Efficient Immersive Video Compression using Screen Content Coding

Jarosław
Samelak

Adrian
Dziembowski

Dawid
Mieloch

Marek
Domański

Maciej
Wawrzyniak

adrian.dziembowski@put.poznan.pl

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

ABSTRACT

The paper deals with efficient compression of immersive video representations for the synthesis of video related to virtual viewports, i.e., to selected virtual viewer positions and selected virtual directions of watching. The goal is to obtain possibly high quality of virtual video obtained from compressed representations of immersive video acquired from multiple omnidirectional and planar (perspective) cameras, or from computer animation. In the paper, we describe a solution based on HEVC (High Efficiency Video Coding) compression and the recently proposed MPEG Test Model for Immersive Video. The idea is to use standard-compliant Screen Content Coding tools that were proposed for other applications and have never been used for immersive video compression. The experimental results with standard test video sequences are reported for the normalized experimental conditions defined by MPEG. In the paper, it is demonstrated that the proposed solution yields up to 20% of bitrate reduction for the constant quality of virtual video.

Keywords

Video compression, video codecs, virtual reality.

1 INTRODUCTION

The recent development of virtual reality applications raises rapidly growing research interests in immersive video [Isg14]. In particular, substantial efforts are made in virtual view synthesis [Ceu18], [Yua18], [Rah18], [Zhu19], virtual navigation and free-viewpoint television [Tan12], [Sta18], [Cha19]. Recently, image-based rendering of virtual views became widely applicable for head-mounted devices and other displays suitable for VR content. The content may be computer-generated or it may be acquired from multiple omnidirectional and perspective (planar) cameras. Such visual content constitutes an immersive video that may have various representations. Recently, great interest is attained by point clouds [Cui19], [Zha20], [Li20], [Sch19], but the representation that is most often used in research is multiview video plus depth (MVD) [Mue11]. Therefore, this paper is focused on multiview video plus depth representations of immersive video. For such representations, depth has to be estimated, and a lot of work has

already been done for depth estimation in the above-mentioned applications, e.g. [Mie20]. Once the representation is estimated, the representation of immersive video needs to be compressed before transmission (cf. Fig. 1).

Obviously, the compression artifacts deteriorate the fidelity of view synthesis. Therefore, in the paper, we consider immersive video compression and the influence of the compression on the quality of the virtual video rendered from compressed data. Moreover, we propose an alternative approach to immersive video compression, and we demonstrate the advantages of this alternative approach. In particular, we demonstrate that our approach results in a reduced bitrate for the same quality of virtual views, i.e., for a constant bitrate, the proposed approach results in the improved quality of the synthesized virtual views as compared to the approaches from [Dom19], [Fle19], [Laf19], and [Wie19].

2 IMMERSIVE VIDEO COMPRESSION

A multiview representation of immersive video may consist of multiple perspective (planar, 2D) views with vastly overlapping fields of views, or it may consist of a few overlapping 360-degree videos. The compression of immersive video takes advantage of the inter-view
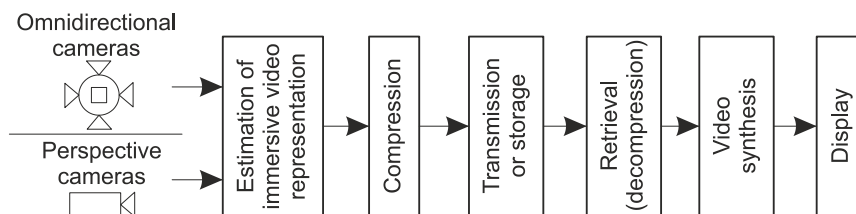
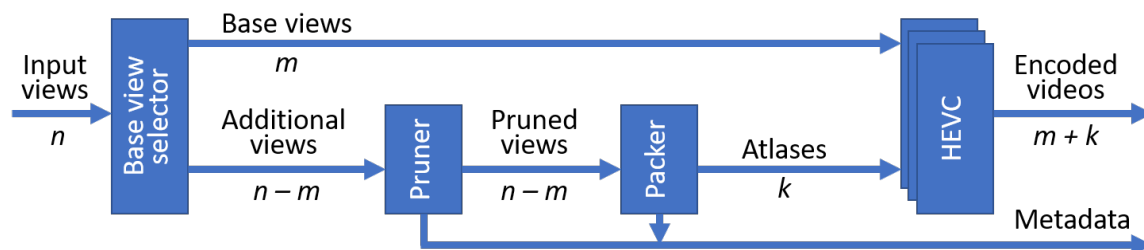**Figure 1: Data flow in immersive video systems.**

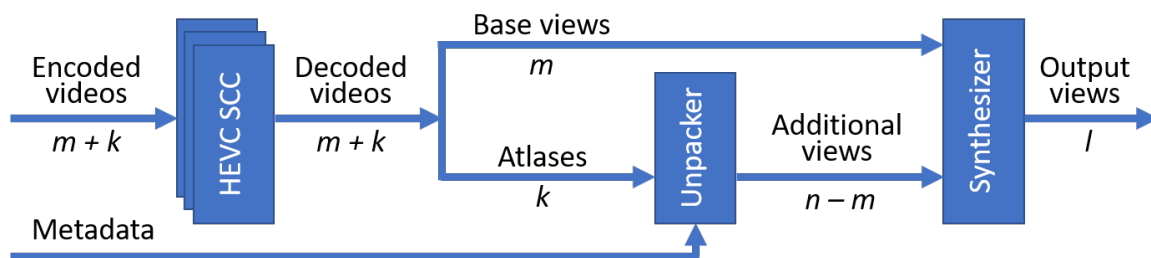**Figure 2: MPEG immersive video encoder (TMIV framework).**

Figure 3: MPEG immersive video decoder (TMIV framework).

redundancy existing in the input multiview representation. Removal of this redundancy will result in decreasing the amount of data required to fully represent the whole three-dimensional scene.

