ASYMMETRIC CIRCULAR PROJECTION FOR DYNAMIC VIRTUAL REALITY VIDEO STREAM SWITCHING

Yueming Wang, Ronggang Wang, Zhenyu Wang, Wen Gao

Shenzhen Graduate School, Peking University
Shenzhen, 518055, China

ABSTRACT

To provide full panoramic video content, the transmission bandwidth and playing complexity of virtual reality (VR) video are much higher than that of traditional video. Dynamic stream switching is an efficient scheme which stores multiple asymmetric projections of panoramic video in server side, and transmits only one asymmetric projection (with much lower resolution than full panoramic video) according to the direction of user's head. In this paper, we propose an asymmetric circular (ASC) projection method to generate the asymmetric projection of panoramic video. ASC uses equal area projection in the main viewpoint area to ensure the high and uniform quality of this area, and a projection with decreasing sampling density in other areas. Experimental results show that the bit rate and playing complexity of VR video are greatly reduced with the proposed ASC method. ASC is superior to state-of-the-art method and has been adopted in the standard of IEEE 1857.9-Immersive Visual Content Coding.

Index Terms**—**virtual reality, video streaming, panoramic video, spherical video projection, IEEE 1857.9

1. INTRODUCTION

Virtual reality (VR) video provides an immersive experience by presenting a 360° panoramic video. Users usually watch VR video through head mounted displays (HMDs), which can track users head direction and render corresponding video content to provide a sense of immersion. Unfortunately, existing encoding and storage techniques do not support the processing of spherical video. Therefore, the spherical panoramic video needs to be projected onto the two-dimensional plane for storage and encoding.

There are many ways to map a spherical panoramic video onto a two-dimensional plane [1]. The most common projection method for full 360-degree video is equirectangular projection [2] (ERP), but ERP has a serious oversampling in the high latitude area and wastes the transmission bandwidth. Afterwards, equal-area cylindrical (reduce the sampling density in latitude direction to compensate for oversampling in longitude direction) [2], cube projection (project the sphere onto the cube) [3], adjusting tiles projection (divide video of equirectangular format into tiles, the tiles are down-sampled based on the content of the video) [4], tile segmentation scheme (divide video of equirectangular format into tiles, then the bipolar areas are projected into circles, and other tiles are down-sampled based on the latitude) [5] were proposed. These projection methods can somewhat reduce the transmission bandwidth of panoramic video. But, since the panoramic video needs to provide 360 degrees of video content, so the transmission bandwidth and the playing complexity of panoramic video are still much larger than that of traditional video. Moreover, with the development of virtual reality video capture devices, the resolution of panoramic video is increasing to 8k and even higher-resolution. In order to address the above issues, the IEEE 1857.9-Immersive Visual Content Coding working group has been set up to standardize efficient panoramic video projections and coding methods [6].

Since one viewer has only one particular viewpoint at one moment. Dynamic stream switching is an effective scheme to reduce both transmission bandwidth and playing complexity. In this scheme, a panoramic video is represented by multiple asymmetric projections with different main viewpoints. At one moment, according to the head direction of viewer, only one asymmetric projection of panoramic video with a main viewpoint that is the nearest to the viewer's viewpoint is transmitted, and at the next moment when the direction of viewer's head is changed, another asymmetric projection of panoramic video with a different main viewpoint is transmitted (if necessary). Usually the asymmetric projection of panoramic video with main viewpoint uses a higher sampling density in the main viewpoint area and a lower sampling density in the other area, and its resolution is much lower than that of panoramic video.

Geert Van der Auwera has proposed an asymmetric projection scheme called truncated pyramid projection (TSP) [7], which is the down-sampling of the cube projection. TSP does not down-sample the front face (which is set as the main viewpoint area) of the cube projection, and performs incremental down-sampling on the other faces of the cube projection, this projection can effectively save the bandwidth and decoding complexity. However, the main viewpoint range of TSP is fixed (about 90°). Additionally, the front face of the cube projection uses perspective projection, and its sampling density is low in the middle and high in the around.

