IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 27, 2019, accepted January 21, 2020, date of publication January 30, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970063

CTU Layer Rate Control Algorithm in Scene Change Video for Free-Viewpoint Video

TAO YAN[1], IN-HO RA[2], HUI WEN[1], MIN-HANG WENG[1], QIAN ZHANG[3], AND YAN CHE[1]

[1]School of Information Engineering, Putian University, Putian 351100, China
[2]School of Computer, Information and Communication Engineering, Kunsan National University, Gunsan 54150, South Korea
[3]School of Information and Electromechanical Engineering, Shanghai Normal University, Shanghai 200234, China

Corresponding author: In-Ho Ra (ihra@kunsan.ac.kr)

This work was supported in part by the Natural Science Foundation of China under Grant 61741111, in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology under Grant 2016R1A2B4013002, in part by the Natural Science Foundation of Fujian under Grant 2019J01816 and Grant 2019J01815, in part by Natural Science Foundation of Jiangxi under Grant 20181BAB202011, in part by the Putian University's Initiation Fee Project for Importing Talents for Scientific Research under Grant 2019003 and Grant 2018088, and in part by the Engineering Research Center of Big Data Application in Private Health Medicine. The work of Tao Yan was supported by the Program for New Century Excellent Talents in Fujian Province University.

ABSTRACT At present, the rate control algorithm for multiview high-efficiency video coding (MV-HEVC) does not have the capability of efficient coding tree unit(CTU) layer bit allocation, and the video quality varies greatly for sequences with sudden scene changes or large motions. To overcome this limitation, this paper proposes a rate control algorithm for MV-HEVC based on scene detection. Firstly, we established ρ domain rate control model based on multi-objective optimization. Then, it uses image similarity to make reasonable bit allocation among viewpoints. If the video scene is switched, the image similarity is recalculated, and then the correlation between the weights of the interview point rates and the correlation between the viewpoints are analyzed. Finally, the frame layer rate control considers the layer B-frame and other factors in allocating the code rate, and the basic unit layer rate control adopts different quantization methods according to the content complexity of the CTU. Experimental results show that the proposed rate control algorithm can maintain good coding efficiency and decrease the average video quality variation by 25.29%.

INDEX TERMS 3D video coding, scene detection, rate control, image similarity analysis, bit allocation.

I. INTRODUCTION

In a 3DTV/FTV system, multiview video coding (MVC) and decoding are common techniques regardless of the method used by the multiview acquisition section and the three-dimensional (3D) stereoscopic display section. Currently, videos are commonly two-dimensional or unrealistic three-dimensional scenes. Multiview video is a new type of video with real three-dimensional stereoscopic and viewpoint interaction functions. Users can select and examine 360-degree 3D scenes from multiple angles with the naked eye. It has become a new research interest for scholars at home and abroad [1]–[7]. Video encoding has also come up with many excellent algorithms. For example, many researchers use machine learning to video coding [8]–[10]. In the case of rate control, when scene changes occur in a video, the time domain redundancy of the video at the changing sections disappears. If the adjacent video frames are switched as part of the same coding unit, 3D video coding rate control is generated. A large error greatly reduces the compression performance of the video, and the actual video has a large number of scene changes. Therefore, it is necessary to perform scene change detection and use image similarity analysis to detect video scene changes for adaptive adjustment of coding length.

A video scene consists of a set of semantically related shots that describe an independent high-level semantic unit. Shots in the same scene are temporally adjacent—they may occur in the same place and share the same background or may be different sides of the same content—and combine to express a complete semantic meaning. The composition of the video

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

scene should conform to the principle of film editing. Based on the analysis of this principle, a scene detection algorithm is proposed. By detecting the lens boundary, the video stream is segmented into shots, and the key frames are dynamically extracted according to changes in the lens content. Lenses with similar time and similar content are combined into a lens class, and the correlation between the lens classes is analyzed to complete the detection of the video scene. Therefore, because of the maturity of lens boundary detection technology, it is meaningful to study the detection (or generation) of video scenes. Current video scene detection methods fall into two broad categories: model-based methods and film-based rules. In the first type of method, a prior model must be built based on the specific application or domain. The existing lens-based video analysis method cannot reflect the semantic relationship of the video because the lens information granularity is too small. Therefore, it is necessary to organize video content according to the high-level semantic unit—the scene. A fast and effective video scene detection method is proposed. According to the principle of film editing, the development mode of video scene content is classified, and the principle of scene construction is given. A new combination method based on a sliding lens window is proposed. Lenses of similar content are organized into a lens class, and a lens class correlation function is defined to measure the correlation between the lens classes and complete the generation of the scene.

MV-HEVC is one of the newer video coding standards [11]. At present, rate control based on MV-HEVC around the word is relatively less than that on the previous rate control technology. Most of the studies are engaged in research on rate control for HEVC [12]–[16]. There have been many studies on rate control for HEVC based on scene switching [17], [18], but most of the rate control algorithms for HEVC are concentrated in the single-channel video coding standard which are not applicable to mutiview video coding. And multi-view video is far more than two or more videos, and its bit allocation and rate control are more complicated. Lim *et al.* proposed a MVC rate control algorithm based on the binomial model [19]. The algorithm used the geometric relationship between the disparity prediction and the motion prediction to divide all video frames into several different coding type frames between the various viewpoints. There are large differences between parallax prediction characteristics, although Fani and Rezaei considered the influence of layered B-frames in the temporal coding layer in multiview video coding [20] and proposed a novel PID-fuzzy video rate controller for high-delay applications of the HEVC standard. The experimental results showed that the algorithm could maintain good coding efficiency, but the average rate control of experimental results given by test sequences had an error of above 1%, which makes it difficult for the algorithm to meet the needs of practical applications. In [21], the inter-view RD model was established by the joint texture and the virtual view point distortion function. However, this method does not utilize the inter-viewpoint correlation, and the code rate control accuracy is lower than the average bit rate error of 1.56%. Yan and Ra [22] proposed slightly allocation algorithm based on similarity analysis among viewpoints for 3D video coding. Vizzotto *et al.* considered only two viewpoints in the rate control algorithm for stereo video coding [23]. Because the algorithm was used in multi-view video coding, this type was based on TM5. The accuracy of the target bit number allocation was deteriorated, resulting in very high rate control error and difficulty in bit allocation.

There are also many scholars around the world engaged in HEVC-based multiview video coding rate control research. Shao *et al.* [24] proposed a fixed allocation ratio to assign code rates between texture depths, but this method does not achieve optimal coding efficiency for different sequences. Xiao *et al.* [25] proposed a scalable rate allocation algorithm to apply different bandwidths. Fang *et al.* [26] proposed an analysis model for estimating virtual visual distortion in 3D videos. The analysis method combined with the frequency domain and time domain was used to estimate the virtual visual distortion caused by the distortion of the depth image. The estimation model was accurate but complex. Pan and Wei [27] proposed a depth-based 3D-HEVC rate control algorithm with a fixed color and depth code rate ratio of 4:1, but this method cannot obtain optimal virtual view quality. Wang *et al.* [28] proposed a 3D-HEVC rate control algorithm based on the binomial R-D model. Because of the direct use of the H.264 medium rate control model, the rate control accuracy is low. In 2018, Lei *et al.* [29] established the 3D-HEVC rate control algorithm using the R-λ model. The experimental results were very satisfactory, but the correlation between the viewpoints was not considered, resulting in high computational complexity. The above three-dimensional video coding rate control algorithm did not consider the scene change problem. Therefore, further research is needed to study the rate control optimization model for MV-HEVC.

