Received September 6, 2019, accepted October 9, 2019, date of publication October 21, 2019, date of current version November 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2948709

Perception-Based CTU Level Bit Allocation for Intra High Efficiency Video Coding

XIAOYAN LIU[1,2], YUN ZHANG[1], (Senior Member, IEEE), LINWEI ZHU[1], AND HUANHUA LIU[1,3]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China
[3]School of Computer Science and Engineering, Central South University, Changsha 410075, China

Corresponding author: Yun Zhang (yun.zhang@siat.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871312, in part by the Guangdong Natural Science Foundation for Distinguished Young Scholar under Grant 2016A030306022, in part by the Key Project for Guangdong Provincial Science and Technology Development under Grant 2017B010110014, in part by the Shenzhen International Collaborative Research Project under Grant GJHZ20170314155404913, in part by the Shenzhen Science and Technology Program under Grant JCYJ20170811160212033 and Grant JCYJ20180507183823045, in part by the Guangdong International Science and Technology Cooperative Research Project under Grant 2018A050506063, and in part by the Membership of Youth Innovation Promotion Association, Chinese Academy of Sciences, under Grant 2018392.

ABSTRACT The conventional rate control algorithm does not fully consider the characteristics of the Human Visual System (HVS). To address this problem, in this paper, a perception based Coding Tree Unit (CTU) level bit allocation algorithm for intra High Efficiency Video Coding (HEVC) is proposed, which aims to minimize the perceptual distortion of each CTU under a given bit rate constraint. Firstly, to better represent perceptual distortion, Perceptually Weighted Mean Squared Error (PWMSE) is adopted instead of Mean Squared Error (MSE). Then, the relationship between Rate (R) and Perceptually Weighted Distortion (D_p) is formulated by the proposed R-D_p model. Finally, the perceptual weighting factor derived from the R-D_p model is used to guide CTU level bit allocation. Experimental results show that the proposed algorithm achieves 6.27% bit rate reduction and 0.38 dB Bjøntegaard Delta Perceptually Weighted Peak Signal to Noise Ratio (BD-PWPSNR) gain on average. Under other seven quality evaluation metrics, the proposed algorithm achieves from 1.57% to 8.95% average bit rate reduction while maintaining the perceptual quality, which significantly outperforms the benchmark schemes. Moreover, the proposed perception-based CTU level bit allocation algorithm also maintains a high rate control accuracy, which reaches 99.997%.

INDEX TERMS Perceptual coding, bit allocation, rate control, high efficiency video coding.

I. INTRODUCTION

With the development of multimedia video technology, Ultra High Definition (UHD), Three-Dimensional (3D), and Virtual Reality (VR) videos are becoming more and more popular. In the foreseeable future, the demand of realistic visual experience will lead to a broader development of Wide Color Gamut (WCG), High Dynamic Range (HDR) and light field technology [1]. Simultaneously, it brings explosive growth in data volume as well, which challenges the multimedia video data storage and transmission. An efficient video coding standard is always desired for academic and industrial societies. H.265/High Efficiency Video Coding (HEVC) [2] standard has doubled the coding efficiency of H.264/Advanced Video Coding (AVC) [3]. This greatly improves the performance of video coding, but more coding gains are supposed to be achieved from the perceptual perspective. Perceptual coding can further remove the perceptual redundancy, and bring better compression efficiency and visual experience, which has been studied by more and more scholars. In video coding, one of the most important modules is rate control, which targets at effectively meeting the dynamic channel bandwidth. Generally, the process of rate control can be divided into two steps. The first one is bit allocation, which determines the optimal target number of bits in Group of Picture (GOP), frame and Coding Tree Unit (CTU) levels. The second one is the value of the Quantization Parameter (QP) determination with the R-Q or R-λ models.

In general, the rate control algorithm is developed according to the mathematical models of R-Q, R-λ, D-λ, etc. A rate control algorithm based on the R-Q model was proposed in [4]. Subsequently, Li *et al*. [5] established R-λ model to

The associate editor coordinating the review of this manuscript and approving it for publication was Byeungwoo Jeon.

determine the values of QP, which significantly improved the rate control accuracy and was incorporated into the HEVC test Model (HM) version 10.0 [6] to replace the R-Q model. However, this bit allocation algorithm was only utilized for non-intra frames. Thus, the proposals of JCTVC-M0257 [7] and JCTVC-M0036 [8] were proposed for intra frames. In JCTVC-M0257, intra frame bit allocation was guided by the linear correlation between the number of bits and the Sum of Absolute Transform Difference (SATD). Recently, a new GOP level bit allocation method was presented in [9] to match the HEVC GOP coding structure accurately. Guo *et al.* [10] proposed an effective frame-level bit allocation method to improve video coding performance and maintain high precision of bit rate control in HEVC. Li *et al.* [11] constructed the D-λ model and developed an optimal bit allocation algorithm.

To further exploit the redundancies in human perception distortion, many scholars began to study perceptual video coding [12]. There are many perceptual quality prediction models, such as Structural Similarity Index Metric (SSIM), Gradient Magnitude Similarity Deviation (GMSD), picture wise Just Noticeable Difference (JND) [13], etc. With SSIM, the game theory was employed [14] to guide the CTU level bit allocation for intra frame. Moreover, SSIM based rate control algorithms were proposed in [15] and [16]. Based on the gradient information, Yang *et al.* [17] utilized gradient magnitude similarity deviation [18] in the CTU level bit allocation. A gradient based R-λ model for the intra frame rate control was proposed by Wang *et al.* [19]. In addition, a bit allocation algorithm based on saliency was also proposed in [20]. Sun *et al.* [21] proposed a new perception-based intra frame coding optimization algorithm that adaptively updated bit allocation for the salient and non-salient regions. Texture complexity and motion information were used to measure the temporal-spatial masking effect, which were used to guide the rate control algorithm in [22]. Although these mentioned metrics could remove the perceptual redundancy to some extent, the performance can be further improved for the compressed image or video from the perceptual quality perspective.

The algorithm of Perceptually Weighted Mean Squared Error (PWMSE) [23] is the state-of-the-art perceptual quality evaluation metric, which is more consistent with Human Visual System (HVS). The metric of PWMSE used a randomness map to measure the masking effect of HVS and simulate the process of the initial part of HVS to remove the imperceptible signal [23], which can better capture the characteristics of HVS. Since the consumer of the video stream is the end-user, the perceptual characteristics should be fully considered. So we employ the PWMSE quality metric in this paper. Therefore, a perception based CTU level bit allocation algorithm for intra High Efficiency Video Coding (HEVC) is proposed.

The major contributions of this paper are

1) A novel CTU-level bit allocation method is proposed, in which the conventional distortions such as MSE, SAD, are replaced by a perceptual distortion to reflect the real human perceived quality. Then, the associated perceptual weighting factor is derived and utilized in HEVC to guide the bit allocation in CTU level for rate control.
2) Four models of Rate (R) and Perceptually Weighted Distortion (D_p) are formulated in a statistical manner, and two of them are separately incorporated into the video codec and compared. In addition, two parameter estimation methods are proposed for model parameters determination.
3) In addition to PWMSE, seven mainstream perceptual metrics are adopted for coding performance evaluation, which illustrate that the proposed algorithm is efficient.

The remainder of this paper is organized as follows. In Section II, the problem and motivations are presented. The proposed perception-based rate and distortion models are illustrated in Section III. In Section IV, two R-D_p models are adopted to derive the perceptual weighting factor of each CTU. In Section V, the perception-based CTU level bit allocation algorithm for HEVC is described in detail. In Section VI, the experimental results and analyses are presented. Conclusions are then drawn in Section VII.

II. PROBLEM AND MOTIVATIONS

At present, the intra frame-based CTU level bit allocation algorithm utilizes the MSE or Mean Absolute Difference (MAD) of the co-located CTU as the weight for bit allocation, but this method measures pixel by pixel difference between original and reconstructed images, which does not fully consider the perception characteristic of HVS. From the subjective perspective, there is a model of visual masking effect for the human eye [21]. As shown in Fig. 1, although the values of MSE are identical in Figs. 1(b) and (c), the visual qualities of these two images are different. The visual quality of Fig. 1(c) is significantly better than that of Fig. 1(b). Because of the distortion in Fig. 1(c) is mainly located in the background of trees and weeds, the distortion perceived by the human eye is relatively small; while the distortion in Fig. 1(b) is mainly located in the foreground of the person and some flat areas. These areas have weak masking effect and are sensitive to distortion, so humans can easily detect the distortion. The perceived image qualities vary with the video contents and regions although they have the same MSE.

