The influence of a lossy compression on the quality of estimated depth maps

Adrian Dziembowski, Marek Domański, Adam Grzelka, Dawid Mieloch, Jakub Stankowski, Krzysztof Wegner
Chair of Multimedia Telecommunications and Microelectronics,
Poznan University of Technology, Poland
{adziembowski, agrzelka, dmieloch, jstankowski, kwegner}@multimedia.edu.pl

Abstract – **The paper considers the influence of a lossy compression of multiview video on the quality of estimated depth maps. We have encoded a set of views with commonly known compression techniques used in compact IP cameras: HEVC (MPEG H part 2 / H.265), AVC (MPEG 4 part 10 / H.264) and MPEG 2 part 2 / H.262. Basing on the decoded data we have estimated depth maps and rendered a set of the intermediate virtual views. A drop of the quality was studied by comparison of virtual views rendered out of the depth maps estimated from an uncompressed video. The obtained results show that with a reasonable compression rate, lower quality of virtual views is acceptable for real implementable FTV systems.**

Keywords – **Free navigation; Free-viewpoint television; Depth map estimation; Virtual view synthesis; Video compression**

I. INTRODUCTION

Currently we observe rapid development of various kinds of multiview systems. In such systems a scene is registered from different directions with a set of cameras. This requires not only to capture multiple views, but also to transmit all of the captured views. An example of special kinds of multiview systems are free viewpoint television systems [1], where an user has an ability to observe a scene from infinite set of continuously placed viewpoints. Capturing and transmitting of a large number of views is unreasonable, therefore, depth maps which are 3D representation of a scene are estimated and transmitted along with the sparse subset of captured views. Any arbitrary selected viewpoint can be recreated with the use of the depth image base rendering (DIBR) technique [2,3].

II. MOTIVATION

A typical way of delivering Free-Viewpoint Television services to a viewer is usage of a scheme called MVD (Multiview Video plus Depth [4]). In that approach, a viewer receives two main parts of the information: multiple views and corresponding depth maps. In an MVD bitstream, the proportion between the multiview texture information and the depth information is ca. 85/15 [5,6,7]. Thereby, if we omit the depth information, the bitstream can be reduced by 15%.

Unfortunately, two problems occur from this approach. First, the depth information has to be generated in a decoder, what implies the necessity of real-time algorithms application. The real-time depth estimation algorithms already exist (e.g. [8]), but the quality of the depth they estimate is worse when compared to more complex algorithms [9]. However, one can expect that constantly increasing performance of the hardware will significantly reduce time of the depth map estimation.

The second problem of sending only texture to a user is the calculation of the depth based on the compressed texture. In a typical approach, the depth maps are estimated before any compression in order to avoid the influence of compression artifacts on the depth estimation. However, there is no practical possibility to send the uncompressed multiview video to the final user. Thereby, if depth maps are to be created in the decoder, the depth estimation should be based on lossy compressed texture.

The question we try to deal with in the paper is whether the multiview sequence could be initially compressed with no significant loss of the quality of estimated depth, thus, if a lossy compression has considerable influence on the quality of synthesized virtual views. Moreover, we try to test what compression ratio of the input sequences maintains proper quality of the depth map estimation.

The usage of lossy-compressed views in the depth map estimation has one significant advantage – it provides a lower cost of the FTV system. If we assume that the captured images can be compressed, we can use simpler consumer cameras (opposite to systems without video compression, e.g. [10]), cheaper cabling (lower throughput), data capturing and gathering.

III. EXPERIMENTS

In order to address the questions mentioned in the previous section, we had to choose a set of test sequences. We decided to use multiview sequences with more than 30 different views. For each sequence we used four particular views for the depth estimation and further for the virtual view synthesis. In positions of remaining 27 views (Fig. 1) we synthesized virtual views and measured the PSNR value for each position.

In the experiments, 7 multiview sequences were used: 5 synthetic and 2 captured by the multicamera FTV system [1]. The testing set with the information about source and used views for each sequence is listed in Table I.

For each sequence we considered evenly distributed real cameras, so the number of reference views between each pair of real views is equal within one sequence.

The presented work has been funded by the Polish Ministry of Science and Higher Education for the status activity consisting of research and development and associated tasks supporting development of young scientists and doctoral students.

Figure 1. Used camera positions

For arc version of BBB Butterfly the real views are views 6, 19, 32 and 45, what corresponds to the MPEG's recommendation for free navigation in that sequence [13]. The choice of views for other sequences was taken to ensure similar maximal disparities in the images.

