Asymmetric Representation for 3D Panoramic Video

Guisen Xu, Yueming Wang, Zhenyu Wang, and Ronggang Wang *

Shenzhen Graduate School, Peking University
*Corresponding author: rgwang@pkusz.edu.cn

Abstract. The 3D panoramic video provides an immersive stereo visual experience by presenting 3D omnidirectional videos of real-world scenes. The key challenge is to develop an efficient representation of 3D panoramic video in order to maximize coding efficiency. In this paper, we propose an asymmetric representation based on the binocular suppression theory that in a stereo sequence, where the sharpness of left-eye and right-eye view differ, the perceived binocular quality of a stereoscopic sequence is rated close to the sharper view. According to the theory, in our method, one view is down-sampled to half size of original sequence in horizontal and vertical direction, and the other keeps the original resolution. To improve the quality of the reconstructed 3D panoramic video, we propose an anti-aliasing method based on detail blurring. Experimental results show that our representation can get significant coding gains (26.13%) over side-by-side and top-and-bottom method with full resolution.

Keywords: Virtual reality; 3D panoramic image; Asymmetric representation.

1 Introduction

Virtual reality (VR) refers to creating an artificial environment with immersive 3D visual experience. Generally, VR with head-mounted displays (HMDs) is associated with gaming applications and computer-generated content. However, based on the ability to display wide-field-of-view content at high resolution, and track user head motion and update the displayed content with low latency, HMDs can be used to provide immersive visual experiences involving real-world scenes. Such applications are called Cinematic VR. In order to provide a fully immersive experience, a real-world environment has to be captured in all directions (e.g., with a camera rig) resulting in an omnidirectional video corresponding to a viewing sphere. To achieve 3D effects, a separate view is generated and presented to each eye, which leads to stereoscopic omnidirectional video with the corresponding parallax between the two views. Generally, the 3D panoramic video consists of two storage formats: side-by-side and top-and-bottom. And it's common to store 3D video in equirectangular format.

Due to the parallax between two views, 3D panoramic video means doubling the amount of raw data compared to 2D video. Therefore, transmitting 3D panoramic video consumes a lot of bandwidth.

In order to save the bandwidth for transmitting 3D video, we represent 3D panoramic video in mixed resolution method (MR) instead of full resolution method (FR) by down-sampling right-eye frames vertically and horizontally to half. Furthermore, we applied anti-aliasing method based on detail blurring as pre-process to the down-sampled right-eye frames, which can further decrease the bit rate without increasing the computational complexity of rendering. Finally, we evaluate the coding efficiency and the perceived binocular quality of our representation.

The contributions of this paper are as follows.

- We propose an asymmetric representation to improve the coding performance of 3D panoramic video.
- In order to improve the quality of reconstructed 3D panoramic video and further decrease bit rate, we propose an anti-aliasing method based on detail blurring.

2 Related Work

In this section, we briefly review the relevant researches on 3D panoramic video representation. Since 3D panoramic videos need to be played in real-time, those complex methods, such as Depth-Image-Based Rendering (DIBR) [1], aren't suitable for 3D panoramic video. In order to find a more efficient representation, some researchers try to exploit the spatio-temporal correlation of the left-eye and right-eye frames of a given 3D panoramic video, as is done in the Multiview Profile of MPEG-2 [2, 3]. The concept of mixed-resolution coding (MR), which was first introduced by Perkins M.G. in 1992 [4], indicates that mixed-resolution stereo image sequences can provide acceptable image quality, although the latter comments were based on informal observations. In 2000, Stelmach, L. *et al.* explored the response of the human visual system to mixed-resolution stereo video sequences, and concluded that spatial filtering of one channel of a stereo video sequence may be an effective means of reducing storage and transmission bandwidth [5]. Therefore, in a stereo sequence, where the sharpness of left-eye and right-eye view differ, the perceived binocular quality of a stereoscopic sequence was rated close to the sharper view.

