Embedded Wavelet Packet Image Coder with Fast Rate-Distortion Optimized Decomposition

Jin Li and C.-C. Jay Kuo

Integrated Media Systems Center and Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, California 90089-2564

Po-Yuen Cheng

Physical Optics Corporation, Applied Technology Division
Torrance, California 90505-5228

ABSTRACT

A fast rate-distortion (R-D) optimized wavelet packet (WP) transform is proposed for image compression in this research. By analyzing the R-D performance of the quantizer and the entropy coder, we show that the coding distortion D can be modeled as an exponentially decaying function as the coding rate R increases. With this exponential R-D model, it is proved that the constant R-D slope criterion for optimum bit allocation is equivalent to the constant distortion criterion, which can be easily implemented via thresholding. Based on this analytical result, we develop a fast wavelet packet decomposition scheme which is optimized in the R-D sense by comparing simple parameters associated with each wavelet packet band such as the 1st or 2nd absolute moments. We have performed extensive experiments to demonstrate the performance of an image coder using the proposed R-D optimized wavelet packet transform, and shown that our scheme is highly competitive with all well known state-of-the-art image coders.

Keywords: image compression, rate-distortion, wavelet packets, wavelet coding, embedded coding.

1 INTRODUCTION

To achieve image compression, it is typical to transform image data from the space domain to the transform domain for energy compaction, and then quantize and encode transform coefficients. Since different images have different characteristics, an adaptive transform finetuned to image characteristics is more effective than a fixed transform. It is of great interest to determine the best transform which meets a certain optimality criterion while keeping the computational complexity under a certain level. In this work, we focus on image compression methods based on wavelet and wavelet packet transforms, which are viewed as fixed and adaptive transforms, respectively. Commonly used criteria for selecting a wavelet packet decomposition scheme include maximum energy compaction, minimum entropy, etc.

For compression, the performance of a transform should be evaluated by the final result of the entire coding. In other words, the best transform should match the quantizer and the entropy coder so that the best rate-distortion (R-D) tradeoff can be achieved. Thus, the R-D performance of the quantizer and the entropy coder plays an important role in the design of the optimal wavelet packet transform. Previous work has been performed to optimize the R-D performance of image and video coders. Strobach [15] developed an adaptive quadtree scene coder for video coding, where motion vectors were estimated with a criterion which minimizes the joint coding rate for motion vectors and quadtree residuals. In wavelet packet (WP) based image coding, Ramchandran *et al.* [12], [18] estimated the coding

distortion and the bit rate with respect to each WP decomposition to search for the best choice.

The major problem with these methods is the high computational complexity required. In [15], coding rates and their associated distortions are obtained by performing the actual quadtree coding on the residuals. In [12] and [18], the coding distortion is evaluated by computing the mean square error (MSE) after quantization, and the coding rate is estimated by calculating the first order entropy of quantized WP coefficients. To minimize the cost function, one still has to compute unknown operating parameters such as the Lagrangian factor λ [12], [18]. The practical use of these methods is greatly hindered by the high computational cost.

In this research, we examine a new image compression scheme which applies the R-D optimized wavelet packet (WP) transform and adopts the highly efficient layered zero coding (LZC) scheme [16] for the quantization and entropy coding of WP coefficients. Our major contribution in this work is the proposal of a new WP decomposition scheme which efficiently determines the best WP transform in the R-D sense. The key idea is the use of an empirical R-D model to characterize the performance of the successive quantizer and the entropy encoder of LZC. This model enables us to determine the R-D optimized WP transform accurately with a very low computational cost. Extensive experimental results will be presented to demonstrate the performance of the new compression scheme and the comparison with that of several state-of-the-art image coders. Generally speaking, our scheme provides a substantial coding gain at the expense of a slightly increased computational cost. For example, it outperforms the embedded zero wavelet coding (EZW) [14] scheme from 1 to 2 dB for different images at various bit rates with an increase of complexity of 60-70% (see Section 6 for more details).

This paper is organized as follows. We briefly review the wavelet packet transform and the layer zero coding scheme in Section 2. Then, an empirical R-D model is presented in Section 3, and a fast R-D optimized wavelet packet decomposition scheme is described in Section 4. The complete compression algorithm, i.e. the layered zero coder with the R-D optimized WP transform, is summarized in Section 5. Experimental results for image and video compression are given in Section 6. Concluding remarks are provided in Section 7.

2 BACKGROUNDS

In this section, we first provide a brief review of the wavelet packet transform and layered zero coding scheme. Then, an overview of our approach is given.

2.1 Wavelet packet transform

Discrete wavelet transform of a sequence is achieved by passing the sequence through a quadrature mirror filter (QMF) consisting of a certain low- and high-pass filter pair followed by a 2-to-1 downsampling operation. Extensive experiments [17] have shown that the biorthogonal 9-7 tap spline filter [1] gives the best result so that it is used throughout this work. The 2-D wavelet transform is constructed by forming the tensor product of two 1-D wavelet transforms along the horizontal and vertical directions. When we extend the above wavelet transform to the multiscale case, there exist many approaches. One approach is to apply the wavelet decomposition recursively only in the lowest frequency bands, which results in the well known pyramid wavelet transform. The pyramid decomposition handles smooth and edge regions well, since it can effectively represent low frequency components and localize high frequency components. However, it is not suitable for texture representation. Textures can be conveniently analyzed by the wavelet packet (WP) transform [2], where the wavelet decomposition is applied adaptively in subbands. Due to its adaptivity, the WP transform is suitable for nonstationary signal analysis and representation. One critical problem in the WP transform is the choice of a good decomposition among all possible wavelet packets with respect to a given signal. In early WP literatures, criteria were chosen based on the measurement of entropy [3] and energy [2]. They are however not directly related to a better compression performance.

