Journal of Theoretical and Applied Information Technology
31st March 2013. Vol. 49 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RATE CONTROL FOR MULTI-VIEW VIDEO CODING BASED ON VISUAL PERCEPTION

[1]YI LIAO, [1,2]MEI YU, [1]XIAODONG WANG, [1,2]GANGYI JIANG, [1]ZONGJU PENG, [1]FENG SHAO
[1] Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
[2]National Key Lab of Software New Technology, Nanjing University, Nanjing, China
E-mail: [1]liaoyi518@163.com, [2]jianggangyi@126.com

ABSTRACT

A rate control algorithm for multi-view video coding based on visual perception is proposed. The algorithm is performed on four levels, namely view level, group-of-picture (GOP) level, frame level and macro-block level. In the view level, we pre-encode a GOP to obtain the bitrates proportion among the views. In the GOP level, the initial quantization parameter and the target bits are calculated for the GOP combined with the feature of hierarchical B pictures (HBP). In the frame level, bits allocation is decided by the predicted complexity of the frame. In the macro-block level, the rate distortion model is adjusted based on the visual perception. Experimental results show that the proposed algorithm can increase the quality of visual sensitive region by 0.18 dB to 0.54 dB and improve the subjective quality compared with the conventional approach.

Keywords: *Rate control, Multi-view video coding, Visual perception, Hierarchical B pictures*

1. INTRODUCTION

Multi-view video is a development direction of the new generation multi-media network [1]-[2]. It is obtained by multiple cameras shooting the same scene from different viewpoints which can bring audiences a feeling of reality to the video scene. However, the huge amount of data makes it difficult for multi-view video to be widely used. The Joint Video Team (JVT) presents the Multi-view Video Coding (MVC) and the joint multi-view video model (JMVM). MVC mainly researches on motion compensation and disparity compensation [3], encoding structure [4], virtual viewpoint synthesis [5] and so on.

Rate control in MVC is still lacking in maturity and no rate control standard has yet been defined for multi-view video. However, some researches of rate control for multi-view video have been published. Kamolrat et al. [6] investigated the effect of bitrates of color and depth on the quality of virtual view generation in depth image-based rendering, then went through all the candidate pair of QP of color and depth as to find the optimum QPs to encode. Liu et al. [7] proposed a jointly coding method and set the target rate proportion between depth and video for 1:4, which can reduce the virtual view PSNR fluctuation between consecutive frames for multi-view video rate control. Yan et al. [8] improved the encoding structure of the JMVM and achieved the continuous controlling of bitrates among viewpoints, then allocated the bits rationally based on the correlation of the inter-view.

This work introduces a rate control algorithm for multi-view video coding based on visual perception. The two referenced frames information are used to predict the complexity of current frame as to allocate bitrates rationally in the frame level. In macro-block level, we extract motion region and segment the foreground and background firstly, then we set four regions based on the different degrees of importance for human visual to adjust the Lagrange multiplier. The proposed rate control method can accurately control the bitrates and provide a good visual effect for multi-view video.

The rest of the paper is organized as follows. In Section 2, related works is introduced. In Section 3, we briefly describe the visual importance division based on the depth and motion characteristic. A rate control algorithm for multi-view video based on the visual importance is proposed in Section 4. Then, the experimental results are analyzed in Section 5. Finally, the conclusions are given in Section 6.

2. RELATED WORKS

In MVC, rate distortion optimization (RDO) technique is adopted to choose the best encoding mode for macro-block (MB) in MVC. RDO goes through all the modes for inter-frame encoding and selects the mode of the least rate distortion cost to encode each MB. However, the RDO cannot choose the mode which is consistent with human visual perception, hence, a number of projects have worked on this problem. In [9], Lagrange multiplier was adjusted for motion estimation at the MB level based on the Lagrange costs of neighboring MBs. In [10], structural similarity index was used to optimize the rate distortion model which can be integrated with human perception. In [11], the Lagrange multiplier was adjusted according to the bitrates of the encoded macro-blocks which can control the rate more accurately.

