Video and Depth Bitrate Allocation in Multiview Compression

Krzysztof Klimaszewski, Krzysztof Wegner, Marek Domański
Chair of Multimedia Telecommunications and Microelectronics, Poznan University of Technology,
Polanka 3 61-131 Poznan, Poland
kklima@et.put.poznan.pl

Abstract - **In the paper, a novel method for determining the optimal ratio between bitrate of texture (video) and depth data in a multiview compression with depth maps is presented. The method provides an easy way of adjusting the bitrate of compressed data stream using only a single parameter, while maintaining the correct balance between video and depth bitrate. The balance is optimized for the best possible quality of the synthesized view that can be generated in the range between input views. The works concentrate on a scenario where the video and depth maps are compressed independently using the contemporary multiview coder.**

Keywords **- 3D video coding, depth coding, multiview compression, multiview with depth, MVC**

I. Introduction

Recently the application scenarios for multiview video (a set of videos from a number of synchronized cameras capturing the same scene from different viewpoints) have been studied, together with an advanced concept of a free viewpoint television [24]. In this paper, the scope of work is multiview video with depth [18] coding.

One of the applications of multiview video with depth maps is to provide necessary data for novel autostereoscopic displays [4]. These displays allow the viewers to perceive three dimensional images without the need of wearing any kind of glasses and to perceive the parallax effect. Those displays require a large number of views (over 20) captured by very closely positioned cameras. Capturing a video from a very dense set of cameras, as well as transmitting such a large number of raw views, turns out to be impractical. Therefore the need for an efficient way of sending necessary data arises. One solution is to transmit only a small number of views together with additional information about geometry of the scene in a form of depth maps, and synthesize (render) dense set of necessary views at the receiver side [18]. In modern video systems the data, prior to being sent to viewers, is compressed with lossy compression algorithms. Those algorithms have an inherent property of introducing a tradeoff between reconstructed video quality and the bitstream size. In multiview video with depth there is an additional tradeoff that needs to be considered. In such a case the proper ratio between bitrate for video and depth needs to be established by correctly setting quantization parameters for video and depth.

In this paper, the way of finding a formula for estimating suitable compression parameter for depth data based on given compression parameter for video is presented. Developed methodology allows to provide close to best possible quality of synthesized views presented to viewers while keeping bitrate at a given level.

II. Quality Measures for Systems Using Multiview Video with Depth

The data format considered in the paper contains two, completely different kinds of data - video and depth. The most popular objective quality metric for video is PSNR. PSNR metric can provide results comparable to other, more complicated metrics, when considering video quality [7]. The problem is the way to asses quality of the depth data. Depth data is not directly presented to the viewer, but rather used for Depth Image Based Rendering [20] of the virtual view. It is only the view synthesized with use of the depth data that is directly presented to the viewer. Thus, as results shown in [7][14] suggest, PSNR of the depth data is not a good indicator of compressed depth quality. For the depth maps it is important how their distortions influence the process of view synthesis and the quality of synthesized virtual views. Not every part of the depth map is equally important in the process of virtual view synthesis. In fact, there exist areas in depth maps, where the distortions in depth map will not influence the virtual view synthesis process at all. Unfortunately, PSNR metric does not provide this kind of information. Therefore the quality of depth maps cannot be considered separately from primary purpose of the depth maps, which is virtual view synthesis [26]. It is the virtual view quality that should be considered when evaluating the quality of the multiview video with depth. This is why such a method was adapted by Moving Pictures Experts Group (MPEG) of the International Standardization Organization (ISO/IEC) in the process of evaluating methods of processing and compression of the multiview video with depth [17]. Adopted method assesses the performance of compression method by evaluating the quality of virtual views in terms of PSNR.

In the presented paper, two different variants of quality measures are considered. The first method compares a virtual view quality against virtual view generated from original uncompressed data [5]. The other comparison is done against a view from a real camera, positioned and oriented exactly in the same point as the virtual camera that would capture the virtual view. The results obtained using second method may reflect better the subjective tests results, as shown in [20]: a saturation of quality measure is typically observed in subjective quality tests.

