PROBABILISTIC MODELING OF WAVELET COEFFICIENTS FOR PROCESSING OF IMAGE AND VIDEO SIGNALS

S. M. Mahbubur Rahman

A thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

Concordia University

Montréal, Québec, Canada

April 2009

Your file Votre référence
ISBN: 978-0-494-63363-2
Our file Notre référence
ISBN: 978-0-494-63363-2

ABSTRACT

Probabilistic Modeling of Wavelet Coefficients for Processing of Image and Video Signals

S. M. Mahbubur Rahman, Ph.D.

Concordia University, 2009

Statistical estimation and detection techniques are widely used in signal processing including wavelet-based image and video processing. The probability density function (PDF) of the wavelet coefficients of image and video signals plays a key role in the development of techniques for such a processing. Due to the fixed number of parameters, the conventional PDFs for the estimators and detectors usually ignore higher-order moments. Consequently, estimators and detectors designed using such PDFs do not provide a satisfactory performance.

This thesis is concerned with first developing a probabilistic model that is capable of incorporating an appropriate number of parameters that depend on higher-order moments of the wavelet coefficients. This model is then used as the prior to propose certain estimation and detection techniques for denoising and watermarking of image and video signals.

Towards developing the probabilistic model, the Gauss-Hermite series expansion is chosen, since the wavelet coefficients have non-compact support and their empirical density function shows a resemblance to the standard Gaussian function. A modification is introduced in the series expansion so that only a finite number of terms can be used for modeling the wavelet coefficients with rendering the resulting PDF to become negative. The parameters of the resulting PDF, called the modified Gauss-Hermite (MGH) PDF, are evaluated in terms of the higher-order sample-moments. It is shown that the MGH PDF fits the empirical density function better than the

existing PDFs that use a limited number of parameters do.

The proposed MGH PDF is used as the prior of image and video signals in designing maximum *a posteriori* and minimum mean squared error-based estimators for denoising of image and video signals and log-likelihood ratio-based detector for watermarking of image signals. The performance of the estimation and detection techniques are then evaluated in terms of the commonly used metrics. It is shown through extensive experimentations that the estimation and detection techniques developed utilizing the proposed MGH PDF perform substantially better than those that utilize the conventional PDFs. These results confirm that the superior fit of the MGH PDF to the empirical density function resulting from the flexibility of the MGH PDF in choosing the number of parameters, which are functions of higher-order moments of data, leads to the better performance. Thus, the proposed MGH PDF should play a significant role in wavelet-based image and video signal processing.

ACKNOWLEDGEMENTS

First, I thank God, the Almighty, for giving me the ability to finish my doctoral thesis. By God's will the completion of this work has become a reality.

I would like to express my deep gratitude to my supervisors Professor M. Omair Ahmad and Professor M. N. S. Swamy for their constant support, encouragement, patience, and invaluable guidance during the span of this research. I am grateful to them for providing me the freedom and motivation to explore new ideas. I also want to thank them for spending many hours with me in discussion about my research. The useful suggestions and comments provided by the members of the thesis examination committee, especially by the external examiner Professor Rui J. P. de Figueiredo of University of California, Irvine, as well as those of anonymous reviewers of my journal papers are also deeply appreciated.

I am indebted to Concordia University, NSERC, and ReSMiQ for their financial supports that were crucial for completing this research. Also, I would like to acknowledge the Department of Electrical and Computer Engineering and the Faculty of Graduate Studies in the Concordia University.

My sincere thanks go to my parents, family members, and relatives for their support, encouragement, and prayers during my research.

Special thanks to my wife, Tamanna Howlader, whose patience, love, and unlimited support during her own doctoral study have made the successful completion of this thesis. I would like to mention my little daughter, Tanzila Rahman, whose beautiful face always refreshes my mind.

Finally, I wish to thank my colleagues in the Center for Signal Processing and Communications notably Dr. M. I. H. Bhuiyan, Dr. S. A. Fattah, Dr. Saad Bouguezel, Dr. Awni Itradat, Ms. Celia Shahnaz, and Mr. Rajashekhar Kakumani for their friendship, fruitful suggestions and constant moral support.

Contents

List of Tables

List of Figures

List of Acronyms

AVC	:	Advanced Video Coding
AWGN	:	Additive White Gaussian Noise
BK	:	Bessel K-form
BVG	:	Bivariate Gaussian
BVMGH	:	Bivariate Modified Gauss-Hermite
CDF-9/7	:	Cohen-Daubechies-Feauveau-9/7
CWT	:	Complex Wavelet Transform
DCT	:	Discrete Cosine Transform
D-8	:	Daubechies-8
DFT	:	Discrete Fourier Transform
DT-CWT	:	Dual-Tree Complex Wavelet Transform
DWT	:	Discrete Wavelet Transform
ED	:	Efficacy of Detector
EM	:	Expectation Maximization
GC	:	Gram-Charlier
GH	:	Gauss-Hermite
GΓ	:	Generalized Gamma
GG	:	Generalized Gaussian
GSP	:	Gaussian Statistical Processor
HMT	:	Hidden Markov Tree
HVS	:	Human Visual System
IFMAP	:	Inter-Frame Maximum A Posteriori

i.i.d.	:	Independent and Identically Distributed
IPSNR	:	Improvement in Peak Signal-to-Noise Ratio
IQR	:	Interquartile Range
JPEG	:	Joint Photographic Experts Group
KLD	:	Kullback-Leibler Divergence
KNN	:	K-Nearest Neighbor
KSD	:	Kolmogorov-Smirnov Distance
LAWMAP	:	Locally-Adaptive Window-based Maximum A Posteriori
LAWML	:	Locally-Adaptive Window-based Maximum Likelihood
MAP	:	Maximum A Posteriori
MGH	:	Modified Gauss-Hermite
ML	:	Maximum Likelihood
MMSE	:	Minimum Mean Squared Error
MoM	:	Method of Moments
MPEG	:	Moving Picture Experts Group
MSE	:	Mean Squared Error
PDF	:	Probability Density Function
PR	:	Perfect Reconstruction
PRS	:	Pseudorandom Sequence
PSNR	:	Peak Signal-to-Noise Ratio
Q2C	:	Quadrature to Complex
ROC	:	Receiver Operating Characteristics
SAR	:	Synthetic Aperture Imagery
$S\alpha S$	:	Symmetric Alpha Stable
S-8	:	Symmlet-8
SMG	:	Scale Mixtures of Gaussians

SNIG	:	Symmetric Normal Inverse Gaussian
SNR	:	Signal-to-Noise Ratio
SPIHT	:	Set Partitioning in Hierarchical Trees
SSIM	:	Structural Similarity
SWT	:	Stationary Wavelet Transform
WDR	:	Watermark-to-Document Ratio

List of Symbols

δ	:	Bin width in histogram
ν	:	Bipolar PRS generated from secret key
α	:	Characteristic exponent in $S\alpha S$ PDF
$\mathrm{erfc}\{\cdot\}$	:	Complementary error function
C	:	Computational complexity
$*$	:	Convolution operator
ρ	:	Correlation
$\mathtt{C_{fa}}, \mathtt{C_m}$	:	Costs of false alarm and miss-detection
$\mathbf{\Sigma}$	:	Covariance matrix
c, p	:	Current and previous frames
τ	:	Detector threshold
d	:	Dimension of signal
$\mathrm{erf}\{\cdot\}$	:	Error function
$\mathcal{F}, \mathcal{F}^{-1}$	:	Forward- and inverse-transform of DFT
$\Gamma(\cdot)$	:	Gamma function
H_r	:	Hermite polynomials of order r
ζ	:	Highest order of H_r in finite GH series
$\mathbf{1}\{\cdot\}$	:	Indicator function
Λ	:	Likelihood ratio
h_ϕ, h_ψ	:	Low- and high-pass filters of DWT
$\boldsymbol{W}_L, \boldsymbol{W}_H$	:	Low- and high-pass transformation matrices of DWT
r, θ	:	Magnitude and phase components of complex coefficient

$E\{\cdot\}$	:	Mathematical expectation
μ, σ^2	:	Mean and variance
$\kappa_r(\cdot)$	:	Modified Bessel function of third kind with order r
M_n	:	Moment of order n
$\mathcal{H}_0, \mathcal{H}_1$	:	Null hypothesis and alternative hypothesis
ℓ	:	Number of decomposition level in DWT
N_b	:	Number of points using bin width of histogram
$\mathcal{L}$	:	Number of previous frames in processing
N	:	Number of samples in a subband
$\aleph$	:	Number of samples in squared-shape local neighborhood $\mathcal{S}$
$D_r(\cdot)$	:	Parabolic cylinder function of order r
a, b	:	Parameters in GΓ PDF
P_{fa}, P_{det}	:	Probabilities of false alarm and detection
$p(\cdot)$	:	Probability density function
ϕ, ψ	:	Scaling and wavelet functions of DWT
$\Omega(\cdot)$	:	Score function
β_r	:	Series coefficients in GH expansion
s, c	:	Shape and width parameters in GG, BK, SNIG PDFs
$N_{1\ell} \times N_{2\ell}$	:	Size of a subband at level ℓ
$N_1 \times N_2 \times N_3$	:	Size of data in 3D space
$\mathtt{w} \times \mathtt{w}$	:	Size of square-shaped window
S, K	:	Skewness and kurtosis
$\mathcal{N}(\mu, \sigma^2)$	:	Standard Gaussian random variable
$(x. y, z)$	:	3D spatial index in pixel-domain
(i, j, k)	:	3D spatial index in wavelet-domain
$\Pr(\cdot)$	:	Total probability

$\mathrm{tr}\{\cdot\}$	:	Trace of a matrix
$\mathcal{U}(\cdot)$	:	Unit step function
u, v	:	Real and imaginary components of complex coefficient
w	:	Watermark embedded in wavelet coefficients
ξ	:	Watermark strength
ε	:	Wavelet coefficient of noise
g	:	Wavelet coefficient of noisy/watermarked image/video
f	:	Wavelet coefficient of true image/video
ω	:	Weight parameter in temporal filter
γ	:	Weight parameter of the MGH PDF

Chapter 1

Introduction

1.1 Introduction

In the era of multimedia technology, digital image and video signals have become important sources of visual information. Development of novel algorithms for various applications of these signals covers a wide area of research effort in the signal processing community. Such algorithms include those for classification, texture analysis, denoising, compression, restoration, watermarking, and target recognition. Denoising is essential for signals such as ultrasound, X-ray, magnetic resonance imagery, mammography, computer tomography, and digital radiology, which are used for medical diagnosis. Denoising is also essential for signals such as the synthetic aperture radar (SAR) images that are used for remote sensing of earth and planetary resources. Compression of visual signals is indispensable for transmission and storage, since the size of the of such signals is usually large. Watermarking is becoming an increasing necessity for copyright protection of digital data.

A number of transforms such as the discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT) are used for processing of image and video signals depending on the requirements of the applications. Some of the notable features due to which the DWT has enjoyed significant success in the development of several efficient algorithms, such as those for denoising, restoration,

compression, watermarking, and target recognition, are:

- high energy compaction efficiency with an acceptable computational complexity,
- flexibility in choosing a basis function,
- excellent time-frequency/space-frequency characteristics of the transformed coefficients,
- efficient modeling of the human visual system (HVS) using the subband decompositions that are very similar to the response of the neurons of the HVS.

In fact, the upcoming image coding standard JPEG 2000[1] [1] uses the DWT as the quality of an image produced by the wavelet transform is better than that produced by the other transforms for a given bitrate. The video coding standard H.264/MPEG-4 AVC[2] [2] uses the Hadamard transform and adopts a scalable feature like the DWT. It has been shown in [3]-[5] that the rate-distortion performance of this codec may be improved by using wavelets in the transform coding. In addition, several new wavelet-based video codecs such as Dirac [6], [7], motion-lower tree wavelet [8], and motion-compensated temporal filtering and motion vector coding [9] have shown to perform better than H.264/MPEG-4 AVC.

Statistical methods have emerged as significant tools in the development of algorithms for interpreting and processing of signals. Such algorithms require probability models for the signals. The requirements of a desirable probability model are:

- The model should capture the essential variability of the signal.
- The model should show a good statistical match with the empirical model.
- The model should be mathematically tractable and computationally efficient.

[1] JPEG stands for Joint Photographic Experts Group.

[2] MPEG stands for Moving Picture Experts Group and AVC stands for Advanced Video Coding

It may be mentioned that in the case of a video, the correlation between two neighboring frames provides a measure of motion that exists between these frames. Hence, incorporation of this correlation is essential in the development of any probability model of a video.

1.2 Motivation

Signal-estimation and signal-detection are the two major components of statistical signal processing. The performance of any wavelet-based estimation or detection technique for an image or video signal is highly dependent on the probability density function (PDF) of the wavelet coefficients of the signal. The PDF of the wavelet coefficients may be estimated using any of the three traditional techniques, namely, non-parametric, parametric, and semi-parametric. Non-parametric density estimation techniques such as reproducing kernel-based density estimation in the weighted Folk space [10] and smoothing splines [11], are not suitable for incorporating any *a priori* knowledge. For example, non-parametric techniques mathematically become very complex in the case of estimation of a PDF from noisy data by incorporating the known noise characteristics. In addition, these estimation techniques may be expensive in terms of memory and computation. Hence, a non-parametric density estimation technique is not a good choice, when the PDF has a known shape. Since the PDFs of the wavelet coefficients of image/video are known to have sharp peaks and long tails, the non-parametric density estimation techniques are not used for this purpose. In general, the parametric density estimation techniques provide acceptable results when the shape of the PDF is known, and hence, several attempts have been made to estimate the PDFs of the wavelet coefficients of an image/video using such a technique. Examples of such PDFs are the generalized Gaussian (GG) [12]-[20], symmetric alpha-stable ($S\alpha S$) [21], Bessel K-form (BK) [22], and symmetric normal

inverse Gaussian (SNIG) [23]. It is to be noted that the parametric PDFs usually have a small number of parameters, leading to relatively simple and straight forward estimation techniques. Two advantages of the parametric density estimation techniques are the tractable incorporation of the *a priori* knowledge and a lower computational complexity. However, in view of the fact that the number of parameters are fixed and cannot be changed, the parametric PDFs often cannot provide a satisfactory fit with the empirical density function of the signal. In such a case, a semi-parametric density estimation technique may be used [24]. Semi-parametric PDFs are also referred to as the scale mixtures of parametric density functions; some examples of such PDFs are the scale mixtures of Gaussians (SMG) [25], [26] and scale mixtures of Laplacians [27]. Two major concerns of the semi-parametric PDFs are in the selection of the number of parameters, particularly the scale parameters, and the estimation of the parameters. It is to be pointed out that the computational complexity of any statistical signal processing technique depends on the number of parameters of the PDF, as well as the computational load of their estimation. In a parametric density estimation technique, the parameters of a PDF may be estimated using several strategies such as the method of moments (MoM) [28], maximum likelihood (ML) method [29], [30], and entropy-matching method [31]. Since in most cases the DWT coefficients of an image or video can be treated as a data set having a large sample size, the moment-based parameter estimation provides a consistent result [32]. In addition, the method of moments is convenient for estimating the parameters from a noisy data. In a semi-parametric density estimation technique, however, the information obtained from the higher-order moments of the data samples cannot be used for estimating the parameters of the PDF. Instead, the parameters are usually estimated by using an iterative algorithm such as the expectation maximization (EM), the computational complexity of which is very high, exponentially increasing with the number of scale parameters.

At present, most of the wavelet-based image and video processing algorithms exploit the parametric PDFs to capture the variability of the true signal coefficient. Computationally expensive semi-parametric PDFs have also been used, but these PDFs fail to incorporate the information obtained from the higher-order moments, which are necessary for obtaining better probability models. Since the existing statistical estimation and detection techniques are developed based on such PDFs, their performance is naturally affected by the mismatch between these PDFs and the empirical PDFs.

1.3 Scope of the Thesis

The objective of the work is to develop a new PDF for the wavelet coefficients of image and video signals using a non-traditional density estimation technique, such as employing a suitable orthogonal series expansion, so that the PDF is able to incorporate an appropriate number of parameters that are functions of higher-order moments of the data to improve the closeness of fit of the PDF to the empirical density function with an appreciable accuracy. In order to investigate the effectiveness of using the new probabilistic model in the statistical estimation and detection techniques, studies for denoising of image and video signals, as well as for watermarking of images are undertaken.

In the proposed density estimation technique, the PDF is expressed as a series expansion of orthogonal polynomials. In practice, the weight function of such a series expansion is represented in terms of a standard PDF. Table 1.1 lists a few orthogonal polynomials along with their corresponding weight functions and regions of support. Among the various possible series expansions, we choose the one that uses the Hermite polynomials for the following two major reasons:

- These polynomials are orthogonal with respect to the Gaussian weight function,

Table 1.1: Weight functions and regions of support for the series expansions of some orthogonal polynomials.

Polynomial	Weight function	Support
Chebyshev	Arcsin PDF†	$(-a, a)$
Bernstein	Empirical PDF	$(0, 1)$
Laguerre	Gamma PDF	$(0, \infty)$
Hermite	Gaussian PDF	$(-\infty, \infty)$
Legendre	Uniform PDF	$(-1, 1)$

†The compact support of the arcsin PDF is $(-a, a)$.

which resembles the empirical PDF; hence, it is more likely that the expanded PDF would match the empirical one in a better manner.

- The Gauss-Hermite (GH) expansion being valid in $(-\infty, \infty)$ is more appropriate for wavelet coefficients that do not have a finite region of support.

In addition, there are certain other advantages in using the GH expansion in the proposed density estimation technique:

- The series has a uniform mean square convergence [33].
- The parameters can be expressed in terms of higher-order moments in closed-form.
- The polynomials can be determined in a recursive manner.

A modification of the GH expansion is proposed wherein an appropriate number of terms can be used so that the resulting PDF matches well with the empirical PDF of the wavelet coefficients of an image/video signal, while ensuring at the same time the resulting PDF to be non-negative. The proposed PDF will be referred to as the modified GH (MGH) PDF. It is shown through experimentations that the MGH PDF fits the empirical density function of the wavelet coefficients better than the other commonly used PDFs do.

A number of statistical estimation and detection techniques, which include the maximum *a posteriori* (MAP) and minimum mean squared error (MMSE)-based denoising algorithms for images and video as well as log-likelihood ratio-based watermark detection algorithm for an image signal, are developed using the proposed MGH PDF as the prior of the wavelet coefficients. Extensive experimentations are carried out on frequently used test images and video sequences to show the effectiveness of these algorithms, that use the proposed PDF as prior.

1.4 Organization of the Thesis

The thesis is organized as follows.

In Chapter 2, a brief review of the classical DWT, stationary wavelet transform (SWT), and dual tree complex wavelet transform (DT-CWT) is presented. This review is intended to facilitate the understanding of the development of the algorithms presented in the thesis. This chapter also includes some preliminaries on the commonly used PDFs of the wavelet coefficients of images and video.

In Chapter 3, the development of the new PDF, the univariate MGH PDF, for the subband and local neighboring wavelet coefficients of the image signal, is presented. Experimental results are provided showing that the MGH PDF fits the empirical density function better than the conventional PDFs do.

Statistical modeling of a Gaussian signal in the DT-CWT domain, in both the Cartesian and polar representations, is presented in Chapter 4. The models proposed in this chapter are used in the denoising algorithms developed in the subsequent chapters.

In Chapter 5, wavelet-based subband- and locally-adaptive image denoising algorithms are developed using the proposed univariate MGH PDF. Performance of the denoising algorithms are studied using various indices such as the peak signal-to-noise

ratio (PSNR), structural similarity (SSIM), and visual quality.

In Chapter 6, another application of the proposed univariate MGH PDF is presented by developing a detector for the wavelet-based image watermarking scheme. Performance of this watermark detector is studied using the receiver operating characteristics (ROC), efficacy, and robustness.

The bivariate MGH PDF for modeling the subband wavelet coefficients of two neighboring frames of a video is developed in Chapter 7. Experimental results on test video sequences are provided to show that the proposed PDF matches the empirical PDF very well.

In Chapter 8, computationally efficient wavelet-based locally-adaptive video denoising algorithms are developed using a simpler form of the bivariate MGH PDF.

Finally, some concluding remarks highlighting the contributions of the thesis and scope of further research are provided in Chapter 9.

Chapter 2

Wavelet Coefficients and Their Traditional PDFs

2.1 Introduction

In visual signal processing, the purpose of a transformation is to be able to use some of the important properties of the signal in the transformed domain that are not available to be used in the pixel-domain. The transform should decorrelate the spatially distributed energy to a fewer number of data samples such that no information is lost. The importance of the wavelet transform in the area of image and video signal processing has been discussed in Chapter 1. Fig. 2.1 shows simple block diagrams of transform-based statistical estimation and detection techniques for an image or video signal. It may be seen from this figure that the first step of such a technique is the transformation of spatial or spatio-temporal information into the coefficient domain. The second step is to process the transformed coefficients. For an estimation technique the inverse transformation is necessary, whereas for a detection technique the inverse transformation is not required. Statistical estimation as well as detection techniques employ the PDF of the wavelet coefficients as the prior of an image or video signal. The wavelet transform has several variants, each suitable for specific applications. For example, the classical DWT, which is also known as the decimated DWT, is suitable for compression and the SWT or simply the shift-invariant form

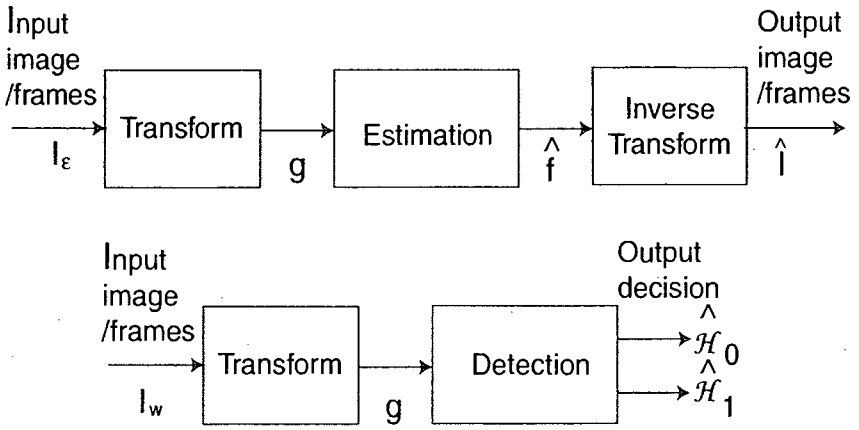


Figure 2.1: Block diagrams of transform-based statistical estimation and detection techniques for an image or video signal.

of the DWT is suitable for denoising and restoration [34]. The decimated DWT is shift-variant as well as it has poor directional selectivity. The SWT is shift-invariant at a cost of huge computational complexity and its directional selectivity is same as that of the decimated DWT. Improved directional selectivity can be obtained using a special type of wavelet transform such as the complex wavelet transform (CWT) [35]-[38], ridgelet [39], curvelet [40], contourlet [41], and steerable pyramid [42], [43]. In general, the probabilistic model that performs well in the classical DWT also provides a similar performance in other variants of the wavelet transform. In this thesis, we concentrate on probabilistic modeling of the coefficients of the classical DWT and the CWT. There are some notable key features of the CWT such as having a near shift-invariance property, an improved directional selectivity, and a low computational complexity.

In this chapter, first a summary of the wavelet theory [44]-[48] is presented. The way of obtaining the one dimensional (1D) wavelet coefficients from a 1D signal using the concept of multiresolution analysis is given. Another way of estimation of the wavelet coefficients using the concept of filter-bank [44], [45], [48] is considered later. An implementation of the 2D DWT is presented followed by which that of the

2D SWT and CWT as well as 3D DWT are discussed. Finally, some preliminaries on the commonly used PDFs along with their moment-based parameter estimation techniques for modeling the wavelet coefficients of the image and video signals are given.

2.2 Preliminaries on DWT

2.2.1 1D DWT

Let $I(x)$, $x = 1, 2, \cdots, N_1$ represent a 1D signal of size $N_1 \times 1$, where x is the 1D index. The DWT of the signal is given by [49]

$$I(x) = \frac{1}{\sqrt{N_1}} \left[\sum_{i=1}^{N_1} f_J^A(i) \phi_J^A(x,i) + \sum_{\ell=1}^{J} \sum_{i=1}^{N_1} f_\ell^{\mathcal{D}}(i) \psi_\ell^{\mathcal{D}}(x,i) \right] \tag{2.1}$$

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

$$f_J^A(i) = \frac{1}{\sqrt{N_1}} \sum_{x=1}^{N_1} I(x) \phi_J^A(x,i) \tag{2.2}$$

and

$$f_\ell^{\mathcal{D}}(i) = \frac{1}{\sqrt{N_1}} \sum_{x=1}^{N_1} I(x) \psi_\ell^{\mathcal{D}}(x,i) \tag{2.3}$$

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

$$\phi(x) = \sqrt{2} \sum_{u=-\infty}^{\infty} h_\phi(u) \varphi(2x - u) \tag{2.4}$$

and

$$\psi(x) = \sqrt{2} \sum_{u=-\infty}^{\infty} h_{\psi}(u)\phi(2x-u) \tag{2.5}$$

where $h_{\phi}(u)$ and $h_{\psi}(u)$, respectively, are called the scaling and wavelet vectors for the forward DWT. Using the above relations, Mallat [45] has shown that the 1D DWT coefficients of adjacent decomposition levels can be estimated very efficiently by using the following equations [45]:

$$f_{\ell+1}^{A}(i) = \sum_{u=-\infty}^{\infty} h_{\phi}(u-2i) f_{\ell}^{A}(u) = h_{\phi}(-u) * f_{\ell}^{A}(u)\Big|_{u=2i, i\geq 0} \tag{2.6}$$

$$f_{\ell+1}^{\mathcal{D}}(i) = \sum_{u=-\infty}^{\infty} h_{\psi}(u-2i) f_{\ell}^{A}(u) = h_{\psi}(-u) * f_{\ell}^{A}(u)\Big|_{u=2i, i\geq 0} \tag{2.7}$$

where $*$ is a convolution operator. The above equations imply that the approximate and detail coefficients at decomposition level $(\ell+1)$ can be obtained from filtering the approximate coefficients at decomposition level ℓ by h_{ϕ} and h_{ψ}, respectively, and down-sampling by 2. The inverse transform of the 1D DWT coefficients of adjacent decomposition levels can be estimated very efficiently by using [45]

$$f_{\ell}^{A}(i) = \tilde{h}_{\phi}(u) * f_{\ell+1}^{A}(u)\Big|_{u=\frac{i}{2}, i\geq 0} + \tilde{h}_{\psi}(u) * f_{\ell+1}^{\mathcal{D}}(u)\Big|_{u=\frac{i}{2}, i\geq 0} \tag{2.8}$$

where $\tilde{h}_{\phi}(u)$ and $\tilde{h}_{\psi}(u)$, respectively, are called the scaling and wavelet vectors for the inverse DWT. The above equation implies that the approximate coefficients at decomposition level ℓ can be obtained by filtering the up-sampled (up-sampling by 2) approximate and detail coefficients at decomposition level $(\ell+1)$ by $\tilde{h}_{\phi}$ and $\tilde{h}_{\psi}$, respectively, and adding the results. Using the concept of perfect reconstruction (PR) in filter-bank theory[1], $h_{\phi}(u)$ and $h_{\psi}(u)$ are often referred to as the dual analysis filters, whereas $\tilde{h}_{\phi}(u)$ and $\tilde{h}_{\psi}(u)$ as dual synthesis filters [48]. A simple block diagram of the 1D forward DWT using the analysis filters and the inverse DWT using the synthesis

[1] In filter-bank theory, the PR means that the output of the filter-bank is identical to the input except for a possible delay and overall scaling factor.

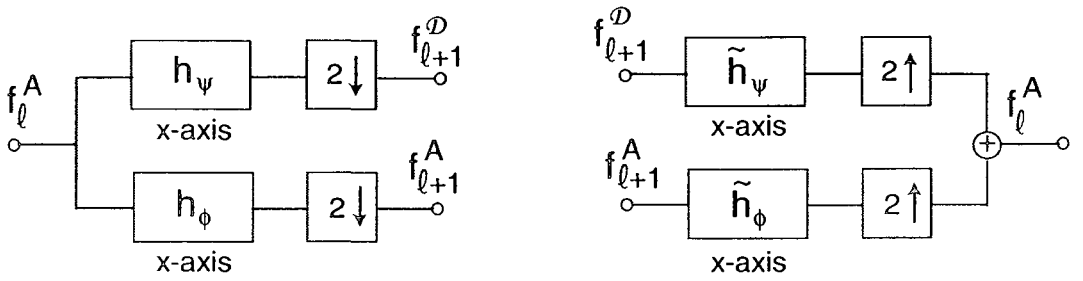


Figure 2.2: Block diagram of the 1D DWT using the analysis and synthesis filters. (a) Forward transform for obtaining the $(\ell+1)$-th level detail coefficients. (b) Inverse transform for obtaining the ℓ-th level approximate coefficients.

filters for adjacent decomposition levels is shown in Fig. 2.2. It is to be noted that there is no data redundancy in the DWT coefficients due to the down-sampling of the filtered coefficients after each level decomposition. In literature, this wavelet transform is referred to as the decimated 1D DWT.

2.2.2 2D DWT

Let $I(x,y)$, $x = 1,2,\cdots,N_1$, $y = 1,2,\cdots,N_2$ represent pixels of a 2D signal such as an image or a video frame of size $N_1 \times N_2$, where (x,y) is the 2D index. The DWT of this signal is given by [49]

$$I(x,y) = \frac{1}{\sqrt{N_1 N_2}}\left[\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_J^A(i,j)\Phi_J^A(x,y,i,j) + \sum_{\ell=1}^{J}\sum_{\mathtt{O}\in\mathtt{H,V,D}}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_\ell^{\mathtt{O}}(i,j)\Psi_\ell^{\mathtt{O}}(x,y,i,j)\right] \tag{2.9}$$

where $f_\ell^{\mathtt{O}}$ $(\mathtt{O} \in \mathtt{H,V,D})$ denotes the detail coefficients in the decomposition level ℓ $(\ell = 1,2,\cdots,J)$ of orientation $\mathtt{O}$, and Φ and Ψ, respectively, are the 2D scaling and wavelet functions. Separable scaling and wavelet functions are common for extending the 1D DWT to the multidimensional DWT. Hence, the 2D scaling and wavelet

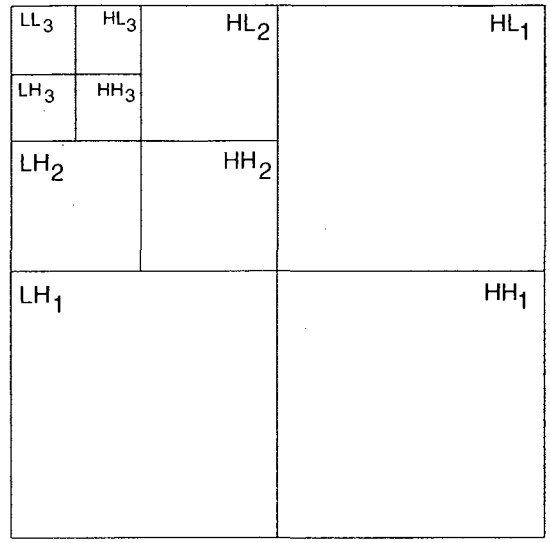


Figure 2.3: Subband representations of a 3-level 2D DWT.

functions can be represented in terms of 1D scaling and wavelet functions as

$$\Phi_J^A(x,y,i,j) = 2^{J/2}\phi(2^J x - i)\phi(2^J y - j) \qquad \Psi_\ell^{\mathrm{H}}(x,y,i,j) = 2^{\ell/2}\psi(2^\ell x - i)\phi(2^\ell y - j)$$

$$\Psi_\ell^{\mathrm{V}}(x,y,i,j) = 2^{\ell/2}\phi(2^\ell x - i)\psi(2^\ell y - j) \qquad \Psi_\ell^{\mathrm{D}}(x,y,i,j) = 2^{\ell/2}\psi(2^\ell x - i)\psi(2^\ell y - j)$$

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

$$f_J^A(i,j) = \frac{1}{\sqrt{N_1 N_2}} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} I(x,y)\Phi_J^A(x,y,i,j) \tag{2.10}$$

and

$$f_\ell^{\mathrm{D}}(i,j) = \frac{1}{\sqrt{N_1 N_2}} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} I(x,y)\Psi_\ell^{\mathrm{D}}(x,y,i,j) \tag{2.11}$$

These coefficients are clustered into groups or subbands of different levels and orientations. The subbands HL_ℓ, LH_ℓ, and HH_ℓ ($\ell = 1, 2, \cdots, J$) contain the detail coefficients of the horizontal (0°), vertical (90°), and diagonal (±45°) orientations, viz., f_ℓ^{H}, f_ℓ^{V}, and f_ℓ^{D}, respectively. The subband LL_J is the lowest resolution residual that contains f_J^A. The subband representation of a typical 3-level 2D DWT is shown in Fig. 2.3. Due to the separability of the scaling and wavelet functions, the 2D DWT may be

implemented using two separate 1D DWTs. Similar to the 1D DWT, fast and efficient implementations of the forward and inverse 2D DWTs are obtained by using the analysis and synthesis filters, respectively. Figs. 2.4 and 2.5 show the simple block diagrams of the forward and inverse 2D DWTs, respectively. It may be seen from these figures that both the forward and inverse transforms use a two-stage filtering operation, one through the columns (i.e., x-axis) and another through the rows (i.e., y-axis).

Orthogonality, symmetry, and length of the analysis and synthesis filters are some of the important characteristics of the 2D DWT. Orthogonality decorrelates the transformed coefficients, and thus minimizes the redundancy. Symmetry provides linear phase and permits a symmetric boundary extension technique that minimizes border artifacts. A short length filter is valuable for reducing truncation artifacts in the reconstructed signal [50]. Since a finite impulse response filter with a length greater than two cannot have both the orthogonal and symmetry properties, the analysis or synthesis filters of the DWT usually do not possess these properties simultaneously[2]. In the DWT having an orthogonal filter-bank, h_ϕ is the time reversal of h_ψ, h_ϕ and h_ψ being themselves time reversals of $\tilde{h}_\phi$ and $\tilde{h}_\psi$, respectively. A biorthogonal or spline filter of length greater than two can be made symmetric. In such a filter, it is not necessary for h_ϕ and h_ψ to be time reversals of each other. However, to maintain the requirement of PR, $\tilde{h}_\phi$ and $\tilde{h}_\psi$ must be time reversals of h_ϕ and h_ψ, respectively. The choice of the analysis and synthesis filters, whether they be orthogonal, biorthogonal or spline, depends on the application. For example, biorthogonal or spline filters are preferable for image compression, since a symmetric boundary extension or low truncation error of the transformed coefficients improves the compression performance. On the other hand, denoising requires a higher decorrelation efficiency and, therefore, orthogonal filters are preferred for this purpose.

[2]The only exception is the Haar wavelet with a filter of length two.

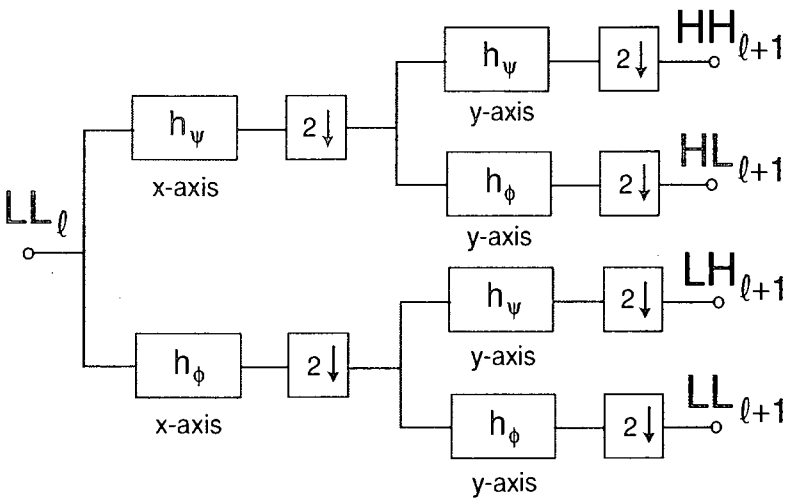


Figure 2.4: Block diagram of the 2D forward DWT using the analysis filters for obtaining the $(\ell + 1)$-th level detail coefficients.

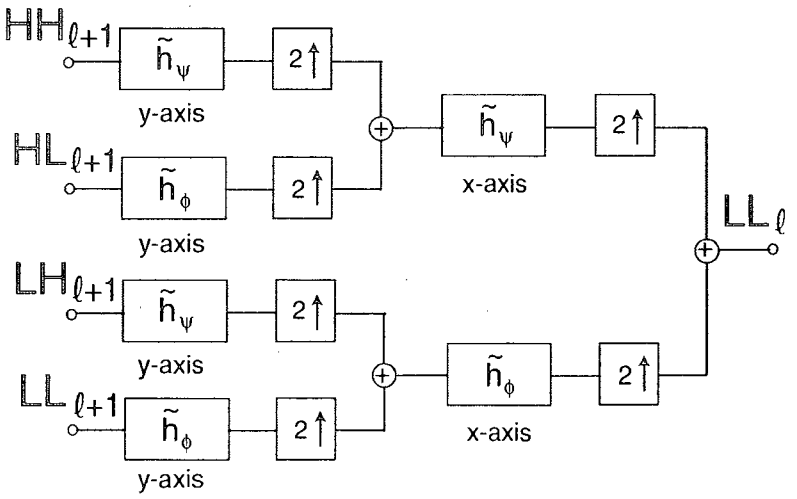


Figure 2.5: Block diagram of the 2D inverse DWT using the synthesis filters for obtaining the ℓ-th level approximate coefficients.

2D SWT

Although down-sampling in the forward DWT allows a PR of the 2D signal [45], some of the possible shifts are neglected [51]. In such a case, a small shift in the input signal may cause significant variations in the distribution of the energy among the decimated 2D DWT coefficients at different decomposition levels. Hence, processing of the decimated 2D DWT coefficients may not provide the desired performance in many applications including signal denoising and restoration. To overcome such a problem, the SWT coefficients are calculated by following a procedure similar to that of the DWT except that the down-sampling is avoided to account for all the possible shifts. Such a transform is also called the overcomplete representation or shift-invariant form or non-decimated form of the DWT. The data redundancy in the SWT is $2^{\ell d}$ for a d dimensional signal [12], [51].

2D CWT

The decimated 2D DWT-based signal processing have two major drawbacks, namely, lack of shift-invariance and poor directional selectivity. The SWT overcomes only the first problem, but at the expense of a huge computational load. The second problem is inherent in the SWT, since it uses the 2D scaling and wavelet functions that are real and separable. In other words, both the decimated DWT and SWT coefficients capture the features of a 2D signal only in the directions of 0°, 90°, and ±45°. In order to improve the directional selectivity, the CWT that has complex valued scaling and wavelet functions, $\Phi_1 + \imath\Phi_2$ and $\Psi_1 + \imath\Psi_2$, such that Φ_1 and Φ_2 as well as Ψ_1 and Ψ_2 form Hilbert pairs, has been introduced. Such a choice of scaling and wavelet functions provides the CWT to have six directional features, namely, $-15°$, $-45°$, $-75°$, $15°$, $45°$, and $75°$. Therefore, the CWT has improved directional property compared to both the decimated DWT and the SWT. Various methods have been

proposed for obtaining the CWT coefficients [35]-[38], [52]-[57]. However, due to the simplicity of implementation and a sufficiently low redundancy, the dual-tree CWT (DT-CWT) proposed by Kingsbury [52] and later generalized by Selesnick [54], is becoming popular [35], [36]. The DT-CWT consists of two trees of DWT in parallel and provides four pairs of subbands, namely, $(LL1_\ell, LL2_\ell)$, $(LH1_\ell, LH2_\ell)$, $(HL1_\ell, HL2_\ell)$, and $(HH1_\ell, HH2_\ell)$, where 1 and 2 refer to the two trees. The implementation of the DT-CWT requires that the first level decomposition, i.e., $\ell = 1$, uses the dual analysis filters $h_{\phi 0}$ and $h_{\psi 0}$ and the higher-level decompositions, i.e., $\ell > 1$, the first tree uses $h_{\phi 1}$ and $h_{\psi 1}$ and the second uses $h_{\phi 2}$ and $h_{\psi 2}$. The filters $h_{\phi 1}$ and $h_{\psi 1}$ operate on the odd numbered data samples, while $h_{\phi 2}$ and $h_{\psi 2}$ on the even numbered data samples. Having a delay of one-half of a sample between these operations ensures that the dual filters of the two trees, namely, tree 1 and tree 2, form a Hilbert pair [35], [54], [55], [58]-[61]. In order to improve the shift-invariance property, the DT-CWT avoids the down-sampling operation in the first level decomposition, whereas the higher-level decompositions use this operation similar to that in the case of the decimated DWT. Thus, the DT-CWT improves the shift-invariance property, but with a reduced redundancy of 2^d, which is less than that of the SWT by a factor 2^ℓ. Finally, an additional building block, hereafter referred to as the quadrature to complex (Q2C) block, is added at the end of each subband for estimating the complex coefficients from the real quadrature-filtered coefficients. Simple block diagrams of the forward transform of the 2D DT-CWT including the Q2C block for the first-level and ℓ-th ($\ell > 1$) level decompositions, respectively, are shown in Figs. 2.6 and 2.7. Detailed analysis of estimating the 2D DT-CWT coefficients has been provided in [35], [52]-[55], [58]-[60].

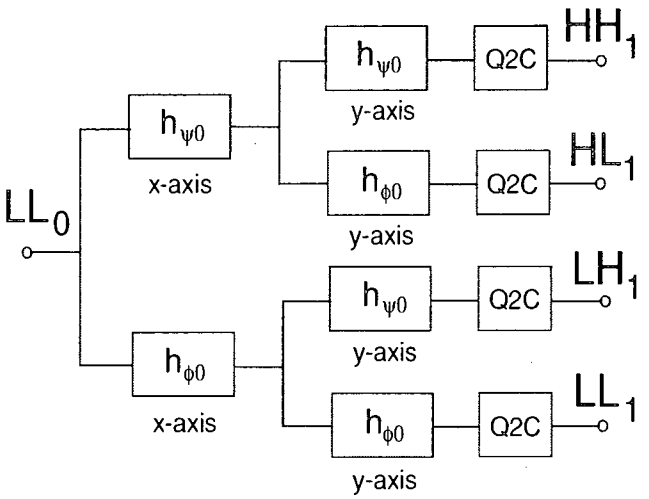


Figure 2.6: Block diagram of the 2D forward DT-CWT for the first level decomposition.

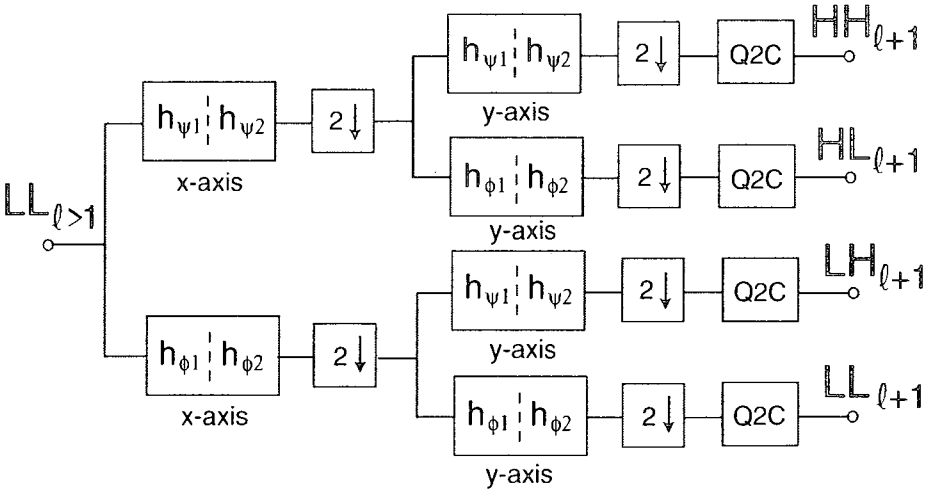


Figure 2.7: Block diagram of the 2D forward DT-CWT for the ℓ-th ($\ell > 1$) level decomposition.

2.2.3 3D DWT

Let $I(x,y,z)$, $x = 1,2,\cdots,N_1$, $y = 1,2,\cdots,N_2$, $z = 1,2,\cdots,N_3$ represent pixels of a 3D signal such as a video of size $N_1 \times N_2 \times N_3$, where (x,y,z) is the 3D index. The DWT of this signal is given by

$$I(x,y,z) = \frac{1}{\sqrt{N_1 N_2 N_3}} \left[\sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{j=1}^{N_3} f_J^A(i,j,k) \bar{\Phi}_J^A(x,y,z,i,j,k) + \sum_{\ell=1}^{J} \sum_{\Theta} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{k=1}^{N_3} f_\ell^\Theta(i,j,k) \bar{\Psi}_\ell^\Theta(x,y,z,i,j,k) \right] \quad (2.12)$$

where f_J^A ($A \in LLL$) denotes the approximate coefficients in the largest level J, f_ℓ^Θ ($\Theta \in \{LLH, LHL, LHH, HLL, HHL, HLH, HHH\}$) denotes the detail coefficients in the decomposition level ℓ ($\ell = 1,2,\cdots,J$) in the direction Θ, and $\bar{\Phi}$ and $\bar{\Psi}$ are the 3D scaling and wavelet functions, respectively. Separable scaling and wavelet functions of the 3D DWT are given by

$$\bar{\Phi}_J^{LLL}(x,y,z,i,j,k) = 2^{J/2}\phi(2^J x - i)\phi(2^J y - j)\phi(2^J z - k)$$

$$\bar{\Psi}_\ell^{LHL}(x,y,z,i,j,k) = 2^{\ell/2}\phi(2^\ell x - i)\psi(2^\ell y - j)\phi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{LLH}(x,y,z,i,j,k) = 2^{\ell/2}\phi(2^\ell x - i)\phi(2^\ell y - j)\psi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{LHH}(x,y,z,i,j,k) = 2^{\ell/2}\phi(2^\ell x - i)\psi(2^\ell y - j)\psi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{HLL}(x,y,z,i,j,k) = 2^{\ell/2}\psi(2^\ell x - i)\phi(2^\ell y - j)\phi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{HHL}(x,y,z,i,j,k) = 2^{\ell/2}\psi(2^\ell x - i)\psi(2^\ell y - j)\phi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{HLH}(x,y,z,i,j,k) = 2^{\ell/2}\psi(2^\ell x - i)\phi(2^\ell y - j)\psi(2^\ell z - k)$$

$$\bar{\Psi}_\ell^{HHH}(x,y,z,i,j,k) = 2^{\ell/2}\psi(2^\ell x - i)\psi(2^\ell y - j)\psi(2^\ell z - k)$$

The approximate and detail coefficients of the 3D DWT can be obtained as

$$f_J^A(i,j,k) = \frac{1}{\sqrt{N_1 N_2 N_3}} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \sum_{z=1}^{N_3} I(x,y,z) \bar{\Phi}_J^A(x,y,z,i,j,k) \quad (2.13)$$

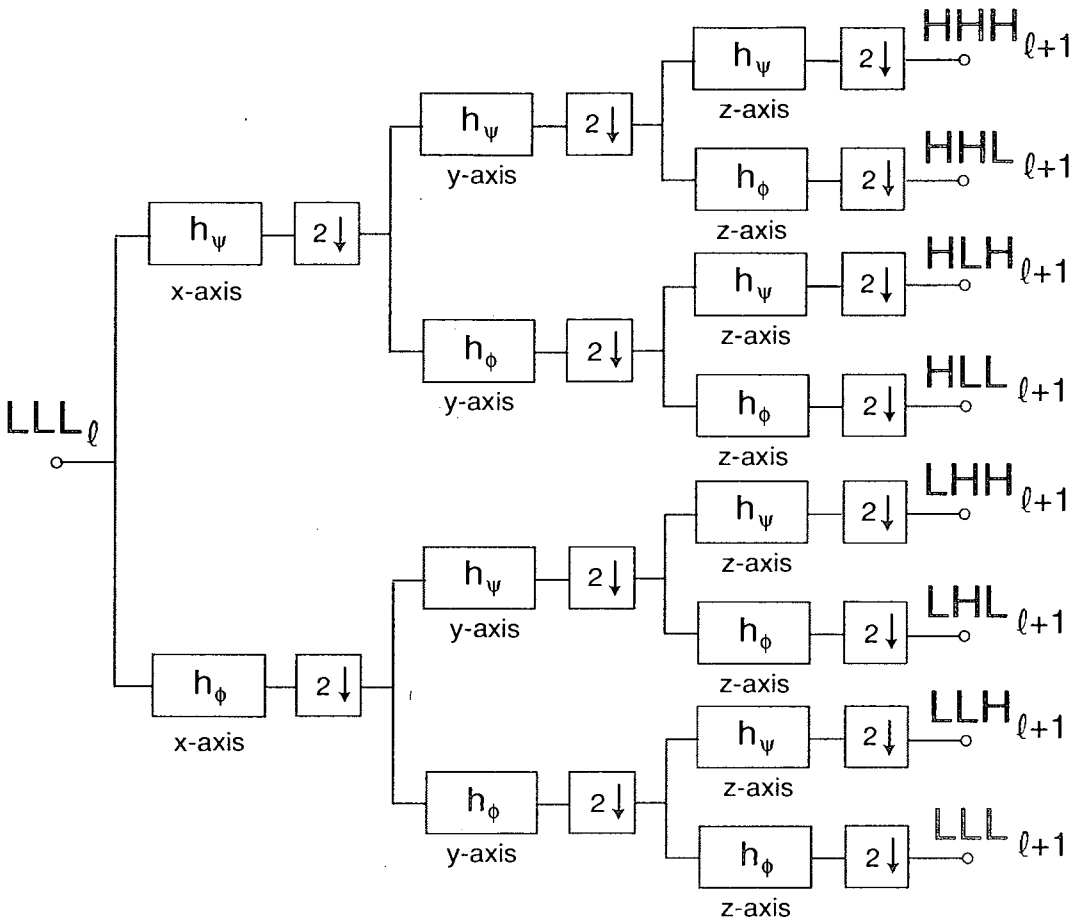


Figure 2.8: Block diagram of the 3D forward DWT using the analysis filters for obtaining the $(\ell+1)$-th level detail coefficients.

and

$$f_\ell^\Theta(i,j,k) = \frac{1}{\sqrt{N_1 N_2 N_3}} \sum_{x=1}^{N_1} \sum_{y=1}^{N_2} \sum_{z=1}^{N_3} I(x,y,z) \bar{\Psi}_\ell^\Theta(x,y,z,i,j,k) \qquad (2.14)$$

The eight subbands of the 3D DWT are LLL_ℓ, LHL_ℓ, LLH_ℓ, LHH_ℓ, HLL_ℓ, HHL_ℓ, HLH_ℓ, and HHH_ℓ $(\ell = 1, 2, \cdots, J)$. A block diagram of the forward transform of the 3D DWT for the ℓ-th level decomposition is shown in Fig. 2.8.

2.3 Existing PDFs

This section presents an introduction to the commonly used PDFs for modeling the DWT coefficients of an image or a video signal. Without loss of generality, let $\boldsymbol{f}$ represent the vector containing all the coefficients of a subband, wherein each coefficient f_{ij} or f_{ijk} is a sample of the subband of the 2D DWT at index (i, j) or that of the 3D DWT at index (i, j, k). Let $p_{\boldsymbol{f}}(\boldsymbol{f})$ represent the PDF of the random variable $\boldsymbol{f}$. A common assumption for modeling the DWT coefficients is that the underlying random process is stationary, i.e., the PDFs of the DWT coefficients are invariant to translation in the coefficient plane. In other words, the DWT coefficients are assumed to be independent and identically distributed (i.i.d.). By observing the nature of the histogram, the following two important properties of the DWT coefficients of image and video signals are revealed:

- The PDFs of the DWT coefficients are invariant with respect to scaling [62].
- The PDFs of the DWT coefficients have a zero mean and are heavy-tailed [63].

Hence, a heavy-tailed PDF is usually chosen to model the subband coefficients of the DWT of an image or a video signal. An important fact is that unlike other transforms, the coefficients of the DWT show a considerable statistical dependency within a subband, and a very weak dependency across subbands [64]. Therefore, in order to account for the intra-subband dependency, the PDFs are refined with respect to the local neighborhood for some applications such as denoising and restoration. A few of the PDFs have been extended to the bivariate [65]-[68] and multivariate forms [69], [70] or used in the hidden Markov tree (HMT) framework [71], [72] to take into consideration the inter-subband dependency. A brief summary of some of the traditional PDFs of the DWT coefficients of image and video signals are given below.

2.3.1 GG PDF

Since the work of Mallat [47], the zero-mean GG PDF[3] with a shape parameter s $(1 < s < 2)$ has been most commonly used to model the DWT coefficients of images and video. The GG PDF is defined as

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{2\frac{c}{s}\Gamma(\frac{1}{s})} e^{-|\boldsymbol{f}/c|^s} \tag{2.15}$$

where c $(c > 0)$ is the width parameter and $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$ is the Gamma function. Let $M_{n\boldsymbol{f}}$ be the n-th order moment of the data samples of $\boldsymbol{f}$. The values of s and c are computed numerically using the second- and fourth-order moments of the data samples as

$$\sigma_{\boldsymbol{f}}^2 = M_{2\boldsymbol{f}} = \frac{c^2\Gamma(\frac{3}{s})}{\Gamma(\frac{1}{s})} \qquad \text{and} \qquad K_{\boldsymbol{f}} = \frac{M_{4\boldsymbol{f}}}{M_{2\boldsymbol{f}}^2} = \frac{\Gamma(\frac{1}{s})\Gamma(\frac{5}{s})}{\Gamma^2(\frac{3}{s})} \tag{2.16}$$

where $\sigma_{\boldsymbol{f}}^2$ is the variance and $K_{\boldsymbol{f}}$ is the kurtosis. It is to be noted that the GG PDF becomes a Gaussian PDF when $s = 2$ and a Laplacian PDF when $s = 1$.

2.3.2 SMG PDF

In the SMG PDF, the random variable $\boldsymbol{f}$ is assumed to be a product of a Gaussian random vector $\mathbf{F}$ and an independent hidden random positive scalar-multiplier $\sqrt{t}$. In [25], [26], [42], [73], the SMG PDF of the DWT coefficients is evaluated as

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{t}\boldsymbol{\Sigma}_{\mathbf{F}}|}} \int \exp\left[-\frac{\boldsymbol{f}^T(\boldsymbol{t}\boldsymbol{\Sigma}_{\mathbf{F}})^{-1}\boldsymbol{f}}{2}\right] p_{\boldsymbol{t}}(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \tag{2.17}$$

where N is the number of data samples of a subband and $\boldsymbol{\Sigma}_{\mathbf{F}} = \boldsymbol{\Sigma}_{\boldsymbol{f}}$ is the covariance matrix, assuming that the mean of $\boldsymbol{t}$ is unity. The most critical part in evaluating the SMG PDF is the estimation of the density function for $\boldsymbol{t}$, i.e., $p_{\boldsymbol{t}}(\boldsymbol{t})$. In [42], $p_{\boldsymbol{t}}(\boldsymbol{t})$ is approximated by a log-normal PDF given by

$$p_{\boldsymbol{t}}(\boldsymbol{t}) = \frac{1}{t\sqrt{2\pi\sigma_L^2}} \exp\left[-\frac{(\log \boldsymbol{t} - \mu_L)^2}{2\sigma_L^2}\right] \tag{2.18}$$

[3] In literature, the GG PDF is also referred to as the generalized Laplacian PDF

where the parameters μ_L and σ_L are calculated in terms of the second- and fourth-order moments of the data samples as

$$\mu_L = -\frac{\log(\sigma_{\boldsymbol{t}}^2 + 1)}{2} \qquad \text{and} \qquad \sigma_L^2 = \log(\sigma_{\boldsymbol{t}}^2 + 1) \tag{2.19}$$

where

$$\sigma_{\boldsymbol{t}}^2 = \frac{M_{4\boldsymbol{f}}}{3\sigma_{\boldsymbol{f}}^2} - 1 \tag{2.20}$$

In [25], the $p_{\boldsymbol{t}}(\boldsymbol{t})$ is approximated by the Jeffrey's prior defined by $p_{\boldsymbol{t}}(\boldsymbol{t}) \propto \frac{1}{\boldsymbol{t}}$. Since it is an improper PDF, a valid region $(0, t_{\min})$ is required to define for this PDF.

2.3.3 BK PDF

The BK PDF has been used for modeling the DWT coefficients in [74]. Similar to the GG PDF, it has only two parameters, viz., the shape parameter $\bar{s}$ ($\bar{s} > 0$), and the width parameter $\bar{c}$ ($\bar{c} > 0$). The BK PDF is defined by

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{\sqrt{\pi}\Gamma(\bar{s})}\left(\frac{\bar{c}}{2}\right)^{-\frac{\bar{s}}{2}-\frac{1}{4}} \left|\frac{\boldsymbol{f}}{2}\right|^{\bar{s}-\frac{1}{2}} \kappa_{\bar{s}-\frac{1}{2}}\left(\sqrt{\frac{2}{\bar{c}}}|\boldsymbol{f}|\right) \tag{2.21}$$

where κ_r is the modified Bessel function of the third kind and of order r, defined by [75]

$$\kappa_r(\bar{z}\boldsymbol{f}) = \frac{\Gamma(r+\frac{1}{2})(2\boldsymbol{f})^r}{\sqrt{\pi}\bar{z}^r} \int_0^\infty \frac{\cos(\bar{z}u)\mathrm{d}u}{(u^2+\boldsymbol{f}^2)^{r+\frac{1}{2}}} \tag{2.22}$$
$$\left(Re\{r\} > -\frac{1}{2}, \quad \bar{z} > 0, \quad |\arg \boldsymbol{f}| < \frac{\pi}{2}\right)$$

The values of the parameters $\bar{s}$ and $\bar{c}$ are computed from the variance and kurtosis of the data samples as

$$\sigma_{\boldsymbol{f}}^2 = \bar{s}\bar{c} \qquad \text{and} \qquad K_{\boldsymbol{f}} = \frac{3}{\bar{s}} + 3 \tag{2.23}$$

The BK PDF reduces to a double exponential PDF for $\bar{s} = 1$. If $\bar{s} \gg 1$, the density function gets closer to a Gaussian PDF [74].

It may be noted that $p_{\boldsymbol{f}}(\boldsymbol{f})$ can be defined by the so called *normal variance-mean mixture* PDF. In such a case, the conditional density function $p_{\boldsymbol{f}|\mathcal{Z}}(\boldsymbol{f}|\mathcal{Z})$ is a Gaussian PDF with mean $\mu_{\boldsymbol{f}} + \mathcal{Z}\mu_{\mathcal{Z}}$ and variance $\mathcal{Z}\mathcal{V}$, where $\mathcal{Z}$ is the mixing variable. If $\mu_{\boldsymbol{f}} = \mu_{\mathcal{Z}} = 0$ and $\mathcal{V} = 1$, the conditional PDF results the SMG PDF, whereas if $\mathcal{Z}$ is a scaled Gamma PDF, then it approaches to BK PDF.

