Improving H.264/AVC Video Coding With Adaptive Coefficient Suppression

Zhengyi Luo, Li Song, and Shibao Zheng
Institute of Image Communication & Information Processing, Shanghai Jiao Tong University
Shanghai, 200240, China

***Abstract*—Video coding has been widely adopted to achieve pleasant video quality at constrained bitrate. In this paper, adaptive frequency coefficient suppression directed by Human Visual System (HVS) is presented for H.264 video coding. Firstly, starting from Just Noticeable Distortion (JND) models for the classic DCT domain, we deduce a JND threshold for the H.264 transform domain with decent adaptation. Then the resultant threshold is used to adaptively suppress the transform coefficients of prediction residuals. It should be noted that our scheme is fully compatible with the H.264 standard. And experimental results show that compared to normal methods, significant bitrate reduction can be obtained by our scheme at similar subjective quality.**

I. INTRODUCTION

With the development of multimedia technologies, versatile video applications have been widely adopted to provide better services. But in light of the huge size of video data and the limited bandwidth or storage, numerous video coding standards have been approved to compress video effectively, of which H.264/AVC [1] is the state-of-the-art one.

Video coding is devoted to produce certain video quality with bits as few as possible. Normally video quality is evaluated by some easy to measure distortion criteria, such as the mean squared error (MSE), which is frequently employed to direct the process of video coding. But considering such objective criteria's inefficiency of quality evaluation in many scenarios, the importance of exploiting human perception has been recognized within the video coding community to compress video further. As far as the H.264 standard is concerned, several coding schemes in considering of Human Visual System (HVS) have been proposed. Cheng et al. [2] introduced a MB-based reduced resolution coding mode to the standardized framework, which reduced the spatial resolution of some MB residuals adaptively while maintaining good subjective video quality. Schuur et al. [3] proposed to remove part of high frequency transform coefficients of prediction residuals alternatively, so that acceptable subjective quality could be obtained with fewer bits. However, they didn't take into account the images' local properties. In the case of Region Of Interest (ROI) coding, Zheng et al. [4] proposed to remove some transform coefficients of prediction residuals in the background, yet bringing little degradation to the attention-drawing ROI regions. Besides the manipulation in the spatial or frequency domain of prediction residuals, recently Yuan et al. [5] proposed to operate in the spatial domain of original images. They first removed part of less perceptible frequency components of original images by pre-filtering at the encoder side. Then at the decoder side images of improved subjective quality could be obtained via final detail completion.

To our knowledge, though human perception has long been utilized in video compression, few specific visual distortion models have been applied for H.264 coding. In this paper, we propose a scheme of H.264 coding using a specific Just Noticeable Distortion (JND) model, which can direct bit allocation better than former schemes. Firstly, in view of the lack of JND models in the H.264 transform domain, we deduce a JND threshold for this domain from existing models for the classic DCT domain with decent adaptation. Then based on the deduced threshold, we perform coefficient suppression of prediction residuals adaptively in the transform domain depending on their impacts to the human perception. It should be noted that our scheme involves no changes to the H.264 standard. And experimental results show that significant bit saving can be achieved if the scheme is implemented in the existing coding framework.

The remainder of this paper is organized as follows. In Section II, a JND threshold in the H.264 transform domain is introduced in detail. Section III describes the adaptive coefficient suppression based on the deduced JND threshold. Experimental results validating the effectiveness of our scheme are shown in Section IV. And Section V draws the conclusion.

II. JND THRESHOLD IN THE H.264 TRANSFORM DOMAIN

JND, which refers to the maximum distortion people cannot perceive typically, can serve as an important guide in perceptual image processing. Owing to the popularity of the DCT transform, significant efforts have been spent on JND models in the DCT domain.

This work was partially supported by National Natural Science Foundation of China (60702044, 60625103 and 60632040).

978-1-4244-5309-2/10/$26.00 ©2010 IEEE

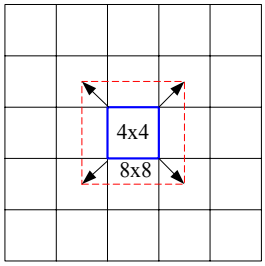


Figure 1. Illustration of block extension.

Unlike the popular 8x8 DCT transform used in previous standards, H.264 introduces a low-complexity 4x4 transform, which can be computed exactly in integer arithmetic so as to avoid transform mismatch problems. Usually building JND models in a new domain is not an easy task. But considering the new transform's origination from the classic 4x4 DCT, we suggest to deduce its JND threshold in two steps. Firstly, a JND model for the classic 4x4 DCT is built through adaptation from existing models of 8x8 DCT. Then based on the resultant model, the JND threshold for the H.264 transform domain is approximated by means of linear transform.

