Multimed Tools Appl (2018) 77:12777–12803
DOI 10.1007/s11042-017-4914-4

A fast and HEVC-compatible perceptual video coding scheme using a transform-domain Multi-Channel JND model

Gang Wang[1] · **Yongfei Zhang**[1,2] · **Bo Li**[1,2] · **Rui Fan**[1] · **Mingliang Zhou**[1]

Received: 20 November 2016 / Revised: 15 April 2017 / Accepted: 5 June 2017 /
Published online: 4 July 2017

Abstract Coding optimization methods incorporating the just noticeable distortion (JND) model, called perceptual video coding (PVC), have drawn much attention in recent years for better video coding performance. To further remove perceptual redundancy in every channel and improve the coding performance, this paper proposes a fast PVC scheme in the latest High Efficiency Video Coding (HEVC) framework based on our proposed variable block-size transform-domain multi-channel JND model. Firstly, through extensive experiments, we find out for the first time that the contrast masking (CM) effects for chroma channels show a lowpass property in frequency, which differs from the luma channel that has a bypass property. Based on this observation, CM effects in chroma blue (Cb) and chroma red (Cr) channels are modeled as a continuous function for variable-sized blocks, respectively. Secondly, since different characteristics of the human visual system (HVS) exhibit quite distinct effects in luma and chroma channels and effects in chroma channels were not well explored, we develop a new JND model through comprehensive consideration for both luma and chroma channels of five typical HVS effects, with especial focus on parameterized modeling of chroma channels in each effect. Finally, to incorporate the proposed JND model into the latest HEVC coding framework, a multi-channel coefficients suppression method based on JND thresholds and quantization parameter (QP) ranges is proposed in the transform and quantization process, which can decrease the computational complexity. Extensive experimental results show that the proposed PVC scheme implemented in HEVC reference software (HM15.0) can yields significant bit saving of up to 25.91% and on average 9.42% with similar subjective quality, compared to HM15.0, and consistently outperforms two PVC schemes with much reduced bitrate and complexity overhead.

✉ Yongfei Zhang
yfzhang@buaa.edu.cn

[1] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing, China

[2] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

Keywords Just noticeable distortion (JND) · Perceptual video coding (PVC) · High efficiency video coding (HEVC)

1 Introduction

The high efficiency video coding (HEVC) standard has been standardized by the Joint Collaborative Team on Video Coding (JCT-VC) [29–31, 36]. On the one hand, HEVC standard provides higher compression efficiency in comparison to the video coding standard H.264/AVC [16, 23, 32]. On the other hand, HEVC standard aims to be applied in the high-quality video coding. As we all know, those video coding standards (e.g., H.264/AVC, HEVC etc.) popularly adopt objective metrics to assess the encoder performance. More specifically, the Sum of Squared Error (SSE) is used as a metric for computing the rate-distortion (RD) costs. However, these objective metrics are not perfectly consistent in visual quality with the perception characteristics of the human visual system (HVS) [4, 24, 26, 35]. Meanwhile, it is hard to further remove statistical correlation redundancies of video signal, such as spatial, temporal, and symbol redundancy, and improve the video coding efficiency in conventional video coding frameworks, which has almost kept unchanged for decades. In HEVC, the quantization matrix is designed by the contrast sensitivity function (CSF) -based quantization which is shown in Fig. 1. The low frequency coefficients use a finer quantization step size, while the higher frequency coefficients apply a larger quantization step size. However, since the quantization of HEVC only coarsely considers the CSF factor in some larger transform blocks, many characteristics of the HVS have not been sufficiently considered to further enhance the encoding performance, with either reduced bitrate or improved visual quality.

Recently, perceptual video coding has been paid increasing attention to improve coding efficiency by combining with the perception characteristics of the HVS. As far as computational models for perceptual thresholds are concerned, PVC schemes can be coarsely classified into three categories: ROI-based methods [8, 28], visual attention guided methods [10, 11, 15, 17, 37] and JND-based methods [6, 14, 18, 20]. The ROI-based methods apply machine vision algorithms to automatically detect regions-of-interest (ROI). Visual attention guided methods exploit computational neuroscience models to predict the human attention regions. The JND models use

Fig. 1 Default quantization matrix for transform

16	16	16	16	17	18	20	24
16	16	16	17	18	20	24	25
16	16	17	18	20	24	25	28
16	17	18	20	24	25	28	33
17	18	20	24	25	28	33	41
18	20	24	25	28	33	41	54
20	24	25	28	33	41	54	71
24	25	28	33	41	54	71	91

knowledge about human psychophysics to compute the perceptual threshold. The JND-based PVC scheme has shown better video coding performance in video compression application.

Among JND-based methods for PVC schemes, many previous works focused on improving JND models. Yang et al. [33, 34] developed a new spatial pixel-domain JND model with the nonlinear additivity model for masking (NAMM) for color image/video in the YCbCr space. Yang's method takes count for the overlapping effect of luminance masking and texture masking at three color channels. Besides, Yang's model integrates temporal masking (TM) into spatio-temporal JND model for the color video. However, the pixel-domain JND model is less accurate compared with the transform-domain JND model for JND thresholds. Wei et al. [25] proposed a spatio-temporal JND model for gray image/video in discrete cosine transform (DCT) domain. In Wei's model, a new luminance adaptation JND (LA-JND) model is introduced by considering the gamma correction. And, a novel fixed block classification is also proposed. Luo et al. [18] proposed a JND-based application which can adjust the quantization by the JND threshold. Due to previous works in [18, 25] applied only to fixed size blocks, Kim et al. [14] proposed an HEVC-compliant PVC scheme which can support variable block-sized transform units. They applied the JND profile for monochrome images in [1] to suppress transform coefficients. However, this method can only suppress luminance component for adopting the single-channel JND model.

For JND-based PVC schemes, many approaches have been developed for different encoding processes in previous studies. In [33, 34], PVC methods can be adopted to suppress residues in pixel-domain before transform. Kim et al. [14] proposed a PVC method for variable block-sized transform kernels after transform and before quantization. Chen et al. proposed a foveation JND-based method that suppress transform coefficients by MB-level QP adjustment [5] in the quantization process. Luo et al. [18] introduced a JND-based PVC scheme by tuning the quantization using a JND-normalized error model after quantization.

In summary, conventional PVC schemes have at least the following limitations:

1) Conventional transform-domain JND-suppression PVC schemes directly apply luma JND thresholds into chroma channels [18, 20, 27], which will result in underestimating or overestimating JND thresholds in Cb and Cr channels. The HVS has different characteristics in luma and chroma channels.
2) Previous traditional PVC schemes incorporate a fix-sized JND model (e.g., 4×4 [6, 18] or 8×8 [25]) to remove perceptual redundancy which is not suitable for HEVC with various block sizes form 4×4 to 32×32. The fix-sized JND model will also bring the computational complexity for the block classification.

In order to fix the above problems of previous works, a new multi-channel JND model is proposed in this paper. In HEVC, a coding tree unit (CTU) contains a luma coding tree block (CTB) and two corresponding chroma coding tree block in HEVC. Then, each coding block (CB) is partitioned into transform blocks (TBs) by the residual quadtree (RQT) structure [9]. In practice, since perceptual redundancy exists not only in the luma channel but also in the chroma channel. Therefore, the model is designed by not only considering the luminance channel perceptual redundancy, but also introducing the chroma channel perceptual redundancy. The JND threshold in each channel has five effects based on the perceptual characteristic of the HVS. Moreover, this paper also proposes a fast perceptual video coding scheme which applies the proposed JND model to further enhance the coding efficiency for HEVC. Our PVC scheme can calculate JND thresholds for every channel by the JND model to suppress

transform coefficients after transform and before quantization process in HEVC. The main contributions of this work are summarized as follows:

- To simultaneously effectively remove perceptual redundancy in luma and chroma channel for YCbCr video format in HEVC, a transform-domain multi-channel JND model in the YCbCr space is proposed, which is based on different characteristics of the HVS in luma and chroma channels. In contrast, JND thresholds of previous conventional PVC schemes calculated in the luma channel guide to remove the perceptual redundancy in the luma or in the chroma channel, thus leading to ineffectively remove chroma redundancy [18, 20].
- We find out for the first time that CM effects in two chroma channels show a lowpass property in frequency, which differs from the luma channel that has a bypass property in frequency**.** Based on this observation, we model a continuous function for CM effects in Cb and Cr channels, respectively. By this method, the CM effect can support the variable block-sized transform units of HEVC, which is in harmony with the luma channel computation [1].
- A multi-channel summation effect function of $S^{\Omega}(L)$ is proposed by deriving the probability summation model. Since the summation effect of the luma channel has lower JND thresholds than chroma channel, two chroma summation effects for Cb and Cr channels are obtained by designed psychophysical experiments to maintain the same distortion for the 4×4 to 16×16 transform block size. In previous color JND models [5, 21], a fixed value for the summation effects is only used for single-size transform units.
- To achieve fast video coding, a multi-channel coefficients suppression method based on JND thresholds and QP ranges is proposed in the transform and quantization process, which can decrease the computational complexity. In previous PVC schemes [14, 25], transform coefficients are only considered the relationship between the luma component and JND thresholds. Moreover, we propose a frame-level texture complexity for TBs to decrease the computational complexity.

This paper is organized as follows: In Section 2, we briefly address the system overview of our proposed fast multi-channel JND-based PVC scheme for HEVC. The Section 3 elaborates the proposed transform-domain multi-channel JND model in the YCbCr space, and describes the computation of summation effects for various block-sized transforms kernels in luminance and chroma channels. In Section 4, we propose a fast and HEVC-compatible perceptual video coding scheme which applies the proposed JND model to further enhance the coding efficiency. And the experiment results of our proposed JND model and PVC scheme are given in Section 5. Finally, the conclusion is drawn in Section 6.

