

动态场景下基于图像分割与融合点线特征的视觉SLAM方法

姚东昊, 袁野, 吴宝磊, 刘娜

(上海理工大学 机器智能研究院, 上海 200093)

摘要: 针对视觉SLAM在动态场景下定位精度差和鲁棒性不足的问题, 提出一种适用于动态场景的视觉SLAM算法SPL-SLAM。首先, 利用YOLOv8-seg实例分割网络来检测先验区域, 用于动态区域识别和语义地图构建; 其次, 通过引入线特征约束来增强系统的鲁棒性; 再次, 利用运动一致性检测和角度检测算法识别动态区域, 并过滤动态点特征和线特征; 最后, 设计了一种自适应权重的误差函数并结合剩余静态点线特征进行位姿估计, 同时创建了一个静态稠密点云地图, 以避免动态物体干扰。使用TUM数据集及真实场景进行实验验证, 结果表明, 相较于ORB-SLAM3及其他相关的动态SLAM算法, SPL-SLAM算法在动态场景下具有更好的定位精度。

关键词: 视觉SLAM; 图像分割; 位姿估计; 点云地图; 动态场景

DOI: 10.11907/rjtk.241357

开放科学(资源服务)标识码(OSID):

中图分类号: TP242.6

文献标识码: A

文章编号: 1672-7800(2025)006-0142-09



Visual SLAM Method Based on Image Segmentation and Fusion of Point and Line Features in Dynamic Scenes

YAO Donghao, YUAN Ye, WU Baolei, LIU Na

(Institute of Machine Intelligence, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Aiming at the problems of poor positioning accuracy and insufficient robustness of visual SLAM in dynamic scenes, a visual SLAM algorithm SPL-SLAM suitable for dynamic scenes is proposed. The algorithm first utilizes the YOLOv8 seg instance segmentation network to detect prior regions for dynamic region recognition and semantic map construction; Secondly, the robustness of the system is enhanced by introducing line feature constraints; Again, use motion consistency detection and angle detection algorithms to identify dynamic regions, and filter dynamic point features and line features; Finally, an adaptive weight error function was designed and combined with residual static point and line features for pose estimation, while creating a static dense point cloud map to avoid dynamic object interference. Experimental verification was conducted using the TUM dataset and real scenes, and the results showed that compared to ORB-SLAM3 and other related dynamic SLAM algorithms, the SPL-SLAM algorithm has better positioning accuracy in dynamic scenes.

Key Words: visual SLAM; image segmentation; pose estimation; point cloud map; dynamic scene

0 引言

同时定位与地图构建(Simultaneous Localization and Mapping, SLAM)是机器人领域的重要研究方向,旨在使机

器人能够同时进行定位与地图构建。近年来,研究人员提出许多开源的SLAM方法,包括Cartographer、Gmapping和Hector SLAM等,均基于激光传感器实现^[1-3]。相较于激光雷达,相机低廉的成本和丰富的图像信息使SLAM方法在动态场景中有着更为广泛的应用。经过几十年的发展,一

收稿日期: 2024-04-10

扫描二维码阅读全文:

基金项目: 国家重点研发计划项目(2023YFC3605800)

作者简介: 姚东昊(2000-),男,上海理工大学机器智能研究院硕士研究生,研究方向为图像处理和视觉SLAM;袁野(1991-),男,博士,上海理工大学机器智能研究院讲师、硕士生导师,研究方向为类脑学习;吴宝磊(2000-),男,上海理工大学机器智能研究院硕士研究生,研究方向为图像处理和视频理解;刘娜(1985-),女,博士,上海理工大学机器智能研究院讲师、硕士生导师,研究方向为面向遥感图像的哈希学习。本文通讯作者: 刘娜。



些成熟的视觉SLAM系统已经出现,如ORB-SLAM2^[4]、LSD-SLAM^[5]、RGB-D SLAM^[6]、DSO^[7]和SVO^[8]。然而,以上系统仍存在一些未解决的问题。大多数SLAM系统通常假设环境是静态的,意味着在地图或跟踪过程中不考虑动态物体。然而,在动态环境中,这一假设不再成立。为解决这一问题,研究人员提出了一些解决方法,大致可分为基于几何的方法和基于语义的方法两类。

基于几何的方法利用数学模型来检测和跟踪环境中的移动物体,这类方法通常依赖于视觉测距或运动估计技术,根据物体在摄像机图像中的二维投影来估计物体的三维运动。例如,Fang等^[9]采用最优估计和均匀采样技术来检测动态物体。Li等^[10]利用帧间关键帧的配准和静态点权重来最小化移动物体的不良影响。Sun等^[11]提出一种增强型RGB-D SLAM方法,在预处理阶段用于过滤与移动物体相关的数据。Wang等^[12]提出一种针对动态室内环境的新型RGB-D SLAM方法,该方法基于数学模型和几何约束建立,可作为SLAM过程中的数据过滤阶段进行整合。Ni等^[13]提出一种改进的基于ORB-SLAM3构建的视觉SLAM方法,引入改进的新地图点筛选策略和重复退出地图点消除策略,并将它们结合起来识别明显的动态地图点。Zhou等^[14]提出一种在室内动态环境中具有两个通道的RGB-D SLAM方法,根据匹配精度将帧分为轻微和高度动态的情况,并且提出一种基于汉明距离的几何约束,以提高匹配精度。然而,由于几何信息不足,大多数基于几何的动态SLAM方法未能达到最佳性能。

基于语义的方法则整合了关于环境的语义信息来检测和跟踪动态物体,涉及使用深度学习技术对场景中的物体进行分类,并预测它们随时间的运动情况。通过利用语义信息,这类方法可以处理复杂场景中的移动物体,这些物体可能部分被遮挡或具有相似的外观。例如,Zhang等^[15]利用YOLO来识别环境中的物体,并创建了一个语义地图来过滤动态特征点,从而提高了系统的准确性。Bescos等^[16]提出一种基于ORB-SLAM2的语义系统,使用Mask R-CNN对每个图像帧进行分割并获得语义掩码。该方法整合了多视图几何和语义掩码,以消除动态特征点并填补被动态物体遮挡的背景区域。Yu等^[17]提出的DS-SLAM通过将语义分割与光流方法相组合来减轻动态物体的影响。Zheng等^[18]提出一种稳健且轻量的动态SLAM方法RLD-SLAM,将目标检测与贝叶斯滤波相结合,在保证高准确性的同时快速获取静态特征点。Islam等^[19]提出一种名为ARD-SLAM的动态对象识别技术,将几何运动信息与未来运动数据巧妙地结合在一起,用于过滤动态物体。Luo等^[20]提出一种结合改进Fast-SCNN与Deepfillv2的FD-SLAM系统,采用语义分割和几何方法识别和消除动态区域;同时基于高效的修补网络,利用自由形式的蒙版来填补被动态对象遮挡的背景。虽然深度学习方法是

