THE COMPLEX GAUSSIAN SCALE MIXTURES

OF COMPLEX WAVELET COEFFICIENTS

AND APPLICATIONS

by

YOTHIN RAKVONGTHAI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2011

ACKNOWLEDGEMENTS

and also for making my life here more pleasurable.

ABSTRACT

THE COMPLEX GAUSSIAN SCALE MIXTURES

OF COMPLEX WAVELET COEFFICIENTS

AND APPLICATIONS

YOTHIN RAKVONGTHAI, Ph.D.

The University of Texas at Arlington, 2011

Supervising Professor: Soontorn Oraintara

In this thesis, the complex Gaussian scale mixture (CGSM), which is an extension of the Gaussian scale mixture (GSM) for real-valued random variables to the complex case, is presented to model the complex wavelet coefficients. Along with some related propositions and miscellaneous results, the probability density functions (pdf) of the magnitude and phase of the complex random variable are presented. Specifically, the closed forms of the magnitude pdf for the case of complex generalized Gaussian distribution (CGGD) and the phase pdf for the general case are presented. Subsequently, the pdf of the *relative phase* is derived.

Moreover, parameter estimation methods in the presence of noise for several magnitude pdf's which are special cases of the magnitude pdf related to the CGSM, and for the relative phase pdf (RP pdf) are proposed. In addition, the parameter estimation of the RP pdf is investigated, and a non-iterative estimator for the relative phase pdf's parameters is also proposed.

To show the usefulness of the proposed model, the CGSM is then applied to image denoising using Bayes least square estimator. The experimental results show that using the CGSM of complex wavelet coefficients visually improves the quality of denoised images from the real case. Moreover, the derived magnitude pdf of the CGGD is then utilized in texture image retrieval that uses complex coefficient magnitude to improve the accuracy rate from using the real or imaginary parts. Finally, the problem of noisy texture retrieval, where the query image is contaminated by noise, is studied. This texture retrieval scheme is based on the proposed parameter estimation methods in the presence of noise. The retrieval results show that using both magnitude and phase information of complex coefficients improves the accuracy rate from solely using the magnitude or phase information, and also from using the real or the imaginary parts. These simulation results are also consistent in several complex transform domains.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Motivation

It is well known that statistical modeling in the wavelet domain is favorable for many image processing applications, such as denoising and compression, because of wavelet's capability of analyzing and representing images [62]. This ability can be further improved by using complex-valued wavelets rather than real-valued wavelets. Fig. 1.1 shows an example of complex-valued decomposition of the Lena image into subbands of complex coefficients. This enhancement is mainly due to the fact that the complex wavelets are based on complex-valued sinusoids constituting an analytic signal [58]. In addition, the advantages of complex wavelets over real wavelets are directly related to the complex magnitude and phase. For example, the magnitude of a complex coefficient possesses the shift-invariance property while a real wavelet coefficient is shift-varying. Furthermore, it is well known that the magnitude of a complex wavelet coefficient better represents a singularity than either the real/imaginary part of the complex coefficient or the value of a real wavelet coefficient. Besides magnitude information, phase information plays a key role in image processing as well. A famous example that shows the importance of phase is in [44], where the Fourier phase is shown to contain more information about image features than the magnitude. More-over, there is also a connection between the phase of a complex wavelet coefficient and image features, such as edges. Specifically, the phase of a complex coefficient in each scale near a singularity varies linearly with its distance to the singularity [4]. In addition to the intrascale phase relationship, the coefficient phases across scales at

1

(a)

(b)

(c)
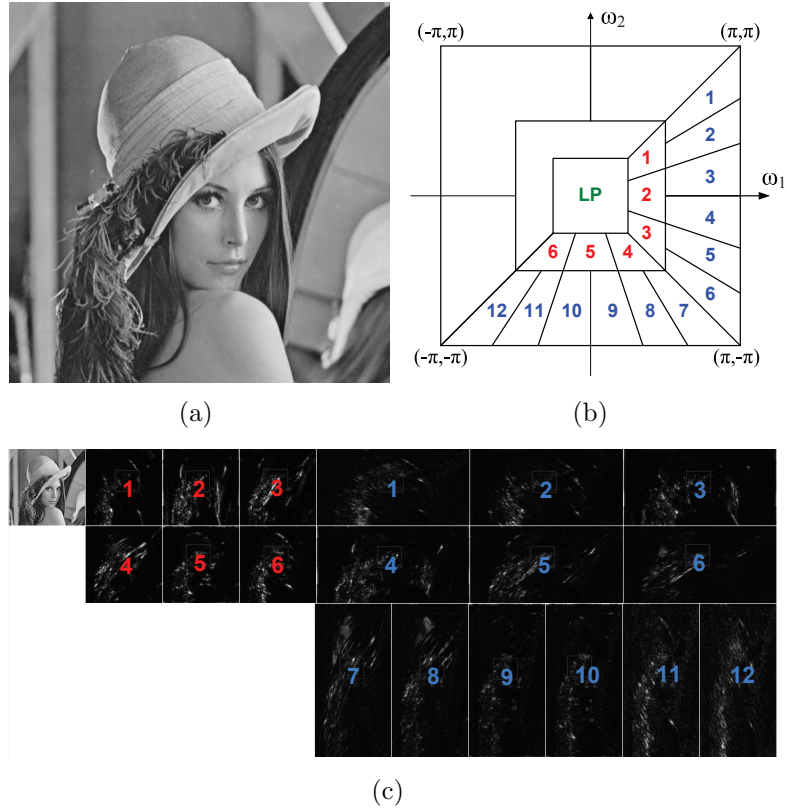
Figure 1.1. Example of complex multiscale decomposition: 2 scales (levels) with 6 and 12 subbands: (a) input image, (b) frequency partition, and (c) subbands of coefficient magnitude.

a singularity are aligned [29, 74]. These intrascale and interscale relationships have been used in some image processing applications, such as feature angle determination [4], feature detection [29], and texture retrieval [69, 70]. All of these point out the significance of the magnitude and phase information of complex coefficients, which leads to a need for statistical modeling in the complex wavelet domain.

1.2   Problem Statement

In the probabilistic framework, an appropriate statistical model for complex random variables is required to fully utilize complex wavelet coefficients as well as their magnitude and phase information. This thesis investigates on:

1. a statistical model that can accurately capture the statistics of complex wavelet coefficients extracted from natural images,

2. the use of the model in image processing in the complex wavelet domain, and

3. the related magnitude and phase pdf's which are stemmed from the model and their applications.

1.3 Statistical Image Modeling in the Wavelet Domain

This section provides a literature review on statistical modeling in the wavelet domain.

The wavelet transform has been shown to be efficient for image modeling [35]. In the wavelet domain, the marginal distribution of wavelet coefficients representing an image follows a peaky and heavy-tailed non-Gaussian distribution. These characteristics can be well modeled by the generalized Gaussian distribution (GGD) [35]. Nevertheless, the GGD model only captures the marginal statistics but does not account for their joint statistics which also exist in the wavelet domain.

To model both marginal and joint statistics in the wavelet domain, the Gaussian scale mixture (GSM) model has been developed [47, 71, 72]. The GSM [5] model characterizes the set of real-valued random vectors that can be expressed as the product of a zero-mean Gaussian random vector and an independent positive random variable, i.e., a GSM random vector is a mixture of a possibly infinite number of zero-mean Gaussian random vectors.

Specifically, a random vector $\mathbf{X}$ with dimension N can be characterized by a Gaussian scale mixture (GSM) if it can be expressed as

$$\mathbf{X} = \sqrt{S}\widetilde{\mathbf{X}} \tag{1.1}$$

where $\widetilde{\mathbf{X}}$ is a zero-mean Gaussian random vector with covariance matrix $\mathbf{C}_{\widetilde{\mathbf{X}}}$, and S is an independent positive scalar random variable called the *multiplier* or *hidden multiplier*. For $\mathbf{X}$ to have the unique GSM representation as in (1.1), the mean of S must be specified. Without loss of generality, assume that S has unit mean [47], i.e., $E[S] = 1$. This implies that the covariance matrix $\mathbf{C}_{\mathbf{X}} = E[S]\mathbf{C}_{\widetilde{\mathbf{X}}} = \mathbf{C}_{\widetilde{\mathbf{X}}}$. Accordingly, the pdf of $\mathbf{X}$ is given by [47]

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \int_0^\infty f_{\mathbf{X}|S}(\mathbf{x}|s) f_S(s)\, ds \\
&= \int_0^\infty \frac{\exp\left(-\frac{\mathbf{x}^T (s\mathbf{C}_{\widetilde{\mathbf{X}}})^{-1}\mathbf{x}}{2}\right)}{(2\pi)^{\frac{N}{2}}\sqrt{|s\mathbf{C}_{\widetilde{\mathbf{X}}}|}} f_S(s)\, ds,
\end{aligned}
\tag{1.2}
$$

for $\mathbf{x} \in \mathbb{R}^N$, where $\mathbf{x}^T$ is the transpose of $\mathbf{x}$, and $|\mathbf{C}_{\widetilde{\mathbf{X}}}|$ is the determinant of $\mathbf{C}_{\widetilde{\mathbf{X}}}$. Note that $f_{\mathbf{X}}(\mathbf{x})$ is a Gaussian pdf when $f_S(s)$ is an impulse function. See [47] also for the choices of the pdf of the multiplier.

The GSM distribution encompasses many known distributions as special cases, such as the Student's t-distribution, the logistic distribution [5], the α-stable distribution, the GGD, and the symmetrized Gamma distribution [72]. The GSM model can accurately capture both marginal and joint statistics of wavelet coefficients which have peaky and heavy-tailed characteristics. Besides the GSM, other wavelet-based image models are also used to capture the characteristics including the hidden Markov tree (HMT) model [13], the bivariate model [60], the Bessel K form density [18], the multivariate Laplacian distribution [57], the Gauss-Hermite expansion [48], multivariate prior models [63], as examples. In addition, the GSM modeling in the wavelet domain has been used for image denoising in [47], which provides a high-quality image denoising algorithm.

Since the emersion of complex wavelets, there are a number of research studies on the statistical modeling based on complex wavelets. For instance, in [10], the

complex hidden Markov tree (CHMT) model is proposed for complex coefficients obtained from the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT) [27]. The bivariate model for complex coefficients and image denoising with the bivariate shrinkage are studied in [60]. In [3], the bivariate α-stable distribution is used for image denoising in the complex wavelet domain. In [49], the Cartesian and polar forms of marginal densities of DT-$\mathbb{C}$WT coefficients due to Gaussian signals are studied. Thanks to the usefulness of the GSM model and the complex wavelets as well as their magnitude and phase information, the modification of the GSM for complex coefficients should be beneficial.

Indeed, the probability density function (pdf) of a complex random vector (as a general form of a complex random variable) is the joint pdf of two real-valued random vectors representing the real and the imaginary parts. In order to express the pdf of a complex random vector as an analytic function of the complex vector itself, we need an additional assumption. For a class of distributions whose pdf's depend only on the covariance matrix such as the Gaussian distribution, the assumption is that the covariance matrices of the real and imaginary parts are equal, and that the sum of the two cross-covariance matrices is zero. With such an assumption, the pdf of a complex random vector whose real and imaginary parts are Gaussian can be expressed as a function of the complex covariance matrix, and has been studied in [19, 38, 75] as the complex Gaussian pdf.

Recently, there are also wavelet-based statistical models extended from the GSM, e.g., the space variant GSM [22], the oriented-adaptive GSM [24], and the GSM for derotated complex coefficients [39]. Nonetheless, those kinds of distributions are outside the scope of this work and are not discussed herein.

1.4   Thesis Outline

Chapter 2 introduces the complex Gaussian scale mixtures (CGSM) for the complex wavelet coefficients. Chapter 3 discusses the magnitude pdf's and the relative phase pdf (RP pdf), which are derived from the CGSM, and their related parameter estimation both in the absence and in the presence of noise. To show the use of the proposed CGSM in practice, Chapter 4 discusses several applications based on complex wavelets of the CGSM and its related magnitude and relative phase pdf's in image denoising, texture retrieval and noisy texture retrieval. Finally, concluding remarks and future directions are presented in Chapter 5.

CHAPTER 2

COMPLEX GAUSSIAN SCALE MIXTURES OF COMPLEX WAVELET
COEFFICIENTS

2.1   Introduction

In this chapter, the complex Gaussian scale mixture (CGSM) is introduced as
an extension of the GSM, which is for real-valued random vectors, to the complex
case and is then used to model complex wavelet coefficients. This chapter begins
with a discussion of modeling the real and imaginary parts of complex coefficients by
the GSM. It is then shown that the corresponding joint covariance matrix satisfies
the circular condition. Assuming that the real part is a GSM, it is proved that the
imaginary part is also a GSM with the same multiplier. Accordingly, the CGSM
of complex wavelet coefficients is proposed. In addition, some related propositions
of the CGSM are presented. Because of the importance of magnitude and phase
information of complex wavelets, examples of using the CGSM to derive two pdf's
related to the magnitude and phase are also discussed. To show the effectiveness of
the CGSM, image denoising is performed using Bayes least squares (BLS) estimator
with the proposed model. The results show that using the CGSM of complex wavelet
coefficients visually improves the quality of denoised images from using the GSM of
real wavelet coefficients.

It is worth noting that the framework herein is not only limited to the DT-ℂWT
[27], which is probably the most widely used complex wavelet transform in image
processing, but also includes other complex-valued multiresolution transforms, such
as the fast discrete curvelet transform (FDCT) [7], the pyramidal dual-tree directional

filter bank (PDTDFB) [43], and the uniform discrete curvelet transform (UDCT) [41, 54]. The derived pdf can be applied and easily extended to other complex transforms because they all behave like complex analytic bandpass filters and thus yield complex subband coefficients whose real and imaginary parts corresponds to wavelets that form a Hilbert transform pair. Therefore, the complex-valued multiresolution transforms are referred to as complex wavelets while the corresponding coefficients are referred to as complex wavelet coefficients or complex coefficients henceforth.

2.2 Statistics of Complex Coefficients

The goal of this section is to study the statistics of the real and imaginary parts of complex coefficients as well as their statistical relationship related to a Hilbert transform pair. In particular, it is shown that the GSM can be used to model the real/imaginary part, and that both parts are related in terms of covariance matrices in Subsections 2.2.1 and 2.2.2, respectively.

2.2.1 Gaussian Scale Mixtures

To begin with, this section shows that the real and imaginary parts are consistent with the GSM model.

To validate the GSM modeling of the real and imaginary parts of complex coefficients, the marginal statistics of complex coefficients produced using the UDCT of the Lena image is studied. The log of normalized marginal histogram of the real/imaginary part of complex coefficients along with its best fitted Student's t-distribution, which is a class of GSM whose hidden multiplier has the inverse Gamma pdf [31], is shown in Fig.'s 2.1(a) and 2.1(b) for the real and imaginary parts, respectively. Fitting is done by minimizing the relative entropy ΔH (the Kullback-Leibler divergence (KLD)) between the histogram and the pdf. It can be seen that the his-
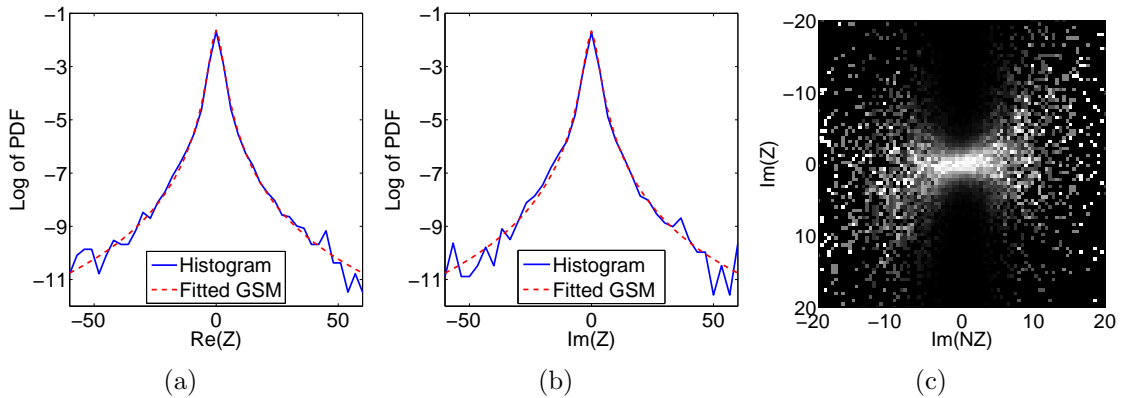
Figure 2.1. Marginal and joint statistics of the real/imaginary part of complex coefficients in one finest subband of the Lena image: (a) Log of normalized marginal histogram and the fitted GSM for the real part; (b) for the imaginary part; and (c) the conditional histogram of the imaginary part of two spatially adjacent coefficients, Z and its adjacent neighbor NZ. Each column has been independently rescaled to fulfill the full range of intensities.

togram is highly non-Gaussian with high peak at zero, and that the fitting result for each histogram is quite good with the ratio between the KLD and the histogram entropy $\Delta H/H = 0.00077$ and 0.00107 for the real and imaginary parts, respectively. Note also that the marginal distributions of both parts are close to each other. In addition to the marginal statistics, the conditional histogram of two spatially adjacent coefficients is shown in Fig. 2.1(c). Only the imaginary part is considered because the result of the real part is also similar. The shape of this conditional histogram is so-called *bow-tie* shape, which implies that two random variables are roughly uncorrelated but highly dependent. These statistical characteristics of the imaginary part of complex coefficients can be modeled by a GSM [47, 71, 72].

2.2.2   Hilbert Transform Pairs and Correlation Relationships

Since the real and imaginary parts of complex coefficients are obtained from sampling two processes which form a Hilbert transform pair, their relationship is crucial to the development of the CGSM, whose pdf is a function of complex variables.

This subsection derives the relationship of a 2-D Hilbert transform pair in terms of autocorrelation and cross-correlation, and relates it to the relationship of the real and imaginary parts of complex coefficients. The details are as follows.

Let $X_1(\mathbf{t})$ be a real 2-D random process, where $\mathbf{t} = [t_1, t_2]^T$. The Hilbert transform of $X_1(\mathbf{t})$, which is denoted by $\mathcal{H}\{X_1(\mathbf{t})\}$, is defined as

$$\mathcal{H}\{X_1(\mathbf{t})\} \triangleq (h * X_1)(\mathbf{t}),$$

where $h(\mathbf{t})$ is defined by its Fourier transform $H(\boldsymbol{\omega})$ as [21]

$$H(\boldsymbol{\omega}) = -j \, \mathrm{sgn}(\boldsymbol{\omega}^T \mathbf{e}) = \begin{cases} -j, \text{if } \boldsymbol{\omega}^T \mathbf{e} > 0; \\ 0, \text{ if } \boldsymbol{\omega}^T \mathbf{e} = 0; \\ j, \text{ if } \boldsymbol{\omega}^T \mathbf{e} < 0, \end{cases} \tag{2.1}$$

where $\boldsymbol{\omega} = [\omega_1, \omega_2]^T$, and $\mathbf{e} = [e_1, e_2]^T$, is a real unit-norm vector. Let $X_2(\mathbf{t})$ be the Hilbert transform of $X_1(\mathbf{t})$, i.e., $X_2(\mathbf{t}) = \mathcal{H}\{X_1(\mathbf{t})\}$. Assume that $X_1(\mathbf{t})$ is wide-sense stationary (WSS). Since $\mathcal{H}$ is linear time-invariant (LTI), $X_1(\mathbf{t})$ and $X_2(\mathbf{t})$ are jointly WSS [45], i.e.,

$$E[X_i(\mathbf{t} + \boldsymbol{\tau})X_j(\mathbf{t})] = R_{X_i X_j}(\boldsymbol{\tau}),$$

where $\boldsymbol{\tau} = [\tau_1, \tau_2]^T$, for $1 \leq i, j \leq 2$. Let $R_{X_i}(\boldsymbol{\tau}) \triangleq R_{X_i X_i}(\boldsymbol{\tau})$. The cross power spectrum of two processes $X_i(\mathbf{t})$ and $X_j(\mathbf{t})$ is defined as the Fourier transform $S_{X_i X_j}(\boldsymbol{\omega})$ of $R_{X_i X_j}(\boldsymbol{\tau})$

$$S_{X_i X_j}(\boldsymbol{\omega}) = \int_{\mathbb{R}^2} R_{X_i X_j}(\boldsymbol{\tau})e^{-j\boldsymbol{\omega}^T\boldsymbol{\tau}}d\boldsymbol{\tau}.$$

When $i = j$, the cross power spectrum becomes the power spectrum (or spectral density) of $X_i(\mathbf{t})$, denoted as $S_{X_i}(\boldsymbol{\omega})$. Note that $S_{X_1 X_2}(\boldsymbol{\omega}) = S_{X_2 X_1}^*(\boldsymbol{\omega})$ since $X_i(\mathbf{t})$'s are real. The relationships of the cross power spectrums are then given by

$$S_{X_2}(\boldsymbol{\omega}) = S_{X_1}(\boldsymbol{\omega})|H(\boldsymbol{\omega})|^2,$$

$$S_{X_1 X_2}(\boldsymbol{\omega}) = S_{X_1}(\boldsymbol{\omega})H^*(\boldsymbol{\omega}), \text{ and}$$

$$S_{X_2}(\boldsymbol{\omega}) = S_{X_1 X_2}(\boldsymbol{\omega})H(\boldsymbol{\omega}). \tag{2.2}$$

From (2.1) and (2.2), it follows that

$$S_{X_1}(\boldsymbol{\omega}) = S_{X_2}(\boldsymbol{\omega}), \text{ and } S_{X_1 X_2}(\boldsymbol{\omega}) = -S_{X_2 X_1}(\boldsymbol{\omega}), \text{ and thus}$$

$$R_{X_1}(\boldsymbol{\tau}) = R_{X_2}(\boldsymbol{\tau}), \text{ and } R_{X_1 X_2}(\boldsymbol{\tau}) = -R_{X_2 X_1}(\boldsymbol{\tau}). \tag{2.3}$$

Consider an N-size neighborhood of each of the real and imaginary parts of complex coefficients. The corresponding neighborhood vectors are given by

$$\mathbf{X} = [X_1, \ X_2, \ldots, \ X_N]^T, \text{ and } \mathbf{Y} = [Y_1, \ Y_2, \ldots, \ Y_N]^T,$$

for the real and imaginary parts, respectively. Since the discrete complex wavelet transform can be considered as a bank of bandpass filters, the two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are zero-mean. As $\mathbf{X}$ and $\mathbf{Y}$ are obtained from sampling two processes which form a Hilbert transform pair, it follows from (2.3) that

$$\mathbf{C_X} = \mathbf{C_Y} \text{ and } \mathbf{C_{XY}} = -\mathbf{C_{YX}}, \tag{2.4}$$

where $\mathbf{C_X} \triangleq E[\mathbf{X}\mathbf{X}^T]$, and $\mathbf{C_{XY}} \triangleq E[\mathbf{X}\mathbf{Y}^T]$.

Note that this condition on the covariance matrices (2.4) is the so-called *circular* [46] or *proper* [40] condition, and a complex random vector that satisfies these conditions on covariance matrices is called *circular* [46] or *proper* [40]. The circular condition allows us to express the jointly Gaussian pdf of the real part $\mathbf{X}$ and the imaginary part $\mathbf{Y}$ as the complex Gaussian pdf of the associated complex number $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$. This will be discussed briefly in Subsection 2.3.2.

To verify that the complex wavelet coefficients satisfy the circular condition, an image is decomposed into three scales using the UDCT, DT-ℂWT, PDTDFB, and FDCT. Then, in each subband, the 9×1 complex random vector is formed from the 3×3-block neighborhood. To quantitatively measure the circularity of the complex random vector, the degree of impropriety [55] which measures how much a complex

Table 2.1. Degrees of impropriety for various complex transforms

Image	Scale	UDCT	DT-$\mathbb{C}$WT	PDTDFB	FDCT
Lena	1	0.0741	0.9375	0.1406	0.1892
	2	0.0435	0.0855	0.1482	0.0795
	3	0.0924	0.2340	0.1495	0.0639
Barbara	1	0.0822	0.8882	0.0905	0.1444
	2	0.0372	0.0854	0.0972	0.1112
	3	0.0614	0.1683	0.0712	0.0814

random vector is non-circular is computed. For a zero-mean complex random vector $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ with complex covariance matrix $\mathbf{C_Z} \triangleq E[\mathbf{Z}\mathbf{Z}^H]$, where $\mathbf{Z}^H = (\mathbf{Z}^*)^T$ is the conjugate transpose of $\mathbf{Z}$, and pseudo-covariance matrix $\mathbf{P_Z} \triangleq E[\mathbf{Z}\mathbf{Z}^T]$, the *degree of impropriety* is defined by [55]

$$d = 1 - \frac{|\overline{\mathbf{R}}|}{|\mathbf{C_Z}|^2}, \tag{2.5}$$

where $\overline{\mathbf{R}} = \begin{bmatrix} \mathbf{C_Z} & \mathbf{P_Z} \\ \mathbf{P_Z^*} & \mathbf{C_Z^*} \end{bmatrix}$. The value of the degree of impropriety, d, is between zero and one, where $d = 0$ when the complex random vector is circular and least circular when $d = 1$.

Table 2.1 shows the degrees of impropriety for the Lena, and Barbara images. For each image, the degree of impropriety of each subband is computed using (2.5). Each entry is the average value of the degrees of impropriety of all subbands in one scale. From Table 2.1, for the UDCT, PDTDFB, and FDCT, it can be seen that the average degrees of impropriety in all three scales of each transform are small, i.e., the UDCT, PDTDFB, and FDCT coefficients can be assumed to satisfy the circular condition. This is also true for the DT-$\mathbb{C}$WT in scales 2 and 3, whereas in the first scale, the degrees of impropriety are quite large ($d > 0.88$ for both images), which implies that the DT-$\mathbb{C}$WT coefficients in the finest scale do not satisfy the circular condition. This is because, in the implementation of the DT-$\mathbb{C}$WT, the analyticity

of the scale 1 filters is sacrificed to achieve the approximate analyticity at other scales [58].

2.3 Complex Gaussian Scale Mixtures

To exploit the advantages of complex wavelets and the usefulness of the complex magnitude and phase in the statistical framework, an appropriate model that can handle complex random variables is needed. The aim of this section is to introduce the pdf of a vector of complex coefficients whose real and imaginary parts are GSM's because of the effectiveness of the GSM model in statistical wavelet modeling as mentioned earlier.