According to the mixed-resolution theory, we down-sampled right-eye view vertically and horizontally at half. In order to perform anti-aliasing in down-sampled right-eye view, we proposed an anti-aliasing method based on detail blurring, which used the structure extraction method proposed by Xu L. *et al*. [6]. First, we divided the down-sampled frame into structure image and detail image. And then we used the Gaussian filter to blur the detail images for anti-aliasing. Finally, left-eye and right-eye view are spliced to a complete 3D panoramic image.

3 Approach

According to the principle of mixed-resolution coding, a 3D panoramic video, in which one view has a reduced resolution (mixed-resolution representation, MR), yields the same subjective quality in comparison to the full-resolution representation (FR). Thus, a lower bit rate at equal quality is achieved for MR. Therefore, we propose an asymmetric representation (mixed resolution with anti-aliasing, MRAA) to improve the coding performance of 3D panoramic video. The framework of our method is shown in **Fig. 1**. First, we keep the left-eye view in full resolution and down-sample the right-eye view by the factor of two in horizontal and in vertical direction. When rendering the right-eye view, the frames are up-sampled to the original resolution again. Unfortunately, up-sampling leads to alias in right-eye view, which influences the perceived binocular quality of a 3D panoramic sequence. In order to solve this problem, we apply structure extraction technology and Gaussian filter to the right-eye view to perform anti-aliasing. Finally, we splice left-eye view and right-eye view together.

In this paper, we down-sampled the right-eye view. In fact, both left-eye and right-eye view can be used as down-sampled view.

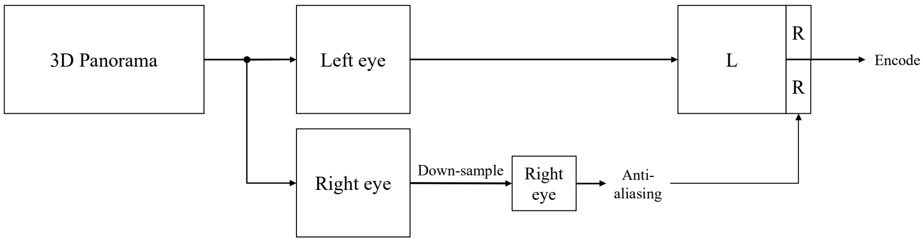

Fig. 1. Framework of asymmetric splicing method for 3D panoramic image

3.1 Anti-alias Method Based on Detail Blurring

When rendering the right-eye view, recovering down-sampled view to original resolution results in aliasing. After analysis, the alias is obvious in Y component of right-eye and almost impossible to be observed in U and V component. Therefore, we mainly solve the problems of aliasing in Y component. Super-resolution image reconstruction is an effective anti-aliasing technology. However, most super-resolution methods are too time-consuming to be employed in 3D panoramic video rendering. In order to render video in real-time, we propose the anti-aliasing method based on detail blurring and applied our method as a pre-process step, which doesn't increase the computation complexity of rendering and saves more bit rate.

Anti-aliasing method Based on Gaussian Filter. Blurring images with Gaussian filter is a simple anti-aliasing method. However, Gaussian filter only considers the spatial correlation between pixels, and does not consider the similarity between pixels. Therefore, the image blurred by Gaussian filter will lose the boundary information. To solve this problem, we propose the anti-aliasing method based on detail blurring.

Anti-aliasing method Based on Detail Blurring. Inspired by the method of structure extraction from texture, proposed by Li Xu *et al.* [6], we utilize structure extraction method to divide down-sampled image into base image and detail image. After blurring detail image by Gaussian filter, we add base image and detail image to a complete right-eye image. Because detail image contains details of down-sampled image, blurring detail image will also blur alias. The framework of anti-aliasing based on detail blurring is shown as **Fig. 2**.