非常强大的工具,但是其在很大程度上依赖于网络质量。在某些情况下,具有简单架构的网络可能无法准确识别动态物体,而复杂的网络架构可能会减慢系统识别速度。

为解决动态物体导致的相机姿态估计不准确以及仅依赖有限数量静态特征点进行定位而引起的鲁棒性不足等问题,本文提出一种名为SPL-SLAM(Segmentation-based Point-Line Feature Fusion SLAM)的视觉SLAM系统,该系统在ORB-SLAM3框架的基础上增加了图像分割模块并融合点线特征进行位姿估计^[21]。本文的主要研究工作为:①使用YOLOv8-seg网络进行图像分割,从而获取先验对象的掩码,并利用TensorRT进行加速,实现整个系统的实时跟踪;②为避免在剔除掉过多动态特征后导致跟踪失败的情况,本文采用LSD(Line Segment Detector)算法进行线特征提取,并引入线特征约束^[22]。通过结合特征点的运动一致性检查和特征线的角度检查识别动态区域,并过滤掉动态特征点和线;③设计了一个自适应权重的误差函数,根据剩余静态特征点和线来估计相机姿态;④构建稠密3D点云地图,并通过添加语义信息使移动机器人在执行导航和避障等任务时具备高级感知能力,同时避免动态物体对地图构建的干扰,保证了地图的可用性。

1 算法框架

SPL-SLAM的总体框架如图1所示。首先,针对输入每个RGB图像,使用YOLOv8-seg实例分割网络提取其先验区域,并提取点特征和线特征。利用点特征的运动一致性检测和线特征的角度检查,结合先验区域对每个点和每条线进行分类,确定其为动态或静态。其次,基于从YOLOv8-seg实例分割网络获取的先验区域以及动态特征点和特征线,根据预定义阈值来识别动态区域。再次,对动态区域内的特征(包括点和线)进行过滤。最后,利用剩余的静态特征点和特征线计算所有点和线的动静态权重。将其与点和线的重投影误差相结合,用于优化相机姿态,同时基于采集的深度图像构建稠密3D点云地图。

1.1 YOLOv8实例分割

鉴于需要像素级动态物体检测和语义信息映射,本文

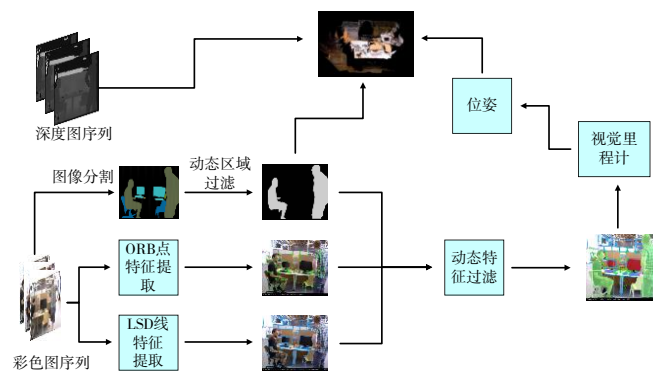


Fig. 1 Overall framework of SPL-SLAM

图1 SPL-SLAM总体框架

采用在 MS COCO 数据集上预训练的 YOLOv8 实例分割模型。该模型能够对日常生活中常见的 80 类物体进行分割。SPL-SLAM 系统采用 YOLOv8 从输入图像中提取实例分割区域,为动态物体检测和语义信息提供先验区域。YOLOv8 主要由以下 3 个模块组成,用于完成物体检测与分类:①骨干网络。该模块采用一系列卷积和反卷积层来提取特征,结合残差连接和瓶颈结构来减小网络大小并增强性能。该部分由 5 个卷积模块、4 个 C2f 模块和 1 个 SPPF (Spatial Pyramid Pooling Fast) 模块组成。与 YOLOv5 的 C3 模块相比,C2f 模块具有更少的参数和更优越的特征提取能力;②颈部。该模块采用多尺度特征融合技术,对来自骨干网络不同阶段的特征图进行融合,以增强特征表征能力。具体来说,YOLOv8 的颈部包括 1 个 SPPF 模块、1 个 PAA (Probabilistic Anchor Assignment) 模块和 2 个 PAN (Path Aggregation Network) 模块;③头部。该模块负责最终的物体检测与分类任务,由 1 个检测头和 1 个分类头组成。其中,检测头包括一系列卷积和反卷积层,用于生成检测结果;分类头采用全局平均池化对每个特征图进行分类。

获取周围环境的语义信息有两个目的:一方面,其能提供先验信息以消除潜在的动态特征。本文采用一种结合先验对象区域与几何约束的方法来过滤动态对象。随后,动态对象上的特征点和线被移除,增强了跟踪线程内姿态估计的准确性以及系统在动态环境中的稳健性。另一方面,SPL-SLAM 将根据过滤后得到的静态对象创建静态 3D 点云地图。如图 2(彩图扫 OSID 码可见,下同)所示,在分割效果和细节处理方面,与 SegNet 和 Mask R-CNN 相比,YOLOv8 获得的掩码边缘更清晰、更完整。

