ISSN 2464-4617 (print) Computer Science Research Notes WSCG2020 Proceedings
ISSN 2464-4625 (DVD) CSRN 3001

Estimation of the Optimum Depth Quantization Parameter for a Given Bitrate of Multiview Video plus Depth in 3D-HEVC Coding

Yasir Al-Obaidi
Institute of Multimedia Telecommunications
Poznań University of Technology
60-965, Poznań, Poland
yasir.al-obaidi@doctorate.put.poznan.pl

Tomasz Grajek
Institute of Multimedia Telecommunications
Poznań University of Technology
60-965, Poznań, Poland
tomasz.grajek@put.poznan.pl

ABSTRACT

One of the most popular data formats to represent natural immersive visual data is the MVD (Multiview Video plus Depth) format. The representation of a three-dimensional scene requires a huge amount of data in the form of a dense set of views accompanied by high-quality depth maps. All of this data need to be transmitted over the network to a viewer. Therefore, a question arises, how to allocate the bitrate between views and depth maps to obtain the maximal quality for a given bit budget. In the paper, a simple model for optimum bitrate allocation between color and depth data in 3D-HEVC coding is proposed. The provided model for quantization parameters allows better bitrate division between color and depth data, which leads to a significant (23-35%) bitrate reduction of the total bitrate of the multiview stream. At the same time, it preserves the same quality of synthesized virtual views in comparison to the common test condition (CTC) recommendation, which is considered as a well-established reference encoder configuration.

Keywords

Multiview video plus depth (MVD), Bit allocation, 3D-HEVC, 3D video coding.

1. INTRODUCTION

Immersive visual media are a top research topic nowadays, since they can provide, for example, real depth perception and realistic vision. Immersive visual data [Isg04] may refer to both computer-generated and natural content.

In immersive applications like free-viewpoint television (FTV) [Tan12, Sta18] additional views of a scene have to be generated based on the recorded ones. This allows a viewer to freely walk through and look around in the recorded scenes. Additional views are commonly generated via view synthesis techniques [Fen04], [Zin10].

One of the most popular data formats to represent natural immersive visual data is the MVD (Multiview Video plus Depth) format [Mul11]. For natural content, MVD means that there are some views acquired with synchronized cameras placed in quite arbitrary positions around a scene and corresponding depth maps. The depth maps may be acquired directly by specialized cameras [Par20, Son17, Zel14] or estimated algorithmically based on captured color images [Liu15, Mie18, Qin17, Weg18].

High-quality immersive views require a huge amount of data in the form of a dense set of views accompanied by high-quality depth maps. All of this data need to be transmitted over the network to a viewer's receiver in order to allow them to choose their own viewpoint of a scene.

In order to efficiently handle the multiview data, many compression technologies have been proposed. Most of them are based on the single-view video compression technology. Therefore, the multiview extension of the very popular AVC compression technology is developed as 3D-AVC [Ann14, Han13]. Similarly, the newer HEVC technique has its own multiview extension in the form of a 3D-HEVC codec [Ann18, Tec16]. Both of them have been developed almost at the same time around 2015 by the experts of the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) [JCT17].

The main improvement over the single-view compression is achieved by extending the interframe motion-compensated prediction mechanism to interview disparity-compensated prediction. Owing to the improvement, as well as to some minor tools, multiview extensions allow a 30% bitrate reduction in comparison to simply coding all of the data in parallel by single-view codecs (simulcast coding) [Tec16].

Additionally, in 3D-HEVC, several tools designed especially for depth coding have been introduced. Some of them allow the prediction of depth data based on color pictures [Mul12], while others exploit specialized prediction of depth [Mul12, Tec16]. All of those tools allow further bitrate reduction. However, as depth data are not a huge component in the compressed data stream, any improvements in depth coding do not translate directly into overall coding performance improvement.

The 3D-HEVC was developed mainly for multi-camera systems with linear camera arrangement (Figure 1a). At the time of 3D-HEVC development, i.e., around 2011-2013, such applications were considered as important, e.g., with respect to the autostereoscopic display technology. Therefore, many tools in 3D-HEVC have been developed especially for linearly arranged views, and thus do not provide any compression gain for other types of multiview content [Sam16, Sta15].

