USC–SIPI REPORT #251

JPEG-based Image Compression Using Wavelet Transform

by

Kwo-Jyr Wong, Po-Yuen Cheng and C.-C. Jay Kuo

March 1994

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

JPEG-based Image Compression Using Wavelet Transform*

Kwo-Jyr Wong† Po-Yuen Cheng‡ C.-C. Jay Kuo‡

March 3, 1994; EDICS - IP 1.1 Coding

Abstract

The block DCT (Discrete Cosine Transform) plays an important role in the JPEG still image compression standard. In this research, we investigate the replacement of the block DCT in the JPEG with the fully-decomposed wavelet-packet transform (FWT) while keeping all other building blocks the same with the objective to improve the performance of the JPEG. Following the FWT, two postprocessing operations, i.e. frequency band swapping and frequency/spatial coefficient rearrangement, are implemented. The JPEG quantization and coding techniques are then applied. The resulting method gives better performance than the DCT-based JPEG scheme in the MSE (mean square error) measure as well as visual appearances.

1 Introduction

The JPEG image compression standard has been widely used in industry. The block DCT (Discrete Cosine Transform) plays an important role in the JPEG. There exist some disadvantages with the block DCT. First, since the error caused by the quantization process is isolated in each spatial block, the block effect becomes obvious at a high compression ratio. Second, energy compaction is only achieved within each local block, and redundancy between blocks cannot be fully utilized. Thus, it is desirable to develop a transform which reduces the block effect, achieves global energy compaction, and allows fast computation. In this research, we consider the replacement of the block DCT in the JPEG with the fully-decomposed wavelet-packet transform (FWT) to improve the MSE measure and visual appearances.

The basic idea of our method can be briefly stated as follows. By performing a two-scale wavelet decomposition, we decompose an image into LL, LH, HL and HH four subbands. Then, by recursively applying the two-scale decomposition to all subbands, we obtain the fully-decomposed wavelet-packet transform (FWT), where each subband has the same block size and corresponds to a particular frequency channel. Following the FWT, we perform two postprocessing operations. The first is to rearrange the positions of subbands obtained from the FWT to have an energy distribution which decreases in a zigzag order. The second is to resample coefficients in the same position from all frequency subbands to form new blocks so that each new block corresponds to a particular spatial region in the original image and

*This work was supported by the National Science Foundation Presidential Faculty Fellow Award ASC-9350309.

†The author is with Chung-Shan Institute of Science and Technology, Lung-Tan, Taiwan 32505.

‡The authors are with the Signal and Image Processing Institute and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, California 90089-2564. E-mail: poyuench@sipi.usc.edu and cckuo@sipi.usc.edu.

each coefficient in the block corresponds to a frequency component in that region. Note that a block formed by this approach has a wider convolution area than that of a DCT block. Finally, we substitute the DCT-transformed block with the new block, and apply the same quantization and coding procedures as the JPEG.

This work is organized as follows. In Section 2, we give a brief review of the JPEG. Section 3 explains the details of replacing the block DCT with the FWT in the JPEG. Experiment results are shown in Section 4, and concluding remarks are given in Section 5.

2 Brief Review of JPEG

As early as 1982, people interested in setting up a standard technique for color image compression form a team in the International Organization for Standardization (ISO). Later, the International Telegraph and Telephone Consultative Committee (CCITT) joined to pursue the same goal. The united team was called the Joint Photographic Experts Group (JPEG). The specifications had been modified many times since 1988 and, after a lot of effort and discussion, the JPEG standard was finalized in 1992.

The operation mode of the JPEG includes both the DCT-based mode for lossy compression and the DPCM-based mode for lossless compression. The original image is decomposed into 8×8 blocks in the spatial domain. A fast discrete cosine transform (DCT) is used to transform the data into 64 transform coefficients. Based on the energy compaction property of the DCT, these coefficients are quantized by the luminance and chrominance (in the color case) quantization tables [9]. Next, the quantized data are encoded for DC and AC coefficients separately. The DPCM technique to applied to the DC coefficients whereas the AC coefficients are traversed in a zigzag order and encoded with a run-length code. Finally, the DC and AC output data are encoded by their own entropy encoders. The decoder system is a mirror version of the encoder. It uses the same quantization tables and entropy coding tables as those used in the encoder. A simplified encoder flowchart and decoder flowchart of JPEG are shown in Figs. 1 and 2, respectively.

