VR HMD Compatible Spherical Stereo Panorama by Single Camera with Regular Lens

Zhouyuan Li
Google Inc.
1600 Amphitheatre Pkwy, Mountain View, CA 94043
lizhouyuandream@gmail.com

Jiafan Yu
Stanford University
450 Serra Mall, Stanford, CA 94305
jfy@stanford.edu

Abstract

A spherical stereo panorama is a pair of panoramic images with horizontal parallax, that both have a complete spherical projection covering 360° in azimuth and 180° in elevation. We present an analyze a new method to generate spherical stereo panorama by using one camera with regular lens, which are compatible for VR HMD display.

1. Introduction

The evolution of virtual reality requires both the innovation in hardware development, such as VR HMDs, as well as the development of a easy way for generating VR compatible contents by publics besides professional producers.

The process of generating traditional panoramic images, which have only one single viewpoint, and usually projected on a extended planer image surface (*e.g.*, iPhone's automatic panoramic mode), has been a mature technology, and there are a bunch of popular and stable algorithms for implementation [3]. To create a regular stereo image, which needs a pair of images from two viewpoints, corresponding to the views from left and right eyes, to get stereo disparities for objects at different distances from the viewpoints.

However, neither the monoscopic panorama nor a regular stereo image are not capable for providing a immersive environment on today's virtual reality (VR) head mounted display (HMD).

For displaying a fixed image with 100% VR immersive experience, the image has to be a panorama with a complete spherical projection to cover arbitrary viewing directions. In addition, for each viewing direction, there must be a pair of images providing horizontal disparities within a certain range, to stimulate the mechanism of human binocular vision, reproducing a credible and consistent perception of depth [4].

In 2001, Peleg *et al*. illustrated that stereo panoramas cannot be photographed by two omnidirectional cameras from two viewpoints [7]. In addition, they provided a method only using one rotating camera to generate planar stereo panorama. Li *et al*. further provides a depth reconstruction method based on planar stereo panorama [6]. Based on the rotating-camera method, Li *et al*. [5] and Bourke [2] used a single fish-eye camera rotating along the axis off the optical center to acquire mosaics and build a spherical stereo based on the mosaics.

For all these approaches mentioned above, they only considered one degree of freedom of camera movement, the rotation about the fixed axis. Since there is no elevation changes for camera, a fish-eye lens is necessary to get images of zenith and nadir to have a spherical panorama.

In 2011, Ainsworth *et al*. used a dual-camera image capture system with a designated dual-camera controller to generate spherical stereo panoramas [1].

Though the dual-camera image capture system enables 2 degrees of freedom for camera movement that can change both azimuth and elevation, the dual-camera system is very complicated and the control system is not easy to implement for publics.

Recently, Nokia and other companies released their commercialized products for spherical stereo panorama capturing devices, such as Nokia OZO, Samsung Beyond. Without any exception, all of them have multiple (¿5) cameras covering almost all directions on a viewing spheres. And the cost for these devices are expensive, which constrains them only for professional use.

In this paper, we propose and implement a simple method to generate so-called spherical stereo panorama, by using only one camera with regular lens, inspired by the ideas from generating planar stereo panorama by one cameras [7].

This course report is organized as follows: section II analyses the practical technical difficulties for spherical stereo panoramas generation; section III discusses photo stitching algorithms implementation; section IV shows the experimental results and section V is the conclusion.

2. Spherical Stereo Panorama Generation by Single Camera

A stereo panorama consists of two images taken from two different view points, with horizontal parallax, for all possible viewing directions. Figure 1 give an intuitive illustration of the spherical projection for stereo images. The two view points represent two eyes. However, by using

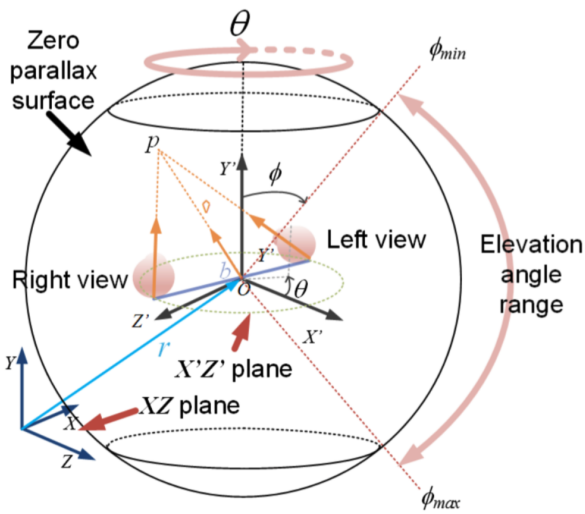


Figure 1: Illustration of Horizontal Parallax for Spherical Projection [4]

