

Thesis for the Degree of Master of Engineering

Image Compression Using Wavelet-Vector Quantization Algorithms

by

Yongri Piao

Department of Telematics Engineering

Graduate School

Pukyong National University

August 2005

Image Compression Using Wavelet-Vector Quantization Algorithms

웨이블릿과 양자화 알고리즘에 의한



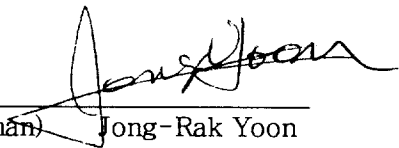
A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
in the Department of Telematics, Graduate School,
Pukyong National University

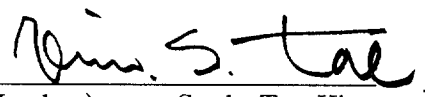
August 2005

Image Compression Using Wavelet-Vector
Quantization Algorithms

A Dissertation
by
Yongri Piao

Approved as to style and content by:


(Chairman) Jong-Rak Yoon


(Member) Seok-Tae Kim


(Member) Moon-Gab Joo

August, 2005

CONTENTS

I.	Introduction	1
II.	Wavelet Transform and Huffman Coding	4
1.	Continuous Wavelet Transform	5
1.1	Definition	5
1.2	Two-Dimensional CWT	7
1.3	Filter Bank Interpretation	8
1.4	Two Dimensional Filter Banks	10
2.	Discrete Wavelet Transform	12
2.1	Subband Coding and Decoding	12
2.2	Fast Wavelet Transform Algorithm	14
2.3	Discrete Wavelet Transform Design.....	16
2.3.1	Scaling Vector	17
2.3.2	Wavelet Vector.....	18
2.4	Two-Dimensional Discrete Wavelet Transform ..	19
2.5	Biorthogonal Wavelet Transform	20
2.5.1	Implementation.....	21
2.5.2	Biorthogonal Wavelets.....	22
3.	Huffman Coding for Image Compression	23
III.	Image Compression Algorithm	29
1.	Pre-Quantizer.....	29
2.	Wavelet Transform	32
3.	Huffman Coding with The Predictor	35
IV.	Simulation Results	37
V.	Conclusion.....	41
	References	42
	Acknowledgements	45

Image Compression using Wavelet-Vector Quantization Algorithms

Yongri Piao

Department of Telematics Engineering, Graduate School
Pukyong National University

Abstract

This paper proposes the method to images of losses using restorable Wavelet Transformation. The algorithm proposed in this work starts from processing the Pre-quantizer on the original images to organize an image that matches the gray level. The Wavelet Transformation filters the original image which is already pre-quantized in order to segment bands. Considering the lowest coding of bands influencing the most to the overall condition of the reconstructed image, it uses the Huffman coding using prediction. Reconstructed images by proposed algorithm shows higher PSNR than coding images of JPEG or non Pre-quantized images. Pre-quantizer can control the peak errors and is expected to be useful at mass image compression.

I. Introduction

With the development of the society, human needs more and more information in exchanging image, voice, and various medias. Using digital techniques to process image and voice, human needs to do a huge amount of work. Saving and transforming the data directly not only costs a lot of money, but also is not supported by equipments. That is because there are many redundancies in data. Employing the current compression techniques can achieve 3 or 4 times of compression rate or tens or hundreds times compared to the original data capacity.

Digital image compression techniques can be divided into lossless and loss compression techniques. Lossless compression technique consists of DPCM (differential pulse code modulation) and PCM (pulse code modulation). Traditional digital image compression technique mainly predicts PCM, DPCM[13-14], BKC (block truncation coding), VQ (vector quantization)[11-12], HC (hierarchical coding), SC (subband coding)[18-20], TC (transform coding)[11-15] and etc. International standard of image compression successfully adopted one or more compound compression techniques, for example, JBIG, JPEG, MEPEG-1, MPEG-2, MPEG-4, and developing MPEG-7.

Generally, using DPCM to compress an image is limited, and is weak

against channel errors. However, JPEG[16-17] is the compression standard to process color or binary images. Utilizing JPEG can achieve higher compression ratio, and also can change the form of images in coding process. In addition, reconstruction images can make Block effect horizontally and vertically. That is because DCT[11-15] transforms and quantizes the original image.

In 1980, a French scientist, Morlet, first presented Wavelet transform [1-10]. Comparing Fourier analysis, Wavelet transform is the local transform of time and frequency. It obtains signals more effectively and analyses local signals. Using Wavelet transform can achieve higher compression ratio, but no block effects. Aside from that, Wavelet transform possesses shortening, and horizontally shifting images and produces various resolution pictures.

Wavelet transform not only has the advantage of Fourier transform, overcomes its drawbacks and conquers some disadvantages of it. Wavelet transform is broadly used in the still image and video compression domain and it has become an important algorithm of the international standard of some image compressions.

This paper proposes the method for images of losses using restorable Wavelet transformation. The algorithm proposed in this work starts from processing the pre-quantizer on the original images to organize an image

that matches the gray level. The Wavelet transformation filters to the original image which is already pre-quantized in order to segment bands. Considering the lowest coding of bands influencing the most to the overall condition of the reconstructed image, it uses only the Huffman coding with prediction. Reconstructed images by proposed algorithm shows higher PSNR than coding JPEG or non pre-quantized images. Applying pre-quantizer can control the peak errors and can also be expected to be useful for mass image compression.

II. Wavelet Transform and Huffman Coding

Considerable interest has arisen in recent years regarding to new transform techniques that address specifically the problems of image compression, edge or feature detection and texture analysis. These techniques come under the headings of multi-resolution analysis, time-frequency analysis, pyramid algorithms, and Wavelet transforms [1].

In this paper, we review some of the limitations of the traditional Fourier and similar transforms and define three types of Wavelet transforms that promise improved performance for certain applications. We trace some of the developments that have led to the current state of Wavelet analysis, nothing similarities that tend to unify these different approaches under the banner of Wavelet transforms. Later in the paper, we illustrate some of the applications of Wavelet transforms.

We restrict ourselves to transforming real-valued, measurable, square-integrable functions of one and two dimensions, since these encompass the signals and images that are of interest to us. As before, we introduce each concept in one dimension for simplicity and then generalize it to two dimensions for application to images. We begin by introducing the three basic types of Wavelet transforms. Then we illustrate some particular Wavelets and applications of Wavelet transforms.

1. Continuous Wavelet Transform

1.1 Definition

If $\psi(x)$ is a real-valued function whose Fourier spectrum, $\psi(x)$, satisfies the admissibility criterion [2,3]

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(s)|^2}{|s|} ds < \infty \quad (1)$$

then, the $\psi(x)$ is called a basic Wavelet. Notice that, due to the s is the denominator of the integrand, it is necessary.

$$\psi(0) = 0 \Rightarrow \int_{-\infty}^{\infty} \psi(x) dx = 0 \quad (2)$$

Furthermore, since $\psi(x) = 0$ as well, we can see that the amplitude spectrum of an admissible wavelet is similar to the transfer function of a bandpass filter. In fact, any bandpass filter impulse response with zero mean [Eq. (2)] that decays to zero fast enough with increasing frequency [Eq. (1)] can serve as a basic Wavelet for this transform.