Inspired by this observation, a perception-based CTU level bit allocation algorithm is proposed in this paper for HEVC in the intra coding scenario. The coding resource of bits is supposed to be allocated according to the masking effect of HVS. A smaller number of bits will be allocated to the coding unit with significant human eye masking effect, *i.e.,* less sensitive, and vice versa.

III. PERCEPTION-BASED RATE AND DISTORTION MODEL

A. PERCEPTUAL DISTORTION MEASUREMENT

There are many existing perceptual image quality metrics, such as SSIM, MS-SSIM, VIF, GSMD, FSIM, and VSI, etc. The perceptual quality model PWMSE proposed in [23],

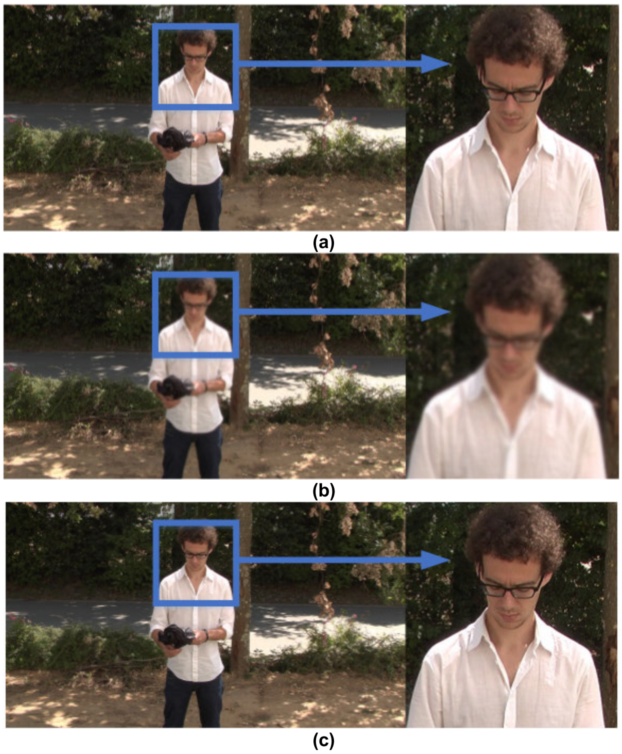

FIGURE 1. Different perceptual qualities under the same value of MSE. (a) Original image, MSE = 0. (b) The distortion located at foreground, MSE = 48. (c) The distortion located at background, MSE = 48.

which captures the perceptual characteristics of the HVS and has more consistency with human perception.

The PWMSE is adopted to represent the perceptual distortion with the randomness value in this paper. In a randomness map, the larger the random value, the stronger the masking effect, the less distortion can be perceived by the HVS and vice versa. The process of the visual signal received by the HVS is represented with a linear transfer function, many researchers treated the Contrast Sensitivity Function (CSF) as the spatial Modulation Transfer Function (MTF), and used it to define characteristics of initial processing in HVS. Hence, perceptual distortion $\Delta\mathbf{D_F}$ [23] is expressed as

$$\Delta\mathbf{D_F} = \mathbf{F} * (\mathbf{I} - \mathbf{I_d}), \tag{1}$$

where $\mathbf{I}$ is the original image, $\mathbf{I_d}$ is the distorted image. The symbol * represents the convolutional operation, and the function $\mathbf{F}$ is the spatial low-pass filter of the CSF model [23], whose element F(u, v) can be presented by

$$F(u, v) = \left(\alpha + \beta\xi_{u,v}\right) e^{-\gamma\xi_{u,v}}, \tag{2}$$

where $\xi_{u,v}$ is the spatial frequency at (u, v) and α, β, γ are constant parameters. PWMSE is calculated in [23] as

$$PWMSE = \ln\left(\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\Delta D_F(i,j)^2 \times S(i,j)\right), \tag{3}$$

where $\Delta D_F(i, j)$ is the element of perceptual distortion $\Delta\mathbf{D_F}$ at (*i*, *j*) position, *S*(*i*, *j*) is the randomness map value at (*i*, *j*) position. *W* and *H* are the width and height of the image, respectively. We use PWMSE to measure the distortion of the reconstructed videos. It is reported in [23] that the performance of PWMSE is much higher than the conventional perceptual quality evaluation metrics. In addition, many coding algorithms in HEVC encoder, such as motion estimation, rate control, intra and inter prediction, were all developed based on the MSE, it is easier and more compatible to integrate the PWMSE to the encoder than other perceptual metrics. So we choose this PWMSE as the distortion metric in HEVC bit allocation.

B. RATE-PERCEPTUALLY WEIGHTED DISTORTION MODEL

Since the optimization objective, i.e., the distortion measurement, in bit allocation of HEVC has been changed, and the corresponding *R-D* model should be changed accordingly. So we need to establish the new relationship between Rate (*R*) and Perceptual Weighted Distortion (D_P) noted as *R*-D_P model. As we know, in the traditional *R-D* model, the hyperbolic function has been widely used when the distortion metric is defined as MSE. For *R*-D_p model, we also consider the hyperbolic function [24] as

$$D_p = \alpha \times R^{-\beta}, \tag{4}$$

where α and β are model parameters, which are related to video content. In addition, the exponential function [24] based model is also considered in modeling the *R*-D_p relationship, which can be presented as

$$D_p = c \times e^{k \times R}, \tag{5}$$

where c and k are parameters, which are related to video content. We also consider the following Simple Polynomial function based Model (SPM) as the *R*-D_p model

$$D_p = a + \frac{b}{R}, \tag{6}$$

where a and b are model parameters related to video content. Besides the above functions, we also consider the following Complex Polynomial function based Model (CPM)

$$D_p = m \times R^n + l, \tag{7}$$

where m, n and l are model parameters related to video content. We perform the experiments to compare the four models. We use the default All Intra (AI) configuration in the platform of HM-16.7 [25], and then fit the *R*-D_p curve, as shown in Fig. 2. Six sequences are selected, including "Kimono", "Johnny", "PartyScene", "RaceHorses", "PeopleOnStreet", and "BasketballPass". 150 frames are encoded for each sequence under fixed QPs (QP = 22, 27, 32, 37). The D_p is represented by the PWMSE of the luminance component, which is calculated by Eq. (3). The *R* is represented by *bpp* (bit per pixel), which can be calculated by

$$bpp = \frac{R}{f \times H \times W}, \tag{8}$$

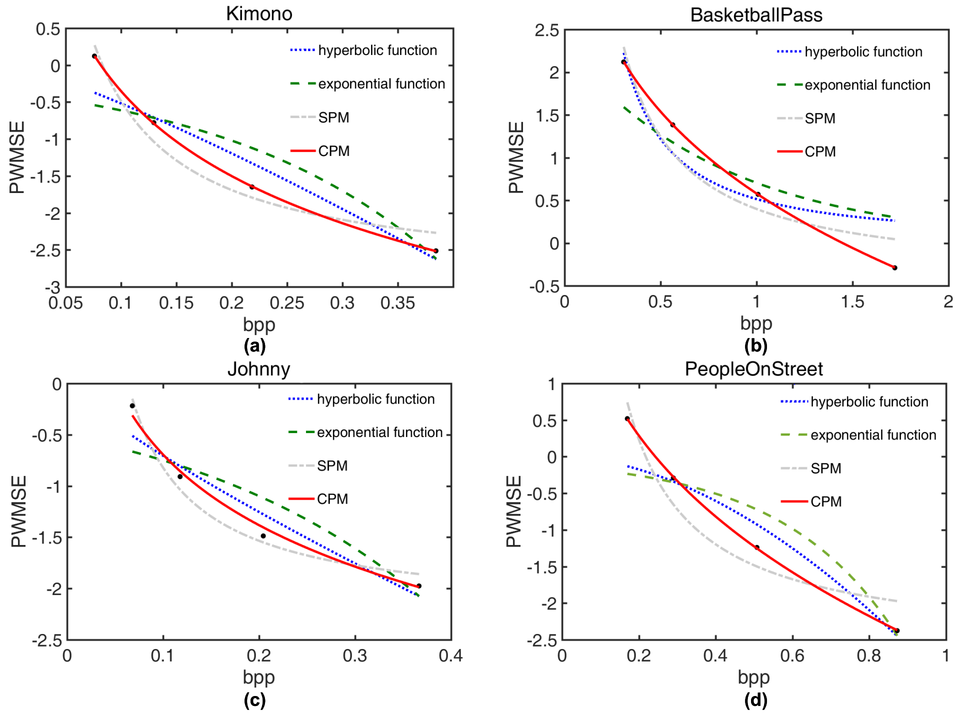

FIGURE 2. The R-D_p fitting curves of four sequences with four models. (a) Kimono. (b) BasketballPass. (c) Johnny. (d) PeopleOnStreet.

