RICE UNIVERSITY

Image Data Compression Using Wavelet Decomposition

by

Dong Wei

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Approved, Thesis Committee:

C. S. Burrus, Chairman
Professor
Electrical and Computer Engineering

R. G. Baraniuk
Assistant Professor
Electrical and Computer Engineering

R. O. Wells, Jr.
Professor
Mathematics

Houston, Texas

May, 1995

Image Data Compression Using Wavelet Decomposition

Dong Wei

Abstract

Wavelet-based multi-resolution representation has become a cutting-edge technology in the area of image data compression. Though the discrete wavelet transform is closely related to the perfect-reconstruction octave-band filter banks used in subband coding schemes, wavelets have provided very promising new ideas and insights for image data compression. Wavelet-based coding techniques have achieved competitive performance compared with other well-known image coding techniques. This thesis develops a general framework for wavelet-based lossy image coding. We discuss two application problems and develop the corresponding fast algorithms using wavelet-based techniques (e.g., optimum wavelet-packet bases selection, wavelet-domain soft-thresholding).

Acknowledgments

Writing a thesis is *never* the work of one person. Therefore, I would like to thank here those who have helped me in one way or another to complete this work.

First, I would like to take this opportunity to thank my advisor Dr. C. S. Burrus for introducing me to the elegant and fruitful wavelet theory, creating the right environment for me to work in this area, and supervising me to do my graduate study. Then, I would like to thank Dr. R. G. Baraniuk, Dr. R. O. Wells, Jr., and Dr. V. Veeravalli, whose advice helped me finish this work.

Further, special thanks to all other members of the Digital Signal Processing (DSP) Group and Computational Mathematics Laboratory (CML), who were always helpful in solving various problems, and created a friendly and pleasant environment for research. In particular, I would like to thank Mr. Hai Tao, who is now with University of Illinois at Urbana-Champaign, for the helpful discussion on several image coding topics.

Also, the funding for this research from Bell Northern Research (BNR) and Texas Advanced Technology Program (TATP) is gratefully acknowledged.

Finally, I should greatly thank my wife Xun Yang, and my parents, for their patience, encouragement, full support, and too many other things to write down.

Contents

Illustrations

Tables

Chapter 1

Introduction

1.1 Motivation

Image data compression is concerned with minimizing the number of bits required to represent an image [15, 17, 23, 24, 32].

The total number of bits required to transmit or store the digital images and the digital image sequences (digital video) can be extremely large. Often this number surpasses the bandwidth of the intended transmission, or the storage capacity of the storage media, or makes certain digital image communication services economically infeasible. Since, however, most visual images and video sequences contain a large amount of statistical redundancy and visual irrelevancy, data compression or source coding methods can be applied to the data so as to reduce the number of bits and make widespread use of the digital imagery practical. The data compression is of importance to many existing and possible future image communication services and systems, such as archival systems for medical images and museums, remote sensing imaging, geological surveys, multi-media applications, video telephony, digital video recording, and new video distribution services such as digital television, high-definition television, and so on. The application of data compression is also possible in the development of fast algorithms where the number of operations required to implement an algorithm is reduced by working with the compressed data.

This thesis discusses one of the relatively recently developed compression techniques and its application to the image data compression. This compression technique uses multiple resolutions of the image and is based on the wavelet theory.

1.2 Fundamentals

1.2.1 Basic Idea

The digital image data can be compressed by eliminating the redundant information so as to provide a more efficient representation of data while preserving the essential information contained in the data. In this thesis, we consider only the lossy compression algorithms. There are two types of redundancy that can be exploited by the image compression systems:

1. *Spatial Redundancy.* It is related to such factors as predictability, randomness and smoothness in the image data. For example, in almost all natural images, the values of neighboring pixels are strongly correlated.

2. *Spectral Redundancy.* In images composed of more than one spectral band, the spectral values for the same pixel location are often correlated.

The removal of spatial and spectral redundancies is often accomplished by the predictive coding or the transform coding. Quantization (including thresholding) is the most important means of *irrelevancy* reduction.

1.2.2 Performance Measures

The following outlines those metrics most often used for evaluating the performance of the compression algorithms. In this thesis, among all the metrics, we mainly consider bit rate and peak signal-to-noise ratio (PSNR).

Bit Rate/Compression Ratio

The *bit rate* is the average number of bits per pixel (bpp) in the compressed image. The *compression ratio* is the ratio between the number of bits used to represent the compressed image and the number of bits used to represent the original image.

Numerical Distortion Measure

For n-bit images, or equivalently, 2^n-gray-scale images, *peak signal-to-noise ratio* in decibels (dB) is computed as

$$\text{PSNR} = 10 \log_{10} \frac{(2^n - 1)^2}{\text{MSE}}$$

where MSE is the mean-squared error defined as

$$\text{MSE} = \|X - \hat{X}\|_2^2 = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} [X(i,j) - \hat{X}(i,j)]^2$$

and, N and M are the width and the height, respectively, of the images in pixels, X and $\hat{X}$ are the original image and the reconstructed image, respectively. The norm most often used to measure "distance" is the l^2-norm, but that is more because this is the easiest norm to handle than for any other reason. All experts agree that the l^2-norm is not a good candidate for a "perceptual" norm, but as far as we know, there is no agreement on a better candidate. Based on the research on the human visual system, it is argued that in most instances the error incurred in image compression should be measured in the mean-absolute sense (l^1-norm) instead of the mean-square sense (l^2-norm) [13].

Perceptual Distortion Measure

The perceptual errors are difficult to quantify mathematically. Also, the numerical performance is not directly related to the subjective image quality. This means that an algorithm may have a larger numerical error but appear to be better in visual image quality. This can happen if one algorithm exhibits greater error in areas or frequencies which the eye is relatively insensitive. Also, in general, the coding artifacts due to the coarse quantization, such as the block effect, the contouring effect, the ringing effect, etc., are *not* clearly indicated by those numerical distortion metrics, but can be viewed easily by human eyes.

Algorithm Complexity and Speed

In general, the computational complexity of the data compression algorithms are evaluated on the basis of the number and the type of operations (multiplies, divides, look-ups, etc.) performed per pixel. These numbers translate to approximate running time for a computer platform bench-marked for the speed of each operation assuming a serial implementation of the algorithm.

Susceptibility to Error

Some algorithms will suffer the localized information loss from the channel bit-error effects, while others may experience total image destruction. This indicates certain algorithms may have the different error detection and correction requirements. The error detection algorithms reduce the loss of information but also increase the overhead, which in turn increase the total bit rate.

1.2.3 Distortion Rate Functions

The rate distortion theory is a branch of the information theory that allows us to calculate the performance bounds without teh consideration of a specific coding method. The *distortion rate function*, $D(R)$, provides a lower bound D on the average distortion for a given rate R, and hence an upper bound on the performance of the practical coders. It can be shown that $D(R)$ is a monotonically decreasing, convex function [6]. Unfortunately, the theory does not provide us with a method on how to construct a practical optimum coder. However, the $D(R)$ bounds provide the characteristics curves of the coders and serve as calibrators of the coding performance (see Appendix).

1.3 Image Coding Techniques

1.3.1 Predictive Coding—DPCM

The philosophy underlying predictive techniques is to remove the mutual redundancy between successive pixels, and encode and transmit only the new information [17, 23, 24, 32]. The *differential pulse coded modulation* (DPCM), one of the most widely used predictive coding techniques, is based on quantizing and coding a prediction error signal (Figure 1.1). The DPCM performs well at high bit rate, but worse than transform coding techniques at low bit rate. Usually, the DPCM is combined with other techniques in many coding systems.

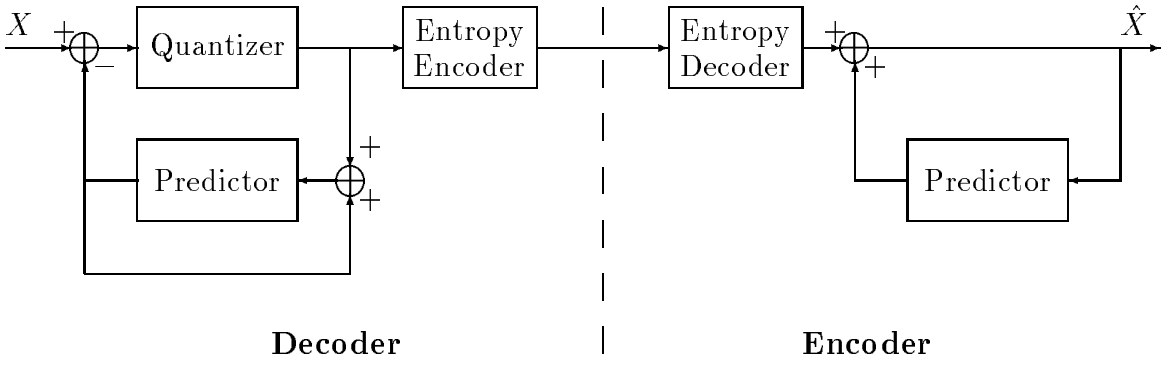

Figure 1.1 A DPCM codec

1.3.2 JEPG/DCT

A joint ISO/CCITT* committee known as JPEG (Joint Photographic Experts Group) has established the first international compression standard for the continuous-tone still images [33, 42]. A generic scheme is specified as a baseline JPEG method for the lossy compression (Figure. 1.2). The method is based on a division of the image

*ISO standards for International Standardization Organization. CCITT standards for International Consultative Committee for Telephone and Telegraph.

into blocks of 8x8 pixels, after which each block is transformed with a discrete cosine transform (DCT). The transform coefficients are quantized, where each frequency is weighted according to its importance to the human visual system (HVS). The quantized DC coefficients are differentially coded using the DPCM method and the quantized AC coefficients are ordered into a zig-zag sequence and coded by a run-length coder and a Huffman coder block-wise. The cosine functions used in the DCT are nearly optimal for stationary signals, but images are rarely stationary. So, the DCT-based algorithms do not tolerate the high compression ratios. Also, the JPEG method cannot efficiently exploit the redundancy among the blocks. Therefore, the JPEG method performs very well at high or medium bit rates and introduces perceptibly annoying blocking artifacts (horizontal and vertical artificial edges due to the block-wise transform coding) at low bit rates.

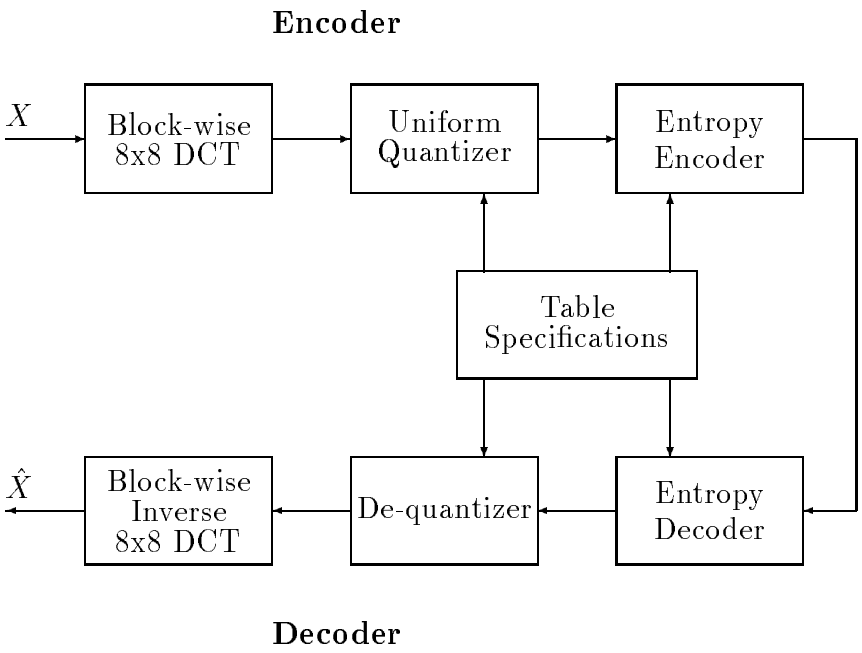

Figure 1.2 JPEG baseline coding algorithm

1.3.3 Vector Quantization

The vector quantization (VQ) is a popular and powerful scheme for compressing the correlated discrete signal sets whose characteristics have been "trained" initially [17, 19, 31]. We define a *vector quantizer* Q of dimension k and size N a mapping from a vector in k-dimensional Euclidean space, $\mathbf{R}^k$, into a finite set $\mathcal{C}$ (called a *codebook*) containing N output or reproduction vectors (called *codewords*), $Q : \mathbf{R}^k \rightarrow \mathcal{C}$, where $\mathcal{C} = (y_1, y_2, \ldots, y_N\}$ and $y_i \in \mathbf{R}^k$ for each $i \in \mathcal{J} \equiv \{1, 2, \ldots, N\}$. The potential advantages of vector quantizers are:

1. the ability to exploit both the linear dependence and the nonlinear dependence among the vector coordinates,

2. the extra freedom in choosing the partition geometry on the vector space,

3. the ability to achieve *fractional* values of resolution (measured in bits per sample or bits per vector component).

