Stacked OmniStereo for Virtual Reality with Six Degrees of Freedom

Jayant Thatte (*jayantt@stanford.edu*) and Bernd Girod
Department of Electrical Engineering

SCIEN
The Stanford Center for Image Systems Engineering

Abstract

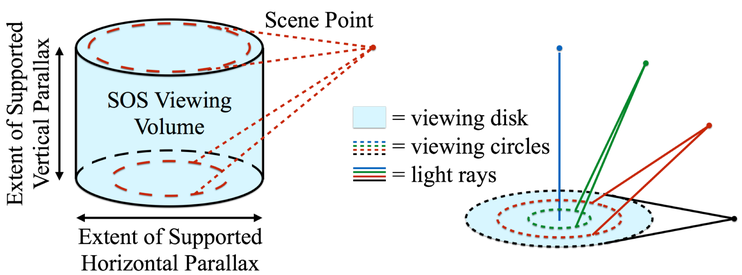
Rendering accurate motion parallax is crucial for a comfortable and immersive virtual reality (VR) experience. Most live-action content available today is recorded and rendered from a fixed viewpoint causing "VR sickness". We introduce Stacked Omnistereo, a novel data representation that can render VR with 6 degrees of freedom

Comparison with Other Representations

Data Representation	Rotation Supported	Translation Supported	Data Footprint (No. of Panoramas)	Handles Speculars
Stereo Pair	No	No	2 (images)	NA
Omnistereo [1]	Yaw	No	2	NA
Concentric Mosaics [2]	Yaw	Horizontal	~ 40	Yes
DASPs [3]	Yes	Horizontal	2 texture, 2 depth	Limited
Light fields	Yes	Yes	~ 50-300	Yes
SOS (Proposed)	Yes	Yes	4 texture, 4 depth	Limited

Stacked OmniStereo (SOS)

- Comprises 4 texture-plus-depth RGBD panoramas
- Maps complete 3D space without "blind regions"
- Novel views can be synthesized, usually without inpainting, for viewpoints within the viewing volume
- The dimensions of the viewing cylinder must therefore exceed the anticipated range of the viewer's head-motion



Left: Stacked OmniStereo construction – scene point P gets recorded in 4 panoramas from 4 distinct viewpoints
Right: Omnistereo modified to map 3D space without blind regions

References

[1] Peleg et al. "Omnistereo: panoramic stereo imaging," *PAMI,* 2001
[2] Shum et al. "Rendering with concentric mosaics," SIGGRAPH 1999
[3] Thatte et al. "Depth augmented stereo panorama for cinematic virtual reality with head-motion parallax," *IEEE Conf. on Multimedia & Expo (ICME)*, 2016
[4] Zbontar et al. "Computing the Stereo Matching Cost with a Convolutional Neural Network", Computer Vision and Pattern Recognition (CVPR), 2015

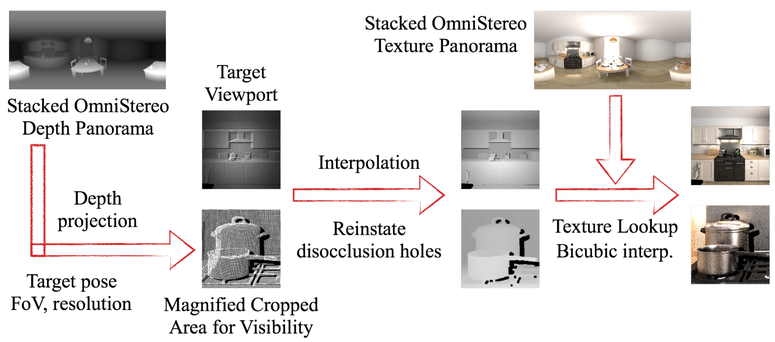
Construction: From Camera Rig to SOS

Comparison

Unlike conventional algorithms, jointly using all input cameras for stitching allows us to generate wide-baseline panoramas and thus support a larger range of head motion

A section of the output panorama: (a-c) Using Facebook Surround 360 algorithm, baselines 6, 12, 18 cm respectively. (d) Using proposed method, baseline 30 cm. Notice the missing region artifacts present in (a-c).

View Synthesis: From SOS to Viewport



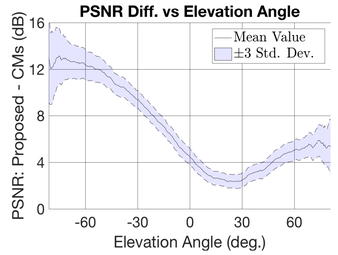
Conclusions

- High fidelity novel views can be synthesized using 4 texture-plus-depth panoramas within a fixed volume
- Outperforms state-of-the-art VR representations by avg. PSNR of up to 3 dB when evaluated for 6-DoF motion
- While SOS can synthesize plausible view-dependent specular highlights, highly non-Lambertian effects such as reflection and transparency are rendered incorrectly

Results

- Quantitative evaluation using 6 synthetic scenes based on 3D models of real environments. We use Blender for photorealistic rendering with specular highlights, reflections, refractions, and indirect illumination.
- Avg. PSNR across 1200 views using 6-DoF motion CM [2]: 29.3 dB, DASP [3]: 33.8 dB, proposed: 34.7 dB

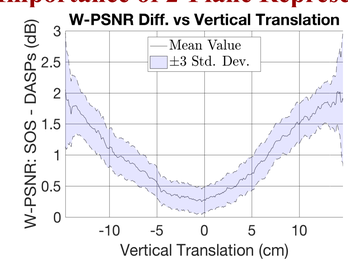
Importance of Depth to Correct Perspective Distortion



PSNR gain for synthesized views using SOS (proposed) over purely image-based representation [2]

Left: Views synthesized using Concentric Mosaics [2] suffer from perspective distortion. Right: SOS (proposed) uses depth to render the correct perspective

Importance of 2-Plane Representation for Vertical Translation



PSNR gain for synthesized views using SOS (proposed) over single-plane representation – DASPs [3]

Left: Artifacts in views from DASPs [3] (single plane); Right: Correct views using SOS

View Dependent Specular Highlights

Notice the changes in specular highlights across the rendered views

Limitations:

Ground truth | Incorrect reflection | Ground truth | Incorrect transparency