

动态场景下融合改进 YOLOv7 的视觉 SLAM 算法<sup>\*</sup>史 涛<sup>1</sup> 校诺政<sup>1</sup> 丁 焱<sup>1</sup> 许金东<sup>2</sup>

(1. 天津理工大学电气工程与自动化学院 天津 300384; 2. 冀东油田勘探开发建设监督中心 唐山 063200)

**摘 要:**针对传统的视觉同步定位与地图构建(SLAM)在动态场景下容易受到运动物体干扰,导致位姿估计精准度和鲁棒性下降的问题,提出了一种基于目标检测网络的视觉 SLAM 算法。该算法通过在 ORB-SLAM2 的跟踪线程中新增动态特征点检测剔除模块,从而利用静态特征点进行位姿估计。首先,选择 YOLOv7 作为目标检测的主干网络,结合 GhostNet 轻量化卷积网络和具有 SE 注意力机制的卷积(Conv\_SE),以有效地检测周围环境;其次,对检测到的物体进行分类处理,剔除动态物体特征点,通过几何约束的方法进一步检测和剔除潜在运动物体;最后,仅利用静态特征点进行特征匹配和位姿估计。在 TUM 数据集上的验证结果表明,与 ORB-SLAM2 相比,提出的算法在动态 Walk 序列下,绝对轨道误差(ATE)的均方根误差平均减少 96.5%,在其他动态序列下也有改进效果。实验证明,该算法在动态场景下能够显著提升系统的定位精度和鲁棒性。

**关键词:**视觉 SLAM;动态场景;目标检测;位姿估计

**中图分类号:** TN2      **文献标识码:** A      **国家标准学科分类代码:** 520.604

## Visual SLAM algorithm for fusing improved YOLOv7 in dynamic scenes

Shi Tao<sup>1</sup> Xiao Nuozheng<sup>1</sup> Ding Yao<sup>1</sup> Xu Jindong<sup>2</sup>

(1. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China;

2. Jidong Oilfield Exploration and Development Construction Supervision Center, Tangshan 063200, China)

**Abstract:** In response to the susceptibility of traditional visual simultaneous localization and mapping (SLAM) to disturbances from moving objects in dynamic scenes, a visual SLAM algorithm based on object detection networks is proposed. This algorithm introduces a module for detecting and rejecting dynamic feature points in the tracking thread of ORB-SLAM2, thereby utilizing static feature points for pose estimation. Firstly, YOLOv7 is chosen as the backbone network for object detection, combined with GhostNet lightweight convolutional networks and convolution with SE attention mechanism (Conv\_SE) for effective environmental detection. Secondly, the detected objects are processed through classification, rejecting feature points associated with dynamic objects, and employing geometric constraints to further identify and remove potential moving objects. Finally, only static feature points are used for feature matching and pose estimation. Validation results on the TUM dataset indicate that compared to ORB-SLAM2, the proposed algorithm achieves an average reduction of 96.5% in the root mean square error (RMSE) of absolute trajectory error (ATE) in dynamic walk sequences and shows improvement in other dynamic sequences as well. Experimental evidence demonstrates that this algorithm significantly enhances the localization accuracy and robustness of the system in dynamic scenarios.

**Keywords:** vision SLAM; dynamic scene; object detection; pose estimation

## 0 引 言

移动机器人同步定位与地图构建<sup>[1-2]</sup> (simultaneous

localization and mapping, SLAM), 是指机器人在自身位置不确定的条件下, 在部分已知或完全未知的环境中运动时, 根据位姿估计和传感器探测数据进行自定位, 并同时

收稿日期: 2024-04-07

<sup>\*</sup> 基金项目: 国家自然科学基金青年科学基金(62103298)项目资助

构建增量式地图。利用地图进行自主定位和导航的能力,被认为是机器人真正实现自主作业的关键和基础。这项技术在自动驾驶、扫地机器人等领域有广泛的应用。

SLAM按照传感器的不同可以分为基于激光雷达的激光SLAM和基于相机的视觉SLAM,视觉SLAM因传感器提供的信息丰富以及使用成本较低从而获得了较多的关注和使用。

