Accurate and Robust Object SLAM with 3D Quadric Landmark Reconstruction in Outdoors

IEEE Robotics and Automation Letters (RA-L)
IEEE International Conference on Robotics and Automation (ICRA)

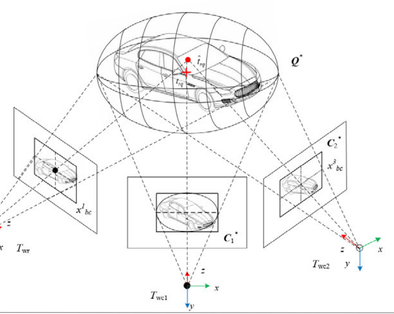
Rui Tian[1], Yunzhou Zhang[1*], Yonghui Feng[1], Linghao Yang[1], Zhenzhong Cao[1], Sonya Coleman[2], Dermot Kerr[2]

Object-oriented and semantic-enhenced map

- To effectively overcome the observation noise, we propose an accurate and robust quadric landmark initialization method based on the SQP algorithm by separating of quadric parameters.
- We proposed an ODA algorithm that combines semantic inliers distribution, Kalman-based motion prediction, and ellipsoidal projection to achieve accurate object data association and pose estimation.
- Jointly estimate the camera pose and the quadric landmarks of objects.
- Based on the proposed algorithms, we implement a real-time stereo visual SLAM with accurate and robust ellipsoids representation of objects, aming to build an object-oriented and semantically-enhanced map for outdoor navigation.

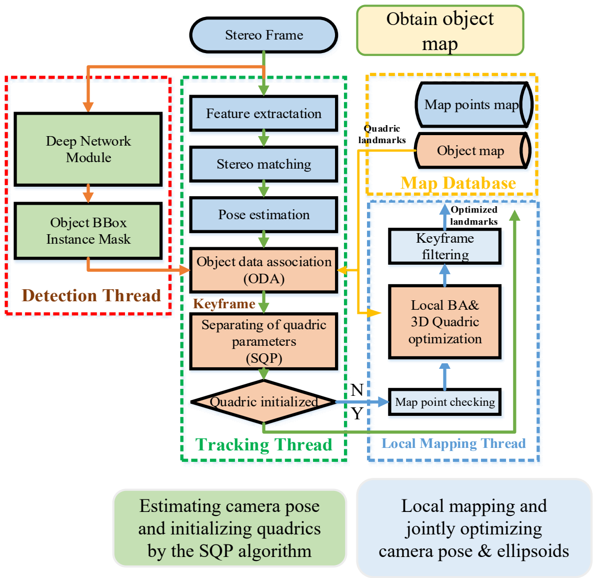
Facing problems

The state-of-the-art quadric-based SLAM algorithms always face observability problems and are sensitive to observation noise, which limits their application in outdoor scenes.t

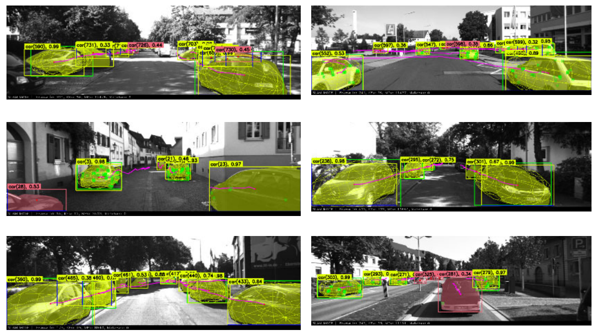


System Overview

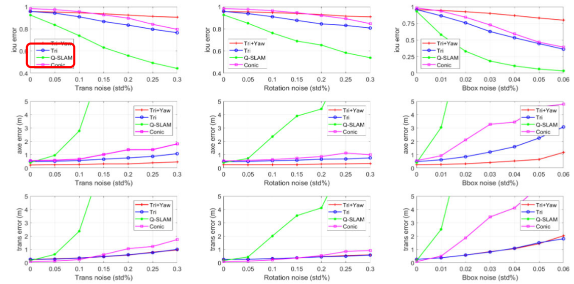
Decoupling of Quadric Parameters（DQP）
Object Data Association （ODA）



Experiments



Robustness to observation noise



Accuracy in initialization success rate, IOU, object pose estimation and localization

TABLE I
SUCCESS RATE COMPARISON IN KITTI RAW DATA DATASET.

Sequence	Ours	Conic [1]	Q-SLAM [6]
09	**0.6912**	0.4468	0.2706
22	**0.6512**	0.3333	0.2923
23	**0.6230**	0.3949	0.2829
36	**0.6047**	0.4096	0.3514
59	**0.5625**	0.3556	0.2558
93	**0.4815**	0.2826	0.3617
Average	**0.6023**	0.3705	0.3024

TABLE II
IoU_{2D} COMPARISON IN KITTI RAW DATA DATASET.

Sequence	Ours	Conic [1]	Q-SLAM [6]
09	**0.7335**	0.7252	0.7031
22	0.7629	**0.7791**	0.7662
23	0.7509	**0.7529**	0.6959
36	**0.7604**	0.7558	0.7127
59	0.6508	**0.6878**	0.6500
93	0.7232	0.6751	**0.7433**
Average	**0.7303**	0.7293	0.7119

TABLE III
TRANSLATION ERROR (M) COMPARISON IN KITTI RAW DATA DATASET.

