Color-dependent pruning in immersive video coding

Dawid Mieloch [1]
dawid.mieloch@put.poznan.pl
Adrian Dziembowski [1]
Marek Domański [1]
Gwangsoon Lee [2]
Jun Young Jeong [2]

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

[2] Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea

ABSTRACT

This paper presents the color-dependent method of removing inter-view redundancy from multiview video. The pruning of input views decides which fragments of views are redundant, i.e., do not provide new information about the three-dimensional scene, as these fragments were already visible from different views. The proposed modification of the pruning uses both color and depth and utilizes the adaptive pruning threshold which increases the robustness against the noisy input. As performed experiments have shown, the proposal provides significant improvement in the quality of encoded multiview videos and decreases erroneous areas in the decoded video caused by different camera characteristics, specular surfaces, and mirror-like reflections. The pruning method proposed by the authors of this paper was evaluated by experts of the ISO/IEC JTC1/SC29/WG 11 MPEG and included by them in the Test Model of MPEG Immersive Video.

Keywords

Immersive video coding, multiview compression, virtual reality.

1. INTRODUCTION

In an immersive video, the viewer can interactively change his/her position in the three-dimensional scene, allowing virtual traversing, e.g., using a virtual reality set [Dom17]. Typically, in order to generate a virtual view, some kind of three-dimensional representation of an acquired scene has to be utilized. The most widespread is the multiview video plus depth representation (MVD) [Mül11]. The so-called depth maps are used to store, in the form of an additional grayscale video, the distance from the camera to the 3D point for each pixel of the corresponding video.

Immersive video applications are gaining recently a large interest both from the video processing researchers and from the standardization community, as dedicated compression standards are emerging [Boy21]. It makes the introduction of such kinds of services to possible consumers much easier.

Independent compression of multiple views and corresponding depth maps (e.g., using HEVC [Sul12]) results in high bitrates [Dom21]. Instead, the compression of the immersive video should take advantage of the inter-view redundancy existing in the multiview representation. For example, input representation may consist of multiple typical camera-acquired videos with vastly overlapping fields of views, or it may consist of a few overlapping omnidirectional (360-degree) videos. The redundancy resulting from the spatial overlap between input views can be used to decrease the size of data required to fully represent the whole three-dimensional scene.

Basic multiview encoders (such as MPEG's Multiview Video Coding – MVC [Mer06] or Multiview extension of High Efficiency Video Coding – MV-HEVC [Tec16]) usually utilize the inter-view prediction, based on motion vectors estimated between neighboring frames (similarly as it is done in the temporal domain in typical inter-frame prediction). However, in this approach, textures and depth maps are encoded independently, so the searching of motion vectors does not utilize the information about three-dimensional geometry of an encoded scene. This improvement was introduced in 3D-HEVC in the form of the depth-based inter-view prediction [Tec16]. While both described types of prediction highly decrease the bitrate of encoded videos (more than 40% reduction in comparison with simulcast encoding), it still requires allocating a part of the bitstream for residual data left after the prediction. Moreover, as the prediction does not fully eliminate the inter-view redundancy, the pixel-rate (understood as the number of pixels to be decoded per second to produce the requested view) is not decreased.

The current state-of-the-art technology for immersive video compression was developed by the MPEG ISO/IEC group under the MPEG Immersive video name (MIV) [ISO22]. The foundations of this standard were built on technologies presented by proponents of Call for Proposals for 3DoF+ video coding [Dom19], [Fle19]. Most of the proposals followed a similar core idea, that few base views that gather most of the information of the scene should be encoded in their entirety, while supplementary information (visible from the remaining viewpoints) can be transmitted in the form of a mosaic of patches, collectively forming an atlas. Fig. 1 presents an example of this representation: an atlas with base views is presented on the left, an atlas with patches is in the middle, while on the right two corresponding atlases with depth maps can be seen. These atlases are then encoded using a typical video encoder, e.g. HEVC [Sul12], while in the decoder, the atlases are used to synthesize the requested viewpoint for a final viewer. The full description of MPEG Immersive video can be found in [Boy21].

