

A virtual reality video stitching system based on mirror pyramids

Ling Zhu, Wei Wang, Yu Liu, ShiMing Lai, Jing Li

College of Information System and Management

National University of Defense Technology

Changsha, China 410073

Email: orange4working@gmail.com

Abstract—This paper introduces a virtual reality video acquisition method based on mirror pyramids. The system is designed to capture a high-definition stereo video, while the mirror pyramids allow users to shoot objects that are located close to the camera. First, we theoretically study the stereoscopic imaging of the specular reflection device and verify its feasibility. Second, we propose a complete stitching process that includes calibration, mask extraction, and fusion for the special imaging device. The mirror reflection of the 3D imaging method allows users to shoot close objects and achieve an excellent 3D effect.

Keywords-Panoramic stereo imaging, mirror pyramids, video stitching

I. INTRODUCTION

Virtual reality technology based on panoramic photography has attracted wide usage along with the continuous development of computer vision technology and the promotion of virtual reality glasses. For instance, users are presented with virtual scenes when viewing the structure of stadiums and browsing street views through a virtual display network system. Accordingly, how to obtain a virtual reality or panoramic stereo video has recently become a hot research topic.

A virtual reality camera can be constructed using two approaches, namely, a refractive method in which only a refractive element (e.g., a lens) is used, and a refraction method in which a reflective member (e.g., a mirror) is used in combination with a refractive element. [1] presents an overview of panoramic imaging techniques. The refractive system includes a camera cluster, a fisheye lens-based system, and a rotating camera [2]. However, this system has a non-zero parallax angle that prevents the seamless stitching of objects that are located close to the camera, thereby preventing the panoramic camera from shooting close up or similar scenes.

Some authors have attempted to use the surface mirror sensor and a single camera to solve the abovementioned parallax angle problem [3], [4], [5], [6]. However, given its low resolution, the generated single camera cannot meet the requirements of a high-definition video.

A multi-mirror reflective panorama, also known as mirror pyramid panorama camera, can be co-located at a single

point within the pyramid by positioning multiple cameras around the mirrors to locate the pyramids, thereby effectively forming a wide-field virtual camera. However, stereoscopic imaging lacks theoretical research.

In this paper, we analyze the hardware structure of a panoramic 3D camera and deduce the mathematical relationships between the optical center distribution of the camera and the angle of the adjacent camera as well as that between the mirror and its structure.

We also propose a complete video stitching process that includes calibration, mask extraction of effective scene information, and fusion. The motion parameters of adjacent cameras are calibrated through the relative motion parameters to be registered and projected into the latitude and longitude panoramic coordinate systems. The mask operation filters the scene out of the mirror and enhances the stitching effect through fusion. The panoramic camera consists of 13 cameras, of which 12 cameras shoot the horizontal scene and 1 camera shoots the scene above. Six of these cameras create a panorama through a stitching process and then create another panorama along with the adjacent and shared cameras. A 3D effect is achieved by using a virtual reality headset to obtain the left and right scenes from these two panoramas.

The rest of the paper is organized as follows. The second chapter reviews the related literature. The third chapter describes the hardware design of the panoramic stereo camera. The fourth chapter describes the splicing process. The fifth chapter analyzes and presents the experimental results.

II. RELATED WORK

Most of the existing methods for capturing ODS panoramas depend on a quiet scene [5], [7] or only have few sports suitable for video textures [8], thereby allowing the pan[8]rama to be captured by rotation.

Some authors have tried to capture a panoramic stereo video directly. Tanaka and Tachi [9] used rotating prism slices to capture the associated rays, but this process would require a complex setup and could only generate low-quality videos.

Many commercially available panoramic machines apply refractive methods to stitch images seamlessly. However,



Figure 1. The virtual reality video stitching system based on mirror pyramids consists of 13 industrial cameras, where the left and right panoramas are composed of six cameras with a shared camera that can record the objects located close to the camera and capture a high-definition stereo video.

the presence of a non-zero parallax prevents those objects that are positioned close to the camera from being stitched. Although many mainstream video stitching algorithms have been proposed to solve this problem, they can bring additional computational overhead. Therefore, this paper attempts to eliminate or reduce such parallax by designing a reflective mirror pyramid structure. A mirror pyramid panoramic camera has many attractive features, including single-view imaging and high-resolution video capture.