where f is the frame rate, W and H are the width and height of the image, respectively. Fig. 2 shows the fitting curves of four sequences, including "Kimono", "Johnny", "BasketballPass", and "PeopleOnStreet". The dots are real data. The blue, green, red, and gray lines indicate the hyperbolic function, the exponential function, SPM, and CPM, respectively. From the Fig. 2, we can see that the CPM in Eq. (7) outperforms the other models for most cases within the practical bitrate range. We also use R-squared to measure the accuracy of these four models. The results are shown in Table 1. The average fitting accuracy of the hyperbolic function in Eq. (4), the exponential function in Eq. (5), the SPM in Eq. (6), and the CPM in Eq. (7) are 0.89, 0.85, 0.92, and 0.99, respectively. The fitting accuracy of the CPM in Eq. (7) is the highest one among the fitting schemes. In addition, the accuracy of SPM is in the second place, which is much better than the Hyperbolic and exponential models. It is a little inferior to CPM, however, it has a simpler form in calculating the optimal bit allocation parameters. Thus, the SPM and CPM are both considered and used in this work.

IV. PERCEPTUAL WEIGHTING FACTORS DETERMINATION FOR CTUS

Before performing bit allocation, we need to get the perceptual weighting factor of each CTU to guide the CTU level bit allocation. In this section, according to the SPM and CPM, the derivations of the perceptual weighting factors for the two models are presented accordingly.

TABLE 1. Curve fitting accuracy comparison under different functions based R-D model.

Sequences	Fitting Accuracy R^2			
	Hyperbolic (Eq.4)	Exponential (Eq.5)	SPM (Eq.6)	CPM (Eq.7)
Kimono	0.91	0.81	0.96	0.98
Johnny	0.91	0.79	0.98	0.98
PartyScene	0.95	0.81	0.87	1.00
PeopleOnStreet	0.89	0.82	0.91	1.00
RaceHorses	0.85	0.94	0.87	1.00
BasketballPass	0.87	0.94	0.91	1.00
Average	**0.89**	**0.85**	**0.92**	**0.99**

A. PERCEPTUAL WEIGHTING FACTOR DETERMINATION FOR SIMPLE POLYNOMIAL FUNCTION BASED MODEL (SPM)

The goal of video coding optimization is to minimize distortion at a certain bit rate. Similarly, the goal of perceptual coding optimization is to minimize perceptually weighted distortion at a certain bit rate. The minimization of perceptually weighted distortion in a frame can be formulated as

$$\min \sum_{i=1}^{N} D_{p_i} \quad \text{s.t.} \sum_{i=1}^{N} R_i \leq R_{pic}, \tag{9}$$

where R_{pic} is the target bit allocated to a frame, R_i is the bit allocated to the i-th CTU, and D_{p_i} is the i-th CTU perceptually weighted distortion calculated by PWMSE. We take the Lagrangian multiplier algorithm to convert this constrained problem into an unconstrained problem, which can be presented as

$$J=\sum_{i=1}^{N}D_{p_i}+\lambda\left(\sum_{i=1}^{N}R_i-R_{pic}\right), \tag{10}$$

where λ is the Lagrange multiplier. Then we solve the unconstrained problem by taking the partial derivatives of R_i and λ in Eq. (10) respectively, and set them to zero

$$\begin{cases}\dfrac{\partial J}{\partial R_i}=\dfrac{\partial D_{p_i}}{\partial R_i}+\lambda=0\\ \dfrac{\partial J}{\partial \lambda}=\sum\limits_{i=1}^{N}R_i-R_{pic}=0.\end{cases} \tag{11}$$

Then according to the relationship between R and D_p, we apply Eq. (6) into Eq. (11) and obtain

$$\begin{cases}\dfrac{\partial J}{\partial R_i}=-\dfrac{b_i}{R_i^2}+\lambda=0\\ \dfrac{\partial J}{\partial \lambda}=\sum\limits_{i=1}^{N}R_i-R_{pic}=0.\end{cases} \tag{12}$$

Solving the Eq. (12), we obtain R_i

$$R_i=\frac{\sqrt{b_i}}{\sum\limits_{j=1}^{N}\sqrt{b_j}}\cdot R_{pic}, \tag{13}$$

where $\sqrt{b_i}$ is the content-related perceptual weight for the i-th CTU. Let $\omega_{CTU}^{P1}(i)$ be the perceptual weighting factor of the i-th CTU, and it is defined by

$$\omega_{CTU}^{P1}(i)=\frac{\sqrt{b_i}}{\sum\limits_{j=1}^{N}\sqrt{b_j}}. \tag{14}$$

According to Eq. (14), once b is given, the corresponding perceptual weighting factor can be obtained. In the traditional model parameter solution, the parameters are initially defined and then updated. In this paper we present a new adaptive parameter solving method. We consider the problem of solving the SPM coefficient b as a problem of solving the linear regression model coefficient. Eq. (6) can be rewritten as

$$D_p=a+b\times R', \tag{15}$$

where $R'=\frac{1}{R}$. We solve linear regression model Eq. (15) by least squares method, and then we get a general solution of the coefficient

$$b=\frac{l_{xy}}{l_{xx}}, \tag{16}$$

where

$$\begin{cases}l_{xy}=\sum\limits_{i=1}^{n}(R_i'-\overline{R'})(D_{p_i}-\overline{D_p})\\ l_{xx}=\sum\limits_{i=1}^{n}(R_i'-\overline{R'})^2\\ \overline{R'}=\dfrac{1}{n}\sum\limits_{i=1}^{n}R_i'\\ \overline{D_p}=\dfrac{1}{n}\sum\limits_{i=1}^{n}D_{p_i}.\end{cases} \tag{17}$$

From the above process, the perceptual weight $\sqrt{b_i}$ of each CTU can be adaptively obtained according to the video content, thereby obtaining the perceptual weighting factor $\omega_{CTU}^{P1}(i)$ of each CTU. In the experiment, n is set as 16, and in order to prevent the bits gap between different CTUs, we empirically clip the perceptual weight $\sqrt{b_i}$ to [6.0, 7.0].

B. PERCEPTUAL WEIGHTING FACTOR DETERMINATION FOR COMPLEX POLYNOMIAL FUNCTION BASED MODEL (CPM)

Similar with the process of Eq. (9) to Eq. (11), we apply the CPM model of Eq. (7) into Eq. (11) and obtain

$$\begin{cases}\dfrac{\partial J}{\partial R_i}=n_i\times m_i\times R_i^{n-1}+\lambda=0\\ \dfrac{\partial J}{\partial \lambda}=\sum\limits_{i=1}^{N}R_i-R_{pic}=0.\end{cases} \tag{18}$$

Different from the SPM, Eq. (18) in the solution of the CPM cannot be explicitly expressed as that of the SPM in Eq. (13). That is to say, the relationship between the number of bits allocated by each CTU and the total number of bits cannot be explicitly formed. Thereby the perceptual weighting factor similar to Eq. (14) cannot be obtained directly. Therefore, we obtain the corresponding number of bits of the CTU by directly solving the model parameters. In the case that the CPM model coefficients n and m are known, we use the idea in paper [26] to solve λ. The specific process is as follows, Eq. (18) can also be expressed as:

$$\begin{cases}R_i=(-\dfrac{\lambda}{m_i\times n_i})^{\frac{1}{n_i-1}}\\ \sum\limits_{i=1}^{N}R_i-R_{pic}=0.\end{cases} \tag{19}$$

