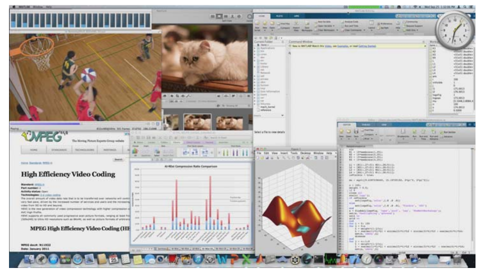
A Novel Rate Control Scheme for Video Coding in HEVC-SCC

Hao Yang, Liquan Shen, Yang Yang, and Weiyao Lin

***Abstract*—The particular characteristics of contents generated by computers and the frequent scene change in screen content videos make the exiting R-λ model in the rate control be not suitable to encode screen content videos. A novel rate control scheme designed for screen content videos coding (SCC) is proposed in this paper. Text blocks, screen image blocks and nature image blocks in screen content videos can result in different bitrate-distortion (R-D) relationships. A content-based rate control method is put forward at Coding Tree Unit (CTU) level for this reason. Three independent parameter update modes are adopted for three different types of CTUs according to their corresponding R-D relationship. Furthermore, frequent scene change in screen content videos needs a novel bit allocation scheme for precise rate control. In view of this, the frames in the screen content videos are classified into scene changed frames and scene unchanged frames. Different bit allocation methods are adopted for different types of frames on the basis of inter frame and intra frame complexity. Besides, a region level bit allocation algorithm considering the inter frame continuity and the numbers of different types of CTUs is added between frame level and CTU level bit allocation schemes. Experimental results show that our proposed rate control algorithm respectively achieves 0.88 dB BDPSNR and 1.54 dB BDPSNR increase under the Low Delay coding structure on average, compared with the default rate control algorithm with hierarchical and non-hierarchical bit allocation in the HEVC-SCC.**

***Index Terms*—Screen content videos, rate control, parameter update mode, frequent scene change, bit allocation.**



(a)

(b)

Fig. 1. The examples of screen content videos. (a) remote desktop; (b) animation.

I. INTRODUCTION

WITH the explosive development of mobile devices, wireless technology, cloud computing and network experiences including video communication, remote desktop, video conference, animation [1], etc., have become more and more indispensable in daily life. These applications introduce a new kind of video named screen content video (SCV) which is shown in Fig. 1. SCV is generated by computers which possesses the characteristics of limited distinct colors, less sensor noises, sharper edges and wider energy distribution compared with traditional nature videos. Usually, it contains texts, graphics and camera-captured nature images. These peculiarities, if properly leveraged, would offer opportunities for great improvements in compression performance over High Efficiency Video Coding (HEVC) standard [2] which is specifically designed for nature videos coding. In the past few years, the popularity of SCVs makes the screen content video coding a hot research topic. The HEVC Screen Content Coding extension (HEVC-SCC) [3] is developed by the Joint Collaborative Team on Video Coding (JCT-VC) [4], which is combined with ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Expert Group (MPEG) to enhance the coding performance of SCVs. Contraposing characteristics of Screen Content Videos, several new coding tools, such as Intra-Block Copy (IBC) [5], Palette Mode (PLT) [6], Adaptive Color Transform (ACT) [7] and Adaptive Motion Vector Resolution (AMVR) [8], are added in HEVC-SCC. These new tools make it possible for HEVC-SCC

Manuscript received April 25, 2019; revised September 3, 2019; accepted September 10, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61671282 and Grant 61931022, through the Shanghai Pujiang Program under Grant 15pjd015, Shanghai Science and Technology Innovation Plan under Grant 18010500200, and Shanghai Shuguang Program under Grant 17SG37. *(Corresponding author: Liquan Shen.)*

H. Yang and Y. Yang are with the Key Laboratory of Specialty Fiber Optics and Optical Access Networks, School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China (e-mail: aidoneus@shu.edu.cn; judith_hi@163.com).

L. Shen is with the Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200072, China (e-mail: jsslq@shu.edu.cn).

W. Lin is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wylin@sjtu.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2019.2954062

to achieve more than 50% bitrate saving over the standard HEVC on average [9]. Apart from the improvement of prediction accuracy and complexity optimization [10]–[11], rate control (RC) for SCV also draws great attention of researchers.

RC is a critical operation in video coding, which is aimed at guaranteeing high quality and steady video propagation within the limited transmission bandwidth. RC is conducted at three levels: group of pictures (GOP) level, frame level and basic unit level. The basic unit is the minimal unit in RC, and in Advanced Video Coding (AVC), the basic unit is Macro Block. In HEVC, the basic unit is represented by Coding Tree Unit (CTU). Generally, the most rate control algorithms includes two steps. One of them is the bit allocation module in which rational bits are assigned to each coding level. The other one is the parameter update module where more accurate parameters are determined. Since the limitation of transmission bandwidth is an intractable issue, it is necessary to conduct the rate control while coding videos.

With the improvement of video coding standard, the mainstream RC algorithm developed. MPEG-2 [12] applies TM5 [13] RC scheme. H.263 [14] applies TMN8 [3], and VM8 [16] is utilized in MPEG-4 [17]. The rate control methods applied in H.264/AVC [18] and H.265/HEVC are specified in [19]–[20] respectively. Usually, these methods are based on three important models: R-Q model [21]–[23], R-ρ model [24]–[25] and R-λ model. R-Q model cannot solve the chicken and egg dilemma, and R-ρ model depends on quantized transform coefficients. Compared with these two models, R-λ model is more appropriate to code high definition and ultra-high definition videos. Therefore, this model has been adopted as the default RC algorithm in HEVC. Li *et al.* [26] proposed a λ-domain RC model based on the bitrate R and the Lagrange multiplier λ which represents the slope of the R-D curve, and established the linear relationship between the Quantization Parameter (QP) and the logarithm of λ in [27].

Basing on the R-λ model, several advanced RC algorithms are put forward. In [28], Sum of Absolute Transformed Difference (SATD) is utilized to replace Mean Absolute Difference (MAD) to guide the bit allocation scheme at the CTU level which improves the accuracy of intra frame rate control. A two-pass method is proposed in [29]. A fast encoder with limited set of coding tools is firstly used to obtain the data used for bit allocation and model parameter initialization, and in the second pass, an adaptive QP adjustment framework is proposed to update parameters. The rate control algorithm in [30] classifies all the CTUs into three types according to their coding modes and R-D relationships, and then a mixed rate control model is raised for different types of CTUs. In the work [31], a learning-based initial QP method is proposed to replace the convention calculative method which improves the rate control performance. Guo *et al.* [32] proposes an optimal bit allocation at frame level and GOP level by utilizing the information of encoded frames in the previous GOP and the temporal R-D dependency among different GOPs. Some works of RC aim at improving the subjective quality. Generally, they can be divided into two main ideas. The first one is to take the visually relevant indicators like Structural Similarity Index Measure (SSIM) instead of the original distortion calculated by Mean Square Error (MSE) as the parameter in the Rate Distortion Optimization (RDO) [33]. The second one is taking the bit allocation scheme on the basis of the saliency. More bits are assigned to the region with the high attention and less bits are allocated to the area with small salience [34]–[37]. Except for the conventional rate control optimization method, a novel approach based on the Convolutional Neural Network (CNN) is proposed in [38] to estimate the model parameters to improve the accuracy.

The existing RC methods achieve significant RC performance for natural contents. However, The computer generated SCV have the characteristics of less sensor noises, limited distinct colors, repeated patterns and sharp edges compared with NCV. These unique features make the relationship between the coded bits and distortion in SCV is different from NCV, which results in performance degeneration when directly applying the natural RC methods for screen content videos. Therefore, more efficient rate control methods are urgent needed for SCC.

Some RC schemes for SCVs are put forward recently. The work in [39] classifies the frames in the SCVs into three categories: stationary frames, continuous frames and abrupt frames according to their disparate complexity, and then assigns bits for them respectively. On the basis of the research in [39], a more reasonable bit allocation strategy considering the buffer overflow and underflow is proposed by designing a pre-analyzer where the information of the proceeding frames is collected in [40]. A flexible slide window is adopted in [41] to take place of the GOP with the fixed size to handle the problem that the complexity of the adjacent frames in SCVs is not continuous. In [42], each frame in the SCV is determined as a key frame or a non-key frame depending its inter-frame correlation. A modified R-Q model is employed to allocate bits and adjust parameters for the key frame and the non-key frame, respectively. In conclusion, most of the existing RC schemes for SCVs are designed according to the feature of a whole frame, but ignore the characteristics in the video content. Additional performance can be achieved by jointly considering inter frame and intra frame characteristics.