One of the possible scenarios assumes the compression of MVD representation using a standard 3D-HEVC video encoder. Its coding techniques use inter-view prediction based on depth maps and statistical dependencies between views and corresponding depth maps. The use of this encoder reduces the required bitrate by up to 50% in comparison with simulcast HEVC [Tec16], which encodes each view and each depth map separately. Other works focus more on the reduction of pixel rate, i.e., the number of pixels that have to be sent during the transmission. An interesting technique described in [Gar19] proposes a decoder-side reconstruction of depth maps using views compressed using simulcast HEVC or MV-HEVC. This solution provides a 50% reduction of the pixel rate (because depth maps do not have to be sent) and up to a 35% reduction of the required bitrate while preserving similar quality of the video.

The state-of-the-art technology for immersive video compression is being developed by ISO/IEC MPEG group [ISO19e]. The MPEG Test Model for Immer-

sive Video (TMIV) is already publicly available as a descriptive and software framework for research [ISO19d], and in the next months, the works on this future video standard are planned to enter one of the final stages of preparation.

The forthcoming standard is built using the technologies presented by proponents in response to the Call for Proposals for 3DoF+ video coding [ISO19a]. Some proposals followed nearly the same basic idea that several base views gathering most of the information of the scene should be encoded in their entirety, while supplementary information (e.g., disocclusions from other views, Fig. 4) can be transmitted in the form of a mosaic of much smaller patches, that all together are grouped into atlases [Dom19], [Fle19]. The main idea of TMIV follows a similar scheme – see Fig. 2 and Fig. 3 for the overview.

First of all, n input views with depth maps are split into two groups: m base views and $n\text{-}m$ additional views. The pruner (cf. Fig. 4), basing on depth, identifies and extracts regions occluded in the base views. These occluded regions are left in additional views, while the rest of the regions are removed. It results in small patches left in the pruned additional views. The packer gathers patches from all additional views into k at-
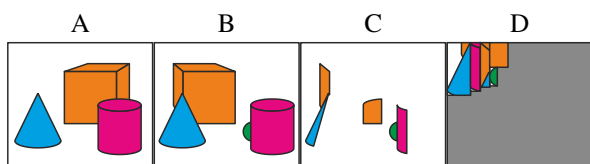
Figure 4: a) Base view, b) additional view, c) pruned additional view (preserved disocclusions), d) atlas.
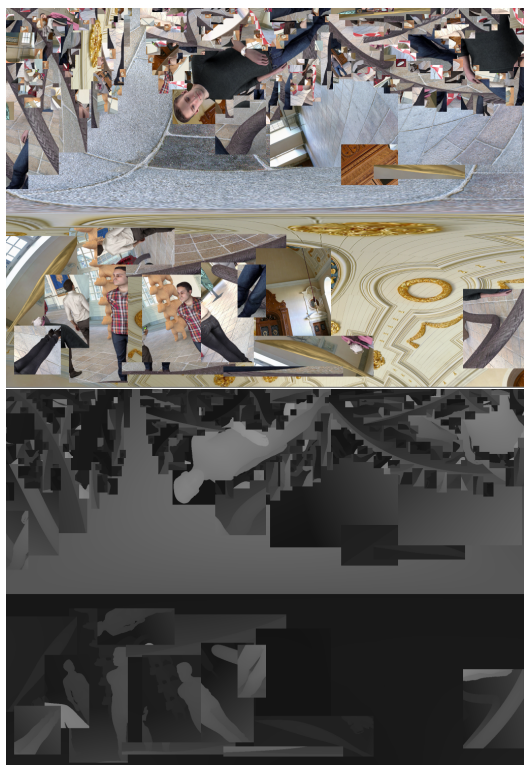

Figure 5: Example of an atlas with a corresponding depth map.

lases. In order to provide better encoding efficiency, the patches in atlases contain all information from their bounding box, as this decreases the number of sharp edges in the encoded atlas. A schematic example of pruning and packing is presented in Fig. 4. For example, an atlas for the TechnicolorMuseum [Dor18] test sequence is presented in Fig. 5. The number of atlases is usually much smaller than the number of additional views, ensuring the reduction of pixel rate, while still preserving the whole representation of the encoded three-dimensional scene. In the end, the base views and atlases are fed to simulcast HEVC encoders.

In the decoder, base views and patches from atlases, together with metadata that contain the initial positions of patches in input views, are used to synthesize l output views, which can be reconstructed input views, or any number of virtual views required by a user of the immersive video system (e.g., a stereopair for a virtual reality headset).

The development of an entirely new competitive compression technology would require huge costs and manpower. Therefore, in order to reduce the time needed to finish the works on the new coding technology and standard, the current approach is to build the immersive video coding technology on top of the general video compression technology [Dom19], [Fle19], [ISO19e].

The common feature of the above-mentioned coding technologies is the use of virtual view synthesis and the application of general video coding techniques like HEVC or even the application of 3D-HEVC that is the specialized coding technology for multiview plus depth video. In the following section, we propose the application of HEVC Screen Content Coding [Xu16b], the technique for computer-generated visual content, in order to increase the quality of virtual view synthesis performed on the compressed representation of the immersive video.

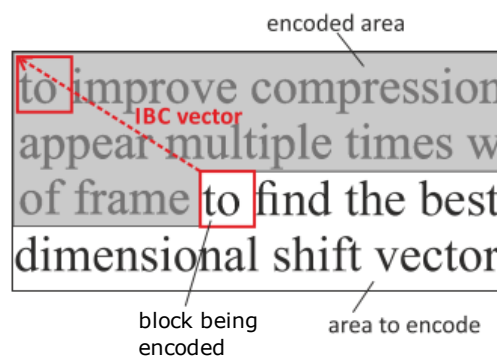## 3 NEW APPROACH TO COMPRESSION OF PATCH ATLASES

Figure 6: Operation of Intra Block Copy.

As mentioned before, for the efficient compression of patch atlases, the authors propose to use HEVC Screen Content Coding [Xu16b] instead of a standard video coding technology like HEVC [ISO15] or 3D-HEVC [Tec16]. Screen Content Coding is developed as an extension of HEVC, dedicated for the compression of computer-generated visual content, such as a remote keyboard, screen recordings or cloud gaming.