目前较为成熟的视觉SLAM系统有Mono-SLAM<sup>[3]</sup>、LSD-SLAM<sup>[4]</sup>以及ORB-SLAM2<sup>[5]</sup>等。这些传统的视觉SLAM方法在静态场景中能够取得令人满意的效果,但是在现实的环境中存在动态物体,并且会影响SLAM系统定位和建图的准确性。针对此问题国内外一些学者从不同的角度提出解决方法。Tan等<sup>[6]</sup>在2013年提出的IS-MAR通过关键帧的特征点比较方法,可以在静态点较少的情况下辨认动态特征。刘建军等<sup>[7]</sup>从特征点提取的角度来解决系统在特殊环境特征点提取不足等问题。邓鹏等<sup>[8]</sup>为了解决SLAM系统运行环境复杂的问题,采用多传感器融合的方法来提高系统的稳定性和鲁棒性。

在深度学习方向,DS-SLAM<sup>[9]</sup>通过SegNet语义分割方法和运动一致性去除环境中的动态物体,提高了SLAM系统在高动态环境下运行的鲁棒性。Dyna-SLAM<sup>[10]</sup>使用MaskR-CNN语义分割并结合多视图几何和的方法来检测动态物体,并通过对被这些动态物体遮挡的背景进行重新绘制来构建静态地图,但是语义分割存在计算量大对设备的硬件要求较高和分割速度慢,不能满足实时性的问题。Wang等<sup>[11]</sup>提出的一种基于深度学习的视觉SLAM算法,结合了实例分割网络和光流法,对动态对象上的特征进行检测剔除。DRSO-SLAM<sup>[12]</sup>是一种用于室内动态场景的SLAM算法,采用语义分割后建立先验信息库,进行粗略的自运动估计,再通过几何约束法来过滤掉动态特征点。RDS-SLAM<sup>[13]</sup>提出了一种基于语义分割可实时的SLAM系统,通过关键帧的选取策略的改变来完成最新的语义信息的处理,从而完成系统实时性的要求。刘钰嵩等<sup>[14]</sup>使用光流和语义信息相结合的方法来解决动态物体干扰问题,最后构建静态语义八叉树地图,但系统的计算量大,导致实时性不足。吴丽凡等<sup>[15]</sup>提出了一种基于YOLO算法去除动态特征点的方法,其采用的是YOLOv3算法在复杂的动态环境下检测精度有待提高。陈敏<sup>[16]</sup>采用改进后的YOLOv3算法来提高复杂环境的检测精度。CFP-SLAM<sup>[17]</sup>提出了一种基于YOLOv5s和静态概率的算法,通过计算物体、关键点的静态概率将其作为权重参与到相机位姿估计当中,在完成实时性的同时减少了计算量。

综上,基于传统几何约束的方法在运动场景下运行效果差;采用深度学习语义分割的方法虽然可以有效的将动态物体去除,并且保留较多的静态特征点来完成位姿估计,但是这种算法对硬件要求比较高,耗时严重,实时性差。采用YOLO目标检测算法,在运行速度上得到了极

大的提升,但是丧失了部分精度,当相机移动速度稍快时,检测算法就会失效,并且没有应对潜在运动物体的方法。

针对以上研究中存在的问题,本文提高了SLAM系统在动态环境下位姿估计的精确性和鲁棒性,基于ORB-SLAM2框架增加了动态特征点剔除模块,采用改进YOLOv7的轻量级目标检测网络检测动态物体,并结合多视角几何约束来完成环境存在的动态物体特征点的识别剔除,然后利用静态特征点对相机进行位姿估计。最后在德国慕尼黑工业大学提供的TUM数据集下,验证了本文所提方法的有效性和可行性。

## 1 动态特征点剔除

### 1.1 目标检测算法

YOLO算法<sup>[18]</sup>作为one-stage目标检测算法最典型的代表,其基于深度神经网络进行对象的识别和定位,运行速度可满足于实时系统。YOLOv7算法<sup>[19]</sup>,使用了一些新技术和优化,进一步提高了检测性能和速度,是目前非常流行的目标检测算法。在准确率和速度上超越了以往的YOLO系列。与以往版本相比,在COCO数据集上,它的平均精度(AP)比YOLOv4<sup>[20]</sup>高出了约5%。YOLOv5<sup>[21]</sup>的运行速度较快,但精度稍低,YOLOv7使用了一种新的设计思路来提高检测速度和精度,YOLOv7的精度更高,考虑到在实际的环境中会出现动态物体运动速度快画面模糊等问题,为了避免错检漏检情况出现,同时满足SLAM实时性,降低算力,本文以YOLOv7作为基础算法。