Nalwa et al. [10] designed a four-sided specular reflector, to which [11] added an upright camera to expand its vertical field of view while theoretically discussing the pyramid for eight cases. Shimamura et al. [12] utilized a pair of upper and lower mirror facets to achieve stereoscopic imaging and experimental implementation. The video content for this projection may be captured by two cameras with curved mirrors. Although the vertical baseline is useful for estimating the depth of the stereo, using this tool to generate a video is not conducive.

Schreer, O. et al. [13] designed a six-sided mirror pyramid device that shoots a 3D panoramic video based on the location of the adjacent camera. This design was then verified in [14].

With a pair of panoramic videos as its output (one for the left eye and the other for the right eye), the transmission requires twice the amount of data of a single video. Most of the common editing operations can be easily performed (i.e., color correction and cross fade), and many tools are available for highly complex stereoscopic perception editing [15].

Many companies, such as Facebook and Nokia, have started to produce 360 panoramic 3D cameras. However, these devices are not open to the external stitching method and have inflexible camera designs, thereby making their systems extremely difficult to assess.

III. DESIGN OF AN OMNI-STEREO VIDEO CAMERA

We analyze the hardware structure of the mirror-reflective stereoscopic imaging system, specifically the optical center distribution, the angle of the camera offset, and the relationship between the size and structure of the mirror.

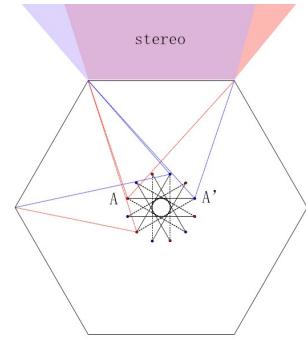


Figure 2. Distribution of the 12 virtual optical centers

A. Optical center

In Figure 2, the 12 points represent 12 virtual optical centers, the red and blue rays represent the range of scenes that are captured by the left and right eyes, respectively, and the scene overlap area is a stereoscopic imaging area. A, A' denotes a pair of left and right eyes for the virtual point of view, while the six prisms of the border represent the six mirrors. A, A', which we set to 70 mm in this paper, simulates the distance of the human eye design. The follow-up design is based on this constraint. Six pairs of left and right eye images are stitched and the real human eye is simulated to observe the world and to obtain a 3D panoramic image.

B. Offset angle

The camera must be offset because its horizontal field angle should cover the real scene reflected by a single mirror. We take a pair of left and right eye virtual views and a mirror for analysis. In Figure 3, A and A' denote a pair of left and right virtual viewpoints, M'N' represents the mirror, h is the distance from the virtual viewpoint to the mirror, r is the distance from the optical center to the mirror, R is the distance of M'N', is the offset angle, and d is the distance of AA'. Given that the virtual view connection is 12 equilateral, the angle OAA' is 15° . The relationship of φ is illustrated as follows:

$$h = |AE| = |CF| \quad (1)$$

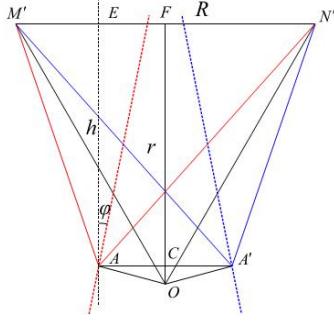


Figure 3. Analysis of the offset angle

$$h = r - \frac{d}{2} \tan \angle OAA' = r - \frac{\sqrt{6} - \sqrt{2}}{8} d \quad (2)$$

$$\varphi = \frac{1}{2} \left(\arctan \frac{R+d}{2h} - \arctan \frac{R-d}{2h} \right) \quad (3)$$

If the distance between the adjacent optical centers is fixed to the human eye pupil distance of 70 mm, then the distance from the virtual optical center to the mirror is 14.8 cm, while the camera offset angle φ is 9.8° . Therefore, the right-hand panorama that has been stitched must be offset by 2 or 19.6° in the opposite direction.

