Modeling and Simulation 建模与仿真, 2024, 13(3), 2295-2304
Published Online May 2024 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2024.133210

Hans汉斯

基于YOLOv5的移动机器人动态视觉SLAM算法研究

李佳星，丛佩超*，刘俊杰，肖宜轩

广西科技大学机械与汽车工程学院，广西 柳州

收稿日期：2024年4月18日；录用日期：2024年5月13日；发布日期：2024年5月20日

摘　要

移动机器人在未知环境中，通过同步定位与地图构建(SLAM)技术，实现了精准的自身定位功能。目前大多数视觉SLAM系统均假设环境是静态的，但在实际应用中，由于大量动态目标的存在，严重影响机器人的定位与建图精度。为改善这一情况，本文基于ORB-SLAM3系统提出一种鲁棒的动态视觉SLAM系统，其融合YOLOv5深度学习方法，以减少动态目标的影响。并在公共TUM数据集和真实场景中测试本文算法的性能，结果表明：本文算法与ORB-SLAM3相比，具有更高的鲁棒性。

关键词

动态视觉SLAM，ORB-SLAM3，深度学习，移动机器人

Research on Dynamic Visual SLAM Algorithm for Mobile Robots Based on YOLOv5

Jiaxing Li, Peichao Cong*, Junjie Liu, Yixuan Xiao

College of Mechanical and Automotive Engineering, Guangxi University of Science and Technology, Liuzhou Guangxi

Received: Apr. 18th, 2024; accepted: May 13th, 2024; published: May 20th, 2024

Abstract

Mobile robots can achieve precise self-localization through Simultaneous Localization and Map-

*通讯作者。

文章引用: 李佳星, 丛佩超, 刘俊杰, 肖宜轩. 基于 YOLOv5 的移动机器人动态视觉 SLAM 算法研究[J]. 建模与仿真, 2024, 13(3): 2295-2304. DOI: 10.12677/mos.2024.133210

ping (SLAM) technology in unknown environments. Most current visual SLAM systems assume that the environment is static, but in practical applications, the presence of a large number of dynamic objects seriously affects the robot's localization and mapping accuracy. To improve this situation, this paper proposes a robust dynamic visual SLAM system based on the ORB-SLAM3 system, which integrates the YOLOv5 deep learning method to reduce the impact of dynamic objects. The performance of the algorithm in this paper is tested on the public TUM dataset and real-world scenarios, and the results show that the algorithm in this paper has higher robustness compared with ORB-SLAM3.

Keywords

Dynamic Visual SLAM, ORB-SLAM3, Deep Learning, Mobile Robot

1. 引言

近年来，随着人工智能和机器人技术的飞速发展，移动机器人的自主导航与定位问题已成为研究热点。目前，SLAM (Simultaneous Localization and Mapping，即同时定位与地图构建)技术是移动机器人领域的核心支柱[1] [2]。作为一种前沿的感知与导航方法，SLAM 在缺乏先验信息的复杂环境中发挥着至关重要的作用。其中视觉 SLAM 技术利用视觉传感器获取图像信息，实现对环境的高精度感知和地图构建。在静态环境中，视觉 SLAM 技术利用环境中的固定特征进行定位和建图，展现出良好的稳定性和可靠性。

在现实场景中，尤其是工作环境，动态元素往往占据主导地位，它的存在使特征点匹配过程易产生错误数据关联，从而影响定位精度与地图构建效果。2023 年，Jin 等人[3]基于 ORB-SLAM3 框架添加语义分割、动态特征剔除和点云稠密地图三个主要过程，在语义分割线程采用 SparseInst 分割网络实现对动态对象的检测，但该语义分割网络计算较复杂，影响系统实时性。因此，视觉 SLAM 技术面临着巨大的挑战[4] [5]。针对以上问题，本文基于 ORB-SLAM3 [6]系统引入 YOLOv5 深度学习方法完成对动态对象的检测与剔除。该系统不仅能排除动态对象的干扰，还利用语义信息重建静态背景，从而提高建图定位精度。本文主要贡献总结如下：