2 System overview of the proposed multi-channel JND-based PVC scheme

Figure 2 shows the framework of proposed multi-channel JND-based PVC scheme. As shown in the top-shaded box of Fig. 2, our multi-channel JND model can independently compute JND thresholds for input video frames in the luma channel, the Cb channel and the Cr channel. In the transform and quantization process, JND profiles will determine whether residual transform coefficients will be suppressed further in every channel. The proposed transform-domain multi-channel JND model in YCbCr space will be elaborated in Section 3. Besides, our proposed PVC scheme is faster than the state-of-the-art PVC scheme by adopting the

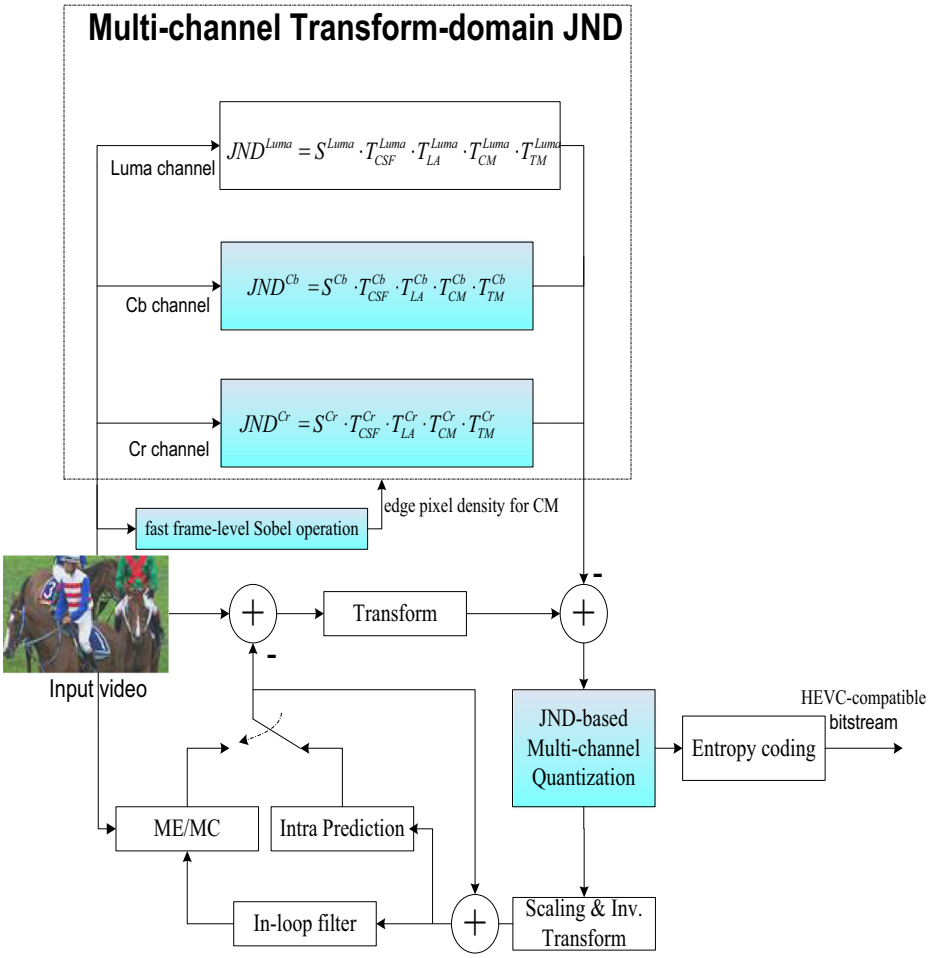


Fig. 2 The framework of the proposed multi-channel JND-based

frame-level Sobel operation and JND-based multi-channel quantization approaches in blue rectangles of Fig. 2. It is also noted in Fig. 2 that, since our PVC scheme suppresses residual transform coefficients without adjusting coding parameters of HEVC, the output result is the standard HEVC-compatible bitstreams. Please refer to Section 4 for our developed fast and HEVC-compatible PVC scheme based on the proposed JND model in Section 3.

3 The proposed transform-domain multi-channel JND model in YCbCr space

Since different characteristics of the HVS exhibit quite distinct effects in luma and chroma channels and effects in chroma channels were not well explored, in this section, we developed a new JND model through comprehensive consideration of both luma and chroma channels of five typical effects, with especial focus on parameterized modeling of each effect in chroma channels.

The proposed transform-domain multi-channel JND model in the YCbCr space is formulated as product form of five factors based on the perceptual characteristic of the HVS by introducing these factors of the typical luma JND model into the chroma JND model: the CSF function, LA effect, CM effect, TM effect and the summation effect, which is expressed by

$$JND^{\Omega}_{spatio-temporal}\left(n,\omega_{i,j},\varphi_{i,j},\mu_p,\tau\right) = S^{\Omega}(L)\cdot T^{\Omega}_{CSF}\left(\omega_{i,j},\varphi_{i,j}\right)\cdot T^{\Omega}_{LA}\left(n,\mu_p\right)\cdot T^{\Omega}_{CM}\left(n,\omega_{i,j},\tau\right)\cdot T^{\Omega}_{TM}\left(n,\omega_{i,j},mv\right) \quad \Omega\in\{Y,Cb,Cr\} \quad (1)$$

where Ω is Y, Cb or Cr channel, n is the index of a TB, $\varphi_{i,j}$ is the directional angle of the (i, j)-*th* DCT coefficient of a DCT block [18] and *mv* is the motion vector of a TB. In Eq. (1), $\omega_{i,j}$ indicates the spatial frequency in cycles per degree for the (i, j)-*th* transform coefficient of a TB, which is given by

$$\omega_{i,j} = (1/2L)\sqrt{(i/\theta_x)^2 + (j/\theta_y)^2} \quad (2)$$

where θ_x and θ_y are horizontal and vertical visual angles of a pixel, respectively, and L is the size of a TB. In Eq. (1), μ_p is the average pixel intensity of a TB. In Eq. (1), τ indicates an edge pixel density as a texture complexity metric in a TB, which is calculated by

$$\tau = \left(1/L^2\right)\sum_{y}^{N}\sum_{x}^{N} edge(x,y) \quad (3)$$

$edge(x, y)$ is an edge flag which is 0 for non-edge and 1 for edge computed by a 3 × 3 Sobel edge operator at *(x, y)* position.

In Eq. (1), $T^{\Omega}_{CSF}\left(\omega_{i,j},\varphi_{i,j}\right)$ is the CSF at $\omega_{i,j}$ and $\varphi_{i,j}$, $T^{\Omega}_{LA}\left(n,\mu_p\right)$ is the factor for the LA effect, $T^{\Omega}_{CM}\left(n,\omega_{i,j},\tau\right)$ is the factor for the CM effect, $T^{\Omega}_{TM}\left(n,\omega_{i,j},mv\right)$ is the factor for the TM effect and $S^{\Omega}(L)$ accounts for the spatial summation effect of the $L \times L$ TB in a Ω channel. For these five effects in the luma channel, we adopt the exiting parameters for luma JND thresholds [14, 20, 25].

Our new JND model through comprehensive consideration of both luma and chroma channels of five typical effects, with especial focus on parameterized modeling of each effect in chroma channels. These detailed experiments will be addressed in the following remainder of this section.

3.1 Contrast sensitivity function for the luma and chroma TBs

The CSF function quantifies how well the human vision perceives a contrast sensitivity in the spatial frequency domain. The luma CSF expresses high-pass characteristics in the transform domain while the chroma CSF shows low-pass characteristics [6]. The HVS is more sensitive to luma contrast variation in the mid-band frequency than those in low-frequency and high-frequency bands. However, the HVS is less sensitive to chroma contrast changes than the luma sensitivity in the whole frequency band. As the spatial frequency increases, the chroma CSF would decrease. So, we will develop the CSF formula for two chroma TBs and adopt the typical CSF for the luma TB according to CSF characteristics.

For the luma effect of a TB, we adopt the typical spatial CSF, which is calculated by [25]:

$$T_{CSF}^{Y}\left(\omega_{i,j},\varphi_{i,j}\right)=\frac{1}{\phi_i\phi_j}\cdot\frac{\exp\left(cw_{ij}\right)/\left(a+bw_{ij}\right)}{r+(1-r)\cdot\cos^2\varphi_{ij}} \tag{4}$$

where ϕ_i and ϕ_j are normalization factors for transform coefficients, the term $1/(r+(1-r)\cdot\cos^2\varphi_{ij})$ describes the oblique effect, where r is empirically set to 0.6 and φ_{ij} stands for the directional angle of corresponding transform coefficients.

Since [1, 14] use the 8 × 8 DCT block for experiments, this paper uses parameter values of $a = 1.33$, $b = 0.11$ and $c = 0.18$ in [25] which are obtained by the 8 × 8 DCT block based on the fitting method.

According to typical contrast sensitivity functions for chroma channels in [6], we develop the CSF formula for two chroma TBs which is given by

$$T_{CSF}^{Cb/Cr}\left(\omega_{i,j},\varphi_{i,j}\right)=\exp\left(cw_{ij}\right)/\left(a+bw_{ij}\right) \tag{5}$$

To estimate parameter values in Eq. (5), we adopt the least mean squared error solution as the fitting method, and $a = 0.13$, $b = 0.11$ and $c = 0.2$ are obtained for Cr channel while $a = 0.1$, $b = 0.11$ and $c = 0.21$ for Cb channel.

3.2 The luminance adaptation effect for chroma TBs

The HVS has different visibility to different background luminance values [12]. According to Weber-Fechner law, the HVS is less sensitive in the dark or bright environment than the middle range of intensities, which can be called the luminance adaptation (LA) effect. Therefore, HVS luminance sensitive characteristics would be utilized by incorporating the LA effect into the JND model. As in Eq. (1), $T_{LA}^{\Omega}\left(n,\mu_p\right)$ reflects the LA effect of a TB for every channel which can be calculated by a quasi-parabola curve [33].

For chroma TBs, human eyes perceive images/videos in Cr or Cb channel which is also influenced by the luminance sensitivity mechanism [6]. This due to the fact that the YCbCr space is a non-uniform space. So, our proposed JND model considers the LA effect for Cr or Cb TBs by adopting the LA-JND model for luma TBs, which is given by

$$T_{LA}^{Cb/Cr}\left(n,\mu_p\right)=\begin{cases}-\mu_p(A-1)/B+A & \mu_p\le B\\ 1 & B<\mu_p<C\\ \left(\mu_p-C\right)(D-1)/\left(2^Q-1-C\right)+1 & \mu_p\ge C\end{cases} \tag{6}$$

In Eq. (6), the values of A, B, C and D are parameter values of the LA-JND model. And Q indicates bit-depth. To find optimal parameter values in Eq. (6), we perform a chroma LA-JND experiment by the JND suppression in Cb and Cr channels, respectively. According to Weber's U-shape curve, we adjust values of A, B, C and D by selecting all HM test sequence to code 10 frames [14, 20, 25]. Finally $A = 3$, $B = 85$, $C = 90$ and $D = 3$ are obtained with similar subjective visual qualities compared to those without JND-suppression.

3.3 Proposed transform-domain CM-JND model for chroma TBs in YCbCr space

The CM effect indicates that the visibility threshold for non-uniform areas is obviously higher than that of uniform areas. In other words, the HVS is more sensitive to the smooth regions

than the edge and texture region. Therefore, the CM effect can be incorporated into the transform-domain JND model (CM-JND) which takes account into the spatial frequency and the texture complexity.