3 JPEG-Based Image Compression Using Wavelet Transform

We show in this section the replacement of the block DCT with the FWT in the JPEG. The implementation of FWT is described in Section 3.1. Following the FWT, two postprocessing operations have to be performed to make the FWT output compatible to the coding requirement of the JPEG, namely, frequency band swapping and spatial-frequency coefficient rearrangement. They are detailed in Sections 3.2 and 3.3. Finally, in addition to the JPEG-suggested quantization table, we discuss another approach to generate a new quantization table by using the optimum bit allocation formula in Section 3.4.

3.1 Fully-decomposed Wavelet-Packet Transform (FWT)

The wavelet transform provides a multiresolution tool for signal analysis. The two-scale discrete wavelet transform of a sequence $f[n]$ can be viewed as passing the signal through a quadrature mirror filter (QMF) bank consisting of a low- and high-pass filter pair denoted, respectively, by $h[k]$ and $g[k]$ with $g[k] = (-1)^k h[1-k]$. Additional constraints can be imposed on $h[k]$ and $g[k]$ so that the resulting wavelet representation has some nice properties [7]. Wavelet transforms with various filter coefficients $h[k]$ have been proposed. They include the Haar basis, the family of Daubechies bases [2], and the spline wavelet basis [5]. For 2D signals, one can generalize the above idea by forming the tensor product of 1-D wavelet transforms along horizontal and vertical directions so that the signals can be decomposed into the low-low (LL), low-high (LH), high-low (HL), and high-high (HH) frequency channels. For details on the relationship between the wavelet transform and filter bank theory, we refer to [8].

There exist many possible ways to generalize the above two-scale wavelet transform to the multiple scale case. The conventional approach is to apply the two-scale decomposition recursively only to the lowest frequency channels. Since it can be efficiently implemented via a pyramidal computational algorithm [5], we call it the pyramid-structured wavelet transform. The second approach is to apply the two-scale decomposition to channels which contain a significant amount of energy or information. The structure of the computational algorithm can be well described by binary and quad trees for 1D and 2D cases, respectively. It is therefore called the tree-structured wavelet transform [1], which is also known as the wavelet-packet transform. Finally, we may consider the recursive application of the two-scale decomposition to all frequency channels, and call it the fully-decomposed wavelet-packet transform (FWT). Note that under the FWT, each subband has the same block size and corresponds to a particular frequency channel. The FWT applied to the Lena image of size 512×512 by using the Haar basis with two decomposition levels is shown in Fig. 3, where gray levels in each block are normalized to be between 0 and 255 for the ease of visualization.

Various wavelet bases and techniques in handling the boundary effect of subimages have been examined in [10]. The study shows the biorthogonal wavelet transform which has both the FIR (or compact support) and linear phase properties and the symmetric boundary extension of the subimage are useful in reducing the pseudo high frequency noise. The advantages of using the linear phase FIR filters and symmetric boundary extension can be explained below.

By transforming a signal from the spatial to the frequency domain, we often introduce different phase terms with respect to different frequency components. The phase terms exist in both the horizontal and vertical directions for 2-D images. It is well known that the phase term in the frequency domain corresponds to a spatial shift in the original image domain. If the linear phase holds in both horizontal and vertical directions, the quantization error in the frequency domain will keep the same shift amount in the 2-D spatial domain so that the spatial shape of the object in the original image can be preserved. In contrast, if the filter does not have a linear phase, the quantization error will introduce different amount of spatial

shift for different frequency components and the object shape in the original image will be distorted. To give an example, the edges of an object become blurred. To use the linear phase FIR filters in the filter bank, we have to choose the biorthogonal wavelet transform.

By performing the wavelet decomposition on an image, we often assume that the image is periodically extended. This is equivalent to a circulation convolution operation. But the circular convolution method will cause the pseudo high frequency along the image boundary in the transform domain. As a result, we may see some artifact in the decompressed image, say, the ringing effect near the edge region. By symmetry boundary extension, we extend the image by treating the image boundary as a mirror. The pseudo high frequency can be significantly reduced with such an extension.

