

MICROARRAY IMAGE DENOISING USING COMPLEX GAUSSIAN SCALE
MIXTURES OF COMPLEX WAVELETS

by

LAKSHMI SRINIVASAN

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2011

Copyright © by LAKSHMI SRINIVASAN 2011

All Rights Reserved

To my husband, Ganesh and our parents,
for their unconditional love and support.

ACKNOWLEDGEMENTS

First and foremost, I offer my sincere gratitude to my supervisor, Dr. Soontorn Oraintara, for believing in me and giving me this wonderful research opportunity. I am grateful for his invaluable advice, encouragement and patience during the course of my Masters studies. He has inspired me with his dedication, attention to detail and immense knowledge in this field. I want to thank the thesis committee members Dr. Michael T. Manry and Dr. Saibun Tjuatja for their interest in my research and for taking time to serve in my committee.

I am grateful to all the teachers who taught me during the years I spent in school, both in India and in United States. I want to thank my friendly and cheerful group of MSP labmates, Yothin Rakvongthai, Jixing Yao and Eshwar for their support and encouragement during my thesis study. In particular, I am very thankful to Yothin for his help and motivation, throughout my work. Being new to research, I have learnt a lot from them. It has been an invaluable learning experience for me.

Finally, I would like to express my deepest gratitude to my husband, Ganesh for his sacrifice, encouragement and advice during my graduate studies. Without him this would not have been possible. I am also extremely grateful to our parents and siblings for their support, patience and inspiration. I also thank several of my friends who have helped me throughout my career.

April 15, 2011

ABSTRACT

MICROARRAY IMAGE DENOISING USING COMPLEX GAUSSIAN SCALE MIXTURES OF COMPLEX WAVELETS

LAKSHMI SRINIVASAN, M.S.

The University of Texas at Arlington, 2011

Supervising Professor: Soontorn Oraintara

The scientific world has witnessed an explosion in the development of comprehensive and high-throughput methods for molecular biology experimentation. High-density DNA microarray technology, allows researchers to monitor the expression levels of thousands of genes in an organism simultaneously, to characterize genetic diseases at the molecular level and to direct new treatment for specific cellular aberrations. The microarray analysis is rapidly becoming a standard research tool. But, the images produced by microarray experiments, are not perfect and includes noisy sources, that contaminate them during the various stages of its formation. These microarray images need to be denoised to ensure reliable and accurate downstream analysis. A major challenge in DNA microarray analysis is to effectively dissociate actual gene expression values from experimental noise. This thesis, focuses on proposing an efficient noise reduction technique for microarray images, by using an appropriate model for the complex wavelet coefficients, obtained by decomposition of these images using a complex transform.

Among the number of filtering and enhancement techniques available for noise reduction, wavelet-based approaches have been more successful as it processes the images in multiresolution. In particular, complex wavelets have been more successful in image denoising due to its shift invariance property and improved directional selectivity. A two- channel cDNA microarray experiment generates two 16-bit red and green channel images that reflect the expression levels of the genes in treatment and control samples respectively. Since the two channel images produced are from the same microarray slide, a significant noise correlation between the microarray images exist and methods that exploit this property of inter-channel signal and noise correlation between the two channel images in the complex wavelet domain, achieve better denoising performance. The Gaussian scale mixtures (GSM) model of wavelet coefficients using Bayesian least square(BLS) estimator has been very effective in noise reduction for natural images. To fully utilize the usefulness of complex wavelet coefficients, complex Gaussian scale mixtures (CGSM) model has been developed as an extension of the GSM for real wavelet coefficients. The CGSM model of complex wavelet coefficients, improves the quality of denoised images from using the GSM of real wavelet coefficients.

In this work, we combine the advantages of using an improved CGSM model of the complex wavelet coefficients, by taking into consideration the inter-channel dependency in the complex coefficients of the image as well as the noise for denoising the red and green channel images. Thus, we propose to jointly denoise the two channel microarray images by modeling the complex coefficients of signal and noise using CGSM, by incorporating the joint statistics of the images into the model to achieve better noise reduction performance.

Extensive experimentations are carried out on a 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 structural similarity (SSIM) index as a measure of signal preservation quality of the denoised images to the original image. To impress the usefulness of the joint model, we have compared the joint denoising of the two channel images with independent denoising of these images using same CGSM model. We find the best window size for denoising these microarray images using our proposed method such that, the PSNR of the output images is maximized. We have also compared the performance of the our algorithm against some existing noise reduction methods in literature. We have used the Dual Tree- Complex Wavelet Transform (DT-CWT), which is probably the most widely used complex wavelet transform in image processing, but have also compared our method with other complex-valued multiresolution transforms, such as the fast discrete curvelet transform (FDCT), the pyramidal dual-tree directional filter bank (PDTDFB), and the uniform discrete curvelet transform (UDCT). Results indicate that the proposed denoising method adapted to microarray images, do indeed, lead to better noise reduction evaluated in terms of PSNR and SSIM. Thus, we expect our proposed model for noise reduction, to play a significant role in improving the reliability of the results obtained from practical microarray experiments.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	v
LIST OF FIGURES	x
LIST OF TABLES	xi
Chapter	Page
1. INTRODUCTION	1
1.1 cDNA and Importance of microarray analysis	1
1.2 Background and Motivation	6
1.3 Organization of the Thesis	9
2. NOISE REDUCTION IN MICROARRAY IMAGES	11
2.1 Introduction	11
2.1.1 cDNA microarray image	11
2.1.2 Applications of Microarrays	15
2.1.3 Need for Noise reduction	16
2.2 Previous work	18
2.2.1 Pixel based methods	19
2.2.2 Transform based methods	19
3. COMPLEX GAUSSIAN SCALE MIXTURE (CGSM)	22
3.1 Introduction	22
3.2 Gaussian scale mixtures (GSM)	22
3.3 CGSM model of complex coefficient	23
3.4 Application of CGSM for image denoising	25

4. CGSM BASED DENOISING FOR MICROARRAY IMAGES	28
4.1 Introduction	28
4.2 Proposed Joint denoising algorithm	28
4.3 Experimental results	33
4.3.1 Experimental Setup	34
4.3.2 <i>Experiment I</i> : Best window size estimation	35
4.3.3 <i>Experiment II</i> : Comparison with other denoising methods . .	39
4.3.4 <i>Experiment III</i> : Comparison with other transforms	42
5. CONCLUSION AND FUTURE WORK	48
5.1 Summary	48
5.2 Conclusion	49
5.3 Future work	50
Appendix	
A. ABBREVIATION LIST	52
REFERENCES	55
BIOGRAPHICAL STATEMENT	59

LIST OF FIGURES

Figure	Page
1.1 Central dogma of molecular biology	3
2.1 Illustration of the steps in cDNA microarray experiment (Duggan D.J. <i>et al</i> [1],1999, Nature Genetics)	13
2.2 A sample microarray image	14
2.3 Samples of noisy microarray images taken from SMD database	16
3.1 Illustration of Transform domain denoising	25
4.1 Illustration of Joint denoising in complex wavelet domain	29
4.2 Conditional histogram of the hidden multipliers between (a) \hat{S}_1 and \hat{S}_2 and (b) Joint \hat{S} and \hat{S}_1	31
4.3 Illustration of the subband neighborhood coefficients used for different window sizes (a) w1 (b) w5 and (c) w9. The number denotes the total number of coefficients used in that neighborhood	33
4.4 Illustration of the use of symmetric window (w5) for joint denoising of red and green channel images	37
4.5 Illustration of the use of asymmetric window (w5) & (w1) for joint denoising of red and green channel images	37
4.6 Plot of the PSNR differences (in dB) of our proposed method, LMMSE and MAP estimators as a function of the input PSNR with respect to Bishrink method. (a) Red channel image (b) Green channel image	43
4.7 Plot of the PSNR differences (in dB) of our proposed method, LMMSE and MAP estimators as a function of the input PSNR with respect to BLS-GSM method. (a) Red channel image (b) Green channel image	44
4.8 3D plot of a spot from a microarray image. (a) Clean spot (b) Noisy spot with noise of $\sigma_\epsilon=1200$ and $\rho = 0$ (c) denoised spot using LMMSE (d) denoised spot using Joint CGSM with window (w5) & (w1).	47

LIST OF TABLES

Table		Page
4.1	PSNR values (in dB) and SSIM indices of denoised images using DTCWT-CGSM with different window sizes	38
4.2	PSNR values (in dB) and SSIM indices of denoised images from several methods	40
4.3	Redundancy ratios of FS and DT-CWT	41
4.4	Redundancy ratios of the transforms used herein	45
4.5	PSNR values (in dB) and SSIM indices of denoised images using Joint CGSM(Window:(5)&(1)) with different transforms	46

CHAPTER 1

INTRODUCTION

DNA microarray technology has been established as one of the most important tool for analyzing the interactions between the genes that make up a genome. With microarray technology, the possibility to simultaneously monitor the expression levels of thousands of genes, under various experimental conditions, has become a reality. With such genome-wide gene expression profiles, microarray technology holds a promise for providing possible clues in treating difficult diseases caused due to genetic abnormalities.

1.1 cDNA and Importance of microarray analysis

The ability of the cells to store and translate the genetic information is very essential to create and maintain a living organism. The entire genetic material in a living organism is called a genome. Every organism, has a genome that contains all of the biological information needed to build and maintain a living organism. The biological information contained in a genome is encoded in its deoxyribonucleic acid (DNA) and is divided into discrete units called genes. These genes are packaged in chromosomes and affect specific characteristics of the organism. A cell relies on its gene products for a wide variety of functions including energy production, macromolecule biosynthesis, cellular architecture maintenance, and response to environmental stimuli. Although DNA is the carrier of genetic information in a cell, proteins are the active working components of the cellular machinery. Proteins function in the cell in various forms like, **enzymes** that catalyze nearly all chemical processes

in cells; **structural components** that give cells their shape and help in mobility; **hormones** that transmit signals throughout the body; **antibodies** that recognize foreign molecules; and **transport molecules** that carry oxygen. Proteins are made up of smaller units called amino acids. DNA is the blueprint that specifies the order and number of amino acids and, therefore, the shape and function of the protein. Expression of the information in DNA, to synthesize proteins is mediated by ribonucleic acid (RNA) molecule. Thus, the cell expresses itself through the proteins it produces, which are coded by genes in the cell.

DNA: The DNA that constitutes a gene is a double-stranded molecule consisting of two chains running in opposite directions giving it a coiled structure called double-helix. These two anti-parallel strands of sugar-phosphate linkages are connected to each other by chemical pairing of each base on one strand to a specific base on the other strand. Adenine, Thymine, Guanine and Cytosine are four amine bases of which DNA is made up of and its sequential order contributes to the functioning of a particular segment of the DNA strand (a gene). The bases exhibit a characteristic and specific bonding known as base pairing. Base pairing (also known as Watson-Crick base pairing) is a chemical bonding process that allows molecular hybridization to occur, which will be discussed later. Between two strands of DNA, the base known as adenine (A) specifically bonds with thymine (T) through two hydrogen bonds, and guanine (G) specifically bonds to cytosine (C) through three hydrogen bonds, in a manner that creates the double helix. The amine base that will form a bonding pair with another amine base (A with T and G with C) is considered to be its complementary base. This complementary base pairing is what makes DNA a suitable molecule for carrying our genetic information: one strand of DNA can act as a template to direct the synthesis of a complementary strand. In this way, the information in a DNA sequence is readily copied and passed on to the next generation of cells.

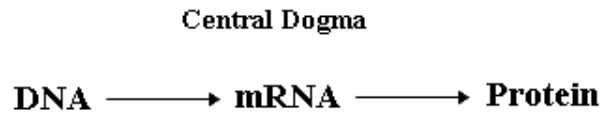


Figure 1.1. Central dogma of molecular biology.

Transcription of DNA to RNA and translation RNA to protein is referred to as the central dogma of molecular biology and is illustrated in figure (1.1) For the cell to express, it must generate the proteins required to carry out the specific functionality. Protein synthesis occurs in two steps : Transcription and Translation. The encoded information in DNA is copied to messenger RNAs (mRNAs) by a process called Transcription. Transcription is the communication of a genetic code from DNA to RNA through the synthesis of a strand of RNA that has sequence of bases complementary to that of the DNA strand from which it is generated. mRNA contains the base uracil in the place of base thymine in DNA. The sequence of the bases in a DNA segment comprises the code or genetic instructions that are passed on from the DNA molecule to the RNA molecule because of the specific pairing that occurs between the bases in DNA and RNA. Genetic transcription is carried out to direct the activity of the cell.

Synthesis of proteins from the information coded in the mRNA is accomplished through a process called Translation. The mRNA then moves out of the nucleus of the cell to the site of protein synthesis, the ribosomes. To convert the mRNA into protein, transfer RNA (tRNA) is used to read the mRNA sequence, three nucleotides (represented by their bases A, U ,G or C) at a time. A sequence of three nucleotides in mRNA is called a Codon. Each of the possible 64 codons in the mRNA codes for an amino acid. The mRNA sequence is matched, three nucleotides at a time to a complementary set of three nucleotides in the anticodon region of the correspond-

ing tRNA molecule. Opposite the anticodon region of each tRNA, an amino acid is attached and as the mRNA is read off, the amino acids on each tRNA are joined together through peptide bonds. The polymer of amino acids thus generated, forms the protein molecule. Thus, we see that the proteins generated by the cell depends upon the codon sequences in mRNA, which are formed by a base sequence complementary to that of DNA strand it is generated from. Though all the cells in the body contain the same genome, the genes expressed in each cell differs and hence each cell produces different proteins giving it a different behavior.

