevant to discriminating qualities of the images. As stated in the introductory chapter of this thesis, the abnormal and normal images contain nonstationary image features which may be used by both the physician and a computing device to differentiate between normal cases and the various pathologies discussed. To see how the DWT can provide utility in this sense, the localization properties will be discussed.

When applying a filter in a directional fashion, as is done when computing the DWT for an image, specific frequency events can be isolated. For instance, highpass filtering in the vertical direction (along columns) localizes any horizontal edge components and highpass fil-

tering along the rows would capture vertical edge information. Consequently, the 2-D DWT can localize specific frequency events as a result of the directional filtering operations. For instance, the HH band isolates diagonal edges, the HL subband captures horizontal edge components, the LH subband describes vertical edges and the LL band contains a blurred and decimated version of the original image. As can be seen in Figure 3.14, where the DWT has been performed on an image which contains a square in the middle, the horizontal sides of the square have been perfectly localized in the HL band, the vertical edges in the LH band and the diagonal corners in the HH band.

These directional, localized events can be used to extract features related to the orienta-

Figure 2.8: (a) Synthetic image $f(i,j)$ (256×256) containing an 80×80 square, (b) DWT of $f(i,j)$ in (a).

tion of texture within the biomedical images. For example, malignant spiculated masses and architectural distortion in mammograms [44], or myelinated tissue fibres and an increase in blood vessels in the retinal images [4] are all examples of oriented texture.

The DWT can also be useful to analyze biomedical image features which are surrounded by and embedded in background noise, such as masses hidden in breast parenchyma, cancerous lesions masked by the tissue wall of the small bowel or the blending of drusens into

the normal background of the retinal images. To explain this, consider the synthetic image

Figure 2.9: (a) Synthetic image $f(i, j)$ (256×256) containing a circle with a radius of 60 embedded in gaussian noise (mean = 0, variance = 1), (b) DWT of $f(i, j)$ in (a).

of Figure 2.9(a) and the corresponding DWT of this image in Figure 2.9. The white circle can be considered as a homogeneous mass, lesion or other pathology which is embedded in a nonuniform/heterogeneous background, like the ones just described for the mammogram, small bowel and retinal images. As can be seen in the high frequency bands (HL , LH , HH), the uniform mass has been segmented from these regions. This is because the lesion has constant contrast which corresponds to low frequency information. Furthermore, a lot of the background noise has been removed in the LL band and the mass is relatively intact. Therefore, if a texture analysis tool is used on these subbands, vital information could be extracted about the uniformity of the mass since a homogenous mass would: (1) create uniform gaps in the high frequency bands and, (2) the mass structure itself would still be present in the low frequency band LL .

Another good example of how the DWT would be advantageous for a biomedical image feature extraction application arises from the multiresolutional properties of the decomposition. As shown in Figure 2.10, the edges of the rectangle are diffused across many subbands

Figure 2.10: (a) Synthetic image $f(i, j)$ (256×256) containing a 50×100 rectangle, (b) DWT of $f(i, j)$ in (a).

and decomposition levels. As the decomposition level increases, the spatial resolution decreases but the frequency resolution increases. Structures within the biomedical images which have low frequency resolution with a large space-resolution and high frequency resolution with small scales can both be realized with this transformation. Since biomedical images have various types of structures, this representation is versatile in dealing with all types of abnormalities present within the images. Furthermore, the diffusion of features or events across subbands, allows features to be captured not only *within* subbands, but also *across* subbands as well.

DWT for Compression

Since medical images are acquired in volumes, a compression scheme dedicated to these types of images must achieve high compression ratios. Therefore, if the DWT is to be useful for a biomedical image compression scheme, the image must be realized in as few non-zero coefficients as possible. This requires the wavelet and scaling basis functions to in some way resemble the image data, which is the case with the wavelet basis functions since they are

found to have the same correlation structure as most nonstationary data [45]. As shown earlier, biomedical images have nonstationary properties. This allows the DWT to provide a more compact representation since it is possible to get an accurate approximation of the image by using only a small number of non-zero wavelet coefficients [45]. By using only a few non-zero wavelet coefficients to describe the biomedical image, a coder such as Huffman [46] [47] or arithmetic [48] [49] may be applied on the coefficients to achieve high compression ratios. As a result of the compact representation that wavelets offer, the DWT has become extremely popular in image coding applications [50] [51] [38].

One of the most important properties of the wavelet transform for compression of biomedical images arises from the features which may be realized with this transformation. For instance, the DWT permits for progressive decoding, which allows the image to be reconstructed incrementally by various resolution types [40] [52]. This paves the way for digital libraries since efficient database retrieval over the Internet from remote locations is possible. Doctors could then access patients' records from their clinic, without having to wait for the report from the hospital or collection facility. Additionally, even local database users would benefit from fast database browsing since the medical images could be viewed as a thumbnail prior to decompressing the image. Other advantageous applications include region of interest (ROI) coding, where a suspicious region may be solely decoded for viewing purposes, and many more.

Chapter 3

Classification of Biomedical Images

3.1 Introduction

As discussed in Chapter 1, the digitization of biomedical imagery paves the way for a multitude of new applications. One such application is Computer-Aided Diagnosis (CAD), where a computer program can act as “a second pair of eyes” to assist the physician with diagnostic results. The results of the CAD system may be used to compliment the doctor’s initial findings as well as point out any suspicious regions which may have been missed. In general, CAD systems aim to: (1) decrease the number of unnecessary biopsies, (2) detect cancer and other diseases in their early stages so that therapies have a greater likelihood of success and (3) generally improve the quality of the diagnosis [53]. All these are beneficial to the patient and would result in a superior form of health care. For such an application to be realized, a feature extraction and classification scheme must be identified which provides results that discriminate between pathologies of the biomedical images.

Another important application arises from the multitude of biomedical images which need to be efficiently stored and retrieved. Text annotations are time consuming and can be prone to human error. If a robust and highly discriminatory feature set is discovered for the biomedical images in question, an automated content-based image retrieval (CBIR) algorithm may be designed, which would allow the images to be retrieved based on their content, not on their text descriptions.

As a way to realize these applications, this Chapter will focus on feature extraction algo-

rithms which exploit textural and other discriminatory characteristics of the mammogram, retinal and small bowel images. Features will be extracted from both the spatial and wavelet domains. The mammogram images are stored in a lossless format, whereas both the retinal and the small bowel images are stored as lossy JPEG images. Feature extraction in the compressed domain has become an important topic recently [54] [55] [56] [57] [58], since the prevalence of images stored in lossy formats far supersedes the number of images stored in their raw, digitized format. Consequently, it will be interesting to see how the feature extraction schemes perform on these images.

In the past, a lot of research has been dedicated to vessel and retinal segmentation as well as the registration of retinal images [59] [60] [61] [62]. As of recently, automated disease detection has become more important and some research has been done in this area. To provide a comparative analysis of the techniques used in this thesis, other research in the area of feature extraction and classification of retinal images is included in Table 3.1.

Previous works for feature extraction and classification of mammogram images is listed in Table 3.2. The main focus is the differentiation between benign and malignant masses, although some work is included which describes the classification of abnormal versus normal mammograms. The authors used a variety of methods which include texture analysis in the wavelet and spatial domains, shape analysis and more.

As there is no other reported works for the feature extraction and classification of bowel images, the research which is discussed in this Chapter is the first reported work in the area. For more information on the technology which collects the small bowel images (PillCamTM), see Section 1.3.2.

Table 3.1: Related works in retinal image feature extraction and classification.

Source	Images	Domain	Features	Classifier	Rate
Wang et al. [32]	54 abnormal (retinopathy), 100 normal	Spatial	Colour	Minimum distance discriminant	¹ 100%, ² 70%
Sinthana-yothin et al. [31]	484 normal, 283 abnormal (retinopathy)	Spatial	Segmentation based on graylevels	Neural networks, physician	³ 80%, ⁴ 71%
Brandon et al. [33]	81 abnormal (62 drusen, 19 miscellaneous), 38 normal	Spatial and Wavelet	Green colour band, wavelet-based signatures, various intensity/density measures and more.	Neural networks, area, region and image level thresholds	⁵ 87%, ⁶ 71%

¹Overall correct classification rate for abnormal (retinopathy) images using enhanced images

²Overall correct classification rate for normal images using enhanced images.

³Overall correct classification rate of abnormal images.

⁴Overall correct classification rate of normal images.

⁵Overall correct classification rate of drusen vs. non-drusen images.

⁶Overall correct classification rate of all types of abnormal images.

Table 3.2: Related works in mammogram image feature extraction and classification.

Source	Images	Domain	Features	Classifier	Rate
El-Faramawy et al. [25]	28 benign, 26 malignant	Spatial, Fourier	Compactness, moments, Fourier descriptors, chord length statistics	Discriminant Analysis	¹ 76%
Rangayyan et al. [26]	28 benign, 26 malignant	Spatial, Fourier	Region-based acutance, compactness, Fourier descriptors, chord length statistics	Discriminant Analysis	² 95%
Pfisterer et al. [63]	40 abnormal (both benign and malignant)	Spatial, Wavelet	Laws texture features, wavelet coefficients thresholds, orientation strength	Radiologist	³ 88%, ⁴ 68%
Lambrou et.al [64]	276 normal and abnormal	Spatial, Wavelet	Angular second moment-sum, entropy-difference, etc.	Minimum distance classifier	⁵ 87%
Mudigonda et al. [65]	28 benign, 26 malignant	Spatial	Gradient/margin sharpness	Discriminant Analysis	⁶ 82%

¹Overall correct classification rate for benign vs. malignant.

²Best overall correct classification rate for benign vs.malignant using acutance only.

³Percentage of true positives (TP) while on average producing 4 false negatives per image using Laws features.

⁴Percentage of true positives (TP) while on average producing 1 false negatives per image using Laws features and wavelet enhancement combined.

⁵Best overall correct classification rate for normal vs. abnormal (malignant, benign, microcalcifications, asymmetry) using the features and classification scheme listed.

⁶Overall correct classification rate for benign vs. malignant using features extracted from mass margins.

The following sections will cover the methods and techniques that will be used to classify the mammogram, small bowel and retinal images. First, the feature extraction problem will be described and the need for invariant features is discussed. To extract shift-invariant features, a translation-invariant DWT is investigated followed by an experiment which proves the shift-invariance of such a representation. The textural properties of the three medical images are examined and several features are extracted. A description of the classification scheme used will also be included, followed by experimental results.

3.2 Feature Extraction Problem Formulation

In a feature extraction problem, it is necessary to quantify important structures, events or clues using highly discriminatory descriptors. The extracted features are then fed into a classifier, which arrives at a decision. For biomedical images, this decision may be related to the diagnosis of the patient. Let $\mathcal{X} \subset \mathbf{R}^n$ represent the signal space which contains all biomedical images with the dimensions of $n = N \times N$. Since the images contained within $\mathcal{X}$ can be expected to have a very high dimensionality (for example mammogram images: $n = 1024 \times 1024$), using all these samples to arrive at a classification result would be prohibitive [66]. Furthermore, the original image space $\mathcal{X}$ is also redundant, which means that all the image samples are not necessary for classification. Therefore, to gain a more useful representation, a *feature extraction operator* f may map the subspace $\mathcal{X}$ into a feature space $\mathcal{F}$

$$f : \mathcal{X} \rightarrow \mathcal{F}, \quad (3.1)$$

where $\mathcal{F} \subset \mathbf{R}^k$, $k \leq n$ and a particular sample in the feature space may be written as a feature vector: $F = \{F_1, F_2, F_3, \dots, F_k\}$. If $k < n$, the feature space mapping would also result in a dimensionality reduction.

Prior to feature extraction, the biomedical images will be transformed to the DWT domain since the primary focus of this thesis is to examine the benefits of MRA for these types of images. Once in the DWT domain, the feature extraction operator f will be applied to the subband samples of the experimental images (retinal, mammogram and small bowel).

In the feature space, the new samples will correspond to a *feature set* which will be used to uniquely characterize the biomedical images. In the case of highly discriminatory features, image objects would map to non-intersecting clusters in the feature space $\mathcal{F}$.

3.2.1 Invariant Features

To provide a diagnosis or insight into the health of a particular patient, discriminatory features are extracted from the biomedical images (retinal, mammogram and small bowel). Although it is important to choose features which provide the maximum discrimination between two or more pathologies, it is also important that these features are robust. A feature is robust if it provides consistent results across the entire application domain [67]. To ensure robustness, the numerical descriptors should be rotation, scale and translation (RST) invariant. In other words, if the image is rotated, scaled or translated, the extracted features should be insensitive to these changes, or it should be a rotated, scaled or translated version of the original features, but not modified [6]. This would be useful for classifying unknown image samples since these test images will not have structures with the same orientation and size as the images in the training set [68].

If a feature is extracted from a transform domain, it is also important to investigate the invariance properties of the transform since any invariance in this domain also translates to an invariance in the features. For instance, the 1-D Fourier spectrum is a well-known translation-invariant feature [69] since any translation in the time domain representation of the signal, does not change the magnitude spectrum in the Fourier domain

$$f(t - \tau_o) \Leftrightarrow F(\omega) \cdot e^{-j\omega\tau_o}, \quad (3.2)$$

for all real values of τ_o . Similarly, scaling in time results in a easily definable reaction in the frequency domain

$$f(\alpha t) \Leftrightarrow \frac{1}{|\alpha|} \cdot F\left(\frac{\omega}{\alpha}\right), \quad (3.3)$$

where α is an integer value. The overall representation achieved by these operations still permits the extraction of a consistent feature set since the shifted or scaled signal is easily

related to the original. By ensuring invariant features, it is possible to account for the natural variations and structures within the retinal, mammographic and small bowel images.

As previously discussed, the DWT offers good space-frequency localization properties, which may be suitable for extracting features from biomedical images. Additionally, the DWT is scale-invariant, which would permit the extraction of features for various sized-pathologies. However, as shown in Section 2.4.2, the DWT is shift-variant, i.e. the coefficients of a circularly shifted image are not a translated versions of the original image's coefficients. For instance, the DWT of an input biomedical image $f(x, y)$ can be shown as:

$$f(x, y) \longrightarrow \boxed{\text{DWT}} \longrightarrow \hat{F}(k_1, k_2, j)$$

where $\hat{F}(k_1, k_2, j)$ are the 2-D DWT coefficients at scale j . A shift of the image will result in a different set of coefficients

$$f(x + \Delta x, y + \Delta y) \longrightarrow \boxed{\text{DWT}} \longrightarrow \hat{F}(k'_1, k'_2, j)$$

where $k'_1 \neq k_1 + a_1 \cdot \Delta x$ and $k'_2 \neq k_2 + a_2 \cdot \Delta y$ for all $(a_1, a_2), (\Delta x, \Delta y) \in \mathbf{Z}$. Since different shifts of the input image results in a completely different set of amplitude values for the coefficients, the extraction of a consistent feature set is difficult [70] [6]. The same space-frequency structure located in two different spatial locations may cause two separate feature sets to be extracted. In order to utilize the DWT for a feature extraction application, a shift-invariant algorithm must be investigated so that a consistent set of coefficients is chosen for all shifts of the input image. For shift-invariant features, it is necessary to utilize a shift-invariant discrete wavelet transform (SIDWT) on the input image $f(x, y)$

$$f(x, y) \longrightarrow \boxed{\text{SIDWT}} \longrightarrow \tilde{F}(k_1, k_2, j)$$

to compute the wavelet coefficients $\tilde{F}(k_1, k_2, j)$. This representation achieved by such a transform would be considered shift-invariant if a shift of the input image $(\Delta x, \Delta y) \in \mathbf{Z}$ results in output coefficients which are exactly the same as $\tilde{F}(k_1, k_2, j)$, or a spatially shifted version of it. This may be shown by

$$f(x + \Delta x, y + \Delta y) \longrightarrow \boxed{\text{SIDWT}} \longrightarrow \tilde{F}(k'_1, k'_2, j)$$

where $k'_1 = k_1 + b_1 \cdot \Delta x$ and $k'_2 = k_2 + b_2 \cdot \Delta y$ for some $(b_1, b_2) \in \mathbb{Z}$. If the coefficients are exactly the same: $b_1 = b_2 = 0$. Therefore, to ensure a dependable feature set is extracted, shift-invariant DWT algorithms are investigated.

3.3 Shift-Invariant Discrete Wavelet Transform (SIDWT)

The shift-variant property of the DWT is widely known and several solutions have been proposed. To achieve a shift-invariant representation, Mallat et. al use an overcomplete, redundant dictionary, which corresponds to filtering without decimation [6] [36]. From the filtered and fully sampled version of the image, local extrema are used for translation invariance since a shift in the input image results in a corresponding shift of the extrema [6] [71]. Since there is no decimation, each level of decomposition contains as many samples as the input image, thus making the algorithm computationally complex. It also requires significant memory bandwidth.

Simoncelli et. al propose an approximate shift-invariant DWT algorithm by relaxing the critical sampling requirements of the DWT [72]. This algorithm is known as the power-shiftable DWT since the power in each subband remains constant. As explained in [36], the shift-variant property is also related to aliasing caused by the DWT filters. The power shiftable transform tries to remedy this problem by reducing the aliasing of the mother wavelet in the frequency domain. The modifications to the mother wavelet result in a loss of orthogonality [73].

The Matching Pursuit (MP) algorithm can also achieve a shift-invariant representation, when the decomposition dictionary contains a large amount of redundant wavelet basis functions [74]. However, the MP algorithm is extremely computationally complex and arriving at a transformed representation causes significant delays [75]. Bradley combines features of the DWT pyramidal decomposition with the *à trous* algorithm [6], which provides a trade off between sparsity of the representation and time-invariance [36]. Critical sampling is only carried out for a certain number of subbands and the rest are all fully sampled. This representation only achieves an approximate shift-invariant DWT [36].

The algorithms discussed either try to minimize the aliasing error by relaxing critical subsampling and/or add redundancy into the wavelet basis set. However, these algorithms either suffer from lack of orthogonality (which is not always an issue for feature extraction), achieve an approximate shift-invariant representation, are computationally complex or require significant memory resources. To combat these downfalls, it is possible to use the SIDWT algorithm proposed by Beylkin, which computes the DWT for all circular shifts, in a computationally efficient manner [35]. Furthermore, the transformation utilizes orthogonal wavelets, thereby resulting in less redundancy in the representation [71]. Belkyn's work has also been extended to 2-D signals by Liang et. al [71] [73] [76] and its performance in a biomedical image feature extraction application will be investigated. To the best of the author's knowledge, this is the first application of such a transform for a medical imaging application.

3.3.1 2-D SIDWT Algorithm

For different shifts of the input image, it was shown that the DWT can produce one of four possible representations after one level of decomposition. These four DWT coefficient sets (cosets) are not translated versions of one another and each coset may be generated in response to one of four shifts of the input: $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, where the first index corresponds to the row shift and the second index is the column shift. All other shifts of the input (at this decomposition level) will result in coefficients which are shifted versions of one of these four cosets. Therefore, in order to account for all possible representations, these four cosets may be computed for each level of decomposition. This requires the LL band from each level to be shifted by the four translates $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$ and each of these new images can be separately decomposed to account for all representations. To compute the coefficients at the j^{th} decomposition level, for the input shift of $(0, 0)$, the subbands LL^j , LH^j , HL^j , HH^j may be found by filtering the previous levels coefficients LL^{j+1} , as shown below:

$$LL_{(0,0)}^j(x, y) = \sum_m \sum_n h_o(m - 2x) h_o(n - 2y) \cdot LL^{j+1}(m, n), \quad (3.4)$$

$$LH_{(0,0)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_o(n-2y) \cdot LL^{j+1}(m, n), \quad (3.5)$$

$$HL_{(0,0)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_1(n-2y) \cdot LL^{j+1}(m, n), \quad (3.6)$$

$$HH_{(0,0)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_1(n-2y) \cdot LL^{j+1}(m, n). \quad (3.7)$$

The subband expressions listed in Equation 3.4 through to Equations 3.7 contain the coefficients which would appear the same if LL^{j+1} is circularly shifted by $\{0, 2, 4, 6, \dots, s\}$ rows and $\{0, 2, 4, 6, \dots, s\}$ columns, where s is the number of row and column coefficients in each of the subbands for the level $j+1$.

The subband coefficients which are the response to a shift of $(0,1)$ in the previous level's coefficients may be computed by

$$LL_{(0,1)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_o(n-2y) \cdot LL^{j+1}(m, n-1), \quad (3.8)$$

$$LH_{(0,1)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_o(n-2y) \cdot LL^{j+1}(m, n-1), \quad (3.9)$$

$$HL_{(0,1)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_1(n-2y) \cdot LL^{j+1}(m, n-1), \quad (3.10)$$

$$HH_{(0,1)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_1(n-2y) \cdot LL^{j+1}(m, n-1), \quad (3.11)$$

which contain all the coefficients for $\{0, 2, 4, 6, \dots, s\}$ row shifts and $\{1, 3, 5, 7, \dots, s-1\}$ column shifts of LL_{j+1} . Similarly, for a shift of $(1,0)$ in the input, the DWT coefficients may be found by

$$LL_{(1,0)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_o(n-2y) \cdot LL^{j+1}(m-1, n), \quad (3.12)$$

$$LH_{(1,0)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_o(n-2y) \cdot LL^{j+1}(m-1, n), \quad (3.13)$$

$$HL_{(1,0)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_1(n-2y) \cdot LL^{j+1}(m-1, n), \quad (3.14)$$

$$HH_{(1,0)}^j(x, y) = \sum_m \sum_n h_1(m-2x)h_1(n-2y) \cdot LL^{j+1}(m-1, n), \quad (3.15)$$

which contain all the coefficients if the previous levels' coefficients LL_{j+1} are shifted by $\{1, 3, 5, 7, \dots, s-1\}$ rows and $\{0, 2, 4, 6, \dots, s\}$ columns. For an input shift of $(1,1)$, the subbands may be computed by

$$LL_{(1,1)}^j(x, y) = \sum_m \sum_n h_o(m-2x)h_o(n-2y) \cdot LL^{j+1}(m-1, n-1), \quad (3.16)$$

$$LH_{(1,1)}^j(x, y) = \sum_m \sum_n h_1(m - 2x) h_o(n - 2y) \cdot LL^{j+1}(m - 1, n - 1), \quad (3.17)$$

$$HL_{(1,1)}^j(x, y) = \sum_m \sum_n h_o(m - 2x) h_1(n - 2y) \cdot LL^{j+1}(m - 1, n - 1), \quad (3.18)$$

$$HH_{(1,1)}^j(x, y) = \sum_m \sum_n h_1(m - 2x) h_1(n - 2y) \cdot LL^{j+1}(m - 1, n - 1). \quad (3.19)$$

Similarly, these subband coefficients account for all DWT representations, which correspond to $\{1, 3, 5, 7, \dots, s-1\}$ row shifts and $\{1, 3, 5, 7, \dots, s-1\}$ column shifts of the input subband LL_{j+1} .

