— Supplementary Material —
Multiway Point Cloud Mosaicking with Diffusion and Global Optimization

Shengze Jin[1] Iro Armeni[2] Marc Pollefeys[1,3] Dániel Baráth[1]
[1] Department of Computer Science, ETH Zurich, Switzerland
[2] Department of Civil and Environmental Engineering, Stanford University
[3] Microsoft Mixed Reality & AI Lab, Zurich, Switzerland

Abstract

The supplementary material contains:
(I) Visualizations for pairwise registration on the 3DMatch and 3DLoMatch datasets,
(II) Visualizations for multiway registration on the NSS dataset,
(III) Registration recall for multiway registration on the four datasets,
(IV) Ablation study on pairwise registration results on the NSS dataset,
(V) The run-times of the methods, and
(VI) A video that gives a summary of our method and results.

1. Visualizations of Pairwise Registration

In Figures 1 and 2, we show pairwise registration results on the *3DLoMatch* and *3DMatch* datasets, respectively. We do not show pairwise results on NSS and KITTI since: (i) the NSS point clouds represent spaces with ceiling information and lack details – such pairwise registration results are hard to interpret even with the ceilings are cut off; and (ii) the results on KITTI are quite saturated with all methods achieving good results. While our ODIN achieves the most accurate registrations (as per Table 1 in the main paper), no significant difference is visible in the pairwise visualizations. We show the results of the three best matchers (according to Table 1 in the main paper), namely ODIN, PEAL [14] and GeoTransformer [8].

1.1. Visualizations on 3DMatch

In Figure 1, we show examples of point cloud registration on the *3DMatch* dataset. We also report the RMSE for all results, which we use to determine if two point clouds are correctly registered. Specifically, per row:

Row (1): This example showcases a particularly challenging pair with a very low overlap in the point clouds. While all methods manage to estimate the correct pose coarsely, both GeoTransformer and PEAL achieve high RMSE. The output of our ODIN is close to the ground truth transformation, with an RMSE that is substantially lower than its competitors.

Row (2): In this case, GeoTransformer fails to find a correct pose, even coarsely. Similar to row (1), PEAL manages to output an acceptable transformation. However, it has a high RMSE, that is far from the ground truth. The registration output of ODIN is almost an order of magnitude more accurate than that of PEAL in terms of RMSE, and, visually, it is very close to the ground truth registration. This success underscores the efficacy of our dual-stream architecture combined with the attention mechanism, which directs the network's focus towards regions of high confidence for more dependable correspondence inference. Additionally, the diffusion model plays a crucial role in eliminating noisy matches, further enhancing the overall precision.

Row (3): In this example, both GeoTransformer and PEAL fail. ODIN has a higher RMSE than in the above examples, however, the registration output is visually acceptable and closer to the ground truth.

While the examples show that there is still room for improvement, ODIN clearly achieves substantially better registrations than the state of the art in 3D Match.

1.2. Visualizations on 3DLoMatch

In Figure 2, we show examples of point cloud registration on the *3DLoMatch* dataset. Similar to 3DMatch, we report the RMSE for all results. Specifically, per row:

Row (1): In this example, we observe that ODIN recovers the pose very accurately, while both GeoTransformer and PEAL fail entirely. Their RMSE is two orders of magnitude higher than that of ODIN. This again highlights the importance of the proposed two-stream architecture and the diffusion-based denoising.

Row (2): Here, ODIN provides a close-to-GT pose. GeoTransformer and PEAL struggle to find a good pose.

Method	NSS				3DMatch			3DLoMatch			KITTI		
	RR (%)↑	RE (°)↓	TE (m)↓		RR (%)↑	RE (°)↓	TE (cm)↓	RR (%)↑	RE (°)↓	TE (cm)↓	RR (%)↑	RE (°)↓	TE (cm)↓
Predator	64.6	13.43	0.65	PEAL	94.1	4.72	15.8	78.8	16.03	50.2	75.7	9.46	11.85
+ Open3d [3]	51.3	12.76	0.64	+ Open3d	81.8	4.72	15.8	68.9	14.23	45.1	83.2	6.21	7.72
+ DeepMapping2 [2]	60.1	11.54	0.64	+ DeepM.	82.7	4.23	14.5	70.1	13.25	39.4	91.5	3.34	6.04
+ LMPR [6]	65.1	11.35	0.62	+ LMPR	82.4	3.98	12.6	70.7	13.07	37.3	81.1	6.79	7.89
+ LIRTS [13]	65.9	11.42	0.61	+ LIRTS	86.9	3.95	12.0	76.2	11.52	36.0	84.7	5.17	6.94
+ RMPR [12]	66.9	10.87	0.62	+ RMPR	95.9	3.57	11.6	83.1	10.18	34.4	86.1	4.69	6.38
+ **Wednesday**	75.6	2.24	0.51	+ **Wednesday**	96.8	2.58	9.4	86.4	7.21	29.1	94.6	2.52	5.92
ODIN + Wednesday	**78.3**	**2.01**	**0.42**		**97.3**	**2.32**	**8.4**	**87.1**	**6.44**	**26.5**	**96.2**	**2.18**	**4.76**

