

一种面向动态环境下视觉同时定位和建图的图像预处理方法

卓桂荣^{1,2}, 卢守义^{1,2}, 熊璐^{1,2}

(1. 同济大学汽车学院, 上海 201804; 2. 同济大学新能源汽车工程中心, 上海 201804)

摘要: 提出了一种用于动态环境下视觉同时定位和建图 (SLAM) 系统的图像预处理方法。该方法可以很容易地集成到现有视觉 SLAM 系统中, 使其在高动态环境下能够稳定、准确和连续的工作。首先, 提出了一种综合使用语义分割网络和光流估计网络的动态物体识别算法, 鲁棒、准确地识别图像中潜在的动态物体。然后, 为了检测与动态物体关联的阴影, 提出了一种基于区域生长的阴影识别算法。最后, 使用图像补全技术对剔除动态物体后的图像进行补全。将该图像预处理方法与双目 ORB-SLAM2 结合, 并在 KITTI 数据集上进行了实验, 实验表明所提出的图像预处理方法显著地提升了视觉 SLAM 系统的定位精度, 并且图像预处理方法中的每一个模块都有着不可替代的作用。

关键词: 图像预处理方法; 视觉同时定位和建图; 动态环境; 定位

中图分类号: U469.79

文献标志码: A

potential dynamic objects in images. Then, to detect shadows associated with dynamic objects, it proposes a shadow recognition algorithm based on region growing. Afterwards, it uses image completion techniques to fill in the gaps left by the removal of dynamic objects from the images. Finally, it combines this image preprocessing method with stereo ORB-SLAM2 and conducts experiments on the KITTI dataset, which demonstrates that the proposed image preprocessing method significantly improves the positioning accuracy of visual SLAM systems. Each module in the image preprocessing method plays an irreplaceable role.

Key words: image pre-processing methods; visual simultaneous localization and mapping; dynamic environment; localization

An Image Pre-Processing Method for Visual Simultaneous Localization and Mapping in Dynamic Environments

ZHUO Guirong^{1,2}, LU Shouyi^{1,2}, XIONG Lu^{1,2}

(1. School of Automotive Studies, Tongji University, Shanghai 201804, China; 2. Clean Energy Automotive Engineering Center, Tongji University, Shanghai 201804, China)

Abstract: This paper proposes an image preprocessing method for visual simultaneous localization and mapping (SLAM) systems in dynamic environments, which can be easily integrated into existing visual SLAM systems to enable stable, accurate, and continuous operation in highly dynamic environments. First, it proposes a dynamic object recognition algorithm that integrates the use of semantic segmentation networks and optical flow estimation networks to robustly and accurately identify

车辆的自主定位技术是自动驾驶技术中的重要组成部分^[1]。传统的基于全球定位系统 (GPS) 的定位方法是智能车辆中最常用的定位方法。但 GPS 在城市环境中易受到遮挡, 并存在多径效应, 会导致定位精度下降, 甚至错误^[2]。为解决上述问题, GPS 经常与惯性测量单元 (IMU) 融合为智能车辆提供更加稳定和精确的定位方式^[3]。但在这种融合方式下, IMU 在匀速运动时会出现加速度计失效的情况, 只能依靠 GPS 对车辆进行定位。因此, 基于 GPS 的定位方法应用在智能车辆定位中并不鲁棒。

同时定位和建图算法 (simultaneous localization and mapping, SLAM) 可以很好地解决基于 GPS 的定位方法失效时智能车辆的精准定位问题^[4]。根据使用传感器的不同, SLAM 算法可以分为视觉 SLAM 和激光 SLAM 两大类, 前者因低成本、可以

收稿日期: 2023-03-05

基金项目: 国家自然科学基金 (52325212), 国家重点研发计划 (2022YFE0117100), 中央高校基本科研业务费专项资助

第一作者: 卓桂荣, 副教授, 工学博士, 主要研究方向为车辆动力学控制、智能车辆组合定位。

E-mail: zhuoguirong@tongji.edu.cn



论文
拓展
介绍

提供丰富的场景信息等优点得到高度的关注和广泛的研究。传统的视觉SLAM可以分为两类:基于特征的方法和直接方法^[5]。基于特征的方法利用在连续图像之间匹配的显著图像特征,通过最小化特征对应的重投影误差的方式获得相机运动。MonoSLAM^[6]、PTAM^[7]、ORB-SLAM2^[8]以及ORB-SLAM3^[9]都是经典的基于特征方法的视觉SLAM系统。与基于特征的方法相比,直接方法通过光度误差直接恢复相机的位姿,而不进行特征提取。DTAM^[10]、LSD-SLAM^[11]、DSO^[12]以及LDSO^[13]都是经典的基于直接方法的视觉SLAM系统。