Performing a full decomposition will result in a tree which contains the DWT coefficients for all N^2 circular translates of an $N \times N$ image. At each level of decomposition, the LL band is shifted four times, and for each shift $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, four new sets of subbands are generated. The decomposition tree is shown in Figure 3.1 and each circular node corresponds to only three subband images: HH , LH and HL , since at each level the LL band is shifted and then further decomposed. The number of coefficients in each node (per decomposition level) remains constant at $3N^2$, and a complete decomposition tree will have $N^2(3\log_2 N + 1)$ elements [71]. To compute the DWT for all N^2 translates of the image costs $O(N^2 \log_2 N)$, due to the periodicity of the rate change operators [73].

After a full decomposition is achieved as described above, it is necessary to come up with an organized way to address each of the coefficients. A proper addressing scheme will help to find the wavelet transform for a particular translate (m, n) , where m is the row shift and n is the column translate of the input image. For a path in the tree, which originates from the root, terminates at a leaf node and corresponds to the translate (m, n) , an expression may be developed which considers all row shifts and all column shifts as binary vectors, where each vector entry can be either 0 or 1. Therefore, the binary expansions may be rewritten as

$$m = \sum_{i=1}^{\log_2 N} a_i 2^{i-1}, \quad (3.20)$$

$$n = \sum_{i=1}^{\log_2 N} b_i 2^{i-1}, \quad (3.21)$$

Figure 3.1: Shift-invariant DWT decomposition tree for three decomposition levels.

where a_i and b_i correspond to the binary symbol which represents the row and column shift at decomposition level i , respectively. In order to find the three subimages (HL , HH and LH) which correspond to the translate (m, n) in the K^{th} decomposition level in the tree, it is necessary to find the S^{th} node which corresponds to this shift, as shown below

$$S = 2 \cdot \sum_{i=1}^K a_i 4^{K-i} + \sum_{i=1}^K b_i 4^{K-i}. \quad (3.22)$$

After the three subimages are located within the tree, to ensure that they correspond to the translate of the input by (m, n) , these three images (HH , LH , HL) must be shifted by $(xShift, yShift)$

$$xShift = \sum_{i=K+1}^{\log_2 N} a_i 2^{i-K-1}, \quad (3.23)$$

$$yShift = \sum_{i=K+1}^{\log_2 N} b_i 2^{i-K-1}. \quad (3.24)$$

This new representation computes the 2-D DWT for all circular translates in a computationally efficient manner. To achieve shift-invariance, a method to select a consistent set

of wavelet coefficients which are independent of the input translation must be determined. Such a method would choose a consistent basis set (which corresponds to a particular translate of the original image) and Coifman and Wickerhauser's best basis selection technique [77] may be used.

Best Basis Paradigm

Coifman and Wickerhauser defined a method to choose a set of basis functions, based on the minimization of a cost function $\mathcal{J}$ [77]. The cost function $\mathcal{J}$ is often called an "information cost" and it evaluates and compares the efficiency of many basis sets [66]. Although there are many choices for cost functions, an additive information cost is preferred so that a fast-divide and conquer tree search algorithm may be used to find the best set of wavelet coefficients [71]. A cost function $\mathcal{J}$ is additive if it maps a sequence $\{x_i\}$ to $\mathbf{R}$ while ensuring that the following properties are always true:

$$\mathcal{J}(0) = 0, \quad (3.25)$$

$$\mathcal{J}(\{x_i\}) = \sum_i \mathcal{J}(x_i). \quad (3.26)$$

To choose a consistent set of wavelet coefficients, an entropy cost function $\mathcal{J}$ is used for best basis determination. Entropy gives insight about the uniformity of the coefficients' representation (maximum energy compaction), which may be used for texture analysis. Furthermore, entropy is beneficial since it can achieve additivity [66]. Shown below is the expression of entropy which is minimized:

$$h_r(x) = \sum_i |x_i|^r \log |x_i|^r, \quad (3.27)$$

where r is usually set to 1 or 2.

To choose the best basis representation, we begin at the bottom of the decomposition tree (see Figures 3.1 and 3.2) and work upwards. For each parent node, there are four child nodes, each containing the high frequency subbands of a particular translate. The cost $\mathcal{A}$ of a particular translate $(p, q) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ at some node is computed by summing

Figure 3.2: Best basis selection corresponding to the minimum cost path.

the cost of the individual high frequency subbands for that shift:

$$A_{(p,q)} = \mathcal{J}(LH_{(p,q)}) + \mathcal{J}(HL_{(p,q)}) + \mathcal{J}(HH_{(p,q)}). \quad (3.28)$$

To minimize entropy, the node with the minimum cost for each parent would be selected at every decomposition level. The path which is connected from the root of the tree all the way down to the leaves, is selected as the the minimum cost path, as shown in Figure 3.2. This path corresponds to the DWT of a particular translate and is chosen as the consistent set of basis functions in order to achieve shift-invariance. An analytical proof of the SIDWT can be found in [78]. However, to demonstrate that the representation achieved by the SIDWT provides a consistent set of coefficients for features extraction, several experiments were conducted and are detailed below. These tests will show that a consistent set of features may be extracted from the DWT domain.

3.3.2 Experimental Proof

In order for a SIDWT algorithm to be shift-invariant, all shifts in the input image must result in coefficients which are exactly the same or translated versions of one another. By choosing the minimum cost path, the best basis selection algorithm chooses a particular set of DWT coefficients from the decomposition tree of Figure 3.1 and these coefficients remain *exactly*

the same for all input shifts. The selected coefficients correspond to a particular translated version of the input image that minimizes the cost $\mathcal{J}$ obtained by the best basis algorithm. This shift is denoted by $(\Delta x_{min}, \Delta y_{min})$ and the wavelet coefficients which correspond to this shift are $\tilde{F}_{min}(k_1, k_2, j)$:

$$f(x, y) \longrightarrow \overbrace{\boxed{\text{SIDWT}}}^{(\Delta x_{min}, \Delta y_{min})} \longrightarrow \tilde{F}_{min}(k_1, k_2, j).$$

All other translations of the original image $(\Delta x, \Delta y) \in \mathbf{Z}$ will have the appropriate basis functions chosen by the best basis algorithm (which correspond to a shift of the input (m, n)), such that they are exactly the same as $\tilde{F}_{min}(k_1, k_2, j)$, as in:

$$f(x + \Delta x, y + \Delta y) \longrightarrow \overbrace{\boxed{\text{SIDWT}}}^{(m, n)} \longrightarrow \tilde{F}_{min}(k_1, k_2, j)$$

for $(m, n) \in \mathbf{Z}$. Note that (m, n) is the translation selected by the best basis algorithm for an input shift of $(\Delta x, \Delta y)$. Therefore, in order for this algorithm to be truly shift-invariant, all shifts must be related by:

$$(\Delta x_{min}, \Delta y_{min}) = (\Delta x, \Delta y) + (m, n), \quad (3.29)$$

where

$$\Delta x_{min} = \Delta x + m, \quad (3.30)$$

$$\Delta y_{min} = \Delta y + n. \quad (3.31)$$

The synthetic image $f(x, y)$ which will be used for testing purposes is shown Figure 3.3. The four cosets (output images which correspond to the DWT of $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$) is shown in Figure 3.4. Note that different structures within these images have a visually different appearance (indicating different coefficient sets).

(a)

Figure 3.3: (a) Synthetic image $f(x,y)$ composed of a circle (radius = 20 pixels) and a square (width = 50 pixels). The image is 256×256 .

The SIDWT algorithm was performed on $f(x,y)$ and the best basis algorithm selected the following translate:

$$(\Delta x_{min}, \Delta y_{min}) = (94, 94). \quad (3.32)$$

For several randomly selected shifts, the parameters of Equation 3.29 to Equation 3.31 are computed. If the representation is truly shift-invariant, Equation 3.29 should be equal to the actual found $(\Delta x_{min}, \Delta y_{min})$ shown in Equation 3.32. The results are tabulated in Table 3.3. As shown in the last two columns of Table 3.3, for each input shift $(\Delta x, \Delta y)$, the correct values of $(\Delta x_{min}, \Delta y_{min})$ were calculated as defined in Equation 3.32, proving the representation is shift-invariant. The SIDWT for the four cosets $f(x + \Delta x, y + \Delta y)$, $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ are shown in Figure 3.5. Note that the structures which were noted to be different in Figure 3.4 now have the same appearance.

DWT of image $f(x,y)$

DWT of image $f(x,y+1)$

DWT of image $f(x+1,y)$

DWT of image $f(x+1,y+1)$

Figure 3.4: DWT decomposition of the image $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$. These are the four cosets.

SIDWT of image $f(x,y)$

SIDWT of image $f(x,y+1)$

SIDWT of image $f(x+1,y)$

SIDWT of image $f(x+1,y+1)$

Figure 3.5: SIDWT decomposition of the image $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0,0), (0,1), (1,0), (1,1)\}$.

Table 3.3: Translation parameters of the SIDWT for various translates $(\Delta x, \Delta y)$ of the input $f(x, y)$.

Input shift $(\Delta x, \Delta y)$	Best basis shift (m, n)	$\Delta x_{min} = \Delta x + m$	$\Delta y_{min} = \Delta y + m$
(0,0)	(94,94)	94	94
(0,1)	(94,93)	94	94
(1,0)	(93,94)	94	94
(1,1)	(93,93)	94	94
(10,20)	(84,74)	94	94
(50,3)	(44,91)	94	94
(7,28)	(87,66)	94	94
(4,70)	(90,24)	94	94
(11,40)	(83,54)	94	94
(36,21)	(58,73)	94	94
(91,93)	(3,1)	94	94

Another method of analyzing the properties of the SIDWT can be found by looking at the histograms of the wavelet coefficients. Since the shift-variant DWT has coefficients which are different for each of the fundamental shifts $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, a shift-invariant representation would have the same coefficient histogram for each one of these shifts of the input. Furthermore, the mean, the variance and the % energy per subband would also remain the same. The computed SIDWT coefficients for all four input translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$ was previously shown in Figure 3.5 and the corresponding histograms for these coefficients are shown in Figure 3.6. As can be seen by the visual appearance of these histograms, they appear to be exactly the same, which shows that the representations achieved by the SIDWT is truly shift-invariant. Table 3.4 contains the values of the mean and variance of the DWT coefficients for each shift and Table 3.5 contains the % energy per subband for the first level of decomposition. The mean μ_X , variance σ_X^2 and % energy ε_Y of subband Y are calculated using Equation 3.33, Equation 3.34 and Equation 3.35 respectively, where $p_X(\cdot)$ is the normalized histogram of X , $c(k_1, k_2, j)$ is set of wavelet coefficients at scale j and ε_j is the total energy of all the

Figure 3.6: The normalized histograms for the DWT coefficients computed for the images $f(x + \Delta x, y + \Delta y)$, where $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

wavelet coefficients at level j .

$$\mu_X = E(X) = \sum_i i \cdot p_X(i), \quad (3.33)$$

$$\sigma_X^2 = E(X^2) - E^2(X) = \sum_i i^2 \cdot p_X(i) - \mu_X^2, \quad (3.34)$$

$$\varepsilon_Y = \frac{\sum_{k_1} \sum_{k_2} c^2(k_1, k_2, j)}{\varepsilon_j}. \quad (3.35)$$

As can be seen by the values listed in Table 3.4 and Table 3.5, the mean μ , variance σ^2 and % energy ε per subband remain constant for each shift, which indicates that the distribution of the coefficients are the same.

Although the visual representation of the coefficients shown in Figure 3.5 is the same and the mean μ , variance σ^2 and % energy ε listed in Table 3.4 and Table 3.5 shows us that the global perspective is the same, it is also necessary to verify that each coefficient is exactly *correlated* with the coefficients in the other SIDWT representations shown in Figure

Table 3.4: Mean μ and variance σ^2 of the SIDWT output for translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$

Input shift $(\Delta x, \Delta y)$	Mean μ	Variance σ^2
(0,0)	0.031433	122.2979
(0,1)	0.031433	122.2979
(1,0)	0.031433	122.2979
(1,1)	0.031433	122.2979

Table 3.5: % Energy ε per subband (1 level) of the SIDWT output for translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$

Input shift $(\Delta x, \Delta y)$	% ε (LL_1)	% ε (LH_1)	% ε (HL_1)	% ε (HH_1)
(0,0)	50.075	19.285	19.289	9.3507
(0,1)	50.075	19.285	19.289	9.3507
(1,0)	50.075	19.285	19.289	9.3507
(1,1)	50.075	19.285	19.289	9.3507

3.5. This may be achieved by computing the correlation coefficient $\rho(X_1, X_2)$

$$\rho(X_1, X_2) = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}}, \quad (3.36)$$

where σ_X is the standard deviation of X and $\text{cov}(X_1, X_2)$ is the covariance between two random variables X_1 and X_2 , which is expressed by

$$\text{cov}(X_1, X_2) = \text{E}[(X_1 - \mu_1)(X_2 - \mu_2)]. \quad (3.37)$$

If the correlation coefficient $\rho(X_1, X_2)$ is equal to 1, the two images X_1 and X_2 are an increasing linear combination of one another and therefore are perfectly correlated. In the case of a decreasing linear relationship, $\rho(X_1, X_2)$ is equal to -1. For complete linear independence (and therefore uncorrelation), $\rho(X_1, X_2)$ is equal to 0 and for all other cases, the value of $\rho(X_1, X_2)$ is between -1 and 1. The closer the value of $\rho(X_1, X_2)$ is to either -1 or 1, the stronger the correlation between the variables. Table 3.6 contains the correlation coefficient

$\rho(X_1, X_2)$ which is computed by correlating two sets of DWT coefficients for a particular translate of the input. As shown in this table, the SIDWT provides a shift-invariant representation since the correlation coefficient is exactly 1, indicating that the coefficients are exact replicas of one another. These results are further confirmed by computing the 2-D cross-

Table 3.6: Correlation coefficient $\rho(X_1, X_2)$ for each combination of two SIDWT outputs X_1 and X_2 . The SIDWT outputs are computed from the translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ of the input image $f(x, y)$. Note that X_1 and X_2 correspond to the SIDWT representations achieved by some translate, as dictated by $(\Delta x, \Delta y)$.

	$X_2 = \text{sidwt}_{(0,0)}$	$X_2 = \text{sidwt}_{(0,1)}$	$X_2 = \text{sidwt}_{(1,0)}$	$X_2 = \text{sidwt}_{(1,1)}$
$X_1 = \text{sidwt}_{(0,0)}$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$
$X_1 = \text{sidwt}_{(0,1)}$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$
$X_1 = \text{sidwt}_{(1,0)}$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$
$X_1 = \text{sidwt}_{(1,1)}$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$	$\rho(X_1, X_2) = 1$

correlation sequence which computes the correlation of two sequences for all translations. When a value of the cross-correlation sequence is a maximum, X_1 and X_2 are maximally correlated at the lag which corresponds to the spatial location of the maximum coefficient. For two images or sets of DWT coefficients to be perfectly correlated, the maximum coefficient value must be located in the center of the image, which corresponds to the lag of (0,0). The cross-correlation sequence $\mathcal{R}_{(X_1, X_2)}$ is found by

$$\mathcal{R}_{(X_1, X_2)}(p, q) = E[x_1(i + p, j + q) \cdot x_2(p, q)], \quad (3.38)$$

where p and q are the lags. The results for a few SIDWT image combinations are plotted in Figure 3.7. Since the maximum value is located at the centers of these 2-D cross-correlation sequences, the representations are maximally correlated at a lag of (0,0), therefore proving shift-invariance.

Figure 3.7: The cross-correlation sequence $\mathcal{R}_{(X_1, X_2)}$ between images $f(x + \Delta x, y + \Delta y)$, which are decomposed by the SIDWT for the circular translates $(\Delta x, \Delta y) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

3.4 Feature Extraction for Biomedical Images

In order to analyze the properties of a biomedical image, it may be important to characterize specific events or qualities of the image based on their textural properties. As was discussed in Chapter 1, the mammogram, retinal and small bowel images all contain regions with different textural characteristics. Therefore, the development of a method to evaluate these texture properties may provide discriminating clues about the medical states of each image. In fact, texture has been one of the most important qualities used to classify images [79] [80].

3.4.1 Image Texture

When a textured surface or image is viewed, the human visual system can discriminate between textured regions quite easily. To try and understand how the human visual system can easily differentiate between textures, Julesz defined textons, which are elementary units of texture [81]. Various textured regions can be decomposed using these textons, which include elongated blobs, lines, terminators and more. It was also found that the frequency content, scale, orientation and periodicity of these textons can provide important clues on how to differentiate between two or more textured areas [81] [6]. Additionally, Julesz also discussed how human perception utilizes different parts of the visual system to process global and localized texture elements. The following is a summary of how humans perceive texture:

- Primitive texture units can be used to decompose complex texture into more basic texture elements.
- Human perception allows for differentiation between local and global texture.
- The positions and orientations of textons contributes to the textural information.

These key ideas and fundamental points have opened up a wide-variety of computer vision applications [69] [79] [82], which aim to mimic the properties of the human visual system when discriminating texture. Since many applications and algorithms are being built to assist humans with their everyday tasks, it only makes sense these computing devices understand images the same way as humans do. A robust texture analysis scheme would take into account all of the discussed points (for instance: global and local window sizes which help to understand the orientation, periodicity, scale or frequency content of the primitive texture elements).

In computer vision applications, texture is considered as the information contained in the spatial relationships of gray levels [80]. Textural features may be derived from the spatial distribution or statistical variance of the gray level values and can be used to discriminate between images. However, since many different types of texture analysis techniques exist, it

is necessary to choose analysis tools which are tuned to the textural-characteristics of the image in question.

Typically, texture can be categorized into two groups: (1) (quasi-) periodic, where texture elements are repeated in a (quasi-) periodic manner or (2) random [44]. Random textures are those that do not contain an elementary texture unit (texton). Some examples of (quasi-) periodic texture are a series of bricks in a wall, a group of bubbles or bunch of straw, as shown in Figure 3.8. Examples of random (aperiodic) texture include clouds, trees or the

Figure 3.8: Some examples of (quasi-) periodic texture images. All images are obtained from the USC Signal and Image Processing Institute. (a) Brodatz - brick wall (D94 H.E), (b) Brodatz - plastic bubbles (D112 H.E), (c) Brodatz - Straw (D15).

facial features of a mandrill. Samples of these types of images are shown in Figure 3.9. As can be seen by these images, different regions can be distinguished by their textural properties. For example, the sky in Figures 3.9(a) and (c) is uniform or constant, while the trees are non-homogeneous and granular. The face of the mandrill can also be segmented into different regions according to texture; the facial hair contains fine and oriented texture, while the nose and face are relatively smooth. Other texture qualities include coarseness, patchiness, being mottled and much more [44].

Since textural characteristics of an image provide discriminatory information about specific regions and objects, the application of such tools could be beneficial for biomedical image analysis. Therefore, the textural properties of the experimental images (mammo-gram, retinal, small bowel) will be investigated.

Figure 3.9: Some examples of aperiodic texture images which contain random texture regions. All images are obtained from the USC Signal and Image Processing Institute. (a) Tree, (b) mandrill, (c) sailboat on a lake.

Small Bowel Texture

As the PillCamTM SB moves through the gastrointestinal tract, it is recording a video of the path it travels. There are many organs, structures and other anatomical objects which are viewed by the PillCamTM SB. As was noticed from the experimental dataset, normal images contain smooth, homogeneous texture elements with very little disruption in uniformity except for folds and crevices. This is shown in Figure 3.10 (a) - (e).

As for abnormal small bowel images (benign and malignant), the types pathologies can vary from a polyp to a Kaposi's sarcoma or even a multifocal carcinoma. These diseases may occur in various sizes, shapes, orientations and locations within the gastrointestinal tract. Although there are many types of diseases, abnormalities of the small bowel seem to have some common textural characteristics. Firstly, the diseased region contains many different textured areas simultaneously and secondly, these diseased areas seem to mostly contain heterogeneous texture components. This may be seen in Figure 3.10.

Retinal Texture

Normal retinal images are easily characterized by their overall homogeneous appearance. Other anatomical properties of a normal eye (such as the retina, fovea or veins) are embedded into this homogeneous texture. This is easily seen in Figure 3.11(a)-(c).

Figure 3.10: Small bowel images captured by the PillCamTM SB, which exhibit textural characteristics. (a) Healthy small bowel, (b) normal neocecal valve, (c) normal colonic mucosa, (d) normal small bowel, (e) normal jejunum, (f) small bowel polyp, (g) small bowel lymphoma, (h) GIST tumor, (i) polypoid mass, (j) small bowel polyp.

Unlike normal eyes, eyes which contain disease do not possess uniform texture qualities. Three cases of abnormal retinal images are shown in Figure 3.11(d)-(f). Diabetic retinopathy is characterized by exudates or lesions which are typically found in random whitish/yellow patches of varying sizes, locations [32] with relatively defined margins [31], as shown in Figure 3.11(a). The exudates occur in very dense, semi-homogeneous patches, which possess directional textural elements that can be differentiated from the homogeneous background. Furthermore, blood vessels rapidly grow in the retina to compensate for reduced blood supply caused by the exudates [34]. This may be characterized by oriented texture which is composed of a high concentration of blood vessels. Another clinical sign of diabetic retinopathy are microaneurysms and haemorrhages. Microaneurysms appear as small, red, round dots and haemorrhages can have ‘dot’, ‘blot’ or ‘frame’ configurations [31]. These two types of pathologies are the first clinically detectable signs of diabetic retinopathy [31] and may be detected with a texture analysis tool since they would disrupt the homogeneity of the background.

Another type of eye disease is known as macular degeneration which can cause blindness

Figure 3.11: Retinal images which exhibit textural characteristics. (a)-(c) Normal, homogeneous retinal images, (d) background diabetic retinopathy (dense, homogeneous yellow clusters), (e) macular degeneration (large, radiolucent drusens with heterogeneous texture properties), (f) central retinal vein occlusion (oriented, radiating texture).

if it goes untreated. Macular degeneration may be characterized by drusens, which appear as yellowish, cloudy blobs, which exhibit no specific size or shape and can appear with a variety of characteristics [33]. This is shown in Figure 3.11(e) by the yellow ‘blobs’ which are scattered throughout the eye. These structures are heterogenous and not as dense (radiopaque) as the exudates previously discussed.

Other diseases include central retinal vein and/or artery occlusion and is shown in Figure 3.11(f). Here, it is easy to notice a oriented texture pattern which radiates from a center location.

