

# Rate Control Algorithm for Multiview Video Coding based on Human Visual Characteristics

Tao Yan<sup>a,\*</sup>, In-Ho Ra<sup>b</sup>, Qiuwen Zhang<sup>c</sup>, Hui Wen<sup>a</sup>, Hang Xu<sup>a</sup>, and Shuqing Chen<sup>a</sup>

<sup>a</sup>*School of Information Engineering, Putian University, Putian, 351100, China*

<sup>b</sup>*School of Computer Information and Communication Engineering, Kunsan National University, Gunsan, 54150, South Korea*

<sup>c</sup>*School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China*

---

## Abstract

In order to satisfy transmission and application, this paper illustrates the drawback of the current rate-distortion model and the characteristics of multi-view video coding (MVC). This paper proposes a bit allocation and rate control scheme for MVC based on frame complexity and human visual characteristics. Firstly, the proposed algorithm improves the quadratic rate-distortion (R-D) model. Then, the proposed algorithm reasonably allocates bit-rate among views based on frame complexity and human visual characteristics. This paper turns the bit allocation among views into a multi-objective optimization problem. Finally, this paper proposes a view layer and frame layer bit allocation algorithm for MVC according to frame complexity and the previously coded information. Simulation results show that the proposed algorithm can effectively control bit rate for MVC, where the average rate control error of our algorithm is 1.08%.

**Keywords:** MVC (multi-view video coding); rate control; human visual characteristics; frame complexity

(Submitted on May 11, 2018; Revised on June 16, 2018; Accepted on July 17, 2018)

© 2018 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

The multi-view video is a new generation of stereoscopic and interactive video. The most prominent features of three-dimensional television (3DTV) are interactivity and stereoscopic effect. Free viewpoint television (FTV)/3DTV has been widely used in the 3D film, advertisements, 3D games, and other fields [1-2]. However, the huge amount of original data of multi-view video brings enormous challenges to its storage, compression and transmission. Therefore, MVC has been intensively studied.

Rate control is an important part of the video coding standard. Any video compression standard leaving the rate control will be limited. It mainly controls the encoding parameters to make the generated bit stream conform to the application requirements. It not only maintains efficient coding efficiency, but also makes the subjective quality between the frames before and after the sequence has no significant fluctuations. Previous video coding standards have their own rate control models such as MPEG-2, MPEG-4, H.263, H.264, MVC and HEVC [3-7].

In China, we started studying rate control for MVC relatively late. However, in recent years, some research teams have begun to study rate control. For example, Tsinghua University, Shanghai University, Zhejiang University, Ningbo University, and Tianjin University have also achieved certain research results for MVC. Among them, the team led by Prof. Zhang Zhaoyang from Shanghai University has done a lot of research on MVC and established related experimental methods and experimental platforms. In general, the visual quality of the entire image is largely determined by the visual quality of the area of interest, while the degradation of the area is often not easily noticeable and has little effect on the visual quality of the entire image. Studies have shown that our visual attention is often influenced by two types of factors: cognitive triggers (advanced factors) and irritating triggers (lower factors). Cognitive triggering is the recognition of objects and patterns by the human brain in terms of cognition (prior knowledge or experience). For example, when observing an image, it gives the observer some hints or instructions. Since cognitive triggers cannot be simply simulated by a model, it is

\* Corresponding author.

E-mail address: [yantaoshu@aliyun.com](mailto:yantaoshu@aliyun.com)

difficult to implement in practical applications. In this paper, the most correlated primary view is taken as the view of the attention of our human eye. The target bit of the main view is increased, and the target bit of the non-main view is lowered, thereby improving the subjective quality of the image.

In multi-view video coding, the parallax prediction characteristics among views have a large difference. Therefore, coded images with the same inter-view prediction relationship may have a different frame complexity. At this time, there will be a certain deviation in the number of target bits calculated by using the same model parameters. To improve the coding quality and reduce the computational complexity, Z Zhao et al [8] propose a rate control algorithm based on the complexity of the video content per frame. Lei et al. propose a rate control algorithm based on R- $\lambda$  model for 3D-HEVC [9]. However, the computer's complexity was too high, which was necessary to further study the optimal model of 3D-HEVC bit rate control. Therefore, the coded images with the same inter-view prediction relationship or inter-temporal prediction relationship may have different coding characteristics. In this case, the target bits calculated by the same model parameters cannot be used.

