**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 MPEG2012/M25966**

**July 2012, Stockholm, Sweden**

|  |  |
| --- | --- |
| **Source** | **Kwangwoon University (KWU) and Simon Fraser University (SFU)** |
| **Status** | **Proposal** |
| **Title** | **3D-HEVC – Rate control for 3D multi-view video coding** |
| **Author** | **Woong Lim (limwoong@kw.ac.kr)**  **Hyomin Choi (hyomin06@kw.ac.kr)**  **Junghak Nam (qejixfyza@kw.ac.kr)**  **Hyunho Jo (idjhh@kw.ac.kr)**  **Donggyu Sim (dgsim@kw.ac.kr)**  **Ivan V. Bajić (ibajic@ensc.sfu.ca)** |

# Abstract

This contribution presents a rate control scheme for multi-view video coding based on the unified rate-quantization (URQ) model for HEVC described in JCTVC-H0213 [1] and JCTVC-I0094 [2]. The proposed rate control scheme is implemented on 3D-HTM3.1. The proposed rate-control scheme includes all features described in JCTVC-H0213 and JCTVC-I0094, which means that it includes GOP, frame, and unit (LCU) level rate control based on the quadratic pixel-based unified rate-quantization model. In addition, this contribution describes initial QP setting for the first frame of extended view, as well as depth map-based inter-view MAD (Mean Absolute Difference) prediction. In the case of extended view, the first frame doesn’t have temporal reference frame but has an inter-view reference frame in the base view, so it is possible to remove the view-axis spatial redundancy utilizing a P-frame coding scheme. Hence, the initial QP for the first frame of the extended view needs to be different from the QP value calculated by using the conventional pixel-based URQ model. For the MAD prediction for extended view, depth map based inter-view MAD prediction from the previously coded and neighbouring view is applied. To evaluate the proposed multi-view rate control scheme, 3D-HTM3.1, which uses hierarchical QP setting, is used for the constant bit rate (CBR) case.

# Introduction

A rate control algorithm for HEVC, described in JCTVC-H0213, was adopted in the reference HM software at the 8th San Jose meeting in February 2012. Some improvements were reported in JCTVC-I0094. While the approach from JCTVC-H0213 and JCTVC-I0094 can be applied to HEVC-based multi-view encoding, some modifications are necessary in order to adapt the rate control procedure to the multi-view scenario.

First, we should consider the initial QP value of the first frame of the extended views. 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. In general, an inter frame generates less bits than an intra frame due to the removal of redundancy between the current frame and reference frame(s). Therefore, the QP value of the first frame of an extended view should be set lower than the QP value of the first frame of the base view. For this purpose, we make use of the predefined QP offsets for hierarchical coding.

Furthermore, for stable generation of bits, we propose inter-view MAD prediction for extended views. In single view encoding, rate control scheme uses the temporal MAD factor to predict the complexity of texture for the current frame. In single-view coding, to predict the complexity of texture, there is no other choice but to use the previously encoded frame along temporal axis to calculate temporal MAD. But in multi-view coding, the previously encoded frame(s) in neighbouring view(s) can also be employed. In addition, depth map can be used to find the corresponding position of the current unit block in the neighbouring view. With these proposed schemes, we found that the average bitrate error of the proposed rate control is about 0.89% (i.e., the accuracy is better than 99%), with an average PSNR degradation about 2.1dB.

# Rate control for the base view

The existing rate control in HEVC creates a bitstream with the bitrate that matches the available channel bitrate for single-view video. The algorithm takes as inputs several pieces of information, such as the available channel bitrate, coding structure, frame rate, GOP size etc.. Specifically, the rate control scheme employs a pixel-based unified rate quantization (URQ) model to produce the bitstream with a specified bitrate. It is fairly suitable for HEVC, which has diverse coding block sizes, as it avoids having a separate model for each block size. In multi-view video coding, the base view is coded as a single-view video, without prediction from neighbouring views. Hence, the rate control algorithm for the base view is the same as HEVC rate control [2].

# Rate control for extended views

The proposed rate control scheme considers two additional issues that are specific to multi-view coding. The first is the decision about initial QP for the first frame in extended view(s). In multi-view video coding, the compression efficiency of extended view(s) is generally higher than that of the base view, because of inter-view prediction. In particular, most gain comes from the fact that the first frame of extended view is coded as an inter frame, rather than intra frame. In the existing HEVC rate control scheme, initial QP setting for the first frame is calculated based on the target bitrate in bits per pixel (bpp), with the underlying assumption that the first frame is coded as an intra frame. Therefore, the initial setting of QP for the first frame in extended view should be lower than the first frame in base view.

The second multi-view coding-specific issue we consider is MAD prediction. Conventionally, texture complexity in rate control is determined based on MAD, which is predicted from temporal reference frames in single-view coding. In multi-view coding, however, we also have the option of using the previously-encoded neighbouring view for MAD prediction. Hence, we propose to predict the MAD value of the current unit by referencing the MAD value of the corresponding region from the previously encoded and decoded base view. The correspondence between the views is determined by using the depth map.

## Initial QP decision for the first frame in an extended view

At the beginning of encoding, the first frame of an extended view is an inter-view frame. For the first frame of the base view (intra frame), the conventional approach to set the initial QP is as follows.

For deriving the QP, *bppinit* should be calculated with available bitrate  by formula (1).

 (1)

*Npixels* is the number of pixels in a frame, *f* indicates the predefined frame rate and the initial QP is set by the value of *bppinit* as follows,

 (2)

This approach works well for the first frame of the base view, which is an intra frame, but it is not appropriate for the first frames of extended views, which are inter frames. To take this into account, we offset the value determined in (2) by the QP offsets among views, which are predefined in the encoder configuration file.

 (3)

*ViewLayerQPOffset* indicates the QP offset between view layers. The first frame in subsequent GOPs employs QP value calculated by formula (3) and the value can be modified by formula (4)

 (4)

where *Nref* indicates the number of reference frames in a GOP, and the index *k* is a frame index among the reference frames. Equation (5) represents clipping of *QPi*(1) to [*QPi*−1(*NGOP*)−2, *QPi*−1(*NGOP*)+2], so that it is within ±2 of the QP of the last frame in the previous GOP, for smoothness. After decision of *QPi*(1), it is modified according to the condition that the value of *Vi*(j)+ *Ii*(j). *Vi*(j) and *Ii*(j) are the occupancy of virtual buffer and the status of initial buffer respectively.

