An HEVC-Compliant Perceptual Video Coding Scheme based on Just Noticeable Difference Models

Jaeil Kim and Munchurl Kim
Korea Advanced Institute of Science and Technology
Daejeon, Korea
Jaeil1203@kaist.ac.kr, mkim@ee.kaist.ac.kr

Abstract—In this paper, we first introduce a spatio-temporal Just Noticeable Difference (JND) model based perceptual video coding (PVC) scheme which is completely compliant with HEVC. We incorporate into HEVC quantization process a JND model that takes into account both the temporal JND characteristics of temporal masking effects and the spatial JND characteristics of the luminance adaptation and contrast masking effects as well as contrast sensitivity. To be HEVC-compliant, the transform coefficients are suppressed based on the spatio-temporal JND model before quantization during HEVC encoding process. The spatio-temporal JND model works in transform domains in the cases of non-transform skip modes, but operates in pixel domain for the transform skip cases in HEVC. To make it effective the JND-based suppression, a distortion compensation factor is also proposed to reflect the perceptual distortion into the rate-distortion optimization (RDO) based encoding process. The proposed spatio-temporal JND-based PVC scheme in HEVC yields remarkable bitrate reductions of maximum 53.37% and average 10.15% with negligible PSNR/SSIM quality loss (average PSNR drop of 1.46dB and average SSIM drop 1.46%), compared to the original HEVC reference software (HM 11.0).

I. INTRODUCTION

Recently, the standardization of a next-generation video coding, called High Efficiency Video Coding (HEVC) [1], has almost been finalized with a goal of improving its coding efficiency more than twice that of H.264/AVC [2]. In order to improve coding performance, advanced techniques are newly adopted for HEVC, which include a flexible structure of Coding Unit (CU) with skip modes, symmetric and asymmetric prediction units (PUs) with advanced motion vector prediction and merge modes, a residual quad-tree transform structure with 4×4, 8×8, 16×16 and 32×32, 33 directional intra prediction with planar and DC modes, and additional in-loop filtering of sample adaptive offset [1]. Most of the video coding standards have been developed towards reducing rate-distortion (RD) costs. However, such efforts of improving the coding efficiency have led to dramatic increases in encoder complexity, and further improvement of coding efficiency becomes more and more difficult in a rate-distortion (MSE; Mean Squared Error) optimization (RDO) sense even with more elaborate coding tools. In spite of much effort on coding efficiency improvements, the reduction in signal distortion does not always coincide with perceptual quality enhancement.

Another axis toward coding efficiency improvement is perceptual video coding (PVC) where perceptual distortion can be incorporated instead of signal distortion such as MSE. Therefore, much effort has been made into this direction [3-6]. In PVC, how to effectively incorporate visual perception models into coding process is one of the most important issues in reducing perceptual redundancy. Among the techniques to reduce the perceptual redundancy, JND is known as an effective model to represent perceptual redundancy of human visual systems (HVS) [3-6]. Recently, several methods have focused on PVC [3-6]. In [3], a JND-adaptive residue preprocessor was proposed to suppress the perceptual redundancy of motion-compensated residue signals in pixel domains with a nonlinear additive model. Chen et al. [4] introduced a foveation JND model to improve the previous spatial and temporal JND models. The proposed JND model in [4] is used for macroblock quantization adjustment in H.264/AVC. Naccari et al. [5] showed a JND model that is incorporated into both encoder and decoder sides of H.264/AVC to avoid additionally required signal bits for the adjustment of quantization step sizes. However, the methods [4, 5] are not H.264/AVC-compliant. Recently, Lue et al. [6] developed a JND based H.264/AVC-compliant PVC by tuning the levels of quantized transformation coefficients.

In this paper, we first propose an HEVC-compliant PVC scheme with a more elaborate new JND model that considers the contrast sensitivity function (CSF) characteristics as well as the luminance adaptation (LA), contrast masking (CM) and temporal masking (TM) effects, which is applicable for variable block-sized Transform Units (TUs) of HEVC. Also, a distortion compensation factor is proposed to reflect the perceptual distortion by the JND models instead of the RDO based encoding. This paper is organized as follows: In Section 2, a JND model based on our previous works is introduced which is integrated into HEVC; Section 3 discusses how to adjust the transformed coefficients to apply the JND model into the quantization process of HEVC and how to derive the distortion compensation factor for rate-perceptual distortion optimization; the experimental results are shown in Section 4; and Section 5 concludes this paper.