3.2 Frequency Band Swapping

The zig-zag scanning used in the JPEG takes advantage of the fact that most images have high energy in the low frequency bands and low energy in the high frequency bands. However, due to the down-sampling and up-sampling operations in the wavelet transform process, the positions of low and high frequency bands have been changed and the property that energy decays from low to high frequency bands is lost. Frequency band swapping is performed to recover the desired energy decay property.

To understand the frequency band swaping, let us examine the 1D case. Consider a signal $x(n)$ down-sampled by 2. The input-output relation can be written as

$$y(n) = x(2n).$$

The corresponding Fourier transform in the frequency domain is

$$Y(e^{j\omega}) = \frac{1}{2}\sum_{k=0}^{1} X(e^{j(\omega/2-2\pi k/2)}),$$

where $Y(e^{j\omega})$ and $X(e^{j\omega})$ are the Fourier transforms of the output $y(n)$ and input $x(n)$, respectively. It is clear that the spectrum of the down-sampled signal has two copies of $X(e^{j\omega})$ with frequency stretched by 2 and shifted by 0 and 2π. Let the original signal consist of a lowpass filtered signal $x_L(n)$ with energy distributed from $-\pi/2$ to $\pi/2$, and a highpass filtered signal $x_H(n)$ with energy distributed from $-\pi$ to $-\pi/2$ and from $\pi/2$ to π. After the downsampling operation, the low-frequency signal $x_L(n)$ will have energy in the interval from π to $-\pi$ with a stretched shape of the original frequency band. However, the high-frequency signal $x_H(n)$ will have energy in the intervals from -2π to $-\pi$ and from π to 2π, which are in fact wrapped around to the baseband with a reversed and stretched frequency shape. Thus, we have to swap the frequency bands for the highpass filtered signal to keep the energy distribution decay smoothly.

For 2D separable FWT, the frequency band swapping can be easily performed with the following rule. Consider a 2-level decomposed image with 16 subbands. Let Qi, $i = 1 \ldots 4$, denote the 4 larger subbands as shown in Fig. 4, corresponding to LL, HL, LH and HH frequency channels due to the 1st level decomposition. At the 2nd level decomposition, we decompose each Qi into another 4 smaller subbands.

We do nothing in $Q1$ because it corresponds to the LL frequency channel. The left and right subbands in $Q2$ are exchanged, since $Q2$ corresponds to the high frequency channel along the horizontal direction. Similarly, the top and bottom subbands in $Q3$ are exchanged. Finally, we swap the subbands of $Q4$ along the diagonal direction. The swapping operation is illustrated in Fig. 4. The frequency block swapping should be recursively performed.

The energy distribution of Lena image before and after frequency band swapping are shown in Fig. 5 (a) and (b), respectively. As expected, the energy distribution becomes approximately monotonically attenuated from d.c. to the high-frequency portion after the swapping operation. Note that to compensate the frequency band swapping applied in encoding, the inverse frequency band swapping has to be applied in decoding.

3.3 Frequency/Spatial Coefficient Rearrangement

In Section 3.2, we showed the swapping of frequency bands so that the energy distribution is approximately monotonically decreasing from the d.c band to the highest frequency band which is the band located in the bottom right corner of the transform domain. However, at this stage, the block still corresponds to a energy band whereas the FWT coefficients within each band corresponds to a spatial area in the original image. In contrast, by using the block DCT, we have each block corresponds to a spatial region and each coefficient corresponds to a frequency component. Thus, in order to be compatible with the coding requirement of the JPEG, we have to rearrange the swapped FWT coefficients so that the newly formed block corresponding to a spatial region with different frequency components.

This rearrangement can be done easily. For simplicity, we use a simple image of size 16×16 to illustrate the idea of frequency/spatial resampling. Suppose the image is decomposed into 16 blocks of size 4×4 via FWT. The transformed coefficients after frequency band swapping is shown in Fig. 6. The coefficients in the same position of each frequency band are collected in the same block as shown in Fig. 7, where the order of the coefficients are kept the same as the original order in their frequency bands. Note that the new block is similar to the transformed DCT block in the JPEG but with a wider convolution range.