III. Multiview with Depth Compression Method

Compression of multiview video with depth can be easily realized with the use of a multiview video coder, such as MVC (Multiview Video Coder), specified in an annex H of the MPEG4/AVC (Advanced Video Coder) standard [10]. Although the MVC was not originally developed for depth map compression, many research works proved its usefulness [12][13] and it was recently successfully adopted by the standardization committee without any modifications, for compression of the depth data as well.

In the presented method, both video and depth maps, are encoded independently with MVC. The appropriate ratio between the bitrates for those two types of data needs to be established, as it, together with total bitrate setting, influences the view synthesis quality.

IV. Bitrate Allocation for Video and Depth

The problem considered in the paper is related to allocation of bitrate for views and depth maps in multiview video with depth compression. Improper balance between those two kinds of data can result in a decrease of virtual view quality that can be achieved at given total bitrate. Therefore, developing a method for adjusting the bitrate ratio between video and depth data is a very important in development of a multiview system.

A. Previous work

In previously published works, the problem of bitrate allocation for views and depth maps has been already noticed. In [19], authors propose to allocate bitrate in a way that the MSE (Mean Squared Error) is kept at the same level for views and depth maps, but do not provide any algorithm for selecting the quantization parameters that satisfy this requirement. In [2] authors present a method for bitrate allocation that is based on adaptive approach that depends on currently used virtual view. Although this method is more accurate than the one presented here, it is also much more complex. Another adaptive algorithm is described in [16]. It is also more complex than the one shown in this paper. In the works documented in [1], authors estimate the optimal bitrate ratio based on two sequences, but do not provide any direct formula for calculating quantization parameters for video and depth data. They only estimate that about 40% to 60% of bitrate should be devoted to depth data. In [15] authors propose a distortion model for estimating the virtual view quality. Authors observe that their method is less complex than the full search method over all possible *QP*-*QD* pairs (*QP* – quantization parameter for views, *QD* – quantization parameter for depth maps). In this aspect, the method presented here is even less complex, as it requires no computations at all during the compression process. In [25] and [9] the authors develop models for determining the proper *QP* and *QD* values and demonstrate good performance of their methods. Those methods have, however, a significant shortcoming – for proper operation they require a few test compression runs for every compressed sequence. In the approach proposed here, no model parameters are required beforehand nor during the compression, therefore no overhead is added over the standard compression process.

There is no documented research on much simpler method, that uses an estimated relation of the quantization parameter for video *QP* and depth *QD*. This paper brings report on such an approach. In the following section, a method for automatic management of bitrate ratio is presented. It allows the reduction of the number of coefficients used for control of bitrate to only one, namely *QP* - just like in the compression of traditional video.

B. Experimental setup

The compression method used in presented work is based on reference MVC coder [11]. Video and depth data for two views are compressed. The same MVC coder, with the same configuration, is used for views and depth maps compression. The compression process in this approach is controlled by two quantization parameters, *QP* (for views) and *QD* (for depth). Each can be set independently. To verify the quality of the reconstructed data, the decompressed video and depth maps are fed to the view synthesis algorithm used in MPEG works [23] and a virtual view is produced. The quality of this virtual view is then evaluated using both methods outlined in Section II.

C. Evaluation of the performance of the described method

The performance of the described compression method was evaluated on a set of multiview test sequences with depth maps recommended by the MPEG for research on multiview video with depth compression methods [6][8][21][3]. Eight cases were considered, as listed in Table I. The number of sequences used in the experiment was limited, since the availability of high quality depth maps that are produced using supervised method of depth map calculation is limited. For each scenario, exhaustive tests were conducted: 42 different values of QP (from 10 to 51) and 42 different values of QD (from 10 to 51) were considered, resulting in 1764 bitstreams for each case.

The results show that there is an optimal bitrate ratio between video and depth map for each total bitrate value (see Fig. 2 and 3 for reference – the solid green lines). With this

TABLE I
Cases Considered in The Experiments

Scenario number	Sequence name	Resolution	Frames per second	Numbers of cameras used	Synthesized camera	Number of frames used
1	Book Arrival	1024 x 768	16.67	8 and 10	9	100
2	Newspaper	1024 x 768	30	4 and 6	5	100
3	Undo Dancer	1920 x 1088	25	1 and 3	2	16
4	Poznan Hall 2	1920 x 1088	25	5 and 7	6	16
5	Poznan Hall 2	1920 x 1088	25	6 and 7	in the middle between 6 and 7	16
6	Poznan Hall 2	1920 x 1088	25	6 and 7	in the middle between 6 and 7	100
7	Poznan Street	1920 x 1088	25	3 and 5	4	16
8	Poznan Street	1920 x 1088	25	3 and 4	in the middle between 3 and 4	16

