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 multi-view 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
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 Immersive 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-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 atl-
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], [Fie19], [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

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.

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

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Content type</th>
<th>Number of base views</th>
<th>Number of atlases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom [Kro18]</td>
<td>O, CG</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Museum [Dor18]</td>
<td>O, CG</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Hijack [Dor18]</td>
<td>O, CG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Kitchen [Boi18]</td>
<td>P, CG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Painter [Doy17]</td>
<td>P, NC</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Frog [Sal19]</td>
<td>P, NC</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Fencing [Dom16]</td>
<td>P, NC</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

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...
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.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Bitrate reduction</th>
<th>Quality improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base view</td>
<td>Atlas</td>
</tr>
<tr>
<td>Classroom</td>
<td>0.78% 1.74%</td>
<td>0.02 dB</td>
</tr>
<tr>
<td>Museum</td>
<td>1.16% 4.00%</td>
<td>0.01 dB</td>
</tr>
<tr>
<td>Hijack</td>
<td>0.00% 6.55%</td>
<td>0.03 dB</td>
</tr>
<tr>
<td>Kitchen</td>
<td>0.49% 7.44%</td>
<td>0.05 dB</td>
</tr>
<tr>
<td>CG</td>
<td>0.61% 4.93%</td>
<td>0.03 dB</td>
</tr>
<tr>
<td>Painter</td>
<td>-0.09% 1.53%</td>
<td>-0.01 dB</td>
</tr>
<tr>
<td>Frog</td>
<td>0.44% 1.39%</td>
<td>0.00 dB</td>
</tr>
<tr>
<td>Fencing</td>
<td>0.30% 0.69%</td>
<td>0.00 dB</td>
</tr>
<tr>
<td>NC</td>
<td>0.22% 1.20%</td>
<td>0.00 dB</td>
</tr>
<tr>
<td>Average</td>
<td>0.44% 3.33%</td>
<td>0.01 dB</td>
</tr>
</tbody>
</table>

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 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-
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 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.

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].

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].
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].

Table 3: Bitrate reduction and quality improvement (compared to the HEVC main profile) for base views and atlases, depth data.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Base view</th>
<th>Atlas</th>
<th>Base view</th>
<th>Atlas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom</td>
<td>11.76%</td>
<td>18.38%</td>
<td>1.85 dB</td>
<td>3.14 dB</td>
</tr>
<tr>
<td>Museum</td>
<td>8.52%</td>
<td>13.12%</td>
<td>0.56 dB</td>
<td>0.62 dB</td>
</tr>
<tr>
<td>Hijack</td>
<td>7.89%</td>
<td>9.25%</td>
<td>1.09 dB</td>
<td>1.43 dB</td>
</tr>
<tr>
<td>Kitchen</td>
<td>15.34%</td>
<td>25.10%</td>
<td>1.28 dB</td>
<td>1.80 dB</td>
</tr>
<tr>
<td>CG</td>
<td>10.88%</td>
<td>16.47%</td>
<td>1.20 dB</td>
<td>1.75 dB</td>
</tr>
<tr>
<td>Painter</td>
<td>4.60%</td>
<td>4.27%</td>
<td>0.27 dB</td>
<td>8.77 dB</td>
</tr>
<tr>
<td>Frog</td>
<td>2.01%</td>
<td>3.16%</td>
<td>0.12 dB</td>
<td>32.32 dB</td>
</tr>
<tr>
<td>Fencing</td>
<td>14.23%</td>
<td>12.22%</td>
<td>0.90 dB</td>
<td>0.87 dB</td>
</tr>
<tr>
<td>NC</td>
<td>6.95%</td>
<td>6.55%</td>
<td>0.43 dB</td>
<td>13.99 dB</td>
</tr>
<tr>
<td>Average</td>
<td>9.19%</td>
<td>12.22%</td>
<td>0.87 dB</td>
<td>6.99 dB</td>
</tr>
</tbody>
</table>

Table 4: BD-rate reduction.

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.

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 (Bjøntegaard 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.
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).

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

7 REFERENCES


