

Adaptive Frame Level Rate Control in 3D-HEVC

Songchao Tan¹, Junjun Si², Siwei Ma², Shanshe Wang³, Wen Gao²

¹ School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China

² Institute of Digital Media, Peking University, Beijing 100871, China

³ School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150080, China

Abstract—In this paper, we propose a new frame level rate control algorithm for the high efficiency video coding (HEVC) based 3D video (3DV) compression standard. In the proposed scheme, a new initial quantization parameter (QP) decision scheme is provided, and the bit allocation for each view is investigated to smooth the bitrate fluctuation and reach accurate rate control. Meanwhile, a simplified complexity estimation method for the extended view is introduced to reduce the computational complexity while improves the coding performance. The experimental results on 3DV test sequences demonstrate that the proposed algorithm can achieve better R-D performance and more accurate rate control compared to the benchmark algorithms in HTM10.0. The maximum performance improvement can be up to 12.4% and the average BD-rate gain for each view is 5.2%, 6.5% and 6.6% respectively.

Index Terms— rate control, initial QP, bits allocation, 3D video coding

I. INTRODUCTION

With today's rapid development of 3D acquisition and display technologies such as auto-stereoscopic 3D TV [1] and free-viewpoint video [2], 3D video has gained increasing interests recently. To better support various stereoscopic or 3D video applications, Motion Picture Experts Group (MPEG) has started the 3D video (3DV) standardization efforts, which are currently continued in the Joint Collaborative Team (JCT) on 3D Video coding (JCT-3V). Both H.264/AVC and HEVC [3] are utilized to develop 3DV standards by JCT-3V. And as one of them, 3D-HEVC which targets at developing tools for both texture views and depth views is based on HEVC.

The typical 3DV is stereo-view video which provides each eye with one video separately at the same time. The disparity between these two videos causes the illusion of depth perception for human. And in 3D-HEVC, due to the need of a more comprehensive view, it involves a more general case of n -view multi-view video. By using the depth image-based rendering (DBIR) technique, we can get large view-angle scene with fewer amounts of data [4].

Rate Control (RC) is employed in video coding to regulate the bitrate meanwhile guarantee good video quality. As for 3D-HEVC, it becomes more complicated because multiple views are involved in coding and within each view there are two kinds of video sequences (i.e. texture and depth map). Two RC algorithms Unified Rate Quantization (URQ) [5] and R-lambda [6] are proposed for 3D-HEVC up to now. In [7], an inter-view mean absolute difference (MAD) prediction based on the depth map is introduced to improve the accuracy

of the MAD prediction. However due to the bit allocation for each view is not suitable for 3D-HEVC, the performance of PSNR, the dramatic fluctuation of generated bits and bits error are unsatisfactory. Aiming at solving the problems of the existing algorithms, a more suitable bit allocation scheme which gives full consideration to the particularity of the new video coding standard is used in the proposed RC algorithm. As an important part of RC algorithm, a new initial quantization parameter (QP) decision scheme is proposed for 3D-HEVC. Through the introduction of the video's characteristic, our initial QP decision scheme can provide a more smooth quality and less bitrate fluctuation. At the same time, in order to achieve a higher quality and higher coding efficiency, we improve the R-Q model by introducing inter-view complexity estimation.

The rest of this paper is organized as follows. In Section II, the algorithm is briefly described. In II-A, the decision of initial QP is proposed. The bit allocation for base and extended view is investigated in II-B. And in II-C, inter-view complexity estimation is designed for the proposed R-D model. In Section III, the experimental results are given to demonstrate the efficiency of the proposed RC algorithm. Finally, Section IV concludes this paper.

II. PROPOSED RATE CONTROL ALGORITHM FOR HEVC-3DV

A. The decision of Initial QP

The initial QP decision is important for video coding. Bit per pixel (bpp) based initial QP decision is adopted as illustrated in (1).

$$bpp = \frac{bitrate}{w \times h \times f} \quad (1)$$

where bpp refer to the bit per pixel, w and h are the width and height of the sequence. f is the frame rate of the video.

