IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 10, 2021, accepted November 25, 2021, date of publication December 3, 2021,
date of current version December 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3132294

Deep Convolutional Feature-Driven Rate Control for the HEVC Encoders

ISMAIL MARZUKI, JONGSEOK LEE, WAHYU WIRATAMA, AND DONGGYU SIM, (Senior Member, IEEE)
Department of Computer Engineering, Kwangwoon University, Seoul 01897, South Korea

Corresponding author: Donggyu Sim (dgsim@kw.ac.kr)

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Centre (ITRC) support program supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP) under Grant IITP-2021-2016-0-00288, and in part by the Research Grant of Kwangwoon University in 2020.

ABSTRACT This work proposes a rate control model based on deep convolutional features to improve the video coding performance of the HEVC encoders under the random access (RA) configuration. The proposed algorithm extracts high-level features from the original and previous coded frames using a pretrained visual geometry group (VGG-16) model by considering characteristics of a different temporal layer for the RA configuration. Subsequently, R–λ parameters (alpha and beta), bit allocation, λ estimation, and quantization parameter decision at frame-level are formulated by utilizing the extracted high-level features to maintain video quality and bitrate accuracy control. In addition, bit allocation at the group-of-picture (GOP)-level rate control is proposed with perceptual-based thresholding to control smooth bitrates and visual quality between adjacent GOPs. The results verify that the proposed algorithm is efficient in coding performance and bit accuracy by keeping visual quality. Compared with the existing R–λ rate model in HM-16.20, the proposed models can achieve an average BD-rate gain of −4.39% and −8.74% in PSNR and MSSSIM metrics for the RA configuration, respectively.

INDEX TERMS Deep neural network, high efficiency video coding (HEVC), perceptual video coding, rate control, video coding.

I. INTRODUCTION

The media services on the wired and wireless internet, such as video streaming and video communication, have gained popularity. More people at diverse locations and situations access media content and services. Thus, delivering high-quality video content while minimizing rates over heterogeneous networks is a consistent requirement of content providers and consumers. In video coding, the rate control has been widely known to play a critical role in transmitting high-quality video under a particular network bandwidth and limited buffer capacity. Hence, the rate control algorithm that can serve high video quality under the bandwidth constraint designated for such video applications becomes essential.

Many rate control studies exist for high-efficiency video coding (HEVC) encoders [1]–[5]. During the standardization process of HEVC, the R–λ rate control model has been adopted for the HEVC encoders [2], [3]. This model aims to (1) control the high accuracy of bitrates, (2) maintain rate-distortion performance, and (3) optimize visual quality by appropriately allocating bits to different rate control stages: group-of-picture (GOP)-level, frame-level, and block-level at a given bitrate. However, to this day, it is challenging to find rate control algorithms designed for these video applications targeting perceptual quality optimization. In HEVC, a random-access (RA) configuration is introduced for the typical video streaming and broadcasting applications. Generally, the RA configuration in HEVC employs a dyadic high-delay hierarchical B-prediction structure. Frames in a video sequence are arranged into different temporal layers for a higher compression rate and visual quality [6]. Such a hierarchical B-prediction structure and advanced coding tools [7]–[14] make developing a RA configuration rate control more challenging.

In contrast, the current rate control model in HEVC for the RA configuration has been more stable than the other coding configurations. However, some limitations of the existing rate control under this configuration need to be addressed

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

appropriately. The bitrate allocation for each frame could still be misallocated. It turns into failing the bitrate accuracy control and causes an imbalance in a compressed video visual quality. Previous studies reveal that the existing rate control model miscarries several rate control parameters that degrade rate-distortion (RD) performance, bit accuracy, and subjective video quality [15]–[21]. The human visual system (HVS) supports that there is much perceptual redundancy in videos [57]. However, the current rate control model in HEVC encoder reference software, e.g., HM-16.20, does not entirely consider perceptual characteristics.

To this end, this work proposes a deep neural network (DNN) feature-based GOP-level and frame-level for the HEVC encoder by extending a perceptual adaptive quantization parameter (QP) selection algorithm [22]. There have been many attempts to leverage the capability of DNN for perceptual quality purposes [22]–[28]. A common conception is that the DNN-based extracted features are primarily implicated in capturing global characteristics of the input images appealing for adaption perceptual quality problems. At the frame-level, the proposed algorithm formulates an estimation model for α and β parameters based on high-level visual features extracted from the original and reconstructed frames using a pretrained visual geometry group (VGG-16) network model [29]. It is expected to improve the quality of α and β decisions, affecting the R–λ relationship quality. Then, the proposed algorithm redefines the R–λ estimation for each frame to minimize distortion and save more bitrates without inducing perceptible visual depressions in the compressed video frames.

Further, the hierarchical B-prediction structure in the RA configuration is considered when extracting the visual features and taking other coding parameters, i.e., temporal layer id, initial QP value at the frame level, and base QP value. These coding parameters maintain the λ value that can improve visual quality. Consequently, adjustment on the QP estimation is also determined based on the proposed visual features. Moreover, a more perceptual-friendly bit allocation model at the GOP-level is proposed by considering the frame rate of sequence information and proposing a perceptual-based threshold to control bitrates smoothness between the adjacent GOPs. Thus, the proposed algorithm demonstrates higher coding gain evaluated under PSNR and MSSSIM by approximately -4.39% and -8.74%, respectively, with some visual quality improved than HM-16.20.

The remainder of this paper is organized as follows. Section 2 briefly discusses rate control in HM and related works. Section 3 details the proposed rate control. In Section 4, an evaluation of the proposed algorithm is described. The conclusion of the work is presented in Section 5.

II. CURRENT STATE OF THE R–λ MODEL

This section discusses the existing R–λ rate control model in HM-16.20 for RA configuration. The λ value at frame-level in the R–λ model [2], [3] is estimated as

$$\lambda = \alpha \times bpp^{\beta}. \tag{1}$$

The bpp term is the estimated bit per pixel, whereas α and β at the frame-level rate control are set as $+6.7542$ and $+1.7860$ for I-frame, and other frames are $\alpha = +3.2003$ and $\beta = -1.367$. The original studies [2], [3] claimed that these values were provided based on video content. However, studies have discovered [30]–[32] that the existing R–λ model in the HM software is not optimal for the coding structure for the RA configuration.

A. EXISTING R–λ RATE CONTROL MODEL IN THE RA CONFIGURATION OF HM ENCODER

The first limitation is that the R–λ rate control model tends to consume more bits for frames at the beginning of a sequence. Since more bits are used to encode the first few frames, fewer bits are available to compress the rest of the frames in a sequence. However, this condition influences the overall rate control performance. Figure 1 depicts the frame-by-frame quality of bitrate, PSNR, and MSSSIM for the "BQMall" generated from HM-16.20 with the rate control enabled for the RA configuration. The PSNR and MSSSIM qualities of each frame are not well maintained when the model attempts to code several last frames. It makes sense that a lack of bitrate forces the encoder to code a frame with a high QP and generate a low quality of a compressed video.

Figure 2 shows the reconstructed frames by the HEVC encoder with the activated rate control. The quality of reconstructed frames is progressively degraded as the available bitrate is getting limited. This figure reflects that an adequate bit allocation for a frame may produce a higher visual quality subjectively than those allocated with insufficient bitrates. This phenomenon exists because the R–λ rate control model fails to provide appropriate α and β, leading to inaccurate bit allocation [16], [30] and deteriorating λ and quality of the subsequent frames [18]–[21]. For further investigation, Table 1 lists the actual bitrates for several test sequences generated by the current rate control model in HM-16.20 for the RA configuration. Notably, the target bitrate is provided to assess how well the model can control the bitrate accuracy with different QP settings. The smaller the bitrate accuracy value is, the better the improvements of the rate control algorithm. However, the rate control bitrate accuracy in HM-16.20 is not well maintained, especially when the test sequences are tested using higher QP settings, such as in QP 32 and 37. In essence, the quality of α, β, λ, bit allocation, and QP decisions impact the whole performance of the rate control model. Hence, it is crucial to reevaluate how Eq. (1) can provide optimum HEVC encoder performance.

B. SOME EFFORTS TO IMPROVE THE EXISTING RATE CONTROL MODEL IN THE RA CONFIGURATION

Many studies have attempted to improve the rate control in HEVC. For GOP-rate control, Wu *et al.* [33] proposed a

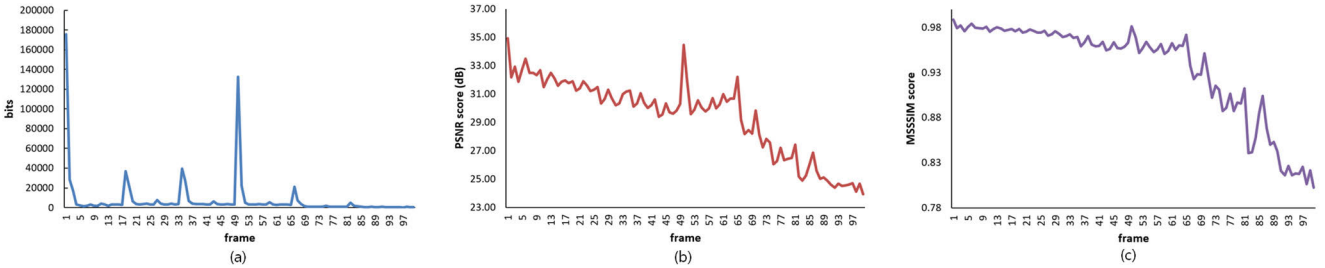


FIGURE 1. Generated (a) bitrate, (b) PSNR, and (c) MSSSIM of the "BQMall" test sequence by the rate control in HEVC.

FIGURE 2. Subjective quality of the "BQMall" test sequence with the rate control in HEVC enabled.

TABLE 1. Comparisons of bpp–λ and the proposed bpp_{CTU}–λ.

Sequence	QP	Target bitrate (Kbps)	Actual bitrate (Kbps)	Bitrate accuracy (%)
C_BasketballDrill	22	3449.62	3528.12	2.26
	27	1661.61	1756.22	5.67
	32	814.32	897.41	10.25
	37	431.97	474.86	9.92
C_PartyScene	22	6811.48	6967.83	2.30
	27	3102.90	3274.88	5.54
	32	1460.49	1600.17	9.60
	37	690.52	775.69	12.26
C_RaceHorses	22	4787.17	4793.99	0.15
	27	2024.75	2076.28	2.53
	32	943.85	991.01	4.98
	37	462.78	494.53	6.81
D_BlowingBubbles	22	1639.64	1658.76	1.14
	27	752.66	788.76	4.75
	32	351.46	384.38	9.51
	37	163.58	187.18	14.14

rate control by exploiting the R-Q model for HM-13.0. The proposed algorithm is designed by considering the temporal prediction structure in HEVC. Although this algorithm aims to tackle only GOP-level rate control, determination on the QP value for the first frame and the coding tree unit (CTU)-level rate control is also presented. However, the proposed algorithm can merely favor a small coding gain. To reiterate the aforementioned Q-domain rate control problem, it still suffers from a general "chicken and egg" dilemma. Rodriguez and Schierl [34] proposed a buffer-constrained rate control for real-time HEVC with the hierarchical GOP structure in HM-9.0. This algorithm operates on three layers, namely the intra period layer, picture-level, and CTU-level. However, this algorithm also runs under the Q-domain rate control model, which might suffer from similar general Q-domain rate control problems. In addition, the same limitation applies to the Wang *et al.* [35] algorithm designed based on the R-Q model rate control for HM-8.0.

