IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 15, 2020, accepted February 13, 2020, date of publication February 24, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976142

Perceptual Adaptive Quantization Parameter Selection Using Deep Convolutional Features for HEVC Encoder

ISMAIL MARZUKI AND DONGGYU SIM
Department of Computer Engineering, Kwangwoon University, Seoul 139701, 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, under the Information Technology Research Center (ITRC) supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP), under Grant IITP-2019-2016-0-00288, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) through the Ministry of Science, ICT & Future Planning under Grant NRF-2018R1A2B2008238.

ABSTRACT In this paper, we propose a perceptual adaptive quantization based on a deep neural network on high efficiency video coding (HEVC) for bitrate reduction while maintaining subjective visual quality. The proposed algorithm adaptively determines frame-level QP values for different picture types of the hierarchical coding structure in HEVC by taking into account the high-level features extracted from the original and previously reconstructed pictures. A predefined model based on the visual geometry group (VGG-16) network is exploited to extract the high-level features for subjective visual characteristics. Furthermore, the Lagrange multiplier for each frame is also adaptively determined by involving the proposed features for deciding the appropriate parameter of the Lagrange multiplier that can be used for rate-distortion optimization during the encoding process. Experimental results reveal that the proposed perceptual adaptive QP selection can facilitate bitrate savings up to 65.73% and 47.68% and improve the BD-rate based on SSIM by approximately 20.68% and 14.27% under low-delay-P and random-access coding structures, respectively, with very minimal visual quality degradation when compared to HM-16.20 without adaptive QP selection.

INDEX TERMS Adaptive quantization parameter, deep neural network, high efficiency video coding (HEVC), perceptual quantization parameter, VGG-16 network, video coding.

I. INTRODUCTION

High-efficiency video coding (HEVC) standard has been widely accepted to achieve better compression performance over H.264/Advanced Video Coding (AVC) by maintaining similar visual quality [1]. It has encompassed various video media services and applies not only to full high definition (FHD) but also to 4K/8K ultra-HD (UHD) [2]–[4]. Since the standard was released, many studies have been conducted for the sake of its advantages of visual quality improvement [5]–[7], computational complexity reduction [8]–[16], bitrate reduction [17], [18], and prospects as a future video coding standard [19]–[26]. Among many coding tools, rate-distortion optimization (RDO) in the HEVC software model (HM) [26]–[28] is used to improve its coding efficiency [30], [31]. It is based on optimization using the global Lagrange multiplier and determines the quantization parameter (QP) value using a QP- λ model. The Lagrange multiplier λ can be termed as a function of the quantization step size, which is closely related to the QP value. It is used for the coding efficiency of each basic unit by selecting the best coding mode under a given QP value, where the basic unit can be a frame, slice, or coding unit (CU). The common test condition (CTC) designed by the Joint Video Experts Team (JVET) employs static quantization parameters for fair comparison in standardization [32]. However, an adaptive QP selection is known to be effective in improving subjective visual quality for practical applications. The adaptive QP should be designed to be harmonized within the RDO process. It can adjust the QP value for a distinctive frame or slice according to different spatial, temporal, or visual aspects. Some studies have discovered approaches to improve the compression rates [33]–[37] or visual quality [38]–[44] with various adaptive QP techniques. Typically, these studies

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

prioritize the determination of optimum QPs for the RDO process to produce better encoding parameters by analyzing the QP- λ relationship or by observing the effectiveness of spatial-temporal dependencies among the basic units. Generally, these studies take into consideration the essential role of λ in the RDO process. Thus, it will be interesting to consider a deep neural network (DNN) for more varied QPs in HEVC. Studies have prevailed benefits of DNN for video coding [45]–[49]. However, there is no existing effective DNN-based algorithm for perceptual adaptive QP purposes.

This study presents a DNN-based QP selection method by the adaptive determination of frame-level perceptual QP for HEVC to achieve bitrate reduction without inducing visual quality degradation. The proposed algorithm is embedded in HM-16.20 and generates QP values adaptively for different picture types and coding structures in HEVC. The proposed algorithm first determines a QP for the first frame in a sequence by averaging the standard deviation value of the original blocks (StD). Then, the proposed algorithm obtains high-level features from the original and reconstructed frames using a pretrained visual geometry group (VGG-16) network model [50]. Based on the extracted high-level features, more visual-friendly QP is then distributed for the next consecutive frames in the encoding order. The algorithm also determines the Lagrange multiplier adaptively for each frame based on the proposed model, which can be used for RDO in the encoding process. As a result, the proposed algorithm demonstrates significant coding gain with minimal visual degradation against HM-16.20 and other existing adaptive QP algorithms.

The rest of this paper is organized as follows. In section 2, we briefly present an overview of the QP decision in HM and related works. In section 3, we discuss the proposed perceptual adaptive QP for HM. In section 4, we review several performance evaluations of the proposed algorithm, and finally, we draw the conclusions and suggest further research directions in section 5.

II. CURRENT STATE OF QP SELECTION AND RELATED STUDIES OF PERCEPTUAL ADAPTIVE QP IN HEVC

The current QP selection within the RDO process in HEVC is not optimal. Many studies have revealed several weaknesses of the QP selection technique in the HEVC encoder. In this section, several adaptive QP techniques for HEVC are discussed as follows.

A. GENERAL QP SELECTION CONCEPT IN HM

QP selection in video coding can be mathematically described as an RDO problem [35], [36] that minimizes the total coding distortion D at a given bitrate R_T as:

$$QP^* = \left(QP_i^*, \cdots, QP_N^*\right) = \arg\min\left(QP\right) \sum_{i=1}^{N} D_i, \quad s.t. \sum_{i=1}^{N} R_i \leq R_T \tag{1}$$

where N denotes the number of basic units, D_i is the coding distortion, and R_i is the coding bitrate of the $i-$th basic unit. Note that the basic unit in HEVC term may be a frame, slice, or CU. D_i and R_i in (1) form on $QP = (QP_i, \cdots, Q_N)$. QP_i refers to the QP value for the $i-$th basic unit and $QP^* = (QP_i^*, \cdots, QP_N^*)$ represents the optimal QP set for the N basic units. Applying the λ method [29] into the following unconstrained form, equation (1) can be rewritten as:

$$QP^* = \arg\min\left(QP\right)\{J\}, \quad J = \sum_{i=1}^{N} D_i + \lambda \cdot \sum_{i=1}^{N} R_i \tag{2}$$

where J stands for the total rate-distortion *(RD)* cost function, and λ represents the trade-off parameter between D_i and R_i. Along with the RDO process, λ in HEVC can be obtained as

$$\lambda = QP_{factor} \cdot 2^{QP/3}, \tag{3}$$

where QP denotes the quantization parameter, and QP_{factor} is a constant parameter related to coding configurations. The QP value in (3) is an integer introduced to represent an actual quantization step size by an exponential mapping function. However, the quantization step size in HEVC tends to be static for complexity reduction in the RDO process. Applying a fixed or predefined QP scheme may cause the compression rate to drop significantly, while HEVC has different coding configurations. Hence, this becomes a major challenge for any QP method design in HEVC. Many QP adjustment methods have been studied for better coding gain. For example, a QP–λ relationship is used to determine the λ value according to an initial QP, and subsequently, the new QP value is recalculated [30], [31]. This algorithm is widely known as a straight-forward algorithm for the RDO scheme in HEVC. Wang *et al.* [33] introduced an improved block-level adaptive QP value that considers previously coded block information. Zhao *et al.* [34] proposed a QP cascading scheme that assigns QP values to different hierarchical temporal picture layers. Similar algorithms were also introduced by Li *et al.* [35] and He *et al.* [36], which presented only an inter-frame dependency technique. As far as we know, these last two algorithms can provide better coding gain for an HEVC encoder. Extensive use of spatial-temporal predictions in HEVC is important for adaptive QP selection in RDO. Although the integration of such propagation effects is desirable, there are not many such studies.

B. EXISTING METHODS OF PERCEPTUAL ADAPTIVE QP SELECTION FOR HM

Determining the QP value for video encoders also affects the entirely visual quality of a video sequence. To improve the subjective quality of adaptive QP, the spatial and temporal features or combination of those may be designed empirically. Open software of × 265 [38] becomes one of several algorithms that developed a perceptual adaptive QP method with spatial and temporal features. However, it still

fails to give promising outcomes if a reference frame has characteristics different from the current coding frame. Test Model 5 (TM5 Model) of MPEG-2 software [39] also uses the method that scales a quantization step according to the spatial activity of one CU relative to a frame-level average of the spatial activity. This method fails when the size of a large CU block needs to be estimated, thus limiting its performance [37]. Similarly, Yeo *et al.* [40] also introduced a block-level adaptive QP selection algorithm. It observes the spatial and temporal pixel characteristics of CU blocks. However, it needs a higher encoding time. Prangnell *et al.* [41] used transform coefficients based on a soft thresholding method. However, the proposed soft thresholding method may still cause fluctuations of the visible quality, resulting in severe visual distortion.

An alternative algorithm was proposed by determining a QP offset based on a $QP - \lambda$ relationship that is formed. Yeo *et al.* [40] has also studied related topics. However, their method utilized only the spatial variance of a block, which is limited for videos with large homogeneous areas [42]. Xiang *et al.* [43] proposed a perceptual motion estimation method using a spatial-temporal just-noticeable-distortion (JND) model for a QP offset design. Rouis *et al.* [44] generated perceptual features temporally as well as CTU visual sensitivity for spatial features. However, both features considered in this algorithm are provided only for an adaptive λ in RDO. As a conclusion, spatial and temporal perceptual features for an adaptive QP decision can provide a better trade-off [43], [44].

