Received September 3, 2019, accepted September 23, 2019, date of publication September 26, 2019, date of current version October 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2943887

Down-Sampling Based Rate Control for Mobile Screen Video Coding

TONG TANG, JIN YANG, BIAO DU, AND LIUWEI TANG
School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
Key Laboratory of Optical Communication and Network, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
Key Laboratory of Ubiquitous Sensing and Networking, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Tong Tang (tangtong@cqupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771082 and Grant 61871062, and in part by the Program for Innovation Team Building, Institutions of Higher Education in Chongqing, under Grant CXTDX201601020.

ABSTRACT With the improvement of hardware capability of mobile devices, mobile devices are more and more widely used. Mobile screen video recording is one of the important applications. However, the generated video files will occupy plenty of memory resources, which are limited in mobile devices. Fortunately, a well designed video coding method could effectively relieve such pressure. Therefore, aimed at the mobile screen video, this paper proposes a down-sampling based rate control algorithm to improve the compression efficiency. Firstly, the source video is down sampled by a factor of 4. Then, the down sampled video is encoded twice and the coding information is stored. Finally, based on the stored coding information, the real encoding process is optimized at bit allocation and bit control. Experimental results show that compared with the default rate control method in high efficiency video coding test model, the proposed method could obviously improve rate distortion performance and bit control accuracy.

INDEX TERMS Rate control, video coding, mobile, screen content, HEVC.

I. INTRODUCTION

Use of mobile devices, such as smart phones, has grown dramatically in the last few years and has penetrated the consumer space [1], [2]. A significant reason is that on these devices, there are more and more applications, one of which is the screen video recording. Screen video recording becomes more popular with the development of hardware capability mobile games. However, the dilemma of mobile screen video recording is that even if after compression, the video file size is still large, but the memory resource of mobile is limited. Sometimes users have to delete favorite videos to free up memory space, degrading the user experience. Therefore, improving the compression efficiency of mobile screen video with better encoder could relieve the pressure of memory resource. Moreover, with the development of Internet of Things [3]–[5] and 5G techniques [6], [7], improving the compression performance of mobile videos could combine the mobile application with communication better [8], [9].

The focus of this paper is optimizing the mobile screen video quality with bitrate constraint. A candidate video compression format for use in such environments is the recently standardized HEVC (High Efficiency Video Coding) [10]. HEVC is the latest video coding standard developed by JCT-VC (Joint Collaborative Team of Video Coding), and it can save about 50% bitrate at similar quality compared with H.264 [11]. After finalization of the HEVC base specification, JCT-VC continued to work on extensions. The screen content coding (SCC) extensions [12] improve compression capability for video containing a significant portion of rendered (moving or static) graphics, text, or animation rather than camera captured video scenes.

During the development of video coding standards, corresponding RC (rate control) algorithms are usually recommended for their test or verification or reference models, such as the TM5 [13] for MPEG-2, the VM8 [14] for MPEG-4, the TMN8 [15] for H.263. For the test model of HEVC, the λ (Lagrangian multiplier) domain rate control has been suggested [16]. RC plays an important role in the process of video coding. RC helps to adopt proper coding parameters, aiming at minimizing the distortion of compressed videos with bitrate constraint. It has been acknowledged that RC can be divided into two steps. The first step is bit allocation. Appropriate bits are allocated to each level in the hierarchical video coding structure, including group of picture (GOP) level, frame level and coding unit (CU) level. The second

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

step is bit control. Appropriate coding parameters (such as coding mode, quantization parameter (QP), etc.) are adopted to achieve the target bitrate.

During the development of the HEVC, RC methods can be classified into three types: Q-domain methods, ρ-domain methods and λ-domain methods. In the early HEVC test model (HM6.0), the Q-domain RC method was suggested [17]. But there are two drawbacks to this method: one is that the Q-domain model cannot accurately achieve the bits for the header information, the other one is the well-known "chicken and egg" dilemma [18] between the QP determination and R-D (rate-distortion) optimization process.

ρ-domain method is firstly proposed in [19], building the approximately linear relationship between R and ρ: $R = \theta \cdot (1 - \rho)$, where θ is a model parameter related to video content, ρ denotes the percentage of zero transformed coefficients. Aimed at HEVC, a quadratic ρ-domain method was developed in [20], achieving better compression efficiency than the Q-domain RC method. However, when transform-domain pictures change, the ρ-domain model are required to change synchronously [21], which is a thorny problem to be addressed.

λ-domain method (λ-RC) was suggested in HEVC starting from HM10.0 [10], [22], the relationship between R and λ was built, i.e. $R - \lambda$ model: $\lambda = \alpha \cdot R^{\beta}$, where α and β are model parameters related to video content. Compared with the Q-domain model (i.e. $R - Q$ model), the $R - \lambda$ model could accurately achieve the bits for coding both the residue information and the header information, meanwhile the$R - \lambda$ model could achieve continuous λ values. Therefore from HM10.0, the λ-domain method has been the recommended RC method in HM and lots of improved λ-RC algorithms were proposed [23]–[28], making contributions in terms of bit allocation and bit control. Aiming at bit allocation strategy, in [23] and [24], the CTU level optimal bit allocation problem was formulated and solved, the optimal CTU level bit allocation schemes were proposed. [27] considered the visual importance and allocated bits according to visual weighting map for every CTU, improving the perceptual quality. Aiming at bit control, [25] built a hyperbolic $R - \lambda$ model based on gradient for intra frame, reducing the bit estimation error. [26] pre-encoded the video and captured the complexity distribution, based on the pre-encoding information the CTU level bit allocation was optimized and the $R - \lambda$ model was adjusted at scene cut frames.

Above mentioned RC methods were designed for conventional camera captured video (CCV), as screen content video (SCV) has significantly different characteristics from camera captured video, as shown in Fig. 1, the existing RC methods aimed at CCV may be not suitable for SCV. Therefore in the past few years, some RC methods aimed at SCV have been developed [29]–[33]. Reference [29] proposed a window based RC method for SCC, optimized frame level bit allocation by pre-processing and optimized bit control by adjusting λ values of CTU. Reference [32] proposed an

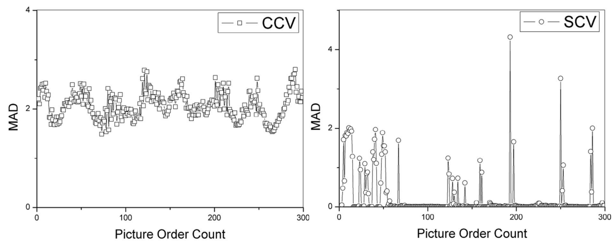


FIGURE 1. Frame MAD(mean absolute difference) distribution comparison between CCV and SCV. CCV used here: *BasketballDrill*, SCV used here: *Slide Editing*.

improved $R - Q$ model based on sum of absolute transformed difference (SATD), improved bit control accuracy. In order to satisfy the low delay coding requirement, [30] classified pictures into two types, set target buffer and controlled bits according to frame type. [30] could prevent buffer overflow and underflow well. Based on [30], [31] classified pictures into three types and set corresponding target buffer, to improve the rate-distortion (R-D) performance. In our previous work [33], we utilized the available buffer space to allocate more bits for complex frames, to reduce quality refinement, and proposed a linear $R - MAD$ model to control bits, achieving better R-D performance and bit control accuracy than [31].

