Application of SY JND Model to SPIHT LLCVD

Day-Fann Shen and Jen-Hsing Sung
Signal Compression and Multimedia Communication Laboratory
Department of Electrical Engineering
Yunlin University of Science and Technology, Douleo, Yunlin, Taiwan

Abstract

The wavelet based SPIHT algorithm has recently attracted significant attentions from researchers in image coding. SPIHT maximizes PSNR (or minimize MSE) for each bit it transmitted, however, it is well known that PSNR is not a good indicator of the perceptual image fidelity. Human visual perception is more sensitive to distortions in lower frequency components than distortions in the higher frequency components. In this paper, we enhance SPIHT by introducing a process called JND_SQ, between wavelet transform and SPIHT. JND_SQ allows SPIHT to maximize perceptual fidelity instead of PSNR for each bit transmitted. We proposed a perceptual fidelity criterion called JND_PSNR based on the SY JND model, the new criterion is more consistent with human perception than PSNR. Experimental results show that JND_SPIHT outperforms SPIHT in terms of JND_PSNR. The JND_SQ process results in a large amount of zeros, which can be further utilized to improve the speed of SPIHT. Finally, JND_SQ can be turned off if the original SPIHT is desired.

1. Introduction

The wavelet based SPIHT[1] algorithm has attracted significant attentions from researchers in image coding. SPIHT maximizes PSNR (or minimize MSE) for each bit it transmitted, however, it is well known that PSNR is not a good indicator of the perceptual image fidelity. Human visual perception is more sensitive to distortions in lower frequency components than distortions in the higher frequency components.

In this paper, we enhance SPIHT by maximizing perceptual fidelity instead of PSNR for each bit transmitted. To achieve this goal, a quantization process called JND_SQ is introduced between wavelet transform and SPIHT. In the JND_SQ, wavelet coefficients in each subband are scalar quantized using a set of quantization step sizes proposed in SY JND model,[2] where lower frequency subbands has smaller step size. Thus, JND_SQ modifies wavelet coefficients to reflect the importance of the sub-bands to human visual perception.

Why do we choose SY JND model? There are many wavelet based JND models proposed in the past years,[2,3,4] among them the JND model by Watson et. al.[3] has been the most referenced. However, up-to-date there is no comparison and evaluation to these visual models. To select the best JND model, we compared and evaluated JND models by SY (Shen and Yan)[1,2] and Watson et. al.[3] based on a restrict subjective observations.[5] The results show that the SY JND model outperforms Watson's,[5] some important data are shown in Appendix at the end of the paper. For this reason, we adopt quantization step sizes proposed in SY JND model.[5] For performance evaluation purpose, we proposed JND_PSNR, which is more consistent with human perception than PSNR. Experimental results show that JND_SPIHT outperforms SPIHT in terms of JND_PSNR. The JND_SQ process results in a large amount of zeros, which can be further utilized to improve the speed of SPIHT. Finally, JND_SQ can be turned off if the original SPIHT is desired.

This paper is organized as follows: JND_SQ and JND_SPIHT are described in section 2; the perceptually turned image fidelity criterion JND_PSNR is illustrated in section 3. Effects of JND_SQ and comparisons between JND_SPIHT and SPIHT is conducted in section 4, conclusions and discussion are given in section 5, and finally, further research topics are presented in section 6.

2. JND_SPIHT and JND_SQ –The Algorithm and The JND based Scalar Quantization

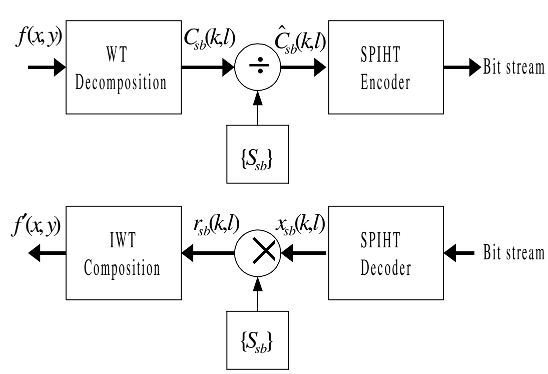

Figure 1. The JND_SPIHT algorithm

Human visual perception emphasizes more on the lower frequency components than the higher frequency components, a larger wavelet coefficient $C_{sb}(k,l)$ in higher frequency subband may be less important to human perception than a smaller one in lower frequency subband. For this reason, transmitting coefficients in the order of the magnitudes, as in SPIHT, can maximize PSNR (or minimize MSE), but does not necessary maximize the perceptual quality. To reflect the importance of a coefficient to visual perception, we introduce the JND_SQ process between the SPIHT and wavelet transform. x_{sb} is the partially decoded coefficient value, which is then multiplied by the corresponding S_{sb} before inverse wavelet transform at the end of each map.

