Interactive Fusion of 360° Images for a Mirrored World

Ruofei Du*, David Li†, and Amitabh Varshney‡ *Fellow, IEEE*

Augmentarium, Department of Computer Science, and University of Maryland Institute for Advanced Computer Studies

University of Maryland, College Park

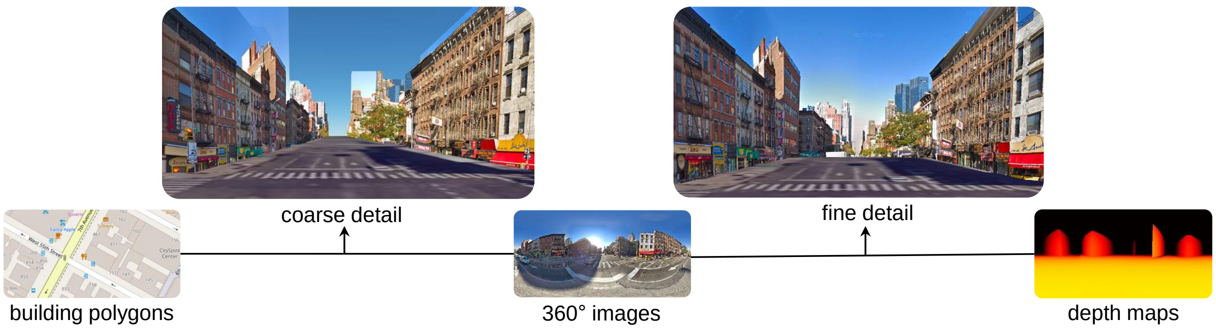

Figure 1: Results and overview of our interactive reconstruction pipeline. Our system is available at `https://geollery.com`.

Abstract

Reconstruction of the physical world in real time has been a grand challenge in computer graphics and 3D vision. In this paper, we introduce an interactive pipeline to reconstruct a mirrored world at two levels of detail. Given a pair of latitude and longitude coordinates, our pipeline streams and caches depth maps, street view panoramas, and building polygons from Google Maps and OpenStreetMap APIs. At a fine level of detail for close-up views, we render textured meshes using adjacent local street views and depth maps. When viewed from afar, we apply projection mappings to 3D geometries extruded from building polygons for a coarse level of detail. In contrast to teleportation, our system allows users to virtually walk through the mirrored world at the street level. We present an application of our approach by incorporating it into a mixed-reality social platform, Geollery, and validate our real-time strategies on various platforms including mobile phones, workstations, and head-mounted displays.

Keywords: virtual reality, 360° image, 3D reconstruction, mixed reality, projection mapping, mirrored world

Index Terms: I.3.3 [Computer Graphics]: Image Manipulation—Image-based rendering I.4.6 [Computer Graphics]: Graphics Systems and Interfaces—Virtual realities

1 Introduction

3D models of the physical world are widely used in a diverse set of applications including virtual tourism [9, 10, 14], geographical education [12], and neighborhood auditing [1]. However, interactive reconstruction of a mirrored world remains a significant challenge. On the one hand, commercial products such as Google Earth[1], offer world-scale, textured meshes at the aerial level, but the texture quality downgrades significantly for close-up views. Moreover, it does not allow users to freely walk in the virtual environments due to occlusion from the satellite imagery. On the other hand, classic high-fidelity approaches to modeling the 3D world have concentrated on generating 3D meshes using raw input data with the structure-from-motion (SfM) pipelines [13, 15, 16]. Despite the effectiveness of these offline systems, their data requirements and processing requirements make them unsuitable for mobile and web applications with processing and bandwidth constraints.

We introduce an interactive pipeline of fusing 360° images for a mirrored world at two levels of detail (Fig. 1). At a fine level of detail for close-up views, we incorporate multiple Google Street View panoramas and depth data to reconstruct textured meshes directly on the GPU. At a coarse level of detail when viewed from afar, we create extruded boxes with the building metadata from OpenStreetMap[2] and texture the meshes with street view panoramas. We contribute a web-based architecture to stream, cache, reconstruct, and render the mirrored world in real time. Our system, Geollery [8], is available at `https://geollery.com`.

2 Rendering Algorithms

At a fine level of detail for close-up views, we reconstruct an approximate geometry based on the depth maps associated with each street view and propose ways of seamlessly aligning the adjacent street view geometries. As illustrated in Fig. 2, this approach takes advantage of the high resolution of the street view images while incorporating low-resolution depth maps to generate an approximate geometry.

