Detail Preserving Image Compression using Wavelet Transform

<u>Tham</u> Jo Yew[†], Student Member, IEEE

Abstract- In this paper, a complete compression and decompression algorithm for low bit rate still image coding is presented. It is a wavelet-based technique which first decomposes an image hierarchically into oriented subbands, and then encodes the wavelet coefficients using a zerotree data structure similar to that proposed by Shapiro [14-16]. We then incorporate four different enhancements to this original method to further improve both the objective measure of peak signal-to-noise ratio (PSNR) and the subjective perceptual quality of the reconstructed images. First, the optimum initial threshold is determined adaptively. Second, a "Quad-EZW" method which further decomposes the higher-frequency subbands is employed. Third, a method for predicting the higher-frequency coefficients is applied. Finally, a novel technique to translate the reconstruction values is introduced. With these enhancements, some performance improvements of up to 3 dB were achieved. It also totally outperforms JPEG, the current international standard for still image compression, especially at low and very low bit rates.

Keywords— wavelet transform, zerotrees, adaptive initial threshold, Quad-EZW, prediction of high-frequency, optimum translational factors, low bit rate compression.

I. INTRODUCTION

THE proliferation of digital technology has not only accelerated the pace of development of many image processing software and multimedia applications, but also motivated the need for better compression algorithms. Traditional transforms such as the discrete cosine transform (DCT) were employed successfully in JPEG [11],[19] (for lossy, continuous-tone gray scale or color still image compression), and in MPEG [8] (for motion picture compression). However, their inherent blocking artifacts are objectionable at higher compression ratios. In the past few years, the use of the wavelet transform for compression has been gaining wide popularity. The main advantages of wavelet-based methods lie in their energy compaction property and multiresolution decomposition capability.

Subband coding (SBC) is a waveform coding method first introduced by Crochiere *et al.* [7] in 1976 for medium rate speech coding. Recently, this technique has been extended to the coding of images and video sequences. The basic idea of subband coding is to split the spectrum of an image into non-overlapping bands of different resolutions (analysis stage). Since each subband has a reduced bandwidth, they may be downsampled. A more optimal bit allocation strategy can now be employed by assigning different number of bits to exploit the statistical properties of each subband. This allows the coding errors to be distributed across the subbands in a more visually optimal manner. Furthermore, different coding techniques can also be used for different subbands to reduce statistical re-

[†] The author is with the Department of Electrical Engineering, National University of Singapore. E-mail: engp5267@nus.sg. dundancies more efficiently. Reconstruction is achieved at the decoder end by adding upsampled and appropriately filtered versions of the subimages (synthesis stage).

The notion of using subband decomposition for image coding via a quadrature mirror filter (QMF) bank was first applied by Woods and O'Neil [20] in 1986. From then on, many variations that exploited this hierarchical tree structure were investigated. Shapiro [14], [15], [16] proposed a very economical means of representing insignificant information across scales by an elegant data structure called the *zerotrees*. He obtained very good image quality (PSNR = 30.23 dB, 0.125 bbp; PSNR = 33.17 dB, 0.25bbp)¹. Said and Pearlman [12],[13] extended this zerotree concept to a more general perspective by coding the state transitions instead of the actual states of the coefficients. A slight performance improvement (PSNR = 33.69 dB, 0.25 bpp) was obtained. Xiong et al.[21], introduced an approach for jointly optimizing both scalar quantization and tree-based quantization of pyramidal image decompositions (PSNR = 28.76 dB, 0.25 bpp, $256 \times 256 \text{ LENA}$). Banham and Sullivan [2] incorporated a quadtree segmentation based solely on the wavelet coefficients to code the image (PSNR = 32.17 dB, 0.54 bpp). Chang and Zakhor [4], [5], [6], and Taubman and Zakhor [17], [18] extended this multiresolution property to develop a scalable video compression scheme. By restricting the coding scheme to 2-D still images, very good results (PSNR = 30.96 dB, 0.125bpp; PSNR = 34.12 dB, 0.25 bpp) were obtained.

This paper is organized into two main parts. The first part (Section II) presents an overview of Shapiro's *Embedded Zerotrees of Wavelet Coefficients* (EZW) algorithm with slight variations. The second part is divided into four sections (Section III - VI), each explaining a different enhancement method of the EZW technique. Section VII concludes this paper with suggested future directions. Subjective improvements using the above enhancements are demonstrated in Section VIII.

II. THEORY OF EZW: A REVIEW

EZW is a new data structure proposed by Shapiro [14],[15],[16] for 2-D still image subband coding using wavelet transform. It was shown to produce excellent compression performance, both in terms of statistical peak signal-to-noise ratio (PSNR) and subjective human perception of the reconstructed image quality. It totally outperforms JPEG especially at low and very low bit rate compressions. The following subsections briefly outline the basic features and motivations of EZW for image com-

¹ All PSNR values quoted in this section is for the luminance component of a 512x512 LENA image, unless specified otherwise.

pression. A more elaborate algorithmic discussion with pseudo-codes and an example for both the encoder and decoder can be found in [10].

A. Basic Features of EZW Compression Method

As in most image compression methods, the EZW method consists of three main stages. They are the transformation stage, EZW with implicit quantization stage, and entropy coding stage.

