NOTE TO USERS

This reproduction is the best copy available.

Wavelet-Based Noise Reduction of cDNA Microarray Images

TAMANNA HOWLADER

A THESIS

IN

THE DEPARTMENT

OF

MATHEMATICS AND STATISTICS

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

CONCORDIA UNIVERSITY

MONTRÉAL, QUÉBEC, CANADA

JUNE 2009

Your file Votre référence
ISBN: 978-0-494-63377-9
Our file Notre référence
ISBN: 978-0-494-63377-9

ABSTRACT

Wavelet-Based Noise Reduction of cDNA Microarray Images

Tamanna Howlader, Ph.D.

Concordia University, 2009

The advent of microarray imaging technology has lead to enormous progress in the life sciences by allowing scientists to analyze the expression of thousands of genes at a time. For complementary DNA (cDNA) microarray experiments, the raw data are a pair of red and green channel images corresponding to the treatment and control samples. These images are contaminated by a high level of noise due to the numerous noise sources affecting the image formation. A major challenge of microarray image analysis is the extraction of accurate gene expression measurements from the noisy microarray images. A crucial step in this process is denoising, which consists of reducing the noise in the observed microarray images while preserving the signal information as much as possible. This thesis deals with the problem of developing novel methods for reducing noise in cDNA microarray images for accurate estimation of the gene expression levels.

Denoising methods based on the wavelet transform have shown significant success when applied to natural images. However, these methods are not very efficient for reducing noise in cDNA microarray images. An important reason for this is that existing methods are only capable of processing the red and green channel images separately. In doing so, they ignore the signal correlation as well as the noise correlation that exists between the wavelet coefficients of the two channels. The primary objective of this research is to design efficient wavelet-based noise reduction algorithms for cDNA microarray images that take into account these inter-channel dependencies by 'jointly' estimating the noise-free coefficients in both the channels. Denoising algorithms are

developed using two types of wavelet transforms, namely, the frequently-used discrete wavelet transform (DWT) and the complex wavelet transform (CWT). The main advantage of using the DWT for denoising is that this transform is computationally very efficient. In order to obtain a better denoising performance for microarray images, however, the CWT is preferred to DWT because the former has good directional selectivity properties that are necessary for better representation of the circular edges of spots. The linear minimum mean squared error and maximum a posteriori estimation techniques are used to develop bivariate estimators for the noise-free coefficients of the two images. These estimators are derived by utilizing appropriate joint probability density functions for the image coefficients as well as the noise coefficients of the two channels.

Extensive experimentations are carried out on a large set of cDNA microarray images to evaluate the performance of the proposed denoising methods as compared to the existing ones. Comparisons are made using standard metrics such as the peak signal-to-noise ratio (PSNR) for measuring the amount of noise removed from the pixels of the images, and the mean absolute error for measuring the accuracy of the estimated log-intensity ratios obtained from the denoised version of the images. Results indicate that the proposed denoising methods that are developed specifically for the microarray images do, indeed, lead to more accurate estimation of gene expression levels. Thus, it is expected that the proposed methods will play a significant role in improving the reliability of the results obtained from practical microarray experiments.

ACKNOWLEDGEMENTS

First, I would like to thank Almighty God for the blessings He has bestowed on me that has made the completion of this thesis possible. I am deeply indebted to my Supervisor, Professor Yogendra P. Chaubey for his guidance, motivation, and unwavering support throughout the course of my research. He has been an excellent mentor and was always available to discuss my work. I would like to thank him for introducing me to microarray research and for granting me the freedom to pursue my own ideas and interests within this area.

Constructive suggestions have been provided by members of the Thesis Examination Committee as well as anonymous reviewers of research publications that have improved the quality of this work. Their contribution is gratefully acknowledged.

I would like to thank the staff at the Department of Mathematics and Statistics, in particular, Ann Marie Agnew and Judy Thykootathil, for their help and support. In addition, the financial support that I received from Concordia University through the Concordia University Graduate Fellowship and the Bourses d'etudes Hydro Quebec de l'Universite Concordia Awards is greatly appreciated.

I would also like to give special thanks to my husband, Dr. S. M. Mahbubur Rahman, for providing me with insightful suggestions and feedback in the area of my research. I am grateful to him for his patience and support during my doctoral studies. The most precious gift that I have received during the time of writing this thesis is the birth of my daughter. She has brought me great joy during times of stress. I also wish to thank the University of Dhaka for granting me study leave to complete my doctoral studies at Concordia University. Finally, I thank my parents for their constant encouragement, loving support and prayers for my well-being.

Contents

List of Tables

List of Figures

List of Acronyms

A	:	Adenine
AGN	:	Additive Gaussian Noise
AWGN	:	Additive White Gaussian Noise
cDNA	:	Complementary DNA
CWT	:	Complex Wavelet Transform
Cy3	:	Cyanine-3
Cy5	:	Cyanine-5
C	:	Cytosine
1D	:	One Dimensional
2D	:	Two Dimensional
db	:	Decibels
Db8	:	Daubechies 8 wavelet function
DCT	:	Discrete Cosine Transform
DFT	:	Discrete Fourier Transform
DNA	:	Deoxyribonucleic acid
DT-CWT	:	Dual-Tree Complex Wavelet Transform
DWT	:	Discrete Wavelet Transform
GG	:	Generalized Gaussian
G	:	Guanine
HMT	:	Hidden Markov Tree
IG	:	Inverse-Gamma
i.i.d.	:	Independent and Identically Distributed

IPSNR	:	Input Peak Signal-to-Noise Ratio
LAWMAP	:	Locally-Adaptive Window-based Maximum A Posteriori
LMMSE	:	Linear Minimum Mean Squared Error
MAE	:	Mean Absolute Error
MAP	:	Maximum A Posteriori
ML	:	Maximum Likelihood
MMAE	:	Minimum Mean Absolute Error
MMSE	:	Minimum Mean Squared Error
MoM	:	Method of Moments
mRNA	:	Messenger Ribonucleic Acid
MSE	:	Mean Squared Error
PCR	:	Polymerase Chain Reaction
PMT	:	Photomultiplier Tube
PDF	:	Probability Density Function
PSNR	:	Peak Signal-to-Noise Ratio
ROI	:	Region of Interest
RNA	:	Ribonucleic Acid
$S\alpha S$	:	Symmetric Alpha Stable
SWT	:	Stationary Wavelet Transform
Sym8	:	Symlet wavelet function of order 8
T	:	Thymine
TIFF	:	Tagged Image File Format

List of Symbols

$\mathcal{W}$	:	Transformation matrices of DWT
W	:	Transformation matrices of CWT
ϕ	:	1D scaling function
ψ	:	1D wavelet function
$x_l^{\mathcal{D}}$	:	Detail coefficients in level l
x_J^A	:	Approximate coefficients in level J
Φ	:	2D scaling function
Ψ	:	2D wavelet function
x_l^O	:	Detail coefficients in level l of orientation O for 2D DWT
T, $T_{n(n=1,2,3)}$	:	Threshold for wavelet shrinkage functions
f	:	Pixel intensity of noise-free microarray image
g	:	Pixel intensity of noisy microarray image
ε	:	Pixel intensity of noise
y	:	DWT coefficient or magnitude component of CWT coefficient of noisy signal
x	:	DWT coefficient or magnitude component of CWT coefficient of noise-free signal
v	:	DWT coefficient or magnitude component of CWT coefficient of noise
μ, σ^2	:	Population mean and variance
$\mathcal{N}(\mu, \sigma^2)$	:	Gaussian probability density function with mean μ and variance σ^2

σ_ε^2	:	Population variance of noise in pixel domain
ρ_ε	:	Population inter-channel correlation of noise in pixel domain
σ_v^2	:	Population variance of noise coefficient
ρ_v	:	Population inter-channel correlation of noise coefficients or magnitude of noise coefficients
r, g	:	Red and green channels
$\{x_r, x_g\}$	:	Noise-free DWT coefficient or magnitude component of CWT coefficient
$\{X_r, X_g\}$	:	Random variables of x_r and x_g
$\{\sigma_r, \sigma_g\}$	:	Population standard deviations of X_r and X_g
ρ	:	Population correlation between X_r and X_g
$\boldsymbol{\Sigma_X}$	:	Population covariance matrix of random vector $\boldsymbol{X}$
$I_\vartheta(\cdot)$	:	Modified Bessel function of order ϑ
$F_1(\cdot)$	:	Gaussian hypergeometric function
$\Gamma(\cdot)$	:	Gamma function
$\mathbf{1}\{\cdot\}$	:	Indicator function
$E\{\cdot\}$	:	Mathematical expectation
$\mathrm{tr}\{\cdot\}$	:	Trace of a matrix
$\mathtt{r}, \theta$	:	Magnitude and phase components of complex coefficient
ℓ	:	Log-likelihood function
N	:	Number of samples in a subband
$\mathcal{S}(k)$	:	Local neighborhood centered at k, i.e., (k_1, k_2)
M	:	Number of samples in squared-shape local neighborhood $\mathcal{S}$
a, b	:	Parameters in IG PDF
λ	:	Parameter in exponential PDF
$p(\cdot)$	:	Probability density function

$\Omega(\cdot)$	:	Score function
$w \times w$	:	Size of square-shaped window
(i, j)	:	2D spatial index in pixel-domain
(k_1, k_2)	:	2D spatial index in wavelet-domain
$\mathcal{O}(\cdot)$	:	Order notation
$\mathcal{L}$	:	Length of scaling or wavelet vector
$x^{\mathrm{r}} \angle x^{\theta}$	:	Magnitude and phase component of CWT coefficient x where x^{r} is the magnitude component and $\angle x^{\theta}$ is the phase component

Chapter 1

Introduction

Complementary DNA (cDNA) microarray image technology has transformed molecular genetics, allowing researchers to study the activity of thousands of genes at a time. By examining all genes of a given organism at the same time and possibly under different conditions, scientists can obtain valuable insights on gene function, gene regulation and gene interaction. The application of this technology is increasing in recent years as it becomes a powerful tool for discovering new types of diseases and for predicting or diagnosing the type of disease based on gene expression measurements.

The raw data of cDNA microarray experiments is a pair of fluorescent intensity images that measure the relative activity of genes under study. DNA microarray image processing is an important information extraction problem occurring in molecular biology and bioinformatics. Since, a microarray experiment is a multi-phased process with various sources of noise, the resulting images are often very noisy. The extraction of accurate gene expression measurements from these noisy images is a challenging problem and an area of intense scientific research.

1.1 cDNA Microarray Images

The cDNA microarray image technology is a tool geared at measuring the 'activity' of a gene. The two-channel cDNA microarray is designed to measure the activity of

a set of genes under two conditions, namely, treatment and control (Zhang, 2006). Very simply, a typical two-channel microarray experiment consists of the following steps. Messenger RNA (mRNA) from control and treatment samples are converted into cDNA, labelled with fluorescent dyes (green Cy3 dye for control, red Cy5 dye for treatment) and mixed together. The mixture is then washed over a slide spotted with probes, which are DNA sequences from known genes. A given strand of cDNA bonds with the spot representing the gene which produced that transcript. Next, the slide is scanned producing two 16-bit images, one for the green channel and another for the red. These image files are the raw data of the microarray experiment. Each spot on the images consists of a number of pixels (typically, 50-400), wherein the brightness of each pixel reflects the amount of Cy3 or Cy5 at the spatial location corresponding to that pixel. Thus, one can identify the genes that are differentially expressed between the two samples by comparing the pixel intensities of each spot in the red and green channel images. A more detailed description of the concepts behind microarray technology and the steps involved in microarray experiments is given in Chapter 2.

Microarray experiments involve a large number of error-prone steps which result in a high level of noise in the resulting images (Ruosaari and Hollmen, 2002; Mastriani and Giraldez, 2006). The sources of noise can be divided into two categories: instrument noise and microarray noise. Instrument noise is produced by the imaging system itself. A microarray scanner is a complex instrument containing electrical, optical, and mechanical parts. Consequently, there are many sources of distortions that affect the final microarray images. The different types of instrument noise are photon noise, laser light reflection, dark current noise, electronic noise and quantization noise (Zhang et al., 2002; Lukac et al., 2004; Mastrogianni et al., 2008). Microarray noise results from local contamination during the printing, hybridization and washing

stages of the experiment, dust on the glass slide and by non-specific hybridization to the probes on the microarray surface (Zhang et al., 2004). The presence of noise causes serious distortions in the pixel intensities and produces variations in the image such as different spot diameters, variable shape or contour, high background and/or low foreground, spatial artifacts, etc.

1.2 Motivation

One of the most important steps of microarray experiments is the conversion of the microarray image into numerical information that quantifies gene expression. This step is known as microarray image analysis. In cDNA microarray experiments, the log-intensity ratio is widely used as a measure of relative abundance of a gene in the treatment and control samples (Jornsten et al., 2002, 2003). This quantity is later used in downstream analysis such as gene clustering (Tseng and Kao, 2005) and sample classification (Xiong et al., 2007). The success of cDNA microarray data analysis is, therefore, critically dependent on the accuracy of the estimated log-intensity ratios.

There are three major steps involved in extracting the log-intensity ratios from microarray images. The first step is gridding, which consists of identifying target areas within the image. Target areas are non-overlapping regions of the image each enclosing one spot (Bozinov and Rahnenfuhrer, 2002). In the second step, pixels within the target area that correspond to the spot must be found. This is performed by segmentation, which consists of partitioning the target area into the spot and its local background. The presence of noise results in inaccurate spot segmentation (Daskalakis et al., 2007). For instance, when histogram-based methods of segmentation are used, noise or artifacts may be classified into the spot leading to errors in the estimated spot intensity values (Zhang et al., 2004). In a noisy image, it is difficult to distinguish low intensity spots from their background and this often results in large

measurement errors. The next step is to calculate the log-intensity ratio from the spot pixel intensities. Common practices for combining pixel data in a spot from both images into an estimate of the log-intensity ratio include forming the simple ratio of average pixel values or forming the ratio of pixel medians.

Significant differences are observed between values of the log-intensity ratio obtained from the noisy images and the estimated noise-free or 'denoised' images. This indicates that the log-intensity ratio is highly sensitive to the presence of noise. The noise must therefore be reduced prior to calculating the log-intensity ratio to prevent erroneous biological conclusions. To address the problem of noise in microarray experiments, one approach has been to develop statistical methods for estimating 'true' gene-expression measurements from the noisy image. In such an approach, error models have been used to describe the relationship between real and observed intensity values while taking into account the cumulative effects of chemical, optical, and computational factors introduced by the microarray technology (Ideker et al., 2000; Goryachev et al., 2001). However, the process of assessing and controlling the relative contributions of the many sources of noise in a microarray experiment is difficult. These methods are therefore complicated and yet to be standardized. Instead of tackling the problem at the analysis stage, alternatively, one could approach the problem from the top. This involves processing the noisy image so as to obtain an estimate of the noise-free image (denoising), from which, the true gene-expression measurements could then be extracted. Once the denoised image is obtained, the process of estimating the true gene-expression measurements becomes a relatively straightforward and easy process. The value of this approach is being increasingly recognized as more and more research articles are being devoted to the development of efficient algorithms for reducing noise in cDNA microarray images (O'Neill et al., 2003; Lukac and Smolka, 2003; Wang et al., 2003b; Lukac et al., 2004; Mastriani and

Giraldez, 2006).

There are some commercially available microarray image processing softwares for estimating the log-intensity ratios such as QuantArray, Spot, and GenePix. These softwares use low-level noise reduction techniques, such as the median filter, for processing the microarray images. However, these methods are insufficient, and the presence of noise results in erroneous segmentation and intensity extraction that leads to inaccurate estimation of the log-intensity ratios (Daskalakis et al., 2007). Thus, use of efficient algorithms for reducing noise is essential for microarray image analysis. Noise removal in microarray images can be performed both in the pixel-domain as well as in the transform domain. For instance, O'Neill et al. (2003) proposed a noise reduction technique that is implemented in the pixel-domain. This method constructs a mask of the artifact, which is used to recreate the image of the noise. The noise image is then subtracted from the observed image to leave the noise-free signal. Lukac and Smolka (2003) proposed a method based on center-weighted vector median filters to attenuate noise in microarray images. In a later paper, Lukac et al. (2004) proposed two types of order-statistic filters for removing noise, one for the spot and another for the background.

Although the pixel-based methods are simpler to implement in general, methods developed using an appropriate transform domain are more efficient in reducing noise. Various kinds of transforms such as the discrete cosine transform (DCT), the discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) are available for processing microarray images. However, among the various transforms, the DWT has enjoyed greater success in reducing the noise in a signal (including an image) due to its space-frequency localization property and the freedom to choose different kinds of basis functions (Mallat, 1999). Unlike the other transforms, the DWT has been used extensively in signal and image denoising due to its high energy compaction proper-

ties (see for example, Donoho and Johnstone, 1994; Gao and Bruce, 1997; Gao, 1998; Donoho and Johnstone, 1995; Chang et al., 2000a; Cai and Silverman, 2001; Johnstone and Silverman, 2005). Moreover, the decimated DWT, which is also known as the classical DWT, is nonredundant[1] and thus denoising methods based on this transform have computational complexity that is comparable to methods based on other transforms such as the DFT and DCT (Mallat, 1999; Sendur and Selesnick, 2002a). It is to be noted that fast implementation of the denoising algorithm is essential for processing the huge data of microarray experiments. However, the decimated DWT lacks the desirable shift-invariance and directional selectivity properties (see Chapter 2 for details) that are important in the context of microarray image denoising. Thus, any variant of the DWT that has these features is expected to perform better.

The success of wavelet-based methods for reducing noise in signals as well as standard images has motivated some research in the context of microarray images as well (see for example, Wang et al., 2003b; Mastriani and Giraldez, 2006; Mastrogianni et al., 2008). Moreover, the DWT being a multiresolution analysis allows one to efficiently process an image at more than one resolution, and thus, it is gaining attention among researchers in the development of new techniques for conducting several tasks of microarray image processing and data analysis, such as, gridding (Athanasiadis et al., 2007), spot recognition (Bidaut et al., 2006; Wang et al., 2003a), and analysis of differential gene expression (Turkheimer et al., 2004, 2006). Hence, an efficient wavelet-based noise reduction algorithm may be seamlessly embedded into the routines of such wavelet-based techniques for analyzing gene expression data so that the entire process of image processing and data analysis becomes faster, automated, and more efficient.

The existing wavelet-based techniques for denoising of signals or standard images (see for example, Donoho, 1995; Donoho and Johnstone, 1995; Gao, 1998; Chang et al.,

[1]Number of transform coefficients is equal to the number of data samples

2000a; Johnstone and Silverman, 2005; Mihçak et al., 1999; Cai and Silverman, 2001) are capable of 'individually' processing the red and green channel images. However, significant positive correlation exists between the pixel intensities as well as DWT coefficients of the two channels at the same spatial location (Zhang et al., 2005b; Davies and Seale, 2005; Howlader and Chaubey, 2009b). In addition, it is expected that there will be significant inter-channel noise-correlation since the factors that affect image formation are similar for both channels. Thus, by processing the images separately, one is, in fact, ignoring the information content in the other channel. For this reason, existing methods for image denoising are less efficient for microarray image processing applications.

1.3 Objectives

The main objective of this thesis work is to develop new noise reduction algorithms for cDNA microarray images in the wavelet domain. Since the purpose of a microarray image is not for visual inspection, but rather, for the extraction of information regarding gene expression levels, our aim is to develop algorithms that reduce noise with minimal loss of information for accurate estimation of gene expression levels.

The theoretical development of these new algorithms is based on the observation that there exists significant correlation between the wavelet coefficients of the red and green channel images of cDNA microarray experiments at the same spatial location. When such an association exists, a 'joint estimation' technique should provide better estimates of the noise-free coefficients than a method that individually processes the two channels. It may be mentioned that correlations may exist with coefficients at other spatial locations as well, however, these correlations will be much weaker. Denoising algorithms that use the wavelet transform consist of three basic steps: (i) calculating the wavelet transform of the noisy signal (ii) modifying the noisy wavelet

coefficients according to some rule, and (iii) computing the inverse transform using the modified coefficients. The problem considered in this research primarily deals with step (ii), i.e., developing efficient bivariate estimators for the noise-free coefficients of the red and green channel images.

The proposed algorithms are implemented both in the decimated DWT domain as well as in the complex wavelet transform (CWT) domain. Although the decimated DWT is frequently used in image denoising, methods are also developed in the CWT domain because the latter has desirable properties not found in the decimated DWT that allow better detection of the circular edges of spots on a microarray image. To construct the bivariate estimators, we consider minimization of various cost functions, such as the mean squared error (MSE) in the case of linear minimum mean squared error (LMMSE) estimation and the uniform cost function in the case of maximum a posteriori (MAP) estimation. For a good estimation performance, the bivariate estimators are derived by considering appropriate joint prior functions for the wavelet coefficients of the two images that take into account the inter-channel signal correlation. The inter-channel noise correlation is also considered in the estimation by defining appropriate joint probability density functions for the noise coefficients in the two channels. To evaluate the performance of the proposed algorithms, extensive simulation experiments are performed using a large set of cDNA microarray images. Since the goal of microarray image denoising is to retrieve accurate estimates of gene-expression levels from noisy microarray images, the performance of a denoising algorithm should be evaluated both in terms of the amount of noise reduction as well as accuracy of the estimated log-intensity ratio. Thus, the proposed methods are compared with existing wavelet-based methods using two performance criteria: denoising performance (in the MSE sense) and accuracy of the estimated log-intensity ratios.

1.4 Organization of the Thesis

This thesis is organized as follows. Some background on the steps of microarray experiments and wavelet-based image denoising are given in Chapter 2. The contents of this chapter facilitate the understanding of the algorithms developed in the subsequent chapters. In Chapters 3 and 4, respectively, we present our proposed algorithms for microarray image denoising in the decimated DWT and the CWT domains. Bivariate wavelet estimators are derived considering correlation between the signal as well as the noise coefficients in the two channels using two separate estimation techniques. Simulation experiments are conducted on a set of cDNA microarray images to evaluate the performance of the proposed algorithms. We conclude with Chapter 5 which summarizes the main theme and overall findings of this research work and provides directions for future study.

Chapter 2

Microarray Experiments and Wavelet-Based Denoising

2.1 Introduction

Microarray imaging technology is one of the recent advances in bioinformatics and molecular biology that provides 'snapshots' of which genes are expressed in cells of various tissues and diseases at a particular point in time. Each microarray experiment consists of a long chain of delicate steps from tissue acquisition to microarray data analysis, and typically, errors and distortions can get magnified in each step. Most of these errors show up as noise in the microarray images. Although there are different types of microarrays, our focus is on the cDNA microarray, also known as the spotted array. In the first part of this chapter, we discuss the concepts behind cDNA microarray experiments and explain some details of the steps that are involved.

Microarray image processing is a critical aspect of microarray experiments. Although the basic goal of extracting the pixel intensities on each spot is straightforward, the presence of noise in the image makes it a complex process. In image signal processing, the signal is often transformed because the transformed data exhibits certain properties that enable one to efficiently handle the problem in question. For instance, denoising can be performed more efficiently in the DWT domain than in the pixel domain because the former being a multiresolution analysis, allows one to

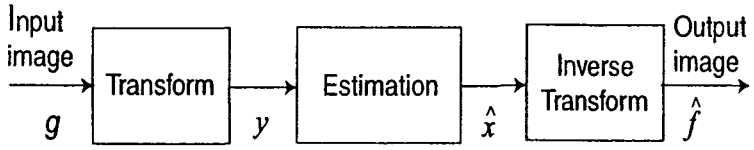


Figure 2.1: Block diagram of a transform-based image denoising algorithm.

process an image at more than one resolution. The motivation for using wavelet transform for reducing noise from a microarray image has been discussed in Chapter 1. Any transform-based, including the DWT-based, image denoising scheme is shown in Fig. 2.1 with a simple block diagram. In this figure, g represents the noisy pixel intensities, y are the noisy transform coefficients, $\hat{x}$ are the estimated noise-free coefficients and $\hat{f}$ are the estimated noise-free pixel intensities. The first step is the transformation of spatial information into the coefficient domain. The second step is to estimate the transform coefficients from their noisy observations. Finally, an inverse transformation is necessary to obtain the denoised estimate of the microarray image. The wavelet transform has several variants, each having certain advantages. For example, the decimated DWT is suitable to obtain a fast denoising algorithm, since such a transform is non-redundant. However, this transform is shift-variant and has poor directional selectivity. The stationary wavelet transform (SWT) is shift-invariant at a cost of huge computational complexity and its directional selectivity is the same as that of the decimated DWT. The CWT is a special type of wavelet transform that overcomes some of the limitations of the decimated DWT and SWT. Its notable features include near shift-invariance property, improved directional selectivity, and low computational complexity.

In the second part of this chapter, a brief introduction to the wavelet transform (Daubechies, 1992; Mallat, 1999; Goswami and Chan, 1999; Strang and Nguyen, 1996) is presented. The mechanism for obtaining the one dimensional (1D) wavelet coefficients from a 1D signal using the concept of multiresolution analysis is shown. Next,

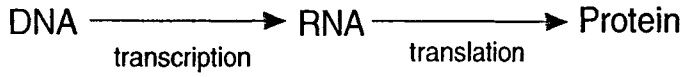


Figure 2.2: Graphical illustration of the conversion of genetic information into proteins.

a procedure for calculating the 2D DWT, SWT, and CWT coefficients by extending the concept of 1D wavelet theory is presented. Finally, a brief review of some of the standard wavelet-based denoising methods is given.

2.2 Overview of cDNA Microarray Experiments

The cells of living organisms contain inheritable (genetic) information stored in a molecule known as deoxyribonucleic acid (DNA). DNA are double-stranded molecules. Each strand is composed of a long string of nucleotides and each nucleotide contains one of four bases [adenine (A), guanine (G), cytosine (C) or thymine (T)]. The two complementary strands of the DNA molecule are held together by the exclusive base pairing of A with T, and G with C. This basic chemical fact of complementarity lies at the basis of each microarray. A gene is a segment of DNA that maps into a specific protein. In other words, the production of proteins can be thought of as the function of a gene. Gene expression is the process by which genetic information at the DNA level is converted to functional proteins. Fig. 2.2 summarizes the two main steps of gene expression: transcription and translation. Transcription is the first step in gene expression in which mRNA (single stranded molecules similar to DNA) is synthesized from a DNA template. Translation is the process by which mRNA are used as patterns to produce protein. Thus, the mRNA delivers DNAs genetic message to the cytoplasm of a cell where proteins are made.

A microarray is an ordered array of microscopic elements (or spots) on a planar substrate (nylon membrane, glass or plastic) that allows the specific binding of genes or gene products. The purpose of a microarray is to measure the amount of message

that was broadcast through the RNA for each gene in the genome. The two-channel cDNA microarray is designed to measure the transcriptional activity of a set of genes under two conditions, namely, treatment and control (Zhang, 2006).

A typical two-channel microarray experiment consists of the following steps. The first step is array fabrication in which a set of previously known cDNA sequences called probes are printed onto the array using a robotic arrayer. The probes are produced by polymerase chain reaction (PCR), which is a technique that 'amplifies' or replicates DNA fragments, and then purifies it to reduce the presence of unwanted components. The selection of the set of probes depends on the experiment; usually genes that are relevant to the biological questions under investigation are selected. When different probes matching all mRNAs in a cell are used, a snapshot of the total mRNA pool of a living cell or tissue can be obtained. The microarray spots are therefore collections of probe molecules (segments of cDNA strands) arranged in grids that allow specific binding of target molecules. In the second step, mRNAs are separately isolated from the experimental and control samples and converted into cDNA through a procedure known as reverse transcription. The cDNA is labelled with a fluorescent green (Cy3) or red (Cy5) dye, the green dye being used for the control and the red dye for the treatment. The third step is known as hybridization in which the two labelled target cDNAs are mixed in equal proportions and then applied to the microarray containing probe cDNAs in the spots. If the probe and target cDNAs are complementary of each other then they should be bound by their base pairs. Thus, a given strand of target cDNA bonds with the spot representing the gene which produced that transcript. The strength of the binding depends on the amount of the gene expression in the target samples. For example, if a gene (spotted on the array) is more expressed in the experimental cell (labelled with Cy3) than in the control cell, then Cy3-molecules should bind more to that gene spot compared to

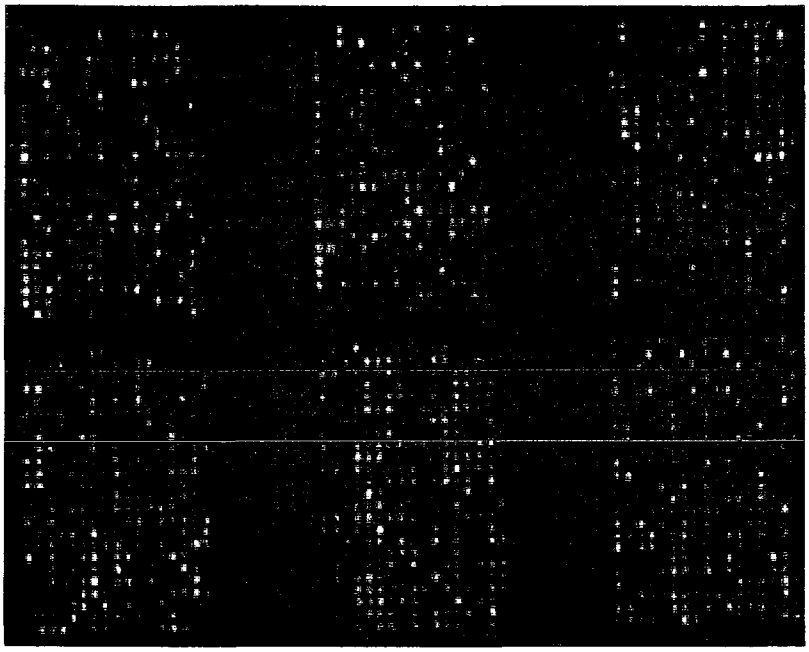


Figure 2.3: Combined false color image for cDNA microarray (Zhang, 2006). Red spots in the combined image correspond to spots that are expressed more in channel one. Green spots correspond to those expressed more in channel two. Yellow spots have a similar level of expression in both channels. Dark spots are low expressed in both channels.

the Cy5-molecules. After sufficient time is allowed for this competitive hybridization, the microarray is carefully washed a number of times so that all the unbound target cDNAs are washed off. The next steps of a microarray experiment are image analysis and data extraction.

In image analysis, a confocal laser microscope is used to scan the array at two channels using two wavelengths, one for the Cy5 fluorescent-tagged sample (say, Channel 1) and another for the Cy3 fluorescent-tagged sample (say, Channel 2). A detector captures the emitted photons, measures and records their levels and subsequently converts the measurements into an electric current by a photomultiplier tube (PMT). This in turn is digitized into pixel intensities and stored in tagged image file format

(TIFF). Two 16-bit TIFF images are generated corresponding to the two samples under investigation. A 16-bit image has a dynamic range of [0, 65535] (Zhang, 2006). Both images consist of spots arranged in regular grid-like patterns. These images are considered as the 'raw' data for the microarray experiment. The images corresponding to Cy3 (i.e., green) and Cy5 (i.e., red) can be overlayed to produce a single false-color image as shown in Fig. 2.3. By analyzing the location and intensity of the fluorescent signals of the red and green channel images, one can determine the level of activity of each gene. For instance, any spot whose intensity is different between the two channels, corresponds to a gene that is differentially expressed in the treated versus control group. Fig. 2.4 summarizes the steps in a microarray experiment. Further details regarding the set-up of microarray experiments may be found in Wit and McClure (2004) and McLachlan et al. (2004) and issues relating to experimental design for microarrays is discussed by Churchill (2002).

