Optimal CTU-Level Rate Control Model for HEVC Based on Deep Convolutional Features

ISMAIL MARZUKI[ID]**, JONGSEOK LEE**[ID]**, AND DONGGYU SIM**[ID]**, (Senior Member, IEEE)**

Department of Computer Engineering, Kwangwoon University, Seoul 01897, South Korea

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

ABSTRACT This paper proposes an optimal rate control model based on deep neural network (DNN) features to improve the coding tree unit (CTU)-level rate control in high-efficiency video coding for conversational videos. The proposed algorithm extracts high-level features from the original and previously reconstructed CTU blocks based on a predefined DNN model of the visual geometry group (VGG-16) network. Then, the correlation of the high-level feature and quantization parameter (QP) values of previously coded CTUs is explored for subjective visual characteristics to estimate the CTU-level rate control model parameters (alpha and beta) and the bit allocation of each CTU. Therefore, this paper also proposes a new model for Lambda estimation for each CTU by improving its relationship with the estimated bits per pixel to control the rate and relative distortion. Furthermore, the Lambda and QP boundary settings were adjusted based on the proposed perceptual model to ensure the rate control accuracy of each CTU. The results of experiments with the proposed algorithm, when compared to the rate control model in HM-16.20, reveal higher bitrate accuracy and an average BD-rate gain based on PSNR, SSIM, and MSSSIM metrics using the low-delay-P configuration.

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

I. INTRODUCTION

Rate control in all video coding applications is important for optimizing visual quality by appropriately allocating bits to each rate control stage at the group-of-picture-level, picture-level, and block-level for a given bitrate condition. An excellent rate control model must be capable of assigning an accurate number of bits to each step by maintaining better visual quality in compressed videos. Conversely, a rate control model can be designed to require a lower bitrate to alleviate bandwidth bottlenecks and maintain the same subjective image quality [1]. Many rate control algorithms, for example, TM5 for MPEG–2 [2], VM8 for MPEG–4 [3], and H.264/AVC, have been studied extensively for different video coding standards [4]. Two rate control models were proposed and implemented on the HM reference software during the standardizing process for high efficiency video coding (HEVC), the latest video coding standard that adopts several advanced technologies [6]–[15]. The pixel-wise unified rate quantization (URQ) model [16], [17] was the first rate-control algorithm adopted in HM-6.0. The URQ model selects a corresponding quantization parameter (QP) that can be applied to the coding tree unit (CTU) block to set a target bitrate at the pixel-level. However, this is only effective when all coding parameters besides the QPs are fixed. Hence, rate-distortion (R–D) performance and bitrate accuracy are not optimal for CTU since the QP computation is acknowledged to suffer from the well-known 'chicken and egg' dilemma. Another rate control approach introduced for HEVC is the R–λ model [18], [19] that was developed to compute a target bitrate for a CTU in the λ–domain and demonstrated better rate control performance than several existing algorithms. However, the R–λ model for HEVC has several challenges in estimating the proper model parameters. At the CTU-level rate control, the R–λ model cannot achieve optimal performance for higher-resolution videos in low bitrate scenarios [20], especially for videos coded under a low-delay configuration [21], [22]. It is confirmed

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim[ID].

that, according to the human visual system (HVS), there is much perceptual redundancy of a video sequence that can be exploited further to improve the coding efficiency as well as the rate control performance of HEVC significantly. Note that the CTU-level rate control is essential for regulating the bitrate allocation, improving the coding performance, and comprehensively controlling the visual quality of a compressed video. In regards to the perceptual redundancy for the coding efficiency improvements, our previous study found that a perceptually adaptive QP at frame-level can significantly improve the coding efficiency of the HM-16.20 by taking advantage of the subjective characteristics of a video sequence [23].

This paper presents a CTU-level rate control algorithm for an HEVC encoder based on a deep neural network (DNN) feature. Investigations confirm that the existing R–λ model estimation does not accurately represent the relationship between the λ value and model parameters (α and β). Hence, the estimation process does not favor a proper R–λ relationship, and this leads to inaccurate rate control and failing to minimize distortion. Several studies have demonstrated that CTU-level rate control efficiency in HEVC can be achieved by diminishing the error due to the selected R–D model [24]–[26] and the error due to inaccurately estimating the model parameters [27]–[30]. In this study, the proposed algorithm establishes a new model for α and β estimations to improve the correlation between them and the estimated λ, based on the high-level features and QP values of previously coded CTUs. As in our previous study [23], high-level features are obtained from the original and reconstructed CTUs using a pre-trained visual geometry group (VGG-16) network model [23], [31]. Based on the proposed models for estimating α and β, the R–λ estimation for each CTU is then reformulated to minimize distortion over the entire frame at a given target bitrate. Consequently, the QP estimation and the boundary settings of λ and QP are also adjusted based on the extracted high-level features to maintain visual quality and ensure the bit accuracy of each CTU. Evaluated against HM-16.20 and other algorithms using the PSNR, SSIM, and MSSSIM indices, the proposed algorithm demonstrates significant coding gains with notable visual quality enhancement.

The remainder of this paper is organized as follows. In Section 2, the CTU-level rate control in HM and related works are briefly discussed. Section 3 details the proposed CTU-level rate control for HM. The proposed algorithm is evaluated in Section 4, and the paper is concluded in Section 5.

## II. CURRENT STATE OF R–λ MODEL FOR CTU-LEVEL RATE CONTROL ALGORITHM IN HEVC ENCODER

The R–λ rate control model is designed by taking λ into account as a crucial parameter during the rate control loop process that is estimated as follows:

$$\lambda = \alpha \times bpp^{\beta} \qquad (1)$$

where bpp is the bit per pixel term for the estimated bitrate for a CTU. α and β are the tuning parameters related to video contents that will be updated after the encoding process for the co-located CTU of the subsequent frames. In HEVC, α has initial values set to $+3.2003$, and β is initialized to -1.367. Different initial values of α and β are claimed to have little impact on the compressed videos, R–D performance, and bitrate accuracy [19]. However, during the updating process, both model parameters are unable to fit the relationship between distortion and bitrate for each CTU [24]–[30] accurately. This is because, in general, the λ_{real} value is not equal to $\alpha \times (bpp_{real})^{\beta}$. Therefore, the interrelationship between α and β towards λ is also critical to the R–λ rate control model performance. Note that bpp_{real} and λ_{real} respectively denote the actual consumed bpp and the actual λ used for calculating the QP after encoding each CTU. Once the bpp value is estimated for each CTU, the λ estimation process can be calculated based on equation (1). Then, the boundary adjustment of the estimated λ, the QP estimation process, and the QP boundary adjustment are affected. Please refer to the discussions in [18], [19], [32], [33].

Many studies have attempted to improve the CTU-level rate control for HEVC. Li et al. [33] proposed a weight-based R–λ rate control by applying the visual variation of pixels within a CTU to enhance the perceptual quality of conversational videos. However, the visual attention model proposed for the R–λ rate control is designed to affect only the region of interest, which may cause imbalanced bit allocations in other regions, resulting in the deterioration of the R–λ model accuracy. Li et al. [28] also proposed an optimal bit allocation (OBA) algorithm using a recursive Taylor expansion by arguing that bit allocation for the low-delay-P (LDP) case of the existing CTU-level R–λ model is not optimal. However, the proposed algorithm does not consider any initial QP scheme that leads to insignificant performance when sequences are coded with higher initial QP values. Li et al. [29] proposed a model parameter estimation algorithm (MPE) for α and β for CTU-level rate control by analyzing the CTU scan order that correlates to the complexity of the CTU contents. However, the proposed algorithm does not examine the bit allocation, λ, and QP estimation models, which limits the MPE algorithm performance. Wang et al. [30] presented a novel rate control algorithm based on the improved λ parameter that inhibits bit fluctuation and improves video quality. However, the proposed algorithm did not identify the characteristics of the test sequences adequately. In short, the algorithm inaccurately estimated the model parameters. Zhou et al. [26] designed an SSIM-based R–D model to improve the CTU-level rate control algorithm. This algorithm formulates the CTU-level bit allocation as a global optimization problem and disregards several model estimation issues present in the CTU-level rate control. This limits the overall performance of the proposed algorithm.