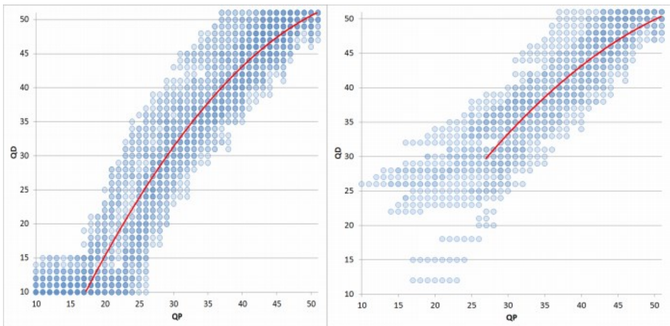
optimal ratio, the best quality of the synthesized views can be obtained for a given bitrate.

For some cases the decrease of QD (increase quality) of the depth data while keeping the same QP (the same quality) of the video may lead to a decrease of virtual view quality in case of comparison with real captured reference view (Fig. 2 bottom). This phenomenon can be attributed to deficiencies of the provided depth maps and/or view synthesis algorithms. Those deficiencies can be slightly masked by depth map encoding artifacts and result in increase of virtual view quality. This phenomenon is noticeable also for a synthetic sequence ("Undo Dancer"). For this sequence a perfect, accurate depth map is provided.

D. Derivation of formula for bitrate distribution

To find the relationship between *QP* and *QD*, the *QD* and *QP* values for points of optimal bitrate distribution were plotted, as shown in Fig. 1a and 1b. Two sets of results were obtained: for the case where quality measure is based on comparison with view synthesized from uncompressed data, (Fig. 1a) and with the original view (Fig. 1b). In Fig. 1a, for *QP* values smaller than 17 the points are concentrated along the horizontal line at the level of *QD* approximately equal to 11 and are excluded from the interpolation process. Similarly, in Fig. 1b, for *QP* values smaller than 27 are also excluded since they are too scattered to allow for a reasonable quality of approximation. The high value of the QP limit is doubtlessly a limiting factor of the method.

In order to estimate relationship between QP and QD it was decided to use the simplest possible regression, and use polynomial approximation. The order of the polynomial was set to two, in order to give the ability to follow the curvature that is evident on the graph, while keeping the degree of approximation polynomial as low as possible. Moreover, by using the second order polynomial, the non-monotonicity of the curve that occurs for higher order polynomials, can be avoided.



(a) (b)

Fig. 1. Approximation curve for *QP*-*QD* relationship. Color intensity corresponds to the number of scenarios for which a given pair lays on the optimal line - the more cases, the more intense the color. Comparison to view synthesized with uncompressed data (a) and real view (b).

The obtained curves are shown in Fig. 1a and 1b as straight red lines. The formula for the curve in Fig. 1a, found using polynomial regression, is:

$$QD = -0.0216 \times QP^2 + 2.6872 \times QP - 29.876, \qquad (1)$$

The coefficient of determination is used to verify the quality of approximation. Its value is R^2=0.8521, indicating good approximation. For comparison with a real view case, the formula is:

$$QD = -0.0155 \times QP^2 + 2.073 \times QP - 14.885, \qquad (2)$$

The value of coefficient of determination is R^2=0.6936, indicating acceptable approximation. It is suggested to use the following formulas, depending on the evaluation method chosen. For comparison with synthesized view:

$$QD = \begin{cases} \lfloor -0.0216 \times QP^2 + 2.6872 \times QP - 29.376 \rfloor, & QP > 16 \\ 11, & QP \leq 16, \end{cases}$$

(3)

and for comparison with real view:

$$QD = \lfloor -0.0155 \times QP^2 + 2.073 \times QP - 14.385 \rfloor \qquad (4)$$

The presented above equations have been obtained by exploration of experimental data.