In previous works for color videos/images with JND-suppression, most of the approaches directly apply luma JND thresholds into chroma channels [20, 27], which will result in underestimating or overestimating JND thresholds in Cb and Cr channels. Chen et al. [6] proposed a chroma CM-JND model in DCT-domain. Since Chen's method only simply considers the CM effect related to DCT coefficients in a 4×4 DCT block, the JND model can hardly accurately estimate HVS thresholds in DCT domain for few frequency samples. Obviously, the method is not suitable in HEVC where different sizes of TBs are used for the transform processing.

In this paper, we propose separately a CM-JND model for Cr and Cb TBs in the YCbCr space with the lowpass property. For each chroma channel, the CM effect for a TB will be modeled as a continuous 2-D function of the spatial frequency and texture complexity. By this way, the proposed CM-JND model can also be used for variable block-sized TBs in HEVC. To obtain the CM effect of every chroma TB, we design a method based on elaborate psychophysical experiments with the 8×8 DCT block size, which can compute the contrast masking effect for different texture patterns to improve the accuracy, in chroma channels. The 8×8 DCT block size can obtain more sufficient frequency samples than the 4×4 DCT block to improve the accuracy of HVS thresholds.

We perform psychophysical experiments to characterize the CM effect for chroma JND in DCT domain with respect to DCT frequency and edge pixel density. Table 1 shows the test conditions for psychophysical experiments. The detailed experimental procedure for detecting distortion is described as follows:

Initialization: Monitor shows a test color image resided in parafovea regions. (**R**2 in Fig. 3)

Step 1: A subject is notified where the distortion will be injected by making the **R**2 region dark. (Fig. 3a)

Step 2: The subject gradually increases the inject noise value for a designated DCT coefficient in the Cb/Cr channel until the viewer starts to perceive the resulting distortion (See in Fig. 3b).

Step 3: *Step* 1 and *Step* 2 are iteratively operated by returning to *Step* 1 with a different participant until all test participants are finished.

Table 1 Expermental setup

Test condition		Value
DI[1)]	Display	Lenovo LT2452p
	Type and size	LED 24″ Wide
	Resolution	1920 × 1200
Number of subjects		10
Viewing distance		0.03dpp[2)](≈1.3 m)
Ambient illumination		300 lx
Test color image size		256 × 256 pixels
Test color patch size		32 × 32 pixels
Bit depth of the test patch		8 bit

[1)] DI: Display information; [2)] dpp: visual angle of a pixel (degree per pixel)

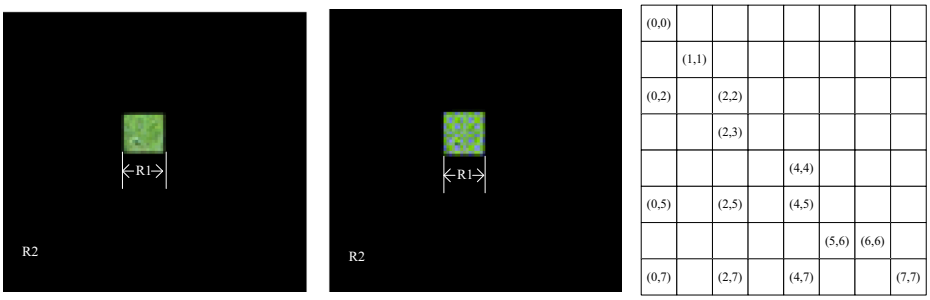


Fig. 3 A test color patch case. (**a**) a test color patch in a fovea region; (**b**) visible chroma distortion by increasing the noise values of the (4,4)-th coefficients of all 8 × 8 DCTs in the patch for Cb channel; (**c**) positions of the selected 15 DCT basis functions

We select the 8 × 8 DCT block size for every chroma channel as the test patch DCT transform size. In psychophysical experiments, the perceptual sensitivity of DCT coefficients is symmetric for upper and lower triangular components. Therefore, we choose 15 DCT chroma coefficients sparsely for perceptual chroma CM-JND measurements in a lower triangle frequency zone of 8 × 8 DCT block [1]. Fig. 3 shows a test color patch exemplar with a central color image patch where distortion is supposed to be injected in a Cb channel.

In this paper, we use average pixel intensity metric and edge pixel density metric for modeling the CM-JND in Cb and Cr channels, respectively. Average pixel intensity metric can be calculated as

$$\mu_{pixel} = \left(1/kL^2\right) \sum_{y}^{N} \sum_{x}^{N} I(x,y) \tag{7}$$

where k is the dynamic range of pixel values (255 for 8 bit-depth image), L is DCT block size and $I(x,y)$ is the pixel intensity at (x,y). Edge pixel density metric employs τ in Eq. (3). We select 7 patches from some test video sequences with YCbCr420 format. To eliminate the LA effect, these patches will be selected of which average pixel intensity values are between 0.35 to 0.8 in two chroma channels, as shown in Table 2.

Table 2 Sequence patches used in chroma JND measurements for CM effects

Sequences	Cb channel		Cr channel		Patches
	τ	μ_{pixel}	τ	μ_{pixel}	
SlideShow	0	0.8	0	0.5	
RaceHorses	0.21	0.55	0.23	0.39	
RaceHorses	0.12	0.58	0.17	0.59	
RaceHorses	0.08	0.7	0.12	0.54	
RaceHorses	0.14	0.44	0.15	0.47	
BQMall	0.2	0.43	0.18	0.47	
RaceHorses	0.09	0.35	0.08	0.47	

The chroma CM-JND is modeled as a normalization factor which is given by

$$T_{CM}^{Cb/Cr}(n,\omega_{i,j},\tau) = JND_{CM}^{Cb/Cr}(n,\omega_{i,j},\tau)/JND_{CM}^{Cb/Cr}(n,\omega_{i,j},\tau=0) \tag{8}$$

$JND_{CMCb/Cr}(n,\omega_{i,j},\tau)$ is the JND value of Cb or Cr channel for the CM effect at $(\omega_{i,j},\tau)$. In Eq. (8), $JND_{CMCb/Cr}(n,\omega_{i,j},\tau=0)$ is only related to CSF which is excluded from LA and CM effects.

Figure 4a and b show the measured T_{CM}^{Cb} values from psychophysical experiments. For the range of edge pixel density metric τ between 0 and 0.1 in Fig. 4a, it is seen that $T_{CM}^{Cb}(n,\omega_{i,j},\tau)$ is decreased as τ increases. $T_{CM}^{Cb}(n,\omega_{i,j},\tau)$ is increased as τ increases in the range of edge between 0.1 and 0.2. For $\tau > 0.2$, $T_{CM}^{Cb}(n,\omega_{i,j},\tau)$ decreases gradually. It is also observed in Fig. 4b that $T_{CM}^{Cb}(n,\omega_{i,j},\tau)$ performs as a low pass filter, which shows larger T_{CM}^{Cb} values in the range of ω between 0 and 5 cpd. So, based on the measured T_{CM}^{Cb} values in Fig. 4a and b, a 2-D T_{CM}^{Cb} is model as a continuous 2-D function of ω and τ which is given by

$$T_{CM}^{Cb}(n,\omega_{i,j},\tau) = \begin{cases} (g_{\tau=0.1}-1)(\tau)/0.1+1 & 0\le\tau<0.1 \\ 10(g_{\tau=0.2}-g_{\tau=0.1})(\tau-0.1)+g_{\tau=0.1} & 0.1\le\tau<0.2 \\ g_{\tau=0.2}+(g_{\tau=0.1}-g_{\tau=0.2})(\tau-0.2)/0.05 & 0.2\le\tau \end{cases} \tag{9}$$

where $g_{\tau CM}(\omega)$ is modeled in a gamma probability density function which is given by

$$g_{\tau}^{CM}(\omega) = \left(m^{v}/\Gamma(0.5)\omega^{(n-1)}\exp(-m\omega)\right)p + t \tag{10}$$

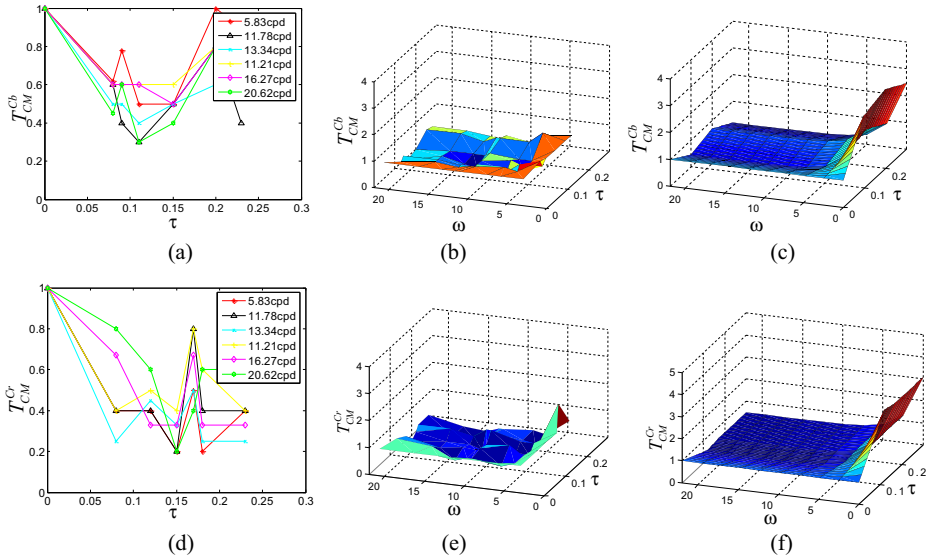


Fig. 4 Psychophysical experiment results and modeled CM effects: (**a**)Measured T_{CM}^{Cb} values projected on τ; (**b**) Measured T_{CM}^{Cb} values; (**c**)Modeled T_{CM}^{Cb} values; (d)Measured T_{CM}^{Cr} values projected on τ; (**e**)Measured T_{CM}^{Cr} values; (**f**)Modeled T_{CM}^{Cr} values

According to psychophysical experiments, parameters are computed based on the least squares solution as:

$$g^{CM}_{\tau=0.1}(\omega) : v = 1, m = 0.5, p = 8, t = 0.45 \tag{11}$$

$$g^{CM}_{\tau=0.2}(\omega) : v = 1, m = 0.5, p = 8, t = 0.7 \tag{12}$$

Similarly, we adopt the same method for the Cr channel with the Cb channel. $T^{Cr}_{CM}(n, \omega_{i,j}, \tau)$ is given by

$$T^{Cr}_{CM}(n, \omega_{i,j}, \tau) = \begin{cases} \left(g^{CM}_{\tau=0.12}(\omega)-1\right)(\tau)/0.12 + 1 & 0 \le \tau < 0.12 \\ 8.33\left(g^{CM}_{\tau=0.17}(\omega)-g^{CM}_{\tau=0.12}(\omega)\right)(\tau - 0.12) + g^{CM}_{\tau=0.12}(\omega) & 0.12 \le \tau < 0.17 \\ g^{CM}_{\tau=0.17}(\omega) & 0.17 \le \tau \end{cases} \tag{13}$$

$$g^{CM}_{\tau=0.12}(\omega) : v = 1, m = 0.5, p = 10, t = 0.4 \tag{14}$$

$$g^{CM}_{\tau=0.17}(\omega) : v = 1, m = 0.5, p = 8, t = 0.6 \tag{15}$$

By adopting elaborate psychophysical experiments, we can obtain CM effects for two chroma channels in different texture regions to accurately estimate JND thresholds. The proposed $T^{Cb}_{CM}(n, \omega_{i,j}, \tau)$ and $T^{Cr}_{CM}(n, \omega_{i,j}, \tau)$ models in (9) and (13) are shown Fig. 4c and f, respectively.