In this paper, the discriminations of screen content videos and conventional natural videos on color, texture and structure has been taken fully consideration. For the reason that the video content generated by computers contains flat blocks with the single color, blocks with text and blocks with common natural images, the relationships of bitrate (R) and distortion (D) of different video contents may be various. A content-based rate control scheme is proposed in this paper. The contributions of our work are summarized as follows:

(1) Since the fact that R-D relationships of Text (T), Screen Image (SI) and Nature Image (NI) appear various peculiarities, the original unitive R-λ model is not suitable for SCVs any more. A content-based and independent parameter update method is proposed for rate control at the CTU level.
(2) Scene change is detected according to the frame complexity and the number of the same type of CTU. The relationship between frame complexity and bits assumed

to encode a frame is established to guide bit allocation at the GOP level and the frame level.

(3) For the reason that different types of CTU adopt different R-λ models, a bit allocation scheme for the region containing the same type of CTUs is needed. Thus, a region level bit allocation scheme is added between the frame level and the CTU level bit allocation. The bits allocated to text region, SI region and NI region are determined according to the coding distortion and CTU number. While the scene change occurs, the region level bit allocation method only depends on the number of CTUs.

The rest of this paper is organized as follows. Section II makes an analysis of the R-D relationships of different video contents, and the proposed content-based parameter update scheme is introduced in detail. In Section III, the bit allocation approach based on the inter-frame and intra-frame complexity is proposed. The experimental results and the conclusion of this paper are presented in Section IV and Section V respectively.

II. Analysis on the R-D Relationship of SCVs and the Proposed Content-Based Parameter Update Method

RC is designed to minimize the distortion D at the given target bits R_t, which can be written as

$$min\ D \qquad s.t.\ R \le R_t, \tag{1}$$

where R is the practical bits consumed to compress the video. The relationship between R and D is significant to the establishment of RC models. The typical hyperbolic R-D model is described as

$$D = cR^{-k}, \tag{2}$$

Basing on the R-D model, the relationship between λ and R is established in work [26] which is named as the R-λ model and adopted in HEVC-SCC.

$$\lambda = -\frac{\partial D}{\partial R} = ckR^{-k-1} = \alpha R^{\beta}, \tag{3}$$

where λ is the slope of R-D curve, α and β are parameters associated with video contents. However, this model is built to encode conventional nature videos which are diverse from screen content videos in many ways, such as coding methods, image complexity and so on. A hypothesis, that the R-λ model will not fit the new video any more, is made in this paper, and an experiment is designed to check it out. In the experiment, videos are encoded with QPs set as 22, 27, 32 and 37, and the fitting R-D curves are drawn in Fig. 2. R and D are expressed by bit per pixel (*bpp*) and Mean Square Error (*MSE*) of Y component respectively. Obviously, in spite of that R-D relationship of SCV still satisfy the typical hyperbolic R-D model [43]–[44], the coefficients of the power function are quite different and the gap is more than 10 times and even up to 10000 times for some videos. Additionally, λ-*bpp* curves can be fitted in the same way, and there is a huge discrepancy among the parameters α and β in λ-*bpp* curves of different kinds of videos.

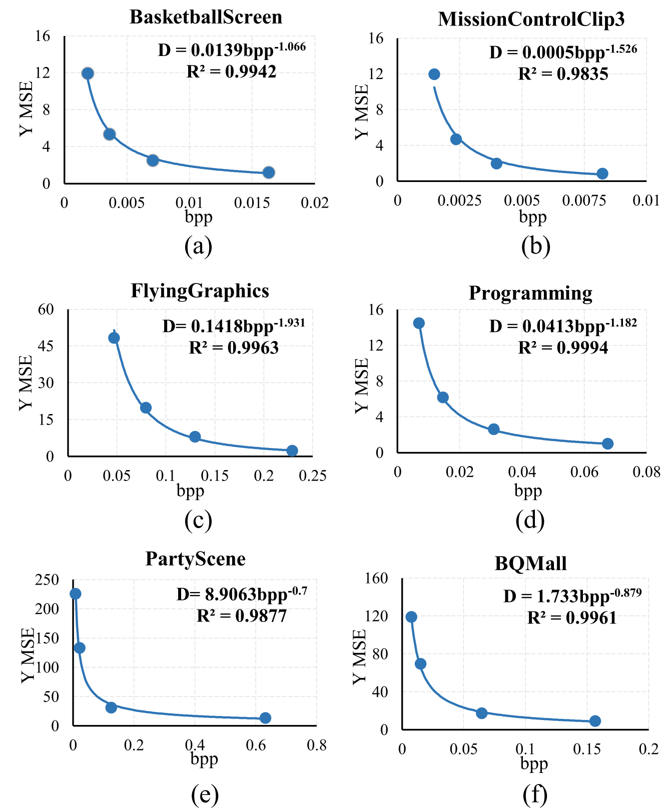


Fig. 2. R-D curves. (a), (b) mixed content videos; (c), (d) pure computer-generated content videos; (e), (f) pure camera-captured content videos.

Since the R-D relationship of SCVs changes, a hypothesis is made like that the R-λ model will not match this video any more and needs modification. Another experiment is made to confirm the hypothesis. First of all, the CTUs in a frame of a SCV is sorted into three types: T-CTUs, SI-CTUs and NI-CTUs. After that the fitting λ-*bpp* curves of three kinds of CTUs are obtained. In general, textural region has monotonous color with sharp texture edges compared with pictorial regions. Therefore, local mean activity measure (LMAM) is utilized to detect textural CTUs (T-CTUs). To calculate the LMAM, we first define the activity of the pixel $P_{i,j}$ at position $(i,\ j)$ as

$$A_{i,j} = \theta A1_{i,j} + (1-\theta)A2_{i,j} \tag{4}$$

$$A1_{i,j} = \Big(\big(P_{i,j} - P_{i-1,j-1}\big)^2 + \big(P_{i,j} - P_{i+1,j-1}\big)^2 + \big(P_{i,j} - P_{i-1,j+1}\big)^2 + \big(P_{i,j} - P_{i+1,j+1}\big)^2\Big)^{1/2} \tag{5}$$

$$A2_{i,j} = \Big(\big(P_{i-1,j-1} - P_{i+1,j+1}\big)^2 + \big(P_{i-1,j+1} - P_{i+1,j-1}\big)^2\Big)^{1/2} \tag{6}$$

where the θ is set to 0.5 according to [47]. $A1$ and $A2$ are calculated as local activities for different directions. Furthermore, the LMAM is defined as

$$LMAM = \frac{1}{W \times H}\sum_{i=1}^{H}\sum_{j=1}^{W} A_{i,j} \tag{7}$$

where W and H are the width and height of the region. $A_{i,j}$ demonstrates the activity value at position $(i,\ j)$.

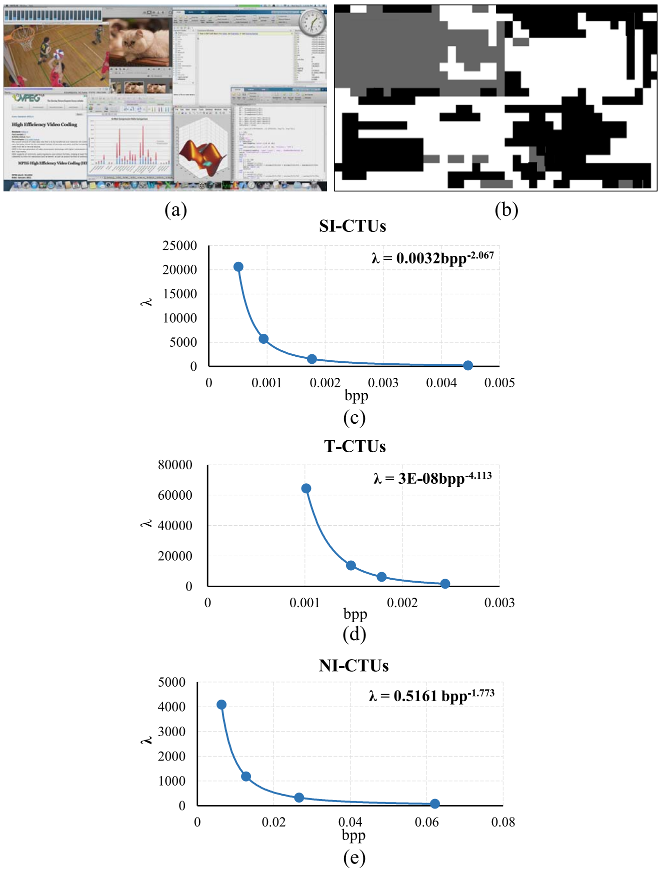


Fig. 3. Different video contents and their corresponding λ-*bpp* curves.