More recently, the optimal WP decomposition to achieve the best rate-distortion (R-D) tradeoff has been examined by researchers [6], [11], [12], [18]. Under this framework, the selection of the best WP in the R-D sense can be

formulated as the solution of the following cost function:

$$J = \min_{\lambda} \min_{\mathbf{F} \in \mathcal{F}} \min_{\mathbf{Q} \in \mathcal{Q}} \sum_{i} \mathcal{D}_i + \lambda \mathcal{R}_i, \tag{1}$$

where $\mathcal{F}$ denotes the set of all possible WP transforms, $\mathcal{Q}$ denotes the set of all allowed spatial-frequency quantization (SFQ) schemes described in [12], the Lagrangian factor λ corresponds to the optimal operating slope, and $\mathcal{D}_i$ and $\mathcal{R}_i$ are the total coding distortion and coding rate for band i with respect to a specific WP decomposition $\mathbf{F}$. In their work, $\mathcal{D}_i$ is estimated by using the mean square error (MSE) between the original and quantized wavelet packet coefficients and $\mathcal{R}_i$ is estimated by the first order entropy of quantized coefficients. To determine the best choice, one has to perform the estimation on $\mathcal{D}_i$ and $\mathcal{R}_i$ for every quantization scheme in $\mathcal{Q}$ and for every WP decomposition in $\mathcal{F}$, which is computationally expensive. Furthermore, to calculate the cost function J, the optimal wavelet packet decomposition $\mathbf{F}$, the corresponding optimal quantization $\mathbf{Q}$ and the Lagrangian factor λ have to be determined jointly. In [6], [11], [12], this procedure is carried out by three embedded loops. In the most inner loop, both λ and $\mathbf{F}$ are fixed for the search of the optimal quantizer $\mathbf{Q}$. In the second loop, λ is fixed for the search of the best combination of $\mathbf{F}$ and $\mathbf{Q}$. Finally, the optimal combination of λ, $\mathbf{F}$ and $\mathbf{Q}$ is determined in the outer loop. Although speed-up techniques such as the gradient-based method can be used to reduce the computational complexity during the search process, the overall computational cost is still very high.

2.2 Layer zero coding (LZC)

Although wavelet-based image coding has been studied for years [1], [9], it is the work of Shapiro [14], known as the embedded zerotree wavelet (EZW) coding, that demonstrates the clear advantage of the wavelet transform approach. In addition to providing a very good rate-distortion performance, EZW has a desirable embedding property where bits are generated in the order of significance so that the scheme is suitable for progressive transmission. Key ingredients in EZW include: successive quantization of wavelet coefficients, classification of wavelet coefficient representation into significant and insignificant coding, and zerotree grouping of insignificant bits.

There has been follow-up work which further improves the R-D performance of EZW, for example, the joint space-frequency quantization (SFQ) proposed by Xiong and Ramchandran [18], and the scheme of set partitioning in hierarchical trees (SPIHT) proposed by Said and Pearlman [13]. Another interesting work known as the layered zero coding (LZC), which exploits interscale prediction among bit layers to improve the R-D performance, was proposed by Taubman and Zakhor [16]. LZC replaces the coding of zerotrees of insignificant bits with a context adaptive arithmetic coder which enables the entropy coding of correlated insignificant bits in both tree and non-tree forms. The context adaptive arithmetic coder is first adopted in the JBIG binary image coding standard [5]. Its properties are highlighted as follows. First, it is a highly efficient entropy coder with its average coding rate close to the entropy of the source. Second, it is computationally inexpensive, since its coding process only involves addition and shifting operations. Third, it is parameter free in the sense that the source probability distribution p_0 and p_1 are estimated on the fly with no training. Finally, it is easy to construct parallel coders for a compound source by assigning an individual coder to a specific source.

LZC treats the bit layer representation of wavelet coefficients as a compound source, where each bit is conditioned by the values of its surrounding bits in the space and subband domains. A context rule is adopted to classify such a compound source into different categories. For example, an n-bit context classifies the source into 2^n categories, which include the zerotree correlation as one special case. With the help of the context adaptive arithmetic coding, LZC achieves a better compression than EZW. For more details of the above discussion, we refer to [16].

2.3 Overview of this work

Our work follows the basic framework of LZC, i.e. the quantizer and the entropy coder of the proposed scheme are the same as those in LZC. However, the pyramid wavelet transform in LZC is modified to be the R-D optimized WP decomposition in this research. Our scheme evaluates the R-D efficiency of each WP decomposition based on an empirical R-D codec model. It is first demonstrated that the coding distortion D decays exponentially with the increase of the coding rate R by applying LZC to wavelet coefficients in any given subband. Then, we show that the constant R-D slope criterion for optimum bit allocation is equivalent to the constant distortion criterion under

the exponentially decaying R-D characteristics, and that the constant distortion criterion can be easily implemented by applying a single significant threshold to all WP bands. Thus, the codec operating point is controlled by a quantity D_{th} called the threshold distortion. The R-D performance of any WP band can be estimated by using D_{th} and simple statistics such as the 1st or 2nd absolute moments associated with each WP band. This enables us to develop a fast algorithm for performing the R-D optimized WP decomposition. The new child-parent band relationship and the scanning order with respect to a given WP decomposition can also be defined. By combining all individual components, we can present a complete compression algorithm consisting of the WP transform, successive quantization and arithmetic entropy coding.

