NESTED POLYGONAL CHAIN MAPPING OF OMNIDIRECTIONAL VIDEO

Kashyap Kammachi-Sreedhar, Miska M. Hannuksela
Nokia Technologies, Tampere, Finland
E-mail: firstname.lastname@nokia.com

ABSTRACT

Conventionally, the equirectangular projection (ERP) has been used for representing 360-degree panorama videos. However, ERP suffers from over stretching in polar regions and thus increases the bitrate and the encoding/decoding complexity. This paper presents the nested polygonal chain packing method where the top and bottom stripes of an ERP picture are resampled sample-row-wise. The sampling ratio is a linear function of the sample row being processed. The height of the top or bottom stripe can be for example a quarter of the height of the ERP picture. The resampled sample rows are arranged in nested polygonal chains.

According to the presented experimental results, the proposed nested polygonal chain packing provides on average 6.1% bitrate reduction compared to ERP.

***Index Terms*—** Mapping, Projection, Panorama video, Virtual Reality

1. INTRODUCTION

Virtual reality (VR) applications aim at providing a complete immersion experience of the real world. Thus, in order to simulate the sense of immersion, VR applications make use of 360-degree panorama videos along with head-mounted-displays (HMDs).

Traditionally, 360-degree panorama videos are represented with the equirectangular projection (ERP). However, ERP horizontally stretches the polar areas of the panorama scene. The sampling density towards the poles is proportionally greater than that in the equatorial region. This results in redundant pixels being encoded and decoded in the areas close to the poles, which in turn increases encoding and decoding complexity. Additionally, the modern video coding techniques, for example, the Advanced Video Coding (H.264/AVC) and the High Efficiency Video Coding (H.265/ HEVC) do not take into account the 360-degree geometry of the panorama video and may therefore result in a decreased rate-distortion (RD) performance. Furthermore, a full immersion experience in VR requires the use of stereoscopic panorama videos with high resolution of at least 3840×1920, per view and high frame rates in the range of at least 60 Hz [1], leading to high bitrates for storage and streaming.

Projections representing the omnidirectional view of the real world have been studied for many years [2][3], but most commonly ERP has been used.

The drawbacks of ERP have driven researchers to invent new projection formats. Additional projection formats, such as cubemap and rhombic dodecahedron projections, have been studied in [4][5]. Since ERP is broadly used in content production tools, alternative projection formats would require projection conversion from ERP prior to compression for the distribution. Such a conversion inherently involves interpolation of sample values and hence may result in picture quality degradation. ERP-based formats are hence preferred to reach the best quality in combination with ERP-based content production.

In [6], Yu et al. divide the ERP video into a number of stripes that are horizontally downsampled prior to encoding. The work in [7] analyzed the performance of three-stripe coding according to [6] as well as an in-loop temporally adaptive variation of such regional downsampling. In [8], Li et al. also divide the ERP pictures into a number of stripes but use cylinder top and bottom faces to represent the top and bottom stripes, respectively. Even though the proposed schemes reduce the stretching in the polar region and reduce the pixel density proportionally, they introduce boundary artifacts between the resampled stripe boundaries and thus affect the visual experience of the user. The boundary artifact can be concealed by using overlapping stripes, which however lead to bitrate overhead and hence reduce the compression performance.

In this paper we propose the nested polygonal chain mapping of ERP panoramas. The mapping methods adjust the sampling density in the polar areas such that it is approximately equal to the sampling density in the equatorial region of the sphere.

This paper has been organized in the following way. In Section 2 the proposed nested polygonal chain mapping schemes is discussed in detail. Section 3 describes the experimental setup and quality assessment methods used for obtaining the results. Finally, conclusions are drawn in Section 4.

2. NESTED POLYGONAL CHAIN MAPPING (NPCM) METHODS

At first, we analyze the sampling on the sphere in comparison to ERP. This is followed by the discussion of the proposed three variants of the NPCM method.