For the frame-level rate control, Pan *et al.* [36] proposed a frame-level rate control based on visual characteristics of the input video for HEVC. This algorithm estimates the grey-level co-occurrence matrix to determine the relationship between the visual video characteristics and rate. However, this algorithm is designed only for the Low-Delay-P configuration, which has a more straightforward hierarchical picture structure than the RA configuration. Gong *et al.* [31] proposed a temporal-layer-motivated R–λ rate control at the frame-level for the RA configuration in HM-14.0. Their work

argues that the R–λ rate control model in HM is not optimal for this configuration. Accordingly, this algorithm facilitates a novel λ estimation by considering the frame-level motion difference. Distinguishing these sequences with slow and fast motions indicates this algorithm also designed the λ estimation, including λ and QP clipped function separately. Zhang *et al.* [37] proposed a frame-level rate control algorithm by observing the GOP-level rate control model quality. However, the proposed algorithm is designed only for low delay hierarchical GOP structure. This algorithm proposed a more adaptable quality dependency model for the current GOP by analyzing the low delay reference structure. Subsequently, the current GOP quality model estimates the frame-level bit allocation with a global rate-distortion optimization formulation built under the total generated bitrate constraint of a GOP. Similarly, Guo *et al.* [38] proposed a frame-level and GOP-level rate control for a low delay structure. The RD characteristics at the frame-level of the encoded frames at the same temporal layer position in adjacent GOPs are analyzed to formulate bitrate estimation for the GOP-level. Then, a global RD optimization was then proposed based on a recursive Taylor expansion model to cover the Lambda estimation, which is then used to benefit the bit budget at frame-level rate control.

For the CTU-level rate control, there have also been many works attempting to improve the rate control performance of the HEVC encoders by considering the perceptual characteristics of CTU blocks. Lim and Sim [39] proposed a luminance adaptation characteristic based on a pixel domain just noticeable difference model to determine bit allocation and QP value for each CTU in a frame. However, the model tends to have a similar performance with the anchor R–λ rate control model. Guo *et al.* [40] proposed an inter-block dependency model to improve the CTU-level rate control by estimating some propagation factors of each 16 × 16 block. However, the algorithm was designed only for a low delay coding structure, which may have some difficulties applying it for RA configuration. Zhou *et al.* [41] formulated a clip function of λ and QP for the CTU-level rate control by proposing a visual difference predictor model for a high dynamic range input video. Raufmehr and Rezaei [42] designed a fuzzy rate controller-based for the scalable HEVC. However, both algorithms were proposed for different use cases from the proposed algorithm. Bosse *et al.* [43] proposed a distortion sensitivity model based on a deep neural network to estimate bitrate at CTU-level. The work shows a significant improvement compared with a constant QP setting under the 'all-intra' configuration of the HEVC encoder. However, it does not provide evaluations against the rate control settings and much information on the neural network architecture. Li *et al.* [44] extracted characteristics of each CTU block to estimate allocation bits at the CTU-level applied for the 360-degree video.

This work proposes novel estimation models for α, β, λ, bit allocation, and QP for the R–λ rate control model in HM-16.20. A deep convolutional feature-based GOP- and frame-level rate control are proposed for the RA configuration. The DNN model for video coding has gained much attention in the video coding community [45]–[52]. Thus, the proposed algorithm obtains some advantages of using high-level perceptual features constructed from the original and reconstructed frames using a predefined VGG-16 model. The proposed algorithm achieves a higher compression rate and improves the compressed quality video against HM-16.20 with the rate control.

III. PROPOSED RATE CONTROL FOR THE RA CONFIGURATION IN THE HEVC ENCODER

As discussed in Section 2, the existing R–λ rate control in the HM software is not optimal for the RA configuration coding structure. From formula (1), the R–λ rate control model is expressed with three main factors, namely the estimated bpp, α, and β parameters. In the frame-level rate control, α and β are initialized differently from those at the CTU-level rate control. Although the original work [2], [3] committed that α and β were based on the characteristics of test video sequences. However, α and β in the existing frame-level rate control cannot correctly impact the bpp–λ relationship. Inaccurate α and β parameters are responsible for inaccurate bitrate control and achieving minimal rate-distortion, as depicted in Figure 1, Figure 2, and Table 1.

This work determines the values of α and β for the frame-level rate control by considering the high-level features of a particular convolutional layer of the VGG-16 architecture. From previous work in [30], it is observed that a strong interrelationship between α–λ and β–λ is vital for improving the quality of the bpp–λ relationship. Consequently, it will also result in increasing the reconstructed video quality subjectively. Specifically, a pre-trained VGG-16 model is aimed to extract perceptual features from the original and reconstructed frames to address the estimation problems at the existing rate control, including the estimation of bit allocation at GOP- and frame-level, parameters of α and β, λ and QP values. In addition, the initial QP of a frame based on [22] is involved for the proposed frame-level rate control to handle the frame-level initial QP decision. Both spatial and temporal perceptual features are also accustomed to these estimation problems to satisfy the RA configuration coding structure. Notably, this proposed algorithm is designed mainly for the B-frame type. Therefore, it is required to check whether the frame being coded is an I-frame or B-frame when estimating the α, β, and λ. The I-frame type is assessed using the existing model as in HM-16.20, whereas the B-frame is based on the proposed ones. Finally, the QP value can be determined based on the design of the proposed algorithm. The CTU-level rate control and the rest process in the unit encoding are conducted as those in HM-16.20.

A. VISUAL FEATURE EXTRACTION BASED ON THE PRE-TRAINED VGG-16 NETWORK

This work begins by using the $D_{VGGfeature}$ symbol for referring to the proposed perceptual loss value for each frame.

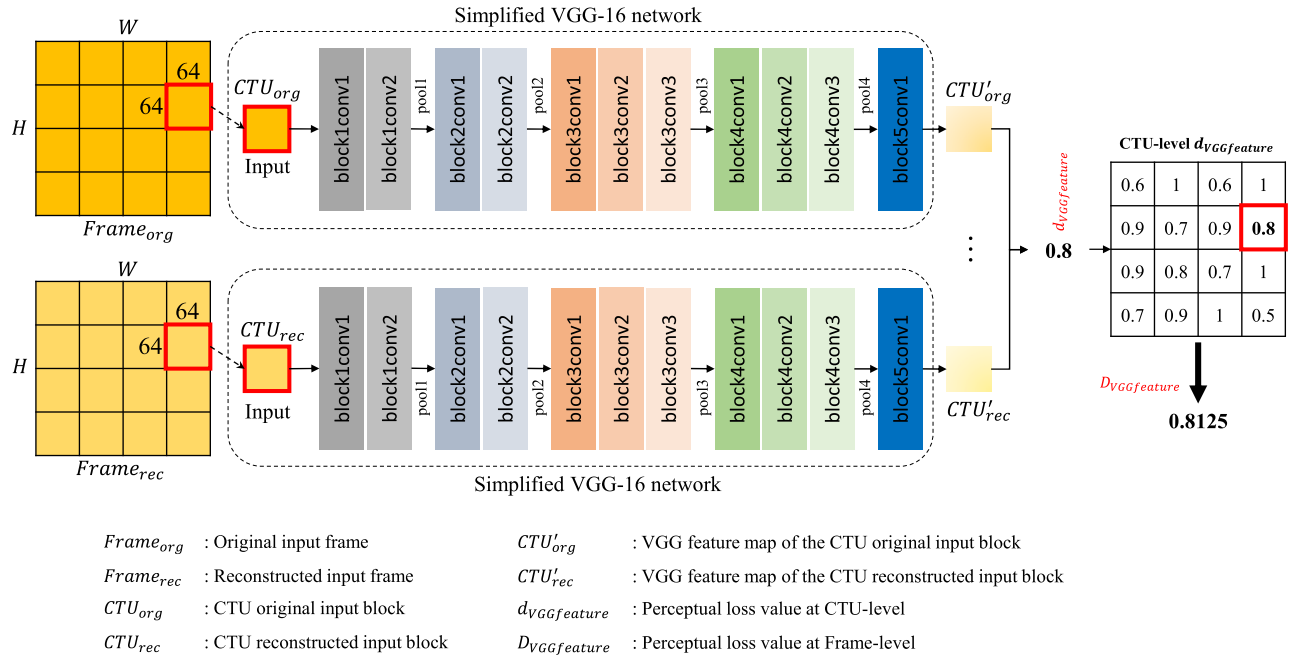


FIGURE 3. Perceptual loss generation based on the double-simplified VGG-16-based architecture from the original and reconstructed frames.

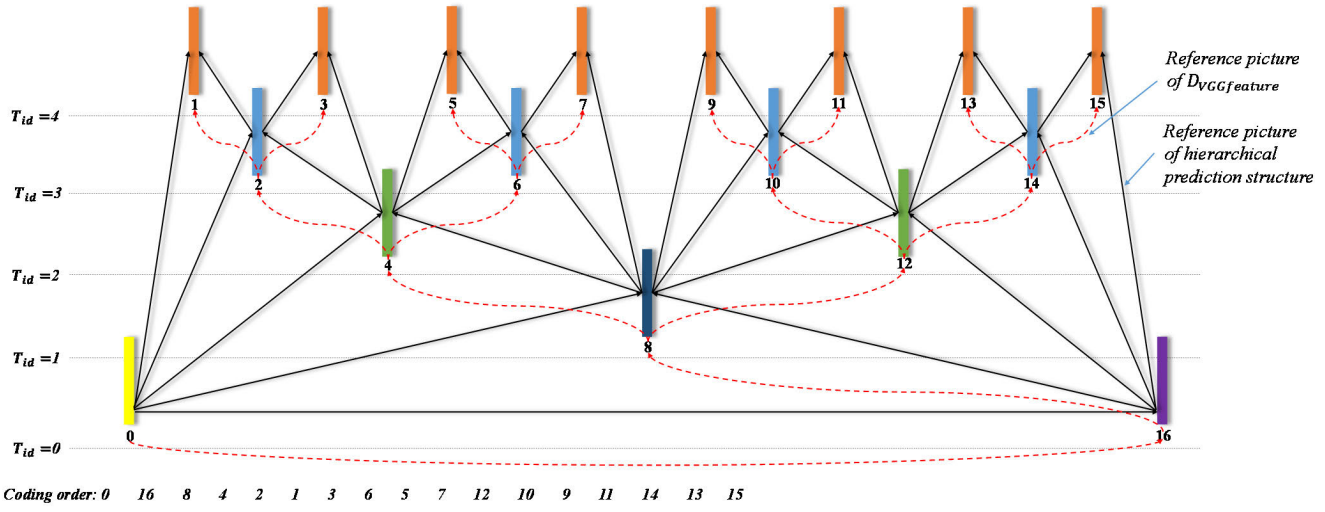


FIGURE 4. Illustration of the $D_{VGGfeature}$ employment on the dyadic B-hierarchical prediction structure of the RA configuration.