However, the bpp-based initial QP strategy widely used in RC algorithm does not work well in 3D-HEVC. Because bpp can just reflect the bitrate, but the characteristic of video is ignored. For instance, for the sequence "Poznan_Hall2" and "Undo_Dancer", as shown in Table I, the bpp of sequence "Poznan_Hall2" when the anchor's initial QP is 25 is same as the bpp of sequence "Undo_Dancer" when the anchor's initial QP is 35. Obviously, by utilizing bpp-based initial QP scheme, the initial QP will be too large for sequence "Poznan_Hall2".

In this paper, by a lot of experimental analysis we propose an improved scheme for the decision of initial QP based on bpp per variance of gradient (BPVG) as follows.

$$\begin{aligned} BPVG &= Bpp \times 10^5 / VG \\ VG &= D(\text{gradient}) \end{aligned} \quad (2)$$

where VG refer to the variance of each pixels' gradient in the first frame.

As shown in Table I, due to the characteristic of video is considered reasonably, the proposed initial QP can solve the above mentioned problem well.

TABLE I INITIAL QP FOR DIFFERENT SEQUENCE

Class	Sequence	Anchor QP	Kbps	bpp	bpp based Initial QP	Proposed Initial QP
1920*1088	Poznan_Hall2	25	810.69	0.0155	35	25
		30	344.89	0.0066	40	30
		35	177.24	0.0034	47	35
		40	98.51	0.0019	49	40
	Undo_Dancer	25	4626.44	0.0923	25	25
		30	2005.90	0.0384	30	30
		35	916.22	0.0175	35	35
		40	426.33	0.0082	40	40

B. Bit allocation for 3D-HEVC

In this section, we first briefly review the bit allocation in video coding and then detail our proposed method. Bit allocation plays a key role in video coding. In transform coding, bit allocation is performed by allocating the bits budget to different groups of transformed coefficients to minimize the overall quantization distortion. In [8], Huang first presents the optimal bit allocation problem. Optimal bit allocation is incorporated with rate control scheme in video coding to improve the coding quality.

In [9], the bit allocation is designed to allocate the bits weighted averagely for each frame and for I-SLICE the target bits are multiplied by five because the quality of I-SLICE has a great impact on the next few frames. For the Low Delay (LD) configuration, the scheme works well, but in the Random Access (RA) configuration, the performance and the bitrate error are unsatisfactory. Through the analysis of the bit allocation, we find the reason is that the scheme is based on rate gop [10] (in RA, a rate gop is composed of eight frames) as,

$$T_{gop} = (R/F) \cdot N \quad (3)$$

where R is the target bitrate, F is the frame rate of the sequence and N is the number of frames in a rate gop. And when a frame is I-SLICE, the target bits for it are multiplied by five. But the extra bits cost by I-SLICE are just managed by overflow/underflow handling strategy instead of considered carefully. That explains why the scheme can work well just in LD which has only one I-SLICE.

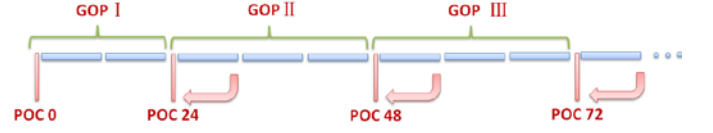


Fig. 1 Proposed Bit Allocation Scheme

To solve the problem, we carried out a series of related research and find that due to the particularity of I-SLICE which means the PSNR of a frame have no relation to the frame before this I-SLICE and in a GOP there is only one I-SLICE. Based on that, we proposed the bit allocation which based on GOP.

As shown in Fig 1, the target bits for a frame is calculated as follow (4),

$$target_{n,i} = \begin{cases} bpf \times (GOP - gop + 1) \times w_i, & n = 1 \\ bpf \times GOP \times w_i, & n \in [2, m-1] \\ bpf \times (frame_last) \times w_i, & n = m \end{cases} \quad (4)$$

$$bpf = \text{bitrate} / \text{framerate} \quad (5)$$

where $target_{n,i}$ is the target bits for n th GOP, i th frame. w_i is the weight of different frame. $frame_last$ is the number of the last GOP.

When overflow or underflow occurs, the difference between target bits and actual bits in a GOP will be distributed equally to the rest waiting for coding GOP. At the same time, the video buffer verifier (V BV) operation model is established to smooth quality fluctuation.

