

# 一种基于多规则约束的SLAM概率物体关联方法研究\*

张建博<sup>1</sup>, 陶庆<sup>1†</sup>, 袁亮<sup>1,2</sup>

(1. 新疆大学 机械工程学院, 新疆 乌鲁木齐 830017; 2. 北京化工大学 信息科学与技术学院, 北京 100029)

**摘要:** 针对单目语义同时定位与地图构建在物体关联上的关联模糊问题, 提出一种基于多规则约束的概率物体关联方法. 利用物体的类别、尺寸和朝向等语义信息进行物体关联, 消除物体关联过程中的模糊性, 从而提高SLAM的定位精度, 进而可以为地图构建提供准确的位姿. 通过室外公共数据集KITTI验证了基于多规则约束的概率物体关联方法能够有效提高SLAM的定位精度.

**关键词:** 同时定位与地图构建; 数据关联; 语义信息

**DOI:** 10.13568/j.cnki.651094.651316.2021.09.25.0001

**中图分类号:** TP242 **文献标识码:** A **文章编号:** 2096-7675(2022)04-0492-06

**引文格式:** 张建博, 陶庆, 袁亮. 一种基于多规则约束的SLAM概率物体关联方法研究[J]. 新疆大学学报(自然科学版)(中英文), 2022, 39(4): 492-497.

**英文引文格式:** ZHANG Jianbo, TAO Qing, YUAN Liang. A probabilistic object association method for SLAM based on multi-rule constraints[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2022, 39(4): 492-497.

## A Probabilistic Object Association Method for SLAM Based on Multi-Rule Constraints

ZHANG Jianbo<sup>1</sup>, TAO Qing<sup>1</sup>, YUAN Liang<sup>1,2</sup>

(1. School of Mechanical Engineering, Xinjiang University, Urumqi Xinjiang 830017, China;

2. School of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

**Abstract:** Aiming at the ambiguous problem of simultaneous localization and mapping(SLAM) in object association, a probabilistic object association method based on multi-rule constraints was proposed. The semantic information of objects such as category, size and orientation is used for object association, eliminating the ambiguity in the process of object association, thus improving the localization accuracy of SLAM, and providing accurate pose for mapping. Based on the outdoor public dataset, it is verified that the multi-rule probabilistic object association method can effectively improve the localization accuracy of SLAM.

**Key words:** simultaneous location and map construction; data association; semantic information

## 0 引言

同时定位与地图构建(Simultaneous Localization And Mapping, SLAM)是未知环境下移动机器人根据自身携带的传感器进行自主定位和环境地图构建,近年国内外对此进行了大量研究<sup>[1-3]</sup>. 语义SLAM在机器人的人机交互、任务级导航以及路标级定位方面起着至关重要的作用,而其最关键的一环是相机连续帧中观测的语义信息与真实物理世界中物体的关联问题,由于观测噪声以及假阳性问题会造成物体关联的模糊,从而严重影响SLAM的定位精度. 经典视觉SLAM分为前端跟踪和后端优化,目前的物体关联方法分为基于前端和基于后端两种方法. Bowman等人基于高斯假设,采用期望最大化(Expectation-Maximization, EM)算法进行迭代优化<sup>[4]</sup>. 而Doherty等人将物体关联看作是非高斯推断问题,并在因子图的基础上增加语义因子进行优化<sup>[5-6]</sup>. 他

\* 收稿日期: 2021-09-25

基金项目: 国家自然科学基金(51865056).

作者简介: 张建博(1996-),男,博士生,从事移动机器人定位与导航的研究, E-mail: jianbo\_zhang@stu.xju.edu.cn.

† 通讯作者: 陶庆(1978-),男,博士,教授,主要从事先进设计与制造技术和康复机器人的研究, E-mail: xjutoao@qq.com.