At a coarse level of detail when viewed from afar, we source data from OpenStreetMap using the Overpass API[3] to obtain 2D polygons for buildings. While these polygons are not as widely available as street view images, we find that in urban areas such as New York City, 2D building polygons often come with useful metadata such as the height in meters or the number of floors for each building. To convert these 2D polygons into 3D, we extrude them to the correct height based on the information provided in the metadata. For instances where metadata is not available, we extrude them to a predefined height of 16 meters to represent a 4-story building. Neither method requires any server-side preprocessing and all client-side processing can be done in background threads for interactive applications.

*e-mail: me@duruofei.com Note that the first two authors contribute equally to the paper. Geollery is developed at University of Maryland, College Park. Ruofei Du is now a Research Scientist at Google LLC.

†e-mail: dli7319@terpmail.umd.edu

‡e-mail: varshney@cs.umd.edu

[1]Google Earth: `https://www.google.com/earth`

[2]OpenStreetMap: `https://openstreetmap.org`

[3]Overpass: `https://wiki.openstreetmap.org/Overpass_API`

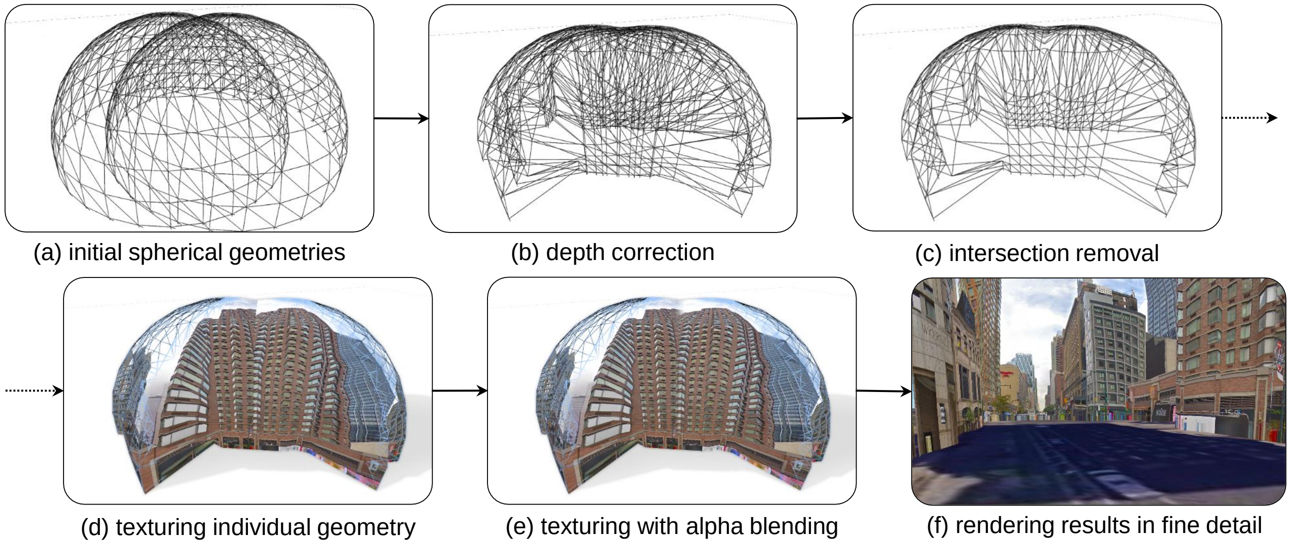

Figure 2: Our rendering pipeline for fusing 360° images in fine detail. (a) We start by generating spherical geometries to represent the 360° images. The segments of the sphere are greatly reduced for visualizing the geometries clearly. (b) In the vertex shader, we correct the depth value for each vertex by sampling the depth maps with its spherical coordinates. (c) In the fragment shader, we discard the pixels in the intersection of the adjacent geometries. (d) Texturing individual spheres with the corresponding 360 images may leave a visible seam. (e) Texturing the spheres with a weighted average according to the position to the camera yields a smoother result. (f) Finally, we interactively texture the ground plane with the corresponding satellite images, apply Gaussian filters in occluded regions, and allow the users to freely walk along the street while streaming the next street views on the go.