In conclusion, the estimation models in the CTU-level rate control are crucial for obtaining high bitrate accuracy in all
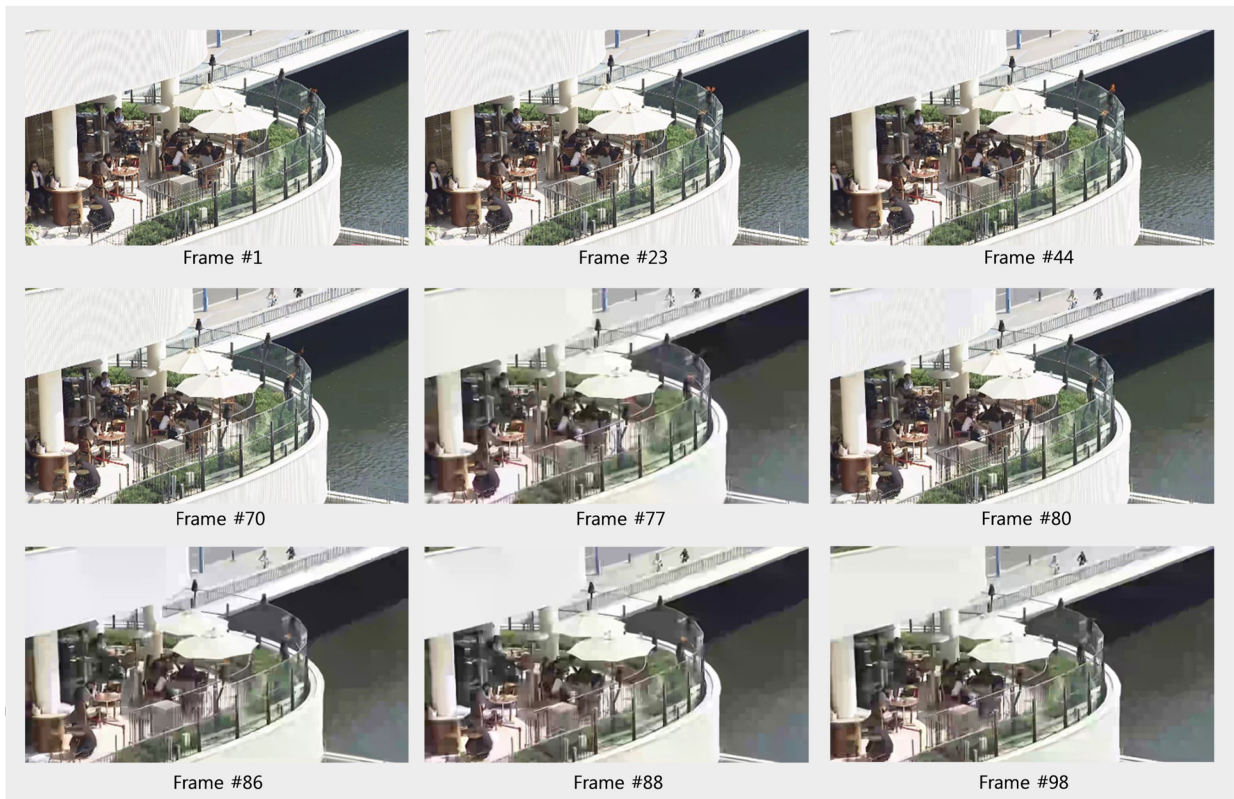
FIGURE 1. Frame-by-frame depiction of the visual quality of the 'BQTerrace' with rate control activated.

rate control models. This paper presents the development of a new estimation model for the α and β parameters as well as estimation models for the bit allocation, parameter λ, QP decision, and boundary adjustment of both λ and QP for the CTU-level rate control in the HEVC encoder. Furthermore, an adaptive QP decision at the frame-level [23] is also used for the proposed CTU-level rate control algorithm. The proposed algorithm is designed to impove the rate control in HM-16.20 reference software as its performance is significantly degraded with rate control enabled. The proposed algorithm achieves around 20% coding gain and enhances the subjective quality over the HM-16.20 with rate control on. The proposed algorithm is based on high-level features extracted from a predefined DNN model [31]. Note that the use of the DNN model for video coding is becoming more appealing to the video coding community [34]–[45]. However, investigations into using the DNN model specifically for CTU-level rate control for conversational videos are still rare. Therefore, a CTU-level rate control algorithm based on a predefined VGG network for HEVC is presented in this paper. The code of the proposed algorithm is available online: *https://bit.ly/2QRxOjA.*

III. PROPOSED CTU-LEVEL RATE CONTROL ALGORITHM FOR THE HEVC ENCODER

The R–λ rate control model in HEVC can be written as Eq. (1). There are three main factors involved: the estimated

bpp and two model parameters, α and β. The bpp term in the CTU-level rate control of HM-16.20 is calculated by including the α and β parameters that may also influence the accuracy of the bpp estimation. Regardless of the block characteristics, for all CTU blocks in the first frame, α and β are initialized with fixed values, respectively set to $+3.2003$ and -1.367. Although the original study [18], [19] argues that the video contents determine α and β, the model parameters in the existing CTU-level rate control are still unable to represent the bpp–λ relationship accurately. Based on our investigation, improper α and β parameters are responsible for the deficiency in achieving minimal distortion and rate control accuracy. Figure 1 shows an example of the visual quality of the 'BQTerrace' sequence encoded by the existing CTU-level rate control model in HM-16.20 under the LDP configuration. The subjective quality of the sequence gradually decreases from the first to the last frames. This phenomenon is caused mainly by the failure of rate control when estimating the model parameters. Therefore, a strong interrelationship between α–λ and β–λ is also crucial for the CTU-level rate control to improve the bpp–λ relationship, which also enhances the visual quality of a sequence.

In this paper, the proposed algorithm explores the use of visual feature extractions based on a particular convolutional layer of a DNN model for CTU-level rate control purposes. This paper proposes the use of a pre-trained DNN model to tackle model parameter estimation for rate control at the
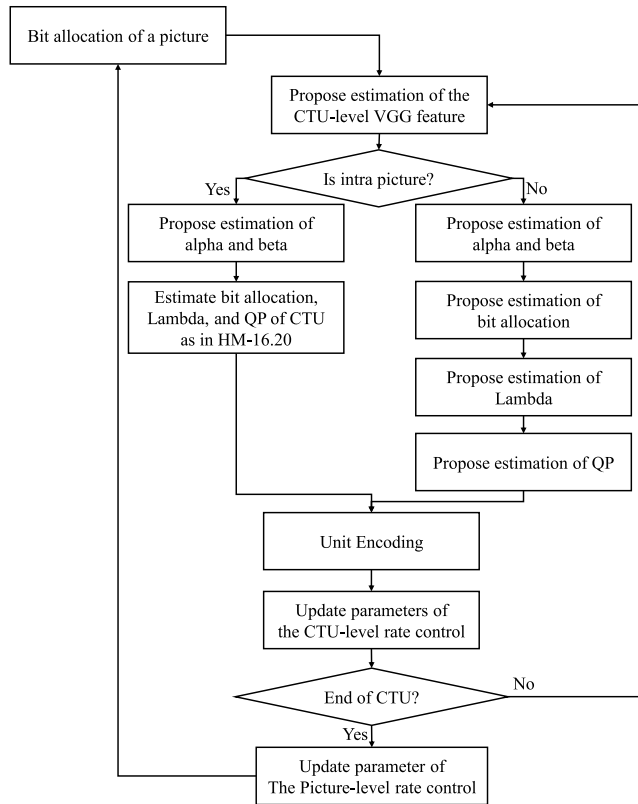
FIGURE 2. Overall flowchart of the proposed CTU-level rate control.

α and β parameters, λ, and QP, as well as the boundary settings of λ and QP for consecutive frames. The designed visual features result in a perceptual loss value based on the Euclidean distance measure.

A detailed description in several sub-categories is given as follows.

A. ESTIMATION OF THE PROPOSED VGG FEATURE

The use of our previously published work here [23], as illustrated in Figure 3, is to investigate further our findings on the visual features extracted from the original and reconstructed CTUs using a pre-trained VGG-16 network. The pre-trained VGG-16 trained on the ImageNet dataset [31] is directly employed for examining the HM-16.20 CTU-level rate control model without conducting a separate training phase. The reason for the use of the pre-trained VGG-16 network has also been discussed in our previously published paper in [23]. Briefly, the VGG-16 has an extremely deep convolutional layer, which results in convolution filters intimately to search universal patterns and generalize them. Therefore, many studies use the predefined VGG-16 model as a feature extraction technique. Note that the visual features in our previous study [23] are based on the averaged VGG feature of all CTUs within a picture used for determining the frame-level adaptive QP, which is different from the feature used for the proposed CTU-level rate control.