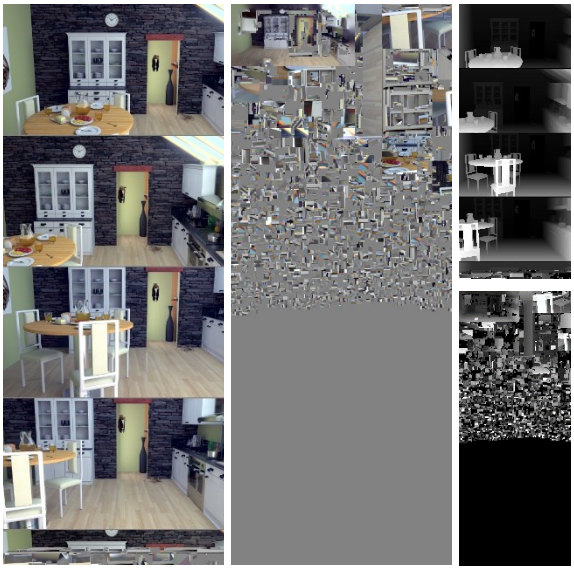

Figure 1. Atlases for sequence Kitchen [Boi18]: two texture atlases and two depth atlases (reduced resolution).

In order to decide which information should be packed into atlases, the so-called pruning process is utilized. It is based on the recognition of the inter-view redundant parts of input views and removing them from further processing (the details are described in Section 2), as these areas do not provide new information about the three-dimensional scene – these fragments were already visible from different views.

This paper introduces the pruning process improvement in which the color information is utilized in a way that minimizes the number of patches required to properly encode the non-Lambertian surfaces present in a 3D scene. The organization of the paper is as follows: Section 2 shows the description of the basic pruning process and the details of the proposed color-dependent pruning method; Section 3 includes the results of the comparison of the proposed method with the basic depth-dependent pruning. In the end, Section 4 summarizes the paper and includes conclusions drawn from the performed experiments.

2. INTER-VIEW PRUNING OF MULTIVIEW VIDEOS

Overview of the pruning process

The simplified process of pruning is shown in Fig. 2. First of all, basic views are inserted into the pruning graph (as root nodes). Then, all pixels of basic views are projected (using the depth information) to each additional view. After creating the pruning mask for each additional view, the additional view with the maximum number of preserved pixels is selected (to prefer larger patches). This selected additional view is then added to the pruning graph (as a child node of other nodes in the graph). The projection of all preserved pixels of the selected view to remaining additional views is repeated and the pruning mask is iteratively updated for each remaining additional view.

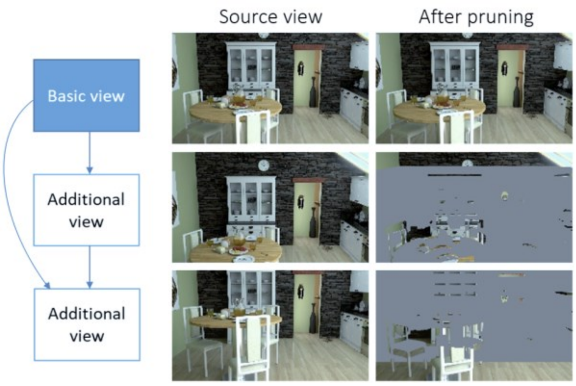

Figure 2. The idea of source views pruning.

The described process is dependent on the condition which has to be met to decide if the element should be pruned from an input view. Initially, the pruning process was based only on depth information, i.e., if the difference between the depth of the transferred point and the point corresponding to it was higher than a depth pruning threshold (10%), then the corresponding point was marked as redundant and removed. This approach eliminated the problem of noisy input views, for which it is much harder to determine the similarity between the points in neighboring views when the decision is based on color. On the other hand, when non-Lambertian surfaces are present in the scene, the depth-based pruning was leaving only one instance of such surface (only from one of the views), because the depth of a reflective surface is the same in all views, so specular reflections are irreversibly lost in the pruning.

