Multimedia Tools and Applications (2022) 81:12495–12518
https://doi.org/10.1007/s11042-022-12146-1

Perceptual importance analysis-based rate control method for HEVC

HongWei Lin[1] · Xiangqun Li[1] · Mingliang Gao[1] · Keyan Deng[1] · Yongsheng Xu[1]

Received: 3 June 2021 / Revised: 1 September 2021 / Accepted: 3 January 2022 /
Published online: 19 February 2022

Abstract
High efficiency video coding (HEVC) has achieved high coding efficiency as the video coding standard. For rate control in HEVC, the conventional R-λ scheme is based on mean absolute difference in allocating bits; however, the scheme does not fully utilize the perceptual importance variation to guide rate control, thus the subjective and objective quality of coded videos has room to improve. Therefore, in this paper, we propose a rate control scheme that considers perceptual importance. We first develop a perceptual importance analysis scheme to accurately abstract the spatial and temporal perceptual importance maps of video contents. The results of the analysis are then used to guide the bit allocation. Utilizing this model, a region-level bit allocation procedure is developed to maintain video quality balance. Subsequently, a largest coding unit (LCU)-level bit allocation scheme is designed to obtain the target bit of each LCU. To achieve a more accurate bitrate, an improved R-λ model based on the Broyden-Fletcher-Goldfarb-Shanno model is utilized to update the R-λ parameter. The experimental results showed that our method not only improved subjective and objective video quality with lower bitrate errors compared to the original RC in HEVC, but also outperformed state-of-the-art methods.

Keywords HEVC · Rate control · Perceptual importance · Bit allocation · Parameter update

1 Introduction

High efficiency video coding (HEVC) is a video compression standard launched by the Joint Collaborative Team on Video Coding in 2013. Compared to H.264/AVC, HEVC can save 50% bitrate while generating similar reconstructed video quality [35]. Improvement in coding

✉ HongWei Lin
linhongwei@xbmu.edu.cn

[1] College of Electrical Engineering, Northwest Minzu University, No.1 Xibeixincun, Lanzhou 730030, China

efficiency is achieved by exploiting several new techniques, such as advanced motion vector, prediction unit, transform unit, quadtree structure-based coding tree unit (LCU) with variable block sizes varying from8 × 8 to 64 × 64, and so on [41]. Due to the above new techniques, HEVC has become one of the most important standards in video applications.

Rate control (RC) plays a key role in video coding systems, and the goal is to maintain good visual quality by matching the constraints of video channel bandwidth. To achieve this objective, most of the video encoding standards incorporate RC algorithms into their encoding frameworks. RC methods can be divided into two steps. The first step is bit allocation, which is performed to achieve optimal rate distortion (R-D) performances through efficient allocation of proper bitrates at different coding levels, including group of picture (GOP) level, frame level, and coding units (CU) level [4]. The second step is choosing proper quantization parameters (QPs) to achieve the allocated bits for each level. The above RC methods is applied in many video coding standards, such as TMN8 [38] applied in H.263 and JVT-N046 [28] applied in H.264/AVC.

Similar to H.264/AVC, HEVC also adopts an RC method to optimize the Lagrangian rate distortion optimization (RDO) [45] performance of coded videos. For example, the latest HEVC RC method, λ domain RC [20, 22], is an important part in HEVC. In the λ domain RC method, the bitrate allocation is still in terms of mean absolute difference (MAD) of LCU. However, the video quality is mostly verified by human eyes, and according to the human visual system (HVS), there is considerable perceptual redundancy in video frames [18]. For instance, when a person watches video frames, a region with people or moving objects (high perceptual importance region (HPIR)) is given more attention than other regions (low perceptual importance regions (LPIR)). Thus, the MAD may not be sufficiently correlated with perceptual quality [13]. Therefore, a large number of bits can be saved by reducing perceptual redundancy in the LPIR region with imperceptible perceptual quality loss.

When people watch videos, more attention is attracted by areas of HPIR, e.g., human faces, figures, and moving objects [18]. The basis of perceptual-based video coding is that the distortion in HPIR is more likely to be perceived; thus, more bits should be assigned to these areas to maintain visual quality. However, the visual quality in HPIR is often lower than that in LPIR. The reason is that HPIR usually contains texture information and fast-moving objects. Thus, compared to LPIR, HPIR achieves lower visual quality improvement with equally allocated bits. Hence, it is necessary to design a perceptual importance-based RC method for HEVC with the aim of providing optimal perceptual video quality under a constraint bandwidth. The perceptual importance-based video coding quality assessment has two requirements. First, the perceptual assessment model should be able to describe HPIR and LPIR with high accuracy. Second, the perceptual assessment model should possess a low complexity property and be easily incorporated into the video coding procedure (e.g., rate control).

The rest of this paper is organized as follows. Section 2 briefly reviews related works, and Sec. 3 describes the proposed perceptual importance classification algorithm. In Sec. 4, a perceptual importance-based RC algorithm is proposed. The experimental results and discussions are presented in Sec. 5. Finally, the conclusion of this paper is provided in Sec. 6.

2 Related works

There have already been several works on RC for video coding standards. For the previous video coding standard H.264/AVC, Liu et al. [30] presented a linear rate quantization (R-Q) model for fitting the relationship between bitrate and the QP. Moreover, a prediction scheme

was proposed to reduce the MAD abruption. An et al. [1] presented a primal-dual decomposition and sub-gradient projection-based method to iteratively calculate the RDO procedure for H.264/AVC RC, which can improve the RD performance of the RC algorithm. Dong et al. [9] proposed a context-adaptive parameter prediction scheme to improve the accuracy of the estimated MAD of the R-Q model used in H.264/AVC.

HEVC is a next-generation video encoding standard, and various RC methods have been proposed for it. For instance, Choi et al. [7, 8] proposed some R-Q methods, where the R-Q model is the recommended RC scheme to be adopted by HM6, the reference software of HEVC. Liang et al. also proposed an HEVC RC scheme based on the R-ρ model [27], where ρ is the percentage of zeroes among the DCT coefficients after quantization. Compared to the R-Q model, the R-ρ model shows a slight improvement in estimating the target bits. By analyzing the residual signal probability distribution of hierarchical quad-tree CUs, Lee et al. [19] proposed an RC algorithm that allocated texture and non-texture with different rate models. Subsequently, Li et al. [21, 25] found that the Lagrange multiplier (λ) is a crucial factor for RC in HEVC, and they proposed the R-λ model for HEVC, which has a lower bitrate mismatch and better R-D performance than the R-Q and R-ρ models. The R-λ model was adopted in the HEVC test software HM due to its outstanding rate accuracy and R-D performance. Gao et al. [10] and Guo et al. [15] proposed a temporal RDO propagation models for HEVC bit allocation procedure, and these methods showed better coding efficiency. Li et al. [26] proposed a LCU-level HEVC bit allocation method, which achieved better R-D performance as it considered the R-D characteristics of each LCU in one frame. In [6], Chen et al. proposed an optimized LCU-level low-delay RC approach for HEVC in which the parameter distributions of the estimated R-D model are considered, thereby efficiently improving the R-D performance. In addition to the above works, an extremely low-delay method was designed for an HEVC intra-frame RC model in [29]. Further, in reference [12], Gao et al. proposed a data-driven RC method, which improved HEVC R-D performance by an effective initial QP-chosen method. A joint machine learning-based RC scheme was also proposed by Gao et al. [11] to improve the performance of the R-λ model. In our previous work [48], a new parameter updating method is proposed to improve the RC performance of the R-λ model. However there is some problems in [48]: i. The work of [48] only utilizes the gradient value as spatial information to guide bit allocation, this guidance mode not consider the temporal information, thus the bit allocation procedure is relatively rough. ii. The gradient information is not very match the video perceptual importance. To address the above issues, in this work, a temporal-spatial combined information is proposed to guide bit allocation and a relationship between bitrate and perceptual importance is formulated. In addition, in order to keep the convergence speed, we use the parameter updating method in [48] to maintain the parameter updating speed in this work. It is noteworthy that, only the parameter updating procedure of this work is referred the corresponding part in [48].