A. JND Adaptation for the 4x4 DCT

As numerous standards employ the 8x8 DCT transform, there exist several excellent JND models in the classic 8x8 DCT domain. Here we adopt the spatial model in the recent literature [6] and adapt it decently for the 4x4 DCT scenario.

The adopted JND is expressed as the product of a base threshold and a modulation factor. Let n denote the index of a block, and i and j denote the DCT coefficients' indices. Then the JND is formulated as

$$\begin{aligned} T_{JND}(n,i,j) &= T_{Basic}(n,i,j) \times F_M(n,i,j) \\ F_M(n,i,j) &= F_{lum}(n) \times F_{contrast}(n,i,j) \end{aligned}, \quad (1)$$

where T_{JND} is the JND threshold, and T_{Basic} is the base threshold. The product of the luminance adaptation factor F_{lum} and the contrast masking factor $F_{contrast}$ constitutes the modulation factor $F_M(n,i,j)$. T_{Basic} accounts for the spatial contrast sensitivity function (CSF) and can be expressed as

$$T_{Basic}(n,i,j) = s \cdot \frac{1}{\phi_i \phi_j} \cdot \frac{\exp\left(c\omega_{ij}\right) / \left(a + b\omega_{ij}\right)}{r + (1-r)\cdot \cos^2 \varphi_{ij}}. \quad (2)$$

Here we calculate it in the 4x4 DCT scenario $(0 \le i, j \le 3)$, and related parameters in (2) scale correspondingly from [6]. F_{lum} accounts for the impact of brightness to human perception. Here we first compute the average intensity value of every 4x4 block. Then F_{lum} is calculated the same as the original model. Usually distortion is easily observed in the smooth or edge areas but not in those with high texture energy. And such masking effect is taken into account in the model via $F_{constrast}$. Firstly, the *Canny* operator is applied to detect the edge pixels for a given image. Then based on the percentage of edge pixels within, 8x8 blocks are classified into types of Plane, Edge or Texture, which are used to calculate $F_{contrast}$ in the next. For the case of 4x4 blocks, to cover the shortage of edge information we propose to examine the percentage of edge pixels in the extended 8x8 blocks as shown in Fig. 1, so that their block types can be determined more accurately. And the remaining operations are similar to the original model. Now with all components available, finally the JND threshold in the classic 4x4 DCT domain can be obtained by (1).

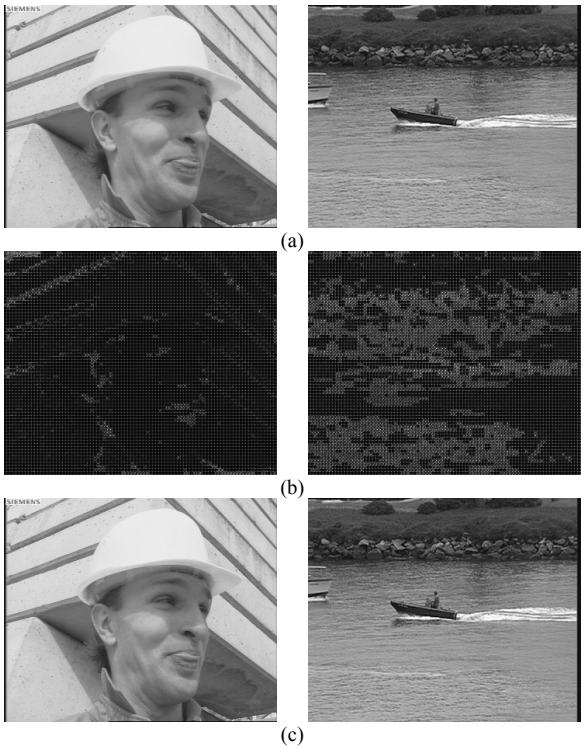


Figure 2. Illustration of the deduced JND: (a) the original frames, (b) JND thresholds in the H.264 transform domain (enhanced for visibility), (c) contaminated frames with random plusminus JND added to each subband.

B. JND Translation to the H.264 Transform Domain

Assume for a 4x4 block the JND threshold in the classic 4x4 DCT domain is $J(4\times4)$. Usually the JND threshold constructs a simply connected domain in the distortion space, of which J itself is a representative boundary point. As the preferred transform matrix H originates from the 4x4 DCT and is near-orthogonal [7], we propose to approximate the JND threshold in the H.264 transform domain via linear transform. Let C denote the 4x4 DCT transform matrix and X denote the point in the image domain corresponding to the JND threshold. Then we have

$$J = C \cdot X \cdot C^T. \quad (3)$$

Let $J^*(4\times4)$ denote the JND threshold in the H.264 transform domain. Assume the preferred transform matrix is applied during encoding, then J^* can be approximated by

$$J^* \approx H \cdot X \cdot H^T = H \cdot C^{-1} \cdot J \cdot \left(C^T\right)^{-1} \cdot H^T. \quad (4)$$

Fig. 2 illustrates the effect of the deduced JND threshold in the H.264 transform domain. It can be seen that the threshold correlates well with the perceptual distortion.