2.3.4 SNIG PDF

The SNIG PDF has been used for modeling the DWT coefficients of log-transformed images that are common in SAR or medical applications [23], [76]. This PDF is a mixture of the inverse Gaussian and standard Gaussian PDFs, and is expressed as

$$p_{\boldsymbol{f}_L}(\boldsymbol{f}_L) = \frac{\tilde{s}\tilde{c}}{\pi\sqrt{\tilde{c}^2 + \boldsymbol{f}_L^2}}\kappa_1\left(\tilde{s}\sqrt{\tilde{c}^2 + \boldsymbol{f}_L^2}\right)e^{\tilde{s}\tilde{c}} \tag{2.24}$$

where $\boldsymbol{f}_L$ denotes the DWT coefficients of a log-transformed image, $\tilde{s}$ ($\tilde{s} > 0$) is the shape parameter, and $\tilde{c}$ ($\tilde{c} > 0$) is the width parameter. The parameters of this PDF are estimated from the variance and kurtosis of the data samples as

$$\sigma^2_{\boldsymbol{f}_L} = \frac{\tilde{c}}{\tilde{s}} \qquad \text{and} \qquad K_{\boldsymbol{f}_L} = \frac{3(1+\tilde{s}\tilde{c})}{\tilde{s}^2} \tag{2.25}$$

The SNIG PDF results in a Gaussian PDF of variance $\frac{\tilde{c}}{\tilde{s}}$, when $\tilde{c} \to \infty$ and $\tilde{s} \to \infty$ such that $\frac{\tilde{c}}{\tilde{s}}$ is finite, and in a Cauchy PDF when $\tilde{s} = 0$.

2.4 Conclusion

In this chapter, a short introduction to the 1D DWT has been presented along with its 2D and 3D extensions. The implementation issues of the DWT have been discussed using the concept of filter-bank. The procedure of obtaining the SWT and DT-CWT using the classical DWT have also been presented in this chapter. Finally, some of the standard PDFs such as the GG, SMG, BK, and SNIG, for modeling the wavelet coefficients have been briefly discussed. The moment-based parameter estimation techniques of these PDFs have also been given.

Chapter 3

Proposed PDF of Wavelet Coefficients for Images

3.1 Introduction

The importance of a suitable PDF of the DWT coefficients in the development of statistical estimation and detection techniques for processing the image signals has been explained in Chapter 1. In this chapter, a new PDF for the DWT coefficients of an image signal is developed using the GH expansion. It is shown that by using the proposed PDF, an appropriate number of parameters that are dependent on higher-order moments, can be chosen for the subband coefficients [77], [78] as well as the local neighboring coefficients [34], [79]. A suitable modification of the GH expansion has been introduced so that the resulting PDF of the subband coefficients is always non-negative. The parameters of the proposed PDF are expressed in a closed-form expression in terms of the higher-order sample-moments of the data. Using extensive simulations, it is shown that the proposed PDF matches the empirical density function better than the standard ones do, both for the subband and local neighboring coefficients.

3.2 PDF of Subband Coefficients

The PDF of the random variable of subband coefficients $\boldsymbol{f}$ may be expressed in terms of the Hermite polynomials, which are orthogonal with respect to a Gaussian weight function, in the form [80]

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \sum_{r=3}^{\infty} \beta_r H_r\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] \tag{3.1}$$

where $\sigma_{\boldsymbol{f}}$ is the standard deviation of $\boldsymbol{f}$ and $H_r(\boldsymbol{f})$ is the Hermite polynomial of order r that satisfies the recurrance relation [75]

$$\begin{aligned} H_0(\boldsymbol{f}) &= 1 \\ H_1(\boldsymbol{f}) &= \boldsymbol{f} \\ H_{r+1}(\boldsymbol{f}) &= \boldsymbol{f}H_r(\boldsymbol{f}) - rH_{r-1}(\boldsymbol{f}) \qquad r \geq 1 \end{aligned} \tag{3.2}$$

The resulting PDF will be called as the GH PDF. A closed-form expression of the series coefficients β_r can be obtained in terms of the n-th order moment $M_{n\boldsymbol{f}}$ as (see Appendix A-1)

$$\beta_r = \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^u}{2^u (r-2u)!u!} \frac{M_{(r-2u)\boldsymbol{f}}}{\sigma_{\boldsymbol{f}}^{r-2u}} \tag{3.3}$$

where $\lfloor u \rfloor$ denotes the largest integer contained in u. The series coefficients β_r are determined only by the even-order moments for even r, and the odd-order moments for odd r. Hence, the parameter estimation of the GH PDF becomes relatively simple and straight forward. Although the PDF of the subband coefficients described in (3.1) is an infinite series expression, in practice, only the first few terms are used for a reasonably good approximation. For example, if the highest value of order r in (3.1) is chosen as 4, the resulting density function reduces to the Gram-Charlier (GC) PDF [79], [81], [82]

$$p_{4\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \frac{S_{\boldsymbol{f}}}{3!} H_3\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right) + \frac{K_{\boldsymbol{f}} - 3}{4!} H_4\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] \tag{3.4}$$

Table 3.1: Average values of the magnitude of the skewness ($|S_f|$) and the kurtosis (K_f) of the subband DWT coefficients for some test images.

Test image	$\|S_f\|$			K_f		
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 1$	$\ell = 2$	$\ell = 3$
Lena	0.09	0.17	0.26	12.81	20.40	14.70
Barbara	0.01	0.16	0.17	18.05	14.31	18.67
Boat	0.12	0.14	0.12	10.84	12.29	10.75
Man	0.09	0.16	0.15	16.49	10.03	7.97
Baboon	0.09	0.05	0.14	7.77	7.02	7.04
Grass	0.02	0.03	0.05	4.27	4.12	3.72
Bridge	0.01	0.07	0.08	7.28	5.82	6.04

where, $S_f = M_{3f}/\sigma_f^3$ and $K_f = M_{4f}/\sigma_f^4$, respectively, are the skewness and kurtosis of f. However, the approximation of the right side of (3.1) by only three terms may introduce some errors in $p_f(f)$, and the error can be reduced by including more number of terms in the series. In order to include a larger number of terms in the series, higher-order moments are necessary for estimating the series coefficients β_r. But, the deviations between the estimated sample-moments and the true-moments are not insignificant when the order becomes high [81]. Therefore, the error introduced by the estimation of β_r is not negligible when a large number of terms of the series is included. In addition, the decay rate of the density function $p_f(f)$ is proportional to $H_r(f/\sigma_f)e^{-f^2/2\sigma_f^2}$ or simply $(f/\sigma_f)^r e^{-f^2/2\sigma_f^2}$ for sufficiently large f. The larger the order r, the slower the decay of $p_f(f)$, thus making it a challenge to fit the tail part of $p_f(f)$ with that of the empirical PDF of the image coefficients. Assuming the largest value of r in (3.1) to be ζ, the right side of (3.1) may introduce real zeros in the expansion of $p_f(f)$, the maximum number of such zeros being ζ. In essence, the PDF defined by the truncated GH series may not be always non-negative. For example, the symmetric (i.e., the skewness is zero) GC PDF is non-negative only if $3 \leq K_f \leq 7$ [83]. Table 3.1 shows the average values of the magnitude of the skewness and the kurtosis of the subband coefficients in the first three levels for some

test images. It can be observed that the skewness of the subband coefficients of image is insignificant. It has also been observed that other odd-order moments of interest are also insignificant. Hence, the PDF of the subband coefficients of image can be treated as symmetric. As a result, all β_r in the infinite series of (3.1) for odd r disappear and the truncated GH PDF for the image wavelet coefficients can be written as

$$p_{\zeta f}(f) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{f^2}{2\sigma_f^2}} \left[1 + \sum_{r=2}^{\zeta/2} \beta_{2r} H_{2r}\left(\frac{f}{\sigma_f}\right) \right] \tag{3.5}$$

The kurtosis of the subband wavelet coefficients of images is usually more than 7 (see Table 3.1); therefore, $p_{4f}(f)$ cannot model these coefficients sufficiently well. In addition, each subband of the image wavelet coefficients usually has a sufficiently large number of data samples, and hence, the higher-order moments would be useful for modeling. Recall that the truncated GH PDF with a sufficiently large value of ζ is not ensured to be non-negative. However, the model of the subband coefficients defined by the truncated GH PDF in (3.5) is an even function, and therefore, the real zeros are symmetrically situated around the mode. Fig. 3.1 shows the probabilistic modeling of the HL_1 band DWT coefficients by the empirical PDF and by the truncated GH PDF with ζ varying from 4 to 14 for the images *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial*. The empirical PDF is obtained from the histogram-based density estimation technique [84], wherein the bin width of the histogram plays a significant role. If the bin width is too small, then the estimated PDF suffers from undersmoothing, i.e., the density function has a small bias but a large variability. On the other hand, if it is too large, then the estimated PDF suffers from oversmoothing, i.e., the density function has a small variability but a large bias. In order to obtain a tradeoff between the bias and variance, an appropriate choice of the bin width is necessary. The bin width can be fixed or variable in size. Variable width may be better, when the density function is asymmetric and multimodal in nature. Since the PDF of the image wavelet coefficients is unimodal and symmetric in nature, a fixed bin width is chosen for the

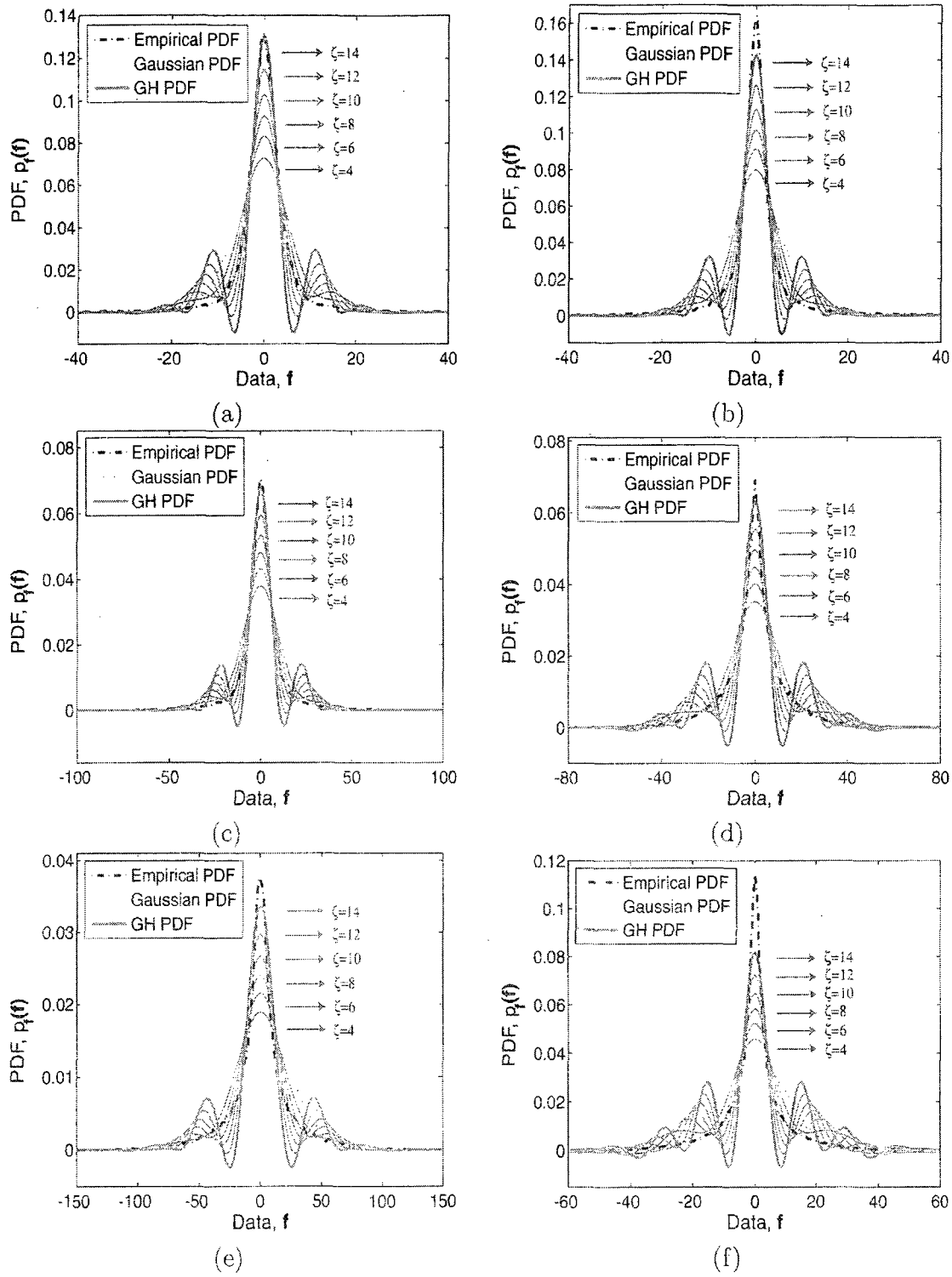


Figure 3.1: PDF of the HL_1 band DWT coefficients in terms of the Gauss-Hermite expansion with increasing value of ζ. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

histogram. The value of this bin width is chosen to be $2(\text{IQR})(N_{1\ell}N_{2\ell})^{-\frac{1}{3}}$, IQR being the interquartile range[1] and $N_{1\ell} \times N_{2\ell}$ the size of the subband, since such a choice asymptotically minimizes the mean integrated squared error of the empirical PDF that is non-Gaussian. It can be observed from Fig. 3.1 that, as ζ increases, the kurtotic behavior of the truncated GH PDF approaches that of the empirical one, even though it results in zeros in the tail part of the PDF. Let $\pm f_0$ be the two zeros closest to the mode for a particular value of ζ. The value of f_0 may be found by using any root finding routine, but we calculate it as

$$f_0 = \arg\min_{f} \left[\frac{1}{2}\Big(p_{\zeta f} + |p_{\zeta f}|\Big)\mathcal{U}(f) == 0 \right] \tag{3.6}$$

where $\mathcal{U}(f)$ is a unit step function and '==' is logical 'equal' operation. It is evident from Fig. 3.1 that when the coefficients lie in the interval $(-f_0, f_0)$, then these coefficients can be modelled well by the proposed GH PDF with an appropriate choice of ζ. Outside this range, i.e., in the tail part, such a truncated GH PDF introduces negative probability, which is not possible. Hence, in the tail part of the truncated GH PDF, a modification is necessary so that the probability becomes non-negative and at the same time the modified GH PDF fits well with the empirical PDF. For such a modification, the use of a GH PDF is preferred to that of using a different PDF, since the estimated parameters and Hermite polynomials can also be used for the tail part of the PDF. Examples of GH PDFs with non-negative probability are the Gaussian PDF and the symmetric GC PDF having $3 \leq K_f \leq 7$. Since the latter has a longer tail than the former and the PDFs of the image wavelet coefficients are known to have long-tails, we choose the GC PDF as the modified tail. In order to ensure that the total area under the curve of the modified GH PDF is unity, we multiply the GC PDF by a weight parameter γ. In this thesis, the density function

[1] The IQR is the difference between the third and first quartiles of the data.

is referred to as the MGH PDF. It can be mathematically formulated as [77]

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \begin{cases} \dfrac{1}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \displaystyle\sum_{r=2}^{\zeta/2} \beta_{2r} H_{2r}\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] & \text{if } |\boldsymbol{f}| < f_0 \\ \dfrac{\gamma}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \dfrac{K_m - 3}{4!} H_4\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] & \text{if } |\boldsymbol{f}| \geq f_0 \end{cases} \tag{3.7}$$

where $0 < \gamma < 1$ is the weight parameter, and $K_m = \max\big(\min(K_{\boldsymbol{f}}, 7), 3\big)$. Since the total area under the MGH PDF is unity, the parameter γ satisfies

$$\frac{2}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} \int_0^{f_0} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \sum_{r=2}^{\zeta/2} \beta_{2r} H_{2r}\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] \mathrm{d}\boldsymbol{f} + \frac{2\gamma}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} \int_{f_0}^{\infty} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \frac{K_m - 3}{4!} H_4\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] \mathrm{d}\boldsymbol{f} = 1 \tag{3.8}$$

After some manipulations, the value of γ can be obtained as (see Appendix A-2)

$$\gamma = \frac{1 - 2\left[I_0\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right) + \sum_{r=2}^{\zeta/2} \beta_{2r} I_{2r}\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right)\right]}{1 - 2\left[I_0\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right) + \frac{K_m - 3}{4!} I_4\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right)\right]} \tag{3.9}$$

where

$$\begin{aligned} I_0\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right) &= \frac{1}{2}\mathrm{erf}\left(\frac{f_0}{\sqrt{2}\sigma_{\boldsymbol{f}}}\right) \\ I_{2r}\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right) &= -\frac{1}{\sqrt{2\pi}} H_{2r-1}\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right) e^{-\frac{f_0^2}{2\sigma_{\boldsymbol{f}}^2}} \qquad u \geq 2 \end{aligned} \tag{3.10}$$

and $\mathrm{erf}(\cdot)$ represents the error function. In summary, the MGH PDF for a particular subband is estimated using the following steps:

1. Select ζ as the highest-order even moment. Estimate the moments M_{2r} $(r = 1, 2, \cdots, \zeta/2)$.

2. Estimate the series parameters β_{2r} $(r = 1, 2, \cdots, \zeta/2)$ by using (3.3).

3. Select the bin width $\delta_{\boldsymbol{f}} = 2(\mathrm{IQR})(N_{1\ell}N_{2\ell})^{-\frac{1}{3}}$. Set the discrete points $\boldsymbol{f}_d = [\boldsymbol{f}_{\min} : \delta_{\boldsymbol{f}} : \boldsymbol{f}_{\max}]$, where $\boldsymbol{f}_{\min}$ and $\boldsymbol{f}_{\max}$, respectively, are the minimum and maximum values of $\boldsymbol{f}$.

4. Estimate the Hermite polynomials $H_{2r}(\cdot)$ $(r = 1, 2, \cdots, \zeta/2)$ by using (3.2) for each $\boldsymbol{f}_d$.

5. Estimate f_0 using (3.5) and (3.6).

6. Estimate γ using (3.9) and (3.10).

7. Obtain the MGH PDF by (3.7).

The proposed MGH PDF requires the estimation of β_{2r}, H_{2r} and γ, and the computational complexities of these are $\mathcal{O}(\zeta N_1 N_2/2)$, $\mathcal{O}(N_b)$ and $\mathcal{O}(N_b + \zeta/2)$, respectively, where N_b is total number of discrete points resulting from the selected bin. Since $N_b \ll N_1 N_2$ and $\zeta \ll N_b$, the computational complexity of the proposed MGH PDF is $\mathcal{O}(\zeta N_1 N_2/2)$. Fig. 3.2 shows the modeling of the HL_1 band DWT coefficients by the MGH PDF in the linear-scale using different values of ζ for the images *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial*. It can be observed from this figure that the MGH PDF is ensured to be non-negative and for a sufficiently large value of ζ, the tail part of the MGH PDF matches very well with that of the empirical PDF. Therefore, in the proposed MGH PDF the choice of ζ is of importance, since it determines the highest-order of the Hermite polynomials and of the moments used in the parameter estimation. In order to make a good choice of ζ, the mean squared error (MSE) between the empirical and proposed MGH PDFs is plotted as a function of the highest-order moments used in (3.7) for the HL_1, HL_2, HL_3, and HL_4 band DWT coefficients for some test images, and shown in Fig. 3.3. It can be observed from this figure that in general the inclusion of an increased number of the higher-order moments in the parameter estimation of the proposed MGH PDF lowers the MSE. Due to the fact that the magnitudes of the wavelet coefficients increases significantly with the level of decomposition and so the spread of the empirical PDF, a lower MSE is obtained at a higher decomposition level. It is seen from Fig. 3.3 that a larger value

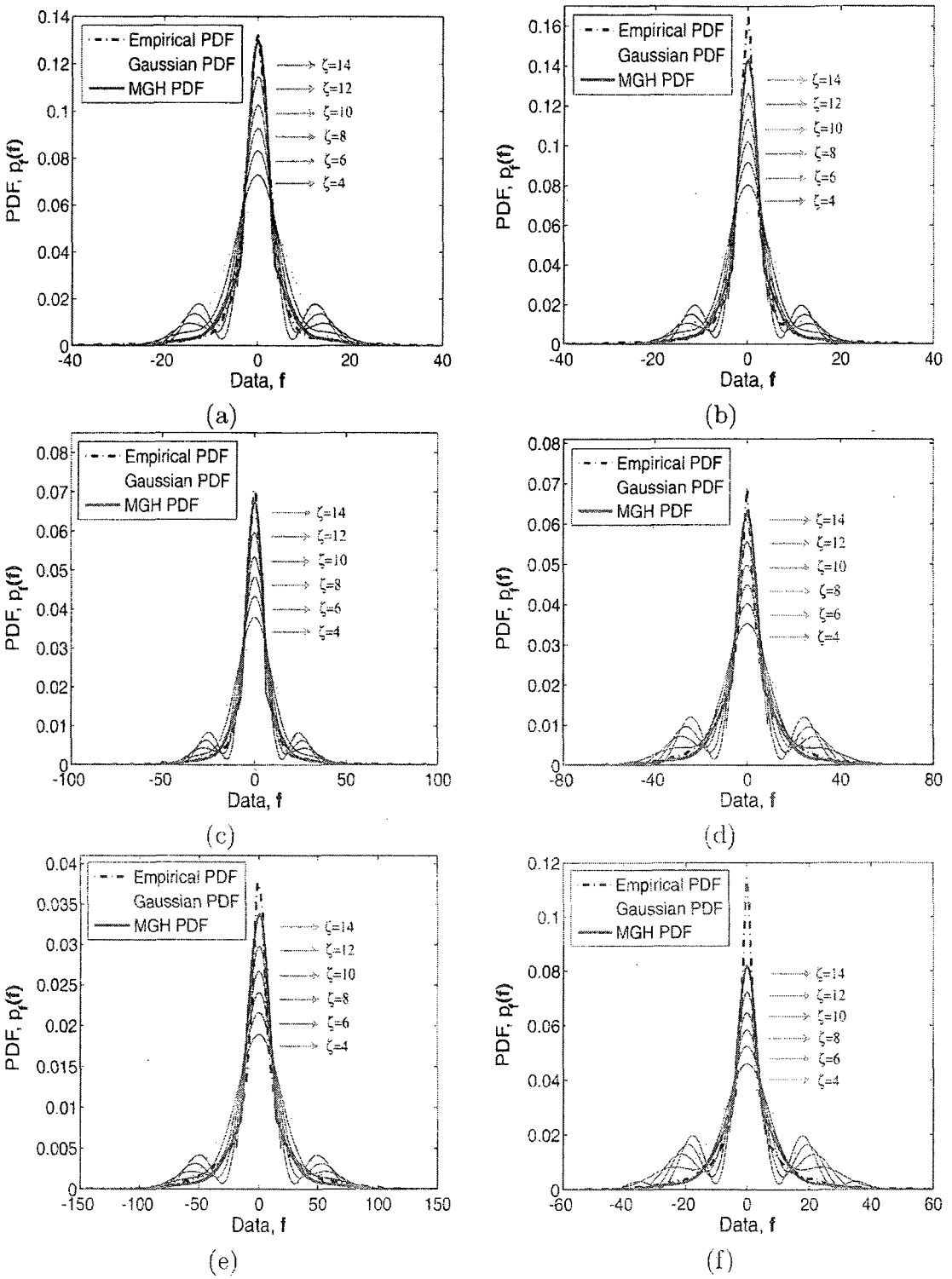


Figure 3.2: Modified Gauss-Hermite PDF of the HL_1 band DWT coefficients with increasing value of ζ. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

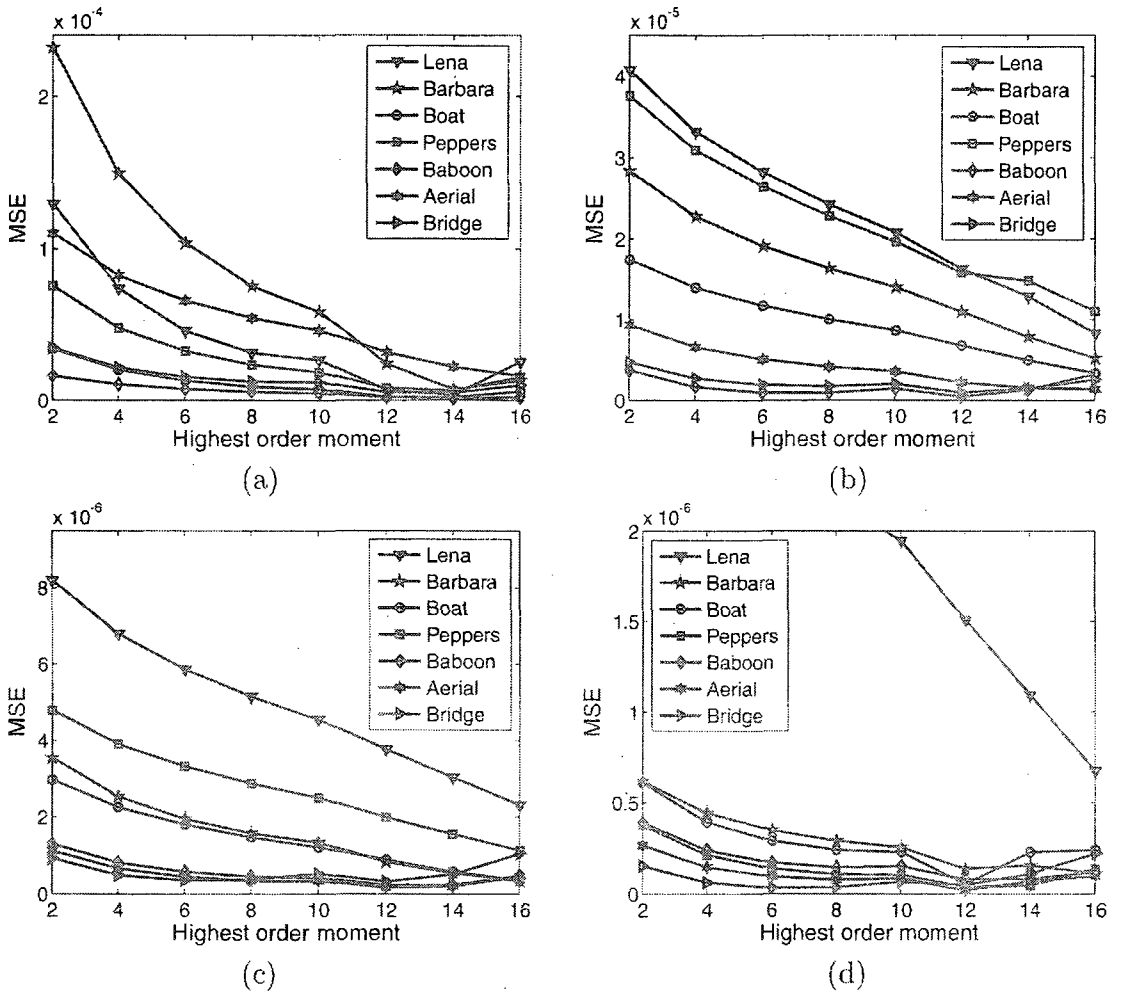


Figure 3.3: Mean squared error between the empirical and the modified Gauss-Hermite PDFs for the DWT coefficients as a function of the highest-order of the moments that has been used for the parameter estimation. The subbands are (a) HL_1, (b) HL_2, (c) HL_3, and (d) HL_4.

of ζ may not always provide a lower MSE due to the significant change in f_0 of the proposed MGH PDF, however, a value of 12, 14 or 16 is a good choice for ζ at any decomposition level. Fig. 3.4 shows, in the log-scale, the modeling performance of three PDFs, namely, the GG, BK, SNIG, and MGH that uses $\zeta = 14$ for the HL_1 band DWT coefficients of the images *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial*. The closeness of the fit of the proposed PDF with the empirical one, both in its peak and tail parts, is obvious as compared to the others. Moreover, to quantify the performance of the PDFs, we use, the Kolmogorov-Smirnov distance (KSD), $\max_{\boldsymbol{f}} |\int [p_{\boldsymbol{f}}^{\text{emp}}(\boldsymbol{f}) - p_{\boldsymbol{f}}(\boldsymbol{f})] \mathrm{d}\boldsymbol{f}|$, and the Kullback-Leibler divergence (KLD), $\int p_{\boldsymbol{f}}(\boldsymbol{f}) \ln \frac{p_{\boldsymbol{f}}(\boldsymbol{f})}{p_{\boldsymbol{f}}^{\text{emp}}(\boldsymbol{f})} \mathrm{d}\boldsymbol{f}$, as the two metrics, where $p_{\boldsymbol{f}}^{\text{emp}}(\boldsymbol{f})$ denotes the empirical PDF. Table 3.2 shows the results concerning the metrics KSD and KLD for the GG, BK, SNIG, and MGH PDFs of the first and second level image DWT coefficients. The metrics are calculated by averaging the results obtained using 96 images, each of size 512×512, from the database given in [85]. Standard deviations of the KLD metric are also included in the table. For the results shown here, the MGH PDF uses the even-order moments from 2 up to 14, whereas the GG, BK, and SNIG PDFs use only the second and fourth order moments in view of their limited number of parameters. From the table, it is evident that in most of the cases the MGH PDF, compared to the GG, BK, and SNIG PDFs, has the lowest distance and divergence from the empirical one. This high degree of matching between the empirical and MGH PDFs is mainly due to the fact that the proposed PDF can incorporate more number of higher-order moments in its parameter estimation process as compared to the others. This feature, along with the modification proposed through the introduction of the parameter γ, results in the proposed MGH PDF to match the empirical PDF very well both in the peak and tail segments. In order to demonstrate the effectiveness of the proposed PDF, a subband-adaptive denoising algorithm is developed in Chapter 5

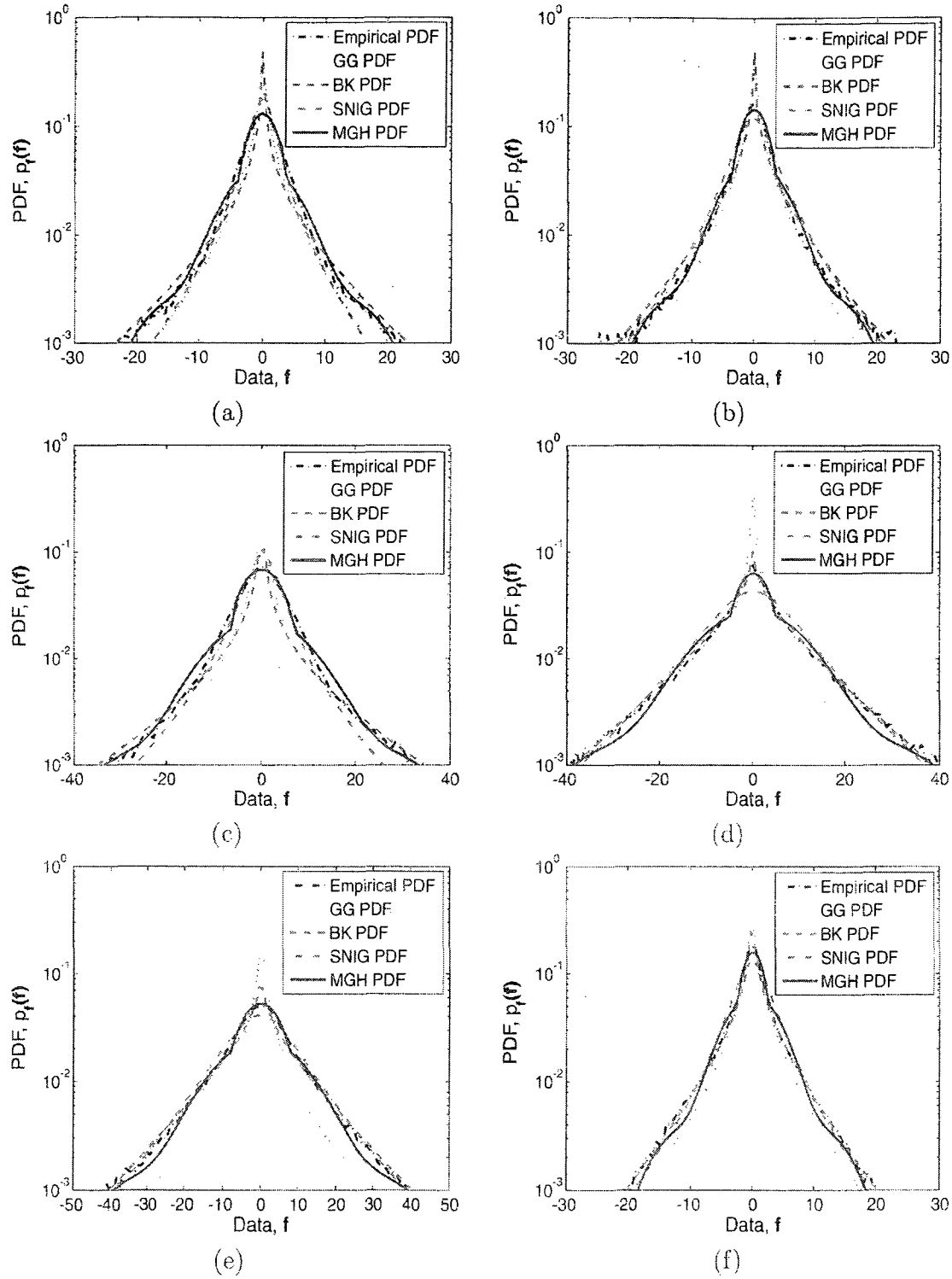


Figure 3.4: Modified Gauss-Hermite PDF in log-scale for the HL_1 band DWT coefficients. Hermite polynomials are up to the fourteenth order. Test images are (a) *Lena*, where $\gamma = 0.588$, (b) *Barbara*, where $\gamma = 0.648$, (c) *Boat*, where $\gamma = 0.609$, (d) *Bridge*, where $\gamma = 0.780$, (e) *Baboon*, where $\gamma = 0.727$, and (f) *Aerial*, where $\gamma = 0.704$.

Table 3.2: Values of the metrics KSD and KLD for the prior function modeling of the subband DWT coefficients of images. The average values of KLD (standard deviations in parentheses) and KSD are obtained from a 96 image database.

Prior model	KSD		KLD	
	$\ell = 1$	$\ell = 2$	$\ell = 1$	$\ell = 2$
	HH			
GG PDF	0.149	0.194	0.331 (0.238)	0.420 (0.327)
BK PDF	0.108	0.133	0.302 (1.932)	0.255 (0.994)
SNIG PDF	**0.086**	0.068	**0.082** (0.153)	0.092 (0.159)
MGH PDF	0.087	**0.067**	**0.082** (**0.045**)	**0.064** (**0.044**)
	LH			
GG PDF	0.146	0.185	0.309 (0.214)	0.326 (0.235)
BK PDF	0.088	0.081	0.122 (0.455)	**0.070** (0.091)
SNIG PDF	0.073	0.110	0.144 (0.373)	0.208 (0.560)
MGH PDF	**0.069**	**0.075**	**0.061** (**0.038**)	0.076 (**0.058**)
	HL			
GG PDF	0.161	0.201	0.352 (0.252)	0.417 (0.454)
BK PDF	0.083	0.084	0.108 (0.228)	**0.069** (0.099)
SNIG PDF	0.078	0.084	0.165 (0.645)	0.153 (0.333)
MGH PDF	**0.070**	**0.073**	**0.064** (**0.042**)	**0.069** (**0.054**)

and a watermark detection algorithm in Chapter 6, wherein the MGH PDF is used as the prior function of DWT coefficients of images.

3.3 PDF of Local Neighboring Coefficients

In some statistical estimation techniques, such as denoising and restoration of images, the dependency of the local neighboring DWT coefficients may be considered for achieving an improved performance. However, the computational complexity of the locally-adaptive techniques increases significantly as compared to that of the corresponding subband-adaptive ones. Very often, the local neighboring DWT coefficients are modelled using the standard Gaussian PDF, since in such a case closed-form expressions can be obtained for many techniques. Thus, the computational complexity of the locally-adaptive techniques using Gaussian PDF becomes relatively lower than

that of using a complex PDF. Since the Gaussian PDF is a specific case of the GG, $S\alpha S$, and SNIG PDFs, in some locally-adaptive techniques the parameters of these PDFs are refined with respect to the local spatial contexts. Note that the Gaussian PDF is also a specific case of the MGH PDF, when $\beta_{r(r\geq 3)} = 0$. By including an appropriate number of parameters in the GH expansion, a PDF better than the Gaussian and other PDFs may be obtained for the local neighboring coefficients similar to a PDF for the subband coefficients. In the case of modeling the local neighboring coefficients, however, the parameters that are estimated from the higher-order moments will not be so effective as that in the case of subband coefficients, since the error introduced by the estimated higher-order sample-moments in the former is not insignificant as in the latter. This is due to the fact that the number of data samples of the local neighboring coefficients is much less than that of the subband coefficients. It is found that the GC PDF, the parameters of which are estimated considering up to the fourth order moment is sufficient for probabilistic modeling of the local neighboring DWT coefficients of images. Similar to the subband coefficients, the skewness of the local neighboring image wavelet coefficients are very small, i.e., $S_{\boldsymbol{f}} \ll 1$. Hence, the PDF of the local neighboring coefficients can be treated as symmetric by considering $S_{\boldsymbol{f}}$ to be zero. Without loss of generality, we represent $\boldsymbol{f}$ as the vector containing all the local neighboring coefficients at the spatial location (i, j). Hence, the PDF of the local neighboring image wavelet coefficients may be expressed in terms of the symmetric GC PDF in the form [81]

$$p_{\boldsymbol{f}}(\boldsymbol{f}) = \frac{1}{\sigma_{\boldsymbol{f}}\sqrt{2\pi}} e^{-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}} \left[1 + \frac{K_{\boldsymbol{f}} - 3}{4!} H_4\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\right] \tag{3.11}$$

Table 3.3 shows the average values of $K_{\boldsymbol{f}}$ for the local neighboring DWT coefficients of the first and second level decompositions of a few test images using the window sizes of 5×5 and 11×11. Note that similar results have been obtained for $K_{\boldsymbol{f}}$ using other test images given in the database [85]. It can be observed from Table 3.3 that

Table 3.3: The average values of kurtosis for the test images obtained from the local neighboring DWT coefficients.

Test image	Kurtosis, K_f					
	Local (5×5)			Local (11×11)		
	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 1$	$\ell = 2$	$\ell = 3$
Lena	3.51	4.51	4.79	4.25	6.19	6.91
Barbara	3.41	4.60	4.72	4.53	6.36	6.25
Boat	3.57	4.80	4.92	4.41	6.30	6.89
Man	4.47	4.93	4.93	5.57	5.56	5.36
Baboon	3.29	3.54	3.52	3.76	4.17	4.21
Grass	3.47	3.48	3.24	3.64	5.56	5.36
Bridge	4.08	4.10	3.96	4.71	4.44	4.71

the kurtosis of the local neighboring coefficients is usually greater than 3 and increases with the increasing window size. These results show that the Gaussian distribution is not a very good probabilistic model for the local neighboring coefficients, specially when the window size is relatively large. It may be mentioned that the consideration of dependency of the local neighboring coefficients is useful, when the window size is 5×5, 7×7, or 9×9 and a further increase in the window size generally does not yield any appreciable improvement in the performance of estimators such as for denoising and restoration. Recall that the symmetric GC PDF is ensured to be positive, if $3 \leq K_f \leq 7$ [83]. Extensive simulations reveal that the kurtosis of the local neighboring coefficients for the window size of interest lies in this range. Hence, the symmetric GC PDF is ensured to be non-negative while modeling the local neighboring coefficients. Fig. 3.5 shows the empirical, Gaussian, and GC PDFs to model the local neighboring coefficients of the HL_1 subband for the test images *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial* using a window size of 9×9. Since the images are of size 512×512 and the HL_1 subband is of size 256×256, each of the PDFs is obtained by averaging the $256 \times 256 = 65536$ PDFs corresponding to each of the coefficients in the HL_1 subband. It is evident from Fig. 3.5 that the proposed PDF matches the empirical one better than the Gaussian PDF does. It is to be noted

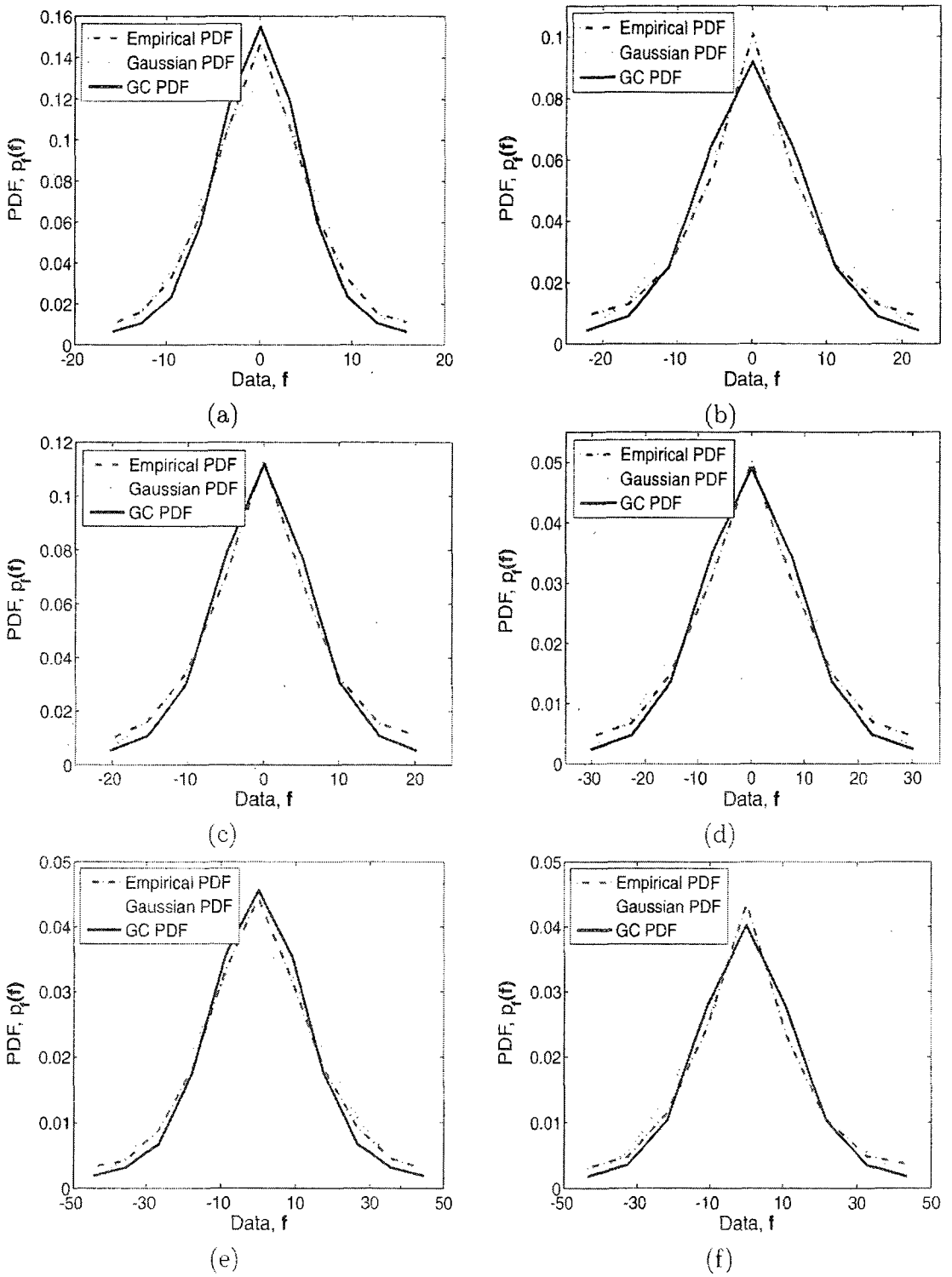


Figure 3.5: GC PDFs for modeling the local neighboring DWT coefficients using a 9×9 window in the HL_1 subband. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

that similar results are observed for other window sizes using various subbands of different test images. Table 3.4 shows the modeling performance of the Gaussian, GG, SNIG, and GC PDFs in terms of the metrics KSD and KLD of a number of test images. It is seen from Table 3.4 that both the KSD and KLD of the GC PDF are lower than that of the Gaussian, GG and SNIG PDFs, showing that the proposed PDF is a better probability model than the others for the local neighboring DWT coefficients. Similar results have been obtained for the other test images given in the database [85]. The effectiveness of the proposed PDF is shown in Chapter 5, by developing Bayesian denoising algorithms using this PDF as prior.

3.4 Conclusion

Traditional PDFs of the subband DWT coefficients of images have a limited number of parameters that are calculated from the first few moments. As a result, these PDFs cannot provide a sufficiently good statistical match for the empirical PDF of the subband DWT coefficients of an image. In this chapter, a new PDF, wherein a more appropriate number of parameters that are functions of higher-order moments can be incorporated, has been developed in order to overcome this shortfall. This PDF uses a series expansion in terms of the Hermite polynomials that are orthogonal with respect to the Gaussian weight function. Such an orthogonal series expansion has been utilized, since the expansion is valid for the non-compact support of the DWT coefficients and the Gaussian weight function of this expansion bears a resemblance with the empirical PDF. A closed-form expression for the parameters of the PDF, which has been referred to as the GH PDF, has been derived in terms of the higher-order sample-moments. A modification has been proposed in the GH PDF so that an appropriate number of parameters can be used to model the DWT coefficients of an image, while keeping the resulting PDF to be non-negative. It has been

Table 3.4: Results concerning the metrics KSD and KLD (standard deviations in parentheses) for prior function modeling of the local neighboring DWT coefficients using a 9×9 window.

Prior model	KSD		KLD	
	$\ell = 1$	$\ell = 2$	$\ell = 1$	$\ell = 2$
	Lena			
Gaussian PDF	0.085	0.089	0.221 (1.136)	0.322 (1.358)
GG PDF	0.165	0.159	0.305 (0.744)	0.348 (0.846)
SNIG PDF	0.082	0.080	0.184 (0.917)	0.247 (1.034)
GC PDF	**0.079**	**0.073**	**0.166 (0.698)**	**0.231 (0.761)**
	Barbara			
Gaussian PDF	0.080	0.100	0.214 (1.186)	0.425 (1.770)
GG PDF	0.153	0.155	0.304 (0.858)	0.408 (1.295)
SNIG PDF	0.076	0.087	0.174 (0.961)	0.325 (1.401)
GC PDF	**0.075**	**0.080**	**0.170 (0.791)**	**0.314 (1.247)**
	Boat			
Gaussian PDF	0.090	0.096	0.247 (1.161)	0.401 (1.526)
GG PDF	0.167	0.156	0.317 (0.716)	0.385 (1.018)
SNIG PDF	0.085	0.084	0.204 (0.932)	0.305 (1.162)
GC PDF	**0.083**	**0.075**	**0.182 (0.705)**	**0.283 (0.901)**
	Bridge			
Gaussian PDF	0.085	0.087	0.288 (1.262)	0.282 (1.225)
GG PDF	0.159	0.160	0.329 (0.772)	0.328 (0.758)
SNIG PDF	0.078	0.080	0.231 (0.996)	0.229 (0.989)
GC PDF	**0.071**	**0.074**	**0.206 (0.756)**	**0.203 (0.755)**
	Baboon			
Gaussian PDF	0.077	0.081	0.204 (1.388)	0.230 (1.150)
GG PDF	0.161	0.162	0.308 (1.048)	0.311 (0.762)
SNIG PDF	0.075	0.078	0.181 (1.290)	0.199 (1.021)
GC PDF	**0.074**	**0.074**	**0.166 (1.003)**	**0.174 (0.759)**
	Aerial			
Gaussian PDF	0.097	0.093	0.445 (1.607)	0.377 (1.359)
GG PDF	0.146	0.151	0.390 (1.010)	0.361 (0.810)
SNIG PDF	0.077	0.077	0.314 (1.118)	0.279 (1.008)
GC PDF	**0.066**	**0.067**	**0.312 (0.942)**	**0.259 (0.804)**

shown that this modified GH (MGH) PDF matches the empirical ones of the test images used, more closely than the commonly used PDFs such as the GG, BK, and SNIG. The proposed PDF has been adapted to model the local neighboring DWT coefficients. It has been shown that this PDF also performs better than the standard PDFs used for local neighboring coefficients. The MGH PDF will be employed as a prior in developing denoising algorithms and a watermark detector for image signals in Chapters 5 and 6, respectively.

Chapter 4

Statistics of 2D DT-CWT Coefficients of a Gaussian Signal

4.1 Introduction

This chapter deals with the statistical properties of the 2D DT-CWT coefficients of a Gaussian distributed signal both in the Cartesian and polar forms. The first level of decomposition of the DT-CWT uses wavelet filters that form a Hilbert-pair only approximately, while those at the higher levels form a Hilbert-pair that is almost exact. Hence, a significant correlation exists between the quadrature-filtered coefficients of the two trees in the first level of decomposition as compared to the other levels. As a consequence, in the Cartesian representation, the real and imaginary components of the complex coefficients are modelled as independent zero-mean Gaussian having unequal variances for the first level of decomposition and equal variances for the higher levels [86]. In the polar representation, the magnitude components are modelled by a generalized Gamma (GΓ) PDF for the first-level decomposition and a Rayleigh PDF for the higher levels [86], [87]. The corresponding phase components are modelled by an analytic PDF. The proposed PDFs of the transformed coefficients match very well with the empirical ones, as evidenced by Monte-Carlo simulations. It is also shown that the moments of the corresponding PDFs closely approximate the estimated sample-moments. Two techniques, namely, MAP-based estimation and

phase-based ridge detection are developed using the proposed PDFs. Simulation studies are carried out showing that the use of the proposed techniques provides improved estimation and detection performance of images in a noisy environment.

4.2 Gaussian Signal and 2D DT-CWT

In many applications of communication and signal processing, 2D signals are statistically modelled using the Gaussian distribution. One of the major reasons for this is that, in most cases, the statistical processors become mathematically tractable, if signals are modelled using the Gaussian distribution. Examples of such Gaussian statistical processors (GSPs) include the Wiener filter for the estimation technique and the Matched filter for the detection technique [88]. According to the central limit theorem, the statistics of a signal can be approximated well by this distribution, if the size of the data samples is large [89]. The Gaussian distribution plays a significant role even when the signal statistics is non-Gaussian. For example, a wide class of elliptically symmetric non-Gaussian distributions including the symmetrized Gamma, $S\alpha S$, Student-t, and GG distributions can be expressed in terms of a SMG distributions [25], [90]-[92]. If the signal is statistically modelled by an SMG distribution, the corresponding statistical processor can be expressed in terms of the scale components and the GSPs. In many applications, the signal is corrupted by noise, for which a commonly employed statistical model is the Gaussian distribution.

In recent years, the DT-CWT is being increasingly used in different applications of 2D signals, reasons for such are explained in Chapter 2. For statistical modeling of the 2D DT-CWT coefficients, a common assumption is that the wavelet filters form a Hilbert-pair [35], [54], [55], [58]-[60], [93]. In such a case, if a Gaussian distributed signal is decomposed by the 2D DT-CWT, the real and imaginary components of the various complex coefficients can be modelled as two independent zero-mean Gaussian

distributions having the same variance. As a consequence of this statistical model, the magnitude components of the various complex coefficients can be modelled by a Rayleigh distribution [94]. Similar to the distribution of the DWT coefficients, the distribution of the 2D DT-CWT coefficients of an image signal shows several characteristics of non-Gaussian behavior, such as high kurtosis and heavy tail. Choi *et al.* [95], Shaffrey *et al.* [96], and Ferrari and Winsor [97] have shown that the real and imaginary components of the 2D DT-CWT coefficients of such image signals can be modelled by two independent SMG distributions, assuming that the wavelet filters of the 2D DT-CWT form a perfect Hilbert-pair. Further, they have shown that the magnitude components of the 2D DT-CWT coefficients can be modelled by a scaled mixture of Rayleigh distributions. But the implementation of the 2D DT-CWT requires that the pair of wavelet filters in the first level of decomposition is different from that of the succeeding levels of decomposition and that the complex coefficients in the first one are approximately analytic [35], [86]. Therefore, a significant amount of correlation exists between the quadrature-filtered coefficients of the two trees of the DT-CWT in the first level of decomposition. Due to this correlation, when a Gaussian distributed signal is decomposed by a 2D DT-CWT, the real and imaginary components of the complex coefficients in the first level of decomposition become independent zero-mean Gaussian distributions with unequal variances. Hence, the magnitude components of such complex coefficients cannot be modelled by a Rayleigh distribution. In addition, in many applications a knowledge of the statistical properties of the corresponding phase components is required, such as edge detection [98], denoising [99], and texture identification [56]. To the best of the our knowledge, in the polar representation of the 2D DT-CWT, no systematic approach exists to statistically model the complex coefficients obtained from decomposing a Gaussian distributed signal. This chapter investigates the statistical properties of the

2D DT-CWT coefficients of a Gaussian distributed signal using the various operations associated with the transform. Since each of the two trees of the 2D DT-CWT uses the DWT, we start with the statistical properties of the 2D DWT coefficients of a Gaussian distributed signal. The statistics of the complex wavelet coefficients are analyzed both in the Cartesian (real and imaginary components) and polar (magnitude and phase components) representations. The corresponding PDFs of the coefficients and their moments are derived. The performance of the proposed PDFs are tested using the Monte-Carlo simulations and the estimation of the sample-moments. It is shown that the proposed density functions match with the empirical ones and the corresponding moments approximate the sample-moments very well. Further, MAP estimation techniques and a phase-based ridge detection technique are developed for a noisy image in order to show the effectiveness of the proposed models.