The basic tool used in HEVC-SCC is Intra Block Copy [Xu16a]. It is designed to improve the compression efficiency of fonts and other repetitive patterns that may appear multiple times within a single frame (cf. Fig. 6). The IBC tool searches the encoded part of the frame in order to find the best match for the unit being currently encoded. This search results in a two-dimensional shift vector with the components being integer multiples of the sampling periods (i.e., the horizontal and vertical sampling periods).

The idea to apply Intra Block Copy to the compression of camera-captured content was presented in [Sam17]
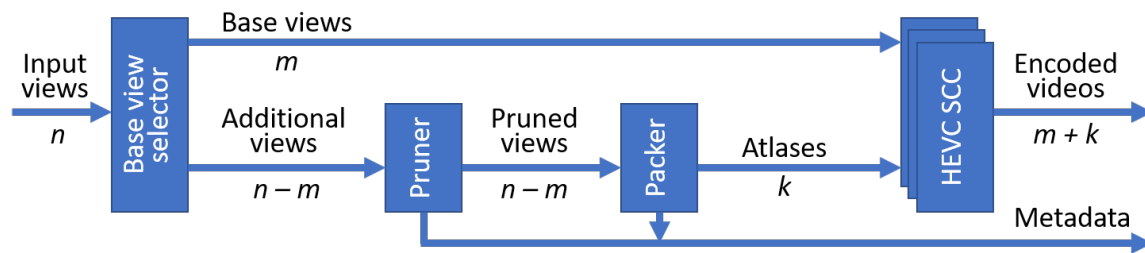
Figure 7: Proposed MPEG immersive video encoder with HEVC Screen Content Coding.

Sequence	Content type	Number of base views	Number of atlases
Classroom [Kro18]	O, CG	1	1
Museum [Dor18]	O, CG	2	2
Hijack [Dor18]	O, CG	1	2
Kitchen [Boi18]	P, CG	1	2
Painter [Doy17]	P, NC	1	4
Frog [Sal19]	P, NC	2	8
Fencing [Dom16]	P, NC	1	3

Table 1: Test sequences. O – omnidirectional, P – perspective, CG – computer generated, NC – natural content.

and [Sam19]. It was proven that IBC can be successfully used to exploit inter-view similarities in frame-compatible stereoscopic videos. The authors now propose to extend this idea onto the compression of patch atlases (Fig. 7). A single atlas often contains similar patches, located in distant parts of a frame. The IBC tool would be an ideal solution for efficient compression in such a case.

Other arguments in favor of using HEVC-SCC for the compression of patch atlases are additional SCC tools – Color Transform [Xu16b] and Palette Mode [Xu16c]. As presented in [Sam17], the influence of these tools on the compression efficiency of camera-captured content is negligible, however, they may provide a significant gain when applied to the compression of depth patch atlases. The results of using HEVC-SCC instead of HEVC are presented in the following section.

4 EXPERIMENTS AND RESULTS

4.1 Methodology of the experiments

The goal of the experiments is to demonstrate the usefulness and efficiency of the standard-compliant Screen Content Coding HEVC extension applied in immersive video coding. In order to present the advantages of such an approach, the recent MPEG Immersive Video encoder – TMIV [ISO19d] is used. The video data generated by TMIV is then encoded using HEVC-SCC. The results are compared to those obtained by the use of HEVC main profile.

The proposed approach is assessed using 7 miscellaneous test video sequences as described in Table 1. These sequences are commonly used in research and standardization activities on immersive video [ISO19b] because of their very diversified characteristics (natural and computer-generated content, omnidirectional and perspective cameras, different resolutions, etc.). For each sequence, 97 frames are used, which refers to 3 full groups of pictures (GOPs).

All common coding parameters (e.g. GOP size, Intra Period, max CU Width, Sample Addaptive Offset, etc.) are exactly the same for both encoders and the same as defined in MPEG recommendations for experiments on immersive video coding [ISO19b], [Yu15]. The same values of QP (Quantization Parameter) are set for both encoders: HEVC and HEVC-SCC. The ΔQP between depth and texture data is set to 10 in order to better preserve depth quality (e.g. when QP for texture was set to 22, QP for corresponding depth was set to 12; experiments were performed for 5 QP values – for texture: 22, 27, 32, 37, and 42), which is crucial for proper view synthesis.

In Section 4.2, the results of encoding of atlases are presented. For each sequence, the bitrate was calculated as a sum of bitrates for all atlases.

The quality (the average difference between atlases before and after encoding) was calculated as the average PSNR of all atlases. The texture and depth atlases are discussed separately.

In Section 4.3 the results of the virtual view synthesis are discussed. For each sequence, the bitrate is calculated as a sum of bitrates for all atlases, including both depth and texture.

The quality of synthesized views was measured using 5 objective quality metrics, which are commonly used in immersive video applications: Weighted-to-Spherically-Uniform PSNR (WS-PSNR) [Sun17], Multi-Scale SSIM (MS-SSIM) [Wan03], Visual Information Fidelity (VIF) [She06], Video Multimethod Assessment Fusion (VMAF) [Li16] and ISO/IEC MPEG's metric for immersive video: IVPSNR [ISO19c].

All used metrics are full-reference ones, therefore in order to estimate quality, the virtual views in positions

Sequence	Bitrate reduction		Quality improvement	
	Base view	Atlas	Base view	Atlas
Classroom	0.78%	1.74%	0.02 dB	0.04 dB
Museum	1.16%	4.00%	0.01 dB	-0.01 dB
Hijack	0.00%	6.55%	0.03 dB	0.22 dB
Kitchen	0.49%	7.44%	0.05 dB	0.11 dB
CG	0.61%	4.93%	0.03 dB	0.09 dB
Painter	-0.09%	1.53%	-0.01 dB	22.86 dB
Frog	0.44%	1.39%	0.00 dB	33.00 dB
Fencing	0.30%	0.69%	0.00 dB	0.02 dB
NC	0.22%	1.20%	0.00 dB	18.63 dB
Average	0.44%	3.33%	0.01 dB	8.03 dB

Table 2: Bitrate reduction and quality improvement for the use of HEVC Screen Content Coding tools instead of the plain HEVC for the base views and atlases. A positive number denotes bitrate reduction or quality index increase for the synthesized videos due to the usage of SCC.

of input views were synthesized using decoded video data. Then, the estimated quality was averaged over all views.

