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/M49963
October 2019, Geneva, Switzerland

Source	**Poznań University of Technology (PUT), Poznań, Poland** **Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea**
Status	**Input**
Title	**PUT/ETRI Response to Immersive Video CE-3: Atlas Preparation**
Author	Adrian Dziembowski*, Dawid Mieloch*, Marek Domański*, Gwangsoon Lee** * – Poznań University of Technology, ** – Electronics and Telecommunications Research Institute

1 Introduction

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

2 Overview of the proposed technique

Atlas preparation algorithm in TMIV 2.0 has one major flaw – there is a lot of redundancy 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 bitstream.

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 should be aligned with the grid ("alignment" parameter in the proposal is equal to 32). 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) it does not work. Therefore, we proposed an additional cluster splitting (Fig. 2).

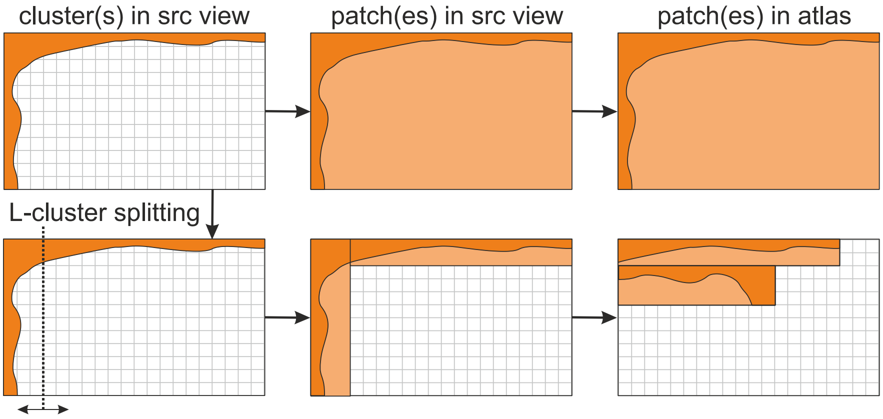

Fig. 1. L-shaped cluster splitting. The alignment grid is colored in grey.

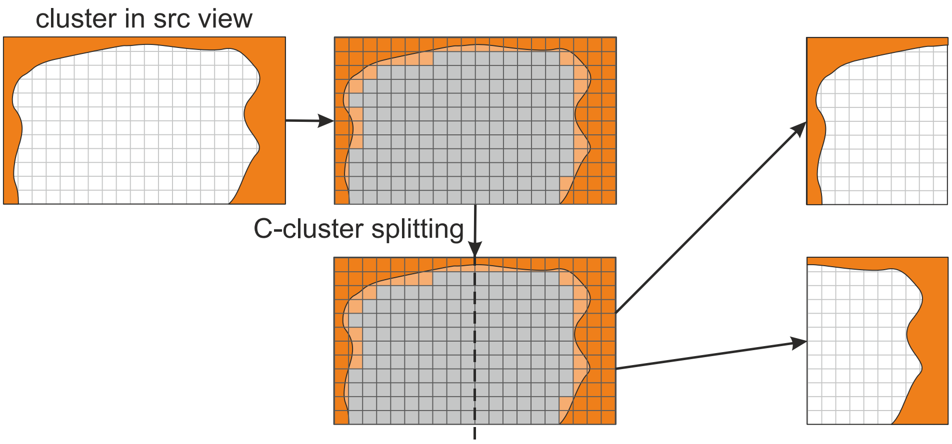

Fig. 2. C-shaped cluster splitting. The alignment grid is colored in grey.

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.

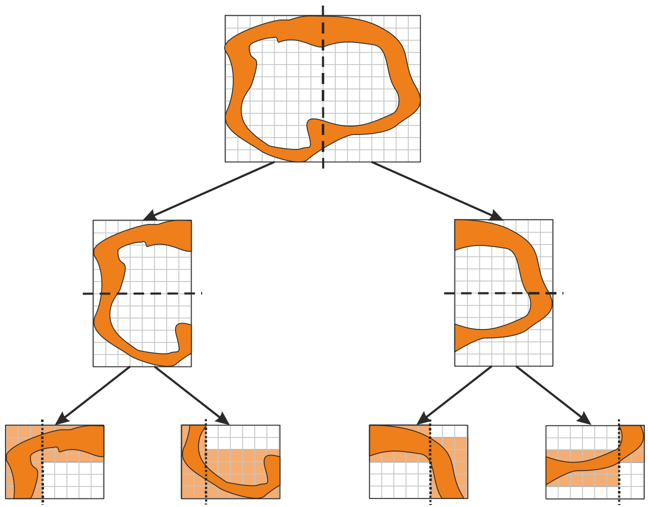

Fig. 3. Recursive splitting of the patch; dashed lines: C-splitting, dotted lines: L-splitting.