In this paper, we propose a new flexible asymmetric projection called asymmetric circular projection (ASC). Compared with TSP [7], our proposed ASC projection has at least the following two advantages:

- ASC uses equal area projection in the main viewpoint area, which ensures that the sampling density in the main viewpoint area is high and uniform.
- The range of main viewpoint area is flexible and can be customized according to the application scenario, network conditions, or other factors.

ASC uses a decreasing sampling density in left area other than main viewpoint area and the further from the center of the main viewpoint, the lower the sampling density. Therefore, the transition from main viewpoint area and the other areas is smooth. Experimental results show that our proposed ASC can greatly save the transmission bandwidth and reduce the playing complexity. ASC is superior to state-of-the-art method of TSP and it has been adopted in the standard of IEEE 1857.9-Immersive Visual Content Coding.

The rest of this paper is organized as follows. Section 2 presents the proposed asymmetric circular projection method. Section 3 provides the experimental results to testify the effectiveness of the proposed method. At last, this paper is concluded in Section 4.

2. PROPOSED METHOD

We divide the sphere of VR video into three areas according to the angle between the area and the main viewpoint center. As shown in Fig.1 (a), without loss of generality, C is supposed as the center of the main viewpoint, the area I is the main viewpoint area ($0°\sim Z_1$), and the area II ($Z_1\sim Z_2$) and area III ($Z_2\sim 180°$) are non-main viewpoint areas. We project them onto the area I (the radius is $0\sim\rho_1$), area II (the radius is $\rho_1\sim\rho_2$) and area III (the radius is $\rho_3\sim 0$) on the two-dimensional square plane in Fig.1 (b), respectively.

Our projection method has the following characteristics:

- The range of main viewpoint area is variable. Z_1, Z_2 and ρ_1 can be customized according to the application scenario, network conditions, or other factors.
- The main viewpoint area uses equal area projection (presented in Section 2.1), which ensures that the sampling density of the main viewpoint is high and uniform.
- The non-main viewpoint area uses a projection method with decreasing sampling density (presented in Section 2.2 and Section 2.3) to saving bit rate and playing complexity.

2.1 Equal area projection in main viewpoint area

Perspective projection and equal area projection are common

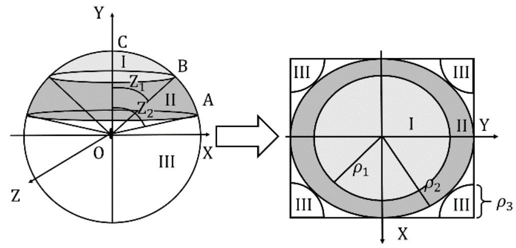


Fig.1 The illustration of projecting sphere video to ASC format

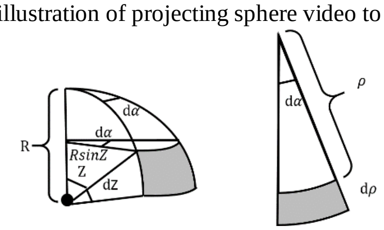


Fig.2 (a) A small area on a three-dimensional sphere; (b) Corresponding area on a two-dimensional circle

projection methods in Surveying and Mapping. TSP uses perspective projection in main viewpoint area, which causes a low sampling density in the middle of main viewpoint area and high sampling density in the around of main viewpoint area. In order to overcome the problem of sampling density difference in main viewpoint area, we use equal area projection in main viewpoint area.

Equal area projection is a type of map projection, the ratio between the area of a graph on the sphere and its corresponding area on the two-dimensional plane is a constant. The gray area in Fig.2 (a) shows a small area on the sphere, and its area size can be expressed as:

$$S_{sphere} = R(sinZ)d\alpha \times RdZ \quad (1)$$

where Z represents the angle between the area and the center of the main viewpoint and R represents the radius of the sphere. The gray area in Fig.2 (b) shows the corresponding area (of the grey area on sphere) on the two-dimensional circle, and its area size can be expressed as:

$$S_{circle} = \rho d\alpha \times d\rho \quad (2)$$

where ρ represents the distance between the area and the center of the circle. In equal area projection, the ratio between to these two areas is a constant, which can be expressed as:

$$\frac{S_{circle}}{S_{sphere}} = \frac{\rho d\alpha \times d\rho}{R(sinZ)d\alpha \times RdZ} = \frac{\rho d\rho}{R^2(sinZ)dZ} = C_0 \quad (3)$$