1) 针对动态对象缺乏语义信息问题，本文引入 YOLOv5 目标检测算法，实时生成场景中动态对象的基本语义信息；

2) 针对特征点剔除问题，本文设计一种特征点剔除策略，该策略不仅降低动态对象的影响，还提高定位建图精度；

3) 结合语义信息，对剔除动态点后的静态点构建稠密三维地图，并在 TUM [7] (慕尼黑工业大学)的 RGB-D 数据集与真实动态场景中评估本文算法。实验结果表明：本文算法优于 ORB-SLAM3 等视觉 SLAM 算法。

2. 算法原理

2.1. 系统概述

ORB-SLAM3 是一种先进的视觉同时定位与地图构建系统，主要由三大线程组成：跟踪线程、局部建图线程和闭环检测线程。跟踪线程的主要任务是负责提取和匹配 ORB 特征点，通过最小化重投影误差

来估算相邻帧之间的相对位姿。局部建图线程则负责管理和优化关键帧，通过集束调整(BA)技术来改进局部地图中所有帧的位姿精度。闭环检测线程则负责在整个地图中识别并检测闭合的回环，通过位姿图优化来不断纠正累积的漂移误差，从而提高整个地图的精度和稳定性。该系统在静态场景中具有较高鲁棒性，但在动态场景中，由于动态物体的存在会导致特征点间产生错误的匹配关系，从而影响定位建图精度。因此，本文在原有 ORB-SLAM3 系统的基础上，添加了基于 YOLOv5 的目标检测线程和动态特征点剔除模块，改进的系统框图如图 1 所示。