2.3 Preliminaries on DWT

2.3.1 1D DWT

Let $f(i)$, $i = 1, 2, \cdots, N_1$ represent a 1D signal of size $N_1 \times 1$, where i is the 1D index. The DWT of the signal is given by (Gonzalez and Woods, 2002)

$$f(i) = \frac{1}{\sqrt{N_1}} \left[\sum_{k_1=1}^{N_1} x_J^A(k_1)\phi_J^A(i,k_1) + \sum_{l=1}^{J} \sum_{k_1=1}^{N_1} x_l^{\mathcal{D}}(k_1)\psi_l^{\mathcal{D}}(i,k_1) \right] \tag{2.1}$$

where x_J^A denotes the approximate coefficients in the largest level J, $x_l^{\mathcal{D}}$ denotes the detail coefficients in a decomposition level l ($l = 1, 2, \cdots, J$), $\phi_J^A(i,k_1) = 2^{J/2}\phi(2^J i - k_1)$, $\psi_l^{\mathcal{D}}(i,k_1) = 2^{l/2}\phi(2^l i - k_1)$, and ϕ and ψ, respectively, are the scaling and wavelet functions. The approximate and detail coefficients can be obtained as

$$x_J^A(k_1) = \frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} f(i)\phi_J^A(i,k_1) \qquad x_l^{\mathcal{D}}(k_1) = \frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} f(i)\psi_l^{\mathcal{D}}(i,k_1) \tag{2.2}$$

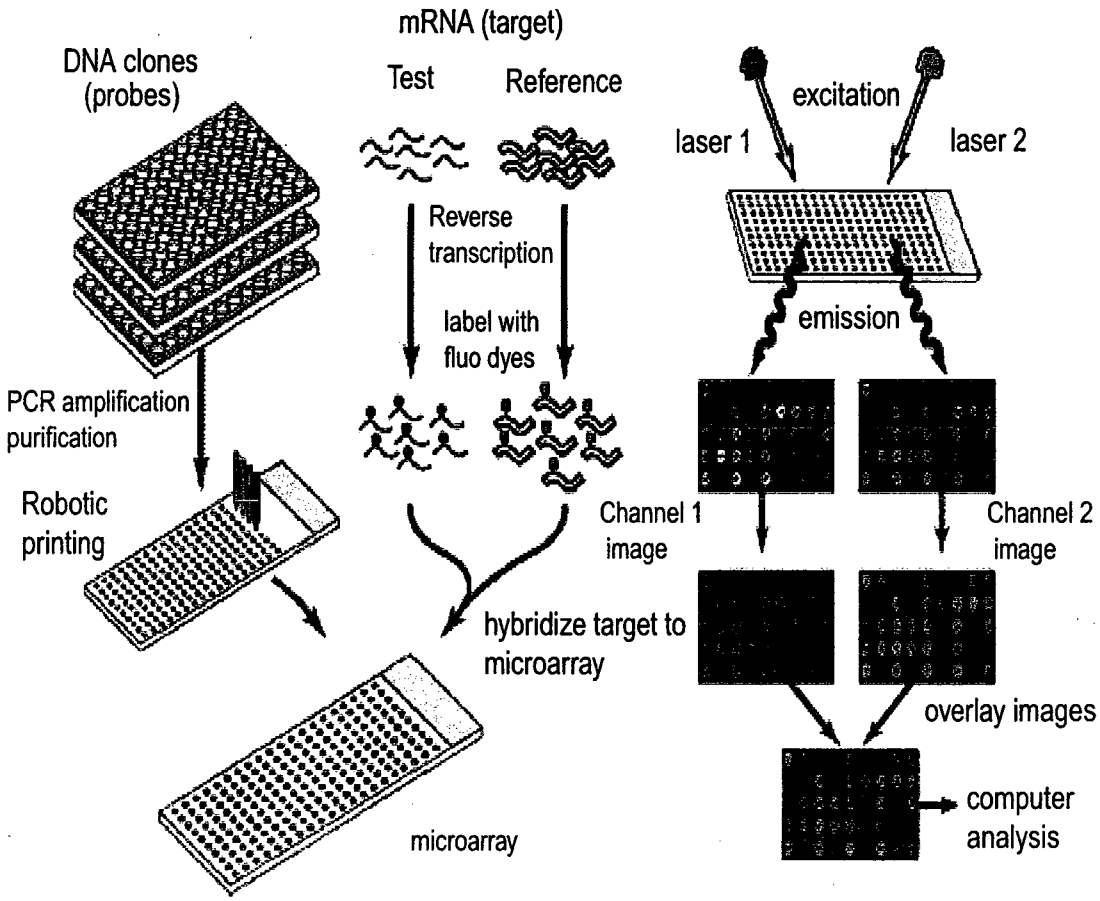


Figure 2.4: Steps in a cDNA microarray experiment (Zhang, 2006).

The functions ϕ and ψ are chosen in such a way that the signal can be reconstructed from the DWT coefficients without error. Since the DWT is a type of multiresolution analysis, the nesting property requires the scaling function ϕ and wavelet function ψ satisfy the two-scale dilation equations

$$\phi(i) = \sqrt{2} \sum_{u=-\infty}^{\infty} h_\phi(u)\phi(2i-u) \qquad \psi(i) = \sqrt{2} \sum_{u=-\infty}^{\infty} h_\psi(u)\phi(2i-u)$$

where $\boldsymbol{h}_\phi$ and $\boldsymbol{h}_\psi$, respectively, are called the scaling and wavelet vectors for forward DWT. Using the above relations, Mallat (1999) has shown that the 1D DWT coefficients of adjacent decomposition levels can be estimated very efficiently by the

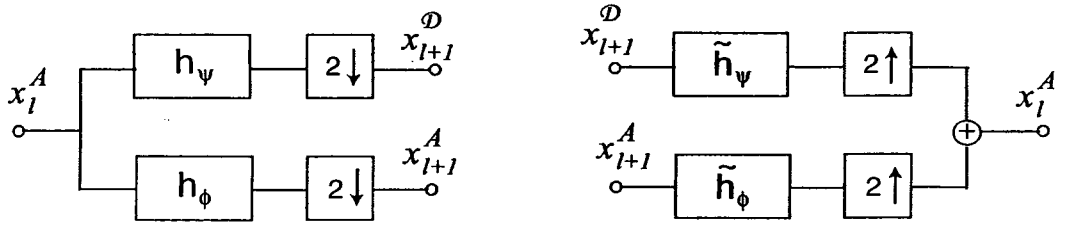


Figure 2.5: Block diagram of the 1D DWT using the scaling and wavelet vectors at the analysis and synthesis stages. (a) Forward transform for obtaining the $(l+1)$-th level detail coefficients. (b) Inverse transform for obtaining the l-th level approximate coefficients.

following

$$x_{l+1}^A(k_1) = \sum_{u=-\infty}^{\infty} h_\phi(u-2k_1)x_l^A(u) = h_\phi(-u) * x_l^A(u)\Big|_{u=2k_1, k_1\geq 0} \tag{2.3}$$

$$x_{l+1}^{\mathcal{D}}(k_1) = \sum_{u=-\infty}^{\infty} h_\psi(u-2k_1)x_l^A(u) = h_\psi(-u) * x_l^A(u)\Big|_{u=2k_1, k_1\geq 0} \tag{2.4}$$

where $*$ is a convolution operator. The above equations imply that the approximate and detail coefficients at decomposition level $(l+1)$, respectively, can be obtained from convolving the approximate coefficients at decomposition level l by $\boldsymbol{h}_\phi$ and $\boldsymbol{h}_\psi$ and down-sampling by 2. The coefficients at the $(l+1)$-th level may also be obtained by performing the matrix operations $\boldsymbol{x}_{l+1}^A = \mathcal{W}_\phi \boldsymbol{x}_l^A$ and $\boldsymbol{x}_{l+1}^{\mathcal{D}} = \mathcal{W}_\psi \boldsymbol{x}_l^A$, where $\mathcal{W}_\phi$ and $\mathcal{W}_\psi$ are circulant matrices having $\boldsymbol{h}_\phi$ and $\boldsymbol{h}_\psi$ as their first columns, respectively, and $\boldsymbol{x}_l^A$ is the vector of approximate coefficients at the l-th level. It can be found that the inverse transform of the 1D DWT coefficients of adjacent decomposition levels can also be estimated very efficiently by the following

$$x_l^A(k_1) = \tilde{h}_\phi(u) * x_{l+1}^A(u)\Big|_{u=\frac{k_1}{2}, k_1\geq 0} + \tilde{h}_\psi(u) * x_{l+1}^{\mathcal{D}}(u)\Big|_{u=\frac{k_1}{2}, k_1\geq 0} \tag{2.5}$$

where $\tilde{\boldsymbol{h}}_\phi$ and $\tilde{\boldsymbol{h}}_\psi$, respectively, are called the scaling and wavelet vectors for inverse DWT. The above equation implies that the approximate coefficients at decomposition level l can be obtained from convolving the up-sampled (by inserting zero in odd or even indices) approximate and detail coefficients at decomposition level $(l+1)$.

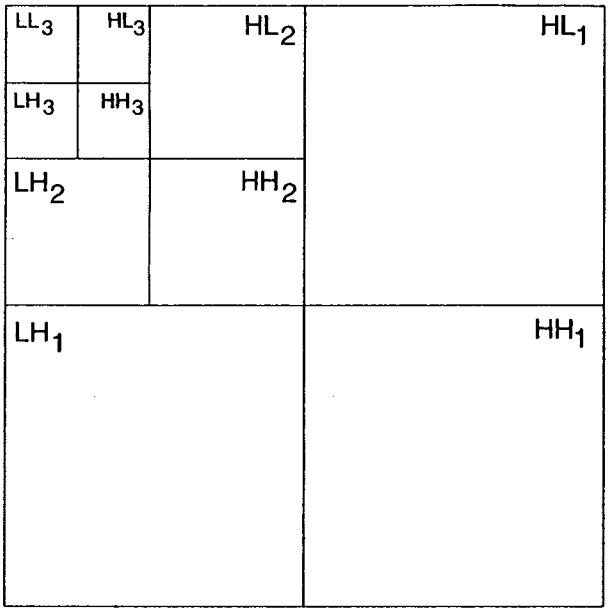


Figure 2.6: Subband representations of a 3-level 2D DWT.

respectively, by $\tilde{\boldsymbol{h}}_\phi$ and $\tilde{\boldsymbol{h}}_\psi$ and adding the results. It may be pointed out that to maintain the perfect reconstruction requirement, $\tilde{\boldsymbol{h}}_\phi$ is time reversal of $\boldsymbol{h}_\phi$ and $\tilde{\boldsymbol{h}}_\psi$ is time reversal of $\boldsymbol{h}_\psi$. A simple block diagram of the forward 1D DWT using the analysis scaling-and-wavelet vectors and inverse 1D DWT using the synthesis scaling-and-wavelet vectors for adjacent decomposition levels is shown in Fig. 2.5. It is to be noted that there is no data redundancy in the DWT coefficients due to the down-sampling of convolved coefficients after each level decomposition. That is, the number of DWT coefficients is the same as the number of data samples. In literature, this is very often referred to as the decimated 1D DWT.

2.3.2 2D DWT

Let $f(i,j), i = 1, 2, \cdots, N_1, j = 1, 2, \cdots, N_2$ represent pixels of 2D signal (e.g., image) of size $N_1 \times N_2$, where (i,j) is the 2D index. The DWT of this signal is given by

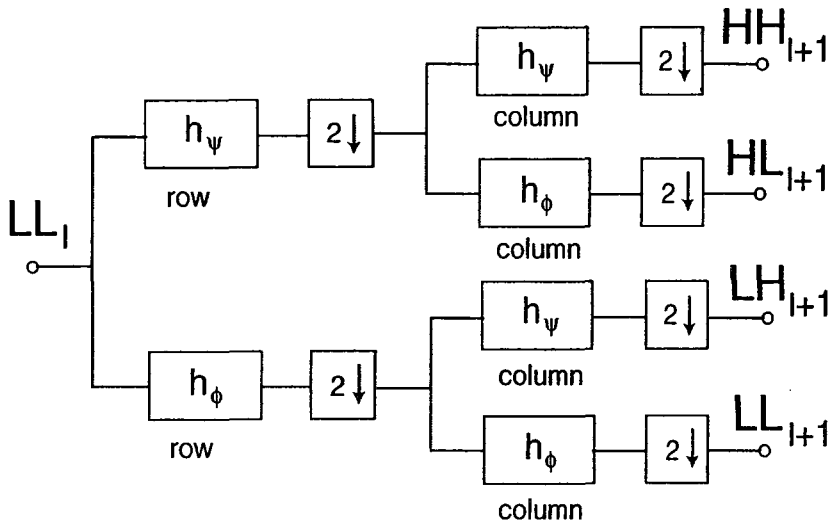


Figure 2.7: Block diagram of the 2D forward DWT using the analysis scaling-and-wavelet vectors for obtaining the $(l+1)$-th level detail coefficients.

(Gonzalez and Woods, 2002)

$$f(i,j) = \frac{1}{\sqrt{N_1 N_2}}\Big[\sum_{k_1=1}^{N_1}\sum_{k_2=1}^{N_2} x_J^A(k_1,k_2)\Phi_J^A(i,j,k_1,k_2) + \sum_{l=1}^{J}\sum_{O\in H,V,D}\sum_{k_1=1}^{N_1}\sum_{k_2=1}^{N_2} x_l^O(k_1,k_2)\Psi_l^O(i,j,k_1,k_2)\Big] \tag{2.6}$$

where x_l^O $(O \in H,V,D)$ denotes the detail coefficients in the decomposition level l $(l = 1,2,\cdots,J)$ of orientation O, and Φ and Ψ, respectively, are the 2D scaling and wavelet functions. Separable scaling and wavelet functions are common for extending the 1D DWT to the multidimensional DWT. Hence, the 2D scaling and wavelet functions can be represented in terms of 1D scaling and wavelet functions as

$$\Phi_J^A(i,j,k_1,k_2) = 2^{J/2}\phi(2^J i - k_1)\phi(2^J j - k_2)$$

$$\Psi_l^H(i,j,k_1,k_2) = 2^{l/2}\psi(2^l i - k_1)\phi(2^l j - k_2)$$

$$\Psi_l^V(i,j,k_1,k_2) = 2^{l/2}\phi(2^l i - k_1)\psi(2^l j - k_2)$$

$$\Psi_l^D(i,j,k_1,k_2) = 2^{l/2}\psi(2^l i - k_1)\psi(2^l j - k_2)$$

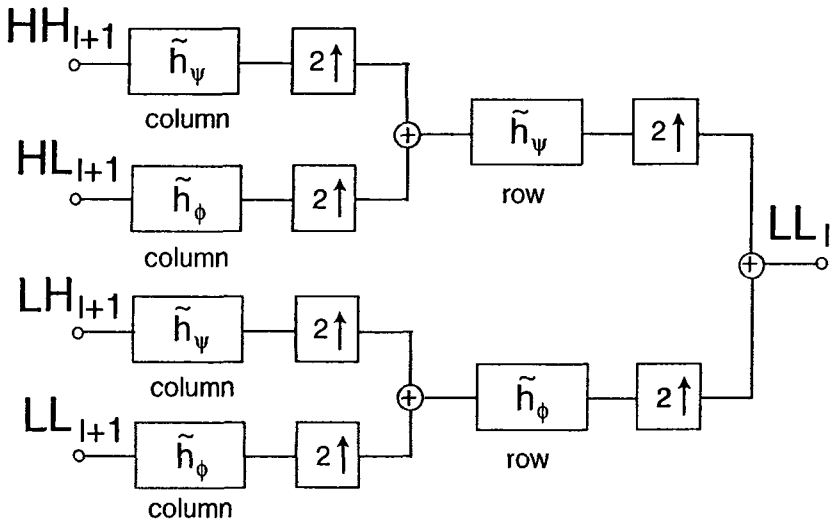


Figure 2.8: Block diagram of the 2D inverse DWT using the synthesis scaling-and-wavelet vectors for obtaining the l-th level approximate coefficients.

The approximate and detail coefficients of the 2D DWT can be obtained as

$$x_J^A(k_1,k_2) = \frac{1}{\sqrt{N_1N_2}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(i,j)\Phi_J^A(i,j,k_1,k_2) \tag{2.7}$$

$$x_l^O(k_1,k_2) = \frac{1}{\sqrt{N_1N_2}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} f(i,j)\Psi_l^O(i,j,k_1,k_2) \tag{2.8}$$

These coefficients are clustered into groups or subbands of different levels and orientations. The subbands HL_l, LH_l, and HH_l ($l = 1,2,\cdots,J$) contain the detail coefficients of the horizontal (0°), vertical (90°), and diagonal (±45°) orientations, viz., x_l^H, x_l^V, and x_l^D, respectively. The subband LL_J is the lowest resolution residual that contains x_J^A. The subband representation of a typical 3-level 2D DWT is shown in Fig. 2.6. Due to the separability of the scaling and wavelet functions, the 2D DWT may be implemented using two separate 1D DWT. Similar to the 1D DWT, the fast and efficient implementation of the forward and inverse of 2D DWT is obtained by using the analysis and synthesis scaling-and-wavelet vectors. Both the forward and inverse transforms use two-stage convolution, one through columns and

another through rows. A simple block diagram of the forward 2D DWT is shown in Fig. 2.7 and the inverse one in Fig. 2.8.

Orthogonality, symmetry, and short support of the wavelet function are some of the desirable characteristics of the 2D DWT. Orthogonality decorrelates the transform coefficients, thereby minimizing the redundancy. Symmetry permits a symmetric boundary extension that minimizes border artifacts. Short support property is valuable for reducing truncation artifacts in the reconstructed images. The choice of wavelet function (e.g., orthogonal, biorthogonal, or spline) is dependent on the type of application. For example, biorthogonal or spline wavelet functions are preferable for image compression, since a symmetric boundary extension or a low truncation error of the transform coefficients improves compression performance. On the other hand, denoising requires higher decorrelation efficiency and, therefore, orthogonal wavelet functions are more preferable for this purpose. Examples of orthogonal wavelet functions that are commonly used in denoising are Daubechies 8 wavelet function (Db8) and the Symlet wavelet function of order 8 (Sym8) (Daubechies, 1992). The Db8 wavelet function consists of the following scaling and wavelet vectors for forward DWT:

$$\begin{aligned}
\boldsymbol{h}_{\phi}^{T} &= \Big[-0.0001, 0.0007, -0.0004, -0.0049, 0.0087, 0.0140, -0.0441, -0.0174, 0.1287, \\
&\qquad 0.0005, -0.2840, -0.0158, 0.5854, 0.6756, 0.3129, 0.0544\Big] \\
\boldsymbol{h}_{\psi}^{T} &= \Big[-0.0544, 0.3129, -0.6756, 0.5854, 0.0158, -0.2840, -0.0005, 0.1287, 0.0174, \\
&\qquad -0.0441, -0.0140, 0.0087, 0.0049, -0.0004, -0.0007, -0.0001\Big].
\end{aligned}$$

The scaling and wavelet vectors of the Sym8 wavelet function are given by:

$$\begin{aligned}
\boldsymbol{h}_{\phi}^{T} &= \Big[-0.0034, -0.0005, 0.0317, 0.0076, -0.1433, -0.0613, 0.4814, 0.7772, 0.3644, \\
&\qquad -0.0519, -0.0272, 0.0491, 0.0038, -0.0150, -0.0003, 0.0019\Big]
\end{aligned} \tag{2.9}$$

$$\boldsymbol{h}_{\psi}^{T} = \Big[-0.0019, -0.0003, 0.0150, 0.0038, -0.0491, -0.0272, 0.0519, 0.3644, -0.7772,$$
$$0.4814, 0.0613, -0.1433, -0.0076, 0.0317, 0.0005, -0.0034\Big].$$

2D SWT

Although down-sampling in the DWT allows perfect reconstruction of 2D signal (Mallat, 1999), the DWT is not shift-invariant (Coifman and Donoho, 1995). In other words, a small shift in the input signal may cause significant variation in the distribution of energy between the DWT coefficients at different decomposition levels. Hence, processing of the 2D DWT coefficients may not provide desired performance in some applications, such as image denoising. To overcome this problem, the SWT coefficients are calculated by the same procedure as the DWT, except the down-sampling is avoided to account for the all possible shifts. Such a transform is often called the overcomplete representation or shift-invariant form or non-decimated form of the DWT. The data redundancy in the SWT increases to a value of 2^{ld}, for d dimensional signal (Chang et al., 2000b; Coifman and Donoho, 1995).

2.3.3 2D CWT

The decimated 2D DWT-based image denoising have two major drawbacks, namely, the lack of shift-invariance and the poor directional selectivity. The SWT resolves only the first problem, however, at the expense of huge computational load. The second problem is inherent for the SWT, since it uses the 2D scaling and wavelet functions that are real and separable. In other words, both the decimated DWT and SWT coefficients capture only three directional features in the 2D signal, namely, the horizontal (0°), vertical (90°), and diagonal (±45°) features. In order to improve the directional selectivity, the CWT has been introduced that has complex valued scaling and wavelet functions, viz., $\Phi_1 + \imath\Phi_2$ and $\Psi_1 + \imath\Psi_2$, such that Φ_1 and Φ_2

as well as Ψ_1 and Ψ_2 form Hilbert pairs[1] (Selesnick et al., 2005; Selesnick, 2001, 2002; Tay et al., 2006). The choice of scaling and wavelet functions of CWT is such that this transform can capture six directional features, namely, -15^o, -45^o, -75^o, 15^o, 45^o, and 75^o of an image. Therefore, the CWT has better directional selectivity as compared to both the decimated DWT and SWT. Various methods has been proposed for obtaining the CWT coefficients (Portilla and Simoncelli, 2000; Kingsbury, 1999, 2001; Selesnick, 2001, 2002; Fernandes et al., 2003; Clonda et al., 2004; Barber and Nason, 2004; Selesnick et al., 2005; Olkkonen et al., 2006), however, due to the simplicity of implementation and sufficiently low redundancy, the dual-tree CWT (DT-CWT) that has been proposed by Kingsbury (1999) and later generalized by Selesnick (2001), is becoming popular. The DT-CWT consists of two trees of DWT in parallel and provides four pairs of subbands, namely, $(LL1_l, LL2_l)$, $(LH1_l, LH2_l)$, $(HL1_l, HL2_l)$, and $(HH1_l, HH2_l)$, where 1 and 2 refer to the two trees. The implementation of the DT-CWT requires that the functions Φ_1 and Ψ_1 operate on the odd numbered data samples and Φ_2 and Ψ_2 on even numbered data samples. Having a delay of one-half of a sample, these operations ensure that Φ_1 and Φ_2 as well as Ψ_1 and Ψ_2 form a Hilbert pair, and thus, the real and imaginary components of the complex coefficients are statistically uncorrelated. In order to improve the shift-invariance property, the DT-CWT avoids down-sampling operation in the first-level decomposition. Hence, the DT-CWT has much lower shift sensitivity than the DWT but a redundancy of 2^d, which is higher than that of the DWT but lower than that of the SWT by a factor of 2^l. Detailed analysis of estimating the 2D DT-CWT coefficients can be found in Selesnick et al. (2005); Kingsbury (1999, 2001); Selesnick (2001, 2002); Gopinath (2003); Chaux et al. (2006) and Tay et al. (2006).

[1]Two functions form a Hilbert pair if one of them is the Hilbert transform of the other.

2.4 Wavelet-Based Denoising: A Review

Owing to the sparseness[2] of the wavelet transform, only a few large detail coefficients in a subband contain information about the underlying image, while small values can be attributed to the noise that uniformly contaminates all subbands (Fadili and Boubchir, 2005). Wavelet-based denoising techniques modify each coefficient of the subband by using a thresholding or shrinkage function, which shrinks the coefficients in magnitude towards zero. Removing the small coefficients then removes most of the noise. In practice, the coefficients of subband LL_J are kept intact. This is because of the fact that the approximate coefficients usually represent important features of the image and any small perturbation yields a significant error.

From the early stages of wavelet-based denoising, a considerable effort has been made to design appropriate shrinkage or thresholding functions for the noisy wavelet coefficients. Some of the well known nonlinear shrinkage functions for estimating a noise-free wavelet coefficient x from its noisy observation y are:

- Hard-thresholding (Donoho and Johnstone, 1994): $\hat{x} = y \cdot \mathbf{1}(|y| > T)$
- Soft-thresholding (Donoho and Johnstone, 1994): $\hat{x} = \text{sgn}(y) \cdot \max(0, |y| - T)$
- Firm-shrinkage (Gao and Bruce, 1997):

$$\hat{x} = \begin{cases} 0 & \text{if } |y| \leq T_1 \\ \text{sgn}(y) \cdot \dfrac{T_2(|y| - T_1)}{T_2 - T_1} & \text{if } T_1 < |y| \leq T_2 \\ y & \text{if } |y| > T_2 \end{cases} \tag{2.10}$$

- Non-negative garrote-shrinkage (Gao, 1998):

$$\hat{x} = \text{sgn}(y) \cdot \max\left(0, 1 - \left(\frac{T}{y}\right)^2\right) \tag{2.11}$$

[2]Most of the coefficients are close to zero.

- Smoothly clipped absolute deviation thresholding for a given constant α (Antoniadis and Fan, 2001):

$$\hat{x} = \begin{cases} \text{sgn}(y) \cdot \max(0, |y| - T) & \text{if } |y| \leq 2T \\ \dfrac{(\alpha - 1)y - \alpha T \text{sgn}(y)}{\alpha - 2} & \text{if } 2T < |y| \leq \alpha T \\ y & \text{if } |y| > \alpha T \end{cases} \tag{2.12}$$

- Piecewise linear thresholding (Stefano et al., 2004):

$$\hat{x} = \begin{cases} \dfrac{T_2}{T_1} y & \text{if } |y| \leq T_1 \\ (y - \text{sgn}(y) \cdot T_1) \dfrac{T_3 - T_2}{T_3 - T_1} + T_2 \cdot \text{sgn}(y) & \text{if } T_1 < |y| \leq T_3 \\ y & \text{if } |y| > T_3 \end{cases} \tag{2.13}$$

- NeighCoeff thresholding (Cai and Silverman, 2001):

$$\hat{x} = y \cdot \max\left(0, 1 - \frac{2 \log N \sigma_v^2}{\sum_{\mathcal{S}_3(k)} y_k^2}\right) \tag{2.14}$$

- Amplitude-scale-invariant Bayes estimator (Figueiredo and Nowak, 2001):

$$\hat{x} = \frac{1}{y} \cdot \max\left(0, y^2 - 3\sigma_v^2\right) \tag{2.15}$$

where T, T_1, T_2, or T_3 is a threshold, σ_v^2 is the noise variance, $\mathbf{1}$ is a indicator function, $\mathcal{S}_3(k)$ is a 3×3 squared-shape local neighborhood centered at spatial location $k = (k_1, k_2)$, and N is the number of data samples in a subband. Various denoising methods have been developed by choosing an appropriate threshold for a given shrinkage function. Some of the successful thresholds are:

- Universal threshold in VisuShrink (Donoho, 1995) method: $T = \sigma_v \sqrt{2 \log(N)}$, where N is the total number of data samples.

- Threshold in translation invariant denoising (Coifman and Donoho, 1995) method: $T = \sigma_v \sqrt{2 \ln(N \log_2 N)}$

- Sure threshold in SureShrink (Donoho and Johnstone, 1995) and GarrotShrink (Gao, 1998) methods: $T = \arg\min_{0 \le T \le \sqrt{2\log N}} \left[N - 2 \cdot \#\{k : |\frac{y_k}{\sigma_v}| \le T\} + \sum_{k=1}^{N} \{\min(|\frac{y_k}{\sigma_v}|, T)\}^2 \right]$
- Threshold in BayesShrink (Chang et al., 2000a) method: $T = \sigma_v^2/\sigma_x$

where σ_x^2 is the variance of the noise-free coefficients. The choice of an appropriate threshold may depend on factors such as the type of image, the denoising efficiency and computational efficiency. For instance, when a fast denoising algorithm is required, the universal threshold may be used since it needs to be calculated only once for all the subbands. However, this threshold is too high for good denoising. Most signals in images show a spatially non-uniform energy distribution, which motivates the choice of a non-uniform threshold. Since a given noisy signal may consist of some parts where the magnitudes of the signal coefficients are below the globally defined threshold and other parts where the noise magnitudes exceed that given threshold, a subband-adaptive threshold (e.g. threshold in BayesShrink and SureShrink) or spatially-adaptive threshold (e.g. threshold in NeighCoef) is preferred for a better denoising performance. However, these thresholds require greater computation time because the threshold needs to be computed separately for each subband in the subband-adaptive methods, or for each coefficient in the spatially-adaptive methods.

The shrinkage function and threshold are also derived using a probabilistic model for the noise-free coefficients. Such a function is given by (Hyvarinen, 1999)

$$\hat{x} = \mathrm{sgn}(y) \cdot \max\left(0, |y| - \sigma_v^2 |\Omega(y)|\right) \tag{2.16}$$

where $\Omega(x) = -\frac{\mathrm{d}}{\mathrm{dx}}[\log(p_X(x))]$ is the score function of x and $p_X(x)$ is its probability density function (PDF). In the specific case, where $p_X(x)$ is defined by the Laplacian distribution, the shrinkage function is obtained as (Sendur and Selesnick, 2002a)

$$\hat{x} = \mathrm{sgn}(y) \cdot \max\left(0, |y| - \frac{\sqrt{2}\sigma_v^2}{\sigma_x}\right) \tag{2.17}$$

Recently, Bayesian estimators are also being used to estimate the noise-free wavelet coefficients from their noisy observations. Standard estimators are the MAP, MMSE, and minimum mean absolute error (MMAE) estimators, which are obtained using the posterior density function $p_{X|Y}(x|y)$ as follows (Papoulis, 1991).

- MAP-based estimation: The denoised coefficients are estimated as the conditional mode given by

$$\hat{x} = \arg\max_{x} p_{X|Y}(x|y) \tag{2.18}$$

- MMSE-based estimation: The denoised coefficients are estimated as the conditional mean given by

$$\hat{x} = \int x p_{X|Y}(x|y)\mathrm{d}x \tag{2.19}$$

- MMAE-based estimation: The denoised coefficients are estimated as the conditional median which is obtained from the relation

$$\int_{-\infty}^{\hat{x}} p_{X|Y}(x|y)\mathrm{d}x = \int_{\hat{x}}^{\infty} p_{X|Y}(x|y)\mathrm{d}x = \frac{1}{2} \tag{2.20}$$

In order to obtain the posterior density function, various prior functions have been assumed for the subband DWT coefficients as well as for local neighboring coefficients. Since sparsity of the DWT coefficients is significant, the histogram of the DWT coefficients of a subband is unimodal, heavy-tailed and sharply peaked in nature (Huang, 2000; Srivastava et al., 2003). Thus, heavy-tailed prior functions are commonly used for the subband DWT coefficients. Examples of such PDFs are Laplacian (Sendur and Selesnick, 2002a), generalized Gaussian (GG) (Mallat, 1989), Bessel K-form (Fadili and Boubchir, 2005), symmetric alpha-stable ($S\alpha S$) (Achim et al., 2003), symmetric normal inverse Gaussian (Solbo and Eltoft, 2004), scale mixtures of Gaussian (SMG) (Chipman et al., 1997; Abramovich et al., 1998; Clyde and George, 2000), and quasi-Cauchy (Johnstone and Silverman, 2005). The parameters of the PDF

can be estimated using different techniques, such as the method of moments (MoM), maximum likelihood (ML), MAP, and entropy-matching method (Aiazzi et al., 1999). However, the prior function for local neighboring DWT coefficients is very often chosen as Gaussian[3] (Mihçak et al., 1999; Kazubek, 2003; Cai and Silverman, 2001). In such a case, the MAP estimator for the noise-free coefficient is given by

$$\hat{x}_k = \frac{\sigma_{x_k}^2}{\sigma_{x_k}^2 + \sigma_v^2} \cdot y_k \tag{2.21}$$

where $\sigma_{x_k}^2$ and σ_v^2 are replaced by their estimates. To improve the performance of locally-adaptive estimators, parameters of the PDF are estimated using more sophisticated methods, such as MAP estimation, in which a prior function for the parameter is defined. Such an approach is found in Mihçak et al. (1999), where the prior of the variance parameter $\sigma_{x_k}^2$ is chosen as the exponential PDF. The MAP estimator of the variance in such a case is

$$\hat{\sigma}_{x_k}^2 = \max\left(0, \frac{M}{4\lambda}\left[-1 + \sqrt{1 + \frac{8\lambda}{M^2}\sum_{\mathcal{S}_w(k)} y_k^2}\right] - \sigma_v^2\right) \tag{2.22}$$

where $\mathcal{S}_w(k)$ is a $w \times w$ squared-shape local neighborhood centered at spatial location $k = (k_1, k_2)$, M is the number of data samples in $\mathcal{S}_w(k)$, and the hyperparameter λ is estimated as $\hat{\lambda} = 1/\sqrt{\mathrm{Var}(\hat{\sigma}_{x_k}^2)}\ \forall k$, in which $\hat{\sigma}_{x_k}^2$ is the ML estimate. When the prior function for the variance parameter is chosen as Jeffrey's non-informative prior, the MAP estimator is obtained as (Figueiredo and Nowak, 2001)

$$\hat{\sigma}_x^2 = \max\left(0, \frac{y^2}{3} - \sigma_v^2\right). \tag{2.23}$$

There exists a nonlinear dependency between the wavelet coefficients of a particular location in a given subband and that of the corresponding location of the coarser subband. Denoising algorithms have also been developed that take into consideration this inter-scale dependency by using hidden Markov tree (HMT) models (Romberg

[3]In the literature, the Gaussian PDF is also called normal PDF.

et al., 2001; Crouse et al., 1998) or an appropriate joint PDF (Sendur and Selesnick, 2002b; Achim and Kuruoğlu, 2005; Zhang et al., 2005a). An example of a bivariate PDF that considers the nonlinear inter-scale dependency (Sendur and Selesnick, 2002b) is

$$p_{XX_c}(x, x_c) = \frac{3}{2\pi\sigma_x^2} \cdot \exp\left(-\frac{\sqrt{3}}{\sigma_x}\sqrt{x^2 + x_c^2}\right), \quad -\infty < x, x_c < \infty \tag{2.24}$$

where x_c is the DWT coefficient of the coarser subband corresponding to the location of x. The MAP estimator using this PDF is

$$\hat{x} = \frac{y}{\sqrt{y^2 + y_c^2}} \cdot \max\left(0, \sqrt{y^2 + y_c^2} - \frac{\sqrt{3}\sigma_v^2}{\sigma_x}\right) \tag{2.25}$$

and the method is referred to as the BiShrink method (Sendur and Selesnick, 2002b).