C. DNN APPROACH TO PERCEPTUAL ADAPTIVE QP SELECTION FOR HM

The use of DNN for video coding has now become possible for the video coding community. Liu *et al.* [45] and Ma *et al.* [46] have presented case studies on deep learning-based video coding. Several researchers such as Choi and Bajic [47] studied a deep learning-based frame prediction using decoded frames to predict the textures of a block. It performs both uni- and bi-directional predictions at various distances from a target frame. Ki *et al.* [48] developed a JND model based on deep learning for the assessment of perceptual distortion in HEVC. Li *et al.* [49] proposed a DNN-based rate control for Intra coded pictures in HEVC that is designed to predict the parameters of the $R - \lambda$ rate control model. Other studies have successfully revealed the benefits of deep learning for video encoding. However, it is still difficult to find one specific deep learning method for a perceptual adaptive QP. In this paper, we present a perceptual adaptive QP based on a predefined VGG network for HEVC.

III. PROPOSED ALGORITHM FOR PERCEPTUAL ADAPTIVE QP SELECTION FOR HEVC ENCODER

The main objective of the proposed algorithm is to achieve significant bitrate savings without inducing noticeable visual distortions in reconstructed video frames. We first observed the current setting of the $QP - \lambda$ relationship in HEVC, as shown in (3). The two main factors involved are the QP_{factor} and QP value. Frame-level QP decision in HM-16.20 is determined with the same QP offset for multiple frames in the same temporal ID layer, while the QP_{factor} denotes for the coding structure parameter is always set static as 0.57, regardless of frame or slice types and coding structures. In HEVC, the different frames form a set of hierarchical structures within a group of pictures, GOP. For example, frames at a higher temporal layer in the same GOP can be predicted from one or more frames at the lower temporal layers. Therefore, giving only the default value of QP offset and QP_{factor} to generalize different frames and coding structures is not perceptually wise for HEVC encoders. Both spatial and temporal features could be sufficient to resolve the issues. However, most of the existing adaptive QP methods mainly concentrate only on one of both elements. In this paper, the proposed algorithm demonstrates visual feature extraction based on a particular convolutional layer of a DNN model for a frame-level adaptive QP. We consider both the spatial and temporal features to generate the adaptive QP and $\boldsymbol{QP_{factor}}$ decision for the proposed algorithm.

Fig. 1 depicts the whole process of the proposed algorithm. As shown in Fig. 1, the proposed algorithm is embedded in the HEVC encoder. The proposed algorithm is processed during the slice initialization. Depending on the slice or frame types, the QP value and $\boldsymbol{QP_{factor}}$ are determined adaptively. Fig. 2 shows the detailed process of the proposed algorithm. For the first frame in a sequence, the proposed algorithm is designed in a straightforward manner by considering the standard deviation values of the original frame to decide upon a QP value and set $\boldsymbol{QP_{factor}}$ as its default value. Then, a pretrained VGG-16 model is employed to extract visual features from the original and reconstructed frames to predict the QP and $\boldsymbol{QP_{factor}}$ for consecutive frames. The designed visual features result in a perceptual loss value based on the Euclidean distance measure, $\boldsymbol{VGG_{feature}}$. The QP and Lagrange multiplier values based on $\boldsymbol{VGG_{feature}}$ are then adaptively estimated by considering the picture types and coding configurations in HEVC. A detailed discussion of this section is divided into several sub-categories as follows. Symbols and descriptions used in the proposed algorithm of the adaptive frame-level perceptual QP for HEVC are tabulated in Table 1.

A. GENERATION OF VISUAL FEATURES FOR THE PROPOSED PERCEPTUAL ADAPTIVE QP ALGORITHM

We propose to adaptively adjust a perceptual QP value per frame by employing a deep learning network, namely, the VGG-16 network [50]. The proposed algorithm employs a pretrained VGG-16 model to construct high-level feature descriptors using a specific convolutional layer. We select VGG-16 for this study due to some of its desirable characteristics. VGG-16 is widely recognized for its remarkable performance on image classification, which classifies over 14 million images to 1000 categories. It has a better image classification accuracy than the AlexNet model [51]. It has

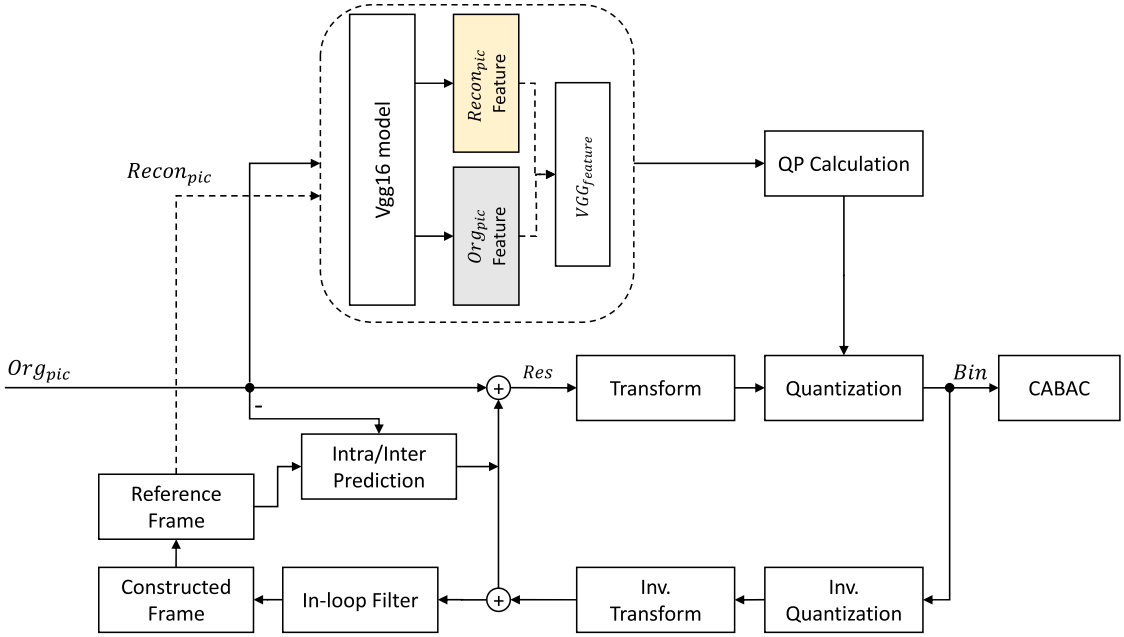

FIGURE 1. Block diagram of the proposed perceptual adaptive QP.

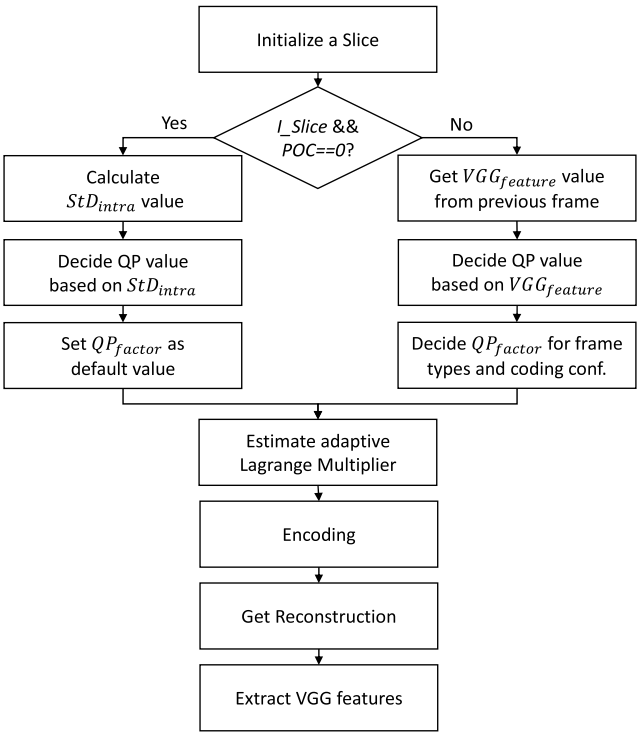

FIGURE 2. Overall flowchart of the proposed perceptual adaptive QP.

a straightforward architecture that is constructed simply by stacking convolution, pooling, and fully connected layers without branches or shortcut connections to reinforce gradient flow. Such a design is versatile and adaptable for different practical purposes. Besides, the VGG-16 has an extremely deep convolutional layer design used to train on an enormous and manifold image dataset, which results in convolution filters that are well suited to search universal patterns and generalize them. It is also widely applied as a feature extraction technique in many computer vision solutions [52], [53]. For the same reason, the proposed algorithm also takes advantage of the VGG-16 convolution layers only for visual feature extraction. In this paper, a simplified VGG-16 network is employed by removing the latest pooling and fully connected layers, as depicted in Fig. 3. In the figure, h and w represent the height and width of the input 64×64 CTU block, respectively. Fortunately, the VGG network can handle any input block size, as long as h and w are multiplication of 32. Hence, the CTU block size can be used directly without necessary prior processing. By examining the visualization of convolution filters and trial–and–error experiments, we selected 'block5conv1', which is the first-fifth convolution layer to build general features for the proposed algorithm. The 'pool5' layer is initially included in the network. However, it is neither considered for the algorithm nor included in the figure. The 'pool5' layer is commonly affected by specific classification objects, which is not favorable for the detection of general features. We mainly consider the generalizability of the VGG network, and thereby, the proposed feature descriptors can search for common and universal patterns.

