Reference Samples Padding for Intra-Frame Coding of Omnidirectional Video

Ning Li*, Shuai Wan* and Fuzheng Yang†
* School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
E-mail: ning_li@mail.nwpu.edu.cn, swan@nwpu.edu.cn.
† State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China
E-mail: fzhyang@mail.xidian.edu.cn

***Abstract*—Due to lack of video coding standard tailored specially to spherical or omnidirectional videos, which are important in Virtual Reality applications, omnidirectional video frames have to be projected onto the 2-Dimension(2D) plane, using the EquiRectangular Projection(ERP) for example, to utilize existing video compression techniques. As there is no information available for prediction outside the left or right border after projection, unexpected artefacts that are generated at the right border of frames degrade the subjective visual quality. Utilizing continuity of the content in the omnidirectional video based on ERP, in this work, an efficient intra-coding method is proposed based on adaptively padding the reference samples at the right border for omnidirectional videos. Using the state of art video encoder, the proposed method can reduce the BD-rate by 0.24% in average for the coding tree units (CTUs) at the right border.**

Fig. 1. Example of ERP projection.

I. Introduction

Providing users with immersive and interactive experiences, Virtual Reality (VR) applications have been gaining its popularity in recent years. Omnidirectional videos or 360 videos play an important role in these applications. To provide a high visual quality and compelling immersive experience, omnidirectional videos are commonly of ultrahigh resolutions and require high transmission bandwidths. Hence efficient compression performance of omnidirectional videos is in crucial need. Existing video coding standards, however, are mostly designed for the planar video, while omnidirectional videos are spherical in nature. It is generally necessary to map omnidirectional videos to the plane domain to adapt to current video coding standards. Typical mapping schemes that have been explored include the EquiRectangular Projection (ERP) [1], [2], Cube-map Projection (CMP) [2], [3], Rhombic Dodecahedron Projection (RDP) [4], etc. ERP that contains one face unfolds the video frames using a longitude and latitude grid. Other projections, like CMP and RDP, contain more than one face. The faces in that projection formats are packed to a 2D rectangular picture with a frame packing method. Among the number of sphere-to-plane projection techniques, the ERP is widely used by most omnidirectional video content providers because of its simplicity. At the same time, the ERP is a format specified in the Omnidirectional Media Applications Format (OMAF) [5] as developed by the Moving Picture Experts Group (MPEG). Fig. 1 gives an example of the ERP projection.

A lot of different compression schemes have been proposed in literature for coding omnidirectional videos to reduce the bitrate, where some use viewpoint-adaptive encoding and streaming approaches. In these schemes [6], [7], [8], [9], view content that is within the viewers areas of interest is encoded and delivered at the highest quality, while the rest of content encoded in a lower quality. Some others improve compression performance by reducing the redundant pixels in projection. In those schemes [10], [11], [12], [13], the ERP is widely studied, in which the redundant pixels in north and south poles are reduced as much as possible. The rest schemes [14], [15], [16], [17], [18] resort to different coding methods for different projection format to improve coding performance. These methods all achieve good compression performance, but they leave out the impact of discontinuous border on the compression performance. As illustrated in Fig. 1, an artificial discontinuity is introduced at the left and right border in the ERP based video, which is essentially continuous in the sphere video, whereas intra prediction in High Efficiency Video Coding (HEVC) [19] is not dedicated to such artificial discontinuities. When an omnidirectional video viewer looks at this region, the visual experience will generally be worse than in the other regions.

In this work, considering the continuity of content at the left and right border in omnidirectional videos based on ERP, we modified the way of padding reference samples at the right border in intra-frame coding in HEVC. When a prediction block (PB) at the right border is being processed, its top-right reference samples can be filled with the reconstructed values at the left border. This method can not only reduce the coding bitrate, but also improve the visual quality.

The rest of this paper is organized as follows. Section II reviews the related work on omnidirectional videos and the process of padding the reference samples in intra prediction. The proposed method for padding the reference samples for intra-frame coding is presented in section III. Section IV describes the experimental results and the conclusion was drawn in section V.