2.5 Conclusion

In this chapter, an overview has been given of the steps involved in a cDNA microarray experiment. A short introduction to the 1D DWT has been presented along with its 2D extension. The relations of the SWT and CWT with the decimated DWT have also been discussed. A brief review of wavelet-based denoising has been presented. The estimation formulae of some of the existing denoising methods have also been given.

Chapter 3

DWT-Based Denoising Algorithms for cDNA Microarray Images

3.1 Introduction

Due to the non-ideal environment of the imaging technology and array construction, microarray images are contaminated by noise. Gene expression measurements obtained from such noisy images are inaccurate. Hence, the noise reduction step is unavoidable prior to any analysis of cDNA microarray images. This chapter begins with some background on noise models for cDNA microarray images and discusses the necessity for algorithms that remove additive Gaussian noise (AGN). A short introduction to traditional DWT-based denoising algorithms for the removal of AGN is given in Chapter 2. An important limitation of the existing algorithms is that they are not efficient for microarray image denoising applications. This is because these algorithms are only capable of processing the red and green channel images separately. But, there exists significant correlation between the DWT coefficients of the red and green channel images at the same spatial location (Howlader and Chaubey, 2009b). Moreover, the underlying processes that generate noise in the two images are similar, and hence, noise coefficients of the two channels are expected to be correlated as well. Thus, by processing the two images separately, conventional denoising algorithms ignore the inter-channel dependencies between the DWT coefficients of the

image as well as the noise. We argue that improved denoising performance could be achieved by exploiting these dependencies and incorporating information from both images during the estimation process.

This chapter uses LMMSE and MAP estimation techniques to develop new bivariate estimators (Howlader and Chaubey, 2009b) for the DWT coefficients of the red and green channel images of cDNA microarrays. Estimation is based on a joint statistical model for the image coefficients of the red and green channel images. The presence of noise correlation between the two channels is also considered in the estimation process. It is shown that both approaches lead to the same bivariate wavelet coefficient estimator. Unknown variance parameters in the estimator are estimated locally using the MAP criterion with different choices for the prior function. Certain denoising algorithms are obtained by choosing suitable prior functions. Extensive simulations are conducted to assess the efficiency of the proposed algorithms with respect to conventional DWT-based denoising methods. Since the purpose of a microarray image is the extraction of information regarding gene expression levels, a good microarray image denoising algorithm removes noise while preserving much of the signal information. Thus, results are presented not only for denoising of the pixels measured in the MSE sense, but also for accuracy of the log-intensity ratios estimated from the denoised images. It is shown that the proposed methods provide an improved noise reduction performance and yield log-intensity ratios that are close to the true values as compared to that of the existing denoising methods.

3.2 Noise Models

Microarray technology is a complex electrical-optical-chemical process involving multiple random factors. Consequently, there are many sources of error that show up as noise in the resulting images. Examples of noise in microarray images include pho-

ton noise, laser light reflection, electronic noise, quantization noise, dust on the glass slide, and so on (Lukac et al., 2004; Zhang et al., 2004; Mastrogianni et al., 2008). If the image is corrupted by additive noise, then the intensity of the noisy image at the spatial location (i, j) is given by

$$g(i,j) = f(i,j) + \varepsilon(i,j) \tag{3.1}$$

where $\varepsilon(i,j)$ is the noise at that spatial location. Noise that corrupts the image may be dependent or independent of the image. In the particular case, when $\varepsilon(i,j)$ is independent of the image and assumed to be i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$, the model in (3.1) is called the additive white Gaussian noise (AWGN) model. As an example, electronic noise in microarray images follows the AWGN model. Other noise models for microarray images include the multiplicative noise model (Rangayyan et al., 1998) as in the case of PMT noise (Zhang et al., 2004) and speckle, and the impulsive noise model (Lukac et al., 2004; Boncelet, 2000). Apart from the Gaussian distribution, the Poisson (Balagurunathan et al., 2002), uniform (Mastrogianni et al., 2008) (e.g., quantization noise) and exponential distributions (Davies and Seale, 2005) have also been used to describe the noise characteristics of microarray images. It is to be noted that many non-additive and non-Gaussian noise models for images can be mathematically remodelled as the additive noise model in (3.1) (Rangayyan et al., 1998). For instance, photon noise is non-additive, signal dependent and can be described by the Poisson process (Rangayyan et al., 1998)

$$p_{\tilde{g}|f}(\tilde{g}(i,j)|f(i,j),\varrho) = \frac{[\varrho f(i,j)]^{\tilde{g}(i,j)} \exp(-\varrho f(i,j))}{\tilde{g}(i,j)!} \tag{3.2}$$

where $\tilde{g}(i,j)$ is the observed image and ϱ is the rate of the Poisson process. The normalized image $g(i,j) = \tilde{g}(i,j)/\varrho$ can be described by the model in (3.1) by considering $\varepsilon(i,j) = g(i,j) - f(i,j)$ as the noise having zero mean and variance $\sigma^2_{\varepsilon(i,j)} = E\{\tilde{g}(i,j)/\varrho^2\}$ (Kuan et al., 1985; Rangayyan et al., 1998). Alternatively, the

Anscombe transform $\mathcal{T}(\tilde{g}(i,j)) = 2\sqrt{(\tilde{g}(i,j) + \frac{3}{8})}$ may be used to convert the image $\tilde{g}(i,j)$ that is corrupted with Poisson noise to the transformed image $g(i,j)$ corrupted by AWGN having $\sigma_\varepsilon^2 = 1$, under the assumption that $E\{\tilde{g}\}$ is large (Starck et al., 1998). Again, consider the multiplicative noise model that is given by

$$\tilde{g}(i,j) = f(i,j)\zeta(i,j) \tag{3.3}$$

where $\zeta(i,j)$ is a stationary noise uncorrelated with the image having mean μ_ζ and variance σ_ζ^2. This model can also be converted to the additive noise model in (3.1) by defining $g(i,j) = \tilde{g}(i,j)/\mu_\zeta$ and expressing $g(i,j)$ in terms of signal plus signal-dependent additive noise as follows

$$g(i,j) = f(i,j) + \left[\frac{\zeta(i,j) - \mu_\zeta}{\mu_\zeta}\right] f(i,j). \tag{3.4}$$

By comparing (3.4) with (3.1), we can identify the additive noise as $\varepsilon(i,j) = (\zeta(i,j)/\mu_\zeta - 1)f(i,j)$ having zero mean and index dependent noise variance (Kuan et al., 1985)

$$\sigma^2_{\varepsilon(i,j)} = \frac{{\sigma_\zeta}^2}{1 + {\sigma_\zeta}^2} \left(\sigma^2_{g(i,j)} + \mu^2_{g(i,j)}\right) \tag{3.5}$$

where $\mu_{g(i,j)}$ and $\sigma^2_{g(i,j)}$, respectively, are the mean and variance of the corrupted intensities at each index (i,j). Alternatively, the log-transformation can be applied to (3.3) to obtain an additive noise model. Since most of the noise types in microarray images can be either modelled or remodelled as additive noise, denoising of microarray images corrupted by additive noise stands as a classical problem in microarray image analysis. In this thesis, we focus on developing efficient algorithms for reducing AGN in cDNA microarray images.

3.3 Image Denoising Using DWT

The DWT is an important tool in the denoising of images corrupted by AGN due to its energy compaction and decorrelation (i.e., subbands are uncorrelated) properties

(Mallat, 1999; Sendur and Selesnick, 2002a). The decimated DWT is non-redundant and this feature make it attractive for fast and efficient image denoising. Moreover, a method which performs well in the decimated DWT domain is expected to perform well using other wavelet-like transforms (e.g., curvelet, ridgelet, contourlet, etc.). The DWT-based image denoising techniques for AGN can be found in Chang et al. (2000a,b); Moulin and Liu (1999); Simoncelli and Adelson (1996); Achim et al. (2003); Fadili and Boubchir (2005); Portilla et al. (2003); Strela et al. (2002); Portilla et al. (2001); Coifman and Donoho (1995); Sendur and Selesnick (2002a); Crouse et al. (1998); Wainwright and Simoncelli (2000); Choi et al. (2000); Mihçak et al. (1999); Kazubek (2003); Donoho (1995); Donoho and Johnstone (1995); Zhang et al. (2000); Cai et al. (2001); Crouse et al. (1999); Fan and Xia (2001).

There are two major approaches to statistical wavelet-based denoising. In the first approach, the wavelet coefficients are modified using certain threshold parameters and nonlinear shrinkage functions (Vidakovic, 1999; Antoniadis et al., 2001). Image denoising methods using such an approach include the VisuShrink (Donoho, 1995), SureShrink (Donoho and Johnstone, 1995), BayesShrink (Chang et al., 2000a), amplitude-scale-invariant Bayes estimator (Figueiredo and Nowak, 2001), NeighCoef (Cai and Silverman, 2001), Spatial-correlation thresholding (Bao and Ma, 2005), and empirical Bayes thresholding (EBayesThresh) (Johnstone and Silverman, 2005). Standard choices for nonlinear shrinkage functions are soft- and hard-thresholding (Donoho and Johnstone, 1994), firm-shrinkage(Gao and Bruce, 1997), and non-negative garrote shrinkage (Gao, 1998). Although computationally simple, arbitrary choice of nonlinearity in these methods leads to certain drawbacks. For instance, the soft-thresholding technique yields biased estimates with moderate variances whereas hard-thresholding yields less biased estimates having higher variances (Gao, 1998). The second and better approach is to design an estimator for the noise-free wavelet coeffi-

cients of images by minimizing a Bayesian risk, typically under MMSE , MMAE , or MAP criterion. For instance, Simoncelli and Adelson (1996) use MMSE estimation to derive the wavelet coefficient estimator, whereas Bhuiyan et al. (2007) and Sendur and Selesnick (2002a) use MMAE and MAP estimation, respectively.

Image denoising algorithms can be broadly classified as subband-adaptive (Pizurica and Philips, 2006) or locally/spatially-adaptive (Chang et al., 2000b; Jansen and Bultheel, 2001). For instance, the methods SureShrink, BayesShrink, and EBayesThresh are subband adaptive techniques, which calculate a different threshold for each subband under the assumption that wavelet coefficients in each subband are independent and identically distributed (i.i.d). However, DWT coefficients of a subband are spatially non-stationary and there exists strong intra-subband and weak interscale dependencies among the coefficients (Liu and Moulin, 2001). To account for intra-subband dependency, locally-adaptive techniques, such as the LAWMAP (Mihçak et al., 1999) and NeighCoef (Cai and Silverman, 2001), estimate a wavelet coefficient from its local neighboring region. The PDF for local neighboring DWT coefficients of natural images is very often chosen as zero-mean Gaussian (Mihçak et al., 1999; Voloshynovskiy et al., 2001). This is mainly due to the fact that the heavy-tailed PDFs used for the subband-adaptive methods approach the Gaussian PDF when the parameters are estimated from local neighboring coefficients. For example, the shape parameter of GG PDF and characteristic exponent of $S\alpha S$ PDF approach 2 when local neighboring coefficients of natural images are considered. The locally-adaptive denoising methods have two major advantages over the subband-adaptive ones. First, they yield better performance since they take into account the intra-subband dependency among the coefficients. Secondly, these methods are computationally efficient because unlike the subband coefficients which have heavy-tailed PDFs, the local neighboring coefficients have simpler PDFs such as the Gaussian. Therefore in most cases, the final estimator

has a closed-form expression and numerical calculation is not required. Further, the estimation of the parameters of the PDF of local neighboring coefficients whether it be by MoM or ML, is much simpler than in the case of a heavy-tailed PDF of the subband coefficients. Thus, in general, for image denoising a locally-adaptive estimator is preferable to the subband-adaptive one. Several methods have also taken into account the inter-scale dependency between the wavelet coefficients (Sendur and Selesnick, 2002a; Achim and Kuruoğlu, 2005). But, information theoretical analysis indicates that only slightly better denoising performance can be achieved by considering the inter-scale dependence in addition to the intra-subband one (Liu and Moulin, 2001). Moreover, such a denoising method is often computationally more intensive.

Existing DWT-based denoising algorithms may work well with the natural images. However, they are not very efficient for reducing noise in cDNA microarray images. An important characteristic of cDNA microarray images is that the DWT coefficients of the red and green channel images are significantly correlated at the same spatial location. Experiments are conducted on several microarray images to verify this important feature. Since, existing denoising algorithms can only be used to process the red and green channel images separately, these methods are incapable of considering the inter-channel dependencies that exist in the microarray images. In the next section, new denoising algorithms are designed for microarray images using joint estimation techniques that exploit the correlation between the coefficients of the two channels.

3.4 Estimation of DWT Coefficients

Let $f_r(i,j)$ and $f_g(i,j)$ be pixels of the red and green channel images, respectively, where $i = 1, \ldots, N_1$ and $j = 1, \ldots, N_2$. Then the noisy pixels may be represented as

$$g_r(i,j) = f_r(i,j) + \varepsilon_r(i,j) \tag{3.6}$$

$$g_g(i,j) = f_g(i,j) + \varepsilon_g(i,j) \tag{3.7}$$

where $\varepsilon_r(i,j)$ and $\varepsilon_g(i,j)$ are noise-samples at the reference location. It is assumed that noise-samples of the red and green channel images are correlated in the pixel domain, and are distributed as i.i.d. zero mean bivariate Gaussian having equal variance σ_ε^2 and correlation coefficient ρ_ε. The standard deviation σ_ε indicates the strength of noise and ρ_ε measures the amount of linear dependency of noise between channels. It is assumed that σ_ε and ρ_ε are known. However, if σ_ε is unknown, it may be estimated from the sample standard deviation of pixels in regions selected from the background of the red or green channel image. Similarly, ρ_ε can be estimated using the sample correlation coefficient by selecting corresponding regions from the backgrounds of the red and green channel images. It is necessary to use the background regions for such an estimation, since these regions contain mostly noise and little or no signal information.

Let $x_r(k_1,k_2)$ and $x_g(k_1,k_2)$ denote the DWT coefficients of the noise-free microarray images in the red and green channels, respectively, at spatial location (k_1,k_2) of a given subband. Since the DWT is a linear transform, the noisy coefficients of the images at that spatial location can be written as

$$y_r(k_1,k_2) = x_r(k_1,k_2) + v_r(k_1,k_2) \tag{3.8}$$

$$y_g(k_1,k_2) = x_g(k_1,k_2) + v_g(k_1,k_2) \tag{3.9}$$

where $v_r(k_1,k_2)$ and $v_g(k_1,k_2)$ are the noise coefficients of the red and green channels, respectively, with equal variance σ_v^2 and correlation coefficient ρ_v. If σ_v and ρ_v are

Table 3.1: Average of local inter-channel correlations of DWT coefficients in various subbands using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
HH_l	0.04	0.19	0.58	0.95
HL_l	0.25	0.67	0.88	0.96
LH_l	0.10	0.54	0.84	0.96

unknown, the DWT coefficients of the noisy image may also be used to estimate them. For example, σ_v may be estimated using the robust median estimator (Donoho and Johnstone, 1995) that is given by

$$\hat{\sigma}_v = \frac{\text{median}(|y_r|)}{0.6745}, \quad \forall\, y_r \in \text{subband } HH_1. \tag{3.10}$$

Since the DWT coefficients of images in a subband are spatially non-stationary (Mihçak et al., 1999; Jansen and Bultheel, 2001; Cai and Silverman, 2001), the random variables of the coefficients are index dependent. Let $x_r(k_1, k_2)$ and $x_g(k_1, k_2)$ be the samples of the random variables $X_r(k_1, k_2)$ and $X_g(k_1, k_2)$, respectively. Similarly, we define the random variables $Y_r(k_1, k_2)$ and $Y_g(k_1, k_2)$. On the other hand, the wavelet coefficients of noise are spatially stationary and, therefore, the corresponding random variables V_r and V_g are index independent. For notational convenience, the indices are suppressed unless stated otherwise.

To develop an efficient estimator for the DWT coefficients of microarray images, a priori knowledge about the random variables X_r and X_g is essential. It is known that the pixel intensities of spots at the same spatial location of the red and green channel images are strongly correlated with each other (Davies and Seale, 2005; Zhang et al., 2005b). Since the DWT is a linear transform, the corresponding transform coefficients of the two channels are expected to be highly correlated as well. We verify this experimentally by computing the correlation between local neighboring coefficients of two channels for 10 typical cDNA microarray images. Table 3.1 shows

Table 3.2: Average percentage of X_r and X_g that are significantly correlated in various subbands using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
HH_l	20	68	93	92
HL_l	33	75	78	87
LH_l	25	72	86	79

average values of these local correlations for a four-level DWT that are estimated using a 7×7 sliding window. It is to be noted that the correlations obtained using other window sizes, such as 3×3, 5×5, and 9×9 are also similar. From the table, it is evident that positive correlation exists between the local neighboring coefficients of the two images, particularly at higher levels of decomposition. Further, it is seen that the correlations for the subbands at any given level are higher than that of the corresponding lower levels. This is expected since wavelet coefficients in the higher levels contain more significant structural features than those in the lower levels (Mallat, 1999), for which, there exists more similarity among the former. To determine if the observed correlations are statistically significant, we perform the standard test which is based on the t-distribution (Goulden, 2007). The test is performed for the correlation between X_r and X_g at a given spatial location by using the wavelet coefficients within a 7×7 window centered at that location. The level of significance used for this test is 5 percent. Table 3.2 gives the average percentage of X_r and X_g in each subband that are significantly correlated wherein the percentages are obtained by averaging over the same set of images used for Table 3.1. It is seen from Table 3.2 that at least 65% coefficients within the subbands of the red and green channel images are significantly correlated with each other at a decomposition level $l \geq 2$. In addition, the percentage of significant correlations in the subbands of decomposition level $l = 1$ are non-negligible. The results obtained using other window

sizes are very similar to those given in Table 3.2. Thus, joint estimation is necessary for obtaining better denoising performance. It is to be mentioned that although the inter-channel correlations in the subbands of level $l = 1$ are lower than that of the higher levels, estimation of the noise-free coefficients of these subbands using a joint estimation technique cannot be expected to be worse than methods assuming zero correlation.

For joint estimation of the noise-free coefficients, a bivariate PDF for X_r and X_g is required that takes into account the inter-channel correlation. In the pixel domain, the distribution of spot intensities of the red or green channel image is approximately Gaussian (Balagurunathan et al., 2002; Davies and Seale, 2005). Moreover, the images follow the same statistical distribution in the pixel- and DWT-domains (Srivastava et al., 2003). Therefore, a good choice for modelling the noise-free coefficients of the red or green channel image at a given spatial location is the zero-mean Gaussian PDF with variance estimated from the local neighboring coefficients. Since the coefficients of the local neighborhoods of the two images are linearly dependent, the joint density function of X_r and X_g is chosen to be the zero-mean bivariate Gaussian PDF that is given by

$$p_{X_r X_g}(x_r, x_g) = \frac{1}{2\pi\sigma_r\sigma_g\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x_r^2}{\sigma_r^2} + \frac{x_g^2}{\sigma_g^2} - 2\rho\frac{x_r}{\sigma_r}\frac{x_g}{\sigma_g}\right\}\right], \quad -\infty < X_r, X_g < \infty \tag{3.11}$$

where $\{\sigma_r^2\ (\sigma_r^2 \geq 0),\ \sigma_g^2\ (\sigma_g^2 \geq 0)\}$ and $\rho\ (-1 \leqslant \rho \leqslant 1)$, are the variance and correlation parameters, respectively, which are estimated using the local neighboring coefficients. Here, the parameter ρ measures the linear dependency between the DWT coefficients of the two channels. Two methods are used to assess bivariate normality to evaluate the appropriateness of the proposed joint PDF for the index-dependent random variables X_r and X_g using a set of local neighboring coefficients centered at that index. First, we consider the well-known Mardia's test (Mardia, 1970, 1974)

Table 3.3: Average percentage of X_r and X_g in each subband that follow the bivariate Gaussian PDF by Mardia's test (Mardia, 1970) using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
HH_l	97.74	95.40	94.04	94.82
HL_l	97.34	94.57	92.73	89.20
LH_l	97.83	94.81	91.55	88.46

that is based on skewness and kurtosis of data, since this test is stable and reliable for assessing multivariate normality (Romeu and Ozturk, 1993; Kim and Timm, 2006). Table 3.3 presents the results concerning Mardia's test using a 7×7 window as the local neighboring region for each coefficient. Level of significance used for this test to reject the null hypothesis of bivariate normality is 5 percent. The values in this table represent the percentage of the coefficients in each subband that follow the bivariate Gaussian PDF by averaging over the same set of images used for Table 3.1. It can be seen that a high proportion of X_r and X_g follow the bivariate Gaussian PDF in all the subbands. Next, chi-square plots (Kim and Timm, 2006) are constructed as a graphical tool to examine bivariate normality of these random variables. Fig. 3.1 shows the average of the chi-square plots obtained for all the random variables X_r and X_g in the subbands HL_1, HL_2, HL_3, and HL_4 using the same settings that are used for Mardia's test. As can be seen from these plots, the squared generalized distances of the DWT coefficients that are measured from the local neighboring region and the chi-square quantiles are approximately equal. The chi-square plots for the other subbands are very similar to those in Fig. 3.1 and are therefore not shown. Since neither the chi-square plots nor the results of Mardia's test indicate significant departures from bivariate normality, we conclude that the joint PDF in (3.11) is an

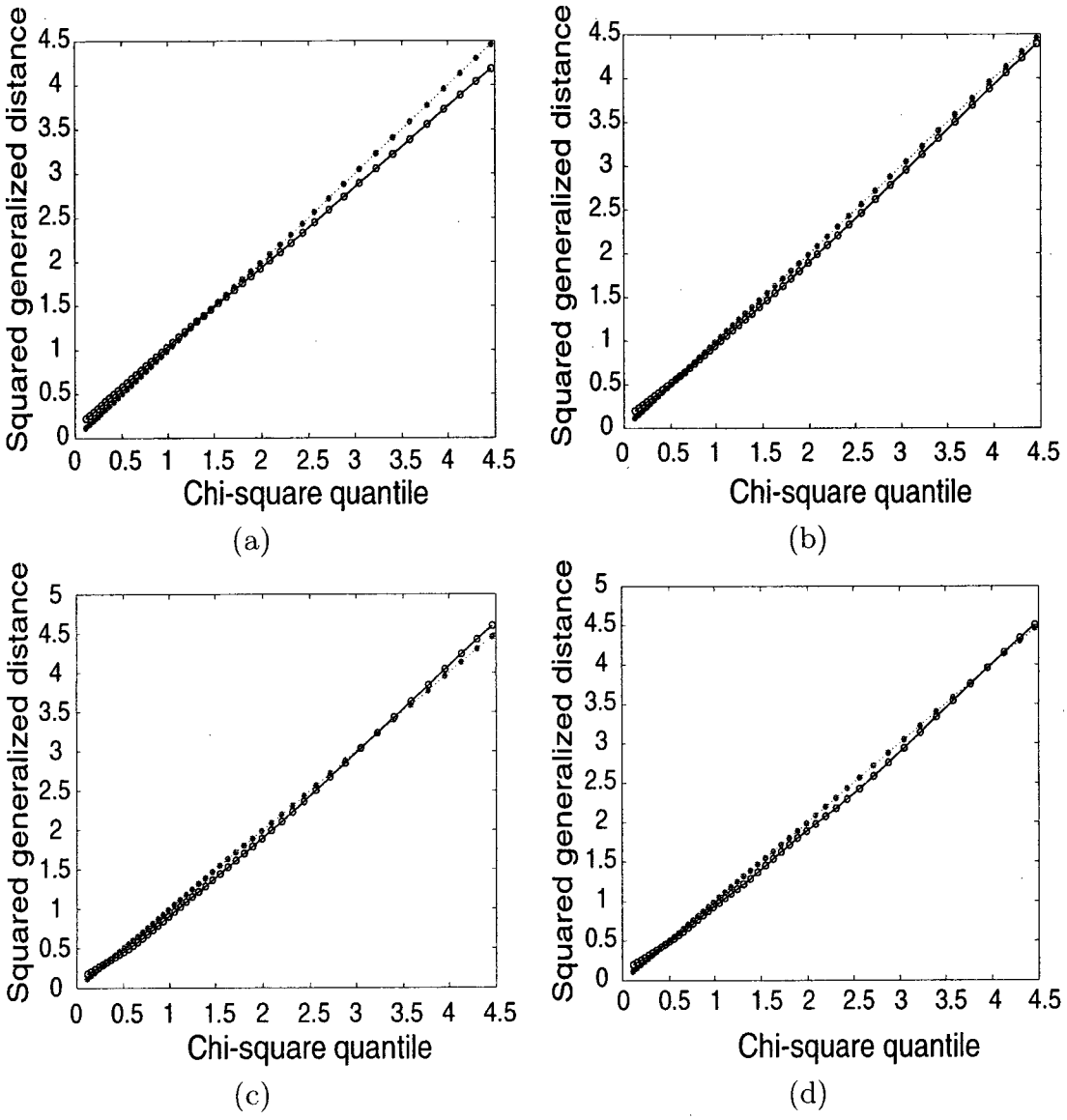


Figure 3.1: Chi-square plots to assess bivariate normality of X_r and X_g in the subbands (a) HL_1 (b) HL_2 (c) HL_3 and (d) HL_4.

appropriate model for the random variables X_r and X_g. As we shall see later, an advantage of being able to use the bivariate Gaussian PDF in the development of an estimator is that it yields a closed-form solution so that the resulting denoising algorithm is fast and efficiently implementable. In the following sections, we use this PDF in LMMSE and MAP estimation to derive bivariate estimators for X_r and X_g.

3.4.1 Bivariate LMMSE Estimator

Let the additive noise model in (3.8) and (3.9) be represented in vector notation as

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v} \tag{3.12}$$

where $\boldsymbol{x} = [x_r, x_g]^T$, $\boldsymbol{y} = [y_r, y_g]^T$, and $\boldsymbol{v} = [v_r, v_g]^T$ are samples of the random vectors $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{V}$, respectively. Then the LMMSE estimator for $\boldsymbol{x}$ given the corrupted observation $\boldsymbol{y}$ can be written as (Kamen and Su, 1999)

$$\hat{\boldsymbol{x}} = \mu_{\boldsymbol{X}} + \Sigma_{\boldsymbol{XY}}\Sigma_{\boldsymbol{Y}}^{-1}(\boldsymbol{y} - \mu_{\boldsymbol{Y}}) \tag{3.13}$$

where $\mu_{\boldsymbol{X}}$ and $\mu_{\boldsymbol{Y}}$ are the mean vectors of the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\Sigma_{\boldsymbol{XY}}$ is the cross-covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Y}$, and $\Sigma_{\boldsymbol{Y}}$ is the covariance matrix of $\boldsymbol{Y}$. The matrices $\Sigma_{\boldsymbol{XY}}$ and $\Sigma_{\boldsymbol{Y}}$ may be written as

$$\Sigma_{\boldsymbol{XY}} = E\{(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{Y} - \mu_{\boldsymbol{Y}})^T\} \tag{3.14}$$

and

$$\Sigma_{\boldsymbol{Y}} = E\{(\boldsymbol{Y} - \mu_{\boldsymbol{Y}})(\boldsymbol{Y} - \mu_{\boldsymbol{Y}})^T\} \tag{3.15}$$

where $E\{\cdot\}$ is the mathematical expectation. Since the DWT coefficients of the image and additive noise have zero means (Mallat, 1999), $\mu_{\boldsymbol{X}} = \mu_{\boldsymbol{V}} = 0$, and $\mu_{\boldsymbol{Y}} = \mu_{\boldsymbol{X}} + \mu_{\boldsymbol{V}} = 0$, where $\mu_{\boldsymbol{V}}$ is the mean vector of $\boldsymbol{V}$. Further, the image and additive

noise are independent so that

$$\begin{aligned}\Sigma_{\boldsymbol{XY}} =& E\{\boldsymbol{XY}^{\boldsymbol{T}}\} \\ =& E\{\boldsymbol{X}(\boldsymbol{X}+\boldsymbol{V})^T\} \\ =& E\{\boldsymbol{XX}^T\} + E\{\boldsymbol{XV}^T\} \\ =& E\{\boldsymbol{XX}^T\} + \mu_{\boldsymbol{X}}\mu_{\boldsymbol{V}}^T \\ =& \Sigma_{\boldsymbol{X}} \end{aligned} \tag{3.16}$$

which is the covariance matrix of the bivariate Gaussian random variable $\boldsymbol{X}$ given by

$$\Sigma_{\boldsymbol{X}} = \begin{bmatrix} \sigma_r^2 & \rho\sigma_r\sigma_g \\ \rho\sigma_r\sigma_g & \sigma_g^2 \end{bmatrix}. \tag{3.17}$$

The covariance matrix of $\boldsymbol{Y}$ is found to be $\Sigma_{\boldsymbol{Y}} = \Sigma_{\boldsymbol{X}} + \Sigma_{\boldsymbol{V}}$, where $\Sigma_{\boldsymbol{V}}$ is the covariance matrix of the joint PDF of V_r and V_g. Thus, the LMMSE estimator of $\boldsymbol{x}$ given in (3.13) becomes

$$\hat{\boldsymbol{x}} = \Sigma_{\boldsymbol{X}}(\Sigma_{\boldsymbol{X}} + \Sigma_{\boldsymbol{V}})^{-1}\boldsymbol{y} \tag{3.18}$$

To find $\hat{\boldsymbol{x}}$, the elements of $\Sigma_{\boldsymbol{V}}$ need to be specified. The noise model in (3.6) and (3.7) assumes that the microarray images are corrupted with additive bivariate Gaussian noise where σ_ε^2 is the noise strength in each channel and ρ_ε is the inter-channel noise correlation. It is known that the DWT coefficients of Gaussian noise in an image follow the i.i.d. zero-mean Gaussian distribution (Mallat, 1999). Then the DWT coefficients of additive bivariate Gaussian noise in microarray images follow the i.i.d. zero-mean bivariate Gaussian PDF with equal variance $\sigma_v^2 = \frac{1}{N}\mathrm{tr}(\mathcal{W}\mathcal{W}^T)^l\sigma_\varepsilon^2$ and correlation parameter $\rho_v = \rho_\varepsilon$, where $\mathcal{W} \in \{\mathcal{W}^O; O \in H, V, D\}$ denotes the transform matrix of the DWT and $\mathrm{tr}\{\cdot\}$ the trace of a matrix. The relation $\rho_v = \rho_\varepsilon$ holds since the covariance of $v_r(k_1, k_2)$ and $v_g(k_1, k_2)$ is given by $\sigma_{v_r v_g} = \frac{1}{N}\mathrm{tr}\{\mathcal{W}\mathcal{W}^T\}\sigma_{\varepsilon_r\varepsilon_g}$. For orthogonal wavelet functions, $\frac{1}{N}\mathrm{tr}\{\mathcal{W}\mathcal{W}^T\} = 1$ and hence in this case, $\sigma_v^2 = \sigma_\varepsilon^2$. On

the other hand, $\rho_v = \rho_\varepsilon$, whether the wavelet function be orthogonal, biorthogonal, or spline. The covariance matrix of $\boldsymbol{V}$ may be written as

$$\Sigma_{\boldsymbol{V}} = \sigma_v^2 \begin{bmatrix} 1 & \rho_v \\ \rho_v & 1 \end{bmatrix}. \tag{3.19}$$

Using the expressions of $\Sigma_{\boldsymbol{X}}$ and $\Sigma_{\boldsymbol{V}}$ in (3.18) and performing the matrix operations, the DWT coefficients of red and green channel images can be obtained as

$$\hat{x}_r \Big| [y_r, y_g] = \frac{\sigma_r^2\sigma_1^2 - \rho\rho_v\sigma_r\sigma_g\sigma_v^2}{\sigma_r^2\sigma_1^2 + \sigma_v^2(\sigma_g^2 + \sigma_v^2) - \sigma_v^2\rho_v(\sigma_v^2\rho_v + 2\rho\sigma_r\sigma_g)} \left[y_r + \frac{\sigma_v^2(\rho\sigma_g - \rho_v\sigma_r)}{\sigma_r\sigma_1^2 - \rho\rho_v\sigma_g\sigma_v^2} y_g \right] \tag{3.20}$$

$$\hat{x}_g \Big| [y_r, y_g] = \frac{\sigma_g^2\sigma_2^2 - \rho\rho_v\sigma_r\sigma_g\sigma_v^2}{\sigma_g^2\sigma_2^2 + \sigma_v^2(\sigma_r^2 + \sigma_v^2) - \sigma_v^2\rho_v(\sigma_v^2\rho_v + 2\rho\sigma_r\sigma_g)} \left[y_g + \frac{\sigma_v^2(\rho\sigma_r - \rho_v\sigma_g)}{\sigma_g\sigma_2^2 - \rho\rho_v\sigma_r\sigma_v^2} y_r \right] \tag{3.21}$$

where $\sigma_1^2 = [1 - \rho^2]\sigma_g^2 + \sigma_v^2$ and $\sigma_2^2 = [1 - \rho^2]\sigma_r^2 + \sigma_v^2$.