Meanwhile, it can be known from the human visual characteristics that audiences tend to be more sensitive to the motion region than motionless region in the video [12], and be more interest in the foreground region than the background [13]. Obviously, it is imperative for us to design the rate control algorithm to allocate more bits for the regions of foreground and motion as to control the video more comfortable for audiences.

3. VISUAL ATTENTION REGION

3.1 Foreground Region Detection

The foreground and background are divided by the OSTU method. For a depth image, let t be a dynamic threshold, $g(t)$ be the inter-class variance, $w_0(t)$ and $w_1(t)$ be the probabilities of foreground and background, respectively, and $u_0(t)$ and $u_1(t)$ be the average luminance of foreground and background, respectively.

$$g(t) = w_{0(t)} \times w_{1(t)} \times \left(u_{0(t)} - u_{1(t)}\right)^2 \qquad (1)$$

Let the value of t go through all the gray scale. When $g(t)$ get the maximum, we define t is the best threshold. If the depth value is larger than t, it is defined as foreground.

3.2 Motion Region Detection

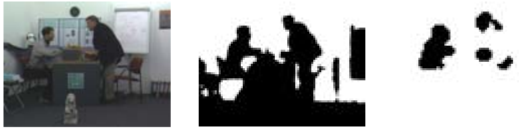
Let $D(j,n)$ be the mean absolute difference between the n-th MB in the j-th frame and the (j-1)-th frame, respectively, and $I_{(j,n)}(w,h)$ be the luminance of the pixel (w,h) in the n-th MB in the j-th frame. $D(j,n)$ can be calculated as

$$D(j,n) = \frac{1}{256}\sum_{h=0}^{15}\sum_{w=0}^{15}\left| I_{(j,n)}(w,h) - I_{(j-1,n)}(w,h) \right| \qquad (2)$$

When $D(j,n)$ is larger than threshold T_D, the current MB is defined as a moving MB and motionless otherwise, and T_D is a constant and its value is 2.5. However, in order to improve the accuracy of motion region detection, the MB of less than five motion pixels is defined as a motionless MB. In addition, if a motion MB is all around for motionless MB, it means that it is not obvious for the global and then it is defined as motionless MB.

3.3 Visual Attention Importance

The foreground and motion extraction results are shown in Figure 1.



(a)Original image *(b)Foreground region* *(c) Motion region*

Figure 1. Foreground and motion region for Leavelaptop

We define four regions based on the different degrees of importance for human visual, as shown in Table 1, where P denotes the importance. The larger of the P, the more important it is for human visual.

Table 1. Importance division based on the depth and motion characteristic

Depth Characteristic	Motion Characteristic	Importance
Foreground	Motion	P =3
	Motionless	P =2
Background	Motion	P =2
	Motionless	P =1

4. THE PROPOSED RATE CONTROL ALGORITHM FOR MVC

After the above discussion, we propose a new rate control algorithm for multi-view video encoding. It consists of four stages: 1) view level rate allocation; 2) group-of-pictures (GOP) rate allocation; 3) frame level rate allocation; 4) macro-block level rate allocation.

4.1 View Level Rate Allocation

In the hierarchical B pictures (HBP) structure, views are classified as I-view, B-view and P-view, these three views are on different degrees of importance for multi-view video encoding. In [8], a statistical method of pre-encoding several frames in each view is as to get the approximate rate proportion among views. In this paper, we pre-encode a GOP of 8 frames on the off-line stage and obtain the proportion to guide the online bits allocation for each view.

