

Advanced HEVC Screen Content Coding for MPEG Immersive Video

Jarosław Samelak, Adrian Dziembowski * and Dawid Mieloch

Institute of Multimedia Telecommunications, Poznan University of Technology, 60-965 Poznań, Poland

* Correspondence: adrian.dziembowski@put.poznan.pl

Abstract: This paper presents the modified HEVC Screen Content Coding (SCC) that was adapted to be more efficient as an internal video coding of the MPEG Immersive Video (MIV) codec. The basic, unmodified SCC is already known to be useful in such an application. However, in this paper, we propose three additional improvements to SCC to increase the efficiency of immersive video coding. First, we analyze using the quarter-pel accuracy in the intra block copy technique to provide a more effective search of the best candidate block to be copied in the encoding process. The second proposal is the use of tiles to allow inter-view prediction inside MIV atlases. The last proposed improvement is the addition of the MIV bitstream parser in the HEVC encoder that enables selecting the most efficient coding configuration depending on the type of currently encoded data. The experimental results show that the proposal increases the compression efficiency for natural content sequences by almost 7% and simultaneously decreases the computational time of encoding by more than 15%, making the proposal very valuable for further research on immersive video coding.

Keywords: immersive video coding; screen content coding; MPEG immersive video (MIV)

Citation: Samelak, J.; Dziembowski, A.; Mieloch, D. Advanced HEVC Screen Content Coding for MPEG Immersive Video. *Electronics* **2022**, *11*, 4040. <https://doi.org/10.3390/electronics11234040>

Academic Editors: Kajetana Snopek, Wojciech Wojtasiak and Grzegorz Pastuszek

Received: 28 October 2022

Accepted: 2 December 2022

Published: 5 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Virtual navigation is an essential functionality of virtual reality systems that lets a user watch a scene from any arbitrary location. The transmission of such immersive content requires an efficient representation, as independent compression of multiple acquired views (and their depth maps) yields not practical ranges of bitrates and pixel rates. In the reasonable range of 5 to 30 Mbps, the simulcast encoding cannot properly handle the multiview content, especially for more sophisticated camera arrangements and types, e.g., omnidirectional content captured by fish-eye or 360-degree cameras [1]. In such cases, large parts of the scene are not transmitted because of the pixel rate constraints (i.e., the total number of samples that can be decoded by a typical consumer device [2]), significantly decreasing the immersion of the immersive video system user.

MPEG immersive video is the first incoming video codec that tries to deal with the compression of content for virtual reality applications [2]. The MIV encoding is based on removing inter-view and temporal redundancies from input videos and efficient packing of remaining information (in the form of small patches) into atlases fed into the video compressor (Figure 1). Since atlases are kept in 2D shape, atlases can be encoded using any video encoder.

The MIV is a new type of video codec that internally utilizes another standard video compression technique (e.g., HEVC [3], Figure 2). This codec-agnosticism future-proofs the overall compression efficiency, as the internal codec can be easily changed when a new compression standard is available. While such an approach can ensure the longevity of the standard, this paper proposes a different approach. In the proposal, the internal codec is also adapted to the characteristics of immersive video. This way, there is no need to wait for the new, more efficient codecs to emerge, as the quality can be improved using

already available solutions (even without any changes to the coder itself, but just with its proper configuration [4]).

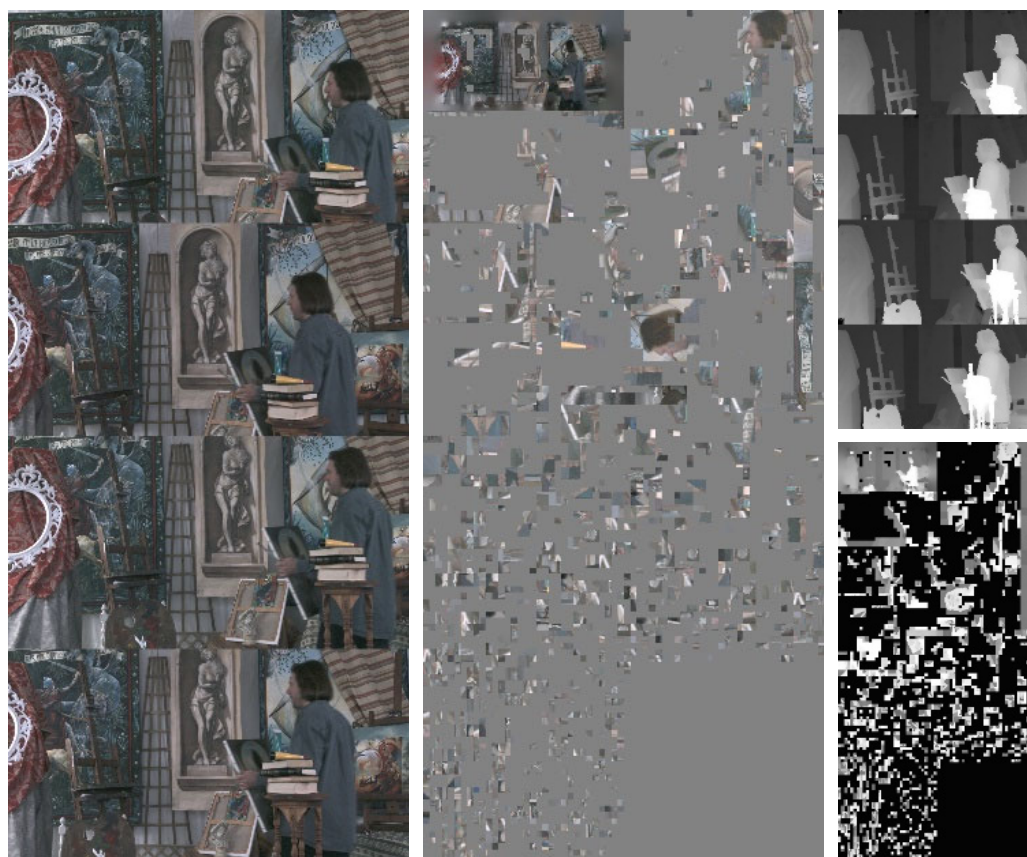


Figure 1. Four atlases for the Painter sequence [5]: two full-resolution texture atlases (left: first texture atlas containing a subset of stacked input views, middle: second texture atlas containing “patches” from other input views), and two reduced-resolution depth atlases (right: two depth atlases, containing depth values which correspond to the objects in both texture atlases).