II. JND MODELS AND A JND BASED HEVC ENCODER ARCHITECTURE

A. Overall Architecture of A JND Based HEVC Encoder

In order to reduce perceptual redundancy by utilizing the JND characteristics of human visual systems (HVS), we incorporate a simple JND model in pixel domain and an elaborated JND model in transform domain into the transform
and the quantization process of HEVC. Fig. 1 shows the overall architecture of our proposed JND-based HEVC-compliant PVC encoder.
texture type, DCT is performed to calculate the $MF_{CM}$ values. In general, this approach has two shortcomings: (i) the $MF_{CM}$ is too roughly modeled in texture complexity perspective where only three types of texture complexity are reflected into the $MF_{CM}$ modeling; (ii) it requires large computational complexity in running the Canny edge operator for image block classification and the DCT process for $MF_{CM}$ calculation, which must be performed repetitively for all TU block sizes when it is applied for RDO based encoding process of HEVC.

In this paper, we propose an elaborate $MF_{CM}$ model as a continuous function of texture complexity. The proposed $MF_{CM}$ model does not require image block classification with DCT computation of image blocks, and uses a simple Sobel edge operator for texture complexity measure. Edge pixel density is employed as a texture complexity metric for image blocks to model $MF_{CM}$. The edge pixel density is calculated as

$$\tau = \left(\frac{1}{N^2}\right) \sum_{i} \sum_{j} \text{edge}(i, j)$$

where $N$ is an TU block size, $\text{edge}(i, j)$ is an edge flag, which is 0 for non-edge and 1 for edge, at $(x, y)$ position, and is calculated using a 3x3 Sobel operator. In order to determine edge flags, threshold value is set to 12. The edge pixel density in (6) has the range from 0 to 1.

In order to model the JND for the CM effect as $MF_{CM}$ in transform domain with respect to the edge pixel density ($\tau$) and the spatial frequency ($\omega$) in cycles per degree, we performed psychophysical experiments according to the same procedure in [8]. Fig. 2 shows the measured $MF_{CM}$ and estimated $MF_{CM}$ values versus spatial frequency ($\omega$) for different texture complexity values. Based on the measured $MF_{CM}$ values, the $MF_{CM}$ is modeled as

$$MF_{CM}(\omega, \tau) = \begin{cases} 
(1.56\tau - 0.15)\omega + e^{2.2\tau} & \omega \leq 4.17 \\
(-0.5\tau)\omega + 5e^{4\tau} - 4.6 & \omega > 4.17 
\end{cases}$$

Fig. 2 shows the $MF_{CM}$ model in (7) where the Pearson Correlation Coefficient (PCC) and Root Mean Square Error (RMSE) values between the measure $MF_{CM}$ values and the estimated $MF_{CM}$ values are 93.18% and 0.46 respectively.
\[
JND_{(i,j)} = \begin{cases} 
JND_{(i,j)} \cdot \left( M^{(i,j)} + f \right) & \text{if } qbits \geq JND_{(i,j)} \\
0 & \text{otherwise}
\end{cases}
\] (10)

where \( JND_{(i,j)} \) denotes the transformed coefficient level of the residue suppressed by the JND models in the \((i,j)\)-th subband of the \(n\)-th TU block after quantization process, and \( f \) is a rounding offset. The JND based suppression in (10) is performed at one go unlike the method [6] where the JND based best suppression is found iteratively, which may not be desirable in hardware implementation.