As video quality is assessed by humans, video coding standards that incorporate well-designed human perceptual RC algorithms have attracted considerable research interests [46]. Chadha et al. [5] proposed a rate aware perceptual preprocessing method, which can enhance visual quality with any codec and bitrates. Zeng et al. [49] developed a perceptual sensitivity scheme to guide bit allocation. Zhu et al. [52] designed a perceptual based RDO scheme, in this a CNN-based on-line training method is first explored to determine the VMAF-related distortion estimation coefficient. Recently, Zhou et al. [51] also established an SSIM-based rate-distortion model, and the model was transformed into a global optimization problem to guide the LCU-level RC of HEVC. A weight-based R-λ perceptual RC scheme was presented

by Li et al. [24] based on the observation that faces draw more attention in conventional video, weight map-based eye tracking was utilized in the bit allocation procedure. In [23], the researchers claim that visual saliency can represent the probability of human attention; hence, graph-based visual saliency was utilized to adjust QP, which assigned less bitrates with a low probability of visual attention. Bai et al. [2] used average saliency to weigh bit allocation for each LCU. The ROI concept is often utilized to guide bit allocation in RC too, in [32], the coding blocks in the ROI were encoded with lower QP to improve perceptual quality. In addition to spatial complexity, the temporal complexity of video sequence is an important factor to measure the perceptual based bit allocation in rate control. Recently, Wei et al. [43] used static and dynamic based perceptual feature to control bit allocation. Wang et al. [42] also proposed a masking effect-based RC method, which considered temporal and spatial information. However, the bitrate accurately of these models are relatively rough. Gong et al. [14] used a temporal-layer-motivated method to guide bit allocation, and achieved better rate control results in random access configuration of HEVC. In [44], Wei et al. used spatial/temporal visual saliency to guide the LCU-level bit allocation in HEVC, and the distortion of each LCU was weighted by the corresponding saliency. These above perceptual based methods all use spatial/temporal human perceptual factors as the weight of each region to represent the perceptual quality, but the factors are is not match the perceptual importance very well. In addition, bit allocation for different regions (such as HPIR and LPIR) is no balance optimization in these algorithms that means the LCU-level bit allocation lacks an optimal global bit allocation between different perceptual regions. Therefore, the LCU bitrate in HPIR might be excessive, causing the perceptual quality of LPIR to be too low, or vice versa. Finally, the RC parameter updating procedure in these methods is all follow the method in [22] that is a first-order convergence model, which means the convergence speed of updating the parameters is relatively slow, that induce low bit rate accuracy.

To address the above problems, in this study, we investigated the LCU-level rate control based on perceptual importance analysis and formulated the LCU-level rate control. The contributions of this study are as follows: a simple but effective perceptual importance analysis algorithm that combines temporal-spatial information to express perceptual importance is proposed. The relationship between bitrate and perceptual importance analysis is established that is further applied in the formulation of bit allocation. A region-level bit allocation which considers a global optimization problem is established that can further balance the video quality of different regions. A new model parameter updating strategy is used in R-λ RC model that is robust to scene variations.

3 Proposed perceptual importance classification algorithm

It should be noted that the high perceptual importance region is not always equal to the more visual attention region in the HVS. For instance, a region with moving objects is likely to attract visual attention. However, once the moving objects are in random texture regions or their speed becomes faster than the noticeable capacity of humans, viewers tend to ignore distortions in these regions. Thus, such regions are perceptually less important [34]. In this section, we describe the proposed perceptual importance classification algorithm for video frames. As the distortions in a region with high perceptual importance are easily noticed by the HVS, the proposed perceptual importance analysis algorithm is mainly composed of three parts: moving analysis model, texture region distinction, and model fusion. The moving degree

of LCUs in a frame is represented numerically using the moving analysis model. Subsequently, a texture region distinction is utilized to separate LCUs based on the texture information intensity. After the LCU classification, a complete perceptual importance weight is decided by combining the results of the moving analysis and texture distinction models. The analysis algorithm is described below.

3.1 Moving analysis model

As mentioned above, people are more sensitive to moving objects, especially in video applications such as video conferencing, video surveillance, and visual telephone. In these applications, viewers typically focus on moving objects, which means moving regions in video frames attract more attention than the stationary regions [37]. Hence, any distortion in a moving area is easily detected. Therefore, the perceptual qualities of moving analysis are crucial to the overall video frame quality. As HEVC adopts LCU as the base coding unit, an LCU-based region moving degree (RMD) method is proposed to indicate the moving magnitude of each LCU. To obtain the RMD, each video frame is first sent to a low-pass filter whose main function is to remove the high-frequency noise in the video frame. To reduce complexity, we use a 3 × 3 averaging filter with a uniform weight of 1/9. It is well-known that the smaller the luminance difference between the LCUs in the same position in two consecutive frames [30], the more similar the two LCUs are. Equivalently, the moving magnitude of the LCUs in the current frame is low, and vice versa. Therefore, we use the luminance difference between two LCUs in the same position in two consecutive frames to describe RMD:

$$MD_n(X,Y) = \sum_{(i,j)\in L_n(X,Y)} (P_n(i,j) - P_{n-1}(i,j)) \tag{1}$$

where $P_n(i,\ j)$and $P_{n\ -\ 1}(i,\ j)$ represent the luminance pixels at location $(i,\ j)$ in the current and previous frames, and $L_n(X,\ Y)$ represents the LCU at location $(X,\ Y)$in frame f_n.

For motion LCUs, the distortion sensitivity of the human eye decreases with an increase in motion speed [17]. Thus, $L_n(X,\ Y)$ is classified as moving at a normal speedLCU_{MNS} if $MD_N(X,\ Y)$is less than a thresholdT_N; otherwise, $L_n(X,\ Y)$ is classified as moving too fast LCU_{MTS}. The threshold T_m is defined as

$$T_m = \alpha \times \frac{1}{N} \sum_{B_n(X,Y)\in f_n} MD_n(X,Y) \tag{2}$$

where, $\alpha$is a scaling factor with a value of 1.2 in our experiments, and N is the number of LCUs in each frame.

After the LCU classification according to motion speed, we comprehensive consider the characteristics of LCU_{MNS} and LCU_{MTS}, and the RMD for each LCU is finally defined as:

$$RMD_m(X,Y) = \begin{cases} T_m / MD_n(X,Y), & \text{if } MD_n(X,Y) > T_m \\ MD_n(X,Y)/T_m, & \text{Otherwise} \end{cases} \tag{3}$$

Defining RMD with Eq. 3 has two advantages. First, as seen from the equation, $RMD_n(X,\ Y)$ is less than 1 for LCU_{MTS}, but greater than 1 for LCU_{MNS}. This character qualitatively shows that

LCU_{MNS} is of more perceptual importance thanLCU_{MTS}. Second, the moving speed degree is also quantitatively reflected from the equation: forLCU_{MTS}, the greater the value of $MD_n(X, Y)$, the closer the value of $RMD_n(X, Y)$ is to 0, which means the LCU is in a region of moving too fast and has less perceptual importance. For LCU_{MNS}, the greater value of $MD_n(X, Y)$ means that the current LCU has a higher magnitude in a normal motion span and more perceptual importance.

3.2 Texture region distinction

The conventional R-λ RC model adopted the MAD value of the LCU in a previous frame in the same level to measure the bit allocation weight [20, 21]. However, the MAD value was weakened to represent the HVS perception, which causes improper bit allocations and degrade perceptual quality [45]. There are several visual quality assessment metrics to measure the HVS perception [36, 39]. The work in [36] reports that there is a strong relationship between the texture characteristics and HVS perception. Hence, in this section, we propose an effective region texture degree (RTD) analysis model to measure the perceptual importance of the HVS in video frames. This not only indicates the HVS perception better than MAD, but is also easier to incorporate into the video compression standard.

An important characteristic of perceptual importance is that people are more likely to be attracted by texture regions containing high spatial contrasts than smooth regions containing low spatial contrasts. However, distortions in the texture regions that contain many edges are usually less noticeable [31]. Thus, more bits should be allocated to the texture region, and fewer bits may be allocated to the edge or smooth regions. Bazen et al. [3] suggested that squared gradients can represent texture, and they can effectively divide texture regions into texture region and edge region characteristics. Therefore, we propose a perceptual texture characteristics presentation method based on squared gradients. First, we adopt the sobel operator to capture the texture information of the current frame. The gradient value of the pixel $P_{i,j}$ at position (i,j) in $L_n(X, Y)$is defined as:

$$\begin{aligned} G_x &= P_{i-1,j+1} + 2P_{i,j+1} + P_{i+1,j+1} - P_{i-1,j-1} - 2P_{i,j-1} - P_{i+1,j-1} \\ G_y &= P_{i+1,j-1} + 2P_{i+1,j} + P_{i+1,j+1} - P_{i-1,j-1} - 2P_{i-1,j} - P_{i-1,j+1} \end{aligned} \tag{4}$$

The squared gradient is then calculated by:

$$G_{ii} = \sum_{(i,j)\in L_n(X,Y)} G_i^2, \quad G_{jj} = \sum_{(i,j)\in L_n(X,Y)} G_j^2, \quad G_{ij} = \sum_{(i,j)\in L_n(X,Y)} G_i G_j \tag{5}$$

The texture coherence of the squared gradient can be calculated by [3]:

$$\mathrm{Coh} = \frac{\sqrt{\left(G_{ii} - G_{jj}\right)^2 + 4G_{ij}^2}}{G_{ii} + G_{jj}} \tag{6}$$

If the Coh value of an LCU is larger than the edge threshold T_e, then the current LCU contains excessive edge information, and the LCU is classified as an edge LCU. In contrast, if the Coh value of an LCU is smaller than the texture threshold T_t, then the current LCU contains little texture information, and the LCU is classified as a smooth LCU. Otherwise, the LCU belongs to texture LCU. This LCU-type decision is given as:

$$LCU\ Type = \begin{cases} Edge\ LCU, & if\ \mathrm{Coh} > T_e \\ Texture\ LCU, & if\ T_t < \mathrm{Coh} \leq T_e \\ Smooth\ LCU, & others \end{cases} \tag{7}$$

After the texture LCU decision, some of the classified LCUs were inconsistent with their neighboring LCUs. This inconsistency causes severe artifacts and significantly degrades video quality. Therefore, after classification, the consistency of each block should be analyzed to rectify the inconsistencies. The rectification process is explained thus: all classified LCUs are examined in a raster scan order; if there are eight edge LCUs around a texture LCU, the predetermined texture LCU should be amended as an edge LCU, and vice versa. Some other cases are amended likewise.