In the transformation stage, a separable discrete 2-D wavelet decomposition is applied to allow a multiresolution analysis of the image. The first level of decomposition splits the image into four subbands consisting of one smooth (LL) subband, and three detail subbands - vertical (HL) subband, horizontal (LH) subband, and diagonal (HH) subband. Subsequent levels of decomposition further split the LL subband into four subbands, thus forming a hierarchical tree of directionally-sensitive subbands. Inherently the wavelet decomposition concentrates more than 95% of the total energy in the LL subband alone and less than 5% is distributed among all the detail subbands. This already suggests data compression because by just coding the few coefficients in this LL subband, one can preserve and transmit most of the information to the receiver/decoder. However, discarding the low-energy coefficients in the detail subbands will correspond to throwing away the high-frequency portion of the signal, and this is manifested as image blurring. This artifact motivates a more efficient and economical way for selecting coefficients that are deemed important for good reconstruction from both the smooth and detail subbands.

The second stage focuses on this critical selection process. An efficient compression algorithm will have to perform a very good job in selecting the significant coefficients² across all the scales. In doing so, the encoder must send both the positions and values of the significant coefficients to the decoder. Notice that such positional information of the significant coefficients can be sent efficiently if we can transmit the positions of the insignificant coefficients with minimal bits. EZW provides an economical zerotree data structure that predicts and implicitly encodes the positions of these insignificant coefficients across the scales. With the formation of zerotrees, many bits are saved for encoding the values of the significant coefficients. Furthermore, EZW does not adopt any explicit quantization technique, but instead, the precision levels of the significant coefficients are successively refined via multiple passes by comparing with certain thresholds. An initial threshold is chosen for the first round and its value is successively halved after each round. This successive approximation approach allows the generation of a single embedded bit stream that supports progressive transmission - a feature particularly useful for large image database browsing. In other words, such an embedded bit stream allows the encoding and the decoding processes to be stopped at any point without indicating the point of termination in the final reconstructed image. From a practical point of view, this feature is also found useful in many rate-constraint and distortion-constraint applications such as photo-journalism.

The entropy encoding stage consists of an adaptivemodel arithmetic encoder. By applying successive approximation, EZW coding produces only short alphabet input symbols to the arithmetic encoder. This allows the adaptive model to track changing symbol probabilities faster and learn quickly to improve its overall compression performance. Generally, arithmetic coders are more robust and can give a better coding performance as compared to Huffman coders which do not perform well for skewed input probability densities.

B. Parent-Child Relationship of Subbands

An interesting characteristic of recursive four-subband decomposition is the formation of spatial orientation trees as depicted in Fig. 1. Consider the shaded pixels in subbands HL3, HL2 and HL1, which form a spatial orientation tree (TREE) with a parent node at HL3. The corresponding four pixels with the same spatial orientation at HL2 will be the children (CHILD) of this parent node at HL3. Similarly, all the (2x2) pixels at HL2 and all the (4x4) pixels at HL1 are the descendants (DESC) of the same parent node at HL3. Notice also that a tree formed with a node at the base band LL3 (BASE) will have three main subtrees with nodes at HL3, LH3 and HH3, as shown by all the boxed pixels in Fig. 1 below. Mathematically, these inter-relationships can be written as:

$$DESC(n) = \bigcup_{m \in CHILD(n)} TREE(m)$$
 (1)

$$TREE(n) = \{n\} \cup DESC(n)$$
 (2)

$$\bigcup_{n \in \text{BASE}} \text{TREE}(n) = \{0, 1, 2, \dots, MN - 1\} \quad (3)$$

for all coefficients n of an $M \ge N$ image.

C. Formation of Zerotrees and A Priori Scanning Order

As described in Sec. II-B, the subbands of a decomposed image can be represented by spatial orientation trees rooted at the base band. EZW algorithm exploits this feature by introducing a data structure called *zerotree*. As pointed out in Sec. II-A, the bit budget should be spent economically to encode as many significant coefficients as possible, and ideally no bits are used to encode the insignificant coefficients. With this idea in mind, zerotrees provide an efficient way in representing these insignificant coefficients with minimal bits. We define a spatial orientation tree as a zerotree (with a root at coefficient c_n) if each element of the tree is deemed insignificant (smaller than the current threshold, T_i). In addition, a zerotree is formed only if it is *not* already part of a previously formed zerotree with root at c_{n-m} in the same pass with $T = T_i$, as described below:

$$\begin{aligned} \operatorname{ZTR}(c_n) &: \quad \forall_{k \in \operatorname{TREE}(c_n)} |c_k| <= T_i, \quad \text{and} \\ & c_n \notin \operatorname{ZTR}(c_{n-m})|_{T=T_i}, \quad m \in \mathcal{Z}^+ \end{aligned}$$

$$(4)$$

 $^{^{2}}$ Those coefficients which have higher energy content and are deemed to contain more information about the image.



Fig. 1. Parent-child relationship of subbands in a 3-scale hierarchical decomposition.

where c_{n-m} is scanned before c_n in the same pass with $T = T_i$. It is obvious that a lot of bits can be saved if all the completely predictable insignificant coefficients of this zerotree are sent to the decoder with only one ZTR symbol at the position of the zerotree root.