3.4.2 Bivariate MAP Estimator

The MAP estimator for $\boldsymbol{x}$ given the corrupted observation $\boldsymbol{y}$ is (Kamen and Su, 1999)

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{x}} p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \tag{3.22}$$

Using the Bayes rule, (3.22) can be written in terms of the PDFs of noise and the noise-free wavelet coefficients as

$$\begin{aligned} \hat{\boldsymbol{x}}(\boldsymbol{y}) &= \arg\max_{\boldsymbol{x}} \left[p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y} \mid \boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) \right] \\ &= \arg\max_{\boldsymbol{x}} \left[p_{\boldsymbol{V}}(\boldsymbol{y} - \boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}) \right] \\ &= \arg\max_{\boldsymbol{x}} \left[\ln(p_{\boldsymbol{V}}(\boldsymbol{y} - \boldsymbol{x})) + \ln(p_{\boldsymbol{X}}(\boldsymbol{x})) \right] \end{aligned} \tag{3.23}$$

where $p_{\boldsymbol{X}}(\cdot)$ is the joint PDF given in (3.11) and $p_{\boldsymbol{V}}(\cdot)$ is the zero-mean bivariate Gaussian PDF given by

$$p_{\boldsymbol{V}}(\boldsymbol{v}) = \frac{1}{2\pi\sigma_v^2\sqrt{(1 - \rho_v^2)}} \cdot \exp\left[-\frac{1}{2\sigma_v^2(1 - \rho_v^2)} \left\{ v_r^2 + v_g^2 - 2\rho_v v_r v_g \right\} \right] \tag{3.24}$$

By using (3.24), (3.23) becomes

$$[\hat{x}_r, \hat{x}_g] = \arg\max_{x_r, x_g} \left[-\frac{1}{2\sigma_v^2(1-\rho_v^2)} \left\{ (y_r - x_r)^2 + (y_g - x_g)^2 - 2\rho_v(y_r - x_r)(y_g - x_g) \right\} + \ln\left(p_{X_r X_g}(x_r, x_g)\right) \right] \tag{3.25}$$

To maximize (3.25), we solve the equations

$$\frac{1}{(1-\rho_v^2)\sigma_v^2}\left[y_r - x_r - \rho_v(y_g - x_g)\right] - \frac{1}{(1-\rho^2)}\left[\frac{x_r}{\sigma_r^2} - \frac{\rho}{\sigma_r}\frac{x_g}{\sigma_g}\right] = 0 \tag{3.26}$$

$$\frac{1}{(1-\rho_v^2)\sigma_v^2}\left[y_g - x_g - \rho_v(y_r - x_r)\right] - \frac{1}{(1-\rho^2)}\left[\frac{x_g}{\sigma_g^2} - \frac{\rho}{\sigma_g}\frac{x_r}{\sigma_r}\right] = 0 \tag{3.27}$$

to obtain the following MAP estimators for x_r and x_g (Howlader and Chaubey, 2009b):

$$\hat{x}_r\Big|[y_r, y_g] = \frac{\sigma_r^2\sigma_1^2 - \rho\rho_v\sigma_r\sigma_g\sigma_v^2}{\sigma_r^2\sigma_1^2 + \sigma_v^2(\sigma_g^2 + \sigma_v^2) - \sigma_v^2\rho_v(\sigma_v^2\rho_v + 2\rho\sigma_r\sigma_g)}\left[y_r + \frac{\sigma_v^2(\rho\sigma_g - \rho_v\sigma_r)}{\sigma_r\sigma_1^2 - \rho\rho_v\sigma_g\sigma_v^2}y_g\right] \tag{3.28}$$

$$\hat{x}_g\Big|[y_r, y_g] = \frac{\sigma_g^2\sigma_2^2 - \rho\rho_v\sigma_r\sigma_g\sigma_v^2}{\sigma_g^2\sigma_2^2 + \sigma_v^2(\sigma_r^2 + \sigma_v^2) - \sigma_v^2\rho_v(\sigma_v^2\rho_v + 2\rho\sigma_r\sigma_g)}\left[y_g + \frac{\sigma_v^2(\rho\sigma_r - \rho_v\sigma_g)}{\sigma_g\sigma_2^2 - \rho\rho_v\sigma_r\sigma_v^2}y_r\right] \tag{3.29}$$

where $\sigma_1^2 = [1-\rho^2]\sigma_g^2 + \sigma_v^2$ and $\sigma_2^2 = [1-\rho^2]\sigma_r^2 + \sigma_v^2$. Thus, LMMSE and MAP estimation lead to the same bivariate estimator. In the special case where noise coefficients in the two channels are uncorrelated (i.e., $\rho_v = 0$), the above estimator reduces to:

$$\hat{x}_r\Big|[y_r, y_g] = \frac{\sigma_r^2\sigma_1^2}{\sigma_r^2\sigma_1^2 + \sigma_v^2(\sigma_g^2 + \sigma_v^2)}\left[y_r + \rho\frac{\sigma_v^2}{\sigma_1^2}\frac{\sigma_g}{\sigma_r}y_g\right] \tag{3.30}$$

$$\hat{x}_g\Big|[y_r, y_g] = \frac{\sigma_g^2\sigma_2^2}{\sigma_g^2\sigma_2^2 + \sigma_v^2(\sigma_r^2 + \sigma_v^2)}\left[y_g + \rho\frac{\sigma_v^2}{\sigma_2^2}\frac{\sigma_r}{\sigma_g}y_r\right]. \tag{3.31}$$

3.4.3 Parameter Estimation

The estimators given in (3.28), (3.29) and (3.30), (3.31) require prior knowledge of the parameters σ_r^2, σ_g^2, and ρ. Since image coefficients in a given subband of the red and

green channel images are spatially non-stationary, these parameters depend on the spatial index (k_1, k_2). For notational simplicity, we shall denote the two-dimensional spatial index (k_1, k_2) by k in the remainder of this chapter. A locally-adaptive method is used to estimate these parameters at the k-th index using the noisy coefficients in a neighboring region $\mathcal{S}(k)$. The local neighborhood $\mathcal{S}(k)$ is defined as all the coefficients within a square-shaped window that is centered at $[y_r(k), y_g(k)]$. Although wavelet coefficients of an image are not i.i.d. within a subband, for the purpose of parameter estimation, it is assumed that coefficients within a local neighborhood are i.i.d. Thus, we assume that the joint PDF of $Y_r(k)$ and $Y_g(k)$ is an i.i.d. zero-mean bivariate Gaussian PDF with parameters $\sigma^2_{y_r}(k)$, $\sigma^2_{y_g}(k)$ and $\rho_y(k)$, which are estimated from the coefficients in $\mathcal{S}(k)$. In other words, $\sigma^2_{y_r}(m) = \sigma^2_{y_r}(k)$, $\sigma^2_{y_g}(m) = \sigma^2_{y_g}(k)$, and $\rho_y(m) = \rho_y(k)$ for all indexes $m \in \mathcal{S}(k)$. We now describe how the parameters at each index may be estimated using ML and MAP estimation.

ML Method

First, the ML method is used to estimate the parameters for each coefficient. We compute the ML estimates of the index-dependent parameters using a sliding window approach and then substitute these values in (3.28) and (3.29) to obtain the MAP estimates for x_r and x_g, respectively. Let $\sigma^2_{y_r}$ and $\sigma^2_{y_g}$ be the variances for $Y_r(m)$ and $Y_g(m)$, respectively, and ρ_y be the correlation between them conditioned on $\mathcal{S}(k)$. The log-likelihood for the noisy coefficients is given by

$$\begin{aligned}\ell(k) &= \sum_{m\in\mathcal{S}(k)} \ln p_{Y_rY_g}\left(y_r(m), y_g(m) \mid \sigma^2_{y_r}, \sigma^2_{y_g}, \rho_y\right) \\ &= -M \ln 2\pi\sigma_{y_r}\sigma_{y_g}\sqrt{1-\rho_y^2} - \frac{1}{2(1-\rho_y^2)}\Bigg[\frac{\sum_m y_r^2(m)}{\sigma^2_{y_r}} + \frac{\sum_m y_g^2(m)}{\sigma^2_{y_g}} \\ &\qquad - 2\rho_y \sum_m \frac{y_r(m)}{\sigma_{y_r}}\frac{y_g(m)}{\sigma_{y_g}}\Bigg] \qquad (3.32)\end{aligned}$$

where $p_{Y_r Y_g}(y_r, y_g \mid \sigma_{y_r}^2, \sigma_{y_g}^2, \rho_y)$ is the zero-mean bivariate Gaussian density with parameters $\sigma_{y_r}^2$, $\sigma_{y_g}^2$, and ρ_y, and M is the total number of coefficients in the neighborhood $\mathcal{S}(k)$. The following ML estimates are obtained by differentiating $\ell(k)$ with respect to the parameters σ_{y_r}, σ_{y_g}, and ρ_y, setting the results to zero, and then solving the likelihood equations.

$$\hat{\sigma}_{y_r}^2(k) = \arg\max_{\sigma_{y_r}^2 \geq 0} \ell(k) = \frac{1}{M} \sum_{m \in \mathcal{S}(k)} y_r^2(m) \tag{3.33}$$

$$\hat{\sigma}_{y_g}^2(k) = \arg\max_{\sigma_{y_g}^2 \geq 0} \ell(k) = \frac{1}{M} \sum_{m \in \mathcal{S}(k)} y_g^2(m) \tag{3.34}$$

$$\hat{\rho}_y(k) = \arg\max_{-1 \leq \rho_y \leq 1} \ell(k) = \frac{\sum_{m \in \mathcal{S}(k)} y_r(m) y_g(m)}{\sqrt{\sum_{m \in \mathcal{S}(k)} y_r(m)^2 \sum_{m \in \mathcal{S}(k)} y_g(m)^2}} \tag{3.35}$$

We then obtain the ML estimates for parameters of the noise-free coefficients as

$$\hat{\sigma}_r^2(k) = \max\left(\frac{1}{M} \sum_{m \in \mathcal{S}(k)} y_r^2(m) - \sigma_v^2, 0 \right) \tag{3.36}$$

$$\hat{\sigma}_g^2(k) = \max\left(\frac{1}{M} \sum_{m \in \mathcal{S}(k)} y_g^2(m) - \sigma_v^2, 0 \right) \tag{3.37}$$

$$\hat{\rho}(k) = \max\left(\min\left(\frac{1}{\hat{\sigma}_r(k)\hat{\sigma}_g(k)} \frac{1}{M} \sum_{m \in \mathcal{S}(k)} y_r(m) y_g(m) - \sigma_v^2 \rho_v, 1 \right), -1 \right) \tag{3.38}$$

These values are substituted in (3.28) and (3.29) to obtain the MAP (or LMMSE) estimates for the noise-free wavelet coefficients of two channels.

MAP Method

The MAP estimation technique yields better estimates than ML estimation provided that correct prior functions for the unknown parameters can be defined. In this section, we investigate the choice of appropriate prior functions for the parameters $\sigma_{y_r}^2$, $\sigma_{y_g}^2$, and ρ_y, and later use these prior functions to derive the MAP estimators. In the Bayesian literature, the inverse-gamma (IG) PDF is often used as the prior density for the variance parameter (Congdon. 2003). More specifically, it is used as a

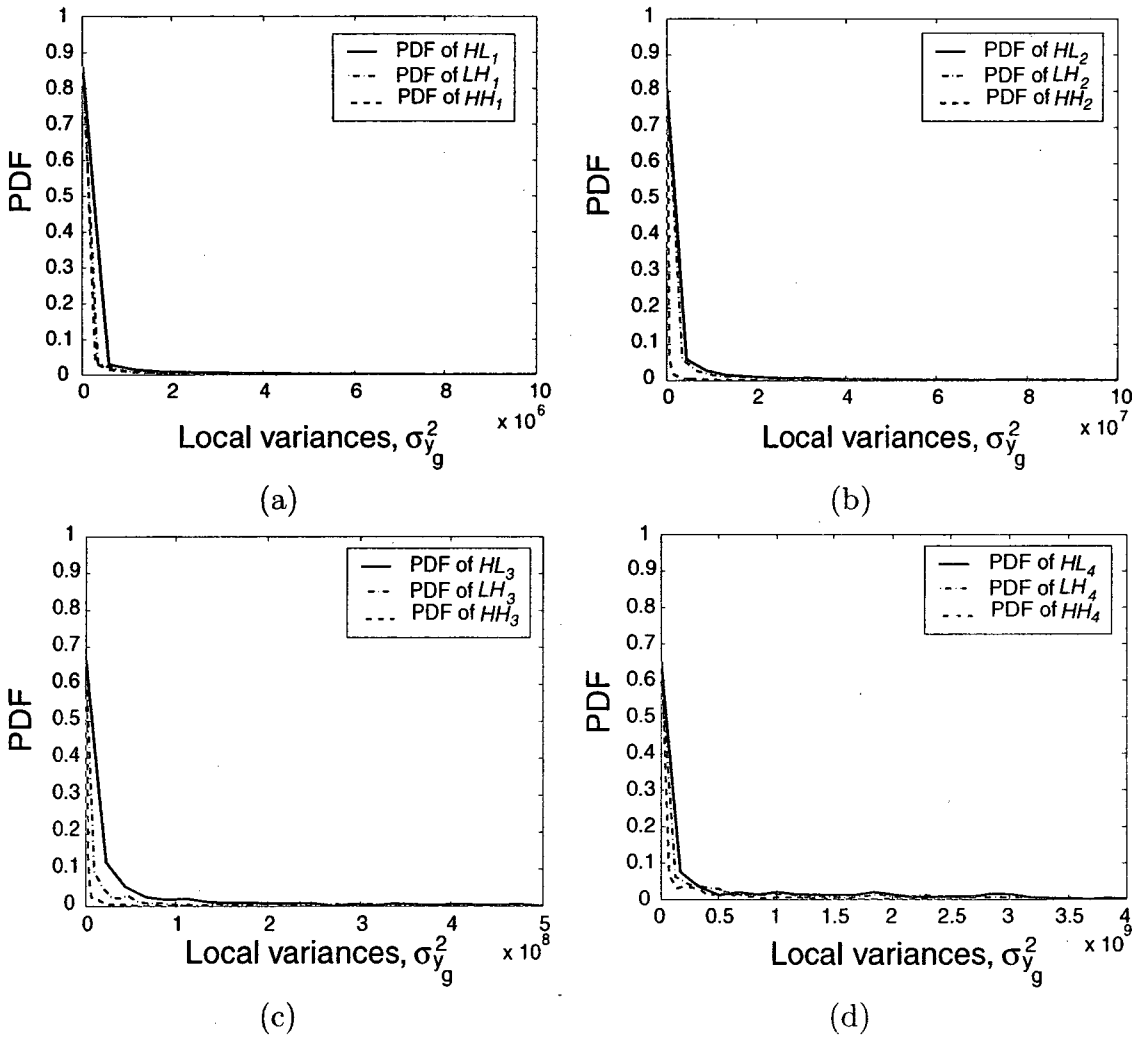


Figure 3.2: Empirical PDFs of local variances of the wavelet coefficients in the subbands of (a) first level (b) second level (c) third level and (d) fourth level decomposition of a typical green channel image.

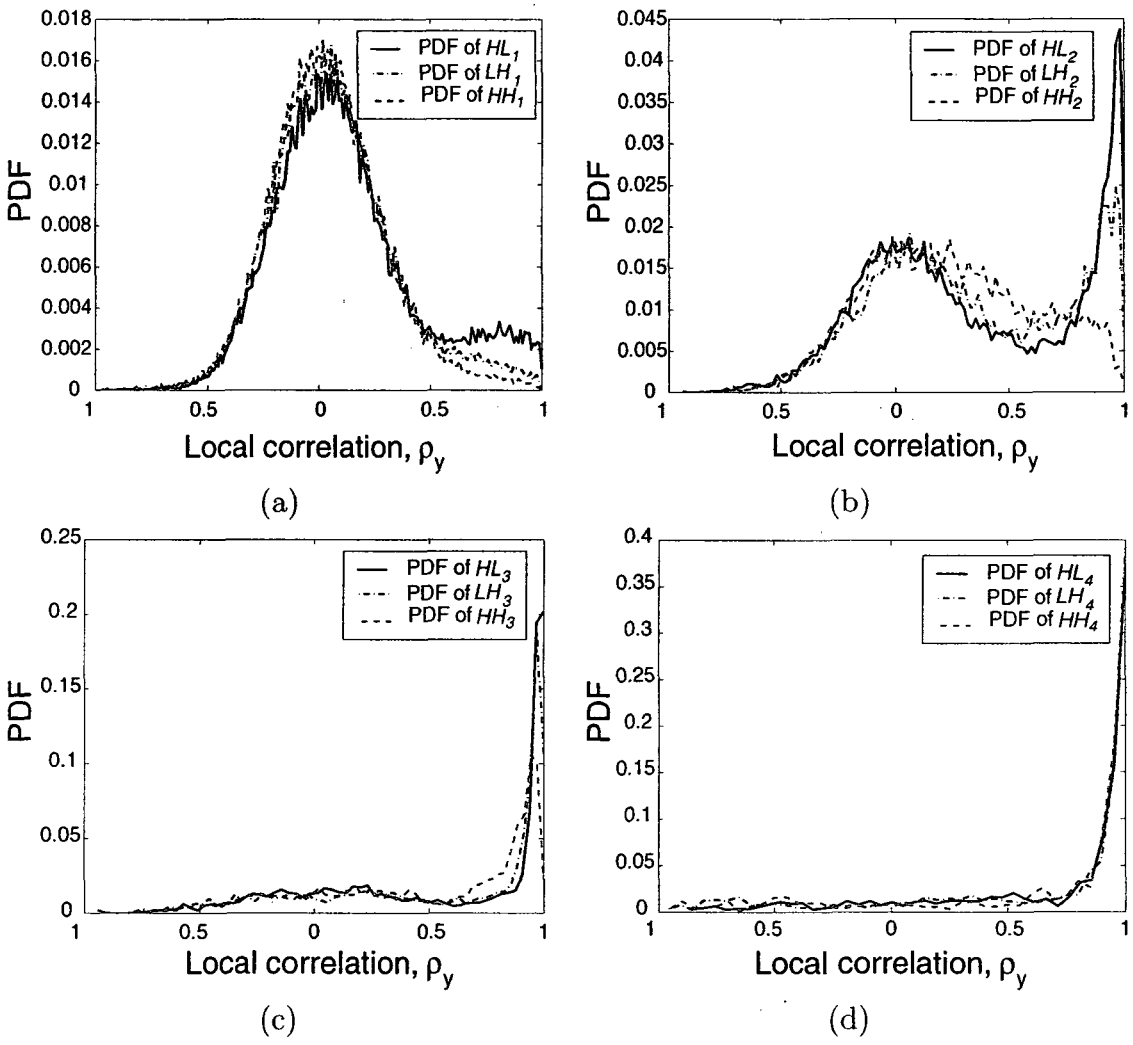


Figure 3.3: Empirical PDFs of local correlation of the wavelet coefficients in the subbands of (a) first level (b) second level (c) third level and (d) fourth level decomposition of red and green channel images.

conjugate prior for the variance in a univariate normal model. Hence, first we choose the IG prior to model the local[1] variances of the wavelet coefficients in the red and green channel images. However, the exponential prior has been shown to be a reasonable model for the local variances of the DWT coefficients for natural images (Mihçak et al., 1999). Fig. 3.2 shows empirical PDFs of the local variances of DWT coefficients in the subbands of the first four levels of decomposition for a typical green channel microarray image. The variances are estimated using a sliding window of size 5×5. Empirical PDFs are also obtained for subbands of the red channel image using different window sizes, but these are similar to Fig. 3.2 and are therefore not shown. From Fig. 3.2 it is seen that the empirical PDFs for $\sigma_{y_g}^2$ obtained in each subband at each level of decomposition are similar in shape. They are roughly convex, bounded at zero and highly skewed to the right. The data therefore suggest that the exponential prior may be an appropriate model for the local variances of the DWT coefficients of cDNA microarray images as well. However, the exponential PDF is a special case of the gamma PDF, which contains both a shape parameter as well as a scale parameter. Intuitively, one would expect the two-parameter gamma PDF to provide a better fit to the data. We fitted the gamma PDF to the local variances of the red and green channel images. For most subbands, the estimated value of the subband-adaptive shape parameter was found to be close to one. Hence, the exponential PDF was chosen as the prior function for modelling $\sigma_{y_g}^2$ and $\sigma_{y_r}^2$.

Although the empirical PDFs for $\sigma_{y_g}^2$ show a consistency in shape among the various subbands, the empirical PDFs of ρ_y vary substantially according to level of decomposition. This can be seen from Fig. 3.3, which shows the empirical PDFs of ρ_y in the subbands of each level of decomposition for a typical two-channel cDNA microarray image. The empirical PDFs in the subbands of the first level appear to be approximately Gaussian and are centered at $\rho_y = 0$. However, the empirical PDFs for

[1]The term local implies that the parameter is estimated using the local neighboring coefficients.

the higher-level subbands are clearly non-Gaussian. For instance, the empirical PDFs for the second and third-level subbands are multi-modal in nature. Moreover, as the level of decomposition increases, the peakedness of the subband PDFs increases near the boundary at $\rho_y = 1$. Thus, there is no consistency in shape of the empirical prior functions of ρ_y in the various levels. Level-specific prior functions might be considered such as a Gaussian PDF for the first level subbands, SMG for the second and third level subbands, and a negatively skewed PDF for the subbands in the fourth level of decomposition. However, such an approach increases the computational complexity of the proposed algorithm significantly. Therefore, we prefer to estimate ρ_y using the ML method as described in the previous subsection. We now describe the MAP estimation method for obtaining estimates of the local variances $\sigma^2_{y_r}$ and $\sigma^2_{y_g}$ using (i) the IG prior, and (ii) the exponential prior.

(i) Inverse-gamma prior

In this section, we derive MAP estimates for the index-dependent parameters $\sigma^2_{y_r}$ and $\sigma^2_{y_g}$ using the following IG priors

$$h_1(\sigma^2_{y_r}) = \frac{{b_1}^{a_1}}{{\sigma^2_{y_r}}^{a_1+1}\Gamma(a_1)} \exp(-b_1/\sigma^2_{y_r}), \quad \sigma^2_{y_r} \geq 0 \tag{3.39}$$

$$h_2(\sigma^2_{y_g}) = \frac{{b_2}^{a_2}}{{\sigma^2_{y_g}}^{a_2+1}\Gamma(a_2)} \exp(-b_2/\sigma^2_{y_g}), \quad \sigma^2_{y_g} \geq 0 \tag{3.40}$$

where a_1, b_1 and a_2, b_2 are the subband-adaptive hyperparameters. Then the 'joint prior' for $\sigma^2_{y_r}$ and $\sigma^2_{y_g}$ is considered to be a product of the marginal priors in (3.39) and (3.40). Using the above prior functions, the log-likelihood function for the noisy coefficients in $\mathcal{S}(k)$ can be written as

$$\ell(k) = \sum_{m \in \mathcal{S}(k)} \ln p_{Y_r Y_g}\left(y_r(m), y_g(m) \mid \sigma^2_{y_r}, \sigma^2_{y_g}, \rho_y\right) + \ln h_1(\sigma^2_{y_r}) + \ln h_2(\sigma^2_{y_g}) \tag{3.41}$$

Maximization of $\ell(k)$ with respect to σ_{y_r} and σ_{y_g} gives the following likelihood equations.

$$\beta_1\sigma_{y_r}^2 + \frac{\rho_y}{1-\rho_y^2}\frac{\sum_m y_r(m)y_g(m)}{\sigma_{y_g}}\sigma_{y_r} - \frac{1}{1-\rho_y^2}\sum_m y_r^2(m) - 2b_1 = 0 \tag{3.42}$$

$$\beta_2\sigma_{y_g}^2 + \frac{\rho_y}{1-\rho_y^2}\frac{\sum_m y_r(m)y_g(m)}{\sigma_{y_r}}\sigma_{y_g} - \frac{1}{1-\rho_y^2}\sum_m y_g^2(m) - 2b_2 = 0 \tag{3.43}$$

where $\beta_1 = M + 2a_1 + 2$ and $\beta_2 = M + 2a_2 + 2$. Solving (3.42) and (3.43) for σ_{y_r} and σ_{y_g}, and using the relations $\sigma_{y_r}^2 = \sigma_r^2 + \sigma_v^2$ and $\sigma_{y_g}^2 = \sigma_g^2 + \sigma_v^2$, we obtain the following MAP estimators for σ_r^2 and σ_g^2:

$$\hat{\sigma}_r^2(k) = \max\left(\frac{1}{4\beta_1^2}\left[-\frac{\rho_y(k)\Psi_3(k)}{\sigma_{y_g}} + \sqrt{\frac{\rho_y^2(k)\Psi_3(k)^2}{\sigma_{y_g}^2} + 4\beta_1\Big(\Psi_1(k) + 2b_1\Big)}\right]^2 - \sigma_v^2, 0\right) \tag{3.44}$$

$$\hat{\sigma}_g^2(k) = \max\left(\frac{1}{4\beta_2^2}\left[-\frac{\rho_y(k)\Psi_3(k)}{\sigma_{y_r}} + \sqrt{\frac{\rho_y^2(k)\Psi_3(k)^2}{\sigma_{y_r}^2} + 4\beta_2\Big(\Psi_2(k) + 2b_2\Big)}\right]^2 - \sigma_v^2, 0\right) \tag{3.45}$$

In the above expressions,

$$\Psi_1(k) = \frac{1}{1-\rho_y^2(k)}\sum_m y_r^2(m) \tag{3.46}$$

$$\Psi_2(k) = \frac{1}{1-\rho_y^2(k)}\sum_m y_g^2(m) \tag{3.47}$$

$$\Psi_3(k) = \frac{1}{1-\rho_y^2(k)}\sum_m y_r(m)y_g(m). \tag{3.48}$$

The MAP estimators $\hat{\sigma}_r^2(k)$ and $\hat{\sigma}_g^2(k)$ contain unknown quantities $\sigma_{y_g}^2(k)$, $\sigma_{y_r}^2(k)$, $\rho_y(k)$, a_1, a_2, b_1, and b_2 that need to be replaced by suitable estimates. We choose the ML estimates for $\sigma_{y_g}^2(k)$, $\sigma_{y_r}^2(k)$ and $\rho_y(k)$. Considering $a_1 > 2$ and $a_2 > 2$, the hyperparameters of the IG priors for each subband are estimated by the following

MoM estimators

$$\hat{a}_1 = \frac{(\bar{\sigma}^2_{y_r})^2}{\mathrm{Var}(\sigma^2_{y_r})} + 2 \qquad b_1 = \bar{\sigma}^2_{y_r}(\hat{a}_1 - 1) \quad \forall\, \hat{\sigma}^2_{y_r} \in \text{subband} \tag{3.49}$$

$$\hat{a}_2 = \frac{(\bar{\sigma}^2_{y_g})^2}{\mathrm{Var}(\sigma^2_{y_g})} + 2 \qquad b_2 = \bar{\sigma}^2_{y_g}(\hat{a}_2 - 1) \quad \forall\, \hat{\sigma}^2_{y_g} \in \text{subband.} \tag{3.50}$$

where $\bar{\sigma}^2_{y_r}$ and $\bar{\sigma}^2_{y_g}$ represent the mean of the ML estimates of the variances of the noisy coefficients in a subband for the red and green channel images, respectively. Finally, the MAP estimates $\hat{\sigma}^2_r(k)$ and $\hat{\sigma}^2_g(k)$, and the ML estimate for $\rho(k)$ are substituted in (3.28) and (3.29) to obtain the MAP estimates for the noise-free wavelet coefficients.

(ii) Exponential prior

In this section, we assume that the 'joint prior' for $\sigma^2_{y_r}$ and $\sigma^2_{y_g}$ is a product of the following exponential priors

$$f_1(\sigma^2_{y_r}) = \lambda_1 \exp(-\lambda_1 \sigma^2_{y_r}), \quad \sigma^2_{y_r} \geq 0 \tag{3.51}$$

$$f_2(\sigma^2_{y_g}) = \lambda_2 \exp(-\lambda_2 \sigma^2_{y_g}), \quad \sigma^2_{y_g} \geq 0 \tag{3.52}$$

where λ_1 and λ_2 are the two subband-adaptive hyperparameters. Using the prior functions given in (3.51) and (3.52), the following likelihood equations are obtained

$$\frac{1}{1-\rho_y^2}\frac{\sum_m y_r^2(m)}{\sigma^2_{y_r}} - 2\lambda_1\sigma^2_{y_r} = M + \frac{\rho_y}{1-\rho_y^2}\sum_m \frac{y_r(m)}{\sigma_{y_r}}\frac{y_g(m)}{\sigma_{y_g}} \tag{3.53}$$

$$\frac{1}{1-\rho_y^2}\frac{\sum_m y_g^2(m)}{\sigma^2_{y_g}} - 2\lambda_2\sigma^2_{y_g} = M + \frac{\rho_y}{1-\rho_y^2}\sum_m \frac{y_r(m)}{\sigma_{y_r}}\frac{y_g(m)}{\sigma_{y_g}} \tag{3.54}$$

From (3.53) and (3.54) we have

$$\begin{aligned} &\frac{1}{1-\rho_y^2}\frac{\sum_m y_r^2(m)}{\sigma^2_{y_r}} - 2\lambda_1\sigma^2_{y_r} = \frac{1}{1-\rho_y^2}\frac{\sum_m y_g^2(m)}{\sigma^2_{y_g}} - 2\lambda_2\sigma^2_{y_g} \\ &\Rightarrow A\sigma^4_{y_r} + B\sigma^2_{y_r} + C = 0 \end{aligned} \tag{3.55}$$

where

$$A = 2\lambda_1, \quad B = \frac{1}{1-\rho_y^2}\frac{\sum_m y_g^2(m)}{\sigma^2_{y_g}} - 2\lambda_2\sigma^2_{y_g}, \quad \text{and} \quad C = -\frac{1}{1-\rho_y^2}\sum_m y_r^2(m)$$

Solving for $\sigma^2_{y_r}$ in (3.55) and using the relation $\sigma^2_{y_r} = \sigma^2_r + \sigma^2_v$, we obtain the MAP estimator for $\sigma^2_r(k)$ as

$$\hat{\sigma}^2_r(k) = \max\left(\frac{1}{4\lambda_1}\left[2\lambda_2\sigma^2_{y_g}(k) - \frac{\Psi_2(k)}{\sigma^2_{y_g}(k)} + \sqrt{\left(2\lambda_2\sigma^2_{y_g}(k) - \frac{\Psi_2(k)}{\sigma^2_{y_g}(k)}\right)^2 + 8\lambda_1\Psi_1(k)}\right] - \sigma^2_v, 0\right) \tag{3.56}$$

Similarly, we obtain the MAP estimator for $\sigma^2_g(k)$ as:

$$\hat{\sigma}^2_g(k) = \max\left(\frac{1}{4\lambda_2}\left[2\lambda_1\sigma^2_{y_r}(k) - \frac{\Psi_1(k)}{\sigma^2_{y_r}(k)} + \sqrt{\left(2\lambda_1\sigma^2_{y_r}(k) - \frac{\Psi_1(k)}{\sigma^2_{y_r}(k)}\right)^2 + 8\lambda_2\Psi_2(k)}\right] - \sigma^2_v, 0\right) \tag{3.57}$$

where $\Psi_1(k)$ and $\Psi_2(k)$ are given in (3.46) and (3.47), respectively. Since $\sigma^2_{y_g}(k)$, $\sigma^2_{y_r}(k)$, $\rho_y(k)$, λ_1 and λ_2 are unknown, they must be replaced by suitable estimates. We use the ML estimates given in (3.33), (3.34), and (3.35) to estimate $\sigma^2_{y_g}(k)$, $\sigma^2_{y_r}(k)$, and $\rho_y(k)$, respectively. Since σ^2_r and σ^2_g are random variables, whereas σ^2_v is a constant, $\mathrm{Var}(\sigma^2_{y_r}) = \mathrm{Var}(\sigma^2_r)$ and $\mathrm{Var}(\sigma^2_{y_g}) = \mathrm{Var}(\sigma^2_g)$. Hence, the hyperparameters λ_1 and λ_2 for each of the subbands are estimated as

$$\hat{\lambda}_1 = \frac{1}{\sqrt{\mathrm{Var}(\hat{\sigma}^2_r)}} \quad \forall\, \hat{\sigma}^2_r \in \text{subband} \tag{3.58}$$

$$\hat{\lambda}_2 = \frac{1}{\sqrt{\mathrm{Var}(\hat{\sigma}^2_g)}} \quad \forall\, \hat{\sigma}^2_g \in \text{subband} \tag{3.59}$$

where $\hat{\sigma}^2_r(k)$ and $\hat{\sigma}^2_g(k)$ are the ML estimates in (3.36) and (3.37), respectively. Finally, the unknown parameters in (3.28) and (3.29) are replaced by the MAP estimates $\hat{\sigma}^2_r(k)$ and $\hat{\sigma}^2_g(k)$ given above and the ML estimate $\hat{\rho}(k)$ given in (3.38) to obtain a better estimate of the noise-free wavelet coefficients. Fig. 3.4 summarizes the steps of the proposed noise reduction scheme.

