QP INITIALIZATION AND INTERVIEW MAD PREDICTION FOR RATE CONTROL IN HEVC-BASED MULTI-VIEW VIDEO CODING

Woong Lim[†], *Ivan V. Bajić*[‡], *and Donggyu Sim*[†]

[†]Kwangwoon University, Korea, [‡]Simon Fraser University, Canada

ABSTRACT

Rate control is an important component of an end-to-end video communication system. Currently, there are several proposals for rate control in the upcoming High Efficiency Video Coding (HEVC) standard, but none specifically for the multi-view extension of the standard. In this paper, we apply one of the HEVC single-view rate control schemes to the multi-view scenario, and propose two improvements. One improvement deals with Quantization Parameter (QP) initialization, and the other deals with interview Mean Absolute Difference (MAD) prediction. Experimental results demonstrate increased accuracy of rate control, reduced fluctuation of instantaneous bitrate, as well as a reduction in PSNR degradation compared to the existing rate control algorithm.

Index Terms— HEVC, multi-view video coding, rate-control, mean absolute difference, depth map

1. INTRODUCTION

High Efficiency Video Coding (HEVC) is an upcoming video compression standard [1] that delivers higher compression efficiency compared to earlier standards, especially for high-resolution video. The work on multi-view extension of HEVC started in March 2011 by the MPEG 3DV group [2]. The goal of the multi-view video coding standard is to provide efficient compression and high quality view reconstruction of an arbitrary number of densely-spaced views [3]. The multi-view standard will address coding of both texture and depth maps, to allow high quality view synthesis between existing views.

Although not a part of the standard, rate control is important for practical deployment of systems and services based on HEVC and its extensions. Rate control refers to the adjustment of coding parameters, especially Quantization Parameter (QP), on a Group-Of-Pictures (GOP), frame, or block-basis, with a goal towards controlling the encoder's output bitrate [4], [5]. This is important for end-to-end video communication system design in applications like broadcasting and streaming, because high fluctuation of bitrate may cause an overflow or underflow of buffers along the communication path(s).

The first rate control scheme adopted in the HM reference software for HEVC was based on a Unified Rate-Quantization (URQ) model [6], and has since been improved [7]. More recently, another HEVC rate control scheme based on the so-called R-lambda model [8] was adopted into HM reference software. Unfortunately, the HM software version incorporating R-lambda rate control was not released until Nov. 27, 2012, so it was not available during the writing of this paper. Our results are therefore reported in the context of the earlier URQ model [6], [7]. It should be noted, however, that the methods proposed in this paper (namely, QP initialization and MAD prediction) are external to the rate control algorithm itself, and are therefore applicable to the R-lambda scheme [8] as well.

The paper is organized as follows. In Section 2, we briefly outline the single-view HEVC rate control based on the URQ model, and provide an overview of rate control in multi-view video coding. Section 3 describes the proposed QP initialization and interview MAD prediction. Section 4 presents experimental results, while Section 5 provides conclusions.

2. PRELIMINARIES

2.1. URQ model-based rate control

The URQ model-based rate control [6], [7] relies on GOP-, frame-, and CTU-level bit budget control. At the GOP level, the available bits for the current GOP are computed based on the balance of the bits spent on encoding previous GOPs relative to the bit budget. Any shortage or excess of bits is carried forward, as illustrated in Fig. 1.

At the frame level, the number of available bits for the current frame is computed, from which a QP value is found using a pixel-based URQ model in equation (1),

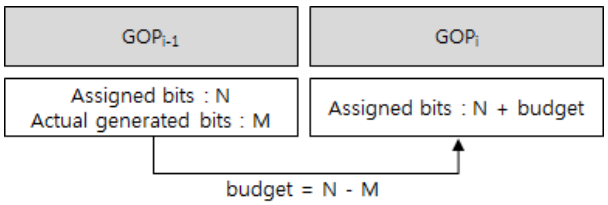

Fig 1. GOP-level bit budget control

$$\frac{T_i(j)}{N_{pixels,i}(j)} = \alpha \cdot \frac{MAD_{pred,i}(j)}{Qstep_i(j)} + \beta \cdot \frac{MAD_{pred,i}(j)}{Qstep_i^2(j)} \quad (1)$$

where $N_{pixels,i}(j)$ is number of pixels in the j-th frame of the i-th GOP, and $T_i(j)/N_{pixels,i}(j)$ is the target average bit rate for this frame in bits per pixels (bpp). To compute the quantizer step size, $Q_{step,i}(j)$, the model needs the Mean Absolute Difference value $MAD_{pred,i}(j)$, which is predicted from a previously encoded frame. At the CTU level, a similar computation based on MAD predicted from previously encoded CTUs is performed to find the appropriate quantizer step size, and subsequently the QP for the current CTU.

