Chapter 2

Wavelet-based Image Compression

Transform based image compression schemes first involve the transformation of spatial information in to another domain. For example, the DCT transforms an image into the frequency domain. The goal of the transformation is a compact, complete representation of the image. The transform should decorrelate the spatially distributed energy into fewer data samples such that no information is lost. Orthogonal transforms have the feature of eliminating redundancy in the transformed image. Compression occurs in the second step when the transformed image is quantized (i.e. when some data samples—usually those with insignificant energy-levels—are discarded).

The inverse transform reconstructs the compressed image in the spatial domain. Since the quantization process is not invertible, the reconstruction cannot perfectly recreate the original image. This type of compression is called *lossy*; Figure 2.1 shows a block diagram of a lossy compression scheme. The remainder of this chapter explores the scalar wavelet transform, its applicability to image compression and a quantization technique that is used for image compression in this thesis.

In transform based image compression, entropy coding typically follows the quantization stage. Entropy coding minimizes the redundancy in the bit stream and is fully invertible at

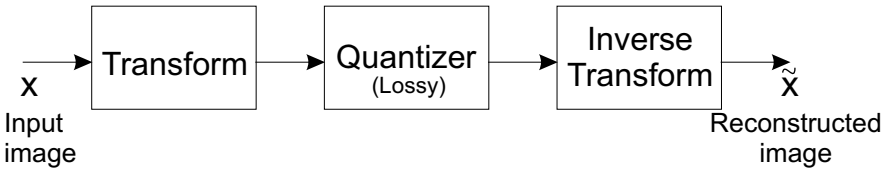

Figure 2.1: Block diagram of a transform based lossy compression system.

the decoding end. So, it is lossless and usually gives about 0.4–0.6 dB gain in the PSNR [18]. Since we are primarily concerned with evaluating the performance of the wavelet transform in compression, this thesis does not study the effects of entropy coding. A detailed description of widely used entropy coding schemes is presented in [32].

2.1 Multiresolution Analysis and the Discrete Wavelet Transform

Multiresolution analysis is the heart of the theory behind wavelets; the analysis enables us to represent a square integrable function (one that exists in L^2) as a linear sum of the weighted orthogonal bases (at different resolutions) of the vector subspaces of L^2. Given a vector space $V_0 \in L^2$, a scaling function $\phi(t)$ is defined such that its translates $\{\phi(t-k),\ k \in \mathcal{Z}\}$ span V_0. Let us define a larger space V_{-1} to be the span of the scaled dilates $2^{1/2}\phi(2^1 t - k)$, $k \in \mathcal{Z}$. Then V_0 is said to be contained in space V_{-1} so that $V_0 \subset V_{-1}$. In general, if we define a space V_j to be the span of functions $\{2^{-j/2}\phi(2^{-j}t-k),\ k \in \mathcal{Z}\}$, we have a hierarchical nesting of spaces given by equation (2.1) and Figure 2.2.

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots. \tag{2.1}$$

Since $V_j \in V_{j-1}$, the bases of V_j and V_{j-1} are not orthogonal. In fact, they are related by the following 2-scale equation,

$$\phi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} h_k \, \phi(2t-k) \tag{2.2}$$

where, $\{h_k\}$ is the set of coefficients corresponding to the projection of the bases $\phi_{j-1,k}(t)$ (the subscripts $(j-1)$ and k denote the scale and shift, respectively) of space V_{j-1} onto the bases $\phi_{j,k}(t)$ of space V_j

$$h_k =< \phi(t), \phi(2t-k) > \quad k \in \mathcal{Z}. \tag{2.3}$$

If we denote that part of V_{j-1} which is not V_j by W_j, we can write

$$V_{j-1} = V_j \oplus W_j \tag{2.4}$$

where,

$$V_j \setminus W_j = \{\emptyset\}. \tag{2.5}$$

We introduce the wavelet function $\psi(t)$ such that its dilated translates $\{2^{-j/2}\psi(2^{-j}t - k)$, $k \in \mathcal{Z}\}$ span the space W_j. Then, $\psi_{j,k}(t)$, (the subscripts j and k denote the scale and shift, respectively) form the bases for W_j. The bases of W_j and V_{j-1} are related by the following equation,

$$\psi(t) = \sqrt{2} \sum_{k=-\infty}^{\infty} g_k \, \phi(2t-k) \tag{2.6}$$

where, $\{g_k\}$ is the set of coefficients corresponding to the projection of the bases $\phi_{j-1,k}(t)$ of space V_{j-1} onto the bases $\psi_j(t)$ of space W_j. This is shown in equation (2.7).

