PERCEPTUALLY-ADAPTIVE PRE-PROCESSING FOR MOTION-COMPENSATED RESIDUE IN VIDEO CODING

X. K. Yang, W. S. Lin, Z. K. Lu, E. P. Ong and S. S. Yao

Institute for Infocomm Research
Agency for Science, Technology and Research, Singapore 119613.
Email: xkyang@ieee.org, {xkyang,wslin,zklu,epong,ssyao}@i2r.a-star.edu.sg

ABSTRACT

We present a perceptually-adaptive pre-processing scheme for motion-compensated residue, based on just-noticeable-distortion (JND) profile. Human eyes cannot sense any changes below the JND threshold around a pixel due to their underlying spatial/temporal masking properties. From the viewpoint for signal compression, smaller variance of signal results in less objective distortion of the reconstructed signal for a given bit-rate. In this paper, the JND profile is incorporated into a motion-compensated residue signal pre-processor for variance reduction towards coding quality enhancement. A solution of adaptively determining the parameter for the residue pre-processor is also proposed. Experimental results show that both perceptual quality and objective quality are enhanced in coded video at a given bit-rate.

1. INTRODUCTION

The goal of video compression and coding should aim at the highest perceptual quality with a given bit-rate, since the ultimate receiver of most decompressed video signal is the human visual system (HVS). It is imperative for us to design a coding algorithm that minimizes perceptual distortion between the original and the decoded visual signal.

Human eyes cannot sense any changes below the just noticeable distortion (JND) threshold around a pixel due to their underlying spatial/temporal sensitivity and masking properties [1]. An appropriate JND model can significantly help to improve the performance of video coding algorithms. Several methods for finding JND have been proposed, based upon intensive research in subbands (discrete cosine transform (DCT), wavelet domain, etc.) [2, 3, 4, 5], as well as some work in image-domain [6, 7, 8].

Perceptual coding has been so far focused mainly on determination of proper quantization steps for image coding with subband JND [2, 6, 5]. A few attempts have been made to non-standard video coding [7, 4]. In [7], image-domain JND has been used as the threshold for inter-frame replenishment with low-motion (like head-and-shoulder) scenes in low bit-rate videophony. In [4], subband JND has been used in the quantization process and also in controlling the block splitting process in variable-size motion search.

This paper aims at proposing a new perceptually-adaptive video coding scheme based on our previous JND profile [8]. It can be applied to any prevalent standardized hybrid video coding (cascading of motion estimation (ME) and DCT), such as H.261/263 and MPEG-1/2/4. If visual signal to be coded has smaller variance, less objective distortion is resulted in the reconstructed frame for a given bit-rate [9, 10]. We attempt to explore the methodology that reduces the variance of motion compensated residues in an inter-coded frame under the JND guidance. A JND-adaptive pre-processing module is proposed to achieve this by adding or subtracting an appropriate quantity regulated by the JND profile. A parameter used to adjust the said quantity is determined by minimizing the overall objective distortion for the current frame, i.e., the sum of the quantization distortion for the pre-processed residue signal and the distortion introduced by the residue signal pre-processor. The pre-processing module provides an extra design option for coding quality control, besides quantization.

2. JND-BASED PRE-PROCESSING OF MOTION-COMPENSATED RESIDUES

The framework for the proposed perceptually-adaptive JND-based pre-processing scheme for motion-compensated residue is given in Figure 1, where $I(x,y)$, $\tilde{I}(x,y)$, $R(x,y)$, $\tilde{R}(x,y)$ and $JND(x,y)$ respectively denote the original intensity, the motion compensated intensity, the residue signal before JND-adaptive pre-processing, the residue signal after JND-adaptive pre-processing and the JND value for a pixel located at (x,y) of the current frame; $JND(x,y)$ can be obtained using the JND estimator proposed in [8].

