

Capturing, Reconstructing, and Simulating: the UrbanScene3D Dataset (Supplementary Material)

Liqiang Lin^①, Yilin Liu^①, Yue Hu, Xinguang Yan, Ke Xie, and Hui Huang^{*①}

Shenzhen University

In this supplementary material, we offer more statistics and evaluations about UrbanScene3D, including the statistics of the volume of the UrbanScene3D benchmark (Sec. 1), the visualization of different observations of the real scenes captured by a drone (Sec. 2), more reconstruction results (Sec. 3), the statistics of reconstruction cost (Sec. 4), and the comparison of different overlaps.

1 Data Volume of the Benchmark

Table 1 shows the data volumes of each scene in the benchmark, which in total provides 16860 Gb images and 656 Gb reconstructed models with 507 Gb textures. Such abundant data not only enables the research for learning-based methods for aerial path planning and urban reconstruction, but also make it possible to evaluate both existing and future methods comprehensively. Notably, thanks to the enormous amount of captured images, our real-scene reconstructions capture more intricate details than existing datasets.

* Corresponding author

Table 1: The statistics of the data volume for each scene in the benchmark. Image(#): total image number of each scene; Image(Gb): the total volume of image sets for each scene; Recon(#): the number of provided reconstruction models; Recon(Gb): the total volume of the reconstruction meshes for each scene; Texture(Gb): the total volume of the reconstruction textures for each scene.

Scene	Image(#)	Image(Gb)	Recon(#)	Recon(Gb)	Texture(Gb)
School	14,897	3,158	25	85	108
Bridge	13,228	2,366	25	76	75
Castle	7,414	1,476	25	20	57
Town	8,948	1,661	25	47	58
Polytech	19,635	5,830	13	258	122
ArtSci	12,261	2,369	13	170	87
Total	76,383	16,860	126	656	507



Fig. 1: Four observations of the scene Shanghai with different weather conditions.

2 Different Observations

UrbanScene3D contains aerial acquisition paths generated with different aerial path planning methods, e.g., oblique photography and the methods proposed by Smith et al. [1], Zhou et al. [3], and Zhang et al. [2], resulting in different observation patterns. The real scenes were captured multiple times, providing more observations with different lighting settings. The platform with UE4 and AirSim also provides different weather effects to simulate the outdoor environment, such as rainy, snowy, foggy, daytime, or night. Fig. 1 gives an example of observations of the same scene with different weather conditions.

3 Reconstruction Results

Here we give more visualization of the reconstructed models with the four different planners. Figs. 2-5 show the reconstructed models and the corresponding reconstruction error maps on the synthetic School, the real scene Polytech, and the real scene ArtSci separately. Table 2 and Table 3 show statistics of reconstruction error of the two real scenes Polytech and ArtSci separately. When capturing the real scenes, an ICP process is performed to better evaluate the reconstruction. Even so, the results from the real scenes still have a more significant deviation than the synthetic scenes. In order to reduce the influence of noise and outliers on the evaluation results, we report the average 80% accuracy and average 90% accuracy for the real scenes in the Table 2 and Table 3.

Similar to the quantitative results, the paths generated by oblique photography produce the roughest reconstructed models. The paths generated by Smith et al. [1] and Zhou et al. [3] produce better reconstruction results.