3.3 LCU perceptual importance weight decision

Based on the above analysis, the perceptual importance of each LCU can have two perceptual characteristics: RMD and RTD. To consider these two degrees comprehensively for perceptual importance characterization, a two-degree fusion scheme is proposed to represent perceptual weighting in the RC procedure.

First, in accordance with the texture analysis model, the RTD importance of LCUs is scored by the perceptual importance level (PIL) (the results are shown in Table 1). As analyzed in Sec. 3.2, for the RTD characteristic, the most important perceptual region is texture LCUs, which is scored by level "3." The perceptual importance of smooth LCUs is one level lower than that of texture LCUs and is scored by level "2." Finally, the edge LCUs are classified as the lowest level of perceptual importance and scored by level "1" as distortions in these regions are less noticeable. If the value of LCUs in consecutive frames changed dramatically, visible flickering artifacts will be produced. To address this problem, temporal consistency should be considered, and the RTD importance of an LCU should be adjusted. Let Δ denote the maximum difference of the PIL value between an LCU and its reference LCU in a consecutive frame. We set Δ to 1 in our experiments.

After the RTD-based perceptual importance level is decided (Table 1), the RMD is incorporated into the fusion scheme. A production-based, two-perceptual-degree fusion method is utilized to compute perceptual weighting factor:

$$W = PIL \times RMD_m(X, Y) \tag{8}$$

For the proposed weighting factor W, if a texture LCU contains relatively slow moving objects assigned to relatively low values of W, then this type of LCUs can achieve more perceptual importance than edge LCUs containing too fast moving objects. This is because the distortion in a region with objects moving too fast or

Table 1 RTD perceptual importance of LCUs

Region texture degree (RTD)	Perceptual importance level
Texture LCUs	3
Smooth LCUs	2
Edge LCUs	1

the edge region is both more likely to be unnoticed. To verify the effectiveness of the proposed LCU perceptual importance weight decision method, an example of the LCU perceptual importance map for the sequence "BasketballDrive" is shown in Fig. 1. The black regions are edge LCUs, grey regions are moving-too-fast or smooth LCUs, white regions are normal-speed or texture LCUs, and lighter regions represent areas of higher perceptual importance. It is observed that the different importance regions can be successfully classified using the proposed method.

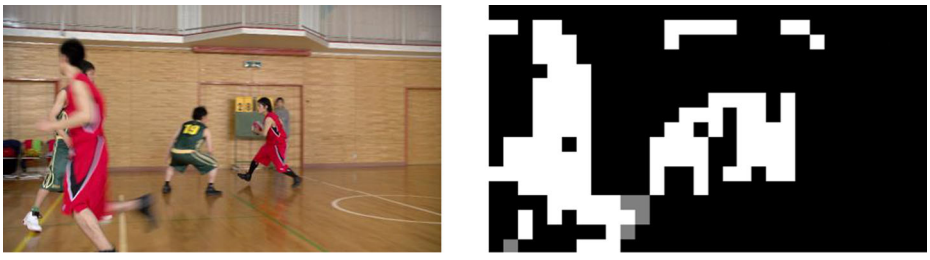
AUC is a standard used to measure the quality of classification model, the AUC value is between 0.5 and 1.0, and a larger AUC represents better classification performance [47]. In order to further evaluate the performance of the proposed perceptual importance classification algorithm more fairly, the AUC of the proposed classification method is tested. The comparison results of AUC with different bitrates and different configuration are tested and presented in Table 2. From Table 2, we can see that most AUC values of the proposed algorithm are bigger than two other algorithms. In addition, the average AUC values of proposed method are also bigger than other comparison methods too. That means the proposed perceptual importance classification algorithm is not only classified the perceptual importance area effectively but also better than the other exiting similar classification algorithms.

4 Proposed perceptual importance based rate control algorithm

4.1 Perceptual importance based bit allocation

4.1.1 Region level bit allocation

In the proposed bit allocation scheme, the GOP level and frame level bit allocations use same method in [22]. As seen from Table 1, the high-level region requires more bits to reduce distortions. However, if the high-level region is allocated too many bits, it will degrade video quality in the low-level region as only too few bits will be left to encode the region. Such quality degradation is inevitably perceived by the human eye. To address the problem, we propose a region-level bit allocation method for the proposed RC algorithm before presenting the LCU-level bit allocation method. For the RTD procedure in Sec. 3, LCUs in one frame are divided into three regions with the same RTD perceptual importance level. The bit allocation for different regions is:



(a) (b)

Fig. 1 Perceptual importance map for BasketballDrive. (a) Original frame; (b) The Perceptual importance map

Table 2 AUC Performance Comparison under LD and RA Configuration and Different Bitrates

Sequence	H Wei [43]				Gong [14]				Propose			
	LD		RA		LD		RA		LD		RA	
	OP=22	QP=32	OP=22	QP=32	OP=22	QP=32	OP=22	QP=32	OP=22	QP=32	OP=22	QP=32
Cactus	0.7655	0.7483	0.7746	0.7553	0.8573	0.7750	0.7961	0.7492	0.8997	0.8301	0.8680	0.8214
RaceHorses	0.6132	0.6094	0.6237	0.5971	0.74522	0.6336	0.6139	0.6260	0.7995	0.6339	0.6966	0.8328
BQSquare	0.6076	0.6280	0.5867	0.6079	0.6302	0.5922	0.6094	0.5824	0.6808	0.6553	0.6747	0.7195
BasketballPass	0.7550	0.7520	0.7264	0.7187	0.7296	0.7544	0.7857	0.7439	0.7707	0.7375	0.8170	0.8351
FourPeople	0.8206	0.8102	0.8211	0.8122	0.8372	0.8585	0.8308	0.8271	0.8657	0.9199	0.8876	0.8606
Johnny	0.8868	0.8847	0.8744	0.8716	0.8691	0.8563	0.8851	0.8794	0.8978	0.8701	0.8883	0.8504
Average	**0.7414**	**0.7387**	**0.7344**	**0.7271**	**0.7781**	**0.7450**	**0.7535**	**0.7347**	**0.8190**	**0.7744**	**0.8053**	**0.8199**

$$T_{Texture} = T_{Fra} \times \frac{Num_{Texture}}{Num_{Frame}} \tag{9}$$

$$T_{Smooth} = T_{Fra} \times \frac{Num_{Smooth}}{Num_{Frame}} \tag{10}$$

$$T_{Edge} = T_{Fra} \times \frac{Num_{Edge}}{Num_{Frame}} \tag{11}$$

where $T_{Texture}$, T_{Smooth}, and T_{Edge} are the target bits of all LCUs in the texture, smooth, and edge regions, respectively; $Num_{Texture}$, Num_{Smooth}, and $Num_{Edge}$are the LCU numbers in the texture, smooth, and edge regions, respectively; T_{Fra} is the target bits of the current frame; and Num_{Frame} is number of LCUs per-frame.

4.1.2 LCU level bit allocation

In the R-λ model, the MAD value and target bits of one whole frame is used to allocate bits in the LCU level. In our method, rather than replacing MAD in [22], the bit allocation of the LCU level follows the proposed perceptual weighting factor W and target bits of the same region. The target bits of each LCU can be formulated as:

$$T_{LCU,R} = T_R \times \frac{W_{i,R}}{\sum_{i=1}^{Num_R} W_{i,R}} \tag{12}$$

where $T_{LCU,R}$is the target bit of the current LCU in the same LCU region; $W_{i,R}$is the weighting factor of the i th LCU, $Num_R$is number of LCUs in the same region; $T_R$is the target bits of the whole region with the same LCU type, R is texture, smooth and edge, which represent texture, smooth and edge regions, respectively.