4.2 GOP Level Rate Allocation

In the GOP level, the total number of bits for each GOP and the QP of the key frame are calculated, In MVC, I-frame, P-frame and the first B-frame in B-view in each GOP are defined as key frames. Before encoding the *i*-th GOP, the bits allocated for the GOP are computed by

$$T_r(i,0)=\frac{u(i,1)}{F_r}*N_{gop}-(\frac{B_s}{8}-B_c(i-1,N_{gop})) \quad (3)$$

where $u(i,1)$ denotes the bandwidth, F_r denotes the frame rate of the video, N_{gop} denotes the total number of frames in each GOP, B_s is the capacity of the buffer and $B_C(i\text{-}1,N_{gop})$ is the buffer fullness after coding the (*i*-1)-th GOP. In the case of constant bandwidth, $T_r(i,j)$ is updated frame by frame as

$$T_r(i,j)=T_r(i,j-1)-A(i,j-1) \quad (4)$$

The hierarchical B pictures structure is taken into consideration in the proposed algorithm, the QP of the key frame is computed by

$$QP_{key}=\frac{S_{BQP}}{N_B}-1-\frac{8T_r(i-1,N_{gop})}{T_r(i,0)}-\frac{N_{gop}}{15} \quad (5)$$

where N_B is the total number of B-frames in a GOP and S_{BQP} is the sum of QPs for all B-frames in the previous GOP.

4.3 Frame Level Rate Allocation

By considering the buffer constraints, the first candidate target bits for the *j*-th frame in the *i*-th GOP are calculated as

$$\tilde{f}(i,j)=\frac{B}{F_r}+\gamma*(TB(i,j)-CB(i,j)) \quad (6)$$

where $CB\ (i,j)$ denotes the current buffer fullness, $TB\ (i,j)$ denotes the target buffer level, B is the constant bandwidth, and γ is a constant and its value is 0.75.

Meanwhile, considering the remained bits, the second candidate target bits for the *j*-th frame in the *i*-th GOP are calculated as

$$C(i,j)=\frac{1}{2}(\frac{R_{r1}}{PSNR_{r1}}+\frac{R_{r2}}{PSNR_{r2}}) \quad (7\text{-a})$$

$$\hat{f}(i,j)=\frac{C(i,j)}{C(i,j)+C_{ave}\times N_{RB}(j-1)}\times T_r(i,j) \quad (7\text{-b})$$

where R_{r1} and R_{r2} are the bitrates of forward reference frame and backward reference frame, respectively, $PSNR_{r1}$ and $PSNR_{r2}$ are the PSNR of forward reference frame and backward reference frame, respectively, $C(i,j)$ denotes the predicted complexity of *j*-th frame in the *i*-th GOP, C_{ave} denotes the average complexity of the coded B-frames, and $N_{RB}(j\text{-}1)$ denotes the number of the remaining B-frames.

The target bitrates for the *j*-th frame in the *i*-th GOP are finally expressed as a weighted combination of $\tilde{f}(i,j)$ and $\hat{f}(i,j)$

$$f(i,j)=\beta*\tilde{f}(i,j)+(1-\beta)*\hat{f}(i,j) \quad (8)$$

where β is a constant and its value is 0.5.

4.4 Macro-block Level Rate Allocation

In the MVC, RDO technique is adopted to choose the best encoding mode for the macro-block (MB), RDO goes through all the modes for inter-frame encoding and selects the mode of the least rate distortion cost to encode each MB. The rate distortion cost is calculated as

$$\begin{aligned}J(s,c,MODE\,|\,\lambda_{MODE})&=SSD(s,c,MODE)\\&+\lambda_{MODE}\cdot R(s,c,MODE)\end{aligned} \quad (9)$$

where s is the original data and c is its reconstruction data, *MODE* denotes the prediction mode, $SSD(s,c,MODE)$ is the sum of absolute differences, and $R(s,c,MODE)$ denotes the number of bits with the *MODE* to encode the MB. λ_{MODE} is the Lagrange multiplier which is calculated by

$$\lambda_{MODE}=0.85\times 2^{-\frac{8}{3}}\times Q_{step}^2 \quad (10)$$

where Q_{step} is the quantization step-size which is calculated by the R-Q quadratic model [14], as the *MAD* get larger, the Q_{step} and the λ_{MODE} get smaller, then the rate distortion model will choose the complex mode and allocate more bits for the MB. But it is not worth allocating too many bits on a MB which is not sensitive to human visual. Consequently, the λ_{MODE} could be adjusted as

$$\lambda_{MODE}=0.85\times 2^{-\frac{8}{3}}\times\alpha_P\times(1+\beta_P)\times Q_{step}^2 \quad (11)$$

