Fast, Robust, Accurate, Multi-Body Motion Aware SLAM

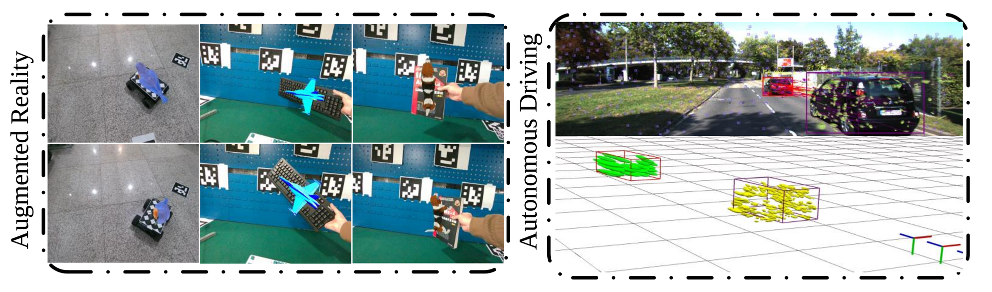
Linghao Yang, Yunzhou Zhang, Rui Tian Shiwen liang, You Shen, Sonya Coleman, Dermot Kerr

IEEE Transactions on Intelligent Transportation Systems(T-ITS Accept)

Overview

Task: Simultaneous Localization, Mapping and Moving Object Tracking (SLAMMOT)

Application:



Contribution:

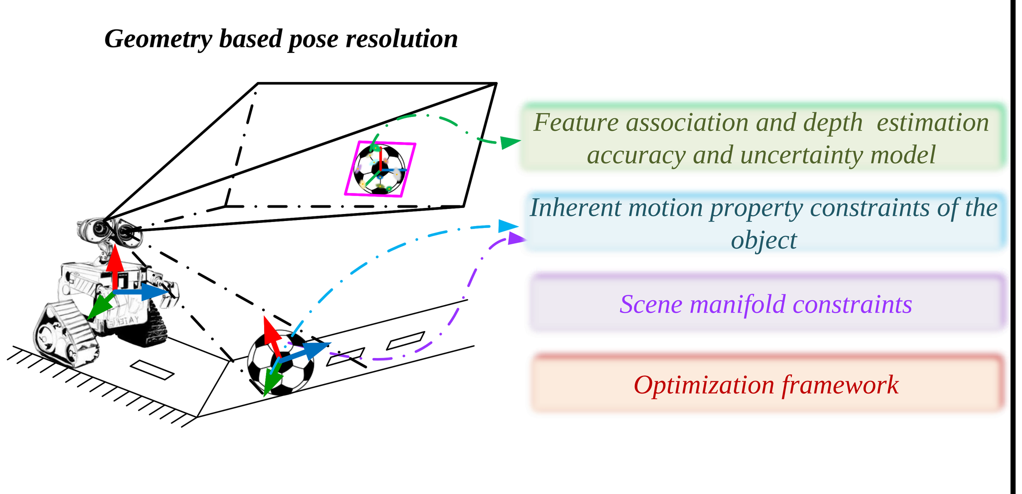
A multi-cues tracker with probabilistic fusion is proposed to enhance the robustness and accuracy of object-level association and 3-D motion tracking by integrating compensation properties from different cues.

A novel fully tightly-coupled state optimization factor graph is constructed, integrating rigid object motion properties, planar manifold priors, and unified geometric observations to enhance the stability and local consistency of multi-objects state estimation.

