SphereFusion: Efficient Panorama Depth Estimation via Gated Fusion

Supplementary Material

1. Panorama Projections

To capture the texture features and avoid distortion and discontinuity, SphereFusion relies on two panorama projections: the equirectangular projection and the spherical mesh. In this section, we describe the details of their conversion, and how to convert the panorama image in spherical mesh to equirectangular for depth map evaluation or re-project it to 3D space for visualization.

1.1. The E2S

Based on the definition of the equirectangular projection and the spherical mesh, we define the E2S (equirectangular projection to spherical projection) module, which converts a panorama image from the equirectangular projection to the spherical projection. Given a triangle center (x, y, z) from the spherical mesh, we first calculate its position on the image plane (u, v) by Eq. 1, and then use a bi-linear sample to capture value on the image plane, finally assign corresponding value to the triangle.

$$\begin{cases} u = (1 + atan(y, x)/\pi) \times W/2 \\ v = (0.5 + atan(z, \sqrt{x^2 + y^2})/\pi) \times H \end{cases} \tag{1}$$

1.2. The S2E

Meanwhile, we can also define the S2E (spherical projection to equirectangular projection) module, which converts a panorama image from the spherical projection to the equirectangular projection. Given a pixel (u, v) on the image plane, we first calculate its position on the sphere surface, then calculate its 3D position to find out the closest triangle on the spherical mesh, and finally assign the value of the triangle to the pixel.

$$\begin{cases} x = cos(longitude)cos(latitude) \\ y = sin(longitude)cos(latitude) \\ z = sin(latitude) \end{cases} \tag{2}$$

1.3. Projection To 3D Point Cloud

For better comparison, we visualize point clouds using different methods. Although we name the result obtained from the panorama depth estimation as the depth map, it represents the distance map, which means the Euclidean distance between a 3D point and the camera center. Therefore, we can reproject a panorama depth map to 3D space by Eq. 3 after converting it into the sphere coordinates, where d is the distance.

$$\begin{cases} X = cos(latitude)cos(longitude) \times d \\ Y = cos(latitude)sin(longitude) \times d \\ Z = sin(latitude) \times d \end{cases} \tag{3}$$

2. Evaluation

To compare with other methods, we convert our depth map in the spherical domain to equirectangular projection. Following BiFuse [7] and SliceNet [5], we use five evaluation metrics, including MAE (mean average error), MRE (mean relative error), RMSE (root mean square error), RMSE(log), and δ^n. Eq. 4 shows how to calculate them, where gt is the ground truth, pr is the predicted depth, V is valid pixels, and N is the number of valid pixels. For MAE, MRE, RMSE, and RMSE(log), the smaller is better. For the δ, the bigger is better. During the evaluation, we set the depth range to $0.1 \sim 10$ meters.

$$\begin{aligned} MAE &= \sum_{i \in V} |gt_i - pr_i| \\ MRE &= \sum_{i \in V} \frac{|gt_i - pr_i|}{gt_i} \\ RMSE &= \sqrt{\frac{\sum_{i \in V} (gt_i - pr_i)^2}{N}} \\ RMSE_{log} &= \sqrt{\frac{\sum_{i \in V} (log_{10}(gt_i) - log_{10}(pr_i))^2}{N}} \\ \delta^n &= \frac{\sum_{i \in V} max(\frac{gt_i}{pr_i}, \frac{pr_i}{gt_i}) < 1.25^n}{N} \end{aligned} \tag{4}$$

3. Visulization

In this section, we add more visualization results on three datasets and compare SphereFusion (ours) with state-of-the-art methods.

On 360D [9], we compare our method with SphereDepth [8] and UniFuse [3]. Figure 1 shows depth maps and point clouds generated by different methods. Our method uses texture features extracted by the 2D encoder to enhance the mesh encoder in the spherical domain and combines the strengths of two encoders. Results show that our method captures more features from the scene and can reconstruct more details in the scene, such as doors, tables, and walls.

On Matterport3D [2] and Stanford2D3D [1], we compare our method with OmniFusion [8] and PanoFormer [3]. Figure 2 shows depth maps generated by different methods on Matterport3D and corresponding point clouds. Figure

3 shows results generated by different methods on Stanford2D3D. Compared with tangent patches, our method directly estimates the panorama depth in the spherical domain and does not need any special mechanism to fuse these patches. The results of OmniFusion have obvious patch gaps. Meanwhile, the results of PanoFormer are smoother but also lose some details. Our method reconstructs more details but suffers from imperfect ground truth [3]. Depth maps and point clouds show that our method achieves competitive results with state-of-the-art methods with a lighter network and higher efficiency.