This paper proposes a three-dimensional video encoding rate control method combined with scene detection. The core of the algorithm is to first use image similarity analysis to detect video scene changes. If the video scene is switched, the group of pictures (GOP) encoding length is adaptively adjusted, and then image similarity analysis is used to analyze the inter-view correlation and frame complexity, such as the inter-view point, frame layer, basic unit layer bit allocation and rate control. Experimental simulation results show that the algorithm can maintain good coding efficiency and greatly improve the accuracy of rate control.

II. ρ MODEL FOR MV-HEVC

By studying the advantages and disadvantages of the classic rate control model for 2D video coding (mainly including the binomial *R-D* model and ρ model), the two aforementioned bit-rate control models were then improved to study the optimal model for MV-HEVC bit-rate control. Previous studies obtained an MVC bit-rate control algorithm based on the binomial R-D model [22] and [30], and the MV-HEVC bit-rate control algorithm is similar to MVC bit-rate control.

The design of an MV-HEVC bit-rate control model can also be considered as a multiobjective optimization problem. It is necessary to adaptively adjust the parameters of the bit-rate control model according to the characteristics of the video content and the requirements of the specific application. The objective of this study was to experimentally determine the relationship between ρ(representing the number of zero coefficients after quantifying a transform coefficient as a percentage of all coefficients) and the bit-rate of the encoded part of the texture. Without loss of generality, it is assumed that ρ and the texture coding bit rate $T(\rho)$ have the following relationship:

$$T(\rho) = \sum_{i=1}^{n} \alpha_i (1-\rho)^i + \gamma \tag{1}$$

The parameter solution of $\alpha_i (i = 1, 2 \cdots n)$ and γ, which can also be seen as a multiobjective optimization problem, is:

$$\begin{cases} Max \quad PSNR = \dfrac{1}{N} \sum_{s=1}^{N} PSNR(s) \\ Min \quad \sigma = \dfrac{1}{N} \sum_{s=1}^{N} (PSNR(s) - PSNR)^2 \\ Other\ objective\ functions \end{cases} \tag{2}$$

with a limiting condition:

$$\frac{1}{N} \sum_{s=1}^{N} E(s) < E \tag{3}$$

σ can reflect the fluctuation degree of the video quality before and after the application of bit-rate control, $E(s)$ denotes the viewpoint bit-rate control error of S, and E denotes the expected set rate control error.

Of course, to find the local optimal solution, the idea of this study was to simplify the aforementioned MV-HEVC bit-rate control model. The preliminary study found that the $a_i (i \geq 3)$ bit-rate control model is relatively high in complexity; thus, we only had to study the condition of $a_i (i \leq 2)$ and $\lim_{\rho \to 1} T(\rho) = 0$. Therefore, the aforementioned MV-HEVC bit-rate control model can be simplified as $T(\rho) = \sum_{i=1}^{2} a_i \cdot (1-\rho)^i$. We can assume that ρ has the following quadratic relationship with the encoding bit rate $T(\rho)$ of the texture.

$$T(\rho) = \alpha \cdot (1-\rho)^2 + \beta \cdot (1-\rho) + \gamma \tag{4}$$

where α, β, γ are the one-dimensional regression coefficients and their initial values are set to 0, R/2, and 0, respectively. After each frame is encoded, it is necessary to refresh during the postencoding stage as follows:

$$\lim_{\rho \to 1} T(\rho) = 0 \tag{5}$$

Then, we study the following $T - \rho$ model:

$$T(\rho) = \alpha (1-\rho)^2 + \beta (1-\rho) \tag{6}$$

where α, β can be provided by the following statistical analysis method. Suppose that $x_1(\rho) = (1-\rho)^2, x_2 = 1-\rho$, and let $(x_{11}, x_{12}, k_1), (x_{12}, x_{22}, k_2), \cdots, (x_{1n}, x_{2n}, k_n)$ be the n existing sample values; thus

$$X = \begin{pmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \\ \vdots & \vdots \\ x_{1n} & x_{2n} \end{pmatrix} \quad K = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{pmatrix} N = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \tag{7}$$

Using multiple regression techniques, the model parameter N can be calculated as follows:

$$N = (X^T \cdot X)^{-1} \cdot X^T \cdot K \tag{8}$$

where X^T is the transpose matrix of X and $(X^T X)^{-1}$ is the inverse matrix of $X^T X$.

III. 3D VIDEO SCENE DETECTION

A video scene consists of a set of semantically related shots that describe an independent high-level semantic unit. Shots in the same scene are temporally adjacent. The shots may occur in the same place sharing the same background or may be different sides of the same content that combine to express a complete semantic meaning. The composition of the video scene should conform to the principle of film editing. Based on the analysis of this principle, a scene detection algorithm is proposed. By detecting the lens boundary, the video stream is segmented into shots, and the key frames are dynamically extracted according to changes in the lens content. The lenses with a similar time and similar content are combined into a lens class. The correlation between the lens classes is analyzed to complete the detection of the video scene. In the case of rate control, when scene changes occur in the video, the time domain redundancy of the video at the point of the changes disappears. If the adjacent video frames are still switched as the same coding unit, 3D video coding rate control is achieved. However, a large error greatly reduces the compression performance of the video, and the actual video has a large number of scene changes. Therefore, it is necessary to perform scene change detection and adopt image similarity analysis to detect the adaptive adjustment coding length of the video scene changes.

A. SCENE DETECTION

Aiming at a large number of scene changes in video frames, a nonconnected point-based scene change detection algorithm is proposed to improve coding performance. The proposed scene detection algorithm has low complexity. At the same time that motion estimation is performed, the video scene change detection is completed. The scene changes will result in a change in the length of the GOP and may occur when the GOP length is too short. An improved adaptive GOP time-domain filtering technique is proposed to avoid degradation of coding performance due to a short GOP. For video scene change detection and segmentation of different lengths of the GOP, an interframe rate control algorithm based on the rate

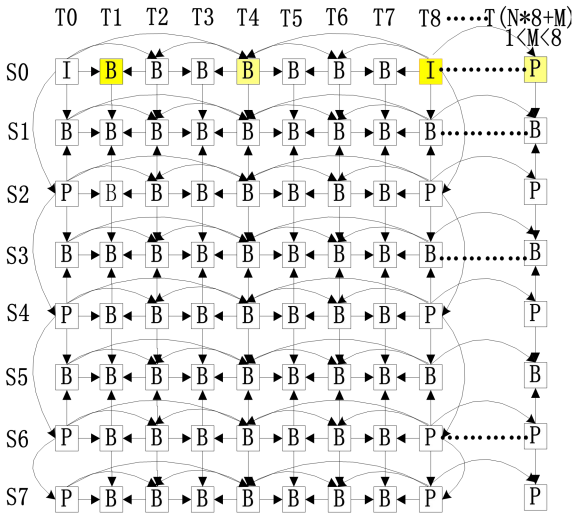


FIGURE 1. MV-HEVC prediction structure.

distortion model is proposed. The interframe rate allocation is performed by using the relationships between video distortion and both code rate and video frame complexity, thus optimizing the total quality of the reconstructed video frames. The experimental results show that the adaptive interframe rate control algorithm based on scene detection can obtain an improved coding performance.