It may be noted that the estimators for the local variances given in (3.44) and (3.45) using the IG prior, and in (3.56) and (3.57) using the exponential prior, yield

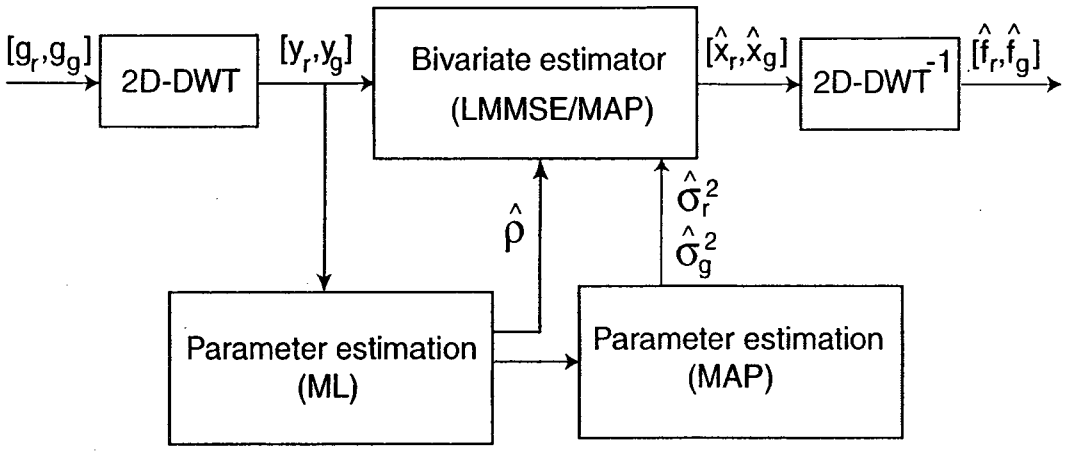


Figure 3.4: Block diagram for obtaining denoised cDNA microarray image using the proposed joint estimation technique.

non-negative real numbers. Moreover, the estimators are such that the variances of the wavelet coefficients for any one image are estimated using noisy coefficients from both the images. By using information from both channels, the dependency between the two images is accounted for, and therefore, it is expected that these estimators will yield good results.

It is to be pointed out that the two sets of MAP estimates for the local variances result in two separate denoising algorithms. The first algorithm, referred to as Proposed Method I, uses the local variance estimators given in (3.44) and (3.45) based on the IG prior. The main steps of this algorithm are given in Fig. 3.5. The second algorithm, which we call Proposed Method II, uses the local variance estimators given in (3.56) and (3.57) and is outlined in Fig. 3.6. To assess the computational efficiency of these algorithms, the computational complexity of each algorithm may be derived. In practice, this complexity refers to the number of elementary computer operations required by the algorithm to solve the given problem expressed as a function of the data size. The computational complexity of the algorithms for Proposed Method I and Proposed Method II may be derived as follows. For an image of size $N_1 \times N_2$, the computational complexity of the forward or inverse DWT may be found

1. Input: $g_r(i,j)$, $g_g(i,j)$, $i = 1, \cdots N_1$, $j = 1, \cdots N_2$
2. Forward transform: Obtain the noisy DWT coefficients $\boldsymbol{y}_r = \mathcal{W}\boldsymbol{g}_r$ and $\boldsymbol{y}_g = \mathcal{W}\boldsymbol{g}_g$, using the forward transform matrix $\mathcal{W}$
3. Estimation of DWT coefficients using the local parameters:
   for $l = 1 : J$ **do**
     for $O = 1 : 3$ $(1 \Rightarrow H, 2 \Rightarrow V, 3 \Rightarrow D)$ **do**
       $\forall k$ in a subband,
       Compute $\hat{\sigma}_r^2(k)$ (refer to (3.44))
       Compute $\hat{\sigma}_g^2(k)$ (refer to (3.45))
       Compute $\hat{\rho}(k)$ (refer to (3.38))
       Compute $\hat{x}_r(k)$ and $\hat{x}_g(k)$ (refer to (3.28) and (3.29))
     end for
   end for
4. Inverse transform: Obtain denoised images $\hat{\boldsymbol{f}}_r = \mathcal{W}^{-1}\hat{\boldsymbol{x}}_r$ and $\hat{\boldsymbol{f}}_g = \mathcal{W}^{-1}\hat{\boldsymbol{x}}_g$ using the inverse transform matrix $\mathcal{W}^{-1}$

Figure 3.5: Steps in the algorithm for Proposed Method I.

as $\mathcal{O}(\frac{16}{3}\mathcal{L}(1 - 4^{-J})N_1N_2)$, where $\mathcal{L}$ is the length of the scaling or wavelet vector, J is the highest-level of decomposition and $\mathcal{O}$ denotes the order of the calculation. Since J is a positive integer, the upper bound of this complexity is $\mathcal{O}(\frac{16}{3}\mathcal{L}N_1N_2)$. Proposed Method I or Proposed Method II requires that for each of the DWT coefficients, total 5 parameters are being estimated locally. These are: three ML estimates for the correlation coefficient and variances in the two channels, and two MAP estimates for the variances. Hence, for a window size of $w \times w$, the computational complexity due to the parameter estimation is $\mathcal{O}(5w^2N_1N_2)$. The complexity of obtaining the denoised coefficients using these local parameters is $\mathcal{O}(4N_1N_2)$. The overall computational complexity for each of the algorithms comprises the complexities of the forward and inverse DWT of two images, parameter estimation, and estimation of noise-free coefficients from the noisy ones. Since $w^2 \gg 1$ and $\mathcal{L} \gg 1$, the computational complexity of any of the proposed DWT-based denoising algorithm may be obtained as

1. Input: $g_r(i,j)$, $g_g(i,j)$, $i = 1, \cdots N_1$, $j = 1, \cdots N_2$
2. Forward transform: Obtain noisy DWT coefficients $\boldsymbol{y}_r = \mathcal{W}\boldsymbol{g}_r$ and $\boldsymbol{y}_g = \mathcal{W}\boldsymbol{g}_g$, using the forward transform matrix $\mathcal{W}$
3. Estimation of DWT coefficients using the local parameters:
   for $l = 1 : J$ **do**
   for $O = 1 : 3$ $(1 \Rightarrow H, 2 \Rightarrow V, 3 \Rightarrow D)$ **do**
   $\forall k$ in a subband,
   Compute $\hat{\sigma}_r^2(k)$ (refer to (3.56))
   Compute $\hat{\sigma}_g^2(k)$ (refer to (3.57))
   Compute $\hat{\rho}(k)$ (refer to (3.38))
   Compute $\hat{x}_r(k)$ and $\hat{x}_g(k)$ (refer to (3.28) and (3.29))
   end for
   end for
4. Inverse transform: Obtain denoised images $\hat{\boldsymbol{f}}_r = \mathcal{W}^{-1}\hat{\boldsymbol{x}}_r$ and $\hat{\boldsymbol{f}}_g = \mathcal{W}^{-1}\hat{\boldsymbol{x}}_g$ using the inverse transform matrix $\mathcal{W}^{-1}$

Figure 3.6: Steps in the algorithm for Proposed Method II.

$\mathcal{O}((5w^2 + \frac{64}{3}\mathcal{L})N_1N_2)$.

3.5 Estimation of Log-Intensity Ratio

The purpose of microarray image denoising is to improve the extraction of statistical information regarding gene expression levels instead of visual enhancement as in traditional image denoising. Therefore, a successful microarray denoising algorithm is the one that reduces noise with minimal loss of information for downstream statistical analysis. In cDNA microarray experiments, gene expression is commonly quantified in terms of the log-intensity ratio that is given by (Jornsten et al., 2002):

$$R = \log_2 \left(\frac{\frac{1}{n}\sum_{S_r \in ROI} S_r - b_r}{\frac{1}{n}\sum_{S_g \in ROI} S_g - b_g} \right) \tag{3.60}$$

where S_r and S_g denote pixel intensities in the red and green channel images, respectively. ROI refers to the region of interest or spot where hybridization occurs, n is

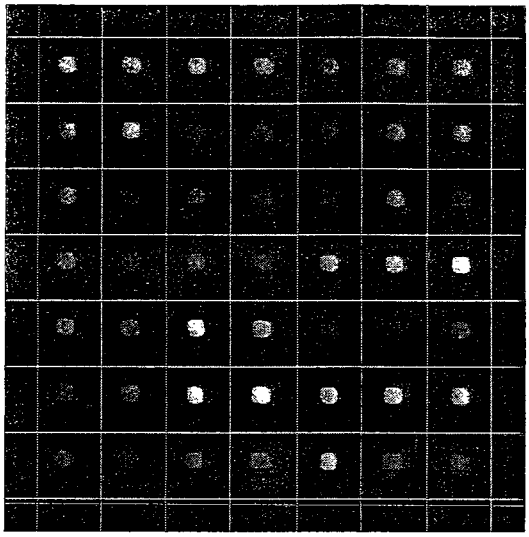


Figure 3.7: Grid identifying target areas for spots in a red channel microarray image.

the number of pixels in the ROI, and b_r and b_g are the median of the pixel intensities of local background corresponding to the ROI of two channels. By comparing the treatment versus control samples on a gene-by-gene basis, the log-intensity ratios identify genes that are activated or repressed by the treatment. Although relative differences in gene expression may also be measured by the ratio of spot intensities in the red and green channel images, the log-intensity ratio is preferred instead. This is because intensity ratios have the disadvantage of treating up- and down-regulated genes differently. For instance, genes up-regulated by a factor of 2 have an intensity ratio of 2, whereas those down-regulated by the same factor have an intensity ratio of -0.5. The logarithm base 2 transform has the advantage of producing a continuous spectrum of values and treating the intensity ratio (or their reciprocals) symmetrically, so that a gene up-regulated by a factor of 2 has a log-intensity ratio of 1, a gene down-regulated by a factor of 2 has a log-intensity ratio of -1, and a gene expressed at a constant level (with a ratio of 1) has a log-intensity ratio equal to zero. (Hayat, 2005). Another advantage is that the distribution of log-transformed intensity ratios is approximately normal. Presence of noise in microarray images can seriously distort

the pixel intensities of a spot and corresponding background, and thus, the estimated log-intensity ratio values. Hence, an effective way to assess the competence of various denoising algorithms for microarray images is to compare the accuracy of the log-intensity ratios that are estimated from the denoised images.

To calculate R, the first step is to identify target areas in the red and green channel images. In general, the target area is a square or rectangular region on an image enclosing one spot. These regions are identified by placing a grid over the entire image. Fig. 3.7 shows an example of such a gridding technique. Since this gridding is made on a regular array, the grid lines are evenly spaced resulting in square-shaped target areas, each containing a spot and the background. The next step, known as segmentation, consists of identifying the pixels that belong to the ROI and background in a target area. There exists several segmentation methods, each having certain advantages and drawbacks. Some of these methods have been compared in Zhang et al. (2004) and Yang et al. (2002). In this chapter, we use a histogram method, which defines the background and ROI as pixels with intensities between the 5th and 20th percentile and between the 80th and 95th percentile, respectively (Bozinov and Rahnenfuhrer, 2002; McLachlan et al., 2004). This segmentation method is chosen for its simplicity, which is why it has been implemented in well known microarray image processing softwares, such as QuantArray (McLachlan et al., 2004). The local background intensity for each spot, b_r or b_g, is estimated using the median of pixel intensities in the background of the square target area for each image.

3.6 Experimental Results

Extensive experimentations are carried out to compare the performance of the proposed algorithms with that of others on several microarray images obtained from real experiments. The experiments are conducted on a large number of microarray images

that have been downloaded from the website of the Stanford MicroArray Database [2]. In order to conduct the experiments original noise-free images are necessary. Since perfectly noise-free images are not available in practice, we choose as noise-free images those that appear to be corrupted with very little noise. The images are selected by inspection in the following manner. First, a synthetic background is created by adding Gaussian noise on a zero-intensity background. This is then compared with the background of the test images by zooming in. Images having backgrounds that do not resemble the synthetically corrupted zero-intensity background and possess almost no artifacts are chosen as the noise-free images. The images that have been used in the experiments are 16-bit TIFF files of size 1000×1000.

Noisy images are synthetically created by adding bivariate Gaussian noise to the noise-free images considering four values of ρ_ε, viz., 0, 0.25, 0.50, and 0.75 and five values of σ_ε, viz., 800, 1000, 1200, 1400 and 1600. The performance of the noise reduction algorithms are compared using two sets of experiments. In the first set of experiments, the denoising performance of each algorithm is quantified in decibels (dB) using the peak signal-to-noise ratio (PSNR) ratio which is defined as[3]

$$\mathrm{PSNR}(\hat{f}) = 10 \log_{10} \left(\frac{65535^2}{\mathrm{MSE}(\hat{f})} \right) \tag{3.61}$$

where $\mathrm{MSE}(\hat{f})$ is the MSE of the denoised image that is given by

$$\mathrm{MSE}(\hat{f}) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \left(\hat{f}(i,j) - f(i,j) \right)^2. \tag{3.62}$$

and f represents the true microarray image of size $N_1 \times N_2$. This measure is inversely proportional to the residual error in an image. Thus, a higher PSNR implies a better noise reduction performance and vice versa. In the second set of experiments, we estimate values of the log-intensity ratio R from the denoised images and compare

[2] http://smd.stanford.edu/index.shtml

[3] The peak of the image signal is $2^{16} - 1 = 65535$.

these values with the estimates obtained from the noise-free images. Distances between these two estimates are measured using the mean absolute error (MAE) given by

$$\mathrm{MAE}(\hat{R}) = \frac{1}{\aleph}\sum_{q=1}^{\aleph} |\hat{R}_q - R_q|. \tag{3.63}$$

where R_q is the log-intensity ratio for the q-th spot estimated using the noise-free image, $\hat{R}_q$ is the estimate obtained from the denoised image, and $\aleph$ is the total number of spots in the image. A lower value of the MAE indicates a better performance of the denoising algorithm due to greater accuracy in the estimation of R.

In Section 3.4.3, MAP estimation of the local variance parameters were described using the IG prior as well as the exponential prior. Thus we implement two denoising algorithms, namely, Proposed Method I and Proposed Method II corresponding to the IG and exponential prior functions, respectively. Apart from the proposed methods, eight different DWT-based denoising methods have been considered in the simulation, namely, the VisuShrink (Donoho, 1995), SureShrink (Donoho and Johnstone, 1995), GarroteShrink (Gao, 1998), BayesShrink (Chang et al., 2000a), EBayesThresh (Johnstone and Silverman, 2005), LAWMAP (Mihçak et al., 1999), NeighCoef (Cai and Silverman, 2001) and BiShrink (Sendur and Selesnick, 2002a). Since these methods do not take into account the presence of correlated noise between the two channels, it is reasonable to compare these methods with the proposed ones under the assumption of uncorrelated noise. Therefore, these methods are compared for the case when $\rho_\varepsilon = 0$. In the experiments, the DWT of the images are obtained using Sym8 (Daubechies, 1992). A four-level decomposition of the DWT is chosen, since any further decomposition level does not produce a significant increase in denoising performance. Among the methods that are considered, the VisuShrink uses a universal threshold parameter for all subbands, the GarroteShrink, BayesShrink, SureShrink, and EBayesThresh use subband-adaptive parameters, whereas the LAWMAP, NeighCoef, BiShrink and the

Table 3.4: Output PSNR values in dB for the DWT-based denoising algorithms at various noise strengths when $\rho_\epsilon = 0$.

σ_ϵ	800	1000	1200	1400	1600
$\text{IPSNR}_g^\dagger$	38.29	36.35	34.77	33.43	32.27
$\text{IPSNR}_r^\dagger$	38.27	36.33	34.75	33.41	32.25
Methods					
VisuShrink	42.78	41.59	40.53	39.73	38.50
	43.44	42.24	41.26	40.35	38.97
SureShrink	44.04	42.82	41.80	40.93	39.88
	44.74	43.46	42.47	41.62	40.44
GarroteShrink	43.29	41.82	40.66	39.72	39.78
	43.55	42.07	40.80	39.78	40.11
BayesShrink	43.07	42.05	41.19	40.42	39.71
	43.81	42.73	41.79	40.98	40.39
EBayesThresh	43.85	42.64	41.49	40.63	39.95
	44.49	43.12	42.02	41.16	40.37
LAWMAP	44.85	43.61	42.58	41.72	40.86
	45.51	44.20	43.14	42.26	41.45
NeighCoef	45.50	44.27	43.18	42.22	41.15
	46.28	44.94	43.81	42.86	41.67
BiShrink	45.24	44.05	43.04	42.17	41.41
	45.96	44.67	43.59	42.68	41.87
Proposed I	45.53	44.21	43.30	42.50	41.78
	46.34	44.96	43.99	43.15	42.41
Proposed II	**45.59**	**44.44**	**43.43**	**42.56**	**41.79**
	46.49	**45.33**	**44.33**	**43.44**	**42.68**

† IPSNR_g and IPSNR_r denote the input PSNRs for the green and red channel images respectively.

proposed methods use locally-adaptive parameters. Only BiShrink considers the inter-scale dependency between the wavelet coefficients in addition to the intra-subband one by using a bivariate circularly symmetric PDF to model the dependency between a coefficient and its parent. The denoising performance of the locally-adaptive methods are tested using several window sizes, such as 3×3, 5×5, 7×7, and 9×9. Except for the NeighCoef, which is designed for a 3×3 window, the results of the remaining three locally-adaptive methods are shown for a 5×5 window, since this size gives the best denoising performance in most of the cases. It is to be noted that

the PSNR results obtained for the proposed methods remain almost the same when 7×7 window size is used.

Table 3.4 shows the output PSNR values for various denoising techniques obtained by averaging the results of twenty-five microarray images. The highest PSNR values among the methods for each of the noise levels are highlighted in boldface. It can be seen from the table that the locally-adaptive methods have a better performance compared to the subband-adaptive methods. Moreover, in terms of output PSNR values, both Proposed Method I and Proposed Method II outperform the other locally-adaptive methods at all noise levels. Comparing between these two methods, we see that Proposed Method II, which uses the exponential prior, gives better denoising performance than Proposed Method I, which uses the IG prior. In the remainder of this chapter, Proposed Method II will be referred to as simply the Proposed Method, and further comparisons will be made with this method only. We see that the NeighCoef, which gives the nearest PSNR values to the Proposed Method for nearly all noise levels, falls short by 0.40 dB on average. The significance of the observed differences is better understood in terms of the MSE, which measures the residual error in an image. The MSE's for the two competing methods, namely, NeighCoef and LAWMAP, are higher than that of the Proposed Method by an average of 35×10^3 units and 53×10^3 units, respectively. These high figures indicate significant improvement in denoising performance of the Proposed Method over the competing ones. It is interesting to note that in spite of accounting for the inter-scale dependencies within an image, BiShrink does not perform as well as the Proposed Method. This suggests that for the denoising of cDNA microarray images, the consideration of inter-channel dependency is more important than the inter-scale one. In addition to good denoising performance, our method is computationally efficient. For example, the average time required to implement the Proposed Method on an image of size 1000×1000 in the

MATLAB environment using a 2 GHz processor with 512 MB RAM is 1.71 seconds. Under the same setup, the two competing methods, viz., LAWMAP and NeighCoef, require 1.01 and 0.93 seconds, respectively. Thus, in terms of computational complexity the Proposed Method is comparable to the other methods. It is observed that the output PSNRs obtained for the red channel images are consistently higher than that of the green channel images for every denoising method. A possible explanation for this is that the signal-to-noise ratios are higher in the red channel images than in the green channel images.

Residual noise images are obtained as the absolute difference between the denoised image and the original image. These images are useful for making visual assessments of denoising performance because the brightness of a pixel on the image reflects the magnitude of the estimation error at that spatial location. Fig. 3.8 shows residual noise images for SureShrink, which is the competing subband-adaptive method, NeighCoef, which is the competing locally-adaptive method, and the Proposed Method. As compared to the other methods, the residual noise image for the Proposed Method contains spots that are less bright. This indicates that the Proposed Method more successfully removes noise from the spots. More detailed comparisons can be made by observing Fig. 3.9, which shows the 3-D representation of a noise-free spot, its noisy version corrupted with a noise sequence having $\sigma_\epsilon = 1200$ and $\rho_\epsilon = 0.25$, and the corresponding denoised versions using the SureShrink, LAWMAP, NeighCoef and the Proposed Method. This figure clearly reveals that the Proposed Method removes most of the noise from both the background and spot while preserving most of the signal magnitudes on the spot. On the other hand, the spot is oversmoothed by the other denoising methods including the NeighCoef. It is to be

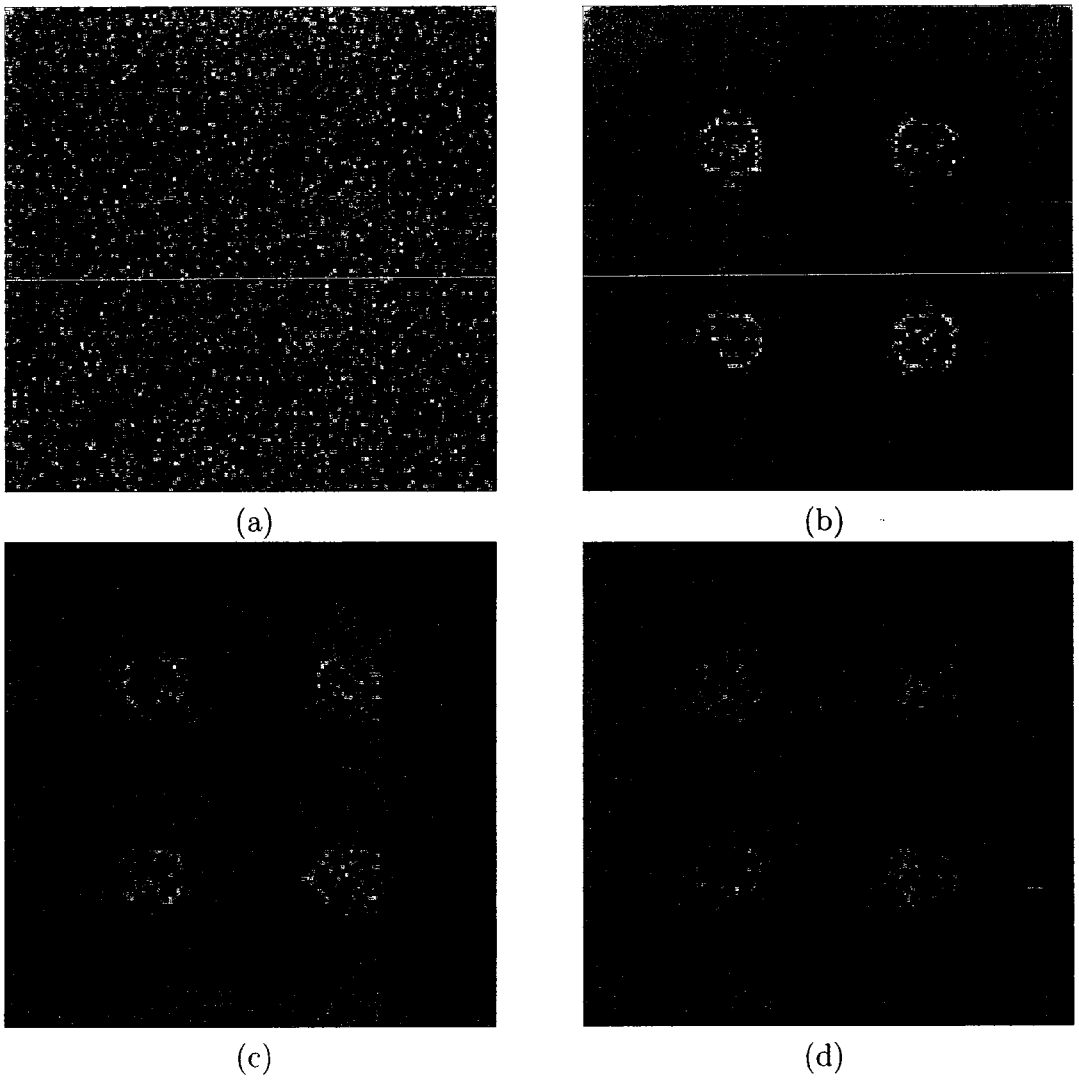


Figure 3.8: Residual noise images using various denoising algorithms with $\sigma_\varepsilon = 1200$ and $\rho_\varepsilon = 0.25$. (a) Noisy image (b) SureShrink (c) NeighCoef, and (d) Proposed Method.

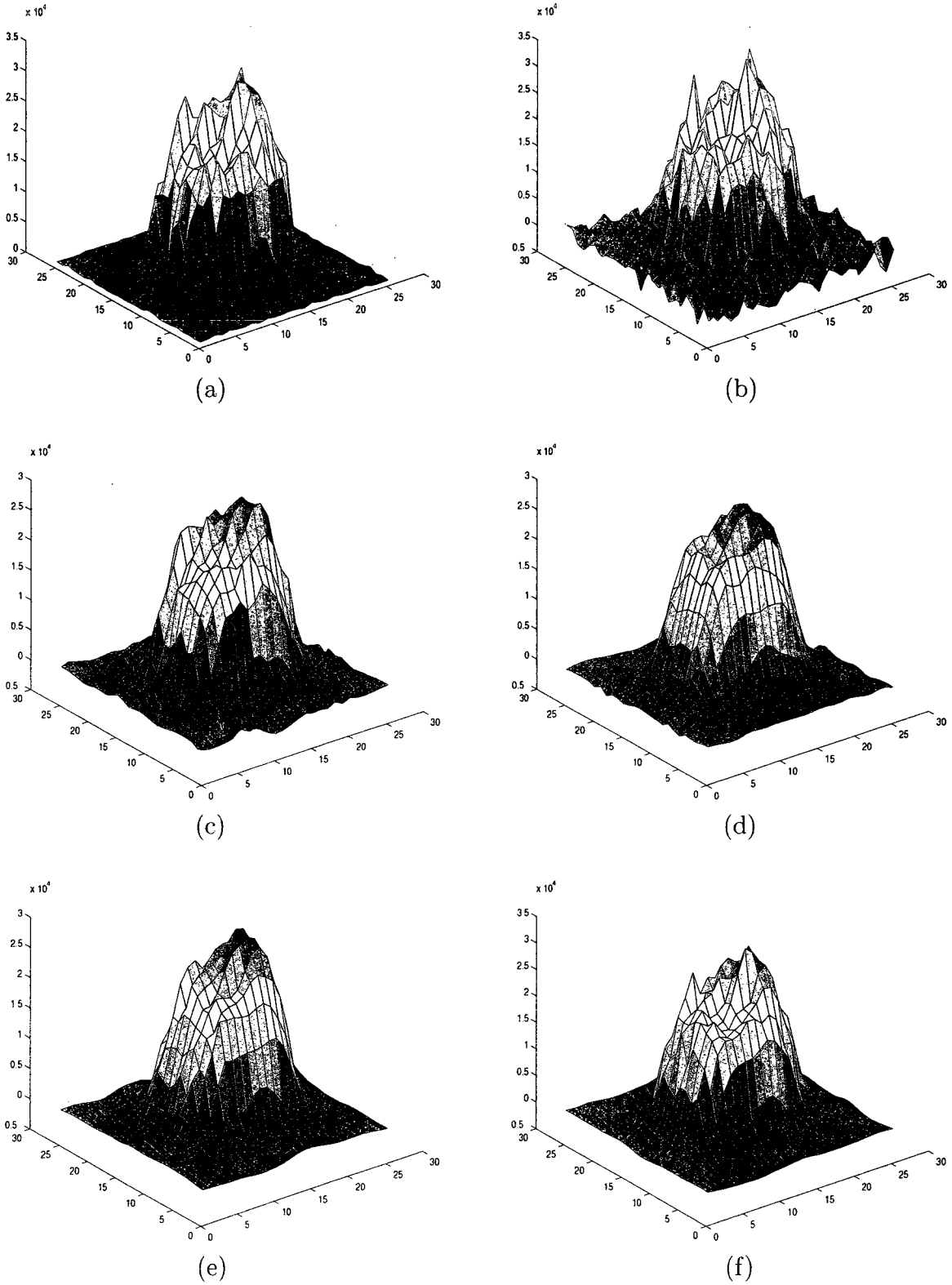


Figure 3.9: 3-D visualization of a spot. (a) Original and (b) noisy version of the spot with $\sigma_\varepsilon = 1200$ and $\rho_\varepsilon = 0.25$. Denoised versions of the spot using (c) SureShrink (d) LAWMAP (e) NeighCoef, and (f) Proposed Method.

Table 3.5: Output MAE values of log-intensity ratios for the DWT-based denoising algorithms at various noise strengths when $\rho_\varepsilon = 0$.

σ_ε	800	1000	1200	1400	1600
(Input MAE)	1.2602	1.4220	1.5972	1.7327	1.8580
Methods					
VisuShrink	0.4075	0.4077	0.4079	0.4088	0.4097
SureShrink	0.2547	0.3047	0.3396	0.3987	0.4056
GarroteShrink	0.4267	0.5327	0.6283	0.7003	0.7363
BayesShrink	0.7336	0.8255	0.8544	0.8802	0.9090
EBayesThresh	0.4603	0.5340	0.6012	0.6681	0.6587
LAWMAP	0.3218	0.4038	0.4687	0.5249	0.5705
NeighCoef	0.3012	0.3165	0.3265	0.3292	0.3328
BiShrink	**0.2312**	0.3080	0.3695	0.4353	0.4925
Proposed	0.2595	**0.2928**	**0.3113**	**0.3229**	**0.3275**

noted that oversmoothing is undesirable because less accurate gene activity measurements result due to the reduced magnitude of spot intensities and inaccurate selection of spot and background pixels in the segmentation stage resulting from the blur in the image.

Table 3.5 shows the values of the output MAE for estimating the log-intensity ratio R using the nine DWT-based denoising methods, wherein the images are corrupted with the five different noise levels considered in this section and with $\rho_\varepsilon = 0$ for each noise sequence. The estimated values of R are obtained by averaging over the same set of test images that are used for Table 3.4. It can be observed from this table that the output MAEs of a noise reduction scheme are lower than the corresponding input MAEs, thus indicating that an efficient denoising algorithm for microarray images is essential for a better accuracy in the estimation of R. The tabular results indicate that the locally-adaptive noise reduction methods provide lower output MAE than the subband-adaptive ones in general. A similar observation has been made from the results based on the PSNR values in Table 3.4. From Table 3.5, it can be further seen that the proposed noise reduction method provides the lowest values of MAE for

Table 3.6: Output PSNR values of green and red channel images using the Proposed Method for different noise correlations and noise strengths.

σ_ε^a	800	1000	1200	1400	1600
$\rho_\varepsilon^b = 0.25$	45.86, 46.73	44.73, 45.53	43.70, 44.54	42.82, 43.67	42.07, 42.91
$\rho_\varepsilon = 0.50$	45.80, 46.65	44.65, 45.43	43.60, 44.42	42.70, 43.54	41.93, 42.76
$\rho_\varepsilon = 0.75$	45.81, 46.62	44.64, 45.37	43.57, 44.34	42.64, 43.45	41.87, 42.66

various noise levels except for one instance, i.e., $\sigma_\varepsilon = 800$. It is to be noted that even for this case, the MAE value given by the Proposed Method is close to the lowest one provided by the BiShrink method. We conclude that among the methods compared, the proposed one is the most efficient, since it consistently provides the lowest MAE at higher noise levels.

Experiments are carried out to investigate the performance of the Proposed Method in the presence of correlation between noise of the two channels. Table 3.6 shows the output PSNR of the Proposed Method for green and red channel images using three different noise correlations, viz., $\rho_\varepsilon = 0.25$, 0.50, and 0.75, at various noise levels. If we compare Table 3.4 and Table 3.6, it can be seen that when there is low correlation of noise between the channels, such as $\rho_\varepsilon = 0.25$, the Proposed Method can provide even better output PSNRs than in the case of zero noise correlation. However, the PSNR values decline with increasing values for ρ_ε, especially at higher noise standard deviations. This observation is consistent with the findings of a recent study (Miller and Kingsbury, 2008) where it has been shown that the MSE of a bivariate estimator increases with the decreasing absolute deviation between the noise and signal correlations. However, from Table 3.6 it is evident that the PSNR values do not decline by more than 0.25 dB. Hence, the Proposed Method provides good quality output images even when the noise sequences of the two channels are correlated.

3.7 Conclusion

In this chapter, we have proposed new methods for jointly estimating the DWT coefficients of two-channel cDNA microarray images corrupted by AGN. We have used both LMMSE estimation as well as MAP estimation for deriving joint estimators for the wavelet coefficients in the two channels. Both estimation techniques lead to the same bivariate estimator. Motivation for using a joint estimation technique arises from the observation that significant correlation exists between DWT coefficients in the two images of cDNA microarrays and thus an estimation technique that uses information from both images is expected to perform better. The bivariate Gaussian PDF is used as the prior function for the noise-free coefficients in both the channels. An important feature of the proposed estimator is that it 'jointly' estimates the noise-free coefficients in the two channels while taking into account the inter-channel dependency of the noise.

