INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG04
MPEG VIDEO CODING

ISO/IEC JTC1/SC29/WG04 MPEG/M56369
April 2021, Online

Source	Poznań University of Technology, Poznań, Poland Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea
Status	Input document
Title	MIV CE2.11: Encoder-side rendering
Author	Adam Grzelka, Adrian Dziembowski, Dawid Mieloch, Marek Domański (PUT) Gwangsoon Lee, Jun Young Jeong (ETRI)

Abstract

This document presents a description of the continuation of the PUT/ETRI experiment on encoder-side rendering (MIV CE2.11). In the proposed solution, the only base view is created by synthesizing a full-ERP view at the position of one of the input views. Thus, the central part of the base view contains non-reprojected information from one view, while both sides of the view are being reprojected from all other views. The rest of the input views are additional ones. The proposed solution is intended for omnidirectional content and decreases bitrate while preserving similar subjective quality.

1 Proposed technique

In TMIV8, the base views are chosen from input views. Such an approach has two major flaws, dependent on the video resolution and projection type combined with camera arrangement.

If the resolution of input views is small (e.g. FullHD), there are many base views. In this case, the inter-view redundancy between these views is not reduced.

For omnidirectional sequences captured by several semispherical cameras, there are not enough base views to capture the entire scene. In this case, there are many very big patches, what can result in the appearance of long, irregular fake edges in the synthesized views (edges between different patches), e.g. in SN sequence.

In the proposed approach, there is always only one base view, acquired by a virtual camera placed at the position of one of the input views. However, the field of view of the virtual camera is higher than the FOV of the input one in order to allow the virtual camera capturing of the whole scene (Figs. 1 and 2). Sequences captured by full-ERP cameras are processed in the same way, as in TMIV8.

Remaining information (i.e. occlusions from all the views) are processed in a typical way because all the input views are labeled as additional ones. Such an approach increases the list of views by one.