Removal of inter-view redundancy based only on depth of a point is also encountered in methods of producing the multiview layered depth image (LDI) [Anj17] and compression of meshes [Tan18] and voxels [Käm16]. Although LDI representation is used for efficient synthesis of virtual views rather than for compression purposes, abovementioned methods show that depth-based type of inter-view redundancy check was widely considered to be state-of-the-art in multiview processing.

The proposed color-dependent pruning

In the approach proposed by the authors of this paper, two types of information are taken into account: depth and texture. Depth information is analyzed in the same way as described above, while color information is analyzed in a point-to-block comparison (Fig. 3). In this example, a pixel from view v0 is reprojected to v1. Depth similarity is being checked only for the colocated pixel (dark blue). The color of the pixel marked in orange is compared to the color of all pixels in the 3×3 neighborhood of the colocated one.

A similar approach was proposed earlier in [Mie21], where the point-to-block matching was used as an inter-view similarity metric for depth estimation purposes. This metric was shown to work especially well for highly compressed input views or when the amount of noise present in a multi-view sequence is significant. Even though the pruning is always performed on uncompressed views, the decreased influence of noise on inter-view matching is desirable and preferable in this part of multiview processing. Color-based matching of objects and points have also been shown to be efficient in single video temporal tracking [Ker10].

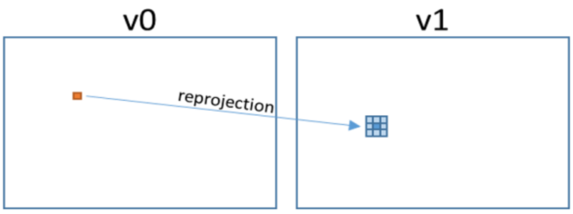

Figure 3. The idea of point-to-block similarity measurement used in color-dependent pruning.

In the proposal, if the minimum color error within a block is lower than a threshold (and the depth-based condition is also met), the pixel of view v1 is being pruned. Using of the 3×3 neighborhood instead of a larger one was based on results from [Mie21] which showed that such block size is the best for inter-view matching in multiview videos. Having this variable fixed, during the preliminary experiments, the pruning threshold was set to 4% of the bit depth of color (i.e., 40 for 10 bits per sample views).

As can be seen in Fig. 4, the same area can reflect the light differently when acquired with the camera facing different directions. The atlas generated with and without proposed color-dependent block-based pruning is presented in Fig. 5. As shown, the proposed solution allows preserving regions with different texture/lighting conditions, which were otherwise pruned.

Moreover, because of block-based characteristics, the proposed pruning is less sensitive to noise and only slightly increases the non-pruned area for such type of content. Fig. 6 presents two consecutive frames of ClassroomVideo sequence and the difference between them (to illustrate the amount of noise), while Fig. 7 presents the difference in the atlas when it is generated with color-dependent pruning.

Nevertheless, in order to further decrease the influence of noise, we included further enhancement to the proposed pruning method, which allows adapting to sequence characteristics.

The threshold for color-based pruning should be increased for noisy sequences to reduce redundancy in atlases. On the other hand, for sequences with negligible noise, the pruning threshold should be smaller to allow preserving also fragments of the scene with slight lighting inconsistencies.

In the proposed solution, the fixed pruning threshold is multiplied by a global inter-view luma standard deviation. This value is calculated for the first frame of each encoded group of pictures (GoP) as a standard deviation of a set A, which contains luma differences between inter-view corresponding pixels. GoP can be changed in the configuration of MIV encoder, by default it is equal to 32 to match the GoP used in HEVC encoder. Calculating this threshold for a whole GoP is sufficient as in MIV the results of pruning are in the end gathered for these frames.

In order to populate set A, first of all, all pixels are reprojected between all pairs of input views. For each pixel, the luma of the pixel is compared with the luma of all pixels in the 3×3 neighborhood of the pixel from another input view. If the smallest difference in this neighborhood is 0, then the inter-view correspondent pixels were found, so the luma difference between the reprojected pixel and the center of the co-located block is included in set A.