In a real RC application, there will be always a mismatch between the allocated target bits and actual encoding bits in each LCU. Thus, if the previous LCUs cost more or fewer bits than the target bits, the target bits of the remaining LCUs should be compensated for to maintain video quality. Thus, $T_{LCU,R}$should be improved as:

$$\begin{aligned} T^{I}_{LCU,R} &= \left\{ T_{R,rem} + \frac{\sum_{j=1}^{i-1}\left(T_{LCU,R,j} - T^{Act}_{LCU,R,j}\right)}{SW} \right\} \times Rat_{LCU,R} \\ Rat_{LCU,R} &= \frac{W_{k,R}}{\sum_{k=i}^{Num_R} W_{k,R}} \end{aligned} \tag{13}$$

where $T^{I}_{LCU,R}$is the improved target bits of the current (i.e., ith) LCU; $T_{R,rem}$is the number of remaining bits used to encode the remaining LCUs in the same LCU type region; $T_{LCU,R,j}$and $T^{Act}_{LCU,R,j}$are the numbers of actual encoded bits and target bits estimated by Eq.12 of the previous LCUs, respectively; and SWdenotes the size of the sliding window. In our experiments, SW= 8.

4.2 Improved R-λ parameter update model

Once the number of the target bits of LCU is determined, the next step is to determine an appropriate QP to achieve the target bits. In our work, the QP of each LCU can be achieved based on the R-λ model [22] as

$$\lambda = \alpha \times bpp^{\beta} \tag{14}$$

$$QP = 4.2005\ln\lambda + 13.7122 \tag{15}$$

where λ is the Lagrange coefficient, bpp is the number of target bits for per pixel, α and β are model parameters and are updated by the previous coded LCUs [22].

In order to adapt the different characteristics of the input video, the value of α and β should be continuously update during the encoding process. In the conventional R-λ model [22], α and β are updated as:

$$\lambda_{comp} = \alpha_{old} bpp_{real}^{\beta_{old}} \tag{16}$$

$$\alpha_{new} = \alpha_{old} + \delta_{\alpha} \times \left(\ln\lambda_{real} - \ln\lambda_{comp}\right) \times \alpha_{old} \tag{17}$$

$$\beta_{new} = \beta_{old} + \delta_{\beta} \times \left(\ln\lambda_{real} - \ln\lambda_{comp}\right) \times \ln bpp_{real} \tag{18}$$

where λ_{comp} and λ_{real} are the predicted and real lambda values, respectively; δ_{α} and δ_{β} are the learning rates set by 0.1 and 0.05, respectively; bpp_{real} represents real consumed bits.

For the notations in [22], the least mean square (LMS) algorithm was adopted by the R-λ model to update the values of α and β. However, the LMS model is a first-order convergence model, which means the convergence speed of updating the parameters of the LMS is relatively slow. Thus, the LMS algorithm cannot always achieve the accurate target bits. In our previous work [48], we proved that the Broyden-Fletcher-Goldfarb-Shanno (BFGS) model used a positive definite matrix, which avoids the trouble of directly calculating the Hessian matrix. Meanwhile, the inverse of the positive definite matrix can be easily obtained. Thus, the BFGS-based parameter updating algorithm can achieve a more global and faster convergence speed than the LMS-based method. Therefore, in this work, we used the BFGS-based model to update the parameters of the R-λ model. The BFGS-based α and β updating procedure is given as:

$$\alpha_{new} = \alpha_{old} + \delta_{amijo} \cdot d_{\alpha} \cdot \alpha_{old} \tag{19}$$

$$\beta_{new} = \beta_{old} + \delta_{armijo} \cdot d_{\beta} \tag{20}$$

where δ_{amijo} is the search step size calculated by a linear search process, d_{α} and d_{β} are search direction vectors of α and β, respectively [48].

Although, the BFGS-based model can update the parameters faster convergence, but the dramatical change in the bits of the LCUs caused by scene change or violent object movement will inevitable cause visible flickering artifacts. Thus, to keep the quality of coded video

consistent, both λ and QP should not change significantly. We proposed a new clipped method. That is: the value of λ and QP for the current LCU should be clipped in a range:

$$\lambda_{curr} = clip\left\{ \begin{array}{c} max\left(2^{\frac{-1}{3}} \cdot \lambda_{pre} / \frac{\sum_{i=1}^{n_{curr}} w_i}{n_{curr}}, 2^{\frac{-2}{3}} \cdot \lambda / \frac{\sum_{i=1}^{n_{curr}} w_i}{n_{curr}}\right), \\ min\left\{\lambda_{pre}, min\left(2^{\frac{1}{3}} \cdot \lambda_{pre} \cdot \frac{\sum_{i=1}^{n_{curr}} w_i}{n_{curr}}, 2^{\frac{2}{3}} \cdot \lambda \cdot \frac{\sum_{i=1}^{n_{curr}} w_i}{n_{curr}}\right)\right\} \end{array} \right\} \tag{21}$$

$$QP_{curr} = clip\left\{ max\left(QP_{pre} - \frac{\sum_{i=1}^{n_{curr}} w_i}{n_{curr}}, QP - \frac{2 \cdot \sum_{i=1}^{n_{curr}} w_i}{n_{curr}}\right)\right\} \tag{22}$$

where λ_{curr} is the λ value of the current LCU, λ_{pre} is the λ value of the previous encoded LCU, QP_{curr} is the QP of the current LCU, QP_{pre} is the QP of the previous encoded LCU. $\lambda_f$is the λ value of the current frame, and $n_{curr}$is the index of the current LCU. In the clip procedure, the λ value is not only adjusted by theλ_f, but also by the perceptual importance weight w_i. Thus, the clip procedure can not only maintain visual quality consistently, but also ensures that more bits are allocated to more perceptual important LCUs.

In order to show the whole proposed rate control algorithm more completely and clearly, we present the summary of the proposed rate control Algorithm in below:

Step 1: Initialize the GOP level and frame level bit allocations by the method in [22]
Step 2: Calculate the perceptual importance of current frame
Step 3: Calculate temporal complexity: classify the LCU of current frame into $LCU_{MNS}$and $LCU_{MTS}$by Eq. (1)–(3)
Step 4: Calculate spatial complexity: classify the LCU of current frame into*Edge LCU*, *Texture LCU* and *Smooth LCU*by Eq. (4)–(7)
Step 5: Decide the perceptual importance of each LCU by Eq. (8) and Table 1
Step 6: According to the perceptual importance results of step 2, allocate target bits for each region-level ($T_{Texture}$,T_{Smooth} and T_{Edge}) of current frame by using Eq. (9)–(11)
Step 7: Under the restriction of the above region-level bit allocation, the bits for each LCU in different regions are allocated by:
Step 8: Calculate bits of LCU ($T_{LCU,R}$) by Eq. (12)
Step 9: Compensate the bit allocation mismatch, the $T_{LCU,R}$ is improved by Eq.(13)
Step 10: Calculate QP for the LCUs by Eq. (14)–(15)
Step 11: Clip the value of λ and QP for the current LCU by Eq. (21)–(22)
Step 12: Encode LCUs with calculated QP
Step 13: Update the parameters by Eq. (19)–(20)

5 Experimental results

5.1 Experiment setup

To evaluate the performance of the proposed perceptual importance-based RC algorithm, named Proposed in this paper. We incorporated the proposed method into the HEVC reference

software HM16.19 [16]; RDOQ and RDOQTS were disabled, but the remaining parts were the same as in HM16.19 [16]. The test conditions were set as two encoder configurations: lowdelay_P_main (LD) and randomaccess_main (RA). Thirteen standard test video sequences from classes B, C, D, and E were selected for evaluation and encoded under four QP values: 22, 27, 32, and 37. The first 300 frames of each test sequences were encoded. Further, the target bitrate was obtained by encoding the video sequences according to the same encoding configuration that disables rate control. We then compared the proposed method with several state-of-the-art RC algorithms, including the default RC scheme in HM 16.10, named LI [22]; three related spatial/temporal based perceptually RC methods: [14, 43, 44], named Wei [44], H Wei [43] and Gong [14]; and our previous work [48], named Ye [48].