In order to obtain improved estimates of the unknown local variances in the proposed estimator, we have investigated the use of inverse-gamma and exponential priors for these parameters in MAP estimation. Separate denoising algorithms were obtained for each choice of the prior. The simulation experiments revealed that better denoising performance is attained when the exponential prior is used. The empirical prior distribution for the correlation coefficient has not shown a consistent behavior in the different subbands. For simplicity in implementation, the ML estimate for this parameter was chosen. The effectiveness of the proposed denoising scheme was tested by estimating the log-intensity ratio, which is the most important output statistic for analysis of cDNA microarray experiments. Simulation results showed that the Proposed Method gives better denoising performance than existing methods both in terms of the PSNR of estimated images and MAE of estimated log-intensity ratios. Even when there exists correlation between noise in the two channels, the Proposed

Method is still found to be highly efficient. As a final comment, the denoising algorithms that been proposed may appear to be computationally intensive at the outset, but they are actually simple and fast to implement due to the closed-form solution that is obtained.

Chapter 4

CWT-Based Denoising Algorithms for cDNA Microarray Images

4.1 Introduction

The success of an image denoising algorithm is largely dependent on the type of transform used. Although the DWT-based methods proposed in Chapter 3 performed very well, further improvements in denoising performance could be achieved by overcoming two major drawbacks of the DWT, namely, poor directional selectivity and lack of shift invariance. In the context of microarray image denoising, an improved directional selectivity and a better shift invariance property of the transform are desirable for better reconstruction of edges in the spots that have approximately circular shape. An introduction to 2D DWT and CWT has been given in Chapter 2. This chapter motivates the use of CWT-based algorithms for microarray image denoising and describes some of its features that overcome limitations of the traditional DWT-based methods.

The CWT coefficients of an image have two types of representations, namely, Cartesian and polar. Since the magnitude components of polar representation have a better shift-invariance property than the individual real and imaginary components of Cartesian representation (Sendur and Selesnick, 2002a,b; Achim and Kuruoğlu, 2005), the former is chosen for denoising instead of the latter. It is observed that

the magnitude components of the CWT coefficients in a given subband are correlated between the red and green channel images (Howlader and Chaubey, 2008, 2009a). It is expected that magnitude components of the noise coefficients in the two channels will be correlated as well. The inter-channel dependencies between magnitude components of the image coefficients as well as the noise coefficients are ignored when the images are processed separately. For this reason, existing CWT-based methods do not provide satisfactory denoising performance when applied for microarray images. In this chapter, new bivariate LMMSE and MAP estimators (Howlader and Chaubey, 2009a) are developed that jointly estimate the magnitude components of the CWT coefficients in the two channels. Experimental results show that by taking into account the inter-channel dependency between magnitude components of the image coefficients as well as the noise, these estimators yield better noise reduction performance and more accurate estimates of the log-intensity ratios than existing CWT-based methods. In addition, the proposed CWT-based methods are contrasted with the DWT-based methods proposed in Chapter 3 both in terms of denoising performance as well as computational efficiency.

4.2 Motivation for Using CWT in Microarray Image Denoising

The DWT used in image denoising can be of many types, such as orthogonal or biorthogonal, real or complex-valued, separable or nonseparable, and decimated or non-decimated. The choice of the type of DWT to be used should depend, among other factors, on the characteristics of the image. In Chapter 3, we proposed noise reduction algorithms for cDNA microarray images based on the decimated DWT that are computationally very efficient. The proposed algorithm was highly successful in meeting the two most important objectives of microarray image denoising: apprecia-

ble reduction of noise level and more accurate estimation of gene expression levels. However, the decimated DWT is not necessarily the 'best' transform to use for microarray images. Microarray images are highly structured and consist of thousands of spots that are roughly circular in shape arranged on a regular grid. The decimated DWT lacks two important properties, namely, shift invariance and good directional selectivity, that are useful in the denoising of images possessing circular features or edges. Due to lack of shift-invariance of the DWT, shifting the signal by one results in different coefficients at the next lower level. The DWT is said to have poor directional selectivity because it is capable of detecting only the horizontal, vertical and diagonal features within an image. Both shift-invariance and good directional selectivity are desirable for microarray image denoising because they enable better reconstruction of the approximately circular edges of the spots.

An improved image denoising performance can be achieved by using a non-decimated or redundant form of the DWT referred to as SWT (Mallat, 1999), which has better shift-invariance property than the decimated one, but suffers from substantially increased computation requirements. In addition, the directional selectivity of the SWT is same as that of the DWT. There are other redundant transforms having better shift-invariance and directional selectivity properties such as the CWT (Kingsbury, 1999, 2001; Selesnick, 2001, 2002; Selesnick et al., 2005; Fernandes et al., 2003; Clonda et al., 2004; Barber and Nason, 2004; Sendur and Selesnick, 2002b; Olkkonen et al., 2006; Achim and Kuruoğlu, 2005), the curvelet (Starck et al., 2002), the contourlet (Eslami and Radha, 2006), the wedgelet, the bandlet, the steerable pyramid (Portilla et al., 2001), the matching pursuit, and the basis pursuit. Methods based on these transforms lead to better denoising performance than that of the decimated DWT, but involve greater computational complexity. However, among these various transforms, the CWT is preferable due to its minimum redundancy, which makes

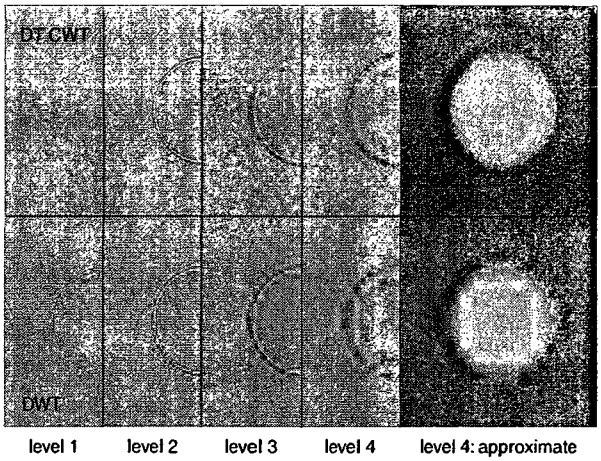


Figure 4.1: Reconstruction of a circular disc image on a dark background using 4-level CWT and DWT (Kingsbury, 1999).

the CWT-based denoising algorithms computationally more efficient than the other redundant transform-based algorithms. A computationally efficient algorithm is desirable in the context of microarray image denoising due to the huge volume of data that needs to be processed.

Unlike the separable DWT coefficients, which capture only three directional features in an image, the wavelet functions of the CWT capture six directional features. Hence this transform has directional selectivity that is better than the classical DWT. The transform is approximately shift-invariant so that there are almost no blurred edges or ripple artifacts in the denoised image. Thus segmentation of the spot from its background can be performed more accurately. In addition, CWT-based denoising methods are computationally more efficient than transforms having similar properties due to its limited redundancy. Fig. 4.1 compares the efficiency of the CWT relative to the DWT for detecting the circular edges of a spot that is typical for a microarray image. The upper row shows the output images reconstructed from the CWT coefficients, while the lower row shows the output images when the decimated DWT is used. In the lower row, there are substantial artifacts in the form of irregular edges. These artifacts are absent in the CWT images illustrating good shift-invariance and

directional selectivity properties of the latter. These results motivate us to propose new algorithms for reducing AGN in cDNA microarray images using the CWT.

4.3 Estimation of CWT Coefficients

Consider the additive noise model in the pixel domain given in Section 3.4. We formulate this model in the CWT domain as follows. Let $x_r(k_1, k_2)$ and $x_g(k_1, k_2)$ denote the polar representation of the CWT coefficients of the noise-free red and green channel images, respectively, at the spatial location (k_1, k_2) of a given subband. Since the CWT is a linear transformation, the noisy coefficients of the images at that spatial location can be written in the polar representation as

$$y_r(k_1, k_2) = x_r(k_1, k_2) + v_r(k_1, k_2) \tag{4.1}$$

$$y_g(k_1, k_2) = x_g(k_1, k_2) + v_g(k_1, k_2) \tag{4.2}$$

where $v_r(k_1, k_2)$ and $v_g(k_1, k_2)$ are the noise coefficients of the red and green channel images, respectively. Let $x_r^{\mathrm{r}}(k_1, k_2)$ and $x_g^{\mathrm{r}}(k_1, k_2)$ be the magnitude components of $x_r(k_1, k_2)$ and $x_g(k_1, k_2)$, respectively, and $x_r^{\theta}(k_1, k_2)$ and $x_g^{\theta}(k_1, k_2)$ be the corresponding phase components. Similar notations may be used for $y_r(k_1, k_2)$, $y_g(k_1, k_2)$, $v_r(k_1, k_2)$ and $v_g(k_1, k_2)$. In the polar representation, both the magnitude and phase components are necessary for reconstructing the denoised image. Several studies have shown that the denoised magnitude components along with the noisy phase components are sufficient for obtaining an appreciable noise reduction performance (see for example, Sendur and Selesnick, 2002a; Achim and Kuruoğlu, 2005; Sendur and Selesnick, 2002b). Motivated by these results, bivariate estimators are derived to reduce the noise from the magnitude components of the CWT coefficients of microarray images. The denoised images are then reconstructed using the inverse CWT of $\hat{x}_r^{\mathrm{r}}(k_1, k_2)\angle y_r^{\theta}(k_1, k_2)$ and $\hat{x}_g^{\mathrm{r}}(k_1, k_2)\angle y_g^{\theta}(k_1, k_2)$. Since estimators are derived for the magnitude components, hereafter in this paper, the superscripts r and θ are dropped

Table 4.1: Average of inter-channel correlations for the magnitude components of CWT coefficients in various subbands using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
$HH1_l$	0.48	0.82	0.88	0.96
$HH2_l$	0.52	0.76	0.94	0.96
$HL1_l$	0.77	0.91	0.95	0.97
$HL2_l$	0.67	0.91	0.96	0.96
$LH1_l$	0.71	0.93	0.96	0.96
$LH2_l$	0.77	0.91	0.95	0.97

for notational simplicity and the resulting notations represent the magnitude components of the CWT coefficients.

As in the case of DWT coefficients of images, the CWT coefficients are considered to be locally non-stationary to take into account the intra-subband dependency (Sendur and Selesnick, 2002b). As a result, the random variables of the magnitude components of the coefficients are index dependent. Let $X_r(k_1, k_2)$ and $X_g(k_1, k_2)$ be the random variables for $x_r(k_1, k_2)$ and $x_g(k_1, k_2)$, and $Y_r(k_1, k_2)$ and $Y_g(k_1, k_2)$ be the random variables for the noisy observations $y_r(k_1, k_2)$ and $y_g(k_1, k_2)$. Since the CWT coefficients of noise are spatially stationary, the random variables V_r and V_g representing the magnitude components of noise are index independent. For notational convenience, the indices are suppressed unless stated otherwise.

The existing CWT-based methods for image denoising can be used to process the red and green channel images separately. In doing so, these methods ignore the inter-channel dependencies that exist between the coefficients. In Chapter 3, experiments were conducted on several microarray images which revealed the presence of significant correlation between the DWT coefficients of the two channels. Since the CWT is a linear transform, and strong dependency is observed between the images in the pixel domain, magnitude components of the CWT coefficients should be highly correlated as well. This is experimentally verified by measuring the correlation between the local

Table 4.2: Average percentage of magnitude components X_r and X_g that are significantly correlated in various subbands using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
$HH1_l$	35	79	100	100
$HH2_l$	34	78	100	100
$HL1_l$	35	87	100	100
$HL2_l$	36	89	100	100
$LH1_l$	36	84	100	100
$LH2_l$	36	84	100	100

neighboring magnitude components of the two channels for 10 typical cDNA microarray images. Table 4.1 shows the average values of the local correlations obtained using the same set-up as in Table 3.1. It is seen from this table that in all subbands, the inter-channel correlations for the magnitude components are much higher than the inter-channel correlations for the DWT coefficients given in Table 3.1. This is because, in general, the magnitude components of the CWT possess more signal features than the DWT coefficients. Tests of significance for the correlation coefficient are performed for each pair of magnitude components X_r and X_g using the same test procedure and experimental set-up as described for Table 3.2. It can be seen from Table 4.2 that for a subband in decomposition level $l > 2$, 100% of the components X_r and X_g are significantly correlated with each other. In addition, a large proportion of the components (more than 75%) in the subbands of decomposition level $l = 2$, and at least one-third of these in $l = 1$ are significantly correlated. Thus, joint estimation of the magnitude components is a necessity for obtaining a satisfactory denoising performance. The bivariate PDF of the local neighboring magnitude components of two channels plays a significant role for such joint estimation.

Both real-valued as well as non-negative valued random variables have been used to model the magnitude components of CWT coefficients in a subband of an image. Examples include, the univariate α-stable distributions (Achim and Kuruoğlu, 2005)

Table 4.3: Average percentage of X_r and X_g in each subband that follow the bivariate Gaussian PDF by Mardia's test (Mardia, 1970) using 10 typical cDNA microarray images.

Subband	Level of decomposition			
	$l=1$	$l=2$	$l=3$	$l=4$
$HH1_l$	94.38	86.44	72.64	94.08
$HH2_l$	94.35	88.88	75.04	95.86
$HL1_l$	93.18	86.44	69.60	89.35
$HL2_l$	93.19	86.80	71.52	89.94
$LH1_l$	94.26	86.96	73.76	90.53
$LH2_l$	94.50	87.60	71.84	87.57

and the univariate Rayleigh mixture model (Ferrari and Winsor, 2005; Shaffrey et al., 2002). Although magnitude components are non-negative random variables, probability models for non-negative data are not appropriate for modelling the magnitude components of CWT coefficients of microarray images. This is because, microarray images contain some smooth regions in the background, which give rise to a non-negligible proportion of zero-valued magnitude components. Probability models for non-negative data, such as the Rayleigh and generalized gamma PDFs, fail to capture this important feature. On the other hand, PDFs such as the α-stable distributions become complicated when extended to their bivariate form. Further, in the locally adaptive methods, use of complicated PDFs results in high computational cost since closed-form solutions cannot be obtained in most of the cases. It is observed that when such PDFs are used in modelling the local neighboring coefficients, the values of the parameters of the fitted model are such that these PDFs can be very closely approximated by the Gaussian distribution. Since it is desirable to use a bivariate PDF that provides satisfactory fit to the data and is mathematically tractable, we choose the joint density function of X_r and X_g to be the bivariate Gaussian PDF

given by

$$p_{X_r X_g}(x_r, x_g) = \frac{1}{2\pi\sigma_r\sigma_g\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)} \times \left\{ \frac{(x_r-\mu_r)^2}{\sigma_r^2} + \frac{(x_g-\mu_g)^2}{\sigma_g^2} - 2\rho\frac{(x_r-\mu_r)}{\sigma_r}\frac{(x_g-\mu_g)}{\sigma_g} \right\} \right], \tag{4.3}$$

where the parameters μ_r ($\mu_r \geq 0$), μ_g ($\mu_g \geq 0$) , σ_r^2 ($\sigma_r^2 \geq 0$), σ_g^2 ($\sigma_g^2 \geq 0$) and ρ ($-1 \leqslant \rho \leqslant 1$) are estimated using the local neighboring coefficients. To examine the appropriateness of this PDF for the index dependent random variables X_r and X_g, Mardia's test (Mardia, 1970, 1974) is performed using a set of local neighboring magnitude components centered at that index. Level of significance used for this test to reject the null hypothesis of bivariate normality is 5 percent. In addition, graphical assessments for bivariate normality are also made by constructing chi-square plots (Kim and Timm, 2006) using the same data. The results concerning the Mardia's test using a 7×7 window as the local neighboring region for each of the components are given in Table 4.3. This table provides the percentage of the magnitude components for each subband that follow the bivariate Gaussian PDF by averaging over the same set of images that were used for Table 4.1. It can be seen from Table 4.3 that a significant proportion (more than 85%) of the random variables X_r and X_g in each subband follow the bivariate Gaussian PDF in $l = 1$, 2, and 4 as well as a sufficient proportion (more than 70%) in $l = 3$. Fig. 4.2 shows the average of the chi-square plots obtained for all the random variables X_r and X_g in the subbands $HL1_1$, $HL1_2$, $HL1_3$, and $HL1_4$ using the same settings that are used for Mardia's test. It is seen that these plots closely resemble the 45^o reference line. The chi-square plots for the other subbands are not shown since they are very similar to those in Fig. 4.2 The findings of the chi-square plots thus reinforce the conclusion that the magnitude components closely follow the bivariate Gaussian PDF.

The results of the Mardia's test given in Table 4.3 and that of the chi-square plots

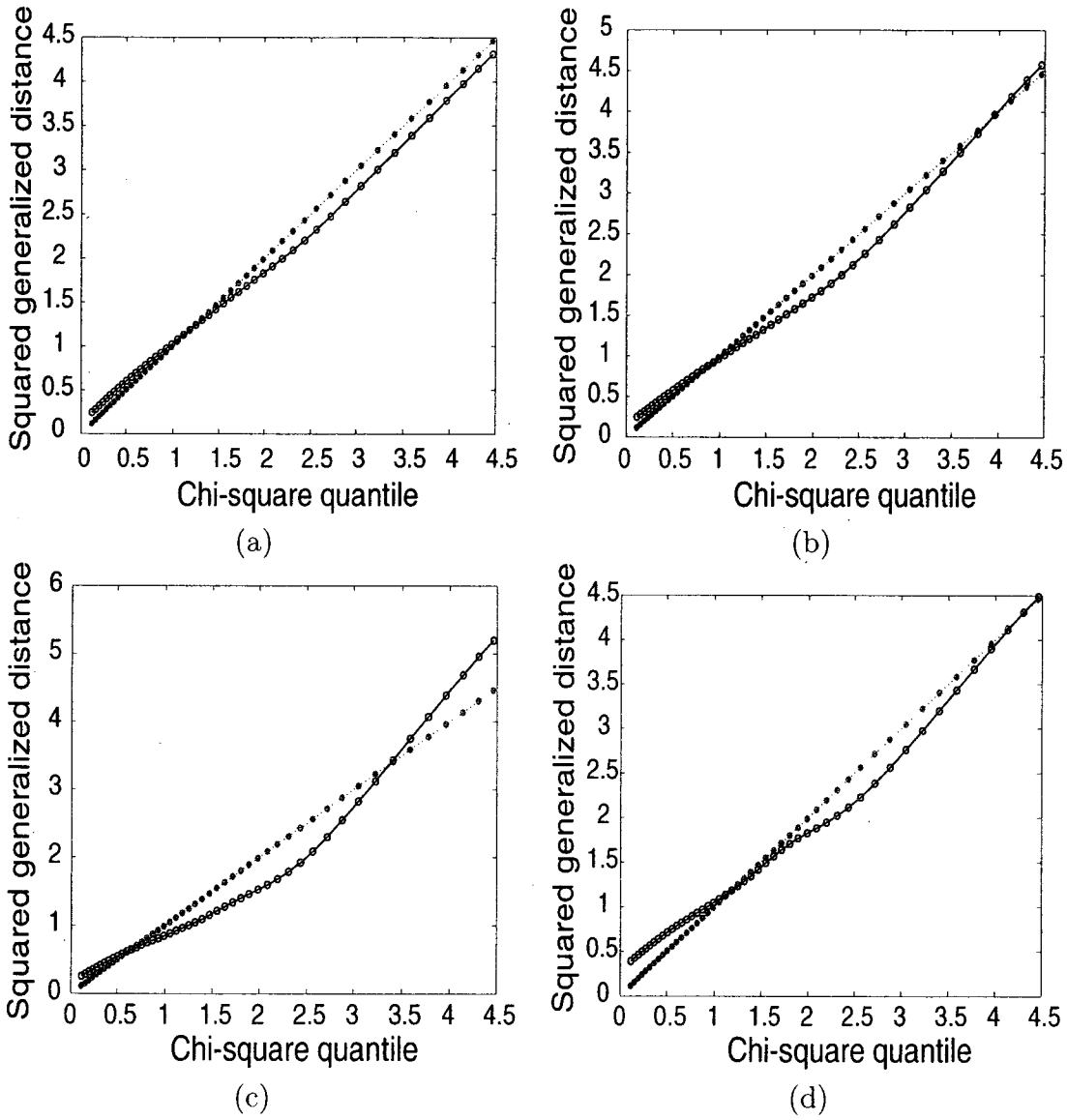


Figure 4.2: Chi-square plots to assess bivariate normality of X_r and X_g in the subbands (a) $HL1_1$ (b) $HL1_2$ (c) $HL1_3$ and (d) $HL1_4$.

in Fig. 4.2 show that the proportion of the random variables X_r and X_g that follow the bivariate Gaussian PDF is the least in $l = 3$. However, a choice of bivariate Gaussian PDF for this level does not degrade overall denoising performance significantly. This is due to the fact that the wavelet coefficients in the lower level subbands are dominated by noise unlike that in the higher levels (Mallat, 1999; Rahman and Hasan, 2003). As a consequence, overall denoising performance is highly dependent on the efficiency of noise removal from the subbands at the lower levels. In other words, it is more important to choose a joint PDF that provides a better fit to the magnitude components of the subbands in a lower level to achieve a satisfactory denoising performance. Hence, for the purpose of denoising, the bivariate Gaussian PDF is an appropriate choice. In the next section, bivariate LMMSE and MAP estimators are derived using this PDF as a joint prior function.

4.3.1 Bivariate LMMSE Estimator

Let $\boldsymbol{x} = [x_r, x_g]^T$, $\boldsymbol{y} = [y_r, y_g]^T$, and $\boldsymbol{v} = [v_r, v_g]^T$ be the samples of the random vectors $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{V}$ representing magnitude components of the noise-free coefficients, noisy coefficients and noise-coefficients, respectively. To find the LMMSE estimator for $\boldsymbol{x}$, we use the formula given in (3.13), which requires knowledge of the mean vectors $\mu_{\boldsymbol{X}}$ and $\mu_{\boldsymbol{Y}}$, the covariance matrix $\Sigma_{\boldsymbol{Y}}$, and the cross-covariance matrix $\Sigma_{\boldsymbol{XY}}$. Since the image and additive noise are independent, it can be shown that $\mu_{\boldsymbol{Y}} = \mu_{\boldsymbol{X}} + \mu_{\boldsymbol{V}}$ and $\Sigma_{\boldsymbol{XY}} = \Sigma_{\boldsymbol{X}}$, where $\mu_{\boldsymbol{V}}$ is the mean vector of $\boldsymbol{V}$, $\mu_{\boldsymbol{X}}^T = [\mu_r, \mu_g]$ is the mean vector of $\boldsymbol{X}$, and $\Sigma_{\boldsymbol{X}}$ is the covariance matrix of $\boldsymbol{X}$ given by

$$\Sigma_{\boldsymbol{X}} = \begin{bmatrix} \sigma_r^2 & \rho\sigma_r\sigma_g \\ \rho\sigma_r\sigma_g & \sigma_g^2 \end{bmatrix} \tag{4.4}$$

where $\boldsymbol{X}$ follows a bivariate Gaussian PDF. The covariance matrix of $\boldsymbol{Y}$ is found to be $\Sigma_{\boldsymbol{Y}} = \Sigma_{\boldsymbol{X}} + \Sigma_{\boldsymbol{V}}$, where $\Sigma_{\boldsymbol{V}}$ is the covariance matrix of the joint PDF of V_r and

V_g. Thus, the LMMSE estimator given in (3.13) becomes

$$\hat{\boldsymbol{x}} = \mu_{\boldsymbol{X}} + \Sigma_{\boldsymbol{X}}(\Sigma_{\boldsymbol{X}} + \Sigma_{\boldsymbol{V}})^{-1}(\boldsymbol{y} - \mu_{\boldsymbol{X}} - \mu_{\boldsymbol{V}}) \tag{4.5}$$

To find the elements of $\Sigma_{\boldsymbol{V}}$, the characteristics of noise in the CWT-domain are to be determined. It is known that when a two-dimensional vector has elements that are zero-mean normally distributed, are uncorrelated, and have equal variance, then the magnitude of the vector will have a Rayleigh distribution (Simon, 2002). Since the real and imaginary components of the CWT coefficients of Gaussian noise in an image follow the i.i.d. zero-mean Gaussian distribution (Shaffrey et al., 2002), the PDF of the magnitude components V_r or V_g will follow the univariate Rayleigh PDF. Further, it can be shown that the real or imaginary components of the CWT coefficients of additive bivariate Gaussian noise in the red and green channel images follow the i.i.d. zero-mean bivariate Gaussian distribution having correlation coefficient ρ_ε and variance $\mathrm{tr}\{\boldsymbol{W}\boldsymbol{W}^T\}^l\sigma_\varepsilon^2$, where $\boldsymbol{W}$ is the transformation matrix of CWT. Hence, from the definition of a bivariate Rayleigh random variable, the magnitude components of the CWT coefficients of additive bivariate Gaussian noise in microarray images follow the i.i.d. bivariate Rayleigh distribution (Simon, 2002) having the parameters $\sigma_v^2 = \mathrm{tr}\{\boldsymbol{W}\boldsymbol{W}^T\}^l\sigma_\varepsilon^2$ and $\rho_v = \rho_\varepsilon$. Using these parameters the joint PDF of V_r and V_g may be written as (Simon, 2002)

$$p_{\boldsymbol{V}}(\underline{\boldsymbol{v}}) = \frac{v_r v_g}{(1-\rho_v^2)\sigma_v^4} \cdot \exp\left(-\frac{v_r^2 + v_g^2}{2(1-\rho_v^2)\sigma_v^2}\right) I_0\left(\frac{v_r v_g |\rho_v|}{\sigma_v^2(1-\rho_v^2)}\right) \tag{4.6}$$

where $I_\vartheta(\cdot)$ denotes the modified Bessel function of order ϑ (Abramowitz and Stegun, 1965). Since the marginal PDF of V_r or V_g is the univariate Rayleigh PDF, the mean and variance of the random variables is given by $\mu_n = \sigma_v\sqrt{\frac{\pi}{2}}$ and $\sigma_n^2 = \sigma_v^2(\frac{4-\pi}{2})$, respectively. Thus, the mean vector of $\boldsymbol{V}$ is $\mu_{\boldsymbol{V}}^T = [\sigma_v\sqrt{\frac{\pi}{2}}, \sigma_v\sqrt{\frac{\pi}{2}}]$ and the covariance matrix is

$$\Sigma_{\boldsymbol{V}} = \sigma_v^2 \begin{bmatrix} \frac{4-\pi}{2} & \frac{\pi}{2}\varphi(\rho_v) \\ \frac{\pi}{2}\varphi(\rho_v) & \frac{4-\pi}{2} \end{bmatrix} \tag{4.7}$$

where $\varphi(\rho_v) = (1-\rho_v^2)^3 {}_2F_1(\frac{3}{2}, \frac{3}{2}; 1, \rho_v^2) - 1$ and ${}_2F_1(\alpha, \beta; \gamma; t)$ is the Gaussian hypergeometric function (Abramowitz and Stegun, 1965). Using the expressions of $\mu_{\boldsymbol{X}}$, $\mu_{\boldsymbol{V}}$, $\Sigma_{\boldsymbol{X}}$, $\Sigma_{\boldsymbol{V}}$ in (4.5) and performing the matrix operations, the magnitude components of the CWT coefficients of red and green channel images can be obtained as

$$\hat{x}_r\Big|[y_r, y_g] = \mu_r + \frac{\sigma_r^2\sigma_1^2 - \omega_1}{\sigma_r^2\sigma_1^2 + \sigma_v^2(\frac{4-\pi}{2})\Big[\sigma_g^2 + \sigma_v^2\big(\frac{4-\pi}{2}\big)\Big] - \omega_2}\Bigg[y_r - \mu_r - \sigma_v\sqrt{\frac{\pi}{2}} + \frac{\sigma_r\sigma_v^2\omega_3}{\sigma_r^2\sigma_1^2 - \omega_1}\big(y_g - \mu_g - \sigma_v\sqrt{\frac{\pi}{2}}\big)\Bigg] \tag{4.8}$$

$$\hat{x}_g\Big|[y_r, y_g] = \mu_g + \frac{\sigma_g^2\sigma_2^2 - \omega_1}{\sigma_g^2\sigma_2^2 + \sigma_v^2(\frac{4-\pi}{2})\Big[\sigma_r^2 + \sigma_v^2\big(\frac{4-\pi}{2}\big)\Big] - \omega_2}\Bigg[y_g - \mu_g - \sigma_v\sqrt{\frac{\pi}{2}} + \frac{\sigma_g\sigma_v^2\omega_4}{\sigma_g^2\sigma_2^2 - \omega_1}\big(y_r - \mu_r - \sigma_v\sqrt{\frac{\pi}{2}}\big)\Bigg] \tag{4.9}$$

where

$$\omega_1 = \frac{\pi}{2}\rho\varphi(\rho_v)\sigma_r\sigma_g\sigma_v^2 \tag{4.10}$$

$$\omega_2 = \frac{\pi^2}{4}\sigma_v^4\varphi^2(\rho_v) + 2\omega_1 \tag{4.11}$$

$$\omega_3 = \rho\sigma_g\Big(\frac{4-\pi}{2}\Big) - \frac{\pi}{2}\sigma_r\varphi(\rho_v) \tag{4.12}$$

$$\omega_4 = \rho\sigma_r\Big(\frac{4-\pi}{2}\Big) - \frac{\pi}{2}\sigma_g\varphi(\rho_v) \tag{4.13}$$

$$\sigma_1^2 = [1-\rho^2]\sigma_g^2 + \sigma_v^2\Big(\frac{4-\pi}{2}\Big) \tag{4.14}$$

$$\sigma_2^2 = [1-\rho^2]\sigma_r^2 + \sigma_v^2\Big(\frac{4-\pi}{2}\Big) \tag{4.15}$$

4.3.2 Bivariate MAP Estimator

The MAP estimator for $\boldsymbol{x}$ given the corrupted observation $\boldsymbol{y}$ is (Kamen and Su, 1999)

$$\begin{aligned}\hat{\boldsymbol{x}}(\boldsymbol{y}) &= \arg\max_{\boldsymbol{x}} p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \\ &= \arg\max_{\boldsymbol{x}} [\ln(p_{\boldsymbol{V}}(\boldsymbol{y}-\boldsymbol{x})) + \ln(p_{\boldsymbol{X}}(\boldsymbol{x}))]\end{aligned} \tag{4.16}$$

where $p_{\boldsymbol{X}}(\cdot)$ is the bivariate Gaussian PDF and $p_{\boldsymbol{V}}(\cdot)$ is the bivariate Rayleigh PDF given in (4.6).Then the bivariate MAP estimator becomes

$$[\hat{x}_r, \hat{x}_g] = \arg\max_{x_r, x_g} \left[\ln(y_r - x_r) + \ln(y_g - x_g) - \frac{(y_r - x_r)^2}{2(1-\rho_v^2)\sigma_v^2} - \frac{(y_g - x_g)^2}{2(1-\rho_v^2)\sigma_v^2} + \ln I_0\left(\frac{(y_r - x_r)(y_g - x_g)|\rho_v|}{\sigma_v^2(1-\rho_v^2)}\right) + \ln\left(p_{X_r X_g}(x_r, x_g)\right) \right] \quad (4.17)$$

