Quantization of Depth in Simulcast and Multiview Coding of Stereoscopic Video plus Depth Using HEVC, VVC and MV-HEVC

Yasir Al-Obaidi
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
yasir.al-obaidi@doctorate.put.poznan.pl

Tomasz Grajek
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
tomasz.grajek@put.poznan.pl

Marek Domański
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology
Poznań, Poland
marek.domański@put.poznan.pl

***Abstract*— Virtual reality, free-viewpoint television, virtual navigation, 360º video are the areas of research and technology that need efficient compression of multiview video plus depth acquired by cameras with arbitrary positions. Astonishingly, proliferation of 3D extensions of AVC and HEVC technology is very low. Therefore, in this paper, we present a study on independent coding of views and depth maps. A simple technique is proposed to estimate quantization step for depth as a function of the quantization step for multiview video. This technique is studied in the context of multiview video plus depth acquired using cameras located around a scene. The approach is based on simple modeling of the relation between quantization parameters for depth and multiview video. The experimental results are obtained for stereoscopic video with two respective depth maps. For standard MPEG test sequences, the results demonstrate usefulness of the approach for HEVC, VVC, MV-HEVC codecs.**

Keywords— Multiview plus depth, HEVC, VVC, Quantization Optimization

I. Introduction

Virtual reality, free-viewpoint television, virtual navigation, 360º video are the areas of research and technology that need efficient compression of multiview video plus depth (MVD) [1]. For the abovementioned applications, a video is acquired with cameras placed on quite arbitrary positions that are not limited to a line. Unfortunately, the existing multiview extensions of AVC [2] and HEVC [3] video compression technologies were prepared for other applications. Owing to the fact that these extensions were adjusted to linear camera arrangement only, research clearly indicates that they are more efficient than simulcast encoding for rectified MVD data acquired by linear systems [4]. However, in the case of video sequences acquired with circular arrangement of cameras, minor gain or even no compression gain can be expected [5], [6].

In 2015, Moving Picture Experts Group (MPEG) issued a Call for Evidence (CfE) on Free-Viewpoint Television: Super-Multiview and Free Navigation, addressing systems with arbitrary positions of cameras [7].

In this paper we focus on circular camera arrangement for such supplications as: FTV, virtual navigation and virtual reality (e.g. [8], [9], [10]). In addition, we assume that the depth maps are available either from depth cameras [11], [12] or from video analysis [13], [14].

MVD data format means that we work with multiple videos and associated depth maps. MVD data may be compressed either by joint video and depth coding standards (e.g. 3D-HEVC [3], 3D-AVC [2], [15], [16]) or by independent coding of videos and depth maps. Because 3D extensions of video compression standards are hardly ever supported in practical applications, therefore, in this paper, we consider independent coding of videos and depth maps.

For the case of independent coding of multiple videos sequences and multiple depth maps sequences, generally, two approaches may be considered: simulcast and multiview. For simulcast coding, standard codecs for each video and depth sequence may be used e.g. [17]. Depth map sequence is considered as a one component (monochromatic) video. In both cases - multiview and simulcast, bitrate control is a complex issue, as only mechanisms for monoscopic video coding are well-developed. Bitrate control for MVD data is considered to be even more challenging, as we focus more on synthesized views than on real ones. In other words, in applications mentioned above, mostly virtual views synthesized from MVD data are used. Consequently, the quality of virtual views is a key factor for bitrate control mechanism [5], [6], [17], [18], [19].

Some simple relations between multiview video and multiple depth maps for AVC and HEVC have already been proposed in literature, e.g. in [18], [19], [20], [21].

In the paper, which should be understood as an extension of [20], the analysis of encoding of MVD data using standard techniques for MVD such as the state-of-the art HEVC [3] and currently being developed VVC [22], [23] is presented. Moreover, a simple technique to estimate quantization step for depth as a function of the quantization step is proposed for multiview video acquired with circular camera systems.