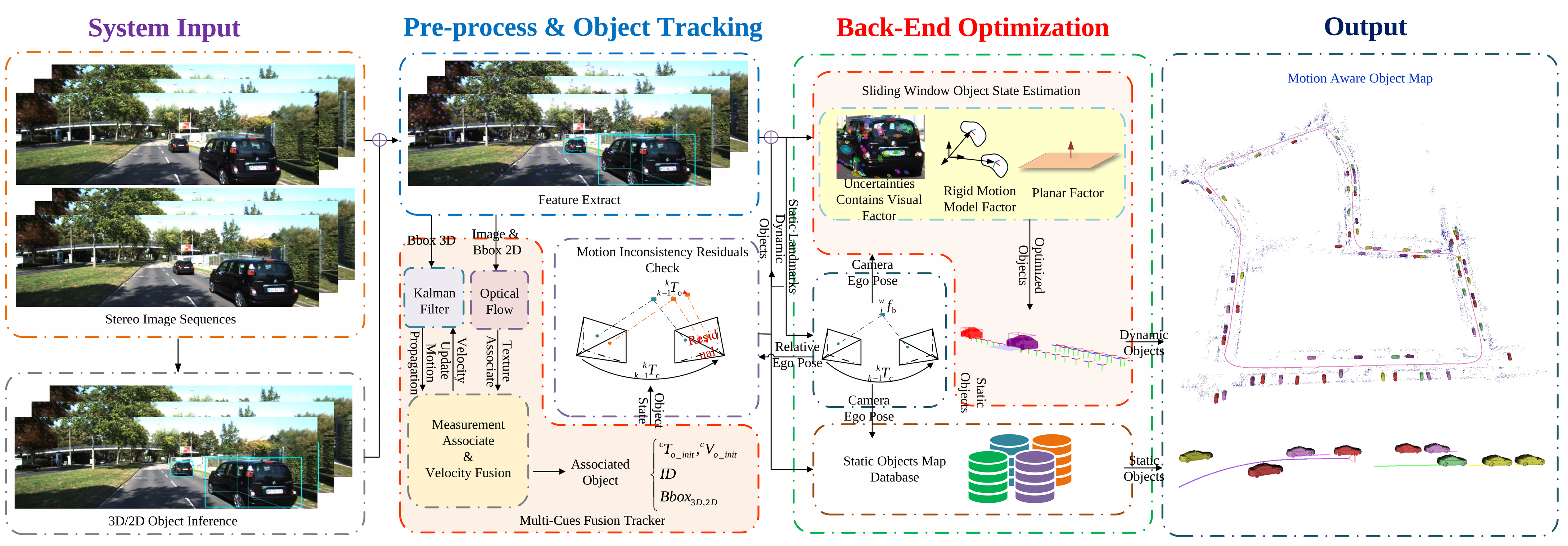
By combining a more dependable **uncertainty model and motion-aided object feature-level tracking algorithm**, we enhanced accuracy and robustness of object state estimation performance.

A vision-based fast, robust and accurate multi-body SLAM system combining filtering and sliding window optimization is proposed and validated using a publicly available benchmark dataset, simulation environment and self-built real-world dataset where it is shown to offer accurate localization and 3-D motion tracking of objects.

Motivation



Method



Experiment

KITTI Dataset

Table I: Camera ego-motion comparison with state-of-the-art systems using the KITTI Tracking Dataset. Best results are highlighted as first, second, and third. first second third

Sequence	ORB-SLAM3(Stereo) [26] ATE	ORB-SLAM3(Stereo) [26] RPE.r	ORB-SLAM3(Stereo) [26] RPE.t	DynSLAM [42] ATE	DynSLAM [42] RPE.r	DynSLAM [42] RPE.t	Li [16] ATE	ClusterSLAM [24] ATE	ClusterSLAM [24] RPE.r	ClusterSLAM [24] RPE.t	ClusterVO [25] ATE	ClusterVO [25] RPE.r	ClusterVO [25] RPE.t	Proposed Approach ATE	Proposed Approach RPE.r	Proposed Approach RPE.t
0926-0009	1.45	0.01	1.81	7.51	0.06	2.17	1.14	0.92	0.03	2.34	0.79	0.03	2.98	1.47	0.04	3.60
0926-0013	0.30	0.01	0.92	1.97	0.04	1.41	0.35	2.12	0.07	5.50	0.26	0.01	1.16	0.23	0.01	0.81
0926-0014	0.84	0.01	1.15	5.98	0.09	2.73	0.51	0.81	0.03	2.24	0.48	0.01	1.04	0.81	0.02	2.81
0926-0051	0.43	0.00	1.08	10.95	0.10	1.65	0.76	1.19	0.03	1.44	0.81	0.02	2.74	0.41	0.04	0.70
0926-0101	3.47	0.03	13.88	10.24	0.13	12.29	5.30	4.02	0.02	12.43	3.18	0.02	12.78	2.74	0.03	8.20
0929-0004	0.44	0.01	1.21	2.59	0.02	2.03	0.40	1.12	0.02	2.78	0.40	0.02	1.77	0.36	0.02	1.56
1003-0047	17.01	0.05	26.86	9.31	0.05	6.58	1.03	10.21	0.06	8.94	4.79	0.05	6.54	1.98	0.03	6.79

