

# 视觉 SLAM 在动态场景下的图像处理方法

游通飞<sup>1,2</sup>, 孔令华<sup>1,2</sup>, 刘文玉<sup>1</sup>, 易定容<sup>3</sup>, 殷江<sup>1,2</sup>

(1. 福建工程学院 机械与汽车工程学院, 福建 福州 350118;

2. 福建工程学院 数字福建工业制造物联网实验室, 福建 福州 350118; 3. 华侨大学 机电及自动化学院, 福建 厦门 361021)

**摘要:** SLAM 一直是机器人领域的研究热点, 近年来取得了万众瞩目的进步, 但很少有 SLAM 算法考虑到动态场景的处理。针对视觉 SLAM 场景中动态目标的处理, 提出一种在动态场景下的图像处理方法。将基于深度学习的语义分割算法引入到 ORB\_SLAM2 方法中, 对输入图像进行分类处理的同时剔除人身上的特征点。基于已经剔除特征点的图像进行位姿估计。在 TUM 数据集上与 ORB\_SLAM2 进行对比, 在动态场景下的绝对轨迹误差和相对路径误差精度提高了 90% 以上。在保证地图精度的前提下, 改善了地图的适用性。

**关键词:** 视觉 SLAM; 动态场景; ORB-SLAM2; 特征点; 剔除

**中图分类号:** TP391      **文献标识码:** A      **文章编号:** 1001-8891(2021)10-0960-08

## Image Processing Method for Visual Simultaneous Localization and Mapping in Dynamic Scene

YOU Tongfei<sup>1,2</sup>, KONG Linghua<sup>1,2</sup>, LIU Wenyu<sup>1</sup>, YI Dingrong<sup>3</sup>, YIN Jiang<sup>1,2</sup>

(1. School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China;

2. Digital Fujian Industrial Manufacturing IoT Lab, Fuzhou 350118, China;

3. College of Mechanical Engineering and Automation, Huaqiao University, Xiamen 361021, China)

**Abstract:** Simultaneous localization and mapping(SLAM) has always been a research hotspot in the robotics field. In recent years, remarkable progress has been made in SLAM research, but few SLAM algorithms have considered the processing of dynamic scenes. Therefore, in this study, an image processing method for dynamic target processing in a visual SLAM scene is proposed. The semantic segmentation algorithm based on deep learning was introduced into the ORB\_SLAM2 method and input image classification processing was accomplished while removing the feature points on the body. Pose estimation was performed based on images with eliminated feature points. Compared to ORB\_SLAM2 on the TUM dataset, the absolute trajectory error and relative path error accuracy were improved by more than 90% in the dynamic scene. To ensure the accuracy of the generated map, the applicability of the map was improved.

**Key words:** visual SLAM, dynamic scene, ORB-SLAM2, feature points, remove

## 0 引言

SLAM (simultaneous localization and mapping), 中文名为同步定位与地图构建, 它从提出到发展已经有三十余年<sup>[1-2]</sup>。根据使用传感器的类型, 将 SLAM 分为激光 SLAM 和视觉 SLAM<sup>[3]</sup>。随着 SLAM 技术的进一步完善, 该方法被广泛应用到机器人技术上。视觉 SLAM 技术因其性价比高、效果佳, 成

为当下的研究热点。目前大多数视觉 SLAM 的研究工作主要是集中在静态环境中展开的, 真实场景下往往都是呈现动态形式<sup>[4]</sup>。这些静态场景下的算法处理起真实环境时, 结果往往不尽人意。为了解决视觉 SLAM 在动态场景下工作的鲁棒性, Kai Wang 等人<sup>[5]</sup>通过识别和处理移动和潜在可移动对象, 增强了视觉 SLAM 中的建图和定位精度。具体来说, 利用前一帧和当前一帧的位姿信息来细化后一帧的

收稿日期: 2020-06-06; 修订日期: 2020-07-28。

作者简介: 游通飞 (1994-), 男, 福建福州人, 硕士研究生, 研究方向: 视觉 SLAM 和语义分割。

通信作者: 孔令华 (1963-), 男, 加拿大人, 教授, 博士, 硕士生导师, 研究方向: 三维视觉和多光谱检测。E-mail: 15392030898@163.com。

基金项目: 国家自然科学基金资助项目 (51775200)。

分割, 确定所有潜在的可移动的对象, 发送到 vSLAM 模块进一步计算相应帧的跟踪和建图, 使分割精度得以提高。DynaSLAM<sup>[6]</sup>中, 作者提出了一种检测动态对象系统。使用 MASK-RCNN 对物体进行实例分割, 然后对具有移动性的对象进行分割。该工作还引入了背景区域填充的方法, 可以修复动态目标遮挡的区域。在 DS-SLAM<sup>[7]</sup>中作者通过检测物体的运动一致性, 判断人是否是动态对象, 并获得语义分割后的人的标签。降低了基于视觉的 SLAM 中动态对象的影响。针对视觉 SLAM 场景中动态目标的问题, 本文在 ORB-SLAM2 处理动态环境基础上, 引入语义分割算法<sup>[8-9]</sup>。根据人身的 ORB 特征点在帧间变化的特点, 判断人是否运动。将语义分割中人的类别作为动态物体进行特征点剔除, 最终达到所建地图不会因人的运动而受到干扰。