The viability of forming zerotrees is based on the hypothesis that if a wavelet coefficient at a coarse scale is insignificant, then all coefficients of the same orientation in the same spatial location at finer scales (i.e., the descendants) are also *likely* to be insignificant. This is directly related to the inherent energy packing property of wavelet decomposition. In order to be consistent with this hypothesis, an a priori scanning order is employed to scan the higher energy coefficients at the coarser scales first. Fig. 2 shows the scanning pattern for a 3-scale decomposition. It is obvious that no child coefficient is scanned before its parent.

D. Dominant and Subordinate Passes

In the following sections, we denote A_i as the state of Ain the *i*th round of the EZW process, where i = 1, 2, 3, To begin, let us define the *i*th round of passes R_i as consisting of a dominant pass DP_i , and a subordinate pass SP_i . In the beginning, all wavelet coefficients c_n are put into a list called the dominant list, $DL_{i=1}$, while another list called the subordinate list, $SL_{i=1}$, is empty. Once the initial threshold $T_{i=1}$ is determined, the first round of passes $R_{i=1}$ will begin. The dominant pass DP_i acts as a discriminating process to determine the significance of each c_n with respect to the current threshold, T_i . It is considered significant if its magnitude $|c_n|$ is larger than T_i , and insignificant otherwise. If c_n is significant, its sign is determined as either positive (POS) or negative (NEG). The encoder encodes this sign and sets its value in DL_i to zero to facilitate the formation of ZTR in subsequent rounds. Its magnitude $|c_n|$ is then transferred into



Fig. 2. A priori scanning order of subbands for both the encoder and decoder of EZW.

the subordinate list, SL_i . However, if c_n is insignificant, its descendant coefficients are checked for a zerotree root (ZTR) as explained earlier. If the zerotree formation fails, then this c_n is coded as an isolated zero (IZ). Notice that zerotrees cannot be formed in the finest scale (FINEST) of the decomposition. A special zero (Z) code is used to replace both IZ and ZTR at this scale. This is done to reduce the number of different possible symbols to be arithmetic encoded, thus further improving the overall compression performance. As a result, a DP_i will give out only three or four different possible codes as follows:

$$\operatorname{POS}(c_n)$$
 : $|c_n| > T_i$ and $c_n > 0$ (5)

$$\operatorname{NEG}(c_n) \quad : \quad |c_n| > T_i \quad \text{and} \quad c_n < 0 \tag{6}$$

$$\mathrm{ZTR}(c_n)$$
 : $orall_{k \in \mathrm{TREE}(c_n)} |c_k| <= T_i, ext{ and }$

$$c_n \notin \operatorname{ZTR}(c_{n-m})|_{T=T_i}, m \in \mathcal{Z}^+$$
 (7)
IZ (c_n) : $|c_n| \leq T_i$ and

$$\exists_{k \in \text{DESC}(c_n)} | c_k | > T_i \tag{8}$$

$$Z(c_n)$$
 : $|c_n| <= T_i$ and $c_n \in \text{FINEST}.$ (9)

After all c_n in the DL_i are discriminated, this DP_i will end and the SP_i begins. Now, each significant coefficient $|c_n|$ in SL_i will have a reconstruction value as will be seen by the decoder. By default, an insignificant coefficient will have zero as the reconstruction value. As a simple case, consider a coefficient with an actual value A that is to be successively approximated via the rounds of passes (see Fig. 3). After the DP_i , the center of the uncertainty interval of XY (i.e., M) is chosen as the reconstruction value having a certain precision. Notice that the uncertainty interval XY before the SP_i is equal in magnitude to the current threshold T_i , since $T_{i+1} = \frac{1}{2}T_i$. The subordinate pass aims to further refine the precision of the reconstruction value. During a SP_i , the uncertainty interval is halved (i.e., |XM| = |MY| = |XY|/2), by dividing XY into two halves. The new reconstruction value is determined as the center of this smaller uncertainty interval (i.e., R_L or R_U) depending on whether $|c_n|$ lies in the upper (UPP) or lower (LOW) half, respectively. In this example, since A lies in the lower half, the new reconstruction value will be R_L . It can be seen that the uncertainty interval now is halved and the reconstruction value is closer to A than before this SP_i . As a result, a SP_i generates only two different possible codes as summarized below:

$$UPP(c_n)$$
 : $|c_n|$ lies in the upper half of the
uncertainty interval, MY (10)

$$LOW(c_n)$$
 : $|c_n|$ lies in the lower half of the
uncertainty interval, XM. (11)



Fig. 3. Uncertainty interval of the reconstruction value as seen by the decoder. |XY| is equal in magnitude with T_i and it is the uncertainty interval for a reconstruction value M before the SP_i . After the SP_i , the uncertainty interval is halved to either |XM|or |MY| and the new refined reconstruction value is either R_L or R_U , respectively.

In this manner, the precision of the reconstruction value can be doubled after each SP_i . After all $|c_n|$ in SL_i are refined, the current SP_i will end. The next round of passes will begin with DP_{i+1} and followed by SP_{i+1} . This process of selection of significant coefficients in DP_{i+1} and refinement of uncertainty intervals in SP_{i+1} will continue until the bit budget is exhausted (rateconstrained), or a certain target distortion is achieved (distortion-constrained). It is apparent that the precision gets higher and higher approaching the exact value as *i* increases via successive approximation, which is rather similar in spirit to bit-plane encoding.