2.2. Rate control for HEVC-based multi-view coding

We incorporated URQ model-based rate control into the current 3D-HTM software, which is the HEVC-based reference software for multi-view video coding used by the MPEG 3DV group [2]. The block diagram of a rate control scheme for multi-view coding is shown in Fig. 2. For the base view, the rate control scheme is the same as that for single-view HEVC. A straightforward way to achieve rate control for extended views is to use single-view rate control on each view separately; this approach will be the benchmark method against which the proposed methods will be compared.

Most rate control schemes use previously generated bits and prediction errors (usually in the form of MAD) to predict the to-be-generated bits for the current coding unit [9], [10], [11]. Fig. 2 indicates with shading the functional blocks proposed in the present paper for multi-view rate control. Since these blocks, namely QP initialization and MAD prediction, are external to the rate control modules, they can be used in conjunction with a variety of rate control schemes, including the new R-lambda scheme for HEVC [8], and possibly others. The URQ model-based scheme was chosen as a platform to test the proposed methods mainly due to its availability in the reference software at the time of writing this paper.

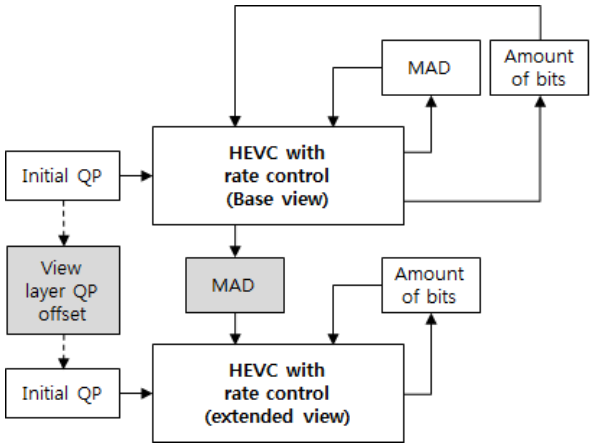

Fig. 2. Rate control for HEVC multi-view extension

3. PROPOSED METHODS

A frame in the extended view exhibits a high degree of similarity compared to the base view frame with the same Picture Order Count (POC), because both frames are acquired at the same time by cameras at slightly different positions. Hence, not only texture, but various coding parameters of the two frames are likely to be highly related. The proposed methods attempt to use base-view parameters to improve rate control in extended views.

3.1. QP initialization in extended views

In the existing HEVC rate control methods [6], [7], [8], at the beginning of encoding, initial QP for the first frame in the base view is decided based on the target bitrate using an empirically obtained table of QP values. One could use the same strategy for extended views as well. However, unlike the base view, the first frame in each GOP in extended views is inter-coded relative to the reconstructed first frame in the base view, so a different set of QP values would seem more appropriate. Fortunately, there is a simple solution. The 3D-HTM configuration file defines QP offsets for hierarchical coding of multi-view video. The default values of these offsets were obtained empirically based on extensive testing under common test conditions of the MPEG 3DV group, and are thought to be appropriate from the rate-distortion point of view. Hence, we initialize the QP of the first frame of each GOP in extended views as

$$QP_{ext,i}(1) = QP_{base,i}(1) + ViewLayerQPOffset \quad (2)$$

where $ViewLayerQPOffset$ is the pre-defined QP offset between view layers, while $QP_{ext,i}(1)$ and $QP_{base,i}(1)$ are QPs of the first frame in the i-th GOP of the extended view and the base view, respectively.

3.2. Depth map-based interview MAD prediction

In order to select a QP value for the current CTU, a rate control algorithm needs to predict the MAD value of the current CTU using previously encoded information. In single-view rate control, this is done by using the MAD of a block in the reference frame, an approach we shall call temporal MAD prediction. While this may work well in the case of low motion, temporal MAD prediction is less accurate as the motion level increases, and becomes completely unreliable in case of scene changes.