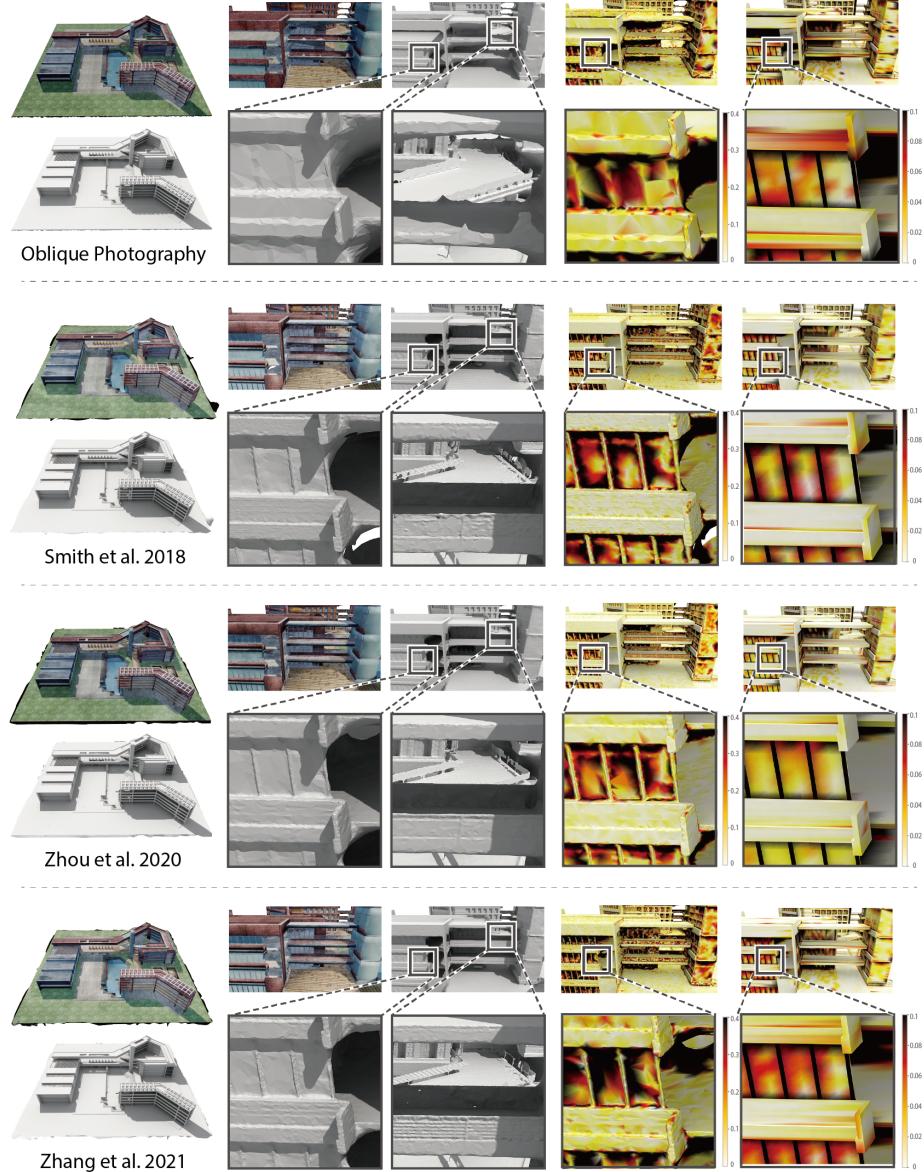


Fig. 2: The comparison of reconstruction results of the synthetic scene school and the corresponding error maps. The error maps in the fourth column show the computed accuracy errors on the reconstructed meshes, and the maps in the last column show the corresponding completeness errors on the ground-truth mesh.