E. Reordering/Prioritization Protocol

By using orthogonal filter banks, any error introduced in quantizing the wavelet coefficients in the transform domain will eventually translate to a proportional amount of error in the reconstructed (spatial) domain. Therefore in order to optimize the utilization of bits, the bit budget should be used carefully to code those coefficients with higher information³ content (in this case, those higher energy coefficients). For example, suppose the decoder receives a transform coefficient of value $|c_n|$, the mean square error will decrease by $|c_n|^2/N$, where N is the image size. This concept motivates the need to reorder the significant coefficients in decreasing order of magnitude.

³Information provides an indication as to how much reduction in distortion is achieved after receiving that part of the coded message.

EZW implements this idea by reordering the significant coefficients in the subordinate list SL_i , before the refinement process is carried out in the subordinate pass SP_i . Notice that this reordering process cannot be done indiscriminately as the decoder will be unable to keep track of the reordered coefficients. Therefore, a prioritization scheme is used to avoid this. It is based on importance by precision, magnitude, scale, and spatial location as listed below:

- **Precision** this primary factor ensures the numerical precision of each significant coefficient. As the encoding process proceeds, the threshold T_i and the uncertainty intervals are getting smaller leading to an increase in precision value. Therefore, all coefficients in SL_i have to be refined to the same precision before any coefficient is refined further.
- Magnitude this refers to the magnitude of the reconstruction value as seen by the decoder. Coefficients with a higher reconstruction value are placed at the top of SL_i . However, those coefficients having the same reconstruction value cannot be moved in the list as these changes can never be known by the decoder. This requires them to be further prioritized according to scale.
- Scale this factor ensures that the positional information is maintained by being consistent with the same a priori scanning order adopted by both the encoder and decoder. This means that coarser scales are given higher priority than finer scales.
- Spatial location this factor is considered if there exist some coefficients having the same reconstruction value and belonging to the same scale. Then, higher priority is given to the coefficient which is scanned first. This is again consistent with the a priori scanning order to implicitly transmit the positional information to the decoder.

The gist of this prioritization protocol is that reordering is done by reconstruction magnitude according to the a priori scanning order. Precision is observed by ensuring that all significant coefficients in SL_i are refined in the current SP_i before they are refined further in the next SP_{i+1} .

F. Performance Comparison of EZW and JPEG

It was pointed out earlier that EZW coding produces very good reconstructed image quality as compared to the current international standard, JPEG, especially at low and very low bit rates. Fig. 4 compares the performance of JPEG against the EZW employing a biorthogonal wavelet filter. It is obvious that $JPEG^4$ always yields a PSNR value which is much lower than that using the EZW method. Figs. 7 and 8 depict the original images. More illustrative comparisons are displayed in Figs. 9, 10, 11 and 12 (in Sec. VIII).

⁴Results were obtained using John Bradley's xview program (version 3.0) [3].



Fig. 4. Performance comparison between JPEG and EZW with a biorthogonal wavelet filter (*without* any enhancements proposed in this paper) using 512x512 LENA image

III. Adaptive Finding of Optimum Initial Threshold

The selection of significant and insignificant coefficients is actually a thresholding process. In order to start the EZW coding, an initial threshold T_1 needs to be chosen. In general, threshold T_i can be written in the form $T_i = \alpha 2^k$, where $\alpha \in (0, 1]$, $k \in \mathbb{Z}$, and $i \in \mathbb{Z}^+$. A simplified approach (as adopted in Shapiro's EZW [14],[15],[16]) is to set $\alpha = 1$ and choose a k (say, k_o) such that:

$$\frac{1}{2}max \leq T_1 \leq max, \qquad T_1 \in \mathcal{R}^+.$$
 (12)

where max is the maximum magnitude of all wavelet coefficients. Dividing T_i by two simply decrements k while leaving α unchanged.

However, this method of determining T_1 is unlikely to be optimum in minimizing any given distortion function, and seems to be highly image dependent. Therefore in this enhancement method, we propose an adaptive approach to find an optimum T_1 that will minimize the error function (e.g., the mean square error) for a given bit rate as implemented in [10]. This corresponds to choosing a more appropriate value for α while using the same k_o as determined above. In the simulations, we divided the interval (as defined in Eq. (12)) into I equal parts, where I is a user-defined number of adaptive levels. For each of these I possible T_1 values, we iterate the EZW encoding process and the T_1 that gives the largest PSNR is finally chosen to be the *optimum* initial threshold in the actual EZW coding.

This adaptive process can be performed in the *trans*form domain as a direct result of energy conservation property using orthogonal wavelet filters. Any quantization errors introduced in the wavelet coefficients (in the transform domain) via successive approximation will correspond to a proportional⁵ amount of error in the re-





Fig. 5. Performance Comparison between Shapiro's EZW and EZW with adaptive initial threshold using 512x512 LENA image

constructed image (in the spatial domain). In this way, there is no need to go back and forth between the image and transform domains. Furthermore, the facts that EZW coding has short encoding and decoding CPU times (about 2.5 and 2.0 seconds respectively) and that no overhead is incurred at the decoder further motivate the practicality of this adaptive process of finding an optimum T_1 .

With this enhancement, PSNR improvement of up to 1 dB can be achieved for a wide range of compression ratios from 8:1 to 1024:1 as illustrated in Fig. 5. More importantly, improved subjective reconstructed image quality was observed in Fig. 14 as compared to the original EZW (without adaptive initial threshold) in Fig. 13.

