PCSTracker: Long-Term Scene Flow Estimation for Point Cloud Sequences

Supplementary Material

6. Number of Iteration

We vary the number of iterative refinement steps to evaluate their impact on overall performance. As shown in Tab. 7, the model achieves its lowest EPE_{3D} on the PointOdyssey3D dataset with K=6 iterations. In practice, we use K=4 as the default setting, which offers a favorable trade-off between accuracy and computational efficiency.

Dataset	Num. Iteration	$EPE_{3D}\downarrow$	$\delta_{3D}^{avg}\uparrow$	$Survival_{3D}^{0.50}\uparrow$	$MAE_{3D}\downarrow$
POD3D	1	0.171	81.16	91.75	0.167
	2	0.136	86.09	94.41	0.130
	4	0.133	86.37	93.65	0.129
	6	**0.124**	**87.70**	**95.05**	**0.118**
	8	0.126	87.52	95.02	0.120
	12	0.132	86.64	94.93	0.125
ADT3D	1	0.482	65.07	79.87	0.315
	2	0.402	71.28	84.66	0.255
	4	0.372	74.44	87.74	0.226
	6	0.370	**74.72**	88.42	**0.220**
	8	**0.369**	74.16	**88.89**	0.221
	12	0.381	71.08	88.15	0.232

Table 7. **Comparison results on PointOdyssey3D (POD3D) and ADT3D datasets with different iteration numbers.**

7. Impact of Point-Voxel Correlation

We evaluate the effectiveness of the point–voxel dual-branch correlation pyramid by individually removing each branch, as shown in Tab. 9. Removing the voxel correlation branch increases EPE_{3D} by 73.5%, whereas removing the point correlation branch results in an 11.3% increase.

8. Impact of the Number of Query Points

Tab. 8 summarizes the effect of varying the number of query points on the final performance. Experiments on both PointOdyssey3D and ADT3D show that accuracy is lowest when tracking a single query point. Increasing the number of query points from 1 to 32 results in a substantial improvement, reducing EPE_{3D} by 76.9% and 53.9%, respectively. Further increasing the number of query points continues to provide stable performance improvements. These results suggest that jointly estimating denser point sets leads to more stable and accurate motion predictions.

9. Robust for Occlusion

We compare the performance of PCSTracker with the SF-baseline under varying occlusion levels to evaluate their robustness to occlusions. As shown in Tab. 10, our method consistently achieves substantially lower EPE_{3D} on both POD3D and ADT3D across all occlusion levels, indicating significantly improved robustness. In addition, leveraging a broad temporal context not only benefits occluded points but also improves the accuracy of non-occluded points.

Dataset	Num. Query	$EPE_{3D}\downarrow$	$\delta_{3D}^{avg}\uparrow$	$Survival_{3D}^{0.50}\uparrow$	$MAE_{3D}\downarrow$
POD3D	1	0.668	56.04	67.66	0.619
	32	0.154	83.13	93.72	0.148
	64	0.143	85.24	93.99	0.137
	256	0.142	84.84	94.01	0.138
	512	0.139	85.16	**94.34**	0.135
	1024	**0.133**	**86.37**	93.65	**0.129**
ADT3D	1	1.279	37.86	53.02	0.969
	32	0.589	59.94	75.08	0.394
	64	0.545	63.20	78.11	0.363
	256	0.501	66.25	80.81	0.327
	512	0.489	66.43	80.96	0.318
	All	**0.372**	**74.44**	**87.74**	**0.226**

Table 8. **Comparison results on PointOdyssey3D (POD3D) and ADT3D datasets with different numbers of query points.**

Correlation Volume	$EPE_{3D}\downarrow$	$\delta_{3D}^{avg}\uparrow$	$Survival_{3D}^{0.50}\uparrow$	$MAE_{3D}\downarrow$
Point-only	0.501	51.07	65.16	0.510
Voxel-only	0.150	84.95	93.22	0.145
Point-Voxel	**0.133**	**86.37**	**93.65**	**0.129**