$$g_k =< \psi(t), \phi(2t-k) > \quad k \in \mathcal{Z}. \tag{2.7}$$

If the subspace V_j is expressed as the union of its orthogonal subspaces, equation (2.4) can be expressed as follows.

$$V_{j-1} = V_n \oplus W_n \oplus W_{n-1} \oplus W_{n-2} \oplus \cdots \oplus W_j. \tag{2.8}$$

Figure 2.2 illustrates this concept of nested spaces. From equation (2.8), the scaling functions $\{\phi_{n,k}(t)$, $k \in \mathcal{Z}\}$ (which span the subspace V_n) and the wavelet functions $\{\psi_{m,k}(t)$, where, $m = n, n-1, \cdots j$ and $k \in \mathcal{Z}\}$, (which span the subspaces W_n, $W_{n-1} \cdots W_j$, respectively) form a complete set of orthonormal bases for the subspace V_{j-1}.

Projecting a function onto these hierarchical subspaces allows us to study its characteristics at different levels of resolution. This *multiresolution analysis* (MRA) allows a function with

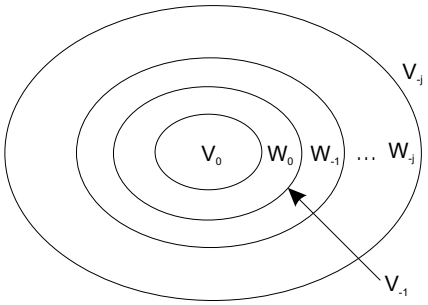

Figure 2.2: Nested spaces for multiresolution analysis.

resolution $(j-1)$ to be represented as the sum of its approximations at some coarsest scale n and its details at scales $(n, n-1, \cdots j)$. In other words, the synthesis of $x(t) \in V_{j-1}$, is given by equation (2.9). This equation is also called the *inverse discrete wavelet transform* (IDWT).

$$x(t) = \sum_k a_{n,k}\phi_{n,k}(t) + \sum_{m=j}^{n} \sum_k b_{m,k}\phi_{m,k}(t)$$

$$= \sum_k a_{n,k}\phi_{n,k}(t) + \sum_k b_{n,k}\phi_{n,k}(t) + \sum_k b_{n-1,k}\phi_{n-1,k}(t) + \cdots \sum_k b_{j,k}\phi_{j,k}(t). \qquad (2.9)$$

In equation (2.9), the coefficients $a_{n,k}$ and $b_{m,k}$ $(m = n, n-1, \cdots j)$ are given by equation (2.10). This equation is referred to as the *discrete wavelet transform* (DWT) and corresponds to the analysis of the input signal $f(t)$.

$$a_{n,k} =< x(t), \phi_{n,k}(t) >$$

$$b_{n,k} =< x(t), \psi_{n,k}(t) > \qquad (2.10)$$

$a_{n,k}$ are generated by the $\phi_{n,k}(t)$; so they correspond to a coarse resolution in the MRA and are called *approximation coefficients* at scale n. The coefficients $b_{m,k}$ are generated by the $\psi_{n,k}(t)$; so they represent the *details* at scale m and are called *wavelet coefficients*.

The DWT naturally lends itself to image compression for the following reasons.

- Many natural images exhibit fractal or self-similar behavior. The DWT is an efficient representation for these signals since all wavelet coefficients can be computed from the wavelet coefficients at one scale.
- The DWT is well suited to progressive encoding due to its pyramid decomposition (Section 2.3).
- The space-scale decomposition in the DWT allows for the incorporation of *HVS* characteristics.

2.2 Implementation of the Discrete Wavelet Transform

The wavelet transform is implemented as a filter bank. Mallat's fast wavelet transform shows how the approximation and wavelet coefficients are calculated by iterating the filter bank [12]. The perfect reconstruction orthogonal wavelet filter bank is shown in Figure 2.3. The coefficients $\{h_k\}$ and $\{g_k\}$ from equations (2.2) and (2.6) correspond to the lowpass and highpass filters in the bank. The input samples $x(n)$ correpond to the projection of $x(t)$ onto some finest scale. The lowpass output from the analysis bank is decomposed iteratively; this gives an octave band decomposition of the input signal. The synthesis bank performs the inverse wavelet transform and restores the input to within a delay.