$D_{VGGfeature}$ is the averaged value of all perceptual loss values of CTUs within a frame estimated from the high-level visual features extracted using the VGG-16 network. The proposed algorithm explores the perceptual loss for a frame $D_{VGGfeature}$ to address some issues of the estimation models at GOP- and frame-level rate control scheme in HM-16.20 for the RA coding structure. In this work, $d_{VGGfeature}$ denotes the perceptual loss generated from a collocated CTU position between the original and reconstructed frames based on the double-simplified VGG-16 network, as depicted in Figure 3 and detailed in [22]. Notably, $d_{VGGfeature}$ is conjectured using the Euclidean distance as a perceptual loss function in the range 0 to 1 to reflect the human HVS attribute. Furthermore, $D_{VGGfeature}$ is fashioned by considering the RA coding configuration temporal structure in the HEVC encoder for a better HVS attribute deliberation in the proposed algorithm, as illustrated in Figure 4.

The proposed double-simplified VGG-16 framework is designed based on the full-reference visual quality approach by excluding the 'pool5' and 'fully connected' layers from the VGG-16 network. Both layers are traditionally favored by specific classification objects, which intuitively influence the quality of generated visual features. Consequently, the proposed framework expects merely developed features

from the last block of the convolution layer 'block5conv1' in Figure 3. A pre-trained VGG-16 network trained on the ImageNet dataset is directly used without separately training it on the network. The VGG-16 network has been well-known to produce a high-quality visual feature for many computer vision applications. To this end, the extremely deep convolutional layer of VGG-16 is exerted for each CTU block of the original and reconstructed frames. Then, it is required to identify universal patterns and generalize them for the perceptual features. Notably, the network is prepared for the input with an RGB color format.

B. BIT ALLOCATION FOR THE GOP-LEVEL RATE CONTROL ALGORITHM

In the HM-16.20 software, the bitrate allocation for GOP is governed by a fixed smooth window size parameter *(SW)* set to 40, which applies to all GOP cases in a sequence. Notably, the *SW* parameter is aimed to supervise the smoothness of bitrates between the adjacent GOPs. However, in HEVC, each test sequence may have different frame rates and varying target bitrates. Accordingly, a static *SW* parameter will not be perceptually canny for different video test conditions with a distinct visual characteristic.

The proposed bit allocation of GOP-level rate control T_{GOP}, expressed in Equation (2), attempts to consider a more visually appealing approach by assessing the perceptual loss value of the last coded frame in a GOP. In contrast to designing the entire bit allocation of the existing model, the proposed algorithm extends the current model by utilizing the frame rate *FR*, the remaining frames going to be encoded *FL*, the average frame bitrate for the entire frames in a sequence R_{picAvg}, the remaining bits after encoding the previous GOP $R_{bitLeft}$, and GOP size information of each test sequence, when the λ value of the last coded frame at a GOP λ_{last} is less than the given threshold. λ_{last} defined in Equation (3) is introduced as a perceptual-based threshold with similar goals as *SW* in HM-16.20 to control the smoothness of bitrates between the adjacent GOPs. However, this control function for the proposed algorithm is provided to be more adaptive than the existing smooth window size parameter in HM-16.20. λ_{last} is computed based on the $D_{VGGfeature}$ value, λ value of the previously coded frame $\lambda_{prevPic}$, and λ value of the current frame $\lambda_{currPic}$. R_{picAvg} in the proposed bit allocation of the GOP-level rate control is calculated as in Equation (4), which is the same as the existing rate control model in HM-16.20, where *TB* and *NF* refer to the given target bitrate for a sequence and the full frames in a sequence, respectively.

$$T_{GOP} = \begin{cases} \dfrac{(R_{bitsLeft} - R_{picAvg} * (FL - SW))}{SW} \\ \times GOP_{size}, \quad if\, \lambda_{last} > 2000 \\ \dfrac{(R_{bitsLeft} - R_{picAvg} * (FL - FR))}{FR} \\ \times GOP_{size}, \quad otherwise \end{cases} \tag{2}$$

$$\lambda_{last} = D_{VGGfeature} \times \lambda_{prevPic} + \left(1.0 - D_{VGGfeature}\right) \times \lambda_{currPic} \tag{3}$$

$$R_{picAvg} = \frac{TB}{NF} \tag{4}$$

C. BIT ALLOCATION FOR FRAME-LEVEL RATE CONTROL ALGORITHM

The rate control algorithm in HM-16.20 formulates bit allocation at the frame-level by firstly estimating bit ratio ω formed by considering the hierarchical B-prediction for the RA configuration. Next, the estimated bit ratio is capped depending on the remaining frames that will be encoded *FL* set to less than 16. When this condition is unsatisfied, the frame-level bit allocation will be estimated based on a static scale threshold set to 0.5, determined empirically. This existing thresholding scheme on the frame-level bitrate estimation model in HM-16.20 does not stand with the visual characteristic of a frame that will be coded. Depending on the feature of a frame, the threshold scale may not favor the rate control performance. Furthermore, it lacks temporal information from the previously coded frame perceptual-wise, predominantly when the model is applied to budgeting bitrate at the frame-level designed under a dyadic high-delay hierarchical prediction structure RA configuration.

$$T_{Propic} = \begin{cases} \dfrac{T_{GOPbitsleft} \times \omega}{\sum_{i=0}^{N-1} \omega_i}, if\ FL < 16 \\ D_{VGGfeature} \times \left(\dfrac{T_{GOPbitsleft} \times \omega}{\sum_{i=0}^{N-1} \omega_i}\right) + \\ \left(1.0 - D_{VGGfeature}\right) \times \left(\dfrac{T_{GOP} \times \omega}{\sum_{i=0}^{N-1} \omega_i}\right), \\ otherwise \end{cases} \tag{5}$$

The proposed frame-level bit allocation considers the perceptual loss value $D_{VGGfeature}$ to alter the thresholding scheme of the existing bit allocation at the frame-level. $D_{VGGfeature}$ is designed for the RA coding structure manner. The proposed attempts to determine each frame bit allocation by creating the existing threshold more adaptive and broader within the range values and getting better exposure on a visual-friendly bit allocation model. Notably, the proposed extension model applies when *FL* is larger than 16. Otherwise, the exact estimation model as in HM-16.20 is applicable. The proposed bit allocation of frame-level rate control T_{Propic} can be described as in Equation (5), where $T_{GOPbitsleft}$ denotes the remaining bits of the current GOP after allocating bits T_{GOP}, ω_i represents the bit ratio of the i-th frame, and N is for the total number of frames in the sequence.

D. ESTIMATION OF A AND B PARAMETERS, Λ, AND QP FOR THE FRAME-LEVEL RATE CONTROL

After calculating each frame target bitrate, the next step is to determine the λ value for each frame based on Equation (1). In HM-16.20, the bpp term is derived with the target bitrate of a picture over the number of pixels in

a frame, or mathematically, $bpp = T_{HMpic}/N_{pixel}$. Notably, Equation (1) is only valid for not Intra frame type. The λ value for an Intra frame λ_{Intra} is decided from Equation (6). The estimation model uses the total cost of an Intra frame $Cost_{intra}$ from the Hadamard transform from CTUs within the Intra frame, which is then capsulated as the mean absolute difference for each pixel within a frame MAD_{pixel} as in Equation (7). $Cost_{intra}$ can adequately yield frame characteristic information for λ_{Intra} for the proposed algorithm that also employs the same manner of λ_{Intra} as in HM-16.20.

$$\lambda_{Intra} = \left(\frac{\alpha}{256}\right) \times \left(\frac{MAD_{pixel}}{bpp}\right)^{\beta} \tag{6}$$

$$MAD_{pixel} = \left(\frac{Cost_{intra}}{N_{pixel}}\right)^{1.2517} \tag{7}$$

For other frame types, the proposed algorithm initially defines bpp_{pro} as the bpp term at the frame-level calculated based on the proposed bit allocation of the frame-level T_{Propic}, described as

$$bpp_{pro} = \frac{T_{Propic}}{N_{pixel}} \tag{8}$$

Subsequently, α_{pro} and β_{pro} are proposed to utilize the proposed $D_{VGGfeature}$ before completing the encoding process to improve the initialization of the model parameters α and β of the existing rate control model parameters. An observation on updating the existing α and β parameters for each frame is conducted based on the existing work [2], [3]. The updated α and β are denoted as α_{new} and β_{new} computed, as in Equations (9) and (10).

$$\alpha_{new} = \alpha_{estimated} + \left(\delta_{\alpha}\left(ln\lambda_{real} - \lambda_{comp}\right) \times \alpha_{estimated}\right) \tag{9}$$

$$\beta_{new} = \beta_{estimated} + \left(\delta_{\beta}\left(ln\lambda_{real} - \lambda_{comp}\right) \times lnbpp_{real}\right) \tag{10}$$

Notably, $\alpha_{estimated}$ and $\beta_{estimated}$ denote the estimated α and estimated β, respectively. δ_{α} and δ_{β} are constants, bpp_{real} represents the bits per pixel required during the encoding process, and λ_{real} and λ_{comp} represent the actual and the estimated λ, respectively.

