

文章编号 1004-924X(2025)08-1313-14

目标检测与光流融合的对象级动态同步定位与 地图构建方法

李曙光¹, 陈沁梅², 史金龙^{1*}, 白素琴³, 王成根¹, 左欣¹

(1. 江苏科技大学 计算机学院, 江苏 镇江 212000;

2. 南京船舶雷达研究所, 江苏 南京 210046;

3. 中国民航大学 民航智慧机场理论与系统重点实验室, 天津 300300)

摘要: 现有大多数视觉同步定位与地图构建(Simultaneous Localization and Mapping, SLAM)方法大部分基于静态环境假设, 导致其在动态环境中的定位精度显著下降。为解决这一问题, 本文提出一种结合目标检测和光流方法的对象级动态SLAM方法。该方法使用目标检测获取对象信息, 结合光流和对象重投影技术来识别对象的动静属性, 并剔除动态对象上的特征点。随后, 寻找检测对象和地图中对象的最佳匹配关系。然后, 在关键帧中优化静态对象, 同时提出一种动态二次曲面优化策略, 用于在对象地图中优化动态二次曲面模型, 并追踪动态对象的运动轨迹。最后, 重建稠密静态背景。在Bonn和TUM数据集上的实验表明, 本文方法的绝对位姿精度提升约44.3%, 相对位姿精度提升约19.0%。实验结果表明, 本文方法在动态场景中能够实现更精确、更稳健的定位。为进一步验证系统的在线性能, 本文还在真实动态场景中对该系统进行了测试, 并达到了预期的结果。

关键词: 动态场景; 同步定位与地图构建; 光流; 二次曲面; 对象检测

中图分类号: TP391 **文献标识码:** A

doi: 10. 37188/OPE. 20253308. 1313

CSTR: 32169. 14. OPE. 20253308. 1313

Object-level dynamic simultaneous localization and mapping method fusing object detection and optical flow

LI Shuguang¹, CHEN Qinmei², SHI Jinlong^{1*}, BAI Suqin³, WANG Chenggen¹, ZHUO Xin¹

(1. School of Computer Science, Jiangsu University of Science and Technology,
Zhenjiang 212000, China;

2. Nanjing Marine Radar Research Institute, Nanjing 210046, China;

3. Key Laboratory of Smart Airport Theory and System, Civil Aviation University of China,
Tianjin 300300, China)

* Corresponding author, E-mail: shi_jinlong@163. com

Abstract: Most existing visual Simultaneous Localization and Mapping (SLAM) systems assume a static environment. This assumption leads to significant degradation in positioning accuracy in dynamic scenes.

收稿日期: 2024-11-11; **修订日期:** 2025-01-15.

基金项目: 浙江省科技计划项目(No. 2024C04056(CSJ)); 国家自然科学基金委员会面上项目(No. 51875270); 中国民航大学民航智慧机场理论与系统重点实验室开放基金资助(No. SATS202207)

To address this limitation, this paper introduced an object-level dynamic SLAM method. The method integrated object detection with optical flow techniques. Object detection was used to acquire detailed semantic information about objects. Optical flow and object reprojection technologies were employed to distinguish between static and dynamic objects. Feature points associated with dynamic objects were subsequently removed. An optimal matching relationship was established between detected objects and map objects. Static objects were optimized within keyframes to improve localization accuracy. A dynamic quadratic surface optimization strategy was introduced. This strategy optimized dynamic quadratic surface models in the object map. It also enabled the tracking of dynamic object trajectories. Finally, the method reconstructed a dense static background. Experiments were conducted on the Bonn and TUM datasets. The results demonstrate significant improvements in accuracy. Absolute pose accuracy improves by 44.3%. Relative pose accuracy improves by 19.0%. These findings confirm that our method can deliver more precise and robust localization in dynamic scenes. To further validate the system's online performance, real-world dynamic scenarios were tested. The experimental results met the expected performance standards. These tests confirmed the system's reliability in practical applications.

Key words: dynamic scenes; SLAM; optical flow; quadratic surface; object detection

1 引言

相较于激光SLAM,视觉SLAM技术以其成本低廉、易于集成及功耗较低等优势,在自动驾驶、移动机器人、增强现实(AR)、虚拟现实(VR)以及无人机等领域展现了显著的应用潜力。视觉SLAM主要利用相机获取场景信息,实现精确的位姿估计和环境建模。然而,传统的视觉SLAM系统大都基于静态场景假设(如参考文献[1]~[4]),其中,ORB-SLAM^[1]通过提取和匹配静态ORB特征点来实现定位与建图,其核心假设是场景中的特征点在连续帧之间保持静止;LSD-SLAM^[2]利用直接法对静态场景进行半稠密重建,其优化过程依赖于场景的静态一致性。而在实际应用中,通常包含动态对象(如行人、移动车辆等),动态对象的存在会导致错误的数据关联,进而影响定位精度和系统的鲁棒性。为了应对动态场景中遇到的挑战,研究人员提出了多种改进方案。

为了应对动态场景中的挑战,研究人员提出了多种改进方案。基于传统框架的动态SLAM算法主要采用纯几何约束或结合其他传感器[5](如惯性测量单元,IMU)来实现动态点和静态点的区分。LC-CRF SLAM^[6]利用GC-RANSAC^[7]提供初始位姿和动态点信息,然后通过条件随机场(Conditional Random Field,

CRF)进一步检测并处理动态点。StaticFusion^[8]和ReFusion^[9]分别使用K均值聚类 and 像素残差计算方法识别动态区域,从而避免在构建地图时受到这些区域的影响。Zhang等人^[10]通过在球面上进行三角化以计算深度信息,并使用贝叶斯滤波方法剔除动态地图点,仅利用静态地图点进行位姿估计。然而,这些方法通常需要对场景有更多的先验知识,并且当动态对象占据图像较大比例时,系统难以有效地去除所有动态干扰因素。

随着深度学习技术的发展,一些研究^[11-15]开始将神经网络与SLAM系统进行集成,利用语义分割、目标检测或光流网络技术增强SLAM系统对动态场景的适应能力。例如,DynaSLAM^[16]结合Mask R-CNN^[17]和多视角几何技术来处理动态点,并修复被遮挡的背景,生成静态的三维地图。RDS-SLAM^[12]在ORB-SLAM3^[18]基础上提出一种可以处理双向关键帧队列的算法,使系统运行速度与传统视觉SLAM相当^[19]。Blitz-SLAM^[20]利用BlitzNet^[21]获得的掩码深度信息和极线约束来分类潜在动态区域内的静态与动态点。OVD-SLAM^[22]使用YOLOv5^[23]检测对象,结合语义信息和深度约束来区分前景和背景,并使用光流识别动态区域,同时计算优化权重以减少动态点的负面影响。IoFusion^[24]利用实例分割与光流技术构

建语义地图,检测并剔除场景中的动态特征点。RSO-SLAM^[25]结合语义分割和光流网络,增强ORB-SLAM2^[3],并使用K-means+Connectivity算法识别运动区域,同时采用光流衰减传播策略优化运动区域的检测与运动概率计算。然而,仅依靠目标检测无法确定潜在动态对象在当前时刻的运动状态;而通过语义信息获得的像素级掩码对于重叠对象不够精确,无法覆盖对象的所有区域,特别是对对象边界部分。