Complementary DNA (cDNA): mRNA produced from transcription directs the production of cellular proteins, and is thus important in experiments of gene expression. Researchers want to observe what cellular proteins are produced and the function of those proteins in particular types of cells (such as tumor cells) or in response to specific external stimuli. So, they are interested in testing the expression patterns of the mRNA. Although protein synthesis and activation are not regulated solely by mRNA levels in a cell, mRNA measurement is used to estimate cellular changes in response to external signals or environmental changes. The molecule of mRNA is relatively fragile and is first chemically bound to a DNA molecule. In order to remove it from the other cellular components, it can easily be broken down by the action of enzymes that are prevalent in biological solutions. So researchers commonly extract the mRNA from the cell and generate a form of DNA that possesses the complementary bases of the mRNA while existing in a more stable state. This form of DNA, known as complementary DNA (cDNA), is created directly from the sample mRNA through a procedure known as reverse transcription (transcribing complementary genetic base sequences from RNA to DNA). cDNA is also called synthetic DNA, since it is formed through reverse transcription from RNA rather than

through self-replication during cell division. cDNA is generally prepared in strand lengths of 500 to 5,000 bases of known sequence.

Microarrays may be used to measure gene expression in many ways, but one of the most popular applications is to compare expression of a set of genes from a cell maintained in a particular condition (treatment) to the same set of genes from a reference cell maintained under normal conditions(control). In two-channel DNA microarray, mRNA molecules from the control and treatment samples to be studied, are isolated and complementary DNA (cDNA), which is called target, is produced from it using Reverse Transcriptase (RT) enzyme. During this process, the control samples are labeled with green (Cy3) and treatment samples with red (Cy5) fluorescent dyes respectively. This mixture is then washed over a microarray slide. A microarray slide has known reference genes called probes that are printed on them in a pre-determined scheme. This results in competitive hybridizations between the target and probe, causing the target to bind to the slide. The scanner then scans the slide with a green and red laser separately to produce two 16-bit images. The spots occupy a small fraction of the image area, and have to be individually located and isolated from the image background prior to the estimation of its mean intensity. The intensity values of each spot in the image represents the level of gene expression (strength of binding), with higher values implying more expression of that gene[2]. It thereby enables us to identify differentially expressed genes in the control and treatment samples. Log intensity ratio of the expression levels of these gene spots in the red and green channel images are extracted and used for further analysis like gene clustering, etc. The DNA microarrays will be discussed in detail in chapter 2.

1.2 Background and Motivation

In this section we first discuss the steps involved in microarray image analysis and the motivation for the proposed work. Once the microarray image has been converted into the digital domain, the first step in the analysis of data is to process these images. The intensity of the spots in red and green channel image give the expression level of the gene of interest in the treatment and control samples respectively. To measure the relative activity of the gene, we take the log intensity ratio of the gene spot intensity of treatment and control sample images. If this quantity is greater than zero, we say the gene is over-expressed or up-regulated in the treatment sample in comparison to control. If less than zero, gene activity is suppressed or down-regulated in the treatment compared to control. If zero, they are equal expressed in both conditions. These transcription values from the image analysis stage are commonly focused at finding similarities between the behavior genes under two different conditions or between known and unknown genes in the initial samples. Such behavior analysis renders the ability to predict the behavior of genes in different conditions or function of unknown genes.

This overall process can be broken down into three main steps [3]:

1. **Addressing:** A typical slide has several gene groupings on its surface, with hundreds of genes per group. Addressing, uniquely identifies the locations of the genes in the image in such a way that, there is no overlap between any two gene locations. The separation of gene spots in the image into distinct cells by constructing grids covering block of spots is called microarray image gridding. Several methods have been proposed for gridding. They can be viewed in terms of automation as manual, semiautomated and fully automated. For example, popular softwares ImaGene [4] uses a semiautomated process while ScanAlyze [5] uses a manual gridding. An

overview of the various gridding methods can be found in [6].

2. **Segmentation:** Once the genes have been localized by the addressing process, the task becomes one of accurately separating the gene spots themselves from their local background area. Various algorithms proposed to solve this problem include fixed circle method in [5], adaptive circle method, hybrid k-means algorithm [7], Markov random field (MRF) modeling for the microarray spot segmentation [8], wavelet based MRF [9], etc. A comparison of some these methods of segmentation is presented in [10].

3. **Feature Extraction:** Feature extraction quantifies the area of interest (gene spot) by intensity measurement along with that of the local background. This process calculates a host of other metrics to facilitate downstream analysis such as gene clustering [11], tumor analysis, etc.

One of the major factors that complicates the image analysis task is that, the microarray images are contaminated with various types of noise (biological and experimental). If this noise is treated improperly, analysis of these images may result in erroneous biological conclusions. Biological noise is intrinsic, it includes the stochastic internal noise of the cell and error sources related to sample preparation, and it induces image blurring[12]. Experimental noise can be subdivided into source noise and detector noise. Source noise is generated during the fabrication and target labeling, whereas detector noise is generated during the amplification and digitization stages. These types of noise produce microarray images, which are corrupted by irregularities in the shape, size, and position of the spots, and are dominated by spatially inhomogeneous noise[12]. The undesirable effect of noise is that, it causes inaccurate spots segmentation, which in turn has a direct effect on the incorrect estimation of the relative mean spots intensities and reduces the reproducibility and validity of the gene expression levels, derived from microarray images. Noise complicates all microarray

image processing tasks (gridding, segmentation, intensity extraction). For noiseless images, these would have been a trivial task even by using simpler segmentation and feature extraction method, but this is not the case[12].

In order to overcome these and other noise related issues, a systematic approach is needed while analyzing the image data. The pre-processing of these images, should be able to reduce the artifact noise without disturbing the signal of interest (the gene spots).

An investigation of the noise reduction algorithms for microarray images reveal an important phenomena.

First, among the methods that are proposed for denoising the images, the wavelet based approaches have been promising due to its ability to analyze images in more than one resolution. In particular, complex wavelets due to improved directional selectivity and shift invariance property are an attractive choice.

Second, methods that account for the inter-channel noise and signal correlation between the red and the green channel images are more efficient in denoising the microarray images.

Third, a recent work [13], which combines both above said advantages by denoising the magnitude of the complex coefficients using linear minimum mean square error (LMMSE) and maximum *a posteriori* (MAP) estimator have shown better denoising performance in terms peak signal to noise ratio (PSNR).

Among the models proposed for wavelet coefficients, we are particularly interested in Gaussian scale mixture (GSM) model. It is a statistical model of the coefficients of an over-complete multiscale oriented basis. Neighborhood coefficients in adjacent positions and scales are modeled as a product of, a Gaussian random vector and a hidden positive scalar multiplier, which modulate the local variance of the neighborhood coefficients. This model is able to account for the observed corre-

lation between the coefficient amplitudes. This GSM model is used for real -valued random vectors. Application of this model for noise reduction in natural images using Bayesian least square (BLS) estimator [14] is very effective. Recently, complex GSM (CGSM) was developed for complex wavelets, as an extension of GSM used for real wavelets. This model completely utilizes the usefulness of complex magnitude and phase of the complex coefficients in the statistical framework for image modeling. It is shown to be more effective in image denoising than using the joint GSM model for real and imaginary parts of complex coefficients.

The above discussed study of microarray image denoising algorithms and the usefulness of CGSM, is the main motivation behind this work. In this proposed work, we want to apply the more appropriate CGSM model of the complex coefficient and also utilize the property of correlation between the signal and noise that exist between the two channel microarray images in denoising them.

1.3 Organization of the Thesis

This thesis is organized as follows. Some background on the microarray experiments and wavelet-based noise reduction schemes are given in Chapter 2. It helps in understanding the proposed algorithm and its advantages. Chapters 3, discusses the details of the CGSM and shows how it models the complex wavelet coefficients. In chapter 4, we present our proposed algorithm for microarray image denoising using the Bayes least square (BLS) estimator in the complex wavelet domain. Using these results, we discuss how we incorporate the inter-channel signal and noise correlations into the BLS estimator of the CGSM to denoise the two channel images. Suitable window sizes for denoising these images is investigated first. Using this optimum neighborhood window size, simulation experiments are conducted on a set of cDNA microarray images to evaluate the performance of the proposed algorithms. Com-

parisons are made between the joint denoising algorithm and independent denoising of these images to show the effectiveness of the proposed method. In addition, we also compare its performance, with some of the existing efficient noise reduction techniques. Finally, we fix our proposed model and vary the complex transform used to capture the effect of choice of transform in denoising these images. 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

NOISE REDUCTION IN MICROARRAY IMAGES

2.1 Introduction

In this chapter, we discuss the various noise reduction techniques for microarray images. To get a good understanding of these methods and the difficulties involved, a clear understanding of the dynamics involved in the experiment should be understood. So, first we focus primarily on technical aspects of cDNA microarrays and the various sources of noise. Next, we impress on the need for an efficient pre-processing of these images to reduce the noise contamination. Among the previous work in this area, we place particular emphasis on wavelet-based techniques as it will give a good basis for us to introduce our proposed work.

2.1.1 cDNA microarray image

High-density DNA microarray technology allows researchers to monitor the interactions among thousands of gene transcripts in an organism on a single experimental medium, which is often a glass microscope slide or nylon membrane. Microarrays can be used to evaluate the dynamic expression of genes in response to normal cellular activity (for example, changes in gene transcription, cell division) or in response to external stimuli (for example, a toxic substance, viral infection). The ability to simulate a large variety of cellular conditions and then translate and process the resulting large quantities of data, provides a systematic way to evaluate cellular function and genetic variations.

The following is a simplification of the complicated biochemical processes and protocols involved in the preparation of nucleic acid materials and microarrays and in the conduct of gene expression studies in the biology laboratory. These were taken from [15]. It is provided as an introduction to the technology.

The microarray array is a solid base on which a grid of spots or droplets of genetic material of known sequence is arranged systematically. The array is commonly a small piece of glass or nylon (similar to a microscope slide), with thousands of spots or wells that can each hold a droplet representing a different cDNA sequence. The cDNA of known sequence that is fixed onto the array after micro-spotting is commonly referred to as the probe, while the polynucleotide of unknown sequence in the biological sample solution is commonly referred to as the target. The robotic machine that applies the droplets of different cDNA strands of known sequence to a spot on the array is called the spotter or arrayer.

Once the slide is prepared, to identify and measure the presence of a nucleotides of unknown sequence (in the target sample) after it binds to the material on the microarray, it is labeled with a fluorescent dye. The dye is incorporated with the molecule during reverse transcription. A different dye color is used for the control sample, and generally the two dye colors used are Cyanine 3 (Cy3) and Cyanine 5 (Cy5) due to the accompanying scanning and imaging requirements to detect fluorescent light of specific wavelengths. These two samples are then mixed and washed over a microarray slide leading to hybridization between the target and the probes. Molecular hybridization is the association of single strands of polynucleotides through their specific base-pairing properties to form a complementary double-stranded molecule. This is the chemical process that occurs between the labeled polynucleotide strands of target tissues and their complementary strands of cDNA of known sequence among the spots on the array. Ideally, if a polynucleotide from the target sample contains a

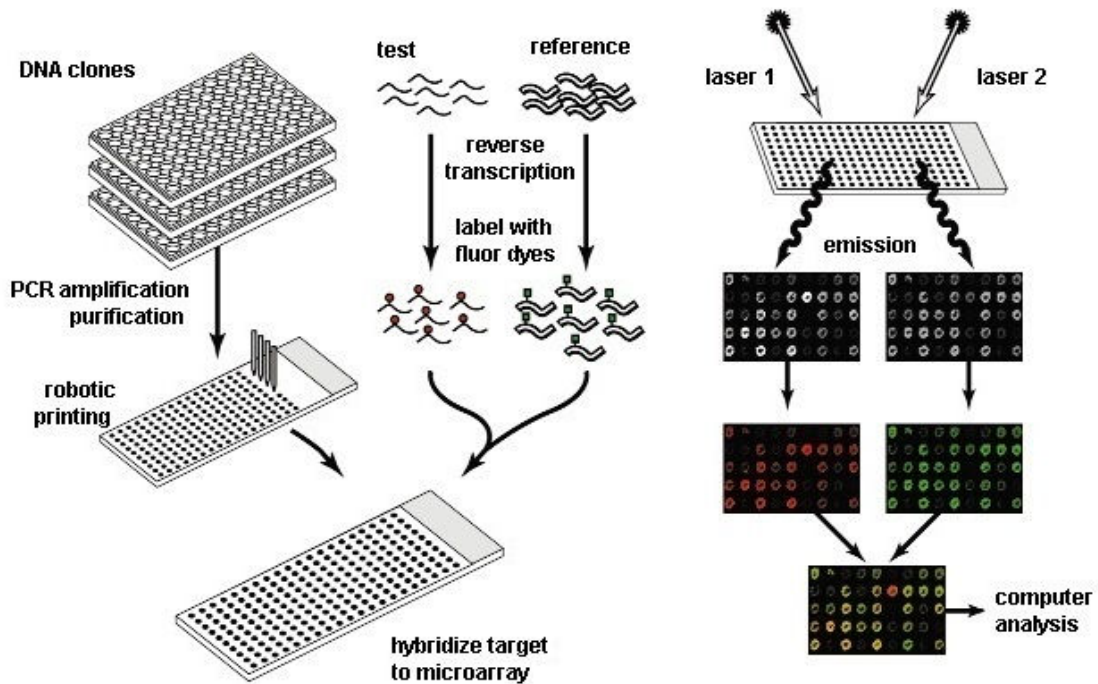


Figure 2.1. Illustration of the steps in cDNA microarray experiment (Duggan D.J. *et al* [1], 1999, Nature Genetics).

base sequence that is complementary to that of a polynucleotide at one spot of the array, it will hybridize to the molecule at that spot. The location of that spot on the array grid will then be detectable by the fluorescent light that is given off during the scanning and imaging processes. When many target polynucleotides hybridize to complementary cDNA probe strands at one spot on the array then the fluorescent signals emitted and detected at that spot will have greater intensity.