## 1 基于 ORB-SLAM2 的图像处理方法

### 1.1 传统的 ORB-SLAM2

ORB-SLAM2<sup>[10]</sup>算法是由跟踪线程、局部建图线程、闭环检测线程这 3 个平行线程和全局 BA (bundle adjustment) 优化线程组成, 如图 1 所示。跟踪线程是将相机收集到的每一帧数据进行 ORB 特征提取, 通过相机位姿跟踪和局部地图跟踪获得局部地图与当前帧的匹配。利用最小化投影误差优化位姿, 判断是否生成关键帧。局部建图线程是对

上一线程所得新关键帧进行处理, 剔除地图中质量偏低的点。通过三角化的方法来恢复出共视程度比较高的关键帧地图点, 检查当前关键帧与相邻帧重复的地图点, 进行地图点融合。当前帧相连的关键帧及地图点做局部 BA 优化, 剔除多余的关键帧。闭环探测和闭环校正组成闭环检测线程。其中闭环探测是通过计算当前帧与相连关键帧的词袋 (bag of words, BoW) 模型最低得分, 判断当前关键帧是否产生闭环并进行连续性检测并计算相似变换。闭环校正主要包含闭环融合和图优化这两个部分。全局 BA 优化线程是对全局的位姿和地图点进行优化, 达到更新地图的目的。

### 1.2 PSPNet 算法

PSPNet (pyramid scene parsing network)<sup>[11]</sup>算法利用基于不同区域的上下文池化来实现全局上下文信息的能力。PSPNet 算法结构如图 2 所示, 指定一个输入图像(a), 首先经过卷积神经网络得到特征图 feature map(b), 特征图分别进入金字塔解析模块和连接层。一部分特征图经过金字塔解析模块处理后的特征图进行向上采样, 向上采样的结果传入到连接层与另一部分特征图进行结合, 获得特征表示, 该特征包含局部和全局上下文信息(c)。金字塔池化模块处理后得到的结果输入卷积层, 语义分割的预测(d)便可呈现。

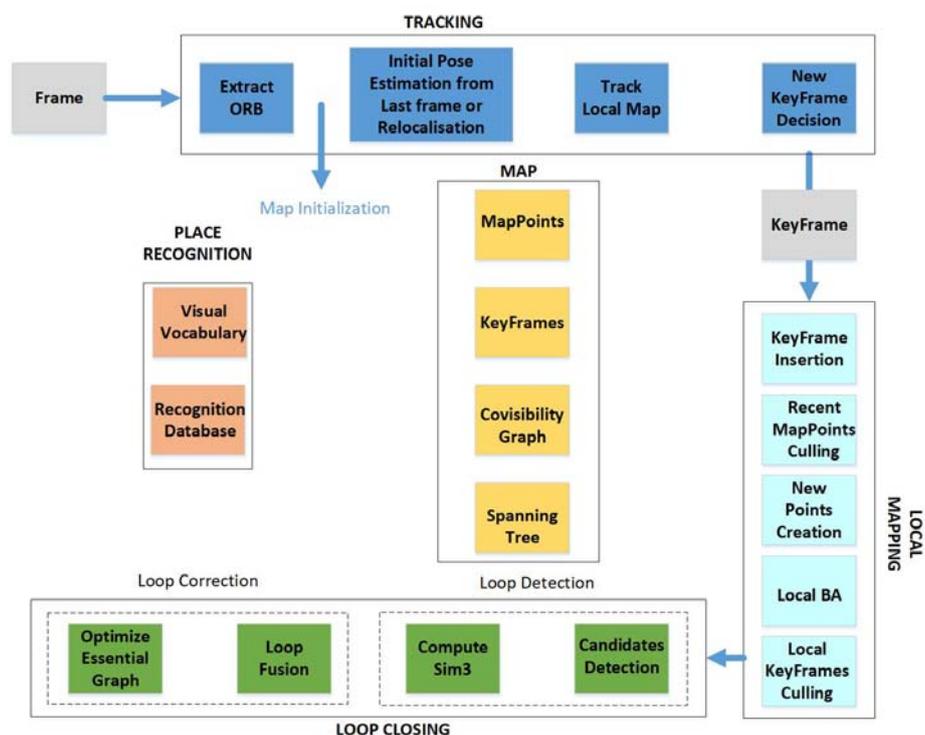


图 1 ORB\_SLAM2 系统线程和结构

Fig.1 ORB\_SLAM2 system threads and structure

金字塔池模块有4种不同的尺度的特征。其中最上层代表生成单个 bin 输出的全局池。下方的其余金字塔将特征图分成不同的子区域，形成不同位置的表示。大小各异特征图是由金字塔池化模块中各层次输出得到，若金字塔的大小为  $N$ ，为使全局特征不变，每一个金字塔层将使用  $1 \times 1$  卷积层去减小上下文的维度到原来的  $1/N$ 。低维度的特征图上采样得到与原始特征图大小相同的特征。各层次的特征进行连接即为金字塔池化全局特征。