尽管研究界已经从不同的角度开发了各种视觉SLAM方法。然而,以上方法都假设相机所运行的是一个静态环境。但现实场景中会不可避免地出现行人、汽车等动态物体。动态物体会导致许多错误或不稳定的数据关联,降低SLAM系统的精度。为此,众多学者对动态场景下的SLAM问题做了大量研究。现有研究成果可分为两类:基于几何的方法和基于语义的方法^[14]。基于几何的方法是使用随机采样一致性(RANSAC)^[15]、鲁棒核函数^[16]以及多视图几何^[17-18]等异常值剔除算法来剔除图像中检测到的动态特征点。但该方法存在无法检测到暂时保持静止的潜在动态物体以及缺乏语义信息等局限性。一些学者也提出了基于深度学习模型的动态SLAM系统。刘钰嵩等^[19]使用光流算法获得的稠密光流场提取动态物体掩码,之后剔除掩码内的动态物体点,来提高视觉SLAM系统的精度。张小勇等^[20]使用Lucas-Kanade稀疏光流和Mask RCNN图像分割网络相结合的方法感知图像中的动态物体。DS-SLAM^[21]中使用SegNet图像分割网络和运动一致性检测相结合的方法剔除动态特征点。Dyna-SLAM^[22]中使用Mask RCNN图像分割网络和多视图几何相结合的方法检测图像中的动态物体并剔除动态特征点。Detect-SLAM^[23]中使用SSD目标检测算法检测图像中的动态对象并根据检测结果设定特征点的移动概率来提高视觉SLAM系统对动态环境的鲁棒性。DM-SLAM^[24]使用DeepLabV3+语义分割算法和多视图几何相结合的方法对图像中的动态特征点进行剔除。综上所述,基于语义的动态SLAM方法大多使用光流估计、语义分割或目标检测等方法获得图像中动态物体的像素级掩码或边界框,之后使用语义信息剔除图像中的动态物体。但光流估计算法无法识别图像中小物体。此外,使用

语义分割来识别图像中的动态物体且十分依赖语义分割网络的精度,当语义分割精度低时,对图像中的动态物体识别不完整,并且剔除动态物体后场景留下的空洞依然会对相机的定位精度产生不小的影响。

针对以上问题,在本文中,提出了一个用于动态环境下视觉SLAM系统的图像预处理方法。该方法使用语义分割网络和光流估计网络鲁棒、精确地检测图像中潜在的动态物体。针对语义分割网络无法检测动态物体阴影的问题,提出了一种基于区域生长的阴影识别算法,对动态物体的阴影进行识别。之后,对图像中剔除动态物体及其阴影的空洞进行补全,获取被动态物体遮挡的特征信息。通过这种方式,不但剔除了图像中的动态物体及其阴影,而且对剔除后图像中的空洞部分进行了补全,增强了图像中的特征信息,以实现更稳健和准确的特征检测和匹配。此外,所提出的图像预处理方法可以很容易地集成到现有视觉SLAM中,以提高其在动态环境下的定位性能。

本文完成的工作总结如下:①提出了一种综合使用语义分割网络和光流估计网络的动态物体识别算法,其具有良好的精度和鲁棒性。②提出了一种基于区域生长的阴影识别算法,其可以准确地检测出图像中动态物体的阴影。实验表明,它是获得高质量空洞补全图像的必备模块并与视觉SLAM的定位精度密切相关。③提出了一个用于动态环境下视觉SLAM系统的图像预处理方法,其可以很容易地集成到现有视觉SLAM中,以提高其在动态环境下的定位精度。

1 提出的方法

图1展示了所提出的图像预处理方法结合视觉SLAM系统的框架。所提出的图像预处理方法由三个模块组成,分别为:动态物体识别模块,阴影检测模块以及图像补全模块,这些模块相互作用以解决动态SLAM问题。

动态物体识别模块的输入是彩色图像序列,模块中设有光流估计网络和语义分割网络,分别对输入的图像进行光流运动估计和语义分割来得到图像对应的光流图和语义分割图。结合图像的光流图和语义分割图,可以得到图像中的动态物体。阴影检测模块首先在原有的语义分割图中将图像中的动态物体进行掩码处理,得到动态物体掩码图。然后根

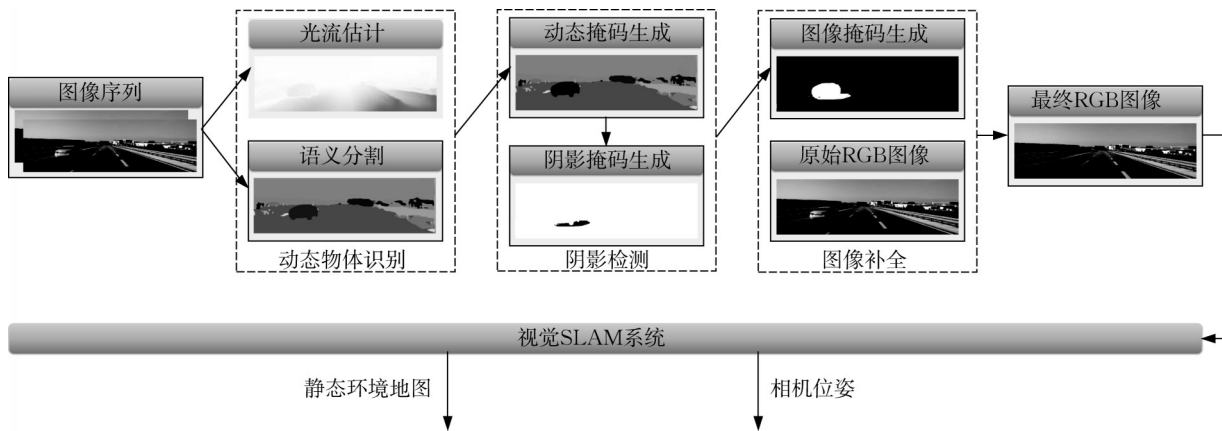


图1 所提出的图像预处理方法结合视觉SLAM系统框架图

Fig. 1 Framework of proposed image pre-processing method combined with the visual SLAM system

据动态物体掩码图将图中动态物体的阴影识别出来并作掩码处理得到阴影掩码图。阴影检测模块使用动态物体掩码图和阴影掩码图进行组合得到图像掩码图。将图像掩码图和相对应的原始RGB(红绿蓝)图像输入到图像补全网络中可得到最终去除动态物体及其阴影的补全图像。将该图像输入到视觉SLAM系统中可得到静态环境地图以及无动态物体干扰的相机位姿估计值。各个模块的具体内容在随后的小节中进行了简要描述。