3.4 Quantization Table Selection

After the above steps, we can follow the same quantization and coding procedures given by the JPEG. They include the quantization with a quantization table, the DPCM coding of the DC coefficients, the zigzag scanning and run-length coding of the AC coefficients, and entropy coding of all data. The quantization table suggested by the JPEG was obtained empirically based on the subjective visual quality. We will discuss another simple approach to choose the quantization table in this subsection.

The JPEG applies a uniform quantizer to quantize the DCT coefficient in the same position of a block (i.e. the same frequency component). However, different step sizes are used for coefficients in different positions, and they are summarized in a quantization table. The table can be defined by the encoder

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Table 1: Standard quantization table used frequently in JPEG.

user, and must be sent to the receiver end accompanied with the compressed image data. Usually the quantization table is generated empirically without taking the rate-distortion compromise into account. Although the JPEG standard does not specify the values in the quantization table, a typical quantization table widely used for the luminance signal is shown in Table 1. This table was obtained based on extensive testing by members of the JPEG committee. It often provides a quite good performance. The user may scale the table and round the values to different integers to achieve either higher compression ratio or lower distortion. By choosing a large step size, we get a high compression ratio at the expense of a poor quality of the decompressed image.

Obviously, Table 1 is not the only good choice for quantization. In general, we want a finer step size at a lower-frequency band and a coarser step size in a higher-frequency band. In this work, we use the idea of optimum bit allocation formula to generate a different quantization table. Before the frequency/spatial coefficient rearrangement, the energy in each frequency band is calculated. Let the energy in band (k,l) be $\sigma^2(k,l)$. Then, we can assign bits to each frequency component with the bit allocation formula [3]

$$B_{k,l} = \bar{B}_a + \frac{1}{2}\log_2\sigma^2_{k,l} - \frac{1}{2(R^2-1)}\sum_k\sum_l \log_2\sigma^2_{k,l},$$

where the value of $\bar{B}_a$ is set to make sure that no negative value appears in $B_{k,l}$. A good choice of $\bar{B}_a$ is 4. Then, we convert the allocated bit value to the quantization step size via

$$L_{k,l} = \frac{\mu}{B_{k,l}},$$

where μ is a constant. A quantization table generated for the Lena image is listed in Table 2. Note that we call it the adaptive quantization table since it is image-dependent.

The value of the constant μ depends on the desired image quality. It plays a similar role as that of the "quality factor" used in the JPEG. When we want to achieve a higher compression ratio, this value is set larger. As a result, the slope of the quantization matrix becomes steeper, and the difference of the quantization step sizes among the low- and high-frequency components becomes larger accordingly. This means the high-frequency components are discarded more easily and the total amount of encoded data is reduced.

16	13	17	19	24	28	32	41
16	16	19	21	27	32	36	46
22	20	23	24	32	39	42	55
27	26	27	25	33	42	40	51
39	35	36	34	43	55	58	70
48	44	47	41	58	71	66	77
49	49	54	43	55	69	57	70
61	67	75	59	76	82	72	85

Table 2: Adaptive quantization table for the Lena image.

4 Experimental Results

We have performed the compression experiment on many test images. We are particularly interested in the comparison of the following three methods:

1. the DCT-based JPEG with the standard quantization table as given in Table 1,
2. the FWT with the standard quantization table,
3. the FWT with the adaptive quantization table.

With methods 1 and 2, we compare the performance difference between the block DCT and FWT. With methods 2 and 3, we study the performance difference due to different quantization schemes.

The results of two typical test images, i.e. the Lena and creek images of size 512×512, are reported in Figs. 8 (a) and (b) where the PSNR (Peak Signal to Noise Ratio) values are plotted at various compression ratios. For these data, the 9-tap biorthogonal quadrature mirror filter bank [6] was used to implement the FWT. We have the following observations based on Figs. 8 (a) and (b). The FWT has a higher PSNR value than the DCT in all ranges of compression ratios. The difference becomes larger as the ratio goes higher. As far as the quantization table is concerned, it is attractive to use the adaptive quantization table only when the compression ratio is low. When the compression ratio is high, the quantization tables makes little difference.

To illustrate the visual appearance of the distortion, we show the decompressed Lena and creek images in Figs. 9 and 10, respectively, by using the JPEG and the FWT. The block effect is very obvious in the smooth regions for the JPEG when the compression ratio is high. With the FWT method, the block effect is significantly reduced. The distortion in the smooth region is much less visible. However, we do see distortion clearly in regions around the edges.