### 1.3 ORB-SLAM2 动态场景下的优化

为解决 ORB-SLAM2 算法在动态场景下的漂移量大、目标丢失等问题，本文提出将 ORB-SLAM2 算法和语义分割算法进行结合。目前，基于深度学习的语义分割算法中，PSPNet 算法表现出优秀的分割性能，是近些年语义分割中具有高精度算法的代表，故选择 PSPNet 算法作为本文的语义分割算法。系统流程图如图 3 所示。其中，PSPNet 是在数据集 PASCAL VOC2012 进行训练，该数据集是目前衡量语义分割的标准数据集之一<sup>[12]</sup>。

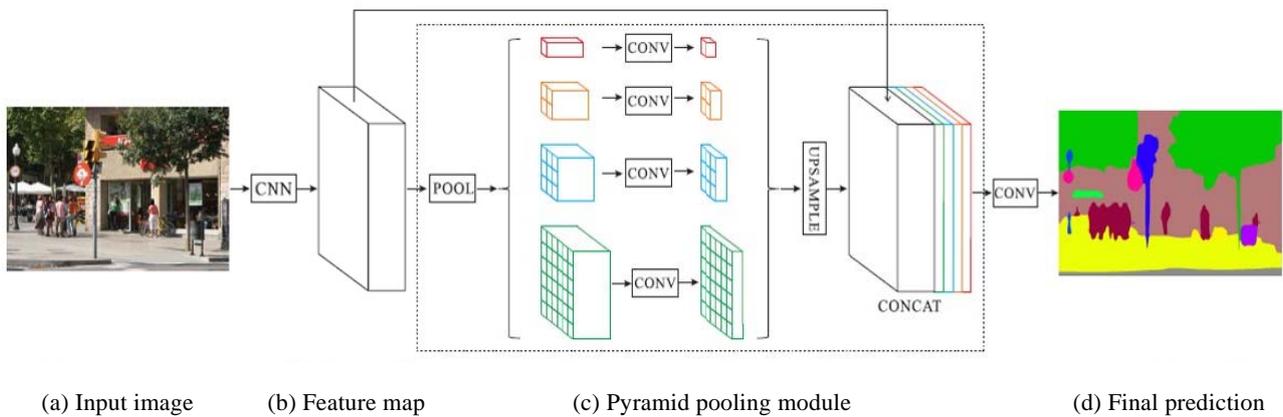


图 2 PSPNet 算法结构

Fig.2 PSPNet algorithm structure

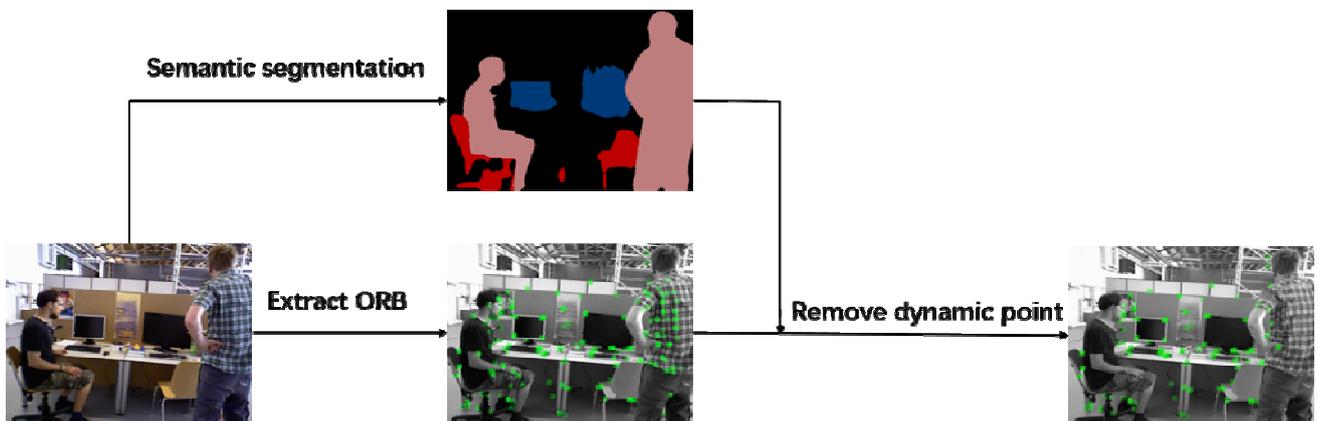


图 3 剔除动态点方法

Fig.3 Remove dynamic point method

本文的方法是将原始图像分别输入至 ORB-SLAM2 算法和 PSPNet 算法中。在 ORB-SLAM2 跟踪线程中进行提取特征点时，经过 PSPNet 网络分割过的图像传入 ORB-SLAM2 中。此时，分割图像中的人作为动态对象，ORB-SLAM2 将剔除分布在人身上的特征点。剔除人身上特征点后的其余特征点继续进行特征匹配和位姿估计。

把人作为动态对象的特征点剔除后，ORB-SLAM2 算法在动态场景下工作的能力得到了明显提升，展示出较好的抗干扰能力。

## 2 实验与分析

本文使用公共数据集 TUM RGB-D<sup>[13]</sup>对 SLAM 系统稳定性进行评估。该数据集是由德国慕尼黑工业大学提出的一个标准 RGB-D 数据集，目前经常