C. Proposed R-Q Model

The trade-off between the output bitrate (R) and the quality (D) of compressed video is determined by quantization step size (Q_i), which is indexed by quantization parameter (Q). The R-Q model has been studied extensively for previous video coding standards such as H.264/AVC and H.265/HEVC. Based on experimental analysis, we propose the R-Q model as,

$$R = \alpha \times X / QP \quad (6)$$

where α is the model parameter. R is the coding rate. X is the complexity estimation for the current picture. QP is the quantization parameter. X is computed as:

$$X = \left(\sum_{i=0}^n (w_i \times SAD_i) / \sum_{i=0}^{n-1} (w_i \times SAD_i) \right)^{1-\lambda} \times R_{n-1} \times QP_{n-1} \quad (7)$$

n is the current frame number. QP_{n-1} is the quantization parameter of the $(n-1)$ th frame. R_{n-1} is the actual bits of the $(n-1)$ th frame. λ is a constant, the recommended value is 0.6. w_i is the weight of SAD values of previously encoded frames.

Due to the relationship between extended view and base view as shown in Fig 2, we can see that there is strong

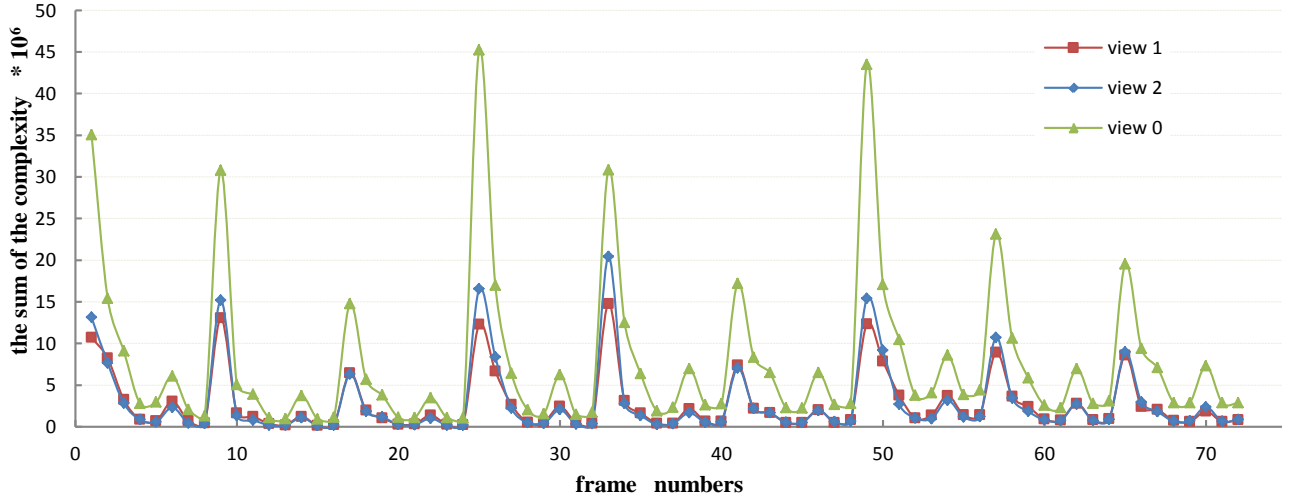


Fig. 2 The complexity (SAD) between view 0, view 1 and view 2

correlation between adjacent viewpoints. Through a lot of experimental analysis, the complexity of extended view's frame can accurate replaced by base view's frame which just finished coding in the same GOP after appropriate scaling. It should be noted that I-SLICE should consider independently, because extended view's frame has no I-SLICE. Based on statistical analysis, we proposed a linear model to estimate the complexity of the dependent views as follow (8),

$$SAD_{view_n} = SAD_{view_0} \times \frac{target_n}{target_0}, \quad (n=1,2) \quad (8)$$

where SAD_{view_n} is the complexity of the n th view, $target_n$ is the target bits of view n .

III. EXPERIMENTAL RESULTS

The experiments are conducted under 5-view scenario, where 3 views are coded, and 2 virtual views are synthesized. The testing sequences include "Kendo", "Balloons", "Newspaper", "GT_Fly", "Poznan_Hall2", "Poznan_Street", "Undo_Dancer" and "Shark". The texture and depth map of three views are separately encoded with HTM10.0 encoder started as I-view, P-view and P-view respectively. For P-frame, the interview prediction is involved. 300 frames are encoded for each view.