在对象检测过程中,为了减少YOLOv7算法的计算成本,方便部署,因此改进了一种基于YOLOv7的轻量化目标检测算法。

将普通的卷积层换成一种更加轻量化的GhostConv卷积,以减少模型参数的数量和计算复杂性,结合原有的批量归一化(batch normalization, BN)层和激活函数(sigmoid linear unit, SiLU),结合成新的GBS模块,来完成所需特征的提取,改进后算法的参数量由原来的3 762万下降至3 601万。

为了增强对动态信息的捕捉能力,将挤压激励注意力机制(squeeze-and-excitation, SE)同时与卷积进行结合提出Conv\_SE,应用在Backbone中特有的聚合网络(effective long-range aggregation network, ELAN)模块中,使网络结构更关注动态物体的特征信息,改进后的检测准确率由95%提高至96.8%。

本文所使用改进后的部分如图1所示,将普通的卷积CBS更换为更加轻量化的GBS模块,应用在普通的聚合网络ELAN模块中。与SE注意力机制结合提出一种新的卷积操作,替换原ELAN模块中的最后一层操作。

### 1.2 几何约束去除潜在动态点

在使用目标检测算法检测出动态物体后,直接去除相应的动态点即可,但是在实际环境中存在一些潜在的运动

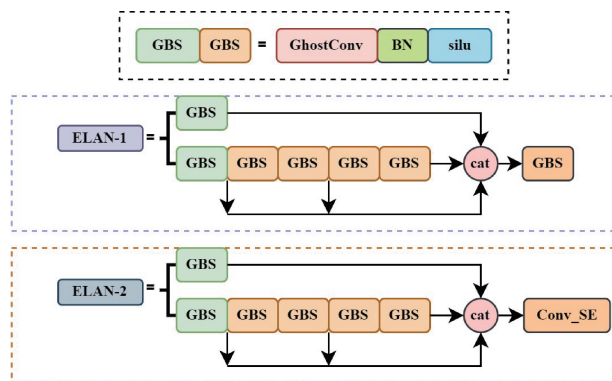


图1 ELAN-1、ELAN-2 结构

Fig. 1 ELAN-1, ELAN-2 structural diagrams

物体,比如被人移动的椅子、手中拿着的书本等,在这种情况下单纯的使用深度学习的方法是不够严谨准确,结合对极几何约束的方法来去除剩余的动态点。

对极几何<sup>[22]</sup> (epipolar geometry) 又叫对极约束,根据图像二维平面信息来估计相机帧间运动的相对位姿关系。将 RGB-D 相机放置在机器设备上,此时计算相机位姿便得到了自身位姿。通过相机的运动可以得到相邻时刻的两帧图像,设  $X_1$  是空间存在的特征点,  $I_1$  和  $I_2$  是相机在前一时刻和当前时刻获得的 RGB 图像,三维空间点  $X_1$  分别在两幅图像  $I_1$ 、 $I_2$  中成像为点  $P_1$ 、 $P_2$ ,  $X_1$  与相机的中心  $C_1$ 、 $C_2$  构成的平面记为  $\pi$ , 则上述对极几何关系如图 2 所示。

相机中心  $C_1$ 、 $C_2$  与空间点  $X_1$  形成的平面为极平面  $\pi$ , 相机中心  $C_1$ 、 $C_2$  的连线成为基线,基线与图像平面的交点  $e_1$ 、 $e_2$  称为极点,则极平面  $\pi$  与图像平面  $I_2$  的交线  $L$  为极线,所需要的约束条件就是在极线上完成的。

如图 2 所示,  $C_1$  和  $C_2$  分别为前一时刻和当前时刻相机对应的光心位置,  $X_1$  是在相邻矩下观测到的映射点,若特征匹配已完成便能得到  $C_1$  处观测处点  $X_1$  的空间坐标,  $P_1 = [x_1, y_1]^T$  在  $I_2$  成像的坐标为  $P_2 = [x_2, y_2]^T$ , 齐次形式分别表示为:

$$P_1 = [x_1, y_1, 1]^T \quad (1)$$

$$P_2 = [x_2, y_2, 1]^T \quad (2)$$

则此时从点  $C_1$  观测到  $X_1$  的位置,得出的极线  $L$  为:

$$L = FP_1 = F \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} A \\ B \\ C \end{bmatrix} \quad (3)$$