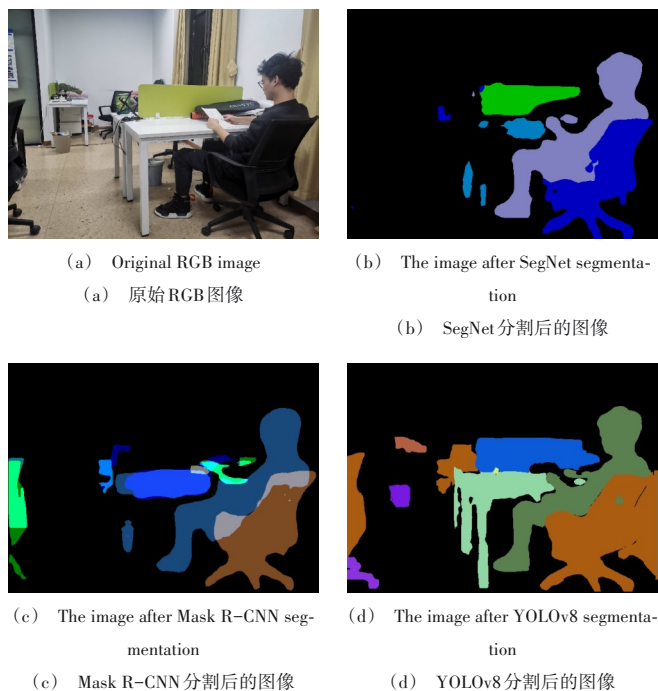


Fig. 2 Comparison of different semantic segmentation networks

图 2 不同语义分割网络比较

1.2 动态区域移除

特征点和特征线的提取质量决定了 SLAM 系统的精度和稳定性。本文使用 YOLOv8 实例分割网络提取先验对象的轮廓区域,并得到掩膜,然后根据特征点的运动一致性和极线约束算法计算特征点到极线的距离。极线约束是在双目或双摄像头系统中的一种几何约束,表示在一帧图像中观察到的特征点与另一帧图像中该特征点可能存在的位置之间的关系。当相机移动时,静态物体的特征点会在图像中保持相对不变的位置,而动态物体的特征点则会在不同图像帧中出现不同的位置。利用极线约束可以滤除那些不符合特定运动模式的特征点,因为这些特征点在不同图像帧中不会落在同一条极线上。具体而言,对于当前帧中的每个特征点,根据相机内参以及初始位姿将其投影至下一帧中,然后计算投影后的点到极线的距离,从而确定其是否属于静态场景。如果特征点在多个帧之间的极线上具有一致的位置,并且距离小于一定的阈值,那么可以将其视为静态特征点,否则可能是动态特征点。同时,SPL-SLAM 系统设计了一个夹角检测模块以检测动态特征线,该模块依赖光流法。本文作出以下合理假设:对于成功提取的线特征,在视觉变化不大且物体没有剧烈运动的情况下,其端点的 3D 位置在跟踪过程中始终保持不变。基于以上假设,线特征也可以像点特征一样使用光流法。

如图 3 所示,线段 l_1 和 l_2 为相邻两帧图像中通过 LBD (Line Band Descriptor) 算法匹配的特征线,点 A_1 、 B_1 和 C_1 分别为 l_1 的起点、中点和终点,点 A_2 、 B_2 和 C_2 同理。根据假设,如果线特征是动态的,则这 3 个点在相同的图像运动中对应的 3 个轨迹与 X 轴之间的夹角,即 θ_1 、 θ_2 和 θ_3 是相等的。本文算法计算这 3 个夹角差值的平均值,如果该值大于一定的阈值,则将线特征视为动态的。

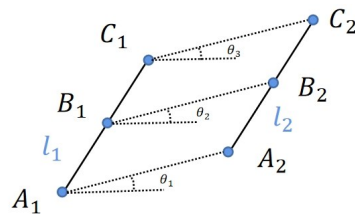


Fig. 3 Detection of angles between different line features

图 3 不同线特征之间的夹角检测

至此,已经在一帧图像中初步判别出动态特征点和特征线,结合 YOLOv8 得到的先验区域,按照以下方法可以得到最终的动态区域。首先,初始化两个空集合 S_1 和 S_2 ,用于存放动态区域;其次,遍历所有先验区域,统计每一个区域中的动态特征点和特征线数量;再次,按照区域中动态点特征的数量判定该区域是否为潜在动态区域,并加入集合 S_1 ,同理根据区域中动态线特征的数量判定该区域是否为潜在动态区域,并加入集合 S_2 ;最后,筛选出同时出现在集合 S_1 和 S_2 中的区域,即为最终动态区域,根据动态区域的掩膜过滤掉动态特征点和特征线。然而,当提取的特

征点和特征线恰巧在掩膜的边缘上时, 很难将其滤除掉。为解决这一问题, SPL-SLAM 对掩膜进行扩充处理。首先检测一帧语义分割后图像中 RGB 值为 (255, 255, 255) 的像素点; 然后将以该像素点为中心, 半径为 10 像素的圆形区域内的像素值设置为 (255, 255, 255); 最后得到扩张后的掩膜。如图 4 所示, 图 4(a) 表示原始图像, 图 4(b) 表示通过 YOLOv8 得到的掩膜, 图 4(c) 为扩张后的掩膜, 图 4(d) 为掩膜叠加在原始图后的结果。可以明显看出, 动态物体的边缘进行了扩张, 使得原本在边缘的特征点和特征线也能被滤除。

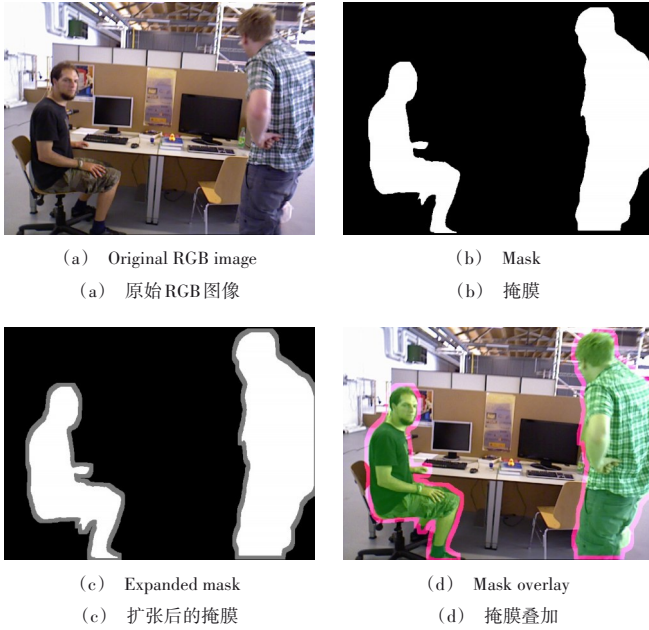


Fig. 4 Expansion of the mask

图 4 掩膜扩张

根据扩张后的动态物体掩膜, 动态区域中的特征点和特征线可以被滤除。如图 5 所示, 图 5(a) 为根据 ORB 特征提取的特征点, 图 5(b) 为根据 LSD 算法提取的线特征, 图 5(c) 为融合点线特征后的特征图, 图 5(d) 为滤除动态特征点和特征线后所剩余的静态特征点和特征线。

1.3 基于自适应权重的优化模型

在移除动态特征后, 剩余的点和线可被视为静态, 可以通过构建一个几何优化模型来迭代计算出相机位姿。将上一帧的特征点和特征线投影到三维空间, 再根据恒速运动模型的初始位姿将 3D 点和线重投影到当前帧, 在整个过程中点和线的重投影误差可以看作关于位姿的函数, 因此可以通过迭代优化来计算出相机位姿的最优值。