5.2 R-D performance

The aim of this paper is to improve video quality; thus, the R-D performance is an important performance evaluation metric for the proposed method. The R-D performances of the methods were measured in terms of the Bjøntegaard delta peak signal-to-noise rate (BD-PSNR), which indicates quality improvement against the benchmark of the same coding bitrate. A positive value means performance gain while a negative value means performance loss. In our experiments, the default rate control scheme in HM 16.10, named LI [22] was set as the benchmark. The R-D curves of different RC sequence methods are shown in Fig. 2. For the difference between the curves, it can be seen that the proposed RC method outperforms other rate control methods at both high and low bitrates. In order to display the R-D performance of the proposed method clearly and comprehensively, the results of each test sequence for four different methods under LD and RA configurations are shown in Table 3. As seen from the table, compared to the benchmark method (LI [22]), the proposed method improves the BD-PSNR performance by 0.48 dB and 0.30 dB on average for LD and RA configurations, respectively. These improvement results are also better than the other methods. Specifically, the BD-PSNR improved by 0.60 dB for the test sequence “RaceHorses” under the LD configuration and 0.49 dB for the test sequence “Kimono1” under the RA configuration. For the “Johnny” sequence, the average BD-PSNR improvement of the proposed scheme is 0.28 dB higher than the method of H Wei [43], 0.36 dB higher than the method of Ye [48] and 0.12 dB higher than the method of Wei [44] under the LD configuration. For the “Kimono1” sequences, the average BD-PSNR improvement of the proposed model is 0.16 dB higher than the method of Gong [14], and 0.10 dB higher than the method of Wei [44] and 0.32 dB higher than the method of Ye [48] under the RA configuration. The reason of quality improvement is that: although Wei [44], H Wei [43] and Gong [14] are all use spatial/temporal information to guide bit allocation, but the spatial/temporal information of the proposed method is represent the perceptual importance better. Second, these methods are all not consider the bit balance of different regions that induce the significantly quality decrease of LPIR. In addition of that, Ye [48] only consider the spatial information to guide bit allocation, thus the R-D performance is not as well as other methods.

As PSNR does not always match perceptual quality very well, the perceptual importance weighted PSNR (EWPSNR) [50] is used to re-evaluate the perceptual video quality. Table 4 presented the BD-EWPSNR. From Table 4, due to the bits are optimally allocated according to perceptual importance, the BD-EWPSNR gaining by the proposed algorithm are all outperform other methods.

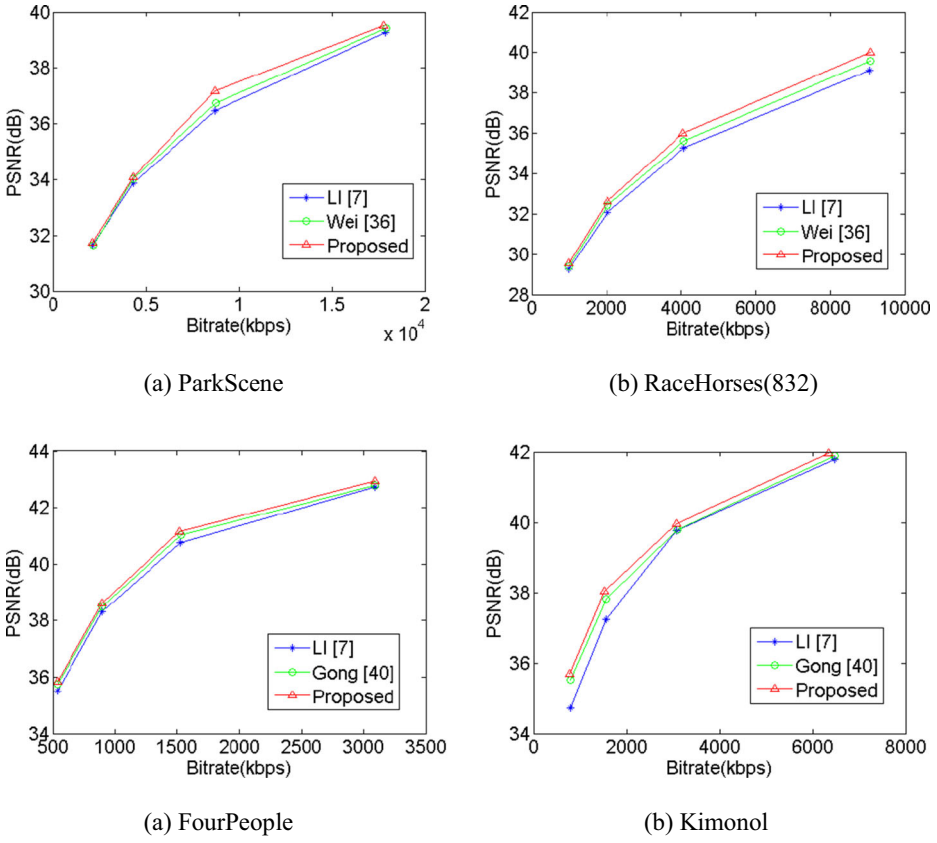


Fig. 2 R-D curves comparison. The (a) (b) are the cases of LD configuration, (c) (d) are the cases of RA configuration

5.3 Subjective quality comparison

As the video quality is finally evaluated by humans, subjective assessment such as the perceptual quality and structure similarity (SSIM) are all important video perceptual quality evaluation metric. The comparisons of the perceptual subjective quality for different sequences are shown in Fig. 3. The sequence "FourPeople" is encoded by QP 32 under LD configuration, and the sequence "Cactus" is encoded by QP 37 under RA configuration. Some selected regions are magnified for better comparison. It can be seen that the visual quality of texture and motion regions encoded by the proposed model is better than the conventional RC methods, especially in the selected regions. Especially, the bottom of the cactus in Fig. 3(b) has more detail than the corresponding part in Fig. 3(a); in addition that, the blocking artifacts on the face in Fig. 3(c) is more obviously than the corresponding part in Fig. 3(d).

Moreover, the SSIM of the proposed algorithm are much higher than those of HEVC anchor [22] direct encoding. There is because that compared to conventional RC methods the selected regions are allocated more bits by the proposed algorithm. Therefore, the visual quality of encoded videos can be improved effectively by the proposed algorithm. In addition, the SSIM results of each test sequence for four different methods under LD and RA configurations are shown in Table 5, where the data are average of four QPs corresponding

Table 3 R-D Performance Comparison In Terms of Y-Component BD_PSNR (dB) Against HEVC Anchor under LD and RA Configuration

Class	Sequence	Ye [48] (dB)		Wei [44] (dB)		H Wei [43] (dB)	Gong [14] (dB)	Propose (dB)	
		LD	RA	LD	RA	LD	RA	LD	RA
B	*BaskteballDrive*	0.11	0.04	0.17	0.11	0.23	0.18	0.26	0.18
	BQTerrace	0.39	0.13	0.52	0.18	0.52	0.22	0.61	0.29
	Cactus	0.27	0.16	0.31	0.20	0.31	0.27	0.45	0.33
	ParkScene	0.29	0.22	0.43	0.25	0.43	0.35	0.56	0.30
	Kimono1	0.41	0.17	0.79	0.39	0.79	0.33	0.62	0.49
C	*PartyScene*	0.14	0.19	0.20	0.02	0.32	0.11	0.36	0.16
	BQMall	0.10	0.15	0.24	0.14	0.27	0.18	0.37	0.21
	RaceHorses	0.21	0.30	0.39	0.27	0.53	0.32	0.63	0.41
	BasketballDrill	0.07	0.03	0.19	0.12	0.40	0.21	0.35	0.19
D	*BQSquare*	0.04	0.06.	0.13	0.07	0.28	0.06	0.33	0.18
	RaceHorses	0.30	0.19	0.41	0.27	0.34	0.28	0.60	0.37
	BlowingBubbles	0.28	0.21	0.39	0.30	0.48	0.32	0.58	0.39
E	*FourPeople*	0.13	*0.09*	*0.31*	*0.20*	0.47	0.24	0.39	0.37
	Johnny	0.23	0.17	0.47	0.29	0.31	0.25	0.59	0.36
	Average	**0.21**	**0.16**	**0.35**	**0.20**	**0.41**	**0.24**	**0.48**	**0.30**

bitrates. From Table 5, the average SSIM value is 0.9125 for our previous work Ye [48], 0.9197 for method of Wei [44] and 0.9249 for the proposed algorithm under LD configuration; the average SSIM value is 0.9027 for our previous work Ye [48], 0.9112 for method of Wei [44] and 0.9179 for the proposed algorithm under RA configuration. From the results, we can see the proposed algorithm outperforms Ye [48] and Wei [44] in the case of subjective quality. These results demonstrate that our method not only has better objective video quality, but also has better perceptual subjective quality when compared with other state-of-the-art RC methods.