In order to calculate the difference between two encoding approaches, the Bjoentegaard Delta [Bjo01] metric was used.

4.2 Efficiency of immersive video coding using HEVC-SCC

In the proposed approach, all videos, i.e., base views, atlases, and corresponding depth maps, are being independently encoded using HEVC-SCC. Therefore, it was possible to split the encoding results depending on the data type.

In Figs. 8 and 9, the rate-distortion curves for views only (excluding depth) are presented. In general, the usage of HEVC-SCC allows to achieve better quality at the same bitrate when compared to the HEVC main profile.

At this point, it has to be mentioned why the quality of the TechnicolorPainter and IntelFrog sequences is astonishingly high. Actually, the PSNR value presented in Fig. 9 was averaged over all encoded base views and atlases. While there were no issues for base views, some of the atlases contain no patches within one or more group of pictures (e.g., within the third GOP of the IntelFrog sequence, where there are fewer occlusions than for the first two GOPs, 5 of 8 atlases are empty thus completely grey).

As discussed in Section 3, HEVC-SCC should perform better on atlases than on base views. Indeed, as the results presented in Table 2 show, the bitrate reduction
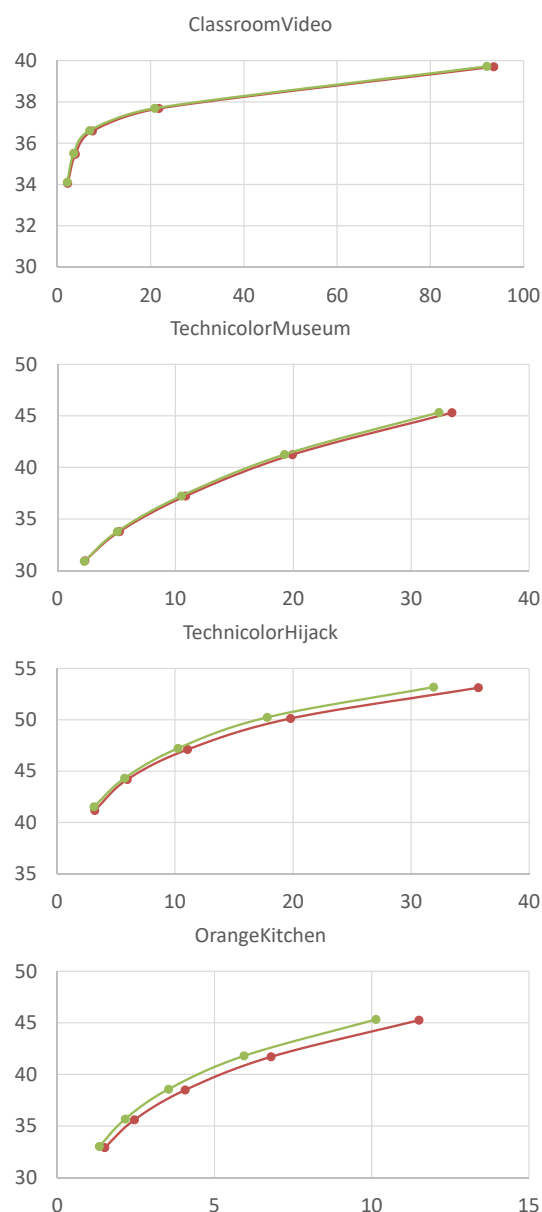

Figure 8: Rate-distortion curves for the immersive video codecs with the HEVC-SCC as compared to the plain HEVC: computer-generated sequences, input views encoding; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].

caused by using HEVC-SCC instead of HEVC is significantly higher for atlases than for base views. In general, also the quality improvement is bigger for atlases, however, the difference between HEVC and HEVC-SCC is really slight (except for the TechnicolorPainter and IntelFrog sequences, where HEVC-SCC performs much better for their almost empty atlases).

The second type of data being encoded is depth maps. The RD curves for depth are presented in Figs. 10 and 11. Compared to the encoding of input views, the en-
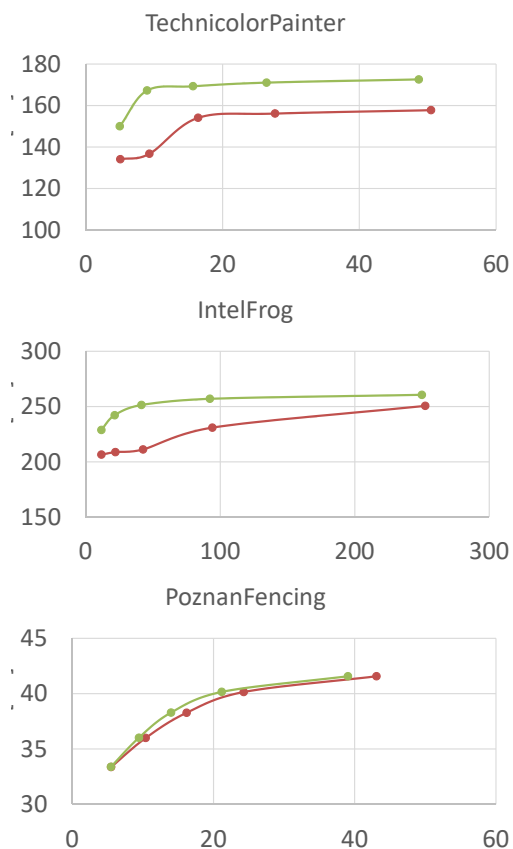
Figure 9: Rate-distortion curves for the immersive video codecs with the HEVC-SCC as compared to the plain HEVC: natural sequences, input views encoding; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].

coding gain in depth maps caused by the application of the SCC extension of HEVC is significantly higher. For all test sequences, HEVC-SCC allows for achieving a significantly better quality of depth maps, while preserving the same bitrates.