The addition of the proposed scheme to a typical hybrid video coding scheme is: before the DCT stage, each motion compensated residue is adjusted by a JND-adaptive pre-processor. The proposed scheme can be completely compatible with all current video coding standards, since it is

0-7803-8554-3/04/$20.00 ©2004 IEEE.

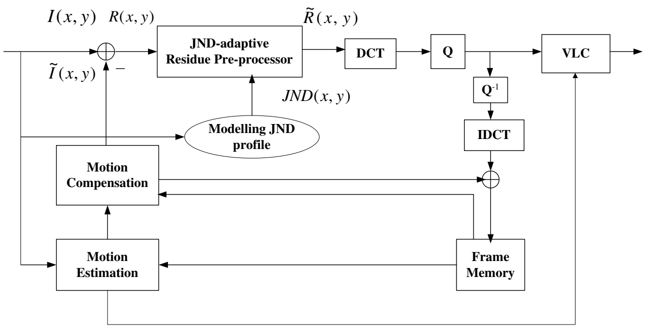

Fig. 1. The framework for the proposed perceptually-adaptive video coding scheme based on JND profile

unnecessary to transmit extra side-information and change the coding syntax.

The JND-adaptive residue pre-processor can be described as the following pseudo-codes:

/* **Algorithm 1**: JND-adaptive residue preprocessing */
if $R(x, y) - \overline{R_B} < -\lambda \cdot JND(x, y)$
  $\tilde{R}(x, y) = R(x, y) + \lambda \cdot JND(x, y)$
  else
    if $|R(x, y) - \overline{R_B}| \leq \lambda \cdot JND(x, y)$
      $\tilde{R}(x, y) = \overline{R_B}$
    else
      $\tilde{R}(x, y) = R(x, y) - \lambda \cdot JND(x, y)$

where $\overline{R_B}$ is the average of the 64 residues in the 8×8 block around (x,y); λ $(\in [0, 1])$ is a parameter to be determined. The constraint of $\lambda \in [0, 1]$ is to avoid introducing perceptual distortion into motion compensated residues in the JND-adaptive pre-processing process.

Signal distortion of a conventional video coder is introduced by the quantizer for DCT coefficients. With the proposed scheme, additional consideration about distortion is the JND-adaptive residue pre-processing. Let D_1 and D_2 respectively denote the distortion of coding the pre-processed residues and the distortion due to JND-adaptive pre-processing.

When the quantized data are entropy-coded, the MSE-based signal distortion is proportional to the variances of DCT coefficients [9, 10] except for very low bit rates (large distortion). It is expected that reduction of the variance of residue signal leads to reduction of the variances of DCT coefficients, and consequently leads to reduction of D_1, since the DCT coefficients are the decomposition of residue signal in frequency-domain.

As can be seen from Algorithm 1, the larger λ, the closer $\tilde{R}(x, y)$ will be to $\overline{R_B}$. Therefore, D_1 can be reduced by compressing the residues pre-processed with a larger λ. On the other hand, a larger λ will result in larger D_2. As the result, a tradeoff exists between D_1 and D_2, and an optimal value for λ exists for minimization of the overall distortion $D = D_1 + D_2$.

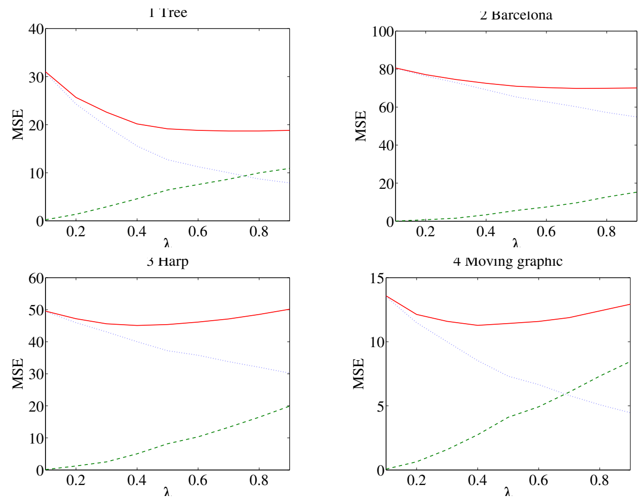

Fig. 2. Tradeoff between D_1 and D_2 with the relationship of λ for the first four sequences in VQEG dataset compressed at 5 Mbps (D_1, D_2 and $D = D_1 + D_2$ are illustrated by the doted, the dashed, and the solid lines respectively).