用来评估 RBG-D 的一个数据集。该数据集是用来评估 SLAM 系统或 VO 算法在具有移动物体的动态环境下性能。数据集中的图像是使用 Kinect 传感器进行获取,包含彩色图像、深度图像和地面真实轨迹的数据,其中彩色图像和深度图像时间戳一一对应。数据集自带了两种评估 SLAM 系统的程序,分别是相对位姿误差 (relative pose error, RPE) 和绝对轨迹误差 (absolute trajectory error, ATE)。

在一定的时间间隔内衡量轨迹的局部精度被称为相对位姿误差 (RPE),用于评价系统的漂移量。

$$E_i := (\mathbf{Q}_i^{-1} \mathbf{Q}_{i+\Delta})^{-1} (\mathbf{P}_i^{-1} \mathbf{P}_{i+\Delta}) \quad (1)$$

式中:  $\mathbf{P} \in \text{SE}(3)$  是估计位姿;  $\mathbf{Q} \in \text{SE}(3)$  是真实位姿;  $i$  代表时间或者帧数;  $\Delta$  表示时间间隔。

通过比较估计和地面真实轨迹 (ground truth) 之间的绝对距离被称为绝对轨迹误差,用于评价轨迹的全局一致性。

$$F_i := \mathbf{Q}_i^{-1} \mathbf{S} \mathbf{P}_i \quad (2)$$

其中  $\mathbf{S} \in \text{SE}$  是估计位姿到真实位姿的转换矩阵。

本文是在 TUM 数据集的 `rgbd_dataset_freiburg3_walking_xyz`、`rgbd_dataset_freiburg3_walking_halfsphere`、`rgbd_dataset_freiburg3_walking_static`、`rgbd_dataset_freiburg3_sitting_static` 四个序列工作性能的对比。表 1、2、3 是 ORB-SLAM2 和本文方法的对比结果。`walking_xyz`、`walking_halfsphere`、`walking_static`、`sitting_static` 绝对轨迹误差图和相对位姿误差图分别如图 4、5、6、7、8、9、10、11 所示。其中,左边图为绝对轨迹误差,右边图为相对位姿误差。

表 1 绝对轨迹误差对比 (ATE)

Table 1 Absolute trajectory error comparison

Sequences	ORB-SLAM2				Ours				Improvement/%			
	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std
walking_xyz	0.5357	0.4964	0.4733	0.2014	0.0269	0.0185	0.0151	0.0196	94.98	96.27	96.81	90.27
walking_halfsphere	0.4318	0.3651	0.3107	0.2305	0.0334	0.0285	0.0243	0.0175	92.26	92.19	92.18	92.41
walking_static	0.3753	0.3398	0.2963	0.1593	0.0076	0.0068	0.0062	0.0034	97.97	98.00	97.91	97.87
sitting_static	0.0082	0.0071	0.0063	0.0041	0.0062	0.0054	0.0047	0.0031	24.39	23.94	25.40	24.39

表 2 相对位姿误差对比 (RPE)

Table 2 Relative pose error comparison

Sequences	ORB-SLAM2				Ours				Improvement/%			
	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std
walking_xyz	0.7856	0.6444	0.5714	0.4493	0.0400	0.0280	0.0222	0.0285	94.91	95.65	96.11	93.66
walking_halfsphere	0.6200	0.4957	0.4705	0.3724	0.0474	0.0414	0.0373	0.0231	92.35	91.65	92.07	93.80
walking_static	0.5354	0.3946	0.1802	0.3618	0.0112	0.0102	0.0096	0.0048	97.91	97.42	94.67	98.67
sitting_static	0.0127	0.0112	0.0101	0.0060	0.0093	0.0081	0.0073	0.0044	26.77	27.68	27.72	26.67

表 3 相对旋转误差对比 (RRE)

Table 3 Relative rotation error comparison

Sequences	ORB-SLAM2				Ours				Improvement/%			
	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std	Rmse	Mean	Median	Std
walking_xyz	14.3695	11.7969	0.1856	8.2046	0.8776	0.6192	0.0082	0.6220	93.89	94.75	95.58	92.42
walking_halfsphere	14.5176	12.0261	0.2177	8.1323	1.0316	0.8956	0.0139	0.5118	92.89	92.55	93.62	93.71
walking_static	9.6864	7.1088	0.0558	6.5796	0.3021	0.2724	0.0044	0.1306	96.88	96.17	92.11	98.02
sitting_static	0.3572	0.3220	0.0054	0.1546	0.3347	0.2981	0.0048	0.1523	6.30	7.42	11.11	1.49

从表1~3可以观察到,在前3个动态环境下,无论相对位姿误差还是绝对轨迹误差,改进后的ORB-SLAM2系统的Rmse、Mean、Median和Std值比传统ORB-SLAM2都提高了90%以上。而在第

4个静态环境下,最高只比ORB-SLAM2提高了27.68%。实验数据表明:改进后的ORB-SLAM2系统能较好地克服动态环境下系统的漂移并达到轨迹全局一致性的效果。