References

[1] Iro Armeni, Sasha Sax, Amir R Zamir, and Silvio Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint arXiv:1702.01105*, 2017. 1

[2] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *arXiv preprint arXiv:1709.06158*, 2017. 1

[3] Hualie Jiang, Zhe Sheng, Siyu Zhu, Zilong Dong, and Rui Huang. Unifuse: Unidirectional fusion for 360 panorama depth estimation. *IEEE Robotics and Automation Letters*, 6 (2):1519–1526, 2021. 1, 2, 3

[4] Yuyan Li, Yuliang Guo, Zhixin Yan, Xinyu Huang, Ye Duan, and Liu Ren. Omnifusion: 360 monocular depth estimation via geometry-aware fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2801–2810, 2022. 4, 5

[5] Giovanni Pintore, Marco Agus, Eva Almansa, Jens Schneider, and Enrico Gobbetti. Slicenet: deep dense depth estimation from a single indoor panorama using a slice-based representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11536–11545, 2021. 1

[6] Zhijie Shen, Chunyu Lin, Kang Liao, Lang Nie, Zishuo Zheng, and Yao Zhao. Panoformer: Panorama transformer for indoor 360 {\deg} depth estimation. *arXiv preprint arXiv:2203.09283*, 2022. 4, 5

[7] Fu-En Wang, Yu-Hsuan Yeh, Min Sun, Wei-Chen Chiu, and Yi-Hsuan Tsai. Bifuse: Monocular 360 depth estimation via bi-projection fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 462–471, 2020. 1

[8] Qingsong Yan, Qiang Wang, Kaiyong Zhao, Bo Li, Xiaowen Chu, and Fei Deng. Spheredepth: Panorama depth estimation from spherical domain. In *2022 Tenth international conference on 3D vision (3DV)*, 2022. 1, 3

[9] Nikolaos Zioulis, Antonis Karakottas, Dimitrios Zarpalas, and Petros Daras. Omnidepth: Dense depth estimation for indoors spherical panoramas. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 448–465, 2018. 1