1.1 动态物体识别

为了检测图像中的动态物体,使用了语义分割和光流估计相结合的方法,并证实了这两种方法在检测动态物体上具有互补效果。

使用短期密集连接(short-term dense concatenate, STDC)网络^[25]作为本研究的语义分割网络,这是语义分割的最新技术。在Cityscapes数据集^[26]上对网络进行了预训练,之后在KITTI数据集^[27]上对训练好的网络进行了细化。STDC网络的输入是一个大小为 $m \times n \times 3$ 的RGB图像,经过该网络后,输出是一个大小为 $m \times n \times l$ 的矩阵,其中 l 是图像中类别的数量。合并所有通道后,可以获得图像中潜在的动态物体(人、自行车、汽车、公共汽车、货车等)的分割。分割结果如图2b所示,从图中可以看出,语义分割算法对图像中远处的动态物体分割效果较好,但由于车辆显示不全的原因对图像中近处的动态物体分割不完整。为了解决此问题,采取了语义分割和光流估计相结合的方法。

使用MaskFlowNet^[28]作为本研究的光流估计网络。在KITTI 2015^[27]和Sintel^[29]数据集上对网络进行了训练。MaskFlowNet将 $t-1$ 时刻和 t 时刻的图像作为输入,输出 t 时刻图像的密集光流,如图2c所

示。动态物体上像素的光流值与静态物体或图像背景上像素的光流值有较大的差异,因此,通过求取密集光流图中每一个的像素光流梯度,来获得动态物体的轮廓,光流梯度的计算方法如下:

$$\begin{cases} g_{-v_{(i,j)}} = f_{(i+1,j)} - f_{(i,j)} \\ g_{-u_{(i,j)}} = f_{(i,j+1)} - f_{(i,j)} \\ g_{(i,j)} = \sqrt{g_{-v_{(i,j)}}^2 + g_{-u_{(i,j)}}^2} \end{cases} \quad (1)$$

式中: $f_{(i,j)}$ 表示在像素位置 (i,j) 处的光流值; $g_{-v_{(i,j)}}$ 表示在像素位置 (i,j) 处的纵向光流梯度值; $g_{-u_{(i,j)}}$ 表示在像素位置 (i,j) 处的横向光流梯度值; $g_{(i,j)}$ 表示在像素位置 (i,j) 处的光流梯度值。根据求出来的光流梯度值,设置合适的阈值即可得到动态物体的轮廓,如图2d所示。从图中可以看出,基于光流梯度的方法可以有效地识别出图像中近处的动态物体,但无法识别图像中远处的动态物体,因此可以和基于语义分割的方法互相弥补。根据识别出的动态物体轮廓,向图像边界填充即可得到图像中近处完整的动态物体,结合语义分割结果,即可得到输入图像的动态物体掩码图,如图2e所示。

1.2 阴影检测

在室外场景行驶时,智能车辆经常处在复杂的照明环境下,这会生成与车辆粘合的阴影。1.1节所述的动态物体识别方法,只是将图像中的动态物体识别出来,无法对与动态物体关联的阴影进行有效的检测。在动态物体移除中,处理动态物体产生的阴影是十分必要的。若只移除图像中的动态物体则会产生糟糕的图像修复结果,因为未移除的阴影不仅保留在修复后的图像中产生移动鬼影,而且还严重误导了空洞的上下文修复,因为阴影往往被选为最佳匹配像素。

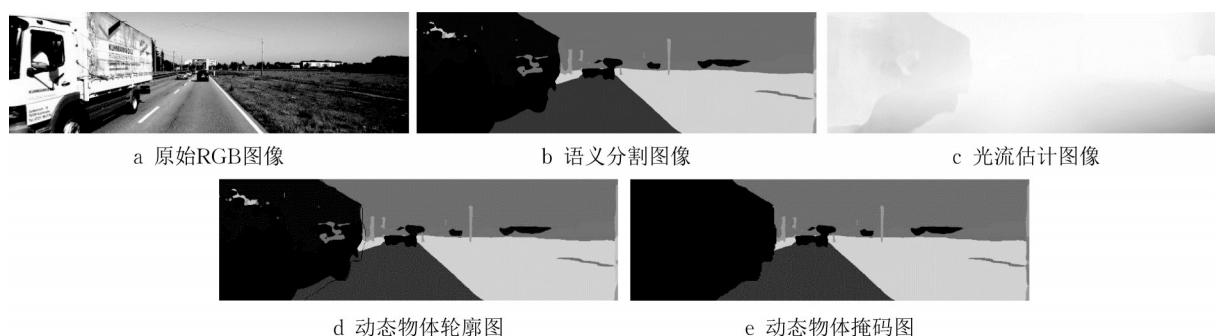


图2 动态物体识别算法

Fig. 2 Dynamic object detection method

阴影问题通常有两种解决方案。第一种解决方案是扩大动态物体掩码范围。然而,通过实验发现掩码范围的增加会显著地降低图像修复结果,因为越接近原始掩码,指导掩码进行修复的信息越多,修复的效果越好。第二种解决方案是使用深度学习的方法。然而基于深度学习的方法泛化能力较差,在使用过程中也不鲁棒。为了能有效地检测出与动态物体关联的阴影,提出了一种基于区域生长的阴影

识别算法。算法使用原始RGB图像以及带有动态物体掩码的语义分割图作为输入,根据设定的初始种子选取规则,选取初始阴影种子点,把满足区域生长条件的阴影种子点添加到增长范围中,然后把把这些新像素当作新的阴影种子点,重复上面的过程,直到没有满足区域生长条件的像素时则结束增长。最终输出图像中的阴影掩码图。算法流程如图3所示。图中,HSL表示色相、饱和度、明度。