Currently, many researchers focus on multiview systems with cameras located either around a scene (Figure 1b), in an idealized case on an arc around a scene. Such a setup allows better impressiveness, wider perspective change during free navigation and more freedom in choosing the view position, while it is not so difficult to build and calibrate such a camera system.

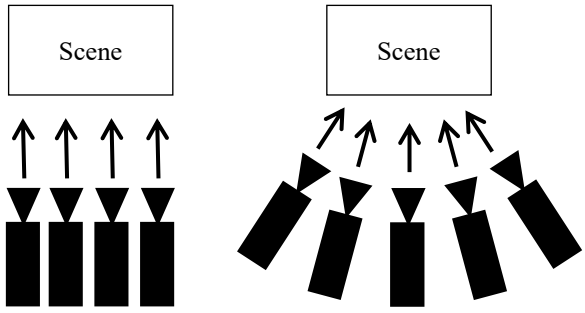

Figure 1. Camera arrangements: (a) linear (b) nonlinear

One of the most important problems with multiview compression stems from different characteristics of its components: color images (views) and depth data. Both can be compressed with different strength resulting from different overall characteristics. This problem can be stated differently: having the number of bits fixed, what number of bits should be allocated to color images (views) and to depth data in order to obtain the highest quality of a virtual view generated from decoded data at the decoder. When the compression of depth data is too strong, spatial relationships are not well-preserved and views synthesized based on the decoded data lose quality, due to spatial distortions. Similarly, too strong compression of color images makes the views generated at the decoder blurry and low in quality. Therefore, a "sweet spot", where bits spent for both components of multiview data are balanced, must exist.

2. PREVIOUS WORKS

Many solutions for the best allocation of bits between color and depth data have already been found. In most of them, the encoder is controlled via setting quantization parameter for color images (*QP*) and depth maps (*QD*) separately. Therefore, many researchers focus on providing pairs of quantization parameters for color and depth data which lead to the best quality virtual views generated out of decoded data at the particular bitrate. Many such pairs can be found in the literature, but the optimum golden standard still remains an open question. In [Bos11, Bos13], the authors have estimated the optimal bitrate ratio based on only two multiview sequences but did not supply any formula for calculating the optimal quantization parameters for video and depth data. In [Kli14a, Kli14b], the authors have proposed formulas that permit optimal bitrate allocation in the multiview video plus depth compression based on only one quantization parameter (*QP*) and estimating the second one based on previously derived formulas. Regretfully, the authors did not report performance analysis (bitrate reduction), especially they did not compare the proposed model with manually selected optimum quantization pairs for the sequences used. In [Sta13a], the authors presented a model showing the relationship between the quantization parameter of the color data (*QP*) and depth data (*QD*) in 3D-AVC coding. But again, the model has not been compared to manually selected optimum pairs of both quantization parameters. The formulas for optimum bitrate allocation in multiview video compression plus depth using simulcast HEVC, simulcast VVC, and MV-HEVC codecs were derived in [Alo18b, Alo19]. Furthermore, bitrate reduction between the proposed model and coding with the straightforward approach keeping both quantization parameters equal (*QP*=*QD*) has been reported.

3. METHODOLOGY

In the experiments, we have simulated a simple FTV system where two views and two depth maps are transmitted and a viewer always selects in-between input views the virtual view position to be synthesized.

A block diagram of the simulated system has been presented in Figure 2. Videos and depth maps have been encoded and then fully decoded using 3D-HEVC. Next, the decoded views with associated depth maps have been used to construct a requested virtual view. The generated virtual view has been compared via PSNR of luminance with the view acquired by the real camera precisely in the same position in 3D space as the generated virtual view. Finally, the measured PSNR of the generated virtual view and total bitrate of the data necessary for constructing it, have been gathered and plotted on a chart.