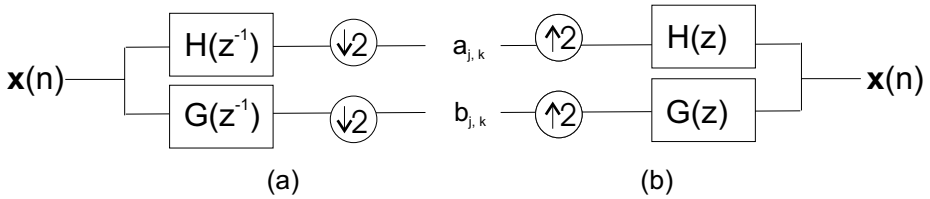

Figure 2.3: The perfect reconstruction scalar wavelet filter bank (a) analysis section, (b) synthesis section.

Orthogonality and symmetry of the basis functions (and the corresponding filters h_k and g_k) are two desirable characteristics in the wavelet transform. Orthogonality decorrelates the

transform coefficients—thereby minimizing redundancy. Symmetry provides linear phase and permits simple symmetric boundary extension techniques that minimize border artifacts (this is discussed further in Section 2.4). For scalar wavelets, the filter bank cannot simultaneously have orthogonality and linear phase [22]. The only exception to this rule is the trivial Haar wavelet with a filter length of two. So orthogonal scalar wavelets cannot be symmetric. By relaxing the orthogonality condition, symmetric wavelet bases can be constructed. The biorthogonal wavelets are one such class of wavelets. The image compression literature indicates that symmetry is a stronger determinant of compressed image quality than orthogonality [1, 30].

In our experiments, we examine image compression performance for the 8-tap orthogonal Daubechies wavelet D_8 [23] and the popular (9/7)-tap $B_{9/7}$ [1] and (22/14)-tap $B_{22/14}$ [30] biorthogonal wavelets.

2.2.1 Vanishing Moments of Wavelets

As mentioned earlier, a desirable wavelet characteristic for image compression is good energy compaction. The number of *vanishing moments* that a wavelet possesses provides a means for evaluating its energy compaction efficiency. The greater the number of vanishing moments, the better the energy compaction. A wavelet is said to have K vanishing moments if

$$\int (t - t_0)^m \psi(t) dt = 0, \tag{2.11}$$

where, $0 \leq m \leq K - 1$. This implies that polynomials up to order $K - 1$ can be exactly represented by the approximation coefficients. The D_8 orthogonal filter bank has 8 vanishing moments. The analysis and synthesis bank wavelets for $B_{9/7}$ have 4 vanishing moments about $t_0 = 0$. For $B_{22/14}$, the analysis and synthesis wavelets have 5 and 7 vanishing moments respectively about $t_0 = \frac{1}{2}$. Based on the number of vanishing moments, we expect D_8 to provide the best energy compaction, followed by $B_{22/14}$ and $B_{9/7}$ (in that order).

2.2.2 DWT of 2-D Signals

Since an image is two dimensional, it requires a 2-D wavelet transform. Separable wavelet transforms are typically used; separability gives separable scaling and wavelet functions so that the 2-D transform is equivalent to two separate 1-D transforms [23]. The 2-D DWT is implemented as a 1-D row transform followed by a 1-D column transform. The transform coefficients are obtained by projecting the 2-D input signal $x(s,t)$ onto a set of 2-D basis functions that are simply expressed as the product of two 1-D basis as shown in equation (2.12). (s and t are the variables that denote the two dimensions–rows and columns).

$$\begin{aligned} \phi(s,t) &= \phi(s)\cdot\phi(t), & \psi_1(s,t) &= \psi(s)\cdot\phi(t), \\ \psi_2(s,t) &= \phi(s)\cdot\psi(t), & \psi_3(s,t) &= \psi(s)\cdot\psi(t). \end{aligned} \tag{2.12}$$