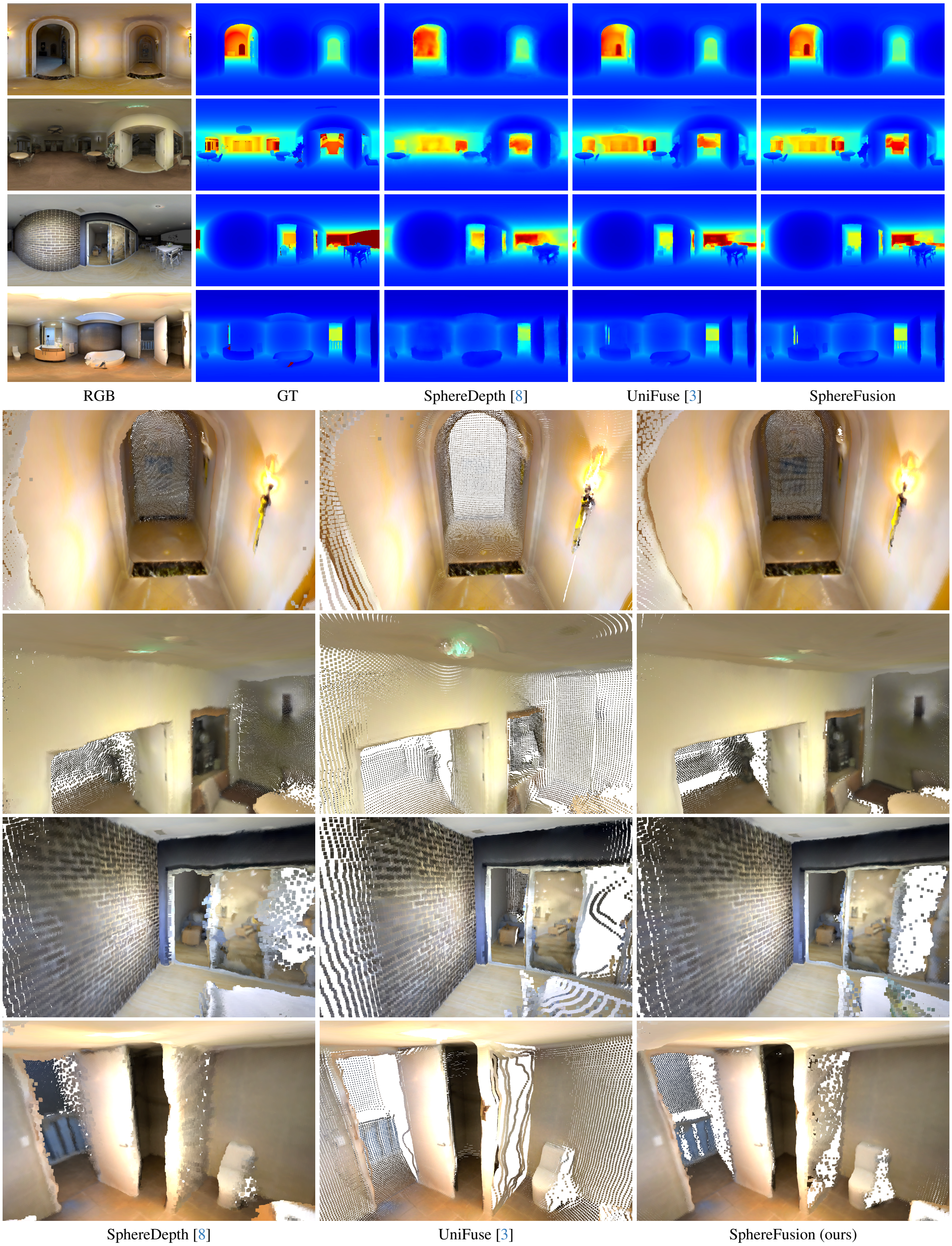


Figure 1. Depth maps and point clouds of 360D. Invalid parts of the depth map are set to red. Our method SphereFusion reconstructs more details of the scene.