式中:  $F$  为对应的基础矩阵。从理论上讲,如果姿态绝对准确,当前时刻地图点的投影将在由极点  $e_2$  和  $P_2$  组成的极线  $L$  上。然而,由于动态物体的运动,所观测的特征点就不会正常投影到极线  $L$  上,动态地图点作为异常数据的投影不满足约束。所以便可用极线与点  $P_2$  的位置关系来判断所观测的特征点是否为动点,再继续下一步的工作。点  $P_2$  到极线  $L$  的距离为:

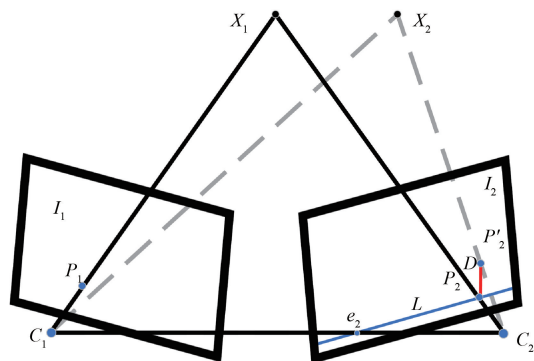


图2 几何约束图

Fig. 2 Figure size sample

$$D = \frac{|P_2^T F P_1|}{\sqrt{\|A\|^2 + \|B\|^2}} \quad (4)$$

理想的情况下静态点到极线的距离  $D$  应该等于 0,但是由于相机在运动过程中受周围环境的影响不能保证静态点就一定投影到极线上。所以设置一定的阈值来减去环境干扰带来的误差,当  $D$  大于设定的阈值时,则判断该特征点为动态点并进一步将其过滤,若  $D$  在阈值的范围之内,便可当静态特征点继续利用。

### 1.3 改进的 ORB-SLAM2 的系统框架

ORB-SLAM2 系统是首个支持单目、双目和 RGB-D 相机的完整开源方案,最大的特点是拥有 3 个并行线程,包括用于特征提取与匹配的跟踪线程、用于更新关键帧和地图点信息的局部建图以及用于检测循环和清除累积误差的闭环检测。能够实时计算出相机位姿并生成三维场景地图。相比较其他视觉 SLAM 系统,ORB-SLAM2 支持多种相机工作,拥有清除累积误差的闭环检测功能,并且系统具有更好的扩展性、稳定性和更高的成熟性。

原系统在周围是静态的环境中能够取得较好的效果,然而实际生活中经常存在动态运动的物体,这对原系统的准确性造成显著的影响。为此,本文基于 ORB-SLAM2 系统进行了改进,在跟踪模块中加入了动态特征点剔除方法。系统提取特征点之前,选择一种轻量级目标检测算法先行对周围环境进行检测识别,再对物体进行分类处理。系统框架如图 3 所示,其中间浅色框内的部分是添加的动态点剔除流程。

针对那些移动性较强,对系统产生较大影响的动态物体,选择直接剔除特征点,不再使用其身上的特征,以最大程度减少对系统的影响。对于自身不会移动但在干预下有潜在运动可能的物体,使用约束法来判断是否在运动,以确定是否保留特征点。最后对于不会移动的物体和未识别的周围环境,采取原始的特征点提取方法,对剔除后的特征点进行下一步位姿估计和局部建图的使用,提高系统在动态环境下的定位精度和鲁棒性。系统使用改进后轻量化的目标检测网络可以准确检测各类物体,使用绿色