In this proposed CTU-level rate control, instead of using the averaged VGG features, individual CTU-level VGG features are applied to favor the model estimations for the proposed CTU-level rate control algorithm. The correlation between the visual feature and the QP value of each CTU isused, as shown in Figure 4. The 'BlowingBubbles' test sequence is coded with QP 22, 27, 32, and 37. The normalized StD values of the original CTU blocks in Figure 4(a) coded under the 'All Intra' configuration were observed to have a strong correlation with the selected QPs. It is emphasized that the StD value from the original CTU block is regarded as the visual characteristic for the proposed CTU-level rate control algorithm of the first frame only. The remaining CTUs in the frames following will apply the proposed VGG feature. Figure 4(b) shows a high correlation between the VGG feature and QP selection per CTU under the IPPP structure.

B. ESTIMATION OF BIT ALLOCATION AND MODEL PARAMETERS FOR THE PROPOSED CTU-LEVEL RATE CONTROL ALGORITHM

The R–λ relationship in Eq. (1) of the CTU-level rate control model in HM-16.20 is shown to be inadequate for optimally allocating bits for each CTU and inaccurately computes the model estimations. The estimated α, β, and bpp are the key parameters to overcome the disadvantages mentioned above. First, the estimated α, β, and λ of each CTU from the first 17 frames of the 'BlowingBubbles' test sequence are collected. Then, the correlations of each estimated α and β parameter with the λ values of all CTUs within a frame are observed using the Pearson product-moment correlation

CTU-level that is intended to improve the rate control performance for conversational video services. Both spatial and temporal features are considered for estimating the α and β parameters that also influence the design of other estimation processes, including the estimations of bit allocation, λ, and QP. In addition, the boundary settings of λ and QP are also fine-tuned to be more perceptual-friendly. Since rate control algorithms also rely on the initial QP of a frame, the proposed CTU-level rate control takes into account our previous study of perceptual adaptive QP decision to handle the frame-level initial QP decision. Please refer to the detailed algorithm, as discussed in [23].

Figure 2 illustrates the flow of the proposed algorithm. The proposed CTU-level rate control algorithm is applied after the frame-level QP initialization in [23] is employed. The VGG feature for the proposed algorithm is first estimated during the CTU-level rate control loop. Note that the estimated VGG feature in the proposed algorithm is different from our previously published work. For the first frame in a sequence, the proposed algorithm is designed straightforwardly. The standard deviation value (StD) of the original CTU block is examined to estimate the model parameters of each CTU within an Intraframe. Then, the estimations of bit allocation, λ, and QP are kept the same as in the HM-16.20 default processes. After encoding the first Intraframe, a pre-trained VGG-16 model is employed to extract visual features from the original and the reconstructed CTU to estimate the
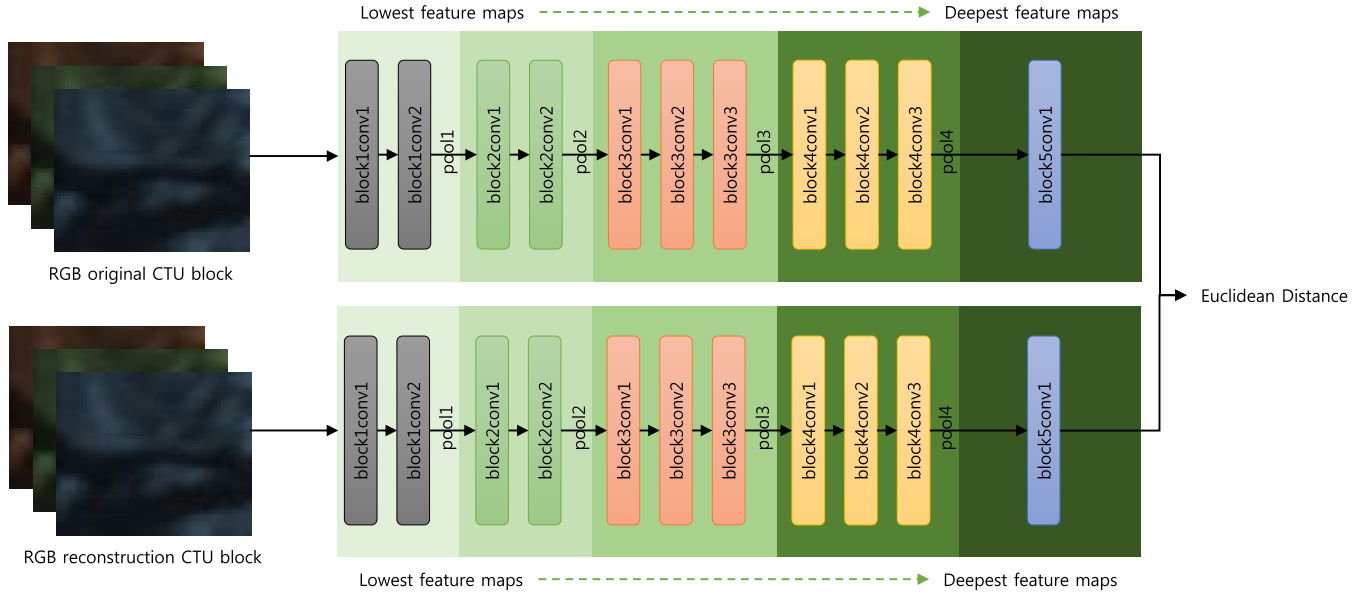
FIGURE 3. Proposed a double-simplified VGG-16 network architecture.
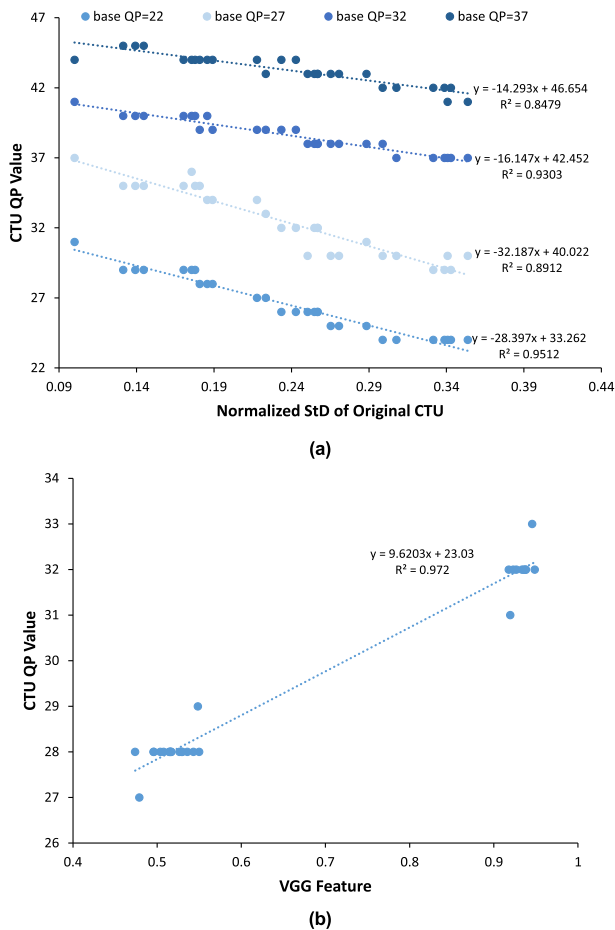

(a)

(b)

FIGURE 4. Correlation between QP values and: (a) Normalized StD and (b) VGG feature.

coefficient, and the correlation results are tabulated per frame, as shown in Table 1. From the table it can be seen that the $\alpha_{estimated} - \lambda$ and $\beta_{estimated} - \lambda$ in the CTU-level rate

control of HM-16.20 are very weak, which is critical for the bpp$-\lambda$ relationship. In particular, when POC $= 0$, $\alpha_{estimated}$ and $\beta_{estimated}$ are set to their default values, which shows no correlation to the estimated λ in Table 1. Inaccurate $\alpha_{estimated}$ and $\beta_{estimated}$ on the first frame lead to the miscalculation of allocating bits, estimating λ, degrading distortion, and decreasing visual quality; all the results may also affect the following frames.