A set of Wavelet basis functions, $\{\psi_{a,b}(x)\}$, can be generated by translating and scaling the basic Wavelet, $\psi(x)$, as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (3)$$

where $a > 0$ and b are real numbers. The variable reflects the scale (width) of a particular basis function, while b specifies its translated position along the x -axis.

The continuous Wavelet transform of $f(x)$ with respect to the Wavelet $\psi(x)$ is then [2,3]

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(x) \psi_{a,b}(x) dx \quad (4)$$

The Wavelet transform coefficients are once again given as inner products of the function being transformed with each of the basis functions.

Grossman and Morlet [2] showed that the inverse continuous Wavelets transform is

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_f(a,b) \psi_{a,b}(x) db \frac{da}{a^2} \quad (5)$$

The scale factor in front of the right-hand side of Eq. (3) ensures that the norms of the Wavelet basis functions are all equal, since

$$\|f(\frac{x-b}{a})\| = \sqrt{\int_{-\infty}^{\infty} |f(\frac{x-b}{a})|^2 dx} = \sqrt{a} \|f(x)\| \quad (6)$$

Since the basic Wavelet has zero mean [Eq. (2)], all scaling and translations of it [Eq. (3)] will be likewise to have zero mean, and the mean of $f(x)$ must be accounted for separately.

1.2 Two-Dimensional CWT

The continuous Wavelet transform $W(a,b)$ of a one-dimensional function $f(x)$ is a function of two variables, one more than $f(x)$. The CWT is said to be overcomplete, as it represents a considerable increase in information content and in the volume required for data storage. For functions of more than one variable, this transform also increases the dimensionality by one.

If $f(x)$ is a function of two dimensions, its continuous Wavelet transform is

$$W_f(a, b_x, b_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \psi_{a, b_x, b_y}(x, y) dx dy \quad (7)$$

where b_x and b_y specify the translation in two dimensions. The inverse two-dimensional continuous Wavelet transform is

$$f(x, y) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_f(a, b_x, b_y) \psi_{a, b_x, b_y}(x, y) db_x db_y \frac{da}{a^3} \quad (8)$$

where

$$\psi_{a, b_x, b_y}(x, y) = \frac{1}{|a|} \psi\left(\frac{x - b_x}{a}, \frac{y - b_y}{a}\right) \quad (9)$$

and $\psi(x, y)$ is a two-dimensional basic Wavelet. The same generalization extends to cover functions of more than two variables.

1.3 Filter Bank Interpretation

The following exercise illustrates one way of viewing the continuous Wavelet transform. We define the general Wavelet basis function at scale a as

$$\psi_a(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x}{a}\right) \quad (10)$$

This is the basic Wavelet scaled by a and normalized by $a^{-1/2}$. It defines a set of functions that becomes broader with increasing a . We also define

$$\psi_a^*(x) = \psi_a^*(-x) = \frac{1}{\sqrt{a}} \psi^*\left(-\frac{x}{a}\right) \quad (11)$$

which is the reflected complex conjugate of the scaled wavelet. If $\psi(x)$ is real and even, as is often the case, the reflection and conjugation have no

effect.

Now we can write the continuous Wavelet transform [Eq. (5)] as

$$W_f(a,b) = \int_{-\infty}^{\infty} f(x) \check{\psi}_a(b-x) dx = f * \check{\psi}_a \quad (12)$$

For fixed a , then, $W_f(a,b)$ is the convolution of $f(x)$ with the reflected conjugate Wavelet at scale a .

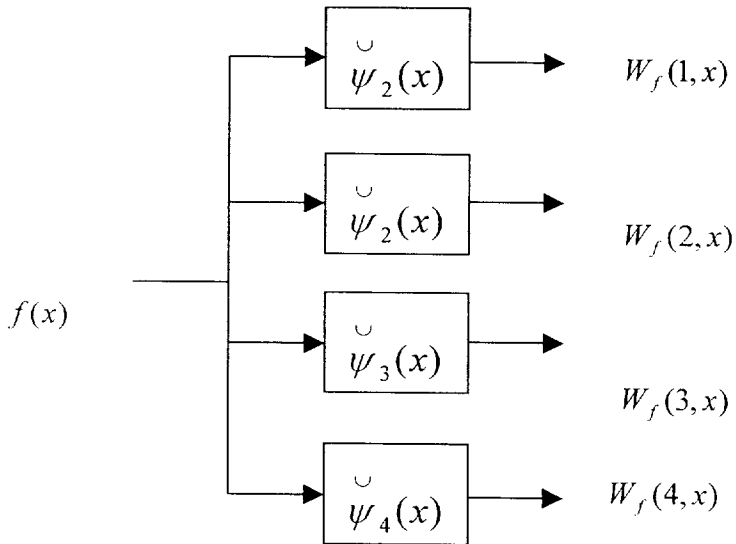


Fig.1. Filter bank analogy for the Wavelet transform of a signal

Figure 1 shows the Wavelet transform as a bank of linear (convolution) filters acting upon $f(x)$. Each value of a defines a different bandpass

filter, and the outputs of all the filters, taken together, comprise the Wavelet transform. Further Eq. (6) becomes

$$\begin{aligned} f(x) &= \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [f * \check{\psi}_a](b) \psi_a(b-x) db \frac{da}{a^2} \\ &= \frac{1}{C_\psi} \int_{-\infty}^{\infty} [f * \check{\psi}_a * \psi_a](x) \frac{da}{a^2} \end{aligned} \quad (13)$$

which implies that the filter outputs, each filtered again by $\psi_a(x)$ and properly scaled, combine to reconstruct $f(x)$. This is a statement of calderon's identity [4,5], which predates Grossman and Morlet by 20 years.

1.4 Two Dimensional Filter Banks

Figure 2 illustrates the filter bank approach in two dimensions. Here, each filter $\psi_a(x, y)$ is two-dimensional impulse response, and its output is a bandpass-filtered version of the image. The stack of filtered images comprises the Wavelet transform.

Again, the redundancy is considerable. In fact, if $\psi(u, v)$, the transfer function of $\psi(x, y)$ is nonzero everywhere except at the origin, one could, theoretically, recover the original image from any one of the filter output by inverse filtering. Alternatively, if the image is bandlimited to an interval

over which at least one $\psi_a(u,v)$ is nonzero, then $f(x,y)$ could be recovered from that filter output alone. The conclusion, then, is that the potential value of the integral Wavelet transform lies not in a compact representation, but in decomposition and analysis of signals and images.