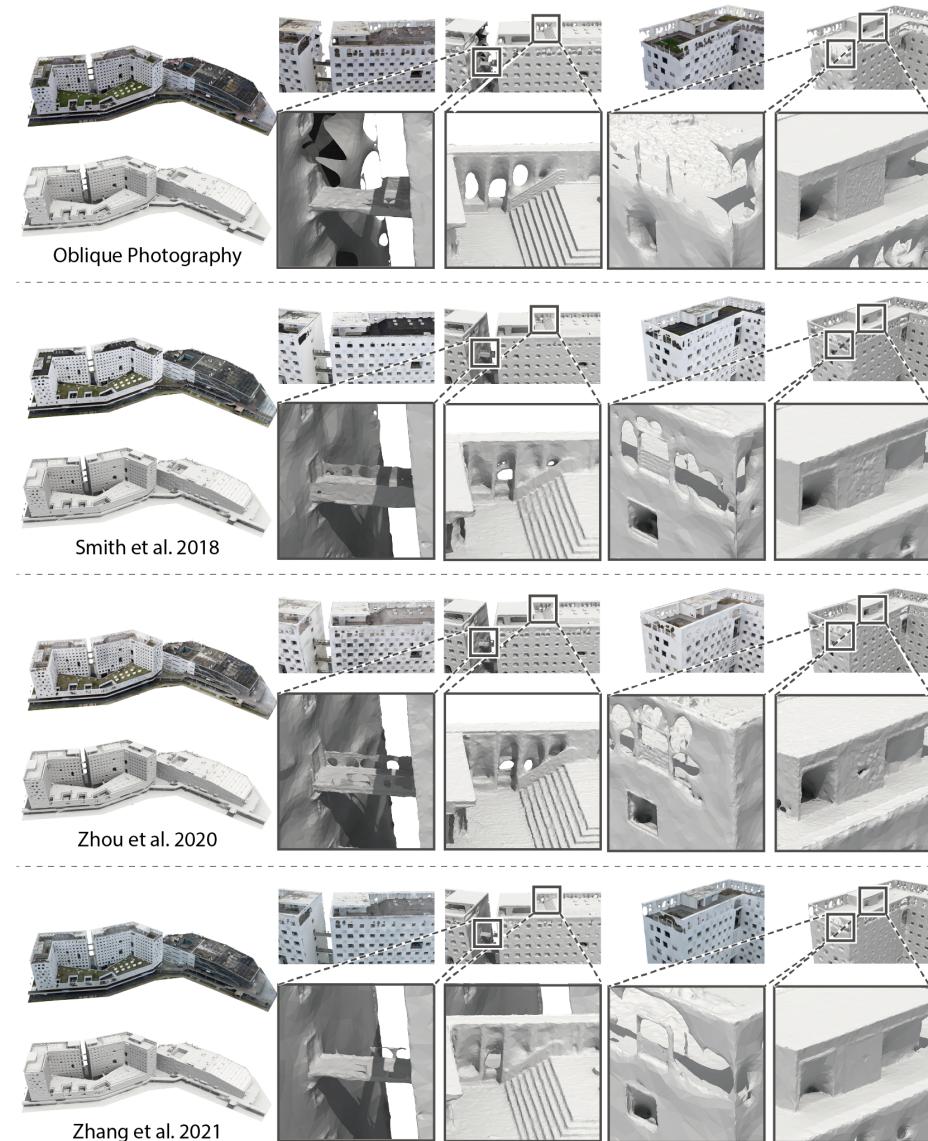


Fig. 3: The comparison of the reconstruction results of the real scene Polytech.

