The Influence of Coding Tools on Immersive Video Coding

Jakub Szekiełda Adrian Dziembowski Dawid Mieloch

adrian.dziembowski@put.poznan.pl

Institute of Multimedia Telecommunications, Poznań University of Technology
Polanka 3, 61-131 Poznań, Poland

ABSTRACT

This paper summarizes the research on the influence of HEVC (High Efficiency Video Coding) configuration on immersive video coding. The research was focused on the newest MPEG standard for immersive video compression – MIV (MPEG Immersive Video). The MIV standard is used as a preprocessing step before the typical video compression thus is agnostic to the video codec. Uncommon characteristics of videos produced by MIV causes, that the typical configuration of the video encoder (optimized for compression of natural sequences) is not optimal for such content. The experimental results prove, that the performance of video compression for immersive video can be significantly increased when selected coding tools are being used.

Keywords

Immersive video coding, immersive video systems, video compression.

1. INTRODUCTION

Many modern multimedia systems that include steps in which the video compression is applied can be described as codec-agnostic. It means that any video codec can be utilized, so the selection of the compression method is completely transparent to the rest of the system. Obvious examples of codec-agnostic video systems and methods are streaming-related methods (e.g., MPEG Dynamic Adaptive Streaming over HTTP [Sod11]) or simple simulcast compression required, e.g., in surveillance or free-viewpoint television systems [Sta18].

Besides these applications, recently a new trend of using existing compression methods as an internal processing tool in new video codecs can be seen. The latest examples are MPEG-5 LCEVC (Low Complexity Enhancement Video Coding) [Mea20] that introduces an enhancement layer which, when combined with a base video encoded with another existing video codec, produces an enhanced video stream, or VPCC (Video-based Point Cloud Coding) [Gra20] or MIV (MPEG Immersive Video) [Boy21] that utilize video compression for dynamic three-dimensional scenes and objects. In immersive video, user can change the viewpoint and is not limited to watch views acquired by cameras located around a scene. While the use of existing state-of-the-art compression methods makes it easier to develop new codecs for more and more new emerging technologies, configuration of internal coder is often not optimized for these applications, as the default configuration usually provides already satisfactory results.

This paper describes a different approach, in which adaptation of the internal coder configuration is performed, while the external one is not changed. The proposed experiments focus on the influence of HEVC configuration on immersive video coding performed by MIV. In MIV, some views (base views) are fully transmitted, while for others (additional views) only the non-redundant information is included in atlases – synthetic videos containing information from many input views (as a mosaic of patches – Fig. 2).

The content of atlases highly varies from the typical video sequences, motivating the need for using a non-standard set of coding tools and testing their configuration.

2. EXPERIMENTS METHODOLOGY

Overview

In the experiments (Fig. 1), the input views together with depth maps were processed in the TMIV (Test Model for MPEG Immersive Video) [MPEG21a] encoder. It outputs 4 atlases: 2 containing texture information (called T0 and T1) and 2 containing depth information with reduced resolution (G0 and G1). Then, each atlas was separately encoded with x265 video encoder [X265] with 5 QP values: 22, 27, 32, 37, and 42 for texture and 4, 7, 11, 15, and 20 for depth.

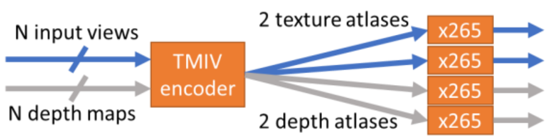

Figure 1. Scheme of an experiment.

The TMIV encoder works in two major configurations: “MIV Atlas” (A17, presented earlier) and “MIV View” (V17, Fig. 3). In the MIV View configuration, a subset of input views is transmitted within atlases. The remaining views are completely skipped thus some information is ignored (e.g., for Group sequence, only 8 of 21 input views are transmitted). We tested both configurations, thus each experiment was run twice.