Yan et al. propose rate control for MVC, but the RCE is too large [10-11]. This paper proposes a rate control method for MVC based on frame complexity and human visual characteristics. Firstly, the proposed algorithm improves the traditional rate control model. Then, the proposed algorithm reasonably allocates bit-rate among views based on frame complexity and human visual characteristics. This paper turns the bit allocation among views into a multi-objective optimization problem. To this end, this paper first analyzes the advantages and disadvantages of existing rate control algorithms. Then this paper proposes to improve the traditional rate control model according to the characteristics and requirements of MVC. Simulation results show that the proposed algorithm can effectively control bit rate for MVC, with the average rate control error of our algorithm is 1.08%.

## 2. Structure of Rate Control

Multiview Video is a new type of video with three-dimensional stereoscopic and viewpoint interaction functions that can meet the user's ability to select and operate audiovisual objects from multiple angles and provide 360-degree scene roaming interactive capability. However, the compression efficiency of existing coding standards is still insufficient for high-definition, ultra-high-definition video applications, and it still requires a more efficient coding compression scheme. Rate control algorithm for MVC is very complicated. Based on the JVT MVC encoder, the proposed RC algorithm includes 4 layers, as shown in Figure 1, including GGOP (a group of GOP), GOP (group of pictures), the frame layer, and the basic-unit layer.

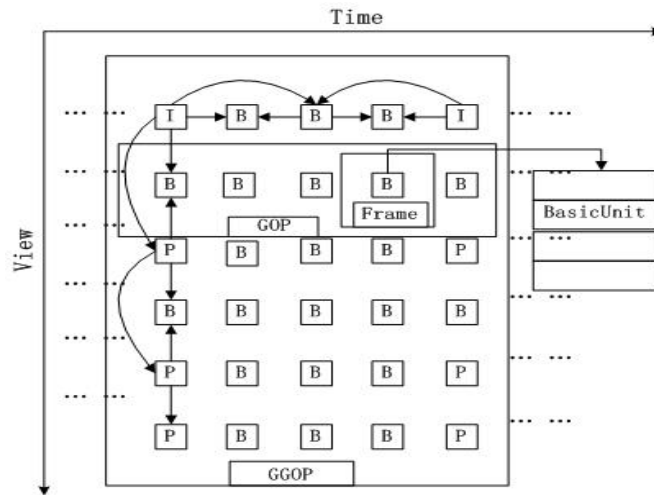


Figure 1. 4-Layer structure of rate control

## 3. Improved R-D Model

The second multi-view coding-specific issue we considered was MAD prediction. This paper first analyzes the advantages and disadvantages of the existing rate control algorithms. Then, this paper proposes to improve the traditional rate control model according to the characteristics and requirements of MVC. The improved MAD prediction is described as follows.

The quadratic rate-quantization (R-Q) model adopted by MVC is as follows:

$$\frac{R_i - H_i}{\sum_{i=1}^N \lambda_i \cdot MAD_i} = \frac{x_i}{Q_i} + \frac{y_i}{Q_i^2} \quad (1)$$

MAD prediction for a macroblock in I frame: In multi-view coding (MVC), we also have the option of using the previously-encoded neighbouring view for MAD prediction. We adopted a criterion for I frame. The improved MAD prediction is described as follows:

$$MAD_i = \sum_{(x,y)} abs(I_{x,y} - \mu_{block(x,y)}) \quad (2)$$

Where  $I_{x,y}$  denotes image gray value of pixel  $(x, y)$  and  $\mu_{block(x,y)}$  denotes image gray value of the  $8 \times 8$  block that the pixel belongs to.

#### 4. Rate Control Algorithm for MVC

Compared with bit allocation and the rate control algorithm for H.264, the rate control algorithm for MVC is very complicated. The existing rate control in H.264 creates a bitstream with the bitrate that matches the available channel bitrate for a single-view video. Based on the work in [12], we should consider the initial QP value of the first frame of the extended views. The main steps of the rate control algorithm for MVC are as follows:

##### 4.1. View Level Rate Control

People will unconsciously be interested in certain areas when they observe them and call these areas “visual attention areas”. In general, the visual quality of the entire image is largely determined by the visual quality of the area of interest, while the degradation of the area is often not easily noticeable and has little effect on the visual quality of the entire image.