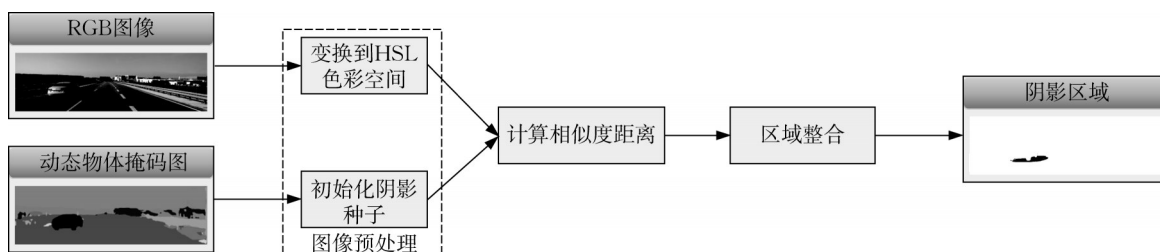


图3 基于区域生长的阴影识别算法流程

Fig. 3 Process of region growing based shadow recognition algorithm

1.2.1 图像预处理

预处理包括将RGB图像转换为HSL色彩空间以及根据带有动态物体掩码的语义分割图选取初始阴影种子两部分。将RGB图像转换到HSL色彩空间,可以根据色相、饱和度、明度(hue saturation lightness, HSL)色彩空间中的亮度值对图像中的阴影进行判别。图3中的RGB图像在HSL色彩空间中亮度值分布直方图如图4所示。图像中阴影部分的亮度会明显低于正常区域的亮度,且图像中的阴影部分只占图像很小的一部分。因此根据图4本文设定亮度阈值为55。

初始种子的选择对于区域增长算法是至关重要的,只有初始种子落在与动态物体关联的阴影内,阴影区域才能正确的生长。为了保证图像中的每一个动态物体关联的阴影都有初始种子,根据1.1节求出的带有动态物体掩码的语义分割图制定了以下初

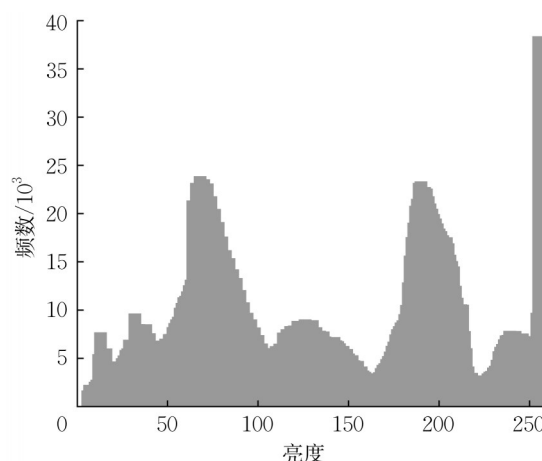


图4 亮度值分布直方图

Fig. 4 Histogram of brightness value distribution

始种子选择策略:①因为阴影总是落在动态物体的下方,因此选取与图像中动态物体掩码最下方所有

像素相邻的像素作为初始种子;②假设初始种子落在路面上,因此通过语义分割图分割出的路面对上一步获得的初始种子进行过滤。③根据设定的亮度阈值与初始种子自身的亮度值对上一步获得的初始种子进一步过滤。通过以上三个步骤可获得最终的初始种子列表。

1.2.2 区域增长策略

当初始种子选取完成后,根据初始种子与其周围8个像素的相似性以及像素的亮度值对阴影区域进行增长。相似性使用像素之间的欧式距离来判断,欧氏距离的计算如式(2)所示:

$$d = \sqrt{(r_{in} - r)^2 + (g_{in} - g)^2 + (b_{in} - b)^2} \quad (2)$$

式中: r_{in}, g_{in}, b_{in} 表示初始种子的R,G,B值; r, g, b 表示与初始种子相邻像素的R,G,B值。因光照等原因,图像中的房屋、树木的部分经常满足以上判断条件,因此,根据语义分割图将以上区域中满足判断条件的像素点去除。

当区域生长停止后,为防止检测到的阴影区域中含有遗漏的像素点,对不满足判断条件的像素点进行了遗漏检测。若不满足判断条件的像素点周围8个像素中有5个像素都属于阴影区域,那么该像素即为被遗漏的像素,也应属于阴影区域。阴影检测算法处理如算法1所示。

Algorithm 1 Shadow detection algorithm

Input: Original RGB image I_R , Semantic image with dynamic object mask I_S

Output: Shadow mask image I_M

```

1: seeds = GetInitialSeeds( $I_S$ )
2:  $I_H = \text{TranslateToHSL}(I_R)$ 
3: while length(seeds)  $\neq 0$  do
4: point  $\leftarrow$  seeds[0]
5: Remove seeds[0] in seeds
6: if point not processed then
7: point marked as processed
8: else

```

```

9: continue
10: end if
11: for adjacentpoint within adjacentpoints do
12: if adjacentpoint do not cross the border then
13: if adjacentpoint not processed then
14: if MeetRegionalGrowth(adjacentpoint,  $I_H$ ) then
15: Add adjacentpoint to seeds
16: else
17: Add adjacentpoint to leakseeds
18: end if
19: else
20: continue
21: end if
22: else
23: continue
24: end if
25: end for
26: end while
27: LeakDetect(leakseeds)

```

1.3 图像补全

在获得最终的动态物体掩码图后,根据动态物体掩码图将动态物体及其阴影在RGB图像上剔除,图像上会产生空洞掩码。之后使用图像补全算法来修补该图像中的空洞,这样就可以合成一幅只包含环境静态结构的图像,这种图像满足SLAM系统对静态环境的假设,对提高相机位姿估计精度有重大意义。