The last feature is particularly important when the optimal bit-allocation algorithms are used. These features indicate that the scalar quantization (SQ) is simply a restricted special case of the VQ. According to Shannon's rate distortion theory, better results are always obtained when vectors rather than scalars are coded. Therefore, the VQ can indeed give superior performance over the SQ. It can be shown that the VQ can approach the rate distortion limit as the vector dimensionality becomes large. However, one obvious disadvantage of the VQ is the high computational cost for the design of a sufficiently good vector quantizer, which grows exponentially with the vector dimensionality.

1.3.4 Subband Coding

The *subband coding* (SBC) first appeared in 1976 [11]. It is an important and widely used type of analysis-synthesis coding systems where the signal is decomposed onto

a set of narrow-band signals that in some sense offer a more fundamental or more primitive representation of the signal [39, 40, 41, 45, 47]. The quantized subband signals are then used to *synthesize* a reproduction of the original signal. Thus, the highly correlated signals are amenable to the efficient removal of redundancy by the SBC. A general one-dimensional SBC scheme is illustrated in Figure 1.3. Most two-dimensional subband decompositions are built by cascading one-dimensional subband filter banks. The input signal X is convolved in a bank of M bandpass analysis filters with the frequency responses $\{H_i(e^{j\omega})\}$ and down-sampled by the factors $\{k_i\}$ that correspond to the bandwidths of $\{H_i(e^{j\omega})\}$. The frequency responses and the down-sampling factors are usually chosen such that the number of samples before and after the subband decomposition is the same, i.e., $\sum_{i=1}^{M} \frac{1}{k_i} = 1$, which ensures that there is no additional redundancy introduced by the subband decomposition. The subband signals are quantized and transmitted using the variable length coders (VLC). At the decoder, the subband signals are decoded by the variable length decoders (VLD), up-sampled by factors $\{k_i\}$ and passed through a bank of synthesis filters with frequency responses $\{G_i(e^{j\omega})\}$. The output signals of the synthesis filters are finally summed up to yield the reconstructed signal. In fact, the DCT can be interpreted as a special subband coding technique.

1.3.5 Wavelet Transform Coding

Theoretical and applied research in the field of wavelets has made tremendous progress in the last few years [12]. The wavelet transform has become a cutting-edge technology in the image data compression research [1, 2, 3, 13, 27, 37]. The wavelet transform, due to the good localization in both the space domain and frequency domain of the basis functions, can handle the non-stationary signals, hence greater compression ratios can be obtained. Image data compression schemes based on wavelets are rapidly gaining maturity, and have already begun to appear in commercial software/hardware systems. The reconstruction quality of the wavelet compressed images

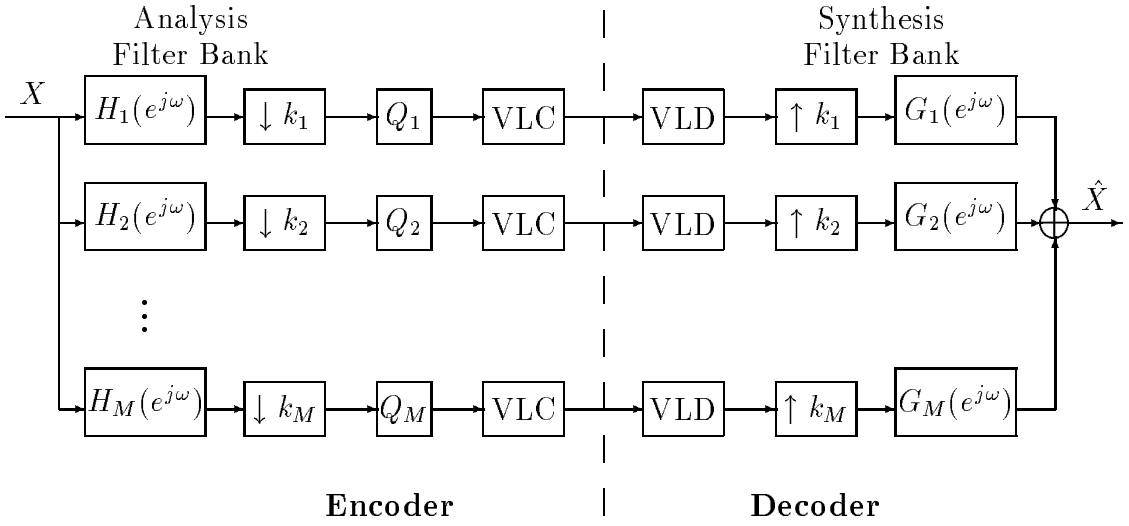

Figure 1.3 A 1-D M-channel SBC scheme

has already moved well beyond the capabilities of the JPEG method. At compression ratios above 30 : 1, the performance of the JPEG coder rapidly deteriorates, while the wavelet coders degrade gracefully well beyond the ratio of 100 : 1. Also, the perceptual quality of the wavelet compressed images is better than that of the JPEG compressed images at low bit rates. In addition, the "zooming" property of the wavelet-based multi-resolution representation is very suitable for the progressive transmission of the image data, which is a good framework for the data management in the scientific visualization when large volume of data is involved. A general wavelet transform encoder is depicted in Figure 1.4. Compression is accomplished by applying a wavelet transform to decorrelate the image data, quantizing the resulting transform coefficients, and coding the quantized values. Image reconstruction is accomplished by inverting the compression operations. We will describe more detailed wavelet-based compression techniques in Chapter 2.

It is interesting to note that the research on the wavelet decomposition coding has had a strong impact on several areas of the numerical analysis, especially in the

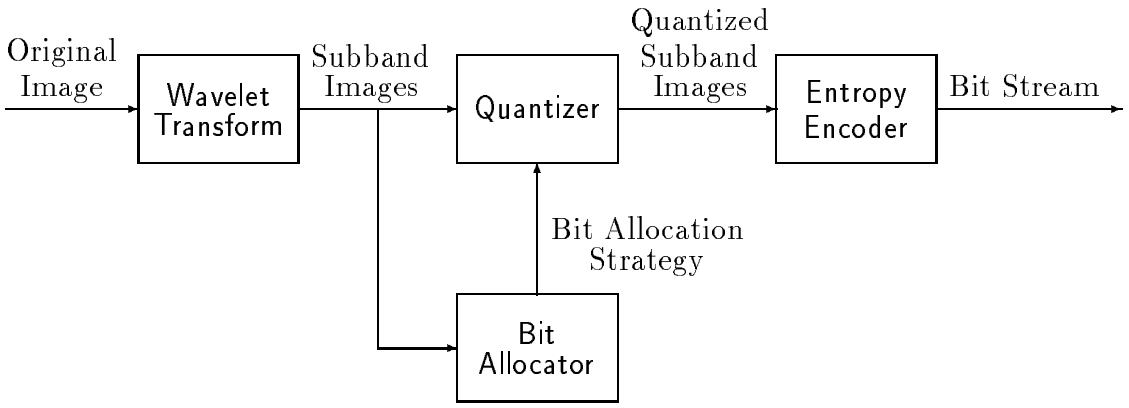

Figure 1.4 A typical wavelet transform encoder

solution of the partial differential equations. The compression of an image, which is just a matrix of intensity values, is not really different from compressing the kernel matrix of a functional operator. The compressed operator is a sparse matrix, and sparse matrix operations can often be performed orders of magnitude faster than their non-sparse counterparts. Undoubtedly, this will lead to new results in the numerical analysis that will impact image data compression, leading to better algorithms in areas such as computer vision, etc.

1.3.6 Other Coding Techniques

There are many other image data compression techniques, such as *fractal coding* [4, 16, 22], *recursive block coding* [15], region-based coding, image coding by exploiting the human visual perception, etc. We do not discuss them in this thesis. In general, each technique has its own merits and drawbacks, and its performance depends on various applications. Therefore, we conjecture that the hybrid coding techniques combining the advantages of the above techniques could be very promising [1, 3, 36, 47].

1.4 Thesis Organization

Having introduced the background of the image data compression, we now describe the layout of this thesis. Chapter 2 contains an explanation of the basic wavelet decomposition theory. In Chapter 3 , we turn our attention to the fundamental transform coding schemes. We then discuss a novel technique for the reduction of quantization noise in Chapter 4. Chapter 5 describes our algorithms for simultaneous noise reduction and data compression. Finally, Chapter 6 gives the summary.

Chapter 2

Wavelet Decomposition

2.1 A Short Review of Wavelet Theory

For the purposes of signal analysis, we would like to have a representation of the signal which contains the information about both time and frequency behavior of the signal. More specially, we want to know the frequency content of the signal at a particular instant in time. However, the resolution in time (Δx) and the resolution in frequency ($\Delta\omega$) cannot both be made arbitrarily small at the same time because their product is lower bounded by the uncertainty principle or the Heisenberg inequality

$$\Delta x \Delta\omega \geq \frac{1}{2}.$$

This means one must trade off the time resolution for the frequency resolution, or vice versa. Thus, it is desirable to obtain a very good resolution in time if one is willing to settle for a low resolution in frequency, or vice versa. The time-frequency resolution of the short-time Fourier transform is fixed over the entire time-frequency plane. The wavelet transform follows the above ideas and allows the resolutions Δx and $\Delta\omega$ to vary in the time-frequency plane in order to obtain a multi-resolution analysis.

Wavelets are functions generated from one single function ψ by dilations and translations

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \qquad a \in \mathbf{R}^+, b \in \mathbf{R}$$

where a is the scaling parameter and b is the shift parameter. The *mother wavelet* ψ must satisfy the admissibility condition

$$\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty$$

where Ψ is the Fourier transform of ψ.

The basic idea of a wavelet transform is to represent any arbitrary function f as a weighted superposition of wavelets

$$f = \sum_{j,k} d_{j,k}(f)\psi_{j,k} \qquad j,k \in \mathbf{Z}$$

where

$$d_{j,k}(f) = \langle \psi_{j,k}, f \rangle = \int_{-\infty}^{+\infty} \psi_{j,k}(x) f(x) dx.$$

For $a = 2^j, b = k2^j$, there exist special choices of ψ such that the $\{\psi_{j,k}\}$ constitute an orthonormal basis. The wavelet is therefore

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$$

where j is the scaling parameter and k is the shift parameter. To introduce the multi-resolution notion, we must define a *scaling function*

$$\phi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k).$$

The projection on this family of functions $\{\phi_{j,k}\}$ gives an approximation of a signal f with resolution 2^{-j}. The wavelet coefficients $\{d_{j,k}\}$ describe the information lost when going from an approximation of f with a resolution 2^{-j+1} to a coarser approximation with a resolution 2^{-j}. All this is translated into the *Mallat algorithm* for the computation of the $d_{j,k}(f) = \langle \psi_{j,k}, f \rangle$ [30]:

$$c_{j,k}(f) = \sum_n h_{2k-n} c_{j-1,n}(f), \qquad d_{j,k}(f) = \sum_n g_{2k-n} c_{j-1,n}(f)$$

where $h_l = 2^{1/2} \int \phi(x-l)\phi(2x)dx$ and $g_l = (-1)^l h_{-l+1}$. If the function f is given in the sampled form, then one can take these samples as the highest order resolution approximation coefficients, $c_{0,k}$, and the Mallat algorithm describes a subband decomposition on these sampled values, with a lowpass filter h and a highpass filter g, called *quadrature mirror filters* (QMFs). Because of their association with the orthonormal

wavelet bases, these filters give *perfect reconstruction* (PR) property, i.e.:

$$c_{j-1,k}(f) = \sum_n [h_{2n-k} c_{j,n}(f) + g_{2n-k} c_{j,n}(f)].$$

The orthonormal compact wavelets corresponding to FIR QMFs have been designed [12].