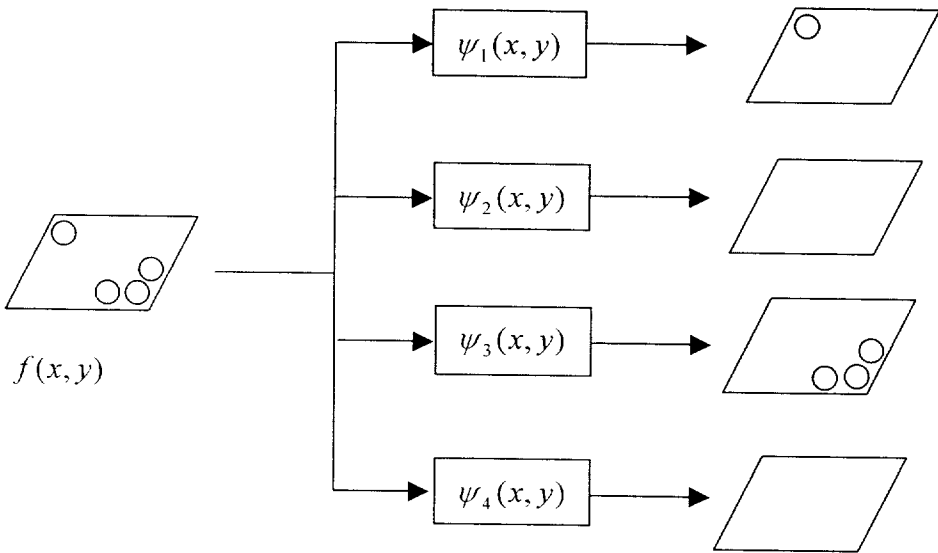


Fig. 2. Filter bank analogy for the Wavelet transform of an image

2. Discrete Wavelet Transform

The DWT most closely resembles the unitary transforms discussed in the previous section. It promises to be the most useful for image compression, processing an analysis. Given a set of orthonormal basis functions, one can compute the discrete Wavelet transform just as one does any other unitary transform. Obtaining a suitable basic wavelet, however, requires further background material.

2.1 Subband Coding and Decoding

Two-channel subband coding, then ,requires only filtering $f(i\Delta t)$ with $h_0(i\Delta t)$ and $h_1(i\Delta t)$, followed by subsampling each output. This yields the two half-length subband signals

$$g_0(k\Delta t) = \sum_i f(i\Delta t)h_0((-i + 2k)\Delta t) \quad (14)$$

and

$$g_1(k\Delta t) = \sum_i f(i\Delta t)h_1((-i + 2k)\Delta t) \quad (15)$$

Reconstruction is effected by upsampling the lower and upper subband signals, interpolating them with $2h_0(i\Delta t)$ and $2h_1(i\Delta t)$, respectively, and adding them together. This is given by

$$f(i\Delta t) = 2 \sum_k [g_0(k\Delta t)h_0((-i+2k)\Delta t) + g_1(k\Delta t)h_1((-i+2k)\Delta t)] \quad (16)$$

and is illustrated in Figure 3.

We have a slight problem at the midfrequency point $s = s_N/2$, since encoding and decoding entails filtering $f(i\Delta t)$ twice, once with $h_0(i\Delta t)$ and once with $h_1(i\Delta t)$, since $H_0(s_N/2) = 1/2$ and $H_1(s_N/2) = 1/2$. In the next section, where we use more general bandpass filters, we handle the situation explicitly.

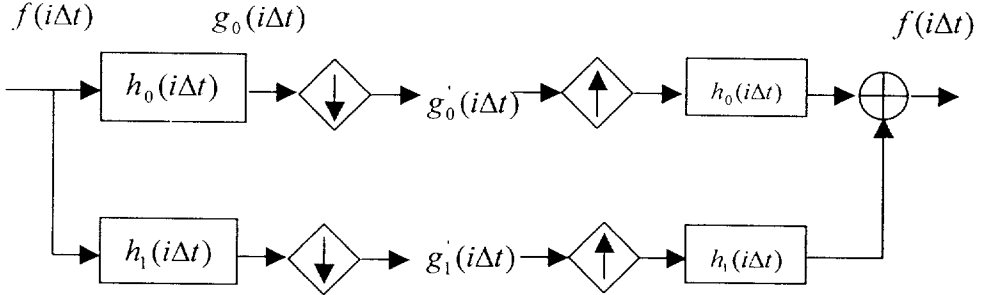


Fig. 3. Two-band subband coding and reconstruction

We could have chosen to partition the frequency axis into M shorter intervals of length $2s_N/M$, producing M subband signals of N/M points each, as is commonly done in subband coding. Different frequency components then show up in separate subband channels. Since we are moving toward the DWT, however, we stick with the choice of two subbands ($M=2$).

2.2 Fast Wavelet Transform Algorithm

Mallat [6] defined a discrete Wavelet transform algorithm that is more efficient than computing a full set of inner products. It applies two-band subband coding in an iterative fashion and builds the Wavelet transform from the bottom to up, which is, computing small-scale coefficients first.

After the first step of subband coding, as outlined in Section A, the lower subband signal, $g_0(i\Delta t)$, is once again subjected to halfband subband coding. This leaves us with the $N/2$ -point upper halfband signal and two $N/4$ -point subband signals corresponding to the first and second quarters the interval $[0, s_N]$.

The process is continued, at each step retaining the upper halfband signal and further encoding the lower halfband signal, until a one-point lowband signal is obtained. The transform coefficients are then the lowband point and the collection of subband-coded upper halfband signals. This is shown in Figure 4. the first $N/2$ coefficients come from the upper halfband of $F(s)$, the next $N/4$ points from the second quarter band, etc.

The impulse response, h_j , doubles in scale at each iteration. Thus, we have an orthonormal wavelet transform. The basic Wavelet is $h(t) = \delta(t) - \text{sinc}(at)$, and the basis function is $\{2^{-j/2} h(2^j t - n)\}$. Thus, subband coding, basically a time-frequency transform technique, has been

employed to define a time-scale Wavelet transform.

The foregoing algorithm is sometimes referred to as the fast Wavelet transform (FWT), or Mallat's herringbone algorithm, due to the appearance of the diagram in Figure 4. The inverse transform is obtained by reversing the process, as shown in Figure 5.