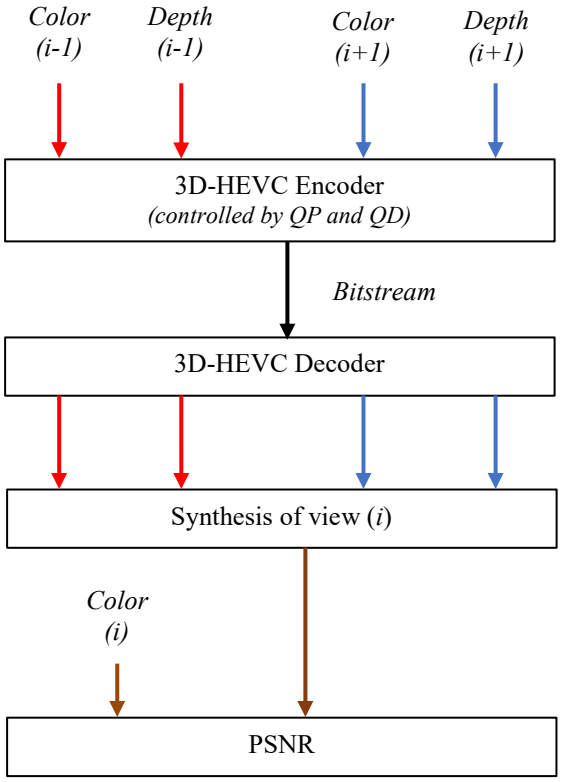

Figure 2. Structure of performed experiments

During the experiments we have examined (encode, decode, and synthesize) all of the possible pairs of quantization parameters in the range of 25 to 51. It results in 27 x 27 encodings and virtual views generated, for which we have gathered the data.

The experiments have been conducted on eight multiview sequences recommended by the Moving Picture Experts Group (MPEG) affiliated by the International Organization for Standardization (ISO). Moreover, all selected sequences have provided high-quality depth maps for all of the views.

In order to test our findings, we have divided the multiview sequences into two groups (Table 1). The first group is called a training group and was used to estimate the parameters of the proposed model. The second group is called the verification group, and was used to assess the performance of the proposed model.

As mentioned at the beginning, three views have been used for each sequence; two views have been used to produce a virtual view, while the third view has been used as a view synthesis reference for quality assessment. For the virtual views synthesis, state-of-the-art synthesis software called View Synthesis Reference Software [Sta13b] has been used. This software package developed by MPEG allows high-quality virtual view rendering based on two videos and two depth maps.

Sequence name	Resolution	Position of the encoded views	Position of the synthesized view
Training set			
Ballet [Zit04]	1024×768	3, 5	4
Breakdancers [Zit04]	1024×768	2, 4	3
BBB Butterfly [Kov15]	1280×768	49, 51	50
BBB Flowers [Kov15]	1280×768	39, 41	40
Kermit [Sal19]	1920×1080	5,7	6
Poznan_CarPark [Dom09]	1920×1088	3,5	4
Verification set			
Poznan_Block2 [Dom16]	1920×1080	2, 6	4
Poznan_Fencing [Dom16]	1920×1080	2, 6	4

Table 1. Test sequences used in experiments

For the experiments, version 3D-HTM 16.2 of the 3D-HEVC reference software [3DHEVC] was used. The encoder was configured according to the MPEG common test conditions for 3D video [Mul13].

To simplify and study bitrate allocation between views and depth maps, we have assumed that the quantization parameter for video (*QP*) is constant for all views, and the quantization parameter for depth maps (*QD*) is also fixed for all depth maps. Thus, two quantization parameters have been used to control the encoder instead of four.

4. THE PROPOSED MODEL

We have assumed that there exists a simple linear relationship between the quantization parameters for color data (*QP*) and depth data (*QD*).

$$QD(QP) = \alpha \cdot QP + \beta \quad (1)$$

For estimating the parameters of the proposed model, first we need to find all optimum quantization parameter (*QP*-*QD*) pairs for multiview sequences from the training set. Similarly to [Alo18a, Alo19], we

have tested all possible quantization parameter pairs (*QP*-*QD* pairs) in the range of 25 to 51.

In Figure 3, we have presented the obtained R-D (rate-distortion) curves. Each red point represents a coding result of one quantization pair (*QP*-*QD* pair). Let us bear in mind that the quality of compression is assessed as the quality of the virtual view generated based on decoded data. Based on this raw data we have selected all the pairs of quantization parameters (*QP*-*QD* pairs) which lead to the best quality of a virtual view generated out of data compressed at the given bitrate (blue lines in Figure 3). The optimum quantization pairs (*QP*-*QD* pairs) belong to an envelope of a raw cloud of PSNR-bitrate points. Those selected optimal pairs of quantization parameters have been presented onto a *QP*-*QD* plain (Figures 4 and 5). As it can be seen, those pairs lie almost on a straight line.

