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

ISO/IEC JTC1/SC29/WG04 MPEG/M55749
January 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-related: Encoder-side rendering
Author	Adrian Dziembowski, Dawid Mieloch, Marek Domański (PUT) Gwangson Lee, Jun Young Jeong (ETRI)

Abstract

This document presents a description of the PUT/ETRI experiment on encoder-side rendering (MIV CE2.11). In the proposed solution, the only base view is the one rendered using the information from all the input views, while the rest of the input views are additional ones.

1 Proposed technique

In TMIV7, 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 which is able to capture the whole scene. 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.

The parameters of the base view are chosen differently for omnidirectional and perspective content. For ERP, the angle of view of the base view is set to 360°×180°. The virtual base camera is placed in the middle of the camera rig. The resolution is set to be the highest possible within pixelrate constraints and 2:1 aspect ratio.

For perspective content, the virtual base camera has the same angle of view as input cameras. In order to allow capturing of the whole scene, the camera is shifted away from the scene (along x-axis). Regarding y and z position, the base camera is placed in the middle of the camera rig.

The x-offset is calculated in a way which allows the base camera to capture the entire scene without adding empty margins at both sides of the base view (Fig. 1).

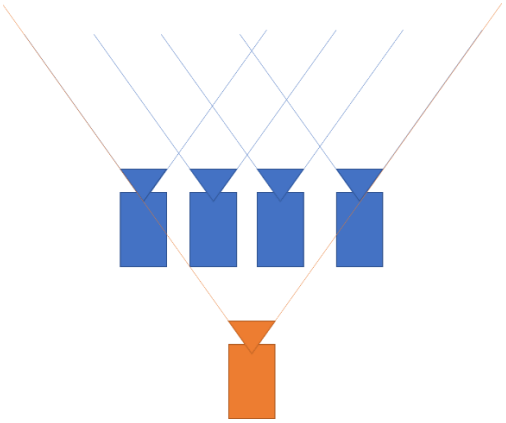


Fig. 1. Placement of the perspective base view.

The resolution of the base view resolution is set to be the highest possible within pixelrate constraints and the aspect ratio resulting from the camera arrangement.