们都在SLAM的后端优化部分对物体约束进行了联合优化,但在大规模SLAM中,计算成本将会非常高,且仍然很难解决物体关联模糊的问题。

Yang等人在线将物体建模为立体框,利用物体检测框中检测的共视特征点的数量进行物体关联<sup>[7]</sup>。Wu等人结合文献[7-8]中对物体的表示,并采用一种集成的物体关联策略<sup>[9]</sup>。上述工作在大部分情况下是轻量且鲁棒的,但由于本质是基于特征点的关联,而在特征点较少或物体间有相互遮挡的环境下很容易发生关联模糊。Ran等人提出了多假设狄利克雷过程的非参方法进行物体关联<sup>[10]</sup>。Zhang等人提出用非参概率模型进行物体关联<sup>[11]</sup>,但仍未考虑多种物体类别以及假阳性观测的问题。

针对上述问题,本文提出一种基于多规则约束的概率物体关联方法。该方法结合德罗内克函数进行多物体的语义类别约束,增加了可关联物体的语义种类,同时分别对物体的欧式距离、尺寸、朝向以及共视特征点数量进行概率建模。然后选择最大关联概率进行观测与路标的物体关联,减少由于单一关联规则造成的关联模糊问题。最后将概率物体关联引入语义SLAM系统,并在后端进行联合优化,从而提高SLAM的定位精度。

## 1 语义SLAM物体关联

### 1.1 语义SLAM

语义SLAM可以描述为根据传感器观测的物体信息推断移动机器人的位姿以及周围环境中语义物体的位置。相比传统基于特征点、线或面的SLAM系统,语义SLAM能够根据物体的多种属性提供更准确的物体关联从而提高定位精度。在未知环境中,搭载的传感器随着机器人的运动,观测到一系列路标 $Z(0:M) = \{z_0, z_1, \dots, z_m\}$ ,最终的目标是推断机器人位姿 $X(0:T) = \{x_0, x_1, \dots, x_t\}$ 和物体的位置 $L(0:N) = \{l_0, l_1, \dots, l_n\}$ ,而在观测 $z_i$ 与路标 $l_j$ 之间的对应关系 $D$ 就是物体关联要解决的问题。所以SLAM本质上可以转换为如下概率推断问题:

$$\hat{X}, \hat{L} = \arg \max_{X, L} p(X, L | Z, \hat{D}) \quad (1)$$

当前的物体关联概率可以通过给定的先验 $\hat{X}, \hat{L}$ 进行推断:

$$\hat{D} = \arg \max_D p(D | \hat{X}, \hat{L}, \hat{Z}) \quad (2)$$

### 1.2 多规则约束

基于视觉的图像包含丰富的语义信息,其中包括物体类别属性、大小尺寸以及朝向。这些属性在物体关联中可以起到至关重要的作用,而传统物体关联都未曾考虑,这在具有物体遮挡等场景下会出现关联模糊问题,从而影响视觉SLAM的定位精度。本文利用物体的多种属性规则建模,对物体关联进行约束。

采用文献[7]的3D立体框表示物体,每个物体由9个参数组成, $O = \{T, d\}$ ,其中 $T = [R, t] \in SE(3)$ 表示物体的6自由度位姿, $d \in R^3$ 表示立体框的三维长度。本文建立五种规则进行鲁棒的物体关联,分别为物体框的语义类别规则、欧氏距离规则、大小尺寸规则、朝向角度规则和包含特征点规则。

物体框的语义类别规则规定只有相同的类别观测才能进行物体关联,采用Kronecker delta函数建模:

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (3)$$

$$p_c = \delta_{ij} \cdot p_r(O_i) \quad (4)$$

式中: $p_c$ 表示物体的观测 $O_i$ 关联到路标 $l_j$ 上的概率,本文采用YOLOv3<sup>[12]</sup>网络进行2D物体框的检测, $p_r$ 表示检测的置信度。当观测类别与物体类别相同时, $\delta_{ij} = 1$ ,语义关联概率取决于物体观测的置信度。而观测类别与物体类别不是同类物体时, $\delta_{ij} = 0$ 、 $p_c = 0$ ,观测不与该路标进行关联。

检测与路标的3D物体框的欧氏距离、大小尺寸以及朝向角度规则具有相似的规律,即二者的差值越小代表物体关联的概率越高。首先定义误差项:

$$e_d = \sqrt{(D_o - D_l)^2} \quad (5)$$

$$e_s = \sqrt{(S_o - S_l)^2} \quad (6)$$

$$e_a = \sqrt{(A_o - A_l)^2} \quad (7)$$

式中:  $e_a$ 表示观测与路标的误差.  $D_o$ 、 $S_o$ 以及 $A_o$ 分别表示物体的位置、大小和朝向角度. 相应的概率可以建模为负指数函数:

$$P_o = e^{-\frac{e_a}{U}} \quad (8)$$

式中:  $P_o$ 分别表示距离概率 $P_d$ 、尺寸概率 $P_s$ 和朝向概率 $P_a$ ;  $U$ 是尺度因子. 此概率模型为单调递减的, 即观测与路标差值越小, 物体关联的概率越大.