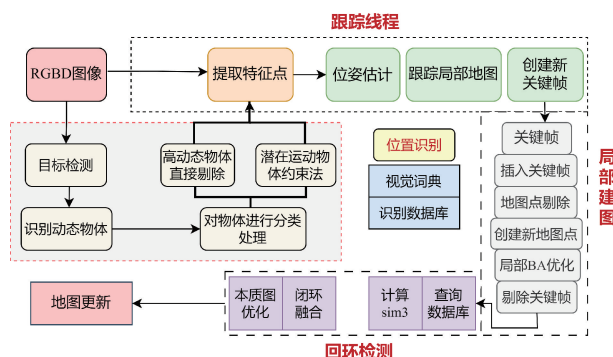


图3 系统框架图

Fig. 3 System framework diagram

或蓝色矩形框显示,如图4所示。系统在运行过程中检测动态物体,并且有效剔除特征点,保留静态物体的特征信息。



图4 剔除动态物体特征点

Fig. 4 Reject dynamic object feature points

## 2 实验结果

对所提出算法的有效性进行验证,使用德国慕尼黑工

业大学提供的 TUM 数据集<sup>[23]</sup>,该数据集包括室内采集的彩色图像和深度图像,在 SLAM 评估位姿中广泛应用。数据集内的 FR3 系列是在室内的办公桌前有两位做动作的人来构建的场景,这些信息类似日常生活场景,符合动态场景信息。视频中的两位实验人员会做不同幅度的动作,分别使用不同的图像序列做对比,根据场景动态变化的程度将其分类,其中“Sitting”类表示两个人坐在椅子上交流,仅手部、头部有简单的动作,被归为低动态序列,而“Walking”序列为两个人在相机前行走,所做的动作幅度更大,被归为高动态序列。相机运动方式如下:1)“static”相机固定在一个位置作为相机的静态状态;2)“xyz”相机沿着 XYZ 三个坐标轴方向做平移运动;3)“halfsphere”相机在一个半径为一米的半球上运动;4)“rpy”相机沿着主轴做旋转、俯拍、仰拍动作。使用绝对轨迹误差(absolute trajectory error, ATE)和相对姿态误差(relative pose error, RPE)作为评估精度的指标,其中 RPE 包括平移漂移和旋转漂移。采用均方根误差(root mean square error, RMSE)、平均误差(mean)和标准差(standard deviation, SD)来表示系统的稳定性和鲁棒性。

### 2.1 与 ORB-SLAM2 算法对比

本文所提出的算法是在 ORB-SLAM2 框架基础上进行的修改,因此将改进后的算法与原框架算法的结果进行了对比。

表1是原算法与改进算法的绝对轨迹估计误差数据比较,包括均方根误差、平均误差和标准差3个数值。在“Walking”的高动态场景下4种不同的相机轨迹,相比 ORB-SLAM2,本文提出的算法在均方根误差均可降低90%以上。由表1可知,本文算法动态场景下具有良好的性能,可以在动态场景中稳定的运行。

表1 绝对轨迹误差结果对比

Table 1 Comparison of absolute trajectory error result

数据集	ORB-SLAM2/m			本文算法/m			提升/%		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
Sitting_static	0.008 6	0.007 5	0.004 3	0.007 1	0.006 1	0.003 6	17.44	18.67	16.28
Walking_xyz	0.703 7	0.659 2	0.330 4	0.016 3	0.014 5	0.007 5	97.68	97.80	97.73
Walking_rpy	0.841 6	0.714 7	0.444 5	0.030 1	0.022 8	0.019 6	96.42	96.81	95.59
Walking_half	0.637 5	0.566 9	0.284 0	0.032 7	0.028 3	0.016 3	94.87	95.01	94.26
Walking_static	0.259 1	0.217 1	0.141 5	0.007 4	0.006 6	0.003 3	97.14	96.96	97.67

表2是相对姿态平移部分和旋转部分的误差数据。原算法与所提出算法相比,在相对轨迹误差方面的降低效果在动态场景下的效果同样明显,在平移部分的均方根误差平均减少90%以上。在低动态“Sitting”序列下相机固定的运动轨迹的误差比原算法效果反而更差,这是由于相机在环境中所检测到的信息少,并且在剔除动态物体的特征点后,系统运行使用的特征点少,对位姿估计的准确性就有所降低,但是误差不存在大幅度的降低。

### 2.2 与其他算法对比

为了验证本文算法的先进性,本文与近年来其他优秀视觉 SLAM 算法对比。首先将相机的轨迹用图像的形式表达,采用“Walking”高动态序列展示结果。图5所示为 ORB-SLAM2 系统的运行轨迹,图6所示为 Dyna-SLAM 系统的运行轨迹,图7所示为本文提出算法的运行轨迹结果。其中黑色和蓝色的线分别代表相机的实际轨迹和估计轨迹,红色线是轨迹的差。图5中轨迹差线段更长甚至