III. ADAPTIVE COEFFICIENT SUPPRESSION IN THE H.264 TRANSFORM DOMAIN

Normally for the nth 4x4 block in an image, if $w_{n,i,j}$ and $l_{n,i,j}$ denote the transform coefficient of residuals in the (i,j)th subband of the H.264 transform domain before and after quantization respectively, the preferred quantization process in this subband can be implemented as [8]

$$\begin{aligned} \left|l_{n,i,j}\right| &= \left(\left|w_{n,i,j}\right| \cdot MF_{i,j} + f\right) \gg qbits \\ \operatorname{sign}\left(l_{n,i,j}\right) &= \operatorname{sign}\left(w_{n,i,j}\right) \end{aligned}, \tag{5}$$

where $\gg$ indicates a binary shift, f is the offset, and $MF_{i,j}$ and $qbits$ respectively are the predefined multiplication factor and the calculated times of shift which both depend on the selected quantization parameter.

Usually the bitrate of H.264 coding highly depends on the number and the absolute value of nonzero coefficients after quantization. For a nonzero coefficient $l_{n,i,j}$, if we suppress it by $k_{n,i,j}$, obviously the introduced error will be

$$e_{n,i,j} = \left|w_{n,i,j}\right| - \left[\left(\left|l_{n,i,j}\right| - k_{n,i,j}\right) \ll qbits\right] \Big/ MF_{i,j} \ . \tag{6}$$

From [9] we find that the perceptual distortion in the (i,j)th subband can be modeled by a function of the JND-normalized error

$$\tau = \frac{e_{n,i,j}}{J^*_{n,i,j}} , \tag{7}$$

where $J^*_{n,i,j}$ is the JND threshold in this subband of the nth block. So from the point of perceptual distortion, we can designate a JND-normalized error threshold T for images. Then we can try to suppress nonzero coefficients, whose resultant perceptual distortion won't exceed the threshold. Specifically, for a nonzero coefficient $l_{n,i,j}$ in the (i,j)th subband of the nth 4x4 block[1], the problem can be formulated as

$$\begin{aligned} &\max\ k_{n,i,j} \ (0 \le i,j \le 3) \\ &s.t.\ 0 \le k_{n,i,j} \le \left|l_{n,i,j}\right| \\ &\tau \le T \end{aligned} \ . \tag{8}$$

In this way, we can flexibly regulate resultant bits based on HVS, and via adaptive coefficient suppression similar quality can be maintained with reduced bits.

[1] As DC coefficients in the Intra16x16 mode are processed separately, they are excluded from suppression in this paper.

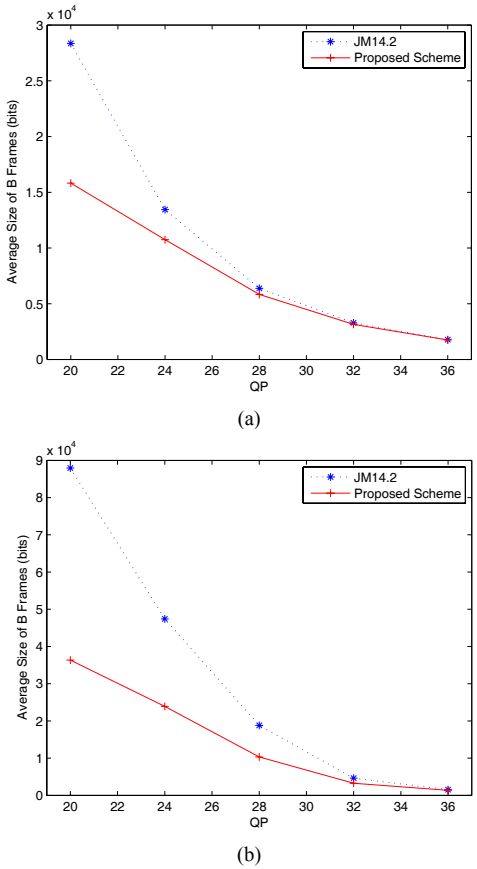


Figure 3. Average size of B frames for (a) "Foreman", (b) "Coastguard".

