An Adaptive Quantization Scheme for 2-D DWT Coefficients

Po-Yueh Chen^{*} and Jia-Yu Chang

Department of Computer Science and Information Engineering, National Changhua University of Education, No. 2 Shi-Da Road, Changhua City, Taiwan

Abstract: Two-dimensional Discrete Wavelet Transform (2-D DWT) is an effective tool for image compression and hence attracts much attention in recent years. In this paper, we propose an adaptive quantization approach which quantizes the DWT coefficients according to their characteristics. Unlike the uniform quantization which uses a fixed quantization interval, the proposed scheme applies two asymmetric interval widths to one single sub-band. According to the experiment results, it leads to a superior image quality while maintaining the same compression ratio.

Keywords: Image compression; quantization; DWT coefficients.

1. Introduction

In 1991, JPEG is proposed as a standard for still images. However, there is still deficiency [1][2] such as the distortion resulted from high compression ratios and the lack of ability to strengthen images locally. Hence, at the beginning of 2000, the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) made the core of JPEG2000 [3][4][5] which adopts DWT as the standard compression tool. As a result, JPEG2000 exhibits several main advantages [6][7] as follows: (1) higher compression ratio while maintaining acceptable image quality, (2) convenience for progressive transmission, (3) better fault-tolerant rate [8], (4) lighter demand for storing devices, and (5) the ability to strengthen the region of interest (ROI) [9].

In JPEG2000, uniform quantization is performed for compressing the DWT coefficients and hence some information gets lost inevitably [10]. Greater quantization interval results in better compression ratios but poorer quality for reconstructed images. Uniform quantization employs symmetrical and equal interval width for quantization. However, when the image energy is asymmetric, the resulting DWT coefficients demonstrate various characteristics between various images and hence the reconstructed image quality can be improved by some adaptive approaches.

As shown in Figure1, a 2-D DWT generates 4 sub-bands, denoted as LL, HL, LH, and HH, which exhibit diversity of energy distributions. The low frequency sub-band LL preserves essential visional features for the original image and we can perform the second level DWT on it if multi-resolution representation is required. Since the role of LL is more crucial than that of the other three sub-bands, finer quantization (more quantization intervals) should be applied on it. This improves the compression ratio without decreasing the image quality. Furthermore, within any single sub-band, if some values like right interval width, left interval width, and *median* can

* Corresponding author; e-mail: <u>pychen@cc.ncue.edu.tw</u>	Receive
	Rev
© 2013 Chaoyang University of Technology, ISSN 1727-2394	Acc

be computed in advance, they assist the implementation for a much more efficient quantization scheme. This is the primary inspiration for the proposed approach and the execution steps are described in section 3 in details.



Figure 1. Four coefficient sub-bands after 2-D DWT transform

The rest of this paper is organized as follows. In section 2, Discrete Wavelet Transform and the uniform quantization is briefly reviewed. Section 3 demonstrates a new adaptive quantization for DWT coefficients with a numerical example. The experiment results and analysis are illustrated in section 4. Section 5 provides some concluding remarks.

2. Related works

2.1. Review of discrete wavelet transform (DWT)

Discrete Wavelet Transform is a method for translating an image from space domain into frequency domain. Haar DWT is the simplest DWT which divides an image into four sub-bands by addition and subtraction [11][12][13]. The procedure for 2-D Haar DWT is described as follows:

- (a) Vertical division: Vertical division divides an image into two parts. The first one is the sums of two adjacent columns, which is stored in left side as low frequency coefficients. The other part is the differences of two adjacent columns, which is stored in right side as high frequency coefficients. Performing vertical division on the 4 by 4 example shown in Figure 2(a), it results in Figure 2(b).
- (b) Horizontal division: After the vertical division, horizontal division further divides the image into four parts. As shown in Figure 2(c), sums of two adjacent rows are stored in upper side as low frequency coefficients where differences of two adjacent rows are stored in lower side as high frequency coefficients.

After these two 1-D divisions, the resulting 4 sub-bands are shown in Figure 2(d).

