AVS Encoding Optimization with Perceptual Just Noticeable Distortion Model

Qi Cai, Li Song
Institute of Image Communication and Information Processing
Shanghai Jiao Tong University
Shanghai, China
Email: {caiqi0132069,song_li}@sjtu.edu.cn

***Abstract*—Integrating efficient visual perceptual cues into standardized video coding framework can improve performance significantly. In this paper we propose to enhance AVS encoder by using the latest just noticeable distortion (JND) model to adjust DCT coefficients of prediction residues in a content adaptive way. To better modeling JND profile in AVS integer DCT domain, we further derive the JND mapping from the classical DCT domain to AVS Integer DCT domain. The experiment shows that the proposed algorithm can reduce the bitrate by about 13% on average, compared to the AVS standard encoder at similar visual quality.**

Keywords-AVS; residues suppression; just noticeable distortion

I. INTRODUCTION

Initiated by Audio and Video Coding Standard Working Group of China, AVS now is an IEEE compression standard for digital audio and digital video. The Part II of AVS (we will refer to this as AVS for simplicity in the rest of the paper), the video part, is competing with H.264/AVC as the next generation of MPEG-2 [1]. Similar to AVC, the video coding part of AVS is mainly designed to reduce the statistical redundancy between neighboring pixels, while the redundancy related to Human Visual System (HVS) is not fully exploited. As HVS is the final receiver of most of the images and videos, if we can utilize the masking effect and remove the less sensitive information during the encoding process, further coding gain could be achieved. One of important HVS masking models, the Just Noticeable Distortion (JND), which refers to the minimum distortion that can be perceived by HVS with respective to the original video, has been widely used to improve video coding performances. Yang et al. [2] used a pixel-wise JND model for residual processing in MPEG-2. In context of H.264/AVC, methods like DCT-domain JND guided coefficients shaping [3][4][5] or adaptive quantization [6] were proposed to remove the unperceived residuals. However, they neglected the specific differences between the H.264/AVC integer DCT and the classical DCT, and the latter is always misused to derive JND threshold. In this paper, we further extend previous works, e.g. [3], to AVS framework by introducing a JND guided soft-thresholding technique to suppress prediction residues before quantization. We further regularize rate distortion optimization (RDO) correspondingly and exploit JND computation in AVS Integral DCT domain. Experiments across different QPs show that bitrate can be reduced 13% on average without degradation of visual quality. The rest of the paper is organized as follows. The JND soft thresholding scheme for AVS encoding structure is introduced in Section 2. The details about JND computation in AVS integer DCT domain are in Section 3. Experiment results are given in Section 4. Section 5 draws a conclusion.

II. THE PROPOSED JND GUIDED SOFT THRESHOLDING FOR AVS ENCODING

The proposed scheme is shown in Fig.1, where we introduce a new component called soft thresholding between AVS integer DCT and quantization to suppress HVS imperceivable prediction residues. Thus our scheme can be treated as a content based adaptive quantization since threshold values are changed with different images. The main difference between

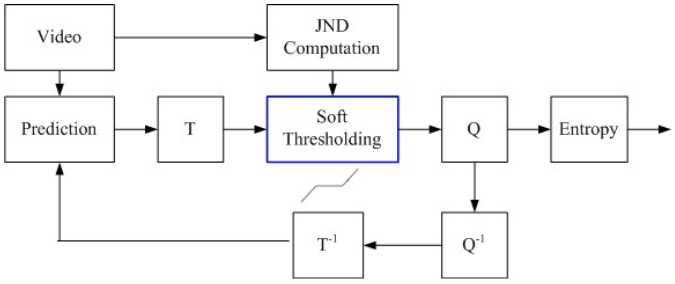


Fig. 1. AVS encoding with JND guided soft thresholding

our scheme and Maks in [3] is that we use soft thresholding instead of hard thresholding. Just like their application in denoising [7], we expect this soft version of thresholding better than hard one as it allows us to remove more redundant residues as shown in a sequel.

