QUANTIZATION OPTIMIZATION IN MULTIVIEW PLUS DEPTH VIDEO CODING

Krzysztof Klimaszewski, Olgierd Stankiewicz, Krzysztof Wegner, Marek Domański

Chair of Multimedia Telecommunications and Microelectronics,
Poznań University of Technology,
Polanka 3, 61-131 Poznań, Poland,
email: kklima@et.put.poznan.pl, [ostank, kwegner]@multimedia.edu.pl

ABSTRACT

In the paper, novel methods for quality control in multiview with depth video compression are described. First, a simple steepest-descent-based algorithm for adjusting quantization parameters for texture and corresponding depth to obtain the maximal quality for a given bitrate is proposed. Then, experimental results are shown, basing on which a mathematical model for calculation of quantization parameters for the depth and texture is proposed. The paper also covers subject of coding depth with different resolution than the texture. The work is presented in context of the most contemporary compression methods related to AVC, but could also be easily adapted to upcoming 3D-video coding extensions of HEVC.

Index Terms— 3D video coding, MVD, MVC+D, 3D-AVC, quantization control, depth maps, video compression

1. INTRODUCTION

This paper concerns 3D video coding systems in the context of multiview plus depth (**MVD**) video representation. In such, multiple views and accompanying depths are coded and transmitted to the receiver, where both texture and depth are used to synthesize virtual views, e.g. for autostereoscopic displays. The depth is not presented to viewers directly and is only used in view synthesis process. Thus, the quality of MVD video is measured by the quality of the output views and synthesized virtual views.

In classical 2D video coding, bitrate and quality are controlled by a single quantization parameter index, typically called QP. Larger values of QP correspond to worse quality and lower bitrate.

In 3D video in the MVD format, the problem is more complex as two components are transmitted and the quantization has to be controlled for texture (**denoted QP** in equations in this paper) and depth (**denoted QD**). Moreover, the resolution of the coded depth, can be different from that of the texture. Therefore, the aim of this paper is to provide an answer to the question: how to control QP and QD in order to maximize perceived output video quality?

Currently, there are two standardized coding technologies that can be used for MVD, both based on ubiquitous AVC [1] and its multiview extension MVC:

- **MVC+D** ("Multiview and Depth video coding" [2]) – a simple extension of MVC for depth which does not introduce any new coding tools, and
- **3D-AVC** ("AVC compatible video-plus-depth extension" [3]) – a more advanced extension, in which depth-to-texture prediction tools are used.

In both of these coding technologies, considered in this paper, QP and QD are used for quality-bitrate control.

The problem of selecting the proper balance between texture and depth bitrates has already been studied in a number of works. In [4] authors model distortions caused by depth and video quantization. Building of a model is time consuming, and cannot be easily applied in coders that encode texture and depth jointly. In [5] authors describe a sophisticated model for assigning QD values based on texture of a compressed image region. This method is applicable only to depth coding. In [6] and [7] a model of distortion caused by depth compression is used for bitrate allocation. In [8] authors use an exponential model for the optimal bitrate distribution between texture and depth. Another approach is presented in [9], where a simple equation is given for establishing QD parameter for depth coding. In those works, the proposed approach to finding quantization parameters that would maximize coding performance is to test, at least for model training purposes, all combinations of quantization parameters for texture and depth views (Fig. 1). Such approach can and have been successfully applied for coders based on unmodified MVC only, because coding of texture and depth is independent and thus the process does not require coding with all combinations of QP-QD pairs, but only texture coding with all QP possibilities and depth coding with all QD possibilities, separately.

Beyond the mentioned works, this paper considers a problem of finding the optimal QP-QD settings in MVC+D and 3D-AVC. Such coding technologies, pose a new challenge in optimization of 3D video coding process, due to even increased computational complexity and different texture-depth coding dependences. Therefore, application of the above-mentioned approaches would be very time-consuming. In this paper we propose two methods: algorithm-based and model-based, which aim at overcoming this disadvantage.