2.1. Sampling on Sphere

An ERP picture of height H and width W, as shown in Figure 1a, covers 360 degrees (2π radians) in horizontal direction and 180 degrees (π radians) in vertical direction. As discussed in [7], each sample row in an ERP picture has a corresponding circle on the spherical surface with a defined latitude coordinates as shown in Figure 1b.

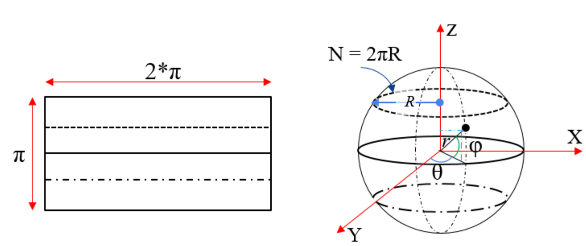

Figure 1. Sampling on a) Equirectangle projection and on b) Spherical surface.

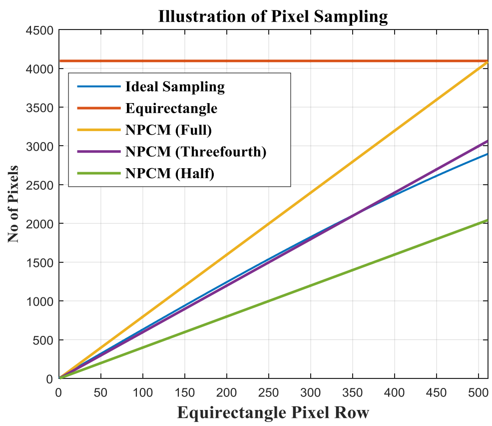

Figure 2. Illustration of pixel sampling on sphere compared to Equirectangle and NPCM methods (top-most 512 sample rows of 4096×2048 ERP picture).

Each sample on a spherical surface can be represented by three parameters namely the radial distance (r), and the two polar coordinates (θ, φ). Considering the circumference of the sphere to be equal to width W of the ERP panorama, the radius r of the sphere is equal to:

$$r = \frac{W}{2\pi} \quad (1)$$

For a given sample row in an ERP panorama, the ideal number of samples that would yield the same sampling density as in the equator along the corresponding circle/latitude of the spherical surface is given by:

$$N = round(2\pi R) \quad (2)$$

R is the radius of the circle/latitude corresponding to the sample row in an ERP panorama and is given by:

$$R = r \times \cos(\varphi) \quad (3)$$

The angle φ is the angle between the radius of circle/latitude and the radial line connecting the center of sphere to any point in the circle. This angle is given by Equation 4.

$$\varphi = \pi \times \left(\frac{n+0.5}{H} - 0.5\right) \quad (4)$$

where n is the number of the sample row in an ERP panorama from the center towards each pole in the range of 1 to H, inclusive. Based on the above argument and as illustrated in Figure 2, it can be seen that an ERP panorama picture of size 4096×2048 ($W\times H$) has samples distributed in the ratio of 1:1 near the equator to 1:4096 near the polar points as compared to sample distribution on the spherical surface. This clearly indicates that a lot of bits is wasted for representing the redundant data in ERP panoramas. Thus, in order to optimize the samples in each row of ERP panorama while still having the same ERP projection for 360-degree panorama representation, we propose the nested polygonal chain mapping which is discussed in Section 2.2.

2.2. NPCM Scheme

In the nested polygonal chain mapping method, an ERP panorama picture is split into three stripes. Each sample row of the top stripe is downsampled row-wise. The resampled top-most sample row is placed in the middle of the packed rectangular region, surrounded by the resampled second top-most sample row, which in turn is surrounded by the resampled third top-most sample row of the top stripe. A similar operation is also performed for the bottom stripe but in an opposite order of sample rows, which means that the bottom-most sample row of the bottom stripe is resampled and placed in the middle of the corresponding packed rectangular region. The process is illustrated in Figure 3a to Figure 3c.

In the case of full mapping, each resampled sample row in the packed rectangular region is completely surrounded by the resampled sample row of the correspondingly consecutive row of the ERP panorama. Thus, the sample rows occupy rectangular region of area ($2h \times 2h = 4h^2$), where h, is the number of sample rows of the ERP panorama packed in the rectangular region.