Specifically, our JND based soft thresholding function is defined as follows

$$\rho_{Th}(z) = \begin{cases} z - Th & \text{if } z \geq Th \\ z + Th & \text{if } z < -Th \\ 0 & \text{if } |z| \leq Th \end{cases} \quad (1)$$

where Th stands for JND value of each DCT coefficient computed from current frame, and z stands for corresponding coefficient.

There are two issues need to be addressed with this new framework. The first one is how to get precise JND value in context of AVS specific integer DCT domain. Another is how to adapt coding modes selection when this new distortion is introduced in addition to classical quantization error. We solve the first issue in Section 3 and discuss the later in the following section.

Generally Lagrangian based RDO is widely used for real video encoder, in which the final prediction mode chosen for one coding block should minimize the following cost function

$$J = D + \lambda R \tag{2}$$

where λ is Lagrange multiplier usually depending on quantization step size; D is distortion metric between original block and reconstructed block which could be the sum of absolute difference (SAD), the sum of squared difference(SSD) or the sum of absolute transformed difference(STAD); R is the actual cost of bits for encoding certain block with corresponding prediction mode.

As shown in [3], now the D in equation (2) should reflect JND thresholding effect when performing RDO during encoding. Let X be the block to be coded, X' be the corresponding reconstructed block, X_p be the prediction of X, and Y the prediction residual. Thus the new distortion using SAD in our framework becomes

$$D = \|T^{-1}(\rho_{Th}(T(X' - X)))\|_1 \tag{3}$$

$$X' = X_p + T^{-1}(Q^{-1}(Q(\rho_{Th}(Y)))) \tag{4}$$

Another important factor to RDO is Lagrange multiplier λ. Traditionally it is controlled by quantization parameter in both H.264 and AVS. Here λ should adapt to new distortion. There are several options in literatures. The simplest solution is to weight the original λ with an empirically function of JND like [4]. A better solution is to estimate an updated λ by computing equivalent quantization step sizes as shown in [5].

Let $D(Qp)$ be the classical distortion from quantization only, and $D'(Qp, Th)$ be the new distortion from both quantization and JND thresholding. In AVS scheme, λ is mainly the function of Qp. To calculate the updated λ, the new distortion $D'(Qp, Th)$ has to be approximated by $D(Qp\prime)$. The parameter Qp' is the equivalent quantization parameter and can be derived as:

$$Qp' = argmin_{Qp \in S_Q} |D'(Qp_0, Th) - D(Qp)| \tag{5}$$

where Qp_0 is initial quantization parameter and S_Q be the set of quantization parameters in AVS. As some design differences remain between AVS and H.264/AVC, full adaptation of [5] to AVS will be completed in our future work.

III. JND COMPUTATION IN AVS INTEGER DCT DOMAIN

A. State of The Art JND Models

The existing JND models generally belong to two categories. One is pixel-wise JND model, the other is sub-band JND model. Compared to pixel wise JND model, sub-band JND model can further incorporate contrast sensitivity function (CSF) which can describe the sensitivity of human vision for each frequency component. This advantage makes sub-band JND model more attractive.

In context of DCT domain JND for video, Jia et al. [8] proposed an efficient DCT domain JND model over previous computational spatio-temporal CSF models, which introduced more visual effects, such as luminance adaption and contrast masking. Later, Wei et al. [9] enhanced the Jia's model by introducing a new temporal modulation factor to Jia's work and further considering different motion directions. Because of better performance, we use this model in the sequel. Briefly, the Wei's JND model can be expressed as:

$$T_{JND} = T_{Basic} \cdot F_{Lum} \cdot F_{Contrast} \cdot F_T \tag{6}$$

where T_{JND} is the JND threshold in DCT domain; T_{Basic} accounts for spatial contrast sensitivity function (CSF); F_{Lum} and $F_{Contrast}$ model the sensitivity of HVS when luminance and texture of background change respectively; F_T is the sensitivity of HVS to temporal contrast in condition of certain spatial frequency.