Mammographic Texture

Some manually segmented mammographic regions containing normal tissue samples, benign circumscribed masses and malignant spiculated masses are shown in Figure 3.12. Histogram

equalized versions of these images are included in Figure 3.13, so that important structures are more visible. The normal regions clearly show that the parenchyma and other glandular tissue possess random textural qualities since they cannot be easily described by an elementary texture unit. Moreover, these normal tissue samples appear to be heterogeneous on a global scale, while localized regions contain homogeneous texture qualities. The benign masses have a rounded appearance with a defined boundary, while the inside of the mass is relatively uniform and radiolucent. This has also been noted by other others, see [20] [26] [65]. In contrast, the malignant masses possess ill-defined boundaries, are of higher density (radiopaque) and have an overall nonuniform appearance in comparison to the benign lesions. Furthermore, spicules from malignant masses cause disturbances in the homogeneity of tissues in the surrounding breast parenchyma [44]. Since benign and malignant masses carry different textural qualities, a system which exploits their textural characteristics will be utilized. Due to the projection nature of mammography [83], overlapping structures and normal breast parenchyma masking the lesions, analysis could be difficult.

3.4.2 Texture Analysis

Using the differences in texture qualities for the types of biomedical images discussed, there needs to unified approach to analyze these images. Typically, there is a 4 problem domain in texture analysis for digital images:

1. Shape from texture,
2. Texture synthesis,
3. Texture segmentation,
4. Texture classification.

Since the end application is to provide discrimination between the biomedical images (retinal, mammogram, small bowel), only texture classification techniques will be discussed.

Haralick defined Graylevel Co-occurrence Matrices (GCMs) as a way to extract textural

Figure 3.12: Manually segmented 128×128 regions of a mammogram. (a)-(c) Normal regions, (d)-(f) circumscribed benign masses, (g)-(i) spiculated malignant masses.

features from an image for classification [80]. GCMs count the the number of two-pixel combinations in the area and are typically normalized so that the matrix may be treated as a probability density function (PDF). Let each entry of the normalized GCM be represented as

$$p(l_1, l_2, d, \theta) = \frac{P(l_1, l_2)}{\sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} P(l_1, l_2)}, \quad (3.39)$$

where $P(l_1, l_2)$ is the number of occurrences of gray level l_1 and l_2 at a distance d and angle θ . Additionally, $\sum_{l_1} \sum_{l_2} P(l_1, l_2)$ is the normalizing factor and L is the maximum number of gray levels in the image. Note that these matrices are symmetric: $p(l_1, l_2, d, \theta) = p(l_2, l_1, d, \theta)$.

Using the GCMs as a probability distribution allows statistical quantities to be computed from the region of interest. Some examples are correlation, inverse difference moment, sum average, sum variance and many more [80] [44]. To exploit the textural characteristics of the retinal, mammogram and small bowel images, it will be necessary to use texture features

Figure 3.13: Histogram equalized version of Figure 3.12

which describe the relative homogeneity or non-uniformity of the images, since these type of texture properties were noticed to differentiate between the normal and abnormal images. The features which will be used are homogeneity (h), which describes how uniform the texture is, contrast c , which measures the amount of contrast variation and entropy e , which is a measure of nonuniformity or the complexity of the texture [44].

Homogeneity

Homogeneity h is computed by the angular second moment

$$h = \sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} p^2(l_1, l_2, d, \theta), \quad (3.40)$$

where $0 \leq h \leq 1$. If the region is homogeneous, there will be only a few rapidly changing graylevel values. As a result, only a small number of high probability values will be found in the GCM matrix, translating into a large h value. Conversely, if the texture is nonuniform,

there will be lots of varying pixel values which would each carry a small probability value. This would cause the value of h to be small.

Contrast

Contrast c may be found by the difference moment

$$c = \sum_{k=0}^{L-1} k^2 \underbrace{\sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1}}_{|l_1-l_2|=k} p(l_1, l_2, d, \theta). \quad (3.41)$$

If there is constant contrast throughout the region of interest, $k = l_1 - l_2 = 0$, which would cause c to go to 0. If there is a lot of rapidly changing pixel values, $k = l_1 - l_2 \gg 0$. Consequently, the value of k^2 would also be large, translating into a large value of c .

Entropy

Entropy e may be found using the traditional information theoretic definition

$$e = \sum_{l_1=0}^{L-1} \sum_{l_2=0}^{L-1} p(l_1, l_2, d, \theta) \log_2(p(l_1, l_2, d, \theta)). \quad (3.42)$$

For complex, random texture structures, the entropy will be large while a completely uniform texture will result in $e = 0$.

Traditionally, features based on the GCM have been extracted from the spatial domain [80] [79]. A weakness of this technique is that it doesn't consider other important qualities which aid in texture discrimination. As Julesz has mentioned in [81], texture discrimination is also dependant on the perception of texture events at different scales, frequencies and orientations. Extracting textural features from the spatial domain can account for oriented structures, since the GCM can be computed at various angles, i.e. 0° can capture horizontal structures, 90° can describe vertical texture components and 45° and 135° for diagonal texture. Furthermore, if the GCM is computed for the biomedical image as a whole, the extracted features would provide useful information about the global texture content. However, to get an even better representation, the other features such as frequency and scale must be considered.

Since the Fourier transform does not permit the extraction of space-localized frequency events, using the Fourier domain for texture analysis would not be sufficient. Although it may reveal which spectral components are present in the image, it would not provide a description of where in the image these events occur. Space-localized frequency structures may be an important clue for differentiating between two textured biomedical images and using the Fourier transform would neglect them from any analysis. As was discussed in Chapter 1, the STFT offers a time-frequency representation, but suffers from various other downfalls, such as a fixed time frequency resolution and computational inefficiency for 2-D signals. Additionally, since an important aspect of texture is scale, the STFT does not offer any insight about this. Therefore, to gain a good representation, texture analysis may be performed in the wavelet domain since both spatial and frequency information is present at different scales. Furthermore, using directional GCMs will also allow for oriented texture features which also exploit scale, space and frequency information.

To capture the global texture content of the biomedical images, GCMs and textural features will be extracted in the spatial domain from the image as a whole. These features will be computed for multiple angles at distance of $d = 1$, for fine texture analysis.

To account for the texture information which is present in both space and frequency for different scales, the texture features will also be extracted from the wavelet domain at various angles at a distance of $d = 1$ to capture fine textural properties. Typically, the DWT isn't used for texture analysis due to its shift-variant property [36]. However, using the SIDWT algorithm described in Section 3.3.1 will permit the extraction of a consistent feature set, thus allowing for multiscale texture analysis.

3.4.3 Multiresolutional Features

To examine texture features at various scales, GCMs $p(l_1, l_2, d, \theta)$ are computed from the wavelet domain for each scale, for a pixel distance of $d = 1$ and several angles θ . Since each subband isolates different frequency components, the directions of the GCMs are selected so that they can capture localized, oriented, edge information. Consider Figure 3.14, which

contains a test image and its DWT for 1 level of decomposition. As shown in the figure, the *HL* band isolates horizontal edge components, the *LH* subband isolates horizontal edges, the *HH* band captures the diagonal high frequency components and *LL* band contains the lowpass filtered version of the original. Consequently, to capture these oriented texture components, the GCM is computed at 0° in the *HL* band, 90° in the *LH* subband, 45° and 135° in the *HH* band and 0° , 45° , 90° and 135° in the *LL* band to account for any directional elements which may still may be present in the low frequency subband. Computing GCMs and features at multiple angles will help to provide a semi-rotational invariant representation. The computed GCMs are summarized in Table 3.7.

From the GCMs listed in Table 3.7, homogeneity h and entropy e are computed for

Figure 3.14: (a) Test image, (b) DWT of (a).

each decomposition level using Equation 3.40 and 3.42. Contrast c isn't computed in the DWT domain since the dynamic range of DWT coefficients is not constant from subband to subband, therefore not permitting for a consistent comparison among features. The series of features generated for a decomposition level j are listed in Table 3.8. Note that the quantity in parenthesis is the angle at which the GCM was computed.

Table 3.7: The graylevel co-occurrence matrices (GCM) computed for each level of decomposition.

Subband	GCM(s) Computed
HL	$p(l_1, l_2, 1, 0^\circ)$
LH	$p(l_1, l_2, 1, 90^\circ)$
HH	$p(l_1, l_2, 1, 45^\circ), p(l_1, l_2, 1, 135^\circ)$
LL	$p(l_1, l_2, 1, 0^\circ), p(l_1, l_2, 1, 45^\circ), p(l_1, l_2, 1, 90^\circ), p(l_1, l_2, 1, 135^\circ)$

For each subband that contains more than one directional feature, these features are averaged so that the features are not biased to a particular orientation of texture. The new features are shown below:

$$\tilde{h}_{HH}^j(45^\circ, 135^\circ) = \frac{1}{2} \cdot (h_{HH}^j(45^\circ) + h_{HH}^j(135^\circ)), \quad (3.43)$$

$$\tilde{e}_{HH}^j(45^\circ, 135^\circ) = \frac{1}{2} \cdot (e_{HH}^j(45^\circ) + e_{HH}^j(135^\circ)), \quad (3.44)$$

$$\tilde{h}_{LL}^j(0^\circ, 45^\circ, 90^\circ, 135^\circ) = \frac{1}{4} \cdot (h_{LL}^j(0^\circ) + h_{LL}^j(45^\circ) + h_{LL}^j(90^\circ) + h_{LL}^j(135^\circ)), \quad (3.45)$$

$$\tilde{e}_{LL}^j(0^\circ, 45^\circ, 90^\circ, 135^\circ) = \frac{1}{4} \cdot (e_{LL}^j(0^\circ) + e_{LL}^j(45^\circ) + e_{LL}^j(90^\circ) + e_{LL}^j(135^\circ)). \quad (3.46)$$

For each decomposition level, the features from each subband are averaged as well so that (1) there is no bias for a particular orientation of texture in the features, (2) the number of features are small so that classification is less prohibitive, and (3) the features may provide a semi-rotational invariant representation. This results in two features per decomposition level j :

$$F_1^j = \frac{1}{4} \cdot (h_{HL}^j(0^\circ) + h_{LH}^j(90^\circ) + \tilde{h}_{HH}^j(45^\circ, 135^\circ) + \tilde{h}_{LL}^j(0^\circ, 45^\circ, 90^\circ, 135^\circ)), \quad (3.47)$$

$$F_2^j = \frac{1}{4} \cdot (e_{HL}^j(0^\circ) + e_{LH}^j(90^\circ) + \tilde{e}_{HH}^j(45^\circ, 135^\circ) + \tilde{e}_{LL}^j(0^\circ, 45^\circ, 90^\circ, 135^\circ)). \quad (3.48)$$

Depending on the size of the image, the number of levels where features are extracted vary. To build a probability distribution (GCM) from a small number of coefficients will result in

Table 3.8: A list of features extracted from each subband for each decomposition level j .

Subband	Homogeneity Features	Entropy Features
HL^j	$h_{HL}^j(0^\circ)$	$e_{HL}^j(0^\circ)$
LH^j	$h_{LH}^j(90^\circ)$	$e_{LH}^j(90^\circ)$
HH^j	$h_{HH}^j(45^\circ), h_{HH}^j(135^\circ)$	$e_{HH}^j(45^\circ), e_{HH}^j(135^\circ)$
LL^j	$h_{LL}^j(0^\circ), h_{LL}^j(45^\circ), h_{LL}^j(90^\circ), h_{LL}^j(135^\circ)$	$e_{LL}^j(0^\circ), e_{LL}^j(45^\circ), e_{LL}^j(90^\circ), e_{LL}^j(135^\circ)$

heavily skewed results since there is not enough samples to gain a true representation of the probability distribution. Therefore, the last three levels of decomposition will be omitted for all the images, regardless of their size. This means that at a minimum, the GCM is being computed from 8×8 blocks.

3.4.4 Spatial Features

To gain insight into the global texture qualities of the medical images, GCMs may be computed for the images as a whole. This allows the features which are extracted to describe the overall homogeneity of the image. Since pathologies within the medical images disturb the homogeneity of the images, extracting global features may also be discriminating.

In the spatial domain, the GCMs are computed for four directions ($0^\circ, 45^\circ, 90^\circ, 135^\circ$), so as not to bias the probability distribution for one particular orientation of texture. All three features previously defined are extracted; homogeneity h (Equation 3.40), contrast c (Equation 3.41) and entropy e (Equation 3.42). For each feature type, a new feature is computed as the average of the features computed for each GCM:

$$F_3 = \frac{1}{4} \cdot (h(0^\circ) + h(45^\circ) + h(90^\circ) + h(135^\circ)), \quad (3.49)$$

$$F_4 = \frac{1}{4} \cdot (e(0^\circ) + e(45^\circ) + e(90^\circ) + e(135^\circ)), \quad (3.50)$$

$$F_5 = \frac{1}{4} \cdot (c(0^\circ) + c(45^\circ) + c(90^\circ) + c(135^\circ)). \quad (3.51)$$

Since a set of features have been defined which will be applied to the medical images (mammogram, small bowel and retinal), a classification scheme needs to be explored so that the discrimination of these features can be evaluated.

3.5 Pattern Analysis/Classification

Using the feature extraction operator f , the image space $\mathcal{X}$ has been mapped to the feature space $\mathcal{F}$, where each image may be described by a feature vector $F = \{F_1^j, F_2^j, F_3^j, F_4, F_5\}$. The individual extracted features $\{F_1^j, F_2^j, F_3^j, F_4, F_5\}$ are those described in Subsections 3.4.3 and 3.4.4, and the overall length of the feature set depends on the number of decomposition levels. Although the dimensionality of this feature space is much less than the dimensions of any of the biomedical images, it is necessary to classify this data into groups, such that a decision can be made about the image data (i.e. whether the biomedical image is normal or abnormal). For each feature set F_i , let $\mathcal{Y} \in \{1, 2, \dots, C\}$ be the class label or category that each of these i feature vectors belong to, where N_C corresponds to the total number of class labels in the given classification problem. For this particular application, N_C is two since discrimination between two types of pathologies is investigated. Therefore, each image may be classified into two groups: $\mathcal{Y} \in \{1, 2\}$. In order to arrive at a classification result, it is necessary to define a *classifier* which maps the feature space into the decision regions $d : \mathcal{F} \rightarrow \mathcal{Y}$.

Prior to choosing a classifier, a series of factors must be considered which are related to the feature set and database size. Firstly, a large number of test samples are required to evaluate a classifier with low error (misclassification) rates [84] since a small database will cause the parameters of the classifiers to be estimated with low accuracy [85]. This requires the biomedical image database to be large, which may not always be the case since acquiring the images can necessitate rigorous ethics reviews and also the number of pathologies may be limited (i.e. data collection will have to continue for a number of years

to get all the desired cases). Secondly, a large number of features used in conjunction with a complex classification rule also requires the number of training samples to be large. Since the biomedical image databases which will be used for experimentation do not all have a large number of samples and the number of features extracted for each image is small, it was shown in [86] that a simple, linear classification scheme is the most effective for this scenario. Additionally, if the extracted features are strong (i.e. the features are mapped into nonoverlapping clusters in the feature space) the use of a simple classification scheme will be sufficient in discriminating between classes. Therefore, linear discriminant analysis (LDA) [87] [66] will be the classification scheme of interest.

In LDA, a training set and a testing set is required. The training set arrives at the decision boundaries which are used to differentiate between classes and the test set uses the predefined rules to arrive at a classification result. To arrive at the discriminating functions, the feature vectors $\{F_i\}_{0 \leq i \leq k}$ of the training set are transformed into canonical discriminant functions D

$$D = F_1 \cdot x_1 + F_2 \cdot x_2 + \cdots F_k \cdot x_k + b, \quad (3.52)$$

where $\{x_i\}_{0 \leq i \leq k}$ are the discriminant function coefficients and b is a constant. The discriminating scores are chosen using an optimization algorithm which maximizes the class separability [87]. Maximizing the class separation is equivalent to maximizing the ratio of the between class scatter matrix to the within class scatter matrix. The between class scatter matrix S_b is defined as

$$S_b = \sum_{i=1}^{N_c} p_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (3.53)$$

where p_i is the a priori probability of class i , N_c is the total number of classes, T is the transpose, μ is the mean of all sample sets as a whole and μ_i is the mean of the i^{th} class. The within scatter matrix S_w is defined as

$$S_w = \sum_{i=1}^{N_c} p_i \cdot \Sigma_i, \quad (3.54)$$

where the Σ_i is the covariance of class i .

To enhance the reliability of LDA, the *Leave One Out Method (LOOM)* is also used. In-

stead of dividing the N input samples into two equal sized sets of training and testing data, one sample is removed from the whole set and the discriminant scores are derived from the remaining $N - 1$ data samples. Then, using these discriminant scores, the left out sample is classified. This procedure is completed for all N samples. Since biomedical image databases are expected to contain a small number of images, using LOOM will allow the classifier parameters to be estimated with higher accuracy, since more samples are used. This is all completed at a cost of increased computational complexity, since the discriminant functions must be recalculated N times [88].

Using the spatial and wavelet features, along with the classification scheme just discussed, the overall performance of the system is tested for all the three biomedical images (mammogram, small bowel and retinal). The results are listed below.

3.6 Results

To test the performance of the proposed feature extraction and classification scheme, all three biomedical images (mammogram, small bowel and retinal) were used. The objective of the proposed system for each of the medical images is as follows:

1. **Small Bowel Images** - Automatically classify various benign and malignant lesions from regions throughout the small bowel tract.
2. **Retinal Images** - To correctly classify between diseased and normal eyes.
3. **Mammogram Images** - To differentiate between benign and malignant tumours of different shapes and sizes.

Table 3.9 lists the characteristics of the biomedical images used for experimentation.

3.6.1 Preprocessing

Prior to computing the SIDWT and extracting features, several preprocessing steps were completed for the mammogram, small bowel and the retinal images, which are discussed below.

Table 3.9: Characteristics of the biomedical images used for experimentation.

	Small Bowel	Retinal	Mammograms
Type	Colour	Colour	Grayscale
Dimensions	256 × 256	700 × 605	1024 × 1024
Pixel Depth	24bpp	24bpp	12bpp
File Format	Lossy (.jpeg)	Lossy (.jpeg)	Raw (.pgm)
Number of Images	75 (41 normal, 34 abnormal (various lesions))	58 (38 normal, 20 abnormal (drusens, retinopathy, central vein/artery occlusion, etc.))	54 (35 benign, 19 malignant)

Retinal and Small Bowel Images

Unlike the mammogram images which are grayscale, the retinal and small bowel images are colour images. To provide a consistent analysis with the mammogram images, grayscale conversion was completed by transforming the images in to the YIQ colour space, where the Y component contains the luminance (grayscale) information and the I and Q components represent the colour information. Only the luminance plane was used in the analysis so as to be consistent with the mammogram images. Since the DWT requires the dimensions of the input image to be a factor of two, zero padding was completed for the retinal images in addition to the grayscale conversion.

Mammogram Images

To try and combat the overlapping structures, masking of lesions by the breast parenchyma and the similar texture characteristics of normal tissue and the benign or malignant masses, a region of 128×128 was manually segmented around the region of interest. This region was large enough to account for all masses in the database used. Additionally, for the same reasons previously mentioned, histogram equalization was performed on the 128×128 region to try and highlight the masses of interest.

3.6.2 Block Diagram

The block diagram of the proposed system for the small bowel and retinal images is shown in Figure 3.15 and Figure 3.16 contains the block diagram of the system used to analyze the mammogram images.

3.6.3 Classification Results

Classification using LDA with the LOOM was performed using the statistical analysis package SPSSTM (Statistical Package for the Social Sciences). The following subsection will present the classification results as a confusion matrix, which shows how many images were

Figure 3.15: System block diagram for the classification of the retinal and small bowel images.

Figure 3.16: System block diagram for the classification of mammogram images.

correctly classified and how many images were incorrectly classified. The results for the small bowel images are shown in Table 3.10, the retinal images in Table 3.11, and the mammogram images are in Table 3.12.

From Table 3.10, small bowel classification results are shown, indicating that the correct classification rates are 80 % (normal as normal) and 70.1% (abnormal as abnormal). Misclassification rates were computed to be 20% (normal as abnormal) and 29.9% (abnormal as normal). An overall correct classification rate of 75% was achieved for the bowel images and an average misclassification rate of 25% was achieved.

As shown in Table 3.11, the classification rates of the retinal images were: 92.1% (normal as normal), and 70% (abnormal as abnormal). Incorrect classification rates included: normal was misclassified as abnormal at a rate of 7.9% and abnormal was misclassified as normal 30% of the time. An overall correct classification rate of 81% was achieved for the retinal images while simultaneously achieving an incorrect classification rate of 19%.

Table 3.10: Confusion matrix containing the number of correctly classified small bowel images as either normal or abnormal.

	Normal	Abnormal
Normal	33 (80%)	8 (20%)
Abnormal	10 (29.9%)	24 (70.1%)

Table 3.11: Confusion matrix containing the number of correctly classified retinal images as either normal or abnormal.

	Normal	Abnormal
Normal	35 (92.1%)	3 (7.9%)
Abnormal	6 (30%)	14 (70%)

Table 3.12: Confusion matrix containing the number of correctly classified mammogram masses as either benign or malignant.

	Benign	Malignant
Benign	21 (60%)	14 (40%)
Malignant	8 (42%)	11 (58%)

As shown in Table 3.12, the percentage of correct classification of benign masses as benign was 60% and malignant as malignant was 58%. Benign masses were misclassified as malignant 40% of the time and malignant masses were incorrectly classified as benign at a rate of 42%. An overall correct classification rate of 59% was achieved for the mammogram images while achieving an incorrect classification rate of 41%.

Comparison and discussions of the above results will be carried out in Chapter 5. In that Chapter, the reasons for the successes and failures of the proposed classification scheme will be investigated in detail.

Chapter 4

Compression of Biomedical Images

4.1 Introduction

As discussed in Chapter 1, the integration of digital mammography into health care facilities has become an important task since the associated benefits of full-field digital mammography make it an attractive alternative to the traditional method of filming. As a result, it is necessary to understand some of the issues which arise from the integration of full-field mammography into hospitals and clinics. For instance, the file size of each mammogram is an important parameter to consider. Since physicians need high resolution images to warrant effective diagnosis, a large number of bits are required to get the desired fidelity. In a telemammography system, where mammograms may be transmitted between hospitals, collection facilities, or mobile devices, large files will decrease system efficiency by requiring too much bandwidth for transmission. Similarly, managing the storage of voluminous mammogram datasets would also pose a significant challenge, especially since medical facilities must keep these records for several years. Additionally, the regular screening programs in Canada [10] and other countries create a large influx of mammograms. Therefore, to combat this, a high performance source encoding algorithm must be identified which reduces the mammogram file size to a more manageable size for easy storage and efficient transmission.