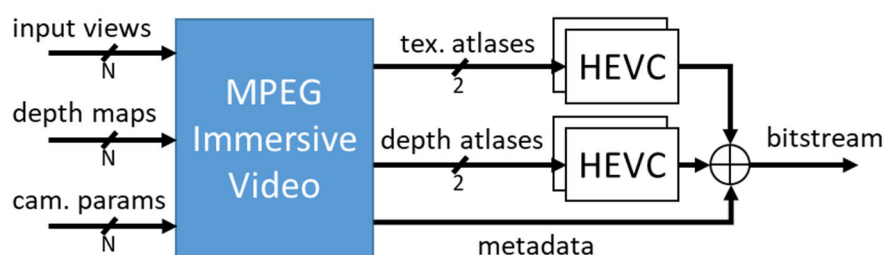


Figure 2. Simplified MPEG Immersive Video coding scheme; input data: N input views, corresponding depth maps, and camera parameters (intrinsic and extrinsic); MIV encoder output—metadata and 4 video streams, including two texture atlases with corresponding depth atlases; each video stream is independently encoded using basic HEVC encoder.

As was shown in the previous research of the authors on HEVC screen content coding [6,7], its use as an internal TMIV codec can significantly increase the final quality of video and simultaneously decrease the required bitrate [8]. Such results made HEVC SCC a good base for further modifications that enhanced its efficiency in immersive video coding.

This paper describes three proposed modifications:

1. The quarter-pel accuracy in Intra Block Copy (IBC) technique (earlier presented in [9], but in the context of multiview video compression, not in immersive video),
2. The use of tiles to allow inter-view prediction inside MIV atlases,
3. Adding the MIV bitstream parser in the HEVC encoder allows selecting the most efficient coding configuration depending on the type of currently encoded data (e.g., different configurations for textures and depth maps).

The works related to improving immersive video coding using MPEG immersive video are presented in Section 2. The above listed proposed modifications of HEVC screen content coding are described in Section 3. Section 4 includes the methodology of the experiments, while Section 5—their results. Section 6 concludes the paper, summarizing performed research and pointing out possible future works.

2. Related Works

The work on the MPEG immersive video coding standard started in 2019. During the three years of development of MIV, multiple researchers worked on increasing its efficiency, and several papers have already been published.

Methods described in the literature increase the efficiency of all essential steps of the MIV coding pipeline [2], including pruning, packing, atlas preparation, and the last step performed at the decoder side—rendering. Two effective techniques enhancing pruning are presented in [10] and [11]. The first one utilizes input depth maps and input view textures, allowing for proper pruning of highly-reflective non-Lambertian surfaces. The technique presented in [11] adds another pruning iteration, significantly increasing the encoding efficiency for high-bitrate systems.

Several papers regard the atlas preparation step, including the patch merging [12], allowing for more efficient packing of non-pruned information within atlases, non-linear depth quantization [13], or adaptation of patch occupancy [14] to the block-based nature of typical video encoders, such as HEVC or VVC. Other works focus on rendering and increasing the quality of synthesized viewports for immersive video [15] and non-Lambertian content [16].

A significant number of MIV-related papers regard the MIV geometry absent profile [1] and the decoder-side depth estimation (DSDE) approach [17], which allows for decreasing the amount of transmitted video data by omitting transmission of depth. Several papers utilize the geometry assistance SEI message [18], using the encoder-derived features of depth maps [17] transmitted within MIV metadata. Such features may be used for increasing the quality of depth maps estimated at the decoder side, decreasing the total decoding time [19], or for efficient encoding of multi-plane images (MPI) [20]. Other interesting approaches were presented in [21], where a possibility of utilizing motion compensation for DSDE was examined, and [22] – where depth estimation is performed for transmitted full views and smaller patches transmitted within texture atlases [2].

All mentioned papers focus on increasing MIV coding efficiency, while none modifies the video encoders used within the MIV encoding pipeline. We believe the modification presented in this paper is particularly beneficial, as it does not exclude using other described methods and can be used simultaneously with them.

Other MIV-related research regards using existing video encoders for efficient transmission of MIV content, e.g., by including MIV bitstream within HEVC SEI [23], optimization of delta QP between texture and depth [24,25], using the VVC frame packing mechanism [26,27], or using the unmodified HEVC SCC—what was studied by the Authors of this paper in [8].

Some of the already presented modifications of HEVC SCC were focused on increasing its efficiency for some applications, e.g., encoding of raw light-field video [28] and implementing multiview video coding equivalent to HEVC Multiview Coding (MV-

HEVC) [9,29,30]. To the authors' best knowledge, this paper presents the first modification of a video compression technique (i.e., HEVC SCC) that makes it more adapted to the characteristics of video produced by another compression standard (MIV).

3. Proposed Modifications of Screen Content Coding

As mentioned before, using HEVC screen content coding as an internal MIV codec can significantly improve the efficiency of immersive video compression when compared to plain HEVC [8]. Based on the characteristics of immersive video processed by the MIV encoder (after pruning and packing into atlases), the authors proposed adding three modifications to increase the compression efficiency further. It should be mentioned that these modifications, contrary to the proposal from [8], are not compliant with the HEVC Screen Content Coding standard. The first two tools are adopted from [9,30], which have already been proposed for multiview video coding.

3.1. Quarter-Pel Accuracy of Intra Block Copy

HEVC Intra Block Copy (IBC) is a compression technique in which the processed block and its prediction candidate (the best matching block of points) belong to the same picture [31]. Such an approach is beneficial, especially for the compression of fonts and other repetitive patterns. The search for the most similar block is initially limited to full-pel (pixel) accuracy. Such restriction is reasonable if the vast majority of the input video is screen content. For compression of camera-captured or realistic computer-generated content, it can be beneficial to increase the block matching accuracy to sub-pel. In other words, IBC was successfully applied with sub-pel accuracy for compression of camera-captured, frame-compatible multiview video [9,30].