However, the existing RC methods for SCV are designed for on-line applications, insufficient information of the whole video sequence could be obtained when encoding the current frame, thus the R-D performance can't achieve global optimal [34]. The target of this paper is encoding the mobile recorded screen video, which is an off-line application and the R-D performance is our focus. Although there have been some multi-pass encoding methods aimed at off-line applications [34]–[37], the encoding complexity is twice or triple of the single-pass methods. Therefore, in this paper, we propose a down-sampling based rate control for mobile screen video coding, the R-D performance is improved 15% while the averagely encoding time increase is only 17%. Concretely, the main contributions of this paper are as follows:

- The correlations between down-sampling videos (various down-sampling factors) and the original video are analyzed, a proper down-sampling factor in terms of R-D characteristic and encoding complexity is chosen.
- A picture classification method is proposed, pictures are classified into complex and simple type, to enhance bit control.
- An exponential $R - Q$ model having high correlation is proposed for complex pictures.
- Based on twice down-sampling encoding results and our $R-Q$ model, an global optimal bit allocation scheme and a bit control scheme are proposed.

The rest of this paper is organized as follows. Section 2 introduces the proposed RC method in detail. Section 3 presents the experimental results. And conclusions are drawn in Section 4.

II. PROPOSED RC METHOD

In this section, firstly the correlation between down-sampling video and the original video is analyzed and the optimal bit allocation scheme is proposed, next the picture classification method is present, then the $R-Q$ model is built and the bit control scheme is introduced, finally the whole proposed RC method is summarized.

A. BIT ALLOCATION

In this part, the correlations between down-sample videos (various down-sampling factors) and original video are analyzed firstly, then the optimal bit allocation scheme is proposed.

1) DOWN-SAMPLING CODING

In order to obtain sufficient information of the whole video to optimize bit allocation, the video needs to be pre-encoded. But as mentioned above, two-pass encoding costs double time of the real encoding process. Thus in this paper, the original video is down-sampled firstly, and the down-sampled video is pre-encoded to obtain the information of the whole video, so that the pre-encoding time could be reduced significantly.

As we know, the larger the down-sampling factor is, the smaller the resolution of down-sampled video is. Smaller video resolution means less encoding time, but meanwhile means more loss of information. In other words, if the original video is down sampled with a large factor, though the pre-encoding time could be reduced extremely, down-sampled video loses much information and the pre-encoding results can not represent the real R-D characteristic of the origin video.

To obtain a proper down-sampling factor, we conducted a simple experiment. Two screen content video sequences (*ClearTypeSpreadsheet* and *WordEditing*) were randomly selected to encode. Concretely, each video was down-sampled with the factor of 2, 4 and 8, abbreviated as D2 video, D4 video and D8 video respectively. The original video and down-sampled videos were encoded with a fixed QP, configured with *encoder_randomaccess_main_scc.cfg*, the bit consumption of every frame and the total encoding time were recorded.

The encoding time consumption was measured as

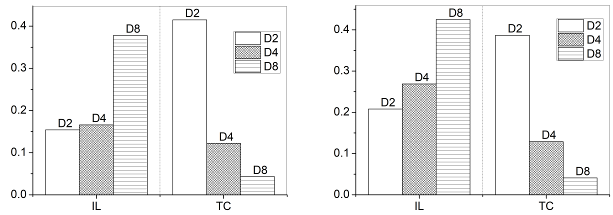
$$TC = \frac{T^s}{T^{ori}} \tag{1}$$

where, T^s is the encoding time of the video down-sampled by the factor of s, T^{ori} is the encoding time of the original video.

The information loss of down-sampled video was defined as

$$IL = \sum_{i=1}^{N} \left| \frac{R^s(i)}{R^s_{all}} - \frac{R^{ori}(i)}{R^{ori}_{all}} \right| \tag{2}$$

where N is the frame number of the whole video, $R^s(i)$ is the bit cost of the $i-th$ frame in the video down-sampled by the factor of s, R^s_{all} is the total bit cost of video down-sampled by the factor of s, $R^{ori}(i)$ is the bit cost of the $i-th$ frame in the original video and R^{ori}_{all} is the total bit cost of the original video. IL measures the difference between the down-sampled video and the original video in terms of the proportion of frame bit consumption over total bits. The bigger the difference is, the more information loss.



(a) *ClearTypeSpreadsheet* (b) *WordEditing*

FIGURE 2. Information loss (*IL*) and encoding time consumption (*TC*) comparisons for videos down-sampled by different factors.

Fig. 2 shows the information loss and encoding time consumption comparisons for videos down-sampled by different factors (2, 4, 8). We can see that D4 video has similar performance with D2 video in terms of information loss, meanwhile the encoding time of D4 video decreases several times compared with D2 video. Moreover, the encoding time of D8 video decreases nearly ten times compared with D2 video, but D8 video has over twice as much information loss as D2 video. Therefore, finally we choose the factor of 4 to conduct downsampling of videos, making a compromise of information loss and encoding time.

2) BIT ALLOCATION SCHEME

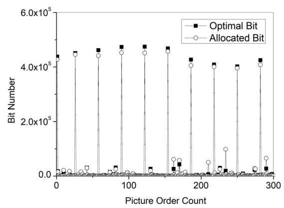
The bit allocation could be optimized based on the coding results of down-sampled video. Because the fixed-QP encoding is nearly optimal, and the information loss of D4 video is similar with D2 video, which is acceptable, thus the fixed-QP encoding result of D4 video is also nearly optimal.