在上一帧中, 某个特征点 p 的三维坐标为 P_w , 根据恒速运动模型设置初始相机位姿为 T, K 为相机内参, 根据以下公式将 P_w 重投影到当前帧, 得到像素坐标 P_{uw} :

$$P_{uw} = f(P_w, T, K) \quad (1)$$

在当前帧中通过特征匹配算法得到点 p 的匹配点 p' 的像素坐标为 p'_{uw} 。通过计算匹配点与重投影点之间的距离

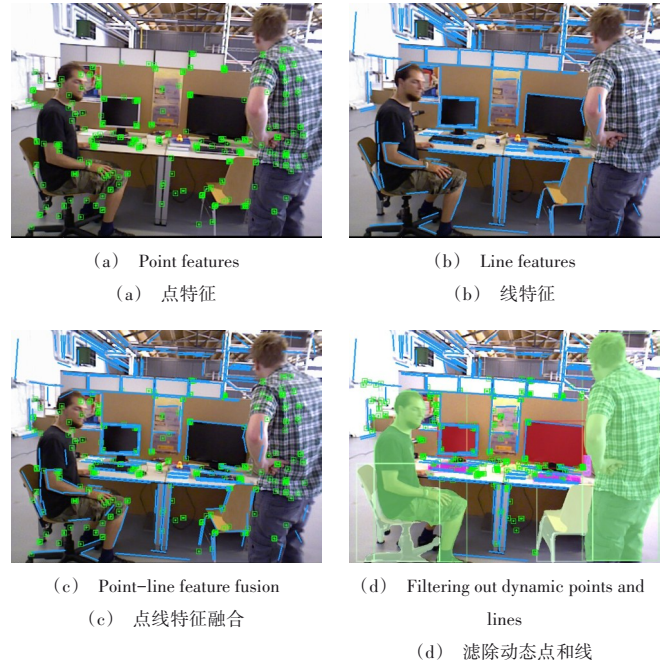


Fig. 5 Dynamic feature filtering

图 5 动态特征过滤

得到特征点的重投影误差。公式为:

$$E_j = P'_{uw} - P_{uw} \quad (2)$$

与特征点相比, 线的重投影误差更加复杂一些。在完成线特征匹配后, 基于前一帧图像的深度图计算前一帧线特征中相应的 3D 坐标。如图 6 所示, 将 3D 线特征的两个端点投影到当前帧, 分别计算两个端点与投影直线的距离, 以构建重投影误差模型。

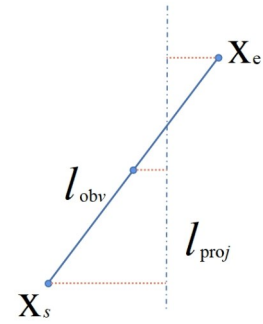


Fig. 6 The reprojection model of the line

图 6 线的重投影模型

图 6 中, X_e 为观测直线 l_{obs} 的终点, X_s 为起点, l_{proj} 为根据线特征匹配得到的直线。当前帧特征线的重投影误差可表示为:

$$E_l = \left| \frac{d(X_s, l')}{d(X_e, l')} \right| = \left| \frac{X_s^T l'}{\|l'\|} \right| \quad (3)$$

式中: l' 为上一帧图像中的特征线 l 经过特征匹配后在当前帧得到的匹配直线。

结合上述两个重投影误差已经可以初步构造优化问

题,并迭代计算相机的最优位姿。然而,这个优化问题非常依赖静态特征点和特征线的提取质量,若错将某些动态特征当作静态特征点或特征线来构建整个几何模型,会导致非常大的误差,从而对位姿计算产生巨大影响,并降低整个系统的定位精度。采用图像分割和几何约束步骤已经滤除了大部分动态区域中的特征点和特征线。然而,在某些局部运动、局部静止的情况下,以上方法并不能完全移除场景中的动态特征。设想以下场景:一个坐在椅子上的人在转头或挥手,其大部分特征是静态的,只有少部分在头部和手部的特征是动态的,若不加约束地对其进行优化,势必会将这些动态特征的误差引入优化模型。为解决这一问题,改进原有的重投影误差函数,并对每个特征引入自适应权重系数。表示为:

$$e = \sum_{k,j} (\alpha_p e_{k,j}^T \Omega_{k,j}^{-1} e_{k,j}) + \sum_{k,l} (\alpha_l e_{k,l}^T \Omega_{k,l}^{-1} e_{k,l}) \quad (4)$$

式中: $e_{k,j}$ 为点的重投影误差函数; $e_{k,l}$ 为线的重投影误差函数; e 为融合点线特征的误差函数; α_p 为特征点的权重系数,可结合几何模型来计算。

首先,遍历图像中剩余的静态特征点,根据极线约束和运动一致性检测计算两帧之间相匹配特征点的距离 d 。如图7所示, p_1 和 p_2 为图像平面上的像素点, P_1^w 和 P_2^w 为像素点根据相机位姿和内参矩阵投影得到的三维点,它们之间的差值即为所求距离 d 。其次,按照距离排序,得到最大距离 $dist_{\max}$ 和最小距离 $dist_{\min}$ 。最后,计算特征点的权重系数,该系数的取值范围为0~1。计算公式为:

$$\alpha_p = \frac{dist_{\max} - d}{dist_{\max} - dist_{\min}} \quad (5)$$

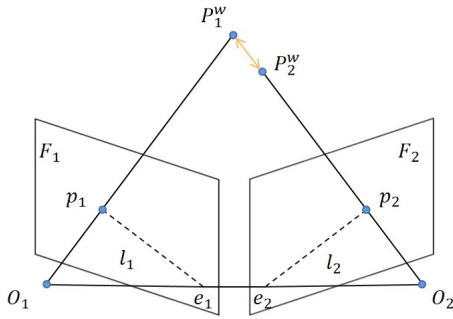


Fig. 7 Epipolar constraint in a dynamic environment

图7 动态环境下的极线约束

匹配特征点对的距离越大,说明该特征点有越大的概率成为动态特征点,在后续优化中可为其重投影误差分配更小的权重系数。同理, α_l 为特征线的权重系数,其计算方式与特征点类似。首先,遍历图像中剩余的静态特征线,根据之前夹角检测计算出的 θ_1 、 θ_2 、 θ_3 ,按照以下公式得到夹角差值的平均值:

$$\theta_{average} = \frac{\|\theta_1 - \theta_2\| + \|\theta_1 - \theta_3\| + \|\theta_2 - \theta_3\|}{3} \quad (6)$$

权重系数的计算公式为:

$$\alpha_l = \cos(\theta_{average}) \quad (7)$$

根据光流法假设,静态特征点在两幅图像中具有相同的运动,因此它们所组成的特征线夹角应为0。当夹角平均值越大时,线特征为动态的概率越大,而对夹角平均值进行余弦运算后的权重越小。同时,余弦运算能确保权重系数的范围在0~1之间。

为方便观测加入自适应权重后的效果,本文可视化不同权重的点特征和线特征。具体如图8所示。图8(a)为原始图片,图8(b)为增加自适应权重后提取到的点线特征,红色的特征点和特征线表示在位姿优化过程中拥有更小的权重系数,有更大的概率为动态的点和线。根据前文的动态区域识别算法,图8中坐在椅子上的两个人由于大多数特征点和特征线为静态,被判别为静态区域。然而,左边人的头部、手部和腿部是局部动态的,右边人的头部和手部也是局部动态的。在SPL-SLAM系统中,这些区域中的特征点和特征线被赋予更小的权重,使其在后续位姿优化中起到更弱的作用,从而使整个系统能充分利用完全静态的特征点和特征线,得到更准确的位姿。