For better features with HVS consideration, we introduce a perceptual loss function with a full-reference visual quality measure that uses the Euclidean distance. It is based on a comparison of different feature maps extracted from original and reconstructed blocks, as depicted in Fig. 4. The reconstructed block fed to the network is derived after the in-loop filter process. The figure shows that the same model of the

TABLE 1. Symbols and descriptions used in the proposed perceptual adaptive QP selection.

Symbols	Description
QP	Quantization parameter
QP_{factor}	Constant parameter related to coding structure parameter
QP_0	QP value of the first frame in a sequence
QP_{init}	Initial QP value set by the encoder
$QP_{perceptual}$	QP decision determined that considers the picture types
StD	Standard deviation value
StD_{intra}	Average *StD* of the total number CTUs within an original Intraframe
σ_i and μ_i	*StD* and mean values of the original $i-$th CTU block
$VGG_{feature}$	Proposed perceptual loss value
$D_{VGG_{pre}}$	VGG feature distortion of a predicted frame
$D_{VGG_{ref}}$	VGG feature distortion of a reference frame
$f(\cdot)$	Relationship between D_{VGGref} and D_{VGGpre}.
Fid	Hierarchical frame index
ΔpQP_{Fid_i}	Parameter of $QP_{perceptual}$ for Fid in P-frame
ΔbQP_{Fid_i}	Parameter of $QP_{perceptual}$ for Fid in B-frame
$D_{ref}^{(I)}$	Extracted feature of the reference Intraframe
$QP_{OffsetModelScale_i}$	Default model scale of QP offset in HM-16.20
$QP_{OffsetModelOffset_i}$	Default model offset of QP offset in HM-16.20
ΔQP_{Offset_i}	QP offset function
ΔpQP_{Offset}	Clip function of QP offset for P-frame case
ΔbQP_{Offset}	Clip function of QP offset for B-frame case
$I_{QP_{factor}}$	QP_{factor} for I-frame
$P_{QP_{factor}}$	QP_{factor} for P-frame
$B_{QP_{factor}}$	QP_{factor} for B-frame
c	Constant parameter to decide $P_{QP_{factor}}$
c_i	Constant parameter to decide $B_{QP_{factor}}$ for different Tid_i
Tid_i	Temporal ID index
POC	Picture order count
GOP_{size}	Group of pictures size

VGG-16 network is utilized for extracting those high-level features. The Euclidean distance is preferred owing to its simplicity in expressing $VGG_{feature}$ as a perceptual loss value. To do this, we first convert the color format of both the original and the reconstructed CTU blocks to the RGB color format. This process is suggested as a requirement of the VGG-16 architecture. Then, the network can operate adequately to obtain visual features from both input blocks. Once a $VGG_{feature}$ is generated, we then use it to determine the QP value and QP_{factor} adaptively for the Lagrange multiplier decision.

B. PERCEPTUAL ADAPTIVE QP DETERMINATION WITH QP-λ RELATIONSHIP

From the formula in (3), the QP value per frame can be derived. However, the λ value in HM-16.20, which represents the Lagrange multiplier is decided later after the QP decision is determined, while the QP value per frame is decided empirically based on the HM configuration. Therefore, finding a proper parameter for predicting a frame-level perceptual adaptive QP is a challenging issue.

Generally, coding errors may propagate from the previous frame to subsequent frames because of the prediction coding scheme in video coding standards. In this study, the proposed algorithm determines the frame-level QP for different picture types by obtaining a perceptual loss value based on high-level features from the original and previously reconstructed pictures. With regards to the first frame in a sequence, the determination of a proper QP value is crucial as it will determine the overall coding performance. However, having only an original picture is not enough to provide a perceptual loss value before the encoding. Hence, we examine whether the standard deviation values *(StD)* of the original blocks can demonstrate the characteristics of a complete picture for frame-level QP decision. We activated rate control to observe the different QP values of every CTU within the intraframe using the 'BasketballPass' test sequence with QP 22, 27, 32, and 37. Subsequently, a relationship between QP and *StD* is presented in Fig. 5. A lower *StD*, which reflects a flat region, tends to have a higher QP, vice versa. Therefore, we can expect some coding gain with lower visual quality depression in this area. However, applying the *StD* value directly to vary λ over the QP_{factor} may lead to high coding loss performance. Therefore, the QP decision in this algorithm is adjusted by firstly normalizing the pixel value of every CTU block in a frame before calculating *StD* and disregarded the λ and QP_{factor} for QP decision. Then, the QP of the first frame can be more visual-friendly provided and can be expressed as:

$$QP_0 = QP_{init} - 3\log_2(StD_{intra}) \tag{4}$$

$$StD_{intra} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i \tag{5}$$

$$\sigma_i = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(x_j - \mu_i)^2} \tag{6}$$

where QP_0 denotes the QP value of the first frame in a sequence, and QP_{init} represents the initial QP value set by the encoder. Since we design the proposed algorithm in CTU wise, the final picture characteristic of the first frame is decided based on the StD_{intra} value, which is the average *StD* of the total number N of the original CTU blocks in an Intra frame. Thus, the symbols σ_i and μ_i become the *StD* and mean values of the original $i-$th CTU block, respectively. M denotes the total number of pixel values x_j.

For the rest of the frames, the quality of the reconstruction frames is generally influenced by a previously coded frame with a certain QP value. In this study, instead of analyzing the distortion of two consecutive frames, we investigate the distortion of VGG features for determining a proper QP value perceptually. Note that the proposed VGG features are extracted from the original and reconstructed frames based on the VGG-16 model. Therefore, the distortion of VGG features of two consecutive frames can be expressed as

$$D_{VGGpre} = f\left(D_{VGGref}\right) \tag{7}$$

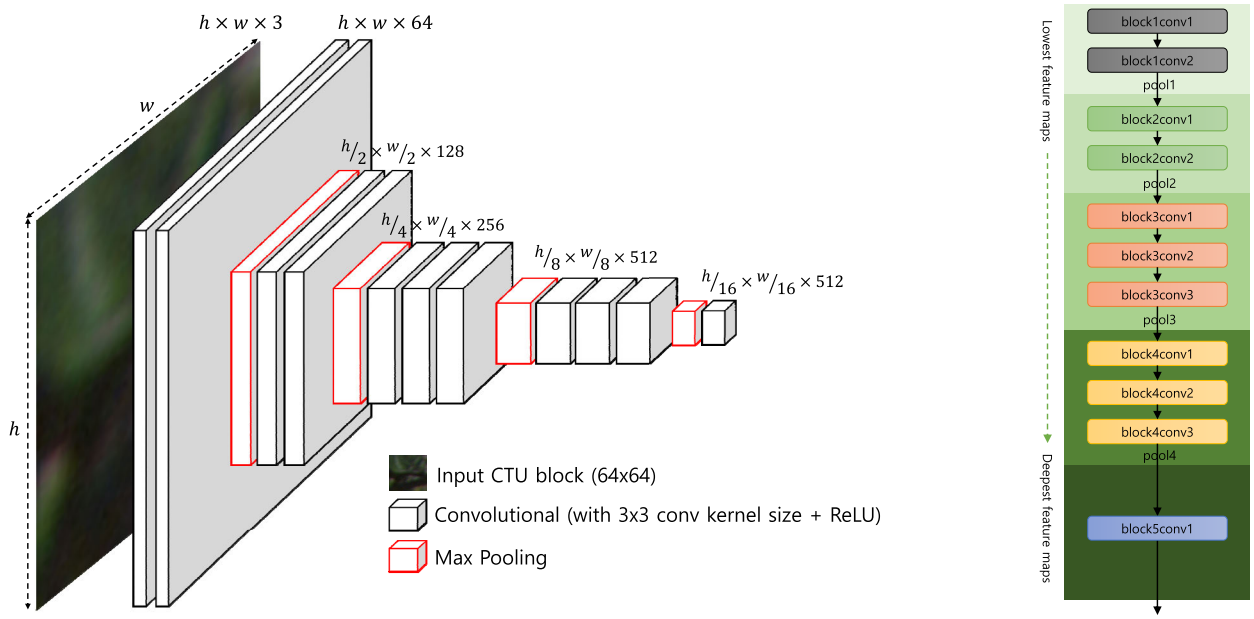

FIGURE 3. Proposed double-simplified VGG-16 network architecture.