Let $\omega_{CTU}^{P2}(i)$ be the perceptual weighting factor of the i-th CTU, and it is defined by

$$\omega_{CTU}^{P2}(i)=(-\frac{\lambda}{m_i\cdot n_i})^{\frac{1}{n_i-1}}\times N^2, \tag{20}$$

where N^2 is the size of each CTU. According to Eq. (20), once m, n, and λ are given, the corresponding perceptual weighting factor can be obtained. For nonlinear regression model Eq. (7), take the logarithm on both sides and obtain

$$\ln(D_p-l)=\ln(m\times R^n), \tag{21}$$

convert the nonlinear regression model in Eq. (7) to a linear regression model, Eq. (21) can be rewritten as

$$D_p^{'} = m^{'} + n \times R^*, \tag{22}$$

where $D_p^{'} = \ln(D_p - l)$, $R^* = \ln R$, and $m^{'} = \ln m$. Before solving the linear regression model Eq. (22), we choose three points (R_1, D_{p1}), (R_2, D_{p2}), (R_3, D_{p3}), where the third point meets $R_3 \approx R_1 \times R_2$, substitute the three points into Eq. (7), and obtain

$$\begin{cases} D_{p_1} - l = m \times R_1^n \\ D_{p_2} - l = m \times R_2^n \\ D_{p_3} - l = m \cdot (R_1 \times R_2)^n, \end{cases} \tag{23}$$

then we obtain

$$l = \frac{D_{p_1} \times D_{p_2} - D_{p_3}^2}{D_{p_1} + D_{p_2} - 2D_{p_3}}. \tag{24}$$

Solve the linear regression model in Eq. (22) by least squares method, and then get a general solution of the coefficients

$$\begin{cases} n = \dfrac{h_{xy}}{h_{xx}} \\ m^{'} = \overline{D^{'}} - n\overline{R^*}, \end{cases} \tag{25}$$

where

$$\begin{cases} h_{xy} = \sum\limits_{i=1}^{n} (R_i^* - \overline{R^*})(D_{p_i}^{'} - \overline{D_p^{'}}) \\ h_{xx} = \sum\limits_{i=1}^{n} (R_i^* - \overline{R^*})^2 \\ \overline{R^*} = \dfrac{1}{n} \sum\limits_{i=1}^{n} R_i^* \\ \overline{D^{'}} = \dfrac{1}{n} \sum\limits_{i=1}^{n} D_{p_i}^{'}. \end{cases} \tag{26}$$

The solution of λ for Eq. (19) can be calculated through the bisection method, which is described in detail in Algorithm 1. The bisection method is used to adjust λ. After the solution of R_i in $R_i = (-\frac{\lambda_j}{m_i \times n_i})^{\frac{1}{n_i - 1}}$ is obtained, substitute this R_i into $D_j = \sum\limits_{i=1}^{N} R_i - R_{pic}$, and then judge whether the summation of the allocated rate can satisfy the given rate constraint or not. If the allocated bits do not satisfy the rate constraint, the value of λ will be adjusted, and the solutions for Ri will be re-calculated until their summation satisfies Eq. (19). From the above process, the perceptual weighting factor $\omega_{CTU}^{P2}(i)$ of each CTU can be adaptively obtained according to the video content. In the experiment, we empirically normalize the perceptual weighting factor to [6.0, 7.0].

V. VISUAL PERCEPTION-BASED CTU LEVEL BIT ALLOCATION

In this section, we perform CTU level bit allocation algorithm for intra frame. The key idea of CTU level bit allocation is to allocate the remaining bits to the remaining CTUs.

Algorithm 1 Solving λ in Eq. (19) Using the Bisection Method

1. **Initialization:** $\lambda_j = 0$, $\lambda_{max} = 1000$, $\lambda_{min} = 0$, $tol = 0.0001$
2. **Given equation**: $R_i = (-\frac{\lambda_j}{m_i \times n_i})^{\frac{1}{n_i - 1}}$
3. **Input:** Initial value of λ_j, Initial iterator $j = 1$, Initial difference $D_0 = 0$
4. **Repeat process**
5. Calculate the difference $D_j = \sum\limits_{i=1}^{N} R_i - R_{pic}$
6. If $D_j > 0$ then
7. $\lambda_{min} \leftarrow \lambda_j$, $\lambda_j \leftarrow (\lambda_j + \lambda_{max})/2$
8. else
9. $\lambda_{max} \leftarrow \lambda_j$, $\lambda_j \leftarrow (\lambda_j + \lambda_{min})/2$
10. update the iterator: $j \leftarrow j + 1$
11. **Until** $D_j < tol$
12. **Output**: Final value of λ_j

We use the SATD calculated by each CTU as the initial bit weight for the current CTU. This initial bit weight is used to calculate the estimated remaining bits. We use the remaining actual bits in the sliding window to compensate the estimated remaining bits. Then, we use the perceptual weighting factor to calculate the optimal bits allocated for each CTU. The specific process is as follows. The initial bits allocated to each CTU is calculated by

$$R_{CTU}(i) = \frac{C_{CTU}(i)}{\sum\limits_{j=1}^{N} C_{CTU}(j)} \times R_{pic}, \tag{27}$$

where $C_{CTU}(i)$ is the SATD value of the i-th CTU, which is the initial bit weight. R_{pic} is the target bit allocated to a frame. $R_{CTU}(i)$ is the initial bit allocated to the i-th CTU. N is the number of CTUs in a frame. The remaining bits R_{CTU}^E are calculated by

$$R_{CTU}^E(i) = \frac{\sum\limits_{k=i}^{N} C_{CTU}(k)}{\sum\limits_{j=1}^{N} C_{CTU}(j)} \times R_{pic}. \tag{28}$$

Then, the actual remaining bits are used to adjust the estimated remaining bits R_{CTU}^O as

$$R_{CTU}^O(i) = R_{CTU}^A(i) + \frac{\left(R_{CTU}^A(i) - R_{CTU}^E(i)\right) \times (N - i)}{SW}, \tag{29}$$

where R_{CTU}^A is the actual remaining bits, i is the number of the encoded CTUs, R_{CTU}^E is the estimated remaining bits, which is calculated by Eq. (28). SW is the size of a smooth window, which is used to make the bitrate change smoother. SW in our experiment is set as 4, which follows the literature [5].

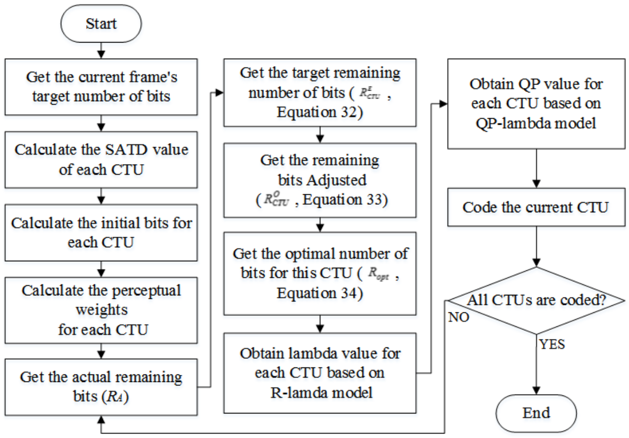

FIGURE 3. The flowchart of the perceptually weighted CTU level bit allocation algorithm.

The perceptual weight of each CTU is used to obtain the optimal bits allocated to each CTU

$$R_{opt}(i) = \frac{\omega^{P}_{CTU}(i)}{\sum_{j=i}^{N} \omega^{P}_{CTU}(j)} \times R^{O}_{CTU}(i), \tag{30}$$

where the perceptual weight $\omega^{P}_{CTU}(i)$ is calculated by Eq. (14) or Eq. (20), and the remaining number of bits to be adjusted R^{O}_{CTU} will be calculated by Eq. (29).

After the bit allocation is performed, the QP is determined by the two relational models of R-λ [5], [24] and QP-λ [27]. The flowchart of the perceptually weighted CTU level bit allocation algorithm is shown in Fig. 3. The SATD value of each CTU is firstly calculated and the ininitial bits of each CTU are allocated. Secondly, the perceptual weighting factor of each CTU is determined. And then the actual remaining number of bits is adopted to update the optimal number of bits. Next, the perceptual weighting factor is used to obtain the optimal number of bits per CTU. Finally, the R-λ and QP-λ models are employed to obtain the QP value of each CTU.

VI. EXPERIMENTAL RESULTS AND ANALYSES

A. EXPERIMENTAL SETUP

To testify the effectiveness of the proposed algorithm, JCTVC-K0103 [5], JCTVC-M0257 [7], HM-16.7 default anchor algorithm and Wang-SPL [19] are adopted as the benchmark schemes. These algorithms are all implemented on the HEVC reference platform HM-16.7. Experimental encoding settings are shown in Table 2, and the rest parameters are default. The target bitrate of each sequence is collected under the platform of HM-16.7 reference software with fixed QPs (QP = 22, 27, 32, 37), following the Common Test Conditions (CTC). Twenty four video sequences with various contents and characteristics are tested. Table 3 shows the properties of the test sequences including resolution, frame rate, and the number of encoded frames. 150 frames of each sequence are encoded.

