Recent Advances in Rate Control: From Optimization to Implementation and Beyond

Xuekai Wei, Mingliang Zhou, Heqiang Wang, Haoyan Yang, Lei Chen, and Sam Kwong, Fellow, IEEE

Abstract—Video coding is a video compression technique that compresses the original video sequence to produce a smaller archive file or reduce the transmission bandwidth under constraints on the visual quality loss. Rate control (RC) plays a critical role in video coding. It can achieve stable stream output in practical applications, especially real-time video applications such as video conferencing or game live streaming. Most RC algorithms either directly or indirectly characterise the relationship between the bit rate (R) and quantisation (Q) and then allocate bits to every coding unit so as to guarantee the global bit rate and video quality level. This paper comprehensively reviews the classic RC technologies used in international video standards of past generations, analyses the mathematical models and implementation mechanisms of various schemes, and compares the performance of recent state-of-the-art RC algorithms. Finally, we discuss future directions and new application areas for RC methods. We hope that this review can help support the development, implementation, and application of RC for new video coding standards.

Index Terms—Video coding, rate control, AVC, HEVC, VVC.

I. INTRODUCTION

Video coding is a technology for converting uncompressed digital video signals into standardised and decodable formats that consume less space. As the speed supported by internet infrastructures, including fixed broadband, mobile networking and Wi-Fi, has increased, video traffic has come to dominate global data traffic; video transmission currently represents approximately 82% of all traffic, and this percentage is still rising [1]. However, uncompressed videos occupy a large quantity of bits, and this high data volume greatly limits various video applications if suitable coding or compression is not applied [2]. To address this issue, multiple generations of video coding standards have been successively proposed through collaboration among several international standards-setting organisations, e.g., the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Video Coding Experts Group (VCEG) and the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) Moving Picture Experts Group (MPEG). The most well-known recent standards are the Advanced Video Coding (AVC) standard [3], the High Efficiency Video Coding (HEVC) standard [4] and the Versatile Video Coding (VVC) standard [5]. Throughout the history of video standards, each generation of standards has offered significantly improved rate–distortion performance [6], with the main objective of either reducing the coding bit rate as much as possible while ensuring a certain video quality or reducing the coding distortion as much as possible while maintaining a certain coding bit rate limit. Usually, the original image is divided into multiple square blocks of pixels [7], which are processed in sequence, followed by intra-frame/inter-frame prediction [8], transformation and inverse transformation [9], quantisation and inverse quantisation [10], loop filtering [11], and entropy coding [12] to finally obtain a video stream.

Rate control (RC) is a mechanism for determining how many bits are to be transmitted during the encoding process, which is useful because the available bandwidth for video transmission is usually limited. To support the effective transmission of video data while guaranteeing the playback quality of the video service under the condition that all relevant channel bandwidth and transmission delay constraints are met, an RC mechanism selects a series of encoding parameters so as to simultaneously ensure that the bit rate of the to-be-encoded video satisfies the required rate limit and that the encoding distortion is as low as possible. The coding parameters usually include partition models, prediction models and quantisation parameters (QPs). Since a large number of intra-frame and inter-frame prediction techniques are applied in video compression algorithms, the levels of rate–distortion performance achieved for the numerous coding units are interdependent, and the QP of each coding unit is directly determined in accordance with rate–distortion optimisation (RDO) technology. The complexity of obtaining straightforward closed-form solutions for the coding parameters is extremely high. Therefore, an actual RC scheme is usually divided into two steps. First,
bits are assigned to each basic coding unit so as to achieve the minimum distortion in accordance with the total bit budget; this process is called bit allocation. Second, in accordance with the relationship model between the coding rate and the QP, the QP is independently determined for each coding unit in accordance with its target number of bits. Because an RC module is a necessary component of any video encoder, all video coding standards have their own recommended RC models, such as TM5 [13] of MPEG-2, TMN8 [14] of H.263, JVT-G012 [15] of H.264/AVC, JCTVC-H0213 [16] and JCTVC-K0103 [17] of H.265/HEVC, and JVET-K0390 [18] of H.266/VVC.

From the perspectives of mathematical models and realisation mechanisms, various possible RC schemes have been extensively explored. These schemes have essentially been developed to explore the relationship between rate (R) and distortion (D); that is, different models are based on different R-D curves. Based on the assumption that the encoder can determine the target bit rate by choosing a suitable Q, R-Q-based RC involves estimating the relationship between R and D in the QP domain; hence, these schemes are called Q-domain RC schemes. Many works have followed this idea. Liu et al. [19] proposed an accurate linear R-Q model to characterise the relationship between the total R for both texture and nontexture information and the QP. Hu et al. [20] proposed a frame-level RC algorithm that employs bit information in the RDO process instead of the mean absolute deviation (MAD) of the residuals to predict the complexity of each frame and uses a self-adaptive exponential R-Q model to apply RC. Choi et al. [16] proposed a unified R-Q (URQ) model that can be employed for RC at any level (groups of pictures (GOPs), frames, or basic units) because it captures the relationship between the target rate R and the QP value for a pixel. R-Q-based RC algorithms are widely used in AVC. However, an investigation of JCTVC-I0426 [21] has revealed that the slope \( \lambda \) of the R-D curve is more important than the QP for bit rate determination. Therefore, Li et al. [17] proposed a \( \lambda \)-domain RC method and implemented it in the HEVC standard. Karczewicz et al. [22] proposed an R-\( \lambda \)-based RC scheme for intra frames/slices based on the sum of absolute transformed differences (SATD) rather than the mean square error (MSE). Fang et al. [23] proposed an R-\( \lambda \)-based RC method with a preencoding process. There is also another kind of RC algorithm that builds an association between R and the percentage of zeros among the quantised transform coefficients (\( \rho \)). He et al. [24] estimated an R-D function with low computational complexity in the \( \rho \) domain and proposed an encoder-based rate-shape-smoothing algorithm. Liu et al. [25] proposed a new linear model to obtain the QP at the frame level, for which the model parameters in the \( \rho \) domain can be adaptively estimated from temporal or interlayer information. Additional milestone papers and studies are also represented in Fig. 1. We can observe that R-Q models are applied most frequently in H.264, whereas R-\( \lambda \) models are widely adopted in H.265 and H.266, reflecting the superiority of the latter. Another observation is that RC methods are usually implemented at the frame or block level based on considerations regarding the effectiveness and complexity of the encoding process.

In this paper, we provide a comprehensive review and survey of RC schemes, from the most widely used methods in engineering to the latest research directions in academia. Moreover, a variety of mathematical models of the relationship between R and Q are introduced and analysed to explore their advantages and disadvantages. Different kinds of RC algorithms are also compared in terms of coding efficiency and time consumption. Finally, recent advances in RC research and future research directions are considered.

The remainder of this paper is organised as follows. Section II introduces and analyses RC models from different perspectives. In Section III, the performance comparison of different RC schemes is discussed. Section IV surveys recent and future RC techniques, and Section V concludes the paper.

### II. Brief Overview of Rate Control

Before we begin exploring recent advances in RC, we introduce the RC structures used in HEVC and VVC. In HEVC, the RC process begins with parameter initialisation; then, bit allocation is performed at three main levels, namely, the GOP level, the frame level and the coding tree unit (CTU) level. The parameters are updated when the features of the coding units change. In particular, QP determination is an important step in RC [4]. Compared with HEVC, the skip mode at the CTU level is modified in VVC [5]. More details are shown in Fig. 2 and 3. This section provides a brief introduction to RC models,

---

**Fig. 1.** History of the main proposals for RC methods.
divided into six parts: R-Q models, exponential R-λ models, R-ρ models, deep learning models, scalable video coding and other types of models.

A. R-Q Methods

The RC algorithms for H.264 adopt a variety of techniques, including the adaptive basic unit layer (ABUL) approach, the fluid traffic model (FTM), a linear MAD model, and a quadratic rate–distortion model [68]. A hierarchical bit rate control strategy considering the GOP level, the frame level and the basic unit level is adopted. In the Joint Video Team (JVT)'s proposal, the JVT-G012 bit rate control algorithm is adopted. This algorithm introduces the concept of a basic unit, which is used to divide each frame into several basic units. The basic unit may be a macroblock, a row of macroblocks, a field or a frame. In frame-level bit rate control, the target number of bits per frame is allocated based on the network bandwidth, buffer usage, buffer size, and remaining bits. In basic-unit-level bit rate control, the target bits are averaged based on the remaining target bits for the frame. The quadratic R-Q model adopted in the JVT-G012 algorithm for H.264 is as follows:

$$R = \text{MAD} \times \left( \frac{X_1}{Q_{\text{step}}} + \frac{X_2}{Q_{\text{step}}^2} \right)$$  \hspace{1cm} (1)

where $R$ represents the number of coding bits required by the encoding quantisation coefficient and $Q_{\text{step}}$ denotes the quantisation step size of the basic units. $X_1$ and $X_2$ are the model coefficients. The MAD is predicted through the following linear prediction model:

$$\text{MAD}_{cb} = a_1 \times \text{MAD}_{pb} + a_2$$  \hspace{1cm} (2)

where $\text{MAD}_{cb}$ and $\text{MAD}_{pb}$ denote the MADs for the current basic unit and the corresponding position in the previous frame, respectively, and $a_1$ and $a_2$ are model coefficients, which are updated through linear regression during the processing of the last macroblock of each basic unit. Following the emergence of quadratic models, the authors of [26], [29], and [31] further developed models of this kind to increase their coding efficiency. Lee et al. proposed a scalable RC scheme with a more accurate second-order R-D model. Xu et al. first offered a solution to the chicken-and-egg problem between RC and RDO in H.264 and assigned different numbers of bits to different modes accordingly to avoid the regime of poor behaviour of quadratic R-D models. Yuan et al. presented an adaptive coding feature prediction method using spatiotemporal correlations to improve the accuracy of R-D modelling. The works in [27] and [30] were both developed based on linear models. He et al. presented a linear RC (LRC) algorithm for the JVT encoder combined with a simple scheme for collecting frame-level statistics. Ma et al. proposed a new R-D model using the real quantisation step size and, on this basis, proposed an improved RC scheme for the H.264/AVC encoder. With the development of R-Q models, the works in [20], [28], and [32] took the complexity of the frame content into consideration in R-Q modelling. Jiang et al. mitigated video distortion due to strong motion or scene changes by using statistics from previously encoded frames to more accurately predict frame complexity. Kwon et al. developed an RC algorithm under a constant bit rate constraint for the H.264 baseline profile encoder. Hu et al. first adopted a two-stage RC scheme to decouple RDO and RC, then used bit information to predict the frame complexity in the mode decision process for RDO, and finally proposed an adaptive exponential R-Q model for RC. The works in [33] and [35] used adaptive coding methods. Wang et al. proposed an R-D-optimised RC algorithm with adaptive initialisation for H.264. In contrast to traditional RC, the area-based RC method proposed by Hu et al. can adaptively control the rate in accordance with the content to attain better subjective and objective quality. Tsai et al. proposed determining the QPs of general intra frames and scene-change frames by means of a rate–quantisation step size (QS) model and a scene-change-aware model based on Taylor series [34].

For HEVC, the work in [36] proposed a logarithmic model to describe the relation between distortion and rate. Liang et al. determined that the rate and QP approximately obey a logarithmic relation and proposed a logarithmic R-Q model for HEVC. The works in [39] and [69] mainly considered the low-delay (LD) case. Wu et al. proposed an RC scheme for LD.
video coding considering the temporal prediction structure of HEVC. Si et al. proposed frame-level RC schemes for HEVC designed for LD and random access (RA) coding individually. Hosking et al. presented an enhanced intrasymbol RC method that can produce more accurate predictions and thus reduce the average mismatch rate [40]. Wang et al. proposed a frame-level RC algorithm for HEVC based on the reference picture set (RPS) mechanism, leading to specialisation of the QP determination and RPS mechanisms in HEVC, which considerably improved the coding efficiency [37]. The works in [38], [70], and [71] considered linear models to describe the relation between distortion and rate. Yoon et al. combined a linear rate model with an R-Q model based on the Cauchy distribution. Si et al. modified a linear model to adjust the QP to the q scale. Tian et al. proposed a rate–complexity–QP (RCQ) model for HEVC intra-frame RC, which includes linear distortion quantisation as well as exponential R-Q and linear rate–complexity models.

For VVC, Mao et al. presented an RC model based on transform coefficient modelling and derived corresponding R-Q and D-Q models [41]. Helmrich et al. proposed an RC design based on a two-step R-Q model and derived the two-pass encoding parameters [42]. Representative R-Q models and their corresponding key ideas are summarised in Table I, which is sorted by the year of publication and divided into three parts: classic models used in H.264, classic models used in H.265 and classic models used in H.266. The category column of this table indicates at which layer the RC algorithm is applied.

### B. \( \lambda \)-Domain Methods

The existing research shows that the slope \( \lambda \) of an R-D model can be obtained from the hyperbolic R-D function [45] as follows:

\[
\lambda = -\frac{\partial D}{\partial R} = C K \cdot R^{-K-1} \triangleq \alpha R^{\beta} \tag{3}
\]

where \( \alpha \) and \( \beta \) are parameters related to the video source. Li et al. first proposed this type of model for use in HEVC [78]. With the development of R-\( \lambda \) models, the works in [22], [49] and [51] took the complexity of the frame content into consideration in R-\( \lambda \) modelling. In JCTVC-M0257, Karczewicz et al. used video content complexity in a model at the intra-frame/slice level. Wang et al. considered the complexity at the intra-frame level and designed a gradient-based R-\( \lambda \) (GRL) model. Building on the previous developments, Zhou et al. added video complexity to a model for both intra frames and inter frames. Wen et al. [79] proposed an R-\( \lambda \)-based RC algorithm with preencoding that improves encoding performance by means of two solutions: one uses only \( 16 \times 16 \) coding units (CUs) for preencoding, while the other uses both a hyperbolic function and an exponential function as an improvement to the R-\( \lambda \) model. The work in [55] took the influence of temporal layers into consideration in the

### Table I

| Author          | Category      | Year | Key idea                                                                 |
|-----------------|---------------|------|--------------------------------------------------------------------------|
| Lee et al. [36] | Macroblock    | 2000 | Proposed a scalable RC scheme with a more accurate second-order R-D model |
| He et al. [27]  | Frame level   | 2003 | Proposed an LRC algorithm for the JVT encoder combined with a simple frame-level statistical acquisition scheme |
| Jiang et al. [28] | Frame level   | 2004 | Used the MAD ratio as a measure of global frame coding complexity         |
| Xu et al. [29]  | Frame level   | 2004 | Assigned different bits to different modes to avoid the regime of poor behaviour of quadratic R-D models |
| Li et al. [29]  | Basic unit    | 2004 | Taking a macroblock, slice or frame as the basic unit, used a linear model to solve the chicken-and-egg problem for RC in H.264 |
| Ma et al. [30]  | Frame level and macroblock | 2005 | Using a newly proposed R-D model, developed an RC scheme with tunable complexity |
| Yuan et al. [31] | Frame level   | 2006 | Proposed an adaptive coding feature prediction method using spatiotemporal correlation to improve R-D modelling accuracy |
| Kwon et al. [32] | Macroblock    | 2007 | Proposed an RC scheme for H.264 video coding using an enhanced R-D model |
| Liu et al. [19] | Macroblock    | 2007 | Established an exact linear R-Q model to describe the relationship between the total number of bits of texture and nontexture information and the QP |
| Wang et al. [33] | Macroblock    | 2008 | Proposed a joint R-D optimisation RC algorithm for H.264 |
| Tsai et al. [34] | Intra frame   | 2010 | Proposed the use of Taylor series rate–QS and scene-change-aware models to determine the QPs of general intra frames and scene-change frames |
| Hu et al. [20]  | Frame level   | 2010 | Proposed an RC model based on an adaptive exponential R-Q model |
| Hu et al. [35]  | Region-based  | 2012 | Studied inter-frame information to objectively divide a frame into regions based on the R-D behaviour of the frame |
| Liang et al. [36] | CTU level     | 2013 | Proposed an R-Q model for the approximate logarithmic relationship between the rate and QP in HEVC |
| Wang et al. [37] | Frame level   | 2013 | Proposed a frame-level RC algorithm with the RPS mechanism for HEVC |
| Tian et al. [38] | Frame level   | 2014 | Proposed an RCQ algorithm for HEVC intra-frame RC |
| Wu et al. [39]  | CTU level     | 2016 | For LD video coding, proposed an RC scheme for HEVC |
| Hosking et al. [40] | Frame level | 2016 | Proposed an improved intra RC method to produce more accurate predictions |
| Mao et al. [41] | Frame level   | 2020 | Proposed a dependency factor describing the relationship between a reference frame and a frame to be encoded |
| Helmrich et al. [42] | Frame level | 2021 | Derived two-pass coding parameters based on a new two-step R-Q model |
R-λ model to ensure that pictures in different time-domain layers are given different levels of importance for prediction. The authors of [52] and [57] constructed R-D models on the basis of optimised bit allocation. Li et al. improved the CTU-level λ-domain model based on optimised bit allocation, and Guo et al. similarly improved the frame-level λ-domain model by means of optimised bit allocation. Based on Lagrangian multiplier (LM) theory, Wang et al. [43] considered optimisation from the GOP level to the CTU level, thereby improving the video quality. Tang et al. [80] proposed a generalised rate–distortion–λ (R-D-λ) optimisation solution for HEVC RC.