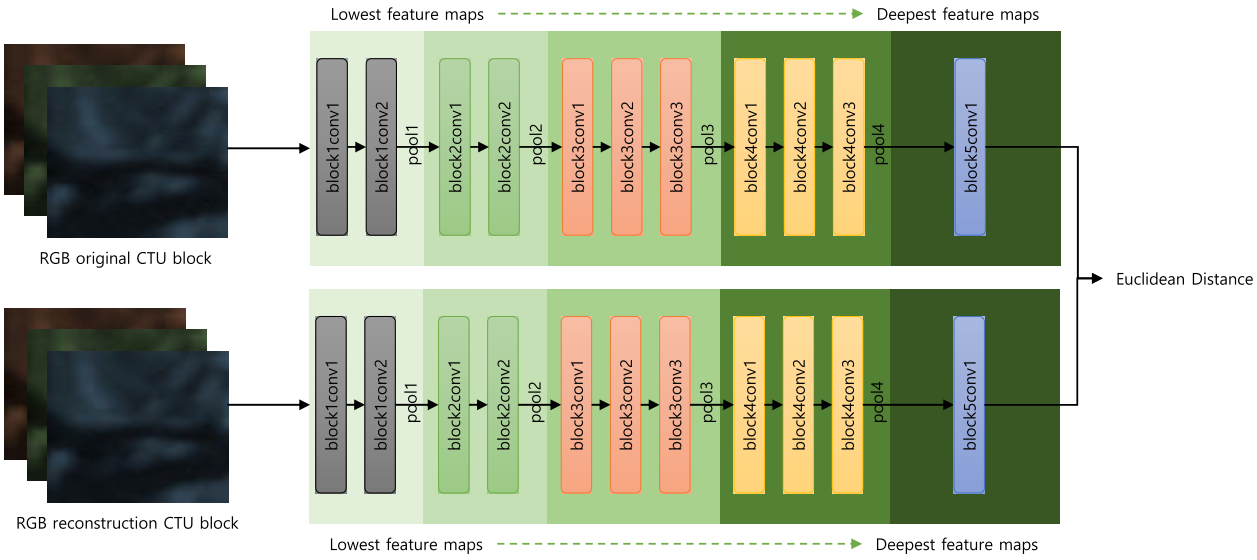

FIGURE 4. Proposed double-simplified VGG-16 network architecture.

where D_{VGGpre} is the VGG feature distortion of a predicted frame, D_{VGGref} denotes the VGG feature distortion of a reference frame, and $f(\cdot)$ is the relationship between D_{VGGref} and D_{VGGpre}.

Fig. 6(a) shows the VGG feature distortion relationship between two consecutive frames of the 'BasketballPass' test sequence. The sequence is encoded under LDP configuration with the coding structure of I-P-P-P-P. Each P frame uses only its previous coded frame as a reference. We set the predicted frame with a fixed QP value of 32 and encoded the first 15 frames. It can be seen that D_{VGGref} influences D_{VGGpre}. A further experiment was also conducted with rate control enabled to support the observations. Fig. 6(b) shows a high correlation between the VGG feature and QP selection per frame. Accordingly, the QP decision for the rest of the frame can be determined by considering the picture types as in (8).

The QP decision for a future intra picture can be determined by using the $VGG_{feature}$ from a previously intra coded picture. With regards to the QP decision for P- and B- frames, we control QP_{init} with ΔpQP_{Fid_i} and ΔbQP_{Fid_i} depending on the hierarchical frame index $i(Fid_i)$ as shown in Table 2. The values of ΔpQP_{Fid_i} and ΔbQP_{Fid_i} are derived

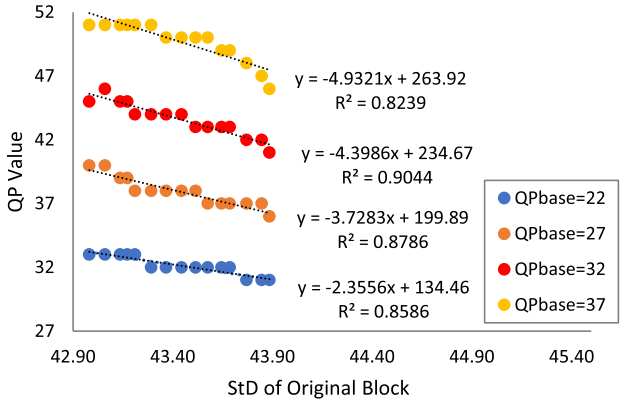

FIGURE 5. Correlation between *StD* value of original blocks and QP values.

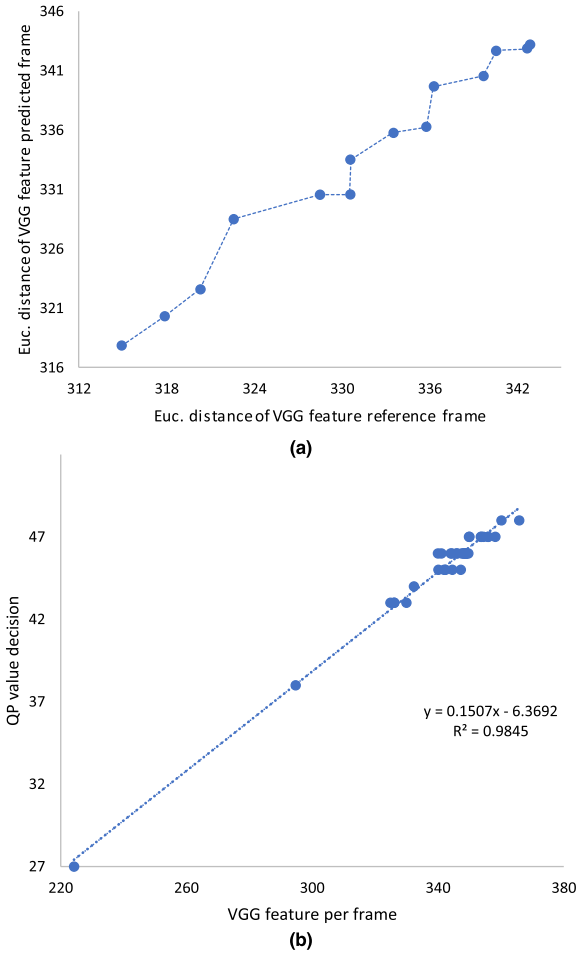

FIGURE 6. Relationship of: (a) VGG feature distortion between reference and predicted frames, and (b) VGG feature and QP selection.

empirically, which also corresponds to the coding structure under the LDP and RA configurations, respectively. For avoiding large fluctuations in quality between neighboring frames, both ΔpQP_{Fid_i} and ΔbQP_{Fid_i} values for different temporal levels should satisfy the conditions described in (9)–(11), where $QP_{OffsetModelScale_i}$ and $QP_{OffsetModelOffset_i}$ are derived as the default settings as in HEVC encoder configurations organized depending on the frame index i. Values of both $QP_{OffsetModelScale_i}$ and $QP_{OffsetModelOffset_i}$ parameters can be found as in Table 3.

$$QP_{perceptual} = \begin{cases} QP_0, \\ \quad if\ I\ frame\ or\ slice,\ POC = 0 \\ QP_{init} - 3\log_2\left(VGG_{feature}\right), \\ \quad if\ I\ frame\ or\ slice,\ POC \neq 0 \\ QP_{init} + \Delta pQP_{Fid_i}, \\ \quad if\ P\ frame\ or\ slice \\ QP_{init} + \Delta bQP_{Fid_i}, \\ \quad if\ B\ frame\ or\ slice \end{cases} \tag{8}$$

$$\Delta pQP_{Offset} = Clip(0.0, 3.0, \Delta QP_{Offset_i}) \tag{9}$$

$$\Delta bQP_{Offset} = \begin{cases} Clip\left(0.0, 3.0, \Delta QP_{Offset_i}\right), \\ \quad if\ Fid = 0 \\ Clip\left(0.0, 3.0, \Delta QP_{Offset_i}\right), \\ \quad if\ Fid = 1 \\ Clip\left(0.0, 6.0, \Delta QP_{Offset_i}\right), \\ \quad if\ Fid = 2 \\ Clip\left(0.0, 7.0, \Delta QP_{Offset_i}\right), \\ \quad if\ Fid = 3 \\ Clip\left(0.0, 9.0, \Delta QP_{Offset_i}\right), \\ \quad if\ Fid = 4 \end{cases} \tag{10}$$

$$\Delta QP_{Offset_i} = QP_{perceptual} \times QP_{OffsetModelScale_i} + QP_{OffsetModelOffset_i} + VGG_{feature} \tag{11}$$

C. PERCEPTUAL ADAPTIVE LAGRANGE MULTIPLIER DETERMINATION WITH QP-λ RELATIONSHIP

For increased bitrate savings while maintaining the visual quality of the proposed adaptive QP decision algorithm, we also aim to determine the Lagrange multiplier by involving the proposed $VGG_{feature}$. Note that the Lagrange multiplier in HM-16.20 is assigned a static QP_{factor} value. Hence, it is essential to provide an adaptive QP_{factor} designed for different picture types and coding structures in HEVC.

1) QP_{factor} DECISION FOR I-FRAMES

First, we searched for the best QP_{factor} of intra coded frames by assigning several constant values of equation (3) through experiments using HM-16.20 under All Intra configurations. 'BasketballPass', 'BQSquare', 'BlowingBubbles', and 'RaceHorses' were used with all the QP settings for the experiment. Fig. 7 depicts the BD-rate based on SSIM performance with the corresponding QP_{factor} values. It shows an approximation of the optimum QP_{factor} for intra frames, which lies in the range of 0.60 to 0.80 with a minimal BD-BR-SSIM gain of approximately −0.2%, while the highest coding gain is approximately −0.5% given by QP_{factor} as 0.65. Accordingly, the QP_{factor} for intra pictures can be

TABLE 2. Initial of ΔpQP_{Fid_i} and ΔbQP_{Fid_i} for different Fid_i.