Different wavelet filters are also compared in our experiment. We show the results of the Lena image in Fig. 11 with several wavelet bases, including biorthogonal spline variant [2], spline variant [4], and 9-tap quadrature mirror filter [6]. All these wavelets have a linear phase filter response. We see that the 9-tap quadrature mirror filter gives the best result for this test image.

5 Conclusion and Extensions

The proposed algorithm have several advantages over the JPEG. First, the FWT gives a better performance than that of the DCT-based JPEG both in the MSE measure and visual appearance. By using the JPEG, the block effect of decompressed images is very obvious at higher compression ratios. We exploit more global information by using the FWT so that the block effect can be reduced significantly. Another potential advantage is that we have the flexibility of choosing different wavelet bases. It may be possible to improve the performance furthermore by designing a more suitable wavelet basis, say, 2D nonseperable wavelet basis. It seems possible to finetune the quantization table based on human perception. This is under our current investigation. This work has focused on the compression of monochrome images. We are interested in extending the proposed method to color images in the near future.

References

[1] T. Chang and C. C. J. Kuo, "A wavelet transform approach to texture analysis," in *Proc. ICASSP-92*, pp. 661–664, 1992.

[2] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Principles and Applications to Speech and Video*, Prentice-Hall Inc, 1984.

[4] P. M. M. Antonini, M. Barlaud and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Processing*, Vol. 1, pp. 205–220, Apr. 1992.

[5] 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.

[6] T. Senoo and B. Girod, "Vector quantization for entropy coding of image subbands," *IEEE Trans. on Image Processing*, Vol. 1, pp. 526–532, Oct. 1992.

[7] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Review*, Vol. 31, pp. 614–627, Dec. 1989.

[8] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Trans. on Signal Processing*, Vol. 40, pp. 2207–2232, 1992.

[9] G. Wallace, "The JPEG Still Picture Compression Standard," *Communications of the ACM*, Vol. 34, No. 4, pp. 31–44, 1991.

[10] K.-J. Wong and C.-C. J. Kuo, "A full wavelet transform (FWT) approach to image compression," 1993 IS&T/SPIE Symposium on Electronic Imaging: Science and Technology, San Jose, California, Jan. 31-Feb. 5, 1993.

Figure Captions

Figure 1: Block diagram of the JPEG encoder.

Figure 2: Block diagram of the JPEG decoder.

Figure 3: The fully-decomposed wavelet-packet transform (FWT) of the Lena image with 2 decomposition levels.

Figure 4: Frequency band swapping.

Figure 5: Energy distribution of the Lena image (a) before and (b) after the frequency band swapping.

Figure 6: Transform coefficients after frequency band swapping.

Figure 7: Transform coefficients after spatial/frequency rearrangement.

Figure 8: Performance comparison of the JPEG and the FWT.

Figure 9: Visual comparison of the decompressed Lena image: (a) JPEG and (b) FWT with compression ratio 28:1 (0.28bpp).

Figure 10: Visual comparison of the decompressed creek image: (a) JPEG and (b) FWT with compression ratio 32:1 (0.25bpp).

Figure 11: Performance comparison of the FWT with different wavelet bases.