Many scanners use a specific frequency of light from a laser (for example, an argon laser) in the ultraviolet region to excite the fluorescent dye attached to the target samples that have hybridized to their complementary probe sequences on the array. The photons emitted by the excited dye are collected at a detector, which measures

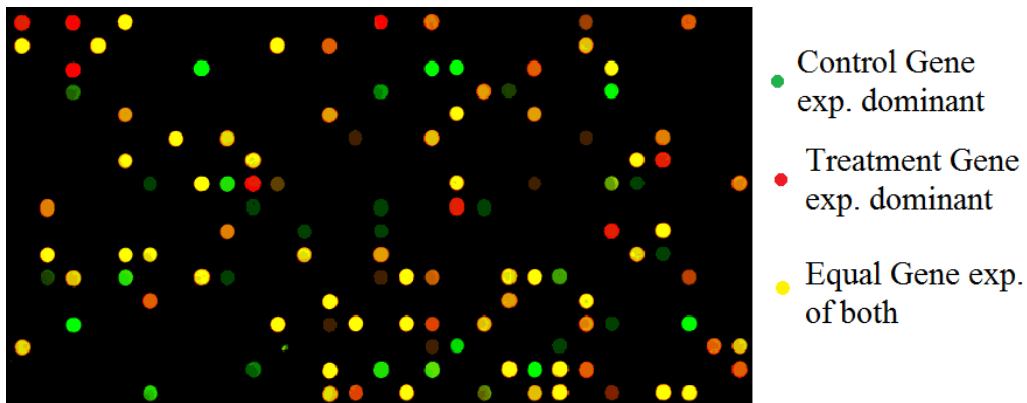


Figure 2.2. A sample microarray image.

and records their levels, converting the measurements to electrical signals. Since two different fluorescent labels are generally used in gene expression studies, each slide is scanned at two wavelengths. The ratio of the fluorescent light emissions between the two different wavelengths (corresponding to the two different dyes used to label the unknown and control polynucleotide target samples) is the indirect measurement of the relative gene transcript expression levels.

A digitized scanned array image is obtained from the microarray scanner and is displayed on a monitor. False coloration of the fluorescent intensities, translated on the computer monitor as pixel intensities, is applied to the image to produce a color image for the analyst to read. If the biochemist tagged the polynucleotide from the unknown experimental sample with a red dye and the control polynucleotide sample with a green dye, and the false colorations mimic the fluorescent tagging, then visualization of a red spot on the final array grid indicates that the unknown polynucleotide hybridized abundantly to the cDNA affixed at that location on the microarray slide. A final green spot indicates that the control polynucleotide hybridized abundantly to the cDNA affixed at that location, a yellow spot indicates that the unknown and the control polynucleotides hybridized in relatively equal amounts at that location

on the microarray, and a black spot indicates that neither sample of polynucleotides hybridized at that location. The standard image format for microarray images is a 16-bit tagged image file format (TIFF). Figure (2.2) illustrates a cDNA microarray experimental procedure. For more details on the specifics of each step refer to [15, 16]

2.1.2 Applications of Microarrays

DNA Microarray technology can help in the identification of new genes, their function and expression levels under different conditions. DNA Microarray technology helps researchers learn more about different diseases such as heart diseases, mental illness, infectious disease and especially the study of cancer. It will be possible for the researchers to further classify the types of cancer on the basis of the patterns of gene activity in the tumor cells. This will tremendously help the pharmaceutical community to develop more effective drugs as the treatment strategies will be targeted directly to the specific type of cancer. Microarray technology also has extensive application in pharmacogenomics and toxicogenomics. Pharmacogenomics is the study of correlations between therapeutic responses to drugs and the genetic profiles of the patients and Toxicogenomics establishes correlation between responses to toxicant and the changes in the genetic profiles of the cells exposed to such toxicant. Comparative analysis of the genes from a diseased and a normal cell will help in the identification of the biochemical constitution of the proteins synthesized by the diseased genes. The researchers can use this information to synthesize drugs which will combat with these proteins and reduce their effect. Microarray technology provides a robust platform for the research of the impact of toxins on the cells and their passing on to the progeny.

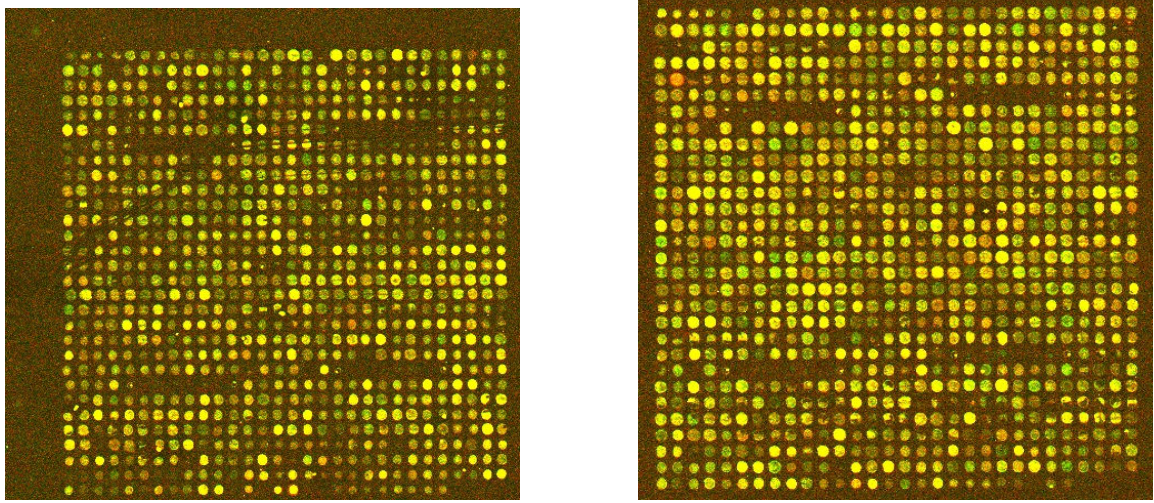


Figure 2.3. Samples of noisy microarray images taken from SMD database.

2.1.3 Need for Noise reduction

The technology for microarray studies is still evolving, and many researchers are conducting their studies using different types of customized microarrays, including home-made array chips. In general, the current technology does not consistently generate robust and reliable data when used in the average laboratory. Ideally, reliable microarray data should exhibit the qualities of accuracy, assessed by the probable error of a measurement, and precision, defined by the reproducibility of a measurement [15]. Samples of noisy microarray images from the SMD database is shown in figure (2.3).

The intensities of the signal represent the amount of fluorescent DNA bound to microarrays and is subject to considerable uncertainty because of large- and small-scale intensity fluctuations within spots, nonadditive background, and fabrication artifacts (Brown et al. 2001), contributing to poor-quality images. Some of the noise sources in the different stages of the experiment are discussed below.

Biological contamination One of the places for the noise to enter the system is within the biological stage itself. This process relies on the biochemical interactions of the underlying four- base nucleotide components. A mutation or corruption of a single nucleotide within a sequence interaction can have a significant impact on the final hybridization results.

Hybridization noise As hybridization is a complex process, it can be affected by conditions like temperature, humidity, salt and formamide concentration. With genes of closely related families, it is not uncommon for cross hybridization to occur between different genes. Such cross hybridization can manifest itself into the resultant image surface as excessive bounding of genetic material or erroneous pairing between sample and reference material.

Washing noise Once the hybridization process is complete, the microarray slide must be washed for two important reasons. First, not all the cDNA within an individual experiment will have bound to the probe's DNA. This excess material must be removed to calculate the gene expression quantity as accurately as possible. Second, cross hybridization between similar species can cause probe to bind with inappropriate target gene. Washing helps in remove these weakly bound target from the probe region and thereby increasing the stringency of the resultant data. A poor wash with excessive wash material or prolonged wash time, can causing the washing material to bind to dry on the slide surface.

Systematic noise This type of noise in microarray usually results from poor slide preparation and manual process. Other sources of systematic variation, internal or external to the sample, include fluctuations in the physical properties of the dyes, efficiency of dye incorporation, probe coupling and processing procedures, background and overshining effects, and scanner settings, among others. The background signal may increase due to dust, fibers, fingerprints, autofluores-

cence of the coated glass and hybridization problems resulting from dehydration near the edge of coverslips (Hess et al. 2001).

Thus, we can summarize that, due to the dynamics involved in the microarray experiment, the microarray images produced are not perfect. It is contaminated with various amounts of artifact noise that could cause an analysis algorithm to make erroneous decisions. The noise originates from different sources during the course of experiment, such as hybridization noise, washing noise, systematic noise, hardware noise, photon noise, electronic noise, laser light reflection, dust on the slide, etc. Additionally, the natural fluorescence of the glass slide and non-specifically bounded DNA or dye molecules add a substantial noise floor to the microarray image. To make the task even more challenging, the microarrays are also afflicted with discrete image artifacts such as highly fluorescent dust particles, unattached dye, salt deposits from evaporated solvents, fibers and various airborne debris[17]. Clearly, given the various sources of noise in the microarray procedure, microarray images need to be denoised to ensure reliable and accurate downstream analysis. Thus, critical first steps in any analysis of gene expression data must include an attempt to clean the data by automatic procedures that can improve image quality, by separating signal from noise.

2.2 Previous work

In this section we investigate the previous research work on noise reduction in microarray images. 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. These are help only in low-level

noise reduction. 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[12]. To achieve accurate estimation of log-intensity ratios from these images, efficient algorithms for reducing noise is essential for microarray image analysis. These noise reduction schemes can be broadly classified as pixel domain and transform domain methods.

2.2.1 Pixel based methods

In pixel domain methods [18], [19], [17], contamination due to noise is removed by a filtering operation using the neighborhood information of the pixels. In [18], nonlinear filtering solutions based on robust order statistics of the pixels is used to remove background and high frequency noise. It consider the two channel images as vectorial inputs and denoises them by minimizing the aggregate distance between the vectorial inputs. It also exploits the correlation that exists in the microarray images. A similar method in [17], proposes use of arithmetic mean filter (AMF) and the vector median filter (VMF) for the suppression of the detected noise in microarray images. [20] proposes a pixel based approach in which possible microarray slide image is recreated, with all the genes removed. The estimate of background thus recreated is subtracted from the original, to calculate the gene ratios with less influence from outliers and other artifacts. But its seen that, these pixel based method, though simpler to implement, are not as efficient as transform based methods.

2.2.2 Transform based methods

In image processing, the image is often transformed because certain properties of the transform domain data enables us to efficiently solve the problem at hand than the pixel domain approach. Various kinds of transforms such as the discrete cosine

transform (DCT), the discrete Fourier transform (DFT), the discrete wavelet transform (DWT) and Complex Wavelet Transform (CWT) are available for processing microarray images. Among these various transforms, wavelets are being extensively used for image denoising as it can efficiently process images in more than one resolution. Denoising of images corrupted by Gaussian noise using wavelet techniques is effective because of its ability to capture the energy of a signal in few energy transform values. In addition it also provides flexibility with choice of basis function, as required for the application, making it an attractive choice. Wavelet transform is superior approach to other time-frequency analysis tools because its time scale width of the window can be stretched to match the original signal, especially in image processing studies. This makes it particularly useful for non-stationary signal analysis, such as noises and transients. Also, wavelets being used for other microarray image analysis tasks like gridding and segmentation[9], it can easily be incorporated into a comprehensive automated microarray image analysis system.

In [21], stationary wavelet transform (SWT) is employed, which is shift invariant, to reduce the noise in these images. It follows a three step procedure, in which the image is first decomposed by the stationary wavelet transform. Then the decomposition coefficients are thresholded by the *SureShrink* rule. These threshold coefficients are then reconstructed to get the denoised image. The main drawback of such an approach is that, the efficiency of this algorithm depends highly on the choice of thresholding function used. The method in [22] is based on decomposing the noisy microarray into wavelet subbands, then applying a smoothing filter within each highest subband, and reconstructing the microarray from the modified wavelet coefficients. This is applied only to the first level of decomposition. The smoothing filter exploits the local coefficient variations in reducing noise. Another similar approach [23], first decomposes the signal by a multiresolution transform and then

accounts for both the multiscale correlation of the subband decompositions and their heavy-tailed statistics. Unfortunately, the discrete wavelet decomposition based on these typical wavelets suffer from two main problems which hamper their use for many image analysis and reconstruction tasks [24]:

- Lack of *shift invariance*, which means small shifts in the input signal can cause major variations in the distributions of energy between the wavelet transform coefficients at different scales.
- Poor *directional selectivity* for diagonal features, because wavelet filters are separable and real.

For better reconstruction of edges of the circular spots of the microarray images, improved directional selectivity and shift invariance of the transform are desirable. Complex wavelets overcome these two key problems by introducing limited redundancy into the transform. They provide both shift invariance and good directional selectivity. Dual-Tree Complex wavelet Transform, contourlet, curvelet and steerable pyramid possess these properties. Dual-Tree Complex wavelet transform (DT-CWT) due to its minimum redundancy and computational efficiency are popularly used. [13] exploits the inter-channel signal and noise correlation that exists between the two channel microarray generated from the same slide. It accounts for this property in the complex wavelet domain and effectively reduces the noise in these images compared to independent denoising of images. With this knowledge as the basis, we present the proposed model in the next chapter.

CHAPTER 3

COMPLEX GAUSSIAN SCALE MIXTURE (CGSM)

3.1 Introduction

The goal of this chapter is to introduce CGSM, which is used to model the complex wavelet coefficients of signal and noise in microarray images. This work has been completely developed in the paper [25]. It exploits the usefulness of the complex magnitude and phase of complex wavelets and proposes an appropriate model to handle complex random variables in the statistical framework.

3.2 Gaussian scale mixtures (GSM)

Noise reduction in images based on a statistical modeling of the wavelet coefficients by GSM was developed in [14]. Gaussian scale mixtures are a class of distributions that can be expressed as a product of a Gaussian random vector, and an independent hidden scalar multiplier. In multiscale orientated representation, the marginal distributions of the wavelet coefficients of images are highly kurtotic and have high correlation between coefficients at similar position, orientation and scale. GSM model can account for both the shape of wavelet coefficient marginal and the strong correlation between the neighborhood amplitudes. The GSM distribution encompasses many known distributions as special cases, like Student's t-distribution, the generalized Gaussian distribution, etc[26].