It is proposed to limit block matching to quarter-pel to make it consistent with the inter-picture prediction in HEVC [3]. Naturally, sub-pixel block matching requires additional interpolation and modifications of the IBC search algorithm. In the proposal, the IBC first finds the best match at full-pel precision. Then, the same interpolation algorithm as in the HEVC motion estimation is applied, and the search result is refined at quarter-pel accuracy.

Introducing more precise vectors is only justified if the gain from more accurate prediction is bigger than the cost of encoding vectors at higher precision. Section 5.1 presents experimental results of increasing intra block copy precision in the compression of computer-generated and natural-content immersive video.

3.2. Tile-Based Intra Block Copy

As presented in Figure 1, an atlas can contain multiple base views combined into a single frame. Using intra block copy for compression of such a sequence can be very efficient because a frame-compatible multiview video forms a somewhat repetitive pattern [9,29,30]. Intra block copy may match the same objects in different views, which is an equivalent of disparity compensated prediction performed in Multiview HEVC [32]. Intra block copy is unaware that the frame is composed of similar views and simply tries to find the best match within the previously encoded area. This approach is obviously not optimal and may have a negative impact on the encoding time.

In the proposal, the authors divided the atlases in such a way that each base view is encoded as a single tile. Tile encoding is an HEVC-compliant procedure and can be set up in the configuration of the HEVC SCC test model [33]. The same division in tiles was applied to both atlases with base views and patches from other views.

On top of tile encoding, the IBC algorithm was also modified. In the proposal, IBC starts the block matching from the collocated position in the previously encoded tile instead of the neighboring block. This way, the best match can be found much faster, and the IBC vector is shorter, which benefits the overall compression efficiency.

Moreover, the in-loop filtering was applied after the compression of each tile instead of filtering the whole slice (as initially done in standard-compliant HEVC [3]). Such an approach improves the quality of tiles used as a reference by the Intra Block Copy technique.

3.3. Parsing of MIV Bitstream

The output of the MIV encoder, which is the input for standard video coding technique (Figures 1 and 2), is composed of atlases, each of which has different characteristics. Therefore, it seems reasonable to apply different compression techniques for each type of input video. In this paper, the authors propose to add a MIV metadata parser and use it to control the configuration of the modified HEVC screen content coding codec (Figure 3). MIV metadata allow the recognition of whether the compressed data contains texture or depth maps, as well as base views or patches, and configures the accuracy of IBC and tile encoding accordingly. Moreover, metadata provide information about the resolution of base views, which can be used to set up the tile size automatically.

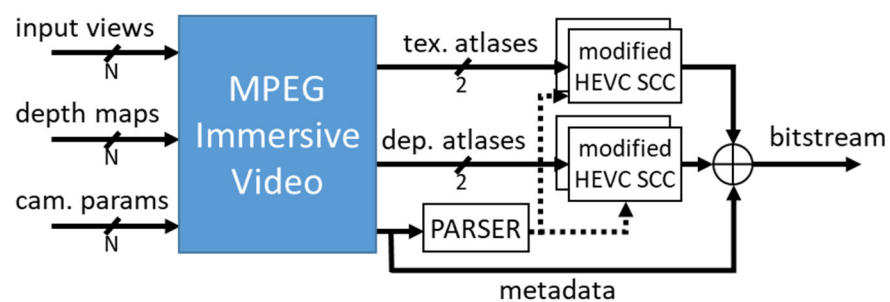


Figure 3. Simplified MPEG immersive video coding scheme with modified HEVC SCC encoder controlled by MIV metadata; input data: N input views, corresponding depth maps, and camera parameters (intrinsic and extrinsic); MIV encoder output—metadata and 4 video streams, including two texture atlases with corresponding depth atlases; each video stream is independently encoded using modified HEVC SCC encoder, encoding parameters (e.g., QP) are controlled by MIV metadata.

The research on how the proposed modifications influence the overall quality of the immersive video was presented in Section 5.

4. Methodology of the Experiments

The proposed extension of the HEVC screen content coding was tested under MPEG immersive video (MIV) common test conditions (CTC) [5], which define the strict framework for the evaluation of new video encoding methods. The experiments were performed using 15 test sequences, including nine computer-generated (CG) and six natural content sequences (NC) captured by real multicamera systems. Seventeen frames were encoded and processed for each test sequence according to the MIV CTC.

Virtual views were synthesized using decoded bitstreams and placed in the same positions as all input views to allow objective quality assessment. The quality was calculated using two full-reference metrics: WS-PSNR [34] and IV-PSNR [35]. Then, for both metrics, the BD rates were calculated [36] in order to quantify the relationship between bitrate and quality. Only the video encoding step of the entire MIV encoding pipeline was modified in the experiments. The MIV encoding (i.e., creation of atlases performed by the test model for MIV: TMIV [37]) and decoding (which includes virtual view synthesis) remained unchanged (cf. Figure 2). Due to SCC being compatible only with 8-bps videos [33], the bit-depth of atlases produced by the TMIV encoder was changed from 10 to 8 bps before HEVC encoding and changed back to 10 bps after HEVC decoding.

Besides the change described above, we have followed the MIV CTC. Therefore, all the detailed information about the framework of experiments can be found in [5], including a description of the used test sequences, chosen start frames, QP values, delta QP between texture and depth components, exact versions of used software, etc.

In total, six experiments were performed to test the performance of the proposed SCC modifications. In the first experiment, the usage of quarter-pel accuracy of the IBC (cf. Section 3.1) was compared to the basic SCC. The second and third experiments evaluated the performance of quarter pel IBC accuracy used only for one data type. In experiment 2, texture atlases were encoded using typical SCC (full-pel accuracy), while for depth atlases, the quarter-pel accuracy was used. In experiment 3, the opposite variant was chosen (full-pel accuracy for depth and quarter-pel accuracy for texture).