 (5)

## Local depth map-based inter-view MAD prediction

The encoder with rate control scheme should generate bits to conform with the pre-specified bitrate. To do this, rate control algorithm calculates the available bits and restricts the amount of bits in a GOP level, frame level and unit level. The algorithm also decides the best QP for the current unit by using the amount of error (MAD) in a collocated unit in the previously encoded frame. This means that the algorithm should predict the to-be-generated bits for the current unit, by making use of the information from the previous frame. However, the temporally referenced information may be inaccurate when the scene changes or fast motions occur between frames. To prevent this problem, we propose the unit-level inter-view MAD prediction based on the depth map

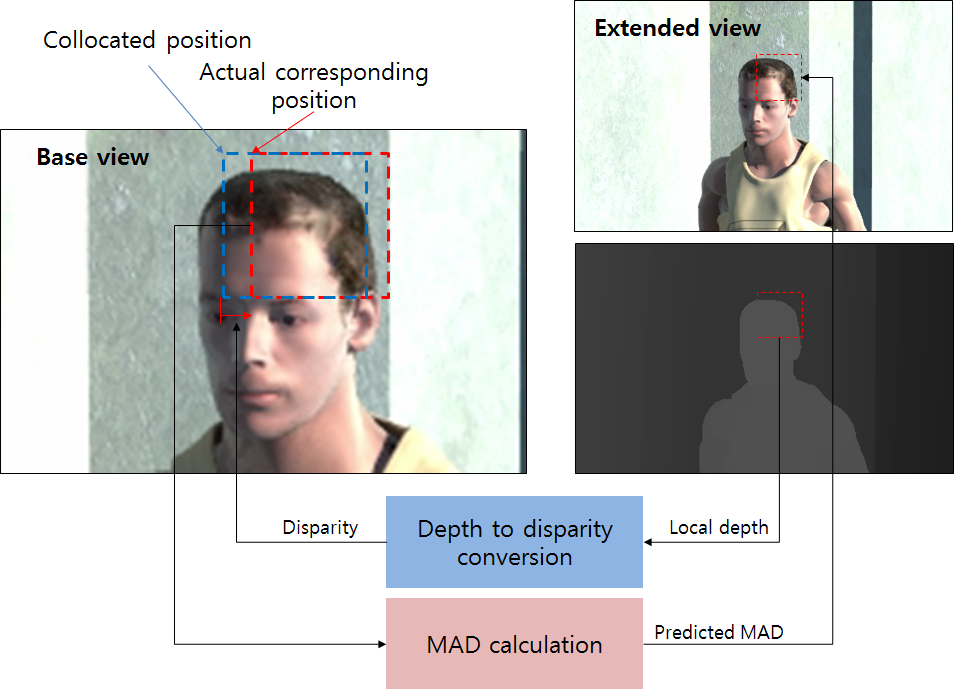


Figure 1. Overall structure of depth map based inter-view MAD prediction for a coding unit

In the current 3D multi-view video encoding scheme, the base view is coded independently, but extended view is dependent on the base view by referencing the reconstructed base view. A frame in the base view and a frame in extended view with the same POC are acquired at the same time. Therefore, the disparities between object locations in these frames depend only on the object’s distance from the cameras. By using the depth map, we can calculate the disparity of a coding unit and find out the corresponding position in the neighbouring view. Since the two frames with the same POC are temporally collocated, scene changes and fast motion do not affect this correspondence. The proposed local depth for the current unit is calculated by the following formula (6).

 (6)

*CamPosleft* and *CamPosright* stand for the camera position of left and right views respectively, while *Camtrans* represents the distance between the views. The perceived depth *Z* can be calculated by formula (7).

 (7)

*Znear*, *Zfar*, *Dmax* and *Davg* stand for the nearest depth, the furthest depth, the maximum depth intensity and the average local depth intensity, respectively. The local depth intensity is calculated by formula (8). Fig. 2 shows the coarse down-sampling of the depth map.

 (8)

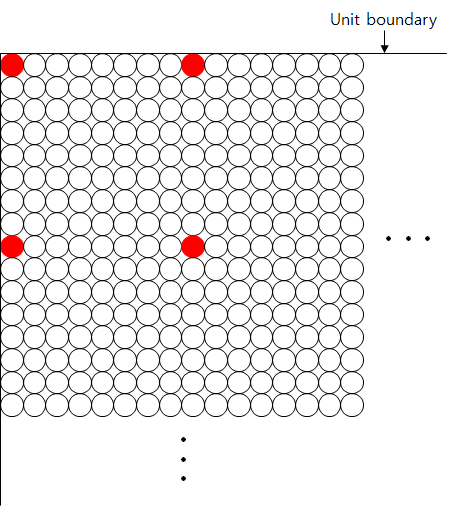


Figure 2. Sub-sampling of the depth map of the corresponding coding unit

*N* and *M* are the down-sampling factors in the vertical and horizontal directions, respectively. They are set to 1/8 of the unit’s width and height, respectively. We can calculate the disparity by using formula (9).

 (9)

The camera parameters (*CamPosleft*, *CamPosright*, *Znear*, *Zfar*, *FocalLen*) are encoded in the high-level syntax in 3D multi-view coding. Hence, based on this information, we can calculate the local depth for the current coding unit.

# Experimental results

The proposed rate control for multi-view video coding is estimated under the CBR condition. Table 1 shows the delta bitrate to show the overall performance. The proposed rate control is evaluated with the same condition of the common random access condition of the current 3D multi-view coding. To show the difference of the number of bits, we use % kbps error defined by formula (10).

 (10)

Table 1. Performance of view-wise actual generated bits of 3D-HTM rate control and the existing 3D-HTM