V. PERFORMANCE OF THE DESCRIBED METHOD

The applicability of the formulas was verified on a sequence from outside the training set. In this case, another multiview sequence recommended by the MPEG [22] – "Dog" – was used. In the test, 100 frames from views 38, 41 were encoded and view 39 was synthesized in order to measure the quality of the reconstruction. The performance of the method is presented in Fig. 2 (where virtual view was used as reference for PSNR estimation) and in Fig. 3 (where a real view was used as reference for PSNR estimation).

The set of green (solid) lines is obtained by performing the process for every possible *QP*-*QD* pair, while the thick red dashed line connects the points obtained by using the formulas proposed - formula (3) for Fig. 2 and formula (4) for Fig. 3.

It can be seen that the red (dashed) line follows the line of optimal *QP*-*QD* pairs very closely. The largest differences can be observed for large bitrates in Fig. 2, where the difference between the quality of virtual view obtained with data compressed using the presented method and the best performance obtained in exhaustive search is still smaller than 0.5 dB. This range of bitrates, however, is rarely used in practice. For most cases, the difference is much smaller, less than 0.1 dB.

VI. CONCLUSIONS

In the paper, we proposed simple formulas that allow bitrate allocation in multiview video with depth compression using MVC coder. The described method provides a good estimate of the quantization parameter *QD* that results in close to optimal bitrate allocation for views and depth for a given quantization parameter for views *QP*. This method allows that a single parameter - quantization parameter for view compression (*QP*) can be used for the control of bitrate of a bitstream with multiview video and depth data. The parameter *QD* for depth map compression is calculated automatically and proper bitrate allocation is automatically established. Notably, no overhead calculations are nececessary during compression.

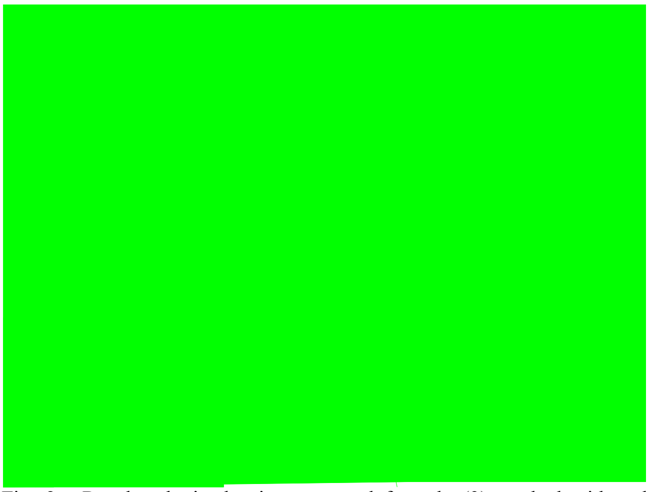


Fig. 2. Results obtained using proposed formula (3) marked with red (dotted) line. Results for every possible pair *QP-QD* in the considered range - green (solid) lines.

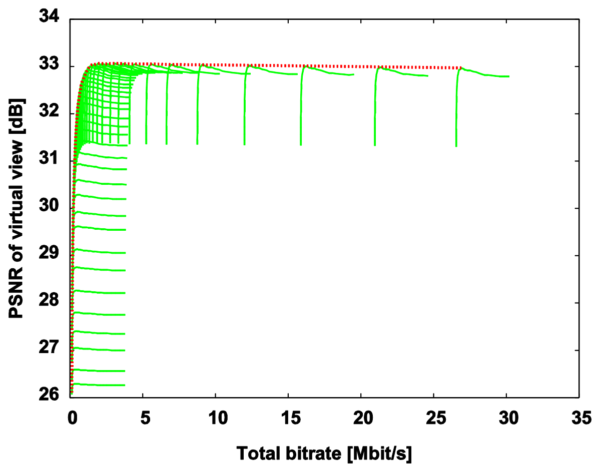


Fig. 3. Results obtained using proposed formula (4) marked with red (dotted) line. Results for every possible pair *QP-QD* in the considered range - green (solid) lines.

ACKNOWLEDGMENT

The research was financed by the National Science Centre Poland, decision number DEC-2012/05/B/ST7/01279.

REFERENCES

[1] 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), 2011,* 16-18 May 2011.

[2] 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.

[3] 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 MPEG Doc. M17050*, 2009.