The fourth experiment evaluated the performance of quarter-pel accuracy and tile-based IBC analysis (cf. Section 3.2). In this experiment, the tile-based IBC analysis was enabled for the first texture atlas, as it contains a subset of stacked full input views, while the second one contains numerous smaller patches (Figure 1). In experiments 5 and 6, the usage of tile-based IBC analysis for more video streams was tested for texture atlases only (experiment 5) and all four atlases (experiment 6).

5. Experimental Results

5.1. Quarter-Pel IBC Accuracy

The basic HEVC SCC uses the full-pel accuracy of the intra block copy. In the first experiment, it was compared to the proposed quarter-pel IBC accuracy.

As presented in Table 1, the proposed modification significantly increases the encoding efficiency, especially for the natural content. For computer-generated sequences, the profit from using the proposed modification is smaller but still noticeable for IV-PSNR. Regarding WS-PSNR, the proposed quarter-pel IBC accuracy does not negatively impact the coding efficiency. Of course, a more precise analysis of IBC vectors requires more operations, thus, higher computational time. On average, the proposed modification increases the encoding time by 7%.

Table 1. Basic SCC vs. SCC with quarter-pel IBC accuracy; BD-rates for two used quality metrics (WS-PSNR and IV-PSNR, a negative value indicates increased coding efficiency) and increase of the encoding time when compared to basic SCC (value higher than 100% indicates longer computational time); BD-rate change higher than 2% was highlighted in red (for efficiency loss) or green (for efficiency gain).

Sequence	BD-Rate WS-PSNR _Y	BD-Rate IV-PSNR	Δ Encoding Time
ClassroomVideo	-0.3%	0.7%	107.2%
Museum	-0.6%	0.9%	108.7%
Fan	-0.3%	0.5%	103.7%
Kitchen	-0.1%	-0.2%	106.3%
Chess	-0.5%	-1.3%	104.3%
Group	-0.6%	0.5%	108.0%
ChessPieces	-1.0%	-3.0%	108.3%
Hijack	-0.1%	-0.5%	113.2%
Mirror	-0.5%	0.0%	101.5%
CG: average	-0.0%	-0.3%	106.8%
Painter	-1.6%	-1.9%	110.4%
Frog	-0.1%	-0.0%	107.5%
Carpark	-1.3%	-1.4%	108.7%
Fencing	-2.0%	-2.4%	107.2%
Hall	-6.0%	-7.3%	106.2%
Street	-2.9%	-3.1%	108.6%
NC: average	-2.3%	-2.7%	108.1%
All: average	-0.9%	-1.2%	107.3%

Table 2 contains bitrates and IV-PSNR values averaged separately for natural content and computer-generated sequences. As presented, the quality of the synthesized virtual views is practically the same for full-pel and proposed quarter-pel IBC accuracy. When considering bitrates, the results for texture atlases are also similar. A significant gain caused by using the quarter-pel accuracy for depth for natural content can be spotted for depth atlases.

Table 2. Basic SCC vs. SCC with quarter-pel IBV accuracy, bitrate, and quality change for different rate points; bitrate change higher than 2% was highlighted in red (for efficiency loss) or green (for efficiency gain).

Seq.	Rate	Texture Bitrate [Mbps]			Depth Bitrate [Mbps]			IV-PSNR [dB]	
		FPel	QPel	Δ	FPel	QPel	Δ	FPel	QPel
CG	R1	58.95	58.38	-1%	7.38	7.35	0%	44.31	44.31
	R2	15.81	15.72	-1%	5.39	5.37	0%	43.04	43.05
	R3	5.01	5.07	1%	3.72	3.71	0%	40.89	40.90
	R4	1.65	1.66	1%	2.78	2.78	0%	38.14	38.14
NC	R1	79.46	79.02	-1%	31.07	28.95	-7%	45.41	45.42
	R2	16.59	16.53	0%	13.81	12.74	-8%	43.92	43.93
	R3	5.42	5.51	2%	6.19	5.90	-5%	41.76	41.75
	R4	2.02	2.04	1%	3.76	3.69	-2%	38.46	38.47

Due to the significant difference between results obtained for depth and texture atlases, the performance of the proposed more precise IBC vectors for different types of videos was evaluated. In Table 3, the results for using the quarter-pel IBC accuracy for depth atlases together with basic full-pel accuracy for texture atlases are presented, together with the results of the alternate experiment, where the quarter-pel accuracy was used for texture, and the depth atlases were encoded using basic HEVC SCC.

Table 3. Basic SCC vs. SCC with mixed full-pel and quarter-pel IBC accuracy; BD-rates for two used quality metrics (WS-PSNR and IV-PSNR, a negative value indicates increased coding efficiency) and increase of the encoding time when compared to basic SCC (value higher than 100% indicates longer computational time); BD-rate change higher than 2% was highlighted in red (for efficiency loss) or green (for efficiency gain).

Sequence	QPel Accuracy for Depth, FPel for Texture			FPel Accuracy for Depth, QPel for Texture		
	BD-Rate WS-PSNR _y	BD-Rate IV-PSNR	Δ Encoding Time	BD-Rate WS-PSNR _y	BD-Rate IV-PSNR	Δ Encoding Time
ClassroomVideo	-0.2%	-0.2%	100.8%	-0.1%	0.9%	107.0%
Museum	-0.0%	-0.1%	100.3%	0.6%	1.0%	108.3%
Fan	0.3%	0.2%	99.1%	-0.0%	0.3%	102.7%
Kitchen	-0.1%	-0.3%	97.2%	0.2%	0.0%	105.9%
Chess	-0.1%	-0.1%	97.6%	-0.4%	-1.1%	103.7%
Group	0.4%	0.2%	101.9%	0.3%	0.3%	107.4%
ChessPieces	-1.0%	-1.5%	99.8%	-0.0%	-1.5%	107.8%
Hijack	0.1%	-0.2%	100.9%	-0.2%	-0.4%	112.0%
Mirror	-0.2%	-0.2%	99.4%	0.7%	0.3%	101.1%
CG: average	-0.1%	-0.2%	99.7%	0.1%	-0.0%	106.2%
Painter	-1.3%	-1.3%	100.4%	-0.3%	-0.6%	108.9%
Frog	-0.5%	-1.0%	101.1%	0.6%	1.0%	106.9%
Carpark	-1.5%	-1.7%	103.2%	0.2%	0.3%	107.8%
Fencing	-1.9%	-2.5%	100.8%	-0.0%	0.0%	106.3%
Hall	-6.3%	-7.1%	101.3%	0.4%	-0.2%	105.9%
Street	-4.5%	-4.6%	100.6%	1.5%	1.5%	108.2%