It should be noted that AVS supports 4×4 and 8×8 transform block [1], similar to H.264/AVC. Therefore some changes are needed for 4×4 block size related parameters. Specific parameters can be found in [5][9].

B. JND Model in AVS Integer DCT Domain

In AVS, DCT is approximated by integer DCT to reduce computational complexity and mismatch. Even though such mismatch is small, and it may not be a big issue for coarse JND estimation as used by previous works [3][4][6], a more precise estimation method does exist. This method is to derive JND mapping between this approximate DCT domain and classical DCT domain.

Denoting the AVS transform matric as T_N (N stands for the transform block size, 4×4 or 8×8), the forward transform can be formulated as:

$$T_N \cdot X \cdot T_N^T = Z \tag{7}$$

where X is the residual matrix and Z refers to integer transform coefficient matrix. The corresponding classical DCT can be formulated by:

$$DCT_N \cdot X \cdot DCT_N^T = Z' \tag{8}$$

where DCT_N is DCT transform matrix in $N \times N$ size; X is residual matrix and Z' is DCT coefficient matrix. Similar to the relationship between the H.264 Integer DCT and the classical DCT, relationship between the AVS Integer DCT and the classical DCT can be represented as:

$$Z' = Z \otimes S_N \tag{9}$$

where S_N is scale matrix, and $\otimes$ stands for element-by-element multiplication between two matrices.

Further more, according to the design technique of integer cosine transform(ICT)[10], the relationship between DCT_N and T_N is

$$DCT_N = T_N \otimes R_N, \tag{10}$$

where R_N is the normalization matrix, each coefficient of which is equal to the reciprocal of 2-norm of the corresponding row vector in T_N. If we substitute the DCT_N component in equation (8) with equation (10), we can get

$$(T_N \cdot X \cdot T_N^T) \otimes (R_N \otimes R_N^T) = Z'. \tag{11}$$

By comparing equation (9) and (11), we can derive the following relationship

$$S_N = R_N \otimes R_N^T. \tag{12}$$

Finally, with this scale matrix, we can easily obtain the JND relationship between two domains since computation in both (6) and (9) are element-wise

$$Th_{DCT} = Th_{AVS} \otimes S_N. \tag{13}$$

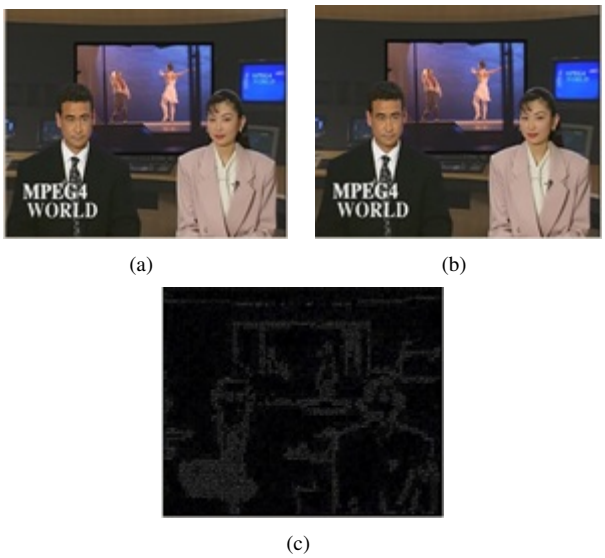
IV. EXPERIMENTS

Our algorithm is implemented on AVS RM09.06 version software [11]. The GOP structure is IBBP with one I frame every 30 frames. The test materials are 4:2:0 YUV sequences with the different resolutions including 4CIF, CIF and QCIF. We deliberately choose different QPs for different sequences and resolution to cover bitrate from high to low.

A. Subjective Video Quality

Video quality of the reference software and the proposed method is compared under the same QP using the Double Stimulus Continuous Quality Scale (DSCQS) method [12]. Experiments were conducted in a room illuminated by fluorescent lights. 15 observers, 10 males and 5 females, were involved in the experiments. The sequences were displayed on a $21'$ displayer (SyncMaster 206BW), and the viewing distance was about four times the image height.