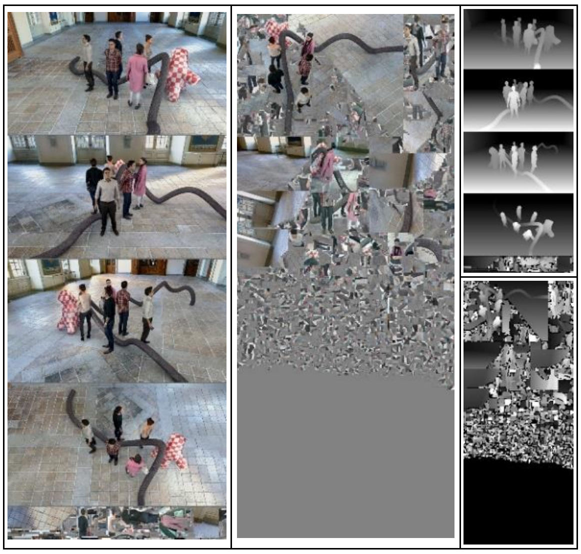

Figure 2. Atlases for Group sequence, MIV Atlas (A17). From left: atlas T0, T1, G0 (top), and G1.

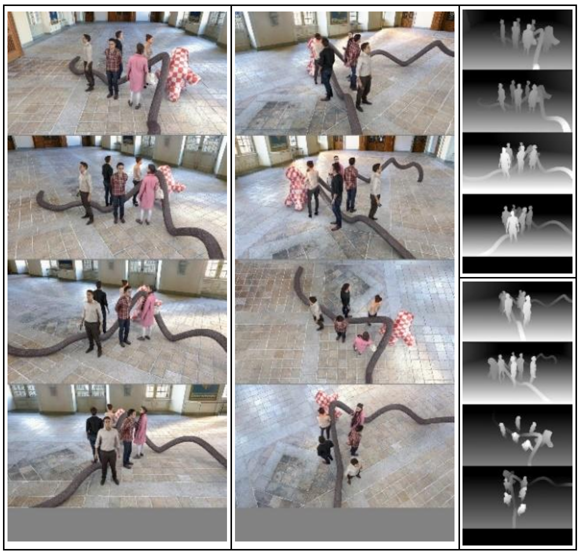

Figure 3. Atlases for Group sequence, MIV View (V17). From left: atlas T0, T1, G0 (top), and G1.

To preserve the readability of the paper, results obtained for all 5 QP values and all 14 test sequences were averaged.

We have decided to use the x265 encoder [X265] because of two main reasons. At first, it allows to flexibly change numerous encoding parameters, so we can easily analyze, which aspects of video encoding influence the immersive video encoding the most. Secondly, x265 is a very fast encoder, it is two orders of magnitude faster than HEVC Test Model (HM) [MPEG17] – the HEVC Test Model [Sze20].

In the experiments, 13 encoding parameters were tested. For some of them, several tests were performed resulting in 22 experiments in total. We have tested parameters, which potentially could improve the encoding efficiency of the immersive video:

1. b-adapt – flexibility of setting the GOP (group of pictures) structure,
2. bframes – maximum number of consecutive B-frames (bidirectional-predicted frames),
3. bframe-bias – probability of choosing B-frames,
4. lookahead-slices – number of threads used for frame cost calculation,
5. max-merge – maximum number of neighboring blocks analyzed in motion prediction,
6. me – motion search method (method of searching of corresponding blocks in previously-encoded frames),
7. no-early-skip – additional analysis of possible modes providing better quality in increased time,
8. rd – rate-distortion optimization level,
9. rdoq-level – level of rate-distortion analysis within quantization step,
10. rect – rectangular motion partitioning,
11. rect amp – rectangular motion partitioning with the possibility of asymmetric partitioning,
12. ref – number of L0 references,
13. subme – subpixel refinement level.