To improve the correlations of $\alpha_{estimated} - \lambda$ and $\beta_{estimated} - \lambda$ in the CTU-level rate control, the estimation process of the α and β parameters is analyzed by taking advantage of the VGG feature and QP value relationship in Figure 4 before completing the encoding process. The updating parameter process of α and β for a CTU based on the existing work in [18], [19] is observed, symbolized as α_{new} and β_{new} and respectively can be computed as per Eq. (2) and Eq. (3)

$$\alpha_{new} = \alpha_{estimated} + \left(\delta_\alpha \left(ln \, \lambda_{real} - \lambda_{comp}\right) \times \alpha_{estimated}\right) \quad (2)$$

$$\beta_{new} = \beta_{estimated} + \left(\delta_\beta \left(ln \, \lambda_{real} - \lambda_{comp}\right) \times ln \, bpp_{real}\right) \quad (3)$$

According to [18], [19], Eq. (3) can also be formulated as

$$\beta_{new} = \beta_{estimated} - \left(\delta_\beta \left(-2 \left(\ln \lambda_{real} - \lambda_{comp}\right) \times \ln bpp_{real}\right)\right) \quad (4)$$

Note that $\alpha_{estimated}$ denotes the estimated α, and $\beta_{estimated}$ corresponds to the estimated β. 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 Eq. (2) and (4), the VGG feature and QP value of each CTU are investigated experimentally to improve $\alpha_{estimated} - \lambda$ and $\beta_{estimated} - \lambda$. The parameters λ_{real}, bpp_{real}, δ_α, and δ_β in Eq. (2) and (4) are disregarded in the experiments to create a new estimation model for α'_{new} and β'_{new} as

TABLE 1. Comparisons of $\alpha_{estimated} - \lambda$ and $\beta_{estimated} - \lambda$ and the proposed $\alpha'_{new} - \lambda$ and $\beta'_{new} - \lambda$.

POC	Picture Type	$\alpha_{estimated}-\lambda$	$\beta_{estimated}-\lambda$	$\alpha'_{new}-\lambda$	$\beta'_{new}-\lambda$
0	I	0.00%	0.00%	**91.28%**	**91.28%**
1	P	0.00%	0.00%	**97.89%**	**97.89%**
2	P	6.10%	11.68%	**74.71%**	**84.99%**
3	P	53.51%	49.85%	**84.59%**	**90.82%**
4	I	15.14%	17.75%	**75.45%**	**84.70%**
5	P	0.28%	23.18%	**63.08%**	**71.65%**
6	P	27.49%	64.72%	**76.65%**	**78.82%**
7	P	36.35%	60.77%	**79.35%**	**80.44%**
8	I	53.88%	57.45%	**77.32%**	**86.80%**
9	P	35.54%	77.16%	**67.59%**	**96.26%**
10	P	48.17%	73.14%	**80.40%**	**91.45%**
11	P	0.25%	64.12%	**73.90%**	**92.44%**
12	I	42.36%	64.16%	**64.00%**	**80.68%**
13	P	39.10%	67.80%	**61.79%**	**69.21%**
14	P	37.74%	63.54%	**67.44%**	**85.31%**
15	P	26.25%	71.77%	**68.24%**	**92.11%**
16	I	15.34%	54.34%	**72.66%**	**63.72%**
17	P	34.91%	64.93%	**79.98%**	**94.32%**

in (5) and (6), before the encoding process, expressed as

$$\alpha'_{new} = \alpha_{estimated} + \left(VGG_{prevCTU} + QP_{prevCTU}\right) \quad (5)$$

$$\beta'_{new} = \beta_{estimated} - \left(VGG_{prevCTU} + QP_{prevCTU}\right) \quad (6)$$

where $VGG_{prevCTU}$ and $QP_{prevCTU}$ represent the VGG feature and QP value of the previously coded CTU. Note that CTUs of POC $= 0$ set the StD value of the original CTU block as the VGG features, and the QP value is set to the frame-level initial QP value. As a result, the proposed model of the estimated α'_{new} and β'_{new}, can demonstrate stronger correlations for the relationship $\alpha'_{new}-\lambda$ and $\beta'_{new}-\lambda$ compared with the existing estimated α and β models in HM-16.20, as shown in Table 1. Consequently, the bit allocation formula is also proposed for the CTU-level rate control based on the estimated α'_{new} and β'_{new}, which can be defined as

$$
\begin{cases}
bpp_{CTU} = \dfrac{T_{CTU}}{N_{CTU}} + \left(VGG_{prevCTU} + QP_{prevCTU}\right) \\
T_{CTU} = T_{pic} \times \eta_{CTU} \Big/ \sum_{\{AllNotCodedCTUs\}} \eta_{CTU} \\
\eta_{CTU} = N_{CTU} \times \left(\dfrac{\lambda_{pic}}{\alpha'_{new}}\right)^{\beta'_{new}}
\end{cases} \quad (7)
$$

where bpp_{CTU} is the bit per pixel allocation of the current CTU, T_{CTU} stands for the current CTU target bits, and N_{CTU} denotes the total number of pixels in a CTU. Other parameters, such as T_{pic} and λ_{pic}, are the target bits and the estimated λ from the frame-level rate control. In (7), η_{CTU} is calculated to represent the weight of the current CTU to satisfy the T_{CTU} constraint.

C. ESTIMATION OF λ AND QP FOR THE PROPOSED CTU-LEVEL RATE CONTROL

Since Eq. (5) to (7) are intended to improve the estimation α, β, and bit allocation of each CTU, and the distortion minimization can also be achieved for the R–D performance.

However, to improve bpp–λ in (1), a new estimated model λ'_{new} is also proposed as

$$\lambda'_{new} = \left(\alpha'_{new} \times \left(bpp_{CTU}\right)^{\beta'_{new}}\right) + \left(\left(VGG_{prevCTU} \times 2.07052\right) + 16.771\right) \quad (8)$$

Note that the term $\left(VGG_{prevCTU} \times 2.07052\right) + 16.771$ is introduced to minimize distortion and be more visual-friendly over the rate, and vice versa.

As a result of the above analysis, the correlation coefficients of bpp_{CTU} and λ'_{new} of each CTU block from several frames of the 'BlowingBubbles' sequence are depicted to observe how well these statistical models fit the experimental analysis. As shown in Figure 5, the $bpp_{CTU}-\lambda'_{new}$ relationship strongly matches the model in (7) and (8) with the given adaptive α'_{new} and β'_{new} parameters. Note that α'_{new} and β'_{new}, the parameters in Figure 5, are found in the average values of all CTUs in a frame. Table 2 shows comparisons between the proposed $bpp_{CTU} - \lambda'_{new}$ relationship and the existing bpp–λ in HM-16.20 applied to the first 17 frames of the 'BlowingBubbles' sequence in percentage. The relationship models of the proposed and HM-16.20 are compared using R^2, illustrated in Figure 5, as the coefficient of determination, which has a value in the range [0, 1]. R^2 value that is closer to 1 is the better model. The proposed models tend to have better correlations of bpp–λ than the existing HM models.

Finally, based on the proposed visual feature, given the newly estimated λ'_{new}, the estimation QP denoted as QP'_{new} can also be further computed as

$$QP'_{new} = 4.2005 \times log\left(\lambda'_{new}\right) + 13.7122 + VGG_{prevCTU} \quad (9)$$

In addition, for every CTU, the boundary settings of λ and QP respectively should also be adjusted to impose an appropriate bit allocation on each CTU based on the visual feature characteristic. Specifically, the λ'_{new} boundary setting can be defined
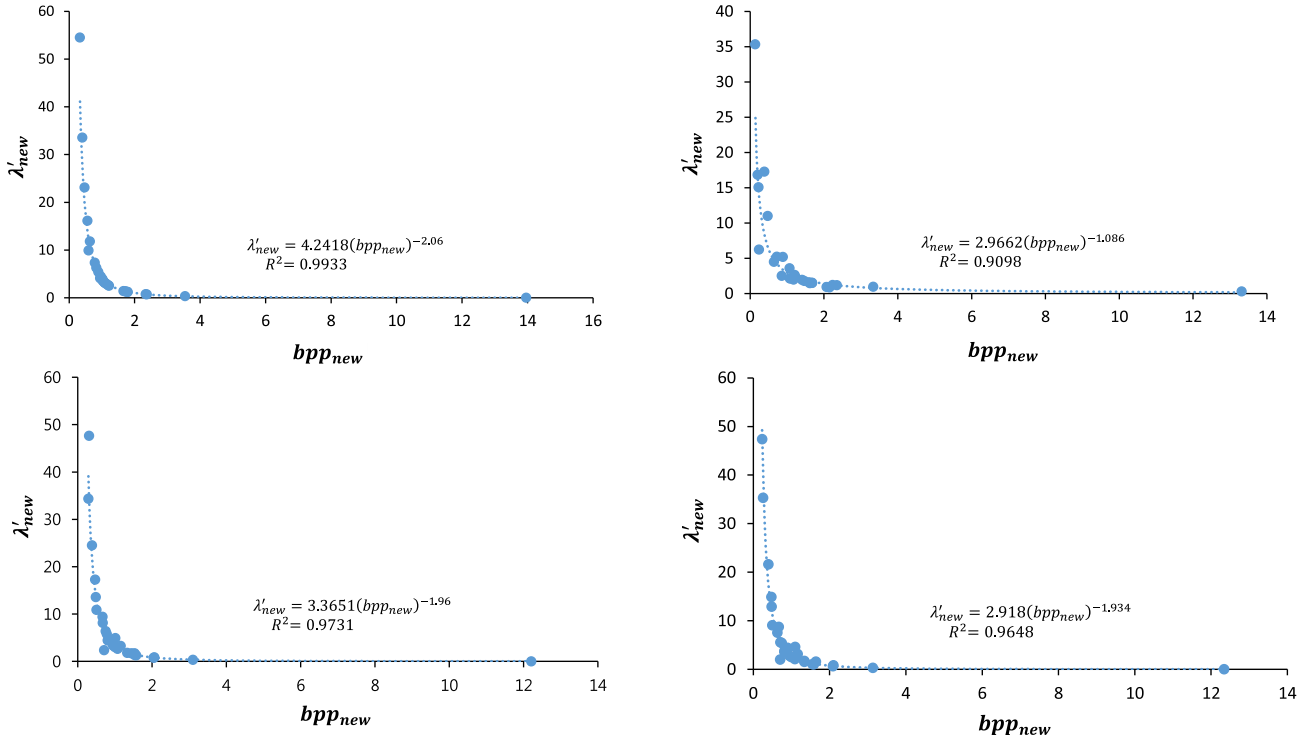
FIGURE 5. The curve fitting of the proposed $bpp_{CTU} - \lambda'_{new}$ relationship from different frames of 'BlowingBubbles'.