For the case of three-fourth mapping, each resampled sample row in the packed rectangular region only partially surrounds the resampled sample row of the consecutive row of the ERP panorama. Thus, only occupies the three-fourth part of the rectangular region with the effective area being $3h^2$ and the remaining area of h^2 may be padded with zeros to make it a rectangular packing.

Finally, in the case of half mapping, each resampled sample row in the packed rectangular region surrounds only half of the resampled sample row of the consecutive row of the ERP panorama. Thus, occupies only half part of the rectangular region having an effective area of $2h^2$.

With all the above three mappings the NPCM method linearly approximates the ideal number of samples along each row of the ERP panorama and is illustrated in Figure 2. Number of samples for the sample row m in the range of 1 to h, inclusive, is given by Equation 5.

$$P_{NPCM}(m) = (m - 0.5) \times C \quad (5)$$

where C, is a constant that takes values of 8, 6 and 4 for the full, three-fourth and half mapping cases, respectively. The values of C

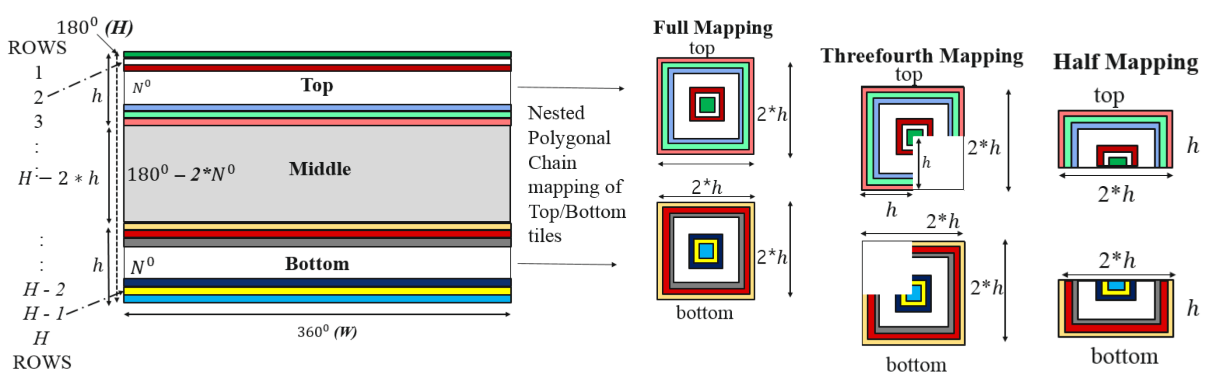

Figure 3. An illustration of Nested Polygonal Chain Mapping of Top and Bottom Stripes of Equirectangle panorama and the corresponding variants in a) Full mapping, b) Three-fourth Mapping and c) Half Mapping.

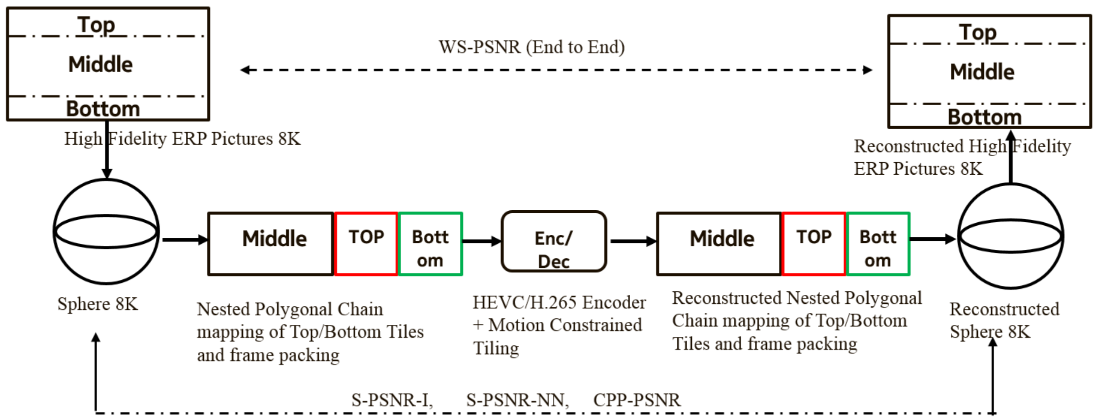

Figure 4. An illustration of simulation and JVET quality assessment methods.

are based on the rectangular geometry of the packed region. As it can be seen in Figure 2, the three-fourth mapping method closely approximates the ideal sampling along each row of the ERP panorama, while the case of full and half mappings deviate along either sides of the ideal curve, respectively.