II. The Approach

Generally, bitrate allocation in modern video encoders is controlled by quantization steps, and these are controlled by the quantization parameters independently for video – *QP* and depth – *QD.*

Quality of synthesized view is considered to be a commonly accepted measure for assessment of performance of MVD data encoding [5], [6], [17].

The goal of the work is to find out a simple and sub-optimal formula for choosing the quantization parameter for

DOI: 10.1109/PCS48520.2019.8954539

depth maps, based on a given value of quantization parameter for videos, that ensures the highest quality of a synthesized view at the given bitrate. In other words, we would like to derive a formula as follows:

$$QD = f(QP) \quad (1)$$

For simplicity, a linear relationship between both quantization parameters will be considered, as indicated by preliminary results obtained by the authors in [20]:

$$QD = \alpha \cdot QP + \beta \quad (2)$$

In order to estimate parameters of the considered model, the optimum *QP*-*QD* pairs for a set of sequences have to be selected from all possible pairs of quantization parameters. Figure 1 shows the exemplary results of the virtual view synthesis with the use of encoded views and depth maps with *QP*-*QD* pairs (represented as points; single color is assigned to a single *QP* value). The envelope over a cloud of PSNR-bitrate points contains the optimum *QP*-*QD* pairs that form the best R-D (rate-distortion) curve (red line). The best R-D curve connects the points with the highest locations on the plot in Fig. 1.

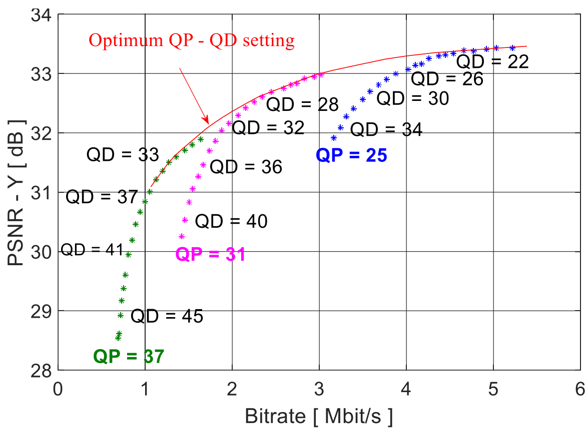

Fig. 1. R-D curve obtained from the optimum *QP*-*QD* pairs for given QP values.

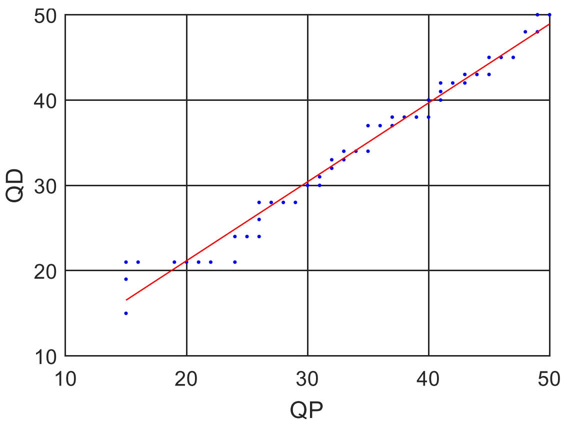

Fig. 2. The optimum *QP*-*QD* pairs (blue dots) and the corresponding linear model (red line).

Consequently, a simple formula for choosing *QD* value will be derived (Fig. 2) for a given *QP* value.

The abovementioned approach has been applied for HEVC, VVC, and MV-HEVC codecs.

III. Methodology - Details

As explained in Section II, we want to estimate parameters of the linear model (Eq. 2). For the experiments we have used eight multiview sequences recommended by the Moving Picture Experts Group (MPEG) affiliated by International Organization for Standardization (ISO). All selected sequences have high quality depth maps for all of the views.

These sequences were divided into two sets: training and verification one (Table I). Training set of sequences was used to estimate parameters of the model, whereas verifying one was used to check how good our model is.

For all sequences, we have selected three views, two of which form stereo pair and the third one, lying in between the selected views, is a view synthesis reference for quality assessment (Table I).

