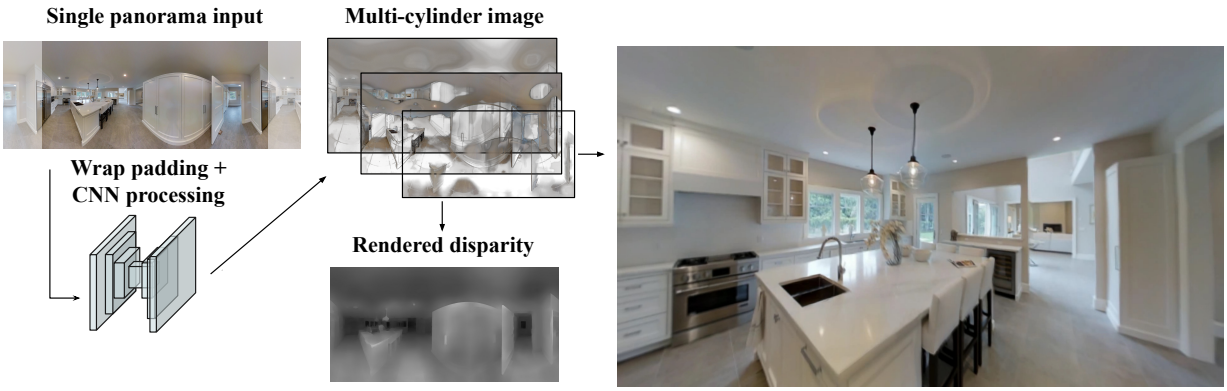
PanoSynthVR: View Synthesis From A Single Input Panorama with Multi-Cylinder Images

Richa Gadgil
Reesa John
California Polytechnic State University
USA
rgadgil@calpoly.edu
rejohn@calpoly.edu

Stefanie Zollmann
University of Otago
New Zealand
stefanie.zollmann@otago.ac.nz

Jonathan Ventura
California Polytechnic State University
USA
jventu09@calpoly.edu



(a) Overview of method

(b) Example view synthesis result

Figure 1: Our method provides free-viewpoint view synthesis from a single input panorama. The multi-cylinder image representation consists of semi-transparent cylindrical layers at varying depths. (a) We produce a multi-cylinder image by processing a horizontally wrap-padded panorama in a convolutional neural network. (b) Real-time view synthesis is achieved by projecting and compositing the textured cylinders with over blending. An animated version of (b) is available online.

ABSTRACT

We introduce a method to automatically convert a single panoramic input into a multi-cylinder image representation that supports real-time, free-viewpoint view synthesis for virtual reality. We apply an existing convolutional neural network trained on pinhole images to a cylindrical panorama with wrap padding to ensure agreement between the left and right edges. The network outputs a stack of semi-transparent panoramas at varying depths which can be easily rendered and composited with over blending. Initial experiments show that the method produces convincing parallax and cleaner object boundaries than a textured mesh representation.

SIGGRAPH '21 Posters, August 09-13, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8371-4/21/08.
https://doi.org/10.1145/3450618.3469144

ACM Reference Format:
Richa Gadgil, Reesa John, Stefanie Zollmann, and Jonathan Ventura. 2021. PanoSynthVR: View Synthesis From A Single Input Panorama with Multi-Cylinder Images. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '21 Posters), August 09-13, 2021.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3450618.3469144

1 INTRODUCTION

Depth cues such as binocular stereo and motion parallax are important aspects of immersion in virtual reality (VR). To achieve these effects when rendering captured imagery and video, the system must be able to provide free-viewpoint view synthesis in real time. Recent work on view synthesis either is targeted to perspective images with a limited field-of-view (FOV) [Shih et al. 2020], or uses stereo [Attal et al. 2020] or multi-view imagery [Broxton et al. 2020; Serrano et al. 2019] as input.

In contrast, our approach requires only a single panorama as input, which can be easily captured with a smartphone or a consumer-level 360 camera, and is thus compatible with the vast amount of existing 360 imagery available on the Internet. In this work we address the challenges inherent in applying a single-view multi-plane image (MPI) network to panoramic images, and present what is to our knowledge the first work capable of producing a multi-cylinder

image (MCI) representation from a single panoramic input. We call our approach PanoSynthVR.

2 METHOD

Our approach builds upon the single-view MPI network [Tucker and Snavely 2020], which automatically converts a perspective image into an MPI representation. The MPI consists of semi-transparent layers at varying depths, which are easily rendered and composited with over blending to achieve view synthesis. However, their results are not suitable for immersive virtual reality experiences because of the limited FOV.

To extend this concept to panoramic input to support immersive viewing, we first considered using a cube map panorama representation, which consists of six square perspective images with 90° FOV arranged into a cube. Since the MPI network was trained on square perspective images, this is a natural representation to use. We ran the MPI network on each cube face individually and then stitched them back into semi-transparent cube map layers. However, we found that there was disagreement in the disparity estimate across the edges of each cube map face, since the network was run on each cube face separately.