3. PARAMETER DETERMINATION

Based on the data collected from 20 video sequences compressed at bit-rate of 5 Mbps, D_1 and D_2 are statistically modelled as follows (Due to the limited space, we cannot present the statistical derivation in details):

$$D_1 = \epsilon^2 e^{-\alpha b} \sigma_{\mathbf{R}}^2 [1 - \psi(\sigma_{\mathbf{R}}^2)^{-\phi} \ln(\lambda^2 P_{JND}^2 + 1)] \quad (1)$$
$$D_2 = \omega \lambda^2 P_{\mathbf{JND}}^2 \ln(\sigma_{\mathbf{R}}^2 + 1) \quad (2)$$

with

$$P_{JND}^2 = \frac{1}{X \cdot Y} \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} (JND(x, y))^2 \quad (3)$$

where b is the number of bits assigned for coding the residue signal, $\sigma_{\mathbf{R}}^2$ is the variance of residue signal $\mathbf{R}$, $\epsilon^2 = 1.2$, $\alpha = 1.386$, $\psi = 0.4$, $\phi = 0.3$, and $\omega = 0.1$.

To explore the relationship between λ and the distortion (D_1, D_2 and D), 20 video sequences are used in experiments. These sequences are from Video Quality Expert Group (VQEG) [11] (listed in Table 1), and span a spectrum of various motion, zooming, color and texture. The first ten sequences are with frame rate of 25 fps and resolution of 720×576 pixels, and the rest are with frame rate of 30 fps and resolution of 720×480 pixels. Figure 2 shows D_1, D_2 and D as the functions of λ for the first four video sequences in VQEG dataset compressed at bit-rate of 5 Mbps. And all 20 sequences inclusive of the first four sequences have the similar behaviors to the curves plotted in Figure 2. It can be concluded from Figure 2 that there exists a minimum for each of these convex curves. As will be shown in

Section 4, the positions of the minima determine how well the proposed pre-processing can perform.

A minimum exists with D versus λ curves (as illustrated in the examples of Figure 5). Let $\frac{\partial D}{\partial \lambda} = 0$, we have

$$(\lambda^*)^2 = \Lambda \quad (4)$$

with

$$\Lambda = \frac{1}{P_{\mathbf{JND}}^2} \cdot \left(\frac{\psi \cdot \epsilon^2 \cdot e^{-\alpha \cdot \bar{b}} \cdot (\sigma_{\mathbf{R}}^2)^{1-\phi}}{\omega \cdot \ln(\sigma_{\mathbf{R}}^2 + 1)} - 1\right) \quad (5)$$

Considering $\lambda \in [0, 1]$, we finally get

$$\lambda^* = \begin{cases} 0, & \Lambda < 0 \\ \sqrt{\Lambda}, & 0 \leq \Lambda \leq 1 \\ 1, & \Lambda > 1 \end{cases} \quad (6)$$

4. OVERALL PERFORMANCE

The proposed JND-based pre-processing scheme has been implemented by incorporating the JND estimator and the JND-adaptive residue pre-processor into the MPEG-2 Test Model 5 (TM5) [12] coder. The twenty test sequences listed in Table 1 are used for experiments, since they represent a good diversity of visual signal. The target bit rate is set as 5 Mbits/s. The GOP setting is N=12 and M=3 (both P and B frames are present).