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sequence** | **target total** | **Target rate V0** | **Target rate V1** | **Target rate V2** | **view0** | **view1** | **view2** | **% kbps error view0** | **% kbps error view1** | **% kbps error view2** | **overall kbps error(%)** |
| **Balloons** | 1540.00 | 876.00 | 316.00 | 348.00 | 868.90 | 313.33 | 344.69 | 0.81 | 0.85 | 0.95 | 0.85 |
| 803.00 | 482.00 | 153.00 | 168.00 | 478.04 | 151.90 | 166.55 | 0.82 | 0.72 | 0.86 | 0.81 |
| 452.00 | 278.00 | 83.00 | 91.00 | 275.86 | 81.87 | 90.07 | 0.77 | 1.36 | 1.02 | 0.93 |
| 267.00 | 168.00 | 47.00 | 52.00 | 166.75 | 46.81 | 51.58 | 0.74 | 0.40 | 0.82 | 0.70 |
| **Kendo** | 1507.00 | 844.00 | 319.00 | 344.00 | 836.37 | 316.70 | 341.17 | 0.90 | 0.72 | 0.82 | 0.85 |
| 765.00 | 451.00 | 151.00 | 163.00 | 448.59 | 149.92 | 162.03 | 0.53 | 0.71 | 0.60 | 0.58 |
| 427.00 | 257.00 | 82.00 | 88.00 | 255.36 | 81.20 | 87.42 | 0.64 | 0.97 | 0.66 | 0.71 |
| 251.00 | 155.00 | 47.00 | 49.00 | 153.83 | 46.69 | 48.87 | 0.76 | 0.67 | 0.26 | 0.64 |
| **Newspaper** | 1663.00 | 950.00 | 333.00 | 380.00 | 939.59 | 326.45 | 373.19 | 1.10 | 1.97 | 1.79 | 1.43 |
| 821.00 | 500.00 | 149.00 | 172.00 | 494.30 | 145.19 | 174.36 | 1.14 | 2.56 | 1.37 | 0.87 |
| 441.00 | 275.00 | 77.00 | 89.00 | 267.32 | 74.69 | 86.80 | 2.79 | 3.00 | 2.48 | 2.76 |
| 253.00 | 160.00 | 44.00 | 49.00 | 154.82 | 43.34 | 48.53 | 3.24 | 1.49 | 0.96 | 2.49 |
| **Gtfly** | 4938.00 | 3668.00 | 634.00 | 636.00 | 3647.57 | 626.73 | 629.89 | 0.56 | 1.15 | 0.96 | 0.68 |
| 2084.00 | 1610.00 | 237.00 | 237.00 | 1600.65 | 233.16 | 235.22 | 0.58 | 1.62 | 0.75 | 0.72 |
| 970.00 | 754.00 | 108.00 | 108.00 | 750.93 | 106.71 | 107.05 | 0.41 | 1.20 | 0.88 | 0.55 |
| 471.00 | 366.00 | 52.00 | 53.00 | 363.31 | 52.48 | 53.56 | 0.74 | 0.92 | 1.05 | 0.35 |
| **PoznanHall2** | 1367.00 | 810.00 | 273.00 | 284.00 | 799.28 | 267.41 | 272.13 | 1.32 | 2.05 | 4.18 | 2.06 |
| 572.00 | 344.00 | 113.00 | 115.00 | 342.55 | 111.89 | 119.67 | 0.42 | 0.99 | 4.06 | 0.37 |
| 299.00 | 178.00 | 60.00 | 61.00 | 176.13 | 58.87 | 61.01 | 1.05 | 1.88 | 0.02 | 1.00 |
| 167.00 | 100.00 | 34.00 | 33.00 | 97.51 | 34.36 | 33.26 | 2.49 | 1.05 | 0.78 | 1.13 |
| **PoznanStreet** | 3935.00 | 2427.00 | 762.00 | 746.00 | 2412.10 | 758.08 | 741.63 | 0.61 | 0.51 | 0.59 | 0.59 |
| 1407.00 | 956.00 | 228.00 | 223.00 | 951.99 | 227.31 | 220.99 | 0.42 | 0.30 | 0.90 | 0.48 |
| 649.00 | 461.00 | 95.00 | 93.00 | 456.35 | 94.01 | 91.92 | 1.01 | 1.04 | 1.17 | 1.04 |
| 327.00 | 240.00 | 45.00 | 42.00 | 239.42 | 43.94 | 41.96 | 0.24 | 2.35 | 0.10 | 0.51 |
| **Dancer** | 6514.00 | 4571.00 | 1005.00 | 938.00 | 4571.22 | 1000.92 | 933.07 | 0.00 | 0.41 | 0.53 | 0.13 |
| 2741.00 | 1996.00 | 387.00 | 358.00 | 1981.67 | 386.42 | 355.75 | 0.72 | 0.15 | 0.63 | 0.63 |
| 1258.00 | 916.00 | 178.00 | 164.00 | 909.13 | 176.83 | 162.69 | 0.75 | 0.66 | 0.80 | 0.74 |
| 593.00 | 429.00 | 85.00 | 79.00 | 426.31 | 85.01 | 79.43 | 0.63 | 0.01 | 0.55 | 0.38 |
|  |  |  |  |  |  |  | **Avg** | 0.94 | 1.13 | 1.09 | **0.89** |

Table 1 shows the percentage error of the proposed rate control given the target bitrates produced by the existing 3D-HTM. The average percentage error across all views and sequences is about 0.89%, meaning that the accuracy is better than 99%.

However, rate control comes with some cost in terms of PSNR degradation. To show the PSNR degradation from the proposed rate control, we computed the ΔPSNR as

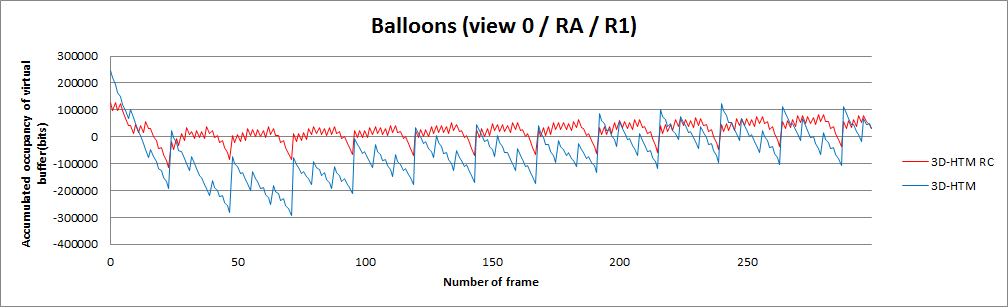
 (11)

Table 2. Performance of view-wise quality of 3D-HTM rate control and the existing 3D-HTM