For multi-view video coding, a more accurate MAD prediction for extended views may be obtained by using the information from the base view. The proposed approach is illustrated in Fig. 3. The basic idea is to use the available depth map to find the position of the current extended-view

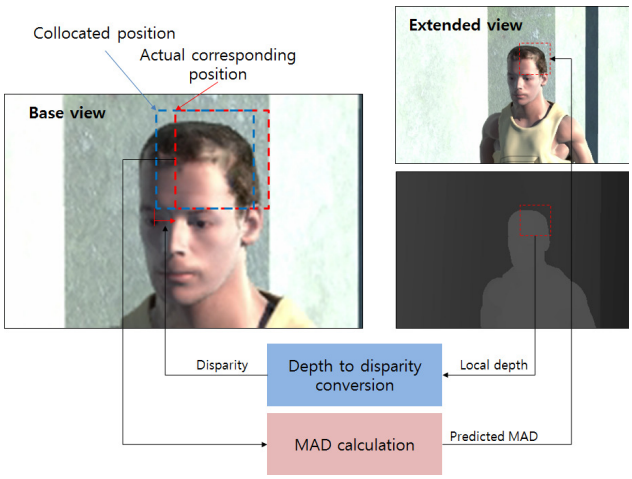

Fig. 3. Proposed interview MAD prediction

CTU within the base-view frame, and use the MAD of the pixels at that position as a predicted value for MAD of the current CTU.

To find the position of a block in the base view that corresponds to the current CTU, we proceed as follows. Let D_{avg} be the average disparity value of a block in the depth map corresponding to the current extended-view CTU. The disparity value D between the extended-view CTU and its position in the base view can be found following [12]

$$Cam_{trans} = CamPos_{left} - CamPos_{right} \tag{3}$$

$$Z = \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) \cdot \frac{D_{avg}}{D_{max}} + \frac{1}{Z_{far}} \tag{4}$$

$$D = FocalLen \cdot Cam_{trans} \cdot Z \tag{5}$$

where $CamPos_{left}$ and $CamPos_{right}$ are the left and right camera position, respectively, Z_{near} and Z_{far} are the actual nearest and furthest depth in the view, D_{max} is the maximum depth map value (usually 255) and $FocalLen$ is the camera's focal length. These parameters are available in the high-level syntax in 3D multi-view coding [12].

The following example illustrates the benefits of interview MAD prediction. Fig. 4 shows the actual CTU-wise MAD labeled as "a-MAD", the MAD predicted temporally (as in conventional single-view rate control) labeled as "t-MAD" and the MAD predicted using the approach described above, labeled as "iv-MAD", on two segments of the 'Dancer' sequence. The top plot corresponds to a segment with slow motion, and the bottom plot corresponds to a segment with faster motion. Observe that in the top plot, all three curves essentially overlap, whereas in the bottom plot, the iv-MAD curve is closer to the a-MAD curve, especially in the cases where the actual MAD is above 6 (i.e., when the motion is high). This is

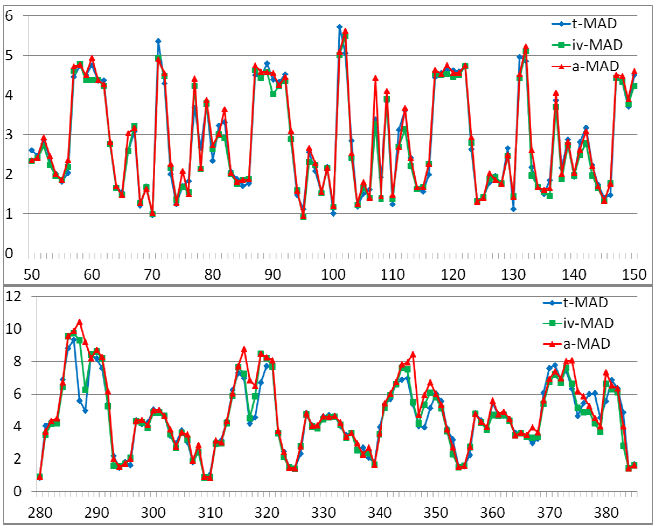

Fig. 4. Horizontal axis: CTU index, vertical axis: actual and predicted MADs. Slow motion (top), and fast motion (bottom)

further confirmed quantitatively. In the top plot, the Pearson correlation between t-MAD and a-MAD is 0.981, while the correlation between iv-MAD and a-MAD is 0.990, both very high values. Meanwhile, in the bottom plot, the correlation between t-MAD and a-MAD is 0.919, while that between iv-MAD and a-MAD is 0.975, a considerably higher value. Similarly, the average prediction Mean Squared Error (MSE) of t-MAD in the top plot is 0.065, and that of iv-MAD is 0.049. In the bottom plot, the average MSE of t-MAD is 0.931, and that of iv-MAD is 0.375, almost three times lower. In summary, interview MAD prediction is more accurate than the conventional temporal MAD prediction, especially in the case of fast motion.