表 2 相对姿态误差平移部分结果对比

Table 2 Comparison of relative pose error translation part result

数据集	ORB-SLAM2/m			本文算法/m			提升/%		
	RMSE	Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD
Sitting_static	0.009 3	0.008 1	0.004 4	0.008 8	0.007 5	0.004 6	5.38	7.41	-4.55
Walking_xyz	0.355 0	0.273 8	0.222 1	0.021 4	0.018 9	0.010 1	93.97	93.10	95.45
Walking_rpy	0.443 4	0.272 6	0.287 5	0.044 1	0.034 1	0.027 8	90.05	87.49	90.33
Walking_half	0.454 7	0.295 0	0.346 0	0.032 1	0.027 3	0.016 9	92.94	90.75	95.12
Walking_static	0.227 8	0.094 4	0.207 3	0.009 6	0.008 6	0.004 3	95.79	90.89	97.93

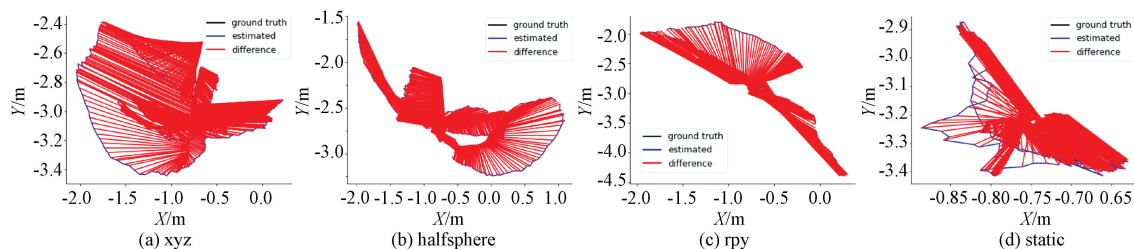


图 5 ORB-SLAM2 算法运行轨迹

Fig. 5 Trajectory of the ORB-SLAM2 algorithm

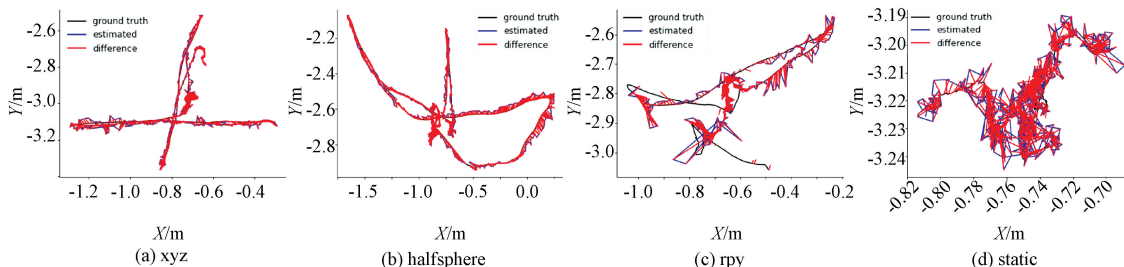


图 6 Dyna-SLAM 算法运行轨迹

Fig. 6 Trajectory of the Dyna-SLAM algorithm

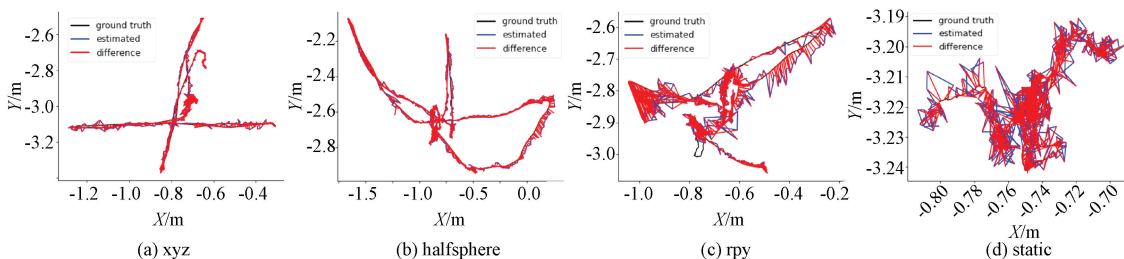


图 7 本文算法运行轨迹

Fig. 7 Trajectory of our method algorithm

占据了图像大部分,这种算法运行的轨迹差值大。由图 6 和 7 可以看出,虽然也有一定的误差但是轨迹差线段都比较短,估计轨迹与实际轨迹基本一致,误差较小,但是 Dyna-SLAM 采取的是分割的方法,在实际运行时对设备的要求较高,运行时间相对较慢。