#	x265 parameter	default value	tested value	other changes
1	--b-adapt	2	0	
2			1	
3	--bframes	4	16	
4	--b-adapt	2	0	--bframes 16
5			1	
6	--bframe-bias	0	100	
7	--lookahead-slices	8	1	
8			4	
9	--max-merge	2	3	
10	--me	hex	dia	
11			umh	
12			star	
13	--no-early-skip	disabled	enabled	
14	--rd	3	1	
15			2	
16			5	
17	--rdoq-level	0	2	
18	--rect	disabled	enabled	--limit-modes
19	--rect --amp	disabled	enabled	
20	--ref	3	4	
21	--subme	2	3	
22	*optimal configuration*			

Table 1. Performed experiments.

Test sequences

The test set contained 14 miscellaneous high-resolution test sequences, including 7 synthetic, computer-generated sequences (Fig. 5):

1. ClassroomVideo, 16 full-360° 4K×2K cameras [Kro18],
2. Museum, 24 semispherical 2K×2K cameras placed on the sphere [Dor18],
3. Hijack, 10 parallel 4K×2K cameras with angle of view 180°×90° [Dor18].
4. Chess, 10 semispherical 2K×2K cameras placed on the sphere [Ilo19],
5. ChessPieces, 10 semispherical 2K×2K cameras placed on the sphere [Ilo20],
6. Kitchen, 25 perspective FullHD cameras placed in the 5×5 matrix [Boi18],
7. Fan, 15 perspective FullHD cameras placed in the 5×3 matrix [Dor20a],
8. Group, 21 perspective FullHD cameras placed on a sphere [Dor20b],

and 6 natural sequences, captured by real multicamera systems containing different numbers of perspective cameras (Fig. 4):

1. Fencing, 10 FullHD cameras placed on an arc [Dom16],
2. Carpark, 9 FullHD cameras placed linearly [Mie20],
3. Street, 9 FullHD cameras placed linearly [Mie20],
4. Hall, 9 FullHD cameras placed linearly [Mie20],
5. Frog, 13 FullHD cameras placed linearly [Sal18],
6. Painter, 16 cameras with resolution 2048×1088 placed in the 4×4 matrix [Doy17].

All sequences are commonly used in immersive video applications, e.g., within ISO/IEC JTC1/SC29/WG04 MPEG Video Coding group [MPEG21]. Each sequence has 17 frames.

Figure 4. Natural sequences. Left column: Frog, Hall, Painter; right: Fencing, Carpark, Street.

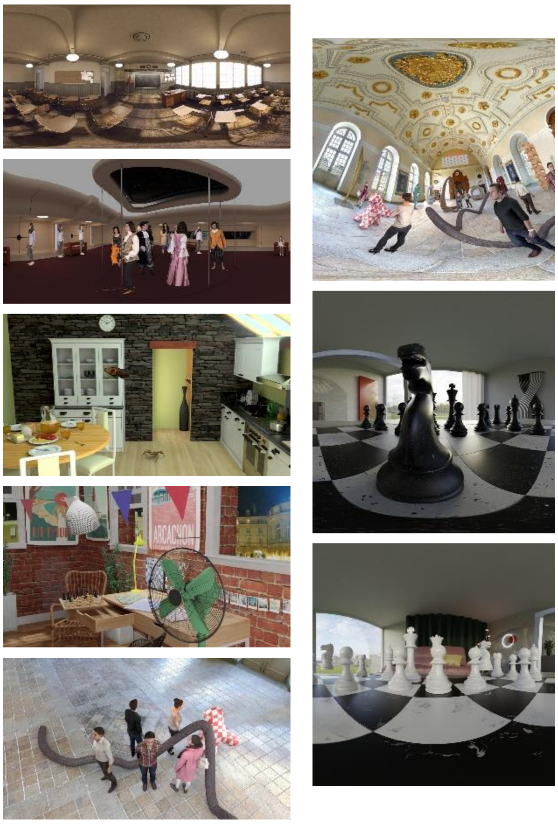

Figure 5. Computer-generated sequences. Left column: ClassroomVideo, Hijack, Kitchen, Fan, and Group; right: Museum, Chess, ChessPieces.