A. R-D Performance and Rate Accuracy

In order to evaluate the performance of the proposed RC algorithm, R-lambda algorithm proposed in [6] is utilized for comparison. First, the R-D curve between proposed and R-lambda is shown in Fig.3. Then the performance compare with RC in the latest reference software HTM10.0 will be shown in Table II.

To evaluate the accuracy of the bit rate achievement, the following measurement is adopted.

$$Error = \frac{|R_{actual} - R_{target}|}{R_{target}} \times 100\% \quad (9)$$

where $Error$ is the bit error, R_{actual} is the total bits used to encode the depth map and texture map of three views; R_{target} is the target bits of that.

The errors of proposed, URQ model and R-lambda model are presented in Table II. We can see that the proposed RC scheme is accurate enough that the mismatch is less than 1.0% on average.

Fig. 3 shows the R-D curves of the two algorithms for sequences "Poznan_Hall2" and "GT_Fly". From that, we can see that the proposed scheme can achieve a larger PSNR value at each target bit rate for each of the four sequences than the R-lambda scheme.

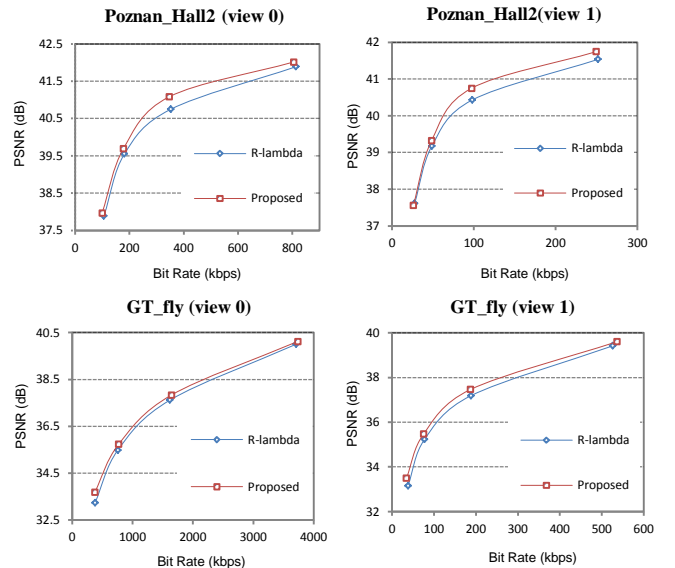


Fig. 3 The rate distortion curves of the proposed rate control scheme as well as the rate control in reference software HTM10.0 (R-lambda).

TABLE II THE PERFORMANCE BETWEEN R-LAMBDA AND PROPOSED RC SCHEME

	Proposed vs R-lambda				Proposed vs URQ				Bitrate Error		
	Video 0	Video 1	Video 2	Video PSNR	Video 0	Video 1	Video 2	Video PSNR	URQ bitrate error	R-lambda bitrate error	Proposed bitrate error
Balloons	-3.2%	-4.5%	-5.2%	-3.9%	-25.3%	-45.0%	-43.4%	-33.0%	0.59%	1.47%	0.26%
Kendo	-4.3%	-4.8%	-5.3%	-4.3%	-33.0%	-50.2%	-50.0%	-39.9%	0.82%	0.58%	0.33%
Newspaper_CC	-1.8%	-2.0%	-2.8%	-2.5%	-27.6%	-55.2%	-42.2%	-37.2%	1.83%	2.61%	0.49%
GT_Fly	-6.8%	-12.5%	-12.3%	-8.4%	-31.1%	-52.4%	-51.8%	-36.4%	0.64%	1.46%	1.58%
Pozan_Hall2	-12.4%	-12.2%	-12.1%	-12.2%	-14.5%	-30.9%	-32.4%	-21.9%	0.69%	2.88%	0.44%
Poznan_Street	-3.9%	-5.2%	-4.3%	-4.1%	-33.4%	-48.7%	-69.4%	-44.7%	0.32%	6.01%	3.53%
Undo_Dancer	-6.6%	-6.1%	-6.5%	-6.3%	-24.8%	-42.2%	-40.9%	-30.4%	0.33%	0.98%	0.31%
Shark	-2.8%	-4.4%	-3.8%	-3.6%	-12.1%	-16.8%	-14.3%	-14.4%	0.51%	1.43%	0.40%
1024*768	-3.1%	-3.7%	-4.4%	-3.6%	-28.7%	-50.1%	-45.2%	-36.7%	1.08%	1.55%	0.36%
1920*1088	-6.5%	-8.1%	-7.8%	-6.9%	-23.2%	-38.2%	-41.7%	-29.6%	0.50%	2.55%	1.25%
average	-5.2%	-6.5%	-6.6%	-5.7%	-25.2%	-42.7%	-43.0%	-32.3%	0.72%	2.18%	0.92%