Figure 4. Two views of multiview test sequence Chess [Ilo19].

Figure 5. Atlas of patches generated for Chess sequence: depth-dependent pruning (left) vs. proposal (right).

Figure 6. Fragments of two consecutive frames of ClassroomVideo sequence and the difference between them (on the right).

Figure 7. Atlas of patches generated for ClassroomVideo sequence: depth-dependent pruning (left) vs. proposal (right).

3. EXPERIMENTAL RESULTS

Methodology and design of experiments

To evaluate the proposed color-dependent pruning, the method was implemented in TMIV 5 [MPEG20c]. Test Model for Immersive Video is implemented as a C++ project which is publicly available. All changes which are being added to the standard and have to be implemented in TMIV, are earlier accepted by its software coordinator to avoid adding low-quality code to the repository.

Two experiments were performed. The first one shows the results for encoding with color-dependent pruning without taking into account the noise characteristics, therefore, for the fixed pruning threshold. The latter experiment evaluates the encoding efficiency with an adaptive threshold automatically calculated for each encoded test sequence.

11 multiview sequences from MIV Common Test Conditions (CTC) [MPEG20a] were used in experiments. Short characteristics of sequences are presented in Table 1, while an example of input view from each sequence was shown in Figs. 8 and 9.

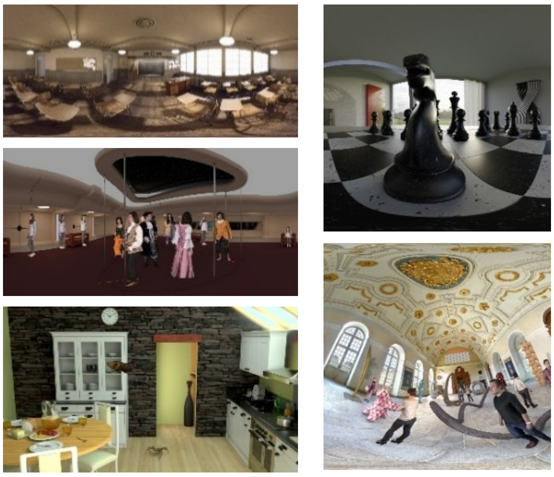

Figure 8. Computer-generated sequences. Left column: ClassroomVideo, Hijack, Kitchen; right column: Chess, Museum.

Figure 9. Natural sequences. Left column: Carpark, Street, Frog; right column: Hall, Fencing, Painter.

Test sequences were encoded with 5 different rate points (RP) using the default configuration of the TMIV encoder, described in the CTC document [MPEG20a]. The final performance was measured using two objective quality metrics: WS-PSNR [Sun17], and IV-PSNR [MPEG20b]. The first of these metrics is based on the peak signal-to-noise ratio but it was adapted to take into account the spherical projection of 360-degree videos, while the latter is adapted to measure the distortion introduced by the virtual view synthesis process.

The results are shown as Bjøntegaard delta (BD-rate [Bjo01]) change for low and high bitrates. The Bjøntegaard delta shows the percentage change in the bitrate required to achieve the same quality for two tested techniques. It was calculated both for the four smallest QPs (high bitrates – High-BR BD-rate) and for the four largest ones (low bitrates – Low-BR BD-rate). A gain or a loss larger than 3% is indicated by a green or red cell respectively.

Sequence name	Views	Type	Resolution	Source	Pruning-related sources of difficulty
Carpark	9	NC/Persp.	1920×1088	[Mie20]	Reflections on cars
Chess	10	CG/ERP	2048×2048	[Ilo19]	Reflections on the floor
Classroom Video	16	CG/ERP	4096×2048	[Kro18]	Highly noticeable noise
Fencing	10	NC/Persp.	1920×1080	[Dom16]	Not inter-view consistent color characteristics
Frog	13	NC/Persp.	1920×1080	[Sal18]	Not inter-view consistent color characteristics
Hall	9	NC/Persp.	1920×1088	[Mie20]	Reflections on the floor
Hijack	10	CG/ERP	4096×2048	[Dor18]	Reflections on clothes
Kitchen	25	CG/Persp.	1920×1080	[Boi18]	Reflections and transparency of kitchen objects
Museum	24	CG/ERP	2048×2048	[Dor18]	-
Painter	16	NC/Persp.	2048×1088	[Doy18]	-
Street	9	NC/Persp.	1920×1088	[Mie20]	Reflections on cars