Table 4 R-D Performance Comparison In Terms of Y-Component BD_EWPSNR (dB) Against HEVC Anchor under LD and RA Configuration

Class	Sequence	Ye [48] (dB)		Wei [44] (dB)		H Wei [43] (dB)	Gong [14] (dB)	Propose (dB)	
		LD	RA	LD	RA	LD	RA	LD	RA
B	*BaskteballDrive*	0.20	0.06	0.25	0.49	0.21	0.27	0.38	0.57
	BQTerrace	0.31	0.09	0.44	0.46	0.35	0.32	0.37	0.5
	Cactus	0.04	0.37	0.37	0.1	0.29	0.41	0.36	0.44
	ParkScene	0.30	0.30	0.19	0.23	0.42	0.33	0.39	0.52
	Kimono1	0.22	0.07	0.33	0.18	0.53	0.59	0.36	0.51
C	*PartyScene*	0.36	0.16	0.36	0.19	0.22	0.39	0.57	0.25
	BQMall	0.28	0.11	0.51	0.14	0.28	0.38	0.39	0.51
	RaceHorses	0.15	0.24	0.23	0.38	0.21	0.34	0.56	0.46
	BasketballDrill	0.20	0.24	0.14	0.31	0.53	0.46	0.61	0.55
D	*BQSquare*	0.33	0.04	0.38	0.33	0.36	0.40	0.42	0.22
	RaceHorses	0.30	0.32	0.22	0.23	0.32	0.30	0.52	0.42
	BlowingBubbles	0.28	0.36	0.47	0.16	0.37	0.46	0.41	0.30
E	*FourPeople*	0.34	0.05	0.36	0.29	0.5	0.59	0.51	0.47
	Johnny	0.33	0.34	0.34	0.5	0.29	0.17	0.42	0.27
	Average	**0.26**	**0.20**	**0.33**	**0.29**	**0.35**	**0.39**	**0.45**	**0.43**

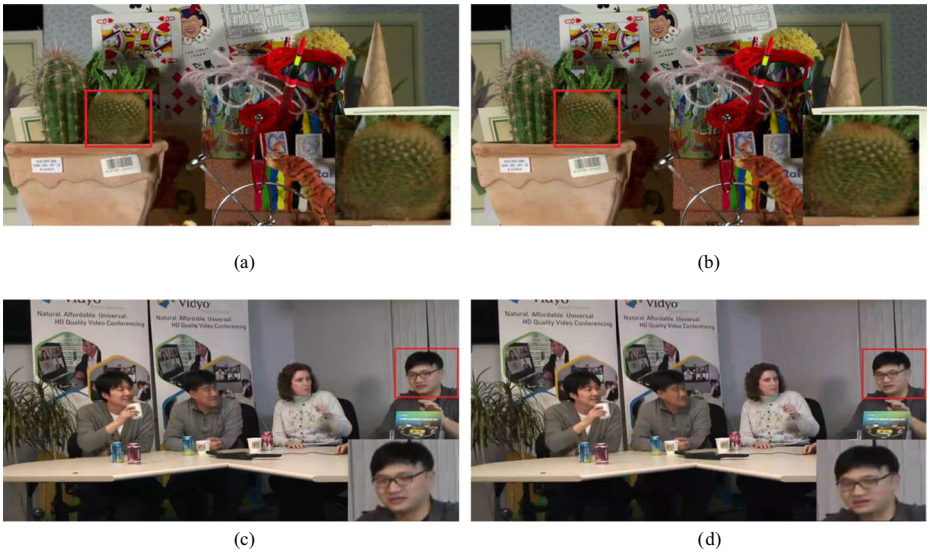


Fig. 3 Subjective results comparisons for HEVC anchor [22] and proposed method. (a) HEVC anchor [22] result for Cactus with QP = 37 under RA configuration, SSIM = 0.9036; (b) Proposed result for Cactus with QP = 37 under RA configuration, SSIM = 0.9137; (c) HEVC anchor [22] result for FourPeople with QP = 32 under LD configuration, SSIM = 0.9269; (d) Proposed result for FourPeople with QP = 32 under LD configuration, SSIM = 0.9428

Besides Fig. 3, we also adopted an assessment proposed by Rec. ITU-R BT.500 to further evaluates the subjective video quality. The assessment is called single stimulus continuous quality scale (SSCQS). Different video resolutions are tested in the assessment procedure. Before each assessment, the observers were required to watch 10 other training videos to help them better understand the video subjective quality assessment procedure. 10 observers were participant in this assessment. All the videos are displayed in their original resolutions, to avoid the influence of scaling operation. Note that the uncompressed reference and test video sequences were displayed with a random order. The quality rate scales for observers to evaluate video quality are excellent (10–8.1), good (8–6.1), fair (6–4.1), poor (4–2.1), and bad (2–0.1). After observers watched the video, difference mean opinion scores (DMOS) were

Table 5 SSIM Performance Comparison of Proposed Method Against Different Other Methods

Class	Sequence	Ye [48] (dB)		Wei [44] (dB)		Propose (dB)	
		LD	RA	LD	RA	LD	RA
B	*BQTerrace*	0.9104	0.8923	0.9158	0.9032	0.9201	0.9067
	Cactus	0.8841	0.8825	0.8923	0.8911	0.9013	0.9139
	ParkScene	0.8804	0.8831	0.8940	0.8980	0.8984	0.9008
	Kimono1	0.924	0.9187	0.9301	0.9262	0.9367	0.9346
C	*PartyScene*	0.9162	0.9039	0.9278	0.9163	0.9347	0.9219
	RaceHorses	0.9461	0.9198	0.9467	0.9289	0.9493	0.9334
D	*BQSquare*	0.9134	0.9083	0.9159	0.9116	0.9213	0.9187
	RaceHorses	0.8913	0.882	0.8973	0.8765	0.9011	0.8821
E	*FourPeople*	0.9369	0.9217	0.9432	0.9248	0.9479	0.9301
	Johnny	0.9223	0.9154	0.9345	0.9320	0.9384	0.9367
	Average	**0.9125**	**0.9027**	**0.9197**	**0.9112**	**0.9249**	**0.9179**

computed to reveal the difference of subjective quality between the compressed and uncompressed videos. The smaller value of DMOS corresponds to better subjective quality of the compressed video sequence. The Table 6 compares the average DMOS values of different compressed video sequences, all the values are average of four tested QPs corresponding bitrates. From this table, we can see that the DMOS values of our scheme are smaller than the perceptual URQ scheme, and much less than the conventional R-λ scheme, especially at high resolutions. Therefore, our scheme can provide higher subjective video quality. The works in [33] [40] demonstrate that there exit a perceptual redundancies degradation tolerance degree that people cannot perceive significant video quality differences. From Table 6, it can observed that the value of DMOS for the Class E video sequences are lower than other Classes video sequences for all methods. The reason is that, Class E test sequences are almost video conference scene, the characteristic of the type video is static background and slow moving objects, that means the perceptual redundancies degradation tolerance degree of this Class video is higher than other Classes.

5.4 Bitrate accuracy comparison

In addition to minimizing coding distortion, the key objection of RC is to make the output bitrate of a video encoder equal to the target bitrate as close as possible. Therefore, bitrate accuracy is another significant evaluation criterion of the RC algorithm, which is measured in terms of bitrate error (BitError). The BitError is defined as:

$$BitError = \frac{|R_{tar} - R_{act}|}{R_{tar}} \tag{23}$$

where R_{tar} and R_{act} are target and actual bitrates, respectively. The smaller the BitError value, the higher the bitrate accuracy achieved.

The averages of the four bitrate errors for each test sequence in the experiments are listed in Table 7. From this table, it can be seen that all four rate control algorithms obtained very high bitrate accuracies. For most of the test videos, our proposed method exceeded four other comparison methods for the two coding configurations. On average, the bitrate error of the proposed method is 0.37% and 2.19% and 2.63% less than Li [22], Wei [44] and H Wei [43] respectively, under the LD configuration. Similarly, 0.48%, 1.55% and 0.26% bitrate error

Table 6 Comparison In Terms Of DMOS Against HEVC and Other Methods

Class	Sequence	Li [22]		Wei [44]		Propose	
		LD	RA	LD	RA	LD	RA
B	*BQTerrace*	53.69	56.01	46.46	48.21	44.04	43.89
	Cactus	52.34	50.71	48.53	54.41	39.18	43.78
	ParkScene	47.69	53.15	48.54	40.86	42.18	38.14
	Kimono1	61.27	65.47	61.51	56.25	45.21	49.10
C	*PartyScene*	53.16	56.93	53.05	46.07	39.62	40.89
	RaceHorses	43.69	48.86	39.90	43.00	30.64	34.98
D	*BQSquare*	53.17	52.31	51.32	51.93	37.71	43.29
	RaceHorses	66.94	73.55	60.05	63.54	52.85	48.19
E	*FourPeople*	43.26	40.80	39.14	42.22	35.67	34.38
	Johnny	37.14	39.64	36.28	34.19	31.32	29.90
	Average	**51.23**	**53.74**	**48.48**	**48.07**	**39.84**	**40.65**

Table 7 Bitrate Accuracy Comparison In Terms Of Bitrate Error (%) Against HEVC Anchor Under LD And RA Configuration