C. Size relationship

We theoretically analyze the size of the mirror frame. In Figure 4, H is the distance from the bottom to the mirror, h is the distance from the virtual center to the mirror, ν is the vertical viewing angle, δ is the loss angle resulting from the blocked view of the camera, θ is the tilt angle of the frame, and a is the diameter of the camera.

$$2h \sin \theta \sin \delta = \frac{a}{2} \sin (\theta - \delta) \quad (4)$$

$$\sin^2 \delta = \frac{\frac{a^2}{8} (1 - \cos 2\theta)}{2h^2 (1 - \cos 2\theta) + ha \sin 2\theta + \frac{a^2}{4}} \quad (5)$$

If the distance between the adjacent optical center is fixed to the human eye pupil distance of 70 mm, then h can be obtained using formula (2). Given that the camera diameter a has a known quantity and that the geometric relations (4) and (5) show that the tilted angle θ is fixed, then the loss angle δ is obtained. The vertical field angle and camera height are computed as follows:

$$\nu = 2 \left(\frac{\pi}{2} - \theta - \delta \right) \quad (6)$$

$$H = r \tan \theta + \frac{h}{\cot(\frac{\nu}{2}) - \cot(\theta)} \quad (7)$$

In Figure 5, H represents the height of the mirror surface device, W represents the maximum cross-sectional area of the device, and H_0 represents the height of the mirror effective area. The overall size of the camera with the frame

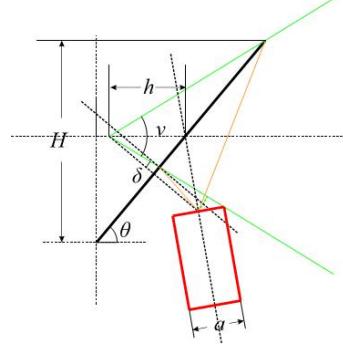


Figure 4. Calculation principle of the relative pose

tilt angle θ increases, while the vertical field of the view is relatively reduced. To balance these two, we set θ to 52.8° , the vertical viewing angle to 60° , the camera effective height to 22.4 cm, and the maximum cross-section to 54.7 cm.

IV. STITCHING

We describe the stitching process of the mirror-reflective stereoscopic imaging system. For special devices, we divide the stitching process into calibration, mask extraction, and integration.

A. Calibration

We find the motion parameters between the cameras by following the traditional way of finding the feature points [16]. Given that the overlapping area is small and that the feature points in the overlapping area are very difficult to detect, we obtain the motion parameters by means of the manual feature points. In the experiment, we use the KOLOR graphical tool to obtain the relative parameters of the camera in the latitude and longitude coordinate systems.

B. Mask operation

The scene outside the mirror will seriously influence the effect of fusion. Given that the camera position is

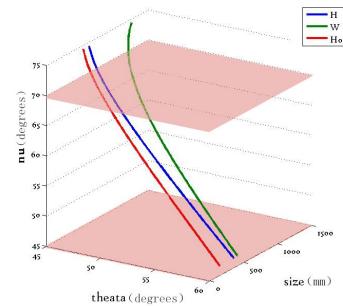


Figure 5. Relationship between the tilt angle θ and the size of the camera

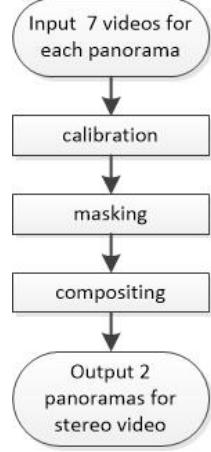


Figure 6. Flow of the stitching process

relatively fixed, we design a mask that allows each camera to capture the scene within the mirror. In the formula (8), p_i denotes the pixel on the original image, while α_i denotes the corresponding pixel in the mask.

Through the mask, the mirror of the scene extracted at the same time filter out the mirror outside the scene. In the experiment, the mask can be manually developed in advance.

$$p_i' = \begin{cases} p_i, & \text{if } \alpha_i = 1 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

In Figure 7, (a) is the original input image, (b) is the mask image, (c) is the effective scene extracted for the mask, and (d) is the mosaic detail before the mask is extracted. The scenes outside the mirror make the image of the mirror residual, thereby greatly improving the mask operation.