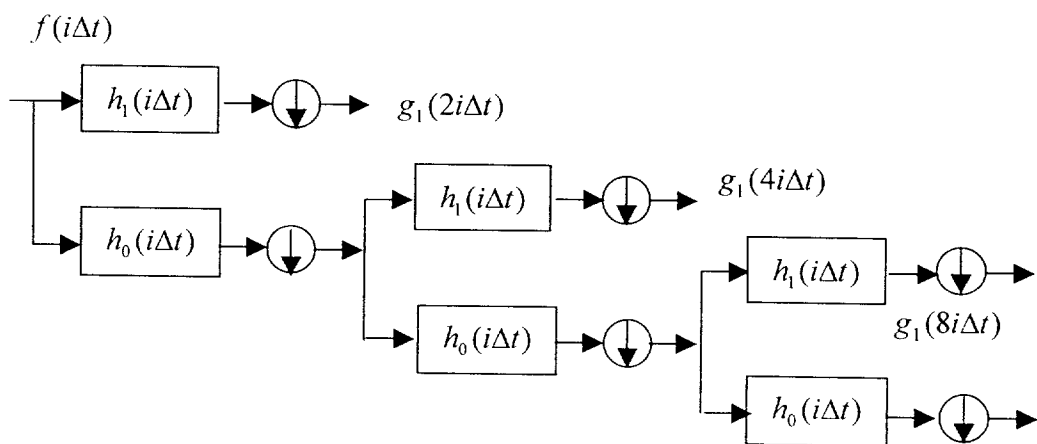


Fig. 4. The discrete Wavelet transform algorithm

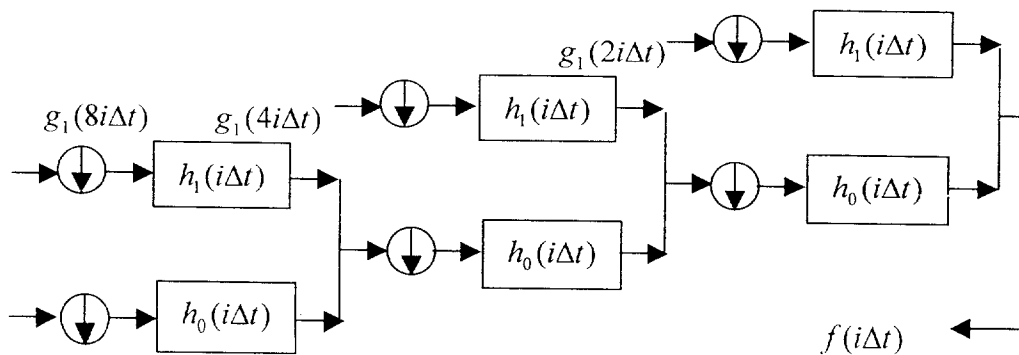


Fig. 5. The inverse discrete Wavelet transform

2.3 Discrete Wavelet Transform Design

We are now prepared to approach the design of a basic Wavelet for use in a discrete Wavelet transform. For the DWT, we can use any pair of subband coding filters that allows Eq. 16 to hold.

Writing Eq. 16 in the frequency-domain, we have

$$\begin{aligned} F(s) &= 2\left[\frac{1}{2}G_0(s)H_0(s) + \frac{1}{2}G_1(s)H_1(s)\right] \\ &= 2\left[\frac{1}{2}F(s)H_0(s)H_0(s) + \frac{1}{2}F(s)H_1(s)H_1(s)\right] \end{aligned} \quad (17)$$

which means that

$$F(s) = F(s)[H_0^2(s) + H_1^2(s)] \quad (18)$$

and the two filter transfer functions must satisfy the condition

$$H_0^2(s) + H_1^2(s) = 1 \quad \text{for } 0 \leq |s| \leq s_N \quad (19)$$

The transfer functions are squared here because $f(t)$ is convolved twice with each filter, once during coding and once during decoding. This resolves the problem that was noted in Section A.

Suppose $H_0(s)$ is a smooth-edged lowpass transfer function that we wish to use in a Wavelet transform. Clearly, the corresponding $H_1(s)$ is given by

$$H_1^2(s) = 1 - H_0^2(s) \quad (20)$$

Thus, a well-selected lowpass filter is all that is required to design a discrete Wavelet transform.

2.3.1 Scaling Vector

To develop a discrete Wavelet transform, we need only a discrete lowpass filter impulse response $h_0(k)$ that meets certain conditions [8]. This impulse response is sometimes called a scaling vector.

From $h_0(k)$ we can generate a related function $\phi(t)$, called the scaling function. We can also generate $h_1(k)$ and, from it and $\phi(t)$, the basic Wavelet, $\psi(t)$. If the scaling vector has only a finite number of nonzero entries, then $\phi(t)$, $\psi(t)$, and the resulting Wavelets will all have compact support [8]. That is, they will be zero outside a relatively short interval on the t -axis.

Let the scaling vector be a sequence such that

$$\sum_k h_0(k) = \sqrt{2} \quad \text{and} \quad \sum_k h_0(k)h_0(k+2l) = \delta(l) \quad (21)$$

then there exists a scaling function

$$\phi(t) = \sum_k h_0(k) \phi(2t - k) \quad (22)$$

2.3.2 Wavelet Vector

Once we have both $\phi(t)$ and $h_0(k)$ in hand, we continue the development by defining a discrete highpass impulse response called the Wavelet vector as

$$h_1(k) = (-1)^k h_0(-k + 1) \quad (23)$$

and, from that, a basic Wavelet

$$\psi(t) = \sum_k h_1(k) \phi(2t - k) \quad (24)$$

from which an orthonormal Wavelet set

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \quad (25)$$

follows.

2.4 Two-Dimensional Discrete Wavelet Transform

The concepts developed for the representation of one-dimensional signals generalize easily to two dimensions [3,5,6,8]. As with unitary image transforms, we consider the case where the two-dimensional scaling function is separable; that is,

$$\phi(x, y) = \phi(x)\phi(y) \quad (26)$$

where $\phi(x)$ is one-dimensional scaling function. If $\psi(x)$ is its companion Wavelet, then the three two-dimensional basic Wavelets

$$\begin{aligned} \psi^1(x, y) &= \phi(x)\psi(y) \\ \psi^2(x, y) &= \psi(x)\phi(y) \\ \psi^3(x, y) &= \psi(x)\psi(y) \end{aligned} \quad (27)$$

establish the foundation for a two-dimensional Wavelet transform. Note that the superscript is used here as an index rather than an exponent. In particular, the set of functions

$$\{\psi_{j,m,n}^l(x, y)\} = \{2^j \psi^l(x - 2^j m, y - 2^j n)\} \quad j \geq 0, l = 1, 2, 3 \quad (28)$$

where j, l, m, n are integers, is an orthonormal basis for $L^2(R^2)$.

2.5 Biorthogonal Wavelet Transform

The functions that qualify as orthonormal Wavelet with compact support lack desirable symmetry properties. It would be convenient, for example, if $\psi(x)$ and $\tilde{\psi}(x)$ —one for decomposition and the other for reconstruction—we can have symmetrical Wavelets with compact support [3,4,9,10,11]. The two Wavelets are duals of each other, and the Wavelet families $\{\psi_{jk}(x)\}$ and $\{\tilde{\psi}_{jk}(x)\}$ are biorthogonal; that is ,

$$\langle \psi_{j,k}, \tilde{\psi}_{l,m} \rangle = \delta_{j,l} \delta_{k,m} \quad (29)$$

then we have

$$c_{j,k} = \langle f(x), \tilde{\psi}_{j,k}(x) \rangle \quad \text{and} \quad d_{j,k} = \langle f(x), \psi_{j,k}(x) \rangle \quad (30)$$

for the decomposition, and

$$f(x) = \sum_{j,k} c_{j,k} \psi_{j,k}(x) = \sum_{j,k} d_{j,k} \tilde{\psi}_{j,k}(x) \quad (31)$$

for the reconstruction. Either Wavelet can be used for the decomposition, provided that the other one is used for the reconstruction. The biorthogonal Wavelet transform allows the use of symmetric Wavelets having compact support.

2.5.1 Implementation

The one-dimensional biorthogonal Wavelet transform requires four discrete filters. We must choose two lowpass filters (scaling vectors), $h_0(n)$ and $\tilde{h}_0(n)$, whose transfer functions satisfy

$$H_0(0) = \tilde{H}_0(0) = 1 \text{ and } H_0(s_N) = \tilde{H}_0(s_N) = 0 \quad (32)$$

where $s_N = 1/2\Delta x$ is the folding frequency. From these, we generate two bandpass filters (wavelet vectors), as before, by half-period shifts of their transfer functions:

$$h_1(n) = (-1)^n h_0(1-n) \text{ and } \tilde{h}_1(n) = (-1)^n \tilde{h}_0(1-n) \quad (33)$$

Now we can implement the FWT herringbone algorithm using these four filters, as shown in Figure 6.