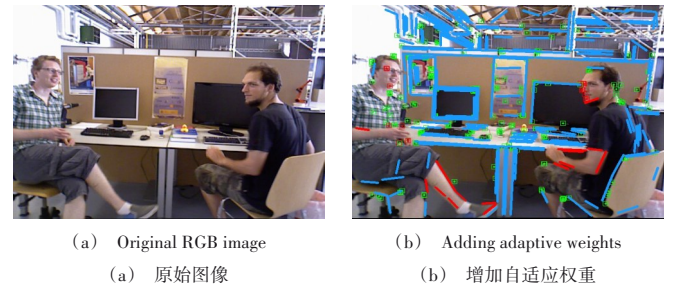


Fig. 8 Feature points and feature lines before and after adding adaptive weights

图8 增加自适应权重前后的特征点和特征线

1.4 稠密3D点云地图构建

点云是一种用于表示三维对象或场景的数据结构,由许多离散点组成,每个点都有其自己的空间坐标以及颜色、法线等附加信息。点云在各个领域都有广泛应用,包括机器人导航与感知、自动驾驶、城市建模、工业制造和医学成像等。点云构建线程专注于生成静态背景的三维点云地图。传统的点云地图构建方法是将RGB-D图像序列和由跟踪线程估算得到的相机姿态作为输入。首先基于图像的每一帧获得局部点云地图,然后使用估算的相机姿态将局部点云拼接成全局点云地图。然而,以这种方式构建的点云地图通常不考虑动态物体的影响。在高动态环境中构建时,这样的点云地图可能包含许多由移动物体引起的噪点。本文结合图像分割得到的动态区域过滤掉因动态物体而产生的噪点,生成更清晰的点云地图。具体如图9所示。

图9(a)为ORB-SLAM3添加稠密建图线程后得到的地图。可以看出,未过滤动态物体的点云地图中存在许多动态噪点,难以应用于实际需求中。伴随着移动机器人导航、交互等任务对地图的要求越来越高,一般形式的地图

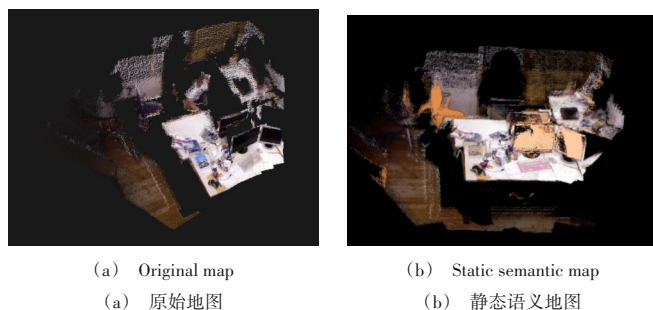


Fig. 9 Construction of dense 3D point cloud maps

图9 稠密3D点云地图构建

已无法满足。如图9(b)所示,为保证移动机器人能够更好地识别地图中的物体,本文对地图中的物体进行语义标注,同时结合图像分割得到的静态物体掩膜构建出只包含静态物体且含有语义信息的点云地图,其将桌面上的电脑标注为黄色,键盘标注为粉色,桌边的椅子标注为橙色,从而提升了机器人对周围环境的感知能力。

2 实验方法与结果分析

2.1 实验环境与参数设置

所有实验均在配备 Intel Core i7-8750H (2.20 GHz) CPU、8 GB RAM、NVIDIA GTX 1050Ti GPU 和 Ubuntu 20.04 系统的笔记本电脑上运行。在实验中,本文对关键参数进行了精细调整,以确保算法的性能和稳定性。具体参数设置如下:ORB特征提取器的特征点数量设为500,特征匹配的阈值设为0.7,滤波器的窗口大小设为5×5,深度图的分辨率设为640×480,相机内参 $f_x=522.0$, $f_y=522.0$, $c_x=317.0$, $c_y=252.0$,相机畸变系数 $k_1=0.0$, $k_2=0.0$, $p_1=0.0$, $p_2=0.0$ 。

2.2 实验数据集

TUM数据集是由德国慕尼黑工业大学计算机视觉实验室提供的一个广泛使用的RGB-D数据集,旨在支持机器人学和计算机视觉领域的研究。该数据集包含多种室内场景的RGB-D图像序列以及精确的相机位姿数据,适用于SLAM和3D重建等研究。其中,fr3/w序列属于Dynamic Objects类别,主要用于评估视觉SLAM算法在动态场景中的性能。该类别包含多个子序列,如fr3/walking_static、fr3/walking_xyz、fr3/walking_halfsphere和fr3/walking_rpy等,这些序列通过模拟不同的动态场景和相机运动,为算法的鲁棒性评估提供了丰富的测试数据。例如,在fr3/walking_static序列中,一个人在行走而另一个人保持静止,华硕Xtion传感器固定在行走者的身上,记录了行走过程中的RGB-D图像和相机位姿数据,用于评估算法在动态场景中对静止对象的定位和建图能力。在fr3/walking_xyz序列中,一个人在行走,同时相机沿x-y-z轴移动,记录了行走过程中相机的三维运动,用于评估算法在动态场景中对复杂运动的鲁棒性。在fr3/walking_half-

sphere序列中,一个人在行走,同时相机沿半球形轨迹移动,记录了行走过程中相机的半球形运动,用于评估算法在动态场景中对复杂轨迹的鲁棒性。在fr3/walking_rpy序列中,一个人在行走,同时相机在原地沿主轴(滚动—俯仰—偏航)旋转,记录了行走过程中相机的旋转运动,用于评估算法在动态场景中对旋转运动的鲁棒性。TUM数据集为研究人员提供了丰富的动态场景数据,有助于评估和改进视觉SLAM算法在复杂环境中的性能。

2.3 评价指标

绝对轨迹误差(Absolute Trajectory Error, ATE)用于衡量算法估计的位置与真实位置之间的差异,是评估整体SLAM算法性能的指标,可以非常直观地反映算法精度和轨迹全局一致性。需要注意的是,估计位姿和真实轨迹通常不在同一坐标系中,需要先将两者对齐。对于双目SLAM和RGB-D SLAM,需要通过最小二乘法计算一个从估计位姿到真实位姿的转换矩阵。单目相机具有尺度不确定性,需要计算一个从估计位姿到真实位姿的相似转换矩阵。第*i*帧的ATE计算公式为:

$$F_i = Q_i^{-1} S P_i \quad (8)$$

此外,也可使用均方根误差(Root Mean Square Error, RMSE)来计算ATE。在已知总数*n*与间隔的情况下,可以采用RMSE统计这个误差,同时结合实际情况综合衡量算法表现。遍历所有间隔的RMSE平均值计算公式为:

$$RMSE(F_{1:n}, \Delta) = \left(\frac{1}{m} \sum_{i=1}^m \|trans(F_i)\|^2 \right)^{\frac{1}{2}} \quad (9)$$