TABLE I. LIST OF TEST SEQUENCES

Sequence name	Sequence source	Used views
BBB[a] Butterfly Arc	Holografika [11]	6, 7, …, 45
BBB Butterfly Linear	Holografika [11]	6, 7, …, 45
Dog	Nagoya Univ. [12]	0, 2, …, 60
BBB Flowers Linear	Holografika [11]	6, 7, …, 45
Pantomime	Nagoya Univ. [12]	0, 2, …, 60
BBB Rabbit Arc	Holografika [11]	6, 7, …, 45
BBB Rabbit Linear	Holografika [11]	6, 7, …, 45

a. Big Buck Bunny

To measure the impact of using different encoders, we decided to perform all the experiments independently for the simulcast MPEG-2, AVC and HEVC. For each encoder, the GOP size was 13. The frame arrangement was: I BB P BB P BB P BB P (except for MPEG-2, where B-frames were replaced by P-frames). We used publicly available optimized encoders: for MPEG2 – mpeg2video encoder from FFmpeg package [14], for AVC the x264 encoder [15] and for HEVC the x265 encoder [16]. All encoders have been configured in "fast" operation mode in order to simulate the real-world low-power embedded encoders.

All the experiments were performed for 7 different QP values (20, 25, 30, 35, 40, 45 and 50) and for uncompressed data to ensure the possibility of comparing the results. For each test sequence the same number of 100 frames was encoded and used for the depth estimation and the view synthesis.

The quality of synthesized views was measured by estimating a PSNR value for pairs made of a synthesized virtual view and an uncompressed reference view. That approach provides the measurement of quality of the whole image processing – both the depth estimation and the view synthesis.

The overview of experiments is shown in Fig. 2. At first, 4 real views are lossy encoded and decoded. Once compressed, views are used for the depth map estimation. Then, the virtual views in positions of all the reference cameras are synthesized – each of them using two neighboring (compressed) views with corresponding depth maps. Finally, the mean PSNR for all virtual views is calculated.

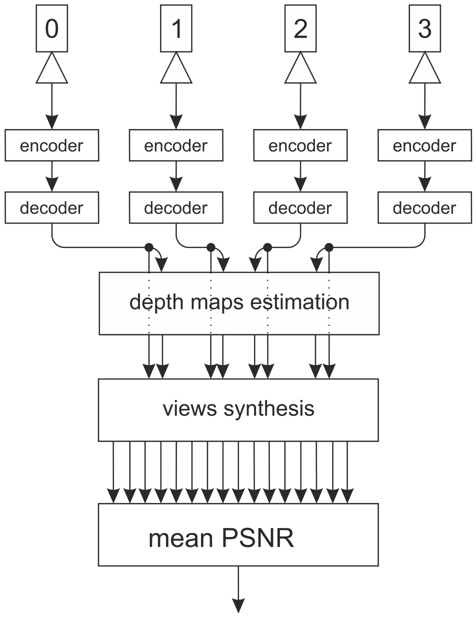

Figure 2. Overview of the experiment

IV. EXPERIMENTAL RESULTS

In Tables II – IV results obtained for 3 analyzed encoders are shown. Each table contains the mean PSNR value for every considered QP for each test sequence. For MPEG2, instead of QP from 20 to 50 we chose the values of quantization parameter to ensure similar quality distribution for all encoders.

TABLE II. MEAN PSNR VALUES FOR EACH TEST SEQUENCE (MPEG-2 ENCODING)

Sequence name	PSNR for original	PSNR [dB] for different Q						
		2	*4*	*6*	*10*	*18*	*30*	*51*
BBB Butterfly Arc	36.9	35.0	34.8	34.5	34.2	33.8	33.3	33.2
BBB Butterfly Linear	35.7	34.7	34.5	34.3	34.1	33.8	33.3	33.2
Dog	30.0	29.9	29.8	29.7	29.6	29.5	29.4	29.3
BBB Flowers Linear	27.5	26.8	26.6	26.5	26.3	25.9	25.6	25.4
Pantomime	30.3	29.9	30.0	29.9	29.8	29.6	29.4	29.4
BBB Rabbit Arc	31.2	30.9	30.5	30.2	29.8	29.0	28.3	28.2
BBB Rabbit Linear	29.8	29.6	29.4	29.3	28.9	28.3	27.7	27.6