C. Compositing

Minor exposure differences and registration errors will create a seam around the stitching. Traditional methods, such as the pyramid fusion method [17], have a high complexity. Given that the effective scene has been extracted through the previous mask operation and that the overlapping area is relatively small, we design the following fusion algorithm:

In the algorithm, the seam finding algorithm can get from opencv. We eventually multiply the weight matrix that has been calculated from the original image. Figure 8(a) shows the obvious patchwork of the overlapping area before the fusion, while Figure 8(b) shows the results after applying the above fusion algorithm. The transition of the stitching process can be smoothed.

V. RESULT

The panoramic camera consists of 13 industrial cameras, where the left and right panoramas are composed of six cameras with a shared camera (Figure 1). The upper and lower

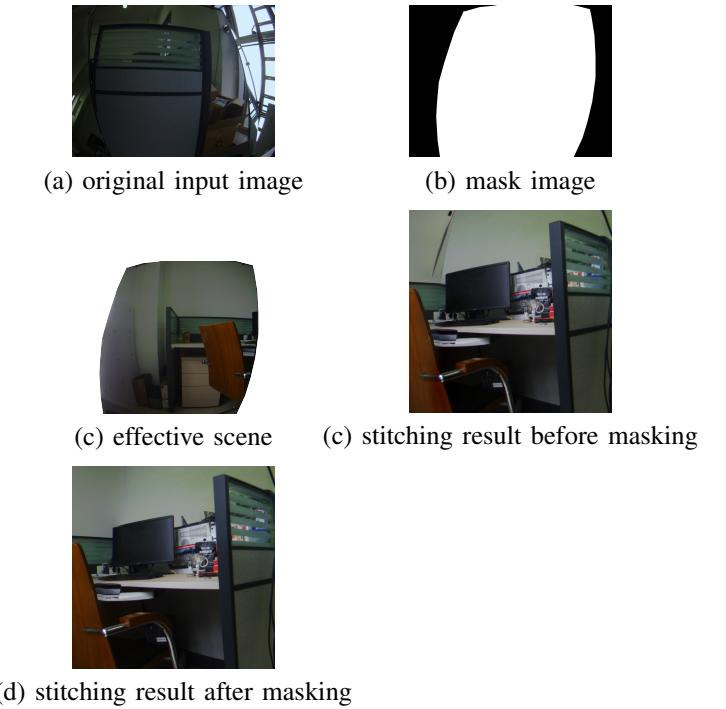


Figure 7. Result of the mask operation

Algorithm 1: Weight Matrix Computation

Input: original picture projected to the panoramic coordinate systems I

Output: the weight matrix W

for Traverse the pixels in the original image I **do**
The closer the suture is, the smaller the weight

$$\text{weight}_i = \frac{\text{blend_width} - \text{norm}(\text{dis}(i, \text{center}))}{\text{blend_width}}$$

end

for Traverse the value in the weight matrix W **do**
Normalize the weight of the overlapping area,
Update W:

$$\text{weight}'_i = \frac{\text{weight}_i}{\text{weight_sum}}$$

end

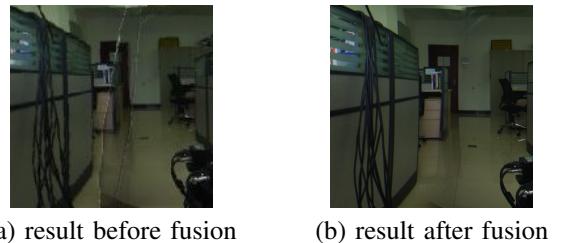


Figure 8. Result of the compositing

parts of Figure 9 show the left and right eye panoramas, respectively. The frame portion is a perspective image that is projected onto the left and right eyes. Even those objects that are positioned close to the camera show an excellent stitching effect. The red and blue maps in Figure 10 denote the left and right eye panoramas, respectively. The parallax becomes larger and smaller when the object is located closer to and farther from the camera, respectively.

The stereoscopic video produced by the specular reflection has a 60 stereoscopic view, but we have no high demand for upward scenes because of their lack of information (i.e., sky). Therefore, we set up a camera and two stereo video pairs to share the scene information and to expand the vertical viewing angle as shown in Figure 11. However, the upward scene has no strong 3D effect. Moreover, given that the optical center of the upward camera is not coincide with other virtual optical centers, some stitching errors may be generated within a close range.

VI. CONCLUSION

This paper presents a virtual reality video acquisition method based on mirror pyramids. Through the mirror reflection of the 3D imaging method, we can shoot close objects and produce an excellent 3D effect.