Inspired by the updating process of α and β in Equations (9) and (10), the $D_{VGGfeature}$ value of each frame is investigated experimentally to improve $\alpha_{estimated}$–λ and $\beta_{estimated}$–λ. Parameters, namely λ_{real}, bpp_{real}, δ_{α}, and δ_{β}, in Equations (9) and (10) are disregarded in the experiments to create a new estimation model for α_{pro} and β_{pro}, as in Equations (11) and (12), before the encoding process.

$$\alpha_{pro} = \alpha_{HM} + D_{VGGfeature} \tag{11}$$

$$\beta_{pro} = \beta_{HM} + D_{VGGfeature} \tag{12}$$

The proposed Equations (11) and (12) aims to improve the estimation of the existing α and β parameters in HM-16.20, denoted as α_{HM} and β_{HM}, respectively. Thus, both parameters are anticipated to strengthen the relationship of α–λ and β–λ based on the proposed visual feature, which can be denoted as α_{pro}–λ and β_{pro}–λ. To this end, improving the relationship between α to λ and β to λ may also enhance the relationship of bpp and λ. To finally be able to do that, a new estimated model for λ based on the proposed visual feature λ_{pro} is also required to be adjusted. The proposed λ_{pro} is defined as:

$$\lambda_{pro} = \left(\alpha_{pro} \times {bpp_{pro}}^{\beta_{pro}}\right) + \frac{TL_{feature}}{bpp_{pro}}, \tag{13}$$

$$TL_{feature} = \frac{StD_{TLid-1} \times (TL_{id} + QP_{base})}{StD_{TLid-1} \times QP_{init}^{TL}}, \tag{14}$$

where $TL_{feature}$ is introduced to perpetuate λ_{pro} for preserving the current frame visual quality according to its temporal layer characteristic. StD_{TLid-1} stands for the standard deviation of the original frame in the temporal layer ID $TL_{id} - 1$, QP_{base} is the given base QP value, and QP_{init}^{TL} denotes the initial QP frame of the current TL_{id}. Notably, QP_{init}^{TL} is determined based on previous work in [22]. This modification of λ_{pro} indicates the QP decision of the existing rate control model based on the proposed $D_{VGGfeature}$, defined as

$$QP_{pro} = 4.2005 \times \log\left(\lambda_{pro}\right) + 13.7122 + D_{VGGfeature} \tag{15}$$

In general, QP decides the quantization step after the transformation determining each predicting mode distortion level and the residual after quantization. The proposed algorithm also uses two constant values in Equation (15), 4.2005 and 13.7122. It was found that both existing values provide the best coding efficiency for the proposed algorithm. However, the QP value generated by the proposed QP_{pro} is forecasted based on the proposed λ_{pro}. In addition, we added $D_{VGGfeature}$ to the formulation of QP_{pro} to bias the default value 13.7122 in HM-16.20. To validate the proposed QP_{pro} performance, the proposed formula in Equation (15) can also be represented as

$$QP_{pro} = QP + D_{VGGfeature}, \tag{16}$$

where QP denote the QP formulation with λ_{pro} without including $D_{VGGfeature}$. Then, Figure 5 is provided to illustrate the performance of the proposed QP_{pro} in terms of BD-BR-PSNR and BD-BR-MSSSIM, resulting in -2.70% and -22.99%, respectively. To validate the proposed QP_{pro}, different formulations of QP_{pro} are observed to adjust the proposed QP_{pro}, such as $QP_{pro} = QP - D_{VGGfeature}$ and $QP_{pro} = QP + 0.5 + D_{VGGfeature}$. It shows that the other formulas tend to worsen BD-BR-MSSSIM while strengthening the BD-BR-PSRN performance. Notably, the proposed algorithm is aimed to maintain or even improve the visual quality of a compressed video. Therefore, BD-BR-MSSSIM is best considered for the proposed algorithm. A positive BD-BR-MSSSIM score implies a low performance of the proposed. In addition, we tried to use the original QP estimation as in HM-16.20 for the proposed algorithm, which resulted in a lower BD-BR-MSSSIM score than the proposed QP_{pro}. However, the proposed algorithm still exhibited significant improvements in terms of objective and subjective quality. To further evaluate the effectiveness of the proposed QP_{pro}, Figure 6 depicts the curve of the QP_{pro}–λ_{pro} relationship. The R^2 coefficients in this figure display a strong

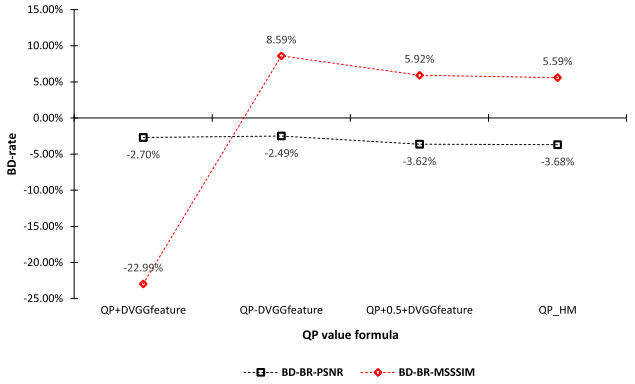


FIGURE 5. Validation of QP_{pro} on BD-BR-PSNR and BD-BR-MSSSIM.

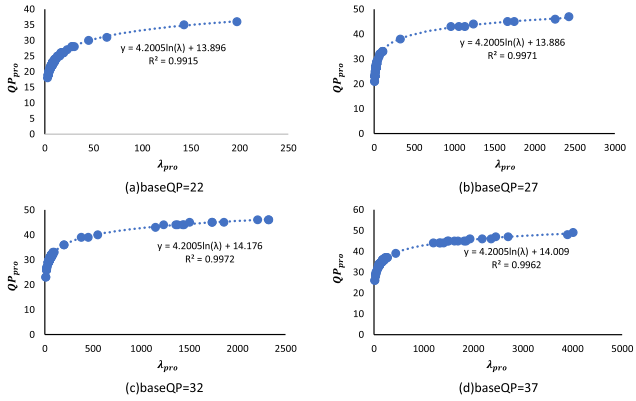


FIGURE 6. Effectiveness of QP_{pro} towards λ_{pro}.

TABLE 2. Comparisons for the relationship of α–λ and β–λ between the existing model in HM-16.20 and the proposed algorithm, denoted as α_{HM} λ_{HM} and β_{HM}–λ_{HM} for those based on the HM-16.20, and α_{pro}–λ_{pro} and β_{pro}–λ_{pro} based on the proposed algorithm, respectively.

Seq.	α_{HM}-λ_{HM}	β_{HM}-λ_{HM}	α_{pro}-λ_{pro}	β_{pro}-λ_{pro}
A	24.46%	32.20%	**67.75%**	**68.01%**
B	21.77%	34.14%	**67.37%**	**55.54%**
C	31.65%	34.04%	**81.48%**	**84.10%**
D	13.69%	38.89%	**68.41%**	**90.30%**
E	13.87%	28.33%	**78.65%**	**91.34%**
F	18.19%	32.45%	**73.39%**	**69.07%**

QP_{pro}–λ_{pro} relationship. Notably, the "BasketballPass" test sequence is used to visualize this result coded with different QP settings.

E. EFFECTIVENESS OF THE PROPOSED RATE CONTROL MODEL

After proposing the estimation models for bit allocation at the GOP- and frame-level, α_{pro}, β_{pro}, λ_{pro}, bpp_{pro}, and QP_{pro}, the effectiveness of those models should be evaluated by comparing them with the existing estimation models in HM-16.20. To reiterate, it is crucial to inspect the relationship of α to λ and β to λ to determine a better

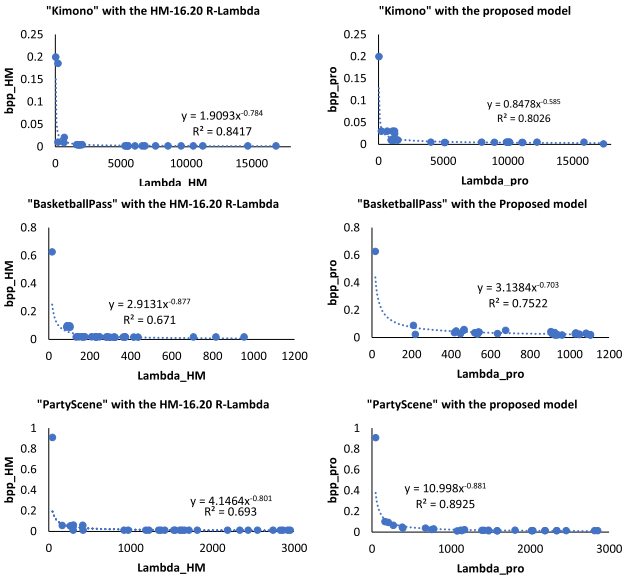


FIGURE 7. Curve fitting comparisons of bpp_{HM}–λ_{HM} and bpp_{pro}–λ_{pro}.

TABLE 3. Comparisons for the relationship of *bpp*–λ between the existing model in HM-16.20 and the proposed algorithm, denoted as bpp_{HM}–λ_{HM} for those based on the HM-16.20, and bpp_{pro}–λ_{pro} based on the proposed.

Seq.	bpp_{HM}–λ_{HM}	bpp_{pro}–λ_{pro}
A	**84.17%**	80.26%
B	**83.43%**	83.06%
C	86.44%	**96.97%**
D	69.30%	**89.25%**
E	55.07%	**82.95%**
F	67.10%	**75.22%**

relationship of bpp to λ. Table 2 tabulates the aforementioned relationship comparisons. The following test sequences are employed: (A) "Kimono", (B) "BasketballDrive", (C) "BQMall", (D) "PartyScene", (E) "BQSquare", and (F) "BasketballPass".

First, the estimated α, β, and λ of the first 17 frames of the test sequence are collected. The correlations of each estimated α and β parameter with the λ values of the tested frames are observed using the Pearson product-moment correlation coefficients. From Table 2, the proposed algorithm tends to yield a stronger relationship of α_{pro}–λ_{pro} and β_{pro}–λ_{pro} than the existing α_{HM}–λ_{HM} and β_{HM}–λ_{HM}. In addition, it is reasonable to mention that the existing α_{HM}–λ_{HM} and β_{HM}–λ_{HM} relationships are critical to the quality of the bpp–λ relationship. In particular, when POC $= 0$, α_{HM} and β_{HM} are set to their default values, observation shows no correlation to the estimated λ. Inaccurate α_{HM} and β_{HM} on the first frame led to the miscalculation of the allocating bits, estimating λ, worsening distortion, and decreasing visual quality — all these results may affect the following consecutive frames. To confirm the quality of

TABLE 4. Experimental environment.

Item	Specification
Processor	AMD Ryzen 7 3700X 8-Core @ 3.59GHz
GPU	GeForce RTX 2080 Ti 16GB
Operating system	Windows 10 Pro 64-bit
Anchor	HM-16.20 (without any modification)
Coding configuration	RA configuration
GOP size	16
Rate control	Enabled
Hierarchical bit	Adaptive ratio bit allocation
Test sequences	Class B (FHD), Class C (WVGA), and Class D (WQVGA)
DNN framework	TensorFlow-GPU with CUDA 9.0

the bpp–λ relationship based on the proposed one, Figure 7 shows some comparisons between the bpp_{HM}–λ_{HM} and bpp_{pro}–λ_{pro} relationship. The ''BQMall'' and ''BQSquare'' test sequences are coded to display this analysis. This figure portraits a stronger bpp–λ relationship generated by the proposed than that in HM-16.20. The R^2 coefficient is used for the determination with a value in the range [0, 1]. R^2 value that is closer to 1 is the better model. Table 3 delineates the percentage portion of the existing bpp_{HM}–λ_{HM} in comparison with the proposed bpp_{pro}–λ_{pro} for other test sequences. This table confirms that the proposed models tend to have a better bpp–λ relationship than those in HM-16.20.

IV. EXPERIMENTAL RESULTS

The evaluation of the proposed rate control algorithm was assessed based on conditions defined in Table 4. Specifically, the proposed algorithm was evaluated under the common test conditions of HEVC [53] to assess bitrate accuracy, coding efficiency, objective visual quality under the peak signal-to-noise ratio (PSNR), and the multi-scale structural similarity (MSSSIM) metrics, and the subjective evaluation. We compared the proposed algorithm with other rate control algorithms in HM-16.20, initially from the R–λ model [2] and Li *et al.* [54] algorithm called model parameter estimation (MPE) under the same experimental conditions.