3. EXPERIMENTAL RESULTS

In total, 22 experiments were performed. In each, only one parameter was changed, while other parameters were set to default values. Each configuration was compared to the default configuration of the x265 encoder. Results of the encoding using default configuration are presented in Tables 2 and 3.

As mentioned earlier, all the results presented in Section 3 were averaged over 5 QP values and 14 sequences.

default configuration	A17		V17	
	bitrate [kbps]	PSNR [dB]	bitrate [kbps]	PSNR [dB]
avg G0	8583	69.86	8563	69.82
avg G1	10131	70.82	10174	70.76

Table 2. Encoding of depth atlases.

default configuration	A17		V17	
	bitrate [kbps]	PSNR [dB]	bitrate [kbps]	PSNR [dB]
avg T0	17781	42.36	17352	42.33
avg T1	10588	45.22	10531	45.26

Table 3. Encoding of texture atlases.

In consecutive subsections, only the most beneficial experiments are described. Results of all performed experiments are included at the end of the section.

Number of B-frames

In experiment #3, the influence of maximum number of consecutive B-frames was tested. By default, this number is set to 4. We have tested 16 B-frames, so the whole tested sequence was treated as one GOP.

bframes 16	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-0.22%	-0.06%	-0.21%	-0.07%
avg G1	0.42%	-0.07%	-0.21%	-0.05%

Table 4. Encoding of depth atlases, --bframes 16 (default: 4).

bframes 16	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-2.41%	-0.06%	-2.13%	-0.07%
avg T1	-0.54%	0.10%	-1.88%	-0.07%

Table 5. Encoding of texture atlases, --bframes 16 (default: 4).

The results obtained for texture and depth atlases are different – an increase of the number of B-frames allows to significantly decrease the bitrate of texture atlases, while for depth the difference is much smaller.

In the immersive video, the cameras are usually stationary, as the change of viewing position is made by the final viewer himself. Therefore, increasing the B-frames number allows to increase the coding efficiency of the video. This assumption is true both for textures and perfect (usually computer-generated) depth maps. The depth maps estimated for natural content are often not stable, thus a smaller number of B-frames is a better solution in this case.

Moreover, for the A17 configuration of the TMIV, the coding gain for atlases T1 and G1 is much smaller than for atlases T0 and G0. For V17 such a difference cannot be spotted.

The reason of this phenomenon is the different characteristics of the first and second atlas. The first atlas in the A17 configuration contains full base views while the second one contains many smaller patches, so it is much less temporally consistent. In the V17 configuration, both atlases contain similar information – full views.

Motion search method

In experiments #10, #11, and #12, different methods of motion search were tested. The default method in the x265 decoder is “hex” – hexagon-shaped search. We have tested three other methods: “dia” – diamond search, “umh” – uneven multi-hexagon, and “star” with 8 directions of searching. An exhaustive search (“full”) was not tested because it is ridiculously slow and impractical.

Diamond search performs worse than the default hexagonal search (see Tables 20 and 21). More sophisticated search methods (“umh” and “star”) allow to increase coding efficiency, especially for depth atlases (Tables 6 and 8).

For texture atlases, the coding gain is smaller but noticeable. It should be noted, that both search methods perform better for atlas T1 than for T0.

The “hex” method is adapted for analyzing the natural videos (i.e., textures), while it performs worse for depth maps. In uneven multi-hexagon and star methods, more prediction directions are being analyzed, allowing to better adapt to different characteristics of depth video (large smooth areas, sharp edges).

me umh	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-1.53%	0.03%	-1.57%	0.03%
avg G1	-1.59%	0.04%	-1.50%	0.04%

Table 6. Encoding of depth atlases, --me umh (default: hex).

me umh	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-0.14%	0.01%	-0.13%	0.01%
avg T1	-0.46%	0.05%	-0.10%	0.00%

Table 7. Encoding of texture atlases, --me umh (default: hex).

me star	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-2.78%	0.06%	-2.91%	0.05%
avg G1	-2.44%	0.07%	-2.79%	0.06%

Table 8. Encoding of depth atlases, --me star (default: hex).

me star	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-0.19%	0.01%	-0.18%	0.01%
avg T1	-0.62%	0.06%	-0.14%	0.01%

Table 9. Encoding of texture atlases, --me star (default: hex).