Let an image be decomposed into oriented subbands at different scales. Then, local neighborhood around each reference coefficient is defined. This may include coefficients from different subbands at nearby scales or orientation as well as from the

same subband. Having defined this, the coefficients in each local neighborhood of a subband is modeled by a GSM. By definition, a random vector is a Gaussian scale mixture if and only if, it can be expressed as the product of a zero-mean Gaussian vector \mathbf{u} and an independent positive scalar random variable \sqrt{z}

$$\mathbf{x} = \sqrt{z}\mathbf{u} \quad (3.1)$$

The variable z is known as the *multiplier*. The vector \mathbf{x} is mixture of Gaussian vectors, whose density is determined by the covariance matrix \mathbf{C}_u of the vector \mathbf{u} with mixing density $p_z(z)$ Then,

$$p_{\mathbf{x}}(\mathbf{x}) = \int p(\mathbf{x}|z)p_z(z)dz \quad (3.2)$$

$$= \int \frac{\exp\left(\frac{-\mathbf{x}^T(z\mathbf{C}_u)^{-1}\mathbf{x}}{2}\right)}{(2\pi)^{N/2}|z\mathbf{C}_u|^{1/2}} p_z z dz \quad (3.3)$$

where, N is dimensionality of the vectors \mathbf{x} and \mathbf{u} . The most widely used choice for $p_z(z)$ is the *Jeffery's prior*, which is obtained as

$$p_z(z) \propto \frac{1}{z} \quad (3.4)$$

More details regarding the choices for the probability density of $p_z(z)$ can be found in [14].

3.3 CGSM model of complex coefficient

Having introduced the GSM model, we extend these to introduce the concept of complex GSM (CGSM) developed in [25].

For image denoising, we model the real subband coefficients of the image using GSM. To handle complex subband coefficients generated by complex wavelets we have two options.

- To model jointly the real and imaginary parts of the complex coefficients by GSM.

- To model the complex coefficients directly as complex GSM. CGSM models the complex subband coefficients as a product of a complex Gaussian pdf and a hidden multiplier.

Of the two, the CGSM is preferred as it fully utilizes the complex coefficients, its magnitude and phase information in the statistical framework.

Distributions whose pdf depend only on the covariance matrix, the pdf of a complex random variables can be expressed as a function of a complex-valued vector itself, if its the real and imaginary parts satisfy the circularity condition i.e., the covariance matrix of real and imaginary parts must be equal and the sum of their cross covariance must be zero.

To start with, let X and Y be two $N \times 1$ zero-mean Gaussian random vectors. Assume that they are jointly Gaussian with the circular condition, i.e.

$$\mathbf{C}_X = \mathbf{C}_Y \quad \text{and} \quad \mathbf{C}_{XY} = -\mathbf{C}_{YX}, \quad (3.5)$$

where $\mathbf{C}_X = E[\mathbf{X}\mathbf{X}^T]$, and $\mathbf{C}_{XY} = E[\mathbf{X}\mathbf{Y}^T]$. Let $U = [X^T, Y^T]^T$, which is Gaussian with covariance matrix $C_U = \begin{bmatrix} C_X & C_{XY} \\ C_{XY} & C_X \end{bmatrix}$. Let $Z = X + jY$ with complex covariance matrix $C_Z = E[ZZ^H]$. Therefore, $C_Z = 2(C_X - jC_{XY})$. Then, we can express the Gaussian pdf of $U = [X^T, Y^T]^T$ as the complex Gaussian pdf of $Z = X + jY$ having the form

$$f_Z(\mathbf{z}) = \frac{\exp(-\mathbf{z}^H \mathbf{C}_Z^{-1} \mathbf{z})}{\pi^N |\mathbf{C}_Z|}, \quad \text{for } \mathbf{z} \in \mathbf{C}^N. \quad (3.6)$$

This circularity condition is shown to be satisfied by complex subband coefficients of the complex wavelet transforms like DT-CWT, FDCT and UDCT in [25] and thus can be modeled as a pdf of complex vector.

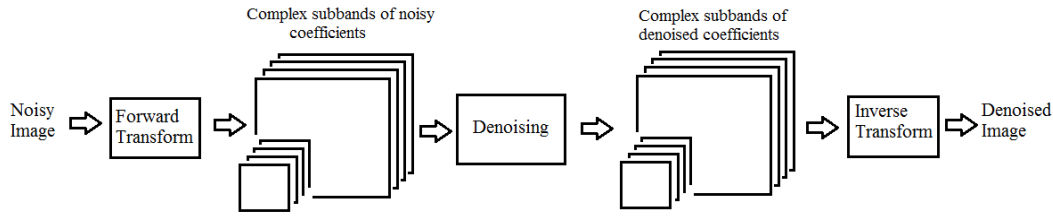


Figure 3.1. Illustration of Transform domain denoising.

Also it's been shown that if \mathbf{X} and \mathbf{Y} are two random vectors of neighborhoods of the real and imaginary parts of complex subband coefficients, respectively, then Y is a GSM, if \mathbf{X} is a GSM, and that both X and Y have the same scalar multiplier.

From the above two results, the pdf of the complex-valued random vector, $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ can be expressed as the product of the zero-mean complex Gaussian random vector $\tilde{\mathbf{Z}} = \tilde{\mathbf{X}} + j\tilde{\mathbf{Y}}$ and the unit mean S as

$$\mathbf{Z} = \sqrt{S}\tilde{\mathbf{Z}} \quad (3.7)$$

Conditioned on S , $\mathbf{Z}|\{S = s\} = \sqrt{s}\tilde{\mathbf{Z}}$ is complex Gaussian with complex covariance matrix $\mathbf{C}_{\mathbf{Z}|S} = s\mathbf{C}_{\tilde{\mathbf{Z}}}$, where $\mathbf{C}_{\tilde{\mathbf{Z}}} = E[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^H] = 2(\mathbf{C}_{\tilde{\mathbf{X}}} - j\mathbf{C}_{\tilde{\mathbf{X}}\tilde{\mathbf{Y}}})$. Hence, \mathbf{Z} is said to be Complex GSM.

3.4 Application of CGSM for image denoising

Formulation of the image denoising problem is illustrated in the figure (3.1).

Assuming the CGSM model for the subband image coefficients, the subband neighborhood coefficients \mathbf{V} of an image corrupted with additive Gaussian noise can be modeled as

$$\mathbf{V} = \mathbf{Z} + \mathbf{W} \quad (3.8)$$

where, Z and W are the subband neighborhood coefficients of the original image and the noise respectively. From equation 3.7, this can be written as,

$$\mathbf{V} = \sqrt{s}\tilde{\mathbf{Z}} + \mathbf{W} \quad (3.9)$$

Both $\tilde{\mathbf{Z}}$ and \mathbf{W} are zero-mean Gaussian vectors, with covariance $C_{\tilde{Z}}$ and C_W . When conditioned on $S = s$, the random vector \mathbf{V} is a zero mean Gaussian with covariance $C_{V|s} = sC_{\tilde{Z}} + C_W$ and its pdf is given by,

$$f_{V|S}(\mathbf{v}|s) = \frac{\exp(-\mathbf{v}^H(sC_{\tilde{Z}} + C_W)^{-1}\mathbf{v})}{(\pi)^N |sC_{\tilde{Z}} + C_W|} \quad (3.10)$$

For each neighborhood, we estimate the reference coefficient Z_c at the center of the neighborhood from \mathbf{V} . Bayes Least Squares (BLS) estimator, estimates Z_c from the observed noisy subband coefficients as,

$$\mathbf{E}\{Z_c|\mathbf{V}\} = \int_0^\infty f_{S|V}(s|\mathbf{v})\mathbf{E}\{Z_c|\mathbf{V} = \mathbf{v}, S = s\}ds \quad (3.11)$$

where, $\mathbf{E}\{.\}$ is the expectation operator and $\mathbf{E}\{Z|\mathbf{V}, S = s\}$ is given by local Wiener estimate as,

$$\mathbf{E}\{Z|\mathbf{V} = \mathbf{v}, S = s\} = sC_{\tilde{Z}}(sC_{\tilde{Z}} + c_W)^{-1}\mathbf{v} \quad (3.12)$$

The distribution of the multiplier in equation 3.11 is computed using Bayes rule as,

$$f_{S|V}(s|\mathbf{v}) = \frac{f_{V|S}(\mathbf{v}|s)f_S(s)}{\int_0^\infty f_{V|S}(\mathbf{v}|\alpha)f_S(\alpha)d\alpha} \quad (3.13)$$

where, we choose the prior $f_S(s)$ to be Jeffrey's noninformative prior as used in [14], and is obtained as

$$f_S(s) \propto \frac{1}{s}$$

and $f_{V|S}(\mathbf{v}|s)$ is given by equation 3.10

In this way, the estimate of the center coefficient is obtained for each neighborhood. The inverse DT-CWT is applied on these estimated subband coefficients to get the denoised image.

CHAPTER 4

CGSM BASED DENOISING FOR MICROARRAY IMAGES

4.1 Introduction

In this chapter, we introduce how we extend the Bayesian least square(BLS) estimator based on the CGSM model, for denoising the two channel microarray images jointly. The advantage of this approach is that, denoised subband coefficients of each image is estimated from the neighborhood coefficients taken from both the images. This gives a better estimate than just using neighborhood of one image because of the existence of hidden correlations between the signal and the noise of the two channel microarray images.

4.2 Proposed Joint denoising algorithm

In this proposed method, we incorporate the joint statistics of the complex coefficients of two channel images in the Bayesian least square estimator using CGSM model and simultaneously denoise them together. This is illustrated in the figure (4.1)

If V_r , Z_r and W_r are the complex neighborhood subband coefficient of the observed noisy image, actual image and the noise of the red channel image respectively,

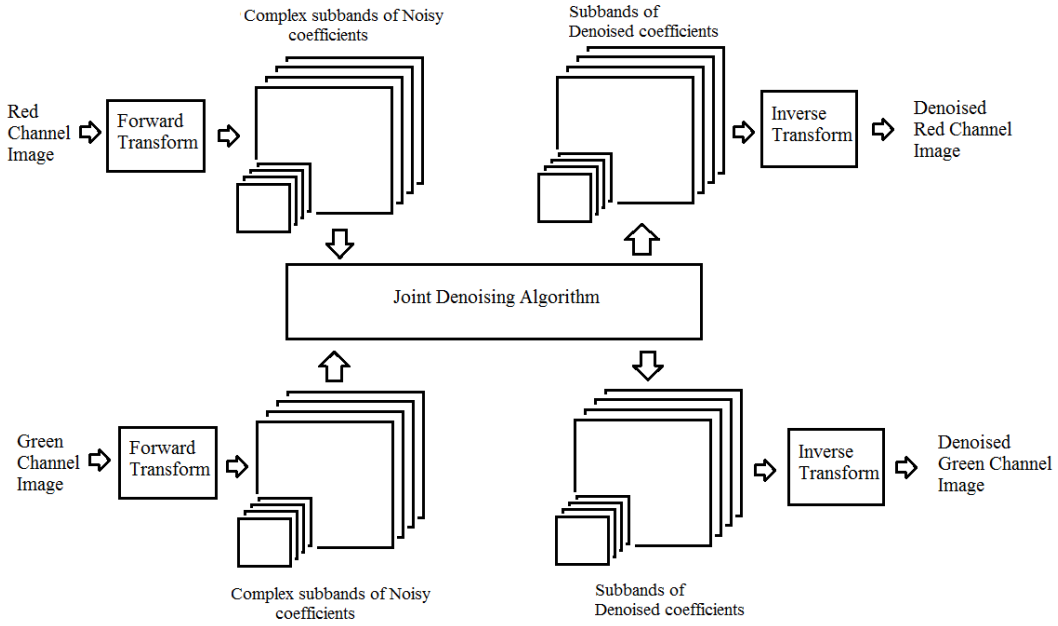


Figure 4.1. Illustration of Joint denoising in complex wavelet domain.

and if V_g , Z_g and W_g are the corresponding quantities for the green channel image, then we can write,

$$V_r = Z_r + W_r$$

$$V_g = Z_g + W_g$$

By CGSM model equation we can write,

$$V_r = \sqrt{S_1} \tilde{Z}_r + W_r$$

$$V_g = \sqrt{S_2} \tilde{Z}_g + W_g$$

where, \tilde{Z}_r and \tilde{Z}_g are zero mean complex Gaussian vectors and S_1 and S_2 are their hidden multipliers respectively. We augment these neighborhood subband coefficients from both the images one below the other to generate a new vector V and W given

by $V = [V_r^T \ V_g^T]^T$ and $W = [W_r^T \ W_g^T]^T$ respectively.

We propose to augmented these neighborhood subband coefficients as

$$V = \begin{bmatrix} V_r \\ V_g \end{bmatrix} \quad W = \begin{bmatrix} W_r \\ W_g \end{bmatrix}$$

and write as

$$V = Z + W$$

Then, by CGSM model equation for the subband coefficients, we can write

$$V = \sqrt{S}\tilde{Z} + W \quad (4.1)$$

where, S is the hidden multiplier for both the images and the \tilde{Z} is the zero mean complex Gaussian pdf. Both \tilde{Z} and W are zero-mean complex Gaussian vectors, with associated complex covariance matrices $C_{\tilde{Z}} = E[\tilde{Z}\tilde{Z}^H]$ and $C_W = E[WW^H]$.

For us to be able to jointly model the subband coefficients as CGSM, S_1 and S_2 are the hidden multipliers of the two images when modeled independently by CGSM, should be equal i.e., $S_1 = S_2$. figure (4.2(a)) shows the plot of the conditional histogram of the estimates \hat{S}_2 given \hat{S}_1 and figure (4.2(b)) between the joint \hat{S} given \hat{S}_1 obtained by the maximum likelihood estimator in each neighborhood.

We observe that $\hat{S}_1 \approx \hat{S}_2$. Thus, we can combine the coefficients from the neighborhood of both the images and model them as a CGSM with the same hidden multiplier S .