To recognize the screen image (SI) CTU and natural image (NI) CTU in pictorial regions, energy distribution is explored for these two kinds of contents. Discrete Cosine Transform (DCT) is utilized to obtain frequency information. For NI, most energy concentrates on the low frequency area, while the energy in SI mostly distributes in the medium-high frequency area. After DCT transform, the ratio (R_{low}) of sum of low frequency (1010 samples at the above-left corner) coefficients to the whole frequency coefficients of the CTU are generated as the measurement to recognize SI-CTUs and NI-CTUs.

$$R_{low} = \sum_{i=1}^{10}\sum_{j=1}^{10} C_{i,j} / \sum_{i=1}^{64}\sum_{j=1}^{64} C_{i,j} \tag{8}$$

where $C_{i,j}$ is the DCT coefficient at position (i, j).

The segmentation results and λ-*bpp* curves are exhibited in Fig. 3. In Fig. 3(b), the white blocks are SI-CTU, NI-CTUs are colored in gray, and the remaining black ones present T-CTUs. It can be seen that the disparity in α and β of different λ-*bpp* curves is also too great to be ignored. Therefore, if adjacent CTUs belong to different CTU types, the parameters α and β utilized to encode the current CTU are not suitable to update the parameters for next CTU. Therefore, a reasonable parameter update strategy is required.

On basis of the analysis above, the original R-λ model can not describe the R-λ relationship of various video contents. Thus, a mixed RC model is set up to solve this problem.

$$\lambda_{\Theta} = \alpha R_{\Theta}{}^{\beta}, \quad (\Theta = T, SI, NI), \tag{9}$$

where Θ denotes the CTU type including T-CTU, SI-CTU and NI-CTU. In this model, the parameters α and β need to be updated separately. That is to say, the renewed parameters α and β which are calculated by current kind of CTU only can be adopted to encode the next same kind of CTU rather than the CTU following unless the following CTU belongs to same kind as the current CTU. The detail process is expressed as follows.

$$\begin{aligned} \lambda_{\Theta comi} &= \alpha_{\Theta i} bpp_{acti}^{\beta_{\Theta i}}, \\ \alpha_{\Theta i+1} &= \alpha_{\Theta i} + \delta_{\alpha} \times (ln\lambda_{acti} - ln\lambda_{\Theta comi}) \times \alpha_{\Theta i}, \\ \beta_{\Theta i+1} &= \beta_{\Theta i} + \delta_{\beta} \times (ln\lambda_{acti} - ln\lambda_{\Theta comi}) \times lnbpp_{acti}, \\ &(\Theta = T, SI, NI) \end{aligned} \tag{10}$$

In foregoing formulas, bpp_{acti} and λ_{acti} stand for the consumed actual bits and actual λ used to code the i-th CTU belonging to Θ-CTU. δ_{α} and δ_{β} are fixed values inherited from work [26] (i.e., 0.1 and 0.05 separately) applying to all CTUs like bpp_{acti} and λ_{acti}. Here we set T to Θ as an example to explain formulas above. α_{Ti} and β_{Ti} are the parameters used to code T-CTUs. The estimated λ_{Tcomi} is computed with formula (4) for the i-th T-CTU. Besides, α_{Ti+1} and β_{Ti+1} represent the updated α_{Ti} and β_{Ti} for the $(i+1)$-th T-CTU. Parameters used to code other types of CTUs can be deduced from this. In this experiment, the initial values of α_{Ti} and β_{Ti} are set to 0.0085 and −2.5341, α_{SIi} and β_{SIi} are set to 0.0872 and −1.9422, and 0.3161 and 1.7730 are valued to α_{NIi} and β_{NIi} by averaging the coefficients and exponents of three kinds of λ-*bpp* curves severally. The proposed update modes of parameters ensure the accuracy of the R-λ model for SCVs containing one or one more constituents generated by computer.

III. Analysis on the R-D Relationship Between Coding Bits and Video Complexity

The bit allocation scheme is a significant part in the rate control algorithm. Generally, in order to obtain stable coding quality, more bits should be allocated to the frames with larger spatial and temporal complexity. In SCVs, frequent scene or content variation is a great characteristic, which increases the temporal complexity of SCVs. In this section, the mean square error (MSE) between two adjacent frames is utilized to demonstrate the inter-frame complexity. MSE is a simple and effective measurement which reflects the difference between the adjacent frames caused by motion and scene change. The *MSE* between the i-th frame and the $(i-1)$-th frames is calculated as,

$$MSE_i = \frac{1}{H \times W} \sum_{n=1}^{H}\sum_{m=1}^{W} (P_i(n, m) - P_{i-1}(n, m))^2 \tag{11}$$

where $P_i(n, m)$ and $P_{i-1}(n, m)$ represent the pixel values at the position (n, m) in i and $i-1$ frame, respectively. Parameters H and W are the height and the width of a frame.

An experiment is conducted on the test model for screen content video coding (HM-16.10+SCM-8.0) with the fixed QP

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

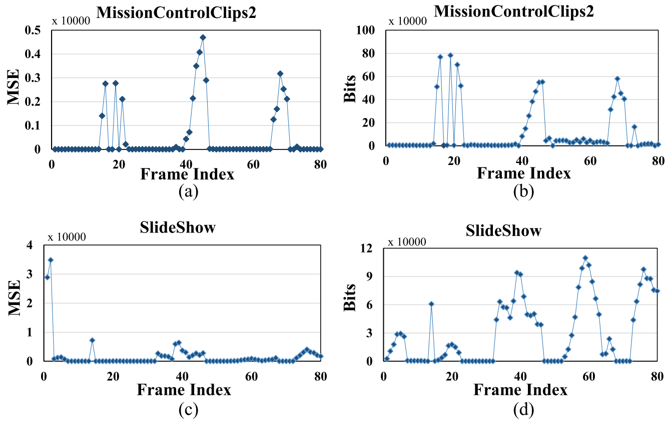


Fig. 4. The distribution of MSEs and bits consumed to encode a fame.

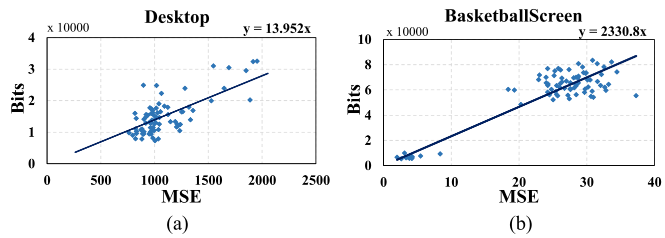


Fig. 5. The R-MSE curves of scene unchanged frames.

to analyze the relationship between MSE and bits consumed to encode a frame. The initial QP in the experiment is set to 22 and the sequences *MissionControlClips2* and *SlideShow* with 160 frames are selected for investigation. These sequences are classical SCVs. In order to enable all test frames be encoded in the same condition, 80 frames encoded with the final QP 25 are chosen to show the distribution of MSE and the ultima coding results which are drawn in Fig. 4.

It can be seen from Fig. 4 that the bits consumed to encode a frame are related to its MSE value. Two conclusions can be obtained from this figure: 1) the MSE of a frame increases a lot when the scene changes abruptly, and the scene changed frame (SCF) needs much more bits to enable the stable video quantity. 2) the bits utilized to encode the SCFs are even hundreds of times more than that used to encode the scene unchanged frames (SUFs). These two points infer that frequent scene change should be taken into consideration when the bit allocation scheme is conducted to SCVs, and a new bit allocation scheme is needed.

The positive correlation between the MSE of the SCF and bits used to encode it can be seen in Fig. 4. Furthermore, we also investigate the relationship between the MSE and bits of the SUFs, and the results are presented in Fig. 5. It can be seen that the amount of bits consumed to encode the SUFs increases while the MSE value increases. A linear relationship can be obtained as,

$$R = k \times MSE, \tag{12}$$

where R represents the coding bits. k is a parameter related to practical coding.