2.4 比较实验

DS-SLAM使用神经网络进行动态特征滤除,将其与本文提出的SPL-SLAM进行比较,以验证所提算法在性能上的改进。表1为SPL-SLAM与DS-SLAM在动态数据集序列中的比较结果。可以看出,SPL-SLAM在动态数据集序列中的表现优于DS-SLAM,特别是在fr3/w/xyz和fr3/w/rpy这样高度动态的数据集中性能改进显著。DS-SLAM在语义分割性能方面的限制导致对一些动态点的无效消除,从而降低了相机姿态估计的准确性。而本文使用YOLOv8实例分割,提供了更加精确的分割能力,有效避免了DS-SLAM的分割缺陷。此外,本文还使用具有自适应权重的误差函数来优化姿态估计,在局部运动环境中表现出更优异的性能。综上所述,SPL-SLAM系统在轨迹估计方面表现出色,并在高度动态的场景中展现出更强的鲁棒性。然而,在fr3/s/rpy这样的低动态序列中,其优势不太明显。

本文还对SPL-SLAM与其他动态SLAM系统进行了比较,包括DynaSLAM^[16]、ORB-SLAM3^[21]、CFP-SLAM^[23]、Blitz-SLAM^[24]、YOLO-SLAM^[25]和PLD-SLAM^[26]。定量实验结果见表2,其中破折号表示原始文献中未提供相应的数据。在5个实验数据序列中,与ORB-SLAM3系统相比,SPL-SLAM系统进行相机姿态估计的准确性提高了76.5%。与DynaSLAM、CFP-SLAM、Blitz-SLAM、YOLO-

Table 1 Absolute trajectory error comparison between SPL-SLAM and DS-SLAM systems

表 1 SPL-SLAM 与 DS-SLAM 系统 ATE 比较 (m)

序列	DS-SLAM				SPL-SLAM			
	RMSE	Mean	Median	STD	RMSE	Mean	Median	STD
fr3/w/xyz	0.028 8	0.020 6	0.016 6	0.020 1	0.012 9	0.011 2	0.010 5	0.006 8
fr3/w/rpy	0.475 2	0.405 1	0.285 6	0.248 4	0.082 7	0.059 2	0.052 6	0.041 5
fr3/w/half	0.028 3	0.024 2	0.020 9	0.014 7	0.031 5	0.026 9	0.022 6	0.016 4
fr3/w/static	0.007 6	0.006 9	0.006 3	0.003 6	0.011 3	0.010 4	0.009 8	0.004 3
fr3/s/rpy	0.021 3	0.015 5	0.011 8	0.014 6	0.019 8	0.014 9	0.010 9	0.012 2

Table 2 Comparison of ATE between SPL-SLAM and other systems

表 2 SPL-SLAM 与其他系统的 ATE 比较 (m)

序列	ORB-SLAM3	Dyna-SLAM	CFP-SLAM	YOLO-SLAM	Blitz-SLAM	PLD-SLAM	Ours
fr3/w/xyz	0.535 9	0.015 8	0.014 1	0.019 4	0.015 3	0.014 4	0.012 9
fr3/w/rpy	0.740 8	0.040 2	0.036 8	0.093 3	0.035 6	0.221 2	0.082 7
fr3/w/half	0.396 2	0.027 4	0.023 7	0.026 8	0.025 6	0.026 1	0.031 5
fr3/w/static	0.354 0	0.008 0	0.006 6	0.009 4	0.010 2	0.006 5	0.011 3
fr3/s/rpy	0.021 4	0.030 2	0.025 3	—	—	0.022 2	0.019 8

SLAM 和 PLD-SLAM 相比, SPL-SLAM 系统在 fr3/w/xyz 和 fr3/s/rpy 数据集上表现更好, ATE 分别提升了 26.5%、15.1%、15.7%、33.5% 和 10.5%, 但在其他数据集上略显不及。

为评估引入线特征和自适应权重对系统的有效性, 本文利用 TUM 数据集提供的轨迹评估工具。如图 10 所示, 真实轨迹以黑色表示, 估计轨迹以蓝色表示, 它们之间的差异由红线表示。与 DS-SLAM 相比, 本文方法在 fr3/w/xyz 和 fr3/w/rpy 的动态序列中均表现出卓越性能。DS-SLAM 在语义分割性能上的局限性导致其无法有效剔除一些动态点, 从而影响了相机姿态估计的准确性。为解决这一问题, 本文方法采用 YOLOv8 实例分割增强分割能力, 有效弥补了 DS-SLAM 在分割方面的不足。在 fr3/w/xyz 和 fr3/w/rpy 序列中, 动态物体运动的幅度较大, 通过上文的动态区域移除算法可有效过滤掉动态物体。在局部运动的环境下, 通过引入带有自适应权重的损失函数可以滤除掉一部分动态特征。因此, 本文算法在这两个数据集中表现出一定优势。而在 fr3/w/static 序列中, 人的运动幅度不是很大, 相邻帧中的人物掩膜轮廓差别也不是很大, 本文算法会将其当作静态物体, 随后再通过自适应权重来区分动态特征和静态特征, 因此在这个数据集中优势不明显。总体来说, SPL-SLAM 系统在轨迹估计方面表现出色, 并在高动态场景中展现出更强的鲁棒性。然而, 在低动态序列(如 fr3/s/static 序列)中, 其优势并不明显, 这可以归因于低动态场景中动态物体的运动幅度较小, 剔除动态特征对相机姿态估计精度的影响相对较小。