TABLE I. Test Sequences Used in the Experiments

Sequence name	Resolution	Encoded views	Synthesized view
Training set of MVD sequences			
Ballet [24]	1024×768	3, 5	4
Breakdancers [24]	1024×768	2, 4	3
BBB Butterfly [25]	1280×768	49, 51	50
BBB Flowers [25]	1280×768	39, 41	40
Kermit	1920×1080	5,7	6
Poznan_CarPark	1920×1088	3,5	4
Verification set of MVD sequences			
Poznan_Block2 [26]	1920×1080	2, 6	4
Poznan_Fencing [26]	1920×1080	2, 6	4

The selected pairs of views (two videos and two depth maps) were encoded with various encoders. Each time those four components were fully decoded and virtual views were rendered at the position of the selected third view. For the virtual views synthesis we have used the state-of-the-art synthesis software called View Synthesis Reference Software [27]. This software package developed by the MPEG allows high quality virtual view rendering based on two videos and two depth maps. Therefore, in our case, the only degradation of quality will result from the compression technique applied to encode videos and depth maps used in the experiments. Quality of the rendered view was measured by the luminance PSNR against the quality of the view from the real camera at the same spatial position.

For the experiment we have chosen recent versions of codecs of popular video coding standards:

- HM codec in version 16.18 [28] of the HEVC coding standard,
- HTM codec in version 16.2 [29] of the multiview extension of HEVC technology, namely MV-HEVC,
- and VTM version 3.0 [30] of the upcoming VVC technology.

Each encoder was configured according to the recommended settings, as described in appropriated MPEG common test condition documents: [31] for HEVC, [32] for MV-HEVC and [33] for VVC.

In order to simplify experiments and study only video/depth allocation, we have assumed that the quantization parameter for videos (*QP*) and the quantization parameter for depth maps (*QD*) will be kept constant for both of the views. Thus, only two quantization parameters, instead of four, are needed: one for video and one for depth map.

In order to find the optimum *QP-QD* settings according to [20], [34], all *QP-QD* pairs were tested (*QP* and *QD* values both from 25 to 50) for all test sequences for all codecs.

IV. Estimation of Model Parameters

The pairs of parameters α and β of the model (Eg. 2) have been estimated using the least squares fitting to the optimum *QP-QD* pairs. The results obtained individually for different codecs averaged over six training test sequences are collected in Table II. Because model's parameters for different codecs turn out to be very similar, last row in Table II presents also parameters of the so called "global model" – the model derived for the optimum *QP-QD* pairs for all training sequences and for all codecs.

TABLE II. Model's Parameters Derived Individually for the Different Codecs and Averaged Over all Coedcs (Global)

Codec	Parameters	
	α	β
HEVC	1.20	-11.27
VVC	1.26	-13.13
MV-HEVC	1.20	-9.41
Average (Global)	1.22	-11.13

V. Assessment of the Proposed Models

In order to verify the developed formulas, we have compared the quality of the synthesized virtual views achieved using the proposed models (Table II) with the straightforward (*QP=QD*) approach and the optimum *QP-QD* pairs for two different test sequences for verification set: *Poznan_Block2* and *Poznan_Fencing*.

Firstly, we assessed models dedicated to the given compression technology (Table III). The coding efficiency of the given technology is assessed by calculating average difference between curves for PSNR (*ΔPSNR*) and bitrate (*ΔBitrate*). These metrics are an extension of the well-known Bjøntegaard metrics (BD-PSNR and BD-Rate) [35] in order to work with more than four points.

TABLE III. *ΔPSNR* and *ΔBitrate* Metrices Calciulated for Encoding Using Straightforward (*QP=QD*) and Optimum *QP-QD* Pairs Substituted with Proposed Model for Different Coding Techniques

Codec	**(*QP=QD*) vs Proposed**		**Optimum vs Proposed**	
	ΔPSNR [dB]	***ΔBitrate [%]***	***ΔPSNR [dB]***	***ΔBitrate [%]***
Poznan_Block2				
HEVC	-0.10	15.89	0.13	-26.70
VVC	-0.08	16.54	0.14	-43.51
MV-HEVC	-0.07	8.34	0.33	-52.10
Poznan_Fencing				
HEVC	-0.04	17.81	0.09	-54.03
VVC	-0.03	14.10	0.06	-28.13
MV-HEVC	-0.02	7.74	0.10	-53.21