Table 1. **Multiway point cloud registration** on the NSS [11], 3DMatch [15], 3DLoMatch [7] and KITTI [5] datasets. The reported metrics are the registration recall (RR), average rotation (RE) and translation errors (TE). For each dataset, we choose the best-performing pairwise estimator from the baselines. We run Predator [7] on NSS and PEAL [14] on the other datasets. The best results are in **bold**.

Method	All spatiotemporal pairs			Only same-stage pairs			Only different-stage pairs		
	RR (%)↑	RTE (m)↓	RRE (°)↓	RR (%)↑	RTE (m)↓	RRE (°)↓	RR (%)↑	RTE (m)↓	RRE (°)↓
FPFH [10]	11.70	2.23	45.32	30.82	2.42	29.35	0.42	4.06	78.01
FCGF [4]	24.43	2.04	39.89	42.86	2.23	32.12	10.52	3.23	53.24
D3Feat [1]	22.73	2.26	33.09	36.51	2.05	27.22	4.76	2.53	40.76
Predator [7]	64.97	0.65	13.52	92.99	0.27	4.83	28.42	1.16	24.85
GeoTransformer [9]	39.07	0.99	22.93	55.59	0.73	17.02	17.51	1.34	30.62
PEAL [14]	58.72	0.71	15.78	88.63	0.32	5.32	19.71	1.22	29.42
ODIN	**69.73**	**0.54**	**11.96**	**95.46**	**0.21**	**4.36**	**36.17**	**0.97**	**21.87**

Table 2. **Pairwise point cloud registration** on the NSS dataset. The reported metrics are the Registration Recall (RR), which measures the fraction of successfully registered pairs; the Relative Rotation Error (RRE); and the Relative Translation Error (RTE). We show ablation results for same-stage and different-stage pairs. The best results are in **bold**.

Row (3): In this example, all methods fail to recover a good pose. However, ODIN still manages a significantly lower RMSE than the other methods. In addition, visually, the output is not far from the ground truth alignment.

2. Visualizations of Multiway Registration

In Figure 3, we show examples of point cloud multiway registration for the *NSS* dataset. We choose to visualize NSS as it is the most challenging dataset. We show results of our proposed method and [12, 13]. We choose [12, 13] as they are – after ours – the next best-performing methods as per Table 2 in the main paper. Specifically, per row:

Rows (1), (2), (3) and (4): LIRST [13] and RMPR [12] fail to achieve an acceptable global registration in these examples. Their outputs are incomprehensible and are far from the expected results. Such outputs are frequent for these methods on this dataset. While our proposed method has inaccuracies, it provides substantially more accurate registrations that are not far from the ground truth. This highlights that the proposed multiway registration pipeline is more robust to such complicated scenarios than the state of the art.

Row (5): In this example, all methods fail to achieve a good registration. As before, both LIRST and RMPR results are incomprehensible. Our method manages to find the structure coarsely, however, there are mistakes, showing that there is still room for improvement.

3. Multiway Registration Recall

We provide the registration recall (RR) for multiway registration on the NSS [11], 3DMatch [15], 3DLoMatch [7], and KITTI [5] datasets in Table 1. RE and TE are taken from Table 2 in the main paper. The successfully registered pairs are defined following the protocol from [9, 11, 12]. For each dataset, we choose the best-performing pairwise estimator from the baselines. We run Predator [7] on NSS and PEAL [14] on the other datasets.