II. Related Work and Padding the Reference Samples in Intra-Frame Coding

A. Related Work

Tile-based streaming of panoramic video is an efficiency way to reduce bandwidth consumption while providing immersive experience to users. In this way, the same video content is encoded into multiple versions at different resolutions, and each version is divided into multiple tiles. According to the viewers viewpoint, a set of tiles are transmitted with the highest resolution, while the remaining parts are transmitted from the low resolution version of the same content. Some researchers have proposed some constructive solutions ([6], [7], [8]) for this field. The work in [9] evaluate the impact of different tiling scheme on the compression efficiency and on bit-rate for transmission.

The ERP format is one of most popular format for omnidirectional video, but there is an obvious flaw in this format, where many redundant pixels are generated in the polar regions due to the stretching effect,. In order to overcome this problem, latitude down-sampling techniques have been proposed, in which the polar regions are down-sampled to reduce the redundant pixels to improve coding efficiency. For example, the work in [10] divides the omnidirectional video frame into tiles, and then resizes each tile to reduce the number of redundant pixels. A similar method is proposed in [11], where the shape of tiles in polar regions changes from rectangular to circle. In [12], the video frame is divided into three tiles: north pole, south pole and equator. The tiles at polar regions will be down-sampled and reconstructed into one tile of the same size as tile at equator. In [13], after down-sampling, the shape of the video frame changes from rectangular to diamond, and then the diamond shape is rearranged into a square.

A number of methods have been devoted to improve the compression performance for omnidirectional video coding. An adaptive QP quantization method has been proposed in [14], in which different CTUs in each coded frame are assigned with different QP offsets according to Weighted-to-Spherically-Uniform peak signal-to-noise ratio(WS-PSNR) weights. A cell based omnidirectional video coding method is proposed in [15], and a method is designed to process these cells in parallel. For different projection formats, authors in [16] propose two kinds of panoramic video compression solutions. One is a content adaptive temporal resolution adaptation scheme for cube map projection. The other is a quantization and rate-distortion optimization scheme for Equirectangular projection. For intra-frame coding and inter-frame coding in HEVC, [17] also proposes two kinds of compression methods for panoramic video based pseudo-cylindrical projection. In the intra-frame coding, the boundary blocks were modified by smoothing the texture in the partial blocks. In the inter-frame coding, non-effective picture area of reference frames at the border is generally filled with the content of effective picture area from the opposite border. In [18], the omnidirectional video is first pre-processed before encoding, which includes two steps. First, the omnidirectional video frame is mapped into panorama via a hyperbolic mirror surface. Next, a resampling step is implemented, in which the image dimension is scaled independently in both the horizontal and vertical direction.

B. Padding the Reference Samples in Intra-Frame Coding

Intra prediction is one of core techniques of HEVC. The first step in intra prediction is to fill the reference samples which come from the adjacent reconstructed blocks. The accuracy of these reference samples will affect the accuracy of subsequent samples prediction.

Given the size of the current prediction block (PB) as $N \times N$, a total of $4N+1$ spatially adjacent reference samples may be used for prediction. The samples come from different coded blocks are partitioned into 5 parts: bottom-left, left, top-left, top, and top-right. The general process of padding reference samples is shown in Fig. 1, where five different colours denote five parts. When the reference samples are all available, they can be used directly, as shown in Fig. 2 (a). When some of the reference samples are not available, the unavailable reference samples are filled with the closest available reference sample in the clockwise direction, as shown in Fig. 2 (b). If all the reference samples are not available, the reference samples are filled with a constant value. If the depth of pixel is 8bit, the constant value is 128, and it is 256 for 10bit.