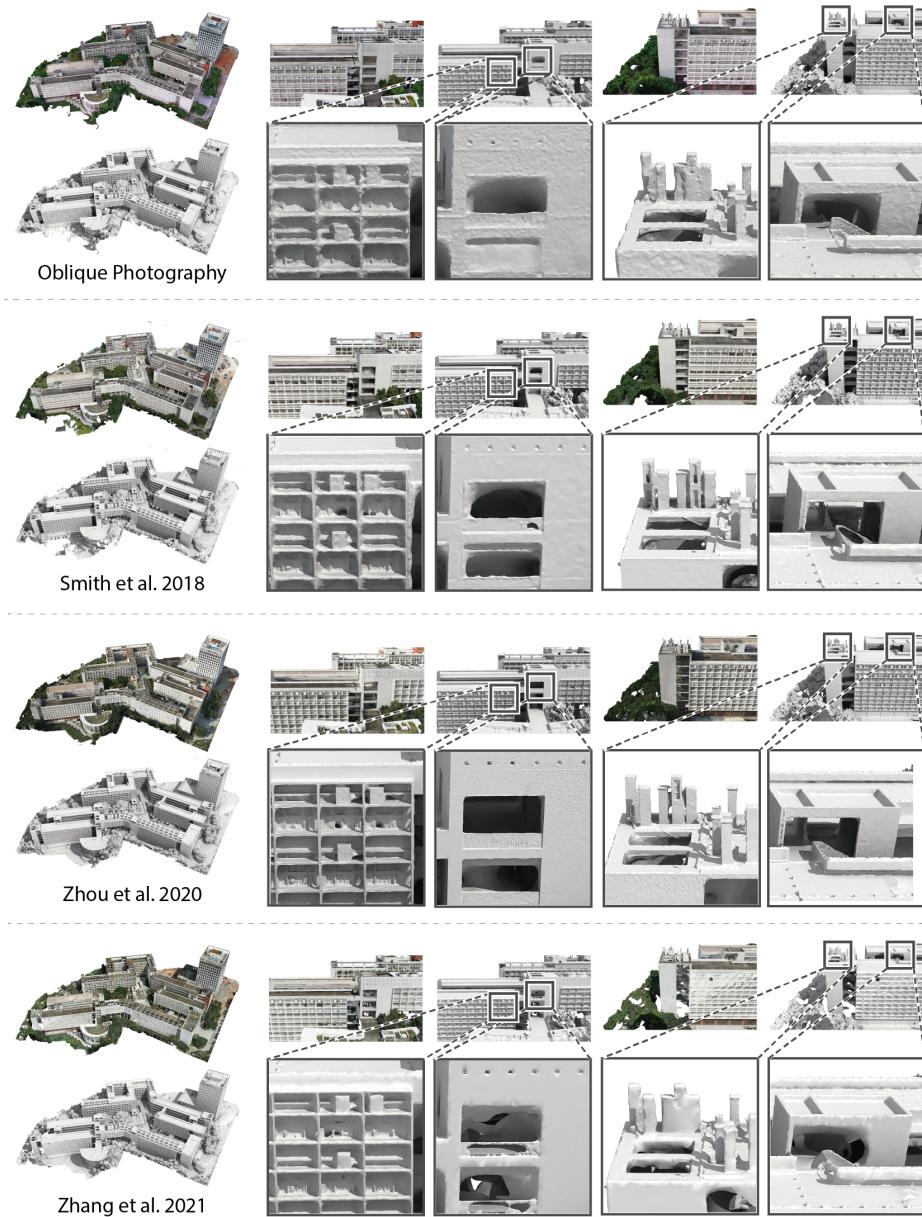


Fig. 4: The comparison of the reconstruction results of the real scene ArtSci.



Fig. 5: The visualization of the accuracy maps and the completeness maps on the two real scenes Polytech and ArtSci. First row: accuracy maps for Polytech; Second row: completeness maps for Polytech; Third row: accuracy maps for ArtSci; Fourth row: completeness maps for ArtSci.

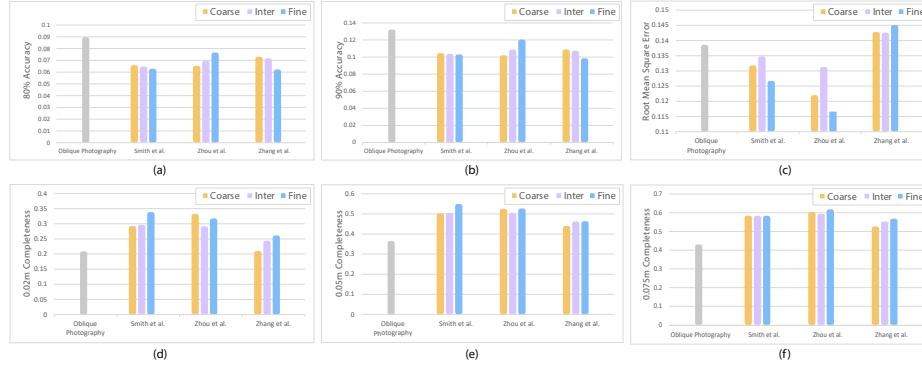


Fig. 6: Reconstruction error of different methods with different proxies on the real scene Polytech. (a): 80% average accuracy ($m \downarrow$); (b): 90% average accuracy ($m \downarrow$); (c): root mean square error ($m \downarrow$); (d): 0.02m completeness (%) \uparrow ; (e): 0.05m completeness (%) \uparrow ; (f): 0.075m completeness (%) \uparrow .