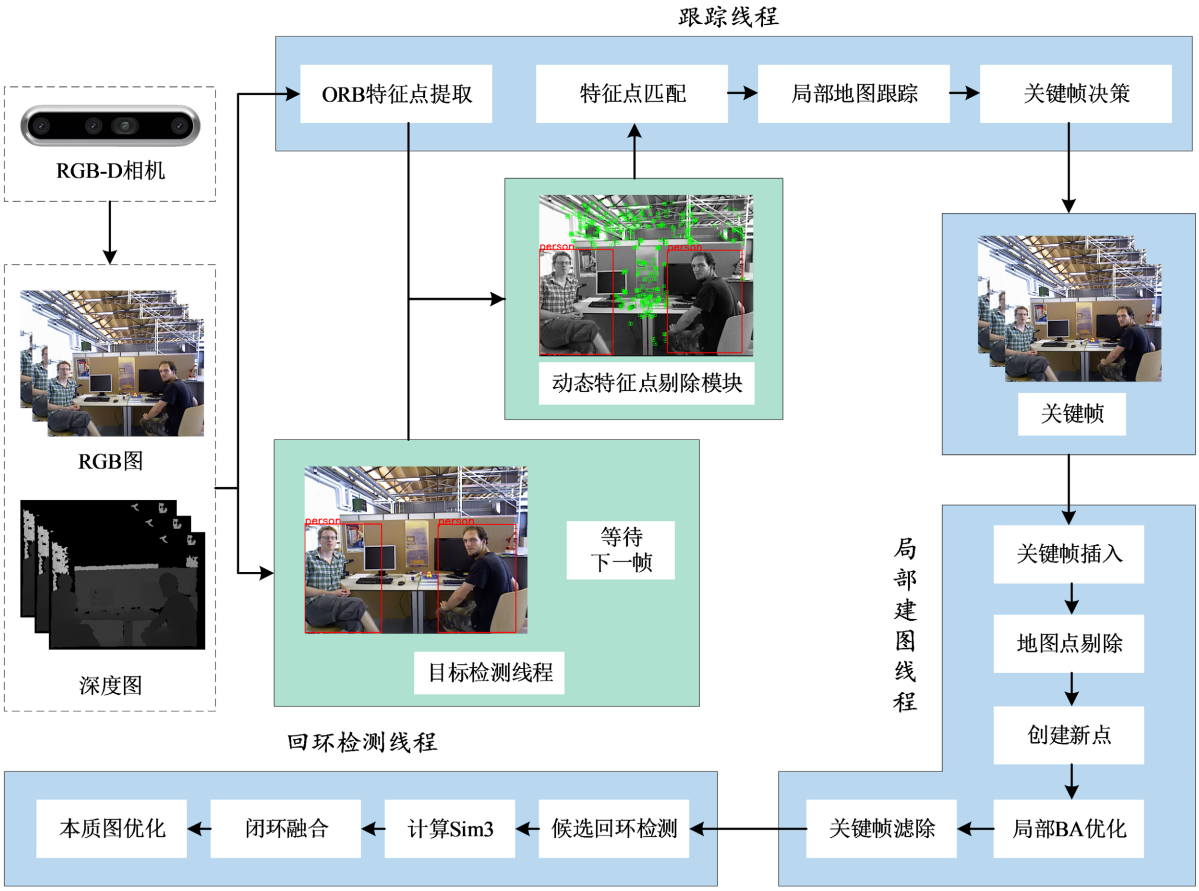


Figure 1. System block diagram
图 1. 系统框图

2.2. YOLOv5 检测动态目标

随着深度学习的发展，很多学者开始将深度学习方法[8] [9]引入到视觉 SLAM 中，以减少动态目标的影响，提高其定位建图精度。常见的深度学习方法分为两类，目标检测和语义分割。目标检测的网络相对简单，但不能清晰表达物体的轮廓信息；而语义分割网络虽能清晰表达物体的轮廓信息，但网络相对复杂，具有更高的计算量。传统视觉语义 SLAM 系统在兼顾语义分割网络的精度和实时性方面面临巨大挑战。尽管 Mask-RCNN 作为一种常用的两阶段实例分割方法具备较高分割精度，但其对图像的处理时间较长，难以满足实时性要求。为提高视觉 SLAM 系统的精度和运行效率，本文采用轻量级检测算法——YOLOv5 算法。相较于 Mask-RCNN 算法，YOLOv5 分割算法具有更高运行效率，其设计更注重实时性。