RDS-SLAM、Dyna-SLAM、DS-SLAM 与本文算法绝对轨迹的均方根误差和标准差结果如表 3 所示。为了更清楚直观的观察到运行结果误差的情况,将对比后均方根

误差最小的数值使用加粗表示。在“xyz”和“rpy”两种相机运动方式下,本文的算法得出的误差是最小的,同样可以看出本文提出算法并不是全部误差数据都优越于其他算法,与最有效的数据相比,提出的算法与其差距很小。DS-SLAM 和 Dyna-SLAM 算法都是采用语义分割的方式来剔除动态物体的特征点,这种方式运行所需的时间更长,在相同的误差情况下,能够满足实时性要求是非常重要的。

表 3 与其他研究结果对比  
Table 3 Comparison of results with other researches (m)

数据集	RDS-SLAM		Dyna-SLAM		DS-SLAM		本文算法	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
Sitting_static	0.008 4	0.004 3	0.010 8	0.005 6	0.006 5	0.003 3	0.007 1	0.003 6
Walking_xyz	0.057 1	0.022 9	0.016 4	0.008 6	0.024 7	0.016 1	0.016 3	0.007 5
Walking_rpy	0.160 4	0.087 3	0.035 4	0.019 0	0.444 2	0.235 0	0.030 1	0.019 6
Walking_half	0.080 7	0.045 4	0.029 6	0.015 7	0.030 3	0.015 9	0.032 7	0.016 3
Walking_static	0.020 6	0.012 0	0.006 8	0.003 2	0.008 1	0.003 6	0.007 4	0.003 3

2.3 实验结果分析

通过上述各项实验数据可知,本文提出算法在动态环境下可取得的良好稳定性和鲁棒性。在采用更加轻量级的目标检测算法识别到动态物体并剔除其特征点后,定位精度更加准确,在“Walking”序列下的绝对轨迹误差和相对姿态误差的均方根数值降低幅度达 90% 以上,各项实验数据验证在动态环境下,无论是相对轨迹误差的平移部分和旋转部分,均比原算法更加准确。在“Sitting”的低动态环境下,由于动态物体动作幅度小,对位姿估计结果产生的影响较少,实验结果表示仍有提升效果。解决动态物体对系统的影响,大大提升了定位的准确性。由图 7 可以清楚地看到,两种轨迹更加接近,误差较小。在有动态物体干扰的环境下,系统可稳定的工作。

3 结 论

针对传统的视觉 SLAM 算法,在动态场景下受到的影响导致定位精度和鲁棒性降低的问题,本文提出一种基于语义信息和几何约束的视觉 SLAM 算法。将 SLAM 系统的前端添加一个动态物体特征点剔除模块,采用一种改进后更加轻量化的 YOLOv7 目标检测算法来识别场景中的物体,再通过对识别到的物体进行动态分类,将动态物体特征点进行剔除,然后通过几何约束的方法剔除掉潜在运动的特征点,最后利用静态特征点进行位姿估计。本文的算法在 TUM 数据集上得到了验证,相比原算法,本文算法在动态场景下 ATE 的均方根误差减少 96.5%,RPE 平移部分均方根误差也降低 93%,并且保证系统的实时性。在未来的工作中,尝试将算法部署到机器人上,采用多传感器融合的方式使系统的应用性更强。

参 考 文 献

[1] 任明宇,陈万米,张圆圆. 融合激光和视觉信息的机器人 SLAM 方法研究[J]. 电子测量技术, 2019, 42(13):92-97.  
REN M Y, CHEN W M, ZHANG Y Y. Service robot SLAM research combining laser and visual information[J]. Electronic Measurement Technology, 2019, 42(13):92-97.  
[2] 宋玉琴,熊高强,曾贺东,等. 多平面点优化的单目