It is well known that the structures of computations in a wavelet transform and in an octave-band filter bank are identical. Therefore, aside from the different views and interpretations that have given to them, the main difference lies in the filter design. Wavelet filters are chosen so as to be *regular* (that is, they have many zeros at $z = -1$). The classical subband filters are not regular, and they have been designed to have good stop-bands and thus are close to being "regular", at least for the first few octaves of subband decomposition. In my opinion, the contribution of the wavelet theory to image coding comprises two parts. Firstly, there is the link of the discrete filtering with the continuous function spaces. This link is formed by the iterated filters and their limit functions, the scaling function and the wavelet. Secondly, there is the regularity of a filter. For image coding this regularity is the only new criterion the wavelet theory has added to the design process of filters. However, it is still not clear whether the regularity is a nice property in image coding and regular filters are more suitable to the transform coding schemes.

Since images are mostly smooth (except for occasional edges), it seems appropriate that a wavelet transform coding scheme for image analysis should correspond to an orthonormal basis with a reasonably smooth mother wavelet. In order to have fast computation, the QMFs should be short (short filters lead to less smoothness, so, they cannot be too short). On the other hand, it is desirable that the FIR QMFs be linear phase (corresponding to symmetry for wavelets), since such filters can be easily cascaded in pyramidal filter structures without the need for phase compensation. In image processing, the phase information is often much more important than the amplitude information [29]. Unfortunately, orthogonality and symmetry are conflict

properties for design of wavelets. Therefore, there are no nontrivial orthonormal linear phase FIR QMFs with the PR property, regardless of any regularity considerations. The only symmetric wavelets are Haar basis,i.e., $h_0 = h_1 = \sqrt{2}$ and $g_0 = -g_1 = \sqrt{2}$, with all other $h_n, g_n = 0$.

2.2 Dyadic Orthogonal Discrete Wavelet Transform

Figure 2.1 illustrates a single step of the two-dimensional (2-D) dyadic discrete wavelet transform (DWT) for an image, which is accomplished by cascading two separate one-dimensional (1-D) DWTs. In fact, this is the most widely used 2-D DWT. The corresponding 2-D bases are separable and obtained from the products of 1-D bases. The image X is first filtered along the horizontal direction using the reverse versions of a pair of QMFs, resulting in a lowpass image and a highpass image. Here, $\bar{h}$ and $\bar{g}$ are the reverse versions of h and g, respectively, i.e., $\bar{h}_n = h_{-n}$, $\bar{g}_n = g_{-n}$. These two subband images are then downsampled by two (dropping every other filter value) along the horizontal direction. Both two suband images are filtered and downsampled along the vertical direction, resulting in four subband images: X_{ll}, X_{lh}, X_{hl} and X_{hh}, which are directionally sensitive. X_{ll} is a coarse version of X, X_{lh} emphasizes the horizontal image features, X_{hl} the vertical features and X_{hh} the diagonal features. It is customary in the wavelet transform coding to recursively decompose the low resolution subband image. The corresponding inverse transform is illustrated in Figure 2.2. The subband images are upsampled, filtered and summed along both two directions to yield a reconstructed image $\hat{X}$.

If the above transform is orthogonal (the corresponding wavelet bases are orthonormal), The filters (h, g) and $(\tilde{h}, \tilde{g})$ are identical.

The probability density function (PDF) of the wavelet coefficients for a given subband image can be modeled reasonably well by the generalized Gaussian distribution

$$p_X(x) = \frac{rg(r,\sigma)}{2\Gamma(\frac{1}{r})} \exp(-|g(r,\sigma)x|^r)$$

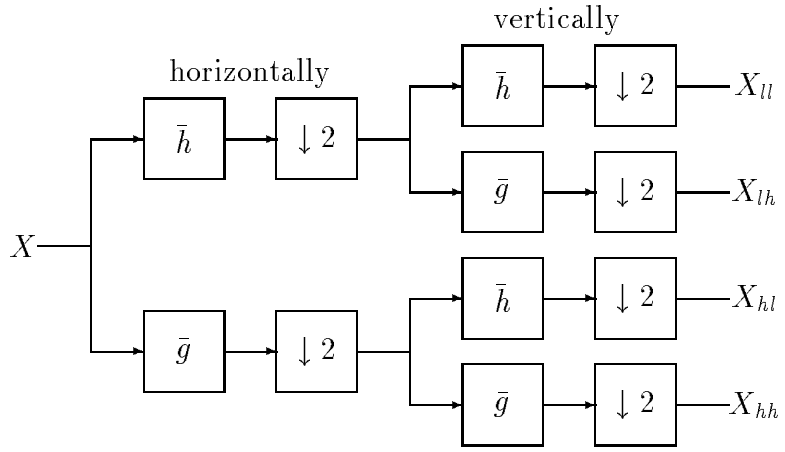

Figure 2.1 Single step of a 2-D dyadic DWT

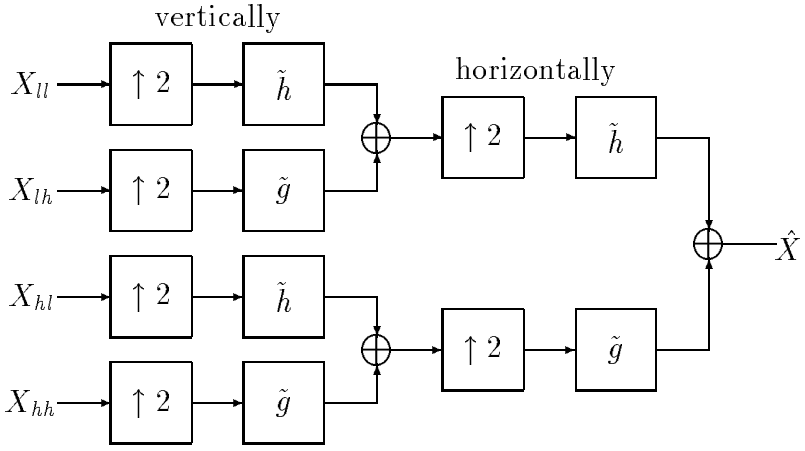

Figure 2.2 Single step of a 2-D dyadic inverse DWT

with

$$g(r,\sigma) = \frac{1}{\sigma}\left[\frac{\Gamma(\frac{3}{r})}{\Gamma(\frac{1}{r})}\right]^{\frac{1}{2}}$$

where σ is the standard deviation of the subband image data, r is the shape parameter describing the exponential rate of decay of the distribution (for the wavelet coefficients, usually $r < 1$), and $\Gamma(\cdot)$ is the Gamma function. Note that $p_X(x)|_{r=1}$ and $p_X(x)|_{r=2}$ are the Laplacian PDF and the Gaussian PDF, respectively. The exception is the subband with the lowest resolution, which is the coarse approximation of the original image. Figure 2.3 depicts an example and indicates that the DWT compacts large amount of energy on very few wavelet coefficients so that it is much easier to compress the wavelet coefficients than the original image.

2.3 Biorthogonal Discrete Wavelet Transform

One can preserve the linear phase by relaxing the orthonormality requirement, and using the biorthogonal bases [8]. It is then still possible to construct examples where the mother wavelets have arbitrarily high regularity [12]. For the biorthogonal DWT, the filters (h, g) and $(\tilde{h}, \tilde{g})$ are different, satisfying

$$\sum_n h_n \tilde{h}_{n+2k} = \delta_{k,0}, \qquad \tilde{g}_n = (-1)^n h_{-n+1}, \qquad g_n = (-1)^n \tilde{h}_{-n+1}.$$

Besides the advantage of the linear phase in the QMFs, the biorthogonal DWT also has the following merits:

1. It is possible for the analysis lowpass filter and the synthesis lowpass filter to have different lengths, which could result in better coding performance than the filters with equal lengths.

2. It is possible for the filters to have odd length, which is impossible for the orthogonal DWT.

3. There exists a family of biorthogonal wavelet bases, whose corresponding QMF coefficients are dyadic rationals (i.e., the denominators have the form $2^n, n \in \mathbf{Z}$)

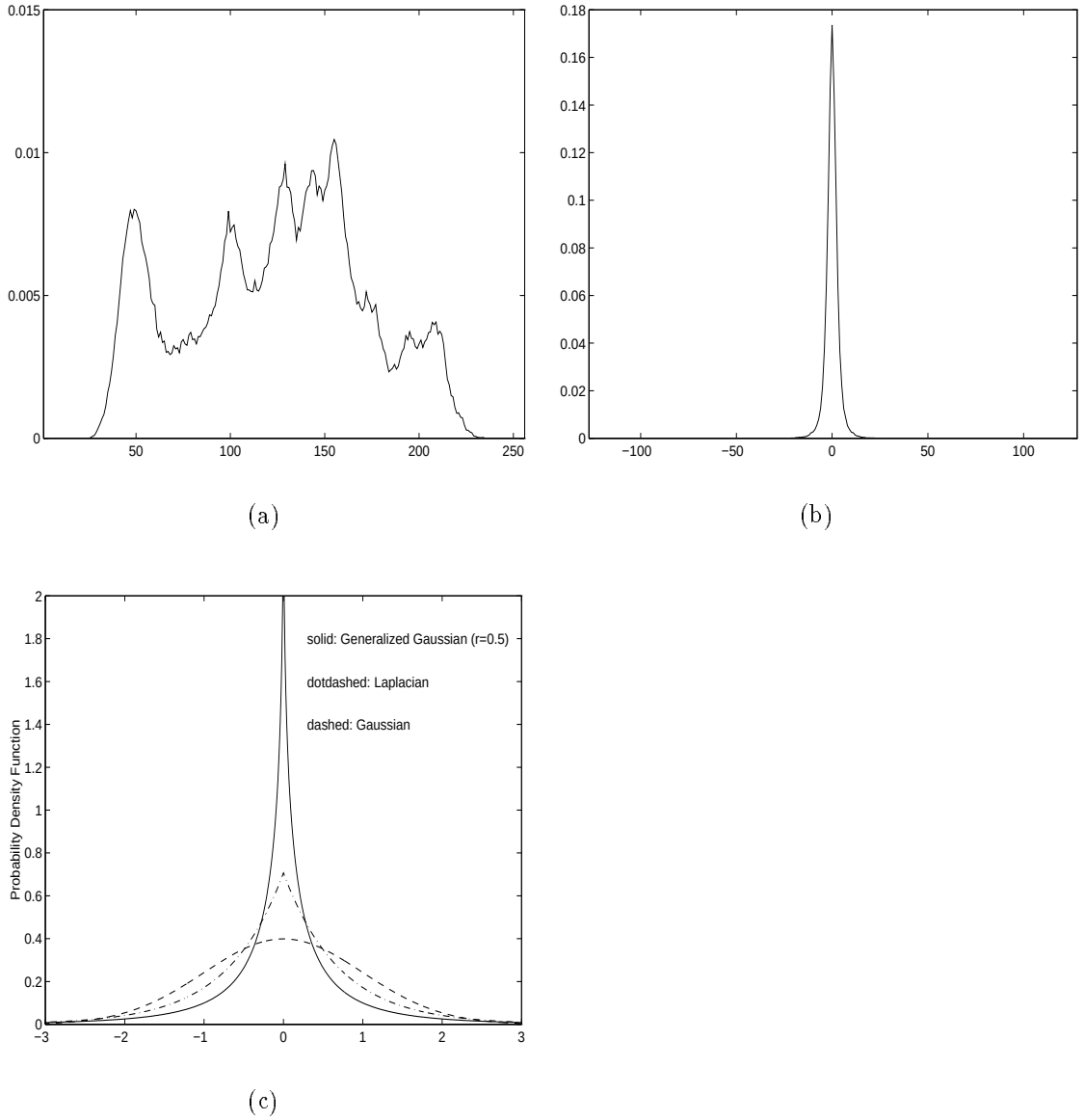

Figure 2.3 (a) Normalized histogram of the original Lenna image (256 gray-scale); (b) Normalized histogram of a subband image from the DWT of Lenna image; (c) PDFs of the generalized Gaussian distribution.