A. JND_SQ – the JND based Quantization

Each wavelet coefficient $C_{sb}(k,l)$ is quantized by a step size S_{sb} of the corresponding subband sb.

$$\begin{cases} \hat{C}_{sb}(k,l) = round\left[\dfrac{C_{sb}(k,l)}{S_{sb}}\right] \\ r_{sb}(k,l) = x_{sb}(k,l)\cdot S_{sb} \end{cases}, \quad S_{sb} = \Gamma_{sb}\cdot\phi .$$

Where Γ_{sb} is the nominal quantization step size for subband sb, two sets of $\{\Gamma_{sb}\}$ proposed by Shen and Yan (SY) and Watson are shown in Appendix (Table 2(a)(b)). It is noted that Γ_{sb} for lower frequency subbands are smaller than that for higher frequency subbands, thus, reflecting the importance of subband sb to human perception. After adjustment by JND_SQ, a larger $\hat{C}_{sb}(k,l)$ is more importance than a smaller one. Furthermore, $\{\hat{C}_{sb}(k,l)\}$ require fewer symbols and has lower entropy, which can be exploited to improve the coding efficiency. ϕ is the compression control factor for trade offs between bit rates and image qualities. It is noted that the value of ϕ must be restricted so that S_{sb} does not smaller than 1. For ϕ =1, the reconstructed image is visually lossless at a nominal viewing distance of 60 cm [2]. Better image quality can be obtained with a smaller value of ϕ, however, $S_{sb} > 1$ should be hold for most cases.

3.JND_PSNR – The JND Based Image Fidelity Criterion

PSNR is common used to indicate the fidelity of the decoded image to its original image. PSNR is normally obtained from the spatial domain. In energy conserving wavelet transforms (with orthogonal basis), PSNR from the transformed domain by summing up the MSEs from all subbands. It is noted that coefficients from all subbands are considered as the same importance in PSNR. The JND_PSNR described below is an indicator reflecting the visual importance.

$$JND\_PSNR = 10\log\frac{255^2}{MSE_{JND}}, \quad MSE_{JND} = \frac{\sum_{sb=1}^{n} JND\_Err_{sb}}{N},$$

Where *sb* is the subband, *n* is the total number of subbands and N is the number of total pixels. And

$$JND\_Err_{sb} = \sum_k\sum_l\left[\Psi(\left|c_{sb}(k,l) - r_{sb}(k,l)\right| - S_{sb}/2)\right]^2 / w_{sb}^2,$$

$$x = \left|c_{sb}(k,l) - r_{sb}(k,l)\right| - S_{sb}/2, \begin{cases}\Psi(x) = x, x > 0 \\ \Psi(x) = 0, x \le 0\end{cases}, \quad w_{sb} = \frac{\Gamma_{sb}}{6.0}.$$

Where w_{sb} is the visual weight for subband *sb,* the weight for the lowest subband is normalized to w_{sb} = 1.0. It is noted that error within $S_{sb}/2$ is considered as zero and that image quality after JND_SQ is infinity in terms of JND_PSNR by the above definition.

4. Comparisons of JND_SPIHT and SPIHT

We compare SPIHT and JND_SPIHT from three aspects: (A) bit rates vs. JND_PSNR (B) number of non-zero coefficients, number of maps and the computational complexity. Daubecies 9/7 filter, 16 wavelet subbands and adopts SY JND model.

A. Bit rates vs. Image fidelity

The bpp vs. JND_PSNR curves for Lena 512x512 in the range in 0.2 and 1.2 are plotted in Figure 2.

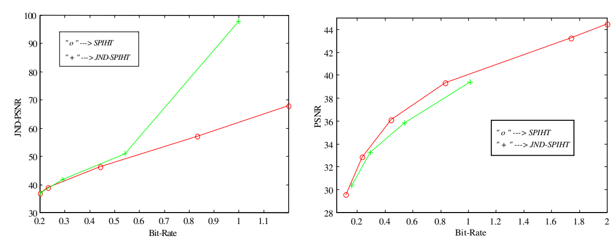

Figure 2. Comparison of bit rates vs. Image fidelity By SPIHT and JND_SPIHT on 512x512 Lena. (a) JND_PSNR (b) PSNR. (uncode).