TABLE 2. Experimental settings.

Profile	Main
Structures	All Intra
CTU Size/Depth	64\4
Intra Period	1
Motion Search Range	64
RDOQ\RDOQTS	1\1
Entropy Coding	CABAC
Rate Control	1
LCULevelRateControl	1
RCLCUSeparateModel	1

TABLE 3. Summary of the testing sequences.

Sequences	Resolution	Frame rate	Frames
Traffic	2560×1600	30	150
PeopleOnStreet		30	150
SteamLocomotiveTrain		60	150
NebutaFestival		60	150
Kimono	1920×1080	24	150
BasketballDrive		50	150
BQTerrace		60	150
Cactus		50	150
ParkScene		24	150
Tennis		24	150
RaceHorses	832×480	30	150
BQMall		60	150
PartyScene		50	150
BasketballDrill		50	150
Flowervase		30	150
Keiba		30	150
Mobisode2		30	150
BasketballPass	418×240	50	150
BlowingBubbles		50	150
BQSquare		60	150
Vidyo3	1280×720	60	150
KristenAndSara		60	150
Johnny		60	150
FourPeople		60	150

B. COMPARISON OF R-D PERFORMANCE USING PWMSE METRIC

We convert PWMSE [23] to Perceptual Weighted PSNR (PWPSNR) to evaluate the quality of compressed video,

$$PWPSNR = 10log_{10}\left(\frac{255^2}{e^{PWMSE}}\right). \tag{31}$$

Then, Bjøntegaard Delta Perceptual Weighted Peak Signal to Noise Ratio (BD-PWPSNR) and Bjøntegaard Delta Bit Rate (BD-BR) [28] with respect to PWPSNR are adopted as the R-D performance evaluation metrics. The negative value of BD-BR with respect to PWPSNR, denoted as

TABLE 4. R-D Performance comparison with JCTVC-K0103 on HM-16.7 when quality is measured with PW-PSNR.

Sequence	M0257 [7]		Wang-SPL [19]		HM-16.7 [25]		Proposed SPM		Proposed CPM	
	BD-BR (PWPSNR)	BD-PWPSNR	BD-BR (PWPSNR)	BD-PWPSNR	BD-BR (PWPSNR)	BD-PWPSNR	BD-BR (PWPSNR)	BD-PWPSNR	BD-BR (PWPSNR)	BD-PWPSNR
BasketballDrill	-0.33%	0.02	3.95%	-0.27	-4.85%	0.34	-5.01%	0.35	-4.92%	0.34
BasketballDrive	-0.09%	0.01	0.92%	-0.05	-7.32%	0.39	-9.96%	0.54	-10.45%	0.56
BQMall	-3.36%	0.16	-4.00%	0.20	-3.68%	0.20	-6.67%	0.36	-6.57%	0.35
BQTerrace	-0.92%	0.05	-6.74%	0.40	-7.95%	0.47	-10.38%	0.62	-14.84%	0.91
Cactus	-0.38%	0.02	-2.25%	0.13	-10.06%	0.58	-11.13%	0.64	-13.61%	0.80
FourPeople	-0.03%	0.00	-1.31%	0.11	-6.84%	0.50	-7.54%	0.55	-7.48%	0.55
Johnny	-0.09%	0.00	0.48%	-0.03	-1.54%	0.07	-3.60%	0.16	-3.77%	0.17
Kimono	-0.20%	0.01	1.89%	-0.13	-1.40%	0.10	-1.01%	0.07	-1.71%	0.12
KristenAndSara	0.02%	0.00	1.28%	-0.07	-2.77%	0.17	-6.02%	0.37	-5.92%	0.36
ParkScene	-0.23%	0.01	-2.30%	0.13	-8.42%	0.52	-9.49%	0.59	-10.10%	0.63
PartyScene	-5.61%	0.21	-3.70%	0.12	-9.33%	0.40	-7.73%	0.36	-7.60%	0.36
PeopleOnStreet	-0.28%	0.02	1.08%	-0.09	-1.60%	0.12	-1.88%	0.14	-2.61%	0.20
RaceHorses	-4.99%	0.22	-8.57%	0.49	-8.93%	0.52	-10.10%	0.61	-10.15%	0.62
Traffic	-0.28%	0.02	-3.53%	0.26	-5.07%	0.36	-5.86%	0.42	-6.31%	0.47
BasketballPass	-0.50%	0.03	2.54%	-0.16	-1.62%	0.10	-0.40%	0.02	-0.36%	0.02
BlowingBubbles	-2.22%	0.12	-1.98%	0.11	-3.17%	0.18	-2.04%	0.11	-2.02%	0.11
BQSquare	-1.64%	0.11	-0.88%	0.05	-2.26%	0.15	-1.57%	0.10	-1.46%	0.09
Tennis	-0.59%	0.03	6.67%	-0.34	-2.91%	0.16	-4.73%	0.26	-6.68%	0.37
Vidyo3	-0.06%	0.00	-0.04%	0.00	-4.31%	0.31	-6.89%	0.49	-6.89%	0.49
Flowervase	-0.04%	0.00	3.92%	-0.27	-8.17%	0.58	-9.62%	0.69	-9.60%	0.69
Keiba	-1.30%	0.08	-0.67%	0.04	-1.45%	0.09	-3.56%	0.23	-3.63%	0.23
Mobisode2	-4.48%	0.19	0.80%	-0.04	-5.66%	0.24	-6.31%	0.27	-5.47%	0.24
SteamLocomotiveTrain	-1.31%	0.08	1.90%	-0.10	-6.38%	0.37	-6.96%	0.41	-7.05%	0.42
NebutaFestival	-1.08%	0.09	1.53%	-0.16	-0.32%	0.02	-0.38%	0.02	-1.38%	0.11
Average	**-1.25%**	**0.06**	**-0.38%**	**0.01**	**-4.83%**	**0.29**	**-5.78%**	**0.35**	**-6.27%**	**0.38**

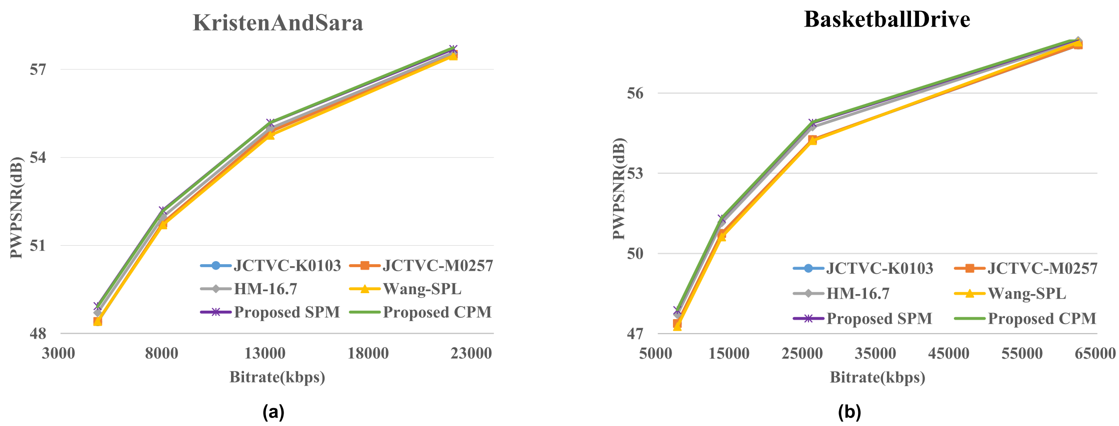

FIGURE 4. The R-PWPSNR curve of two test sequences. (a) KristenAndSara. (b) BasketballDrive.

BD-BR (PWPSNR), and the positive value of BD-PWPSNR indicates coding gain and vice versa.