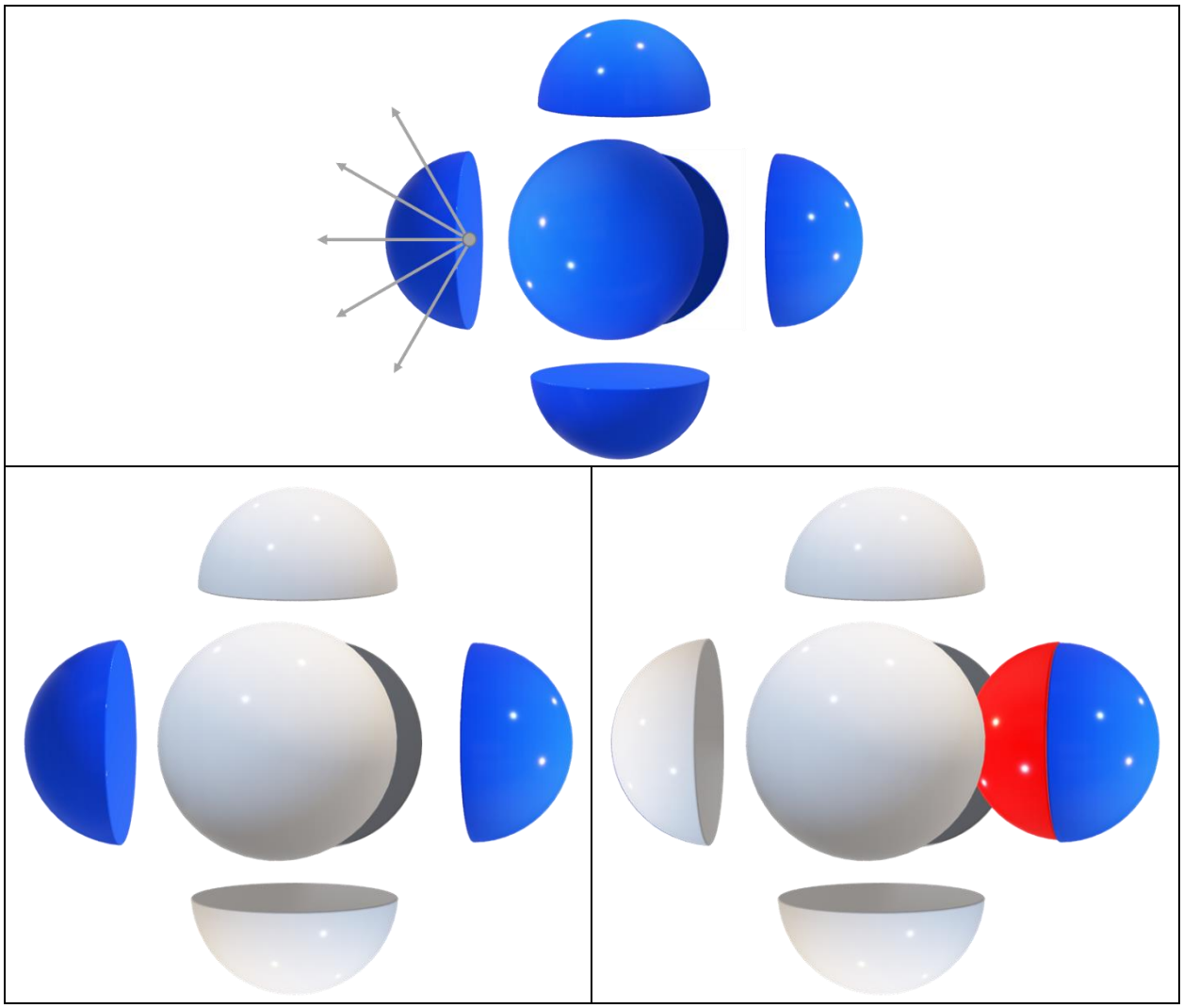


Fig. 1. Top row: input half-ERP views (cameras directed outwards, e.g. like in SN), Bottom left: basic (blue) and additional (white) views in TMIV8. Bottom right – proposed: basic full-ERP view containing original information (blue) and resynthesized information (red).

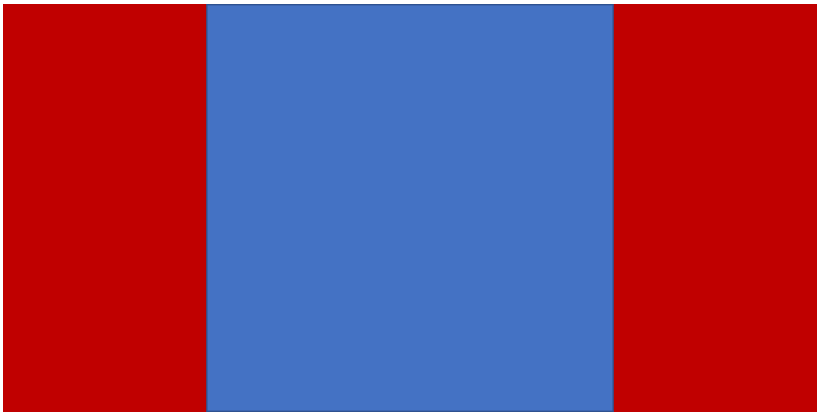


Fig. 2. Basic view, containing original information from input view (blue) and resynthesized information from other views (red).

The base view is chosen based on the analysis of the depth range of all the views. The view with the highest depth range is chosen as the basic one, as it probably contains the most

desirable/interesting information for the viewer. If several views have the same depth range, the one chosen by TMIV's view selector/labeler is selected among them.