[8]. Thus, the multiplication and the division operations can be simplified to the addition and the subtraction for the floating-point numbers or to shifting for the integer numbers. This makes the transform much faster. This property also makes lossless coding possible, which is usually impossible for the transform coding.

One disadvantage for the biorthogonal DWT is that such a nonorthogonal transform does not preserve the l^2 norm of the quantization error. Therefore, some optimization schemes, such as the best bases selection, which are based on the orthogonal transforms, do not apply to the biorthogonal DWT.

Figure 2.4 illustrates the energy compaction properties of some wavelet filters with comparable lengths. We choose the Daubechies orthogonal wavelet filter of length 8 (Daub8) [12], and two implementations of the biorthogonal wavelet filter pair of lengths 9 and 7 (CDF97 and CDF79) designed by Cohen, Daubechies and Feauveau [8]. The implementation of using the 9-tap filter as the analysis lowpass filter and using the 7-tap filter as the synthesis lowpass filter leads to the best performance. The opposite implementation using the same filters results in even worse performance than the Daub8 filter. The reason is that the long filter will compact more energy and the short filter will accumulate less quantization error.

2.4 Shift-Invariant Discrete Wavelet Transform

The lack of shift invariance has been acknowledged to be one of the serious drawbacks of the orthogonal DWT [12], i.e., there is no "simple" relationship between the wavelet coefficients of the original signal and the shifted signal. It has been noticed that the shift sensitivity is caused by the aliasing due to the down-sampling operators in each subband. By relaxing the requirement for critical sampling in the subbands, which results in a *redundant* DWT and loses the orthogonality of the transform, shift-invariance can be established. One way of computing a shift-invariant DWT (SIDWT)

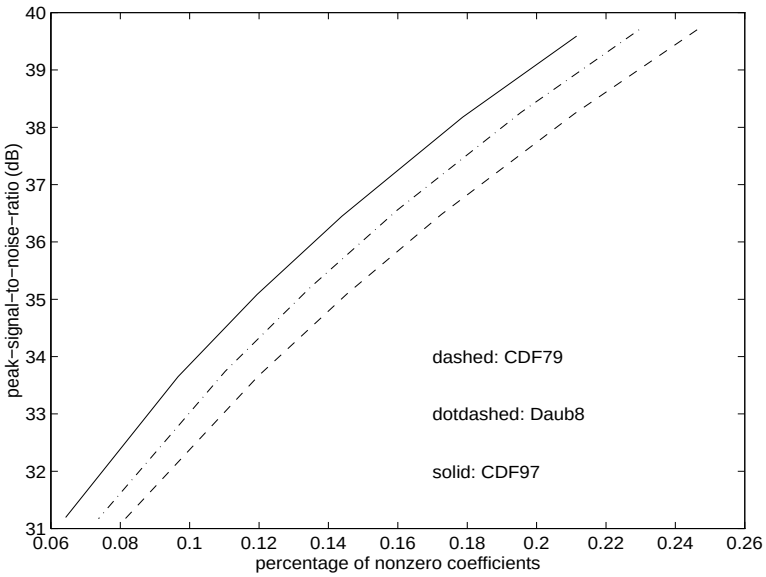

Figure 2.4 Energy compaction property of DWT

is to compute the DWT of *all* shifts. It has been shown that there are only $N \log N$ different coefficient values among those corresponding to all shifts of the input signal. In [28], a quad-tree search algorithm is proposed for searching the best representation among the DWTs of all the shifted images with a given cost criterion.

2.5 Quincunx Discrete Wavelet Transform

The *quincunx* DWT (QDWT) uses non-separable and non-oriented 2-D filters [25]. A non-separable downsampling by a factor of two of an image $X(n_1, n_2)$ is obtained by retaining only samples satisfying

$$(n_1, n_2) = (m_1 + m_2, m_1 - m_2) \qquad m_1, m_2 \in \mathbf{Z}.$$

The resulting samples are located on a quincunx sub-lattice of $\mathbf{Z}^2$. Similar to its 1-D counterpart, the QDWT provides only two subband images at each resolution level. The original image is decomposed with a resolution scale of $\sqrt{2}$ (and not 2 as in the

separable case), which means that the QDWT will be twice as fine as the dyadic DWT for the same decomposition level. Also, the QDWT is more isotropic. The McClellan transform can be used to design 2-D filters from 1-D filters. In addition, other downsampling schemes can be applied.

The dyadic DWT is a degenerate 2-D DWT as it does not make full use of the 2-D characteristics in an image. However, it has a simple implementation with low computational complexity due to the separability of the filters. The QDWT is more computationally expensive, and the design of 2-D filters with good performance is more complicated.

2.6 Wavelet Packets Transform

Wavelet packet (WP) bases [9] were introduced recently as a collection of orthonormal bases for the discrete functions of $\mathbf{R^N}$. This library contains the well-known wavelet basis, Short-Time-Fourier-Transform(STFT)-like basis, Walsh functions, and smooth versions of Walsh functions called wavelet packets. The wavelet packets represent a generalization of the multi-resolution decomposition methods, and comprise the entire family of subband coded (tree) decompositions. The library of wavelet packet bases organizes itself into a homogeneous tree (e.g., for two-dimensional signals such as images, this library has a structure of a complete quad-tree.), which can be efficiently searched for a "best basis" under some optimality criterion. The entire WP tree can be obtained by the recursive decomposition of both the low-pass subband and the high-pass subband using the quadrature mirror filters. Each admissible basis appears as a subtree formed by pruning the whole tree (Figure 2.5).

Due to the non-stationary behavior of images, any particular choice of subband decompositions, including the widely used wavelet decomposition, doesn't provide the optimal wavelet packet bases that best adapt to the images. On the other hand, it can be shown that there are more than 2^N wavelet packet bases for a given signal , where N is the size of the signal. Therefore, fast search algorithms are essential. A

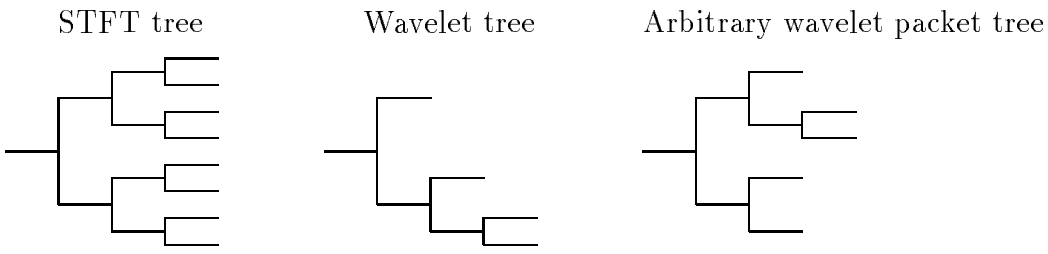

Figure 2.5 Some 1-D WP decompositions of level 3

key point for this class of "best-bases" algorithms is that the cost functional $M(\cdot)$ should be additive, i.e., $M(0) = 0$ and $M(\sum_i X_i) = \sum_i M(X_i)$, so that it can split nicely across the Cartesian products. Thus the search is a fast divide-and-conquer.

An entropy-based algorithm for best-basis selection was proposed [10]. It chooses the basis with the minimum Shannon-Weaver entropy, which is a measure of the energy distribution of an unquantized vector and is not directly related to quantization and coding. Therefore, this criterion does not guarantee the optimality in the rate-distortion sense, which is the measure of the "true" performance of data compression. Ramchandran and Vetterli [34] generalized the above algorithm using a rate-distortion framework. Their algorithm is indeed a combination of the concept of orthonormal tiling of the spatial-frequency plane using wavelet packets and the discipline of rate-distortion optimal bit allocation.

However, neither of the two algorithms indicates how to efficiently construct the rate-distortion curves, whose computational complexity counts for a significant percentage of the total cost of the searching algorithm.

Chapter 3

Transform Coding Schemes

3.1 Quantization

The purpose of quantization is to achieve further compression by representing the transform coefficients with no greater precision than is necessary to achieve the desired image quality. Stated another way, the goal of this processing step is to discard the information which is not visually significant. Quantization is a many-to-one mapping, and therefore is fundamentally lossy. It is the principal source of lossiness in the transform coders.

3.1.1 Scalar Quantizer

We define an N-level scalar (one-dimensional) quantizer (SQ), Q, as a mapping $Q : \mathbf{R} \rightarrow \mathcal{C}$ where $\mathbf{R}$ is the real line and $\mathcal{C} \equiv \{y_1, y_2, \ldots, y_N\} \subset \mathbf{R}$, where $y_1 < y_2 < \ldots < y_N$, is the output set of *reconstruction levels* or the "codebook" with size $|\mathcal{C}| = N$. This mapping is generally a staircase function and the quantization rule is as follows: Define $t_k, k = 1, 2, \ldots, N+1$, as a set of increasing *decision levels*. If the input variable lies in the interval $[t_k, t_{k+1})$, then it is mapped to y_k, the kth reconstruction level.

Uniform Scalar Quantizer

The most common and simplest of all scalar quantizers is the *uniform scalar quantizer* (USQ), in which

1. the decision levels $t_1, t_2, \ldots, t_{N+1}$ are equally spaced on the real line;

2. the reconstruction levels $y_1, y_2, \ldots, y_N$ are the midpoints of the decision intervals.

Llyod-Max Quantizer

Given the fixed number of scalar quantizer levels, the minimum mean-squared error (MMSE) scalar quantizer, or the Llyod-Max quantizer (LMQ) [23, 24], satisfies the following two conditions:

1. the optimum decision levels are halfway between neighboring reconstruction levels;

2. a reconstruction level should be the centroid of the PDF in the appropriate interval.

The LMQ can be designed by using *Llyod iteration algorithm* [17].

It can be shown that, if the quantized variables are entropy coded by a variable-length coding scheme such as the Huffman coding, the USQ gives a better performance than the LMQ. It has been found that the uniform quantizer is quite a good approximation of the "optimum quantizer" based on entropy versus mean square distortion criterion, if the quantization step size is optimized with respect to this criterion.

Entropy-Constrained Scalar Quantizer

The *entropy-constrained scalar quantizer* (ECSQ) [7] minimizes the mean-square error, subject to a constraint on the entropy of the quantized variables. The ECSQ is optimal in the rate-distortion sense. Again, the Llyod iteration algorithm combined with the Lagrange multiplier method can be used to design the ECSQ. The LMQ is thus a special case of ECSQ, when the corresponding Lagrangian multiplier is set to be zero. It is clear that the both the LMQ and the ECSQ are more computationally expensive than the USQ.