We contribute to the literature by verifying the theoretical feasibility of the stereoscopic imaging of mirror-reflex devices and by constructing a special device that can achieve a complete stitching process. The hardware is fully equipped with real-time stitching capabilities to reduce the parallax effect and to avoid complex stitching operations. We also add an upward camera to expand the vertical viewing angle of the specular reflection device and to improve its visual effect.

Although the method proposed in this paper is simple and effective, there is still a lot of work to be done in the future. First, because of the reflected light, it can not be guaranteed that the virtual pupil distance corresponding to the scene in each direction equals 70mm, so the stereo effect is not equivalent everywhere. A special projection mode needs to be studied to improve this phenomenon. Secondly, there still remain a parallax effect in the vertical direction. we will try to solve this problem in the future.

REFERENCES

- [1] K. H. Tan, H. Hua, and N. Ahuja, “Multiview panoramic cameras using mirror pyramids,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 26, no. 7, pp. 941–6, 2004.
- [2] R. Anderson, D. Gallup, J. T. Barron, J. Kontkanen, N. Snavely, S. Agarwal, and S. M. Seitz, “Jump: virtual reality video,” *Acm Transactions on Graphics*, vol. 35, no. 6, p. 198, 2016.
- [3] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya, “Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors,” in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, 2002, pp. 485–489 vol.1.
- [4] S. Baker and S. K. Nayar, “A theory of single-viewpoint catadioptric image formation,” *International Journal of Computer Vision*, vol. 35, no. 2, pp. 175–196, 1999.
- [5] S. Peleg, M. Ben-Ezra, and Y. Pritch, “Omnistereo: panoramic stereo imaging,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 23, no. 3, pp. 279–290, 2001.
- [6] S. K. Nayar, “Catadioptric omnidirectional camera,” in *Conference on Computer Vision and Pattern Recognition*, 1997, p. 482.
- [7] C. Richardt, Y. Pritch, H. Zimmer, and A. Sorkine-Hornung, “Megastereo: Constructing high-resolution stereo panoramas,” in *Computer Vision and Pattern Recognition*, 2013, pp. 1256–1263.
- [8] V. Couture, M. S. Langer, and S. Roy, “Panoramic stereo video textures,” in *IEEE International Conference on Computer Vision*, 2011, pp. 1251–1258.
- [9] K. Tanaka and S. Tachi, “Tornado: omnistereo video imaging with rotating optics,” *IEEE Transactions on Visualization & Computer Graphics*, vol. 11, no. 6, p. 614, 2005.
- [10] V. S. Nalwa, “A true omnidirectional viewer,” *Bell Labs.holmdel Nj Tech.rep*, pp. pgs. 476–488, 1996.
- [11] ———, “Panoramic viewing system with offset virtual optical centers,” 2001.
- [12] J. Shimamura, N. Yokoya, H. Takemura, and K. Yamazawa, “Construction of an immersive mixed environment using an omnidirectional stereo image sensor,” in *Omnidirectional Vision, 2000. Proceedings. IEEE Workshop on*, 2000, pp. 62–69.
- [13] O. Schreer, P. Kauff, P. Eisert, and C. Weissig, “Geometrical design concept for panoramic 3d video acquisition,” in *Signal Processing Conference*, 2012, pp. 2757–2761.
- [14] C. Weissig, O. Schreer, P. Eisert, and P. Kauff, “The ultimate immersive experience: panoramic 3d video acquisition,” in *International Conference on Advances in Multimedia Modeling*, 2012, pp. 671–681.
- [15] S. J. Koppal, C. L. Zitnick, M. Cohen, S. B. Kang, B. Ressler, and A. Colburn, “A viewer-centric editor for 3d movies,” *IEEE Computer Graphics & Applications*, vol. 31, no. 1, pp. 20–35, 2011.
- [16] Brown, Matthew, Lowe, and G. David, “Automatic panoramic image stitching using invariant features,” *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.
- [17] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, “Pyramid methods in image processing,” *Rca Engineer*, 1984.



Figure 9. Projection from the left and right panorama



Figure 10. Red and blue stereo panoramic result



Figure 11. Result for increasing the vertical viewing angle