We solve the differential equation to get the relationship between ρ and Z:

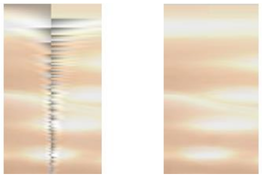
$$\rho = Csin\frac{Z}{2} \text{ (where } C = 2R\sqrt{C_0}\text{)} \quad (4)$$

C_0 is solved by the following boundary condition:

$$\rho_1 = Csin\frac{Z_1}{2} \quad (5)$$

2.2 Down-sampling in non-main viewpoint areas

In the non-main viewpoint areas, we let the ratio between the areas on two-dimensional circle and sphere is a decreasing



(a) (b)

Fig.3 (a) The seam on the boundary between area II and area III after inverse projection; (b) The seam on the boundary between area II and area III is eliminated after improvement

function of the variable Z, which ensures that the further from the center of the main viewpoint, the lower the sampling density.

Among the many descending functions, we choose $cos(Z/2)$, its value gradually decreases and the decreasing speed gradually increases in the range of 0 to π. So the ratio of the areas between the sphere and the circle can be expressed as:

$$\frac{S_{circle}}{S_{sphere}} = \frac{\rho d\rho}{R^2(sinZ)dZ} = C_0 cos\frac{Z}{2} \quad (6)$$

We solve the differential equation to get the relationship between ρ and Z:

$$\rho = \sqrt{C\left(cos\frac{Z}{2}\right)^3 + C'} \text{ (where } C = -\frac{8}{3}R^2C_0\text{)} \quad (7)$$

The coefficients C_0 and C' of the II region and the III region are different. For the II region, the coefficients C_0 and C' are solved by the boundary condition:

$$\rho_1 = \sqrt{C\left(cos\frac{Z_1}{2}\right)^3 + C'}$$
$$\rho_2 = \sqrt{C\left(cos\frac{Z_2}{2}\right)^3 + C'} \quad (8)$$

For the III region, the coefficients C_0 and C' are solved by the boundary condition:

$$\rho_3 = \sqrt{C\left(cos\frac{Z_2}{2}\right)^3 + C'}$$
$$0 = \sqrt{C\left(cos\frac{180°}{2}\right)^3 + C'} \quad (9)$$

2.3 Padding area

In the ASC, some areas in the two-dimensional square plane are unused (areas other than the area I, II, and III in Fig. 1(b)), and we refer to these areas as padding areas. The padding areas are filled with pixel value of 128. Due to interpolation in inverse projection (project ASC back to ERP) or rendering, the pixels at the edges of areas II and III are affected by the pixel values in the padding areas(As shown in Fig.3 (a)).

To address the above issue, a few lines of pixels (In the experiment, we used 5 lines of pixels) outside the edge of areas II and III are projected according to the projection method of the II or III areas. In addition, these lines of pixels can also avoid the impact of filtering in encoding and decoding on the area II and area III. The Fig.3 (b) shows the effect after the above improvement.

Fig.4 Sample frames of the test sequence set

3. EXPERIMENTS

We evaluate the coding efficiency of the proposed projection method on the test sequence set provided by LETIN VR [8], which have been adopted as the test sequences of IEEE 1857.9-Immersive Video Content Coding [6]. The test sequence set contains 6 omnidirectional sequences in equirectangular (ERP) format with 4096x2048 resolution, and the duration and frame rate of each video is 10 seconds and 30 frames per second, respectively. Fig.4 shows some sample frames of the test sequence set.

3.1 Windowed S-PSNR

Recently, S-PSNR[9] has been widely used to evaluate the distortion of panoramic video. The distortion measure is calculated by the sampling points distributed uniformly on a sphere of panoramic videos. S-PSNR can calculate the distortion of panoramic video and the distortion of videos in different projection formats.