4.3 Statistics of Gaussian Signal in 2D DT-CWT

Let $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$ represent an i.i.d. Gaussian signal, where μ_ϵ and σ_ϵ^2, respectively, are the mean and variance. If the signal is transformed using a linear transformation matrix $\boldsymbol{W}$, then the transformed coefficient is also i.i.d. Gaussian having a variance

$$\hat{\sigma}^2 = \mathrm{tr}\{\boldsymbol{W}^T\boldsymbol{W}\}\sigma_\epsilon^2 \tag{4.1}$$

where $\mathrm{tr}\{\cdot\}$ represents the trace of a matrix. The 2D DWT uses two linear transformation matrices that are circulant in nature, wherein one of the matrices is obtained from the low-pass filter coefficients and the other from the high-pass filter coefficients. Let $\boldsymbol{W}_L$ and $\boldsymbol{W}_H$ denote, respectively, the low- and high-pass transformation matrices of a given decomposition level ℓ. If the transformation matrices are orthogonal, i.e., the condition $\mathrm{tr}\{\boldsymbol{W}_L^T\boldsymbol{W}_L\} = \mathrm{tr}\{\boldsymbol{W}_H^T\boldsymbol{W}_H\} = 1$ is maintained, then the signal variance has no scaling in any of the four subbands, namely, LL_ℓ, LH_ℓ, HL_ℓ, and HH_ℓ in the DWT domain. In such a case, the distribution function in the 2D DWT domain is

$\mathcal{N}(0, \sigma_\epsilon^2)$ [45]. On the other hand, if the transformation matrices are biorthogonal, i.e., $\text{tr}\{\boldsymbol{W}_L^T \boldsymbol{W}_L\} \neq \text{tr}\{\boldsymbol{W}_H^T \boldsymbol{W}_H\} \neq 1$, then the distribution function of the transformed coefficients is $\mathcal{N}(0, \hat{\sigma}^2)$, where the variance can be shown to be

$$\hat{\sigma}^2 = \begin{cases} [\text{tr}\{\boldsymbol{W}_L^T \boldsymbol{W}_L\}]^{2\ell} \sigma_\epsilon^2 & \text{for } LL_\ell \\ [\text{tr}\{\boldsymbol{W}_L^T \boldsymbol{W}_L\}]^{2\ell-1} [\text{tr}\{\boldsymbol{W}_H^T \boldsymbol{W}_H\}] \sigma_\epsilon^2 & \text{for } LH_\ell, HL_\ell \\ [\text{tr}\{\boldsymbol{W}_L^T \boldsymbol{W}_L\}]^{2\ell-2} [\text{tr}\{\boldsymbol{W}_H^T \boldsymbol{W}_H\}]^2 \sigma_\epsilon^2 & \text{for } HH_\ell \end{cases} \tag{4.2}$$

The 2D DT-CWT uses either biorthogonal or orthogonal filter coefficients in its four pairs of subbands, namely, ($LL1_\ell$, $LL2_\ell$), ($LH1_\ell$, $LH2_\ell$), ($HL1_\ell$, $HL2_\ell$), and ($HH1_\ell$, $HH2_\ell$). Except for the first-level decomposition of the transform, the wavelet filters form almost a Hilbert-pair [35], [54], [55], [60]. Let $(\mathbb{W}_L, \mathbb{W}_H)$ denote the pair of transformation matrices in the first level and $(\boldsymbol{W}_L, \boldsymbol{W}_H)$ the pair transformation matrices in a level $\ell (\ell > 1)$ through which the wavelet filters of the two trees form a Hilbert pair. The procedure of estimation of the 2D DT-CWT coefficients of a given signal using the wavelet filters has been given in Section 2.2.2. It may be found that the estimated signal variance in each subband is not only dependent on the transformation matrices, but also on the operation of Q2C. To calculate the complex coefficients of the 2D signal from the quadrature values, the filtered coefficients are partitioned into four sub-2D coefficients, termed as $q_m (m = 1, 2, 3, 4)$, corresponding to odd-odd, odd-even, even-odd and even-even indices, respectively. The coefficients in q_m are obtained from a 2D DWT-type linear transformation. Therefore, each of the sub-2D coefficients of the Gaussian random variables denoted as $\boldsymbol{q}_m (m = 1, 2, 3, 4)$ are $\mathcal{N}(0, \sigma_{\boldsymbol{q}}^2)$, and it can be easily shown that the variances of the coefficients after

the 2D signal passes through ℓ-stage filtering are

$$\sigma_q^2 = \begin{cases} [\mathrm{tr}\{\mathbb{W}_L^T\mathbb{W}_L\}]^2[\mathrm{tr}\{\boldsymbol{W}_L^T\boldsymbol{W}_L\}]^{2\ell-2}\sigma_\varepsilon^2 & \text{for } LL_{\ell(\ell>0)} \\ [\mathrm{tr}\{\mathbb{W}_L^T\mathbb{W}_L\}]^2[\mathrm{tr}\{\boldsymbol{W}_L^T\boldsymbol{W}_L\}]^{2\ell-3}[\mathrm{tr}\{\boldsymbol{W}_H^T\boldsymbol{W}_H\}]\sigma_\varepsilon^2 & \text{for } LH_{\ell(\ell>1)}, HL_{\ell(\ell>1)} \\ [\mathrm{tr}\{\mathbb{W}_L^T\mathbb{W}_L\}]^2[\mathrm{tr}\{\boldsymbol{W}_L^T\boldsymbol{W}_L\}]^{2\ell-4}[\mathrm{tr}\{\boldsymbol{W}_H^T\boldsymbol{W}_H\}]^2\sigma_\varepsilon^2 & \text{for } HH_{\ell(\ell>1)} \\ [\mathrm{tr}\{\mathbb{W}_L^T\mathbb{W}_L\}][\mathrm{tr}\{\mathbb{W}_H^T\mathbb{W}_H\}]\sigma_\varepsilon^2 & \text{for } LH_1, HL_1 \\ [\mathrm{tr}\{\mathbb{W}_H^T\mathbb{W}_H\}]^2\sigma_\varepsilon^2 & \text{for } HH_1 \end{cases} \tag{4.3}$$

4.3.1 Cartesian Form of 2D DT-CWT Coefficients

The complex coefficients of any pair of subbands of the 2D DT-CWT in the Cartesian form, namely, $\varepsilon_{u1} + \imath\varepsilon_{v1}$ and $\varepsilon_{u2} + \imath\varepsilon_{v2}$, are calculated in terms of certain algebraic operations of $q_m(m = 1, 2, 3, 4)$. According to the filtering and partitioning processes, the real and imaginary components of the complex coefficients can be expressed as [35]

$$\varepsilon_{u1} = \frac{1}{\sqrt{2}}(q_1 - q_4) \qquad \varepsilon_{v1} = \frac{1}{\sqrt{2}}(q_2 + q_3) \tag{4.4}$$

$$\varepsilon_{u2} = \frac{1}{\sqrt{2}}(q_1 + q_4) \qquad \varepsilon_{v2} = \frac{1}{\sqrt{2}}(q_2 - q_3) \tag{4.5}$$

Let the components ε_{un} and ε_{vn}, n = 1, 2, represent the samples of random variables ε_u and ε_v, respectively. The filtered coefficients of the 2D DT-CWT are designed in such a way that the random variables ε_u and ε_v that are obtained from the random variables $\boldsymbol{q}_m(m = 1, 2, 3, 4,)$ are uncorrelated [95], [100]. In addition, each of the $\boldsymbol{q}_m(m = 1, 2, 3, 4)$ is zero-mean Gaussian; therefore, the PDFs of ε_u and ε_v are independent zero-mean Gaussian. Thus, the joint PDF of ε_u and ε_v can be written as

$$p_{\varepsilon_u\varepsilon_v}(\varepsilon_u, \varepsilon_v) = p_{\varepsilon_u}(\varepsilon_u)p_{\varepsilon_v}(\varepsilon_v) \tag{4.6}$$

$$p_{\varepsilon_u}(\varepsilon_u) = \frac{1}{\sqrt{2\pi\sigma_u^2}}\exp\left[-\frac{\varepsilon_u^2}{2\sigma_u^2}\right] \tag{4.7}$$

$$p_{\varepsilon_v}(\varepsilon_v) = \frac{1}{\sqrt{2\pi\sigma_v^2}}\exp\left[-\frac{\varepsilon_v^2}{2\sigma_v^2}\right] \tag{4.8}$$

where σ_{u}^2 and σ_{v}^2 are, respectively, the variances of the real and imaginary components of the complex coefficients. The shape of any PDF is characterized by parameters, such as the skewness and kurtosis, defined by its moments [89]. All the odd-order moments of a zero-mean Gaussian signal are zero and the even-order moments can be expressed in terms of the standard deviation. In general, the n-th order moments are calculated from the expected values of the real and imaginary components that are [89]

$$E\{\varepsilon_{\mathrm{u}}^n\} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \sigma_{\mathrm{u}}^n \prod_{m=1}^{\frac{n}{2}} (2m-1) & \text{for } n \text{ even} \end{cases} \tag{4.9}$$

and

$$E\{\varepsilon_{\mathrm{v}}^n\} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \sigma_{\mathrm{v}}^n \prod_{m=1}^{\frac{n}{2}} (2m-1) & \text{for } n \text{ even} \end{cases} \tag{4.10}$$

where $E\{\cdot\}$ is the mathematical expectation. In each subband, the variances of the real and imaginary components are dependent on the equal-valued variances of $\boldsymbol{q}_m (m = 1, 2, 3, 4)$, and equal-valued covariances between the two pairs of the random variables, $(\boldsymbol{q}_1, \boldsymbol{q}_4)$ and $(\boldsymbol{q}_2, \boldsymbol{q}_3)$. It can be shown that the variances of the real and imaginary components are

$$\sigma_{\mathrm{u}1}^2 = \sigma_{\mathrm{v}2}^2 = (1 - \rho_\epsilon)\sigma_{\boldsymbol{q}}^2 \tag{4.11}$$

$$\sigma_{\mathrm{v}1}^2 = \sigma_{\mathrm{u}2}^2 = (1 + \rho_\epsilon)\sigma_{\boldsymbol{q}}^2 \tag{4.12}$$

where ρ_ϵ $(-1 \leq \rho_\epsilon \leq 1)$ is the correlation coefficient between either $\boldsymbol{q}_1$ and $\boldsymbol{q}_4$, or $\boldsymbol{q}_2$ and $\boldsymbol{q}_3$. The value of ρ_ϵ is a constant that depends on the set of filter coefficients. In the first level of decomposition, the 2D DT-CWT uses the wavelet filters that approximately form a Hilbert-pair [35]. Hence, for the first-level decomposition the value of ρ_ϵ is non-zero. Fig. 4.1 shows, in the first level of decomposition, the estimated values of the covariance between $\boldsymbol{q}_1$ and $\boldsymbol{q}_4$ as a function of the variance

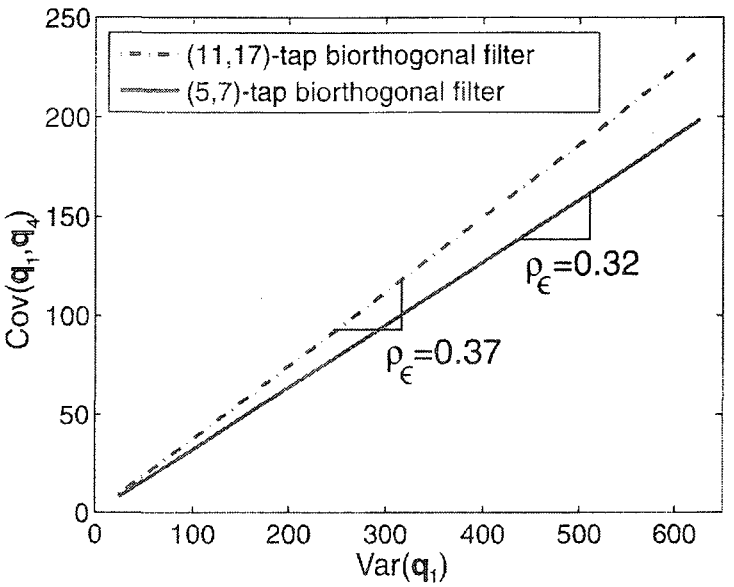


Figure 4.1: Estimation of ρ_ϵ using the variance of $\boldsymbol{q}_1$ and the covariance between $\boldsymbol{q}_1$ and $\boldsymbol{q}_4$ in the first-level decomposition of the 2D DT-CWT, wherein the filters form an approximate Hilbert-pair.

of $\boldsymbol{q}_1$ for the two sets of biorthogonal filters considered in [53]. It can be seen that ρ_ϵ, the ratio of the covariance and the variance, is constant for a particular pair of filters and increases with the length of the filter. It is to be noted that, under such a condition, when the wavelet filters approximately form a Hilbert-pair, the value of ρ_ϵ depends mainly on two factors, namely, the parameters of the wavelet filters (e.g., filter length, vanishing moments, etc.) and the covariance among the neighboring coefficients that occurs in all redundant representations [101]-[103]. What is important here is that the value of ρ_ϵ is non-zero, and hence, the original Gaussian distributed signal is transformed to two independent zero-mean Gaussian distributed components with unequal variances in the first level of decomposition. The second- and higher-level decompositions of the DT-CWT use wavelet filters that almost form a Hilbert-pair [35], [36], [54], [55], [60], and hence, the corresponding value of ρ_ϵ is close to zero. Assuming this ρ_ϵ to be zero in (4.11) and (4.12), we see that the variances of the real and imaginary components of the complex coefficients for the second- and

higher-level decompositions are

$$\sigma_{\mathrm{u}1}^2 = \sigma_{\mathrm{v}1}^2 = \sigma_{\mathrm{u}2}^2 = \sigma_{\mathrm{v}2}^2 = \sigma_q^2 \qquad \ell > 1 \tag{4.13}$$

Thus, in the Cartesian form of the 2D DT-CWT, a Gaussian distributed signal is transformed to two independent zero-mean Gaussian distributed components with equal variances in the second- and higher-level decompositions, and with unequal variances in the first-level decomposition.

4.3.2 Polar Form of 2D DT-CWT Coefficients

The complex coefficients of any pair of subbands of the 2D DT-CWT in the polar form, namely, $\varepsilon_{\mathrm{r}1}\angle\varepsilon_{\theta 1}$ and $\varepsilon_{\mathrm{r}2}\angle\varepsilon_{\theta 2}$, are calculated in terms of the real and imaginary components of the corresponding coefficients. The magnitude and phase components of the complex coefficients are

$$\varepsilon_{\mathrm{r}1} = \sqrt{\varepsilon_{\mathrm{u}1}^2 + \varepsilon_{\mathrm{v}1}^2} \qquad \varepsilon_{\theta 1} = \arctan\left(\frac{\varepsilon_{\mathrm{v}1}}{\varepsilon_{\mathrm{u}1}}\right) \tag{4.14}$$

$$\varepsilon_{\mathrm{r}2} = \sqrt{\varepsilon_{\mathrm{u}2}^2 + \varepsilon_{\mathrm{v}2}^2} \qquad \varepsilon_{\theta 2} = \arctan\left(\frac{\varepsilon_{\mathrm{v}2}}{\varepsilon_{\mathrm{u}2}}\right) \tag{4.15}$$

Let the components $\varepsilon_{\mathrm{rn}}$ and $\varepsilon_{\theta \mathrm{n}}$, $\mathrm{n} = 1, 2$, represent the samples of the random variables ε_{r} and ε_θ, respectively. Since the real and imaginary components of the complex coefficients in the first level of decomposition are independent zero-mean Gaussian having unequal variances, say σ_{u}^2 and σ_{v}^2 ($\sigma_{\mathrm{u}} \neq \sigma_{\mathrm{v}}$), the magnitude components of the coefficients have no closed-form PDF. An approximation of such a PDF is the GΓ in the form [104]

$$p_{\varepsilon_{\mathrm{r}}}(\varepsilon_{\mathrm{r}}) = \frac{2a^a \varepsilon_{\mathrm{r}}^{2a-1}}{\Gamma(a) b^{2a}} \exp\left[-a\left(\frac{\varepsilon_{\mathrm{r}}}{b}\right)^2\right] \quad \varepsilon_{\mathrm{r}} > 0, a > 0, b > 0 \tag{4.16}$$

where

$$b = \sqrt{\sigma_{\mathrm{u}}^2 + \sigma_{\mathrm{v}}^2} \qquad a = \frac{(\sigma_{\mathrm{u}}^2 + \sigma_{\mathrm{v}}^2)^2}{2(\sigma_{\mathrm{u}}^4 + \sigma_{\mathrm{v}}^4)} \tag{4.17}$$

and the n-th order expected value of the density function is [105]

$$E\{\varepsilon_{\mathrm{r}}^n\} = \frac{\Gamma(a+n/2)}{\Gamma(a)}\left(\frac{b}{\sqrt{a}}\right)^n \tag{4.18}$$

The density function of the phase components also does not have a closed-from expression. A pseudo-analytical form of the PDF of the phase components in such a situation can be obtained from the results in [104] as

$$p_{\varepsilon_\theta}(\varepsilon_\theta) = \frac{\sigma_{\mathrm{u}}\sigma_{\mathrm{v}}}{2\pi(\sigma_{\mathrm{u}}^2\sin^2\varepsilon_\theta + \sigma_{\mathrm{v}}^2\cos^2\varepsilon_\theta)} \quad -\pi < \varepsilon_\theta < \pi \tag{4.19}$$

Assuming that the values of σ_{u}^2 and σ_{v}^2 are close to each other, it can be shown that the n-th order expected value of this PDF is

$$E\{\varepsilon_\theta^n\} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \dfrac{2\sigma_{\mathrm{u}}\sigma_{\mathrm{v}}}{\sigma_{\mathrm{u}}^2+\sigma_{\mathrm{v}}^2}\dfrac{\pi^n}{n+1} & \text{for } n \text{ even} \end{cases} \tag{4.20}$$

The real and imaginary components of the complex coefficients of the second- and higher-level decompositions are modelled as independent zero-mean Gaussian having equal variances, i.e., $\sigma_{\mathrm{u}}^2 = \sigma_{\mathrm{v}}^2 = \sigma_{\mathrm{c}}^2$ (say). Hence, the magnitude components of such coefficients can be statistically modelled by the Rayleigh PDF in the form [89]

$$p_{\varepsilon_{\mathrm{r}}}(\varepsilon_{\mathrm{r}}) = \frac{\varepsilon_{\mathrm{r}}}{\sigma_{\mathrm{c}}^2}\exp\left[-\frac{\varepsilon_{\mathrm{r}}^2}{2\sigma_{\mathrm{c}}^2}\right] \quad \varepsilon_{\mathrm{r}} > 0 \tag{4.21}$$

where the n-th order expected value of this density function is given by [94]

$$E\{\varepsilon_{\mathrm{r}}^n\} = (2\sigma_{\mathrm{c}}^2)^{\frac{n}{2}}\Gamma\left(1+\frac{n}{2}\right) \tag{4.22}$$

and the phase components are modelled by the uniform PDF in the form [89]

$$p_{\varepsilon_\theta}(\varepsilon_\theta) = \frac{1}{2\pi} \quad -\pi < \varepsilon_\theta < \pi \tag{4.23}$$

where the n-th order expected value of this density function is given by

$$E\{\varepsilon_\theta^n\} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \dfrac{\pi^n}{n+1} & \text{for } n \text{ even} \end{cases} \tag{4.24}$$

It can be observed that the expressions for the PDFs and moments for the second- and higher-level decompositions given by (4.21), (4.22), (4.23), and (4.24) can also be obtained from the corresponding expressions for the first-level decomposition given by (4.16), (4.18), (4.19), and (4.20), respectively, by considering $\sigma_{\mathtt{u}} = \sigma_{\mathtt{v}} = \sigma_{\mathtt{c}}$.

4.4 Estimation and Detection

In this section, an estimation technique and a detection technique are developed for image signal processing using the statistical models that have been proposed in the previous section for the Cartesian and polar representations of the 2D DT-CWT coefficients. In both the cases, the image is assumed to be corrupted by additive white Gaussian noise (AWGN). Since the DT-CWT is a linear transform, in each subband the complex wavelet coefficient of the noisy image at the spatial location (i, j) can be written as

$$g_{ij} = f_{ij} + \varepsilon_{ij} \tag{4.25}$$

where f_{ij} and ε_{ij} are, respectively, the wavelet coefficients of the true image and noise. Let each of the wavelet coefficients g_{ij}, f_{ij}, and ε_{ij} represent the samples of the random variables $g_{\mathtt{d}}$, $f_{\mathtt{d}}$, and $\varepsilon_{\mathtt{d}}$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}, \mathtt{r}, \theta$), respectively. Here the suffixes u, v, r, and θ, respectively, denote the real, imaginary, magnitude, and phase components of the complex coefficients. The random variable $\varepsilon_{\mathtt{d}}$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}, \mathtt{r}, \theta$) can be statistically modelled as described in Section 4.3. Next, using the statistics of $\varepsilon_{\mathtt{d}}$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}, \mathtt{r}$) and the MAP criterion, we formulate the estimators for f_{ij}. Also, using the statistics of ε_{θ}, a ridge detection technique is proposed. For notational simplicity, f_{ij} in the Cartesian form is denoted as $f_{\mathtt{u}} + \imath f_{\mathtt{v}}$ and in the polar form as $f_{\mathtt{r}} \angle f_{\theta}$. Similar notations are used for g_{ij} and ε_{ij}.

4.4.1 MAP Estimation

The MAP estimators for $f_{\mathtt{d}}$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}, \mathtt{r}$) can be written as (see Section 5.3.2)

$$\begin{aligned}\hat{f}_{\mathtt{d}}(g_{\mathtt{d}}) &= \arg\max_{f_{\mathtt{d}}} p_{f_{\mathtt{d}}|g_{\mathtt{d}}}(f_{\mathtt{d}}|g_{\mathtt{d}}) \\ &= \arg\max_{f_{\mathtt{d}}} \left[p_{\epsilon_{\mathtt{d}}}(g_{\mathtt{d}} - f_{\mathtt{d}}) \cdot p_{f_{\mathtt{d}}}(f_{\mathtt{d}})\right] \\ &= \arg\max_{f_{\mathtt{d}}} \left[\ln p_{\epsilon_{\mathtt{d}}}(g_{\mathtt{d}} - f_{\mathtt{d}}) + \ln p_{f_{\mathtt{d}}}(f_{\mathtt{d}})\right] \end{aligned} \tag{4.26}$$

where $p_{\epsilon_{\mathtt{d}}}(\cdot)$ and $p_{f_{\mathtt{d}}}(\cdot)$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}, \mathtt{r}$) are the corresponding PDFs for the noise and true image, respectively. In this case, $p_{f_{\mathtt{d}}}(\cdot)$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}$) are defined as locally i.i.d. zero-mean Gaussian with conditional variance $\mathcal{N}(0, \sigma^2_{f_{\mathtt{d}}})$ [106]-[109]. Since $f_{\mathtt{r}}$ is a non-negative random variable, $p_{f_{\mathtt{r}}}(\cdot)$ is defined as locally i.i.d. Gaussian distribution with a conditional non-negative mean and variance $\mathcal{N}(\mu_{f_{\mathtt{r}}}, \sigma^2_{f_{\mathtt{r}}})$ ($\mu_{f_{\mathtt{r}}} \geq 0$).

In the Cartesian representation, $p_{\epsilon_{\mathtt{d}}}(\cdot)$ ($\mathtt{d} \in \mathtt{u}, \mathtt{v}$) are zero-mean Gaussian PDFs with subband-adaptive variances. Therefore, the estimated real and imaginary components of the image coefficients are

$$\hat{f}_{\mathtt{u}} = \frac{\sigma^2_{f_{\mathtt{u}}}}{\sigma^2_{f_{\mathtt{u}}} + \sigma^2_{\epsilon_{\mathtt{u}}}} \cdot g_{\mathtt{u}} \qquad\qquad \hat{f}_{\mathtt{v}} = \frac{\sigma^2_{f_{\mathtt{v}}}}{\sigma^2_{f_{\mathtt{v}}} + \sigma^2_{\epsilon_{\mathtt{v}}}} \cdot g_{\mathtt{v}} \tag{4.27}$$

where $\sigma^2_{\epsilon_{\mathtt{u}}}$ and $\sigma^2_{\epsilon_{\mathtt{v}}}$ are obtained using (4.11)-(4.13). The variances, $\sigma^2_{f_{\mathtt{u}}}$ and $\sigma^2_{f_{\mathtt{v}}}$, of the true image components are estimated locally from the noisy observation using a ML method as [106]

$$\sigma^2_{f_{\mathtt{u}}} = \max\left(\frac{1}{\aleph}\sum_{\mathcal{S}_{ij}} g^2_{\mathtt{u}} - \sigma^2_{\epsilon_{\mathtt{u}}}, 0\right) \tag{4.28}$$

$$\sigma^2_{f_{\mathtt{v}}} = \max\left(\frac{1}{\aleph}\sum_{\mathcal{S}_{ij}} g^2_{\mathtt{v}} - \sigma^2_{\epsilon_{\mathtt{v}}}, 0\right) \tag{4.29}$$

where $\mathcal{S}_{ij}$ is a square-shaped local neighborhood region centered at g_{ij} and $\aleph$ is the total number of data samples in this region. The estimated image is obtained from the inverse transformation of $\hat{f}_{\mathtt{u}} + \imath\hat{f}_{\mathtt{v}}$. In the polar representation, the GΓ PDF with

subband-adaptive parameters is sufficient to define $p_{\epsilon_r}(\cdot)$. In such a case, to maximize (4.26), we need to solve the equation

$$\frac{1-2a}{g_r - f_r} + \frac{2a}{b^2}(g_r - f_r) - \frac{f_r - \mu_{f_r}}{\sigma^2_{f_r}} = 0 \tag{4.30}$$

Thus, a closed-form expression for the MAP estimator of f_r can be obtained as

$$\hat{f}_r = \max\left(\frac{1}{4a\sigma^2_{f_r} + 2b^2}\Big((4a\sigma^2_{f_r} + b^2)g_r + b^2\mu_{f_r} - \sqrt{b^4(g_r - \mu_{f_r})^2 - 4b^2\sigma^2_{f_r}(1-2a)(2a\sigma^2_{f_r} + b^2)}\Big), 0\right) \tag{4.31}$$

It can be shown that the parameter a lies between $\frac{1}{2}$ and 1. Therefore, the value of $\hat{f}_r$ given in (4.31) is real and non-negative. The MAP estimator considering that $p_{\epsilon_r}(\cdot)$ as a Rayleigh PDF is obtained as a specific case of (4.31), when $a = 1$ and $b = \sqrt{2}\sigma_c$. Thus, the MAP estimator for the second- or higher-level decomposition is given by

$$\hat{f}_r = \max\left(\frac{1}{2(\sigma^2_{f_r} + \sigma^2_c)}\Big((2\sigma^2_{f_r} + \sigma^2_c)g_r + \mu_{f_r}\sigma^2_c - \sqrt{\sigma^4_c(g_r - \mu_{f_r})^2 + 4\sigma^4_{f_r}\sigma^2_c + 4\sigma^2_{f_r}\sigma^4_c}\Big), 0\right) \tag{4.32}$$

Note that an estimator similar to (4.32), which has been given in [110], can be obtained by letting $\mu_{f_r} = 0$. The mean μ_{f_r} and variance $\sigma^2_{f_r}$ of the magnitude components of the true image wavelet coefficients are estimated locally from the noisy observation using a ML method as

$$\mu_{f_r} = \max\left(\frac{1}{\aleph}\sum_{S_{ij}} g_r - \mu_{\epsilon_r}, 0\right) \tag{4.33}$$

$$\sigma^2_{f_r} = \max\left(\frac{1}{\aleph}\sum_{S_{ij}} (g_r - \mu_{g_r})^2 - \sigma^2_{\epsilon_r}, 0\right) \tag{4.34}$$

where μ_{g_r} is the mean of noisy samples,

$$\mu_{\epsilon_r} = \begin{cases} \dfrac{b}{\sqrt{a}}\left(\dfrac{\Gamma(a+\frac{1}{2})}{\Gamma(a)}\right) & \text{for } \ell = 1 \\ \sqrt{\dfrac{\pi}{2}}\cdot\sigma_c & \text{for } \ell > 1 \end{cases} \tag{4.35}$$

and

$$\sigma^2_{\varepsilon_{\mathbf{r}}} = \begin{cases} b^2\left(1 - \frac{1}{a}\left(\frac{\Gamma(a+\frac{1}{2})}{\Gamma(a)}\right)^2\right) & \text{for } \ell = 1 \\ \frac{4-\pi}{2}\cdot\sigma^2_{\mathbf{c}} & \text{for } \ell > 1 \end{cases} \tag{4.36}$$

Şendur and Selesnick [66], and Achim and Kuruoğlu [67] have shown that the estimated magnitude components of the DT-CWT, along with the noisy phase components, are sufficient for a good noise reduction performance. The estimated image is, therefore, obtained from the inverse transformation of $\hat{f}_{\mathbf{r}}\angle g_\theta$.

4.4.2 Ridge Detection

The phase of the 2D DT-CWT carries significant information near the image discontinuities, such as ridges and edges. Thus, the phase statistics of the complex wavelet coefficients can be used to detect a ridge in an image. In particular, we are interested in detecting the ridge of an image corrupted by a random Gaussian signal. In such a case, we need to distinguish the phase components of the ridges from that of the smooth areas. In the smooth areas, the phase characteristics are totally random due to the Gaussian noise. Therefore, the sample-variance [111] of the phase components in this region is equal to or greater than the phase variance of the noise. On the other hand, in a region where a ridge exists, the phase components from the true signal are much stronger than that of the noise. Hence, in such a region, the sample-variance of the phase components is smaller than that of the noise. Taking into account the inter-scale phase consistency [99], we use the phase components of the parent subband to improve the detection. Finally, the phase of a ridge in a subband of the ℓ-th decomposition can be detected as

$$\angle\hat{f}_{\theta_\ell} = \angle g_{\theta_\ell}\cdot \mathbf{1}_\ell\{\sigma^2_{g_\theta} < \sigma^2_{\varepsilon_\theta}\} - 2\angle g_{\theta_{\ell-1}}\cdot \mathbf{1}_{\ell-1}\{\sigma^2_{g_\theta} < \sigma^2_{\varepsilon_\theta}\} \tag{4.37}$$

where $\mathbf{1}_\ell\{\cdot\}$ is an indicator function that takes a value of 1 or 0 depending on whether its logical value is true or false. $\sigma^2_{g_\theta}$ is the sample-variance of the noisy phase compo-

nents in the local neighborhood $\mathcal{S}_{ij}$, and $\sigma^2_{\epsilon_\theta}$ is the phase variance of the noise that can be calculated as

$$\sigma^2_{\epsilon_\theta} = \begin{cases} \frac{\pi^2}{3} \cdot \frac{2\sigma_{\epsilon_u}\sigma_{\epsilon_v}}{\sigma^2_{\epsilon_u} + \sigma^2_{\epsilon_v}} & \text{for } \ell = 1 \\ \frac{\pi^2}{3} & \text{for } \ell > 1 \end{cases} \tag{4.38}$$

4.5 Experimental Results

To test the performance of the proposed statistical models, two sets of experiments are conducted. In the first set of experiments, Monte-Carlo simulations are carried out to test how well the proposed PDFs match the empirical ones. The closeness of fit of the sample-moments and proposed theoretical moments are also examined through these experiments. In the second set of experiments, we evaluate the performance of the estimators and detector developed in Section 4.4 as applications of the proposed PDFs. The complex coefficients in all these experiments are obtained by employing a 3-level 2D DT-CWT, wherein an (11,17)-tap biorthogonal filter in the first level of decomposition and 6-tap Q-shift orthogonal filters in the second- and higher-level decompositions [53] are used.

4.5.1 Experiments on Monte-Carlo Simulations

In these experiments, each of the Monte-Carlo simulations uses one thousand runs. In each of the runs, a Gaussian distributed signal $\mathcal{N}(10, 100)$ of size 512×512 is generated. Fig. 4.2 shows the contour plots of the joint PDFs of the real and imaginary components of the coefficients of the HL subbands of the first- and second-level decompositions, namely, $HL1_1$, $HL2_1$, $HL1_2$, and $HL2_2$. It can be seen from this figure that the real and imaginary components of the complex coefficients for each subband are independent zero-mean Gaussian having unequal variances for the first level of decomposition, and have equal variances for the second-level decomposition.

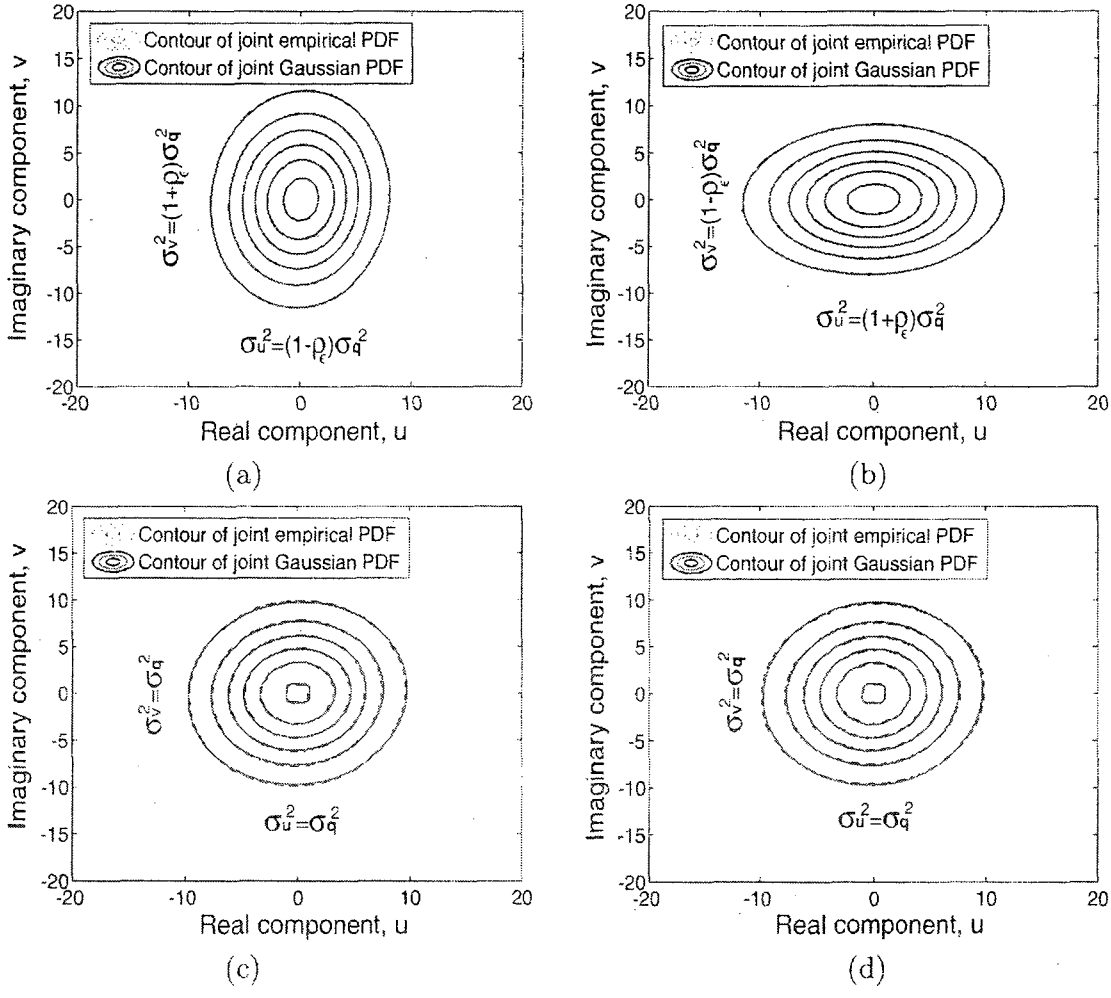


Figure 4.2: Contour plots of the joint PDFs of the real and imaginary components of the 2D DT-CWT coefficients for the subbands (a) $HL1_1$, (b) $HL2_1$, (c) $HL1_2$, and (d) $HL2_2$.

It may be pointed out that the variances of the real and imaginary components are not equal in the other subbands of the first level of decomposition. Except in the first-level decomposition, the real and imaginary components in all the subbands at levels 2 and 3 show equal variances. From Fig. 4.2, it is also seen that the variances of the real and imaginary components of any subband are dependent on σ_q^2 and ρ_ϵ, which themselves are dependent on the filter coefficients, as described in Section 4.3. Further, it is observed that the contours of the proposed joint PDFs very closely match those of the empirical ones.

Fig. 4.3 shows the PDFs of the magnitude components of the complex coefficients for the above mentioned subbands. In the case of the subbands of the first-level decomposition, the magnitude components of the coefficients are modelled using the GΓ PDF, whereas that of the second-level by using the Rayleigh PDF. It is observed form Figs. 4.3(c) and 4.3(d) that the Rayleigh PDF matches with the corresponding empirical PDF very well both in the peak and tail regions. This is consistent with the fact that the Rayleigh PDF, an exact density function, has been used in modeling the magnitude components of the coefficients in the second level of decomposition. Similar result holds for the other subbands of the second level of decomposition as well as for all the subbands of the third level of decomposition. On the other hand, the GΓ PDF, which is an approximate density function that has been used in the first-level decomposition, matches very well in the tail part of the empirical one, but has a slight mismatch in the peak region. It is to be pointed out that similar result holds for the other subbands in the first level of decomposition. In many applications of signal processing, the tail part is more important than the peak part of the empirical PDF [67]. Moreover, the GΓ PDF is mathematically tractable. Hence, the GΓ is an appropriate PDF for modeling the magnitude components of the 2D DT-CWT coefficients in the first-level decomposition, provided the signal being

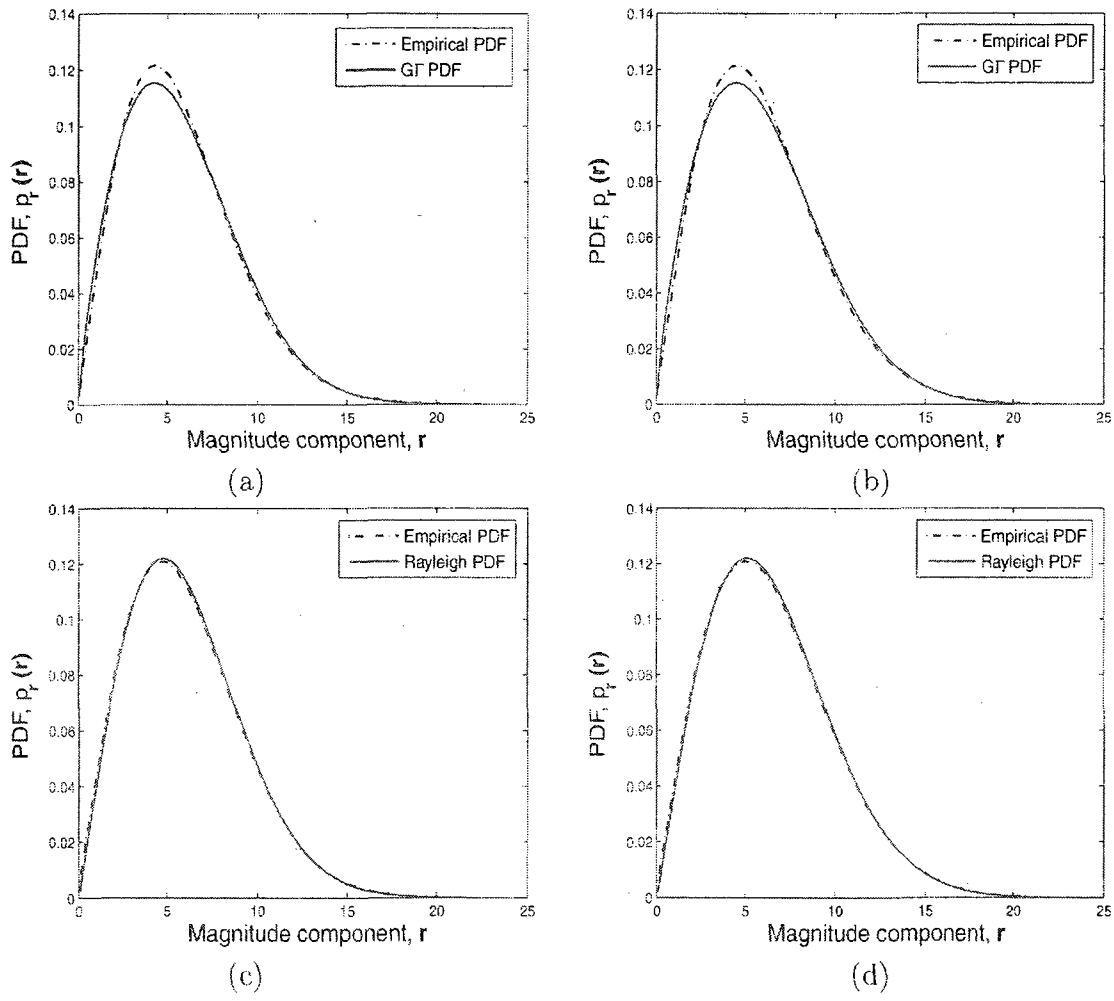


Figure 4.3: PDFs of the magnitude components of the 2D DT-CWT coefficients for the subbands (a) $HL1_1$, (b) $HL2_1$, (c) $HL1_2$, and (d) $HL2_2$.

decomposed follows the Gaussian distribution. Fig. 4.4 shows the corresponding PDFs of the phase components of the coefficients in polar coordinates. It is observed that the pseudo-analytic PDF given by (4.19) very closely approximates the distribution of the phase components of the 2D DT-CWT coefficients of the subbands under consideration in the first-level decomposition. It may be mentioned that similar performance has been observed for this PDF to model the phase components of the complex coefficients of the other subbands in the first-level decomposition. It is seen from Figs. 4.4(c) and 4.4(d) that the uniform PDF used in the second-level decomposition matches very well with the empirical ones for the coefficients of the subbands $HL1_2$ and $HL2_2$. Similar results hold for the other subbands in level 2 and all the subbands in level 3. Fig. 4.5 shows the non-zero sample-moments [89] as well as that of the proposed theoretical moments up to the fourth order for the different subbands, where we have used a Gaussian distributed signal having a mean of 0.1 and a variance of unity. Such a choice of the variance renders the values of all the higher-order moments of the real, imaginary, and magnitude components to be less than unity, and thus, a single plot may be used to compare all the higher-order moments of a particular component. In Fig. 4.5, subbands 1 to 6 are obtained from the first-level decomposition, 7 to 12 from the second-level decomposition, and the rest from the third-level decomposition. It is observed from this figure that for the different subbands, the proposed theoretical moments approximate the sample-moments very well for both the Cartesian and polar representations of the DT-CWT coefficients. As the level of the decomposition increases, the number of data samples decreases. As a consequence, the deviation between the sample and proposed moments increases. In addition, this deviation also increases with the order of the moments. These two observations resulting from our experiments are consistent with the fact that the deviation between the sample and theoretical moments increases with

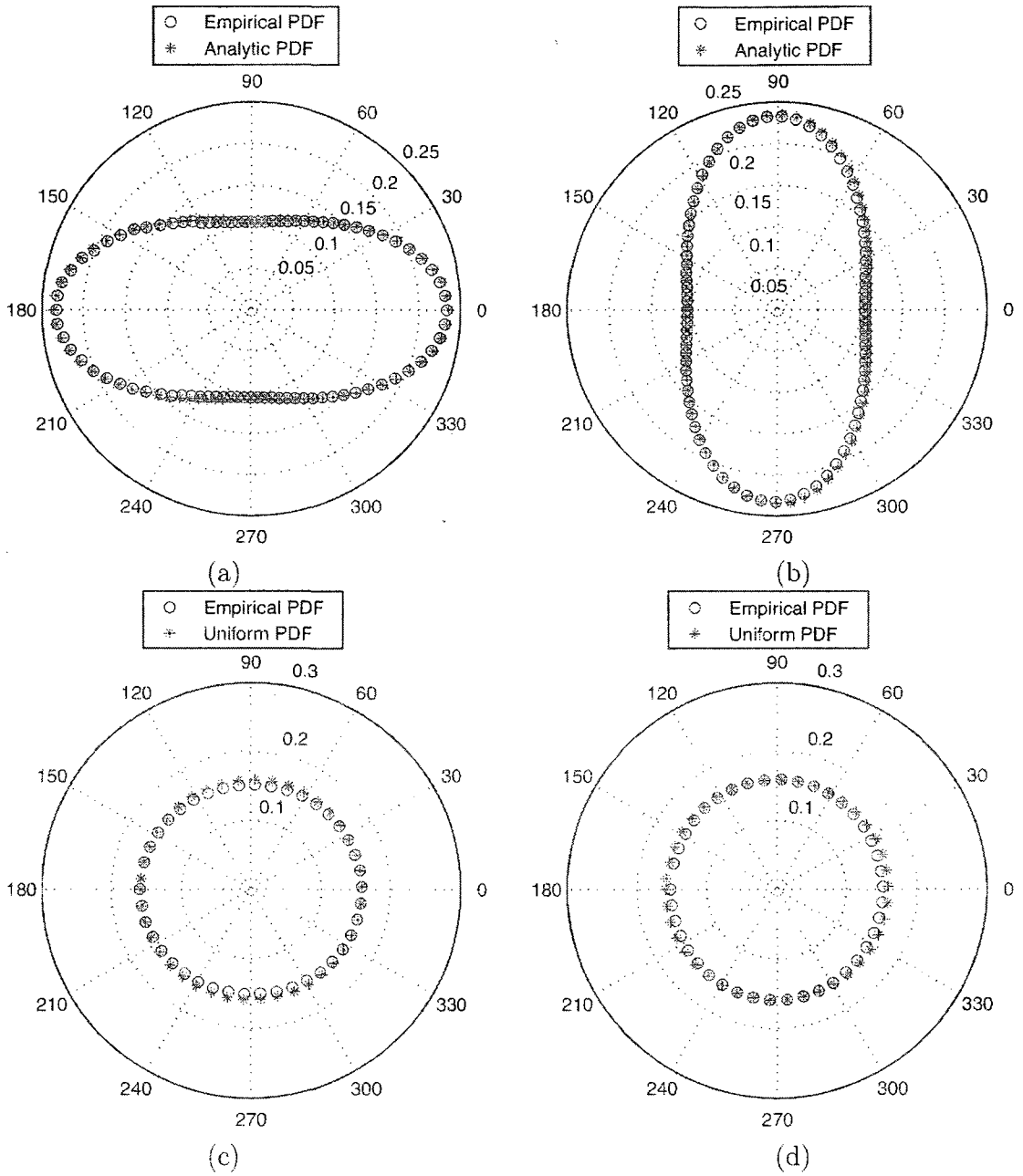


Figure 4.4: PDFs of the phase components of the 2D DT-CWT coefficients for the subbands (a) $HL1_1$, (b) $HL2_1$, (c) $HL1_2$, and (d) $HL2_2$.

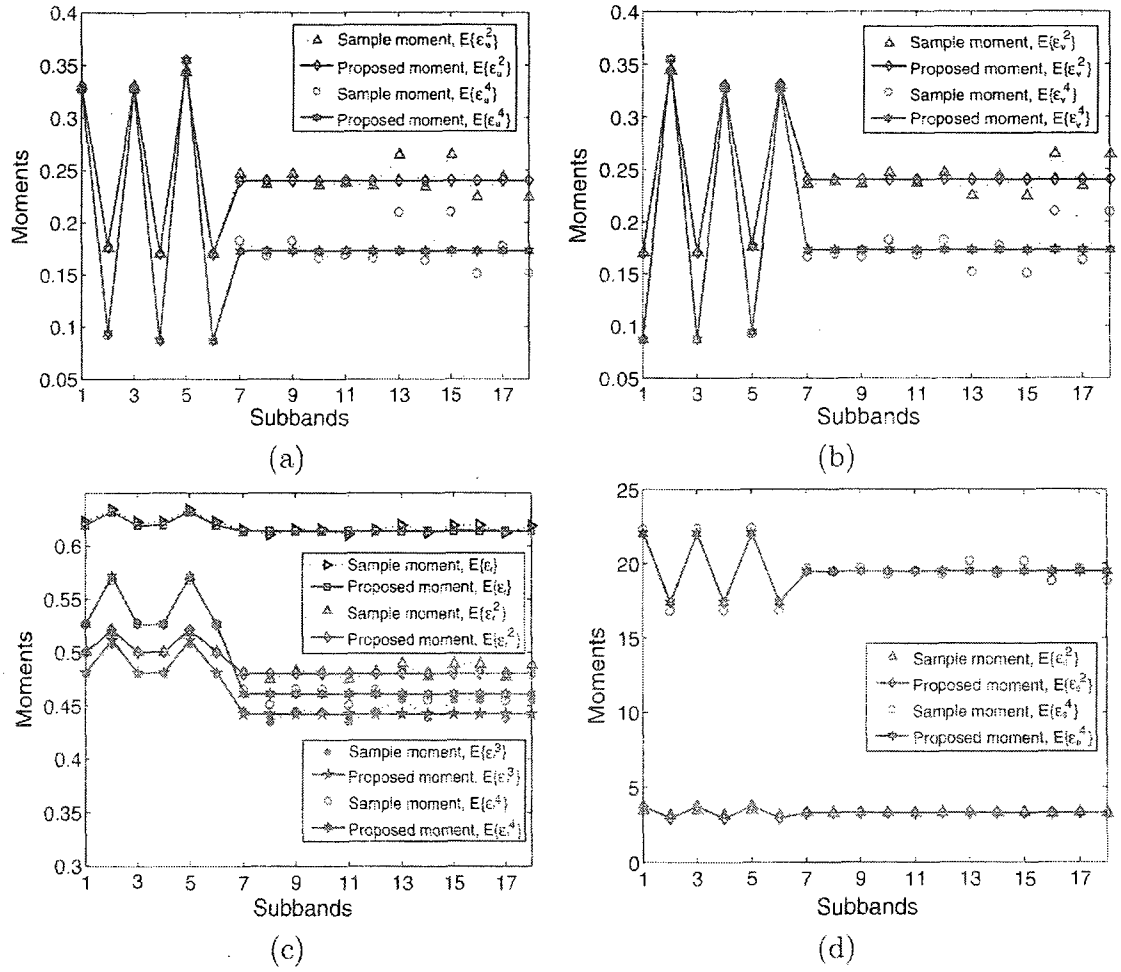


Figure 4.5: Moments of the 2D DT-CWT coefficients for different subbands of a 3-level decomposition. The components of the coefficients are (a) real, (b) imaginary, (c) magnitude, and (d) phase.

decreasing number of data samples as well as with increasing order of moments [111]. In our experimental study, however, it has been observed that for the moments considered, the average deviation is less than 5% for any of the decomposition levels. If we consider all the subbands of the 3-level DT-CWT, then the average deviation is even less than 0.2%. Thus, the results of Fig. 4.5 concerning the moments show that the proposed PDFs can model different components of the DT-CWT coefficients of the Gaussian distributed signal very well.

4.5.2 Experiments on Estimation and Detection

In these experiments, we use several grayscale test images of size 512×512, such as *Lena*, *Boat*, *Barbara*, and *Bridge*. The images are corrupted by additive Gaussian noise sequences with standard deviations of 10, 15, 20, 25, 30, and 50. The window sizes used for calculating the local statistics of the components of the complex coefficients are 3×3, 5×5, 7×7, and 9×9. Here, we present the MAP estimation results only for the image *Lena* with a noise standard deviation of 20 using a 7×7 window and the phase-based ridge detection results only for the image *Boat* with a noise standard deviation of 10 using the same window size. In the former case, we have chosen the image *Lena*, as it is one of the most widely used test images that possesses a good mixture of edges (horizontal, vertical, diagonal, curve, etc.) and textures, whereas for latter case, we have chosen the image *Boat* as it possesses ridges with various orientations.

Fig. 4.6 shows a cropped section of the original image *Lena*, its noisy version, and the output images of the MAP estimators both in the Cartesian and polar forms. Fig. 4.6(c) shows the output image from the estimator in the Cartesian form considering the variances of the real and imaginary components to be all equal in each level, while Fig. 4.6(d) shows the corresponding image for the case where they are unequal in the first level and equal in the succeeding two higher levels. The PSNR is 30.71 dB in the

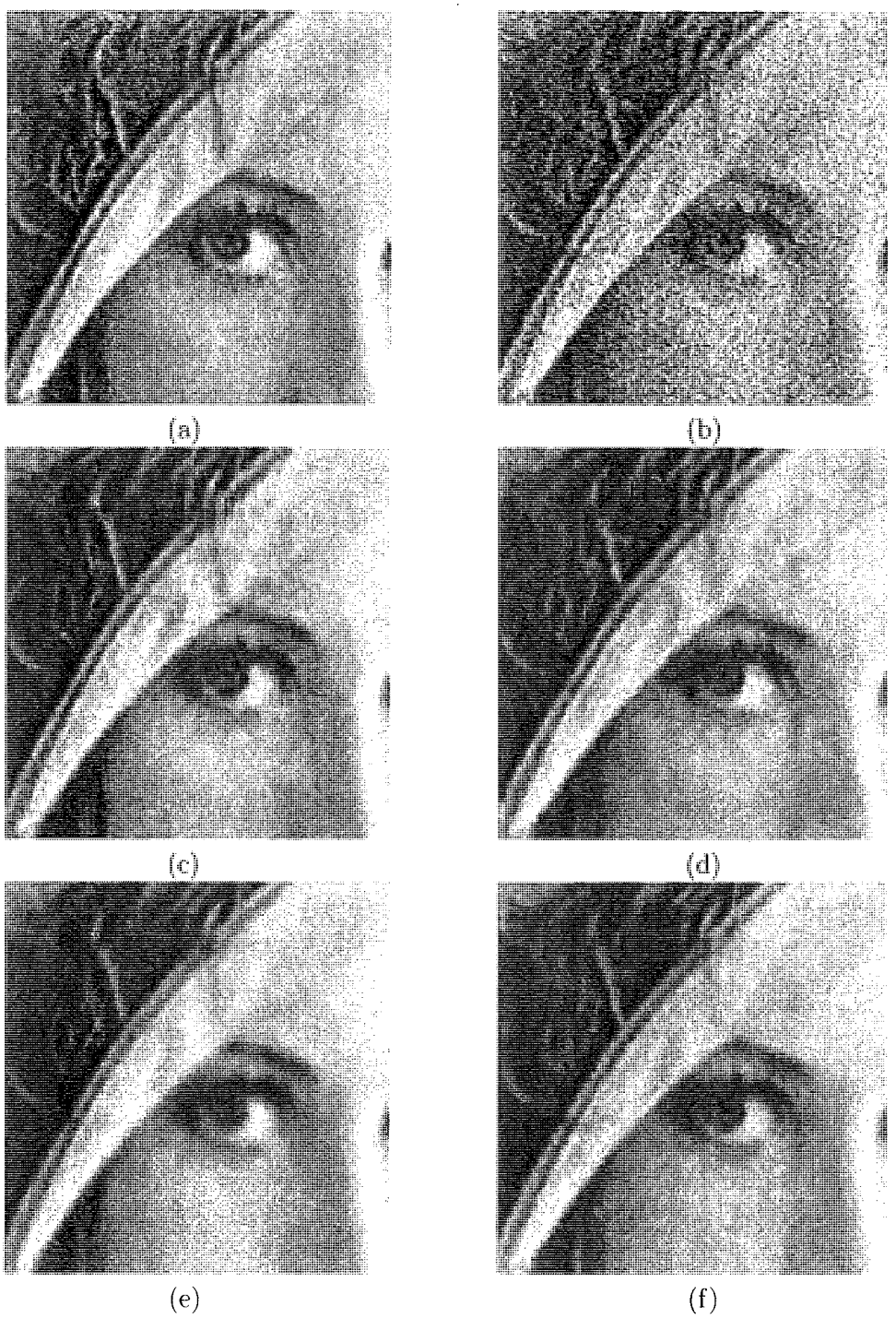


Figure 4.6: Output of the cropped image *Lena* using various MAP estimators, wherein the noise standard deviation is 20. (a) Original image and (b) noisy image. (c) Estimated image of (b) using equal variance for the real and imaginary components in $\ell = 1$ (PSNR=30.71 dB). (d) Estimated image of (b) using unequal variances for the real and imaginary components in $\ell = 1$ (PSNR=31.74 dB). (e) Estimated image of (b) using Rayleigh PDF for the magnitude components in $\ell = 1$ (PSNR=31.40 dB). (f) Estimated image of (b) using GΓ PDF for the magnitude components in $\ell = 1$ (PSNR=31.95 dB).

former case and 31.74 dB in the latter case. The corresponding output images from the estimators in the polar form are shown in Figs. 4.6(e) and 4.6(f), respectively. It is found that the estimator using the Rayleigh PDF as the noise model in the first level provides an overall PSNR of 31.40 dB, whereas that using the GΓ PDF provides a PSNR of 31.95 dB. The improvement of the estimation performance, such as the proposed estimator using the polar form preserving the edges on the hat and eye-lid of *Lena* more significantly than the estimator using the Rayleigh PDF as the noise model in the first level, can also be observed from the output images (see Figs. 4.6(e) and 4.6(f)). These results show that significant improvements in the estimation performance can be obtained using the proposed statistical models both in the Cartesian form and in the polar form. In the polar form, however, the estimator provides PSNR values that are higher than that in the Cartesian form, since the magnitude components are more shift-invariant than the individual real and imaginary components [66], [67]. Hence, an estimator using the magnitude components usually provides a better estimate of a noisy image with less amount of Gibb's artifact as compared to the one using the individual real and imaginary components. The reduced amount of artifacts in the output obtained in the polar form as compared to that in the Cartesian form is evident, when we compare the images in Figs. 4.6(e) and 4.6(f) with that in Figs. 4.6(c) and 4.6(d). The improvement in the performance of the estimation of the wavelet coefficients, whether in Cartesian form or in polar form, is mainly due to the fact that the content of the clean signal increases as the scale increases [45], [107]. Therefore, a significant improvement of the estimation performance can be achieved at a lower level of decomposition than that at higher ones. In other words, the statistical model in the first level of decomposition is more crucial for any estimation technique.

Fig. 4.7(a) shows a cropped section of the original image *Boat*. The noise-free phase

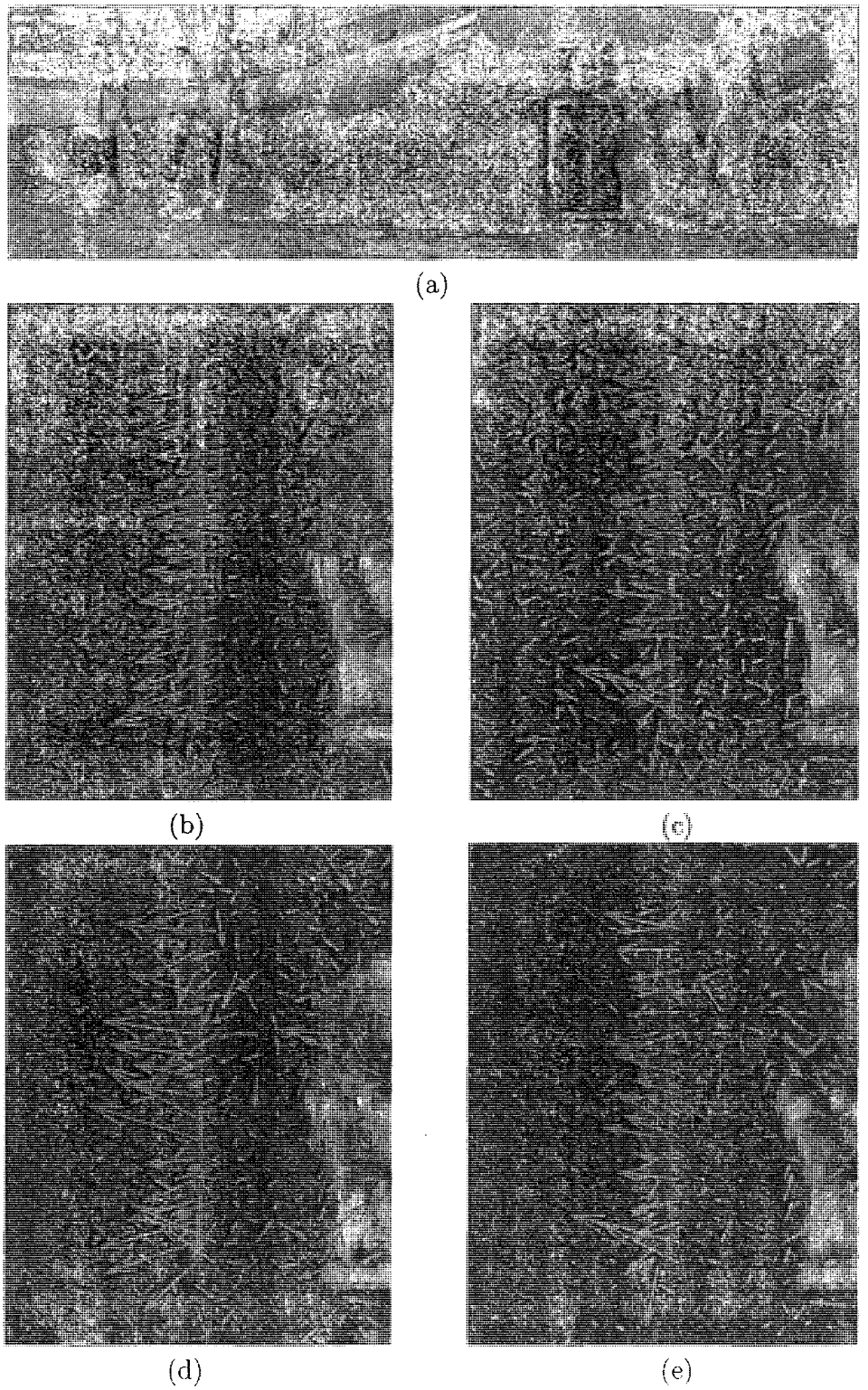


Figure 4.7: Ridge detection by phase for the cropped image *Boat*, wherein the noise standard deviation is 10. (a) Original image. (b) Phase of the zoomed-in section of (a). (c) Phase of the corresponding zoomed-in section of the noisy image. (d) Detected phase from (c) using uniform PDF in $\ell = 1$ (MSE=41.49 dB). (e) Detected phase from (c) using the proposed PDF in $\ell = 1$ (MSE=41.21 dB).

vectors of the ridges inside the rectangle of the original image are shown in Fig. 4.7(b) and the corresponding noisy ones in Fig. 4.7(c). In order to detect the phase components, we use the complex coefficients of the $HL1_1$ and $HL1_2$ subbands of the 2D DT-CWT. Fig. 4.7(d) shows the detected phase vectors assuming a uniform PDF for the phase components in the first two levels, while Fig. 4.7(e) shows the corresponding detected vectors by considering the proposed analytic PDF in the first level and uniform PDF in the second level. The MSE between the original and estimated phase components is 41.49 dB in the former case, and 41.21 dB in the latter case. Thus, in terms of the MSE, an improved detection performance is achieved using the proposed PDF instead of using the uniform PDF. The phase components of the complex wavelet coefficients, specially those in the lower level of decompositions, are sensitive to noise [112], [113]. This is due to the fact that the phase components are more shift-variant than the magnitude components and the lower level of decompositions are contaminated by noise more significantly than the higher level ones. Hence, a more accurate statistical model of the phase components for the first-level decomposition can produce a better quality ridge detection. The improved ridge detection performance can be observed from Figs. 4.7(d) and 4.7(e), in which the random phase vectors appearing in the noisy image outside the ridges are removed significantly by the proposed detector as compared to that using the uniform PDF.

4.6 Conclusion

In this chapter, probabilistic models of a Gaussian distributed signal have been investigated in the 2D DT-CWT domain. It has been shown that in the Cartesian form, the real and imaginary components of the complex coefficients are independent zero-mean Gaussian having unequal variances in the first level of decomposition and equal variances in the second- and higher-level decompositions. The variances of the

transformed signals are scaled by a factor that depends on the transformation matrices and the correlation between the quadrature-filtered coefficients of the two trees. Since the first-level decomposition of the 2D DT-CWT uses wavelet filters that form an approximate Hilbert-pair, the correlation between the quadrature-filtered coefficients of the two trees cannot be ignored in this level. Consequently, the variances of the real and imaginary components of the complex coefficients in the first level of decomposition are not the same. In the polar form, the magnitude components of the complex coefficients have been modelled by a GΓ distribution for the first-level decomposition and by a Rayleigh distribution for the second- and higher-level decompositions. The corresponding phase components have been modelled by an analytic PDF. Monte-Carlo simulations have shown that the proposed statistical models very closely match the empirical models. The estimates of the moments and their close proximity to the sample-moments further reinforce the accuracy of the proposed statistical models. In addition, the proposed statistical models are used to obtain MAP estimators and a phase-based ridge detector for noisy images. Experimental results have shown that the proposed statistical models play a significant role in these applications, particularly when the input signal is characterized by a Gaussian distribution. It is to be pointed out that the statistical models of Cartesian and polar representations are further used for reducing AWGN of images and video in Chapters 5 and 8, respectively.