A. OBJECTIVE PERFORMANCE EVALUATIONS

A total of 13 test sequences are encoded for the RA configuration with the GOP size set to 16, which is a typical case in practical applications, using all QP parameters: 22, 27, 32, and 37. Subsequently, all the experiments for each QP were summarized by averaging the results of every test sequence. Notably, the bitrate accuracy *BA* for this evaluation is below.

$$BA = \frac{TB - AB}{TB} \tag{17}$$

Based on Equation (17), a lower *BA* score indicates a better performance. *TB* represents the given target bitrate as listed in Table 10. *AB* is for the actual generated bitrate. *BA* is to check how accurately a rate control model can satisfy the given *TB*. Regarding the coding efficiency performance (BD-BR), the evaluation was obtained to measure the bitrate reduction by both proposed and conventional algorithms.

In addition, it is calculated while maintaining the same video quality measured by the PSNR and MSSSIM metrics, denoted as BD-BR-PSNR and BD-BR-MSSSIM. A negative value of BD-BR implies an improvement of BD-BR. For the objective visual quality, the assessments were applied by observing the difference between the generated PSNR and MSSSIM metrics of the proposed algorithm, called as $PSNR_{pro}$ and $MSSSIM_{pro}$, respectively, compared with other algorithms, $PSNR_{con}$ and $MSSSIM_{con}$, described as

$$\Delta Y_{PSNR} = PSNR_{pro} - PSNR_{con} \tag{18}$$

$$\Delta Y_{MSSSIM} = MSSSIM_{pro} - MSSSIM_{con} \tag{19}$$

ΔY_{PSNR} and ΔY_{MSSSIM} provide the difference values of the objective visual quality from the PSNR and MSSSIM metrics, respectively. A positive result from Equations (18) and (19) advises a higher visual quality than other algorithms.

As shown in Table 5, the proposed rate control algorithm conveys a better bitrate accuracy than the existing rate control models, with approximately 3.23% bitrate accuracy on average. Both the tested conventional algorithms slightly suffer in terms of bit accuracy by approximately 5%. The objective quality of the proposed algorithm also yields higher PSNR and MSSSIM scores than the tested traditional algorithms. It is observed that the proposed algorithm demonstrates 34.90 dB and 0.97644 for the PSNR and MSSSIM metrics on average, respectively. In addition, the proposed algorithm portraits consistent improvements in objective quality in almost all the test sequences. Table 6 confirms the quality difference of PSNR and MSSSIM for each class test sequence of the proposed rate control algorithm against the MPE and R–λ models. The proposed algorithm also yields better bitrate accuracy than the MPE algorithm by approximately −3.88%, BD-BR-PSNR by −7.85%, and BD-BR-MSSSIM by −4.39%. It generates better bitrate accuracy on BD-BR-PSNR and −8.74% when compared with the R–λ model. Accordingly, the proposed rate control algorithm based on deep convolutional features can perform better than the other two rate control algorithms.

B. SUBJECTIVE PERFORMANCE EVALUATIONS

For the subjective quality comparison, Figures 8 and 9 depict several regions from different test sequences generated by the proposed algorithm and the R–λ model of HM-16.20. In Figure 8, a visual quality comparison of the reconstructed ''BQMall'' test sequence from frame number 85 is displayed. The proposed algorithm quality commits to producing slightly higher visual quality subjectively than the one generated by the R–λ model in HM-16.20. In addition, the proposed algorithm can demonstrate a moderately lower bitrate of approximately 452.59 Kbps than the R–λ model at 456.74 Kbps. The proposed algorithm can testify to the high video quality of a compressed video while decreasing the generated bitrates. Accordingly, the proposed algorithm

TABLE 5. Objective comparisons of the existing rate control and proposed algorithms for the RA configuration.

Class	Sequence	R–λ			MPE			Proposed		
		BA	PSNR	MSSSIM	BA	PSNR	MSSSIM	BA	PSNR	MSSSIM
B	Kimono	3.37%	37.40	0.96536	0.57%	**38.13**	**0.97356**	**0.07%**	37.30	0.96485
	ParkScene	6.19%	35.78	0.97299	7.32%	35.79	0.97301	**3.42%**	**35.82**	**0.97384**
	Cactus	7.83%	35.30	0.97529	7.64%	35.32	0.97530	**7.22%**	**35.34**	**0.97576**
	BasketballDrive	2.88%	36.45	0.97466	**2.58%**	36.48	0.97472	6.51%	**36.68**	**0.97597**
	BQTerrace	2.29%	34.37	0.98043	2.29%	34.39	0.98046	**-6.23%**	**34.41**	**0.98088**
C	BasketballDrill	7.03%	**36.84**	0.97991	7.14%	**36.84**	0.97986	**2.97%**	36.71	**0.98007**
	BQMall	0.63%	33.85	0.97084	0.43%	33.94	0.97394	**0.01%**	**34.04**	**0.97791**
	PartyScene	**7.42%**	31.97	0.97621	7.70%	32.03	0.97629	8.81%	**32.28**	**0.97760**
	RaceHorses	3.62%	32.96	0.96501	4.01%	32.95	0.96454	**3.50%**	**33.04**	**0.96655**
D	BasketballPass	3.98%	37.99	0.98272	4.12%	**38.01**	0.98359	**1.95%**	38.00	**0.98678**
	BQSquare	3.04%	32.81	0.98717	**1.97%**	32.82	0.98747	3.59%	**33.16**	**0.98869**
	BlowingBubbles	7.39%	**33.73**	0.97804	8.02%	33.72	0.97798	**4.07%**	**33.73**	**0.97872**
	RaceHorses	4.79%	32.98	0.96425	**3.35%**	32.94	0.96358	6.13%	**33.19**	**0.96607**
	Average	4.65%	34.80	0.97484	4.40%	34.87	0.97572	**3.23%**	**34.90**	**0.97644**

TABLE 6. Average PSNR, MSSSIM, BD-BR-PSNR, and BD-BR-MSSSIM comparisons between the existing rate control and proposed algorithms for the RA configuration.

Class	R–λ vs. Proposed				MPE vs. Proposed			
	Y_{PSNR}	Y_{MSSSIM}	BD-BR-PSNR	BD-BR-MSSSIM	Y_{PSNR}	Y_{MSSSIM}	BD-BR-PSNR	BD-BR-MSSSIM
B	0.05	0.00051	-4.66%	-6.15%	-0.11	-0.00115	-4.14%	-6.03%
C	0.11	0.00254	-3.89%	-9.01%	0.08	0.00188	-3.32%	-7.90%
D	0.14	0.00202	-4.64%	-11.14%	0.15	0.00191	-4.18%	-9.62%
Average	0.10	0.00169	-4.39%	-8.74%	0.04	0.00088	-3.88%	-7.85%

(a) (b)

FIGURE 8. Visual quality comparisons of (a) the existing R–λ model at 456.74 Kbps and (b) the proposed algorithm at 452.59 Kbps on frame #85 of the "BQMall" sequence.

can achieve a higher coding gain than the existing R–λ model in HM-16.20.

Moreover, Figure 9 presents a visual quality generated by the proposed and the existing one in the HM software

FIGURE 9. Visual quality comparisons of (a) the existing R–λ model at 207.01 Kbps and (b) the proposed algorithm at 211.33 Kbps on frame #98 of the "BasketballPass" sequence.

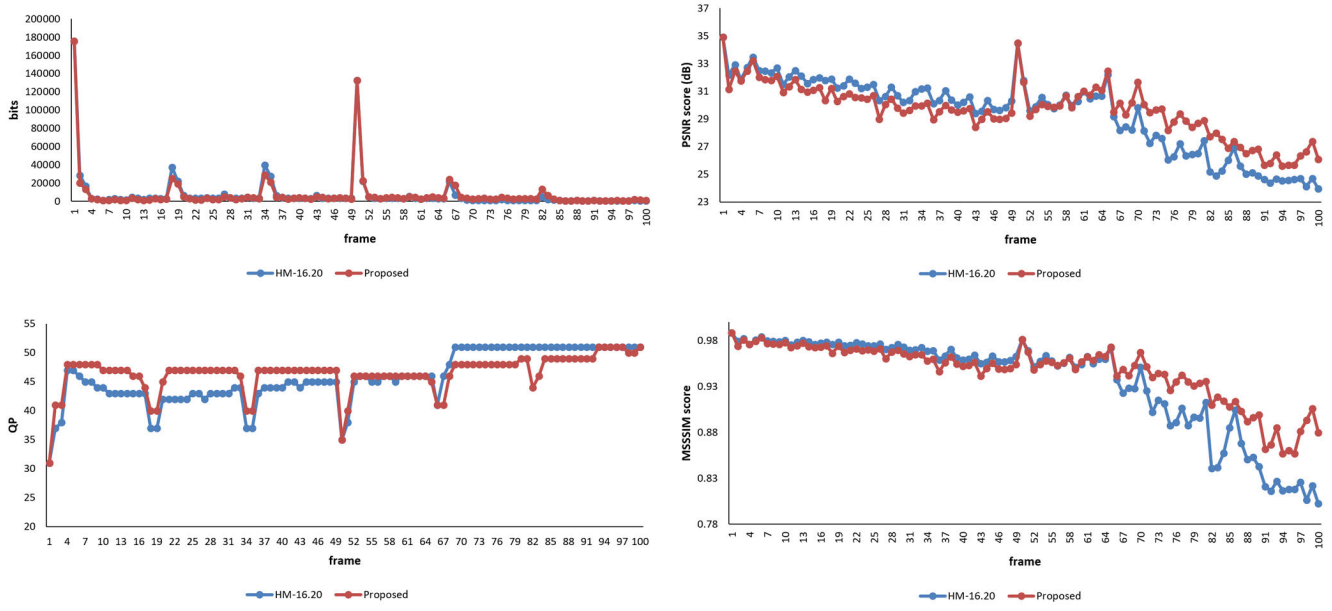


FIGURE 10. From left to the right: frame-by-frame bits, PSNR, QP, and MSSSIM comparisons between the R–λ model in HM-16.20 and proposed algorithms for the "BQMall" test sequence.

from frame number 98 of the "BasketballPass" test sequence. The proposed algorithm witnesses a slightly higher actual bitrate by approximately 211.33 Kbps than the R–λ model in HM-16.20 at 207.01 Kbps. However, this increased bitrate is also followed by more improvements in the visual quality of the proposed algorithm. Hence, the proposed algorithm can still contribute to the overall coding gain. Additional observations are then provided to confirm these improvements to