2 Experimental results – A97

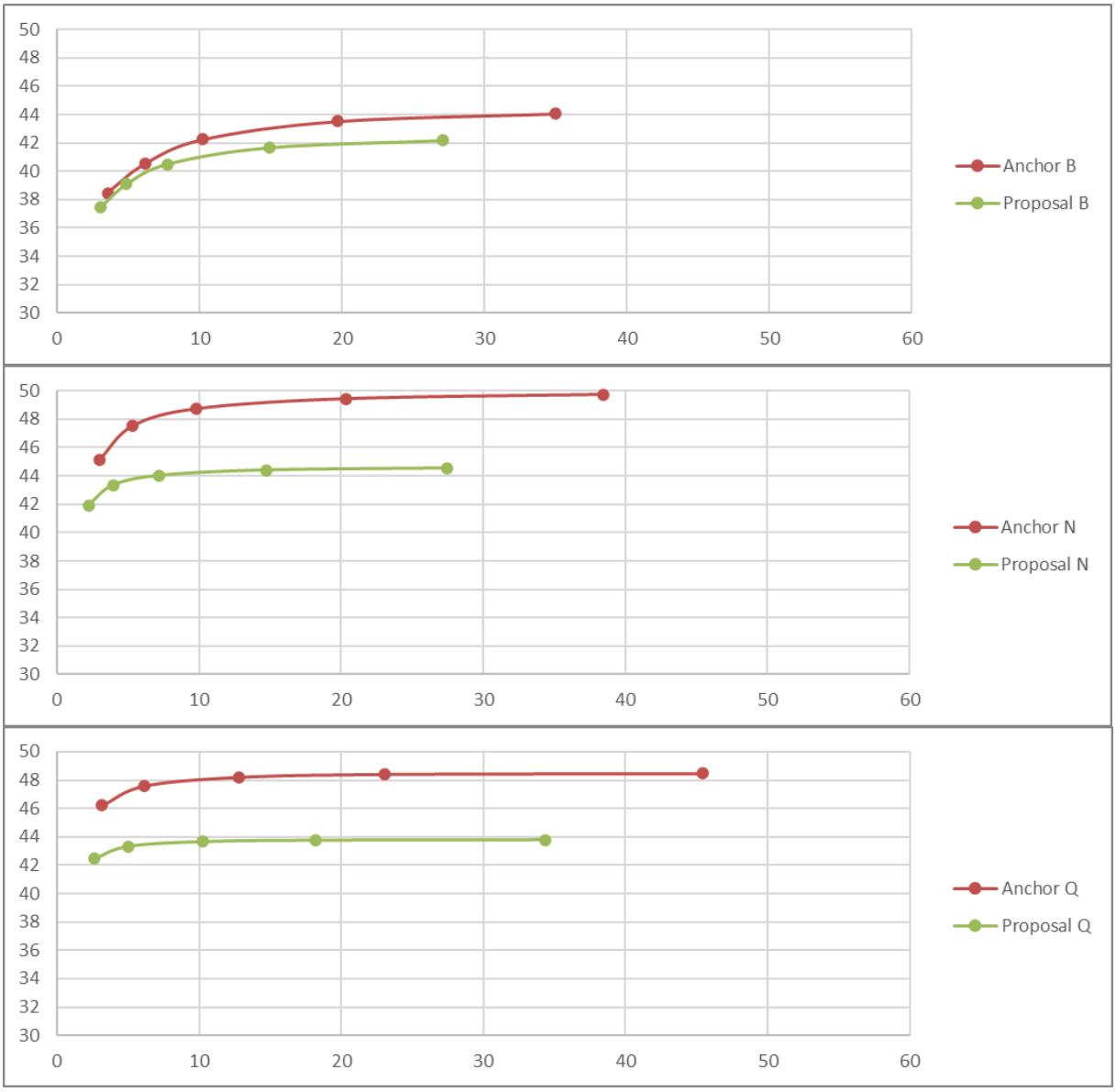


Fig. 3. IVPSNR BD-rates for sequences captured by half-ERP cameras.

Comments:

- the total bitrate was reduced by 25% for QP1 and 20% for QP5,
- characteristics and quantity of artifacts seem to be similar in the anchor and proposed approach,
- arrangement of artifacts is different because of different basic views/patches,
- the experiment was not launched for SA, because the input views are already full-ERP, thus nothing would change compared to TMIV8.

Table 1. Bitrate comparison: anchor (left) vs. proposed (right).