Chapter 5

Wavelet-Based Image Denoising Using Proposed MGH PDF

5.1 Introduction

In this chapter, it is shown that wavelet-based image denoising problem can be well addressed by the MGH PDF developed in Chapter 3. First, we provide a brief introduction to show that various image degradation models can be represented by an AWGN model. The role of DWT in reducing AWGN from a corrupted image is also discussed. Next, statistical estimation techniques for the noisy wavelet coefficients are developed using two standard criteria, viz., MMSE [77], [78] and MAP [34], [79]. Both the subband- and locally-adaptive image denoising methods that use the MGH PDF are considered. Experiments are carried out using the decimated DWT and DT-CWT coefficients of an image to evaluate the performance of the denoising methods developed. It is shown that the proposed MGH PDF allows us to set an appropriate amount of shrinkage of the noisy wavelet coefficients to obtain an output image that meets the requirement of a desired level of PSNR, SSIM or visual quality. It is also shown that introduction of the proposed PDF in the wavelet-based image denoising methods improves the estimation performance as compared to the existing denoising methods.

5.2 Image Denoising and DWT

The distortion of images by noise is common during its acquisition, processing, compression, storage, transmission, and reproduction. The aim of denoising is to remove the noise while keeping the signal features as much as possible. Traditional algorithms perform image denoising in the pixel-domain. However, the application of the DWT in image denoising, compression, etc. has shown remarkable success over the last decade. The noise that corrupts the signal may be signal dependent or signal independent [114]. If the noise is independent of the signal, the DWT coefficient of the noisy image at the spatial location (i, j) of a subband is given by

$$g_{ij} = f_{ij} + \varepsilon_{ij} \tag{5.1}$$

where ε_{ij} is the DWT coefficient of the additive noise at the reference location. In most of the cases, the noise is assumed to be zero-mean white Gaussian (e.g., thermal noise). with a variance σ_{ε}^2. In the literature, this type of degradation model is very often referred to as the AWGN model. For signal dependent noise, such as the photon noise, film-grain noise, and speckle noise, the DWT coefficients of the noisy image can also be represented as AWGN model, but with some modifications. For example, in a low light imaging system the photon noise can be assumed as a Poisson process. In such a case, the degradation model can be treated as AWGN [115] by considering $g_{ij} = \tilde{g}_{ij}/\varrho$ and the noise variance to be index dependent, $\sigma_{\varepsilon_{ij}}^2 = E\{\tilde{g}_{ij}/\varrho^2\}$, where $\tilde{g}_{ij}$ is the wavelet coefficient of the observed image and ϱ is a constant of proportionality. When images recorded on a photographic film are digitized, a noise called the film-grain noise corrupts the image. Froehlich *et al.* [116] have shown that the corruption by film-grain noise can be modelled as

$$g_{ij} = f_{ij} + \mathrm{k}(f_{ij})^{\chi} g_{1_{ij}} + g_{2_{ij}} \tag{5.2}$$

where $g_{1_{ij}}$ and $g_{2_{ij}}$ are, respectively, the samples from two zero-mean i.i.d. Gaussian random processes with variances $\sigma^2_{g_1}$ and $\sigma^2_{g_2}$, k is a proportionality factor, and χ is an exponent usually having a value of 0.5. It has been shown that the degradation by film-grain noise can be modelled as AWGN by considering the index dependent noise variance $\sigma^2_{\varepsilon_{ij}} = \mathrm{k}^2 E\{g_{ij}\}\sigma^2_{g_1} + \sigma^2_{g_2}$ [115]. Another type of noise encountered in a practical imaging system is the speckle noise, which is generated when the system uses coherent radiation pattern. Examples of such noise can be found in the ultrasound, SAR, laser, and sonar images. The most widely used model for speckle noise is the multiplicative model, which is given by [115], [117]

$$g_{ij} = f_{ij} v_{ij} \tag{5.3}$$

where v_{ij} is a stationary noise uncorrelated with the image having mean μ_v and variance σ^2_v. Depending on the application, distribution of this noise can be assumed to be Gaussian as in SAR images, Rayleigh as in ultrasound images, etc. Such a multiplicative degradation model can also be converted to the AWGN model by considering the index dependent noise variance to be [115]

$$\sigma^2_{\varepsilon_{ij}} = \frac{{\sigma_v}^2}{1 + {\sigma_v}^2} \left(\sigma^2_{g_{ij}} + \mu^2_{g_{ij}} \right) \tag{5.4}$$

where $\mu_{g_{ij}}$ and $\sigma^2_{g_{ij}}$, respectively, are the mean and variance of the degraded DWT coefficients at each index (i, j). Homomorphic approach [78] is another way of converting the multiplicative degradation model to the additive one, wherein the log-transformation is used in (5.3). Since most of the practical noise encountered can be either modelled or remodelled as additive Gaussian noise, denoising of images corrupted by AWGN stands as a classical problem in signal-processing.

The DWT-based signal denoising techniques for AWGN can be found in [12]-[15], [21], [22], [25], [26], [42], [51], [65], [72], [73], [95], [106]-[108], [118]-[124], where the wavelet shrinkage method proposed by Donoho [119], [120] is the pioneering work. The

method described in [120] provides a mini-max optimal solution. Another criterion commonly used is the MSE, even though it does not match well with the characteristics of the HVS [125]. However, since it is a measure of the noise power [114], and is mathematically convenient in the context of optimization, it is employed as a denoising criterion in many image denoising applications. The DWT used in image denoising can be of many types, such as orthogonal/biorthogonal, real/complex-valued, separable/non-separable, or decimated/non-decimated. Due to the shift-invariance property, the non-decimated separable DWT [45] improves the image denoising performance in MSE approximately by 1 dB as compared to that of the decimated representation [12], [71], [126], [127]. To reduce the computational load of the non-decimated wavelet transform as well as the Gibbs artifacts, cycle-spinning has been proposed [51]. Image denoising by using transforms with improved directional selectivity, such as the DT-CWT [52], ridgelet [39], curvelet [40], contourlet [41], and steerable pyramid [42], [43] have also been introduced.

There are two major approaches for statistical wavelet-based denoising [128], [129]. The first approach is to design a statistically optimal threshold parameter for some nonlinear thresholding or shrinkage functions. Image denoising methods using such an approach include the VisuShrink [119], SureShrink [120], BayesShrink [13], amplitude-scale-invariant Bayes estimator [130], NeighShrink [131], Spatial-correlation thresholding [132], and empirical Bayes thresholding (EBayesThresh) [133]. Standard choices for nonlinear thresholding or shrinkage functions are soft- and hard-thresholding [134], firm-shrinkage [135], and non-negative garrote shrinkage [136]. Methods using such shrinkage functions are computationally simple, but have certain drawbacks in view of the fact that the functions representing the nonlinearity are arbitrarily chosen. For example, the soft-thresholding technique yields a biased estimate with a moderate variance, whereas the hard-thresholding technique yields a less biased estimate but

with a higher variance [136]. In the second approach, instead of using an arbitrary function to represent the nonlinearity, the shrinkage function is designed by minimizing a Bayesian risk, typically under the MMSE criterion [15], minimum mean absolute error criterion [137] or MAP criterion [65], [138]. The development of an optimal threshold or shrinkage function is usually based on the PDF or simply the prior function of the image wavelet coefficients. The marginal PDF of a true image and its self-similar subband wavelet coefficients are unimodal and sharply-peaked in nature [74], [139]. Therefore, higher-order statistics are necessary to describe the sharp central cusp of the density function of the image coefficients. The parameters of a PDF and their estimation play a significant role in the Bayesian denoising technique, since the computational complexity of a subband-adaptive Bayesian denoising is $\mathcal{O}(\eta N_1 N_2)$, η being the number of parameters of the PDF and $N_1 \times N_2$ the size of the image. Subsequent to the work of Mallat [47], the GG PDF with the shape parameter s $(1 < s < 2)$ has been commonly used to model the image wavelet coefficients [12]-[15], [69], [126], [127], [140]. For example, the threshold factor in the BayesShrink method [13] and the parameter 'signal of interest' in the ProbShrink method [127] assume that the DWT coefficients in each subband are i.i.d. random variables having a GG distribution. Under the simplest prior GG with $s = 2$, and the i.i.d. assumption, both the MMSE and MAP estimators provide the same closed-form solution, the well known Wiener filter. For the purpose of denoising, the SMG prior function is used in [24], [25], [42], [70]-[72], [141], [142], wherein the parameters are estimated using the EM algorithm. In [92], it has been shown that four to eight state variables of the SMG PDF are sufficient for a good approximation of the prior function of the subband wavelet coefficients of images. In order to obtain a Bayes estimator for image denoising, Figueiredo and Nowak [130] have used Jeffreys' prior as a statistical model for the wavelet coefficients, since this prior is free of any parameter. In [21] and [143], the

$S\alpha S$ PDF with characteristic exponent α $(1 < \alpha < 2)$ has been proposed as a prior of the image coefficients. This PDF has no closed-form expression even though its characteristic function has one, and its parameter estimation is poor, specially in the presence of noise [92]. The BK PDF has been applied in the wavelet-domain MMSE estimator by Fadili and Boubchir [22]. The SNIG PDF for reducing the multiplicative noise of a log-transformed image (e.g., SAR and medical ultrasound images) has been used in a DWT-based MAP estimator in [23], [76]. At present, most of the PDFs used for modeling the image wavelet coefficients use a fixed number of parameters that are usually derived from the second and fourth order moments of the image coefficients. As a result, in most of the cases these PDFs become much more leptokurtic than the empirical one (see Chapter 3). Since the DWT coefficients are locally stationary [64], the parameters of some of the PDFs are refined with respect to the local spatial context within a subband [12], [13], [25], [42], [79], [106], [107], [118], [131]. In such a case, when the denoising technique is locally-adaptive within a subband, the computational complexity increases to $\mathcal{O}(\eta \mathtt{w}^2 N_1 N_2)$, $\mathtt{w} \times \mathtt{w}$ being the size of a squared-shape local window. To take into consideration the inter-subband dependency, in some cases PDFs have been extended to the bivariate [65]-[68] and multivariate forms [69], [70] or used in the HMT framework [71], [72]. In such a case, the computational load of a denoising technique increases exponentially with the number of subbands, and this is substantially higher than the corresponding locally-adaptive method.

In this chapter, novel DWT-based denoising methods are developed using the univariate MGH PDF that has been proposed for the DWT coefficients of images. This PDF is used in view of the fact that it matches the empirical density function of the wavelet coefficients very well, which would have an effect on the denoising performance. The proposed PDF is first applied as a prior function in the subband- and locally-adaptive Bayesian MMSE-based noise reduction methods. Next, a locally-adaptive Bayesian

MAP-based noise reduction method is developed. The subband- and locally-adaptive methods are considered for denoising, in view of the fact that they offer lower computational burden as compared to the method using inter-subband dependency. It is shown that the introduction of the more accurate prior in the proposed noise reduction techniques results in a performance better than that of the subband- and locally-adaptive wavelet-based image denoising methods that use conventional PDFs, in terms of the standard indices of noise reduction.

5.3 Denoising Algorithms

In this section, we assume that the image pixels are corrupted by AWGN with a known variance σ_{ε}^2. If σ_{ε}^2 is unknown, it may be estimated by applying the median-absolute-deviation method [120] in the highest frequency subband of the noisy wavelet coefficients. We propose both MMSE- and MAP-based denoising, wherein the Bayes-risk estimator uses the proposed MGH PDF. In a Bayesian framework, g_{ij}, f_{ij}, and ε_{ij} in (5.1) are considered as samples of the independent random variables $\boldsymbol{g}$, $\boldsymbol{f}$, and $\boldsymbol{\varepsilon}$, respectively. The signal component $\boldsymbol{f}$ is modelled as the i.i.d. MGH random variable. The random variable of the noise component $\boldsymbol{\varepsilon}$ is modelled as a zero-mean Gaussian random variable with a subband-adaptive variance $\tilde{\sigma}_{\varepsilon}^2$, which depends on σ_{ε}^2, $\boldsymbol{W}$, and the arithmetic operations of the corresponding wavelet-domain (see Chapter 4).

5.3.1 MMSE Estimator

The Bayes-risk estimator for the denoised wavelet coefficient $\hat{f}_{ij}$ using the MMSE criterion[1] is the conditional mean and is given by

$$\hat{f}(g) = \int p_{f|g}(f|g) f \mathrm{d}f \tag{5.5}$$

[1]MMSE criterion minimizes the MSE cost function

where $p_{\boldsymbol{f}|\boldsymbol{g}}(\boldsymbol{f}|\boldsymbol{g})$ is the posterior density function conditioned on the wavelet coefficients under consideration. According to the Bayes' rule, the prior function $p_{\boldsymbol{f}}(\boldsymbol{f})$ of the image wavelet coefficients can be used to compute the posterior density function as

$$\begin{aligned} p_{\boldsymbol{f}|\boldsymbol{g}}(\boldsymbol{f}|\boldsymbol{g}) &= \frac{p_{\boldsymbol{g}|\boldsymbol{f}}(\boldsymbol{g}|\boldsymbol{f})p_{\boldsymbol{f}}(\boldsymbol{f})}{p_{\boldsymbol{g}}(\boldsymbol{g})} \\ &= \frac{p_{\boldsymbol{g}|\boldsymbol{f}}(\boldsymbol{g}|\boldsymbol{f})p_{\boldsymbol{f}}(\boldsymbol{f})}{\int p_{\boldsymbol{g}|\boldsymbol{f}}(\boldsymbol{g}|\boldsymbol{f})p_{\boldsymbol{f}}(\boldsymbol{f})\mathrm{d}\boldsymbol{f}} \\ &= \frac{p_{\boldsymbol{\epsilon}}(\boldsymbol{g}-\boldsymbol{f})p_{\boldsymbol{f}}(\boldsymbol{f})}{\int p_{\boldsymbol{\epsilon}}(\boldsymbol{g}-\boldsymbol{f})p_{\boldsymbol{f}}(\boldsymbol{f})\mathrm{d}\boldsymbol{f}} \end{aligned} \tag{5.6}$$

where $p_{\boldsymbol{\epsilon}}(\cdot)$ is the density function of the wavelet coefficients corresponding to the noise component. It can be seen from (5.5) and (5.6) that the MMSE-based shrinkage function $\hat{\boldsymbol{f}}(\boldsymbol{g})$ is independent of the weight parameter γ of the MGH PDF that has been estimated in Chapter 3. An analytical expression of the proposed MMSE-based shrinkage function can be obtained as (see Appendix B)

$$\hat{\boldsymbol{f}}(\boldsymbol{g}) = \frac{\Delta_{\text{num}}}{\Delta_{\text{den}}} \tag{5.7}$$

where

$$\begin{aligned} \Delta_{\text{num}} = \sum_{r=0}^{\varsigma/2} \beta_{2r} \sum_{u=0}^{r} &(2r-2u+1)\varphi(r,u)(\lambda\tilde{\sigma}_{\boldsymbol{\epsilon}})^{2r-2u+2} \\ &\cdot \left[D_{-(2r-2u+2)}\left(-\frac{\lambda}{\tilde{\sigma}_{\boldsymbol{\epsilon}}}\boldsymbol{g}\right) - D_{-(2r-2u+2)}\left(\frac{\lambda}{\tilde{\sigma}_{\boldsymbol{\epsilon}}}\boldsymbol{g}\right) \right] \end{aligned} \tag{5.8}$$

$$\begin{aligned} \Delta_{\text{den}} = \sum_{r=0}^{\varsigma/2} \beta_{2r} \sum_{u=0}^{r} &\varphi(r,u)(\lambda\tilde{\sigma}_{\boldsymbol{\epsilon}})^{2r-2u+1} \\ &\cdot \left[D_{-(2r-2u+1)}\left(-\frac{\lambda}{\tilde{\sigma}_{\boldsymbol{\epsilon}}}\boldsymbol{g}\right) + D_{-(2r-2u+1)}\left(\frac{\lambda}{\tilde{\sigma}_{\boldsymbol{\epsilon}}}\boldsymbol{g}\right) \right] \end{aligned} \tag{5.9}$$

$\varphi(r,u) = \frac{(-2)^u(2r)!}{u!\sigma_{\boldsymbol{f}}^{2r-2u}}$, $\lambda = \left(\frac{\sigma_{\boldsymbol{f}}^2}{\sigma_{\boldsymbol{f}}^2+\tilde{\sigma}_{\boldsymbol{\epsilon}}^2}\right)^{\frac{1}{2}}$, and $D_r(\cdot)$ is the parabolic cylinder function [144]. It is to be noted that for the subband-adaptive MMSE estimator when $|\boldsymbol{g}| > f_0$, $\varsigma = 4$ and $\beta_4 = \frac{K_m-3}{4!}$. The shrinkage function given by (5.7) may be directly used for estimating each of the noisy wavelet coefficients. The computational complexity of this

function is $\mathcal{O}\left(4C_D N_1 N_2 \sum_{r=0}^{\zeta/2} \frac{(r+1)(r+2)}{2}\right) = \mathcal{O}\left(C_D(\zeta+2)(\zeta+4)(\zeta+6)N_1 N_2/12\right)$, where C_D is the computational complexity of $D_r(\cdot)$. Since the computational load increases exponentially with ζ, an implementation of such a function may not be practical in view of the excessive computational effort involved. Hence, we represent the shrinkage function in terms of a linear convolution as

$$\hat{\boldsymbol{f}}(\boldsymbol{g}) = \frac{p_{\boldsymbol{\epsilon}}(\boldsymbol{g}) * (\boldsymbol{g}p_{\boldsymbol{f}}(\boldsymbol{g}))}{p_{\boldsymbol{\epsilon}}(\boldsymbol{g}) * p_{\boldsymbol{f}}(\boldsymbol{g})} \tag{5.10}$$

Taking the computational advantage of the DFT, the shrinkage function can be expressed as

$$\hat{\boldsymbol{f}}(\boldsymbol{g}) = \frac{\mathcal{F}^{-1}\left[\mathcal{F}[p_{\boldsymbol{\epsilon}}(\boldsymbol{g})] \cdot \mathcal{F}[\boldsymbol{g}p_{\boldsymbol{f}}(\boldsymbol{g})]\right]}{\mathcal{F}^{-1}\left[\mathcal{F}[p_{\boldsymbol{\epsilon}}(\boldsymbol{g})] \cdot \mathcal{F}[p_{\boldsymbol{f}}(\boldsymbol{g})]\right]} \tag{5.11}$$

where $\mathcal{F}$ is the forward DFT and $\mathcal{F}^{-1}$ is the inverse DFT. Using a fast implementation of the DFT, the complexity of computing the shrinkage function can be reduced to $\mathcal{O}(5N_b \log N_b)$. The function given in (5.11) provides the estimates of the noisy wavelet coefficients at the discrete points $\boldsymbol{f}_d$, where the probabilities of the MGH PDF are estimated. Therefore, an interpolation function is required to estimate each of the noisy wavelet coefficients. In our case, we choose the cubic Hermite spline [145] as an interpolation function so that the estimated Hermite polynomials for the MGH PDF can also be used for interpolation. This spline function requires the estimation of four parameters and has a computational complexity of $\mathcal{O}(4(N_b-1)N_1N_2)$. Since $N_b \ll N_1N_2$ and $\zeta \ll N_b$, the computational complexity of the shrinkage function given by (5.11) for subband-adaptive denoising is $\mathcal{O}(4N_bN_1N_2)$. From simulation results, we have observed that the denoising performances using the two shrinkage functions, given by (5.7) and (5.11), are essentially the same. Hence, we prefer to use (5.11) for the proposed Bayesian denoising in view of its lower computational complexity.

The proposed MMSE estimation requires that the parameters of the proposed prior

function be estimated as accurately as possible from the noisy observations. Since the parameters of the MGH prior function are expressed in terms of $M_{n\boldsymbol{f}}$, the estimation of the n-th order moment of the signal coefficients from the noisy observations is sufficient for such a parameter estimation. The noise is independent of the signal and additive in nature, since it is AWGN. Therefore, the n-th order moment of the signal coefficients is estimated as $\hat{M}_{n\boldsymbol{f}}$ by utilizing, $M_{n\boldsymbol{g}}$ and $M_{n\boldsymbol{\varepsilon}}$, which are, respectively, the n-th order moments of the noisy signal and noise coefficients. Since the noise is zero-mean Gaussian distributed, the odd-order moments of the noise coefficients are zero and the even-order moments can be expressed in terms of the noise standard deviation $\tilde{\sigma}_{\boldsymbol{\varepsilon}}$. In general, the n-th order moment of the noise coefficients is [89]

$$M_{n\boldsymbol{\varepsilon}} = \begin{cases} 0 & \text{for } n \text{ odd} \\ \tilde{\sigma}_{\boldsymbol{\varepsilon}}^{n} \prod_{u=1}^{\frac{n}{2}} (2u-1) & \text{for } n \text{ even} \end{cases} \tag{5.12}$$

Therefore, the n-th order moment of the signal coefficients is evaluated as

$$\hat{M}_{n\boldsymbol{f}} = \begin{cases} M_{n\boldsymbol{g}} & \text{for } n \text{ odd} \\ \max\left(M_{n\boldsymbol{g}} - \sum_{u=1}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{2u} \hat{M}_{(n-2u)\boldsymbol{f}} M_{(2u)\boldsymbol{\varepsilon}} - M_{n\boldsymbol{\varepsilon}}, 0 \right) & \text{for } n \text{ even} \end{cases} \tag{5.13}$$

The parameter estimation of the MGH prior function is relatively simple, since these $\hat{M}_{n\boldsymbol{f}}$ can be substituted in (3.3) to obtain $\hat{\beta}_{2r}$. On the other hand, the parameter estimation of traditional prior functions that have fewer number of parameters are often not so simple. For example, the shape and width parameters of the GG prior function are estimated numerically that involves iteration. In practice, the convergence of the iterative procedure depends on the initial choice of the parameter. Fig. 5.1 shows an example of the non-linear shrinkage function $\hat{\boldsymbol{f}}(\boldsymbol{g})$ obtained from the proposed denoising technique that uses three different values of ζ in the MGH prior function. It can be observed from this figure that the amount of shrinkage can be controlled by changing ζ. This feature of being able to vary the amount of shrinkage

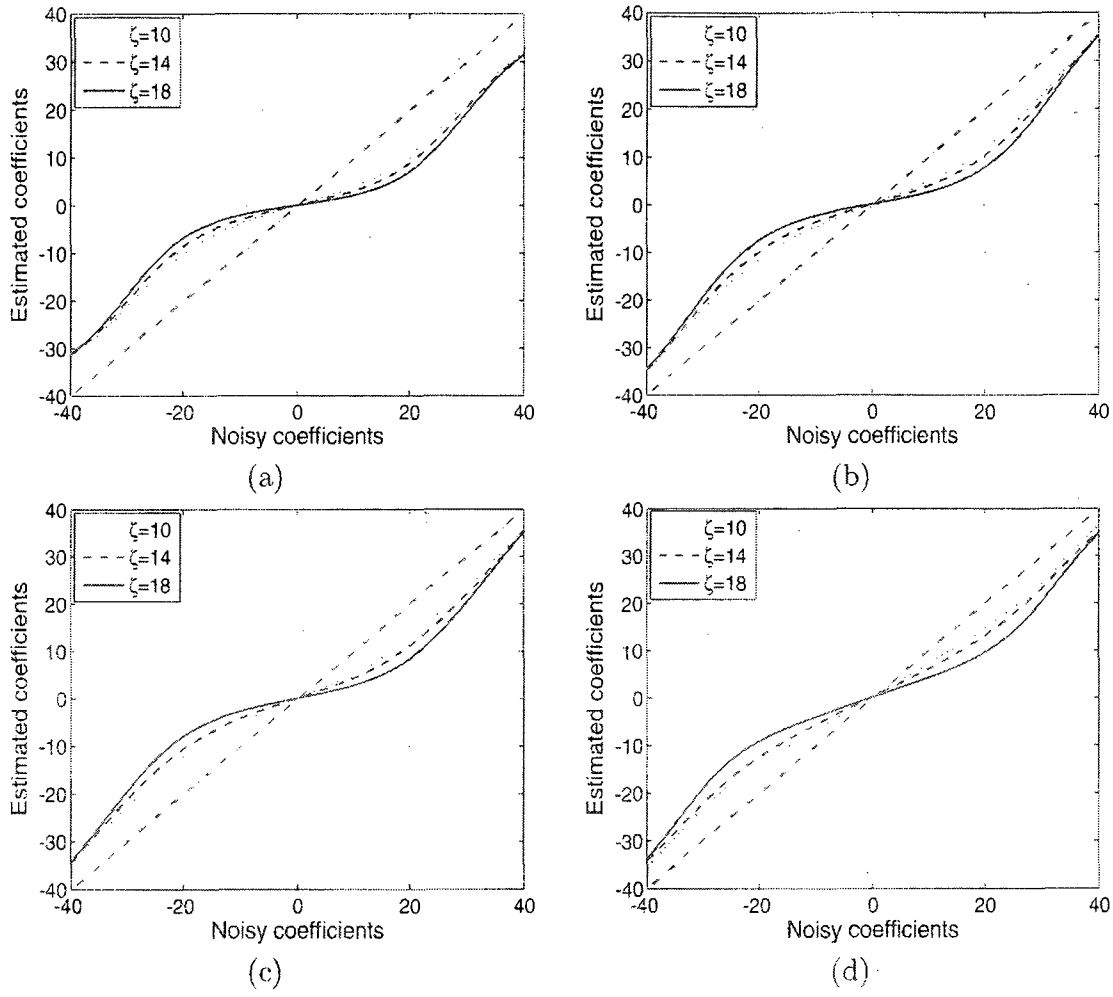


Figure 5.1: Shrinkage function obtained from the modified Gauss-Hermite probability density function for the LH_2 subband DWT coefficients as a function of ζ with $\sigma_\epsilon = 10$. The test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, and (d) *Bridge*.

may be useful, in particular, in producing an output image that meets the requirement of a desired level of denoising performance from the user's perspective. Since the wavelet coefficients of images are locally stationary, an improved performance of the Bayesian denoising method can be obtained by considering the proposed prior function to be locally-adaptive during the MMSE estimation. As explained in Chapter 3, the higher-order moments of data samples are not so effective for modeling the local neighboring coefficients as compared to that for the subband coefficients. It has been shown that the symmetric GC PDF performs better than the Gaussian, GG, and SNIG PDFs for modeling the local neighboring wavelet coefficients of images. A better quality MMSE estimator for image denoising can, therefore, be obtained by using the symmetric GC PDF as the prior function in the proposed MMSE-based shrinkage function. Extensive simulations reveal that the difference in denoising performance in terms of the MSEs of the MMSE estimators designed using the MGH PDF with $\zeta = 6$ and $\zeta = 4$ is no more than 0.25%. Hence, in order to save the computational effort, MGH PDF with $\zeta = 4$, i.e., the GC PDF, is chosen as the prior function for the case of locally-adaptive denoising. In such a case, the number of parameters of the MGH prior function is the same as that of the conventional prior functions, such as the GG, BK, and SNIG. The computational complexity of the locally-adaptive version of the proposed MMSE-based denoising method is $\mathcal{O}(4\mathtt{w}^2 N_{\mathtt{w}} N_1 N_2)$, where $N_{\mathtt{w}}$ is the average number of discrete points in a window obtained from the selected bin.

5.3.2 MAP Estimator

The Bayes-risk estimator for the denoised wavelet coefficient $\hat{f}_{ij}$ using the MAP criterion[2] is the conditional mode of the posterior density function and is given by [66]

$$\hat{f}(\boldsymbol{g}) = \arg\max_{\boldsymbol{f}} p_{\boldsymbol{f}|\boldsymbol{g}}(\boldsymbol{f}|\boldsymbol{g}) \tag{5.14}$$

[2]MAP criterion minimizes the delta cost function

Using the definition of the posterior density function given in (5.6), the denoised wavelet coefficient using the MAP criterion can be written as

$$\begin{aligned}\hat{f}(\boldsymbol{g}) &= \arg\max_{\boldsymbol{f}} \left[p_{\boldsymbol{\varepsilon}}(\boldsymbol{g}-\boldsymbol{f}) \cdot p_{\boldsymbol{f}}(\boldsymbol{f})\right] \\ &= \arg\max_{\boldsymbol{f}} \left[\ln p_{\boldsymbol{\varepsilon}}(\boldsymbol{g}-\boldsymbol{f}) + \ln p_{\boldsymbol{f}}(\boldsymbol{f})\right] \end{aligned} \tag{5.15}$$

Except for a few number of prior functions, e.g., Gaussian and Laplacian, the MAP estimation rarely provides closed-form solutions for a relatively complex prior. In MAP estimation, it is desirable that the prior function be continuously differentiable, since the exact estimation of $\hat{f}_{ij}$ at the non-differentiable points of $p_{\boldsymbol{f}}(\boldsymbol{f})$ is not possible. In this section, we develop the MAP-based Bayesian denoising method that is locally-adaptive because of the following two major reasons:

- The prior function $p_{\boldsymbol{f}}(\boldsymbol{f})$ in the locally-adaptive case, defined by the symmetric GC PDF, is continuously differentiable.
- The locally-adaptive estimators provide a better denoising performance as compared to the subband-adaptive ones, even though the computational load of the former is much higher than that of the latter.

Similar to most of the other prior functions, it is not possible to obtain an exact closed-form expression for the MAP estimator using the proposed GC prior. Hence, we have used the approach proposed by Hyvarinen [146] to obtain an approximate solution of (5.15), which can be found as

$$\hat{f}_{ij}(\boldsymbol{g}) = \text{sign}(g_{ij}) \cdot \max\left(|g_{ij}| - \tilde{\sigma}_{\varepsilon}^2 |\Omega(\boldsymbol{g})|, 0\right) \tag{5.16}$$

where $\Omega(\boldsymbol{f}) = -\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{f}}\left[\ln p_{\boldsymbol{f}}(\boldsymbol{f})\right]$ is called the score function of $\boldsymbol{f}$. Using the identity $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{f}}\left[H_r(\boldsymbol{f})\right] = r H_{r-1}(\boldsymbol{f})$, the score function of the symmetric GC PDF can be obtained as

$$\Omega(\boldsymbol{f}) = \frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}^2} - \frac{\frac{\kappa_{\boldsymbol{f}}-3}{3!} H_3\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)}{1 + \frac{\kappa_{\boldsymbol{f}}-3}{4!} H_4\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)} \cdot \frac{1}{\sigma_{\boldsymbol{f}}} \tag{5.17}$$

The evaluation of $\Omega(\boldsymbol{f})$ requires that the variance and kurtosis of the noise-free wavelet coefficients be calculated from the noisy condition. In other words, it is necessary to estimate the second and fourth order moments of the noise-free image coefficients from the noisy coefficients. Since the image is corrupted by AWGN, the second and fourth order moments can be estimated from (5.13) as

$$\begin{aligned} \hat{M}_{2\boldsymbol{f}} &= \max(\hat{M}_{2\boldsymbol{g}} - \tilde{\sigma}_{\boldsymbol{\varepsilon}}^2, 0) \\ \hat{M}_{4\boldsymbol{f}} &= \max(\hat{M}_{4\boldsymbol{g}} - 6\hat{M}_{2\boldsymbol{f}}\sigma_{\boldsymbol{\varepsilon}}^2 - 3\tilde{\sigma}_{\boldsymbol{\varepsilon}}^4, 0) \end{aligned} \tag{5.18}$$

It is to be noted that the computational complexities of estimation of $M_{n\boldsymbol{f}}$ and H_r for the proposed estimator are $\mathcal{O}(\mathtt{w}^2 N_1 N_2)$ and $\mathcal{O}(N_1 N_2)$, respectively. Since $\mathtt{w}^2 \gg 1$, the computational complexity of the proposed MAP estimator is $\mathcal{O}(2\mathtt{w}^2 N_1 N_2)$.

5.4 Experimental Results

Extensive experimentations have been carried out in order to compare the performance of the proposed denoising methods with that of the other methods. We give in this section results concerning five 512×512 grayscale images, namely, *Lena*, *Barbara*, *Boat*, *Bridge*, and *Baboon* and two types of wavelet representations, namely, the decimated DWT and redundant DT-CWT. The first three test images are obtained from the same sources as mentioned in [42] and the last two from [85]. The DWT-based experiments use the orthogonal wavelet filter Symmlet-8 (S-8), whereas the DT-CWT-based ones use the $(11, 17)$-tap biorthogonal filters at level $\ell = 1$, and 6-tap Q-shift orthogonal filters at $\ell \geq 2$, as have been proposed in [53]. The proposed algorithms are applied to the 4-level decomposition of the DWT and the 3-level decomposition of the DT-CWT, since any further decomposition beyond these levels does not produce a significant increase in the denoising performance. The performance of the Bayesian denoising method is tested using two sets of experiments. In the first set of experiments, the MGH prior function is used for the subband-adaptive

denoising. In the second set of experiments, the symmetric GC prior function, i.e., MGH prior function with $\zeta = 4$, is used for locally-adaptive denoising.

5.4.1 Subband-Adaptive Bayesian Denoising

Seven different DWT-based Bayesian denoising methods, which are subband-adaptive in nature, have been considered in this set of experiments. These are the classical Wiener filter, SureShrink [120], BayesShrink [13], EBayesThresh [133], ProbShrink [127], BKShrink [22] and proposed method. As pointed out in Section 5.2, these methods enjoy a computational complexity that is lower than that of the corresponding locally-adaptive one by a factor of $\mathtt{w}^2$; and that the computational complexity of those methods that employ inter-subband dependency is even higher than those of the corresponding locally-adaptive ones. To the best of our knowledge the BKShrink and ProbShrink methods provide the best results amongst the subband-adaptive denoising methods. The denoising performance of the various methods are evaluated using two indices, namely, the PSNR and SSIM [125]. In this context, it is to be noted that the authors in [125] have claimed that SSIM can assess the similarity between the original and processed images better than the PSNR. Denoising performance of the proposed method is tested by using the two shrinkage functions, given by (5.7) and (5.11) in Section 5.3. The evaluation of the parabolic cylinder function for the analytic expression in (5.7) is performed by the same source code as mentioned in [22]. It is to be noted that the processing time of (5.11) is very fast as compared to that of (5.7). For example, in a 3 GHz and 512 MB machine, the processing time of (5.11) using the MATLAB code is only 6.6 seconds for $\zeta = 14$ and the time is virtually the same for values of ζ. On the other hand, the processing time of (5.7) is 1 hour and 24 minutes for $\zeta = 4$ and the time increases exponentially with ζ. It is also observed that for a given value of ζ, the denoising performances of both (5.7) and (5.11) are essentially the same. Hence, we provide the results of the proposed method that uses

the shrinkage function given by (5.11). From our experiments, we have observed that $\zeta = 14$ provides the best performance in terms of the PSNR, whereas $\zeta = 16$ provides the best performance in terms of the SSIM, for the proposed method. These values of ζ are consistent with the values of ζ mentioned in Chapter 3 as providing the best statistical modeling of the image wavelet coefficients by the MGH PDF. The output PSNR as a function of the input PSNR obtained from various DWT-based denoising methods are shown in Figs. 5.2, 5.3, 5.4, 5.5, and 5.6, respectively, for the test images *Lena*, *Barbara*, *Boat*, *Bridge*, and *Baboon*. From these figures, it can be observed that the proposed denoising method with $\zeta = 14$ provides the highest output PSNR consistently as compared to the other methods. According to the output PSNR values, the improvement of the denoising performance provided by the proposed method is substantial, when the input PSNR is low, i.e., the noise strength is significant. Fig. 5.7 shows a comparison between the output PSNR values obtained from the SureShrink [120], BayesShrink [13], EBayesThresh [133], ProbShrink [127] methods employing the SWT and proposed methods using the Cartesian representation of the DT-CWT, when the images are corrupted with $\sigma_\varepsilon = 20$. From this figure, it can also be seen that the output PSNR values of the proposed method are consistently higher than that of the other methods while redundant wavelet transforms are used. Table 5.1 shows the SSIM values obtained from the various DWT-based denoising methods for three different noise standard deviations. It can be observed from the table that, except for two values of σ_ε in the case of the *Barbara* image, the proposed method with $\zeta = 16$ provides the best SSIM compared to that provided by the other denoising methods. It should be noted that the PSNR values for $\zeta = 16$ are about the same as those for $\zeta = 14$, the difference being not more than 0.1 dB, and that these values are at least as good as those provided by the other methods. Fig. 5.8 shows the original cropped image *Lena*, its noisy version with $\sigma_\varepsilon = 25$, and the DWT-based

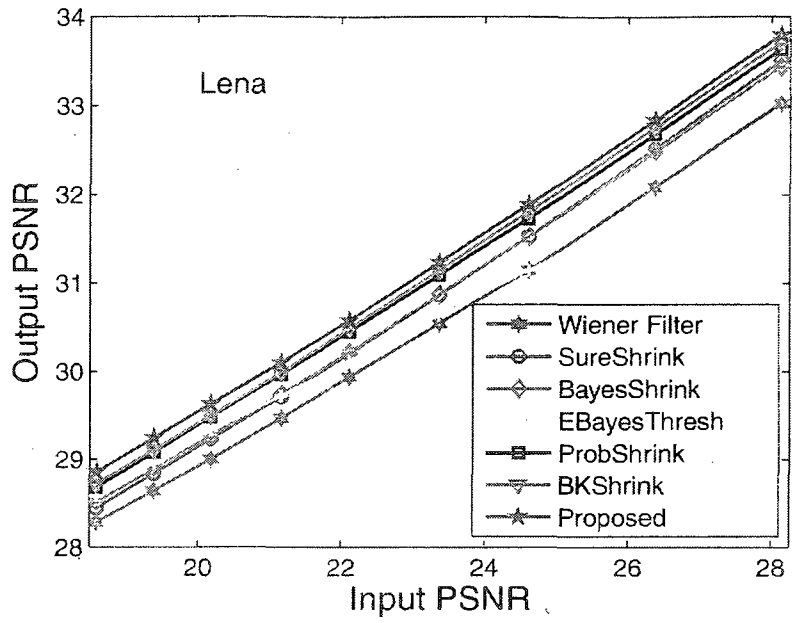


Figure 5.2: Output PSNR as a function of the input PSNR for different subband-adaptive DWT-based Bayesian denoising algorithms using the test image *Lena* ($\zeta = 14$ for the proposed method).

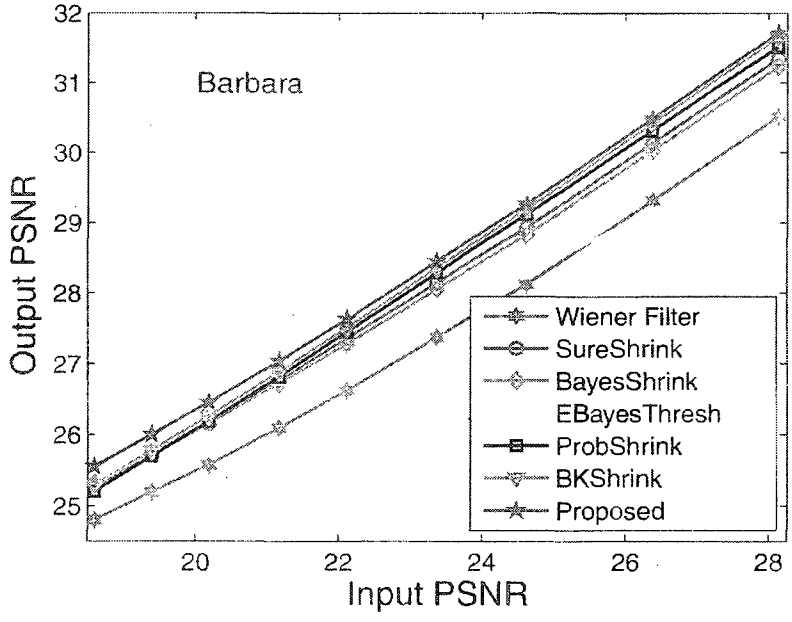


Figure 5.3: Output PSNR as a function of the input PSNR for different subband-adaptive DWT-based Bayesian denoising algorithms using the test image *Barbara* ($\zeta = 14$ for the proposed method).

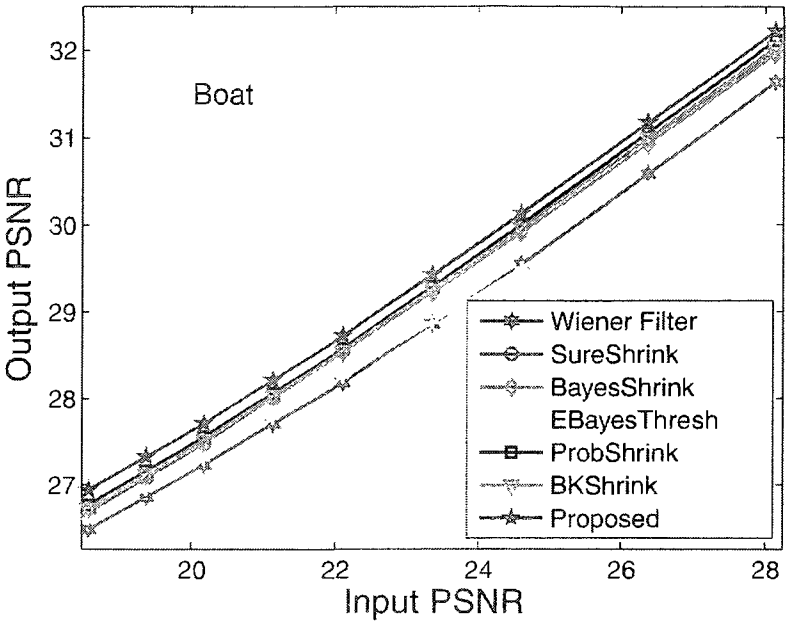


Figure 5.4: Output PSNR as a function of the input PSNR for different subband-adaptive DWT-based Bayesian denoising algorithms using the test image *Boat* ($\zeta = 14$ for the proposed method).

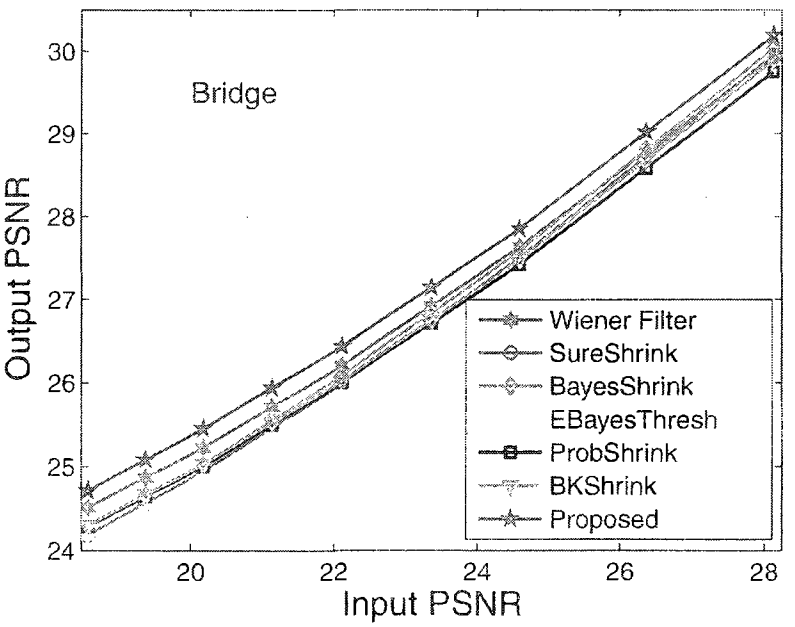


Figure 5.5: Output PSNR as a function of the input PSNR for different subband-adaptive DWT-based Bayesian denoising algorithms using the test image *Bridge* ($\zeta = 14$ for the proposed method).

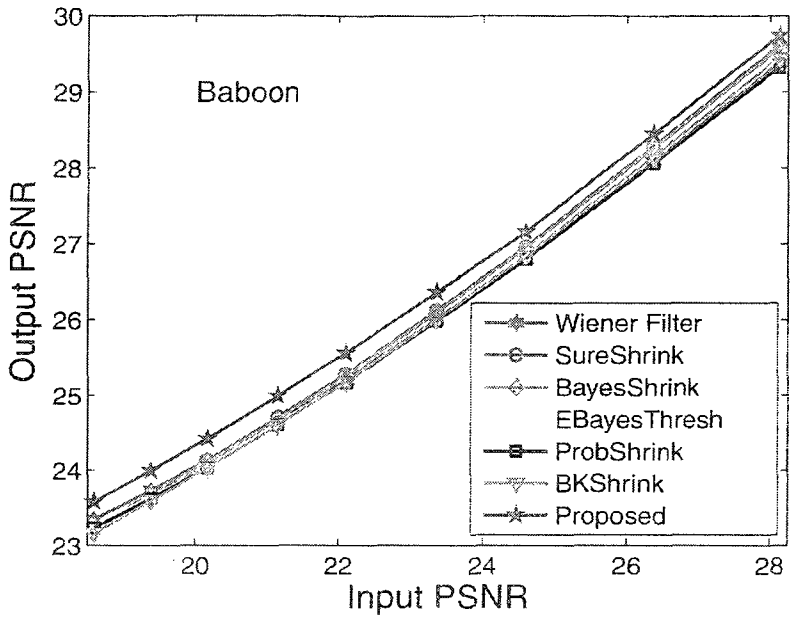


Figure 5.6: Output PSNR as a function of the input PSNR for different subband-adaptive DWT-based Bayesian denoising algorithms using the test image *Baboon* ($\zeta = 14$ for the proposed method).

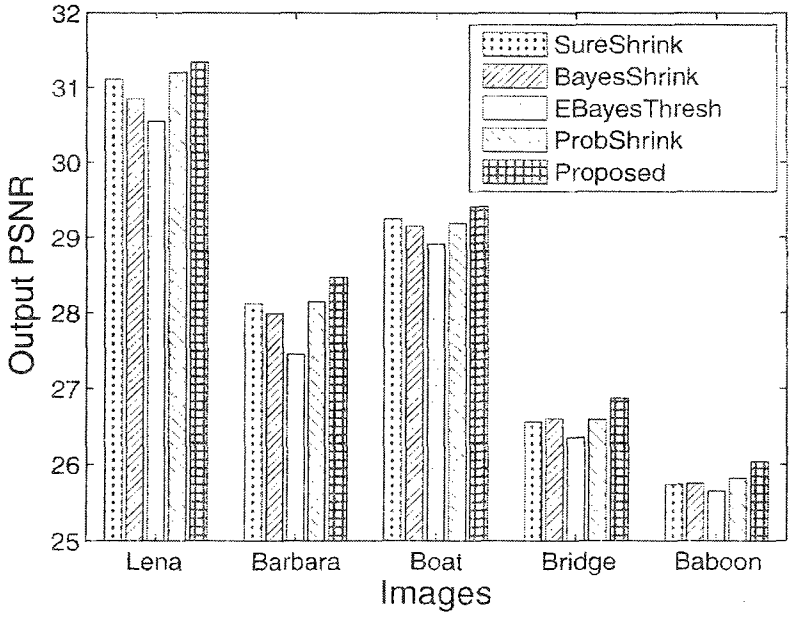


Figure 5.7: Output PSNR of subband-adaptive Bayesian denoising algorithms using redundant DWT, wherein test images are corrupted with $\sigma_\varepsilon = 20$ ($\zeta = 14$ for the proposed method).

Table 5.1: Values of the index SSIM for various subband-adaptive Bayesian denoising methods using the DWT.

	Noisy	Sure Shrink in [120]	Bayes Shrink in [13]	EBayes Thresh in [133]	Prob Shrink in [127]	BK Shrink in [22]	Proposed
Lena							
$\sigma_\epsilon = 10$	0.8733	0.9442	0.9379	0.9310	0.9505	0.9523	**0.9528**
$\sigma_\epsilon = 20$	0.6788	0.8836	0.8786	0.8861	0.8992	0.9015	**0.9039**
$\sigma_\epsilon = 30$	0.5326	0.8306	0.8318	0.8440	0.8530	0.8586	**0.8617**
Barbara							
$\sigma_\epsilon = 10$	0.9122	0.9501	0.9437	0.9320	0.9550	0.9562	**0.9565**
$\sigma_\epsilon = 20$	0.7645	0.8804	0.8726	0.8716	0.8916	**0.8943**	0.8915
$\sigma_\epsilon = 30$	0.6303	0.8134	0.8119	0.8122	0.8242	**0.8311**	0.8245
Boat							
$\sigma_\epsilon = 10$	0.9002	0.9420	0.9394	0.9309	0.9465	0.9478	**0.9486**
$\sigma_\epsilon = 20$	0.7325	0.8665	0.8650	0.8673	0.8793	0.8805	**0.8833**
$\sigma_\epsilon = 30$	0.5917	0.8034	0.8029	0.8099	0.8192	0.8226	**0.8242**
Bridge							
$\sigma_\epsilon = 10$	0.9544	0.9622	0.9625	0.9490	0.9593	0.9611	**0.9637**
$\sigma_\epsilon = 20$	0.8501	0.8933	0.8953	0.8708	0.8896	0.8894	**0.8996**
$\sigma_\epsilon = 30$	0.7355	0.8228	0.8275	0.7988	0.8193	0.8166	**0.8369**
Baboon							
$\sigma_\epsilon = 10$	0.9450	0.9546	0.9544	0.9342	0.9527	0.9555	**0.9566**
$\sigma_\epsilon = 20$	0.8301	0.8753	0.8763	0.8493	0.8757	0.8765	**0.8817**
$\sigma_\epsilon = 30$	0.7121	0.7966	0.7990	0.7613	0.7978	0.7954	**0.8060**

denoised versions provided by the BayesShrink method [13], ProbShrink method [127], BKShrink method [22], and proposed method with $\zeta = 16$. From this figure, it can be observed that the proposed method can reduce the noise significantly with the least amount of artifacts (e.g., in the eye and the edge of the hat) as compared to the other competing methods.

5.4.2 Locally-Adaptive Bayesian Denoising

Seven different DWT-based locally-adaptive denoising methods have been considered in this set of experiments. These are the BivariateShrink [66], locally-adaptive

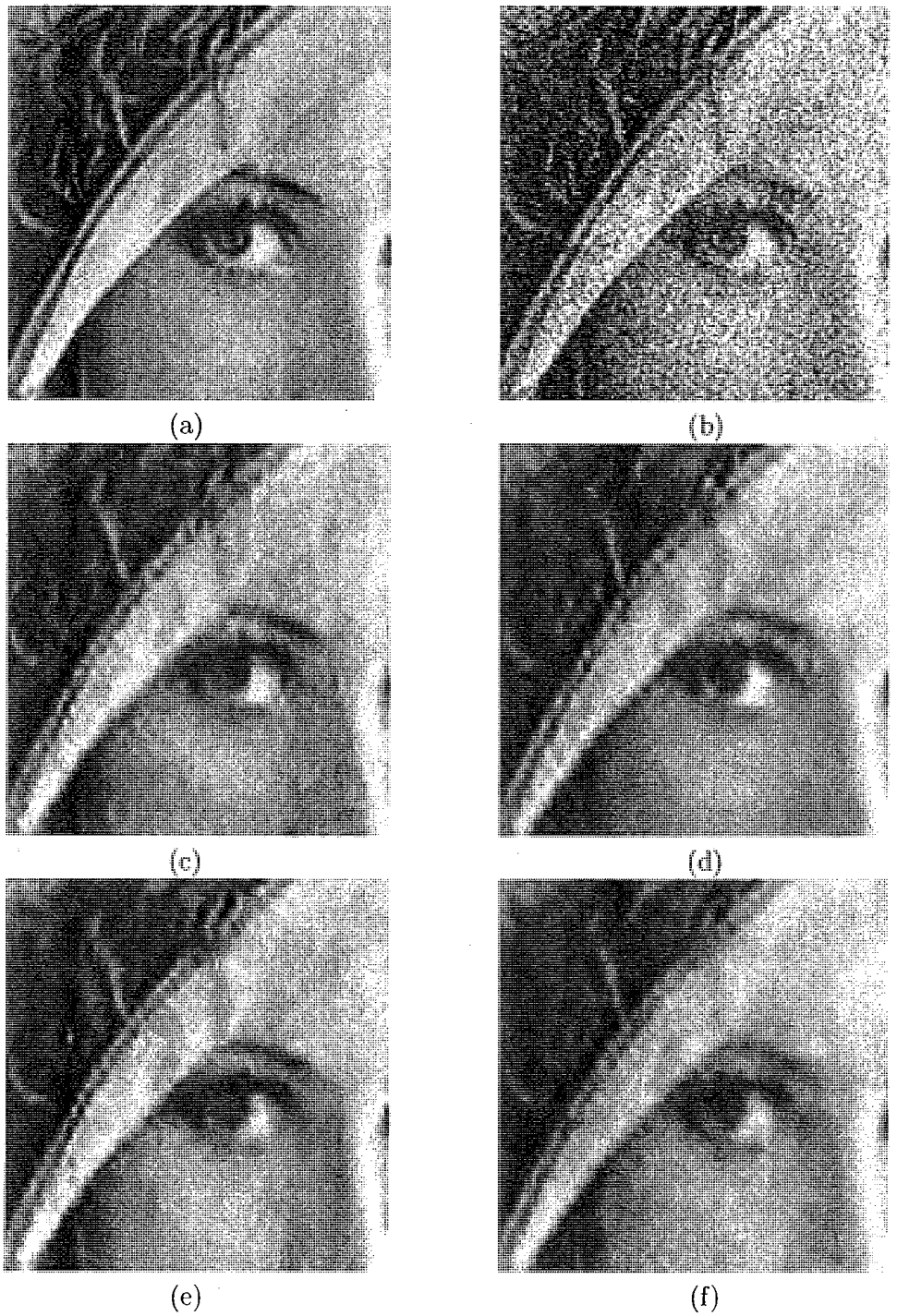


Figure 5.8: Comparison of the denoised images obtained from the different subband-adaptive Bayesian DWT-based denoising algorithms on *Lena* with $\sigma_\epsilon = 25$: (a) original image, (b) noisy image (PSNR=20.18 dB, SSIM=0.6053), and denoised images using (c) BayesShrink method (PSNR=29.27 dB, SSIM=0.8552), (d) ProbShrink method (PSNR=29.52 dB, SSIM=0.8771), (e) BKShrink method (PSNR=29.50 dB, SSIM=0.8801), and (f) proposed method with $\zeta = 16$ (PSNR=29.58 dB, SSIM=0.8823).

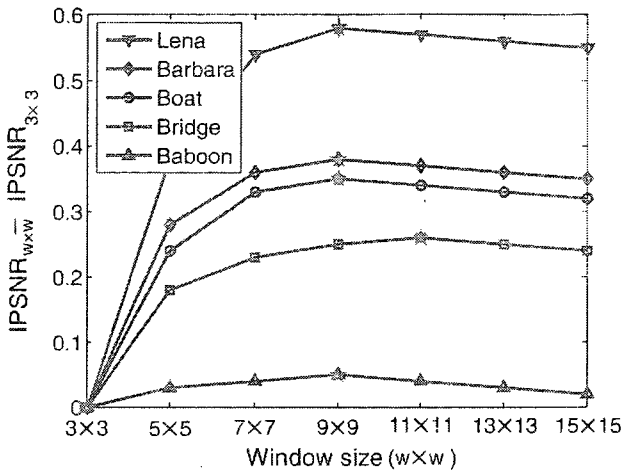


Figure 5.9: The IPSNR values in dB as a function of the window size relative to the IPSNR values corresponding to a 3×3 window.

window-based MAP (LAWMAP) [106], NeighShrink [131], ProbShrink [127], SMG method [42] and proposed MMSE- and MAP-based methods. In the first method, both the inter- and intra-subband dependencies are considered. The second, third, and fourth methods inherently use only the intra-subband dependency. In the case of the fifth, we provide the results obtained from the intra-subband dependency only in order to make a fair comparison with the proposed methods. The parameters of the symmetric GC prior function for the proposed denoising methods are refined locally using different sizes of square-shaped neighborhood regions. Fig. 5.9 shows the improvement in the PSNR (IPSNR) values[3] as a function of the window size with a reference of 3×3 window, i.e., $\text{IPSNR}_{\mathtt{w}\times\mathtt{w}} - \text{IPSNR}_{3\times 3}$ as a function of $\mathtt{w} \times \mathtt{w}$, for the MMSE-based denoising of different test images being corrupted with a noise having $\sigma_\epsilon = 20$. From this figure, it can be observed that as the window size is increased from 3×3, the IPSNR value increases first and then drops slightly when the window exceeds a certain size. This is due to the fact that a sufficient number of samples are required for improved parameter estimation of the prior function, but, if the neighbor-

[3]IPSNR is the difference between output PSNR and input PSNR.

hood region becomes relatively large, the estimation process loses its local contextual information. Similar results are obtained for other noise standard deviations as well as for the MAP-based denoising. The experiments reveal that a 9×9 window provides the highest IPSNR values in most cases. Considering this fact, we report the results provided by our proposed denoising methods using a 9×9 window.

Table 5.2 shows the output PSNR values obtained from the various locally-adaptive denoising methods for the test images considered here. From the tabular results, it can be observed that among all the DWT-based denoising methods, the proposed methods provide the highest output PSNR values as compared to the other methods. Between the two proposed denoising methods, viz., MAP-based and MMSE-based, the output PSNR values of the latter are higher than that of the former, the difference being not more than 0.15 dB. It is to be noted that the proposed MAP-based denoising method is computationally very fast because of its simple thresholding function. For example, in a 3 GHz and 512 MB machine, the processing time of this method is only 1.6 seconds, while that of the SMG method [25] and ProbShrink method [127] are 12.7 and 14.2 seconds, respectively. On the other hand, the proposed locally-adaptive MMSE-based method is relatively expensive in terms of computation, since it requires a processing time of 15.12 minutes for the above mentioned example. However, this computational cost is acceptable, since the MMSE-based method consistently provides the highest output PSNR. Fig. 5.10 shows the original cropped image *Barbara*, its noisy version with $\sigma_\varepsilon = 20$, and the DWT-based denoised versions provided by four different denoising methods, namely, the ProbShrink method [127], SMG method [42], and proposed MAP- and MMSE-based methods. This figure also reveals that the proposed methods can reduce noise significantly with the least amount of artifacts (e.g., in the eyes and scarf of *Barbara*) as compared to the others.

Table 5.2: Output PSNR, $20\log_{10}(255/\sigma_e)$ in dB, where σ_e is the error standard deviation, for different DWT-based locally-adaptive denoising methods.