Therefore, in this paper, the bit cost of fixed-QP encoding of D4 video is used to allocate bits for every frame as

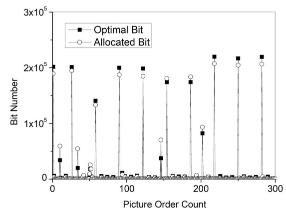
$$R_T(i) = \frac{R^D(i)}{\sum_{j=i}^{N} R^D(j)} \cdot R^{seq}_{left} \tag{3}$$

where, $R_T(i)$ is the allocated bit of the $i-th$ frame, $R^D(i)$ is the bit cost of the $i-th$ frame in the D4 video, N is the frame number of the whole video and R^{seq}_{left} is the left bits of the real encoding video sequence. It should be noted that the pre-encoding process for D4 video is conducted twice, and the encoding QP values are 20 and 40 respectively. Thus R^D in Eq. (3) is the averaged bit cost of twice pre-encoding results. The reason we use twice pre-encoding results is that in the bit control scheme (will be addressed in the following section), twice pre-encoding is needed, so the averaged result is used to make the bit allocation scheme more robust. The reason we use QP of 20 and 40 is that in the common test condition, the QP range is [22, 37], thus the calculated $R-Q$ model by twice pre-encoding results could be effective for common test condition.

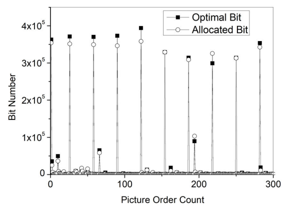
Fig. 3 shows the performance of the bit allocation scheme in Eq. (3). We can see that for different video sequences,



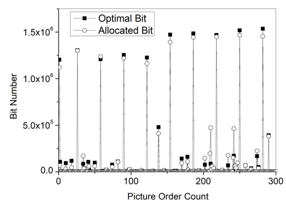
(a) *ClearTypeSpreadsheet*



(b) *sc_video_conference*



(c) *Slide Editing*



(d) *ChineseDocumentEditing*

FIGURE 3. The performance of the bit allocation scheme. Optimal Bit: the consumed bits of the original video encoded with fixed QP of 32. Allocated Bit: obtained by Eq. (3).

the proposed bit allocation scheme could achieve nearly optimal, which verify the effectiveness of the proposed bit allocation scheme.

B. BIT CONTROL

In this part, a highly correlated $R - Q$ model is proposed firstly, then a picture classification scheme is introduced, finally the bit control scheme is present.

1) $R - Q$ MODEL

As shown in Fig. 1, for screen content videos, most of the bits are consumed by a few frames, called bit-consuming frames. Therefore, if the bits of these bit-consuming frames are well controlled, then the bits of the whole video are well controlled.

Thus we randomly selected several screen content video sequences to encode with fixed QP, configured with *encoder_randomaccess_main_scc.cfg*, a few bit-consuming frames were selected to investigated the relationship between the frame-level bit cost and the frame QP.

Fig. 4 shows relationship between frame-level bit cost BPP (bit per picture) and frame QP for different pictures from different video sequences, where POC means picture order count. We can see that BPP and QP obeys good exponential relationship, and the Adj. R-Square (degree-of-freedom adjusted coefficient of determination) values are approximate to 1, which verify the good fitness of the exponential model.

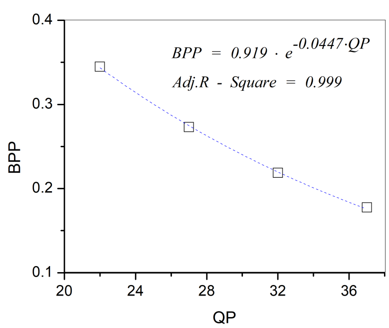
Therefore, aimed at the bit-consuming frames, an exponential $R - Q$ model is proposed as

$$R = a \cdot e^{b \cdot QP} \tag{4}$$

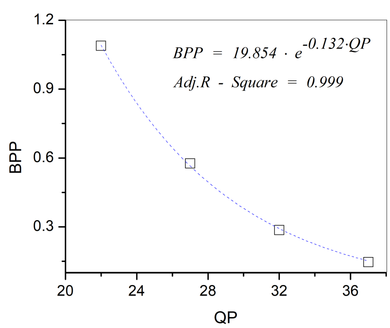
where, a and b are model parameters, which are related to the video contents.

2) PICTURE CLASSIFICATION

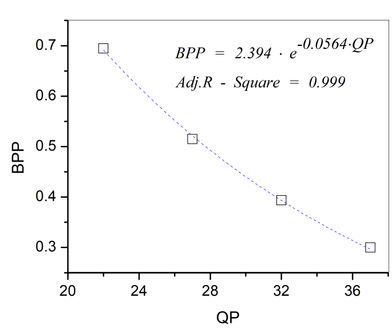
As shown in Fig. 4, different pictures have different model parameters (a, b). But in the real encoding process,



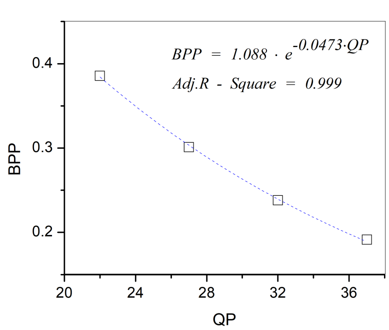
(a) *sc_video_conference*, POC = 0



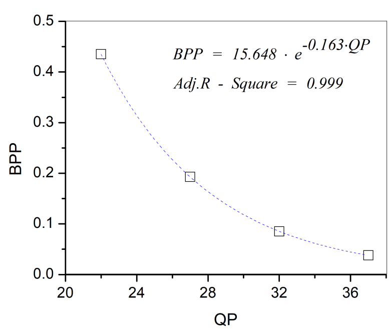
(b) *sc_robot*, POC = 0



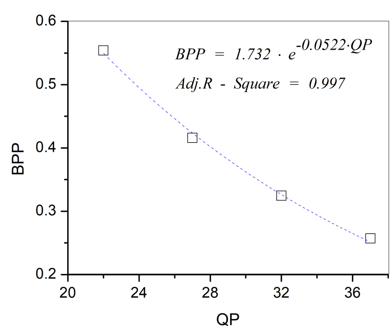
(c) *Slide Editing*, POC = 0



(d) *sc_video_conference*, POC = 288



(e) *sc_robot*, POC = 136



(f) *Slide Editing*, POC = 224

FIGURE 4. The performance of the proposed $R - Q$ model.