The proposed NPCM methods come with the following benefits: 1) the sampling density towards poles is closer to that in the equator. 2) The spatial resolution of the packed rectangular region increases gradually in a sample-row-wise manner so that there is no visible boundary between the middle stripe and the top and bottom stripes. Therefore, it avoids the overhead of overlapping between the stripes. 3) The packed top and bottom stripes are rectangular and hence suitable for block-based video coding standards. However, additional processing consists only of one-dimensional resampling on top of a conventional ERP content creation and rendering processing chains.

3. EXPERIMENTAL SETUP

The simulation pipeline is based on the common test conditions and evaluation procedure for 360-degree video as defined by the Joint Video Exploration Team (JVET) [9] and is illustrated in Figure 4.

As suggested in [9], the following ten test sequences were used in simulations: Train_le, SkateboardingTrick_le, SkateboardInLot, ChairLift, KiteFlite, Harbor, PoleVault_le, AerialCity, DrivingInCity, DrivingInCountry. The first two sequences have 600 frames each with a frame rate of 60 Hz, while the remaining sequences have 300 frames with a frame rate of 30 Hz. The first six sequences have a resolution of 8192×4096 and the remaining four sequences have a resolution of 3840×1920.

As a pre-processing step the high fidelity resolution (HF-ERP) version of ERP videos were resampled to the recommended frame sizes of the coding ERP videos and thus form the anchor for comparing the proposed NPCM methods. Since the HF-ERP format is used for generating all the proposed schemes it has been decided by JVET [9] to resample the HF-ERP videos in order to reduce the bias towards ERP format. As a second anchor method we used the segment sphere projection (SSP) [8] with three stripes where the top and bottom stripes had 45 degree coverage. For the case of NPCM methods the HF-ERP videos were resampled such that the effective picture area in NPCM is within 5% range of the coding ERP. As per [9], the effective picture area excluded the count of black sample area, if present in any of the mapped representations. The ERP was divided into three regions such that the top and bottom stripes cover 45 degrees each and the middle region covering the central 90 degree of the ERP. It was also made sure that the resolutions of each region was aligned to the size of a multiple of coding tree unit (CTU) of the High Efficiency Video Coding (HEVC) standard [10]. The resolutions of the respective mappings are given in Table I. w and h indicate the width and height of the resample top/bottom stripes, respectively. W_F and H_F represent the width and height of the packed frame, respectively.

As shown in Figure 4, all the encodings were performed with HEVC reference software (HM) version 16.7 [11] patched with the motion constrained tiling concept of [12]. However, de-blocking filter was disabled across the tile boundaries. Motion constrained coding enables us in packing the three regions (middle stripe, top/bottom packed regions) into a single image frame and thus facilitates the use of single decoders, which could be helpful for the existing multimedia devices.

Each video sequence was coded with random access configuration with 10-bit main profile and a search range of 256 and GOP size of 8 was used. The following four quantization parameter values were used: 22, 27, 32 and 37. For all the conversion between projections the 360Lib software from JVET test conditions were used, which internally made use of 6-taps Lanczos filter for luma and 4-taps Lanczos filter for chroma samples, respectively.

After decoding, the objective quality was measured between the respective reconstructed coded mappings re-projected to HF-

TABLE I. CODING RESOLUTIONS OF ERP, SSP AND NPCM METHODS.