IV. QUAD-EZW CODING WITH HIGHER-FREQUENCY SUBBAND DECOMPOSITION

As a result of energy compaction using wavelet decomposition, most of the encoded significant coefficients will come predominantly from the smooth (LL) subband, and almost none from the finer-scale subbands especially at very low bit rates. This results in poor edge quality and is manifested as severe image ringing and blurring. This level of quality is not very acceptable in most applications where high frequency details are perceptually important. Also, in some applications, edge-like information may be more desirable than the texture information. For instance, it will be very useful to count the number of levels of a building, or the number of vehicles on a street, even when the images are compressed at significantly low bit rates. In this context, it is less important to know the texture of the walls of the buildings, or the color of the vehicles.

As described in Sec. II-A, the original image is made up of four subbands (hence, called the *quadrants*) after the first level of conventional octave-band decomposition.

thogonal QMF filter [1] used by Shapiro) exhibits a discrepancy in PSNR values of only about 1% between the spatial and transform domains; hence, this still justifies the application of this adaptive process in the transform domain.

Subsequent L - 1 levels of decomposition will recursively decompose the LL quadrant only, while the other three quadrants are not further decomposed (see Fig. 6(a)). This enhancement method proposed here aims to preserve more edge information by further decomposing and encoding these three higher-frequency quadrants independently. As a result, each of these four quadrants will form its own hierarchical oriented tree structure in which *four* independent EZW processes can now be applied.

We should notice that the data in these higherfrequency quadrants are not as correlated as those in the LL (smooth) quadrant. As a result, further decomposition may not concentrate the energy in their respective smooth subbands. Three different methods of decomposing these higher-frequency quadrants were investigated. The first method (METHOD 1) performs the conventional octaveband decomposition to each of these three quadrants (see Fig. 6(b)). The second method (METHOD 2) also recursively decomposes each of the three higher-frequency quadrants for L-1 levels⁶, but this time, each quadrant is decomposed along its respective orientation direction (see Fig. 6(c)). For example, the HL quadrant is recursively decomposed along its HL subband direction. This directionally sensitive full-band decomposition is motivated by the idea of energy compaction in the respective orientation direction. It was found that the energy is not really concentrated in any particular subband when further decomposition is applied. Theoretically, if the original image has a very well-defined directional texture, then METHOD 2 is found to be more appropriate. The third method (METHOD 3) further decomposes only the HL and LH quadrants using 1-D subband decomposition (see Fig. 6(d)). For example, the HL subband is decomposed column-wise only as it contains predominantly vertical information. Similarly, the horizontal information in the LH subband is decomposed row-wise only. The HH subband is not decomposed any further as it has diagonal information, to which arguably, our eyes are insensitive. In this case, the HH quadrant can either be discarded⁷, or included into the EZW process for the LL quadrant.

Another major consideration in adopting this "Quad-EZW" approach is selecting the appropriate resolution level (i.e., scale) for further decomposition. Simply choosing the finest-scale (i.e., scale = 1) is definitely *not* the most suitable choice for different compression rates. As explained, predominantly only the smooth subbands of the LL quadrant are encoded at low bit rates. Suppose that, at such a low bit rate, the three quadrants at scale = 1 were chosen for further decomposition, this will leave a "gap" of which the finer scales of the LL quadrant are *not* encoded. In other words, only the smooth subbands of the LL quadrant and the three quadrants at scale = 1 are encoded. Since subbands at scale = 1 correspond to very fine textures of the original image, this will create "line-

Fig. 6. (a) Top Left: Conventional octave-band decomposition (b) Top Right: METHOD 1 (c) Bottom Left: METHOD 2 (d) Bottom Right: METHOD 3

drawing" artifacts around the edges of a blurred image. In order to ameliorate this problem, a more appropriate scale for further decomposition must be determined based on the target compression ratio. Generally speaking, at low bit rate compressions, a higher scale (e.g., scale = 2) should be chosen to avoid such a "gap". This is intuitively justifiable as each finer scale in this hierarchical decomposition can be regarded as representing the edge information of the previous coarser scale in the pyramid.

After further decomposition, each quadrant is now analyzed for its perceptual significance according to the requirements of the decoded images. A general guide is to compare the proportion of energy in each quadrant, and then allocate the available bits accordingly. In this manner, the original total bit budget can be allocated more "optimally" to preserve the desired edge information. Another more subjective approach is to allow the user to perform the bit allocation via an input interface. This, in fact, provides a useful feature for a more flexible compression scheme in which the user can have control over the bit allocation procedure to meet different needs.

Note also that since each EZW process is applied *in-dependently*, their initial thresholds are different. This allows the higher-frequency coefficients to be encoded despite their lower energy contents. By doing so, it actually contradicts the original aim of EZW to encode and transmit the more significant coefficients first. However, these low-energy coefficients should not be discarded as they are perceptually important for some applications. Although this *may* not maximize the PSNR value at a given bit rate, employing the "Quad-EZW" method preserves more edge information. For example it is easier to count the number of levels of the buildings in Fig. 16 (using METHOD 1) as compared to that in Fig. 15 which applies the original EZW algorithm.