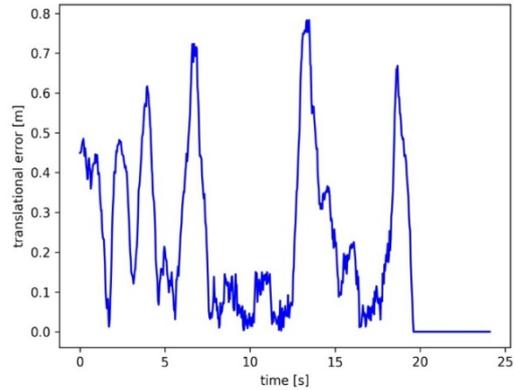
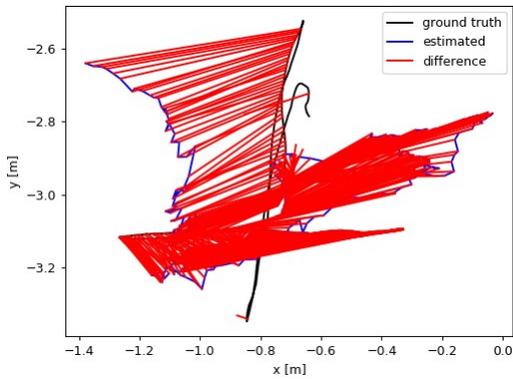


图4 ORB-SLAM2在walking\_xyz下的误差(左:轨迹误差;右:相对位姿误差)

Fig.4 Error of ORB-SLAM2 under walking\_xyz(left: absolute trajectory error; right: Relative pose error)

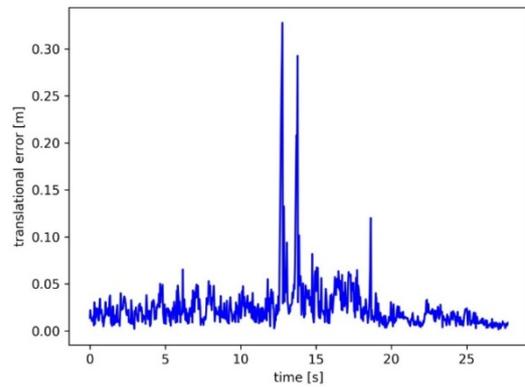
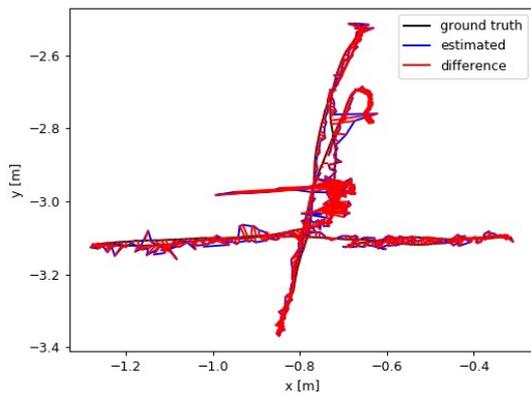


图5 本文方法在walking\_xyz下的误差(左:轨迹误差;右:相对位姿误差)

Fig.5 Error of our method under walking\_xyz(left: absolute trajectory error; right: relative pose error)

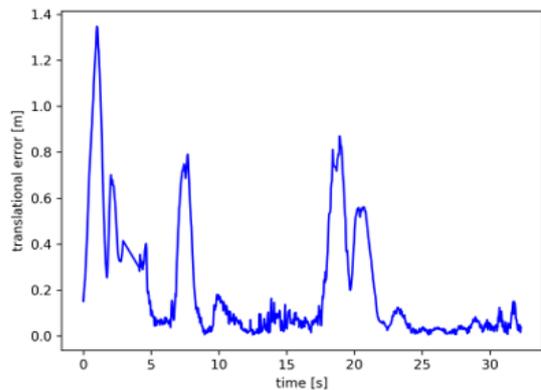
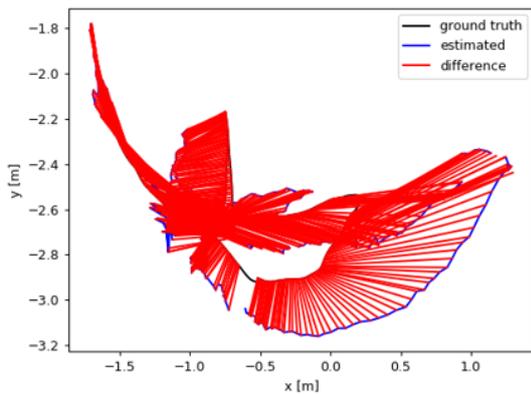


图6 ORB-SLAM2在walking\_halfsphere下的误差(左:轨迹误差;右:相对位姿误差)

Fig.6 Error of ORB-SLAM2 under walking\_halfsphere(left: absolute trajectory error; right: relative pose error)