The pdf of the observed neighborhood vector conditioned on s is zero-mean complex Gaussian with covariance

$$\mathbf{C}_{V|s} = s\mathbf{C}_{\tilde{Z}} + \mathbf{C}_W \quad (4.2)$$

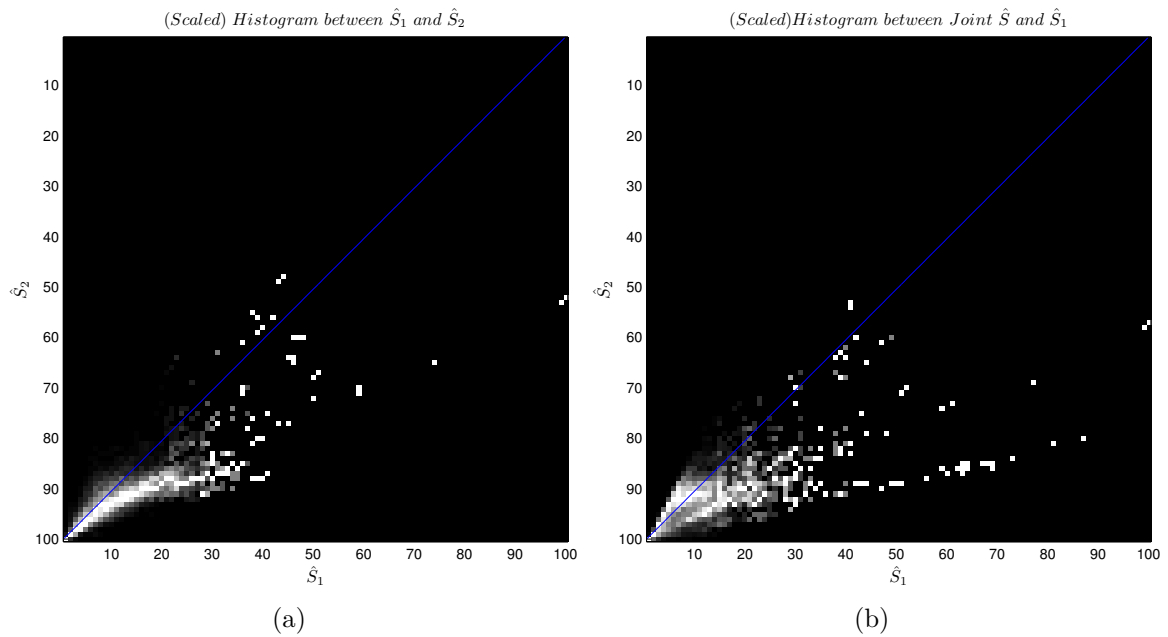


Figure 4.2. Conditional histogram of the hidden multipliers between (a) \hat{S}_1 and \hat{S}_2 and (b) Joint \hat{S} and \hat{S}_1 .

is given by

$$f_{V|S}(v|s) = \frac{\exp(-\mathbf{v}^H(\mathbf{s}\mathbf{C}_{\hat{\mathbf{z}}} + \mathbf{C}_{\mathbf{w}})^{-1}\mathbf{v})}{\pi^N |z\mathbf{C}_{\hat{\mathbf{z}}} + \mathbf{C}_{\mathbf{w}}|}. \quad (4.3)$$

From (4.1), the covariance of observed noisy vector C_V can be computed from $C_{V|S}$ in (4.2) by taking expectations over S :

$$C_V = \mathbf{E}\{S\}C_{\hat{\mathbf{z}}} + C_{\mathbf{w}}$$

If we set $\mathbf{E}\{S\} = 1$, we get,

$$C_{\hat{\mathbf{z}}} = C_V - C_{\mathbf{w}} \quad (4.4)$$

Then BLS estimator, as discussed, is the conditional expectation given by,

$$\hat{\mathbf{Z}} = E[\mathbf{Z}|\mathbf{V}] = \int_0^\infty f_{S|V}(s|\mathbf{v})E[\mathbf{Z}|\mathbf{V} = \mathbf{v}, S = s] ds, \quad (4.5)$$

where $f_S(s)$ is the pdf of the positive scalar random variable S . In the implementation, the integration in (4.5) is computed numerically, where K is the number of points for s , by

$$E[\mathbf{Z}|\mathbf{V}] = \sum_{k=1}^K f_{S|\mathbf{V}}(s_k|\mathbf{v})E[\mathbf{Z}|\mathbf{V} = \mathbf{v}, S = s_k], \quad (4.6)$$

When conditioned on S and \mathbf{V} , the conditional expectation is obtained by the local Wiener estimate as

$$E[\mathbf{Z}|\mathbf{V} = \mathbf{v}, S = s] = s\mathbf{C}_{\hat{\mathbf{Z}}}(s\mathbf{C}_{\hat{\mathbf{Z}}} + \mathbf{C}_{\mathbf{W}})^{-1}\mathbf{v}, \quad (4.7)$$

To estimate \mathbf{Z} , $f_{S|\mathbf{V}}(s|\mathbf{v})$ as in (4.6) is computed as follows

$$f_{S|\mathbf{V}}(s|\mathbf{v}) = \frac{f_{\mathbf{V}|S}(\mathbf{v}|s)f_S(s)}{\int_0^\infty f_{\mathbf{V}|S}(\mathbf{v}|\alpha)f_S(\alpha) d\alpha}, \quad (4.8)$$

where we choose the prior $f_S(s)$ to be Jeffrey's noninformative prior [14] for the experiments in this paper.

Substituting equation (4.7), (4.3), (4.8) in equation (4.5), we estimate the denoised subband complex coefficients. The estimated $\hat{\mathbf{Z}}$ vector has estimates of neighborhood of subband coefficients for both the images. We extract the two center coefficients, one for each neighborhood of subband coefficients of each image. Thus, we get two estimates of coefficients for each of the images from one neighborhood. Having computed these denoised coefficients for all the neighborhood, we get denoised red and green channel images from them by inverse DT-CWT.

Summary of our denoising algorithm:

1. Decompose the red and the green channel images into subbands.
2. For each subband except the lowpass residual band
 - (a) Compute the joint neighborhood noise covariance, $\mathbf{C}_{\mathbf{W}}$, by augmenting the coefficients from the subband of red and green channel image, one below the other. This is computed for each combination of the values of σ_ϵ and ρ of noise in pixel domain.
 - (b) Estimate the joint observed noisy neighborhood covariance, $\mathbf{C}_{\mathbf{V}}$ similarly.

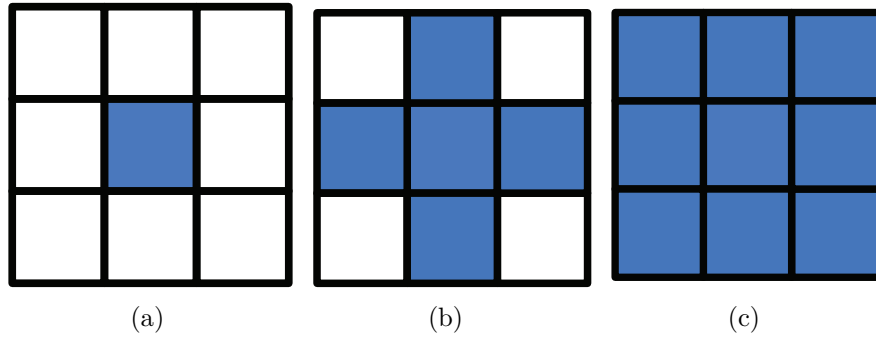


Figure 4.3. Illustration of the subband neighborhood coefficients used for different window sizes (a) w_1 (b) w_5 and (c) w_9 . The number denotes the total number of coefficients used in that neighborhood.

- (c) Estimate $C_{\bar{Z}}$ from $C_{\mathbf{W}}$ and $C_{\mathbf{V}}$ using (4.4).
 - (d) For each neighborhood:
 - i. For each value of $S = s_k$ in the summation range of (4.6):
 - A. Compute $\mathbf{E}\{Z|V, s\}$ using (4.7).
 - B. Compute $f_{V|s}$ using (4.3).
 - ii. Compute $f_{s|V}$ using (4.8) with Jeffery's non-informative prior.
 - iii. Compute $\mathbf{E}\{Z|V\}$ using (4.6) and extract Z_{cr}, Z_{cg} , the center coefficients of the neighborhood for red and green channel subband, respectively.
3. Reconstruct the denoised images of red and green channel from the processed subbands and their lowpass residuals.

4.3 Experimental results

The goal of this section is to present results of the experiments to demonstrate the effectiveness of cDNA microarray denoising using the proposed method. We first discuss the experimental setup, then we study our results using three experiments. In experiment I, we fix the transform (DT-CWT) and vary the window sizes to find

best local neighborhood size for denoising the microarray images, in experiment II we compare different methods of microarray image denoising, and in experiment III, we fix our proposed model and vary the complex wavelet transform used.

4.3.1 Experimental Setup

For denoising performance analysis, we need the original noise-free images. Since these are not available in practice, we pick images that visually appear to be corrupted with very little noise, as the noise-free images. We used ten such pairs of two channel microarray images downloaded from the website of the Stanford Microarray Database (SMD)[27] and selected them as described in[13] i.e., 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. These images are in 16-bit TIFF format and were cropped to a size of 1024 x 1024. Noisy images were then created by adding bivariate Gaussian noise to each of the selected red and green channel images, considering four values of ρ of 0, 0.25, 0.50 and 0.75 and three values of σ_ϵ , 800, 1200 and 1600. The covariance matrix for each subband coefficients of the noise image (noise added to the zero intensity image) for various combinations of σ_ϵ and ρ were computed offline over 100 iterations and saved. The performance of the noise reduction algorithm is evaluated in terms of the Peak Signal to Noise Ratio (PSNR) and Structural Similarity(SSIM) index[28]. In this paper, the PSNR value (in dB) is defined as

$$\text{PSNR} = 20 \log_{10}(65535/\sqrt{(MSE)})$$

where MSE is the mean square error of the denoised image. Denoising algorithm should be able to reproduce the original clean image with least mean squared error. Both PSNR and SSIM, quantitatively define the ability of the denoising algorithm to produce an image as close to the original image.

4.3.2 *Experiment I: Best window size estimation*

We compare different window sizes for the proposed joint CGSM based method. The goal of this exercise is to find the optimum size of the window for denoising these microarray images. For this we have considered different combination of window sizes both symmetric and asymmetric. Symmetric and asymmetric are classified on the basis of number of coefficients used from the two subband images to denoise each image. For symmetric case, equal number of subband coefficients are taken from subbands of both the images while for the asymmetric case, unequal number of subband coefficients are taken to denoise. The different coefficient window sizes are depicted in figure (4.3)

As explained earlier, the coefficients in the window are augmented one below the other for defining the neighborhood for denoising. For example, in joint denoising using a 3×3 window, the 9 coefficients of green channel subband image are augmented below those of the red channel subband image. For (w5) & (w1) case, to denoise red channel image 5 coefficients in plus-shaped neighborhood ((i.e., center and the adjacent vertical and horizontal coefficients, denoted henceforth as w5)) of red channel is combined with the center coefficient (w1) of the green channel (denoted henceforth as w5 & w1), thereby using 6 coefficients to denoise the red channel and vice versa for green channel image. This is illustrated in the figures (4.4) and (4.5) for an example window sizes of (w5) and (w5)&(w1).

We have considered here six different window sizes for proposed joint CGSM based denoising:

1. (w9)& (w9): 3×3 window of coefficients from both images.
2. (w5)& (w5): plus-shaped window(center, top,bottom,left and right coefficients) from both images.
3. (w1)& (w1): center coefficient alone from both images.
4. (w9)& (w5): 3×3 window from image to be denoised and plus-shaped window from other image.
5. (w9)& (w1): 3×3 window from image to be denoised and plus-shaped window from other image.
6. (w5)& (w1): plus-shaped window from image to be denoised and center coefficient from other image.

The PSNR and SSIM values for the joint CGSM using different window sizes are as shown in TABLE 4.1. We find that for the images considered joint window (w5)& (w1) performs the best. As we increase the window size for either image, the PSNR decreases. We suspect that, including more coefficients in the neighborhood, adds redundant information and does not help in denoising.

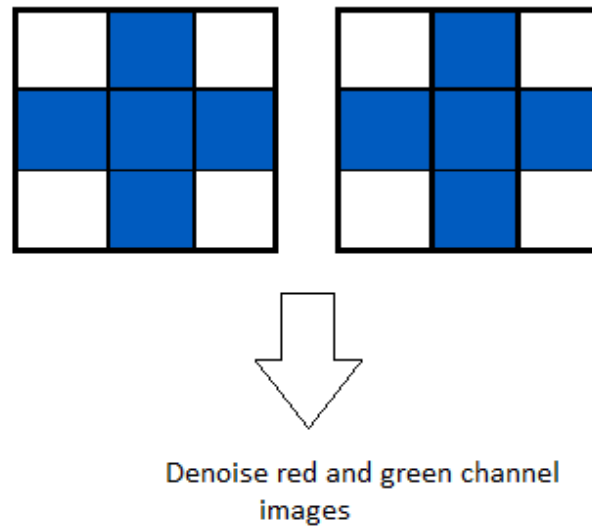


Figure 4.4. Illustration of the use of symmetric window (w_5) for joint denoising of red and green channel images.

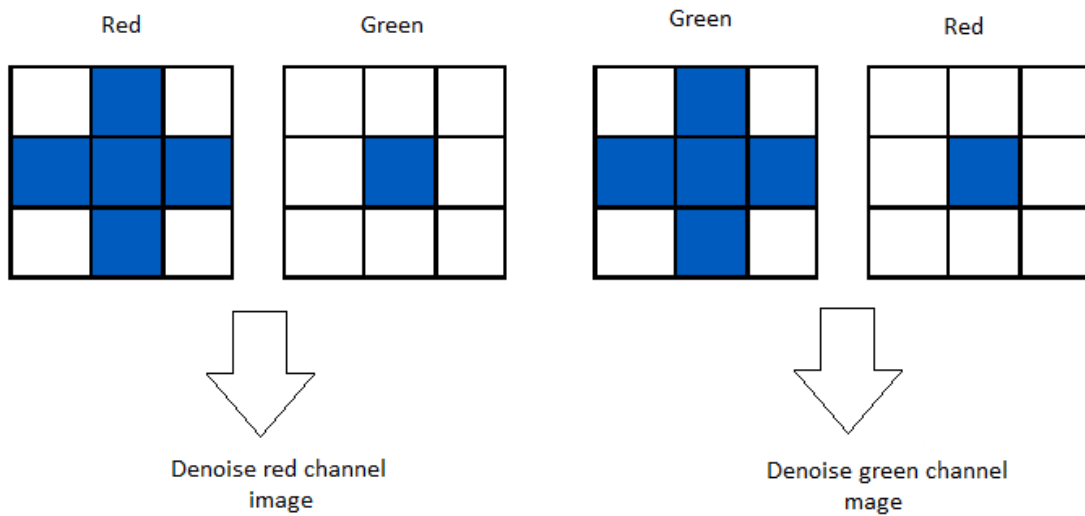


Figure 4.5. Illustration of the use of asymmetric window (w_5) & (w_1) for joint denoising of red and green channel images.