The focus of this chapter is to evaluate wavelet-based compression schemes for digital mammograms to see if the DWT can offer a good enough representation of these images to give way to high coding gains and other beneficial features that aid in database storage

and telemammography applications. The particular wavelet-based compression scheme of interest is the state-of-the-art codec known as JPEG 2000, which has recently received a lot of attention [89] [40] [50]. To evaluate the performance of JPEG 2000, the qualities and features of the JPEG 2000 coding engine must be compared to several other lossless coders for digital mammograms. Unlike the previous chapter, which uses all three images for experimentation, this chapter only applies compression on mammogram images since the other images (retinal and gastrointestinal) are already in compressed form.

Some work on wavelet-based compression schemes for digital mammograms has already been done. For instance, Penedo et al. performed region-based wavelet compression on digital mammograms [90]. The wavelet coefficients generated were supplied to an object-based embedded coder, where either Set Partitioning Into Hierarchical Trees (SPIHT) [91] or Set Partitioned Embedded Block Coder (SPECK) [92] was employed. The embedded bitstream is truncated at several rates and the compression performance is compared to lossy JPEG 2000 and the original SPIHT for the same experimental conditions. The work of Penedo et. al only compares lossy JPEG 2000 and SPIHT for mammogram images and neglects to consider lossless compression. Similarly, Wanigasekara et. al. performs a quality evaluation only for the lossy performance of JPEG 2000 on mammogram images and compares it to the traditional JPEG algorithm based on the Discrete Cosine Transform (DCT) [93]. For 30 mammograms, after comparing the compression rates and corresponding PSNR values of JPEG 2000 and the original JPEG algorithm, it was found that JPEG 2000 outperforms JPEG at all bitrates. Da Silva et. al used Delaunay Triangulation for prediction and the residual image was coded using the JPEG-LS approach [89]. For a database of 10 mammograms, the achieved results were compared to the performance of various lossless JPEG compressors (including JPEG 2000), and it was found that the proposed method offered a significant improvement in compression performance. However, the algorithm is computationally complex (even after some modifications) and the predictor fails in noisy image regions. Additionally, to gain a better understanding of the true compression performance of the proposed method, a larger mammogram dataset is required.

To gain better knowledge of the compression performance of wavelet-based techniques for digital mammograms, the following will be addressed: (1) the codec will be lossless to ensure diagnosing information is not removed, (2) the database size will be large (100 mammograms) to ensure the true compression performance is reflected in the results and (3) a comprehensive overview of both JPEG 2000 and other codecs will be given in order to understand the advantages and disadvantages of such schemes for mammograms.

To investigate the benefits of a source encoding scheme, it is important to understand that every scheme's performance is dependant on various properties of the input image. For instance, a model-based coder could use pixel correlation to arrive at an accurate model description, like a Markov model. Therefore, in a mammogram compression application, it will be important to identify which image characteristic each compression scheme exploits to achieve high compression. It is also important to evaluate the features and qualities of a compressor that benefits both a telemammography and database storage application. Arriving at an optimal compression scheme for mammograms requires an evaluation and investigation of these issues. This is the objective of this work and a strong motivation.

4.2 Image Model

Prior to processing any digital image, it is necessary to classify the source's behaviour. By appropriately classifying an image model for the mammograms, one can understand how certain image processing techniques exploit the unique traits this source possesses. For instance, pixels within a digital image are found to exhibit a high degree of correlation with their neighbouring pixels [94]. By accurately modeling these pixel interdependencies, it is possible to employ image processing techniques, such as compression, which utilizes this property.

A digital mammogram is an information source just like any other digital signal and therefore it can be modeled using either a memory or memoryless source model. If a mammogram is a memoryless source, the occurrence of a particular pixel value does not increase the likeliness of encountering a similar valued pixel within the same region. In contrast, if the

mammogram images are memory sources, each pixel would exhibit a statistical dependency with a series of pixels previously emitted from the source. Since pixels within any digital image are found to exhibit a high degree of correlation with its neighbouring pixels, we can model each mammogram as a memory source.

Regions of a grayscale mammogram are found to be highly concentrated with one shade and gradually fade or blend into another shade, proving the occurrence of one pixel clearly increases the likeliness of encountering a similar-valued pixel in the same neighbourhood. As a result, the mammograms of interest will be modeled as memory sources. Figure 4.1 is a closeup of the breast boundary within a mammogram and is included to illustrate the correlated behaviour which exists between neighbouring pixels.

An excellent statistical description of a memory source is given by a Markov source. A

Figure 4.1: Zoomed in region (14×14) of a mammogram demonstrating pixel correlation.

k th order Markov source models the dependence between symbols as a probability distribution, which considers the relationship between a present symbol with k symbols previously emitted. Furthermore, the current symbol's statistics are completely independent of the symbols emitted prior to these k symbols. Since the mammograms were found to exhibit the properties of a memory source, it is appropriate to model each image as a Markov source.

4.3 Performance Measures

Usually, the compression performance of a source encoding scheme is described by a compression ratio CR [95], where a high value of CR indicates good compression. The calculation of CR is found by

$$CR = \frac{\text{Size before compression}}{\text{Size after compression}}, \quad (4.1)$$

Another measure which quantifies the compression performance of a lossless coder is the bitrate (R). The bitrate (R) specifies the average number of output bits that are used to represent each input symbol. The lower the value of R , the better compression is. The expression for R (in bits/pixel (bpp)) can be found below:

$$R = \frac{\text{Number of output bits}}{\text{Number of input symbols}} \quad (\text{bpp}), \quad (4.2)$$

A quantity which differentiates between the compression speed of two source encoders will be defined here as the compression rate (μ). The compression rate (μ) of a source encoder expresses the number of input symbols that are compressed in 1 second. For real-time applications where (nearly) instantaneous processing is required, this quantity will give an indication as to the relative encoding delays that transmitters or database software would face when using this compression scheme for a mammogram application. Although this value is platform dependant, it will give an indication of the relative computational complexity of each source encoder for a mammogram image compression.

$$\mu = \frac{\text{Number of input symbols}}{\text{Time to encode a mammogram}} \quad (\text{symbols/s}). \quad (4.3)$$

4.4 Source Encoder Requirements

Before a source encoder is deemed to be a favourable and optimal compression scheme for mammograms, a series of factors must be considered. Firstly, because mammograms are used for diagnosis, it is crucial that these images be delivered in their entirety. To guarantee

this, only lossless compression techniques are employed to ensure no loss of information. Secondly, the lossless compressor of choice must offer a high degree of compression, thereby reducing each mammogram to the smallest possible file size. Lastly, if this compression scheme is to be widely used within the medical industry, it must be portable to a multitude of computing devices. Computational complexity and memory requirements can limit the implementation of a particular source encoder on devices with limited resources.

4.5 Lossless Source Encoders for Mammograms

The following subsections describe the features and functionality of several lossless compression schemes for mammogram images. Properties of the mammograms which may be exploited by each scheme will be discussed, as well as other features which assist in database applications or telemammography.

4.5.1 Adaptive Huffman

The adaptive Huffman compressor is a modification of the conventional static Huffman algorithm, where both are variable length coders assigning an integral number of bits. Instead of gathering symbol statistics first, the adaptive Huffman algorithm uses a dynamic tree that counts the probability of occurrence and assigns a codeword to each symbol, in real time. The algorithm is called adaptive because it is “learning” the characteristics of the source as each new input arrives, unlike static Huffman codes which require an initial file scan to obtain a statistical model of the source.

Traditionally, there are two popular versions of the adaptive Huffman scheme: Algorithm FGK [96] proposed by N. Faller, R.G. Gallager and D.E. Knuth, and Algorithm A [97] proposed by J. Vitter. Although both algorithms perform encoding and decoding in a very similar fashion, there is a difference in the way each of these respective algorithm’s dynamic tree is updated and managed. Algorithm FGK depends on the sibling property to update and reshuffle the tree [46]. Vitter’s Algorithm A uses implicit node numbering

and minimizes the tree height and external path lengths in the tree for all incoming symbols [97]. In either algorithm, the tree is updated such that the Huffman tree reflects the current symbol's distribution and all codes are still instantaneous [98], [99].

As stated earlier, adaptive Huffman algorithms (both Algorithm FGK and Λ) manage encoding using a dynamic tree structure. The tree consists of a 0-node, which is the root of the tree and represents the $n - k$ symbols which have not arrived yet, where n is the total message length and k is the number of symbols in the alphabet already accounted for. Every other node in the tree has two child nodes, where each represent two things: (1) a unique symbol (one of the k symbols in the alphabet that have arrived) and (2) the number of times this particular symbol has occurred. The associated codeword is described by the path from the root of the tree to the terminating leaf node which represents the symbol being encoded. As a new symbol arrives, the encoder increments the symbol's corresponding leaf weight, transmits the codeword for this symbol and reshuffles the tree according to the criteria imposed by the particular algorithm. If a symbol arrives which has not been added to the encoding tree yet, it is added to the tree and an escape code is generated and transmitted to the decoder along with the original symbol and its corresponding codeword [99]. Upon receipt of the escape code, the decoder updates the decoding tree to account for the newly encountered symbol and decodes the codeword assigned to it. The key to successfully decoding an adaptive Huffman code is to maintain synchronization between the encoding and decoding trees. An example of both the FGK and Λ encoding trees for the sequence *fgh* is included in Figure 4.2.

For a mammogram compression application, real-time compression proves to be one of the main advantages of adaptive Huffman algorithms. The algorithm requires $O(l)$ time for the encoding and decoding operations, where l is the length of the codeword assigned to the current symbol being processed [96]. For a telemammography application, the overall transmission time can be minimized since no buffering is required (only a single file scan is necessary for encoding). Consequently, as a symbol is encoded, it can be immediately sent to the receiver for decoding, thus permitting both encoding and decoding operations to be

Figure 4.2: Dynamic Huffman trees of (a) Algorithm FGK, (b) Algorithm Δ , both operating on sequence $\{fgh\}$.

performed on-line.

Both the encoder and decoder separately manage their own dynamic Huffman trees, which means that the decoder does not require a source model for decoding. Therefore, unlike static Huffman codes which are heavily dependant on the source's probability model for decoding, adaptive Huffman codes optimize bandwidth utilization by requiring minimal transmission overhead [96]. Furthermore, the encoding and decoding trees must remain synchronized at all times, to ensure correct decoding. If a single bit is received in error, the decoder will lose synchronization with the encoder, causing the rest of the pixels to be decoded incorrectly. Lack of error resiliency is a common problem with many adaptive algorithms. This is a serious problem for a mammography application since such distortion would not permit accurate diagnosis.

As a result of its low computational complexity and minimum memory requirements needed to manage the dynamic tree structures, adaptive Huffman compression could be easily implemented on devices with limited resources, such as a PDA. Portability is an important aspect of a mammogram image application. For breast cancer detection, mammograms have proved to be the best method to identify this disease, making it extremely important that the images are received in a timely manner [10].

Any Huffman algorithm yields the best possible representation using discrete codewords, where each codeword is composed of an integral number of bits. As a consequence of this, all symbol probabilities are being approximated by negative powers of two [47] and can take up to one extra bit per symbol than is actually necessary [49]. Therefore, for highly correlated sources such as mammograms or any other source with an irregular probability distribution (for example: bi-level images), Huffman coding may not be the most suitable compression algorithm.

4.5.2 Arithmetic Codes

For sources with non-fixed distributions, the rough approximation used by Huffman codes will not accurately describe the source's behaviour, thus reducing coding gains. As a way to combat this deficiency, arithmetic compressors perform additional preprocessing to obtain a more accurate model of the source's distribution prior to compression. The more precise the probability estimation is, the better compression will be [48]. By adding a supplementary system block which models the source's symbol frequencies before compression, potentially higher coding efficiencies can be gained.

Another improvement over Huffman codes is the way arithmetic compression schemes assign bits to codewords. Since Huffman compressors use an integral number of bits for each codeword, each symbol will be encoded with at least one bit per symbol. In order to encode more efficiently, arithmetic compressors relax this requirement by using a non-integral number of bits to represent the codeword [49].

As a result of the initial step used to compute the probabilities, arithmetic coders are necessarily comprised of two parts: (1) a probability modeler and (2) the arithmetic coding engine. This is depicted in Figure 4.3.

By obtaining the probability model of the source separately and prior to compression, arithmetic codes provide extra utility by permitting flexibility in the choice of the source model used. As previously stated, compression performance is greatly dependant on the accuracy of the estimation of the source's distribution. Therefore, a model can be cho-

Figure 4.3: Components of a traditional arithmetic compressor.

sen which is more inline with the mammogram's actual Markovian characteristics, possibly giving way to higher coding gains. Generally, a probability model can be either adaptive (gathering source statistics as each symbol arrives), semi-adaptive (using an initial file scan to determine the symbol's frequencies where a higher order Markov model may be used) or non-adaptive, where a fixed model is used for all files. It has been shown that non-adaptive models provide the poorest compression performance [48].

The central principle of any arithmetic code is the use of a floating point interval between $(0,1]$ to represent the entire incoming message. As new symbols arrive, the interval is refined according to a predefined rule and the number of bits needed to represent this interval is increased upon each refinement. For larger message sizes, it can be expected that the final interval will be narrow and consequently will require more precision to represent the entire message.

Initially, the encoding interval E is initialized to the half-open unit interval $(0,1]$. Let the lower end of the interval be represented by L and the higher end by H . Therefore, prior to encoding, the following is true:

$$E = [L, H) = [0, 1), \quad (4.4)$$

or equivalently

$$0 \leq E < 1. \quad (4.5)$$

For the first incoming symbol a_j , the encoding interval E (which is initialized to $(0,1]$), is subdivided into a series of subintervals. Each subinterval represents every possible symbol that may arrive next and the size of the subintervals is proportional to each of the symbol's

probability of occurrence. The symbols' probabilities are those that are computed prior to compression.

Based on the symbol which actually arrives next, this symbol's subinterval is chosen as the new encoding interval E , and the new range may be calculated by

$$E = [L + P_C(H - L), L + P_N(H - L)), \quad (4.6)$$

where P_C is defined as the cumulative probability function and P_N is the next cumulative probability distribution. The cumulative probability function P_C is the sum of the original probabilities of all previously encoded symbols, and next cumulative probability distribution P_N is the summation of both the current symbol's probability and all the symbols already encoded. For an arbitrary symbol a_j , the cumulative probability P_C and the next cumulative probability function P_N may be computed by using the following formulas:

$$P_C = \sum_{k=1}^{j-1} p_k, \quad (4.7)$$

$$P_N = P_C + p_j = \sum_{k=1}^j p_k, \quad (4.8)$$

where p_k is the probability of occurrence for some symbol a_k .

Once the new encoding interval E has been selected, it is then further subdivided into a series of subintervals (as done before) to account for all the symbols which may occur next. Interval refinement continues until all symbols are encoded. The resultant encoded message can be uniquely represented by any real number contained between the upper and lower bounds of the final floating point interval. The final interval's precision depends on the number of refinements the original interval has undergone. Consequently, intervals that require a great deal of precision will also require a large number of bits for representation. An example of an arithmetic algorithm operating on *baccf*, out of a complete alphabet of $\{a, b, c, d, e, f\}$, is shown in Figure 4.4. The corresponding precalculated probabilities and ranges for all symbols of the alphabet are shown in Table 4.1.

The main benefit of an arithmetic compression scheme is its flexibility. Dependant on

Symbol	Frequency	Initial Range
<i>a</i>	0.2	[0,0.2)
<i>b</i>	0.3	[0.2,0.5)
<i>c</i>	0.1	[0.5,0.6)
<i>d</i>	0.2	[0.6,0.8)
<i>e</i>	0.1	[0.8,0.9)
<i>f</i>	0.1	[0.9,1.0)

Table 4.1: Symbol probabilities and initial encoding ranges assigned for the symbol alphabet $\{a, b, c, d, e, f\}$. The values of this table are used in conjunction with the arithmetic coding example shown in Figure 4.4.

the application and the data source to be encoded, different source distributions may be used to generate symbol probabilities, which is important since compression performance is dependant on the approximation of the source's behaviour. In this work, the desired outcome is compression of digital mammograms for a telemedicine-based or mass-storage application. Since the digital mammograms were shown earlier to be memory sources that exhibit Markovian correlation, a sophisticated higher-order Markov model used in conjunction with an arithmetic coder should produce high coding efficiencies. Unlike Huffman codes, which use a fixed probability distribution, maximum compression can be achieved by exploiting the correlation that exists between adjacent pixels of the digital mammograms.

The advantages of a model-based encoding paradigm can be offset by the associated computational complexity and memory requirements needed to manage and create the source model. Model creation can necessitate an additional file scan, which can add a noticeable delay in the compression process for large mammogram files (which are fast becoming the norm). File-scan delays and model buffering will restrict a real-time encoding operation, which is critical in a telemammography system, remote access to digital libraries and other Internet-based applications. A compression scheme solely dedicated to digital mammograms should permit the ease of information exchange and management for diagnosis.

Further implications arise from model management. For large symbol sets, like those of high resolution mammograms (i.e. with 2^{12} gray levels), the time required to search the look

Figure 4.4: Arithmetic coding: interval subdivision process for the encoding of *bacc* for a complete alphabet $\{a, b, c, d, e, f\}$. Initial probabilities and ranges are listed in Table 4.1.

up table to find the probabilities can be substantial, especially if higher order Markovian data models were used. Additionally, the memory required to store such an extensive model may limit the implementation on devices with limited memory resources.

In a telemammography application, both the encoder and decoder require knowledge of the mammogram's data model, which necessitates its transmission over the transmission link. For large models, which can be expected for mammogram images, this will add significant transmission overhead, thus reducing the optimization of bandwidth usage. For a mammogram image compression scheme to be optimal for a medical application, it should permit easy implementation as well as maximum compression benefits for both magnetic storage of voluminous mammogram datasets as well as their transmission.

Not only does model creation and manipulation use up a lot of computational resources, but so does the compression process itself. Most implementations which use full precision arithmetic codes require one multiplication per event, while other popular implementations require two multiplications and two divisions per event [100]. Again, this may limit the implementation of such compression schemes on devices with limited processing power.

Arithmetic codes lack error resiliency, especially when adaptive models are used [100]. A

single pixel received in error will cause the decoder to lose synchronization, propagating the error throughout the entire mammogram image. A severely distorted mammogram would not be useful for diagnosis purposes.

4.5.3 Lempel-Ziv Welch (LZW) Codes

One of the simplest ways to compress data is the dictionary-based approach. When using such a scheme, the objective is to look for repeating patterns in the source data and assign a codeword to each pattern. These patterns and their codes can then be recorded in a dictionary for encoding purposes. In theory, if the source data contains enough repeating patterns and there is enough data, then it is possible to compress this data by replacing repeated patterns with their unique code words. Since these dictionary-based encoding schemes depend on repeating patterns within a source message, it is easy to notice that this type of algorithm relies heavily on correlated data.

The Lempel-Ziv Welch algorithm, more commonly known as LZW, is one such dictionary-based compression scheme. LZW is a derivative of the LZ78 and the LZ77 compression algorithm initially created by Lempel and Ziv in 1977 [101]. Unlike Huffman codes, which assign a variable length codeword to a fixed number of bits, LZW codes assign a fixed number of bits to a variable length string.

Generally, the LZW algorithm is a table search and symbol parsing algorithm. If each unique input symbol is L bits long, the dictionary will initially contain a predefined list of the 2^L possible symbols in the alphabet and the corresponding codewords. Additional string patterns located within input message (besides the initial standard and predefined entries) are dynamically added to the dictionary and are assigned unique, fixed-length codewords.

As an initial symbol a_i is read in from the input, its existence in the dictionary is checked. If there isn't a matching string in the dictionary which includes this symbol as a prefix, the codeword for a_i is sent. Then, the next symbol a_{i+1} is read in, the new extended string $a_i a_{i+1}$ is added to the dictionary and a unique codeword is assigned. On the other hand, if the

symbol a_i is found to be a prefix for another entry of the table, the appended string $a_i a_{i+1}$ is checked for its presence within the existing dictionary patterns. If $a_i a_{i+1}$ is not present and is not a prefix for another string, the codeword is sent for $a_i a_{i+1}$ and $a_i a_{i+1} a_{i+2}$ is added to the dictionary. This process continues until all entries in the dictionary have been filled. For a N -bit dictionary, a fully populated dictionary would contain 2^N unique string patterns. The rest of the input message is encoded using a search (of the dictionary) and parse (of the input) algorithm by using the already present dictionary entries to assign codewords. An example of the dynamic dictionary for the alphabet $\{a, b, c\}$, which would be built to encode the string *aabacaaaba* and the corresponding output codestream is shown in Table 4.2

Since codewords are appointed to a given symbol only if it is found in the dictionary,

Symbol String	Codeword
<i>a</i>	001
<i>b</i>	002
<i>c</i>	003
<i>aa</i>	004
<i>ab</i>	005
<i>ba</i>	006
<i>ac</i>	007
<i>ca</i>	008
<i>aaa</i>	009
<i>aba</i>	010

Coded String	Sent Codeword
<i>a</i>	001
<i>a</i>	001
<i>b</i>	002
<i>a</i>	001
<i>c</i>	003
<i>aa</i>	004
<i>ab</i>	005
<i>a</i>	001

Table 4.2: The LZW dictionary which would be dynamically constructed for the encoding of *aabacaaaba* is on the left. Note that the original alphabet symbols $\{a, b, c\}$ would be initialized in the dictionary prior to any encoding. The right table contains The LZW coded output for *aabacaaaba* using the dictionary shown on the left side.

LZW codes heavily rely on repeating patterns. This can be seen by the coding example shown in Table 4.2. Since only a few repeating strings were found in the input message, 10 symbols were reduced to only 8 codewords. Depending on the number of bits used to represent each input symbol and the number of bits assigned to every codeword, this may actually result in an expansion. As long as the dictionary is long enough and the input

data is highly correlated, LZW compressors should find a large number of repeating patterns to produce high coding gains. However, small dictionary tables and sources which contain only a few recurring patterns will not achieve desirable compression ratios. This is clearly indicated by coding example presented in the tables shown in Table 4.2.

Mammogram images, like many natural images, are highly correlated data sources [94] and are found to contain many areas of constant contrast. Consequently, repetitive pixel patterns can be located within each mammogram and LZW compressors should perform reasonably well for an appropriately-sized dictionary. Compression performance is dependant on the dictionary size for two reasons: (1) larger dictionaries will allow for more repeating patterns to be discovered, thus permitting longer strings of pixel patterns to be encoded with fewer bits and (2) repeating pixel patterns with a longer period than the maximum dictionary buffer length will not be detected, resulting in a less efficient representation. Therefore, choosing the appropriate dictionary buffer length is important.