Our proposed **Wednesday** (without ODIN) consistently improves upon *all* state-of-the-art algorithms and gains 0.9% to 8.7% in RR compared to the second-best method on the four datasets. Our full pipeline, **ODIN** + **Wednesday**, achieves additional improvements on RR on all datasets.

4. Pairwise Ablation on the NSS Dataset

We show an ablation of the pairwise point cloud registration results on the NSS dataset in Table 2. The reported metrics are the Registration Recall (RR), which measures the fraction of successfully registered pairs; the Relative Rotation Error (RRE) in degrees (°); and the Relative Translation Error (RTE) in meters (m). Specifically, we ablate the results for same-stage pairs and different-stage pairs as defined in the original paper [11], which evaluates independently the performance of pairs of point clouds from the same (w/o change) or different (w/ change) temporal stages.

The first three columns (first block) show the results on all pairs regardless of being from same or different stages. This is the same as in Table 1 in the main paper. For the same-stage pairs (second block), the results improve for all methods compared to the previous case and follow the same trend in the order of performance. However, only three algorithms provide very high performance (above 88%; Predator, PEAL, and ODIN), with ODIN being the most accurate. The rest follow at below 55% of performance. On the different-stage pairs, there is a substantial difference (*i.e.*, 7.7% RR) between ODIN and the second best method, Predator. While ODIN is significantly better than all competitors, it is important to note that its RR on the different-stage pairs is still far from 100%. This highlights that further improvements are needed to robustly solve such complicated scenarios exhibiting temporal changes.

We find it interesting that, while PEAL falls short compared to Predator, our ODIN significantly outperforms both, while building on similar architectural blocks as PEAL. This demonstrates the importance of the proposed two-stream attention learning architecture coupled with the diffusion denoising module. PEAL's effectiveness heavily relies on the initial pose provided by the GeoTransformer. It struggles to correct this initial pose if it is too inaccurate. In contrast, our method does not rely on an initial pose and, thus, it identifies correct correspondences more robustly. It effectively filters out erroneous correspondences, retaining only those with high confidence.

5. Processing Times

We evaluate the runtime on a computer with Intel(R) Xeon(R) CPU E3-1284L v4 @ 2.90GHz and GeForce RTX 3090 GPU. In Table 3, we provide the total and average times in seconds of pairwise registration methods on the 3DMatch dataset. The total time represents the cumulative runtime for pairwise registration across the entire scene, while the average time denotes the mean duration expended for each individual pair. The results show that the proposed ODIN runs at a similar speed to its less accurate alternatives. Specifically, it is marginally slower than GeoTransformer and PEAL, and it is twice as fast as Predator.

In Table 4, we provide the total and average times of multiway registration methods on the 3DMatch dataset. For this experiment, we compare using the same methods as those listed in Table 2 of the main paper. The proposed method, Wednesday, falls in the middle in terms of runtime.

In conclusion, there is no trade-off when using the proposed ODIN and Wednesday. They obtain state-of-the-art results while running at a similar speed as the baselines.

Method	Total Time (s)	Average Time (s)
Predator	460	0.26
GeoTr.	159	0.09
PEAL	212	0.12
ODIN	248	0.14

Table 3. Total and average time of pairwise point cloud registration pipelines on the 3DMatch dataset.

Method	Total Time (s)	Average Time (s)
PEAL	212	0.12
PEAL + Open3d	283	0.16
PEAL + DeepMapping2	7399	4.18
PEAL + LMPR	301	0.17
PEAL + LIRTS	425	0.24
PEAL + RMPR	244	0.14
PEAL + Wednesday	389	0.22
ODIN + Wednesday	425	0.24

Table 4. Total time and average time per point cloud pair of multiway point cloud registration pipelines on the 3DMatch dataset.

References

[1] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In *CVPR*, 2020. 2

[2] Chao Chen, Xinhao Liu, Yiming Li, Li Ding, and Chen Feng. Deepmapping2: Self-supervised large-scale lidar map optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9306–9316, 2023. 2

[3] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust reconstruction of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5556–5565, 2015. 2

[4] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *ICCV*, 2019. 2

[5] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012. 2

[6] Zan Gojcic, Caifa Zhou, Jan D Wegner, Leonidas J Guibas, and Tolga Birdal. Learning multiview 3d point cloud registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1759–1769, 2020. 2