2 Experimental results – A17

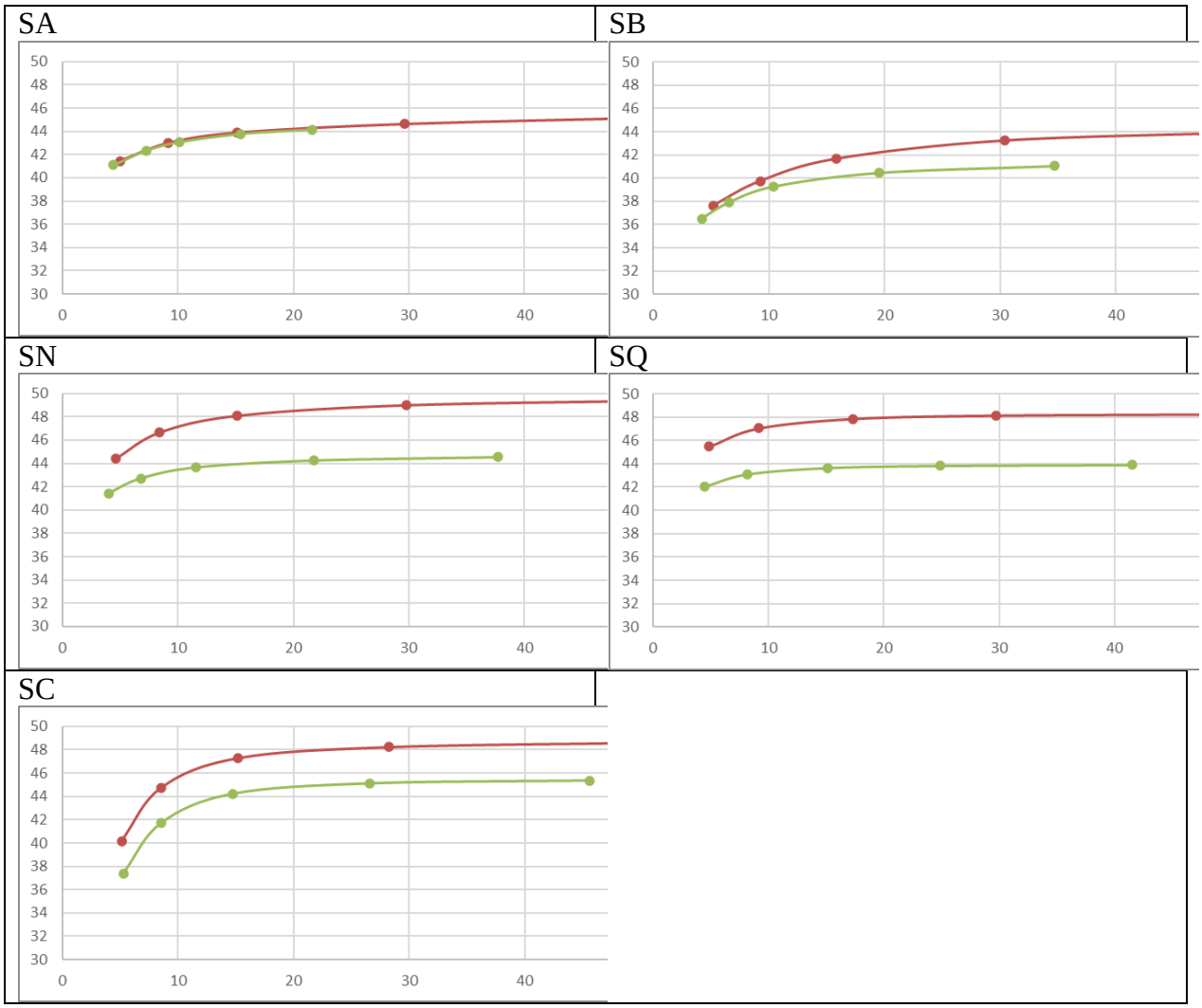
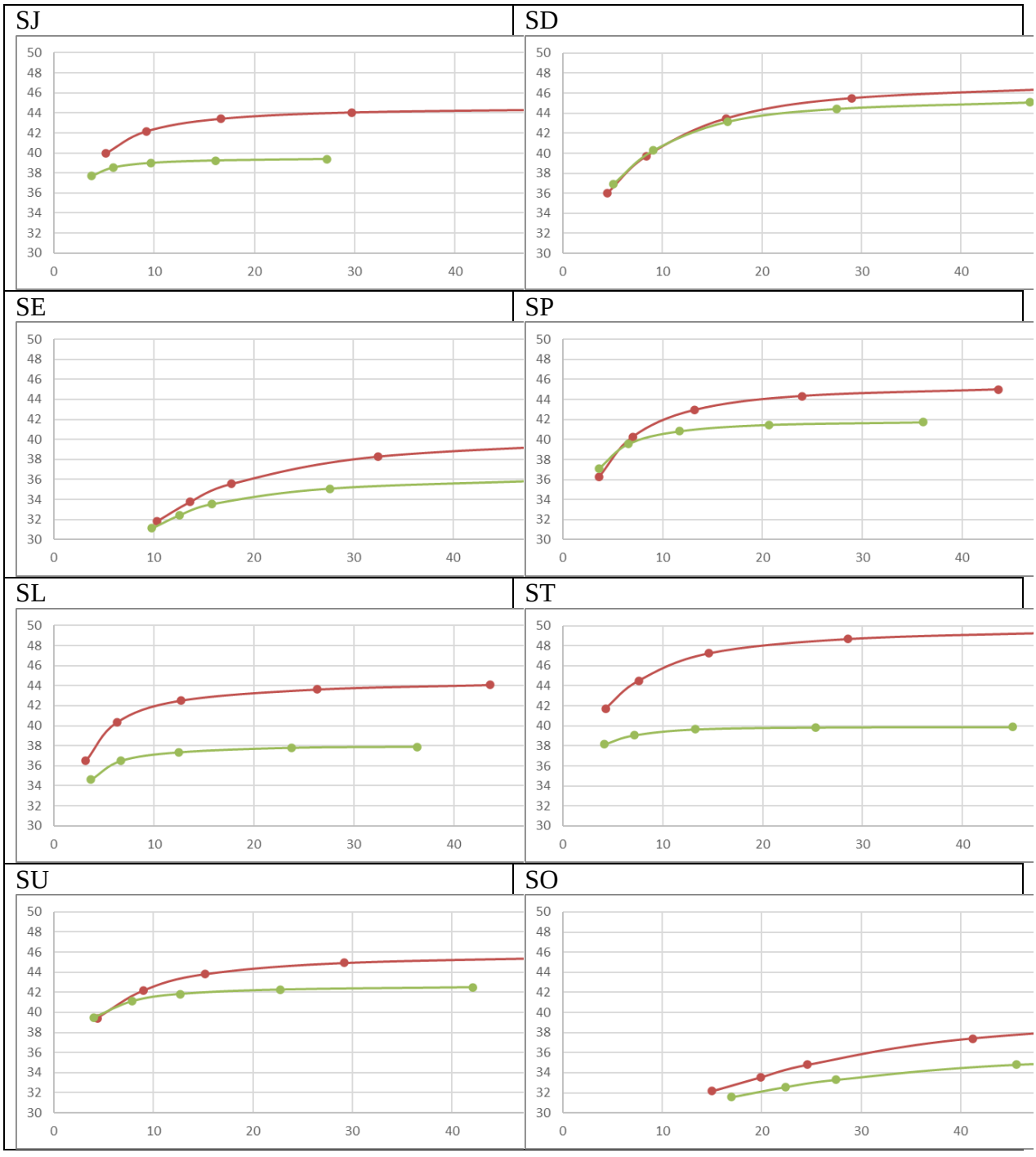


