Generating VR Live Videos with Tripod Panoramic Rig

Feng Xu* Tsinghua University
Tianqi Zhao† Tsinghua University
Bicheng Luo‡ Tsinghua University
Qionghai Dai§ Tsinghua University

ABSTRACT

Recent breakthrough in consumer-level virtual reality (VR) devices brings an increasing demand of VR live content. As converting real life content into VR need complex computations, current techniques can not synthesize 360° 3D VR content with high performance, not to mention real time. We propose an end-to-end system that records a scene using a tripod panoramic rig and broadcasts 360° stereo panorama videos in real time. The system performs a panorama stitching technique which pre-compute 3 stitching seam candidates for dynamic seam switching in the live broadcasting. This technique achieves high frame rates (>30fps) with minimum foreground cut-off and temporal jittering artifacts. Stereo vision quality is also better preserved by a proposed weighting-based image alignment scheme. We demonstrate the effectiveness of our approach on a variety of videos delivering live events. And our system has been successfully used in broadcasting live shows to mobile phone users on a professional live broadcasting platform with about 390 million user visits per month.

Keywords: VR live video, 360° scene representation, video stitching, image-based rendering

Index Terms: Computing methodologies—Artificial intelligence—Computer vision—Image and video acquisition; Computing methodologies—Computer graphics—Image manipulation—Image-based rendering

1 INTRODUCTION

In recent years, Virtual reality (VR) has become a hot topic in both academia and industry. With more and more VR equipment developed and produced, high quality VR content is eagerly required. Nowadays, the majority of VR content is rendered by computer graphics techniques, whose quality is still far from real scenes. On the other hand, there exists rich real content in our daily lives. The content bottleneck could be solved if we can transfer real content into VR content. Thus the key problem is how to perform the transfer with high quality and low latency. If this problem can be solved, more applications like VR live broadcasting, VR tourism and VR remote operation could be enabled.

Current techniques for converting real content to be displayed on VR equipment majorly relay on stereo panoramas, where two panoramas give both stereo visions and 360° viewing experiences to users. In industry, Jaunt ONE [7], OZO [16], Surround 360 [4] and Google Jump [1] are proposed to generate stereo panoramas from camera rigs. However, all these techniques are not real-time as complex computations are required to synthesize stereo panoramas from multiple frames recorded by rig cameras. On the other hand, NextVR [15] proposes a live VR broadcasting system which records the scene with two high quality wide angle cameras and streams the two videos to users. As their technique almost does not require any computation to do data processing but directly use the recorded videos, they achieve real-time performance. However, they only give stereo visions in the range of about 180°. Thus it does not provide 360° stereo visions.

Figure 1: Live stereo panorama video streamed to a user using our system (The top-left thumbnail: the stereo views of the user. The bottom-left thumbnail: the tripod panoramic rig and the live scene).

Different from all the aforementioned techniques, we propose an end-to-end system that records a scene and converts it to be displayed on VR equipment with 360° stereo visions in real time. As Fig. 1 and 2 shows, our key idea is to use three stereo pair cameras to directly record high quality stereo video pairs on three main directions, each with 120° apart. In this manner, on the three main directions, high quality stereo views are directly obtained without requiring any calculations. As we use wide angle cameras to record the scene, the three right (or left) eye images contain common regions, so they can be stitched together to get a right (or left) eye panorama. With the obtained stereo panoramas, we render stereo visions for the non-main directions, so the 360° stereo visions are generated. Even though it is still impossible to interact with a scene, our technique opens the door for generating full VR content of the scene as it provides visual information.

There are some technique challenges in building a system as described. First, if we set the stitching seam to be fixed in the common regions, cut-off artifact will emerge when objects pass though the seam (e.g. Samsung Gear 360 [5] and VUZE [19]). On the other hand, if the seam is estimated on every frames, previous techniques [3, 6, 8, 10–12] can not achieve the real-time performance required by VR live broadcasting. And processing every frames independently will cause noticeable temporal jittering artifact in the result. Second, traditional stitching requires to align the input images to minimize their differences in common regions. However, this alignment may individually move the two images of a stereo pair in their corresponding panoramas, which may destroy the stereo vision experience when a user views the pair.