⁶In fact, any suitable number of levels can be chosen. If L-1 is chosen, then each of the four main quadrants will have the same L-1 levels of decomposition.

 $^{^{7}\}mathrm{If}$ the energy in this subband is significantly lower as compared to the other subbands.

V. PREDICTION OF HIGH-FREQUENCY COEFFICIENTS

A direct consequence of conventional octave-band decomposition is that the more informative base band (higher energy content) is located at the coarsest scale, while the upper bands (lower energy content) at the finer scales contain the higher spatial frequencies. As explained in the previous section, these high frequencies are perceptually very important in certain regions although they contribute only a small proportion (about 1-3% for most natural images) of the total signal energy. Since these coefficients make up three quarters of the original image size, it is very expensive to code them. Nevertheless, discarding most of this high-pass information creates visible artifacts around the edges. This motivates us to preserve as many high-frequency coefficients as possible with minimum overhead.

Since the subbands are filtered and decimated versions of the same original image, this enhancement method predicts the locations of high-frequency coefficients by exploiting the relationship between subbands at different scales. This is done by analyzing the activity of the coefficients in the base band. Various activity indices, including simple edge detection techniques such as Sobel operator [9], were investigated. However in this enhancement method, we used an activity index based on the maximum amplitude difference between the greatest and smallest coefficients within a 3 x 3 neighborhood. An even better prediction is obtained using a directional activity measurement, which is motivated by the directionally-sensitive subbands. For example, prediction of active coefficients in the high-low (vertical) subband can be made by measuring the maximum amplitude difference along the rows of a 3 x 3 window. A pixel is classified as "active" if the maximum amplitude difference is larger than a predetermined threshold, and as "quiet" if it is smaller than that threshold. In general, more than two classes can be used. For example by specifying two different thresholds, three classes can be obtained (this is employed in our simulations). We also discovered that the average energy of the upper band pixels corresponding to locations classified as "active" can be up to 10 times higher than those classified as "quiet". Hence, better edges can be obtained by selectively encoding the predictably more energetic highfrequency coefficients.

This enhancement is used not only to predict the positions of the predictably active pixels, but also their reconstruction values. Each class is assigned *one* reconstruction value, which is chosen to be the mean of the coefficients belonging to the same class. The essence of this prediction method is that only the signs of the active pixels are sent explicitly to the decoder, while the positions are sent implicitly via a pre-determined scanning order. As the magnitude of each pixel is not transmitted, a number of bits are saved. However, more distortion could be introduced if the difference between the predicted reconstruction value and the actual magnitude is greater than the corresponding difference between the actual magnitude and zero (or the value of the coefficient without using this prediction method). Therefore the effectiveness of this prediction depends largely on the accuracy of the classifications, and the accuracy of the predicted reconstruction value for each class.

Some performance improvements can be obtained by employing this enhancement method. For example, we can observe better boundary or edge quality around the hat, shoulder, face and eyes of LENA in Fig. 18 as compared to Fig. 17 in which no prediction of high-frequency coefficients is made.

VI. Application of A Posteriori Translational Factors to EZW Coding

By applying EZW coding, significant coefficients are found and each coefficient will have a reconstruction value after the decoding process. Because of the fact that these reconstruction values are *not* the same as the actual values, distortion is introduced in the decoded image. We propose here a method to translate these reconstruction coefficients uniformly with respect to their corresponding actual wavelet coefficients in order to reduce the mean square error, and hence improve the decoded image quality.

As a simple example, consider the ideal case that all the reconstruction coefficients, $\hat{f}(i, j)$, (in the wavelet domain) of a particular subband are merely *shifted* by a certain fixed amount, k, from their corresponding actual wavelet coefficients, f(i, j) such that:

$$\hat{f}(i,j) = f(i,j) - k, \qquad (13)$$

where *i* and *j* denote the indices of the wavelet coefficients in the subband (or block of coefficients) of interest. In this case, the reconstruction values in the decoder will be *exact* if this entire block of reconstruction coefficients is translated back by this fixed amount, *k*, after⁸ the decoding process, but before the inverse wavelet transform is applied. In this section, we propose an enhancement method to find the optimum value of *k* that will reduce this mean error between $\hat{f}(i, j)$ and f(i, j). By keeping track of the reconstruction values (as seen by the decoder) in the encoder, these optimum values of *k* can be determined and transmitted as side information to the decoder.

It can be shown that this approach has the direct consequence of minimizing the mean square error (MSE), and hence maximizing the PSNR value. Recall that the definition of PSNR for an $M \ge N$ 8-bit (i.e., pixel intensity from 0-255) image is:

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i,j) - \hat{f}(i,j)]^2}$$
(14)

where f(i, j) and $\hat{f}(i, j)$ have the same notations as defined above. Again, we assume the *orthogonality principle* here as an error introduced in the wavelet coefficients will translate to a proportional amount of error in the spatial domain.