LZW codes do not require any *a priori* knowledge of the source's statistical characteristics. Unlike Huffman codes which assume a fixed distribution regardless of the message source, the LZW algorithm adapts to the source's statistics, gaining better representation as encoding proceeds [102]. As a result, only a single file scan is needed to encode the entire mammogram image, thereby reducing total encoding times and permitting real-time compression. As soon as a single pixel or string of pixels have been encoded, the codeword can be immediately transmitted over the wired or wireless link in a telemedicine-based application.

As previously stated, model creation, management and transmission has various downsides for a mammogram compression application, such as: increased memory requirements to store the model, increased search times for finding the correct frequency count and obvious transmission overhead required to transmit the model to the decoder. This is especially true for high resolution mammogram images, which are needed since doctors do not expect quantization to remove important diagnosing artifacts from the mammograms. Since both the encoder and decoder of LZW coders manage their own dynamically changing dictionary, only the compressed codestream needs to be transmitted to the decoder [101], thus minimiz-

ing the downfalls associated with model management. Therefore, in a telemedicine-based application, where mammograms may be transmitted between hospitals or mobile devices, bandwidth usage can be optimized.

Another important consideration is the error resiliency of LZW codes. Since the encoder and decoder must remain synchronized at all times to ensure correct decoding, a single code-word received in error can cause serious decoding errors. This is highly undesirable for a medical image application, since mammogram integrity is necessary to ensure correct diagnosis. Bit errors may cause the radiologist to detect microcalcifications or masses that are actually nonexistent, or it also may mask present lesions.

4.5.4 Lossless-JPEG (LJPEG) Compression

The ITU-T (a sector of the International Telecommunications Union (ITU)) and the International Organization for Standardization (ISO) united forming the Joint Photographic Experts Group (JPEG), to create a uniform compression standard for still color and grayscale images. By 1992, JPEG, a lossy compression algorithm based on the Discrete-Cosine Transform (DCT), became an International Standard (IS) for the encoding of still images [50]. Although lossy JPEG compression received the most attention, within the same technical specification, a lossless coding mode was also defined [103]. The lossless JPEG coder (LJPEG) was created to support images with resolutions between 2 and 16 bits per pixel, which is sufficient for processing most digital mammograms.

The LJPEG coder is comprised of two primary components: (1) a predictor and (2) an entropy coder. The entropy coder can either be a Huffman or an arithmetic compressor. If an arithmetic codec is used, an additional preprocessing step is required to calculate the source's distribution. The LJPEG encoder is shown in Figure 4.5.

LJPEG makes use of seven simple linear predictor functions to predict the intensity value for a pixel X using the context shown in Figure 4.6. A prediction P_X is computed using the values of three already known pixel intensities I_A , I_B and I_C and the list of the

Figure 4.5: LJPEG block diagram.

seven possible prediction functions are shown in Table 4.3. Note that the $\lfloor \cdot / 2 \rfloor$ should be implemented using right shift arithmetic operations to ensure reversibility. After an appro-

Figure 4.6: LJPEG context.

priate prediction function P_X is selected and applied, the prediction error ϵ_X for a pixel X is then calculated using

$$\epsilon_X = I_X - P_X, \quad (4.9)$$

where I_X is the actual pixel intensity of pixel X . The error residual is then entropy coded.

Linear prediction schemes predict the value of the current pixel based on the linear combination of the intensities of pixels located in the same neighbourhood (as shown by the predictor functions in Table 4.3). The overall benefit of utilizing a prediction scheme in conjunction with an entropy coder is the removal of redundancy between successive pixels by encoding the differences between the actual pixel value I_X and the predicted value P_X [104]. A good prediction scheme will produce small residuals for the prediction error ϵ_X and the net effect will be a dynamic range reduction of the data to be encoded, thereby requiring

Table 4.3: The seven predictor functions as specified by the LJPEG standard.

Selection	Prediction Function
0	No Prediction
1	$P_X = I_A$
2	$P_X = I_B$
3	$P_X = I_C$
4	$P_X = I_A + I_B - I_C$
5	$P_X = I_A + \lfloor (I_B - I_C)/2 \rfloor$
6	$P_X = I_B + \lfloor (I_A - I_C)/2 \rfloor$
7	$P_X = \lfloor (I_A + I_B)/2 \rfloor$

less bits to encode.

Common linear prediction techniques for 2-D imagery are known to remove majority of the redundancy except for the edge information [105]. The edge information remains as a result of an inaccurate prediction (large prediction residual), which occurs since pixel values at edges change abruptly. As stated, edges create large prediction residuals and this consequently increases the dynamic range to be encoded. A larger dynamic range translates into a larger number of bits for coding, ultimately reducing compression efficiency. Since mammograms contain significant amounts of edge information, including the breast outline, breast tissue and blood vessels, it will be important to see if the lack of edge decorrelation affects compression performance.

According to the LJPEG standard, for each pixel, the residual error ϵ is coded using either modified Huffman or arithmetic compression. The modified Huffman coder collects the prediction errors ϵ with similar magnitude values into categories and builds a code tree for each category [106]. For each coded residual, extra bits are appended to the category's codeword to identify the actual value of ϵ which occurred. For the specific ranges of each ϵ category, the interested reader is referred to [103].

If arithmetic coding is used, ϵ statistics are gathered first using the value of ϵ for pixels immediately to the left and to the top of the pixel in question. Similar to the modified Huffman scheme mentioned above, the differences are grouped into five categories: zero (0),

small positive (+S), small negative (-S), large positive (+L) and large negative (-L). Using these five categories and the context described earlier, 25 different conditioning states can be generated from which probability estimates are calculated. After the model has been calculated, arithmetic compression is performed on the residuals ϵ .

When examining LJPEG's usefulness in a mammogram image compression scheme, one must consider the efficiency of the prediction scheme as well as the advantages and disadvantages of the entropy coder used. If the modified Huffman coder is used, it has similar qualities to those previously explained for the adaptive Huffman compressor, when used in a telemammography and mass-storage application. For instance, Huffman coders use an integral number of bits, which assumes a fixed distribution of the source data. This assumption often leads to poor compression results. Additionally, the modified Huffman algorithm is computationally efficient (like most Huffman codes), allowing for on-line transmission and easy implementation on a variety of devices, even those which lack resources. Since the codebook categories are predefined ahead of time, both the encoder and decoder can store this information, not requiring it to be transmitted over the link. This will reduce transmission overhead.

If arithmetic coding is used, ϵ statistics are gathered first. Therefore, just like any other arithmetic coder, LJPEG would also require an initial scan to gather source statistics. This translates into coding and transmission delays as well as transmission overhead to transmit the model. Furthermore, arithmetic coding by itself is computationally complex. A detailed discussion on the benefits and associated downfalls of both arithmetic and Huffman coding schemes are listed in Sections 4.5.2 and 4.5.1, respectively.

Throughout the coding process, the linear predictor remains the same while encoding the entire mammogram image. A more intelligent system would adapt to the mammogram properties. For instance, one predictor function would process the constant contrast regions, like the background, while the remaining high-frequency regions would be handled with a predictor function dedicated to remove the redundancy of edge information. However, this is not the case with the present LJPEG standard and its non-adaptive nature can be expected

to reduce coding gains.

The last consideration to be discussed is the computational requirements of the predictor scheme. The worst case scenario would arise from a predictor function that considers the largest amount of neighbourhood pixels. In the LJPEG standard, it is possible to choose one of seven predictor functions, where the most complex one uses three adjacent pixels to arrive at a prediction. For a mammogram with dimensions of 1024×1024 , to calculate predictions for each pixel (using the most complex predictor) would take approximately three times as long as the simplest predictor, which uses a single pixel for prediction. An increase in computation time can be accounted to the additional memory accesses and increased number of arithmetic operations. A prediction function that considers a larger number of past pixels will produce a smaller error residual ϵ at a cost of increasing in computational complexity. Of course, any additional encoding time added by the predictor functions are also superimposed with the delays created by the arithmetic or Huffman codes.

4.5.5 JPEG-LS Compression

After some time, the demand for an efficient lossless compression technique encouraged the JPEG group to revisit lossless coding. Consequently, JPEG-LS, a compression algorithm that supplies lossless and nearly lossless compression to continuous-tone, colour and grayscale still images, was born. Nearly lossless compression means an image may be encoded and decoded at a compression rate that is predefined by the maximum allowable error rate [107]. Only the lossless codec will be investigated here to ensure reversibility so that the mammogram is delivered without distortion.

The lossless JPEG-LS compressor combines the powerfulness of both prediction techniques and context modeling in a computationally efficient manner. All model and coding considerations were ultimately made to reflect good compression at very low complexity. The functionality of the JPEG-LS coder can be grouped into three major categories: 1) prediction 2) context modeling and 3) coding. A simplified block diagram for the encoder

of JPEG-LS is shown in Figure 4.7.

Figure 4.7: JPEG-LS encoder block diagram.

Prediction

Using the local gradient information, if it is found that the region of the mammogram contains edge information, the predictor is initiated. The fixed predictor of JPEG-LS is nonlinear and makes use of prior knowledge through elementary edge detecting capabilities [108]. Using the causal template shown in Figure 4.8, this non-linear prediction scheme is listed below:

$$P_X = \left\{ \begin{array}{ll} \min(I_A, I_B), & \text{if } I_C \geq \max(I_A, I_B), \\ \max(I_A, I_B), & \text{if } I_C \leq \min(I_A, I_B), \\ I_A + I_B - I_C, & \text{otherwise.} \end{array} \right\}, \quad (4.10)$$

where P_X is the predicted intensity value of pixel X and I_A , I_B and I_C are the values of pixels A , B and C , respectively. This simple median edge-detector is used for each pixel X within the mammogram to determine whether a vertical or horizontal edge is located in the vicinity. If such an edge is found, the fixed predictor will chose the pixel value that is not geographically located on the edge [109], aiming to decorrelate the edge information. Although mammogram images contain constant regions, like the radiological background, the breast itself is highly textured and contains significant edge information. Reduction of redundancy

in the high frequency regions of the breast will improve compression performance, since the overall dynamic range of prediction residuals will be lower.

After the mammogram's error image is computed, context models are gathered to create

C	B	D
A	X	

Figure 4.8: JPEG-LS: contextual template used for prediction, context and probability model generation.

a statistical model which governs the behaviour of the prediction residuals. The statistics and prediction errors are passed to the entropy coder. It has been shown by [109] that the compression performance of JPEG-LS is strictly dependant on the accuracy of the statistical model (since it determines the entropy coder's output code length) and not necessarily by small values of the prediction error. If high frequency edge information (from the mammogram) was not thoroughly decorrelated using the prediction scheme, JPEG-LS's context modeler will incorporate this information into its statistical model to reduce the overall code length. Therefore, for a mammogram image which is rich with edgy regions, this prediction scheme and context modeler should lead to an efficient representation.

Context Modeling

Once the error image been computed for the mammogram in question, the *context* in which the errors occur is determined. Using the learned contextual information, a probability distribution is then created for each prediction residual $\epsilon_X = I_X - P_X$. This is known as context-based statistical modeling.

It is a widely accepted fact that prediction residuals of continuous-tone images (such as mammograms) are modeled well by a two-sided geometric distribution (TSGD) centered around zero [110]. According to the TSGD distribution, the prediction error is proportional

to $\theta^{|\epsilon|}$, where $\theta \in \{0, 1\}$ determines the decay rate of the distribution. However, due to an inherent DC offset present in error predictions signals [111], the TSGD parametric class $P_{(\theta, \mu)}$ used by the JPEG-LS standard for the prediction error at each context, is found by computing

$$P_{(\theta, \mu)}(\epsilon) = C\theta^{|\epsilon - \mu|}, \quad \epsilon = 0, \pm 1, \pm 2, \dots, \quad (4.11)$$

where μ is the offset parameter and C is a normalization factor.

Using the TSGD premise (and many other simplifying assumptions [111]), the parameters of the probability model for each prediction error (and hence context) are estimated using the local gradients ($g_1 = I_D - I_B$, $g_2 = I_B - I_C$, $g_3 = I_C - I_A$), which govern the statistical behaviour of the prediction errors [111]. Therefore, the error probabilities of each mammogram can be efficiently represented using two parameters, namely the rate of decay and the mean of the TSGD. This allows the higher-order dependencies of each mammogram to be modeled in a simple manner without incurring a large increase in model cost, which is expected for higher-order probability models [111]. Consequently, computational complexity is significantly reduced and the memory requirements needed to store the two parameters (versus a large probability table, like those used in arithmetic codes) is minimal. This efficient representation also allows for fast access to the appropriate statistical parameters and reduced memory requirements. Therefore, for devices with limited hard-disk space and computational power, JPEG-LS can be easily implemented on them, permitting maximum interchange of digital mammograms between doctors and other health care workers. Utilizing the TSGD also facilitates a telemammography application, since transmission overhead is significantly reduced (large probability models do not have to be sent with the encoded mammogram). In a telemedicine scheme for digital mammograms, bandwidth usage will be minimized.

Coding

Using the gradient information, either run-length or probability-based entropy coding is initiated. The Run-Length Encoder (RLE) is activated in “flat” or constant-contrast regions and in areas where the variances of the pixel intensity vary significantly, the entropy coder is employed. This will help to efficiently encode both the high and low frequency information of the mammograms. The particular entropy coder used is a Golomb-Rice coder, which is optimal for exponentially decaying (geometric) probability distributions of non-negative numbers (a one-sided geometric distribution (OSGD)) [110]. In order to match the optimality of the Golomb-Rice coder to the prediction error distribution, which is a TSGD, several provisions are made to make the coder optimal for a *two-sided* exponentially decaying distribution, thus accounting for negative integer values as well [111].

JPEG-LS uses the Golomb-Rice coder to code the prediction error and it consists of two parts: a binary number representing the context and a unary number. These two numbers are combined and then they are encoded. The family of Golomb codes are a subset of Huffman codes that do not require the recourse of looking up values in code tables. For a detailed review of the Golomb-Rice coding algorithm used in JPEG-LS, see [110] or [111]. An important thing to note here is that the Golomb-Rice code is generated “on the fly” and each code word is sent immediately to the decoder for decoding. In a hospital or busy clinic, this efficient scheme will allow the mammogram to be coded on-line, in a real-time environment. Furthermore, only the coded bitstream needs to be transmitted, thereby allowing for efficient bandwidth use in a telemammography setting.

To ensure that the mammogram is delivered to the diagnosing professional with minimum distortion, the compression algorithm should not be extremely susceptible to channel errors. However, this is not the case with JPEG-LS. Firstly, if the unary bit of the encoded word is received in error (i.e. its value is flipped), it can cause complete destruction of the decoded image [112]. Additionally, if the binary part of the encoded word is received in error, this will cause a complete failure during decompression since this represents the context in which the pixel appeared [112]. Therefore, Golomb-Rice codes are heavily reliant on synchronization

between the the decoder and encoder and any loss in synchronization will result in a severely distorted mammogram, which is no longer appropriate for diagnosis.

4.5.6 JPEG 2000 Compression

After some time, JPEG's competitors were coming up with algorithms which offered great features like lossy and lossless compression using a single codestream [107]. Motivated by these algorithms, the JPEG committee proposed JPEG 2000, a brand new algorithm which utilizes the DWT and performs both lossless and lossy compression within one unified coding framework. [52]. Part 1 of JPEG 2000 specifies the standard for JPEG 2000's core coding system and became an International Standard (IS) in December of 2000.

Although JPEG 2000 offers both lossless and lossy coding within one unified framework, this paper only examines JPEG 2000's lossless compression mode to ensure the mammograms are delivered to the physician without distortion. For lossless compression of grayscale mammograms, JPEG 2000's encoder and decoder are shown in Figure 4.9.

Figure 4.9: The block diagram of JPEG 2000 for the compression of mammogram images.

Tiling

Some image processing algorithms become extremely complex when performed on a large image, such as a mammogram. Consequently, these algorithms experience long execution times and require a lot of memory resources to perform specific tasks. To reduce computational overhead and memory requirements of some of the more demanding components of JPEG 2000, the standard permits tiling, which tiles the mammogram of interest into non-overlapping, rectangular segments. Future processing is then completed on the individual tile components reducing both the overall computational complexity as well as the memory requirements of the computing device [113]. This allows maximum interchange between devices with limited memory resources. Therefore, tiling gives doctors and other health care workers the added versatility to manage, transmit and receive mammograms on a variety of devices.

Each tile component can be extracted and reconstructed independently, allowing the physician or other sections of the codec to have random, spatial access to specific portions of the bitstream. Diagnosing professionals can then extract a suspicious region within the mammogram, zoom in and examine it more closely, all without decoding other portions of the compressed bitstream. Each tile component can be extracted and reconstructed independently, giving the user or other sections of the codec random, spatial access to specific portions of the bitstream. This is useful to doctors if a specific region within a mammogram needs to be zoomed in and examined more closely. Tiling is demonstrated in Figure 4.10, where an input image of arbitrary size has been segmented into 20 tiles, $\{T_1, T_2, T_3, \dots, T_{20}\}$.

DC-Level Shifting

Grayscale, mammogram images contain pixels that only take on unsigned values. Since a DC component does not affect the gray-level variance of an image, it can be removed by DC-level shifting.

To remove a DC bias from the mammogram data, which uses L bits of precision, the

Figure 4.10: Tiling of an image into 20 non-overlapping rectangles.

value of 2^{L-1} must be subtracted from every pixel value. In fact, the JPEG 2000 coding engine expects the input mammogram's nominal range to be centrally located around zero, so that the coder can make many simplifying assumptions about context models and numerical overflows [52].

Discrete Wavelet Transform (DWT)

Unlike its predecessor JPEG-LS, JPEG 2000 makes use of a state of the art wavelet decomposition method to decorrelate the image data prior to entropy coding. In a compression application, the DWT is beneficial since it can efficiently represent nonstationary images (such as digital mammograms) using only a few nonzero coefficients [82]. Therefore, the dynamic range of the DWT coefficients is significantly less than that of the original mammogram, requiring less bits for representation. For specific details on the implementation of the DWT used for lossless compression in JPEG 2000, the interested reader is referred to [51]. An important thing to report, however, is that JPEG 2000 uses the 5/3 wavelet for decomposition. This 5/3 wavelet was also used for feature extraction (Chapter 3).

In terms of compression efficiency and other utilities in a mammogram application, a fundamental and key addition to the JPEG 2000 algorithm is the DWT. As discussed, the DWT is a signal decomposition technique which localizes non-stationary events within the mammogram and organizes them into subbands. Wavelet basis functions have been shown

to have good time-frequency resolution, allowing for the efficient representation of localized time-frequency structures. These structures may provide important clues about the health of the patient's breast. For instance, the DWT is employed separably along rows and columns, yielding subbands which reveal information about localized horizontal, vertical and diagonal edge components within the mammogram. Edge information can aid in spiculated mass detection and segmentation within mammogram images. Furthermore, the DWT offers an excellent multi-resolution representation, which can be ideal for texture analysis [114]. Localized texture features can be used to discriminate between benign and malignant masses or microcalcifications. Another important point to mention is that large valued wavelet coefficients offer insight into the sharp transitions present in the original mammogram. This can be used for microcalcification detection or benign and malignant mass segmentation, which would help a radiologist to detect areas of concern within the breast tissue. Therefore, in addition to providing a compact representation, the DWT coefficients can serve as inputs into a CAD system, where mass identification and classification for the mammograms can be performed.

Spatially consistent regions of a subband (at the same decomposition level) correspond to the same spatial location within the original image. Therefore, correlation not only exists between adjacent wavelet coefficients, but also between coefficients across subbands. To aid in error resiliency, which is needed for a telemammography scheme, the entropy coder only utilizes the relationships that exist between wavelet coefficients within a subband [115].

An example of 3-scale (dyadic) wavelet decomposition for a mammogram image can be seen in Figure 4.11, where contrast enhancement was performed for visualization purposes. Notice how the mammogram is still visible in the LL_3 subband, indicating most of the energy or information is contained within the lower frequencies. As shown by the figure, the output of the DWT is a series of subband samples that have a probability distribution that is heavily skewed towards lower amplitude values [115] and therefore, should give way to good compression results since the average dynamic range of the mammogram is significantly reduced. This would allow fewer bits to be used.

Figure 4.11: Placement of DWT coefficients for three levels of decomposition and the corresponding mammogram output.

Quantization

Part 1 of the JPEG 2000 standard uses scalar quantization with a dead-zone about the origin and adheres to the formula:

$$q_b(u, v) = \text{sign}(a_b(u, v)) \times \left\lfloor \frac{|a_b(u, v)|}{\Delta_b} \right\rfloor, \quad (4.12)$$

where $a_b(u, v)$ is a wavelet coefficient located at (u, v) of subband b , Δ_b is the quantization step of subband b and $q_b(u, v)$ is the respective quantized output.

For lossless compression, Δ_b is defaulted to the value of 1, essentially bypassing the quantization operation. This ensures no information loss and the mammogram will be delivered to the medical professional in its entirety.

Bit-Plane Coding

After the mammogram has been transformed and quantized, the wavelet coefficients are passed to a bit-plane coder, where they are first segmented into non-overlapping rectangular blocks. Three of these spatially consistent rectangles, known as code blocks, at different

resolution levels, one in each subband, constitute a packet partition location or a precinct as shown in Figure 4.12. Each precinct is coded sequentially.

Precincts give random (spatial) access to the bitstream at a much finer granularity

Figure 4.12: Packet partition location (precinct) and codeblocks of a single tile or for an entire image.

than tiling, which aids in efficient memory implementations for devices that lack memory resources [40]. This allows mammograms to be exchanged freely between many different devices. Additionally, packet partition locations are utilized in a progressive transmission application, where mammograms are viewed incrementally as they are arriving, instead of viewing the image after it has been completely received. To facilitate the transfer and remote archival and retrieval of mammograms, progressive transmission utilities can be set up to allow the users of the system to maximize their bandwidth efficiency. This would make remote access to digital libraries for mammogram images an even more suitable option.

A bit-plane coder gathers contextual (statistical) information about the input data and organizes the data according to significance. Each code block within the precinct is visited in raster order and each is independently coded using bit-plane coding. For each bitplane within the codeblock, three passes are initiated: the significance pass, the refinement pass and the cleanup pass. Each pass aims to determine the significance of the bits and orders them according to their determined significance. For a detailed description of the bit-plane

coding algorithm, which is a modified version of the EBCOT algorithm, see [116]. The Embedded Zero Wavelet (EZW) algorithm [117] and Set Partitioning in Hierarchical Trees (SPIHT) [91] are the algorithms in which EBCOT is derived from.

