International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:7, 2013

A New H.264-Based Rate Control Algorithm for Stereoscopic Video Coding

Yi Liao, Wencheng Yang, and Gangyi Jiang

Abstract**—**According to investigating impact of complexity of stereoscopic frame pairs on stereoscopic video coding and transmission, a new rate control algorithm is presented. The proposed rate control algorithm is performed on three levels: stereoscopic group of pictures (SGOP) level, stereoscopic frame (SFrame) level and frame level. A temporal-spatial frame complexity model is firstly established, in the bits allocation stage, the frame complexity, position significance and reference property between the left and right frames are taken into account. Meanwhile, the target buffer is set according to the frame complexity. Experimental results show that the proposed method can efficiently control the bitrates, and it outperforms the fixed quantization parameter method from the rate distortion perspective, and average PSNR gain between rate-distortion curves (BDPSNR) is 0.21dB.

Keywords**—**Stereoscopic video coding, rate control, stereoscopic group of pictures, complexity of stereoscopic frame pairs.

I.INTRODUCTION

WITH the rapid development of video technology and human' increasingly demands for visual entertainment, stereoscopic video is stepping into our lives gradually. Rate control (RC) plays an essential role in stereoscopic video coding which has been widely researched in recent years [1], [2]. Researchers have proposed several rate control algorithms for conventional video coding standards, such as MPEG-2 TM5, H.263 TMN8, MPEG-4 VM8 and H.264 G012. However, these algorithms were designed for 2D video coding system which cannot effectively and directly applied to stereoscopic video coding.

Recently a number of projects have begun work on RC in stereoscopic video research areas, Zhu et al. [3] computed the quantization parameter (QP) for I frame based on frame rate, bitrates and image type and then improved the quadric rate-distortion model based on the cyclopean perception. Lu et al. [4] defined the main view and the assistant view; the frames in the two views were classified into six types and given different weights. Shao et al. [5], [6] confirmed a certain poor quality is allowed between left and right views at the same stereo video subjective quality based on subjective test and proposed a distortion-quantization relationship between left and right views to take full advantage of the binocular psycho-visual redundancy. Besides, Liu et al. [7] establish the virtual view quality model (VVQM) to assess the right virtual view qualities under different rate combinations of video and depth, which was fit for mobile 3D video coding. Wang et al. [8] used the weighted average video quality metric (VQM) of the left and right views as the stereoscopic video distortion metric and built a cubic polynomial RD model. A hot research topic in recent years is the multi-view rate control, many scholars and institutions involved in research in this area [9]-[11]. However, these researches did not consider the complexity of left and right frames which may be leads to the bits allocation irrationally.

In this paper, we predict the frame complexity on temporal and spatial in order to allocate bits more rationally in stereoscopic frame (SFrame) level. Position significance is taken into account in frame level to allocate bits between the left frame and the right frame. The proposed method can accurately control the bitrates and provide a better rate distortion performance and subjective quality.

The rest of the paper is organized as follows. In Section II, the frame complexity on temporal and spatial is described in detail. A rate control algorithm for stereoscopic video is proposed in Section III. Then, the experimental results are analyzed in Section IV. Finally, the conclusions are given in Section V.

II. FRAME COMPLEXITY

Frame complexity is one of important factors in video coding, how to measure and take advantage of frame complexity becomes extremely important in rate control. JVT-G012 rate control algorithm uses MAD (Mean Absolute Difference) to measure the complexity of the frame, which is defined as

$$MAD = \frac{1}{M \times N} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} \left| I_{x,y} - I'_{x,y} \right| \tag{1}$$

where M and N denote the height and width of the frame, respectively, $I_{x,y}$ denotes the luminance value at pixel (x, y) in a frame, $I'_{x,y}$ denotes the predicted value of $I_{x,y}$, and $|I_{x,y}-I'_{x,y}|$ denotes the predicted residuals.