The coding results of SUFs with different fixed QPs are shown in Fig. 6. It can be seen that, the parameter k in the

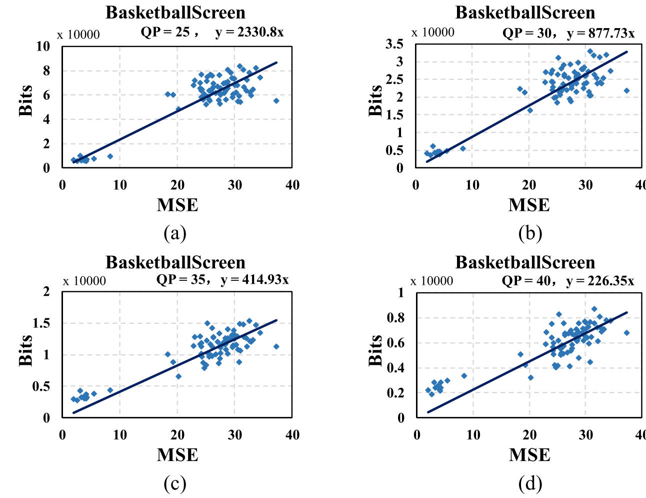


Fig. 6. The R-MSE curves for scene unchanged frames under different QPs.

linear relationship is different when the QP is different, so is the bit R. In our experiment, the bitrate obtained by fixed QP coding is set as the target bitrate for rate controlling. Thus, a fixed QP can correspond to a target bitrate. On basis of this fact, the correlation between the parameter k and QP can be replaced by the correlation between k and the target bpp, and the relationship can be described as

$$k = a \times bpp + b \tag{13}$$

where bpp is used to represent the target bitrate for the video coding. a and b are parameters related to video content. The content related parameters a and b are determined by both inter complexity and intra complexity. In the experiment, the inter complexity is measured by average MSE of all encoded SUFs which is denoted by C_{inter}, and the intra complexity C_{intra} is represented by the average local variance of pixels in the current frame.

$$C_{intra} = \frac{\sum_{i=1}^{N} \sum_{n=1}^{64} \left(\sum_{m=1}^{64} P_i(n,m) - \bar{P}_i \right)^2}{N} \tag{14}$$

where the average local variance C_{intra} is calculated for the CTU with the size of 64×64. $P_i(n,m)$ is the pixel value at the position (n, m) of the i-th CTU in a frame, and $\bar{P}_i$ is the average value of all pixels in the i-th CTU. N represents the total number of CTUs in a frame. On the basis of the experiment analysis, the relationship between parameters a, b and inter-complexity C_{inter}, intra-complexity C_{intra} can be written as

$$f(C) = e^{\frac{150}{\sqrt{C_{intra}}}} + \frac{1000}{logC_{inter}} \tag{15}$$

$$a = 88052 \times logf(C) - 428791 \tag{16}$$

$$b = -0.974 \times f(C) + 203.062 \tag{17}$$

Therefore the relationship between inter and intra complexity of the video and bits consumed to encode it can be established according to formulas (12)-(17). The more reasonable bit allocation scheme based on the relationship above will achieve better coding performance.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IV. Bit Allocation Approach Based on Inter-Frame and Intra-Frame Complexity

Frequent scene change is an important characteristic of SCVs, and quantities of bits consumed to encode SCFs and SUFs are quite different. For this reason, the bit allocation scheme for SCVs needs to be respectively conducted for the two types of frames. It is necessary to detect the scene change before allocating bits for each coding level.

A. Scene Change Detection

The scene change detection result is significant for bit allocation methods at GOP level and frame level. Two factors are adopted to detect scene change. The first one is MSE_i which is calculated in formula (6). The threshold of MSE_i which is used to determine the frame type is set as 2500 or 100 times of average MSE of all encoded SUFs (100avgMSE). That is to say, if the MSE_i is larger than 2500 or 100avgMSE, the i-th frame can be determined as SCF, otherwise, the frame is SUF.

Apart from MSE, the ratio of the same kind of CTUs in the current frame and the last frame is also adopted to detect scene change. The ratio is calculated as

$$Ra_i = \frac{Num_{Maxi}}{Num_{Lasti}} \tag{18}$$

where Num_{Maxi} is the amount of the type of CTUs which occupies the largest areas in the current frame. Num_{Lasti} represents the number of the same kind of CTUs in the last frame. If Ra_i is larger than 1.3 or smaller than 0.7, the i-th frame can be determined as SCF, otherwise, the frame is SUF.

B. GOP Level Bit Allocation

In the practical coding processing, the first frame of a video is special for it does not need referent frames. In generally, the size of the first GOP is set as 1. That is to say, the first frame is the first GOP. The size of the remaining common GOP is 4 as set in the configure file. The average bits are assigned to the first frame which is calculated as

$$R_1 = \frac{R_{Total}}{F_{Total}} \tag{19}$$

where R_1 is the target bits allocated to the first frame of the video. R_{Total} and F_{Total} are the target bitrate and the frame rate of the video.

The number of SCFs is important to be taken into consideration when allocating bits for a common GOP with the size of 4. Briefly, if there are more SCFs in a GOP, more bits should be allocated to the GOP. The detail allocation method is introduced in the formula

$$R_{GOP} = \frac{R_{Left}}{F_{Left} + N_{SCF}} \times (S_{GOP} + N_{SCF}) \tag{20}$$

where R_{GOP} is the target bits assigned to the current GOP. R_{Left} and F_{Left} are left bits and frames after encoding the last GOP, respectively. N_{SCF} represents the number of SCFs in the current GOP which is obtained accord to the scene change detection method. Besides, S_{GOP} is the size of the GOP. In terms of this bit allocation approach, more bits are allocated to the GOP with more SCFs, which ensures the coding quality.

C. Frame Level Bit Allocation

In SCVs, the scenes change frequently. The frames in SCVs are classified into SCFs and SUFs, and two different bit allocation schemes are adopted to encode the SCF and the SUF.

As shown in Fig. 4, both the values of MSE and bits consumed to code a frame change abruptly while the scene change happens. Moreover the variation tendencies of the two parameters are conformably. In view of this feature, the determination of target bits for SCFs is related to the ratio of MSE of each frame in the GOP. So, if the i-th frame in the current GOP is a SCF, the allocated target bits R_{ti} is computed as

$$R_{ti} = \frac{MSE_i}{\sum_{i=1}^{S_{GOP}} MSE_i} \times R_{GOP} \tag{21}$$

where MSE_i represents the MSE of the i-th frame in the GOP.

If the i-th frame in the current GOP is a SUF, according to the analysis of the relationship of MSE and coding bits in Section III, the target bits R_{comi} can be computed as

$$R_{comi} = k \times MSE_i \tag{22}$$

where k is a parameter related to target bits, inter complexity and intra complexity, and the detail calculation method refers to formulas (12)-(17). However, there is a great difference between the target bits R_{comi} and consumed bits in the practical coding processing. For, SUFs, the variation tendencies of R_{comi} and practical coding bits are in good agreement. So, the allocated target bits for the SUF in the GOP, R_{ti}, is computed as

$$R_{ti} = \frac{R_{comi}}{\sum_{i=1}^{n_{SUF}} R_{comi}} \times (R_{GOP} - R_{SCF}) \tag{23}$$

where n_{SUF} represents the number of SUFs in the current GOP. R_{SCF} is the sum of preset target bits of all SCFs in the GOP. In general, more bits are allocated to SCFs via the bit allocation method at the frame level proposed above, which keeps the stability of coding quality even the scene changes. Meanwhile, it is more reasonable to allocate the coding bits for SCVs by considering both inter and intra complexity.