RDO level

In experiments #14, #15, and #16, the different levels of the rate-distortion optimization were tested. In x265, the level of the RDO may vary from 1 to 5, where 5 is the full RDO analysis and 1 is the least exhaustive optimization. By default, RDO level is set to 3.

We have tested three levels of RDO: 1, 2, and 5 (4 was omitted as it produces exactly the same results as 3).

Setting the RDO level to 1 or 2 speeds up the computations, but it greatly decreases the coding efficiency (see Tables 20 and 21).

The results of using full RDO are presented in Tables 10 and 11. For encoding of texture atlases, full RDO not only increases the computational time but also decreases the coding efficiency. The opposite results were obtained for depth atlases, where full RDO allows to decrease the bitrate by more than 14%.

The video encoders are adapted for natural video encoding, while, as mentioned earlier, depth maps have significantly different characteristics. The encoding tools, including partial rate-distortion optimization, are optimized for texture encoding. When the full RDO process is performed, these texture-aimed simplifications are disabled.

rd5	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-14.06%	0.74%	-14.15%	0.70%
avg G1	-14.09%	0.94%	-13.83%	0.75%

Table 10. Encoding of depth atlases, --rd 5 (default: 3).

rd5	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	1.64%	0.22%	1.21%	0.22%
avg T1	0.72%	0.12%	1.25%	0.23%

Table 11. Encoding of texture atlases, --rd 5 (default: 3).

Rectangular motion partitioning

In experiments #18 and #19, the influence of the rectangular motion partitioning was tested. By default, x265 uses square partitioning only. In order to keep the short computational time, the --limit-modes parameter was used. This parameter limits number of modes analyzed for each CU, significantly decreasing the computational time while having a slight impact on coding efficiency.

In #18, only the symmetric motion partitions were permitted, thus the block can be split into two N×2N or 2N×N blocks.

The results gathered in Tables 12 and 13 indicate, that enabling the rectangular motion partitioning allows increasing coding efficiency, both for depth and texture atlases. However, the bitrate reduction achieved for depth atlases is significantly higher.

As mentioned earlier, depth video contains large smooth regions that often include only constant depth values. These regions are separated by sharp edges and many of them are horizontal or vertical. Enabling of the rectangular partitioning allows to better fit the CU grid to objects in the depth video thus significantly increase the coding efficiency. For textures, rectangular partitioning also allows to increase the efficiency, but the gain is much smaller because of video characteristics (fewer smooth areas, highly textured objects, etc.).

In #19, also asymmetric partitioning was permitted, so each block can be split also in 25%/75% or 75%/25% proportion (both vertically and horizontally. As presented in Tables 14 and 15, it allows to additionally increase the coding efficiency, but the difference is slight.

rect	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-21.87%	0.54%	-22.44%	0.51%
avg G1	-22.71%	0.70%	-22.06%	0.53%

Table 12. Encoding of depth atlases, --rect --limit-modes.

rect	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-0.70%	0.05%	-0.75%	0.05%
avg T1	-1.39%	-0.12%	-0.72%	0.05%

Table 13. Encoding of texture atlases, --rect --limit-modes.

rect amp	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-21.94%	0.54%	-22.50%	0.52%
avg G1	-22.71%	0.70%	-22.11%	0.53%

Table 14. Encoding of depth atlases, --rect --amp --limit-modes.

rect amp	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-0.76%	0.06%	-0.83%	0.06%
avg T1	-1.54%	0.02%	-0.79%	0.06%

Table 15. Encoding of texture atlases, --rect --amp --limit-modes.