Studies have shown that our visual attention is often influenced by two types of factors: cognitive triggers (advanced factors) and irritating triggers (lower factors). Cognitive triggering is the recognition of objects and patterns by the human brain in terms of cognition (prior knowledge, experience). For example, when observing an image, give the observer some hints or instructions. Since cognitive triggers cannot be simply simulated by a model, it is difficult to implement in practical applications. In this paper, the most correlated primary view is taken as the view of the attention of our human eye. The target bit of the main view is increased, and the target bit of the non-main view is lowered, thereby improving the subjective quality of the image.

Although the first frames of extended views do not have a temporal reference, they are still encoded as inter frames, not intra frames, because they can use already-encoded initial frames of the base view and potentially other already-encoded views. The multi-view video is taken from different angles by the camera array. It contains multiple viewpoint images in the same scene, so there is information redundancy among the viewpoints. Therefore, compared with the traditional single-view video coding, in addition to the use of space-time redundancy to achieve compression coding purposes, it is more important to eliminate spatial prediction redundancy among different views.

Based on this rate model, this paper proposes a rate control and bit allocation algorithm for MVC based on human visual characteristics and frame complexity. We first study the HVS (Human visual system) characteristics, and then consider the sensitive and non-sensitive regions of the human eye for more reasonable code rate allocation. Finally, we increase the number of allocated bits in sensitive regions and reduce the number of allocated bits in non-sensitive regions. The results show that the algorithm can improve the quality of sensitive areas and improve the subjective visual effects. We can put the three-dimensional video point of view between the bit allocation, so the problem boils down to the cost equation for the rate-distortion optimization problem.

The viewpoint bit allocation problem can be expressed with the following formula:

(i) Maximum:

$$Q = \frac{1}{SG} \sum_{s=0}^{S-1} \sum_{g=1}^G Q_{sg} \quad (3)$$

(ii) Restrictions:

$$(R_i + \sum_{s=0, s \neq I}^{S-1} R_s) \leq R_{Total} \quad (4)$$

$$R_i = \sum_{g=1}^G R_{i,g} \quad (5)$$

$$R_i = \alpha_i \cdot R_{Total} \quad (6)$$

$$\sum_{i=0}^S \alpha_i \quad (7)$$

Where,  $Q_{sg}$  is the  $g$ th image quality in the  $s$ th viewpoint,  $Q$  represents the average image quality.  $R_{Total}$  is the actual bits generated by the  $i^{\text{th}}$  GOP.  $R_i$  and  $R_s$  are the actual bits generated by the main viewpoint and auxiliary viewpoint, respectively. Parameter  $\alpha_i$  obtains a local optimal solution by precoding the  $i^{\text{th}}$  GOP. Then, we solve the optimal value by precoding several sets of GOP.

#### 4.2. Target Bit Rate for Frames

Multi-view video coding employs a hierarchical B-frame prediction structure. In the hierarchical B frame prediction structure, B frames of different time layers are independent of each other. Therefore, the frame layer rate control must solve the B-frame bit allocation and frame complexity of different time layers. Lei et al. [13] proposed the following optimal frame target bit allocation method according to the residual energy of the coded frame.

$$T(j) = \frac{MAD_j}{MAD_a} \cdot \frac{(T - \sum_{m=1}^M C_m)}{M} + C_j \quad (8)$$

All parameter values are shown in Reference [13]. In general, the more intense the video sequence is, the larger the scene change of the image content of each frame is, and the more bits are needed for video encoding. On the contrary, the smaller the image content scene of each frame is, the less bits are needed for video encoding. In order to improve rate control error, the current frame target bit is calculated by Equation (9).

$$T_r(j) = \left[ \frac{MAD_{j-1}}{MAD_a} \cdot \phi + \zeta \right] \cdot \frac{MAD_{j-1} \cdot FD(j)^2}{MAD_a \cdot \frac{1}{j-1} \cdot \sum_{k=1}^{k=j-1} FD(k)^2} + \eta \quad (9)$$

In Equation (9),  $FD(j)$  is the temporal activity for the  $j^{\text{th}}$  frame.  $FD(j-1)$  is the temporal activity for the  $(j-1)^{\text{th}}$  frame.  $\zeta$  and  $\eta$  are experimental parameters.

#### 4.3. Macroblock Layer Rate Control

The basic unit can include the entire frame image at the maximum and only contains one macro block at the minimum. Multi-view video coding employs a hierarchical B-frame prediction structure. In the hierarchical B frame prediction structure, B frames of different time layers are independent of each other. Therefore, the macroblock layer rate control must solve the basic unit complexity of different time layers. For I-frames, the use of the same quantization parameter for the entire frame generally does not require the basic unit layer rate control. The following focuses on bit rate control of the basic unit layer of a B-frame or P-frame. The main point of this is: first, the number of target bits is determined for each B-frame or P-frame. Then, assign these bits to each basic unit, and finally, select the quantization parameter according to the binomial model.