本研究的图像补全算法选择E2FGVI算法^[30],这是图像补全的最新技术。E2FGVI算法分为光流计算、特征传播以及图像间依赖关系建模三个可训练的模块。三个模块可以联合优化,从而实现更高效和准确的图像修复,E2FGVI的算法框架如图5所示。将当前RGB图像及其动态物体掩码图以及之前的6帧RGB图像作为算法的输入,算法输出对当前RGB图像去除动态物体后的补全图像。



图5 E2FGVI算法框架

Fig. 5 Algorithm framework of E2FGVI

2 实验

2.1 实验环境和数据集

在本节中,将本研究提出的图像预处理方法与经典的视觉 SLAM 系统双目 ORB-SLAM2 相结合并和原始双目 ORB-SLAM2 比较,以验证本研究所提出的图像预处理方法在动态环境中对视觉 SLAM 系统的有效性。此外,还进行了消融试验,以验证图像预处理方法中的每一个模块都是不可缺少的。所有实验均在配备 AMD Ryzen 7 4800H CPU、RTX2060 GPU 和 16GB 内存的计算机上进行。

KITTI 数据集是用于评估车载视觉 SLAM 系统定位精度的优秀数据集。该数据集由 22 个双目序列组成,其中前 11 个序列提供准确的真值。在本节中,使用 KITTI 数据集中的 5 个序列来评估性能,即高动态场景序列:KITTI 01、KITTI 09,低动态场景序列:KITTI 02、KITTI 04、KITTI 05。

为了定量评估该算法的性能,使用表示轨迹全局一致性的绝对轨迹误差(absolute trajectory error, ATE)和测量平移和旋转漂移的相对位姿误差(relative pose error, RPE)来评估系统的整体性能。均方根误差(root mean square error, RMSE)比均值和中值更能反映系统的准确性和鲁棒性,最大值

(MAX)和最小值(MIN)能反映系统的稳定性。因此,本文通过分别处理每个序列来获得 ATE 和 RPE 的 RMSE 值、MAX 值和 MIN 值,以评测位姿估计精度和系统稳定性。

2.2 实验结果

2.2.1 阴影检测模块对图像补全效果的影响

为了验证阴影检测模块对图像补全效果的影响,对阴影检测模块进行了消融实验,即分别将未对阴影掩码的动态物体掩码图和对阴影掩码的动态物体掩码图作为图像补全算法的掩码图输入,补全后的图像如图 6 所示。从图中可以看出,阴影检测模块是获得干净的修复背景的有效方式。虽然消除与动态物体关联的阴影的缺点是空洞被扩大了。众所周知,较大的空洞会使修复效果变差。但是,去除阴影后会降低找到匹配像素的难度,减少不同亮度的错误匹配块的影响。如图 6 所示,从不带有阴影检测模块生成的补全图像中可以看出,因需要补全的空洞附近是动态物体的阴影,所以补全后的部分大多是黑色的,而去除阴影后在空洞处的补全内容则更贴合实际。同时,阴影检测模块可以去除在连续图像之间移动的阴影鬼影。添加阴影检测模块也可以提高视觉 SLAM 的定位精度,此点将在下面的实验中详细说明。



图 6 阴影检测模块对图像补全效果影响对比图

Fig. 6 Comparison of impact of shadow detection module on image completion performance

2.2.2 整体系统评估

在本节中,对提出的图像预处理方法对视觉 SLAM 系统的提升效果进行了测试。使用 KITTI 数据集中具有动态物体的序列:KITTI 01、KITTI 09, KITTI 02、KITTI 04、KITTI 05 作为评估数据集。在上述数据集中对带本文提出的图像预处理方法的双目 ORB-SLAM2 系统和原始双目 ORB-SLAM2 系统进行测试。表 1 展示了在绝对轨迹误

差方面的对比结果,图 7 展示了在 KITTI 01、KITTI 05 以及 KITTI 09 序列上两种方法估计的轨迹与真值轨迹之间的差异图。从以上实验结果可以得到:本文所提出的图像预处理方法极大地提高了视觉 SLAM 系统在动态环境中的定位精度。对比在高动态场景和低动态场景数据集上的实验结果,可以看出场景中的动态物体越多提升效果越明显,这也说明了本文提出的图像预处理方法的有效性。以序列

05 为例,绘制了每隔 10、20、30、40 m 以及 50 m 的相对位姿误差,如图 8 所示。从图中可以看出,提出的图像预处理方法有效抑制了视觉 SLAM 系统在平移和旋转上的漂移,极大的提升了视觉 SLAM 系统的定位精度。

2.2.3 消融试验

为探索图像预处理方法中的每一个模块对视觉 SLAM 定位精度的影响,对阴影检测模块和图像补全模块进行了消融试验。即将不带有阴影检测模块的图像预处理方法处理后的图像和将不带有图像补全模块的图像预处理方法处理后的图像作为视觉 SLAM 系统的输入。两种消融试验输入图像的示例如图 9 所示。表 2 和表 3(最好的结果被加粗)分别展示了两种消融试验在绝对轨迹误差方面的对比结果。从表中可以看出,无论是去除阴影检测模块还是去除图像补全模块都对视觉 SLAM 系统的精度产生了影响,在一些序列上不仅对定位精度没有提升,反而会降低精度。结合图 9 可以得到,对于不带有图像补全模块的图像预处理方法处理后的图像,图像中的掩码部分的边缘易被检测为特征点,从而降低了精度。而对于不带有阴影检测模块的图像预处理方法处理后的图像,将阴影检测模块去除后,补