2.3.1 Equivalence of the Hidden Multipliers of the Real and Imaginary Parts

Lemma 2.3.1. *Let $X(\mathbf{n})$ and $Y(\mathbf{n})$ be two random variables of the real and imaginary parts of the complex subband coefficient at position $\mathbf{n}$, respectively. If $X(\mathbf{n})$ is a GSM with constant scalar multiplier in the neighborhood of $\mathbf{n}$, then $Y(\mathbf{n})$ can be modeled by a GSM with the same scalar multiplier.*

Proof. Assume that $X(\mathbf{n})$ is a GSM. From (1.1), $X(\mathbf{n})$ can be written as

$$X(\mathbf{n}) = \sqrt{S_1(\mathbf{n})}\widetilde{X}(\mathbf{n}),$$

where $\widetilde{X}(\mathbf{n})$ is zero-mean Gaussian, $S_1(\mathbf{n})$ is unit-mean and independent to $\widetilde{X}(\mathbf{n})$. Since $Y(\mathbf{n})$ is a Hilbert transform pair of $X(\mathbf{n})$, we have

$$Y(\mathbf{n}) = \sum_{\mathbf{m}\in\mathbb{Z}^2} X(\mathbf{n}-\mathbf{m})h(\mathbf{m}),$$

where $h(\mathbf{m})$ is the impulse response of a 2D Hilbert transform. Hence $Y(\mathbf{n})$ can be approximated by

$$Y(\mathbf{n}) = \sum_{\mathbf{m}\in\mathbb{Z}^2} X(\mathbf{n}-\mathbf{m})h(\mathbf{m}) \approx \sum_{\mathbf{m}\in\mathcal{A}} X(\mathbf{n}-\mathbf{m})h(\mathbf{m}),$$

where $\mathcal{A}$ is the set of points in the neighborhood of $\mathbf{n}$. Since $\mathbf{m} \in \mathcal{A}$, $S_1(\mathbf{n} - \mathbf{m}) = S_1(\mathbf{n})$. Hence,

$$Y(\mathbf{n}) \approx \sum_{\mathbf{m} \in \mathcal{A}} \sqrt{S_1(\mathbf{n} - \mathbf{m})} \widetilde{X}(\mathbf{n} - \mathbf{m}) h(\mathbf{m})$$
$$= \sqrt{S_1(\mathbf{n})} \sum_{\mathbf{m} \in \mathcal{A}} \widetilde{X}(\mathbf{n} - \mathbf{m}) h(\mathbf{m}) = \sqrt{S_1(\mathbf{n})} \widetilde{Y}(\mathbf{n}),$$

where $\widetilde{Y}(\mathbf{n}) = \sum_{\mathbf{m} \in \mathcal{A}} \widetilde{X}(\mathbf{n} - \mathbf{m}) h(\mathbf{m})$ is zero-mean Gaussian and independent to $S_1(\mathbf{n})$. $\qquad \square$

Proposition 2.3.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two random vectors of neighborhoods of the real and imaginary parts of complex subband coefficients, respectively. If $\mathbf{X}$ is a GSM, then so is $\mathbf{Y}$ and both $\mathbf{X}$ and $\mathbf{Y}$ have the same scalar multiplier.*

Proof. Assume that $\mathbf{X}$ is a GSM. From (1.1), $\mathbf{X}$ can be written as $\mathbf{X} = \sqrt{S_1} \widetilde{\mathbf{X}}$, where $\widetilde{\mathbf{X}}$ is zero-mean Gaussian, S_1 are unit-mean and independent to $\widetilde{\mathbf{X}}$. It follows directly from Lemma 2.3.1 that

$$\mathbf{Y} = \sqrt{S_1} \widetilde{\mathbf{Y}},$$

where $\widetilde{\mathbf{Y}}$ is zero-mean Gaussian independent to S_1, i.e., $\mathbf{Y}$ is also a GSM with the same hidden multiplier. $\qquad \square$

This result is demonstrated in Fig. 2.2, which considers the conditional histogram of S_2 given S_1, $f_{S_2|S_1}(s_2|s_1)$, where S_1 and S_2 are defined according to $\mathbf{X} = \sqrt{S_1} \widetilde{\mathbf{X}}$ and $\mathbf{Y} = \sqrt{S_2} \widetilde{\mathbf{Y}}$. Due to their hidden structure, the hidden multipliers S_1 and S_2 cannot be observed. However, each of them can be estimated by the maximum likelihood estimator in each neighborhood [71]:

$$\hat{S}_1(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{C}_{\widetilde{\mathbf{X}}}^{-1} \mathbf{x}}{N} \quad \text{and} \quad \hat{S}_2(\mathbf{y}) = \frac{\mathbf{y}^T \mathbf{C}_{\widetilde{\mathbf{Y}}}^{-1} \mathbf{y}}{N},$$

where $\mathbf{x}$ and $\mathbf{y}$ are the real and imaginary parts of coefficients in that neighborhood of size N. In Fig. 2.2, the conditional histograms corresponding to the real and imaginary parts of complex coefficients from one finest subband of the Lena (Fig. 2.2(a)),
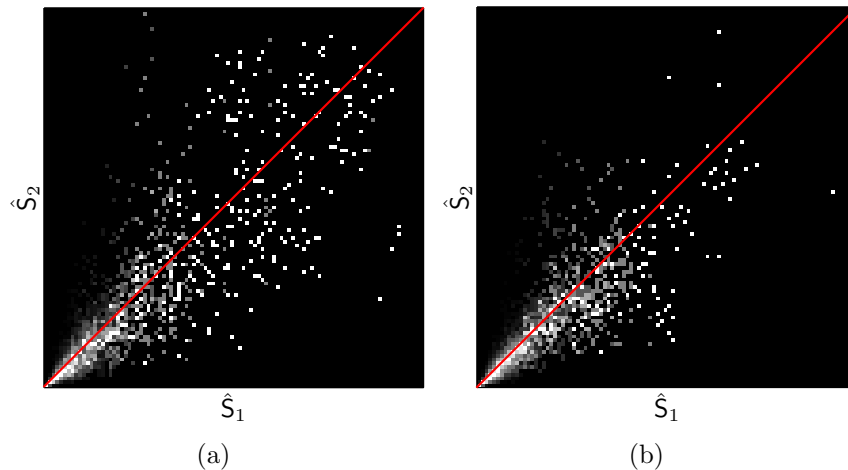
Figure 2.2. Conditional histograms of the estimates $\hat{S}_1$ and $\hat{S}_2$ of the hidden multipliers of the real and imaginary parts of complex coefficients from one finest subband of the (a) Lena, and (b) Barbara images. Each column has been independently rescaled to fulfill the full range of intensities. The diagonal line in each histogram represents $\hat{S}_1 = \hat{S}_2$.

and Barbara (Fig. 2.2(b)) images are shown, where the images are decomposed using the UDCT. It can be seen that the conditional histograms are densely concentrated around the diagonal line $\hat{S}_1 = \hat{S}_2$ especially when $\hat{S}_1$ is small.

2.3.2 Complex Gaussian Distribution

Before the CGSM is introduced, this subsection discusses the complex Gaussian distribution, which is an important class of the CGSM. Particularly, this subsection presents how the pdf of jointly Gaussian random vectors with the circular condition can be written as the pdf of a complex Gaussian vector. For simplicity, the zero-mean case is assumed.

To start with, let $\mathbf{X}$ and $\mathbf{Y}$ be two $N{\times}1$ zero-mean Gaussian random vectors. Assume that they are jointly Gaussian with the circular condition. Let $\mathbf{U} = [\mathbf{X}^T,\ \mathbf{Y}^T]^T$, which is Gaussian with covariance matrix $\mathbf{C_U} = \begin{bmatrix} \mathbf{C_X} & \mathbf{C_{XY}} \\ -\mathbf{C_{XY}} & \mathbf{C_X} \end{bmatrix}$, with the pdf given by

$$f_{\mathbf{U}}(\mathbf{u}) = \frac{1}{(2\pi)^{\frac{2N}{2}}\sqrt{|\mathbf{C_U}|}} \exp\left(-\frac{\mathbf{u}^T \mathbf{C_U}^{-1}\mathbf{u}}{2}\right), \quad \text{for } \mathbf{u} \in \mathbb{R}^{2N}. \tag{2.6}$$

Moreover, its characteristic function has the form

$$\Phi_{\mathbf{U}}(\boldsymbol{v}) \triangleq E[e^{j(\boldsymbol{\alpha}^T\mathbf{X}+\boldsymbol{\beta}^T\mathbf{Y})}] = \exp\left(-\frac{\boldsymbol{v}^T\mathbf{C_U}\boldsymbol{v}}{2}\right), \tag{2.7}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^N$, and $\boldsymbol{v} = [\boldsymbol{\alpha}^T, \ \boldsymbol{\beta}^T]^T$. Let $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ with complex covariance matrix $\mathbf{C_Z} = E[\mathbf{ZZ}^H]$. Therefore, $\mathbf{C_Z} = 2(\mathbf{C_X} - j\mathbf{C_{XY}})$, and satisfies the following conditions: [26, 45]

1) $\mathbf{u}^T \mathbf{C_U}^{-1}\mathbf{u} = 2\mathbf{z}^H\mathbf{C_Z}^{-1}\mathbf{z}$, where $\mathbf{u} = [\mathbf{x}^T, \mathbf{y}^T]^T$ and $\mathbf{z} = \mathbf{x} + j\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$.

2) $\boldsymbol{v}^T\mathbf{C_U}\boldsymbol{v} = \frac{1}{2}\boldsymbol{\omega}^H\mathbf{C_Z}\boldsymbol{\omega}$, where $\boldsymbol{v} = [\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T]^T$ and $\boldsymbol{\omega} = \boldsymbol{\alpha} + j\boldsymbol{\beta}$ for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^N$.

3) $2^{2N}|\mathbf{C_U}| = |\mathbf{C_Z}|^2$.

Using statements 1) and 3), the Gaussian pdf of $\mathbf{U} = [\mathbf{X}^T, \ \mathbf{Y}^T]^T$ in (2.6) can be expressed as the complex Gaussian pdf of $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ having the form

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\exp\left(-\mathbf{z}^H\mathbf{C_Z}^{-1}\mathbf{z}\right)}{\pi^N|\mathbf{C_Z}|}, \quad \text{for } \mathbf{z} \in \mathbb{C}^N. \tag{2.8}$$

Using statement 2), the characteristic function of $\mathbf{U}$ in (2.7) can be written as the characteristic function of $\mathbf{Z}$ which is obtained by

$$\Phi_{\mathbf{Z}}(\boldsymbol{\omega}) \triangleq E[e^{j(\boldsymbol{\alpha}^T\mathbf{X}+\boldsymbol{\beta}^T\mathbf{Y})}] = \exp\left(-\frac{\boldsymbol{\omega}^H\mathbf{C_Z}\boldsymbol{\omega}}{4}\right), \tag{2.9}$$

where $\boldsymbol{\alpha}, \ \boldsymbol{\beta} \in \mathbb{R}^N$, and $\boldsymbol{\omega} = \boldsymbol{\alpha} + j\boldsymbol{\beta} \in \mathbb{C}^N$. We refer to [26, 45] for the proofs and more comprehensive details.

2.3.3   Complex Gaussian Scale Mixtures

Proposition 2.3.1 and the circular condition (2.4) allow us to express the joint pdf of two real-valued random vectors representing the real and imaginary parts as the pdf of the complex-valued random vector. The details are as follows.

To begin with, let $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} = \sqrt{S}\widetilde{\mathbf{Z}}$, where $\widetilde{\mathbf{Z}} = \widetilde{\mathbf{X}} + j\widetilde{\mathbf{Y}}$ is a zero-mean complex Gaussian random vector and the unit-mean S is independent to $\widetilde{\mathbf{Z}}$. Therefore,

the complex covariance is obtained by $\mathbf{C_Z} = E[\mathbf{Z}\mathbf{Z}^H] = E[S]\mathbf{C_{\widetilde{Z}}} = \mathbf{C_{\widetilde{Z}}}$. It follows from the results in Subsection 2.3.2 that $\mathbf{C_{\widetilde{Z}}} = E[\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^H] = 2(\mathbf{C_{\widetilde{X}}} - j\mathbf{C_{\widetilde{X}\widetilde{Y}}})$. From (2.8),

$$f_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}) = \frac{\exp\left(-\widetilde{\mathbf{z}}^H \mathbf{C}_{\widetilde{\mathbf{z}}}^{-1}\widetilde{\mathbf{z}}\right)}{\pi^N |\mathbf{C}_{\widetilde{\mathbf{z}}}|},$$

where N is the neighborhood size. Conditioned on S, $\mathbf{Z}|\{S = s\} = \sqrt{s}\widetilde{\mathbf{Z}}$ is complex Gaussian with complex covariance matrix $\mathbf{C}_{\mathbf{Z}|S} = s\mathbf{C}_{\widetilde{\mathbf{z}}}$. Hence, the pdf of $\mathbf{Z}$ is given by

$$
\begin{aligned}
f_{\mathbf{Z}}(\mathbf{z}) &= \int_0^\infty f_{\mathbf{Z}|S}(\mathbf{z}|s)f_S(s) \ ds \\
&= \int_0^\infty \frac{\exp\left(-\mathbf{z}^H(s\mathbf{C}_{\widetilde{\mathbf{z}}})^{-1}\mathbf{z}\right)}{\pi^N |s\mathbf{C}_{\widetilde{\mathbf{z}}}|} f_S(s) \ ds,
\end{aligned}
\tag{2.10}
$$

for $\mathbf{z} \in \mathbb{C}^N$. Note that $f_{\mathbf{Z}}(\mathbf{z})$ is a complex Gaussian pdf when $f_S(s)$ is an impulse function. Therefore, the complex random vector $\mathbf{Z}$ is called a *complex Gaussian scale mixture* (CGSM) because of its behavior as a complex Gaussian when conditioned on S.

2.4 Some Miscellaneous Results

This section discusses miscellaneous results which stem from the CGSM. Particularly, properties of the CGSM are addressed together with some distributions that are derived using the CGSM.

2.4.1 Related Properties

This subsection presents some properties related to the CGSM proposed in Section 2.3. To start with, the characteristic function of the CGSM is discussed.

Proposition 2.4.1. *If $\mathbf{Z} = \mathbf{X} + j\mathbf{Y}$ is an N-dimensional CGSM, then the characteristic function of $\mathbf{Z}$ is given by*

$$\Phi_{\mathbf{Z}}(\boldsymbol{\omega}) = E[e^{j(\boldsymbol{\alpha}^T\mathbf{X}+\boldsymbol{\beta}^T\mathbf{Y})}] = \int_0^\infty \exp\left(-\frac{\boldsymbol{\omega}^H(s\mathbf{C}_{\mathbf{Z}})\boldsymbol{\omega}}{4}\right) f_S(s) \ ds,$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \in \mathbb{R}^N$, $\boldsymbol{\omega} = \boldsymbol{\alpha} + j\boldsymbol{\beta}$, and $\mathbf{C_Z} = E[\mathbf{Z}\mathbf{Z}^H]$.

Proof. Given $S = s$, $\mathbf{Z}|S$ is complex Gaussian with complex covariance matrix $\mathbf{C_{Z|S}} = s\mathbf{C_{\widetilde{Z}}} = s\mathbf{C_Z}$. The proof follows from the fact that the conditional characteristic function of $\mathbf{Z}$ is given by

$$\Phi_{\mathbf{Z}|S}(\boldsymbol{\omega}) = \exp\left(-\frac{\boldsymbol{\omega}^H(s\mathbf{C_Z})\boldsymbol{\omega}}{4}\right),$$

which follows from (2.9). $\qquad\square$

Proposition 2.4.2. *If $\mathbf{Z}$ is an N-dimensional CGSM, then $\mathbf{Z}^*$ is also an N-dimensional CGSM and $\mathbf{AZ}$ is an M-dimensional CGSM, where $\mathbf{A}$ is an $M \times N$ complex constant matrix.*

Proof. The proof is trivial. $\qquad\square$

Before presenting a property of the CGSM, this subsection discusses an analogous property of the GSM that is also mentioned in [23].

Proposition 2.4.3. *If X_1 and X_2 are independent univariate GSM's, then $X_3 = X_1 + X_2$, $X_4 = X_1X_2$, and $X_5 = \frac{X_1}{X_2}$ are also GSM's.*

Proof. Let $X_i = \sqrt{S_i}\widetilde{X}_i$ for $i = 1, 2$, where the unit-mean positive random variable S_i and the zero-mean Gaussian random variable $\widetilde{X}_i$ are independent. Assume further that X_1 and X_2 are independent.

Given $S_1 = s_1$ and $S_2 = s_2$, $X_3|\{S_1 = s_1, S_2 = s_2\}$ is zero-mean Gaussian with variance $s_1\text{Var}[\widetilde{X}_1] + s_2\text{Var}[\widetilde{X}_2]$. Therefore, X_3 can be written as $X_3 = \sqrt{S_3}\widetilde{X}_3$, where $S_3 = \frac{S_1\text{Var}[\widetilde{X}_1] + S_2\text{Var}[\widetilde{X}_2]}{\text{Var}[\widetilde{X}_1] + \text{Var}[\widetilde{X}_2]}$ is unit-mean and $\widetilde{X}_3 \sim \mathcal{N}(0, \text{Var}[\widetilde{X}_1] + \text{Var}[\widetilde{X}_2])$ independent to S_3.

To show that X_4 is a GSM, consider $X_4 = X_1X_2 = \sqrt{S_1S_2}\widetilde{X}_1\widetilde{X}_2$. Hence, X_4 can be written as $X_4 = \sqrt{S_4}\widetilde{X}_4$, where $S_4 = \frac{S_1S_2}{\text{Var}[\widetilde{X}_2]}\widetilde{X}_2^2$ is unit-mean and $\widetilde{X}_4 = \sqrt{\text{Var}[\widetilde{X}_2]}\text{sgn}(\widetilde{X}_2)\widetilde{X}_1$ is zero-mean Gaussian independent to S_4.

Finally, write $X_5 = \frac{X_1}{X_2} = \frac{\sqrt{S_1}\widetilde{X}_1}{\sqrt{S_2}\widetilde{X}_2}$. Similarly, it can be seen that $X_5 = \sqrt{S_5}\widetilde{X}_5$, where $S_5 = \frac{S_1}{S_2 E[\frac{1}{S_2}]} \frac{\frac{1}{\widetilde{X}_2^2}}{\mathrm{Var}[\frac{1}{\widetilde{X}_2}]}$ is unit-mean, and $\widetilde{X}_5 = \sqrt{\mathrm{Var}[\frac{1}{\widetilde{X}_2}]E[\frac{1}{S_2}]}\mathrm{sgn}(\frac{1}{\widetilde{X}_2})\widetilde{X}_1$ is zero-mean Gaussian independent to S_5. Hence, $X_3, X_4,$ and X_5 are also GSM's. $\qquad\square$

Like the GSM, a similar property for the CGSM is found in the following proposition.

Proposition 2.4.4. *If Z_1 and Z_2 are independent univariate CGSM's, then $Z_3 = Z_1 + Z_2$, $Z_4 = Z_1 Z_2$, and $Z_5 = Z_1/Z_2$ are also CGSM's.*

Proof. Let $Z_i = \sqrt{S_i}\widetilde{Z}_i$ for $i = 1, 2$ where the unit-mean positive random variable S_i and the zero-mean complex Gaussian random variable $\widetilde{Z}_i$ are independent. For $i = 1, 2$, write $Z_i = X_i + jY_i$, where X_i and Y_i are GSM's with $X_i = \sqrt{S_i}\widetilde{X}_i$ and $Y_i = \sqrt{S_i}\widetilde{Y}_i$. Assume further that Z_1 and Z_2 are independent. Therefore, $[X_1, Y_1]$ and $[X_2, Y_2]$ are independent. Write

$$Z_3 = Z_1 + Z_2 = (X_1 + X_2) + j(Y_1 + Y_2) \triangleq X_3 + jY_3,$$

$$Z_4 = Z_1 Z_2 = (X_1 X_2 - Y_1 Y_2) + j(X_1 Y_2 + X_2 Y_1) \triangleq X_4 + jY_4,$$

$$Z_5 = Z_1/Z_2 = \frac{X_1 X_2 + Y_1 Y_2}{X_2^2 + Y_2^2} + j\frac{X_2 Y_1 - X_1 Y_2}{X_2^2 + Y_2^2} \triangleq X_5 + jY_5.$$

From Proposition 2.4.3 and the fact that $[X_1, Y_1]$ and $[X_2, Y_2]$ are independent, it is easy to show that X_i and Y_i are orthogonal GSM's with the same variance and the same hidden multiplier, for $i = 3, 4, 5$. As a result, $Z_3, Z_4,$ and Z_5 are also CGSM's. $\qquad\square$

In the next two subsections, the usefulness of the proposed CGSM for deriving two pdf's related to the magnitude and phase of complex coefficients is exemplified.

2.4.2   Complex Generalized Gaussian Distribution and Its Magnitude PDF

Based on the CGSM model, there is in general no closed form for the magnitude pdf owing to the presence of the hidden multiplier. This leads to difficulties in some applications such as texture retrieval where the closed form for the KLD between two distributions is desirable [50]. Therefore, this subsection focuses on a special case of the CGSM, the complex generalized Gaussian distribution (CGGD), whose real version, the GGD, is widely used to model real-valued wavelet coefficients [?, 34]. In this subsection, the magnitude pdf of a complex wavelet coefficient characterized by the CGGD is derived. The CGGD has been utilized to model complex SAR images in [14].

Let a complex coefficient $Z = X + jY$ be characterized by the CGSM when $N = 1$, i.e., $[X, Y]^T$ is a GSM. If $f_S(s)$ is such that $f_{X,Y}(x, y)$ is the pdf of the bivariate GGD [9, 50], then the distribution of Z is a CGGD with the pdf [14]

$$
\begin{aligned}
f_Z(z) = f_{X,Y}(x, y) &= \int_0^\infty \frac{\exp\left(-\frac{x^2+y^2}{2s\sigma^2}\right)}{2\pi s\sigma^2} f_S(s) \, ds \\
&= \frac{\gamma \Gamma\left(\frac{2}{\gamma}\right)}{2\pi \Gamma^2\left(\frac{1}{\gamma}\right)\sigma^2} e^{-\left(\frac{\Gamma\left(\frac{2}{\gamma}\right)}{2\Gamma\left(\frac{1}{\gamma}\right)} \frac{x^2+y^2}{\sigma^2}\right)^\gamma} \\
&= \frac{\gamma \Gamma\left(\frac{2}{\gamma}\right)}{2\pi \Gamma^2\left(\frac{1}{\gamma}\right)\sigma^2} e^{-\left(\frac{\Gamma\left(\frac{2}{\gamma}\right)}{2\Gamma\left(\frac{1}{\gamma}\right)} \frac{|z|^2}{\sigma^2}\right)^\gamma},
\end{aligned}
$$

where $\gamma > 0$ and $\sigma^2 > 0$ are the parameters, and $\Gamma(z) \triangleq \int_0^\infty e^{-t} t^{z-1} dt$, $z > 0$, is the Gamma function. Note that, since the GGD is also known as the exponential power distribution [28], this distribution of Z can be called as the complex exponential power distribution. Let $X = R\cos\theta$ and $Y = R\sin\theta$, where $R \geq 0$ and $-\pi \leq \theta < \pi$. Then, the pdf of R is given by [50]

$$f_R(r) = \int_{-\pi}^{\pi} f_{R,\Theta}(r, \theta) \, d\theta = \int_{-\pi}^{\pi} r f_Z(r \cos \theta + jr \sin \theta) \, d\theta$$

$$= \frac{\beta}{\alpha^2 \Gamma\left(\frac{2}{\beta}\right)} r e^{-\left(\frac{r}{\alpha}\right)^\beta}, \tag{2.11}$$

where $\alpha = \sigma \sqrt{\frac{2\Gamma(1/\gamma)}{\Gamma(2/\gamma)}}$, and $\beta = 2\gamma$ are the two parameters.

2.4.3 Derivation of the Relative Phase PDF

Although the complex wavelet phase holds crucial information, the coefficient phase in a subband is uniformly distributed, which leads to difficulty in exploiting phase information in the statistical approach. This suggests us to utilize the information of a multivariate phase model, such as a bivariate model, rather than a univariate one.

This subsection uses the CGSM to derive the pdf of the relative phase. The relative phase [69] Φ at a spatial location (k, l) is defined as the phase difference of two neighboring complex wavelet coefficients within a particular subband:

$$\Phi = \Theta_1 - \Theta_2,$$

where Θ_1 is the phase of $Z(k, l)$, the coefficient at position (k, l), and Θ_2 is the phase of $Z(k, l+1)$ (or $Z(k+1, l)$). The relative phase can be derived by assuming that the two adjacent complex coefficients, say Z_1 and Z_2, are characterized by the CGSM [70]. Using $N = 2$ for the CGSM pdf in (2.10), the pdf of the complex random vector $\mathbf{Z} = [Z_1, Z_2]^T$ is given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \int_0^\infty \frac{\exp\left(-\mathbf{z}^H (\mathbf{C}_{\mathbf{Z}|S})^{-1} \mathbf{z}\right)}{\pi^2 |\mathbf{C}_{\mathbf{Z}|S}|} f_S(s) \, ds.$$

Consequently, the conditional polar form pdf is described by

$$f_{\mathbf{R},\Theta|S}(r_1, r_2, \theta_1, \theta_2 | s) = r_1 r_2 \frac{\exp\left(-\mathbf{z}^H (\mathbf{C}_{\mathbf{Z}|S})^{-1} \mathbf{z}\right)}{\pi^2 |\mathbf{C}_{\mathbf{Z}|S}|},$$

where $Z_i = R_i e^{j\Theta_i}$, for $i = 1, 2$, and $\mathbf{C_{Z|S}} = E[\mathbf{ZZ}^H|s] = s\mathbf{C_{\tilde{Z}}} = s\begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{bmatrix}$. Let

$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{12}^* & b_{22} \end{bmatrix}$ be the inverse of $\mathbf{C_{Z|S}}$. Then, the conditional joint pdf of the phases is given by [38]

$$
\begin{aligned}
f_{\Theta|S}(\theta_1, \theta_2|s) &= \frac{1}{\pi^2}|\mathbf{B}| \int_0^\infty \int_0^\infty r_1 r_2 e^{-r_1^2 b_{11} - r_2^2 b_{22}} \\
&\quad \times e^{-2r_1 r_2 |b_{12}| \cos(\theta_1 - \theta_2 - \angle b_{12})} \, dr_1 dr_2 \\
&= \frac{1 - \lambda^2}{4\pi^2(1 - \gamma^2)}\left(1 - \frac{\gamma \arccos(\gamma)}{\sqrt{1 - \gamma^2}}\right),
\end{aligned}
$$

$$
\text{where} \quad \lambda = \frac{|b_{12}|}{\sqrt{b_{11} b_{22}}} = \frac{|sa_{12}|}{\sqrt{s^2 a_{11} a_{22}}} = \frac{|a_{12}|}{\sqrt{a_{11} a_{22}}},
$$

$$
\mu = \angle(-b_{12}) = \angle sa_{12} = \angle a_{12},
$$

and $\gamma = \lambda \cos(\theta_1 - \theta_2 - \mu + \pi)$. Therefore, the joint pdf is obtained by

$$
\begin{aligned}
f_\Theta(\theta_1, \theta_2) &= \int_0^\infty f_{\Theta|S}(\theta_1, \theta_2|s) f_S(s) \, ds \\
&= \frac{1 - \lambda^2}{4\pi^2(1 - \gamma^2)}\left(1 - \frac{\gamma \arccos(\gamma)}{\sqrt{1 - \gamma^2}}\right).
\end{aligned}
$$

Hence, the pdf of the relative phase $\Phi = \Theta_1 - \Theta_2$ is given by

$$
f_\Phi(\phi) = \frac{1 - \lambda^2}{2\pi(1 - \gamma^2)}\left(1 - \frac{\gamma \arccos(\gamma)}{\sqrt{1 - \gamma^2}}\right), \tag{2.12}
$$

where μ and λ are the two parameters of the pdf, and $\gamma = \lambda \cos(\phi - \mu + \pi)$. It should be noted that this pdf (2.12) is independent from the prior $f_S(s)$, and is therefore in the same form as the relative phase pdf for the complex Gaussian. The relative phase pdf (2.12) is used to parameterize the relative phase which extracts phase information of complex wavelet coefficients. It is then applied to texture image retrieval and improves the accuracy rate in [70].

2.5 Summary

In this chapter, the complex Gaussian scale mixture (CGSM) model has been proposed for modeling complex wavelet coefficients. The proposed CGSM can treat complex random variables appropriately. It has been shown that the real and imaginary parts of complex wavelet coefficients are related in such a way that their joint pdf can be expressed as a function of complex coefficients. Some related propositions and results of the CGSM are also discussed. Moreover, the pdf's related to the magnitude and phase based on the CGSM assumption are derived.

CHAPTER 3

THE MAGNITUDE AND RELATIVE PHASE DISTRIBUTION FOR COMPLEX
WAVELET MODELING

3.1 Introduction

Complex-valued wavelets have become more widely used in image processing
in the last decade because of several properties that do not exist in the real wavelet
domain [58]. As discussed in Chapter 1, the properties of complex wavelets which
are the advantages over the real ones directly relate to the magnitude and phase
information of the complex coefficients.

This chapter studies the statistical models for modeling the magnitude and
phase of complex wavelet coefficients in order to fully utilize the nice properties of
complex wavelets. To begin with, the generalized Gaussian distribution (GGD), which
is widely-used to capture the peaky and heavy-tail characteristics, is discussed for
modeling the real and imaginary parts of the complex coefficient. Furthermore, the
magnitude pdf that results from the CGSM which includes several special cases is
derived for magnitude modeling.

Despite the importance of the phase information, in the probabilistic framework,
most of statistical models based on complex wavelets utilize only the magnitude
information of the complex coefficients. It is mainly because the phase of complex
subband coefficients is uniformly distributed, which leads to difficulties to use the
phase information in statistical image processing. This suggests that a multivariate
phase model (e.g. a bivariate model) should be employed.

Recently in [68], the relative phase, which is defined as the difference between two complex wavelet phases, is introduced to extract the phase information from subband complex coefficients, and is then used for texture retrieval and segmentation. To model the relative phase, this chapter derives the relative phase pdf (RP pdf) based on the CGSM assumption.

In the statistical framework, the estimation of the pdf's parameters is central to many applications where it is required to extract the information from a set of random samples [26]. One of the most often used estimators is the maximum likelihood (ML) estimator, which is the estimator that maximizes the likelihood function, i.e., the ML estimator most likely caused the samples to occur once the sample set is given. For the RP pdf, an iterative method to find the ML estimator is needed because the corresponding log-likelihood equations are highly nonlinear. This is undesirable in some situations when the computational time is a primary concern. Even though the mean direction of the samples can be used as a non-iterative estimator for the pdf's location parameter, an iterative technique to estimate the concentration parameter is still necessary. Therefore, this chapter also proposes a parameter estimation method for the concentration parameter of the RP pdf which yields an estimator without using an iterative algorithm.

In addition, the problem of signal in noise is central to statistical signal processing theory [26]. Therefore, a method to estimate the parameters of several magnitude pdf's and the RP pdf's parameters when the underlying input image is contaminated by additive white Gaussian noise are also proposed.

3.2 Statistical Models of Complex Wavelet Subbands

The goal of this section is to discuss the statistical models that are used to model the subband coefficient in the complex wavelet domain. Specifically, this sec-

tion discusses the distributions which are suitable with the real/imaginary parts, the magnitude, and relative phase of complex coefficients. Moreover, for each distribution, its parameter estimation and Kullback-Leibler Divergence (KLD) which are used later in Section 4.4 to obtain features from images in the database and to obtain the distance between two images, respectively, are also presented.