3.1.2 Vector Quantizer

We have introduced the VQ in Section 1.3.3. Direct use of the unconstrained VQ suffers from a serious complexity barrier that greatly limits its practical use as a complete and self-contained coding technique. Thus, several techniques have been developed which apply various constraints to the structure of the VQ codebook and yield a correspondingly altered encoding algorithm and design technique. This means that the code vectors cannot have arbitrary locations as points in k-dimensional space but are distributed in a restricted manner that allows easier (lower complexity) search. These methods generally compromise the performance achievable with unconstrained VQ, but often provide very useful and favorable trade-offs between the performance and the complexity. The clever design of such structure-constrained VQ schemes is part of what makes engineering an art that goes beyond the realm of straightforward use of mathematics to analyze engineering problems. The often used VQs include the LBG VQ, the entropy-constrained VQ (ECVQ), the lattice VQ (LVQ), etc. We do not discuss an encyclopedic listing of the VQ techniques in this thesis and leave them to the literature.

3.2 Bit Allocation

In general, the subband images resulting from multi-resolution decomposition are not of equal significance. In most cases, the total number of available bits is inevitably limited. Therefore, it is desirable to allocate more bits to those subband images containing more information so as to code those subband images more accurately than others. The *bit allocation* problem is the task of distributing a given quota of bits to the various subband images to optimize the overall coder performance.

3.2.1 Problem Formulation

Assume that the original image is decomposed into K subband images. Define the overall rate, R, and the overall distortion, D, as functions of the bit allocation vector, $\mathbf{r} = (r_1, r_2, \ldots, r_K) \in \mathbf{Z}^{k+}$, according to $D(\mathbf{r}) = \sum_{i=1}^{K} d_i(r_i)$ and $R(\mathbf{r}) = \sum_{i=1}^{K} r_i$, where $d_i(r_i)$ is the distortion incurred in optimally quantizing the ith subband image with r_i bits. In fact, $d_i(r_i)$ is the distortion-rate function of the ith subband image. The bit allocation problem is

$$\min_{\mathbf{r}} D(\mathbf{r}) \qquad s.t. \qquad R(\mathbf{r}) \leq R_c$$

where R_c is the given number of available bits.

3.2.2 Approximate Optimal Solution

The classical closed-form solution [17, 23, 24] to the bit allocation problem using the high resolution quantization approximations is

$$r_i^* = \frac{R_c}{K} + \frac{1}{2}\log_2 \frac{\sigma_i^2}{\rho^2}$$

where σ_i^2 is the variance of the ith subband image and $\rho^2 = (\prod_{i=1}^{K} \sigma_i^2)^{\frac{1}{K}}$ is the geometric mean of these variances. For simplicity, here we do not consider the nonnegativity and integer constraints for r_i. In practice, re-optimization procedures are required to make r_i a nonnegative integer.

The drawback of this solution is that the assumption of fine quantization is invalid for low bit-rate coding, where the coarse quantization is used.

3.2.3 Rate-Distortion Optimal Solution

The "hard" constrained problem can be converted to a relatively "easy" equivalent unconstrained problem by using the *generalized Lagrange multiplier method* [38]. The following theorem gives the basic idea underlying this solution.

Theorem: For any $\lambda \geq 0$, the solution $\mathbf{r}^*(\lambda)$ to the unconstrained problem

$$\min_{\mathbf{r}}\{D(\mathbf{r}) + \lambda R(\mathbf{r})\}$$

is also the solution to the constrained problem with the constraint $R_c = R(\mathbf{r}^*(\lambda))$.

We need to find that λ for which the constraint $R(\mathbf{r}^*(\lambda))$ is equal or nearly equal to the given rate constraint R_c. Thus, a repeated solution of the unconstrained problem for different but carefully chosen values of λ is needed to locate the best attainable solution for $R(\mathbf{r}^*(\lambda))$ that is less than B but as close to B as possible. In fact, the unconstrained problem can be efficiently solved for a given value of λ by noting that the objective function can be written as the sum:

$$D(\mathbf{r}) + \lambda R(\mathbf{r}) = \sum_{i=1}^{K} [d_i(r_i) + \lambda r_i]$$

and for each i, the ith term can be separately minimized to find the best r_i. The key point of implementing this scheme is the computation of the distortion rate functions for subband images, $d_i(r_i)$.

3.3 Entropy Coding

The *entropy coding* is a technique for encoding discrete symbols into variable length codewords in an invertible fashion so as to reduce the average number of bits per symbol while suffering no loss of fidelity. the entropy codes are often used in conjunction with quantizers to achieve extra compression beyond quantization. The entropy coding has been extensively studied in the literature and the detailed accounts may be found in many books and papers. Thus, we leave this topic to the references. The most popular techniques for entropy coding include the Huffman codes [21], the arithmetic codes [46], the Ziv-Lempel codes [48, 49] and the Rice codes [35].

Chapter 4

Reduction of Quantization Noise

In the entire wavelet decomposition coder, if the perfect-reconstruction filter banks and the lossless entropy coders are used, then the quantizer is the only source that results in error (or noise). Coarse quantization of wavelet coefficients often results in some undesirable artifacts in the reconstructed image, such as the ringing effect, the contouring effect and the block effect, especially at very low bit rate. Thus, the design of a decoder in the image coding can be viewed as a statistical estimation problem of reconstructing the original image signal from the decompressed image, a noisy observation, using the classical model of *signal plus additive noise*. Then, various noise-removal methods can be used to suppress the reconstruction noise (including reduction of those annoying artifacts), and therefore improve the coding performance.

4.1 Statistical Model of Quantization Noise

The quantizer error sequence e_n resulting from applying a quantizer Q to an input x_n has been defined by

$$e_n = y_n - x_n \qquad \text{where} \qquad y_n = Q(x_n).$$

We can write

$$y_n = x_n + e_n;$$

that is, the output of the quantizer can be written as the input signal plus the quantization error or *quantization noise*, e_n. This representation is often called the *additive noise model* of a quantizer and allows us to view the quantization process as the addition of a noise term to the original signal. Note that we have not (yet)

made any assumptions or approximations; the additive noise model is always valid in the sense that the above expression always holds. The model is often carried further, however, by making specific assumptions as to the statistical nature of the quantization noise so that the model can be handled analytically as a classical *signal plus noise* system in communications or signal processing.

4.1.1 Uniform Scalar Quantizer

For a USQ with a quantization step size q, the quantization noise is thus an additive noise lying in the interval $[\frac{-q}{2}, \frac{q}{2}]$. It has been shown [17] that, if the quantization resolution is high, then the quantization noise $\mathbf{e}$ is independent of the input, and can be modeled as an uncorrelated sequence of random variables with a uniform distribution:

$$f_{\mathbf{e}}(e) = \begin{cases} \frac{1}{q} & \text{if } \frac{-q}{2} \leq e \leq \frac{q}{2} \\ 0 & \text{otherwise} \end{cases}$$

with a zero-mean and a variance $\frac{q^2}{12}$. Note that here only the *granular noise* (the noise in the finite range $[\frac{-q}{2}, \frac{q}{2}]$) is considered and the *overload noise* (the noise with a magnitude above $\frac{q}{2}$ that occurs at the boundaries of the quantizer when a USQ is applied to a random variable with an infinite support probability density function) is ignored.

4.1.2 Lloyd-Max Quantizer

For LMQ, it can be shown [24] that *the quantization noise is orthogonal to the output*, i.e., $E[\mathbf{ey}] = 0$. Therefore,

$$E[\mathbf{ex}] = E[\mathbf{e}(\mathbf{y} - \mathbf{e})] = E[\mathbf{ey}] - E[\mathbf{e}^2] = -\sigma_{\mathbf{e}}^2$$

i.e., *the quantization noise and the input are correlated.* Thus, a purely input-independent additive noise model is invalid. A better model for LMQ is therefore one with a less-than-unity gain component and an additive uncorrelated component,

according to $\mathbf{y} = \alpha\mathbf{x} + \mathbf{r}$, where α is the multiplication (gain) factor such that $\alpha\mathbf{x}$ represents the signal part of the quantized signal $\mathbf{y}$ and $\mathbf{r}$ is the uncorrelated additive noise term. By making a proper choice for the gain α the assumption of uncorrelation between $\mathbf{r}$ and $\mathbf{x}$ can be justified.

$$E[\mathbf{xe}] = E[\mathbf{x}(\mathbf{y} - \mathbf{x})] = E[\mathbf{x}(\alpha\mathbf{x} + \mathbf{r} - \mathbf{x})] = (\alpha - 1)\sigma_{\mathbf{x}}^2 + E[\mathbf{xr}].$$

If we choose $\alpha = 1 - \sigma_q^2/\sigma_x^2$, then $E[\mathbf{xr}] = 0$. Note that the value $\sigma_q^2 = E[(\mathbf{y} - \mathbf{x})^2]$ is the mean-squared error of the LMQ. Therefore, α is equal to one minus the normalized quantizer distortion.

4.1.3 Entropy-Constrained Scalar Quantizer

Since the design of the ECSQ using the Lloyd iteration algorithm is purely a numerical recipe, there is no explicit model for the ECSQ noise. However, since the USQ is a good approximation of minimum entropy quantizer if the quantization resolution is high, we conjecture that the ECSQ noise can be modeled as the USQ noise.

4.1.4 Vector Quantizer

The statistical model of the vector quantization noise is much more complicated, for almost all practical design algorithms for VQ are based on some special structure constraints to alleviate the computational cost. Also, most VQ algorithms are purely numerical. Therefore, in general, the VQ noise sequence is strongly correlated to the input, the output and itself.

4.2 De-Noising via Wavelet Shrinkage

There are many methods that have been used to remove the noise:

1. the lowpass filtering, which usually introduces a distortion in the form of blurring;

2. the median filtering, which works very well only for the impulsive noise;

3. the Wiener filtering, which is the minimum mean-square-error method if the statistical models of the signal and the noise are known (However, in image coding, the statistical model of the original image is unknown for the decoder).

Therefore, all of the above methods have their limitations.

Donoho and Johnstone [14] proposed a nonlinear, universal method for reconstructing an unknown signal from noisy data. The method attempts to reject noise by damping or thresholding in the wavelet domain.

Suppose we wish to recover an unknown signal $\mathbf{x}$ from noisy data $\mathbf{y}$,

$$y_i = x_i + \sigma e_i, \qquad i = 0, 1, \ldots, n-1$$

where $e_i \overset{iid}{\sim} \mathcal{N}(0,1)$ is a white Gaussian noise, and σ is the noise level. Let $\hat{\mathbf{x}}$ be the estimate of $\mathbf{x}$. Our goal is to optimize the mean-squared error

$$\frac{1}{n} E[\| \hat{\mathbf{x}} - \mathbf{x} \|_2^2] = \frac{1}{n} \sum_{i=0}^{n-1} E[(\hat{x}_i - x_i)^2]$$

subject to the side condition that with high probability, $\hat{\mathbf{x}}$ is at least as smooth as $\mathbf{x}$. The simple wavelet-domain-soft-thresholding method has three steps:

1. Compute the discrete wavelet transform of the noisy data $\mathbf{y}$, obtaining the wavelet coefficients;

2. Apply the soft-thresholding nonlinearity

$$\eta_t(v) = sgn(v)(|v| - t)_+ = \begin{cases} v - t & \text{for } v > t \\ 0 & \text{for } -t \le v \le t \\ v + t & \text{for } v < -t \end{cases}$$

   to the wavelet coefficients with a specially-chosen threshold t;

3. Perform the inverse wavelet transform, recovering the estimate $\hat{\mathbf{x}}^*$.

It has been shown that $\hat{\mathbf{x}}^*$ achieves almost the minimax mean square error over every one of a wide range of smoothness classes, including many classes where traditional linear estimators do not achieve the minimax rate. Unlike most minimum-mean-square methods (e.g. the Wiener filtering), this method does not exhibit any noise-induced structures, such as the spurious oscillations. Obviously, the computational complexity of this non-linear operator is $O(n)$. In addition, the experimental results showed that this approach provides better visual quality than the procedures based on mean-squared error alone. Thus, it can also be used to remove the artifacts (e.g. blocking) of the transform coding at low bit-rate [18].