In [7], Auwera et al proposed a square windowed (the window is defined by latitude and longitude) S-PSNR to measure the distortion of main viewport area. Windowed S-PSNR defined a window around the center of main viewport, windowed S-PSNR is calculated by using the sampling points within the window.

In our experiments, we use both square window and circular window to calculate the windowed S-PSNR. Circular window is defined by the angle between the point on the sphere and the center of the main viewpoint. For example, when the size of circular window is 90°, windowed S-PSNR is calculated using the sampling points on the sphere that the angle between the point and the center of the main viewpoint is less than 45° (The angles of the two directions add up to 90°).

3.2 Test conditions

In our experiments, the BD-rates are computed by using ERP as the anchor, TSP and ASC projections are tested.

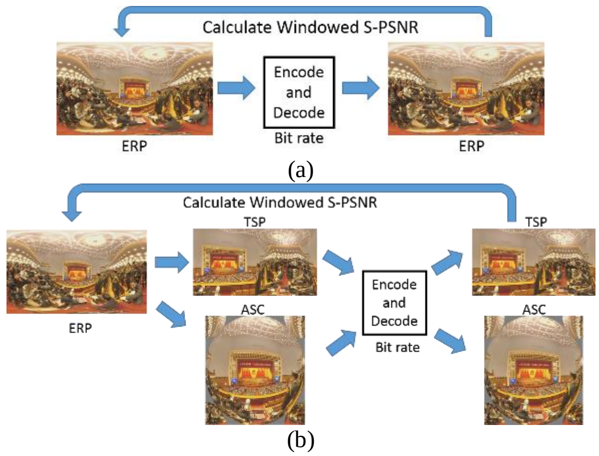


Fig.5 (a) Test process of the anchor; (b) Test process of the tested methods

The parameters in the experiment are set as follows. The center of the main viewpoint is the point where both the longitude and latitude are zero. The resolution of TSP is 2048 ×1024 (same as [7]). The resolution of ASC is 1536×1536, and Z_1, Z_2 and ρ_1 are set as 60°, 90° and 665, respectively.

The test process of anchor is set as follows. HM16.0 (default configuration of RA) was used to encode and decode the sequence of the ERP format, and then we compute the windowed S-PSNR between decoded sequence and original sequence. We re-projected the sequences in ERP format into ASC and TSP formats, respectively. And HM16.0(default configuration of RA) was used to encode and decode the sequences in ASC and TSP format, then we calculated the windowed S-PSNR between decoded sequence and original sequence. The test process is illustrated in Fig.5.

3.3 Experimental results

In our experiments, we use both square window and circular window to calculate windowed S-PSNR. Tab.1 shows the results of BD-rates in terms of windowed S-PSNR with different window sizes, and Fig.6 shows the corresponding BD-rate curve of Y component.

For ASC, there is a BD-rate gain of 51% in main viewpoint area. In non-main viewpoint area, the gain decreases gradually with the increase of window size.

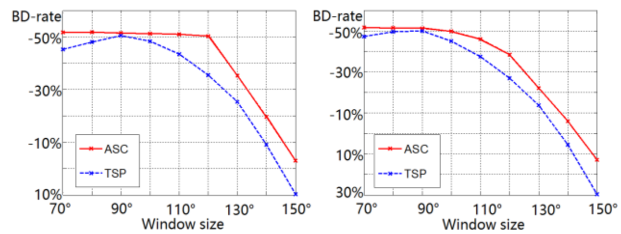
Tab.1 BD-rates for TSP and ASC compared with ERP

(a) Using circular window to calculate Windowed S-PSNR

ERP(anchor)	ASC BD-rate (%)			TSP BD-rate (%)		
Win. Size	Y	U	V	Y	U	V
70°	-51.8	-54.5	-54.2	-45.3	-50.4	-50.4
80°	-51.7	-54.2	-53.8	-48.0	-52.5	-52.4
90°	-51.5	-53.7	-53.5	-50.4	-54.6	-54.6
100°	-51.3	-53.2	-53.0	-48.4	-53.3	-53.5
110°	-51.0	-52.6	-52.4	-43.4	-49.5	-49.8
120°	-50.3	-51.6	-51.4	-35.4	-44.2	-44.7
130°	-35.2	-42.6	-43.0	-25.3	-37.9	-39.3
140°	-19.6	-33.3	-35.6	-9.0	-30.3	-33.8
150°	-3.0	-24.4	-28.3	9.7	-21.6	-26.8