3 Conclusion and Future Vision

We have briefly introduced two real-time algorithms of interactive fusion of 360° images for a mirrored world. These approaches have been integrated into Geollery, a novel mixed-reality social platform.

With rapid advances in virtual and augmented reality, we envision a system that fuses a variety of multimedia data [2] to create a vivid mirrored world, in which it fuses not only 360° images and geotagged social media [7, 9], but also live surveillance videos [3], the user's telepresence from multiview videos [4, 5], and elaborately reconstructed architectures. One may adopt foveated rendering [11] to accelerate the rendering process and leverage visual cryptography [6] for transferring secure messages in XR. Such a system may eventually change the way we communicate with each other and consume the information throughout the world in immersive and mobile environments.

References

[1] P. Clarke, J. Ailshire, R. Melendez, M. Bader, and J. Morenoff. Using Google Earth to Conduct a Neighborhood Audit: Reliability of a Virtual Audit Instrument. *Health & Place*, 16(6):1224–1229, 2010. doi: 10.1016/j.healthplace.2010.08.007

[2] R. Du. *Fusing Multimedia Data Into Dynamic Virtual Environments*. PhD thesis, University of Maryland, College Park, Nov. 2018.

[3] R. Du, S. Bista, and A. Varshney. Video Fields: Fusing Multiple Surveillance Videos Into a Dynamic Virtual Environment. In *Proceedings of the 21st International Conference on Web3D Technology*, Web3D, pp. 165–172. ACM, Jul. 2016. doi: 10.1145/2945292.2945299

[4] R. Du, M. Chuang, W. Chang, H. Hoppe, and A. Varshney. Montage4D: Interactive Seamless Fusion of Multiview Video Textures. In *Proceedings of ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D, pp. 124–133. ACM, May. 2018. doi: 10.1145/3190834.3190843

[5] R. Du, M. Chuang, W. Chang, H. Hoppe, and A. Varshney. Montage4D: Real-Time Seamless Fusion and Stylization of Multiview Video Textures. *Journal of Computer Graphics Techniques*, 1(15):1–34, Jan. 2019.

[6] R. Du, E. Lee, and A. Varshney. Tracking-Tolerent Visual Cryptography. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, VR, p. 2, Mar. 2019.

[7] R. Du, D. Li, and A. Varshney. Experiencing a Mirrored World With Geotagged Social Media in Geollery. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA, p. 4, May. 2019.

[8] R. Du, D. Li, and A. Varshney. Geollery: a Mixed Reality Social Media Platform. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI)*, CHI, p. 13. ACM, May. 2019. doi: 10.1145/3290605.3300915

[9] R. Du and A. Varshney. Social Street View: Blending Immersive Street Views With Geo-Tagged Social Media. In *Proceedings of the 21st International Conference on Web3D Technology*, Web3D, pp. 77–85. ACM, Jul. 2016. doi: 10.1145/2945292.2945294

[10] R. Du and A. Varshney. Systems, Devices, and Methods for Generating a Social Street View. US Patent App. 15/559,955, Mar. 2016.

[11] X. Meng, R. Du, M. Zwicker, and A. Varshney. Kernel Foveated Rendering. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 1(5):1–20, May. 2018. doi: 10.1145/3203199

[12] T. C. Patterson. Google Earth As a (not Just) Geography Education Tool. *Journal of Geography*, 106(4):145–152, 2007. doi: 10.1080/00221340701678032

[13] B. Russell and R. Martin-Brualla. 3D Wikipedia: Using Online Text to Automatically Label and Navigate Reconstructed Geometry. *ACM Transactions on Graphics (TOG)*, 32(6):1–10, 2013. doi: 10.1145/2508363.2508425

[14] N. Snavely, S. M. Seitz, and R. Szeliski. Photo Tourism: Exploring Photo Collections in 3D. *ACM Transactions on Graphics (TOG)*, 25(3):835–846, 2006. doi: 10.1145/1179352.1141964

[15] A. Torii, M. Havlena, and T. Pajdla. From Google Street View to 3D City Models. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 2188–2195. IEEE, 2009. doi: 10.1109/ICCVW.2009.5457551

[16] J. Xiao, T. Fang, P. Zhao, M. Lhuillier, and L. Quan. Image-Based Street-Side City Modeling. *ACM Transactions on Graphics (TOG)*, 28(5):114:1–114:12, 2009. doi: 10.1145/1661412.1618460