A Method of Patch Merging for Atlas Construction in 3DoF+ Video Coding

Sung-Gyune Im, Hyun-Ho Kim, Gwangsoon Lee, and Jae-Gon Kim
Korea Aerospace University
{rbs293, hhkim}@kau.kr, gslee@etri.re.kr, jgkim@kau.ac.kr

Abstract

MPEG-I Visual group is actively working on enhancing immersive experiences with up to six degree of freedom (6DoF). In virtual space of 3DoF+, which is defined as an extension of 360 video with limited changes of the view position in a sitting position, looking at the scene from another viewpoint (another position in space) requires rendering additional viewpoints using multiple videos taken at the different locations at the same time. In the MPEG-I Visual workgroup, methods of efficient coding and transmission of 3DoF+ video are being studied, and they released Test Model for Immersive Media (TMIV) recently. This paper presents the enhanced clustering method which can pack the patches into atlas efficiently in TMIV. The experimental results show that the proposed method achieves significant BD-rate reduction in terms of various end-to-end evaluation methods.

1. Introduction

Recently, with the increased commercial interests in deploying Virtual Reality (VR) applications, 360 video has become popular as a new media type giving immersive experiences. In order to enhance immersive experiences with up to six degrees of freedom (6DoF), MPEG-I Visual Group is actively working on it [1], [2]. In VR space of 3DoF+ which is considered in MPEG-I, looking at the scene from another viewpoint (another position in space) requires rendering additional viewpoints using multiple videos taken at the different locations at the same time. These multiple videos of source views that form a 3DoF+ video have very large volume to support high resolutions such as 4K or 8K, etc. To efficiently compress these 3DoF+ videos, various compression methods are being studied in the MPEG-I Visual Group and they released Test Model for Immersive Media (TMIV) recently. In this paper, we propose enhanced method for clustering in TMIV to reduce the amount of information needs to be transmitted while increasing rendering performance.

2. Test Model for Immersive Media

In the MPEG-I Visual workgroup, they released the working draft (WD) [3] and the test model for immersive video (TMIV) [4] to efficiently compress and transmit large volume of 3DoF+ video. The main idea on the compression of 3DoF+ video in TMIV is to reduce the total amount of pixel data to be transmitted by eliminating the redundancy between all available source views which are highly correlated in advance of compression.

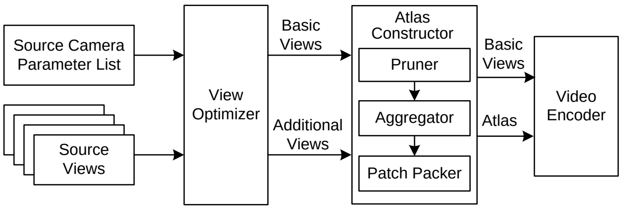

Figure 1. Overall architecture of the TMIV encoder

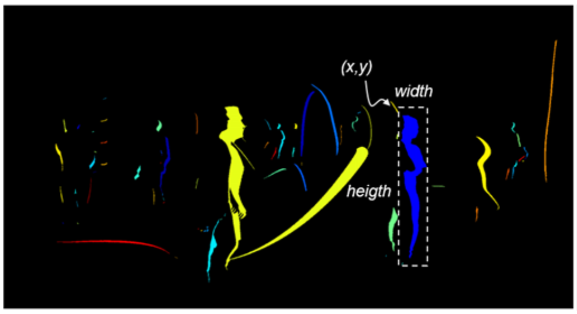

Figure 2. An example of pruned view and generated patches

Fig. 1 shows an overall architecture of the TMIV encoder. A set of input source views with different view positions are classified into 'basic view' or 'additional view' in a view optimizer firstly. After that, in an atlas constructor, which contains a pruner, an aggregator, and a patch packer, additional views are pruned into multiple patches each of which containing remaining regions after removing the redundant parts between the basic view and each additional view.

Then, the remaining regions are clustered into multiple separated regions each of which is bounded by a rectangular box called patch. Therefore, in each rectangular patch, 'invalid region' which contains redundant pixels are included as well as 'valid region' which contains real residual pixels that should be transmitted. Fig. 2 shows an example of pruned view, in which each colored region representing the aggregated and clustered region will be generated into a rectangular patch.

