Enhancing Visual Quality of Multi-view 360 Video Compression Pipeline

Junyoung Yun[3], Hong-Chang Shin[2], Gwangsoon Lee[2], Jong-Il Park[1]

[1]Dept. of Computer and Software, Hanyang University, Seongdong-gu, Seoul, 04763, Korea
Tel.: 82-2-2220-0368, E-mail: jipark@hanyang.ac.kr
[2]Electronics and Telecommunications Research Institute, Yuseong-gu, Daejeon, 34129, Korea
[3]Dept. of Computer and Software, Hanyang University, Seongdong-gu, Seoul, 04763, Korea
Keywords: 360 video, 3DoF plus, virtual view synthesis

ABSTRACT

A three degrees of freedom plus(3DoFP) video formatting pipeline was presented at MPEG-I Visual. A 3DoFP video gives motion parallax for users' slight translational movement as well as rotation. The given 3DoFP pipeline is based on virtual view synthesis using multiple view color and depth images on which visual redundancies among the given view images are removed. Extracted necessary image areas from redundancy removal process are packed, transmitted and reconstructed to show contents to end users. However, the early researches on view synthesis uses all redundant information, the impact of removed redundant area is not explored much. In this work, we present a method for enhancing final synthesized image quality of the given pipeline dealing with redundancy removal.

1 INTRODUCTION

With the advent of HMD (Head Mounted Display), the interest for immersive media which provide high level of immersion and reality is growing. Immersive contents with virtual spaces created by computer graphics technology are actively produced and used in many areas. However, the difficulty of acquiring real world data and the large size of the acquired data are major problems when producing immersive media for real world scenes. Therefore, immersive contents of real world are not being used much. Even though these problems exist, there are several attempts to provide real world immersive media. The 360 video format which gives users the freedom of rotation, and virtual view synthesis which provides motion parallax caused by translational movement are representative. In the case of 360 videos, data acquisition of real world is carried out using 360 cameras or 360 panoramic stitching and the acquired data is represented in conventional rectangular image formats using equirectangular projection or cube map projection. Data represented in rectangular form are finally encoded using HEVC (High Efficiency Video Coding) encoder to reduce data size. Virtual view synthesis is a method that obtains images which will be shown at virtual viewpoints. This could be achieved by three-dimensional warping of color images

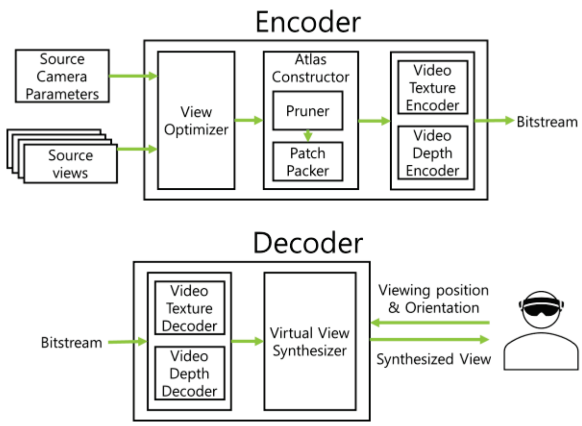

Fig. 1 Brief structure of MPEG 3DoFP Pipeline.

and depth images captured at various viewpoints. When a view image is warped to virtual viewpoint, there could be areas not visible by occlusion at the original viewpoint. These areas could be covered by other view images. This method only uses small amount of multi-view images to show the captured environment.