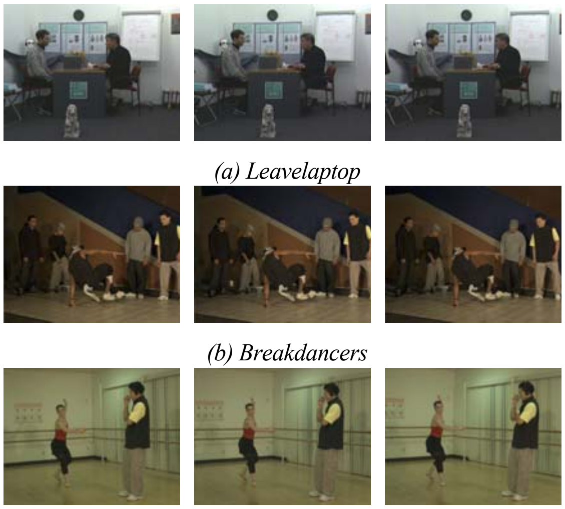
where β_P denotes the probabilities of *P* region, it is used to avoid that the *P* is too large to lead to the bit allocation sharp inequalities between two types of region, and α_P is a parameter and it is decided by the visual importance *P* described in Table 1.

$$\alpha_P=\begin{cases}0.6 & P=3\\0.7 & P=2\\3.0 & P=1\end{cases} \quad (12)$$

5. EXPERIMENTAL RESULTS

In order to demonstrate the advantage of the proposed rate control algorithm, it was taken to be compared with JMVC-RC which was JMVC7.0

added by JVT-G012 [14]. Several experiments were performed with multi-view video sequences of "Leavelaptop", "Breakdancers" and "Ballet" with the size of 1024×768.



(a) Leavelaptop

(b) Breakdancers

(c) Ballet

Figure 2. Three views of multi-view video test sequences

In the experiments, the three views (view5, view 6, view 7) for "Leavelaptop", the three views (view0, view 1, view 2) for "Breakdancers", the three views (view4, view 5, view 6) for "Ballet", Figure 2 shows the three test sequences. The total number of frames in each view was 81.

Table 2 indicates that the absolute inaccuracy of the JMVC-RC method and the proposed method are within 1.03% and 1.01%, respectively, they both can provide certain degree of rate control accuracy. In the region of *P*=3, the PSNR of the proposed method can raise by an average of 0.32 dB because there are more meticulous modes and more bits set for these MBs.

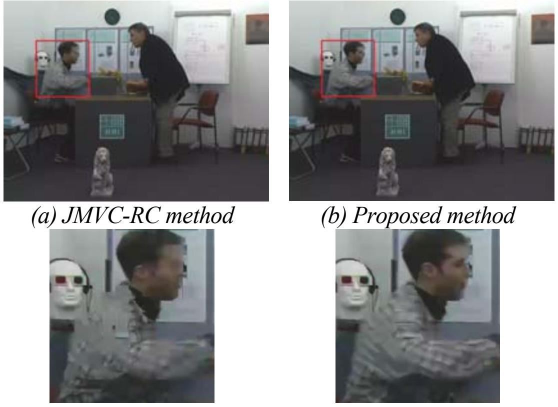
The subjective visual of the reconstructed images are shown in Figure 3-5, where the local images are the regions of both foreground and motion. In the Figure 3, it is clear that the clothes and the face of the sitting man are more distinct in the proposed method. Figure 4 shows that the edge of the dancing man is more clear in the proposed method. Figure 5 shows that the proposed method can achieve better texture of the ballet dancer compared with the JMVC-RC. Therefore, the proposed algorithm can provide more comfortable visual image quality for audiences.