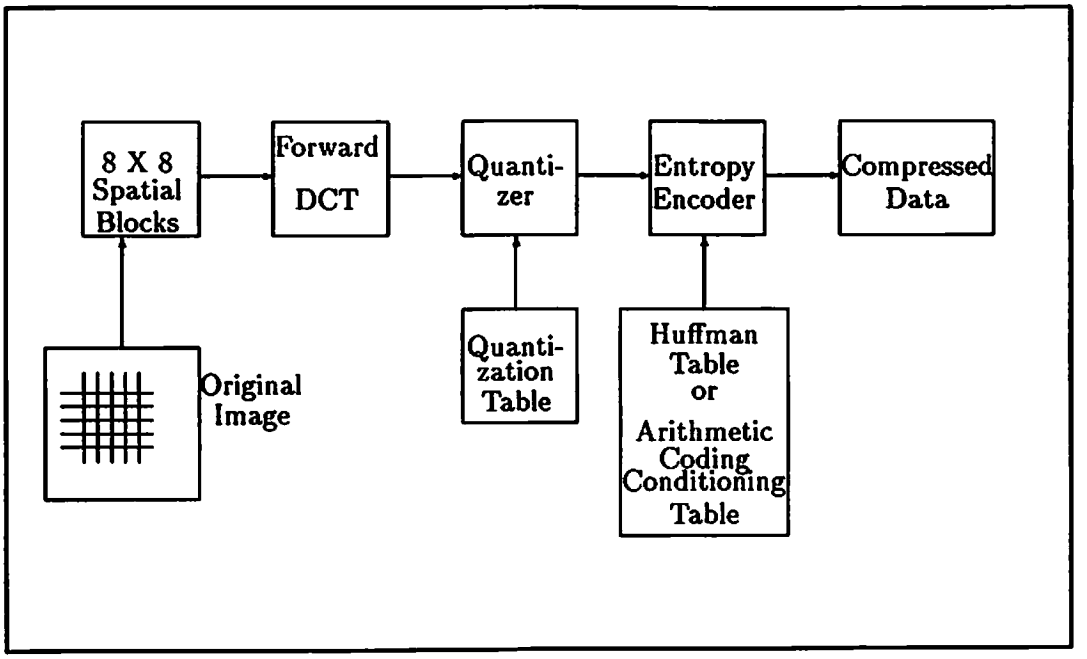


Figure 1: Block diagram of the JPEG encoder.

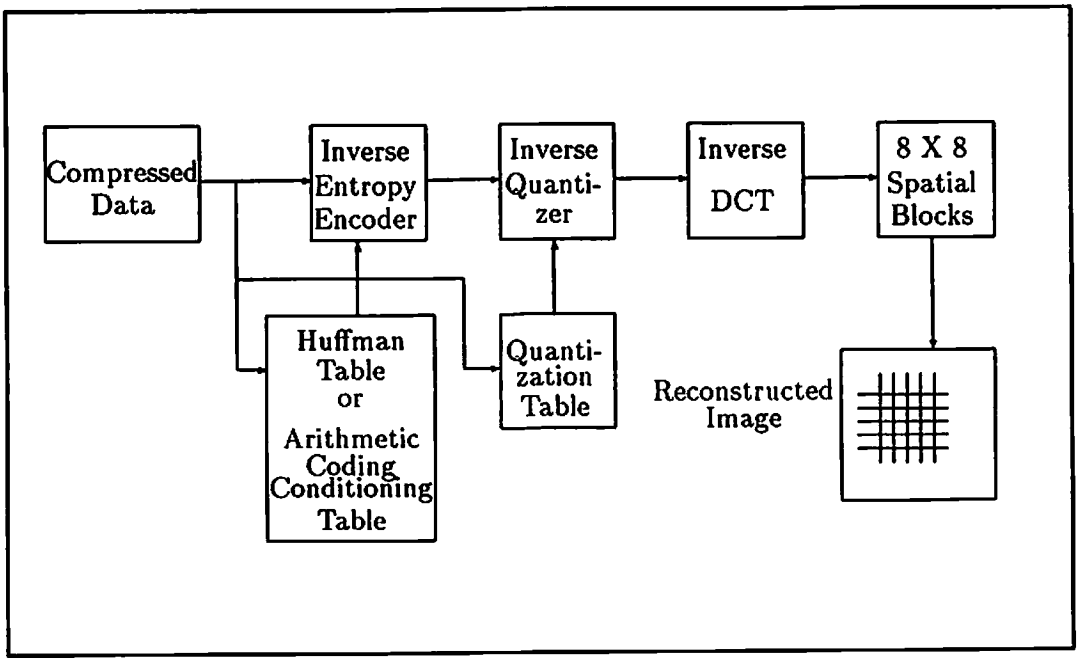


Figure 2: Block diagram of the JPEG decoder.

Figure 3: The fully-decomposed wavelet-packet transform (FWT) of the Lena image with 2 decomposition levels.