此外,动态对象的表示方法也是研究的重要方向之一,主要可以分为三类:隐式表示、参数化几何模型以及点云聚类方法。DSP-SLAM^[26]利用基于深度学习的符号距离函数(Signed Distance Function, SDF)网络提出一种隐式对象表示法;NodeSLAM^[27]通过网络中的自动编码器构造对象并优化其位姿。参数化模型通常使用基本几何形状如立方体或二次曲面来表示对象^[28]。OA-SLAM^[29]使用二次曲面作为对象模型,并提出一种参数细化策略以识别并重建三维对象模型。此外,VDO-SLAM^[30]和ClusterVO^[31]使用点云聚类来表示对象,其中ClusterVO引入了CRF来聚合属于同一刚体对象的点云。与立方盒和聚类相比,对偶二次曲面可以通过对称矩阵来表示对象。上述具有二次曲面公式的对象SLAM系统在世界坐标中初始化,仅能够用于静态二次曲面,这些方法通常只是简单地移除场景中的动态对象,而不对其进行三维建模或追踪其运动轨迹。因此,在某些特定任务中,当需要利用被剔除的动态信息时,这些策略则显得不足。

为解决上述问题,本文基于ORB-SLAM3算法,提出了一种结合目标检测和光流方法的对象级动态SLAM方法。本文的主要贡献如下:

(1)提出一种高效且实时的动态对象识别方法,通过卡方检验去除光流异常点,检查移动一致性以识别动态点,从而避免求解基本矩阵的大量计算,同时结合对象的重投影进一步识别动态对象。

(2)使用二次曲面模型表示三维场景中的对象,并提出一种动态二次曲面优化策略来优化动态对象,同时追踪其在三维地图中的运动轨迹。

(3)在ORB-SLAM3系统中集成了对象检测和对象优化方法。并在广泛使用的Bonn和TUM基准数据集中进行了评估和方法对比。测试结果表明,本文方法具有出色的定位性能。

(4)使用移动机器人在动态环境中对本文SLAM系统进行了测试,结果表明本文方法能够在复杂的真实世界动态环境中实现可靠定位。

2 系统框架

本文在ORB-SLAM3的基础上进行扩展,新增对象检测、对象关联、对象优化以及稠密建图模块,系统框架如图1所示。与现有方法相比,本文的创新点在于通过对象级的动态处理策略,显著提升动态场景下的定位精度。具体模块设计如下:

(1)对象检测模块:该模块包括Yolov5目标检测、光流检测和对象重投影。本文将场景中检测到的对象分为三类:将频繁移动的对象视为动态对象,直接剔除其特征点;将没有发生移动的对象视为静态对象,保留其稳定的特征,用于位姿优化;将可能移动的对象视为潜在动态对象,使用光流检测和对象重投影技术进一步识别其动静属性。与DynaSLAM^[16]和RDS-SLAM^[12]相比,本文方法通过分层处理策略,有效降低动态对象对系统性能的干扰,同时避免过度剔除特征点的问题。

(2)对象关联模块:该模块的主要任务是对齐地图中的三维对象和目标检测的结果。本文通过匈牙利算法^[32]建立检测对象与地图对象之间的最优匹配关系,能够更准确地实现对象对齐,避免传统方法中因动态对象导致的错误关联问题。与ClusterVO^[31]和VDO-SLAM^[30]相比,本文在对象关联过程中引入了更高效的匹配策略,进一步提升对象关联的准确性。

(3)对象优化模块:该模块包括对象初始化、静态对象优化和动态对象优化。本文提出一种动态二次曲面优化策略,能够在对象地图中优化动态对象的参数模型,并实时追踪其运动轨迹。与EAO-SLAM^[28]和OA-SLAM^[29]相比,本文的动态二次曲面优化策略不仅提高了系统的定位

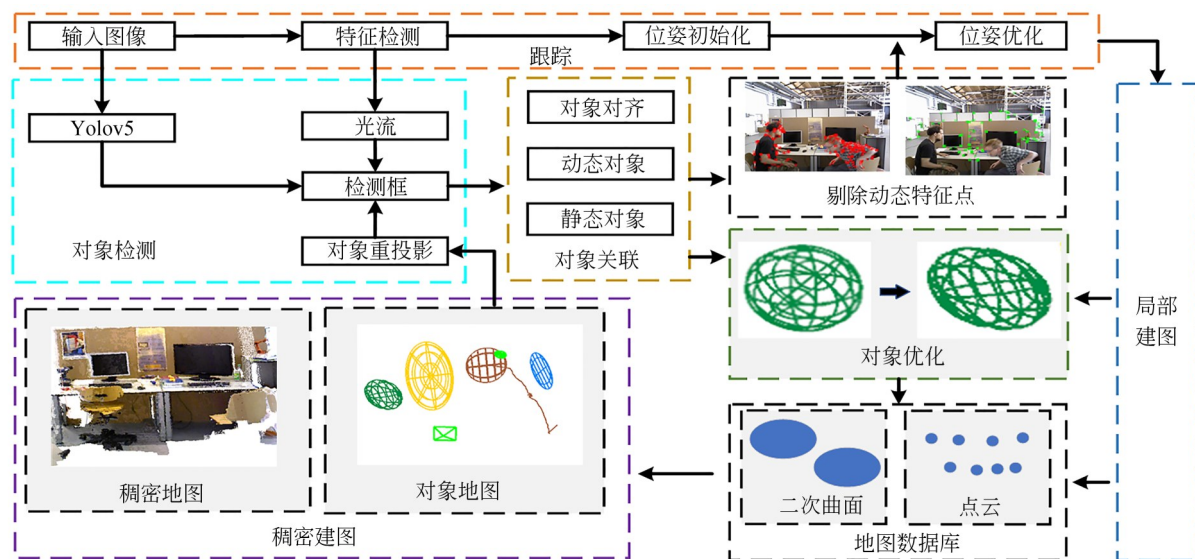


图 1 系统框架图

Fig. 1 System framework diagram

精度,还为动态对象的建模与追踪提供了新思路。

(4)稠密建图模块:该模块用于构建场景的静态稠密背景和对象地图。本文通过分离静态背景和动态对象,实现了更高质量的稠密建图。具体而言,静态背景通过多视图立体匹配技术进行重建,而动态对象则通过二次曲面模型进行表示和优化。

3 对象检测

系统将 RGB-D 图像输入到对象检测模块,首先使用目标检测网络获取对象信息,同时将该图像提取到的特征点和对象关联;然后使用光流检测来识别动态点,并结合对象重投影方法来判断是否为动态对象。本文的对象检测策略可以准确识别出动态对象并剔除动态特征点,显著减少动态特征对系统的影响,同时保留更多稳定的静态特征用于位姿优化。

3.1 光流检测

光流检测通过识别光流向量中的异常点来判定对象属性,从而避免计算基础矩阵,并减少计算量。首先,通过目标检测网络获取对象的检测框,本文将检测框外的特征点视为静态点,同时将检测到的对象视为潜在动态对象;然后通过判断特征点是否位于检测框内,将特征点和检测