To tackle these challenges, we first propose an image alignment scheme which jointly considers the viewing directions and the stitching errors. To be specific, the positions of the three images in the panorama are first predefined based on their viewing directions. And then the alignment is not performed on the whole images but majorally on the image boundary (regarding to the horizontal direction) region. In this manner, the central parts of the images are not moved

*e-mail: xufeng2003@gmail.com
†e-mail: 94255282@qq.com
‡e-mail: luobc14@mails.tsinghua.edu.cn
§e-mail: qhdai@tsinghua.edu.cn

2018 IEEE Conference on Virtual Reality and 3D User Interfaces
18-22 March, Reutlingen, Germany
978-1-5386-3365-6/18/$31.00 ©2018 IEEE

(a) (b)

Figure 2: The tripod panoramic rig of our system. (a) The top view shows our triangle tripod. (b) The front view shows a pair of stereo cameras on one side of the tripod.

and thus the stereo vision is kept, while the boundary parts, which are not that important to the stereo vision, are moved to minimize the misalignment. Then for the stitching, we propose a seam switching method to tradeoff between the cut-off and the temporal jittering artifacts. We first pre-calculate three best seams in three partitions of the common region. Then we switch among the three seams to avoid objects passing though the seam. Thus the cut-off can be reduced. Also, the scheme will try to minimize the frequency for switching, which will benefit to avoid temporal jittering.

2 METHOD

In this section, we first develop a hardware, a tripod panoramic rig, to record 360° data for our purpose (Sec. 2.1). Then we propose our algorithm for converting the recorded image sequences to stereo panorama videos, which includes two steps: template generation (Sec. 2.2) and real-time panorama generation (Sec. 2.3). The former outputs a panoramic stitching template with three groups of seams in the three common regions for each panorama. The later switches in the candidate seams to stitch the stereo panorama video in real time.

2.1 Tripod Panoramic Rig

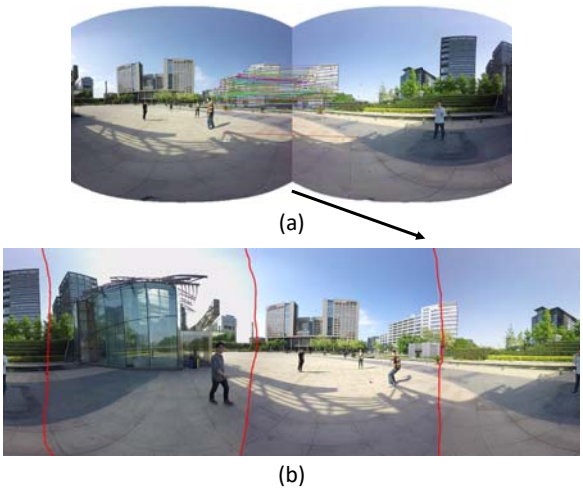
Our tripod panoramic rig consists of a regular triangle tripod on which three stereo pair cameras are attached. The stereo pair cameras on each side of the tripod are a pair of wide-angle cameras. All of the 6 cameras are connected to a controller PC with HDMI capture cards installed. Fig. 2 shows an example of our capture device.

2.2 Template Generation

We first set up our rig in the scene. Then, 6 images of the 6 cameras are captured to compute the stitching template for this scene.

Pre-process Based on our system setup, the left eye panorama is stitched by images of the three cameras placed on the left side of each pair. The template generation scheme for either the two eyes are exactly the same. Thus, we only discuss the algorithm for one eye. Before the stitching, we first pre-process the images by fish eye image undistortion [13] by warping images to keep straight lines and color correction [2] to linearly transform the color of one image to be coherent with that of another image.

Image Alignment Firstly, the positions of the three images in the panorama are predefined based on their viewing directions. As our system is delicately fabricated and the cameras are pre-calibrated. The center of the three images are uniform distributed on the horizontal (y) direction in the panorama. Then we perform SIFT [14] feature extraction and matching between each pair of the three images (Fig. 3(a) shows an example.). The alignment is performed by MLS method [17] where the matched features move to the same image coordinates while other pixels around the matched features move as rigid as possible. The target positions of the features are carefully designed to be the weighted combinations of the matched two positions. The weight is determined by the distance from the position to the corresponding center position of the image on the y direction. The larger the distance is, the smaller the weight is. In this case, the center region of an image will keep its original position in the panorama, as this part is more important to the stereo vision and the movement in this region may destroy the stereo vision. The boundary on the y direction will moves to minimize the misalignment.