Such results are highly expected because of the characteristics of depth maps which contain mostly no texture, but large, smooth, semi-repeatable regions which can be efficiently encoded using SCC tools.

The efficiency of HEVC-SCC for base views and atlases is compared in Table 3. While for the input views encoding results were similar for natural and computer-generated sequences, the results for depth encoding are different for both sequence types. For computer-generated sequences, HEVC-SCC performs significantly better for atlases than for base views. However, for natural sequences, there is no significant difference between both types of data. The reason is the quality of depth maps, since for computer-generated sequences the depth is smooth within the objects' interior and sharp at their edges, while depth maps for natural content were algorithmically estimated based
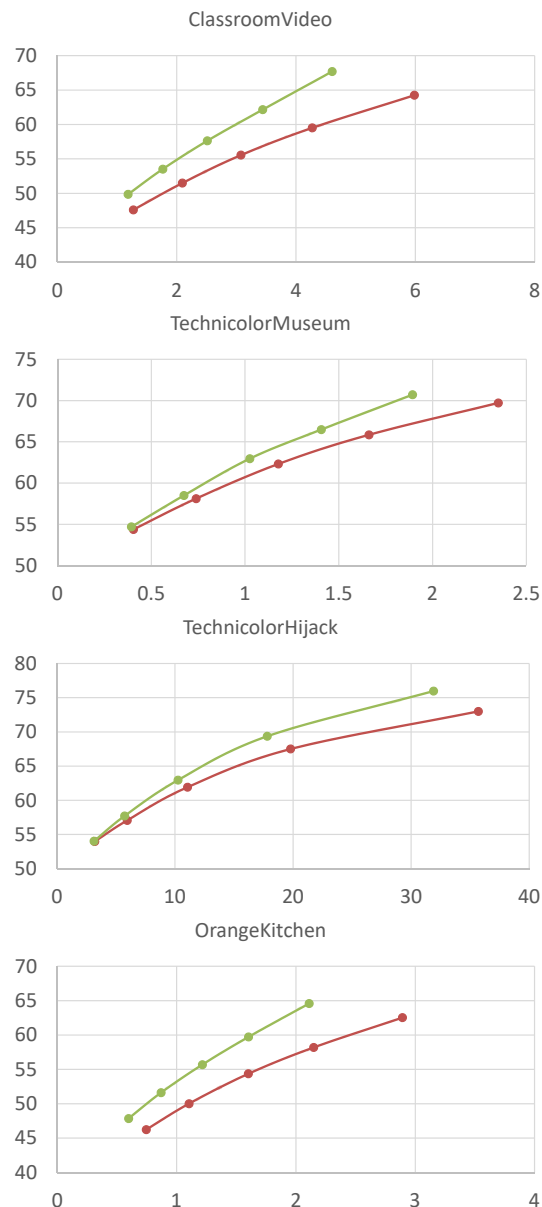

Figure 10: Rate-distortion curves for immersive video codecs with HEVC-SCC as compared to plain HEVC: computer-generated sequences, depth maps encoding; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].

on input views, therefore, they contain artifacts, such as blurred edges or grained objects. As a result, the atlases contain many small, different patches that negatively influence the HEVC-SCC encoding efficiency.

However, despite the problems described above, for depth data, HEVC-SCC performs much better than plain HEVC (even for natural sequences), helping reduce the bitrates and slightly increase the quality of decoded views.
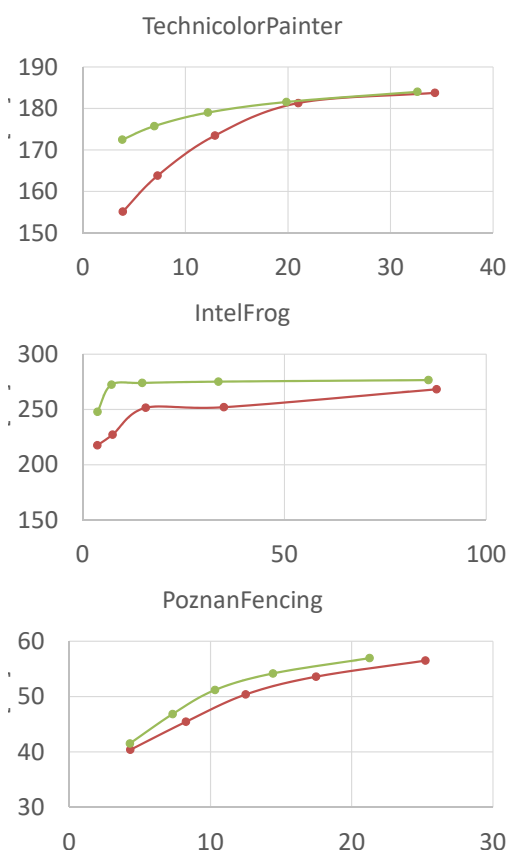
TechnicolorPainter

IntelFrog

PoznanFencing

Figure 11: Rate-distortion curves for immersive video codecs with HEVC-SCC as compared to plain HEVC: natural sequences, depth maps encoding; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].

Sequence	WSPSNR	VIF	VMAF	SSIM	IVPSNR
Classroom	22.36%	10.87%	16.42%	10.28%	10.24%
Museum	7.82%	4.07%	8.77%	4.48%	4.78%
Hijack	19.83%	14.91%	20.06%	15.88%	9.20%
Kitchen	22.64%	16.35%	30.19%	16.30%	5.65%
CG	18.16%	11.55%	18.86%	11.74%	7.47%
Painter	3.37%	3.75%	2.92%	3.37%	3.33%
Frog	3.85%	2.70%	5.04%	4.14%	1.48%
Fencing	11.41%	11.18%	10.23%	10.31%	9.46%
NC	6.21%	5.88%	6.06%	5.94%	4.76%
Average	13.04%	9.12%	13.38%	9.25%	6.31%

Table 4: BD-rate reduction.