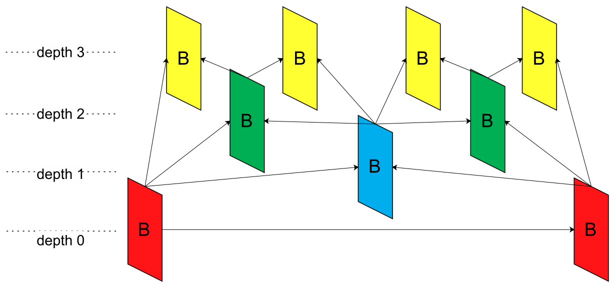
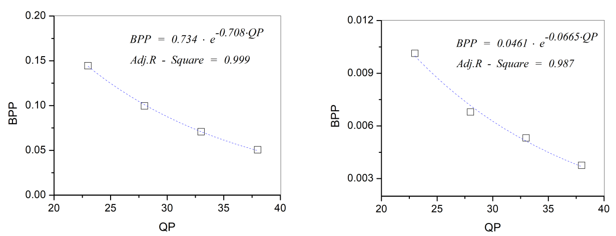


FIGURE 5. An example of an RA (random access) prediction structure with a GOP size of 8.



(a) *Slide Editing*, POC = 72 (b) *Slide Editing*, POC = 80

FIGURE 6. R-D characteristics comparison of two adjacent pictures at the same depth (depth = 1) in ***Slide Editing***.

model parameters of the current frame cannot be get before encoding the current frame. Thus, in conventional RC methods, model parameters of the current frame are predicted by previous encoded frames [13], [14], [22]. Since the hierarchical GOP structure was adopted from H.264 (as shown in Fig. 5), adjacent pictures at different depths have obviously different R-D characteristics. Therefore, in conventional RC methods aimed at H.264 or HEVC, pictures at the same depth share one set of model parameters [22], and the model parameters are updated with a smooth update strategy [16]. In other words, in conventional RC methods, pictures at the same depth are classified into the same type, having similar RC characteristics.

However, the conventional picture classification method and smooth update strategy are based on the premise that adjacent pictures have continuous content, which is the feature of the conventional camera captured video. And for screen content video, the video contents are discontinuous, thus though two adjacent pictures are at the same depth, their R-D characteristics may be obviously different, as shown in Fig. 6. We can see that though two adjacent pictures are both at depth 1, their model parameters (a, b) have great differences.

Therefore, similar with our previous work [33], in this paper, pictures are classified according to the complexity. Based on the down-sampling results, the frame-level bit cost of encoded D4 video is used to measure picture complexity, and the picture classification scheme is as

$$type[i] = \begin{cases} complex, & if\ R^D(i) > \beta \cdot R^D_{all} \\ simple, & otherwise \end{cases} \tag{5}$$

where, $R^D(i)$ is the bit cost of the $i-th$ frame in the encoded D4 video, R^D_{all} is the total bit cost of the whole encoded D4 video and β is the threshold.

From Eq. (5), if β is too small, most pictures will be classified into "complex", the correlation of the $R-Q$ model applied to complex picture will very bad, causing the bit costs of complex pictures out of control. If β is too large, few pictures will be classified into "complex", maybe the correlation of the $R-Q$ model applied to complex picture will be very good, but even if the bit costs of all the complex pictures are controlled well, the bit costs of the whole video may be not controlled well, because most bits are consumed by the simple frames.

Therefore, the value of β is important for the bit control of the whole video. And the selection of β value will be addressed in the next section II-B.3.

3) BIT CONTROL SCHEME

Based on the $R-Q$ model in Eq. (4) and the picture classification scheme in Eq. (5), the bit control scheme is proposed. Actually, the focus of bit control is the determination of the parameters a and b in $R-Q$ model.

As shown in Fig. 1, the contents of screen content videos change dramatically. Thus adjacent pictures of the same type may still have different R-D characteristics, which means that the model parameters of previous encoded adjacent picture (updated with smooth strategy) cannot serve as the model parameters of the current picture.

In this paper, the model parameters of the current picture are predicted based on the encoding results of the down-sampled video (one down-sampled video is encode twice with the QP of 20 and 40, to obtain two model parameters). Concretely, the model parameters are predicted as

$$\begin{cases} \tilde{a}^{ori}_{comp}(i) = \dfrac{a^D_{comp}(i)}{a^D_{comp}(i-1)} \cdot a^{ori}_{comp}(i-1) \\ \tilde{b}^{ori}_{comp}(i) = \dfrac{b^D_{comp}(i)}{b^D_{comp}(i-1)} \cdot b^{ori}_{comp}(i-1) \end{cases} \tag{6}$$

where, $\tilde{a}^{ori}_{comp}(i)$ is the predicted parameter a of the $i-th$ complex picture in the original video, $a^D_{comp}(i)$ is the actual parameter a of corresponding picture in the D4 video, $\tilde{b}^{ori}_{comp}(i)$ is the predicted parameter b of the $i-th$ complex picture in the original video and $b^D_{comp}(i)$ is the actual parameter b of corresponding picture in the D4 video.

Obviously, the prediction accuracy of Eq. 6 is influenced by the picture classification result, which is decided by the value of β. In order to select a proper β, we randomly selected several screen content videos to encode with fixed QP, configured with *encoder_randomaccess_main_scc.cfg*. Then we changed β, calculated the corresponding predicted model parameter $\tilde{a}$ and actual model parameter a of every complex picture, parameter prediction error is averaged as in Eq. (7), where M is the total number of the complex pictures, $a^{ori}_{comp}(i)$ is the actual parameter a of the $i-th$ complex picture in the original video. Moreover, we calculated the

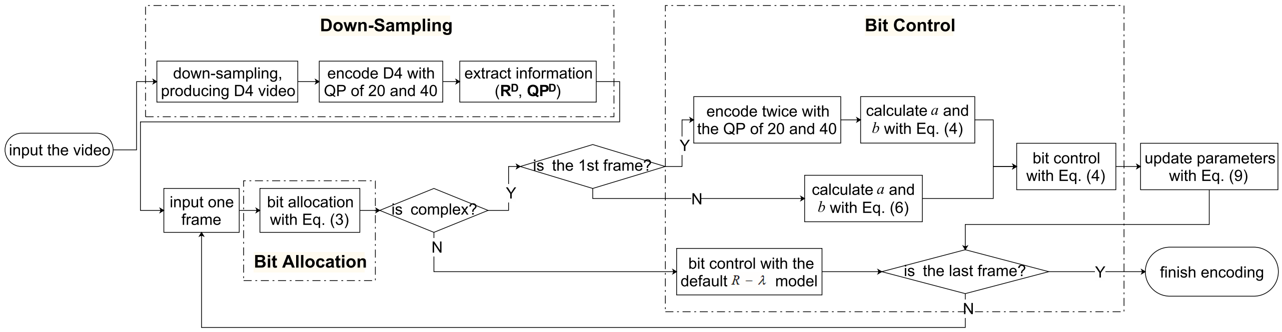


FIGURE 7. Flowchart of the proposed RC method.