In Table 4, the BD-BR and BD-PWPNSR are calculated when compared with benchmark rate control scheme JCTVC-K0103. We can observe that JCTVC-M0257 reduce the bit rate from −0.02% to 5.61% and 1.25% on average. Wang-SPL and the HM16.7 achieve −0.38% and −4.83% in terms of BD-BR on average, respectively. As compared with the JCTVC-K0103, the proposed SPM algorithm saves 5.78% bit rate and achieves 0.35 dB coding gain. For the best case, the bit rate reduction reaches 11.13% and the value of BD-PWPSNR is 0.64dB, which is much better than the benchmark schemes. Compared with the JCTVC-K0103, the proposed CPM algorithm saves 6.27% on average bit rate and achieves 0.38 dB coding gain. For the best case, bit rate reduction is 14.84% and the value of BD-PWPSNR is 0.91dB, which is much better than the benchmark schemes and the proposed SPM scheme. The performance of our proposed two perception based bit allocation algorithms can achieve better performances in terms of BD-BR and BD-PWPSNR than those of other algorithms. Therefore, the CPM model performs better and is more capable of characterizing the $R\text{-}D_p$ model. Moreover, the R-D curves are illustrated in Fig. 4. It can be observed that the R-D curve of the proposed algorithms are located above of JCTVC-K0103, JCTVC-M0257, HM-16.7 default anchor, and Wang-SPL, which indicates that the proposed algorithm is the best. As shown in Fig. 5, the image quality of each

TABLE 5. R-D performance comparisons when quality is measured with PWMSE_VIDEO [31] and IFC [27].

Sequence	M0257 [7]		Wang-SPL [19]		HM-16.7 [25]		Proposed SPM		Proposed CPM	
	BD-BR (PWMSE_VIDEO)	BD-BR (IFC)	BD-BR (PWMSE_VIDEO)	BD-BR (IFC)	BD-BR (PWMSE_VIDEO)	BD-BR (IFC)	BD-BR (PWMSE_VIDEO)	BD-BR (IFC)	BD-BR (PWMSE_VIDEO)	BD-BR (IFC)
BasketballDrill	-0.30%	-0.14%	2.16%	0.12%	-3.60%	-2.69%	-3.20%	-2.63%	-2.98%	-2.48%
BasketballDrive	0.07%	-0.04%	1.65%	0.35%	-2.49%	-2.29%	-3.33%	-2.81%	-3.17%	-2.72%
BQMall	-2.89%	-1.67%	-6.58%	-2.59%	-4.41%	-1.80%	-6.39%	-2.38%	-6.35%	-2.31%
BQTerrace	-0.87%	-0.42%	-3.01%	-0.71%	-4.01%	-1.98%	-4.83%	-2.18%	-6.48%	-2.93%
Cactus	-0.44%	-0.24%	2.12%	0.66%	-3.36%	-1.48%	-4.01%	-1.60%	-4.97%	-2.03%
FourPeople	-0.04%	-0.01%	-1.51%	0.31%	-5.48%	-1.93%	-5.72%	-1.70%	-5.60%	-1.62%
Johnny	-0.10%	-0.08%	-0.64%	1.10%	-1.70%	0.17%	-3.67%	-0.25%	-3.66%	-0.19%
Kimono	0.10%	-0.11%	0.16%	0.82%	-1.42%	-0.53%	-1.83%	-0.66%	-1.74%	-0.67%
KristenAndSara	-0.06%	0.01%	-0.68%	1.01%	-2.40%	-0.28%	-4.71%	-0.86%	-4.67%	-0.78%
ParkScene	-0.21%	-0.07%	0.46%	-0.63%	-4.70%	-2.43%	-5.35%	-2.66%	-5.43%	-2.66%
PartyScene	-6.19%	-4.43%	-5.31%	-4.71%	-8.69%	-6.22%	-8.91%	-6.85%	-8.92%	-7.01%
PeopleOnStreet	-0.29%	-0.13%	-0.82%	-0.58%	-1.32%	-1.02%	-1.67%	-1.17%	-2.34%	-1.41%
RaceHorses	-3.03%	-2.66%	-6.14%	-3.75%	-6.93%	-4.14%	-6.68%	-4.04%	-6.63%	-4.04%
Traffic	-0.28%	-0.14%	-1.93%	-0.75%	-3.35%	-1.76%	-4.17%	-2.03%	-4.63%	-2.16%
BasketballPass	-0.42%	-0.26%	-3.32%	0.19%	-1.97%	-1.17%	-0.95%	-0.63%	-0.83%	-0.61%
BlowingBubbles	-1.17%	-1.25%	-0.61%	-1.12%	-2.39%	-1.86%	-2.41%	-1.76%	-2.29%	-1.70%
BQSquare	-1.25%	-0.24%	1.36%	0.74%	0.52%	-0.11%	3.14%	0.86%	3.47%	1.00%
Tennis	-1.03%	-0.44%	-0.08%	1.03%	-3.49%	-2.30%	-4.51%	-2.52%	-4.93%	-2.50%
Vidyo3	-0.07%	-0.05%	-2.60%	-0.33%	-3.98%	-1.24%	-6.20%	-1.64%	-6.05%	-1.50%
Flowervase	0.06%	-0.02%	-3.66%	0.12%	-7.95%	-3.18%	-10.25%	-4.20%	-10.44%	-4.22%
Keiba	-0.72%	-0.70%	-1.63%	-0.75%	-2.05%	-1.35%	-2.39%	-1.46%	-1.66%	-1.04%
Mobisode2	-4.58%	-2.10%	-4.88%	-0.57%	-6.72%	-1.99%	-6.44%	-1.50%	-6.20%	-1.02%
SteamLocomotiveTrain	-2.98%	-0.50%	-1.37%	0.28%	-2.00%	-2.21%	-1.75%	-2.52%	-1.90%	-2.72%
NebutaFestival	3.29%	-0.81%	1.38%	-0.71%	3.96%	-0.82%	3.34%	-1.37%	1.47%	-2.09%
Average	**-0.98%**	**-0.69%**	**-1.48%**	**-0.44%**	**-3.33%**	**-1.86%**	**-3.87%**	**-2.02%**	**-4.04%**	**-2.06%**

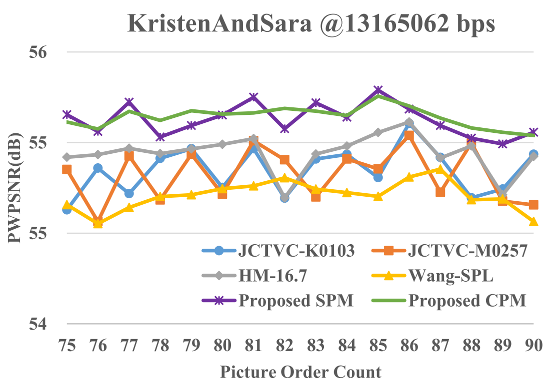

FIGURE 5. PWPSNR per picture for the sequence KristenAndSara.

frame of our proposed algorithm is significantly better than that of other algorithms. Because we consider the perceptual characteristics of the HVS, and allocate fewer bits to the coding unit with a significant masking effect and allocate relatively more bits to other coding units. It removes the perceptual redundancy and achieves better coding performance. Overall, the comparative experimental results show that the coding performances of our algorithms are significantly better than those of the JCTVC-K0103, JCTVC-M0257, HM-16.7 default anchor and Wang-SPL.

C. COMPARISON OF R-D PERFORMANCE USING DIFFERENT QUALITY METRICS

Besides the PWPSNR, seven other popular quality metrics are also utilized for performance evaluation, including IFC [29], MSSSIM [30], PWMSE_VIDEO [31], SSIM [32], UQI [33], FSIM [34], and VIF [35]. It should be noted that these metrics of IFC, MSSSIM, SSIM, and UQI are computed in case of the luminance component.

The experimental results of these metrics are shown in Tables 5, 6, and 7. From Table 5, the BD-BR with respect to PWMSE_VIDEO, denoted as BD-BR (PWMSE_VIDEO) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −0.98%, −1.48%, −3.33%, −3.87%, −4.04%, respectively. In case of IFC, the values of BD-BR (IFC) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −0.69%, −0.44%, −1.86%, −2.02%, and −2.06%, respectively. From Table 6, taking SSIM as the quality evaluation metric, the values of the BD-BR (SSIM) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −2.52%, −0.64%, −6.20%, −7.40%, −8.00%, respectively. In case of using UQI, the values of BD-BR (UQI) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −1.72%, −2.85%, −6.71%, −8.47%, and −8.95%, respectively. Among these quality metrics, the proposed rate control with SPM and CPM have better visual quality or more bit rate savings than the benchmark schemes. Therefore, we can conclude that the performance of our proposed algorithms are better than the state-of-the-art algorithms. In addition, the bit rate savings of proposed CPM are more than that of proposed SPM. The results of the other three perceptual evaluation methods are shown in Table 7. Due to the page limitation, the data in the table is the mean BD-BR of the above test sequences. We can

TABLE 6. R-D performance comparisons when quality is measured with SSIM [32] and UQI [33].