3.4 The TM effect for chroma TBs

The TM effect must be considered in the spatio-temporal JND model. It is because the HVS more easily detects distortions for slow moving objects than for still and fast moving objects. The TM effect can also be incorporated into our JND model in Eq. (1). In Cr and Cb channel, we adopt the TM-JND model [18, 25], which is given by

$$T^{Cb/Cr}_{TM}(n, \omega_{i,j}, mv) = \begin{cases} 1 & f_s < 5cpd \& f_t < 10Hz \\ 1.07^{(f_t-10)} & f_s < 5cpd \& f_t \ge 10Hz \\ 1.07^{f_t} & f_s \ge 5cpd \end{cases} \tag{16}$$

where f_s is the spatial frequency related to $\omega_{i,j}$ of chroma TBs, f_t is the temporal frequency related to mv of chorma TBs and the eye movement.

3.5 The summation effect for variable block sizes TBs

Since the distortion of a block is obtained by the probability summation model for individual JND threshold of transform coefficients over spatial neighborhood, a summation effect factor must be considered in the JND model. In previous transform-domain JND models, the summation effect only uses a constant value as a summation effect factor for a fix-sized block.

In these works, the summation effect factor value of luma TBs is set to 0.25, and chroma TBs is set to 1 [9, 25]. However it is unsuitable for variable-sized TBs using various sizes transform core in HEVC to apply the constant summation effect value for the fix-sized block. Therefore, we propose a summation effect function about various sizes for chroma TBs, and adopt the summation effect function for luma TBs in [14]. In order to deduce the summation effect function for chroma TBs, we briefly introduce the probability summation model and the summation effect function for luma TBs in [1, 14].

According to the probability summation model in [22], the probability detecting the distortion at a spatial frequency ω by injecting noise is modeled based on psychometric function, which is given by

$$P(\omega) = 1-2^{(-|\Delta C(\omega)/JND(\omega)|^{\eta})} \tag{17}$$

where $\Delta C(\omega)$ is difference between the original and the distortion transform coefficient, $JND(\omega)$ is the measured JND threshold at ω for 50% of observers detect the distortion from experiments. In Eq. (17), the value of η is used to fit the normalized histogram of observers' JND perceiving.

The model in Eq. (17) is extend to the whole detection probability for all distorted transform coefficients of a $L \times L$ block, based on the following two assumptions [22]: 1)The whole distortion of a $L \times L$ block is detected if the least one distorted transform coefficient is perceived, and 2)The distortion detection of every noise-injected transform coefficient in a $L \times L$ block is uncorrelated.

According to two assumptions above, the summation detection probability of a $L \times L$ block, represented asB_L, can be expressed by

$$P(B_L) = 1- \prod_{\omega \in B_L} (1-P(\omega)) \tag{18}$$

Substituting (17) into (18) yields

$$P(B_L) = 1-2^{-D_{B_L}^{\eta}} \tag{19}$$

where

$$D_{B_L} = \left(\sum_{\omega \in B_L} \left| \frac{\Delta C(\omega)}{JND(\omega)} \right|^{\eta} \right)^{1/\eta} \tag{20}$$

where D_{B_L} is the whole distortion in a $L \times L$ block.

It can be seen that as L of the block size increases, D_{B_L} increases resulting in increasing $P(B_L)$ in Eq. (19). This will lead to making the distortion of the noise-injected block more easily visible. Consequently, to maintain the identical detection probability for various transform sizes when obtaining corresponding JND thresholds of TBs, D_{B_L} must maintain identically for different sizes TBs. For this, the summation effect function of $S(L)$ is proposed to compensate the JND threshold of individual transform coefficient for different transform block sizes. By this way, D_{B_L} for variable block-sized TBs will be identical.

Based the single channel case in [1], we derive the multi-channel summation effect function of $S^{\Omega}(L)$ which $S(L)$ can be extend to the proposed multi-channel $S^{\Omega}(L)$ in Eq. (1). In order to model the summation effect function for every channel, we inject the noise separately into transform

coefficients $C(\omega)$ of different channel TBs to meet the condition with $\Delta C(\omega) = JND^{\Omega}_{spatio-temporal}$ in Eq. (1) for every channel. So putting Eq. (1) into Eq. (20), we can obtain

$$D^{\Omega}_{B_L} = \left(\sum_{\omega \in B_L} \left| S^{\Omega}(L) \right|^{\eta^{\Omega}} \right)^{1/\eta^{\Omega}} \tag{21}$$

Eq. (21) can be rewritten as

$$S^{\Omega}(L) = D^{\Omega}_{B_L} \cdot L^{-2/\eta^{\Omega}} \tag{22}$$

The D_{B_L} and η for Ω channel need to be parameterized in Eq. (22). For Cb and Cr channels, we perform psychophysical experiments under the same experimental condition in Table 1. For $S^{Y}(L)$ in the luma channel, we adopt parameters in [14].

Since the HVS color perception depends on the concentration of cones in retina, the fovea region is more sensitive to color perception where there is a high density of cones. The fovea region covers approximately 2° in visual angle. So, in our psychophysical experiments, we select 32×32 pixels color block as the test patch which can be fully projected on the fovea region. To support three different transform block sizes of 16×16, 8×8 and 4×4 in Cb and Cr channels, $S^{Cr}(L)$and $S^{Cb}(L)$ are measured. To well estimate JND thresholds under two assumptions in [22], a homogeneous test color patch with μ_{pixel}=0.8, τ =0 for Cr channel, and μ_{pixel} =0.5, τ =0 for Cb channel is used for experiments. We inject JND thresholds adjusted by $S(L)$ in Eq. (1) to all coefficients of each corresponding Cb and Cr transform block in the test color patch to estimate the parameters. Ten viewers are involved in experiments. Each viewer gradually increases JND values by adjusting $S(L)$ from zero until the viewer starts to perceive the distortion for 16×16, 8×8 and 4×4 Cb and Cr TBs. When 50% of viewers perceive the distortion for the block, $S(L)$ for Cb and Cr channels would be obtained separately. From $S(L)$ for Cb and Cr channels, we obtain parameters of $D^{Cb}_{B_L}$=2.24, $D^{Cr}_{B_l}$=12.02, η^{Cb}=2.15 and η^{Cr}=1.33 based on the least square solution. At last, we can obtain the summation effect function of $S^{Cb}(L)$ and $S^{Cr}(L)$. Figure 5 shows the accuracy of measured $S(L)$ and modeled $S(L)$ for Cb and Cr channels, whose JND thresholds are higher than the luma channel.

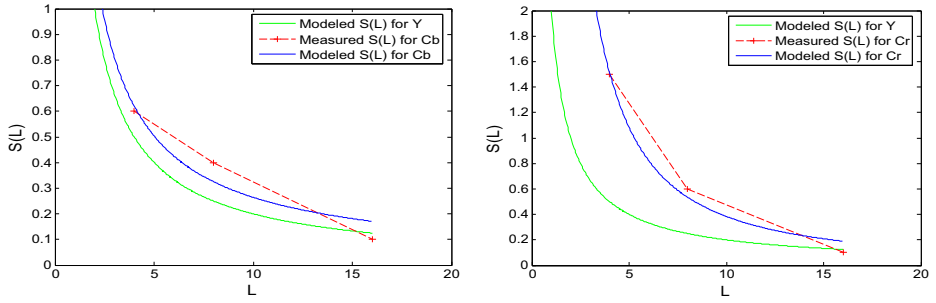


Fig. 5 Psychophysical experiment results and modeled summation effects: (**a**) the modeled $S^{Y}(L)$, measured $S^{Cb}(L)$ and modeled $S^{Cb}(L)$; (**b**) the modeled $S^{Y}(L)$, measured $S^{Cr}(L)$ and modeled $S^{Cr}(L)$

4 Fast and HEVC-compatible PVC scheme based on the proposed JND model

This section will introduce our PVC scheme which incorporates the proposed JND model into HM 15.0 [13]. Since previous HEVC-based PVC scheme [14] has two limitations in decreasing the computational complexity: 1) TU-level Sobel edge detection is performed repeatedly for the iterative TB partition based on the RDO process to computing the texture complexity, and 2) transform coefficients are only considered the relationship between the luma component and JND thresholds, leading to increasing memory access for chroma components. To fix aforementioned problems, our fast PVC scheme is proposed by performing the frame-level overall edge detection as will be presented in Section 4.1 and using the JND-based multi-channel transform coefficients suppression method as will be elaborated in Section 4.2.

4.1 The frame-level texture complexity computation for TBs

HEVC uses repeatedly the iterative TB partition to obtain the best R-D performance. Therefore, the frame-level overall edge detection for reducing complexity and improving performance in the encoder side have two advantages: 1) decrease duplicate operations of the Sobel edge detection for the iterative TU partition based on the RDO process, and 2) the Sobel edge operator employed for local edge detection in every block is more ineffective to achieve the frame-level overall edge profile corresponding the HVS. It is because edge pixels of blocks can hardly obtain pixels of neighbor blocks to compare their intensity values. In summary, the frame-level overall detection can be more effective than the TU-level detection by experiments.