By means of linear regression, we have estimated the parameters of a linear model connecting the quantization parameter of depth data with the quantization parameter of color images. The estimated parameters have been gathered in Table 2. Our model allows us to calculate one quantization parameter based on the other, while maintaining the optimal bitrate allocation.

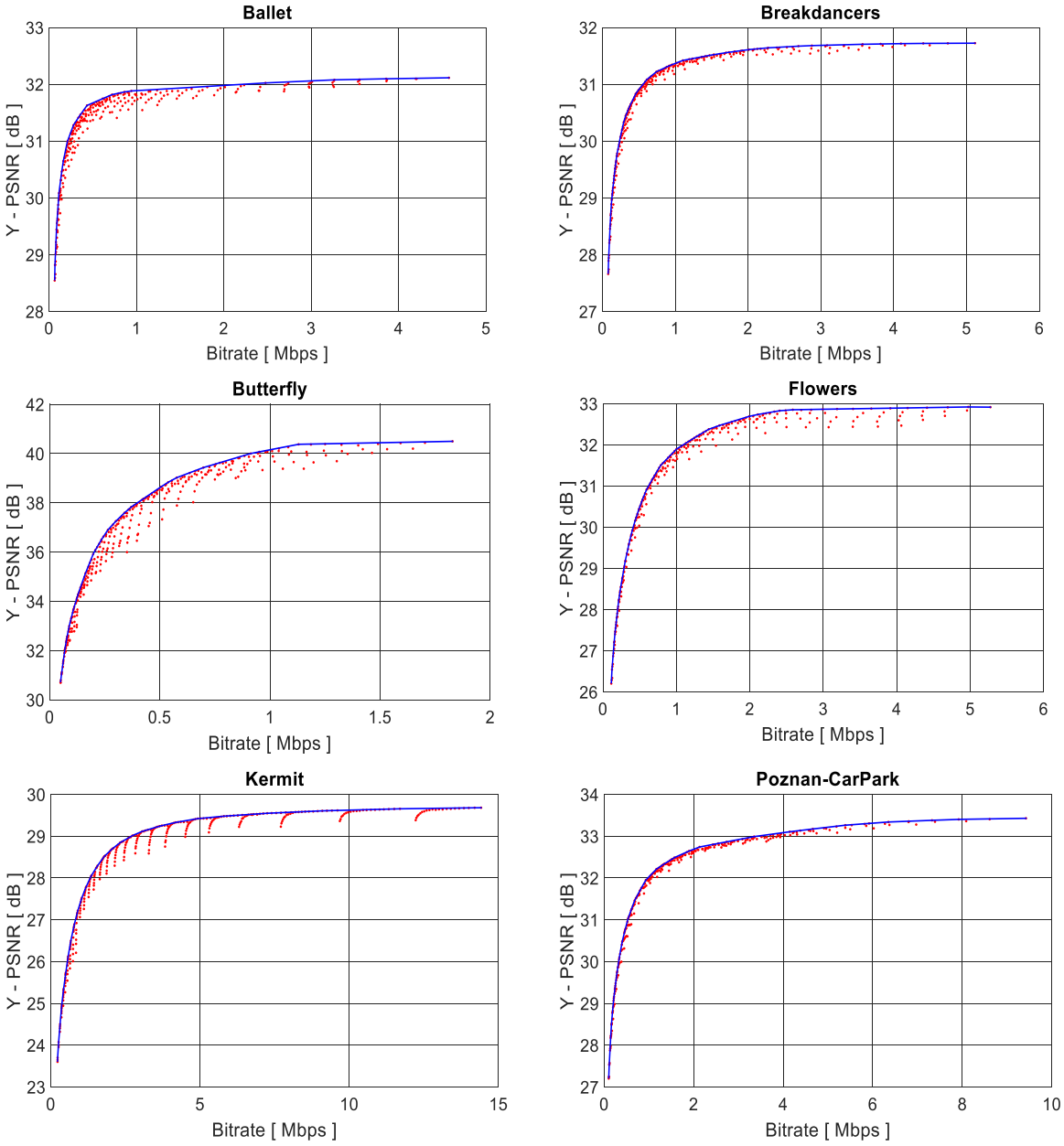

Figure 3. The best curve calculated by coding a video with all *QP*-*QD* pairs

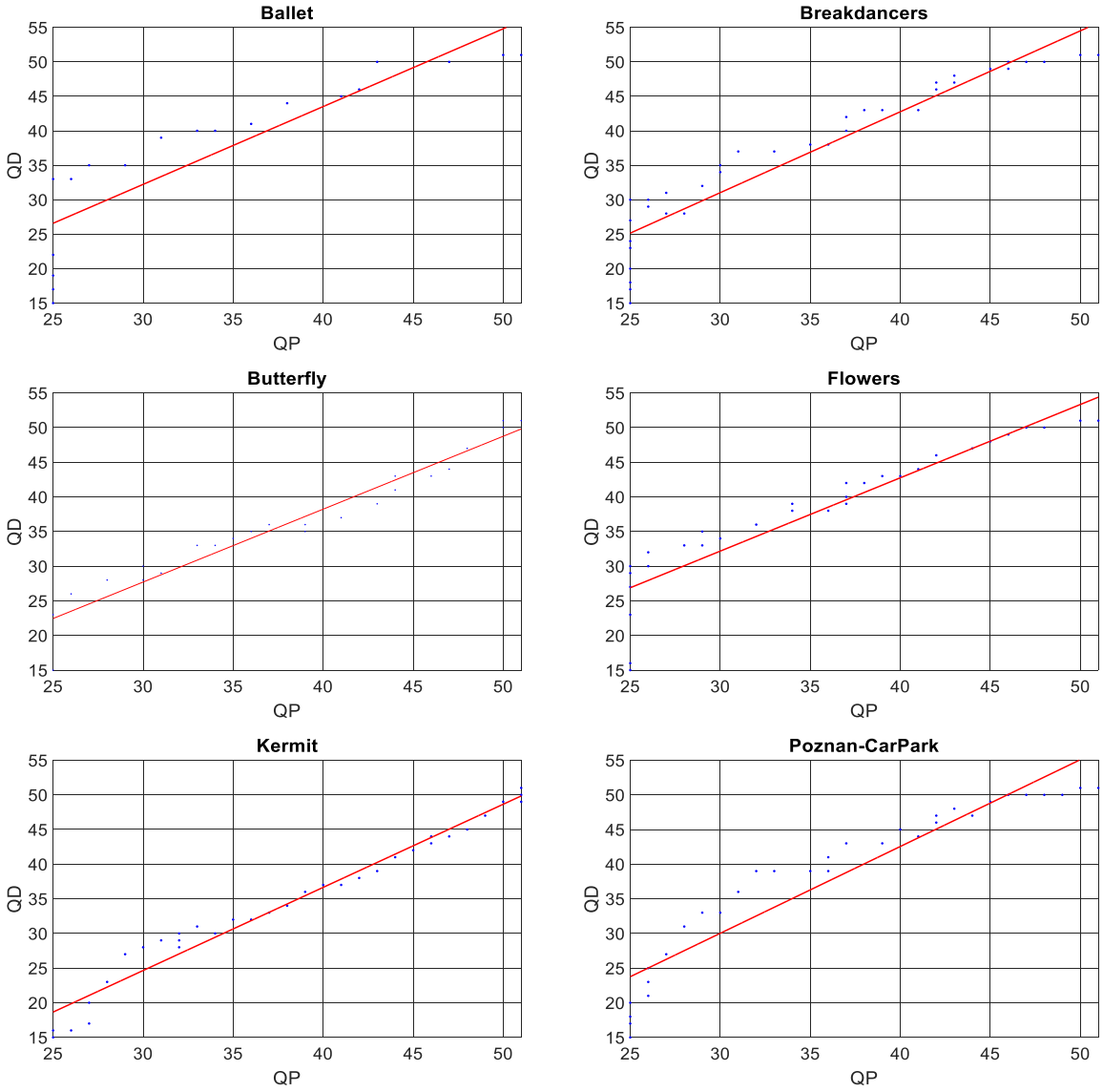

Figure 4. The approximate relationship between *QP* and *QD* for the optimum pairs for each training sequence with the use of linear regression