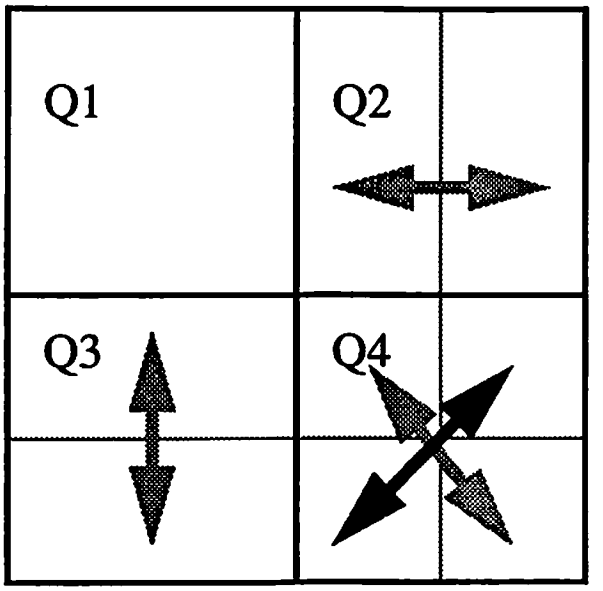


Figure 4: Frequency band swapping.

(a) (b)

Figure 5: Energy distribution of the Lena image (a) before and (b) after the frequency band swapping.

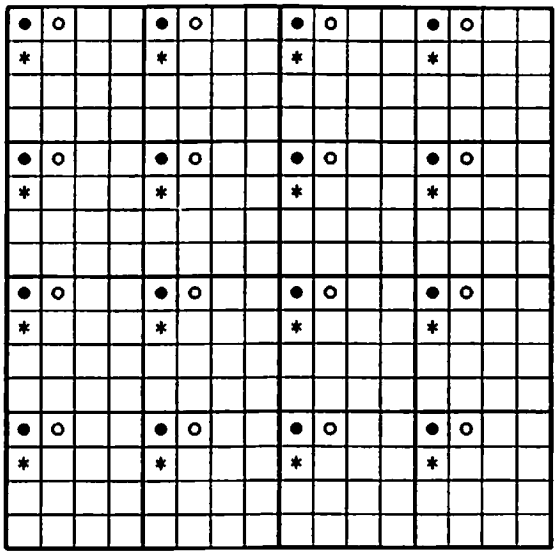


Figure 6: Transform coefficients after frequency band swapping.

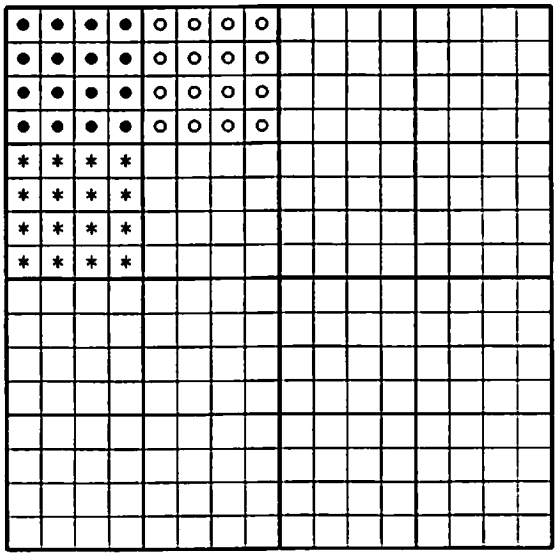
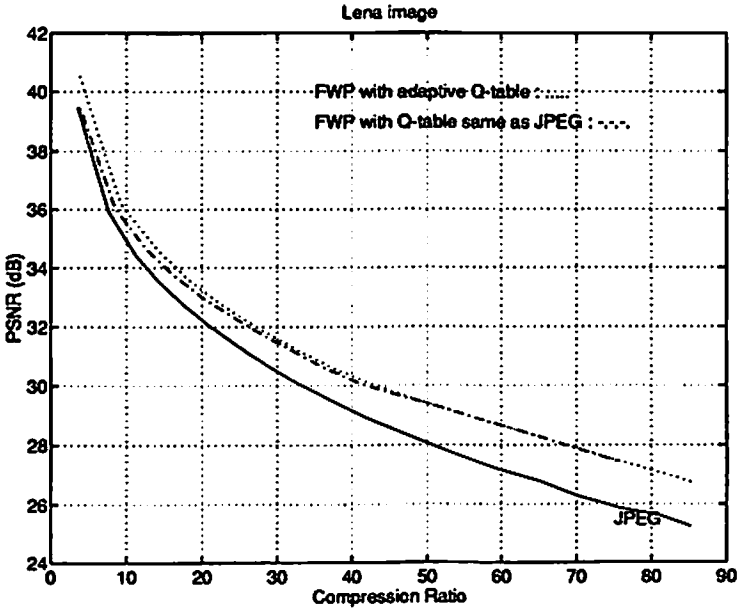
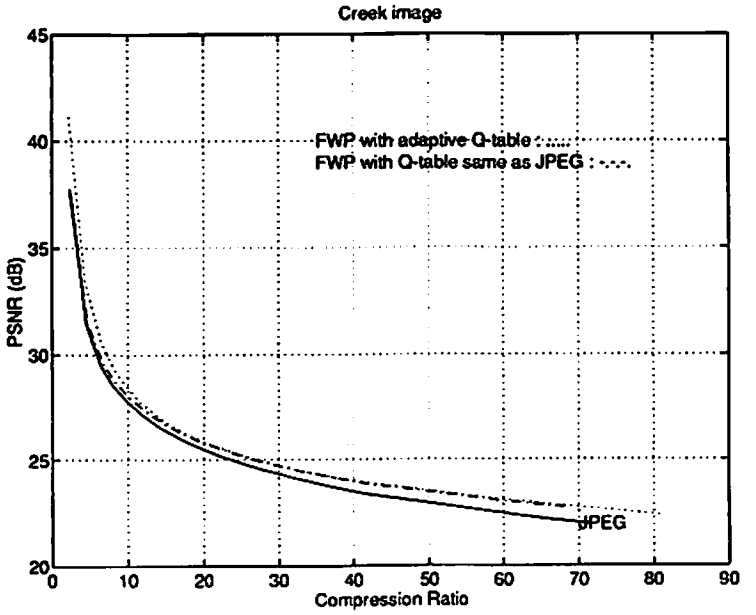