corresponding bit cost proportion of the complex pictures compared with the total bit budget of the original video as in Eq. (8), where $R^{ori}_{comp}(i)$ is the bit cost of the $i-th$ complex picture in the original video and R^{ori}_{all} is the total bit budget of the original video.

$$Err = \sum_{i=1}^{M} \left| \tilde{a}^{ori}_{comp}(i) - a^{ori}_{comp}(i) \right| / M \tag{7}$$

$$BP = \sum_{i=1}^{M} R^{ori}_{comp}(i) / R^{ori}_{all} \tag{8}$$

Fig. 8 shows two example of the relationship between β and prediction error (Err), and bit cost proportion of complex pictures (BP). We cam see that along with the increase of β, generally, Err and BP both decrease, which verifies our above analysis. And our target is to select a proper β to achieve small Err and large BP, thus finally we set β to 0.018 in our experiment.

Moreover, after encoding one complex picture, its model parameters will be updated as

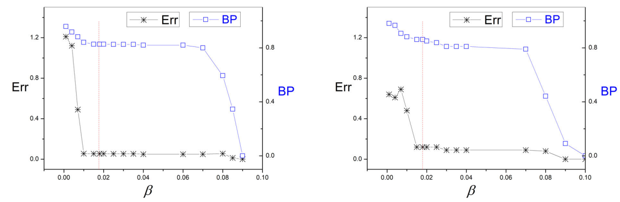
$$\begin{cases} a_{new} = a_{old} + \delta_a \cdot (BPP_{real} - BPP_{calc}) \\ b_{new} = b_{old} + \delta_b \cdot (BPP_{real} - BPP_{calc}) \cdot QP \end{cases} \tag{9}$$

where, a_{new} and b_{new} are the updated parameters, a_{old} and b_{old} are the old parameters, BPP_{real} is the real bit cost of the current frame, BPP_{calc} means the predicted bit cost calculated by the $R-Q$ model, δ_a and δ_b are constants, measuring the updating speed, set to 0.1 and 0.05 respectively in our experiment, and QP is the average QP of the current frame.

It should be noted that above bit control scheme aims at controlling the bit cost of the complex frames, after adopting our $R-Q$ model and picture classification scheme, the bit cost of complex frames could be controlled well and the complex frames consume most bits of the whole video sequence. For simple frames, we continue to use the default $R-\lambda$ model in HM to control the bit cost.

C. SUMMARY

The proposed RC procedure could be summarized as in Fig. 7. Firstly, the original video is down-sampled with the factor of 4, producing the D4 video. Next, the D4 video



(a) *ClearTypeSpreadsheet* (b) *ChineseDocumentEditing*

FIGURE 8. Relationship between β and prediction error (***Err***), and bit cost proportion of complex pictures (***BP***).

is encoded twice with the QP of 20 and 40 respectively, the encoding information (R^D, QP^D) is restored. Then, based on the bit cost of D4 video R^D, current frame is allocated bits according to Eq. (3). Then, if the current frame is the first frame, it is encoded twice with the QP of 20 and 40 and its model parameters (a and b) are obtained by Eq. (4). Otherwise if the current frame is complex frame, its model parameters are obtained by Eq. (6), and the bit of the current frame is controlled by Eq. (4). Otherwise if the current frame is simple frame, its bit is controlled by the default $R-\lambda$ model. Finally, after encoding the current frame, model parameters are updated with Eq. (9).

III. EXPERIMENTAL RESULT

In this section, firstly the experimental design is introduced, then several important indexes to measure RC method are compared including video quality, bit control accuracy and coding complexity.

A. EXPERIMENTAL DESIGN

To verify the effectiveness of the proposed RC method, we implemented it into the HEVC reference software (HM-16.10+SCM-8.0) [38]. The test platform is a PC with an twelve-core Intel Core i7 @3.20GHz CPU, 8GB RAM, and x64 Windows7. Our testing materials are the recommended SCC sequences of JCT-VC proposals [39]–[41], to simplify the following introduction, every test sequence is defined an abbreviation, as shown in Tab. 1. As this paper targets the off-line application, the R-D performance is the most important consideration. And RA encoding mode could achieve superior R-D performance than intra-only and

TABLE 1. Screen content testing sequences.

Type	Resolution	Format	Sequence name	Abbreviation	Frame number	FPS
TGM	1920 × 1080	YUV 4:2:0	ChineseDocumentEditing_1920x1080_30_8bit_420.yuv	*CDE*	300	30
	1920 × 1080	YUV 4:2:0	ClearTypeSpreadsheet_1920x1080_30_8bit_420.yuv	*CTS*	300	30
	1920 × 1080	YUV 4:2:0	EnglishDocumentEditing_1920x1080_30_8bit_420.yuv	*EDE*	300	30
	1920 × 1080	YUV 4:2:0	sc_desktop_1920x1080_60_8bit_420.yuv	*DESK*	600	60
	1280 × 720	YUV 4:2:0	SlideEditing_1280x720_30.yuv	*SE*	300	30
M	1280 × 720	YUV 4:4:4	sc_video_conferencing_doc_sharing_1280x720_30_300_8bit_444.yuv	*VCDS*	300	30
	1280 × 720	YUV 4:4:4	sc_web_browsing_1280x720_30_8bit_300_444_r1.yuv	*WB*	300	30
	1920 × 1080	YUV 4:2:0	BigBuck_1920x1080_60_8bit_420.yuv	*BB*	400	60
A	1280 × 720	YUV 4:2:0	sc_robot_1280x720_30_8bit_300_420.yuv	*ROBOT*	300	30
	1024 × 768	YUV 4:2:0	ChinaSpeed_1024x768_30.yuv	*CS*	500	30

TGM: text and graphics with motion; M: mixed content; A: animation; FPS: frames per second

TABLE 2. Main encoding parameters.

Codec	HM-16.10+SCM-8.0 [38]
Encoding structure	Random-access configuration
profile	Main_scc
IntraPeriod	Approximately one second [40]
GOPsize	8
SearchRange	64
Deblock Filter	ON
SAO	ON

low-delay encoding modes, due to the RA mode's more effective inter picture prediction [42]. Therefore, all the testing sequences are encoded with RA mode, configured with *encoder_randomaccess_main_scc.cfg*. The main encoding parameters are listed in Tab. 2.