2.2. Review of uniform quantization in JPEG2000

2.2.1. Quantization encoding

Assume *I* is an $M \times N$ gray image, its pixels can be described as follows. I = { $x_{ij} | 1 \le i \le M, 1 \le j \le N, x_{ij} \in \{0, 1, 2, \dots, 255\}$ } (1) After performing DWT on *I*, we obtain four sub-bands LL, HL, LH, and HH of size $\frac{M}{2} \times \frac{N}{2}$.

Conventionally, JPEG2000 performs the uniform quantization on the resulting DWT

coefficients sub-bands. For most images, after subtracting the average of maximum and minimum, the distribution of coefficients is similar to a zero-mean Laplacian and hence uniform quantization adopts value 0 as the center of quantization. The uniform quantization process for JPEG2000 is formulated by Equation (2).

$$Q(i, j) = \text{sign} (H(i, j)) \left[\frac{|H(i, j) - \omega|}{\Delta b} \right]$$
(2)

where Q(i, j) is the quantized result at position (i, j), H(i, j) represents the original DWT coefficient at position (i, j), Δb stands for the interval width for quantization, and ω is the average of maximum and minimum. Figure 3 demonstrates a simple numerical example for this quantization process. As shown in the figure (ω =0 and Δb =10), the quantization result of -24 is -2 since it lies between -20 and -30. Besides the so-called "dead zone" (-10, 10) which is of width $2\Delta b$, all intervals are of the same width through the whole sub-band.

А	В	С	D		A+B	C+D	A-B	C-D
Ε	F	G	Н		E+F	G+H	E-F	G-H
I	J	К	L		l+J	K+L	I-J	K-L
Μ	N	0	Ρ		M+N	O+P	M-N	O-P
	(a)						(b)	
(A+B)+ (E+F)	(C+D)+ (G+H)	(A-B) +(E-F)	(C-D) +(G-H)				Ц	
(I+J)+ (M+N)	(K+L)+ (O+P)	(I-J)+ (M-N)	(K-L)+ (O-P)		L	.L		L
(A+B)- (E+F)	(C+D)- (G+H)	(A-B)- (E-F)	(C-D)- (G-H)		LH		LI	ш
(I+J)- (M+N)	(K+L)- (O+P)	(I-J)- (M-N)	(K-L)- (O-P)				п	П
	(b)			-			(d)	

Figure 2. Harr Discrete Wavelet Transform process (a) original data (b) vertical division (c) horizontal division (d) four sub-bands



2.2.2. Quantization decoding

Uniform de-quantization is formulated in Equation (3).

$$R_{Q(i,j)} = \begin{cases} (Q(i,j)+r)\Delta b + \omega, Q(i,j) > 0\\ (Q(i,j)-r)\Delta b + \omega, Q(i,j) < 0\\ 0, \quad otherwise \end{cases}$$
(3)

where $R_{Q(i,j)}$ and Q(i, j) stand for the reconstructed coefficient and the quantized value respectively. $0 \le r \le 1$ is an optional parameter for controlling the recovering position within a quantization interval. The default value in JPEG2000 is r = 0.5. For the same example shown in Figure 3, -2 will be decoded as -25 if we set r = 0.5.

Finally, perform the IDWT (Inverse Discrete Wavelet Transform) on the recovered coefficients in all 4 sub-bands to obtain the reconstructed image I'.

$$I = \{ x_{ij} \mid 1 \le i \le M, 1 \le j \le N, x_{ij} \in \{0, 1, 2, \dots, 255\} \}$$
(4)

3. The Adaptive Quantization Scheme

In the proposed scheme, asymmetric characteristics of images are manipulated to improve the image quality [14][15][16]. A simple numerical example is utilized in the following sub-sections for explaining the algorithm in convenience.

3.1. Adaptive quantization encoding

Four sub-bands of DWT coefficients are obtained after performing DWT on the original image. Figure 4 depicts a 4 by 4 numerical example for such a single sub-band.

255	210	198	177
155	25	200	0
153	199	174	180
150	200	45	207

Figure 4. A numerical example of a 4 by 4 sub-band

To precisely exhibit the data aggregation, *median* is more adequate than *mean* (the average of maximum and minimum). Hence, unlike the uniform quantization, the adaptive quantization employs two various interval widths while adopting median as the center of data.

To perform adaptive quantization, the right interval width, left interval width, and *median* for each sub-band have to be computed in advance. As shown in Figure 5, the *median* (which is $\frac{177 + 180}{2} \approx 179$ in this example) is subtracted from every coefficient to get a displacement

matrix.