TABLE 2. Comparisons of bpp−λ and the proposed $bpp_{CTU} - \lambda$.

POC	Picture Type	$bpp - \lambda$	$bpp_{CTU} - \lambda'_{new}$
0	I	52.38%	**96.82%**
1	P	43.17%	**99.33%**
2	P	13.06%	**89.89%**
3	P	0.03%	**81.37%**
4	I	24.42%	**90.98%**
5	P	0.47%	**72.99%**
6	P	0.64%	**87.22%**
7	P	1.44%	**97.31%**
8	I	11.51%	**68.72%**
9	P	0.22%	**88.19%**
10	P	1.68%	**84.13%**
11	P	0.14%	**65.15%**
12	I	0.07%	**84.24%**
13	P	16.83%	**96.48%**
14	P	4.10%	**75.82%**
15	P	4.28%	**81.48%**
16	I	1.86%	**79.49%**
17	P	1.95%	**63.26%**

in Eq. (10), as shown at the bottom of the next page, and correspondingly, the QP'_{new} boundary smoothing is modified as Eq. (11), as shown at the bottom of the next page. Note that $\lambda'_{prevCTU}$ and $QP'_{prevCTU}$ stand for the newly estimated λ and QP of the previously coded CTU block. From (10) and (11), the proposed CTU-level rate control promises more variations for λ'_{new} and QP'_{new} to facilitate better visual quality for the coded CTU based on the proposed visual feature. Therefore, CTUs that have a higher $VGG_{prevCTU}$ may have more considerable latitude for visual quality improvements.

IV. EXPERIMENTAL RESULTS

The proposed rate control algorithm was evaluated under experimental conditions in Table 3. Specifically, the proposed algorithm was assessed based on objective and subjective performance evaluations under the common test conditions of HEVC [46]. For objective performance, several assessments were carried out, such as bitrate accuracy, bitrate error, coding efficiency, and objective visual quality based on three different metrics: PSNR, SSIM, and MSSSIM. The subjective evaluation was performed by conducting the mean opinion scores (MOS) test and calculating the difference MOS (DMOS) scores. The assessments were completed by comparing the proposed algorithm with the HM-16.20, URQ model [17], OBA model [28], and MPE model [29].

A. OBJECTIVE PERFORMANCE EVALUATIONS

The same experimental environment was set for both the anchor CTU-level rate control algorithm in HM-16.20 and the proposed algorithm to obtain fair comparisons, as listed in Table 3. All the objective performance measures are listed in Table 4. A total of 16 test sequences are encoded under the LDP configuration with the IPPP structure, which is a typical case in practical applications [19], using all QP parameters: 22, 27, 32, and 37. All experiments of each QP were then summarized by averaging the results of every test sequence. The bitrate error ΔBE is defined as

$$\Delta BE = BA_{proposed} - BA_{HM} \qquad (12)$$

where $BA_{proposed}$ and BA_{HM} denote the bitrate accuracy produced by the proposed algorithm and the CTU-level rate control model of the anchor software, respectively. The smaller the ΔBE value, the better the improvements. Note that $BA_{proposed}$ and BA_{HM} are calculated based on the same target bitrate expressed as

$$BA = \frac{TB - AB}{TB} \qquad (13)$$

where BA represents the result of the bitrate accuracy, TB denotes the given target bitrate, and AB stands for the actual bitrate generated by the tested rate control algorithm. Thus, AB may differ according to the rate control model that is being tested. The main objective of the BA evaluation is to check how accurately the tested models can meet the given target bitrate TB. In terms of the objective visual quality, the assessments were applied by observing the difference between the generated PSNR, SSIM, and MSSSIM metrics of the proposed algorithm (symbolized as $PSNR_{proposed}$, $SSIM_{proposed}$, and $MSSSIM_{proposed}$) compared with the HM-16.20 software (symbolized as $PSNR_{HM}$, $SSIM_{HM}$, and $MSSSIM_{HM}$), defined as

$$\Delta Y_{PSNR} = PSNR_{proposed} - PSNR_{HM} \qquad (14)$$

$$\Delta Y_{SSIM} = SSIM_{proposed} - SSIM_{HM} \qquad (15)$$

$$\Delta Y_{MSSSIM} = MSSSIM_{proposed} - MSSSIM_{HM} \qquad (16)$$

ΔY_{PSNR}, ΔY_{SSIM}, and ΔY_{MSSSIM} denote the difference values of the objective visual quality under the PSNR, SSIM, and MSSSIM metrics, respectively. A positive value for (14) to (16) indicates that the objective visual quality of the proposed algorithm is better than that of HM-16.20. Finally, the coding efficiency performance (BD-BR) of the proposed algorithm was also measured against the anchor algorithm with the PSNR, SSIM, and MSSSIM metrics, denoted as BD-BR-PSNR, BD-BR-SSIM, and BD-BR-MSSSIM, respectively. A negative value of BD-BR indicates gains over the anchor CTU-level rate control algorithm.

From Table 4, the proposed algorithm can surpass the objective performance evaluations of the existing HM-16.20 algorithm. Primarily, test sequences that have larger background areas, many homogeneous regions, and slow motions, such as 'BQTerrace,' 'BQSquare,' 'FourPeople,' 'Johnny,' 'KristenAndSara,' etc., the proposed algorithm can facilitate significant coding gains in all objective quality aspects. For instance, in terms of the BD-BR PSNR, BD-BR-SSIM, and BD-BR-MSSSIM, the proposed CTU-level rate control algorithm distributes gains up to -57.08%, -76.29%, and

TABLE 3. Experimental environment.

Item	Specification
Processor	Intel(R) Core(TM) i7-4770K CPU @ 3.50GHz
GPU	GeForce GTX 1080 with 4GB RAM
Operating system	Windows 10 64-bit
HEVC version (anchor)	HM-16.20 (without any modification)
Coding configuration	IPPP structure under the Low-Delay-P (LDP)
Rate control	Enabled
Hierarchical bit	Adaptive ratio bit allocation
Test sequences	Class B (FHD), Class C (WVGA), Class D (WQVGA), and Class E (HDTV)
DNN framework	TensorFlow-GPU with CUDA 9.0

-75.57%, respectively. Conversely, the proposed algorithm can yield moderate coding improvements for 'PartyScene,' 'Kimono,' and 'RaceHorses' that have more moving textures and more motions.

Comparisons of the objective performances of the proposed and other algorithms in the HM-16.20, URQ, OBA, and MPE models are presented in Tables 5–8. The BD-BR performances of the proposed algorithm and other algorithms applying the PSNR, SSIM, and MSSSIM metrics are compared in Table 5. From the table, the proposed algorithm outperforms the coding efficiency in all measured metrics against all other CTU-level rate control algorithms. Specifically, in terms of BD-BR-PSNR, the proposed algorithm provides compelling coding improvements, on average, by approximately -25.34% compared with the R–λ model in HM-16.20, -35.72% coding gain over the URQ model, -19.71%, and -21.87% gain against the OBA and MPE models, respectively. For BD-BR-SSIM and BD-BR-MSSSIM performances in ascending order, the proposed algorithm achieves an average gain of approximately $\{-29\%, -35\%, -43\%, -49\%\}$ and $\{-27\%, -35\%, -42\%, -49\%\}$ over the OBA, MPE, HM-16.20, and URQ, respectively. From Table 5, Class E contributes the highest BD-BR gain in all metrics that significantly influences the averaged BD-BR results of the proposed algorithm. The visual characteristics of sequences classified to Class E, i.e., 'FourPeople,' 'Johnny,' and 'KristenAndSara,' play a prominent role in obtaining higher objective measures for the proposed algorithm.