4.2 JND-based multi-channel suppression

For the JND-based suppression of conventional PVC schemes, transform coefficients are only considered the relationship between the luma component and JND thresholds, leading to increasing memory access for chroma components in the multi-channel case. So the complexity can be further decreased. HEVC transform in the codec can be expressed by

$$\begin{aligned}\mathbf{Y}_{N\times N} &= \beta(N)\cdot\left\{\chi(N)\cdot[\psi(N)\cdot(\mathbf{C}_{N\_HEVC}\times\mathbf{X}_{N\times N})]\times\mathbf{C}_{N\_HEVC}{}^{T}\right\}\\ &= \beta(N)\cdot\mathbf{Y}'_{N\times N}\end{aligned} \tag{23}$$

where $\mathbf{Y}_{N\times N}$ is the finite precision approximations, $\mathbf{X}_{N\times N}$ is the residual coefficients matrix, $\mathbf{C}_{N\_HEVC}$ is the real-valued DCT matrix, $\mathbf{Y}'_{N\times N}$ is the output residual transform coefficients in HM codec, $\beta(N)=2^{-7+\log_2 N}$, $\chi(N)=2^{-\log_2 N-6}$ and $\psi(N)=2^{-\log_2 N+1}$.

The quantization process in HEVC can be expressed as

$$\left|l_{i,j}\right| = \left(\left|d_{i,j}\right|\times f[QP\%6]+offset\right) >> b \quad i,j=0,\cdots,N-1 \tag{24}$$

$$sign\left(l_{i,j}\right) = sign\left(d_{i,j}\right) \tag{25}$$

where $d_{i,j}$ is the element of the matrix $\mathbf{Y}'_{N\times N}$, $l_{i,j}$ is the element of the quantization matrix, $f[QP\%6]$ is the table with 6 entries, $b=21-N+QP/6$ and the offset is a rounding offset which is between $2^{20-N+QP/6}$ and $2^{21-N+QP/6}$.

It is noted in Eq. (24) that if $\mathbf{Y}'_{N\times N}$ will be suppression with JND thresholds, JND thresholds need be scaled. It is because that JND thresholds are obtained from the classic DCT matrix, while $\mathbf{Y}'_{N\times N}$ is calculated by the integer DCT matrix.

Thus, Eq. (23) can be rewritten by

$$\hat{\mathbf{Y}}_{N\times N} = \beta(N)\cdot\left(\mathbf{Y}'_{N\times N} - \gamma(N)\cdot\mathbf{JND}_{N\times N}\right) \tag{26}$$

where $\hat{\mathbf{Y}}_{N\times N}$ is a JND suppression matrix of $\mathbf{Y}'_{N\times N}$, $\gamma(N)$ is the JND threshold scaling factor between the real-valued DCT matrix and finite precision approximations matrix, which is equal to $\beta^{-1}(N)$ corresponding to the block size N, $\mathbf{JND}_{N\times N}$ is calculated by the integer DCT matrix.

Accordingly, the JND-based luma-channel quantization with coefficient suppression can be derive from Eq. (24), which can be expressed by

$$\left|l^*_{i,j}\right| = \begin{cases} \left(\left(\left|d_{i,j}\right| - \gamma(N)\left|JND_{i,j}\right|\right) \times f[QP\%6] + offset\right) >> b & \left|d_{i,j}\right| > \gamma(N)\left|JND_{i,j}\right| \\ 0 & otherwise \end{cases} \tag{27}$$

where $d_{i,j}$ is the element of the matrix $\mathbf{Y}'_{N\times N}$, $l_{i,j*}$ is the quantization level of the transform coefficient $d_{i,j}$ suppressed by the JND threshold, $JND_{i,j}$ is the element of the matrix $\mathbf{JND}_{N\times N}$.

The QP will be different in luma and chroma channels when QP is greater than or equal to 30. It is because that the large quantization step used in chroma channel results in color drift. Therefore, Eq. (27) can be used in the luma channel, but not be used directly in chrome channels when QP is greater than or equal to 30. So, we derive the JND-based chroma-channel quantization when QP is greater than or equal to 30, which can be given by

$$\left|l^*_{i,j}\right| = \begin{cases} \left(\left(\left|d_{i,j}\right| - \gamma(N)\left|JND_{i,j}\right|\right) \times f[QP_{chroma}\%6] + offset\right) >> b & \left|d_{i,j}\right| > \gamma(N)\left|JND_{i,j}\right| \\ 0 & otherwise \end{cases} \tag{28}$$

where QP_{chroma} is chroma QP which can be obtained from QP.

Combining Eqs. (27) and (28), the JND-based chroma-channel quantization with coefficient suppression can be derive as

$$\left|l^*_{i,j}\right| = \begin{cases} \left(\left(\left|d_{i,j}\right| - \gamma(N)\left|JND_{i,j}\right|\right) \times f[QP\%6] + \text{offset}\right) >> b & \left|d_{i,j}\right| > \gamma(N)\left|JND_{i,j}\right| \quad and \quad QP < 30 \\ \left(\left(\left|d_{i,j}\right| - \gamma(N)\left|JND_{i,j}\right|\right) \times f[QP_{\text{chroma}}\%6] + \text{offset}\right) >> b & \left|d_{i,j}\right| > \gamma(N)\left|JND_{i,j}\right| \quad and \quad QP \geq 30 \\ 0 \quad \text{otherwise} \end{cases} \tag{29}$$

where $l_{i,j*}$ is the quantization coefficient level of the transform coefficient $d_{i,j}$ suppressed by the JND threshold, $f[QP\%6]$ is the quantization array, $offset$ is a rounding offset, and $\gamma(N)$ is a JND threshold scaling factor corresponding to the block size N. Eqs. (27) and (29) can support transform coefficients with the variable-sized block in HEVC. Moreover, it can be employed effectively in three channels with fewer duplication operation and memory access.

4.3 Summary of the proposed PVC scheme

In our proposed PVC scheme, we firstly incorporate the proposed transform-domain multi-channel JND model into the codec. And our proposed fast method is also integrated into the codec. It is noted that the proposed PVC scheme is only suppresses coefficients in the

encoding side, the output result bitstreams is HEVC-compatible. The detail description of proposed PVC scheme in HM15.0 is given in 'Algorithm 1'.

Algorithm 1 hierarchical encoding steps with the proposed PVC

1 For each frame of the video sequence
2 Edge detection for binaryzation
2.1 Run the frame-level Sobel edge operator to produce a binaryzation map
3 For each CTU of the frame
4 For each CU encoding of the CTU
5 For each TU of the CU
6 Calculate multi-channel JND thresholds
6.1 Calculate the texture complexity metric τ based on the binaryzation map and the average pixel intensity μ_{pixel} for the luma TB using (3) and (7)
6.2 Calculate the CSF factor T_{CSF}^{Y} and the summation effect factor $S^{Y}(L)$ using (4) and (22)
6.3 Calculate the texture complexity metric τ based on the binaryzation map and the average pixel intensity μ_{pixel} for two chroma TBs using (3) and (7)
6.4 Calculate the CSF factor $T_{CSF}^{Cb/Cr}$, the LA factor $T_{LA}^{Cb/Cr}$, the CM factor T_{CM}^{Cb}, T_{CM}^{Cr}, the TM factor $T_{TM}^{Cb/Cr}$ and the summation effect factor $S^{Cb/Cr}(L)$ using (5), (6), (9), (13), (16) and (22), respectively
7 JND-based multi-channel quantization
8 Compute RD cost
9 End TU-level encoding
10 Select the best mode
11 End CU-level encoding
12 End CTU-level encoding
13 End frame-level encoding

5 Experimental results

This section will evaluate performance of our proposed CM-JND model and HM-based PVC scheme, respectively. The performance of our CM-JND model can further verify CM effects in chroma channels. The performance of the PVC scheme can evaluate gains in the codec by incorporating the proposed multi-channel JND model.

5.1 The performance of the proposed CM-JND model

In order to evaluate the effectiveness of the proposed multi-channel CM-JND model, we use the adjectival categorical judgment method [1, 19] that shows reference images and the distorted images in a side-by-side way. The quality assessment consists of seven score values from better quality to worse quality: 3 for "*Much better*", 2 for "*Better*", 1 for "*Slightly better*", 0 for "*The same*", −1 for "*Slightly worse*", −2 for "*Worse*", and −3 for

"*Much worse*". The view test condition is the same as the one in Table 1. Considering the model used in the YCbCr space and related to the content of sequences instead of the size for video, we select the video frame from eight typical HM video sequences with YCbCr420 format as shown in Fig. 6 [2, 3]. The CM-JND model of Luo's model is employed in the luma channel and taken advantage of corresponding JND thresholds at 4×4 level in chroma channels [18]. Chen's model proposed a chroma CM-JND model in DCT-domain which only simply considers the CM effect related to DCT coefficients in a 4×4 DCT block [6].

To compare Luo's multi-channel CM-JND model for color video coding with the proposed multi-channel CM-JND model, noise will be injected to each coefficient of DCT blocks for every channel in video frames according to the JND threshold as follows [7, 27]:

$$\hat{C}(n,i,j) = C(n,i,j) + \varepsilon S_{rand}(n,i,j) \times JND(n,i,j) \tag{30}$$

where $C(n,i,j)$ is the *(i,j)-th* DCT coefficient, $C(n,i,j)$ is the noise-injected coefficient, ε regulars the magnitude of the JND noise and $S_{rand}(n,i,j)$ takes the value of +1 or −1 randomly, to avoid generating a fixed pattern of changes. Generally, a better model will obtain higher JND thresholds for the region which is insensitive to the HVS, while lower JND thresholds for the region which is sensitive to the HVS. From (30), a better JND model will inject more noise into the insensitive region and less noise into the sensitive region. Therefore, if we inject the same level noise into the original image, a more accurate JND model will obtain better quality than a less accurate JND model. At the same subjective quality, a better model will obtain higher JND thresholds and lead to lower PSNR. Table 3 shows the comparison of the proposed CM-JND model and Luo's model in terms of peak signal to nosie ratio (PSNR) and mean opinion score (MOS) values.