The conducted experiments showed that proposed formulas derived independently for different compression techniques (Table III) led to decrease of the total bitrate and improved the virtual view quality for sequences, when compared to the straightforward (*QD=QP*) approach. However, as might be expected, comparison with the optimum *QD-QP* pairs led to an increase of total bitrate and a decrease of virtual view quality. Figures from 3 to 5 present R-D (rate-distortion) curves for *Poznan_Block2* sequence for HEVC, VVC and MV-HEVC codecs, respectively.

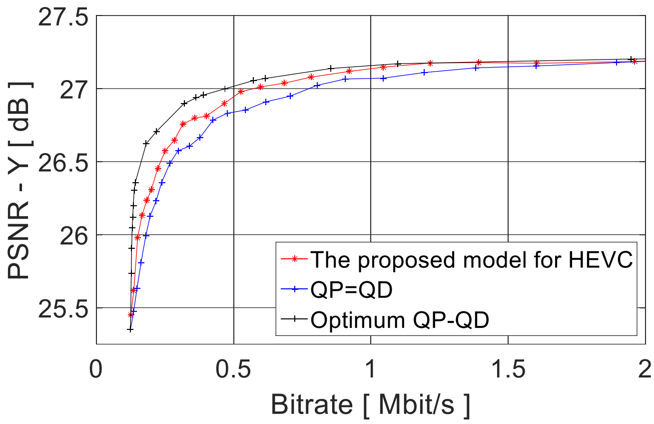

Fig. 3. R-D curves comparison between the proposed model, the straightforward approach (*QP=QD*), and the optimum *QP-QD* pairs for HEVC codec for *Poznan_Block2* sequence .

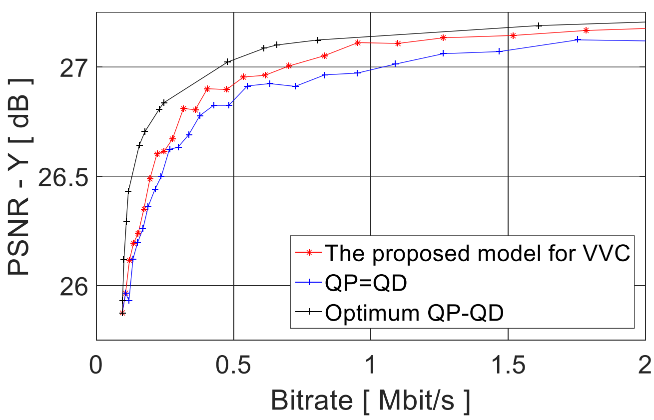

Fig. 4. R-D curves for the proposed model, the straightforward approach (*QP=QD*), and the optimum *QP-QD* pairs for VVC codec for *Poznan_Block2* sequence .

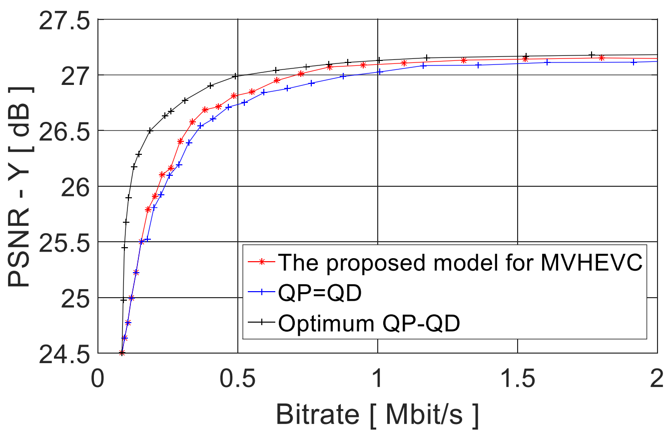

Fig. 5. R-D curves for the proposed model, the straightforward approach (*QP=QD*), and the optimum *QP-QD* pairs for MV-HEVC codec for *Poznan_Block2* sequence.