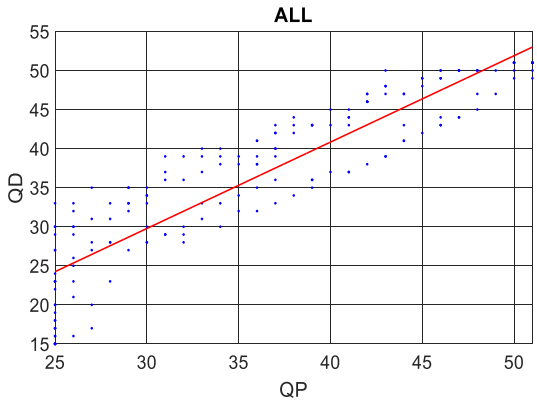

Figure 5. Optimum *QP-QD* pairs for all training sequences (blue points) and approximation line for *QP-QD* relationship (red line)

Sequence	α	β
Ballet	1.13	-1.64
Breakdancers	1.17	-4.19
Butterfly	1.05	-3.84
Flowers	1.06	-0.44
Poznan_CarPark	1.25	-7.55
Kermit	1.20	-11.34
Average	**1.11**	**-3.40**

Table 2. Parameters α and β for the linear regression model approximation for optimum *QP-QD* pairs.

5. PERFORMANCE OF THE PROPOSED MODEL

We have tested the proposed model in order to compare the optimum pairs of quantization parameters with another test set (verification set). As mentioned before, it was composed of two multiview sequences: Poznan_Block2 and Poznan_Fencing.

We have encoded both sequences with quantization parameters according to our proposed model (according to average values from Table 2) and with the quantization parameter pairs provided in the common test condition (CTC) document used during the development of 3D-HEVC [Mul13]. CTC specifies in detail the encoder configuration, including quantization parameter pairs for 3D-HEVC testing.

After the encoding, we have compared the total resulting bitrate and the quality of the virtual view generated based on the decoded data.

The comparison of the proposed model with the CTC conditions has been performed by calculating the average difference between the curves for PSNR (ΔPSNR) and bitrate (ΔBitrate). The calculated ΔPSNR and ΔBitrate are just a simple extension of the well-known Bjøntegaard metric [Bjo01] to work with more than four points. The obtained results can be found in Tables 3 and 4. Figure 6 presents R-D (rate-distortion) curves for one of the verification multiview sequences - Poznan_Block2.

	Tested / Reference	CTC	Proposed model	Optimum *QP-QD* pairs
Poznan_Block2	CTC	N/A	**-34.36%**	-51.99%
	Model	52.35%	N/A	-27.10%
	Optimum	108.29%	37.17%	N/A
Poznan_Fencing	CTC	N/A	**-23.05%**	-44.44%
	Model	31.58%	N/A	-28.22%
	Optimum	79.99%	40.26%	N/A

Table 3. Comparison of Bjøntegaard ΔBitrate metrics for optimum *QP-QD* pairs, common test condition (CTC) and the proposed model for 3D-HEVC

Experiments conducted on a verification set have shown that the proposed model for calculating quantization parameter pairs (Table 3) led to a decrease of total bitrate and improved the virtual view quality of sequences when compared to the quantization pairs recommended in the common test condition (CTC) document (23-35%). However, as always in modeling, the usage of the model leads to an increase in total bitrate and a decrease of virtual view quality in comparison to the usage of optimum *QD-QP* pairs. It is worth noticing that optimum *QP-QD* pairs for a given sequence are not known in advance, especially when new content is used. This is why the proposed model is an improvement over the recommendation included in the CTC document.