3 Rate-distortion model for wavelet packet decomposition

3.1 Exponential R-D model

Consider the decomposition of an image into N subband B_i, $i = 1, 2, \cdots, N$, with the wavelet or wavelet packet transform. It is assumed that wavelet coefficients are decorrelated in each subband, and the distribution can be modeled by a random variable X_i with the generalized Gaussian probability density function (pdf):

$$\begin{aligned} P_{X_i}(x) &= a_i e^{-[b_i|x|]^{\gamma_i}}, \qquad (2) \\ a_i = \frac{b_i \cdot \gamma_i}{2\Gamma_i(1/\gamma_i)}, &\qquad b_i = \sigma_i^{-1}\left[\frac{\Gamma_i(3/\gamma_i)}{\Gamma_i(1/\gamma_i)}\right]^{1/2}, \end{aligned}$$

where γ_i is a shape parameter and parameters a_i and b_i are functions of γ_i and variance σ_i^2. Note that a generalized Gaussian pdf is completely specified by the shape parameter γ and variance σ^2. Also, the Gaussian and Laplacian pdfs are special cases of the generalized Gaussian pdf with $\gamma = 2$ and 1, respectively. The R-D performance of coding such a source will be discussed in this section. We will focus on the theoretical performance bound and then the performance of a practical wavelet coder.

The theoretical Shannon lower bound (SLB) of a statistical source X is of the form [4]

$$R = h(X) - \frac{1}{2}\log_2(2\pi e D) \quad \text{or} \quad D = \frac{2^{2h(X)}}{2\pi e}2^{-2R}, \qquad (3)$$

where $h(X)$ is the entropy of random variable X. It is proved in [4] that SLB is achievable theoretically for the Gaussian source. Even though SLB may not be achievable for other sources, it still provides a good performance bound for practical coders. Equation (3) shows that coding distortion D decays exponentially with the increase of coding rate R. This important property is called the *exponential R-D* property. It is expected that a reasonably good coder should follow SLB closely and, therefore, satisfy the exponential R-D property.

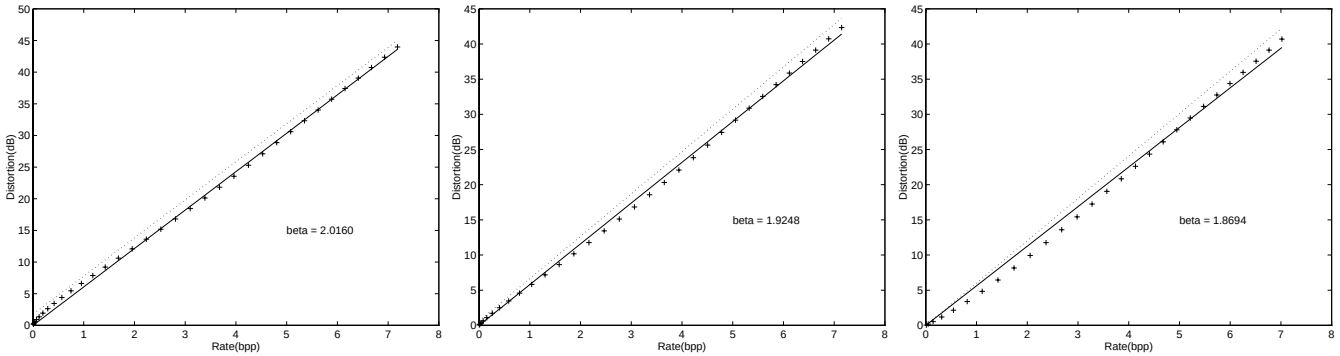
Next, the R-D performance of a practical wavelet coder is examined. In particular, we apply the layered zero coding (LZC) scheme [16] to a generalized Gaussian distributed source with different shape parameters. The experimental R-D operating points for $\gamma = 0.7$, 1.0 (Laplacian) and 2.0 (Gaussian) are shown with '+' in Figs. 1 (a)-(c), respectively. These R-D operating points can be well approximated by the exponential model

$$\text{Assumption A:} \qquad D = D_{\max}2^{-\beta R}, \qquad (4)$$

where the parameter β characterizes the exponentially decaying rate and can be calculated via least-square fitting, and $D_{\max}$ is the maximum coding distortion at coding rate $R = 0$, which equal to the variance of the source. Equation (4) is adopted as the first assumption in our model.

It is worthwhile to point out that, for the coding of the generalized Gaussian source with LZC, β is only related to the shape parameter γ_i and independent of the variance σ_i^2. This can be explained below. Suppose that the variance of the source is increased by a factor of k^2, i.e. $\sigma_i^{'2} = k^2\sigma_i^2$. The maximum coding distortion also increases proportionally $D_{\max}^{'} = k^2 D_{\max}$. By adopting a new quantization threshold $Q_i^{'} = kQ_i$, we can encode the new generalized Gaussian source at the same coding rate $R_i^{'} = R_i$ and the new distortion becomes $D_i^{'} = k^2 D_i$. Consequently, we have

$$D_i^{'} = D_{\max}^{'}2^{-\beta R_i^{'}},$$



(a) (b) (c)

Figure 1: Rate-distortion performance of coding of the generalized Gaussian pdf with shape parameter (a) $\gamma = 0.7$, (b) $\gamma = 1$ (Laplacian), (c) $\gamma = 2$ (Gaussian), where the dotted line is the Shannon Lower Bound (SLB) of the source, the '+' symbols denote experimental points, and the solid line indicates the approximating rate-distortion function with $\beta = 2.0160, 1.9248, 1.8694$, respectively.

and conclude that β is independent of the variance of the source.