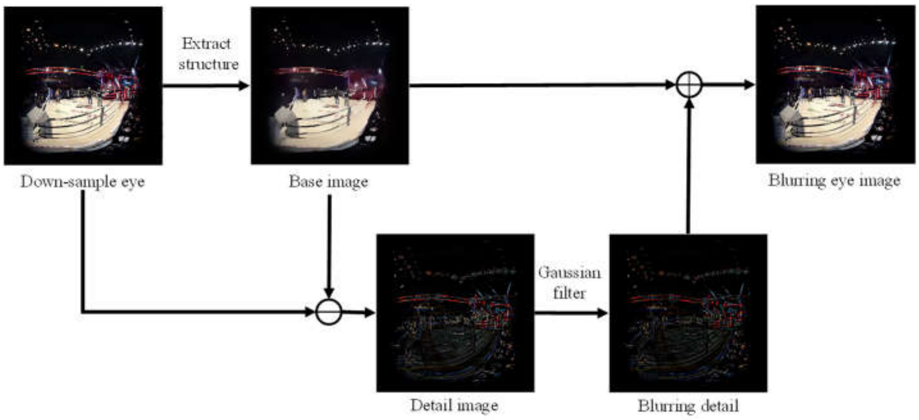

Fig. 2. Framework of anti-alias based detail blurring

3.2 The Method of Splicing Left-eye and Right-eye View

After down-sampling and anti-aliasing, we splice left-eye and right eye view together. Our method can be applied to side-by-side and top-and-bottom 3D panoramic videos. The method of splicing is shown in **Fig. 3**. Supposed A and B are the view of two eyes. The splicing method for the side-by-side 3D panoramic image is shown in **Fig. 3** (a) and (b). First, we divide B into B1 and B2. After down-sampling, B1′ and B2′ are spliced in the vertical direction. And then A and B are spliced in the horizontal direction. The splicing method for the top-and-bottom 3D panoramic image is shown in **Fig. 3** (c) and (d). First, we divide B into B1 and B2. After down-sampling, B1′ and B2′ are spliced in the horizontal direction. And then A and B are spliced in the vertical direction.

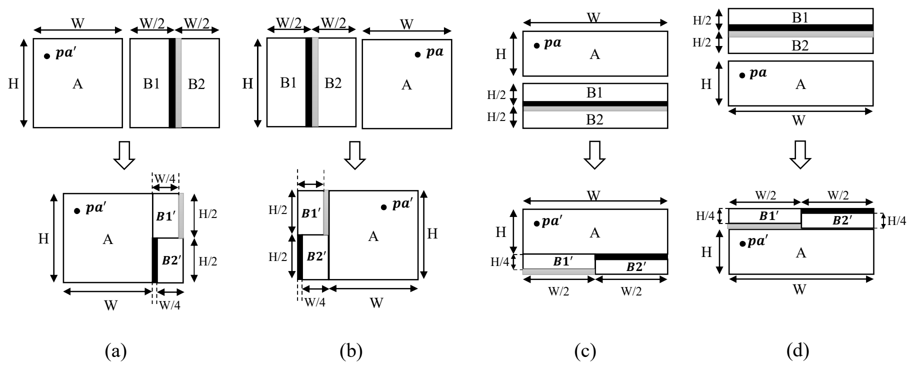

Fig. 3. (a), (b) Splicing method for side-by-side. (c), (d) Splicing method for top-and-bottom

3.3 Padding

According to splicing method mentioned in Section 3.2, the down-sampled view B isn't a continuous image after splicing. Due to filtering in encoding and decoding, the pixels at the right edge of B1′ and the left edge of B2′ are affected. Taking **Fig. 3** (a) as an example, the right-edge of B1 and the left-edge of B2 should be continuous. However, the splicing method leads to that the right edge of B1′ and the left edge of B2′ isn't continuous, which causes seam in right-eye view after encoding and decoding. In order to solve the problem, we pad 8-line pixels in the right edge of B1′ to the left edge of B2′; pad 8-line pixels in the left edge of B2′ to the right edge of B1′. Padding can avoid the impact of filtering in encoding and decoding, as shown in **Fig. 4**.