Suppose we define J_k as the criterion function to be minimized in order to maximize the value of PSNR, and k_{opt}

⁸Hence, we term it as the *a posteriori* translational factor.

as the translational factor that will optimize/minimize J_k as follows:

$$J_k = \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i,j) - (\hat{f}(i,j) + k)]^2.$$
(15)

By taking the gradient of J_k with respect to k, and setting to zero, the value of k_{opt} that minimizes the criterion function, J_k , is given by:

$$k_{opt} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i,j) - \hat{f}(i,j)].$$
(16)

It is shown that the value of k_{opt} is given by the mean difference between the actual and reconstructed wavelet coefficients. This result is, in fact, very intuitive because k_{opt} can be thought of as the center of gravity that best represents the data in the sense of minimizing the sum of squared distance from $\hat{f}(i, j)$ to f(i, j). Actually, this enhancement method is as good as introducing another quantization level⁹ to the wavelet coefficients.

We should also appreciate that the application of such a k_{opt} value can have two direct consequences. It can shift or translate a reconstruction value towards or away from the actual value. Since the value of k_{opt} is a "global" translational factor that shifts all the reconstruction values of a particular block of coefficients as one entity, some values will be shifted away, and some towards, the actual values. Nevertheless, such a shift with k_{opt} will, on the average, reduce the error between the actual and reconstructed signals, as shown in Equation (16) above. However, in some cases, the application of such a translation can result in noise-like artifacts. This could be explained by the fact that some coefficients are closer to their actual values if they are *not* translated than after the translation. By scaling down the values of k_{opt} by a certain factor (say, a factor of two or less), we can reduce the effect of shifting although, in doing so, the improvement in PSNR value will also drop.

However, it is not meaningful to calculate only one value of k_{opt} for the entire transformed image. This is because the decorrelated signal generally has a probability distribution that resembles a Laplacian distribution centered at zero. Graphically, we can view both the trailing ends of this distribution as corresponding to the significant coefficients found by the EZW process, while the center portion corresponds to the insignificant coefficients which have zero as the reconstruction values. Since the majority of the coefficients are still *insignificant*¹⁰ after the EZW encoding, the translational factor computed for the entire transformed image will have a value very close to zero as k_{opt} measures the mean error between these signals. As a result, we classify the coefficients of each subband according to their signs and significance into four subblocks with four translational factors - POS_SIG, NEG_SIG, POS_INSIG, and NEG_INSIG. This approach is motivated by the fact that the mean error of each subblock now is non-zero (especially the higher-frequency subbands which have mostly a zero as the reconstruction values).

Since at the end of the EZW decoding process, only the locations and signs of significant coefficients are known to the decoder, an additional overhead is needed to transmit the signs of *insignificant* coefficients to the decoder. Nevertheless, this side information can be efficiently encoded using run-length coding, followed by Huffman encoding. One may argue that this additional overhead can be avoided if only the significant coefficients are translated. This argument is definitely valid, but shifting only those few significant coefficients alone will not have much effect on the overall PSNR and subjective decoded image quality. Our simulations concluded that by translating the majority of the insignificant coefficients with the optimum translational factors (instead of a zero as the reconstruction values) yields a very significant improvement of up to 3 dB in terms of PSNR value and subjectively more pleasing decoded image quality due to reduced ringing effects and sharper edges. This improvement is obvious by comparing Figs. 19 and 20. However, it should be noted that the actual compression ratio for the decoded image in Fig. 20 is less than 64:1 due to the extra overhead.

VII. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we first reviewed the Embedded Zerotrees of Wavelet Coefficients (EZW) algorithm as proposed by Shapiro. A complete compression and decompression program for still image coding was presented. Slight modifications such as using a biorthogonal wavelet filter were made. Then we proposed four enhancements to the EZW coding to further improve both the objective measure of PSNR values by up to 3 dB and the subjective quality of decoded images. The first enhancement method adaptively searched for the optimum initial threshold to eventually yield a better reconstructed image when EZW coding is applied. The second method aimed to improve the high-frequency information by further decomposing the higher-frequency subbands and applying a "Quad-EZW" approach to code the significant coefficients of both the coarse and fine resolutions. The third method also improved the edge quality of the decoded image by predicting and encoding the higher-frequency subbands based on the activity in the coarser scales. The fourth method produced significant improvements, both in terms of higher PSNR values and better reconstructed image quality, by shifting blocks of wavelet coefficients with optimum translational factors prior to the application of inverse wavelet transform.

Many future directions can be investigated. A direct approach is to extend these methods to the coding of color images in which both the luminance and chrominance components can be coded. In addition, the visual masking effects based on the human visual system (HVS) can also be exploited. For example, the various subbands can be weighted by different factors according to their

⁹ This method of determining the quantization level that minimizes the MSE is similar to the Lloyd-Max quantizer [9], except that the quantization level computed using this enhancement method is tailored to a specific block of wavelet coefficients.

¹⁰Only about 2% of the total number of wavelet coefficients are significant for a compression ratio of 64:1.

perceptual significance before the EZW is applied. Another interesting direction is to combine both enhancement methods 3 and 4 (as described above), and assign the optimum translational factors (using method 4) only to the predictably active high-frequency coefficients obtained using method 3. This can reduce the extra overhead as well as the noise-like artifacts as mentioned in Sec. VI. In fact, these four enhancements can also be combined and analyzed. We have also extended this EZW method to video sequence coding using a motion-compensated 3dimensional zerotree data structure. Moreover this class of video compression algorithm is both multirate and multiresolution scalable. It was shown to retain very good reconstruction image quality even for low (about 128 kbps) and very low (less than 64 kbps) bit rates.