For the generalized Gaussian source with shape parameter $\gamma = 0.7$, 1.0, 2.0, we have $\beta = 2.0160$, 1.9248 (Laplacian) and 1.8694 (Gaussian), respectively. The fitted curves with these β values are plotted with solid lines in Figs. 1 (a)-(c). SLB is also plotted in the same figures with dotted lines for comparison. One can clearly see that LZC gives a performance very close to SLB. With shape parameter γ changes from 2.0 to 0.7, β_i only changes from 1.8694 to 2.0160, which is less than 10%. Since β_i is also insensitive to the change of the generalized Gaussian shape parameter γ_i, we adopt the following approximation as the second assumption in our model:

$$\text{Assumption B:} \qquad \beta_i \approx \beta, \tag{5}$$

for band B_i, $1 \le i \le N$.

3.2 Model with constant exponentially decaying rate

For band B_i, $1 \le i \le N$, we denote the number of wavelet coefficients, the average distortion (per coefficient), the average bit rate (per coefficient) and the maximum coding distortion by S_i, D_i, R_i and $D_{\max,i}$, respectively. It is clear that the total number of coefficients for the entire image is $S = \sum_i^N S_i$. For a total bit budget R, we want to optimally allocate the average bit rates $R_1, R_2, \cdots, R_N$ so that

$$\begin{cases} \sum_i^N S_i R_i = R, & \text{(constrained bit budget)} \\ \min \sum_i^N S_i D_i, & \text{(minimized distortion)} \end{cases} \tag{6}$$

This problem can be solved with the Lagrangian method by requiring

$$\frac{\partial D_i}{\partial R_i} = -\lambda \quad \text{(constant)}, \quad 1 \le i \le N, \tag{7}$$

where λ is known as the Lagrangian multiplier. By assuming that the coder satisfies (4), we have

$$\frac{\partial D_i}{\partial R_i} = D_{\max,i} 2^{-\beta_i R_i} \cdot (-\beta_i \ln 2) = -D_i \beta_i \ln 2. \tag{8}$$

Under assumptions (4) and (5) and by comparing (7) and (8), it is easy to see that the constant R-D slope criterion for optimum bit allocation can be converted to the constant distortion criterion:

$$D_i = D_{th} \quad \text{(another constant)}, \quad 1 \le i \le N, \tag{9}$$

which is much easier to implement in comparison with the constant slope criterion (7).

3.3 Nonnegative bit rate constraint

By rearranging (4), we have

$$R_i = \frac{1}{\beta}\log_2 \frac{D_{\max,i}}{D_i} = \frac{1}{\beta}\log_2 \frac{\sigma_i^2}{D_i}, \tag{10}$$

where the fact that the maximum coding distortion $D_{\max,i}$ is equal to the variance σ_i^2 of WP band B_i is used. Taking the nonnegative bit rate constraint $R_i \geq 0$ into consideration, the solution to (6) has to be modified from (9) to:

$$D_i = \begin{cases} D_{th} & \text{(constant)}, \quad D_{th} < \sigma_i^2, \\ \sigma_i^2, & \quad D_{th} \geq \sigma_i^2, \end{cases} \tag{11}$$

We call D_{th} the *threshold distortion* (or known as the water-filled distortion in [4]). Parameter D_{th} controls the operating point of the coder. Equation (11) implies that, to optimally encode a WP decomposed image at a certain bit rate, we have to select D_{th} so that the band is not encoded at all if its variance is smaller than D_{th} and with distortion equal to D_{th} after coding if its variance is greater than D_{th}.

For the layered zero coder (LZC), D_{th} can be controlled by adjusting the value of the finest quantization step size T_S (see also discussion in Section 5). We analyzes the coding distortion (MSE) versus the finest quantization step size T_S for several generalized Gaussian sources with variance $\sigma^2 = 1$. The behavior of LZC changes dramatically at the point $T_{th} = \sqrt{12\sigma^2}$.

$$D = \begin{cases} T_S^2/12, & T_S^2/12 < \sigma^2 \\ \sigma^2, & T_S^2/12 \geq \sigma^2 \end{cases} \tag{12}$$

For $T_S > T_{th}$, the coding distortion can be well approximated by σ^2 and is independent of the value of T_S. On the other hand, for $T_S < T_{th}$, the coding distortion can be approximated by $T_S^2/12$. By comparing (11) and (12), we have the association:

$$D_{th} = \frac{T_S^2}{12}.$$

A smaller value of T_S implies a smaller value of D_{th} and a coded image of a better quality. The mean squared error D_{mse} is very close to D_{th} when D_{th} is small. Thus, we can use the threshold distortion D_{th} to control the coded image quality as well as to represent the mean squared error. Applying (11) to (10) and using D_{th} as the coding control parameter, we have the average bit rate for band B_i as:

$$R_i = \frac{1}{\beta}\log_2 \max\left(1, \frac{\sigma_i^2}{D_{th}}\right) = \frac{1}{\beta}\log_2 \frac{\kappa_i^2}{D_{th}}, \qquad \text{with} \qquad \kappa_i^2 = \max(D_{th}, \sigma_i^2) \tag{13}$$

denotes the *bounded variance* for WP band i. By substituting (13) in (6), we can write the total bit budget for the entire image as:

$$R = \frac{1}{\beta}\sum_{i=1}^{N} S_i \log_2 \frac{\kappa_i^2}{D_{th}}, \tag{14}$$

Thus, (14) can be viewed as the rate-distortion function of the WP decomposition. The optimal bit allocation scheme only depends on the threshold distortion D_{th} and the variance σ_i^2 of WP band B_i.

4 FAST R-D OPTIMIZED WAVELET PACKET DECOMPOSITION

Based on the R-D model presented in the previous section, we will derive a fast wavelet packet decomposition scheme in this section.