We intend to apply the soft-thresholding scheme to suppress the reconstruction noise due to the wavelet-domain quantization, while maintaining the relatively sharp features (e.g. edges) of the original image. From the analysis of the statistical models of the quantization noise, we know that in each subband of the wavelet decomposition, the quantization error is recursively upsampled and filtered through the synthesis filters. Therefore, the reconstruction error is *not* a white Gaussian noise. However, we can still successfully apply this nonlinear scheme for the following reasons. The wavelet transform tends to whiten the data so that the wavelet coefficients of the reconstructed noise might be much more uncorrelated than the noise itself if the wavelets (or the corresponding filter banks) used for compression and de-noising are different, i.e., different frames or tight frames are used for the wavelet expansion. Secondly, it is shown that if the noise is bounded (which is obviously the case for most image processing problems), then the soft-thresholding is optimal. Finally, we propose to apply the adaptive wavelet soft-thresholding to compensate for the data correlation.

Besides the low computational complexity, our universal scheme is completely independent on the coding scheme, therefore independent of the various statistical models of the quantization noise.

The only parameter in this nonlinear operator, the threshold, has a significant effect on the de-noising performance. In practice, the unknown parameter σ, the variance of noise, must be estimated from the data according to Donoho's original method. In general, both under-thresholding and over-thresholding cause more distortion between the de-noised signal and the original signal. Thus, in each subband, there exists an optimal threshold to minimize the distortion.

It is interesting to note that if we choose the RDWT instead of the classical DWT, then we obtain better results for de-noising [26].

4.3 Dithering Method

Due to the coarse quantization, the quantization error and the input image are correlated. This correlation is reflected by some perceptually undesirable image-dependent patterns in the reconstructed image, such as the contouring effects, teh ringing effects, teh block effects, etc. By adding some appropriate high-frequency perturbation signals, such as pseudo-random noise, to an image prior to quantization, it is possible to break up these undesirable patterns. This classical technique is called *dithering* [24]. The basic idea of dithering is, by adding a pseudo-random signal to the image before it is quantized and subtracting the same signal from the quantized value, to effectively replace the image-dependent quantization noise with the image-independent noise which is less annoying to human eyes. Figure 4.1 is the block diagram of a dithered quantizer system.

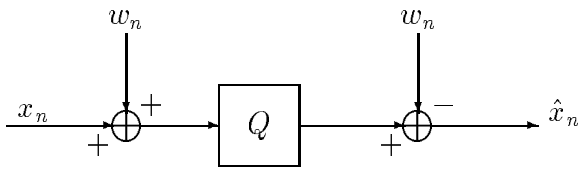

Figure 4.1 A dithered uniform scalar quantizer

Let q denote the step-size of the uniform scalar quantizer Q. It has been shown that if w_n is a white noise uniformly distributed in $[-q/2, q/2]$, then the quantizer output $\hat{x}_n$ can be modeled approximately as

$$\hat{x}_n = x_n + e_n$$

where e_n is a white noise that is independent of x_n and has a uniform distribution in $[-q/2, q/2]$. Such a white noise is less visible and easier to remove using noise-reduction techniques such as the Wiener filtering, the median filtering, the wavelet shrinkage, etc.

4.4 Experimental Results

From our experiments, we obtain the following results:

1. The reconstruction noise in a wavelet compressed image is highly correlated with the original image. Thus, it is harder to remove than the speckle noise in synthetic aperture radar images, which can be modeled as a white Gaussian process. Since the statistical characteristics of the original image and the reconstruction noise is unknown in practice, it is hard to perform de-noising using those classical techniques such as the Wiener filtering. The wavelet shrinkage is a powerful method that can attenuate the reconstruction noise with low computational complexity and independent on the noise statistics.

2. Because in general the reconstruction noise energy is not uniformly distributed across all scales, the scale-adaptive thresholding scheme performs better than the uniform thresholding scheme.

3. While the classical subband coding and the JPEG method suffer from teh contouring effect and the block effect, respectively, at low bit rates, the wavelet transform coding indicates less artifacts and results in better visual quality, though the numerical performance is the same. However, the transform domain

dithering does not improve the performance of the wavelet transform coding. The resulting error after using dithering technique is still correlated to the original image. This is due to the aliasing effect of the QMFs when the quantization error of the wavelet coefficients exists.

Figure 4.2 illustrates a toy example of improving coding performance in terms of both the numerical measure and the visual quality via the dithering and the wavelet shrinkage. The results indicate that the wavelet shrinkage can improve the PSNR by attenuating the signal-independent noise in the reconstructed image, but cannot remove the contouring effect corresponding to the signal-dependent noise (Figure 4.2(b)). If the original image is dithered before quantization, then the noise in the quantized image is approximately signal-independent and hence less visually annoying (Figure 4.2(c)). Furthermore, the wavelet shrinkage can be used as a post-processing scheme to efficiently remove the signal-independent noise and improve both the numerical measure and the perceptual quality (Figure 4.2(d)). In fact, a compression system can be viewed as a generalized quantizer so that the dithering method and the wavelet shrinkage method can be applied as the pre-processing scheme and the post-processing scheme, respectively, to suppress the reconstruction noise and remove the artifacts.

(a) (b)

(c) (d)

Figure 4.2 De-Noising via dithering and wavelet shrinkage: (a) spatially uniform quantized image at 3 bpp (PSNR = 28.62dB); (b) wavelet shrinkage de-noised image (PSNR = 30.15dB); (c) dithered and spatially uniform quantized image (PSNR = 28.82dB); (d) de-noised image via dithering and wavelet shrinkage (PSNR = 31.85dB)

Chapter 5

Simultaneous Noise Reduction and Data Compression

In this chapter, we propose a novel method for simultaneous noise reduction and data compression based on shrinking, quantizing and coding the wavelet packet coefficients [43]. A fast dynamic programming pruning algorithm is used to efficiently choose the best basis from the entire library of the admissible wavelet packet bases, which jointly optimizes the bit allocation and the quantization in the rate-distortion sense. Soft-thresholding in wavelet domain can significantly suppress noise, e.g., the speckles of the synthetic aperture radar (SAR) images, while maintaining bright reflections for subsequent detection and recognition. Our method can be viewed as not only a generalization but also a combination of the methods used by Guo et al [20] and Werness et al [44].

5.1 Background

Synthetic aperture radar is an active coherent all-weather imaging system that operates in the microwave region of the spectrum. This imagery is well suited to the task of remote ground mapping in many applications, such as surveillance, oceanography, glaciology, and agriculture. There are two problems in the practical SAR applications. One is the speckle phenomena. The speckle results from the necessity of creating the image with the coherent radiation. A fully developed speckle pattern appears chaotic and unordered. When the detail information in the image is important, the speckle can be viewed as a noise that causes degradation of the image. Therefore, speckle reduction is an essential procedure before the procedures of automatic target de-

tection and recognition. Another problem is the large amount of SAR data. The data collected and processed by a SAR system are inherently complex. Thus, data compression is desirable for quick transmission of the collected information.

The wavelet transform is a relatively new technique for multi-resolution decomposition of images and is widely used in both noise reduction and data compression of SAR images [20, 44]. Thus, it is very efficient to combine the procedures of noise reduction and data compression in a single process of decomposition and reconstruction. As a generalization of the wavelet basis, the wavelet packet (WP) bases, which is a rich family of the orthonormal bases, can be expected to be more suited to match the non-stationary statistics of the images. Therefore, it is desirable to fast select an optimal wavelet packet basis under some criterion and to achieve better performance of de-noising and compression.

5.2 Bit Allocation and Quantization

The optimal (in rate-distortion sense) bit-allocation scheme is embedded in the best WP bases selection algorithm. In our algorithm, we use the uniform scalar quantizer (USQ). There are several advantages of using the USQ:

1. It is very easy and fast to implement a USQ alleviating the computational complexity for the rate-distortion functions. The trade-off between the computational complexity and the performance is very crucial in the applications of compression of very large data sets.

2. It can be shown that, if the quantized variables are entropy coded by a variable-length coding scheme such as the Huffman coding, then the USQ gives a better performance than the Lloyd-Max quantizer [23]. It has been found that the uniform quantizer is a quite good approximation of the "optimum quantizer" based on entropy versus mean square distortion criterion, if the quantization step size is optimized with respect to this criterion.

3. It is easy to combine the procedure of the wavelet domain soft-thresholding, which helps both de-noising and data compression, with the procedure of the uniform scalar quantization.

In our application, each subband is assigned a finite set of quantizers with different quantization steps. Hence, the rate-versus-distortion functions of the subbands can be easily computed. We use the mean-squared error (MSE) and teh first-order entropy as the measures of the distortion and the rate, respectively. The weighted MSE criterion related to the model of the human vision system (HVS) can also be used to further improve the visual quality of the reconstructed images.

5.3 Basic Idea

Ideally, we want to jointly optimize the performance of compression and detection. However, this is an intractable optimization problem due to the non-existance of a good measure for both data compression and noise reduction. So, our goal here is to optimally compress the SAR image in rate-distortion sense while maintaining the performance of de-noising.

Assume that the optimal wavelet packet subtree from the node n "onwards" to the full tree-depth is known. Then by Bellman's optimality principle [5], we know that all possible paths passing through the node n must invoke this same optimal "finishing" path. At each non-leaf node of the tree, there are two contenders for the "surviving path", the parent and its children, with the winner having the lower Lagrangian cost. According to this sub-optimality, we can apply a dynamic programming to construct the optimal subtree starting from the leaf nodes upwards. When we reach the root node, the best basis with the minimum Lagrangian cost is known.

Figure 5.1 describes the basic structure of our algorithm. We first perform the de-noising. Then, we initialize a slope value λ and compute the rate-distortion relations and the minimum Lagrangian costs of all subbands with the admissible quantizer sets.

We then start from the leave nodes of the entire WP tree and compare each node with its four child nodes. For each node, if its Lagrangian cost is less than the sum of those costs of its four child nodes, we mark this node as a "merge" node; otherwise, we mark this node as a "split" node and update its information (Lagrangian cost and the corresponding rate and distortion) with those of its child nodes. We recursively do the comparisons until we reach the root node. Then, we can curve out the optimal sub-tree for this λ by pruning all the sub-trees rooted at those "merge" nodes and the corresponding rates and distortions. We can repeat the above process using a bisection search to find the proper λ so that the resulting rate is identical to the desired bit budget.

Since we include the de-noising procedure in our coding algorithm, which obviously causes some error between the de-noised image and the original image, we use the MSE between the coefficients of the de-noised image and those of the decompressed image as the distortion.

5.4 A Pruning Method

The computational complexity of computing the subband distortion-rate curves counts for a significant percentage of the total cost of our algorithm. We propose a pruning

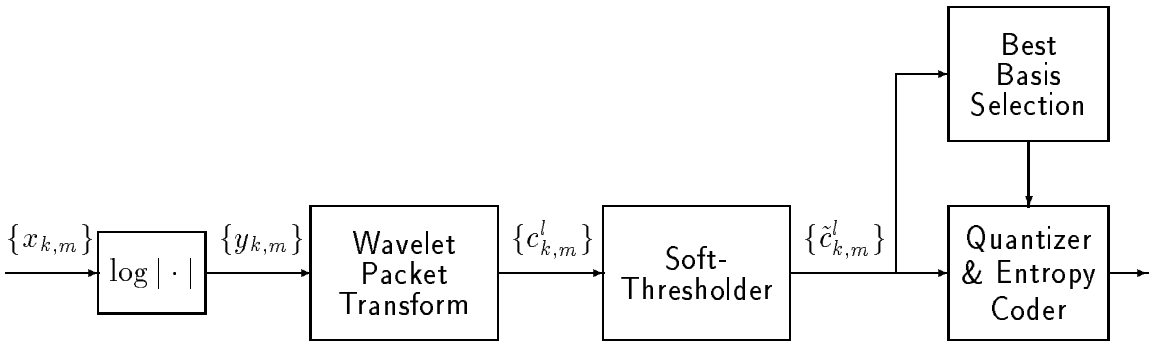

Figure 5.1 Block diagram of the algorithm for simultaneous de-speckling and compression

method based on the fact that the distortion-rate function is monotonically decreasing and convex, to fast search for the minimum Lagrangian cost in each subband image.