Let $R_{MAD}(j)$ be the relative complexity between the j-th SFrame and the encoded SFrames before the j-th SFrame in i-th SGOP (Stereoscopic Group of Pictures), which is calculated as

$$R_{MAD}(i,j) = \frac{P_{MAD}(i,j)}{A_{MAD}(i,j)} \tag{2}$$

Yi Liao, Mei Yu, Wencheng Yang, and Gangyi Jiang are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. Gangy Jiang is also as a visiting researcher with National Key Lab of Software New Technology, Nanjing University, Nanjing, China. (e-mail: liaoyi518@ 163.com, yumei2@ 126.com, jianggangyi@ 126.com).

where $P_{MAD}(i,j)$ denotes the predicted *MAD* of the *j*-th SFrame in the *i*-th SGOP, $A_{MAD}(i,j)$ denotes the average *MAD* of the encoded SFrames before the *j*-th SFrame, which is calculated as

$$A_{MAD}(i,j) = \frac{1}{j-1}\sum_{k=1}^{j-1}(MAD_L(i,k) + MAD_R(i,k)) \quad (3)$$

where $MAD_L(i, k)$ and $MAD_R(i, k)$ denote the *MAD* of the left and right frames in the *k*-th SFrame, respectively.

As we know, MAD reflects the temporal relation between the current frame and the previous frame after motion compensation, but it does not consider the spatial relativity in a frame. In order to express the complexity in an accurate way, we first compute the frame gradient as

$$Grad = \frac{1}{M \times N}\sum_{y=0}^{N-1}\sum_{x=0}^{M-1}\left|I_{x,y} - I_{x+1,y}\right| + \left|I_{x,y} - I_{x,y+1}\right| \quad (4)$$

Let $R_{Grad}(i, j)$ be the relative gradient between the current SFrame and the encoded SFrames, it is calculated as

$$R_{Grad}(i,j) = \frac{P_{Grad}(i,j)}{A_{Grad}(i,j)} \quad (5)$$

where $P_{Grad}(i,j)$ denotes the predicted *Grad* of the *j*-th SFrame in the *i*-th SGOP, $A_{Grad}(i,j)$ denotes the average *Grad* of the encoded SFrames ,which is calculated as

$$A_{Grad}(i,j) = \frac{1}{j-1}\sum_{k=1}^{j-1}(Grad_L(i,k) + Grad_R(i,k)) \quad (6)$$

where $Grad_L(i, k)$ and $Grad_R(i, k)$ denote the *Grad* of the left and right frames in the *k*-th SFrame, respectively.

Let $COMP(i, j)$ be the temporal-spatial complexity of the *j*-th SFrame in the *i*-th SGOP, which can be expressed as a weighted combination of $R_{MAD}(i,j)$ and $R_{Grad}(i,j)$

$$COMP\ (i,j) = \alpha \times R_{MAD}\ (i,j) + (1-\alpha) \times R_{Grad}\ (i,j) \quad (7)$$

where α is a constant and its value is 0.5.

III.Rate Control Algorithm for Stereoscopic Video Based On H.264

The algorithm is proposed based on the H.264 compression standard which is compatible with 2D video communication. Meanwhile, stereoscopic video is composed by two views, left view is set as main view and the right view is set as assistant view, redundancy is exist between the two views. Motion estimation is adopted in left view while motion estimation and disparity estimation are combined in right view.

A.Prediction Structure for Stereoscopic Video Coding

In all the prediction structures for stereoscopic video coding, disparity estimation and motion estimation are combined makes a good efficiency [12], as shown in Fig. 1, in which the L and R denote the left view and right view, respectively. All the frames between the two I frames comprise a stereoscopic group of pictures (SGOP), the two frames in one time comprise a SFame. I-frames are completely intra-coded, the P-frames and B-frames in left view are coded by motion estimation, and the P-frames and B-frames in right view are coded by motion estimation and disparity estimation.

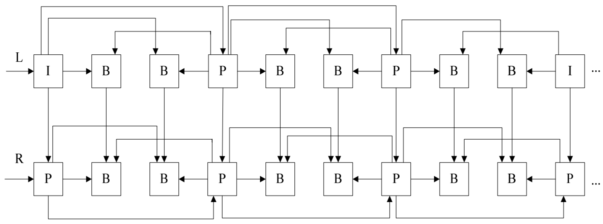


Fig. 1 Prediction structure for stereoscopic video coding

B.Algorithm Description

Based on the above discussion, we propose a new rate control algorithm. It is performed on three levels: 1) SGOP level rate allocation; 2) SFrame level rate allocation; 3) frame level rate allocation. In the proposed method, the target bitrates for frame are allocated by the frame complexity, position significance and reference property, and the target buffer is set by considering frame complexity.