Fid_i	ΔpQP_{Fid_i}	ΔbQP_{Fid_i}			
		$QP_{init} < 24$	$24 \leq QP_{init} < 29$	$29 \leq QP_{init} < 37$	$QP_{init} < 37$
0	-	+1	+2	+2	+3
1	+5	+1	+2	+2	+3
2	+4	+3	+4	+5	+6
3	+5	+4	+5	+6	+7
4	+1	+6	+7	+8	+9

TABLE 3. Default value of $QP_{OffsetModelScale_i}$ and $QP_{OffsetModelOffset_i}$ for different Fid_i.

Fid_i	LDP Configuration		RA Configuration	
	$QP_{OffsetModelScale_i}$	$QP_{OffsetModelOffset_i}$	$QP_{OffsetModelScale_i}$	$QP_{OffsetModelOffset_i}$
0	-	-	0	0
1	0.2590	-6.5	0.2061	-4.8848
2	0.2590	-6.5	0.2286	-5.7476
3	0.2590	-6.5	0.2333	-5.9000
4	0.0	0.0	0.3	-7.1444

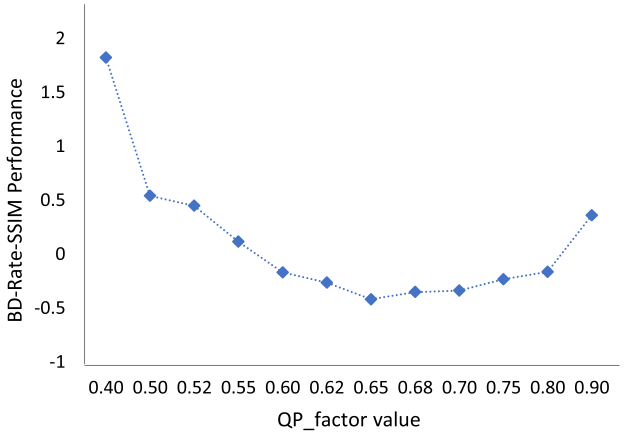

FIGURE 7. QP_{factor} decision and BD-rate-SSIM of intra coded frames.

determined as

$$I_{QP_{factor}} = \begin{cases} 0.57, & POC = 0 \\ \dfrac{StD_{Intra} + VGG_{feature}}{2}, & POC \neq 0 \end{cases} \quad (12)$$

where $I_{QP_{factor}}$ must satisfy $0.57 \leq I_{QP_{factor}} \leq 0.80$, POC denotes the picture order count, and $VGG_{feature}$ is a perceptual loss value from the original and previously intra coded pictures based on the VGG-16 model.

2) QP_{factor} DECISION FOR P-FRAMES

In the Inter picture coding framework under the LDP configuration, the quality of the reconstruction frames is generally influenced by the coding structure factor (or QP_{factor} as previously mentioned). As a result, the distortion of one frame with a certain QP value may affect both the visual quality and RD performance of future frames in encoding order according to the given QP_{factor}. Based on the previous observation illustrated in Fig. 6(a), the VGG feature of a predicted frame D_{VGGpre} increases linearly with the VGG feature of a reference frame D_{VGGref}. Note that the λ values among different frames in the same GOP should be set differently, although they are coded with the same QP value. Hence, deciding the QP_{factor} for different frames in a different temporal layer is desirable, and relationship in (7) can be approximated as

$$P_{QP_{factor}} = D_{VGGpre} \approx c \times D_{VGGref} + D_{ref}^{(I)} \quad (13)$$

where $P_{QP_{factor}}$ stands for the QP_{factor} of P-frame, and c is the linear coefficient, i.e., the slope of the approximated linear distortion relationship between D_{VGGpre} and D_{VGGref}. $D_{ref}^{(I)}$ is added to the linear relationship to represent the feature extraction of the reference frame coded under all intra mode. The $D_{ref}^{(I)}$ value in the proposed algorithm is used to maintain gaps of bit distributions among inter-coded pictures in the same GOP and set as

$$D_{ref}^{(I)} = \frac{StD_{intra}}{(GOP_{size} - Fid_i)} \quad (14)$$

where GOP_{size} and Fid_i denote the GOP size for LDP, which is set to 4 and the frame index listed in the same GOP, respectively. An illustration of how $P_{QP_{factor}}$ is provided for P-frames under the LDP coding structure can be seen in Fig. 8. Then, the combination of (13) and (14) can be expressed as

$$P_{QP_{factor}} = D_{VGGpre} \approx c \times D_{VGGref} + \left(\frac{StD_{intra}}{(GOP_{size} - Fid_i)}\right) \quad (15)$$

Since D_{VGGREF} is the same as $VGG_{feature}$ for the perceptual retention purposes in $P_{QP_{factor}}$, (15) can be further adjusted as in (16), where the parameter c is empirically set as 0.45 in this study.

$$P_{QP_{factor}} = c \times Vgg_{feature} + \left(\frac{StD_{intra}}{(GOP_{size} - Fid_i)}\right) \quad (16)$$

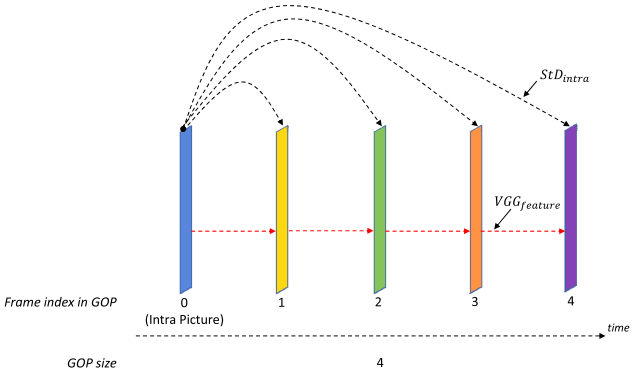

FIGURE 8. Example of the proposed adaptive QP_{factor} for LDP case.

3) QP_{FACTOR} DECISION FOR B-FRAMES

For RA configuration, the QP_{factor} decision uses a similar concept as those in the LDP case with further adjustments. We first analyzed the hierarchical B coding structure under RA configuration in the HEVC depicted in Fig. 9. Both the coding distortion and visual quality of the higher temporal layers are affected by those of the lower temporal levels. For the first frame in a GOP coded as an I-frame, its coding distortion and visual quality will depend only on the spatial operation. However, those pictures coded as B-frames, including the frame with temporal ID = 0 but not an I-frames, need to be treated in Interframe fashion with its corresponding reference frames. Table 4 shows the POC difference between the current POC and its reference pictures to their temporal ID. This algorithm is designed to enable proper feature extraction for the coding frames. However, we used only the reference frame nearest to the current coded picture in the RA coding structure.

As we follow a similar concept in LDP configuration, thus, the formula in (17) for the RA case can be

TABLE 4. Pattern of POC difference between the current POC and its reference POCs.

Tid_i	POC Difference							
0 (Not I_Slice)	16	32						
1	8	-8	16					
2	4	-4	-12	12				
3	2	-2	-6	6	-10	10		
4	1	-1	-3	3	-5	5	-9	9

expressed as

$$B_{QP_{factor}} = c_i \times Vgg_{feature} + \left(\frac{StD_{intra}}{(GOP_{size} - Tid_i)}\right) \quad (17)$$

where $B_{QP_{factor}}$ represents the QP_{factor} for the B-frame, and $VGG_{feature}$ denotes the VGG feature extraction of the reference frames. StD_{intra} is given from the I-frame depending on the intra period of each sequence configuration. GOP_{size} is the GOP size of the RA case, which is set to 16, and Tid_i is the temporal ID of frames in the same GOP. Parameter c_i is a constant value of the $i-$th temporal ID that determines the $B_{QP_{factor}}$ of each frame in different temporal IDs. We first searched the best c per Tid_i empirically with the default QP setting as in HM-16.20. Fig. 10 depicts the results of the BD-BR-SSIM with the selected c values for different temporal IDs. The 'BasketballPass' and 'RaceHorses' test sequences are used for testing all the QP settings. According to Fig. 10, it can be seen that the optimum c values for temporal ID-1 (T_1) is 0.20, and for T_2 to T_4 have the best c values 0.30, 0.40, and 0.42, respectively. In this test, the c_i values increase with the temporal IDs; hence, we set the c values as 0.12 for the Interframe having temporal ID = 0. Accordingly, the c values for different temporal ID in (17) can be

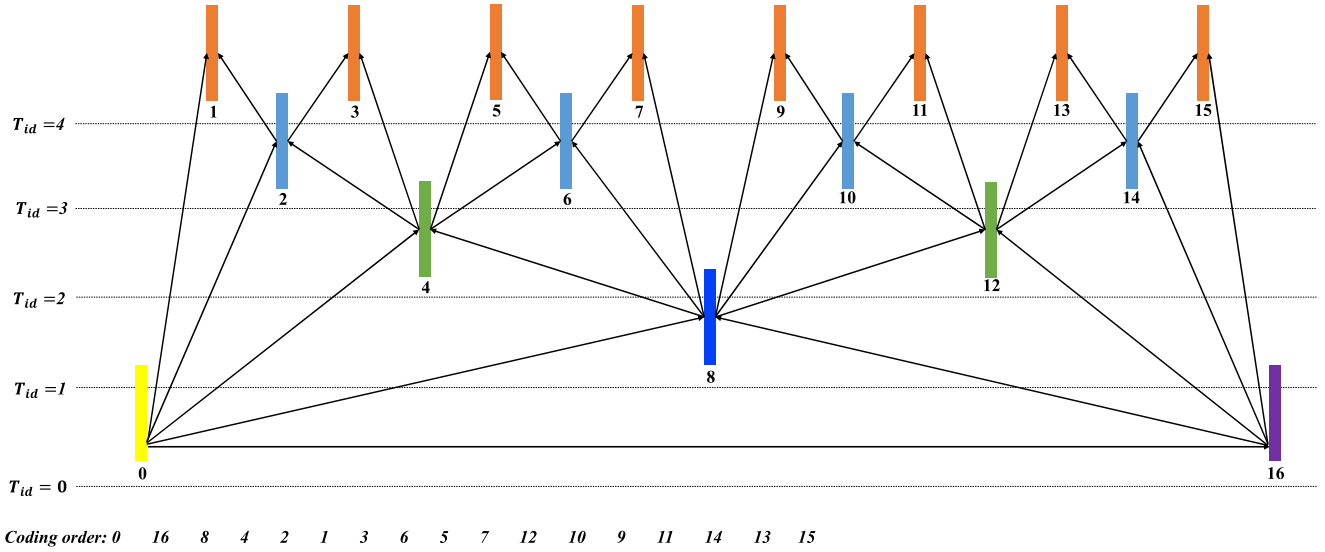

FIGURE 9. Hierarchical B coding structure under RA configuration.