Sequence	M0257 [7]		Wang-SPL [19]		HM-16.7 [25]		Proposed SPM		Proposed CPM	
	BD-BR (SSIM)	BD-BR (UQI)	BD-BR (SSIM)	BD-BR (UQI)	BD-BR (SSIM)	BD-BR (UQI)	BD-BR (SSIM)	BD-BR (UQI)	BD-BR (SSIM)	BD-BR (UQI)
BasketballDrill	-0.62%	-0.14%	0.48%	-0.85%	-7.25%	-7.32%	-7.84%	-8.13%	-7.77%	-8.11%
BasketballDrive	-0.47%	-0.14%	4.70%	-2.44%	-3.65%	-6.29%	-5.42%	-8.39%	-5.97%	-8.86%
BQMall	-5.14%	-4.06%	-3.47%	-4.54%	-4.97%	-5.20%	-5.32%	-6.57%	-5.21%	-6.44%
BQTerrace	-2.34%	-0.73%	-2.95%	-8.37%	-8.29%	-7.66%	-10.10%	-10.01%	-14.62%	-13.11%
Cactus	-0.90%	-0.18%	0.23%	-4.08%	-7.26%	-7.87%	-8.54%	-9.39%	-10.19%	-11.03%
FourPeople	-0.19%	0.00%	2.06%	-3.53%	-4.88%	-8.97%	-4.89%	-9.82%	-4.82%	-9.86%
Johnny	-0.01%	-0.17%	6.04%	0.06%	-1.47%	-2.78%	-2.57%	-6.87%	-2.57%	-6.85%
Kimono	-0.29%	-0.20%	1.47%	-0.51%	-2.81%	-3.26%	-3.97%	-4.51%	-3.97%	-4.88%
KristenAndSara	-0.06%	0.02%	1.68%	-1.61%	-3.66%	-4.00%	-7.66%	-7.98%	-7.10%	-7.97%
ParkScene	-0.64%	-0.39%	-0.87%	-0.77%	-7.34%	-8.41%	-8.38%	-10.10%	-8.47%	-10.33%
PartyScene	-12.62%	-10.33%	-10.09%	-9.60%	-14.70%	-12.96%	-15.13%	-14.08%	-15.13%	-14.28%
PeopleOnStreet	-0.34%	-0.08%	-7.54%	-6.23%	-7.16%	-4.99%	-7.31%	-5.87%	-12.50%	-9.03%
RaceHorses	-2.74%	-2.68%	0.01%	-0.12%	-2.81%	-3.97%	-3.42%	-4.88%	-3.11%	-4.66%
Traffic	-11.84%	-7.59%	-11.10%	-10.00%	-12.18%	-10.45%	-12.09%	-11.24%	-11.86%	-11.19%
BasketballPass	-1.01%	-0.11%	-2.85%	-4.08%	-8.68%	-7.37%	-11.46%	-9.60%	-11.83%	-10.28%
BlowingBubbles	-4.96%	-4.49%	-4.16%	-3.92%	-6.81%	-7.60%	-6.88%	-8.58%	-6.76%	-8.43%
BQSquare	-4.51%	0.30%	-4.31%	-0.13%	-7.16%	-4.44%	-7.42%	-5.92%	-7.27%	-5.79%
Tennis	-0.45%	-0.90%	7.65%	0.41%	-3.26%	-8.08%	-4.25%	-8.91%	-5.49%	-8.96%
Vidyo3	-0.45%	-0.17%	-0.42%	-3.25%	-4.76%	-6.74%	-6.14%	-9.95%	-5.73%	-9.88%
Flowervase	-0.30%	-0.06%	7.00%	1.20%	-6.41%	-10.30%	-7.00%	-12.63%	-6.93%	-12.37%
Keiba	-2.18%	-1.53%	-0.07%	-3.63%	-0.42%	-2.56%	-2.53%	-5.10%	-2.24%	-5.28%
Mobisode2	-0.20%	-2.65%	7.11%	-0.02%	3.52%	-3.98%	3.64%	-5.32%	4.52%	-4.42%
SteamLocomotiveTrain	-4.07%	-0.77%	-3.12%	-1.52%	-19.31%	-10.63%	-20.77%	-12.93%	-22.63%	-14.73%
NebutaFestival	-4.14%	-4.32%	-2.83%	-0.76%	-7.14%	-5.23%	-12.07%	-6.52%	-14.43%	-8.00%
Average	**-2.52%**	**-1.72%**	**-0.64%**	**-2.85%**	**-6.20%**	**-6.71%**	**-7.40%**	**-8.47%**	**-8.00%**	**-8.95%**

TABLE 7. BDBR performance comparison when the image quality is measured with FSIM [34], VIF [35] and MSSSIM [30].

algorithms	BD-BR (FSIM)	BD-BR (VIF)	BD-BR (MSSSIM)
M0257	-4.37%	-1.19%	-3.67%
Wang-SPL	-1.42%	0.31%	-1.49%
HM-16.7	-6.44%	-1.58%	-6.43%
Proposed SPM	-6.74%	-1.63%	-7.56%
Proposed CPM	-6.57%	-1.57%	-7.85%

observe that the BD-BR (FSIM) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −4.37%, −1.42%, −6.44%, −6.74%, −6.57%, respectively. The values of the BD-BR (FSIM) of the algorithm JCTVC- M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −1.19%, 0.31%, −1.58%, −1.63%, −1.57%, respectively. The values of the BD-BR (MSSSIM) of the algorithm JCTVC-M0257, HM-16.7 default anchor, Wang-SPL, proposed SPM, and proposed CPM are −3.67%, −1.49%, −6.43%, −7.56%, −7.85%, respectively. It can be found that the performance of the proposed CPM is not as good as the proposed SPM, in BD-BR (FSIM) and BD-BR (VIF). The reason is that our objective aims to minimize the PWMSE subject to the bit rate constraint, and the characteristics of the two evaluation metrics are a little different from the PWMSE, resulting in some differences in the results with other metrics. But overall, the performance of our proposed two perception based bit allocation algorithms can achieve significant R-D performance improvement than other algorithms.

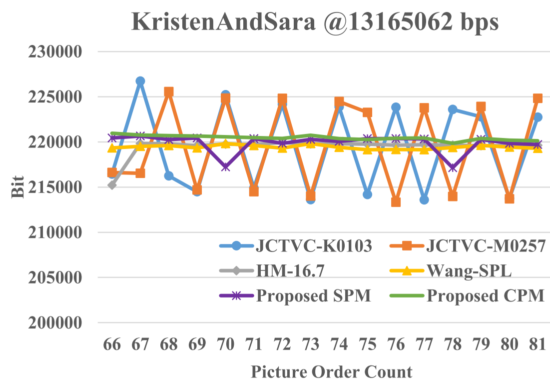

FIGURE 6. Actual bits per picture for the sequence KristenAndSara.

D. BIT RATE ACCURACY COMPARISONS

Besides the R-D performance comparison, we use bitrate error rate to measure the accuracy of the rate control algorithm. The average bitrate error rate (E) is calculated as

$$E = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|R_{T,i} - R_{A,i}\right|}{R_{T,i}} \cdot 100\%, \tag{32}$$

where $R_{T,i}$ is the target bitrate and $R_{A,i}$ is actual bit rate, N is the number of bit rate point, which is 4 in this paper. The average of the bitrate error of the four bit rate points is regarded as the final bitrate error of the sequence. The average bitrate error rates (E) of different coding schemes are shown in Table 8. The bitrate error of JCTVC-0257, Wang-SPL, proposed SPM, and proposed CPM are 0.014%, 0.032%, 0.003%, 0.005%, respectively. Fig.6 illustrates the

TABLE 8. Comparison of bitrate error rate (%) between proposed perception-based rate control algorithm and other algorithms.