(b) Using square window to calculate Windowed S-PSNR

ERP(anchor)	ASC BD-rate (%)			TSP BD-rate (%)		
Win. Size	Y	U	V	Y	U	V
70°	-51.7	-54.1	-53.8	-47.4	-52.2	-52.1
80°	-51.5	-53.6	-53.4	-49.6	-54.3	-54.1
90°	-51.2	-53.0	-52.9	-50.1	-54.3	-54.4
100°	-49.8	-51.3	-51.4	-45.0	-49.8	-50.3
110°	-46.0	-48.0	-48.3	-37.4	-44.5	-45.4
120°	-38.4	-41.7	-42.4	-27.0	-37.8	-39.5
130°	-22.1	-32.3	-34.8	-13.7	-30.0	-33.5
140°	-5.9	-23.7	-28.4	5.5	-21.7	-27.7
150°	12.9	-15.9	-22.8	29.7	-13.3	-22.1



(a) (b)

Fig.6 BD-rate comparison between ERP (anchor) and tested methods (a) Using circular window to calculate Windowed S-PSNR; (b) Using square window to calculate Windowed S-PSNR

Our method has a higher BD-rate gain than TSP, when the circular window is used to calculate windowed S-PSNR, the BD-rate gain of ASC is 1%~15% over TSP, and when the square window is used to calculate windowed S-PSNR, the BD-rate gain of ASC is 1%~11% over TSP. What's more, the main viewpoint area range of ASC can be customized while that of TSP is fixed.

4. CONCLUSION

In this paper, we proposed a new asymmetric projection scheme, called ASC, which is used for dynamic VR video stream switching. ASC uses equal area projection in the main viewpoint area, which ensures that the sampling density in the main viewpoint area is high and uniform. ASC uses a projection method with decreasing sampling density in non-main viewpoint areas and the further away from the main viewpoint area, the lower the sampling density. Experimental results show that ASC can get a BD-rate gain of 51% in main viewpoint area, and ASC is superior to state-of-the-art method of TSP in terms of coding efficiency and flexibility.

ACKNOWLEDGMENTS

Thanks to National Natural Science Foundation of China 6167206, 61370115, China 863 project of 2015AA015905, Shenzhen Peacock Plan, Shenzhen Research Projects of JCYJ20160506172227337 and JCYJ20150331100658943, and Guangdong Province Projects of 2014B010117007 and 2014B090910001 for funding.

REFERENCES

[1] Denis Zorin and Alan H Barr, “Correction of geometric perceptual distortions in pictures,” in Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM, 1995, pp. 257–264.

[2] Aljoscha Smolic and David McCutchen, “3DAV exploration of video-based rendering technology in MPEG,” Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 3, pp. 348–356, 2004.

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

[4] Matt Yu, Haricharan Lakshman, and Bernd Girod, “Content adaptive representations of omnidirectional videos for cinematic virtual reality,” in Proceedings of the 3rd International Workshop on Immersive Media Experiences. ACM, 2015, pp. 1–6.

[5] Jisheng Li, Ziyu Wen, Sihan Li, Yikai Zhao, Bichuan Guo, Jiangtao Wen, “Novel tile segmentation scheme for omnidirectional video,” IEEE International Conference on Image Processing. IEEE, 2016, pp. 370-374.

[6] http://www.ieee1857.org/1857.9.asp

[7] Geert Van der Auwera, Muhammed Coban, Hendry, Marta Karczewicz, “AHG8: Truncated Square Pyramid Projection (TSP) For 360 Video,” 4th JVET Meeting, Doc. JVET-C0021, Chengdu, China, October 2016.

[8] http://www.air360china.com

[9] Matt Yu, Haricharan Lakshman, and Bernd Girod, “A framework to evaluate omnidirectional video coding schemes,” in Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on. IEEE, 2015, pp. 31–36.