(a) (b)

Figure 3: Image alignment and panorama stitching. (a) Feature matching between two neighboring frames. (b) The stitched panorama using the middle seams (red line) generated by [9].

Our image alignment scheme keeps the stereo vision better than free aligning schemes. We know that the three stereo pairs record ideal stereo visions because the camera distance is set close to the pupil distance of human. In our alignment method, for the recorded images, we determine their positions in the panorama based on the viewing directions of the stereo pairs, and the central parts of the images are not moved in the aligning. Thus ideal stereo visions in these directions can be guaranteed as shown by the red curve in Fig. 4. However, if the images are freely moved in the panorama, the stereo vision can not be kept in the directions of the stereo cameras, leading to the blue curve in Fig. 4.

Seam Calculation Then, the overlapped three images are fed into a stitching algorithm [9]which compute the stitching seams by minimizing the color difference crossing the seams. Different from [9], we segment the overlapped region of two images into three subregions with the same image width, and find the optimal seam in each sub-region. Fig. 3(b) shows the stitched panorama according to the middle stitching seams. while Fig. 5(a, b) show the panoramas stitched with the three left and right seams, respectively.

2.3 Real-Time Panorama Generation

The technique in this subsection is to switch among the three seams when foreground objects emerge in the scene during the live broadcasting. First, by comparing the color differences between an input image and the pre-recorded background image [18], we find the foreground object and locate its center in the panoramas, as shown in Fig. 6. Then, we compute three distances between the foreground center and the three seams. When the distance to the current seam is smaller than a threshold d (20 pixels in all our experiments), we switch to use the farthest seam. Fig. 7 shows one example of seam switching. In this case, the green seam is the current seam, and the character on the left image is shown in the panorama. However, the character is about to cross the seam and the cut-off artifact will be generated. At this time, our method switches to the blue seam to avoid the artifact. And the character on the right appears in the panorama.

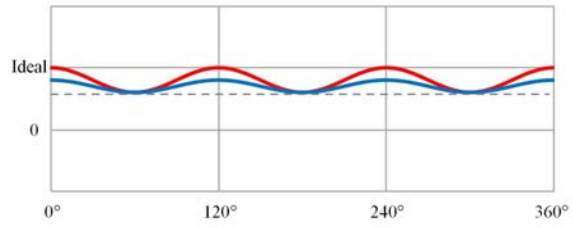


Figure 4: Stereo vision quality changes with the viewing directions. The red curve indicates the stereo vision quality achieved by our method while the blue line illustrates the possible situation when the image positions in the panorama change freely in the alignment. Here the ideal stereo vision quality is achieved at the directions of 0° (360°), 120° and 240°, because the stereo cameras in our rig are facing to these directions.

(a)

(b)

Figure 5: Panorama stitching using different seams. (a) The panorama stitched by three left seams. (b) The panorama stitched by three right seams.

By defining a proper d, seam switching only happens when the foreground object is approaching the current seam, thus minimum frequency of switching is guaranteed to avoid temporal jittering as much as possible. When switching does need to be performed, choosing the farthest one has the minimum possibility to generate cut-off artifact and lowest potential to perform switching again in the near future. So we design the switching strategy like this.

3 Results

In this section, we first discuss the system setup and the performance of our experiments. Then we compare our method with the stitching method based on fixed seams. The effectiveness of our system is discussed in the application subsection with more results in the accompanying video.

Setup Our experiments are performed using the structured panoramic rig with 6 wide-angle (185°) *GoPro Hero 4* cameras. Each camera outputs frames with 1920×1440 resolution in 60fps. Our algorithms runs on a PC with *Intel Core i7-6700* CPU, 16GB memory, and an *Nvidia GTX 960* graphics card. Our system is capable to stream the stereo panorama video live with 2048×2048 resolution in 30fps with 4Mbps network bandwidth.

Performance The pre-computation time for template generation take nearly 20 seconds. In the online stage, our algorithm for generating one stereo panorama takes only 25 milliseconds in average, while the frame rate of broadcasting live videos is limited to 30fps because of the network bandwidth constraints. Thus, the performance of our system satisfies the requirement of real-time live broadcasting.