Our proposed method is compared with the recommended RC scheme [16] in HEVC. In [16], there are three bit allocation strategies, i.e. equal bit allocation (EBA), fixex ratio bit allocation (FRBA) and adaptive ratio bit allocation (ARBA). ARBA is the improved method of FRBA, adding adaption by considering video encoding characteristics. Since the screen content video has discontinuous contents, in our experiment, EBA based RC method (EBA-RC) and ARBA based RC method (ARBA-RC) are compared with the proposed method.

According to the CTC proposal [40], four QP values 22, 27, 32, 37 are set to the base QP of the constant-QP encoding method (CQP), the final bitrate cost of CQP is set to the target bitrate of other RC methods.

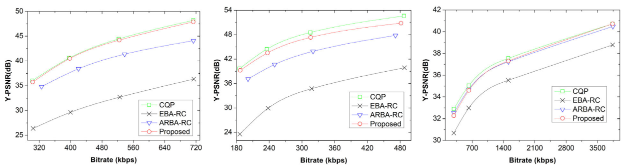
B. VIDEO QUALITY COMPARISON

The coding efficiency is measured by BD-rate and BD-PSNR [43], which represent the average bit rate reduction, the anchor of three RC methods (EBA-RC, ARBA-RC and the proposed RC method) is the CQP method. Tab. 3 shows the R-D performance of three RC methods.

TABLE 3. R-D performance comparison.

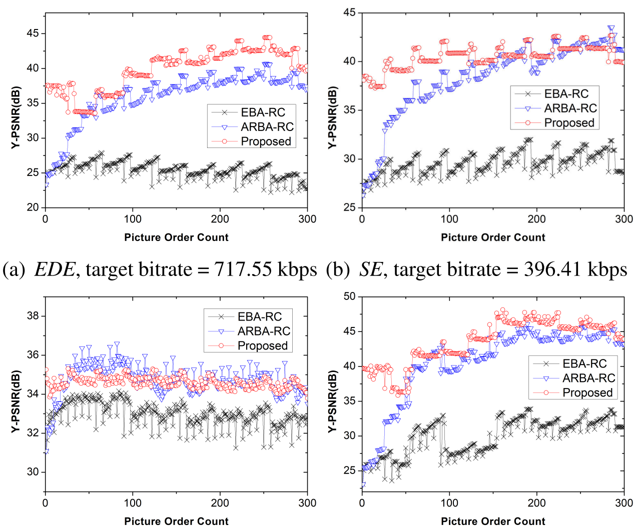
Type	Sequence	BD-rate (%)		
		EBA-RC	ARBA-RC	Proposed
TGM	*CDE*	111.65	21.04	14.21
	CTS	145.36	56.14	15.87
	EDE	222.39	38.55	6.48
	DESK	71.76	18.78	14.96
	SE	131.33	26.10	1.61
M	*VCDS*	158.82	40.10	8.93
	WB	227.02	43.95	4.96
	BB	189.87	30.19	18.12
A	*ROBOT*	96.58	11.02	9.32
	CS	33.82	4.14	3.23
Average		138.86	29.00	9.77

We can see that firstly compared with the CQP method, the proposed RC method has averagely 9.77% bit rate increase, which is much better than the EBA-RC (138.86%) and ARBA-RC (29.00%). Then, the proposed RC method could achieve better R-D performance than other two RC methods for all the testing sequences in different types. The significant R-D performance improvement of the proposed method mainly owes to the global optimal bit allocation scheme in section II-B.3. Finally, the proposed method could obtain more improvement against other two RC methods on the videos with the type of TGM and M, which is depicted in Fig. 9. Because for TGM and M videos, the video contents are discontinuous, thus the proposed bit allocation scheme could obviously outperform the conventional hierarchical bit allocation scheme in ARBA-RC and the equal bit allocation scheme in EBA-RC. While for A videos (animation),



(a) *SE (TGM)* (b) *VCDS (M)* (c) *ROBOT (A)*

FIGURE 9. RD curves comparison.



(a) *EDE*, target bitrate = 717.55 kbps (b) *SE*, target bitrate = 396.41 kbps

(c) *ROBOT*, target bitrate = 625.38 kbps (d) *VCDS*, target bitrate = 235.69 kbps

FIGURE 10. Comparison of three RC methods in terms of PSNR distribution for different video sequences.

the picture complexity distribution is similar to conventional camera captured videos, though the video contents are generated by computer, the video contents are continuous. Therefore the conventional hierarchical bit allocation scheme and the proposed bit allocation scheme could both achieve approximately optimal, having similar R-D performance.

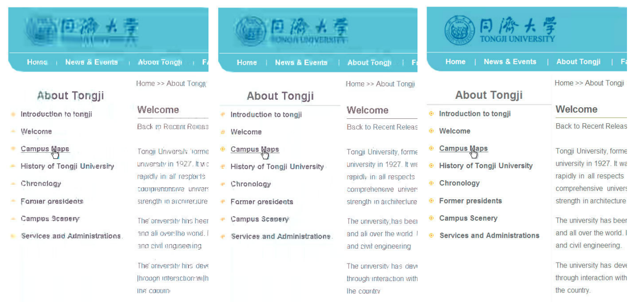
Fig. 10 shows the comparison of three RC methods in terms of PSNR distribution. We can see that for different video sequences, the proposed RC method could achieve the best consistent quality and the highest quality, which verifies the effectiveness of our proposed bit allocation scheme and the bit control method.

Fig. 11 shows the comparison of three RC methods in terms of subjective quality. We can see that the proposed RC method could achieve the best quality, obviously better than other two RC methods, which owes to the proposed bit allocation scheme.

C. BIT CONTROL ACCURACY COMPARISON

The sequence-level bit control error is defined as

$$\Delta R^{seq-L} = \frac{\left|R^{seq}_{actual} - R^{seq}_{tar}\right|}{R^{seq}_{tar}} \times 100\% \tag{10}$$



(a) EBA-RC (b) ARBA-RC (c) Proposed

FIGURE 11. Subjective quality comparison of three RC methods for the $17-th$ frame in sequence *WB*, encoded with the target bitrate of 216.08 kbps.

TABLE 4. Sequence-level bit control accuracy comparison.

