

DOI:10.11918/202001039

结合道路结构化特征的语义 SLAM 算法

李琳辉,张溪桐,连 静,周雅夫,郑伟娜

(工业装备结构分析国家重点实验室(大连理工大学),辽宁 大连 116024)

摘要: 视觉 SLAM(simultaneous localization and mapping)是智能车辆领域的研究热点,在包含运动目标干扰或近景特征不显著的场景中,容易产生帧间位姿估计结果精度不足或失效问题. 为此,本文提出一种结合场景语义信息和路面结构化特征的 SLAM 算法. 首先,针对上述特殊场景中运动目标干扰的情况,设计带有改进金字塔池化模块的语义分割神经网络,得到图像中各像素对应的目标类别,作为剔除运动像素点的依据,从而避免运动点参与特征匹配导致的位姿计算准确性下降问题;然后,针对有效近景特征点不足的情况,基于 V 视差算法确定图像中的道路平面区域并拟合出精确的视差方程,以计算路面上像素点的精确视差值,并提出一种基于路面结构化特征(车道线、马路边界、路面交通标记等)的位姿计算方法;最后通过场景实验得出,本文提出的改进算法计算结果的绝对轨迹误差小于原算法. 证明该方法能够在存在运动目标干扰或缺乏近景特征的场景中具有较高的位姿估计精度,建立了有效的包含语义信息的稠密点云地图,具有良好的环境适应性.

关键词: 智能车辆;SLAM;语义分割;V 视差;结构化特征

中图分类号: U469.79

文献标志码: A

文章编号: 0367-6234(2021)02-0175-09

Semantic SLAM algorithm combined with road structured features

LI Linhui, ZHANG Xitong, LIAN Jing, ZHOU Yafu, ZHENG Weina

(State Key Laboratory of Structural Analysis for Industrial Equipment (Dalian University of Technology),
Dalian 116024, Liaoning, China)

Abstract: Vision-based simultaneous localization and mapping (SLAM) is a research hotspot in the field of intelligent driving. However, for the scenes that contain moving targets or inconspicuous close-range features, it is easy to cause ineffective or inaccurate pose estimation between frames. To solve this problem, this paper proposes a SLAM algorithm based on road structured features and scene semantic information. First, for the problem of target moving, a semantic segmentation neural network with an improved pyramid pooling module was designed to obtain the target category corresponding to each pixel in the image. The segmentation results were taken as basis for the elimination of moving points, which avoids the problem of low accuracy of pose calculation caused by moving points participating in feature matching. Then, in view of the lack of effective feature points, the road area in the image was determined based on v-disparity algorithm, and disparity function was obtained to calculate the accurate disparity value of the pixels on the road. Furthermore, a pose calculation method based on road structured features (e.g., lane lines, road boundaries, pavement traffic markings) was proposed. Finally, scene experiments were carried out and results show that the absolute trajectory error of the improved algorithm proposed in this paper was smaller than that of the original algorithm, which proves that the proposed method has higher pose estimation accuracy in scenes with moving targets or inconspicuous close-range features. In addition, an effective dense point cloud map containing semantic information was established, which has good environmental adaptability.

Keywords: intelligent vehicle; SLAM; semantic segmentation; v-disparity; structured feature

SLAM(simultaneous localization and mapping,同时定位及地图构建)是实现智能车辆自主导航的关键,受到国内外学者的广泛重视^[1]. 根据传感器的不同,SLAM 可分为基于激光传感器的 LSLAM^[2]和基于视觉传感器的 VSLAM^[3]. 其中,基于视觉传感

器的 SLAM 具有成本低、获得的环境信息丰富等优点,是当前智能驾驶领域的研究热点.