D. Region Level Bit Allocation

Since the various characteristics in three different kinds of CTUs result in different R-D relationships, the bit allocation schemes for regions containing different CTUs should be designed respectively. For SUFs, the region including more CTUs should be assigned more bits, and vice versa. Additionally, much more complex texture and sharper edges in T-CTU and NI-CTU make the regions containing these two CTUs need more bits to keep coding quality. The preset bits for different regions according to the CTU number is calculated as

$$\begin{aligned} R_{P\Theta-SCF} &= \omega_\Theta \times \frac{Num_\Theta}{Num_{Total}} \times R_{Pic}, \\ \omega_\Theta &= \begin{cases} 1.1 & \Theta = T \\ 1.2 & \Theta = NI \\ 0.8 & \Theta = SI \end{cases} \end{aligned} \tag{24}$$

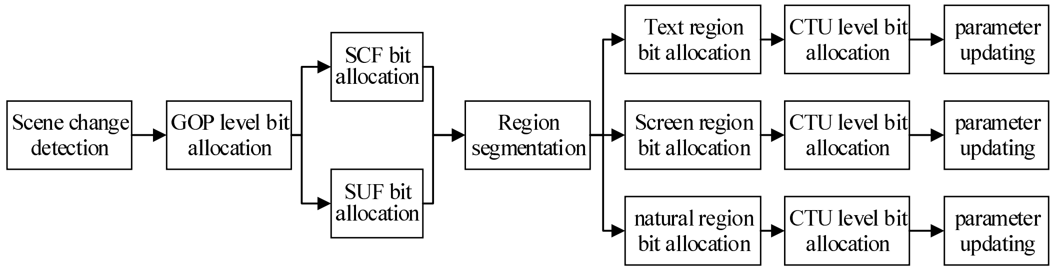


Fig. 7. The flowchart of the proposed RC algorithm.

where Num_Θ refers to the number of Θ-CTUs in the current frame, and Num_{Total} refers to the total number of all CTUs in the frame. R_{Pic} is the target bits of the frame which is calculated through the formula (21) or (23). Then, we can get the total preset bits ($R_{PTotal-SCF}$) for a frame based on three regions as

$$R_{PTotal-SCF} = \sum_{\Theta} R_{P\Theta-SCF} \quad \Theta \in \{T, SI, NI\} \tag{25}$$

The final target bits ($R_{\Theta-SCF}$) allocated to each region for SCF can be expressed as

$$R_{\Theta-SCF} = R_{P\Theta-SCF} + \frac{Num_\Theta}{Num_{Total}} \times (R_{Pic} - R_{PTotal-SCF}) \tag{26}$$

In general, the refinement bits are added to the preset bits to eliminate the bits error for a whole SCF.

For SUFs, the contents in the adjacent frames is continuous. If the region in the last frame generates more distortions, more bits should be allocated to the same type of the region to reduce the propagation of distortion. As for the region with less distortion, reasonable reduction of bits will not generate many distortions. Based on the distortion situation in the last frame and the number of CTUs in the current frame, the bit allocation scheme at region level for SUFs can be designed as follows,

$$R_{\Theta-SUF} = \omega_d \times \frac{D_{\Theta_{Last}}}{D_{Total_{Last}}} \times R_{Pic} + (1 - \omega_d) \times R_{\Theta-region} \tag{27}$$

where $R_{\Theta-SUF}$ is the target bits for the region only containing Θ-CTUs in SUFs. ω_d is the weight to combine the influence of the propagation of temporal distortion and the video contents, which is set to 0.6 in our experiments. $D_{\Theta_{Last}}$ represents the sum of the distortion generated by encoding all Θ-CTUs in the last frame, and $D_{Total_{Last}}$ is the total distortion of the last frame. $R_{\Theta-region}$ is a target bit component calculated according to the number of Θ-CTUs, and its calculation approach is the same as that in the formulas (24)-(26).

E. CTU Level Bit Allocation

At CTU level, the bits allocated to each CTU are in connection with the target bits of the region containing Θ-CTUs. According to the R-λ model, the preset target bits for the i-th Θ-CTU are calculated as

$$r_{\Theta pi} = \left(\frac{\lambda}{\alpha_{\Theta i}}\right)^{\frac{1}{\beta_{\Theta i}}} \tag{28}$$

TABLE I
IMPORTANT CODING PARAMETERS USED IN OUR EXPERIMENTS

GOP size	4
CTU size	64 × 64
Maximum CU depth	4
Hierarchical bit allocation	1/0
Frames to be encoded	160
Video format	YUV/RGB

where the Lagrange multiplier λ is computed at the frame level. Considering the target bits for the Θ-region, the bits assigned to the i-th Θ-CTU are modified as

$$r_{\Theta i} = \frac{r_{\Theta pi}}{\sum_{i=1}^{Num_\Theta} r_{\Theta pi}} \times R_\Theta, \; (R_\Theta = R_{\Theta-SCF}, R_{\Theta-SUF}) \tag{29}$$

Similarly, Θ is set as T, SI or NI.

F. The Overall Flowchart of the Proposed RC Method

To better explain the process of the proposed SCC rate control algorithm, the overall SCC rate control framework is presented in Fig. 7. For an input video sequence, scene change detection will be conducted first for each GOP. Each frame of GOP is categorized into either the scene changed frame (SCF) or the scene unchanged frame (SUF). Then, GOP level bit allocation is performed to get the target bits of each GOP. Based on the frame category (SCF or SUF), the GOP level target bits are allocated to each frame. Before CTU level bit allocation, frames are divided into three parts, including T-CTUs, SI-CTUs and NI-CTUs. Then, the target bits for each region and CTU are determined. Parameter updating for different contents are conducted at CTU level, respectively.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Several experiments are conducted to test the efficiency of the rate control algorithm proposed in this paper. Firstly, the experiment setup is introduced in Section V-A, and then both objective quality and subjective quality are verified in the remaining sections.

A. Experiment Setup

All rate control algorithms tested in our experiment are conducted in the HEVC-SCC test model HM-16.10+SCM-8.0, and Low Delay is selected as the coding structure. Some important coding parameters are shown in Table I.

TABLE II
AVERAGE BDPDNR AND BDBR OF OUR PROPOSED RC ALGORITHM COMPARED WITH SCM-8.0-NH AND SCM-8.0-H

Video Format	Sequence	Ours vs SCM-8.0-NH		Ours vs SCM-8.0-H	
		BDPSNR(dB)	BDBR(%)	BDPSNR(dB)	BDBR(%)
YUV	BasketballScreen	2.25	-25.14	0.84	-9.97
	Console	-0.31	1.09	0.79	-4.25
	Desktop	2.72	-15.93	0.65	-4.16
	FlyingGraphics	0.50	-5.15	0.28	-2.81
	Kimono1	0.04	-2.36	0.19	-7.79
	Map	1.48	-19.70	0.49	-5.76
	MissioncontrolClip2	0.89	-12.02	0.33	-4.87
	MissioncontrolClip3	2.85	-25.43	1.45	-14.17
	Programing	2.28	-33.06	1.20	-19.07
	Robot	1.25	-37.00	0.66	-21.37
	SlideShow	3.98	-38.82	0.51	-7.41
	Web_browsing	2.16	-19.29	0.86	-6.95
RGB	Console	0.74	-3.04	0.04	-0.23
	Desktop	0.01	-0.02	1.10	-6.06
	FlyingGraphics	-0.35	3.52	-0.04	0.18
	Kimono1	0.68	-24.76	0.52	-19.05
	Map	2.40	-23.27	4.85	-40.80
	Programing	2.38	-38.52	2.00	-35.56
	Robot	0.90	-24.44	0.36	-12.14
	SlideShow	3.22	-25.11	1.09	-0.14
	Web_browsing	2.30	-14.73	0.39	-3.65
Average		**1.54**	**-18.25**	**0.88**	**-10.76**

Several rate control algorithms are chosen to verify the efficiency of our proposed method. In order to facilitate the description, the name of all rate control algorithms are simplified. Firstly, if *Hierarchical bit allocation* in the test model HM-16.10+SCM-8.0 is set as 1, the rate control method is named as "SCM-8.0-H". In this method, the bit allocation scheme at frame level is conducted with the ratio introduced in [26]. While *Hierarchical bit allocation* in the test model HM-16.10+SCM-8.0 is set as 0, the method is named as "SCM-8.0-NH", and the target bits for each frame are allocated equally. The approach proposed in our paper is denoted as "Ours". The test video sequences in this paper are recommend by JCT-VC. All the videos belong to four video classifications, and they are "text and graphics with motion (TGM)", "mixed content (M)", "animation (A)" and "camera-captured content (CC)" [45]. In order to guarantee the video coding quality within a wide range, the video is encoded with four fixed QPs. The four fixed QPs are set as 22, 27, 32 and 37, respectively. The final bitrate obtained by coding with the fixed QP is set as the target bit rate.