4. EXPERIMENTAL RESULTS

The proposed methods were implemented in the 3D-HTM 4.0.1 reference software and evaluated under the conditions listed in Table 1 for the Constant Bit Rate (CBR) case. Each sequence was encoded at four target bitrates as follows. First, the 3D-HTM 4.0.1 reference encoder was used to encode the sequences with four different fixed QP values, as specified in the common test conditions [13]. The resulting bitrates for the three views were then used as target bitrates for the rate-control-enabled encoder. The total target bitrates

Table 1. Sequences and encoding conditions

Sequence	Resolution	fps	Num of views	Coding structure
Balloons	1024×768	30	3	Random access
Kendo	1024×768	30	3	Random access
Newspaper	1024×768	30	3	Random access
GTFly	1920×1088	25	3	Random access
Poznanhall2	1920×1088	25	3	Random access
Poznanstreet	1920×1088	25	3	Random access
Dancer	1920×1088	25	3	Random access

(all views together) were in the range of 250-1700 kbps for 1024×768 sequences, and in the range of 165-6550 kbps for 1920×1088 sequences. The reason for the larger range of rates at higher resolution is the larger variability of motion content among these sequences.

Table 2 shows the average percentage error in the actual generated bitrate compared to the target bitrate. We show the results for the two extended views (view 1 and view 2) for two cases: when the proposed methods are switched off (i.e, when using the benchmark rate control scheme by itself), and when they are switched on. As seen in the table, using the proposed methods reduces the average error to below 1%.

Rate control without bit allocation usually degrades PSNR performance, because quantization decisions are made based on the buffer occupancy alone. Table 3 shows the PSNR degradation (relative to the 3D-HTM coder without rate control) in the extended views for the two cases. As seen in the table, the proposed methods result in lower degradation in most cases, with an average reduction of PSNR degradation of 0.06 dB and 0.01 dB in views 1 and 2, respectively.

The purpose of rate control is not only to control the total average bitrate, but also to limit fluctuations in the instantaneous bitrate. To assess this aspect of the proposed methods, in Table 4 we show the standard deviation of the bits generated per second for the two cases in extended views. As seen in the table, the proposed methods reduce the standard deviation in instantaneous bitrate compared to the benchmark method. The average reduction of instantaneous bitrate fluctuations under the test conditions, across all sequences, was about 3300 bps, as indicated in the rightmost column in the table.

A graphical illustration of instantaneous bit generation is provided in Fig. 5, which shows the accumulated occupancy of the virtual buffer vs. frame index for view 1 of the 'Balloons' sequence with the target bitrate of 311 kbps. In this graph, the value 0 on the vertical axis means that upon encoding the current frame, the total number of bits spent is exactly as required by the target bitrate. As seen in the figure, proposed methods lead to a much reduced fluctuation in the buffer occupancy compared to the benchmark rate control method.

5. CONCLUSIONS

We proposed two methods to improve rate control of extended views in multi-view video coding based on HEVC. The proposed methods were implemented and tested on the 3D-HTM 4.0.1 reference software. It was demonstrated that interview MAD prediction leads to more accurate MAD prediction compared to conventional methods, especially in the case of fast motion. The two proposed methods led to better rate control accuracy, reduced fluctuation of instantaneous bitrate, as well as a small reduction of PSNR degradation, compared to the benchmark method.