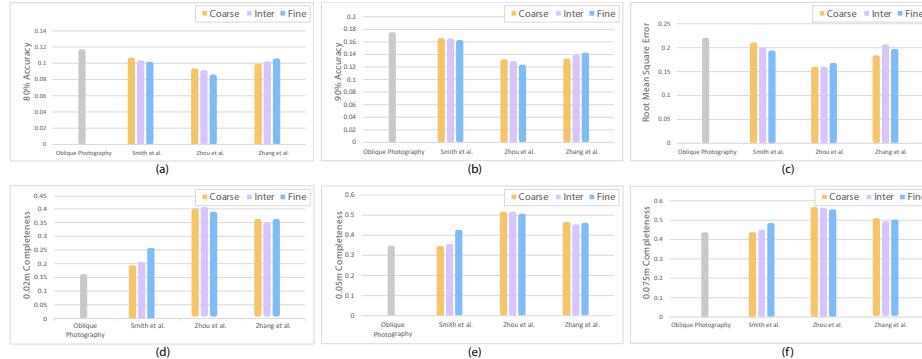


Fig. 7: Reconstruction error of different methods with different proxies on the real scene ArtSci. (a): 80% average accuracy ($m \downarrow$); (b): 90% average accuracy ($m \downarrow$); (c): root mean square error ($m \downarrow$); (d): 0.02m completeness (%) \uparrow ; (e): 0.05m completeness (%) \uparrow ; (f): 0.075m completeness (%) \uparrow .

4 Reconstruction Cost

Fig. 8 shows the time consumption of aerotriangulation and reconstruction for the four different planners. The time consumption of aerotriangulation and reconstruction is related to the number of input images but are not proportional to the number of images. As we can see in Fig. 8, the path generated by oblique photography captures the most images but does not result in the longest aerotriangulation and reconstruction time. The path generated by Smith et al. [1] captures a bit more images than the methods proposed by Zhou et al. [3] and Zhang et al. [2] and consume more aerotriangulation time and reconstruction time. The path generated by Zhang et al. [2] captures the comparable num-

Table 2: Reconstruction error of different methods with different proxies on the real scenes Polytech. Proxy: different proxies used for path planning (#); Image: the number of captured images (#); AE 80%: average 80% reconstruction accuracy (mm); AE 90%: average 90% reconstruction accuracy (mm); Comp 0.02m: 0.02m completeness of reconstruction models (%); Comp 0.05m: 0.05m completeness of reconstruction models (%); Comp 0.075m: 0.075m completeness of reconstruction models (%).

methods	Proxy	Image	AE	AE	RSME	Comp 0.02m	Comp 0.05m	Comp 0.075m
			80%	90%				
Oblique photography	-	2510	0.0896	0.132	0.139	21.90	36.54	43.07
Smith et al. [1]	Coarse	777	0.0659	0.105	0.132	29.25	50.29	58.73
	Inter	678	0.0646	0.104	0.135	29.66	50.52	58.77
	Fine	685	0.0629	0.103	0.127	33.84	54.96	58.73
Zhou et al. [3]	Coarse	1410	0.0653	0.102	0.122	33.89	52.56	60.11
	Inter	1316	0.0697	0.109	0.132	29.92	50.71	59.59
	Fine	1238	0.0766	0.121	0.117	32.05	52.90	61.81
Zhang et al [2]	Coarse	764	0.0730	0.109	0.143	21.00	44.01	53.58
	Inter	813	0.0718	0.108	0.137	24.73	44.79	54.38
	Fine	770	0.0622	0.098	0.145	26.13	46.39	56.11

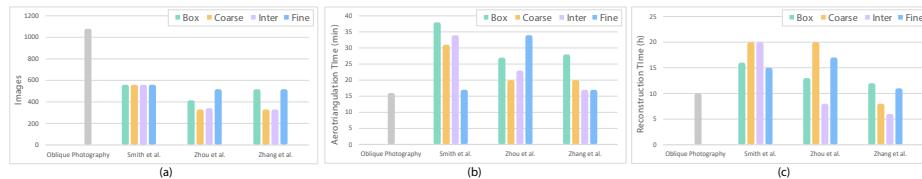


Fig. 8: The statistics of the reconstruction time for the four different planners on the synthetic scene School with 70% overlap. (a): The number of captured images for reconstruction; (b): The time consumption of aerotriangulation; (c): The time consumption of reconstruction.

ber of images with the path generated by Zhou et al. [3], but consumes less aerotriangulation and reconstruction time.