表 1 完整系统在绝对轨迹误差方面与原始 ORB-SLAM2 系统的对比结果

Tab. 1 Results of complete system in terms of absolute trajectory error compared to original ORB-SLAM2 system

数据集	指标	原始 ORB-SLAM2	图像预处理方法+ORB-SLAM2
高动态场景	01 MAX	16.265	11.063
	MIN	0.753	0.516
	RMSE	8.108	6.139
	09 MAX	6.861	2.276
	MIN	0.773	0.108
	RMSE	2.861	0.962
低动态场景	02 MAX	6.991	6.419
	MIN	0.293	0.321
	RMSE	3.213	2.967
	04 MAX	0.266	0.321
	MIN	0.042	0.030
	RMSE	0.156	0.133
05	MAX	1.005	0.787
	MIN	0.140	0.079
	RMSE	0.634	0.330

全后图像的质量大大下降,但总归是消除了动态物体,并对掩码部分进行了补全,使得 SLAM 系统不在遭受动态物体的干扰,但地上的阴影部分,会提供被稳定检测的特征点,使定位精度降低。

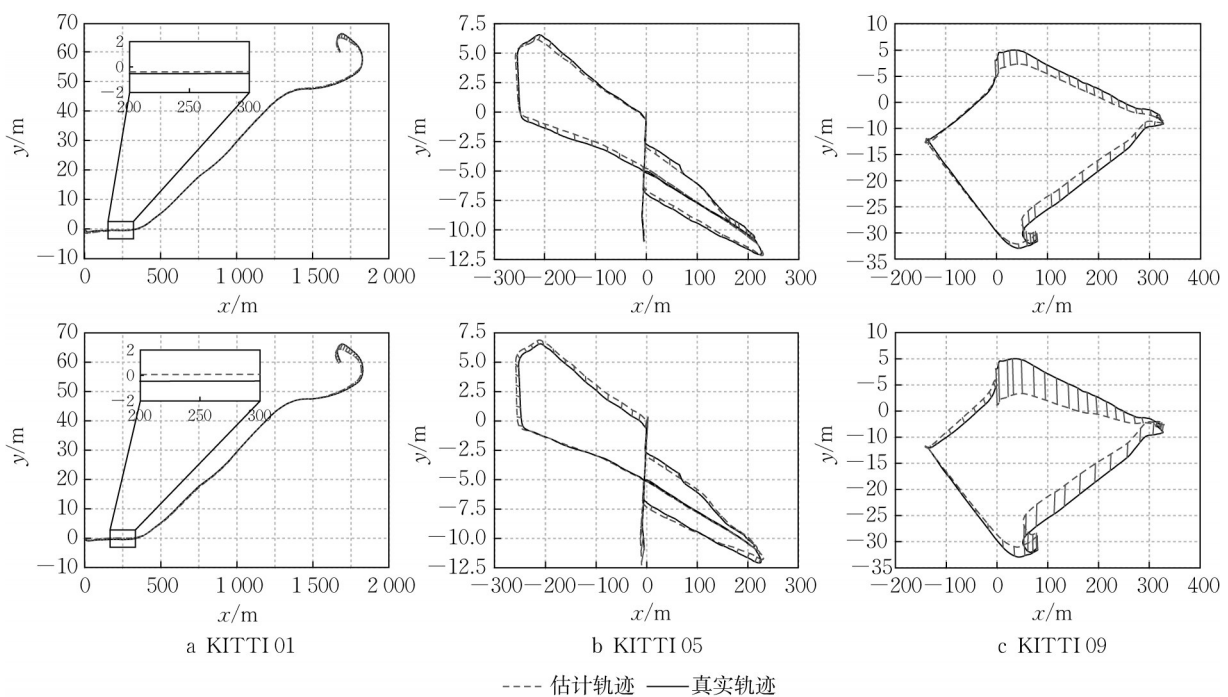


图 7 在高动态场景 KITTI01、KITTI09 以及低动态场景 KITTI05 上,带有本文提出的图像预处理方法的双目 ORB-SLAM2 系统(图中第一行)与原始双目 ORB-SLAM2 系统(图中第二行)的 ATE 图

Fig. 7 ATE diagram of stereo ORB-SLAM2 system with image pre-processing method proposed in this paper (first row) and original stereo ORB-SLAM2 system (second row) on high dynamic scenes KITTI01, KITTI09, and low dynamic scene KITTI05