Table 9. **Impact of point-voxel correlation**

Dataset	Method	Number of Occlusion Frames			
		0	2	4	>4
POD3D	SF-baseline	0.1950	0.2421	0.2714	0.3511
	Ours	**0.0754**	**0.1034**	**0.1200**	**0.1736**
ADT3D	SF-baseline	0.3404	0.4188	0.4690	0.5233
	Ours	**0.0961**	**0.1228**	**0.1303**	**0.1464**

Table 10. **Comparison performance under different occlusion levels.** POD3D denotes the PointOdyssey3D dataset.

10. Discussion of Limitation

Although our method effectively estimates long-term scene flow directly from raw point-cloud sequences, it still exhibits an important limitation. The core challenge lies in its sensitivity to the distribution of input coordinates—particularly geometric scale and scene-dependent

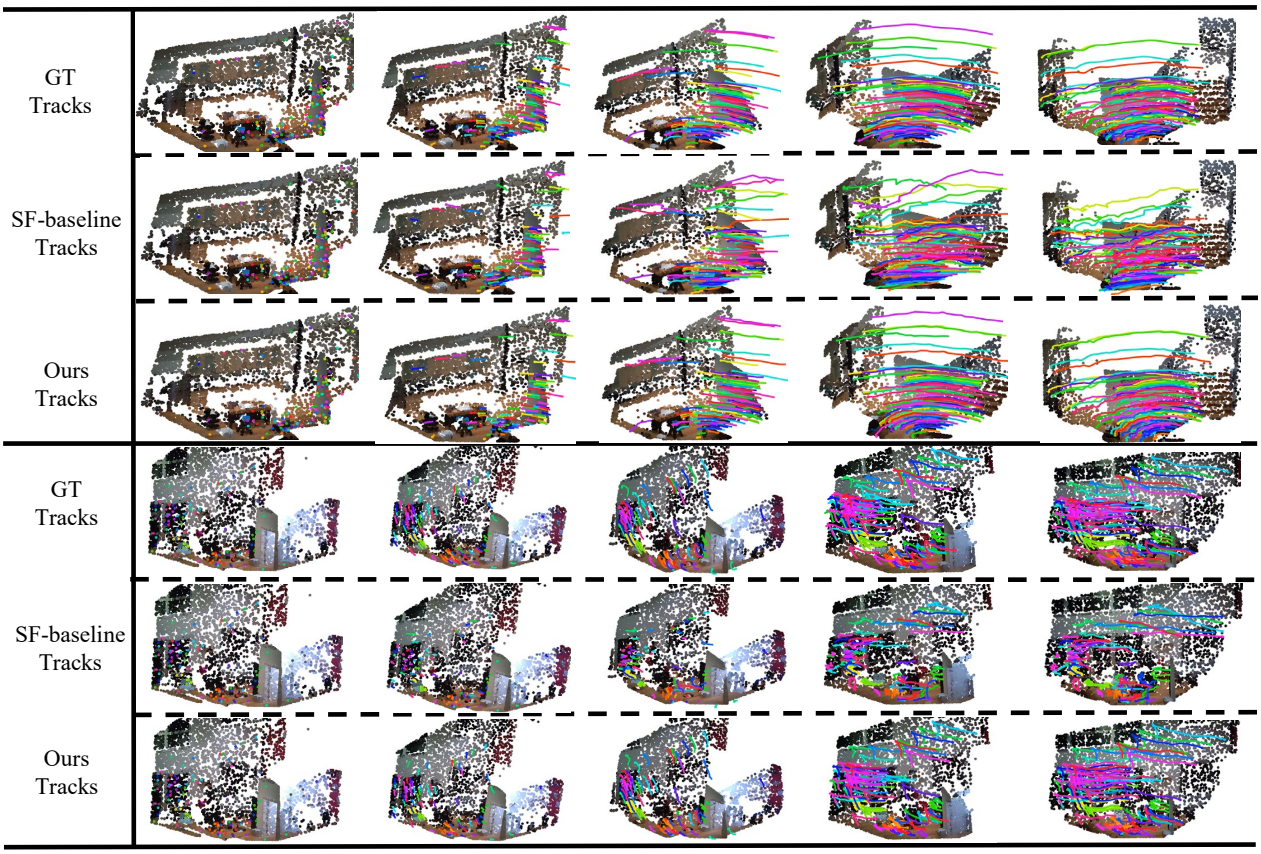
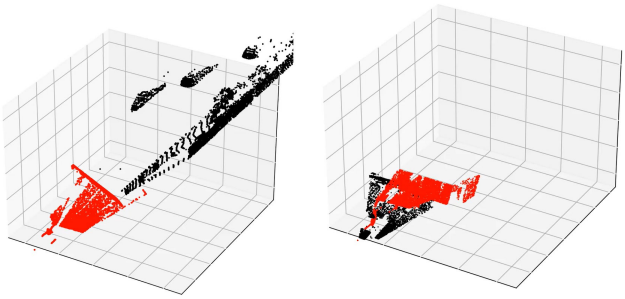