TABLE III. MEAN PSNR VALUES FOR EACH TEST SEQUENCE (AVC ENCODING)

Sequence name	PSNR for original	PSNR [dB] for different QP						
		20	*25*	*30*	*35*	*40*	*45*	*50*
BBB Butterfly Arc	36.9	36.3	36.1	35.7	35.1	34.2	33.0	30.6
BBB Butterfly Linear	35.7	35.4	35.2	34.9	34.4	33.7	32.5	30.4
Dog	30.0	29.5	29.5	29.4	29.2	28.8	28.1	26.9
BBB Flowers Linear	27.5	26.8	26.6	26.5	26.2	25.8	25.3	24.6
Pantomime	30.3	30.0	30.0	29.9	29.9	29.5	28.7	27.6
BBB Rabbit Arc	31.2	30.8	30.6	30.2	29.7	29.0	27.8	26.4
BBB Rabbit Linear	29.8	29.6	29.4	29.2	28.8	28.2	27.3	26.2

TABLE IV. MEAN PSNR VALUES FOR EACH TEST SEQUENCE (HEVC ENCODING)

Sequence name	**PSNR for original**	**PSNR [dB] for different QP**						
		20	***25***	***30***	***35***	***40***	***45***	***50***
BBB Butterfly Arc	36.9	35.8	35.4	35.1	34.7	33.9	33.1	30.3
BBB Butterfly Linear	35.7	35.4	35.1	34.8	33.9	33.2	31.4	30.1
Dog	30.0	29.5	29.5	29.4	29.2	28.8	28.1	26.9
BBB Flowers Linear	27.5	26.7	26.2	26.2	25.9	25.5	25.1	24.4
Pantomime	30.3	30.0	29.9	29.9	29.7	29.3	28.9	28.1
BBB Rabbit Arc	31.2	31.2	31.1	30.8	30.4	29.5	28.1	26.2
BBB Rabbit Linear	29.8	29.8	29.7	29.5	29.2	28.8	27.5	25.9

Fig. 3 shows PSNR values from Tables II – IV, averaged for all the sequences. In general, the presented quality distribution over different QP values is typical – for lower QP the quality is high and while increasing QP value, the quality decreases.

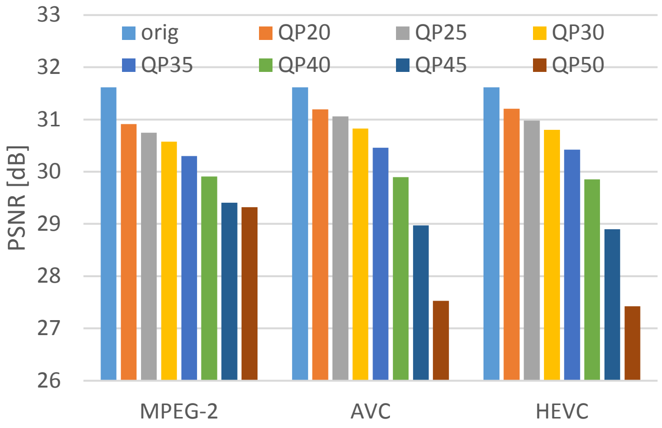

Figure 3. Mean PSNR value for different QP values for 3 used encoders

In Fig. 3 an interesting fact is visible, that between uncompressed video quality and the quality of synthesis using video compressed using two lowest QP values is only very slight. The difference between PSNR for uncompressed video and QP25 is only 0.5dB. That 0.5dB is the cost of decreasing the bitrate 1000 times: from over 1000Mbps to 1000kbps (Table V).

TABLE V. AVERAGE BITRATES FOR USED SEQUENCES FOR EACH QP

	Bitrates [kbit/s] for uncompressed		
	1250000		
QP	**Bitrates [kbit/s] for different encoders**		
	MPEG-2	***AVC***	***HEVC***
20	4317	2265	1898
25	2024	1297	1057
30	1370	754	599
35	889	449	341
40	592	273	194
45	452	170	106
50	435	114	57

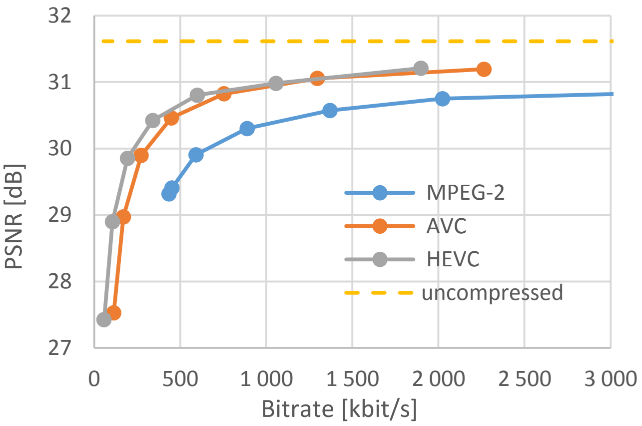

Figure 4. RD curves for MPEG-2, AVC and HEVC

In Fig. 4 the RD curves for all analyzed encoders are presented. The dashed horizontal line represents the quality of views synthesized using uncompressed real views.