1.SGOP Level Rate Allocation

In the SGOP level, we calculate the target bits and the initial QP for each SGOP. The initial bits for the *i*-th SGOP is calculated as

$$T(i,0) = \frac{u(i,0)}{F_r} \times N - (\frac{B_s}{8} - B_C(i,0)) \quad (8)$$

where $u(i,0)$ denotes the initial bandwith when encoding the *i*-th SGOP, F_r denotes the frame rate and N denotes the number of the SFrames in the SGOP. B_s denotes the buffer size and $B_c(i,0)$ denotes the buffer occupancy after coding the previous SGOP.

Let $T(i, j)$ (j>1) be the remaining bits after coding the *j*-th SFrame in the *i*-th SGOP, and $T(i, j)$ is updated as

$$T(i,j) = T(i,j-1) + \frac{u(i,j) - u(i,j-1)}{F_r} \times (N-j) - A(i,j-1) \quad (9)$$

where $u(i,j)$ denotes the bandwith when encoding the *j*-th frame in the *i*-th SGOP, $A(i, j\text{-}1)$ denotes the number of bits generated by the (j-1)-th SFrame in the *i*-th SGOP.

The QP for the first frame is calculated by

$$QP_{st} = \frac{S_{PQP}}{N_P} - \frac{8T(i-1,N)}{T(i,0)} - \frac{N}{15} - 1 \quad (10)$$

where S_{PQP} and N_p denote the sum of QP of the P- frames and the total number of P-frames in previous SGOP, respectively.

2.SFrame Level Rate Allocation

The task in this stage is to establish the target bits allocated to a SFrame. The target bits of a SFrame are determined based on the buffer level and remained bits in the SGOP.

By considering the remained bits, the first candidate target bits for the *j*-th SFrame in the *i*-th SGOP are calculated by

$$\hat{T}(i,j) = \frac{T(i,j)}{N-j} \times COMP\ (i,j) \tag{11}$$

where the $COMP(i,j)$ is calculated as in Section II, the larger $COMP(i,j)$ is, the more bits should be allocated to the *j*-th SFrame in the *i*-th SGOP.

Meanwhile, the second candidate target bits for the *j*-th SFrame in the *i*-th SGOP are calculated by considering the buffer constraints. Let $TB\ (i, j)$ be the target buffer level, it is calculated as

$$TB(i,j) = TB(i,j-1) - \frac{TB(i,2)}{N-j-1} \times COMP(i,j) \tag{12}$$

the larger $COMP(i,j)$ is, the larger buffer space should be leave for the *j*-th SFrame in the *i*-th SGOP.

Although the bigger buffer size can bear the big swings of bitrates, too big buffer will cause much buffer delay. As a result, we should maintain suitable buffer control accuracy to prevent buffer from overflow and underflow. Meanwhile, the front frames are the reference for the back frames, it means that the front frames are important than the back ones, so it is necessary to allocate more bits for the front ones. The target bits for the subsequent SFrames are determined by

$$\tilde{T}(i,j) = W_P \times \frac{u(i,j)}{F_r} \times \frac{N-j}{N_P} + \gamma \times (TB(i,j) - CB(i,j)) \tag{13}$$

where W_P is an adjustable factors and its value is 2, $TB\ (i,j)$ denotes the target buffer level, and $CB\ (i,j)$ denotes the current buffer fullness.

The target bits for a SFrame are finally expressed as a weighted combination of $\hat{T}(i,j)$ and $\tilde{T}(i,j)$

$$T\ (i,j) = \beta * \hat{T}(i,j) + (1-\beta) * \tilde{T}\ (i,j) \tag{14}$$

where β is a constant and its value is 0.5.

3.Frame Level Rate Allocation

In this stage, the QPs for all frames are calculated. B-frames and P-frames are performed on different approach.