B									B									
Test		Bitrate [Mbps]				Fraction [%]			Test		Bitrate [Mbps]				Fraction [%]			%
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	
B	QP1	32.826	2.084	0.067	34.978	94%	6%	0%	B	QP1	24.780	2.246	0.075	27.101	91%	8%	0%	-22.5%
B	QP2	17.994	1.635	0.067	19.697	91%	8%	0%	B	QP2	13.009	1.815	0.075	14.899	87%	12%	1%	-24.4%
B	QP3	8.899	1.286	0.067	10.253	87%	13%	1%	B	QP3	6.216	1.460	0.075	7.750	80%	19%	1%	-24.4%
B	QP4	4.962	1.125	0.067	6.154	81%	18%	1%	B	QP4	3.485	1.299	0.075	4.859	72%	27%	2%	-21.0%
B	QP5	2.533	0.950	0.067	3.550	71%	27%	2%	B	QP5	1.886	1.091	0.075	3.052	62%	36%	2%	-14.0%
N									N									
Test		Bitrate [Mbps]				Fraction [%]			Test		Bitrate [Mbps]				Fraction [%]			%
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	
N	QP1	35.732	2.635	0.094	38.462	93%	7%	0%	N	QP1	25.383	1.973	0.075	27.430	93%	7%	0%	-28.7%
N	QP2	17.644	2.635	0.094	20.373	87%	13%	0%	N	QP2	12.675	1.973	0.075	14.722	86%	13%	1%	-27.7%
N	QP3	7.643	2.022	0.094	9.759	78%	21%	1%	N	QP3	5.549	1.521	0.075	7.144	78%	21%	1%	-26.8%
N	QP4	3.593	1.607	0.094	5.295	68%	30%	2%	N	QP4	2.654	1.209	0.075	3.938	67%	31%	2%	-25.6%
N	QP5	1.557	1.300	0.094	2.951	53%	44%	3%	N	QP5	1.168	0.987	0.075	2.230	52%	44%	3%	-24.4%
Q									Q									
Test		Bitrate [Mbps]				Fraction [%]			Test		Bitrate [Mbps]				Fraction [%]			%
Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	Anchor	Test point	*Texture*	*Depth*	*Metadata*	*Total*	*Texture*	*Depth*	*Metadata*	
Q	QP1	42.802	2.539	0.096	45.436	94%	6%	0%	Q	QP1	32.265	2.029	0.077	34.371	94%	6%	0%	-24.4%
Q	QP2	20.450	2.539	0.096	23.084	89%	11%	0%	Q	QP2	16.107	2.029	0.077	18.213	88%	11%	0%	-21.1%
Q	QP3	10.174	2.539	0.096	12.809	79%	20%	1%	Q	QP3	8.170	2.029	0.077	10.276	80%	20%	1%	-19.8%
Q	QP4	4.189	1.848	0.096	6.133	68%	30%	2%	Q	QP4	3.405	1.488	0.077	4.970	69%	30%	2%	-19.0%
Q	QP5	1.700	1.358	0.096	3.154	54%	43%	3%	Q	QP5	1.394	1.110	0.077	2.581	54%	43%	3%	-18.2%