Our study found that when the video scene is switched, the similarity of the adjacent images drops sharply. Therefore, this paper uses image similarity analysis to detect video scene changes and adaptively adjust the encoding length. P_i represents the scene change probability of the i[th] frame, as expressed in formula (9).

$$P_i = \left| 1 - \frac{Coff_{i,i-1}}{\left(\frac{1}{i-2}\right)\sum_{j=1}^{i-2} Coff_{i-1,j}} \right| \geq \beta \tag{9}$$

where $Coff_{i,j}$ is the correlation between the i[th] frame and the j[th] frame, which can be obtained by formula (10), and β is a set threshold. Let A and B be two images acquired at the time of the two viewpoints. $L(A,B)$, $C(A,B)$ and $S(A,B)$ represent the brightness comparison, contrast comparison and structural information comparison of the two images, respectively. Then, the correlation between images A and B is determined by formula (10) as follows:

$$Coff(A,B) = \eta \cdot (L(A,B))^{\phi} \cdot (C(A,B))^{\varphi} \cdot (S(A,B))^{\gamma} \tag{10}$$

$\phi, \varphi,$ *and* γ can adjust the proportion of brightness, contrast and structural information, respectively. η is an adjustment parameter.

B. ANALYZE THREEE POINTS OF SCENE SWITCHING

Figure 1 shows the MV-HEVC prediction structure, where the horizontal axis represents the time direction. It is assumed that the GOP length is 8 and the encoding length L is N^*8+R. Scene switching is divided into the following three points.

1) Assume that the scene change occurs at frame X. $X = Y^*8,\ 0 <= Y <= N$, which is exactly a GOP multiple. It uses I-frame encoding. I-frames are intra-coded using intra-frame prediction. Its reference image is an image that has been encoded by itself. Therefore, scene switching in the I frame does not affect the encoding of the current frame or the encoding of subsequent frames. Therefore, The GOP is defined and does not change during transmissions.

2) Assume that the scene change occurs at frame X. $X = Y^*8 + J,\ 0 <= Y <= N-1,\ 0 < J < 8$. As shown in the figure, it uses B-frame encoding, and B-frame uses inter-prediction to perform inter prediction. Or refer to the self-encoded image, so at least one frame is in the same scene as its reference frame, so it does not affect the encoding of the B frame. Therefore, The GOP is defined and does not change during transmissions.

3) Assume that the scene change occurs at frame X. $X = N^*8 + J,\ 0 < J <= R$. As shown in the figure, it uses P-frame coding. When scene switching occurs, the correlation between adjacent I-frames and P-frames is greatly reduced. The spatial correlation of the CTU is greater than the temporal correlation. After the CTU is selected by the coding mode, the intra prediction coding mode is used. The actual bits generated by the coding are much larger than those generated by the inter prediction mode. The reduction of encoding resources affects the encoding quality of subsequent frames and consumes unnecessary inter-frame prediction time. Because of the last few frames, the impact is not significant and can be ignored. Therefore, The GOP is defined and does not change during transmissions.

That being the case, why do we need scene detection in this paper? The scene switching proposed in this paper has two main advantages. The first advantage. Previously, most scene switching algorithms used inter-frame correlation for scene detection without considering inter-view correlation. The second advantage. We can use the correlation between viewpoints to detect scene switching, and we can also use the correlation between viewpoints to allocate bits between viewpoints. The difficulty of rate control for MV-HEVC is how to allocate bits between views. Previous rate control algorithms used correlation among viewpoints to allocate reasonable bits among viewpoints. The encoded information is used to predict the weight of the bit allocation of the current viewpoint. When the scene is switched, the correlation is greatly reduced and continued use of this method will produce a large error. Therefore, this paper first examines scene switching. If a scene change occurs, the inter-view correlation is recalculated.

IV. 3D VIDEO CODING RATE CONTROL ALGORITHM

The multiview video sequence is obtained by continuously shooting the same scene with a camera array, and there is a very high content correlation between viewpoints. The size of the inter-view correlation reflects the redundancy between

the viewpoints. When inter-view point rate allocation is performed for different multiview videos, the difference in the correlation between the viewpoints must be considered.

To be compatible with the latest coding of standard HEVC, the bit allocation and rate control proposed in this paper is based on the HEVC rate control algorithm. The key step of the proposed algorithm is to perform reasonable rate allocation between each viewpoint according to multiview video coding. The purpose is to ensure a balance of video quality between viewpoints. Based on previous research [22] and [30], the proposed algorithm is further improved. The key steps of the algorithm are described as follows.

A. VIEW LEVEL RATE CONTROL

Before encoding, the bits between the viewpoints are difficult to allocate. In this paper, the bits are allocated reasonably at different viewpoints according to the correlation between the viewpoints. In the MVC encoding architecture, the primary view is first encoded. Since the main view does not refer to other viewpoints, the generated code rate is independent of the main view. The primary view refers to the coded image frame of the same view and the image of the coded view for predictive coding, and the generated code rate has a close relationship with the reference view. The weight w_k is used to indicate the degree of importance of the viewpoint k. The larger w_k is, the more important the viewpoint. The total number of bits allocated in the first viewpoint in the *GOP* is obtained by formula (11) as follows:

$$T_k = T_{total} \cdot w_k \tag{11}$$

The weighting factor w_k of each viewpoint GOP is calculated, and then the target number of bits of the current viewpoint is obtained. The specific steps are described by formulas (12)-(14) as follows:

$$Coff(S_k, S_0) = Coff(A, B) \tag{12}$$

$$\xi_k = \varphi_k \cdot Coff(S_k, S_0) \tag{13}$$

$$w_k = \frac{\xi_k}{\sum_{i=0}^{N_{view}-1} \xi_i} \tag{14}$$

where $Coff(S_k, S_0)$ represents the correlation factor between viewpoint k and viewpoint 0; A and B represent the viewpoint and the two corresponding images at the same time; *L(A, B), C(A, B)* and *S(A, B)* represent the brightness comparison, contrast comparison and structural information comparison of the two images, respectively; $\phi, \varphi,$ *and* γ indicate the adjustable proportions of brightness, contrast and structural information, respectively; ξ_k is a composite factor representing a viewpoint; φ_k is an djustment factor; and w_k is a weighting factor of a viewpoint, that is, an importance degree.

Our previous research has shown that when the scene does not switch, w_k can be predicted based on the encoded information [30]. When the scene is switched, the correlation is greatly reduced and continued use of this method will produce a large error. Therefore, this paper first examines scene switching. If a scene change occurs, the inter-view correlation is recalculated.

B. TARGET BIT RATE FOR FRAMES

In the MV-HEVC frame layer rate allocation, the bit allocation per frame is determined by the target buffer capacity, the frame rate, and the actual buffer size. The residual energy of the current coded frame is not taken into account, which may cause image quality degradation or frame skipping. The six different frame types in MVC comprise intraframe coding (I-frame), one-way prediction in the time direction, bidirectional prediction in the time direction, one-way prediction between viewpoints, bidirectional prediction between viewpoints, and both time and inter-view prediction.