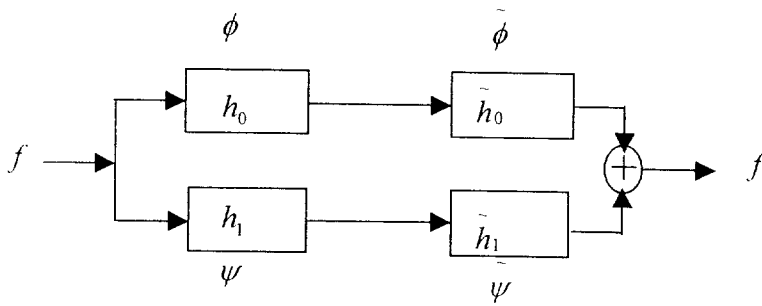


Fig. 6. One decomposition step and one reconstruction step of the biorthogonal Wavelet transform

2.5.2 Biorthogonal Wavelets

The conditions upon biorthogonal Wavelet filters are

$$\sum_n h_0(n) = \sum_n \tilde{h}_0(n) = \sqrt{2} \quad \text{and} \quad \sum_n h_1(n) = \sum_n \tilde{h}_1(n) = 0 \quad (34)$$

and the perfect reconstruction property requires that

$$\begin{aligned} & H_0(s) \tilde{H}_0(s) + H_1(s) \tilde{H}_1(s) \\ &= H_0(s) \tilde{H}_0(s) + H_0(s - s_N) \tilde{H}_0(s - s_N) \\ &= 1 \end{aligned} \quad (35)$$

The two scaling functions are given, in the frequency domain, by

$$\begin{aligned} \Phi(2s) &= \tilde{H}_0(s) \Phi(s) = \prod_{n=0}^{\infty} \tilde{H}_0(s/2^n) \\ \tilde{\Phi}(2s) &= H_0(s) \tilde{\Phi}(s) = \prod_{n=0}^{\infty} H_0(s/2^n) \end{aligned} \quad (36)$$

and the Wavelets are then

$$\begin{aligned} \psi(x) &= \sqrt{2} \sum_n h_1(n+1) \phi(2x-n) \\ \tilde{\psi}(x) &= \sqrt{2} \sum_n \tilde{h}_1(n+1) \tilde{\phi}(2x-n) \end{aligned} \quad (37)$$

3. Huffman Coding for Image Compression

In lossless image compression, there is an intrinsic limitation to how much an image can be compressed. Compression past this point will eliminate some of the information necessary to recreate the original full in its exact form.

The entropy of an image is a measure of this limit. An image's entropy is a measure of its information content. If the entropy is high, an image's information tends to be highly unpredictable. Stating in another way, a high-entropy image's information contains a lot of randomness and has little redundancy. If the entropy is low, an image's information is more predictable - it contains little randomness and its redundancy is high.

We can compute an image's entropy and the probability of its occurrence. This is displayed as a number representing the number of bits necessary to represent that probability. For any random image, this would be as follows:

$$\text{Entropy} = \text{Number of Pixels} \times \text{Number of Lines} \times \text{Number of Bits per Pixel}$$

Because images are rarely made up of totally randomly varying brightnesses, the actual entropy of a normal image will generally be something less than the calculation above. This is because the raw image

data quantity will always be higher than the average information data quantity. The actual entropy of an image is the average information quantity of the image. The form of compression that does this is called entropy coding.

Another common entropy-coding compression technique is Huffman coding. Huffman coding converts the pixel brightness values in the original image to new variable-length codes, based on their frequency of occurrence in the image. In this way, brightnesses that occur frequently are assigned shorter codes, and brightnesses that occur infrequently are assigned longer codes. The result is that the compressed image will require fewer overall bits to describe the original image.

The Huffman compression scheme begins by looking at the brightness histogram of an image. With the histogram, the frequency of occurrence for all brightness in the image is available. By ordering the brightness values by their frequencies of occurrence, we are left with a list where the first value is found the most often in the image, and the last value is found the least often in the image. With this list, the Huffman coder assigns new codes to each brightness value. The assigned codes are of varying lengths; the shortest codes are assigned to the first (most frequent) values in the list and, eventually, the longest codes are assigned to the last (least frequent)

values in the list. Finally, the compressed image is created by simply substituting the new variable-length brightness-value codes for the original 1-byte brightness-value codes. Of course, the Huffman code list that couples original brightness values to their new Huffman variable codes must be appended to the image for use by the Huffman decompression operation, as shown in Figure 7.

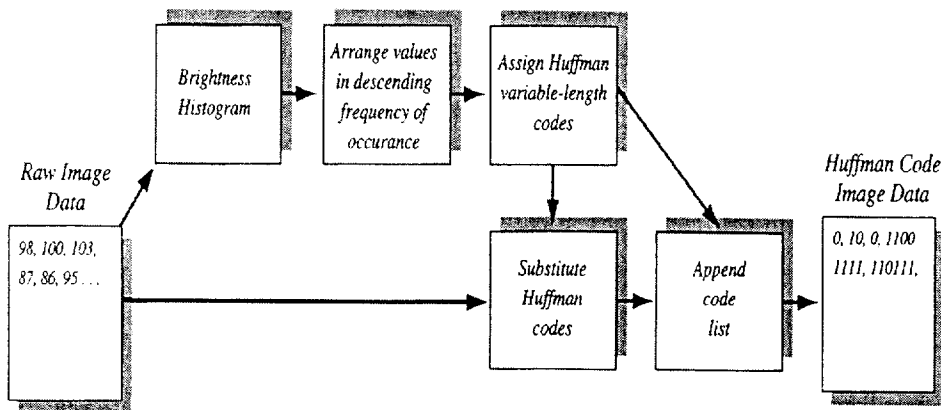


Fig. 7. The flow of the Huffman coding operation

Huffman codes are assigned by creating a Huffman tree that pairs the brightness values based on their combined frequencies of occurrence. The Huffman tree ensures that the longest codes get assigned to the least frequent brightnesses and vice versa. Using the brightness ranked in order of their frequencies of occurrence, the two at the bottom of the list (least frequent) are paired together and labeled 0 and 1. The brightness is represented by their combined frequencies of occurrence. Then, the next two lowest frequencies of occurrence made up occurrence. This continues until all brightness and paired brightness have been paired. The result is a tree that, when followed, indicates the new Huffman binary code for each brightness in the image.