76	31	19	-2
-24	-154	21	-179
-26	20	-5	1
-29	21	-134	28

Figure 5. The displacement matrix

As shown in Figure 6, two asymmetric quantization regions are obtained for this displacement matrix. In this example, the left region is (-179, 0) with length 179 and the right one is (0, 76) with length 76.



Figure 6. Two asymmetric quantization regions

According to the demanded compression ratio, choose an appropriate number of intervals for quantization (set as 10 in this example). As a result, the maximum and minimum quantization values are 10 and -10 respectively. Equations (5) and (6) illustrate the quantization in the right region and in the left one respectively.

$$Q'(i,j) = sign(H'(i,j)) \left[\frac{|H(i,j) - \omega|}{\Delta b_R} \right]$$
(5)

$$Q'(i, j) = sign(H'(i, j)) \left[\frac{|H(i, j) - \omega|}{\Delta b_L} \right]$$
(6)

where $\Delta b_{\rm L}$ and $\Delta b_{\rm R}$ represent the widths of quantization interval for the left region and for the right region respectively, and ω =*median*. In the numerical example, $\Delta b_{\rm L} = 17.9$ and $\Delta b_{\rm R} = 7.6$ to ensure both sides have 10 intervals. Figure 7 records the final quantization results.

10	4	3	0
-1	-9	3	-10
-1	3	0	0
-2	3	-7	4

Figure 7. Quantization results

3.2. Adaptive quantization decoding

The de-quantization process is formulated in Equation (7).

$$\mathbf{R}^{\prime}_{\mathcal{Q}(i,j)} = \begin{cases} (\mathcal{Q}(i,j)+r)\Delta b + \omega, \mathcal{Q}(i,j) > 0\\ (\mathcal{Q}(i,j)-r)\Delta b + \omega, \mathcal{Q}(i,j) < 0\\ 0, \quad otherwise \end{cases}$$
(7)

where $R'_{Q(i,j)}$ stands for the reconstructed coefficient of Q'(i,j), and r is the optional parameter mentioned in section 2.2. Figure 8 exhibits the de-quantization results of the numerical example (r=0).

76	30	23	0
-18	-158	23	-176
-18	23	0	0
-35	23	-123	30

Figure 8. The de-quantization results

Finally, the recovered values are right-shifted by the *median* as shown in Figure 9. Notice that the decoder requires three extra bytes (right interval width, left interval width, and *median*) to carry on the processes of de-quantization and shifting. Although trivial, these three extra bytes actually induce some overhead to the compression ratio. This effect is discussed in next section.

255	209	202	179
161	21	202	3
161	202	179	179
144	202	56	209

Figure 9. The final reconstructed results

4. Experiment results and analysis

In this section, we present the experiment results obtained from adaptive quantization and compare them with those obtained from uniform quantization. Seven 512 by 512, 8-bit gray images, Lena, Pepper, Boat, Elaine, Frog, Couple and Hill, are adopted as samples for simulations. For simplicity, one level 2-D Haar DWT is applied in the experiments. However, the proposed scheme can be conveniently extended for multiple level scenarios.

The number of quantization intervals has significant influence on reconstructed image quality. Therefore, choosing an appropriate number for quantization intervals is an essential issue. *Median*, the 2^{nd} Quartile (Q₂), is adopted as the quantization center because it divides data into two equal sizes. When determining the number of quantization intervals, the 1^{st} Quartile (Q₁) and 3^{rd} Quartile (Q₃) are also taken into consideration. Q₁, *median* (Q₂) and Q₃ divide data into four parts of the same amount. According to the distance between these quartiles, we recognize how intensive the DWT coefficients are distributed.