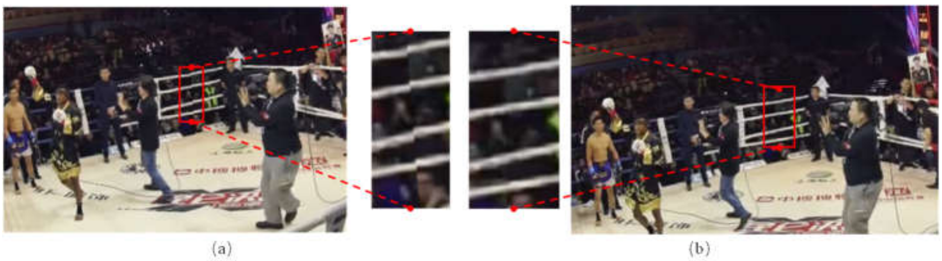

Fig. 4. (a) Before padding, the seam in the middle of right-eye view; (b) After padding, the seam in the middle of right-eye view is eliminated

4 Experiments and Analysis

We evaluated the coding efficiency of the proposed representation on test sequence set. The test set contains two 180°side-by-side 3D panoramic sequences in the equirectangular format at 4096×2048 resolution, the duration of each sequence is 10 seconds and the frame rate of each sequence is 30 fps. **Fig. 5** shows some sample frames of the test set. Finally, a subjective test is carried out to evaluate the perceived binocular quality of 3D panoramic sequences.

(a) (b)

Fig. 5. Sample frames of the test sequence set. (a) Frame of Kunlunjue_1. (b) Frame of Wuxiubo_1

4.1 Experiment for Evaluating the Coding Efficiency

In our experiments, we compared the bit rate of our method (MRAA) with that of MR using FR as the anchor.

The standard deviation of Gaussian filter was set to 0.8. Parameters of structure extraction method [6] were set to follows: $\lambda = 0.01$, $\delta = 3$, $\text{sharpness} = 0.03$, $\text{maxIter} = 4$. λ controls the degree of smooth, increasing λ causes more blurriness. To make the image sharper, λ was set as a very small value (0.01). The factors of down-sampling in horizontal and in vertical direction were both 2. To improve the fairness of comparison, we utilized HM16.0 to encode anchor sequences and testing sequences. The result is shown in **Table 1**.

Table 1. The saving bit rate of two methods compared with full-resolution sequences

FR (anchor)	QP	Save bit rate (%)	
		MR	MRAA
Kunlunjue_1	22	21.70	24.13
	27	21.75	24.36
	32	22.75	25.30
	37	24.27	26.57
Wuxiubo_1	22	26.81	30.30
	27	25.95	30.21
	32	25.46	30.81
	37	26.45	31.29
Average	--	24.39	27.87

According to **Table 1**, both MR and MRAA save bit rate. For MR method, the average proportion of saving bit rate is 24.39%. For MRAA method, the average proportion is 27.87%, which saves more bit rate (3.48%) than MR. Because the anti-aliasing pre-process utilizes the structure extraction and Gaussian filter, which will blur the image. Therefore, it can save bit rate when encoding.

4.2 Subjective Test

Under the same bit rate, the sampled frames of FR, MR and MRAA are shown in **Fig. 6**. The results are captured from VR player in the mobile phone. Due to all left-eye views are full-resolution in the same sequence, we only show the right-eye views. To show more detail of right-eye, we captured the main region of right-eye image to show the results. According to **Fig. 6** (c) and (d), our anti-aliasing method based on detail blurring can slightly improve the quality of down-sampled view.