2. PROPOSED ALGORITHM

In a straight-forward method for finding the optimal QP-QD settings, all possible QP-QD combinations should be tested, resulting in a cloud of PSNR-bitrate points. This process is very time-consuming and inefficient, but the peak envelope over such cloud PSNR-bitrate points (Fig. 1) can be used to find the best R-D (rate-distortion) curve [9].

In this paper we propose a novel algorithmic method in which the shape of the optimal curve is found in iterative steepest-descent manner. It is performed by tracking the peak of coding performance on the QP-QD surface, defined by QP_i and QD_i values attained at successive iterations.

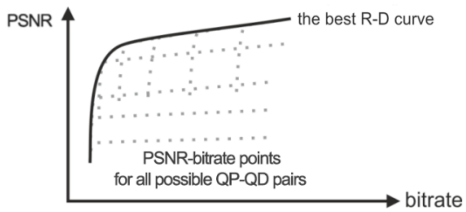

Fig 1. The best R-D curve calculated by encoding video with all possible QP-QD pairs.

The algorithm starts in iteration $i = 0$, with the largest assumed value of the both quantization parameters ($QP_{i=0} = QP_{MAX}; QD_{i=0} = QD_{MAX}$) which relate to the lowest quality and the smallest bitrate (bottom-left end of the R-D curve, Fig. 1). Then, at each next iteration $i + 1$, two possibilities of improving quality of the coded MVD video are tested:

a) increased quality of depth views (decreased quantization parameter for depth views) and unchanged quality of texture views: $QP^a_{i+1} = QP_i \quad ; \quad QD^a_{i+1} = QD_i - 1$,
b) increased quality of texture views (decreased quantization parameter for texture views) and unchanged quality of depth views: $QP^b_{i+1} = QP_i - 1 \quad ; \quad QD^b_{i+1} = QD_i$.

Therefore, either way, the output bitrate increases (denoted by $bitrate^a_i$ and $bitrate^b_i$), but it is used for different purposes in options (a) and (b): for improving quality of depth or for improving quality of texture, respectively.

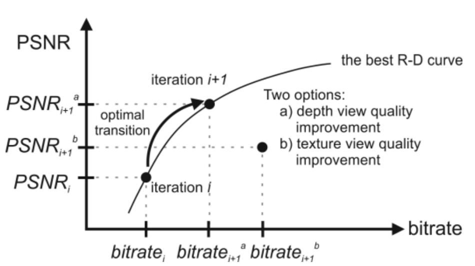

Fig. 2. Proposed steepest-descent optimization of quantization parameters for texture (QP) and depth (QD).

The two considered options ("a" or "b") are then compared with respect to their R-D performance (Fig. 2). For that, calculated are: the total bitrate of all views (denoted $bitrate_i$) and average luminance PSNR ($PSNR^a_i$ and $PSNR^b_i$) of 9 views: 3 coded views and (placed in between of them) 6 synthesized virtual views (Fig. 3). The better option "x" ("a" or "b") which has higher $quality\text{-}vs\text{-}bitrate\text{-}ratio$ is chosen and used in the next iteration:

$$quality\text{-}vs\text{-}bitrate\text{-}ratio^x = \frac{PSNR^x_i - PSNR_{i-1}}{bitrate^x_i - bitrate_{i-1}} \quad . \quad (1)$$

Such steepest-descent process stops when either of two quantization parameters reaches the lowest assumed quantization value (QP_{MIN} for texture or QD_{MIN} for depth).

Although the allowed range of QP parameter value in AVC standard is 1 to 51, in the experiments we have chosen the practical range of quantization parameter values: from $QP_{MIN} = QD_{MIN} = 10$ to $QP_{MAX} = QD_{MAX} = 50$.