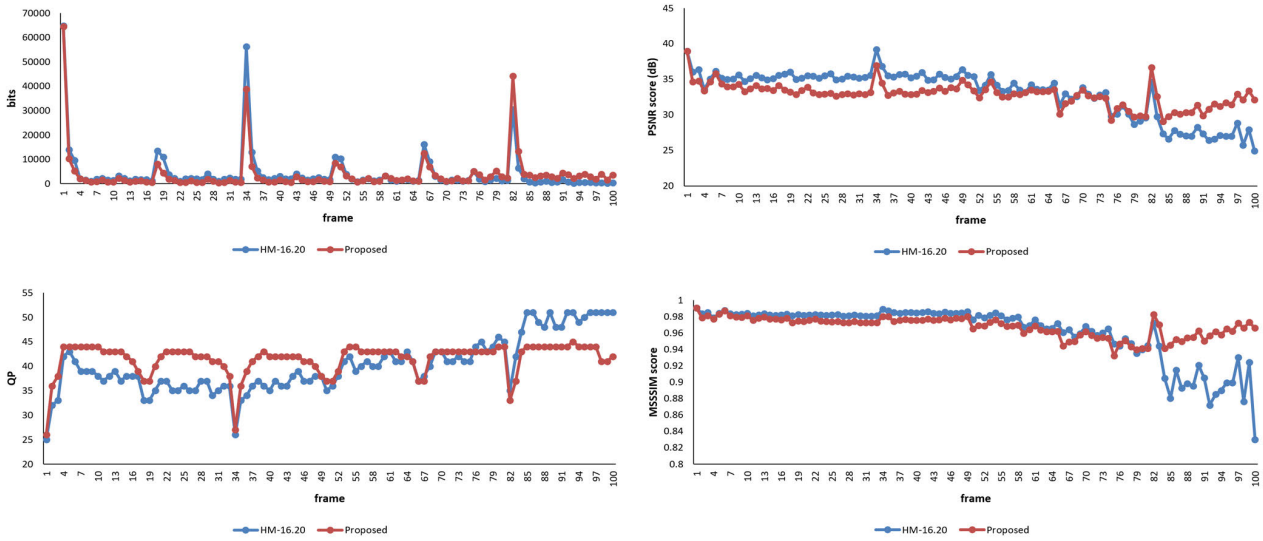


FIGURE 11. From left to the right: frame-by-frame bits, PSNR, QP, and MSSSIM comparisons between the R–λ model in HM-16.20 and proposed algorithms for the "BasketballPass" test sequence.

observe the fluctuation changes of bits, PSNR, MSSSIM, and QP of the "BQMall" and "BasketballPass" test sequences, as shown in Figures 10 and 11, respectively. In these figures, the generated bitrate by the proposed algorithm displays a significant impact on the entire performance of the proposed rate control algorithm. The proposed algorithm can allocate lower bitrates than that in the R–λ model of HM-16.20 at the beginning frames of the test sequences while maintaining relatively similar reconstructed video qualities. This case is also confirmed by the generated PSNR and MSSSIM fluctuation changes of both test sequences. In addition, the induced QP decision comparisons between the proposed algorithm and the R–λ model are depicted. These figures show that the proposed algorithm can result in even more stable QP determination from the first to the last frames of a test sequence. Notably, a stable QP decision is a crucial parameter to indicate the quality of the rate control algorithm as it can measure both bitrate and visual qualities.

C. COMPLEXITY PERFORMANCE EVALUATIONS

The proposed rate control algorithm is designed based on deep learning features to improve the existing rate control overall performance in the R–λ model of HM-16.20 for the RA configuration. Higher time consumption is required to extract the deep-learning-based visual features as a tradeoff with its performance. The complexity is mainly from the original and reconstructed frame visual feature extraction under the VGG-16 network employment. However, the proposed algorithm complexity is still relatively similar to other existing perceptual rate control models [27] and [28]. With the same classes of test sequences used by the proposed algorithm, the algorithms in [27] and [28] require 15.2× and 15.3× more encoding time than the anchor algorithm. Furthermore, the proposed algorithm can produce comparably higher coding gain and visual quality gains than the conventional works in [27] and [28]. It is

FIGURE 12. Sample of the scene change test sequences: (a) Kimono, (b) MarketPlace, (c) RitualDance, and (d) Mobisode2.

typical for a perceptually based rate control algorithm to have a higher complexity than those that do not consider any perceptual feature approaches as a tradeoff of their coding performance.

The proposed algorithm still has room for the feature extractions optimization in parallel. Hence, throughput can be enhanced with a parallel machine, such as GPU, that suppresses more encoding time. The proposed algorithm requires 15× encoding time higher than the R–λ model of

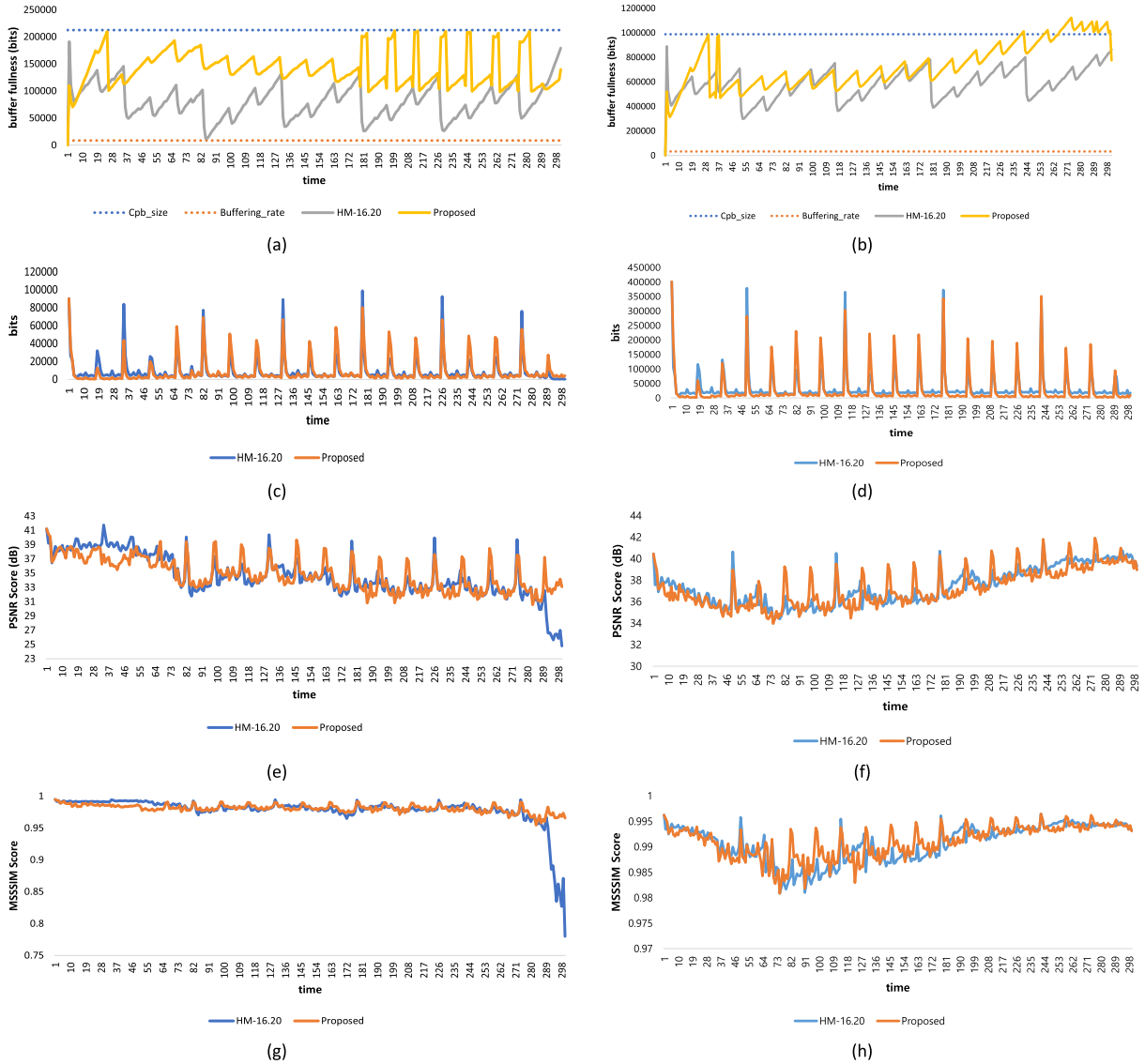


FIGURE 13. Frame-by-frame of buffer occupancy fullness, generated bits, PSNR, and MSSSIM comparisons between the R–λ model in HM-16.20 and proposed algorithms for the "BasketballPass" (left column at 427 Kbps vs. 414 Kbps) and "BQMall" (right column at 1978 Kbps vs. 1611 Kbps) test sequence.

TABLE 7. Test sequences for the scene change case.

No	Sequence	Resolution	Total frames	Total scene changes
1	Kimono	1920x1080	240	1
2	MarketPlace	1920x1080	600	2
3	RitualDance	1920x1080	600	2
4	Mobisode	832x480	300	6
5	Mobisode	416x240	300	6

TABLE 8. BD-rate comparisons between the rate control algorithm in HM-16.20 and the proposed algorithm for the scene change case.

Sequence	HM-16.20 vs. Proposed	
	BD-BR-PSNR	BD-BR-MSSSIM
RitualDance	-9.96%	-14.94%
Kimono	-5.84%	-6.88%
MarketPlace	-2.26%	-4.58%
Mobisode2	-5.29%	-7.61%
Mobisode2	-8.05%	-17.76%
Average	-6.28%	-10.35%

HM-16.20 or 8× higher than the MPE model. Notably, all the experiments were conducted with the rate control enabled.

D. EVALUATION ON SCENE CHANGE CASES

An additional evaluation of several test sequences with scene change cases was also carried out to evaluate the effectiveness of the proposed algorithm. Scene changes imply that the retention of frames in which frame scene content is significantly different from the previously retained frames. The definition of a scene change is generalized to include the abrupt