The macroblock layer rate control first considers the local motion degree, texture activity, edge strength of the macroblock, and quickly partitions the local motion macroblock, edge macroblock, texture macroblock, and smooth macroblock. Then, based on the difference of the degree of attention of human eyes in various types of macroblocks, this paper adjusts the quantization parameters of macroblocks.

In order to be compatible with H.264, the macroblock layer rate control scheme for MVC is similar to H.264. In order to be compatible with H.264, the macroblock layer rate control scheme for MVC is simplified by dividing the allocated bits of each frame evenly to each basic unit layer of the frame. However, even macroblocks in the same basic unit have great differences in image content, texture, activity time domain, and complexity. Therefore, to determine the bit rate more accurately, different quantization values are used based on the difference of the degree of attention of human eyes in various types of macroblocks. This paper adjusts the quantization parameters of macroblocks. The RC method of the basic unit is similar to Reference [11]. The detailed block diagram is shown in Figure 2.

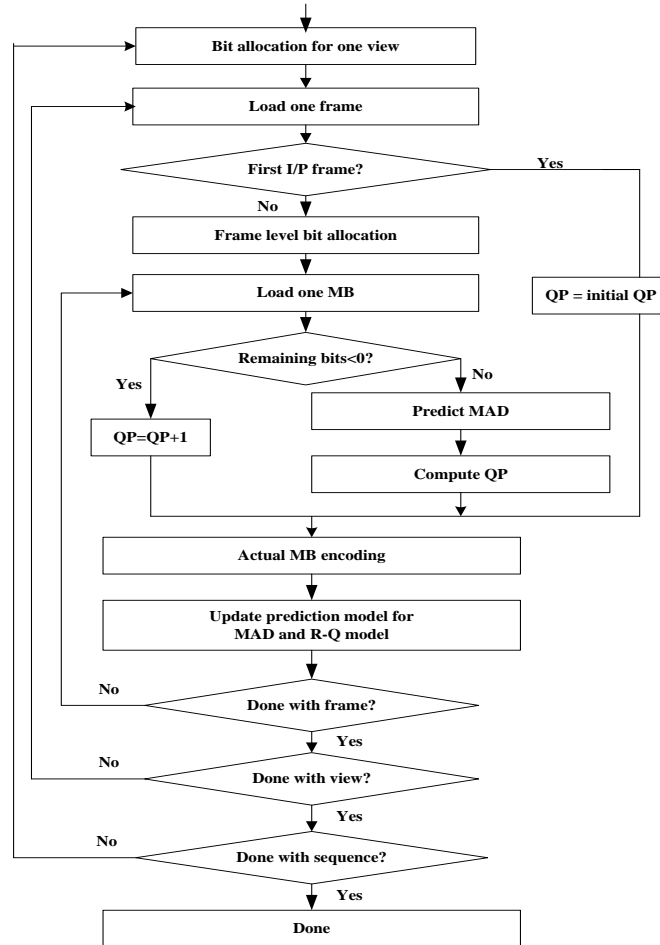


Figure 2. Proposed macroblock-layer bit allocation and rate control algorithm

## 5. Experimental Results

In order to test the effectiveness of the algorithm, the experiments were conducted using the test sequences Kendo, Newspaper, Dancer, PoznalHall2, Lovebird1 and Gtfly provided by the HHI Institute and Nagoya University/Tanimoto Lab. The resolution of the sequence includes 1920 x 1088 pixels or 1024 x 768 pixels. Other coding parameters of the algorithm mainly include entropy coding CABAC, RDO enable, reference frame 2, fast search, and search range 64 pixels. The test platform is an Intel(R) Duo CPU 2.40GHz (dual) with 8GB of memory.

In the latest test platform, this algorithm is compared with the previous two algorithms. In this paper,  $\sigma$  is used to denote the rate control error (REC), which is obtained from Equation (10).

$$\sigma = \frac{|R_{target} - R_{actual}|}{R_{target}} \times 100\% \quad (10)$$

Where  $R_{target}$  and  $R_{actual}$  denote the target bit-rates and the actual coding bit-rates of the sequence, respectively.

