Similarity-Based Patch Packing Method for Efficient Plenoptic Video Coding in TMIV

HyunHo Kim, and Yong-Hwan Kim
Intelligent Image Processing Research Center, Korea Electronics Technology Institute

Abstract

As immersive video contents have started to emerge in the commercial market, research on it is required. For this, efficient coding methods for immersive video are being studied in the MPEG-I Visual workgroup, and they released Test Model for Immersive Video (TMIV). In current TMIV, the patches are packed into atlas in order of patch size. However, this simple patch packing method can reduce the coding efficiency in terms of 2D encoder. In this paper, we propose patch packing method which pack the patches into atlases by using the similarity of each patch for improving coding efficiency of 3DoF+ video. Experimental result shows that there is a 0.3% BD-rate savings on average over the anchor of TMIV.

1. Introduction

With the increase of commercial interest in deploying virtual reality (VR) and augmented reality (AR), technology for creating and transmitting the immersive content is also evolving. In this immersive media, the image changes in real time according to the movement of the user, and it has to transmit several times more information than the existing 2D video. Because of this, the importance of efficient compression and transmission of immersive media is also growing. In order to support this, ISO/IEC Moving Picture Experts Group (MPEG) is actively working on standardization of “Coded Representation of Immersive Media,” called MPEG-I [1], [2] that aims at advanced immersive VR and AR applications.

In MPEG-I, efficient encoding technology is being discussed for each characteristic of various types of immersive media. Among them, 3DoF+ video provides enhanced immersive visual experience to viewers with slight body and head movements in a sitting position to look around in various directions. To support interactive parallax feature, limited alterations of view position in virtual space of 3DoF+ requires rendering the virtual views at any point of view and depth-based rendering is used. Because of this, 3DoF+ video is captured by 360 camera rig or 2D multi-view videos captured with n by m camera array which called plenoptic video or windowed 6DoF. To efficiently compress and transmit this large volume of 3DoF+ media, MPEG-I Visual workgroup developed an immersive media encoder called Test Model for Immersive Video (TMIV) [3].

2. Test Model for Immersive Video (TMIV)

In current TMIV, multi-view video is compressed through at method of transmitting only the remaining areas called patches by eliminating the redundancy between all available source views which are highly correlated based on a few basic views among them. Then, the selected basic views and the patches are packed into one or more frames called atlases. These atlases are encoded by using 2D video encoder such as HEVC or VVC. This TMIV's compression method can reduce bits by up to 80% compared to simulcast coding the entire view.

In packing process, the patches are sorted by size first and sequentially packed into atlas in a raster scan order to occupy the packing area as small as possible. Because the patch is generated by bounding the remaining areas with rectangular shape, it contains 'valid region' which includes real residual pixels that should be transmitted and 'invalid region' which contains redundant pixels. Therefore, patch packing can be proceeded while allowing rotation of a patch and overlapping each other unless valid regions are invaded.

However, simply packing in order of the patch size, the result of the packed atlas is a seemingly quite disordered collection of patches. And this can reduce the compression efficiency of the codec. Thus, by changing the patch packing result to be more 2D video encoder-friendly, the compression efficiency will be improved.

3. Proposed patch packing method

Plenoptic video which is one of 3DoF+ contents, is captured by arranging multiple cameras n by m. In these plenoptic videos, the similarity between view is high because the distance between each view is close. This means that it has multi-view information on the same object. And for this reason, when plenoptic video is compressed through TMIV, a large number of patches with similar size, position and color information are generated. Therefore, if patches with high similarity can be gathered and packed closely, the coding efficiency can be increased.

There are several ways to calculate similarity by comparing colors or shapes between patches. However, exact comparison is impossible because the size and ratio of the patches to be compared are different for each patch. And it also increases the complexity too much because the patch to be packed has to be compared to every other non-packed patch. For this purpose, a simple method of calculating the average YUV value of each patch and comparing it with each patch was selected in this paper. In addition, we reduce the complexity of the comparison process by comparing only patches that satisfy the following conditions before the comparison operation. 1) Have different viewId, 2) Within a certain search range (for example, ±256 pixel around patch), 3) Have similar size (for example, ±10%).