Since human eyes are more sensitive to low frequency signals, LL sub-band is the most crucial part and hence it is natural to allocate more quantization intervals on LL sub-band. We perform six various numbers of intervals on Boat to demonstrate such a phenomenon (Table 1 and Figure 10 summarize the corresponding image quality, where format *x-y-y-y* depicts that *x* intervals are allocated to LL sub-band and *y* intervals are allocated to each of the other sub-bands). Let us examine two sets of data in Table 1, (16-4-4-4, 32-4-4-4) and (32-2-2-2, 32-4-4-4). In the first one, (16-4-4-4, 32-4-4-4), we observe that using one more bit in every coefficient of LL sub-band improves the PSNR by 2.14dB. In the other set, (32-2-2-2, 32-4-4-4), we apply one more bit in each coefficient of LH, HL and HH sub-bands (three times the cost compared to the previous scenario) and obtain a inferior improvement (2.06dB). Hence, we conclude that increasing the number of intervals in LL sub-band is more efficient than increasing those in other sub-bands. Therefore, the number of intervals for the LL sub-band should be greater than those of the other three sub-bands. The numbers of quantization intervals for LL sub-band should be greater than those of the other three sub-bands are determined using Equations (8) and (9) respectively.

Number of Intervals	16-4-4-4	32-2-2-2	32-4-4-4	32-8-8-8	32-16-16-16	32-32-32-32
PSNR (dB)	29.72	29.80	31.86	34.27	36.20	37.34

Table 1. Number of intervals and PSNR



Figure 10. Numbers of intervals and PSNR

$$exp_LL = \left\lceil \log_2 \left(Maximum \left(\frac{Q_1 - \min}{Q_2 - Q_1}, \frac{MAX - Q_s}{Q_s - Q_2} \right) \right) \right\rceil + quality$$
(8)

exp_others = *quality*

(9)

where *quality* is a user-defined quantity which ranges from 1 to 5. Setting *quality*=1 provides the minimum image quality but superior compression ratio while setting *quality*=5 provides the maximum image quality but inferior compression ratio. According to various image quality demands, different numbers of quantization intervals can be set by selecting an appropriate value for *quality*. For LL sub-bands, Equations (8) and (9) ensure that region between Q_1 and *median* (or region between *median* and Q_3) is quantized with 2^{exp_LL} intervals and the other three sub-bands are quantized with fewer intervals (2^{exp_others}). Table 2 exhibits the resulting PSNR and compression ratio for Boat under different numbers of intervals.

Quality	Number of intervals	PSNR(dB)	Compression Ratio (%)
1	16-2-2-2	28.35	34.3796
2	32-4-4-4	31.86	46.8796
3	64-8-8-8	35.84	59.3796
4	128-16-16-16	40.03	71.8796
5	256-32-32-32	44.20	84.3796

 Table 2. PSNR and compression ratio under various quality settings

In the following, we adopt two specific examples to demonstrate the utilizing of Equations (8) and (9). The distribution for LL sub-band in Lena is shown in Figure 11 where we can see that min=109, $Q_1=358$, median=515, $Q_3=629$ and MAX=956. The distances between these five points are 249, 157, 114 and 327 and hence we observe that the coefficients are centralized between Q_1

and Q₃ (Half of the coefficients locate within two smaller distances, 157 and 114). The major concern of Equation (8) is to ensure that we allocate more intervals in data-intensive region. The computation of *exp_LL* for Lena using Equation (8) is as follows: *exp_LL* = $\lceil \log_2(Maximum(1.59, 2.87)) \rceil + quality = 2 + quality.$



As an alternative example, the distribution for LL sub-band in Boat is shown in Figure 12 where we can see that *min*=64, Q₁=439, *median*=633, Q₃=688 and *MAX*=915. The distances between these five points are 375, 194, 55 and 227 and the computation of *exp_LL* for Boat using Equation (8) is as follows: $exp_LL = \left\lceil \log_2(Maximum(1.93, 4.13)) \right\rceil + quality = 3 + quality$.



Figure 12. The distribution of LL coefficients in boat

Setting the *quality=3*, the parameter set (*exp_LL*, *exp_others*) for Lena and Boat are (5, 3) and (6, 3), respectively. As a result, the quantization intervals settings for them are 32-8-8-8 and 64-8-8-8, respectively. Recall that Δb_L and Δb_R represent the widths of quantization interval for the left region and for the right region respectively. In these two specific examples, Lena (Δb_L , Δb_R) = (12.7, 14.8) and Boat (Δb_L , Δb_R) = (8.9, 4.4). We observe that when the distribution is uniform-like (Figure 11), the resulting Δb_R and Δb_L are close where they differ significantly when the distribution is one-sided displaced (Figure 12).