In general, when the motion of the video sequence is more intense, the active time domain of each frame is larger. That is, as the number of image content scene changes increases, more bits are needed for encoding. The active time domain of each frame is the image content scene change. Thus, when the encoding is smaller, fewer bits are needed for encoding. To make the MV-HEVC rate control more precise, the current frame target bit is calculated by equation (15):

$$T_r(j) = T_r'(j-1) \cdot \frac{\sum_{u=1}^{L} W(u) \cdot 2^n}{\sum_{u=1}^{L} \frac{w^B}{w^{I/P}} \cdot W(l) + \sum_{u=1}^{L} W_B(u) \cdot (2^n - 1)} + T_j \tag{15}$$

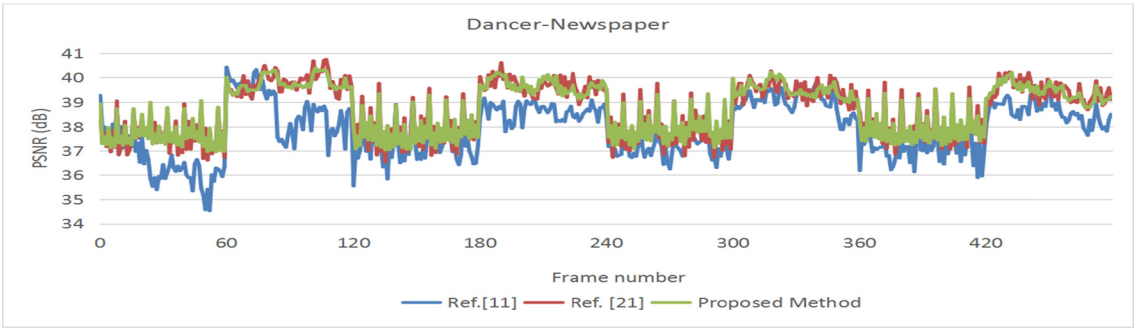
In equation (15), T_j is the bit consumed by the frame header information of frame j, n represents the current time level, $W(u)$ represents the weight of each frame complexity, and $W_B(u)$ represents the weight of the B-frame.

The number of bits allocated to the current frame can be determined according to the number of remaining bits, the number of remaining B-frames per layer, and the scaling factor of the current GOP. $T_{l,i}^B$ represents the number of bits allocated to the l^{th} hierarchical B-picture of the current GOP, and $T^{I/P}$ represents the number of bits to which the I-picture or P-picture is allocated, which can be expressed by equations (16) and (17), respectively:

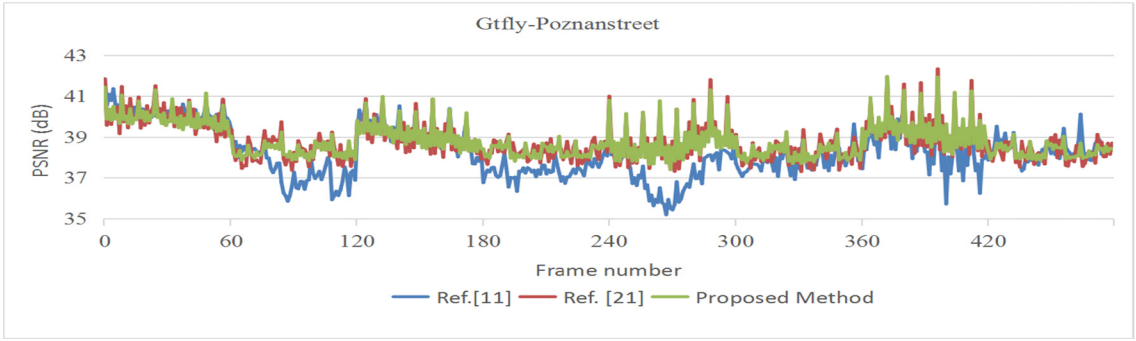
$$T_{l,i}^B = \frac{\beta \cdot w_l \cdot B_l(i)}{\sum_{k=l}^{d-1} w_k \cdot N_B^i(k)} \tag{16}$$

$$T^{I/P} = \frac{\varphi \cdot w^{I/P} \cdot B_{tot}(i)}{W^{I/P} + \sum_{j=0}^{d-1} w_j \cdot N_B^{I/P}(j)} \tag{17}$$

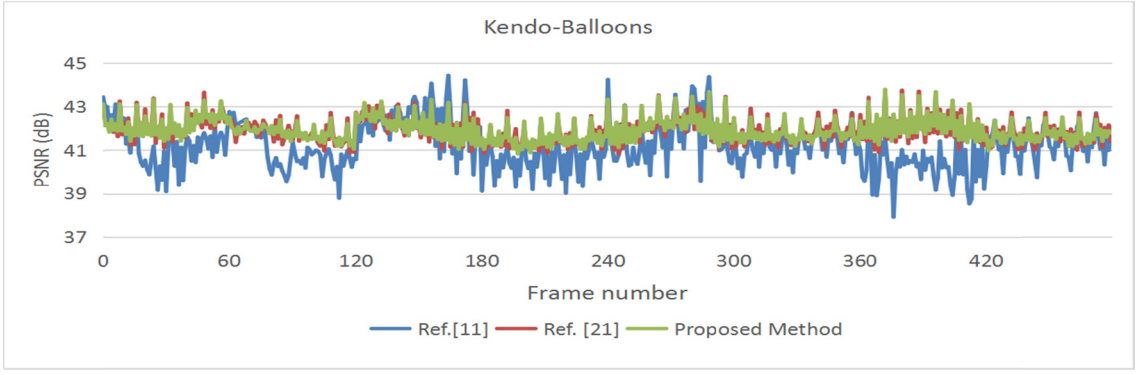
where $B_l(i)$ represents the l^{th} hierarchical i^{th} B-picture when encoding the number of remaining bits in the image in the current GOP, $N_B^i(k)$ represents the number of B-pictures that have not been encoded in the k^{th} layer, B_{tot} represents the total number of bits allocated by the current GOP, and $\omega^{I/P}$ and ω_k represent the scaling factor and the k^{th} point of the I-picture or P-picture, respectively.



(a) Experimental PSNR results of Dancer-Newspaper



(b) Experimental PSNR results of Gtfly-Poznanstreet



(c) Experimental PSNR results of Kendo-Balloons

FIGURE 2. Objective quality fluctuations of sequences.

C. CTU LAYER RATE CONTROL

On CTU level of rate control, considering the complex information of the CTU and the whole frame, the number of bits is allocated according to the relative complexity of the CTU, and the appropriate quantification parameters are determined to achieve the purpose of bit-rate control. The core of the code rate control algorithm based on the ρ domain model is to determine the quantification parameters of the current unit based on the number of bits available in the texture of the uncoded part of the current frame. The number of bits of available texture corresponding to the uncoded unit is allocated according to the total number of bits. The textural part of the uncoded part is the percentage of the total number of bits that are not encoded in the preanalysis process.