PROJECTION	HF-ERP (8192×4096)		HF-ERP (3840×1920)	
	W_F×H_F	w×h	W_F×H_F	w×h
ERP	4096 × 2048	-	3840 × 1920	-
SSP	7296 × 1216	1216 × 1216	6144 × 1024	1024 × 1024
NPCM (FULL)	6912 × 1152	1152 × 1152	5760 × 960	960 × 960
NPCM (3/4)	7296×1216	1216 × 1216	6144 × 1024	1024 × 1024
NPCM (1/2)	7608 × 1280	1280 × 640	5120 × 1024	1024 × 512

TABLE II. COMPRESSION PERFORMANCE OF NPCM METHODS COMPARED TO ERP AND SSP IN BD-RATE(%) AND WS-PSNR (END TO END).

SEQ	(ANCHOR) ERP VS NPCM			(ANCHOR) SSP VS NPCM		
	FULL	THREE-FOURTH	HALF	FULL	THREE-FOURTH	HALF
1	4,2%	4,5%	5,8%	-0,9%	-0,7%	0,5%
2	1,5%	1,8%	1,4%	-1,0%	-0,8%	-1,3%
3	-11,0%	-12,9%	-13,7%	2,2%	0,2%	-0,5%
4	-11,5%	-10,7%	-7,1%	-0,5%	0,3%	4,5%
5	-8,1%	-8,9%	-5,6%	0,0%	-1,2%	2,8%
6	-6,6%	-5,4%	-4,5%	-2,6%	-1,6%	-0,4%
7	-8,9%	-7,4%	-3,5%	-1,6%	0,9%	5,2%
8	-1,9%	-2,9%	1,7%	0,4%	-0,5%	4,2%
9	-4,1%	-3,1%	-1,5%	-1,8%	-0,2%	1,5%
10	-14,6%	-15,3%	-12,9%	3,1%	3,2%	6,5%
AVG	**-6,1%**	**-6,0%**	**-4,0%**	**-0,3%**	**0,0%**	**2,3%**

ERP and the original HF- ERP videos. The quality assessment metric of WS-PSNR (end to end) [14] has been used for comparisons. Other quality metrics namely: CPP-PSNR of [15] and two variants of S-PSNR in [13] have also been defined by JVET in [9] but they help in only evaluating the coding tools and not with comparing different mappings and hence have been omitted.

Table II shows the compression bitrate performance of the NPCM methods as compared to the ERP and SSP. The measured bitrate savings and compression performance measurements are expressed in terms of Bjøntegaard Delta-rate (BD-rate) criterion [16] for luma pictures. It can be observed that in a majority of the test sequences the rate-distortion performance of NPCM is better in all mapping cases as compared to ERP, while a minor bitrate increase was obtained in Train_le and SkateboardingTrick_le sequences. These two sequences do not have any content captured in the top/bottom stripes and hence were filled with pixels to give a blurring effect. Thus, the objective metric performance of ERP in these two sequences do not necessarily reflect the subjective qualities. Also, as seen in Table II, the SSP BD-rate performance is hand in hand with the NPCM methods, however with sequence wise variations due to different content favoring different mappings. Additionally, it should also be noted from Figure 5, that the SSP introduces boundary artifacts near tile boundaries as compared to the proposed NPCM methods and hence would require overlapping tiles to conceal the artifacts resulting in a higher bitrate.

4. CONCLUSION

In this paper, we proposed three variants of nested polygonal chain mapping (NPCM) method namely: a) Full NPCM, b) Three-fourth NPCM and c) Half NPCM. The proposed methods linearly approximate the ideal sampling density on the sphere surface. Experimental results showed that the proposed full, three-fourth and half NPCM mapping methods can reduce the bitrate by 6.1%, 6.0% and 4.0%, respectively, on average when compared to the conventional equirectangular projection.

5. REFERENCES

[1] Oculus notes on VR Best Practices. https://developer3.oculus.com/documentation/intro-vr/latest/concepts/bp_intro/ (January-17-2017).

[2] J. P. Snyder, “Flattening the earth: two thousand years of map projections”, University of Chicago Press, ISBN: 9780226767475, December. 1997.

[3] L. Zelnik-Manor, G. Peters and P. Perona, “Squaring the circle in panoramas”, Tenth IEEE International Conference on Computer Vision, October. 2005.