最后一项规则是包含特征点的属性, 由于每个物体上都有对应特征点, 传统物体关联方法就是根据观测与路标物体上包含的共视特征点数量进行判断<sup>[7]</sup>. 本文利用逻辑回归模型对此规则进行建模:

$$p_g = \frac{1}{1 + \exp(-a(x - b))} \quad (9)$$

式中:  $a$ 、 $b$ 分别为形状因子和阈值. 这一规则的特征主要体现在: 当共视特征点个数 $x$ 接近阈值 $b$ , 关联概率近似接近线性; 当个数少于拐点值, 则关联概率接近0, 而当大于稳定值时对应概率接近1, 即保留关联的模糊性而让其它规则进行判断.

总的概率物体关联模型如下:

$$P = p_c p_d p_s p_a p_g \quad (10)$$

由此获得当前观测与各路标之间的关联概率, 根据最大似然估计(Maximum Likelihood Estimation, MLE), 选择最大关联概率进行关联.

## 2 实验验证

本文对提出的方法在室外公共数据集KITTI<sup>[13]</sup>上进行了对比验证. 对比了ORB-SLAM2<sup>[14]</sup>和CubeSLAM<sup>[7]</sup>, 根据单目视觉SLAM系统, 我们设定输入为环境的图像信息, 并保持前端跟踪、后端优化与CubeSLAM一致, 将物体关联部分的算法作为唯一的变量, 用以验证物体关联算法, 从而使SLAM的定位精度可以反映物体关联的性能.

实验平台是Intel i5-7500 CPU、3.4 GHz、NVIDIA GeForce GT 1030显卡. 为验证大范围场景下物体关联的性能, 在实验中选择公共数据集KITTI进行验证实验, KITTI数据集包含了室外场景下的城镇和乡村、长距离和短距离等丰富的场景, 可以较好地验证系统的泛化性. 此外, 由于序列01、03和04仅包含极少的物体目标, 无法用来验证本文所提的物体关联方法, 故采用其它8个序列进行验证. 基于前人工作<sup>[7]</sup>, 也基于水平面假设, 故假设检测物体垂直且相接于地面.

### 2.1 定量实验

实验计算了轨迹的均方根误差(Root Mean Squared Error, RMSE). 对比了ORB-SLAM2<sup>[14]</sup>、CubeSLAM<sup>[7]</sup>和MHDP<sup>[10]</sup>, 其中CubeSLAM和本文方法都是基于单目ORB-SLAM2系统, 由于增加了物体约束, 从而证明了物体级语义SLAM的精度高于传统基于特征点的SLAM方法. 在实验中, 对四种算法都关闭了回环检测功能, 以验证仅依靠物体级定位的精度.

表 1 均方根误差

序列	00	02	05	06	07	08	09	10
ORB-SLAM2	80.03	35.78	41.05	51.86	17.93	50.72	55.04	26.69
CubeSLAM	13.90	26.20	4.75	6.98	2.67	10.70	10.70	8.37
MHDP	9.14	-	3.66	-	2.35	5.20	9.47	6.59
本文SLAM	4.00	17.22	2.42	4.25	1.48	4.57	4.11	1.80

表1展示了KITTI数据集中序列00、02、05、06、07、08、09和10上的均方根误差. 其中CubeSLAM和MHDP的结果来源于文献<sup>[7]</sup>和文献<sup>[10]</sup>. 由表1可知, 语义SLAM相比传统的基于特征点的SLAM定位精度提升了76.53%. 这是由于物体级的语义信息具有更加鲁棒的特性, 为SLAM提供了更稳定的约束, 从而提高了定位精度. 与

CubeSLAM相比, 由于本文算法加入了多规则约束的物体关联方法, 减少了物体关联模糊的问题, 所以定位精度提高了51.12%。与同类的概率物体关联方法MHDP相比(其采用距离作为关联规则), 本文定位精度提高了49.52%。上述结果间接证明了物体关联结果对语义SLAM具有显著的影响。