4.1 Algorithm

Consider two WP decompositions $\mathbf{F_1}$ and $\mathbf{F_2}$, where $\mathbf{F_1}$ decomposes the image into N_1 bands with sizes $S_{1,1}, \cdots, S_{1,N_1}$ and variances $\sigma_{1,1}^2, \cdots, \sigma_{1,N_1}^2$ and $\mathbf{F_2}$ decomposes the image into N_2 bands with sizes $S_{2,1}, \cdots, S_{2,N_2}$

and variances $\sigma^2_{2,1}, \cdots, \sigma^2_{2,N_2}$. With the same D_{th}, the difference in the coding budget for decompositions $\mathbf{F_1}$ and $\mathbf{F_2}$ is:

$$R_1 - R_2 = \frac{1}{\beta}\left[\sum_{i=1}^{N_1} S_{1,i} \log_2 \kappa^2_{1,i} - \sum_{i=1}^{N_2} S_{2,i} \log_2 \kappa^2_{2,i}\right]. \tag{15}$$

Note that (15) still depends on D_{th} since $\kappa^2_{j,i} = \max(D_{th}, \sigma^2_{j,i})$. However, it is not sensitive to a small variation of D_{th}. This is especially true for a very large or small value of D_{th}.

The proposed WP decomposition algorithm is a merge-based method. Since the WP subtree of an R-D optimized WP decomposition is optimized in the R-D sense itself, the merge-based algorithm gives the R-D optimized wavelet packet decomposition. We first transform an image with a fully-decomposed wavelet transform to the maximum depth d_m. This corresponds to a fully-decomposed quadtree with 4^{d_m} leaf nodes, each of which consists of the same number of coefficients. The merge step starts from the bottom of the quadtree. Consider the WP subtree $F_{d_m-1,i,j}$ at scale $d_m - 1$. The variance of the undecomposed WP subtree $F_{d_m-1,i,j}$ is denoted by $\sigma^2_{d_m-1,i,j}$ and its size by S_{d_m-1}. It can either remain the same or be further decomposed into four WP bands $F_{d_m,2i,2j}$, $F_{d_m,2i+1,2j}$, $F_{d_m,2i,2j+1}$, $F_{d_m,2i+1,2j+1}$ at scale d_m with size $S_{d_m} = S_{d_m-1}/4$. With (15), we can calculate the bit saving due to further decomposition as:

$$R_s = \frac{S_{d_m-1}}{\beta}\left[\log_2 \frac{\kappa^2_{d_m-1,i,j}}{[\kappa^2_{d_m,2i,2j} \cdot \kappa^2_{d_m,2i+1,2j} \cdot \kappa^2_{d_m,2i,2j+1} \cdot \kappa^2_{d_m,2i+1,2j+1}]^{1/4}}\right]. \tag{16}$$

When the saving is greater than the decomposition overhead R', the WP decomposition is accepted. Otherwise, it is rejected. Since each WP subtree needs 1 bit to indicate whether it is further decomposed, the number R' of overhead bits is equal to 4. For practical implementation, we actually enlarge the threshold to be around 10 to avoid many small WP bands. After the decision, the equivalent bounded variance is calculated for future use:

$$\kappa'^2_{d_m-1,i,j} = \begin{cases} \kappa^2_{d_m-1,i,j}, & \text{decomp. rejected} \\ \left[\kappa^2_{d_m,2i,2j} \cdot \kappa^2_{d_m,2i+1,2j} \cdot \kappa^2_{d_m,2i,2j+1} \cdot \kappa^2_{d_m,2i+1,2j+1}\right]^{1/4} 2^{\frac{\beta R'}{S_{d_m-1}}}, & \text{decomp. accepted} \end{cases} \tag{17}$$

Note that the effect of decomposition overhead bits R' is included in (17), if the WP decomposition is accepted. The same procedure iterates from the bottom to the top of the fully-decomposed quadtree.

4.2 Choice of two parameters

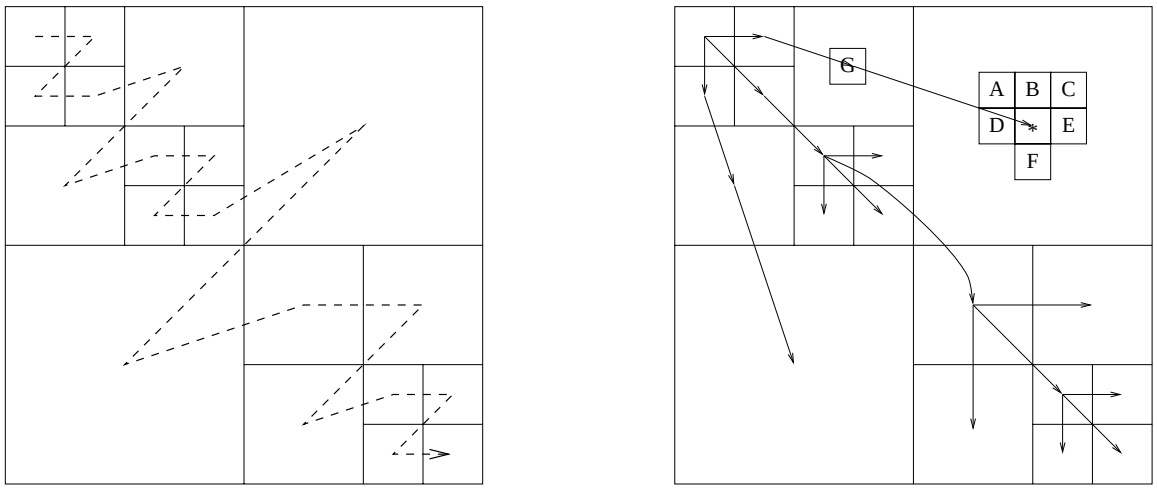
In the R-D optimized WP decomposition, we use the threshold distortion D_{th} to control the desired WP decomposition. D_{th} can be well approximated by the coding MSE D_{mse}. Thus, for a target PSNR value, we can compute D_{th} as

$$D_{th} = 255^2 10^{-\frac{\mathrm{PSNR}}{10}}.$$

The resulting wavelet packet transform is insensitive to a small variation of D_{th}. Generally speaking, if $D_{th,1} > D_{th,2}$, the WP decomposition $\mathbf{F_1}$ associated with $D_{th,1}$ can be derived by merging some of WP subtrees in $\mathbf{F_2}$. In particular, if D_{th} is smaller or greater than all variances of WP bands, the resulting WP decomposition is independent of D_{th}. Therefore, even if we do not choose the exact threshold distortion D_{th} for a specific coding rate, the resulting WP decomposition differs from the optimal one in only a few nodes which should be either merged or decomposed.