1) CALCULATION OF THE ρ VALUE

The quantitative formula used is:

$$\hat{x}_{ij}(QP) = Round(x_{ij} \cdot M(i,j) + \lambda \cdot 2^{qbits}) \gg qbits \quad (18)$$

where "$\gg$ " indicates binary right shift, $\hat{x}_{ij}$ denotes the corresponding quantized value and QP is the quantized parameter. $M(i,j)$ depends on the Qp value. The Qp value for every 6, $M(i,j)$ is the same. λ is the dead band threshold, which is 1/3 when the CTU is in the intraframe coding mode and 1/6 when the CTU is in the interframe coding mode. Denoting the distribution of transform coefficients by $D(x)$, the calculation of ρ can be expressed by the following formula:

$$\rho(QP) = \frac{1}{L} \sum_{|x_{ij}| \le (1-\lambda) \cdot 2^{qbits/M(i,j)}} D(X) + \zeta \quad (19)$$

where L is the number of total transformation coefficients. ζ is an adjustment parameter.

2) DETERMINATION OF CTU LEVEL QUANTIZATION PARAMETERS

$\rho - QP$ relationship expressed by equation (19) calculates QP. When calculating QP, the target QP_{target} is estimated by $p_{t\arg et}$. Because the $\rho - QP$ relationship is monotonically increasing, there is no need to draw the entire QP range on the $\rho - QP$ line. If ρ_{target} is greater than ρ_0, then $QP_{t\arg et}$ is greater than the initial QP, and vice versa. Using the monotonically increasing relation of $\rho - QP$, the target QP is determined by an effective amount of computation.

This process can be applied to the following three situations to determine the quantitative parameters:

a) For the first unit, the quantization parameter takes the average quantization variance of each unit of the previous frame as follows:

$$QP_{0,i}(j+1) = \overline{QP(j)} \tag{20}$$

b) When $T_{r,i}(j+1) < 0$, the quantized parameters of the current unit should be greater than the quantized parameters of the previous unit, that is:

$$QP_{l,i}(j+1) = QP_{l-1,i}(j+1) + 1 \tag{21}$$

c) In other cases, the target bit of the current basic CTU is allocated according to the predicted value of the MAD. The method is simple; it is to divide the bits allocated per frame among the basic unit layers of the frame. In this manner, different CTUs in the same basic unit layer are encoded with the same quantitative parameters.

$$QP_{CTU} = \begin{cases} QP_{CTU} - \alpha & \alpha = \left\lfloor \dfrac{MAD_{CTU}}{MAD_{unit}} \right\rfloor \\ QP_{CTU} + \alpha \end{cases} \tag{22}$$

In equation (22), MAD_{CTU} represents the predicted value of the MAD of the i-th CTU in the current basic unit layer and MAD_{unit} represents the predicted MAD value of the current basic unit. QP_{CTU} represents the quantization parameter of the i-th CTU in the current basic unit layer, and QP_{unit} represents the quantitative parameters of the current basic unit.

V. EXPERIMENTAL RESULTS

To verify the proposed rate control algorithm, this section uses five different nature video sequences to simulate the CTU layer rate control algorithm based on scene detection. Several video sequences with different resolutions, frame rates, and motion activities are tested to evaluate the performance of the proposed rate control scheme. We select seven sequences ''Gtfly'', ''PoznanStreet'' ''Kendo'', ''Newspaper'', ''PoznanHall2'', ''Balloons'', and ''Dancer''. Six sequences ''Gtfly-PoznanStreet,'' ''PoznanStreet-Kendo,'' ''Newspaper-PoznanHall2'', ''Dancer-Newspaper'', ''Kendo-Balloons'', ''Balloons-Newspaper'' with scene cuts, which are generated by cascading two sequences ''Gtfly'' and ''PoznanStreet'', ''PoznanStreet'' and ''Kendo'' ''Newspaper'' and ''PoznanHall2'', ''Dancer'' and ''Newspaper'', ''Kendo'' and ''Balloons'', ''Balloons'' and ''Newspaper'' ''News'' and ''Silent,'' ''Foreman'' and ''Silent,'' and ''Coastguard'' and ''Foreman''. The test conditions are listed in Table 1, and the experiment results are given in Table 2 and Table 3.

TABLE 1. The test conditions.

Experience result	Scene change	GOP Length	The position of scene change
Table. 2	The first point	8	80,160,240,320, 400,480
	The second point	8	82,162,242, 322,402
	The third point	8	482
Table. 3	The first point	12	60,120,180,240, 300,360,420,480

In the latest test platform, this algorithm is compared with the previous two algorithms. In this paper, σ denotes the rate control error (σ_{error}), which is calculated as

$$\sigma_{error} = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \tag{23}$$

where $R_{t\arg et}$ and R_{actual} denote the target bit-rates and the actual coding bit-rates of the sequence, respectively.

The performance measures include the PSNR and the PSNR variation (θ_{PSNR}), which is obtained from equation (24).

$$\theta_{PSNR} = \frac{1}{N} \cdot \sum_{i=1}^{N} (PSNR_i - \frac{1}{N} \cdot \sum_{i=1}^{N} PSNR_i)^2 \tag{24}$$

where N denotes the number of total encoded frames.

Previous rate control algorithms used correlation among viewpoints to allocate bits among viewpoints. The encoded information is used to predict the weight of the bit allocation of the current viewpoint. When the scene is switched, the correlation is greatly reduced and continued use of this method will produce a large error. If a scene change occurs, the interview correlation is recalculated. We implemented our proposed rate control algorithm on MV-HEVC and compared it with [30] using different test sequences with various target bit rates. Table 2 shows some experimental results. Compared with the previous algorithm, the algorithm proposed in this paper can significantly reduce video quality fluctuations while improving encoding efficiency, and the average reduction in PSNR fluctuations is 40.71%. Therefore, this paper first examines scene switching

We implemented our proposed rate control algorithm on MV-HEVC and compared it with [21] using different test

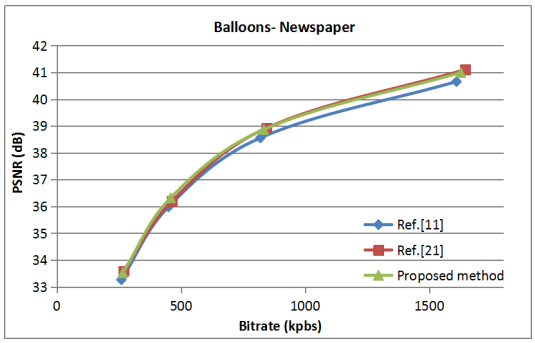
TABLE 2. Performance comparison with other rate control algorithm for three points.