The B-frames are taken inter prediction and the QP could be set larger than the adjacent I-frames or P-frames as to save more bits for I-frames and P-frames. Meanwhile, in order to keep the fluctuation of video quality more stable, the QP difference between the adjacent frames should be set as smaller than 2.

$$QP_B = \frac{QP_1 + QP_2 + 2}{2} \tag{15}$$

where QP_1 and QP_2 are the P-frames in front and back of the B-frame.

For the P-frames, it can be seen from the reference relationship of the coding structure, the left view is the main view which places a greater impact on video quality, so it is given a right weight η to adjust the rate allocation value. The target bits for the left frame and right frame are calculated as

$$T_{left}(i,j) = \eta \times T_{LR}(i,j) \times \frac{Grad_L(i,j)}{Grad_L(i,j) + Grad_R(i,j)} \tag{16}$$

$$T_{right}(i,j) = T_{LR}(i,j) - T_{right}(i,j) \tag{17}$$

where η is a constant and its value is 1.2.

Finally, QPs for left and right frames are computed based on the quadratic R-Q model [13] according to the corresponding target bits, respectively.

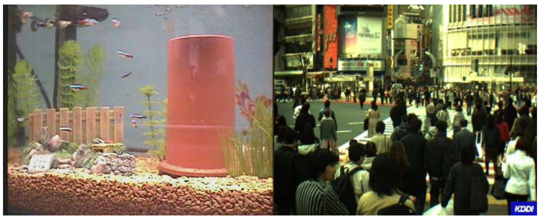
IV. Experimental Results and Analyses

To evaluate the performance of the proposed rate control algorithm, six stereoscopic video sequences with different spatial resolutions, i.e., 'Aquarium' and 'Crowd' with the size of 320×240, 'Akko', 'Rena' and 'Ballroom' with size of 640×480, 'Soccer' with size of 720×480, were used in the experiments. The six test sequences are shown in Fig. 2, we compare the RCE (Rate Control Error), RD (Rate Distortion) performance and subject quality.

RCE is calculated as

$$RCE = \left| \frac{R_{acture} - R_{target}}{R_{target}} \right| \tag{18}$$

where R_{actual} is the bitrates generated by the test sequence, and R_{target} is the target bitrates which is generated by coding the test sequences with fixed QP, and the fixed QP is set to 22, 27, 32 and 37, respectively. The test conditions are shown in Table I.



(a) Aquarium (b) Crowd

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:7, 2013

(c) Akko (d) Ballroom

(e) Rena (f) Soccer

Fig. 2 Test sequences

TABLE I
EXPERIMENT CONDITIONS

SGOP length	15
Frame rate	30
Frames to be encoded	240
MV search range	32
Reference Frames	2
Entropy coding method	CABAC
RD optimization	used
Hadamard transform	used

TABLE II
ACCURACY OF THE PROPOSED ALGORITHM

Sequence	Fixed QP	Target bitrates (kbps)	Actual Bitrates(kbps)	RCE (%)
Aquarium	22	4222.68	4224.28	0.04
	27	1110.57	1112.26	0.15
	32	423.32	425.65	0.55
	37	212.04	214.07	0.96
Crowd	22	5102.33	5093.88	0.17
	27	2783.58	2782.61	0.04
	32	1429.50	1430.88	0.10
	37	709.04	711.06	0.29
Akko	22	3663.85	3664.41	0.02
	27	1653.55	1655.51	0.12
	32	877.63	878.36	0.08
	37	494.92	495.90	0.20
Ballroom	22	7675.14	7675.24	0.01
	27	2967.13	2966.04	0.04
	32	1486.56	1487.52	0.07
	37	796.32	799.52	0.40
Rena	22	3351.03	3351.9	0.03
	27	1465.66	1466.48	0.06
	32	720.53	721.51	0.14
	37	372.46	373.16	0.19
Soccer	22	6698.66	6704.00	0.08
	27	2866.06	2866.05	0.00
	32	1301.12	1303.90	0.21
	37	637.68	638.94	0.20