The state-of-the-art embedded code proposed by Taubman (EBCOT) is a key element of JPEG 2000 since it allows the mammogram to be reconstructed incrementally, either by frequency resolution, spatial resolution or quality levels. This allows devices with limited display and memory resources to view the mammograms, aids in remote digital library access, progressive transmission applications and fast database browsing.

During the bit-plane coding process, samples which become significant give rise to the determination of many other significant coefficients, which tend to propagate along joined image features such as edges [115]. In a second reader system, this could be used for texture analysis or edge detection of suspicious masses of microcalcifications.

Entropy Coding

The entropy coder employed by JPEG2000 is a context-based adaptive binary arithmetic coder known as the MQ-coder, as specified by the JBIG-2 standard [118].

The block diagram for a binary adaptive arithmetic encoder is shown in Figure 4.13, where D is the output symbols (binary decisions) from bit-plane coding, CX is the context developed for that symbol during one of the bitplane coding passes and CD is the resultant, compressed output data. CX is used to select the probability estimate that will be used to code D .

Figure 4.13: The components of a binary adaptive arithmetic encoder.

interval subdivision of Elias coding. This topic is treated extensively in [118] and [119].

JPEG 2000 uses a small number of contexts, which allows for rapid probability adaptation and reduces the cost of independently coded segments [40]. The context bins are re-initialized at the beginning of a new code block and the arithmetic code word is terminated at the end of each code block. This increases error resiliency since errors are isolated or contained within the codeblock. In a telemedicine-based application, where mammograms may be transmitted wirelessly, the mammograms would be subjected to channel errors which could cause serious decoding errors. This is especially true since the MQ-coder relies on previous states for proper decoding.

Packet and Layer Formation

After each codeblock has been compressed via bitplane and entropy coding, the compressed bitstream is organized into packets. A packet includes the entire compressed bitstream from every codeblock in a packet partition location.

If a series of packets are grouped together, one from each packet partition location at each resolution level, a layer is formed. Since a packet can be thought of as a single increment of quality for one resolution level at one spatial location (since packet partition locations roughly correspond to spatial locations), a layer can be thought of as one quality increment for the entire image, at full resolution.

If all layers contained within the image are grouped together, they form the final code stream, where each layer sequentially and monotonically improves the image quality. Layers provide the decoder with a mechanism to decode each codeblock in some predefined order, which is useful for progressive decoding [50]. The code stream and its elementary components are illustrated in Figure 4.14.

Figure 4.14: Organization of packets, layers, codeblocks and code stream

Remarkable Features

The remarkable features of JPEG 2000 are plentiful and make it a very attractive compression scheme for mammogram images. Some of these features include: region of interest (ROI) coding, SNR and spatial scalability, error resilience, and a new file format which permits the inclusion of content-based descriptions.

ROI coding allows a user-defined region within the mammogram to be encoded with higher fidelity than the background. This ROI may be extracted and later processed using a feature extraction and classification algorithm to determine the type of mass or masses present within the breast tissue.

Spatial scalability can be utilized for fast database accessing applications since the user can view each mammogram as a thumbnail prior to selecting the mammogram of interest. Spatial scalability can also deliver different resolutions to other computing devices depending on their display and bandwidth capabilities [51].

SNR scalability is great for Internet-based applications since the user has the option to view and reject the image as it is arriving. Therefore, bandwidth usage can be optimized and remote digital libraries can be an efficient way to access mammograms.

Error resiliency is achieved at the entropy coding level and the packet level. Arithmetic coding by itself is very susceptible to channel errors since a single error can cause a loss of

synchronization at the decoder. By independently coding each code block, the error will be restricted to that particular code block, allowing the rest of the mammogram to be decoded successfully.

At the packet level, an additional packet is defined and created with a resynchronization marker to allow for spatial partitioning and resynchronization, providing a simple solution to keep track of or locate any packets.

The new file format of JPEG 2000 permits the inclusion of metadata, which efficiently presents details regarding image properties, content and vendor information. Metadata enables the user to access this information without having to view the image. For instance, a content-based database archival and retrieval algorithm based on feature extraction and classification of malignant and benign masses can sort, locate and store mammograms based on their diagnosis. Additionally, the metadata may contain text indicating the date, diagnosis, type of treatment expected and the patient's name. This would be ideal for a mammographic imaging application since it would warrant efficient image archival and fast text-based searches in a database, permitting doctors or other healthcare workers to search for a patient's name or diagnosis, without having to search the entire database to locate the mammogram of interest.

4.6 Compression Results

All compression tests were conducted on a 2.4GHz processor (desktop computer) and were supplied the same set of 100 mammograms, which included 22 benign, 8 malignant and 70 normal cases. As described in Chapter 3, each mammogram has a resolution of 12 bpp (grayscale) and is 1024 pixels wide and 1024 pixels high.

To verify which lossless compression scheme is optimal for a mammogram compression application, experiments were conducted using various software implementations. The lossless source encoders tested were: JPEG 2000 (Jasper implementation [120]), JPEG-LS (LOCO-I implementation [121]), lossless-JPEG (LJPEG) with Huffman coding, adaptive Huffman

(Algorithm A implementation), arithmetic with a zero order and a first order probability model and Lempel-Ziv Welch (LZW) with 12 and 15 bit dictionaries. The average compression results for the 100 mammogram data set are shown in Table 4.4 and plotted in Figure 4.15 and Figure 4.16, where the compression ratios CR were calculated using Equation 4.1, the bitrate R was calculated using Equation 4.2 and the compression rate μ was computed using Equation 4.3. It should be noted here that good compression is dictated by large values of CR or small values of R and computationally efficient schemes are indicated by high values of μ . As is shown in Table 4.4, Figure 4.15 and Figure 4.16, JPEG 2000 offers the best compression ($CR = 9.319$, $R = 1.288$ *bpp*) but suffers from computational complexity ($\mu = 501.3$ KSymbols/s). Detailed discussions of the compression results can be found in Chapter 5.

Table 4.4: Average compression results for 100 mammograms.

Source Encoder	CR	R (bpp)	Compression Time/Image (s)	μ (KSymbols/s)
Adaptive Huffman	2.454	4.891	0.529	1936.6
Arithmetic Order 0	2.219	5.410	0.503	2035.3
Arithmetic Order 1	2.753	4.360	3.339	306.6
LZW 12bit	2.636	4.553	0.442	2319.1
LZW 15bit	4.675	2.567	0.316	3241.1
LJPEG	2.580	4.653	0.724	1413.6
JPEG-LS	9.151	1.311	0.135	7585.2
JPEG 2000	9.319	1.288	2.043	501.3

Figure 4.15: Compression ratios CR and bitrates R for 100 mammograms. Note that the x axis corresponds to a particular coder: 1 - Adaptive Huffman, 2 - Arithmetic Order 0, 3 - Arithmetic Order 1, 4 - LZW 12bit, 5 - LZW 15bit, 6 - LJPEG, 7 - JPEG-LS, 8 - JPEG 2000.

Figure 4.16: Compression time/image and compression rate μ for 100 mammograms. Note that the x axis corresponds to a particular coder: 1 - Adaptive Huffman, 2 - Arithmetic Order 0, 3 - Arithmetic Order 1, 4 - LZW 12bit, 5 - LZW 15bit, 6 - LJPEG, 7 - JPEG-LS, 8 - JPEG 2000.

Chapter 5

Conclusions

5.1 Discussion of Classification Results

A unified feature extraction and classification scheme was developed using MRA techniques for biomedical images. Textural features were extracted from the spatial and wavelet domains in order to obtain both global and localized numerical descriptors. As described in Section 3.4.1, perception of texture requires analysis of spatial, scale and frequency properties of the texture elements. To ensure the DWT representation was suitable for consistent extraction of features, a shift-invariant discrete wavelet transform (SIDWT) was computed. As per the author's knowledge, this is one of the first applications of the SIDWT for medical image analysis.

To evaluate the classification performance of the purposed system, three groups of real-world images were used: small bowel images, retinal images and mammograms. Given below are discussions of the proposed system's performance on each type of image.

5.1.1 Small Bowel Image Classification

As discussed in Section 3.6, a total of 75 abnormal and normal bowel images were used for experimentation and these images were correctly classified (normal as normal and abnormal as abnormal) at rate of 75.1%, while achieving a misclassification rate of 24.9% (see the confusion matrix in Table 3.10). These images are snapshots of a video taken by the PillCamTM

SB as it travels through the gastrointestinal tract. Out of 75 images, 41 contained various normal regions (pylorus, jejunum, ileo-cecal valve, etc.), and the other 34 contained several benign and malignant pathologies (polyps, carcinoid tumours, submucosal masses, sarcomas, etc.) all within different regions of the gastrointestinal tract. It is very interesting to see that the classification rates are so high even though: (1) the angle of the camera is different from image to image, (2) the abnormal and normal images which were compared came from different regions within the gastrointestinal tract, (3) the pathologies were not restricted to a specific type, but in fact included many diseases and (4) the masses and lesions were of various sizes and shapes. The success of the system can be accounted to several factors. As was shown in Figure 3.10 and explained in Section 3.4.1, the abnormal regions contained either global or localized heterogeneous texture elements, where as the normal images were relatively uniform. Therefore, the SIDWT provided the means to extract consistent localized texture properties from the images in question, which helped to describe the degree of homogeneity found in localized areas. Additionally, features extracted from the spatial domain took into consideration the global heterogeneous qualities. The combination of these two features resulted in the differentiation between most of the normal and abnormal images, even though the texture qualities of the abnormal images were both local and global.

Another important factor which can be attributed to the success of the proposed classification system on the small bowel images is that masses and lesions of varying sizes were captured by the DWT, due to its scale-invariant property. Even though pathologies can come in various sizes, the DWT captured all of them, regardless of the size.

The reason that the system was relatively robust to the angle of the camera is due to the fact that features were collected at various angles (0° , 45° , 90° , 135°) and then averaged. This offered a semi-rotational invariant representation which could account for the textural properties at various angles within the gastrointestinal tract.

Since this is the first work in the area of small bowel image classification, the results are promising and show great potential for applications such as CAD and CBIR. This is especially true since all features were extracted in a fully-automated manner without any in-

tervention or assistance from a gastroenterologist. This means that such a system could in fact be used as a tool which could either (1) sort the 8 hours of film and highlight suspicious regions or (2) automatically retrieve a specific region or mass, without having to use text annotations.

5.1.2 Retinal Image Database

The retinal image database contained 38 normal images and 20 abnormal images. The results are listed in Table 3.11, where it is shown that the classification rates are quite high even though there is a number of different pathologies present in the group of abnormal retinal images. The average correct classification rate was found to be 81% while a misclassification rate of 19% was obtained. Some diseases/pathologies which are present in the retinal images include: central vein/artery occlusion, retinopathy, large drusens and more. Although the number of different pathologies which were grouped as abnormal were numerous, the proposed system could still differentiate between normal and abnormal images with a high success rate. This could be attributed to reasons similar to the ones just described for the bowel images. For instance, the retinal images were successfully classified by the system since both local and global homogeneity measures were considered. As shown by Figure 3.11, abnormalities seem to possess relatively nonuniform texture qualities both on large scales (artery occlusion) and in localized regions (retinopathy, drusens), which were properly characterized by the selected feature set.

Additionally, the scale-invariant property of the DWT also aided in extracting features for all types of pathologies, regardless of the size, which is important since the possible set of pathologies of the human eye do not come in a predefined size. Since many of the abnormal retinal images possess oriented texture elements (central vein/artery occlusion), the overall success of the system can also be accounted to the fact that the oriented texture properties were properly characterized as a result of (1) computing the features for different angles to characterize the oriented textural properties (both global and local) and averaging them to

obtain a semi-rotational invariant representation, (2) extracting features from the 2-D DWT, which isolates localized directional texture elements as a result of the directional filtering operations.

As shown by other works in the area of retinal imaging (Section 3.1), the authors' schemes were tuned to only detect a specific type of abnormality (i.e. drusens OR retinopathy) and use predefined classification thresholds which may vary from database to database. Additionally, if each of the corresponding literature is read in detail, it will be noticed that a number of processing steps are taken prior to feature extraction (i.e. blood vessel segmentation, image thresholding, neural networks). In the proposed work, no prior segmentation algorithms or other preprocessing is performed (aside from grayscale conversion) and the major success of this scheme is that numerous types of pathologies (not just one) are differentiated from the normal images, with high classification rates. These results could also be applied on other databases to get similar results.

The good classification results (for various types of abnormalities) of the proposed system for retinal images are an indication that the extracted textural features are providing excellent discrimination between the normal and abnormal retinal images. This would allow such a system to be implemented in an ophthalmologist's office, either as a CAD application or for a CBIR scheme.

5.1.3 Mammogram Images

For a total of 54 mammogram images, which contained 35 benign and 19 malignant masses of all sizes and shapes, the proposed system achieved a classification rate of 59% and a misclassification of 41%; see the confusion matrix of Table 3.12. Unlike the other two images, the classification rates for the mammograms are unfortunately quite low. Although contrast enhancement and manual region segmentation was performed, the overlapping structures and breast parenchyma must have obstructed the true behaviour of the masses. Additionally, some normal tissue patterns have the same properties as some of the lesions in the

database which would have caused the features to be similar for both types of pathologies. This results in high misclassification rates since the features are not providing good enough discrimination for these scenarios. The correctly classified regions can be accounted to some of the same reasons which were previously discussed for both the retinal and small bowel images. For instance, global and local homogeneous texture properties were expressed in the spatial and wavelet domain feature set, which captured both the overall properties of the mass, as well as the surrounding tissue (this would sometimes be helpful since some masses cause distortion or nonhomogeneity in the tissue surrounding them). The scale-invariant DWT representation was utilized to capture all mass sizes and the isolation of directional edge components at various angles could have helped to characterize the malignant lesions with radiating spicules, since they possess oriented texture qualities. In general, the overlapping and background tissue, which possessed both homogeneous and nonhomogeneous qualities, were too significant and caused a severe degradation of classification performance.

As was shown in Table 3.2, some other authors' have achieved good results in a mammogram feature extraction application. This is because some of the authors' either (1) incorporated shape information into the analysis, (2) used mass margin information, (3) used a radiologist for classification (not an automated classifier), (4) only compared normal to abnormal (which includes both malignant and benign). Since the proposed system only tried to describe the textural properties to differentiate between benign and malignant lesions in conjunction with an automated classifier, the feature set must not have been robust enough to classify the abnormalities which are embedded in and interleaved with normal breast parenchyma.

5.1.4 Overall Discussion on Classification Results

Overall, any disruption of the classification rates for all three images can be accounted to cases where there is a lack of differentiation between the texture uniformity of the abnormal images and normal images. It remains impossible to perfectly characterize the properties of

all abnormalities, since they can come in any shape or form. Additionally, normal tissue and regions can also sometimes assume the properties of abnormal regions. For instance, for a patient who has many veins in their eyes, these regions could be characterized as abnormal since the concentration of normal veins may be considered as oriented heterogeneous texture elements, which is typically classified as abnormal. Furthermore, a small bowel image may be of a region that has more than the average amount of wrinkles and folds, which would also cause the system to fail, since it would possess similar nonuniform properties as abnormal regions.

Another important consideration about the classification results arise from the sizes of the databases. As was stated in Section 3.5, the number of images used for classification can determine the accuracy of the estimated classifier parameters. Since the number of images which were used in each test scenario were relatively small, any degradation of classification performance may also be accounted to the lack of proper estimation of the classifiers parameters. Additionally, finding the right trade off between number of features and database size is an ongoing research topic and has yet to be perfectly defined [86] [85] [84].

A last point for discussion is the fact that features were successfully extracted from the compressed domain in the retinal and small bowel images. Since many forms of multi-media are being stored in lossy formats (.mp3, .mpeg, .jpeg), it is important that classification systems may also be successful when utilized in the compressed domain.

5.2 Discussion of Compression Results

From Table 4.4 and Figure 4.15, it is easily noticed that JPEG 2000 and JPEG-LS both offer a significantly higher degree of compression than the other lossless coders, since their compression ratios CR are significantly higher than the other coders. Additionally, although the compression performance of JPEG 2000 and JPEG-LS are very similar (since their compression ratios CR only differ by 1.81%), JPEG 2000 is performing slightly better.

The high compression yielded by JPEG 2000 can be accounted to three factors: use of an MRA technique (DWT), bit-plane coding and the context-based adaptive binary arithmetic

coder. The DWT decomposed the mammogram into a multiresolutional representation which only uses a few non-zero coefficients to represent the entire mammogram. This significantly reduces the average dynamic range of the input into the bitplane coder. The bitplane coder utilizes the relationships that exists between the bits of wavelet coefficients, resulting in a efficiently represented embedded code and respective context models. The 18 contexts used by JPEG 2000 and the respective output binary symbols are passed to a context-based adaptive binary arithmetic coder for compression. Context-based adaptive binary arithmetic coding achieves good compression results by: (1) selecting probability models for each symbol according to the symbol's context, (2) adaptation of the probability estimates based on local statistics and (3) using arithmetic coding [122]. Rapid probability adaption is achieved as a result of the small number of contexts used by the JPEG 2000 coding engine. Although the compression achieved is significant, JPEG 2000's speed of compression is relatively slower. In fact, as shown in Table 4.4 and Figure 4.16, JPEG 2000 is the second slowest compression scheme, next to arithmetic codes with an order 1 model. By adding all the additional features to the JPEG 2000 codec, unfortunately there is also an associated cost in terms of computational efficiency.

The JPEG-LS algorithm achieved excellent results for both compression and speed. The compression performance of JPEG-LS is a result of the nonlinear prediction scheme used to generate the error distribution and decorrelate edge information, an accurate statistical descriptor (TSGD) used to model the prediction residuals and the Golomb-Rice coder, which is an entropy coder specifically tuned to the prediction error distribution. Also, all computational complexity considerations that went into creating JPEG-LS (i.e. efficient statistical model parameters, different coding modes, etc.) has managed to make this a very efficient compressor. As is shown in Table 4.4 and Figure 4.16, the compression speed is significantly lower than any of the other codecs.

As shown in Table 4.4 and Figure 4.15, LJPEG performed quite poorly. Although mammogram images consist of regions with slowly varying contrast, a large portion of the mammogram image still contains high frequency edge content, like the breast outline and the

inner breast tissue and blood vessels. As a result of this, the linear prediction techniques used by LJPEG can remove the redundancy of the “flat” regions, but from the compression results, it is obvious that the simple predictor function was not able to decorrelate the edge information. Unlike JPEG-LS, which uses a probability distribution and contexts to account for the redundancy that is not removed by the predictor, LJPEG directly codes the error image. Consequently, large prediction residuals are found at edges and the overall dynamic range wasn’t reduced enough to warrant good compression. Additionally, coding gains also suffered since Huffman coding is used. Arithmetic coding was avoided due to the drastic increase in computational complexity.

When examining the other compression results, it is obvious that adaptive Huffman, LZW with a 12 bit dictionary and arithmetic with a zero order probability performed very poorly. A downfall of any Huffman compressor and the reason for the adaptive Huffman compressor’s poor performance is due to the fact that maximum compression is achieved when the symbol probabilities are a power of $\frac{1}{2}$ [47]. Obviously, as seen by the compression results, the mammograms cannot be efficiently modeled using the assumptions of the Huffman-based coder.

Since codewords are appointed to a given symbol or pattern of symbols only if they are found in the dictionary, LZW codes heavily rely on repeating patterns. As long as the dictionary is long enough and the mammogram data is highly correlated, LZW compressors should find a large number of repeating patterns to produce high coding gains. Although the mammogram’s pixels are highly correlated and the file size is large enough to gather large string patterns, LZW with a 12 bit dictionary performed poorly. This is because the dictionary’s maximum buffer length was not long enough to capture the longer symbol patterns. For LZW with a 15 bit dictionary, the 15 bit buffer was long enough to recognize a large number of patterns in the input mammogram and this resulted in reasonably good compression results. Since the *exact* statistical behaviour of each mammogram is not known to the LZW compressor prior to compression, choosing a suitable dictionary buffer length can be difficult, thus ultimately affecting compression performance.

The differing compression performance of arithmetic with a zero order and arithmetic with a first order probability model is a result of the differing probability models. Arithmetic coding that employs a higher-order probability model exploits correlation between successive pixels in order to gain a more accurate description of the source. Since the arithmetic coder with a zero-order probability model assumes the mammogram's pixels to be stochastically independent (which is inaccurate since mammograms can be modeled by a Markov source) compression performance suffered. It is also important to mention how slow compression becomes by introducing a higher-order model for source description, as is noticed when comparing the compression speeds of arithmetic with a zero-order model, to that of an arithmetic coder with a first-order model. Although increased compression efficiency is achieved, one must decide if the increase in computational complexity warrants the savings in terms of bits.

JPEG 2000 and JPEG-LS both offer great lossless compression performance for mammograms compared to the other coders. Therefore, one of these two schemes is the best choice for a compression scheme for mammograms. Since the compression performances are quite similar, other variables must be considered, such as: the memory requirements needed to manage these schemes, their computational complexity, error resilience and the range of features that aid in a effective source encoding algorithm for the transmission and mass-storage of digital mammograms.

LOCO-I (LOW COMplexity LOSSless COMpression for Images), produced by Hewlett Packard, is the software-based implementation of JPEG-LS used for experimentation and was designed to have low complexity [110]. Upon comparison of the encoding and decoding delays of LOCO-I (JPEG-LS) and JasPer (JPEG 2000), the execution times of JPEG 2000 were longer than those experienced by JPEG-LS (as shown in Table 4.4 and Figure 4.16). The coding delays of JPEG 2000 are a result of the computationally complex computation of the DWT, bit-plane and entropy coding. Not only are these portions of the JPEG 2000 algorithm computationally complex, but they also require significant memory bandwidth. The JPEG 2000 standard tries to combat these downfalls by providing options such as tiling,

which segments the image into smaller portions and allows computationally and memory intensive portions of the code to be preformed on individual tiles. This significantly reduces the computational and memory requirements of the computing device and also makes way for implementation on devices with limited resources, such as a PDA. Furthermore, since these components of the algorithm (when implemented in software) are extremely demanding of the computing device, for a more efficient and speedy implementation, these sections could be implemented by a hardware accelerator.

Although JPEG-LS may perform at a faster speed than JPEG 2000 and requires little storage for its probability models, this algorithm is very susceptible to channel errors, making it poor choice for communication-based applications [123]. If a compression algorithm is to be utilized by the medical profession, it should not only provide good compression results, but it also should be robust enough to withstand transmission errors to a certain degree. This would allow maximum exchange of mammogram images to health care facilities in an effective, rapid manner, reducing wait times for diagnosis. JPEG 2000 presents an improvement in this area by adding error resiliency capabilities to the bitstream. By re-initializing and terminating the arithmetic codestring at the beginning and the end of each code block respectively, any error caused by channel effects will be solely contained within the code block and will not propagate throughout the image. Additionally, packet markers allow for the retransmission of lost packets.