For the PSNR, SSIM, and MSSSIM quality comparisons in Table 6, the proposed algorithm can produce substantially more quality over the other algorithms, mainly against HM-16.20, OBA, and MPE. In other words, the proposed algorithm cannot attain the PSNR difference from the URQ model because URQ tends to require more bitrates than the given target bits abnormally during the encoding process.

$$\lambda'_{new} = max \left\{ max \left(\lambda'_{prevCTU} \times 2^{\frac{-1.0}{3.0}} \Big/ \left(2 \times VGG_{prevCTU}\right), \lambda_{pic} \times 2^{\frac{-2.0}{3.0}} \Big/ 2 \times VGG_{prevCTU} \right), \right.$$

$$\left. min \left\{ \lambda'_{new}, min \left(\lambda'_{prevCTU} \times 2^{\frac{1.0}{3.0}} \Big/ \left(2 \times VGG_{prevCTU}\right), \lambda_{pic} \times 2^{\frac{2.0}{3.0}} \Big/ 2 \times VGG_{prevCTU} \right) \right\} \right\} \qquad (10)$$

$$QP'_{new} = max \left\{ max \left(QP'_{prevCTU} - 11, QP_{pic} - 12 \right), min \left\{ QP'_{new}, min \left(QP'_{prevCTU} + 11, QP_{pic} + 12 \right) \right\} \right\} \qquad (11)$$

TABLE 4. Objective comparisons between rate control of HM-16.20 and the proposed algorithm.

Class	Sequence	ΔBE	ΔY-PSNR	ΔY-SSIM	ΔY-MSSSIM	BD-BR-PSNR	BD-BR-SSIM	BD-BR-MSSSIM
	Kimono	-0.10	-0.02583	0.00263	0.00205	-10.18%	-16.09%	-6.91%
	ParkScene	-0.01	0.87695	0.01691	0.03592	-21.35%	-35.95%	-41.19%
B	Cactus	0.08	1.11895	0.00888	0.03371	-21.11%	-41.43%	-41.08%
	BasketballDrive	-0.14	0.39003	0.00411	0.00246	-22.86%	-32.52%	-18.76%
	BQTerrace	-0.07	1.27088	0.01388	0.04052	-33.11%	-57.47%	-57.82%
	BasketballDrill	-0.02	0.73988	0.02374	0.02677	-14.33%	-25.41%	-23.18%
C	BQMall	-0.02	0.94168	0.02052	0.03940	-14.44%	-28.85%	-32.04%
	PartyScene	-0.02	0.66030	0.02046	0.03938	-9.69%	-18.06%	-20.64%
	RaceHorses	-0.09	0.52687	0.01250	0.01792	-19.25%	-36.36%	-31.22%
	BasketballPass	-0.22	-0.20038	0.02558	0.02571	-19.58%	-48.36%	-45.97%
D	BQSquare	0.25	2.18133	0.03464	0.05534	-28.01%	-64.56%	-67.88%
	BlowingBubbles	-0.06	0.34032	0.01661	0.04179	-14.93%	-37.05%	-41.92%
	RaceHorses	-0.17	-0.32453	0.00895	0.00920	-10.70%	-21.33%	-14.41%
	FourPeople	-0.01	2.41808	0.03341	0.05509	-53.35%	-76.29%	-75.57%
E	Johnny	-0.01	2.45150	0.00994	0.02446	-57.08%	-75.86%	-72.57%
	KristenAndSara	-0.04	2.61673	0.01246	0.03557	-55.47%	-76.29%	-73.09%

TABLE 5. BD-BR comparisons under the PSNR, SSIM, and MSSSIM metrics of the proposed algorithm and other existing rate control algorithms.

Metric	Class	HM-16.12 vs. Pro	URQ vs. Pro	OBA vs. Pro	MPE vs. Pro
	B	-21.72%	-23.16%	-19.90%	-22.28%
	C	-14.43%	-32.72%	-10.09%	-10.01%
BD-BR-PSNR	D	-18.31%	-41.55%	-9.67%	-21.27%
	E	-55.30%	-52.89%	-45.61%	-37.80%
	Average	**-25.34%**	**-35.72%**	**-19.71%**	**-21.87%**
	B	-36.69%	-55.35%	-29.23%	-34.04%
	C	-27.17%	-39.08%	-12.97%	-22.85%
BD-BR-SSIM	D	-42.82%	-51.24%	-17.83%	-41.76%
	E	-76.15%	-60.02%	-63.10%	-43.60%
	Average	**-43.24%**	**-49.08%**	**-28.67%**	**-34.96%**
	B	-33.15%	-52.09%	-24.09%	-35.86%
	C	-26.77%	-40.79%	-11.71%	-23.33%
BD-BR-MSSSIM	D	-42.54%	-52.87%	-22.80%	-40.04%
	E	-73.74%	-59.03%	-60.17%	-42.28%
	Average	**-41.52%**	**-48.86%**	**-27.44%**	**-34.98%**

TABLE 6. Y-PSNR, Y-SSIM, and Y-MSSSIM comparisons of the proposed algorithm and other existing rate control algorithms.

Metric	Class	HM-16.12 vs. Pro	URQ vs. Pro	OBA vs. Pro	MPE vs. Pro
	B	0.72620	-1.17399	0.76270	1.03013
	C	0.71718	-0.36693	0.50504	0.70983
BD-BR-PSNR	D	0.49919	-0.84811	0.18094	0.65527
	E	2.49543	-0.34276	1.88977	1.74074
	Average	**0.99892**	**-0.73490**	**0.76417**	**0.98958**
	B	0.00928	0.00480	0.00840	0.00955
	C	0.01931	0.00317	0.01183	0.01918
BD-BR-SSIM	D	0.02144	0.00316	0.01190	0.02106
	E	0.01860	0.00183	0.01489	0.01475
	Average	**0.01658**	**0.00342**	**0.01135**	**0.01581**
	B	0.02293	0.00377	0.01975	0.01217
	C	0.03087	0.00344	0.01835	0.01717
BD-BR-MSSSIM	D	0.03301	-0.00039	0.02318	0.01631
	E	0.03837	0.00017	0.03073	0.01227
	Average	**0.03033**	**0.00197**	**0.02232**	**0.01447**

Therefore, the giher PSNR scores can be obtained. Consequently, the URQ model bit rate accuracy and bitrate error are also less precise than the other algorithms, as shown in Table 7 and Table 8, respectively. Fortunately, among all existing models for the SSIM and MSSSIM comparisons, the proposed algorithm is the most effective. However, a relatively small SSIM and MSSSIM rate is achieved when compared with the URQ model, on average, at 0.00342 and 0.00197, respectively. These effects are mainly caused by the above issues of the URQ model. Accordingly, the proposed

TABLE 7. Bitrate accuracy comparisons between the proposed and other algorithms.

Class	HM-16.12	URQ	OBA	MPE	Proposed
B	0.03106	0.82214	0.02208	-0.02767	-0.01609
C	-0.03203	0.55009	-0.03482	-0.07626	-0.06605
D	-0.02604	1.24161	-0.04288	-0.02781	-0.07608
E	0.50896	2.48993	0.51166	0.46054	0.48988
Average	**0.09062**	**1.17171**	**0.08341**	**0.05169**	**0.05129**

TABLE 8. Bitrate error comparisons of the proposed algorithm and other existing algorithms.