The works in [46], [54], and [81] addressed optimised RC in accordance with the video content. Sanchez et al. proposed a context-based model that incorporates predictive coding technology and uses a piecewise linear function to approximate the R-D curve. Meddeb et al. developed an advanced algorithm for video coding that separates the video content into regions of interest and regions of noninterest and increases the bit allocation for regions of interest while maintaining the overall bit rate. Li et al. optimised bit allocation for multiview texture videos based on interview dependency and spatiotemporal correlation. The works in [56], [59], and [60] mainly considered high-dynamic-range (HDR) video. Mir et al. enhanced the λ-QP relation for HDR video coding to solve the problem of compression performance degradation caused by different coding standards. Perez-Daniel et al. developed a λ-domain model into a multi-R-λ model that considers the wider range of luma values found in HDR video. Zhou et al. considered the issue of visual differences in HDR video and built an R-D model based on this issue. The works in [50], [61], and [63] considered perceptual RC methods. Zeng et al. incorporated human visual acuity into the video coding process by separating the input video into perceptually sensitive areas and less perceptually sensitive areas and increasing the bit allocation for perceptually sensitive areas. Lim et al. proposed a perceptual luminance-adaptive single-loop encoding method. Zhou et al. adopted a just-noticeable distortion (JND) factor to build a λ-domain model that can well describe the distortion of the perceptual field to be used for bit allocation. Li et al. developed an advanced λ-domain model by using an intra-CTU RC scheme [58] that considers the influence of the drift from earlier CTUs to subsequent CTUs in RC. Lee et al. incorporated texture and nontexture factors into frame-level RC to build a λ-domain model [44]. The works in [47], [53], [62], and [82] considered the LD configuration. Wang et al. designed an RC model for the LD case that can adapt to rapid variations in coding efficiency. Chen et al. introduced the variance into the R-D model with the aim of minimising the distortion of the CTUs. Guo et al. presented a λ-domain frame-level RC scheme. Yang et al. adjusted the R-D model to correct for buffer overflow and underflow issues at low latency. Existing RC methods typically optimise the MSE between the distorted image \( Z_i \) and the original image \( Z_j \), i.e., the following cost function:

\[
\frac{1}{P} \sum_{p=1}^{P} (Z_j(p) - a Z_i(p) - b)^2
\]  

(4)

In this function, the optimal values of \( a \) and \( b \) are computed as:

\[
\begin{align*}
\alpha &= \frac{\text{cov}_{Z_i,Z_j}}{\sigma_{Z_i}^2} \\
b^* &= \mu_{Z_j} - \alpha \mu_{Z_i}
\end{align*}
\]

(5)

where \( \text{cov}_{Z_i,Z_j} \) is given by \( \frac{1}{P} \sum_{p=1}^{P} (Z_i(p) - \mu_{Z_i}) (Z_j(p) - \mu_{Z_j}) \). Zhou et al. adjusted the R-D model by replacing the MSE-based distortion evaluation in the R-D model with a distortion based on the structural similarity index measure (SSIM), which is related to visual quality [48].

For VVC, the works in [18], [64], and [67] introduced some adjustments to λ-domain algorithms. Li et al. proposed a three-part scheme: splitting skip and nonskip areas at the picture level, changing the update strategy, and modifying the GOP size to 16. Liu et al. proposed the use of an adaptive λ ratio estimation algorithm. Hyun et al. adjusted the R-λ model in VVC to address textured and nontextured regions simultaneously. Liu et al. researched the relationship between distortion and λ to achieve a balance among bit rate, distortion and video quality. The works in [83], [84], and [85] made use of a quality dependency factor (QDF) to improve the coding efficiency. Liu et al. used QDF-based bit allocation to improve the coding efficiency. Liu et al. also proposed an extension of RC to achieve the configuration in JVET-M0600. Ren et al. proposed an extension of the QDF to a low frame rate. The works in [65] and [66] introduced model modifications in accordance with the coding structure in VVC. Chen et al. presented an R-λ relationship using a quadratic R-D model for VVC, and Li et al. presented an RC model based on the influence of skip blocks to adjust the parameter update strategy. Representative R-λ models and their corresponding key ideas are summarised in table II, which is sorted by the year of publication and divided into three parts: classic models used in H.264, classic models used in H.265 and classic models used in H.266. The category column of this table indicates at which layer the RC algorithm is applied. Notably, for RC strategies, the selection of a suitable R-D model is crucial. Packetwise exponential and hyperbolic models are the two most commonly used models to describe the relationship between bit rate and distortion. Studies on HEVC RC have proven the hyperbolic model to be the better option for that standard. However, the R-D relationship has evolved with the advent of the VVC standard and the integration of new coding tools, necessitating a new R-D model to determine the optimal coding QPs.

C. Models in the ρ Domain

The percentage of zero coefficients, ρ, increases monotonically with the QP when the distribution of the transform coefficients is known, meaning that there is a one-to-one correspondence between ρ and the QP. Hence, the relationship between R and the QP can be established through ρ. Shin et al. modelled the rate–ρ and QP–ρ relationships and adopted a linear approximation scheme to model the rate–ρ relationship [74]. Experiments have shown that the ρ of the quantised transform coefficients has a good linear relation with
the bit rate $R$ \cite{72,73}:

$$R(\rho) = \theta(1 - \rho)$$  \hspace{1cm} (6)$$

where $\theta$ is a constant and $\rho$ is a model parameter that is related to the video content. The authors of \cite{75} argued that a Laplacian distribution is not sufficiently precise to capture the true distribution arising from a quadtree prediction structure. Therefore, a mixed Laplacian distribution was applied to describe this distribution. The authors of \cite{77} proposed an RC algorithm of decreased complexity in the $\rho$ domain, which predicts the encoding parameters for each collocated CTU. Wang et al. proposed a more accurate mixed Laplacian distribution to capture the transform coefficients in HEVC \cite{76}. Representative $\rho$-domain models and their corresponding key ideas are summarised in table III, which is sorted by the year of publication. The category column of this table indicates at which layer the RC algorithm is applied.

### D. Classic Methods in Deep Learning

Deep learning is an effective approach for solving decision-making problems, and thus, it has recently attracted great interest in the video coding community. Hu et al. \cite{87} adjusted the QP by balancing the relationship between texture complexity and coding rate to achieve lower distortion at the CTU level. Gao et al. \cite{86} extracted features from previous frames and built a more accurate R-D model using machine learning. Cooperative game theory has also been introduced into the bit allocation process to increase coding efficiency and quality. Zhou et al. \cite{88} adjusted the QP via a deep neural network when processing dynamic video sequences to reduce distortion and bit rate fluctuations. Wei et al. \cite{89} integrated reinforcement learning and game theory into tile-level bit allocation for 360-degree streaming to increase the quality and coding efficiency.

For VVC, Li et al. presented a convolutional neural network (CNN)-based R-$\lambda$ RC approach for intra-frame coding \cite{97} that reuses the $\lambda$-domain model used for inter-frame RC in the VVC Test Model (VTM) and trained a CNN to simultaneously predict the two model parameters, alpha and beta. Using a multilayer perceptron (MLP) neural network \cite{90}, Raufmehr et al. presented a video bit rate controller that completely conforms to the constraints of real-time applications. Farhad et al. \cite{91}
presented a nonlinear relationship by using a neural network to balance the relationship among the bit rate, buffer size and QP.

Wang et al. designed an RC algorithm [92] that extracts four highly descriptive features to capture the relationship between the video content and the R-D model. Representative deep learning models and their corresponding key ideas are summarised in table IV, which is sorted by the year of publication and divided into two parts: classic models used in H.265 and classic models used in H.266. The category column of this table indicates at which layer the RC algorithm is applied.

With the development of deep learning, Lu et al. presented the first end-to-end deep video compression model [98] that jointly optimises all components for video compression.

E. Classic Methods in Scalable Video Coding

The aim of scalable video coding (SVC) is to allow partial streams to be obtained on the decoding side while encoding the video signal only once. Three types of scalability, namely, temporal scalability, spatial scalability and quality scalability, are desired to meet different application requirements in terms of rate or resolution. Hu et al. introduced a frame-level RC method based on a linear R-Q model and a linear D-Q model [96]; the formulation can be expressed as follows:

$$Q_i^i = \frac{k_i m_i}{\lambda_i, i = 0, \ldots, N}$$

(7)

where $m_i$ is the predicted MAD of the remaining frames at level $i$ and $k_i$, $\lambda_i$, and $\gamma_i$ are parameters. Xu et al. introduced an effective RC method for a scalable video model (SVM) that inherits features of the sophisticated hybrid RC schemes of JVT [93]. Liu et al. proposed a switched model for predicting the MAD of the residual texture using the MAD information available from the previous frame in the same layer [94]. Pitrey et al. presented a new RC scheme for SVC based on a simple yet attractive bit rate modelling framework in the $\rho$ domain [95]. Representative SVC models and their corresponding key ideas are summarised in table V, which is sorted by the year of publication and divided into two parts: classic models used in H.264 and classic models used in H.265. The category column of this table indicates at which layer the RC algorithm is applied.