Table 2. Percentage error in the actual bitrate (%)

Sequence	Proposed methods off		Proposed methods on	
	average view1	average view2	average view1	average view2
Balloons	0.35	0.49	0.48	0.71
Kendo	0.68	0.86	0.98	0.69
Newspaper	2.48	2.51	2.00	2.10
GTFly	6.40	2.84	0.96	0.99
Poznanhall2	0.96	1.79	0.49	0.96
Poznanstreet	4.04	1.25	0.27	0.46
Dancer	1.03	0.67	0.66	0.69
Average	**2.28**	**1.49**	**0.83**	**0.94**

Table 3. PSNR degradation in extended views (dB)

Sequence	Proposed methods off		Proposed methods on	
	average view1	average view2	average view1	average view2
Balloons	1.64	1.53	1.62	1.52
Kendo	2.29	1.43	1.86	1.16
Newspaper	1.51	1.40	1.65	1.64
GTFly	0.94	0.98	0.84	0.86
Poznanhall2	0.56	0.48	0.53	0.49
Poznanstreet	1.39	1.36	1.53	1.59
Dancer	1.15	1.15	0.98	0.98
Average	**1.35**	**1.19**	**1.29**	**1.18**

Table 4. Standard deviation of bits per second (bps)

Sequence	Proposed methods off		Proposed methods on		Avg. diff.
	avg view1	avg view2	avg view1	avg view2	
Balloons	8303	9365	5843	6449	2688
Kendo	5582	6478	4709	5416	967
Newspaper	12001	11618	8900	8648	3036
GTFly	16735	15604	13616	13372	2676
Poznanhall2	6722	7648	7548	8132	–655
Poznanstreet	24353	27542	16044	18906	8473
Dancer	24116	21025	17025	15276	6420
Average	**13973**	**14183**	**10526**	**10886**	**3372**

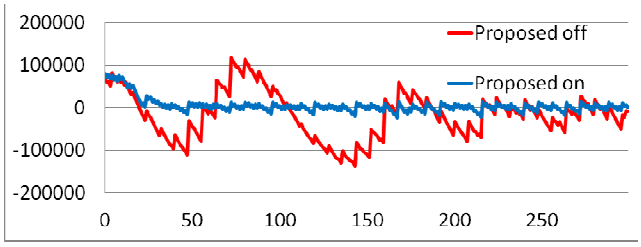

Fig. 5. Virtual buffer occupancy vs. frame index

6. REFERENCES

[1] G. J. Sullivan, J. R. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," to appear in *IEEE Trans. Circuits Syst. Video Technol.*, 2013.

[2] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on 3D Video Coding Technology," *Doc. N12036*, Geneva, Switzerland, Mar. 2011.

[3] ISO/IEC JTC1/SC29/WG11, "Applications and Requirements on 3D Video Coding," *Doc. N12035*, Geneva, Switzerland, Mar. 2011.

[4] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuit Syst. Video Technology*, vol. 13, no. 7, pp. 688-703, Jul. 2003.

[5] J. W. Woods, *Multidimensional Signal, Image, and Video Processing and Coding*, Second Edition, Elsevier/Academic Press, 2012.

[6] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajić, "Rate control based on unified RQ model for HEVC," JCT-VC H0213 (m23088), JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, San José, CA, USA, Feb. 2012.

[7] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajić, "Improvement of the rate control based on pixel-based URQ model for HEVC," JCTVC-I0094 (m24333), ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Apr.-May 2012.

[8] B. Li, H. Li, L. Li and J. Zhang, "Rate control by R-lambda model for HEVC," *JCT-VC K0103*, JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Shanghai, CN, Oct. 2012.

[9] S. Ma, W. Gao and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuit Syst. Video Technology*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.

[10] Y. Liu, Z. Li and Y. Soh, "A novel rate control scheme for low delay video communication of H.264/AVC standard," *IEEE Trans. Circuit Syst. Video Technology*, vol. 17, no. 1, pp. 68-78, Jan. 2007.

[11] S. Zhou, J Li, J Fei and Y. Zhang, "Improvement on rate-distortion performance of H.264 rate control in low bit rate," *IEEE Trans. Circuit Syst. Video Technology*, vol. 17, no. 8, pp. 996-1006, Aug. 2007.

[12] D. Rusanovskyy and M. M. Hannuksela, "Suggestion for a depth-enhanced multiview video coding extension to H.264 Annex A: Nokia 3DV Test Model (3DV-TM) Codec Description and Simulation Results," ITU-T SG16 VCEG-AR14, San José, CA, USA, Feb. 2012.

[13] D. Rusanovskyy, K. Müller, and A. Vetro, "Common Test Conditions of 3DV Core Experiments," JCT2-A1100, ISO/IEC JTC1/SC29/WG11, Stockholm, Sweden, Jul. 2012.