When bit rates under 0.3bpp, JND_SPIHT performs about the same as SPIHT, because the coefficients selected are primarily in the lower frequency subbands. Those important coefficients select by JND_SPIHT are also selected by SPIHT, thus no big difference in the decoded image quality. However, JND_SPIHT outperforms SPIHT in terms of JND_PSNR for bit rates exceed 0.3bpp, because JND_SPIHT is designed to maximize JND_PSNR while SPIHT is design to maximize PSNR for each bit transmitted. At the end of the 9th map, the JND_SPIHT decoder receives 0.54bpp and the reconstructed image fidelity is 50.3db in JND_PSNR. JND_SPIHT completes the coding process at bit rate of 1.0143bpp in the 10th Map. By definition, the JND_PSNR reaches infinity when all nonzero coefficients are received by the decoder, the corresponding PSNR is 39.43dB. How good is the subjective image quality when JND_SPIHT completes the transmission? We use a subjective SPS technique[5] to examine the JND_SPIHT reconstructed Lena (SY model with ϕ =1, JND_PSNR=infinity, PSNR=39.43). For 6 human testers, the average $cd_{VLL}(Lena, \phi = 1)$ (Critical Distance of Visually Loss less) is 74 cm (which has some deviation from the nominal 60 cm), excellent image fidelity for most applications! At similar bit rate of 1.014bpp, the SPIHT decoded image is 62.87dB in JND_PSNR, 40.70 db in PSNR, and average $cd_{VLL}(Lena)$ is 82 cm. *This illustrates a fact that JND_SPIHT reaches better Subjective image quality at the same bit rate*. The results from the subjective tests are consistent with JND_PSNR, indicating JND_PSNR is more consistent with human perception than PSNR.

Table 1. Number of non-zero coefficients in each subband after JND_SQ.

Subband \ Test image	Before JND_SQ	F16g After JND_SQ	Lena After JND_SQ	Salesman After JND_SQ
HH1	65536	17	119	2
LH1	65536	7081	3182	699
HL1	65536	7876	7082	1331
HH2	16384	4150	3564	179
LH2	16384	9245	7315	6546
HL2	16384	8893	9046	7563
HH3	4096	2577	2395	2134
LH3	4096	3207	2768	3251
HL3	4096	3004	3257	3294
HH4	1024	862	856	906
LH4	1024	948	919	960
HL4	1024	907	962	971
HH5	256	242	246	248
LH5	256	251	246	254
HL5	256	247	253	252
LL5	256	256	256	256
Total	262144	49763	42466	28846
		19 %	16 %	11 %

(a)

Map	F16g Before JND_SQ	F16g After JND_SQ	Lena Before JND_SQ	Lena After JND_SQ	Salesman Before JND_SQ	Salesman After JND_SQ
1	233	129	55	170	24	122
2	23	123	174	78	160	106
3	14	10	30	38	70	33
4	96	29	78	106	39	92
5	368	180	212	365	197	325
6	1245	637	769	1090	612	994
7	3454	1949	1880	2444	1544	1993
8	5103	3230	3976	4469	2784	3796
9	9649	5864	7386	8564	4592	6824
10	17239	27229	14398	25132	7456	14561
11	29785	*	34433	*	12226	*
12	47968	*	63096	*	20199	*
13	53418	*	59434	*	30171	*
Total	168595	39380	185921	42456	80074	28846
		23.36 %		22.84 %		36.02 %

(b)

SPIHT continues to transmit the remaining coefficients in three extra maps. The bit rates and corresponding JND_PSNR pairs are (0.83bpp, 57.2dB), (1.73bpp, 85.6dB) (3.07bpp, infinity) for the 11th, 12th, 13th map respectively. It is noted that transmission of these coefficients in the last maps contributes only slightly to the visual perception, but very costly in bit rate. We use a Pentium 600Mhz PC and MATLAB program in the comparisons of computational complexity. To complete the whole process, 20396 and 24795 seconds are required for coding and decoding by SPIHT, while only 1850 and 1693 seconds (or 1/10 of the SPIHT) are required by JND_SPIHT. As expected, Figure 2(b) shows that SPIHT outperforms JND_SPIHT in terms of PSNR for all bit rates.

B. Number of non-zero coefficients and maps

Some detailed analysis is described here: Coefficients smaller than $S_{sb}/2$ are set to zero after JND_SQ as shown in Table 1 (a) and (b). The number of non-zero coefficients is reduced significantly. We noted that most coefficients in lower frequency subbands are kept intact for the lowest frequency subband. Large amount of coefficients are set to zeros in the higher frequency subbands, because those coefficients are smaller in magnitude and their corresponding quantization step size S_{sb} is larger. Note that only 19%, 16% and 11% of coefficients are sufficient to reconstruct visually lossless images at a nominal viewing distance of 60 cm for F16, Lena and Salesman respectively, the remaining coefficient are considered unimportant and are set to zero by JND_SQ. This is an advantage to be exploited for a faster algorithm.