Three points	Sequence	Target bit rate (kbps)	Ref. [30]				The proposed method				The proposed method vs. Ref. [30]	
			Actual bit rate (kbits)	Average PSNR (dB)	θ_{PSNR}	σ_{error} (%)	Actual bit rate (kbits)	Average PSNR (dB)	θ_{PSNR}	σ_{error} (%)	θ_{PSNR} reduction (%)	PSNR gain (dB)
The first point	Gtfly-PoznanStreet	4450	4565.26	38.90	1.94	2.59	4508.74	39.10	0.96	1.32	50.52	0.20
		1730	1794.88	36.87	2.41	3.75	1760.45	37.22	1.01	1.76	58.09	0.35
		800	817.36	34.74	2.56	2.17	805.92	35.10	1.18	0.74	53.91	0.36
		400	411.08	32.79	2.22	2.77	403.32	33.01	1.15	0.83	48.20	0.22
	PoznanStreet-Kendo	3000	3096.30	39.98	1.61	3.21	3055.20	40.25	0.86	1.84	46.58	0.27
		1000	1033.90	38.25	1.32	3.39	1020.40	38.54	0.91	2.04	31.06	0.29
		500	512.40	36.38	1.77	2.48	503.50	36.63	0.85	0.70	51.98	0.25
		300	308.07	33.74	1.57	2.69	302.31	34.27	0.82	0.77	47.77	0.53
	Newspaper-PoznanHall2	1550	1594.95	40.73	0.68	2.90	1578.52	40.95	0.49	1.84	27.94	0.22
		700	728.42	39.06	0.80	4.06	711.62	39.23	0.54	1.66	32.50	0.17
		400	419.44	37.07	1.37	4.86	410.12	37.34	0.69	2.53	49.64	0.27
		200	206.46	35.01	0.85	3.23	204.20	35.26	0.55	2.10	35.29	0.25
The second point	Dancer-Newspaper	4000	4148.00	38.84	1.05	3.70	4056.00	39.02	0.49	1.40	53.33	0.18
		1800	1852.74	36.24	0.76	2.93	1829.34	36.53	0.58	1.63	23.68	0.29
		900	930.15	33.84	0.98	3.35	915.39	34.11	0.56	1.71	42.86	0.27
		400	415.32	31.37	0.97	3.83	408.40	31.76	0.55	2.10	43.30	0.39
	Kendo-Balloons	1500	1542.60	41.74	0.98	2.84	1521.90	41.93	0.76	1.46	22.45	0.19
		800	826.08	39.96	1.12	3.26	810.88	40.11	0.98	1.36	12.50	0.15
		450	464.67	37.27	1.31	3.26	456.26	37.65	0.98	1.39	25.19	0.38
		250	257.53	34.68	2.02	3.01	252.65	34.97	0.88	1.06	56.44	0.29
The third point	Balloons-Newspaper	1600	1652.48	40.83	1.03	3.28	1619.68	41.11	0.52	1.23	49.51	0.28
		800	832.88	38.73	0.97	4.11	814.16	38.96	0.59	1.77	39.18	0.23
		500	519.05	35.90	0.72	3.81	510.30	36.44	0.51	2.06	29.17	0.54
		300	312.30	33.41	1.15	4.10	307.56	33.65	0.65	2.52	43.48	0.24
	PoznanHall2-Dancer	4000	4113.20	39.64	0.99	2.83	4061.60	39.75	0.55	1.54	44.44	0.11
		1600	1648.96	37.58	1.08	3.06	1613.76	37.91	0.69	0.86	36.11	0.33
		800	826.48	35.70	0.94	3.31	814.40	36.04	0.66	1.80	29.79	0.34
		400	410.88	33.70	1.29	2.72	402.92	34.03	0.58	0.73	55.04	0.33
	Average	1254.64	1294.35	36.89	1.30	3.27	1273.55	37.17	0.73	1.53	40.71	0.28

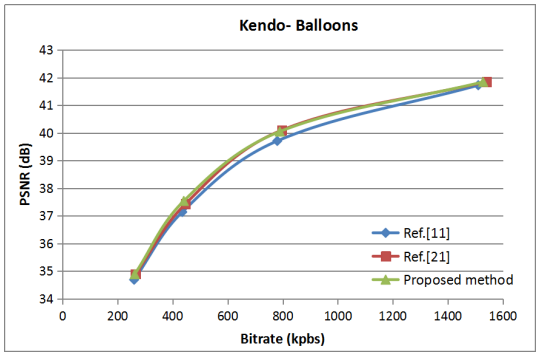
TABLE 3. Performance comparison with other rate control algorithm.

Sequence	Target bit rate (kbps)	The proposed method				Ref.[21]				The proposed method vs. Ref. [21]	
		Actual bit rate (kbits)	Average PSNR (dB)	θ_{PSNR}	σ_{error} (%)	Actual bit rate (kbits)	Average PSNR (dB)	θ_{PSNR}	σ_{error} (%)	θ_{PSNR} reduction (%)	PSNR gain (dB)
Balloons-Newspaper	1610	1627.22	41.00	0.41	1.07	1645.41	41.11	0.48	2.20	15.46	-0.11
	820	833.13	38.87	0.48	1.60	844.94	38.93	0.55	3.04	12.77	-0.06
	450	458.45	36.32	0.41	1.88	464.08	36.20	0.51	3.13	19.59	0.12
	260	266.03	33.53	0.52	2.32	268.74	33.60	0.67	3.36	21.93	-0.07
Kendo-Balloons	1510	1527.74	41.84	0.63	1.18	1540.28	41.85	0.74	2.01	14.54	-0.01
	780	789.10	40.05	0.78	1.17	796.35	40.10	0.92	2.10	14.94	-0.05
	435	440.37	37.53	0.79	1.24	446.16	37.42	1.03	2.57	23.74	0.11
	260	262.11	34.88	0.78	0.81	265.54	34.89	1.24	2.13	36.94	-0.01
Newspaper-PoznanHall2	1525	1548.79	40.88	0.30	1.56	1557.18	40.92	0.34	2.11	10.28	-0.04
	705	714.91	39.16	0.36	1.41	725.20	39.23	0.56	2.87	35.42	-0.07
	375	383.89	37.26	0.49	2.37	389.48	37.22	0.61	3.86	20.33	0.04
	215	219.16	35.15	0.44	1.94	220.62	35.19	0.57	2.62	23.49	-0.04
Gtfly-PoznanStreet	4430	4480.87	39.03	0.81	1.15	4517.64	39.15	1.18	1.98	31.14	-0.12
	1725	1751.16	37.11	0.89	1.52	1771.34	37.01	1.96	2.69	54.41	0.10
	790.5	794.93	35.02	1.04	0.56	800.23	34.98	1.89	1.23	44.79	0.04
	390	392.57	32.95	0.96	0.66	396.79	32.89	2.02	1.74	52.42	0.06
PoznanHall2-Dancer	3955	4005.28	39.70	0.43	1.27	4040.47	39.76	0.54	2.16	20.84	-0.06
	1655	1666.55	37.84	0.52	0.70	1688.73	37.79	0.68	2.04	23.95	0.05
	780	792.40	35.95	0.49	1.59	799.58	35.88	0.58	2.51	15.05	0.07
	380	381.95	33.92	0.39	0.51	387.08	33.86	0.56	1.86	30.08	0.06
PoznanStreet-Kendo	2705	2746.71	40.18	0.67	1.54	2771.60	40.16	0.82	2.46	17.74	0.02
	1060	1079.67	38.49	0.73	1.86	1087.09	38.45	0.93	2.56	21.32	0.04
	515	517.85	36.52	0.74	0.55	523.77	36.64	1.03	1.70	28.63	-0.12
	275	276.69	34.16	0.64	0.62	280.51	34.01	1.10	2.01	41.43	0.15
Dancer-Newspaper	4110	4161.35	38.92	0.38	1.25	4213.96	39.03	0.45	2.53	14.98	-0.11
	1790	1814.26	36.44	0.42	1.36	1825.72	36.36	0.49	2.00	14.68	0.08
	855	867.05	33.99	0.40	1.41	873.97	34.03	0.47	2.22	15.57	-0.04
	425	432.90	31.67	0.40	1.86	438.04	31.63	0.59	3.07	31.52	0.04
Average	1242.34	1258.33	37.08	0.58	1.32	1270.73	37.08	0.84	2.38	25.29	0.00