Figures 3-6 shows the experimental results for Newspaper and Dancer. The experimental results show that the target bit curve assigned by each frame of the algorithm is consistent with the frame complexity. Motion strenuous frames allocate a relatively large number of bits, and motion smoothing frames allocate a correspondingly smaller number of bits. Compared with the average distribution of JVT-G012, this algorithm has a greater advantage. The experimental results show that this algorithm improves the signal-to-noise ratio of drastic motion frames and reduces the signal-to-noise ratio of flat frames. Therefore, the video quality fluctuations between frames are significantly improved.

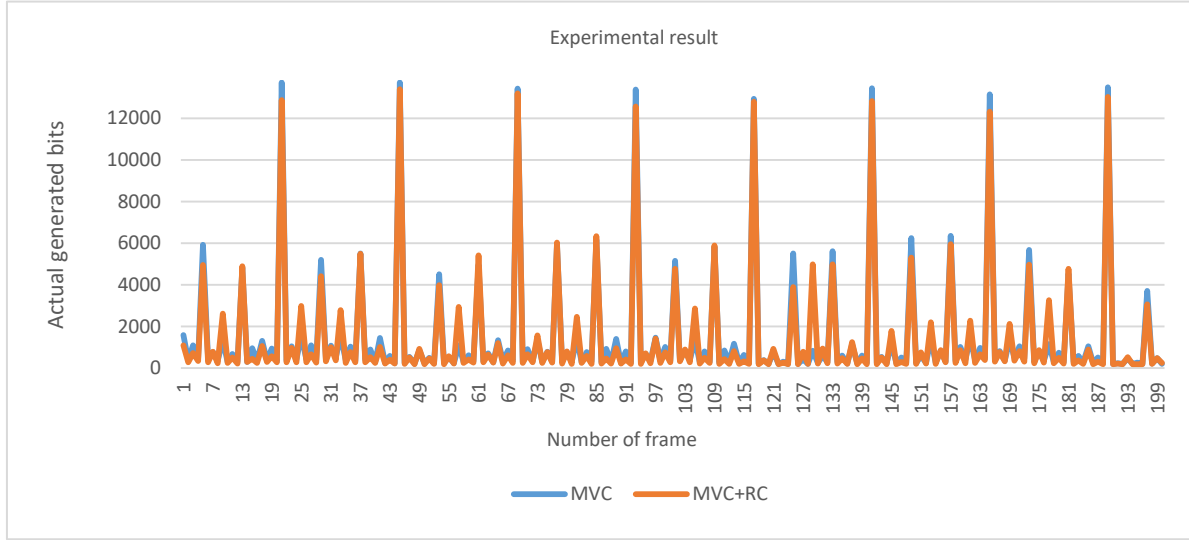


Figure 3. Experimental results of actual generated bits for Newspaper

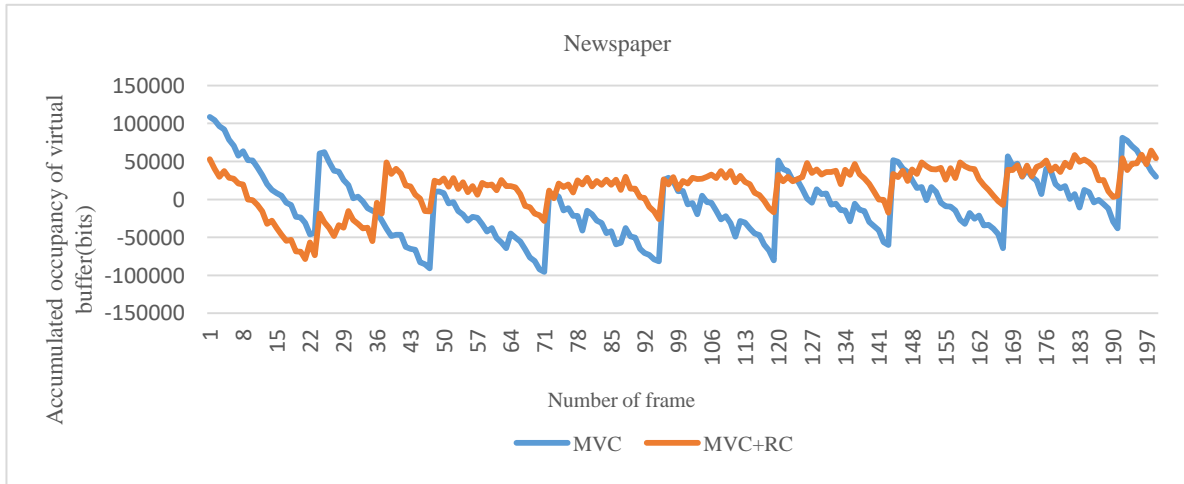


Figure 4. Experimental results of virtual buffer for Newspaper

Figure 7 shows the experimental results of the PoznanHall2 sequences. Compared with previous research results, the PSNR of the rate control algorithm proposed in this paper has improved. In summary, although the algorithm proposed in this paper can maintain high coding efficiency, the rate control deviation is relatively high. In order to get the bit rate more accurately, different quantization values are used based on the difference of the degree of attention of human eyes in various types of macroblocks. This paper adjusts the quantization parameters of macroblocks.