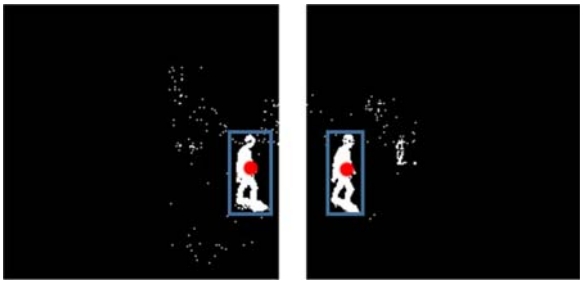


Figure 6: Foreground objects in the common region. The left and right are two images stitched in the panorama. The foreground detection is visualized as white pixels. The foreground center (red point) is localized using the bounding boxes (blue boxes).

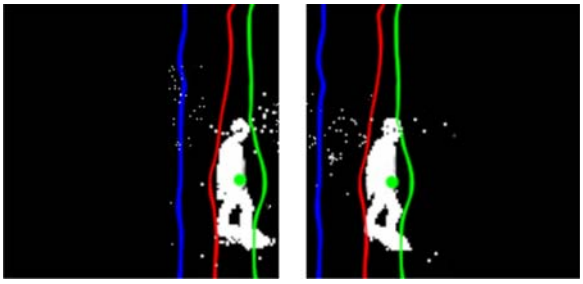


Figure 7: Illustration of seam switching. The left and right are two images stitched in the panorama. The current seam is denoted by green and the seam to switch to is denoted by blue.

3.1 Comparisons

Fig. 8 shows the comparison between fixed seam stitching and our method. For the same input frames, the fixed seam stitching causes apparent cut-off artifact when there exists a foreground object crossing the stitching seam. However, our method provides a much better stitching quality with no foreground cut-off because it detects the foreground and changes the seam in real time.

The first row of Figure 8 provides two moments of an outdoor scene where several people are playing shuttlecock. When there comes a person running through the current seam in the common region, our algorithm detects the foreground and changes the current seam into another one which does not cross the person. The mechanism here makes the foreground person in continuity without cut-off artifacts, while the results of the fixed seam method do have the artifacts.

The left side in the second row of Figure 8 shows an indoor scene where a person is walking around. Similar to the first scene, clear cut-off artifacts exist in the result using the fixed seam. Because of the different depth of the foreground person, the fixed seam solution can not preserve the consistency of the panorama. However, our method pre-computes three seams for stitching candidates switching in real time, which leads to a better result in common regions.

The last scene in Figure 8 shows the panorama of a cross talk performance. The fixed seam method gives ghosting artifacts on the face of a performer, because the stitching happens exactly in the position of his face. Our method eliminates the ghosting artifacts by applying a different seam.

According to the comparison between the fixed seam method and our method, we can see the dynamic seam switching eliminates the stitching artifacts of the panorama video. And our method based