The basic idea is that when the camera is rotated about an axis behind the camera, the images create a volume that for each direction, we can find a projection

It has been analyzed in [7] that, stereo panoramas can be created with single cameras which is rotated about an axis behind the camera, which is shown in Fig. 2

2.1. 2-Degree of Freedom Rotation

In order to generate a spherical stereo panorama by using one camera, the camera must be rotating about a vertical axis to get images at different azimuth angle, to get a bunch of vertical slits at each image representing different viewpoints. Besides, the camera is also needed to be able to rotate about a horizontal axis, to get images at different elevation up/down to zenith/nadir. The most convenient tool for enabling 2-degree of freedom rotation is a regular tripod. The head of tripod could rotate horizontally about the axis behind the camera screen and the optical center, when the elevation angle is 0. The camera can also rotate for different elevation, about the horizontal axis downside the camera body.

However, since the horizontal axis does not pass through the optical center of the camera, the distance between the

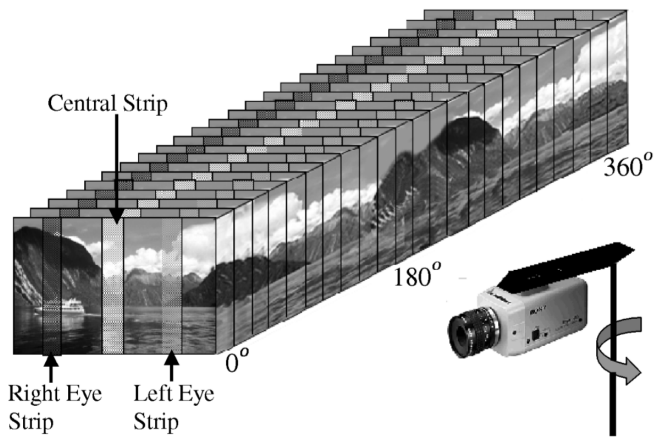
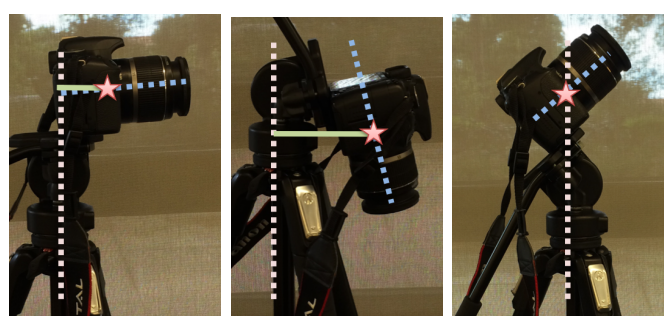


Figure 2: Illlustration of Rotating Camera for Stereo Panorama [7]

optical center and the vertical axis could change dramatically for different elevations. For example, as shown in Fig. 3a, when the elevation is 0, the distance between the camera's optical center and the vertical rotation axis is about 10 cm. When the elevation angle is changed to be $-70°$, the distance between optical center and rotating axis is about 15 cm, which is shown in Fig. 3b. However, when we lift th camera to 50° elevation, the optical center of the camera and the axis is overlapped, so that we can only get stereo images from extracting the strips for elevation lower than 50° for our setup.



(a) 0° Elevation (b) $-70°$ Elevation (c) 50° Elevation

Figure 3: Illustration of distances between optical center and vertical rotation axis for different elevations

Fig. 4 shows the geometric relationship of the setup. When we want to focus on objects at distance S, in other words, the 0-disparity plane is at distance S, the spanning angle of objects at 0-disparity plane is 2α, while the angle between camera's optical axis and the imaging light field is β. By denoting the distance between strip and picture center as v, the distance between rotating center and optical center