B. Coding Performance Comparison Between Ours and Default Rate Control Methods in the Test Model

1) R-D Performance: The R-D performance is one of the most objective evaluation indexes for the rate control algorithm. In this section, two evaluation indicators, BDPSNR and BDBR, are utilized to demonstrate the R-D performance of all the rate control algorithms. The BDPSNR of our proposed method which is calculated by comparing with the algorithms SCM-8.0-NH and SCM-8.0-H is shown in the Table II. In comparison with SCM-8.0-NH, the BDPSNR of Ours is 1.54 dB. Besides, taking the algorithm, SCM-8.0-H with hierarchical bit allocation, as an anchor, the mean BDPSNR of Ours is 0.88 dB. In conclusion, our proposed rate control method is not only better than the SCM-8.0-NH, but also can obtain better performance than the algorithm SCM-8.0-H. Moreover, it can be seen from this table that the performance of SCM-8.0-H is better than that of SCM-8.0-NH. This result indicates that the bit allocation scheme with reasonable ratio at frame level will lead to better coding performance than the equal bit allocation approach. Besides, these data also verify that the optimization of the bit allocation via making the most of the inter frame relationship can improve the coding efficiency.

In the experiment, BDBR which represents the reduction of bitrate is adopted to compare the R-D performance of different rate control algorithms. The BDBRs of our proposed method are exhibited in the Table II. Obviously, our rate control algorithm can save some bitrate when compared with SCM-8.0-NH and SCM-8.0-H. Taking the algorithm SCM-8.0-NH as an anchor, the average BDBR of all 21 videos encoded with Ours is -18.25%. Besides, in comparison with R-λ model with unequal bit allocation, the average BDBR of Ours is -10.76%. The saving of the bitrate proves the superiority of our proposed rate control scheme. The experimental results also confirm that our method can improve the utilization of bandwidth to some degree.

TABLE III
RATE CONTROL ACCURACY COMPAISON FOR DIFFERENT RC ALGORITHMS

Video Format	Sequence	TargetBitrate(kbp)	SCM-8.0-NH		SCM-8.0-H		Ours	
			Bitrate(kbp)	Error(%)	Bitrate(kbp)	Error(%)	Bitrate(kbp)	Error(%)
YUV	BasketballScreen	5573.1	5578.1	0.09	5340.1	4.18	5575.9	0.05
	Console	5053.4	5268.4	4.25	5127.4	1.46	5060.5	0.14
	Desktop	1129.9	1152.7	2.02	1127.9	0.18	1126.9	0.26
	FlyingGraphics	33967.7	33954.8	0.04	33963.7	0.01	33966.2	0
	Kimono1	14534.2	14535.4	0.01	14535.8	0.01	14631.6	0.67
	Map	3646.7	5358.3	46.93	4701.3	28.92	4002.2	9.75
	MissioncontrolClip2	8842.3	9224.4	4.32	9858.1	11.49	8807.7	0.39
	MissioncontrolClip3	1875.7	1874.4	0.07	1794.7	4.32	1874.3	0.08
	Programing	4915.2	4917.0	0.04	4918.4	0.07	4995.2	1.63
	Robot	5147.6	5147.9	0.01	5148.1	0.01	5147.9	0.01
	SlideShow	798.8	805.3	0.81	929.8	16.4	799.3	0.06
	Web_browsing	185.6	196.5	5.89	185.7	0.05	186.1	0.28
RGB	Console	3896.0	3937.9	1.08	3980.6	2.17	3918.8	0.58
	Desktop	1170.2	1200.2	2.56	1178.1	0.67	1229.9	5.09
	FlyingGraphics	48011.0	47977.9	0.07	48013.8	0.01	48022.3	0.02
	Kimono1	9172.9	9173.4	0.01	9174.0	0.01	9173.1	0.00
	Map	1314.8	1396.2	6.19	1399.0	6.40	1256.4	4.44
	Programing	991.9	992.6	0.08	995.8	0.40	999.0	0.72
	Robot	4133.4	4133.7	0.01	4133.9	0.01	4133.9	0.01
	SlideShow	237.5	239.2	0.71	244.1	2.80	244.2	2.83
	Web_browsing	246.5	257.2	4.33	246.3	0.09	251.3	1.92
Average Bitrate Error			—	**3.79**	—	**3.79**	—	**1.38**

2) Rate Control Accuracy: One of the significant goals of rate control is to make the error between the practical bitrate and target bitrate as small as possible. Thus, the rate control accuracy is a key evaluation criterion of the rate control algorithm. The target bitrate, the actual bitrate and the error between them are tabulated in the Table III. The bitrate error, *Error*, is calculated as

$$Error = \frac{|R_a - R_t|}{R_t} \times 100\% \tag{30}$$

where R_a is the actual bitrate obtained in the practical video coding process. R_t is the target bitrate preset before coding. The rate control errors of both the rate control algorithms in the test model HM-16.10+SCM-8.0 with or without hierarchical bit allocation at frame level are 3.79%. The average bitrate error of our algorithm is 1.38%. The comparison in the Table III indicates that the parameter update mode based on the video content and bit allocation method considering the characteristics of SCVs can help reduce the error of rate control and achieve the aim of rate control.

3) Quality Fluctuation: If the difference between the qualities of all frames in the video is large, the visual experience will be influenced. Hence, the quality stability of all frames in the video should be ensured. In this section, the variance of all encoded frames is computed to measure the quality fluctuation. The variance of Y-PSNR is calculated as

$$var = \frac{1}{N}\sum_{i=1}^{N}\left(PSNR_i - \overline{PSNR}\right)^2 \tag{31}$$

where *Var* is the variance. N is the total number of all encoded frames. $PSNR_i$ represents the Y-PSNR value of the i-th frame, and $\overline{PSNR}$ is the average Y-PSNR of all frames. In general, larger *Var* represents larger quality fluctuation and smaller *Var* represents more stable coding quality. Table IV exhibits the values of *Var* of all videos encoded by three different rate control algorithms. For the algorithms SCM-8.0-NH, SCM-8.0-H and Ours, the values of *Var* are 12.29, 9.58 and 9.55 respectively, and our method achieves the smallest variance value. This result indicates that our method reduces the quality fluctuation and enable the visual effects of encoded videos.

C. Coding Performance Comparison Between Ours and Popular Rate Control Algorithms

Apart from the rate control methods adopted in the test model, we also make comparisons between our proposed scheme and other popular rate control algorithms to verify the superiority of Ours. A classical conventional nature videos based rate control method [46] is used for comparison, which denote as DCC_Wen in our experiment. Another algorithm adopted for comparison is a typical rate control scheme designed for SCVs [40], which denote as VCIP_Xiao. As shown in Fig. 8, the coding performance of three algorithms is compared from four aspects. The first one is the average gain of BDPSNR. It can be seen from Fig. 8(a) that the average BDPSNR of Ours is 0.22 dB, larger than that of VCIP_Xiao and 0.80 dB larger than that of DCC_Wen when taking SCM-8.0-H as an anchor. When compared with

TABLE IV
VARIANCE OF Y-PSNR (dB) COMPARISON FOR DIFFERENT RC ALGORITHMS

Video Format	Sequence	SCM-8.0-NH	SCM-8.0-H	Ours
YUV	BasketballScreen	11.35	9.17	6.61
	Console	10.12	13.22	12.92
	Desktop	6.26	24.56	7.70
	FlyingGraphics	1.54	3.55	6.42
	Kimono1	0.14	0.95	0.13
	Map	6.87	7.60	16.96
	MissioncontrolClip2	8.69	9.47	30.22
	MissioncontrolClip3	11.69	16.77	10.80
	Programing	5.39	7.94	5.59
	Robot	0.40	19.72	0.15
	SlideShow	63.41	11.87	23.04
	Web_browsing	14.37	12.65	10.82
RGB	Console	8.62	4.67	5.91
	Desktop	6.49	4.78	8.37
	FlyingGraphics	1.93	3.58	1.91
	Kimono1	0.01	0.17	0.52
	Map	8.36	6.91	8.98
	Programing	6.24	6.32	6.91
	Robot	0.06	0.31	0.33
	SlideShow	66.69	26.69	24.09
	Web_browsing	19.47	10.29	12.25
Average		12.29	9.58	9.55

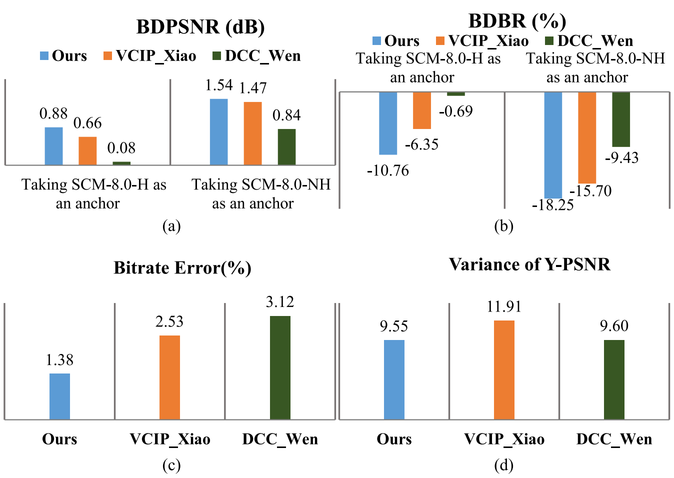


Fig. 8. Coding performance comparison between Ours and other popular rate control algorithms.