NC: average	-2.7%	-3.0%	101.2%	0.4%	0.3%	107.3%
All: average	-1.1%	-1.4%	100.3%	0.2%	0.1%	106.6%

As presented, the quarter-pel accuracy for the IBC vector search is more effective for depth atlases of the natural content. Such depth maps have very similar characteristics to typical screen content, but due to their estimation performed using natural textures, their quality is much lower than for CGI content (e.g., the edges of objects are much less stable). Quarter-pel accuracy of the IBC helps to identify better prediction candidates, decreasing the bitrate required to achieve the same quality. For texture atlases, the BD-rate difference is negligible, both for WS-PSNR and IV-PSNR metrics.

5.2. QPel + Tile-Based IBC Analysis

In the fourth experiment, both proposed modifications of SCC were used. However, the tile-based IBC analysis was enabled only for the first texture atlas and disabled for the second texture and both depth atlases. In this case, other atlases were encoded in the same way as in the previous section (SCC with quarter-pel accuracy).

The tile-based IBC analysis restricts the intra block copy vectors to be found in another part of the atlas (the similar blocks are searched in different views stacked within the atlas, not within the same view). Such a restriction significantly decreases the time required for encoding (Table 4).

Table 4. Basic SCC vs. SCC with quarter-pel IBC accuracy and tile-based IBC analysis for first texture atlas; BD-rate change higher than 2% was highlighted in red (for efficiency loss) or green (for efficiency gain).

Sequence	BD-Rate WS-PSNR _r	BD-Rate IV-PSNR	Δ Encoding Time
ClassroomVideo	2.6%	-5.2%	54.4%
Museum	1.3%	2.1%	78.5%
Fan	0.7%	1.4%	80.7%
Kitchen	1.6%	2.8%	90.4%
Chess	0.3%	-0.5%	90.6%
Group	0.8%	0.3%	77.1%
ChessPieces	-0.6%	-2.4%	93.5%
Hijack	0.6%	1.4%	113.0%
Mirror	2.4%	2.8%	91.4%
CG: average	1.1%	0.3%	85.5%
Painter	-2.3%	-4.2%	95.7%
Frog	-0.7%	-5.6%	75.8%
Carpark	-6.1%	-8.2%	76.0%
Fencing	-1.5%	-3.5%	82.5%
Hall	-6.7%	-7.7%	90.7%
Street	-10.5%	-12.3%	76.8%
NC: average	-4.6%	-6.9%	82.9%
All: average	-1.2%	-2.6%	84.5%

Moreover, the tile-based IBC analysis further increases the coding efficiency for the natural sequences. The IV-PSNR BD-rate gain caused by quarter-pel accuracy was equal to 2.7%. When both tools are enabled, it is much higher: 6.9%.

The tile-based IBC analysis decreases the encoding time for computer-generated content, but the coding efficiency is worse than without enabling this tool. The reason for such a discrepancy between videos of different origins is the lower complexity of the noiseless textures in CG sequences. It is easier to find a good prediction candidate inside the tile, and the cost of this prediction is decreased.

However, the information about the content type is indirectly signaled in MIV, as the depth atlases for sequences with good quality depth maps (i.e., CG sequences) are

differently quantized as atlases for sequences with imperfect depth maps (i.e., natural content). Therefore, it is possible to enable tiling for natural sequences and disable this tool for CG ones.

As presented in Table 4, enabling the tile-based IBC analysis for the first texture atlas increases coding efficiency while decreasing computational time. In Table 5, the results of two additional experiments are presented. In these experiments, the tile-based IBC analysis was also enabled for other video streams: texture atlases in experiment 5, and all atlases, including depth, in experiment 6.

Table 5. Basic SCC vs. SCC with quarter-pel IBC accuracy and tile-based IBC analysis for other atlases; BD-rates for two used quality metrics (WS-PSNR and IV-PSNR, a negative value indicates increased coding efficiency) and increase of the encoding time when compared to basic SCC (value higher than 100% indicates longer computational time); BD-rate change higher than 2% was highlighted in red (for efficiency loss) or green (for efficiency gain).

Sequence	Tile-Based IBC Analysis for both Texture Atlases			Tile-based IBC analysis for all atlases		
	BD-Rate WS-PSNR _Y	BD-Rate IV-PSNR	Δ Encoding Time	BD-Rate WS-PSNR _Y	BD-Rate IV-PSNR	Δ Encoding Time
ClassroomVideo	3.7%	−4.5%	39.7%	9.0%	−1.8%	41.6%
Museum	2.2%	2.9%	47.3%	4.4%	4.7%	50.9%
Fan	1.6%	2.5%	51.8%	6.4%	6.8%	43.6%
Kitchen	3.8%	4.9%	57.6%	8.4%	8.7%	63.4%
Chess	2.1%	0.9%	56.6%	7.4%	5.3%	60.4%
Group	1.7%	1.1%	46.0%	6.9%	4.5%	45.2%
ChessPieces	1.4%	0.4%	56.1%	7.7%	4.9%	59.7%
Hijack	1.8%	2.8%	58.0%	8.5%	8.9%	53.7%
Mirror	4.3%	4.6%	55.8%	7.0%	7.1%	54.0%
CG: average	2.5%	1.7%	52.1%	7.3%	5.5%	52.5%
Painter	−1.9%	−4.0%	52.7%	2.0%	−0.2%	39.0%
Frog	−0.1%	−5.6%	44.4%	3.3%	−2.6%	33.7%
Carpark	−5.6%	−7.9%	63.8%	−2.6%	−5.1%	55.7%
Fencing	−0.7%	−3.0%	54.0%	4.6%	1.9%	44.8%
Hall	−6.3%	−7.2%	71.8%	3.5%	2.5%	54.0%
Street	−10%	−12%	66.3%	−6.2%	−8.0%	58.4%
NC: average	−4.1%	−6.6%	58.8%	0.8%	−1.9%	47.6%
All: average	−0.1%	−1.6%	54.8%	4.7%	2.5%	50.5%

The average results gathered in Table 4 are presented as RD curves in Figure 4. The proposed approach outperforms the basic MIV + HEVC-SCC scenario for natural content by decreasing the total bitrate and preserving the quality. For computer-generated content (right side of Figure 4), the RD curves are very close, indicating that the proposed approach does not significantly change coding efficiency.