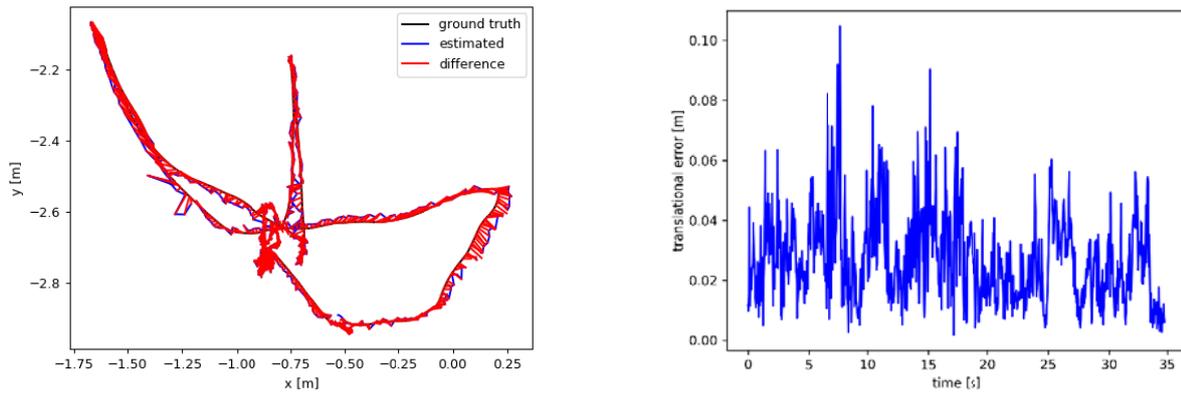


图7 本文方法在 walking\_halfsphere 下的误差(左: 轨迹误差; 右: 相对位姿误差)

Fig.7 Error of our method under walking\_halfsphere(left: absolute trajectory error; right: relative pose error)

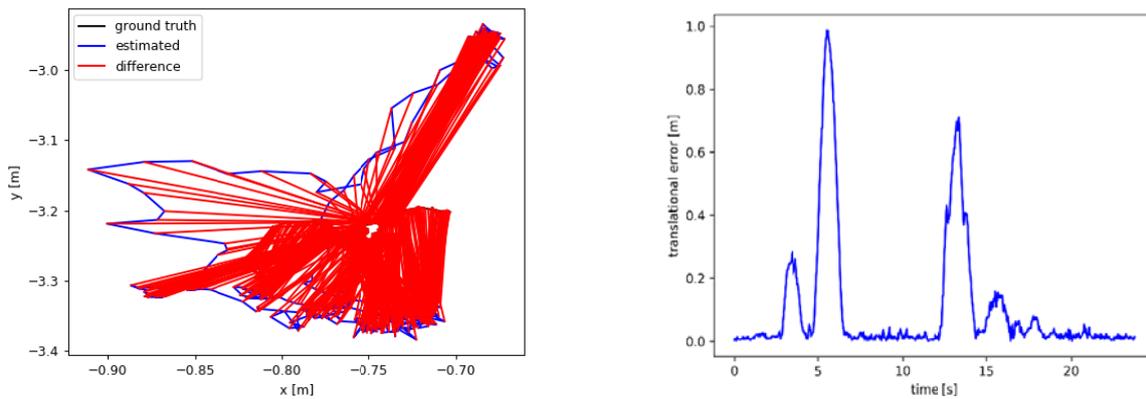


图8 ORB-SLAM2在 walking\_static 下的误差(左: 轨迹误差; 右: 相对位姿误差)

Fig.8 Error of ORB-SLAM2 under walking\_static(left: absolute trajectory error; right: relative pose error)

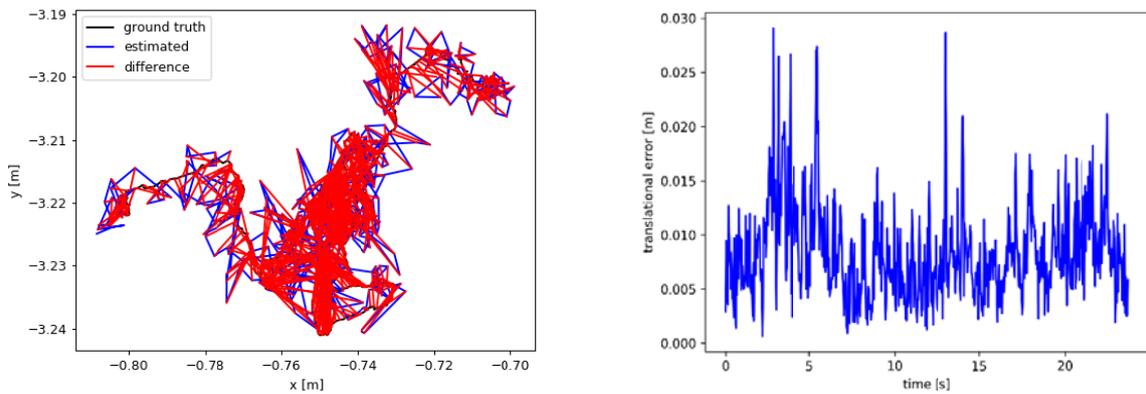


图9 本文方法在 walking\_static 下的误差(左: 轨迹误差; 右: 相对位姿误差)

Fig.9 Error of our method under walking\_static(left: absolute trajectory error; right: relative pose error)