Another parameter to be estimated is the variance σ_i^2 of each WP band, which is equal to the maximum coding distortion $D_{\max,i}$ defined in (4). Since WP coefficients have a zero mean in each subband (except the lowest frequency band), one straightforward way to estimate the variance is to compute the mean squared error (MSE) of the WP coefficients in each band before coding, i.e.

$$\hat{\sigma^2}_{mse,i} = \frac{1}{S_i}\sum_{j=1}^{S_i} x^2_{i,j}. \tag{18}$$



(a) (b)

Figure 2: Layered zero coding of wavelet packet, (a) Depth-first scan order for wavelet packet, (b) Context for arithmetic coding. Among the context, 'A' to 'F' are intraband context, 'G' is interband context. The parent-childhood relationship of the wavelet packet bands is shown with arrows.

where $x_{i,j}$ denote the jth wavelet coefficient in band B_i.

Estimation formula (18) is based on the estimation of sample variance. The accuracy of variance estimation can be improved by considering a more accurate pdf model of WP coefficients. We know from experiments that the pdf of WP coefficients can be well approximated by the Laplacian density. It is not difficult to prove that an unbiased sufficient statistics for the variance of the Laplacian pdf is

$$\hat{\sigma^2}_{mae,i} = \frac{1}{S_i} \sum_{j=1}^{S_i} \mid x_{i,j} \mid, \tag{19}$$

which is the mean absolute error (MAE) of WP coefficients before coding. The performance comparison between the above two variance estimation formulas will be given in Section 6.1.

5 LAYERED ZERO CODING WITH WAVELET PACKET TRANSFORM

The layered zero coding (LZC) scheme [16] can be easily generalized from the pyramidal wavelet transform to the wavelet packet (WP) transform. The complete coding algorithm is presented in this section.

Step 1: WP Decomposition
We perform the WP decomposition as described in the previous section.

Step 2: Successive Quantization
The maximum absolute value of the whole WP decomposed image is searched and denoted by

$$T_0 = \max_{i,j} \mid x_{i,j} \mid,$$

where $x_{i,j}$ denotes the jth coefficient of band B_i. The quantization step size (or called the significance threshold) of layer $s = 1, 2, \cdots, S$ is set to

$$T_s = T_0 \cdot 2^{-s}.$$

Note that the significance threshold value is reduced by one half for each further refinement so that successive quantization can be conveniently performed. T_S corresponds to the finest quantization step.

Step 3: Layered Arithmetic Coding
The whole WP image is scanned using a depth-first order as shown in Fig. 5(a). We start with the first layer

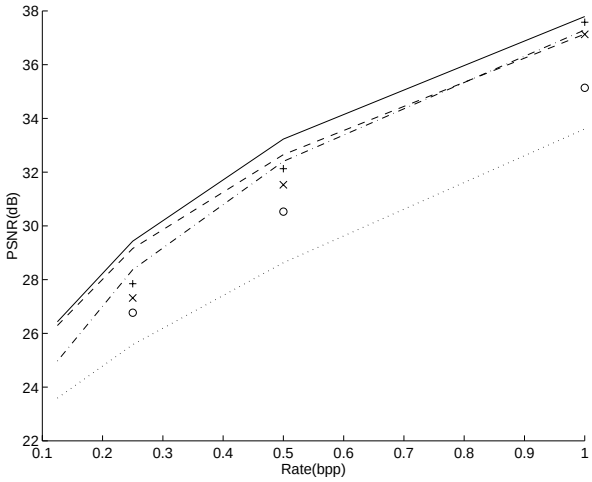


Figure 3: Performance comparison of Barbara image for (a) R-D optimized wavelet packet (WP) coder with $\hat{\sigma^2}_{mae}$ (solid line), (b) R-D optimized WP coder with $\hat{\sigma^2}_{mse}$ (dashed line), (c) LZC (dash dotted line), (d) JPEG (dotted line), (e) EZW ('o'), (f) single tree SFQ ('x'), and (g) double tree SFQ ('+').

of the wavelet coefficients. The scanning order is determined by the band order LL, LH, HL and HH within the same scale and by the raster line order within each WP band. It should be pointed out that the scanning order is completely specified by the WP decomposition so that no additional information has to be transmitted besides the WP decomposition structure. The most significant bit of each scanned wavelet coefficient

$$s_i = \begin{cases} 0, & |\, x_i \,| \leq T_1, \\ 1, & |\, x_i \,| > T_1 \end{cases}$$

is encoded with a context adaptive arithmetic coder. Generally speaking, the context consists of 7 bits as denoted by 'A' to 'G' in Fig. 5(b). Among the 7 bits, 'A'-'F' represent the intraband context. They are characterized by the status of the neighboring wavelet coefficients in the same WP band. Note also that 'A' to 'D' are scanned before the current coding position '*' so that they are values from the current coding layer while 'E' and 'F' are scanned after '*' so that their values are actually obtained from the previous coding layer. The context value 'G' represents the interband relationship. It contains the bit value of the wavelet coefficient of the same layer at the same space location but in the parent band of the current coding position. The parent band is defined to be the WP subband at the same direction but at a coarse scale. One example of the parent-child relationship of the WP bands is depicted in Fig. 5(b). The parent-child relationship is uniquely specified by the WP decomposition so that no overhead information has to be transmitted.