The 2-D DWT can be expressed as the set of equations (2.13). The scaling function $\phi(x,y)$ and wavelets $\psi_1(s,t)$, $\psi_2(s,t)$ and $\psi_3(s,t)$ (and the corresponding transform coefficients $F(N,j,n)$, $F^{(1)}(i,j,n)$, $F^{(2)}(i,j,n)$ and $F^{(3)}(i,j,n)$) correspond to different subbands in the decomposition. $F(N,j,n)$ are the coarse coefficients that constitute the LL subband. $F^{(1)}(N,j,n)$ coefficients contain the vertical details and correspond to the LH subband. $F^{(2)}(N,j,n)$ coefficients contain the horizontal details and correpond to the HL subband. $F^{(3)}(N,j,n)$ coefficients represent the diagonal details in the image and constitute the HH subband. Thus, a single-level decomposition has four subbands of coefficients as shown in Figure 2.4.

$$\begin{aligned} F(N,j,n) &= \int\int x(s,t)2^{N}\phi(2^{N}s-j)\phi(2^{N}t-n)\mathrm{d}s\mathrm{d}t \Rightarrow LL, \\ F^{(1)}(N,j,n) &= \int\int x(s,t)2^{N}\phi(2^{N}s-j)\psi(2^{N}t-n)\mathrm{d}s\mathrm{d}t \Rightarrow LH, \\ F^{(2)}(N,j,n) &= \int\int x(s,t)2^{N}\psi(2^{N}s-j)\phi(2^{N}t-n)\mathrm{d}s\mathrm{d}t \Rightarrow HL, \\ F^{(3)}(N,j,n) &= \int\int x(s,t)2^{N}\psi(2^{N}s-j)\psi(2^{N}t-n)\mathrm{d}s\mathrm{d}t \Rightarrow HH. \end{aligned} \tag{2.13}$$

The synthesis bank performs the 2-D IDWT to reconstruct $x(s,t)$. The 2-D IDWT is given by equation (2.14).

$$\begin{aligned} x(s,t) = & \sum_{j}\sum_{n} F(N,j,n)\cdot 2^{N}\phi(2^{N}s-j)\phi(2^{N}t-n) \\ & + \sum_{i=-\infty}^{N}\sum_{j}\sum_{n} [F^{(1)}(i,j,n)\cdot 2^{i}\psi(2^{i}s-j)\phi(2^{i}t-n) \\ & \quad + F^{(2)}(i,j,n)\cdot 2^{i}\phi(2^{i}s-j)\psi(2^{i}t-n) \\ & \quad + F^{(3)}(i,j,n)\cdot 2^{i}\psi(2^{i}s-j)\psi(2^{i}t-n)] \end{aligned} \tag{2.14}$$

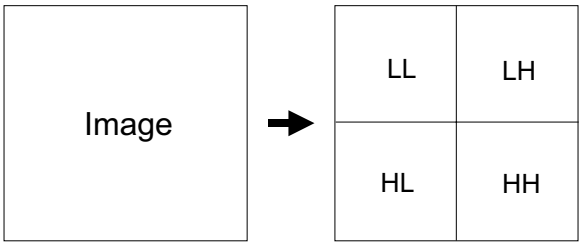

Figure 2.4: Single-level 2-D scalar wavelet decomposition.

2.3 Significance Tree Quantization

When the *LL* subband is iteratively decomposed, a pyramid of subbands is formed from a coarsest scale at the top to a finest scale at the bottom. Figure 2.5 illustrates the three-level pyramid resulting from a three-level decomposition of the image *Lighthouse* .

In this pyramid decomposition, each wavelet coefficient in a coarser subband is the parent of four children arranged in the form of a 2×2 block in the finer subband immediately below it in the pyramid. This is shown in Figure 2.6 where each arrow points from a parent to its group of four children. We can identify hierarchical trees in the decomposition with wavelet coefficients at different scales constituting its nodes at different levels. Images usually contain more energy in the lower frequencies. As a result, coarse coefficients typically

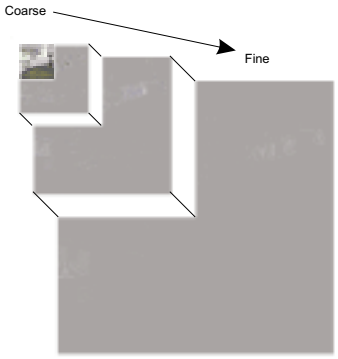

Figure 2.5: Pyramid of subbands in a three-level decomposition

have large magnitudes and the energy diminishes as we move from the tip of the pyramid to the bottom. Thus in each tree, the energy of a parent is indicative of the energy of its descendents. Quantizers like *EZT* [20] and *SPIHT* [18] exploit this spatial relationship within trees.