The key parameter for image quality comparison is the PSNR (Peak Signal to Noise Ratio) which is defined in Equations (10) and (11). Setting *quality=3* in Equations (9), we compute the corresponding *exp_LL* for each sample using Equation (8). The experimental results are

illustrated in Table 3 and Figure 13. The second column in Table 3 is obtained by replacing the *median* by *mean* as the center of quantization [17].

$$PSNR = 10\log_{10}\frac{255^2}{MSE}$$
(10)

$$MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x_{ij} - x_{ij}^{'})^{2}$$
(11)

Samples	Adaptive Quantization by median (dB)	Adaptive Quantization by mean (dB)	Uniform Quantization(dB)	Number of Intervals
Lena	35.99	35.77	35.85	32-8-8-8
Pepper	34.50	34.27	33.38	32-8-8-8
Boat	35.84	36.43	29.1	64-8-8-8
Elaine	35.35	35.14	34.39	32-8-8-8
Frog	31.38	30.76	28.89	64-8-8-8
Couple	32.53	32.40	31.82	64-8-8-8
Hill	34.03	34.56	26.81	32-8-8-8

Table 3. PSNR with seven samples, using quality=3



As shown in Figure 13, both adaptive approaches outperform the uniform quantization for all samples. However, Boat and Hill which possess a one-sided distribution (see Figure 12) produce greater improvements than the other samples. As a heuristic rule, *median* and *mean* should be chosen as the center of quantization while the images under compression are uniform-like distributed and one-sided distributed, respectively.

The calculation of compression ratio is demonstrated as follows. For the 32-8-8-8 (*quality=3*) case, the average bit number per pixel is (6+4+4+4) / 4 = 4.5 (bits required for LL, HL, LH, HH are 6, 4, 4, 4 respectively). Hence, the compression ratio for uniform quantization is $(4.5 \times 512 \times 512 + 64) / (8 \times 512 \times 512) = 56.2531\%$ (recall that there are also 8 bytes overhead for uniform quantization to determine Δb and ω for all 4 sub-bands). For Adaptive Quantization, it is $(4.5 \times 512 \times 512 \times 512 + 96) / (8 \times 512 \times 512) = 56.2546\%$ since the overhead is now increased to 96 bits per image (3 bytes per sub-band and hence 12 bytes totally). The experimental results are depicted in Table 4 and Figure 14.

	Adaptive Quantization (%)	Uniform Quantization (%)	Number of Intervals		
Lena	56.2546	56.2531	32-8-8-8		
Pepper	56.2546	56.2531	32-8-8-8		
Boat	59.3796	59.3781	64-8-8-8		
Elaine	56.2546	56.2531	32-8-8-8		
Frog	59.3796	59.3781	64-8-8-8		
Couple	59.3796	59.3781	64-8-8-8		
Hill	56.2546	56.2531	32-8-8-8		

Table 4. Compression Ratio with seven samples, using quality=3



Figure 14. Compression Ratio with seven samples, using quality=3

As demonstrated in the above tables, AQ outperforms the uniform quantization in average when the compression ratio is almost the same. The inevitable overhead for AQ is that we have to store and transmit three extra values (right interval width, left interval width, and *median*) for each sub-band. As a result, 12 extra memory bytes are required for one image.

Using adaptive quantization and uniform quantization to process seven samples, the image quality, compression ratio and resulting images are summarized in Table 5.

Samples	Adaptive Quantization	Uniform Quantization		
Boat				
	PSNR = 35.84(dB)	PSNR = 29.10(dB)		
	CR = 59.3796%	CR = 59.3781%		
Couple				
	PSNR =32.53 (dB)	PSNR = 31.82(dB)		
	CR = 59.3796%	CR =59.3781 %		



Elaine		
	PSNR = 35.35(dB)	PSNR = 34.39(dB)
	CR = 56.2546%	CR = 56.2531%
Frog		
	PSNR = 31.38(dB)	PSNR = 28.89(dB)
	CR = 59.3796%	CR = 59.3781%

Hill		
	PSNR = 34.03(dB)	PSNR = 26.81(dB)
	CR = 56.2546%	CR = 56.2531%
Lena		
	PSNR = 35.99(dB)	PSNR = 35.85(dB)
	CR = 56.2546%	CR = 56.2531%



Obviously, test samples like Boat, Frog, and Hill get great improvement in both PSNR values as well as naked eyes. This is because they possess one-sided distributed histograms while other ones distribute more uniformly.