IV. EXPERIMENTAL RESULTS

Now after being implemented in the JM14.2 software [10], the proposed scheme is extensively evaluated for many standard sequences. For space limits, here we only show two representative results in which the first 100 frames of CIF sequences "Foreman" and "Coastguard" at 30fps are encoded. Group of Pictures (GOP) of IBBPBBP… structure with one I frame inserted every 30 frames are considered. Two reference frames and CABAC are used during encoding.

To inhibit error propagation, the proposed scheme is applied only for the luminance component of B frames in the experiments. Rate control over fixed quantization parameters (QPs) is enabled to make I and P frames exactly the same for fair comparison. And the JND-normalized threshold is set to 10 empirically here.

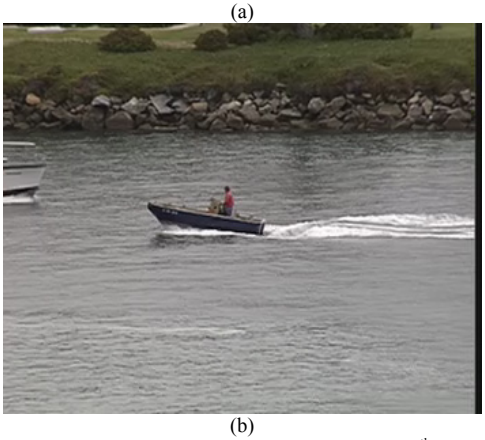
Fig. 3 shows the average size of B frames in dependence of QP for normal processing and the proposed scheme. It can be seen that as much as about 50% of bits for B frames can be saved by our scheme. When QP is low, many nonzero coefficients are available after quantization, leaving us great margin of bit saving via coefficient suppression. On the contrary, when QP is high, there exist fewer nonzero quantized coefficients for suppression, which explains the performance degradation in this case. As far as "Foreman" is concerned, being able to be better predicted during encoding, it has fewer prediction residuals than "Coastguard".

(a)

(b)

Figure 4. Comparison of subjective quality for the 8th frame of "Foreman" when QP is 20: (a) JM 14.2, 33.6kbits, 41.77dB, SSIM=0.9763; (b) the proposed scheme, 17.8kbits, 40.18dB, SSIM=0.9720.

(a)



(b)

Figure 5. Comparison of subjective quality for the 5th frame of "Coastguard" when QP is 24: (a) JM 14.2, 48.3kbits, 37.28dB, SSIM=0.9599; (b) the proposed scheme, 22.8kbits, 34.67dB, SSIM=0.9388.

Furthermore, less JND distortion can be endured for "Foreman" due to its fewer textures. So relatively "Coastguard" enjoys much more bit reduction in the experiments.

As our scheme prefers small QP, in this case the comparison of subjective quality with and without our scheme is illustrated in Fig. 4 and Fig. 5. It can be observed that though a wild gap of objective quality between normal processing and the proposed scheme exists, the subjective quality is still comparable as we regulate bits based on human perception. So our scheme is quite effective for bit saving while maintaining similar visual quality.

V. CONCLUSION

In this paper, a scheme of H.264 coding using adaptive coefficient suppression is proposed. We first deduce a JND threshold in the H.264 transform domain from existing models in the classic DCT domain. Then based on the threshold, transform coefficients of prediction residuals are suppressed adaptively according to their impacts to the perceptual distortion. With our scheme resultant bits can be regulated flexibly based on HVS. And experimental results show that significant bit saving can be achieved at similar subjective quality.

REFERENCES

[1] ISO/IEC, Advanced Video Coding for Generic Audiovisual Services, 14496-10, Mar. 2005.

[2] H. Cheng, A. Kopansky, and M.A. Isnardi, "Reduced resolution residual coding for H.264-based compression system," Circuits and Systems, IEEE International Symposium on, 2006.

[3] B. Schuur, T. Wedi, S. Wittmann, and T. Palfner, "Frequency selective update for video coding," Image Processing, IEEE International Conference on, 2006.

[4] Y.Y. Zheng, X. Tian, and Y.W. Chen, "Adaptive frequency coefficient suppression for ROI-Based H.264/AVC video coding," Networking, Sensing and Control, IEEE International Conference on, 2008.

[5] Z. Yuan, H.K. Xiong, L. Song, and Y. F. Zheng, "Generic video coding with abstraction and detail completion," Acoustics, Speech and Signal Processing, IEEE International Conference on, 2009.

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

[7] H.S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, pp. 598-603, 2003.

[8] I. Richardson, "Transform & quantization," H.264/MPEG-4 Part 10 White Paper, http://www.vcodex.com, 2003.

[9] I. Hontsch and L.J. Karam, "Adaptive image coding with perceptual distortion control," Image Processing, IEEE Transactions on, vol. 11, pp. 213-222, 2002.

[10] "JM 14.2 Reference Software," Available: http://iphome.hhi.de/suehring/tml/download.