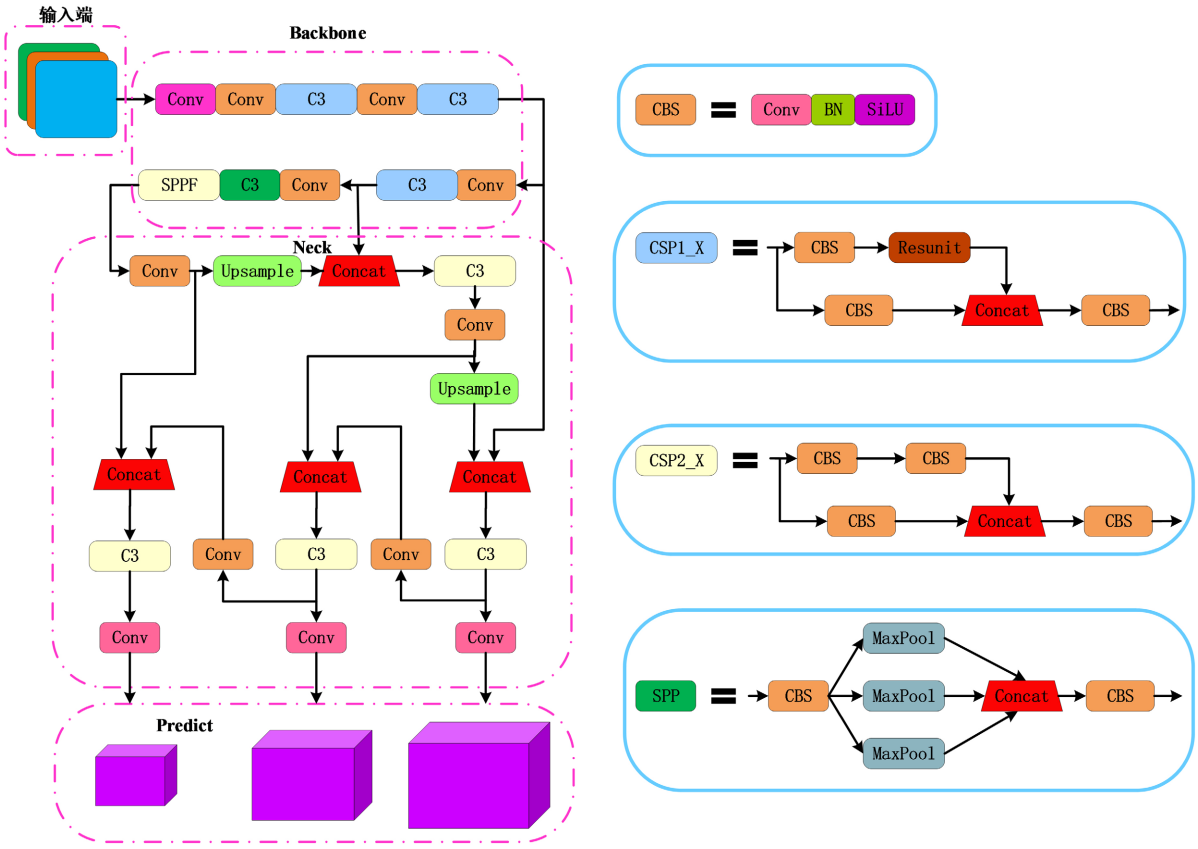


Figure 2. YOLOv5 network structure diagram
图 2. YOLOv5 网络结构图

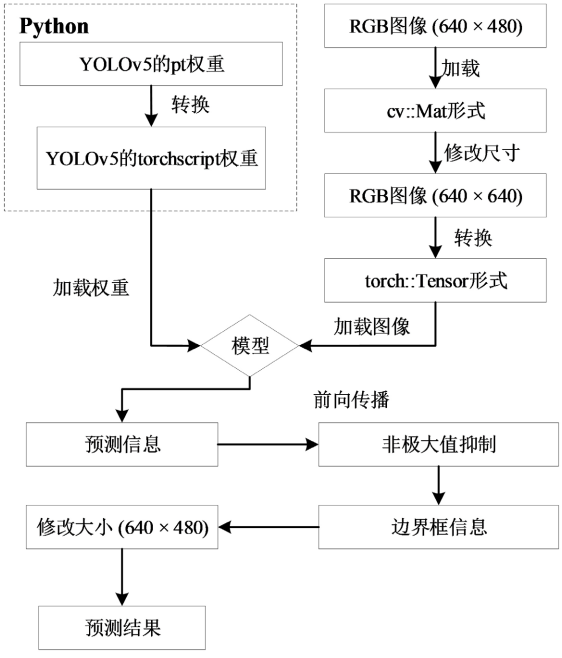


Figure 3. Deployment process (The dashed box is implemented in Python, while the outside of the dashed box is implemented in C++)
图 3. 部署流程(虚线框内为 Python，虚线框外为 C++实现)

YOLOv5 是当前较为先进的单阶段目标检测算法，在资源受限的系统中更具优势。其拥有 x，l，m，s 和 n 五个不同的网络模型，其中 YOLOv5s 在平衡精度和实时性方面具有更佳的性能。因此，选择 YOLOv5s 作为本文的目标检测网络是合理的选择，网络结构图如图 2 所示。但 YOLOv5 是基于 Python 实现的，而 ORB-SLAM3 是基于 C++实现的，为将 YOLOv5 更好融入 ORB-SLAM3 中，本文设计了一个部署模块，如图 3 所示。其中，在 Python 部分获得可供 C++加载的权重文件，并利用前向传播推导出网络模型，输出相关的检测框信息。具体的检测效果如图 1 目标检测线程所示。

2.3. 动态特征点剔除

在 2.2 节已获得相关的动态物体检测框信息，本节将对检测框内的特征点进行剔除。在跟踪线程添加一个特征点剔除模块，根据动态物体检测框的信息判断特征点的动静态属性，若特征点在检测框范围内，则将该特征点视为动态特征点，否则，将其视为静态特征点。对所有属性为动态的特征点，将其从特征点集里删除，并剔除其对应的地图点信息。具体的剔除效果如图 1 动态特征点剔除模块中所示。

3. 实验与分析

在本节中，将在公共 TUM 的 RGB-D 数据集和真实场景中演示本文算法，以评估其在动态场景的精度和鲁棒性，并将本文算法与原始 ORB-SLAM3 算法进行比较。本文的仿真实验是在一台台式机上进行的，其 CPU 型号为 i5-9600KF，主频为 3.70GHz，显卡为 NVIDIAGTX 1650；实际场景测试是在型号为 y700 的联想笔记本电脑上进行，其 CPU 型号为 i5-6300HQ，主频为 2.30GHz，显卡为 NVIDIAGTX 960M。为评估相机轨迹数据和真实估计数据之间差异，本文选用两个常用评估指标，即绝对轨迹误差(ATE)和相对位姿误差(RPE)。ATE 用于描述估计位姿和真实位姿之间的绝对误差，能直观地反映算法精度和轨迹的全局一致性；RPE 则用于描述相邻两帧之间估计位姿和真实位姿变化的差异，其中包括旋转误差和平移误差两个部分。