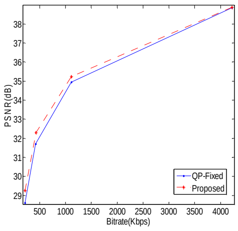
TABLE III
THE PSNR COMPARISONS OF THE TWO METHODS

Sequence	Fixed QP	PSNR(dB)			BDPSNR (dB)
		Fixed QP	Proposed	Gain	
Aquarium	22	38.85	38.87	0.02	0.32
	27	34.94	35.23	0.29	
	32	31.70	32.31	0.61	
	37	28.62	29.25	0.63	
Crowd	22	39.55	39.58	0.03	0.20
	27	35.34	35.42	0.08	
	32	31.26	31.55	0.29	
	37	27.65	28.08	0.43	
Akko	22	43.28	43.32	0.04	0.27
	27	40.38	40.59	0.21	
	32	37.21	37.61	0.40	
	37	34.18	34.66	0.48	
Ballroom	22	40.59	40.62	0.03	0.20
	27	37.57	37.72	0.15	
	32	34.75	35.08	0.33	
	37	32.03	32.41	0.38	
Rena	22	45.48	45.70	0.22	0.14
	27	42.62	42.71	0.09	
	32	39.50	39.67	0.17	
	37	36.55	36.76	0.21	
Soccer	22	41.24	41.26	0.02	0.14
	27	38.31	38.45	0.14	
	32	35.52	35.71	0.19	
	37	32.90	33.10	0.20	

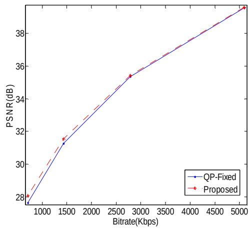
Table II shows that the actual bitrates in proposed method are close to the target bitrates, the average control error is 0.17%. Generally speaking, RCE control to achieve their goals is in less than 5%, it means that the proposed method can control the bitrates accurately. The PSNR comparisons of the proposed method and the fixed QP method are shown in Table III, the proposed method can rise by a maximum of 0.63dB which shows that the proposed method can improve the objective image quality.

The rate distortion (RD) performance comparison results between the proposed method and the Fixed QP are shown in Fig. 3. The curve of the proposed method fluctuates higher than that of the Fixed QP method, the BDPSNR [14] which reflect the RD performance are also shown in Table III, the proposed method can improve the BDPSNR by an average of 0.21dB. Consequently, the proposed method can achieve the better RD performance.

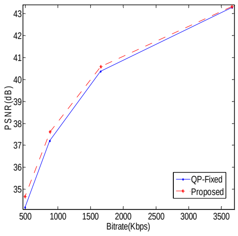
We also compare the subjective image quality between the two methods. Fig. 4 shows that the whole image is more clearly in the proposed method, it is because the scene changes obviously from the 47-th frame to the 48-th frame, the proposed method allocates more bits for the frame of higher complexity. In Fig. 5, distortion of advertising board in the proposed method is smaller than that in the Fixed QP method. That is because the proposed algorithm considered frame complexity in SFrame level bits allocation, bits allocated more reasonable and get better subjective quality of the reconstructed image.



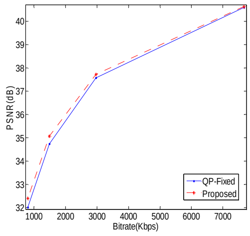
(a) Aquarium



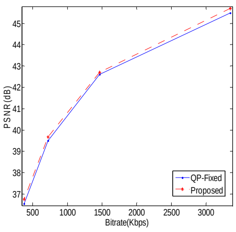
(b) Crowd



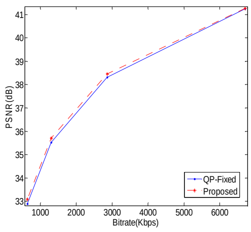
(c) Akko



(d) Ballroom



(e) Rena



(f) Soccer

Fig. 3 The RD performance comparisons between the two methods

(a) Fixed QP (PSNR 37.05dB)

(b) Proposed (PSNR 37.76dB)

(c) Fixed QP

(d) Proposed

Fig. 4 Subjective visual comparison of the two methods for Akko

(a) Fixed QP (PSNR 35.55dB)