Class	Sequence	Li [22] (%)		Wei [44] (%)		H Wei [43] (%)	Gong [14] (%)	Propose (%)	
		LD	RA	LD	RA	LD	RA	LD	RA
B	*BaskteballDrive*	1.07	1.71	1.32	0.91	*3.46*	0.03	*0.27*	0.47
	BQTerrace	0.32	0.34	2.48	2.19	*2.48*	1.04	*0.24*	0.21
	Cactus	0.54	0.21	3.79	3.06	*3.79*	0.31	*0.20*	0.14
	ParkScene	0.33	0.23	4.06	2.56	*4.06*	0.91	*0.13*	0.09
	Kimono1	0.90	0.89	3.75	1.22	*3.75*	0.08	*0.31*	0.13
C	*PartyScene*	0.71	0.57	1.14	0.76	*1.35*	0.35	*0.79*	0.41
	BQMall	0.66	1.09	2.25	2.02	*1.49*	0.80	*0.25*	0.16
	RaceHorses	0.46	0.69	0.74	0.49	*2.67*	0.04	*0.32*	0.37
	BasketballDrill	0.74	1.08	3.83	2.76	*3.47*	0.49	*0.16*	0.16
D	*BQSquare*	1.20	1.02	2.29	1.45	*1.90*	0.23	*0.47*	0.33
	RaceHorses	0.83	0.49	0.98	0.52	*2.38*	0.14	*0.27*	0.68
	BlowingBubbles	0.74	0.66	1.66	1.27	*3.64*	1.85	*0.36*	0.18
E	*FourPeople*	0.85	*1.61*	2.87	2.33	*2.87*	0.31	*0.52*	0.54
	Johnny	0.34	0.23	4.04	4.28	*4.04*	1.06	*0.17*	0.20
	Average	**0.69**	**0.77**	**2.51**	**1.84**	**2.95**	**0.55**	**0.32**	**0.29**

reduction were obtained under the RA configuration in comparison to Li [22], Wei [44] and Gong [14], respectively. As spatial/temporal based RC methods, Gong [14] achieved a lower bitrate error than Li [22] and Wei [44]. However, the bitrate error of proposed method is also less than that of Gong [14]. In general, the proposed RC method not only achieves better R-D performance, but also obtains the smallest bitrate error as the bitrate errors are only 0.32% and 0.29% on average under the LD and RA configurations, respectively. The first reason for this is that in our proposed RC method, the conventional LMS-based parameter updating procedure is replaced by the BFGS-based parameter updating method, which has a faster convergence speed in the parameter update process. This means that the updating procedure of the proposed method may achieve a higher global and faster convergence speed than the LMS based-method. The more important reason is that, the proposed perceptual based region level and LCU level bit allocation method can effectively balance the bit rate between HPIR and LPIR that can significantly reduce the bit shortage of bit allocation procedure and improve the bit accuracy of each LCU. Thus, the proposed method achieves more accurate bitrates than other state-of-the-art methods.

5.5 Complexity comparison

The additional computational cost of the proposed RC algorithm mainly comes from the perceptual importance weight decision. The computational complexity can be measured by the encoding time ratio of the proposed rate control algorithm against the RC method [22]. It is expressed as

$$\Delta T = \frac{\left|T_{Prop} - T_{Anch}\right|}{T_{Anch}} \times 100\% \tag{24}$$

where T_{Prop} and T_{Anch} denote the encoding time of the proposed algorithm and the RC method [22] anchor, respectively. If ΔT is greater than 100%, then the encoding complexity increases,

Table 8 Complexity Comparison of Proposed Method Against RC Method [22] Anchor Under LD And RA Configurations

Sequence	LD		RA	
	Wei [44] (%)	Propose (%)	Wei [44] (%)	Propose (%)
BQTerrace	5.26	8.14	1.65	6.01
Cactus	4.32	7.31	6.21	9.23
ParkScene	3.64	4.24	4.67	6.34
Kimono1	3.71	4.47	2.50	5.41
FourPeople	2.35	3.45	2.36	5.36
Johnny	3.21	4.13	1.49	3.72
Average	**3.75**	**5.29**	**3.15**	**6.01**

and vice versa. The average values of the encoding time of all test sequences were utilized to calculate ΔT, and the results are shown in Table 8. As seen from the table, the proposed RC method slightly increases the encoding time as the complexity of perceptual importance weight decision constitutes a small portion of the complexity of the entire encoding process. The average$\Delta T$of the proposed method is a little higher than the algorithm of Wei [44]. This is because the proposed method uses BFGS to update the RC parameters, which requires extra computational cost due to the use of iterative algorithm.

6 Conclusions

In this paper, we present a perceptual importance-based RC scheme for HEVC. Based on the HVS theory, a formulation of spatial and temporal perceptual importance is developed in a low complexity cost. A fusion method is then utilized to build a comprehensive perceptual importance model, which is formulated as a weighted factor to represent the perceptual importance of each LCU. Furthermore, a new RC scheme is designed from the region level to LCU level. To improve the bitrate accuracy of the proposed RC method, a BFGS-based parameter updating method is utilized to replace the conventional R-λ parameter updating procedure. The experimental results verified that the proposed RC method possesses better properties compared to the state-of-the-art RC methods. This superiority is both in the subjective visual quality and objective quality. Particularly, compared to the conventional HEVC RC R-λ model, the proposed method not only maintains a lower bit error, but it achieved 0.48 and 0.30 dB BD-PSNR gains under the LD and RA coding configuration, respectively. In addition, the proposed method only increases negligible coding complexity.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 62041109, No. 61861038), the Fundamental Research Funds for the Central Universities (Grant No. 31920210073, 31920180115, 31920190039) and Gansu Province Natural Sciences Fund (21JR1RA206).

References

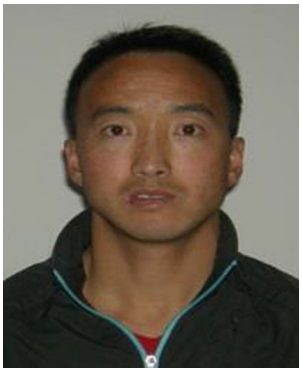
1. An C, Nguyen TQ (2008) Iterative rate-distortion optimization of H. 264 with constant bit rate constraint. IEEE Trans Image Process 17(9):1605–1615

2. Bai L, Song L, Xie R, Xie J, Chen M (Dec 2016) Saliency based rate control scheme for high efficiency video coding. In: *Proc. IEEE Int.* Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 1–6
3. Bazen AM, Gerez SH (2002) Systematic methods for the computation of the directional fields and singular points of fingerprints. IEEE Trans Pattern Anal Mach Intell 24(7):905–919
4. Bross B, (Oct. 2012) High efficiency video coding (HEVC) text specification draft 9 (SoDIS). In: 11th JCT-VC meeting
5. Chadha A, Andreopoulos Y, (2021) Deep Perceptual Preprocessing for Video Coding," In: *Proc. IEEE Int.* Computer Vision and Pattern Recognition (CVPR), pp. 14852–14861
6. Chen Z, Pan X (2019) An optimized rate control for low-delay H. 265/HEVC. IEEE Trans Image Process 28(9):4541–4552
7. Choi H, Nam J, Yoo J, Sim D, Bajic IV, (2012) Rate control based on unified RQ model for HEVC, ITU-T SG16 contribution, JCTVC-H0213, 1–13
8. Choi H, Nam J, Yoo J, Sim D, Bajic IV, (April 2012) Improvement of the rate control based on pixel-based URQ model for HEVC, ITU-T/ISO/IEC JCT-VC Document In JCT-VC I0094
9. Dong J, Ling N (2009) A context-adaptive prediction scheme for parameter estimation in H. 264/AVC macroblock layer rate control. IEEE Trans Circuits Syst Video Technol 19(8):1108–1117
10. Gao Y, Zhu C, Li S, Yang T (2017) Temporally dependent rate-distortion optimization for low-delay hierarchical video coding. IEEE Trans Image Process 4457-4470:4457–4470
11. Gao W, Kwong S, Jia Y (2017) Joint machine learning and game theory for rate control in high efficiency video coding. IEEE Trans Image Process 26(12):6074–6089
12. Gao W, Kwong S, Jiang Q, Fong CK, Wong PH, Yuen WY (2018) Data-driven rate control for rate-distortion optimization in HEVC based on simplified effective initial QP learning. IEEE Trans Broadcasting 65(1):94–108
13. Girod B (1993) What's wrong with mean-squared error? Digital images and human vision[J]. AB Watson ed, pp 207–220
14. Gong Y, Wan S, Yang K, Wu HR, Liu Y (2019) Temporal-layer-motivated lambda domain picture level rate control for random-access configuration in H.265/HEVC. IEEE Trans Circuits Syst Video Technol 29(1):156–170
15. Guo H, Zhu C, Xu M, Li S (2019) Inter-Block Dependency-Based CTU Level Rate Control for HEVC. *IEEE Trans Broadcasting* 66(1):113–126
16. svn_HEVCSoftware. HM Reference Software 16.19. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.19/. Accessed 2018
17. ITU-R, Methodology for the subjective assessment of the quality of television pictures, ITU-R Recommendation BT.500–10. 2000.
18. Lee JS, Ebrahimi T (2012) Perceptual video compression: A survey. IEEE J Sel Top Signal Process 6(6): 684–697
19. Lee B, Kim M, Nguyen TQ (2013) A frame-level rate control scheme based on texture and nontexture rate models for high efficiency video coding. IEEE Trans Circuits Syst Video Technol 24(3):465–479
20. Li B, Li H, Li L, Zhang J, (Oct. 2012) Rate control by R-lambda model for HEVC, ITU-T/ISO/IEC JCT-VC document JCTVC-K0103, Shanghai, CN
21. Li B, Li H, Li L, (2013) Adaptive bit allocation for R-lambda modelrate control in HM, ITU-T/ISO/IEC JCT-VC document JCTVC-M0036, Incheon, Korea
22. Li B, Li H, Li L, Zhang J (2014) λ domain rate control algorithm for high efficiency video coding. IEEE Trans Image Process 23(9):3841–3854
23. Li Y, Liao W, Huang J, He D, Chen Z, (July 2014) Saliency based perceptual HEVC. In: *Proc. IEEE Int. Multimedia and Expo Workshops (ICMEW),* pp. 1–5.
24. Li S, Xu M, Deng X, Wang Z (2015) Weight-based R-λ rate control for perceptual HEVC coding on conversational videos. Signal Process Image Commun 38:127–140
25. Li L, Li B, Li H, Chen CW (2016) λ-Domain optimal bit allocation algorithm for high efficiency video coding. IEEE Trans Circuits Syst Video Technol 28(1):130–142
26. Li S, Xu M, Wang Z, Sun X (2016) Optimal bit allocation for CTU level rate control in HEVC. IEEE Trans Circuits Syst Video Technol 27(11):2409–2424
27. Liang X, Wang Q, Zhou Y, Luo B, Men, A, (Nov 2013) A novel RQ model based rate control scheme in HEVC. In: *Proc. IEEE Int. Visual Communications and Image Processing (VCIP)*, pp. 1–6
28. Lim KP, Sullivan G, Wiegand T, (2005) Text description of joint model reference encoding methods and decoding concealment methods. JVT-O079, Busan, Korea
29. Lin H, He X, Teng QZ, Fu W, Xiong S (2016) Adaptive bit allocation scheme for extremely low-delay intraframe rate control in high efficiency video coding. *Journal of Electronic Imaging 25*(4):043008