Denoising algorithms	Noise standard deviation σ_ϵ				
	10	15	20	25	30
	Lena				
BivariateShrink [66]	34.47	32.63	31.30	30.30	29.49
LAWMAP [106]	34.35	32.40	31.06	30.02	29.22
NeighShrink [131]	34.46	32.52	31.04	29.88	28.88
ProbShrink [127]	34.30	32.41	31.05	30.02	29.25
SMG [42]	34.57	32.72	31.41	30.36	29.57
Proposed method (MAP)	34.67	32.84	31.55	30.52	29.75
Proposed method (MMSE)	**34.81**	**32.95**	**31.63**	**30.61**	**29.78**
	Barbara				
BivariateShrink [66]	32.69	30.35	28.75	27.58	26.63
LAWMAP [106]	32.58	30.27	28.73	27.60	26.72
NeighShrink [131]	32.86	30.44	28.78	27.53	26.54
ProbShrink [127]	32.51	30.10	28.50	27.31	26.38
SMG [42]	32.84	30.44	28.81	27.68	26.72
Proposed method (MAP)	33.00	30.62	29.02	**27.90**	26.96
Proposed method (MMSE)	**33.04**	**30.69**	**29.09**	**27.90**	**26.97**
	Boat				
BivariateShrink [66]	32.48	30.61	29.28	28.24	27.40
LAWMAP [106]	32.36	30.48	29.09	28.07	27.16
NeighShrink [131]	32.72	30.64	29.15	28.05	27.08
ProbShrink [127]	32.53	30.57	29.17	28.10	27.27
SMG [42]	32.78	30.82	29.44	28.38	27.52
Proposed method (MAP)	32.78	30.83	29.44	28.40	27.56
Proposed method (MMSE)	**32.91**	**30.90**	**29.51**	**28.44**	**27.59**
	Bridge				
BivariateShrink [66]	30.41	27.93	26.40	25.33	24.58
LAWMAP [106]	30.42	28.06	26.58	25.55	24.77
NeighShrink [131]	30.35	27.91	26.36	25.31	24.50
ProbShrink [127]	30.14	27.75	26.27	25.25	24.50
SMG [42]	30.31	27.83	26.37	25.34	24.58
Proposed method (MAP)	30.47	28.01	26.57	25.55	**24.81**
Proposed method (MMSE)	**30.59**	**28.15**	**26.63**	**25.58**	**24.81**
	Baboon				
BivariateShrink [66]	30.03	27.48	25.84	24.65	23.77
LAWMAP [106]	30.06	27.53	25.93	24.78	23.94
NeighShrink [131]	29.97	27.44	25.79	24.61	23.68
ProbShrink [127]	29.78	27.27	25.65	24.48	23.61
SMG [42]	29.92	27.38	25.76	24.60	23.68
Proposed method (MAP)	30.10	27.61	26.01	24.85	24.04
Proposed method (MMSE)	**30.22**	**27.71**	**26.09**	**24.93**	**24.05**

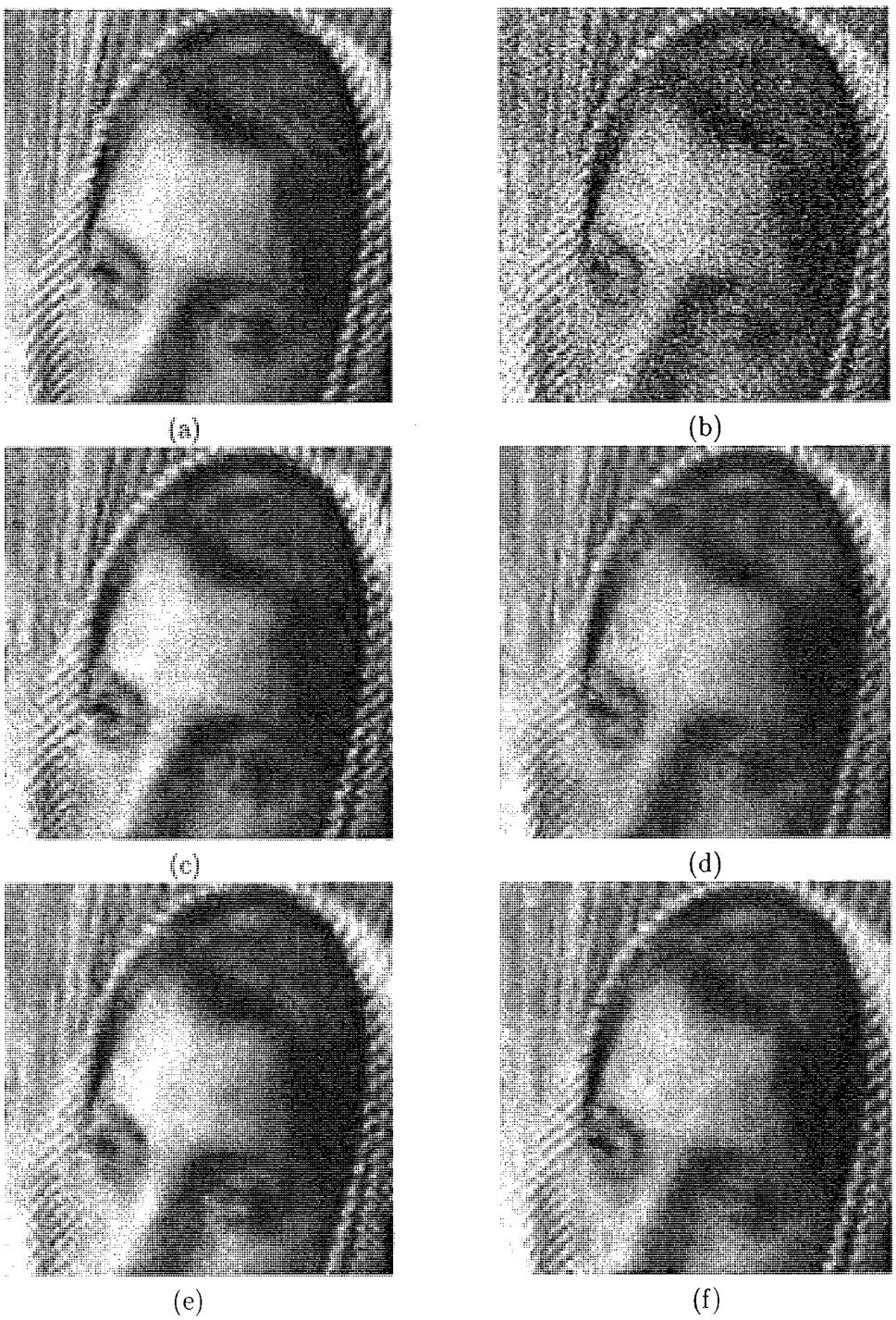


Figure 5.10: Comparison of the denoised images obtained from the different locally-adaptive DWT-based denoising algorithms on *Barbara* with $\sigma_\epsilon = 20$: (a) original image, (b) noisy image (PSNR=22.11 dB), and denoised images using the (c) Prob-Shrink method (PSNR=28.50 dB), (d) SMG method (PSNR=28.81 dB), (e) proposed MAP method (PSNR=29.02 dB) and (f) proposed MMSE method (PSNR=29.09 dB).

5.5 Conclusion

In order to overcome the shortcomings of the traditional PDFs of the wavelet coefficients of images, the MGH PDF was proposed in Chapter 3. In this chapter, in order to demonstrate the efficacy of the proposed MGH PDF in an estimation technique, we have designed and tested the MMSE- and MAP-based shrinkage functions for its use in an image noise reduction technique as a prior function. Both the subband- and locally-adaptive wavelet-based denoising techniques have been developed. The parameters required for the prior function have been estimated from the noisy observations. Extensive simulations have been carried out using a number of standard test images and the simulation results have shown that the performance of the proposed denoising methods, both in the subband- and locally-adaptive conditions, is better than that of the existing methods that use conventional PDFs for probabilistic modeling of the wavelet coefficients.

Chapter 6

Wavelet-Based Image Watermark Detection Using Proposed MGH PDF

6.1 Introduction

In this chapter, it is shown that the incorporation of the univariate MGH PDF, proposed in Chapter 3, in the development of a detector for a DWT-based image watermarking provides an improved detection performance [147], [148]. First, a brief introduction to image watermarking is given. Next, we consider an embedding technique for the proposed DWT-based additive image watermarking scheme. A detector is developed for the DWT-based watermarking using the MGH PDF as the prior, and its decision threshold and ROC derived. Experiments are carried out on a number of test images, and the results show that the proposed watermark detector performs better than the other standard detectors such as the Gaussian and GG do, in terms of the probabilities of detection and false alarm as well as the efficacy. It is also shown that detection performance of the proposed detector is more robust than the competitive GG detector in the case of compression, AWGN, filtering, or geometric attack.

6.2 Image Watermarking and DWT

A digital watermark is a distinguishing piece of information that is embedded into the data with the intention of protecting the intellectual property rights, verifying the authenticity of the data, annotating the data, etc. [149]-[153]. Very often, a watermarking scheme is seen as a secured communication task consisting of two steps, watermark embedding and watermark retrieval. In watermark embedding, the signal, i.e., the watermark, is transmitted through the host data that acts as a channel, whereas in watermark retrieval, the signal is received and extracted from the marked data [152]. The security of the watermarking is maintained in the same way as in cryptography by using a secret key. In contrast to encrypted data, a watermarked data can still be used with the embedded watermark. The legal owner knows the exact embedding process, which is based on a secret key, and, hence, can extract the watermark, while it is not possible do so for an unauthorized party. Thus, watermarking does not necessarily prevent the copying of the marked data; rather, the scheme identifies the original data source and its intended legitimate destination, so that copyright violations can at least be detected.

Digital image watermarking can be of two types: imperceptible and perceptible. Very often, the imperceptible watermarking is preferred to the perceptible one, since the distortion of the original image is minimum in the former than that in the latter, even in the case of embedding multiple watermarks. The imperceptibility constraint in the image watermarking is achieved by taking into account the properties of the HVS during the embedding process [90], [154]-[157]. There are several methods of watermark embedding, such as additive [19], [158]-[160], multiplicative [20], [90], [161], [162], and through quantization [163]-[165]. Whatever be the embedding process, the watermark travels with the marked image and therefore, remains protected until it is removed. In order to resolve the rightful ownership of an image, detection of the watermark is

sufficient assuming that the owner of the marked image possesses the exact secret key and has the knowledge of the embedding process [166]. If a marked image has a single watermark, the detection of the watermark becomes simple provided the original host image is available to the watermark detector. But, for a marked image having multiple watermarks there is ambiguity in resolving the rightful ownership, even when original host image is available to the watermark detector [167]. In addition to this fact, in many applications such as data monitoring and tracking on the Internet, the original host data is not always available and in many multimedia settings such as video watermarking, the use of unmarked host data is not practical due to its huge volume. As a result, it is necessary to design a blind watermark detector that does not require the original host data in the detection process [168].

One of the essential requirements of watermarking is its robustness [17], [18], [155], [169], [170], which means that the watermark can be detected reliably when the marked data goes through any intentional or unintentional standard manipulations such as filtering and geometric attacks. Robustness of a watermarking scheme can be improved during the embedding process as well as during the detection process. The simplest way to improve the robustness during embedding is to increase the amount of embedded information, referred to as the payload, in the host data. In image watermarking, maximum possible payload can be embedded along with the imperceptibility constraint by taking into account the properties of the HVS. Theoretical studies on information embedding suggest that the payload may be increased, if the host data is treated as a side information in the embedder [20]. Nevertheless, the spread spectrum watermarking scheme is the most popular one, since it possesses a very high level of security and robustness [19], [149], [150], [162], [171]-[174]. In such a scheme, the watermark is embedded in a transformed domain and various transforms have been employed for this purpose. Some of these are the

DFT [149], [150], [153], [175], the DCT [19], [20], [152]-[154], [166], [169], and the DWT [17], [18], [90], [132], [148], [150], [155]-[157], [161], [164], [176]. Dual transformed domains have also been used for embedding the watermark, using the DWT-DFT [177] and DCT-DWT [178]. It is well known [157], [158] that the recent image coding standard, JPEG 2000 [1], strongly relies on the DWT for obtaining a good quality image having a low bit rate. Moreover, different subbands of the DWT facilitate the embedding of the watermark by exploiting the HVS more efficiently than the DCT or DFT coefficients do. In such a case, the DWT-based algorithm can embed the watermark with a higher strength, yet maintaining the imperceptibility requirement [179]-[181]. The embedded payload of a DWT-based watermarking system may be increased by using balanced multiwavelets [182]-[184]. Thus, in general, the DWT-based watermark techniques are more robust against any attack [156], [176], [185], [186].

Over the last decade, an appreciable number of watermarking schemes have been developed, wherein the watermark is embedded into the DWT coefficients of the image. The most commonly used detector for these watermarking schemes is the correlation detector [90], [150], [153]-[157], [161], [164], [166], [169], [176], [187], which is optimal only if the data samples follow the Gaussian distribution function [188]. Since the DWT coefficients of an image are non-Gaussian, such a Gaussian detector is neither optimal nor robust. In order to obtain an optimal detector that is also robust, an appropriate approach is to treat the watermark detection as a statistical detection problem, wherein a binary hypothesis test is formulated using a more accurate PDF of the DWT coefficients of the image. Various types of binary hypothesis testings have been used for image watermark detection, and some of these are the Bayesian log-likelihood ratio test [19], [152], [160], [168], [171], [189]-[192], locally most powerful test [159], [160], [174], [193], [194] and Rao test [160], [180], [192], [195]. The choice of the test depends on various factors, such as the number of data samples, availability

and mathematical tractability of the prior function, and strength of the watermark signal. The Bayesian log-likelihood ratio test shows an asymptotically optimal performance that is equivalent to that of the Rao test, when the number of data samples are large, the prior probability distribution function of the signal is known, and the signal to be detected is weak [88], [180]. Since the subbands of the DWT of an image have a sufficiently large number of data samples, the prior of the watermark signal is very often known, and the strength of watermark is weaker than that of the image signal, the Bayesian log-likelihood ratio test is a good choice as a binary hypothesis test for the detection of a DWT-based watermarking scheme. The performance of the statistical detector can be analyzed theoretically by relating the probability of detection and the probability of false alarm, and measuring them through experiments. It is essential to keep the probability of detection at a high level for a given rate of false alarm to increase the reliability of detection. Statistical detectors for watermarking have been developed using various PDFs, such as the Laplacian [191], Student-t [192], and GG [18], [159], [160], [180], [194] for the prior of the DWT coefficients of a given image. Thus, statistical detectors employed for DWT-based image watermarking incorporate a PDF having a fixed number of parameters that depend on the first few moments of the image coefficients. In fact, each subband of the DWT of an image has a sufficiently large number of data samples, and higher-order moments are useful for statistical detection.

In this chapter, a novel DWT-based watermark detector is developed using the proposed univariate MGH PDF due to the fact that this PDF provides an excellent performance in matching the empirical PDF of image coefficients. This is mainly due to the fact that the MGH PDF allows the incorporation of an appropriate number of parameters that depend on the higher-order moments for the probabilistic modeling of the DWT coefficients of an image. Since a more accurate PDF of the image

coefficients would improve the watermark detection performance in terms of the probabilities of detection and false alarm, a statistical detector based on the MGH PDF is developed. The performance of the proposed MGH detector is analyzed using the ROC and efficacy of detector (ED). Extensive experimentations for a wide range of watermark strength are carried out to show the effectiveness of the proposed MGH detector in terms of the probabilities of detection and false alarm as well as to study its robustness, when the marked images are attacked by compression, AWGN, filtering, or geometric distortion.

6.3 Watermark Embedding

In the embedding process of a DWT-based additive spread spectrum watermarking scheme, the watermark signal is added to the DWT coefficients of an image that is to be marked. The watermark is generated using a pseudorandom sequence (PRS) generator, which is initialized with a seed that depends on a secret key. This PRS spreads the spectrum of the original watermark signal over many coefficients making it difficult to be detected. In order to maximize the security and robustness of the watermarking scheme, the PRS needs to have white-noise like properties [171], [196]. Let such a two dimensional PRS, for a given subband, be denoted ν_{ij}, where it takes the values $+1$ or -1 with equal probability; thus, the mean is zero and the autocorrelation is a Kronecker delta function. In order to obtain a robust watermarking system, the watermark signal should have maximum strength but should not affect the perceptual quality of the image; for this purpose, ν_{ij}'s are usually multiplied by a visual mask that takes into consideration the HVS [90], [156], [157], [173]. Here, we use a subband dependent parameter ξ as the visual mask [173] to obtain the watermark signal as $w_{ij} = \xi\nu_{ij}$. Then, the DWT coefficient of the watermarked image at the spatial location (i, j) of a subband, under the assumption that the image and the

watermark are independent of each other, is given by

$$g_{ij} = f_{ij} + w_{ij} \tag{6.1}$$

The parameter ξ is selected in such a way that each of the subbands has the same level of watermark-to-document ratio (WDR) defined as [160], [171], [193]

$$\text{WDR} = 10\log\left(\frac{\sigma_w^2}{\sigma_f^2}\right) = 10\log\left(\frac{\xi^2}{\sigma_f^2}\right) \tag{6.2}$$

where σ_w^2 is the variance of the embedded watermark. The watermarked DWT coefficients are then inverse transformed to get the marked image. A block diagram of the watermark embedding process is shown in Fig. 6.1.

6.4 Watermark Detection

In general, a watermarking scheme for copyright protection has a known watermark; hence, the verification of its existence, i.e., the detection of the watermark, is sufficient for checking the violation of a copyright. The DWT-based watermark detection process that is being proposed considers the watermark w_{ij} as the desired signal, and the subband image coefficient f_{ij} as the unknown additive noise. The verification of the existence of w_{ij} in the subband coefficients of an image can be formulated as a binary hypothesis test given by

$$\mathcal{H}0 : g_{ij} = f_{ij} \tag{6.3a}$$

$$\mathcal{H}1 : g_{ij} = f_{ij} + w_{ij} \tag{6.3b}$$

where $\mathcal{H}0$ and $\mathcal{H}1$, respectively, are the null and alternative hypotheses. In a Bayesian framework, g_{ij}, f_{ij}, and w_{ij} in (6.3) are considered as samples of the independent random variables $\boldsymbol{g}$, $\boldsymbol{f}$, and $\boldsymbol{w}$, respectively. Let $p_{\boldsymbol{g}}(\boldsymbol{g})$ and $p_{\boldsymbol{w}}(\boldsymbol{w})$, respectively, be the PDFs of the random variables $\boldsymbol{g}$ and $\boldsymbol{w}$. Then, the Bayesian log-likelihood ratio,

$\ln\Lambda$ ($-\infty < \ln\Lambda < \infty$), for N ($N = N_{1\ell}N_{2\ell}$) number of samples in a subband is given by [188]

$$\ln\Lambda(\boldsymbol{g}) \equiv \ln\frac{\prod_{u=1}^{N} p_{g_u}(\boldsymbol{g}_u|\mathcal{H}1)}{\prod_{u=1}^{N} p_{g_u}(\boldsymbol{g}_u|\mathcal{H}0)} \gtrless_{\mathcal{H}0}^{\mathcal{H}1} \ln\frac{\mathtt{C}_{\mathtt{fa}}\Pr(\mathcal{H}0)}{\mathtt{C}_{\mathtt{m}}\Pr(\mathcal{H}1)} \equiv \tau \tag{6.4}$$

where $\mathtt{C}_{\mathtt{fa}}$ is the cost of false alarm (i.e., watermark being detected although not embedded), $\mathtt{C}_{\mathtt{m}}$ is the cost of miss-detection (i.e., watermark being not detected although present), $\Pr(\mathcal{H}0)$ is the probability of null hypothesis, $\Pr(\mathcal{H}1)$ is the probability of alternative hypothesis, and τ is the decision threshold. The threshold τ is obtained by using the Neyman-Pearson criterion that minimizes the probability of miss-detection for a bounded false alarm probability [188]. It is to be noted that the threshold τ, which is sufficient to detect the watermark, is totally defined by the associated costs and probabilities of the two hypotheses. The decision about the existence of the watermark is, therefore, made blindly, i.e., without the use of the unmarked image coefficients. A block diagram of the watermark detection process is shown in Fig. 6.2.

6.4.1 Proposed Watermark Detector

In order to obtain an optimal detector using (6.4), an accurate statistical model of the random variable $\boldsymbol{f}$ is crucial, since the PDF of the random variable $\boldsymbol{w}$ is often known. Thus, greater the accuracy of the PDF of the image coefficients incorporated in the detector, the higher the reliability of detection of the watermark at a predefined false alarm rate. In Chapter 3, it has been shown that the MGH PDF provides a match with the empirical PDFs of the image coefficients that is better than that provided by the other standard ones, such as the GG, BK, and SNIG. It has also been shown in Chapter 5 that the MGH prior function results in an improved performance for the DWT-based Bayesian denoising. Motivated by these results, a statistical detector is developed for the DWT-based image watermarking using the MGH PDF as the prior function $p_f(\boldsymbol{f})$, and hereafter the detector is referred to as the MGH detector. For

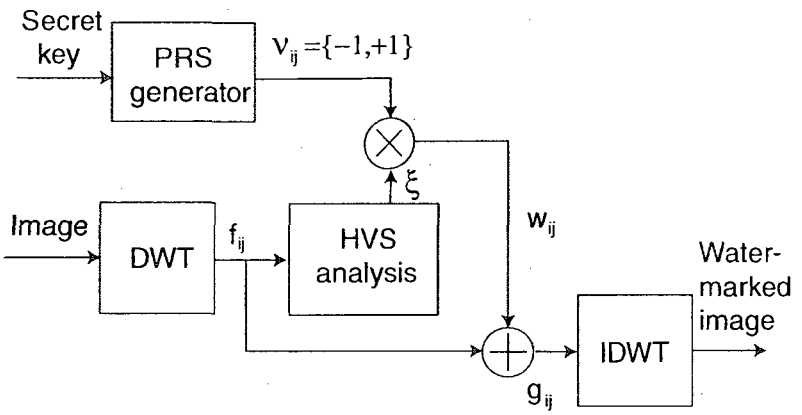


Figure 6.1: Block diagram of the embedding process for the DWT-based additive watermarking.

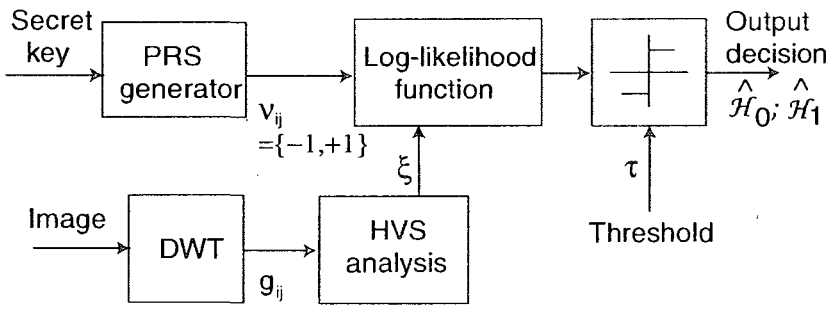


Figure 6.2: Block diagram of the detection process for the DWT-based additive watermarking.

the MGH detector, the log-likelihood ratio given in (6.4) can be written as [147]

$$\begin{aligned}
\ln\Lambda(\boldsymbol{g}) &= \ln\frac{\prod_{u=1}^{N} p_{\boldsymbol{f}_u}(\boldsymbol{g}_u-\boldsymbol{w}_u)}{\prod_{u=1}^{N} p_{\boldsymbol{f}_u}(\boldsymbol{g}_u)} \\
&= \ln\frac{\prod_{\substack{u=1\\|\boldsymbol{f}_u|<f_0}}^{N'} p_{\zeta\boldsymbol{f}_u}(\boldsymbol{g}_u-\boldsymbol{w}_u)\prod_{\substack{u=1\\|\boldsymbol{f}_u|\geq f_0}}^{N''} p_{4\boldsymbol{f}_u}(\boldsymbol{g}_u-\boldsymbol{w}_u)}{\prod_{\substack{u=1\\|\boldsymbol{f}_u|<f_0}}^{N'} p_{\zeta\boldsymbol{f}_u}(\boldsymbol{g}_u)\prod_{\substack{u=1\\|\boldsymbol{f}_u|\geq f_0}}^{N''} p_{4\boldsymbol{f}_u}(\boldsymbol{g}_u)} \\
&= \frac{1}{2\sigma_{\boldsymbol{f}}^2}\left[\sum_{u=1}^{N'}\boldsymbol{g}_u^2-\sum_{u=1}^{N'}(\boldsymbol{g}_u-\boldsymbol{w}_u)^2\right]+\ln\frac{\prod_{\substack{u=1\\|\boldsymbol{f}_u|<f_0}}^{N'}\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{g}_u-\boldsymbol{w}_u}{\sigma_{\boldsymbol{f}}}\right)\right]}{\prod_{\substack{u=1\\|\boldsymbol{f}_u|<f_0}}^{N'}\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{g}_u}{\sigma_{\boldsymbol{f}}}\right)\right]} \\
&\quad+\frac{1}{2\sigma_{\boldsymbol{f}}^2}\left[\sum_{u=1}^{N''}\boldsymbol{g}_u^2-\sum_{u=1}^{N''}(\boldsymbol{g}_u-\boldsymbol{w}_u)^2\right]+\ln\frac{\prod_{\substack{u=1\\|\boldsymbol{f}_u|\geq f_0}}^{N''}\left[1+\beta_{4m}H_4\left(\frac{\boldsymbol{g}_u-\boldsymbol{w}_u}{\sigma_{\boldsymbol{f}}}\right)\right]}{\prod_{\substack{u=1\\|\boldsymbol{f}_u|\geq f_0}}^{N''}\left[1+\beta_{4m}H_4\left(\frac{\boldsymbol{g}_u}{\sigma_{\boldsymbol{f}}}\right)\right]} \\
&= \frac{1}{2\sigma_{\boldsymbol{f}}^2}\left[\sum_{u=1}^{N}\boldsymbol{g}_u^2-\sum_{u=1}^{N}(\boldsymbol{g}_u-\boldsymbol{w}_u)^2\right]+\sum_{\substack{u=1\\|\boldsymbol{f}_u|<f_0}}^{N'}\ln\frac{\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{g}_u-\boldsymbol{w}_u}{\sigma_{\boldsymbol{f}}}\right)\right]}{\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{g}_u}{\sigma_{\boldsymbol{f}}}\right)\right]} \\
&\quad+\sum_{\substack{u=1\\|\boldsymbol{f}_u|\geq f_0}}^{N''}\ln\frac{\left[1+\beta_{4m}H_4\left(\frac{\boldsymbol{g}_u-\boldsymbol{w}_u}{\sigma_{\boldsymbol{f}}}\right)\right]}{\left[1+\beta_{4m}H_4\left(\frac{\boldsymbol{g}_u}{\sigma_{\boldsymbol{f}}}\right)\right]} \qquad (6.5)
\end{aligned}$$

where N' and N'' ($N=N'+N''=N_{1\ell}N_{2\ell}$), respectively, are the number of subband coefficients that satisfy $|\boldsymbol{f}|<f_0$ and $|\boldsymbol{f}|\geq f_0$. It can be seen from (6.5) that the log-likelihood ratio is independent of the weight parameter γ of the MGH PDF mentioned in Section 3.2. The proposed MGH detector requires the estimation of β_{2r} and H_{2r}, and the computational complexities of these are $\mathcal{O}(\zeta N/2)$ and $\mathcal{O}(N)$, respectively. Thus, the computational complexity of the proposed MGH detector is $\mathcal{O}\Big((4+\zeta/2)N\Big)$. It may be mentioned that assuming the subband DWT coefficients of an image follow the GG PDF, the log-likelihood ratio of the GG detector is given by [19], [152], [159], [160], [168], [180]

$$\ln\Lambda(\boldsymbol{g})=\frac{1}{c^s}\left[\sum_{u=1}^{N}|\boldsymbol{g}_u|^s-\sum_{u=1}^{N}|\boldsymbol{g}_u-\boldsymbol{w}_u|^s\right] \qquad (6.6)$$

The computational complexity of estimating the parameters of GG PDF (i.e., s and c) is $\mathcal{O}(2\jmath N)$, where $\jmath$ is the number of iterations for the maximum likelihood-based

parameter estimation [30] or that of the entries in a lookup table for the MoM or entropy matching-based parameter estimation [31]. Thus, the computational complexity of the GG detector is $\mathcal{O}\Big(2(\jmath+1)N\Big)$. The log-likelihood ratio given in (6.6) reduces to that of the Gaussian detector when $s=2$ and $c=\sigma_{\boldsymbol{f}}\sqrt{2}$, and that of the Laplacian detector when $s=1$ and $c=\sigma_{\boldsymbol{f}}/\sqrt{2}$. Thus, the computational complexity of both of these detectors is $\mathcal{O}(3N)$.

6.4.2 Performance Analysis of MGH Detector

The analytical expression of $\ln\Lambda$ given in (6.5) allows the theoretical measurement and experimental verification of the performance of the proposed MGH detector, since the ROC can be derived from this expression. From (6.5), it is seen that for a particular subband, $\ln\Lambda$ is a sum of statistically independent random variables that have a sufficiently large number of data samples. In such a case, according to the central limit theorem [89], the distribution of $\ln\Lambda$ can be approximated by a Gaussian distribution. In order to derive an expression for the ROC, it is required to estimate the mean and the variance of $\ln\Lambda$ conditioned on the two hypotheses, viz., $\mathcal{H}0$ and $\mathcal{H}1$. As explained in Section 6.4, during the embedding process, the watermark that is added to the image coefficient at a given index of a subband is equal to either $+\xi$ or $-\xi$, each having a probability of 1/2. We now define two functions

$$\mathcal{X}_{\zeta}(\boldsymbol{f}_u,\xi)=\frac{\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{f}_u-\xi}{\sigma_{\boldsymbol{f}}}\right)\right]}{\left[1+\sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\frac{\boldsymbol{f}_u}{\sigma_{\boldsymbol{f}}}\right)\right]} \tag{6.7}$$

$$\mathcal{Y}(\boldsymbol{f}_u,\xi)=\frac{\boldsymbol{f}_u^2-(\boldsymbol{f}_u-\xi)^2}{2\sigma_{\boldsymbol{f}}^2} \tag{6.8}$$

Then, the mean of the log-likelihood function conditioned on $\mathcal{H}0$ can be obtained as

$$\mu_{\Lambda 0} = E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}0\right] = -\frac{N\xi^2}{2\sigma_{\boldsymbol{f}}^2} + \frac{1}{2}\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\left[\ln \mathcal{X}_\zeta(\boldsymbol{f}_u,\xi) + \ln \mathcal{X}_\zeta(\boldsymbol{f}_u,-\xi)\right] + \frac{1}{2}\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\left[\ln \mathcal{X}_{4m}(\boldsymbol{f}_u,\xi) + \ln \mathcal{X}_{4m}(\boldsymbol{f}_u,-\xi)\right] \quad (6.9)$$

where $E_{\boldsymbol{w}}[\cdot]$ is the expectation operator on $\boldsymbol{w}$ and $\mathcal{X}_{4m} = \mathcal{X}_4|_{\beta_4=\beta_{4m}}$. The variance of the log-likelihood function conditioned on $\mathcal{H}0$ can be obtained as

$$\sigma_{\Lambda 0}^2 = E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}0\right]^2 - \mu_{\Lambda 0}^2 \quad (6.10)$$

where

$$E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}0\right]^2 = E_{\boldsymbol{w}}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\mathcal{Y}(\boldsymbol{f}_u,\boldsymbol{w}) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\ln \mathcal{X}_\zeta(\boldsymbol{f}_u,\boldsymbol{w})\right]^2 + E_{\boldsymbol{w}}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\mathcal{Y}(\boldsymbol{f}_u,\boldsymbol{w}) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\ln \mathcal{X}_{4m}(\boldsymbol{f}_u,\boldsymbol{w})\right]^2 \quad (6.11)$$

that can be written in the form

$$\begin{aligned} E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}0\right]^2 = {} & \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\mathcal{Y}(\boldsymbol{f}_u,\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\ln \mathcal{X}_\zeta(\boldsymbol{f}_u,\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\mathcal{Y}(\boldsymbol{f}_u,-\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\ln \mathcal{X}_\zeta(\boldsymbol{f}_u,-\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\mathcal{Y}(\boldsymbol{f}_u,\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\ln \mathcal{X}_{4m}(\boldsymbol{f}_u,\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\mathcal{Y}(\boldsymbol{f}_u,-\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\ln \mathcal{X}_{4m}(\boldsymbol{f}_u,-\xi)\right]^2 \end{aligned} \quad (6.12)$$

Similarly, the mean of the log-likelihood function conditioned on $\mathcal{H}1$ can be obtained as

$$\mu_{\Lambda 1} = E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}1\right] = \frac{N\xi^2}{2\sigma_{\boldsymbol{f}}^2} + \frac{1}{2}\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'}\left[\ln \mathcal{X}_{\zeta}^{-1}(\boldsymbol{f}_u;\xi) + \ln \mathcal{X}_{\zeta}^{-1}(\boldsymbol{f}_u,-\xi)\right] + \frac{1}{2}\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''}\left[\ln \mathcal{X}_{4m}^{-1}(\boldsymbol{f}_u;\xi) + \ln \mathcal{X}_{4m}^{-1}(\boldsymbol{f}_u,-\xi)\right] \tag{6.13}$$

and the variance as

$$\sigma_{\Lambda 1}^2 = E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}1\right]^2 - \mu_{\Lambda 1}^2 \tag{6.14}$$

where

$$\begin{aligned} E_{\boldsymbol{w}}\left[\ln \Lambda(\boldsymbol{g})|\mathcal{H}1\right]^2 = {} & \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'} -\mathcal{Y}(\boldsymbol{f}_u,\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'} \ln \mathcal{X}_{\zeta}^{-1}(\boldsymbol{f}_u,\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'} -\mathcal{Y}(\boldsymbol{f}_u,-\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|<f_0}}^{N'} \ln \mathcal{X}_{\zeta}^{-1}(\boldsymbol{f}_u,-\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''} -\mathcal{Y}(\boldsymbol{f}_u,\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''} \ln \mathcal{X}_{4m}^{-1}(\boldsymbol{f}_u,\xi)\right]^2 \\ & + \frac{1}{2}\left[\sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''} -\mathcal{Y}(\boldsymbol{f}_u,-\xi) + \sum_{\substack{u=1 \\ |\boldsymbol{f}_u|\geq f_0}}^{N''} \ln \mathcal{X}_{4m}^{-1}(\boldsymbol{f}_u,-\xi)\right]^2 \end{aligned} \tag{6.15}$$

It can be seen from (6.9)-(6.15) that

$$\mu_{\Lambda 1} = -\mu_{\Lambda 0} \qquad \sigma_{\Lambda 1}^2 = \sigma_{\Lambda 0}^2 \tag{6.16}$$

For the GG detector, the mean and variance of the conditional log-likelihood function can be obtained as [19], [152], [159], [160], [168], [180]

$$\mu_{\Lambda 0} = -\mu_{\Lambda 1} = \frac{1}{c^s}\sum_{u=1}^{N}|\boldsymbol{f}_u|^s - \frac{1}{2c^s}\sum_{u=1}^{N}(|\boldsymbol{f}_u+\xi|^s + |\boldsymbol{f}_u-\xi|^s) \tag{6.17}$$

$$\sigma_{\Lambda 0}^2 = \sigma_{\Lambda 1}^2 = \frac{1}{4c^{2s}}\sum_{u=1}^{N}(|\boldsymbol{f}_u+\xi|^s - |\boldsymbol{f}_u-\xi|^s)^2 \tag{6.18}$$

and for the Gaussian detector, they can be obtained as

$$\mu_{\Lambda 0} = -\mu_{\Lambda 1} = -\frac{N\xi^2}{2\sigma_f^2} \qquad \sigma_{\Lambda 0}^2 = \sigma_{\Lambda 1}^2 = \frac{N\xi^2}{\sigma_f^2} \tag{6.19}$$

In a blind watermark detection technique, it is reasonable to assume that the probabilities of occurrences of the two hypotheses to be equal, i.e., $\Pr(\mathcal{H}0) = \Pr(\mathcal{H}1)$. In such a case, the probability of false alarm and that of detection, for a particular value of τ, can be calculated from the estimated values of the mean and variance of the conditional log-likelihood function as [188]

$$P_{fa} = Q\left(\frac{\tau - \mu_{\Lambda 0}}{\sigma_{\Lambda 0}}\right) \qquad P_{det} = Q\left(\frac{\tau - \mu_{\Lambda 1}}{\sigma_{\Lambda 1}}\right) \tag{6.20}$$

where $Q(u) = \frac{1}{2}\mathrm{erfc}(u/\sqrt{2})$ and $\mathrm{erfc}(\cdot) = 1 - \mathrm{erf}(\cdot)$ is the complementary error function. For a predetermined P_{fa}, the decision threshold τ can be expressed as

$$\tau = \mu_{\Lambda 0} + \sigma_{\Lambda 0} Q^{-1}\left(P_{fa}\right) \tag{6.21}$$

and P_{det} as

$$P_{det} = Q\left(Q^{-1}\left(P_{fa}\right) - 2\sqrt{\frac{\mu_{\Lambda 1}^2}{\sigma_{\Lambda 1}^2}}\right) \tag{6.22}$$

By defining the signal-to-noise-ratio, $\mathrm{SNR} \triangleq \mu_{\Lambda 1}^2/\sigma_{\Lambda 1}^2$ as in [19], [168], [171], [188], we obtain the relation between the P_{fa} and P_{det} as

$$P_{det} = Q\left(Q^{-1}\left(P_{fa}\right) - 2\sqrt{\mathrm{SNR}}\right) \tag{6.23}$$

which is known as the ROC of the detector. It can be seen from (6.23) that the ROC of a detector depends only on the SNR for a fixed value of P_{fa}. Hence, the ROC of different detectors can be compared by calculating the corresponding SNR of the detectors. It is to be noted that in ROC, a larger value of SNR implies a better detection performance, as this will lead to a higher probability of detection for a given P_{fa}. This fact can be further explained by using the conditional PDFs of $\ln \Lambda$ for the

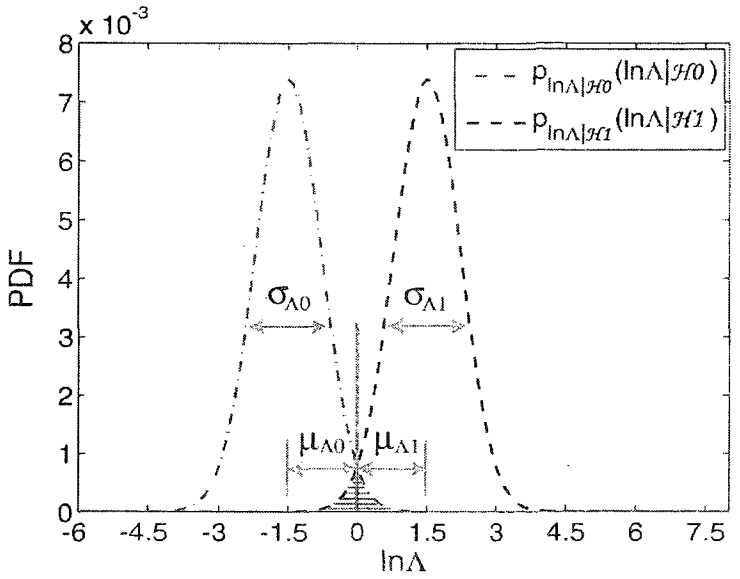


Figure 6.3: Effect of conditional mean and variance of log-likelihood ratio on the performance of a detector.

two hypotheses, viz., $\mathcal{H}0$ and $\mathcal{H}1$ as shown in Fig. 6.3. From this figure, it is seen that the higher the value of the ratio between $\mu_{\Lambda 1}^2$ and $\sigma_{\Lambda 1}^2$, i.e., the SNR, the lower the probability of erroneous detection, which is shown as the union of the areas under the two conditional PDFs. Another measure of performance often used in statistical detection problems, where conditional $\ln \Lambda$ is considered to be normally distributed, is the ED defined by [197]

$$\mathrm{ED} = \frac{\mu_{\Lambda 1} - \mu_{\Lambda 0}}{\sigma_{\Lambda 0}} \tag{6.24}$$

Thus, in addition to the ROC, the detectors can be compared using the ED values, where a higher value of efficacy corresponds to a better detector.

Since the proposed watermark detection is a blind one, it is required that the mean and the variance of the conditional $\ln \Lambda$ that depend on the parameters of the MGH PDF, namely, σ_f and β_{2r} $(r = 2, 3, \cdots, \zeta/2)$, be estimated from the received image, which may or may not contain a watermark. However, this is not a serious issue, as the watermark embedding does not change these parameters significantly in view of

Table 6.1: Parameters of the MGH PDF estimated from the DWT-coefficients of the marked and unmarked images.

	γ	σ_f	β_4	β_6	β_8	β_{10}	β_{12}	β_{14}
	Lena							
Marked	0.3514	41.8020	0.9581	-0.5380	2.5035	-0.7852	4.1475	-1.0161
Unmarked	0.3507	41.8356	0.9562	-0.5328	2.4721	-0.7687	4.0459	-0.9838
	Barbara							
Marked	0.6140	46.5555	0.8329	-0.3431	1.3054	-0.3259	1.0700	-0.4546
Unmarked	0.6146	46.5914	0.8314	-0.3392	1.2950	-0.3235	1.0565	-0.4526
	Boat							
Marked	0.6375	47.5901	0.6882	-0.1238	0.6232	-0.1334	0.5045	-0.3670
Unmarked	0.6357	47.6403	0.6872	-0.1228	0.6199	-0.1332	0.5009	-0.3691
	Bridge							
Marked	0.6873	54.6063	0.4767	-0.0775	0.3016	-0.2162	0.3357	-0.4173
Unmarked	0.6896	54.6304	0.4772	-0.0781	0.3025	-0.2156	0.3361	-0.4171
	Baboon							
Marked	0.6892	59.0495	0.4597	-0.1033	0.3018	-0.2163	0.3270	-0.4182
Unmarked	0.6921	59.1677	0.4594	-0.1035	0.3019	-0.2162	0.3269	-0.4181
	Aerial							
Marked	0.8073	70.5873	0.5736	-0.2847	1.8360	-1.2243	9.1274	-3.9755
Unmarked	0.8046	70.6509	0.5729	-0.2831	1.8158	-1.2056	8.9319	-3.8815

the fact that the signal of the watermark has a much lower energy compared to that of the original image. That this is indeed so is verified by estimating the parameters of MGH PDF both from the marked and unmarked images. Table 6.1 shows the average values of γ, σ_f, and β_{2r} $(r = 2, 3, 4, 5, 6, 7)$ obtained from the subbands of a 4-level orthogonal DWT using six 512×512 grayscale images, viz., *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial*, wherein for watermarking we use WDR $= -30$ dB. It is seen from this table that the deviations between the parameters of the MGH PDF for the marked and unmarked images are negligible. Hence, for estimating $\mu_{\Lambda 0}$, $\mu_{\Lambda 1}$, $\sigma^2_{\Lambda 0}$, and $\sigma^2_{\Lambda 1}$ of the detector that uses the MGH PDF, one can use the received image without any consideration as to whether it is marked or unmarked.

6.5 Experimental Results

Extensive simulations are carried out on a large number of images in order to compare the performance of the proposed MGH detector with that of the others, namely, the Gaussian [90], [150], [153]-[157], [161], [164], [166], [169], [176] and GG [19], [152], [159], [160], [168], [180] detectors in terms of the ROC and the corresponding ED values. Here, we give results concerning only six 512×512 grayscale images, namely, *Lena*, *Barbara*, *Boat*, *Bridge*, *Baboon*, and *Aerial*. The orthogonal wavelet filter Daubechies-8 (D-8) and biorthogonal spline wavelet filter Cohen-Daubechies-Feauveau-9/7 (CDF-9/7) with 4-level decompositions are used. Monte-Carlo tests are performed to experimentally validate the estimated ROC and ED values using 1000 runs. Each of the 1000 runs starts with a unique key to randomly generate the bipolar watermark signals. For a particular run, the same key is used to generate the bipolar watermark signals that are added to all the detail subbands, namely, HL_ℓ, LH_ℓ, and HH_ℓ ($\ell = 1, 2, 3, 4$). No watermark is added to the approximate subband LL_4. The watermarks are embedded in each of the test images for various values of WDR. For a particular WDR, we obtain $3 \times 4 \times 1000 = 12000$ values of P_{det} for a given P_{fa} in view of the fact that the experimentation uses 1000 runs and a 4-level DWT. The ROC of a statistical detector for a test image is calculated by averaging these P_{det}'s for a given P_{fa} and repeating this procedure for various values of P_{fa} ranging from 10^{-4} to 1. Similarly, the ED is calculated by averaging its 12000 estimated values. The value of ζ for calculating the log-likelihood ratio of the proposed MGH detector is chosen to be 14 for the DWT using filter D-8, since such a value is a good choice for modeling the prior function of the image DWT coefficients using an orthogonal filter (see Chapter 3). In our experiments, it is found that the MGH PDF with $\zeta = 18$ provides an appreciable modeling performance while the filter CDF-9/7 is used in the DWT, and hence, in such a case the value of ζ for the proposed MGH detector is

(a) (b) (c)

Figure 6.4: Watermark embedding on *Lena* with a WDR of = −45 dB. (a) Cropped section of the *Lena* image. (b) Watermark in HL_3 band. (c) Watermarked version of (a).

chosen to be 18. The parameters of the GG PDF are estimated using the MoM, since each of the subbands has a sufficiently large number of data samples and in such a situation, MoM provides consistent parameter estimation [32]. The performance of the proposed statistical detector is tested using two sets of experiments. In the first set of experiments, watermarked images have been considered without any kind of attack. In the second set of experiments, the marked images are considered to have been attacked by compression, AWGN, filtering, or geometric distortion.

6.5.1 Detection Performance Without Attack

In this set of experiments, the test images are considered to be modified only by the additive bipolar watermarks in the DWT-domain. The *Lena* image is considered as an example and Fig. 6.4 shows a cropped section of this image, a typical bipolar watermark signal, and the watermarked version of the cropped *Lena* image, when WDR = −45 dB. From this figure, it can be observed that there is no noticeable difference between the original and watermarked images, and hence, the imperceptibility requirement is fulfilled during the embedding process. For a given watermark strength, the performance of the proposed MGH detector is compared with that of

the Gaussian and GG detectors in terms of the ROC and ED. Figs. 6.5(a)-6.5(f) show the ROC of the three detectors for the six test images for $10^{-4} \leq P_{fa} \leq 1$ and WDR $= -40$ dB, while the filters used in the DWT are D-8 and CDF-9/7. It is to be pointed out that the ROC of the Gaussian detector is independent of the test image as well as the wavelet filter, in view of the fact that for this detector P_{det} is calculated using SNR $\triangleq 0.25N(10)^{\mathrm{WDR}/10}$, which depends only on the size of the subband and WDR. It is seen from Fig. 6.5 that for all the test images, the proposed MGH detector has the highest probability of detection for any given value of false alarm. In order to compare the performance of the detectors for varying watermark strengths, we consider the values of WDR in the range -45 dB to -35 dB. For each of the test images, the P_{det}'s of the three detectors are obtained as the average of the detection probabilities calculated from all the twelve subbands for the entire range by fixing the P_{fa} value at 10^{-2}. Figs. 6.6(a)-6.6(f) show these average values obtained from 1000 runs for the six test images, while the filters D-8 and CDF-9/7 are used in the DWT. From these figures, it can be observed that P_{det} increases with the strength of the watermark for all the detectors. In addition, except in the case of the *Boat* image with a value of WDR greater than -37 dB when the wavelet filter D-8 is used, the detection probabilities of the proposed MGH detector are higher than or at least equal to that of the GG and Gaussian detectors at any level of the watermark strength; even in the case of the *Boat* image, the detection probability of the GG detector is not significantly higher than that of the proposed one. It may be found from the estimated P_{det}'s of the MGH detector, shown in Figs. 6.5 and 6.6, that the wavelet filter D-8 provides a better detection performance for the images having minute details (*Baboon* and *Aerial*), whereas for other test images the use of CDF-9/7 results in a better performance.

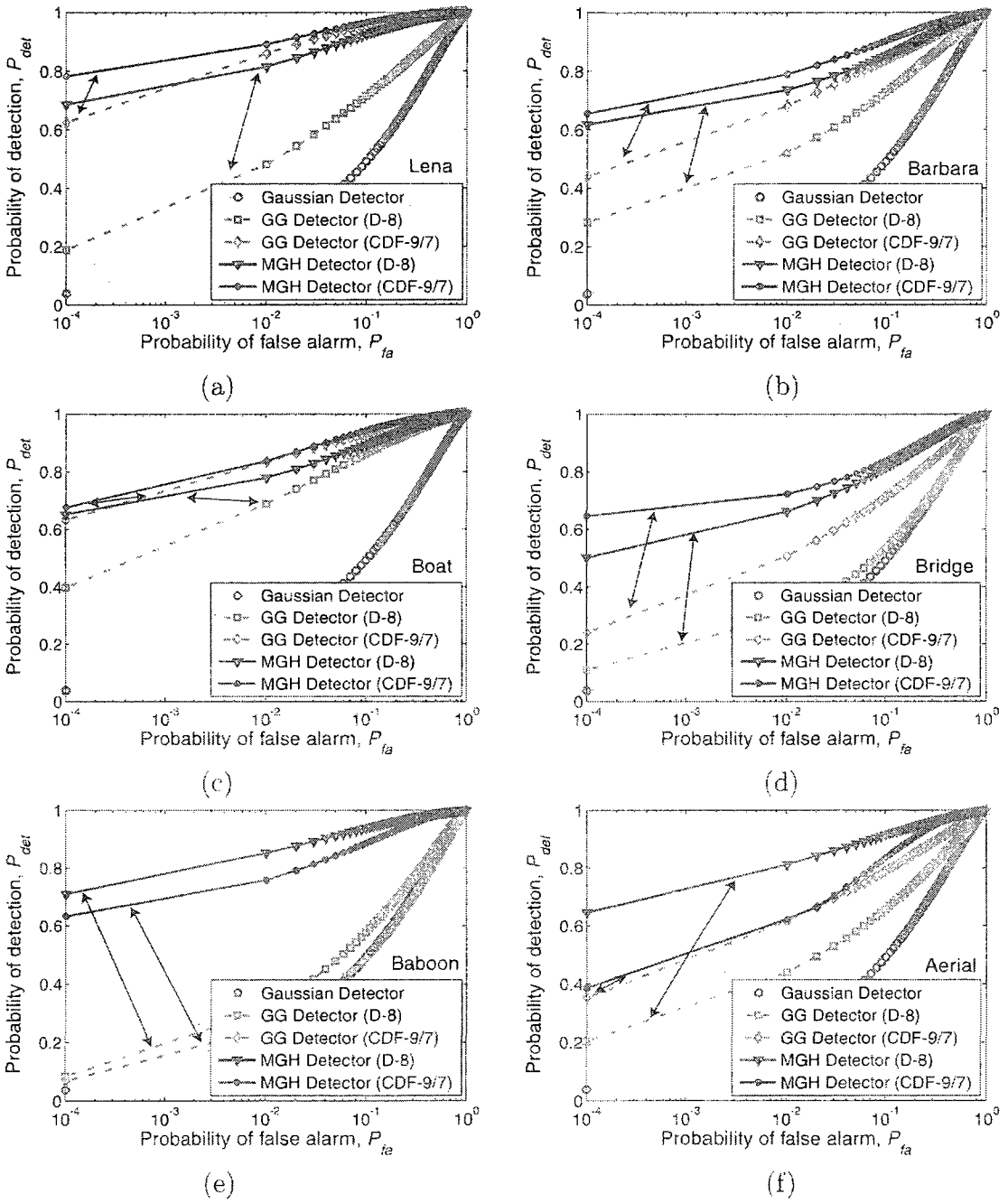


Figure 6.5: ROC of the three statistical detectors, when the images are marked with a WDR of −40 dB. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

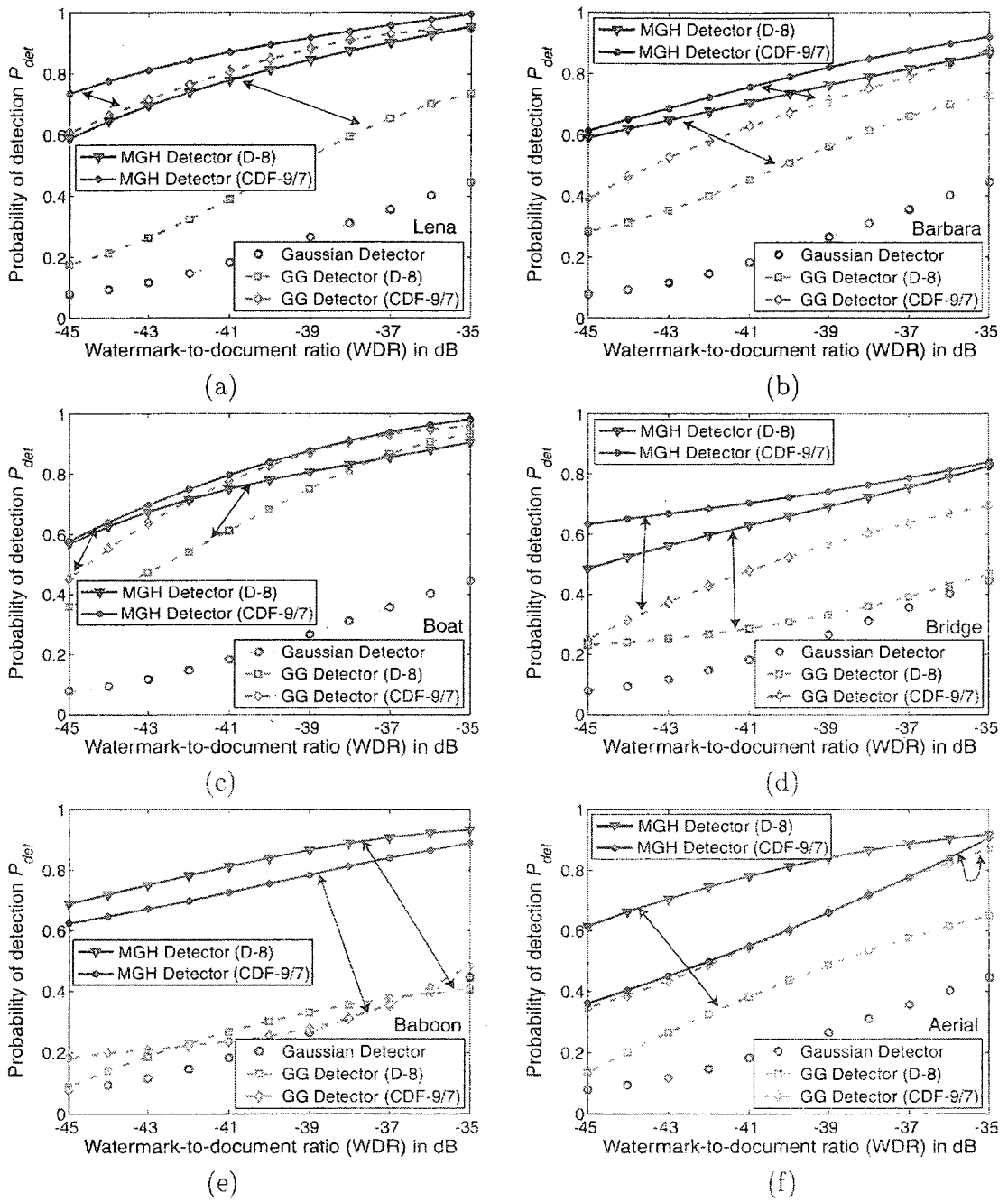


Figure 6.6: Effect on P_{det} of the three statistical detectors for varying strength of watermarks by using 1000 runs and setting $P_{fa} = 10^{-2}$. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

Table 6.2: Comparisons between the target and actual probabilities of false alarm as well as the empirical and actual probabilities of detection of the MGH detector using the subband HL_2 of image *Lena* that is marked with a WDR of -40 dB.

Total Marks	Target P_{fa}	Test Setting Number 1		Test Setting Number 2		
		False Alarms	Actual P_{fa}	Empirical P_{det}	Detected Marks	Actual P_{det}
100	10^{-1}	15	1.5×10^{-1}	0.91	89	0.89
1000	10^{-2}	18	1.8×10^{-2}	0.79	763	0.76
10000	10^{-3}	22	2.2×10^{-3}	0.74	6891	0.69
100000	10^{-4}	28	2.8×10^{-4}	0.68	63427	0.63

To make comparisons between the theoretical and experimental P_{fa}'s and P_{det}'s of the proposed MGH detector, two test settings are considered for a target P_{fa} and a given set of watermarks embedded in a subband. In the first setting, the number of false alarm is estimated by looking for a randomly chosen watermark in the set of subbands of the marked images by comparing the estimated log-likelihood ratio given in (6.5) with the detector thresholds at the target P_{fa}. In the second setting, the detection probability is estimated in a similar way by comparing the log-likelihood ratio of each of the watermarks with the corresponding detector threshold at the target P_{fa}. Table 6.2 shows the results obtained for these two test settings using the subband HL_2 of the image *Lena*. From this table it is evident that the analytical P_{fa}'s and P_{det}'s of the MGH detector approximate very closely the actual ones.

Tables 6.3 and 6.4 show the ED values of the three detectors obtained from the six test images, wherein the DWT employs D-8 and CDF-9/7, respectively, and WDR varies from -40 dB to -35 dB. It is to be pointed out that the ED values of the Gaussian detector are independent of the test image and the wavelet filter employed, in view of the fact that for this detector ED $= \sqrt{N}(10)^{\mathrm{WDR}/20}$, which depends only on the size of the subband and WDR. It is seen from Tables 6.3 and 6.4 that for all the images, the ED increases with the strength of the watermark and further, the MGH detector provides the highest efficacy at any value of WDR. It is to be noted

Table 6.3: ED values of the statistical detectors using the wavelet filter D-8 for various values of WDR.

WDR in dB	-40	-39	-38	-37	-36	-35
	Any Test Image					
Gaussian	1.20	1.35	1.51	1.70	1.90	2.13
	Lena					
GG	2.11	2.34	2.63	2.97	3.30	3.70
MGH	**7.72**	**9.61**	**12.43**	**15.74**	**20.01**	**24.96**
	Barbara					
GG	3.01	3.45	4.01	4.52	5.16	5.96
MGH	**6.57**	**7.45**	**8.76**	**10.07**	**12.07**	**14.98**
	Boat					
GG	3.25	3.74	4.28	4.93	5.63	6.49
MGH	**4.97**	**5.77**	**6.77**	**8.07**	**9.28**	**11.12**
	Bridge					
GG	1.67	1.82	2.11	2.41	2.59	2.92
MGH	**4.32**	**4.96**	**5.65**	**6.39**	**7.44**	**8.89**
	Baboon					
GG	1.33	1.49	1.57	1.64	1.77	1.95
MGH	**8.42**	**9.72**	**11.59**	**13.30**	**15.12**	**17.42**
	Aerial					
GG	2.32	2.85	3.36	3.83	4.27	4.87
MGH	**2.80**	**3.17**	**3.61**	**4.07**	**4.68**	**5.34**

that even in the case of the *Boat* image, the MGH detector performs better than the GG detector in terms of the ED for all values of WDR. Thus, taking into account both the performance measures, namely, P_{det} and ED, the proposed MGH detector is superior to the Gaussian and GG detectors.

6.5.2 Detection Performance With Attacks

As seen from Figs. 6.5 and 6.6 as well as from Tables 6.3 and 6.4, the performance of the Gaussian detector is very much inferior to that of the GG or the proposed one, whether it be in terms of ROC or ED. Hence, we will study the robustness of only the GG and proposed detectors, when the marked images are modified by four standard

Table 6.4: ED values of the statistical detectors using the wavelet filter CDF-9/7 for various values of WDR.