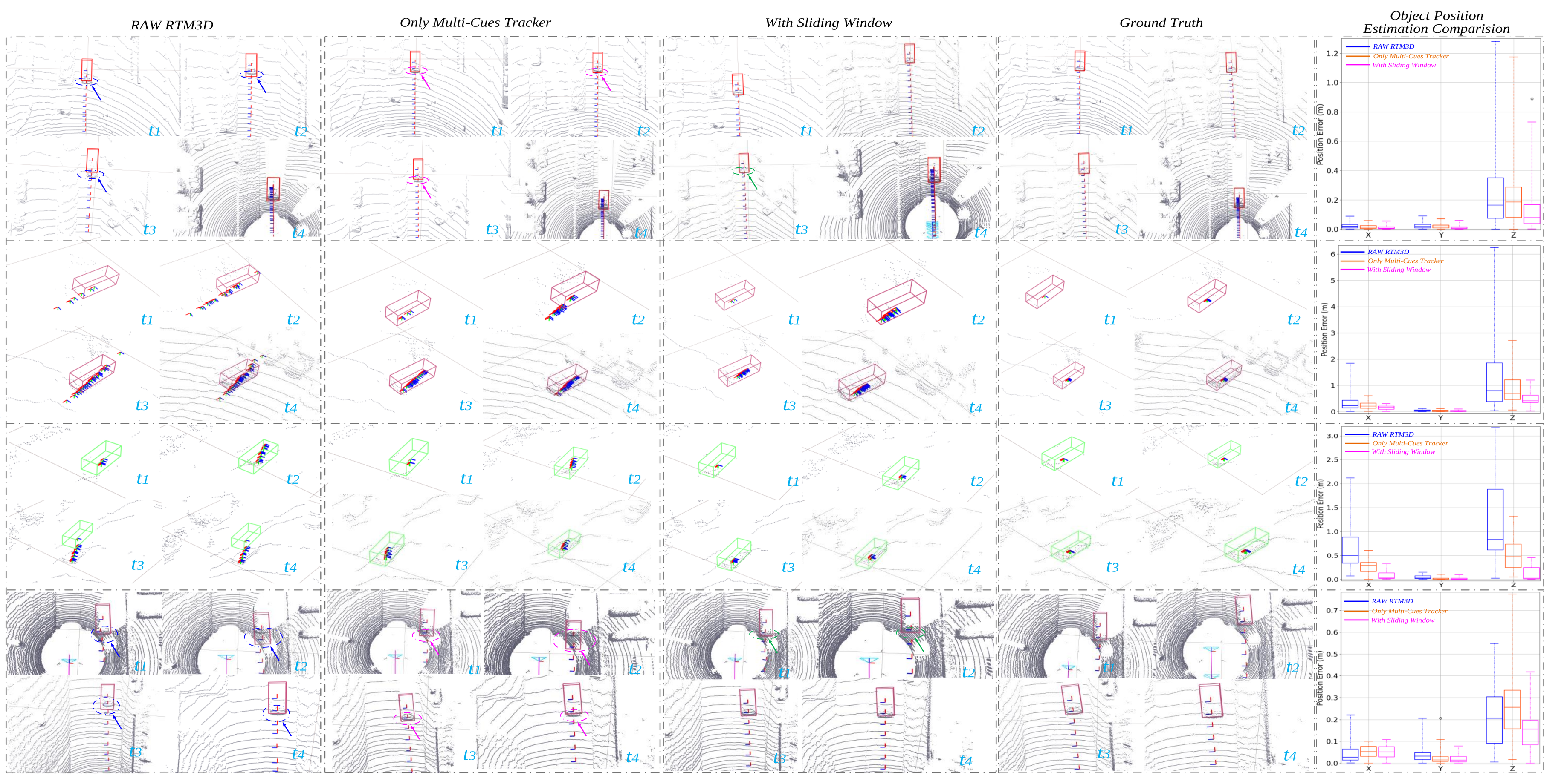
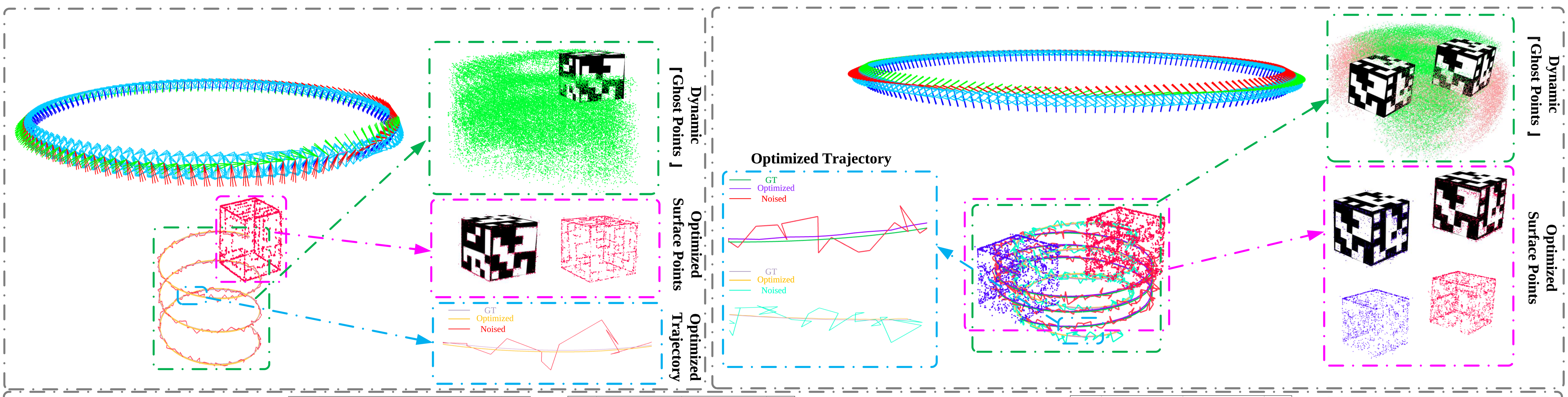