With such an assumption, in the proposed approach, maximally 2×41 (two coding options, $QP, QD \in [10,50]$) coder passes are sufficient to find quantization parameter pairs that maximize coding performance, instead of 41^2 coding passes (all possible QP-QD pairs). This indicates reduction of number of coding passes required for finding the optimal QP-QD pairs by a factor of 20. It is important to note that this reduction of the required computational time is obtained at no cost in terms of accuracy of the method.

Fig. 3. The arrangement of the views in quality evaluation: the coded views are marked in black, while the synthesized virtual ("v") views are marked in gray.

3. ALGORITHM VERIFICATION

The algorithm has been tested in MVC+D and 3D-AVC with use of reference model software 3D-ATM v.8.1. For the experiments, so called Common Test Conditions (CTC) [10] have been used. This set of conditions has been created by JCT-3V group (ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development) in order to allow reproducibility of the research.

One of these conditions is quantization parameters setting, which is $QD = QP$. Such setting will be used **as a reference** for the results of papers proposals.

The experiments have been performed on 7 multiview video test sequences [11-15]. For each of them, 3 views (along with 3 depths) have been coded. The output quality of the coded MVD video has been measured basing on averaged luminance PSNR of 9 views: 3 coded views and 6 virtual views synthesized in between of the coded views in uniformly placed spatial positions (Fig. 3).

The tests have been performed in two configurations regarding resolution of the coded depth:

- "**full depth resolution**" – where coded depth maps have the same resolution as the coded texture views,
- "**half depth resolution**" – where coded depth maps are 2×2 times smaller compared to texture views.

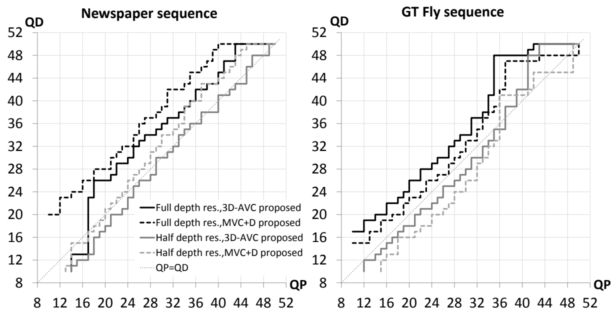

Fig. 4. Optimized quantization parameters pairs for texture (QP) and depth (QD) views for exemplary video test sequences.

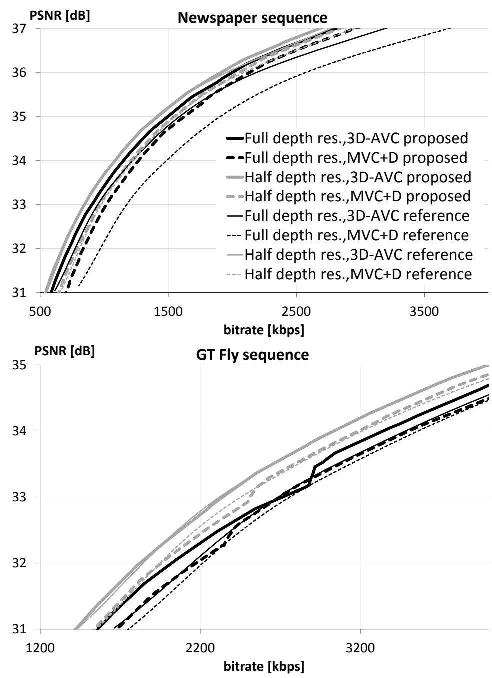

Fig. 5. R-D curves for MVD coding with algorithmically optimized quantization parameters (proposed) and with $QD = QP$ (reference).

The attained results have been presented in the form of QP-QD curves (Fig. 4) showing quantization parameter pairs trajectories that optimize coding performance and also in the form of rate-distortion curves (Fig. 5) showing the bitrate reductions of proposed QP-QD optimization versus the reference $QD = QP$ setting. The results are also summarized in Table 1 in the form of averaged bitrate reductions calculated with the use of Bjøntegaard metric[16].