Type	Sequence	Bit control error (%)		
		EBA-RC	ARBA-RC	Proposed
TGM	*CDE*	2.47	8.64	2.05
	CTS	0.82	3.41	0.24
	EDE	0.18	2.37	2.37
	DESK	0.69	2.56	2.40
	SE	0.54	4.14	0.23
M	*VCDS*	0.55	4.75	0.58
	WB	0.61	1.56	0.93
	BB	0.36	1.97	0.75
A	*ROBOT*	0.089	0.10	0.0037
	CS	0.037	0.013	0.0013
Average		0.63	2.95	0.95

where, R^{seq}_{actual} is the actual bit cost of the whole sequence, R^{seq}_{tar} is the target bits of the whole sequence.

The frame-level bit control error is defined as

$$\Delta R^{frame-L} = \frac{\sum_{i=1}^{N}\left|R^{i}_{actual} - R^{i}_{tar}\right|}{R^{seq}_{tar}} \times 100\% \tag{11}$$

where, R^{i}_{actual} is the actual bit cost of the $i-th$ frame, R^{i}_{tar} is the target bits of the $i-th$ frame.

The sequence-level bit control accuracy comparisons are shown in Tab. 4 and the frame-level bit control accuracy comparisons are shown in Tab. 5. We can see that in terms of sequence-level bit control accuracy, the proposed method has averagely 0.95% bit control error, which is close to EBA-RC (0.63%) and has obvious improvement against ARBA-RC method (2.95%). In terms of frame-level bit control accuracy, the proposed method has averagely 0.21% bit control error, which is better than EBA-RC (0.35%) and ARBA-RC (0.67%). And the proposed method could achieve the best accurate bit control at frame level for all the sequences, verifying the effectiveness of the proposed bit control scheme.

TABLE 5. Frame-level bit control accuracy comparison.

Type	Sequence	Bit control error (%)		
		EBA-RC	ARBA-RC	Proposed
TGM	*CDE*	0.50	0.83	0.26
	CTS	0.46	0.75	0.28
	EDE	0.38	0.71	0.28
	DESK	0.32	0.60	0.32
	SE	0.46	0.89	0.14
M	*VCDS*	0.48	1.00	0.34
	WB	0.44	0.88	0.27
	BB	0.25	0.72	0.15
A	*ROBOT*	0.087	0.19	0.042
	CS	0.10	0.13	0.07
Average		0.35	0.67	0.21

TABLE 6. Algorithm complexity comparison.

Type	Sequence	complexity increase (%)		
		EBA-RC	ARBA-RC	Proposed
TGM	*CDE*	128.4	51.6	62.2
	CTS	80.6	42.7	52.8
	EDE	114.6	59.6	36.1
	DESK	34.5	4.6	39.5
	SE	124.0	52.8	74.9
M	*VCDS*	155.2	47.6	62.7
	WB	138.4	49.8	70.8
	BB	86.2	23.5	40.2
A	*ROBOT*	26.1	1.8	14.6
	CS	6.9	2.6	15.6
Average		89.5	33.6	46.9

D. ALGORITHM COMPLEXITY COMPARISON

The coding complexity of three RC methods is compared by using CQP as the anchor. And compared with CQP method, the encoding time increase ΔT is defined as

$$\Delta T = (T^{RC} - T^{CQP})/T^{CQP} \times 100\% \tag{12}$$

where, T^{RC} is the total encoding time of the RC method to be compared, T^{CQP} is the total encoding time of CQP method.

The complexity of three RC methods is compared in Tab. 6. We can see that all the three RC methods have encoding complexity increase compared with CQP method. Because firstly the QP determination process of RC method consumes time. Then, the bit allocation schemes of EBA-RC and ARBR-RC are not optimal, causing the low quality of key frames, which will result in quality refinement [44], and quality refinement will increase encoding complexity. Finally, though the proposed RC method uses nearly optimal bit allocation scheme, the bit control is not absolutely accurate, thus quality refinement still exist. Moreover, in the proposed RC method, twice pre-encoding processes of down-sampled video are needed, which causing averagely 19.3% complexity increase. In summary, the quality refinement is most serious in EBA-RC, least serious in the proposed method, but the proposed method needs twice pre-encoding processes of down-sampled video. Therefore, EBA-RC has the highest complexity (89.5% increase), the proposed method has the secondary high complexity (46.9% increase) and ARBA-RC has the lowest complexity (33.6% increase).

Meanwhile, we can see from Tab. 6 that three RC methods have the least complexity increase for A type videos. Because as mentioned above, animation type video has continuous content, which is similar to conventional camera captured video. Thus the bit allocation scheme in three RC methods will cause much less quality refinement and much less complexity increment.

IV. CONCLUSION

Screen content videos have been widely used in mobile applications, such as mobile video recording. And screen content videos have significantly different characteristics from camera-captured videos. To reduce the storage space of screen content coding, in this paper, a down-sampling based rate control for mobile screen video coding is proposed. Firstly, the down-sampling size is determined by compromise between information loss and encoding complexity. Nextly, the original video is down-sampled and encoded twice, the encoding information is stored. Then, based on the encoding information of down-sampled video, the global optimal bit allocation scheme, an accuracy $R - Q$ model, a picture classification scheme and a bit control scheme are proposed. Finally, experimental results verify the effectiveness of the proposed RC method in terms of video quality and bit control accuracy.

REFERENCES

[1] R. Garcia, V. Adzic, and H. Kalva, "Adapting low-bitrate skip mode in a mobile environment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 2, pp. 352–365, Feb. 2017.

[2] Y. Zhang, S. Kwong, G. Zhang, Z. Pan, H. Yuan, and G. Jiang, "Low complexity HEVC INTRA coding for high-quality mobile video communication," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1492–1504, Dec. 2015.

[3] D. Wu, H. Shi, R. Wang, H. Fang, and H. Wang, "A feature-based learning system for Internet of Things applications," *IEEE Internet Things J.*, vol. 6, no. 8, pp. 1928–1937, Apr. 2019.

[4] Z. Li, Y. Jiang, Y. Gao, D. Yang, and L. Sang, "On buffer-constrained throughput of a wireless-powered communication system," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 2, pp. 283–297, Feb. 2019.

[5] P. Zhang, X. Kang, D. Wu, and R. Wang, "High-accuracy entity state prediction method based on deep belief network toward IoT search," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 492–495, Apr. 2019.

[6] D. Wu, Z. Zhang, S. Wu, R. Wang, and J. Yang, "Biologically inspired resource allocation for network slices in 5G-enabled Internet of Things," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2888543.

[7] Z. Zhang and L. Wang, "Social tie-driven content priority scheme for D2D communications," *Inf. Sci.*, vol. 480, pp. 160–173, Apr. 2019.