### 2.2 定性实验

图1~图4展示了本文方法的定性结果, 各图中左图展示的是轨迹在 $x$ - $z$ 轴上的轨迹( $y$ 轴垂直于地面), 中图和右图展示了连续图像帧在 $x$ 轴、 $y$ 轴以及 $z$ 轴上的位移误差和旋转误差。从KITTI数据集的轨迹误差、平移误差以及旋转误差上可以看出, 本文方法误差更小、曲线更接近真实值。而没有回环检测的ORB-SLAM2具有大的累计漂移, 语义SLAM中通过物体特征的约束, 减少了累计误差。但CubeSLAM的物体关联方法在室外大范围场景下鲁棒性较差, 而基于改进的多规则约束的方法提高了物体关联的准确率, 从而更接近真实值。

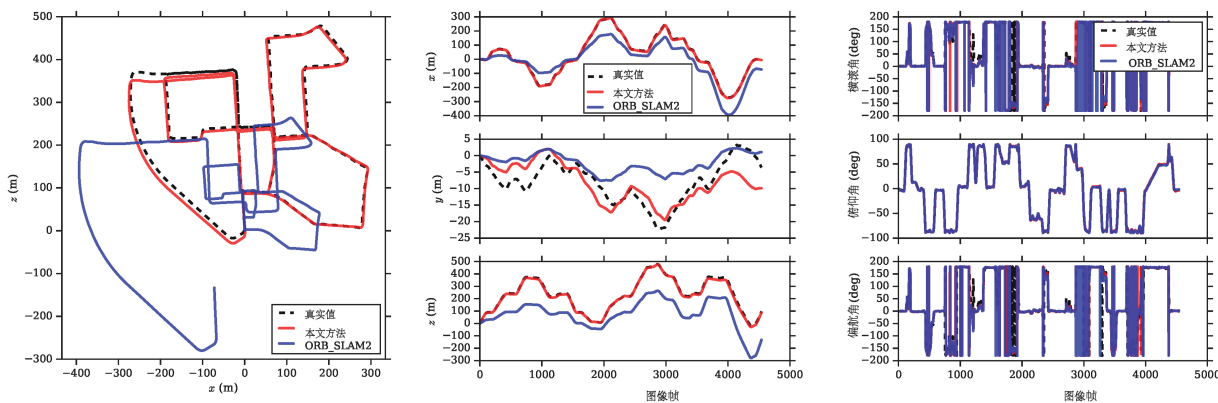


图 1 序列00的轨迹、平移和旋转误差曲线

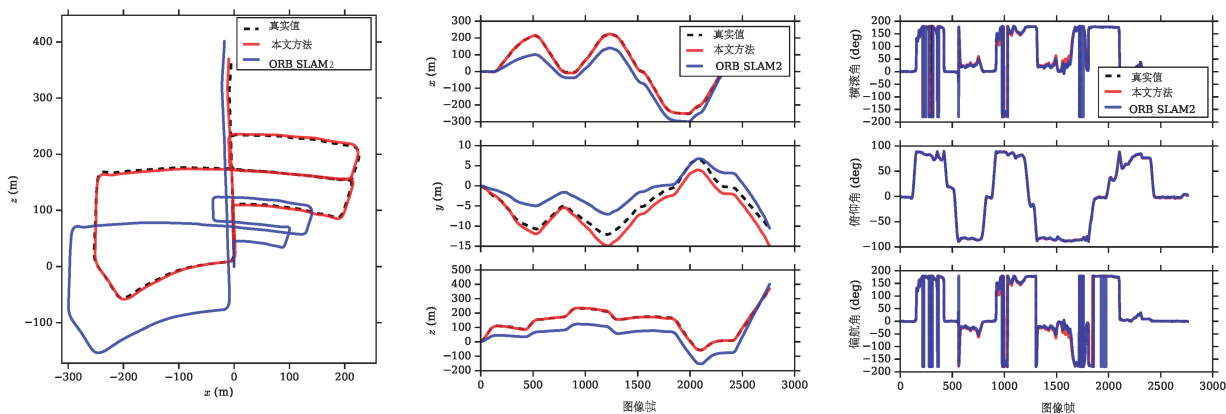


图 2 序列05的轨迹、平移和旋转误差曲线