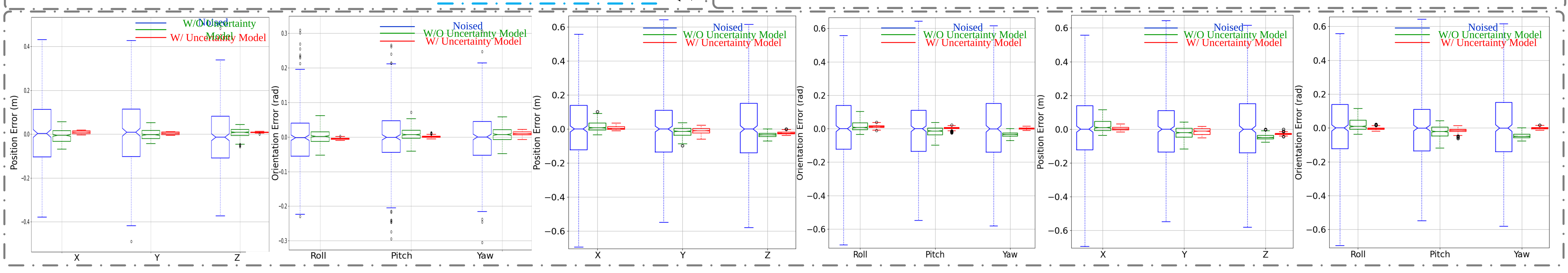