[8] Z. Li, J. Chen, and Z. Zhang, "Socially aware caching in D2D enabled fog radio access networks," *IEEE Access*, vol. 7, pp. 84293–84303, 2019.

[9] D. Wu, L. Deng, H. Wang, K. Liu, and R. Wang, "Similarity aware safety multimedia data transmission mechanism for Internet of vehicles," *Future Gener. Comput. Syst.*, vol. 99, pp. 609–623, Oct. 2019.

[10] 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.

[11] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[12] W.-H. Peng, F. G. Walls, R. A. Cohen, J. Ostermann, A. MacInnis, T. Lin, and J. Xu, "Overview of screen content video coding: Technologies, standards, and beyond," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 4, pp. 393–408, Dec. 2016.

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

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

[15] J.-C. Tsai and C.-H. Shieh, "Modified TMN8 rate control for low-delay video communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 864–868, Jun. 2004.

[16] 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.

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

[18] X. Jing, L.-P. Chau, and W.-C. Siu, "Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control," *IEEE Signal Process. Lett.*, vol. 15, no. 3, pp. 373–376, Mar. 2008.

[19] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[20] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP based rate control for High Efficiency Video Coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1101–1111, Dec. 2013.

[21] 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. 465–479, Mar. 2014.

[22] B. Li, H. Li, L. Li, and J. Zhang, *Rate Control by R-Lambda Model for HEVC*, document JCTVC-K0103, ISO/IEC, ITU-T, Shanghai, China, 2012.

[23] M. Wang and K. N. Ngan, "Optimal bit allocation in HEVC for real-time video communications," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Quebec City, QC, Canada, Sep. 2015, pp. 2665–2669.

[24] S. Li, M. Xu, and Z. Wang, "A novel method on optimal bit allocation at LCU level for rate control in HEVC," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Turin, Italy, Jun./Jul. 2015, pp. 1–6.

[25] 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.

[26] J. Wen, M. Fang, M. Tang, and K. Wu, "R-λ model based improved rate control for HEVC with pre-encoding," in *Proc. IEEE Data Comp. Conf. (DCC)*, Snowbird, UT, USA, Apr. 2015, pp. 53–62.

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

[28] D. Wu, Q. Liu, H. Wang, Q. Yang, and R. Wang, "Cache less for more: Exploiting cooperative video caching and delivery in D2D communications," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1788–1798, Jul. 2019.

[29] Y. Guo, B. Li, S. Sun, and J. Xu, "Rate control for screen content coding in HEVC," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCS)*, Lisbon, Portugal, May 2015, pp. 1118–1121.

[30] Y. Guo, B. Li, S. Sun, and J. Xu, "Rate control for screen content coding based on picture classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Singapore, Dec. 2015, pp. 1–4.

[31] J. Xiao, B. Li, S. Sun, and J. Xu, "Rate control with delay constraint for screen content coding," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, St. Petersburg, FL, USA, Dec. 2017, pp. 1–4.

[32] S. Wang, J. Li, S. Wang, W. Gao, and S. Ma, "A frame level rate control algorithm for screen content coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCS)*, Florence, Italy, May 2018, pp. 1–4.

[33] T. Tang and L. Li, "A low delay rate control method for screen content coding," *Multimedia Tools Appl.*, vol. 78, no. 19, pp. 28231–28256, 2019. doi: 10.1007/s11042-019-07910-9.

[34] G. Cao, X. Pan, Y. Zhou, Z. Chen, and Y. Li, "Two-pass rate control for constant quality in high efficiency video coding," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Taichung, Taiwan, Dec. 2018, pp. 1–4.

[35] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "SSIM-motivated two-pass VBR coding for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2189–2203, Oct. 2017.

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

[37] Z. Zhang, T. Zeng, X. Yu, and S. Sun, "Social-aware D2D pairing for cooperative video transmission using matching theory," *Mobile Netw. Appl.*, vol. 23, no. 3, pp. 639–649, Feb. 2018.

[38] *HEVC Reference Software Tags (Revision 4998)*. Accessed: Jul. 18, 2016. [Online]. Available: https://hevc.hhi.fraunhofer.de/trac/hevc/browser#tags

[39] C. Rosewarne, K. Sharman, and D. Flynn, *Common Test Conditions and Software Reference Configurations for HEVC Range Extensions*, document JCTVC-P1006, ISO/IEC, ITU-T, San Jose, CA, USA, 2014.

[40] H. Yu, R. Cohen, K. Rapaka, and J. Xu, *Common Test Conditions for Screen Content Coding*, document JCTVC-X1015, ISO/IEC, ITU-T, Geneva, Switzerland, 2016.

[41] V. Baroncini, H. Yu, and R. Joshi, *Draft of Final Report on SCC Verification Test*, document JCTVC-AA0040, ISO/IEC, ITU-T, Hobart, TAS, Australia, 2017.

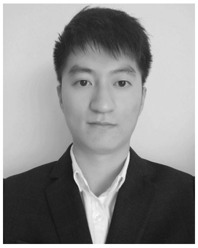
[42] Y. Gong, S. Wan, K. Yang, H. R. Wu, and Y. Liu, "Temporal-layer-motivated lambda domain picture level rate control for random-access configuration in H.265/HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 156–170, Jan. 2019.

[43] G. Bjøntegaard, *Calculation of Average PSNR Differences Between R-D Curves*, document VCEG-M33, ITU-T SG-16, Video Coding Experts Group (VCEG), 2001.

[44] T. Tang, L. Li, and J. Li, "Improved hierarchical quantisation parameter setting method for screen content coding in high efficiency video coding," *IET Image Process.*, vol. 13, no. 8, pp. 1382–1390, Jun. 2019.

TONG TANG received the B.S. and Ph.D. degrees from the Department of Automation, University of Science and Technology of China, Hefei, China, in 2013 and 2018, respectively. He is currently a Lecturer with the Communication Institute, Chongqing University of Posts and Telecommunications. His research interests include video coding, processing, and image processing.

JIN YANG received the B.S. degree from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2018, where she is currently pursuing the M.S. degree with the Communication Institute. Her research interests include video coding and mobile edge computing.



BIAO DU received the B.S. degree from Shenyang Aerospace University, Shenyang, China, in 2018. He is currently pursuing the M.S. degree with the Communication Institute, Chongqing University of Posts and Telecommunications. His research interests include video coding and machine learning.

LIUWEI TANG received the B.S. degree from Chongqing Jiaotong University, Chongqing, China, in 2018. He is currently pursuing the M.S. degree with the Communication Institute, Chongqing University of Posts and Telecommunications. His research interests include video coding and big data.