References

- Adelson, E. H., Simoncelli, E., and Hingorani, R., "Orthogonal pyramid transforms for image coding," in *Proc. SPIE*, Cambridge, MA, vol. 845, pp. 50-58, Oct. 1987.
- [2] Banham, M. R., and Sullivan, B. J., "A wavelet transform image coding technique with a quadtree structure," in *Proc IEEE ICASSP*, 1992.
- Bradley, J.,
 "XView Programming Manual," 3rd. Edition, Sept.
 1991 (latest update Aug. 1993). Also refer to URL
 http://www.unipress.com/cat/xview.html for more information about the XView Toolkit 3.0.
- [4] Chang, E., and Zakhor, A., "Subband video coding based on velocity filters," in *IEEE Inter. Conf. on* CAS, San Diego, pp. III:229-232, Mar. 1992.
- [5] Chang, E., and Zakhor, A., "Velocity-based architectures for 3-D subband video coding," in *IEEE Visual* Signal Processing Workshop, pp. 245-251, Sept. 1992.
- [6] Chang, E., and Zakhor, A., "Scalable video coding using 3-D subband velocity coding and multirate quantization," in *Proc. IEEE ICASSP*, Minneapolis, MN, vol. 5, pp. 574-577, Apr. 1993.
- [7] Crochiere, R. E., Webber, S. A., and Flanagan, J. L.,
 "Digital coding of speech in subbands," in *Bell. Syst. Tech. J.*, vol. 55, pp. 1069-1085, Oct. 1976.
- [8] ISO/IEC CD 13818-2, "Generic coding of moving pictures and associated audio," Nov. 1993.
- Jain, A. K., "Fundamentals of digital image processing," pp 99-131, Prentice-Hall, Eaglewood Cliffs, NJ, 1989.
- [10] Jo Yew, Tham, "Detail preserving image compression using wavelet transform," in *Electrical Engineering* Undergraduate Thesis, National University of Singapore, 1995.
- [11] Pennebaker, W. B., and Mitchell, J. L., "JPEG: Still image data compression standard," in New York, NY: Van Nostrand Reinhold, 1993.
- [12] Said, A., and Pearlman, W. A., "Image compression using the spatial-orientation tree," in *Proc. Int. Symp. Circuits, Syst.*, Chicago, IL, vol. 1, pp. 279-282, May 1993.

- [13] Said, A., "An improved zero-tree method for image compression," in *Report IPL TR-122*, Image Processing Lab., Rensselaer Polytechnic Institute, Troy, NY, Nov. 1992.
- [14] Shapiro, J. M., "An embedded wavelet hierarchical image coder," in *Proc. IEEE ICASSP*, San Francisco, CA, vol. IV, pp. 657-660, March 1992.
- [15] Shapiro, J. M., "Embedded image coding using zerotrees of wavelet coefficients," in *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445-3462, 1993.
- [16] Shapiro, J. M., "An embedded hierarchical image coder using zerotrees of wavelet coefficients," in Proc. Data Compression Conf., Snowbird, Utah, March 1993.
- [17] Taubman, D., and Zakhor, A., "Orientation adaptive subband coding of images," in *IEEE Trans. Image Proc.*, vol. 3, no. 4, pp. 421-437, July 1994.
- [18] Taubman. D., and Zakhor, A., "Multirate 3-D subband coding of video," in *IEEE Trans. Image Proc.*, vol. 3, no. 5, Sept. 1994.
- [19] Wallace, G. K., "The JPEG still picture compression standard," in Commun. ACM, vol. 34, pp. 30-44, April, 1991.
- [20] Woods, J. W., and O'Neil, S. D., "Subband coding of images," in *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 34, pp. 1278-1288, Oct. 1986.
- [21] Xiong, Z., Galatsanos, N. P., and Orchard, M. T., "Marginal analysis prioritization for image compression based on a hierarchical wavelet decomposition," in *Proc. IEEE ICASSP*, vol. V, pp. 546-549, 1993.

ACKNOWLEDGMENT

The author would like to express his sincere gratitude to his project supervisor, Dr. Surendra Ranganath, and his co-supervisor, Dr. Anthony Ng Peng Chew, for all their helpful guidance and advice.



Tham Jo Yew was born in Kuala Lumpur, Malaysia, in 1971. After obtaining his SPM (G.C.E. O-Levels equivalent) in 1988, he was awarded the PSC ASEAN scholarship to pursue his G.C.E. A-Levels in Hwa Chong Junior College, Singapore. He obtained his B.Eng. (Honours) degree in 1995 from the National University of Singapore (NUS). Currently he is a postgraduate research scholar in NUS, leading to a M.Eng. (Electrical) degree. Between 1989 and 1990 he was enthusiasti.

Between 1989 and 1990, he was enthusiastically involved in numerous mathematics com-

petitions. He received recognition in the Australian Mathematics Competition for the Westpac Awards. Having distinctive performance in the American High School Mathematics Competitions (AHSME), he represented Singapore in the American Invitational Mathematics Examination (AIME), and emerged as the champion team. In 1995, he received the NUS Engineering Innovation Award for his work on "Detail Preserving Image Compression using Wavelet Transform." He is also a member of the Malaysian MENSA Society since 1990. His current research interests involve wavelet-based compression techniques for very low bit rate still image and video sequence coding, video scalability, *M*-channel and multiwavelet filter designs and implementations.