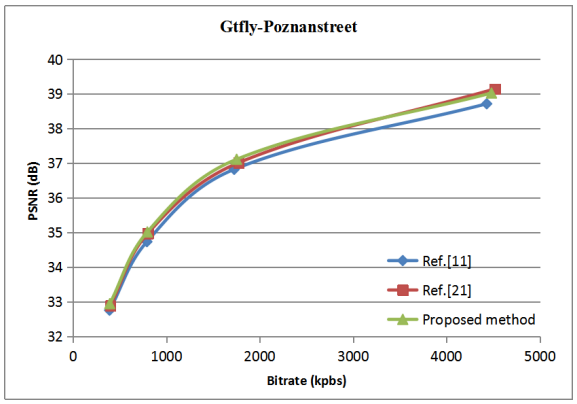
sequences with various target bit rates. Due to the limited length of the paper, this paper only considers the first point of scene change. Table 3 shows the simulation results of the CTU layer rate control algorithm based on scene detection, which was compared with [21]. The rate control errors of the two algorithms calculated from Table 3 are 1.32% and 2.38%, respectively. Compared with [21], the rate control algorithm proposed in this paper has a more accurate rate control and a smaller rate deviation. Figure 2 shows the PSNR fluctuations of the Dancer-Newspaper sequence. The method in [21] is low in accuracy and has large fluctuations, and the PSNR variation of the proposed rate control algorithm is smaller compared to that of the other two algorithms, thus ensuring a smooth transition of image quality and no visual detection of obvious image quality changes. The target bit allocation scheme is obtained according to the average allocation and does not take into account the actual complexity of the viewpoints. In fact, the complexity of each viewpoint in the sequence is different, and the bits required for each viewpoint are also different. In particular, it is necessary to



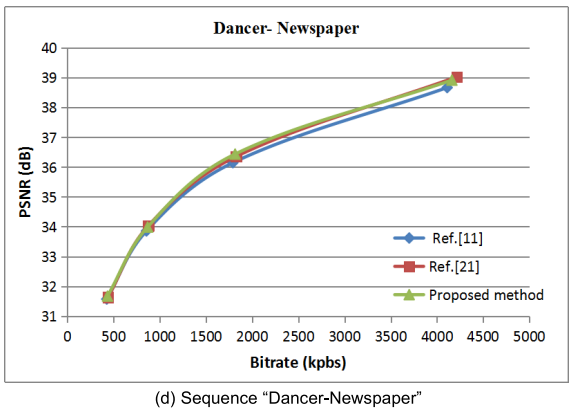
(a) Sequence "Balloon-Newspaper"



(b) Sequence "Kendo-Balloons"



(c) Sequence "Gtfly-Poznanstreet"



(d) Sequence "Dancer-Newspaper"

FIGURE 3. RD-curve of sequences.

allocate more bits for the scene change viewpoints and the motion viewpoints and fewer bits for without changing scenes and the smooth-motion viewpoints and frames.

Figure 3 shows the experimental PSNR results at various target bit rates. After each scene change frame (frame numbers 60, 120, 180, 240, 360 and 480), the PSNR of the [21] rate control algorithm will drop sharply, and the proposed scheme slows this drop significantly. This is mainly because the (i-1)th frame and the i-th frame are greatly different when the scene change occurs at the i-th frame, and thus most of the CTUs in the i-th frame cannot be predicted by the previous frame and will be in intraframe mode. Since the intraframe mode produces more bits than the interframe mode, which seriously deviates from the target buffer occupancy, [21] greatly reduces the number of allocated bits in subsequent frames until the target is reached. The buffer occupancy results in a serious degradation in the quality of subsequent frames, perhaps dozens of frames. Therefore, the scene detection algorithm based on image similarity analysis is proposed to adaptively adjust the GOP encoding length. It is clear that the PSNR value of the scene change frame is smoother than that of the [21] algorithm, and the PSNR value of each frame does not appear to be ''up and down''. The overall quality of the video sequence of the proposed algorithm has been greatly improved, indicating that the proposed optimization algorithm is effective.

Figure 3 shows the experimental results of the sequences. Compared to [11] and [21], the proposed method can effectively control the bit-rate of MV-HEVC and while maintaining a high coding efficiency.

In summary, the proposed rate control algorithm is more accurate than that of [21]. For example, the code rate is more accurate, the code rate deviation is smaller, and the PSNR fluctuation is smaller, which greatly improves the visual effect. The simulation results show that the proposed basic unit layer rate control algorithm can not only greatly reduce video quality fluctuations but also decrease the average fluctuation of the PSNR by 25.29%. The reasons for the excellent performance are that reasonable bit-rate control is performed among viewpoints and that rate control is performed well at the frame layer and the basic unit layer.

VI. CONCLUSION

For MV-HEVC rate control, the current research has not been carried out in depth. After analyzing the characteristics of the existing rate control and multiview video coding, a new rate control algorithm based on scene detection is presented in this paper. The algorithm basically involves the entire rate control process, thus ensuring the accuracy of the MV-HEVC rate control algorithm. The experimental results show that the bit allocation and the rate control algorithm for 3D video coding can effectively control the rate based on the given coding parameters. In the future, we will further study the correlation between the viewpoints to improve the rate control algorithm more effectively. Exploring the perceptual characteristics of the human eye to different viewpoints and the hiding characteristics of the human eye to the auxiliary viewpoints, different coding methods and corresponding code rate allocation strategies can be adopted for different viewpoints in the

coding structure to improve coding efficiency and maintain the interview points. The fluctuation of video quality is limited to the scope of user perception or system requirements. That is, to ensure video quality balance between viewpoints; to explore the visual field, time complexity of multi-source video and human visual perception of different views, and to explore target bit allocation scheme that is reasonable in different viewpoints and time domains.

ACKNOWLEDGMENT

The authors would like to thank the editors and the reviewers for their professional suggestions.

REFERENCES

[1] L. Shen, K. Li, G. Feng, P. An, and Z. Liu, "Efficient intra mode selection for depth-map coding utilizing spatiotemporal, inter-component and inter-view correlations in 3D-HEVC," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4195–4206, Sep. 2018.

[2] D. Liu, P. An, R. Ma, W. Zhan, and L. Ai, "Scalable omnidirectional video coding for real-time virtual reality applications," *IEEE Access*, vol. 6, pp. 56323–56332, 2018.

[3] J. Lei, J. Duan, F. Wu, N. Ling, and C. Hou, "Fast mode decision based on grayscale similarity and inter-view correlation for depth map coding in 3D-HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 3, pp. 706–718, Mar. 2018.

[4] G. Tech, Y. Chen, K. Muller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016.

[5] C. Lin, Y. Zhao, J. Xiao, and T. Tillo, "Region-based multiple description coding for multiview video plus depth video," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1209–1223, May 2018.