The 7-bit context arithmetic coder uses the 7 scale/space neighboring bit informations (i.e. 'A'-'G') to classify the current position '*' into $2^7 = 128$ categories, and assign a separate adaptive arithmetic coder to each category. The context adaptive arithmetic coder is much more efficient than the coder using only the zerotree structure [14] since the latter can be viewed as a special case of using only context 'G'.

For bit layer $s > 1$, we further divide the coding of the bits into two phases: significance and refinement codings, since they have very different rate-distortion (RD) characteristics. We always put the significance coding phase before the refinement coding phase in the same layer for better coding performances[8]. If the current position is encoded with all zeros in the previous layers, we have to identify in the current layer whether the coefficient becomes non-zero, or significant, and the coding is called the significance coding. In the significance coding phase, we encode the bit in the current layer s in the same way as we do for layer 1. For coefficient which becomes significant (non-zero), its sign ('+' or '-') is arithmetically encoded as well. For the position which is already identified as significant in a certain previous layer, the coding of its current '0' or '1' gives the refinement subinterval of coefficient residuals and is called the refinement coding. It is also arithmetically encoded with a special context.

The above coding procedure repeats until the allocated coding bit rate is reached, or the desired coding layer is

(a) (b)

Figure 4: Experimental results for Barbara of (a) the coded image with LZC, and (b) the coded image with R-D optimized wavelet packet transform.

Table 1: Coding performance comparison of the Babara image.

Rate(bpp)	WP (MSE)	WP (MAE)	JPEG	EZW	LZC	ST + SFQ	DT + SFQ
0.125	26.29	26.43	23.59	-	24.97	-	-
0.25	29.16	29.43	25.58	26.77	28.37	27.32	27.85
0.5	32.65	33.23	28.63	30.53	32.40	31.53	32.13
1.0	37.14	37.79	33.61	35.14	37.29	37.13	37.58

reached, or the MSE of the quantized wavelet coefficient is smaller than a certain assigned threshold. The coding bitstream of the proposed WP coder has the embedding property. That is, the bit stream can be truncated at any point to result in a decoded image without truncation artifact.

6 EXPERIMENTAL RESULTS

Experimental results are given in this section to demonstrate the performance of the proposed R-D optimized wavelet packet coder.

6.1 Still image compression

We compare the performance of the proposed WP coder with a number of state-of-the-art compression algorithms, including JPEG [7], the embedded zero tree wavelet coding (EZW) [14], the layered zero coding (LZC) [16] and the space-frequency quantization (SFQ) WP coding [12]. Both decomposition algorithms examined in [12], i.e. the single tree (ST) and the double tree (DT) decompositions, are included in the comparison.

The first test image is "Barbara" of size 512×512. We compare the proposed WP image coder with the comparison coders mentioned above. For WP decompositions, two R-D models $D = D_{\max} 2^{-\beta R}$ with $D_{\max}$ estimated by the mean square error (MSE) $\hat{\sigma^2}_{mse}$ and the mean absolute error (MAE) $\hat{\sigma^2}_{mae}$ of wavelet coefficients in a WP band are compared. Experimental results are given in Table 1 and Fig. 3. We see from the figure that the MAE estimate always outperforms the MSE estimate, which implies that wavelet coefficients can be well modeled by a Laplacian density function. Therefore, in the following experiments, $D_{\max}$ is chosen to be $\hat{\sigma^2}_{mae}$. It is not surprising to see

Table 2: Coding performance for test images: Lena, Baboon, Boat and Creek.

Rate(bpp)	PSNR(dB) of our scheme for				PSNR(dB) of LENA for		
	Lena	Baboon	Boat	Creek	EZW	SFQ	LZC
0.0625	28.34	19.75	26.54	22.12	-	-	-
0.125	31.17	20.65	29.11	23.33	30.23	-	30.96
0.25	34.23	22.19	32.15	24.99	33.17	34.24	34.12
0.5	37.28	24.14	36.07	27.09	36.28	37.31	37.25

Table 3: Comparison of video coding algorithms (PY: pyramid wavelet, WP: wavelet packet).

Video sequence	Average PSNR (dB)			Average Bit Rate (bpp)
	MPEG1	PY	WP	
Flower	26.10	27.93	28.12	0.625
Mobile	25.59	27.61	28.23	0.797
Football	29.74	31.54	31.54	0.266
Table Tennis	28.87	29.85	29.86	0.219
Cheer	27.07	27.96	28.35	0.535
Bicycle	26.91	28.49	28.92	0.638

JPEG performs the worst among the methods in comparison. It basically serves as a reference point for the coding performance improvement. The new R-D optimized WP coder outperforms EZW and LZC by 2.5-2.8 dB and 0.5-1.0 dB, respectively. Since the layered zero coder used in our WP coder is very similar to the one in LZC, the gain of the new WP coder over LZC is mainly due to the R-D optimized WP decomposition. An enlarged portion of the encoded Barbara image with LZC and our coder is compared in Fig. 4(a) and (b). The image encoded by the R-D optimized WP coder provides a better visual quality at the texture dominant regions such as the trousers and table clothes. The performance advantage of our algorithm over the SFQ WP coder is 0.2dB at 1.0bpp and 1.6dB at 0.25bpp while the computational complexity of our coder is much less. Note also that the SFQ WP coder does not have the embedding property.