Table 1. Test sequences used in experiments: ERP – Equirectangular Projection, CG – Computer-Generated, NC – Natural Content.

The evaluation of color-dependent pruning with a fixed threshold

The results for encoding with color-dependent pruning with a fixed threshold in comparison with the pruning dependent only on depth are presented in Table 2. As it can be observed, for most test sequences, the proposed pruning method significantly decreases the BD-rate, i.e., the same quality of a final decoded image can be obtained for reduced bitrate. On average, the bitrate is decreased by 20% for high bitrates.

Fig. 10 shows the plot of the peak signal to noise ratio (PSNR) for different bitrates of encoded Carpark sequence, for which the results are the most similar to the averaged ones. It can be seen that the bitrate was slightly increased for all 5 rate points, what is the result of the increased number of patches observed in atlases. However, adding these patches caused a significant increase in the quality of the encoded video, what, in the end, results in increased compression efficiency.

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
Carpark	-19.8%	-8.6%	-17.2%	-6.7%
Chess	-64.0%	-46.9%	-61.8%	-47.8%
ClassroomVideo	6.8%	15.4%	9.7%	15.3%
Fencing	---	---	---	-39.4%
Frog	-32.3%	11.5%	-23.1%	14.7%
Hall	-51.1%	-33.5%	-40.8%	-29.4%
Hijack	-19.8%	-10.2%	-24.9%	-15.9%
Kitchen	-33.1%	-9.0%	-35.1%	-14.0%
Museum	0.7%	1.5%	0.2%	1.2%
Painter	1.3%	2.0%	1.7%	2.0%
Street	-10.9%	-2.5%	0.5%	4.3%
Average	**-20.2%**	**-7.3%**	**-17.3%**	**-10.5%**

Table 2. BD-rate savings of MIV encoding with color-dependent pruning for a fixed threshold over MIV encoding with depth-dependent pruning.

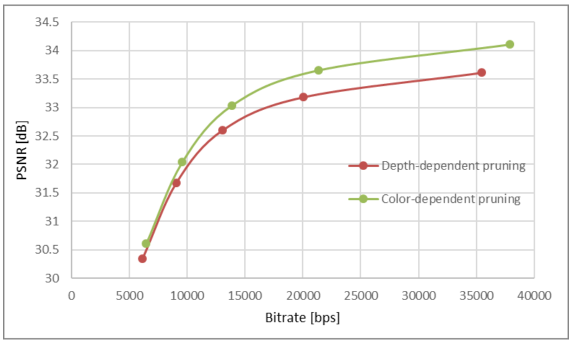

Figure 10. PSNR values for Carpark sequence encoded for 5 rate points using MIV encoder with depth-dependent pruning and with MIV encoder with color-dependent pruning.

As it was shown in Table 3, the proposal increases the runtime of MIV encoding and decoding, what is the result of the increased number of patches included in atlases. The runtime of HEVC encoding was increased insignificantly.

Sequence	MIV encoding	HEVC encoding	MIV decoding
Carpark	100.5%	107.5%	103.0%
Chess	109.4%	103.4%	109.6%
ClassroomVideo	103.4%	95.0%	102.5%
Fencing	101.8%	107.2%	101.4%
Frog	113.3%	108.2%	115.3%
Hall	101.2%	93.2%	102.8%
Hijack	106.9%	115.0%	109.3%
Kitchen	110.7%	103.1%	109.2%
Museum	112.3%	105.0%	114.2%
Painter	101.2%	104.7%	102.0%
Street	103.4%	94.9%	105.5%
Average	106.9%	101.9%	107.7%

Table 3. Runtime ratio of MIV and HEVC encoding and MIV decoding with color-dependent pruning for a fixed threshold over depth-dependent threshold.