the algorithm SCM-8.0-NH, the average BDPSNR of Ours is 1.54 dB which is 0.07 dB larger than that of the algorithm VCIP_Xiao. Meanwhile, the BDPSNR of DCC_Wen is 0.84 dB. Then the BDBR is exhibited in Fig. 8(b). Obviously, the average bitrate reduction of our proposed rate control scheme is much larger than that of VCIP_Xiao and DCC_Wen for both taking SCM-8.0-H or SCM-8.0-NH as anchors. This indicates that the video coding with our rate control algorithm needs less bits than that with VCIP_Xiao or DCC_Wen to obtain the same video quality. As shown in Fig. 8(c), we also compare the coding bitrate error of the three rate control methods. The bitrate error of our proposed algorithm is 1.38%, the bitrate error of VCIP_Xiao is 2.53%, and the bitrate error is 3.12%. Moreover, the comparison of video quality fluctuation reflected by the variance of Y-PSNR is demonstrated in Fig. 8(d). It can be observed that the variances of frame quality obtained by coding with three algorithms, Ours, VCIP_Xiao and DCC_Wen, are 9.55, 11.91 and 9.60, respectively. Since larger variance represents lager fluctuation, and smaller variance represents smaller fluctuation, our proposed method can receive more stable coding quality. Fig. 8 confirms the superiority of our algorithm from BDPSNR, BDBR, bitrate error and quality fluctuation which indicates that making best of the video characteristics of SCVs can improve the coding performance of SCVs.

D. Intuitional Coding Performance Comparison Between Four Algorithms

1) R-D Curves: Except for the indexes BDPSNR and BDBR, R-D curves are also drawn to intuitionally present the coding performance. As shown in Fig. 10, four R-D curves are obtained by encoding with the algorithms SCM-8.0-NH, SCM-8.0-H, VCIP_Xiao and Ours, respectively. Four videos chosen to reflect the coding performance belongs to four kinds of video contents recommended by JCT-VC. In detail, video *MissionControlClips3* belongs to mixed content. Video *Programming* belongs to text and graphics with motion content. Vidoes *Robot* and *Kimono1* belong to Animation and Class camera-captured content, respectively. It can be seen from Figs. 10 (a), (b), (d), the R-D performances of sequences *MissionControlClips3*, *Programming* and *Robot* improve a lot. The quality improvement of the sequence *Kimono1* shown in Fig. 10(c) is inapparent. It is because that our rate control algorithm is specially designed for SCVs, which makes full

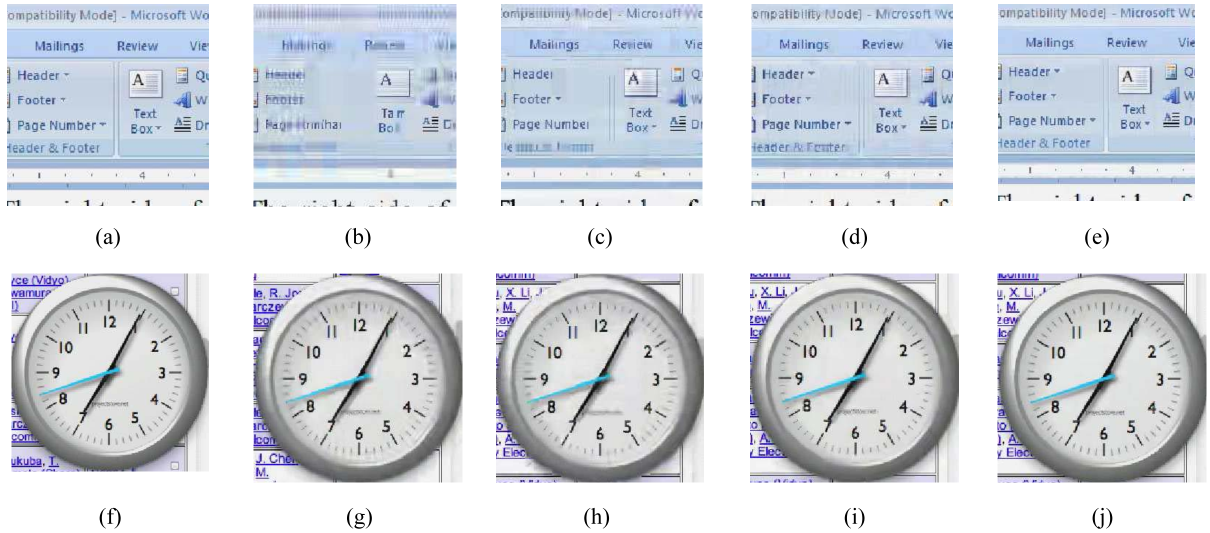


Fig. 9. Subjective quality comparison for different RC algorithms. (a), (f) the original frames; (b), (g) images compressed by SCM-8.0-NH; (c), (h) images compressed by SCM-8.0-NH; (d), (i) images compressed by VCIP_Xiao; (e), (j) image compressed by ours.

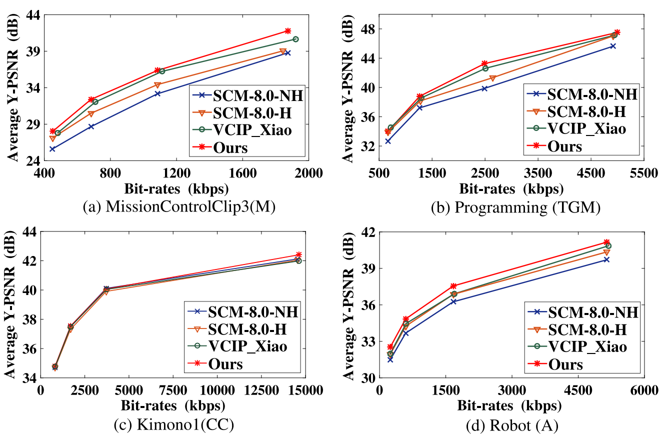


Fig. 10. R-D curves comparisons for different RC algorithms.

use of the characteristics of SCVs. The sequence *Kimono1* is a video captured by camera with characteristics of natural contents.

2) Subjective Quality: The quality of encoded videos can not only be reflected by objective quality, but also by subjective quality. In this section, a frame in the encoded video is randomly selected to reflect the subjective quality. The test sequences *DeskTop* and *MissionControlClips3* are taken as examples to show the subjective quality in Fig. 9. The images in each row from left to right are the original image, the SCM-8.0-NH coded image, the SCM-8.0-H coded image, the VCIP_Xiao image and the image coded by our method, respectively. Obviously, the image encoded by Ours method is clearer than other images, which indicates that our proposed method can improve the subjective quality of video coding.

3) Complexity Analysis: To analysis the computational complexity of the proposed method, a sequence is respectively coded by the original SCM encoder and the proposed encoder at the same hardware platform under the same configuration.

TABLE V
THE COMPLEXITY INCREASE OF THE PROPOSED METHOD COMPARED WITH SCM-8.0

Sequence	Category	BDPSNR(dB	ETI(%))
Missioncontril3	MC	2.85	1.78
Basketballscreen	MC	2.25	10.55
Desktop	TGM	2.72	25.74
Map	TGM	1.48	8.89
Programming	TGM	2.28	0.68
Webbrowsing	TGM	2.16	18.31
Console	TGM	-0.31	6.02
Robot	AMT	1.25	-1.25
Kimono	CC	0.14	3.55

Then, we measure the computation complexity of our method by encoding time increment (ETI), which is defined as

$$ETI = \frac{Time_{Proposed} - Time_{Org}}{Time_{Org}} \times 100\% \qquad (32)$$

where $Time_{Proposed}$ is the encoding time of the proposed method and $Time_{Org}$ is the encoding time of original encoding method. The ETI results of the proposed algorithm with different contents are presented in Table V.