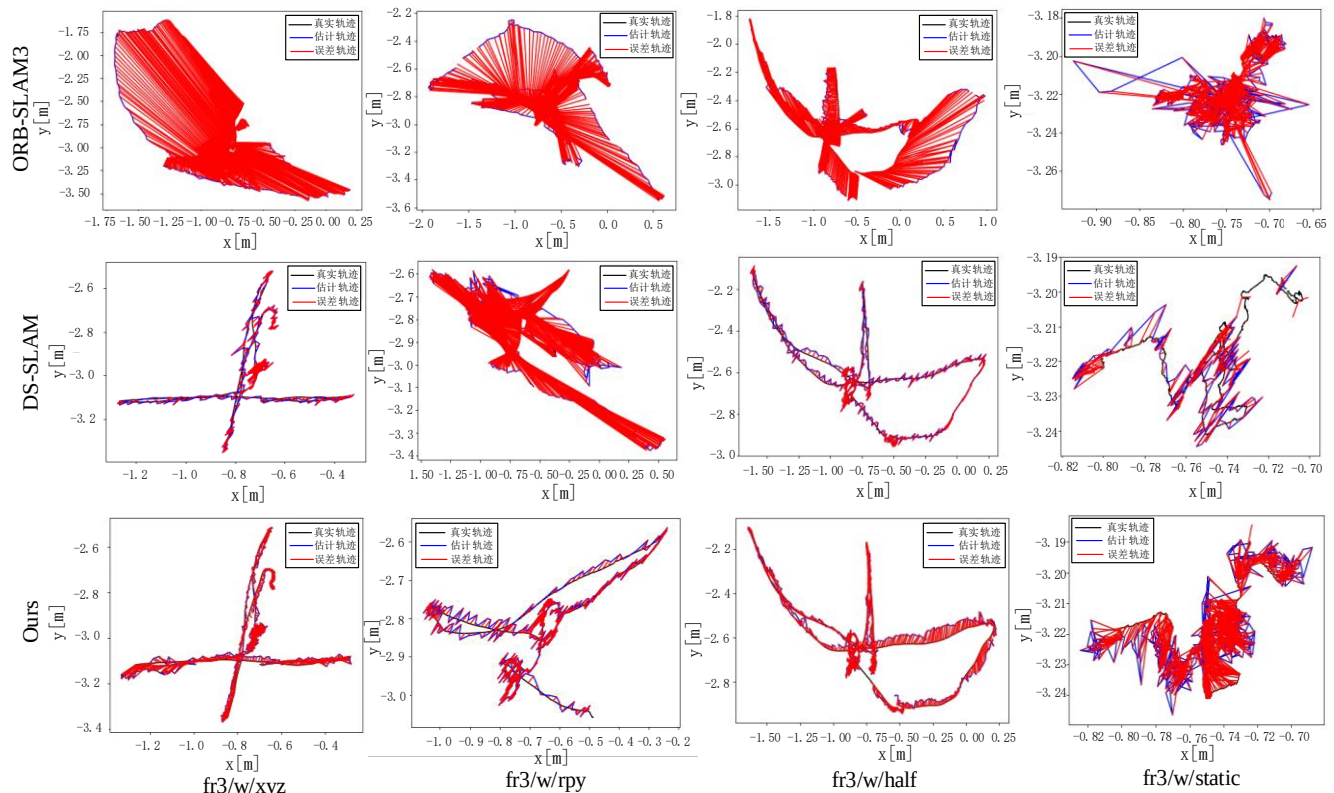


Fig. 10 Comparison of the estimated trajectories

图 10 估计轨迹比较

图 11 为不同网络的分割效果比较。可以看出, 基于 YOLOv8-seg 实例分割网络开发的 SPL-SLAM 在分割性能上优于 DS-SLAM。具体来说, 在一些图片边缘, YOLOv8-

seg 处理得更加平滑, 而 DS-SLAM 会在边缘丢失一些信息。同时, YOLOv8-seg 对一些小型、不完整或有遮挡物体的分割更加准确, 例如图 11 中的键盘和被遮挡的椅子。

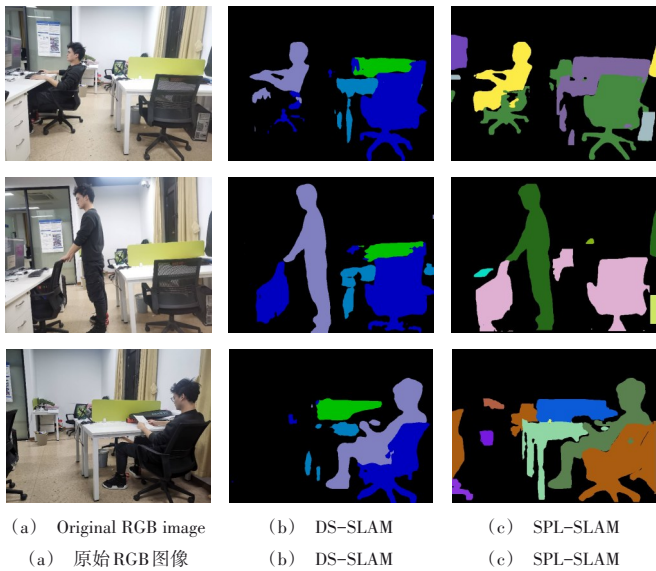


Fig. 11 Comparison of segmentation effects among different networks
图 11 不同网络分割效果比较

2.5 消融实验

为进一步验证利用实例分割和带自适应权重的点线特征误差函数进行相机姿态估计的性能,本文在 TUM RGB-D 数据集上进行消融实验,采用 ATE 评估轨迹误差。结果见表 3。可以看出,当依次引入 LSD 线特征算法和自适应权重时,本文系统在 fr3/w/xyz、fr3/w/static 和 fr3/w/rpy 序列上的 ATE 值逐渐降低,使相机姿态估计的准确性提升了 1.2%、2.2% 和 0.8%。这是由于 LSD 线特征算法能够提取图像中的线段特征,这些特征在某些场景下比点特征更具稳定性和区分性,尤其是在处理具有明显直线结构的场景时。线特征的引入为相机姿态估计提供了更多几何约束,从而提高了估计的准确性。而自适应权重能够根据特征的可靠性和重要性动态调整权重,使模型更加关注那些对姿态估计影响较大的特征,这种动态调整机制能有效提高特征匹配精度,提升相机姿态估计的准确性。

Table 3 Results of ablation experiment

表 3 消融实验结果 (m)			
序列	YOLOv8	YOLOv8 + LSD	YOLOv8 + LSD + 自适应权重
fr3/w/xyz	0.012 976	0.012 945	0.012 909
fr3/w/rpy	0.082 932	0.082 801	0.082 748
fr3/w/half	0.031 625	0.031 523	0.031 549
fr3/w/static	0.011 356	0.011 332	0.011 324

2.6 真实环境下特征点和特征线滤除实验

为进一步评估 SPL-SLAM 系统在真实环境下的性能,使用 Kinect2 相机捕捉动态室内场景中的图像,并按照 TUM 数据集格式处理收集到的图像。图 12 为 DS-SLAM 和 SPL-SLAM 在真实场景中的静态特征提取效果比较。图中描绘了一个场景,其中一个人从左边桌子旁的椅子上站起来,朝右边的桌子走去,然后坐下来。实验结果表明,本文系统在消除动态特征点以及有效利用环境中的线特