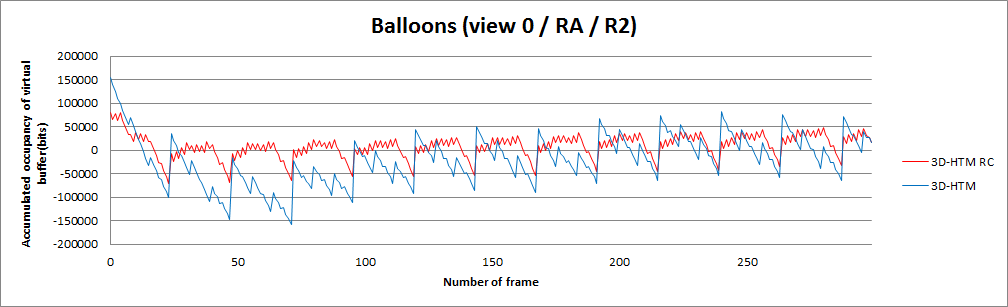
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sequence** | **PSNR V0** | **PSNR V1** | **PSNR V2** | **view0** | **view1** | **view2** | **Δ PSNR V0** | **Δ PSNR V1** | **Δ PSNR V2** | **average Δ PSNR** |
| **Balloons** | 41.97 | 40.34 | 40.10 | 43.10 | 42.29 | 41.95 | 1.14 | 1.95 | 1.85 | 1.65 |
| 39.28 | 37.43 | 37.23 | 41.07 | 39.99 | 39.73 | 1.80 | 2.56 | 2.50 | 2.29 |
| 36.14 | 34.43 | 34.97 | 38.51 | 37.41 | 37.18 | 2.37 | 2.98 | 2.21 | 2.52 |
| 32.97 | 32.31 | 32.20 | 35.68 | 34.67 | 34.47 | 2.72 | 2.36 | 2.27 | 2.45 |
| **Kendo** | 42.44 | 40.88 | 40.52 | 43.62 | 42.75 | 42.37 | 1.17 | 1.87 | 1.85 | 1.63 |
| 39.73 | 38.32 | 38.03 | 41.46 | 40.43 | 40.17 | 1.73 | 2.11 | 2.14 | 2.00 |
| 37.02 | 35.92 | 36.01 | 38.99 | 38.01 | 37.81 | 1.97 | 2.09 | 1.80 | 1.95 |
| 34.25 | 33.71 | 33.53 | 36.38 | 35.48 | 35.31 | 2.13 | 1.77 | 1.77 | 1.89 |
| **Newspaper** | 39.70 | 37.58 | 38.06 | 41.35 | 40.30 | 39.89 | 1.66 | 2.72 | 1.83 | 2.07 |
| 36.87 | 34.81 | 34.40 | 39.00 | 37.98 | 37.54 | 2.13 | 3.16 | 3.14 | 2.81 |
| 34.05 | 32.28 | 32.76 | 36.46 | 35.57 | 35.10 | 2.42 | 3.30 | 2.33 | 2.68 |
| 31.22 | 30.34 | 30.18 | 33.91 | 33.17 | 32.60 | 2.69 | 2.83 | 2.42 | 2.65 |
| **Gtfly** | 38.84 | 38.04 | 38.06 | 40.19 | 39.68 | 39.66 | 1.35 | 1.64 | 1.60 | 1.53 |
| 36.01 | 35.35 | 35.40 | 37.78 | 37.43 | 37.40 | 1.77 | 2.08 | 2.00 | 1.95 |
| 33.61 | 33.22 | 33.24 | 35.60 | 35.36 | 35.32 | 1.99 | 2.15 | 2.08 | 2.07 |
| 31.44 | 31.38 | 31.35 | 33.46 | 33.29 | 33.24 | 2.02 | 1.92 | 1.89 | 1.94 |
| **PoznanHall2** | 40.30 | 39.39 | 39.29 | 42.13 | 41.86 | 41.71 | 1.82 | 2.48 | 2.42 | 2.24 |
| 37.16 | 36.71 | 36.86 | 41.15 | 40.86 | 40.77 | 3.99 | 4.16 | 3.91 | 4.02 |
| 34.93 | 35.97 | 36.12 | 39.76 | 39.45 | 39.41 | 4.83 | 3.48 | 3.29 | 3.87 |
| 34.46 | 35.25 | 35.19 | 38.05 | 37.71 | 37.73 | 3.59 | 2.47 | 2.53 | 2.86 |
| **PoznanStreet** | 38.95 | 37.44 | 36.60 | 39.86 | 38.99 | 38.92 | 0.91 | 1.55 | 2.32 | 1.59 |
| 36.23 | 35.24 | 34.22 | 37.91 | 37.14 | 36.95 | 1.68 | 1.90 | 2.73 | 2.10 |
| 33.68 | 33.08 | 32.17 | 35.90 | 35.27 | 35.00 | 2.22 | 2.19 | 2.82 | 2.41 |
| 31.19 | 31.17 | 30.21 | 33.78 | 33.31 | 33.00 | 2.59 | 2.15 | 2.80 | 2.51 |
| **Dancer** | 38.18 | 37.16 | 37.36 | 38.76 | 37.97 | 38.14 | 0.57 | 0.81 | 0.78 | 0.72 |
| 35.22 | 34.43 | 34.43 | 35.89 | 35.35 | 35.55 | 0.67 | 0.91 | 1.12 | 0.90 |
| 32.70 | 31.88 | 32.14 | 33.38 | 32.95 | 33.16 | 0.68 | 1.07 | 1.01 | 0.92 |
| 30.50 | 30.08 | 30.25 | 31.10 | 30.76 | 30.93 | 0.60 | 0.67 | 0.68 | 0.65 |
|  |  |  |  |  |  | **Avg** | 1.97 | 2.19 | 2.15 | 2.10 |

Table 2 shows the average PSNR degradation is about 2.1dB.

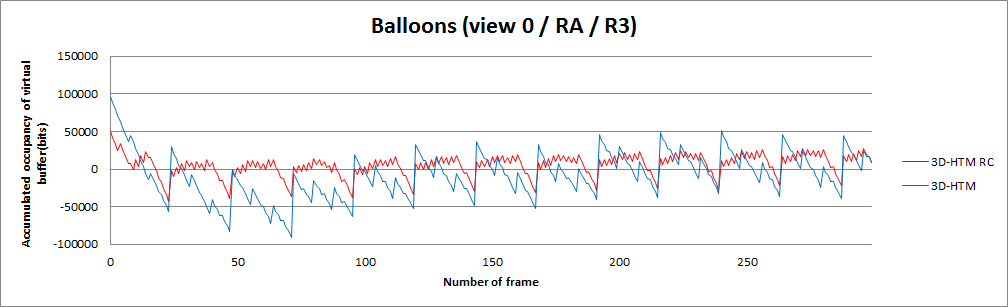
Figures 3 to 9 show the accumulated occupancy of virtual buffer for base view of each sequence.