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.

2.2 Temporal redundancy reduction

In TMIV 2.0, the shape of each patch does not change during the entire GOP. However, if there is any movement in the scene, some pixels are redundant and do not have to be copied for all the frames but only for the frames where they are needed.

In the proposed approach, only the blocks that contain pixels belonging to the cluster for the current frame are copied from the source views to the atlases. In Fig. 4 the comparison between TMIV 2.0 (left) and proposed method (right) is presented. In TMIV, the patch is fully copied for all frames. In proposed approach, major part of the patch's area is empty (thus easier to encode).

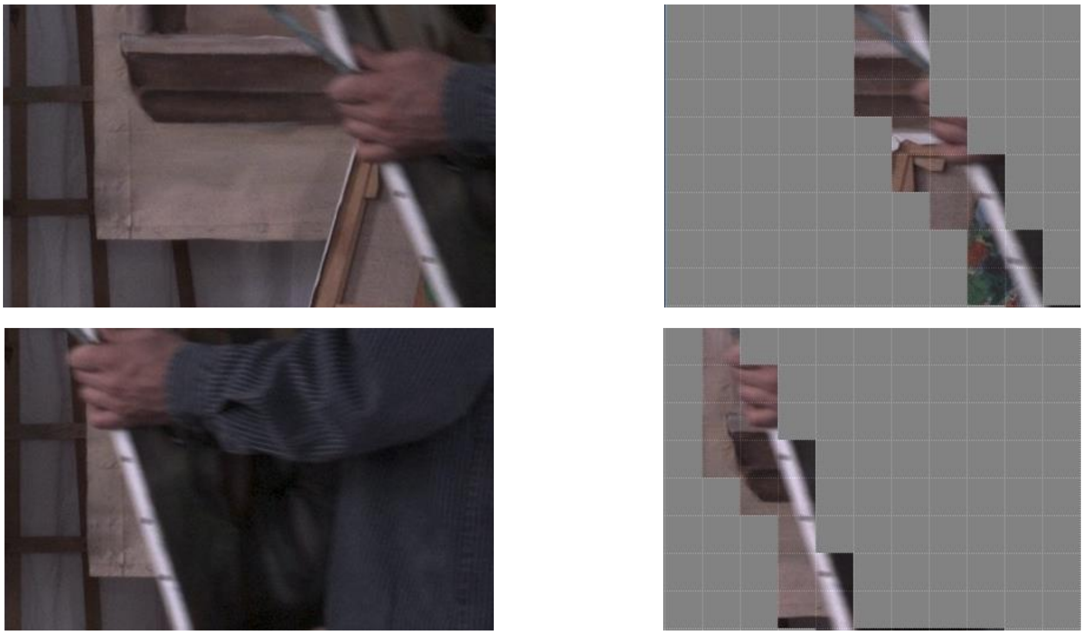

Fig. 4. TMIV 2.0 (left) vs. proposed method (right), frame 36 (top) and 47 (bottom).

2.3 Comments

Redundancy reduction revealed the flaw of TMIV syntheser. In general, the syntheser blends all (both foreground and background) objects together (using a weighted average). When there are fewer repetitions of foreground pixels in atlases, the background is more visible (foreground objects in the synthesized views become transparent).

In order to fix this, we slightly modifying the syntheser by adding a blending threshold. If normalized weights calculated for two collocated pixels are similar enough, the blending is performed without changes. If the absolute difference between them is higher than 0.1, the pixel with a higher weight is copied to the synthesized view.

Moreover, for sequences which do not follow OMAF (SJ and SL) the weights used in blending are calculated improperly. It results in choosing further pixels instead of closer ones. We discovered that changing "depthParameter" from 20 to 200 decreases the influence of this issue.

3 Experimental results

The results of the proposed enhancements are presented in the table below. Note that value 0.0% represents value of BD-rate that could not be calculated (because two of measured sets of values did not overlap). If such value is green, then the proposal achieved much better results than the anchor, if value is red, then the anchor was better.