VSLAM 可分为以 SVO^[4]和 LSD-SLAM^[5]为代表的直接法 SLAM,以及以 SOFT-SLAM^[6]和 ORB-SLAM2^[7]为代表的特征点法 SLAM. 其中,特征点法 SLAM 适用于光照条件较差和相机发生大尺度位移或旋转的场景,具有更强的鲁棒性. 该方法首先提取图像中的显著特征,计算特征的描述子,再根据两帧图像中特征的描述子距离进行匹配,最后根据匹配关系用对极约束、PnP 或 ICP 算法求解出两帧之间

收稿日期: 2020-01-08

基金项目: 国家自然科学基金(61976039,51775082);中央高校基本科研业务费专项基金(DUT19LAB36,DUT17LAB11)

作者简介: 李琳辉(1981—),男,副教授;

连 静(1980—),女,副教授,博士生导师

通信作者: 连 静,lianjing@dlut.edu.cn

的位姿。

特征点法 SLAM 根据图像中包含的世界模型点来反推相机位姿,处于运动状态的特征点会使计算的准确性受到影响,因此不能解决动态场景中位姿计算和地图构建的精度下降问题。如何区分特征点的动静状态是消除运动物体对位姿计算带来不利影响的重点。针对动态场景中缓慢运动的非刚性物体,Agudo 等^[8]使用物理先验算法在动态场景中实现了姿态估计;Ahuja 等^[9]利用 SfM 算法对场景中的运动对象进行实时 3D 重建。针对动态场景中存在快速运动的物体(如自行车、汽车等),Yu 等^[10]提出一种光流跟踪帧间图像的一致性检验方法来找到运动的像素点,能检测到场景中快速运动的物体。从地图构建和方便智能车环境理解的角度而言,如果单纯从动静点区分的角度改善建图精度,一方面很难适应动点多于静点的场景,另一方面会将暂时静止的车辆、行人等可移动目标囊括到地图中。所以,在 SLAM 过程中结合场景的语义信息^[11-12]显得尤为必要。

目前,以 DeepLab^[13]、FastFCN^[14]、SDCnet^[15]以及 PSPNet^[16]为代表的基于深度学习的语义分割方法具有较高的准确性和鲁棒性。其中,Zhao 等^[16]提出的 PSPNet(pyramid scene parsing network)利用带有空洞卷积的特征提取和金字塔池化模块,有效提取图像的上下文信息,能够进一步提高语义分割的精度。

然而,针对某些近景特征不足、光照条件恶劣或存在大量运动目标的场景,在剔除运动目标后,特征点数量不足会导致位姿计算失效。针对该问题,研究者利用场景中线特征^[17]、角特征^[18]等其他特征与特征点结合,共同参与位姿计算。然而这些特征只有在特殊环境下才比较常见,在一些开阔的大型场景下仍存在特征不足的问题。Jeong 等^[19]就大型室外场景提出了 Road-SLAM,该算法利用路面结构化特征对 SLAM 的回环检测部分进行了优化。这些特征在光照条件差的场景下也容易被检测,并且分布在路面这一单一平面上,易于通过立体匹配计算其准确深度,可以用于解决位姿计算过程中近景有效特征不足的问题。除此之外,路面结构化特征还可以应用于车辆车道保持、路径规划等驾驶任务。

基于以上分析,本文提出了一种结合场景语义信息和路面结构化特征的 SLAM 算法。针对复杂交通场景,设计带有改进金字塔池化模块的语义分割网络,得到图像中各像素对应的目标类别,作为运动点剔除的依据,从而避免运动点参与特征匹配而导致的位姿计算准确性下降问题;针对包含大量运动

物体、近景特征点不足、光照条件差等恶劣场景,基于 V 视差确定图像中的路面区域并拟合出精确的视差值,提出一种基于路面结构化特征的位姿计算方法;最后,通过场景实验验证所提出方法的有效性,并建立包含语义信息的稠密点云地图,为复杂的驾驶任务提供尽可能全面的环境信息。

1 算法架构

如图 1 所示,本文设计的 SLAM 算法整体架构参考 ORB-SLAM 算法,共包含 4 个主要模块:位姿计算、稀疏地图构建、回环检测和语义重建。其中稀疏地图构建模块、回环检测模块沿用 ORB-SLAM 的基础框架。在此基础上,针对该 SLAM 算法在包含动态物体以及近景特征不足场景下的精度下降问题,对其位姿计算模块进行改进。同时,为给自动驾驶任务提供更丰富的环境信息,增加了语义重建模块。主要研究工作如下。