	Tested / Reference	CTC	Proposed model	Optimum *QP-QD* pairs
Poznan_Block2	CTC	N/A	**0.28 dB**	0.51 dB
	Model	-0.28 dB	N/A	0.22 dB
	Optimum	-0.51 dB	-0.22 dB	N/A
Poznan_Fencing	CTC	N/A	**0.05 dB**	0.12 dB
	Model	-0.05 dB	N/A	0.06 dB
	Optimum	-0.12 dB	-0.06 dB	N/A

Table 4. Comparison of Bjøntegaard ΔPSNR metrics for optimum *QP-QD* pairs, common test condition (CTC) and the proposed model for 3D-HEVC

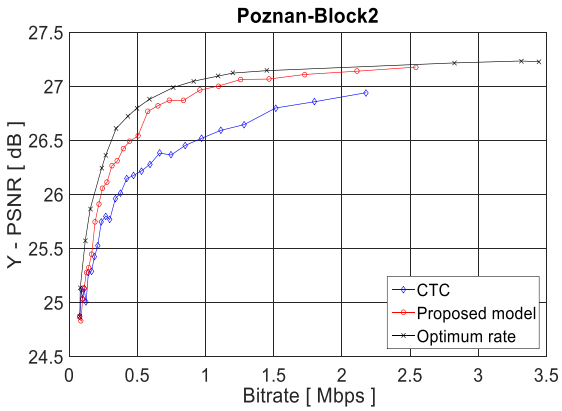

Figure 6. R-D curves comparison between the proposed model, CTC approach, and optimum *QP-QD* pairs for 3D-HEVC coding for the Poznan_Block2 sequence

6. CONCLUSIONS

In the paper, we have proposed a simple model for optimal bitrate allocation between color and depth data in 3D-HEVC coding. The provided model allows better bitrate division between color and depth data, which leads to a significant (23-35%) bitrate reduction of the total bitrate of the multiview stream. At the same time it preserves the same quality of synthesized virtual views in comparison to the common test condition (CTC) recommendation, which is considered as a well-established reference encoder configuration.

Moreover, based on the proposed model, we can control multiview compression by using only one quantization parameter instead of two parameters.

7. ACKNOWLEDGMENTS

The research was supported by the Ministry of Science and Higher Education of the Republic of Poland.

8. REFERENCES

[Alo18a] 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.

[Alo18b] Y. Al-Obaidi, T. Grajek, O. Stankiewicz, M. Domanski, "Bitrate allocation for multiview video plus depth simulcast coding," Int. Conf. Systems, Signals, and Image Proc. (IWSSIP), Maribor, Slovenia, 2018.

[Alo19] Y. Al-Obaidi, T. Grajek, M. Domanski, "Quantization of Depth in Simulcast and Multiview Coding of Stereoscopic Video plus Depth Using HEVC, VVC and MV-HEVC," Picture Coding Symposium (PCS), Ningbo, China, Nov. 2019.

[Ann14] Annex I "Multiview and Depth video coding" of ISO/IEC 14496-10, Int. Standard "Generic coding of audio-visual objects – Part 10: Advanced Video Coding," 8th Ed., 2014, also: ITU-T Rec. H.264, Edition 12.0, 2017.

[Ann18] Annex I "3D high efficiency video coding" of ISO/IEC 23008-2, Int. Standard "High efficiency coding and media delivery in heterogeneous environments - Part 2: High efficiency video coding," 4th Ed., 2018, also: ITU-T Rec. H.264, Edition 7.0, 2019.

[Bjo01] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16, Doc. VCEG-M33, USA, 2001.

[Bos11] 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, June 2011.

[Bos13] E. Bosc, F. Racape, V. Jantet, P. Riou, M. Pressigout, L. Morin, "A study of depth/texture bit-rate allocation in multi-view video plus depth compression," Annals of telecommunications, Vol. 68, Issue 11–12, pp 615–625, 2013.

[Dom09] M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, K. Wegner, "Poznań multiview video test sequences and camera parameters," ISO/IEC JTC1/SC29/WG11 MPEG2009, M17050, Xi'an, China, Oct. 2009.

[Dom16] 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, Switzerland, 2016.

[Feh04] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," Proc. SPIE, 2004.

[Han13] M. Hannuksela, D. Rusanovskyy, W. Su, L. Chen, R. Li, P. Aflaki, D. Lan, M. Joachimiak, H. Li, M. Gabbouj, "Multiview-video-plus-depth coding based on the advanced video coding standard," IEEE Trans. Image Proc., Sep. 2013.

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

[Isg04] F. Isgro, E. Trucco, P. Kauff, O. Schreer, "Three-dimensional image processing in the future of immersive media," IEEE Trans Circuits Syst. Video Techn., vol. 14, pp. 288 – 303, 2004.