Table 4.1. PSNR values (in dB) and SSIM indices of denoised images using DTCWT-CGSM with different window sizes

σ	PSNR of Red and Green channel images			SSIM of Red and Green channel images		
	800	1200	1600	800	1200	1600
Noisy	38.27,38.27	34.75 ,34.75	32.25 ,32.25	0.873, 0.872	0.760, 0.758	0.649, 0.64
Methods	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G
Window :(w9)&(w9)						
$\rho_\epsilon = 0.0$	45.26,45.09	42.71, 42.52	40.95, 40.73	0.969, 0.966	0.945, 0.939	0.919, 0.910
$\rho_\epsilon = 0.25$	45.00,44.83	42.42, 42.22	40.63, 40.40	0.968, 0.965	0.942, 0.936	0.915, 0.906
$\rho_\epsilon = 0.50$	44.79,44.63	42.19, 42.00	40.39, 40.17	0.966, 0.964	0.939, 0.934	0.911, 0.903
$\rho_\epsilon = 0.75$	44.64,44.51	42.02, 41.87	40.21, 40.04	0.965, 0.963	0.937, 0.933	0.908, 0.901
Average	44.92,44.77	42.33, 42.15	40.55, 40.34	0.967, 0.964	0.941, 0.935	0.913, 0.905
Window :(w5)&(w5)						
$\rho_\epsilon = 0.0$	45.66,45.49	43.09, 42.90	41.31, 41.09	0.972, 0.969	0.949, 0.943	0.924, 0.916
$\rho_\epsilon = 0.25$	45.39,45.22	42.79, 42.60	40.98, 40.77	0.970, 0.967	0.946, 0.941	0.920, 0.912
$\rho_\epsilon = 0.50$	45.19,45.03	42.56, 42.39	40.74, 40.53	0.969, 0.966	0.944, 0.939	0.917, 0.909
$\rho_\epsilon = 0.75$	45.05,44.92	42.42, 42.27	40.58, 40.41	0.968, 0.966	0.942, 0.937	0.914, 0.907
Average	45.32,45.17	42.71, 42.54	40.90, 40.70	0.970, 0.967	0.945, 0.940	0.919, 0.911
Window :(w1)&(w1)						
$\rho_\epsilon = 0.0$	45.64,45.47	43.00, 42.75	41.17, 40.85	0.973, 0.971	0.951, 0.946	0.928, 0.919
$\rho_\epsilon = 0.25$	45.32,45.13	42.65, 42.37	40.81, 40.46	0.972, 0.969	0.948, 0.943	0.923, 0.914
$\rho_\epsilon = 0.50$	45.10,44.91	42.41, 42.13	40.55, 40.20	0.970, 0.968	0.946, 0.940	0.919, 0.911
$\rho_\epsilon = 0.75$	45.03,44.86	42.32, 42.09	40.45, 40.15	0.970, 0.967	0.944, 0.940	0.917, 0.910
Average	45.27,45.09	42.59, 42.33	40.74, 40.42	0.971, 0.969	0.947, 0.942	0.922, 0.913
Window :(w9)&(w5)						
$\rho_\epsilon = 0.0$	45.42,45.24	42.88, 42.65	41.13, 40.85	0.970, 0.967	0.946, 0.940	0.921, 0.912
$\rho_\epsilon = 0.25$	45.15,44.95	42.58, 42.34	40.80, 40.52	0.968, 0.966	0.943, 0.938	0.916, 0.908
$\rho_\epsilon = 0.50$	44.94,44.75	42.34, 42.11	40.54, 40.27	0.967, 0.965	0.940, 0.935	0.912, 0.905
$\rho_\epsilon = 0.75$	44.80,44.64	42.19, 42.00	40.38, 40.15	0.966, 0.964	0.938, 0.934	0.910, 0.903
Average	45.08,44.90	42.50, 42.28	40.71, 40.45	0.968, 0.965	0.942, 0.937	0.915, 0.907
Window :(w9)&(w1)						
$\rho_\epsilon = 0.0$	45.42,45.26	42.88, 42.66	41.12, 40.85	0.970, 0.968	0.945, 0.941	0.920, 0.913
$\rho_\epsilon = 0.25$	45.17,44.99	42.60, 42.36	40.82, 40.52	0.968, 0.966	0.943, 0.939	0.916, 0.909
$\rho_\epsilon = 0.50$	44.97,44.80	42.38, 42.15	40.59, 40.30	0.967, 0.965	0.940, 0.936	0.912, 0.906
$\rho_\epsilon = 0.75$	44.85,44.70	42.24, 42.04	40.44, 40.19	0.966, 0.965	0.938, 0.935	0.909, 0.904
Average	45.10,44.94	42.53, 42.30	40.74, 40.47	0.968, 0.966	0.942, 0.938	0.914, 0.908
Window :(w5)&(w1)						
$\rho_\epsilon = 0.0$	45.76,45.61	43.19, 42.99	41.40, 41.15	0.972, 0.970	0.949, 0.945	0.925, 0.918
$\rho_\epsilon = 0.25$	45.51,45.34	42.91, 42.69	41.10, 40.83	0.970, 0.969	0.946, 0.942	0.921, 0.914
$\rho_\epsilon = 0.50$	45.32,45.16	42.70, 42.49	40.87, 40.61	0.969, 0.968	0.944, 0.940	0.917, 0.911
$\rho_\epsilon = 0.75$	45.21,45.07	42.57, 42.39	40.74, 40.51	0.968, 0.967	0.943, 0.939	0.915, 0.910
Average	45.45,45.29	42.84, 42.64	41.03, 40.78	0.970, 0.968	0.946, 0.942	0.920, 0.913

4.3.3 *Experiment II*: Comparison with other denoising methods

4.3.3.1 Experiment II-A

We compare seven methods of image denoising:

1. *Separate BLS-CGSM*[25]: CGSM based denoising in DTCWT domain of each channel image independently. It considers a w5 window for the image to be denoised.
2. *Separate BLS-GSM*[14]: GSM based denoising using Full Steerable(FS) transform of each channel image independently. It also considers a w5 window for the image to be denoised.
3. *Bishrink* [29]: It uses the default window size of 3x3 (9 coefficients).
4. *Separate DWT BLS-GSM*: GSM based denoising using Discrete Wavelet Transform (DWT) of each channel image independently. It also considers a w5 window for the image to be denoised.
5. *Joint DWT BLS-GSM*: GSM based denoising in DWT domain using joint information from the two channel images. It considers the joint window size of w5& w1.
6. *Joint BLS-GSM*: GSM based denoising in FS domain using joint information from the two channel images. It also considers a window of w5 & w1 for denoising.
7. *Joint BLS-CGSM(Proposed)*: CGSM based denoising in DTCWT domain using joint information from the two channel images using a window size of w5 & w1.

In each of these methods, the number of levels of decomposition and the neighborhood window size were chosen such that they yield the best results. DWT methods were implemented with 5 levels of decomposition, which was best for these images. BLS-CGSM methods and Bishrink using DTCWT were implemented with 5 levels of

Table 4.2. PSNR values (in dB) and SSIM indices of denoised images from several methods

σ	PSNR of Red and Green channel images			SSIM of Red and Green channel images		
	800	1200	1600	800	1200	1600
Noisy	38.27,38.27	34.75,34.75	32.25,32.25	0.873, 0.872	0.760, 0.758	0.649, 0.646
Methods	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G	Im-R Im-G
Bishrink:	44.59,44.50	41.87,41.73	39.96,39.80	0.965,0.964	0.934,0.933	0.901,0.899
Separate DWT:(w5)	43.91,43.91	41.13,41.11	39.21,39.16	0.961,0.961	0.929,0.928	0.895,0.894
Separate FS-GSM:(w5)	44.87,44.64	42.33,42.09	40.54,40.30	0.966,0.965	0.940,0.937	0.912,0.908
Separate CGSM:(w5)	45.03,44.97	42.35,42.27	40.48,40.38	0.967,0.967	0.940,0.939	0.911,0.909
Joint DWT: (w5)&(w1)						
$\rho_1 = 0.0$	44.66,44.54	42.00, 41.80	40.16, 39.89	0.967, 0.965	0.939, 0.935	0.911, 0.903
$\rho_1 = 0.25$	44.40,44.28	41.71, 41.51	39.85, 39.58	0.965, 0.963	0.936, 0.932	0.906, 0.899
$\rho_1 = 0.50$	44.21,44.09	41.50, 41.31	39.62, 39.36	0.963, 0.962	0.934, 0.930	0.902, 0.896
$\rho_1 = 0.75$	44.11,44.00	41.37, 41.21	39.49, 39.27	0.963, 0.961	0.932, 0.929	0.899, 0.894
Average	44.35,44.23	41.65, 41.46	39.78, 39.53	0.964, 0.963	0.935, 0.931	0.904, 0.898
Joint GSM: (w5)&(w1)						
$\rho_1 = 0.0$	45.58,45.26	43.12, 42.77	41.39, 41.02	0.971, 0.968	0.949, 0.943	0.926, 0.917
$\rho_1 = 0.25$	45.33,45.00	42.85, 42.48	41.11, 40.72	0.970, 0.967	0.946, 0.941	0.921, 0.913
$\rho_1 = 0.50$	45.14,44.82	42.64, 42.29	40.90, 40.51	0.968, 0.965	0.944, 0.939	0.918, 0.910
$\rho_1 = 0.75$	45.02,44.76	42.52, 42.22	40.76, 40.44	0.967, 0.965	0.942, 0.938	0.916, 0.909
Average	45.27,44.96	42.78, 42.44	41.04 , 40.67	0.969, 0.966	0.945,0.940	0.920 ,0.912
Joint CGSM: (w5)&(w1)						
$\rho_1 = 0.0$	45.76,45.61	43.19, 42.99	41.40, 41.15	0.972, 0.970	0.949, 0.945	0.925, 0.918
$\rho_1 = 0.25$	45.51,45.34	42.91, 42.69	41.10, 40.83	0.970, 0.969	0.946, 0.942	0.921, 0.914
$\rho_1 = 0.50$	45.32,45.16	42.70, 42.49	40.87, 40.61	0.969, 0.968	0.944, 0.940	0.917, 0.911
$\rho_1 = 0.75$	45.21,45.07	42.57, 42.39	40.74, 40.51	0.968, 0.967	0.943, 0.939	0.915, 0.910
Average	45.45,45.29	42.84 , 42.64	41.03, 40.78	0.970 , 0.968	0.946 , 0.942	0.920 , 0.913

decomposition and 6 orientations in each level while BLS-GSM using FS was implemented with 5 level decomposition with 8 orientations. For all GSM/CGSM based methods, that denoise images independently, the optimized window size was found to w5. For the joint denoising schemes based on GSM/CGSM, a plus-shaped window for the image to denoised and the center coefficient of the other channel image (w5 & w1) was found to be best. TABLE 4.2 shows the output PSNR and SSIM values for the seven denoising methods obtained by averaging the results of ten microarray images. From this data, the following observations can be made.

First, among the various methods, the proposed joint CGSM based method outperforms its independent counterpart both in terms of PSNR and SSIM values. This is because, the proposed model considers both intra-subband and inter-channel dependency between the complex wavelet coefficients while denoising. These indicate that proposed joint model is better fit for the subband coefficients of microarray images.

Second, in both separate and joint methods, the complex wavelets perform better than real wavelets. This is because, for microarray image denoising shift invariance property and ability to better represent singularity are desirable which are provided by complex wavelets.

Third, though joint GSM method using FS, gives results close to our proposed joint CGSM, the complexity of our method is comparatively very less with faster computational time. A comparison of the redundancy and computational time taken by these transforms reveal that, FS is approximately four time more redundant than DT-CWT and take four times the computational time of DT-CWT as shown.

Table 4.3. Redundancy ratios of FS and DT-CWT

Transform	FS	DT-CWT
Redund. Ratio	≈ 18.67	4

$$\frac{\textit{Computation time using FS}}{\textit{Computation time using DT - CWT}} = \frac{466s}{108s} \approx 4$$

Fourth, from the experiments I and II-A we observe that, the inter-channel correlation is more stronger than the inter-scale dependency for these images. This is evident from the comparison of the performance of joint CGSM with window (w1)&(w1) and Bishrink.

4.3.3.2 Experiment II-B

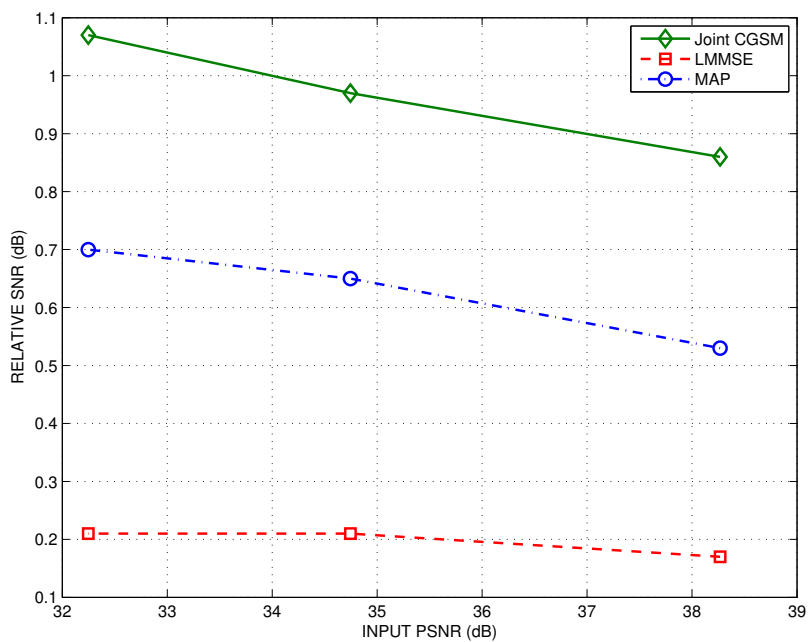
Next, we now compare the performance of our proposed method with linear minimum mean square error (LMMSE) and Maximum *a posteriori* (MAP) estimators proposed in [13]. The methods are compared under two benchmarks.

