Verification of Compression Architecture for 3DoF+ Immersive Video Delivery

Gwangsoon Lee, Hong-Chang Shin, Homin Eum, and Jeongil Seo

Electronics and Telecommunications Research Institute, Dajeon Korea

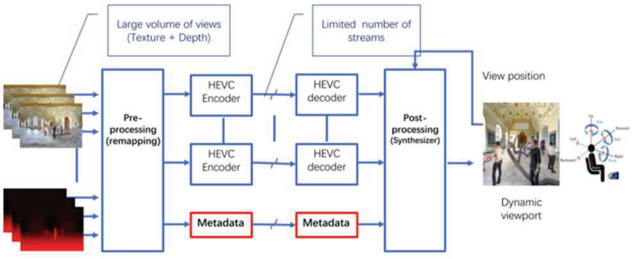
Keywords: Immersive media, 360 video, motion parallax, 6DoF

ABSTRACT

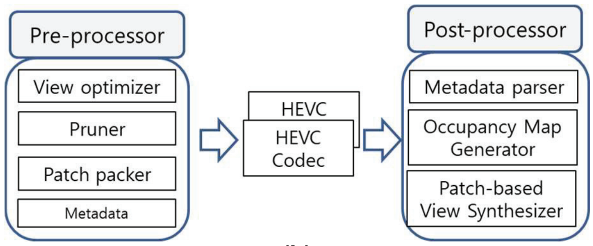
This paper introduces a compression architecture for delivering 3DoF+ immersive video, which can be applied to existing video encoder. Specifically, this paper includes a pruning algorithm that can reduce the redundancy among multiple views while maintaining the higher image quality of rendered view.

1. INTRODUCTION

As a primitive immersive video technology, 3DoF 360 video currently can render viewport images dependent on rotational movements of viewer [1]. However, rendering flat 360° video, i.e. supporting head rotations only, may generate visual discomfort especially when objects close to the viewer are rendered. 3DoF+ enables head movements for a seated person (adding horizontal, vertical and depth translations). It is noted that 3DoF+ 360 video is positioned between 3DoF and 6DoF, which can realize the motion parallax with relatively simple VR software in HMD. First version of the Test Model for Immersive Video (TMIV) for standardization of 3DoF+ video compression was defined at the beginning of this year [2], together with the TMIV-SW 1.0 reference software. As shown in Fig. 1 3DoF+ compression architecture aims to use existing video compression infrastructure by just adding a preprocessor to encoder side and a postprocessor to decoder side. Instead, additional metadata contribute to control the behavior of both preprocessor and postprocessor. Here, the preprocessor is consists of three main steps: the view optimizer, the pruner, and the patch packer & metadata composer. As inputs to the TMIV encoder, multiple views (texture & depth) representing projections of 3D real or virtual scene are currently made available.



(a)



(b)

Fig. 1 (a) Architecture of 3DoF+ video compression, (b) block diagrams of preprocessor/postprocessor

The view optimizer selects one or multiple basic views and additional views from all the source views. The basic views are selected in the consideration of the direction deviation, field of view, and distance and overlap between views. On the other hand, additional views are pre-processed prior to existing video encoder through mainly pruning process and patch packing process. As shown in Fig. 2, the pruning process is to extract the occluded regions across basic views and additional views, which leads the packing process to generate each patches. These two pre-process plays main role in reducing redundant image pixel among multiple views. In this paper, we will propose and verify new pruning algorithm that can reduce effectively 3DoF+ video data prior to existing video encoder while maintain the rendered view image quality.

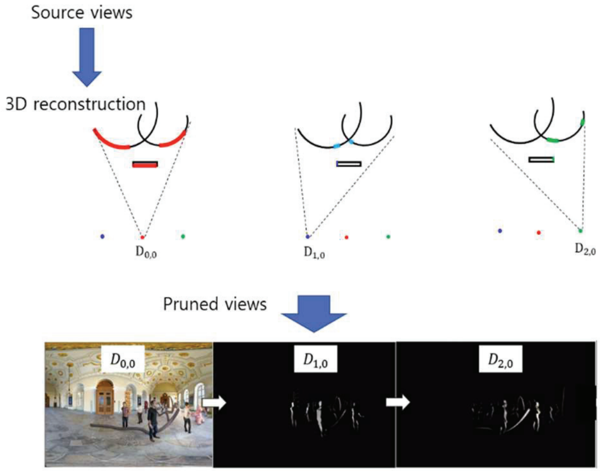


Fig. 2 Concept of pruning process

2. Proposed Algorithms

During core experiment, it was experimentally found that the 3DoF+ video technology based on pruned views may arouse seam effects in rendered viewport image, which is estimated to happen due to the strict process of patch packing. Simple way of reducing this phenomenon is to make adjacent patches overlapped at the each boundary, however, which leads to the increase of redundancy among views again. Therefore, it is clear that one of the issues in pruning process is how to reduce the redundancy among whole view while maintaining the image quality of rendered view as much as the case of view synthesis based on multiple views. Please note that multiple views has an advantage in terms of view synthesis, compared to pruned view case. We here propose an algorithm that can generate the overlapped area among adjacent pruning area while refraining redundancy pixels from being increased. To do this, we adaptively control the size of pruning mask that designates validate pixels within each patches. Specifically, we develop the algorithm to dilate the pruning mask in only direction where the depth difference between inside and outside of the pruning mask are beyond certain level, which is illustrated in Fig. 3. Moreover, the width of dilation is decided by this difference of depth. Our assumption is originated from the fact the possibility of generating big hole in the view synthesis can be increased in the boundary of objects with big depth difference. Therefore, overlapped pixel would be required in specially areas where big hole can occur because such overlapped area from several patches can contribute to the blending effect in the view synthesis. Note that the blending technique can fill certain size of holes as final procedure of view synthesis.