As it was mentioned, all the experiments were performed for frame arrangement I BB P BB P BB P BB P. Thereby, it is necessary to describe the impact of usage of different types of frames on the quality of synthesized views.

For all used encoders, the results were similar, so we focused on one of them – the state-of-art in video compression – HEVC. In the Fig. 5 the mean PSNR (averaged over all sequences) for I-, P- and B-frames was presented. To preserve chart clarity, only three particular QP values were chosen to be shown.

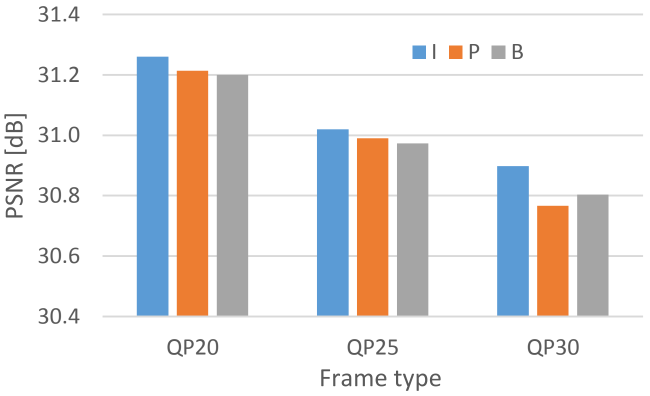

Figure 5. PSNR differences between I-frames, P-frames and B-frames (HEVC)

For all QP values, the leftmost bar (representing I-frames) is the highest. The quality of inter-frames (both the P- and B-frames) is slightly worse.

The information about the negative influence of inter-frames on the quality of synthesized views for every considered QP is presented in Fig. 6.

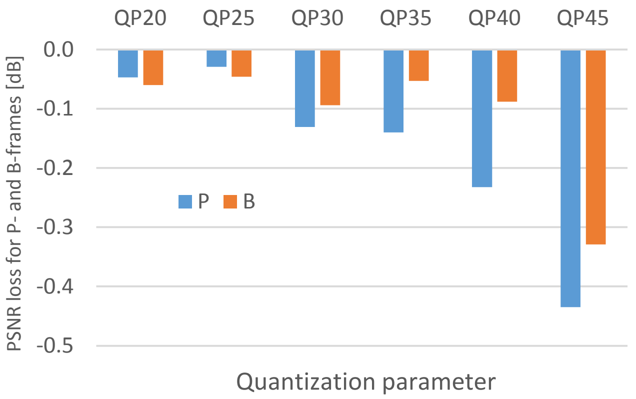

Figure 6. PSNR values for P- and B-frames compared to PSNR for I-frames (HEVC)

Presented data shows clearly, that the quality of the virtual view synthesis using intra-frames is (on average) higher, than the quality for inter-frames. The difference between P- and B-frames is negligibly small. For lower QP values, the PSNR loss for inter-frames is very slight (only 0.04dB for QP = 20 and 25, 0.07dB for QP30). For higher QP, the difference between intra- and inter-frames starts to be noticable.

In Fig. 7 the RD curves for HEVC encoder in All-Intra mode and the mode mentioned in section III are compared.