Figure 5. **Appendix Results on ADT3D.** Each row shows the point clouds and the motion trajectories of the query points at different timestamps within a given scene (each point cloud is colored with its RGB information for better visualization).

distance variations. When the distribution of test data deviates substantially from that of the training set, the model's performance deteriorates sharply.

To illustrate this issue, we construct DriveTrack3D, an autonomous-driving dataset for long-term point cloud scene flow estimation, built from LiDAR sequences in the Waymo Open Dataset and trajectory annotations from TAPVid-3D. As shown in Tab. 11, our model fails entirely on DriveTrack3D, despite performing well on ADT3D datasets. Fig. 6 further visualizes the spatial distribution gap: the training data and ADT3D share broadly similar geometric scales, whereas DriveTrack3D exhibits significantly larger depth ranges and outdoor spatial layouts.

This phenomenon raises an important research question for the community: how to achieve robust and generalizable long-term point cloud scene flow estimation across real-world environments with diverse spatial distributions and significant domain variations.



(a) PointOdyssey3D VS DriveTrack3D (b) PointOdyssey3D VS ADT3D

Figure 6. **Spatial distribution gap across datasets.** red points denote the training dataset (POD3D), while black points represent the datasets used for generalization evaluation.

Metric	DriveTrack3D	ADT3D
$EPE_{3D}(m)$	9.5	0.372

Table 11. **Generalization Results on Different Datasets.**

11. Visualization

In this section, we present additional visualization results to further demonstrate the effectiveness of our method and its accurate understanding of 3D motion. Fig. 5, 7, and 8 show more qualitative results on the ADT3D dataset, high-