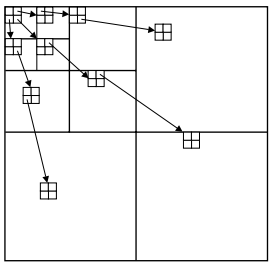

Figure 2.6: Hierarchical trees in multi-level decomposition.

2.3.1 SPIHT

This thesis uses the *SPIHT* quantizer for compression. *SPIHT* is a fully embedded wavelet coding algorithm that progressively refines the most significant coefficients in the order of decreasing energy levels. It apportions the available bit budget between encoding the tree map (which indicates the location of the significant coefficients) and the significance infor-

mation itself (which indicates the energy of the coefficients). It allows for precise rate control and affordable computational complexity. The *SPIHT* quantizer is reported to surpass other coding techniques such as DCT, EZT, etc.

2.4 Symmetric Extension

When finite length signals are processed, the abrupt drop-off of energy at the signal boundaries gives rise to high-energy wavelet coefficients in high-frequency bands. At low bit rates, these coefficients undergo severe quantization. Thus, the reconstruction from the truncated coefficients produces discernable artifacts around the image borders.

The symmetric biorthogonal wavelets have a solution to this problem: symmetric extension. For symmetric biorthogonal wavelets with even-length filters, an input signal of length N is doubled to $2N$ by flipping it about the boundaries. For wavelets with odd-length filters, the signal is extended to length $2N-1$ by flipping it about the boundary but without repeating the boundary sample [23]. The result is a signal (of length N) with even symmetry at the output of the filter bank after one stage of decomposition. Due to the symmetry, only half of the output coefficients, i.e. $\frac{N}{2}$ samples need to be retained. Hence, this technique is non-expansive. Figure 2.7 shows the generation of the LL subband in a one-level decomposition of a 256×256 using an even-tap biorthogonal wavelet.

In the filter bank, the filtering is implemented as an FFT; the FFT treats the input as a periodic signal. So, effectively, this is a periodic, symmetric extension technique.

Symmetric extension results in an improvement of 0.4–0.6 dB over plain periodic extension (for the same biorthogonal wavelet) at the cost of slightly increased computation [4]. Symmetric extension eliminates border artifacts and usually gives perceptible improvement in the image quality.

Symmetric extension is applicable only to symmetric filters and cannot be used with orthog-

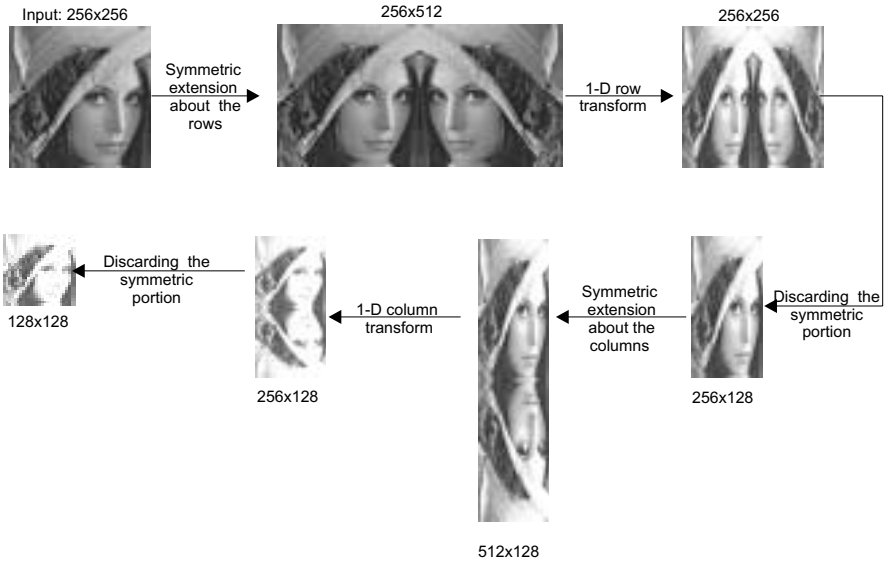

Figure 2.7: Symmetric extension for an even-tap biorthogonal scalar wavelet.

onal wavelets. In fact, symmetric extension is one of the primary reasons for the success of biorthogonal wavelets over orthogonal wavelets in image compression.