Fig. 2. IVPSNR: omnidirectional sequences (red: anchor, green: proposed).

Comments:

- the bitrate of SA is significantly reduced because of noise reduction caused by a synthesis of the base view,
- color artifacts caused by reflections are more visible in the proposed approach, however, they are more stable in time (less flickering) and space (long straight edges),
- both methods cause slightly different artifacts, some artifacts of the anchor were removed when using the proposed approach,
- SC – the angle of view of the base view is not fitted to the SC characteristics, the quality of the anchor is subjectively better (more details, better edges).



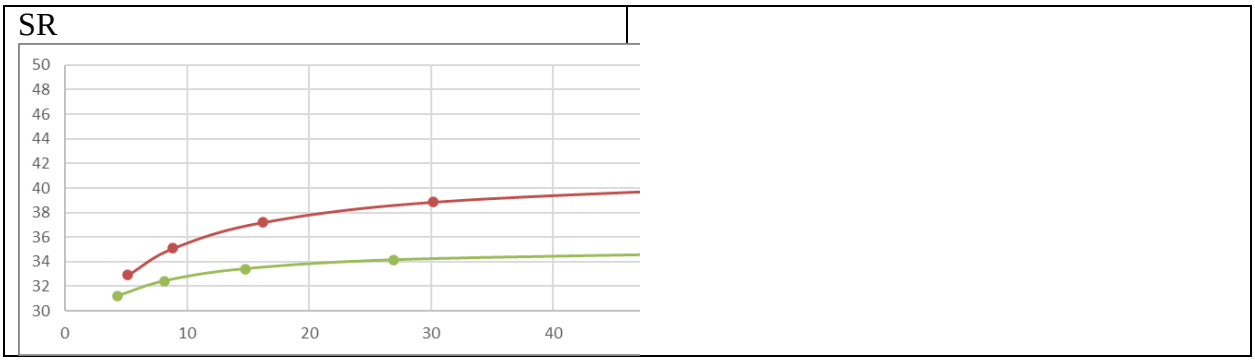


Fig. 3. IVPSNR: omnidirectional sequences (red: anchor, green: proposed).

Comments:

- quality of SD, SP, and SU is similar for both methods,
- for SE the textures are less sharp, but the overall quality is similar,
- SJ – there is a problem with a side wall of the cabinet, it is parallel to the optical axis of the base camera thus invisible in the base view; its depth is also similar to the front wall of the cabinet so it is pruned in additional views,
- SL – the objective quality of the synthesized base view is poor because of depth quality; in anchor, virtual views at the decoder side are synthesized using 4 input views labelled as base views,
- SR – cameras are possibly not OMAF-compatible, shifting the base camera along x-axis results in the wrong position thus base view does not contain valuable information (it contains some information thus increases bitrate, but it also requires sending many patches from other views).

Table 1. Delta-PSNR between views with highest and lowest quality.