| Class	HM-16.12 vs. Pro	URQ vs. Pro	OBA vs. Pro	MPE vs. Pro
B	-0.04715	-0.83823	-0.03817	0.01158
C	-0.03401	-0.61614	-0.03123	0.01021
D	-0.05005	-1.31769	-0.03320	-0.04828
E	-0.01909	-2.00005	-0.02178	0.02933
Average	**-0.03933**	**-1.12042**	**-0.03212**	**-0.00040**
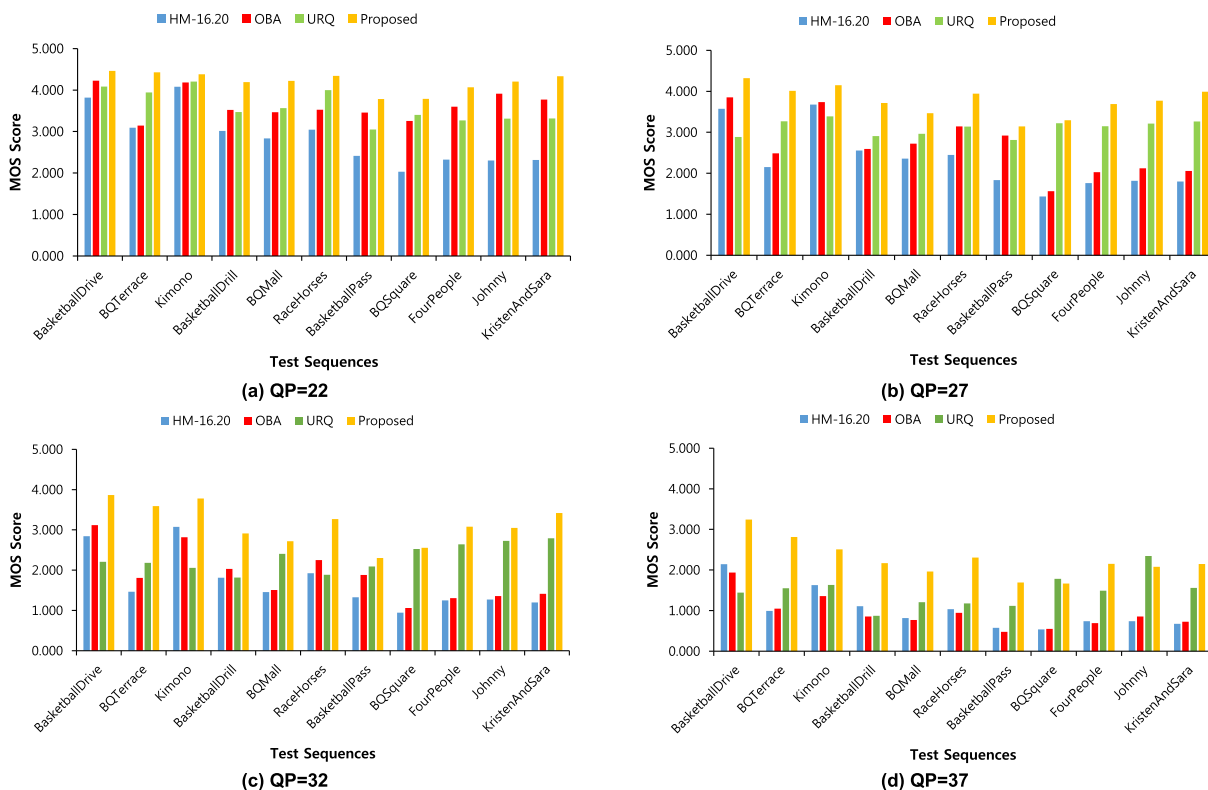
FIGURE 6. MOS comparisons of HM-16.20, OBA, URQ, and the proposed algorithms.

algorithm exhibits the highest performance for bitrate error comparisons of approximately -1.12042 against the URQ model, as shown in Table 8.

B. SUBJECTIVE PERFORMANCE EVALUATIONS

The main goal of the subjective quality evaluation is to compare the proposed CTU-level rate control algorithm with other algorithms, including HM-16.20, URQ [17], and OBA [28]. For all the test sequences, the double stimulus continuous quality scale (DSCQS) method [47] was performed. Sixteen reviewers participated in the test, of which 11 were in the relative field, and the rest were naïve in image processing. Simple demonstrations were conducted to introduce the evaluation process to the reviewers. The reconstructed

frames from the proposed algorithm, HM-16.20, URQ, and OBA, were randomly displayed twice with all the QP values for each participant. Then, the observers were instructed to give MOS values on a continuous scale ranging from 1 to 5. Finally, the MOS values were processed and depicted in the results, as shown in Figure 6. As shown in Figure 6, the proposed algorithm produces higher visual quality at all QP settings compared to conventional algorithms. Moreover, the proposed algorithm controls a better tradeoff between the rate and distortion with QP = 37 than the competing models and still maintains visual quality. In figure 6, the URQ model is the competing algorithm with the smallest MOS against the proposed algorithm. This is mainly due to the inability of the URQ to control the overflow bitrates. In our analy-

BQTerrace (1920x1080)
PartyScene (832x480)

R– model

URQ model

OBA model
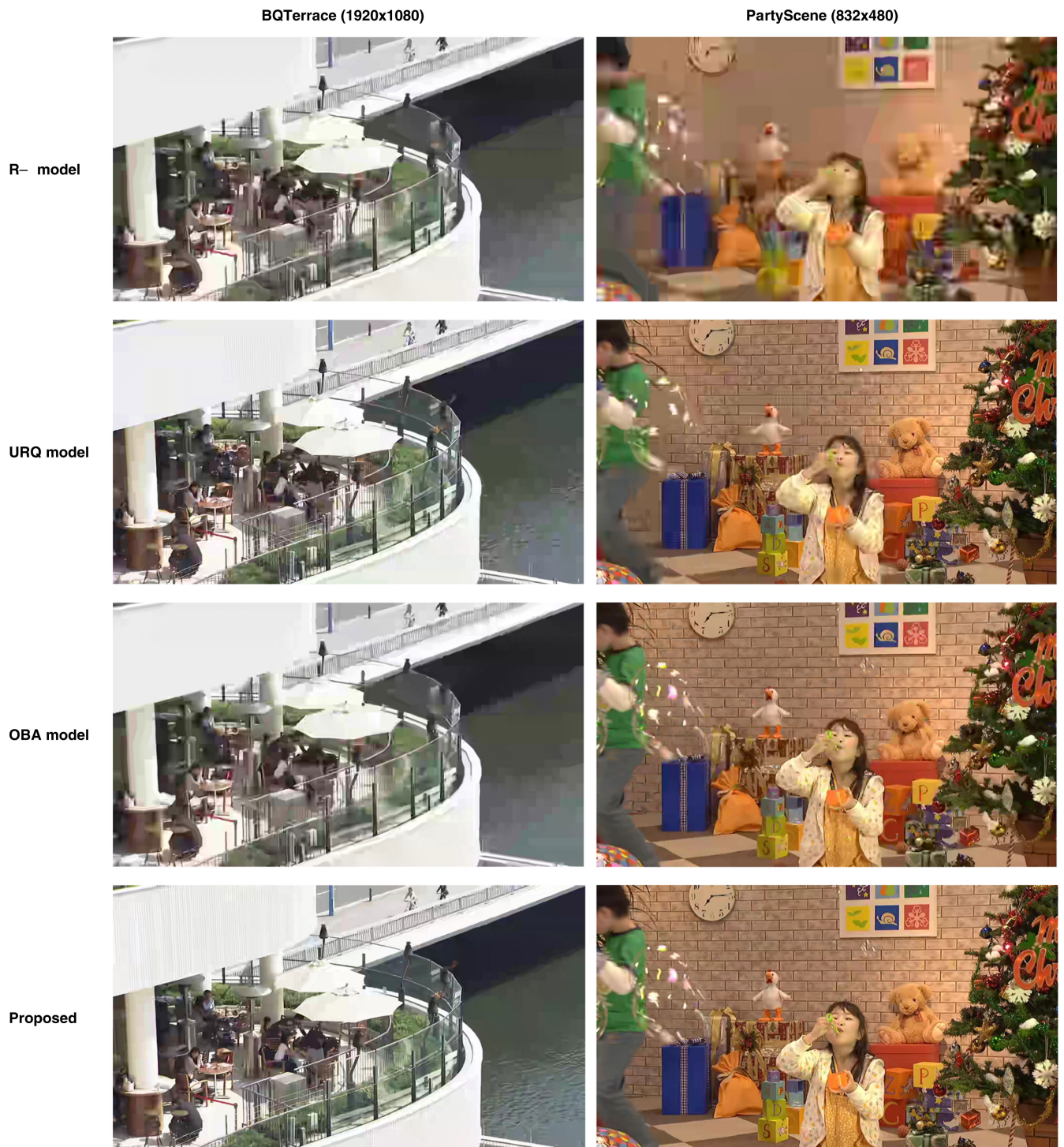
Proposed

FIGURE 7. Visual quality comparisons of the proposed and other existing algorithms on P-frame #77 of 'BQTerrace' sequence coded at 2,352 Kbps (left) and P-frame #89 of 'PartyScene' sequence coded at 1,559 Kbps (right).

sis, the URQ model tends to require more bits and abnormally exceeds the given target bitrate. In the reconstruction frames, the model ensures better visual quality than the other algorithms. In Figure 7, two example frames from POC = 77 of 'BQTerrace' coded at 2,352 Kbps and POC = 89 of 'PartyScene' coded at 1,559 Kbps are presented to show the visual quality comparisons of the proposed and the other existing algorithms. From the visual quality comparisons depicted in Figure 7, the proposed algorithm achieves better visual quality in the two frame examples than in those produced by the HM-16.20, URQ, and OBA rate control models.