The evaluation of color-dependent pruning with an adaptive threshold

The adaptive pruning threshold is based on the global inter-view luma standard deviation, calculated at the beginning of the encoding. The values of standard deviations for tested sequences and the resulting pruning thresholds can be found in Table 4.

Table 5 presents the efficiency of adaptive color-dependent pruning in comparison with fixed-threshold color-dependent pruning. The introduction of an adaptive threshold further improves the compression efficiency for almost all test sequences.

Fig. 11 presents the fragments of the final encoded views for the encoder with depth-dependent pruner (left) and adaptive color-dependent pruner (right). These results prove that the proposal increases the quality of video presented to a viewer by eliminating the errors caused not only by specular surfaces (see Chess sequence), and mirror-like reflections (Hall) but also by different camera characteristics (Fencing). The possibility of increasing the quality in cases where camera color characteristics are not matched can be seen as surprising, as the presented color-based pruning is based only on luma value, however, results presented in [Dzi21] show that inter-view and temporal fluctuations of luma and chromas are correlated in natural sequences.

Detailed results for Painter indicate that the PSNR was very slightly increased (0.03 dB), but the bitrate was increased by 2.5%. This sequence provided very high quality even without the proposal and does not include any reflective surfaces, what explains lower quality, as with the proposal we increased the amount of sent data and increased quality to a very minor degree.

Sequence	Calculated standard deviation	Calculated pruning threshold for 10 bps video
Carpark	0.9337	38
Chess	0.3891	16
ClassroomVideo	0.9555	39
Fencing	0.8132	33
Frog	1.6505	68
Hall	0.2664	11
Hijack	0.2084	9
Kitchen	0.8698	36
Museum	0.8711	36
Painter	0.5670	23
Street	0.8560	35

Table 4. The average global sequence inter-view luma standard deviation and calculated pruning threshold for used dataset.

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
Carpark	-0.7%	0.0%	0.0%	0.3%
Chess	-37.3%	-4.9%	-25.4%	-0.3%
ClassroomVideo	0.4%	0.6%	1.1%	1.0%
Fencing	-16.8%	-5.3%	-6.9%	-0.7%
Frog	-5.5%	-8.3%	-7.4%	-9.4%
Hall	-34.4%	-12.4%	-10.9%	1.4%
Hijack	-11.7%	5.0%	-18.0%	3.2%
Kitchen	-20.4%	-10.8%	-18.7%	-10.5%
Museum	0.5%	0.5%	0.3%	0.4%
Painter	3.2%	5.2%	4.1%	5.6%
Street	-9.4%	-2.8%	-7.9%	-2.6%
Average	**-12.0%**	**-3.0%**	**-8.1%**	**-1.0%**

Table 5. BD-rate savings of MIV encoding with color-dependent pruning for an adaptive threshold over MIV encoding with color-dependent pruning for a fixed threshold.

Table 6 shows the increase of runtime in comparison with the fixed threshold. As it can be observed, the increase of the runtime is correlated with the adaptive threshold – if the threshold is small, then the time of processing is increased, as the number of patches included in atlases is further increased. However, as indicated by the results of the quality of decoded views, these additional patches are required to achieve high-quality reconstruction in the decoder.

Sequence	MIV encoding	HEVC encoding	MIV decoding
Carpark	96.9%	96.9%	99.8%
Chess	118.8%	118.9%	117.0%
ClassroomVideo	104.4%	101.7%	103.4%
Fencing	94.3%	94.1%	99.0%
Frog	95.5%	95.2%	95.1%
Hall	146.0%	146.3%	148.5%
Hijack	122.9%	123.4%	121.6%
Kitchen	108.6%	107.1%	109.7%
Museum	123.1%	119.8%	123.4%
Painter	144.9%	145.6%	149.9%
Street	106.4%	106.2%	111.0%
Average	124.8%	123.8%	128.1%

Table 6. Runtime ratio of MIV and HEVC encoding and MIV decoding with color-dependent pruning for an adaptive threshold over color-dependent pruning for a fixed threshold.