Furthermore, the magnitude of $\{\hat{C}_{sb}(k,l)\}$ is only $1/S_{sb}$ of the original $C_{sb}(k,l)$, the effects are two folds: (1) The number of maps is reduced from 13 to 10 and (2) the number of symbols required to represents $\hat{C}_{sb}(k,l)$ is reduced, thus entropy coding can be adopted to exploit this advantage.

5. Conclusion and Discussion

JND_SPIHT is characterized by inserting JND_SQ process between the SPIHT and wavelet transform; it maximizes JND_PNSR while SPIHT maximizes PSNR. Visually lossless image fidelity is an option in JPEG2000, when desired; JND_SQ may be turned on to obtain a visually lossless image with image quality adjusted by compression control factor ϕ. JND_SPIHT improves the perceptual performance by maximizing JND_PSNR instead of PSNR. In addition, JND_SPIHT completes the coding and decoding process at the speed about 1/10 of SPIHT. Even SPIHT truncates the last few maps to improve the efficiency; JND_SPIHT still gets better image quality in terms of JND_PSNR as shown in Figure 2(a) and subjective measurements.

6. Further Research

Two advantages created by the JND_SQ can be exploited further. (1) A large amount of non-significant coefficients are quantized to zero, only those significant coefficients are kept (19%, 16% and 11% for F16, Lena and Salesman for $\phi = 1$. Furthermore, SPIHT (therefore JND_SPIHT) does not transmit the LBS bit for all coefficient, thus, coefficients JND_SQ with value

This fact can be exploited for a faster algorithm, where zero coefficients be ignored (2) the number of symbols for representing $\hat{C}_{sb}(k,l)$ are reduced, so is the entropy. Certain type of efficient entropy coding can be introduced to improve the coding efficiency.

(PSNR or JND_PSNR are inadequate for very high quality images) We realized that JND_PSNR is not a good indicator for human perception when quality is high. For example: JND_MSE = 1.0, JND_PSNR=48 db, JND_MSE = 0.1, JND_PSNR=68 db, JND_MSE = 0.01, JND_PSNR = 88db, JND_MSE = 0.001, JND_PSNR = 108db, JND_MSE = 0.0001, JND_PSNR = 128dB. For JND_MSE of 0.01 and 0.001, their image qualities are basically the same, but their JND_PSNR difference is 20 dB.

Our research goal is to find a new objective image fidelity criterion (like JND_PSNR which can be calculated by computer without resorting to human testers) for high quality images. The new criterion should be a good predictor of the subjective CDVLL (critical distance of visually lossless) – a precise subjective image fidelity criterion.[5]

7. Acknowledgement

This paper is supported by NSC89-2213-E-224-039 of Taiwan government. Special thanks to W.C. Fan, M.Z. Lin, J.G. Fun, Z.H. Shen, F.K. Tsai and Y.Z. Shin for their efforts in subjective testings.

8. References

1. Said, A., Pearlman W. A. " A new fast and efficient image codec based on set partitioning in hierarchical trees " IEEE Trans. on circuits and systems for video technology, Vol. 6 June 1996.
2. Day-Fann Shen and Loon-shan Yan, " JND measurements and Wavelet based image coding ", SPIE International Optoelectronics Exposition, July 1998.
3. Andrew B. Watson, Gloria Y. Yang, Joshua A.Solomon, and John Villasenor, " Visibility of Wavelet Quantization Noise " IEEE Trans. image Processing, Vol. 6, pp, 1164-1175, 1997.
4. Rajala J Rober E. Van Dyck and Sarah A. Rajala, Subband/VQ Coding of Color Images With Perceptually Optimal Bit Allocation, IEEE Transactions on circuits and systems for video technology, vol 4.NO.1. February,1994
5. Shen, D. F, Sung, J. S. Fan, W.C, M.Z. Lin, J.G. Fun, Z.H. Shen, F.K. Tsai and Y.Z. Shin "VLLCVD Subjective Image Fidelity Criterion and Its Application to Evaluation of Visual Models" submitted to PICS 2001, Quebec, Montreal, Canada.

Biography

Day-Fann Shen received his diploma in EE from Taipei Institute of Technology, Taiwan in 1976. MS and PhD in Computer Engineering from University of Cincinnati and North Carolina State University in 1983 and 1992 respectively. 1988-1992 he worked for IBM, Research Triangle Park, NC. Since 1992 he has been with EE department, Yunlin University of Science and Technology, Taiwan. His work has primarily focused on human visual property, image/video processing/coding and watermarking. He is a member of SPIE and IEEE.