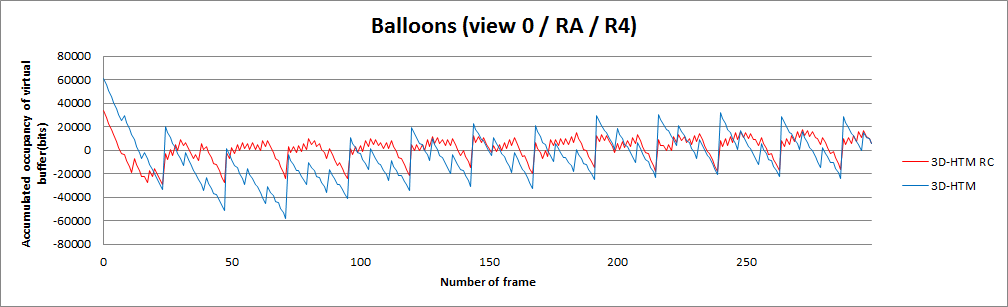
(a)



(b)

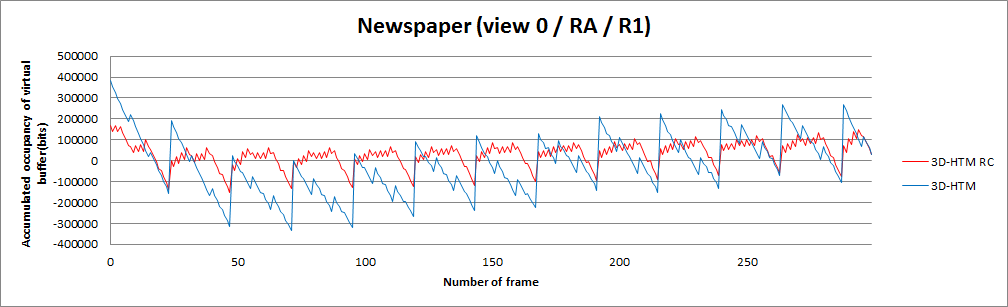


(c)

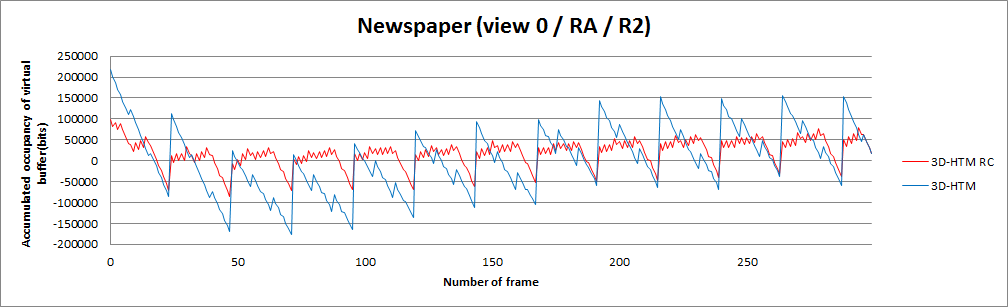


(d)

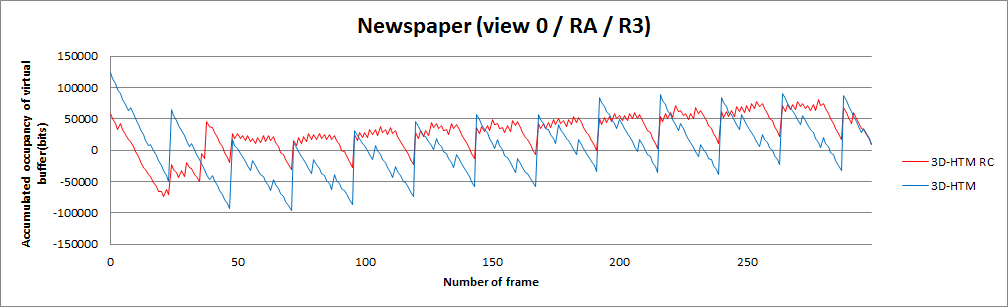
Figure 3. The accumulated occupancy of virtual buffer for ‘Balloons’ sequence. (a)~(d) : rate1 ~ rate4



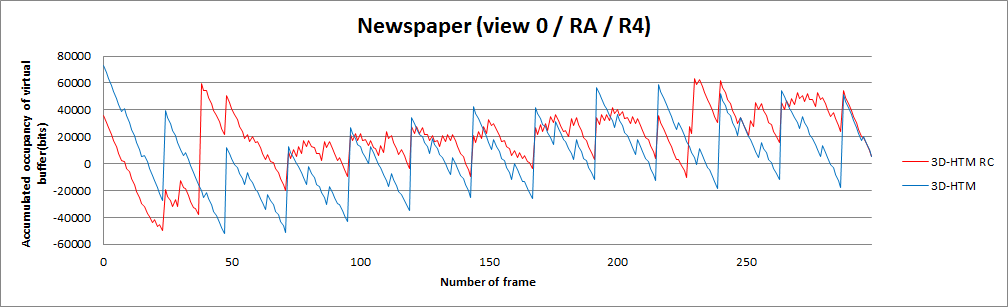
(a)



(b)

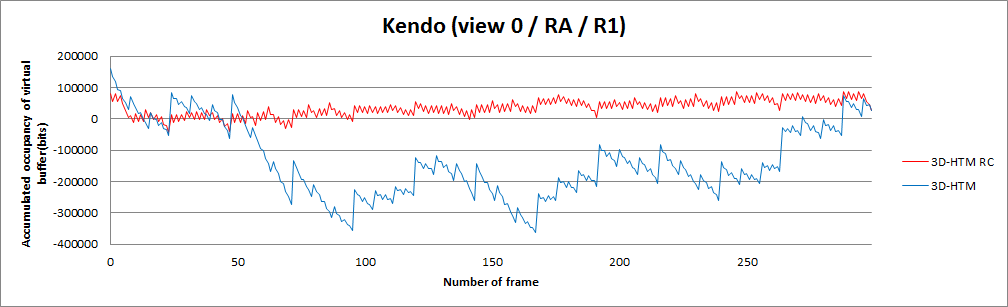


(c)

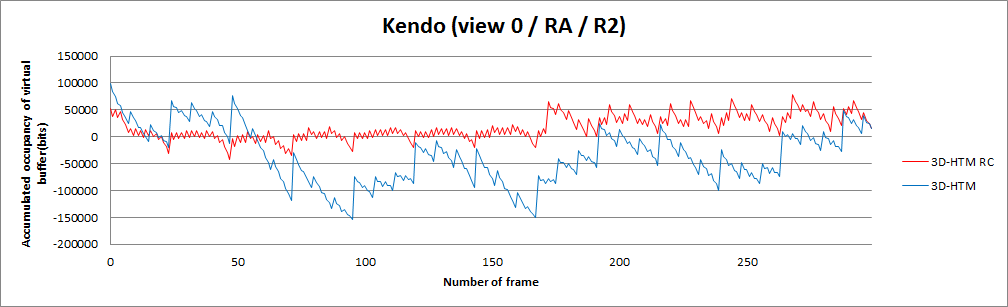


(d)