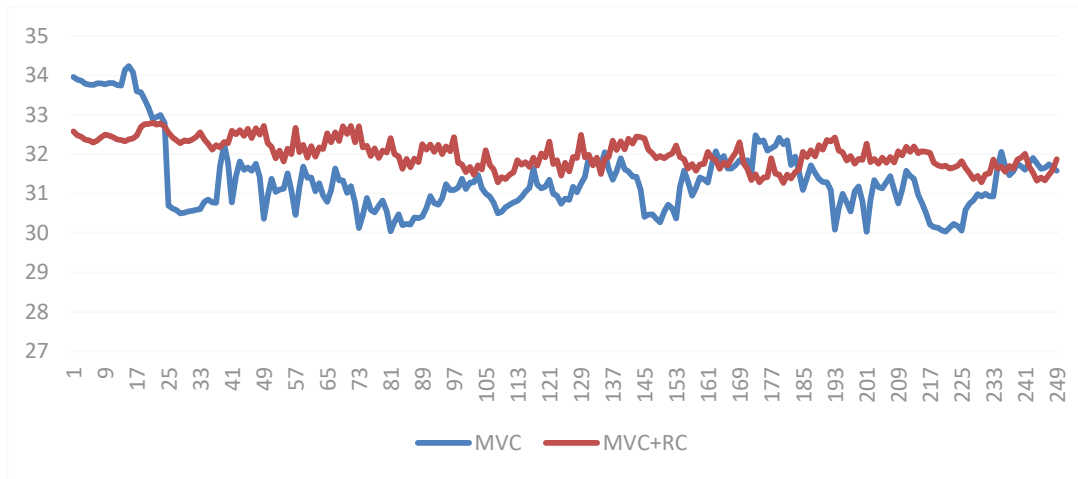


Figure 5. Experimental results for Newspaper

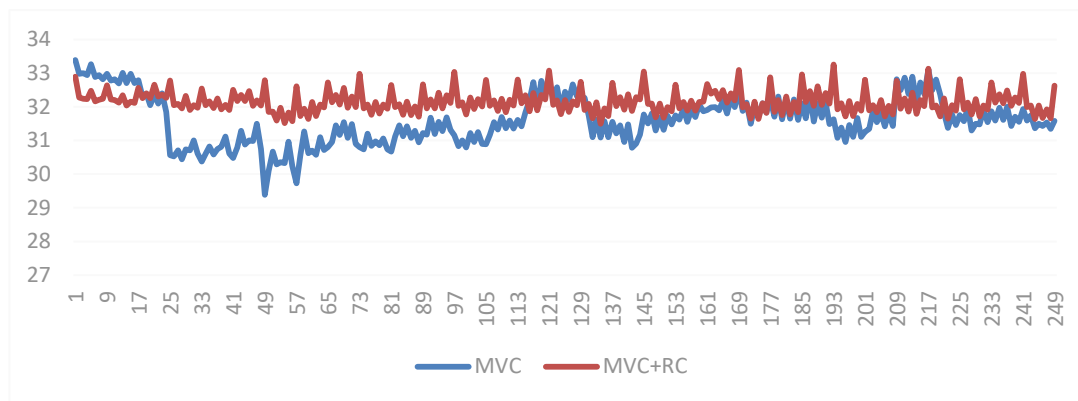


Figure 6. Experimental results for Dancer

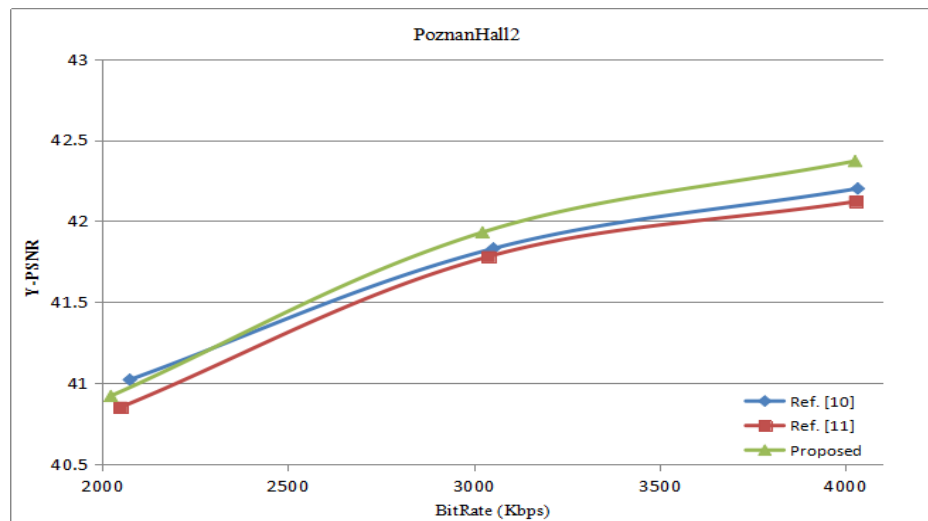


Figure 7. Experimental results for PoznanHall2

From Table 1, it can be seen that the average RCE of our algorithm is 1.08%. Among them, the error rate control of the Lovebird1 sequence is relatively large. The main reason may be that the movement of the Lovebird1 sequence is more intense. If the encoded information is used to predict the inter-view bit allocation, the prediction is inaccurate. Compared with the average distribution of JVT-G012, this algorithm has a greater advantage. The experimental results show that this algorithm improves the signal-to-noise ratio of drastic motion frames and reduces the signal-to-noise ratio of flat frames. Therefore, the video quality fluctuations between frames are significantly improved.

Table 1. Simulation results

Sequence	Target bits /(kbps)	Actual generated bits (kbps)			RCE (%)		
		Ref. [10]	Ref. [11]	Proposed	Ref. [10]	Ref. [11]	Proposed
Kendo	2000	2019.61	2018.82	2018.87	0.98	0.94	0.94
	3000	3028.17	3026.15	3019.02	0.94	0.87	0.63
	4000	4048.29	4033.51	4039.47	1.21	0.84	0.99
Dancer	1256	1277.42	1271.15	1273.42	1.71	1.21	1.39
	2736	2768.76	2758.47	2754.33	1.20	0.82	0.67
	6539	6642.45	6629.56	6603.18	1.58	1.38	0.98
Lovebird1	2000	2086.78	2048.51	2042.47	4.34	2.43	2.12
	3000	3115.56	3073.47	3035.03	3.85	2.45	1.17
	4000	4112.72	4072.52	4031.87	2.82	1.81	0.80
PoznanHall2	2000	2074.27	2049.59	2023.36	3.71	2.48	1.17
	3000	3052.32	3038.91	3022.82	1.74	1.30	0.76
	4000	4032.1	4026.47	4025.08	0.80	0.66	0.63