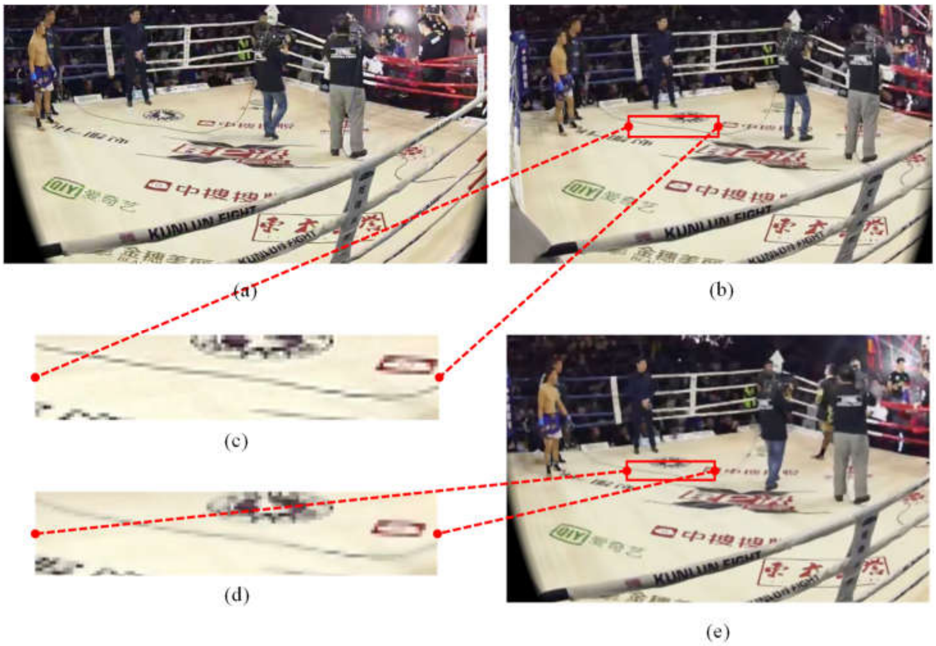

Fig. 6. (a) Right-eye full resolution image. (b) Right-eye mixed resolution image. (c) Alias in (b). (d) Alias in (e). (e) Right-eye mixed resolution image with anti-aliasing.

To assess the binocular quality, we carried out the subjective tests. Because MRAA can save more bit rate than MR, we compare the perceived binocular quality between MRAA and FR. The decoded 3D panoramic sequences with the following total bit rates were used: Kunlunjue: 1.5M/s, 2M/s, 3.5M/s. In our experiment, Samsung Gear VR headset was used to display test sequences. The results of these tests are shown in **Table 2**. It can be seen that MRAA has a slightly better binocular quality than FR. Comparing the results of different bit rate, we found that the smaller the bit rate is, the more obvious the difference between MRAA and FR.

Table 2. Results of subjective tests with coded sequences (sum of votes)

bit rate	MRAA better	No difference	FR better
1.5M/s	5	3	2
2M/s	2	5	3
3.5M/s	1	6	3

5 Conclusion

In this paper, we proposed an asymmetric representation of 3D panoramic video based on the mixed resolution theory. In order to solve the problems of aliasing, we proposed an anti-aliasing method based on detail blurring, which is a pre-processing step. Objective and subjective evaluations for 3D panoramic videos with mixed and full resolution have been carried out. The objective evaluation showed that mixed resolution method is an effective representation, and the average proportion of saving bit rate, compared to full resolution, is 26.13%. Due to the anti-aliasing pre-process based on the structure extraction and Gaussian filter, MRAA save more bit rate (3.48%). The subjective evaluation showed that our method (MRAA) has a slightly better binocular quality than FR.

Acknowledge

Thanks to National Natural Science Foundation of China 61672063，61370115，China 863 project of 2015AA015905, Shenzhen Peacock Plan,Shenzhen Research Projects of JCYJ20160506172227337 and GGFW2017041215130858，and Guangdong Province Projects of 2014B010117007 and 2014B090910001 for funding.

References

1. Fehn, C.: Depth-Image-Based Rendering (DIBR), Compression, and Transmission for a Flexible Approach on 3DTV. In: Electronic Imaging 2004, International Society for Optics and Photonics, pp. 93-104. (2004)
2. ISO/IEC 13818-2. Draft Amendment 3, Sept. (1996)
3. Puri, A., Kollarits, B.G., Haskell, B.G.: Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4. In: Signal Processing: Image Communication, 10(1-3), pp. 201–234. (1997)
4. Perkins M.G.: Data compression of stereopairs. IEEE Transactions on communications. 40(4), 684-696. IEEE (1992)
5. Stelmach, L., Tam, W.J., Meegan, D., Vincent, A.: Stereo image quality: effects of mixed spatio-temporal resolution. IEEE Transactions on circuits and systems for video technology. 10(2), 188-193. IEEE (2000)
6. Xu, L., Yan, Q., Jia, J.: Structure extraction from texture via relative total variation. ACM Transactions on Graphics (TOG), 31(6), 139 ACM (2012)