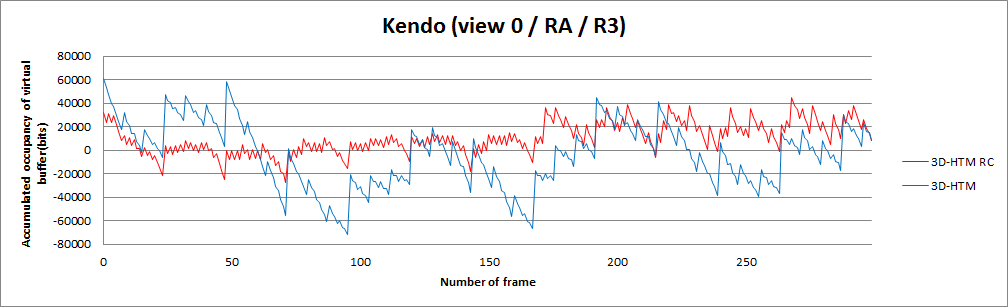
Figure 4. The accumulated occupancy of virtual buffer for ‘Newspaper’ sequence. (a)~(d) : rate1 ~ rate4



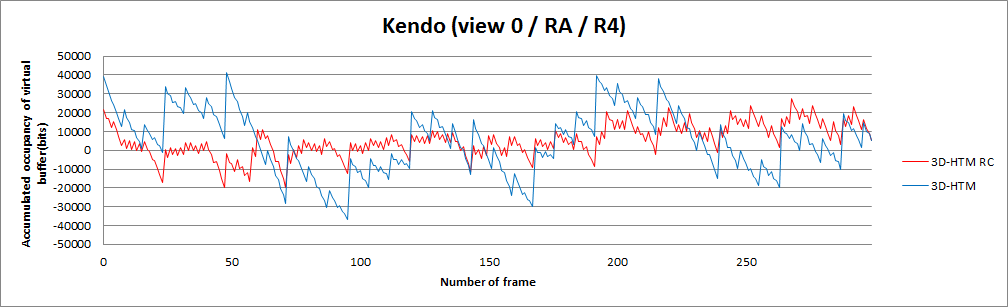
(a)



(b)

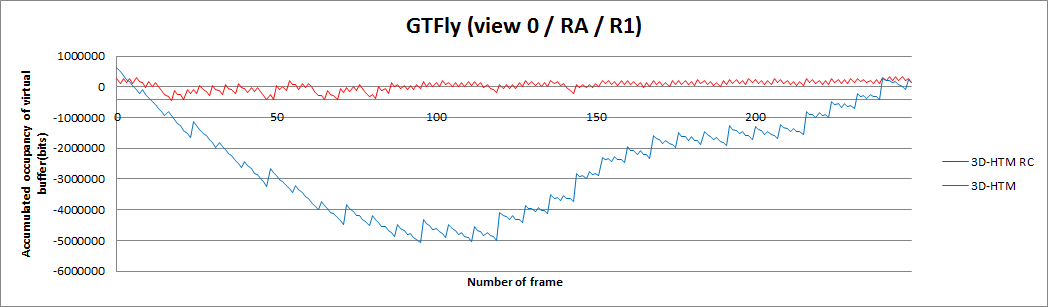


(c)

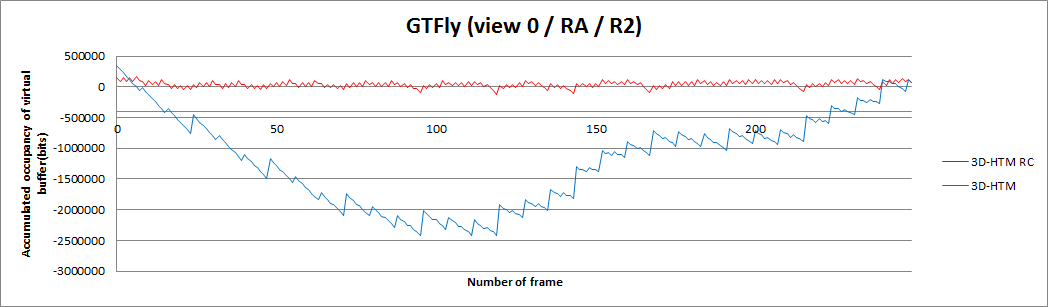


(d)

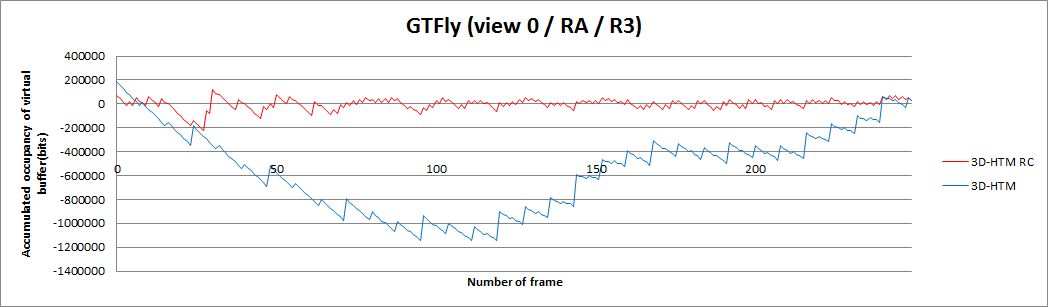
Figure 5. The accumulated occupancy of virtual buffer for ‘Kendo’ sequence. (a)~(d) : rate1 ~ rate4



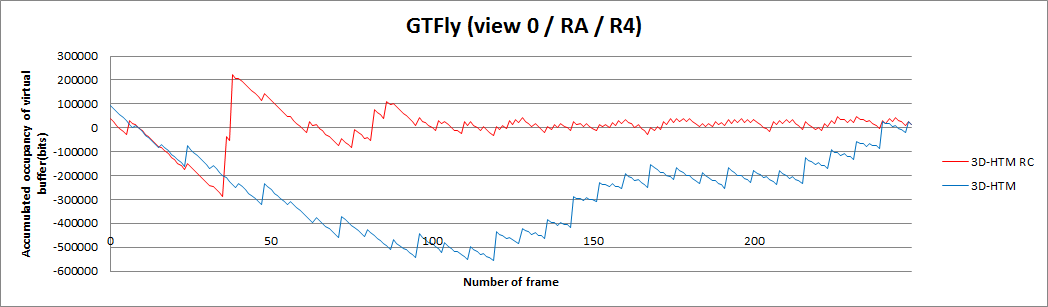
(a)



(b)

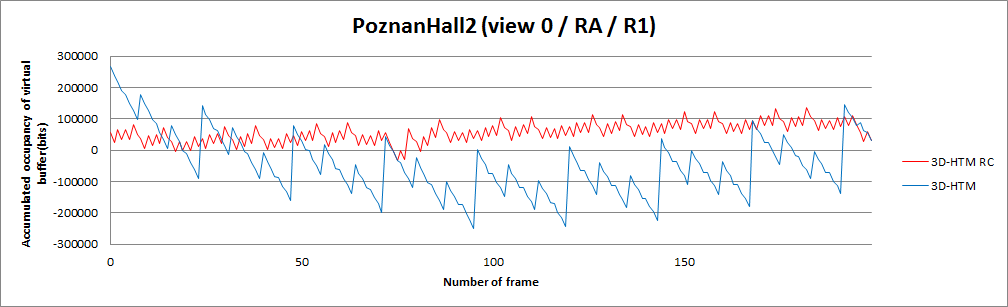


(c)