To address the issue of disparity disagreement at the seams, we instead used a cylindrical panorama representation, in which the entire scene (except the very top and bottom) is represented in a single image. Although the cylindrical projection introduces distortions, locally it appears similar to a pinhole image. However, we still found disagreement in the disparity estimate across the vertical seam on the left and right edges of the panorama.

To address this, we added horizontal wrap padding to the input image. We concatenate the left half of the image on the right and vice versa. After processing the padded image in the network, we crop the excess on the left and right sides of each layer. With the addition of wrap padding, we no longer observed disagreement across the seam, and the MPI network is able to produce a consistent representation of the entire scene.

Our final processing pipeline is summarized in Figure 1. We created a real-time renderer for our multi-cylinder image representation in WebXR as a proof-of-concept. The renderer can be used in VR headsets such as the Oculus Quest.

3 RESULTS

For our experiments we selected a set of ten panoramas from the Matterport3D dataset [Chang et al. 2017], which consists of color panoramas and associated depth maps captured mostly indoors in a real estate context. The single-view MPI network [Tucker and Snavely 2020] was trained on videos from homes for sale on the internet and thus we expected the network to generalize well to these images. We used 32 cylinders, uniformly sampled in inverse depth from depth 1 to 100, in all experiments. More view synthesis results can be seen in the video provided in supplementary material, and our WebXR renderer is also available online[1].

We compared our MCI results to a textured mesh representation. To create the textured mesh, we extracted the disparity map estimate from our results and used it as a displacement map on a cylinder geometry. For fair comparison and to ensure real-time rendering, we restricted each representation to have the same number of vertices. The textured mesh result has noticable artifacts such as texture "stretching" at the edges of objects. However, the MCI is more blurry than the textured mesh in some areas, which is an acknowledged issue in the single-view MPI network output [Tucker and Snavely 2020].

3.1 Feasibility and Preliminary User Feedback

We demonstrated our system to four users in a preliminary study using an Oculus Quest headset with Oculus Link. Users were shown VR renderings of 9 randomized scenes with three randomized conditions 1) PanoSynthVR, 2) Plain 360 panorama and 3) Textured mesh (depth map). Initial feedback showed that users experienced less distortions in the PanoSynthVR results compared to the textured mesh: e.g. mentioning "less wrinkles in the window areas." For the majority of scenes, users rated PanoSynthVR as their preferred option when comparing to the textured mesh. However, they also mentioned the reduced resolution of PanoSynthVR.

4 CONCLUSIONS AND FUTURE WORK

In this work we introduced the MCI representation for real-time, immersive free-viewpoint view synthesis, and demonstrated an initial pipeline to extract an MCI from a single input panorama.

While in this preliminary work we applied an existing network trained on perspective images, we plan to explore training or fine-tuning the network on cylindrical images. Future work could explore how to extract a layered mesh representation [Broxton et al. 2020] from the MCI to decrease storage requirements and possibly improve the sharpness of the results.

We also planning to conduct a more complete user study. In particular, we are interested in analyzing the impact of PanoSynthVR on presence and depth perception, as well as further evaluation of visual quality compared to plain panoramas and the textured mesh.

ACKNOWLEDGMENTS

This work was partially supported by NSF Award No. 1924008 and the New Zealand Marsden Council through Grant UOO1724.

REFERENCES

Benjamin Attal, Selena Ling, Aaron Gokaslan, Christian Richardt, and James Tompkin. 2020. MatryODShka: Real-time 6DoF Video View Synthesis using Multi-Sphere Images. In *European Conference on Computer Vision (ECCV)*. https://visual.cs.brown.edu/matryodshka

Michael Broxton, John Flynn, Ryan Overbeck, Daniel Erickson, Peter Hedman, Matthew DuVall, Jason Dourgarian, Jay Busch, Matt Whalen, and Paul Debevec. 2020. Immersive Light Field Video with a Layered Mesh Representation. 39, 4 (2020), 86:1–86:15.

Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. 2017. Matterport3D: Learning from RGB-D Data in Indoor Environments. *International Conference on 3D Vision (3DV)* (2017).

Ana Serrano, Incheol Kim, Zhili Chen, Stephen DiVerdi, Diego Gutierrez, Aaron Hertzmann, and Belen Masia. 2019. Motion parallax for 360 RGBD video. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 1817–1827.

Meng-Li Shih, Shih-Yang Su, Johannes Kopf, and Jia-Bin Huang. 2020. 3d photography using context-aware layered depth inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8028–8038.

Richard Tucker and Noah Snavely. 2020. Single-view View Synthesis with Multiplane Images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[1] https://jonathanventura.github.io/PanoSynthVR