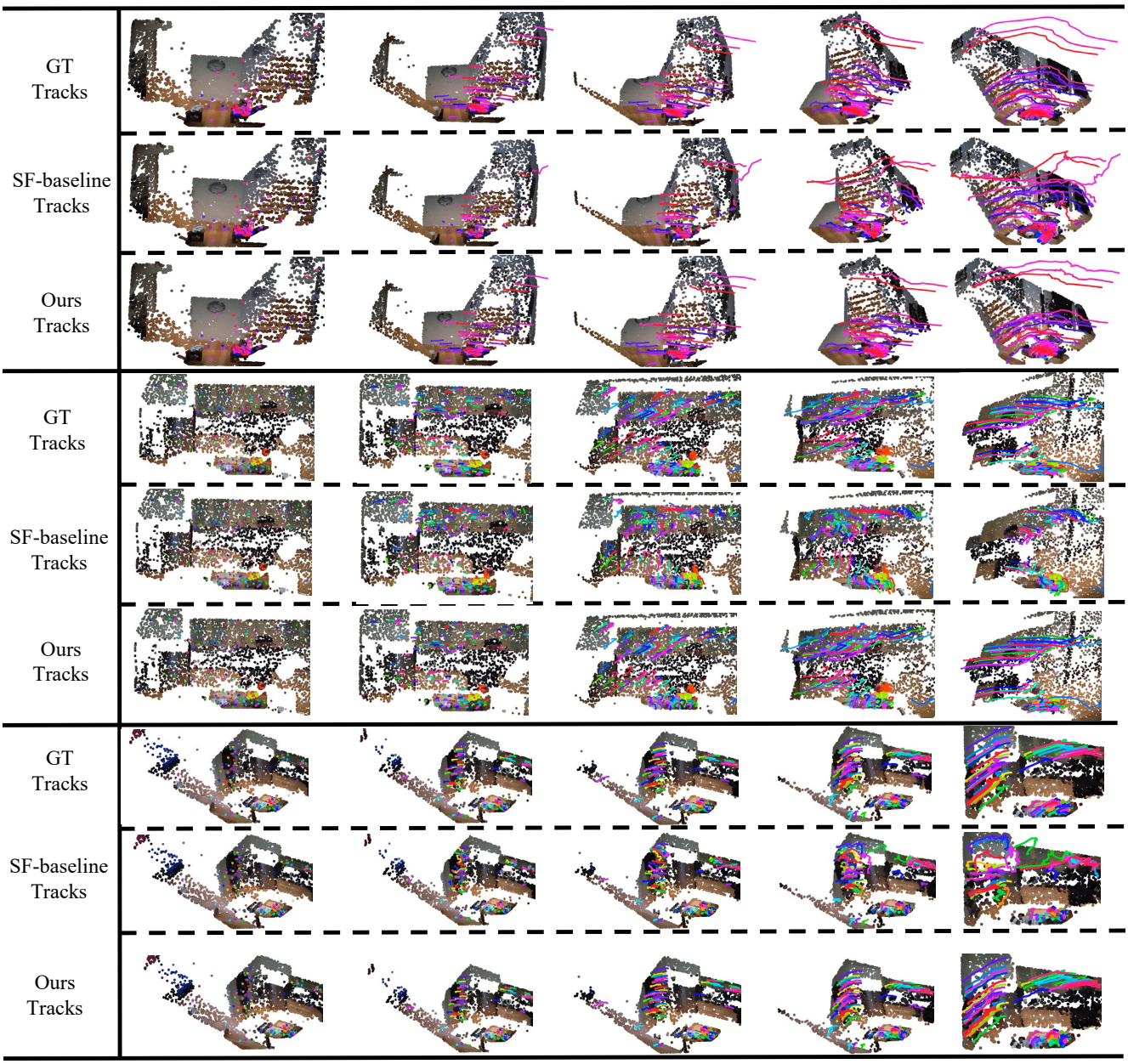


Figure 7. **Appendix Results on ADT3D.** Each row shows the point clouds and the motion trajectories of the query points at different timestamps within a given scene (each point cloud is colored with its RGB information for better visualization).

lighting the strong generalization ability of our approach in real-world scenarios. Fig. 9 further compares trajectories predicted by different methods, showing that our model produces more accurate and geometrically consistent point-level 3D motion, thereby validating its effectiveness and reliability in 3D motion understanding.

