Virtual Reality Motion Parallax with the Facebook Surround-360 using a Depth Augmented Stereo Panorama

David Lindell and Jayant Thatte

February 11, 2017

1 Motivation

Recent interest in virtual reality has led to the development of 360-degree media platforms. In general, 360-degree images or videos can be observed using a head-mounted display (HMD) which tracks head position and renders images for each eye with low latency. Three-dimensional scenes can be portrayed by such displays by incorporating depth cues important to human perception of 3D position. Such cues include binocular disparity, occlusions, and head-motion parallax. Binocular disparity is achieved by presenting separate views of a scene, separated by some baseline, to each eye. Head-motion parallax is accomplished by re-rendering scenes as head-orientation changes and previously occluded objects become visible. Though such visual cues are fairly straightforward to incorporate into computer-generated content (where the position and texture of all scene objects are already known), accurately portraying real-world content in a way that is convincing and comfortable to the virtual-reality consumer is more difficult.

In order to generate omnidirectional views of real-world content, special camera rigs (e.g. the Facebook Surround 360[1]) have been developed which typically incorporate multiple cameras which can together capture a full 360-degree view. From the acquired images, a monocular or stereo view can be rendered, which portrays the scene from a single vantagepoint. However, with only a single viewpoint, or head position, such scenes cannot support head motion parallax and are less convincing and less comfortable for a viewer. Although capturing many images from multiple viewpoints would enable rendering of views across a range of head motion, the resulting increase in the amount of data to be stored and transmitted makes such an approach less appealing.

Other current approaches, such as novel view synthesis or structure from motion techniques require extensive post-processing on scene capture from multiple viewpoints or motion of the 3D camera rig to infer an unknown viewpoint. Our approach enables head-motion parallax by converting the data acquired from a stationary camera rig into a succinct representation which incorporates the 3D scene information.

2 Related Work

Multiple methods exist for acquiring real-world data which can be used to generate scenes supporting head-motion parallax. Data from multiple viewpoints can be captured directly, stereo views can be synthesized using image processing or machine learning techniques, or, the raw data acquired by a camera rig can be converted to some suitable format which incorporates 3D information about the scene.

One way of directly capturing more scene viewpoints to enable motion parallax is the technique of acquiring concentric mosaics [1]. In this method, many panoramas are acquired by moving a camera around a set of concentric circles. The data can then support motion parallax for head-motion withim the region of the concentric camera captures. This method, however, has the drawback of requiring large amounts of data capture.

Other methods of synthesizing novel views in for views not directly captured by a camera include that of [2], which uses image processing to synthesize a novel view. The procedure segments an input image into many areas of roughly the same size, depth, and texture. Then, each segment is warped to the new view, with a segment-shape preserving constraint. In this method, a few different views are typically warped to the novel view, and the holes which occur in any particular warping are filled in with another warped view. The reconstructed novel view may have warping distortions, and the segmentation approach also fails to capture thin objects. Deep learning has also been used to generate novel views, as in [3], where a network is trained on multiple views of a scene, and then generates an unseen view. This method, however, requires a network to be trained on multiple neighboring views before it can infer the new view.

Another approach to providing motion parallax for real-world scenes uses a monoscopic 3D video (where the viewpoint moves) to infer the 3D geometry of a scene using structure from motion. As the

[1] https://facebook360.fb.com/facebook-surround-360/

video is played back, the scene is warped according to the viewer head position to achieve the motion parallax effect [4].

Our proposed solution is to format the raw data from the camera rig into the Depth Augmented Stereo Panorama (DASP) representation [5, 6]. In this representation, the images from the camera rig are converted a stereo pair of panoramas along with corresponding depth maps. If the left and right stereo pairs are separated by a baseline which exceeds that of the typical human interpupillary distance, the stereo pairs can be used to interpolate scene stereo views for a human baseline, along a range of head positions within the larger baseline. This approach reduces the amount of data that must be transferred off the rig to two stereo panorama images (one for texture and one for depth), while yet enabling motion parallax in a low-complexity rendering pipeline.

3 Project Overview

In this project, we demonstrate motion parallax for immersive virtual reality using the Depth-Augmented Stereo Panorama (DASP) representation of [5] and [6]. In order to achieve motion-parallax on a head mounted display at interactive rates, we use data from the open-sourced Facebook Surround 360 camera rig and alter the open-source processing pipeline.

The Facebook Surround 360 camera rig consists of a tripod on which 17 cameras are mounted. Fourteen side-looking cameras with wide-angle lenses are mounted on a 46 cm diameter ring, one camera with a fisheye lens is mounted looking up, and two cameras with fisheye lenses are mounted looking down (enabling the tripod pole to be processed out of the final panorama). While Facebook does not commercially produce the rig, they provide documentation on parts and assembly and also open source code for capturing data and rendering[2]. Raw captured data from the rig are also available.

While the open source code enables rendering stereo panoramas with a baseline of 6.4 cm, which corresponds to the human interpupillary distance, the DASP representation for motion parallax requires panoramas with a much larger baseline, on the order of 20-30 cm. Since increasing the baseline with the default code results in holes in the output panoramas, further processing techniques must be developed to expand the baseline and still produce tenable panoramas. The size of the camera rig ring (46 cm) ultimately limits the extension of the baseline. Additionally, scene depth must be determined and output along with the wide-baseline stereo image panoramas.

Once a DASP representation can be produced from the camera rig, a 3D point cloud of the scene can be assembled (see [6]). For any eye position within a viewing circle determined by the DASP baseline, a viewport can be rendered by projection of the 3D point cloud. The head-motion parallax effect can then be achieved on a HMD either by pre-computing stereo viewports at finely sampled head positions, or optimizing the processing to run at real-time rates, e.g. 30 frames per second or higher. The motion parallax effect can then be achieved on the head-mounted display.

4 Milestones and Goals

The ultimate goal for the project is to demonstrate motion parallax for a scene on a HMD at real-time rates. Ideally all processing would be done in real-time to render the stereo viewports given the head position provided by the HMD. Milestones are summarized in the below table.

Milestone	**Projected Date Finished**	**Finished?**
Modify Facebook code to produce output panoramas with expanded baseline	Jan. 30	✓
Modify Facebook code to produce output depth panoramas with expanded baseline	Feb. 6	✓
Render viewports from DASP using MATLAB code	Feb. 17	
Demonstrate monocular (or stereo) head motion parallax on HMD using precomputed viewports	Feb. 25	
Find a way to improve the depth map by enforcing optical flow consistency across cameras	Mar. 3	
Port MATLAB rendering code to OpenGL to render viewports at real-time rates	Mar. 6	
Real-time motion parallax demo ready	Mar. 14	

[2] https://github.com/facebook/Surround360

References

[1] H.-Y. Shum and L.-W. He, "Rendering with concentric mosaics," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pp. 299–306, 1999.

[2] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation," *ACM Trans. Graph.*, vol. 32, pp. 30:1–30:12, July 2013.

[3] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deepstereo: Learning to predict new views from the world's imagery," *CoRR*, vol. abs/1506.06825, 2015.

[4] J. Huang, Z. Chen, D. Ceylan, and H. Jin, "6-DOF VR videos with a single 360-camera," in *CVPR*, 2017 (submitted).

[5] J. Thatte, J. B. Boin, H. Lakshman, G. Wetzstein, and B. Girod, "Depth augmented stereo panorama for cinematic virtual reality with focus cues," in *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1569–1573, Sept 2016.

[6] J. Thatte, J. B. Boin, H. Lakshman, and B. Girod, "Depth augmented stereo panorama for cinematic virtual reality with head-motion parallax," in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, July 2016.