SLAM 方法[J]. 国外电子测量技术, 2021, 40(10): 40-45.  
SONG Y Q, XIONG G Q, ZENG H D, et al. Monocular SLAM method for multi-plane point optimization [J]. Foreign Electronic Measurement Technology, 2021, 40(10): 40-45.  
[3] DAVISON A J, REID I D, MOLTON N D, et al. Mono SLAM: Real-time single camera SLAM [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.  
[4] ENGEL J, SCHOPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM [C]. European Conference on Computer Vision, 2014: 834-849.  
[5] 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.  
[6] TAN W, LIU H, DONG Z, et al. Robust monocular SLAM in dynamic environments [C]. International Symposium on Mixed and Augmented Reality (ISMAR), 2013: 209-218.  
[7] 刘建军, 卢大威, 胡雪花, 等. 基于点线特征的快速单目惯性 SLAM 算法 [J]. 国外电子测量技术, 2022, 41(3): 14-19.  
LIU J J, LU D W, HU X H, et al. Fast monocular inertial SLAM algorithm based on point line features[J]. Foreign Electronic Measurement Technology, 2022, 41(3): 14-19.  
[8] 邓鹏, 罗静. 复杂环境下机器人多传感器融合定位方法 [J]. 电子测量与仪器学报, 2023, 37(12): 48-57.  
DENG P, LUO J. Robot multi-sensor fusion localization method in complex environment [J]. Foreign Electronic Measurement Technology, 2023, 37(12): 48-57.  
[9] YU C, LIU Z, LIU X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments [C]. International Conference on Intelligent Robots and Systems, 2018: 1168-1174.  
[10] BESCOS B, FACIL J M, CIVERA J, et al. Dyna

- SLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.
- [11] WANG Y, DUAN X, SUN Y, et al. A visual SLAM algorithm based on image semantic segmentation in dynamic environment[C]. International Conference on Pattern Recognition and Artificial Intelligence (PRAI), 2021: 401-405.
- [12] YU N, GAN M, YU H, et al. Drso-SLAM: A dynamic RGBD SLAM algorithm for indoor dynamic scenes[C]. Chinese Control and Decision Conference (CCDC), 2021: 1052-1058.
- [13] LIU Y, MIURA J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods [J]. IEEE Access, 2021, 9: 23772-23785.
- [14] 刘钰嵩, 何丽, 袁亮, 等. 动态场景下基于光流的语义 RGBD-SLAM 算法 [J]. 仪器仪表学报, 2022, 43(12): 139-148.
- LIU Y S, HE L, YUAN L, et al. Semantic RGBD-SLAM in dynamic scene based on optical flow [J]. Chinese Journal of Scientific Instrument, 2022, 43(12): 139-148.
- [15] 吴丽凡, 魏东岩, 袁洪. 基于 YOLO 的复杂环境视觉 SLAM 优化方法 [J]. 计算机应用, 2021, 41(S2): 208-213.
- WU L F, WEI D Y, YUAN H. YOLO-based SLAM optimization method for complex environment vision [J]. Journal of Computer Applications, 2021, 41(S2): 208-213.
- [16] 陈敏. 基于改进 YOLO 与光流约束的视觉 SLAM 算法[D]. 南昌: 华东交通大学, 2022.
- CHEN M. Visual SLAM algorithm based on improved YOLO and optical flow constraints [D]. Nanchang: East China Jiaotong University, 2022.
- [17] 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]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022: 4399-4406.
- [18] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [20] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv: 2004. 10934, 2020.
- [21] ZHU X, LU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [22] 艾青林, 刘刚江, 徐巧宁. 动态环境下基于改进几何与运动约束的机器人 RGB-D SLAM 算法 [J]. 机器人, 2021, 43(2): 167-176.
- AI Q L, LIU G J, XU Q N. An RGB-D SLAM algorithm for robot based on the improved geometric and motion constraints in dynamic environment[J]. Robot, 2021, 43(2): 167-176.
- [23] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012: 573-580.

## 作者简介

史涛, 博士, 讲师, 主要研究方向为类脑智能机器人、机器人视觉和生物启发的智能计算。

E-mail: st99@email.tjut.edu.cn

校诺政(通信作者), 硕士研究生, 主要研究方向为智能机器人 SLAM。

E-mail: xnz2022@stud.tjut.edu.cn

丁垚, 硕士研究生, 主要研究方向为图像处理。

E-mail: 2439360904@qq.com