F. Other Models

1) Segmented R-Q Model: Based on the understanding that the transformation coefficients obey a Laplace distribution with $\sigma^2$, the following segmented model can be obtained [102]:

$$R(Q) = \begin{cases} \frac{1}{2} \log_2 \left(2e^2 \cdot \frac{\sigma^2}{Q^2}\right), & \sigma^2 > \frac{1}{2e} \\ \frac{e^2}{\ln 2} \cdot \frac{\sigma^2}{Q^2}, & \sigma^2 \leq \frac{1}{2e} \end{cases}$$

(8)

where $\sigma^2/Q^2 > 1/2e$ corresponds to the high-bit-rate situation and $\sigma^2/Q^2 \leq 1/2e$ corresponds to the low-bit-rate situation.
TABLE V
REPRESENTATIVE METHODS BASED ON SVC MODELS

| Author | Category | Year | Key idea |
|--------|----------|------|----------|
| Xu et al. [93] | Spatial layer | 2005 | Proposed performing BDO for temporal subband image coding only on low-pass subband images while applying RC to each spatial layer individually |
| Liu et al. [94] | Base layer | 2008 | Predicted the MAD of the residual texture using the available MAD information from the previous frame in the same layer and the base layer for the same frame |
| Pitera et al. [95] | Inter layer | 2009 | Proposed a simple and attractive bit rate modelling framework in the $\rho$ domain |
| Hu et al. [96] | Frame level | 2011 | Proposed a frame-level RC algorithm based on a linear R-Q model and a linear D-Q model |

2) D-Q Model: Seo et al. proposed a D-Q model to determine the target distortion for QP generation, as follows [103]:

$$D = \alpha \sum_{i=0}^{N_{\text{Depth}}} w_i \left( 1 - P_{l,s} \right)$$

$$\times \left( \frac{2Q}{\lambda_{i,NS}^2 + \lambda_{i,NS}^2 e^{\frac{1}{2} \lambda_{i,NS} Q - e^{\frac{1}{2} \lambda_{i,NS} Q}}} \right) + P_{l,s} \frac{2}{\lambda_{i,NS}}$$

(9)

where $D$ is the MSE for a frame, $\alpha$ is a model parameter that compensates for the difference between the actual and estimated distortions, $P_{l,s}$ is the proportion of CUs at depth $i$ for which the skip mode is adopted, $Q$ is the $q$ step size, $N_{\text{Depth}}$ is the maximum CU depth, $\omega_i$ is a weighting factor, and $\lambda_i$ denotes the model parameter for the $i$-th CU depth.

3) Adaptive R-Q Model: The method proposed by Jing et al. aims to select accurate QPs for intra-coded frames in accordance with a target $R$. The parameters of the model can be adaptively adjusted by considering a gradient-based frame complexity measure [99].

4) Two-Pass Methods: The general idea of two-pass RC is to further optimise the QP of each frame in the second encoding pass in accordance with scene complexity statistics computed in the first pass. For AVC, Lie et al. [104] proposed performing frame-level rate allocation in the second pass using content-aware models constructed in the first pass. For HEVC, Wang et al. [105] proposed a SSIM-inspired two-pass RC scheme. The algorithm proposed by Ma et al. [30] consists of a one-pass process and a partial two-pass process at the frame and block levels. Zupancic et al. [106] further proposed a two-pass RC method targeting quality improvement for ultrahigh-definition television (UHDTV) delivery. For VVC, Helmrich et al. proposed an RC design based on a two-step R-Q model and derived the two-pass encoding parameters [42].

5) Visual Attention Models: Liu et al. utilised the mechanism of human visual attention to guide the RC process by incorporating an attention model of motion. They calculated multilayer saliency maps of motion, which were used to adjust the frame-level bit allocation, resulting in quality improvement [100]. Shen et al. proposed an innovative $R$ method that considers human visual attention, in which the stronger the local motion attention is, the greater the $R$ that is assigned to the frame [101].

6) Multithreaded Coding Methods: At the frame level, multicores with a small number of cores can take advantage of multithreaded coding capabilities, e.g., parallelisation, especially when there is little or no dependency between images, as in the case of images in the same temporal layer or even intra frames [107]. In such cases, parallelisation is simple to implement and incurs relatively small coding efficiency losses. However, the gain that can be achieved through frame-level parallelisation is limited by the GOP size, and such processing increases latency despite improving the processing frame rate [108]. In addition to frame-level parallelisation, slice-level parallelisation can be another way to improve performance [109]. Slices partitioned within a picture are independent of each other [110], [111], apart from potential in-loop filtering dependencies that may exist at the slice boundaries. Slices need not be associated with each other during the execution of most processes performed during coding, such as prediction and transformation, which are applied across slices, and slice-level parallelisation can enable a dramatic enhancement in coding efficiency due to motion prediction processing [112]. Nevertheless, among the RC methods discussed in this article, almost none use the slice-level RC approach; therefore, we do not present experimental statistics on it. A more fine-grained, block-level parallelisation technique is most widely applied [113], [114], [115]. However, this approach is more difficult to implement because block-level parallelisation requires a more elaborate scheduling algorithm to ensure the correct ordering of the macroblocks due to their multiple spatial dependencies. Furthermore, wavefront parallel processing (WPP) is a commonly used intra-frame parallelisation method, although it has the drawbacks of limitations on the parallelism that can be achieved and unbalanced computational complexity for RC optimisation. To overcome these drawbacks, Joose et al. [116] proposed two real-time RC algorithms with parallelisation techniques, a $\lambda$-domain (LD) algorithm and an $R$-$\lambda$ model (R-LM) algorithm. An adaptive intra-frame parallelisation method was also proposed in [117] that guarantees higher intra-frame parallelism and more accurate control of parallelisation. A hardware architecture strategy was additionally explored in [118] to improve the parallel acceleration of HEVC hardware.

Representative models discussed in this subsection and their corresponding key ideas are summarised in table VI, which is sorted by the year of publication. The category column of this table indicates at which layer the RC algorithm is applied.
III. PERFORMANCE COMPARISON OF DIFFERENT RC SCHEMES

The bases for measuring the advantages and disadvantages of an RC algorithm do not solely concern how many bits are saved and how much the visual quality is improved; coding/decoding complexity is also an essential consideration for some real-time video applications. Modern encoders, such as AVC, HEVC and VVC, are designed following a framework of square-block-based hybrid coding, which provides the opportunity for performance gains on machines with multithreading capabilities and even multicore processors [124], [125], [126]. Usually, the performance gain is calculated as the execution time of the improved algorithm divided by the execution time of the original algorithm. Many related techniques have been proposed in recent video coding standards, some of which are mentioned in the following. All comparisons were performed under fair and well-controlled test conditions. In the evaluation, the general sequences from each class defined in the Common Test Conditions (CTC) [127] were chosen, with possible coding QPs of [22], [27], [32], and [37]. The total number of test sequences in each class is as follows: Class A1 contains three sequences, Class A2 contains three sequences, Class B contains five sequences, Class C contains four sequences, Class D contains four sequences, and Class E contains three sequences [127]. The resolution of the sequences in each class is as follows: Class A1 has a resolution of 3840 × 2160, Class A2 has a resolution of 3840 × 2160, Class B has a resolution of 1920 × 1080, Class C has a resolution of 832 × 480, Class D has a resolution of 416 × 240, and Class E has a resolution of 1280 × 720 [127]. The total number of frames per sequence is taken to be 300, and for sequences that are fewer than 300 frames, the results are normalised accordingly with respect to the actual number of frames. The resolution of the video sequences varies from the Common Intermediate Format (CIF) to 4K, and the bit depth is 8 bits. The target bit rate was set to the actual bit rate of the original algorithm. Many related techniques have been proposed in recent video coding standards, some of which are mentioned in the following. All comparisons were performed under fair and well-controlled test conditions. In the evaluation, the general sequences from each class defined in the Common Test Conditions (CTC) [127] were chosen, with possible coding QPs of [22], [27], [32], and [37].

The performance comparison of different RC schemes is presented in Table VI. The table includes the authors, the category, the year, and the key idea of each method. The key idea column contains representative methods based on other models.

| Author          | Category   | Year | Key idea                                                                 |
|-----------------|------------|------|--------------------------------------------------------------------------|
| Jing et al. [99]| Intra frame| 2008 | Proposed a method for selecting accurate QPs for intra-coded frames based on a target bit rate; by considering a gradient-based frame complexity measure, the model parameters can be adaptively updated |
| Lin et al. [100]| 4x4 block level | 2009 | Focusing on RC at the frame and macroblock levels, proposed adjusting the number of bits allocated to each frame and each macroblock in accordance with the motion saliency |
| Shen et al. [101]| Macroblock | 2013 | Proposed allocating more bits to visually important macroblocks at the frame level and conversely allocating fewer bits to nonimportant macroblocks |