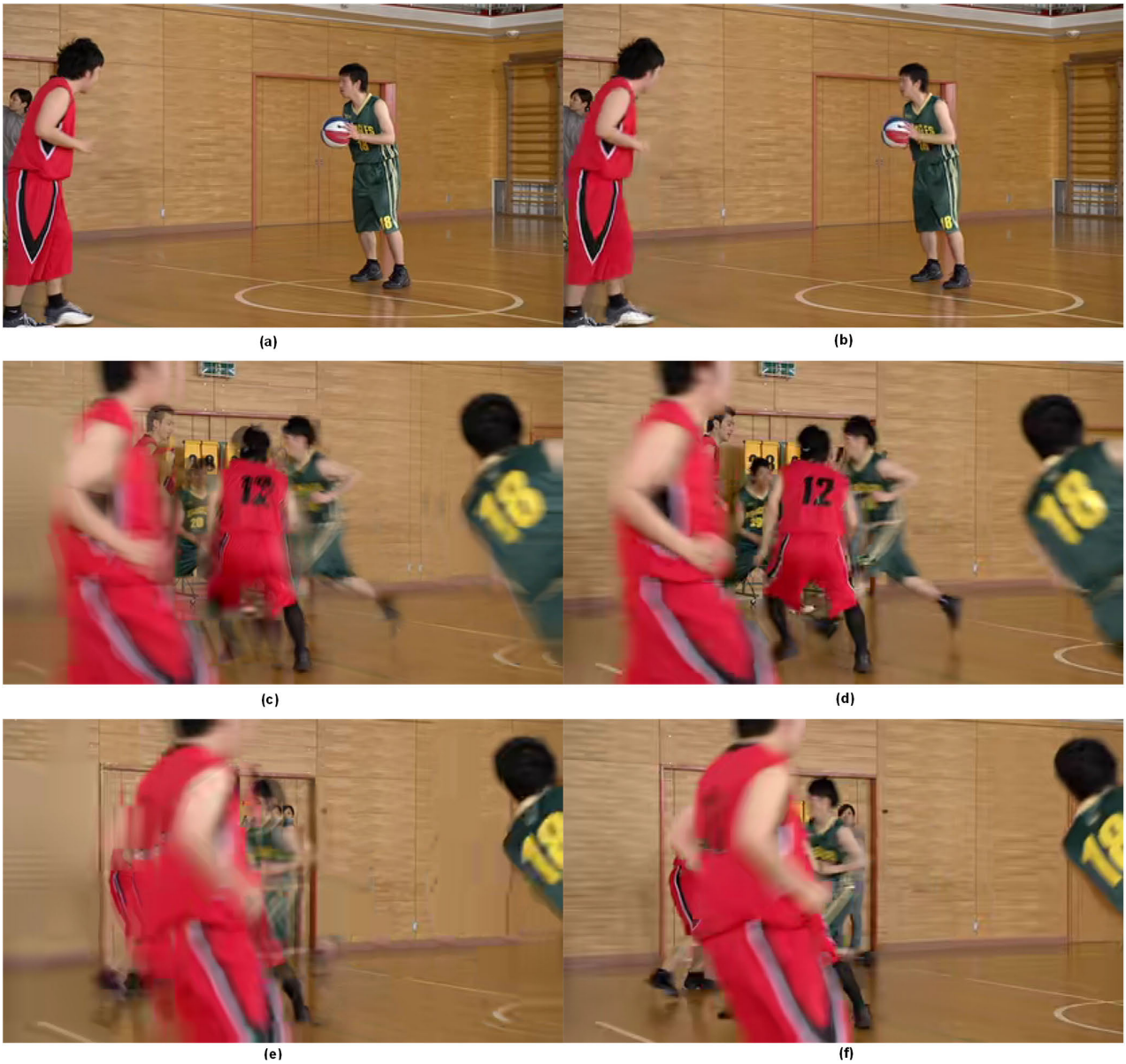


FIGURE 14. Visual quality comparisons of "BasketballPass" between HM-16.20 at 427 Kbps (left column: a, c, e) and the proposed rate control algorithm at 414 Kbps (right column: b, d, f), where a and b are frame #30, c and d are frame #291, and e and f are frame #297.

transitions between shots and gradual transitions between images resulting from the video editing modes and inter-shot changes induced by the camera operations. It is argued that measuring the significance of a change in the content of the video frames is subjective. Figure 12 outlines several test sequences with their scene-changing part. These test sequences are detailed in Table 7.

The proposed algorithm outperforms the BD-BR-PSNR and BD-BR-MSSSIM performances against the HM-16.20 by approximately -6.28% and -10.35%, respectively. This is mainly due to the ability of the proposed algorithm to maintain the generated video quality while reducing the bitrate. To this end, the proposed frame-level QP determination based on [44] should be acknowledged. Therefore, the proposed algorithm has demonstrated better performance for the entire rate control performances than the R–λ rate control model in HM-16.20, e.g., maintaining higher bitrate accuracy, maintaining visual quality, and improving the bitrate efficiency for the RA configuration.

E. EVALUATION UNDER THE HYPOTHETICAL REFERENCE DECODER (HRD) CONSTRAINT

Buffer occupancy analysis is essential for any rate control algorithm for overflow and underflow prevention. In the HEVC encoder, such cases can be handled by enabling the HRD constraint along with the R–λ rate control model as discussed in [2], [55], and [57]. The buffer size, B_{size} is defined as

$$B_{size} = D_{time} \times B_{width} \tag{20}$$

TABLE 9. Comparisons of objective quality between HM-16.20 and the proposed rate control algorithm under the HRD constraint.

Sequence	HM-16.20			Proposed			BA	ΔPSNR	ΔMSSSIM	BD-BR-PSNR	BD-BR-MSSSIM
	Kbps	PSNR	MSSSIM	Kbps	PSNR	MSSSIM					
BQTerrace (1920x1080)	64353	39.24	0.99130	61857	39.23	0.99101	-0.04	0.00	-0.00029	-6.16%	-12.30%
	9666	36.12	0.98598	8662	36.11	0.98591	-0.10	-0.01	-0.00007		
	2354	34.19	0.97810	1815	34.16	0.97805	-0.23	-0.03	-0.00004		
	846	31.58	0.95947	765	31.86	0.96173	-0.10	0.14	0.00226		
BQMall (832x480)	4590	40.10	0.99221	4327	39.76	0.99503	-0.06	-0.34	-0.00018	-5.89%	-7.56%
	1979	37.37	0.99064	1611	37.25	0.99098	-0.19	-0.13	0.00034		
	937	34.49	0.98103	761	34.05	0.97892	-0.19	-0.45	-0.00211		
	471	31.83	0.96557	393	31.17	0.95699	-0.17	-0.65	-0.00863		
BasketballPass (416x240)	1771	42.22	0.99674	1764	42.16	0.99701	0.00	-0.07	0.00026	-5.55%	-11.84%
	876	38.30	0.99161	839	37.89	0.99226	-0.04	-0.41	0.00066		
	427	34.17	0.97786	414	34.78	0.98094	-0.03	0.61	0.00308		
	218	30.21	0.95625	210	31.26	0.95691	-0.03	1.05	0.00066		

TABLE 10. Target bitrate (TB, in kbps) for each test sequence in class B, class C, and class D.

Sequences of Class B	TB (Kbps)	Sequences of Class C	TB (Kbps)	Sequences of Class D	TB (Kbps)
Kimono	5479	BasketballDrill	3911	BasketballPass	1769
	2473		1803		874
	1193		853		425
	587		435		215
ParkScene	8430	BQMall	4587	BQSquare	2741
	3264		1975		904
	1363		935		335
	588		469		130
Cactus	22718	PartyScene	9281	BlowingBubbles	2138
	6019		3690		859
	2649		1559		360
	1302		657		152
BasketballDrive	21464	RaceHorses	6057	RaceHorses	1375
	7154		2359		643
	3268		1019		300
	1659		462		145
BQTerrace	64351				
	9663				
	2352				
	844				

where D_{time} and B_{widht} represent the time delay and bandwidth, respectively. The buffer occupancy (CPB_{est}) is determined by the current state of the coded picture buffer (CPB_{state}) and the buffering rate of the current frame being coded (BR), expressed as

$$CPB_{est} = CPB_{state} + BR \tag{21}$$

Table 9 shows objective quality comparisons between the anchor HM-16.20 and the proposed algorithm. In BD-BR-PSNR and BD-BR-MSSSIM, the proposed algorithm yields BD-rate saving of approximately -6.16% and -12.30% for "BQTerrace", -5.89% and -7.56% for "BQMall", and "BasketballPass" at -5.55% and -11.84%, respectively. Figure 13 depicts comparisons of the frame-by-frame fluctuation of buffer occupancy fullness, bitrate, PNSR, and MSSSIM between HM-16.20 and the proposed algorithm from the "BQMall" and "BasketballPass" test sequences. For the "BQMall" sequence, although the buffer analysis shows overflow at the last frames, the proposed algorithm saves bitrate by -18.61% or controls the bitrate mismatch much lower than the anchor. In another case, the proposed algorithm maintains buffer occupancy for the "BasketballPass" by preventing overflow/underflow while minimizing the generated bitrates. The generated bitrate keeps lower at 414 Kbps than in HM-16.20 at 427 Kbps. Furthermore, the proposed algorithm exhibits more stable PSNR and MSSSIM fluctuations than HM-16.20, particularly in the last frames of the "BasketballPass" sequence.

Subjectively, the proposed algorithm confirms that the "BasketballPass" reconstructed frames are perceptually higher than the video quality produced by HM-16.20. The visual quality comparisons between the proposed algorithm and HM-16.20 are depicted in Figure 14 at frame numbers 30, 291, and 297. The right column (frame a, c, and e) contains reconstructed frames by HM-16.20, and the left column (frame b, d, and f) is from the proposed rate control algorithm. In figure (b), the visual quality difference made by the proposed algorithm is negligibly decreased against the one in figure (a) by HM-16.20. However, as the reconstructed frame goes to the end of the sequence, the visual quality generated

by HM-16.20 suffers even more due to inaccurate bit allocation problems at the frame-level. As depicted in (c) and (e), the quality of reconstructed frames from HM-16.20 is significantly improved in (d) and (f) by the proposed rate control.

V. CONCLUSION

This work proposes an improved rate control algorithm based on a deep convolutional feature for the RA configuration in the HEVC encoder. Note that the current version of the R–λ rate control model is not optimal for the RA configuration setting, especially in its estimation models. Therefore, the proposed algorithm aims to design a novel estimation model for the model parameter estimations (α and β), bit allocation, λ, and QP at the frame-level by considering perceptual characteristic information of a previously coded frame. In this work, the proposed algorithm employs a full-reference visual quality approach by employing a pretrained VGG-16 architecture to extract the visual feature from the original and previously coded frames. The proposed algorithm controls higher bitrate accuracy, thereby improving the coding efficiency and visual quality of the current R–λ rate model in HM-16.20. The proposed algorithm achieves better BD-rate performance on average at −4.39% and −8.74% based on BD-BR-PSNR and BD-BR-MSSSIM, respectively. The proposed algorithm is also robust in scene changes cases. Furthermore, the proposed algorithm demonstrates significant improvements over the HM-16.20 by controlling the overall performance quality and preventing buffer under/overflow when the HRD option is activated. For future work, the proposed algorithm will be adjusted to benefit rate control model for the RA configuration of Versatile Video Coding encoder.

REFERENCES

[1] H. Choi, J. Yoo, J. Nam, D. Sim, and I. V. Bajic, "Pixel-wise unified rate-quantization model for multi-level rate control," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1112–1123, Dec. 2013.

[2] B. Li, H. Li, L. Li, and J. Zhang, *Rate Control by R-Lambda Model for HEVC*, document JCTVC-K0103, Joint Collaborative Team on Video Coding, 2012.

[3] 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.

[4] I. Zupancic, M. Naccari, M. Mrak, and E. Izquierdo, "Two-pass rate control for improved quality of experience in UHDTV delivery," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 167–179, Feb. 2017.

[5] I. Marzuki, Y.-J. Ahn, and D. Sim, "Tile-level rate control for tile-parallelization HEVC encoders," *J. Real-Time Image Process.*, vol. 16, no. 6, pp. 2107–2125, Sep. 2017, doi: 10.1007/s11554-017-0720-5.

[6] 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.

[7] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1697–1706, Dec. 2012.

[8] W. Wiratama, Y.-J. Ahn, I. Marzuki, and D. Sim, "Adaptive Gaussian low-pass pre-filtering for perceptual video coding," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 5, pp. 366–377, Oct. 2018.

[9] I. Marzuki and D. Sim, "Overview of potential technologies for future video coding standard (FVC) in JEM software: Status and review," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 1, pp. 22–35, Feb. 2018.

