INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 MPEG2019/M51602
January 2020, Brussels, Belgium

Source **Poznań University of Technology (PUT), Poznań, Poland**
Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea
Status **Input**
Title **Immersive Video CE3.1: Patch splitting**
Author Adrian Dziembowski*, Dawid Mieloch*, Adam Grzelka*, Jakub Stankowski*, Marek Domański*, Gwangsoon Lee**
* – Poznań University of Technology,
** – Electronics and Telecommunications Research Institute

1 Introduction

This document presents a technical description of one of the PUT/ETRI experiments on the atlas preparation (Immersive Video CE-3 [1]).

2 Overview of the proposed technique

Atlas preparation algorithm in TMIV still has one major flaw – there is a lot of redundant parts of input views in the atlases. The same information is copied within numerous patches, thus repeated many times. There are two reasons of that redundancy:

1. spatial – patches are filled by copying information in the entire bounding box of the cluster,
2. temporal – size and shape of each patch is aggregated for the whole GOP.

The proposed approach reduces influence of both issues, significantly reducing data redundancy, thus the size of the final bitstream. Here, the proposed patch splitting technique that removes the spatial redundancy, will be presented.

2.1 Spatial redundancy reduction

If the shape of the cluster is similar to the rectangle, copying information from the entire bounding box does not introduce much redundancy. However, if the cluster is L-shaped, the quantity of data copied from the source view could be significantly reduced (Fig. 1).

In order to decide how to split an L-shaped cluster, the total area of two subpatches is being minimized. The split line is always parallel to the shorter side of the patch. If splitting would not decrease the total area more than 10%, the split is not performed.

This approach allows to efficiently divide an L-shaped cluster. However, for other cluster shapes (e.g. C-shape), such approach does not result in the division of a cluster. Therefore, we proposed an additional cluster splitting (Fig. 2).

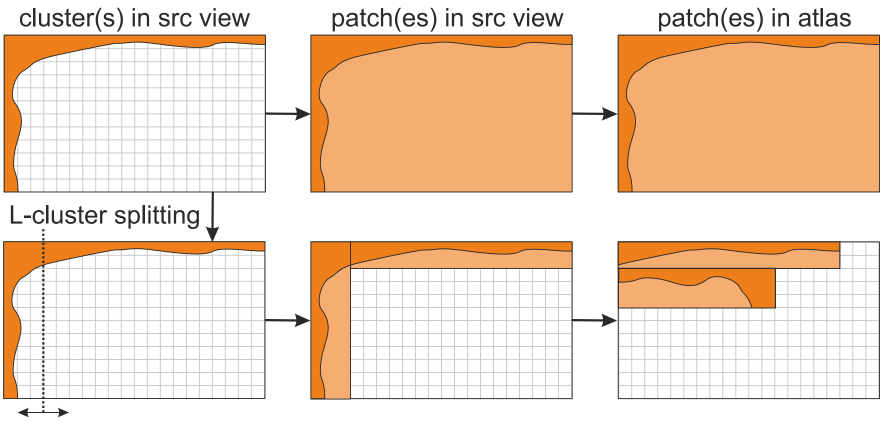

Fig. 1. L-shaped cluster splitting. The alignment grid is colored in grey.

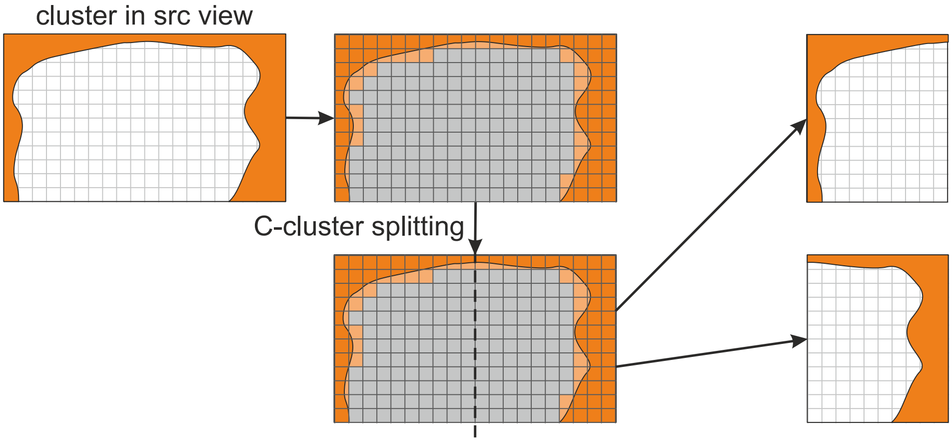

Fig. 2. C-shaped cluster splitting. The alignment grid is colored in grey.

Within the entire bounding box of the cluster, we calculate the number of 32×32 blocks that contain pixels belonging to the cluster (orange blocks in Fig. 2). Then, calculated number is divided by the total number of blocks within the analyzed bounding box. If that ratio is smaller than 30%, the cluster is split in half. Splitting of C-shaped cluster usually results in two L-shaped clusters.