Gtfly	976	1016.34	1005.07	992.62	4.13	2.98	1.70
	2091	2147.49	2127.72	2115.72	2.70	1.76	1.18
	4949	5058.61	5026.28	5003.17	2.21	1.56	1.09
Average					2.26	1.57	1.08

In summary, although the algorithm proposed in this paper can maintain high coding efficiency, the rate control deviation is relatively high.

## 6. Conclusions

Stereoscopic videos are the principle of binocular parallax using the human eye. The binoculars independently receive left and right images of specific photographic points from the same scene. The left eye and the right eye view the respective corresponding images to form the binocular parallax and then merge the brain to obtain a stereoscopic image with a strong sense of depth and realism. However, the stereoscopic stereoscopic stereoscopic viewing angle range is too narrow, and it is difficult for multiple users to simultaneously view. To overcome the limitations of binocular stereo video system applications, it is extended to multi-view video systems. Compared with binocular stereo video, multi-view video can describe the scene structure more accurately, which can greatly expand the stereoscopic view angle and also facilitate the user to interactively select viewpoints, thus having a wider application. The multi-view video is taken from different angles by the camera array. It contains multiple viewpoint images in the same scene, so there is information redundancy among the viewpoints. Therefore, compared with the traditional single-view video coding, in addition to the use of space-time redundancy to achieve compression coding purposes, it is more important to eliminate spatial prediction redundancy among different views.