5. Conclusions

In this paper, we propose a simple yet efficient scheme which applies two different interval widths to a DWT sub-band. Although some extra memory bytes are required, the amount is negligible compared to the size of original images. According to the experimental results, the proposed approach (Adaptive Quantization) leads to a superior image quality while maintaining the same compression ratio. For images with larger variance values, the adaptive quantization scheme performs even more efficiently and provides more significant improvements. Furthermore, higher resolution applications, which demand more quantization intervals, are also suitable scenarios to apply such an adaptive quantization scheme.

References

- [1] Pennebaker, W. B. and Mitchell, J. L. 1993. "JPEG: Still Image Data Compression Standard". Van Nostrand Reinhold. NY.
- [2] ISO/IEC 10918-3 and ITU-T Recommendation T. 81, 1996. Information technology- digital compression and coding of continuous-tone still image: extensions.
- [3] ISO/IEC 15444, 2000. Information Technology: JPEG2000 image coding system.
- [4] Balster, E. J., Fortener, B. T., and Turri, W. F. 2010. "Rate-distortion development for optimal truncation in JPEG2000 imagery". Proceeding of the 4th International Symposium on Communications, Control and Signal Processing (ISCCSP).

- [5] Tsai, P. S. and Suzuki, R. 2008. "Graphics Image Compression Using JPEG2000". IEEE computer society Congress on Image and Signal Processing.
- [6] Christopoulos, C., Skodras, A., and Ebrahimi, T. 2000. The JPEG 2000 still image coding system: an overview. *IEEE Transactions on Consumer Electronics*, 46: 1103-1127.
- [7] Diego, S. C. and Ebrahimi, T. 2000. An Analytical Study of JPEG2000 Functionalities. *IEEE Conference on Image Processing*, 2:49-52.
- [8] Moccagatta, I., Soudagar, S., Liang, J., and Chen, H. 2000. Error Resilient Coding in JPEG2000 and MPEG4. *IEEE Journal on Selected Areas in Communications*, 18: 899-914.
- [9] Christopoulos, C., Askelof, J., and Larsson, M. 2000. Efficient Methods for Encoding region of Interest in the Upcoming JPEG2000 Still Image Coding Standard. *IEEE Signal Processing Letters*, 7: 247-249.
- [10] Biligin, A., Sementilli, P. J., and Marcellin, M. W. 1999. Progressive Image Coding Using Trellis Coded Quantization. *IEEE Transactions on Image Processing*, 8: 1638-1643.
- [11] Jiang, W. and Ortega, A. 2001. Lifting Factorization-Based Discrete Wavelet Transform Architecture Design. *IEEE Transactions on Circuits and Systems for Video Technology*, 11: 651-657.
- [12] Chen, P. Y. and Liao, E. C. 2002. A New Algorithm for Haar Wavelet Transform, *IEEE International Symposium on Intelligent Signal Processing and Communication System*.
- [13] Li, B. F., Dou, Y., and Shao, Q. 2008. Efficient Memory Subsystem for High Throughput JPEG2000 2D-DWT Encoder. *2008 Congress on Image and Signal Processing*, 1: 529-533.
- [14] Albanesi, M. G. and Guerrini, F. 2003. An HVS-based adaptive coder for perceptually lossy image compression. *Pattern Recognition*, 36: 997-1007.
- [15] Fujiyoshi, M. and Kiya, H. 2004. An image-quality guaranteed method for quantization-based watermarking using a DWT. 2004 International Conference on Image Processing (ICIP), 4: 2629-2632.
- [16] Phadikar, A., Maity, S. P., and Kundu, M. K. 2008. Quantization Based Data Hiding Scheme for Efficient Quality Access Control of Images Using DWT via Lifting. *Sixth Indian Conference on Computer Vision, Graphics & Image Processing.*
- [17] Chen, P. Y., Wu, S. H., and Chen, G. F. 2007. An Adaptive Quantization Scheme for JPEG2000. 2007 6th Symposium on Management of Information Technology and Personal Training.