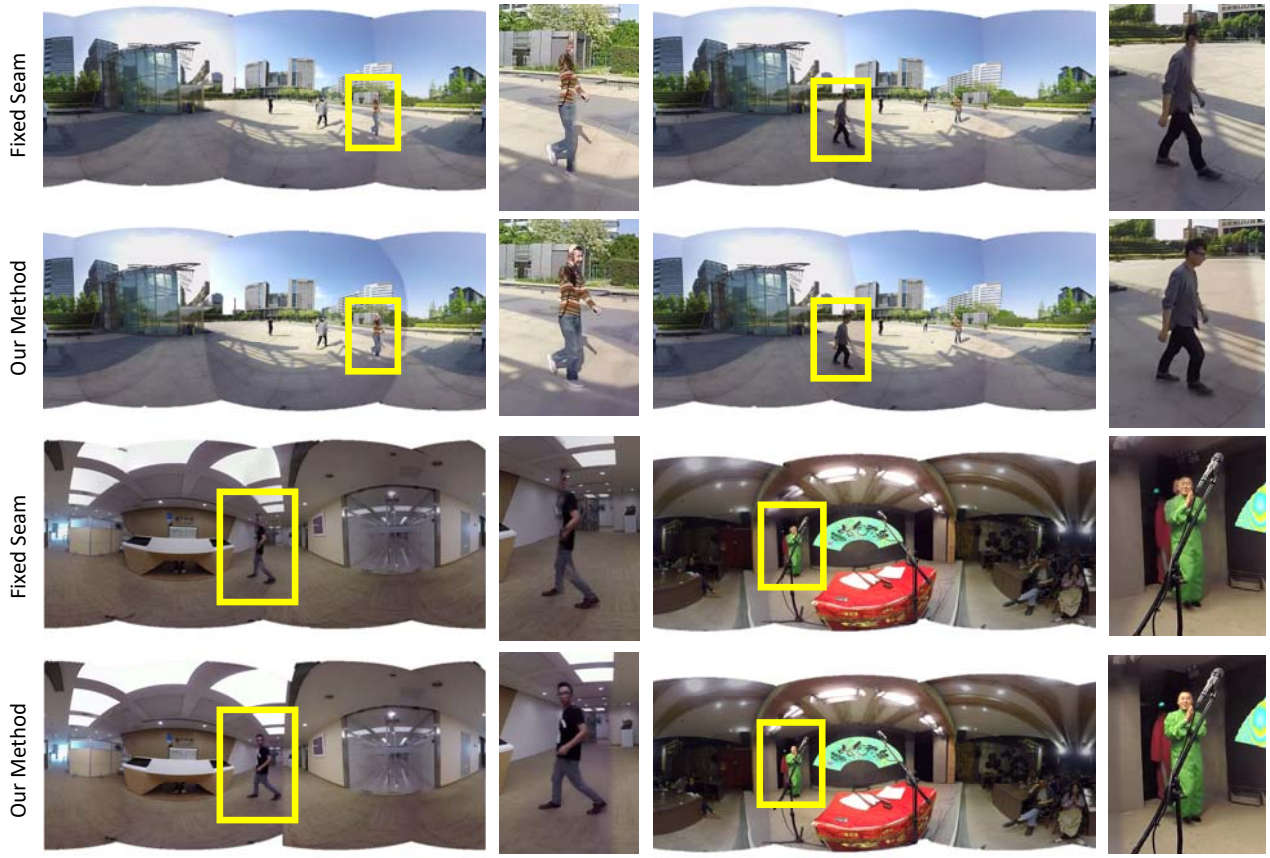


Figure 8: Comparisons between fixed seam stitching and our method. The yellow rectangle shows that a foreground object crosses the seam. The small images on the right side of the panorama show the enlarged details.

on pre-computed mapping templates makes the switching operation runs in real time, which outperforms the current live panorama video delivering systems (e.g. Samsung Gear 360 [5] and VUZE [19]).

User Study Our technique generates a temporal jump in the common region of two images when switching seams, but keep the foreground objects clear and complete. The single seam solution does not suffer the jump but makes the foreground objects ghosted or incomplete. Our observation is that the moving foreground objects usually attracts the attention of the audiences, and thus keeping them visually comfortable makes more sense. To evaluate this, we conduct a user study. We show our 360° video and the video generated with single seam to participants consecutively, and ask the participants to score the quality from 0 (the lowest quality) to 5 (the highest quality). We have 30 participants in total, and each participant watches all the tree experimental clips shown in the accompanying video. The videos of the two different methods are presented to the participants with a random order. As both the scores for our methods and the single seam-based method satisfy normal distributions (Shapiro-Wilk test with $W(30) = 0.90$, $p \leq 0.01$ for our setup and $W(30) = 0.90$, $p \leq 0.01$ for the single seam method), we performed a two-tailed t test on the data. The test showed that the mean results are significantly different ($t(58) = 4.44$, $p = 0.000041$). We find that our method has higher mean score ($M = 3.53$, $SD = 0.60$) than the mean score of the single seam-based method ($M = 2.83$, $SD = 0.62$).

3.2 Applications

Our system has been tested in many scenes to deliver events in VR for commercial use. Besides the cross talk performance shown in Fig. 8, we have also converted several live shows such as talk shows, basketball events, and natural scenarios into stereo panorama videos, and then broadcasted them to mobile phones of end users on a platform with about 390 million user visits per month. The real-time property of our system can be verified in these applications. More results of our method can be seen in the accompanying video. Note that in live broadcasting, we always use one stereo pair to face the interesting region in the scene. As the artifacts of our system always happen in non-interesting regions, our system can satisfy users in most times.

4 Limitations and Future Works

We propose a dynamic seam switching mechanism for stitching panorama videos in real time. However, the generated stereo panorama video will have temporal discontinuities when the stitching seam switches (The frequency of this artifacts is reduced by our switching strategy.). This change will become less noticeable if we implement a more sophistical color correction, e.g. using color checkers, to eliminate the difference among views.