Let Z denote the expression within square brackets on the right hand side of (4.17). To find the values for x_r and x_g that maximize Z, it is necessary to solve the pair of equations $\frac{\partial Z}{\partial x_r} = 0$ and $\frac{\partial Z}{\partial x_g} = 0$, where

$$\frac{\partial Z}{\partial x_r} = -\frac{x_r - \mu_r}{(1-\rho^2)\sigma_r^2} + \frac{\rho}{1-\rho^2}\frac{(x_g - \mu_g)}{\sigma_r \sigma_g} - \frac{1}{y_r - x_r} + \frac{y_r - x_r}{(1-\rho_v^2)\sigma_v^2} - \frac{I_1(u)}{I_0(u)}\frac{(y_g - x_g)|\rho_v|}{\sigma_v^2(1-\rho_v^2)} \quad (4.18)$$

$$\frac{\partial Z}{\partial x_g} = -\frac{x_g - \mu_g}{(1-\rho^2)\sigma_g^2} + \frac{\rho}{1-\rho^2}\frac{(x_r - \mu_r)}{\sigma_r \sigma_g} - \frac{1}{y_g - x_g} + \frac{y_g - x_g}{(1-\rho_v^2)\sigma_v^2} - \frac{I_1(u)}{I_0(u)}\frac{(y_r - x_r)|\rho_v|}{\sigma_v^2(1-\rho_v^2)} \quad (4.19)$$

and $u = \frac{(y_r - x_r)(y_g - x_g)|\rho_v|}{\sigma_v^2(1-\rho_v^2)}$. These nonlinear equations may be solved using the following numerical iterative procedure. Let $\hat{\boldsymbol{x}}^{(0)}$ be the initial value of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}^{(\nu)}$ be its value after the ν-th iteration. The estimate is updated at each iteration using the relation $\hat{\boldsymbol{x}}^{(i+1)} = \hat{\boldsymbol{x}}^{(i)} - \delta \boldsymbol{H}^{-1}\boldsymbol{z}$, where $\delta (0 < \delta \leq 1)$ is a constant that controls the step size, $\boldsymbol{z}^T = \left[\frac{\partial Z}{\partial x_r}, \frac{\partial Z}{\partial x_g}\right]\Big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}^{(\nu)}}$ and $\boldsymbol{H}$ represents the Hessian matrix (Eliason, 2000) evaluated at $\boldsymbol{x} = \hat{\boldsymbol{x}}^{(\nu)}$. The elements of $\boldsymbol{H}$ may be obtained from the following relations

$$\frac{\partial^2 Z}{\partial x_r^2} = -\frac{1}{(1-\rho^2)\sigma_r^2} - \frac{1}{(y_r - x_r)^2} - \frac{1}{(1-\rho_v^2)\sigma_v^2} - \frac{(y_g - x_g)|\rho_v|}{\sigma_v^2(1-\rho_v^2)}\left[\frac{I_1'^2(u)}{I_0^2(u)} - \frac{I_1'(u)}{I_0(u)}\right] \quad (4.20)$$

$$\frac{\partial^2 Z}{\partial x_r \partial x_g} = \frac{\partial^2 Z}{\partial x_g \partial x_r} = \frac{\rho}{(1-\rho^2)\sigma_r\sigma_g} + \frac{I_1(u)}{I_0(u)}\frac{|\rho_v|}{\sigma_v^2(1-\rho_v^2)} - \frac{(y_g - x_g)|\rho_v|}{\sigma_v^2(1-\rho_v^2)}\left[\frac{I_1'^2(u)}{I_0^2(u)} - \frac{I_1'(u)}{I_0(u)}\right] \quad (4.21)$$

$$\frac{\partial^2 Z}{\partial x_g^2} = -\frac{1}{(1-\rho^2)\sigma_g^2} - \frac{1}{(y_g - x_g)^2} - \frac{1}{(1-\rho_v^2)\sigma_v^2} - \frac{(y_r - x_r)|\rho_v|}{\sigma_v^2(1-\rho_v^2)}\left[\frac{I_1'^2(u)}{I_0^2(u)} - \frac{I_1'(u)}{I_0(u)}\right] \quad (4.22)$$

where $I_1'(u) = \frac{1}{2}[I_0(u) + I_2(u)]$. Convergence is attained when $||\hat{\boldsymbol{x}}^{(\nu+1)} - \hat{\boldsymbol{x}}^{(\nu)}|| \leq \epsilon$ where $||\cdot||$ denotes the Euclidean norm and ϵ is a very small quantity. It is desirable

1. Input: $g_r(i,j)$, $g_g(i,j)$, $i = 1, \cdots N_1$, $j = 1, \cdots N_2$
2. Forward transform: Obtain noisy CWT coefficients $y_r^r \angle y_r^\theta$ and $y_g^r \angle y_g^\theta$
3. Estimation of CWT coefficients using the local parameters:
   for $l = 1 : J$ **do**
      for $O = 1 : 6$ ($1 \Rightarrow HL1, 2 \Rightarrow HH1, 3 \Rightarrow LH1, 4 \Rightarrow HL2,$
      $5 \Rightarrow HH2, 6 \Rightarrow LH2$) **do**
         $\forall k$ in a subband,
         Compute $\hat{\mu}_r^2(k)$ (refer to (4.25))
         Compute $\hat{\mu}_g^2(k)$ (refer to (4.26))
         Compute $\hat{\sigma}_r^2(k)$ (refer to (4.27))
         Compute $\hat{\sigma}_g^2(k)$ (refer to (4.28))
         Compute $\hat{\rho}(k)$ (refer to (4.29))
         Compute $\hat{x}_r(k)$ and $\hat{x}_g(k)$ (refer to (4.8) and (4.9))
      end for
   end for
4. Inverse transform: Obtain denoised images using $\hat{x}_r^r \angle y_r^\theta$ and $\hat{x}_g^r \angle y_g^\theta$

Figure 4.3: Steps in the LMMSE-based algorithm.

that the iterative algorithm use a good initial estimate $\hat{\boldsymbol{x}}^{(0)}$. For this reason, we have chosen a scenario assuming that no correlation exists between the image coefficients as well as the noise coefficients of the two channels. In such a case, the following initial values for x_r and x_g are obtained (Howlader and Chaubey, 2008)

$$\hat{x}_r^{(0)} = \max\left(\frac{1}{2(\sigma_r^2+\sigma_v^2)}\left((2\sigma_r^2+\sigma_v^2)y_r + \mu_r\sigma_v^2 - \sqrt{\sigma_v^4(y_r-\mu_r)^2 + 4\sigma_r^4\sigma_v^2 + 4\sigma_r^2\sigma_v^4}\right), 0\right) \tag{4.23}$$

$$\hat{x}_g^{(0)} = \max\left(\frac{1}{2(\sigma_g^2+\sigma_v^2)}\left((2\sigma_g^2+\sigma_v^2)y_g + \mu_g\sigma_v^2 - \sqrt{\sigma_v^4(y_g-\mu_g)^2 + 4\sigma_g^4\sigma_v^2 + 4\sigma_g^2\sigma_v^4}\right), 0\right) \tag{4.24}$$

1. Input: $g_r(i,j)$, $g_g(i,j)$, $i = 1, \cdots N_1$, $j = 1, \cdots N_2$
2. Forward transform: Obtain noisy CWT coefficients $y_r^{\mathrm{r}} \angle y_r^{\theta}$ and $y_g^{\mathrm{r}} \angle y_g^{\theta}$
3. Estimation of CWT coefficients using the local parameters:
   for $l = 1 : J$ **do**
   for $O = 1 : 6$ ($1 \Rightarrow HL1, 2 \Rightarrow HH1, 3 \Rightarrow LH1, 4 \Rightarrow HL2,$
   $5 \Rightarrow HH2, 6 \Rightarrow LH2$) **do**
   $\forall k$ **in a subband,**
   Compute $\hat{\mu}_r^2(k)$ (refer to (4.25))
   Compute $\hat{\mu}_g^2(k)$ (refer to (4.26))
   Compute $\hat{\sigma}_r^2(k)$ (refer to (4.27))
   Compute $\hat{\sigma}_g^2(k)$ (refer to (4.28))
   Compute $\hat{\rho}(k)$ (refer to (4.29))
   Initialize $\hat{\boldsymbol{x}}$ by $\hat{x}_r^{(0)}$ and $\hat{x}_g^{(0)}$ (refer to (4.23) and (4.24))
   Set $\nu = 0$
   while $||\hat{\boldsymbol{x}}^{(\nu+1)} - \hat{\boldsymbol{x}}^{(\nu)}|| > \epsilon$
   Update $\nu = \nu + 1$
   Compute $\boldsymbol{z}$ (refer to (4.18) and (4.19))
   Compute $\boldsymbol{H}$ (refer to (4.20), (4.21) and (4.22))
   Update $\hat{\boldsymbol{x}}^{(\nu+1)} = \hat{\boldsymbol{x}}^{(\nu)} - \boldsymbol{H}^{-1}\boldsymbol{z}$
   end while
   $\hat{x}_r^{\mathrm{r}} = \hat{x}_r^{(\nu+1)}$ and $\hat{x}_g^{\mathrm{r}} = \hat{x}_g^{(\nu+1)}$
   end for
   end for
4. Inverse transform: Obtain denoised images using $\hat{x}_r^{\mathrm{r}} \angle y_r^{\theta}$ and $\hat{x}_g^{\mathrm{r}} \angle y_g^{\theta}$

Figure 4.4: Steps in the MAP-based algorithm.

4.3.3 Parameter Estimation

Both the bivariate LMMSE and MAP estimators require the parameters μ_r, μ_g, σ_r^2, σ_g^2, and ρ to be estimated from the observed noisy magnitude components of the CWT coefficients of red and green channel images. We use the locally-adaptive approach described in Section 3.4.3, in which, parameters at the k-th index are estimated from the components of a local neighborhood $S(k)$ assuming that coefficients within this neighborhood are i.i.d. Here, k represents the two-dimensional spatial index (k_1, k_2).

Then the parameters may be estimated using the ML method as (Giri, 1993)

$$\hat{\mu}_r(k) = \max\left(\frac{1}{M}\sum_{m\in\mathcal{S}(k)} y_r(m) - \mu_n, 0\right) \tag{4.25}$$

$$\hat{\mu}_g(k) = \max\left(\frac{1}{M}\sum_{m\in\mathcal{S}(k)} y_g(m) - \mu_n, 0\right) \tag{4.26}$$

$$\hat{\sigma}_r^2(k) = \max\left(\frac{1}{M}\sum_{m\in\mathcal{S}(k)} (y_r(m) - \bar{y}_r)^2 - \sigma_n^2, 0\right) \tag{4.27}$$

$$\hat{\sigma}_g^2(k) = \max\left(\frac{1}{M}\sum_{m\in\mathcal{S}(k)} (y_g(m) - \bar{y}_g)^2 - \sigma_n^2, 0\right) \tag{4.28}$$

$$\hat{\rho}(k) = \max\left(\min\left(\frac{1}{\hat{\sigma}_r(k)\hat{\sigma}_g(k)}\frac{1}{M}\sum_{m\in\mathcal{S}(k)} (y_r(m) - \bar{y}_r)(y_g(m) - \bar{y}_g), 1\right), -1\right) \tag{4.29}$$

where M is the total number of coefficients in $\mathcal{S}(k)$, and $\bar{y}_r$ and $\bar{y}_g$ are the sample-means of the noisy magnitude components of the red and green channel images. The algorithms for the proposed LMMSE- and MAP-based methods are given in Fig. 4.3 and Fig. 4.4, respectively. Further, the major steps involved in implementing these proposed algorithms are summarized using block diagrams in Fig. 4.5 and Fig. 4.6.

We derive the computational complexities of both algorithms as follows. The upper bound of the computational complexity of the forward or inverse CWT may be found as $\mathcal{O}(\frac{32}{3}\mathcal{L}(N_1N_2))$, where $N_1 \times N_2$ is the size of an image. The proposed LMMSE-based method estimates 5 local parameters for each of the magnitude components of the CWT. Thus, the computational complexity of the parameter estimation of the Proposed LMMSE method is $\mathcal{O}(10w^2N_1N_2)$. For this method, the complexity of estimating the noise-free coefficients is $\mathcal{O}(8N_1N_2)$. Since $\mathcal{L} \gg 1$ and $w^2 \gg 1$, the overall computational complexity of the Proposed LMMSE method can be found as $\mathcal{O}((10w^2 + \frac{128}{3}\mathcal{L})N_1N_2)$. In the case of MAP-based method, each of the denoised

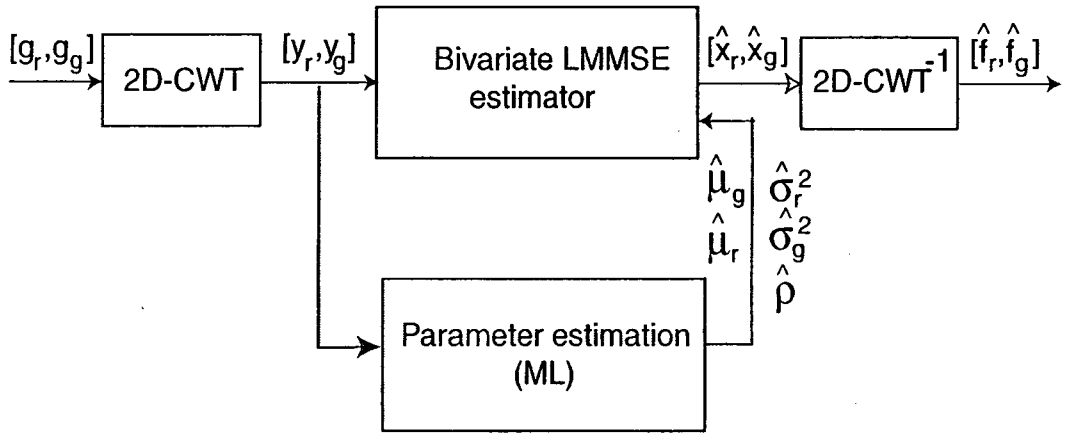


Figure 4.5: Block diagram for obtaining denoised cDNA microarray image using the proposed bivariate LMMSE estimator.

coefficients are updated using an iterative process, wherein each update requires an inverse of a 2×2 Hessian matrix. The computational complexity of the parameter estimation is the same as that for the Proposed LMMSE method. Thus, the combined complexity of the parameter estimation and coefficient estimation of the MAP-based method is $\mathcal{O}((10w^2 + 6C_B\mathcal{I})N_1N_2)$, where C_B ($C_B \gg 1$) is the complexity of the modified Bessel function and $\mathcal{I}$ ($\mathcal{I} \gg 1$) is the number of iterations required for convergence. By including the complexities of the forward and inverse of CWT, the overall computational complexity of the Proposed MAP method is found as $\mathcal{O}((10w^2 + 6C_B\mathcal{I} + \frac{128}{3}\mathcal{L})N_1N_2)$.

4.4 Estimation of Log-Intensity Ratio

In Section 3.5, we described the steps used to extract the log-intensity ratios from the microarray images. Among these steps, a crucial step for the estimation of the log-intensity ratio is segmentation. In Chapter 3, a histogram segmentation method was used for identifying a spot and its local background. The method was based entirely on the distribution of the pixel intensities within the target area and did not use any local spatial information. In this chapter, we propose a more accurate

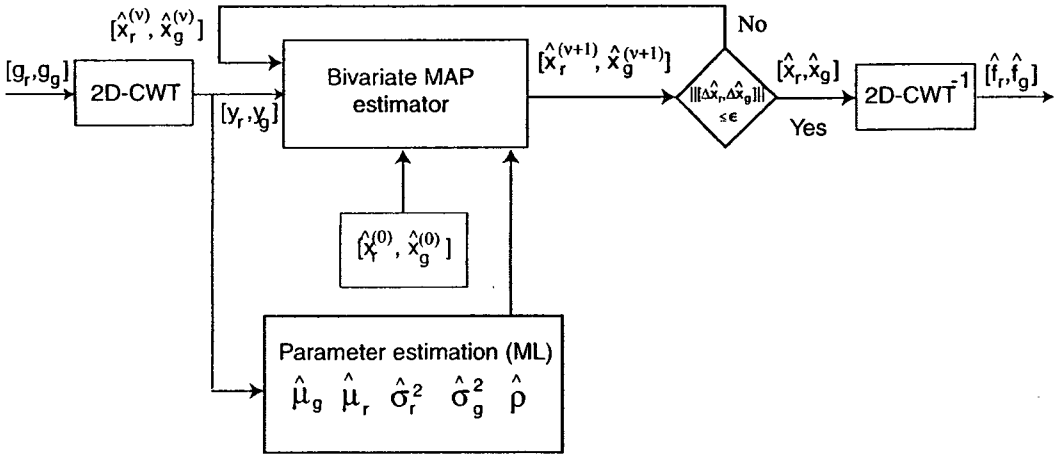
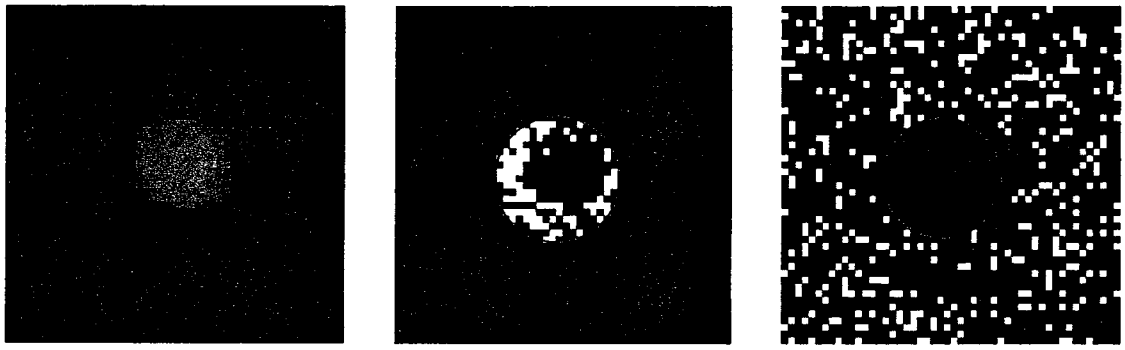


Figure 4.6: Block diagram for obtaining denoised cDNA microarray image using the proposed bivariate MAP estimator.

method of segmentation. We use an adaptive segmentation technique that is a hybrid of the fixed circle-based (Bozinov and Rahnenfuhrer, 2002) and histogram-based (McLachlan et al., 2004) segmentation methods, in order to benefit from the advantages of both these methods. Segmentation is performed separately for the red and green channel images in the following manner. First, the center of mass is located from the pixels within a target area. Then using this point as the center, a circle is drawn having a fixed radius that depends on the size of the pin tip of the robotic arrayer. We define the *ROI* as those pixels within the circle having intensities between the 60-th and 95-th percentiles, and the background as the pixels outside the circle having intensities between the 5-th and 20-th percentiles. It may be noted that these percentiles are calculated based on all the pixels within the target area.

The hybrid segmentation method has the following two notable features that make it superior to the traditional fixed circle and histogram-based segmentation methods. First, the *ROI* selected by this method does not assume any perfectly circular shape at the center of target area as in the case of fixed circle-based method. Thus, the hybrid method provides a better separation of foreground when the spots have varying radii, irregular shapes, or spatial offsets from the center of the target area. Secondly,



(a) (b) (c)

Figure 4.7: Selection of the spot and background pixels from the observed image for calculating the log-intensity ratio. (a) A target area. (b) Pixels selected for the spot are shown as white. (c) Pixels selected for the background are shown as white.

in selecting the pixels for a spot and its corresponding background, the hybrid segmentation method considers the spatial contexts that is ignored in the histogram-based method. Fig. 4.7 shows an example of such selections of the pixels for a spot and its corresponding background in a target area using the hybrid segmentation method. It may be seen from this figure that the method is capable of detecting the irregular shapes of *ROI* and background considering the spatial contexts.

4.5 Experimental Results

Extensive experimentations are carried out to evaluate the performance of the proposed bivariate LMMSE and MAP-based denoising algorithms as compared to that of the other CWT-based algorithms. We have compared only CWT-based algorithms, since methods implemented in the CWT-domain, in general, yield a better performance than those implemented in the DWT-domain (Kingsbury, 1999, 2001; Chaux et al., 2006). The performance of the noise reduction algorithms are compared both in terms of the PSNR and the MAE of the estimated log-intensity ratios. The experiments are conducted on the same set of microarray images that have been used to evaluate the performance of the DWT-based algorithms in Chapter 3. The com-

plex coefficients in the experiments are obtained by employing a 4-level 2D DT-CWT, wherein an (11,17)-tap biorthogonal filter in the first level of decomposition and 6-tap Q-shift orthogonal filters in the second- and higher-level decompositions (Kingsbury, 2001) are used.

Five denoising methods that use the CWT coefficients are considered in the experiments. These are the complex multiwavelet style (CMWS) method (Barber and Nason, 2004), NeighCoef (Cai and Silverman, 2001), BiShrink (Sendur and Selesnick, 2002a), and the proposed bivariate LMMSE and MAP-based denoising methods. Except for the proposed methods, each of the other methods estimate the red and green channel images independently, i.e., ignoring the correlation between the two channels. The CMWS is a subband-adaptive method, while the remaining methods are locally-adaptive. It may be mentioned that NeighCoef, which was originally proposed for denoising the DWT coefficients, has been adapted for implementation in the CWT-domain. The NeighCoef has been chosen since it is one of the competing DWT-based methods. The locally-adaptive methods are implemented using the same settings for the window size mentioned in Section 3.6.

Table 4.4 shows the output PSNR values for the five denoising methods obtained by averaging over the same set of images that have been used for Table 3.4. Since the proposed methods are capable of considering the inter-channel noise correlation, the PSNR values are obtained for different values of ρ_ε at each noise strength σ_ε. The average of these PSNRs at a given σ_ε is compared with the corresponding output PSNRs of the CMWS, NeighCoef and BiShrink methods that are incapable of considering the inter-channel noise correlation. The highest PSNRs corresponding to each σ_ε are highlighted in bold face. It can be seen from Table 4.4 that the proposed bivariate MAP-based denoising method outperforms the others. The average output PSNR values of the proposed bivariate LMMSE-based denoising method also exceeds that

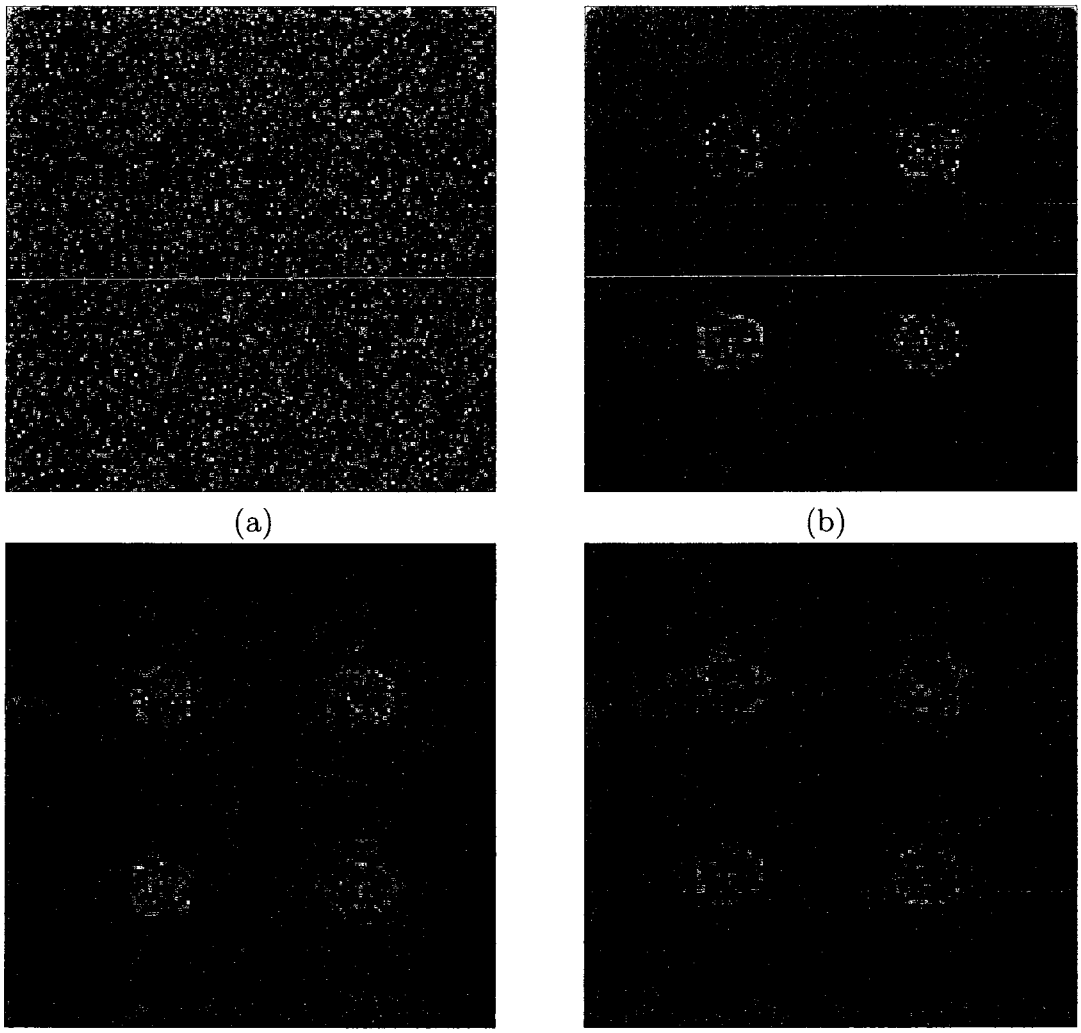
Table 4.4: Output PSNR values in dB for the CWT-based denoising algorithms at various noise strengths.

σ_ϵ	800	1000	1200	1400	1600
$\text{IPSNR}_g^\dagger$	38.29	36.35	34.77	33.43	32.27
$\text{IPSNR}_r^\dagger$	38.27	36.33	34.75	33.41	32.25
Methods					
CMWS	45.16	43.82	42.64	41.63	40.72
	45.86	44.38	42.12	42.00	41.06
NeighCoef	45.62	44.33	43.20	42.23	41.40
	46.40	44.98	43.80	42.80	41.97
BiShrink	45.54	44.29	43.23	42.29	41.50
	46.26	44.95	43.84	42.92	42.09
LMMSE					
$\rho_\epsilon = 0.00$	45.78	44.57	43.53	42.63	41.82
	46.69	45.38	44.27	43.31	42.47
$\rho_\epsilon = 0.25$	45.77	44.56	43.51	42.61	41.80
	46.67	45.36	44.25	43.29	42.45
$\rho_\epsilon = 0.50$	45.72	44.50	43.45	42.54	41.73
	46.60	45.28	44.17	43.22	42.37
$\rho_\epsilon = 0.75$	45.57	44.32	43.25	42.32	41.49
	46.36	45.01	43.87	42.88	42.02
Average	(45.71)	(44.49)	(43.44)	(42.53)	(41.71)
	(46.58)	(45.26)	(44.14)	(43.18)	(42.33)
MAP					
$\rho_\epsilon = 0.00$	46.15	45.00	43.98	43.08	42.27
	47.15	45.91	44.81	43.85	43.01
$\rho_\epsilon = 0.25$	46.12	44.96	43.95	43.05	42.25
	47.12	45.87	44.79	43.84	43.01
$\rho_\epsilon = 0.50$	46.06	44.89	43.87	42.98	42.19
	47.01	45.74	44.66	43.74	42.94
$\rho_\epsilon = 0.75$	45.94	44.76	43.73	42.85	42.07
	46.83	45.54	44.46	43.54	42.73
Average	**(46.07)**	**(44.90)**	**(43.88)**	**(42.99)**	**(42.20)**
	(47.03)	**(45.77)**	**(44.68)**	**(43.74)**	**(42.92)**

† IPSNR_g and IPSNR_r denote the input PSNRs for the green and red channel images, respectively.

of the NeighCoef, CMWS and BiShrink methods. A significant result that can be found from Table 4.4 is that the output PSNR of the BiShrink method, which considers the inter-scale dependency of the magnitude components of the microarray images, falls short of that of the proposed LMMSE- and MAP-based methods that consider the inter-channel dependency of the components by 0.25 dB and 0.73 dB, respectively. This result reinforces the findings of Chapter 3 that consideration of the inter-channel dependency is more important than the inter-scale one for denoising of cDNA microarray images. Another observation that can be made from this table is that with the increasing noise correlation between the two channels the PSNR declines both in the case of LMMSE- and MAP-based denoising methods. A similar relationship between noise correlation and output PSNR was observed in Table 3.6. However, it is evident from Table 4.4 that the PSNR values of the proposed methods do not decline much to fall short of the other methods. In other words, the proposed methods reduce noise better than the others do, even when the noise sequences of the two channels are correlated.

Residual noise images are shown in Fig. 4.8 for the NeighCoef, BiShrink and the proposed MAP method. The NeighCoef appears to be most efficient in reducing noise in the background. However, it is the least efficient in reducing noise from the spots as indicated by the larger number of bright pixels within the area of the spots. Comparing the residual noise images for the BiShrink and MAP methods, we see that the latter has a darker background and the spots within the image contain pixels that are less bright. Thus, among the competing methods, the proposed bivariate MAP method has better overall denoising performance. Fig. 4.9 shows a 3D visualization of a typical noise-free spot, its noisy version corrupted with a noise sequence having $\sigma_\varepsilon = 1400$ and $\rho_\varepsilon = 0.25$, and the corresponding denoised versions using the NeighCoef, BiShrink, proposed LMMSE and proposed bivariate MAP methods. This figure shows



(a) (b)

(c) (d)

Figure 4.8: Residual noise images using various denoising algorithms with $\sigma_\varepsilon = 1400$ and $\rho_\varepsilon = 0.25$. (a) Noisy image (b) NeighCoef (c) BiShrink, and (d) Proposed MAP method.

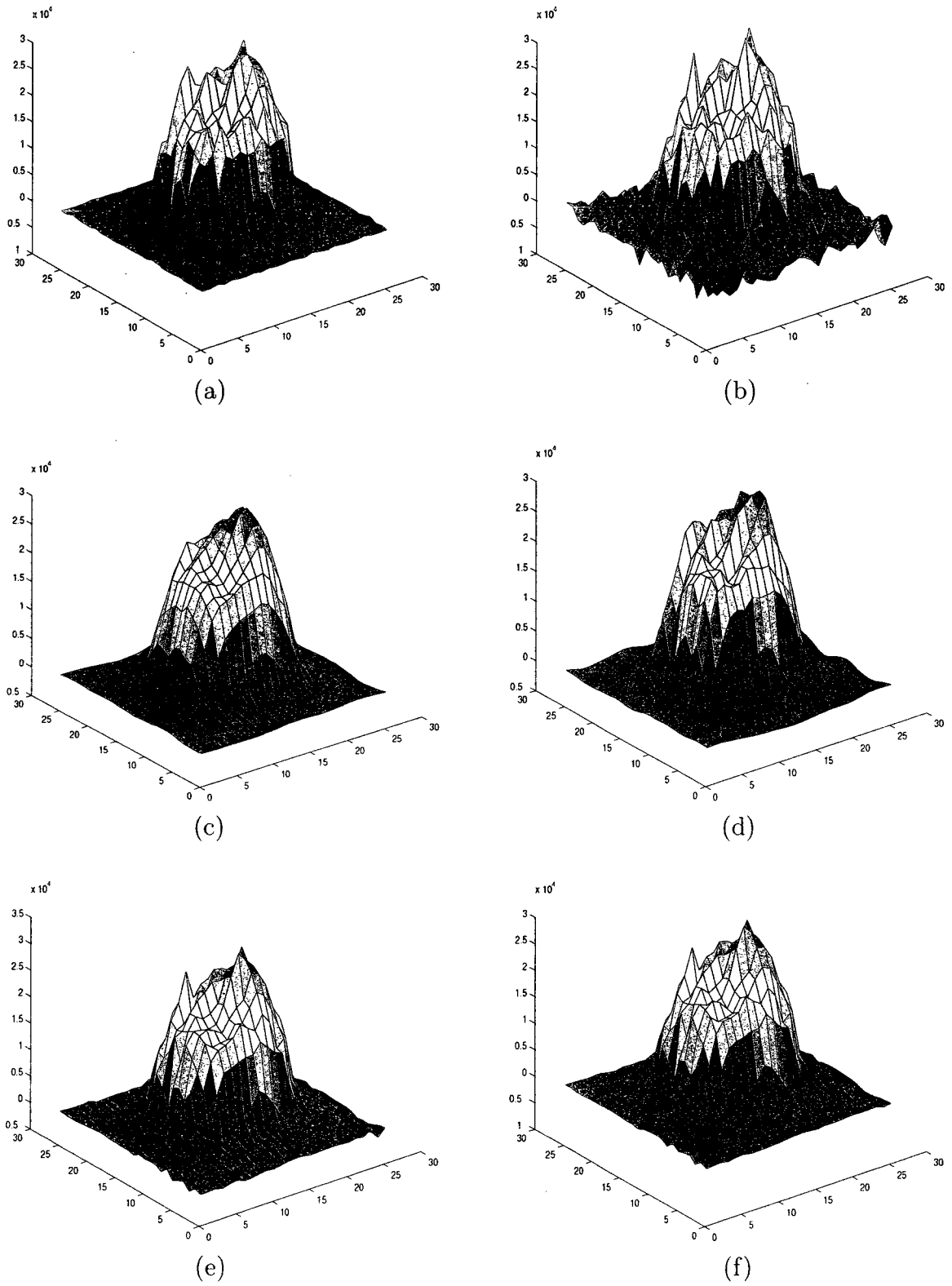


Figure 4.9: 3-D visualization of a spot. (a) Original and (b) noisy version of the spot with $\sigma_\varepsilon = 1400$ and $\rho_\varepsilon = 0.25$. Denoised versions of the spot using (c) NeighCoef (d) BiShrink (e) Proposed bivariate LMMSE method, and (f) Proposed bivariate MAP method.