Max number of L0 references

In experiment #20, the maximum number of L0 references was tested. HEVC specification allows up to 8 references, the default value in x265 is set to 3. In the experiment, 4 L0 references were tested.

As shown in Tables 16 and 17, the coding gain is slight, but noticeable for all the data types. In general, it decreases the bitrate of the atlases without having any impact on the quality.

ref 4	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-0.17%	0.00%	-0.17%	0.00%
avg G1	-0.15%	0.00%	-0.14%	0.00%

Table 16. Encoding of depth atlases, --ref 4 (default: 3).

ref 4	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-0.08%	0.00%	-0.08%	0.00%
avg T1	-0.13%	0.01%	-0.07%	0.00%

Table 17. Encoding of texture atlases, --ref 4 (default: 3).

Subpixel refinement level

In experiment #21, the number of subpixel refinements was changed. By default, it is set to 2. We have tested value 3, which increases the number of half-pel iterations and allows using the chroma residual in the motion estimation decisions.

As presented in Table 18, such a change allows reducing bitrate of depth by 0.3% without quality change. For texture atlases, increasing the subpixel refinement level slightly increases the bitrate.

subme 3	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-0.32%	0.01%	-0.31%	0.01%
avg G1	-0.28%	0.01%	-0.27%	0.01%

Table 18. Encoding of depth atlases, --subme 3 (default: 2).

subme 3	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	0.05%	0.00%	0.05%	0.00%
avg T1	0.20%	0.03%	0.06%	0.00%

Table 19. Encoding of texture atlases, --subme 3 (default: 2).

All tests

Parameter	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
b-adapt 0	1.12%	-0.04%	1.47%	-0.04%
b-adapt 1	2.06%	0.06%	2.43%	0.03%
bframes 16	0.13%	-0.07%	-0.21%	-0.06%
b-adapt 0 bframes 16	5.74%	-0.32%	5.96%	-0.30%
b-adapt 1 bframes 16	2.00%	0.05%	2.35%	0.02%
bframes bias	1.73%	-0.19%	1.26%	-0.19%
lookahead-slices 1	-0.01%	0.00%	0.00%	0.00%
lookahead-slices 4	-0.01%	0.00%	-0.01%	0.00%
max-merge	0.00%	0.00%	0.00%	0.00%
me dia	0.77%	-0.02%	1.05%	-0.02%
me umh	-1.56%	0.04%	-1.54%	0.03%
me star	-2.59%	0.06%	-2.85%	0.06%
no-early skip	0.37%	0.13%	0.58%	0.17%
rd 1	53.04%	-1.40%	50.88%	-1.24%
rd 2	20.01%	-0.89%	21.96%	-0.86%
rd 5	-14.08%	0.84%	-14.00%	0.73%
rdoq-level 2	1.60%	0.85%	1.87%	0.59%
rect limit-modes	-22.32%	0.62%	-22.27%	0.52%
rect amp limit-modes	-22.36%	0.62%	-22.32%	0.53%
ref 4	-0.16%	0.00%	-0.16%	0.00%
subme 3	-0.30%	0.01%	-0.29%	0.01%

Table 20. Encoding of depth atlases.

Other tested parameters decrease the coding efficiency, or their impact is negligible. In Tables 20 and 21, the results for all tested parameters are presented. To preserve the readability of Table 20, values for G0 and G1 were averaged. For the same reason, we have averaged values for T0 and T1 in Table 21.