1. Bishrink[29]
2. BLS-GSM[14]

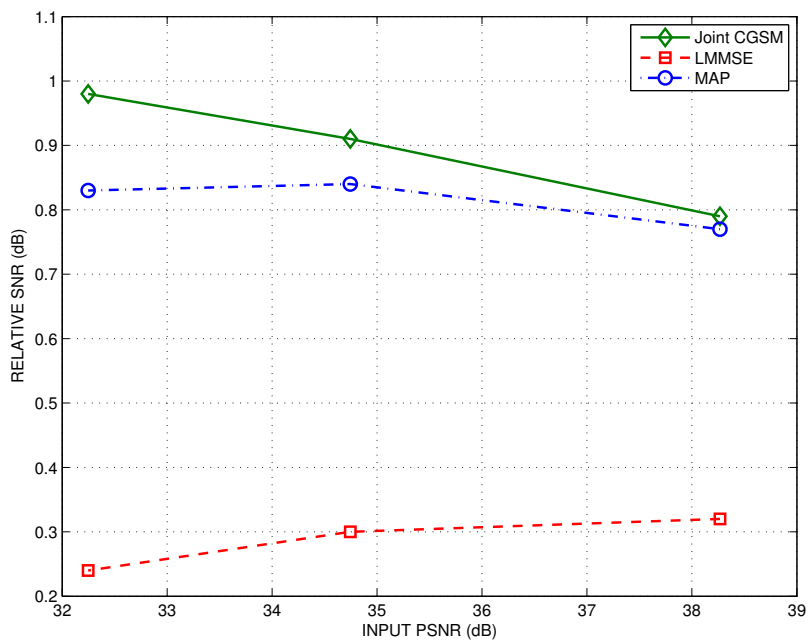
In this analysis, we compared the PSNR differences of the LMMSE and MAP estimators and our proposed method with respect to Bishrink and BLS-GSM for the red and green channel images. We see from the plot that our proposed method, consistently performs better than both the estimators as shown in the figures (4.3.3.2) and (4.3.3.2). This is because, these estimators consider only magnitude information and ignore phase information of complex coefficients while denoising. Also it assumes a different model by considering a Gaussian distribution for the magnitude of complex coefficients while our proposed method models the complex coefficients as complex GSM. It also assumes an equivalence between the magnitude of observed subband coefficients and the additive model of the magnitude of complex coefficients of original signal and noise, while in reality this equivalence is true only for their complex counterparts. From this we conclude that, though LMMSE and MAP methods also considers the inter-channel dependency between subband coefficients, our model is better representation for complex subband coefficients of these images.

4.3.4 *Experiment III*: Comparison with other transforms

In this experiment, we fix our proposed model i.e., joint CGSM with best window size of (w5)& (w1) and vary the transform. We have used 4 different complex wavelet transforms:

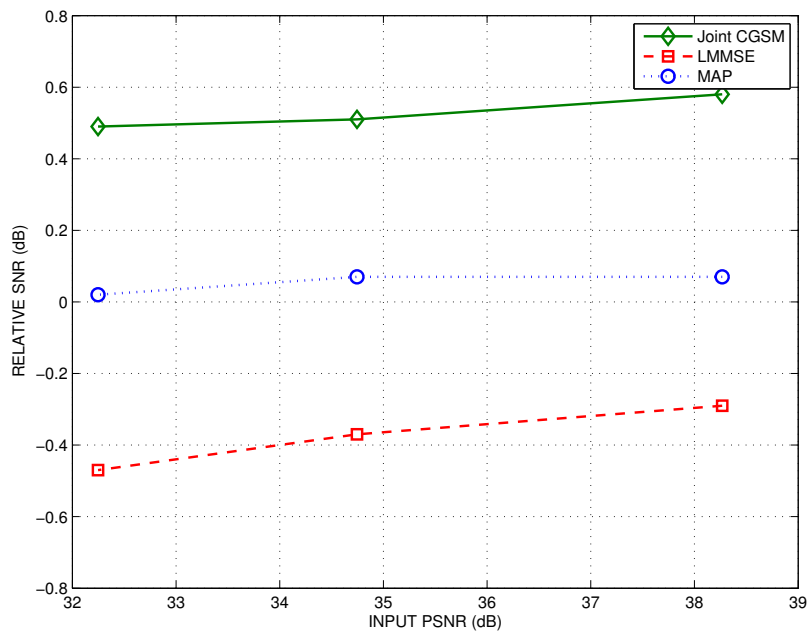


(a)

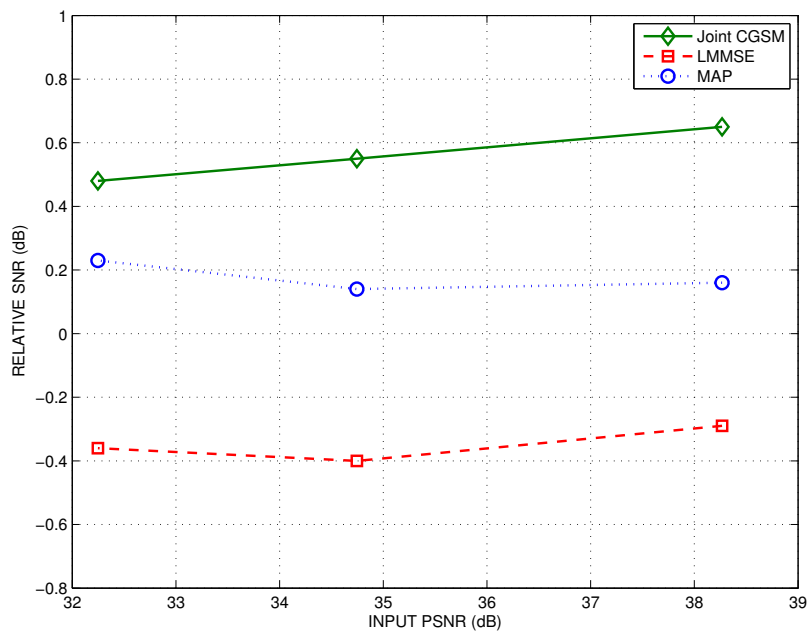


(b)

Figure 4.6. Plot of the PSNR differences (in dB) of our proposed method, LMMSE and MAP estimators as a function of the input PSNR with respect to Bishrink method. (a) Red channel image (b) Green channel image.



(a)



(b)

Figure 4.7. Plot of the PSNR differences (in dB) of our proposed method, LMMSE and MAP estimators as a function of the input PSNR with respect to BLS-GSM method. (a) Red channel image (b) Green channel image.

1. Dual Tree - Complex Wavelet Transform (DT-CWT)[30].
2. Fast Discrete Curvelet Transform (FDCT)[31].
3. Uniform Discrete Curvelet Transform (UDCT)[32]
4. Pyramidal Dual-Tree Directional Filter Bank(PDTDFB)[33].

The PSNR and SSIM values for the joint CGSM using different complex transforms are as shown in TABLE 4.5. For each of these transforms, the level of decomposition and orientation were chosen to give the best result in terms of PSNR and SSIM. As it can be seen, DT-CWT performs the best compared to all other transforms. This is a very interesting observation. We suspect this behavior is because of the nature of images. Unlike natural images, microarray images are made up spots which are very small in size (approximately of size 20 or less) and are isotropic in nature. The dictionary of wavelets is made up of roughly isotropic elements occurring at all scales and positions. The area of expertise of wavelet and curvelet transform has been studied by many and in particular it has been pointed out in [34] that curvelets are computationally efficient for geometric features with line and surface singularities, while wavelets efficiently represent small features (such as eyes of Lena). We suspect that this weakness of the curvelets is witnessed here, where the small features of the image (spots) are well represented by wavelets. Thus, choice of the transform can affect the denoising performance of this model.

The redundancy ratios of the various transforms used in as show in TABLE 4.4

Table 4.4. Redundancy ratios of the transforms used herein

Transform	FS	UDCT	DT-CWT	PDTDFB	FDCT
Redund. Ratio	≈ 18.67	4	4	≈ 2.67	≈ 14.53

Table 4.5. PSNR values (in dB) and SSIM indices of denoised images using Joint CGSM(Window:(5)&(1)) with different transforms

σ	PSNR values Red and Green channel images						SSIM values Red and Green channel images					
	800		1200		1600		800		1200		1600	
Noisy	38.27,38.27		34.75 ,34.75		32.25 ,32.25		0.873, 0.872		0.760, 0.758		0.649, 0.646	
Methods	Im-R	Im-G	Im-R	Im-G	Im-R	Im-G	Im-R	Im-G	Im-R	Im-G	Im-R	Im-G
DTCWT :(5)&(1)												
$\rho_\epsilon = 0.0$	45.76, 45.61		43.19, 42.99		41.40, 41.15		0.972, 0.970		0.949, 0.945		0.925, 0.918	
$\rho_\epsilon = 0.25$	45.51, 45.34		42.91, 42.69		41.10, 40.83		0.970, 0.969		0.946, 0.942		0.921, 0.914	
$\rho_\epsilon = 0.50$	45.32, 45.16		42.70, 42.49		40.87, 40.61		0.969, 0.968		0.944, 0.940		0.917, 0.911	
$\rho_\epsilon = 0.75$	45.21, 45.07		42.57, 42.39		40.74, 40.51		0.968, 0.967		0.943, 0.939		0.915, 0.910	
Average	45.45, 45.29		42.84, 42.64		41.03, 40.78		0.970, 0.968		0.946, 0.942		0.920, 0.913	
UDCT :(5)&(1)												
$\rho_\epsilon = 0.0$	44.53, 44.08		42.16, 41.67		40.52, 39.99		0.963, 0.959		0.937, 0.928		0.910, 0.897	
$\rho_\epsilon = 0.25$	44.30, 43.84		41.92, 41.41		40.27, 39.72		0.962, 0.957		0.934, 0.925		0.905, 0.893	
$\rho_\epsilon = 0.50$	44.12, 43.69		41.73, 41.24		40.08, 39.54		0.960, 0.956		0.931, 0.923		0.902, 0.890	
$\rho_\epsilon = 0.75$	44.01, 43.63		41.62, 41.18		39.95, 39.47		0.959, 0.955		0.929, 0.922		0.899, 0.888	
Average	44.24, 43.81		41.86, 41.38		40.20, 39.68		0.961, 0.957		0.933, 0.925		0.904, 0.892	
CDFB :(5)&(1)												
$\rho_\epsilon = 0.0$	44.55 44.25		42.07 41.69		40.37 39.92		0.964 0.960		0.937 0.929		0.910 0.898	
$\rho_\epsilon = 0.25$	44.29 43.97		41.79 41.39		40.08 39.61		0.962 0.958		0.934 0.926		0.905 0.893	
$\rho_\epsilon = 0.50$	44.09 43.79		41.58 41.20		39.86 39.41		0.961 0.957		0.932 0.924		0.901 0.890	
$\rho_\epsilon = 0.75$	43.96 43.71		41.45 41.12		39.72 39.34		0.960 0.956		0.930 0.923		0.899 0.889	
Average	44.22 43.93		41.72 41.35		40.01 39.57		0.962 0.958		0.933 0.926		0.904 0.893	
FDCT :(5)&(1)												
$\rho_\epsilon = 0.0$	44.66, 44.38		42.04, 41.72		40.17, 39.83		0.961, 0.958		0.926, 0.922		0.888, 0.882	
$\rho_\epsilon = 0.25$	44.45, 44.15		41.83, 41.49		39.97, 39.60		0.960, 0.957		0.925, 0.921		0.886, 0.880	
$\rho_\epsilon = 0.50$	44.29, 44.00		41.67, 41.34		39.81, 39.45		0.959, 0.957		0.924, 0.920		0.885, 0.879	
$\rho_\epsilon = 0.75$	44.19, 43.95		41.58, 41.30		39.72, 39.41		0.959, 0.956		0.923, 0.919		0.884, 0.879	
Average	44.40, 44.12		41.78, 41.46		39.92, 39.57		0.960, 0.957		0.925, 0.920		0.886, 0.880	

Figure (4.8) shows the 3-D visualization of a clean spot, a noisy spot with noise of stand deviation $\sigma_\epsilon=1200$ and $rho = 0$, its denoised version using LMMSE and our proposed method. This plot clearly shows that our method removes most of the noise while preserving the edge information of the spots. The spot appears smoothed at the edges by the LMMSE method. Thus, our method effectively preserves the signal information while reducing the noise content in the images as desired.