In Table 3, MOS values of the proposed CM-JND model are closer to zero than Luo's model and Chen's model. In average PSNR from the Table 3, the proposed model obtains 27.21 dB which is 0.21 dB lower than Luo's model in the luma channel. And the proposed model also is 0.24 dB lower than Chen's model in the luma channel. This is because the proposed model adopts a luma CM-JND model without the block classification to improve the CM effect [34]. It is noticed that the proposed model yields 27.28 dB and 27.29 dB which are 0.13 dB and 0.12 dB lower than Luo's model in Cb and Cr channels, respectively. This due to the fact that the proposed model adopts an elaborate method to estimate CM effects in two chroma channels. Also, the proposed model is 0.05 dB and 0.02 dB lower than Chen's model in Cb and Cr channels, respectively. Moreover the average MOS value of the proposed model is higher than Chen's model. This is because the proposed

Fig. 6 The 8 test frames used in the subjective assessments

Table 3 Comparison of three CM-JND models

Test frame	LUO's [18]				CHEN's [6]				Proposed			
	PSNR			*MOS*	*PSNR*			*MOS*	*PSNR*			*MOS*
	Y	*Cb*	*Cr*		*Y*	*Cb*	*Cr*		*Y*	*Cb*	*Cr*	
Ball	27.44	27.43	27.43	−1.4	27.92	27.53	26.84	−1.1	27.41	27.02	27.09	**−0.6**
Horses	27.08	27.07	27.07	−0.4	27.02	27.01	27.13	−0.2	26.7	26.93	27.03	**−0.2**
Mall	27.1	27.01	27.01	−0.5	27.24	26.97	26.92	−0.3	26.92	27.05	27.00	**−0.3**
Party	27.78	27.79	27.79	−0.9	27.65	27.59	27.66	−0.8	27.75	27.54	27.59	−0.9
Text	27.35	27.36	27.36	−1.3	27.24	27.31	27.39	−0.6	27.16	27.34	27.19	**−0.4**
Flower	27.53	27.52	27.52	−0.9	27. 6	27.3	27.32	−1.2	27.06	27.28	27.22	**−0.4**
Keiba	27.69	27.68	27.68	−1.2	27.69	27.62	27.64	−0.9	27.45	27.81	27.67	−1.1
Mobi	27.4	27.39	27.39	−0.2	27.41	27.3	27.67	−0.2	27.21	27.24	27.57	−0.2
avg	27.42	27.41	27.41	−0.85	27.45	27.33	27.32	−0.66	27.21	27.28	27.30	**−0.51**

model uses the block size which is larger than Chen's model to improve the accuracy of JND thresholds in the spatial frequency.

Figure 7 shows the subjective performance comparison of three multi-channel CM-JND models for *Ball* video frame. Figure 7b, c and d show the noise-injected frame by Luo's, Chen's and proposed model, respectively. As can be seen, Luo's model shows more visible distortions over the frame, especially, in the some smooth regions (e.g., the hoop region). This is because JND thresholds are overestimated in these smooth regions. Chen's model shows more discrete visible distortions over the frame. This is because Chen's model estimates JND profiles in low frequency, so the density of noise is low. Figure 7d shows that the proposed CM-JND model achieves a better quality with almost the same PSNR value in the luma channel and lower PSNR values in chroma channels compared with Luo's and Chen's model. It is seen that JND values are more accurately estimated in the hoop region. And, the texture region (e.g., the cord net region) shows less visible distortions with higher JND noise in chroma channels. Similar results can be observed on other test video frame, as shown in Fig. 8. In all, our proposed multi-channel CM-JND model consistently outperforms two models, in items of both PSNR and MOS values.

Fig. 7 The '*Ball*' video frame of the original and the distortion-injected by the three multi-channel CM-JND models: (**a**) The origin frame; (**b**) Luo's model(PSNR Y:27.44, PSNR Cb:27.43, PSNR Cr:27.43, MOS: -1.4) (**c**) Chen's model(PSNR Y:27.92, PSNR Cb:27.53, PSNR Cr:26.84, MOS: -1.1) (**d**) The proposed model(PSNR Y:27.41, PSNR Cb:27.02, PSNR Cr:27.09, MOS: -0.6)

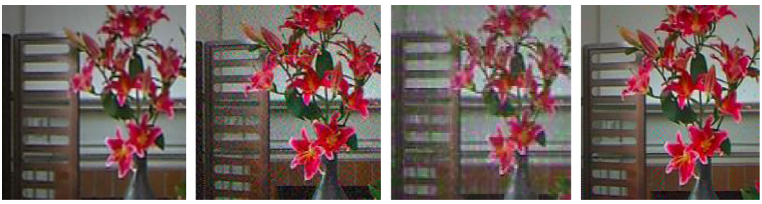


Fig. 8 The *'Flower'* video frame of the original and the distortion-injected by the three multi-channel CM-JND models: (**a**) The origin frame; (**b**) Luo's model(PSNR Y:27.53, PSNR Cb:27.52, PSNR Cr:27.52, MOS: -0.9) (**c**) Chen's model(PSNR Y:27.28, PSNR Cb:27.65, PSNR Cr:27.64, MOS: -1.2) (**d**)The proposed model(PSNR Y:27.06, PSNR Cb:27.28, PSNR Cr:27.22, MOS: -0.4)

5.2 The performance of the proposed PVC scheme

In order to evaluate performance of the proposed PVC scheme, our PVC scheme is incorporated into HM15.0. For performance comparison purpose, the state-of-the-art PVC scheme for HEVC, ie., Kim's PVC scheme specified in [14], which improves Luo's method implemented in H.264/AVC is also incorporated into HM15.0 for fair comparison. Moreover, we incorporate Chen's model into HM15.0 for fair comparison. Therefore, the proposed PVC scheme will be compared with the original HM15.0 the recent Kim's PVC scheme and Chen's PVC scheme under the same encoding configuration.

Since Kim's PVC scheme uses the Sobel operation in the TU level of HM11.0 to compute the complexity of TUs, we apply the operation in TUs of HM15.0 for the Kim's method to make fair comparison. Furthermore, the Kim's method uses two times RDOQ to adjust the RDO process for the mode selection in HM11.0, so we also calculate two times RDOQ to use the same RDO process in HM 15.0 for fair comparison when Kim's method is incorporated into HM15.0.

To evaluate the coding performance of the proposed PVC scheme and compared fairly it with state-of-the-art PVC schemes with Default Lower Delay (LD) and Random Access (RA) configurations [2, 3, 14]. These two configurations are used for all experiments. Four Full-HD (1920 × 1080) and two WVGA (832 × 480) test sequences with different motion characteristics, pixel intensity, background and object complexity are selected in our experiment, which are harmonizing with Kim's test sequences. It is because many video services support high-resolution videos currently. Moreover, our proposed PVC scheme will be mainly used for high-resolution video contents. Experiments are performed on recommended sequences under different QPs (22, 27, 32, 37). The first 100 frames of test sequences are encoded. The simulation is operated on a PC with an Intel 3.4 GHz and 4.0 GB memory.

A better PVC scheme should be more bitrate reduction at the same visual quality and lower computational complexity. To evaluate the objective performance of the proposed PVC scheme compared with the original HM15.0, Kim's PVC scheme [14] and Chen's PVC scheme [6], we use the Multiple Scale-Structural Similarly Index (MS-SSIM) measure, bitrate reduction and encoding time. The bitrate reduction and the encoding time are used for measuring performance of three PVC schemes, compared to HM15.0.

Table 4 shows the object performance of the original HM15.0, Kim's PVC scheme [14], Chen's PVC scheme [6] and the proposed PVC scheme under LD configuration. It

Table 4 Preformance comaprison of the original HM15.0, Kim's PVC scheme, Chen's PVC scheme and proposed PVC scheme under LD configuration

Sequences	QP	MS-SSIM				bitrates(kb/s)				△bitrate(%)			△time(%)		
		HM15	Kim	Chen	Prop.	HM15	Kim	Chen	Prop.	Kim	Chen	Prop.	Kim	Chen	Prop.
Ca	22	0.9976	0.9924	0.9933	0.9912	49,838	39,166	39,628	36,924	−21.41	−20.49	−25.91	29.36	14.22	12.41
	27	0.9863	0.9846	0.9849	0.9851	9505	8583	8732	8450	−9.7	−8.13	−11.1	39.06	20.57	16.39
	32	0.9763	0.9714	0.9714	0.9733	3779	3517	3524	3482	−6.93	−6.75	−7.86	40.52	21.81	20.74
	37	0.9556	0.9439	0.9492	0.9491	1756	1669	1671	1658	−4.95	−4.84	−5.58	27.93	12.39	11.32
BD	22	0.9915	0.9889	0.9899	0.9888	35,088	31,046	31,848	30,055	−11.52	−9.23	−14.34	26.74	15.98	14.01
	27	0.9884	0.9842	0.9874	0.9828	9778	9449	9513	9345	−3.36	−2.71	−4.43	30.09	19.31	17.26
	32	0.9796	0.9733	0.9771	0.9709	4332	4225	4188	4163	−2.47	−3.32	−3.9	34.8	23.41	22.81
	37	0.9603	0.9546	0.9584	0.9476	2125	2065	2095	2045	−2.82	−1.41	−3.76	29.41	18.06	15.98
Ki	22	0.9942	0.9933	0.9939	0.9932	9758	8893	9141	8622	−8.86	−6.32	−11.64	32.11	15.92	14.77
	27	0.9901	0.9876	0.9891	0.9874	4643	4310	4361	4254	−7.18	−6.08	−8.39	45.7	28.19	26.6
	32	0.9792	0.9708	0.9769	0.9706	2297	2072	2149	2053	−9.78	−6.44	−10.60	48.57	34.46	31.94
	37	0.9544	0.9372	0.9478	0.9377	1108	1016	1033	1005	−8.31	−6.79	−9.28	49.92	41.9	39.29
PS	22	0.9927	0.9918	0.9921	0.9916	13,659	12,288	12,453	11,924	−10.04	−8.83	−12.70	41.1	25.53	22.41
	27	0.9862	0.9844	0.9851	0.9841	5170	4729	4822	4603	−8.53	−6.73	−10.97	47.8	38.27	35.32
	32	0.9732	0.968	0.9719	0.968	2134	1926	2065	1886	−9.74	−3.24	−11.62	50.58	40.57	39.7
	37	0.9482	0.9372	0.9441	0.9369	908	813	822	807	−10.49	−9.46	−11.16	51.66	46.32	43.59
BM	22	0.9963	0.996	0.9961	0.9959	7624	7323	7340	7045	−3.94	−3.72	−7.59	47.96	42.84	41.24
	27	0.9934	0.9928	0.993	0.9927	3275	3114	3124	3070	−4.91	−4.61	−6.26	49.4	45.69	43.91
	32	0.9871	0.9856	0.9863	0.9854	1547	1465	1478	1442	−5.30	−4.47	−6.81	63.77	49.58	48.78
	37	0.9728	0.9691	0.9711	0.9688	769	734	731	715	−4.55	−4.94	−7.02	69.71	7.11	5.2
RH	22	0.9951	0.9925	0.9932	0.9949	10,006	9436	9521	9014	−5.70	−4.85	−9.91	42.71	29.81	25.05
	27	0.9861	0.986	0.986	0.9886	4107	3889	3891	3780	−5.31	−5.26	−7.96	41.36	29.72	28.03
	32	0.9705	0.9636	0.9661	0.9735	1719	1620	1627	1578	−5.77	−5.37	−8.22	42.22	31.45	29.79
	37	0.9617	0.9584	0.9608	0.938	766	709	712	695	−7.40	−7.01	−9.23	43.77	33.28	30.43
Average						-	-	-	-	**−7.46**	**−6.31**	**−9.43**	**42.76**	**28.6**	**26.54**

Ca Cactus, *BD* BasketballDrive, *Ki* Kimono, *PS* ParkScene, *BM* BQMall, *RH* RaceHorses

can be noticed in Table 4 that the average bitrate reduction is 7.46% for Kim's PVC scheme, 6.31% for Chen's PVC scheme and 9.42% for the proposed PVC scheme, compared to the original HM 15.0 with similar MS-SSIM values. The similar MS-SSIM value means the video viewer can rarely perceive visible distortion among reconstructed frames with four encoding approaches. As can be seen, the proposed PVC scheme outperforms both Kim's and Chen's with more bitrate reduction, of 1.97% and 3.12%, respectively. What's more, the proposed scheme achieves the maximum bitrate reduction of 25.91% for the *Cactus* sequence at QP = 22 with the negligible reduced MS-SSIM value. It is because that the *Cactus* sequence exhibits plenty of texture and color regions with hardly perceive visible distortion by JND-suppression. For the encoding time, the proposed PVC scheme increases 28.62%, compared to the original HM 15.0, which is 14.1% faster than Kim's PVC scheme. It is because that Kim's method adopts TU-level Sobel edge detection performed repeatedly for the iterative TU partition and HEVC-based single-channel quantization.