[4] I.T. Doser, F.Hassenpflug, R.Zwing, "The Glasses Are Half Empty: Turn Your Sight to Autostereoscopic 3-D," *IEEE Consumer Electronics Magazine* vol.1, no.1, pp.49-60, Jan. 2012.

[5] N.A. El-Yamany, K. Ugur, M.M. Hannuksela, M. Gabbouj (2010) "Evaluation of depth compression and view synthesis distortions in multiview-video-plus-depth coding systems," *3DTV-Conf: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*, 7-9 June 2010.

[6] I. Feldmann, M. Muller, F. Zilly, R. Tanger, K. Muller, A. Smolic, P. Kauff, T. Wiegand, "HHI Test Material for 3D Video," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M15413*, 2008.

[7] C. Hewage, S. Worrall, S. Dogan, S. Villette, A. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE Journal of Sel. Topics in Sig. Proc.* vol.3, no.2, pp.304-318, April 2009.

[8] Y.S. Ho, E.K. Lee, C. Lee, "Multiview video test sequence and camera parameters," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15419*, 2008.

[9] S. Hu, S. Kwong, Y. Zhang, C.-C.J. Kuo, "Rate-Distortion Optimized Rate Control for Depth Map-Based 3-D Video Coding," *IEEE Transactions on Image Processing,* vol.22, no.2, pp.585-594, Feb. 2013.

[10] *Information technology – Coding of audio-visual objects – Part 10: Advanced Video Coding*, International standard ISO/IEC 14496-10, 2010.

[11] JMVC reference codec (April, 2011) [Online] cvs: garcon.ient.rwth-aachen.de/cvs/jvt

[12] K. Klimaszewski, K. Wegner, M. Domański, "Distortions of synthesized views caused by compression of views and depth maps," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009*, 4-6 May 2009.

[13] K. Klimaszewski, K. Wegner, M. Domański, "Influence of views and depth compression onto quality of synthesized views," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M16758*, 2009.

[14] G. Leon, H. Kalva, B. Furht, "3D video quality evaluation with depth quality variations," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, pp.301-304, 28-30 May 2008.

[15] Y. Liu, Q. Huang, S. Ma, D. Zhao, W. Gao, "Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model," *Signal Processing: Image Communication* ,vol. 24, no. 8, pp. 666-681, Sept. 2009.

[16] 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.

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

[18] K. Muller, P. Merkle, T. Wiegand, "3-D video representation using depth maps," *Proceedings of the IEEE* , vol.99, no.4, pp.643-656, April 2011.

[19] K. Muller, A. Smolic, K. Dix, P. Merkle, T. Wiegand, "Coding and intermediate view synthesis of multiview video plus depth," Proc. of *16th IEEE International Conference on Image Processing*, pp.741-744, 7-10 Nov. 2009.

[20] G. Nur, S. Dogan, H.K. Arachchi, A.M. Kondoz, "Impact of depth map spatial resolution on 3D video quality and depth perception," *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010*, 7-9 June 2010.

[21] D. Rusanovskyy, P. Aflaki, M.M. Hannuksela, "Undo Dancer 3DV sequence for purposes of 3DV standardization," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M20028*, 2011.

[22] M. Tanimoto, T. Fujii, N. Fukushima, "1D parallel test sequences for MPEG-FTV," *ISO/IECJTC1/SC29/WG11 MPEG Doc. M15378*, 2008.

[23] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, Y. Mori, "Reference softwares for depth estimation and view synthesis," *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M15377*, 2008.

[24] M. Tanimoto, M. Panahpour Tehrani, T. Fujii, T. Yendo, "FTV for 3-D spatial communication," *Proceedings of the IEEE* , vol.100, no.4, pp.905,917, April 2012.

[25] H. Yuan, Y. Chang, J. Huo, F. Yang, Z. Lu, "Model-Based Joint Bit Allocation Between Texture Videos and Depth Maps for 3-D Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.21, no.4, pp.485-497, April 2011.

[26] Q. Zhang, P. An, Y. Zhang, Z. Zhang, "Efficient rendering distortion estimation for depth map compression," Proc. of *18th IEEE International Conference on Image Processing, 2011*, pp.1105-1108, 11-14 Sept. 2011.