1.1 位姿计算模块改进

1) 动态特征点剔除:为消除场景中动态物体对位姿计算精度的影响,本文通过设计的带有改进的带孔金字塔池化模块^[20](advanced atrous spatial pyramid pooling, ad-ASPP)的神经网络对输入的图片进行语义分割,目标类别包括天空、建筑物、机动车、行人、自行车等。同时,对图像进行 ORB 特征点检测。若场景中某目标包含运动的可能性,则将其定义为动态目标(机动车、行人和自行车),属于动态目标类别的像素点为动态点。本文根据语义分割的结果,将动态点从特征点检测结果中剔除。

2) 立体匹配有效范围外特征点剔除:根据立体匹配原理,场景中相机光心距离过远的点的计算结果会失效,需要对候选点进行筛选,剔除不在立体匹配有效范围内的特征点。

3) 道路结构化特征检测与匹配:进行过两次筛选后,剩余特征点的数量如果达到计算位姿所需点数量的阈值,则直接进行位姿计算过程;若特征点数量未达到阈值,则利用本文提出的基于路面结构化信息的方法进行帧间位姿计算。该方法对图像中的道路结构化特征进行检测与匹配,通过增加路面结构化特征点解决参与 SLAM 位姿计算的有效近景特征不足的问题。

1.2 增加语义重建模块

ORB-SLAM 原本的稀疏地图构建模块构建的点云地图是稀疏的,缺失了大部分的环境信息。因此,本文增加了一个语义重建模块。在系统完成位姿计算和回环检测模块之后得到最终的位姿估计结果。根据稀疏地图构建模块筛选添加后得到的关键帧结

果,以及回环检测模块优化后的位姿计算结果,将当前关键帧图像中的所有像素三维重构成世界坐标系下的点云;同时根据卷积神经网络获取到的语义分

割结果,对点云进行类别属性的定义,从而构建出3D稠密语义点云地图.该地图同时包含了环境的几何信息和语义信息.

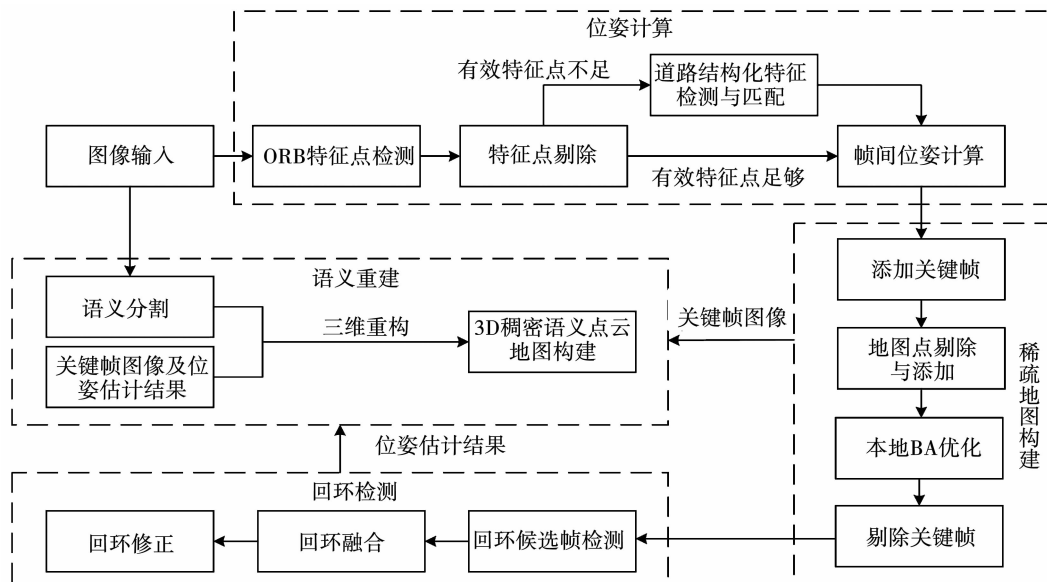


图 1 算法整体流程

Fig. 1 Overall flow chart of algorithm

2 语义分割网络

2.1 主网络结构设计

本文语义分割网络的总体结构见图 2。网络主要分为编码器和解码器两个部分。其中编码器部分采用残差神经网络 (ResNet) 作为特征提取基础网络^[21], 解密码器部分采用改进的带孔金字塔池化模块 (ad-ASPP)。输入的图像经过 ResNet50 特征提取

模块的卷积操作后,得到相较于原图大小1/8的特征图,然后与本文所提出的改进的带孔金字塔池化模块(ad-ASPP)相连,最后得到原图大小的语义分割结果图.将 ad-ASPP 模块置于残差网络之后,可以利用深度网络提取得到更加抽象的语义信息,扩大高级特征信息感受野,这些特征比低级的轮廓、纹理信息更加有用.