4.3 Rendered video quality from compressed data using the standard and proposed approaches

As presented in the previous section, HEVC-SCC allows for decreasing the total bitrate of immersive video data. However, the user of the immersive video system is not concerned about the quality of atlases or corresponding depth maps but pays attention to the final quality of the video he or she is watching. Therefore, in this section, the quality of synthesized virtual views is considered.

In Figs. 12 and 13 the RD-curves for synthesized virtual views are presented. On the horizontal axis, the total bitrate (base views + depth maps and atlases + depth maps) is presented, on the vertical one – the average value of WS-PSNR for luma component of synthesized video. As presented, the proposed approach allows for increasing the quality of synthesized views (compared to HEVC main profile) while preserving the total bitrate.

For each sequence, the average bitrate reduction (Bjoentegaard Delta – BD) between two curves was also estimated. The BD-rate measures the average bitrate change. The same calculations are performed also for 4 other, commonly-used quality metrics. All these values are gathered in Table 4.

As presented, HEVC-SCC performs better for computer-generated sequences. The encoding efficiency for natural sequences is lower, however, even for that type of content, HEVC-SCC works better than HEVC main profile.

In Fig. 14 fragments of virtual views synthesized using data compressed by two encoders are compared with fragments of input views. Note shifted and ragged edges generated by HEVC main (at the middle column).

In general, HEVC-SCC clearly outperforms plain HEVC for all the test sequences and all calculated quality metrics. Therefore, HEVC-SCC is a good choice for immersive video coding.

| | Bitrate reduction | | Quality improvement | |
Sequence	Base view	Atlas	Base view	Atlas
Classroom	11.76%	18.38%	1.85 dB	3.14 dB
Museum	8.52%	13.12%	0.56 dB	0.62 dB
Hijack	7.89%	9.25%	1.09 dB	1.43 dB
Kitchen	15.34%	25.10%	1.28 dB	1.80 dB
CG	10.88%	16.47%	1.20 dB	1.75 dB
Painter	4.60%	4.27%	0.27 dB	8.77 dB
Frog	2.01%	3.16%	0.12 dB	32.32 dB
Fencing	14.23%	12.22%	0.90 dB	0.87 dB
NC	6.95%	6.55%	0.43 dB	13.99 dB
Average	9.19%	12.22%	0.87 dB	6.99 dB

Table 3: Bitrate reduction and quality improvement (compared to the HEVC main profile) for base views and atlases, depth data.
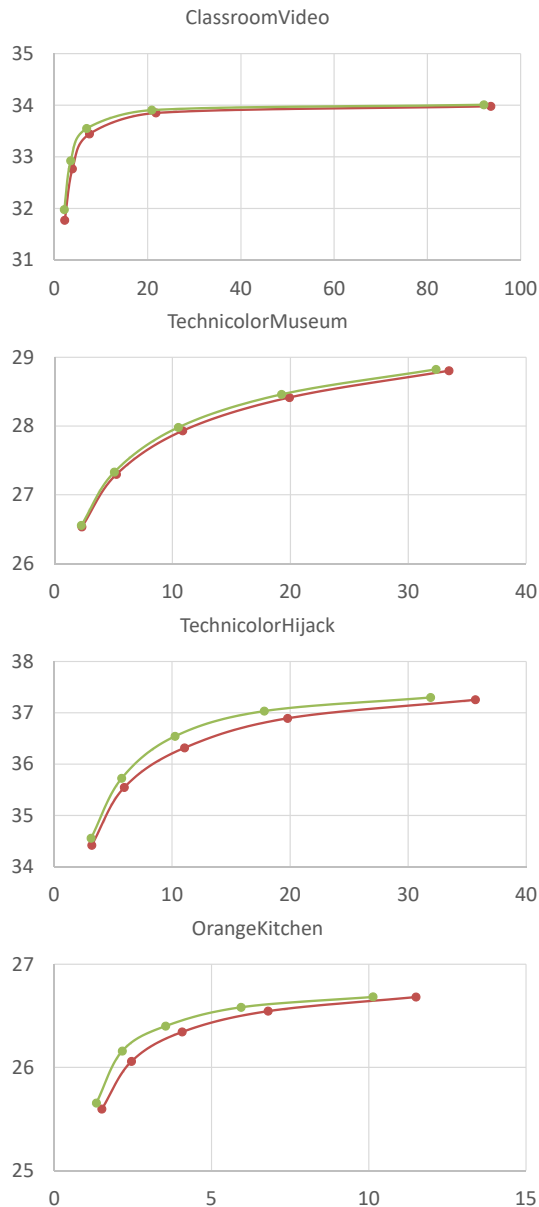
Figure 12: Rate-distortion curves for video synthesis from immersive video codecs with HEVC-SCC as compared to HEVC main profile: computer-generated sequences; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].
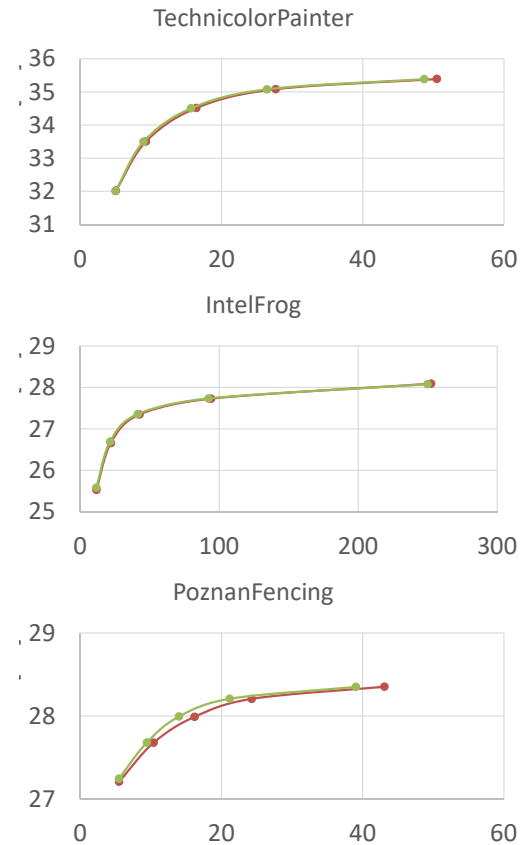