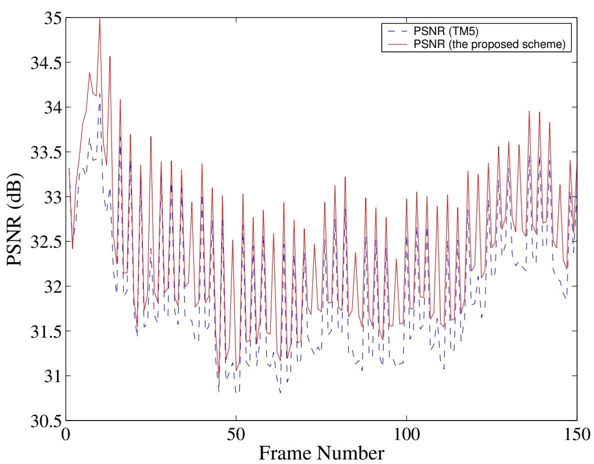

Fig. 3. PSNR comparison of the sequences compressed by the MPEG-2 TM5 coder and the coder with the proposed JND-adaptive pre-processor individually for *Harp*.

4.1. PSNR comparison

Figure 3 illustrates the overall PSNR comparison of the *Harp* sequence, between the original TM5 coder and the improved method with the proposed scheme. It can be seen that PSNR is improved with the proposed scheme. The results for all 20 video sequences are listed in Table 1. The PSNR is significantly improved for all other sequences except for the "Fries" sequence, in which a slight PSNR loss of 0.087 dB occurs; an average PSNR gain of 0.505 dB is achieved by the proposed scheme in comparison with the original TM5 coder.

Table 1. The overall signal distortion (PSNR) and perceptual distortion rating scores (DMOS) of the motion-compensated frames encoded by the original MPEG-2 TM5 coder and the coder with the proposed JND-adaptive pre-processor, respectively

Video sequence	PSNR (dB)		DMOS	
	TM5	The proposed	TM5	The proposed
1.Tree	35.124	36.961	6.4	4.3
2.Barcelona	30.281	30.862	38.5	24.3
3.Harp	31.994	32.411	31.0	21.4
4.Moving graphic	37.645	38.650	31.5	22.4
5.Canoa Valsesia	29.261	29.481	64.8	56.5
6.F1 Car	30.637	31.005	49.5	41.9
7.Fries	37.279	37.192	30.0	20.3
8.Horizontal scrolling2	30.857	31.812	41.0	29.9
9.Rugby	28.401	28.667	37.9	34.3
10.Mobile&calendar 1	29.453	29.880	35.6	29.7
11.Baloon-pops	33.586	33.796	34.2	26.7
12.NewYork 2	39.597	39.930	22.9	9.0
13.Mobile&Calendar 2	27.479	27.779	39.5	26.4
14.Betes pas betes	42.451	42.667	13.6	6.7
15.Le point	34.247	34.536	46.7	36.3
16.Autumn leaves	37.677	38.422	13.3	5.6
17.Football	31.664	31.952	37.3	32.5
18.Sailboat	36.222	37.371	10.2	6.4
19.Susie	41.212	41.446	11.1	6.2
20.Tempete	32.118	32.471	13.3	9.7

There is close correlation between the PSNR gain and the location of the minimum in D-versus-λ curve in Figure 2. If the minimum occurs around $\lambda = 1$ for a sequence, the resultant PSNR gain is substantial; if the minimum occurs close to $\lambda = 0$ for a sequence, the resultant PSNR gain is small; and there are a number of cases in which λ and the resultant PSNR gain lie in between. This demonstrates that the proposed scheme is able to determine λ^* adaptively with different visual signal contents.

4.2. Subjective Quality Evaluation

In order to further confirm the coding quality improvement by the proposed scheme, we performed subjective quality evaluation. Double Stimulus Continuous Quality Scale method, as in Rec. ITU-R BT.500 [13], was used to evaluate the subjective quality of a decoded sequence relative to its original sequence. A decoded sequence is obtained with the MPEG-

2 TM5 coder or the improved coder with the proposed JND-adaptive scheme. Each display session for an original sequence and an associated decoded sequence is: Video Sequence 1, two seconds of grey screen, Video Sequence 2, two seconds of grey screen. The display repeats twice before the viewers are requested to vote for the quality of each sequence. Both the display order of the sequences in a session and the order of the 20 test sequences were randomized for viewers. The Mean Opinion Score (MOS) scales for viewers to vote for the quality after viewing are: Excellent (100-80), Good (80-60), Fair (60-40), Poor (40-20) and Bad (20-0).