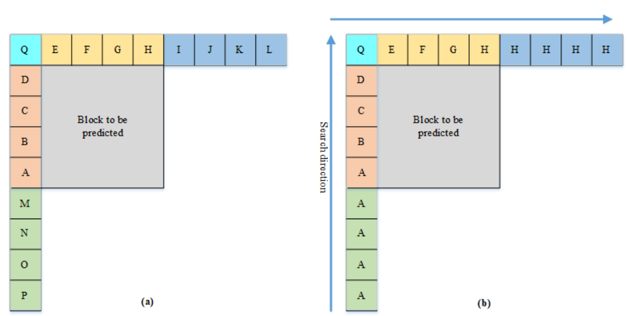

Fig. 2. The process of padding reference samples. (a) the reference samples are all available; (b) the reference samples are partially available (i.e.,bottom-left and top-right unavailable).

III. Proposed Method

One frame image is first split into many CTUs before encoding, and the CTU encoding direction at the same row is from left to right. As shown in Fig. 3, a block with the solid line denotes one CTU, and the arrow represents the encoding direction in one row. During encoding, the CTU in the previous row is always available for CTU encoding in the current row. When encoding the last CTU in a row, there is no information that can be used on the right, therefore the above-right reference samples of PBs at the right border have to be filled with the same values. When the prediction mode is between 27 and 34, these identical reference samples will be used to predict part of pixels values of PB, which will degrade the accuracy of prediction values and lead to the increase in the distortion and bitrate.

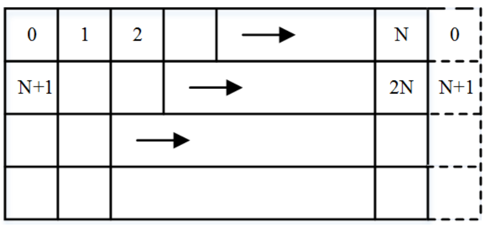

Fig. 3. Illustration of CTU segmentation and encoding direction.

Considering the fact that the content of an omnidirectional video in ERP is continuous at the left and right borders, and according to the encoding direction, the CTU on the left is encoded prior to the right. Based on these two observations, we can regard the encoded CTUs at the left border as the reference for intra-prediction for that at the right border, as shown in dotted lines in Fig. 3. That is, the block with the dotted line denotes the CTU used as reference for intra-prediction for the last CTU in that row, which is identical to the CTU at the left border whereas does not really exist at the right border. In this way, not only the CTUs on the top-right can be used, but also the CTUs on the left can be utilized. In this design, the reference samples of PBs in the CTU at the right border can be properly filled with values, which will greatly increase the accuracy of prediction. Next, we will describe the proposed process of padding the reference samples in detail.

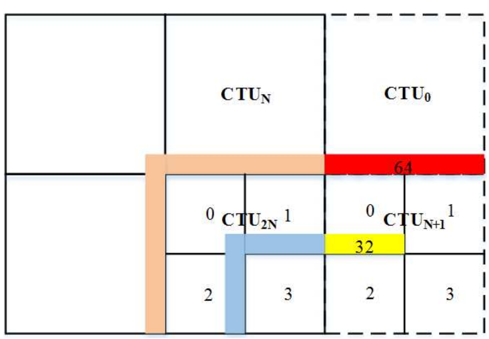

Fig. 4. Illustration of padding reference samples.

As shown in Fig. 4, CTU_N and CTU_{2N} denote the last CTU at the right border in their rows, respectively. CTU_{2N} represents the CTU to be encoded, which has no physically adjacent CTU at the right border. In the proposed work, CTU_0 and CTU_{N+1} which are the first CTUs at the left border in their rows, respectively, will be padded at the end of the rows for reference purposes. In this design, when the size of PB is equal to the size of CTU, the reconstructed values at the bottom of CTU_0 are used as the reference samples for CTU_{2N}, as shown in the red rectangle block. When CTU_{2N} is split into four 32×32 blocks, the top-right reference samples of PB_1 are filled with the first half pixels in the red rectangle block. The top-right reference samples of PB_3 are filled with the reconstructed values at the bottom of the coding unit (CU) with the number 0 in CTU_{N+1}, as shown in the yellow rectangle block. When the CU continues to be split, the top-right reference samples of PB at the right border is always available. Fig. 5 gives a flowchart of padding the top-right reference samples in practice, where the implementation steps are given as follows.