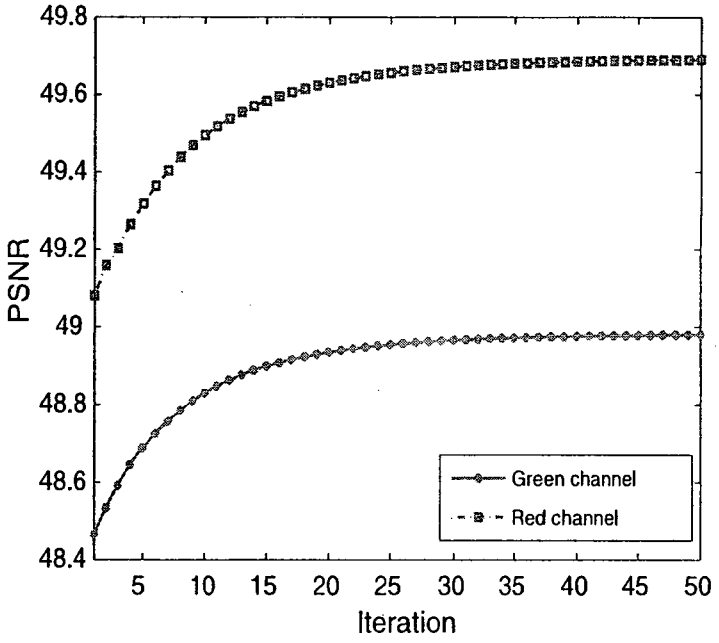


Figure 4.10: Number of iterations required for convergence of the proposed MAP-based algorithm in a given subband of the red and green channel images.

clearly that the spot is oversmoothed by the NeighCoef and BiShrink methods. In contrast, the proposed LMMSE and MAP methods preserve the significant signal intensities on the spot while removing most of the noise from the background.

To study the convergence properties of the MAP-based iterative algorithm, plots were obtained for the output PSNR versus iteration number for each subband of the red and green channel images. Figure 4.10 shows a typical plot obtained in a given subband. In most cases, the algorithm converges within 40 iterations. Moreover, significant improvement is seen in the final output PSNR values over the initial values.

Table 4.5 shows the output MAE values of the estimated log-intensity ratios obtained using the same denoised images that are used for Table 4.4. It can be seen that the output MAEs for a noise reduction scheme are lower than the corresponding input MAEs. In addition, among the methods considered in this experiment, the proposed bivariate MAP estimator provides the lowest average MAEs, i.e., the most accurate estimates of R. The bivariate LMMSE estimator provides higher MAEs as

Table 4.5: Output MAE values of log-intensity ratios for the CWT-based denoising algorithms at various noise strengths.

σ_ε	800	1000	1200	1400	1600
(Input MAE)	0.149	0.153	0.157	0.169	0.180
Methods					
CMWS	0.134	0.142	0.151	0.156	0.160
NeighCoef	0.140	0.149	0.153	0.161	0.175
BiShrink	0.110	0.118	0.133	0.133	0.131
LMMSE					
$\rho_\varepsilon = 0.00$	0.113	0.123	0.121	0.124	0.126
$\rho_\varepsilon = 0.25$	0.107	0.119	0.115	0.125	0.128
$\rho_\varepsilon = 0.50$	0.102	0.115	0.123	0.122	0.129
$\rho_\varepsilon = 0.75$	0.101	0.110	0.118	0.126	0.127
Average	(0.106)	(0.117)	(0.119)	(0.124)	(0.125)
MAP					
$\rho_\varepsilon = 0.00$	0.106	0.113	0.121	0.131	0.131
$\rho_\varepsilon = 0.25$	0.097	0.109	0.117	0.118	0.122
$\rho_\varepsilon = 0.50$	0.100	0.104	0.110	0.114	0.124
$\rho_\varepsilon = 0.75$	0.096	0.098	0.109	0.112	0.116
Average	**(0.100)**	**(0.106)**	**(0.114)**	**(0.119)**	**(0.123)**

compared to the MAP estimator, but these are still lower than those provided by the other methods.

As seen from Tables 4.4 and 4.5, the denoising performance measured in terms of PSNRs and MAEs of the estimated log-intensity ratio are better for the proposed bivariate LMMSE and MAP estimation methods than for the other methods. Among the two methods proposed, denoising performance of the MAP-based method are better than the LMMSE-based one, but the former being an iterative method requires a higher computation time than the latter. For example, the average time required to implement the proposed LMMSE- and MAP-based methods on an image of size 1000×1000 in the MATLAB environment using a 2 GHz processor with 512 MB RAM are 4.57 seconds and 3.86 minutes, respectively. Thus, one needs to weigh the importance of accuracy and computation time when choosing between the LMMSE- and MAP-based denoising methods.

Table 4.6: Performance of competitive DWT- and CWT-based methods in terms of PSNR and computational efficiency.

	$\text{APSNR}_g^\dagger$ $\text{APSNR}_r^\dagger$	Computational complexity	Computation time
DWT-based methods			
LAWMAP	42.72, 43.31	$\mathcal{O}((4w^2 + \frac{64}{3}\mathcal{L})N_1N_2)$	1.01 s
NeighCoef	43.26, 43.91	$\mathcal{O}((18 + \frac{64}{3}\mathcal{L})N_1N_2)$	0.93 s
BiShrink	43.18, 43.75	$\mathcal{O}((2w^2 + \frac{64}{3}\mathcal{L})N_1N_2)$	0.44 s
Proposed Method I	43.46, 44.17	$\mathcal{O}((5w^2 + \frac{64}{3}\mathcal{L})N_1N_2)$	1.94 s
Proposed Method II	43.56, 44.45	$\mathcal{O}((5w^2 + \frac{64}{3}\mathcal{L})N_1N_2)$	1.71 s
CWT-based methods			
NeighCoef	43.36, 43.99	$\mathcal{O}((72 + \frac{128}{3}\mathcal{L})N_1N_2)$	1.38 s
BiShrink	43.37, 44.01	$\mathcal{O}((8w^2 + \frac{128}{3}\mathcal{L})N_1N_2)$	2.56 s
Proposed LMMSE	43.68, 44.57	$\mathcal{O}((10w^2 + \frac{128}{3}\mathcal{L})N_1N_2)$	4.57 s
Proposed MAP	44.10, 44.95	$\mathcal{O}((10w^2 + 6C_B\mathcal{I} + \frac{128}{3}\mathcal{L})N_1N_2)$	3.86 m

† APSNR_g and APSNR_r denote the average PSNRs for the green and red channel images, respectively.

4.6 Comparisons Between DWT- and CWT-based Methods

In this section, comparisons are made between the CWT-based algorithms and the DWT-based algorithms presented in Chapter 3. In particular, we have considered the proposed and the competitive algorithms. These algorithms are compared both in terms of denoising performance as well as computational efficiency, where the latter is analyzed by the computational complexity of an algorithm and time required for implementation in a given setup. Table 4.6 summarizes the average PSNR, the computational complexity, and implementation time[1] for the proposed and compet-

[1]The methods are implemented on an image of size 1000 × 1000 in the MATLAB environment using a 2 GHz processor with 512 MB RAM.

itive DWT- and CWT-based methods. For each of these methods, the window size that gives the best PSNR results has been used. From this table, it is evident that the CWT-based methods yield better denoising performance than the DWT-based methods. For instance, the proposed MAP method, which gives the best denoising performance among the CWT-based methods, shows an average 0.52 dB improvement in PSNR compared to Proposed method II, which gives the best result among the DWT-based methods. This translates to an average decline of 20×10^3 units in the MSE when the CWT-based MAP method is used, thus indicating a significant improvement in denoising performance. Similarly, improvements in denoising performance are also seen for NeighCoef and BiShrink when these methods are implemented in the CWT-domain instead of the DWT-domain. However, the CWT-based methods have increased computational complexity compared to the DWT-based methods (see Table 4.6). This is mainly due to the fact that the DWT is a non-redundant transform, whereas the CWT has a redundancy of 4. As a result of higher computational complexity, the time required to implement the CWT-based methods is, in general, higher than that of the DWT-based methods. For instance, the CWT-based Proposed MAP method requires 3.86 minutes, whereas the DWT-based Proposed Method II requires only 1.71 seconds. Thus, a tradeoff needs to be made between denoising performance and implementation time when choosing between the DWT- and CWT-based methods.

4.7 Conclusion

In this chapter, two CWT-based denoising methods have been developed for microarray images using the standard MAP and LMMSE estimation criteria. The motivation for CWT-based algorithms is that, in contrast to the DWT, the CWT has certain desirable features such as a near shift-invariance property and improved directional

selectivity that are important for denoising of microarray images which have spots that are approximately circular in shape. In addition, the proposed bivariate estimators incorporate the correlation information of the images as well as noise between the red and green channels. Extensive experiments have been carried out on a large set of microarray images to evaluate the performance of the proposed LMMSE- and MAP-based denoising methods. Simulation results show that the proposed methods provide better denoising performance than the DWT-based algorithms given in Chapter 3 as well as the existing CWT-based methods.

Chapter 5

Conclusion

5.1 Concluding Remarks

In the age of biotechnology, cDNA microarray experiments have played a significant role in advancing genomic and medical research. The success of these experiments relies to a great extent on the ability to extract accurate gene expression measurements from the microarray images that are produced. The task of extracting accurate information from microarray images poses significant challenges because the images are contaminated with a high level of noise. Microarray image processing is therefore an important area of research in the field of bioinformatics. This thesis is concerned with the specific problem of developing methods for reducing noise in cDNA microarray images for the purpose of extracting accurate information regarding gene expression levels.

The wavelet transform has certain properties that allow efficient handling of the image denoising problem. In the literature, many algorithms have been proposed in the wavelet domain for reducing noise in standard images. Unfortunately, none of these methods are well adapted for use with microarray images. It is well known that the pixel intensities of the red and green channel images of microarray experiments are linearly dependent. We have found that such linear dependency exists between the two channels in the wavelet domain as well. The traditional wavelet-based image

denoising algorithms process the red and green channel images separately and treat the images as if they are completely independent. As a result, these methods do not have very good denoising performance. More efficient algorithms could be designed if the correlation between the images could be considered in the estimation process. Thus, in this thesis, new algorithms have been proposed for joint estimation of the noise-free coefficients in the red and green channel images. By defining appropriate joint PDFs for the image coefficients as well as the noise, these algorithms are capable of taking into account the signal correlation and noise correlation between the two channels. For estimation of the noise-free images, two separate estimation techniques have been used, namely, LMMSE and MAP estimation. The algorithms have been designed both in the decimated DWT domain as well as in the CWT domain. The motivation for CWT-based algorithms is that these methods exploit the good directional selectivity and shift-invariance properties of the transform that are not available for the DWT. These properties ensure better denoising performance and more accurate representation of the circular edges of spots within the microarray image.

The proposed DWT- and CWT-based methods have been compared with the existing methods using several real microarray images. It is seen that the new methods show significantly better performance than the existing ones both with respect to denoising and accuracy in estimation of the gene expression measurements. We draw some important conclusions based on these findings. First, joint estimation of the red and green channel images of cDNA microarray experiments is necessary for a more efficient noise reduction performance. Secondly, when dealing with microarray images, consideration of inter-channel dependency is more important than the inter-scale one. Thirdly, the choice of the transform determines the efficiency of the noise reduction algorithm. For instance, in terms of denoising performance, the CWT-based algorithms are more efficient because they provide an average of 0.35 dB improvement

in PSNR as compared to the DWT-based ones. On the hand the DWT-based algorithms are computationally more efficient than the CWT-based ones, since the former is a non-redundant transform while the redundancy of the latter is four. When fast implementation of the algorithm is of major importance, the DWT-based denoising algorithms should be used instead. However, in situations where accuracy of the estimated log-intensity ratios is of greater concern, it is worthy to sacrifice the increase in computation time for a better denoising performance by using the CWT-based algorithms.

5.2 Future Work

There are some additional problems related to this thesis work that could be undertaken for future research. In particular, we would like to investigate the following.

1. In our experiments, the MAE was used as an indicator of the accuracy of the log-intensity ratios that were estimated from the denoised images. This measure is simple and good for preliminary comparisons since it measures the average distance between the log-intensity ratios obtained from the denoised images, and the ratios obtained from the noise-free images. However, a detailed analysis requires the use of more sophisticated methods that provide a clearer picture of the effect of denoising on log-intensity ratio estimation. For instance, one may use a new measure such as Kolmogorov-Smirnov distance or Kullback-Leibler divergence that depends on the PDF of the log-intensity ratios. Further, one might perform standard parametric or nonparametric tests of significance to determine if the two samples of log-intensity ratios obtained from the noise-free and denoised images arise from the same population. In addition to examining the effect on the log-intensity ratio, one may also investigate the effect of denoising on downstream analysis, such the performance of clustering algorithms.

2. The algorithms proposed in this thesis have been developed considering a joint prior function for the noise-free coefficients in both the channels. More specifically, we have considered the bivariate Gaussian PDF to model the noise-free DWT coefficients and magnitude components of the CWT coefficients of microarray images. In Chapter 4, it was seen that the proportion of magnitude components in a subband i.e., X_r and X_g, which satisfied the bivariate Gaussian assumption was comparatively low in the subbands of the third level of decomposition. Although this does not seriously degrade the overall denoising performance, one may investigate whether better results could be obtained by defining a more appropriate joint prior function for the subbands in the third level. We suggest two approaches for specifying the level-specific joint prior functions. First, a more appropriate 'parametric' prior may be chosen for the magnitude components of the CWT coefficients. In the second approach, MMSE estimators for the magnitude components may be derived by considering bivariate 'non-parametric' prior functions for the local neighboring magnitude components. In order to obtain such a prior function, nonparametric density estimation techniques, such as, the kernel density estimator or the smooth estimator of the density function proposed by Babu and Chaubey (2006) based on Bernstein polynomials are required.

3. In this thesis, we have constructed algorithms for reducing AGN in microarray images. However, these images may be corrupted with mixtures of various noise types such as a combination of Gaussian noise and Poisson noise. In future, we would like to develop unified wavelet-based algorithms for simultaneously reducing the mixtures of noise in images.

4. Although the wavelet coefficients in the red and green channel images are known to have a linear dependency, other types of dependencies may exist as well. A

thorough investigation is necessary to determine what other types of associations exist between the two images and how these relationships could be exploited in developing more efficient joint estimators for the wavelet coefficients.

5. Noise contamination is a type of point degradation. A closely related problem is blurring, which is a type of spatial degradation that occurs in a microarray image when the size of the physical space represented by the pixel (i.e., pixel size) is smaller than the spot size of the laser (Stekel, 2003). Blurring is a weakness in the practical imaging system and is undesirable because the pixel intensities in a blurred image are distorted. Thus, an important problem in bioinformatics is the development of an effective wavelet-based algorithm that performs denoising and deblurring of the image simultaneously.

References

Abramovich, F., Sapatinas, T., and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society B*, 60(4):725–749.

Abramowitz, M. and Stegun, I., editors (1965). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.* Dover, NY, 10th edition.

Achim, A. and Kuruoğlu, E. E. (2005). Image denoising using bivariate α-stable distributions in the complex wavelet domain. *IEEE Signal Processing Letters*, 12(1):17–20.

Achim, A., Tsakalides, P., and Bezerianos, A. (2003). SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 41(8):1773–1784.

Aiazzi, B., Alparone, L., and Baronti, S. (1999). Estimation based on entropy matching for generalized Gaussian PDF modeling. *IEEE Signal Processing Letters*, 6(6):138–140.

Antoniadis, A., Bigot, J., and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: A comparative simulation study. *Journal of Statistical Software*, 6(6):1–83.

Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–955.

Athanasiadis, E., Cavouras, D., Spyridonos, P., Kalatzis, I., and Nikiforidis, G. (2007). An automatic microarray image gridding technique based on continuous wavelet transform. In *Proceedings of the 12th International Conference on Computer Analysis of Images and Patterns*, pages 864–870, Austria.

Babu, G. J. and Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors. *Statistics and Probability Letters*, 76:959–969.

Balagurunathan, Y., Dougherty, E. R., Chen, Y., Bittner, M. L., and Trent, J. M. (2002). Simulation of cDNA microarrays via a parameterized random signal model. *Journal of Biomedical Optics*, 7(3):507–523.

Bao, P. and Ma, X. (2005). Image adaptive watermarking using wavelet domain singular value decomposition. *IEEE Transactions on Circuits and Systems Video Technology*, 15(1):96–102.

Barber, S. and Nason, G. P. (2004). Real nonparametric regression using complex wavelets. *Journal of the Royal Statistical Society B*, 66(4):927–939.

Bhuiyan, M. I. H., Ahmad, M. O., and Swamy, M. N. S. (2007). Spatially adaptive wavelet-based method using the Cauchy prior for denoising the SAR images. *IEEE Transactions on Circuits and Systems Video Technology*, 17(4):500–507.

Bidaut, G., Manion, F. J., Garcia, C., and Ochs, M. F. (2006). Waveread: Automatic measurement of relative gene expression levels from microarrays using wavelet analysis. *Journal of Biomedical Informatics*, 39:379–388.

Boncelet, C. (2000). Image noise models. In Bovik, A. C., editor, *Handbook of Image and Video Processing*, NY. Academic Press.

Bozinov, D. and Rahnenfuhrer, J. (2002). Unsupervised technique for robust target separation and analysis of dna microarray spots through adaptive pixel clustering. *Bioinformatics*, 18:747–756.

Cai, T. and Silverman, B. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhya: The Indian Journal of Statistics*, 63:127–148.

Cai, Z., Cheng, T. H., Lu, C., and Subramanium, K. R. (2001). Efficient wavelet-based image denoising algorithm. *Electronics Letters*, 37(11):683–685.

Chang, S. G., Yu, B., and Vettereli, M. (2000a). Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546.

Chang, S. G., Yu, B., and Vettereli, M. (2000b). Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transactions on Image Processing*, 9(9):1522–1531.

Chaux, C., Duval, L., and Pesquet, J.-C. (2006). Image analysis using a dual-tree M-band wavelet transform. *IEEE Transactions on Image Processing*, 15(8):2397–2412.

Chipman, H. A., Kolaczyk, E. D., and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92(440):1413–1421.

Choi, H., Romberg, J. K., Baraniuk, R. G., and Kingsbury, N. G. (2000). Hidden Markov tree modeling of complex wavelet transforms. In *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 133–136, Istanbul, Turkey.

Churchill, G. A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nature Genetics*, 32(Suppl):490–495.

Clonda, D., Lina, J. M., and Goulard, B. (2004). Complex Daubechies wavelets: properties and statistical image modelling. *Signal Processing*, 84:1–23.

Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *Journal of the Royal Statistical Society B*, 62(4):681–698.

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant denoising. In Antoniadis, A. and Oppenheim, G., editors, *Wavelets and Statistics*, Berlin, Germany. Springer-Verlag.

Congdon, P. (2003). *Applied Bayesian Modelling.* John Wiley & Sons, NJ.

Crouse, M. S., Nowak, R. D., and Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902.

Crouse, M. S., Nowak, R. D., and Baraniuk, R. G. (1999). Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Information Theory*, 45:846–862.

Daskalakis, A., Cavouras, D., Bougioukos, P., Kostopoulos, S., Glotsos, D., Kalatzis, I., Kagadis, G. C., Argyropoulos, C., and Nikiforidis, G. (2007). Improving gene quantification by adjustable spot-image restoration. *Bioinformatics*, 23(17):2265–2272.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, first edition.

Davies, S. W. and Seale, D. A. (2005). DNA microarray stochastic model. *IEEE Transactions on Nanobioscience*, 4:248–254.

Donoho, D. L. (1995). Denoising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224.

Eliason, S. R. (2000). *Maximum Likelihood Estimation: Logic and Practice.* SAGE, 5th edition.

Eslami, R. and Radha, H. (2006). Translation-invariant contourlet transform and its application to image denoising. *IEEE Transactions on Image Processing*, 15(11):3362–3374.

Fadili, J. M. and Boubchir, L. (2005). Analytical form for a Bayesian wavelet estimator of images using the Bessel K-form densities. *IEEE Transactions on Image Processing*, 14(2):231–240.

Fan, G. and Xia, X. G. (2001). Image denoising using local contextual hidden Markov model in the wavelet domain. *IEEE Signal Processing Letters*, 8(5):125–128.

Fernandes, F. C. A., Spaendonck, R. L. C. V., and Burrus, C. S. (2003). A new framework for complex wavelet transforms. *IEEE Signal Processing*, 51(7):1825–1837.

Ferrari, R. J. and Winsor, R. (2005). Digital radiographic image denoising via wavelet-based hidden Markov model estimation. *Journal of Digital Imaging*, 18(2):154–167.

Figueiredo, M. A. T. and Nowak, R. D. (2001). Wavelet-based image estimation: An empirical Bayes approach using Jeffreys' noninformative prior. *IEEE Transactions on Image Processing*, 10(9):1322–1331.

Gao, H.-Y. (1998). Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational Graphics and Statistics*, 7(4):469–488.

Gao, H.-Y. and Bruce, A. G. (1997). Waveshrink with firm shrinkage. *Statistica Sinica*, 7(4):855–874.

Giri, N. C. (1993). *Introduction to Probability and Statistics.* CRC Press, 2nd edition.

Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing.* Pearson Education Asia, Delhi, India, second edition.

Gopinath, R. A. (2003). The phaselet transform - An integral redundancy nearly shift-invariant wavelet transform. *IEEE Transactions on Signal Processing*, 51(7):1792–1805.

Goryachev, A. B., Macgregor, P. F., and Edwards, A. M. (2001). Unfolding of microarray data. *Journal of Computational Biology*, 8(4):443–461.

Goswami, J. C. and Chan, A. K. (1999). *Fundamentals of Wavelets, Theory, Algorithms, and Application.* John Wiley & Sons, NY, first edition.

Goulden, C. H. (2007). *Methods of Statistical Analysis.* READ BOOKS.

Hayat, M. A. (2005). *Handbook of Immunohistochemistry and in Situ Hybridization of Human Carcinomas: Molecular Genetics: Liver and Pancreatic Carcinomas.* Academic Press, 1st edition.

Howlader, T. and Chaubey, Y. P. (2008). Bivariate estimator for cDNA microarray images using complex wavelets. In *Proceedings of the International Conference on Bioinformatics and Computational Biology*, volume 1, pages 138–144, Las Vegas, NV.

Howlader, T. and Chaubey, Y. P. (2009a). Noise reduction of cDNA microarray images using complex wavelets. *Submitted to IEEE Transactions on Image Processing, Submission no. TIP-05024-2009.*

Howlader, T. and Chaubey, Y. P. (2009b). Wavelet-based noise reduction by joint statistical modeling of cDNA microarray images. *Journal of Statistical Theory and Practice*, 3(2):349–370.

Huang, J. (2000). *Statistics of natural images and models.* PhD thesis, Brown Univ., Providence, RI.

Hyverinen, A. (1999). Sparse code shrinkage: Denoising of non-Gaussian data by maximum likelihood estimation. *Neural Computation*, 11(7):1739–1768.

Ideker, T., Thorsson, V., Siegel, A. F., and Hood, L. E. (2000). Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *Journal of Computational Biology*, 7(6):805–817.

Jansen, M. and Bultheel, A. (2001). Empirical Bayes approach to improving wavelet thresholding for image noise reduction. *Journal of the American Statistical Association*, 96(3):629–639.

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, 33(4):1700–1752.

Jornsten, R., Wang, W., Yu, B., and Ramchandran, K. (2003). Microarray image compression: SLOCO and the effect of information loss. *Signal Processing*, 83(4):859–869.

Jornsten, R., Yu, B., Wang, W., and Ramachandran, K. (2002). Microarray image compression: Sloco and the effects of information loss. *EURASIP Signal Processing Journal, Special issue on genomic signal processing*, 83:859–869.

Kamen, E. W. and Su, J. (1999). *Introduction to Optimal Estimation.* Springer-Verlag, NY, first edition.

Kazubek, M. (2003). Wavelet domain image denoising by thresholding and Wiener filtering. *IEEE Signal Processing Letters*, 10(11):324–326.

Kim, K. and Timm, N. (2006). *Univariate and Multivariate General Linear Models: Theory and Applications with SAS.* CRC Press, 2nd edition.

Kingsbury, N. G. (1999). Image processing with complex wavelets. *Phil. Trans. R. Soc. London A*, 357(1760):2543 – 2560.

Kingsbury, N. G. (2001). Complex wavelets for shift invariance analysis and filtering of signals. *Applied and Computational Harmonic Analysis*, 10(3):234–253.

Kuan, D. T., Sawchuk, A. A., Strand, T. C., and Chavel, P. (1985). Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis And Machine Intelligence.*, PAMI-7(2):165–177.

Liu, J. and Moulin, P. (2001). Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients. *IEEE Transactions on Image Processing*, 10(11):1647–1658.

Lukac, R., Plataniotis, K. N., Smolka, B., and Venetsanopoulos, A. N. (2004). A multichannel order-statistic technique for cDNA microarray image processing. *IEEE Transactions on Nanobioscience*, 3:272–285.

Lukac, R. and Smolka, B. (2003). Application of the adaptive center-weighted vector median framework for the enhancement of cDNA microarray images. *International Journal of Applied Mathematics and Computer Science*, 13(3):369–383.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, CA, second edition.

Mardia, K. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika,*, 57:519–530.

Mardia, K. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhya B,*, 36:115–128.

Mastriani, M. and Giraldez, A. E. (2006). Microarrays denoising via smoothing of coefficients in wavelet domain. *International Journal of Biomedical Sciences*, 1:7–14.

Mastrogianni, A., Dermatas, E., and Bezerianos, A. (2008). Microarray image denoising using spatial filtering and wavelet transformation. In *Proceedings of the 13th International Conference on Biomedical Engineering*, pages 594–597, Singapore.

McLachlan, G. J., Do, K., and Ambroise, C. (2004). *Analyzing Microarray Gene Expression Data.* John Wiley & Sons, Ltd, NJ, first edition.

Mihçak, M. K., Kozintsev, I., Ramchandran, K., and Moulin, P. (1999). Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Processing Letters*, 6(12):300–303.

Miller, M. and Kingsbury, N. (2008). Image modeling using interscale phase properties of complex wavelet coefficients. *IEEE Transactions on Image Processing*, 17(9):1491–1499.

Moulin, P. and Liu, J. (1999). Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Transactions on Information Theory*, 55:909–919.

Olkkonen, H., Pesola, P., Olkkonen, J., and Zhou, H. (2006). Hilbert transform assisted complex wavelet transform for neuroelectric signal analysis. *Journal of Neuroscience Methods*, 151:106–113.

O'Neill, P., Magoulas, G. D., and Liu, X. (2003). Improved processing of microarray data using image reconstruction techniques. *IEEE Transactions on Nanobioscience*, 2(4):176–183.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, NY, third edition.

Pizurica, A. and Philips, W. (2006). Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising. *IEEE Transactions on Image Processing*, 15(3):654–665.

Portilla. J. and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coeffieints. *International Journal of Computer Vision*, 40:49–71.

Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. P. (2001). Adaptive Wiener denoising using a Gaussian scale mixure model. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 37–40, Thessaloniki, Greece.

Portilla, J., Strela, V., Wainwright, M., and Simoncelli, E. P. (2003). Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351.

Rahman, S. M. M. and Hasan, M. K. (2003). Wavelet-domain iterative center weighted median filter for image denoising. *Signal Processing*, 83:1001–1012.

Rangayyan, R. M., Ciuc, M., and Faghih, F. (1998). Adaptive-neighborhood filtering of images corrupted by signal-dependent noise. *Applied Optics*, 37:4477–4487.

Romberg, J. K., Choi, H., and Baraniuk, R. G. (2001). Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10(7):1056–1068.

Romeu, J. L. and Ozturk, A. (1993). A comparative study of goodness-of-tests for multivariate normality. *Journal of Multivariate Analysis,*, 46:309–334.

Ruosaari, S. and Hollmen, J. (2002). Image analysis for detecting faulty spots from microarray images. In *Proceedings of the 5th International Conference on Discovery Science (DS 2002)*, volume 2534, pages 259–266.

Selesnick, I. W. (2001). Hilbert transform pairs of wavelet bases. *IEEE Signal Processing Letters*, 8(6):170–173.

Selesnick, I. W. (2002). The design of approximate Hilbert transform pairs of wavelet bases. *IEEE Signal Processing*, 50(5):1140–1152.

Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151.

Sendur, L. and Selesnick, I. W. (2002a). Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing*, 50(11):2744–2756.

Sendur, L. and Selesnick, I. W. (2002b). Bivariate shrinkage with local variance estimation. *IEEE Signal Processing Letters*, 9(12):438–441.

Shaffrey, C. W., Kingsbury, N. G., and Jermyn, I. H. (2002). Unsupervised image segmentation via Markov trees and complex wavelets. In *Proceedings of IEEE International Conference on Image Processing*, volume 3, pages 801–804, Rochester, NY.

Simon, M. K. (2002). *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists.* Kluwer, Norwell, MA, 1st edition.

Simoncelli, E. P. and Adelson, E. (1996). Noise removal via Bayesian wavelet coring. In *Proceedings of IEEE International Conference on Image Processing*, volume 1, pages 279–382, Lusanne, Switzerland.

Solbo, S. and Eltoft, T. (2004). Homomorphic wavelet-based statistical despeckling of SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 42(4):711–721.

Srivastava, A., Lee, A. B., Simoncelli, E. P., and Zhu, S. C. (2003). On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18:17–33.

Starck. J., Murtagh, F., and Bijaoui, A. (1998). *Image Processing and Data Analysis: The Multiscale Approach.* Published by Cambridge University Press.

Starck, J. L., Candès, E. J., and Donoho, D. L. (2002). The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684.

Stefano, A. D., White, P. R., and Collis, W. B. (2004). Selection of thresholding scheme for image noise reduction on wavelet components using Bayesian estimation. *Journal of Mathematical Imaging and Vision*, 21:225–233.

Stekel, D. (2003). *Microarray Bioinformatics.* Cambridge University Press.

Strang, G. and Nguyen, T. (1996). *Wavelets and Filter Banks.* Wellesley-Cambridge Press, Wellesley, MA, first edition.

Strela, V., Portilla, J., and Simoncelli, E. P. (2002). Image denoising via a local Gaussian scale mixure model in the wavelet domain. In *Proceedings SPIE 45th Annual Meeting*, San Diego, CA.

Tay, D. B. H., Kingsbury, N. G., and Palaniswami, M. (2006). Orthonormal Hilbert-pair of wavelets with (almost) maximum vanishing moments. *IEEE Signal Processing Letters*, 13(9):533–536.

Tseng, V. S. and Kao, C. (2005). Efficiently mining gene expression data via a novel parameterless clustering method. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(4):355–365.

Turkheimer, F. E., Duke, D. C., Moran, L. B., and Graeber, M. B. (2004). Wavelet analysis of gene expression (WAGE). In *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*, volume 2, pages 1183–1186.

Turkheimer, F. E., Roncaroli, F., Hennuy, B., Herens, C., Nguyen, M., Martin, D., Evrard, A., Bours, V., Boniver, J., and Deprez, M. (2006). Chromosomal patterns of gene expression from microarray data: methodology, validation and clinical relevance in gliomas. *BMC Bioinformatics*, 7:526.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* John Wiley & Sons, NY, first edition.

Voloshynovskiy, S., Pereira, S., Iquise, V., and Pun, T. (2001). Attack modelling: Towards a second generation watermarking benchmark. *Signal Processing*, 81:1177–1214.

Wainwright, M. and Simoncelli, E. P. (2000). Scale mixtures of Gaussians and the statistics of natural images. *Advances in Neural Information Processing Systems*, 12.

Wang, X. H., Istepanian, R. S. H., and Song, Y. H. (2003a). Application of wavelet modulus maxima in microarray spots recognition. *IEEE Transactions on Nanobioscience*, 2(4):190–192.

Wang, X. H., Istepanian, R. S. H., and Song, Y. H. (2003b). Microarray image enhancement by denoising using stationary wavelet transform. *IEEE Transactions on Nanobioscience*, 2:184–189.

Wit, E. and McClure, J. (2004). *Statistics for Microarrays: Design, Analysis and Inference.* John Wiley & Sons, Ltd, England, first edition.

Xiong, H., Zhang, Y., and Chen, X. (2007). Data-dependent kernel machines for microarray data classification. *IEEE/ACM Transactions on Computational Biology and Bioinformatics,*, 4(4):583–595.

Yang, Y., Buckley, M., Dudoit, S., and Speed, T. (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, 11:108–136.

Zhang, A. (2006). *Advanced Analysis of Gene Expression Microarray Data.* World Scientific, Singapore, first edition.

Zhang, H., Nosralinia, A., and R. O. Wells, J. (2000). Image denoising via wavelet-domain spatially adaptive FIR Wiener filtering. In *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2179–2182, Istanbul, Turkey.

Zhang, L., Bao, P., and Wu, X. (2005a). Multiscale LMMSE-based image denoising with optimal wavelet selection. *IEEE Transactions on Circuits and Systems Video Technology.*, 15(4):469–481.

Zhang, W., Shmulevich, I., and Astola, J. (2004). *Microarray Quality Control.* John Wiley & Sons, NJ, first edition.

Zhang, X. Y., Chen, F., Zhang, Y., Agner, S. C., Akay, M., Lu, Z., Waye, M. M. Y., and Tsui, S. K. (2002). Signal processing techniques in genomic engineering. *Proceedings of the IEEE*, 90(12):1822–1833.

Zhang, Y., Parthe, R., and Adjerohy, D. (2005b). Lossless compression of DNA microarray images. In *IEEE Computational Systems Bioinformatics Conference, 2005. Workshops and Poster Abstracts*, pages 128 - 132.