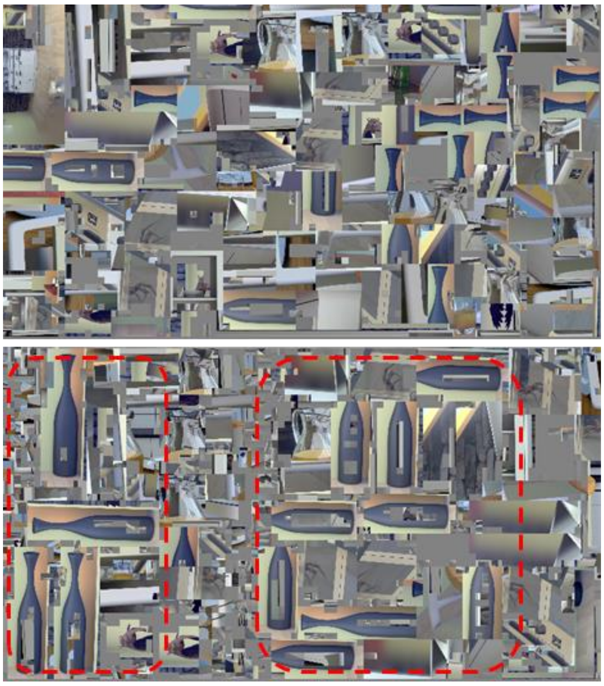

Figure 1. An example of generated atlas (above: Anchor, bottom: proposed)

Figure 1 shows the comparison of generated atlas between the anchor and proposed method with same region. As shown in figure, it is noted that similar patches are gathered in close by proposed method.

Also, patches with high similarity and derived from the same object show similar motion within the them. If an object in a patch moves in certain direction, the motion in another patch generated from same object also has the same direction. Considering these characteristics, the proposed method also unifies the motion of similar patches by using the same rotation information to pack the patches. This will help to make more use of motion information during video coding.

4. Experimental Results

To evaluate the performance of the proposed method in terms of coding efficiency, the end-to-end performance which compares the source views and rendered views synthesized at the same view positions was checked. In the experiments, quality evaluations were carried out with PSNR according to the common test conditions (CTCs) for immersive video [4]. And we used 3 plenoptic contents which are provided by MPEG-I [5]. Table 1 shows the specification about the test sequences which we used.

Table 1. Test sequence specification

Sequence	Resolution	Number of View	Color Format (texture & depth)
OrangeKitchen	3840 x 2160	25 (5x5)	YUV420p10le
ETRIChef	3840 x 2160	25 (5x5)	YUV420p10le
OrangeShaman	3840 x 2160	25 (5x5)	YUV420p10le

Table 2. Performance of proposed method under CTC condition

Sequence	BD-rate
OrangeKitchen	-0.30%
ETRIChef	-0.22%
OrangeShaman	-0.38%
Average	-0.30%

Table 2 shows the end-to-end coding performance of 3DoF+ videos with the atlas constructed by the proposed method in comparison with TMIV as an anchor in terms of Bjontegaard-Delta rate (BD-rate). The proposed method was implemented on the TMIV10.0 [3], and the VVenC [6] was used to compress the constructed atlas according to the CTCs.

It can be observed that there is a minor coding gain in the proposed method over the anchor of TMIV, on average 0.3% BD-rate savings on luma. In detail, most of the gain came from bitrate reduction and there was almost no PSNR difference between reconstructed view results.

5. Conclusion

In this paper, we have presented the new patch packing method in TMIV. Proposed method gathers similar patches and pack into the atlas closely. And we also unify the motion of similar patches by using the same rotation information to pack the patches. The experimental results show that the proposed method gives minor gain in terms of BD-rate.

ACKNOWLEDGEMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00920, Development of Ultra High Resolution Unstructured Plenoptic Video Storage/Compress/Streaming Technology for Medium to Large Space).

[References]

[1] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," ISO/IEC JTC1/SC29/WG11, w16768, Apr. 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] B. Salahieh, J. Jung, A. Dziembowski, "Test model 11 for MPEG Immersive Video," ISO/IEC JTC1/SC29/WG11, w20923, Oct. 2021.

[4] J. Jung, B. Kroon, "Common Test Conditions for MPEG Immersive Video," ISO/IEC JTC1/SC29/WG11, w21230, Oct. 2021.

[5] MPEG content, https://mpegfs.int-evry.fr/mpegcontent

[6] VVenC, https://github.com/fraunhoferhhi/vvenc