Figure 13: Rate-distortion curves for video synthesis from immersive video codecs with HEVC-SCC as compared to HEVC main profile: natural sequences; red: HEVC, green: HEVC-SCC. Vertical axis: PSNR [dB], horizontal: bitrate [Mbps].
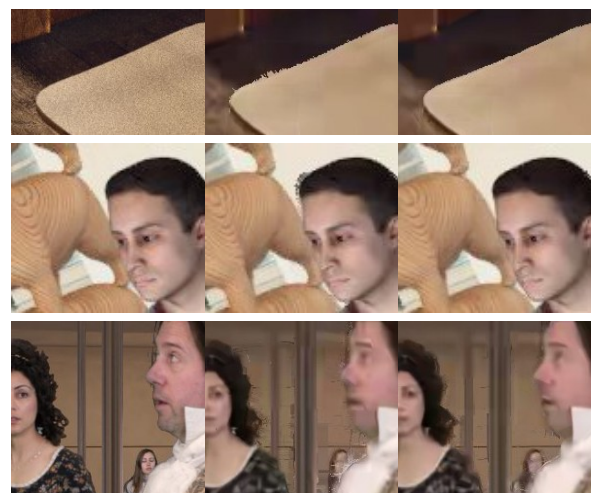
Figure 14: Fragments of: input views (left), views synthesized using data encoded using HEVC main profile (middle) and views synthesized using data encoded using HEVC-SCC (right). From top: ClassroomVideo, TechnicolorMuseum, TechnicolorHijack.

5 CONCLUSIONS

Immersive Video Coding is a new compression technology that is currently in the process of well-advanced standardization. The technology provides a solution for the generation of video sequences and parameters that represent immersive video. The video sequences may be then compressed using standard video coding techniques. In the process of development of this technology and its standardisation, HEVC coding was considered along with some experiments with VVC (Versatile

Video Coding) [Che20]. For Immersive Video Coding, both technologies proved to be useful and efficient according to the results obtained by MPEG. In this paper, we demonstrate that the overall rate-distortion coding efficiency may be even further improved by the use of the standard HEVC-SCC (HEVC Screen Content Coding). The proposed usage of HEVC-SCC requires no modification of the current draft for the standard on Immersive Video Coding [ISO20].

The novelty of the paper also consists in the application of the Screen Content Coding (SCC) technique for the compression of atlases that represent the immersive video. It is a new use of Screen Content Coding that was developed for completely other applications, i.e., with the aim to compress computer-generated images, like those transmitted to remote screens. This technique with the Intra Block Copy tool was never meant as a tool for the compression of immersive video content, in particular, natural immersive content acquired using cameras. The abovementioned application of Screen Content Coding was never described in the references. To our best knowledge, such an application is described for the first time in this paper.

In the paper, the application of Screen Content Coding to immersive video compression is experimentally tested in the framework of the Test Model for Immersive Video [ISO19d] that was recently developed by MPEG as a framework for the forthcoming international standard of immersive video compression [ISO20]. Currently, in the immersive video community, the research is executed using HEVC or 3D-HEVC codecs within the Test Model for Immersive Video. The idea of the paper is to replace HEVC or 3D-HEVC by another standard profile of the HEVC video codec, i.e., HEVC-SCC. It is worth to underline that the application of SCC (like HEVC-SCC) does not interfere with the general structure of the Test Model proposed for the future standard. For the standard test video sequences and the normalized experimental conditions used in the research on immersive video coding, the experimental data demonstrate that the application of HEVC-SCC is significantly more efficient than the traditional application of HEVC or 3D HEVC codecs for the compression of atlases representing the immersive video. This is clearly demonstrated for all MPEG test immersive video sequences available together with their reference data.

The quality improvement of the virtual views corresponds to the bitrate reduction of up to 20%. This quite a high value if we keep in mind that the whole HEVC technology has brought about 50% of the bitrate reduction. The experimental data (cf. Section 4.2) indicate that the main improvement yielded by the application of SCC is related to the higher fidelity of the decoded depth maps, and it is well-known that

the fidelity of depth significantly influences the quality of the virtual views. For the approach proposed, the quality improvement of virtual views is also higher for computer-generated content than for natural content acquired from cameras.

6 ACKNOWLEDGMENTS

7 REFERENCES

[Bjo01] G. Bjoentegaard. Calculation of average PSNR differences between RD-Curves. ITU-T VCEG Meeting, Austin, USA, 2001.

[Boi18] P. Boissonade and J. Jung. [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.

[Ceu18] B. Ceulemans et al. Robust Multiview Synthesis for Wide-Baseline Camera Arrays. IEEE Tr. on Multimedia, 2018.

[Cha19] J. Chakareski. UAV-IoT for next generation virtual reality. IEEE Tr. on Image Proc., 2019.

[Che20] J. Chen et al. The Joint Exploration Model (JEM) for Video Compression With Capability Beyond HEVC. IEEE Tr. on Circuits and Systems for Video Technology, 2020.

[Cui19] L. Cui et al. Point-Cloud Compression: Moving Picture Experts Group's New Standard in 2020. IEEE Cons. Electronics Magazine, 2019.

[Dom16] M. Domański et al. Multiview test video sequences for free navigation exploration obtained using pairs of cameras. ISO/IEC JTC1/SC29/WG11/M38247, Geneva, 2016.

[Dom19] M. Domański et al. Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by PUT and ETRI. ISO/IEC JTC1/SC29/WG11/M47407, Geneva, 2019.