(b) Proposed (PSNR 35.91dB)

(c) Fixed QP

(d) Proposed

Fig. 5 Subjective visual comparison of the two methods for soccer

V. Conclusions

Bit allocation and Rate control are the key technology in the stereoscopic video encoding and transmission. Based on the stereoscopic characteristics of the video, the proposed algorithm allocated bit according to the temporal-spatial complexity, position significance and reference property between the left and right frames. The experiment results show that the proposed method can control the bitrates accurately and have the better performance and subjective quality compared with the Fixed QP method. In future probe, more efforts will be focused on consideration of the binocular redundancy to improve the encoding efficiency.

Acknowledgment

This work was supported by the Natural Science Foundation of China (Grant Nos. 61071120, 61111140392), the Zhejiang Scientific and Technical Key Innovation Team Project (Grant No. 2010R50009), and Projects of Ningbo government (Grant No. 2011B81002).

References

[1] P. Surman, “Stereoscopic and Autostereoscopic Displays,” *3D-TV system with depth-image-based rendering, Springer New York*, pp.375-411, 2013.
[2] A. Aggoun, E. Tsekleves, D. Zarpalas, A.Dimou, P. Daras, P. Nunes, and L. Soares, “Immersive 3D holoscopic video system,” *IEEE Trans. MultiMedia*, vol.20, no.1, pp.28-37, 2013.
[3] Z. Zhu, F. Liang, G. Jiang, and Q. Zhang, “Bit allocation and rate control for stereoscopic video coding,” *in Proc. of 3DTV Conf.*, Kos, Island, 2007, pp.1-4.
[4] S. Lu, J. Sun, and J. Wang, “Multiview video rate control scheme based on human stereo visual feature,” *Journal of Image and Graphics*, vol.14, no.11, pp. 2396-2400, 2009.
[5] F. Shao, G. Jiang, X. Wang, M. Yu, and K. Chen, “Stereoscopic video coding with asymmetric luminance and chrominance qualities,” *IEEE Trans. Consumer Electronics*, vol.54, no.6, pp.2460-2468, 2010.
[6] F. Shao, G. Jiang, M. Yu, Q. Zheng, and K. Chen, “A novel rate control technique for asymmetric-quality stereoscopic video,” *IEEE Trans. Consumer Electronics*, vol.57, no.4, pp.1823-1829, 2011.
[7] Y. Liu, G. Peng, Y. Hu, S. Ci, and H. Tang, “A multi-pass VBR rate control method for video plus depth based mobile 3D video coding,”

Advances in Multimedia Information Proc.-PCM 2010, *Springer Berlin Heidelberg*, pp.156-166, 2011.

[8] Q. Wang, L. Zhuo, J. Zhang, and X. Li, "Frame layer rate control method for stereoscopic video coding based on a novel rate-distortion model," *The Era of Interactive Media*, *Springer New York*, pp. 205-216, 2013.

[9] S. Park and D. Sim, "An efficient rate control algorithm for multi-view video coding," *The 13th IEEE International Sym. Consumer Electronics (ISCE)*, kyoto, Janpan, 2009, pp.115-118.

[10] 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 Trans. on Broadcasting*, vol. no. 2, pp. 562-571, 2011.

[11] Q. Zheng, M. Yu, G. Jiang, F. Shao, and Z. Peng, "Rate control for multi-View video coding based on statistical analysis and frame complexity estimation," *International Conf. Computer, Informatics, Cybernetics and Applications*, Hangzhou, China, 2011, pp.1089-1097.

[12] S. Li, M .Yu, G. Jiang, T. Choi, Y. Kim, and Y. Deak, "Approaches to H.264-Based stereoscopic video coding," *2004 IEEE First Sym. Multi-Agent Security and Survivability*, Hongkong, 2004, pp.365-368.

[13] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6, "Adaptive basic unit layer rate control for JVT," *Doc. JVT-G012r1*, Pattaya, Thailand, 2003, pp.23-27.

[14] G. Bjontegard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33, the 13th VCEG Meeting*, Austin, Texas, USA, 2001.