Figure 7: Transform coefficients after spatial/frequency rearrangement.



(a) Lena



(b) Creek

Figure 8: Performance Comparison of the JPEG and the FWP

(a)

Figure 9: Visual comparison of the decompressed Lena image: (a) JPEG and (b) FWT with compression ratio 28:1 (0.28bpp).

(b)

(a)

Figure 10: Visual comparison of the decompressed creek image: (a) JPEG and (b) FWT with compression ratio 32:1 (0.25bpp).

(b)

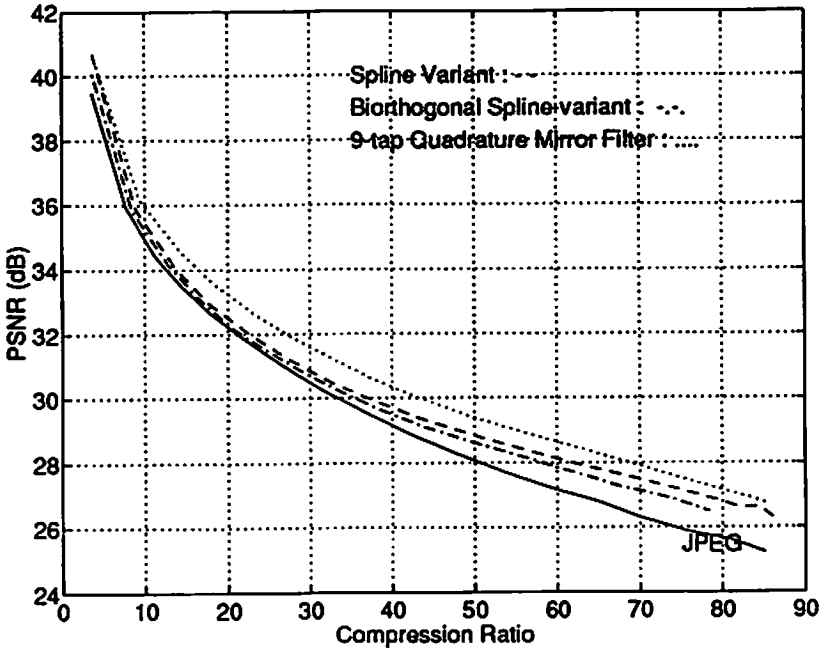


Figure 11: Performance comparison of the FWT with different wavelet bases.