在耗时上, 将 ORB-SLAM2 和改进后的 ORB-SLAM2 分别在 TUM 数据集上运行并取 3 次结果求平均值, 如表 4 所示。ORB-SLAM2 平均耗时 57.439 ms, 改进后的 ORB-SLAM2 平均耗时

79.681 ms。对比改进前后的 ORB-SLAM2 耗时, 改进后的耗时比传统 ORB-SLAM2 多用了 22.242 ms, 这是因为改进后的 ORB-SLAM2 引入了语义分割, 增加了系统的耗时。

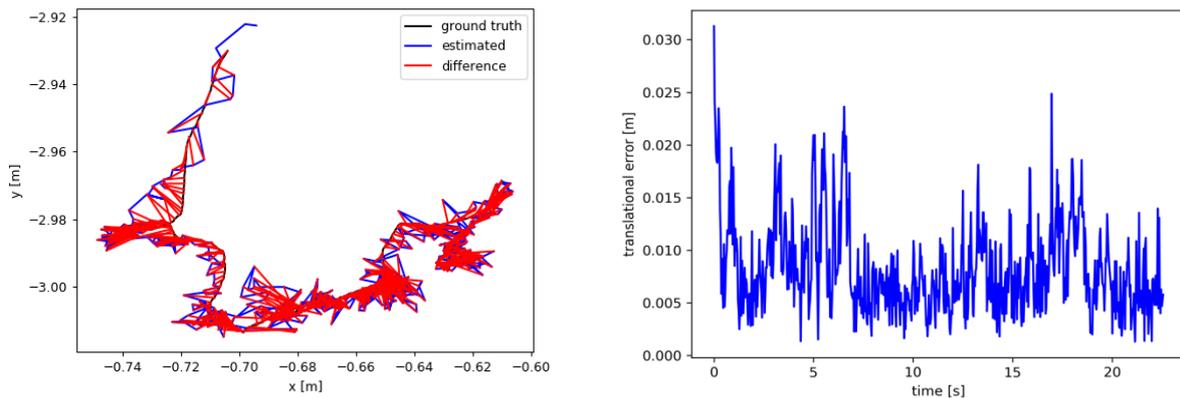


图 10 ORB-SLAM2 在 sitting\_static 下的误差 (左: 轨迹误差; 右: 相对位姿误差)

Fig.10 Error of ORB-SLAM2 under sitting\_static(left: absolute trajectory error; right: relative pose error)

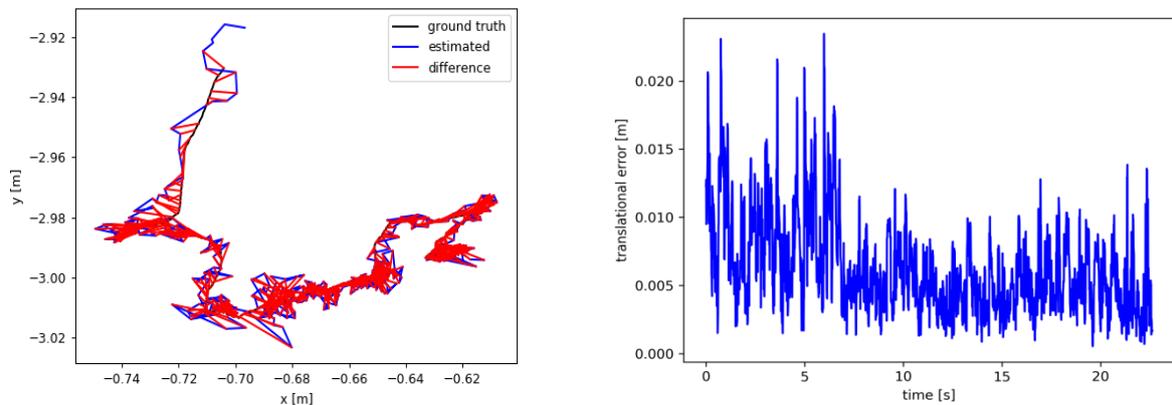


图 11 本文方法在 sitting\_static 下的误差 (左: 轨迹误差; 右: 相对位姿误差)

Fig.11 Error of our method under sitting\_static(left: absolute trajectory error; right: relative pose error)

表 4 两种方法在 TUM 数据集的耗时

Table 4 Time consuming of the two methods in TUM dataset

Methods	1	2	3	Average
ORB-SLAM2	54.314	58.629	59.373	57.439
Ours	81.241	79.298	78.505	79.681

### 3 结论

本文基于 SLAM 系统中 ORB-SLAM2 算法, 针对动态场景下的目标, 提出了 ORB-SLAM2 算法和语义分割算法进行融合的方法。在场景中把人看作是动态对象, 剔除人身上的特征点。在 TUM 数据集上进行实验验证, 结果表明: 本文提出的方法在动态场景下的绝对轨迹误差和相对路径误差比原 ORB-SLAM2 算法至少提高了 90% 以上的精度, 同时系统耗时增加 22.242 ms, 该方法在动态环境中能够保持系统的鲁棒性。

#### 参考文献:

[1] Durrant-Whyte H, Bailey Tim. Simultaneous localization and mapping: part I[J]. *IEEE Robotics & Automation Magazine*, 2006, 13(2): 99-10.