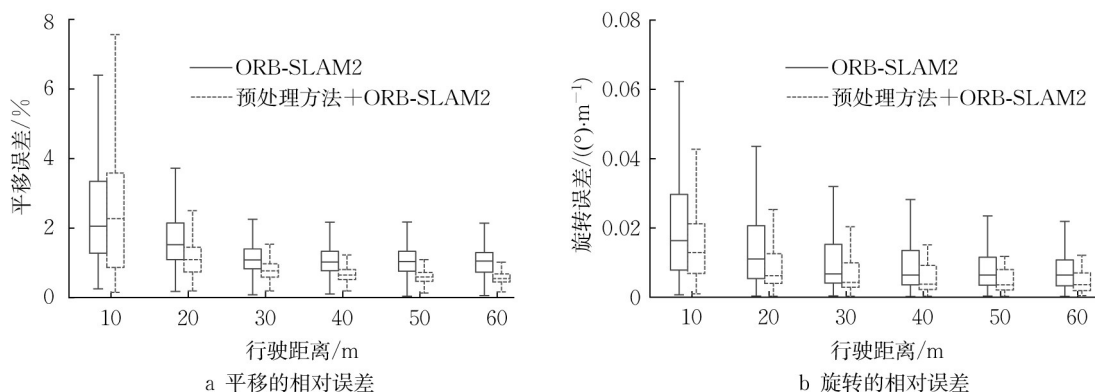


图8 KITTI 05序列中的相对位姿误差
Fig. 8 Relative pose errors in KITTI 05

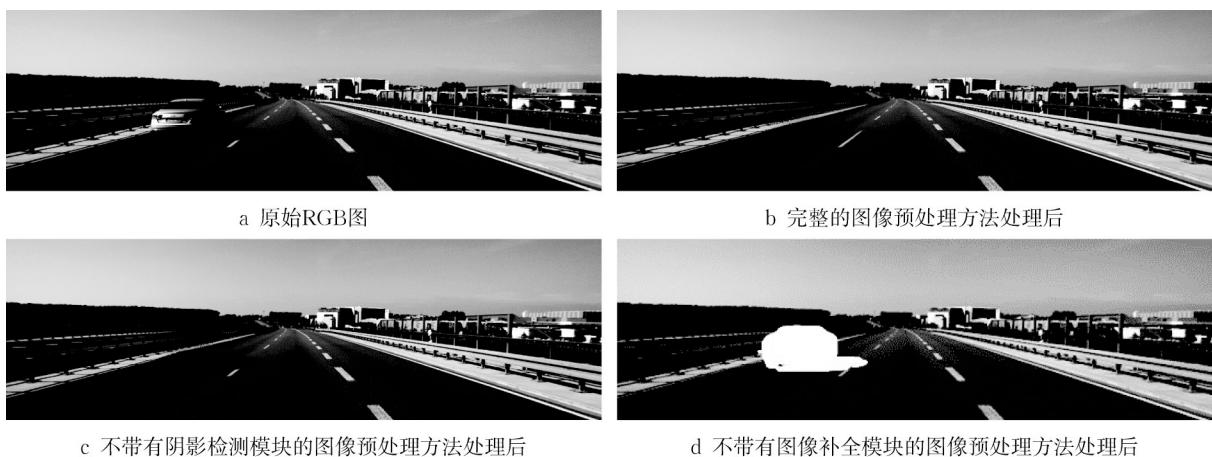


图9 消融试验输入图像示例

Fig. 9 Example image of ablation study input

表2 不带图像补全模块的系统在绝对轨迹误差方面与原始ORB-SLAM2系统的对比结果

Tab. 2 Results of the system without the image complement module in terms of absolute trajectory error compared to the original ORB-SLAM2 system

数据集	指标	原始ORB-SLAM2	不带图像补全模块的预处理方法+ORB-SLAM2
高动态场景	01	MAX	16.265
		MIN	0.753
		RMSE	8.108
	09	MAX	6.861
		MIN	0.773
		RMSE	2.861
低动态场景	02	MAX	6.991
		MIN	0.293
		RMSE	3.213
	04	MAX	0.266
		MIN	0.042
		RMSE	0.156
	05	MAX	1.005
		MIN	0.140
		RMSE	0.634

2.2.3 运行时间

以KITTI数据集01序列为例,对本文所提出的图像预处理算法中主要模块的运行时间进行了评估。进行了5次实验并计算平均值。评估结果如表

4所示。动态物体识别模块由光流估计算法和语义分割算法组成,这两个算法可以并行运行,其中光流估计算法的运行时间为0.11 s,语义分割算法的运行时间为0.005 s,因此该模块的总运行时间为0.11

表 3 不带阴影检测模块的系统在绝对轨迹误差方面与原始 ORB-SLAM2 系统的对比结果

Tab. 3 Results of the system without shadow detection module compared to original ORB-SLAM2 system in terms of absolute trajectory error

数据集	指标	原始 ORB-SLAM2	不带阴影检测模块的预处理方法+ORB-SLAM2
高动态场景	01	MAX	16.265
		MIN	18.201
		RMSE	0.534
	09	MAX	8.108
		MIN	6.861
		RMSE	8.740
低动态场景	02	MAX	1.187
		MIN	0.773
		RMSE	2.861
	04	MAX	4.013
		MIN	10.174
		RMSE	0.263
	05	MAX	4.621
		MIN	0.266
		RMSE	0.320

s。阴影检测模块由动态掩码生成和阴影掩码生成组成,阴影掩码由本文所提出的基于区域生长的阴影识别算法获得。区域生长是一种迭代的算法,因此时间开销会比较大。图像补全模块由图像掩码生成和图像补全算法组成,该模块的总运行时间为 0.19 s。综上所述,本文所提出的图像预处理方法的总运行时间为 0.67 s,其中阴影检测模块是最耗时的模块。未来可以通过硬件支持来缩短运行时间。

表 4 在 KITTI 数据集 01 序列上算法中各模块平均运行时间

Tab. 4 Average runtime of modules in proposed method on KITTI dataset Seq.01

模块名称	动态物体识别	阴影检测	图像补全
运行时间	0.11	0.37	0.19

3 结语

本文介绍了一种用于动态环境下视觉 SLAM 系统的图像预处理方法,该方法可以很容易地集成到现有视觉 SLAM 系统中,使视觉 SLAM 系统能够在动态环境中稳健、准确和连续的工作。该方法首先使用语义分割网络和光流估计网络鲁棒、精确地检测图像中潜在的动态物体。针对语义分割网络无法检测动态物体阴影的问题,提出了一种基于区域生长的阴影识别算法,对动态物体的阴影进行识别。之后,对图像中剔除动态物体及其阴影的空洞进行补全,获取被动态物体遮挡的特征信息。将本文所提出的图像预处理方法与双目 ORB-SLAM2 结合,

并在 KITTI 数据集上进行了实验,实验表明所提出的图像预处理方法显著地提升了视觉 SLAM 系统的定位精度,并且图像预处理方法中的每一个模块都有着不可替代的作用。在未来,计划采用多任务学习框架,以简化预处理方法中的网络架构以及采用硬件加速的方式来优化算法的运行时间。

作者贡献声明:

- 卓桂荣:研究思路指导、资源获取。
- 卢守义:算法设计及调试、试验数据获取、试验数据分析、论文撰写。
- 熊 璐:研究思路指导、资源获取。

参考文献:

[1] KUUTTI S, FALLAH S, KATSAROS K, *et al.* A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications [J]. IEEE Internet of Things Journal, 2018, 5(2): 829.