Recently, MPEG tries to integrate the advantages of 360 video and virtual view synthesis. MPEG presented the concept of 3DoFP video format and test model[1] where basic 3DoFP formatting pipeline is implemented. The brief structure of the pipeline is shown in Fig.1. The pipeline consists of encoder and decoder. In encoder, basic views which are representative views for a given multi-view video content are selected in view optimizer. Other views except basic views are called additional view. After basic view selection, the pruner performs redundancy removal between multi-view 360 videos. This process is called pruning and result of the pruning process is pruned views where redundant information is removed. Pruned views split into patches and rotate to be packed in rectangular video frame which is called atlas in patch packer and finally encoded by HEVC encoder. The decoder reconstructs pruned views from given decoded atlases and synthesizes view by warping and blending reconstructed pruned views to target

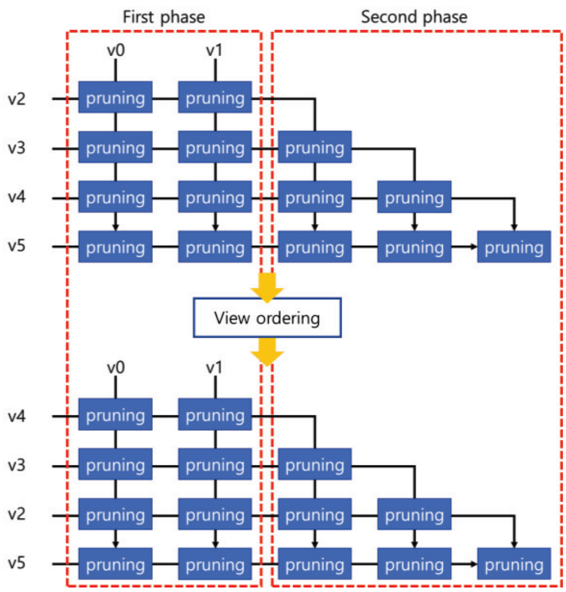

Fig. 2 Concept of changing pruning order.

viewpoint.

The visual quality of 3DoFP pipeline is related with pruning method, because 3DoFP pipeline synthesizes virtual viewpoint image using pruned views. However, early researches on virtual view synthesis usually focuses on the quality of synthesized images while fully taking advantage of redundant areas among multi-view images. In this work, we introduced a new pruning method to enhancing visual quality of 3DoFP pipeline.

2 PRUNING METHOD

In this section we describe the difference between original pruning method included in the MPEG test model and proposed method in this paper. The main purpose of pruning process implemented in the test model is to remove redundancy information that exists between basic views and additional views or across additional views. Individual views are pruned one by one by 3d warping. Early pruned additional views or basic view are warped to pruning target view, then the region which has similar depth value with warped view is removed from target view. The pruning process consists of two phases. In the first phase, redundancy information between basic views and additional views are removed. In the second phase, redundancy information across additional views are removed. As shown in fig.2, the pruned view result is dependent of pruning order in second phase, because the related information corresponding to the additional views at higher order remains more and affects the general shape of the pruned views. Therefore, pruning order could affect visual quality of synthesized view and we introduced pruning order to increase visual quality based on this idea.

2.1 Minimum Overlapping Based View Ordering

The method to determine the pruning view ordering we tested is by using the total amount of overlapping area between each of basic views or early pruned views and additional views. The key idea of this method is that additional views which are not much overlapped by early pruned views may have more visual information of the scene which are not visible in other views. Therefore, fewer numbers of patches in larger scale could be generated instead of large numbers of smaller scale patches. Overlap area between two views is illustrated in Fig. 3. Overlapping area between two views is calculated by following expression

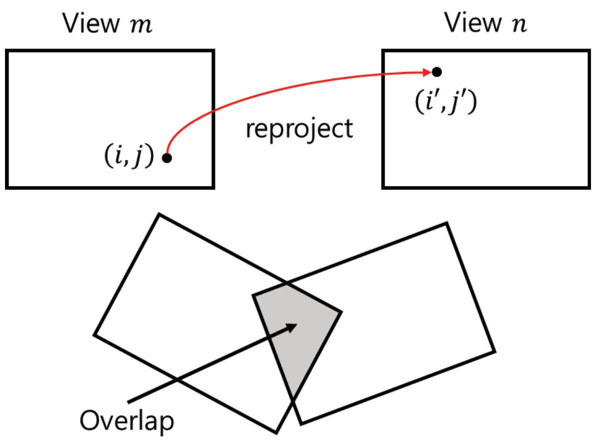

Fig. 3 Illustration of the overlapped area between two views.