到的对象进行关联,使用光流检测每个对象上的特征点,通过检查特征点的光流向量与场景整体运动的一致性,判断特征点是否为异常点(动态点),若对象上的动态点数量大于 30%,则将该对象视为动态对象,否则是静态对象。OVD-SLAM 采用 10% 作为动态点判定阈值,虽然能够更快地识别动态对象,但也可能导致误判,尤其是在光流检测不稳定的情况下。本文通过实验将阈值设置为 30%,能够在保证动态对象识别准确性和计算速度的同时,减少误判率。具体的方法如下:

通过光流获取 $t-1$ 帧和 t 帧对象上匹配的特征点 $F^{t-1}=\{f_i^{t-1}|i=1,2,\dots,L\}$ 和 $F^t=\{f_i^t|i=1,2,\dots,L\}$, L 表示成功跟踪的特征点的数量,选取一定数量的匹配静态点和检测对象上的匹配特征点构成高斯分布模型,然后利用卡方检验判断动态点。

本文将坐标点的协方差矩阵定义为:

$$\Sigma = \begin{bmatrix} (s^2)^m & 0 \\ 0 & (s^2)^m \end{bmatrix}. \quad (1)$$

在本文中,设 $s=2, m=3$ 。

将两个匹配点之间向量的卡方值定义为 δP :

$$\delta P = (P - \frac{1}{L} \sum_{i=1, t}^L P_{i-1, t}^{\text{static}}) \Sigma^{-1} ((P - \frac{1}{L} \sum_{i=1, t}^L P_{i-1, t}^{\text{static}})), \quad (2)$$

其中: P 表示 $t-1$ 帧和 t 帧中第 i 对匹配点之间的向量,以 f_i^{t-1} 为起点, f_i^t 为终点; P_{t-1}^{static} 表示两个匹配的静态点之间的向量。设阈值 $P_{\text{th}} = x_{0.95}^2$,当 $\delta P > P_{\text{th}}$,为动态点,反之,为静态点。

3.2 对象重投影

对象重投影通过将三维地图中的现有对象投影到当前帧并计算重投影误差,以更准确地识别动态对象,从而避免漏检和错检。

首先,通过目标检测网络获取第 t 帧对象检测框 $BBox_i^t$ (其中 $i \in \{0 \cdots M\}$, M 为检测到的对象的数量);

然后,将对象地图中的第 j 个对象 O_j 的二次曲面模型 Q_j^* (其中 $j \in \{0 \cdots N\}$, N 表示对象地图中对象的数量, Q^* 表示二次曲面的参数模型,如公式3所示)投影到当前帧中,得到投影二次曲线 $C_j^* = \{a_x, a_y, \theta_x, \theta_y, \gamma_x, \gamma_y\}$,其中 a, θ, γ 分别表示轴长,方向,中心点等6个参数, x, y, z 表示三个坐标轴, C^* 如公式(5)所示:

$$Q^* =$$

$$\{a_x, a_y, a_z, \theta_x, \theta_y, \theta_z, \gamma_x, \gamma_y, \gamma_z, \xi_{t,t-1}\}^T \in R^{15 \times 1}, \quad (3)$$

其中, Q^* 共有15个自由度,包含轴长、方向、中心点坐标以及从 $t-1$ 时刻到 t 时刻的位姿 $\xi_{t,t-1}$,如式(4)所示:

$$\xi_{t,t-1} = \{\tau_x, \tau_y, \tau_z, \partial_x, \partial_y, \partial_z\}^T \in R^{6 \times 1}, \quad (4)$$

其中: τ 表示平移, ∂ 表示旋转。

$$C^* = K T_c^w Q^* (K T_c^w)^T = \text{proj}(Q^*), \quad (5)$$

其中: K 为相机内参, T_c^w 为相机位姿。

最后计算 C_j^* 的外接包围框和 $BBox_i^t$ 的重投影误差 E_{object}

$$E_{\text{object}} = \text{IoU}(BBox_i^t, B(\text{proj}(Q_j^*))), \quad (6)$$

其中:函数 IoU 表示二维区域的交并比;函数 $\text{proj}(Q_j^*)$ 表示计算 Q_j^* 的投影二次曲线; B 表示计算外接包围框。

设动态检测阈值为 E_{th} :若 $E_{\text{object}} < E_{\text{th}}$ 则将 O_i 视为动态对象,否则为静态对象。OA-SLAM将对象的重投影误差阈值设为0.3,考虑到动态对象的影响,在经过实验后,将 E_{th} 设为0.9。

4 对象关联

对象关联模块的核心任务是通过建立检测对象与地图对象之间的最优匹配关系,确保数据

关联的准确性和一致性。本文引入一种更高效的匹配策略,能够有效避免因动态对象导致的误匹配问题。该模块在本文方法中起到关键作用,不仅确保了检测对象与地图对象之间的准确对应关系,还为后续的对象优化提供可靠的基础。具体做法是,在对象检测完成后,系统使用匈牙利算法将对象地图中的二次曲面模型和二维图像上的对象进行关联,同时剔除动态对象上的特征点,保留稳定的特征用于优化位姿。对象关联过程如下:

通过代价矩阵 $A \in R^{M \times N}$ 找到 M 个二次曲面 Q_j^* 和 N 个二维检测框 $BBox_i$ 之间的最佳匹配关系,代价矩阵 A 由具有不同加权距离的元素 a_{ij} 组成,如式(7)所示:

$$a_{ij} = \lambda_1 a_{ij}^s + \lambda_2 a_{ij}^d. \quad (7)$$

参考文献[33],将 λ_1 和 λ_2 分别设置为2和1; a_{ij}^s 表示检测框内点距离,为在当前帧检测框 $BBox_i^t$ 内光流特征集 $\{x_j^{t-1}\}$ 的数量和特征集 $\{x_i^t\}$ 数量的比值,如式(8)所示:

$$a_{ij}^s = \frac{\text{size}(\text{In}(BBox_i^t, x_j^{t-1}))}{\text{size}(\{x_i^t\})}, \quad (8)$$

其中: $\text{size}()$ 表示计算数量; $\text{In}(BBox_i^t, x_j^{t-1})$ 表示检查特征 x_j^{t-1} 是否位于 $BBox_i^t$ 中。

a_{ij}^d 表示 Q_j^* 的投影二次曲线的外接包围框和 $BBox_i^t$ 之间的IoU距离,如式(9)所示:

$$a_{ij}^d = \text{IoU}(BBox_i^t, B(\text{proj}(Q_j^*))). \quad (9)$$

5 对象优化

对象优化模块根据对象关联的结果来优化对象地图中的二次曲面模型。对于未关联的对象,系统将其视为新对象并进行初始化;对于静态对象,系统在位姿更加准确的关键帧中优化;对于动态对象,系统通过动态二次曲面优化策略优化其参数,并在对象地图中追踪其运动轨迹。对象优化的主要任务是优化二次曲面的参数,并对动态对象进行运动轨迹追踪。本文通过优化静态对象,能够提升系统定位精度;通过动态二次曲面优化策略,能够实时追踪动态对象的运动轨迹,并优化其参数模型。