3.1. TUM 数据集性能评估

本文算法是基于 ORB-SLAM3 算法改进而来，为凸显改进效果，将其与 ORB-SLAM3 算法进行详细对比。选取动态数据集的 4 个高动态序列进行评估，评估结果如表 1~3 所示。由表中数据可知，在所有情况下，本文算法的精度均比 ORB-SLAM3 算法的精度高，即具有更高的精度和鲁棒性。为更直观展示两种算法在精度上的差异，对表中的数据进行可视化，可视化效果如图 4 所示。蓝线表示 ORB-SLAM3 算法 ATE 值，绿线表示本文算法 ATE 值。在处理 walking 系列高动态数据集时，由于人的移动特征更加明显，原始 ORB-SLAM3 算法产生许多错误匹配的特征点对，从而导致计算出错误的相机位姿。相比之下，本文算法采用有效动态特征点剔除策略，仅保留可靠静态特征点，极大地减少特征点误匹配情况，从而显著提高算法鲁棒性。

Table 1. ATE-RMSE (m) evaluation of ORB-SLAM3 and the algorithm proposed in this paper
表 1. ORB-SLAM3 和本文算法的 ATE-RMSE (m)评估

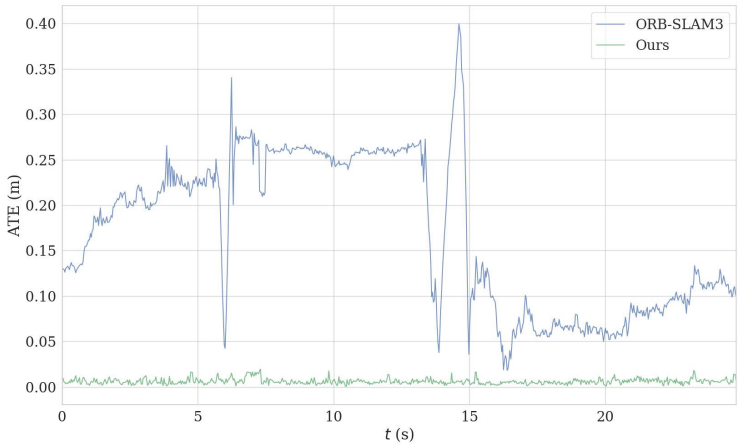
序列	ORB-SLAM3	本文算法
fr3/walking_static	0.1882	**0.0067**
fr3/walking_xyz	0.7218	**0.0158**
fr3/walking_half	0.3283	**0.0308**
fr3/walking_rpy	0.7022	**0.2903**

Table 2. RPE-RMSE (m/s) evaluation of ORB-SLAM3 and the algorithm proposed in this paper (RPE translation part)
表 2. ORB-SLAM3 和本文算法的 RPE-RMSE (m/s)评估(RPE 平移部分)

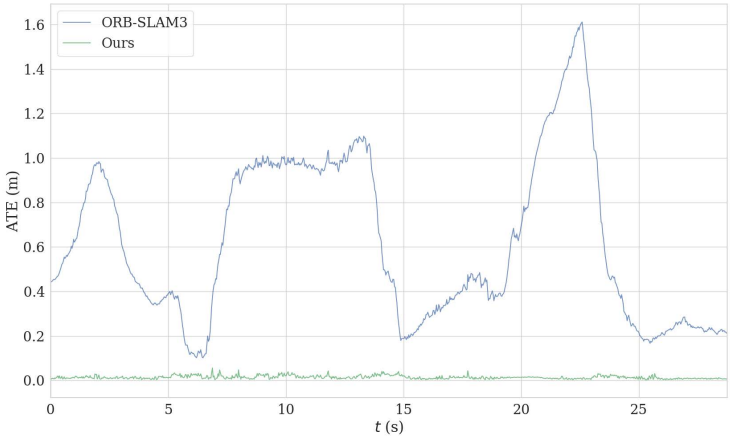
序列	ORB-SLAM3	本文算法
fr3/walking_static	0.0156	**0.0059**
fr3/walking_xyz	0.0258	**0.0116**
fr3/walking_half	0.0227	**0.0135**
fr3/walking_rpy	0.0302	**0.0212**