3.2.1 The Generalized Gaussian Distribution for the Real/Imaginary Parts

Due to its peaky and heavy-tailed characteristics, the generalized Gaussian distribution (GGD) [34] has been commonly used to fit the marginal density of the subband coefficient in the real wavelet domain. The pdf of the GGD is defined as

$$f(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^{\beta}},$$

where $-\infty < x < \infty$, $\Gamma(\cdot)$ is the Gamma function [2], and $\alpha, \beta > 0$ are the pdf's parameters. Since the real and imaginary parts of a complex wavelet coefficient are real-valued and also behave like a real wavelet coefficient, it is reasonable to model the marginal density of the real or imaginary part of the complex coefficient using the GGD.

To estimate the model's parameters we use as features in our framework, we use the maximum likelihood (ML) estimation as discussed in details in [17]. We then provide a brief overview of the main results. Let $x_1, x_2, ..., x_N$ be a sample drawn from the GGD. The ML estimators for α and β are given by

$$\hat{\alpha} = \left(\frac{\hat{\beta}}{N} \sum_{i=1}^{N} |x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}},$$

where $\hat{\beta}$ is the solution of the equation

$$1 + \frac{\Psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^{N} |x_i|^{\hat{\beta}} \ln |x_i|}{\sum_{i=1}^{N} |x_i|^{\hat{\beta}}} + \frac{\ln\left(\frac{\hat{\beta}}{N} \sum_{i=1}^{N} |x_i|^{\hat{\beta}}\right)}{\hat{\beta}} = 0,$$

which can be solved by using a numerical algorithm such as the Newton-Raphson method. For the ML estimation of the GGD and other distributions in this work, we use the accuracy of 10^{-6} as the stopping criterion in the iterative method.

The similarity measurement between two GGD subbands can be obtained by computing the KLD which is given by [17]

$$
\begin{aligned}
D_{\mathrm{GG}}(f_1 \| f_2) \;=\; & \ln\left(\frac{\beta_1 \alpha_2 \Gamma(1/\beta_2)}{\beta_2 \alpha_1 \Gamma(1/\beta_1)}\right) + \\
& \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}.
\end{aligned}
$$

3.2.2 The Magnitude Models

Modeling the magnitude of complex coefficients is crucial in order to fully utilize its nice properties as discussed in Section 3.1. In this subsection, based on the CGSM assumption on the complex coefficient, the pdf for the coefficient magnitude is derived. In addition, several distributions which are special cases of the derived pdf are discussed.

To begin with, assume that the complex subband coefficient X is a CGSM random variable, i.e.

$$
X = \sqrt{S}\widetilde{X},
$$

where S is a unit-mean positive random variable, and $\widetilde{X}$ is zero-mean complex Gaussian with complex variance σ_X^2, and independent to S. Let R be the coefficient magnitude, i.e. $R = |X|$. Given $S = s$, $R|\{S = s\} = \sqrt{s}|\widetilde{X}| = \sqrt{s}\widetilde{R}$, where $\widetilde{R}$ follows the Rayleigh distribution whose pdf is defined by

$$
f_{\widetilde{R}}(\tilde{r}) = \frac{\tilde{r}}{\sigma_R^2}\, e^{-\frac{\tilde{r}^2}{2\sigma_R^2}},
$$

where σ_R is the Rayleigh parameter. Therefore, $R|\{S = s\}$ follows the Rayleigh distribution with parameter $\sqrt{s}\sigma_R$. Consequently, the pdf of R is given by

$$
\begin{aligned}
f_R(r) &= \int_0^\infty f_{R|S}(r|s)f_S(s)ds \\
&= \int_0^\infty \frac{r}{s\sigma_R^2}\, e^{-\frac{r^2}{2s\sigma_R^2}} f_S(s)ds.
\end{aligned}
$$

Note that $f_R(r)$ is the pdf of a Rayleigh random variable when $f_S(s)$ is an impulse function. Therefore, R is called a *Rayleigh scale mixture* (RSM) because of its behavior as a scaled version of a Rayleigh random variable when conditioned on S, and call this $f_R(r)$ the Rayleigh scale mixture pdf (RSM pdf). It should be noted that this pdf is also called the envelope pdf of the spherically-invariant random process (SIRP) in the context of wireless communication fading [77].

Unfortunately, there is generally no closed form for the KLD between two RSM models due to its hidden structure. This RSM pdf encompasses a variety of magnitude models, including Rayleigh, Weibull, and Gamma distributions [77], whose closed forms for the KLD are available. In the following, along with with the parameter estimation and the KLD expression, these three distributions which are used model the complex coefficient magnitude in this work are discussed. It should be noted that modeling the magnitude of the complex subband coefficient by these kinds of distributions is also discussed in [30]. Henceforth, $\{r_1, r_2, ..., r_N\}$ are defined as a sample drawn from the corresponding distribution in the parameter estimation for each of the models.

3.2.2.1 Rayleigh Distribution

To model the magnitude which is a non-negative quantity, a distribution with non-negative support is needed. One of the simplest and most commonly-used distri-

butions which belong to that category is the Rayleigh distribution, which is defined by

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}},$$

where $0 \leq r < \infty$, and $\sigma > 0$ is the parameter.

The ML estimator for σ has the closed-form expression which is given by

$$\hat{\sigma} = \sqrt{\frac{1}{2N} \sum_{i=1}^{N} r_i^2}.$$

The KLD between two Rayleigh subbands is given by

$$D_{\mathrm{Rl}}(f_1 \| f_2) = \left(\frac{\sigma_1}{\sigma_2}\right)^2 + 2\ln\left(\frac{\sigma_2}{\sigma_1}\right) - 1,$$

which is a closed-form expression as well. Even though modeling the coefficient magnitude by the Rayleigh distribution is convenient, it may not fit well with the magnitude histogram since there is only one pdf's parameter.

3.2.2.2 Weibull Distribution

To improve the model fitting from using the one-parameter distribution, we use the two-parameter models which have more degrees of freedom, i.e. more flexibility in modeling. One of widely-used two-parameter distributions is the Weibull distribution, whose pdf is given by

$$f_R(r) = \frac{\beta r^{\beta-1}}{\alpha^\beta} e^{-\left(\frac{r}{\alpha}\right)^\beta},$$

where $0 \leq r < \infty$, and $\alpha, \beta > 0$ are the two parameters.

To obtain the ML estimator, an alternative method is required to estimate Weibull's parameters α and β as suggested in [30]. Since $\ln R$ follows a Gumbel distribution if R follows a Weibull distribution, the parameter estimation of the Gumbel distribution and the transformation method are used to obtain the parameter estimates of the Weibull Distribution. The details are as follows.

To start with, let $T = \ln R$. It follows from the transformation method that the pdf of T is given by

$$f_T(t) = \frac{1}{\sigma} e^{\frac{t-\mu}{\sigma}} e^{-e^{\frac{t-\mu}{\sigma}}},$$

where $\mu > 0$ and $\sigma > 0$ are the two parameters. The ML estimators for μ and σ are given by

$$\hat{\mu} = \hat{\sigma} \ln \left(\frac{1}{N} \sum_{i=1}^{N} e^{\frac{t_i}{\hat{\sigma}}} \right),$$

where $\hat{\sigma}$ is the solution of the equation

$$\frac{1}{N} \sum_{i=1}^{N} s_i - \hat{\sigma} - \frac{\sum_{i=1}^{N} t_i e^{\frac{s_i}{\hat{\sigma}}}}{\sum_{i=1}^{N} e^{\frac{t_i}{\hat{\sigma}}}} = 0,$$

which can be solved numerically. Consequently, the estimators of α and β are given by

$$\hat{\alpha} = e^{\hat{\mu}} \quad \text{and} \quad \hat{\beta} = \frac{1}{\hat{\sigma}}.$$

The KLD between two Weibull subbands is given by

$$
\begin{aligned}
D_{\mathrm{Wbl}}(f_1 || f_2) &= \Gamma\left(\frac{\beta_2}{\beta_1} + 1\right)\left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} + \ln\left(\frac{\beta_1 \alpha_2^{\beta_2}}{\beta_2 \alpha_1^{\beta_1}}\right) \\
&\quad + (\beta_1 - \beta_2)\ln(\alpha_1) + \frac{\gamma_{\mathrm{const}}\alpha_2}{\alpha_1} - \gamma_{\mathrm{const}} - 1,
\end{aligned}
$$

where $\gamma_{\mathrm{const}} = 0.57721...$ is the Euler-Mascheroni constant.

3.2.2.3 Gamma Distribution

Another widely-used two-parameter model for the magnitude is the Gamma distribution. The Gamma pdf is defined as

$$f_R(r) = \frac{r^{\beta-1}}{\alpha^\beta \Gamma(\beta)} e^{-\left(\frac{r}{\alpha}\right)},$$

where $0 \leq r < \infty$, and $\alpha, \beta > 0$ are the two parameters. The ML estimators for α and β of the Gamma distribution are given by

$$\hat{\alpha} = \frac{\frac{1}{N} \sum_{i=1}^{N} r_i}{\hat{\beta}},$$

where $\hat{\beta}$ is the solution of the equation

$$\ln(\hat{\beta}) - \Psi(\hat{\beta}) - \ln\left(\frac{1}{N}\sum_{i=1}^{N}r_i\right) - \frac{1}{N}\ln\left(\sum_{i=1}^{N}r_i\right) = 0,$$

which can be solved numerically using the Newton-Raphson method. To compute the distance between two Gamma subbands, the KLD is given by

$$
\begin{aligned}
D_{\text{Gam}}(f_1||f_2) &= (\beta_1 - \beta_2)\Psi(\beta_1) - \beta_1 + \ln\left(\frac{\Gamma(\beta_2)}{\Gamma(\beta_1)}\right) \\
&\quad + \beta_2\ln\left(\frac{\alpha_2}{\alpha_1}\right) + \frac{\beta_1\alpha_1}{\alpha_2},
\end{aligned}
$$

where $\Psi(\cdot) = \frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$ is the digamma function [2].

In addition to three aforementioned distributions, the magnitude pdf of the CGGD we derived in Chapter 2 also belongs to a class of the RSM pdf. For the sake of completeness, this section also discusses this magnitude distribution of the CGGD and its parameter estimation and KLD expression.

3.2.2.4   Magnitude PDF of the CGGD

From Chapter 2, the magnitude pdf of the CGGD is defined by

$$f(r) = \frac{\beta r}{\alpha^2 \Gamma\left(\frac{2}{\beta}\right)} e^{-\left(\frac{r}{\alpha}\right)^{\beta}}$$

where $\alpha = \sigma\sqrt{\frac{2\Gamma\left(\frac{1}{\gamma}\right)}{\Gamma\left(\frac{2}{\gamma}\right)}}$, and $\beta = 2\gamma$ are the two parameters. The ML estimators for α and β are given by

$$\hat{\alpha} = \left(\frac{\hat{\beta}}{2N}\sum_{i=1}^{N}r_i^{\hat{\beta}}\right)^{1/\hat{\beta}},$$

where $\hat{\beta}$ is the solution of the equation

$$1 + \frac{2\Psi(2/\hat{\beta})}{\hat{\beta}} - \frac{2\sum_{i=1}^{N}r_i^{\hat{\beta}}\ln r_i}{\sum_{i=1}^{N}r_i^{\hat{\beta}}} + \frac{2\ln\left(\frac{\hat{\beta}}{2N}\sum_{i=1}^{N}r_i^{\hat{\beta}}\right)}{\hat{\beta}} = 0.$$

The KLD between two pdf's of this kind is given by

$$
\begin{aligned}
D_{\mathrm{GR}}(f_1 \| f_2) \ = \ & \ln\left(\frac{\beta_1 \alpha_2^2 \Gamma(2/\beta_2)}{\beta_2 \alpha_1^2 \Gamma(2/\beta_1)}\right) + \\
& \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2} \frac{\Gamma((\beta_2+2)/\beta_1)}{\Gamma(2/\beta_1)} - \frac{2}{\beta_1}.
\end{aligned}
$$

3.2.3 The Relative Phase Distribution

To fully utilize the advantages of the complex coefficient, one should not only consider the magnitude information but the phase information as well. To employ the phase information, the relative phase pdf (RP pdf) is introduced in [51, 67] as a statistical model for the relative phase which is the difference between two complex wavelet phases. The details of the RP pdf are as follows.

Let Θ_{X_1} be the phase of $X(k,l)$, the complex coefficient at position (k,l), and Θ_{X_2} be the phase of $X(k, l+1)$ (or $X(k+1, l)$). The relative phase [68] Θ at a spatial location (k,l) is defined as the phase difference of neighboring complex wavelet coefficients within a particular subband:

$$
\Theta = \Theta_{X_1} - \Theta_{X_2}. \tag{3.1}
$$

The RP pdf can be derived by assuming that the two adjacent complex coefficients, say X_1 and X_2, are characterized by the CGSM [52]. After some manipulation, the pdf of the relative phase $\Theta = \Theta_{X_1} - \Theta_{X_2}$ is obtained by

$$
f_\Theta(\theta; \mu, \lambda) = \frac{1-\lambda^2}{2\pi(1-\alpha^2)}\left(1 - \frac{\alpha \arccos(\alpha)}{\sqrt{1-\alpha^2}}\right), \tag{3.2}
$$

where $-\pi \le \mu < \pi$ and $0 \le \lambda \le 1$ are the two pdf's parameters and $\alpha = -\lambda\cos(\theta-\mu)$.

Let $\theta_1, \theta_2, ..., \theta_N$ be a random sample drawn from the RP distribution. The estimators for μ is given by

$$
\hat{\mu} = \arctan\left(\frac{\sum_{i=1}^{N}\sin\theta_i}{\sum_{i=1}^{N}\cos\theta_i}\right),
$$

(a)　　　　　　　　　　　(b)
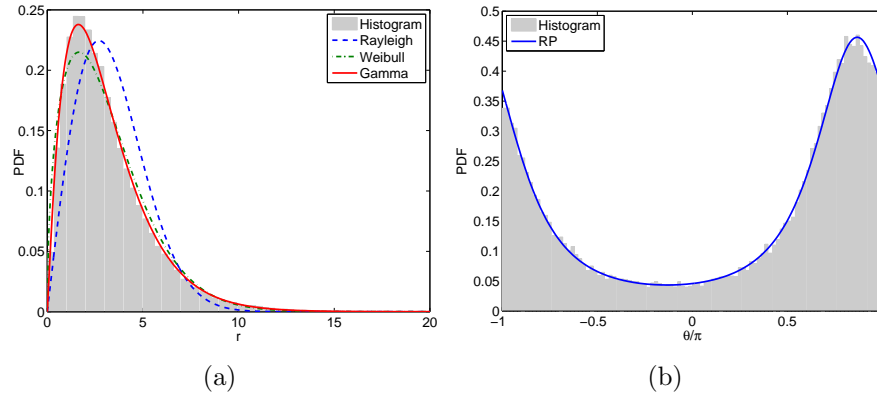
Figure 3.1. Examples of fitting the models with the histograms of the magnitude and relative phase from one finest-scale subband of the UDCT coefficients taken from the Fabric.0004 image: (a) magnitude, and (b) relative phase.

where $\arctan(\cdot)$ is the four-quadrant inverse tangent. On the other hand, the estimator of $\hat{\lambda}$ is the solution of the equation

$$
\begin{aligned}
0 = & \sum_{i=1}^{N} h(\hat{\lambda}, -1) + h(\hat{\lambda}, 1) - 1.5(h(\hat{\lambda}, -t_i) + h(\hat{\lambda}, t_i)) \\
& + \frac{b'(\hat{\lambda}, t_i)}{b(\hat{\lambda}, t_i)},
\end{aligned}
$$

where $t_i = -\cos(\theta_i - \hat{\mu})$, $h(\hat{\lambda}, t_i) = \frac{t_i}{1+\hat{\lambda}t_i}$, $b(\hat{\lambda}, t_i) = \sqrt{1 - \hat{\lambda}t_i^2} - \hat{\lambda}t_i \arccos(\hat{\lambda}t_i)$, and $b'(\hat{\lambda}, t_i) = -t_i \arccos(\hat{\lambda}t_i)$. See [67] for more details. Unfortunately, the closed-form expression of the KLD between two RP pdf's have not been found yet. Therefore, they are computed using a numerical method in this thesis.

To show how well the models discussed above fit with the empirical histograms, examples of the relevant fitted pdf's and the histograms of the magnitude and the relative are displayed in Fig. 3.1 for the Fabric.0004 image of the Vistex database [1]. The complex coefficients are obtained from a finest-scale subband of the UDCT [42] decomposition.

3.3   A Non-Iterative Estimator for the Concentration Parameter of the Relative Phase Distribution

This section proposes a parameter estimation method for the concentration parameter of the RP pdf which yields an estimator without using an iterative algorithm. First, a circular pdf that can well approximate the RP pdf is presented. To find the estimator, the log-likelihood equation is approximated by another equation with simpler form. Moreover, one parameter in the equation is further approximated so that it can be solved for the estimator immediately. This results in a non-iterative estimator for the concentration parameter. Furthermore, the parameter estimation of the simulated random samples is performed to compare the proposed estimator and the ML estimator in the aspect of accuracy and the computational time. Finally, the proposed method is used to fit the RP pdf to the histogram of relative phase samples extracted from some standard test images to see how well it can be used in practice.

3.3.1   An Approximation by Jones-Pewsey's PDF

The goal of this section is to propose an approximation of the RP pdf which can be used later in the parameter estimation. The pdf which is used to approximate the RP pdf is given by

$$g(\theta; \mu, \lambda) = \frac{(1 - \lambda^2)^{\frac{3}{4}}}{2\pi P^0_{-\frac{3}{2}}\left(\frac{1}{\sqrt{1-\lambda^2}}\right)} \frac{1}{(1 - \lambda \cos(\theta - \mu))^{\frac{3}{2}}}, \tag{3.3}$$

where $-\pi \leq \mu < \pi$, $0 \leq \lambda \leq 1$, and $P^0_{-\frac{3}{2}}(z)$ is the associated Legendre function of the first kind of degree $-\frac{3}{2}$ and order 0 [20]. This approximation is a special case of the symmetric circular pdf proposed by Jones and Pewsey [25].[1] As can be seen, the

---

[1]The pdf in [25] is defined by

$$f_\psi(\theta; \mu, \kappa, \psi) = \frac{(\cosh(\kappa\psi) + \sinh(\kappa\psi)\cos(\theta - \mu))^{\frac{1}{\psi}}}{2\pi P^0_{1/\psi}(\cosh(\kappa\psi))}, \tag{3.4}$$
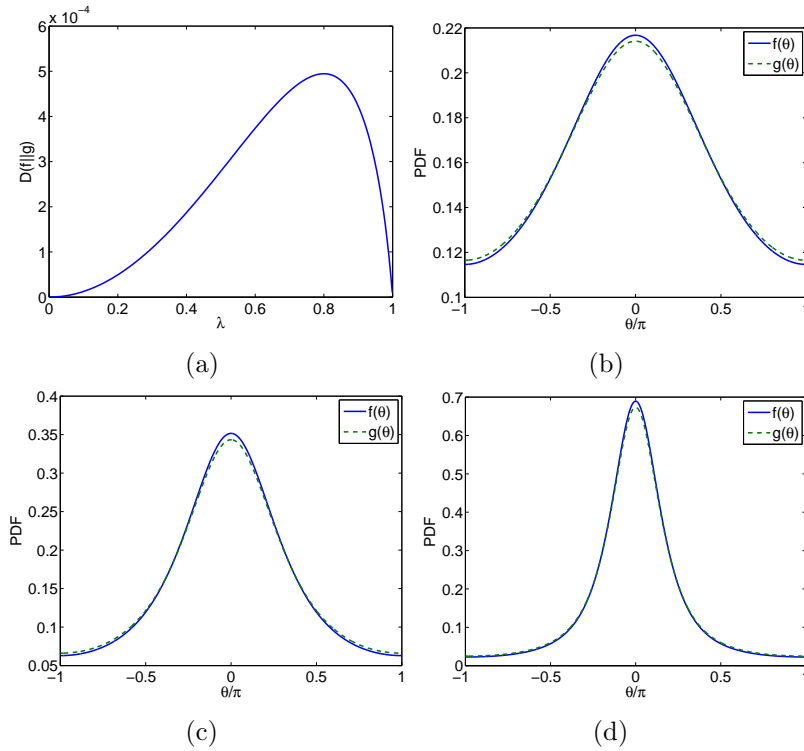
Figure 3.2. Comparison of the RP pdf $f(\theta; \lambda)$ and $g(\theta; \lambda)$: (a) KLD between f and g versus λ; (b) $f(\theta; \lambda)$ and $g(\theta; \lambda)$ when $\lambda = 0.2$; (c) when $\lambda = 0.5$; and (d) when $\lambda = 0.8$.

pdf in (3.3) is Jone-Pewsey's pdf with $\psi = -2/3$ and $\kappa = \frac{3}{2}\mathrm{arctanh}(\lambda)$. According to [25], the $n-$th trigonometric moment of the pdf in (3.3) is therefore obtained by

$$
\mathrm{E}[\cos n(\Theta - \mu)] = \frac{\Gamma\left(\frac{3}{2} - n\right) P^n_{-\frac{3}{2}}\left(\frac{1}{\sqrt{1-\lambda^2}}\right)}{\Gamma\left(\frac{3}{2}\right) P^0_{-\frac{3}{2}}\left(\frac{1}{\sqrt{1-\lambda^2}}\right)}.
\tag{3.5}
$$

It should be also noted that Jones-Pewsey's pdf including the pdf in (3.3) needs a numerical algorithm in the ML parameter estimation [25].

The comparison between $f(\theta; \mu, \lambda)$ and $g(\theta; \mu, \lambda)$ when $\mu = 0$ is shown in Fig. 3.2. The Kullback-Leibler distance (KLD) between $f(\theta; \lambda)$ and $g(\theta; \lambda)$ is rather small for all range of values of λ $(D(f||g) < 0.0006)$ as plotted in Fig. 3.2(a), which implies that $g(\theta; \lambda)$ is close to $f(\theta; \lambda)$. In addition, when $\lambda = 0$ or λ approaches one, the

where $P^\nu_{1/\psi}(z) \triangleq \frac{1}{\pi} \int_0^\pi \cos(\nu\theta)(z + \sqrt{z^2 - 1}\cos\theta)^{\frac{1}{\psi}} d\theta$, is the associated Legendre function of the first kind of degree $1/\psi$ and order ν [20].

two pdf's are the same because when they become the uniform pdf when $\lambda = 0$, and behave like an impulse function when λ approaches one. Moreover, it can be seen from the plots of $f(\theta; \lambda)$ and $g(\theta; \lambda)$ for $\lambda = 0.2$, 0.5, and 0.8 in Fig.'s 3.2(b) - 3.2(d) that the pdf $g(\theta; \lambda)$ can well approximate the RP pdf $f(\theta; \lambda)$.

3.3.2 Approximation of the Maximum Likelihood Estimator for the Concentration Parameter

The aim of this subsection is to use the approximation of the RP pdf discussed in Section 3.3.1 to obtain an approximation of the ML estimator. This subsection begins with a discussion of the ML estimation of the RP pdf.

The corresponding log-likelihood function of the samples $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$ is given by $L(\Theta; \mu, \lambda) = \ln \prod_{i=1}^{N} f(\theta_i; \mu; \lambda)$. Then, the ML estimators $\hat{\mu}_{\mathrm{ML}}$, and $\hat{\lambda}_{\mathrm{ML}}$ can be found as follows:

$$[\hat{\mu}_{\mathrm{ML}}, \hat{\lambda}_{\mathrm{ML}}] = \arg\max_{[\mu,\lambda]} L(\Theta; \mu; \lambda) = \arg\max_{[\mu,\lambda]} \sum_{i=1}^{N} \ln f(\theta_i; \mu, \lambda). \tag{3.6}$$

Consequently, the log-likelihood equations are given by

$$\frac{\partial L(\Theta; \mu, \lambda)}{\partial \mu} = 0, \quad \text{and} \quad \frac{\partial L(\Theta; \mu, \lambda)}{\partial \lambda} = 0.$$

These equations can be solved numerically to find the two estimators. Furthermore, the estimator $\hat{\mu}_{\mathrm{ML}}$ can be well approximated by the mean direction [70], which is given by

$$\hat{\mu}_{\mathrm{ML}} \approx \tilde{\mu} = \arctan \left(\frac{\sum_{i=1}^{N} \sin \theta_i}{\sum_{i=1}^{N} \cos \theta_i} \right), \tag{3.7}$$

where arctan is the four quadrant inverse tangent. Therefore, the problem reduces to finding the ML estimator for the concentration parameter λ, i.e., to find a solution of

$$g(\lambda) = \sum_{i=1}^{N} \left[\frac{-1}{1-\lambda} + \frac{1}{1+\lambda} - \frac{3}{2} \left(\frac{-x_i}{1-\lambda x_i} + \frac{x_i}{1+\lambda x_i} \right) \right.$$
$$\left. - \frac{x_i \arccos(\lambda x_i)}{\sqrt{1-\lambda^2 x_i^2} - \lambda x_i \arccos(\lambda x_i)} \right] = 0,$$

where $x_i = -\cos(\theta_i - \tilde{\mu})$. However, there is no closed form for $\hat{\lambda}_{\mathrm{ML}}$, which requires an iterative method.

To obtain an estimator for λ without numerical algorithm, one way to do is to approximate the likelihood equation. Without loss of generality, assume from now on that $\mu = 0$. The details are as follows.

To begin with, consider the RP pdf

$$
\begin{aligned}
f(\theta; \lambda) &= \frac{1-\lambda^2}{2\pi(1-\alpha^2)}\left(1 - \frac{\alpha\arccos(\alpha)}{\sqrt{1-\alpha^2}}\right) \\
&= \frac{1-\lambda^2}{2\pi}\frac{\partial^2}{\partial\alpha^2}\left(\arccos(\alpha)\right)^2 \\
&= \frac{1-\lambda^2}{2\pi}\int_0^\infty xK_0(x)e^{-\alpha x}dx,
\end{aligned}
\tag{3.8}
$$

where $\alpha = -\lambda\cos\theta$, $K_0(z) \triangleq \int_0^\infty e^{-z\cosh t}dt$ is the modified Bessel function of the second kind with order zero [2], and the last step follows from [32]. Using this expression of the RP pdf in (3.8), the log-likelihood function is given by

$$
L(\Theta; \mu, \lambda) = \ln\prod_{i=1}^N \frac{1-\lambda^2}{2\pi}\int_0^\infty xK_0(x)e^{\lambda\cos\theta_i x}dx.
$$

The corresponding log-likelihood equation is given by

$$
\begin{aligned}
0 &= \frac{\partial}{\partial\lambda}\sum_{i=1}^N \ln\left(\frac{1-\lambda^2}{2\pi}\int_0^\infty xK_0(x)e^{\lambda\cos\theta_i x}dx\right) \\
&= \frac{-2\lambda N}{1-\lambda^2} + \sum_{i=1}^N \frac{\cos\theta_i \int_0^\infty x^2 K_0(x)e^{\lambda\cos\theta_i x}dx}{\int_0^\infty xK_0(x)e^{\lambda\cos\theta_i x}dx}.
\end{aligned}
\tag{3.9}
$$

From [2], an approximation related to $K_0(x)$ is given by

$$
xK_0(x) \approx \sqrt{\frac{\pi}{2}}\sqrt{x}e^{-x}.
\tag{3.10}
$$

Although this approximation (3.10) is for large x, ($x \geq 2$), it is acceptable for small x too. Using (3.10) in (3.9), in order to obtain the estimator for λ, one has to solve the equation

$$
\begin{aligned}
0 &= \frac{-2\lambda N}{1-\lambda^2} + \sum_{i=1}^{N} \frac{\cos\theta_i \int_0^\infty \sqrt{\frac{\pi}{2}} x^{\frac{3}{2}} e^{-x} e^{\lambda\cos\theta_i x} dx}{\int_0^\infty \sqrt{\frac{\pi}{2}} \sqrt{x} e^{-x} e^{\lambda\cos\theta_i x} dx} \\
&= \frac{-2\lambda N}{1-\lambda^2} + \sum_{i=1}^{N} \frac{\cos\theta_i \Gamma(\frac{5}{2})(1-\lambda\cos\theta_i)^{-\frac{5}{2}}}{\Gamma(\frac{3}{2})(1-\lambda\cos\theta_i)^{-\frac{3}{2}}}.
\end{aligned}
$$

Equivalently, one needs to solve

$$
\frac{\lambda}{1-\lambda^2} = \frac{3}{4N} \sum_{i=1}^{N} \frac{\cos\theta_i}{1-\lambda\cos\theta_i}. \tag{3.11}
$$

Unfortunately, this equation still yields a non closed-form solution. To overcome this problem, the value of λ on the right-hand side of (3.11) should be approximated. One simple guess would be the value of λ from the method of moments (MM). However, the closed form expression of λ in terms of the moments is not available because of the complicated expression of the trigonometric moment for the RP pdf, which is given by

$$
\mathrm{E}[\cos n(\Theta - \mu)] = (-1)^n \lambda^n (1-\lambda^2) \frac{\Gamma^2(1+\frac{1}{2}n)}{\Gamma(1+n)} F(1+\tfrac{1}{2}n, 1+\tfrac{1}{2}n; 1+n; \lambda^2),
$$

where $F(a,b;c;z) \triangleq \frac{\Gamma(c)}{\Gamma(c-b)\Gamma(b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a} dt$ is the hypergeometric function, and $\Gamma(z) \triangleq \int_0^\infty e^{-t} t^{z-1} dt$ is the gamma function [2]. From Section 3.3.1, the RP pdf can be approximated by the pdf in (3.3) with the same value of λ. Therefore, instead of the trigonometric moment of the RP pdf, the moment of the pdf in (3.3) will be matched to approximate the value of λ on the right-hand side of (3.11). From [20],

$$
P_{-\frac{3}{2}}^2(z) + 2\frac{z}{\sqrt{z^2-1}} P_{-\frac{3}{2}}^1(z) = \frac{3}{4} P_{-\frac{3}{2}}^0(z). \tag{3.12}
$$

Using (3.5) and substituting z in (3.12) by $\frac{1}{\sqrt{1-\lambda^2}}$,

$$
-\frac{\mathrm{E}[\cos 2\Theta]}{4} + \frac{\mathrm{E}[\cos\Theta]}{\lambda} = \frac{3}{4}.
$$

Therefore, the approximation the estimator of λ from the method of moments (MM) is given by

$$
\hat{\lambda}_{\mathrm{AMM}} = \frac{4\sum_{i=1}^{N}\cos\theta_i}{3N + \sum_{i=1}^{N}\cos 2\theta_i}. \tag{3.13}
$$

Substituting $\hat{\lambda}_{\text{AMM}}$ for the right-hand side λ in (3.11), it reduces to a quadratic equation of λ and the estimator for λ is given by

$$\hat{\lambda}_{\text{AML}} = \frac{-1 + \sqrt{1 + 4C^2}}{2C}, \tag{3.14}$$

where $C = \frac{3}{4N} \sum_{i=1}^{N} \frac{\cos\theta_i}{1 - \hat{\lambda}_{\text{AMM}}\cos\theta_i}$.