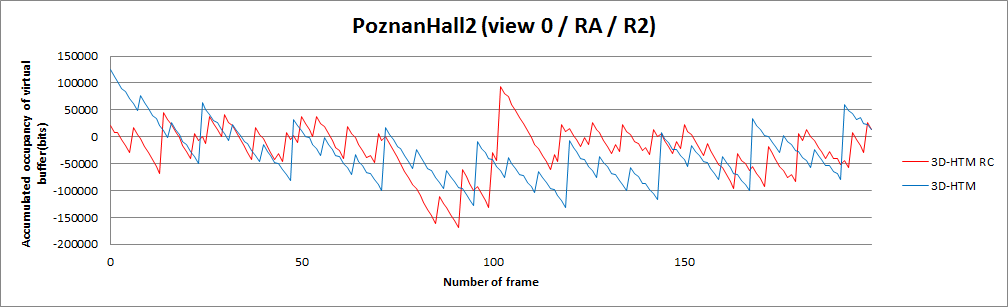


(d)

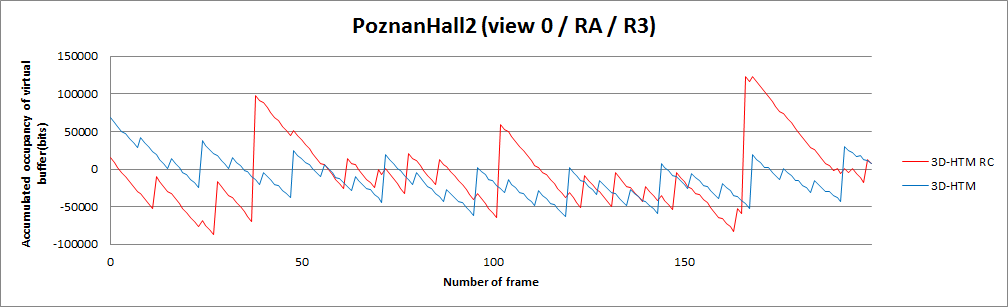
Figure 6. The accumulated occupancy of virtual buffer for ‘GTFly’ sequence. (a)~(d) : rate1 ~ rate4



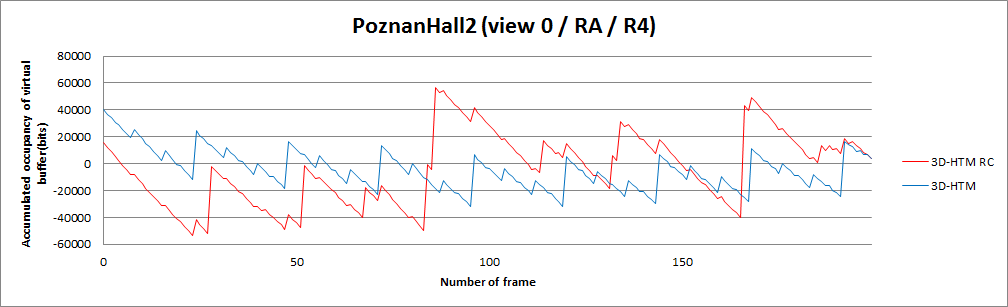
(a)



(b)

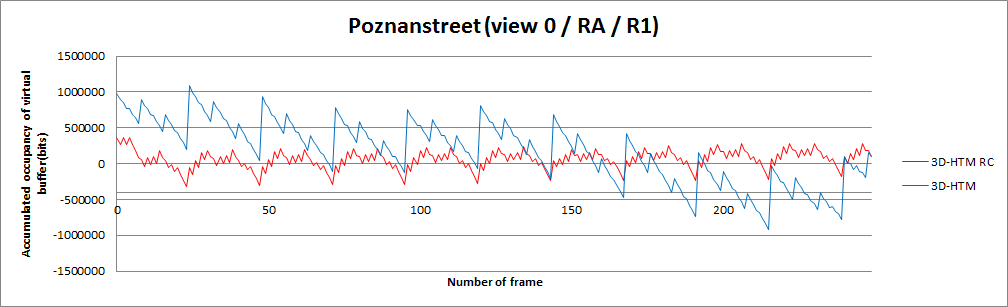


(c)

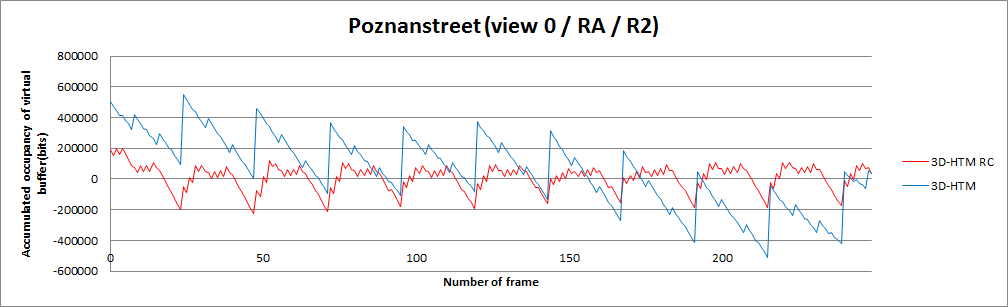


(d)

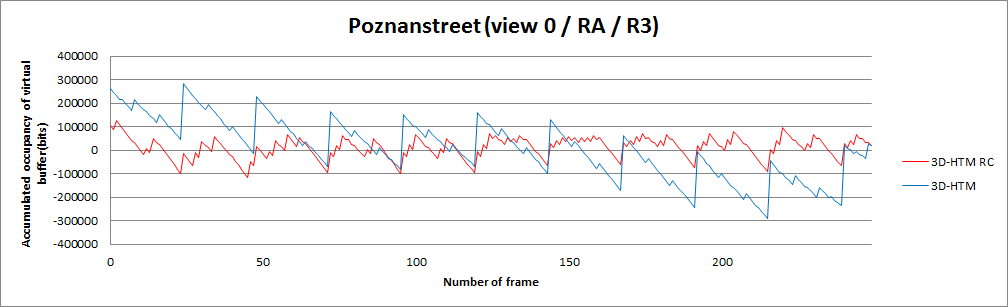
Figure 7. The accumulated occupancy of virtual buffer for ‘PoznanHall2’ sequence. (a)~(d) : rate1 ~ rate4