[7] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4267–4276, 2021. 2, 6

[8] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, 2022. 1

[9] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, Slobodan Ilic, Dewen Hu, and Kai Xu. Geotransformer: Fast and robust point cloud registration with geometric transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[10] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (FPFH) for 3D registration. In *ICRA*, 2009. 2

[11] Tao Sun, Yan Hao, Shengyu Huang, Silvio Savarese, Konrad Schindler, Marc Pollefeys, and Iro Armeni. Nothing stands still: A spatiotemporal benchmark on 3d point cloud registration under large geometric and temporal change, 2023. 2, 7

[12] Haiping Wang, Yuan Liu, Zhen Dong, Yulan Guo, Yu-Shen Liu, Wenping Wang, and Bisheng Yang. Robust multiview point cloud registration with reliable pose graph initialization and history reweighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9506–9515, 2023. 2

[13] Zi Jian Yew and Gim Hee Lee. Learning iterative robust transformation synchronization. In *2021 International Conference on 3D Vision (3DV)*, pages 1206–1215. IEEE, 2021. 2

[14] Junle Yu, Luwei Ren, Yu Zhang, Wenhui Zhou, Lili Lin, and Guojun Dai. Peal: Prior-embedded explicit attention learning for low-overlap point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17702–17711, 2023. 1, 2