[2] 余卓平,高乐天,夏新,等.基于轮胎半径自适应的智能车辆组合定位[J]. 同济大学学报(自然科学版),2022,50(4):504.

YU Zhuoping, GAO Letian, XIA Xin, *et al.* Intelligent vehicle combined positioning based on adaptive tire radius [J]. Journal of Tongji University (Natural Science), 2022, 50(4):504.

[3] FAKHARIAN A, GUSTAFSSON T, MEHRFAM M. Adaptive Kalman filtering based navigation: An IMU/GPS integration approach [C]//2011 International Conference on Networking, Sensing and Control. [S. l.]: IEEE, 2011: 181-185.

[4] 周苏,李伟嘉,郭军华.基于激光雷达的停车场车辆定位算法[J]. 同济大学学报(自然科学版),2021,49(7):1029.

ZHOU Su, LI Weijia, GUO Junhua. Parking lot vehicle positioning algorithm based on LIDAR [J]. Journal of Tongji

- University (Natural Science), 2021, 49(7):1029.
- [5] CHEN W, SHANG G, JI A, *et al.* An overview on visual SLAM: From tradition to ssemantic [J]. Remote Sensing, 2022; 14(13):3010.
 - [6] DAVISON A J, REID I D, MOLTON N D, *et al.* MonoSLAM: Real-time single camera SLAM [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052.
 - [7] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C]//2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. [S. l.]:IEEE, 2007: 225-234.
 - [8] 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.
 - [9] 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.
 - [10] NEWCOMBE R A, LPVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time [C]//2011 International Conference on Computer Vision. [S. l.]: IEEE, 2011: 2320-2327.
 - [11] ENGEL J, SCHOPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM [C]//European Conference on Computer Vision. [S. l.]:Springer, 2014: 834-849.
 - [12] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 611.
 - [13] GAO X, WANG R, DEMMEL N, *et al.* LDSO: Direct sparse odometry with loop closure [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S. l.]:IEEE, 2018: 2198-2204.
 - [14] SAPUTRA M R U, MARKHAM A, TRIGONI N. Visual SLAM and structure from motion in dynamic environments: A survey [J]. ACM Computing Surveys (CSUR), 2018, 51(2): 1.
 - [15] DERPANIS K G. Overview of the RANSAC Algorithm [J]. Image Rochester NY, 2010, 4(1): 2.
 - [16] VYSOTSKA O, STACHNISS C. Exploiting building information from publicly available maps in graph-based SLAM [C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S. l.]:IEEE, 2016: 4511-4516.
 - [17] RICHARD H, ANDREW Z. Multiple view geometry in computer vision [M]. 2nd ed. Cambridge, U.K.: Cambridge University Press, 2003.
 - [18] KUNDU A, KRISHNA K M, SIVASWAMY J. Moving object detection by multi-view geometric techniques from a single camera mounted robot [C]//2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. [S. l.]:IEEE, 2009: 4306-4312.
 - [19] 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.
LIU Yusong, HE Li, YUAN Liang, *et al.* Semantic RGBD-SLAM algorithm based on optical flow in dynamic scenes [J]. Journal of Instrumentation, 2022, 43(12):139.
 - [20] 张小勇,张洪,高清源,等.室内动态场景下基于稀疏光流与实例分割的视觉SLAM算法[J].东华大学学报:自然科学版, 2023, 49(6):111.
ZHANG Xiaoyong, ZHANG Hong, GAO Qingyuan, *et al.* Visual SLAM algorithm based on sparse optical flow and instance segmentation in indoor dynamic scenes [J]. Journal of Donghua University (Natural Science), 2023, 49(6):111.
 - [21] 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). [S. l.]:IEEE, 2018: 1168-1174.
 - [22] BESCOS B, FACIL J M, CIVERA J, *et al.* DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes [J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076.
 - [23] ZHONG F, WANG S, ZHANG Z, *et al.* Detect-SLAM: Making object detection and SLAM mutually beneficial [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). [S. l.]:IEEE, 2018: 1001-1010.
 - [24] CHENG J, WANG Z, ZHOU H, *et al.* DM-SLAM: A feature-based SLAM system for rigid dynamic scenes [J]. ISPRS International Journal of Geo-Information, 2020, 9(4): 202.
 - [25] FAN M, LAI S, HUANG J, *et al.* Rethinking BiSeNet for real-time semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2021: 9716-9725.
 - [26] CORDTS M, OMRAN M, RAMOS S, *et al.* The cityscapes dataset for semantic urban scene understanding [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2016: 3213-3223.
 - [27] GEIGER A, LENZ P, STILLER C, *et al.* Vision meets robotics: The kitti dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231.
 - [28] ZHAO S, SHENG Y, DONG Y, *et al.* Maskflownet: Asymmetric feature matching with learnable occlusion mask [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2020: 6278-6287.
 - [29] BUTLER D J, WULFF J, STANLEY G B, *et al.* A naturalistic open source movie for optical flow evaluation [C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 611-625.
 - [30] LI Z, LU C Z, QIN J, *et al.* Towards an end-to-end framework for flow-guided video inpainting [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE, 2022: 17562-17571.