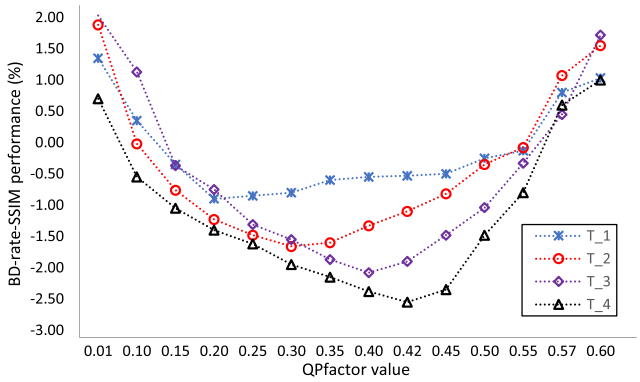

FIGURE 10. *c* parameter decision for each Tid_i under RA configuration.

expressed as:

$$c_i = \begin{cases} 0.12, & if\ Tid = 0\ and\ POC \neq 0 \\ 0.20, & if\ Tid = 1 \\ 0.30, & if\ Tid = 2 \\ 0.40, & if\ Tid = 3 \\ 0.42, & if\ Tid = 4 \end{cases} \tag{18}$$

IV. EXPERIMENTAL RESULTS

The test configuration used for evaluating the proposed algorithm is listed in Table 5. Coding efficiency evaluation was performed under a common test condition for HEVC [32] with the SSIM term [54]. In addition, subjective evaluation was done using the difference mean opinion scores (DMOS). The assessments were conducted by comparing the proposed algorithm against HM-16.20 as an anchor software and also against other existing works [40], [42].

TABLE 5. 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	Low-Delay-P (LDP) and Random-Access (RA)
Rate control	Disabled
DNN framework	TensorFlow-GPU with CUDA 9.0

A. CODING PERFORMANCE EVALUATIONS

We conducted several evaluations of the coding performance to assess the objective quality of the proposed algorithm. All the objective quality measures are tabulated in Table 6. First, we checked the SSIM difference, $\Delta SSIM$ between the proposed algorithm and the anchor. It is defined by

$$\Delta SSIM = SSIM_{PRO} - SSIM_{HM} \tag{19}$$

where $SSIM_{PRO}$ and $SSIM_{HM}$ denote the luma SSIM quality of the proposed algorithm and the anchor, respectively. For (19), a negative value means that the SSIM quality of the proposed algorithm is worse than that of HM-16.20. We also evaluate the bitrate reduction, $\Delta Bitrate$ towards the anchor software, which can be denoted by

$$\Delta Bitrate = \left(\frac{R_{PRO} - R_{HM}}{R_{HM}}\right) \times 100\% \tag{20}$$

where R_{PRO} and R_{HM} represent the output bitrate of the proposed and anchor algorithms, respectively. The proposed algorithm is also evaluated against the anchor in BD-BR with the SSIM metric (BD-BR-SSIM) [54], [55]. For bitrate reduction and BD-BR-SSIM measures, a negative value indicates gains over the anchor. We used HEVC video test sequences with the LDP and RA configurations for several QPs: 22, 27, 32, 37. As shown in Table 6, the proposed algorithm demonstrates a very negligible SSIM degradation of approximately -0.00541 and -0.00656 on average against HM-16.20 without a perceptual adaptive QP method, respectively. In terms of bitrate reduction, the proposed algorithm increases bitrate saving, on average, by approximately -42.67% for LDP and -33.93% for RA configurations over the HM-16.20. For the 'BQTerrace' test sequence, the proposed algorithm achieves the highest bitrate reduction of -66% for the LDP case and -48% for the RA case. Note that the sequence has large flat regions over its frames that benefit the proposed algorithm both spatially and temporally. In terms of the coding efficiency, the proposed algorithm yields better BD-BR-SSIM scores than the anchor about -20.68% and -14.27% for LDP and RA configurations, respectively. The proposed algorithm can also simulate better performance for test sequences with higher resolutions. In the case of LDP, Class B and Class E provide an average coding gain of approximately -21% and -28%, respectively. In the case of RA, Class A also gives a coding gain of approximately -15%.

According to Table 6, the proposed algorithm can achieve better objective performances under the LDP configuration than RA. For the sake of visual quality, the number of intra coded pictures in the LDP case indicates that the proposed algorithm has an essential role in maintaining the quality of the reconstructed frames. Better quality of the reconstructed frames can provide better prediction modes for the future inter coded frames, as well as better visual features for the proposed QP and Lagrange multiplier selections. Considering both spatial and temporal visual features for the proposed algorithm results in significant bitrate reduction while retaining the visual quality of the test videos. For test sequences that have many homogeneous regions, slow motions, and larger background areas than the moving objects in a frame, the proposed algorithm can play a prominent role in obtaining higher objective measures. The visual characteristics of such test sequences can be seen in 'BQTerrace', 'Johnny', 'FourPeople', 'Cactus', 'KristenAndSarra' videos, etc., in which the most significant coding gains are obtained in perceptual terms. On the other hand, the proposed algorithm can contribute only moderate coding improvements for 'Kimono' and 'RaceHorses' that have more textures and fast or more motions.

TABLE 6. Objective quality comparisons between the proposed algorithm and HM-16.20.

Class	Sequence	LDP Configuration			RA Configuration		
		ΔSSIM	ΔBitrate	BD-BR-SSIM	ΔSSIM	ΔBitrate	BD-BR-SSIM
A	PeopleOnStreet				-0.00735	-30.94%	-7.27%
	Traffic				-0.00284	-37.14%	-22.78%
B	Kimono	-0.00792	-36.75%	-8.01%	-0.00941	-30.53%	-5.25%
	ParkScene	-0.00608	-46.41%	-22.49%	-0.00548	-35.78%	-17.51%
	Cactus	-0.00526	-46.61%	-28.56%	-0.00559	-35.03%	-16.50%
	BasketballDrive	-0.00575	-41.04%	-16.28%	-0.00718	-34.84%	-9.68%
	BQTerrace	-0.00369	-65.73%	-28.81%	-0.00298	-47.68%	-17.41%
C	BasketballDrill	-0.00662	-36.57%	-17.26%	-0.00669	-31.82%	-16.90%
	BQMall	-0.00508	-39.89%	-18.40%	-0.00611	-31.12%	-10.68%
	PartyScene	-0.00691	-44.78%	-22.85%	-0.00738	-36.87%	-17.33%
	RaceHorses	-0.01037	-42.64%	-10.77%	-0.01268	-37.13%	-7.74%
D	BasketballPass	-0.00620	-32.67%	-14.96%	-0.00652	-27.69%	-13.62%
	BQSquare	-0.00204	-45.44%	-23.05%	-0.00089	-31.20%	-23.97%
	BlowingBubbles	-0.00526	-37.03%	-20.39%	-0.00535	-30.70%	-14.40%
	RaceHorses	-0.00990	-32.99%	-14.73%	-0.01202	-30.55%	-13.02%
E	FourPeople	-0.00184	-42.81%	-29.90%			
	Johnny	-0.00161	-46.95%	-29.92%			
	KristenAndSara	-0.00202	-44.41%	-24.48%			
	Average of Class A				-0.00510	-34.04%	-15.02%
	Average of Class B	-0.00574	-47.31%	-20.83%	-0.00613	-36.77%	-13.27%
	Average of Class C	-0.00725	-40.97%	-17.32%	-0.00821	-34.23%	-13.16%
	Average of Class D	-0.00585	-37.03%	-18.28%	-0.00620	-33.20%	-12.34%
	Average of Class E	-0.00183	-44.73%	-28.10%			
	Overall	**-0.00541**	**-42.67%**	**-20.68%**	**-0.00656**	**-33.93%**	**-14.27%**

TABLE 7. DMOS comparisons between the proposed algorithm and HM-16.20.