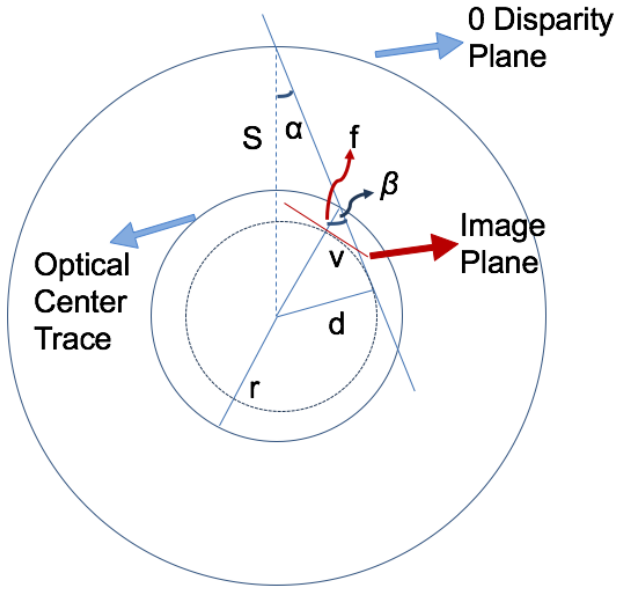


Figure 4: Geometric Relationship of Rotating Camera

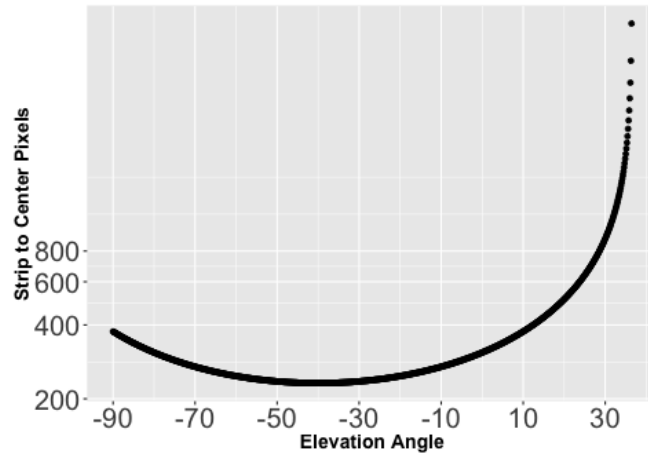


Figure 5: Strip Separation Adjustment for Different Elevations

as r, we have the relationship between the spanning angle and the strip separation:

$$\alpha = \sin^{-1}\left(\frac{r\sin(\tan^{-1}(v/f))}{S}\right).$$

It is clear that we can move the 0 disparity plane arbitrarily by changing the relative position of two panoramas from two strip set.

Furthermore, since S, d, r are also functions of the elevation, we can further plot the relationship between the elevation and corresponding strip separation in order to maintain the same 0-disparity plane distance and the spanning angle for 0-disparity plane distance, which is shown in Fig. 5.

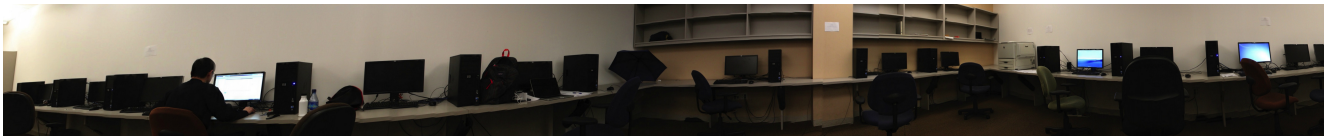
It could be clearly seen that when the elevation angle approaches 50°, the projected strip separation becomes to ∞, at this situation the optical center and the rotation axis overlaps and we are not able to generate stereo strip pairs.

3. Image Stitching of Photo Strips

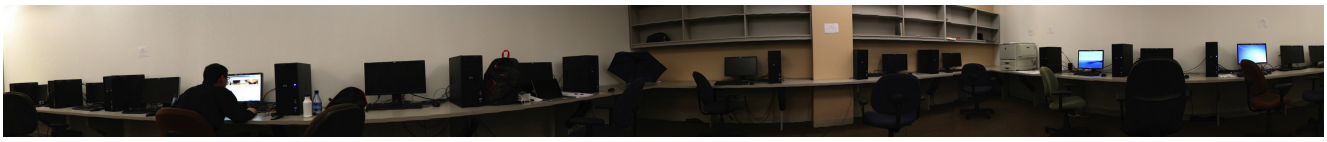
Image stitching pipeline is the core part of our panorama. We take photos from different angles with reasonable amount of overlapping. The output is a stitched wide angle spherical image. Brown and Lowe described this pipeline very clearly in their paper [3] and we followed his idea in our project.