The second set of test images consists of four images of "Lena", "Baboon", "Boat" and "Creek" of size 512×512. Results are summarized in Table. 2. For image "Lena", our algorithm outperforms EZW by about 1.0dB, slightly outperforms the LZC by 0.03-0.21dB, and but is inferior to SFQ by 0.01-0.03dB. Note that the WP decomposition of "Lena" is close to the pyramidal structure. However, the R-D optimized WP decomposition still provides a gain over LZC with the pyramidal wavelet decomposition.

6.2 Video compression

For video compression, we apply the overlapped block motion compensation (OBMC) technique [10] and use the R-D optimized WP coder to encode the residue. Since the residue error after motion compensation has substantial middle frequency components, it is expected that an adaptive WP decomposition would perform better than a fixed pyramidal decomposition. Six test images sequences, i.e. "Flower", "Mobile", "Football", "Table Tennis", "Cheer" and "Bicycle" of the CIF format (360×240 pixels) were used in the experiment. The first frame of each sequence is coded with the intra-coding mode (i.e. treated as the I frame) while the remaining frames are coded with forward prediction (i.e. treated as the P frame). The performance comparison of MPEG1, the pyramid-structured wavelet residual coder (PY) and the R-D optimized WP residual coder is summarized in Table 3. Since both wavelet coders are embedded coders, we can adjust their bit rates to be the same as that of MPEG1. The table shows that the WP coder significantly outperforms MPEG1 with a gain of 1.0 to 2.7dB. The WP coder also outperforms the PY coder by 0.2dB, 0.6dB, 0.4dB, 0.4dB for "Flower", "Mobile", "Cheer" and "Bicycle" sequences, respectively. For "Football" and "Table Tennis", the performances of the WP and the PY coders are about the same. We conclude

that the performance of the WP coder is often better than that of the pyramid wavelet coder for motion compensated residual coding.

7 CONCLUSIONS

A new rate-distortion (R-D) optimized wavelet packet (WP) decomposition has been studied in this work. We analyzed the R-D performance of the wavelet coefficient coding and showed that it satisfies an exponential R-D relationship with a nearly constant decaying parameter β. Thus, the constant R-D slope criterion for optimum bit allocation can be converted to the constant distortion criterion, which can be implemented more conveniently. The layered zero coding technique is used to encode the coefficients obtained from the R-D optimized WP decomposition. The superior performance of the R-D optimized WP coder for still image and video coding is demonstrated by extensive experimental results. It turns out to be competitive to any state-of-the-art coders.

8 ACKNOWLEDGEMENT

This work was supported by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, and National Science Foundation Presidential Faculty Fellow Award ASC-9350309.

9 REFERENCES

[1] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, Vol. 1, pp. 205–230, Apr. 1992.

[2] T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree structured wavelet transform," *IEEE Trans. on Image Processing*, Vol. 2, No. 4, pp. 429–441, Oct. 1993.

[3] R. R. Coifman and M. V. Wickerhauser, "Entropy based algorithms for best basis selection," *IEEE Trans. on Inform. Theory*, Vol. 38, pp. 713–718, Mar 1992.

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley-Interscience, 1991.

[5] D. L. Duttweiler and C. Chamzas, "Probability estimation in arithmetic and adaptive huffman entropy coders," *IEEE Trans. on Image Processing*, Vol. 4, pp. 237–246, Mar. 1995.

[6] C. Herley, J. Kovvacevic, K. Ramchandran, and M. Vetterli, "Tilings of the time-frequency plane: construction of arbitrary orthogonal bases and fast tiling algorithms," *IEEE Trans. on Signal Processing*, Vol. 41, pp. 3341–3358, Dec. 1993.

[7] ISO/IEC-JTC1/SC2/WG8, "Digital compression and coding of continuous-tone still images," ISO Standard CD 10918, ISO, Jun. 1991.

[8] J. Li, P.-Y. Cheng, and C.-C. J. Kuo, "On the Improvements of Embedded Zerotree Wavelet (EZW) Coding," in *SPIE: Visual Communication and Image Processing'95*, vol. 2501, (Taipei, Taiwan), pp. 1490–1501, May 1995.

[9] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. on Acoustic, Speech, and Signal Processing*, Vol. 37, pp. 2091–2110, Dec. 1989.

[10] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: an estimation-theoretic approach," *IEEE trans. on Image Processing*, Vol. 3, No. 3, Sep. 1994.

[11] K. Ramchandran and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. on Image Processing*, Vol. 3, pp. 160–175, Apr. 1993.

[12] K. Ramchandran, Z. Xiong, K. Asai, and M. Vetterli, "Adaptive transforms for image coding using spatially-varying wavelet packets," *Submitted to the IEEE Trans. on Image Processing*, 1995.

[13] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 6, pp. 243–250, Jun 1996.

[14] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Processing*, Vol. 41, No. 12, pp. 3445–3462, Dec. 1993.

[15] P. Strobach, "Tree-structured scene adaptive coder," *IEEE Trans. on Communications*, Vol. 38, No. 4, pp. 477–486, Apr. 1994.

[16] D. Taubman and A. Zakhor, "Multirate 3-D subband coding of video," *IEEE Trans. on Image Processing*, Vol. 3, No. 5, pp. 572–588, Sep. 1994.

[17] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Trans. on Image Processing*, pp. 1053–1060, Aug. 1995.

[18] Z. Xiong, K. Ramchandran, M. T. Orchard, and K. Asai, "Wavelet packet-based image coding using joint space-frequency quantization," in *First IEEE International Conference on Image Processing*, (Austin, Texas), Nov. 13-16, 1994.