Table 3. RPE-RMSE (deg/s) evaluation of ORB-SLAM3 and the algorithm proposed in this paper (RPE rotation part)
表 3. ORB-SLAM3 和本文算法的 RPE-RMSE (deg/s)评估(RPE 旋转部分)

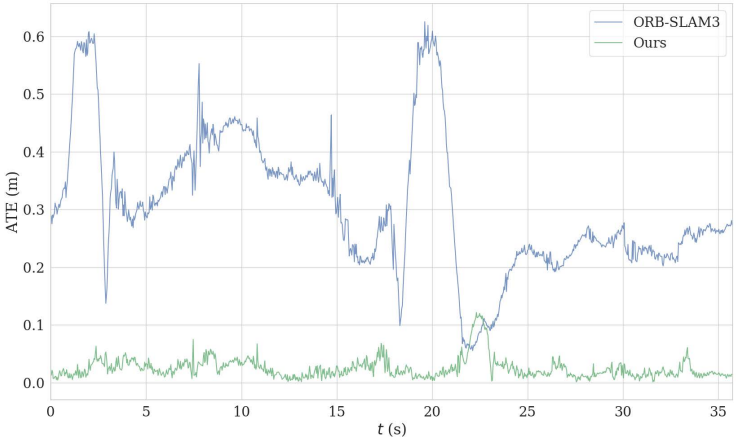
序列	ORB-SLAM3	本文算法
fr3/walking_static	0.3157	**0.1688**
fr3/walking_xyz	0.6281	**0.3879**
fr3/walking_half	0.5719	**0.3956**
fr3/walking_rpy	0.6892	**0.5225**



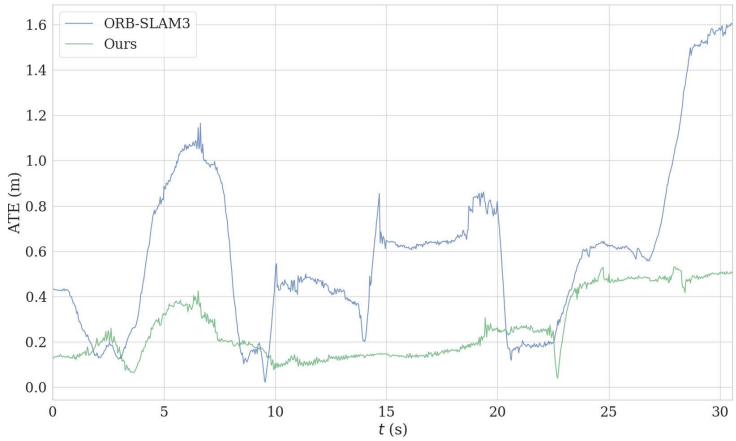
(a) fr3/walking_static



(b) fr3/walking_xyz



(c) fr3/walking_half



(d) fr3/walking_rpy

Figure 4. ATE distribution of ORB-SLAM3 (blue) and the algorithm proposed in this paper (green)
图 4. ORB-SLAM3 (蓝色)和本文算法(绿色)的 ATE 分布