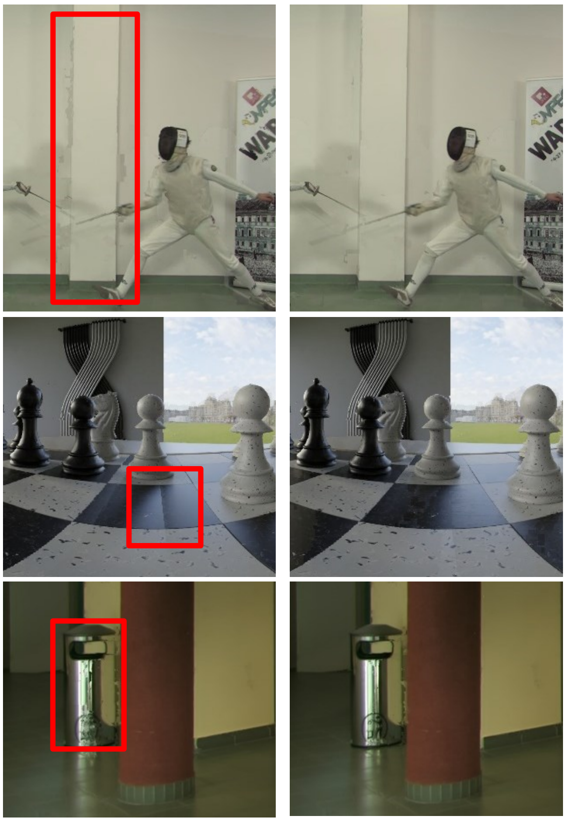

Figure 11. The fragments of decoded views of Fencing, Chess, and Hall sequences for MIV with depth-dependent pruning (left) and MIV with color-dependent pruning (right).

4. CONCLUSIONS AND FUTURE WORKS

This paper presented the color-dependent method of removing inter-view redundancy from multiview video. In order to decide which fragments of views are redundant, i.e., do not provide new information about the three-dimensional scene, as these fragments were already visible from different views, the proposal uses both color and depth. The paper introduces the adaptive pruning threshold which increases the robustness against the noisy input.

The proposed pruning method was shown to provide significant improvement in the quality of videos encoded using a modified Test Model of MPEG Immersive video codec. The largest improvement was due to the elimination of errors caused by different camera characteristics, specular surfaces, and mirror-like reflections, therefore, the proposed method highly increases the efficiency of encoding for complex multiview sequences.

The pruning method proposed by the authors of this paper was evaluated by experts of the ISO/IEC JTC1/SC29/WG 11 MPEG and included by them in the next version Test Model. As works on the new edition of MPEG Immersive video coding standard are starting in mid-2022, the authors of the proposal are planning to evaluate the color-based pruning in this new framework. MIV ed. 2 is planned to include the step of determining the non-Lambertian areas in the scene [MPEG22], so the proposed method can be potentially used for this application.

5. ACKNOWLEDGMENTS

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory).

6. REFERENCES

[Anj17] Anjos R., Madeiras Pereira J., Gaspar J., Fernandes C. Multiview layered depth image. Journal of WSCG, vol. 25. pp.115-122, 2017.

[Bjo01] Bjøntegaard, G. Calculation of average PSNR differences between RD986 curves. ISO/IEC JTC1/SC29/WG11, MPEG/M15378, Austin, TX, 2001.

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

[Boy21] Boyce J., et al. MPEG Immersive Video Coding Standard. Proceedings of the IEEE, vol. 109, no. 9, pp. 1521-1536, Sep. 2021.

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

[Dom17] Domański M., Stankiewicz O., Wegner K. and Grajek T. Immersive visual media – MPEG-I: 360 video, virtual navigation and beyond. 2017 IEEE International Conference on Systems, Signals and Image Processing (IWSSIP), Poznań, pp. 1–9, 2017.