Parameter	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
b-adapt 0	0.84%	-0.58%	0.70%	-0.03%
b-adapt 1	5.80%	0.23%	6.49%	0.06%
bframes 16	-1.71%	0.02%	-2.01%	-0.07%
b-adapt 0 bframes 16	3.75%	-0.54%	2.12%	-0.29%
b-adapt 1 bframes 16	5.78%	-0.14%	6.48%	0.06%
bframes bias	-0.50%	-0.21%	0.03%	-0.23%
lookahead-slices 1	0.00%	0.00%	0.00%	0.00%
lookahead-slices 4	-0.02%	0.00%	0.00%	0.00%
max-merge	0.00%	0.00%	0.00%	0.00%
me dia	0.03%	0.00%	0.01%	0.00%
me umh	-0.26%	0.03%	-0.12%	0.00%
me star	-0.35%	0.04%	-0.16%	0.01%
no-early skip	0.74%	0.17%	0.66%	0.17%
rd 1	14.36%	-0.69%	8.44%	-0.36%
rd 2	1.89%	-0.26%	0.44%	-0.08%
rd 5	1.30%	0.17%	1.23%	0.22%
rdoq-level 2	3.60%	-0.14%	3.04%	-0.38%
rect limit-modes	-0.96%	-0.04%	-0.73%	0.05%
rect amp limit-modes	-1.05%	0.04%	-0.81%	0.06%
ref 4	-0.10%	0.00%	-0.08%	0.00%
subme 3	0.11%	0.02%	0.05%	0.00%

Table 21. Encoding of texture atlases.

Optimal combination of parameters

The analysis of the results of 21 experiments allowed to find the set of optimal parameters for encoding the immersive video.

In the last, 22nd experiment, the encoder was configured in order to provide the best encoding efficiency. For depth atlases, we have used: --me star --rd 5 --rect --amp --limit-modes --ref 4 --subme 3. For texture encoding: --bframes 16 --me star --rect --amp --limit-modes --ref 4.

The results are presented in tables 22 and 23.

optimal configuration	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg G0	-27.57%	1.01%	-28.07%	0.96%
avg G1	-30.27%	1.26%	-27.50%	1.00%

Table 22. Encoding of depth atlases, --me star --rd 5 --rect --amp --limit-modes --ref 4 --subme 3.

optimal configuration	A17		V17	
	dBitrate	dPSNR	dBitrate	dPSNR
avg T0	-3.41%	0.01%	-3.15%	-0.01%
avg T1	-2.74%	0.11%	-2.86%	0.00%

Table 23. Encoding of texture atlases, --me star --bframes 16 --rect --amp --limit-modes --ref 4.

4. CONCLUSIONS

The paper presents the research on the influence of HEVC configuration on immersive video coding. MPEG Immersive Video is agnostic to the internally used codec, therefore, the proposed experiments tested a set of coding tools in order to find the optimal configuration, i.e., adapted to the characteristics of immersive video.

The experimental results showed that the configuration of the internal codec (one of the available HEVC implementations was tested – x265) has a significant impact on the overall coding efficiency, confirming the importance of performing the proposed optimization. The proper configuration allowed to reduce the bitstream on average by 3% for textures and even by 30% for depth maps encoding.

Moreover, the noticed differences in encoding of textures and depth maps draw interesting general conclusions that can be utilized not only in the coding of immersive videos but can be used in future research on coding adapted to typical characteristics of depth maps.

5. ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of Republic of Poland.

6. REFERENCES

[Boi18] Boissonade, P., and Jung, J. [MPEG-I Visual] Proposition of new sequences for Windowed-6DoF experiments on compression, synthesis, and depth estimation. ISO/IEC JTC1/SC29/WG11 MPEG, M43318, Ljubljana, Slovenia, 2018.

[Boy21] Boyce, J., Doré, R., Dziembowski, A., Fleureau, J., Jung, J., Kroon, B., Salahieh, B., Malamal Vadakital, V.K., and Yu, L. MPEG Immersive Video Coding Standard. Proceedings of the IEEE, pp. 1-16, Early Access, 2021.

[Dom16] Domański, M., Dziembowski, A., Grzelka, A., Łuczak, A., Mieloch, D., Stankiewicz, O., and Wegner, K. Multiview test video sequences for free navigation exploration obtained using pairs of cameras. ISO/IEC JTC1/SC29/WG11 MPEG, M38247, Geneva, Switzerland, 2016.