5.1 二次曲面初始化

本文首先使用等轴长(球体)初始化二次曲面,然后随着观测到更多的检测,不断细化二次

曲面。其中,二次曲面中心点通过检测框的中心进行三角测量获得;半轴长 a 由反向投影到二次曲面中心位置的包围框大小确定,如式(10):

$$a = \frac{1}{4l} \sum_{i=1}^l d_i^i \left(\frac{w_i^i}{f_x} + \frac{h_i^i}{f_y} \right) I^{3 \times 3}, \quad (10)$$

其中: d_i^i, w_i^i, h_i^i 分别表示第 t 帧中第 i 个对象检测框中心的深度,像素的宽度和高度; f_x, f_y 表示相机的焦距; l 表示要初始化的对象相关联的观测帧的数量,本文设 $l=15$,即只有当对象被观测到 15 帧以上才会被初始化; $I^{3 \times 3}$ 表示 3×3 的单位矩阵。

5.2 静态二次曲面优化

静态对象优化是利用具有精确相机位姿的关键帧来细化二次曲面的参数,本文计算检测框内接二次曲线和相关联的三维对象的投影二次曲线之间的 Wasserstein 距离,并将其作为目标函数。具体的做法是:

首先,对于存储在对象地图中的二次曲面模型 Q^* ,通过式(3)获得投影 C^* ,并将其描述为 2D 高斯分布 $N(\mu, \Sigma)$,如式(11)所示:

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = 1, \quad (11)$$

其中: $\mu = (\gamma_x, \gamma_y)^T$ 表示二次曲线的中心; Σ^{-1} 计算方法如式(12)所示:

$$\Sigma^{-1} = R(\theta) \begin{pmatrix} \frac{1}{a_x^2} & 0 \\ 0 & \frac{1}{a_y^2} \end{pmatrix} R(\theta)^T. \quad (12)$$

然后,计算三维对象的投影二次曲线和检测框内接二次曲线的 2D 高斯分布 $N_1(\mu_1, \Sigma_1)$ 和 $N_2(\mu_2, \Sigma_2)$ 之间的 Wasserstein 距离公式,如式(13)所示:

$$w_2^2(N_1, N_2) = \|\mu_1, \mu_2\|_2^2 + Tr(\Sigma_1 + \Sigma_2 - 2(\Sigma_1^{\frac{1}{2}} \Sigma_2 \Sigma_1^{\frac{1}{2}})^{\frac{1}{2}}), \quad (13)$$

其中, $\|\cdot\|_2$ 表示计算二范式。

最后,构造重投影误差 E_{distance} ,优化二次曲线参数并更新地图中的对象,如式(14)所示:

$$E_{\text{distance}} = \sum_{j=1}^l s_j^{-1} w_2^2(\pi(BBox_j), \text{proj}(Q_j^*)), \quad (14)$$

其中: s_j 表示关键帧中第 j 个对象的检测得分;

$\pi(BBox_j)$ 表示计算检测框的内接二次曲线; l 表示关键帧数量。

5.3 动态二次曲面优化

动态二次曲面的模型参数通过一个变换矩阵 T 来更新,如式(15)所示:

$$Q^* = T Q^* T^T. \quad (15)$$

本文在动态二次曲面优化框架中优化二次曲面参数和对象位姿,主要使用 SLAM 中常用的图优化方法。图优化可以拆分为两个任务:构建图和优化图,构建的图主要由顶点(vertex)和边(edge)组成,一般将优化项作为顶点,约束项作为边;优化图就是调整优化项,尽量去满足边的约束关系。在动态二次曲面优化框架中,将变换矩阵 T 作为顶点,将运动模型约束、二次曲面重投影约束以及二次曲面尺寸约束作为边,动态二次曲面优化的核心就是调整变换矩阵 T ,尽可能的满足由上述三个约束组成的边。

5.3.1 运动模型约束

类似于 ORB-SLAM3,在连续帧观测中使用恒速运动模型,这种先验知识有助于估计更平滑的运动轨迹,将连续帧对象变换矩阵定义为 $H_{t-1,t}^w$,如式(16)所示:

$$H_{t-1,t}^w = T_t^{o,w} T_{t-1}^{w,o}, \quad (16)$$

其中: $T_t^{o,w}$ 表示第 t 幅图像在世界坐标系中观测到的对象的位姿; $T_{t-1}^{w,o}$ 表示对象坐标系到世界坐标系的变换矩阵。

将运动模型约束能量函数定义为 E_{motion} ,如公式(17)所示:

$$E_{\text{motion}} = \|\log((H_{t-1,t}^w)^{-1} H_{t,t+1}^w)\|_2 \quad (17)$$

5.3.2 二次曲面重投影约束

获取 Q^* 的投影并计算外接包围框,然后通过式(18)计算二次曲面重投影约束能量函数 E_{proj} :

$$E_{\text{proj}} = \|B - B(\text{proj}(Q^*))\|_2. \quad (18)$$

5.3.3 二次曲面尺寸约束

本文通过使用 5.1 节中的对象先验尺寸信息来约束二次曲面。

$$E_{\text{size}} = \|D - \tilde{D}\|_2, \quad (19)$$

其中: $D \in \mathbf{R}^{3 \times 3}$ 是由二次曲面轴长的平方组成的对角矩阵; $\tilde{D}$ 是由二次曲面初始化的半轴长 a 的平方组成的对角矩阵。

5.3.4 联合优化

通过式(20)在动态二次曲面优化框架中构造联合约束能量函数 E_{Dynamic} ,联合优化动态二次曲面。

$$E_{\text{Dynamic}} = E_{\text{motion}} + E_{\text{proj}} + E_{\text{size}}. \quad (20)$$

6 实验对比与分析

本文实验平台配置为 AMD Ryzen 7 6800H CPU(3.20 GHz),16 G 内存和 RTX 3060 GPU 卡,操作系统采用 Ubuntu 20.04。使用由 TensorRT 加速的 YOLOv5 算法作为独立线程专用于目标检测。在真实场景的实验中,本文分别使用手持相机和将相机搭载在阿克曼小车采集数据,两种方式均使用 Realsense D435i RGB-D 相机。

6.1 实验数据集

本文使用 Bonn 数据集和 TUM 数据集进行实验,这两个数据集是广泛用于测试动态 SLAM 的公开数据集,且都是室内场景,并提供每帧相机位姿的真值。两个数据集中都既包含非刚性对象(如人体),也包含刚性对象(如椅子、气球、小车、箱子等)。

Bonn 数据集包含 24 个序列,包含人体走动、人与箱子交互、人与气球交互等多种动态场景。其中 Balloon 系列数据集中的动态对象主要是人和气球,Balloon Tracking 系列涉及气球的运动轨迹追踪;Crowd 和 Synchronous 系列数据集涉及多人在场景中的运动,由于人群密集,动态物体占据图像的大部分区域;Box 系列数据集涉及从场景中移除盒子或者从场景外移入盒子等。

TUM 数据集包含 8 个动态序列,根据人体的运动程度划分为 4 个低动态场景(标记为“s”,代表静止)和 4 个高动态场景(标记为“w”,代表行走)。在高动态场景中,人在办公室内走动导致较大的运动区域;相反,在低动态场景中,人坐在椅子上进行交流,产生较小的运动区域。