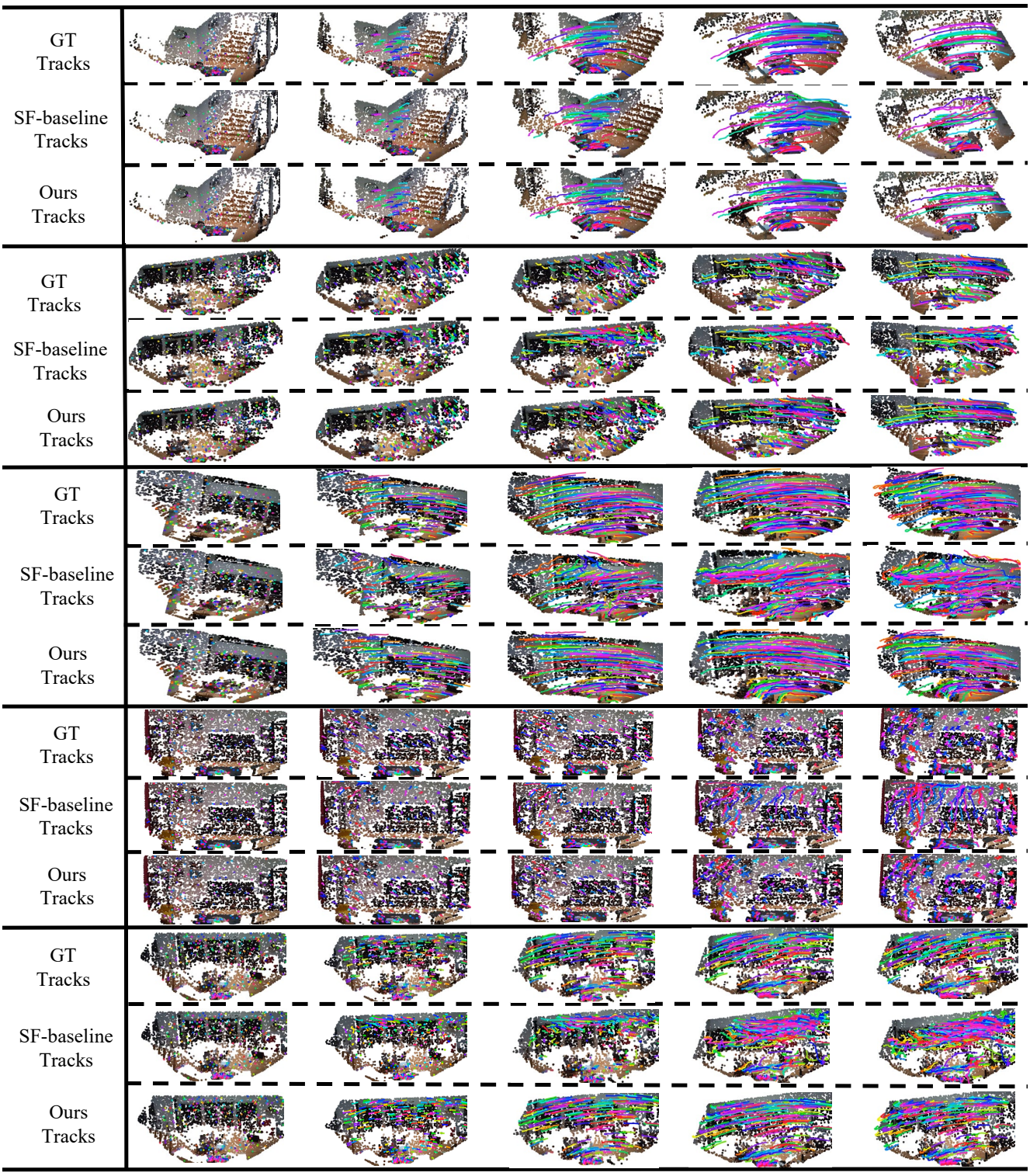


Figure 8. **Appendix Results on ADT3D.** Each row shows the point clouds and the motion trajectories of the query points at different timestamps within a given scene (each point cloud is colored with its RGB information for better visualization).