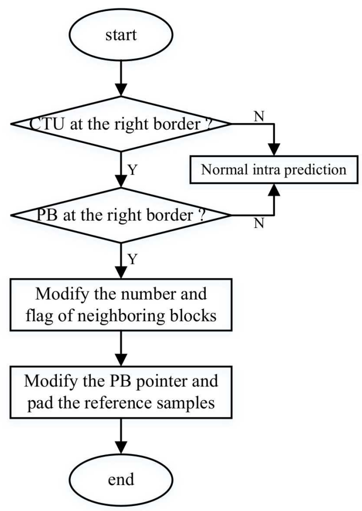

Fig. 5. Flowchart of padding the top-right reference samples.

First, we determine whether a CTU is on the right border. A CTU is numbered in the raster scan order, starting from 0. If the modulo value of the current CTU number plus one and total number of CTU in the frame width is equal to zero, it shows that the current CTU is on the right border, and then we determine whether the current PB is on the right border. If the current CTU is not on the right border, the normal intra-frame coding is implemented.

Second, a CTU can be split into 256 CUs with the minimum size, which is numbered in the Z-order scan. Whether a PB is on the right border is judged based on the width of PB and the index number of the top-left CU with the minimum size. If the current PB is on the right border, the next step is implemented. If not, the normal intra-frame coding is executed.

Third, because the top-right CU is available in this case, the number of neighboring blocks should be increased and the corresponding flag should be set true.

Finally, the data pointer of the reference samples is modified. When the data pointer exceeds the width of the current

PB, it indicates that the top-right reference samples will be filled. The pointer is needed to move to the starting position of the frame image in this line. The moved data pointer position is:

$$piRef = piOri - iPic - (uiCTU \cdot uiCTU_{num} - uiWid). \quad (1)$$

where $piRef$ and $piOri$ represent the starting pointer position of the reference samples that will be filled with and that of the current PB, respectively. Here $iPic$ and $uiCTU$ denote the width of the extended frame image and the size of CTU, respectively. $uiCTU_{num}$ represents the total number of CTU in frame width, and $uiWid$ denotes the width of the PB.

IV. Experimental Results

The proposed method was implemented in the HEVC reference software version 16.16 [20]. 300 frames of each omnidirectional videos were encoded using intra main configuration and with four quantization parameters (QP), i.e., 22, 27, 32, and 37. The coding performance and bitrate saving are measured in terms of the Bjϕntegaard delta BD-rate metric [21], where the negative values represent how much the bitrate is reduced for the same peak-signal to noise ratio (PSNR). In order to better show the coding performance of the method, the total bits and PSNR values of the CTUs with the maximum size at the right border were counted. As it can be observed from table 1 that the proposed method improves the coding performance in the BD-rate by 0.24% in average for the luminance component for those CTUs.

TABLE I
Experimental Results

YUV	**BD-rate**		
	Y	U	V
AerialCity	-0.18%	-0.39%	-0.12%
DrivingInCity	-0.24%	-0.22%	-0.01%
DrivingInCountry	-0.31%	-0.48%	-1.30%
Habor	-0.48%	-0.94%	-0.19%
KiteFlite	-0.05%	-0.26%	-0.18%
PoleVaultle_le	-0.15%	-0.29%	-0.38%
SkateboardTrick_le	-0.34%	-0.91%	-1.44%
Train_le	-0.13%	-0.38%	-0.37%
Average	-0.24%	-0.48%	-0.50%

V. Conclusions

In this paper, we proposed an efficient reference padding method for intra-coding of the ERP-based omnidirectional videos. The proposed method provides reasonable reference samples for a PB at the right border, where content continuity in the omnidirectional video is considered. Experimental results demonstrated the effectiveness of the proposed method.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 61371089 and Grant 61571337.

References

[1] J. P. Snyder, "Flattening the earth, two thousand years of map projections," *University of Chicago Press*, Dec. 1998.

[2] Yan Ye et al, "Algorithm description of projection format conversion in 360lib," JVET-E0084, Geneva, January 2017.