Eleven observers (five of them are with average image processing knowledge and the rest are naive) were involved in the experiments. Their eyesight is either normal or has been corrected to be normal with spectacles. The subjective visual quality assessment was performed in a typical laboratory environment with normal fluorescent ceiling light, using a 21" EIZO T965 professional color monitor with resolution of 1600×1200. The viewing distance is approximately six times of the image height.

Difference Mean Opinion Scores (DMOS) are calculated as the difference of MOSs between the original video and the processed video. The smaller the DMOS is, the higher perceptual quality of the processed video has when compared with the original video. Table 1 compares the averaged DMOSs over the all 11 observers for the 20 sequences. From the table, we can see that the subjective rating is consistently better for the decoded sequences with the proposed scheme, and an average subjective quality gain of 7.9 measured in DMOS is achieved by the proposed scheme.

5. CONCLUSIONS

Based on an image-domain JND profile, a perceptually adaptive pre-processor for motion-compensated residues in video encoder has been proposed, demonstrated with a wide variety of test video sequences and subjective viewing. The proposed technique can be applied to any standardized video coding scheme, such as H.261/263 and MPEG-1/2/4. It improves both objective coding quality (PSNR) and perceptual quality of the decoded images for a given bit-rate. The major contributions of this paper include: 1) The JND profile has been incorporated into a residue signal pre-processor to achieve effective reduction of the variance for the residue signal after motion compensation. The resultant video coding process has an extra designing option of quality improvement, besides quantization. 2) A method of determining the optimum parameter for the pre-processor has been also devised for improvement of both PSNR and perceptual quality.

6. REFERENCES

[1] N. S. Jayant, J. D. Johnston, and R. J. Safranek, "Signal compression based on models of human perception," Proc. IEEE, vol. 81, pp. 1385–1422, 1993.

[2] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in Proc. SPIE Interntional Conference on Human Vision, Visual Processing and Digital Display, vol. 87, 1913 1993.

[3] A. P. Bradley, "A wavelet visible difference predictor," IEEE Trans. Image Processing, vol. 8, no. 5, pp. 717–730, May 1999.

[4] J. Malo, J. Gutierrez, I. Epifanio, F. Ferri, and J. M. Artigas, "Percetual feedback in multigrid motion estimation using an improved dct quantization," IEEE Trans. Image Processing, vol. 10, no. 10, pp. 1411–1427, October 2001.

[5] I. Höntsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," IEEE Trans. Image Processing, vol. 11, no. 3, pp. 213–222, March 2002.

[6] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-d subband image codec for video communication over wireless channels," IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 2, pp. 143–156, 1996.

[7] Y. J. Chiu and T. Berger, "A software-only videocodec using pixelwise conditional differential replenishment and perceptual enhancement," IEEE Trans. Circuits Syst. Video Technol., vol. 9, no. 3, pp. 438–450, April 1999.

[8] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images," in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2003), vol. 3, Hong Kong, April 2003, pp. 609–612.

[9] T. Berger, Rate Distortion Theory. Englewood Cliffs, NJ: Prentice, 1971.

[10] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application–part i: Fundamental theory," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 2, pp. 287–298, April 1997.

[11] VQEG, "Final report from the video quality experts group on the validation of objective models of video qualtiy assessment," [Online], available: www.its.bldrdoc.gov/vqeg/, March 2000.

[12] ISO/IEC JTC1/SC29 WG11, MPEG-2 Test Model 5, ISO Std., Apr 1993.

[13] ITU-R, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R Rec. BT. 500-9, Std., 1999.