6.2 相机位姿精度对比与分析

相机位姿精度是衡量 SLAM 系统性能的重要指标。本文采用绝对轨迹误差(ATE-RMSE)

和相对轨迹误差(RPE-RMSE)评估相机位姿精度,其中绝对轨迹误差通常用于评估相机轨迹的全局一致性,反映每一帧相机位姿估计值与真值之间的差异;相对轨迹误差通过量化局部旋转或平移误差来评估预测的轨迹精度,测量在连续帧之间得出的相对位姿估计值与真实位姿之间的不一致程度;RMSE 被用作每个误差度量的评估指标。

首先与其他算法进行对比,实验结果如表 1~表 3 所示,表中加粗数据表示该序列的相机位姿精度最优值,TOP-1 表示最优的序列数量。

如表 1 所示,在 Bonn 数据集上,将本文方法与 ORB-SLAM3^[18],OA-SLAM^[29],OVD-SLAM^[22],IoFusion^[24]和 RSO-SLAM^[25]进行对比(分别用 O3,OA,OS,IF 和 RS 表示),本文方法在 13 个动态序列中达到了最先进的位姿精度。具体而言,本文在 Crowd 和 Synchronous 系列数据集上均取得了最优的位姿精度,表明本文方法对动态场景占据图像大部分区域时的处理能力更加优越;在 Box 的 11 个数据集中有 7 个取得最优的位姿精度,表明本文方法在针对盒子的动态对象上取得较好的结果;但是在 Balloon 数据集上,RSO-SLAM 优于本文,是因为该方法采用光流网络和语义分割网络,可以更精确地剔除动态对象上的特征,使得位姿计算更加准确。

如表 2~表 3 所示,在 TUM 数据集上,将本文方法与 Dyna-SLAM^[16],ORB-SLAM3^[18],Blitz-SLAM^[20],OVD-SLAM^[22]和 RSO-SLAM^[25]进行对比(分别用 DS,O3,BS,OS 和 RS 表示)。实验结果表明,本文方法在高动态和低动态的场景中均表现出显著的定位精度,在 TUM 的 8 个序列上实现了最优的绝对轨迹误差和相对轨迹误差结果,优于 RSO-SLAM,同时表明本文方法在复杂的动态场景(Walking 系列)的优越性。这归因于本文能够准确地识别场景中的动态对象,剔除动态对象上的特征,保留更准确的静态特征,使参与相机位姿优化的特征点更可靠,提高位姿精度。

实验还计算 RSO-SLAM 和本文方法在 Bonn 数据集和 TUM 数据集上的绝对轨迹误差

表 1 Bonn数据集 ATE 对比

Tab. 1 Comparison of ATE on the Bonn Dataset (m)

Sequence	O3 ^[18]	OA ^[29]	OS ^[22]	IF ^[24]	RS ^[25]	Ours
	2021	2022	2023	2024	2024	/
Balloon	0.042	0.254	0.036	0.028	0.025	0.031
Balloon2	0.072	0.219	0.035	0.029	0.024	0.031
Balloon Tracking	0.038	0.151	0.039	0.041	0.030	0.037
Balloon Tracking2	0.075	0.262	0.078	0.059	0.029	0.073
Crowd	0.386	0.144	0.020	0.027	0.017	0.016
Crowd2	0.704	0.080	0.024	0.028	0.023	0.021
Crowd3	0.339	0.156	0.026	0.038	0.031	0.023
Kidnapping	0.030	0.384	0.029	0.029	0.025	0.028
Kidnapping2	0.026	0.348	0.025	0.025	0.024	0.023
Moving No Box	0.469	0.288	0.024	0.025	0.020	0.025
Moving No Box2	0.041	0.367	0.033	0.037	0.033	0.041
Moving O Box	0.364	0.436	0.220	0.262	0.266	0.058
Moving O Box2	1.063	0.391	0.352	0.134	0.269	0.077
Person Tracking	0.422	0.519	0.051	0.041	0.037	0.044
Person Tracking2	0.834	0.636	0.055	0.056	0.037	0.061
Placing No Box	0.773	0.148	0.022	0.021	0.016	0.019
Placing No Box2	0.023	0.104	0.021	0.022	0.018	0.018
Placing No Box3	0.040	0.187	0.027	0.043	0.025	0.024
Placing O Box	0.320	0.278	0.246	0.177	0.248	0.051
Removing No Box	0.023	0.135	0.022	0.017	0.015	0.021
Removing No Box2	0.023	0.132	0.022	0.025	0.020	0.020
Removing O Box	0.384	0.147	0.357	0.197	0.288	0.057
Synchronous	1.072	0.084	0.012	0.012	0.010	0.006
Synchronous2	1.250	0.066	0.007	0.009	0.006	0.006
TOP-1	0	0	0	0	12	13

表 2 TUM数据集 ATE 对比

Tab. 2 Comparison of ATE on the TUM Dataset (m)

Sequence	DS ^[16]	O3 ^[18]	BS ^[20]	OS ^[22]	RS ^[25]	Ours
	2018	2021	2022	2023	2024	/
fr3/s/static	/	0.007	/	0.008	0.006	0.006
fr3/s/xyz	0.013	0.011	0.015	0.013	0.009	0.008
fr3/s/half	0.019	0.019	0.016	0.023	0.015	0.015
fr3/s/rpy	/	0.026	/	0.041	0.019	0.019
fr3/w/static	0.007	0.014	0.010	0.006	0.006	0.005
fr3/w/xyz	0.016	0.115	0.015	0.017	0.015	0.013
fr3/w/half	0.030	0.158	0.026	0.025	0.023	0.019
fr3/w/rpy	0.035	0.535	0.036	0.033	0.033	0.029
TOP-1	0	0	0	0	3	8

表 3 TUM 数据集 RPE 对比

Tab. 3 Comparison of RPE on the TUM Dataset (m)

Sequence	DS ^[16]	O3 ^[18]	BS ^[20]	OS ^[22]	RS ^[25]	Ours
	2018	2021	2022	2023	2024	/
fr3/s/static	/	0.008	/	/	0.008	0.006
fr3/s/xyz	0.014	0.013	0.014	0.012	0.012	0.008
fr3/s/half	0.024	0.023	0.017	0.021	0.019	0.015
fr3/s/rpy	/	0.032	/	0.062	0.024	0.022
fr3/w/static	0.009	0.020	0.013	0.008	0.008	0.005
fr3/w/xyz	0.022	0.070	0.020	0.018	0.019	0.013
fr3/w/half	0.028	0.092	0.025	0.024	0.026	0.019
fr3/w/rpy	0.045	0.206	0.047	0.045	0.042	0.040
TOP-1	0	0	0	0	0	8

表 4 精度提升-ATE 和 RPE

Tab. 4 Precision improvement-ATE and RPE (m)

Method	ATE	RPE
RSO-SLAM ^[26]	1.662	0.158
Ours	0.925	0.128
Accuracy Enhancement	44.3%	19.0%