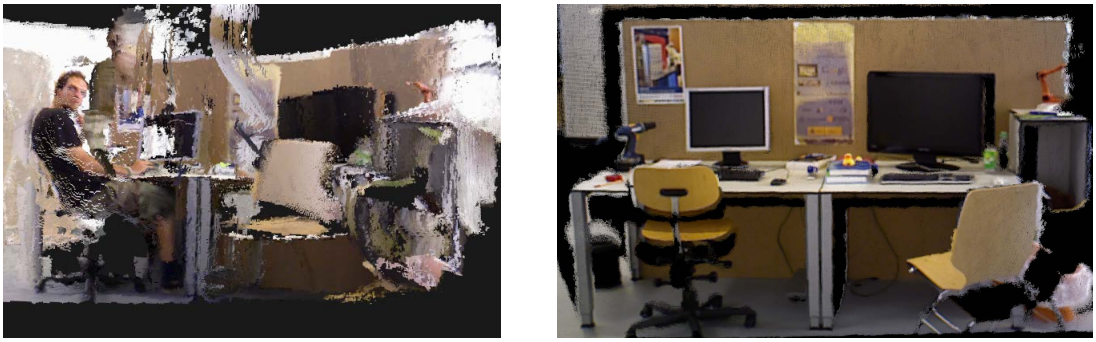
3.2. 稠密建图

稠密地图相对于稀疏地图，具有更高的精度和细节表现力，能更好地描述环境结构、表面属性和几何形态。本文通过结合原始 RGB 图、深度图以及关键帧等信息完成静态三维稠密地图的构建。对 ORB-SLAM3 算法和本文算法在 TUM 数据集进行测试，测试结果如图 5 所示。图中左侧图像为采用 ORB-SLAM3 没有剔除动态对象的结果，其中包含大量动态对象重影，而右侧图像为本文算法建图结果，不包含明显动态对象。

3.3. 真实场景

为充分验证本文算法在真实场景鲁棒性，本文搭建如图 6 所示的实验平台。移动平台为 Agilex Robotics-BUNKER，升降平台上放置 Lenovo y700 笔记本电脑和 Intel RealSense D455 深度相机。实验场景为室内办公区，其中人作为主要移动对象，移动机器人将按照环形路线行驶。如图 7 所示，选取两个视角对比 ORB-SLAM3 算法和本文算法，可以看出 ORB-SLAM3 算法未能剔除动态物体特征点，而本文算法具有良好的特征点剔除效果，且不会误剔除背景静态特征点。图 8 对二者进行轨迹对比，可以看出本文算法轨迹效果更佳，而 ORB-SLAM3 算法因未剔除动态特征点造成匹配点之间的错误数据关联，以致

轨迹漂移严重。由于本文使用的升降平台结构不稳定，导致移动机器人在行驶过程中相机发生晃动，轨迹局部也产生抖动。



(a) ORB-SLAM3　　(b) 本文算法

Figure 5. Dense mapping
图 5. 稠密建图

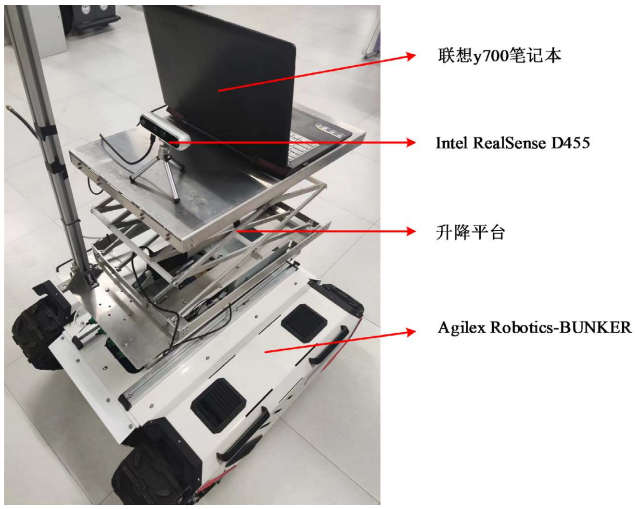


Figure 6. Experimental platform
图 6. 实验平台

(a) ORB-SLAM3　　(b) 本文算法

Figure 7. Actual dynamic scene effect diagram
图 7. 实际动态场景效果图

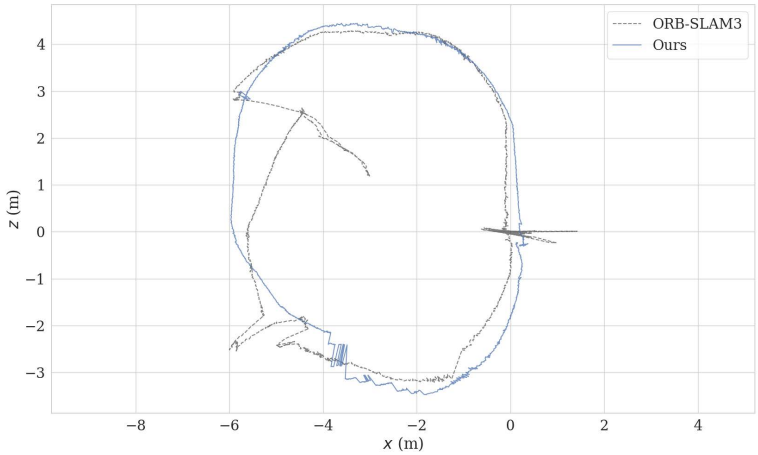


Figure 8. Comparison of indoor dynamic scene circular trajectories
图 8. 室内动态场景环形轨迹对比