Test class	Sequence	Anchor	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	Classroom Video	A1 (MIV)	34.2%	18.5%	0.33	-28.8%	-7.0%	8.4%	5.9%	-9.2%	-3.5%	0.00%
		A2 (MIV view)	0.0%	16.7%	0.98	226.0%	-30.6%	-38.5%	-50.1%	-53.9%	-57.0%	-77.78%
	Technicolor Museum	B1 (MIV)	1.5%	4.0%	0.45	-14.9%	-9.8%	-1.1%	-0.7%	-13.4%	-9.5%	-25.00%
		B2 (MIV view)	-6.9%	-17.0%	0.67	-14.8%	-28.4%	-22.2%	-35.8%	-48.3%	-51.5%	-62.50%
	Technicolor Hijack	C1 (MIV)	-17.2%	-4.4%	1.69	-57.4%	-44.9%	-15.5%	-9.0%	-36.8%	-26.3%	-33.33%
		C2 (MIV view)	158.7%	109.2%	1.43	1.0%	4.9%	55.9%	38.7%	57.9%	38.7%	-60.00%
	Orange Kitchen	J1 (MIV)	0.0%	-68.5%	-2.75	0.0%	0.0%	-76.4%	-42.0%	0.0%	0.0%	0.74%
		J2 (MIV view)	0.0%	0.0%	-3.03	0.0%	51.8%	103.4%	18.0%	0.0%	36.4%	-66.42%
		MIV	4.6%	-12.6%	-0.07	-25.3%	-15.4%	-21.2%	-11.4%	-14.8%	-9.8%	-15.76%
		MIV view	38.0%	27.2%	0.01	53.0%	-0.6%	24.6%	-7.3%	-11.1%	-8.3%	-67.47%
		All anchors	21.3%	7.3%	-0.03	13.9%	-8.0%	1.7%	-9.4%	-13.0%	-9.1%	-41.61%

Test class	Sequence	Anchor	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
NC	Technicolor Painter	D1 (MIV)	10.7%	2.9%	1.08	-33.0%	-25.4%	-3.7%	-5.8%	-22.7%	-19.4%	0.00%
		D2 (MIV view)	0.0%	108.3%	-1.19	107.6%	43.5%	62.5%	32.9%	25.3%	13.8%	-37.50%
	IntelFrog	E1 (MIV)	0.0%	99.3%	3.79	59.0%	-7.8%	284.2%	24.0%	0.0%	35.4%	-19.41%
		E2 (MIV view)	0.0%	0.0%	2.27	0.0%	325.3%	0.0%	0.0%	0.0%	0.0%	15.13%
	Poznan Fencing	L1 (MIV)	0.0%	111.4%	0.89	0.0%	36.6%	0.0%	134.5%	-10.2%	-16.5%	0.74%
		L2 (MIV view)	0.0%	0.0%	-0.09	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-19.41%
		MIV	3.6%	71.2%	1.92	8.6%	1.1%	93.5%	50.9%	-11.0%	-0.2%	-6.71%
		MIV view	0.0%	36.1%	0.33	35.9%	122.9%	20.8%	11.0%	8.4%	4.6%	-16.61%
		All anchors	1.8%	53.7%	1.13	22.3%	62.0%	57.2%	30.9%	-1.3%	2.2%	-11.80%

Test class	Sequence	Anchor	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	4.2%	23.3%	0.78	-10.7%	-8.3%	28.0%	15.3%	-13.2%	-5.7%	-11.99%
		MIV view	21.7%	31.0%	0.15	45.7%	52.4%	23.0%	0.5%	-2.7%	-2.8%	-51.31%
		All anchors	12.9%	27.2%	0.47	17.5%	22.0%	25.5%	7.9%	-7.9%	-4.2%	-34.54%

As the results show, use of the proposal decreases the pixel rate on average by 12% in comparison with MIV anchor. The reduction is higher for CG sequences, as their atlases contained many large patches. After the use of the proposal, such large patches were split into smaller ones and successfully placed in free space present in other atlases.

The very small increase of the pixel rate for SE, SJ and SL is the result of the increased resolution of atlases. When the alignment is set to 32, then the TMIV does not work if height and width of atlases cannot be divided by 32. Therefore, because height of atlases for these sequences was set as 1080 for anchor, we increased it to 1088.

The proposal provided a 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 the proposal achieved better BD-rate for 6 of 7 sequences, while for VMAF for 5 sequences. The worst results were achieved for natural content sequences with low quality depth maps (SE and SL), on the other hand, the case of SD shows that the proposal also can provide the decrease of BD-rate for natural content.

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

As it was stated in section 3, 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 sequence shows that the further decrease of BD-rate can be achieved even for estimated (not generated) depth maps.

We recommend to fix the not-OMAF-compatible camera parameters for SL and SJ.

We also recommend to fix an error in TMIV: when the alignment is set to 32, then the TMIV does not work if height and width of atlases cannot be divided by 32.

Considering much better results than for TMIV 2.0 anchor, we suggest to include our technique into TMIV 3.0.

6 References

[1] Renaud Doré, "Description of Immersive Video Core Experiments 3: Atlas Preparation", ISO/IEC JTC1/SC29/WG11 MPEG/ N18707, July 2019, Gothenburg, Sweden.