The first stage is Feature Detection. We use SIFT feature because Lowe provided source code for SFIT feature extraction on their website and it is very easy for us to use. We first extract the SIFT features on all the images. After that, common features in overlapping regions of the captured images are matched and the optimal homography matrices between pairs of images are calculated for all possible pairs. For a particular feature in image A, we can find its nearest and second nearest neighbor in image B. If its nearest neighbor is much closer than its second nearest neighbor, then the feature in image A and its nearest neighbor in image B is a good match. In this way, we can get all pairs of matched features between each pair of two images. To estimate the pairwise Homography matrix, it has been proved that for all points that lie on the same plane in 3D coordinate system, their projected coordinates onto 2 cameras satisfy a simple homography transformation and that homography matrix has eight degrees of freedom. To estimate the homography matrix we need 4 pairs of matching features, but actually we have much more. Each group of 4 matched features corresponds to a homography matrix and we need to use RANSAC (random sample consensus) algorithm to pick out the best. The best estimated homography matrix after RANSAC contains inliers and outliers. The higher the inlier and the lower the outlier, the more confident we are that the two images contain overlapping region. Each pair of two images has a homography matrix. The homography matrices estimated with higher confidence than a threshold are valid. The homography matrices estimated with lower confidence than a threshold are invalid.

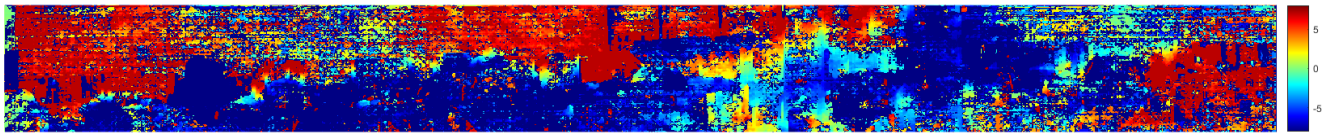
For the bundle adjustment, we have both naive approach and better approach. In our mono panorama case, I wrote MATLAB code to do bundle adjustment heuristically. Before bundle adjustment, the stitched together image can drift dramatically. I wrote a matching function which drags drifted partial image back to horizontal line periodically. The works well for mono panorama but doesnt work



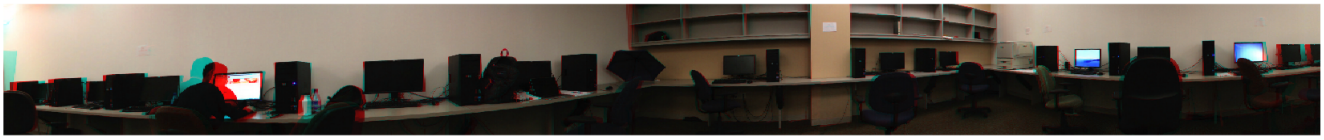
(a) Left Strips Stitched Planar Panorama



(b) Right Strips Stitched Planar Panorama



(c) Disparity Map of Two Panoramas



(d) Disparity Map of Two Panoramas

Figure 6: Planar Stereo Panorama

for spherical. In the spherical case, we utilized the same method as Lowe. After bundle adjustment is image warping. In the mono panorama case, I defined a transformation matrix for all images to the image coordinate of the first image in MATLAB so that all images can warp together.

Our mono panorama experiment ends here, but our spherical panorama has two additional steps: Gain compensation and Multi-band blending. Ideally, the overlapping areas between two images should have the same intensity. However, this is not the case in reality. The idea of gain compensation is that different images will have different intensity so that when we warp then together, we should make the intensity to be roughly equal.

After gain compensation, image edges are still visible. Boundary pixels aren't very likely to have the same intensity. In order to smooth the boundaries, we can perform blending of all the images. Blending essentially means assigning weights for different images on the overlapping area. For a particular image, a naive approach of setting weights would be setting 1 to the center pixel and vary the weights to zero at the boundary. Boundary effect is almost eliminated by the approach. However sine blending is somewhat like a low pass filter, high frequency component is also filtered out. To preserve the sharpness of images, we can perform multi-band blending. The idea of multi-band is to blend over a short range for high frequency component, and over a large range for low frequency component.