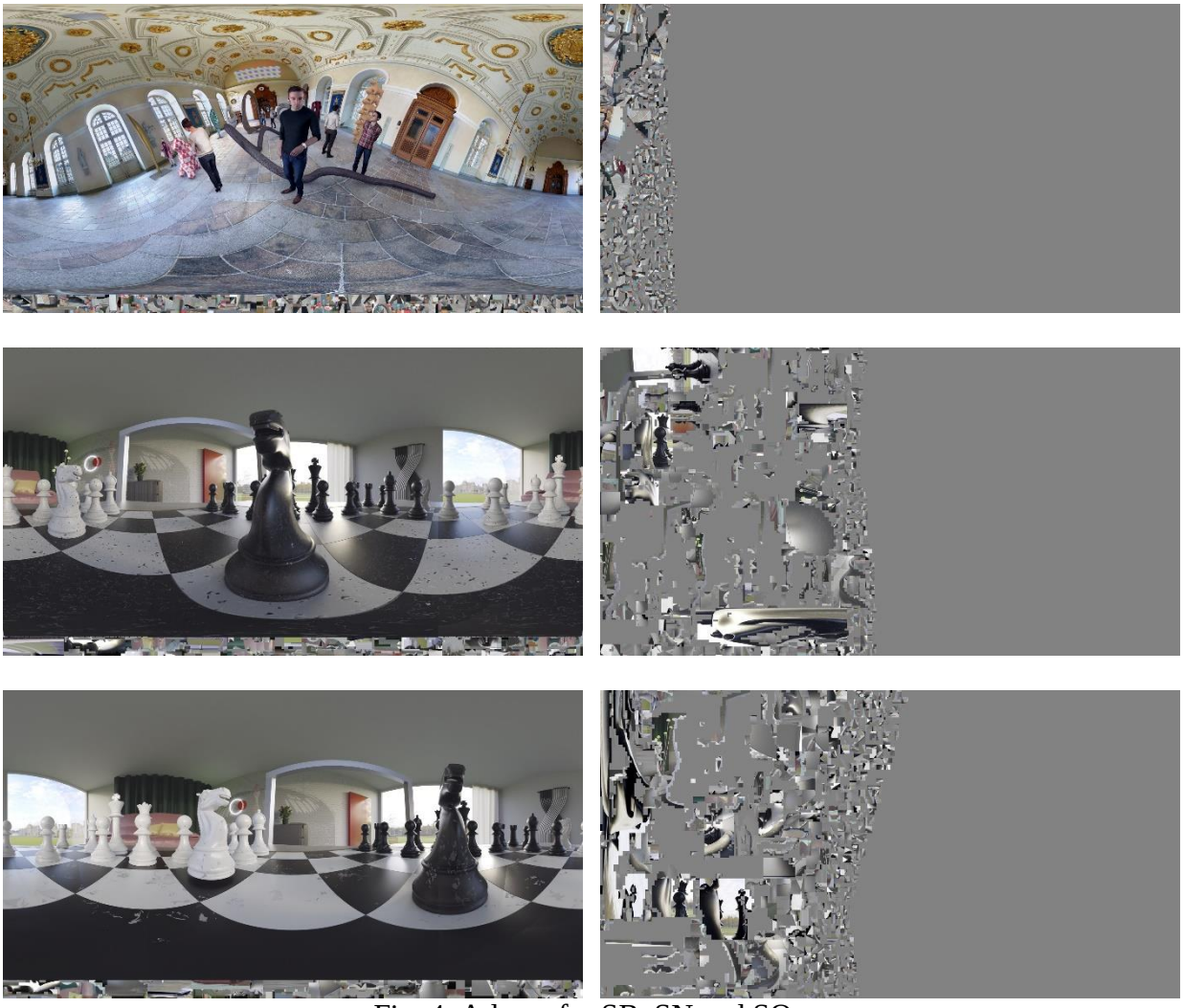
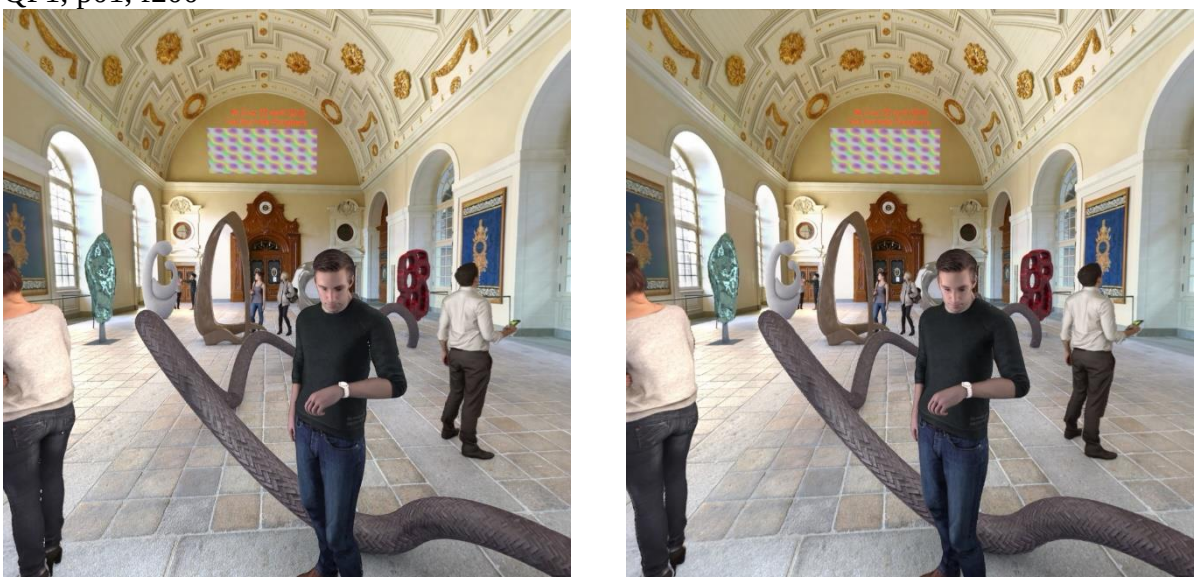
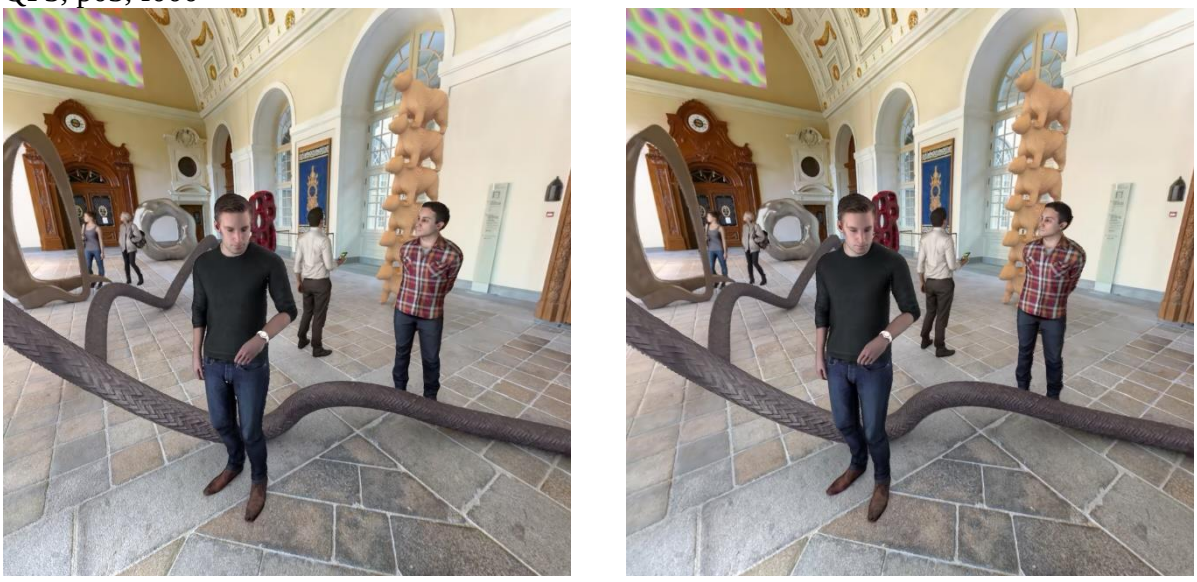