A. Implementations at the Frame Level

In the vast majority of cases, the working environment of the encoder is the CPU; accordingly, many optimisations have been analysed and implemented in modern CPUs. Thus, unless otherwise mentioned, the following experiments were carried out on the CPU. More detailed quantitative comparisons are shown in table VII. When using a CPU, the key to improving performance lies in reducing data dependencies and operational dependencies to improve parallelisation, and at the frame level, these requirements are naturally satisfied. To build an accurate R-Q model, Lee et al. [44] separately constructed three Laplacian probability models for low-textured, medium-textured, and high-textured CUs; the BD-rate of this method is -3.11%, and the BD-PSNR is 0.11 dB (LDP). Wang et al. [75] considered new coding tools for use in HEVC and adopted a hierarchical RC architecture to maintain the video quality of keyframes. In comparisons considering all three configurations, the BD-rate ranges from -3.13% to -6.00%, with corresponding BD-PSNR values ranging from 0.11 dB to 0.23 dB. These authors also proposed D-Q and R-Q models for finding the interframe dependency between to-be-coded frames and the reference frame. Accordingly, a mixed Laplacian distribution ρ-domain-based rate–GOP model was proposed. Wang et al. [37] proposed an efficient hierarchical bit allocation scheme based on a new mechanism for HEVC, i.e., the RPS mechanism, which achieves the highest BD-rate among the compared methods. Moreover, they proposed an innovative header bit ratio prediction method to improve RC accuracy and used a quadratic R-Q model to calculate the QP. Xu et al. [119]
focused on video sequences depicting discontinuous scenes and proposed a novel bit allocation algorithm by building the correlation between the intensity of a scene change and the bit allocation. A BD-rate of -1.5% can be achieved in this way, and the BD-PSNR is 0.14 dB. Song et al. [120] addressed the issue that the RC method used in AVC is no longer suitable for HEVC due to the differences in the GOP coding structures; accordingly, they proposed a new GOP-level bit allocation method that can achieve more accurate RC and lower bit fluctuation, resulting in slightly better R-D performance, as shown in table VII. Guo et al. [57] considered the temporal R-D dependency in GOP-level bit allocation; on this basis, an advanced frame-level R-D model that can more completely use the information of the coded frames was introduced to further enhance R-D performance. Thus, an equation for bit allocation at the frame level was proposed for the optimal Lagrange multiplier approach, which can be solved by employing a recursive Taylor expansion (RTE) scheme. Two comparative experiments were performed using [45] and [121] as benchmarks, referred to as fixed-weight bit allocation (FWA) and adaptive-weight bit allocation (AWA), respectively. The comparison results show that the BD-rate ranges from -4.30% to -4.70% for FWA and -2.90% to -3.20% for AWA. As an alternative to traditional methods that regard the entire RC process as deterministic, some methods treat the variables and parameters of RC as random variables to re-examine the RC process. Hyun et al. [64] observed that the inaccuracy of existing linear rate estimation models causes a decline in RC performance. Thus, they adopted a method called recursive Bayesian estimation (RBE) to precisely estimate rates. Performance comparisons with all three methods show that the BD-rate ranges from -0.20% to -0.30%, with slight improvements in the BD-PSNR. Raufmehr et al. [90] proposed a video bit rate controller to meet the demands of real-time applications. Due to the elimination of traditional models and the utilized network structure, the model complexity is lower, resulting in a large latency reduction. However, these frame-level methods also have some drawbacks, such as poor scalability or high memory occupation. Fortunately, block-based implementations can address these problems [130].

### B. Implementations at the Block Level

RC evaluation at the block level was carried out using the JM reference software for H.264 and the HM reference software for H.265. In table VIII, the RC methods proposed by Kwon et al. [32], Liu et al. [19], Wang et al. [33] and Liu et al. [100] are all compared and analysed. In the LDP mode of [32], the BD-PSNR is improved by 0.35 dB over the baseline. In the scheme of [19], the BD-rate is -5.10%, and the BD-PSNR is 0.33 dB. It can be seen that the algorithm in [33] can achieve better PSNR results than JVT-G012 [15], while the bit rate performance of the two algorithms is similar. In the LDP mode of [100], the BD-rate is -1.02%, the BD-PSNR is 0.07 dB, and the latency reduction is 2.80% on average. [52] proposed a new scheme named the optimal bit allocation (OBA) scheme and presented a detailed comparison based on [45]; the BD-rate of this method is -5.10%, and the BD-PSNR is 0.15 dB. Compared with the method of HM 16.20, the average BD-rate of [62] can reach -3.30%, and the BD-PSNR can reach 0.10 dB. For the method of [60], the BD-rate and BD-PSNR for LDB/LDP/RA are -4.60%/-2.00% and 5.40 dB/1.00 dB/1.90 dB, respectively, which are much higher than those of the baseline HM 16.20 coding algorithms. A recently proposed

### Table VII: Performance Comparison of Frame-Level RC Methods

| Method       | Anchor/Configuration | BD-rate (%) | BD-PSNR (dB) | RC accuracy (%) | Latency reduction (%) |
|--------------|----------------------|-------------|--------------|------------------|-----------------------|
| Lee 2013     | HM 16.20, LDP        | -3.11       | 0.11         | 0.99             | 0.98                  |
| Wang 2013    | HM 16.20, LDB/LDP/RA | -3.13       | 0.11         | 0.99             | 0.84                  |
| Wang 2013    | HM 16.20, LDP/HE     | -2.91       | 0.13         | 0.99             | 0.84                  |
| Xu 2015      | HM 16.20, LDP        | -1.5        | 0.14         | 0.99             | 1.78                  |
| Song 2017    | HM 16.20, RA/LDP     | -0.20       | 0.02         | 0.99             | 0.98                  |
| Guo 2018     | FWA [45], LDP/HE    | -4.70       | 0.21         | 0.99             | 3.22                  |
|             | LDP [121], LDP/HE   | -2.90       | 0.13         | 0.99             | 2.97                  |
| Hyun 2020    | HM 16.20, AL/LDP/RA | -0.30       | 0.02         | 0.99             | 0.22                  |
| Raufmehr 2021 | VTM 17.0, LDP       | -3.05       | 0.11         | 0.99             | 36.61                 |

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
TABLE VIII

| Method    | Anchor | Configuration | BD-rate (%) | BD-PSNR (dB) | RC accuracy (%) | Latency reduction (%) |
|-----------|--------|---------------|-------------|--------------|------------------|-----------------------|
| Kwon 2007 [32] | JM 9.4 | Li [65]/LDP | -5.40        | 0.35         | 97.29            | 2.50                  |
| Liu 2007 [19]  | JM 9.4 | JVT-G012 [15]/LDP | -5.10        | 0.33         | 90.22            | 3.07                  |
| Wang 2008 [33]  | JM 9.4 | JVT-G012 [15]/LDB | -0.63        | 0.41         | 93.47            | 2.10                  |
| Liu 2009 [100]   | JM 9.4 | JVT-G012 [15]/LDP | -0.12        | 0.07         | 95.33            | 2.80                  |
| Li 2016 [52]    | HM 16.20 | Li [45] | -5.10        | 0.15         | 93.57            | 0.50                  |
| Chen 2019 [62]   | HM 16.20 | LDP | -3.30        | 0.10         | 97.27            | 0.10                  |
| Zhou 2019 [60]   | HM 16.20 | LDB | -4.60        | 5.40         | 97.56            | 2.90                  |
|                |        | LDP | -1.20        | 1.00         | 97.38            | 1.84                  |
|                |        | RA  | -2.00        | 1.90         | 99.87            | 2.50                  |
| Zhou 2020 [63]   | HM 16.20 | LDB | -1.35        | 0.01         | 97.08            | 0.01                  |
|                |        | LDP | -3.30        | 0.11         | 94.14            | 0.07                  |
|                |        | RA  | -2.75        | 0.09         | 94.89            | 0.77                  |

TABLE IX

| Method    | Anchor | Configuration | BD-rate (%) | BD-PSNR (dB) | RC accuracy (%) | Latency reduction (%) |
|-----------|--------|---------------|-------------|--------------|------------------|-----------------------|
| Xu 2005 [93] | SVM3.0 [122] | Default | -7.20        | 1.50         | 99.50            | -3.02                 |
| Ma 2005 [30] | TMS | Default | -2.30        | 0.20         | 96.50            | -2.22                 |
|           | AVC-TM [123] | Default | -2.20        | 0.20         | 96.70            | -2.21                 |
| Wang 2009 [43] | JVT-G012 [15] | LDP | -4.20        | 0.98         | 95.49            | -0.45                 |
| Li 2014 [45] | HM 16.20 | LD(noH)/LD/H/RA | -3.10/-5.50/8.90 | 0.29/0.55/1.08 | 99.94/99/99/99 | -3.10/0.10/3.20 |
| Yang 2014 [47] | HM 16.20 | LDB | -2.20        | 0.31         | 99.92            | 1.30                  |
| Zhou 2020 [88] | Hu et al. [57] | HM 16.20/AI | -5.00        | 0.50         | 99.95            | -114.00 (train)/-1.80 (no train) |
| Gao et al. [86] | HM 16.20/LDB | -5.20        | 0.50         | 99.86            | -1048.20 (train)/-1.20 (no train) |
| Sanchez 2018 [54] | HM 16.20 | Default | -1.00        | 0.01         | 99.93            | -1048.20 (train)/-1.20 (no train) |

JND-based [131], [132] perceptual RC method [63] is also included in the comparison, and it can be seen that this method effectively reduces the bit rate without compromising the encoded video quality, as confirmed by objective metrics such as the PSNR. In addition, in terms of coding control, the actual $R$ after encoding is closer to the target $R$. As seen from these comparisons, different methods are suitable for different situations. The OBA scheme is appropriate for achieving a lower bit rate with good quality. The method of [60] is suitable for improving the BD-rate and BD-PSNR for LDB/LDP/RA, while the JND-based perceptual RC method is appropriate for reducing the bit rate without compromising the encoded video quality. Other methods also have their own advantages in achieving a higher BD-PSNR or lower BD-rate.