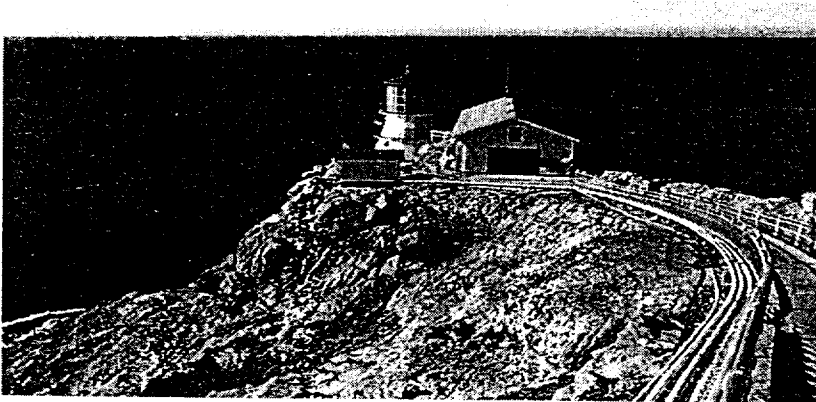


Fig. 8a. Original image, with 3-bit brightness values

Figure 8 shows a 640 by 480 line image, where each pixel is represented, for simplicity, by a 3-bit brightness value. The image's histogram shows the actual number of pixels in the image with each of the eight brightness values. The brightness is ordered based on their frequencies of occurrence and then original image were coded as 3-bit brightness values. The Huffman codes are as small as 1 bit and can be as large as 7 bits. The longest Huffman codes can never be greater than the number of different brightness in the image (in this case, eight) minus 1, even though frequencies of occurrence are always statistically low.

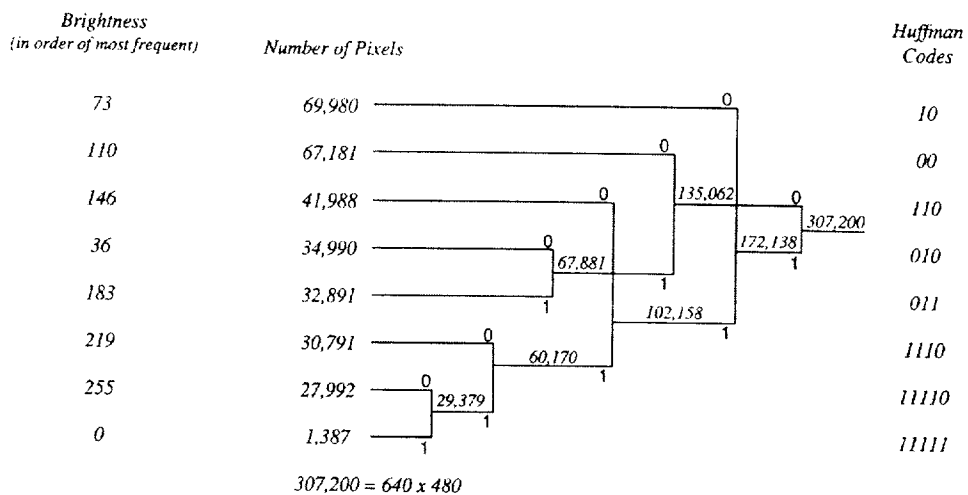


Fig. 8b. The creation of the Huffman tree

The data size of the original image can be computed as 640 by 480 by 3 bits. The Huffman-coded image data size can be computed as the sum of the eight frequencies of occurrence multiplied by the respective number of bits in their code.

Huffman image decompression reverses the compression process by substituting the original fixed-length 1-byte-long brightness values for the variable-length Huffman-coded values. The original image is exactly recreated. Huffman image compression will generally provide compression ratios around 1.5:1 to 2:1.

Modified versions of the Huffman coding scheme can be used to allow code changes throughout the image. This is done by computing image histograms over regions of the image, rather than over the entire image. Codes are then assigned based on the brightness frequencies of occurrence in each region. This modification of the basic Huffman coding scheme allows the codes to change based on the brightness distributions found in different parts of the image. The codes are therefore more efficiently adapted to the regions of the image, yielding improved compression ratios. Of course, every time the Huffman codes are changed, the changes must be appended to the image for later decompression.

III. Image Compression Algorithm

1. Pre-Quantizer

Pre-quantization in spatial domain is optionally employed for bit-rate control and peak error control. Conventionally a single quantizer is used for pre-quantization, by which gray levels as well as precision are reduced. For example, if one applies a pre-quantization with peak error 1 to an image with 256 gray levels, the number of gray levels after pre-quantization may be reduced from 256 down to 86. The reduction of gray may introduce additional visual degradation in addition to the quantization. This visual degradation would be serious in such regions where gray levels are changing slowly. If one uses a single quantizer, the quantized output may be expressed as

$$P(x) = \left[\frac{x+e}{2e+1} \right] \quad (38)$$

where x is input, e is absolute peak error, and $[z]$ is for function which takes the largest integer number less than z . Reconstruction or dequantization may be expressed as

$$P^{-1}(x) = P(x) \cdot (2e+1) \quad (39)$$

To overcome the visual degradation due to the reduction of the gray levels, a novel pre-quantization scheme using multiple quantizers instead of single

quantizer is proposed. The quantized output using multiple quantizers is expressed by

$$Q_i(x) = \left[\frac{x+i}{2e+1} \right] \quad (40)$$

and

$$Q_i^{-1}(x) = Q_i(x)(2e+1) + (e-i), \quad i=0,1,2,\dots,2e \quad (41)$$

where i is the index identifying the quantizer among multiple quantizers.

The quantizer may be chosen by a selection rule on a pixel by pixel basis.

The index may be chosen in cyclic or random orders, or by a context based rule.

Since each quantizer has different representation values as seen in Eq. (41), the number of gray levels in the pre-quantized image with the multiple quantizers are the same as that in the original image.



Fig. 9. Pre-quantized Lena image

Thus visual degradation due to the reduction of number of gray levels with single quantizer and multiple quantizers are exemplary show in Fig.9 for a better visualization of the degradation due to the reduction of gray levels. In this example, absolute peak error was 2. The context-based rule is given by

$$i = \text{mod}((a + b - c), 7) \quad (42)$$

where a , b and c are reconstructed gray levels of the nearest neighbor pixels, i.e., upper, left, and upper left pixels of the current pixel. Note that the index is given to match the gray level of the current pixel by a

prediction. Due to a decrease of number of gray levels with conventional single quantizer, artifact is observed in gradually changing regions. However in the proposed multiple quantizers, no such artifact is observed in either random selection of the quantizer or selection by a context-based rule. The PSNR and bit rate of the pre-quantized images by the proposed multiple quantizers appear similar to that with a convention single quantizer.

2. Wavelet Transform

Wavelet-based compression is one type of transform-based compression. In general, transform-based compression is done according to the scheme shown in Fig. 10.

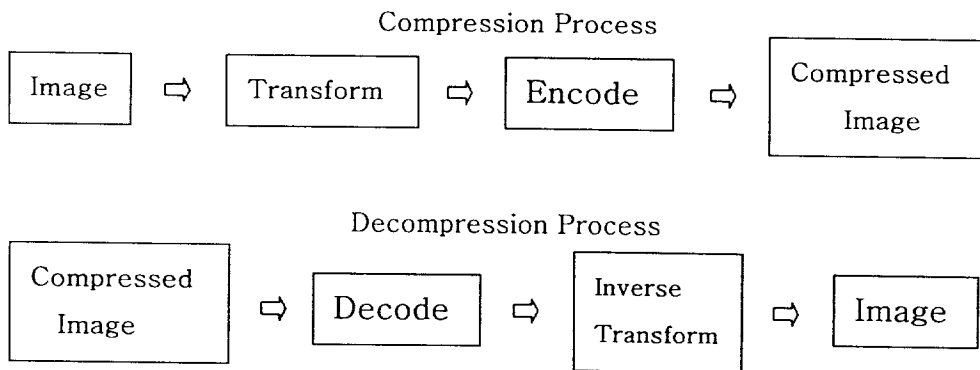


Fig. 10. Transform-based Compression

For some applications of Wavelet transform, where Wavelet coefficients are lossless encoded, Wavelet coefficients in 7-9 tap Daubechies [22] are useful. Furthermore, if the Wavelet coefficients are properly encoded, image can be progressively reconstructed from loss to lossless [23] in a unified manner with one bit stream. For this purposes, several Wavelet transforms whose Wavelet coefficients are Daubechies 7-9 tap have been used. Table 1 shows the 7-9 tap Wavelet filter coefficients.