C. Distortion model for JND-based HEVC

In order to effectively apply the JND models for determining the best CU/PU/TU modes during the HEVC encoding, the RDO based mode decision of the HM must be adjusted to take into account the JND-based suppression. In the previous works [5-6], the Lagrange multiplier, \( \lambda \), is updated by reflecting the JND-based quantization. However, since HEVC supports the SKIP modes in various CU levels for which no residual coding are performed, the SKIP modes are preferably to be selected in RDO sense. Owing to this, the JND based suppression for the non-TSM cases are not likely to be performed. In order to remedy this, we propose a distortion compensation factor that is used to reflect the perceptual distortion based on the JND models so that the JND based suppression for the non-TSM cases can fairly be performed. For this, the distortion \( D_{\text{JND-HM}}(q_{\text{HM}}, JND, \lambda) \) produced by the proposed JND-based suppression is normalized by the distortion \( D_{\text{HM}}(q_{\text{HM}}, \lambda) \) based on the RDO optimized coding in the HM 11.0 as

\[
r = \frac{D_{\text{JND-HM}}(q_{\text{HM}}, JND, \lambda)}{D_{\text{HM}}(q_{\text{HM}}, \lambda)}
\] (11)

where \( r \) is a distortion compensation factor, which is then used to compensate the difference between \( D_{\text{JND-HM}} \) and \( D_{\text{HM}} \) during the HEVC encoding process.

IV. EXPERIMENT RESULTS

In order to verify the effectiveness of the proposed JND-based PVC scheme in HEVC, the JND models in (1) and (5) were implemented into the HM 11.0 and compared with the original HM 11.0 [10]. Also, the famous quality assessment metric, SSIM, were performed to compare the encoded bitstream by the original HM 11.0 and the proposed JND-based PVC scheme encoded 13 test sequences with various spatial and temporal characteristics; one test sequences of 2560×1600, *PeopleOnStreet* (PS), three test sequences of 1920×1080p, *Kimono* (K), *ParkScene* (PK) and *ElFuenite* (EI), two test sequences of 832×480, *BQMall* (BQ) and *PartyScene* (PS), two test sequences of 416×240, *RaceHorses* (RH) and *BasketballPass* (BP), two test sequences of 1280×720, *MobileCalendar* (MC) and *City* (Ci), three test sequences of 352×288, *Akiyo* (Ak), *HallMonitor* (HM) and *Silent* (Si). Note that the Random Access configuration is used for all experiments.

Table II shows the overall performance of the proposed JND-based PVC in HM 11.0 in terms of bitrate reduction (\% ) and PSNR drop (dB) and SSIM drop (%). The bitrate reduction (\( \Delta B \)) is calculated as

\[
\Delta B = \frac{\text{Bitrate}_{\text{proposed}} - \text{Bitrate}_{\text{HM}}}{\text{Bitrate}_{\text{HM}}} \times 100
\] (11)

The SSIM drop for luminance (\( \Delta Y-S \)) is given by

\[
\Delta Y - S = \frac{\text{SSIM}_{\text{proposed}} - \text{SSIM}_{\text{HM}}}{\text{SSIM}_{\text{HM}}} \times 100
\] (12)

The time increment between HM and proposed scheme is given by

\[
\text{Time} = \left( \text{time}_{\text{proposed}} - \text{time}_{\text{HM}} \right) \times 100
\] (13)

As shown in Table II, the average bitrate reduction is 8.91% and the maximum bitrate reduction amounts to 52.72% at \( QP = 22 \) even with PSNR and SSIM drop of 3.46 dB for the City sequence.
compliant, it can be effectively combined to SSIM drop 1.46% with compliant compensation encoding of HEVC by incorporating the distortion distortion is also appropriately combined in TSM and proposed PVC scheme.

Fig. 4 shows the comparison of obtained at lower QP values due to larger amounts of residues. Also in Table II, the JND suppression works more significantly obtained at lower QP values due to larger amounts of residues. Fig. 4 shows the comparison of two magnified image regions by the HM 11.0 and the proposed method.

V. CONCLUSION AND FUTURE WORK

We proposed a JND-based HEVC-compliant PVC scheme which utilizes the perception characteristics of HVS. The proposed PVC scheme effectively incorporates the spatio-temporal JND model of CSF, LA, CS and TM effects for non-TSM and the spatial JND model only with the LA effect for TSM into the quantization process of HEVC and the perceptual distortion is also appropriately combined into RDO based encoding of HEVC by incorporating the distortion compensation factor. The proposed JND-based HEVC-compliant PVC scheme yielded remarkable bitrate reductions with maximum 53.37% and average 10.15% with PSNR and SSIM quality loss (average PSNR drop of 1.46dB and average SSIM drop 1.46%). Since the proposed PVC scheme is HEVC-compliant, it can be effectively combined to design very effective HEVC encoders.

REFERENCES