Table II: Object Motion comparison with state-of-the-art systems using the KITTI Tracking Dataset. Best results are highlighted as first, second, and third.

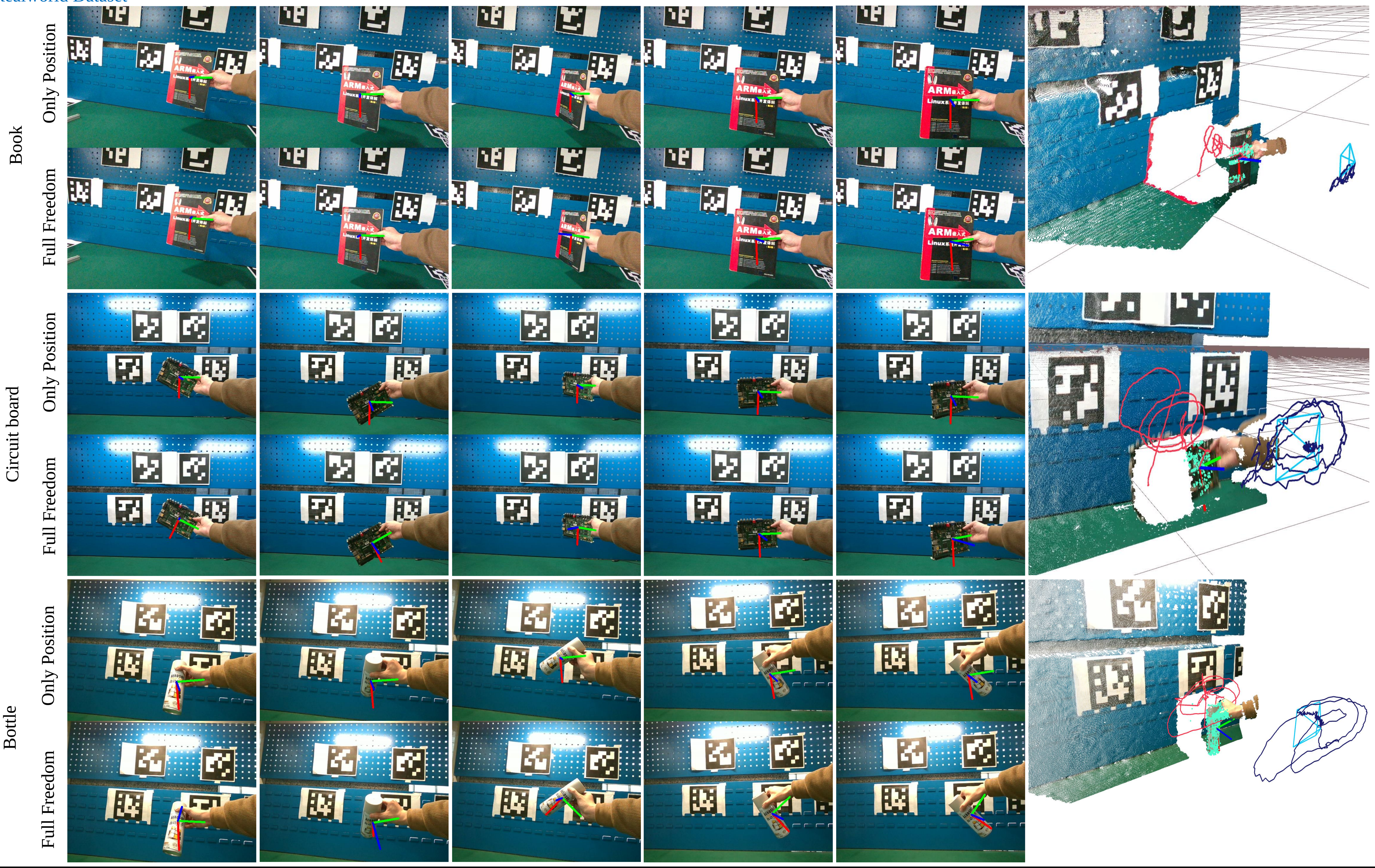
Sequence	DynaSLAM2 [46] ATE	RPE.t	RPE.r	2D TP	BV TP	3D TP	3D MOTP	Twist SLAM [33] ATE	RPE.t	RPE.r	2D TP	BV TP	3D TP	3D MOTP	Proposed Approach ATE	RPE.t	RPE.r	2D TP	BV TP	3D TP	3D MOTP
0013-01	0.69	0.34	1.84	50	39.34	38.53	48.2	0.31	0.1	0.28	58.02	58.02	58.02	60	0.179	0.197	0.035	90.90	66.05	64.55	75.25
0005-31	0.51	0.26	13.5	28.96	14.48	11.45	34.2	0.35	0.19	0.58	30.84	30.84	30.84	35	0.277	0.155	0.013	100.0	69.33	65.45	91.52
0010-00	0.95	0.40	2.84	81.63	70.41	68.37	40.28	0.77	0.21	1.98	7.20	6.10	5.80	2.80	0.20	0.151	1.03	100.0	87.17	86.55	89.65
0011-00	1.05	0.43	12.51	72.65	61.66	52.28	47.35	0.17	0.23	0.23	29.61	29.61	29.61	32.5	0.221	0.254	0.018	100.0	70.129	66.16	87.39
0011-35	1.25	0.89	16.64	53.17	19.05	6.35	26.02	0.1	0.03	0.11	65.00	65.00	65.00	67.5	0.873	0.714	0.032	85.96	12.62	12.12	76.34
0018-02	1.10	0.30	9.27	86.36	67.05	62.12	34.8	0.21	0.27	0.66	84.67	84.67	84.67	87.5	0.199	0.101	0.042	90.23	88.65	86.94	98.42
0018-03	1.13	0.55	20.05	53.33	21.75	16.84	35.8	0.15	0.21	0.56	28.19	28.19	28.19	30	0.303	0.333	0.635	95.33	81.59	80.46	89.27