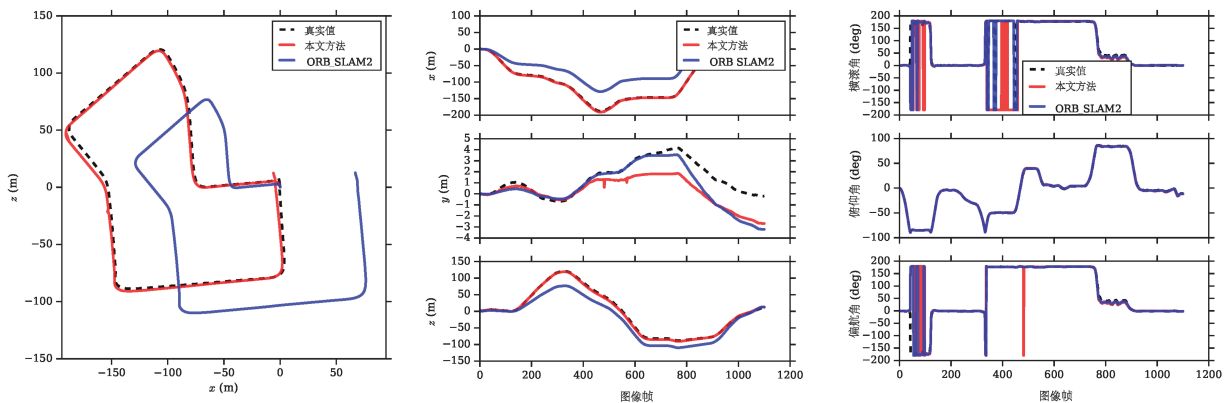


图 3 序列07的轨迹、平移和旋转误差曲线

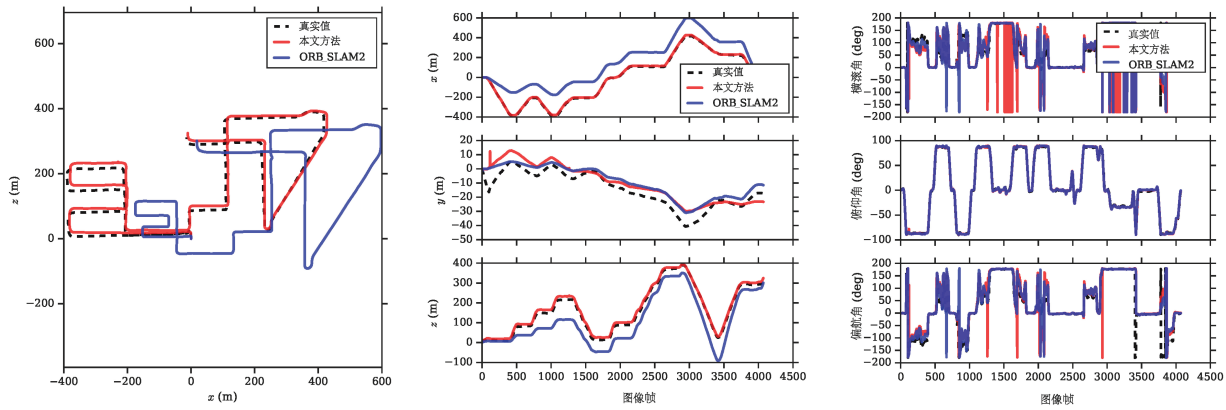


图 4 序列10的轨迹、平移和旋转误差曲线

### 2.3 消融实验

针对本文提出的多规则约束模型, 通过对比MHDP<sup>[10]</sup>并进行消融实验来验证不同规则对于整体系统的影响. 表2展示了KITTI数据集中四个序列的消融实验结果, 选择两个城镇道路(00和07序列)以及两个乡村道路(09和10序列)作为代表, 其中字母“S”“D”“O”和“P”分别表示语义类别、欧氏距离、尺寸朝向(合为一类)、共视特征点四种规则, “+”表示多规则的组合, 即本文方法也可以看作“S+D+O+P”. 由表2可知, “S+D+O”与“O+P”对定位精度的提升有更为明显的影响, 分别提高了49.42%和49.82%, 从而反映出物体关联结果更好.