WDR in dB	-40	-39	-38	-37	-36	-35
	Any Test Image					
Gaussian	1.20	1.35	1.51	1.70	1.90	2.13
	Lena					
GG	4.63	4.97	5.42	5.86	6.41	7.00
MGH	**20.21**	**23.73**	**27.24**	**32.05**	**38.66**	**45.88**
	Barbara					
GG	3.34	3.77	4.39	4.95	5.62	6.59
MGH	**5.02**	**5.97**	**7.08**	**8.50**	**10.84**	**14.16**
	Boat					
GG	4.71	5.23	5.89	6.51	7.35	8.35
MGH	**8.04**	**9.40**	**10.67**	**12.52**	**14.23**	**16.26**
	Bridge					
GG	2.33	2.60	2.86	2.99	3.29	3.66
MGH	**8.75**	**11.29**	**14.89**	**18.29**	**21.11**	**25.46**
	Baboon					
GG	1.00	1.19	1.46	1.76	2.09	2.59
MGH	**17.75**	**21.61**	**24.31**	**27.58**	**31.93**	**35.37**
	Aerial					
GG	3.27	3.75	4.27	4.89	5.78	6.65
MGH	**4.88**	**5.16**	**5.48**	**5.86**	**6.27**	**6.78**

attacks, viz., compression, AWGN, nonlinear filtering (e.g., median filtering), and geometric distortion (e.g., rotation) [170], [198]. The results are presented only for the wavelet filter D-8; very similar results have been obtained when CDF-9/7 is used. Since the proposed detector deals with the DWT-based watermarking, for compression attack we use a popular DWT-based lossy compression algorithm, namely, the set partitioning in hierarchical trees (SPIHT) algorithm [199], [200]. The marked images are compressed at two different rates, 0.7 bpp and 0.4 bpp, the latter rate indicating a strength of compression attack that is higher than the former. Figs. 6.7(a)-6.7(f) show the deviations of the ROC of the GG and MGH detectors when the images, marked with a WDR of -40 dB, are compressed at these two rates from their uncompressed

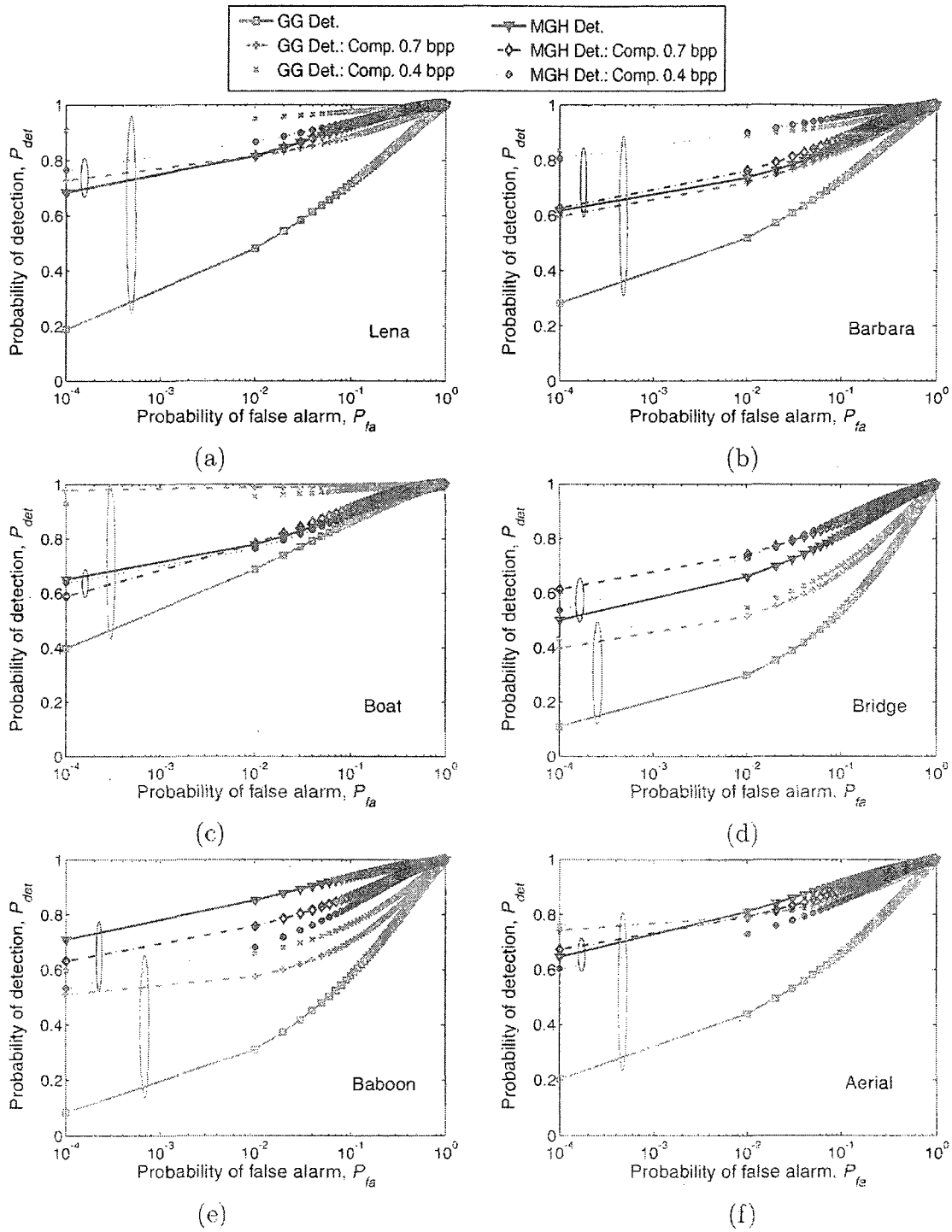


Figure 6.7: Effect of compression on the ROC of the GG and MGH detectors, when the images are marked with a WDR of -40 dB. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

versions. It is seen from these figures that the deviations of the ROC of the MGH detector are much less than that of the GG detector for all the test images. The smaller deviations of the ROC reveal that the MGH detector is more robust than the GG detector in case of a compression attack. An interesting observation that can be made from Fig. 6.7 is that, with an increasing compression attack, the detection probability of the MGH detector increases for the images having sharp details (*Lena* and *Barbara*) and decreases for the images having minute details (*Baboon* and *Aerial*). This may be due to the fact that, for sharp images, the image details are removed significantly than that of the embedded watermark as the strength of the compression attack is increased. In such a case, the effective WDR increases, and hence, the detection probability of the estimated ROC. On the other hand, for the images having minute details, the signals correspond to the embedded watermark are removed significantly than that of the image details. Hence, the effective WDR decreases and consequently, the detection probability of the estimated ROC. However, no such conclusion can be drawn for images having a mix of both sharp and minute details (*Boat* and *Bridge*). To study the robustness of the two detectors when a marked image is attacked by an AWGN, the marked images are corrupted with two noise standard deviations, viz., $\sigma_\varepsilon = 10$ and 20, the latter parameter indicating a higher strength of the AWGN attack. Figs. 6.8(a)-6.8(f) show the deviations of the ROC of the GG and MGH detectors when the images, marked with a WDR of -40 dB, are corrupted by these two noise strengths from their uncorrupted versions. It is seen from Fig. 6.8 that the deviations of the ROC of the proposed MGH detector are significantly less than that of the GG detector for all the test images, and hence, the former is more robust than the latter in case of an AWGN attack. The robustness of the two detectors in the case of nonlinear filtering attack is studied by median filtering the marked images using two squared-shape windows of size 3×3 and 5×5,

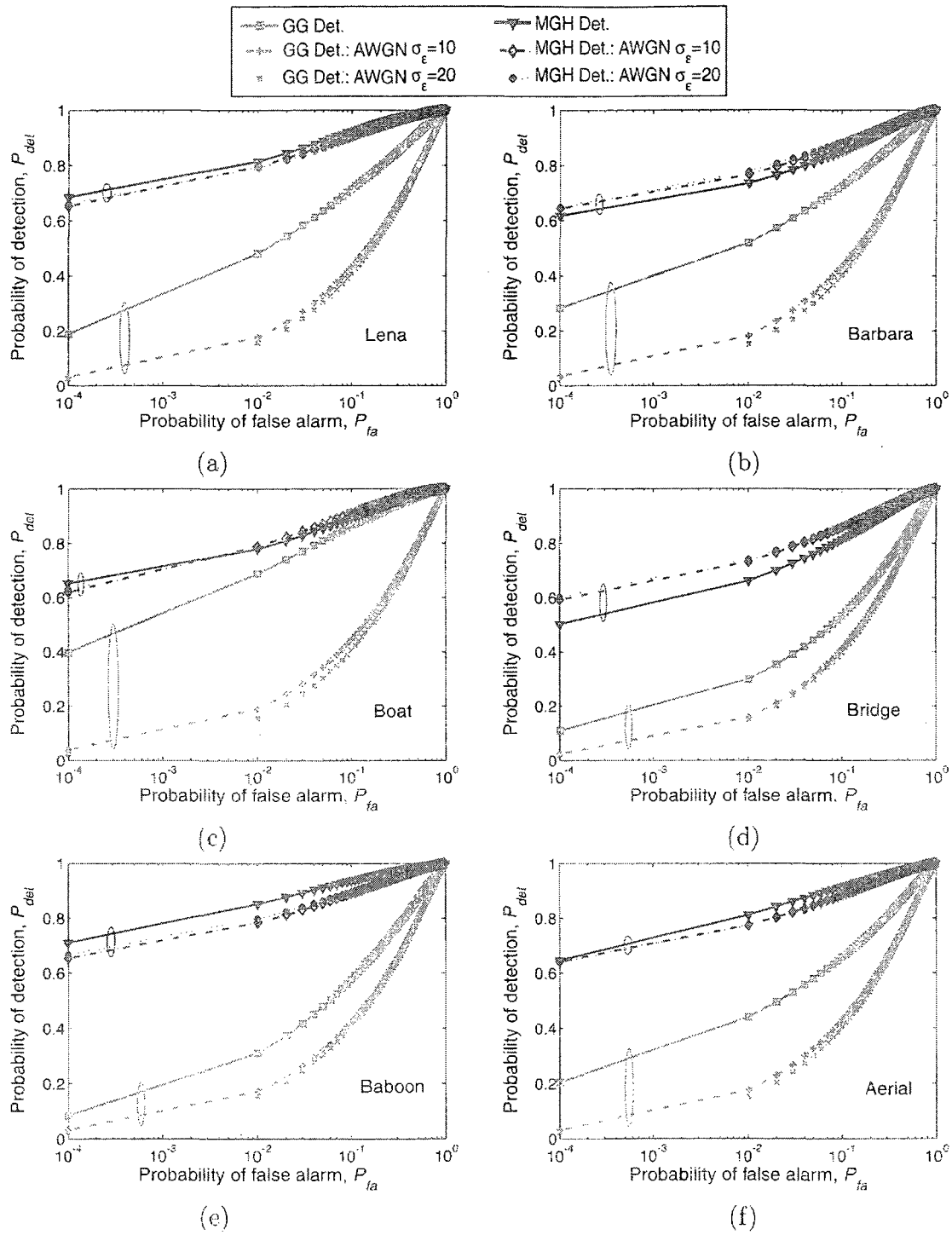


Figure 6.8: Effect of AWGN on the ROC of the GG and MGH detectors, when the images are marked with a WDR of −40 dB. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

the latter size indicating a higher strength of the filtering attack. Figs. 6.9(a)-6.9(f) show the deviations of the ROC of the GG and MGH detectors when the marked images are subjected to these attacks. It is seen from Fig. 6.9 that in most of the cases the deviations of ROC of the MGH detector is the lowest. Thus, in general, the proposed MGH detector is more robust than the GG detector for a median filtering attack. The robustness of these detectors is further studied in the case of a geometric attack by rotating the marked images counter clockwise by two angles, viz., $\theta = 0.25^0$ and $\theta = 1.00^0$, the latter angle indicating a higher strength of this geometric attack. Fig. 6.10 shows the deviations of the ROC of these detectors when the marked images are rotated. It is seen from this figure that the deviations of the ROC of the MGH detector is always lower than that of the GG detector. Hence, the proposed MGH detector is more robust than the GG detector for such a geometric distortion attack. The percentage changes in the ED values of the GG and MGH detectors for the compression, AWGN, median filtering, and rotation attacks considered in the experiments are shown in Table 6.5. From this table it is seen that for all the test images, the change in the ED values of the MGH detector due to these attacks is significantly less than that of the GG detector, showing that the former is more robust than the latter for the compression, AWGN, filtering, and geometric attacks.

6.6 Conclusion

Conventional PDFs of the DWT coefficients of images used in statistical detectors of watermarking schemes have a fixed number of parameters that are estimated from the first few moments. Since these PDFs do not provide a sufficiently good statistical match to the empirical PDF of the image coefficients, the statistical detectors that are designed using such PDFs show inadequate detection performance. In this chapter, a new statistical detector for the DWT-based additive watermarking scheme has been

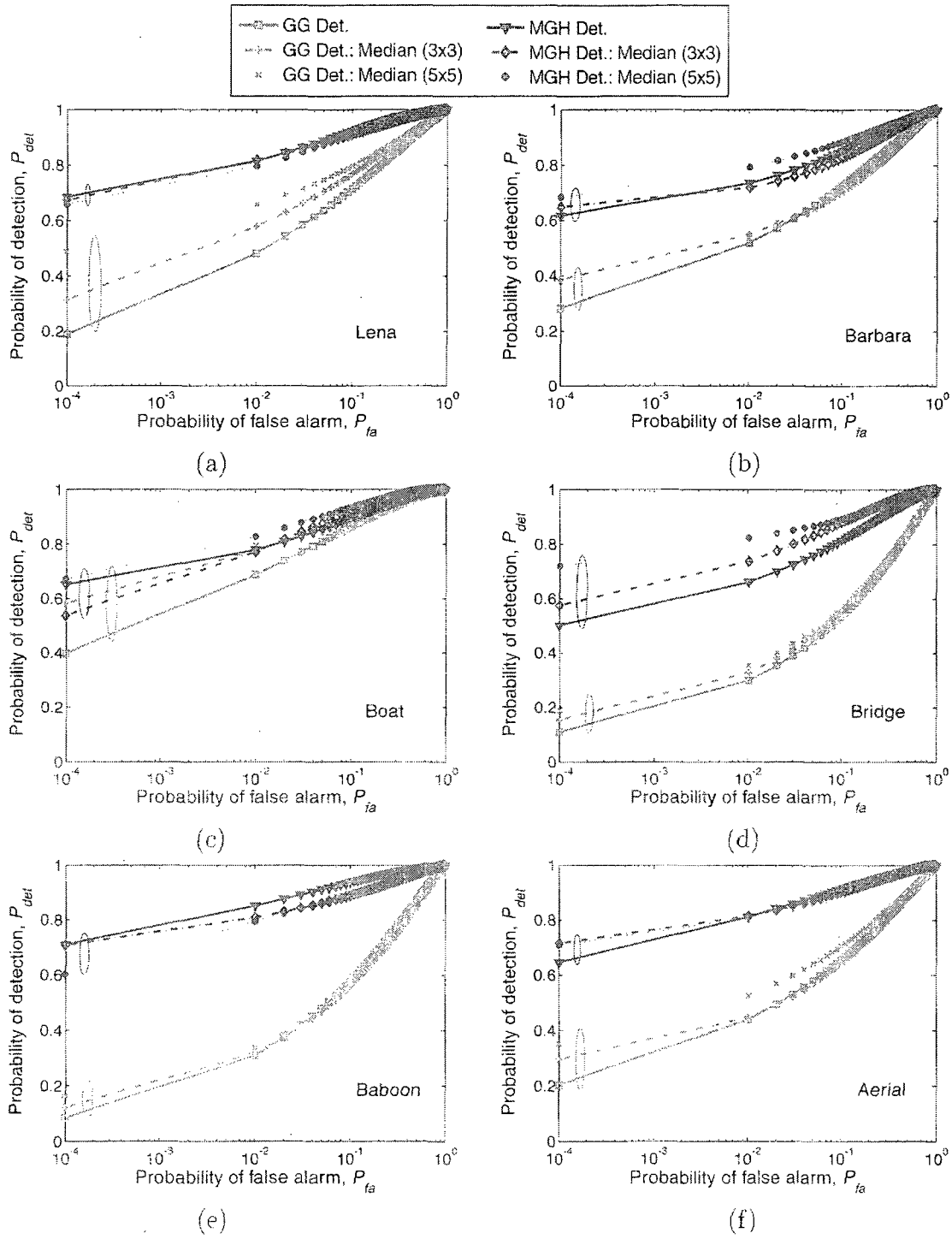


Figure 6.9: Effect of median filtering on the ROC of the GG and MGH detectors, when the images are marked with a WDR of -40 dB. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

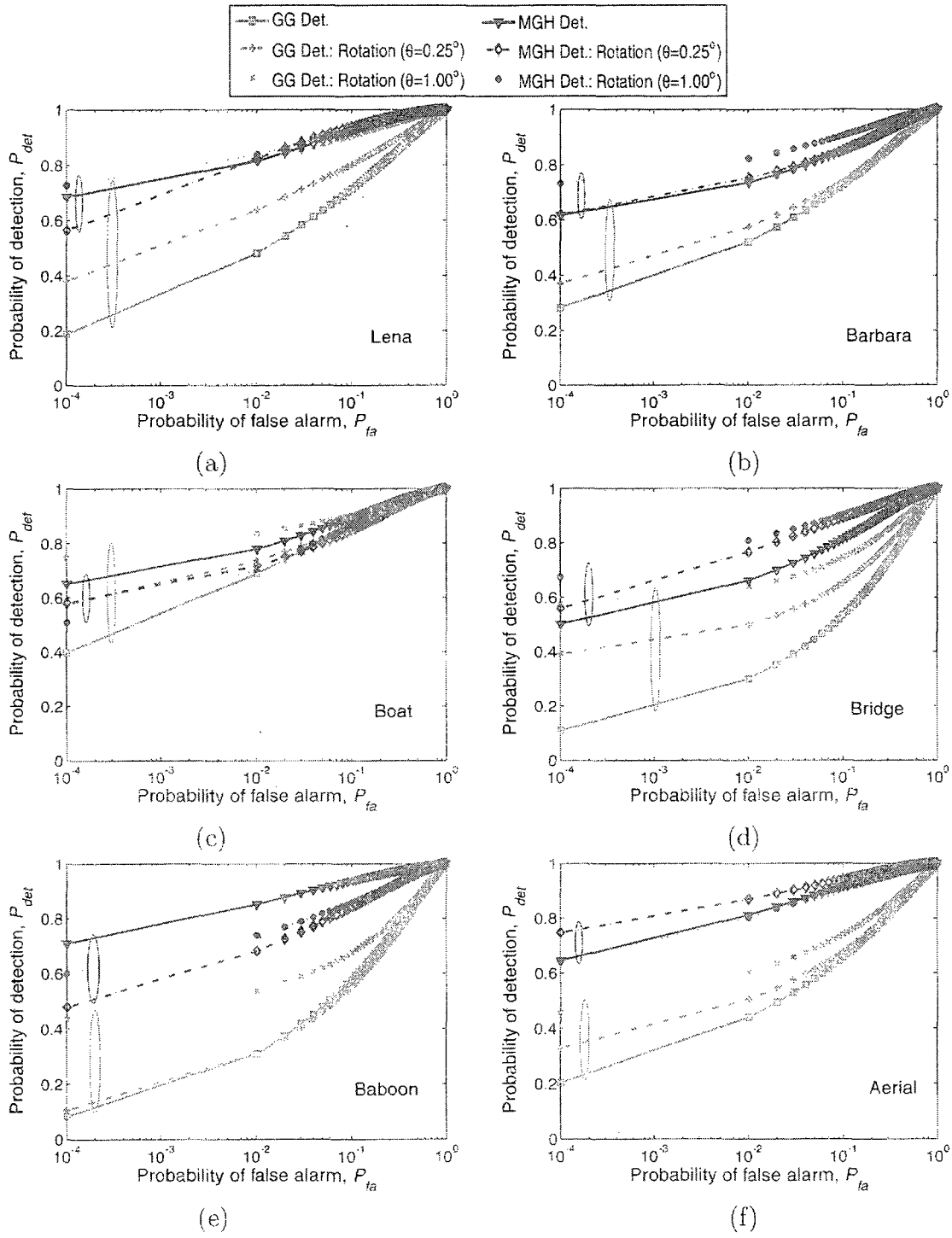


Figure 6.10: Effect of rotation on the ROC of the GG and MGH detectors, when the images are marked with a WDR of −40 dB. Test images are (a) *Lena*, (b) *Barbara*, (c) *Boat*, (d) *Bridge*, (e) *Baboon*, and (f) *Aerial*.

Table 6.5: Percentage change in the ED values of the GG and MGH detectors under the compression, AWGN, median filtering, or rotation attack when the images are marked with a WDR of -40 dB.

Attack	Compression		AWGN		Median Filtering		Rotation	
	0.7 bpp	0.4 bpp	$\sigma_\epsilon = 10$	$\sigma_\epsilon = 20$	3×3	5×5	$\theta = 0.25^0$	$\theta = 1.00^0$
	Lena							
GG	517.0	802.7	65.1	72.1	34.3	83.9	54.2	701.1
MGH	**11.4**	**13.6**	**21.6**	**19.2**	**30.1**	**2.7**	**46.8**	**23.6**
	Barbara							
GG	653.7	829.8	74.5	81.8	27.0	40.9	25.0	580.6
MGH	**14.9**	**64.5**	**2.2**	**3.0**	**9.0**	**23.1**	**16.6**	**14.3**
	Boat							
GG	749.7	790.5	73.7	82.7	46.0	112.6	53.9	422.1
MGH	**8.5**	**57.2**	**28.3**	**39.7**	**34.5**	**107.0**	**31.9**	**12.3**
	Bridge							
GG	178.7	206.4	63.5	73.7	**4.3**	**20.6**	209.2	1497.4
MGH	**27.6**	**29.5**	**16.5**	**28.0**	14.0	26.2	**59.8**	**16.4**
	Baboon							
GG	793.9	1160.7	49.7	60.0	**27.8**	**74.3**	17.6	1021.5
MGH	**39.4**	**31.1**	**28.4**	**27.1**	21.1	69.6	**16.6**	**14.8**
	Aerial							
GG	269.4	295.6	70.5	78.9	25.6	82.2	64.6	794.1
MGH	**61.7**	**68.3**	**42.0**	**52.9**	**19.7**	**12.4**	**60.2**	**121.4**

developed using the proposed MGH PDF, in view of the fact that this PDF provides a better statistical match to the empirical PDF by utilizing an appropriate number of parameters estimated from higher-order moments of the image coefficients. The decision threshold and the ROC of the statistical watermark detector have been derived using the MGH PDF. The experimental results on test images have shown that the detection performance of the proposed MGH detector is superior to that of the Gaussian or GG detector, in terms of the probabilities of detection and false alarm as well as the efficacy. The proposed detector has been shown to be more robust than the GG detector under the attacks of compression, AWGN, filtering, and geometric distortion.

Chapter 7

Proposed PDF of Wavelet Coefficients for Video

7.1 Introduction

An appropriate PDF for the DWT coefficients of a video signal, just as in the case of an image signal, is essential for almost all statistical video processing algorithms. The DWT of a video signal may be represented by one of the following two ways:

- A group of frames decomposed by 3D DWT.
- Individual frames decomposed by 2D DWT.

Video processing based on the 3D DWT avoids the procedure of motion estimation or detection. However, the main drawbacks of the 3D DWT are its inability to make use of the asymmetry of the space and time resolutions that exist in a video, and its long-time latency in view of the memory requirement being constrained by the length of the wavelet filter coefficients [201]. Hence, we would like to use the 2D DWT for video signal processing, wherein the estimation of the motion between frames plays a significant role. The estimation of the motion vector in the data domain is an ill-posed problem [202] that gets worse in the DWT domain due to the shift- and rotation-variance properties of the transform [203]. Hence, an indirect motion estimation during a DWT-based statistical processing of a video signal is preferable.

In other words, capturing the motion information through an appropriate PDF of the DWT coefficients of the video is essential. In this chapter, the bivariate Gaussian PDF is first proposed as a joint PDF of the DWT coefficients of two neighboring frames of a video signal, wherein the correlation parameter of the Gaussian PDF is used as an index of motion between two neighboring frames [87], [138]. Motivated from the excellent probabilistic modeling performance of the univariate MGH PDF, developed in Chapter 3, for the DWT coefficients of an image signal, the bivariate MGH PDF is also developed as a more appropriate joint PDF of the DWT coefficients of two neighboring frames. Similar to the univariate MGH PDF, bivariate MGH PDF allows the incorporation of an appropriate number of parameters using the higher-order moments of the DWT coefficients of the video signal. Experiments are carried out on a number of test video sequences using the proposed bivariate PDFs to show the effectiveness of the PDFs in probabilistic modeling of the DWT coefficients.

7.2 Joint PDF of Inter-Frame DWT Coefficients

Let $f_{k(ij)}$ represent the DWT coefficients for a given subband of the k-th frame at a spatial index (i, j). The wavelet coefficients of the previous neighboring frames are denoted as $f_{l(ij)}$, $l = k-1, k-2, \cdots, k-\mathcal{L}$, $\mathcal{L}$ being the total number of the neighboring frames under consideration. Hereafter, the k-th frame is referred to as the *current frame* and the l-th frame as a *previous frame*. Let $f_{k(ij)}$ and $f_{l(ij)}$ be the samples of two identical and zero-mean random processes $\boldsymbol{f}_c$ and $\boldsymbol{f}_p$, respectively. The correlation coefficient ρ ($-1 < \rho < 1$) between the two random processes is defined by

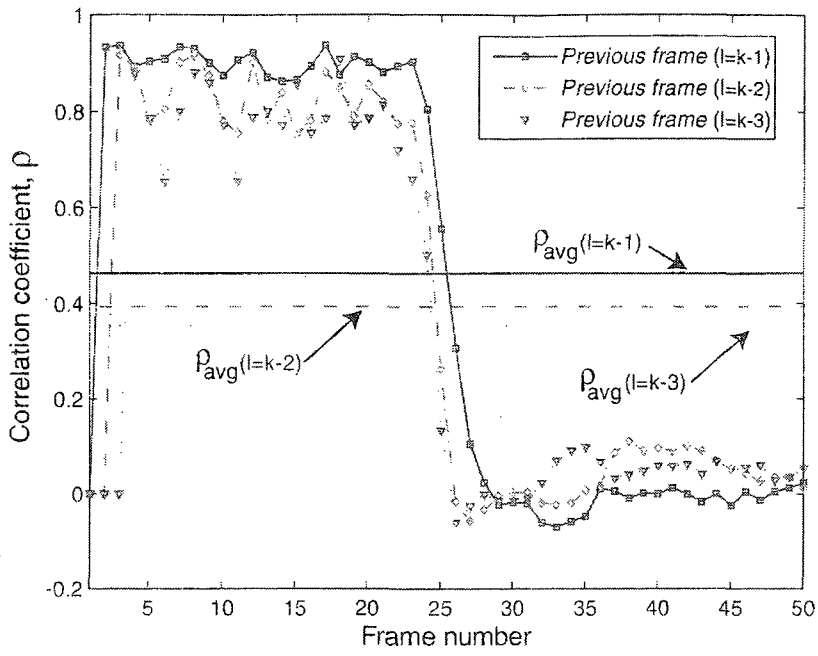
$$\rho = \frac{E\{\boldsymbol{f}_c \boldsymbol{f}_p\} - E\{\boldsymbol{f}_c\}E\{\boldsymbol{f}_p\}}{\sqrt{\left[E\{\boldsymbol{f}_c^2\} - E\{\boldsymbol{f}_c\}^2\right]\left[E\{\boldsymbol{f}_p^2\} - E\{\boldsymbol{f}_p\}^2\right]}} \tag{7.1}$$

The value of ρ represents the linear relationship between the two random processes. Therefore, for a given DWT coefficient of a video, the amount of motion that exists

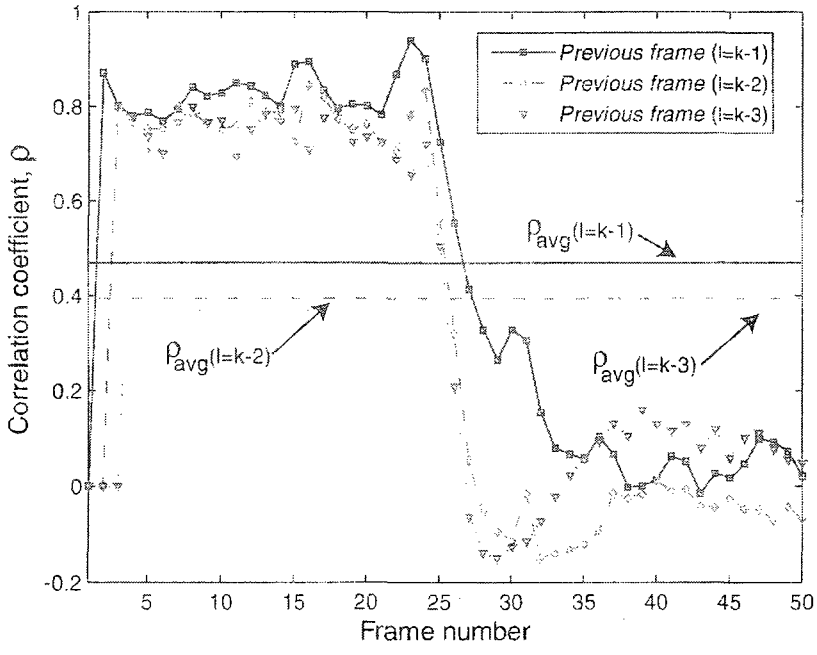
between any two frames can be indirectly measured by ρ. Since the estimation of the motion trajectory of any DWT coefficient is not robust [203], it is preferable to use the parameter ρ as an index of the motion of a coefficient. The higher the value of ρ, the lower the amount of motion between the coefficients of the two neighboring frames, and vice versa. Fig. 7.1 shows the values of the correlation coefficients of a given decomposition level between the *current frame* and any one of the three *previous frames* for the first 50 frames of the typical video sequence, *Tennis*. From this figure, it can be observed that in the first 25 frames there is very little motion, whereas for the rest of the frames, there is considerable motion between the neighboring frames. It can also be noticed that the average value of the proposed motion index ρ_{avg} decreases gradually when the relative motions between the *current frame* k and the *previous frames*, namely, $k-1$, $k-2$, and $k-3$, are considered. The effect of motion on the empirical joint PDF, denoted as $p^{\text{emp}}_{\boldsymbol{f}_c \boldsymbol{f}_p}(\boldsymbol{f}_c, \boldsymbol{f}_p)$, for the HL_1 band DWT coefficients is shown in Fig. 7.2. It can be seen that when the motion between the *current-frame* and *previous-frame* is low, the contour of the joint PDF is nearly elliptic, whereas the contour is nearly circular when the motion is relatively high. The bivariate Gaussian PDF is elliptic or circular in nature provided the dependency between two random processes is strong or weak [204]. Therefore, the joint PDF of the wavelet coefficients for the *current frame* and any of the *previous frames* may be expressed as the bivariate Gaussian (BVG) PDF given by

$$p_{\boldsymbol{f}_c \boldsymbol{f}_p}(\boldsymbol{f}_c, \boldsymbol{f}_p) = \frac{1}{2\pi\sigma_c\sigma_p\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{\boldsymbol{f}_c^2}{\sigma_c^2} + \frac{\boldsymbol{f}_p^2}{\sigma_p^2} - 2\rho\frac{\boldsymbol{f}_c}{\sigma_c}\frac{\boldsymbol{f}_p}{\sigma_p}\right\}\right] \quad -1 < \rho < 1 \tag{7.2}$$

where σ_c and σ_p are the standard deviations of the wavelet coefficients of the *current frame* and the corresponding *previous frame*, respectively. Fig. 7.3 shows the BVG PDFs to model the empirical PDFs that are shown in Fig. 7.2. It may be seen from this figure that the BVG PDFs can provide a satisfactory approximation



(a)



(b)

Figure 7.1: Motion index ρ for the DWT coefficients of two neighboring frames of the video sequence *Tennis*: (a) for the first level decomposition and (b) for the second level decomposition. Note: Here we plot the mean value of ρ obtained from the HL, LH, and HH subbands.

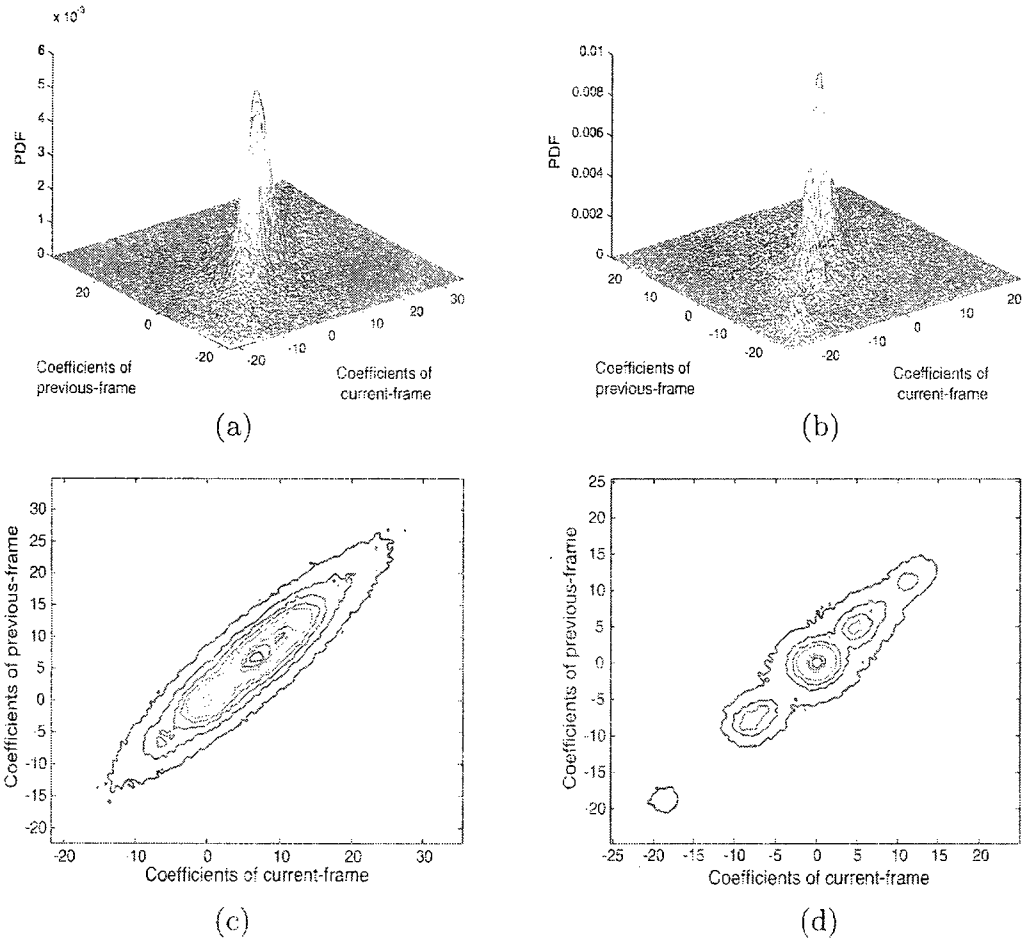


Figure 7.2: Joint PDF of *current frame* and *previous frame* for the HL_1 band DWT coefficients of the video sequence *Tennis*. (a) Empirical PDF obtained from frames 1 to 25. (b) Empirical PDF obtained from frames 26 to 50. (c) Contour of (a). (d) Contour of (b)

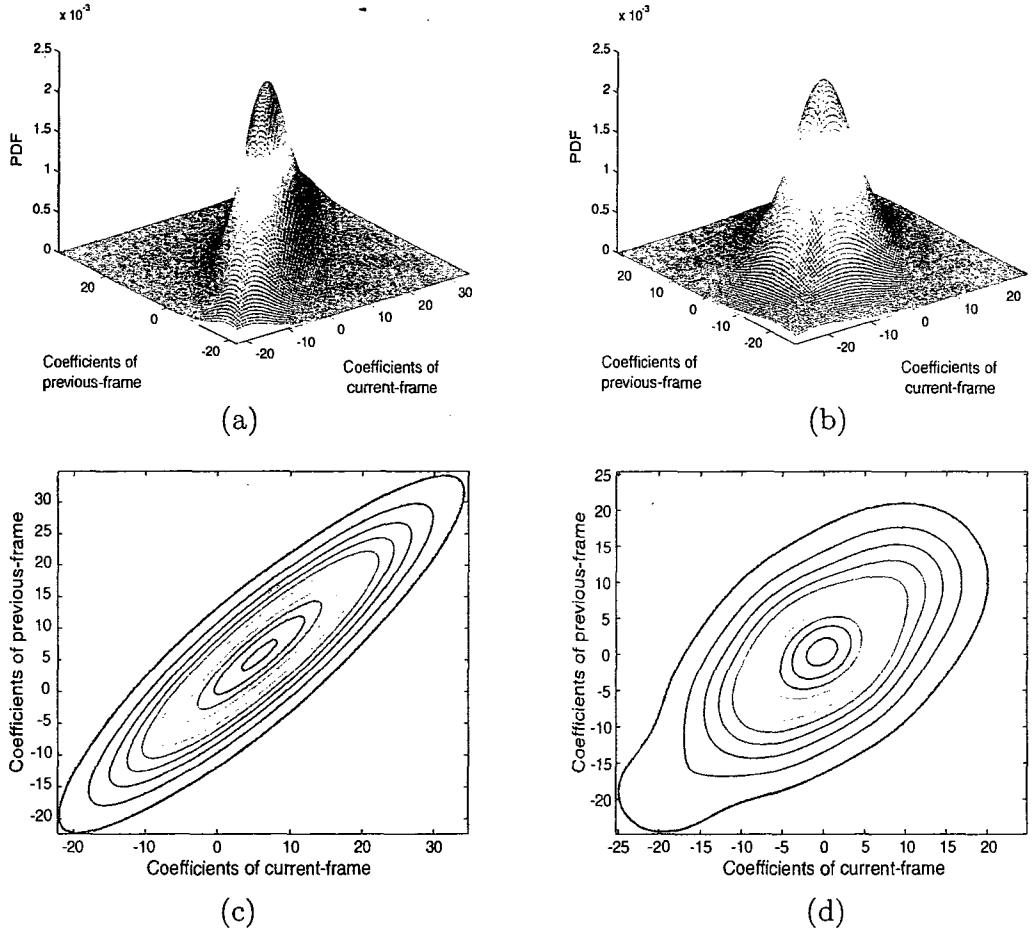


Figure 7.3: Joint PDF of *current frame* and *previous frame* for the HL_1 band DWT coefficients of the video sequence *Tennis*. (a) Bivariate Gaussian PDF obtained from frames 1 to 25. (b) Bivariate Gaussian PDF obtained from frames 26 to 50. (c) Contour of (a). (d) Contour of (b)

of the empirical PDFs that are shown in Fig. 7.2. In view of the mathematical tractability of the BVG prior in MAP estimation techniques, two locally-adaptive MAP-based video denoising algorithms are developed in Chapter 8 using the BVG PDF.

In Chapter 3, the univariate MGH PDF that provides a superior performance in modeling the wavelet coefficients of an image signal was developed. A question naturally arises as to whether one could develop a corresponding bivariate form of the MGH PDF for modeling the wavelet coefficients of two neighboring frames of a video. In order to investigate this query, the bivariate form of the GH PDF is developed and this PDF is expressed as (see Appendix C)

$$p_{f_c f_p}(f_c, f_p) = \frac{1}{2\pi\sigma_c\sigma_p} e^{-\frac{1}{2}\left(\frac{f_c^2}{\sigma_c^2}+\frac{f_p^2}{\sigma_p^2}\right)} \left[1 + \sum_{r=1}^{\infty} \left\{ \beta_{rc} H_r\left(\frac{f_c}{\sigma_c}\right) + \beta_{rp} H_r\left(\frac{f_p}{\sigma_p}\right) + \frac{\rho^r}{r!} H_r\left(\frac{f_c}{\sigma_c}\right) H_r\left(\frac{f_p}{\sigma_p}\right) \right\}\right] \tag{7.3}$$

where $\{\beta_{rc}, \beta_{rp}\}$ are the series parameters and $\{\sigma_c^2, \sigma_p^2\}$ are the variances of the *current* and *previous* frames. The joint PDF given in (7.3) is an infinite series expansion, however, only the first few terms may be used for a reasonably good approximation. Let the maximum value of the order r in the truncated series be ζ. Fig. 7.4 shows the expanded bivariate PDFs obtained by using $\zeta = 12$ to model the empirical PDFs that are shown in Fig. 7.2. It may be seen from this figure that the bivariate GH PDF with a sufficiently large number of series terms approximates well the peak region of the joint PDF. On the other hand, similar to the univariate GH PDF, the tail region of the expanded bivariate PDF is not ensured to be non-negative. Let $\mathcal{P}(f_c, f_p)$ represent the set of coefficients of the bivariate GH PDF which are bounded by the zeros closest to the central mode. A modification is necessary for the bivariate GH PDF to model those coefficients that are in the complementary set of $\mathcal{P}(f_c, f_p)$, i.e., $\overline{\mathcal{P}}(f_c, f_p)$. Let the coefficients of $\overline{\mathcal{P}}(f_c, f_p)$ be modelled by the BVG PDF, which is

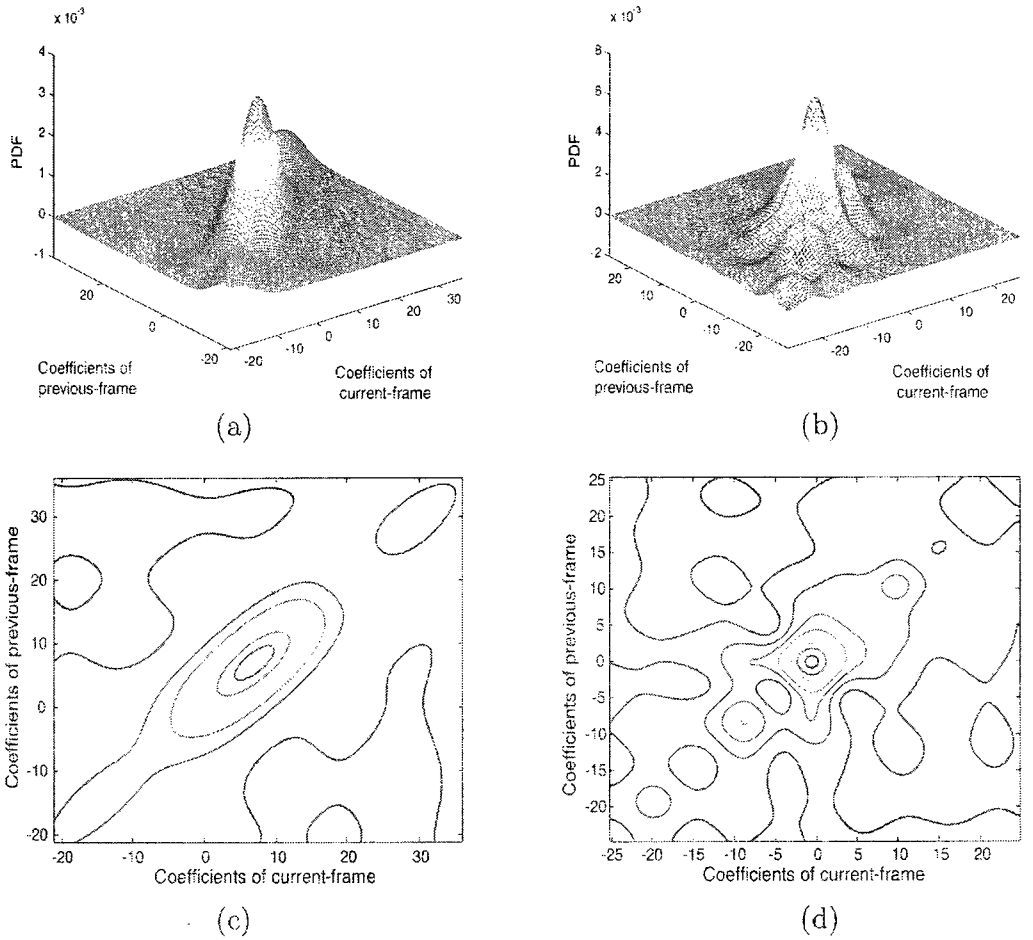


Figure 7.4: Joint PDF of *current frame* and *previous frame* for the HL_1 band DWT coefficients of the video sequence *Tennis*. (a) Bivariate GH PDF obtained from frames 1 to 25. (b) Bivariate GH PDF obtained from frames 26 to 50. (c) Contour of (a). (d) Contour of (b).

ensured to be non-negative, multiplied by the weight parameter γ $(0 < \gamma < 1)$ to ensure that the total volume under the surface of the bivariate modified GH PDF is unity. Finally, the bivariate MGH (BVMGH) PDF can be obtained as

$$p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)=\begin{cases}\dfrac{1}{2\pi\sigma_c\sigma_p}e^{-\frac{1}{2}\left\{\frac{\boldsymbol{f}_c^2}{\sigma_c^2}+\frac{\boldsymbol{f}_p^2}{\sigma_p^2}\right\}}\left[1+\sum\limits_{r=1}^{\varsigma}\left\{\beta_{rc}H_r\left(\dfrac{\boldsymbol{f}_c}{\sigma_c}\right)+\beta_{rp}H_r\left(\dfrac{\boldsymbol{f}_p}{\sigma_p}\right)\right.\right. & \\ \left.\left.+\dfrac{\rho^r}{r!}H_r\left(\dfrac{\boldsymbol{f}_c}{\sigma_c}\right)H_r\left(\dfrac{\boldsymbol{f}_p}{\sigma_p}\right)\right\}\right] & \text{if } (\boldsymbol{f}_c,\boldsymbol{f}_p)\in\mathcal{P}(\boldsymbol{f}_c,\boldsymbol{f}_p) \\ \dfrac{\gamma}{2\pi\sigma_c\sigma_p\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{\boldsymbol{f}_c^2}{\sigma_c^2}+\frac{\boldsymbol{f}_p^2}{\sigma_p^2}-2\rho\frac{\boldsymbol{f}_c}{\sigma_c}\frac{\boldsymbol{f}_p}{\sigma_p}\right\}} & \text{if } (\boldsymbol{f}_c,\boldsymbol{f}_p)\notin\mathcal{P}(\boldsymbol{f}_c,\boldsymbol{f}_p)\end{cases} \tag{7.4}$$

The value of the parameter γ can be numerically calculated from the following constraint

$$\frac{1}{2\pi\sigma_c\sigma_p}\iint_{\mathcal{P}}e^{-\frac{1}{2}\left\{\frac{\boldsymbol{f}_c^2}{\sigma_c^2}+\frac{\boldsymbol{f}_p^2}{\sigma_p^2}\right\}}\left[1+\sum_{r=1}^{\varsigma}\left\{\beta_{rc}H_r\left(\frac{\boldsymbol{f}_c}{\sigma_c}\right)+\beta_{rp}H_r\left(\frac{\boldsymbol{f}_p}{\sigma_p}\right)+\frac{\rho^r}{r!}H_r\left(\frac{\boldsymbol{f}_c}{\sigma_c}\right)\right.\right.$$
$$\left.\left.H_r\left(\frac{\boldsymbol{f}_p}{\sigma_p}\right)\right\}\right]\mathrm{d}\boldsymbol{f}_c\mathrm{d}\boldsymbol{f}_p+\frac{\gamma}{2\pi\sigma_c\sigma_p\sqrt{1-\rho^2}}\iint_{\overline{\mathcal{P}}}e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{\boldsymbol{f}_c^2}{\sigma_c^2}+\frac{\boldsymbol{f}_p^2}{\sigma_p^2}-2\rho\frac{\boldsymbol{f}_c}{\sigma_c}\frac{\boldsymbol{f}_p}{\sigma_p}\right\}}\mathrm{d}\boldsymbol{f}_c\mathrm{d}\boldsymbol{f}_p=1 \tag{7.5}$$

Fig. 7.5 shows the proposed BVMGH PDFs to model empirical PDFs shown in Fig. 7.2. It can be seen from this figure that the proposed joint PDF is non-negative, matches very well the empirical PDF, and captures the multi-modal nature of the empirical PDF. In order to quantify the performance of the closeness of fit of the proposed bivariate PDF with the empirical one, we use KSD and KLD as the metrics, where they are defined by

$$\mathrm{KSD}=\max_{\boldsymbol{f}_c,\boldsymbol{f}_p}\left|\iint\left[p^{\mathrm{emp}}_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)-p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)\right]\mathrm{d}\boldsymbol{f}_c\mathrm{d}\boldsymbol{f}_p\right| \tag{7.6}$$

and

$$\mathrm{KLD}=\iint p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)\ln\frac{p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)}{p^{\mathrm{emp}}_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)}\mathrm{d}\boldsymbol{f}_c\mathrm{d}\boldsymbol{f}_p \tag{7.7}$$

Table 7.1 shows the results concerning the metrics KSD and KLD for the BVG and BVMGH PDFs of the first level DWT coefficients of six video sequences, namely,

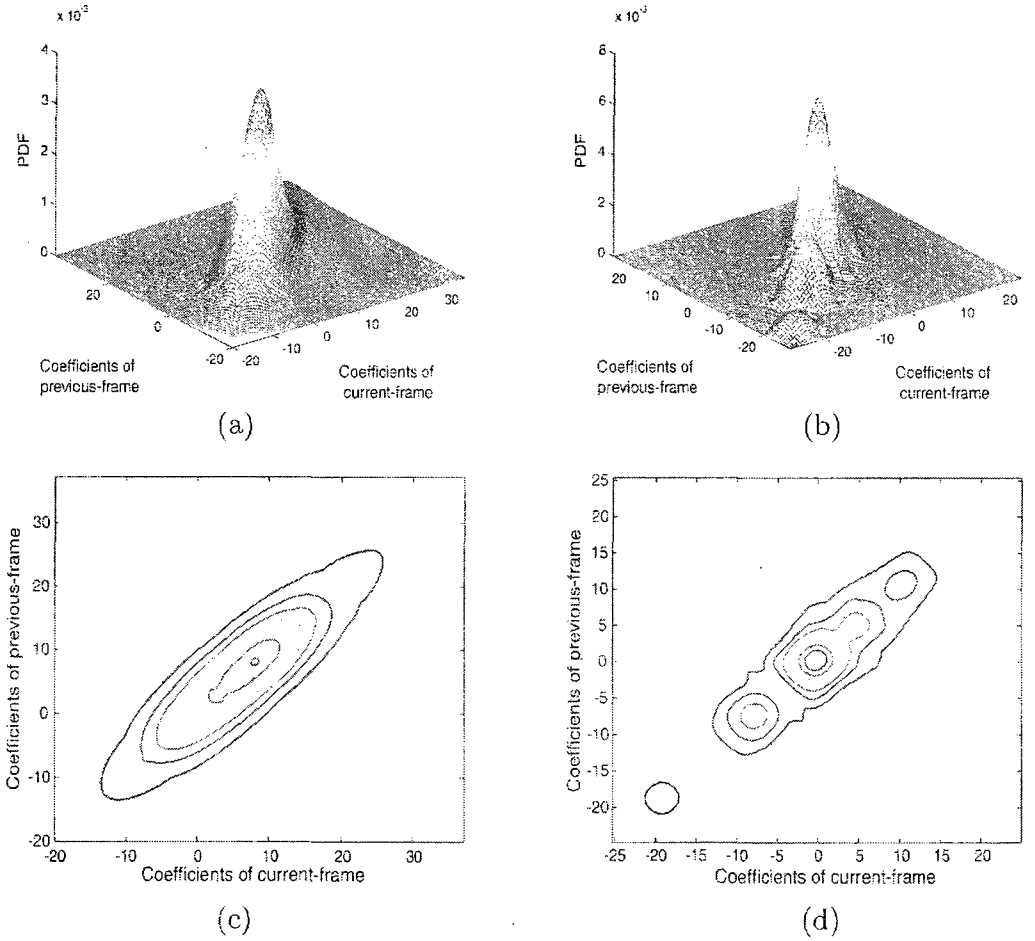


Figure 7.5: Joint PDF of *current frame* and *previous frame* for the HL_1 band DWT coefficients of the video sequence *Tennis*. (a) Bivariate MGH PDF obtained from frames 1 to 25. (b) Bivariate MGH PDF obtained from frames 26 to 50. (c) Contour of (a). (d) Contour of (b).

Table 7.1: Values of the metrics KSD and KLD for the prior function modeling of the DWT coefficients of video sequences. The average values of KLD (standard deviations in parentheses) and KSD are obtained from a first 100 frames of a sequence.

Prior Model	KSD ($\times 10^{-3}$)			KLD		
	HL_1	LH_1	HH_1	HL_1	LH_1	HH_1
	Tennis					
BVG PDF	11.2	17.7	22.8	6.74 (3.01)	3.34 (1.02)	2.75 (1.24)
BVMGH PDF	**10.3**	**14.3**	**13.1**	**5.48 (2.53)**	**2.24 (1.01)**	**1.54 (1.03)**
	Salesman					
BVG PDF	15.3	30.4	**57.0**	2.27 (0.17)	2.52 (0.44)	**7.34 (0.45)**
BVMGH PDF	**11.1**	**27.4**	63.9	**1.38 (0.06)**	**2.07 (0.20)**	7.43 (0.48)
	Coastguard					
BVG PDF	23.2	13.4	99.1	3.70 (1.31)	5.26 (0.92)	4.25 **(1.09)**
BVMGH PDF	**20.3**	**7.2**	**85.7**	**1.88 (0.50)**	**1.33 (0.22)**	**4.22** (1.23)
	Football					
BVG PDF	6.9	8.9	19.5	4.66 (0.99)	5.17 (1.09)	2.87 (0.77)
BVMGH PDF	**5.0**	**6.3**	**8.1**	**3.06 (0.62)**	**3.39 (0.87)**	**0.90 (0.20)**
	Foreman					
BVG PDF	36.9	67.1	113.0	3.81 (1.00)	1.99 (0.62)	4.36 (0.75)
BVMGH PDF	**33.2**	**51.6**	**67.7**	**2.39 (0.51)**	**0.79 (0.20)**	**2.43 (0.66)**
	Garden					
BVG PDF	11.9	28.9	38.7	6.53 (0.37)	4.74 (0.37)	4.68 (0.60)
BVMGH PDF	**11.6**	**27.5**	**36.0**	**5.66 (0.28)**	**3.29 (0.17)**	**2.43 (0.14)**

Tennis, *Salesman*, *Coastguard*, *Football*, *Foreman*, and *Garden*. The metrics are calculated by averaging the results obtained using the first 100 frames of each sequence. Standard deviations of the KLD metric are also included in the table. For the results shown here, the BVMGH PDF uses the moments from 2 up to 12, whereas the BVG PDF can use only the second order moments. From the table, it is evident that in most of the cases, the BVMGH PDF has the lowest distance and divergence from the empirical one, compared to that of the BVG PDF. The closeness of fit between the empirical and BVMGH PDFs is mainly due to the fact that the proposed PDF can incorporate a larger number of higher-order moments in its parameter estimation process. This feature, along with the modification proposed through the introduction of the parameter γ, results in the proposed BVMGH PDF to match the empirical

PDF very well both in the peak and tail segments. Thus, the BVMGH prior function would be very effective for DWT-based video signal processing.

7.3 Conclusion

In this chapter, the correlation between the DWT coefficients of any two neighboring frames of a video has been proposed to be an indirect measure of the motion that exists between these frames. First, the BVG PDF has been considered to be a joint PDF of the inter-frame DWT coefficients of a video signal. Due to the mathematical tractability of the BVG PDF, locally-adaptive MAP-based denoising algorithms are developed using this PDF as prior in Chapter 8. In view of the fact that the incorporation of the higher-order moments of data samples in the parameter estimation provides a better probabilistic model, as seen in the development of the univariate MGH PDF, the BVMGH PDF has been developed as a joint PDF of the inter-frame DWT coefficients of a video signal. The modeling performance of both the proposed joint PDFs, namely, the BVG and BVMGH, are tested on a number of test video sequences. Simulation results show that these PDFs can provide an approximation of the empirical PDF, however, the closeness of fit of the latter is better than that of the former.

Chapter 8

Wavelet-Based Video Denoising Using a Joint PDF

8.1 Introduction

In this chapter, two wavelet-based video denoising techniques are developed by using a joint PDF of the wavelet coefficients of two neighboring frames. The denoising techniques use the correlation parameter of the joint PDF as an index of motion. Due to mathematical tractability, the BVG PDF proposed in Chapter 7 is employed as the joint PDF in the development of bivariate MAP estimators that are used for spatial filtering of the noisy decimated DWT coefficients [138], as well as that of the noisy magnitude components of the DT-CWT coefficients [87]. A recursive time averaging of the spatially filtered coefficients and components is adopted for further noise reduction. Experiments are carried out on a number of test video sequences and the results show that the proposed denoising algorithms provide improved performance both in terms of the PSNR and the perceptual quality as compared to that provided by the other denoising algorithms.

8.2 Video Denoising and DWT

Video processing is becoming increasingly important with the rapid growth of multimedia technology. Among the various video processing tasks, the reduction of noise

is an important one, since distortion of a video is inevitable during its acquisition, processing, storage, transmission, and reproduction [205]. A noisy video has not only an unpleasant visual effect, but it also restricts the user to extract the true content. The aim of video denoising is to improve the perceptual quality by removing the noise, while preserving the signal features as much as possible. The reduction of noise in a video is also a pre-processing step for various applications such as compression, enhancement, and target recognition. Thus, the development of an advanced video denoising scheme is essential.

There exists a high correlation among the neighboring frames of a video, since the motions among such frames are small. Hence, video denoising techniques can be considered as extensions of image denoising techniques, by providing temporal filtering taking into account the correlation between the neighboring frames. In order to reduce the computational effort, most of the video noise reduction techniques use the spatio-temporal filters in the pixel-domain rather than in the transform-domain [202], [203], [206]-[213]. A 3D window along the estimated motion trajectory is generally used for the processing of a pixel value, since one can assume the stationarity of the pixel values along that trajectory. Some examples include the adaptive weighted averaging filter [206], the adaptive recursive least square filter [207], and the motion compensated Kalman filter [208]. However, the estimation of the motion trajectory in the data domain is an ill-posed problem that gets aggravated further in the presence of noise [202], and hence, an indirect motion estimation is preferable. For example, the 3D rational filter proposed in [209], indirectly incorporates the motion information of a video in some threshold parameters, the values of which are chosen in an *ad hoc* manner. Spatio-temporal filter based on the order-statistics [214] is another approach for denoising. One such filter is the K-nearest neighbor (KNN) filter proposed in [210]. The filter orders the pixel values of a 3D local neighborhood and

averages the clustered ones. Although the pixel-domain spatio-temporal filters provide a slight increase in the computational speed compared to the transform-domain ones, they usually cannot suppress the noise sufficiently without the edges of the video signal being smeared.

In signal denoising, the DWT has shown considerable success over the last decade. There are two broad approaches for the DWT-based spatio-temporal filtering for video denoising. First approach is the thresholding of the coefficients of the 3D DWT of a noisy video [215]-[217]. The second and more appropriate approach is to represent each frame by the 2D DWT and perform spatial filtering [12], [65], [72], [106], [119], [126], [218]: this is followed by temporal filtering by taking into account the strong temporal correlation that exists in a video. Due to the shift-invariance property, the non-decimated DWT [45] improves the video denoising performance approximately by 1dB as compared to that of the decimated representation [201], [219], [220]. In order that the video denoising algorithms have a reduced computational load as well as denoised frames have a better directional features, the DT-CWT may be used instead of the non-decimated DWT. Nevertheless, the decimated DWT-based algorithms show the least computational load, an important issue for video processing. In this chapter, we consider both the decimated DWT and DT-CWT for video denoising.