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

[Dom21] Domański M., Al-Obaidi Y., and Grajek T. Universal Modeling of Monoscopic and Multiview Video Codecs with Applications to Encoder Control. 2021 IEEE International Conference on Image Processing (ICIP), pp. 2144-2148, 2021.

[Dor18] Doré R. Technicolor 3DoF+ Test Materials. Document ISO/IEC JTC1/SC29/WG11 MPEG/M42349, San Diego, CA, USA, Apr. 2018.

[Doy18] Doyen D., Boisson G., and Gendrot R. [MPEG-I Visual] New Version of the Pseudo-Rectified Technicolorpainter Content. Document ISO/IEC JTC1/SC29/WG11 MPEG/M43366, Ljublana, 2018.

[Dzi21] Dziembowski A., Mieloch D., Różek S., Domański M. Color Correction for Immersive Video Applications. IEEE Access, vol. 9, pp. 75626-75640, 2021.

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

[Ilo19] Ilola L., Vadakital V.K.M., Roimela K., and Keränen. New test content for Immersive Video – Nokia Chess. Document ISO/IEC JTC1/SC29/WG11 MPEG/M50787, Geneva, Switzerland, Oct. 2019.

[ISO22] Standard ISO/IEC FDIS 23090-12. Information technology – Coded representation of immersive media – Part 12: MPEG Immersive video. 2022.

[Käm16] Kämpe V., et al. Exploiting coherence in time-varying voxel data. Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D '16). pp. 15–21, 2016.

[Ker10] Kerdvibulvech C. Real-time augmented reality application using color analysis. 2010 IEEE Southwest Symposium on Image Analysis & Interpretation (SSIAI), pp. 29-32, 2010.

[Kro18] Kroon B. Test sequence ClassroomVideo. Document ISO/IEC JTC1/SC29/WG11 MPEG/M42415, San Diego, CA, USA, Apr. 2018.

[Mer06] Merkle P., Muller K., Smolic A., and Wiegand T. Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC. 2006 IEEE International Conference on Multimedia and Expo, pp. 1717-1720, 2006.

[Mie20] Mieloch D., Dziembowski A., and Domański M. [MPEG-I Visual] Natural Outdoor Test Sequences. Document ISO/IEC JTC1/SC29/WG11 MPEG/M51598, Brussels, Jan. 2020.

[Mie21] Mieloch D., Klóska D., Woźniak, M. Point-to-Block Matching in Depth Estimation, International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2021.

[MPEG20a] Common Test Conditions for Immersive Video. Document ISO/IEC JTC1/SC29/WG11 MPEG, N19214, Alpbach, Austria, Apr. 2020.

[MPEG20b] Software manual of IV-PSNR for Immersive Video. Document ISO/IEC JTC1/SC29/WG04 MPEG VC, N0013, Online, Oct. 2020.

[MPEG20c] Test Model 5 for Immersive Video. Document ISO/IEC JTC1/SC29/WG11 MPEG, N19213, Online, Apr. 2020.

[MPEG22] Use cases and requirements for MIV. Edition-2 (final), ISO/IEC JTC1/SC29/WG02 MPEG Technical requirements, N00157, Online, January 2022.

[Mül11] Müller K., Merkle P., and Wiegand T. 3-D Video Representation Using Depth Maps. 2011 Proceedings of the IEEE, vol. 99, no. 4, pp. 643-656, 2011.

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

[Sul12] Sullivan G.J., Ohm J., Han W., and Wiegand T. Overview of the High Efficiency Video Coding (HEVC) Standard. IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[Sun17] Sun Y., Lu A., and Yu L. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. IEEE Signal Processing Letters, vol. 24, no. 9, pp. 1408-1412, Sep. 2017.

[Tan18] Tang D., et al. Real-time compression and streaming of 4D performances. ACM Trans. Graph, vol. 37, no. 6, pp 1-11, Dec. 2018.

[Tec16] Tech G., et al. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, Jan. 2016.