C. Implementations Addressing Joint Layer Optimisation

Often, multiple layers can be jointly optimised to improve the coding efficiency. Recall that the RC process can be divided into two main parts: one is bit allocation, while the other is determining how to realise the target bit allocation for a CU through the R-D model. As shown in table IX, the algorithm proposed by Xu et al. [93] is simultaneously implemented at the GOP, frame and basic unit levels, and experiments in SVM3.0 [122] show that the mismatch between the target $R$ and real $R$ does not exceed 0.5% and that the BD-PSNR on average is approximately 1.50 dB. The algorithm proposed by Ma et al. [30] consists of a one-pass process and a partial two-pass process at the frame and block levels, and experiments show that compared with TM5 and AVC-TM [123], the BD-PSNRs are 0.20 dB and 0.20 dB, respectively. The algorithm proposed by Wang et al. [43] is a joint three-layer (JTL) model implemented in JM 9.4 [133]. Compared with JVT-G012 [15], the average BD-PSNR is 0.98 dB, while the average latency reduction is $-0.45\%$. The algorithm proposed by Li et al. [45] is implemented at the frame and CTU levels, and testing in HM 16.20 shows that relative to [16] in the LD(noH), LD(H) and RA configurations, the RC accuracy gains are 0.06%, 0.10% and 0.20%, respectively; the average BD-PSNRs are 0.29 dB, 0.55 dB and 1.08 dB, respectively; and the latency reductions are $-3.10\%$, 0.10%, and $-3.20\%$, respectively. The algorithm proposed by Yang et al. [47] is implemented at the frame and CTU levels, and experiments in HM 16.20 in the LDB configuration show that the bit rate at the same quality is 2.20% lower than that of the RC algorithm in HM 16.20. The algorithm proposed by Zhou et al. [88] is implemented at the frame and CTU levels, and experiments in HM 16.20 show that compared with [86]...

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
and [87] under the all-intra (AI) and LDB coding structures, the RC accuracies are 99.95% and 99.86%, respectively; the BD-PSNR is 0.50 dB in both cases; and the latency reduction is increased by 114% and 1048.2%, respectively, when the time needed to train the model parameters is included and by 1.80% and 1.20%, respectively, when the time needed to train the model parameters is not included. The algorithm proposed by Sanchez et al. [54] was also implemented, and experiments in HM 16.20 show that the RC accuracy is 99.93% with a slight BD-PSNR increase, while the encoding time is increased if the training time is included. The above comparisons indicate that the various algorithms all show improvements in coding efficiency, with different strengths and limitations. Xu et al.’s algorithm minimises the mismatch between the target and real R values, while Ma et al.’s algorithm is effective for one-pass and partial two-pass encoding. A significant PSNR gain along with a slight latency reduction can be guaranteed with Wang et al.’s algorithm. Significant gains in RC accuracy and PSNR can also be seen in Li et al.’s algorithm. Zhou et al.’s and Sanchez et al.’s algorithms are suitable for non-real-time encoding solutions.

From the above evaluations, it can be seen that RC modelling methods have evolved over time, with each model having its own set of advantages and disadvantages. The simple and widely used R-Q models cannot accurately reflect the quality for diverse video content. More accurate R-λ models have been proposed to better reflect the R-D relationship, but these models require frequent and complex updates. An R-ρ model, while improving the linear approximation of the R-D relationship, still establishes only an indirect relationship between R and the QP. Model-free RC methods offer greater flexibility but require more computational resources and have not yet been widely adopted. Overall, RC modelling has become more sophisticated over time, and models should be selected flexibly to balance coding efficiency and performance.

IV. FUTURE WORKS

As efficient tools for compressing video information, video codecs allow service providers to compress video files so that they will occupy minimal storage space and can be efficiently delivered over a range of networks. The purpose of bit rate control is to achieve stable and high-quality video compression to the greatest possible extent under specified bit rate constraints. By removing redundant information, RC methods aim to maintain the original video quality while reducing the amount of data sent over the network or occupying storage space. With the development of virtual and augmented reality applications, video users have gained the ability to interact with and influence objects in immersive three-dimensional (3D) simulated environments that emulate reality with the help of interactive devices, thereby providing an experience equivalent to that of an objective natural environment. This situation has driven the further development of video streaming media in the directions of ultrahigh definition, high dynamics, high frame rates, and high depth [134], [135], [136].

These developments have resulted in ever-increasing demands on video coding techniques for current applications to achieve higher compression efficiency, lower computational complexity, and more intelligent integration into video analysis systems. We believe that the future directions of related research can be summarised from six perspectives: RC methods based on machine learning and deep learning, development from RC methods to quality control methods, perceptual and depth-aware RC methods, RC methods for various advanced video sources, RDO and RC in depth coding, and point cloud RC methods. Future work is expected to focus on topics such as image integration, video capture, encoding, processing, analysis, and understanding, with the objective of guiding a new generation of codecs to effectively and intelligently model the human visual system. A promising possibility is that through the dynamic selection of different RC models, such as R-Q, R-λ, or R-ρ models, or model-free methods in accordance with different video content or coding conditions, the trade-off between coding efficiency and RC accuracy can be optimised [137]. For instance, an R-λ model performs best in terms of coding efficiency and RC accuracy in texture regions of HDR content, but it may not be suitable for non-texture regions. On the other hand, an R-Q model, which is relatively simple, may be more appropriate for simpler motions or textures. By switching between different models based on region characteristics or coding conditions, better RC accuracy and general quality can be achieved while maintaining high coding efficiency. Therefore, more sophisticated algorithms and strategies for dynamically selecting and adapting RC models in real-time video coding applications should be developed. The specific directions of anticipated future development are discussed as follows.

A. Learning-Based Visually Enhanced RC Methods

Machine learning (ML) and deep learning (DL) techniques are widely acknowledged as important tools for analysing and processing massive amounts of weakly correlated or high-dimensional data [138], [139], [140], [141], [142]. As new technologies for video applications (e.g., virtual reality, augmented reality, and point clouds) revolutionise the video coding industry, the heterogeneity and complexity of the captured data are presenting increasing challenges for the efficient compression of these data. Based on the above review of the methods applied to date for video RC in the ML and DL domains, this paper argues that future ML and DL techniques can help to achieve smarter video coding.

Based on the specific requirements of video coding tasks, learning-based RC methods aim to achieve intelligent RC with low complexity, high coding efficiency, and high visual quality. First, ML and DL techniques can be used to combine analysis and recognition tasks with encoding tasks, enabling intelligent RC through the effective reuse of video information and reducing the complexity of video encoding. Second, various learning methods, such as active learning, reinforcement learning, and transfer learning, can be introduced to establish self-updating mechanisms in the relevant models to solve the more complex RC decision-making problems arising in new generations of video coding standards. Third, for the current ML and DL methods applied for video RC, there is still a need to solve the
problems of their relatively high computational complexity and cost. Thus, another important direction of future development for ML and DL RC methods will be to investigate how to realise low-hardware and low-cost implementation solutions based on DL.

B. Perceptual and Depth-Aware RC Methods

Perceptual-based video coding that exploits perceptual redundancy is a promising research area worthy of future consideration [143], [144], [145]. Related research aims to realise an optimal RC mechanism by constructing the relationship between video quality assessment (VQA) results and the coding rate. When a VQA algorithm is to be applied to determine the quality target in a video coding module, it is necessary to adapt the VQA algorithm from image/video-based assessment to block-based assessment and to rate–distortion theory. A convenient method is to model the relationship between the VQA metric of interest and the MSE, given that the MSE is the basis of RDO. In addition, balancing the uniqueness and particularities of perceptual redundancy is also a fundamental difficulty that needs to be overcome. Since the perception of the human visual system for static, dynamic, stereo, and omnidirectional video varies from person to person, it is a daunting challenge to implement a general VQA algorithm suitable for different applications. In addition, the introduction of VQA makes RC complexity control particularly challenging. The computational complexity increases substantially with the adoption of more advanced feature extraction tools and learning-based classifiers for quality prediction. Frequent calls to learning-based VQA algorithms in intensely learning-based RDO schemes can result in extremely complex encoding algorithms. Thus, it will be worth investigating how to integrate VQA, especially better-performing learning-based VQA, into video coding while maintaining a desirable complexity. Future work on perceptual RC will be aimed at developing an objective visual perception model that is broadly advantageous in terms of accuracy, complexity, and adaptability.