Sequences	JCTVC-M0257[7]	Wang-SPL [19]	Proposed SPM	Proposed CPM
BasketballDrill	0.006%	0.006%	0.004%	0.002%
BasketballDrive	0.015%	0.007%	0.002%	0.002%
BQMall	0.012%	0.006%	0.010%	0.008%
BQTerrace	0.019%	0.002%	0.001%	0.003%
Cactus	0.010%	0.002%	0.001%	0.011%
FourPeople	0.005%	0.006%	0.002%	0.007%
Johnny	0.011%	0.001%	0.003%	0.003%
Kimono	0.006%	0.003%	0.002%	0.004%
KristenAndSara	0.006%	0.006%	0.003%	0.003%
ParkScene	0.004%	0.009%	0.001%	0.005%
PartyScene	0.014%	0.001%	0.001%	0.002%
PeopleOnStreet	0.009%	0.003%	0.001%	0.002%
RaceHorses	0.007%	0.001%	0.005%	0.006%
Traffic	0.009%	0.006%	0.001%	0.003%
BasketballPass	0.017%	0.034%	0.004%	0.008%
BlowingBubbles	0.006%	0.011%	0.001%	0.001%
BQSquare	0.008%	0.011%	0.004%	0.012%
Tennis	0.012%	0.099%	0.007%	0.006%
Vidyo3	0.009%	0.023%	0.005%	0.003%
Flowervase	0.007%	0.085%	0.003%	0.002%
Keiba	0.011%	0.053%	0.009%	0.008%
Mobisode2	0.078%	0.027%	0.010%	0.005%
SteamLocomotiveTrain	0.038%	0.318%	0.001%	0.001%
NebutaFestival	0.016%	0.040%	0.000%	0.004%
Average	**0.014%**	**0.032%**	**0.003%**	**0.005%**

bit rate comparison among different rate control scheme for the sequence KristenAndSara. We observe that at the same bit rate, the bit fluctuation of Wang-SPL, and our proposed SPM and CPM algorithms are more stable than those of other algorithms. Our proposed algorithm is much better than the state-of-the-art schemes and maintains good performance in terms of bit rate control accuracy.

VII. CONCLUSION

In this paper, in order to further improve visual quality and the coding performance, a perception based Coding Tree Unit (CTU) level bit allocation algorithm for intra High Efficiency Video Coding (HEVC) is proposed. Perceptually Weighted Mean Squared Error (PWMSE) is adopted to represent the video perceptually weighted distortion. Then, we present four models to formulate the relationship between Rate (R) and Perceptually Weighted Distortion (D_p), and derive the perceptual weighting factors for guiding bit allocation. Compared with other advanced CTU level bit allocation methods, our proposed algorithm achieves a great gain in terms of bit rate reduction and video quality improvement. Besides, the proposed perception-based CTU level bit allocation algorithm maintains high performance in bit rate accuracy.

REFERENCES

[1] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[4] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajić, *Rate Control Based on Unified RQ Model for HEVC*, document ITU-T/ISO/IEC, JCT-VC JCTVC-H0213, San Jose, CA, USA, 2012.

[5] B. Li, H. Li, L. Li, and J. Zhang, *Rate Control by R-Lambda Model for HEVC*, Document ITU-T/ISO/IEC JCT-VC, JCTVC-K0103, Shanghai, China, 2012.

[6] *HM Reference Software 10.0*. Accessed: Jan. 2013. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-10.0/

[7] M. Karczewicz and X. Wang, *Intra frame rate control based on SATD*, document ITU-T/ISO/IEC, JCT-VC JCTVC-M0257, Incheon, Korea, 2013.

[8] B. Li, H. Li, and L. Li, *Adaptive Bit Allocation for R-Lambda Model Rate Control in HM*, document ITU-T/ISO/IEC and JCT-VC JCTVC-M0036, Incheon, KR, 2013.

[9] F. Song, C. Zhu, Y. Liu, Y. Zhou, and Y. Liu, "A new GOP level bit allocation method for HEVC rate control," in *Proc. BMSB*, Cagliari, Italy, Jun. 2017, pp. 1–4.

[10] H. Guo, C. Zhu, S. Li, and Y. Gao, "Optimal bit allocation at frame level for rate control in HEVC," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 270–281, Jun. 2019.

[11] L. Li, B. Li, and C. W. Chen, "*Lambda*domain optimal bit allocation algorithm for high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 1, pp. 130–142, Jan. 2018.

[12] Y. Zhang, S. Kwong, and S. Wang, "Machine learning based video coding optimizations: A survey," *Inf. Sci.*, vol. 506, pp. 395–423, Jan. 2020.

[13] H. Liu, Y. Zhang, H. Zhang, C. Fan, S. Kwong, C.-C. J. Kuo, and X. Fan, "Deep learning based picture-wise just noticeable distortion prediction model for image compression," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 641–656, Dec. 2020.

[14] W. Gao, S. Kwong, Y. Zhou, and H. Yuan, "SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 988–999, Jun. 2016.

[15] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "SSIM-motivated two-pass VBR coding for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2189–2203, Oct. 2017.

[16] M. Zhou, X. Wei, S. Wang, S. Kwong, C. Fong, P. Wong, W. Yuen, and W. Gao, "SSIM-based global optimization for CTU-level rate control in HEVC," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 1921–1933, Aug. 2019.

[17] A. Yang, H. Zeng, L. Ma, J. Chen, C. Cai, and K.-K. Ma, "A perceptual-based rate control for HEVC," in *Proc. IPTA*, Oulu, Finland, Dec. 2016, pp. 1–5.

[18] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[19] M. Wang, K. N. Ngan, and H. Li, "An efficient frame-content based intra frame rate control for High Efficiency Video Coding," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 896–900, Jul. 2015.

[20] H. Wei, W. Zhou, R. Bai, and Z. Duan, "A rate control algorithm for HEVC considering visual saliency," in *Proc. APSIPA ASC*, Honolulu, HI, USA, Nov. 2018, pp. 36–42.

[21] X. Sun, H. Ma, W. Zuo, and M. Liu, "Perceptual-based HEVC intra coding optimization using deep convolution networks," *IEEE Access*, vol. 7, pp. 56308–56316, 2019.

[22] H. Wang, L. Song, R. Xie, Z. Luo, and X. Wang, "Masking effects based rate control scheme for high efficiency video coding," in *Proc. ISCAS*, Florence, Italy, May 2018, pp. 1–5.

[23] S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C.-C. J. Kuo, "Compressed image quality metric based on perceptually weighted distortion," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5594–5608, Dec. 2015.

[24] B. Li, H. Li, L. Li, and J. Zhang, "λ Domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sep. 2014.

[25] *HM Reference Software 16.7*. Accessed: Oct. 2015. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.7/

[26] Z. Chen and X. Pan, "An optimized rate control for low-delay H.265/HEVC," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4541–4552, Sep. 2019.

[27] B. Li, D. Zhang, H. Li, and J. Xu, *QP Determination by Lambda Value*, Document ITU- T/ ISO/IEC, JCT-VC JCTVC-I0426, Geneva, Switzerland, 2012.

[28] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," in *Proc. VCEG-M33 Meeting*, Austin, Texas, USA, Apr. 2001, pp. 2–4.

[29] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. ACSSC*, Nov. 2003, pp. 1398–1402.

[31] S. Hu, L. Jin, H. Wang, Y. Zhang, S. Kwong, and C.-C. J. Kuo, "Objective video quality assessment based on perceptually weighted mean squared error," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 9, pp. 1844–1855, Sep. 2017.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[33] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[34] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[35] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

XIAOYAN LIU received the B.S. degree in digital media technology from Yuncheng University, Yuncheng, China, in 2017. She is currently pursuing the M.S. degree with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, and the Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen. Her research interests include perceptual image or video coding.

YUN ZHANG (M'12–SM'16) received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2014, he was a Postdoctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong. From 2010 to 2017, he served as an Assistant Professor and an Associate Professor with the Shenzhen Institutes of Advanced Technology (SIAT), CAS, where he has been a Full Professor, since 2017. His research interests include video compression, 3D video processing, and visual perception.

LINWEI ZHU received the B.S. degree in applied physics from the Tianjin University of Technology, China, in 2010, the M.S. degree in signal and information processing from Ningbo University, China, in 2013, and the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Hong Kong, in 2019. He is currently a Postdoctoral Fellow with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS). His research interests mainly include depth image-based rendering, depth estimation, and machine learning-based video coding/transcoding.

HUANHUA LIU received the B.S. degree in computer science and technology from Changsha University, China, in 2009, and the M.S. degree in computer science and technology from the University of South China, China, in 2013. He is currently pursuing the Ph.D. degree with Central South University. Since 2017, he has been with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. His current research interests include learning-based image/video quality assessment and video coding.