Proposed cluster splitting is a recursive method. Example of the recursive splitting of an irregularly-shaped cluster is presented in Fig. 3.

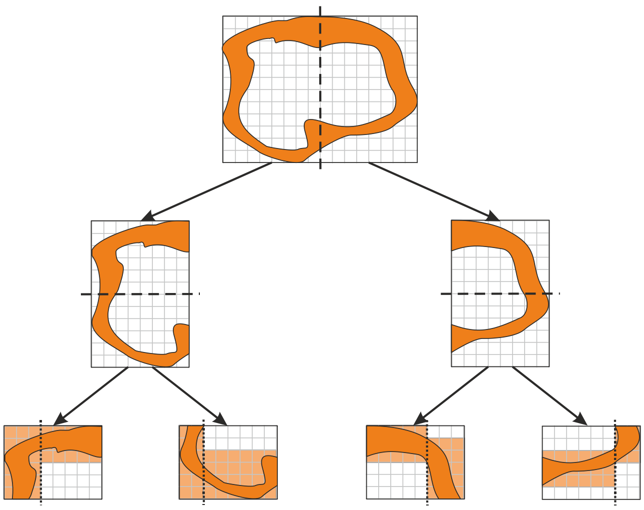

Fig. 3. Recursive splitting of the patch; dashed lines: C-splitting, dotted lines: L-splitting.

We decided that clusters smaller than 64×64 should not be split. It would result in a large number of very small clusters, smaller than a CU block, heavily increasing the required bitrate in HEVC encoding.

3 Experimental results

The results of the proposed enhancement are presented in the table below.

Test class	Sequence	Anchor (ff)	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate MS-SSIM	Low-BR BD rate MS-SSIM	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate ratio
CG	ClassroomVideo	AA97 (MIV)	-1.7%	4.9%	4.23	1.7%	7.2%	6.3%	8.3%	1.6%	5.4%	0.00%
	TechnicolorMuseum	BA97 (MIV)	-4.4%	-6.3%	12.29	-8.2%	-9.9%	-7.3%	-8.7%	-14.8%	-14.2%	-13.51%
	TechnicolorHijack	CA97 (MIV)	5.7%	6.6%	11.67	-0.5%	-0.5%	5.6%	5.5%	1.9%	4.1%	-20.00%
	OrangeKitchen	JA97 (MIV)	-1.3%	-4.1%	13.90	-5.1%	-7.1%	-1.5%	-4.0%	-2.1%	-5.2%	-11.11%
	NokiaChess	NA97 (MIV)	12.2%	7.3%	15.36	9.7%	4.9%	7.4%	4.5%	12.2%	7.0%	-13.51%
		MIV	-0.4%	0.3%	13.90	-3.0%	-2.6%	0.8%	0.3%	-3.3%	-2.5%	-11.44%
		All anchors	-0.4%	0.3%	13.90	-3.0%	-2.6%	0.8%	0.3%	-3.3%	-2.5%	-11.44%

NC	TechnicolorPainter	DA97 (MIV)	-1.0%	-0.4%	6.19	0.9%	0.8%	-0.4%	0.0%	-2.1%	-1.2%	0.00%
	IntelFrog	EA97 (MIV)	1.1%	-1.4%	8.97	-2.1%	-2.9%	-2.3%	-3.4%	0.6%	-2.0%	0.00%
	PoznanFencing	LA97 (MIV)	11.4%	9.2%	13.17	9.4%	5.7%	17.7%	10.5%	6.8%	6.3%	0.00%
		MIV	3.8%	2.5%	13.17	2.7%	1.2%	5.0%	2.4%	1.8%	1.0%	0.00%
		All anchors	3.8%	2.5%	13.17	2.7%	1.2%	5.0%	2.4%	1.8%	1.0%	0.00%

Test class	Sequence	Anchor (ff)	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	Max delta Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate MS-SSIM	Low-BR BD rate MS-SSIM	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR	Pixel rate ratio
All		MIV	1.4%	1.2%	10.06	-0.6%	-1.0%	2.6%	1.2%	-1.2%	-1.0%	-6.71%
		All anchors	1.4%	1.2%	10.06	-0.6%	-1.0%	2.6%	1.2%	-1.2%	-1.0%	-6.71%

As the results show, use of the proposal decreases the pixel rate for CG sequences on average by 11% in comparison with MIV anchor. The visual comparison of atlases for TechnicolorHijack is presented in Fig. 4. As it can be seen, in TMIV many areas are present in both created atlases (e.g. floor). In the proposal, such redundancy is vastly decreased. The pixel rate decrease was not achieved for sequences with group-based TMIV encoding.

The proposal provided a negligible loss of BD-rate for Y-PSNR, however, use of this quality rate is not optimal in case of the use of view synthesis. For IV-PSNR and VMAF the proposal achieved better BD-rate than the anchor.