3.3.3 Experiments in Parameter Estimation

To show the effectiveness of the proposed estimator, its comparison with the ML estimator for estimating the parameters of the RP pdf is shown by running two experiments; the first experiment to show the estimation accuracy and the computational time, and the second experiment to see how well the estimator can fit with a histogram of samples taken from standard test images. The experiments are run on Matlab implementation on 1.73GHz Pentium Dual-Core.

In the first experiment, relative phase samples generated with $\mu = 0$ and ten values of λ $(0, 0.1, ..., 0.9)$ are fitted using the ML estimator $[\hat{\mu}_{\text{ML}}, \hat{\lambda}_{\text{ML}}]$, the approximation of the ML (AML) estimator $[\tilde{\mu}, \hat{\lambda}_{\text{AML}}]$, and the approximation of the MM (AMM) estimator $[\tilde{\mu}, \hat{\lambda}_{\text{AMM}}]$, as in (3.6), (3.7), (3.14), and (3.13), respectively. For each given value of λ, 1000 independent trials are run to compensate the random nature of the simulated data. For the iterative step in the ML estimation method, the Newton-Raphson algorithm is used with $[\tilde{\mu}, \hat{\lambda}_{\text{AMM}}]$ as the initial guess. For each given value of λ in each trial, the absolute error is the absolute value of the estimate subtracted by the exact value of λ, i.e.,

$$\text{Absolute Error} = |\lambda e^{j\mu} - \hat{\lambda}e^{j\hat{\mu}}|,$$

where $[\hat{\mu}, \hat{\lambda}] = [\hat{\mu}_{\text{ML}}, \hat{\lambda}_{\text{ML}}], [\tilde{\mu}, \hat{\lambda}_{\text{AML}}]$, and $[\tilde{\mu}, \hat{\lambda}_{\text{AMM}}]$ for the case of the ML, AML, and AMM estimators, respectively. For each given number of samples (N), the errors are averaged all over the ten values of λ and 1000 trials.

Table 3.1. Comparison of the average absolute error and the average computational time of the ML, the approximation of the ML, and the approximation of the MM methods where N is the number of samples

N	Avg. Abs. Error/Avg. Time in sec.		
	ML: $[\hat{\mu}_{\mathrm{ML}}, \hat{\lambda}_{\mathrm{ML}}]$	AML: $[\tilde{\mu}, \hat{\lambda}_{\mathrm{AML}}]$	AMM: $[\tilde{\mu}, \hat{\lambda}_{\mathrm{AMM}}]$
100	$0.0887/5.5{\times}10^{-3}$	$0.0922/5.0{\times}10^{-4}$	$0.0978/4.5{\times}10^{-4}$
500	$0.0395/1.7{\times}10^{-2}$	$0.0414/1.1{\times}10^{-3}$	$0.0451/8.6{\times}10^{-4}$
1000	$0.0277/3.2{\times}10^{-2}$	$0.0292/1.8{\times}10^{-3}$	$0.0329/1.4{\times}10^{-3}$
5000	$0.0125/1.5{\times}10^{-1}$	$0.0137/7.9{\times}10^{-3}$	$0.0187/5.6{\times}10^{-3}$

The results of the first experiment are shown in Table 3.2. It can be seen that the average absolute errors due the proposed AML method are higher than those of the ML method but lower than the average absolute errors due to the AMM method for all cases of N. Also as expected, the error decreases while the computational time increases when we increase the number of samples N. It should be also noted that, when N increases, the error from the AML method approaches the error from the ML method. The difference between the average absolute errors from the two methods is only 0.0012 when $N = 5000$. In terms of the computational complexity, the time used in the AML method is less than the time used in the ML method (approximately around 10-20 times), and is approximately the same as the time used in the AMM method. This is because the proposed estimation method does not require an iterative method. These imply that the proposed AML estimator is acceptable in terms of accuracy, and efficient in terms of computation in the parameter estimation compared with the ML estimator.

In the second experiment, the proposed estimator is tested with the phase samples from the complex coefficients extracted from eight standard test gray-scale images of size 512×512 using the uniform discrete curvelet transform (UDCT) [41]. Eighteen subbands in the first three finest scales of the decomposition are used. The

values of the normalized relative entropy ($\Delta H/H$), which is defined as the ratio of the relative entropy (or KLD) between the fitted pdf and the histogram and the histogram entropy, are averaged for all eighteen subbands and displayed in Fig. 3.3(a). Moreover, the histogram fitting results of some images, which are the Lena, and the Barbara images, are shown in Fig's. 3.3(b)-3.3(c). From Fig. 3.3(a), it can be seen that the normalized relative entropy when using the ML estimator is the smallest for all images. It is because the ML estimator maximizes the likelihood function, and thus minimizes the KLD between the fitted pdf and the histogram. However, the difference between using the ML estimator and using the AML estimator is not much as it can be seen in the examples of fitting results in Fig.'s 3.3(b)-3.3(c). Therefore, the proposed AML estimator is also comparable with the ML estimator to fit the pdf with the histogram of phase samples of complex coefficients extracted from the standard images.

3.4 Estimating Parameters From Noisy Coefficients for the Real/Imaginary Parts and the Magnitude

In Section 3.2, the distributions used to model the complex coefficient are discussed. This section presents methods to estimate the parameters of those distributions when the underlying image is contaminated by additive white Gaussian noise (AWGN). The aim of this section is to have parameter estimation methods that are used in the feature extraction step of the noisy texture retrieval scheme in Section 4.4

The problem of an image in the presence of AWGN can be formulated as

$$\mathbf{y}_{\text{image}} = \mathbf{x}_{\text{image}} + \mathbf{w}_{\text{image}}, \tag{3.15}$$

where $\mathbf{y}_{\text{image}}$, $\mathbf{x}_{\text{image}}$ and $\mathbf{w}_{\text{image}}$ are the noisy, the clean, and AWGN images, respectively. The additive noise $\mathbf{w}_{\text{image}}$ is assumed to be independent from the clean image
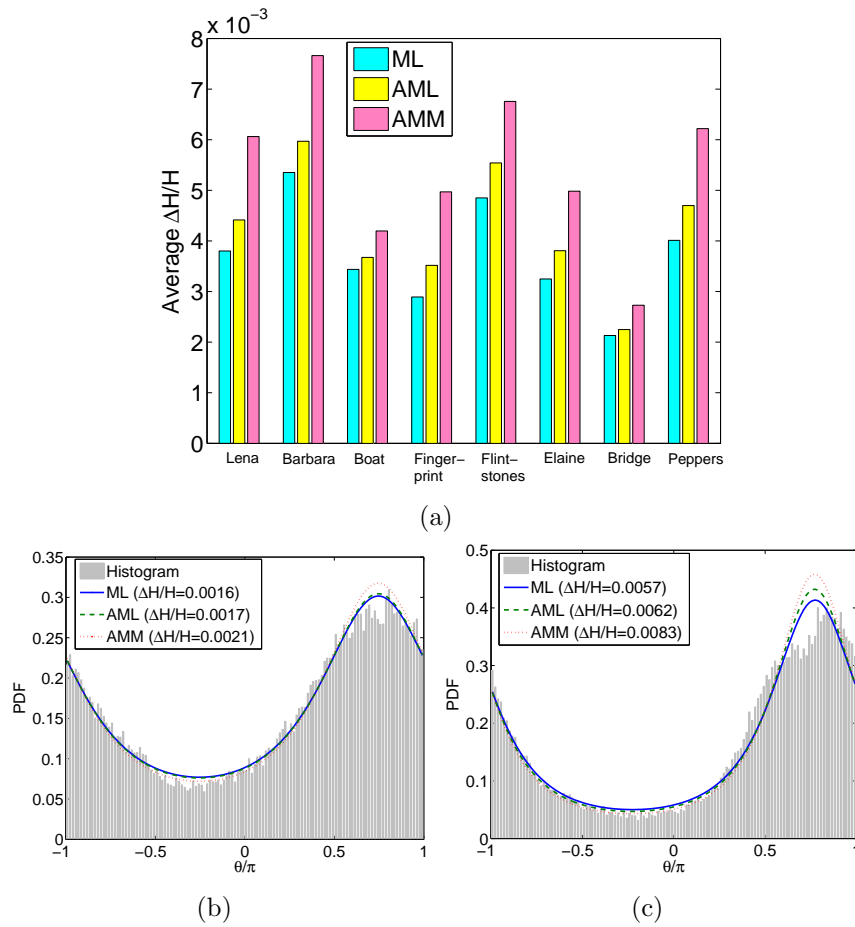
Figure 3.3. The histogram fitting results using the ML, the proposed AML, and the AMM estimators for several test images: (a) average normalized relative entropy for the eight test images; (b) example of fitting for the Lena image, and (c) for the Barbara image. The phase samples are from a particular finest subband.

$\mathbf{x}_{\mathrm{image}}$, and has a known variance σ_{n}^2. In the complex wavelet domain, the problem is given by

$$Y = X + W, \tag{3.16}$$

where Y, X and W are the noisy coefficient, clean coefficient, and complex Gaussian noise, respectively.

For the real-valued and magnitude models, the moment matching method is used to estimate the parameters of the clean coefficient from the noisy coefficient.

For the RP pdf, the same method however cannot be applied, which results in that a more complicated method to estimate the clean parameters is needed.

3.4.1   The GGD model

To begin with, the moment matching method for the GGD parameters is presented For the case of the real part of the coefficient, it is given by

$$Y_\mathrm{R} = X_\mathrm{R} + W_\mathrm{R},$$

where Y_R, X_R and W_R are the real parts of Y, X and W in (3.18), respectively. One can obtain the second and fourth moments of clean coefficients from noisy coefficients by

$$
\begin{aligned}
m_2 \triangleq E[X_\mathrm{R}^2] &= E[Y_\mathrm{R}^2] - \sigma_{W_\mathrm{R}}^2, \\
m_4 \triangleq E[X_\mathrm{R}^4] &= E[Y_\mathrm{R}^4] - 6E[Y_\mathrm{R}^2]\sigma_{W_\mathrm{R}}^2 + 3\sigma_{W_\mathrm{R}}^4,
\end{aligned}
$$

where $\sigma_{W_\mathrm{R}}^2$ is the variance of W_R, which can be computed offline.

For the GGD, the following relations between the two moments and the GG parameters can be obtained:

$$m_2 = \frac{\alpha^2 \Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})}, \quad m_4 = \frac{\alpha^4 \Gamma(\frac{5}{\beta})}{\Gamma(\frac{1}{\beta})}.$$

Therefore, the estimate of β can be obtained from

$$\frac{m_2^2}{m_4} = \frac{\Gamma^2(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})\Gamma(\frac{5}{\beta})}$$

by using the look-up table and linear interpolation. Consequently, the estimate of α is given by

$$\hat{\alpha} = \sqrt{\frac{m_4}{m_2} \frac{\Gamma(3/\hat{\beta})}{\Gamma(5/\hat{\beta})}}.$$

3.4.2   The Magnitude Models

In the case of complex-valued coefficients, one can obtain the second and fourth moments of the magnitude of clean coefficients from noisy coefficients by

$$m_2 \triangleq E[|X|^2] \;\; = \;\; E[|Y|^2] - \sigma_W^2,$$

$$m_4 \triangleq E[|X|^4] \;\; = \;\; E[|Y|^4] - 4E[|Y|^2]\sigma_W^2 + 2\sigma_W^4,$$

where σ_W^2 is the complex variance of the complex Gaussian noise W, which can be compute offline. For the Rayleigh distribution, once the second moment estimate m_2 is obtained, the estimate of the clean parameter is given by

$$\hat{\sigma} = \sqrt{\frac{m_2}{2}}.$$

Next, the two-parameter magnitude models are considered. For the Weibull distribution, the two parameters and the two moments have the following relationship:

$$m_2 = \alpha^2 \Gamma\left(1 + \frac{2}{\beta}\right), \quad m_4 = \alpha^4 \Gamma\left(1 + \frac{4}{\beta}\right).$$

Therefore, the estimate of β can be obtained from

$$\frac{m_2^2}{m_4} = \frac{\Gamma^2(1 + \frac{2}{\beta})}{\Gamma(1 + \frac{4}{\beta})}$$

by using the look-up table and linear interpolation. Consequently, the estimate of α is given by

$$\hat{\alpha} = \sqrt{\frac{m_4}{m_2} \frac{\Gamma(1 + 2/\hat{\beta})}{\Gamma(1 + 4/\hat{\beta})}}.$$

For the Gamma distribution, the Gamma parameters and the two moments have the following relationship:

$$m_2 = \alpha^2 \frac{\Gamma(\beta + 2)}{\Gamma(\beta)}, \quad m_4 = \alpha^4 \frac{\Gamma(\beta + 4)}{\Gamma(\beta)}.$$

Therefore, the estimate of β can be obtained from

$$\frac{m_2^2}{m_4} = \frac{\Gamma^2(\beta+2)}{\Gamma(\beta)\Gamma(4+\beta)}$$

by using the look-up table and linear interpolation. Consequently, the estimate of α is given by

$$\hat{\alpha} = \sqrt{\frac{m_4}{m_2}\frac{\Gamma(\hat{\beta}+2)}{\Gamma(\hat{\beta}+4)}}.$$

3.5 Estimating the Relative Phase Parameters of Complex Wavelet Coefficients in Noise

This section proposes a method to estimate the RP pdf's parameters when the underlying input image is contaminated by additive white Gaussian noise. First, the pdf of the relative phase of the noisy coefficients is dervied. Then, the connection between the RP pdf's parameters and the complex covariance matrix of the corresponding coefficient vector are utilized to relate the considered RP pdf's parameters to the derived pdf. Subsequently, the RP pdf's parameters of the clean coefficients are estimated by using the maximum likelihood method. The simulation results are also displayed to illustrate the effectiveness in estimation of the proposed method.

3.5.1 The relative phase mixture pdf (RPM pdf)

The problem of an image in the presence of additive white Gaussian noise can be formulated as

$$\mathbf{y}_{\text{image}} = \mathbf{x}_{\text{image}} + \mathbf{w}_{\text{image}}, \qquad (3.17)$$

where $\mathbf{y}_{\text{image}}$, $\mathbf{x}_{\text{image}}$ and $\mathbf{w}_{\text{image}}$ are the noisy, the clean, and additive white Gaussian noise images, respectively. The additive noise $\mathbf{w}_{\text{image}}$ is assumed to be independent from the clean image $\mathbf{x}_{\text{image}}$, and has a known variance σ_{n}^2.

Transforming (3.17) into the complex wavelet domain, one can obtain the following equation in each neighborhood of complex wavelet coefficients as

$$\mathbf{Y} = \mathbf{X} + \mathbf{W}, \tag{3.18}$$

where $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{W}$ are the noisy coefficient, clean coefficient, and additive complex Gaussian noise vectors (but not necessarily white), respectively. [2]

In case of the neighborhood of two coefficients, (3.18) reduces to

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}.$$

Assume that $\mathbf{X}$ is a CGSM, i.e. $\mathbf{X} = \sqrt{S}\widetilde{\mathbf{X}}$ where $\widetilde{\mathbf{X}}$ is complex Gaussian, S is a unit-mean positive random variable independent from $\widetilde{\mathbf{X}}$. Let $\mathbf{C}_{\widetilde{\mathbf{X}}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{bmatrix}$ and $\mathbf{C_W} = \begin{bmatrix} b_{11} & b_{12} \\ b_{12}^* & b_{22} \end{bmatrix}$ be the complex covariance matrices of $\widetilde{\mathbf{X}}$ and $\mathbf{W}$, respectively. Therefore,

$$\mathbf{C_Y} - \mathbf{C_W} = \mathbf{C_X} = E[S]\mathbf{C}_{\widetilde{\mathbf{X}}} = \mathbf{C}_{\widetilde{\mathbf{X}}}. \tag{3.19}$$

Let $\Theta = \angle X_1 - \angle X_2$ be the relative phase of the clean coefficient vector $\mathbf{X}$ as defined in (3.1). From (3.2), the pdf of Θ is given by

$$f_{\Theta}(\theta; \mu_x, \lambda_x) = \frac{1 - \lambda_x^2}{2\pi(1 - \alpha_x^2)}\left(1 - \frac{\alpha_x \arccos(\alpha_x)}{\sqrt{1 - \alpha_x^2}}\right),$$

where $\alpha_x = -\lambda_x \cos(\theta - \mu_x)$, and μ_x and λ_x are the two parameters.

In order to find the parameters μ_x and λ_x, one needs to determine their relationship with the pdf of the noisy relative phase. To begin with, the noisy relative

---

[2]The cross-covariance matrix of two vectors obtained from sampling two wide-sense stationary processes which form a Hilbert transform pair needs not be zero [52]. As a result, the complex covariance matrix of a white noise vector in the complex wavelet domain may not be a multiple of the identity matrix.

phase Φ is defined as the relative phase which belongs to the noisy coefficient vector $\mathbf{Y}$:

$$\Phi = \angle Y_1 - \angle Y_2.$$

If the noisy coefficient vector $\mathbf{Y}$ is also a CGSM (i.e., the noisy relative phase Φ has the RP pdf), then it is straightforward to determine the relationship between the RP parameters of Φ and μ_x, λ_x using (2.12) and (3.19). Unfortunately, $\mathbf{Y}$ is not necessarily a CGSM even though $\mathbf{X}$ and $\mathbf{W}$ are independent CGSM's. The conditions which ensure that $\mathbf{Y}$ is also a CGSM are discussed in the following proposition.

Proposition 3.5.1. *The sum of two CGSM's $\mathbf{X}_1 = \sqrt{S_1}\widetilde{\mathbf{X}}_1$ and $\mathbf{X}_2 = \sqrt{S_2}\widetilde{\mathbf{X}}_2$ with $[S_1, \widetilde{\mathbf{X}}_1]$ independent from $[S_2, \widetilde{\mathbf{X}}_2]$ and complex covariance matrices $\mathbf{C}_{\mathbf{X}_1}$ and $\mathbf{C}_{\mathbf{X}_2}$, respectively, is a CGSM if and only if one of the following two statements is true:*

1) $\mathbf{C}_{\mathbf{X}_2} = a\mathbf{C}_{\mathbf{X}_1}$ for some positive constant a,

2) $\mathbf{X}_1$ and $\mathbf{X}_2$ are both complex Gaussian. $\qquad\square$

Proof: See Appendix B. $\qquad\blacksquare$

It should be noted that a statement similar to Proposition 3.5.1 is mentioned for the case of elliptically distributed random vectors in [78]. From Proposition 3.5.1, the noisy coefficient vector $\mathbf{Y}$ does not have to be a CGSM, and the noisy relative phase Φ may not have the RP pdf.

To find the pdf of Φ, consider the conditional pdf of $\mathbf{Y}$ given S. Since $\mathbf{X}$ and $\mathbf{W}$ are independent, given $S = s$, $\mathbf{Y}|\{S = s\}$ is complex Gaussian whose pdf is given by

$$f_{\mathbf{Y}|S}(\mathbf{y}|s) = \frac{1}{\pi^2|\mathbf{C}_{\mathbf{Y}|S}|}e^{-\mathbf{y}^H\mathbf{C}_{\mathbf{Y}|S}^{-1}\mathbf{y}},$$

where $\mathbf{C}_{\mathbf{Y}|S} = s\mathbf{C}_{\widetilde{\mathbf{X}}} + \mathbf{C}_{\mathbf{W}} = \begin{bmatrix} sa_{11} + b_{11} & sa_{12} + b_{12} \\ sa_{12}^* + b_{12}^* & sa_{22} + b_{22} \end{bmatrix}$, and $|\mathbf{C}_{\mathbf{Y}|S}|$ is the determinant

of $\mathbf{C}_{\mathbf{Y}|S}$. Since the complex Gaussian distribution is a special case of the CGSM distribution, from (3.2) and (2.12), the conditional pdf of Φ can be obtained by

$$f_{\Phi|S}(\phi; \mu(s), \lambda(s)|s) = \frac{1-\lambda^2(s)}{2\pi(1-\alpha^2(s))}\left(1 - \frac{\alpha(s)\arccos(\alpha(s))}{\sqrt{1-\alpha^2(s)}}\right), \qquad (3.20)$$

where $\alpha(s) = -\lambda(s)\cos(\phi - \mu(s))$, $\mu(s) = \angle(sa_{12} + b_{12})$ and $\lambda(s) = \frac{|sa_{12} + b_{12}|}{\sqrt{(sa_{11}+b_{11})(sa_{22}+b_{22})}}$ are the pdf's parameters dependent on s. Hence, the pdf of Φ is then given by

$$\begin{aligned} f_\Phi(\phi) &= \int_0^\infty f_{\Phi|S}(\phi|s) f_S(s) ds \\ &= \int_0^\infty \frac{1-\lambda^2(s)}{2\pi(1-\alpha^2(s))}\left(1 - \frac{\alpha(s)\arccos(\alpha(s))}{\sqrt{1-\alpha^2(s)}}\right) f_S(s) ds. \end{aligned} \qquad (3.21)$$

This pdf (3.21) is called the relative phase mixture pdf (RPM pdf). Note that the RPM pdf becomes the RP pdf when $f_S(s) = \delta(s-1)$ or when $a_{ij} = cb_{ij}$ for all $1 \le i, j \le 2$, where c is a positive constant.

An example of the relative phase of the noisy coefficients $\mathbf{Y}$ and the fitted RP and the fitted RPM pdf's is displayed in Fig. 3.4 for the Brick.0001 image of the Vistex database [1]. The complex coefficients are obtained by using the uniform discrete curvelet transform (UDCT) [42]. The hidden multiplier S is assumed to follow the log-normal distribution. How to estimate the parameters of the log-normal pdf from the complex coefficients is discussed in Subsection 4.4.2. As can be seen, the relative phase of the noisy coefficients fits well with the RPM pdf, and differs from the relative phase of the clean coefficients.

3.5.2 Experiments in Parameter Estimation in Gaussian Noise

This section presents a method to estimate the RP parameters of clean coefficients from noisy coefficients. To estimate the RP parameters associated with $\mathbf{X}$,
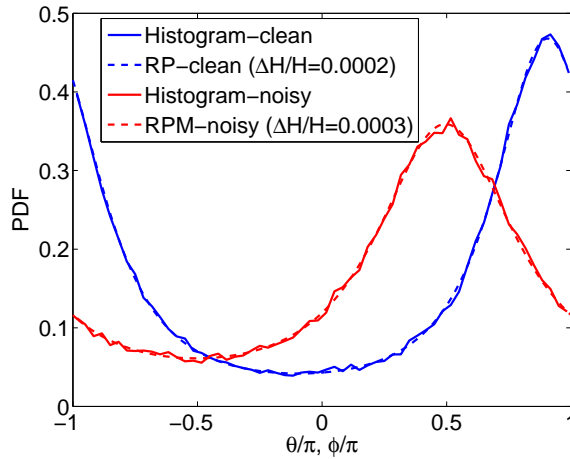
Figure 3.4. Example of fitting the histogram of clean relative phase with the fitted RP pdf, and noisy relative phase samples with the fitted RP pdf and the fitted RPM pdf of one finest-scale subband of the Brick.0001 image. The noise standard deviation is $\sigma_{\mathrm{n}} = 30$.

μ_x and λ_x are represented as a complex parameter $\eta = \eta_1 + j\eta_2 = \lambda_x e^{j\mu_x}$. It follows from (2.12) that $\sqrt{a_{11}a_{22}}\eta = a_{12}$. Therefore, the conditional pdf of Φ in (3.20) can be expressed as

$$f_{\Phi|S}(\phi; \eta|s) = \frac{1 - \lambda^2(\eta, s)}{2\pi(1 - \alpha^2(\eta, s))} \left(1 - \frac{\alpha(\eta, s)\arccos(\alpha(\eta, s))}{\sqrt{1 - \alpha^2(\eta, s)}}\right),$$

$$\text{where} \quad \alpha(\eta, s) = -\lambda(\eta, s)\cos(\phi - \mu(\eta, s)),$$

$$\mu(\eta, s) = \angle(s\sqrt{a_{11}a_{12}}\eta + b_{12}),$$

$$\text{and} \quad \lambda(\eta, s) = \frac{|s\sqrt{a_{11}a_{22}}\eta + b_{12}|}{\sqrt{(sa_{11} + b_{11})(sa_{22} + b_{22})}}.$$

As a result, the pdf of Φ is rewritten by

$$f_{\Phi}(\phi; \eta) = \int_0^{\infty} f_{\Phi|S}(\phi; \eta|s) f_S(s) ds.$$

Hence, the pdf of Φ can be viewed as a function of the complex parameter η, i.e. a function of the vector $\boldsymbol{\eta} = [\eta_1, \eta_2]^T$. To find the estimator of $\boldsymbol{\eta}$ from the observation set of Φ, $\{\phi_1, \phi_2, \ldots, \phi_N\}$, the maximum likelihood method is used. The log-likelihood function for $f_{\Phi}(\phi)$ is given by

$$L(\Phi; \boldsymbol{\eta}) = \ln \left(\prod_{i=1}^{N} f_\Phi(\phi_i; \boldsymbol{\eta}) \right).$$

Consequently, the maximum likelihood (ML) estimator for $\hat{\boldsymbol{\eta}}$ can be found as follows:

$$\begin{aligned} \hat{\boldsymbol{\eta}}_{\mathrm{ML}} &= \operatorname*{argmax}_{\boldsymbol{\eta}} L(\Phi; \boldsymbol{\eta}) = \operatorname*{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{N} \ln f_\Phi(\phi_i; \boldsymbol{\eta}) \\ &= \operatorname*{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{N} \ln \left(\int_0^\infty f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s) f_S(s) ds \right). \end{aligned}$$

An example of the log-likelihood function $L(\Phi; \boldsymbol{\eta})$ corresponding to the examples in Fig. 3.4 is shown in Fig. 3.5. As can be seen, the log-likelihood function is concave, i.e. this is a convex optimization problem.

In practice, one needs to find

$$\hat{\boldsymbol{\eta}}_{\mathrm{ML}} = \operatorname*{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{N} \ln \left(\sum_{n=1}^{N_s} f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s_n) p_n \right),$$

where $p_n = C f_S(s_n)$, C is a constant such that $\sum p_n = 1$, which can be done by using an iterative method to solve for $\hat{\boldsymbol{\eta}}_{\mathrm{ML}}$. See Appendix C for more details.   Once we have $\hat{\boldsymbol{\eta}}_{\mathrm{ML}}$, the estimators for λ_x and μ_x are given by

$$\hat{\lambda}_x = ||\hat{\boldsymbol{\eta}}_{\mathrm{ML}}||, \quad \text{and} \quad \hat{\mu}_x = \angle(\hat{\eta}_{\mathrm{ML},1} + j\hat{\eta}_{\mathrm{ML},2}). \tag{3.22}$$

It is worth noting that if we model the relative phase by other widely used circular pdf's such as the von Mises or the wrapped Cauchy pdf's [37], it is difficult to relate the pdf's parameters of the noisy coefficients to those of the clean coefficients because, unlike the RP pdf, these circular pdf's are not derived based on modeling the complex coefficients by a complex-valued model. This is one of the advantages of modeling the relative phase of complex coefficients by the RP pdf (3.2).