0019-63	0.86	1.45	48.80	35.26	29.48	26.48	33.89	0.28	2.17	1.08	65.93	65.93	36.26	20.64	0.95	1.01	0.88	100.0	28.93	28.45	28.01
0019-72	0.99	1.12	3.36	29.11	29.43	29.43	39.81	0.16	0.05	0.34	16.92	16.92	16.92	20.00	0.78	0.32	0.33	95.56	28.39	28.63	78.22
0020-00	0.56	0.45	1.3	63.68	43.78	31.84	46.15	0.17	0.2	0.72	84.75	84.75	84.75	87.5	0.216	0.169	0.024	88.96	86.36	84.68	91.03
0020-12	1.18	0.4	6.19	42.77	37.64	36.23	40.81	0.24	0.2	1.54	14.24	13.91	13.04	17.25	0.668	0.154	0.325	100.0	17.42	17.42	61.66
0020-122	0.87	0.72	5.75	34.9	34.51	29.02	44.43	0.17	0.02	0.07	84.94	84.75	84.75	87.5	1.065	0.991	0.006	97.59	17.26	13.53	81.17

Simulation Dataset





Realworld Dataset



Conclusion

We introduce an innovative vision-basesd multi-body SLAM system. We make up rigid environment as a unified whole to assist state decoupling by integrating high level semantic information, ultimately enabling simultaneous multi-state estimation. A novel framework is developed for integrating different complementary constraints. It makes it possible for accurate 3-D motion tracking of arbitrary unmodelled, rigid and textured objects and better performance of VSLAM systems in dynamic scenes. Comparable results to state-of-the-art multi-body state estimation solutions using a public benchmark, self-built simulation and real-world datasets demonstrate the effectiveness of our system.