Class	Sequence	LDP Configuration					RA Configuration				
		22	27	32	37	Avg.	22	27	32	37	Avg.
A	PeopleOnStreet						0.02	0.12	0.03	-0.13	0.01
	Traffic						0.01	-0.13	-0.16	-0.27	-0.14
B	Kimono	-0.02	0.01	-0.11	-0.27	-0.10	-0.02	0.00	0.05	-0.22	-0.05
	ParkScene	-0.06	-0.01	-0.06	-0.26	-0.10	0.11	-0.05	0.01	-0.14	-0.02
	Cactus	-0.01	-0.01	-0.23	-0.16	-0.10	-0.06	-0.13	-0.02	-0.08	-0.07
	BasketballDrive	-0.01	-0.06	-0.11	-0.19	-0.10	-0.10	-0.09	0.08	-0.08	-0.05
	BQTerrace	0.11	0.09	-0.02	-0.21	-0.01	-0.01	0.08	0.00	-0.01	0.02
C	BasketballDrill	-0.05	-0.01	0.10	-0.26	-0.06	-0.06	-0.02	0.11	-0.28	-0.06
	BQMall	0.01	0.01	-0.08	-0.04	-0.02	0.04	0.01	-0.04	0.03	0.01
	PartyScene	0.00	0.00	0.08	-0.10	-0.01	0.00	0.04	0.00	-0.13	-0.02
	RaceHorses	0.00	-0.14	-0.08	-0.12	-0.09	0.03	-0.14	-0.04	-0.24	-0.10
D	BasketballPass	-0.07	0.11	0.00	-0.03	0.00	-0.08	0.15	-0.02	-0.08	-0.01
	BQSquare	0.00	0.00	0.07	-0.21	-0.03	0.00	0.00	0.09	-0.05	0.01
E	FourPeople	-0.03	-0.07	-0.14	0.10	-0.03					
	Johnny	0.11	-0.07	-0.01	-0.09	-0.01					
	KristenAndSara	0.01	-0.24	-0.11	-0.13	-0.12					

TABLE 8. Average of DMOS comparisons.

Class	LDP Configuration	RA Configuration
A		-0.07
B	-0.08	-0.03
C	-0.04	-0.04
D	-0.02	0.00
E	-0.05	
Average	**-0.05**	**-0.04**

B. SUBJECTIVE PERFORMANCE EVALUATIONS

Subjective quality assessment was performed to compare the proposed algorithm and HM-16.20 for all the test sequences by following the double stimulus continuous quality scale (DSCQS) method [55]. There are 18 observers among which 11 are in the relative field, and the rest are naïve in image processing. Before the test, we conducted simple demonstrations for the observers to introduce the evaluation process. For each participant, the reconstructed frames from the proposed algorithm and HM-16.20 were randomly shown twice with all the QP values. Then, the observers were asked to provide MOS values in the continuous scale ranging from 1 to 5. Finally, we processed the MOS values to produce the DMOS scores between MOS_{PRO} and MOS_{HM}, which denotes the luma MOS quality of the proposed algorithm and the anchor, respectively. DMOS scores are defined by

$$DMOS = MOS_{pro} - MOS_{HM} \tag{21}$$

Table 7 shows the DMOS of all the test sequences under LDP and RA configurations. For convenience, we introduced the average of DMOS per each sequence for all the QP values to see visual quality judgments of the generated reconstruction frames. Minus values indicate that the video quality of

TABLE 9. BD-rate-SSIM comparisons of the proposed algorithm and other existing algorithms.

Class	Sequence	Xiang [42]	Yeo [40]	Proposed
A	PeopoleOnStreet	-4.32%	4.10%	-10.30%
	Traffic	-6.51%	-7.46%	-13.12%
B	Kimono	-0.56%	-0.46%	-12.91%
	ParkScene	-5.38%	-5.51%	-17.00%
	Cactus	-1.41%	-2.77%	-17.98%
	BasketballDrive	-5.65%	-5.51%	-15.13%
	BQTerrace	-2.07%	-6.64%	-21.64%
C	BasketballDrill	-11.78%	-8.95%	-16.09%
	BQMall	-2.37%	-3.46%	-18.54%
	PartyScene	-4.33%	-2.82%	-19.26%
	RaceHorses	-1.68%	-0.35%	-16.58%
D	BasketballPass	-6.29%	-5.63%	-4.74%
	BQSquare	-2.50%	-0.68%	-16.24%
	BlowingBubbles	-7.05%	-8.32%	-17.15%
	RaceHorses	-2.52%	-3.26%	-7.82%
Average of Class A		-5.42%	-1.68%	-11.71%
Average of Class B		-3.01%	-4.18%	-16.93%
Average of Class C		-5.04%	-3.90%	-17.62%
Average of Class D		-4.59%	-4.47%	-11.49%
Overall		**-4.51%**	**-3.56%**	**-14.44%**

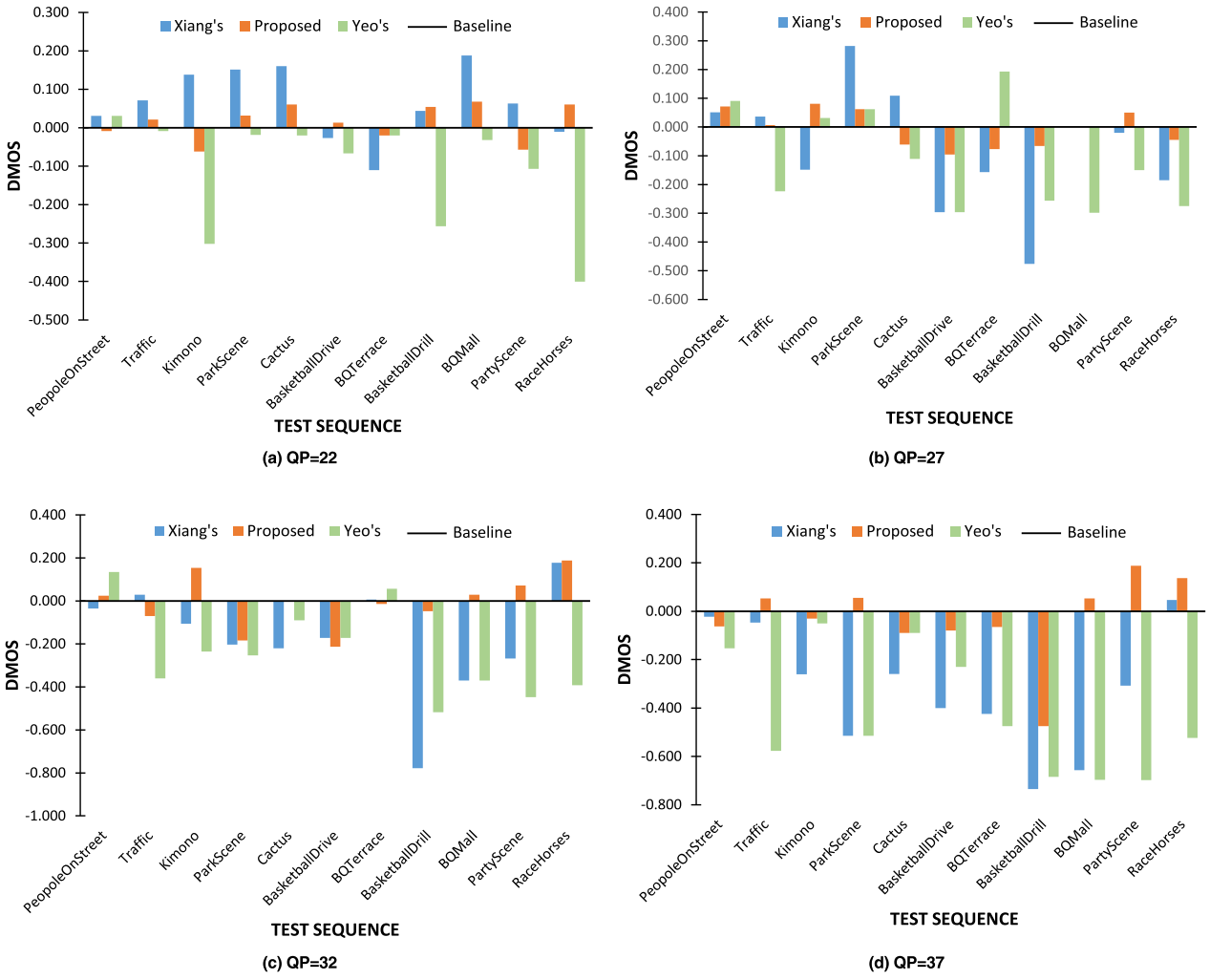

FIGURE 11. DMOS comparisons of Xiang's, Yeo's, and the proposed algorithms.

the proposed algorithm is subjectively worse than that of the anchor ones. As presented, DMOS scales for the entire test sequences are quite close to 0. It means that the proposed algorithm can code nearly visually identical output over those by HM-16.20. For several video sequences, as shown in Table 7, the visual quality of the proposed algorithm is even

slightly better than that of the anchor, such as in 'PeopleOnStreet', 'BQTerrace', 'BQMall', and 'BQSquare', primarily when they are generated under the RA coding structure. This similarity in video quality between the proposed algorithm and HM-16.20 can be seen for all the video sequence classes. We can see that the proposed algorithm degrades visually based on the DMOS test very slightly compared to its anchor, by only about −0.05 and −0.04 for LDP and RA configurations, respectively, as shown in Table 8.