Fig. 4. Atlases for SB, SN and SQ.

QP1, p01, f200



QP3, p03, f000



QP5, p03, f200

Fig. 5. Decoded views (SB): anchor (left) vs. proposed (right).

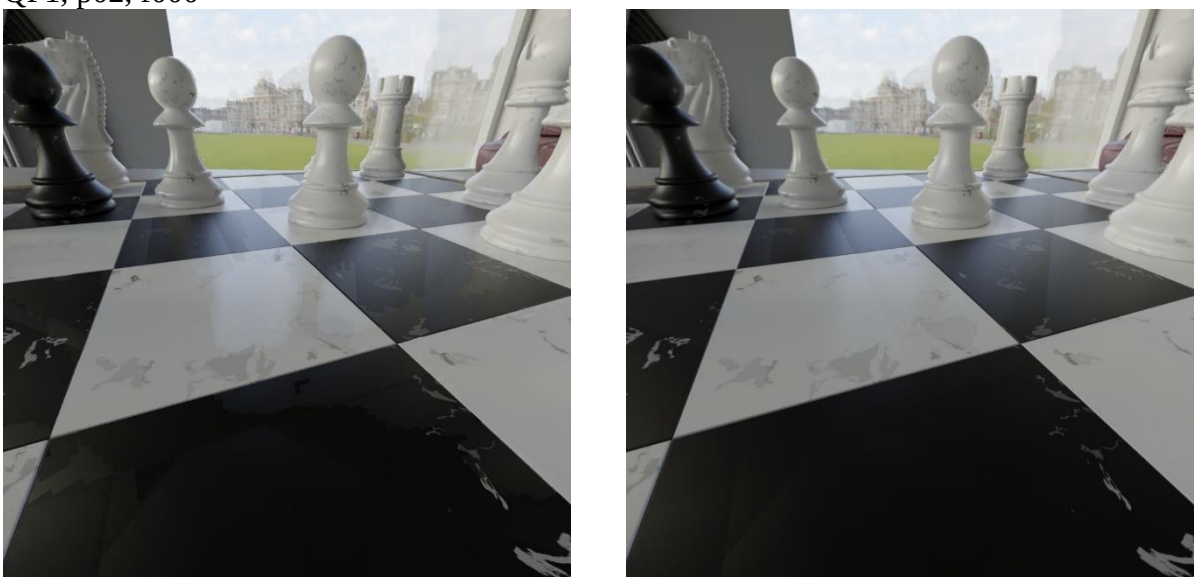
QP1, p01, f000

QP1, p01, f100

QP5, p03, f093

Fig. 6. Decoded views (SN): anchor (left) vs. proposed (right).

QP1, p02, f000



QP5, p03, f000

QP5, p03, f299

Fig. 7. Decoded views (SQ): anchor (left) vs. proposed (right).

3 Conclusions

Proposed approach:

- decreases the bitrate by reducing the inter-view redundancy occurring in TMIV8,
- preserves similar subjective quality.

4 Recommendations

We recommend to:

- continue the MIV CE2.11 to find an effective encoder-side rendering method for perspective content,
- watch provided posetraces.

5 Acknowledgement

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).