Natural content	Computer-generated content
-----------------	----------------------------

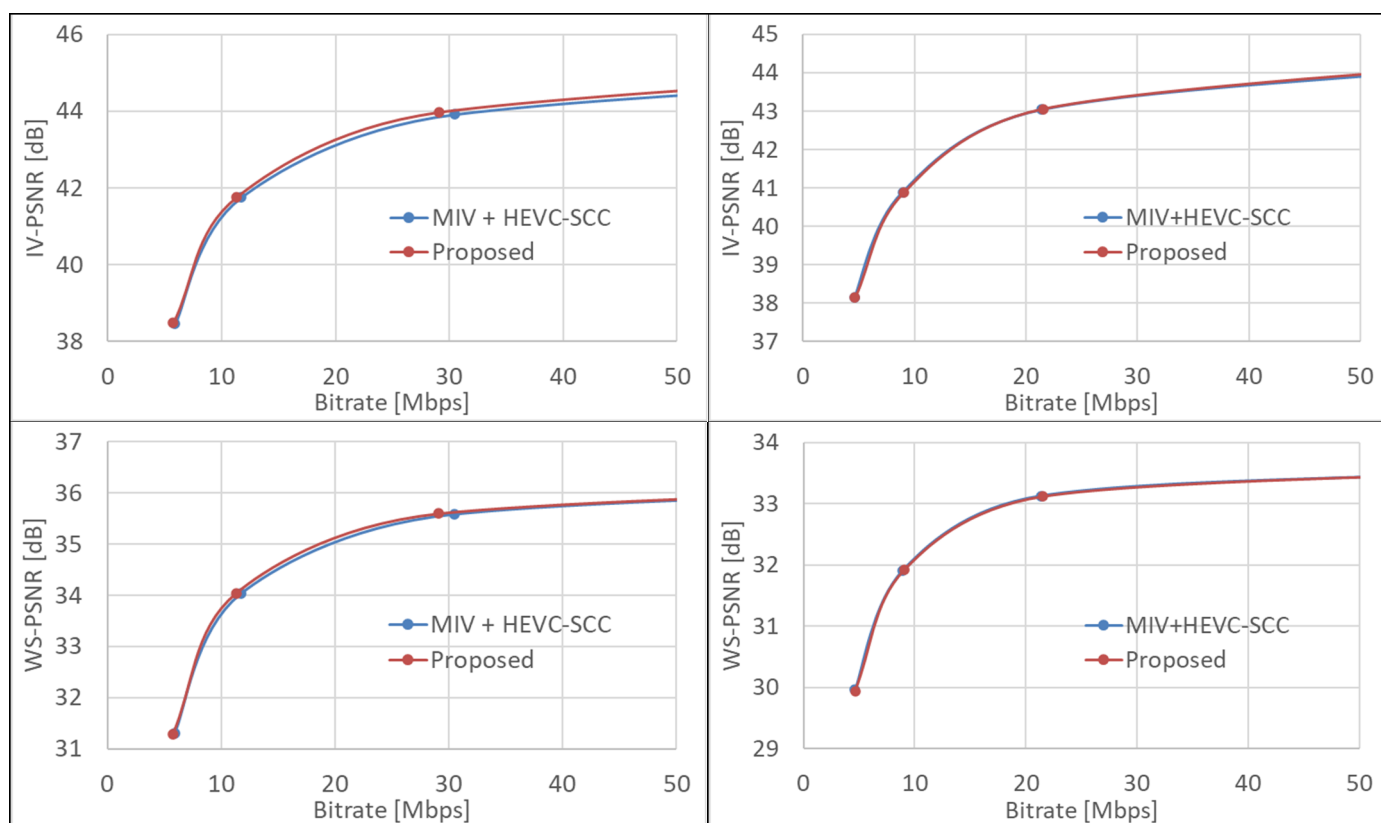


Figure 4. IV-PSNR and WS-PSNR RD-curves for natural content (left) and computer-generated content (right); comparison of basic HEVC-SCC and proposed SCC with quarter-pel IBC accuracy and tile-based IBC analysis for first texture atlas.

The results in Table 5 show that the tile-based IBC analysis is much less effective when used for the second atlas. Such results were expected, as the atlas does not contain separate full views in each tile but much smaller, randomly arranged patches (Figure 1). Therefore, the proposed tool reduces the efficiency of the IBC because its vectors are searched in different tiles; thus, the neighboring patches are omitted, despite having similar textures.

The results are even worse if the tile-based patch analysis is also used for depth atlases. In general, this tool does not increase the efficiency of depth coding. The reason for this is the entirely different characteristics of depth maps when compared to the texture. Depth maps (thus also depth atlases) simultaneously contain sharp edges between objects and large, smooth, self-similar areas, which often include constant depth values. Therefore, it is much more efficient not to restrict the IBC vector searching because the most similar blocks are often arranged close to each other, so very short IBC vectors are sufficient. However, as presented in Table 5, tile-based IBC analysis performed for all video streams decreases the encoding time twice. Therefore, it can be useful for specific use cases that require fast encoding, such as live streaming.