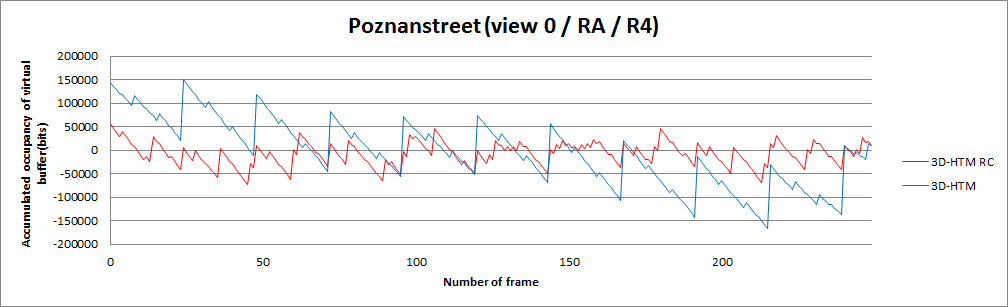
(a)



(b)

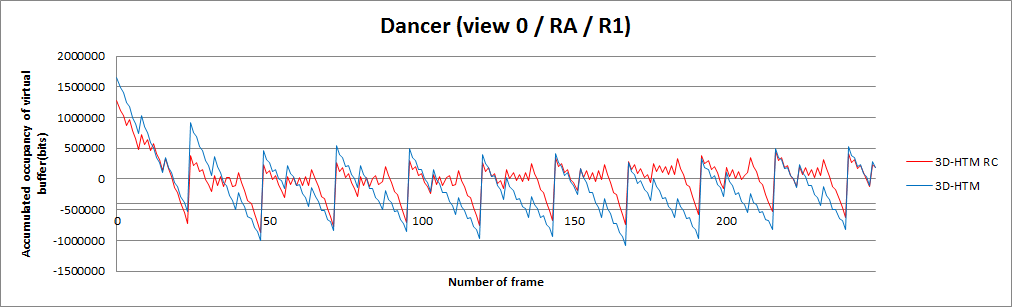


(c)

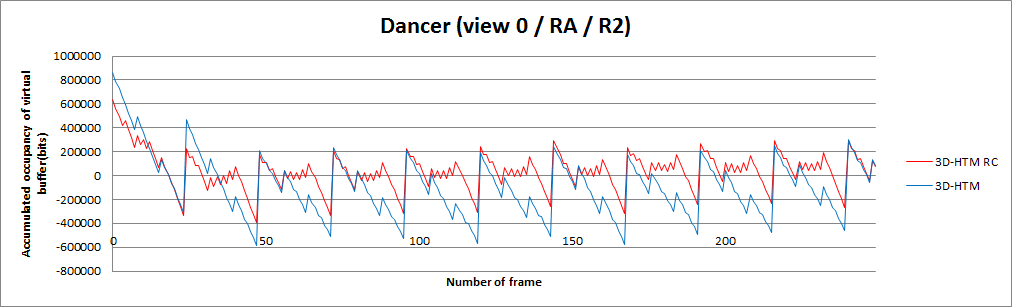


(d)

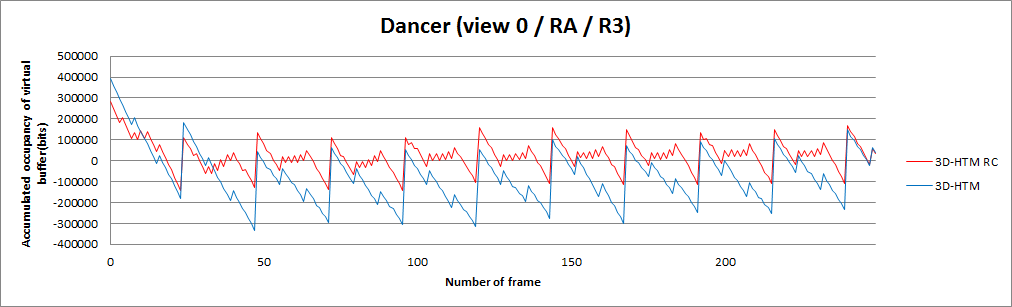
Figure 8. The accumulated occupancy of virtual buffer for ‘PoznanStreet’ sequence. (a)~(d) : rate1 ~ rate4



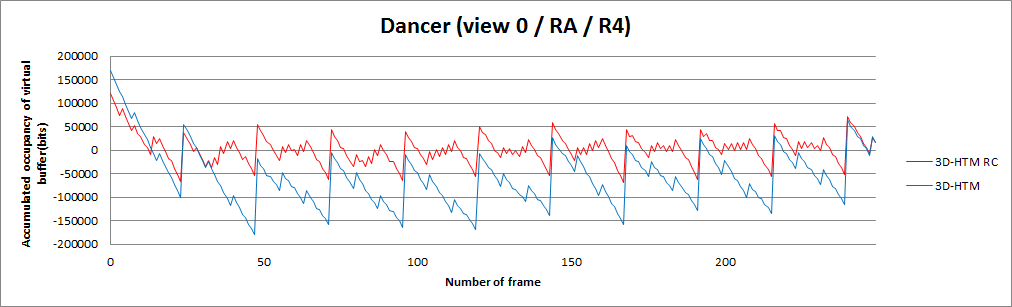
(a)



(b)



(c)



(d)

Figure 9. The accumulated occupancy of virtual buffer for ‘UndoDancer’ sequence. (a)~(d) : rate1 ~ rate4

According to the above figures, general fluctuation of the accumulated occupancy of the virtual buffer for the base view is reasonable, and the deviation from the target occupancy is much smaller than without rate control. However, in some cases in certain sequences, there are occasional sudden spikes in the number of generated bits. We are currently examining the reasons behind these spikes. The results for the extended views are presented in the additional Excel file included with this document.

# Conclusion

We proposed a rate control scheme for 3D multi-view video coding. The proposed rate control is based on the existing rate control in HEVC. Specifically, the HEVC scheme is applied to the base view directly. In addition, we proposed initial setting of QP for the first frame of extended view, and depth map-based inter-view MAD prediction for extended view. By using this rate control for 3D multi-view video coding, the average bitrate accuracy was over 99%, with about 2.1dB average PSNR degradation. We are currently examining the reasons behind occasional sudden spikes in the number of generated bits in some cases. These problems should be fixed soon, and we plan to update the scheme in the near future.

# Reference

1. 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
2. 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,” JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-I0094 (m24333), Geneva, Switzerland, Apr.-May. 2012.

# Patent rights declaration(s)

**KWU and SFU do not have any current or pending patent rights relating to the technology described in this contribution.**