Table 5 shows the object performance of the original HM15.0, Kim's PVC scheme [14], Chen's PVC scheme [6] and the proposed PVC scheme under RA configuration. It can be seen in Table 5 that the average bitrate reduction is 7.01% for Kim's PVC scheme, 6.22% for Chen's PVC scheme and 8.34% for the proposed PVC scheme, compared to the original HM 15.0 with similar MS-SSIM values. The proposed PVC scheme outperforms both Kim's and Chen's with more bitrate reduction, of 1.33% and 2.12%, respectively. The maximum bitrate reduction achieves 18.14% for the *Cactus* sequence at QP = 22 with the negligible reduced MS-SSIM value. For the encoding time, the proposed PVC scheme increases 32.07%, compared to the original HM15.0, which is 24.86% faster than Kim's PVC scheme.

The Double-Stimulus Continuous Quality Scale (DSCQS) method is used for subjective quality assessment [19]. The DSCQS method is a classic standard for subjective quality assessment of video contents [2, 14, 18]. Two reconstructed PVC test sequences are compared in term of the DMOS (different mean opinion score) value as

$$DMOS = MOS_{PVC} - MOS_{ORI} \tag{33}$$

where MOS_{PVC} and MOS_{ORI} denote the MOS values of reconstructed sequences from a PVC scheme and the original HM15.0, respectively.

The display condition for subjective quality assessment experiments is the same as those in Table 1. The view distance is set to 1.3 m (about 3 times the screen height). Fifteen viewers including ten male and five female participate subjective quality assessment experiments.

Figure 9 shows the subjective performance of Kim's, Chen's and the proposed PVC scheme under LD configuration. As can be seen, MOS values of Kim's, Chen's and the proposed PVC scheme for all test sequences at four QP values under LD configuration are not seriously different. Generally, there are less residual coefficients in RA case than LD case. Therefore, the JND-based method will obtain less residual reduction under RA configuration. So the subjective performance under RA configuration is similar with LD case at least.

Figure 10 shows subjective quality comparisons for reconstructed frames and their residual distribution at QP = 27. The test sequence *BQMall* with abundant color data can effectively varify subjective quality by using our scheme. The Fig. 10a, b, c and d show encoding a video

Table 5 Preformance comaprison of the original HM15.0, Kim's PVC scheme,Chen's PVC scheme and proposed PVC scheme under RA configuration

Sequences	QP	MS-SSIM				bitrates(kb/s)				△bitrate(%)			△time(%)		
		HM15	Kim	Chen	Prop.	HM15	Kim	Chen	Prop.	Kim	Chen	Prop.	Kim	Chen	Prop.
Ca	22	0.9921	0.991	0.9913	0.9902	40,820	34,641	35,673	33,417	−15.14	−12.61	−18.14	42.39	21.41	20.92
	27	0.9871	0.9856	0.9866	0.9847	9151	8548	8631	8489	−6.59	−5.68	−7.23	53.11	29.06	28.27
	32	0.9799	0.9797	0.9797	0.9791	3927	3655	3655	3653	−6.93	−6.93	−6.97	60.56	33.99	32.89
	37	0.966	0.9613	0.963	0.9603	1756	1746	1744	1737	−0.58	−0.68	−1.09	49.77	29.53	27.13
BD	22	0.9916	0.9892	0.9905	0.9872	22,085	19,160	19,424	18,209	−13.25	−12.05	−17.55	44.26	25.42	24.74
	27	0.9873	0.984	0.9851	0.9826	7023	6370	6403	6209	−9.30	−8.83	−11.59	53.68	31.68	30.34
	32	0.9723	0.9718	0.9719	0.9702	3332	3051	3071	3050	−8.43	−7.83	−8.46	68.02	40.17	39.79
	37	0.9654	0.9513	0.9646	0.9504	1514	1511	1502	1510	−0.19	−0.79	−0.24	58.38	38.63	37.01
Ki	22	0.9941	0.9921	0.9927	0.9916	8658	7421	7938	7163	−14.29	−8.32	−17.26	53.33	37.72	36.58
	27	0.9898	0.9861	0.989	0.9856	4117	3870	3942	3828	−6.00	−4.25	−7.03	60.61	32.55	31.71
	32	0.9794	0.97	0.9739	0.9694	1969	1809	1857	1801	−8.14	−5.69	−8.53	62.63	36.28	34.15
	37	0.9577	0.9373	0.9419	0.9378	934	896	898	884	−4.15	−3.85	−5.44	63.65	36.6	36.48
PS	22	0.9929	0.9919	0.9922	0.9918	11,340	9892	9926	9684	−12.76	−12.47	−14.6	52.81	29.84	27.85
	27	0.9874	0.9851	0.9841	0.9849	4450	4005	4062	3976	−9.99	−8.72	−10.65	59.94	34.35	33.23
	32	0.9762	0.9702	0.9752	0.9699	1877	1756	1835	1737	−6.45	−2.24	−7.44	63.87	36.47	36.91
	37	0.9545	0.9507	0.9533	0.9404	812	801	806	784	−1.32	−0.74	−3.4	68.58	39.44	39.26
BM	22	0.9965	0.9963	0.9957	0.9961	6348	5739	5809	5443	−9.60	−8.49	−14.26	51.26	27.06	25.15
	27	0.994	0.9932	0.9928	0.9929	2841	2630	2641	2629	−7.42	−7.04	−7.46	58.62	31.72	30.6
	32	0.9884	0.9876	0.9877	0.986	1397	1319	1321	1304	−5.58	−5.44	−6.7	64.36	35.68	34.09
	37	0.9758	0.9705	0.9722	0.9697	710	675	677	671	−4.98	−4.65	−5.41	68.81	38.49	37.35
RH	22	0.9958	0.9951	0.9954	0.9949	8905	8215	8124	8159	−7.75	−8.77	−8.38	44.94	25.75	23.71
	27	0.9864	0.9861	0.9855	0.9856	3637	3391	3472	3386	−6.76	−4.54	−6.89	50.2	30.23	29.28
	32	0.9708	0.9691	0.9696	0.9689	1574	1536	1503	1519	−2.44	−4.51	−3.53	54.04	35.41	34.15
	37	0.9624	0.9617	0.962	0.9609	733	699	702	686	−4.56	−4.23	−6.41	58.55	39.91	38.2
Average						-	-	-	-	**−7.01**	**−6.22**	**−8.34**	**56.93**	**33.22**	**32.07**

Ca Cactus, *BD* BasketballDrive, *Ki* Kimono, *PS* ParkScene, *BM* BQMall, *RH* RaceHorses

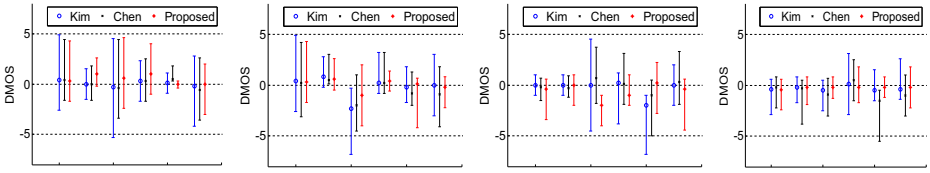


Fig. 9 the average DMOS comparison of Kim's PVC,Chen's PVC and proposed PVC with respect to HM15.0 at (**a**) QP22; (**b**) QP27; (**c**)QP32; (**d**)QP37 for *Cactus*, *BasketballDrive, Kimono, ParkScene, BQMall, RaceHorses* sequences

frame by four approaches which cost 50,408 bits, 46,800 bits, 46,256 bits, 44,624 bits, respectively. Obviously, their residual distribution can also further address the bitrate reduction in blue rectangles of Fig. 10a, b, c and d where the black region denotes no residue while others denotes exiting residue. It can be seen that the blue rectangle region in Fig. 10d has more black data than others, which implies less residual coefficients in the same red rectangle region with hardly perceive luma and chroma distortions.

6 Conclusion

In this paper, we have addressed problems of conventional PVC schemes and proposed a transform-domain multi-channel JND model in the YCbCr space, which can estimate more precisely JND thresholds in every channel for the different sized transform block to remove perceptual redundancy. Firstly, since most of JND models focus on the gray image or video, which results in JND thresholds calculated coarsely in the chroma channel, our transform-domain multi-channel JND model can solve this problem by the modeling method based on psychophysical experiments. Then we proposed a fast and HEVC-compatible perceptual video coding scheme using our proposed JND model.