表 2 消融实验

序列	00	07	09	10	平均	提高/%
S	6.28	2.15	8.58	2.64	4.91	28.70
S+D	6.62	2.87	3.26	2.12	3.72	46.04
S+O	5.96	3.76	3.60	2.91	4.06	41.11
S+P	4.92	2.77	5.35	2.30	3.84	44.34
S+D+O	5.56	3.56	3.01	1.81	3.49	49.42
S+D+P	7.94	2.54	9.71	2.05	5.56	19.30
S+O+P	7.82	3.75	7.02	2.00	5.15	25.29
D	6.53	3.14	3.29	2.48	3.86	43.98
D+O	7.62	2.44	4.73	6.39	5.30	23.15
D+P	8.22	1.71	2.89	3.91	4.18	39.30
D+O+P	9.47	2.04	6.10	3.15	5.19	24.67
O	4.55	3.26	7.23	2.64	4.42	35.85
O+P	4.84	1.51	4.75	2.73	3.46	49.82
P	4.99	1.84	5.21	4.67	4.18	39.37
MHDP	9.14	2.35	9.47	6.59	6.89	0
本文	4.00	1.48	4.11	1.80	3.37	58.63

通过消融实验可以得出如下结果:(1) 多规则约束的效果高于较少或单规则约束的效果;(2) 从单种规则的影响力上看, 欧式距离规则“D”最高, 包含特征点规则“P”次之, 尺寸朝向规则“O”第三, 最后是语义类别规则“S”;(3) 各规则之间会产生相互影响, 即会出现不满足第一条结果的特殊情况, 如“D”规则, 造成这种问题的原因是由于“D”的约束性更强, 应该分配不同的权重, 而权重的比例又来自最终的结果, 即各规则之间无法获得自适应的权重. 目前仅采用最简单的平均权重的方法, 经过大量的实验表明, 虽然伴有少量的随机性, 但整体SLAM系统基本符合上述第一条和第二条结果.

## 3 结论

本文提出一种多规则约束的模型用于语义SLAM的物体关联, 通过物体的语义类别、大小尺寸、朝向角度以及共视特征点的数量约束物体关联, 克服了由于单一策略引起的物体关联模糊问题, 以提高语义SLAM的

定位精度. 通过在真实公共数据集上与ORB-SLAM2和CubeSLAM的对比, 验证了语义SLAM定位精度高于传统SLAM方法, 同时也证明本文提出的基于多规则约束的物体关联的准确性, 能够有效提高物体级语义SLAM的定位精度. 在未来的工作中, 将考虑多规则之间的自适应权重分配问题, 以实现更鲁棒和更高精度的物体关联.

### 参考文献:

- [1] 邢志伟, 朱晓蕊, 何超. 基于行人行为学习的机器人同时定位与可通行区域制图[J]. 机械工程学报, 2019, 55(11): 36-45.
- [2] 韩涛, 黄友锐, 周宁亚, 等. 基于膜计算粒子群优化的FastSLAM算法改进[J]. 新疆大学学报(自然科学版)(中英文), 2020, 37(2): 156-162.
- [3] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [4] BOWMAN S L, ATANASOV N, DANILIDIS K, et al. Probabilistic data association for semantic SLAM[C]. Piscataway: IEEE International Conference on Robotics and Automation(ICRA), IEEE, 2017.
- [5] DOHERTY K, FOURIE D, LEONARD J. Multimodal semantic SLAM with probabilistic data association[C]. Piscataway: IEEE International Conference on Robotics and Automation(ICRA), IEEE, 2019.
- [6] DOHERTY K J, BAXTER D P, SCHNEEWEISS E, et al. Probabilistic data association via mixture models for robust semantic SLAM[C]. Piscataway: IEEE International Conference on Robotics and Automation(ICRA), IEEE, 2020.
- [7] YANG S, SCHERER S. CubeSLAM: monocular 3D object SLAM[J]. IEEE Transactions on Robotics, 2019, 35(4): 925-938.
- [8] NICHOLSON L, MILFORD M, SÜNDERHAUF N. Quadric SLAM: dual quadrics from object detections as landmarks in object-oriented SLAM[J]. IEEE Robotics and Automation Letters, 2019, 4(1): 1-8.
- [9] WU Y, ZHANG Y, ZHU D, et al. EAO-SLAM: monocular semi-dense object SLAM based on ensemble data association[C]. Piscataway: IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), IEEE, 2020.
- [10] RAN T, YUAN L, ZHANG J B, et al. Not only look but infer: multiple hypothesis clustering of data association inference for semantic SLAM[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9.
- [11] ZHANG J B, YUAN L, RAN T, et al. Bayesian nonparametric object association for semantic SLAM[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 5493-5500.
- [12] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. IEEE Computer Vision and Pattern Recognition(CVPR), 2018, <https://doi.org/10.48550/arXiv.1804.02767>.
- [13] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]. Piscataway: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, 2012.
- [14] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

责任编辑: 张自强