After that, the patches generated from each additional view are packed into a single frame called an atlas in the unit of frame. In this way, an atlas is constructed by packing patches of all views in each frame. This packing process sequentially pack each patch into an atlas in a raster scan order and overlapping each other unless valid regions are invaded.

3. Patch Merging

In TMIV, each clustered region generates rectangular shape patch one-by-one. Because of this process, when a large clustered area is made of one patch and even if the small clustered region is included in that patch, a patch for the small region is generated separately. As a result, these created patches are packed into atlas separately and take up more space, even if the small patches can be found in other patches. Also, the total amount of metadata which contains each patch's information will be increased. Fig 3 shows an example of creating patches in pruned view.

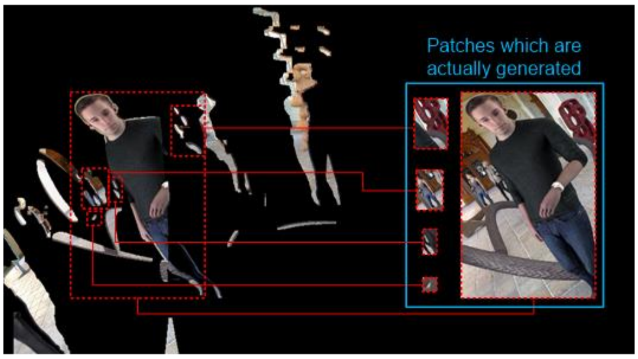

Figure 3. An example of creating patches in the pruned view

To address this problem, we eliminated the unnecessary patch and its metadata by merging patches into one patch if there is another patch included in the created patch's internal area. When the included patches are not fully included in the parent patch, we didn't merge them and created that patch separately. This is because when the patch is cut across the parent patch boundary, visual artifacts may be generated at that position when the view is reconstructed.

4. Experimental Results

Table 1 shows the end-to-end coding performance with our patch merging method in comparison with the anchor in terms of Bjontegaard-Delta rate (BD-rate). The proposed method was implemented on the TMIV 2.0.2 [4], and HM 16.16 [5] was used to compress the constructed atlas according to the common test condition (CTC) for immersive video [6]. It is noted that there is significant coding gain in all sequences in terms of objective quality as well as subjective quality.

Table 1. Experimental results on the compression of 3DoF+ videos with the proposed method

Sequence	WS-PSNR (Y)	VMAF	MS-SSIM	IV-PSNR
ClassroomVideo	-1.8%	-2.3%	-0.5%	-0.5%
TechnicolorMuseum	#####	#####	######	#####
TechnicolorHijack	#####	#####	######	#####
OrangeKitchen	#####	#####	######	#####
TechnicolorPainter	-9.9%	-6.8%	-5.3%	-5.1%
IntelFrog	#####	#####	######	#####
PoznanFencing	-65.1%	-57.4%	-43.9%	-32.4%

Table 2 shows comparison of the number of patches generated with anchor and proposed method. It can be observed that there is a significant decrease in all sequence, especially on the *TechnicolorPainter* and *IntelFrog* sequence which are captured in real world.

Table 2. Number of patches generated with anchor and proposed method

Sequence	Anchor	Proposed
ClassroomVideo	424	416
TechnicolorMuseum	238	219
TechnicolorHijack	415	327
OrangeKitchen	439	407
TechnicolorPainter	648	408
IntelFrog	1,411	544
PoznanFencing	574	532

5. Conclusions

This paper, we proposed a method of merging patches to enhance coding efficiency of the atlas of 3DoF+ video by reducing the number of patches and merging high-correlated patches into one. The experimental results showed that the proposed method reduce the number of patches and gave significant coding gain in terms of objective quality measures of WS-PSNR and IV-PSNR as well as subjective quality measures of VMAF and MS-SSIM, respectively.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT).

References

[1] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," ISO/IEC JTC1/SC29/WG11, w16768, April 2017.

[2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, "Standardization Status of Immersive Video Coding," IEEE Jour. Emerg. Select. Topics Circuits Syst., vol. 9, no. 1, pp. 5-17, Mar. 2019.

[3] J. Boyce, R. Dore, V. Vadakital, "Working Draft 2 of Immersive Video," ISO/IEC JTC1/SC29/WG11, w18576, July 2018.

[4] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, w18577, July 2019.

[5] HM reference software, [Online]. Available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[6] J. Jung, B. Kroon, J. Boyce, "Common Test Conditions for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18443, March 2019.