Sequence	Ours	Conic [1]	Q-SLAM [6]
09	**2.5456**	2.7819	3.6110
22	2.1769	**1.9552**	1.9651
23	**2.3341**	5.6605	8.7088
36	**1.8594**	2.7175	6.9814
59	**1.3276**	1.6883	1.3874
93	**2.5226**	5.4130	4.0654
Average	**2.1277**	3.3694	4.4532

TABLE IV
AXIAL LENGTH ERROR (M) COMPARISON IN KITTI RAW DATA DATASET.

Sequence	Ours	Conic [1]	Q-SLAM [6]
09	**0.6271**	1.2618	1.1400
22	**0.5565**	0.7233	0.8356
23	**0.5494**	1.7886	0.6837
36	**0.7121**	1.2797	0.8799
59	**0.6706**	1.8467	1.2908
93	**0.7357**	1.3156	0.8574
Average	**0.6419**	1.3693	0.9479

TABLE V
RMS ATE (M) ON THE KITTI RAW DATA DATASETS.

Sequence	-09	-22	-23	-27	-36
BA with quadrics	**1.09**	1.82	0.64	**1.08**	**4.25**
BA w/o quadrics	1.11	**1.76**	**0.61**	1.18	4.57
Q-SLAM	5.73	4.32	2.10	7.91	14.74

Real-time performance with GPU

TABLE VI
RUNNING TIME (MS) OF THE MAIN MODULES OF THE PROPOSED SYSTEM ON THE KITTI RAW DATA DATASET.

	Sequence	-09	-22	-23	-36
Settings	Resolution	1242x375	1242x375	1242x375	1242x375
	Camera FPS	10 Hz	10 Hz	10 Hz	10 Hz
	ORB Features	2000	2000	2000	2000
Detection	Segmentation	57.61± 6.75	53.51±5.69	59.86±5.56	54.36±6.38
Tracking	ORB Extraction	59.67±7.93	59.15±8.37	60.99±8.84	56.94±8.25
	ODA	2.04±3.34	0.94±0.77	2.12±0.78	0.93±1.01
	SQP	0.17±0.21	0.24±0.15	0.14±0.12	0.11±0.31
	Total time	**90.14**±8.47	**86.61**±9.31	**93.30**±10.74	**85.88**±9.12
Local Mapping	Local BA	220.06±119.05	211.92±93.86	204.23±68.10	208.33±89.22
	Total time	**327.29**±126.71	**314.45**±114.46	**321.18**±79.24	**309.98**±92.21

Separating of Quadric Parameters（SQP）initialization algorithm

$$
\begin{aligned}
Q_r^* &= T_{rq} Q_q^* T_{rq}^T \\
&= \begin{bmatrix} R_{rq} & t_{rq} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0^T & -1 \end{bmatrix} \begin{bmatrix} R_{rq}^T & 0 \\ t_{rq}^T & 1 \end{bmatrix} \\
&= \begin{bmatrix} R_{rq} D R_{rq}^T - t_{rq} t_{rq}^T & \boxed{-t_{rq}} \\ -t_{rq}^T & -1 \end{bmatrix} \\
&= \begin{bmatrix} \boxed{Q_{r33}^*} & -t_{rq} \\ -t_{rq}^T & -1 \end{bmatrix}
\end{aligned}
$$

Independently estimate quadric center translation

$$
R_{rq} = \boxed{\begin{bmatrix} cos\theta_y & 0 & sin\theta_y \\ 0 & 1 & 0 \\ -sin\theta_y & 0 & cos\theta_y \end{bmatrix}}
$$

$$
Q_r^* = \begin{bmatrix} a_x^2 cos^2\theta_y + a_z^2 sin^2\theta_y - t_{rq}^{x2} & -t_{rq}^x t_{rq}^y & cos\theta_y sin\theta_y (a_z^2 - a_x^2) - t_{rq}^x t_{rq}^z & -t_{rq}^x \\ -t_{rq}^x t_{rq}^y & a_y^2 - t_{rq}^{y2} & -t_{rq}^y t_{rq}^z & -t_{rq}^y \\ cos\theta_y sin\theta_y (a_z^2 - a_x^2) - t_{rq}^x t_{rq}^z & -t_{rq}^x t_{rq}^y & a_x^2 sin^2\theta_y + a_z^2 cos^2\theta_y - t_{rq}^{z2} & -t_{rq}^z \\ -t_{rq}^x & -t_{rq}^y & -t_{rq}^z & -1 \end{bmatrix}
$$

Yaw rotation assumption

Object Data Association （ODA）

$$
a_{ij} = \alpha a_{ij}^p + \beta a_{ij}^g + \gamma a_{ij}^k
$$

Hungarian assignment algorithm

Semantic inliers points distance + Intersection of union distance + Prior object size distance

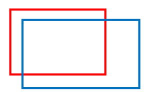
Object Observation Optimization

$$
\begin{aligned}
T, X, q = \arg\min_{T,X} \sum_i \sum_j & \boxed{\|f(T_i, X_j) - u_{ij}\|_{\Sigma_i}^2} + \\
\arg\min_{T,q} \sum_i \sum_j \{ & \boxed{\|e_b(B_{O_i}, B_{D_j})\|_{\Sigma_b}^2} + \\
& \boxed{\|e_a(O_i)\|_{\Sigma_a}^2} + \boxed{\|e_p(\Pi_{D_j}, \Pi_{O_i})\|_{\Sigma_p}^2} \}
\end{aligned}
$$

Projection error

2D detection error

B(D_j)



B(O_i)

Texture plane error

Prior axial length error