[Dor18] Doré, R. Technicolor 3DoF+ test materials. ISO/IEC JTC1/SC29/WG11 MPEG, M42349, San Diego, CA, USA, 04.2018.

[Dor20a] Doré, R., Briand, G., and Thudor, F. [MPEG-I Visual] InterdigitalFan0 content proposal for MIV. ISO/IEC JTC1/SC29/WG11 MPEG, M54732, Online, 07.2020.

[Dor20b] Doré, R., Briand, G., and Thudor, F. [MPEG-I Visual] InterdigitalGroup content proposal. ISO/IEC JTC1/SC29/WG11 MPEG, M54731, Online, 07.2020.

[Doy17] Doyen, D., Langlois, T., Vandame, B., Babon, F., Boisson, G., Sabater, N., Gendrot, R., and Schubert, A. Light field content from 16-camera rig. ISO/IEC JTC1/SC29/WG11 MPEG, M40010, Geneva, Switzerland, 2017.

[Gra20] Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., and Tabatabai, A. An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). APSIPA Transactions on Signal and Information Processing, 9, E13, 2020.

[Ilo19] Ilola, L., Vadakital, V.K.M., Roimela, K., and Keraenen, J. New test content for immersive video – Nokia Chess. ISO/IEC JTC1/SC29/WG11 MPEG, M50787, Geneva, Switzerland, 10.2019.

[Ilo20] Ilola, L., and Vadakital, V. [MPEG-I Visual] [MIV] Improved NokiaChess sequence. ISO/IEC JTC1/SC29/WG11 MPEG, M54382, 2020.

[Kro18] Kroon, B. 3DoF+ test sequence ClassroomVideo. ISO/IEC JTC1/SC29/WG11 MPEG, M42415, San Diego, CA, USA, 04.2018.

[Mea20] Meardi, G., Ferrara, S., Ciccarelli, L., Cobianchi, G., Poularakis, S., Maurer, F., Battista, S., and Byagowi, A. MPEG-5 Part 2: Low Complexity Enhancement Video Coding (LCEVC): Overview and performance evaluation. Proc. of SPIE 11510, Applications of Digital Image Processing XLIII, 115101C, 08.2020.

[Mie20] Mieloch, D., Dziembowski, A., and Domański, M. [MPEG-I Visual] Natural outdoor test sequences. ISO/IEC JTC1/SC29/WG11 MPEG, M51598, Brussels, Belgium, 01.2020.

[MPEG17] High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description Update 9, ISO/IEC JTC1/SC29 /WG11 MPEG/N17047, July 2017, Torino, Italy.

[MPEG21] Common Test Conditions for MPEG Immersive Video. ISO/IEC JTC1/SC29/WG04 MPEG VC, N0051, Online, Jan. 2021.

[MPEG21a] Test Model 8 for MPEG Immersive Video. ISO/IEC JTC1/SC29/WG04 MPEG VC, N0050, Online, Jan. 2021.

[Sal18] Salahieh, B., et al, J. Kermit test sequence for Windowed 6DoF Activities. ISO/IEC JTC1/SC29/WG11 MPEG, M43748, Ljubljana, Slovenia, 2018.

[Sod11] Sodagar, I. The MPEG-DASH Standard for Multimedia Streaming Over the Internet. IEEE MultiMedia, vol. 18, no. 4, pp. 62-67, April 2011.

[Sta18] Stankiewicz, O., Domański, M., Dziembowski, A., Grzelka, A., Mieloch, D., and Samelak, J. A Free-viewpoint Television system for horizontal virtual navigation. IEEE Transactions on Multimedia, vol. 20, no. 8, pp. 2182-2195, 2018.

[Sze20] Szekiełda, J., Stankowski, J., Dziembowski, A., and Mieloch, D. On the computational time of MIV-related experiments. ISO/IEC JTC1/SC29 /WG11 MPEG, M54897, Online, 10.2020.

[X265] MulticoreWare Inc. x265 HEVC Encoder/ H.265 Video Codec. Available on http://x265.org