C. Interoperable End-to-End RC Methods for Hyperrealistic and High-Dimensional Videos

The rapid evolution of immersive video applications, such as augmented reality (AR), virtual reality (VR) and HDR videos, is presenting new challenges for end-to-end RC methods. Such immersive applications require calculating the positions and angles of camera images in real time and generating corresponding artificial images, with the aim of simulating or supporting interaction with the real world [146], [147], [148], [149], [150], [151]. A VR system uses computer simulation to generate a 3D space, presenting users with a visual and interactive simulation without restrictions. A much more advanced technology is mixed reality (MR), which mixes the real-world environment with AR and VR technologies. This unique experience requires redefining the perceptual quality indicators used in the video encoding process, implementing end-to-end RDO, and outputting stable and high-quality code streams while taking viewing interoperability into consideration. Interoperability can be defined as the ability to achieve a stable viewing angle bit rate through interaction during the encoding process, involving the sharing of user viewing data and encoding information between the client and the server through data exchange via the head-mounted device (HMD). RC interoperability for immersive video is closely related to interoperability in broader human–computer interaction. The need for interoperability between immersive video delivery systems increases the practical value of RC coding. This interoperability can be achieved in two ways: through competition for more network bandwidth and by enabling data-driven decisions based on users’ viewing habits. The users of such immersive experiences will also benefit from improved data quality and a better immersive interactive experience. Therefore, an important direction of development for hyperrealistic and high-dimensional video technology will be to implement an R-D model based on HDR video characteristics in order to improve coding performance and achieve a globally optimal rate allocation scheme in the subsequent RC mechanism. In addition, video content prediction has become very popular in recent years due to its ability to learn from previous viewing behaviour to construct the forthcoming video. Such predicted video content can be widely used in decision-making, autonomous driving, video comprehension, etc. Investigating how best to achieve RC for this type of video will be very important, as all of the abovementioned related tasks demand smooth and high-quality streaming video.

D. Beyond 5G/6G-Powered Quality Control Methods

With the popularisation of 5G networking, high-data-rate and low-latency network connections are increasingly expected to ensure a smooth and high-quality playback experience for video users [152], [153], [154]. Since visual quality is an important aspect of the user experience in many media applications, high quality must be guaranteed while pursuing high smoothness. These issues may become critical as the demand for high-quality video transmission becomes more widespread. The objective of quality control is to keep the video quality within a certain high range under the premise of high bandwidth, which can be achieved by using variable bit rates. However, most existing techniques may not provide constant visual quality and/or efficient compression. It will be more critical for future quality control methods to consider the quality difference between frames as a measure of frame complexity in order to model the relationship among the target bit rate, distortion, and QP. Another essential direction of research will be to pursue the implementation of a low-latency quality control scheme to achieve quality stability.

V. Conclusion

This paper comprehensively reviews the latest progress in RC techniques for the H.265/HEVC and H.266/VVC video coding standards and discusses relevant development prospects. More specifically, this paper first introduces and compares different kinds of RC methods based on various R-D models as well as emerging DL-based schemes. Then, the implementation schemes of RC methods on different hardware platforms are reviewed. Finally, we discuss RC methods
based on ML and DL, the evolution from RC to quality control, RC methods considering human visual perception and depth perception, RC methods for 360-degree/HDR video, and RC methods for predicted video content. A comprehensive summary of directions of future work on topics such as RC methods and RDO in depth coding is also presented. The aim is to provide valuable guidance for the improvement, implementation, application, and continuous development of the current and next generations of RC standards.

REFERENCES

[1] Cisco Annual Internet Report White Paper. Accessed: Mar. 9, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] T. Smith, M. Obrist, and P. Wright, “Live-streaming changes the (video) game,” in Proc. 11th Eur. Conf. Interact. TV Video, Jun. 2013, pp. 131–138.

[3] D. Marpe, T. Wiegand, and G. J. Sullivan, “The H.264/MPEG-4 advanced video coding standard and its applications,” IEEE Commun. Mag., vol. 44, no. 8, pp. 134–143, Aug. 2006.

[4] G. J. Sullivan, J. Ohm, W. 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.

[5] B. Bross et al., “Overview of the versatile video coding (VVC) standard and its applications,” IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 10, pp. 3736–3764, Oct. 2021.

[6] T. Berger, “Rate-distortion theory,” in Wiley Encyclopedia of Telecommunications. Hoboken, NJ, USA: Wiley, 2003.

[7] N. Tang et al., “Fast CTU partition decision algorithm for VTC intra and inter coding,” in Proc. IEEE Asia Pacific Conf. Circuits Syst. (APCCAS), Nov. 2019, pp. 361–364.

[8] Y. Huang and R. P. N. Rao, “Predictive coding,” Wiley Interdiscip. Rev.: Comput. Syst. Sci., vol. 2, no. 5, pp. 580–593, Sep./Oct. 2011.

[9] H. S. Malvar, A. Hallapuro, M. Karueczw, and L. Kirofsky, “Low-complexity transform and quantization in H.264/AVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 598–603, Jul. 2003.

[10] M. Budagavi, A. Fulseth, and G. Bjontegaard, “Hevc transform and quantization,” in High Efficiency Video Coding (HEVC). Berlin, Germany: Springer, 2014, pp. 141–169.

[11] A. Norkin et al., “HEVC deblocking filter,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[12] Y. See 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.

[13] L. Wang, “Rate control for MPEG video coding,” Signal Process., Image Commun., vol. 15, no. 6, pp. 493–511, Mar. 2000.

[14] J.-C. Tsai and C.-H. Hsieh, “Modified TMNS rate control for low-delay video communications,” IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 6, pp. 864–868, Jun. 2004.

[15] Z. Li and F. Pan. Adaptive Basic Unit Layer Rate Control for JVT, document JVT-G012, Joint Video Team (JVT), 7th Meeting, Mar. 2003.

[16] H. Choi, J. Nam, J. Yoo, and D. Sim, “Rate Control Based on Unified RQ Model for HEVC,” document JCTVVC-00123, Joint Collaborative Team Video Coding (JCT-VC), 8th Meeting, San José, CA, USA, Feb. 2012.

[17] B. Li, H. Li, and L. Li, “Rate Control by R-Lambda Model for HEVC,” IEEE Trans. Circuits Syst. Video Technol.,vol. 20, no. 6, pp. 878–894, Sep. 2000.

[18] B. Li, H. Li, and 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.

[19] Z. He and T. Chen, “Linear rate control for JVT video coding,” in Proc. Int. Conf. Inf. Technol. Res., Educ., 2003, pp. 65–68.

[20] M. Jiang, X. Yi, and N. Ling, “Improved frame-layer rate control for H.264 using MAD ratio,” in Proc. IEEE Int. Symp. Circuits Syst., May 2004, p. 813.

[21] J. Xu and Y. He, “A novel rate control for H.264,” in Proc. IEEE Int. Symp. Circuits Syst., May 2004, p. 809.

[22] S. Ma, W. Gao, and Y. Lu, “Rate-distortion analysis for H.264/AVC video coding and its application to rate control,” IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[23] W. Yuan, S. Lin, Y. Zhang, W. Yuan, and H. Luo, “Optimum bit allocation and rate control for H.264/AVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 16, no. 6, pp. 705–715, Jun. 2006.

[24] D.-K. Kwon, M.-Y. Shen, and C.-C.-J. Kuo, “Rate control for H.264 video with enhanced rate and distortion models,” IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 5, pp. 517–529, May 2007.

[25] H. Wang and S. Kwong, “Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 10, pp. 140–144, Jan. 2008.

[26] W.-J. Tsai and T.-L. Chou, “Scene change aware intra-frame rate control for H.264/AVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 12, pp. 1882–1886, Dec. 2010.

[27] H. He, B. Li, W. Lin, W.-H. M. Siu, “Region-based rate control for H.264/AVC for low bit-rate applications,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 11, pp. 1564–1576, Nov. 2012.

[28] X. Liang, Q. Wang, Y. Zhou, B. Luo, and A. Men, “A novel R-Q model based rate control scheme in HEVC,” in Proc. Vis. Commun. Image Process. (VCIP), Nov. 2013, pp. 1–6.

[29] S. Wang, S. Ma, L. Zhang, S. Wang, D. Zhao, and W. Gao, “Multi-layer based rate control algorithm for HEVC,” in Proc. IEEE Int. Symp. Circuits Syst. (ISCAS), May 2013, pp. 41–44.

[30] L. Tian, Y. Zhou, and X. Cao, “A new rate-complexity-QP algorithm (RCQA) for HEVC intra-picture rate control,” in Proc. Int. Conf. Comput., Netw. Commun. (ICNC), Feb. 2014, pp. 375–380.

[31] W. Wu, J. Liu, and L. Feng, “Novel rate control scheme for low delay video coding of HEVC,” ETRI J., vol. 38, no. 1, pp. 185–194, Feb. 2016.

[32] B. Hosking, D. Agrafiotis, D. Bull, and N. Eastern, “An adaptive resolution rate control method for intra coding in HEVC,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2016, pp. 1486–1490.