In the 2D DWT-based or DT-CWT-based video denoising algorithms, spatial and temporal filtering are carried out separately [16], [140], [201], [219]-[223]. Pizurica *et al.* [220], [221], have used DWT-based image denoising methods for spatial filtering of individual frames and then applied a recursive temporal filtering in the pixel-domain. The same authors have developed in [223] another denoising algorithm, where a DWT-domain temporal noise reduction is performed followed by a spatial averaging filter for further noise reduction. These DWT-based algorithms outperform the traditional

pixel-domain spatio-temporal denoising techniques, such as the ones proposed in [209] and [210]. However, the advantages of the DWT are not fully exploited in the denoising methods of [220], [221] and [223], since the wavelets are used either in the spatial domain or in the temporal domain. Balster *et al.* [219] have used a DWT-based image denoising technique developed in [224] for spatial filtering of a noisy frame. The temporal noise reduction is performed by using compactly-supported Haar wavelet-based thresholding on the spatially-denoised frames. In [201], a recursive temporal filtering is performed on the spatially filtered [106] significant coefficients along the estimated motion trajectory.

In recent years, the statistical approach has emerged as a new tool for DWT-based spatio-temporal filtering. The basic idea is to model the DWT coefficients of a video with a pre-specified PDF, and consider the video denoising problem as the estimation of the noise-free coefficients using a statistical estimation technique. Some of the commonly used PDFs for the subband coefficients of the DWT of images are the GG [47], BK [74], and SNIG [23]. A reasonable assumption in the use of any of these PDFs is the i.i.d. nature of the subband coefficients of the DWT. But, it is known that the DWT coefficients of images have a considerable intra-subband statistical dependency and a weak inter-subband dependency [64]. In order to tackle the intra-subband dependency, the parameters of the PDFs of the DWT coefficients are estimated locally. The Gaussian PDF having locally-adaptive parameters is a popular choice for modeling the local neighboring coefficients, since this PDF is mathematically convenient for developing an estimation technique [106], [118], [122]. Improved image denoising performance is achieved in several algorithms using such a PDF and by exploiting a better estimate of the local parameters [107], [225]. The inter-subband dependency is considered in various ways, which include the selective thresholding [122], the use of spherically invariant bivariate PDF in the MAP estimation technique [65], [66], and

modeling the DWT coefficients by the HMT model [72]. It is obvious that the way in which the statistical dependency in the spatial filtering of the DWT coefficients of video frames is incorporated, plays a significant role in the denoising performance. Therefore, a spatial filtering that considers the statistical dependency of the inter-frame DWT coefficients of a video should improve the denoising performance further. Such an effort can be found in [16], wherein Gupta *et al.* model the DWT coefficients of a *current frame* and its difference from the *previous frame* as i.i.d. GG for an MMSE-based denoising algorithm. The algorithm ignores the dependency of the neighboring frames except for the previous one, and there is no closed-form expression for the ultimate MMSE-based estimator.

In this chapter, video denoising algorithms are developed by using a joint probabilistic model that takes into account the statistical dependency of the decimated DWT or DT-CWT coefficients of any two neighboring frames of a video. The joint PDF is chosen as the BVG, since the correlation parameter of this PDF may be used as a measure of motion between the wavelet coefficients of the two neighboring frames and the development of estimation techniques using such PDF is mathematically tractable. The intra-subband dependencies are taken into account by considering that the parameters of the BVG PDF vary locally. The proposed PDF is utilized to develop bivariate MAP estimators for spatial filtering of a noisy video. It is shown that the introduction of the proposed probabilistic model for the video wavelet coefficients results in a performance better than that provided by other denoising algorithms, both in terms of the visual quality and PSNR.

8.3 Denoising Algorithms

In this section, we assume that pixels of the video frames are corrupted by AWGN with a known variance σ_ε^2. If the variance is unknown, it may be estimated by applying

the median absolute deviation method [120] in the highest frequency subband of the noisy wavelet coefficients. Since the noise is uncorrelated with the true signal, the corresponding noisy wavelet coefficient of a frame (say, k-th frame) at a spatial location (i, j) of a subband is given by

$$g_{k(ij)} = f_{k(ij)} + \varepsilon_{k(ij)} \qquad k = 1, 2, \cdots, N_3 \tag{8.1}$$

where, $f_{k(ij)}$ and $\varepsilon_{k(ij)}$, respectively, are the wavelet coefficients of the original frame and additive noise at that reference location and N_3 is the total number of frames in the video sequence. For notational simplicity, $f_{k(ij)}$ in the decimated DWT is denoted as f_k and in the polar form of the DT-CWT as $f_{rk} \angle f_{\theta k}$. Similar notations are used for $g_{k(ij)}$ and $\varepsilon_{k(ij)}$. The proposed wavelet-based noise reduction algorithms operate in two steps. First, a bivariate MAP estimator estimates the decimated DWT coefficients, denoted as $\tilde{f}_k$, or the magnitude components of DT-CWT coefficients, denoted as $\tilde{f}_{rk}$, of the *current frame* taking the correlation information of the l-th *previous frame* into account. The mean, variances, and covariances of the joint prior function are estimated from the bivariate ML estimator. In the second step, the estimated coefficients $\tilde{f}_k$ or components $\tilde{f}_{rk}$ are passed through a recursive temporal averaging filter for additional noise reduction to obtain the final estimate $\hat{f}_k$ or $\hat{f}_{rk}$. Since the estimated magnitude components along with the noisy phase components are sufficient for a good noise reduction performance [66], [67], the denoised DT-CWT coefficients are obtained as $\hat{f}_{rk} \angle g_{\theta k}$. Finally, the denoised wavelet coefficients are inverse transformed to obtain the denoised video frame. The block diagram of the proposed DWT-based noise reduction scheme is shown in Fig. 8.1.

8.3.1 Spatial Filtering

In this section, bivariate MAP estimators are developed to spatially filter the decimated DWT coefficients or the magnitude components of the DT-CWT coefficients

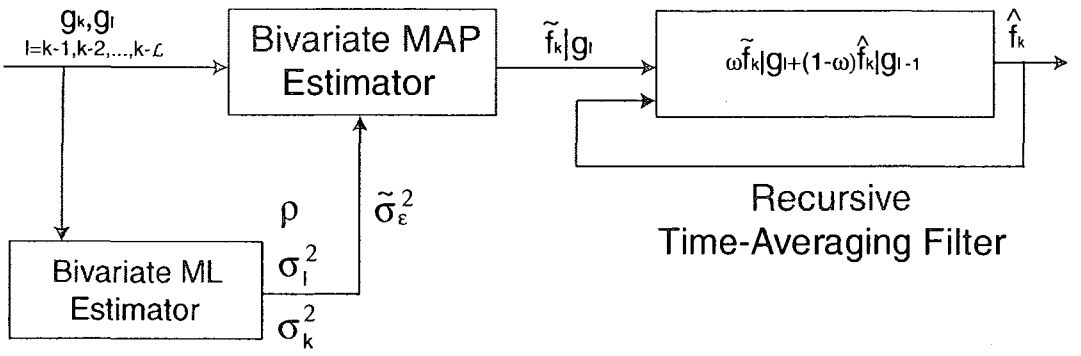


Figure 8.1: Block diagram of the proposed DWT-based denoising scheme.

of a noisy *current frame* using the coefficients or components of a *previous frame*. These estimators consider appropriate noise statistics and the proposed joint prior function for the noise-free coefficients or components.

Bivariate MAP Estimator in Decimated DWT Domain

In the decimated DWT domain, the AWGN is modelled as a zero-mean Gaussian random variable with a subband-adaptive variance $\tilde{\sigma}_\varepsilon^2$ (see Chapter 5). The bivariate MAP estimator for estimating the wavelet coefficients of the *current frame* from the noisy versions of the *current frame* and the l-th *previous frame* can be written as

$$\tilde{\boldsymbol{f}}_k(\boldsymbol{g}_k, \boldsymbol{g}_l) = \arg\max_{\boldsymbol{f}_k, \boldsymbol{f}_l} \left[-\frac{(\boldsymbol{g}_k - \boldsymbol{f}_k)^2}{2\tilde{\sigma}_\varepsilon^2} - \frac{(\boldsymbol{g}_l - \boldsymbol{f}_l)^2}{2\tilde{\sigma}_\varepsilon^2} + \ln p_{\boldsymbol{f}_k \boldsymbol{f}_l}(\boldsymbol{f}_k, \boldsymbol{f}_l) \right] \tag{8.2}$$

where $p_{\boldsymbol{f}_k \boldsymbol{f}_l}(\boldsymbol{f}_k, \boldsymbol{f}_l)$ is the joint prior function of $\boldsymbol{f}_k$ and $\boldsymbol{f}_l$. In the case of decimated DWT, the joint prior function of the local neighboring coefficients, $p_{\boldsymbol{f}_k \boldsymbol{f}_l}(\boldsymbol{f}_k, \boldsymbol{f}_l)$, is chosen as

$$p_{\boldsymbol{f}_k, \boldsymbol{f}_l}(\boldsymbol{f}_k, \boldsymbol{f}_l) = \frac{1}{2\pi\sigma_k\sigma_l\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{\boldsymbol{f}_k^2}{\sigma_k^2} + \frac{\boldsymbol{f}_l^2}{\sigma_l^2} - 2\rho \frac{\boldsymbol{f}_k}{\sigma_k} \frac{\boldsymbol{f}_l}{\sigma_l} \right\} \right] \quad -1 < \rho < 1 \tag{8.3}$$

where σ_k ($\sigma_k \geq 0$) and σ_l ($\sigma_l \geq 0$) are the standard deviations of the signal coefficients of the k-th and l-th frames, respectively, and ρ is the correlation between these

coefficients. In order to maximize (8.2), we need to solve the equations

$$\frac{\boldsymbol{g}_k - \boldsymbol{f}_k}{\tilde{\sigma}_\varepsilon^2} - \frac{1}{1-\rho^2}\left\{\frac{\boldsymbol{f}_k}{\sigma_k^2} - \frac{\rho \boldsymbol{f}_l}{\sigma_k \sigma_l}\right\} = 0 \tag{8.4}$$

$$\frac{\boldsymbol{g}_l - \boldsymbol{f}_l}{\tilde{\sigma}_\varepsilon^2} - \frac{1}{1-\rho^2}\left\{\frac{\boldsymbol{f}_l}{\sigma_l^2} - \frac{\rho \boldsymbol{f}_k}{\sigma_k \sigma_l}\right\} = 0 \tag{8.5}$$

Finally, from (8.4) and (8.5), the estimated wavelet coefficients of the *current frame* $\tilde{\boldsymbol{f}}_k$, based on the l-th *previous frame* are [138]

$$\tilde{\boldsymbol{f}}_k \mid \boldsymbol{g}_l = \frac{\sigma_k^2 \bar{\sigma}^2}{\sigma_k^2 \bar{\sigma}^2 + \tilde{\sigma}_\varepsilon^2[\sigma_l^2 + \tilde{\sigma}_\varepsilon^2]}\left[\boldsymbol{g}_k + \rho \frac{\sigma_l}{\sigma_k}\frac{\tilde{\sigma}_\varepsilon^2}{\bar{\sigma}^2}\boldsymbol{g}_l\right] \tag{8.6}$$

where, $\bar{\sigma}^2 = [1-\rho^2]\sigma_l^2 + \tilde{\sigma}_\varepsilon^2$, for $l = k-1, k-2, \cdots, k-\mathcal{L}$. The closed-form expressions obtained in (8.6) offer an easy implementation of the spatial filtering scheme. Fig. 8.2 shows the shrinkage functions obtained from the proposed bivariate MAP estimator at different values of the correlation coefficient. It may be seen from this figure that if ρ is non-zero, then the noisy wavelet coefficients of the *previous frame* have a significant contribution for estimating the denoised coefficients of the *current frame*. Thus, ρ serves as a measure of the motion between two coefficients at a particular spatial location of subbands of two neighboring frames of a video.

The variance and correlation of the decimated DWT coefficients of the *current* and *previous* frames are estimated from the noisy observation using the ML method as [138]

$$\sigma_k^2 = \max\left(\frac{1}{\aleph}\sum_{\mathcal{S}} \boldsymbol{g}_k^2 - \tilde{\sigma}_\varepsilon^2, 0\right) \tag{8.7}$$

$$\sigma_l^2 = \max\left(\frac{1}{\aleph}\sum_{\mathcal{S}} \boldsymbol{g}_l^2 - \tilde{\sigma}_\varepsilon^2, 0\right) \tag{8.8}$$

$$\rho = \max\left(\min\left(1, \frac{1}{\sigma_k \sigma_l}\left(\frac{1}{\aleph}\sum_{\mathcal{S}} \boldsymbol{g}_k \boldsymbol{g}_l\right)\right), -1\right) \tag{8.9}$$

where $\mathcal{S}$ is a square-shaped local neighborhood region centered at $\boldsymbol{g}_k$ or $\boldsymbol{g}_l$ and $\aleph$ is the total number of data samples in this region.

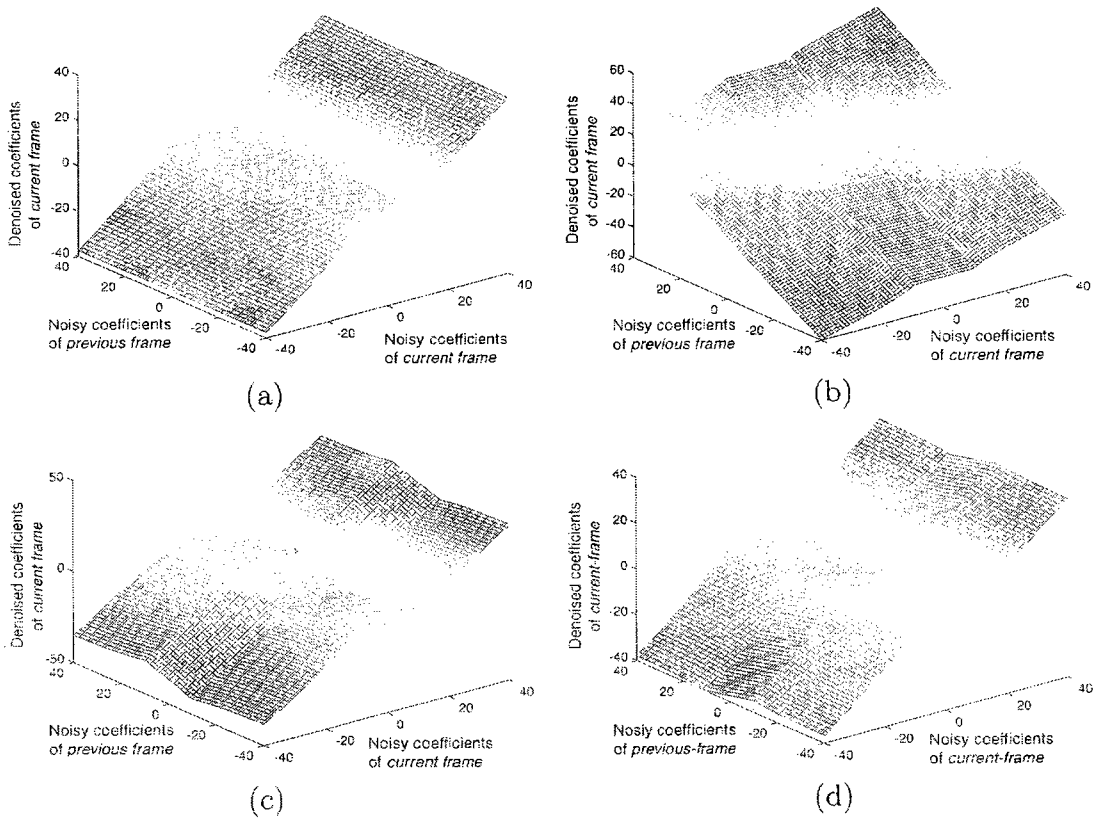


Figure 8.2: Shrinkage functions of the decimated DWT coefficients using the bivariate MAP estimator at different values of the correlation coefficient ρ: (a) $\rho = 0$, (b) $\rho = 1.00$, (c) $\rho = -0.25$, and (d) $\rho = 0.50$.

Bivariate MAP Estimator in DT-CWT Domain

In the DT-CWT domain, the magnitude components of the Gaussian noise in the first-level decomposition can be modelled very well by the GΓ distribution (see Chapter 4). It is also shown in Chapter 4 that in the second and higher level decompositions, these components follow the Rayleigh distribution, which is a particular case of the GΓ distribution. Hence, the bivariate estimator for the magnitude components of the DT-CWT coefficients considers that the magnitude components of the noise follow the GΓ distribution. The bivariate MAP estimator for estimating the magnitude components of the DT-CWT coefficients of the *current frame* from the noisy versions of the *current frame* and the l-th *previous frame* can be written as

$$\tilde{\boldsymbol{f}}_{\mathrm{r}k}(\boldsymbol{g}_{\mathrm{r}k}, \boldsymbol{g}_{\mathrm{r}l}) = \arg\max_{\boldsymbol{f}_{\mathrm{r}k}, \boldsymbol{f}_{\mathrm{r}l}} \Big[(2a-1)\left(\ln(\boldsymbol{g}_{\mathrm{r}k} - \boldsymbol{f}_{\mathrm{r}k}) + \ln(\boldsymbol{g}_{\mathrm{r}l} - \boldsymbol{f}_{\mathrm{r}l})\right) - \frac{a}{b^2}\left((\boldsymbol{g}_{\mathrm{r}k} - \boldsymbol{f}_{\mathrm{r}k})^2 + (\boldsymbol{g}_{\mathrm{r}l} - \boldsymbol{f}_{\mathrm{r}l})^2\right) + \ln p_{\boldsymbol{f}_{\mathrm{r}k}\boldsymbol{f}_{\mathrm{r}l}}(\boldsymbol{f}_{\mathrm{r}k}, \boldsymbol{f}_{\mathrm{r}l}) \Big] \tag{8.10}$$

where $p_{\boldsymbol{f}_{\mathrm{r}k}\boldsymbol{f}_{\mathrm{r}l}}(\boldsymbol{f}_{\mathrm{r}k}, \boldsymbol{f}_{\mathrm{r}l})$ is the joint prior function of $\boldsymbol{f}_{\mathrm{r}k}$ and $\boldsymbol{f}_{\mathrm{r}l}$. In the case of DT-CWT, the joint prior function of the local neighboring magnitude components, $p_{\boldsymbol{f}_{\mathrm{r}k}\boldsymbol{f}_{\mathrm{r}l}}(\boldsymbol{f}_{\mathrm{r}k}, \boldsymbol{f}_{\mathrm{r}l})$, is chosen as

$$p_{\boldsymbol{f}_{\mathrm{r}k}\boldsymbol{f}_{\mathrm{r}l}}(\boldsymbol{f}_{\mathrm{r}k}, \boldsymbol{f}_{\mathrm{r}l}) = \frac{1}{2\pi\sigma_{\mathrm{r}k}\sigma_{\mathrm{r}l}\sqrt{1-\rho_{\mathrm{r}}^2}} \exp\left[-\frac{1}{2(1-\rho_{\mathrm{r}}^2)} \left\{ \frac{(\boldsymbol{f}_{\mathrm{r}k} - \mu_{\mathrm{r}k})^2}{\sigma_{\mathrm{r}k}^2} + \frac{(\boldsymbol{f}_{\mathrm{r}l} - \mu_{\mathrm{r}l})^2}{\sigma_{\mathrm{r}l}^2} - 2\rho_{\mathrm{r}} \frac{(\boldsymbol{f}_{\mathrm{r}k} - \mu_{\mathrm{r}k})(\boldsymbol{f}_{\mathrm{r}l} - \mu_{\mathrm{r}l})}{\sigma_{\mathrm{r}k}\sigma_{\mathrm{r}l}} \right\} \right] \quad -1 \le \rho_{\mathrm{r}} \le 1 \tag{8.11}$$

where $\{\mu_{\mathrm{r}k}, \sigma_{\mathrm{r}k}\}$ $(\mu_{\mathrm{r}k} \ge 0, \sigma_{\mathrm{r}k} \ge 0,)$ and $\{\mu_{\mathrm{r}l}, \sigma_{\mathrm{r}l}\}$ $(\mu_{\mathrm{r}l} \ge 0, \sigma_{\mathrm{r}l} \ge 0)$, respectively, are the parameters, mean and standard deviation, of the magnitude components of the DT-CWT coefficients of the k-th and l-th frames, respectively, and ρ_{r} is the correlation between these components. Thus, to maximize (8.10) we need to solve

the following equations

$$\frac{1-2a}{\boldsymbol{g}_{\mathrm{r}k}-\boldsymbol{f}_{\mathrm{r}k}}+\frac{2a}{b^2}(\boldsymbol{g}_{\mathrm{r}k}-\boldsymbol{f}_{\mathrm{r}k})-\frac{\boldsymbol{f}_{\mathrm{r}k}-\mu_{\mathrm{r}k}}{\sigma_{\mathrm{r}k}^2(1-\rho_{\mathrm{r}}^2)}+\frac{\rho_{\mathrm{r}}(\boldsymbol{f}_{\mathrm{r}l}-\mu_{\mathrm{r}l})}{\sigma_{\mathrm{r}k}\sigma_{\mathrm{r}l}(1-\rho_{\mathrm{r}}^2)}=0 \tag{8.12}$$

$$\frac{1-2a}{\boldsymbol{g}_{\mathrm{r}l}-\boldsymbol{f}_{\mathrm{r}l}}+\frac{2a}{b^2}(\boldsymbol{g}_{\mathrm{r}l}-\boldsymbol{f}_{\mathrm{r}l})-\frac{\boldsymbol{f}_{\mathrm{r}l}-\mu_{\mathrm{r}l}}{\sigma_{\mathrm{r}l}^2(1-\rho_{\mathrm{r}}^2)}+\frac{\rho_{\mathrm{r}}(\boldsymbol{f}_{\mathrm{r}k}-\mu_{\mathrm{r}k})}{\sigma_{\mathrm{r}l}\sigma_{\mathrm{r}k}(1-\rho_{\mathrm{r}}^2)}=0 \tag{8.13}$$

Finally, from (8.12) and (8.13), the magnitude components of the DT-CWT coefficients of the *current frame* $\tilde{\boldsymbol{f}}_{\mathrm{r}k}$, based on the l-th *previous frame*, are estimated in the form of a quadratic solution given by [87]

$$\tilde{\boldsymbol{f}}_{\mathrm{r}k}|\boldsymbol{g}_{\mathrm{r}l}=\max\left(\frac{1}{2\mathcal{A}_k}\left(-\mathcal{B}_k-\sqrt{\mathcal{B}_k^2-4\mathcal{A}_k\mathcal{C}_k}\right),0\right) \tag{8.14}$$

where

$$\mathcal{A}_k=2\left(a\sigma_{\mathrm{r}k}^2+\frac{b^2}{2(1-\rho_{\mathrm{r}}^2)}\right) \tag{8.15}$$

$$\mathcal{B}_k=\boldsymbol{g}_{\mathrm{r}k}\left(2a\sigma_{\mathrm{r}k}^2-\frac{b^2}{2(1-\rho_{\mathrm{r}}^2)}\right)+\mathcal{T}(\boldsymbol{f}_{\mathrm{r}l}) \tag{8.16}$$

$$\mathcal{C}_k=(1-2a)\frac{\sigma_{\mathrm{r}k}^2b^2}{2}+a\boldsymbol{g}_{\mathrm{r}k}^2\sigma_{\mathrm{r}k}^2+\boldsymbol{g}_{\mathrm{r}k}\mathcal{T}(\boldsymbol{f}_{\mathrm{r}l}) \tag{8.17}$$

$$\mathcal{T}(\boldsymbol{f}_{\mathrm{r}l})=\left(\rho_{\mathrm{r}}\frac{\sigma_{\mathrm{r}k}}{\sigma_{\mathrm{r}l}}(\boldsymbol{f}_{\mathrm{r}l}-\mu_{\mathrm{r}l})+\mu_{\mathrm{r}k}\right)\frac{b^2}{2(1-\rho_{\mathrm{r}}^2)} \tag{8.18}$$

The mean, variance, and correlation of the magnitude components of the DT-CWT coefficients of the *current* and *previous* frames are estimated from the noisy observation using the ML method as [87]

$$\mu_{\mathrm{r}k}=\max\left(\frac{1}{\aleph}\sum_{\mathcal{S}}\boldsymbol{g}_{\mathrm{r}k}-\mu_{\varepsilon_{\mathrm{r}}},0\right) \tag{8.19}$$

$$\mu_{\mathrm{r}l}=\max\left(\frac{1}{\aleph}\sum_{\mathcal{S}}\boldsymbol{g}_{\mathrm{r}l}-\mu_{\varepsilon_{\mathrm{r}}},0\right) \tag{8.20}$$

$$\sigma_{\mathrm{r}k}^2=\max\left(\frac{1}{\aleph}\sum_{\mathcal{S}}(\boldsymbol{g}_{\mathrm{r}k}-\bar{\boldsymbol{g}}_{\mathrm{r}k})^2-\sigma_{\varepsilon_{\mathrm{r}}}^2,0\right) \tag{8.21}$$

$$\sigma_{\mathrm{r}l}^2=\max\left(\frac{1}{\aleph}\sum_{\mathcal{S}}(\boldsymbol{g}_{\mathrm{r}l}-\bar{\boldsymbol{g}}_{\mathrm{r}l})^2-\sigma_{\varepsilon_{\mathrm{r}}}^2,0\right) \tag{8.22}$$

$$\rho_{\mathbf{r}} = \max\left(\min\left(1, \frac{1}{\sigma_{\mathbf{r}k}\sigma_{\mathbf{r}l}}\left(\frac{1}{\aleph}\sum_{\mathcal{S}} g_{\mathbf{r}k}g_{\mathbf{r}l} - \mu_{\mathbf{r}k}\mu_{\mathbf{r}l} - \mu_{\mathbf{r}k}\mu_{\epsilon_{\mathbf{r}}} - \mu_{\mathbf{r}l}\mu_{\epsilon_{\mathbf{r}}} - \mu^2_{\epsilon_{\mathbf{r}}}\right)\right), -1\right) \tag{8.23}$$

where $\mathcal{S}$ is a square-shaped local neighborhood region centered at $g_{\mathbf{r}k}$ or $g_{\mathbf{r}l}$, $\aleph$ is the total number of data samples in this region, and $\{\bar{g}_{\mathbf{r}k}, \bar{g}_{\mathbf{r}l}\}$ are the mean parameters of the noisy samples $\{g_{\mathbf{r}k}, g_{\mathbf{r}l}\}$. The mean and variance of the magnitude components of the DT-CWT coefficients of the noise are obtained from the formulas given in Chapter 4. An initial estimate of $f_{\mathbf{r}l}$ is required to calculate $\mathcal{T}(f_{\mathbf{r}l})$. In our case, we obtain this estimate as

$$\bar{f}_{\mathbf{r}l} = \max\left(\frac{1}{2\bar{\mathcal{A}}_l}\left(-\bar{\mathcal{B}}_l - \sqrt{\bar{\mathcal{B}}_l^2 - 4\bar{\mathcal{A}}_l\bar{\mathcal{C}}_l}\right), 0\right) \tag{8.24}$$

where $\bar{\mathcal{A}}_l = \mathcal{A}_k|_{k=l,\rho_{\mathbf{r}}=0}$, $\bar{\mathcal{B}}_l = \mathcal{B}_k|_{k=l,l=k,\rho_{\mathbf{r}}=0}$, and $\bar{\mathcal{C}}_l = \mathcal{C}_k|_{k=l,l=k,\rho_{\mathbf{r}}=0}$.

8.3.2 Temporal Filtering

The estimated decimated DWT coefficients using (8.6) or the magnitude components of the DT-CWT coefficients using (8.14) considers the *current frame* and any one of the neighboring *previous frames*. In order to account for contributions from the several neighboring *previous frames*, a temporal filtering is essential. For the temporal processing, a relatively large value of $\mathcal{L}$ ($\mathcal{L} > 1$) is preferable. However, as the time difference between the k-th and l-th frames increases, the expected value of ρ decreases (see Fig. 7.1). Hence, with the increase of $\mathcal{L}$, the value of g_l or $g_{\mathbf{r}l}$ appears to have very little effect in the estimation of $\tilde{f}_k$ or $\tilde{f}_{\mathbf{r}k}$. A recursive weighted averaging of the estimated coefficients $\tilde{f}_k|g_l$ is, therefore, proposed for the final estimation as

$$\hat{f}_k \mid g_l = \omega\tilde{f}_k \mid g_l + (1-\omega)\hat{f}_k \mid g_{l-1} \tag{8.25}$$

where, $\hat{f}_k \mid g_{k-\mathcal{L}} = \tilde{f}_k \mid g_{k-\mathcal{L}}$, and ω $(0 \leq \omega \leq 1)$ is the weight parameter. A similar recursive temporal filter can be used for the final estimation of $\hat{f}_{\mathbf{r}k} \mid g_{\mathbf{r}l}$. Intuitively,

one can say that the recursive filter should accumulate and average the estimated $\tilde{\boldsymbol{f}}_k$ or $\tilde{\boldsymbol{f}}_{rk}$ from all the *previous $\mathcal{L}$-frames* with the highest weight assigned to the nearest frame and the lowest to the farthest one, where the weight value is determined by ω. If the motion between the neighboring frames is very high, the error between $\tilde{\boldsymbol{f}}_k \mid \boldsymbol{g}_l$ and $\hat{\boldsymbol{f}}_k \mid \boldsymbol{g}_{l-1}$ or $\tilde{\boldsymbol{f}}_{rk} \mid \boldsymbol{g}_{rl}$ and $\hat{\boldsymbol{f}}_{rk} \mid \boldsymbol{g}_{r(l-1)}$ will be large, and ω could be set to a value close to unity. On the other hand, for a very small motion among the neighboring frames, ω could be set close to zero. Here, we have chosen the value of ω to be 0.5, taking into account the fact that a video can have frames with both high and low motion contents.

Note that in the proposed noise reduction schemes, the variance or the mean and variance for each of the coefficients or components for all the frames are calculated only once. For processing the coefficients or components of the *current frame*, it is necessary to calculate the correlation parameter for each of the coefficients or components of the *previous $\mathcal{L}$-frames*. Therefore, for a video of size $N_1 \times N_2 \times N_3$, the computational complexity of the decimated DWT-based denoising algorithm is $\mathcal{O}\left(\mathtt{w}^2(1+\mathcal{L})N_1N_2N_3\right)$ and that of the DT-CWT based algorithm is $\mathcal{O}\left(4\mathtt{w}^2(2+\mathcal{L})N_1N_2N_3\right)$, where $\mathtt{w} \times \mathtt{w}$ is the size of a squared-shape local window.

8.4 Experimental Results

In order to compare the performance of the proposed denoising methods with that of the other methods, we give in this section results concerning experimental studies that have been conducted using the six grayscale video sequences, viz., *Tennis*, *Salesman*, *Coastguard*, *Football*, *Foreman*, and *Garden*. The orthogonal wavelets S-8 with a 4-level decomposition are used for the decimated DWT, whereas $(11, 17)$-tap biorthogonal filters at level $\ell = 1$, and 6-tap Q-shift orthogonal filters at $\ell = 2$, 3, and 4 are used for the DT-CWT. The number of *previous-frames*, $\mathcal{L}$, depends on the

overall motion content of the video sequence. In our experimental study, however, we have found that $\mathcal{L} = 3$ is sufficient for an appreciable denoising performance.

Eight different denoising methods have been considered in our experimentations, where two of them work in the pixel-domain and the rest in the wavelet-domain. The pixel-domain denoising methods use the 3D KNN filter [210] and the 3D rational filter [209]. Wavelet-based denoising methods are the adaptive Bayes-Shrink method of the 3D wavelet coefficients [217], the combined wavelet-domain and temporal denoising method of Pizurica *et al.* [220], the BivariateShrink method [66], the locally-adaptive window-based ML (LAWML) method [106], and the proposed interframe MAP (IFMAP) methods. The proposed IFMAP methods have been tested using several square-shaped local neighborhoods for estimating the local parameters (e.g., mean, variance, and correlation). It is found that the 5×5 or 7×7 window size provides a satisfactory denoising performance. All the results for the proposed IFMAP methods given in this section are obtained using a 5×5 window unless otherwise mentioned and for other methods we use the default settings. Table 8.1 shows the average PSNR values obtained from the various pixel-based and the wavelet-based denoising techniques for the first 120 frames of the test video sequences, wherein we have used $\sigma_\epsilon = 10$, 20, and 40. In this table, for the combined wavelet-domain and temporal filter of [220], we have used the temporal threshold parameter to be 23 and the weight factor 0.6, since these parameters yield the maximum PSNR gain, as mentioned in [220]. According to the PSNR results, the performance of the decimated DWT-based video denoising methods are always superior to that of the simple pixel-domain spatio-temporal filter, and the DT-CWT based methods are superior to the corresponding decimated DWT-based methods. It can be observed from Table 8.1 that, except in one instance, i.e., *Foreman* sequence denoised by the Bayes-Shrink method of the 3D DWT [217], the proposed decimated DWT-based algorithm provides

Table 8.1: Output average PSNR, $20\log_{10}(255/\sigma_e)$ in dB, where σ_ϵ is the error standard deviation, for various denoising methods using frames 1-120 of the test video sequences.

Video sequences	*Tennis*	*Salesman*	*Coastguard*	*Football*	*Foreman*	*Garden*
Noise level	$\sigma_\epsilon = 10$	$\sigma_\epsilon = 10$	$\sigma_\epsilon = 20$	$\sigma_\epsilon = 20$	$\sigma_\epsilon = 40$	$\sigma_\epsilon = 40$
	Pixel-based Methods					
3D Rational [209]	29.86	31.65	27.23	26.24	24.74	18.73
3D KNN [210]	29.14	32.13	26.62	24.95	25.70	18.12
	Decimated DWT-based Methods					
Bayes-3D [217]	30.89	34.18	27.84	26.28	**27.35**	20.13
Pizurica [220]	30.76	34.11	27.42	26.62	26.75	20.00
Bi-Shrink [66]	31.05	32.77	27.82	26.61	26.40	20.35
LAWML [106]	31.40	32.77	28.06	26.77	26.50	20.72
Proposed IFMAP	**32.05**	**34.28**	**28.40**	**27.06**	27.16	**20.92**
	DT-CWT-based Methods					
Bi-Shrink [66]	31.23	33.14	27.93	26.88	27.56	20.39
Proposed IFMAP	**32.07**	**34.38**	**28.67**	**27.40**	**27.59**	**21.20**

average PSNR values, which are higher than that provided by the other decimated DWT-based methods. It may be noted that even in the case of *Foreman* sequence the difference between the PSNRs of the proposed and Bayes-Shrink methods is not more 0.20 dB. As can be seen from Table 8.1 that the proposed decimated DWT-based method which considers the inter-frame dependency shows approximately 0.9 dB improvement in the average PSNR value as compared to that provided by the BivariateShrink method of [66], which considers the inter-subband dependency when applied to the individual frames of the video. In addition, the proposed DWT-based method shows approximately 0.7 dB improvement in the average PSNR value as compared to that provided by the LAWML method [106], which considers only the intra-subband dependency. Clearly, the parameter ρ has a significant effect on the denoising performance in terms of the average PSNR values. Table 8.1 also shows the average PSNR values obtained from the two DT-CWT-based methods, namely, the BivariateShrink [66] and proposed IFMAP. According to these PSNR values the

proposed IFMAP performs better than the BivariateShrink method.

Fig. 8.3 shows the frame-by-frame performance of the various decimated DWT-based methods in terms of the output PSNR values. From this figure, it can be seen that the proposed denoising method provides the highest PSNR value in most cases. Moreover, the variation of the output PSNR over all the 120 frames is the least in the proposed method compared to the other denoising methods. This implies that the proposed denoising method has a better capability to adapt in accordance with the motion within the video without affecting the denoising performance significantly as compared to the other methods. It is also seen from Fig. 8.3 that the proposed method provides better PSNR values for all the frames of a given video sequence as compared to that provided by the BivariateShrink method [66], indicating that in video denoising, it is more important to consider inter-frame dependency than inter-subband dependency, even when the motion is strong. Figs. 8.4(a) and 8.4(b) show the original 60-th frame and its corrupted version with $\sigma_\varepsilon = 10$ for the sequence *Salesman*. Figs. 8.4(c) and 8.4(d) show the denoised versions using the combined wavelet-domain and temporal filter of [220] and the proposed decimated DWT-based IFMAP filter, respectively; the absolute value of the errors occurring in these two denoising algorithms are shown in Figs. 8.4(e) and 8.4(f), respectively. Although the denoised frames from the two algorithms appear to be the same, the error images clearly show the difference between the two schemes. Specifically, the error of the denoised frame obtained from the combined wavelet-domain and temporal denoising of [220] is significant as compared to the proposed method in those spatial locations where motion exists, revealing that the proposed method adapts better with motion of a video. Corresponding results for the sequence *Football* are given in Fig. 8.5, where $\sigma_\varepsilon = 20$ has been used and the proposed decimated DWT-based IFMAP method is compared with the adaptive Bayes-Shrink method of the 3D DWT coefficients [217]. It can be observed that the

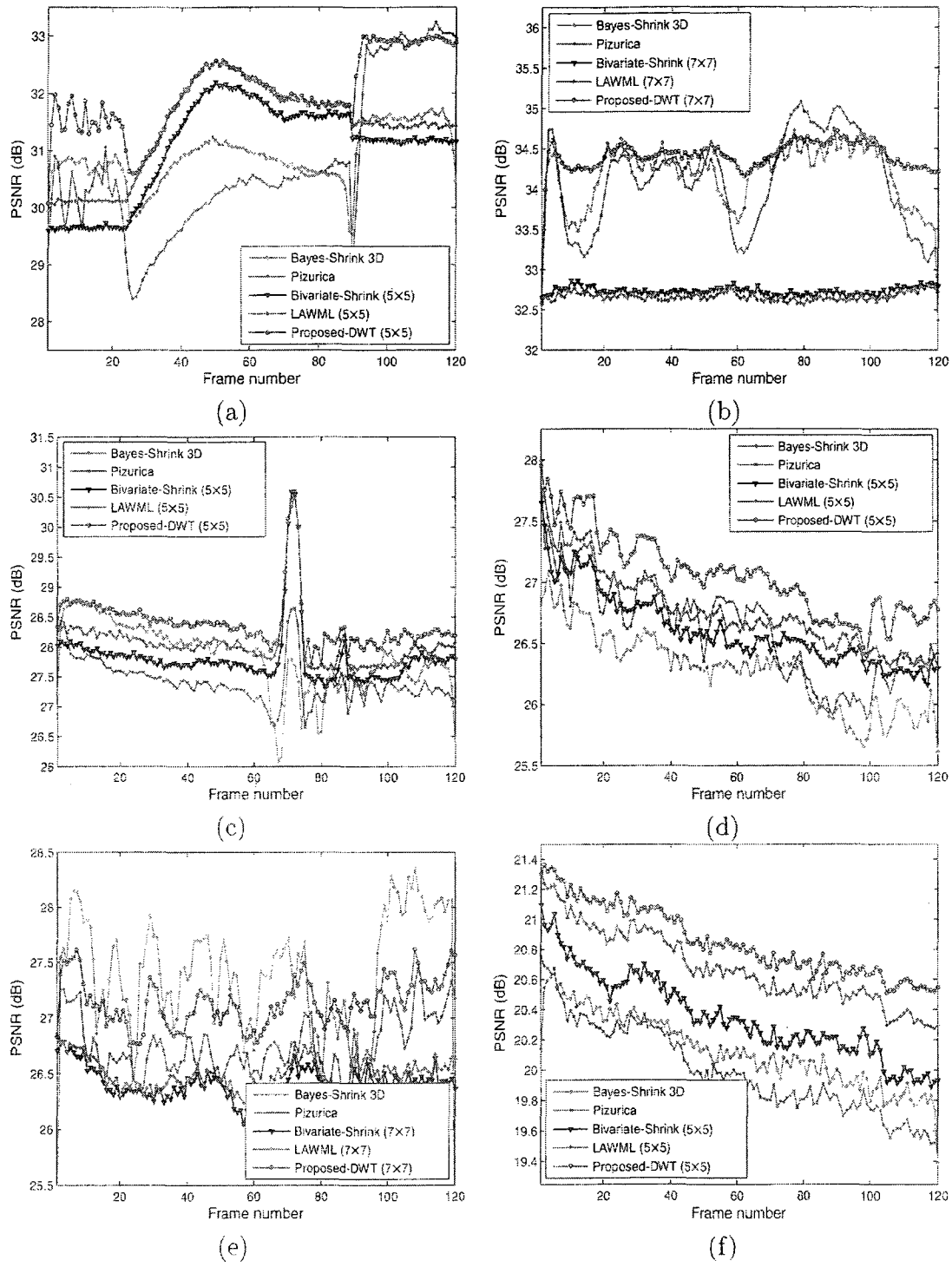


Figure 8.3: Output PSNR values for the different decimated DWT-based denoising schemes for the frames 1-120 of the video sequences. (a) *Tennis* with $\sigma_\varepsilon = 10$. (b) *Salesman* with $\sigma_\varepsilon = 10$. (c) *Coastguard* with $\sigma_\varepsilon = 20$. (d) *Football* with $\sigma_\varepsilon = 20$. (e) *Foreman* with $\sigma_\varepsilon = 40$. (f) *Garden* with $\sigma_\varepsilon = 40$.

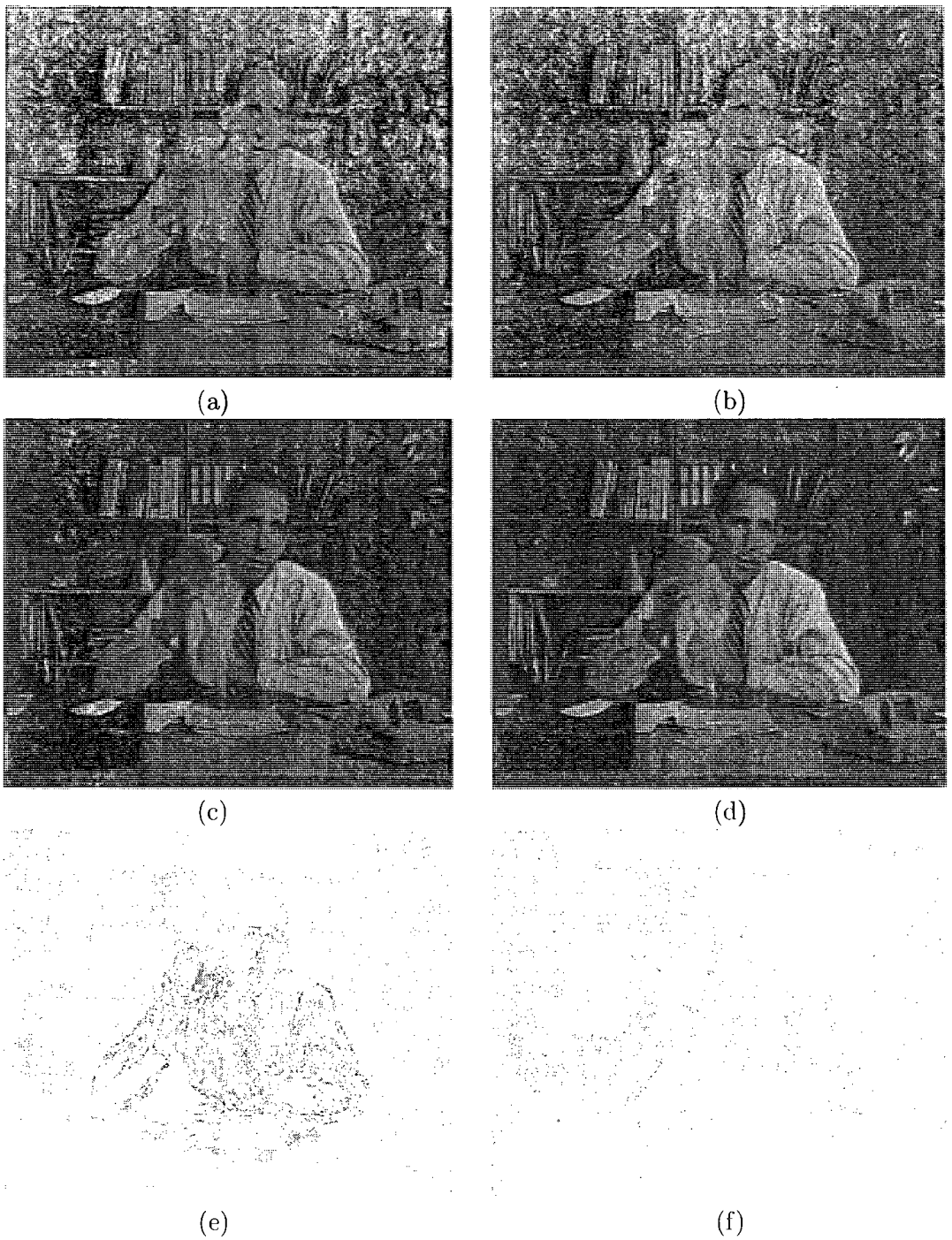


Figure 8.4: Filtered output of the 60-th frame of *Salesman* with $\sigma_\epsilon = 10$: (a) original frame, (b) noisy frame, and denoised frames using (c) the combined wavelet-domain and temporal filter and (d) the proposed decimated DWT-based IFMAP method. (e) Absolute error of (a) and (c). (f) Absolute error of (a) and (d).

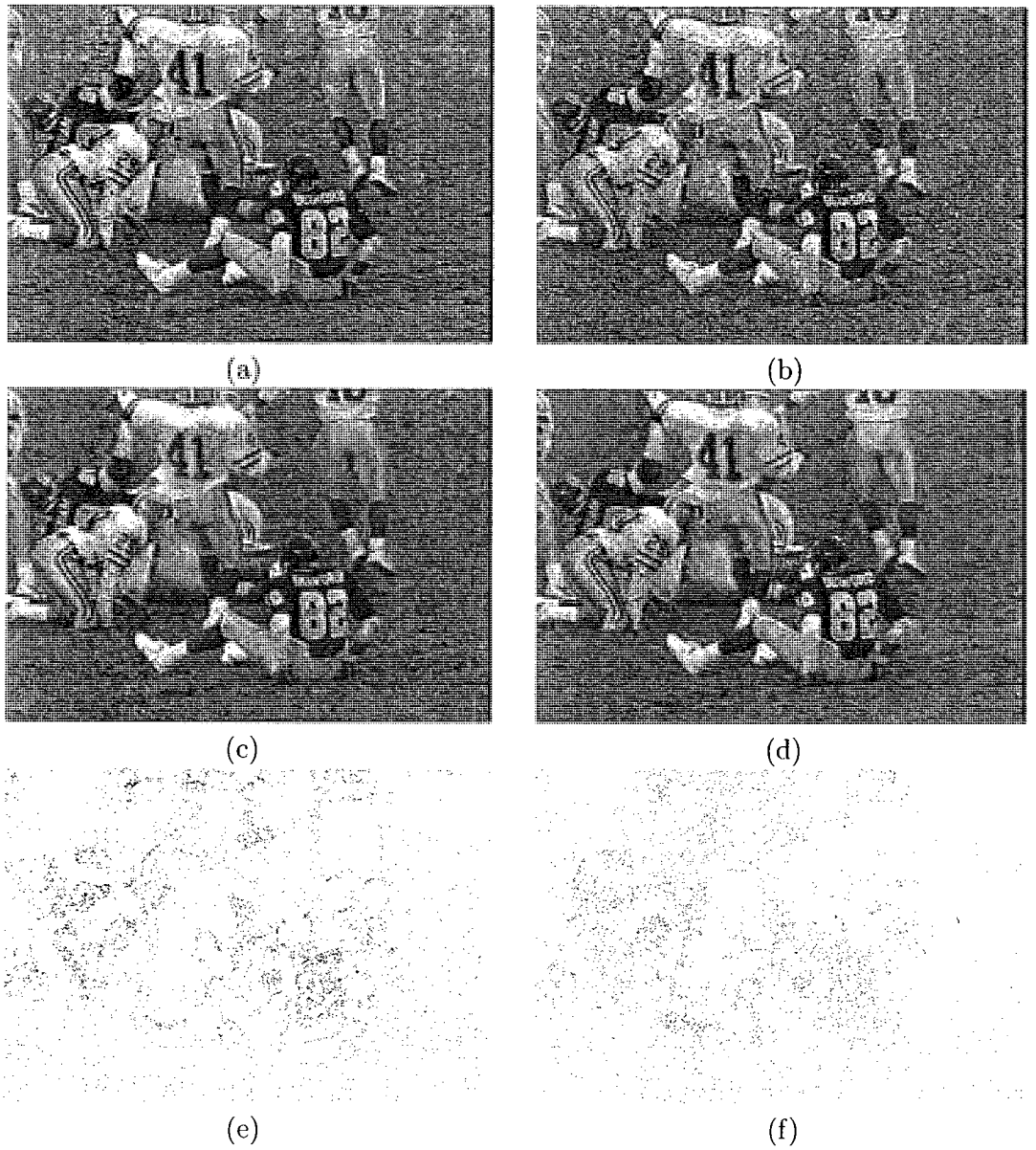


Figure 8.5: Filtered output of the 60-th frame of *Football* with $\sigma_\epsilon = 20$: (a) original frame, (b) noisy frame, and denoised frames using (c) the adaptive Bayes-Shrink of the 3D DWT coefficients and (d) the proposed decimated DWT-based IFMAP method. (e) Absolute error of (a) and (c). (f) Absolute error of (a) and (d).

proposed filter outperforms the 3D DWT-based Bayes-Shrink method, especially in the area of the frame where motion occurs. Another set of results for the sequence *Garden* are given in Fig. 8.6, where $\sigma_\varepsilon = 40$ has been used and the proposed DT-CWT-based IFMAP method is compared with the DT-CWT-based BivariateShrink method [66]. It can also be observed that the proposed filter outperforms the BivariateShrink method, especially in the edge of the objects where motion occurs. From all these figures, it appears that the proposed IFMAP methods result in a perceptual quality that is better than that provided by the other methods. It may be mentioned that this perceptual quality including the ability to reduce the amount of flicker provided by the proposed methods, appears to be even better than that provided by the other methods, when the entire 120 frames of the sequences are viewed rather than just one frame.

8.5 Conclusion

In this chapter, we have used the BVG PDF for the decimated DWT coefficients or the magnitude components of the DT-CWT coefficients of any two neighboring frames for reducing the AWGN in a video. It has been shown that the correlation coefficient of the PDF can be very effectively used as a measure of the motion between any two frames for denoising. Specifically, the BVG PDF is employed for spatial filtering of the noisy decimated DWT coefficients or magnitude components of DT-CWT coefficients of a video. For this purpose, we have developed closed-form bivariate MAP estimators. The spatially filtered coefficients or components are then passed through a recursive time averaging filter for additional noise reduction. The simulation results using six test video sequences have shown superior denoising performance of the proposed inter-frame MAP (IFMAP) denoising methods both in terms of the PSNR and the

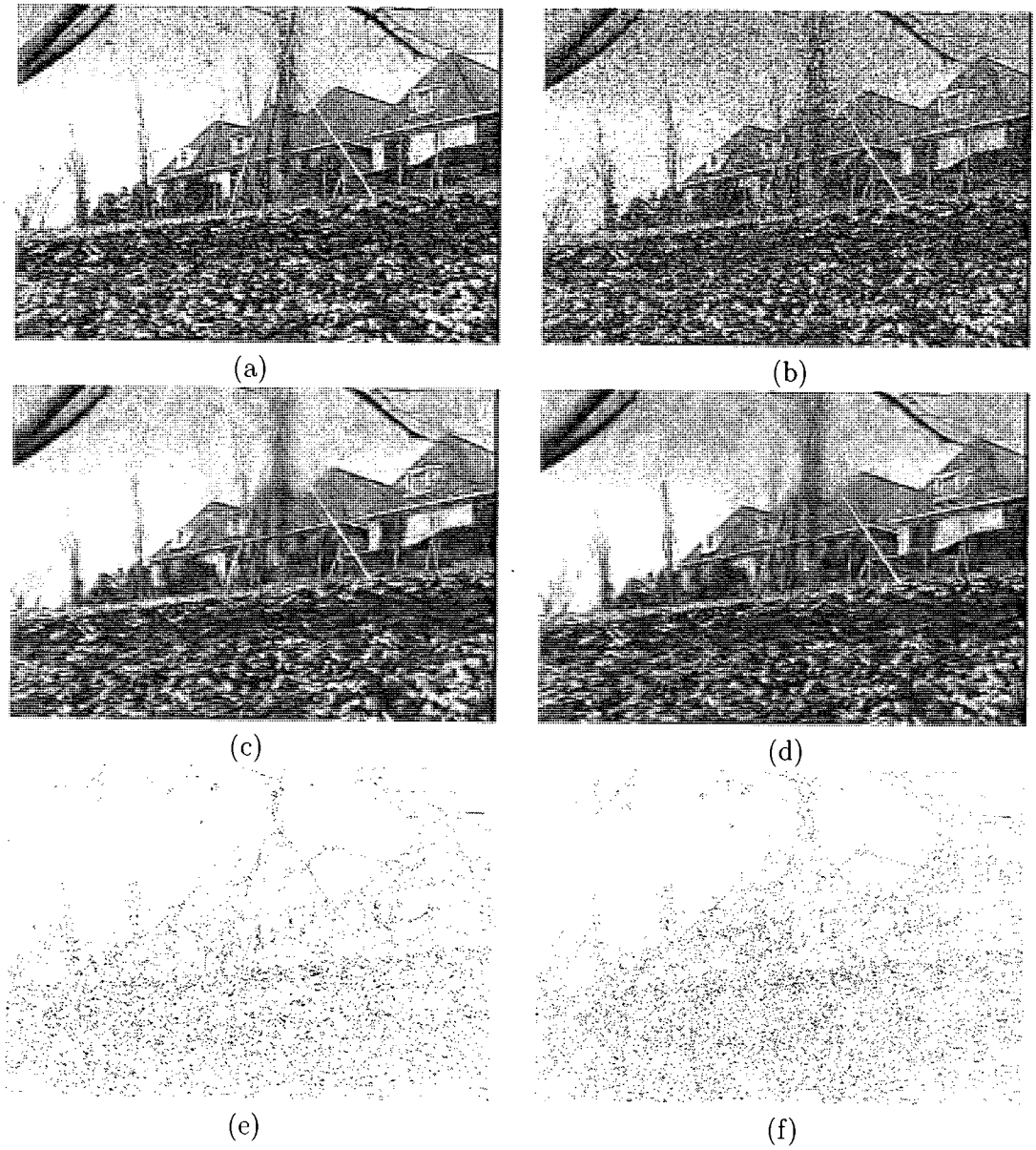


Figure 8.6: Filtered output of the 60-th frame of *Garden* with $\sigma_\epsilon = 40$: (a) original frame, (b) noisy frame, and denoised frames using (c) the DT-CWT-based BivariateShrink method and (d) the proposed DT-CWT-based IFMAP method. (e) Absolute error of (a) and (c). (f) Absolute error of (a) and (d).

perceptual quality compared to that provided by other denoising methods. Finally, it is worth mentioning that a significant feature of the proposed IFMAP-based video denosing methods is in their simplicity of implementation.

Chapter 9

Conclusion

9.1 Concluding Remarks

Probability density functions play a significant role in the development of statistical estimation and detection techniques for processing of signals. In general, a PDF should capture the essential variability of the signal such as the motion in a video, possess a good statistical match to the empirical PDF, and be mathematically tractable and computationally efficient. Studies have shown that the PDFs of the subband coefficients of the DWT of an image or video have sharp peaks and long tails. Since the shape of such PDFs is known, the parametric or semi-parametric techniques are used for the density estimation. An important advantage of the parametric density estimation techniques is that the PDF can effectively use *a priori* knowledge. Common parametric PDFs that are used for the subband coefficients of the DWT of an image or video signal include the GG, BK, SNIG, and $S\alpha S$. In some applications, such as noise reduction, an improved performance in estimation may be obtained by statistically modeling the local neighboring coefficients. An effective way to tackle such a scenario is to estimate the parameters of the PDFs using the local neighboring coefficients. It may be noted that the parametric PDFs possess a fixed number of parameters. The number of parameters of a semi-parametric PDF is not restricted. However, the estimation of parameters of such a PDF cannot make use of the information obtained

from the higher-order moments of the data samples; instead, this estimation adopts a computationally expensive iterative algorithm. An important outcome of not using an appropriate number of parameters estimated from the higher-order moments of the data samples, is that the conventional PDFs show an inadequate statistical match with the empirical PDFs. As a consequence, the existing statistical estimators and detectors that use any of such PDFs very often provide estimation and detection performance that are not satisfactory.

In order to overcome the shortfalls of the traditional wavelet-based statistical estimation and detection techniques for an image or video, this study has introduced a new PDF for the wavelet coefficients of such a signal. The new PDF has been developed using the Gauss-Hermite (GH) series expansion, since such an expansion allows the estimated PDF to have an appropriate number of parameters that can be expressed in terms of higher-order moments of the data samples. Estimation of such a PDF, however, requires a truncation of the GH series function and the calculation of the error prone higher-order sample-moments. Due to these reasons, the expanded PDF is not ensured to be non-negative. In order to maintain the non-negativity constraint of a PDF, a modification in the series has been proposed. Closed-form expressions of the parameters of the resulting PDF, termed as the modified Gauss-Hermite (MGH) PDF, for the wavelet coefficients have been derived in terms of higher-order moments of the data samples. Depending on the application scenario, the MGH PDF is capable of using a suitable number of parameters. For example, the PDF of the subband coefficients may use the parameters obtained from a relatively large number of higher-order moments, since in such a case the number of data samples is considerably large. On the other hand, the PDF of the local neighboring coefficients may use the parameters obtained from a smaller number of higher-order moments, as is the case when the number of data samples is small. It has been shown that univariate MGH PDFs

match very well the empirical PDFs of the subband coefficients as well as the local neighboring coefficients of the test images as compared to the commonly used PDFs such as the GG, BK, and SNIG do. In the process of developing a probabilistic model for the wavelet coefficients of a video, it has been shown that the correlation parameter of a joint PDF may be used as a measure of the motion between two neighboring frames. Bivariate Gaussian PDF is a simple choice for modeling the wavelet coefficients of such neighboring frames. In order to obtain a better probabilistic model, the bivariate MGH PDF has also been developed as an inter-frame joint PDF of the wavelet coefficients of a video. By using a number of test video sequences, it has been shown that the bivariate MGH PDFs match the empirical joint PDFs very well.

The proposed MGH PDF is used as the prior of the wavelet coefficients of images and video to develop the MAP and MMSE-based estimators for denoising of image and video signals and log-likelihood ratio-based detector for watermarking of image signals. Both the subband- and locally-adaptive denoising algorithms have been considered for images, whereas for video we consider the latter, since the latter provides a better denoising performance than the former. To obtain computationally efficient denoising algorithms for video, the bivariate Gaussian PDF has been used as a joint prior function for the local neighboring wavelet coefficients. In order to study the denoising performance, the proposed algorithms have been compared with the existing algorithms in terms of various indices such as the PSNR, SSIM, and visual quality. It has been shown that the proposed denoising algorithms provide a performance superior to that of the others. The proposed watermark detector is compared with the conventional detectors in terms of the probabilities of detection, probabilities of false alarm, and efficacy, as well as in terms of robustness. It has been shown that the proposed MGH detector provides credible watermark detection which is better than that provided by the existing detectors.