Additionally, a difference between values of quantization parameters for video and depth map sequences for models derived for individual codecs have been analyzed. Fig. 6 presents relationship between *ΔQD=QP-QD* and *QP* parameter for models dedicated for different codecs.

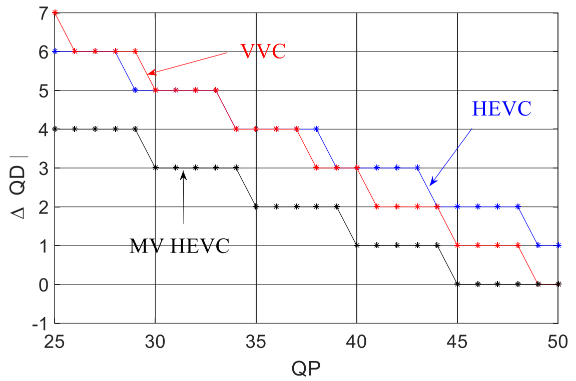

Fig. 6. Relationship between *ΔQD=QP-QD* and *QP* for models dedicated for different codecs .

Fig. 6 clearly shows that the higher the value of *QP* parameter the lower the difference between *QP* and *QD*. In other words, for high bitrates depth maps are quantized significantly weaker, i.e. with smaller quantization steps, whereas for very low bitrates (high QP values) videos and depth maps are quantized with almost the same quantization steps as for video.

Moreover, we were interested in an influence of the applied model on bitrate size for the same quality of virtual view. In order to assess this influence, the following measure was calculated:

$$\Delta B = \frac{B_{ref} - B_{model}}{B_{ref}} * 100\%, \quad (3)$$

where ΔB is the bitrate reduction, B_{ref} is a bitrate for given quality for *QP=QD* approach, and B_{model} is a bitrate for given quality for model dedicated for given codecs. Result of this analysis for *Poznan_Block2* sequence is presented in Fig. 7.

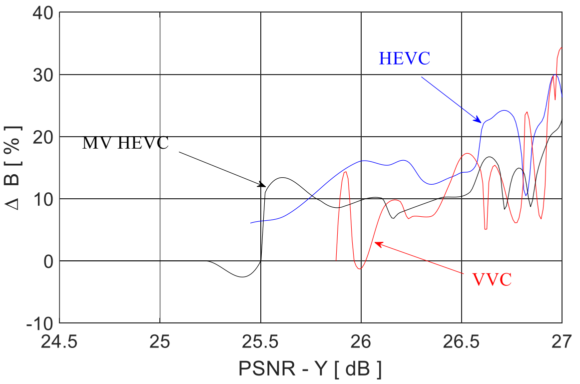

Fig. 7. Bitarte reduction resulted from appling model dedicated for given codec for *Poznan_Block2* sequence.

Finally, the global model (last row of Table II) has been assessed, in exactly the same way, as models dedicated for individual codecs were (Table IV).

In general, the conducted experiments confirm previous conclusions for models dedicated for individual codecs, but it must be taken into account that application of global model led to slightly worse results for HEVC and VVC codecs, whereas performance of MV-HEVC codec was improved.

TABLE IV. *ΔPSNR* AND *ΔBITRATE* METRICES CALCIULATED FOR ENCODING USING (*QP=QD*) AND OPTIMUM *QP-QD* PAIRS SUBSTITUTED WITH GLOBAL PROPOSED MODEL

Codec	(*QP=QD*) vs Global		Optimum vs Global	
	ΔPSNR [dB]	*ΔBitrate [%]*	*ΔPSNR [dB]*	*ΔBitrate [%]*
Poznan_Block2				
HEVC	-0.07	10.71	0.16	-31.86
VVC	-0.07	16.33	0.14	-44.70
MV-HEVC	-0.12	14.04	0.28	-42.91
Poznan_Fencing				
HEVC	-0.03	16.03	0.10	-55.96
VVC	-0.03	12.50	0.06	-30.50
MV-HEVC	-0.04	15.12	0.09	-41.91