5 Different Overlaps

Fig. 9 shows the reconstruction error of the reconstruction results using the four different planners with different overlaps on the synthetic scene school. Higher overlaps means more sample points on the proxy, thus preserving more geometry details of the proxy. For the completeness of the reconstructed models, all four planners obtain higher completeness when the overlaps go larger. When it comes to the accuracy of the reconstructed models, only the method proposed by Zhou et al. [3] achieves higher 90% accuracy, 95% accuracy, and less RMSE when using

Table 3: Reconstruction error of different methods with different proxies on the real scenes ArtSci. Proxy: different proxies used for path planning (#); Image: the number of captured images (#); AE 80%: average 80% reconstruction accuracy (mm); AE 90%: average 90% reconstruction accuracy (mm); Comp 0.02m: 0.02m completeness of reconstruction models (%); Comp 0.05m: 0.05m completeness of reconstruction models (%); Comp 0.075m: 0.075m completeness of reconstruction models (%).

methods	Proxy	Image	AE	AE	RMSE	Comp	Comp	Comp
			80%	90%				
Oblique photography	-	3620	0.117	0.175	0.220	17.40	35.62	43.84
Smith et al. [1]	Coarse	2357	0.107	0.166	0.211	19.22	35.29	42.95
	Inter	2471	0.104	0.165	0.200	20.06	36.13	44.10
	Fine	2900	0.102	0.163	0.194	26.93	42.80	48.11
Zhou et al. [3]	Coarse	966	0.094	0.132	0.160	40.78	52.60	56.80
	Inter	1202	0.092	0.130	0.160	41.58	52.32	56.41
	Fine	1237	0.086	0.124	0.168	38.80	51.20	55.63
Zhang et al [2]	Coarse	1311	0.100	0.133	0.184	36.49	47.64	51.71
	Inter	1314	0.103	0.139	0.207	35.95	45.94	49.75
	Fine	1319	0.106	0.143	0.198	36.69	46.58	50.28

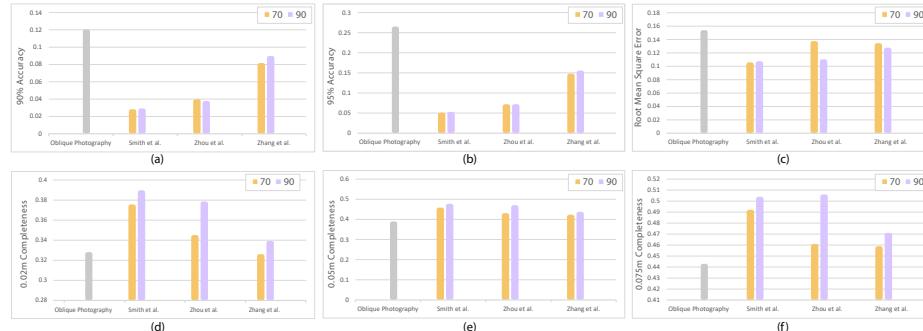


Fig. 9: Reconstruction error of different methods with different overlaps on the synthetic scene School with inter proxy. (a): 90% accuracy (m) ↓; (b): 95% Accuracy (m) ↓; (c): root mean square error (m) ↓; (d): 0.02m completeness (%) ↑; (e): 0.05m completeness (%) ↑; (f): 0.075m completeness (%) ↑.

larger overlap, which means the method proposed by Zhou et al. [3] reflects the geometry characteristics of the proxy better.

Note that using a higher overlap works similarly with using a finer proxy, since they both provide a scene prior with more geometry details. Nonetheless, even with a very high overlap, the sampled points cannot present the details of the proxy completely, which remains as a problem to be solved.

Acknowledgements. This work was supported in parts by NSFC (62161146005, U21B2023, U2001206), GD Talent Program (2019JC05X328), DEGP Key Project (2018KZDXM058, 2020SFKC059), Shenzhen Science and Technology Program (RCJC20200714114435012, JCYJ20210324120213036), and Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ).

References

1. Smith, N., Moehrle, N., Goesele, M., Heidrich, W.: Aerial path planning for urban scene reconstruction: a continuous optimization method and benchmark. ACM Trans. on Graphics (Proc. SIGGRAPH Asia) pp. 183:1–183:15 (2018)
2. Zhang, H., Yao, Y., Xie, K., Fu, C.W., Zhang, H., Huang, H.: Continuous aerial path planning for 3D urban scene reconstruction. ACM Trans. on Graphics (Proc. SIGGRAPH Asia) pp. 225:1–225:15 (2021)
3. Zhou, X., Xie, K., Huang, K., Liu, Y., Zhou, Y., Gong, M., Huang, H.: Offsite aerial path planning for efficient urban scene reconstruction. ACM Trans. on Graphics (Proc. SIGGRAPH Asia) pp. 192:1–192:16 (2020)