JPEG-LS does not utilize the bit-plane coding algorithm proposed by Taubman, which adds different types of scalability to the bitstream. The embedded bitstream gives way to progressive transmission, which allows the user to view the mammogram as it is arriving and reject it (if desired) before it is completely downloaded making remote access to digital libraries a viable option. The bit-plane codes also enable the user to view all the mammograms as a thumbnail prior to selecting it, thus aiding in fast database browsing. Additionally, the DWT can be used for feature extraction and classification, offering the physician a second opinion about suspicious regions such as microcalcifications or benign and malignant masses. These features and the other remarkable features such as region of interest coding, tiling, the

inclusion of metadata, make JPEG 2000 the most applicable algorithm for the compression of medical images.

5.3 Summary of Results

This thesis contains the successful comparison and application of the DWT for two very different and important applications for biomedical images (retinal, small bowel and mammogram): classification and compression. In the proposed classification scheme, textural features were extracted from both the DWT and spatial domains, which describe the relative homogeneity of local and global areas, respectively. As described by in Section 3.4.1, perception of texture requires evaluation spatial graylevel variances as well as the scale and frequency of the texture elements.

Since the DWT is shift-variant, a shift-invariant DWT was explored to ensure that a robust, consistent feature set was extracted. To combat the small database sizes, LDA classification was used in conjunction with the leave one out method in order to try and gain the true approximation of the classifier's parameters. High classification rates were achieved for the small bowel and retinal images in the compressed domain due to the correct characterization of the abnormalities by the textural descriptors. However, classification rates of the mammogram regions suffered since the lesions were masked by normal breast parenchyma.

Compression in the wavelet domain was explored using JPEG 2000 for mammogram images. JPEG 2000 was found to achieve the best compression results at a cost of high computational complexity. However, since the features which arise from the DWT are so useful in a medical image compression scheme (region of interest coding, bitstream scalability, etc.), JPEG 2000 is the best choice for the compression of mammogram images.

5.4 Future Work

Some of the following points could be addressed in future research endeavors:

1. The exploration of feature extraction in the colour components for the retinal and small bowel images. Since these images were converted to grayscale for consistent comparison with the mammogram images, colour information was neglected from the analysis. These colour planes may contain discriminating information.
2. For the mammogram images, a segmentation algorithm may be applied prior to feature extraction. This would allow only the true characteristics of the lesion to be examined, without including the background tissue into analysis.
3. The effect of the number of features extracted as well as the database size on classification performance will also be examined.
4. The exploration of decomposition techniques which may simultaneously capture local and global feature events, such as Non-Negative Matrix Factorization techniques [124].
5. Involvement of radiologists to provide expert clinical guidance.

Appendix A

Filters for DWT Decomposition

The filter coefficients $h_o(n)$ and $h_1(n)$ are related to one another by Equation 2.21. By performing some manipulations and taking the Fourier transform of this equation, $h_o(n)$ and $h_1(n)$ have the following relationship in the frequency domain:

$$H_1(\omega) = H_o(\omega - \pi). \quad (\text{A.1})$$

For wavelet decomposition using filters, $h_o(n)$ is a lowpass filter which represents the scaling functions and $h_1(n)$ is a highpass filter which represents the wavelet functions [17] [19]. Furthermore, both lowpass $h_o(n)$ and highpass $h_1(n)$ filters are half-band filters [38].

If the DWT is computed using filterbanks, then the highpass and lowpass filters will split the original signal into half band low frequency and half band high frequency components. To achieve higher levels of decomposition, the lowest frequency subband is recursively filtered using $H_o(z)$ and $H_1(z)$, followed by subsampling. Therefore, the DWT may be considered as a filtering technique which isolates particular frequency bands. This is shown in Figure A.1, where DWT decomposition is demonstrated using the spectral content of an input signal $f(t)$, where $f(t) \in \mathcal{V}_2$. The first plot of Figure A.1 shows the spectrum of the original signal (space $\mathcal{V}_2$), the second plot shows how the spectrum is split into low and high frequency bands (spaces $\mathcal{V}_1$ and $\mathcal{V}_1$) and the third plot is the spectrum after two levels of decomposition.

For digital implementations, it is possible to use finite impulse response (FIR) or infinite impulse response (IIR) filters. Since FIR filters can realize exact linear phase response and are always stable, FIR filters have become the industry standard for wavelet decomposition

Figure A.1: The frequency domain representation of a two level decomposition using the DWT.

techniques [38]. For perfect reconstruction, the sum of the squares of the magnitude response of both filters must equal 1, as in:

$$\begin{aligned}
 |H_o(\omega)|^2 + |H_1(\omega)|^2 &= 1, \\
 |H_o(\omega)|^2 + |H_o(\omega - \pi)|^2 &= 1.
 \end{aligned}
 \tag{A.2}$$

Filters which meet the above requirements of perfect reconstruction are known as quadrature mirror filters (QMF's) [17].

Appendix B

Discrete Wavelet Transform: Lifting Based Scheme

Although discussion of wavelet transforms using decomposition methods and filterbank techniques provides an intuitive description of what the DWT is and does, it is necessary to investigate techniques which allow efficient implementation of the DWT. One such method is called the lifting based approach [125] [126] [127]. To increase computation speed, lifting based approaches make optimal use of similarities which exist between the lowpass ($H_1(z)$) and highpass ($H_o(z)$) filters. All 1-D implementations will be later extended to 2-D implementations by 'lifting' both the columns and the rows separately.

The lifting based DWT is an efficient scheme since it aims to implement complicated functions with simple and invertible stages [128]. Compared to the filterbank method, the lifting based DWT method offers a less computationally expensive solution to compute the DWT [128] [127].

B.1 Split, Predict and Update

The lifting based scheme relies on three operations to achieve the discrete wavelet transform: 1) split, 2) predict and 3) update. These three operations which comprise the 1-D lifting scheme, is shown in Figure B.1, where $\mathcal{S}$ is the splitting function, $\mathcal{P}$ is the predictor function and $\mathcal{U}$ is the update operation. As shown by Figure B.1, the scaling and wavelet coefficients ($c_j(n)$ and $d_j(n)$) are still from the previous level's coefficients, $c_{j+1}(n)$. Lifting may be also

applied separately to the rows and columns of an image to arrive at a 2-D DWT.

Figure B.1: Generalized 1-D lifting based implementation of the DWT.

B.1.1 Splitting

The splitting operation divides the 1-D input string into even and odd samples, as denoted by $c_{j+1}(2n)$ and $c_{j+1}(2n+1)$, respectively. Using digital signal processing, the even samples may be obtained by decimating the original signal by a factor of 2, and the odd samples may be obtained by subsampling a time shifted (single unit of time) version of the original signal by 2. This is often referred to as the Lazy Wavelet Transform [125].

B.1.2 Prediction

In order to compute the wavelet coefficients $d_j(n)$, a lifting scheme uses a predictor to interpolate the odd-indexed coefficients from the previous scale ($c_{j+1}(2n+1)$). The prediction is subtracted from the original odd-indexed signal to produce the wavelet coefficients $d_j(n)$. This may expressed as

$$d_j(n) = c_{j+1}(2n+1) - \mathcal{P}(c_{j+1}(2n+1)), \quad (\text{B.1})$$

where $\mathcal{P}(\cdot)$ is the predictor function. As stated earlier, the wavelet coefficients correspond to the high frequency components which makes this operation equivalent to highpass filtering.

A good predictor function would produce small valued wavelet coefficients (ideally zero), since the predicted version of the signal would be identical to the original. However, for

nonstationary signals (such as biomedical images) that have properties which change over time, it is not possible to exactly predict the signal [128] and non-zero wavelet coefficients can be expected. There are many different predictor functions which may be used [105] [104] [106], however, in order to implement the forward wavelet transform, the interpolation function is chosen such that it relates to the wavelet $\psi(t)$ [128].

B.1.3 Updating

In a lifting based DWT implementation, the scaling coefficients $c_j(n)$ are computed as the sum of the even-indexed samples ($c_{j+1}(2n)$) and an updated version of the wavelet coefficients $d_j(n)$ as shown below:

$$c_j(n) = c_{j+1}(2n) + \mathcal{U}(d_j(n)), \quad (\text{B.2})$$

where $\mathcal{U}(\cdot)$ is the update function. This operation isolates the low frequency components within the original signal.

B.1.4 2-D Implementation

For biomedical imagery, it is necessary to investigate the lifting based DWT for two dimensions. As stated earlier, 1-D wavelet transforms may be applied separably to the images in order to gain a 2-D DWT representation. This also applies to lifting based schemes as well. By sequentially applying Equations 2.32 and 2.33 first to the rows and then to the columns of an image, the forward transformation is achieved. To reverse this, Equations 2.34 must be applied first along columns and then along rows. The forward operation is depicted in Figure B.2.

Figure B.2: Lifting-based implementation of the DWT for two dimensional signals.

Appendix C

Shift-Variance of the DWT

Recall from Equations 2.25 and 2.26 that the scaling $c_j(k)$ and wavelet $d_j(k)$ coefficients may be computed by filtering $c_{j+1}(k)$ with time delayed versions of $h_o(n)$ and $h_1(n)$, respectively. These filtering equations will be used to examine the shift-variance of the DWT. It is sufficient to only analyze the properties and characteristics for the computation of $c_j(k)$, since all operations also apply to $d_j(k)$. Therefore, if $c_j(k)$ is shift-variant, than so is $d_j(k)$.

For $k \in \{0, 1, 2, \dots, \frac{N}{2} - 1\}$, where $N = 2^{j+1}$ is the length of the signal at scale $j + 1$, it is possible to compute $c_j(k)$ to get the following expressions

$$\begin{aligned} c_j(0) &= \sum_m h_o(m) \cdot c_{j+1}(m), \\ c_j(1) &= \sum_m h_o(m - 2) \cdot c_{j+1}(m), \\ c_j(2) &= \sum_m h_o(m - 4) \cdot c_{j+1}(m), \\ &\vdots \\ c_j\left(\frac{N}{2} - 1\right) &= \sum_m h_o\left(m - 2\left(\frac{N}{2} - 1\right)\right) \cdot c_{j+1}(m), \end{aligned}$$

which shows that $c_j(k)$ may be found by the multiplication and sum of time shifted versions of $h_o(n)$ with coefficients at higher levels. Therefore, define a filtering matrix $\mathbf{H}$ of size

$\left[\frac{N}{2} \times N\right]$, as shown below

$$\mathbf{H} = \begin{bmatrix} h_o & h_1 & h_2 & \cdots & h_L & 0 & 0 & \cdots \\ 0 & 0 & h_o & h_1 & h_2 & \cdots & h_L & 0 \\ 0 & 0 & 0 & 0 & h_o & h_1 & h_2 & \cdots \\ \vdots & & & & & & & \\ h_L & 0 & 0 & \cdots & 0 & h_o & h_1 & h_2 \\ \vdots & & & & & & & \vdots \\ h_2 & \cdots & h_L & 0 & 0 & \cdots & h_o & h_1 \end{bmatrix}, \quad (\text{C.1})$$

where each row is a time shifted (by two) and zero padded version of the lowpass filter $h_o(n) = \{h_o, h_1, h_2, \dots, h_L\}$ and L is the length of the filter. The filtering matrix $\mathbf{H}$ allows the DWT to be computed using matrices

$$\begin{bmatrix} \mathcal{C}^j(0) \\ \mathcal{C}^j(1) \\ \mathcal{C}^j(2) \\ \vdots \\ \mathcal{C}^j(\frac{N}{2} - 1) \end{bmatrix} = \mathbf{H} \cdot \begin{bmatrix} \mathcal{C}^{j+1}(0) \\ \mathcal{C}^{j+1}(1) \\ \mathcal{C}^{j+1}(2) \\ \vdots \\ \mathcal{C}^{j+1}(N - 1) \end{bmatrix},$$

which can be rewritten as

$$\mathbf{C}_{[\frac{N}{2} \times 1]}^j = \mathbf{H}_{[\frac{N}{2} \times N]} \cdot \mathbf{C}_{[N \times 1]}^{j+1}, \quad (\text{C.2})$$

where the subscript of a matrix $[\cdot]$ contains the dimensions, $\mathbf{C}^j$ is the matrix representation of the scaling coefficients at the current scale j and $\mathbf{C}^{j+1}$ contains the scaling coefficients at the higher scale $j + 1$.

To examine the shift-variant behaviour of the DWT, redefine the input $\mathbf{C}^{j+1}$ to be a 2-D array and allow each of the columns to contain a circularly shifted version of the coefficients $c_{j+1}(k)$. Therefore, $\mathbf{C}^{j+1}$ may be rewritten as

$$\mathbf{C}_{[N \times N]}^{j+1} = \begin{bmatrix} \mathcal{C}^{j+1}(0) & \mathcal{C}^{j+1}(N - 1) & \mathcal{C}^{j+1}(N - 2) & \cdots & \mathcal{C}^{j+1}(1) \\ \mathcal{C}^{j+1}(1) & \mathcal{C}^{j+1}(0) & \mathcal{C}^{j+1}(N - 1) & \cdots & \mathcal{C}^{j+1}(2) \\ \mathcal{C}^{j+1}(2) & \mathcal{C}^{j+1}(1) & \mathcal{C}^{j+1}(0) & \cdots & \mathcal{C}^{j+1}(3) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathcal{C}^{j+1}(N - 1) & \mathcal{C}^{j+1}(N - 2) & \mathcal{C}^{j+1}(N - 3) & \cdots & \mathcal{C}^{j+1}(0) \end{bmatrix}; \quad (\text{C.3})$$

Consequently, the DWT would be computed by

$$\mathbf{C}_{[\frac{N}{2} \times N]}^j = \mathbf{H}_{[\frac{N}{2} \times N]} \cdot \mathbf{C}_{[N \times N]}^{j+1}, \quad (\text{C.4})$$

demonstrating that the first column of $\mathbf{C}_{[\frac{N}{2} \times N]}^j$ is the response to no shift of the input signal ($c_{j+1}(k)$), the second column is the response to a single shift in the input signal ($c_{j+1}(k-1)$), the third column is the response to a double shift of the input signal ($c_{j+1}(k-2)$), and so on. To examine the response to each shift m in $c_{j+1}(k-m)$, define the output coefficients to be $c_o^j(k)$. Therefore, for no shift in the input coefficients ($m=0$), the value of the first output coefficient ($k=0$) may be computed by

$$c_o^j(0) = c^{j+1}(0) \cdot h(0) + c^{j+1}(1) \cdot h(1) + \dots + c^{j+1}(N-1) \cdot h(N-1), \quad (\text{C.5})$$

which will be set to an arbitrary value A_o . Since the filter coefficients $h_o(n) = \{h_o, h_1, h_2, \dots, h_L\}$ will be present for the computation of every coefficient, Equation C.5 may be rewritten as a set of vector indices, where each index t contains the coefficient which is multiplied by the filter coefficient h_t

$$c_o^j(0) = (c^{j+1}(0), c^{j+1}(1), \dots, c^{j+1}(N-1)) = A_o. \quad (\text{C.6})$$

Similarly, the rest of coefficients for the shift of $m=0$ may be found by

$$\begin{aligned} c_o^j(1) &= (c^{j+1}(2), c^{j+1}(3), \dots, c^{j+1}(1)) = A_1, \\ c_o^j(2) &= (c^{j+1}(4), c^{j+1}(5), \dots, c^{j+1}(3)) = A_2, \\ c_o^j(3) &= (c^{j+1}(6), c^{j+1}(7), \dots, c^{j+1}(5)) = A_3, \\ &\vdots, \\ c_o^j\left(\frac{N}{2}-1\right) &= (c^{j+1}(N-2), c^{j+1}(N-1), \dots, c^{j+1}(N-3)) = A_{\frac{N}{2}-1}. \end{aligned} \quad (\text{C.7})$$

Therefore, $c_o^j(k)$ is equal to

$$c_o^j(k) = [A_o, A_1, A_2, A_3, \dots, A_{\frac{N}{2}-1}]. \quad (\text{C.8})$$

Similarly, for a shift of $m=1$, the response to the input signal $c_{j+1}(k-1)$ may be found by

$$\begin{aligned} c_1^j(0) &= (c^{j+1}(N-1), c^{j+1}(0), \dots, c^{j+1}(N-2)) = B_o, \\ c_1^j(1) &= (c^{j+1}(1), c^{j+1}(2), \dots, c^{j+1}(0)) = B_1, \end{aligned} \quad (\text{C.9})$$

$$\begin{aligned}
c_1^j(2) &= (c^{j+1}(3), c^{j+1}(4), \dots, c^{j+1}(2)) = B_2, \\
&\vdots \\
c_1^j\left(\frac{N}{2} - 1\right) &= (c^{j+1}(N-3), c^{j+1}(N-2), \dots, c^{j+1}(N-4)) = B_{\frac{N}{2}-1}.
\end{aligned}$$

For simplicity, $c_1^j(k)$ may be rewritten as

$$c_1^j(k) = [B_o, B_1, B_2, B_3, \dots, B_{\frac{N}{2}-1}]. \quad (\text{C.10})$$

Using the principals of Equation C.4 through to Equation C.10, the expressions for the output coefficients $c_m^j(k)$ for all input shifts m may be expressed. For instance, for even shifts of m ($m = \{2, 4, 6, \dots, N-2\}$), the coefficients $c_m^j(k)$ are shifted versions of $c_o^j(k)$ as shown below:

$$\begin{aligned}
c_2^j &= [A_{N-1}, A_o, A_1, \dots, A_{N-2}] = c_o(k-1), \\
c_4^j &= [A_{N-2}, A_{N-1}, A_o, \dots, A_{N-3}] = c_o(k-2), \\
&\vdots \\
c_{N-2}^j &= [A_2, A_3, A_4, \dots, A_o] = c_o\left(k - \left(\frac{N-2}{2}\right)\right).
\end{aligned} \quad (\text{C.11})$$

For the odd shifts of m ($m = \{1, 3, 5, \dots, N-1\}$), the coefficients $c_m^j(k)$ are shifts of $c_1^j(k)$ as shown below:

$$\begin{aligned}
c_3^j &= [B_{N-1}, B_o, B_1, \dots, B_{N-2}] = c_1(k-1), \\
c_5^j &= [B_{N-2}, B_{N-1}, B_o, \dots, B_{N-3}] = c_1(k-2), \\
&\vdots \\
c_{N-1}^j &= [B_2, B_3, B_4, \dots, B_o] = c_1\left(k - \left(\frac{N-2}{2}\right)\right).
\end{aligned} \quad (\text{C.12})$$

Therefore, as a consequence of Equations C.11 and C.12, it is shown that the DWT produces two representations depending on the shift of the input signal. One distribution of amplitude coefficients corresponds to the even shifts of the input and another distribution of amplitude coefficients corresponds to the odd shifts. For even input shifts $M \in \{2, 4, \dots, N-2\}$, it is possible to arrive at the general formula

$$c_M^j(k) = c_o^j\left(k - \frac{M}{2}\right), \quad (\text{C.13})$$

and similarly for odd input shifts $M \in \{3, 5, \dots, N - 1\}$

$$c_M^j(k) = c_1^j \left(k - \frac{M - 1}{2} \right). \quad (\text{C.14})$$

Since it is possible to arrive at two different representations, it is shown that the DWT is *shift-variant*. For two different shifts of the original signal (or previous level coefficients), we arrive at a completely different set of output coefficients.

Although the above derivations and discussions were focused on 1-D signals and implementation, the results can be easily extended to 2-D, for images, since the 1-D filtering operations are applied separably to rows and columns. The coefficients from each of the four subbands at level j may be computed by filtering the LL band from the previous level $j + 1$

$$HH^j = LL^{j+1} \cdot H_1^T \cdot H_1, \quad (\text{C.15})$$

$$HL^j = LL^{j+1} \cdot H_o^T \cdot H_1, \quad (\text{C.16})$$

$$LH^j = LL^{j+1} \cdot H_1^T \cdot H_o, \quad (\text{C.17})$$

$$LL^j = LL^{j+1} \cdot H_o^T \cdot H_o. \quad (\text{C.18})$$

where T is the transpose, H_1 is the highpass filtering matrix, H_o is the lowpass filtering matrix and LL_{j+1} is the lowpass coefficients from the previous level. Note that LL_{j+1} would be the input image for the first level of decomposition.

By filtering both the rows and columns separately, the 2-D DWT can be expected to be shift-variant in both directions. As shown earlier for 1-D signals, the two different distributions of coefficients may be computed from shifting the input by $m = 0$ and $m = 1$. All other shifts of the input results in one of these two representations (for $m = 0$ and $m = 1$) (i.e. the amplitude distribution is the same). Therefore, in 2-D, it is necessary to consider the two possible output representations (even/odd shifts) for both the rows and columns. Consequently, a total of four different representations would be achieved for images, corresponding to the following shifts: (0,0), (0,1), (1,0), (1,1), where the first index corresponds to the row shift and the second index corresponds to the column shift. For any other input shift, the DWT will produce an amplitude distribution which is the same as one of the four, fundamental shifts.

Appendix D

Graphical User Interface

Figure D.1: Graphical user interface created for the classification of biomedical imagery.

For classification of the biomedical images, a graphical user interface (GUI) was created in Matlab to ease image loading, conversion, translation as well as the computation of the shift-invariant discrete wavelet transform (SIDWT) and the discrete wavelet transform (DWT). Other options included: feature extraction and best basis calculation. A sample view of the GUI is shown in Figure D.1, where the image on the left is a contrast enhanced mammogram region containing a benign mass and the image on the right is of one level of DWT decomposition of the image on the right. Code for this GUI can be obtained from the Signal Analysis Research (SAR) group, Ryerson University.

Publications

Journals

- A. Khademi, S. Krishnan. A Comprehensive Overview of Lossless Compression Schemes for Digital Mammograms, under review, *Computer Methods and Programs in Biomedicine Journal*, June 2006.
- A. Khademi, S. Krishnan. Shift-Invariant Discrete Wavelet Transform Analysis for Retinal Image Classification, under review, *Medical & Biological Engineering & Computing (MBEC) Journal*, Sept. 2006.
- A. Khademi, S. Krishnan. Small Bowel Medical Image Classification using Wavelet Analysis, under review, *IEEE Transactions on Medical Imaging*, Sept. 2006.