To validate the proposed estimation method, two experiments in estimating the RP parameters μ_x and λ_x of $\mathbf{X}$ from the relative phase samples of $\mathbf{Y}$ are performed;
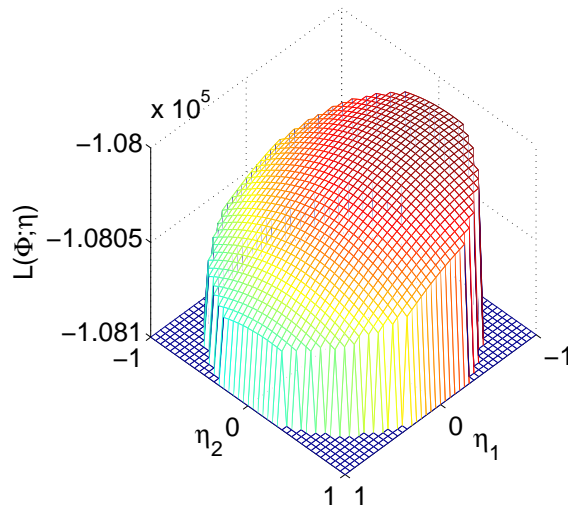
Figure 3.5. Example of the log-likelihood function $L(\Phi; \boldsymbol{\eta}), \boldsymbol{\eta} = [\eta_1, \eta_2]^T, ||\boldsymbol{\eta}|| < 1$, where phase samples are from a finest-scale subband extracted from the Brick.0001 image when $\sigma_{\mathrm{n}} = 30$.

the first experiment with simulated samples, and the second experiment with the actual samples taken from standard test images. We run experiments using Matlab implementation on a 2.13GHz Core 2 Duo with 2GB RAM.

In each experiment, the proposed method which we call the RPM method is compared with the other two methods described as follows.

1. Noisy method: This method blindly assumes that the noisy observation from $\mathbf{Y}$ is clean and estimate the RP pdf of the noisy coefficients, i.e.,

$$\hat{\lambda}_x = \hat{\lambda}_y \quad \text{and} \quad \hat{\mu}_x = \hat{\mu}_y. \tag{3.23}$$

$\hat{\mu}_y$ and $\hat{\lambda}_y$ can be estimated from the samples of the relative phase associated with $\mathbf{Y}$ by using the ML method [70].

2. RP method: This method assumes that the pdf for the relative phase of $\mathbf{Y}$ is the RP pdf, and then predict μ_x and λ_x using the RP pdf parameters of $\mathbf{Y}$ and $\mathbf{W}$. The RP pdf parameters $\mu_y, \lambda_y, \mu_w, \lambda_w$ can be estimated using the ML estimation. Note that, given the complex transform, the estimators $\hat{\mu}_w$ and $\hat{\lambda}_w$

of the RP pdf's parameters of $\mathbf{W}$ can be computed offline. Let $\mathbf{C_Y} = \begin{bmatrix} c_{11} & c_{12} \\ c_{12}^* & c_{22} \end{bmatrix}$.

Therefore, $c_{12} - b_{12} = a_{12}$ because $\mathbf{C_Y} - \mathbf{C_W} = \mathbf{C_X}$. Using the relationship in (2.12), μ_x and λ_x can be estimated by

$$
\hat{\lambda}_x = \begin{cases} \dfrac{|\hat{\lambda}_y e^{j\hat{\mu}_y} - \hat{\lambda}_w e^{j\hat{\mu}_w}|}{\sqrt{(c_{11} - b_{11})(c_{22} - b_{22})}}, & \text{if } \hat{\lambda}_x \leq 0.999; \\ 0.999, & \text{if } \hat{\lambda}_x > 0.999, \end{cases}
$$
$$
\text{and } \hat{\mu}_x = \angle(\hat{\lambda}_y e^{j\hat{\mu}_y} - \hat{\lambda}_w e^{j\hat{\mu}_w}), \tag{3.24}
$$

where the condition on $\hat{\lambda}_x$ is presented to prevent computational instability since the value of λ_x satisfies $0 \leq \lambda_x \leq 1$.

In the first experiment, the clean and noise vectors are simulated, and their summation is the noisy vector. The RP parameters μ_x and λ_x are then estimated using the Noisy, the RP, and the proposed RPM methods as in (3.23), (3.24), and (3.22), respectively. Specifically, the data is generated as follows.

1. Generate $\widetilde{\mathbf{X}}$ and $\mathbf{W}$ as 2×1 complex Gaussian random vectors with $\mathbf{C_{\widetilde{X}}} = \begin{bmatrix} \sigma_x^2 & \sigma_x^2 \lambda_x e^{j\mu_x} \\ \sigma_x^2 \lambda_x e^{-j\mu_x} & \sigma_x^2 \end{bmatrix}$ and $\mathbf{C_W} = \begin{bmatrix} 1 & \lambda_w e^{j\mu_w} \\ \lambda_w e^{-j\mu_w} & 1 \end{bmatrix}$, respectively. Hence, the signal-to-noise ratio (SNR) is equal to σ_x^2. The values of μ_x and μ_w are set as $\mu_x = 0.25\pi$ and $\mu_w = 0.75\pi$.

2. Set $\mathbf{X} = \sqrt{S}\widetilde{\mathbf{X}}$, where S is log normal with unit mean.

3. Obtain $\mathbf{Y} = \mathbf{X} + \mathbf{W}$.

The parameters $\hat{\mu}_x$ and $\hat{\lambda}_x$ are obtained using the three estimation methods mentioned above. For the RPM method, the number of points S of the interval $[s_{\min}, s_{\max}]$ is set as $N_s = 20$. The interval $[s_{\min}, s_{\max}]$ are chosen such that $F_S(s_{\min}) = 0.005$, and $F_s(s_{\max}) = 0.995$, where $F_S(s)$ is the cumulative density function (cdf) which

is defined by $F_S(s) = \int_{-\infty}^{s} f_S(t)dt$. This is to assure that the range of integration sufficiently covers the support of $f_S(s)$.

For each given value of SNR, λ_x and λ_w, the estimation experiment is run for 100 independent trials. The values λ_x and λ_w are taken from $\{0.25, 0.5, 0.75\}$, which result in nine pairs of (λ_x, λ_w). To show the performance of the estimation methods, the root mean squared error (RMSE) between the true parameter and the estimate is computed, i.e.,

$$\text{RMSE} = \sqrt{\frac{1}{100} \sum_{i=1}^{100} |\lambda_x e^{j\mu_x} - \hat{\lambda}_x e^{j\hat{\mu}_x}|^2}.$$

For each value of the number of samples (N) and each value of SNR, the absolute errors are averaged all over the nine pairs of (λ_x, λ_w) and 100 trials.

The hidden multiplier S in Step 2 is set to be log normal with unit mean, i.e.,

$$f_S(s; \mu_{\text{L}}, \sigma_{\text{L}}) = \frac{1}{s\sigma_{\text{L}}\sqrt{2\pi}} \exp\left(-\frac{(\ln s - \mu_{\text{L}})^2}{2\sigma_{\text{L}}^2}\right),$$

where $\mu_{\text{L}} = -\sigma_{\text{L}}^2/2$. To estimate the parameters of $f_S(s)$ from the noisy coefficients, the moment matching method is used, and is given by

$$\begin{aligned}
\hat{\mu}_{\text{L}} &= -\frac{1}{2}\ln(E[S^2]), \quad \hat{\sigma}_{\text{L}}^2 = \ln(E[S^2]), \\
E[S^2] &= \frac{E[|Y|^4] - 4\sigma_Y^2\sigma_W^2 + 2\sigma_W^4}{2(\sigma_Y^2 - \sigma_W^2)^2}.
\end{aligned} \tag{3.25}$$

The expression of $E[S^2]$ in (4.5) can be computed from $Y = X + W = \sqrt{S}\widetilde{X} + W$, where Y, X, $\widetilde{X}$ and W are the univariate version of $\mathbf{Y}$, $\mathbf{X}$, $\widetilde{\mathbf{X}}$ and $\mathbf{W}$, respectively, as defined in (3.18), and σ_Y^2 and σ_W^2 are the variances of Y and W, respectively. Once $\hat{\mu}_{\text{L}}$ and $\hat{\sigma}_{\text{L}}$ are obtained, the limits of the interval $[s_{\min}, s_{\max}]$ can be obtained by $s_{\min} = \exp(\hat{\mu}_{\text{L}} - 2.80703\hat{\sigma}_{\text{L}})$ and $s_{\max} = \exp(\hat{\mu}_{\text{L}} + 2.80703\hat{\sigma}_{\text{L}})$.

The results are shown in Table 3.2, where S has the log normal pdf with $\mu_{\text{L}} = -\frac{1}{2}$ and $\sigma_{\text{L}} = 1$. Apparently, the proposed RPM method performs better than the RP

Table 3.2. Comparison of the average RMSE between the true parameter and the estimates of the three estimation methods where N is the number of samples, and SNR is the signal-to-noise ratio. The unit of the average computational time is sec

N	SNR	RMSE		
		Noisy	RP	RPM
100	0.5	0.5489	0.3222	0.3226
	1	0.4566	0.2515	0.2395
	5	0.2261	0.1482	0.1322
	10	0.1538	0.1141	0.1104
	Time Avg.	3.3×10^{-3}	3.4×10^{-3}	4.8×10^{-2}
500	0.5	0.5434	0.2005	0.1645
	1	0.4472	0.1790	0.1232
	5	0.2100	0.1119	0.0751
	10	0.1339	0.0797	0.0672
	Time Avg.	1.1×10^{-2}	1.1×10^{-2}	2.0×10^{-1}
1000	0.5	0.5420	0.1775	0.1257
	1	0.4459	0.1657	0.0929
	5	0.2080	0.1059	0.0621
	10	0.1320	0.0748	0.0544
	Time Avg.	1.9×10^{-2}	1.9×10^{-2}	3.8×10^{-1}
5000	0.5	0.5401	0.1539	0.0811
	1	0.4454	0.1568	0.0684
	5	0.2068	0.1017	0.0345
	10	0.1301	0.0703	0.0282
	Time Avg.	7.1×10^{-2}	7.1×10^{-2}	1.5×10^{0}

method, and much better than the Noisy method. The average absolute error from the proposed method is less than those from the other two methods. This is because the Noisy method does not consider the fact that the observation data of $\mathbf{Y}$ is noisy, while the RP method, though concerns about the presence of noise, ignores the fact that the relative phase of $\mathbf{Y}$, Φ, has the RPM pdf, not the RP pdf. In terms of the number of samples (N), one can see that the performance of the proposed estimation method is improved when the number of samples increases from 100 to 5000 for a fixed value of SNR. If the number of samples N is fixed, the average absolute error of the

proposed method reduces when the value of SNR increases from 0.5 to 10. However, in the aspect of computational complexity, the RPM method is quite computationally complex as can be seen from the time average in Table 3.2. It is due to the non closed form of the pdf of Φ.

3.6   Summary

In this chapter, the statistical models for modeling the magnitude and relative phase of complex wavelet coefficients have been studied. Based on the CGSM assumption on the complex coefficient, the general cases of the magnitude and relative phase pdf's are derived. Several well-known special cases of the derive magnitude pdf and the relative phase pdf (RP pdf) are shown to fit well with the histograms obtained from the complex coefficients. Moreover, a parameter estimation method for the concentration parameter of the RP pdf which yields an estimator without using an iterative algorithm is proposed. Furthermore, the phase pdf related to the CGSM called the relative phase pdf (RP pdf) is discussed and its non-iterative estimator is proposed as an alternative for the exact maximum likelihood (ML) estimator which requires numerical algorithm. Besides, an estimation method is presented to estimate the magnitude pdf's parameters and the RP pdf's parameters when the underlying input image is contaminated by additive white Gaussian noise. The simulation results show that the non-iterative estimator gives acceptably accurate estimates with less complexity, and that the proposed estimation method in noise is satisfactory for the case of simulated data.

CHAPTER 4

APPLICATIONS

4.1 Introduction

This chapter provides some applications of statistical modeling using the CGSM in the complex wavelet domain. Firstly, the CGSM model is applied to image denoising. Secondly, the magnitude pdf of the CGGD is used to model the coefficient magnitude and is applied to the application of texture retrieval. Finally, the parameter estimation methods in the presence of noise that are proposed in Section 3 are utilized in the application of noisy texture retrieval. To commence with, this chapter discusses the application of the CGSM in image denoising.

4.2 Image Denoising

To show the effectiveness, the proposed CGSM is used to model the complex subband coefficients of an image for image denoising application. The details are as follows.

4.2.1 Bayes Least Squares Estimator for CGSM

One of the best methods for image denoising is the Bayes least squares (BLS) estimator based on the GSM model presented in [47]. For each neighborhood, the reference coefficient at center of the neighborhood is estimated from the set of observed coefficients. The subband coefficients are real, and the pdf is a function of the real variable. The BLS method based on the GSM model is used to estimate the real subband coefficients. However, if one uses a complex transform which decomposes

an image into subbands of complex coefficients, then an algorithm that can handle complex numbers is needed. Therefore, by the analogy with the BLS algorithm for the GSM, the BLS estimator based on the CGSM for estimating the complex coefficients is developed.

Let $\mathbf{V}$ be the random vector corresponding to a neighborhood of N observed complex coefficients

$$\mathbf{V} = \mathbf{Z} + \mathbf{W},$$

where $\mathbf{Z}$ is an original complex coefficient vector and $\mathbf{W}$ is a complex noise vector in the transform domain. Suppose that $\mathbf{W}$ is a zero-mean complex Gaussian and $\mathbf{Z}$ is a CGSM random vector as shown in (2.10). It is well known that the Bayes least squares estimation is the conditional expectation when $\mathbf{Z}$ and $\mathbf{V}$ are real random vectors as follows

$$\widehat{\mathbf{Z}} = E[\mathbf{Z}|\mathbf{V}], \tag{4.1}$$

$$\text{and} \quad E[\mathbf{Z}|\mathbf{V}] = \int_0^\infty f_{S|\mathbf{V}}(s|\mathbf{v})E[\mathbf{Z}|\mathbf{V} = \mathbf{v}, S = s]\,ds, \tag{4.2}$$

where $f_S(s)$ is the pdf of the positive scalar random variable S. In this implementation, the integration in (4.2) is computed numerically, where K is the number of points for s, by

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

When conditioned on S and $\mathbf{V}$, the conditional expectation is obtained by

$$E[\mathbf{Z}|\mathbf{V} = \mathbf{v}, S = s] = s\mathbf{C}_{\widetilde{\mathbf{Z}}}(s\mathbf{C}_{\widetilde{\mathbf{Z}}} + \mathbf{C}_{\mathbf{W}})^{-1}\mathbf{v}, \tag{4.3}$$

where the complex covariance matrices are given by $\mathbf{C}_{\widetilde{\mathbf{Z}}} = E[\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^H]$ and $\mathbf{C}_{\mathbf{W}} = E[\mathbf{W}\mathbf{W}^H]$.

The neighborhood noise covariance matrix $\mathbf{C}_{\mathbf{W}}$ is obtained by decomposing a random noise image which has Gaussian distribution with mean zero, standard

deviation σ and has the same dimension as the original image into subbands. Given $\mathbf{C_W}$, the covariance matrix $\mathbf{C_{\widetilde{Z}}}$ can be computed from the observation covariance matrix $\mathbf{C_V}$, i.e., $\mathbf{C_V} = E[S\widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^H] + E[\mathbf{W}\mathbf{W}^H] = E[S]\mathbf{C_{\widetilde{Z}}} + \mathbf{C_W}$. Since $E[S] = 1$, $\mathbf{C_{\widetilde{Z}}} = \mathbf{C_V} - \mathbf{C_W}$. To force the complex covariance matrix to be positive semidefinite, an eigenvalue decomposition of $\mathbf{C_{\widetilde{Z}}}$ is performed and any possible negative eigenvalues are set to be zero.

The pdf of the observed neighborhood vector conditioned on S is zero-mean complex Gaussian with covariance $\mathbf{C_{V|S}} = s\mathbf{C_{\widetilde{Z}}} + \mathbf{C_W}$,

$$f_{\mathbf{V}|S}(\mathbf{v}|s) = \frac{\exp\left(-\mathbf{v}^H(s\mathbf{C_{\widetilde{Z}}} + \mathbf{C_W})^{-1}\mathbf{v}\right)}{\pi^N |s\mathbf{C_{\widetilde{Z}}} + \mathbf{C_W}|}.$$

To estimate $\mathbf{Z}$, $f_{S|\mathbf{V}}(s|\mathbf{v})$ as in (4.2) is computed as follows

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

where the prior $f_S(s)$ is chosen to be Jeffrey's noninformative prior [6] for the experiments in this paper. According to Jeffrey's rule, $f_S(s)$ satisfies

$$f_S(s) \propto \sqrt{I(s)},$$

where $I(s)$ is Fisher's information measure which is defined by [6]

$$I(s) = E\left[-\frac{\partial^2 \ln f_{\mathbf{Z}|S}(\mathbf{z}|s)}{\partial s^2}\right].$$

For the CGSM model in (2.10),

$$-\frac{\partial^2 \ln f_{\mathbf{Z}|S}(\mathbf{z}|s)}{\partial s^2} = \frac{\partial^2}{\partial s^2}\left[N\ln(s) + \ln|\mathbf{C_{\widetilde{z}}}| + \frac{\mathbf{z}^H \mathbf{C_{\widetilde{z}}}^{-1}\mathbf{z}}{s}\right]$$

$$= \frac{-N}{s^2} + \frac{2\mathbf{z}^H \mathbf{C_{\widetilde{z}}}^{-1}\mathbf{z}}{s^3}.$$

Since $E[\mathbf{z}^H\mathbf{C_{\widetilde{Z}}}^{-1}\mathbf{z}] = s$, by taking the square root of the expectation, Jeffrey's prior is obtained by

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

Note that, in particular, Jeffrey's prior of the CSGM obtained herein is the same as that of the GSM obtained in [47]. Even though this $f_S(s)$ is an improper pdf, in this implementation, $f_S(s)$ is set to be zero in the interval $[0, s_{\min})$ with a small positive value of $s_{\min}$ to cope with this fact. The points for s are sampled with logarithmically uniform spacing, where 13 points of $\ln(s)$ over an interval $[\ln(s_{\min}), \ln(s_{\max})] = [-20.5, 3.5]$ with steps of size two are used. These parameters for the prior are selected according to [47].

4.2.2   Experiment 1

In this subsection, the BLS method described in Section 4.2.1 is implemented for image denoising in the UDCT domain. First, an image is decomposed into subbands using the uniform discrete curvelet transform (UDCT) [41]. The decomposition consists of oriented bandpass bands at five scales: twelve directions in the finest scale, and six directions for the other coarser scales. Each subband except the lowpass subband is denoised by using the BLS-CGSM estimator as described above. Then, the denoised image is reconstructed from the processed subbands. The image is corrupted by additive white Gaussian noise with known variance. The three standard test images of size 512×512, Lena, Barbara, and Boats, are used in this denoising experiment.

The results are displayed in Table 4.1 for the peak signal-to-noise ratio (PSNR) values, and in Table 4.2 for the structure similarity (SSIM) indices [73], where the standard deviation of the input noise is varied between $\sigma = 10$ and $\sigma = 100$. In this paper, the PSNR value is defined as $\text{PSNR} = 20 \log_{10}(255/\sigma_\epsilon)$, where σ_ϵ is the error standard deviation.

First, the proposed CGSM method is compared with the other three methods in the UDCT domain. The three methods are the hard thresholding (HT) method,

Table 4.1. PSNR values (in dB) of denoised images using various denoising methods in the UDCT domain

σ	Noisy PSNR	UDCT HT	UDCT GSMsp	UDCT GSMag	UDCT CGSM
		Lena			
10	28.13	33.58	34.10	34.13	35.49
15	24.61	31.93	31.87	31.83	33.82
20	22.11	30.85	30.38	30.11	32.59
25	20.17	29.99	29.11	28.81	31.62
50	14.15	27.23	24.74	24.28	28.56
75	10.63	25.68	21.94	21.94	26.87
100	8.13	25.07	19.85	21.42	25.73
		Barbara			
10	28.13	31.79	33.19	33.06	33.85
15	24.61	29.62	30.81	30.58	31.83
20	22.11	28.23	29.20	28.84	30.43
25	20.17	27.15	27.91	27.41	29.36
50	14.15	23.83	23.15	22.78	26.21
75	10.63	22.36	20.83	20.47	24.44
100	8.13	21.64	19.00	19.15	23.19
		Boats			
10	28.13	31.35	32.61	32.75	33.37
15	24.61	29.66	30.27	30.36	31.54
20	22.11	28.48	28.68	28.75	30.27
25	20.17	27.55	27.58	27.50	29.30
50	14.15	24.86	23.71	23.29	26.36
75	10.63	23.51	21.11	20.96	24.76
100	8.13	22.57	19.53	19.53	23.74

the GSM method of the real and the imaginary parts of complex coefficients separately (GSMsp), and the method using GSM of the augmented vector of the real and imaginary parts of complex coefficients (GSMag). For the HT method, any complex coefficient whose magnitude is below the threshold is set to zero, where the threshold in each subband is 3 times of the noise standard deviation of that subband. The neighborhood used in the CGSM and GSM methods is hand-optimized to be the

Table 4.2. SSIM indices of denoised images using various denoising methods in the UDCT domain

σ	Noisy SSIM	UDCT HT	UDCT GSMsp	UDCT GSMag	UDCT CGSM
			Lena		
10	0.614	0.885	0.877	0.876	0.910
15	0.451	0.860	0.819	0.813	0.887
20	0.344	0.841	0.761	0.747	0.868
25	0.272	0.825	0.702	0.684	0.851
50	0.113	0.764	0.463	0.435	0.777
75	0.061	0.724	0.315	0.315	0.728
100	0.037	0.695	0.226	0.301	0.698
			Barbara		
10	0.716	0.900	0.906	0.900	0.926
15	0.579	0.858	0.855	0.842	0.899
20	0.478	0.820	0.803	0.782	0.874
25	0.401	0.785	0.751	0.722	0.851
50	0.196	0.657	0.504	0.477	0.744
75	0.111	0.581	0.361	0.341	0.661
100	0.070	0.545	0.265	0.273	0.593
			Boats		
10	0.692	0.894	0.856	0.860	0.878
15	0.537	0.831	0.793	0.795	0.842
20	0.426	0.788	0.733	0.734	0.812
25	0.346	0.752	0.678	0.675	0.785
50	0.155	0.618	0.453	0.434	0.679
75	0.086	0.565	0.309	0.306	0.609
100	0.054	0.532	0.226	0.242	0.564

3×3 block without the parent. From Tables 4.1 and 4.2, one can see that the CGSM significantly outperforms the HT, and the GSMsp methods for all three test images in terms of both PSNR values and SSIM indices. It is because, in the GSMsp method, the real and the imaginary parts are denoised separately, where the fact that they are the real and the imaginary parts of complex coefficients into account.
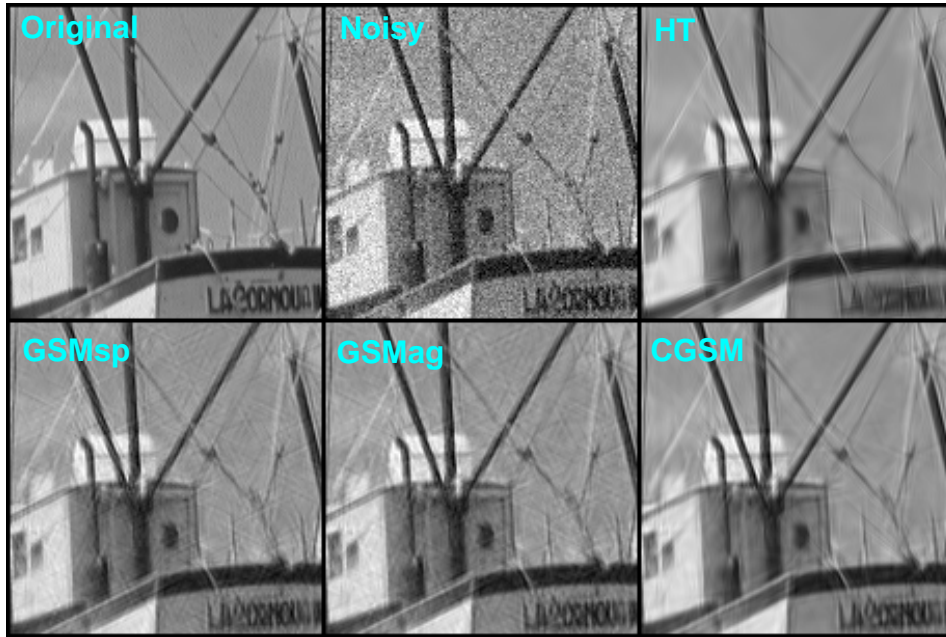
Figure 4.1. Denoising performance comparison of the Boats image (cropped to 150×150) with the UDCT. **Top-left**: original image, **Top-middle**: noisy image PSNR 22.11dB ($\sigma = 20$), **Top-right**: UDCT-HT PSNR 28.48dB, **Bottom-left**: UDCT-GSMsp PSNR 28.68dB, **Bottom-middle**: UDCT-GSMag PSNR 28.75dB, and **Bottom-right**: UDCT-CGSM PSNR 30.27dB.

Compared with the GSMag method, the CGSM method performs considerably better than the GSMag method. According to the shift-invariance property of complex wavelet transforms, it is suggested that the magnitude (or the complex form) of a complex coefficient should be used, i.e., treating complex coefficients as in the CGSM method is suggested for the better results. As a matter of fact, there are two equivalent ways to treat circular complex random vectors: treating them as having complex covariance matrices (as in CGSM), or as having real covariance matrices (as in GSMag). It appears that the results of the GSMag method should be as good as those of the CGSM method. However, this is not true as will be explained in what follows.

In covariance matrix calculation, once the signal covariance matrix is obtained, it is forced to be positive semidefinite. For the case where complex coefficients are treated as having complex covariance matrices, an eigenvalue decomposition of the complex covariance matrix $\mathbf{C}_{\widetilde{\mathbf{Z}}}$ is performed and any possible negative eigenvalues are set to be zero. On the other hand, if the $N \times 1$ complex random vector $\widetilde{\mathbf{Z}}$ is treated as the $2N \times 1$ augmented vector of the real and imaginary part vectors $[\widetilde{\mathbf{X}}^T, \widetilde{\mathbf{Y}}^T]^T$ having the real covariance matrix, one needs to set any possible negative eigenvalues of the real covariance matrix $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\widetilde{\mathbf{X}}} & \mathbf{C}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}} \\ \mathbf{C}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{X}}} & \mathbf{C}_{\widetilde{\mathbf{Y}}} \end{bmatrix}$ to be zero. To see this, let $\lambda_1, \lambda_2, ..., \lambda_{2N}$ be the eigenvalues of $\mathbf{C}$ in non-increasing order $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{2N}$, and $\rho_1, \rho_2, ..., \rho_N$ be the eigenvalues of $\mathbf{C}_{\widetilde{\mathbf{Z}}}$ with $\rho_1 \geq \rho_2 \geq ... \geq \rho_N$. If the circular condition is satisfied, i.e, if $\mathbf{C}_{\widetilde{\mathbf{X}}} = \mathbf{C}_{\widetilde{\mathbf{Y}}}$ and $\mathbf{C}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}} = -\mathbf{C}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{X}}}$, then $\lambda_{2i-1} = \lambda_{2i}$ for $1 \leq i \leq N$, which means that its eigenvalues have even multiplicity [56]. Therefore, the number of possibly negative eigenvalues of $\mathbf{C}$ is even. It also follows from the circularity that there is a relationship: $\rho_i = 2\lambda_{2i-1} = 2\lambda_{2i}$, for $1 \leq i \leq N$ [56]. Therefore, the number of negative eigenvalues of $\mathbf{C}$ is an even number, and twice the number of negative eigenvalues of $\mathbf{C}_{\widetilde{\mathbf{Z}}}$.

However, in practice where the circular condition is not ideally satisfied, this condition on the number of negative eigenvalues might not be true. One example of this possibility is given, where the Barbara image is corrupted by additive white Gaussian noise with standard deviation $\sigma = 20$. The neighborhood coefficients are taken from 3×3 blocks in one finest subband of the UDCT decomposition, which results in the 9×9 complex covariance matrix $\mathbf{C}_{\widetilde{\mathbf{Z}}}$, and the 18×18 real covariance matrix $\mathbf{C}$. All nine eigenvalues (ρ_i) of $\mathbf{C}_{\widetilde{\mathbf{Z}}}$ and eighteen eigenvalues (λ_i) of $\mathbf{C}$ are tabulated in Table 4.3. As can be seen, there are five negative eigenvalues of $\mathbf{C}$, i.e., $\lambda_i, 14 \leq i \leq 18$, are set to zero while only one eigenvalue ρ_9 of $\mathbf{C}_{\widetilde{\mathbf{Z}}}$ is set to

Table 4.3. Example of eigenvalues of $\mathbf{C}$ and $\mathbf{C}_{\widetilde{\mathbf{z}}}$. See text for details.

Matrix	Eigenvalues					
	105.18	92.80	54.09	43.41	42.97	26.32
$\mathbf{C}$	20.90	19.88	15.75	12.75	6.38	5.89
	3.47	-0.86	-1.09	-3.18	-7.34	-18.23
$\mathbf{C}_{\widetilde{\mathbf{z}}}$	189.92	77.73	53.92	41.18	30.73	12.19
	8.80	6.58	-1.95			

zero. As a consequence, the results from using the GSMag method are not as good as (actually much worse than) those results using the CGSM method. Therefore, it is less appropriate to treat the real and imaginary parts of complex coefficients as the augmented vectors having real covariance matrices at least in image denoising framework as in this thesis.