6. Conclusions

The conducted experiments prove that parsing the MIV bitstream is crucial for efficient encoding using the proposed modifications of HEVC SCC. Without the parsing, it would not be possible to differentiate texture and depth atlases or first and second texture atlas. If the additional information about the atlas is decoded, the SCC encoder can perform the proposed tile-based IBC analysis and quarter-pel IBC only for the data for which they are genuinely efficient.

The proposed modifications of the HEVC SCC encoder increased the encoding efficiency of natural sequences by 5–7% (depending on the analyzed quality metric, cf.

Table 4). The proposed tile-based IBC starts the block matching from the collocated position in the previously encoded tile instead of the neighboring block, making the search significantly faster. Even if the second proposed technique of sub-pixel block matching requires additional interpolation, the total encoding time is reduced by more than 15%.

The outstanding performance of the proposal in encoding natural content sequences is very valuable for further research on immersive video coding, as compressing such sequences is much more difficult than in the case of computer-generated content with almost-perfect depth maps. Moreover, the simultaneous, significant reduction of encoding time makes the proposal even more valuable, keeping in mind that any compound processing of multiview video is a very time-consuming, complex task.

The studies presented in this paper are based on the HEVC encoder and its extension. However, in the future, similar research can be performed for newer, more efficient video coding standards, such as VVC or AV1.

Author Contributions: Conceptualization, J.S., A.D., and D.M.; methodology, A.D. and D.M.; software, J.S. and A.D.; validation, J.S.; investigation, D.M.; writing—original draft preparation, review, and editing, J.S., A.D., and D.M.; visualization, A.D.; All authors have read and agreed to the published version of the manuscript.

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

Data Availability Statement: The research was conducted under ISO/IEC JTC1/SC29 WG04 MPEG Video Coding Common Test Conditions for MPEG Immersive Video: <https://mpeg-miv.org/>, accessed on 2022-11-01.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Vadakital, V.K.M.; Dziembowski, A.; Lafruit, G.; Thudor, F.; Lee, G.; Rondao Alface, P. The MPEG immersive video standard – current status and future outlook. *IEEE MultiMed.* **2022**, *23*, 101–111.
- Boyce, J.; Doré, R.; Dziembowski, A.; Fleureau, J.; Jung, J.; Kroon, B.; Salahieh, B.; Vadakital, V.K.M.; Yu, L. MPEG Immersive Video coding standard. *Proc. IEEE* **2021**, *109*, 1521–1536. <https://doi.org/10.1109/JPROC.2021.3062590>.
- Sullivan, G.J.; Ohm, J.; Han, W.J.; Wiegand, T. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *22*, 1649–1668. <https://doi.org/10.1109/TCSVT.2012.2221191>.
- Szekielda, J.; Dziembowski, A.; Mieloch, D. The Influence of Coding Tools on Immersive Video Coding. In Proceedings of the 29th International Conference on Computer Graphics, Visualization and Computer Vision WSCG, Online, 17–20 May 2021. <https://doi.org/10.24132/CSRN.2021.3002.21>.
- Common Test Conditions for MPEG Immersive Video. *Standardization document: ISO/IEC JTC1/SC29/WG04 MPEG VC N0051*, Online, Jan. 2021.
- Xu, X.; Liu, S. Overview of Screen Content Coding in Recently Developed Video Coding Standards. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 839–852. <https://doi.org/10.1109/TCSVT.2021.3064210>.
- Xu, J.; Joshi, R.; Cohen, R.A. Overview of the Emerging HEVC Screen Content Coding Extension. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 50–62. <https://doi.org/10.1109/TCSVT.2015.2478706>.
- Samelak, J.; Dziembowski, A.; Mieloch, D.; Domański, M.; Wawrzyniak, M. Efficient Immersive Video Compression using Screen Content Coding. In Proceedings of the 29th International Conference on Computer Graphics, Visualization and Computer Vision WSCG, Prague, Czech Republic, 17–20 May 2021. <https://doi.org/10.24132/CSRN.2021.3002.22>.
- Samelak, J.; Domański, M. Multiview Video Compression Using Advanced HEVC Screen Content Coding. *arXiv* **2021**, arXiv:2106.13574. <https://doi.org/10.48550/arXiv.2106.13574>.
- Mieloch, D.; Dziembowski, A.; Domański, M.; Lee, G.; Jeong, J.Y. Color-dependent pruning in immersive video coding. *J. WSCG* **2022**, *30*, 91–98. <https://doi.org/10.24132/JWSCG.2022.11>.
- Shin, H.C.; Jeong, J.Y.; Lee, G.; Kakli, M.U.; Yun, J.; Seo, J. Enhanced pruning algorithm for improving visual quality in MPEG immersive video. *ETRI J.* **2022**, *44*, 73–84. <https://doi.org/10.4218/etrij.2021-0211>.
- Kim, H.H.; Lim, S.G.; Lee, G.; Jeong, J.Y.; Kim, J.G. Efficient Patch Merging for Atlas Construction in 3DoF+ Video Coding. *IEICE Trans Inf. Syst.* **2021**, *3*, 477–480. <https://doi.org/10.1587/transinf.2020EDL8119>.
- Park, D.; Lim, S.G.; Oh, K.J.; Lee, G.; Kim, J.G. Nonlinear Depth Quantization Using Piecewise Linear Scaling for Immersive Video Coding. *IEEE Access* **2022**, *10*, 4483–4494. <https://doi.org/10.1109/ACCESS.2022.3140537>.