[15] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1802–1811, 2017. 2, 5

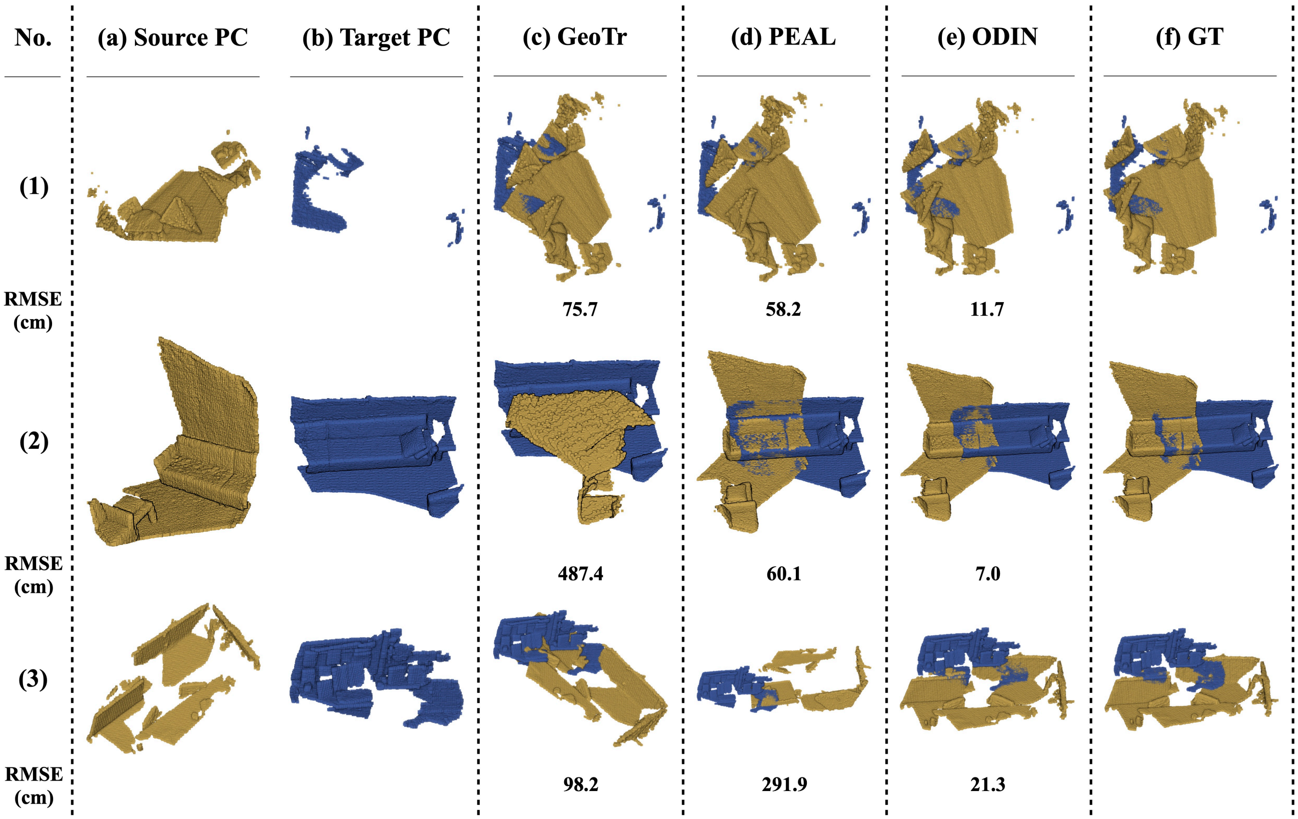

Figure 1. **Qualitative Results for the *3DMatch [15]* dataset.** See Section 1.1 for an explanation of the results. *Best viewed in screen.*

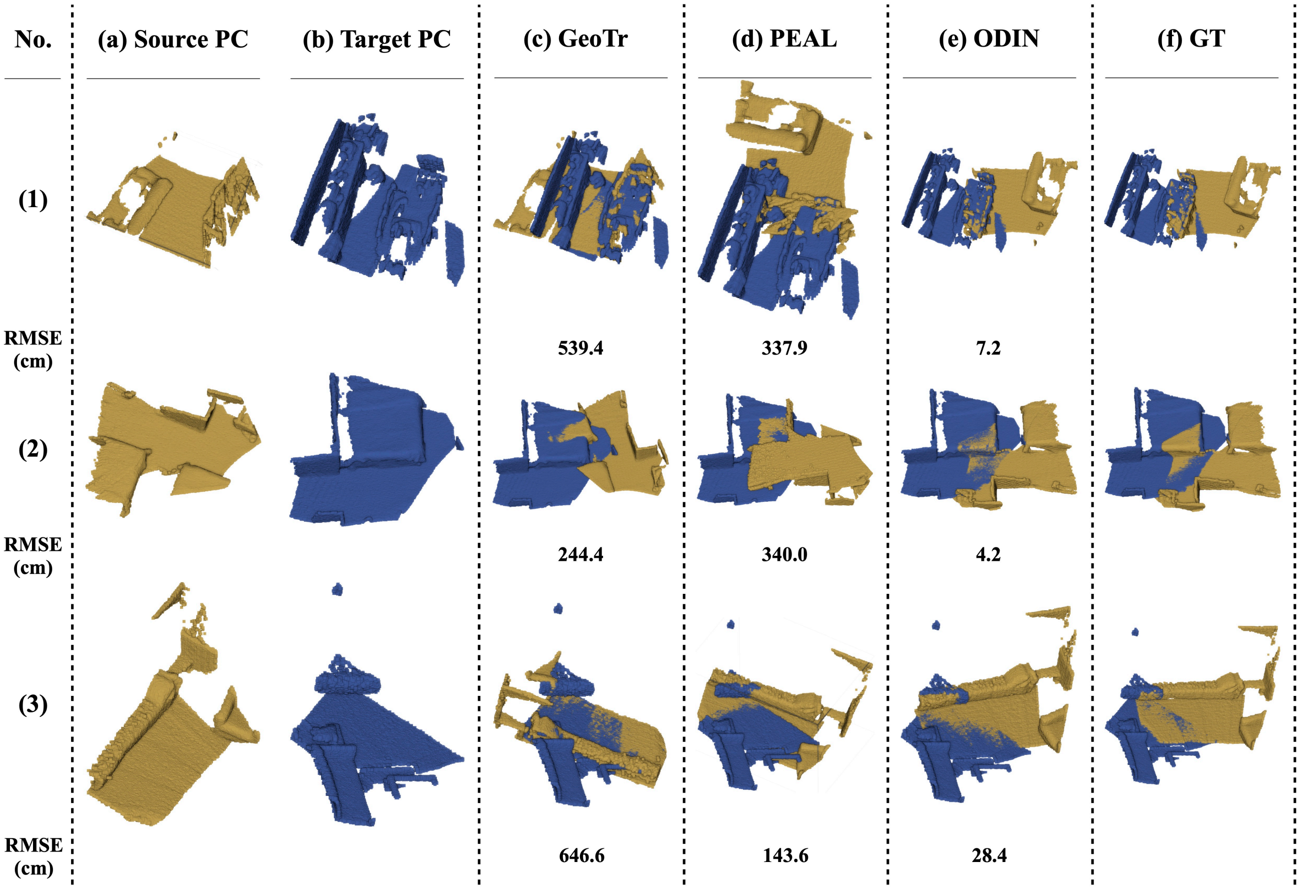

Figure 2. **Qualitative Results for the *3DLoMatch [7]* dataset.** See Section 1.2 for an explanation of the results. *Best viewed in screen.*