An example of visual denoising performance using these four methods for the Boats image when $\sigma = 20$ is illustrated in Fig. 4.1. Obviously, the CGSM method yields a visually better denoised image as it preserves the image details with less artifacts. It should be also noted that the GSMsp and GSMag methods, which treat complex coefficients inappropriately, yield even worse results for most of noise levels than the simple HT method for the Lena image.

4.2.3 Experiment 2

In this experiment, image denoising is performed using CGSM with other complex transforms: DT-$\mathbb{C}$WT, PDTDFB, and FDCT. For all complex-valued transforms using the CGSM method, the neighborhood is the 3×3 block without the parent as in the first experiment.

For PDTDFB and FDCT, images are decomposed into four scales to produce the best results. For DT-$\mathbb{C}$WT, five scales of decomposition in denoising are used.

Table 4.4. PSNR values (in dB) of denoised images of the FS-GSM, and DT-ℂWT BiShrink methods, and the CGSM methods with various complex transforms

σ	Noisy PSNR	FS GSM	DT-ℂWT BiShrink	UDCT CGSM	DT-ℂWT CGSM	PDTDFB CGSM	FDCT CGSM
Lena							
10	28.13	**35.60**	35.30	35.49	35.50	35.33	35.46
15	24.61	**33.90**	33.58	33.82	33.72	33.57	33.72
20	22.11	**32.67**	32.33	32.59	32.40	32.34	32.43
25	20.17	**31.69**	31.36	31.62	31.35	31.40	31.42
50	14.15	**28.61**	28.27	28.56	28.06	28.29	28.19
75	10.63	**26.88**	26.45	26.87	26.32	26.48	26.35
100	8.13	25.70	25.20	**25.73**	25.15	25.20	25.07
Barbara							
10	28.13	34.02	33.30	33.85	34.01	33.89	**34.24**
15	24.61	31.83	31.17	31.83	31.79	31.87	**32.16**
20	22.11	30.27	29.66	30.43	30.25	30.47	**30.70**
25	20.17	29.07	28.53	29.36	29.07	29.41	**29.58**
50	14.15	25.42	25.21	26.21	25.50	26.15	**26.28**
75	10.63	23.61	23.53	24.44	23.60	24.32	**24.46**
100	8.13	22.58	22.48	**23.19**	22.46	23.03	23.16
Boats							
10	28.13	**33.58**	32.99	33.37	33.49	33.28	33.38
15	24.61	**31.69**	31.24	31.54	31.51	31.36	31.48
20	22.11	**30.36**	29.95	30.27	30.13	30.01	30.14
25	20.17	**29.33**	28.94	29.30	29.09	28.99	29.13
50	14.15	26.32	25.97	**26.36**	25.95	26.12	26.08
75	10.63	24.73	24.42	**24.76**	24.33	24.55	24.44
100	8.13	23.72	23.39	**23.74**	23.31	23.50	23.36

The corresponding PSNR values and SSIM indices are tabulated in Tables 4.4 and 4.5, respectively.

Comparing all complex transforms with the CGSM method, the UDCT outperform than the other three transforms for the Lena and Boats images in terms of both PSNR value and SSIM index. For the Barbara image, the UDCT performs better than the DT-ℂWT, and approximately the same as the PDTDFB-CGSM in terms

of PSNR value, but in terms of SSIM index the UDCT rather dominates the two transforms. Considering the Barbara image denoising results of the FDCT, one can see that it outperforms the UDCT in both aspects but it has higher redundancy as shown in Table 4.6.

The CGSM method is also compared with the GSM method using full steerable pyramid (FS-GSM) with eight directions [47], and the bivariate shrinkage method using DT-ℂWT (DT-ℂWT-BiShrink) [61]. For the denoising using FS, an image is decomposed into five scales. The FS-GSM method outperforms other methods including the UDCT-CGSM in terms of both PSNR value and SSIM index especially for the Lena and Boats images. However, for the Barbara image, all complex transforms except the DT-ℂWT, which provides approximately the same results, yield better denoising results in both PSNR and SSIM aspects, where the FDCT gives the best results. This is because in the Barbara image contains oriented line-type structures which can be better represented the curvelet-based transforms. When comparing the DT-ℂWT BiShrink method with the proposed CGSM method, the DT-ℂWT is used for a fair comparison. Based on the DT-ℂWT, the BiShrink method performs approximately the same as the CGSM method for the Lena image, but is outperformed for the other two images. Nevertheless, it should be mentioned that in order to use the CGSM method in the DT-ℂWT domain the DT-ℂWT coefficients in the finest scale is assumed to satisfy the circular condition although this is not true as it is shown in Section . According to the results, this approximation is acceptable. Furthermore, modeling the DT-ℂWT coefficients in the first scale can improve the results but it is outside the scope of this work. However, the UDCT-CGSM method still outperforms the DT-CWT BiShrink method for all three images. Fig. 4.2 shows the visual performance comparison of the Barbara image with different methods. Apparently, the CGSM produces the better denoised image in terms of visual quality compared

Table 4.5. SSIM indices of denoised images of the FS-GSM, and DT-ℂWT BiShrink methods, and the CGSM methods with various complex transforms

σ	Noisy SSIM	FS GSM	DT-ℂWT BiShrink	UDCT CGSM	DT-ℂWT CGSM	PDTDFB CGSM	FDCT CGSM
				Lena			
10	0.614	**0.911**	0.906	0.910	0.910	0.907	0.910
15	0.451	**0.887**	0.882	**0.887**	0.885	0.880	0.886
20	0.344	**0.868**	0.861	**0.868**	0.862	0.858	0.864
25	0.272	0.850	0.843	**0.851**	0.842	0.843	0.844
50	0.113	**0.781**	0.768	0.777	0.761	0.772	0.768
75	0.061	**0.731**	0.710	0.728	0.709	0.716	0.713
100	0.037	0.691	0.665	**0.698**	0.669	0.668	0.669
				Barbara			
10	0.716	**0.930**	0.919	0.926	**0.930**	0.925	**0.930**
15	0.579	0.899	0.888	0.899	0.900	0.897	**0.903**
20	0.478	0.869	0.858	0.874	0.870	0.871	**0.878**
25	0.401	0.839	0.830	0.851	0.841	0.846	**0.854**
50	0.196	0.710	0.713	0.744	0.710	0.737	**0.749**
75	0.111	0.622	0.629	0.661	0.617	0.666	**0.668**
100	0.070	0.563	0.570	0.593	0.551	0.597	**0.601**
				Boats			
10	0.692	**0.883**	0.865	0.878	0.881	0.878	0.879
15	0.537	**0.845**	0.829	0.842	0.840	0.838	0.840
20	0.426	**0.813**	0.797	0.812	0.805	0.803	0.807
25	0.346	0.784	0.769	**0.785**	0.774	0.772	0.777
50	0.155	0.678	0.664	**0.679**	0.660	0.668	0.668
75	0.086	**0.611**	0.596	0.609	0.586	0.602	0.596
100	0.054	**0.564**	0.548	**0.564**	0.535	0.552	0.545

to the GSM and Bishrink methods. In summary, the proposed CGSM method can achieve high quality image denoising, especially for line-typed or oriented details as in the Barbara image. For the Lena and Boats images, the UDCT-CGSM method yields comparable results (≈ -0.05dB and ≈ -0.002 for SSIM index in average) to those of the FS-GSM method but with lower redundancy as shown in Table 4.6.
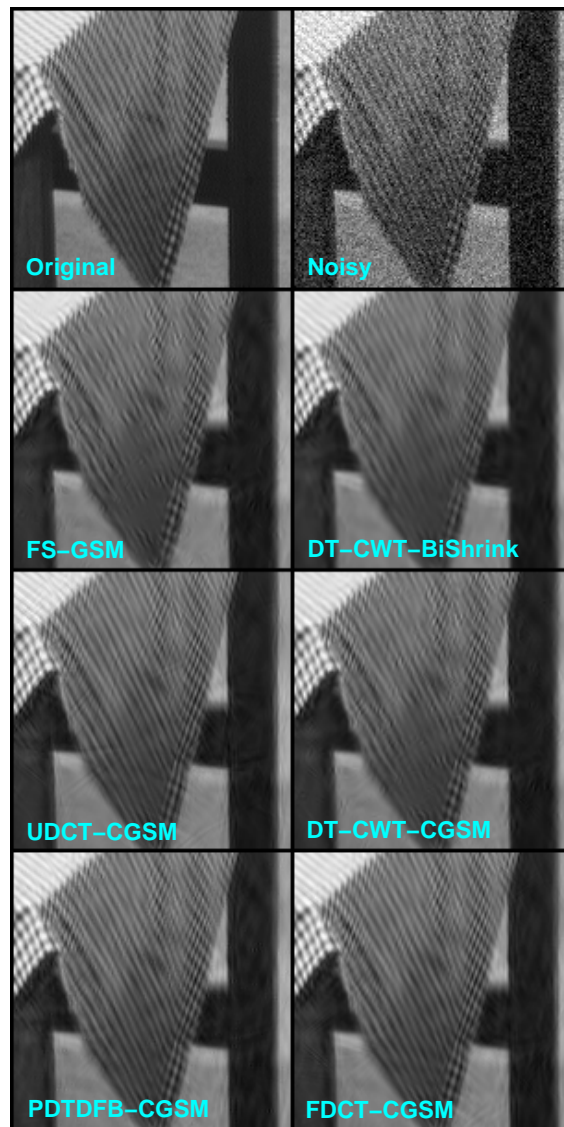
Figure 4.2. Denoising performance comparison of the Barbara image (cropped to 150×150). **Top-left**: original image, **Top-right**: noisy image PSNR 22.11dB ($\sigma = 20$), **Second-left**: FS-GSM PSNR 30.27dB, **Second-right**: DT-ℂWT-Bishrink PSNR 29.66dB, **Third-left**: UDCT-CGSM PSNR 30.43dB, **Third-right**: DT-ℂWT-CGSM 30.25dB, **Bottom-left**: PDTDFB-CGSM PSNR 30.47dB, and **Bottom-right**: FDCT-CGSM PSNR 30.70dB.

4.3 An Application in Texture Retrieval

To show the usefulness of the magnitude pdf of the CGGD as discuss in Chapter 2, the magnitude of the complex coefficients of texture images is modeled by the magnitude pdf for the purpose of texture retrieval.

Table 4.6. Redundancy ratio comparison of the transforms used in the denoising experiment

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

4.3.1 Texture Image Database and Feature Extraction

In the experiments, forty images of size 512×512 from the VisTex database are used as in [70]. Each of the forty images is partitioned into sixteen subimages of size 128×128. Consequently, there are 640 subimages in the database of forty classes where the subimages in the same class belong to the same image of the forty original texture images. Henceforth, each of the 640 subimages is referred to as an image while each of the original forty images is referred to as a class. After normalization to have zero mean and unit variance for all images in the database, each of them is decomposed by the uniform discrete curvelet transform (UDCT) [41]. Here, an image is decomposed into three scales where there are six subbands in each scale. Hence there are eighteen subbands for each image. To extract the feature vector for an image, the coefficients in each subband are fitted with the derived pdf using the parameter estimation method discussed in Section 3.3.2. Consequently, the parameters obtained from all subbands will form the feature vector used in the query process.

4.3.2 Distance Measurement

Once the feature vector of each image is obtained, the distance or similarity between the query image and each image in the database is measured using their feature vectors in order that one can retrieve or select the N nearest images. In this work, the distance between two images is the sum over subbands of the distance between the distributions of the coefficient magnitude in each subband of the two

images. Here, the Kullback-Leibler distance (KLD) is used to measure the distance between two distributions. The KLD between two pdf's $f(x)$ and $g(x)$ is defined by [12]

$$D(f(x)||g(x)) \triangleq \int f(x) \ln \left(\frac{f(x)}{g(x)} \right) \, dx.$$

The KLD $D(f(x)||g(x))$ is nonnegative and is equal to zero when $f(x) \equiv g(x)$. If two pdf's $f_1(r)$ and $f_2(r)$ are described by (2.11), we obtain the closed form expression of the KLD between the two pdf's as

$$
D(f_1(\cdot; \alpha_1, \beta_1)||f_2(\cdot; \alpha_2, \beta_2)) = \ln \left(\frac{\beta_1 \alpha_2^2 \Gamma \left(\frac{2}{\beta_2} \right)}{\beta_2 \alpha_1^2 \Gamma \left(\frac{2}{\beta_1} \right)} \right)
$$
$$
+ \left(\frac{\alpha_1}{\alpha_2} \right)^{\beta_2} \frac{\Gamma \left(\frac{\beta_2+2}{\beta_1} \right)}{\Gamma \left(\frac{2}{\beta_1} \right)} - \frac{2}{\beta_1}. \tag{4.4}
$$

Therefore, the overall distance between two feature vectors F_1 and F_2 extracted from M subbands is given by

$$
d(F_1, F_2) = \sum_{i=1}^{M} D \left(f_R(\cdot; \alpha_1^{(i)}, \beta_1^{(i)}) || f_R(\cdot; \alpha_2^{(i)}, \beta_2^{(i)}) \right).
$$

The retrieval rate of a query image is defined as the number of images among the N nearest images, excluding itself, which belong to the same class divided by fifteen. The average retrieval rate is the retrieval rate of each image averaged over all 640 images in forty classes.

4.3.3   Experimental Results

The retrieval performance is compared when the derived pdf is used in magnitude modeling with the performance of the other three methods. For the first method, two statistics which are the mean and the standard deviation (STD) of the coefficient magnitude in each subband are used as the features. For the second method, the magnitude of the coefficients is modeled by the Rayleigh distribution, where the Rayleigh

pdf parameter is used as the feature for each subband. The third method models each of the real and the imaginary parts of the coefficients by the GGD model [17], where the two parameters of the GGD are the features for each subband. For the first method, the conventional Euclidean distance is used the measure the distance between two feature vectors, while the KLD is used for the latter two.

Table 4.10 shows the average retrieval rates when $N = 15$ with features extracted from six subbands in the finest scale, and from eighteen subbands in all three scales. As can be seen, using the derived pdf to model the magnitude outperforms the other three methods when using one scale and three scales. Using the derived magnitude pdf yields higher retrieval rate than the best of using the GGD for the real part or for the imaginary part by 0.5% and 1.1% when using one scale and three scales, respectively.

Fig. 4.3 displays the average retrieval rate with features extracted from six subbands in the finest scales in Fig. 4.3(a), and from eighteen subbands in all three scales in Fig. 4.3(b). When the number of selected images N is from fifteen to sixty, the proposed method consistently outperforms the other three methods in both cases. Even though the information obtained from the real part or the imaginary part of the coefficients is the same as the information obtained from the magnitude of the coefficients since the real and the imaginary parts are assumed to be identically distributed and uncorrelated, the performance of using the magnitude pdf is better than that of using the pdf of the real part or the imaginary part of coefficients. This is due to the (approximately) shift-invariance property of the complex wavelets, while this is not true for the real wavelets
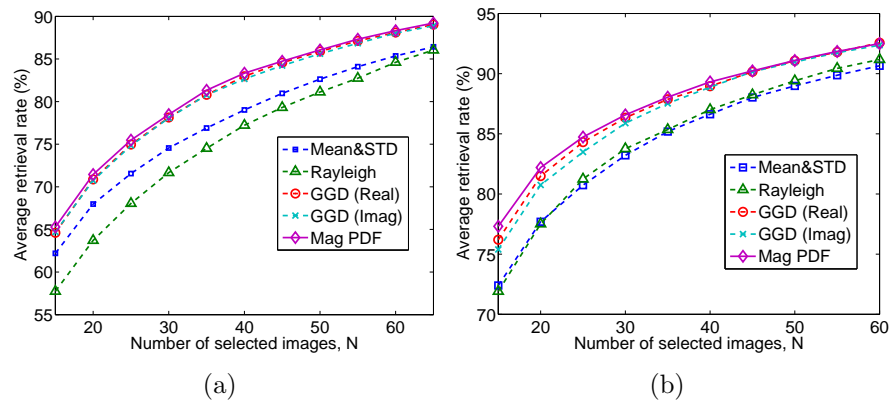
Figure 4.3. Percentage of the average retrieval rate versus the number of selected images N: (a) use only the finest scale, (b) use all three scales.

Table 4.7. Percentage of the average retrieval rate using UDCT and several feature extraction using the finest scale and using all three scales when the number of selected images $N = 15$

1 Scale	Mean&STD	Rayl	GG-Re	GG-Im	Mag PDF
Feat. length	12	6	12	12	12
Avg. rate (%)	62.20	57.74	64.60	64.70	**65.28**
3 Scales	Mean&STD	Rayl	GG-Re	GG-Im	Mag PDF
Feat. length	36	18	36	36	36
Avg. rate (%)	72.40	71.91	76.20	75.39	**77.31**

4.4 Texture Retrieval in Noisy Environment

The content-based image retrieval (CBIR) has been in need in the past years since the rapid growth of digital image data in multimedia database. It is a technique to browse and retrieve images relevant to the user's query image from the database by analyzing the image contents such as shapes and textures. A simple image retrieval system consists of two steps: feature extraction, where a set of features is obtained to represent each image in the database, and similarity measurement, where the distance between the query image and each image in the database is computed in order to retrieve the K most similar images. This section focuses on the texture

image retrieval since the features used in CBIR systems are low-level image features such as texture [17].

One of the most popular techniques for texture retrieval is to use the filter-based or wavelet-based approaches for feature extraction [8, 36, 53, 65]. There have been several research studies using feature extraction in the transform domain. For example, in [36], the mean and standard deviation of the magnitude of Gabor co-efficients are used and the features, and the Euclidean distance is employed in the similarity measurement step. Instead of using the mean and standard deviation of coefficient magnitude and the Euclidean distance, in [17], by modeling the coefficient of the discrete wavelet transform (DWT) as a generalized Gaussian distribution (GGD), the GGD parameters are computed as features while the Kullback-Liebler distance (KLD) between two distributions is used a similarity measure to improve the retrieval accuracy. The texture retrieval scheme we focus on in this work is based on the framework in [17]. Note that the idea of using this similar statistical approach for image retrieval was also introduced in [66]. Because of the effectiveness of using the KLD in similarity measurement and the emersion of the complex wavelets, there has been a number of research work focusing on modeling complex wavelet coefficients and using it in feature extraction for texture retrieval. In [30], together with the computational analysis, the Weibull and Gamma distributions are used to model the magnitude of complex coefficients from the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT) for texture retrieval, which improves the accuracy rate from using the mean and standard deviation features [15]. Moreover, the generalized Gamma distribution is used to model the coefficient for texture retrieval in the case of the DWT (two-sided version of the distribution) in [11], and the coefficient magnitude in the case of the DT-$\mathbb{C}$WT in [33]. The extension of texture retrieval systems to be robust to image rotation has also been studied (see [16, 64] for examples).

The quality of the query images is crucial in the retrieval performance. However, in practice, users may obtain a noisy query image because of the lack of acquisition expertise or an unavoidable noisy environment. Typically, there are two ways to deal with this problem: (1) perform a denoising method first to obtain a denoised query image which features can be extracted or (2) estimate the features of the clean query image from the noisy one. Indeed, there are many promising wavelet-based image denoising methods (see [47] for example). However, performing image denoising before extracting the features from query image might not be computationally efficient since two more steps required in forward and inverse transforms have to be done in order to obtain a denoised query image. Therefore, such a method to deal with noisy texture retrieval will not be discussed here. In this paper, instead of be extracted from a clean query image, feature vectors are extracted from a noisy one by a method of parameter estimation in noise. There are research studies that consider the presence of noise in texture retrieval in the transform domain based on the singular value decomposition (SVD)-based modeling [59], and based on local binary patterns (LBP) [76]. However, these are outside the scope of this work, and will not be discussed herein. There has been no work focusing on using the statistical wavelet approach based on the subband coefficient modeling and the KLD as in [17] for texture retrieval when the query image is contaminated by noise.

In this section, statistical texture retrieval in a noisy environment in the complex wavelet domain is performed by using the parameter estimation methods from noisy coefficients as proposed in Section 3.4. The query image in this scenario is contaminated by additive white Gaussian noise. The proposed methods are used to estimate the pdf's parameters of the original (clean) image from the noisy one, which are then used in computing the retrieval rate.

4.4.1   Texture Image Database and Feature Extraction

In this application, the images from the same database as in Subsection 4.3.1 are used. After normalization to have zero mean and unit variance for all images in the database, white Gaussian noise with a known variance is added. Subsequently, each of them is decomposed by the UDCT into three scales where there are six subbands in each scale. The features of each image in the database (which is clean) are obtained by using the parameter estimation methods described in Section 3.2.

4.4.2   Feature Extraction from the Noisy Query Image

To extract the feature vector for an image from subbands of noisy coefficients, the parameter estimation methods discussed in Section 3.4 are used. In the cases of estimating the parameter β of the GGD, Weibul and Gamma distributions, a look-up table consists of 100 uniformly-spaced points starting the minimum value to the maximum values of the estimates of that parameters obtained from the clean database is utilized. In the case of RP model, the log-normal prior with unit mean is chosen to model the hidden multiplier S in (3.21) which is empirically found to produce good results, i.e.,

$$f_S(s; \mu_{\mathrm{L}}, \sigma_{\mathrm{L}}) = \frac{1}{s\sigma_{\mathrm{L}}\sqrt{2\pi}} \exp\left(-\frac{(\ln s - \mu_{\mathrm{L}})^2}{2\sigma_{\mathrm{L}}^2}\right),$$

where $\mu_{\mathrm{L}} = -\sigma_{\mathrm{L}}^2/2$. To estimate the parameters of $f_S(s)$ from the noisy coefficients, the moment matching method is used, and is given by

$$\begin{aligned}
\hat{\mu}_{\mathrm{L}} &= -\frac{1}{2}\ln(E[S^2]), \quad \hat{\sigma}_{\mathrm{L}}^2 = \ln(E[S^2]), \\
E[S^2] &= \frac{E[|Y|^4] - 4\sigma_Y^2\sigma_W^2 + 2\sigma_W^4}{2(\sigma_Y^2 - \sigma_W^2)^2}.
\end{aligned} \tag{4.5}$$

The expression of $E[S^2]$ in (4.5) can be computed from $Y = X + W = \sqrt{S}\widetilde{X} + W$, where Y, X, $\widetilde{X}$ and W are the univariate version of $\mathbf{Y}$, $\mathbf{X}$, $\widetilde{\mathbf{X}}$ and $\mathbf{W}$, respectively, and σ_Y^2 and σ_W^2 are the variances of Y and W, respectively. The number of points

Table 4.8. The average retrieval rates (%) of several methods using the UDCT when the number of selected images $K = 16$

σ_n	Scales	Assume noisy query image is clean					
		Mean&SD	GG-Re	GG-Im	Rl	Wbl	Gam
0.1	1	57.94	54.54	55.15	53.02	55.94	53.76
	$1, 2$	70.97	68.36	68.39	66.20	69.77	67.70
	$1, 2, 3$	74.08	72.69	71.70	71.13	74.74	72.96
0.2	1	40.55	33.16	34.20	33.35	34.41	32.09
	$1, 2$	63.36	48.67	49.43	48.86	50.34	47.21
	$1, 2, 3$	70.04	56.75	56.50	56.95	58.22	55.47
0.3	1	23.32	18.07	19.79	18.70	19.25	17.50
	$1, 2$	47.50	34.26	35.27	33.85	35.63	32.93
	$1, 2, 3$	61.27	43.26	44.08	42.05	45.28	42.41
Number of features (3 scales)		36			18	36	

S of the interval $[s_\mathrm{min}, s_\mathrm{max}]$ is set as $N_s = 15$. The interval $[s_\mathrm{min}, s_\mathrm{max}]$ are chosen such that $F_S(s_\mathrm{min}) = 0.005$, and $F_s(s_\mathrm{max}) = 0.995$, where $F_S(s)$ is the cumulative density function (cdf) which is defined by $F_S(s) = \int_{-\infty}^{s} f_S(t)dt$. This is to assure that the range of integration sufficiently covers the support of $f_S(s)$. Once $\hat{\mu}_\mathrm{L}$ and $\hat{\sigma}_\mathrm{L}$ are obtained, the limits of the interval $[s_\mathrm{min}, s_\mathrm{max}]$ can be obtained by $s_\mathrm{min} = \exp(\hat{\mu}_\mathrm{L} - 2.80703\hat{\sigma}_\mathrm{L})$ and $s_\mathrm{max} = \exp(\hat{\mu}_\mathrm{L} + 2.80703\hat{\sigma}_\mathrm{L})$.

4.4.3 Distance Measurement

Once the feature vector of each image is obtained, the distance or similarity between the query image and each image in the database is measured using their feature vectors so that one can retrieve or select the K nearest images. The distance between two images in this thesis is the sum over subbands of the distance between the distributions corresponding the the type of features in each subband of the two images. Here, the symmetric version of the Kullback-Leibler distance (KLD), D_sym,

Table 4.9. The average retrieval rates (%) of several methods using the UDCT when the number of selected images $K = 16$

σ_{n}	Scales	Estimate clean parameters from noisy query image					
		L2&L4	GG-Re	GG-Im	Rl	Wbl	Gam
0.1	1	59.10	62.44	63.77	60.60	54.01	55.45
	$1, 2$	68.78	74.51	74.49	71.26	69.34	71.21
	$1, 2, 3$	71.60	77.09	75.86	74.59	76.39	77.08
0.2	1	58.74	60.64	61.79	60.23	53.47	54.59
	$1, 2$	68.59	73.00	73.43	71.03	68.42	70.95
	$1, 2, 3$	71.48	75.54	74.85	74.70	75.52	76.71
0.3	1	58.41	59.90	59.57	59.17	52.09	53.42
	$1, 2$	68.44	72.57	72.50	70.02	67.53	70.01
	$1, 2, 3$	71.38	75.46	74.34	73.83	74.69	76.01
Number of features (3 scales)		36			18	36	

is used to measure the distance between two distributions. The symmetric KLD is defined by

$$D_{\mathrm{sym}}(f, g) \triangleq \frac{1}{2} \left(D(f\|g) + D(g\|f) \right),$$

where $D(f\|g)$ is the KLD of two pdf's f and g which is defined by [12]

$$D(f\|g) \triangleq \int f \ln \left(\frac{f}{g} \right) \tag{4.6}$$

Therefore, the overall distance between two feature vectors F_1 and F_2 extracted from N_b subbands is given by

$$d(F_1, F_2) = \sum_{i=1}^{N_b} D_{\mathrm{sym}} \left(f_1^{(i)}, f_2^{(i)} \right).$$

Since a closed form of the KLD between two RP pdf's is under investigation, a numerical method with 128 bins is used to estimate (4.6) in this paper. The retrieval rate of a query image is defined as the number of images among the K nearest images, which belong to the same class divided by sixteen. The average retrieval rate is the retrieval rate of each image averaged over all 640 images in forty classes.
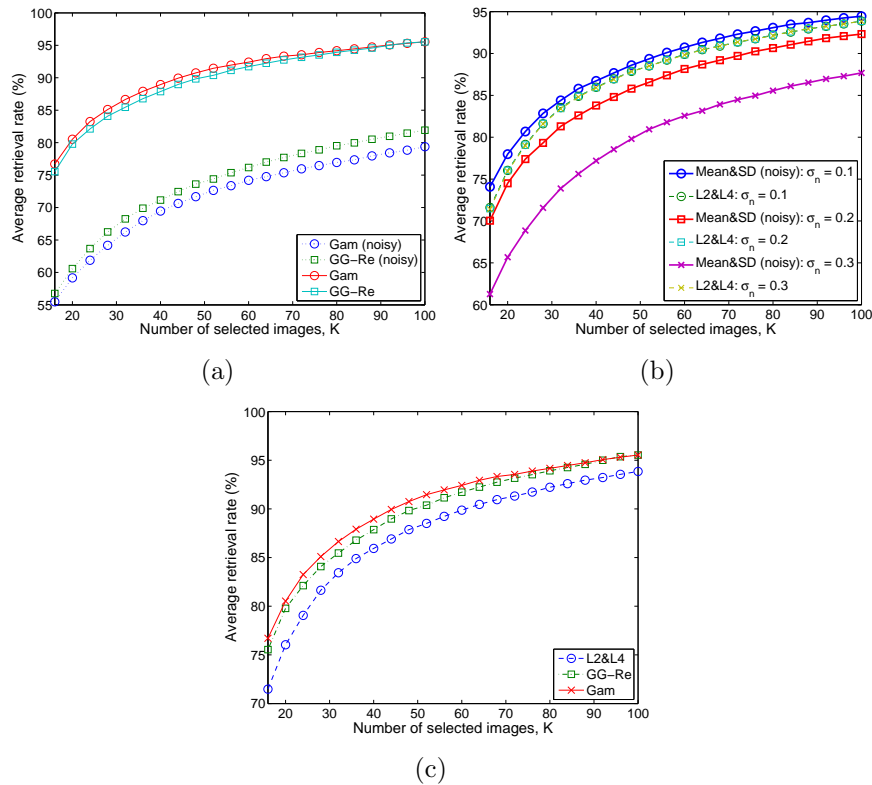
Figure 4.4. Comparison of the average retrieval rate in several aspects: (a) awareness of noisy query ($\sigma_n = 0.2$), (b) Mean&SD (noisy) vs. L2&L4, and (c) Nonmodel-based vs. model-based ($\sigma_n = 0.2$).