征方面优于 DS-SLAM。

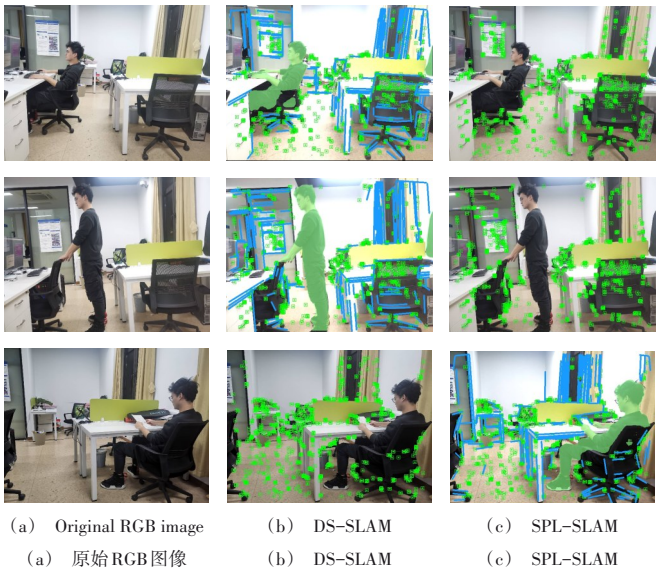


Fig. 12 Comparison of static feature extraction effects in real scenes
图 12 真实场景中静态特征提取效果比较

3 结语

本文提出一种新颖的语义 RGB-D SLAM 系统 SPL-SLAM,用于动态环境中的机器人实时定位。为提高相机位姿估计的准确性和系统的鲁棒性,系统首先利用 YOLOv8 获取精确的轮廓和语义信息,以识别人物作为动态对象标签;同时通过引入线特征、极线约束和角度检测过滤潜在的动态点特征。如果分割对象包含潜在的动态点特征,则认为该对象是动态的。在这种情况下,该对象的区域被标记为动态,该动态区域的点特征和线特征被移除。其次,设计了一种自适应权重的误差函数,并结合剩余静态点特征和线特征来优化相机位姿。最后,结合优化的位姿和深度信息从深度图像中创建静态点云地图,为机器人或自主系统提供更丰富、更高级别的场景理解和感知能力。为评估 SPL-SLAM 系统的性能,本文在公开可用的 TUM RGB-D 数据集以及实际场景上进行了实验。实验结果表明,该系统提高了相机位姿估计的准确性和系统的鲁棒性。虽然本文使用深度学习改进了 ORB-SLAM3 的部分模块,但后端优化仍然依赖传统的数学模型,因此后续考虑使用端到端的神经网络构造整个 SLAM 系统。

参考文献:

[1] HESS W, KOHLER D, RAPP H, et al. Real-time loop closure in 2D LIDAR SLAM[C]//2016 IEEE International Conference on Robotics and Automation (ICRA), 2016: 1271-1278.
[2] GRISETTI G, STACHNISS C, BURGARD W. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling[C]//Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005: 2432-2437.

- [3] KOHLBRECHER S, VON STRYK O, MEYER J, et al. A flexible and scalable SLAM system with full 3D motion estimation[C]//2011 IEEE International Symposium on Safety, Security, and Rescue Robotics, 2011: 155–160.
- [4] MUR-ARTAL R, TARDOS 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.
- [5] ENGEL J, SCHOPS T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM [C]//European Conference on Computer Vision, 2014: 834–849.
- [6] ENDRES F, HESS J, ENGELHARD N, et al. An evaluation of the RGB-D SLAM system[C]//2012 IEEE International Conference on Robotics and Automation, 2012: 1691–1696.
- [7] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 611–625.
- [8] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: fast semi-direct monocular visual odometry [C]//2014 IEEE International Conference on Robotics and Automation (ICRA), 2014: 15–22.
- [9] FANG Y, DAI B. An improved moving target detecting and tracking based on optical flow technique and kalman filter[C]//2009 4th International Conference on Computer Science & Education, 2009: 1197–1202.
- [10] LI S, LEE D. RGB-D SLAM in dynamic environments using static point weighting [J]. IEEE Robotics and Automation Letters, 2017, 2(4): 2263–2270.
- [11] SUN Y, LIU M, MENG M Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach[J]. Robotics and Autonomous Systems, 2017, 89: 110–122.
- [12] WANG R, WAN W, WANG Y, et al. A new RGB-D SLAM method with moving object detection for dynamic indoor scenes[J]. Remote Sensing, 2019, 11(10): 1143.
- [13] NI J, WANG L, WANG X, et al. An improved visual SLAM based on map point reliability under dynamic environments[J]. Applied Sciences, 2023, 13(4): 2712.
- [14] ZHOU Y, TAO F, FU Z, et al. RGB-D SLAM in indoor dynamic environments with two channels based on scenario classification[J]. Measurement Science and Technology, 2023, 34(11): 115007.
- [15] ZHANG L, WEI L, SHEN P, et al. Semantic SLAM based on object detection and improved octomap [J]. IEEE Access, 2018, 6: 75545–75559.
- [16] BESCOS B, FACIL J M, CIVERA J, et al. DynaSLAM: tracking, map**, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076–4083.
- [17] YU C, LIU Z, LIU X J, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018: 1168–1174.
- [18] ZHENG Z, LIN S, YANG C. RLD-SLAM: a robust lightweight VI-SLAM for dynamic environments leveraging semantics and motion information[J]. IEEE Transactions on Industrial Electronics, 2024, 71(11): 14328–14338.
- [19] ISLAM Q U, IBRAHIM H, CHIN P K, et al. ARD-SLAM: accurate and robust dynamic SLAM using dynamic object identification and improved multi-view geometrical approaches [J]. Displays, 2024, 82: 102654.
- [20] LUO Y, RAO Z, WU R. FD-SLAM: a semantic SLAM based on enhanced fast-SCNN dynamic region detection and DeepFillv2-Driven background inpainting[J]. IEEE Access, 2023, 11: 110615–110626.
- [21] CAMPOS C, ELVIRA R, RODRIGUEZ 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.
- [22] VON GIOI R G, JAKUBOWICZ J, MOREL J M, et al. LSD: a line segment detector[J]. Image Processing On Line, 2012, 2: 35–55.
- [23] HU X, ZHANG Y, CAO 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), 2022: 4399–4406.
- [24] FAN Y, ZHANG Q, TANG Y, et al. Blitz-SLAM: a semantic SLAM in dynamic environments[J]. Pattern Recognition, 2022, 121: 108225.
- [25] WU W, GUO L, GAO H, et al. YOLO-SLAM: a semantic SLAM system towards dynamic environment with geometric constraint [J]. Neural Computing and Applications, 2022, 34(8): 6011–6026.
- [26] ZHANG C, HUANG T, ZHANG R, et al. PLD-SLAM: a new RGB-D SLAM method with point and line features for indoor dynamic scene[J]. ISPRS International Journal of Geo-Information, 2021, 10(3): 163.

(责任编辑:尹晨茹)