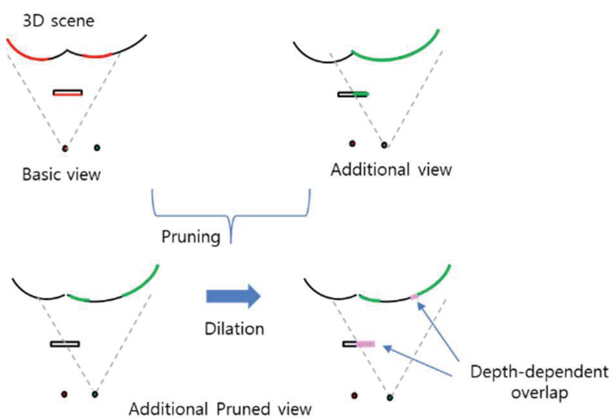


Fig. 3 Proposed pruning algorithm based on adaptive dilation of mask

3. Experimental Results

We experimentally implemented the proposed algorithm that can incorporate into existing TMIV reference software. As the test contents, we select three kinds of 3DoF+ contents with equirectangular projection (ERP), or perspective projections. For the verification of our approach, we first check that the pruning mask area can be dilated as a function of the depth variation, which fully corresponds to intention of proposed algorithm. Fig.4 compares the dilation of mask in three case: anchor case, fixed dilation case and proposed case that dilates the mask as a function of depth. As we can see this figure, initial mask were mainly dilated at the area with large depth in the proposed case. Finally we compared the rendered view in the view synthesis process compared to existing case that just creates the pruning mask without any consideration of depth. In the Fig. 5, it is confirmed that many patches are blended well enough to render the viewport image in good quality through view synthesis process, compared to existing TMIV reference software.

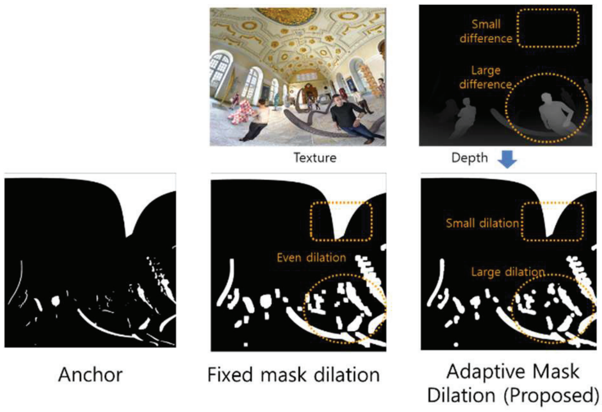


Fig. 4 Comparison of pruning masks using anchor, fixed dilation and proposed pruning algorithm

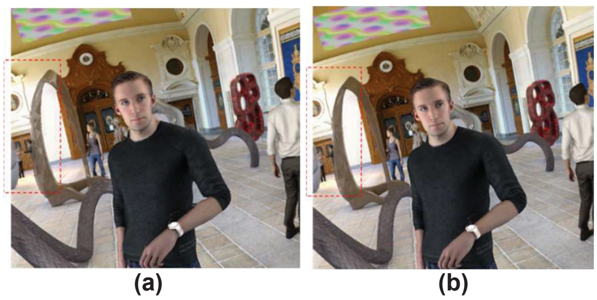


Fig. 5 Comparison of rendered views based on (a) anchor and (b) proposed pruning Algorithm

4. Conclusion

In this paper, we introduced the architecture of 3DoF+ video compression and one core algorithm related to the preprocessor. And we showed the proposed algorithm can enhance subjective quality in the rendered view of decoder. We are currently verifying the coding efficiency by incorporating end-to-end blocks including HEVC Main 10 profile. Therefore, we will be able to introduce overall efficiency of 3DoF+ video compression scheme together with proposed algorithm in near future. However we believe that the viewer will be able to experience more realistic immersive media as the proposed algorithm can be well harmonized in the existing infrastructure for 3DoF+ video compression.

Acknowledgements

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).

REFERENCES

[1] M. Champel, R.Koenen, G. Lafruit Schenkel, M.Budagavi, “Working Draft 0.4 of TR: Technical Report on Architectures for Immersive Media”, ISO/IEC JTC1/SC29/WG11 MPEG2017/m41083, Macau, China, Oct. 2017.

[2] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), “Test model for Immersive Video”, ISO/IEC JTC1/SC29/WG11 MPEG/N18470, Mar. 2019, Geneva, Switzerland.

[3] J. Boyce, R. Doré, V. Kumar Malamal Vadakital (Eds.), “Working Draft 1 of Metadata for Immersive Media (Video)”, ISO/IEC JTC1/SC29/WG11 MPEG/w18464, Mar. 2019, Geneva, Switzerland.