[JCT17] JCT-3V. JCT-3V Document Management System, Sep. 2017.

Accessed: http://phenix.int-evry.fr/jct2

[Kli14a] K.Klimaszewski, K.Wegner, M.Domański, "Video and Depth Bitrate Allocation in Multiview Compression," Int. Conf. Systems, Signals, and Image Proc. (IWSSIP), Dubrovnik, Croatia, June 2014.

[Kli14b] K. Klimaszewski, O. Stankiewicz, K. Wegner, M. Domański, "Quantization optimization in multiview plus depth video coding," IEEE Int. Conf. on Image Proc. (ICIP), Paris, France, Oct. 2014.

[Kov15] P. Kovacs, "[FTV AHG] Big Buck Bunny light-field test sequences," ISO/IEC JTC1/SC29/WG11, Doc. MPEG M35721, Geneva, Switzerland, 2015.

[Mie18] D. Mieloch, A. Grzelka, "Segmentation-based method of increasing the depth maps temporal consistency," Int. Journal of Electronics and Telecommunications, Warsaw, Poland, 2018.

[Liu15] C. Liu, W. Zhang, Z. Qi, L. Shi, "A robust temporal depth enhancement method for dynamic virtual view synthesis," 23rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, WSCG 2015 - Full Papers Proceedings, pp. 191-200, Plzen, Czech Republic, 2015.

[Mul11] K. Muller, P. Merkle, T. Wiegand, "3-D video representation using depth maps," Proc. IEEE, April 2011.

[Mul12] K. Müller, P. Merkle, G, Tech, T, Wiegand, "3D video coding with depth modeling modes and view synthesis optimization," Proceedings of The 2012 Asia Pacific Signal and Information

Processing Association Annual Summit and Conference, Hollywood, USA, 2012

[Mul13] 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.

[Par20] K. Park, S. Kim, K. Sohn, "High-precision depth estimation using uncalibrated LiDAR and stereo fusion," IEEE Trans. Intelligent Transportation Systems, Jan 2020.

[Qin17] Y. Qin, X. Jin, Y. Chen, Q. Dai, "Enhanced depth estimation for hand-held light field cameras," IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP), New Orleans, LA, 2017.

[Sal19] B. Salahieh, B. Kroon, J. Jung, M. Domański, "Test Model for Immersive Video," ISO/IEC JTC1/SC29/WG11 MPEG/N18470, Geneva, Switzerland, April 2019.

[Sam16] 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.

[Son17] Y. Song, Y. Ho, "High-resolution depth map generator for 3D video applications using time-of-flight cameras," IEEE Trans. Consumer Electronics, vol. 63, Nov. 2017.

[Sta13a] O. Stankiewicz, K. Wegner, M. Domański, "AHG14: Optimized QP/QD curve for 3D coding with half and full resolution depth maps," ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Doc. JCT-3V E0269, Vienna, Austria, Aug. 2013.

[Sta13b] 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, Switzerland, 2013.

[Sta15] 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.

[Sta18] O. Stankiewicz, M. Domański, A. Dziembowski, A. Grzelka, D. Mieloch, J. Samelak, "A free-viewpoint television system for horizontal virtual navigation," IEEE Trans. Multimedia, Jan. 2018.

[Tan12] M. Tanimoto, M. Panahpour, T. Fujii, T. Yendo, "FTV for 3 D spatial communication," Proc. IEEE, Feb. 2012.

[Tec16] 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 Technology, Jan. 2016.

[Weg18] K. Wegner, O. Stankiewicz, T. Grajek, M. Domański, "Depth estimation from stereoscopic 360-degree video," IEEE Int. Conf. Image Proc. (ICIP), Athens, Greece, Oct. 2018.

[Zel14] N. Zeller, F. Quint, L. Guan, "Kinect based 3D scene reconstruction," 22nd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, WSCG 2014, pp. 73-81, Plzen, Czech Republic, 2014.

[Zin10] S. Zinger, L. Do, P. H. N. de With, "Free-viewpoint depth image based rendering," J. Vis. Commun. Image Represent., vol. 21, no. 5, pp. 533–541, Jan. 2010.

[Zit04] 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.