[2] 张金凤, 石朝侠, 王燕清. 动态场景下基于视觉特征的 SLAM 方法[J]. *计算机工程*, 2020, 46(10): 95-102.

ZHANG Jinfeng, SHI Chaoxia, WANG Yanqing. SLAM method based on visual features in dynamic scene[J]. *Computer Engineering*, 2020, 46(10): 95-102.

[3] 江国来. 共融移动服务机器人导航与交互关键技术研究[D]. 深圳: 中国科学院大学(中国科学院深圳先进技术研究院), 2019.

JIANG Guolai. Research on Key Technologies of Navigation and Interaction of Inclusive Mobile Service Robot[D]. Shenzhen: University of Chinese Academy of Sciences, 2019.

[4] HAN Shuangquan, XI Zhihong. Dynamic scene semantics SLAM based on semantic segmentation[J]. *IEEE Access*, 2020, 8: 43563-43570.

[5] WANG Kai, LIN Yimin, WANG Luowei, et al. A unified framework for mutual improvement of SLAM and semantic segmen-

- tation[C]//*International Conference on Robotics and Automation (ICRA) of IEEE*, 2019: 5224-5230.
- [6] Bescos Berta, Fácil José M, Civera Javier, et al. Dyna SLAM: tracking, mapping and in painting in dynamic scenes[J]. *IEEE Robotics and Automation Letters*, 2018, **3**(4): 4076-4083.
- [7] YU Chao, LIU Zuxin, LIU Xinjun, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//*IEEE International Conference on Intelligent Robots and Systems(IROS)*, 2018: 1168-1174.
- [8] 王召东, 郭晨. 一种动态场景下语义分割优化的 ORB\_SLAM2[J]. *大连海事大学学报*, 2018, **44**(4): 121-126.
- WANG Zhaodong, GUO Chen. An improved ORB\_SLAM2 in dynamic scene with semantic segmentation[J]. *Journal of Dalian Maritime University*, 2018, **44**(4): 121-126.
- [9] 王晨, 汤心溢, 高思莉. 基于深度卷积神经网络的红外场景理解算法[J]. *红外技术*, 2017, **39**(8): 728-733.
- WANG Chen, TANG Xinyi, GAO Sili. Infrared scene understanding algorithm based on deep convolutional neural network[J]. *Infrared Technology*, 2017, **39**(8): 728-733.
- [10] Mur-Artal Raul, D Tardos Juan. ORB-SLAM2: an opensource slam system for monocular, stereo, and rgb-d cameras[J]. *IEEE Transactions on Robotics*, 2017, **33**(5): 1255-1262.
- [11] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, et al. Pyramid scene parsing network[C]//*IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6230-6239.
- [12] 李瀚超, 蔡毅, 王岭雪. 全局特征提取的全卷积网络图像语义分割算法[J]. *红外技术*, 2019, **41**(7): 595-599, 615.
- LI Hanchao, CAI Yi, WANG Lingxue. Image semantic segmentation based on fully convoluted network with global feature extraction[J]. *Infrared Technology*, 2019, **41**(7): 595-599.
- [13] Sturm Jürgen, Engelhard Nikolas, Endres Felix, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]//*RSJ Int. Conf. Intell. Robots Syst. of IEEE*, 2012: 573-580.
- [14] Usamentiaga R, Venegas P, Guerediaga J, et al. Infrared thermography for temperature measurement and non-destructive testing[J]. *Sensors*, 2014, **14**(7): 12305-12348.
- [15] 武汉凯尔文光电技术有限公司产品中心. JQ-D70Z 热成像人体测温黑体[EB/OL]. [2021-08-19]. [https://http://www.whkelvin.com/pros\\_show.php?id=84](https://http://www.whkelvin.com/pros_show.php?id=84).
- Product Center of Wuhan Kelvin Photoelectric Technology Co. Ltd. JQ- D70Z blackbody of screening thermograph[EB/OL]. [2021-08-19]. [https://http://www.whkelvin.com/pros\\_show.php?id=84](https://http://www.whkelvin.com/pros_show.php?id=84).
- [16] Pascoe D, Ring E, Mercer J, et al. International standards for pandemic screening using infrared thermography[C]//*Proc. of SPIE, Medical Imaging 2010: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2010, **7626**: 76261Z.
- (上接第 924 页)
- [12] 吴继平, 张桂玲, 杨楚明. 工业检测型红外热像仪: GB/T19870—2018[S]. 全国工业过程测量控制和自动化标准化技术委员会 [2018-05-14].
- WU Jiping, ZHANG Guiling, YANG Chuming. Industrial inspecting thermal imagers: GB/T19870—2018[S]. National Industrial Process Measurement Control and Automation Standardization Technical Committee [2018-05-14].
- [13] 白廷柱, 金伟其. 光电成像原理与技术[M]. 北京: 北京理工大学出版社, 2006.
- BAI Tingzhu, JIN Weiqi. *Principle and Technology of Photoelectric Imaging*[M]. Beijing: Beijing University of Technology Press, 2006.