B. Verification the Smoothness of Quality

In this section, we verify the effectiveness of the proposed bits allocation scheme in Section II-B. Because of the quality is unacceptable comparing with the other two algorithms. We choose the sequence *Kendo*, initial QP 25, the generated bits curve between proposed RC algorithm and R-lambda algorithm in HTM10.0. As shown below in Fig. 4, R-lambda model occur bit fluctuation more severe, which can make some transmission delay. Thus, for strict real-time applications proposed scheme is more suitable than R-lambda model.

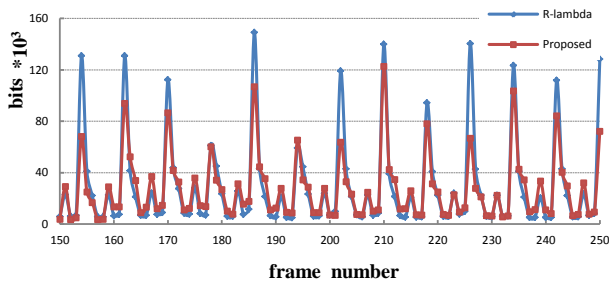


Fig. 4 generated bits curve of "Kendo"

IV. CONCLUSIONS

In this paper, we proposed a frame level RC algorithm to achieve the better quality for 3D-HEVC. Based on our models for the R-Qs and the bits allocation scheme, an R-D optimized RC algorithm is derived. Experiments are conducted on different video sequences and the results demonstrate the effectiveness of the proposed algorithm.

ACKNOWLEDGEMENT

This work was supported in part by the National High-tech R&D Program of China (863 Program, 2012AA010805), National Science Foundation of China (61322106, 61390515), which are gratefully acknowledged.

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529-551, Apr. 1955.
- [2] "3D-HEVC Test Model Description draft 1," JCT on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11/JCT2-A1005, Sep. 2012.
- [3] B. Bross, W.J. Han, J. Ohm, G. Sullivan and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 8," *JCTVC-J1003*, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-TSG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-J1003_d2.doc, 10th Meeting: Stockholm, Sweden, 11-20 July, 2012.
- [4] Siwei Ma, Shiqi Wang, and Wen Gao, "Zero-Synthesis View Difference Aware View Synthesis Optimization For HEVC Based 3D Video Compression", *Visual Communication and Image Processing*, San Diego, USA, Nov.2012.
- [5] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajić, "Rate control based on unified RQ model for HEVC," *JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCT-VC H0213 (m23088)*, San José, CA, USA, Feb. 2012.
- [6] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," *ITU-T SG16 Contribution, JCTVC-K0103*, Shanghai, pp.1-5, Oct. 2012.
- [7] W. Lim, D. Sim and I. V. Bajić, "JCT3V – Improvement of the rate control for 3D multi-view video coding," *ISO/IEC JTC1/SC29/WG11, JCT3V-C0090*, Geneva, Switzerland, Jan. 2013.
- [8] J. J. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 289 - 296, Sept. 1963.
- [9] J. Si, S. Ma, W. Gao, "JCT-VC-Adaptive rate control for HEVC" *ISO/IEC JTC1/SC29/WG119, JCTVC-I0433*, Geneva, CH, 27 April-7 May 2012.
- [10] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP Based Rate Control for High Efficiency Video Coding", *IEEE Journal of Selected Topics in Signal Processing*, vol.7, no.6, pp.1101-1111, Dec.2013.