Conferences

- A. Khademi, S. Krishnan. Multiscale Texture Analysis and Classification of Retinal Images, under review, *IEEE's International Symposium of Biomedical Imaging (ISBI)*, Sept. 2006.
- A. Khademi, S. Krishnan. Multiresolution Analysis and Classification of Small Bowel Medical Images, under review, *IEEE's International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Sept. 2006.
- A. Khademi, S. Krishnan. Comparison of JPEG 2000 to Other Lossless Compression Schemes for Digital Mammograms. *In the proceedings of IEEE's Engineering in Medicine and Biology Conference (EMBC)*, Sept. 2005, pp. 3771 - 3774.
- A. Khademi, D. Hosseinzadeh, S. Krishnan. Source and Channel Coding Techniques for the Faithful Transmission of Digital Mammograms. *In the proceedings of IEEE's International Conference of Upcoming Engineers (ICUE)*, 2004. Won "Best Undergraduate Technical Paper".

Other Scholarly Articles

- A. Khademi. Stochastic Image Processing - Book Review. *IEEE Canadian Review*, Edition 49, pp. 21, Winter 2005.
- A. Khademi, D. Hosseinzadeh, S. Krishnan. Source and Channel Coding Techniques for the Faithful Transmission of Digital Mammograms. *IEEE Canadian Review*, Edition 48, pp. 8-12.

Bibliography

- [1] Yale University. *Section of Cardio Vascular Medicine*. World Wide Web. <http://info.med.yale.edu/intmed/cardio/>.
- [2] Given Imaging Ltd. *PillCamTM Image Atlas*. World Wide Web, 2006. <http://www.givenimaging.com/>.
- [3] J. Suckling et al. *The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica*. World Wide Web. <http://peipa.essex.ac.uk/info/mias.html>.
- [4] M. Goldbaum. *STARE - STructured Analysis of the Retina*. World Wide Web, 2002. <http://www.parl.clemson.edu/stare/>.
- [5] J.P. Hornak. *The Basics of MRI*. World Wide Web, 2006. <http://www.cis.rit.edu/htbooks/mri>.
- [6] S. Mallat. *Wavelet Tour of Signal Processing*. Academic Press, USA, 1998.
- [7] J. Liang and T.W. Parks. A translation-invariant wavelet representation algorithm with applications. *IEEE Transactions on Signal Processing*, 44(2):225 – 232, Feb. 1996.
- [8] American College of Radiology. ACR technical standard for digital image data management. ACR technical standard, 2001.
- [9] American College of Radiology. Practice Guideline for the Performance and Interpretation of Magnetic Resonance Imaging of Bone and Soft Tissue Tumors. ACR technical standard, 2005.

- [10] Health Canada. It's your health: Breast cancer. *World Wide Web*, 2004. <http://www.hc-sc.gc.ca>.
- [11] Canadian Interventional Radiology Association. *Innovative Care Through Minimally Invasive Image-Guided Therapy*. World Wide Web, 2005. <http://www.car.ca/cira/>.
- [12] E. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. Baum, S. Acharyya, E. Conant, L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, and M. Rebner. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New England Journal of Medicine*, 353(17):1773–1783, Oct. 2005.
- [13] S. Feig, L. Bassett, M. Yaffe, C. Vyborny, and P. Butler. ACR practice guideline for the performance of whole breast digital mammography. Technical report, American College of Radiology, 2001.
- [14] L. Kennedy. *Digital Mammography Does not Improve Breast cancer rates, study finds*. Health Behaviour News Service, <http://www.hbns.org/news/mammography06-10-05.cfm>, June 2005.
- [15] S. Ross. *Introduction to Probability Models*. Academic Press, USA, California, 2003.
- [16] C. Burrus, R. Gopinath, and H. Guo. *Introduction to Wavelets and Wavelet Transforms - A Primer*. Prentice Hall International, Inc., Houston, Texas, 1998.
- [17] T.C. Wang and N.B Karayiannis. Detection of microcalcifications in digital mammograms using wavelets. *IEEE Transactions on Medical Imaging*, 17:498 – 509, Aug. 1998.
- [18] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.
- [19] M. Vetterli and C. Herley. Wavelets and filter banks: theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207 – 2232, Sept. 1992.

- [20] C.B.R. Ferreira and D.L. Borges. Automated mammogram classification using a multiresolution pattern recognition approach. pages 76 – 83. Proceedings of XIV Brazilian Symposium on Computer Graphics and Image Processing, Oct. 2001.
- [21] H.D. Cheng, X.J. Shi, R. Min, L.M Hu, X.P. Cai, and H.N. Du. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognition*, 39(4):646–668, April 2006.
- [22] D. Wei, H.P. Chan, M.A. Helvie, B. Sahiner, N. Petrick, D.D. Adler, and M.M. Goodsitt. Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis. *Medical Physics*, 22(9):1501–1513, Sept. 1995.
- [23] N. Day and R. Warren. Mammographic screening and mammographic patterns. *Breast Cancer Research*, 2(4):247–251, June 2000. University of Cambridge, Rosie Maternity Hospital, Cambridge, UK.
- [24] L. Shen, R.M. Rangayyan, and J.E.L. Desautels. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*, 13(2):263 – 274, June. 1994.
- [25] N.M El-Faramawy, R.M. Rangayyan, J.E.L Desautels, and O.A. Alim. Shape factors for analysis of breast tumors in mammograms. volume 1, pages 355 – 358. Canadian Conference on Electrical and Computer Engineering, May 1996.
- [26] 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*, 16(6):799 – 810, Dec. 1997.
- [27] B. Kim, S. Park, C. Jee, and S. Yoon. An earthworm-like locomotive mechanism for capsule endoscopes. pages 2997 – 3002. IEEE/RSJ International Conference on Intelligent Robots and Systems, Aug. 2005.

- [28] Given Imaging Ltd. *PillCam™ SB Capsule Endoscopy - Product Information Guide*. World Wide Web, 2006. <http://www.givenimaging.com/Cultures/en-US/Given/English/Products/CapsuleEndoscopy/>.
- [29] Given Imaging Ltd. *Patient Brochure*. World Wide Web, 2006. <http://www.givenimaging.com/Cultures/en-US/Given/English/Professionals/>.
- [30] Given Imaging Ltd. *Patient Information Guide*. World Wide Web, 2006. <http://www.givenimaging.com/Cultures/en-US/Given/English/Patients/>.
- [31] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija. Automated screening system for diabetic retinopathy. volume 2, pages 915 – 920. Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis, Sept. 2003.
- [32] H. Wang, W. Hsu, K. Goh, and M. Lee. An effective approach to detect lesions in color retinal images. volume 2, pages 181–186. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 2000.
- [33] L. Brandon and A. Hoover. Drusen detection in a retinal image using multi-level analysis. volume 1, pages 618–625. MICCAI, 2003.
- [34] S.F. Barrett, H. Grady Rylander, and A.J. Welch III. Automated lesion data base building for the treatment of retinal disorders. volume 1, pages 426 – 430. Proceedings IEEE International Conference Image Processing, Nov. 1994.
- [35] G.Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM Journal of Numerical Analysis*, 29:1716–1740, 1992.
- [36] A. Bradley. Shift-invariance in the discrete wavelet transform. pages 29 – 38. Digital Image Computing: Techniques and Applications, Dec. 2003.
- [37] A.L. Graps. An introduction to wavelets. *IEEE Computational Sciences and Engineering*, 2(2):50–61, 1995.

- [38] S. Lawson and J. Zhu. Image compression using wavelets and JPEG2000: a tutorial. *Electronics & Communication Engineering Journal*, 14(3):112–121, 2004.
- [39] M. D. Adams and R. K. Ward. Symmetric-extension-compatible reversible integer-to-integer wavelet transforms. *IEEE Transactions on Signal Processing*, 51(10):2624–2636, Oct. 2003.
- [40] M.W. Marcellin, A. Bilgin, M.J. Gormish, and M.P. Boliek. An overview of JPEG-2000. In *Proceedings of the IEEE Data Compression Conference*, page 523. IEEE Computer Society, 2000.
- [41] H. Zhang and J. Fritts. An overview of JPEG-2000. pages 1333–1340. *Proceedings of SPIE - Visual Communications and Image Processing*, Jan. 2004.
- [42] S. Liu, C.F. Babbs, and E.J. Delp. Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Transactions on Image Processing*, 10(6):874–884, June 2001.
- [43] C. Pun and M. Lee. Extraction of shift invariant wavelet features for classification of images with different sizes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1228 – 1233, Sept. 2004.
- [44] R.M. Rangayyan. *Biomedical Image Analysis*. CRC Press LLC, United States of America, 2005.
- [45] W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In A. F. Laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing III*, pages 68–79. *Proceeding of SPIE*~2569, 1995.
- [46] R.G. Gallager. Variations on a theme by Huffman. *IEEE Transactions in Information Theory*, 24:668–674, 1978.
- [47] A. Moffat, R.M. Neal, and I.H. Witten. Arithmetic coding revisited. *ACM Transactions on Information Systems*, 16(3):256–294, 1998.

- [48] P.G. Howard and J.S Vitter. Arithmetic coding for data compression. *Proceedings of IEEE*, 82:857–865, 1994.
- [49] I.H. Witten, R.M. Neal, and J.G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30:520–540, 1987.
- [50] C. Christopoulos, T. Ebrahimi, and A. Skodras. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, 18(5):36–58, 2001.
- [51] C. Christopoulos, T. Ebrahimi, and A. Skodras. The JPEG 2000 still image coding system: An overview. *IEEE Transactions on Consumer Electronics*, 46(4):1103–1127, 2000.
- [52] M.D. Adams. The JPEG-2000 still image compression standard. Technical report, University of Victoria, Dept. of Electrical and Computer Engineering, 2002.
- [53] M.N. Mughal and W. Ikram. Early lung cancer detection by classifying chest CT images: a survey. pages 67 – 72. Proceedings of INMIC - Multitopic Conference, Dec. 2004.
- [54] C. Chiu, H. Wong, and H.H.-S. Ip. Compressed domain feature transformation using evolutionary strategies for image classification. volume 1, pages 429 – 432. International Conference on Image Processing, Oct. 2004.
- [55] Z. Xiong and T.S. Huang. Wavelet-based texture features can be extracted efficiently from compressed-domain for JPEG2000 coded images. volume 1, pages 481–484. International Conference on Image Processing, Sept. 2002.
- [56] S. Chang. Compressed-domain techniques for image/video indexing and manipulation. volume 1, pages 314 – 317. International Conference on Image Processing, Oct. 1995.
- [57] A. Armstrong and J. Jiang. An efficient image indexing algorithm in JPEG compressed domain. pages 350–351. International Conference on Image Processing, June 2001.

- [58] G. Voulgaris and J. Jiang. Texture-based image retrieval in wavelets compressed domain. volume 2, pages 125 – 128. International Conference on Image Processing, Oct. 2001.
- [59] E. Zhang, Y. Zhang, and T. Zhang. Automatic retinal image registration based on blood vessels feature point. volume 4, pages 2010 – 2015. Proceedings of the International Conference on Machine Learning and Cybernetics, Nov. 2002.
- [60] G.K. Matsopoulos, N.A. Mouravliansky, K.K. Delibasis, and K.S. Nikita. Automatic retinal image registration scheme using global optimization techniques. *IEEE Transactions on Information Technology in Biomedicine*, 3(1):47 – 60, Mar. 1999.
- [61] B. Raman, M. Wilson, I. Benche, and P. Soliz. A JAVA-based system for segmentation and analysis of retinal images. pages 336 – 339. Proceedings of the IEEE Symposium Computer-Based Medical Systems, June 2003.
- [62] M. Foracchia, E. Grisan, and A. Ruggeri. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Transactions on Medical Imaging*, 23(10):1189 – 1195, Oct. 2004.
- [63] R. Pfisterer and F. Aghdasi. Detection of masses in digitised mammograms. pages 115 – 120. Proceedings of the 1998 South African Symposium on Communications and Signal Processing, Sept. 1998.
- [64] T. Lambrou, A.D. Linney, R.D Speller, and A. Todd-Pokropek. Statistical classification of digital mammograms using features from the spatial and wavelet domains. pages 1–4. Medical Image Understanding and Analysis Conference, 2002.
- [65] N. Mudigonda, R.M. Rangayyan, and J.E.L Desautels. Gradient and texture analysis for the classification of mammographic masses. *IEEE Transactions on Medical Imaging*, 19(10):1032–1043, Oct. 2000.

- [66] R.R. Coifman and N. Saito. Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5:337–358, 1995.
- [67] S.E. Umbaugh, Y.S. Wei, and M. Zuke. Feature extraction in image analysis. a program for facilitating data reduction in medical image classification. *IEEE Engineering in Medicine and Biology Magazine*, 16:62–73, Jul. 1997.
- [68] M.M Leung and A.M. Peterson. Scale and rotation invariant texture classification. volume 1, pages 461 – 465. Conference Record of The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers, Oct. 1992.
- [69] T.N. Tan. Scale and rotation invariant texture classification. pages 3/1 – 3/3. IEE Colloquium on Texture Classification: Theory and Applications, 1994.
- [70] M. Unser and A. Aldroubi. A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 84(4):626 – 638, Apr. 1996.
- [71] J.Liang and T.W. Parks. A two-dimensional translation invariant wavelet representation and its applications. volume 1, pages 66 – 70. IEEE International Conference on Image Processing, Nov. 1994.
- [72] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, and D.J. Heeger. Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, 38:587 – 607, Mar. 1992.
- [73] J.Liang and T.W. Parks. Image coding using translation invariant wavelet transforms with symmetric extensions. *IEEE Transactions on Image Processing*, 7:762 – 769, May 1998.
- [74] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397 – 3415, Dec. 1993.
- [75] I.Cohen, S.Raz, and D.Malah. Orthonormal shift invariant wavelet packet decomposition and representation. *Elsevier: Signal Processing*, 57:251–270, 1997.

- [76] J.Liang and T.W. Parks. Translation invariant wavelet transforms with symmetric extensions. *IEEE Digital Signal Processing Workshop*, pages 69–72, 1996.
- [77] R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38:713 – 718, Mar. 1992.
- [78] H. Sari-Sarraf and D. Brzakovic. A shift-invariant discrete wavelet transform. *IEEE Transactions on Signal Processing*, 45(10):2621 – 2630, 1997.
- [79] S. Aksoy and R.M. Haralick. Textural features for image database retrieval. pages 45 – 49. *Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998.
- [80] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on System, Man, Cybernetics*, 3(6):610–621, Nov. 1973.
- [81] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, Mar. 1981.
- [82] F. Falzon and S. Mallat. Analysis of low bit rate image transform coding. *IEEE Transactions on Signal Processing*, 46(4):1027–1042, 1998.
- [83] G.M te Brake and N. Karssemeijer. Single and multiscale detection of masses in digital mammograms. *IEEE Transactions on Medical Imaging*, 18(7):628 – 639, July 1999.
- [84] S.J. Raudys and A.K. Jain. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):252 – 264, Mar. 1991.
- [85] S.J. Raudys and A.K. Jain. Small sample size effects in statistical pattern recognition: recommendations for practitioners and open problems. volume 1, pages 417 – 423. *Proceedings of the 10th International Conference on Pattern Recognition*, June 1990.

- [86] K. Fukunaga and R.R. Hayes. Effects of sample size in classifier design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):873 – 885, Aug. 1989.
- [87] S. Balakrishnama, A. Ganapathiraju, and J. Picone. Linear discriminant analysis for signal processing problems. pages 78 – 81. Proceedings of IEEE Southeastcon, 1999.
- [88] U.M. Braga-Neto and E.R. Dougherty. Is cross-validation valid for small-sample microarray classification? *Bioinformatics*, 20(3):374380, 2004.
- [89] L.S. Da Silva and J. Scharcanski. A lossless compression approach for mammographic digital images based on the delaunay triangulation. volume 2, pages 758 – 761. IEEE International Conference on Image Processing, 2005.
- [90] M. Penedo, W.A. Pearlman, P.G. Tahoces, M. Souto, and J.J. Vidal. Region-based wavelet coding methods for digital mammography. *IEEE Transactions on Medical Imaging*, 22:1288 – 1296, October 2003.
- [91] A. Said and W.A. Pearlman. A new, fast, and efficient image codec based on set partitioning in heirarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, 6(3):243–250, 1996.
- [92] A. Islam and W.A. Pearlman. An embedded and efficient low-complexity hierarchical image coder. *Proceedings of SPIE: Visual Communications and Image Processing*, 3653:294–305, 1999.
- [93] N.R. Wanigasekara, S. Ding, Z. Yan, and Y. Zeng. Quality evaluation for JPEG 2000 based medical image compression. volume 2, pages 1019 – 1020. Proceedings of the Second Joint EMBS/BMES Conference, Oct. 2002.
- [94] C.Won and R.M. Gray. *Stochastic Image Processing*. Kluwer Academic/Plenum Publishers, New York, 2004.

- [95] M. Kozuch and A. Wolfe. Compression of embedded system programs. pages 270 – 277. IEEE International Conference on Computer Design: VLSI in Computers and Processors, Oct. 1994.
- [96] R.L. Milidiu, E.S. Laber, and A. Pessoa. *Improved Analysis of FGK Algorithm*. MCC39/97, Departamento de Informatica, 1997.
- [97] J.S Vitter. Design and analysis of dynamic Huffman codes. *Journal of the Association for Computing Machinery*, 34:825–845, 1987.
- [98] L. Lee, L. Liu, J. Wang, and J. Lee. Dynamic mapping technique for adaptive Huffman code. volume 3, pages 653–656. IEEE Conference on Computer, Communication, Control and Power Engineering, 1993.
- [99] M.Y. Javed and A. Nadeem. Data compression through adaptive Huffman coding scheme. volume 2, pages 187–190. Proceedings of TENCON 2000, 2000.
- [100] P.G. Howard and J.S Vitter. Practical implications of arithmetic coding. Technical Report 92-18, Brown University - Department of Computer Science, 1992.
- [101] R.H. Greenfield and W. Kinser. The lempel-ziv-welch (LZW) data compression algorithm for packet radio. pages 225–229. WESCANEX '91 'IEEE Western Canada Conference on Computer, Power and Communications Systems in a Rural Environment, 1991.
- [102] M. Abdat, K. Ouaisa, and P. Plume. Adaptive limitation of the dictionary size in LZW data compression. page 18. Proceedings IEEE International Symposium on Information Theory, 1995.
- [103] International Telegraph and Telephone Consultative Committee (CCITT). Digital compression and coding of continuous-tone still images - Requirements and Guidelines. ISO/IEC 10918-1, ITU T.81, 1993.

- [104] T. Haijiang, K. Sei-Ichiro, T. Kazuyuki, and K. Masa-Aki. Lossless image compression via multi-scanning and adaptive linear prediction. volume 1, pages 81–84. The 2004 IEEE Asia-Pacific Conference on Circuits and Systems, 2004.
- [105] P. Maragos, R. Mersereau, and R. Shafer. Two-dimensional linear prediction and its application to adaptive predictive coding of images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 32:1213–1229, 1984.
- [106] K. Denecker, S. Van Assche, W. Phillips, and I. Lemahieu. State of the art concerning lossless medical image coding. pages 129–136. PRORISC IEEE Benelux Workshop on Circuits, Systems and Signal Processing, 1997.
- [107] JPEG Committee. *JPEG-LS Standard*. World Wide Web, 2004. <http://www.jpeg.org/jpeg/jpegl.html>.
- [108] S. Van Assche, D. De Rycke, W. Philips, and I. Lemahieu. Exploiting interframe redundancies in the lossless compression of 3D medical images. pages 521–527. Data Compression Conference, 2000.
- [109] J. Jiang, B. Guo, and S.Y Yang. Revisiting the JPEG-LS prediction scheme. *IEE Proceedings; Vision, Image and Signal Processing*, 147:575 – 580, 2000.
- [110] G. Sapiro, G. Seroussi, and M. Weinberger. LOCO-I: A low complexity, context-based, lossless image compression algorithm. pages 1–8. IEEE Data Compression Conference, 1996.
- [111] G. Sapiro, G. Seroussi, and M. Weinberger. The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Transactions in Image Processing*, 9:1309–1324, 2000.
- [112] O. Shalev, L. Schveiger, I. Sutskov, and R. Bar-Sella. A modified JPEG-LS image coder with increased channel error robustness. pages 245–248. The 21st IEEE Convention of the Electrical and ELectronic Engineers in Israel, 2000.

- [113] S. Lawson and J. Zhu. Image compression using wavelets and JPEG2000: a tutorial. *Electronics & Communications Engineering Journal*, 14(3):112–121, 2004.
- [114] P. Schumacher and J. Zhang. Texture classification using neural networks and discrete wavelet tranform. volume 3, pages 903–904. IEEE International Conference on Image Processing, 1994.
- [115] D. Taubman, E. Ordentlich, M. Weinberger, and G. Seroussi. Embedded block coding in JPEG2000. *Signal Processing: Image Communication*, 17(1):49–72, 2002. Preprint obtained from David Taubman’s home page: <http://maestro.ee.unsw.edu.au/~taubman/>.
- [116] D. Taubman. High performance scalable image compression with EBCOT. *IEEE Transactions on Image Processing*, 9:1158–1170, 2000.
- [117] J.M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41:3445–3462, 1993.
- [118] JBIG Committee. 14492 final committee draft - coded representation of picture and audio information - lossy/lossless coding of bi-level images. ISO/IEC JTC1/SC 29/WG1 N1359, 1999.
- [119] JBIG and JPEG Committee. JPEG 2000 Part I final committee draft version 1.0 - JPEG 2000 image coding system. ISO/IEC JTC1/SC29 WG1, 2000.
- [120] M.D. Adams. *The JasPer Project Home Page*. World Wide Web, <http://www.ece.uvic.ca/~mdadams/jasper/>, 2005.
- [121] Hewlett Packard. *HP Labs LOCO-I/JPEG-LS Home Page*. World Wide Web, <http://www.hpl.hp.com/loco/>, 2005.
- [122] I.E.G. Richardson. H.264/MPEG-4 Part 10 White Paper. World Wide Web, 2002. www.vocodex.com.

- [123] D. Santa-Cruz, R. Grosbois, and T. Ebrahimi. JPEG 2000 performance evaluation and assessment. *Signal Processing: Image Communication*, 17(1):113–130, 2002.
- [124] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788791, 1999.
- [125] G. Fernández, S. Periaswamy, and W. Sweldens. LIFTPACK: A software package for wavelet transforms using lifting. pages 396–408. Proceedings of the SPIE - Wavelet Applications in Signal and Image Processing IV, 1996.
- [126] W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In A. F. Laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing III*, pages 68–79. Proc. SPIE~2569, 1995.
- [127] W. Sweldens. Wavelets and the lifting scheme: A 5 minute tour. *Z. Angew. Math. Mech.*, 76 (Suppl. 2):41–44, 1996.
- [128] Y. Zhang and M. Zeytinoglu. Improved lifting scheme for block subband coding. pages 487 – 490. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Aug. 1999.