[6] D. Liu, P. An, R. Ma, and L. Shen, "Hybrid linear weighted prediction and intra block copy based light field image coding," *Multimed Tools Appl.*, vol. 77, no. 24, pp. 31929–31951, Dec. 2018.

[7] J.-Y. Lee, J.-K. Han, J.-G. Kim, and T. Q. Nguyen, "Efficient inter-view motion vector prediction in multi-view HEVC," *IEEE Trans. Broadcast.*, vol. 64, no. 3, pp. 666–680, Sep. 2018.

[8] X. Liu, Y. Li, D. Liu, P. Wang, and L. T. Yang, "An adaptive CU size decision algorithm for HEVC intra prediction based on complexity classification using machine learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 144–155, Jan. 2019.

[9] L. Zhu, Y. Zhang, S. Kwong, X. Wang, and T. Zhao, "Fuzzy SVM-based coding unit decision in HEVC," *IEEE Trans. Broadcast.*, vol. 64, no. 3, pp. 681–694, Sep. 2018.

[10] Y. Liu and Y. F. Zheng, "Soft SVM and its application in video-object extraction," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3272–3282, Jul. 2007.

[11] T. Ikai, K. Kei, and T. Suzuki, *JCT3V-MV-HEVC and 3D-HEVC Conformance Draft 4*, documnet ISO/IEC JTC 1/SC 29/WG, JCT3V-N0008, San Diego, CA, USA, vol. 11, Mar. 2016, pp. 22–26.

[12] M. Zhou, X. Wei, S. Wang, S. Kwong, C.-K. Fong, P. H. W. Wong, W. Y. F. Yuen, and W. Gao, "SSIM-based global optimization for CTU-level rate control in HEVC," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 1921–1933, Aug. 2019.

[13] 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.

[14] S. Li, M. Xu, Z. Wang, and X. Sun, "Optimal bit allocation for CTU level rate control in HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2409–2424, Nov. 2017.

[15] 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.

[16] L. Li, B. Li, D. Liu, and H. Li, "λ-domain rate control algorithm for HEVC scalable extension," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2023–2039, Oct. 2016.

[17] H. Amer and E.-H. Yang, "Scene-based low delay HEVC encoding framework based on transparent composite modeling," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 809–813.

[18] Y. Eom, S. Park, S. Yoo, J. S. Choi, and S. Cho, "An analysis of scene change detection in HEVC bitstream," in *Proc. IEEE 9th Int. Conf. Semantic Comput.*, Feb. 2015, pp. 470–474.

[19] J. Lim, J. Kim, K. Ngan, and K. Sohn, "Advanced rate control technologies for 3D-HDTV," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 1498–1507, Nov. 2003.

[20] D. Fani and M. Rezaei, "Novel PID-fuzzy video rate controller for high-delay applications of the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1379–1389, Jun. 2018.

[21] H. Roodaki, Z. Iravani, M. R. Hashemi, and S. Shirmohammadi, "A View-Level Rate Distortion Model for Multi-View/3D Video," *IEEE Trans. Multimedia*, vol. 18, no. 1, pp. 14–24, Jan. 2016.

[22] T. Yan and I. H. Ra, "Bit allocation algorithm based on SSIM video coding," *Int. J. Performability Eng.*, vol. 15, no. 7, p. 1813, Jul. 2019.

[23] B. B. Vizzotto, B. Zatt, M. Shafique, S. Bampi, and J. Henkel, "A model predictive controller for frame-level rate control in multiview video coding," in *Proc. 12nd IEEE Int. Conf. Multimedia Expo*, Melbourne, VIC, Australia, Jun. 2012, pp. 485–490.

[24] F. Shao, G. Jiang, W. Lin, M. Yu, and Q. Dai, "Joint bit allocation and rate control for coding multi-view video plus depth based 3D video," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1843–1854, Dec. 2013.

[25] J. Xiao, M. M. Hannuksela, T. Tillo, M. Gabbouj, C. Zhu, and Y. Zhao, "Scalable bit allocation between texture and depth views for 3-D video streaming over heterogeneous networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 1, pp. 139–152, Jan. 2015.

[26] L. Fang, N.-M. Cheung, D. Tian, A. Vetro, H. Sun, and O. C. Au, "An analytical model for synthesis distortion estimation in 3D video," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 185–199, Jan. 2014.

[27] G. Pan and X. Wei, "Rate-distortion optimized mode switching for error-resilient multi-view video plus depth based 3-D video coding," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 1797–1808, Nov. 2014.

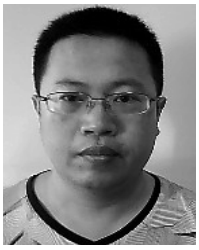
[28] X. Wang, S. Kwong, H. Yuan, Y. Zhang, and Z. Pan, "View synthesis distortion model based frame level rate control optimization for multiview depth video coding," *Signal Process.*, vol. 112, pp. 189–198, Jul. 2015.

[29] J. Lei, X. He, H. Yuan, F. Wu, N. Ling, and C. Hou, "Region adaptive R-λ model-based rate control for depth maps coding," *IEEE Trans. Circuits Syst. Video Technol*, vol. 28, no. 6, pp. 1390–1405, Jun. 2018.

[30] T. Yan, P. An, L.-Q. Shen, and Z.-Y. Zhang, "Bit allocation and rate control algorithm for MVC," *Imag. Sci. J.*, vol. 59, no. 4, pp. 202–210, Aug. 2011.

TAO YAN was born in 1981. He received the Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China, in 2010. He has been with the faculty of the School of Information Engineering, Putian University, where he is currently an Associate Professor. He has authored or coauthored more than 20 refereed technical articles in international journals and conferences in the field of video coding and image processing. His major research interests include multiview high efficiency video coding, rate control, and video codec optimization.

IN-HO RA received the Ph.D. degree in computer engineering from Chung-Ang University, Seoul, South Korea, in 1995. He has been with the faculty of the School of Computer, Information and Communication, Kusan National University, where he is currently a Professor. His major research interest includes block chain.



HUI WEN was born in 1981. He received the Ph.D. degree from the College of Information Engineering, Shenzhen University, China, in 2018. He is currently a Lecturer with the School of Information Engineering, Putian University, China. His current research interests include machine learning and neural networks.

MIN-HANG WENG was born in Kaohsiung, Taiwan, in 1971. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, in 1994, 1996, and 2000, respectively. His research interests include the design and fabrication of RF passive and active circuits, image processing, and solar cells.

QIAN ZHANG was born in 1982. She received the Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China, in 2011. She has been with the faculty of the School of Information and Electromechanical Engineering, Shanghai Normal University, where she is currently an Associate Professor. She has authored or coauthored more than 30 refereed technical articles in international journals and conferences in the field of video coding and image processing. Her major research interests include multiview high efficiency video coding.

YAN CHE was born in 1970. She received the M.E. degree in computer application from Xiamen University, Xiamen, China, in 2005, and the Ph.D. degree in control theory and engineering from Donghua University, Shanghai, China, in 2016. She is currently a Professor with Putian University. Her research interests include intelligent information processing and analysis, networked control systems, robust control/filter theory, and their applications.