4. 结论

本文基于 ORB-SLAM3 系统提出一个鲁棒的动态视觉 SLAM 系统。为避免场景中动态对象产生的错误数据关联问题，我们设计了一个基于 YOLOv5 的目标检测线程和一个动态特征点剔除模块。思路是将动态物体利用 YOLOv5 深度学习网络检测，并将检测框信息输入到跟踪线程中，根据特征点和检测框的位置关系对特征点的属性进行判断，从而将动态特征点从系统中删除。然后，将剔除动态特征点的关键帧结合 PCL 库构建静态稠密三维地图，并在具有挑战性的 TUM RGB-D 数据集以及真实场景中对本文算法进行了评估。对于 TUM 数据集，相比于 ORB-SLAM3 算法，本文算法具有更低的绝对轨迹误差和相对位姿误差；对于真实场景，本文算法对动态特征点进行有效剔除，并具有更高的定位精度和鲁棒性。实验结果表明：与 ORB-SLAM3 算法相比，本文算法在动态场景中效果更佳，具有更高的精度和鲁棒性。

基金项目

中央引导地方科技发展专项资金项目(桂科 ZY19183003)；广西重点研发计划项目(桂科 AB20058001)。

参考文献

[1] Chen, W., Shang, G., Ji, A., *et al*. (2022) An Overview on Visual Slam: From Tradition to Semantic. *Remote Sensing*, **14**, Article 3010. https://doi.org/10.3390/rs14133010

[2] Macario Barros, A., Michel, M., Moline, Y., Corre, G. and Carrel, F. (2022) A Comprehensive Survey of Visual SLAM Algorithms. *Robotics*, **11**, Article 24. https://doi.org/10.3390/robotics11010024.

[3] Jin, J., Jiang, X., Yu, C., *et al*. (2023) Dynamic Visual Simultaneous Localization and Mapping Based on Semantic Segmentation Module. *Applied Intelligence*, **53**, 19418-19432

[4] Sharafutdinov, D., Griguletskii, M., Kopanev, P., Kurenkov, M., Ferrer, G., Burkov, A., Gonnochenko, A. and Tsetserukou, D. (2023) Comparison of Modern Open-Source Visual SLAM Approaches. *Journal of Intelligent & Robotic Systems*, **107**, Article No. 43. https://doi.org/10.1007/s10846-023-01812-7

[5] Zhang, Q., Yu, W., Liu, W., *et al*. (2023) A Lightweight Visual Simultaneous Localization and Mapping Method with a High Precision in Dynamic Scenes. *Sensors*, **23**, Article 9274. https://doi.org/10.3390/s23229274

[6] Campos, C., Elvira, R., Rodríguez, J.J.G., *et al*. (2021) ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap Slam. *IEEE Transactions on Robotics*, **37**, 1874-1890. https://doi.org/10.1109/TRO.2021.3075644

[7] Sturm, J., Engelhard, N., Endres, F., *et al*. (2012) A Benchmark for the Evaluation of RGB-D SLAM Systems. 2012 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura-Algarve, 7-12 October 2012, 573-580. https://doi.org/10.1109/IROS.2012.6385773

[8] Bescos, B., Campos, C., Tardós, J.D., *et al*. (2021) DynaSLAM II: Tightly-Coupled Multi-Object Tracking and SLAM. *IEEE Robotics and Automation Letters*, **6**, 5191-5198. https://doi.org/10.1109/LRA.2021.3068640

[9] Fang, B., Mei, G., Yuan, X., *et al*. (2021) Visual SLAM for Robot Navigation in Healthcare Facility. *Pattern Recognition*, **113**, Article ID: 107822. https://doi.org/10.1016/j.patcog.2021.107822