[34] Y. Mao, M. Wang, S. Wang, and S. Kwong, “High efficiency rate control for versatile video coding based on composite Cauchy distribution,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 4, pp. 2371–2384, Apr. 2022.

[35] C. R. Helmrich, I. Zapancic, J. Brandenburg, V. George, A. Wieckowski, and B. Bross, “Visually optimized two-pass rate control for video coding using the low-complexity XPSNR model,” in Proc. Int. Conf. Vis. Commun. Image Process. (VCIP), Dec. 2021, pp. 1–5.

[36] M. Wang and B. Yan, “Lagrangian multiplier based joint three-layer rate control for H.264/AVC,” IEEE Signal Process. Lett., vol. 16, no. 8, pp. 679–682, Aug. 2009.

[37] B. Lee, M. Kim, and T. Q. Nguyen, “A frame-level rate control scheme based on texture and nontexture rate models for high efficiency video coding,” IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 3, pp. 403–416, Mar. 2014.

[38] 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.
M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, “Region-of-interest-based rate control scheme for high-efficiency video coding,” *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. 1, p. e16, 2014.

Z. Yang, L. Song, Z. Luo, and X. Wang, “Low delay rate control for HEVC,” in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2014, pp. 1–5.

M. Zhou et al., “SSIM-based global optimization for CTU-level rate control in HEVC,” *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 1921–1933, Aug. 2019.

M. Wang, K. N. Ngan, and H. Li, “An efficient frame-content based intra frame rate control for high efficiency video coding,” *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 896–900, Jul. 2015.

H. Zeng, A. Yang, K. N. Ngan, and M. Wang, “Perceptual sensitivity-based rate control method for high efficiency video coding,” *Multimedia Tools Appl.*, vol. 75, no. 17, pp. 10383–10396, Sep. 2016.

F. Wang, Y. Zeng, B. Li, and H.-M. Hu, “Complexity-based intra frame rate control by jointing inter-frame correlation for high efficiency video coding,” *J. Vis. Commun. Image Represent.*, vol. 42, pp. 46–64, Jan. 2017.

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. 24, no. 7, pp. 1240–1244, Nov. 2014.

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.

V. Sanchez, “Rate control for HEVC intra-coding based on piecewise linear approximations,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 1782–1786.

Y. Gong, S. Wan, K. Yang, H. R. Wu, and Y. Liu, “Temporal-layer-mixture and luma-domain picture level rate control for random-access configuration in H.265/HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 156–170, Jan. 2019.

K. R. Perez-Daniel and V. Sanchez, “Luma-aware multi-model rate-control for HDR content in HEVC,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 1022–1026.

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

W. Li, P. Ren, E. Zhang, and F. Zhao, “Rate control for HEVC intra-coding with a CTU-dependent distortion model,” *Signal, Image Video Process.*, vol. 13, no. 1, pp. 17–25, Feb. 2019.

J. Mir, D. S. Talagala, and A. Fernando, “Optimization of HEVC λ-domain rate control algorithm for HDV video,” in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2018, pp. 1–4.

M. Zhou et al., “Global rate-distortion-optimization-based rate control for HEVC HDR coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4648–4662, Dec. 2020.

W. Lim and D. Sim, “A perceptual rate control algorithm based on luminance adaptation for HEVC encoders,” *Signal, Image Video Process.*, vol. 14, no. 5, pp. 887–895, Jul. 2020.

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

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

M. H. Hyun, B. Lee, and M. Kim, “A frame-level constant bit-rate control using recursive Bayesian estimation for versatile video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 2275–22769, 2020.

Y. Chen, S. Kwong, M. Zhou, S. Wang, G. Zhu, and V. Wang, “Intra frame rate control for versatile video coding with quadratic rate-distortion modelling,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4422–4426.

Y. Li, Z. Liu, Z. Chen, and S. Liu, “Rate control for versatile video coding,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 1170–1178.

F. Liu and Z. Chen, “Multi-objective optimization of quality in VVC rate control for low-delay video coding,” *IEEE Trans. Image Process.*, vol. 30, pp. 4706–4718, 2021.

Z. G. Li, F. Pan, K. P. Lim, X. Lin, and S. Rahardja, “Adaptive rate control for H.264,” in *Proc. Int. Conf. Image Process.*, 2004, pp. 745–748.

J. Si, S. Ma, X. Zhang, and W. Gao, “Adaptive rate control for high efficiency video coding,” in *Proc. Vis. Commun. Image Process.*, Nov. 2012, pp. 1–6.
[141] K. Fischer, C. Herglotz, and A. Kaup, “On versatile video coding at UHD with machine-learning-based super-resolution,” in Proc. 12th Int. Conf. Quality Multimedia Expo (QoMEX), May 2020, pp. 1–6.

[142] N. Le, H. Zhang, F. Cricri, R. Ghaznavi-Youvalar, H. R. Tavakoli, and E. Rahtu, “Learned image coding for machines: A content-adaptive approach,” in Proc. IEEE Int. Conf. Multimedia Expo (ICME), Jul. 2021, pp. 1–6.

[143] H. Wang, L. Yu, J. Liang, H. Yin, T. Li, and S. Wang, “Hierarchical predictive coding-based JND estimation for image compression,” IEEE Trans. Image Process., vol. 30, pp. 487–500, 2021.

[144] Y. Li and X. Mou, “Joint optimization for SSIM-based CTU-level bit allocation and rate distortion optimization,” IEEE Trans. Broadcast., vol. 67, no. 2, pp. 500–511, Jun. 2021.

[145] A. Nakhaei and M. Rezaei, “Scene-level two-pass video rate controller for H.265/HEVC standard,” Multimedia Tools Appl., vol. 80, no. 5, pp. 7023–7038, Feb. 2021.

[146] J. Chen, X. Luo, M. Hu, D. Wu, and Y. Zhou, “Sparkle: User-aware viewport prediction in 360-degree video streaming,” IEEE Trans. Multimedia, vol. 23, pp. 3853–3866, 2021.

[147] F. Chiariotti, “A survey on 360-degree video: Coding, quality of experience and streaming,” Comput. Commun., vol. 177, pp. 133–155, Jan. 2020.

[148] T. Zhao, J. Lin, Y. Song, X. Wang, and Y. Niu, Game Theory-Driven Rate Control for 360-Degree Video Coding. New York, NY, USA: Association for Computing Machinery, 2021, pp. 3998–4006.

[149] V. Sanchez, “Rate control for predictive transform screen content video coding based on RANSAC,” IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 11, pp. 4422–4438, Nov. 2021.

[150] K. Perez-Daniel, F. Garcia-Ugalde, and V. Sanchez, “Scene-based imperceptible-visible watermarking for HDR video content,” in Proc. 7th Int. Workshop Biometrics Forensics (IWBF), May 2021, pp. 1–6.

[151] H. Yuan, Q. Wang, Q. Liu, J. Huo, and P. Li, “Hybrid distortion-based rate-distortion optimization and rate control for H.265/HEVC,” IEEE Trans. Consum. Electron., vol. 67, no. 2, pp. 97–106, May 2021.

[152] Q. Zhang, H. Meng, Z. Feng, and Z. Han, “Resource scheduling of time-sensitive services for BSG/6G connected automated vehicles,” IEEE Internet Things J., early access, Nov. 28, 2022, doi: 10.1109/JIOT.2022.3224927.

[153] Y. Cui et al., “Dual identities enabled low-latency visual networking for UAV emergency communication,” in Proc. GLOBECOM IEEE Global Communications Conf., Dec. 2022, pp. 474–479.

[154] A. M. Girgis, J. Park, M. Bennis, and M. Debbah, “Predictive control and communication co-design via two-way Gaussian process regression and AoI-aware scheduling,” IEEE Trans. Commun., vol. 69, no. 10, pp. 7077–7093, Oct. 2021.

Xuekai Wei received the bachelor’s degree in electronic information science and technology and the master’s degree in communication and information systems from Shandong University in 2014 and 2017, respectively, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, China, in 2021. He was a Post-Doctoral Researcher with the School of Artificial Intelligence, Beijing Normal University, Beijing, China, from 2021 to 2022. He is currently an Associate Professor with the School of Computer Science, Chongqing University, Chongqing, China. His current research interests include video coding/transmission and machine learning.

Heqiang Wang received the bachelor’s degree in mechanical design manufacturing and automation from Southwest University. He is currently pursuing the M.S. degree in computer science with Chongqing University. His research interests include perceptual image quality assessment and video coding.

Haoyan Yang received the bachelor’s degree in software engineering from Nanjing Audit University. He is currently pursuing the M.S. degree in electronic information with Chongqing University. His research interests include video coding.

Sam Kwong (Fellow, IEEE) received the B.S. degree from the State University of New York at Buffalo, Buffalo, NY, USA, in 1983, the M.S. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996. From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada, Mississauga, ON, Canada. He joined Bell Northern Research Canada, Ottawa, ON, Canada, as a member of Scientific Staff. In 1990, he became a Lecturer with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China, where he is currently a Professor with the Department of Computer Science. His research interests include video and image coding, and evolutionary algorithms. Prof. Kwong serves as an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS and the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.