By analyzing the obtained Differential Mean Opinion Score (DMOS), we observe that the average DMOS is quite close to zero and deviation of DMOS is very small too (less than 3/100), which shows the proposed method can produce visually similar video quality to the reference software at the same QP. To save space, here we only give an illustrative example as shown in Fig.2. The figures are from the CIF resolution sequence named news with QP= 22. Fig.2(c) is illustration of JND value of a frame, corresponding to Fig. 2(a) and 2(b), where lighter regions correspond to regions with more complex texture, higher luminance and more drastic motion. It is known that HVS is less sensitive to these regions. Comparing Fig. 2(a) and 2(b), we can find that the perceptual quality is the same.



(a) (b)

(c)

Fig. 2. Illustrative example, (a) RM reconstructed frame; (b) reconstructed frame with our method; (c) corresponding JND threshold map

B. Bitrate Saving and Objective Video Quality

Bitrate saving and objective quality evaluated by PSNR and SSIM [13] are presented in Table 1. According to these results, although there is obvious PSNR performance degradation because of JND process, the loss of SSIM value is very small, and bitrate reduction is significant, with 13.71% on average. These results show that our method can effectively exploit the perceptual redundancy without impairing visual quality.

V. CONCLUSION

In this paper we introduce a JND based soft thresholding technique into state-of-the-art AVS encoder to enhance performance. RDO related issues and AVS specific JND computation are addressed in detail. Experiments validate that the proposed algorithm can save over 13% bitrate without impairing the perceptual quality of reconstructed sequence. Meanwhile the proposed scheme is capable of producing bit stream fully compatible to AVS standard.

ACKNOWLEDGMENT

This work was supported by 863 project (2012AA011703), NSFC (61221001), the 111 Project (B07022) and the Shanghai Key Laboratory of Digital Media Processing and Transmission.

REFERENCES

[1] GB/T20090.2, *"Information technologyładvanced audio video coding standard Part2: Video,"* China, AVS Video Expert Group, 2006.

[2] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, *"Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile,"* IEEE Transactions on Circuits and Systems for Video Technology, Vol.15, no.6, pp. 743-752, 2005.

[3] C. Mak, K. Ngan, *"Enhancing compression rate by just noticeable distortion model for H.264/AVC,"* Proceeding of International Symposium on Circuit and Systems, pp.609-612, Taibei, 2009.