SA		SB		SC		SD		SE		SJ		SL	
anc	prop	anc	prop	anc	prop	anc	prop	anc	prop	anc	prop	anc	prop
1.4	0.8	16.1	3.4	9.3	5.1	7.7	2.9	6.3	2.5	16.1	4.7	10.5	5.2
1.0	0.8	10.9	3.3	7.9	4.9	6.1	2.7	4.7	2.4	12.8	4.7	9.8	5.2
0.9	0.7	6.1	3.1	5.6	4.1	3.7	2.2	2.9	2.1	9.1	4.4	8.2	4.9
0.9	0.7	4.3	2.9	2.9	2.5	1.7	1.4	2.2	1.8	5.9	3.9	5.1	4.1
0.7	0.6	3.1	2.8	1.3	1.0	1.3	1.1	1.9	1.6	3.7	3.3	2.4	2.9
SN		SO		SP		SQ		SR		ST		SU	
anc	prop	anc	prop	anc	prop	anc	prop	anc	prop	anc	prop	anc	prop
14.3	7.2	8.0	2.4	7.2	2.6	14.7	7.8	11.8	9.1	9.4	4.3	8.8	2.3
13.1	7.1	6.1	1.9	6.0	2.6	14.3	7.8	8.4	8.4	8.3	4.3	7.3	2.3
10.6	6.8	4.1	1.7	4.5	2.4	13.6	7.7	6.1	7.2	6.4	4.2	5.7	2.3
7.8	6.1	3.3	1.7	2.7	2.1	11.6	7.5	4.5	6.0	3.7	4.0	4.2	2.2
5.1	5.2	2.7	1.7	1.3	1.6	8.0	6.6	3.3	4.9	2.2	3.8	2.3	1.8

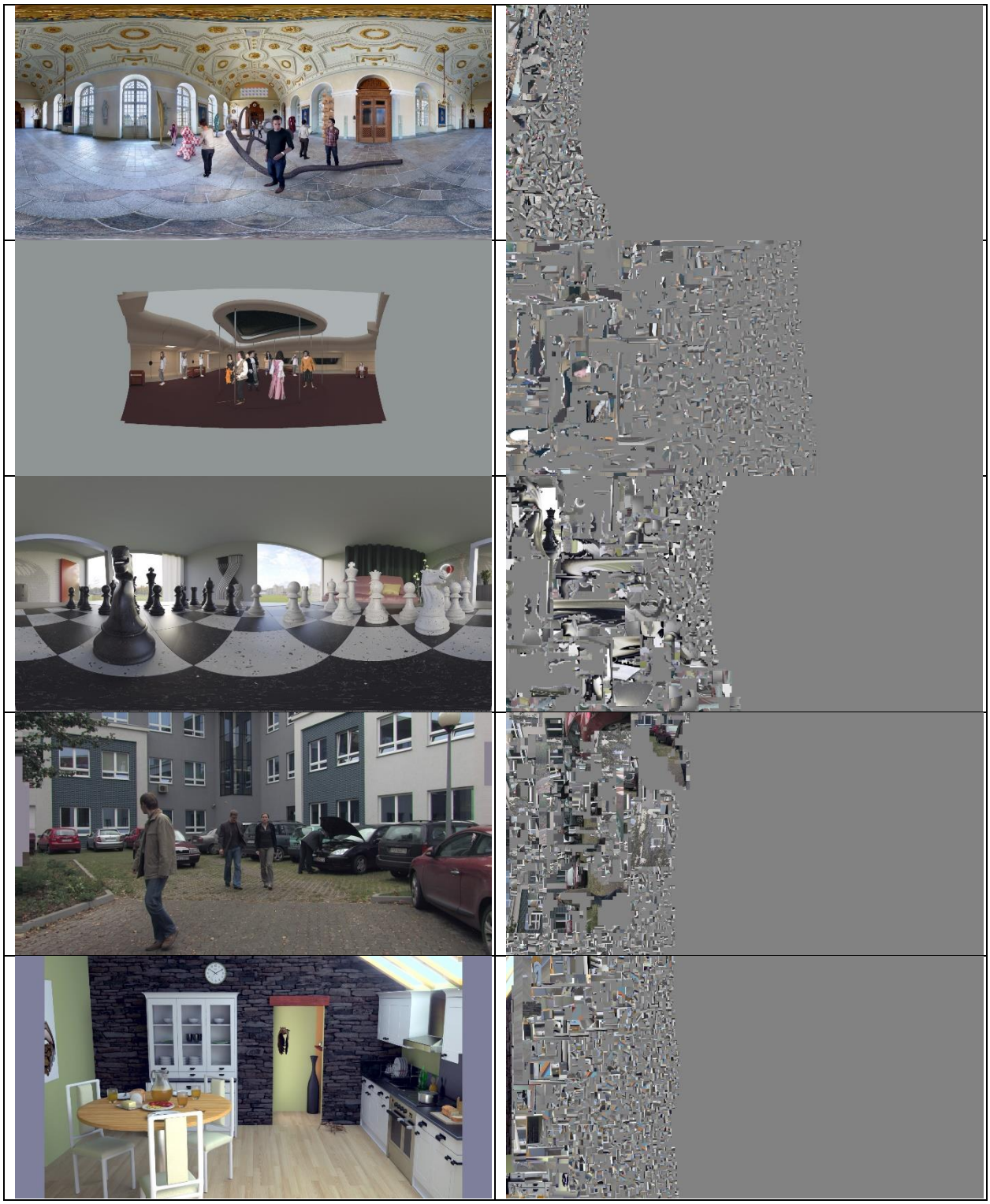


Fig. 4. Atlases for a subset of sequences.