4.4.4 Experimental Results

4.4.4.1 Using Magnitude Only

In this thesis, the noisy texture retrieval experiments are tested with three values of the noise level $\sigma_n = 0.1, 0.2, 0.3$. Tables 4.8 and 4.9 show the average retrieval rates when $K = 16$ with features extracted from subbands in three scales of the UDCT decomposition using five methods according to the features extracted, which are the GG parameters of the real (GG-Re) and the imaginary parts (GG-Im) of the complex coefficient, and the Rayleigh parameter (Rl), the Weibull parameters (Wbl) the Gamma parameters (Gam) of the magnitude. The results for these methods are obtained in two situations: (1) the noisy query image is assumed to be clean (in
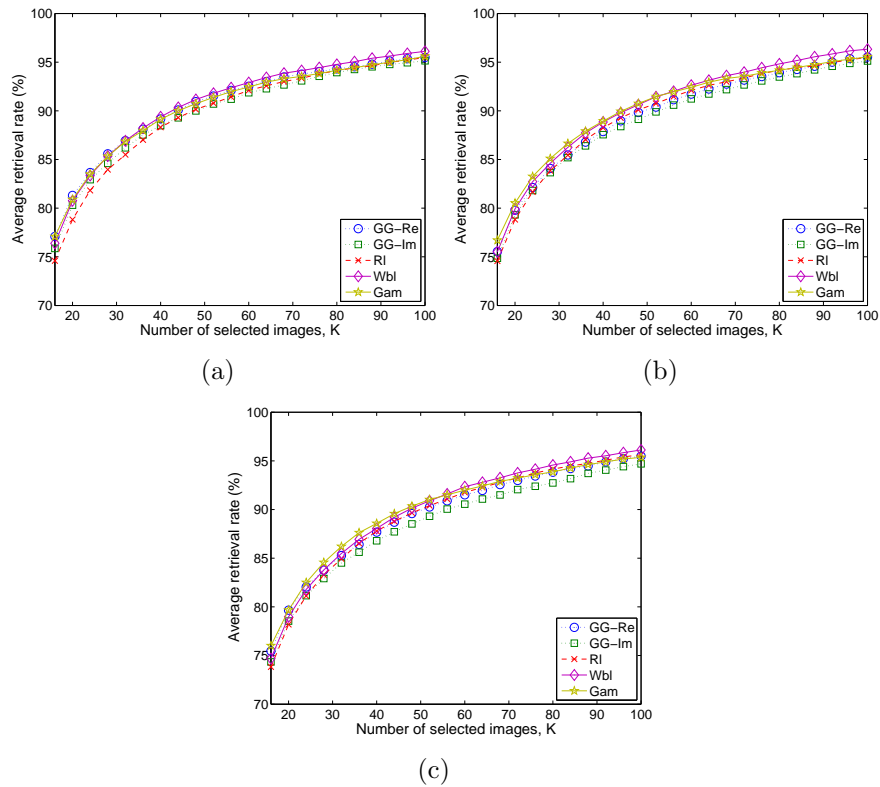
Figure 4.5. The average retrieval rate for several model-based methods: (a) $\sigma_n = 0.1$, (b) $\sigma_n = 0.2$, and (c) $\sigma_n = 0.3$.

Table 4.8), and (2) the presence of noise is taken into consideration and the clean parameters from the noisy coefficient (in Table 4.9) are estimated as described in Sections 3.4. As can be seen from Table 4.8, ignoring the presence of noise degrades the retrieval performance drastically in all of the six methods when $K = 16$. This is also consistent for all values of K from 16 to 100 as illustrated in Fig. 4.4(a) for the Gam and GG-Re methods with three scales as examples. In all plots in Fig.'s 4.4 and 4.5, features from all three scales are used.

Moreover, using the mean and standard deviation of the magnitude of subband coefficient (Mean&SD) as features is also compared. This is probably the simplest and widely-used nonmodel-based method for texture retrieval in the case where it is assumed that the noisy query is clean. Note that the performance of this Mean&SD

method is even better than the performance of the other five model-based methods when noise levels are 0.2 and 0.3 in the situation where the noisy query image is assumed to be clean, which implies that using the mean and standard deviation is more resilient to noise compared to the model-based approaches given that any estimation method of clean parameters is not performed.

When the presence of noise is considered, the estimates of the L_2-norm and the L_4-norm (L2&L4) are used as an example of a nonmodel-based method since one cannot estimate the mean of the magnitude of clean coefficient from the noisy coefficient The L_2-norm and L_4-norm are estimated by

$$L_2 = \sqrt{m_2}, \quad \text{and} \quad L_4 = m_4^{1/4},$$

where m_2 and m_4 are the estimates of the second and the third moment of the coefficient magnitude as described in Section 3.4. For these two nonmodel-based methods, the similarity measure between two images is the Euclidean distance as defined in [36]. The retrieval rates of the Mean&SD and L2&L4 methods are shown in Fig. 4.4(b) as a comparison of the nonmodel-based methods in both situations. For all three levels of noise except when $\sigma_n = 0.1$, the L2&L4 method when one is aware of noise performs consistently better than we blindly use the Mean&SD method. From now on, all the methods are in the situation that the presence of noise is taken into consideration.

Next, the nonmodel-based and the model-based methods are compared. Fig. 4.4(c) displays the accuracy rates of the L2&L4, GG-Re and Gam methods. Apparently, the model-based methods outperform the nonmodel-based methods (with the same number of features which is 36) as one expects. Fig. 4.5 compares the retrieval rates of the five model-based methods. It can be seen that both Wbl and Gam methods which use the coefficient magnitude performs better (or approximately the same

Table 4.10. The average retrieval rates (%) using several complex transforms when the number of selected images $K = 16$

σ_n	Scales	UDCT			DT-$\mathbb{C}$WT			PDTDFB		
		Noisy	RP	RPM	Noisy	RP	RPM	Noisy	RP	RPM
0.1	1	36.04	51.49	57.25	35.11	55.77	60.30	44.44	54.69	58.33
	1, 2	44.80	57.86	61.07	52.37	69.52	71.36	51.67	60.16	62.56
	1, 2, 3	48.29	56.45	57.25	61.29	69.56	71.32	48.28	53.41	55.11
0.2	1	17.52	41.51	56.15	16.71	44.13	54.09	26.69	45.21	52.06
	1, 2	27.20	47.81	58.15	33.02	57.56	67.73	36.48	52.49	59.04
	1, 2, 3	35.51	50.51	56.05	44.03	62.61	69.24	38.61	49.30	52.88
0.3	1	9.00	35.90	46.74	8.54	37.66	49.75	16.33	39.14	48.75
	1, 2	20.01	43.39	54.73	21.87	49.85	64.95	28.53	45.81	56.04
	1, 2, 3	28.93	45.98	54.78	33.28	55.86	68.18	33.25	44.00	51.34

when $\sigma_n = 0.1$) than the GG-Re and GG-Im methods which use the real and the imaginary parts of the coefficient respectively. This is because of the shift-invariance property of the magnitude of complex coefficient. The Rl method which models the magnitude by the Rayleigh distribution, which is a one-parameter distribution, is outperformed when $\sigma_n = 0.1$ by other methods even by the GG-Re and GG-Im methods, which use the real/imaginary parts since the Rl method uses 18 features while the others use 36 features. However, when the noise level increases, the Rl method seems to be less sensitive since it involves with only estimation of the second moment while the other four methods also involve with the estimation of the fourth moment.

4.4.4.2 Using Phase Only

Table 4.10 shows the average retrieval rates when $K = 16$ with features extracted from the relative phase in coefficient subbands in three scales of the UDCT decomposition. As can be seen, using the RPM method in estimating the RP parameters outperforms the Noisy and RP methods for all scales and all three values of noise level. Using the first two scales when $\sigma_n = 0.1$, the RPM method improves the

Figure 4.6. The average retrieval rate versus the number of selected images K using the UDCT: (a) $\sigma_n = 0.1$, (b) $\sigma_n = 0.2$, and (c) $\sigma_n = 0.3$.

accuracy rate from the Noisy method, where the noisy coefficients are blindly used as clean coefficients, by 16.27% while the RP method improves the rate from the Noisy method by 13.06%. Fig.'s 4.6(a)-4.6(c) display the average retrieval rate with features extracted from twelve subbands in the first two scales versus the number of selected images K when the noise level σ_n is from 0.1 to 0.3. When the number of selected images K is from 16 to 100, the proposed RPM method consistently outperforms the other two methods in all three noise levels.

In addition to the UDCT, texture retrieval in a noisy environment is also performed with other two complex transforms: the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT), and the pyramidal dual-tree directional filter bank (PDTDFB). For both transforms, an image is decomposed into three scales, with six directions in each scale
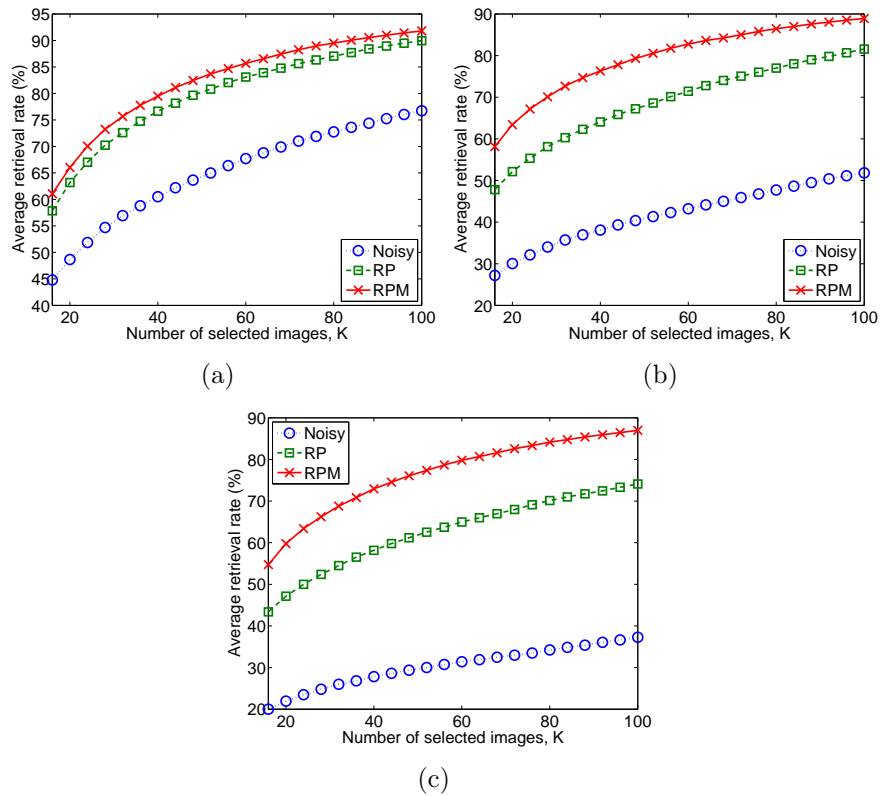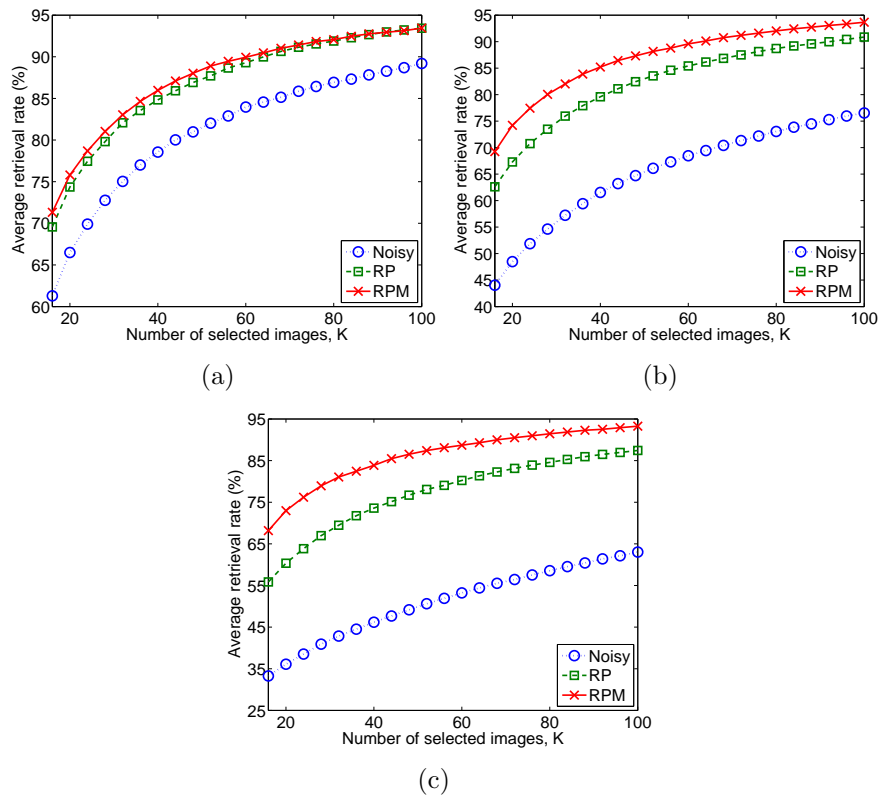
Figure 4.7. The average retrieval rate versus the number of selected images K using the DT-ℂWT: (a) $\sigma_n = 0.1$, (b) $\sigma_n = 0.2$, and (c) $\sigma_n = 0.3$.

in the case of the DT-ℂWT, and with eight directions in each scale in the case of the PDTDFB. The results for the two complex transforms when $K = 16$ are also displayed in Table 4.10. The RPM method still outperforms the other two in estimating the RP parameters as in the case of the UDCT for all scales, and all three noise levels. Fig.'s 4.7 and 4.8 illustrate the average retrieval rate versus the number of selected images with three noise levels for the DT-ℂWT and PDTDFB, respectively. For the DT-ℂWT, three scales are used while only the first two scales are used for the PDTDFB, which produce the highest retrieval rate for each transform. Like in the case of the UDCT, the parameters obtained by the proposed RPM methods yield higher retrieval rates than those from the other two methods.
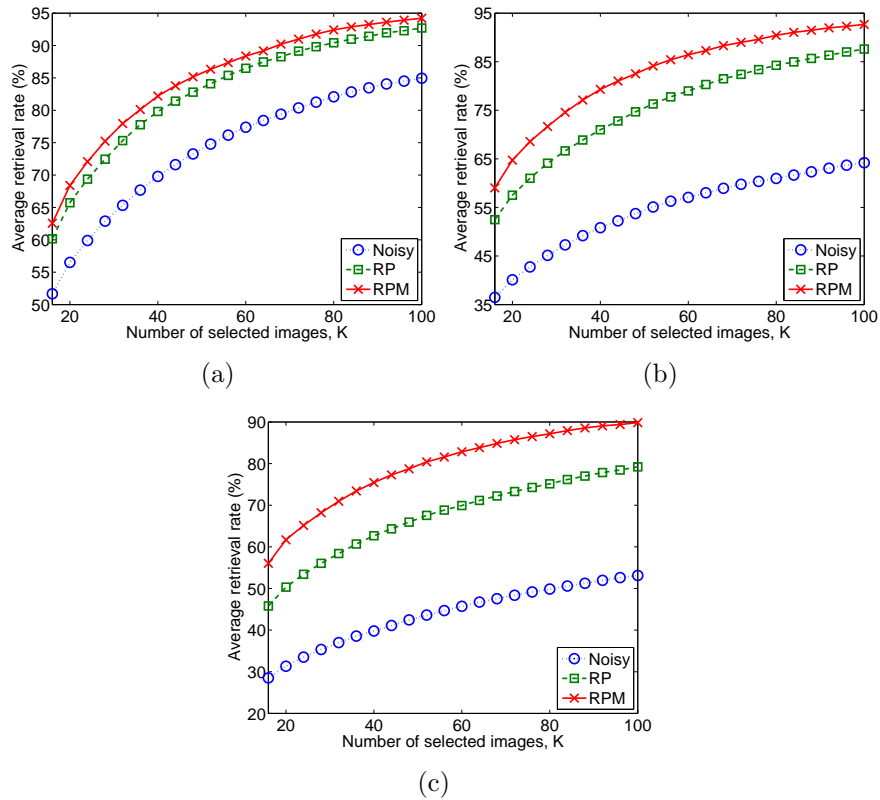
Figure 4.8. The average retrieval rate versus the number of selected images K using the PDTDFB: (a) $\sigma_{\mathrm{n}} = 0.1$, (b) $\sigma_{\mathrm{n}} = 0.2$, and (c) $\sigma_{\mathrm{n}} = 0.3$.

4.4.4.3   Using Both Magnitude and Phase

Furthermore, the phase information in the form of the relative phase is also incorporated to the magnitude information to improve the accuracy rate. Features from the coefficient magnitude in three scales are used while the phase information from two scales is used, which yields the best result. To incorporate the phase information to the magnitude information, one needs to define the combined distance measurement where in this paper we use

$$d(F_1, F_2) = a d_{\mathrm{MG}}(F_1, F_2) + (1 - a) d_{\mathrm{RP}}(F_1, F_2),$$

where $d_{\mathrm{MG}}(F_1, F_2)$ and $d_{\mathrm{RP}}(F_1, F_2)$ are the distance between F_1 and F_2 according the magnitude and the relative phase respectively, $0 \leq a \leq 1$ is the weight which

Table 4.11. The average retrieval rates (%) using magnitude and phase information when the number of selected images $K = 16$

σ_{n}	Magnitude (Scales 1,2,3)		Phase (Scales 1,2)	Magnitude&Phase	
	Wbl	Gam	RP	Wbl&RP	Gam&RP
0.1	76.39	77.08	61.07	80.78	80.23
0.2	75.52	76.71	58.15	79.37	78.93
0.3	74.69	76.01	54.73	77.55	77.64

Table 4.12. The average retrieval rates (%) using several transforms when the number of selected images $K = 16$

σ_{n}	UDCT		DTCWT		PDTDFB		DWT
	Wbl&RP	Gam&RP	Wbl&RP	Gam&RP	Wbl&RP	Gam&RP	GG
0.1	80.78	80.23	81.56	81.26	82.06	82.07	73.48
0.2	79.37	78.93	76.66	75.76	80.64	80.97	70.27
0.3	77.55	77.64	71.83	71.18	80.08	79.76	65.46
Number of features	60	60	60	60	80	80	18

is adjusted such that the combined distance yields the best results when there is no noise.

As can be seen from Table 4.11, incorporating the phase information improves the retrieval rate when $K = 16$ by 2.86% to 4.39% for the Wbl & RP method, and by 1.63% to 3.15% for the Gam & RP method. Fig. 4.9 shows that using both magnitude and phase information consistently improves the retrieval rate from using either the magnitude of the phase information independently, and also better than the best of the real or imaginary parts.

In addition to the UDCT, texture retrieval in a noisy environment is performed with other two complex transforms: the dual-tree complex wavelet transform (DT-CWT), and the pyramidal dual-tree directional filter bank (PDTDFB), and also the conventional DWT, in which we use the GG parameters as the features. For all
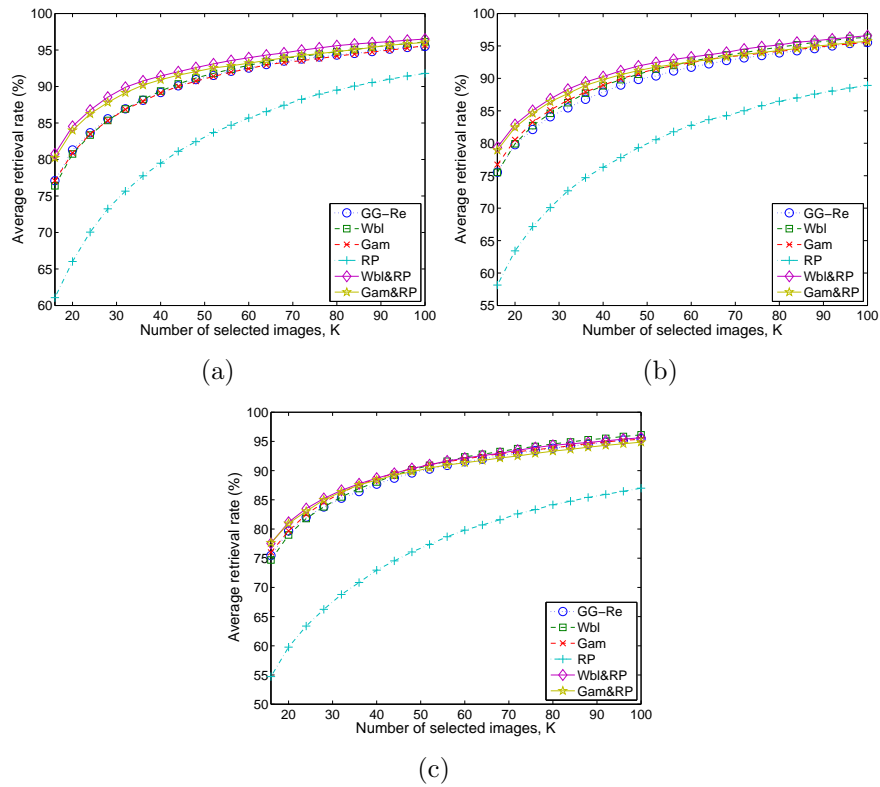
Figure 4.9. Comparison of the average retrieval rate using magnitude-phase information: (a) $\sigma_n = 0.1$, (b) $\sigma_n = 0.2$, and (c) $\sigma_n = 0.3$.

transforms, an image is decomposed into three scales, with six directions in each scale in the case of the DT-$\mathbb{C}$WT, and with eight directions in each scale in the case of the PDTDFB, and three directions each scales for the case of the DWT.

The results for the three transforms when $K = 16$ are also displayed in Table 4.12. Fig.'s 4.10 illustrates the retrieval rate with three noise levels for all transforms. Apparently, using complex-valued transforms yields better results for noisy texture retrieval than using the real-valued DWT, which confirm the advantage of the complex-valued wavelets to the conventional real-valued wavelets. Comparing all three complex transforms, it can be seen that the PDTDFB performs better than the other two transforms because of high directional selectivity. In terms of computational complexity, the average values of computational time for all transforms are
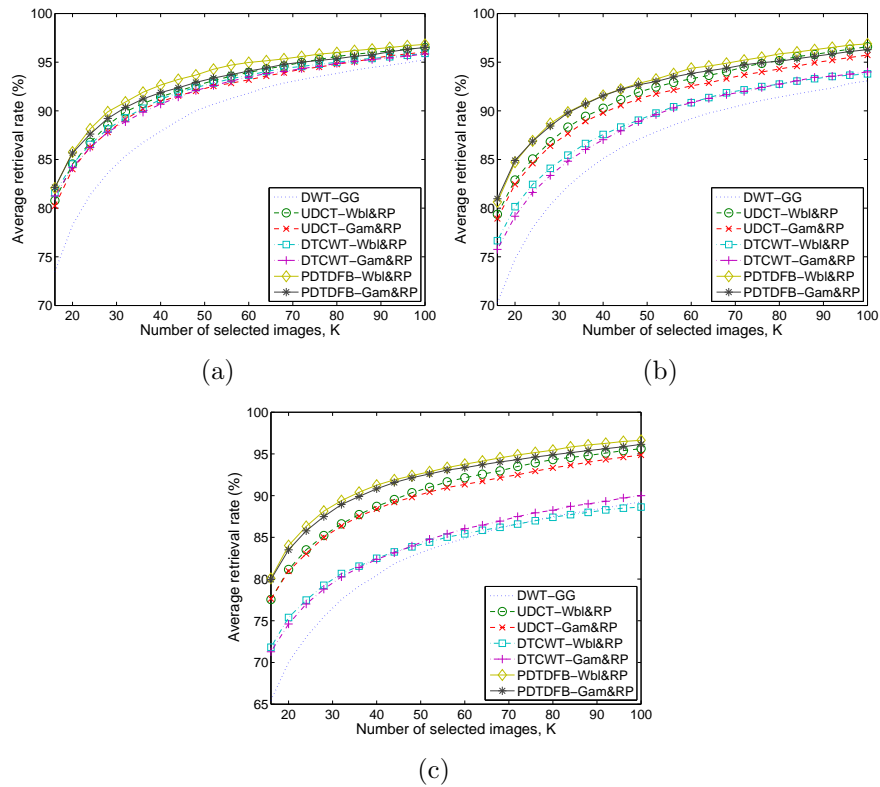
Figure 4.10. The average retrieval rate using several transforms: (a) $\sigma_n = 0.1$, (b) $\sigma_n = 0.2$, and (c) $\sigma_n = 0.3$.

tabulated in Table 4.13, where we show the time of the Wbl&RP method for the complex transforms, and the GG method for the DWT. As can be expected, it requires much more computational time in incorporating the phase information to improve the retrieval rate in both feature extraction and similarity measurement since the RP method requires an iterative method in both steps. Moreover, the PDTDFB is more computationally expensive than the UDCT and the DT-$\mathbb{C}$WT while the DWT requires least computational time.

4.5 Summary

To illustrate the usefulness of the CGSM, the proposed model is applied to three applications: image denoising, texture retrieval and texture retrieval in a noisy

Table 4.13. Computational time (Matlab on intel Core i7 with 2.8 GHz and 8GB RAM) and number of features for various transforms

Details	UDCT		DTCWT	PDTDFB	DWT
Methods	Wbl	Wbl&RP	Wbl&RP	Wbl&RP	GG
Num. of feat.	36	60	60	80	18
FE time (s)	0.017	0.855	0.922	1.117	0.008
SM time (ms)	0.3	5.0	5.1	6.7	0.4

environment. In image denoising, the proposed CGSM has been applied to obtained a denoised image from a noisy one using Bayes least squares (BLS) estimator. The denoising results indicate that the CGSM is possibly a more appropriate model for the complex wavelet coefficients than the joint GSM model of their real and imaginary parts. Moreover, the CGSM allows us to improve the denoising results from those results using the GSM with the real transforms or obtain comparable results by using a complex transform with less redundancy ratio. In addition to image denoising, the application of complex wavelet-based texture retrieval using the derived magnitude pdf of the CGGD is performed. From the experimental results, using the derived magnitude pdf for feature extraction improves from the other three methods which use the mean and the standard deviation, the Rayleigh model, and the GGD model of the real part and the imaginary part.

Furthermore, the problem of texture retrieval in noisy environment where the query image is noisy using complex wavelets in a statistical framework has also been studied. Specifically, the features are extracted from the noisy query image by assuming a statistical model on the complex subband coefficients of the clean image. This feature extraction based on the parameter estimation in noise of the model of the complex coefficients. The retrieval rate is then computed based on the similarity measurement which is performed by computing the Kullback-Leibler distance (KLD)

between the estimated model from the noisy query image and images in the database. The awareness of the presence of noise is validated by the results that the retrieval rate with parameter estimation in noise is higher that the rate when the noisy query image is assumed to be clean. In addition to magnitude information, phase information is also incorporated to improve the retrieval rate. The experimental results show that using magnitude and phase improves the accuracy rate from using either magnitude or phase alone, and that using complex-valued wavelets yield better results than using real-valued wavelets.

The simulation results in both applications are consistent among several complex multiresolution transforms including the dual-tree complex wavelet transform (DT-$\mathbb{C}$WT).

CHAPTER 5

CONCLUDING REMARKS AND FUTURE DIRECTIONS

5.1   Concluding Remarks

In this thesis, the complex Gaussian scale mixture (CGSM) model for complex wavelet coefficients has been proposed. The CGSM, which is the complex version of the Gaussian scale mixture (GSM), can be used to model and treat the complex coefficients appropriately. It has also been shown that the real and imaginary parts of complex wavelet coefficients are related in such a way that their joint pdf can be expressed as a function of complex coefficients. Moreover, some related propositions and results of the CGSM are discussed. As consequences from modeling the complex coefficient by the CGSM, the general forms of magnitude and the phase pdf's have been derived. Since the general form of the magnitude pdf has no closed form, several widely-used magnitude models which are its special cases are discussed. Unlike the case of magnitude, the phase pdf in general case has the closed form, which is called the relative phase pdf (RP pdf). The parameter estimation methods in the presence of noise of the related magnitude and RP pdf's have also been presented. For the case of the RP pdf, a parameter estimation method with requires no iterative algorithm is also presented. The simulation results show that the non-iterative estimator gives acceptably accurate estimates with less complexity, and that the proposed estimation method in noise is satisfactory for the case of simulated data.

To show the effectiveness of the CGSM, the proposed model is used in three applications: image denoising, texture retrieval, and noisy texture retrieval. The proposed CGSM is applied to an application in image denoising using Bayes least

squares (BLS) estimator. The denoising results indicate that the CGSM is an appropriate model for the complex wavelet coefficients. Moreover, the CGSM allows the improvement of the denoising results from those results using the GSM with the real transforms or comparable results by using a complex transform with less redundancy ratio.