It can be seen from Table V that the coding time increase of the proposed method is 8.25% on average from -1.25% to 25.74% compared with SCM-8.0. Furthermore, the encoding time increase of proposed RC method for different sequences is various. The reason is that the proposed method finally reallocates different target bits. The target bits affect the coding parameters for CTUs including lambda and QP, which have great influence on encoding time. In general, the complexity increase of the proposed method is acceptable with significant video quality improvement at the same bitrate.

VI. CONCLUSION

In this paper, a novel rate control scheme has been proposed. Two important characteristics are made the best to optimize the rate control performance. The first one is that different video contents result in different R-D relationships. The second one is that scenes in screen content videos change frequently. On the basis of the two important features, independent parameter update mode and bit allocation method are put forward. The rate control performance is evaluated from four aspects including R-D performance, rate control accuracy, quality fluctuation and subjective quality. All these evaluation criteria verify the efficiency of our proposed rate control scheme.

REFERENCES

[1] H. Yu, K. McCann, R. Cohen, and P. Amon, *Requirements for an Extension of HEVC for Coding of Screen Content*, document MPEG2014/N14174, ISO/IEC JTC1/SC 29/WG 11, San Jose, CA, USA, Jan. 2014.

[2] 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.

[3] W.-P. Peng *et al.*, "Overview of screen content video coding: Technologies, standards, and beyond," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 393–408, Dec. 2016.

[4] ITU-T and I. JTC1, *Joint Call for Proposals on Video Compression Technology*, ITU-T SG16/Q6 document VCEG-AM91, ITU-T, Kyoto, Japan, Jan. 2010.

[5] C. Pang *et al.*, *CE2 Test1: Intra Block Copy and Inter Signaling Unification*, document JCTVC-T0094, Joint Collaborative Team Video Coding, Geneva, Switzerland, Feb. 2015.

[6] W. Pu *et al.*, "Palette mode coding in HEVC Screen content coding extension," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 6, no. 4, pp. 420–432, Dec. 2016.

[7] P. Lai, S. Liu, and S. Lei, *AHG6: On Adaptive Color Transform (ACT) in SCM 2.0*, document JCTVC-S0100, Joint Collaborative Team Video Coding, Strasbourg, France, Nov. 2014.

[8] B. Li, J. Xu, G. J. Sullivan, Y. Zhou, and B. Lin, *Adaptive Motion Vector Resolution for Screen Content*, document JCTVC-S0085, Joint Collaborative Team Video Coding, Strasbourg, France, Oct. 2014.

[9] J. Xu, R. Joshi, and R. A. Cohen, "Overview of the emerging HEVC screen content coding extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 50–62, Jan. 2016.

[10] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 465–470, Feb. 2013.

[11] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2019.2904198.

[12] "Generic coding of moving pictures and associated audio information-Part 2: Video," Int. Telecommun. Union, Geneva, Switzerland, ITU-T Recommendation H.262, Nov. 1994.

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

[14] "Video coding for low bit-rate communication," Int. Telecommun. Union, Geneva, Switzerland, ITU-T Recommendation H.263, Nov. 1995.

[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] J. I. Ronda, M. Eckert, F. Jaureguizar, and N. Garcia, "Rate control and bit allocation for MPEG-4," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1243–1258, Dec. 1999.

[17] Video Group, *MPEG-4 Video Verification Model Version 2.0, Coding of Moving Pictures and Associated Audi MPEG 96/1260*, document JTC1/SC29/WG11, ISO/IEC, Florence, Italy, 1996.

[18] *Advanced Video Codingfor Generic Audiovisual Services*, ITU-T, Geneva, Switzerland, ITU-T recommendation H.262, May 2003.

[19] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, *Adaptive Basic Unit Layer Rate Control for JVT*, document JVT-G012rl, ITU-T/ ISO/IEC JCV-VC, Pattaya, Thailand, Mar. 2003.

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

[21] H. Choi, J. Nam, and J. Yoo, *Rate Control Based on Unified RQ Model for HEVC*, document JCTVC-H0213, ITU-T/ ISO/IEC JCV-VC, San Jose, CA, USA, Feb. 2012.

[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] 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.

[24] 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.

[25] M. Liu, Y. Guo, H. Li, and C. W. Chen, "Low-complexity rate control based on ρ-domain model for scalable video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Hong Kong, 2010, pp. 1277–1280.

[26] 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.

[27] B. Li, J. Xu, D. Zhang, and H. Li, "QP refinement according to lagrange multiplier for high efficiency video coding," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Beijing, China, 2013, pp. 477–480.

[28] J. Si, S. Ma, and W. Gao, "Efficient bit allocation and CTU level rate control for high efficiency video coding," in *Proc. Picture Coding Symp. (PCS)*, San Jose, CA, USA, Dec. 2013, pp. 89–92.

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

[30] W. Gao, S. Kwong, and Y. Jia, "Joint machine learning and game theory for rate control in high efficiency video coding," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6074–6089, Dec. 2017.

[31] W. Gao, S. Kwong, Q. Jiang, C.-K. Fong, P. H. W. Wong, and W. Y. F. Yuen, "Data-driven rate control for rate-distortion optimization in HEVC based on simplified effective initial QP learning," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 94–108, Mar. 2019.

[32] 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.

[33] H. Zhao, W. Xie, Y. Zhang, L. Yu, and A. Men, "An SSIM-motivated LCU-level rate control algorithm for HEVC," in *Proc. Picture Coding Symp. (PCS)*, San Jose, CA, USA, Dec. 2013, pp. 85–88.

[34] Z. Zhang, T. Jing, J. Han, Y. Xu, and F. Zhang, "A new rate control scheme for video coding based on region of interest," *IEEE Access*, vol. 5, pp. 13677–13688, Mar. 2017.

[35] M. Xu, X. Deng, S. Li, and Z. Wang, "Region-of-interest based conversational HEVC coding with hierarchical perception model of face," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 475–489, Jun. 2014.

[36] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-interest based rate control scheme for high efficiency video coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Florence, Italy, 2014, pp. 7338–7342.

[37] M. Paul, "Efficient multiview video coding using 3-D coding and saliency-based bit allocation," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 235–246, Jun. 2018.

[38] Y. Li, B. Li, D. Liu, and Z. Chen, "A convolutional neural network-based approach to rate control in HEVC intra coding," in *Proc. IEEE Vis. Commun. Image Process.*, St. Petersburg, FL, USA, 2017, pp. 1–4.

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

[40] 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, 2017, pp. 1–4.

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

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

[43] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.

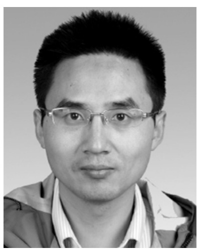
[44] M. Dai, D. Loguinov, and H. Radha, "Rate-distortion modeling of scalable video coders," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, vol. 2. Singapore, 2004, pp. 1093–1096.

[45] H. Yu, R. Cohen, K. Rapaka, and J. Xu, *Common Condition for Screen Content Coding Tests*, document JCTVCQ1015, ITU-T/ISO/IEC JCT-VC, Valencia, Spain, Mar. 2014.

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

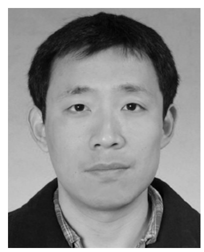
[47] H. Yang, S. Wu, C. Deng, and W. Lin, "Scale and orientation invariant text segmentation for born-digital compound images," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 519–533, Mar. 2015.

Hao Yang received the B.S. degree in communication engineering from Shanghai University, China, in 2015, where he is currently pursuing the Ph.D. degree. He has been working on versatile video coding. His research interests include learning based image/video processing, rate control, and video codec optimization.



Liquan Shen received the B.S. degree in automation control from Henan Polytechnic University, Jiaozuo, China, in 2001, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. Since 2008, he has been a Faculty Member with the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. From November 2013 to November 2014, he was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville. He has authored or coauthored more than 100 refereed technical papers in international journals and conferences in the field of video coding and image processing. He holds ten patents in the areas of image/video coding and communications. His major research interests include high efficiency video coding, perceptual coding, video codec optimization, 3DTV, and video quality assessment.

Yang Yang received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2016, and the M.S. degree from Shanghai University in 2019. Her research interests include image/video coding.



Weiyao Lin received the B.E. and M.E. degrees from Shanghai Jiao Tong University, China, in 2003 and 2005, respectively, and the Ph.D degree from the University of Washington, Seattle, USA, in 2010. He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University, China. His research interests include image/video compression, multimedia understanding, and computer vision.