[4] C. Gruenheit, A. Smolic, and T.Wiegand, “Efficient representation and interactive streaming of high-resolution

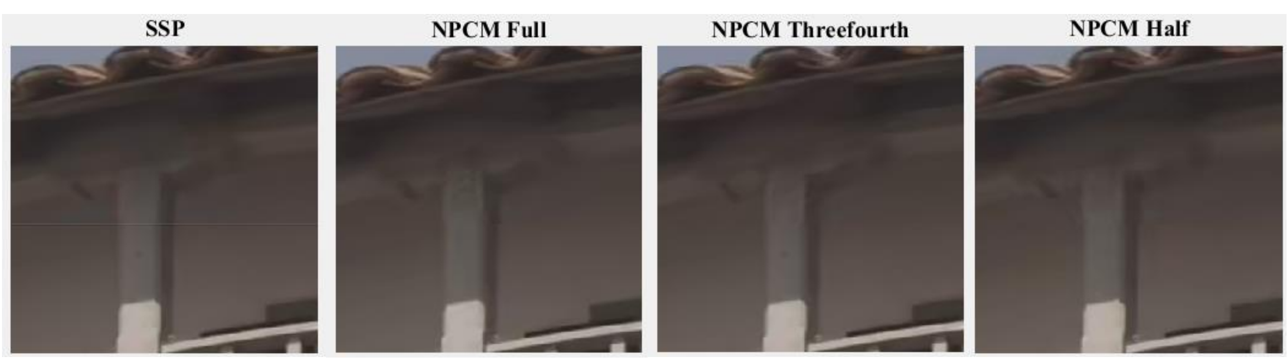

Figure 5. Comparison of tile boundaries in KiteFlite sequence between a) SSP b) NPCM Full, c) NPCM Three-fourth and d) NPCM Half for QP = 37.

panoramic views." IEEE International Conference on Image Processing, 2002.

[5] C. Fu, L. Wan, T. Wong, and C. Leung, "The Rhombic Dodecahedron Map: An Efficient Scheme for Encoding Panoramic Video", IEEE Transaction on Multimedia, vol. 11, no. 4, April 2009.

[6] M. Yu, H. Lakshman, and B. Girod, "Content adaptive representations of omnidirectional videos for cinematic virtual reality", in Proceedings of the 3rd International Workshop on Immersive Media Experiences, pp.1–6, ACM, 2015.

[7] R. Ghaznavi-Youvalari, A. Aminlou, M. M. Hannuksela, "Analysis of Regional Down-Sampling Methods for Coding of Omnidirectional Video", Picture Coding Symposium, December, 2016.

[8] J. Li, Z. Wen, S. Li, Y. Zhao, B. Guo, J. Wen, "Novel Tile Segmentation Scheme for Omnidirectional Video", IEEE International Conference on Image Processing, 2016.

[9] J. Boyce, E. Alshina, A. Abbas, Y. Ye, "JVET common test conditions and evaluation procedures for 360° video", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D1030_v2, JVET 4th Meeting: Chengdu, 15–21 October 2016.

[10] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, December. 2012.

[11] High Efficiency Video Coding (HEVC), Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, DOI= https://hevc.hhi.fraunhofer.de/, (April- 2016)

[12] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," proceedings of ACM MM 2016.

[13] M. Yu, H. Lakshman, B. Girod, "A frame-work to evaluate omnidirectional video coding schemes, in Mixed and Augmented Reality", IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2015.

[14] Y. Sun, A. Lu, L. Yu (Zhejiang Univ.), "AHG8: WS-PSNR for 360 video objective quality evaluation", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0040, 4th Meeting, October. 2016.

[15] V. Zakharchenko, E. Alshina, A. Singh, A. Dsouza (Samsung), "AhG8: Suggested testing procedure for 360-degree video", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0027, 4th Meeting, October. 2016.

[16] G. Bjøntegaard. "Calculation of average PSNR differences between RD-curves". ITU-T Video Coding Experts group document VCEG-M33, April.2001.