ATE 之和,以及相对轨迹误差 RPE 之和。如表 4 所示,与 RSO-SLAM 相比,本文方法的绝对位姿精度提升约 44.3%,相对位姿精度提升约

19.0%。

此外,图 2 和图 3 分别展示了本文方法(第 2 行)和经典算法 ORB-SLAM3(第 1 行)在 Bonn 数据集和 TUM 数据集部分序列的绝对轨迹误差 (ATE)结果图和相对估计误差(RPE)结果图。其中,黑色线条代表数据集的相机位姿真值;蓝色线条代表估计的相机位姿;红色线条代表两者误差(彩图见期刊电子版)。

由图 2 和图 3 可以看出,与 ORB-SLAM3 相比,本文方法估计的相机位姿与真实轨迹的偏差

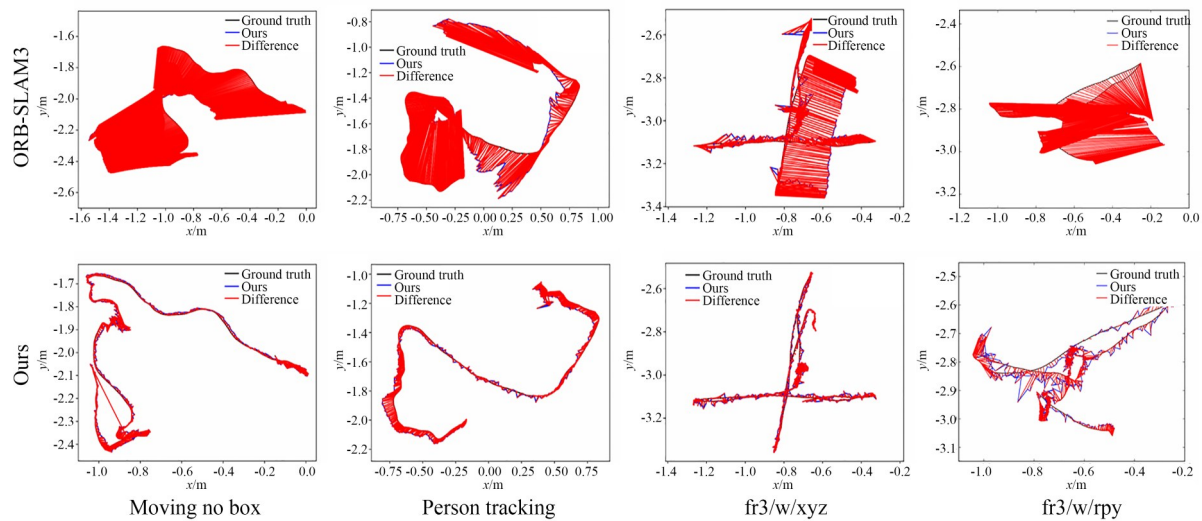


图 2 ORB-SLAM3 和本文方法的绝对轨迹误差图

Fig. 2 Absolute trajectory error diagram for ORB-SLAM3 and ours

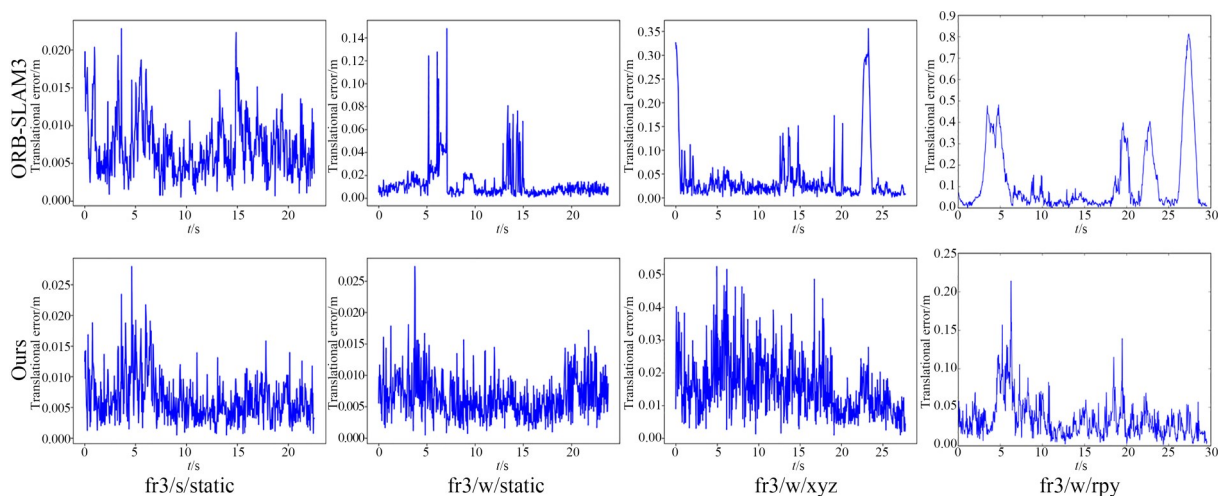


图3 ORB-SLAM3和本文方法的相对轨迹误差图

Fig. 3 Relative pose error diagram for ORB-SLAM3 and ours

均较小,再次验证了本文方法在各场景中的优越表现。

6.3 二次曲面优化实验分析

本实验使用 Bonn 数据集 Removing No Box 序列进行效果展示,将本文方法和 OA-SLAM 进行对比,结果如图 4 所示,分别展示了两种方法在对象移动前、后的结果。其中,第 1 行为三维地图中的二次曲面模型;第 2 行为将二次曲面投影到当前帧上的效果图。由于人是非刚性对象,容易发生大的变形,不利于构建二次曲面优化,因此实验中没有构建人的二次曲面模型;此外,实验

将人搬动的箱子视为动态对象,其余如椅子、设备等视为静态对象。

从图 4 可以看出,对于静态对象,OA-SLAM 重建的对象数量明显少于本文方法;对于动态对象,OA-SLAM 追踪失败,而本文方法不仅追踪成功,并在对象地图中绘制了其运动轨迹;此外,在构建二次曲面模型时,OA-SLAM 出现了明显的位置和尺寸误差,而本文方法能够准确计算对象的位姿,从而在地图中提供了更为精确的对象模型表示。实验结果验证了动态二次曲面优化策略的有效性和优越表现。