Table 2. Simulation results of the proposed method and the JMVC-RC method

Sequence	Target bitrates (kbps)	Actual bitrates (kbps)		Inaccuracy (%)		PSNR(dB) Region *P*=3		
		JMVC-RC	Proposed	JMVC-RC	Proposed	JMVC-RC	Proposed	Gain
Leavelaptop	3420.72	3426.32	3424.26	0.16	0.10	38.74	38.94	0.20
	1412.12	1415.96	1415.02	0.27	0.21	35.71	36.02	0.31
	712.59	713.80	713.83	0.17	0.17	32.64	32.87	0.23
	401.45	401.53	401.11	0.02	0.09	29.47	29.84	0.37
Breakdancers	6020.66	6046.94	6052.76	0.44	0.53	38.48	38.66	0.18
	2167.81	2185.73	2188.01	0.83	0.93	36.74	36.96	0.22
	1033.32	1043.94	1043.80	1.03	1.01	34.65	34.94	0.29
	591.02	595.13	595.02	0.70	0.68	32.14	32.58	0.44
Ballet	2324.61	2332.66	2334.45	0.35	0.42	38.81	39.10	0.29
	1036.23	1040.36	1040.41	0.40	0.40	36.92	37.26	0.34
	574.52	576.35	576.29	0.32	0.31	34.97	35.46	0.49
	348.73	350.71	350.11	0.57	0.39	32.75	33.29	0.54

6. CONCLUSIONS

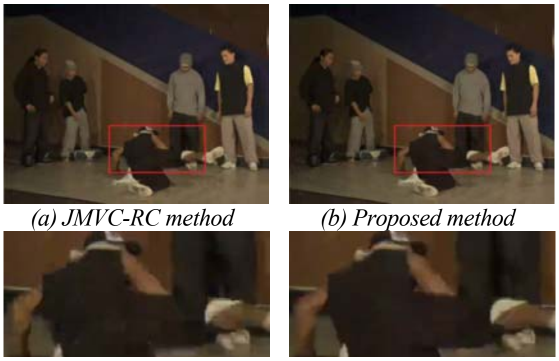
This paper presented a rate control algorithm for multi-view video coding with four levels. In the view level, the bitrates proportions among the views are decided by pre-encoding a group-of-pictures (GOP). In the GOP level, the initial quantization parameter and the target bits are calculated for the GOP. In the frame level, the characteristics of the two reference frames are used to predict the complexity of current frame. In the macro-block level, the foreground and the motion region are allocated more bits and can achieve good visual quality. Experimental results show that the proposed rate control method can accurately control the bitrates and provide a good visual effect for multi-view video. In future probe, more efforts will be focused on consideration of bit allocation among the views to improve the accuracy of the rate controlling.



(a) JMVC-RC method (b) Proposed method

(c) Local Area in Figure 3(a) (d) Local Area in Figure 3(b)

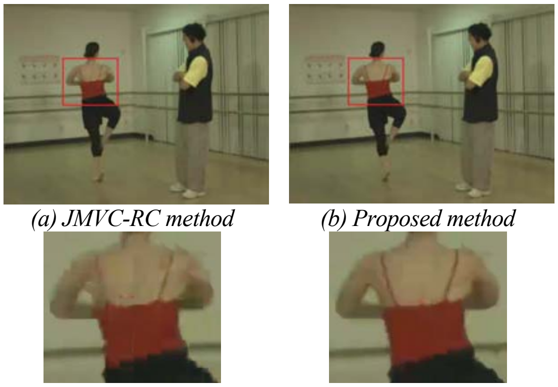
Figure 3. Subjective visual comparison of the two methods for Leavelaptop



(a) JMVC-RC method (b) Proposed method

(c) Local Area in Figure 4(a) (d) Local Area in Figure 4(b)