[10] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform coefficient coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1765–1777, Dec. 2012.

[11] M. Ismail, H. Jo, and D. Sim, "Fast intra mode decision for HEVC intra coding," in *Proc. 18th IEEE Int. Symp. Consum. Electron. (ISCE)*, Jeju Island, South Korea, Jun. 2014, pp. 1–2.

[12] A. Norkin, G. Bjontegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. V. Auwera, "HEVC deblocking filter," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[13] C. M. Fu, E. Alshina, A. Alshin, Y. W. Huang, C. Y. Chen, C. Y. Tsai, C. W. Hsu, S. M. Lei, J. H. Park, and W. J. Han, "Sample adaptive offset in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[14] V. Sze and M. Budagavi, "High throughput CABAC entropy coding in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1778–1791, Dec. 2012.

[15] I. Marzuki, J. Ma, Y.-J. Ahn, and D. Sim, "A context-adaptive fast intra coding algorithm of high-efficiency video coding (HEVC)," *J. Real-Time Image Process.*, vol. 16, no. 4, pp. 883–899, Mar. 2016, doi: 10.1007/s11554-016-0571-5.

[16] X. Wang and M. Karczewicz, *Intra Frame Rate Control Based on SATD*, document JCTVC M0257, Joint Collaborative Team on Video Coding, 2013.

[17] M. Zhou, B. Li, and Y. Zhang, "Content-adaptive parameters estimation for multi-dimensional rate control," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 204–218, Jan. 2016.

[18] 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.

[19] L. Li, B. Li, H. Li, and C. W. Chen, ""λ-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.

[20] S. Li, M. Xu, Z. Wang, and X. Sun, "Optimal bit allocation for CTU level rate control in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2409–2424, Nov. 2017.

[21] M. Wang, K. N. Ngan, and H. Li, "Low-delay rate control for consistent quality using distortion-based Lagrange multiplier," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2943–2955, Jul. 2016.

[22] I. Marzuki and D. Sim, "Perceptual adaptive quantization parameter selection using deep convolutional features for HEVC encoder," *IEEE Access* vol. 8, pp. 37052–37065, 2020, doi: 10.1109/ACCESS.2020.2976142.

[23] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, "Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 130–141, Nov. 2017.

[24] T. Tariq, O. T. Tursun, M. Kim, and P. Didyk, "Why are deep representations good perceptual quality features?" in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Dec. 2018, pp. 1–16.

[25] T. Tariq, J. L. G. Bello, and M. Kim, "A HVS-inspired attention to improve loss metrics for CNN-based perception-oriented super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–9.

[26] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1–18.

[27] J. Kim, S. H. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1786–1800, Nov. 2015.

[28] M. Zhou, X. Wei, S. Kwong, W. Jia, and B. Fang, "Just noticeable distortion-based perceptual rate control in HEVC," *IEEE Trans. Image Process.*, vol. 29, pp. 7603–7614, 2020.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

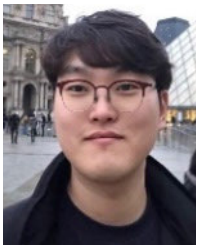
[30] I. Marzuki, J. Lee, and D. Sim, "Optimal CTU-level rate control model for HEVC based on deep convolutional features," *IEEE Access*, vol. 8, pp. 165670–165682, 2020, doi: 10.1109/ACCESS.2020.3022408.

[31] Y. Gong, S. Wan, K. Yang, H. R. Wu, and Y. Liu, "Temporal-layer-motivated lambda domain picture level rate control for random-access configuration in H.265/HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 156–170, Jan. 2019, doi: 10.1109/TCSVT.2017.2769703.

[32] J. He and F. Yang, "Efficient frame-level bit allocation algorithm for H.265/HEVC," *IET Image Process.*, vol. 11, no. 4, pp. 245–257, Apr. 2017.

[33] W. Wu, J. Liu, and L. Feng, "Novel rate control scheme for low delay video coding of HEVC," *ETRI J.*, vol. 38, no. 1, pp. 185–194, Feb. 2016.
[34] S. Sanz-Rodriguez and T. Schierl, "A rate control algorithm for HEVC with hierarchical GOP structures," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 1719–1723.
[35] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP based rate control for high efficiency video coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1101–1111, Dec. 2013.
[36] Z. Pan, X. Yi, Y. Zhang, H. Yuan, F. L. Wang, and S. Kwong, "Frame-level bit allocation optimization based on video content characteristics for HEVC," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 1, pp. 1–20, 2020.
[37] M. Zhang, W. Zhou, H. Wei, X. Zhou, and Z. Duan, "Frame level rate control algorithm based on GOP level quality dependency for low-delay hierarchical video coding," *Signal Process., Image Commun.*, vol. 88, pp. 1–11, Oct. 2020.
[38] 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.
[39] W. Lim and D. Sim, "A perceptual rate control algorithm based on luminance adaptation for HEVC encoders," *Signal, Image Video Process.*, vol. 14, no. 5, pp. 887–895, Jul. 2020, doi: 10.1007/s11760-019-01620-3.
[40] H. Guo, C. Zhu, M. Xu, and S. Li, "Inter-block dependency-based CTU level rate control for HEVC," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 113–126, Mar. 2020.
[41] M. Zhou, X. Wei, S. Wang, S. Kwong, C.-K. Fong, P. H. W. Wong, and W. Y. F. Yuen, "Global rate-distortion optimization-based rate control for HEVC HDR coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4648–4662, Dec. 2020.
[42] F. Raufmehr and M. Rezaei, "SSIM-based fuzzy video rate controller for variable bit rate applications of scalable HEVC," *J. Inf. Syst. Telecomunication*, vol. 7, no. 3, pp. 193–203, 2020.
[43] S. Bosse, M. Dietzel, S. Becker, C. R. Helmrich, M. Siekmann, H. Schwarz, D. Marpe, and T. Wiegand, "Neural network guided perceptually optimized bit-allocation for block-based image and video compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 126–130.
[44] L. Li, N. Yan, Z. Li, S. Liu, and H. Li, "λ-domain perceptual rate control for 360-degree video compression," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 1, pp. 130–145, Dec. 2020.
[45] Y. Li, B. Li, D. Liu, and Z. Chen, "A convolutional neural network-based approach to rate control in HEVC intra coding," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, St. Petersburg, FL, USA, Dec. 2017, pp. 1–4.
[46] S. Liu, B. Choi, K. Kawamura, Y. Li, L. Wang, P. Wu, and H. Yang, *JVET AHG Report: Neural Networks in Video Coding*, document JVET-L0009, Joint Video Experts Team, 2018.
[47] L. Zhou, X. Song, J. Yao, L. Wang, and F. Chen, *Convolutional Neural Network Filter for Intra Frame*, document JVET-I0022, Joint Video Experts Team, 2018.
[48] Y. Wang, Z. Chen, and Y. Li, *AHG9: Dense Residual Convolutional Neural Network Based in-Loop Filter*, document JVET-K0391, Joint Video Experts Team, 2018.
[49] I. Marzuki and D. Sim, "A deep learning-based rate control for HEVC intra coding," in *Proc. Korean Inst. Broadcast Media Eng.*, Seoul, South Korea, Nov. 2019, pp. 180–181.
[50] S. Ki, S.-H. Bae, M. Kim, and H. Ko, "Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3178–3193, Jul. 2018.
[51] D. Liu, Y. Li, J. Lin, H. Li, and F. Wu, "Deep learning-based video coding: A review and a case study," 2019, *arXiv:1904.12462*.
[52] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wang, "Image and video compression with neural networks: A review," 2019, *arXiv:1904.03567*.
[53] F. Bossen, "Common HM test conditions and software reference configurations," document JCTVC-L1100, Joint Collaborative Team on Video Coding, 2013.
[54] B. Li, M. Zhou, Y. Zhang, X. Lin, and W. Guo, "Model parameter estimation for CTU-level rate control in HEVC," *IEEE Multimedia Mag.*, vol. 25, no. 3, pp. 79–91, Jul. 2018, doi: 10.1109/MMUL.2018.112142602.
[55] Y.-J. Ahn, X. Wu, W. Lim, and D. Sim, *Target Bits Saturation to Avoid CPB Overflow and Underflow Under the Constraint of HRD*, document JCTVC-U0132, Joint Collaborative Team on Video Coding, 2015.
[56] Y.-J. Ahn, X. Wu, D. Sim, and H. Ryu, *Improvement of Coding Efficiency for Rate Control Under the Constraint of HRD*, document JCTVC-V0078, Joint Collaborative Team on Video Coding, 2015.
[57] H. Yuan, Q. Wang, Q. Liu, J. Huo, and P. Li, "Hybrid distortion-based rate-distortion optimization and rate control for H.265/HEVC," *IEEE Trans. Consum. Electron.*, vol. 67, no. 2, pp. 97–106, May 2021.

ISMAIL MARZUKI received the B.S. degree in informatics from the UIN Sultan Syarif Kasim Riau, Indonesia, in 2011, and the M.S. and Ph.D. degrees in computer engineering from Kwangwoon University, Seoul, South Korea, in 2015 and 2020, respectively. He joined the Image Processing Systems Lab (IPSL), in 2013. He was involved in video compression standardization activities, including JCTVC and JVET. In 2021, he joined the Video and Image Computing Lab, School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, as a Postdoctoral Researcher. His research interests include related to high-efficiency video coding (HEVC/x265) techniques, fast coding, rate control, versatile video coding (VVC), and perceptual video coding.



JONGSEOK LEE received the B.S. and M.S. degrees in electronic engineering from Kwangwoon University, South Korea, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in computer engineering. He is involved in video compression, including the standardization of the H.266/VVC. His research interests include video coding, video processing, computer vision, spiking neural networks, and deep learning.

WAHYU WIRATAMA received the B.S. degree in computational science and the M.S. degree in informatics from Telkom University, Bandung, Indonesia, in 2013 and 2015, respectively, and the Ph.D. degree from Kwangwoon University, Seoul, South Korea, in 2020. He is currently working as a Senior Research Engineer at CONTEC Company Ltd. His research interests include deep learning, remote sensing, computer vision, and high-efficiency video compression algorithms.

DONGGYU SIM (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree from Sogang University, South Korea, in 1993, 1995, and 1999, respectively. He was with Hyundai Electronics Company Ltd., from 1999 to 2000, being involved in MPEG-7 standardization. From 2000 to 2002, he was a Senior Research Engineer at Varo Vision Company Ltd., working on MPEG-4 wireless applications. He worked with the Image Computing Systems Lab (ICSL), University of Washington, as a Senior Research Engineer, from 2002 to 2005. His research focused on ultrasound image analysis and parametric video coding. Since 2005, he has been with the Department of Computer Engineering, Kwangwoon University, Seoul, South Korea. In 2011, he joined Simon Frasier University, as a Visiting Scholar. He was elevated to an IEEE Senior Member, in 2004. He is one of the leading inventors in many essential patents licensed to MPEG-LA for HEVC standards. His current research interests include video coding, video processing, computer vision, and video communication.