It can be seen that usage of optimized quantization parameters gives better results than usage of $QD = QP$ reference curve. The bitrate reductions are respectively about 1.5% for MVC+D half depth resolution, 1.25% for 3D-AVC half depth resolution, about 8.5% for MVC+D full depth resolution and about 8% for 3D-AVC full depth resolution. This means that the bitrate reductions attained by usage of the proposed coding methodology are greater in the case of coding with full depth resolution that in the case of coding with half depth resolution.

Table 1. Bitrate reductions due to usage of the proposed algorithmically optimized QP-QD quantization pairs related to coding with $QD = QP$ (reference).

Sequence	**Half depth resolution**		**Full depth resolution**	
	MVC+D	**3D-AVC**	**MVC+D**	**3D-AVC**
Poznan Hall2	0.30%	0.29%	3.43%	3.87%
Poznan Street	0.99%	0.70%	9.38%	8.87%
Undo Dancer	1.57%	2.22%	2.08%	2.03%
GT Fly	1.90%	0.49%	1.88%	2.54%
Kendo	3.05%	3.23%	15.78%	17.55%
Balloons	1.65%	1.78%	12.43%	14.70%
Newspaper	1.49%	0.06%	14.57%	6.00%
Average	**1.57%**	**1.25%**	**8.51%**	**7.94%**

4. DEPTH RESOLUTION SELECTION

It can be noted that, counter-intuitively, results presented in Table 1 do not indicate which depth resolution coding variant (half depth resolution of full depth resolution) is more efficient. In order to asses that, we have compared the best-performing curves resulting from proposed algorithmic QP-QD optimization (in MVC+D and 3D-AVC, adequately) of half and full depth resolution coding. The results, presented in Table 2 show that coding with half depth resolution can be about 7% more efficient than coding with full depth resolution, both in MVC+D and 3D-AVC profiles. This confirms the claims that are present in state-of-the art coding outlines [10].

Table 2. Bitrate reduction due to usage of half depth resolution coding, related to full depth resolution, both with use of the proposed algorithmically optimized QP-QD quantization pairs.

Sequence	**MVC+D**	**3D-AVC**
Poznan Hall 2	7.72%	10.18%
Poznan Street	5.50%	6.45%
Undo Dancer	7.81%	9.45%
GT Fly	8.56%	9.75%
Kendo	6.56%	8.11%
Balloons	4.59%	5.52%
Newspaper	5.34%	6.02%
Average	**6.58%**	**7.93%**

5. PROPOSED MODEL

The QP-QD optimization algorithm presented above provides substantial bitrate reductions at moderate computational cost, compared to straight-forward method of encoding with all possible quantization parameter pairs. It is however worth to notice that there are applications where performing any encoding dry-runs is not desired, e.g. in low-latency encoding systems. In such applications, even an approximate QP -to- QD relation is of great interest.

To cater for that need, we present a mathematical model of $QD(QP)$ function, so that problem of QP-QD selection is significantly reduced. We model the shape of the optimal QP-QD curve with the use of linear regression, so that:

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

The pairs of coefficients α and β have been estimated with the use of least squares fitting to the optimal QP-QD pairs optimal generate by the proposed algorithm. The attained results are gathered in Table 3. As it can be seen, α coefficient (the slope of the QP-QD curve) is about 1, while β (the offset of the curve) is $-10.50 \leq \beta \leq 7.84$, depending on the case. In average, in the case of half depth resolution β is about -3.5 (which means that QD should be smaller than QP) and in the case of full depth resolution β is about 3 (which means that QD should be larger than QP).

Table 3. Parameters α and β for linear model approximation (Eq. 2) of algorithmically optimized QP-QD curve, estimated with linear regression with minimization of least squares line fitting.