VI. CONCLUSIONS

In the paper, we proposed simple and sub-optimal formulas to estimate quantization parameter (*QD)* for depth coding based of quantization parameter (*QP*) for video component in multiview plus depth (MVD) video compression in the case of nonlinear camera arrangement. The proposed formulas significantly simplify the process of controlling MVD compression. The proposed models assure sub-optimal bitrate split between video and depth at any requested bitrate (for all *QP* range/bitrate range). In comparison to the naive approach with equal quantization parameters (*QD=QP*), the proposed models results in 19-33% bitrate reduction across all test sequences. The results stay the same regardless compression technology used (HEVC, VVC, 3D-HEVC).

Derived formulas needs to be yet confirmed for other resolutions and a wider set of MVD test video sequences.

ACKNOWLEDGMENT

This work at Chair of Multimedia Telecommunications and Microelectronics, Poznań University of Technology was supported by the Ministry of Science and Higher Education of Republic of Poland.

REFERENCES

[1] K. Muller, P. Merkle and T. Wiegand, “3-D video representation using depth maps,” Proc. IEEE, vol. 99, pp. 643-656, April 2011.

[2] ISO/IEC 14496-10, MPEG-4 Part 10, Generic coding of audio-visual objects, Advanced Video Coding. Edition 8, 2014, also: ITU-T Rec. H.264 Edition 12.0, 2017.

[3] ISO/IEC IS 23008-2, MPEG-H Part 2, High Efficiency Video Coding, Edition 3, 2017, also: ITU-T Rec. H.265 Edition 5.0, 2018.

[4] G. Tech, Y. Chen, K. Müller, J. Ohm, A. Vetro and Y. Wang, “Overview of the multiview and 3D extensions of High Efficiency Video Coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, Jan. 2016.

[5] J. Samelak, J. Stankowski, M. Domański, “Adaptation of the 3D-HEVC coding tools to arbitrary locations of cameras,” Int. Conf. on Signals and Electronic Systems (ICSES), Kraków, Poland, Sep. 2016.

[6] J. Stankowski, Ł. Kowalski, J. Samelak, M. Domański, T. Grajek, K. Wegner, “3D-HEVC extension for circular camera arrangements,” 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV- Con 2015), Lisbon, Portugal, Jul. 2015.

[7] “Call for Evidence on Free-Viewpoint Television: Super-Multiview and Free Navigation”, ISO/IEC JTC1/SC29/WG11 MPEG2015/N15348, Poland, Warsaw, June 2015.

[8] M. Tanimoto, M. Panahpour, T. Fujii, T. Yendo, “FTV for 3 D spatial communication,” Proc. IEEE, vol. 100, pp. 905-917, Feb. 2012.

[9] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, J. Samelak, “A free-viewpoint television system for horizontal virtual navigation,” IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2182-2195, Jan. 2018.

[10] K. Wegner, O. Stankiewicz, T. Grajek and M. Domański, "Depth estimation from stereoscopic 360-degree video," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, 2018, pp. 2945-2948.

[11] K. Park, S. Kim and K. Sohn, "High-precision depth estimation using uncalibrated LiDAR and stereo fusion," in IEEE Transactions on Intelligent Transportation Systems.

[12] Y. Song and Y. Ho, "High-resolution depth map generator for 3D video applications using time-of-flight cameras," in IEEE Transactions on Consumer Electronics, vol. 63, no. 4, pp. 386-391, November 2017.

[13] Y. Qin, X. Jin, Y. Chen and Q. Dai, "Enhanced depth estimation for hand-held light field cameras," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, 2017, pp. 2032-2036.

[14] D. Mieloch, A. Grzelka, “Segmentation-based method of increasing the depth maps temporal consistency,” International Journal of Electronics and Telecommunications, Vol. 64, No. 3, Warsaw, pp. 293-298, ISSN: 0867-6747, 2018.

[15] A. Vetro, T. Wiegand, G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard", Proceedings of the IEEE, vol. 99, no. 4, pp. 626-642, April 2011.