Class	Sequence	HM-16.12 vs. Pro	URQ vs. Pro	OBA vs. Pro
	BasketballDrive	0.879	1.318	0.689
B	BQTerrace	1.786	0.975	1.589
	Kimono	0.589	0.882	0.679
	BasketballDrill	1.125	0.982	0.996
C	BQMall	1.226	0.557	0.976
	RaceHorses	1.354	0.916	1.000
D	BasketballPass	1.193	0.463	0.546
	BQSquare	1.589	0.093	1.220
	FourPeople	1.732	0.611	1.342
E	Johnny	1.745	0.378	1.214
	KristenAndSara	1.975	0.739	1.480
	Average of Class B	1.085	1.058	0.986
	Average of Class C	1.235	0.818	0.991
	Average of Class D	1.391	0.278	0.883
	Average of Class E	1.817	0.576	1.345
	Overall	**1.382**	**0.719**	**1.067**

To verify the visual quality comparisons from the existing algorithms and the proposed algorithms, the DMOS scores were then calculated, respectively symbolized as $MOS_{otherModels}$ and $MOS_{proposed}$, which can be defined by

$$DMOS = MOS_{proposed} - MOS_{otherModels} \qquad (17)$$

Table 9 shows all the DMOS test sequences. The average DMOS per sequence for all the QP values is listed to quickly find the visual quality comparisons of the generated reconstruction frames. Positive values indicate that the video quality of the proposed algorithm is subjectively better than that of existing algorithms. As tabulated in Table 9, the DMOS scales for the entire test sequences are moderately outperformed over the HM-16.20, URQ, and OBA models.

C. COMPLEXITY PERFORMANCE EVALUATIONS

The proposed algorithm requires a tradeoff between a significant performance and running time in comparison with the existing rate control algorithm in HM-16.20. The additional complexity originates from constructing the high-level features from the original and reconstructed CTUs using the VGG-16 network. However, the proposed algorithm can be fully-optimized in parallel to speed up CTU-by-CTU feature extractions. Hence, throughput can be enhanced with a parallel machine such as GPU that suppresses more encoding time. In addition, the proposed CTU-level rate control with the predefined VGG-16 employment requires about $24\times$ more encoding time over the HM-16.20 reference software with rate control enabled. The proposed algorithm also takes a higher running time of about $18\times$ than the MPE rate control model.

V. CONCLUSIONS

In this paper, a deep-learning feature-based CTU-level rate control is proposed to obtain better objective and subjective coding performance for HEVC under the low-delay-P configuration. The proposed algorithm utilizes a predefined model of the VGG-16 network to extract features from both the original and reconstructed CTU blocks. The proposed algorithm

was designed by exploring a perceptual loss function based on the extracted features combined with the QP value of each CTU to remodel the estimation functions of the existing CTU-level rate control of HM-16.20. Compared to the anchor, the proposed algorithm conludes better rate control performances by enhancing the visual quality through significant coding improvements. For future work, the proposed algorithm in [23] will be advanced with several more adjustments that benefit the rate control model for the random-access configuration of HEVC.

REFERENCES

[1] 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.

[2] *MPEG-2 Reference Software*. Accessed: Sep. 2020. [Online]. Available: https://mpeg.chiariglione.org/technologies/reference-software/mpeg-2-reference-software

[3] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[4] K.-P. Lim, G. J. Sullivan, and T. Wiegand, *Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods*, document JVT-N046, Joint Video Team, 2005.

[5] 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.

[6] 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.

[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] 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.

[9] 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.

[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] 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.

[12] 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.

[13] 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.

[14] 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.

[15] M. Ismail, J. Ma, and D. Sim, "Full depth RQT after PU decision for fast encoding of HEVC," in *Proc. 18th IEEE Int. Symp. Consum. Electron. (ISCE)*, Jeju Island, South Korea, Jun. 2014, pp. 1–2.

[16] H. Choi, J. Nam, J. Yoo, D. Sim, and I. Bajic, *Rate Control Based on Unified RQ Model for HEVC*, document JCTVC-H0213, Joint Collaborative Team on Video Coding, 2012.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] M. Zhou, Y. Zhang, B. Li, and X. Lin, "Complexity correlation-based CTU-level rate control with direction selection for HEVC," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 4, pp. 1–23, Aug. 2017.

[22] J. Yoo, J. Nam, J. Ryu, and D. Sim, "Rate control to reduce bitrate fluctuation on HEVC," *IEIE Trans. Smart Process. Comput.*, vol. 1, no. 3, pp. 152–160, Dec. 2012.

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

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

[25] 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.

[26] M. Zhou, X. Wei, S. Wang, S. Kwong, C.-K. Fong, P. H. W. Wong, W. Y. F. 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.

[27] L. Li, B. Li, H. Li, and C. Wen 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.

[28] 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.

[29] B. Li, M. Zhou, Y. Zhang, X. Lin, and W. Guo, "Model parameter estimation for CTU-level rate control in HEVC," *IEEE MultimediaMag.*, vol. 25, no. 3, pp. 79–91, Jul. 2018, doi: 10.1109/MMUL.2018.112142602.

[30] 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.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[32] B. Li, H. Li, and L. Li, *Adaptive Bit Allocation for R-λ Model Rate Control in HM*, document JCTVC-M0036, Joint Collaborative Team on Video Coding, 2013.

[33] S. Li, M. Xu, X. Deng, and Z. Wang, "Weight-based R-λ rate control for perceptual HEVC coding on conversational videos," *Signal Process., Image Commun.*, vol. 38, pp. 127–140, Oct. 2015.

[34] 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.

[35] 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.

[36] 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.

[37] J. Yao, X. Song, S. Fang, and L. Wang, *AHG9: Convolutional Neural Network Filter for Inter Frame*, document JVET-J0043, Joint Video Experts Team, 2018.

[38] T. Hashimoto and E. S. T. Ikai, *AHG9: Separable Convolutional Neural Network Filter With Squeeze-and-Excitation Block*, document JVETK0158, Joint Video Experts Team, 2018.

[39] Y.-L. Hsiao, C.-Y. Chen, T.-D. Chuang, C.-W. Hsu, Y.-W. Huang, and S.-M. Lei, *AHG9: Convolution Neural Network Loop Filter*, document JVET-K0222, Joint Video Experts Team, 2018.

[40] 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.

[41] 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.

[42] 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.

[43] 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*. [Online]. Available: http://arxiv.org/abs/1904.12462

[44] 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*. [Online]. Available: http://arxiv.org/abs/1904.03567

[45] H. Choi and I. V. Bajic, "Deep frame prediction for video coding," 2019, *arXiv:1901.00062*. [Online]. Available: http://arxiv.org/abs/1901.00062

[46] F. Bossen, *Common HM Test Conditions and Software Reference Configurations*, document JCTVC-L1100, Joint Collaborative Team on Video Coding, 2013.

[47] S. Pateux and J. Jung, *An Excel Add-in for Computing Bjontegaard Metric and Its Evolution*, document VCEG-AE07, Video Coding Experts Group, 2007.

ISMAIL MARZUKI received the B.S. degree in informatics from UIN Sultan Syarif Kasim Riau, Indonesia, in 2011, and the M.S. degree in computer engineering from Kwangwoon University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree. He joined the Image Processing Systems Lab (IPSL), in 2013. He is involved in video compression standardization meetings, including JCTVC and JVET. His research interests are related to high-efficiency video coding (HEVC/x265) techniques, fast coding, rate control, and versatile video coding (VVC), and deep learning.
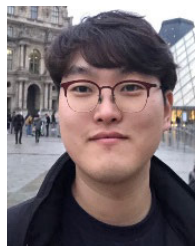
JONGSEOK LEE received the B.S. and M.S. degrees in electronic engineering from Kwangwoon University, South Korea, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in computer engineering. He is involved in video compression, including the standardization of the H.266 /VVC video codec. His current research interests are video coding, video processing, computer vision, spiking neural networks, and deep learning.

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 the Hyundai Electronics Company Ltd., from 1999 to 2000, being involved in MPEG-7 standardization. He was a Senior Research Engineer with Varo Vision Company Ltd., working on MPEG-4 wireless applications, from 2000 to 2002. He worked at the Image Computing Systems Laboratory. (ICSL), University of Washington, as a Senior Research Engineer from 2002 to 2005. He researched 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 the Simon Frasier University as a Visiting Scholar. He is one of the leading inventors in many essential patents licensed to MPEG-LA for HEVC standard. His current research interests are video coding, video processing, computer vision, and video communication.

● ● ●