Our system will fail when all the three seams are intersecting with foreground objects. This happens when multiple objects approach the seam from different directions. In this case, we fix the used seam and the system gives the same results as the method with single seam.

In this work, we use three seams because it is a good tradeoff between performance and quality. For other setups with different viewing ranges or resolutions of cameras, more or less number of seams may also be applicable.

In our experiments, we always face one stereo pair to the interesting region of the scene. However, when the scene contains multiple interesting regions, it is possible that some of them can not be faced by stereo pairs. In this case, bad visual results will be generated.

The stereo vision is not perfectly guaranteed in all directions in our system. As we have fixed the offset to be 0 for the central part of images, users will have correct stereo vision when viewing in these three recording directions. It is still an open problem about how to improve the stereo vision for other directions.

Our technique assumes the rig to be fixed in the recording and the common region between cameras to contain background only in the first frame of the recording. These assumptions can not be satisfied always, but for live broadcasting of a show, these assumptions can be usually satisfied.

5 Conclusions

In this paper, we propose a real-time image stitching technique based on pre-computation. The proposed system minimizes the cut-off artifact and the temporal jittering artifact and still keeps real-time performance. A user study indicates that the proposed technique gives better viewing experience to users than the traditional stitching solution. Based on our method, we design an end-to-end system that records a scene using a tripod panoramic rig and broadcasts 360° stereo panorama videos to end users in real time. Our system has been successfully used in broadcasting many live shows on a professional live broadcasting platform with 390 million user visits per month.

Acknowledgments

This work was supported by the NSFC (No.61727808, 61671268). We wish to express our thanks to the reviewers for their insightful comments. We also would like to thank Yuanfa Cai and Chao Jiang for their help in this work. Tianqi Zhao is the corresponding author.

References

[1] R. Anderson, D. Gallup, J. T. Barron, J. Kontkanen, N. Snavely, C. Hernández, S. Agarwal, and S. M. Seitz. Jump: virtual reality video. *ACM Transactions on Graphics (TOG)*, 35(6):198, 2016.

[2] M. Brown, D. G. Lowe, et al. Recognising panoramas. In *ICCV*, vol. 3, p. 1218, 2003.

[3] Y.-S. Chen and Y.-Y. Chuang. Natural image stitching with the global similarity prior. In *European Conference on Computer Vision*, pp. 186–201. Springer, 2016.

[4] Facebook. Facebook360. `https://facebook360.fb.com/facebook-surround-360/`. 2017.

[5] Gear360. Gear360. `http://www.samsung.com/global/galaxy/gear-360/`. 2017.

[6] B. He and S. Yu. Parallax-robust surveillance video stitching. *Sensors*, 16(1):7, 2015.

[7] Jaunt. Jauntvr. `https://www.jauntvr.com/jaunt-one/`. 2017.

[8] W. Jiang and J. Gu. Video stitching with spatial-temporal content-preserving warping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 42–48, 2015.

[9] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick. Graphcut textures: image and video synthesis using graph cuts. In *ACM Transactions on Graphics (ToG)*, vol. 22, pp. 277–286. ACM, 2003.

[10] C.-C. Lin, S. U. Pankanti, K. Natesan Ramamurthy, and A. Y. Aravkin. Adaptive as-natural-as-possible image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1155–1163, 2015.

[11] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *European Conference on Computer Vision*, pp. 370–385. Springer, 2016.

[12] K. Lin, S. Liu, L.-F. Cheong, and B. Zeng. Seamless video stitching from hand-held camera inputs. In *Computer Graphics Forum*, vol. 35, pp. 479–487. Wiley Online Library, 2016.

[13] Y. Ling and Y. CHENG. The designing methods of fish-eye distortion correction using latitude-longitude projection [j]. *Journal of Engineering Graphics*, 6:005, 2010.

[14] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157. Ieee, 1999.

[15] NextVR. Nextvr. `http://www.nextvr.com/`. 2017.

[16] OZO. Ozo. `https://ozo.nokia.com/vr/`. 2017.

[17] S. Schaefer, T. McPhail, and J. Warren. Image deformation using moving least squares. In *ACM transactions on graphics (TOG)*, vol. 25, pp. 533–540. ACM, 2006.

[18] N. Singla. Motion detection based on frame difference method. *International Journal of Information & Computation Technology*, 4(15):1559–1565, 2014.

[19] VUZE. Vuze. `http://vuze.camera/`. 2017.