14. Dziembowski, A.; Mieloch, D.; Domański, M.; Lee, G.; Jeong, J.Y. Spatiotemporal redundancy removal in immersive video coding. *J. WSCG* **2022**, *30*, 54–62. <https://doi.org/10.24132/JWSCG.2022.7>.
15. Salahieh, B.; Bhatia, S.; Boyce, J. Multi-Pass Renderer in MPEG Test Model for Immersive Video. In *2019 Picture Coding Symposium (PCS)*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5. <https://doi.org/10.1109/PCS48520.2019.8954515>.
16. Fachada, S.; Bonatto, D.; Xie, Y.; Rondao Alface, P.; Teratani, M.; Lafruit, G. Depth Image-Based Rendering of Non-Lambertian Content in MPEG Immersive Video. In Proceedings of the 2021 International Conference on 3D Immersion (IC3D), Brussels, Belgium, 8 December 2021; pp. 1–6. <https://doi.org/10.1109/IC3D53758.2021.9687263>.
17. Garus, P.; Henry, F.; Jung, J.; Maugey, T.; Guillemot, C. Immersive Video Coding: Should Geometry Information Be Transmitted as Depth Maps? *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 3250–3264. <https://doi.org/10.1109/TCSVT.2021.3100006>.
18. *Standard ISO/IEC DIS 23090-12*; Information Technology—Coded Representation of Immersive Media—Part 12: MPEG Immersive Video.
19. Szydelko, B.; Dziembowski, A.; Mieloch, D.; Domański, M.; Lee, G. Recursive block splitting in feature-driven decoder-side depth estimation. *ETRI J.* **2022**, *44*, 38–50. <https://doi.org/10.4218/etrij.2021-0308>.
20. Garus, P.; Henry, F.; Maugey, T.; Guillemot, C. Decoder Side Multiplane Images using Geometry Assistance SEI for MPEG Immersive Video. MMSp 2022—IEEE 24th International Workshop on Multimedia Signal Processing, Shanghai, China, 26–28 September 2022; pp. 1–6.
21. Garus, P.; Henry, F.; Maugey, T.; Guillemot, C. Motion Compensation-based Low-Complexity Decoder Side Depth Estimation for MPEG Immersive Video. In Proceedings of the MMSp 2022—IEEE 24th International Workshop on Multimedia Signal Processing, Shanghai, China, 26–28 September 2022; pp.1–6.
22. Milovanović, M.; Henry, F.; Cagnazzo, M.; Jung, J. Patch Decoder-Side Depth Estimation in Mpeg Immersive Video. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, Toronto, ON, Canada, 6–11 June 2021; pp. 1945–1949. <https://doi.org/10.1109/ICASSP39728.2021.9414056>.
23. Chen, M.; Salahieh, B.; Dmitrichenko, M.; Boyce, J. Simplified carriage of MPEG immersive video in HEVC bitstream. In *Applications of Digital Image Processing XLIV*; SPIE: Bellingham, WA, USA, 2021. <https://doi.org/10.1117/12.2593869>.
24. Lee, S.; Jeong, J.B.; Ryu, E.S. Atlas level rate distortion optimization for 6DoF immersive video compression. In Proceedings of the 32nd Workshop on Network and Operating Systems Support for Digital Audio and Video, Athlone, Ireland, 17 June 2022; pp. 78–84. <https://doi.org/10.1145/3534088.3534354>.
25. Jeong, J.B.; Lee, S.; Ryu, E.S. Delta QP Allocation for MPEG Immersive Video. In Proceedings of the 13th International Conference on ICT Convergence, Jeju, Korea, 19–21 October 2022.
26. Santamaria, M.; Vadakital, V.K.M.; Kondrad, L.; Hallapuro, A.; Hannuksela, M.M. Coding of volumetric content with MIV using VVC subpictures. In Proceedings of the IEEE 23rd International Workshop on Multimedia Signal Processing (MMSp), Tampere, Finland, 6–8 October 2021; pp. 1–6. <https://doi.org/10.1109/MMSP53017.2021.9733465>.
27. Jeong, J.B.; Lee, S.; Ryu, E.S. VVC Subpicture-Based Frame Packing for MPEG Immersive Video. *IEEE Access* **2022**, *10*, 103781–103792. <https://doi.org/10.1109/ACCESS.2022.3207154>.
28. Tsang, S.; Chan, Y.; Kuang, W. Standard-Compliant HEVC Screen Content Coding for Raw Light Field Image Coding. In Proceedings of the 13th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, Australia, 16–18 December 2019; pp. 1–6. <https://doi.org/10.1109/ICSPCS47537.2019.9008587>.
29. Samelak, J.; Stankowski, J.; Domański, M. Efficient frame-compatible stereoscopic video coding using HEVC Screen Content Coding. In proceedings of the IEEE International Conference on Systems, Signals and Image Processing IWSSIP 2017, Poznań, Poland, 22–24 May 2017. <https://doi.org/10.1109/IWSSIP.2017.7965587>.
30. Samelak, J.; Domański, M. Unified Screen Content and Multiview Video Coding—Experimental results. *Standardization document: Joint Video Exploration Team (JVET) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 13th Meeting: Doc. JVET-M0765*, Marrakech, MA, Jan. 2019.
31. Xu, X.; Liu, S.; Chuang, T.D.; Huang, Y.W.; Lei, S.M.; Rapaka, K.; Pang, C.; Seregin, V.; Wang, Y.-K.; Karczewicz, M. Intra Block Copy in HEVC Screen Content Coding Extensions. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2016**, *6*, 409–419. <https://doi.org/10.1109/JETCAS.2016.2597645>.
32. Tech, G.; Chen, Y.; Müller, K.; Ohm, J.R.; Vetro, A.; Wang, Y.K. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 35–49. <https://doi.org/10.1109/TCSVT.2015.2477935>.
33. JCT-VC, HEVC Screen Content Coding Reference Software Repository. Available online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9+SCM-8.0 (accessed on 19 September 2022).
34. Sun, Y.; Lu, A.; Yu, L. Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video. *IEEE Signal Process. Lett.* **2017**, *24*, 1408–1412. <https://doi.org/10.1109/LSP.2017.2720693>.
35. Dziembowski, A.; Mieloch, D.; Stankowski, J.; Grzelka, A. IV-PSNR—The objective quality metric for immersive video applications. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 7575–7591. <https://doi.org/10.1109/TCSVT.2022.3179575>.
36. Bjøntegaard, G. Calculation of Average PSNR Differences between RD-curves. *Standardization document: ITU-T SG16, Doc. VCEG-M33*, Austin, USA, Apr. 2001.
37. Test Model 8 for MPEG Immersive Video. *Standardization document: ISO/IEC JTC1/SC29/WG04 MPEG VC N0050*, Online, Jan. 2021.