Furthermore, the application complex wavelet-based texture retrieval using the derived magnitude pdf of the CGGD, a special case of the CGSM, is performed. From the experimental results, using the derived magnitude pdf for feature extraction improves from the other three methods which use the mean and the standard deviation, the Rayleigh model, and the GGD model of the real part and the imaginary part.

In addition, the application of noisy texture retrieval in a statistical framework has been studied. In this framework, the features are extracted from the image based on statistical modeling in the complex wavelet domain, where the magnitude and relative phase models which are derived from the CGSM are employed. The retrieval rate is then computed based on the similarity measurement which is performed by computing the Kullback-Leibler distance (KLD) between the estimated model from the noisy query image and images in the database. According to the results, the method using the magnitude pdf of coefficient magnitude outperforms the method using the pdf of either the real or imaginary parts of a coefficient. In addition to magnitude information, phase information has been incorporated to improve the retrieval rate. The experimental results show that using magnitude and phase improves the accuracy rate from using either magnitude or phase alone, and that using complex-valued wavelets yield better results than using real-valued wavelets.

5.2   Future Research Directions

One future research direction is the use of the proposed CGSM model in other image processing tasks including image deconvolution in the complex wavelet domain. According to promising image denoising results, applying the CGSM as a sparsity constraint in image deconvolution problem will allow a lower complexity method with better or comparable quality of the reconstructed images.

A second research direction is the computational complexity reduction in noisy texture retrieval using the proposed framework, where an iterative method to estimate parameters of the RP pdf is required. Using a non-iterative parameter estimation method will improve the speed of the algorithm, which results in a more practical texture retrieval system.

Another direction is to utilize the magnitude and phase information of complex wavelets in other applications. Because most of research studies in image processing based on complex wavelets use only the magnitude information, incorporating the phase information could improve the performance, and also provides another viewpoint to solve related image processing tasks.

APPENDIX A

ABBREVIATION LIST

2-D	Two-Dimensional.
AML	Approximate Maximum Likelihood.
AMM	Approximate Moment Method.
AWGN	Additive White Gaussian Noise.
BiShrink	Bivariate Shrinkage.
BLS	Bayes Least Squares.
BLS-GSM	Bayes Least Squares estimator - Gaussian Scale Mixture Model.
CGGD	Complex Generalized Gaussian Distribution.
CGSM	Complex Gaussian Scale Mixture.
DT-$\mathbb{C}$WT	Dual-Tree Complex Wavelet Transform.
DWT	Discrete Wavelet Transform.
FDCT	Fast Discrete Curvelet Transform.
FS	Full Steerable Pyramid.
GGD	Generalized Gaussian Distribution.
GSM	Gaussian Scale Mixture.
GSMag	Gaussian Scale Mixture of the Augmented Vector.
GSMsp	Gaussian Scale Mixture of the Vectors of the Real and Imaginary Parts Separately.
HMT	Hidden Markov Tree.
HT	Hard Thresholding.
KLD	Kullback-Leibler Distance.
ML	Maximum Likelihood.
MLE	Maximum Likelihood Estimator.
PDF	Probability Density Function.
PDTDFB	Pyramidal Dual-Tree Directional Filter Bank.

PSNR	Peak Signal-to-Noise Ratio.
RMSE	Root Mean Squared Error.
RP	Relative Phase.
RPM	Relative Phase Mixture.
RPM PDF	Relative Phase Mixture Probability Density Function.
RP PDF	Relative Phase Probability Density Function.
RSM	Rayleigh Scale Mixture.
SD	Standard Deviation.
SIRP	Spherically-Invariant Random Process.
SNR	Signal-to-Noise Ratio.
SSIM	Structural SIMilarity index.
SVD	Singular Value Decomposition.
UDCT	Uniform Discrete Curvelet Transform.

APPENDIX B

PROOF OF PROPOSITION 3.5.1

Let $\mathbf{X}_i = \sqrt{S_i}\widetilde{\mathbf{X}}_i$ be a CGSM with $E[S_i] = 1$, for $i = 1, 2$, and $\mathbf{X}_3 = \mathbf{X}_1 + \mathbf{X}_2$, where $[S_1, \widetilde{\mathbf{X}}_1]$ and $[S_2, \widetilde{\mathbf{X}}_2]$ are independent. Therefore, $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent.

($\Leftarrow$) If $\mathbf{C}_{\mathbf{X}_2} = a\mathbf{C}_{\mathbf{X}_1}$ for some constant $a > 0$, then the characteristic function of $\mathbf{X}_3$ is given by

$$\Phi_{\mathbf{X}_3}(\boldsymbol{\omega}) = \Phi_{\mathbf{X}_1}(\boldsymbol{\omega})\Phi_{\mathbf{X}_2}(\boldsymbol{\omega})$$
$$= \int_0^\infty \int_0^\infty e^{-\frac{\boldsymbol{\omega}^H((s_1+as_2)\mathbf{C}_{\mathbf{X}_1})\boldsymbol{\omega}}{4}} f_{S_1}(s_1)f_{S_2}(s_2)ds_1 ds_2,$$

where the characteristic function of $\mathbf{X}_i, i = 1, 2$, are given by [52]

$$\Phi_{\mathbf{X}_i}(\boldsymbol{\omega}) = \int_0^\infty \exp\left(-\frac{s_i\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_i}\boldsymbol{\omega}}{4}\right) f_{S_i}(s_i)ds_i.$$

Let $S_3 = \frac{S_1+aS_2}{1+a}$. Since S_1 and S_2 are independent, it can be shown that $\Phi_{\mathbf{X}_3}(\boldsymbol{\omega}) = \int_0^\infty e^{-(\boldsymbol{\omega}^H(s_3\mathbf{C}_{\mathbf{X}_3})\boldsymbol{\omega})/4} f_{S_3}(s_3)ds_3$, where $E[S_3] = 1$ and $\mathbf{C}_{\mathbf{X}_3} = (1+a)\mathbf{C}_{\mathbf{X}_1}$. Hence, we can represent $\mathbf{X}_3 = \sqrt{S_3}\widetilde{\mathbf{X}}_3$, where $\widetilde{\mathbf{X}}_3$ is complex Gaussian with the complex covariance matrix $\mathbf{C}_{\mathbf{X}_3}$, and is independent from S_3, i.e. $\mathbf{X}_3$ is a CGSM. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are both complex Gaussian, then $\mathbf{X}_3 = \mathbf{X}_1 + \mathbf{X}_2$ is also complex Gaussian, and thus a CGSM.

($\Rightarrow$) Suppose $\mathbf{X}_3 = \mathbf{X}_1 + \mathbf{X}_2$ is a CGSM. We can express $\mathbf{X}_3 = \sqrt{S_3}\widetilde{\mathbf{X}}_3$, where $\widetilde{\mathbf{X}}_3$ is complex Gaussian and S_3 is a unit-mean positive random variable independent from $\widetilde{\mathbf{X}}_3$. Since $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, $\mathbf{C}_{\mathbf{X}_3} = \mathbf{C}_{\mathbf{X}_1} + \mathbf{C}_{\mathbf{X}_2}$. Since $\mathbf{X}_i$ are CGSM's, write $\Phi_{\mathbf{X}_i}(\boldsymbol{\omega}) = g_{\mathbf{X}_i}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_i}\boldsymbol{\omega})$. Therefore,

$$g_{\mathbf{X}_3}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_3}\boldsymbol{\omega}) = g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})g_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega}). \tag{B.1}$$

Dividing the complex gradient $\nabla_{\boldsymbol{\omega}}$ of (B.1) by (B.1), we have

$$\frac{g'_{\mathbf{X}_3}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_3}\boldsymbol{\omega})}{g_{\mathbf{X}_3}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_3}\boldsymbol{\omega})}\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_3} = \frac{g'_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}{g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}$$
$$+ \frac{g'_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}{g_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}. \tag{B.2}$$

It follows from (B.2) that, for any complex vector $\mathbf{v}$ which is $\mathbf{C}_{\mathbf{X}_3}$−orthogonal to $\boldsymbol{\omega}$, i.e. $\mathbf{v}^H \mathbf{C}_{\mathbf{X}_3}\boldsymbol{\omega} = 0$, we have

$$\mathbf{v}^H \left(\frac{g'_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}{g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_1} + \frac{g'_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}{g_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_2}\right)\boldsymbol{\omega} = 0.$$

Since this is true for all $\boldsymbol{\omega}$ and for all $\mathbf{v}$ which is $\mathbf{C}_{\mathbf{X}_3}$–orthogonal to $\boldsymbol{\omega}$, we obtain, for all $\boldsymbol{\omega}$,

$$\frac{g'_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}{g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_1}+ \frac{g'_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}{g_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_2} \propto \mathbf{C}_{\mathbf{X}_3}. \tag{B.3}$$

We consider two cases:

1) $\mathbf{C}_{\mathbf{X}_1} \propto \mathbf{C}_{\mathbf{X}_2}$: If $\mathbf{C}_{\mathbf{X}_1} \propto \mathbf{C}_{\mathbf{X}_2}$, i.e. $\mathbf{C}_{\mathbf{X}_2} = a\mathbf{C}_{\mathbf{X}_1}$ for some constant $a > 0$, then it is trivial to show (B.3) is true for all $\boldsymbol{\omega}$.

2) $\mathbf{C}_{\mathbf{X}_1}$ *is not proportional to* $\mathbf{C}_{\mathbf{X}_2}$: Consider all complex vectors $\boldsymbol{\omega}$ such that $\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega} = K$ is a constant. Therefore, $\frac{g'_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}{g_{\mathbf{X}_2}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_2}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_2}$ is a constant matrix. It follows from (B.3) that $\frac{g'_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}{g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}\mathbf{C}_{\mathbf{X}_1}$ must be a constant matrix, i.e. $\frac{g'_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}{g_{\mathbf{X}_1}(\boldsymbol{\omega}^H \mathbf{C}_{\mathbf{X}_1}\boldsymbol{\omega})}$ is a constant for such $\boldsymbol{\omega}$. It can be further deduced that $\frac{g'_{\mathbf{X}_1}}{g_{\mathbf{X}_1}}$ and $\frac{g'_{\mathbf{X}_2}}{g_{\mathbf{X}_2}}$ are both constant for all $\boldsymbol{\omega}$ when $\mathbf{C}_{\mathbf{X}_1}$ is not proportional to $\mathbf{C}_{\mathbf{X}_2}$, which implies that $\mathbf{X}_1$ and $\mathbf{X}_2$ are both complex Gaussian. Q.E.D.

APPENDIX C

MAXIMUM LIKELIHOOD ESTIMATOR FOR η

We need to solve for the ML estimator for $\boldsymbol{\eta}$, i.e.,

$$\hat{\boldsymbol{\eta}}_{\mathrm{ML}} = \underset{\boldsymbol{\eta}}{\operatorname{argmax}} \sum_{i=1}^{M} \log \left(\sum_{n=1}^{N_s} f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s_n)p_n \right).$$

The Newton-Raphson iteration for this problem is given by

$$\boldsymbol{\eta}^{[k]} = \boldsymbol{\eta}^{[k-1]} - \left[\nabla_{\boldsymbol{\eta}}^2 J(\boldsymbol{\eta}^{[k-1]})\right]^{-1} \left[\nabla_{\boldsymbol{\eta}} J(\boldsymbol{\eta}^{[k-1]})\right]^T,$$

$$\text{where } \nabla_{\boldsymbol{\eta}} J(\boldsymbol{\eta}) = \sum_{i=1}^{M} \left[\frac{F_1^{(i)}}{F^{(i)}}, \frac{F_2^{(i)}}{F^{(i)}}\right],$$

$$\nabla_{\boldsymbol{\eta}}^2 J(\boldsymbol{\eta}) = \sum_{i=1}^{M} \begin{bmatrix} \frac{F^{(i)}F_{11}^{(i)}-(F_1^{(i)})^2}{(F^{(i)})^2} & \frac{F^{(i)}F_{12}^{(i)}-F_1^{(i)}F_2^{(i)}}{(F^{(i)})^2} \\ \frac{F^{(i)}F_{12}^{(i)}-F_1^{(i)}F_2^{(i)}}{(F^{(i)})^2} & \frac{F^{(i)}F_{22}^{(i)}-(F_2^{(i)})^2}{(F^{(i)})^2} \end{bmatrix},$$

$$F^{(i)} = \sum_{n=1}^{N_s} f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)p_n,$$

$$F_j^{(i)} = \sum_{n=1}^{N_s} \frac{\partial f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_j} p_n; j = 1, 2,$$

$$F_{jj}^{(i)} = \sum_{n=1}^{N_s} \frac{\partial^2 f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_j^2} p_n; j = 1, 2,$$

$$F_{12}^{(i)} = \sum_{n=1}^{N_s} \frac{\partial^2 f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_1 \partial \eta_2} p_n,$$

$$\frac{\partial f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_j} = \frac{1}{2\pi}\Big[(1-\lambda^2(s))h(c_i(\boldsymbol{\eta}, s))A_i(x_j)$$
$$+g(c_i(\boldsymbol{\eta}, s))P(\eta_j, y_j)\Big]; j = 1, 2,$$

$$\frac{\partial^2 f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_j^2} = \frac{1}{2\pi}\Big[2h(c_i(s))A_i(x_j)P_i(\eta_j, y_j)$$
$$+(A_i(x_j))^2(1-\lambda^2(s))h'(c_i(s))$$
$$+g(c_i(s))P(1, 0)\Big]; j = 1, 2,$$

$$\frac{\partial^2 f_{\Phi|S}(\phi_i; \boldsymbol{\eta}|s)}{\partial \eta_1 \partial \eta_2} = \frac{1}{2\pi}\Big[\big(P(\eta_1, y_1)A_i(x_1)$$
$$+P(\eta_2, y_2)A_i(x_2)\big)h(c_i(s))$$
$$+A_i(x_1)A_i(x_2)(1-\lambda^2(s))h'(c_i(s))\Big],$$

$$A_i(x) = \frac{s\sqrt{a_{11}a_{22}}x}{\sqrt{(sa_{11}+b_{11})(sa_{22}+b_{22})}},$$

$$P(x,y) = -\frac{2s^2 a_{11}a_{22}x + 2s\sqrt{a_{11}a_{22}}y}{(sa_{11}+b_{11})(sa_{22}+b_{22})},$$

$$x_1 = -\cos(\phi_i), x_2 = -\sin(\phi_i), y_1 = \mathrm{Re}\{b_{12}\}, y_2 = \mathrm{Im}\{b_{12}\},$$

$$h(x) = \frac{p(x)}{(1-x^2)^{\frac{5}{2}}},$$

$$h'(x) = \frac{(1-x^2)^{5/2}q(x) + 5x(1-x^2)^{3/2}p(x)}{(1-x^2)^5},$$

$$q(x) = \frac{4 - 4x^2 - 4x\sqrt{1-x^2}\arccos(x)}{\sqrt{1-x^2}},$$

$$p(x) = 3x\sqrt{1-x^2} - (1+2x^2)\arccos(x).$$

We use $\boldsymbol{\eta}^{[0]} = \frac{a_{12}}{\sqrt{a_{11}a_{22}}}$ as an initial guess. The stopping criterion used in the paper is $||\boldsymbol{\eta}^{[k]} - \boldsymbol{\eta}^{[k-1]}|| < 10^{-6}$.

APPENDIX D

RELATED PUBLICATIONS

Journal Papers

• Y. Rakvongthai and S. Oraintara, "Noisy Statistical Texture Retrieval Using Complex Wavelets," in preparation.

• Y. Rakvongthai and S. Oraintara, "Estimating the Relative Phase Parameters of Complex Wavelet Coefficients in Noise," under peerreview for IEEE Transactions on Signal Processing, Jan. 2011.

• Y. Rakvongthai, A. P.N. Vo and S. Oraintara, "Complex Gaussian Scale Mixtures of Complex Wavelet Coefficients," IEEE Transactions on Signal Processing, vol. 58, no. 7, pp. 3545–3556, Jul. 2010.

Conference Papers

• Y. Rakvongthai and S. Oraintara, "A Non-iterative Estimator for the Concentration Parameter of the Relative Phase Distribution," IEEE ICASSP, Dallas, USA, Mar. 2010.

• Y. Rakvongthai and S. Oraintara, "An Estimation Method for the Relative Phase Parameters of Complex Wavelet Coefficients," IEEE ICASSP, Dallas, USA, Mar. 2010.

• Y. Rakvongthai and S. Oraintara, "Statistical Image Modeling with the Magnitude Probability Density Function of Complex Wavelet Coefficients," IEEE ISCAS, Taipei, Taiwan, May 2009.

REFERENCES

[1] MIT Vision and Modeling Group. http://vismod.media.mit.edu.

[2] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions.* National Bureau of Standards, June 1964.

[3] Alin Achim and Ercan E. Kuruoğlu. Image denoising using bivariate α-stable distributions in the complex wavelet domain. *IEEE Signal Processing Lett.*, 12(1):17–20, January 2005.

[4] Ryan Anderson, Nick Kingsbury, and Julian Fauqueur. Determining multiscale image feature angles from complex wavelet phases. In *Proc. ICIAR'05*, Toronto, Canada, September 2005.

[5] David F. Andrews and Colin L. Mallows. Scale mixtures of normal distributions. *J. Roy. Stat. Soc. B*, 36(1):99–102, 1974.

[6] George E.P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA, 1973.

[7] Emmanuel J. Candès, David L. Donoho, and Lexing Ying. Fast discrete curvelet transform. *Multiscale model. and simul.*, 5(3):861–899, September 2006.

[8] Tianhorng Chang and C.C. Jay Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Processing*, 2:429–441, October 1993.

[9] Dongwook Cho and Tien D. Bui. Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Processing: Image Communications*, 20(1):77–89, Jan. 2005.

[10] Hyeokho Choi, Justin K. Romberg, Richard G. Baraniuk, and Nick G. Kingsbury. Hidden Markov tree modeling of complex wavelet transforms. In *Proc. IEEE ICASSP'00*, volume 1, pages 133–136, June 2000.

[11] S. K. Choy and C. S. Tong. Statistical wavelet subband characterization based on generalized Gamma density and its application in texture retrieval. *IEEE Trans. Image Processing*, 19(2):281–289, February 2010.

[12] Thomas Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

[13] Matthew S. Crouse, Robert D. Nowak, and Richard G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Processing*, 46:886–902, April 1998.

[14] Michael S. Davis, Patrick Bidigare, and Daniel Chang. Statistical modeling and ML parameter estimation of complex SAR imagery. In *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, pages 500–502, Nov. 2007.

[15] Peter de Rivaz and Nick G. Kingsbury. Complex wavelet features for fast texture image retrieval. In *Proc. IEEE ICIP'99*, pages 109–113, 1999.

[16] Minh N. Do and Martin Vetterli. Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models. *IEEE Trans. Multimedia*, 4(4):517–527, December 2002.

[17] Minh N. Do and Martin Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Trans. Image Processing*, 11(2):146–158, Feb. 2002.

[18] Jalal M. Fadili and Larbi Boubchir. Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities. *IEEE Trans. Image Processing*, 14(2):231–240, Feb. 2005.

[19] Nathaniel. R. Goodman. Statistics analysis based on certain multivariate complex distribution (an introduction). *Ann. of Math. Stat.*, 34(1):152–177, 1963.

[20] I. S. Gradshteyn and I. M. Ryzhik. *Tables of Integrals, Series, and Products.* Academic Press, June 2007.

[21] Gösta H. Granlund and Hans Knutsson. *Signal Processing for Computer Vision.* Kluwer Academic, Dordrecht, The Netherlands, 1995.

[22] Jose A. Guerrero-Colón, Luis Mancera, and Javier Portilla. Image restoration using space-variant gaussian scale mixture in overcomplete pyramids. *IEEE Trans. Image Processing*, 17(1):27–41, Jan. 2008.

[23] Hasan Hamden. Characterizing and approximating infinite scale mixtures of normals. *Commun. in Stat.-Theory and Methods*, 35(3):407–413, 2006.

[24] David K. Hammond and Eero P. Simoncelli. Image modeling and denoising with orientation-adapted Gaussian scale mixtures. *IEEE Trans. Image Processing*, 17(11):2089–2101, Nov. 2008.

[25] M.C. Jones and Arthur Pewsey. A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, 100(472):1422–1428, 2005.

[26] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Upper Saddle River, NJ, 1993.

[27] Nick G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. and Computat. Harm. Anal.*, 10(3):234–253, May 2001.

[28] Samuel Kotz, Tomasz J. Kozubowski, and Krzysztof Podgórski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance.* Birkhuser, Boston, MA, 2001.

[29] Peter D. Kovesi. Image features from phase congruency. *J. Comput. Vis. Res.*, 16(3), Summer 1999.

[30] Roland Kwitt and Andreas Uhl. Lightweight probabilistic texture retrieval. *IEEE Trans. Image Processing*, 19(1):241–253, January 2010.

[31] Siwei Lyu and Eero P. Simoncelli. Nonlinear image representation using divisive normalization. In *Proc. IEEE CVPR'08*, pages 1–8, Anchorage, USA, June 2008.

[32] David K. C. Macdonald. Some statistical properties of random noise. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 368–372, July 1949.

[33] A. D. E. Maliani, M. E. Hassouni, N. Lasmar, and Y. Berthoumieu. Texture classificatio based on the generalized Gamma distribution and the dual tree complex wavelet transform. In *Proc. ISVC'10*, pages 1–4, 2010.

[34] Stéphane G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11:674–693, July 1989.

[35] Stéphane G. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, CA, 1 edition, 1998.

[36] B. S. Manjunath and W. Y. Ma. *IEEE Trans. Pattern Anal. Machine Intell.*

[37] Kanti V. Mardia and Peter E. Jupp. *Directional statistics.* John Wiley & Sons, Chichester, England, 2000.

[38] Kenneth S. Miller. Complex Gaussian processes. *SIAM Rev.*, 11(4):544–567, Oct. 1969.

[39] Mark Miller and Nick Kingsbury. Image denoising using derotated complex wavelet coefficients. *IEEE Trans. Image Processing*, 19(9):1500–1511, September 2008.

[40] Fredy D. Neeser and James L. Massey. Proper complex random processes with applications to information theory. *IEEE Trans. Inform. Theory*, 39(4):1293–1302, July 1997.

[41] Truong T. Nguyen and Hervé Chauris. Uniform discrete curvelet transform for seismic processing. In *EAGE Conf. and Tech. Exhib. (EAGE'08)*, Rome, Italy, June 2008.

[42] Truong T. Nguyen and Hervé Chauris. Uniform discrete curvelet transform. *IEEE Trans. Signal Processing*, 58(7):3618–3634, Jul. 2010.

[43] Truong T. Nguyen and Soontorn Oraintara. The shiftable complex directional pyramid-part I: Theoretical aspects. *IEEE Trans. Signal Processing*, 56(10):4651–4660, Oct. 2008.

[44] Alan V. Oppenheim and Jae S. Lim. The importance of phase in signals. *Proc. of the IEEE*, 69(5):529–541, May 1981.

[45] Athanasios Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, NY, 2002.

[46] Bernard Picinbono. Second-order complex random vectors and normal distributions. *IEEE Trans. Signal Processing*, 44(10):2637–2640, Oct. 1996.

[47] Javier Portilla, Vasily Strela, Martin J. Wainwright, and Eero P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Processing*, 12:1338–1351, November 2003.

[48] S.M. Mahbubur Rahman, M. Omair Ahmad, and M.N.S. Swamy. Bayesian wavelet-based image denoising using the Gauss-Hermite expansion. *IEEE Trans. Image Processing*, 17(10):1755–1771, Oct. 2008.

[49] S.M. Mahbubur Rahman, M. Omair Ahmad, and M.N.S. Swamy. Statistics of 2-D DT-CWT coefficients for a Gaussian distributed signal. *IEEE Trans. Circuits Syst. I*, 55(7):2013–2025, Aug. 2008.

[50] Yothin Rakvongthai and Soontorn Oraintara. Statistical image modeling with the magnitude probability density function of complex wavelet coefficients. In *Proc. IEEE ISCAS'09*, pages 1879–1882, Taipei, Taiwan, May 2009.

[51] Yothin Rakvongthai and Soontorn Oraintara. Estimating the relative phase parameters of complex wavelet coefficients in noise. *IEEE Trans. Signal Processing*, Jan. 2011. Submitted.

[52] Yothin Rakvongthai, An P.N. Vo, and Soontorn Oraintara. Complex Gaussian scale mixtures of complex wavelet coefficients. *IEEE Trans. Signal Processing*, 58(7):3545–3556, Jul. 2010.

[53] Trygve Randen and John H. Husoy. Filtering for texture classification: A comparative study. *IEEE Trans. Pattern Anal. Machine Intell.*, 21:291–310, April 1999.

[54] K. R. Rao, D. N. Kim, and J. J. Hwang. *Fast Fourier Transforms - Algorithms and Applications*. Springer, 2010.

[55] Peter J. Schreier. Bounds on the degree of impropriety of complex random vectors. *IEEE Signal Processing Lett.*, 15:190–193, 2008.

[56] Peter J. Schreier and Louis L. Scharf. Second-order analysis of improper complex random vectors and processes. *IEEE Trans. Signal Processing*, 51(3):714–725, March 2003.

[57] Ivan W. Selesnick. The estimation of Laplace random vectors in addtive white Gaussian noise. *IEEE Trans. Signal Processing*, 56(8):3482–3496, August 2008.

[58] Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Mag.*, 22:123–151, November 2005.

[59] Srinivasan Selvan and Srinivasan Ramakrishnan. SVD-based modeling for image texture classification using wavelet transformation. *IEEE Trans. Image Processing*, 16(11):2688–2696, November 2007.

[60] Levent Şendur and Ivan W. Selesnick. Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Trans. Signal Processing*, 50(11):2744–2756, November 2002.

[61] Levent Şendur and Ivan W. Selesnick. Bivariate shrinkage with local variance estimation. *IEEE Signal Processing Lett.*, 9(12):438–441, December 2002.

[62] Gilbert Strang and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley - Cambridge, 1997.

[63] Shan Tan, Licheng Jiao, and Ioannis A. Kakadiaris. Wavelet-based Bayesian image estimation: from marginal and bivariate prior models to multivariate prior models. *IEEE Trans. Image Processing*, 17(4):469–481, Apr. 2008.

[64] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides. Rotation invariant texture retrieval with Gaussianized steerable pyramids. *IEEE Trans. Image Processing*, 15(9):2702–2718, September 2006.

[65] Michael Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. Image Processing*, 4(4):1549–1560, November 1995.

[66] Nuno Vasconcelos and Andrew Lippman. A unifying view of image similarity. In *Proc. ICPR'00*, volume 1, pages 38–41, 2000.

[67] An Vo, Soontorn Oraintara, and Nha Nguyen. Vonn distribution of relative phase for statistical image modeling in complex wavelet domin. *Signal Processing*, 91(1):114–125, January 2011.

[68] An P.N. Vo and Soontorn Oraintara. A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation. *Signal Processing: Image Communication*, 25(1):28–46, January 2010.

[69] An P.N. Vo, Soontorn Oraintara, and Truong T. Nguyen. Using phase and magnitude information of the complex directional filter bank for texture image

retrieval. In *Proc. IEEE ICIP'07*, volume 4, pages IV–61–IV–64, San Antonio, USA, September 2007.

[70] An P.N. Vo, Soontorn Oraintara, and Truong T. Nguyen. Statistical image modeling using distribution of relative phase in the complex wavelet domain. In *EUSIPCO'08*, Lausanne, Switzerland, Aug. 2008.

[71] Martin J. Wainwright and Eero P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. *Adv. Neural Inform. Process. Syst.*, 12:855–861, 2000.

[72] Martin J. Wainwright, Eero P. Simoncelli, and Alan S. Willsky. Random cascades on wavelet trees and their use in analying and modeling natural images. *Appl. and Computat. Harm. Anal.*, 11:89–123, 2001.

[73] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Processing*, 13(4):600–612, April 2004.

[74] Zhou Wang and Eero P. Simoncelli. Local phase coherence and the perception of blur. *Adv. Neural Inform. Process. Syst.*, 16, 2004.

[75] Robin A. Wooding. The multivariate distribution of complex normal variables. *Biometrika*, 4(1-2):212–215, 1956.

[76] Ruihong Wu Zhiyong Meng Yan Shang, Weimin Hou. Antinoise rotation invariant texture classification based on LBP features of dominant curvelet subbands. In *Proc. IITA'08*, pages 365–369, 2008.

[77] Kung Yao, Marvin K. Simon, and Ezio Biglieri. A unified theory on wireless communication fading statistics based on sirp. In *IEEE Workshop Signal Process. Adv. Wireless Commun.*, pages 135–139, Lisboa, Portugal, July 2004.

[78] Steeve Zozor and Christophe Vignat. Some results on the denoising problem in the elliptically distributed context. *IEEE Trans. Signal Processing*, 58(1):134–150, January 2010.

BIOGRAPHICAL STATEMENT

Yothin Rakvongthai earned his B.Eng. degree (with first class honors) from Chulalongkorn University, Thailand in 2004, and M.S. degree from the University of California Los Angeles (UCLA) in 2005, all in electrical engineering. He received a bronze medal from the 41st International Mathematical Olympiad in Taejon, Korea, and a Lucent Global Science Scholarship in 2000 and 2001, respectively. His current research interests include statistical image modeling in the complex wavelet domain and its applications to image processing.