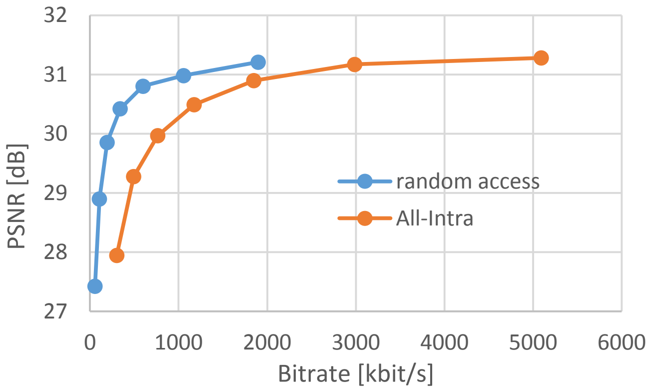

Figure 7. RD curves for All-Intra and random access mode (HEVC)

V. CONCLUSIONS

In the paper, the influence of lossy compression of input views on the estimated depth maps quality was tested. Performed experiments have shown that low compression of video does not reduce the quality of synthesized virtual views significantly. For all tested encoders (MPEG2, AVC and HEVC), the difference between a quality of virtual view synthesized using uncompressed and compressed video was not greater than 1 dB even for QP30. Moreover, for lower QP values the difference was smaller, e.g. for QP20 was around 0.5 dB (both for AVC and HEVC).

The quality of virtual view synthesis was the best for the HEVC encoder. Nevertheless, for AVC, the most popular encoder used in compact IP cameras, the quality was comparable.

The results indicate that when the state-of-the-art depth estimation and the virtual view synthesis algorithms are used for multiview video processing, a recorded sequence can be initially compressed with no significant loss of the quality of estimated depth, even for bitstream reduced 1000 times. Therefore, free-viewpoint television systems cameras do not have to offer an uncompressed video. It can significantly reduce the cost and the complexity of the required network infrastructure and cameras themselves.

REFERENCES

[1] M. Tanimoto, M.P. Tehrani, T. Fujii, T. Yendo, “FTV for 3-D spatial communication”, Proceeding of the IEEE, vol. 100, no. 4, pp. 905-917, 2012.

[2] M. Domański, M. Gotfryd, K. Wegner, “View synthesis for multiview video transmission”, The 2009 International Conference on Image Processing, Computer Vision, and Pattern Recognition, pp. 1-4 Las Vegas, USA, 13-16 July 2009.

[3] K. Wegner, O. Stankiewicz, M. Domański, “Fast view synthesis using platelet based depth representation”, 21st International Conference on Systems, Signals and Image Processing, IWSSIP 2014, Dubrovnik, Croatia, 12-15 May 2014.

[4] K. Müller, P. Merkle, T. Wiegand, “3D Video Representation Using Depth Maps”, Proceedings of the IEEE, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[5] K. Klimaszewski, K. Wegner, M. Domański, “Video and depth bitrate allocation in multiview compression”, 21st International Conference on Systems, Signals and Image Processing, IWSSIP 2014, Dubrovnik, Croatia, 12-15 May 2014.

[6] Y. Morvan, D. Farin, P.H.N. de With, “Joint depth/texture bit-allocation for multi-view video compression,” Proc. of 26th Picture Coding Symposium, 7-9 Nov. 2007.

[7] G. Cheung, V. Velisavljević, A. Ortega, “On dependent bit allocation for multiview image coding with depth-image-based rendering,” IEEE Trans. Image Proc. vol.20, no.11, pp.3179-3194, Nov. 2011.

[8] M. Domański, J. Konieczny, M. Kurc, A. Łuczak, J. Siast, O. Stankiewicz, K. Wegner, “Fast Depth Estimation on Mobile Platforms and FPGA Devices”, 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, Lisbon, Portugal, 8-10 July 2015.

[9] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domański, “Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television”, ISO/IEC JTC1/SC29/WG11 Doc. MPEG M31518, Geneva, CH, Oct.2013.

[10] M. Domański, A. Dziembowski, D. Mieloch, A. Łuczak, O. Stankiewicz, K. Wegner, “A practical approach to acquisition and processing of free viewpoint video”, Picture Coding Symposium PCS, Cairns, pp. 10-14, 2015.

[11] P. Kovacs, “[FTV AHG] Big Buck Bunny light-field test sequences”. ISO/IEC JTC1/SC29/WG11, Doc. M35721, Geneva, 2015.

[12] M. Tanimoto, T. Fujii, N. Fukushima, „1D Parallel Test Sequences for MPEG-FTV”, ISO/IEC JTC1/SC29/WG11, Doc. M15378, Archamps, 2008.

[13] G. Lafruit, K. Wegner, M. Tanimoto, “Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation”, ISO/IEC JTC1/SC29/WG11 Doc. MPEG2015/N15348, Warsaw, Poland 2015

[14] FFmpeg encoder available at: www.ffmpeg.org

[15] x264 encoder available at: www.videolan.org/developers/x264.html

[16] x265 encoder available at: www.x265.org