The worst results were achieved for very complex natural content sequences with estimated depth maps (SL). As the results for the rest of NC sequences show, the proposed technique increases the performance of TMIV.

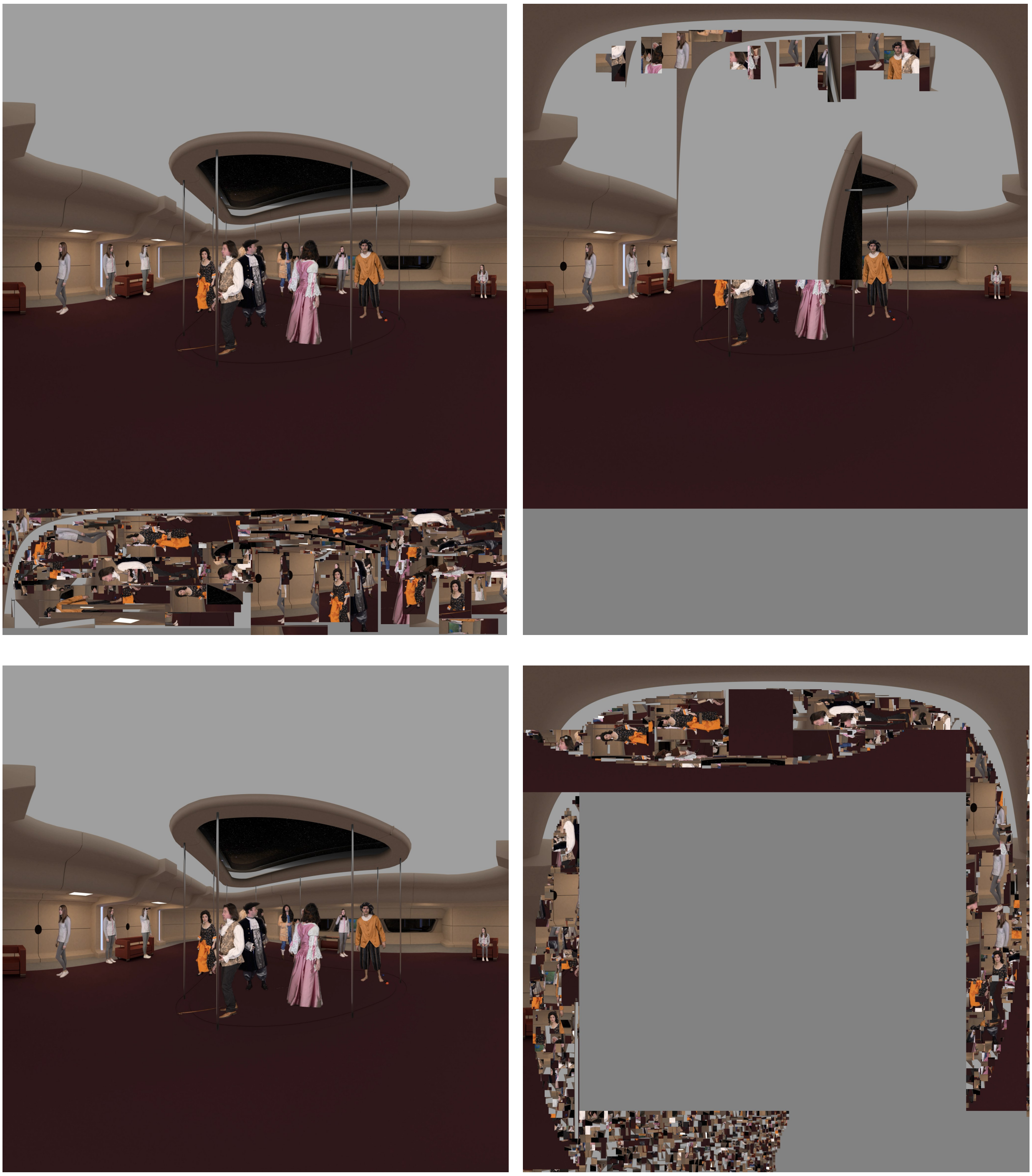

Fig. 4. The visual comparison of atlases for TechnicolorHijack sequence made by TMIV 3.0 (top) and for TMIV 3.0 with the proposed technique (bottom).

4 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).

We would like to thank Hyunho Kim from KAU for thorough crosscheck of the presented results.

5 Recommendations

We recommend to continue the Core Experiment 3.

We recommend the group to focus on the quality of depth maps for natural content, as the SD and SE sequences shows that the further decrease of BD-rate can be achieved even for estimated (not generated) depth maps.

Considering the pixel rate reduction in comparison with TMIV 3.0, we suggest to include our technique into TMIV 4.0.

6 References

[1] Renaud Doré, "Description of Immersive Video Core Experiments 3 (Atlas preparation)", ISO/IEC JTC1/SC29/WG11 MPEG/N18934, October 2019, Geneva, CH.