4. Experimental Results

In order to achieve the final spherical panorama generation goal, we manage to make progresses on two milestones:

- Generating stereo panorama but for cylindrical projection

  For cylindrical projection stereo panorama, since we fix the elevation angle at 0°, we do not need to worry about the strip separation adjustment. Furthermore, for our convenience to not to introduce too much disparity, we choose to take a series of photos in in-door area, so that the depths of different objects do not have dramatic change.

  We take 90 photos for a 360° rotation along the vertical axis behind the camera, where the angular distance for each image is 4°. The original size of the image is 1200 pixels by 800 pixels, while the sensor size of the camera (Canon 500D) is 22.3 x 14.9 mm. And we choose the separation between the strip center to the image center is 300 pixels, while each strip width is 200 pixels. Then we have two sets of 200 x 800 strips, each sets contains 90 pictures. Since we only need to

Figure 7: Spherical Monoscopic Panorama for 0 elevation

Figure 8: Spherical Monoscopic Panorama for -10° elevation

focus on the photo stitching on one direction, it is relative easy for us to generate two panoramas from each strip sets. The two panoramas representing different viewpoints are shown in Fig. 6a and Fig. 6b. The associated disparity map for the two images are shown in Fig. 6c. The positive disparity for the walls which are far away from the camera and the negative disparity for computers and chairs which are closer to the camera are clearly illustrated in the disparity map. Finally, the anaglyph 3D graph is shown in Fig. **??** which could be seen from a pair of 3D red cyan glasses.

- Generating spherical projection panorama but for monoscopic panorama

  Compared with cylindrical projection for strip set stitching, it is much more difficult to construct a spherical projection panorama, especially from narrow strips. Firstly we try to construct cylindrical panorama for each elevation angle, as shown in Fig. 7 for 0 elevation, and Fig. 8 for -10° elevation. However, it is extremely difficult for us to use same algorithm to stitch the two cylindrical panorama to a unified image, because the very strong distortion along the top and bottom of the photos. Several improvements have been made, including increasing the width of the strips and using external package to stitch photos for different azimuths and elevations in one time, instead of the 2-step process. The final results for a monoscopic spherical panorama is shown in Fig. 9.

The final integration of the two partial progresses, that generating spherical stereo panorama seamlessly using narrow strips are under development.

5. Conclusion

Inspired by the planar stereo panorama generation by using single camera, we implement the method, and extend the technology to spherical projection by using a 2-degree of freedom tripod to allow both azimuth rotation and elevation rotation.

Two milestones have been achieved, including generating stereo panoramas for cylindrical projection for a fixed elevation, as well as generating monoscopic spherical projection panorama from narrow image strips.

The major challenges of finishing the final integration are:

- The difficult of 2 dimensional photo stitching for narrow strips, and
- The handling of zenith and nadir.

It is promising to use well calibrated 2-degree rotation system to solve these challenges. By using a finely tuned system with calibration, we can accurately record the elevation and azimuth angle for each strip, so that we do not need to implement photo stitching algorithms by finding matching features, and just need to directly mosaic the photo strips and project them to a spherical coordinate system.

References

[1] R. A. Ainsworth, D. J. Sandin, J. P. Schulze, A. Prudhomme, T. A. DeFanti, and M. Srinivasan. Acquisition of stereo panoramas for display in vr environments. In *IS&T/SPIE Electronic Imaging*, pages 786416–786416. International Society for Optics and Photonics, 2011.

[2] P. Bourke. Capturing omni-directional stereoscopic spherical projections with a single camera. In *Virtual Systems and*

Figure 9: Spherical Monoscopic Panorama for All Available Elevations

Multimedia (VSMM), 2010 16th International Conference on, pages 179–183. IEEE, 2010.

[3] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007.

[4] L. E. Gurrieri and E. Dubois. Acquisition of omnidirectional stereoscopic images and videos of dynamic scenes: a review. *Journal of Electronic Imaging*, 22(3):030902–030902, 2013.

[5] S. Li and K. Fukumori. Spherical stereo for the construction of immersive vr environment. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*, pages 217–222. IEEE, 2005.

[6] Y. Li, H.-Y. Shum, C.-K. Tang, and R. Szeliski. Stereo reconstruction from multiperspective panoramas. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(1):45–62, 2004.

[7] S. Peleg, M. Ben-Ezra, and Y. Pritch. Omnistereo: Panoramic stereo imaging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(3):279–290, 2001.