9.2 Scope for Further Work

The univariate and bivariate MGH PDFs developed in this thesis may be extended to the multivariate MGH PDF for processing the wavelet coefficients of a group of neighboring frames of a video. The proposed MGH PDF may be used in a variety of signal processing applications such as compression, pattern recognition, registration, stabilization, and segmentation, in addition to denoising and watermarking that have been considered in this thesis. Some of the possible avenues of research using this MGH PDF are as follows:

- A video denoising scheme may be developed by using a bivariate MGH PDF that considers higher-order moments in its parameter estimation.
- A video watermarking scheme for the H.264/MPEG-4 AVC standard may be developed using the MGH PDF. In this scheme, the motion compensated multiple frames may be modelled by the multivariate MGH PDF for embedding and detecting a watermark.
- An efficient compression scheme may be developed by estimating the quantization steps of the 2D DWT coefficients of images and 3D DWT coefficients of video by using the MGH PDF.
- An efficient registration scheme of medical images can be developed by using the bivariate MGH PDF. In this scheme, the cross-entropy of two images may be measured using the bivariate PDF, which is capable of capturing the multimodal nature of the empirical PDF.
- A registration scheme to overlay certain groups of frames of video taken from several sensors may be developed using the multivariate MGH PDF.

- A pattern recognition scheme may be developed by using the MGH PDF in a multi-hypotheses Bayesian classification method.
- A scheme may be developed for segmentation of image features by using MGH PDF to obtain an appropriate Bayesian threshold for the 2D DWT coefficients.

References

[1] Official web-site. [Online]. Available: http://www.jpeg.org/

[2] Official web-site. [Online]. Available: http://www.chiariglione.org/mpeg/

[3] X. Jin, X. Sun, F. Wu, G. Zhu, and S. Li, "H.264-compatible spatially scalable video coding with in-band prediction," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 1, Genoa, Italy, 2005, pp. 489–492.

[4] R. G. de Oliveira and R. L. de Queiroz, "Intra prediction versus wavelets and lapped transforms in an H.264/AVC coder," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, San Diego, CA, 2008, pp. 137–140.

[5] A. Aghagolzadeh, S. Meshgini, M. Nooshyar, and M. Aghagolzadeh, "A novel video compression technique for very low bit-rate coding by combining H.264/AVC standard and 2-D wavelet transform," in *Proc. IEEE Int. Conf. on Signal Processing (ICSP)*, Beijing, China, 2008, pp. 1251–1254.

[6] Official web-site. [Online]. Available: http://diracvideo.org/

[7] M. Tun, K. K. Loo, and J. Cosmas, "Rate control algorithm based on quality factor optimization for Dirac video codec," *Signal Process.: Image Commun.*, vol. 23, pp. 649–664, 2008.

[8] O. M. López, M. O. M. Rach, P. Pinol, M. P. Malumbres, and J. Oliver, "M-LTW: A fast and efficient intra video codec," *Signal Process.: Image Commun.*, vol. 23, pp. 637–648, 2008.

[9] A. Golwelkar and J. W. Woods, "Motion-compensated temporal filtering and motion vector coding using biorthogonal filters," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 17, no. 4, pp. 417–428, 2007.

[10] R. J. P. de Figueiredo, "A generalized Fock space framework for nonlinear system and signal analysis," *IEEE Transactions on Circuits and Systems*, vol. 30, no. 9, pp. 637–647, 1983.

[11] C. Gu and C. Qiu, "Smoothing spline density estimation: Theory," *Ann. Stat.*, vol. 21, no. 1, pp. 217–234, 1993.

[12] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1522–1531, 2000.

[13] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.

[14] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors," *IEEE Trans. Info. Theory*, vol. 55, pp. 909–919, 1999.

[15] E. P. Simoncelli and E. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, Lusanne, Switzerland, 1996, pp. 279–382.

[16] N. Gupta, E. I. Plotkin, and M. N. S. Swamy, "Bayesian algorithm for video noise reduction in the wavelet domain," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, Kobe, Japan, 2005, pp. 4943–4946.

[17] M. K. Mihçak, "Information hiding codes and their applications to images and audio," Ph.D. dissertation, Univ. of Illinois at Urbana-Champaign, IL, 2002.

[18] P. Loo, "Digital watermarking using complex wavelets," Ph.D. dissertation, Univ. of Cambridge, UK, 2002.

[19] J. R. Hernández, M. Amado, and F. P. González, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. Image Processing*, vol. 9, no. 1, pp. 55–68, 2000.

[20] J. Cannons and P. Moulin, "Design and statistical analysis of a hash-aided image watermarking system," *IEEE Trans. Image Processing*, vol. 13, no. 10, pp. 1393–1408, 2004.

[21] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, 2003.

[22] J. M. Fadili and L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K-form densities," *IEEE Trans. Image Processing*, vol. 14, no. 2, pp. 231–240, 2005.

[23] S. Solbo and T. Eltoft, "Homomorphic wavelet-based statistical despeckling of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 4, pp. 711–721, 2004.

[24] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 92, no. 440, pp. 1413–1421, 1997.

[25] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[26] V. Strela, J. Portilla, and E. P. Simoncelli, "Image denoising via a local Gaussian scale mixture model in the wavelet domain," in *Proc. SPIE 45th Annu. Meeting*, San Diego, CA, 2002.

[27] H. Rabbani and M. Vafadust, "Image/video denoising based on a mixture of Laplace distributions with local parameters in multidimensional complex wavelet domain," *Signal Process.*, vol. 88, pp. 158–173, 2008.

[28] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 5th ed. NJ: Prentice Hall, 1995.

[29] T. T. Pham and R. J. P. de Figueiredo, "Maximum likelihood estimation of a class of non-Gaussian densities with application to l_p deconvolution," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 73–82, 1989.

[30] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Processing*, vol. 11, no. 2, pp. 146–158, 2002.

[31] B. Aiazzi, L. Alparone, and S. Baronti, "Estimation based on entropy matching for generalized Gaussian PDF modeling," *IEEE Signal Processing Lett.*, vol. 6, no. 6, pp. 138–140, 1999.

[32] L. P. Hansen, "Large sample properties of generalized method of moments estimator," *Econometrica*, vol. 50, no. 4, pp. 1029–1054, 1982.

[33] G. G. Walter, "Properties of Hermite series estimation of probability density," *Ann. Stat.*, vol. 5, no. 6, pp. 1258–1264, 1977.

[34] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Improved image restoration using wavelet-based denoising and Fourier-based deconvolution," in *IEEE International Mid-West Symposium on Circuits and Systems (MWSCAS)*, Knoxville, TN, 2008, pp. 249–252.

[35] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Mag.*, vol. 22, no. 6, pp. 123–151, 2005.

[36] F. C. A. Fernandes, R. L. C. V. Spaendonck, and C. S. Burrus, "A new framework for complex wavelet transforms," *IEEE Signal Processing*, vol. 51, no. 7, pp. 1825–1837, 2003.

[37] D. Clonda, J. M. Lina, and B. Goulard, "Complex Daubechies wavelets: Properties and statistical image modelling," *Signal Process.*, vol. 84, pp. 1–23, 2004.

[38] S. Barber and G. P. Nason, "Real nonparametric regression using complex wavelets," *J. R. Stat. Soc. B*, vol. 66, no. 4, pp. 927–939, 2004.

[39] G. Y. Chen and B. Kégl, "Image denoising with complex ridgelets," *Pattern Recognition*, vol. 40, pp. 578–585, 2007.

[40] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[41] R. Eslami and H. Radha, "Translation-invariant contourlet transform and its application to image denoising," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3362–3374, 2006.

[42] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, Thessaloniki, Greece, 2001, pp. 37–40.

[43] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Trans. Image Processing*, vol. 14, no. 7, pp. 948–959, 2005.

[44] I. Daubechies, *Ten Lectures on Wavelets*, 1st ed. Philadelphia, PA: SIAM, 1992.

[45] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA: Academic Press, 1999.

[46] J. C. Goswami and A. K. Chan, *Fundamentals of Wavelets, Theory, Algorithms, and Application*, 1st ed. NY: John Wiley & Sons, 1999.

[47] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[48] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, 1st ed. Wellesley, MA: Wellesley-Cambridge Press, 1996.

[49] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Delhi, India: Pearson Education Asia, 2002.

[50] M. Unser, "Ten good reasons for using spline wavelets," in *Proc. SPIE, Wavelets: Applications in Signal and Image Processing V*, vol. 3169, San Diego, 1997, pp. 422–431.

[51] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer-Verlag, 1995.

[52] N. G. Kingsbury, "Image processing with complex wavelets," *Phil. Trans. R. Soc. London A*, vol. 357, no. 1760, pp. 2543 – 2560, 1999.

[53] N. G. Kingsbury, "Complex wavelets for shift invariance analysis and filtering of signals," *Applied Comput. Harmon. Anal.*, vol. 10, no. 3, pp. 234–253, 2001.

[54] I. W. Selesnick, "Hilbert transform pairs of wavelet bases," *IEEE Signal Processing Lett.*, vol. 8, no. 6, pp. 170–173, 2001.

[55] I. W. Selesnick, "The design of approximate Hilbert transform pairs of wavelet bases," *IEEE Signal Processing*, vol. 50, no. 5, pp. 1140–1152, 2002.

[56] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vision*, vol. 40, pp. 49–71, 2000.

[57] H. Olkkonen, P. Pesola, J. Olkkonen, and H. Zhou, "Hilbert transform assisted complex wavelet transform for neuroelectric signal analysis," *J. Neurosci. Methods*, vol. 151, pp. 106–113, 2006.

[58] R. A. Gopinath, "The phaselet transform - An integral redundancy nearly shift-invariant wavelet transform," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1792–1805, 2003.

[59] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual-tree M-band wavelet transform," *IEEE Trans. Image Processing*, vol. 15, no. 8, pp. 2397–2412, 2006.

[60] D. B. H. Tay, N. G. Kingsbury, and M. Palaniswami, "Orthonormal Hilbert-pair of wavelets with (almost) maximum vanishing moments," *IEEE Signal Processing Lett.*, vol. 13, no. 9, pp. 533–536, 2006.

[61] H. Olkkonen and J. T. Olkkonen, "Half-delay B-spline filter for construction of shift-invariant wavelet transform," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 7, pp. 611–615, 2007.

[62] D. L. Ruderman and W. Bailek, "Scaling of natural images: Scaling in the woods," *Physic. Rev. Lett.*, vol. 73, no. 6, pp. 814–817, 1994.

[63] D. J. Field, "Relation between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, 1987.

[64] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Processing*, vol. 10, no. 11, pp. 1647–1658, 2001.

[65] L. Şendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.

[66] L. Şendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Lett.*, vol. 9, no. 12, pp. 438–441, 2002.

[67] A. Achim and E. E. Kuruoğlu, "Image denoising using bivariate α-stable distributions in the complex wavelet domain," *IEEE Signal Processing Lett.*, vol. 12, no. 1, pp. 17–20, 2005.

[68] L. Zhang, P. Bao, and X. Wu, "Multiscale LMMSE-based image denoising with optimal wavelet selection," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 15, no. 4, pp. 469–481, 2005.

[69] D. Cho and T. D. Bui, "Multivariate statistical modeling for image denoising using wavelet transforms," *Signal Process.: Image Commun.*, vol. 20, pp. 77–89, 2005.

[70] P. Scheunders and S. D. Backer, "Wavelet denoising of multicomponent images using Gaussian scale mixture models and a noise-free image as priors," *IEEE Trans. Image Processing*, vol. 16, no. 7, pp. 1865–1872, 2007.

[71] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Trans. Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[72] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[73] M. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," *Adv. Neural Inform. Process. Syst.*, vol. 12, 2000.

[74] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S. C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vision*, vol. 18, pp. 17–33, 2003.

[75] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, 10th ed., Dover, NY, 1965.

[76] M. I. H. Bhuiyan, M. O. Ahmad, and M. N. S. Swamy, "Wavelet-based despeckling of ultrasound images with the symmetric normal inverse Gaussian prior," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 1, Hawaii, USA, 2007, pp. 721–724.

[77] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Bayesian wavelet-based image denoising using the Gauss-Hermite expansion," *IEEE Trans. Image Processing*, vol. 17, no. 10, pp. 1755–1771, 2008.

[78] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Bayesian wavelet-based despeckling of ultrasound medical images using the Gauss-Hermite expansion," in *Proc. Joint Statistical Meetings*, Salt Lake City, UT, 2007, pp. 1718–1725.

[79] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Locally adaptive wavelet-based image denoising using the Gram-Charlier prior function," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, San Antonio, TX, 2007, pp. 549–552.

[80] A. Zayezdny, D. Tabak, and D. Wulich, *Engineering Applications of Stochastic Processes*, 1st ed. NY: John Wiley & Sons, 1989.

[81] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 1, Distribution Theory*, 4th ed. London: Charles Griffin, 1977.

[82] E. Jondeau and M. Rockinger, "Gram-Charlier densities," *J. Econ. Dyn. Control*, vol. 25, pp. 1457–1483, 2001.

[83] D. E. Barton and K. E. Dennis, "The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal," *Biometrika*, vol. 39, no. 3/4, pp. 425–427, 1952.

[84] J. S. Simonoff, *Smoothing Methods in Statistics*, 1st ed. NY: Springer-Verlag, 1996.

[85] [Online]. Available: http://decsai.ugr.es/cvg/dbimagenes/index.php

[86] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Statistics of 2D DT-CWT coefficients for Gaussian distributed signal," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 7, pp. 2013–2025, 2008.

[87] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "A new bivariate MAP estimator for DT-CWT-based video denoising," in *IEEE Int. Symp. Circuits and Systems (ISCAS): Accepted - paper number 2223*, Taipei, Taiwan, ROC, 2009.

[88] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, 1st ed. NJ: Prentice Hall, 1998.

[89] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. NY: McGraw-Hill, 1991.

[90] Y. S. Kim, O. H. Kwon, and R. H. Park, "Wavelet based watermarking method for digital images using the human visual system," *Electron. Lett.*, vol. 35, no. 6, pp. 466–468, 1999.

[91] M. J. Wainwright, E. P. Simoncelli, and A. W. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *Applied Comput. Harmon. Anal.*, vol. 11, pp. 89–123, 2001.

[92] L. Boubchir and J. M. Fadili, "A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior," *Pattern Recognition Lett.*, vol. 27, pp. 1370–1382, 2006.

[93] P. Abry and P. Flandrin, "Multiresolution transient detection," in *Proc. Int. Symp. Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, 1994, pp. 225–228.

[94] M. K. Simon, *Probability Distributions Involving Gaussian Random Variables : A Handbook for Engineers and Scientists*, 1st ed. London: Kluwer Press, 2002.

[95] H. Choi, J. K. Romberg, R. G. Baraniuk, and N. G. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, Istanbul, Turkey, 2000, pp. 133–136.

[96] C. W. Shaffrey, N. G. Kingsbury, and I. H. Jermyn, "Unsupervised image segmentation via Markov trees and complex wavelets," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, Rochester, NY, 2002, pp. 801–804.

[97] R. J. Ferrari and R. Winsor, "Digital radiographic image denoising via wavelet-based hidden Markov model estimation," *J. Digital Imaging*, vol. 18, no. 2, pp. 154–167, 2005.

[98] J. K. Romberg, H. Choi, and R. Baraniuk, "Multiscale edge grammars for complex wavelet transforms," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 1, Thessaloniki, Greece, 2001, pp. 614–617.

[99] M. Miller and N. Kingsbury, "Statistical image modelling using interscale phase relationships of complex wavelet coefficients," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 2, Toulouse, France, 2006, pp. 789–792.

[100] J. Romberg, H. Choi, R. Baraniuk, and N. Kingsbury, "Multiscale classification using complex wavelets and hidden Markov tree models," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, Vancouver, BC, 2000, pp. 371–374.

[101] B. Wang, Y. Wang, I. Selesnick, and A. Vetro, "An investigation of 3D dual-tree wavelet transform for video coding," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, Singapore, 2004, pp. 1317–1320.

[102] C. Chaux, L. Duval, and J.-C. Pesquet, "Noise covariance properties in dual-tree wavelet decompositions," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 680–700, 2007.

[103] T. Aach, "Comparative analysis of shift variance and cyclostationarity in multirate filter banks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 5, pp. 1077–1087, 2007.

[104] A. Cavallini, R. Langella, A. Testa, and F. Ruggiero, "Gaussian modeling of harmonic vectors in power systems," in *Proc. 8th IEEE Int. Conf. on Harmonics and Quality of Power*, vol. 2, Athens, Greece, 1998, pp. 1010–1017.

[105] E. W. Stacy, "A generalization of the Gamma distribution," *Annals of Math. Stat.*, vol. 33, no. 3, pp. 1187–1192, 1962.

[106] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Lett.*, vol. 6, no. 12, pp. 300–303, 1999.

[107] S. M. M. Rahman and M. K. Hasan, "Wavelet-domain iterative center weighted median filter for image denoising," *Signal Process.*, vol. 83, pp. 1001–1012, 2003.

[108] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Wavelet-domain image denoising algorithm using series expansion of coefficient P.D.F. in terms

of Hermite polynomials," in *Proc. 3rd. Int. IEEE-NEWCAS Conf.*, Québec, Canada, 2005, pp. 271–275.

[109] M. I. H. Bhuiyan, M. O. Ahmad, and M. N. S. Swamy, "New spatially adaptive wavelet-based method for the despeckling of medical ultrasound images," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 1, New Orleans, 2007, pp. 2347–2350.

[110] S. Gupta, R. C. Chauhan, and S. C. Sexana, "Locally adaptive wavelet domain Bayesian processor for denoising medical ultrasound images using speckle modelling based on Rayleigh distribution," *IEE Proc. Vis. Image Signal Process.*, vol. 152, pp. 129–135, 2005.

[111] H. Cramer, *Mathematical methods of statistics*, 1st ed. NJ: Princeton University Press, 1946.

[112] J. M. Niya and A. Aghagolzadeh, "Edge detection using directional wavelet transform," in *Proc. IEEE Mediterranean Electrotech. Conf.*, vol. 1, Croatia, 2004, pp. 281–284.

[113] Z. Xiao and Z. Hou, "Phase based feature detector consistent with human visual system characteristics," *Pattern Recognition Lett.*, vol. 25, pp. 1115–1121, 2005.

[114] C. Boncelet, "Image noise models," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. NY: Academic Press, 2000.

[115] R. M. Rangayyan, M. Ciuc, and F. Faghih, "Adaptive-neighbourhood filtering of images corrupted by signal-dependent noise," *Appl. Opt.*, vol. 37, no. 20, pp. 4477–4487, 1998.

[116] H. H. Arsenault, C. Gendron, and M. Denis, "Transformation of film-grain noise into signal-independent gaussian noise," *Appl. Opt.*, vol. 23, pp. 845–850, 1984.

[117] J. S. Lim and H. Nawab, "Techniques for speckle noise removal," *Opt. Eng.*, vol. 21, pp. 472–480, 1981.

[118] M. Kazubek, "Wavelet domain image denoising by thresholding and Wiener filtering," *IEEE Signal Processing Lett.*, vol. 10, no. 11, pp. 324–326, 2003.

[119] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Info. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[120] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[121] H. Zhang, A. Nosralinia, and J. R. O. Wells, "Image denoising via wavelet-domain spatially adaptive FIR Wiener filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 5, Istanbul, Turkey, 2000, pp. 2179–2182.

[122] Z. Cai, T. H. Cheng, C. Lu, and K. R. Subramanium, "Efficient wavelet-based image denoising algorithm," *Electron. Lett.*, vol. 37, no. 11, pp. 683–685, 2001.

[123] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Info. Theory*, vol. 45, pp. 846–862, 1999.

[124] G. Fan and X. G. Xia, "Image denoising using local contextual hidden Markov model in the wavelet domain," *IEEE Signal Processing Lett.*, vol. 8, no. 5, pp. 125–128, 2001.

[125] 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. 3: no. 4, pp. 600–612, 2004.

[126] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.

[127] A. Pizurica and W. Philips, "Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising," *IEEE Trans. Image Processing*, vol. 15, no. 3, pp. 654–665, 2006.

[128] B. Vidakovic, *Statistical Modeling by Wavelets*, 1st ed. NY: John Wiley & Sons, 1999.

[129] A. Antoniadis and J. Bigot, "Wavelet estimators in nonparametric regression: A comparative simulation study," *Journal of Statistical Software*, vol. 6, no. 6, 2001.

[130] M. A. T. Figueiredo and R. D. Nowak, "Wavelet-based image estimation: An empirical Bayes approach using Jeffreys' noninformative prior," *IEEE Trans. Image Processing*, vol. 10, no. 9, pp. 1322–1331, 2001.

[131] T. Cai and B. Silverman, "Incorporating information on neighboring coefficients into wavelet estimation," *Sankhya: The Indian Journal of Statistics*, vol. 63, pp. 127–148, 2001.

[132] P. Bao and X. Ma, "Image adaptive watermarking using wavelet domain singular value decomposition," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 15, no. 1, pp. 96–102, 2005.

[133] I. M. Johnstone and B. W. Silverman, "Empirical Bayes selection of wavelet thresholds," *Ann. Stat.*, vol. 33, no. 4, pp. 1700–1752, 2005.

[134] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[135] H.-Y. Gao and A. G. Bruce, "Waveshrink with firm shrinkage," *Statistica Sinica*, vol. 7, no. 4, pp. 855–874, 1997.

[136] H.-Y. Gao, "Wavelet shrinkage denoising using the non-negative garrote," *J. Comp. Graph. Stat.*, vol. 7, no. 4, pp. 469–488, 1998.

[137] M. I. H. Bhuiyan, M. O. Ahmad, and M. N. S. Swamy, "Spatially adaptive wavelet-based method using the Cauchy prior for denoising the SAR images," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 17, no. 4, pp. 500–507, 2007.

[138] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Video denoising based on inter-frame statistical modeling of wavelet coefficients," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 17, no. 2, pp. 187–198, 2007.

[139] J. Huang, "Statistics of natural images and models," Ph.D. dissertation, Brown Univ., Providence, RI, 2000.

[140] A. D. Stefano, P. R. White, and W. B. Collis, "An innovative approach for spatial video noise reduction using a wavelet based frequency decomposition," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 3, Vancouver, BC, 2000, pp. 281–284.

[141] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. R. Stat. Soc. B*, vol. 60, no. 4, pp. 725–749, 1998.

[142] M. Clyde and E. I. George, "Flexible empirical Bayes estimation for wavelets," *J. R. Stat. Soc. B*, vol. 62, no. 4, pp. 681–698, 2000.

[143] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imaging*, vol. 20, no. 8, pp. 772–783, 2001.

[144] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed., A. Jeffrey, Ed. NY: Academic Press, 1994.

[145] C. D. Boor, *A Practical Guide to Splines*, 1st ed., F. John, L. Sirovich, J. P. Lasalle, and G. B. Whitham, Eds. NY: Springer-Verlag, 1978.

[146] A. Hyverinen, "Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation," *Neural Computation*, vol. 11, no. 7, pp. 1739–1768, 1999.

[147] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "A new statistical detector for DWT-based additive image watermarking using the Gauss-Hermite expansion," *in review of IEEE Trans. Image Processing - Submission number: TIP-04415-2008.R1.*

[148] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Statistical detector for wavelet-based image watermarking using modified GH PDF," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Seattle, WA, 2008, pp. 712–715.

[149] G. Voyatzis and J. Pitas, "Image watermarking for copyright protection and authentication," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. NY: Academic Press, 2000.

[150] G. C. Langelaar, I. Setyawaan, and R. L. Lagendijk, "Watermarking digital image and video data: A state-of-the-art overview," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, 2000.

[151] P. Moulin and R. Koetter, "Data-hiding codes," *Proc. IEEE*, vol. 93, no. 12, pp. 2083–2126, 2005.

[152] J. R. Hernández and F. P. González, "Statistical analysis of watermarking schemes for copyright protection of images," *Proc. IEEE*, vol. 87, no. 7, pp. 1142–1166, 1999.

[153] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," in *Proc. IEEE*, vol. 87, no. 7, 1999, pp. 1108–1126.

[154] C. I. Podilchuk and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE J. Selected Areas Commun.*, vol. 16, no. 4, pp. 525–539, 1998.

[155] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Multiresolution scene-based video watermarking using perceptual models," *IEEE J. Selected Areas Commun.*, vol. 16, no. 4, pp. 540–550, 1998.

[156] N. Kaewkamnerd and K. Rao, "Wavelet based image adaptive watermarking scheme," *Electron. Lett.*, vol. 36, no. 4, pp. 312–313, 2000.

[157] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 783–791, 2001.

[158] P. C. Su, H. J. M. Wang, and C. C. J. Kuo, "An integrated approach to image watermarking and JPEG-2000 compression," *J. VLSI Signal Processing*, vol. 27, no. 1-2, pp. 35–53, 2001.

[159] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp: 273–284, 2001.

[160] A. K. Mairgiotis, N. P. Galatsanos, and Y. Yang, "New additive watermark detectors based on a hierarchical spatially adaptive image model," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 29–37, 2008.

[161] W. Zhu, Z. Xiong, and Y. Q. Zhang, "Multiresolution watermarking for images and video," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 9, no. 4, pp. 545–550, 1999.

[162] X. Huang and B. Zhang, "Statistically robust detection of multiplicative spread-spectrum watermarks," *IEEE Trans. Information Forensics and Security*, vol. 2, no. 1, pp. 1–13, 2007.

[163] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Info. Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[164] S. H. Wang and Y. P. Lin, "Wavelet tree quantization for copyright protection watermarking," *IEEE Trans. Image Processing*, vol. 13, no. 2, pp. 154–165, 2004.

[165] O. E. Okman and G. B. Akar, "Quantization index modulation-based image watermarking using digital holography," *J. Opt. Soc. Amer.*, vol. 24, no. 1, pp. 243–252, 2007.

[166] W. Zeng and B. Liu, "A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images," *IEEE Trans. Image Processing*, vol. 8, no. 11, pp. 1534–1548, 1999.

[167] S. Craver, N. Memon, B. L. Yeo, and M. M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications," *IEEE J. Selected Areas Commun.*, vol. 16, no. 4, pp. 573–586, 1998.

[168] Q. Cheng and T. S. Huang, "Blind digital watermarking for images and videos and performance analysis," in *Proc. IEEE Int. Conf. on Multimedia and Expo (ICME)*, vol. 1, NY, 2000, pp. 389–392.

[169] J. J. Eggers and B. Girod, "Quantization effects on digital watermarks," *Signal Process.*, vol. 81, pp. 239–263, 2001.

[170] F. A. P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 58–64, 2000.

[171] A. Briassouli, P. Tsakalides, and A. Stouraitis, "Hidden messages in heavy-tails: DCT-domain watermark detection using alpha-stable models," *IEEE Trans. Multimedia*, vol. 7, no. 4, pp. 700–715, 2005.

[172] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[173] H. J. M. Wang, P. C. Su, and C. C. J. Kuo, "Wavelet-based digital image watermarking," *Opt. Express.*, vol. 3, no. 12, pp. 491–496, 1998.

[174] W. Liu, L. Dong, and W. Zeng, "Optimum detection for spread-spectrum watermarking that employs self-masking," *IEEE Trans. Information Forensics and Security*, vol. 2, no. 4, pp. 645–654, 2007.

[175] V. R. Doncel, N. Nikolaidis, and I. Pitas, "An optimal detector structure for the Fourier descriptors domain watermarking of 2D vector graphics," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 5, pp. 851–863, 2007.

[176] M. S. Hsieh, D. C. Tseng, and Y. H. Huang, "Hiding digital watermarks using multiresolution wavelet transform," *IEEE Trans. Industrial Electronics*, vol. 48, no. 5, pp. 875–882, 2001.

[177] X. Kang, J. Huang, Y. Q. Shi, and Y. Lin, "A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 13, no. 8, pp. 776–786, 2003.

[178] Y. Zhao, P. Campisi, and D. Kundur, "Dual domain watermarking for authentication and compression of cultural heritage images," *IEEE Trans. Image Processing*, vol. 13, no. 3, pp. 430–448, 2004.

[179] V. M. Potdar, S. Han, and E. Chang, "A survey of digital image watermarking techniques," in *IEEE Int. Conf. Industrial Informatics (INDIN)*, Perth, WA, Australia, 2005, pp. 709–716.

[180] A. Nikolaidis and I. Pitas, "Asymptotically optimal detection for additive watermarking in the DCT and DWT domains," *IEEE Trans. Image Processing*, vol. 12, no. 5, pp. 563–571, 2003.

[181] C. T. Hsu and J. L. Wu, "Multiresolution watermarking for digital images," *IEEE Trans. Circuits Sys. II - Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1097–1101, 1998.

[182] L. Ghouti, A. Bouridane, M. K. Ibrahim, and S. Boussakta, "Digital image watermarking using balanced multiwavelets," *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1519–1536, 2006.

[183] C. V. Serdean, M. K. Ibrahim, A. Moemeni, and M. M. Al-Akaidi, "Wavelet and multiwavelet watermarking," *IET Image Processing*, vol. 1, no. 2, pp. 223–230, 2007.

[184] N. Bi, Q. Sun, D. Huang, Z. Yang, and J. Huang, "Robust image watermarking based on multiband wavelets and empirical mode decomposition," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 1956–1966, 2007.

[185] K. M. Parker and J. E. Fowler, "Redundant-wavelet watermarking with pixel-wise masking," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 1, Genoa, Italy, 2005, pp. 685–688.

[186] A. A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme," *Pattern Recognition Letters*, vol. 26, pp. 1019–1027, 2005.

[187] A. Giannoula, N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, "Watermark detection for noisy interpolated images," *IEEE Trans. Circuits Sys. II: Express Briefs*, vol. 53, no. 5, pp. 359–363, 2006.

[188] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. NY: Springer-Verlag, 1994.

[189] M. Barni, F. Bartolini, A. D. Rosa, and A. Piva, "A new decoder for the optimum recovery of nonadditive watermarks," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 755–766, 2001.

[190] T. M. Ng and H. K. Garg, "Maximum likelihood detection in image watermarking using generalized Gamma model," in *Proc. Thirty-Ninth Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 2005, pp. 1680–1684.

[191] T. M. Ng and H. K. Garg, "Maximum-likelihood detection in DWT domain image watermarking using Laplacian modeling," *IEEE Signal Processing Lett.*, vol. 12, no. 4, pp. 285–288, 2005.

[192] A. K. Mairgiotis, G. Chantas, N. P. Galatsanos, K. Blekas, and Y. Yang, "New detectors for watermarks with unknown power based on Student-t image priors," in *Proc. IEEE Int. Workshop on Multimedia Signal Processing*, Crete, Greece, 2007, pp. 353–356.

[193] A. Briassouli and M. G. Strintzis, "Locally optimum nonlinearities for DCT watermark detection," *IEEE Trans. Image Processing*, vol. 13, no. 12, pp. 1604–1617, 2004.

[194] J. Wang, G. Liu, Y. Dai, J. Sun, Z. Wang, and S. Lian, "Locally optimum detection for Barni's multiplicative watermarking in DWT domain," *Signal Process.*, vol. 88, no. 1, pp. 117–130, 2008.

[195] X. Y. Liu, G. Kun, and W. F. Chen, "A blind watermarking optimal detection based on the wavelet transform domain," in *Proc. IEEE Int. Conf. Machine Learning and Cybernetics*, Hong Kong, 2007, pp. 1779–1783.

[196] J. K. Su, J. J. Eggers, and B. Girod, "Analysis of digital watermarks subjected to optimum linear filtering and additive noise," *Signal Process.*, vol. 81, pp. 1141–1175, 2001.

[197] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, 1st ed. NY: Springer-Verlag, 1988.

[198] S. Voloshynovskiy, S. Pereira, V. Iquise, and T. Pun, "Attack modelling: Towards a second generation watermarking benchmark," *Signal Process.*, vol. 81, pp. 1177–1214, 2001.

[199] A. Said and A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 6, no. 3, pp. 243–250, 1996.

[200] [Online]. Available: http://www.cipr.rpi.edu/research/SPIHT/spiht0.html

[201] F. Jin, P. Fieguth, and L. Winger, "Wavelet video denoising with regularized multiresolution motion estimation," *EURASIP J. Applied Signal Processing*, vol. 2006, no. 72705, pp. 1–11, 2006.

[202] J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos, and R. L. Lagendijk, "Noise reduction filters for dynamic image sequences: A review," in *Proc. IEEE*, vol. 83, no. 9, 1995, pp. 1272–1292.

[203] R. L. Lagendijk, P. M. B. V. Roosmalen, and J. Biemond, "Video enhancement and restoration," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. NY: Academic Press, 2000.

[204] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 1st ed. NJ: Prentice-Hall, 1982.

[205] A. C. Kokaram, *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*, 1st ed. London: Springer-Verlag, 1998.

[206] M. K. Ozkan, M. I. Sezan, and A. M. Tekalp, "Adaptive motion-compensated filtering of noisy image sequences," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 3, no. 4, pp. 277–290, 1993.

[207] R. P. Kleihorst, R. L. Lagendijk, and J. Biemond, "Noise reduction of image sequences using motion compensation and signal decomposition," *IEEE Trans. Image Processing*, vol. 4, no. 3, pp. 274–284, 1995.

[208] J. W. Woods and J. Kim, "Motion-compensated spatiotemporal Kalman filter," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds. MA: Kluwer Press, 1993.

[209] F. Cocchia, S. Carrato, and G. Ramponi, "Design and real-time implementation of a 3-D rational filter for edge preserving smoothing," *IEEE Trans. Consumer Electronics*, vol. 43, no. 4, pp. 1291–1300, 1997.

[210] V. Zlokolica, W. Philips, and D. V. D. Ville, "A new non-linear filter for video processing," in *Proc. IEEE Benelux Signal Processing Symposium*, vol. 2, Belgium, 2002, pp. 221–224.

[211] P. Rieder and G. Scheffler, "New concepts on denoising and sharpening of video signals," *IEEE Trans. Consumer Electronics*, vol. 47, no. 3, pp. 666–671, 2001.

[212] L. Tenze, S. Carrato, and S. Olivieri, "Design and real-time implementation a low-cost noise reduction system for video applications," *Signal Process.*, vol. 84, no. 3, pp. 453–466, 2004.

[213] T. W. Chan, O. C. Au, T. S. Chong, and W. S. Chau, "A novel content-adaptive video denoising filter," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, Philadelphia, PA, 2005, pp. 649–652.

[214] H. D. Tagare and R. J. P. de Figueiredo, "Order filters," *Proc. IEEE*, vol. 73, no. 1, pp. 163–165, 1985.

[215] P. M. B. V. Roosmalen, S. J. P. Westen, R. L. Lagendijk, and J. Biemond, "Noise reduction for image sequences using an oriented pyramid thresholding technique," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 1, Lausanne, Switzerland, 1996, pp. 375–378.

[216] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Proc. SPIE, Wavelets: Applications in Signal and Image Processing X*, vol. 5207, San Diego, 2003, pp. 607–618.

[217] N. Rajpoot, Z. Yao, and R. Wilson, "Adaptive wavelet restoration of noisy video sequences," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 2, Singapore, 2004, pp. 24–27.

[218] S. M. M. Rahman, M. O. Ahmad, and M. N. S. Swamy, "Wavelet-based video denoising using Gauss-Hermite density function," in *IEEE International Mid-West Symposium on Circuits and Systems (MWSCAS)*, San Juan, PR, 2006, pp. 592–595.

[219] E. J. Balster, Y. F. Zheng, and R. L. Ewing, "Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising," *IEEE Trans. Circuits Sys. Video Tech.*, vol. 16, no. 2, pp. 220–230, 2006.

[220] A. Pizurica, V. Zlokolica, and W. Philips, "Noise reduction in video sequences using wavelet-domain and temporal filtering," in *Proc. SPIE, Wavelet Applications in Industrial Processing*, vol. 5266, Providence, RI, 2003, pp. 48–59.

[221] A. Pizurica, V. Zlokolica, and W. Philips, "Combined wavelet domain and temporal video denoising," in *IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, Miami, FL, 2003, pp. 334–341.

[222] H. Y. Cheong, A. M. Tourapis, J. Liach, and J. Boyce, "Adaptive spatio-temporal filtering for video de-noising," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, vol. 2, Singapore, 2004, pp. 965–968.

[223] V. Zlokolica, A. Pizurica, and W. Philips, "Video denoising using multiple class averaging with multiresolution," in *Int. Workshop on Very Low Bitrate Video Coding*, Madrid, Spain, 2003, pp. 172–179.

[224] E. J. Balster, Y. F. Zheng, and R. L. Ewing, "Fast, feature-based wavelet shrinkage algorithm for image denoising," in *Proc. IEEE Int. Conf. on Integration of Knowledge Intensive Multi-Agent Systems*, vol. 1, 2003, pp. 722–728.

[225] I. K. Eom and Y. S. Kim, "Wavelet-based denoising with nearly arbitrarily shaped windows," *IEEE Signal Processing Lett.*, vol. 11, no. 12, pp. 937–940, 2004.

Appendix A

PDF of DWT Coefficients of Images Using Gauss-Hermite Expansion

A-1 Series Parameters in GH Expansion

Let $\bar{f}$ be the normalized version of f with the relation $\bar{f} = f/\sigma_f$ and its PDF is $p_{\bar{f}}(\bar{f})$. Since the Hermite polynomials are orthogonal with respect to the zero-mean unit variance Gaussian weight function, $p_{\bar{f}}(\bar{f})$ can be written as [81]

$$p_{\bar{f}}(\bar{f}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{f}^2}{2}} \sum_{r=0}^{\infty} \beta_r H_r(\bar{f}) \tag{A-1-1}$$

where β_r are the series coefficients. Multiplying both sides of (A-1-1) by $H_r(\bar{f})$ and integrating from $-\infty$ to ∞, β_r can be evaluated as

$$\beta_r = \frac{1}{r!} \int_{-\infty}^{\infty} p_{\bar{f}}(\bar{f}) H_r(\bar{f}) \mathrm{d}\bar{f} \tag{A-1-2}$$

Using the fact that $\bar{f}$ has a zero mean and unit variance, one can obtain $\beta_0 = 1$ and $\beta_1 = \beta_2 = 0$. Finally, changing the variable of $\bar{f}$ to f in (A-1-1) yields [80]

$$p_f(f) = \frac{1}{\sigma_f \sqrt{2\pi}} e^{-\frac{f^2}{2\sigma_f^2}} \left[1 + \sum_{r=3}^{\infty} \beta_r H_r\left(\frac{f}{\sigma_f}\right) \right] \tag{A-1-3}$$

The Hermite polynomials $H_r(f)$ can be expressed as

$$H_r(f) = \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^u r!}{2^u (r-2u)! u!} f^{r-2u} \tag{A-1-4}$$

Using the identity $p_{\boldsymbol{f}}(\boldsymbol{f})\mathrm{d}\boldsymbol{f} = p_{\bar{\boldsymbol{f}}}(\bar{\boldsymbol{f}})\mathrm{d}\bar{\boldsymbol{f}}$ and the expression of $H_r(\boldsymbol{f})$, the series coefficients β_r in (A-1-2) can be obtained in a generalized form

$$\begin{aligned}\beta_r &= \frac{1}{r!}\int_{-\infty}^{\infty} p_{\boldsymbol{f}}(\boldsymbol{f}) H_r\Big(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\Big)\mathrm{d}\boldsymbol{f} \\ &= \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^u}{2^u(r-2u)!u!}\frac{1}{\sigma_{\boldsymbol{f}}^{r-2u}}\int_{-\infty}^{\infty} \boldsymbol{f}^{r-2u} p_{\boldsymbol{f}}(\boldsymbol{f})\mathrm{d}\boldsymbol{f} \\ &= \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor} \frac{(-1)^u}{2^u(r-2u)!u!}\frac{M_{(r-2u)\boldsymbol{f}}}{\sigma_{\boldsymbol{f}}^{r-2u}} \end{aligned} \tag{A-1-5}$$

where $M_{n\boldsymbol{f}}$ is the n-th order moment of $\boldsymbol{f}$.

A-2 Parameter γ for MGH PDF

Using the normalized variable $\bar{\boldsymbol{f}}$, we can solve (3.8) as $\gamma = (1-2\Upsilon_1)/(1-2\Upsilon_2)$, where

$$\Upsilon_1 = \frac{1}{\sqrt{2\pi}}\int_0^{\bar{f}_0} e^{-\frac{\bar{f}^2}{2}}\left[1 + \sum_{r=2}^{\zeta/2}\beta_{2r}H_{2r}\left(\bar{\boldsymbol{f}}\right)\right]\mathrm{d}\bar{\boldsymbol{f}} \tag{A-2-1}$$

$$\Upsilon_2 = \frac{1}{\sqrt{2\pi}}\int_0^{\bar{f}_0} e^{-\frac{\bar{f}^2}{2}}\left[1 + \frac{K_m - 3}{4!}H_4\left(\bar{\boldsymbol{f}}\right)\right]\mathrm{d}\bar{\boldsymbol{f}} \tag{A-2-2}$$

Using the identity $\frac{\mathrm{d}}{\mathrm{d}\bar{\boldsymbol{f}}}\Big[H_{r+1}(\bar{\boldsymbol{f}})\Big] = (r+1)H_r(\bar{\boldsymbol{f}})$ and the recurrance relation in (3.2), we obtain

$$\int e^{-\frac{\bar{f}^2}{2}}H_{2r}(\bar{\boldsymbol{f}})\mathrm{d}\bar{\boldsymbol{f}} = -e^{-\frac{\bar{f}^2}{2}}H_{2r-1}(\bar{\boldsymbol{f}}) \tag{A-2-3}$$

Hence, for $r \geq 2$,

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_0^{\bar{f}_0} e^{-\frac{\bar{f}^2}{2}}H_{2r}(\bar{\boldsymbol{f}})\mathrm{d}\bar{\boldsymbol{f}} &= -\frac{1}{\sqrt{2\pi}}\left[e^{-\frac{\bar{f}_0^2}{2}}H_{2r-1}(\bar{f}_0) - H_{2r-1}(0)\right] \\ &= -\frac{1}{\sqrt{2\pi}}e^{-\frac{f_0^2}{2\sigma_{\boldsymbol{f}}^2}}H_{2r-1}\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right), \quad \text{since } H_{2r-1}(0) = 0 \\ &= I_{2r}\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right), \quad \text{(Say)} \end{aligned} \tag{A-2-4}$$

$$\text{Also,} \quad \frac{1}{\sqrt{2\pi}}\int_0^{\bar{f}_0} e^{-\frac{\bar{f}^2}{2}}\mathrm{d}\bar{\boldsymbol{f}} = \frac{1}{2}\mathrm{erf}\left(\frac{f_0}{\sqrt{2}\sigma_{\boldsymbol{f}}}\right) = I_0\left(\frac{f_0}{\sigma_{\boldsymbol{f}}}\right), \quad \text{(Say)} \tag{A-2-5}$$

Using the above, Υ_1 and Υ_2, and hence, the parameter γ can be expressed as

$$\gamma = \frac{1 - 2\left[I_0\left(\frac{f_0}{\sigma_f}\right) + \sum_{r=2}^{\zeta/2} \beta_{2r} I_{2r}\left(\frac{f_0}{\sigma_f}\right)\right]}{1 - 2\left[I_0\left(\frac{f_0}{\sigma_f}\right) + \frac{K_m - 3}{4!} I_4\left(\frac{f_0}{\sigma_f}\right)\right]} \tag{A-2-6}$$

Appendix B

MMSE Shrinkage Function for Image DWT Coefficients Using Gauss-Hermite Expansion

Using (5.5) and (5.6), the shrinkage function can be written as

$$\hat{\boldsymbol{f}}(\boldsymbol{g}) = \frac{\int p_{\boldsymbol{\varepsilon}}(\boldsymbol{g}-\boldsymbol{f})\hat{p}_{\boldsymbol{f}}(\boldsymbol{f})\boldsymbol{f}\mathrm{d}\boldsymbol{f}}{\int p_{\boldsymbol{\varepsilon}}(\boldsymbol{g}-\boldsymbol{f})\hat{p}_{\boldsymbol{f}}(\boldsymbol{f})\mathrm{d}\boldsymbol{f}} = \frac{\sum_{r=0}^{\zeta/2}\beta_{2r}\mathcal{I}_{nr}}{\sum_{r=0}^{\zeta/2}\beta_{2r}\mathcal{I}_{dr}} \tag{B-1-1}$$

where the integrals $\mathcal{I}_{nr}$ and $\mathcal{I}_{dr}$ are given by

$$\mathcal{I}_{nr} = \frac{1}{2\pi\sigma_{\boldsymbol{f}}\tilde{\sigma}_{\boldsymbol{\varepsilon}}}e^{-\frac{\boldsymbol{g}^2}{2\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}}\int_{-\infty}^{\infty}e^{\frac{\boldsymbol{g}\boldsymbol{f}}{\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}-\frac{\boldsymbol{f}^2}{2\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}}H_{2r}\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\boldsymbol{f}\mathrm{d}\boldsymbol{f} \tag{B-1-2}$$

$$\mathcal{I}_{dr} = \frac{1}{2\pi\sigma_{\boldsymbol{f}}\tilde{\sigma}_{\boldsymbol{\varepsilon}}}e^{-\frac{\boldsymbol{g}^2}{2\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}}\int_{-\infty}^{\infty}e^{\frac{\boldsymbol{g}\boldsymbol{f}}{\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}-\frac{\boldsymbol{f}^2}{2\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}-\frac{\boldsymbol{f}^2}{2\sigma_{\boldsymbol{f}}^2}}H_{2r}\left(\frac{\boldsymbol{f}}{\sigma_{\boldsymbol{f}}}\right)\mathrm{d}\boldsymbol{f} \tag{B-1-3}$$

Let us define $\lambda = \left(\frac{\sigma_{\boldsymbol{f}}^2}{\sigma_{\boldsymbol{f}}^2+\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}\right)^{\frac{1}{2}}$, $\vartheta = -\frac{\boldsymbol{g}}{\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}$, and $\varpi = \frac{1}{2\lambda^2\tilde{\sigma}_{\boldsymbol{\varepsilon}}^2}$. Using the expression of $H_{2r}(\boldsymbol{f})$, the integral $\mathcal{I}_{nr}$ can be rewritten as

$$\mathcal{I}_{nr} = \frac{1}{2\pi\sigma_{\boldsymbol{f}}\tilde{\sigma}_{\boldsymbol{\varepsilon}}}e^{\frac{\vartheta\boldsymbol{g}}{2}}\sum_{u=0}^{r}\frac{(-1)^u(2r)!}{2^u(2r-2u)!u!}\frac{1}{\sigma_{\boldsymbol{f}}^{2r-2u}}\int_{-\infty}^{\infty}e^{-\vartheta\boldsymbol{f}-\varpi\boldsymbol{f}^2}\boldsymbol{f}^{2r-2u+1}\mathrm{d}\boldsymbol{f} \tag{B-1-4}$$

Since $Re(\varpi) > 0$ and $Re(2r-2u+2) > 0\ \forall_r$, using the integral formula given in [144] (p. 382), $\mathcal{I}_{nr}$ in (B-1-4) can be evaluated as

$$\mathcal{I}_{nr} = \frac{1}{2\pi\sigma_{\boldsymbol{f}}\tilde{\sigma}_{\boldsymbol{\varepsilon}}}e^{\frac{\vartheta\boldsymbol{g}}{2}}\sum_{u=0}^{r}\frac{(-1)^u(2r)!(2r-2u+1)}{2^{r+1}u!\sigma_{\boldsymbol{f}}^{2r-2u}\varpi^{r-u+1}} \cdot\left[D_{-(2r-2u+2)}\left(\frac{\vartheta}{\sqrt{2\varpi}}\right)-D_{-(2r-2u+2)}\left(-\frac{\vartheta}{\sqrt{2\varpi}}\right)\right] \tag{B-1-5}$$

where $D_r(\cdot)$ is the parabolic cylinder function. Similarly we can obtain

$$\mathcal{I}_{dr} = \frac{1}{2\pi\sigma_f\tilde{\sigma}_\epsilon} e^{\frac{\vartheta g}{2}} \sum_{u=0}^{r} \frac{(-1)^u(2r)!}{2^{r+\frac{1}{2}}u!\sigma_f^{2r-2u}\varpi^{r-u+\frac{1}{2}}} \cdot \left[D_{-(2r-2u+1)}\left(\frac{\vartheta}{\sqrt{2\varpi}}\right) + D_{-(2r-2u+1)}\left(-\frac{\vartheta}{\sqrt{2\varpi}}\right)\right] \quad \text{(B-1-6)}$$

Substituting the values of ϑ and ϖ in (B-1-5) and (B-1-6), and then using (B-1-1), the final expression for the shrinkage function can be obtained as

$$\hat{\boldsymbol{f}}(\boldsymbol{g}) = \frac{\Delta_{\text{num}}}{\Delta_{\text{den}}} \quad \text{(B-1-7)}$$

where

$$\Delta_{\text{num}} = \sum_{r=0}^{\zeta/2} \beta_{2r} \sum_{u=0}^{r} (2r-2u+1)\varphi(r,u)(\lambda\tilde{\sigma}_\epsilon)^{2r-2u+2} \cdot \left[D_{-(2r-2u+2)}\left(-\frac{\lambda}{\tilde{\sigma}_\epsilon}\boldsymbol{g}\right) - D_{-(2r-2u+2)}\left(\frac{\lambda}{\tilde{\sigma}_\epsilon}\boldsymbol{g}\right)\right] \quad \text{(B-1-8)}$$

$$\Delta_{\text{den}} = \sum_{r=0}^{\zeta/2} \beta_{2r} \sum_{u=0}^{r} \varphi(r,u)(\lambda\tilde{\sigma}_\epsilon)^{2r-2u+1} \cdot \left[D_{-(2r-2u+1)}\left(-\frac{\lambda}{\tilde{\sigma}_\epsilon}\boldsymbol{g}\right) + D_{-(2r-2u+1)}\left(\frac{\lambda}{\tilde{\sigma}_\epsilon}\boldsymbol{g}\right)\right] \quad \text{(B-1-9)}$$

where $\varphi(r,u) = \frac{(-2)^u(2r)!}{u!\sigma_f^{2r-2u}}$.

Appendix C

Joint PDF of Inter-Frame DWT Coefficients of Video Using Bivariate Gauss-Hermite Expansion

Let the random variables $\bar{f}_c$ and $\bar{f}_p$ be the normalized versions of f_c and f_p with the relation $\bar{f}_c = f_c/\sigma_c$ and $\bar{f}_p = f_p/\sigma_p$, where σ_c and σ_p are the standard deviations of the data samples of the *current frame* and the *previous frame*, respectively. The marginal PDFs of these normalized random variables, viz., $p_{\bar{f}_c}(\bar{f}_c)$ and $p_{\bar{f}_p}(\bar{f}_p)$, may be expressed in terms of univariate form of the Gauss-Hermite expansion as

$$p_{\bar{f}_c}(\bar{f}_c) = \mathcal{G}_c(\bar{f}_c)\sum_{r=0}^{\infty}\beta_{rc}H_{rc}(\bar{f}_c) \qquad \mathcal{G}_c(\bar{f}_c) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\bar{f}_c^2}{2}} \tag{C-1-1}$$

$$p_{\bar{f}_p}(\bar{f}_p) = \mathcal{G}_p(\bar{f}_p)\sum_{r=0}^{\infty}\beta_{rp}H_{rp}(\bar{f}_p) \qquad \mathcal{G}_p(\bar{f}_p) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\bar{f}_p^2}{2}} \tag{C-1-2}$$

where β_{rc} and β_{rp} are the series coefficients given by

$$\beta_{rc} = \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor}\frac{(-1)^u}{2^u(r-2u)!u!}\frac{M_{(r-2u)c}}{\sigma_c^{r-2u}} \qquad \beta_{rp} = \sum_{u=0}^{\lfloor \frac{r}{2} \rfloor}\frac{(-1)^u}{2^u(r-2u)!u!}\frac{M_{(r-2u)p}}{\sigma_p^{r-2u}} \tag{C-1-3}$$

M_{nc} and M_{np} being the n-th order moments of the data samples of the *current frame* and the *previous frame*, respectively. Let the bivariate Gaussian PDF with unit variances and having a correlation coefficient ρ be

$$\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\bar{f}_c^2 + \bar{f}_p^2 - 2\rho\bar{f}_c\bar{f}_p\right\}\right] \tag{C-1-4}$$

We now show that the joint PDF the random variables $\bar{f}_c$ and $\bar{f}_p$, denoted as

$p_{\bar{f}_c\bar{f}_p}(\bar{f}_c, \bar{f}_p)$, can be expressed as

$$\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) + \mathcal{G}_c(\bar{f}_c)\mathcal{G}_p(\bar{f}_p)\sum_{r=1}^{\infty}\left[\beta_{rc}H_{rc}(\bar{f}_c) + \beta_{rp}H_{rp}(\bar{f}_p)\right] \tag{C-1-5}$$

Using the fact that each of the two marginal PDFs, viz., $p_{\bar{f}_c}(\bar{f}_c)$ and $p_{\bar{f}_c}(\bar{f}_c)$, can be obtained by integrating the joint PDF with an appropriate random variable, we have

$$\int_{-\infty}^{\infty} p_{\bar{f}_c\bar{f}_p}(\bar{f}_c, \bar{f}_p)\mathrm{d}\bar{f}_p = p_{\bar{f}_c}(\bar{f}_c) \tag{C-1-6}$$

$$\int_{-\infty}^{\infty} p_{\bar{f}_c\bar{f}_p}(\bar{f}_c, \bar{f}_p)\mathrm{d}\bar{f}_c = p_{\bar{f}_p}(\bar{f}_p) \tag{C-1-7}$$

Now integrating (C-1-5) with respect to the variable $\bar{f}_p$, we get

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left[\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) + \mathcal{G}_c(\bar{f}_c)\mathcal{G}_p(\bar{f}_p)\sum_{r=1}^{\infty}\left[\beta_{rc}H_{rc}(\bar{f}_c) + \beta_{rp}H_{rp}(\bar{f}_p)\right]\right]\mathrm{d}\bar{f}_p \\
&= \int_{-\infty}^{\infty}\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p)\mathrm{d}\bar{f}_p + \mathcal{G}_c(\bar{f}_c)\sum_{r=1}^{\infty}\beta_{rc}H_{rc}(\bar{f}_c)\int_{-\infty}^{\infty}\mathcal{G}_p(\bar{f}_p)\mathrm{d}\bar{f}_p \\
&\qquad + \mathcal{G}_c(\bar{f}_c)\int_{-\infty}^{\infty}\mathcal{G}_p(\bar{f}_p)\sum_{r=1}^{\infty}\beta_{rp}H_{rp}(\bar{f}_p)\mathrm{d}\bar{f}_p \\
&= \mathcal{G}_c(\bar{f}_c)\int_{-\infty}^{\infty}\mathcal{G}_p(\bar{f}_p)\left[1 + \sum_{r=1}^{\infty}\beta_{rp}H_{rp}(\bar{f}_p)\right]\mathrm{d}\bar{f}_p + \mathcal{G}_c(\bar{f}_c)\sum_{r=1}^{\infty}\beta_{rc}H_{rc}(\bar{f}_c) \\
&= \mathcal{G}_c(\bar{f}_c)\left[1 + \sum_{r=1}^{\infty}\beta_{rc}H_{rc}(\bar{f}_c)\right] = p_{\bar{f}_c}(\bar{f}_c)
\end{aligned} \tag{C-1-8}$$

In a similar way, integrating (C-1-5) with respect to the variable $\bar{f}_c$, we get

$$\int_{-\infty}^{\infty}\left[\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) + \mathcal{G}_c(\bar{f}_c)\mathcal{G}_p(\bar{f}_p)\sum_{r=1}^{\infty}\left[\beta_{rc}H_{rc}(\bar{f}_c) + \beta_{rp}H_{rp}(\bar{f}_p)\right]\right]\mathrm{d}\bar{f}_c = p_{\bar{f}_p}(\bar{f}_p) \tag{C-1-9}$$

From (C-1-6) to (C-1-9), we have the result that the joint PDF can be written as

$$p_{\bar{f}_c\bar{f}_p}(\bar{f}_c, \bar{f}_p) = \mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) + \mathcal{G}_c(\bar{f}_c)\mathcal{G}_p(\bar{f}_p)\sum_{r=1}^{\infty}\left[\beta_{rc}H_{rc}(\bar{f}_c) + \beta_{rp}H_{rp}(\bar{f}_p)\right] \tag{C-1-10}$$

Now, the bivariate Gaussian PDF with unit variances and having a correlation coefficient ρ can be expressed in terms of the orthogonal Hermite polynomials as [80]

$$\mathcal{G}_{cp}(\bar{f}_c, \bar{f}_p) = \mathcal{G}_c(\bar{f}_c)\mathcal{G}_p(\bar{f}_p)\left[1 + \sum_{r=1}^{\infty}\frac{\rho^r}{r!}H_{rc}(\bar{f}_c)H_{rp}(\bar{f}_p)\right] \tag{C-1-11}$$

Hence, the joint PDF of the normalized random variables given in (C-1-10) may be written as

$$p_{\bar{\boldsymbol{f}}_c\bar{\boldsymbol{f}}_p}(\bar{\boldsymbol{f}}_c,\bar{\boldsymbol{f}}_p)=\mathcal{G}_c(\bar{\boldsymbol{f}}_c)\mathcal{G}_p(\bar{\boldsymbol{f}}_p)\Bigg[1+\sum_{r=1}^{\infty}\Big\{\beta_{rp}H_{rp}(\bar{\boldsymbol{f}}_p)+\beta_{rp}H_{rp}(\bar{\boldsymbol{f}}_p)+\frac{\rho^r}{r!}H_{rc}(\bar{\boldsymbol{f}}_c)H_{rp}(\bar{\boldsymbol{f}}_p)\Big\}\Bigg] \qquad \text{(C-1-12)}$$

The joint PDF of the random variables $\boldsymbol{f}_c$ and $\boldsymbol{f}_p$ can be expressed in terms of $p_{\bar{\boldsymbol{f}}_c\bar{\boldsymbol{f}}_p}$ as

$$p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)=p_{\bar{\boldsymbol{f}}_c\bar{\boldsymbol{f}}_p}\Big(\bar{\boldsymbol{f}}_c=\frac{\boldsymbol{f}_c}{\sigma_c},\bar{\boldsymbol{f}}_p=\frac{\boldsymbol{f}_p}{\sigma_p}\Big)\Big|\mathcal{J}(\boldsymbol{f}_c,\boldsymbol{f}_p)\Big| \qquad \text{(C-1-13)}$$

where $\mathcal{J}(\boldsymbol{f}_c,\boldsymbol{f}_p)=\frac{1}{\sigma_c\sigma_p}$ is the Jacobian of $\{\boldsymbol{f}_c,\boldsymbol{f}_p\}$.

Finally, from (C-1-12) and (C-1-13), the joint PDF of the random variables $\boldsymbol{f}_c$ and $\boldsymbol{f}_p$ can be written in the form

$$p_{\boldsymbol{f}_c\boldsymbol{f}_p}(\boldsymbol{f}_c,\boldsymbol{f}_p)=\frac{1}{2\pi\sigma_c\sigma_p}e^{-\frac{1}{2}\left(\frac{f_c^2}{\sigma_c^2}+\frac{f_p^2}{\sigma_p^2}\right)}\Bigg[1+\sum_{r=1}^{\infty}\Big\{\beta_{rc}H_r\Big(\frac{\boldsymbol{f}_c}{\sigma_c}\Big)+\beta_{rp}H_r\Big(\frac{\boldsymbol{f}_p}{\sigma_p}\Big)+\frac{\rho^r}{r!}H_r\Big(\frac{\boldsymbol{f}_c}{\sigma_c}\Big)H_r\Big(\frac{\boldsymbol{f}_p}{\sigma_p}\Big)\Big\}\Bigg] \qquad \text{(C-1-14)}$$