Sequence	Half depth resolution		Full depth resolution	
	MVC+D	3D-AVC	MVC+D	3D-AVC
Poznan Hall 2	α= 1.32 β= -10.50	α= 1.20 β= -7.17	α= 1.20 β= -1.44	α= 1.21 β= -1.94
Poznan Street	α= 1.25 β= -5.51	α= 1.20 β= -4.64	α= 1.01 β= +7.83	α= 1.16 β= +3.27
Undo Dancer	α= 1.12 β= -7.72	α= 1.08 β= -6.69	α= 1.33 β= -6.09	α= 1.28 β= -5.60
GT Fly	α= 1.24 β= -9.56	α= 1.08 β= -3.42	α= 1.12 β= +0.11	α= 1.12 β= +2.89
Kendo	α= 1.23 β= -3.02	α= 1.21 β= -2.48	α= 1.11 β= +5.98	α= 1.06 β= +7.25
Balloons	α= 1.22 β= -3.31	α= 1.20 β= -2.79	α= 1.06 β= +7.02	α= 1.03 β= +7.84
Newspaper	α= 1.23 β= -4.48	α= 1.13 β= -5.12	α= 0.98 β= +9.31	α= 1.16 β= -0.12
Average	**α= 1.15 β= -3.97**	**α= 1.09 β= –2.80**	**α= 1.11 β= +3.42**	**α= 1.13 β= +2.44**

6. MODEL VERIFICATION

Pairs of QP and QD values, calculated from the model based on the averaged values from Table 3, have been used to run coding experiments, similar to the one presented above. The aim was to assess the performance loss of usage of the proposed model, instead of the accurately trained values generated with the proposed steepest-descent algorithm directly. The results – bitrate reductions (calculated with Bjøntegaard metric [16]), related to reference $QD = QP$ curve, are shown in Table 4. Compared with Table 1 it can be seen that the approximation of the model introduces minor deterioration of coding efficiency for training set, about 0.2 percent point in half depth resolution coding and about 0.75 p.p. in full depth resolution coding. This is relatively not much, especially as usage of the model allows QP-QD control without performing any dry-runs and without any computational burden imposed on the encoding process. The results for verification sequences (Poznan Car Park [11] and Shark [17]) prove the good performance of proposed method. The bitrate reduction for those test sequences follows the similar trend as for training set, with gains up to almost 28% of bitrate for full depth resolution.

Table 4. Bitrate reductions for the proposed model-based QP-QD quantization pairs related to coding with $QD = QP$ (reference).

Sequence	Half depth resolution		Full depth resolution	
	MVC+D	3D-AVC	MVC+D	3D-AVC
Training set average:	1.19%	1.12%	7.62%	7.40%
Verification set results:				
Shark	0,59%	0,13%	15,72%	16,12%
Poznan CarPark	0,77%	0,04%	27,63%	22,18%

7. CONCLUSSIONS

In this paper, a novel approaches to estimation of optimized $QD(QP)$ relation has been presented. First, it is shown that with the use of the proposed algorithmic method, the number of coding passes required for finding the optimal QP-QD pairs can be vastly reduced by a factor of about 20.

Also, bitrate reductions attained with the use of algorithmically optimized QP-QD pairs have been tested in MVC+D and 3D-AVC. It has been shown that in case of half depth resolution coding, usage of optimized QP-QD pairs provides about 1.5% of bitrate reduction (measured with Bjøntegaard metric) and in full depth resolution the bitrate reduction is about 8%, in both cases relatively to the usage of straight-forward reference $QD = QP$ setting.

To cater for needs of low-latency encoding systems, a mathematical model of $QD(QP)$ relation has been proposed which allows quality-bitrate control in MVC+D/3D-AVC codecs without performing the proposed algorithmic QP-QD optimization. It has been shown, that although the proposed model has slightly lower performance than the fully-trained algorithm, the difference is relatively small – about 0.2~0.75 p.p. of loss in bitrate reduction.

The paper also reports a performance comparison of coding with full and half depth resolution, in both MVC+D and 3D-AVC. It can be generalized that when optimized QP-QD values are used (independently from the exact proposal) in both variants of AVC-based MVD video coding, it is about 7% more efficient to use half depth resolution coding than to use full depth resolution.