TABLE I
THE PERFORMANCE COMPARISON BETWEEN THE REFERENCE AND THE PROPOSED AVS ENCODER

Sequences (resolution)	Qp	AVS reference encoder			Proposed method			Gain		
		R(kbps)	Y_PSNR(dB)	SSIM	R(kbps)	Y_PSNR(dB)	SSIM	ΔR	$\Delta PSNR$	$\Delta SSIM$
City (4CIF)	34	1742.04	35.52	0.937	1470.36	34.61	0.930	15.60	-0.91	-0.010
	36	1389.91	34.78	0.926	1200.61	33.97	0.916	13.62	-0.81	-0.010
	39	1009.42	33.62	0.906	906.72	32.99	0.896	10.17	-0.63	-0.010
	45	573.41	31.16	0.846	539.76	30.82	0.837	5.87	-0.34	-0.009
Soccer (4CIF)	35	1888.82	36.36	0.919	1634.39	34.84	0.904	13.47	-1.52	-0.015
	38	1418.70	35.09	0.896	1245.58	33.88	0.880	12.20	-1.21	-0.016
	42	968.86	33.33	0.855	865.34	32.48	0.839	10.68	-0.85	-0.016
	48	546.39	30.84	0.778	509.72	30.38	0.768	6.71	-0.46	-0.010
Paris (CIF)	28	975.19	38.74	0.979	701.94	33.18	0.963	28.02	-5.56	-0.016
	36	475.46	34.62	0.956	385.85	31.64	0.941	18.85	-2.98	-0.015
	38	394.00	33.58	0.947	328.68	31.10	0.932	16.58	-2.48	-0.015
	42	273.15	31.59	0.924	236.86	29.90	0.909	13.29	-1.69	-0.015
Foreman (CIF)	28	834.09	39.01	0.962	686.17	36.39	0.951	17.73	-2.62	-0.011
	33	494.35	36.86	0.944	438.34	35.16	0.934	11.33	-1.70	-0.010
	36	367.00	35.61	0.929	332.73	34.26	0.920	9.34	-1.35	-0.009
	40	247.04	33.89	0.905	229.29	32.93	0.896	7.19	-0.96	-0.009
Mobile (CIF)	37	869.66	31.89	0.957	678.90	29.24	0.940	21.94	-2.65	-0.017
	42	500.67	29.52	0.931	411.51	27.80	0.914	17.81	-1.72	-0.017
	45	364.32	28.08	0.909	313.27	26.83	0.892	14.01	-1.25	-0.017
	48	266.52	26.73	0.879	238.01	25.82	0.862	10.70	-0.91	-0.017
Container (QCIF)	22	198.87	43.09	0.980	165.42	36.05	0.966	16.82	-7.04	-0.014
	26	138.43	41.12	0.970	116.03	35.43	0.956	16.18	-5.69	-0.014
	31	82.97	38.47	0.953	71.81	34.47	0.941	13.45	-4.00	-0.012
	33	68.63	37.51	0.946	59.88	34.01	0.935	12.75	-3.50	-0.011
News (QCIF)	27	170.82	41.18	0.984	135.07	34.82	0.972	20.93	-6.36	-0.012
	30	133.40	39.49	0.979	109.46	34.38	0.967	17.95	-5.11	-0.012
	37	76.95	35.89	0.961	66.61	32.89	0.950	13.44	-3.00	-0.011
	39	64.85	34.79	0.953	57.34	32.30	0.942	11.58	-2.49	-0.011
Silent (QCIF)	27	190.05	40.15	0.978	160.71	36.24	0.963	15.44	-3.91	-0.015
	31	135.46	38.07	0.966	119.50	35.32	0.952	11.78	-2.75	-0.014
	37	79.82	34.96	0.935	73.78	33.51	0.924	7.57	-1.45	-0.012
	39	66.66	33.99	0.921	62.77	32.85	0.910	5.84	-1.14	-0.011
Average								13.71%	-2.47	-0.013

[4] Z. Chen and C. Guillemot, *"Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model,"* IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, no. 6, pp. 806-819, 2010.

[5] Z. Luo, L. Song, S. Zheng, N. Ling, *"H.264/Advanced video control perceptual optimization based on JND-directed coefficient suppression,"* IEEE Transactions on Circuits and System for Video Technology, vol.23, no.6, pp.935-948, 2013.

[6] M. Naccari and F. Pereira, *"Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment,"* IEEE Transactions on Circuits and Systems for Video Technology, vol.21, no. 6, pp. 766-782, 2011.

[7] D. Donoho, *"Denoising by soft thresholding,"* IEEE Transaction on Information Theory, vol.41, no.3, pp.613-627, 1995.

[8] Y. Jia, W. Lin, and A. Kassim, *"Estimating just-noticeable distortion for video,"* IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 7, pp.820C829, 2006.

[9] Z. Wei and K. Ngan, *"Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain,"* IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, pp. 337-346, 2009.

[10] C. Zhang, L. Yu, J. Lou, W. Cham and J. Dong, *"The Technique of Prescaled Integer Transform: Concept, Design and Applications,"* IEEE Transactions on Circuits and Systems for Video Technology, vol.18, no.1, pp 84-97, 2008.

[11] http://www.avs.org.cn/

[12] *Methodology for the Subjective Assessment of the Quality of Television Pictures,* Recommendation ITU-R BT.500-12, September 2009.

[13] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *"Image quality assessment: From error visibility to structural similarity,"* IEEE Transactions on Image Processing, vol. 13, no .4, pp. 600-612, 2004.