Figure 4. Subjective visual comparison of the two methods for Breakdancers



(a) JMVC-RC method (b) Proposed method

(c) Local Area in Figure 5(a) (d) Local Area in Figure 5(b)

Figure 5. Subjective visual comparison of the two methods for Ballet

ACKNOWLEDGMENTS

This work is supported by Natural Science Foundation of China (61171163, 61071120, 61271270, 61271021, 61111140392), Natural Science Foundation of Zhejiang Province (Y1101240), and The projects of Ningbo Government (2011A610200, 2011A610197, 2011B81002).

REFRENCES:

[1] P. Merkle, K. Muller and T. Wiegand, "3D Video: Acquisition, Coding, and Display", *IEEE International Conference on Consumer Electronics*, Vol.56, No.2, January 2010, pp.946-950.

[2] F. Shao, G. Jiang, M. Yu, K. Chen and Y. Ho, "Asymmetric Coding of Multi-view Video Plus Depth Based 3-D Video for View Rendering", *IEEE Transactions on Multimedia*, Vol.14, No.1, February 2012, pp.157-167.

[3] Y. Chang and B. Chung, "An Adaptive Search Range Algorithm for Multi-view Motion and Disparity Estimation", *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, July 2-5, 2012, pp. 550-554.

[4] M. Park and G. Hoon, "Realistic Multi-view Scalable Video Coding Scheme", *IEEE Transactions on Consumer Electronics*, Vol.58, No.2, May 2012, pp.535-543.

[5] M. Solh and G. Alregib, "Hierarchical Hole-Filling for Depth-Based View Synthesis in FTV and 3D video", *IEEE Journal of Selected Topics in Signal Processing*, Vol.6, No.5, September 2012, pp.495-504.

[6] B. Kamolrat, W. Fernando and M. Mrak, "Rate Controlling for Color and Depth 3D Video Coding", *Proceedings of the SPIE: Application of Digital Image Processing XXXI*, San Diego, CA, USA, August 2008.

[7] Y. Liu, Q. Huang, S. Ma, D. Zhao and W. Gao, "A Novel Rate Control Technique for Multi-view Video Plus Depth Based 3D Video Coding", *IEEE Transactions on Broadcasting*, Vol.57, No.2, June 2011, pp.562-571.

[8] T. Yan, P. An, L. Shen, Q. Zhang and Z. Zhang, "Rate Control Algorithm for Multi-view Video Coding Based on Correlation Analysis", *2009 Symposium on Photonics and Optoelectronics (SOPO)*, August 14-16, 2009, pp.1-4.

[9] J. Zhang, X. Yi, N. Ling and W. Shang, "Context Adaptive Lagrange Multiplier (CALM) for Rate-distortion Optimal Motion Estimation in Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.20, No.6, June 2010, pp.820-828.

[10] T. Ou, Y. Huang and H. Chen, "SSIM-Based Perceptual Rate Control for Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.21, No.5, May 2011, pp.682-691.

[11] Q. Lin and G. Feng, "The Bit Allocation and RDO Mode Based Rate Control Algorithm", *2010 International Conference on Anti-Counterfeiting Security and Identification in Communication (ASID)*, Quanzhou, China, July 18-20,2010, pp.154-157.

[12] S. Park and M. Kim, "Extracting Moving/Static Objects of Interest In Video", *Lecture Notes in Computer Science*, Vol.4261, November 2006, pp.722-729.

[13] Y. Zhang, G. Jiang and M. Yu, "Stereoscopic Visual Attention-Based Regional Bit Allocation Optimization for Multi-view Video Coding", *EURASIP Journal on Advances in Signal Processing Archive*, Vol. 2010, No. 60, February 2010.

[14] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, "Adaptive Basic Unit Layer Rate Control for JVT", *Doc. JVT-G012r1*, Pattaya, Thailand, March 7-15, 2003, pp.23-27.