$$overlap(m,n) = \frac{\sum w(i,j) \cdot v(i,j) \cdot Fov_m}{\sum w(i,j)} \quad (1)$$

$w(i,j)$ is the spherical weight of 360 video at each position (i,j), $v(i,j)$ means visibility where $v(i,j) = 1$ if (i,j) is visible by both view m and view n. Fov_m is field of view of view m. If there are n basic views or pruned additional views $j_0, j_1, \cdots, j_k, \cdots, j_n$, the total amount of overlapping area l of the selected additional view i can be calculated as below.

$$l = \sum_{k=1}^{n} overlap(i, j_k) \quad (2)$$

Then, the pruning priority is assigned with the inverse order of value of l.

3 EXPERIMENTAL RESULTS

For the evaluation of the proposed pruning method, test sequences including ClassroomVideo (A), TechnicolorMuseum (B), TechnicolorHijack (C), TechnicolorPainter (D), and InterlKermit (E) are used. Table 2 indicates the total number of patches from TechnicolorMuseum sequence individually generated by original test model and minimum overlapping (MinOver) based pruning. It is observed that MinOver could reduce number of patches by about 13.92% with respect to test

Table 1. Objective comparison result of MinOver pruning.

Sequence	High-bitrate BD rate Y-WSPSNR	Low-bitrate BD rate Y-WSPSNR	High-bitrate BD rate VMAF	Low-bitrate BD rate VMAF	High-bitrate BD rate SSIM	Low-bitrate BD rate SSIM	Pixel rate ratio
A	-20.9%	-8.9%	-12.4%	-3.6%	-4.9%	-2.0%	0.00%
B	-4.2%	-4.8%	0.0%	-3.0%	-3.6%	-6.1%	-25.00%
C	34.5%	31.2%	-11.2%	-9.1%	17.0%	12.8%	0.00%
D	-48.3%	-36.1%	-44.1%	-31.4%	-28.5%	-22.4%	0.00%
E	4.8%	0.4%	-6.3%	-6.3%	7.7%	0.6%	-20.00%
Overall	**-6.8%**	**-3.6%**	**-14.8%**	**-10.7%**	**-2.4%**	**-3.4%**	**-6.25%**

Table 2. Total number of patches with MinOver.

	Test Model	MinOver
# of Patch	237	204

Table 3. Total number of atlases generated by MinOver.

Test content	Number of atlases	
	Test Model	MinOver
A	2	2
B	4	3
C	3	3
D	5	4
E	9	7

(a) Test Model

(b) MinOver

Fig. 4 Comparison of atlas maps generated with different view ordering method.

model. Fig. 4 and Table 3 indicates the appearance and the total number of atlas map generated with each method. In case of MinOver, the number of atlas map is reduced across different test contents (B, D and E). Table 1. shows the objective comparison result. Throughout multiple contents, correct evaluation couldn't be performed since the number of generated atlases is different from comparison data. However, in case of contents with the same number of atlases, it can be observed that certain amount of BD-Rate reduction achieved with similar pixel rate.

4 CONCLUSIONS

Based on the inference that the pruning order of views affects the total pixel rate of patch atlas and overall viewing quality. It can be confirmed that the shape of patch atlas and rendering quality are differed according to pruning order. However, changing pruning order can not ensure local visual quality of synthesized views since current method prunes view by view, not local patches. Further researches need to be conducted to deal with this problem.

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

[2] Penner, Eric, and Li Zhang. "Soft 3D reconstruction for view synthesis." ACM Transactions on Graphics (TOG) 36.6 (2017): 235.

[3] Chaurasia, Gaurav, et al. "Depth synthesis and local warps for plausible image-based navigation." ACM Transactions on Graphics (TOG) 32.3 (2013): 30.

[4] Seitz, Steven M., et al. "A comparison and evaluation of multi-view stereo reconstruction algorithms." 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 1. IEEE, (2006).