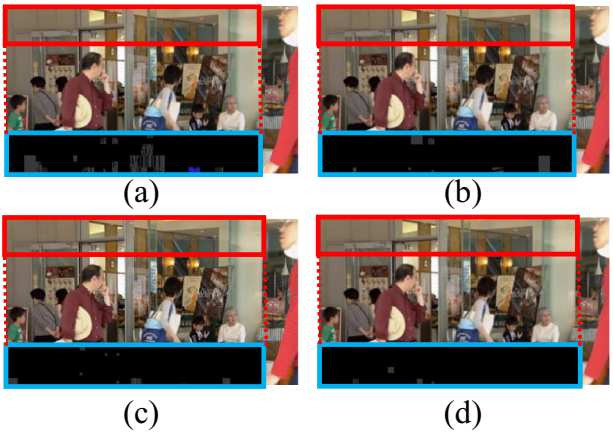


Fig. 10 Comparison of reconstructed 26th frames and their residual distribution at QP27: (**a**) the reconstructed frame by the original HM15.0(Cost 50,408 bits); (**b**) the reconstructed frame by Kim's PVC (Cost 46,800 bits); (**c**) the reconstructed frame by Chen's PVC (Cost 46,256 bits); (**d**) the reconstructed frame by the proposed PVC (Cost 44,624 bits)

To decreasing the computational complexity on the encoding side of our PVC scheme, we improved the frame-level method for evaluating the texture complexity to avoid duplication in the iterative TU partition process and proposed a multi-channel transform coefficients suppression method based on JND thresholds and QP ranges to reduce the encoding time. Extensive experimental results show the proposed PVC scheme yields significant bit saving of up to 25.91% and on average 9.42% with similar subjective quality under LD configuration, compared to the original HM15.0, and consistently outperforms two PVC schemes with much reduced bitrate and complexity overhead. In future work, the encoding complexity of our PVC scheme should be further decreased in the hardware codec.

Acknowledgments This work is partially supported by the National Key Research and Development Plan (Grant No.2016YFC0801001) and the NSFC Key Project (No. 61632001).

References

1. Bae S, Kim M (2014) A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images. IEEE Trans Image Process 23(8):3227–3240
2. Bae S, Kim J, Kim M (2016) HEVC-based Perceptually adaptive video coding using a DCT-based local distortion detection probability model. IEEE Trans Image Process 25(7):3343–3357
3. Bossen F, Common test conditions and software reference configurations, JCTVC-I1100, JCTVC, May 2012
4. Chang HW, Zhang QW, Wu QG, Gan Y (2015) Perceptual image quality assessment by independent feature detector. Neurocomputing 151:1142–1152
5. Chen ZZ, Guillemot C (2010) Perceptually-Friendly H.264/AVC Video Coding Based on Foveated JND Model. IEEE Trans Circuits Syst Vid Technol 20(6):806–819
6. Chen H, Hu R M, Hu J H, Wang Z Y (2010) Temporal color just noticeable distortion model and its application for video coding, in: Proceedings of IEEE ICME2010, pp. 713–718
7. Chou C-H, Li Y-C (1995) A Perceptually tuned Subband image coder based on the measure of just-noticeable-distortion profile. IEEE Trans Circuits Syst Vid Technol 5(6):467–476
8. Deng X, Xu M, Wang ZL (2013) A ROI-based Bit Allocation Scheme for HEVC towards Perceptual Conversational Video Coding, in: Proceedings of 2013 Sixth International Conference on Advanced Computational Intelligence (ICACI), pp. 19–21
9. Flynn D, Marpe D, Naccari M, Nguyen T, Rosewarne C, Sharman K, Sole J, Xu J (2016) Overview of the range extensions for the HEVC standard: Tools, profiles, and performance. IEEE Trans Circuits Syst Vid Technol 26(1):4–19
10. Guraya FFE, Alaya Cheikh F (2015) Neural networks based visual attention model for surveillance videos. Neurocomputing 149:1348–1359
11. Itti L, Koch C (2001) Computational modeling of visual attention. Nat Rev Neurosci 2:194–203
12. Jarsky T, Cembrowski M, Logan SM, Kath WL, Riecke H, Demb JB, Singer JH (2011) A synaptic mechanism for retinal adaptation to luminance and contrast. J Neurosci 31(30):11003–11015
13. Kim I-K High efficiency video coding (HEVC) test model 15 (HM15) encoder description, document JCTVC-Q1002 of ITU-T/ISO/IEC, JCTVC, Apr. 2014
14. Kim J, Bae S, Kim M (2015) An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels. IEEE Trans Circuits Syst Vid Technol 25(11):1786–1800
15. Li Z, Qin S, Itti L (2011) Visual attention guided bit allocation in video compression. Image Vis Comput 29(1):1–14
16. Lin Y-C, Lai J-C, Cheng H-C (2016) Coding unit partition prediction technique for fast video encoding in HEVC. Multimed Tool Appl 75(16):9861–9884
17. Liu T, Zheng N, Ding W, Yuan Z (2008) Video attention: learning to detect a salient object sequence, in: Proceedings of ICPR, pp.1–4.
18. Luo ZY, Song L, Zheng SB, Nam L (2013) H.264/Advanced Video Control Perceptual Optimization Coding Based on JND-Directed Coefficient Suppression. IEEE Trans Circuits Syst Vid Technol 23(6):935–948
19. Methodology for the subjective assessment of the quality of television pictures, ITU-R BT.500–11, 2002

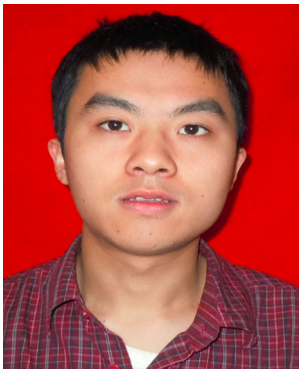
20. Naccari M, Pereira F (2011) Advanced H.264/AVC-Based Perceptual Video Coding: Architecture, Tools, and Assessment. IEEE Trans Circuits Syst Vid Technol 21(6):766–782
21. Peterson HA, Ahumada AJ, Watson AB, Improved detection model for DCT coefficient quantization, in: Proceedings of SPIE, Human Vision, Visual Processing, and Digital Display IV. 1993, 191–201
22. Robson J, Graham N (1981) Probability summation and regional variation in contrast sensitivity across the visual field. Vis Res 21(3):409–418
23. Sullivan GJ, Ohm JR, Han WJ, Wiegand T (2012) Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circuits Syst Vid Technol 22(12):1649–1668
24. Tsai D-S, Chen Y-C (2014) Visibility bounds for visual secret sharing based on JND theory. Multimed Tool Appl 70(3):1825–1836
25. Wei ZY, Ngan KN (2009) Spatio-temporal just noticeable distortion profile for Grey Scale image/video in DCT domain. IEEE Trans Circuits Syst Vid Technol 19(3):337–346
26. Wu HR, Rao KR (2005) Digital video image quality and perceptual coding. CRC Press, Boca Raton
27. Wu JJ, Lin WS, Shi GM, Wang XT, Fu L (2013) Pattern masking estimation in image with structural uncertainty. IEEE Transactions Image Processing 22(12):4892–4904
28. Xu M, Xu M, Wang ZL (2014) Region-of-interest based conversational HEVC coding with hierarchical perception model of face. IEEE J Select Topic Signal Process 8(3):475–489
29. Yan CG, Zhang YD, Xu JZ, Dai F, Li L, Dai QH, Wu F (2014) A highly Parallel framework for HEVC coding unit partitioning tree decision on many-core processors. IEEE Signal Process letts 21(5):573–576
30. Yan CG, Zhang YD, Xu JZ, Dai F, Zhang J, Dai QH, Wu F (2014) Efficient Parallel framework for HEVC motion estimation on many-core processors. IEEE Trans Circuits Syst Vid Technol 24(12):2077–2089
31. Yan CG, Zhang YD, Dai F, Wang X, Li L, Dai QH (2014) Parallel deblocking filter for HEVC on many-core processor. Electron Lett 50(5):367–368
32. Yan CG, Zhang YD, Dai F, Zhang J, Li L, Dai QH (2014) Efficient Parallel HEVC intra prediction on many-core processor. Electron Lett 50(5):805–806
33. Yang XK, Ling WS, Lu ZK, Ong EP, Yao SS (2005) Just noticeable distortion model and its applications in video coding. Signal Process Image Commun 20(7):662–680
34. Yang X, Lin W, Lu Z, Ong E, Yao S (2005) Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile. IEEE Trans Circuits Syst Vid Technol 15(6):742–752
35. Zeng HQ, Yang PAS, Ngan KN, Wang MH (2016) Perceptual sensitivity-based rate control method for high efficiency video coding, Multimed Tool Appl 75(17):10383–10596
36. Zhong GY, He HX, Qing LB (2015) Yue li, a fast inter-prediction algorithm for HEVC based on temporal and spatial correlation. Multimed Tool Appl 74(24):11023–11043
37. Zhong SH, Liu Y, Ng TY, Liu Y (2016) Perception-oriented video saliency detection via spatio-temporal attention analysis. Neurocomputing:1–11

Gang Wang is currently a Ph.D. candidate in the School of Computer Science and Engineering,Beihang University, Beijing, China. He received his M.S. degree from Zhejiang Sci-Tech University, Hangzhou, China, in 2012. His current research interests include multimedia compression, image/video processing, machine learning and hardware video codec design.

Yongfei Zhang received the B.S. degree in Electrical Engineering and the Ph.D. degree in Pattern Recognition and Intelligent Systems from Beihang University, Beijing, China, in 2005 and 2011 respectively. He is currently an Associate Professor with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, and the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing,China. He was a visiting research scholar in Video Processing and Networking Lab, University of Missouri, Columbia, U.S, from 2007 to 2009. His current research interests include image/video processing, compression and network transmission, and video surveillance.

Bo Li received the B.S. degree in computer science from Chongqing University, Chongqing, China, in 1986, the M.S. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1989, and the Ph.D. degree in computer science from Beihang University, Beijing, China, in 1993. He joined the School of Computer Science and Engineering, Beihang University, in 1993, where he has been a Full Professor since 1997. In 2002, he visited the University of Washington, Seattle, WA, USA, as a Senior Visiting Scholar for one year. He is currently the Director of the Digital Media Laboratory, School of Computer Science and Engineering, and the Director of the Professional Committee of Multimedia Technology of the China Computer Federation. He has published more than 100 conference proceedings and journal papers in diverse research fields, including digital video and image compression, video analysis and understanding, remote sensing image fusion, and embedded digital image processors.

Rui Fan received the B.S. degree in computer science and technology from Beijing University of Post and Telecommunication, Beijing, China, in 2011. He is currently a Ph.D. candidate in the School of Computer Science and Engineering,Beihang University, Beijing, China. His research interests include multimedia compression and transmission, image/video process and hardware optimization for video encoder.



Mingliang Zhou received the B.S. degree in applied the School of Computer Science, Sichuan University of Science and Engineering, Zigong, China, in 2008, and the M.S. degree from the School of Computer Science,Guizhou University, Guiyang, China, in 2011. He is currently pursuing the Ph.D degree in computer science and engineering from Beihang University, Beijing, China. His current research interests include video processing and high-efficiency video coding.