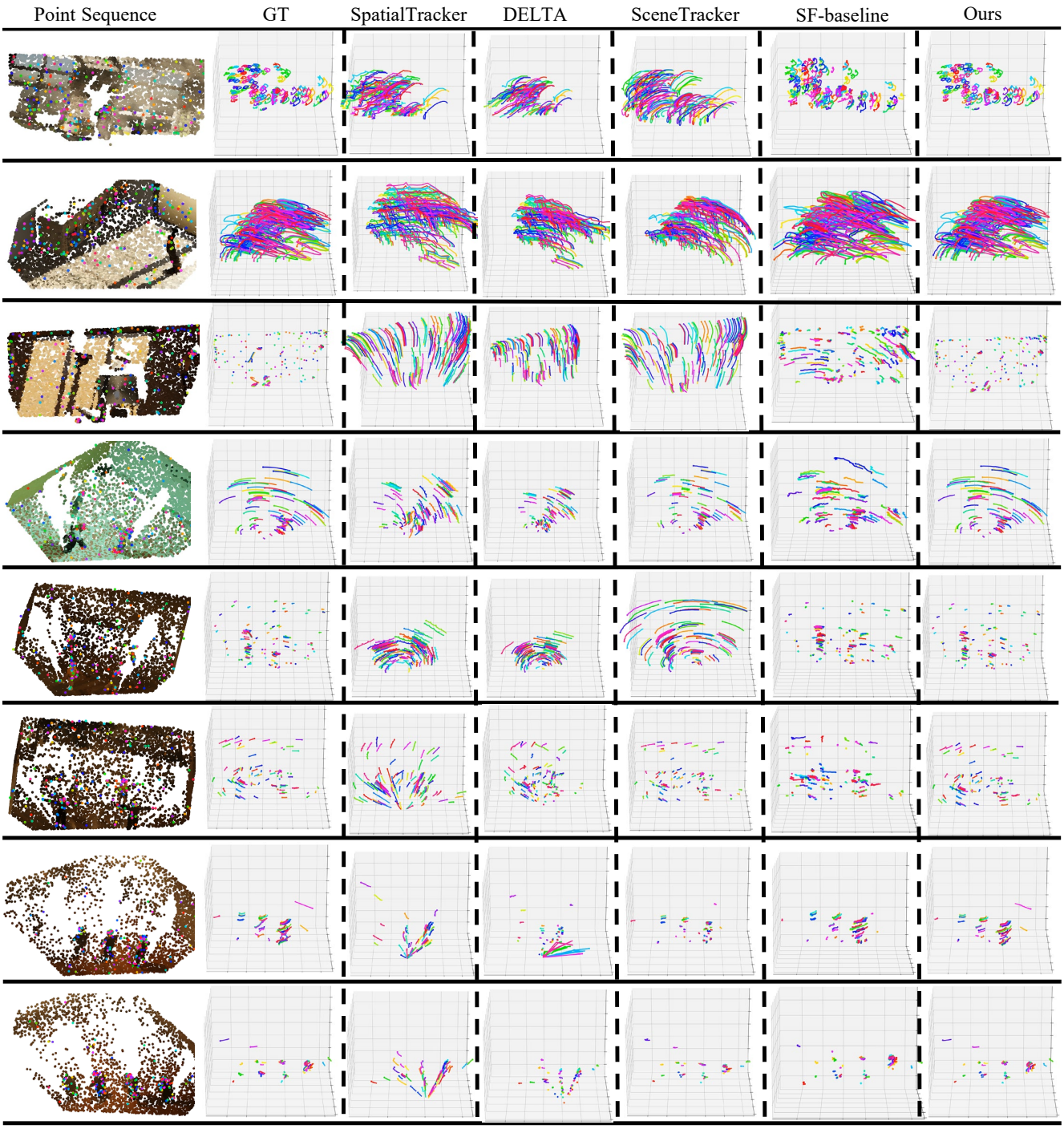


Figure 9. **Appendix Results on PointOdyssey3D.** The first column shows the input point cloud sequence (colored by RGB for visualization) and the corresponding query points. Columns 2–7 compare ground-truth trajectories with predictions from different methods.