Assume that $\{R_i\}$ is a set of admissible bit-rates for a certain subband image, satisfying

$$R_1 < R_2 < \cdots < R_{k-1} < R_k < R_{k+1} < \cdots.$$

Define the Lagrangian cost $J(R_i) = D(R_i) + \lambda R_i$, where $D(\cdot)$ is the distortion-rate function and λ is the Lagrangian multiplier. The goal is to find

$$R^* = \arg\min_{R_i} J(R_i) = \arg\min_{R_i}[D(R_i) + \lambda R_i].$$

We claim the following two pruning conditions to accelerate the search:

Condition 1: $J(R_{k-1}) < J(R_k) \Rightarrow J(R_{k-1}) < J(R_{k+1})$;

Condition 2: $J(R_k) > J(R_{k+1}) \Rightarrow J(R_{k-1}) > J(R_{k+1})$.

We will give the proof in the Appendix.

5.5 Algorithm

The following procedure is used to de-speckle and compress the images.

Step 1: Take the logarithmic operation on the magnitude of the complex SAR image $\{x_{k,m}\}$ to obtain $\{y_{k,m}\}$.

Step 2: Perform a full wavelet packet decomposition (up to a desired level L) on the image y to obtain the coefficients $\{c^l_{k,m}\}$ for the entire WP library.

Step 3: Apply soft-thresholding to $\{c^l_{k,m}\}$ and obtain the de-noised WP coefficients $\{\tilde{c}^l_{k,m}\}$.

Step 4: For each subband $n_{i,j}^l$, quantize the de-noised WP coefficients using all admissible quantizers, record the resulting rates $\{R_{q\in Q}(n_{i,j}^l)\}$, and the corresponding distortions $\{D_{q\in Q}(n_{i,j}^l)\}$, and form the rate-versus-distortion curve.

Step 5: Initialize an operating slope value λ, populate all the nodes $\{n_{i,j}^l\}$ of the WP tree with their minimum Lagrangian costs $J_{i,j}^l(\lambda)$, i.e,

$$J_{i,j}^l(\lambda) = min_{q\in Q}[D_q(n_{i,j}^l) + \lambda R_q(n_{i,j}^l)].$$

Step 6: Initialize $l \leftarrow L$. Let $q^*(n_{i,j}^l) = argmin(J_{i,j}^l(\lambda))$. Initialize

$$\tilde{J}_{i,j}^l \leftarrow J_{i,j}^l, \qquad \tilde{R}_{i,j}^l \leftarrow R_{i,j}^l \qquad (\text{where } R_{i,j}^l = R_{q^*(n_{i,j}^l)}(n_{i,j}^l)),$$
$$\tilde{D}_{i,j}^l \leftarrow D_{i,j}^l \qquad (\text{where } D_{i,j}^l = D_{q^*(n_{i,j}^l)}(n_{i,j}^l)).$$

Step 7: $l \leftarrow l-1$. If $l < 0$, go to Step 10.

Step 8: $\forall i,j = 1,2,\cdots,2^l$ at the lth tree level:

$$\begin{aligned}
\text{if} \quad & (J_{i,j}^l(\lambda) < \tilde{J}_{2i-1,2j-1}^{l+1} + \tilde{J}_{2i-1,2j}^{l+1} + \tilde{J}_{2i,2j-1}^{l+1} + \tilde{J}_{2i,2j}^{l+1}) \\
\text{then} \quad & \{\text{split}(n_{i,j}^l) \leftarrow \text{FALSE}; \\
& \tilde{R}_{i,j}^l \leftarrow R_{i,j}^l; \tilde{D}_{i,j}^l \leftarrow D_{i,j}^l; \tilde{J}_{i,j}^l \leftarrow J_{i,j}^l\} \\
\text{else} \quad & \{\text{split}(n_{i,j}^l) \leftarrow \text{TRUE}; \\
& \tilde{R}_{i,j}^l \leftarrow \tilde{R}_{2i-1,2j-1}^{l+1} + \tilde{R}_{2i-1,2j}^{l+1} + \tilde{R}_{2i,2j-1}^{l+1} + \tilde{R}_{2i,2j}^{l+1}; \\
& \tilde{D}_{i,j}^l \leftarrow \tilde{D}_{2i-1,2j-1}^{l+1} + \tilde{D}_{2i-1,2j}^{l+1} + \tilde{D}_{2i,2j-1}^{l+1} + \tilde{D}_{2i,2j}^{l+1}; \\
& \tilde{J}_{i,j}^l \leftarrow \tilde{J}_{2i-1,2j-1}^{l+1} + \tilde{J}_{2i-1,2j}^{l+1} + \tilde{J}_{2i,2j-1}^{l+1} + \tilde{J}_{2i,2j}^{l+1}\}.
\end{aligned}$$

Step 9: Go to Step 7.

Step 10: Perform a tree-walk, starting from the root of the WP tree, $n_{0,0}^0$, and carve out the optimal subtree corresponding to the best wavelet packet basis with its associated optimal choice of quantizer set. $\tilde{R}_{0,0}^0$ and $\tilde{D}_{0,0}^0$ are the rate and the distortion of the optimal subtree for the given λ, respectively.

Step 11: Iterate from Step 5 to Step 10 towards the optimal operating slope λ^* until the given target bit rate is achieved.

From the above implementation, we find that the computational complexity is mainly determined by the size of the nested loops. In most applications, we only decompose the image up to $4 \sim 6$ levels. The number of quantizers of each subband is also a small constant compared to the size of image. Also, the number of iterations for the bisection search of a convex curve is relatively small. Therefore, the total computational complexity of our algorithm is $O(N)$, where N is the size of image.

5.6 Experimental Results

We apply the above algorithm to the application of synthetic aperture radar (SAR) images. The data used here were collected near Stockbridge, NY, by the Lincoln Laboratory MMW SAR. We apply our algorithm to a polarimetric whitening filtered (PWF) image. We choose four types of clutter regions in the image: trees, scrub, grass and shadows. The following four statistics are used to evaluate the performance of our algorithm:

Standard-deviation-to-mean ratio (s/m) : The quantity s/m (both in power) is a measure of the image speckle in a homogeneous region. We computed the s/m ratio for each type of the clutter region to quantify the speckle reduction capacity of our algorithm.

Log standard deviation: The standard deviation of the clutter data (in dB). This is an important quantity that directly affects the target detection performance of a standard two-parameter constant-false-alarm-ratio (CFAR) algorithm.

Target-to-clutter ratio (t/c): The difference between the target and the clutter means (in dB). It measures how the target stands out of the surrounding clutter.

Deflection ratio: This is the two-parameter CFAR detection statistic.

$$M = \frac{y - \hat{\mu}_y}{\hat{\sigma}_y}$$

where y is the scalar pixel value of the cell, $\hat{\mu}_y$ is the estimated mean of y, and $\hat{\sigma}_y$ is the estimated standard deviation of y. After speckle reduction, M should be higher at the known reflector points and lower elsewhere.

Table 5.1, Table 5.2 and Table 5.3 illustrate the numerical results of those four values for the original and the processed images at four typical regions. The large reductions of standard-deviation-to-mean ratio (s/m) and log-standard-deviation indicate that a significant amount of speckle has been removed. Figure 5.2 shows the perceptual results. We can see from the images that speckle is greatly reduced while sharp features are maintained. After speckle reduction, the deflection ratio should be higher at known reflector points and lower elsewhere. As shown in Table 5.3, the deflection ratio is much higher in the wavelet processed images than that in the PWF processed image. Elsewhere in the image, the deflection ratio values are roughly the same for both methods. This strongly indicates the advantage of our method, and suggests a big improvement in detection performance. Also, cleaner images suggest potential improvements for classification and recognition. Comparing the de-noised and uncompressed image to the de-noised and decompressed image, we achieve a very high peak-signal-to-noise-ratio (PSNR) and the good visual quality. Thus, we have compressed the original image at very low bit rate, 0.2 bits per pixel (bpp), while maintaining the high performance of de-noising.

Table 5.1 s/m for clutter data

	Trees	Scrub	Grass	Shadow
Original Image	1.3025	0.8229	0.6545	0.7020
De-noised Image	0.9245	0.4478	0.3098	0.3611
De-noised & Decompressed Image	0.9251	0.4470	0.3101	0.3618

Table 5.2 $log - std$ for clutter data

	Trees	Scrub	Grass	Shadow
Original Image	4.9398	3.4301	2.9543	2.9008
De-noised Image	3.8335	1.8822	1.3271	1.3081
De-noised & Decompressed Image	3.8337	1.8825	1.3268	1.3077

Table 5.3 Target-to-clutter ratio(t/c) and Deflection ratio for clutter data

	Target-to-clutter ratio(t/c)	Deflection ratio
Original Image	34.0269	11.1842
De-noised Image	30.1037	18.2339
De-noised & Decompressed Image	30.0989	18.2401

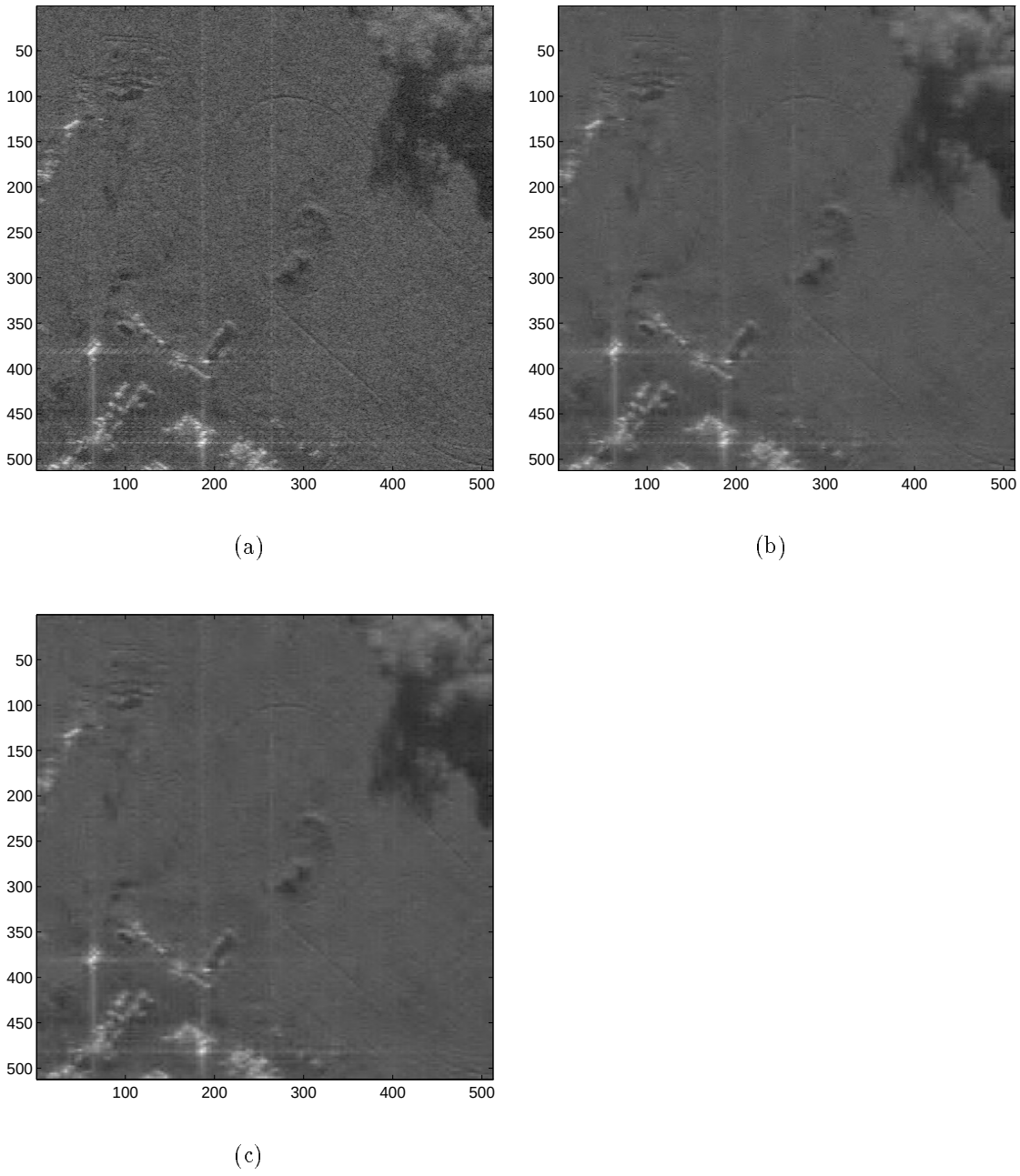

Figure 5.2 SAR images of a farm area (PWF processed): (a) Original image; (b) De-noised image; (c) De-noised and decompressed image (Rate:0.2bpp; PSNR:39.6dB).