Both the proposals for finding optimized $QD(QP)$ relation, the algorithm and the model, can be easily extended to the context of HEVC-based 3D or other coding technologies, which will be the aim of further works.

8. ACKNOWLEDGEMENTS

Research project was supported by National Science Centre, Poland, according to the decision DEC-2012/05/B/ST7/01279.

O. Stankiewicz and K. Wegner are scholarship holders of Scholarship support for Ph.D. students specializing in majors strategic for Wielkopolska's development, Sub-measure 8.2.2 Human Capital Operational Programme, co-financed by EU under the European Social Fund.

REFERENCES

[1] ISO/IEC 14496-10, Int. Standard “Generic coding of audio-visual objects – Part 10: Advanced Video Coding”, 7th Ed., 2012, also: ITU-T Rec. H.264, Edition 7.0, 2012.

[2] “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., 2013, also: ITU-T Rec. H.264, Edition 8.0, 2013.

[3] M. Hannuksela, Y. Chen, T. Suzuki, J.-R. Ohm, G. Sullivan, „Text of ISO/IEC 14496-10:2012/PDAM3 AVC compatible video-plus-depth extension”, *ISO/IEC JTC1/SC29/WG11 MPEG*, Doc. N13142, Shanghai, Oct. 2012.

[4] G. Zhu, G. Jiang, M. Yu, F. Li, F. Shao, Z. Peng, "Joint video/depth bit allocation for 3D video coding based on distortion of synthesized view", *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (BMSB), 2012, 27-29 June 2012.

[5] Y. Zhang, S. Kwong, L. Xu, S. Hu, G. Jiang, C.-C.J. Kuo, "Regional Bit Allocation and Rate Distortion Optimization for Multiview Depth Video Coding With View Synthesis Distortion Model", *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3497-3512, Sept. 2013.

[6] 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.

[7] H. Yuan, Y. Chang, J. Huo, F. Yang, Zh. 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.

[8] F. Shao, G. Jiang, W. Lin, M. Yu, Q. Dai, "Joint Bit Allocation and Rate Control for Coding Multi-View Video Plus Depth Based 3D Video", *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1843-1854, Dec. 2013.

[9] K. Klimaszewski, K. Wegner, M. Domański, "Video and Depth Bitrate Allocation in Multiview Compression", accepted for *International Conference on Systems, Signals and Image Processing* (IWSSIP) 2014, May 2014.

[10] D. Rusanovskyy, K. Müller, A. Vetro, "Common Test Conditions of 3DV Core Experiments", *Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP3 and ISO/IEC JTC 1/SC 29/WG 11:* Doc. JCT3V-D1100, Incheon, South Korea, 20–26 April 2013.

[11] 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, Xian, China, October 2009.

[12] Yo-Sung Ho, E.-K. Lee, C. Lee, “Video Test Sequence and Camera Parameters”, *ISO/IEC JTC1/SC29/WG11 MPEG*, Doc. M15419, Archamps, France, April 2008.

[13] J. Zhang, Ri Li, H. Li, D. Rusanovskyy, M. M. Hannuksela, “Ghost Town Fly 3DV sequence for purposes of 3DV standardization”, *ISO/IEC JTC1/SC29/WG11 MPEG*, Doc. M20027, Geneva, Switzerland, March 2011.

[14] D. Rusanovskyy, P. Aflaki, M. M. Hannuksela, “Undo Dancer 3DV sequence for purposes of 3DV standardization”, *ISO/IEC JTC1/SC29/WG11 MPEG,* Doc. M20028, Geneva, Switzerland, March 2011.

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

[16] G. Bjøntegaard, “Calculation of Average PSNR Differences between RD-curves”, *ITU-T SG16,* Doc. VCEG-M33, April 2001.

[17] National Institute of Information and Communications Technology (NICT), “NICT-3D CG Test Sequences (Shark and MicroWorld)”, *ftp://ftp.merl.com/pub/tian/NICT-3D/Shark/*