3DV
#302

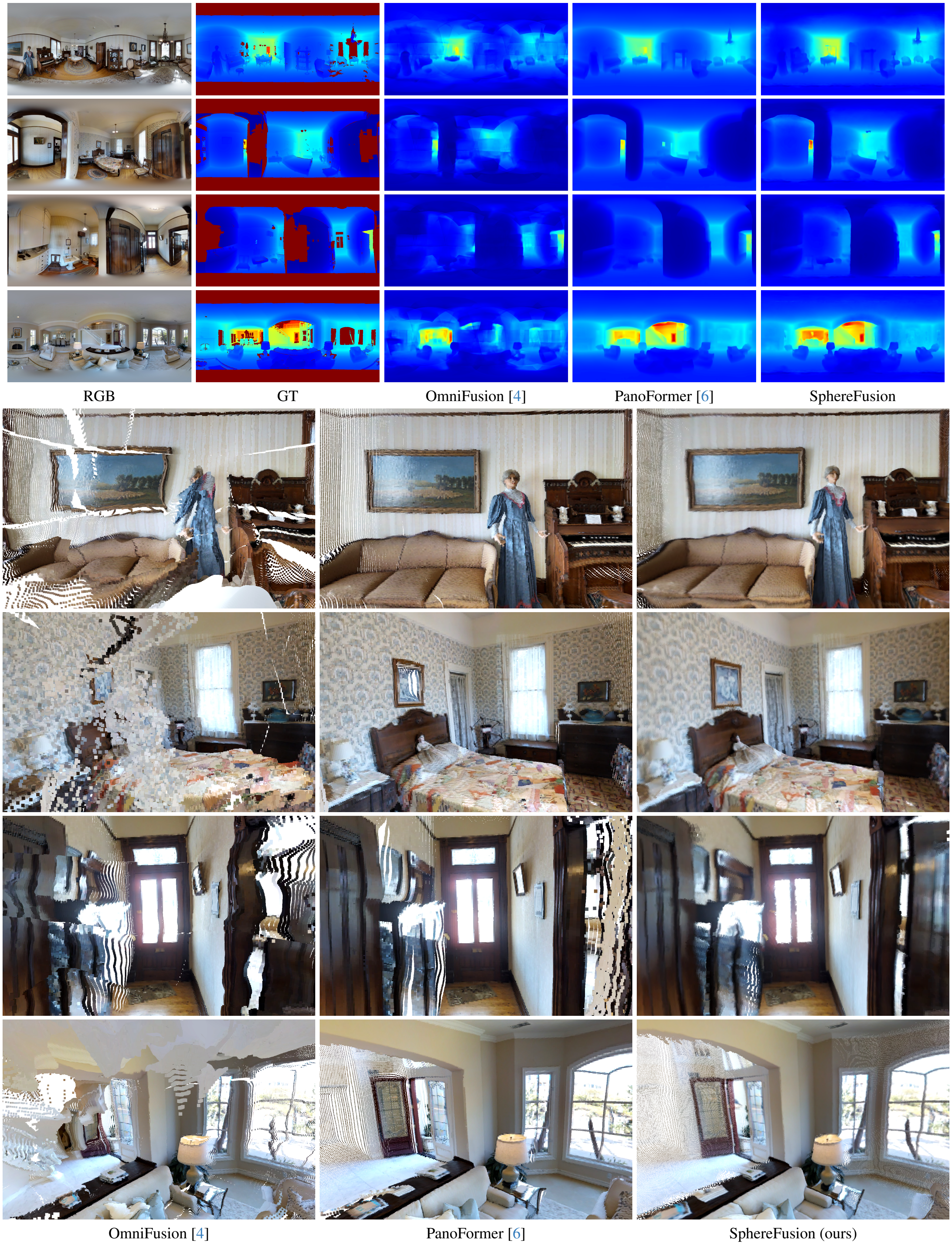
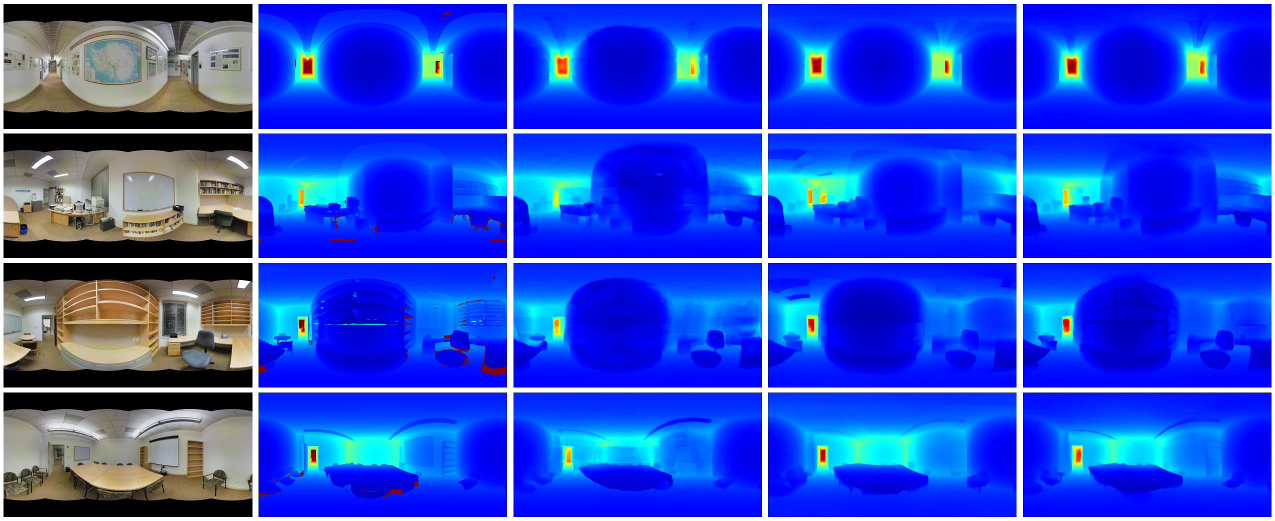


Figure 2. Depth maps and point clouds of Matterport3D. Invalid parts of the depth map are set to red. Our method has less noise and maintains the structure of the scene.



RGB | GT | OmniFusion [4] | PanoFormer [6] | SphereFusion

OmniFusion [4] | PanoFormer [6] | SphereFusion (ours)

Figure 3. Depth maps and point clouds of Stanford2D3D. Invalid parts of the depth map are set to red. Our method does not suffer from discontinuity.