Chapter 6

Summary

Perhaps the biggest potential of wavelets has been claimed for the data compression. Since the discrete wavelet transform is essentially a subband filtering system, and since teh subband coders have been successful in speech and image data compression, it is obvious that the wavelets have found its immediate application in the compression problems. It is also clear that the drastic improvements of compression will not be achieved so easily simply because the wavelets are used. However, the wavelets bring new ideas and new insights. In this respect, the use of the wavelet decomposition in connection with *other* techniques [1, 3, 37, 34] are the promising compression techniques which make good use of the elegant theory of wavelets.

Though the wavelet-based compression techniques have outperformed the only international standard for image data compression, JPEG, it will probably take a long time to standardize these new techniques, or it might be impossible for the wavelet techniques to replace the JPEG. However, the wavelet decomposition coding will still be a very important member in the entire family of the data compression techniques.

Appendix A

Rate Distortion Theory

The rate distortion theory [6] is based on two central concepts: *mutual information* and *distortion*. The mutual information is a symmetric measure of the information that symbols u and v convey about each other. The average mutual information between the ensembles U and V, representing the amplitude-discrete vector-valued random variables u and v with the joint probability density function $p_{UV}(u,v)$ and the marginal probability density functions $p_U(u)$ and $p_V(v)$ is defined as

$$I(U;V) = \sum_u \sum_v p_{UV}(u,v) \log \frac{p_{UV}(u,v)}{p_U(u)p_V(v)}.$$

The distortion measure $d(u,v)$ is a nonnegative scalar quantity that reflects the fidelity of the reproduction of an original image u by an image v. The average distortion then is

$$D_{UV} = \sum_u \sum_v p_{UV}(u,v)d(u,v).$$

The *rate distortion function* (RDF) is defined as

$$R(D) = \inf_{p_{UV}} \{I(U;V) : D_{UV} \leq D\},$$

i.e., $R(D)$ is the greatest lower bound of the average mutual information, subject to the constraint that the average distortion D_{UV} may not exceed D. It can be shown that $R(D)$ is a monotonically decreasing convex function.

The RDF is a performance bound that no source coder can beat. Conversely, rate distortion theory shows that a source coder with a performance arbitrarily close to the RDF exists.

Appendix B

Test Image

Figure B.1 Original Lenna image (512x512, 256 grey-scale)

Appendix C

Proofs

Here, we give the proof of the two pruning conditions in Chapter 5.

Proposition Assume that $D : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a monotonically decreasing and convex function. Define $J(R) = D(R) + \lambda R$, where λ is a positive constant. Assume that $0 < R_1 < R_2 < R_3$. Then,

$$J(R_1) < J(R_2) \Rightarrow J(R_1) < J(R_3);$$

$$J(R_2) > J(R_3) \Rightarrow J(R_1) > J(R_3).$$

Proof:

Define $\alpha = (R_3 - R_2)/(R_3 - R_1)$. Then,

$$R_2 = \alpha R_1 + (1-\alpha) R_3 \qquad \text{and} \qquad \frac{R_2 - R_1}{R_3 - R_2} = \frac{1-\alpha}{\alpha}.$$

Since D is a monotonically decreasing and convex function, we have

$$D(R_1) > D(R_2) > D(R_3) \qquad \text{and} \qquad \alpha D(R_1) + (1-\alpha) D(R_3) > D(R_2).$$

Therefore,

$$\frac{D(R_1) - D(R_2)}{D(R_2) - D(R_3)} > \frac{1-\alpha}{\alpha}, \qquad \text{i.e.} \qquad \frac{D(R_1) - D(R_2)}{D(R_2) - D(R_3)} > \frac{R_2 - R_1}{R_3 - R_2},$$

or

$$\frac{D(R_1) - D(R_2)}{R_2 - R_1} > \frac{D(R_2) - D(R_3)}{R_3 - R_2}.$$

Since $J(R_1) < J(R_2)$, we have

$$\frac{D(R_1) - D(R_2)}{R_2 - R_1} < \lambda.$$

Thus,

$$\frac{D(R_2) - D(R_3)}{R_3 - R_2} < \lambda, \qquad \text{or} \qquad J(R_2) < J(R_3),$$

which implies $J(R_1) < J(R_3)$.

Similarly, we can prove the second condition. □

Bibliography

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using the wavelet transform. *IEEE Trans. on Image Processing*, 2(2), April 1992.

[2] M. Barlaud. *Wavelets in Image Communication.* Elsevier, P.O. Box 211, 1000AE Amsterdam, The Netherlands, 1994.

[3] M. Barlaud, P. Solé, T. Gaidon, M. Antonini, and P. Mathieu. Pyramidal lattice vector quantization for multiscale image coding. *IEEE Trans. on Image Processing*, 3(4):367–381, July 1994.

[4] M. F. Barnsley and L. Hurd. *Fractal Image Compression.* AK Peters, Wellsley, MA 02181, 1993.

[5] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, 1957.

[6] T. Berger. *Rate Distortion Theory: A Mathematical Basis For Data Compression.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[7] P. A. Chou, T. L. Lookabaugh, and R. M. Gray. Entropy-constrained vector quantization. *IEEE Trans. ASSP*, 37(1):31–42, January 1989.

[8] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, 1992.

[9] R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser. Signal processing and compression with wave packets. preprint.

[10] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inform. Theory*, 38(2):1713–1716, 1992.

[11] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. In *Proc. Int. Conf. Information Science and Systems*, pages 443–446, Patras, Greece, August 1976.

[12] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992. Notes from the 1990 CBMS-NSF Conference on Wavelets and Applications at Lowell, MA.

[13] R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Inform. Theory*, 38(2):719–746, March 1992.

[14] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 1994. Also Tech. Report, Department of Statistics, Stanford University,1992.

[15] P. M. Farrelle. *Recursive Block Coding for Image Data Compression.* Springer-Verlag, New York, NY, 1990.

[16] Y. Fisher. *Fractal Image Compression: Theory and Application.* Springer-Verlag, New York, NY, 1995.

[17] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[18] R. A. Gopinath, M. Lang, H. Guo, and J. E. Odegard. Wavelet-based post-processing of low bit rate transform coded images. Technical Report CML TR94-15, Computational Mathematics Laboratory, Rice University, Houston, TX, February 1994. accepted for ICIP 1994, Austin, TX.

[19] R. M. Gray. Vector quantization. *IEEE ASSP Magazine*, 1:4–29, April 1984.

[20] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. Selesnick, and C. S. Burrus. Speckle reduction via wavelet shrinkage with application to SAR based ATD/R. In *SPIE Math. Imagaging: Wavelet Applications in Signal and Image Processing*, volume 2303, San Diego, CA, July 1994. Also Tech report CML TR94-03, Rice University, Houston, TX.

[21] D. A. Huffman. A method for the construction of minimum redundancy codes. *Proceedings of the IRE*, 40:1098–1101, 1952.

[22] A. E. Jacquin. Fractal image coding: A review. *Proc. IEEE*, 81:1451–1466, October 1993.

[23] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[24] N. S. Jayant and P. Noll. *Digital Coding of Waveforms.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1st edition, 1984.

[25] J. Kovačević and M. Vetterli. Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n. *IEEE Trans. Inform. Theory*, 38(2):533–555, March 1992.

[26] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr. Nonlinear processing of a shift-invariant DWT for noise reduction. In *SPIE conference on wavelet applications*, volume 2491, Orlando, FL, April 1995.

[27] A. S. Lewis and G. Knowles. Image compression using the 2-d wavelet transform. *IEEE Trans. on Image Processing*, 1(2):244–250, April 1992.

[28] J. Liang and T. W. Parks. A two-dimensional translation invariant wavelet representation and its applications. In *Proc. Int. Conf. Image Processing*, volume 1, pages 66–70, Austin, TX, November 1994. IEEE.

[29] J. S. Lim. *Two-Dimensional Signal and Image Processing.* Prentice Hall, Englewood Cliffs, NJ, 1990.

[30] S. Mallat. A theory for multiscale signal decomposition: The wavelet representation. *IEEE Trans. on Pattern and Machine Intelligence*, 11(7):674–693, 1989.

[31] N. M. Nasrabadi and R. A. King. Image coding using vector quantization: A review. *IEEE Trans. Commun.*, 36:957–971, August 1988.

[32] A. N. Netravali and B. G. Haskell. *Digital Pictures: Representation and Compression.* Plenum Press, New York, NY, 1988.

[33] W. B. Pennebaker and J. L. Mitchell. *JPEG - Still Image Data Compression Standard.* Van Nostrand Reinhold, New York, 1993.

[34] K. Ramchandran and M. Vetterli. Best wavelet packets in a rate-distortion sense. *IEEE Trans. on Image Processing*, 1993. 160-175.

[35] R. F. Rice. Some practical universal noiseless coding techniques. Technical report, JPL Publication 79-22, 1979.

[36] T. Senoo and B. Girod. Vector quantization for entropy coding of image suubands. *IEEE Trans. on Image Processing*, 1(4):526–532, October 1992.

[37] J. M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. SP*, 41(12):3445–3462, December 1993.

[38] Y. Shoham and A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. ASSP*, 36:1445–1453, September 1988.

[39] M. J. Smith and T. P. Barnwell III. Exact reconstruction techniques for tree-structured subband coders. *IEEE Trans. ASSP*, 34:434–441, 1986.

[40] M. Vetterli. Multi-dimensional sub-band coding: Some theory and algorithms. *Signal Processing*, 6(2):97–112, February 1984.

[41] M. Vetterli. Multirate filter banks for subband coding. Technical Report CU/CTR/TR 165-90-02, Center for Telecommunications Research, Columbia Unversity, New York, NY, 1989. Also in *Subband Coding of Images*, edited by John Woods and published by Kluwer.

[42] G. K. Wallace. The JPEG still picture compression standard. *Commun. ACM*, 34(4):31–44, April 1991.

[43] D. Wei, H. Guo, J. E. Odegard, M. Lang, and C. S. Burrus. Simultaneous speckle reduction and data compression using best wavelet packet bases with application to SAR based ATD/R. In *SPIE conference on wavelet applications*, volume 2491, Orlando, FL, April 1995.

[44] S. A. Werness, S. C. Wei, and R. Carpinella. Experiments with wavelets for compression of SAR data. *IEEE Trans. Geoscience and Remote Sensing*, 32(1):197–201, January 1994.

[45] P. H. Westerink, D. E. Boekee, J. Biemond, and J. W. Woods. Subband coding of images using vector quantization. *IEEE Trans. Commun.*, 36:713–719, June 1988.

[46] I. H. Witten, R. M. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Communications of the ACM*, 30:520–540, 1987.

[47] J. W. Woods and S. D. O'Neil. Subband coding of images. *IEEE Trans. ASSP*, 34, October 1986.

[48] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inform. Theory*, 23:337–343, 1977.

[49] J. Ziv and A. Lempel. Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory*, 24:530–536, 1978.