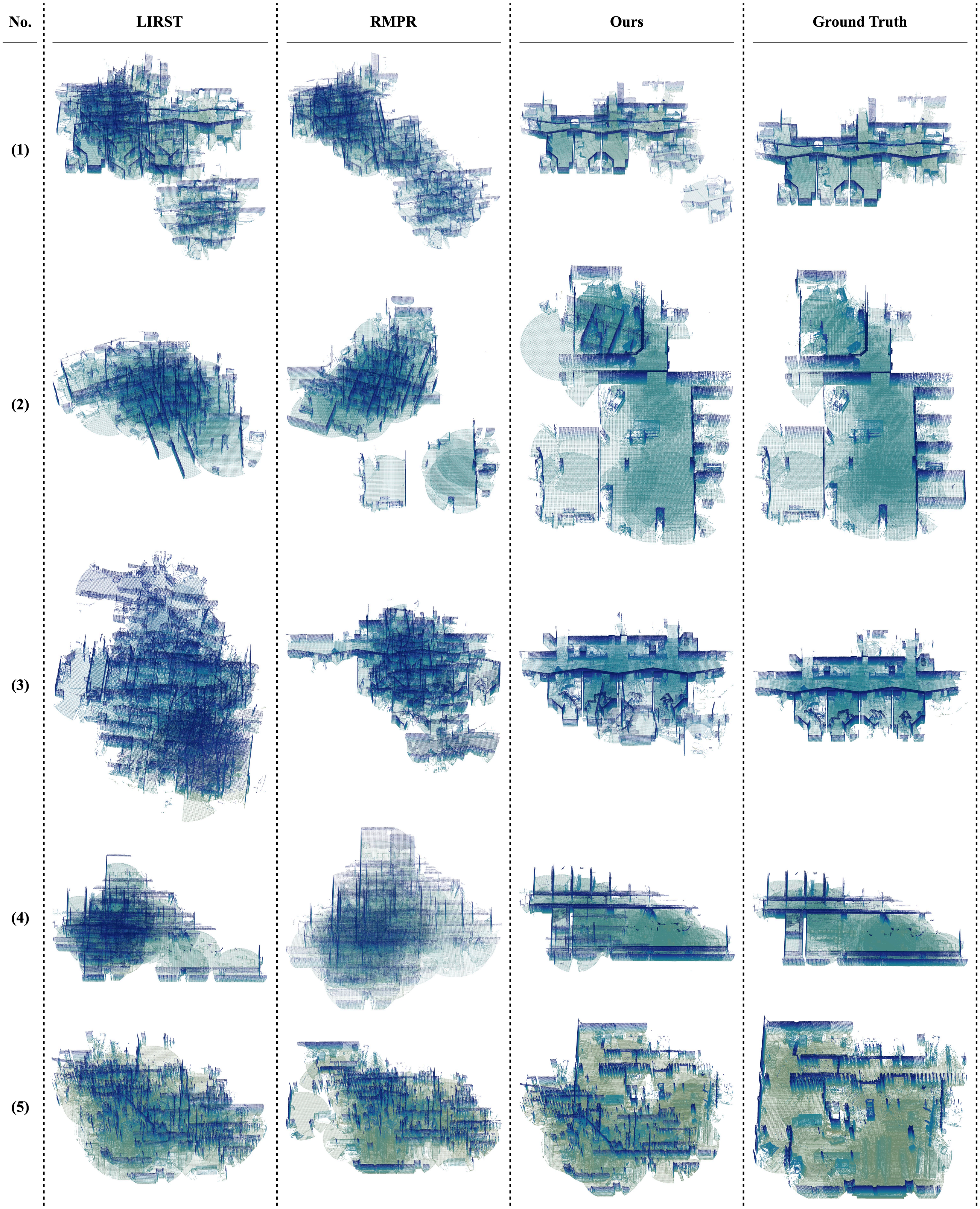

Figure 3. **Qualitative Results for the** ***NSS [11]*** **dataset.** See Section 2 for an explanation of the results. *Best viewed in screen.*