C. COMPARISONS WITH EXISTING ALGORITHMS

After we presented both objective and subjective comparisons between the proposed algorithm and HM-16.20, we can conclude that the perceptual adaptive QP at the frame-level demonstrates its capability to maintain visual quality with better coding efficiency performances in the perceptual term. In this sub-section, we present the same comparisons (objective and subjective comparisons) of the proposed algorithm against other existing algorithms. Table 9 shows the SSIM-based BD-rate comparisons of Yeo *et al.* [40], Xiang *et al.* [42], and the proposed algorithms. As both existing algorithms were integrated into HM-16.0, we also implemented the proposed algorithm in the same software version to meet fair comparisons. As shown in Table 9, we can see that the proposed algorithm in the downgraded version can still outperform two existing algorithms in perceptual coding efficiency. Overall, we can achieve a coding gain of approximately −14.44%, while Xiang's and Yeo's are −4.51% and −3.56%, respectively. Note that all the presented results in Table 9 were generated under random-access configuration with all the quantization parameter values.

Furthermore, we also performed the MOS test to evaluate the subjective visual quality of all the algorithms. Fig. 11 presents the average DMOS results of Xiang's, Yeo's, and the proposed algorithms in the RA structure. The performance of the baseline, which refers to the HM software, is set to zero for the visual similarity evaluation of the three algorithms. DMOS scores that are close to the zero baseline indicate visual similarity to the anchor. From the experimental results, most of the test sequences tested under the proposed algorithm can stand more DMOS points closer to zero, followed by the Xiang's and Yeo's algorithms. This means that the proposed algorithm can give better quality subjectively than the two existing algorithms.

V. CONCLUSION

In this work, we propose a perceptual adaptive QP algorithm at the frame-level to obtain better subjective coding performance for HEVC. The proposed algorithm utilizes a predefined model of the VGG-16 network for feature extractions from the original and previously reconstructed pictures. We designed the proposed algorithm by developing a perceptual loss function based on the extracted features. The proposed algorithm adaptively determines perceptual QP values for different picture types of the hierarchical coding structure in HEVC. Results of approximately −21% and −14% coding gains in SSIM, are yielded by the proposed algorithm, compared with the HM-16.20, for LDP and RA, respectively. The subjective quality evaluation shows that the proposed algorithm can produce comparable visual quality against the anchor with significant bitrate-saving.

REFERENCES

[1] 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.
[2] 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.
[3] C. C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl, "Parallel scalability and efficiency of HEVC parallelization approaches," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1827–1838, Dec. 2012.
[4] H. Jo and D. Sim, "Bitstream decoding processor for fast entropy decoding of variable length coding-based multiformat videos," *Opt. Eng.*, vol. 53, no. 6, Jun. 2014, Art. no. 063102, doi: 10.1117/1.OE.53.6.063102.
[5] Y.-J. Yoon, H. Kim, S.-J. Baek, and S.-J. Ko, "Largest coding unit level rate control algorithm for hierarchical video coding in HEVC," *IEIE Trans. Smart Process. Comput.*, vol. 1, no. 3, pp. 171–181, Dec. 2012.
[6] J. Kim and M. Kim, "Analysis of the JND-suppression effect in quantization perspective for HEVC-based perceptual video coding," *IEIE Trans. Smart Process. Comput.*, vol. 4, no. 1, pp. 22–27, Feb. 2015.
[7] 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.
[8] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, "Reducing complexity of HEVC: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5044–5059, Oct. 2018.
[9] B. Lee and M. Kim, "A CU-level rate and distortion estimation scheme for RDO of hardware-friendly HEVC encoders using low-complexity integer DCTs," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3787–3800, Aug. 2016.
[10] 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.
[11] Q. Hu, X. Zhang, Z. Shi, and Z. Gao, "Neyman-pearson-based early mode decision for HEVC encoding," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 379–391, Mar. 2016.
[12] 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.
[13] Y.-J. Ahn and D. Sim, "Square-type-first inter-CU tree search algorithm for acceleration of HEVC encoder," *J. Real-Time Image Process.*, vol. 12, no. 2, pp. 419–432, Feb. 2015, doi: 10.1007/s11554-015-0487-5.
[14] J. Gu, M. Tang, J. Wen, and Y. Han, "Adaptive intra candidate selection with early depth decision for fast intra prediction in HEVC," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 159–163, Feb. 2018.
[15] K. Yang, Y. Gong, M. Ma, and H. R. Wu, "An efficient rate-distortion optimization method for low-delay configuration in H.265/HEVC based on temporal layer rate and distortion dependence," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 4, pp. 1230–1236, Apr. 2019.
[16] 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.
[17] W. Lee, J. Lee, D. Sim, and S.-J. Oh, "A deep learning based inter-layer reference picture generation method for improving SHVC coding performance," *J. Broadcast Eng.*, vol. 24, no. 3, pp. 401–410, May 2019.
[18] W. Lim and D. Sim, "Determination of optimum quantization parameters in residual quad-tree of HEVC based on perceptual quality," *J. Imag. Sci. Technol.*, vol. 62, no. 2, pp. 205021–205028, Mar. 2018.
[19] V. Barocini, J.-R. Ohm, and G. J. Sullivan, *Report of Results From the Call for Proposals on Video Compression With Capability Beyond HEVC*, document JVET-J1003, Joint Video Experts Team, 2018.
[20] 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.

[21] 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.
[22] 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.
[23] T. Hashimoto and E. Sasaki T. Ikai, *AHG9: Separable Convolutional Neural Network Filter With Squeeze-and-Excitation Block*, document JVET-K0158, Joint Video Experts Team, 2018.
[24] 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.
[25] 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.
[26] 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.
[27] HM. *HEVC Test Model*. [Online]. Available: http://hevc.hhi.fraunhofer.de/svn/svnHEVCSoftware/
[28] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
[29] A. Ortego and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.
[30] B. Li, D. Zhang, H. Li, and J. Xu, *QP Determination By Lambda Value*, document JCTVC-I0426, Joint Collaborative Team on Video Coding, 2012.
[31] B. Li, J. Xu, D. Zhang, and H. Li, "QP refinement according to Lagrange multiplier for high efficiency video coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Beijing, China, May 2013, pp. 447–480.
[32] F. Bossen, *Common HM Test Conditions and Software Reference Configurations*, document JCTVC-L1100, Joint Collaborative Team on Video Coding, 2013.
[33] M. Wang, K. N. Ngan, H. Li, and H. Zeng, "Improved block level adaptive quantization for high efficiency video coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Lisbon, Portugal, May 2015, pp. 509–512.
[34] T. Zhao, Z. Wang, and C. W. Chen, "Adaptive quantization parameter cascading in HEVC hierarchical coding," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2997–3009, Jul. 2016.
[35] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 117–129, Jan. 2016.
[36] J. He, E.-H. Yang, F. Yang, and K. Yang, "Adaptive quantization parameter selection for H.265/HEVC by employing inter-frame dependency," *IEEE Trans. Circuits Syst. for Video Technol.*, vol. 28, no. 12, pp. 3424–3436, Dec. 2018.
[37] T.-D. Chuang, C.-Y. Chen, Y.-L. Chang, Y.-W. Huang, and S. Lei, *AhG Quantization: Sub-LCU Delta QP*, document JCTVC-E051, Joint Collaborative Team on Video Coding, 2011.
[38] *X265/HEVC Reference Software*. [Online]. Available: http://hg.videolan.org/x265
[39] *MPEG-2 Test Model 5, Rate Control and Quantization Control Chapter 10*. [Online]. Available: http://www.mpeg.org/MPEG/MSSG/tm5/Ch10/Ch10.html
[40] C. Yeo, H. L. Tan, and Y. H. Tan, "SSIM-based adaptive quantization in HEVC," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 1690–1694.
[41] L. Prangnell, V. Sanchez, and R. Vanam, "Adaptive quantization by soft thresholding in HEVC," in *Proc. Picture Coding Symp. (PCS)*, Cairns, QLD, Australia, May 2015, pp. 35–39.
[42] G. Xiang, H. Jia, M. Yang, J. Liu, C. Zhu, Y. Li, and X. Xie, "An improved adaptive quantization method based on perceptual CU early splitting for HEVC," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2017, pp. 362–365.
[43] G. Xiang, H. Jia, M. Yang, X. Zhang, X. Huang, J. Liu, and X. Xie, "A perceptually temporal adaptive quantization algorithm for HEVC," *J. Vis. Commun. Image Represent.*, vol. 50, pp. 280–289, Jan. 2018.
[44] K. Rouis, M.-C. Larabi, and J. B. Tahar, "Perceptually adaptive Lagrangian multiplier for HEVC guided rate-distortion optimization," *IEEE Access*, vol. 6, pp. 33589–33603, Jun. 2018, doi: 10.1109/ACCESS.2018.2843384.
[45] 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
[46] S. Ma, X. Zhang, C. Jia, Z. Zhao, S. Wang, and S. Wanga, "Image and video compression with neural networks: A review," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2019.2910119.
[47] H. Choi and I. V. Bajic, "Deep frame prediction for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2019.2924657.
[48] 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.
[49] 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.
[50] 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
[51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[52] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[53] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 105–114.
[54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[55] 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 the 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 Laboratory (IPSL), in 2013. His research interests are related to high-efficiency video compression (HEVC/x265) techniques, fast coding, rate control, and versatile video coding (VVC), and deep learning.

DONGGYU SIM received 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 for 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 main 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.