Table 1. 7-9 tap Wavelet filter coefficients

n	0	± 1	± 2	± 3	± 4
h_n	0.602949	0.266864	-0.078223	-0.016864	0.026749
\tilde{h}_n	0.557543	0.295363	-0.028772	-0.045636	0

We have decomposed the image Lena (Fig. 11) with 7-9 tap Daubechies Wavelet filter. The results are presented in Fig.12. In Fig. 12 we can see the normalized detail subimages at resolution level $m=3$ (Wavelet coefficients).

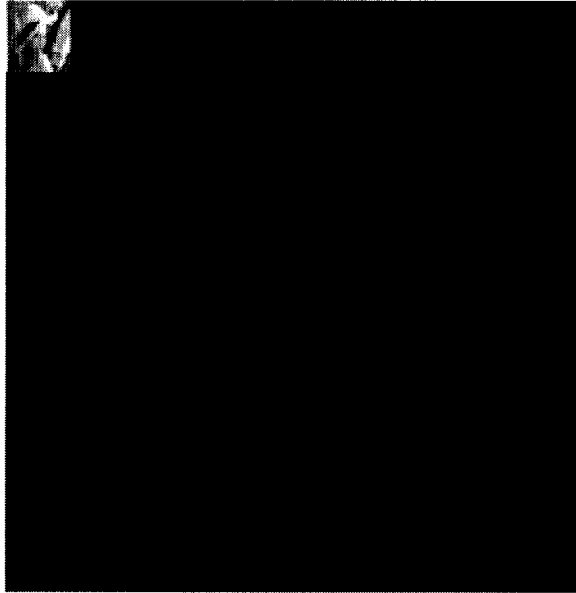


Fig. 11. Decomposed Lena image

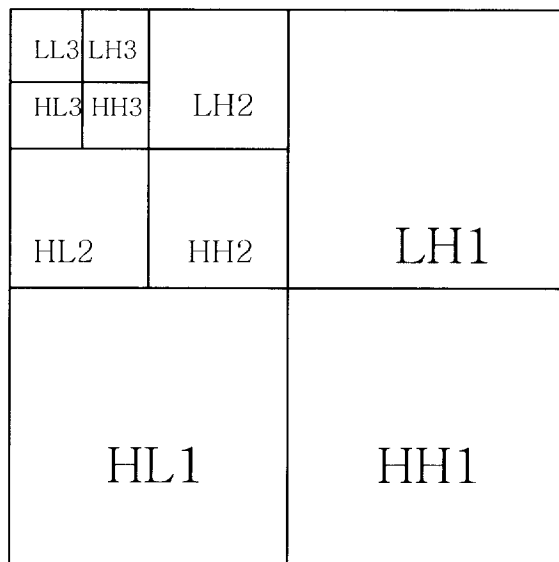


Fig. 12. Subimage at resolution level $m=3$

3. Huffman Coding with The Predictor

In this paper we use 7-9 tap Daubechies Wavelet filter to transform the image pre-quantized images. After Wavelet transforming the image, we can divide the image into two parts, and then code them.

We consider that energy of images mainly is centralized in LL3 area, and the quantization of LL3 area directly affects the quality of reconstruction image. So we use Huffman coding [18] with the predictor in LL3 area. The predictor is defined as follows:

$$\hat{a} = b + c - d \quad (43)$$

where \hat{a} is predicted value of pixel, b, c and d are predicted pixels of the nearest neighbor pixels, i.e., upper, left, and upper left pixels of current pixel.

Because to other domains' signal coding, there is a great deal of edge of original images in those domains, signals in those domains are zigzag scanned, and then we obtain the one- dimension signal. The Fig. 13 shows the direction of the scan.

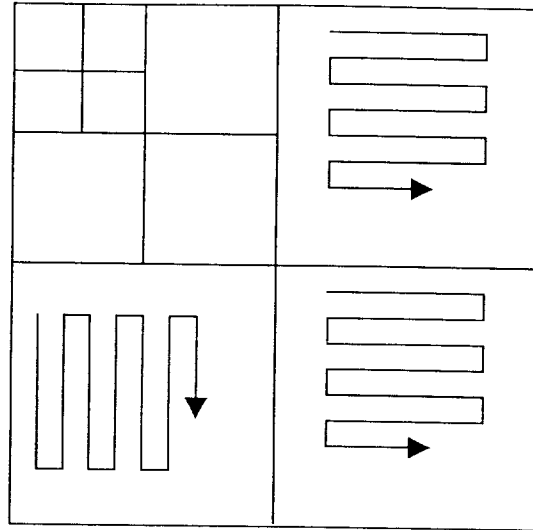


Fig. 13. Direction of zigzag scanning

Because we consider that the great mass of values of one- dimensional signal of the realigned scan is equal to zero, the number of zeroes presented in signals and values consists of a couple of signals. Then the couple of signals are Huffman coded.

In the light of steps, we can accomplish the algorithm of the image compression mentioned in this paper.

IV. Simulation Results

The proposed algorithm is applied to several test images. The proposed algorithm has unique feature adequate for image compression, i.e., compression from near lossless, to lossless compression. Furthermore peak error can be controlled in near lossless compression using a pre-quantizer in spatial domain. Table 2 shows peak errors and peak signal-to-noise ratio (PSNR) between the lossless proposed compression schemes with pre-quantization and lossy compression without pre-quantization. It is as seen in Table 2 with a pre-quantization at near lossless compression. With larger peak error, however, performance of the pre-quantization in spatial domain becomes worse than those by the bit rate control in the Wavelet domain. The test image used in Table 2 was Lena in 256×256 matrix size with 8 bit gray scales.

In Table 3, performances of the proposed coding schemes are compared to those of a float-point based wavelet transform coding and the JPEG standard. For the float-point based wavelet coding, 7-9 tap Daubechies Wavelet filter was used. 256×256 , 8 bit gray scale images was used for the simulation of the lossless and near lossless compression. As seen in Table 3, the proposed Daubechies wavelet transform coding with a pre-quantizer for near lossless compression provides higher PSNR compared to

the float-point based Wavelet transform coding and the JPEG standard. The peak errors by the proposed coding scheme are also much lower than those by the float-point based Wavelet transform coding or the JPEG standard. The capability of the peak error control with the pre-quantizer and the Daubechies Wavelet transform coding is a useful in widely applied in the image compression. Fig. 14-16 shows reconstructed images.