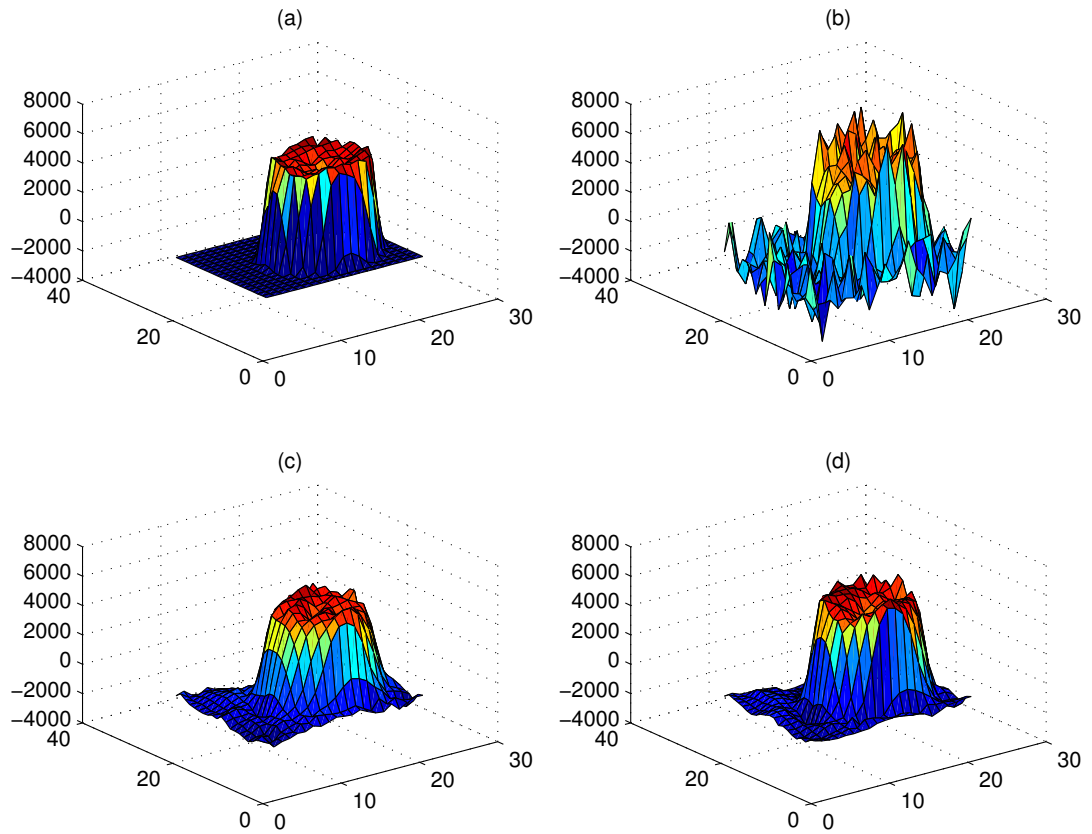


Figure 4.8. 3D plot of a spot from a microarray image. (a) Clean spot (b) Noisy spot with noise of $\sigma_{\epsilon}=1200$ and $\rho = 0$ (c) denoised spot using LMMSE (d) denoised spot using Joint CGSM with window (w5) & (w1). We observe the edge of the circular spot is preserved in our proposed method. .

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Summary

This thesis is concerned with the problem of developing an effective method for noise reduction in cDNA microarray images for the purpose of extracting accurate information regarding gene expression levels. The outcome of two- channel cDNA microarray, are two images, which show the expression level of genes in control and treatment samples in terms of the intensity levels of the gene spots. Measurement of relative gene expression from these samples helps in identifying differentially expressed genes and find the root cause of their behavior. But the success for these experiments depends on extracting meaningful information from these experiments. This poses a great challenge in microarray data analysis as these images are corrupted with high degree of noise accumulated during its synthesis. Unless we can effectively disassociate the required signal from the noise, the intensity measurements from these images will have no significance. Since an error in measurement at this stage propagates down to further stages of gene expression analysis, it is very essential to reduce the noise content in them and this is the main motivation for our work.

Microarray images owing to the way they are generated, carry significant signal and noise correlation between the two channel images. Studies have shown that methods accounting for such correlations are more successful than those that ignore. Among the successful wavelet based methods for denoising, complex wavelets have more desirable properties for denoising microarray images and consequently the CGSM model in the complex wavelet domain provides high quality image denois-

ing. The proposed method in this paper, effectively utilizes the advantages of CGSM model in the framework of noise reduction in microarray images, by incorporating the inter-channel signal and noise correlation that exists between the red and green channel images.

5.2 Conclusion

The three factors that led to the development of this algorithm are as follows. First, the success of wavelet based approaches in image denoising over other transform based approaches or any pixel based approach. Second, among wavelets, complex wavelets due to its redundancy and shift invariance property over the Discrete Wavelet Transform (DWT) are more suitable for noise reduction application. Also, both magnitude and phase of the complex coefficients can be exploited in the statistical framework when working with complex wavelets. Third, algorithms that account for the hidden correlations between the gene signals and the noise in the two channel microarray images have improved performance over those that treat them independently for noise reduction. Thus, we propose to use the CGSM model of the complex coefficients and denoise the two channel microarray images using a Bayesian least square estimator by incorporating the joint statistics of the images into the algorithm. Joint statistics accounts for the correlations that exists between the signal and the noise and the denoised subband coefficient of the both the images are estimated simultaneously. Simulation results show that the proposed method provides better denoising performance in terms of PSNR and SSIM than the existing methods being compared.

From our analysis, we see that, first, our joint CGSM model of complex coefficients of the two channel images performs better than the CGSM model based denoising of these channel images independently, for denoising application. This im-

plies, joint estimation of the denoised subband coefficients is essential for effective noise reduction in these images. The two channel images have inter-channel correlation that must be exploited to efficiently reduce noise. This is accomplished in our algorithm by incorporating the joint covariances of the observed signal and noise of the two channel images. Second, the proposed joint model of complex coefficients is a better fit than the model of magnitude of subband coefficients proposed in [13] as it leads to more effective noise reduction in terms of PSNR and SSIM. Though both methods consider the inter-channel dependency in estimating the denoised subband coefficients, it is important to consider both magnitude and phase information of the subband coefficients for estimating them. Third, the inter-channel correlation is more stronger than the inter-scale dependency for these images. Methods like bishrink, that exploit the inter-scale dependency do not yield high PSNR as the methods that exploit the inter-channel dependency like joint CGSM. Fourth, the choice of the transform used for decomposition is also very important to yield effective noise reduction. From our experimental data, we found DT-CWT to be best suited for these images. Thus, we conclude that our proposed model for noise reduction, will play a significant role in improving the reliability of the results obtained from practical microarray experiments.

5.3 Future work

There are some additional analysis related to this thesis work which are worth investigating. In particular, we would like to investigate the following.

1. We want to apply this algorithm to the denoising of real noisy microarray images from the database by computing the noise characteristics from the higher subbands of the complex transform domain.

2. We observe that the GSM based denoising of microarray images using Full Steerable (FS) performs very close to that CGSM model based on DT-CWT coefficients. We would like to evaluate the performance of our proposed model using complex Full Steerable for noise reduction.

3. In our algorithm, we aim to reduce the additive white Gaussian noise present in the microarray images. But microarray images, in reality, could be contaminated with mixture of noises of varying distribution like impulse noise, Poisson noise, etc and densities. It is necessary to investigate the various noise sources and their distributions. In future, we would like to develop a more general model that can account for these various noise components in these images and help in denoising them more effectively.

APPENDIX A
ABBREVIATION LIST

A	Adenine
AGN	Additive Gaussian Noise
AWGN	Additive White Gaussian Noise
BLS	Bayesian Least Square
CGSM	Complex Gaussian Scale Mixtures
cDNA	Complementary DNA
CWT	Complex Wavelet Transform
Cy3	Cyanine-3
Cy5	Cyanine-5
C	Cytosine
dB	Decibels
db4	Daubechies 4 wavelet function
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DNA	Deoxyribonucleic acid
DT-CWT	Dual-Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
FDCT	Fast Discrete Curvelet Transfrom
FS	Full Steerable
G	Guanine
GSM	Gaussian Scale Mixtures
LMMSE	Linear Minimum Mean Squared Error
MAP	Maximum A Posteriori
mRNA	Messenger RNA
MSE	Mean Squared Error
PDF	Probability Density Function
PDTDFB	Pyramidal Dual-Tree Directional Filter Bank
PSNR	Peak Signal to Noise Ratio

RNA	Ribonucleic Acid
RT	Reverse Transcriptase
SMD	Stanford Microarray Database
SSIM	Structural Similarity Index
SWT	Stationary Wavelet Transform
T	Thymine
TIFF	Tagged Image File Format
tRNA	Transfer RNA
U	Uracil
UDCT	Uniform Discrete Curvelet Transform

REFERENCES

- [1] D. Duggan, M. Bittner, Y. Chen, P. Meltzer, and J. Trent, “Expression profiling using cdna microarrays.”
- [2] A. A. Donald, Y. Zhang, and R. Parthe, “On denoising and compression of dna microarray images,” *Pattern Recognition*, vol. 39, no. 12, pp. 2478–2493, December 2006.
- [3] X. L. Karl Fraser, Zidong Wang, *Microarray Image Analysis An Algorithmic Approach*. Chapman & Hall, 2010.
- [4] (2007) Biodiscovery Inc. Imagene. [Online]. Available: <http://www.biodiscovery.com/imagene.asp>
- [5] (2002) Eisen MB: ScanAlyze. [Online]. Available: <http://rana.lbl.gov/EisenSoftware.htm>
- [6] P. Bajcsy, “An overview of dna microarray grid alignment and foreground separation approaches.”
- [7] B. D. Rahnenfhrer J., “Hybrid clustering for microarray image analysis combining intensity and shape features.”
- [8] O. Demirkaya, M. H. Asyali, and M. M. Shoukri, “Segmentation of cdna microarray spots using markov random field modeling,” vol. 21, pp. 2994–3000, 2005.
- [9] E. I. Athanasiadis, D. A. Cavouras, D. T. Glotsos, P. V. Georgiadis, I. K. Kalatzis, and G. C. Nikiforidis, “Segmentation of complementary dna microarray images by wavelet-based markov random field model,” *IEEE Trans. Inform. Technol. Biomed.*, vol. 13, pp. 1068–1074, 2009.

- [10] A. Lehmussola, P. Ruusuvuori, and O. Yli-Harja, “Evaluating the performance of microarray segmentation algorithms,” vol. 22, pp. 2910–2917, 2006.
- [11] S. A. Q. J, and T. Z, “Genesis: cluster analysis of microarray data,” vol. 18, pp. 207–208, 2002.
- [12] A. Daskalakis, D. Cavouras, P. Bougioukos, D. G. S. Kostopoulos, I. Kalatzis², G. C. Kagadis, C. Argyropoulos, and G. Nikiforidis, “Improving gene quantification by adjustable spot-image restoration.”
- [13] T. Howlader and Y. P. Chaubey, “Noise reduction of cdna micorarray images using complex wavelets,” *IEEE Trans. Image Processing*, vol. 19, no. 8, 2010.
- [14] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Trans. Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.
- [15] G. J. McLachlan, K.-A. Do, and G. J. C.A. McLachlan, *Analyzing microarray gene expression data*. Hoboken, N.J. : Wiley-Interscience, c2004, 2004.
- [16] S. Russell, L. A. Meadows, and R. Russell, *Microarray technology in practice*. Academic Press, Elsevier, 2009.
- [17] K. P. B. Smolka, R. Lukac, “Fast noise reduction in cdna microarray images.”
- [18] B. S. R. Lukac, K. N. Plataniotis and A. N. Venetsanopoulos, “A multichannel order-statistic technique for cdna microarray image processing,” *IEEE Trans. Nanobiosci*.
- [19] R.Lukac and B.Smolka, “Application of the adaptive centerweighted vector median framework for the enhancement of cdna microarray,” *International Journal of Applied Mathematics and Computer Science*, vol. 13, no. 3, pp. 369–383, 2003.
- [20] P. ONeill, G. D. Magoulas, and X. Liu, “Improved processing of microarray data using image reconstruction techniques,” *IEEE Trans. Nanobiosci*.

- [21] X. H. Wang, R. S. H. Istepanian, and Y. H. Song, "Microarray image enhancement by denoising using stationary wavelet transform," *IEEE Trans. Nanotechnol.*, vol. 2, no. 4, pp. 184–189, 2003.
- [22] M. Mastriani and A. E. Giraldez, "Microarrays denoising via smoothing of coefficients in wavelet domain," *Journal of Biomedical Sciences*, vol. 1, pp. 1306–1216, 2006.
- [23] H. Stefanou, T. Margaritis, D. Kafetzopoulos, K. Marias, and P. Tsakalides, "Microarray image denoising using a two-stage multiresolution technique," pp. 383–389, 2007.
- [24] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. Royal Society London A*, pp. 2543–2560, Sept. 1999.
- [25] Y. Rakvongthai, A. Vo, and S. Oraintara, "Complex gaussian scale mixtures of complex wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 58, no. 7, pp. 3545–3556, 2010.
- [26] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. Roy. Stat. Soc. B*, vol. 36, no. 1, pp. 99–102, 1974.
- [27] [Online], <http://smd.stanford.edu/index.shtml>.
- [28] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [29] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Lett.*, vol. 9, no. 12, pp. 438–441, Dec. 2002.
- [30] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Appl. and Computat. Harm. Anal.*, vol. 10, no. 3, pp. 234–253, May 2001.

- [31] E. J. Candès, D. L. Donoho, and L. Ying, “Fast discrete curvelet transform,” *Multiscale model. and simul.*, vol. 5, no. 3, pp. 861–899, Sept. 2006.
- [32] T. T. Nguyen and H. Chauris, “Uniform discrete curvelet transform for seismic processing,” in *EAGE Conf. and Tech. Exhib. (EAGE’08)*, Rome, Italy, June 2008, paper 11.3.4.
- [33] T. T. Nguyen and S. Orintara, “The shiftable complex directional pyramid-part I: Theoretical aspects,” *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 4651–4660, Oct. 2008.
- [34] J.-L. Starcka, M. K. Nguyenb, and F. Murtagh, “Wavelets and curvelets for image deconvolution: A combined approach.”

BIOGRAPHICAL STATEMENT

Lakshmi Srinivasan was born in Tamilnadu, India in 1985. She received her Bachelor of Engineering (B.E.) degree in Electronics & communication Engineering from the Visvesvaraya Technological University, India in May 2007. From 2007 to 2008 she worked in Samsung India Software Operation (SISO) Bangalore, India developing multimedia audio codecs. She graduated with a Masters degree in Electrical Engineering from University of Texas at Arlington, USA. She was member of Multirate Signal Processing lab under the guidance of Dr. Soontorn Oraintara. Her research interests are image processing, bioinformatics, wireless communication and multimedia applications.