30. Liu Y, Li ZG, Soh YC (2006) A novel rate control scheme for low delay video communication of H. 264/AVC standard. IEEE Trans Circuits Syst Video Technol 17(1):68–78
31. Ma YF, Zhang HJ, (Nov 2003) Contrast-based image attention analysis by using fuzzy growing. In: *Proc. the eleventh ACM international conference on Multimedia*, pp. 374–381
32. Meddeb M, Cagnazzo M, Pesquet-Popescu B, (2014) Region-of-interest-based rate control scheme for high-efficiency video coding. APSIPA Transactions on Signal and Information Processing, 3
33. Nami S, Pakdaman F, Hashemi MR, (2020) Juniper: A Jnd-Based Perceptual Video Coding Framework to Jointly Utilize Saliency and JND. In: *Proc. IEEE Int. Multimedia & Expo Workshops (ICMEW),* July 2020, pp. 1–6.
34. Oh H, Kim W (2012) Video processing for human perceptual visual quality-oriented video coding. IEEE Trans Image Process 22(4):1526–1535
35. Ohm JR, Sullivan GJ, Schwarz H, Tan TK, Wiegand T (2012) Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC). IEEE Trans Circuits Syst Video Technol 22(12):1669–1684
36. Ou YF, Ma Z, Liu T, Wang Y (2010) Perceptual quality assessment of video considering both frame rate and quantization artifacts. IEEE Trans Circuits Syst Video Technol 21(3):286–298
37. Park S, Kim M, (Nov 2006) Extracting moving/static objects of interest in video," In: *Proc.* Pacific-Rim Conference on Multimedia, pp. 722–729
38. Ribas-Corbera J, Lei S (1999) Rate control in DCT video coding for low-delay communications. IEEE Trans Circuits Syst Video Technol 9(1):172–185
39. Seshadrinathan K, Bovik AC (2009) Motion tuned spatio-temporal quality assessment of natural videos. IEEE Trans Image Process 19(2):335–350
40. Takeuchi M, Saika S, Sakamoto Y, Nagashima T, Cheng Z, Kanai K, Wei X (2018) Perceptual quality driven adaptive video coding using JND estimation. In: *Proc. IEEE Int. Picture Coding Symposium (PCS),* pp. 179–183
41. Wang Z, Zeng H, Chen J, Cai C (June 2014) Key techniques of high efficiency video coding standard and its extension. In: *Proc.* IEEE Int. Industrial Electronics and Applications, pp. 1169–1173.
42. Wang H, Song L, Xie R, Luo Z, Wang X (May 2018) Masking Effects Based Rate Control Scheme for High Efficiency Video Coding. In: *Proc. IEEE Int.* Symposium on Circuits and Systems (ISCAS), pp. 1–5
43. Wei H et al (2018) A Rate Control Algorithm for HEVC Considering Visual Saliency. *asia pacific signal and information processing association annual summit and conference,* 36–42.
44. Wei H, Zhou W, Zhou X, Bai R, Duan Z (2018) Saliency-based coding tree unit-level rate control for high-efficiency video coding. J Electron Imaging 27(4):043009
45. Wiegand T, Schwarz H, Joch A, Kossentini F, Sullivan GJ (2003) Rate-constrained coder control and comparison of video coding standards. IEEE Trans Circuits Syst Video Techno 13(7):688–703
46. Wong CW, Au OC, Meng B, Lam HK (2003) Perceptual rate control for low-delay video communications. Proc IEEE Int Multimedia Expo 3:III–361
47. Yang Z, Xu Q, Bao S, Cao X, Huang Q (2021) Learning with Multiclass AUC: Theory and Algorithms. IEEE Transs Pattern Anal Machine Intelligence PP:1
48. Ye Y, He X, Teng Q, Qing L, Lin H, Xia D (2018) Adaptive gradient information and BFGS based inter frame rate control for high efficiency video coding. Multimed Tools Appl 77(12):14557–14577
49. Zeng H, Yang A, Ngan KN, Wang M (2016) Perceptual sensitivity-based rate control method for high efficiency video coding. Multimed Tools Appl 75(17):10383–10396
50. Zhang W, Martin RR, Liu H (2017) A saliency dispersion measure for improving saliency-based image quality metrics. IEEE Trans Circuits Syst Video Technol 28(6):1462–1466
51. Zhou M, Wei X, Wang S, Kwong S, Fong CK, Wong P, Gao W (2019) SSIM-based global optimization for CTU-level rate control in HEVC. IEEE Trans Multimedia 21(8):1921–1933
52. Zhu C, Huang Y, Xie R, Song L (2021) HEVC VMAF-oriented Perceptual Rate Distortion Optimization using CNN. In: *Proc. IEEE Int. Picture Coding Symposium (PCS),* pp. 1–5.

Hongwei Lin received the MS degree in communication and information System from Xidian University, China in 2011. In 2019, he received his Ph.D. degree in communication and information system from the Sichuan University, China. He is a lecturer with the Electrical Engineering at Northwest Minzu University. His main research interests include image processing, video compression and communication. E-mail: linhongwei@xbmu.edu.cn

Xiangqun Li received the Ph.D. degree in electronic information science and technology from Sichuan University, China in 2017. Now, He is a lecture with the College of Electrical Engineering at Northwest Minzu University. His main research interests include image processing, video coding and transmission. E-mail: xiangqunli@xbmu.edu.cn



Mingliang Gao received Ph.D. degree in electronic information science and technology from Sichuan University, China in 2017. He is currently an Associate Professor with the College of Electrical Engineering at Northwest Minzu University. His research interests include 2D/3D digital image processing, image communication, and pattern recognition. E-mail: mingliang_Gao@xbmu.edu.cn

Keyan Deng received the M.S. degree from the department of electronic and information engineering, Lanzhou Jiaotong university, Lanzhou, China in 2006; the Ph.D. degree from the Institute for Signals and Information Processing, Lanzhou University, Lanzhou, China in 2017. His research interests include video coding and wireless communications. E-mail: dky0603@163.com

Yongsheng Xu received M.S. degree in control theory and control engineering from Lanzhou University of Technology in 2004. He is currently a senior experimentalist with the College of Electrical Engineering at Northwest Minzu University. His main research interests include image inpainting, and 2D/3D digital image processing. E-mail: 563359741@qq.com