Currently, rate control and bit allocation algorithms for MVC are not yet mature. In order to reduce REC, this algorithm performs effective rate control for the inter-view, frame, and base unit layers. Among them, the error rate control of the Lovebird1 sequence is relatively large. The main reason may be that the movement of the Lovebird1 sequence is more intense. If the encoded information is used to predict the inter-view bit allocation, the prediction is inaccurate. In the future, we will further study the correlation among views so that we can improve the algorithm for MVC. In addition, multi-view video scene switching has not been considered in this paper. If the video scene switches, the algorithm will fail. This will also be one of the focuses of future research.

## Acknowledgments

This work was supported by the Natural Science Foundation of China (No. 61741111, No. 61771432, No. 61302118) and in part by the National Research Foundation of Korea (NRF), a grant funded by the Korean government (MSIP) (No. 2016R1A2B4013002), it was also supported in part by the Natural Science Foundation of Jiangxi (No. 20161BAB212031), the Fujian University New Century Talents Program Funding, the Funding Scheme of Young Key Teacher of Henan Province Universities (No. 2016GGJS-087), and the Putian University Nursery Fund Project (2014057).



## References

1. C. Yang, P. An, L. Shen, and D. Liu, "Adaptive Bit Allocation for 3D Video Coding," *Circuits Systems and Signal Processing*, Vol. 36, No. 5, pp. 2102-2124, May 2017
2. P. Ren, X. Zhang, H. Bi, H. Sun, and N. Zheng, "Toward an Efficient Multiview Display Processing Architecture for 3DTV," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 64, No. 6, pp. 705 - 709, June 2017
3. S. Ma, W. Gao, and Y. Lu, "Rate Control on JVT Standard," in *Proceedings of Joint Video Team of ISO/IEC and ITU-T the 4th Meeting*, pp. 22-26, Klagenfurt, Austria, July 2002
4. Z. Li, F. Pan, K. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive Rate Control for JVT," *Joint Video Team of ISO/IEC and ITU-T the 4th Meeting*, pp. 5-15, Awaji, Japan, December 2002
5. T. Yan, L. Shen, P. An, H. Wang, and Z. Zhang, "Frame-layer Rate Control Algorithm for Multi-view Video Coding," in *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, pp. 1025-1028, Shanghai, China, June 2009
6. T. Yan, P. An, L. Shen, and Z. Zhang, "Bit Allocation and Rate Control Algorithm for MVC," *Journal of the Imaging Science*, Vol. 59, No. 4, pp. 202-210, April 2011
7. T. H. Tsai, S. Su, T. Y. Lee, "Fast Mode Decision Method based on Edge Feature for HEVC Inter-prediction," *IET Image Processing*, Vol. 12, No. 5, pp. 644-651, April 2018
8. Z. Zhao, L. Shen, Q. Hu, and Z. Zhang, "Algorithm Optimization based on the Complexity of HEVC Rate Control," *Chinese Journal of Optoelectronics Laser*, Vol. 25, No. 9, pp. 1715-1720, September 2014
9. J. Lei, X. He, H. Yuan, F. Wu, N. Ling, and C. Hou, "Region Adaptive R- $\lambda$  Model based Rate Control for Depth Maps Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 28, No. 6, pp. 1390-1405, June 2018
10. T. Yan, P. An, L. Shen, Q. Zhang, and Z. Zhang, "Rate Control Algorithm for Multi-view Video Coding based on Correlation Analysis", in *Proceedings of 2009 Symposium on Photonics and Optoelectronics*, pp. 1-4, Wuhan, China, August 2009
11. T. Yan, P. An, L. Shen, Z. Li, H. Wang, and Z. Zhang, "Rate Control Algorithm based on Frame Complexity Estimation for MVC", in *Processings of SPIE International Conference on Visual Communications and Image*, pp. 77-85, Huang Shan, An Hui, China, July 2010
12. T. Yan, P. An, L. Shen, Z. Li, H. Wang, and Z. Zhang, "Frame-layer Bit Allocation for Multi-view Video Coding based on Frame Complexity Estimation," *Journal of Shanghai university*, Vol. 14, No. 1, pp. 50-54, February 2010
13. J. R. Corbera and S M Lei, "A Frame-layer Bit Allocation for H.263+," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 7, pp. 1154-1158, October 2000