[Dor18] R. Doré. Technicolor 3DoF+ test materials. ISO/IEC JTC1/SC29/WG11 MPEG/M42349, San Diego, USA, 2018.

[Doy17] D. Doyen et al. Light field content from 16-camera rig. ISO/IEC JTC1/SC29/WG11 MPEG, M40010, Geneva, Switzerland, 2017.

[Fle19] J. Fleureau et al. Technicolor-Intel Response to 3DoF+ CfP. ISO/IEC JTC1/SC29/WG11 MPEG/M47445, Geneva, Switzerland, 2019.

[Gar19] P. Garus et al. Bypassing Depth Maps Transmission For Immersive Video Coding. 2019 Picture Coding Symposium (PCS), 2019.

[Isg14] F. Isgro et al. Three-dimensional image processing in the future of immersive media. IEEE Tr. on Circuits and Systems for Video Tech., 2014.

[ISO15] ISO/IEC. High efficiency coding and media delivery in heterogeneous environment – Part 2: High efficiency video coding. ISO/IEC Int. Standard 23008-2, 2015.

[ISO19a] ISO/IEC MPEG. Call for Proposals on 3DoF+ Visual. ISO/IEC JTC1/SC29/WG11 MPEG/N18145, Marrakech, 2019.

[ISO19b] ISO/IEC MPEG. Common Test Conditions for Immersive Video. ISO/IEC JTC1/SC29/WG11/N18789, Geneva, 2019.

[ISO19c] ISO/IEC MPEG. Software manual of IV-PSNR for Immersive Video. ISO/IEC JTC1/SC29/WG11/N18709, Goeteborg, 2019.

[ISO19d] ISO/IEC MPEG. Test Model 3 for Immersive Video. ISO/IEC JTC1/SC29/WG11 MPEG/N18795, Geneva, Switzerland, 2019.

[ISO19e] ISO/IEC MPEG. Working Draft 3 of Immersive Video. ISO/IEC JTC1/SC29/WG11 MPEG/N18794, Geneva, Switzerland, 2019.

[ISO20] ISO/IEC MPEG. Text of ISO/IEC CD 23090-12 MPEG Immersive Video. ISO/IEC JTC1/SC29/WG11/N19482, Online, 2020.

[Kro18] B. Kroon. 3DoF+ test sequence ClassroomVideo. ISO/IEC JTC1/SC29/WG11 MPEG/M42415, San Diego, USA, 2018.

[Laf19] G. Lafruit et al. Understanding MPEG-I Coding Standardization in Immersive VR/AR Applications. SMPTE Motion Imaging Journal, 2019.

[Li16] Z. Li et al. Toward a practical perceptual video quality metric. Netflix Technology Blog, 2016.

[Li20] L. Li et al. Advanced 3D Motion Prediction for Video-Based Dynamic Point Cloud Compression. IEEE Tr. on Image Processing, 2020.

[Mie20] D. Mieloch et al. Depth Map Estimation for Free-Viewpoint Television and Virtual Navigation. IEEE Access, 2020.

[Mue11] K. Mueller et al. 3-D video representation using depth maps. Proc. of the IEEE, 2011.

[Rah18] D. Rahaman and M. Paul. Virtual view synthesis for free viewpoint video and multiview video compression using Gaussian mixture modelling. IEEE Tr. on Image Processing, 2018.

[Sal19] B. Salahieh et al. Kermit test sequence for Windowed 6DoF Activities. ISO/IEC JTC1/SC29/WG11/M43748, Ljubljana, 2019.

[Sam17] J. Samelak et al. Efficient frame-compatible stereoscopic video coding using HEVC Screen Content Coding. IWSSIP 2017, Poznan, 2017.

[Sam19] J. Samelak and M. Domaański. Unified Screen Content and Multiview Video Coding - Experimental results. ISO/IEC JTC1/SC29/WG11/M46332, Marrakech, 2019.

[Sch19] S. Schwarz et al. Emerging MPEG Standards for Point Cloud Compression. IEEE J. on Emerging and Sel. Topics in Circuits and Systems, 2019.

[She06] H. Sheikh and A. Bovik. Image information and visual quality. IEEE Tr. on Image Proc., 2006.

[Sta18] O. Stankiewicz et al. A free-viewpoint television system for horizontal virtual navigation. IEEE Tr. on Multimedia, 2018.

[Sun17] Y. Sun et al. Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video. IEEE Signal Processing Letters, 2017.

[Tan12] M. Tanimoto et al. FTV for 3-D spatial communication. Proc. of the IEEE, 2012.

[Tec16] G. Tech et al. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. IEEE Tr. Circuits and Syst. for Vid. Tech., 2016.

[Wan03] Z. Wang et al. Multiscale structural similarity for image quality assessment. The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers, 2003.

[Wie19] M. Wien et al. Standardization Status of Immersive Video Coding. IEEE J. on Emerging and Selected Topics in Circuits and Systems, 2019.

[Xu16a] X. Xu et al. Intra Block Copy in HEVC Screen Content Coding Extensions. IEEE J. on Emerging and Selected Topics in Circuits and Systems, 2016.

[Xu16b] J. Xu et al. Overview of the Emerging HEVC Screen Content Coding Extension. IEEE Tr. on Circuits and Systems for Video Technology, 2016.

[Xu16c] X. Xu et al. Palette Mode Coding in HEVC Screen Content Coding Extension. IEEE J. on Emerging and Sel. Top. in Cir. and Syst., 2016.

[Yu15] H. Yu et al. Common Test Conditions for Screen Content Coding. JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11: Doc. JCTVC-U1015r2, Warsaw, Poland, 2015.

[Yua18] Y. Yuan et al. Object shape approximation and contour adaptive depth image coding for virtual view synthesis. IEEE Tr. on Circuits and Systems for Video Technology, 2018.

[Zha20] J. Zhang et al. Point Cloud Normal Estimation by Fast Guided Least Squares Representation. IEEE Access, 2020.

[Zhu19] S. Zhu et al. An improved depth image based virtual view synthesis method for interactive 3D video. IEEE Access, 2019.