[16] Y. Chen, M. M. Hannuksela, T. Suzuki, S. Hattori, “Overview of the MVC+D 3D video coding standard,” Journal of Visual Communication and Image Representation, vol. 25, no. 4, pp. 679-688, May 2014.

[17] B. Kroon, M. M. Hannuksela, R. Doré, M.-L. Champel, J. Boyce, “Common test conditions on 3DoF+ visual,” ISO/IEC JTC 1/SC 29/WG11, Doc. MPEG2018/N17467, Gwangju, Korea, 2018.

[18] E. Bosc, V. Jantet, M. Pressigout, L. Morin, C. Guillemot, “Bit-rate allocation for multi-view video plus depth,” 3DTV Conf.: The True Vision -Capture, Transmission and Display of 3D Video (3DTV-CON), Antalya, Turkey, 2011.

[19] K. Klimaszewski, K. Wegner, M. Domański, “Video and depth bitrate allocation in multiviewcCompression,” Int. Conf. on Systems, Signals and Image Processing (IWSSIP), Dubrovnik, Croatia, 2014.

[20] Y. Al-Obaidi, T. Grajek, O. Stankiewicz, M. Domanski, “Bitrate allocation for multiview video plus depth simulcast coding,” Int. Conf. on Systems, Signals and Image Proc. (IWSSIP), Maribor, Slovenia, 2018.

[21] K. Klimaszewski, O. Stankiewicz, K.Wegner, M. Domański, “Quantization optimization in multiview plus depth video coding,” IEEE International Conference on Image Processing ICIP 2014, Paris, France, pp. 3223-3227, 27-30 October 2014.

[22] B. Bross, et al. "Versatile Video Coding (Draft 5)" Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 doc. JVET-N1001, Geneva, CH, 2019.

[23] J. Chen ,Y. Ye, S. Kim, " Algorithm description for Versatile Video Coding and Test Model 5 (VTM 5)" Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 doc. JVET-N1002, Geneva, CH, 2019.

[24] L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, “High-quality video view interpolation using a layered representation,” ACM SIGGRAPH, pp. 600-608, 2004.

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

[26] M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, O. Stankiewicz, K. Wegner, “Multiview test video sequences for free navigation exploration obtained using pairs of cameras,” ISO/IEC JTC1/SC29/WG11, Doc. MPEG M38247, Geneva, 2016.

[27] O. Stankiewicz, K. Wegner, M. Tanimoto, M. Domański, “Enhanced View Synthesis Reference Software (VSRS) for Free-viewpoint Television,” ISO/IEC JTC 1/SC 29/WG11, Doc. M31520, Geneva, 2013.

[28] 2D HEVC reference codec available online https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.18.

[29] 3D HEVC reference codec available online https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-16.3

[30] G. Tech, Y. Chen, K. Müller, J. Ohm, A. Vetro, Y.Wang, “Overview of the multiview and 3D extensions of high efficiency video coding,” IEEE Trans. Circuits Sys. Video Techn., vol. 26, pp. 35-49, Jan. 2016.

[31] F. Bossen, “Common test conditions and software reference configurations”, Joint Collaborative Team on Video Coding (JCT-VC) of ITUT SG16 WP3 and ISO. IEC JTC1/SC29/WG11, Doc. JCTVC-J1100, Stockholm, Sweden, 2012.

[32] K. Müller, A. Vetro, “Common test conditions of 3DV core experiments,” ISO/IEC JTC1 SC29/WG11 and ITU-T SG 16 WP 3, Doc. JCT3V G1100, San José, Jan. 2014.

[33] F. Bossen, et al. "JVET common test conditions and software reference configurations for SDR video." Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11 doc. JVET-N1010, Geneva, CH, 2019.

[34] Y. Al-Obaidi, T. Grajek, “Influence of depth map fidelity on virtual view quality,” Int. Conf. on Signals and Electronic Systems (ICSES), Kraków, Poland, Sep. 2018.

[35] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves,” ITU-T SG16,Doc. VCEG-M33, USA, 2001.