Fig. 4. Atlases for SR.

As shown in Figs. 4 and 5, parameters of the base views were calculated properly for most of the content, excluding SC (too wide angle of view) and SR (wrong shift). Therefore, for both sequences, the second atlas contains more patches (especially for SR) than for other content.

In Fig. 6, second texture atlases from anchor and proposed are compared. In general, when using encoder-side rendering, the second atlas is more empty (exception: SP, SU and ST, where there are only 9 cameras, and 4 of them are base views in anchor – in this case disocclusions from other views form a small number of patches).

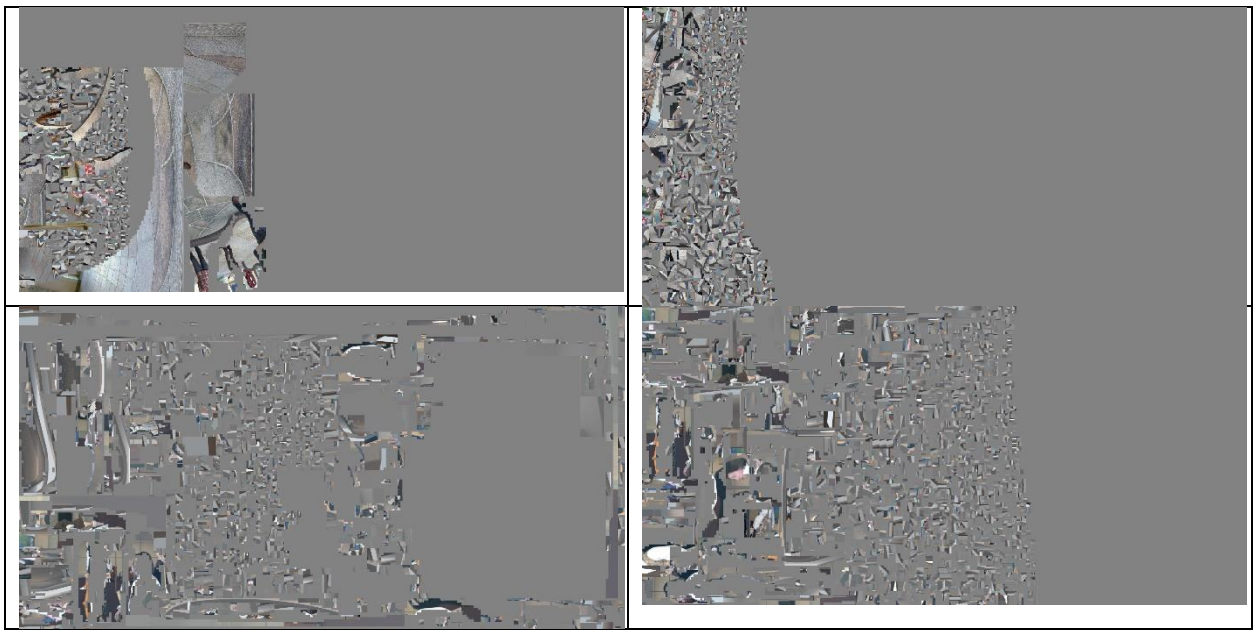


Fig. 6. Atlas 1: anchor (left) vs. proposed (right).

In Fig. 7 the influence on decoding quality is presented. In general, views synthesized using the proposed approach seem to be more smooth (e.g. less noise and reduced ghost edges around the lamps in SA or less visible artifacts in SN). However, some issues have to be eliminated before considering to use encoder-side rendering in TMIV.

Fig. 7. Decoded views (QP1).

3 Conclusions

Proposed approach:

- decreases the bitrate by reducing the inter-view redundancy occurring in TMIV7,
- significantly decreases the delta-PSNR between views with the highest and lowest quality, allowing the viewer the most stable experience when navigating within the scene,
- requires further improvements.

4 Recommendations

We recommend to:

- Continue the MIV CE2.

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).