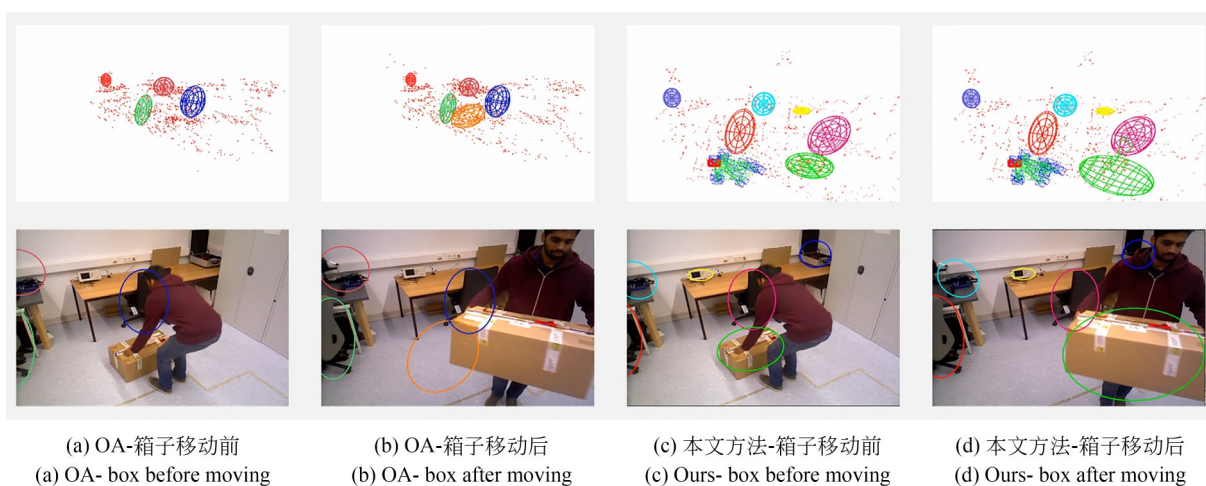


图4 二次曲面优化-Removing No Box

Fig. 4 Quadratic surface optimization - removing no box

6.4 系统运行时间分析与对比

由于不同序列中每帧的跟踪时间不同,因

此,本实验比较了几种动态 SLAM 算法的平均耗时,结果如表 5 所示。另外,本文使用了高效且实

时的动态对象识别方法,避免求解基本矩阵的大量计算,使得系统运行帧率维持在 34~41 fps 之间,表明本方法可以实时运行。

表5 本文方法和动态SLAM时间对比

Tab. 5 Time comparison between proposed method and dynamic SLAM (ms)

Method	Experimental Platform	Average Duration
Dyna-SLAM ^[17]	CPU+TeslaM40	>350
OVD-SLAM ^[23]	i7+RTX2060	27.8
RSO-SLAM ^[26]	i7+RTX3070	45.6
Ours	R7+RTX3060	26.4

6.5 稠密建图

本实验构建了 Bonn 数据集和 TUM 数据集的静态稠密点云地图。如图 5 所示,本文方法可以有效去除动态对象对稠密地图的影响,重建出准确的静态稠密地图,进一步证明了系统的准确性和稳定性。

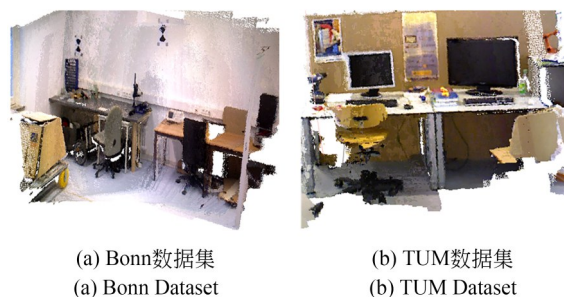


图5 稠密静态地图

Fig. 5 Dense Static Map

6.6 真实场景

为了评估本文方法在真实场景中的效果,本文拍摄了两段不同的序列。

图 6 将 D435i 相机搭载在阿克曼小车上,拍摄了实验室过道里场景。其中,图 6(b)检测并剔除了动态对象(主要是人)的特征;图 6(c)显示了整个大场景的相机轨迹图。图 7 为手持 D435i 在室内实验室拍摄的动态场景,包括移动的行人,搬动的箱子等动态对象,也包括椅子、显示器等静态对象。图 7 两行分别显示了对象地图和投影效果图。

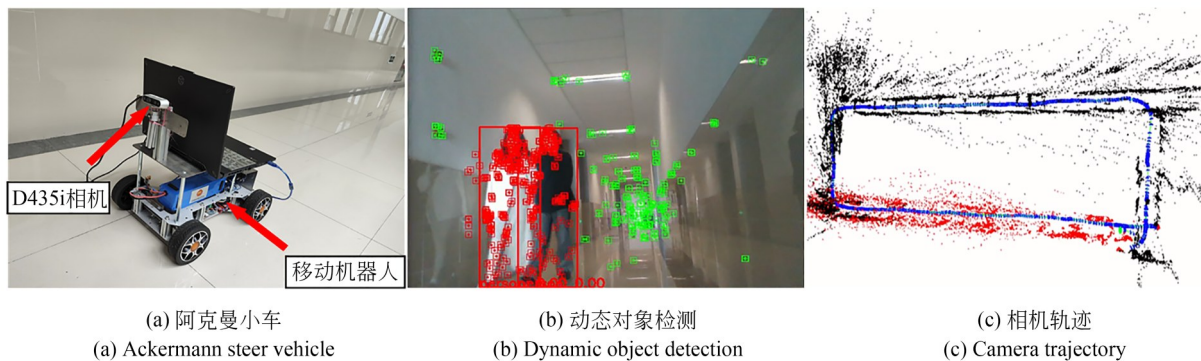


图6 真实场景-序列 01

Fig. 6 Real Scene-Sequence 01

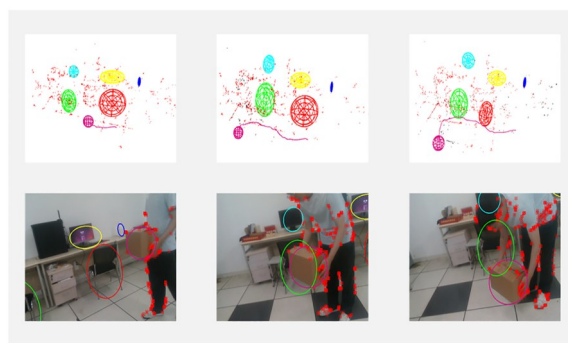


图7 真实场景-序列 02

Fig. 7 Real Scene-Sequence 02

由图 6 和图 7 可以看出,在真实场景中,本文方法可以成功检测到动态对象,并去除其动态特征点;如图 6(c)所示,在大场景的定位测试中,相机在最后成功检测到回环并优化相机轨迹;如图 7 所示,在真实场景中,可以准确追踪动态对象的运动轨迹,验证了动态二次曲面优化策略的优越表现。

7 结 论

本文提出的对象级动态 SLAM 系统展示了

在动态场景下的显著优势,通过结合目标检测、光流算法和对象重投影,系统能够有效地识别动态对象并排除动态特征的影响。该方法不仅提高了在复杂动态场景中的定位准确性,而且增强了系统的整体性能。特别的,本文提出的一种动态二次曲面优化策略,能够在构建的对象地图中优化二次曲面模型,从而更好地追踪动态对象。在 Bonn 和 TUM 标准数据集上的测试表明,与最先进的 RSO-SLAM 相比,绝对位姿精度和相对位姿精度提升分别为 44.3% 和 19.0%。由此可以验证本方法的准确性和定位精度,同时本文还将本方法应用到真实场景,证明了本方法在实际应用中的潜力。

参考文献:

- [1] 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.
- [2] ENGEL J, SCHÖPS T, CREMERS D. *LSD-SLAM: Large-Scale Direct Monocular SLAM* [M]. Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 834-849.
- [3] 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.
- [4] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(3): 611-625.
- [5] 贾晓雪,赵冬青,张乐添,等. 基于自适应惯导辅助特征匹配的视觉 SLAM 算法[J]. *光学精密工程*, 2023, 31(5): 621-630.
JIA X X, ZHAO D Q, ZHANG L T, *et al.* A visual SLAM algorithm based on adaptive inertial navigation assistant feature matching[J]. *Opt. Precision Eng.*, 2023, 31(5): 621-630. (in Chinese)
- [6] DU Z J, HUANG S S, MU T J, *et al.* Accurate dynamic SLAM using CRF-based long-term consistency[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2022, 28(4): 1745-1757.
- [7] BARATH D, MATAS J. Graph-cut RANSAC [C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018: 6733-6741.
- [8] SCONE R, JAIMEZ M, PETILLOT Y R, *et al.* StaticFusion: background reconstruction for dense RGB-D SLAM in dynamic environments[C]. 2018 *IEEE International Conference on Robotics and Automation (ICRA)*. May 21-25, 2018, Brisbane, QLD, Australia. IEEE, 2018: 3849-3856.
- [9] PALAZZOLO E, BEHLEY J, LOTTES P, *et al.* ReFusion: 3D reconstruction in dynamic environments for RGB-D cameras exploiting residuals[C]. 2019 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019: 7855-7862.
- [10] 张裕,张越,张宁,等. 基于逆深度滤波的双目折反射全景相机动态 SLAM 系统[J]. *光学精密工程*, 2022, 30(11): 1282.
ZHANG Y, ZHANG Y, ZHANG N, *et al.* Dynamic SLAM of binocular catadioptric panoramic camera based on inverse depth filter[J]. *Opt. Precision Eng.*, 2022, 30(11): 1282. (in Chinese)
- [11] RUNZ M, BUFFIER M, AGAPITO L. MaskFusion: real-time recognition, tracking and reconstruction of multiple moving objects [C]. 2018 *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. October 16-20, 2018, Munich, Germany. IEEE, 2018: 10-20.
- [12] LIU Y B, MIURA J. RDS-SLAM: real-time dynamic SLAM using semantic segmentation methods [J]. *IEEE Access*, 2021, 9: 23772-23785.
- [13] DAI W C, ZHANG Y, LI P, *et al.* RGB-D SLAM in dynamic environments using point corre-

作者贡献声明:

李曙光:论文构思,方法论,编程和撰写;
陈沁梅:获取资助;
史金龙:项目管理和提供资源;
白素琴,左欣:审核与编辑写作;
王成根:调查研究和指导。

- lations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 373-389.
- [14] HU X G, ZHANG Y Z, CAO Z Z, *et al.* CFP-SLAM: a real-time visual slam based on coarse-to-fine probability in dynamic environments[C]. 2022 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. October 23-27, 2022, Kyoto, Japan. IEEE, 2022: 4399-4406.
- [15] 王成根, 史金龙, 诸皓伟, 等. 基于实例分割与光流的动态场景RGB-D SLAM[J]. *光学精密工程*, 2024, 32(6): 857-867.
- WANG C G, SHI J L, ZHU H W, *et al.* RGB-D SLAM method of dynamic scene based on instance segmentation and optical flow [J]. *Opt. Precision Eng.*, 2024, 32(6): 857-867. (in Chinese)
- [16] BESCOS B, FÁCIL J M, CIVERA J, *et al.* DynaSLAM: tracking, mapping, and inpainting in dynamic scenes[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4): 4076-4083.
- [17] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN[C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2980-2988.
- [18] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, *et al.* ORB-SLAM3: an accurate open-source library for visual, visual - inertial, and multimap SLAM [J]. *IEEE Transactions on Robotics*, 2021, 37(6): 1874-1890.
- [19] JI T T, WANG C, XIE L H. Towards real-time semantic RGB-D SLAM in dynamic environments [C]. 2021 *IEEE International Conference on Robotics and Automation (ICRA)*. May 30 - June 5, 2021, Xi'an, China. IEEE, 2021: 11175-11181.
- [20] FAN Y C, ZHANG Q C, TANG Y L, *et al.* Blitz-SLAM: a semantic SLAM in dynamic environments [J]. *Pattern Recognition*, 2022, 121: 108225.
- [21] DVORNIK N, SHMELKOV K, MAIRAL J, *et al.* BlitzNet: A real-time deep network for scene understanding[C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 4174-4182.
- [22] HE J M, LI M R, WANG Y Y, *et al.* OVD-SLAM: an online visual SLAM for dynamic environments [J]. *IEEE Sensors Journal*, 2023, 23(12): 13210-13219.
- [23] JOCHER G, CHAURASIA A, STOKEN A, *et al.* V7.0-YOLOV5 SOTA realtime instance segmentation[J]. *Zenodo*, 2022.
- [24] ZHU H W, BAI S Q, SHI J L, *et al.* IOFusion: instance segmentation and optical-flow guided 3D reconstruction in dynamic scenes [J]. *The Visual Computer*, 2025, 41(2): 799-813.
- [25] QIN L, WU C, CHEN Z Y, *et al.* RSO-SLAM: a robust semantic visual SLAM with optical flow in complex dynamic environments [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(10): 14669-14684.
- [26] WANG J W, RÜNZ M, AGAPITO L. DSP-SLAM: object oriented SLAM with deep shape priors [C]. 2021 *International Conference on 3D Vision (3DV)*. December 1-3, 2021, London, United Kingdom. IEEE, 2021: 1362-1371.
- [27] SUCAR E, WADA K, DAVISON A. NodeSLAM: neural object descriptors for multi-view shape reconstruction [C]. 2020 *International Conference on 3D Vision (3DV)*. November 25-28, 2020, Fukuoka, Japan. IEEE, 2020: 949-958.
- [28] WU Y M, ZHANG Y Z, ZHU D L, *et al.* EAO-SLAM: monocular semi-dense object slam based on ensemble data association [C]. 2020 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. October 24 - January 24, 2021, Las Vegas, NV, USA. IEEE, 2020: 4966-4973.
- [29] ZINS M, SIMON G, BERGER M O. OA-SLAM: Leveraging objects for camera relocation in visual SLAM [C]. 2022 *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. October 17-21, 2022, Singapore, Singapore. IEEE, 2022: 720-728.
- [30] ZHANG J, HENEIN M, MAHONY R, *et al.* VDO-SLAM: a visual dynamic object-aware SLAM system [EB/OL]. 2020: 2005.11052. <https://arxiv.org/abs/2005.11052v3>.
- [31] HUANG J H, YANG S, MU T J, *et al.* ClusterVO: clustering moving instances and estimating visual odometry for self and surroundings [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 2165-2174.

- [32] KUHN H W. The Hungarian method for the assignment problem [J]. *Naval Research Logistics Quarterly*, 1955, 2(1): 83-97.
- [33] TIAN R, ZHANG Y Z, CAO Z Z, *et al.* Object SLAM with robust quadric initialization and mapping for dynamic outdoors[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(10): 11080-11095.

作者简介:



李曙光(1999—),男,安徽合肥人,硕士研究生,主要从事计算视觉、SLAM方面的研究。E-mail: lishuguang0103@163.com

通讯作者:



史金龙(1976—),男,黑龙江宾县人,博士,教授,硕士生导师。主要从事计算机视觉、人工智能、软件工程方向的研究与教学工作,主持或参与科技部国家重点研发计划子课题、江苏省科技厅产学研前瞻项目、军工课题、企业课题等20余项。E-mail: shi_jinlong@163.com