[3] King To Ng, Shing Chow Chan, and Heung Yeung Shum, "Data compression and transmission aspects of panoramic videos," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, 2005, pp. 82–95.

[4] Chi Wing Fu, Liang Wan, Tien Tsin Wong, and Chi Sing Leung, "The rhombic dodecahedron map: an efficient scheme for encoding panoramic video," in *IEEE Transactions on Multimedia*, vol. 11, no. 4, 2009, pp. 634–644.

[5] "Omnidirectional media application format (omaf)," *[Online]. https://mpeg.chiariglione.org/standards/mpeg-a/omnidirectional-media-application-format*.

[6] Alireza Zare, Alireza Aminlou, Miska M Hannuksela, and Moncef Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," in *Proceedings of the 2016 ACM on Multimedia Conference* 2016, pp. 601–605.

[7] Xavier Corbillon, Gwendal Simon, Alisa Devlic, and Jacob Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *IEEE International Conference on Communications*, 2017, pp. 1–7.

[8] Mohammad Hosseini and Viswanathan Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *IEEE International Symposium on Multimedia*, 2016, pp. 107–110.

[9] Ramin Ghaznavi-Youvalari, Alireza Zare, Huameng Fang, Alireza Aminlou, Qingpeng Xie, Miska M. Hannuksela, and Moncef Gabbouj, "Comparison of HEVC coding schemes for tile-based viewport-adaptive streaming of omnidirectional video," in *IEEE International Workshop on Multimedia Signal Processing*, 2017, pp. 1–6.

[10] Matt Yu, Haricharan Lakshman, and Bernd Girod, "Content adaptive representations of omnidirectional videos for cinematic virtual reality," in *The International Workshopon on Multimedia Signal Processing*, 2015, pp. 1–6.

[11] Jisheng Li, Ziyu Wen, Sihan Li, Yikai Zhao, Bichuan Guo, and Jiangtao Wen, "Novel tile segmentation scheme for omnidirectional video," in *IEEE International Conference on Image Processing*, 2016, pp. 370–374.

[12] Ramin Ghaznavi Youvalari, Alireza Aminlou, and Miska M. Hannuksela, "Analysis of regional down-sampling methods for coding of omnidirectional video," in *2016 Picture Coding Symposium*, 2016, pp. 1-5.

[13] S. H. Lee, S. T. Kim, E. Yip, B. D. Choi, J. Song, and S. J. Ko, "Omnidirectional video coding using latitude adaptive down-sampling and pixel rearrangement," in *Electronics Letters*, vol. 53, no. 10, 2017, pp. 655–657.

[14] Y. Sun and L. Yu, "Coding optimization based on weighted-to-spherically-uniform quality metric for 360 video," in *IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4.

[15] Jaejoon Kim, Taeho Kim, Daegyu Lee, and Youngback Kim, "On realization of modified encoding/decoding for high capacity panoramic video," in *IEEE International Conference on Digital Information Management*, 2009, pp. 122–127.

[16] Minhao Tang, Yu Zhang, Jiangtao Wen, and Shiqiang Yang, "Optimized video coding for omnidirectional videos," in *IEEE International Conference on Multimedia and Expo*, 2017, pp. 799–804.

[17] Ramin Ghaznavi Youvalari, Alireza Aminlou, Miska M. Hannuksela, and Moncef Gabbouj, "Efficient coding of 360-degree pseudo-cylindrical panoramic video for virtual reality applications," in *IEEE International Symposium on Multimedia*, 2017, pp. 525–528.

[18] Ingo Bauermann, Matthias Mielke, and Eckehard Steinbach, "H. 264 based coding of omnidirectional video," in *Computer Vision and Graphics*, Springer Netherlands, 2006.

[19] G. J. Sullivan, J. Ohm, Woo Jin Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, 2012, pp. 1649–1668.

[20] "high efficiency video coding (hevc) reference software hm," *[Online], Available: https://hevc.hhi.fraunhofer.de/*.

[21] Gisle Bjontegaard, "Calculation of average psnr differences between rd-curves," *ITU-T VCEG-M33*, April, 2001.