Table 2. Peak error and PSNR in the compressed images with and without pre-quantizer in the proposed coding scheme

Bit rate [bits/pixel]	With pre-quantizer		Without pre-quantizer	
	PSNR [dB]	Peak error	PSNR [dB]	Peak error
2.89	49.9	1	47.8	5
2.37	45.1	2	44.9	6
2.10	42.1	3	44.0	7

Table 3. Simulation Results






영 상	분류	비트율	압축율	PSNR[dB]
	Wavelet	1 bpp	8	40.09 dB
	JPEG	1 bpp	8	38.32 dB
	Wavelet	1 bpp	8	43.54 dB
	JPEG	1 bpp	8	41.81 dB
	Wavelet	1 bpp	8	40.96 dB
	JPEG	1 bpp	8	38.10 dB
	Wavelet	1 bpp	8	39.13 dB
	JPEG	1 bpp	8	36.94 dB
	Wavelet	1 bpp	8	38.67 dB
	JPEG	1 bpp	8	35.93 dB



Fig. 14. Lena Images and reconstructed Images (1 bpp, PSNR=40.09 dB)



Fig. 15. Camera man original Images and reconstructed Images (1 bpp, PSNR=43.54 dB)



Fig. 16. Hat girl original images and reconstructed Images (1 bpp, PSNR=38.67 dB)

V. Conclusion

In this paper, in order to improve compression ratio, firstly, I pre-quantize the original image, and then gain the image which is the same as the gray-level of the original image.

Based on pre-quantization, 7-9 tap Wavelet filter and Huffman coding with a predictor can compress images. We consider that energy of images mainly is centralized in LL3 area, and the quantization of LL3 area directly affects the quality of reconstruction image. So we use Huffman coding with the predictor in LL3 area, and use runlength coding in other areas.

In order to test the capability of the proposed algorithm, we adopt the image of 256 by 256 size of 8 bit gray level in the simulation. Apparently, we can gain from the simulation results that under the same compression ratio, the algorithm we proposed is better than JPEG, and PSNR of the algorithm is higher than that of JPEG by 2.8 dB. So the algorithm can be widely applied to the image compression.

In the future research, we will continue to explore in the new domain of the image compression. The CNN Theory is widely applied in image processing. The direction of the research is how to use the theory of CNN to apply in the image compression in future, although this is a huge challenge to us!

References

- [1] S. Pittner, J. Schneid, and C. W. Ueberhuber, "Wavelet Literature Survey, " Technical University of Vienna, Austria, 1993.
- [2] A. Grossman and J. Morlet, "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape, " SIAM J. Math Anal., Vol. 15, pp.723-736, 1984.
- [3] C. K. Chui, "An Introduction to Wavelets, " Academic Press, San Diego, 1992.
- [4] A. P. Calderon, "Intermediate Spaces and Interpolation. The Complex Method, " Studia Math, pp.113-190, 1964.
- [5] Y. Meyer (transl. by R. D. Ryan), " Wavelets: Algorithms and Applications, " Society for Industrial and Applied Mathematics, Philadelphia, 1993.
- [6] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, " IEEE Trans., PAMI-11, pp. 674-693, 1989.
- [7] R. A. Gopinath and C. S. Burrus, "Wavelet Transforms and Filter Banks, " Wavelets-A Tutorial in Theory and Applications, pp.603-654, Academic Press, 1992.
- [8] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelet, " Commun. on Pure and Appl. Math., pp.909-996, 1988.

- [9] A. I. Cohen, I. Daubechies, and J. C. Feauveau, "Bi-Orthogonal Bases of Compactly Supported Wavelets, " *Comm. Pure and Applied Math.*, pp.485-560, 1992.
- [10] M. Vetterli and C. Herley, "Wavelets and Filter Banks: Theory and Design, " *IEEE Trans.. SP* 40(9): pp.2207-2232, 1992.
- [11] R. M. Gray, "Vector Quantization," *IEEE ASSP Magazine*, Vol.1, No.2, pp. 4-29, April 1984.
- [12] H. Abut, ed., " Vector Quantization." *IEEE Reprint Collection*, Piscataway, New Jersey: IEEE Press, May 1990.
- [13] Gonzalez, R. C. and Woods, " Digital Image Processing." Reading, Massachusetts: Addison-Wesley, R. E., 1992.
- [14] Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall, A. K. 1989.
- [15] Pratt, " Digital Image Processing," 2nd ed. New York: John Wiley & Sons. W. K 1991.
- [16] J. Shirani, S. Kossentini, " On RD Optimized Progressive Image Coding Using JPEG" *IEEE Trans. Imag.*, Vol. 8, pp.1630-1638, 1999.
- [17] Y. Noguchi, H. Kobayashi, H. Kiya, " A Method of Extracting Embedded Binary Data from JPEG Bitstreams Using Standard JPEG Decoder," *IEEE Trans. Imag.*, Vol. 1, pp. 577-580, 2000.

- [18] D. Taubman, A. Zakhor. " Rate and Resolution Scalable Subband Coding of Image", IEEE ICASSP-93, Circuits and Systems, Vol.1, pp. 271-274, 1993.
- [19] M. J. Smith and D. P. Barnwell, " Exact Reconstruction for Tree-Structured Subband Coders. " IEEE Trans. Acoust., Speech, Signal Proc., Vol. ASSP-34, pp. 434-441, 1986.
- [20] M. Vetterli, " Splitting a Signal into Subsampled Channels Allowing Perfect Reconstruction. " in Proc. IASTED Conf. Appl. Signal Processing Digital Filtering, Paris, France, June 1985.
- [21] C. B. Ahn, H. H. Koh, Y. H. Ahn, and Y. C. Song, " A Unified Lossy-to-Lossless Medical Image Compression Using Integer-Based Wavelet Transform " Imaging Science & Technology, Vol. 4, No.1, 1998.
- [22] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies " Image Coding Using Wavelet Transform, " IEEE Trans. Imag. Proc. Vol. 1, pp. 205-203 ,1992.
- [23] A. Said and W. Pearlman " An Image Multiresolution Representation for Lossless and Lossy Compression, " IEEE Trans. Imag. Proc. Vol. 5, pp. 1303-1310, 1996.

Acknowledgements

I would like to first thank Prof. Seok-Tae Kim for his invaluable guidance and endless patience during my tenure in Korea. Without his encouragement and support, I would not have finished this work. I am pleased that Prof. Jong-Rak Yoon and Prof. Moon-Gab Joo kindly acted as my committee members, their attentive correction is greatly appreciated.

The fulfillment of this work without the continual assistance and support of my fellow colleagues is unimaginable, in particular, Eun-Hee Choi, Soo-Myong Lee, Jong-Hyon Moon, Seok-Jin Lee, Jin-Hoo Huang, etc. in Pukyong National University. I also wish to thank Fu-Shou Jin, Chun-Dan Lin and all other Chinese students in Pukyong National University. they have always been so supportive.

There are so many more names I should mention, each one of which would remind me of so many unforgettable moments during the past two years. The space is limited, but I would always cherish them in my memory of Korea.

This thesis is dedicated to my parents and siblings, their blessings have always been my impetus. My deepest appreciation is expressed to my special friend, Miao-Zhang, for her patience, understanding and inspiration.