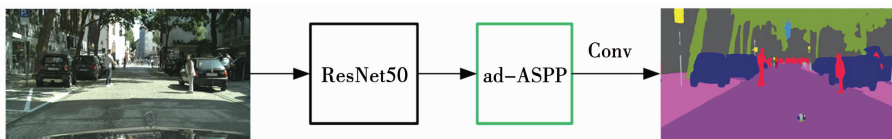


图2 语义分割主网络结构

Fig. 2 Main architecture of semantic segmentation network

2.2 特征提取模块

本文使用的语义分割网络选用允许网络拓展深度而不会产生梯度爆炸和梯度消失的 ResNet50 作为特征提取的主体. ResNet50 主要由 5 个模块组成, 其中各个模块的参数构造见表 1. 其中包括每个模块的输出特征图像大小, 以及每一个模块中各个卷积层的卷积核参数. 特征提取模块的输入是大小为 512×1024 的 RGB 图片, 输出得到大小为 32×64 的特征图像.

2.3 改进的带孔金字塔池化模块(ad-ASPP)

在卷积神经网络中,空洞卷积^[22]通过对卷积核进行插零来扩大每一次卷积的感受野,能够提取到

原本非近邻像素的特征信息,从而获取图像的上下文信息,这一信息在语义分割任务中能够起到重要的作用. 本文根据相关研究提出了一种改进的空洞卷积金字塔池化模型,其主体结构见图 3.

1) 受经典网络 ResNet 的启发,加入了一个直连支路,将该模块的输入直接传递给输出,可以有效保证图片的原始信息不因卷积过程而丢失.

2)不合理地设置空洞卷积的扩张率会造成各种问题,当扩张率过大时,感受野的过大会造成提取到的信息关联性较差,不利于图像中细小物体的特征提取。为合理设置空洞卷积的扩张率,保证信息之间的关联性,本文还增加了一个级联空洞卷积的神

表 1 ResNet-50 模型结构
Tab. 1 ResNet-50 model structure

模块	输出大小(width × height)	卷积核参数
模块 1	256 × 512	$7 \times 7, 64, S_{\text{stride}} = 4$ $3 \times 3, \text{maxpooling}, S_{\text{stride}} = 2$
模块 2	128 × 256	$\left\{ \begin{array}{l} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{array} \right\} \times 3$
模块 3	64 × 128	$\left\{ \begin{array}{l} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{array} \right\} \times 4$
模块 4	64 × 128	$\left\{ \begin{array}{l} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1\,024 \end{array} \right\} \times 6$
模块 5	32 × 64	$\left\{ \begin{array}{l} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2\,048 \end{array} \right\} \times 3$

经网络支路. 该结构在原始的空洞卷积基础上增加了感受野的尺寸,同时消除了不合理的扩张率设置对特征信息关联性的不良影响. 假设未被级联前的两个空洞卷积层的感受野大小分别为 N_1 和 N_2 , 则将二者级联后其感受野尺寸变为 $N = N_1 + N_2 - 1$. 本文设计的级联空洞卷积扩张率分别设置为 6、12、18、24.

3) 受 DeepLab^[23] 网络和 PSPNet^[16] 启发, 将该模块设计成多支路并联的形式, 每条支路提取不同尺寸感受野内图像的特征, 利用 concat 模块将这些支路的特征提取结果合并. 不同感受野的融合使网络在关注图像局部细节的同时把握全局的结构特征.

4) 空洞卷积层的串并联共存结构可以使图像中的每一个像素都更积极地参与计算, 在一定程度上对于“棋盘格效应”^[24] 起到了削弱的作用.

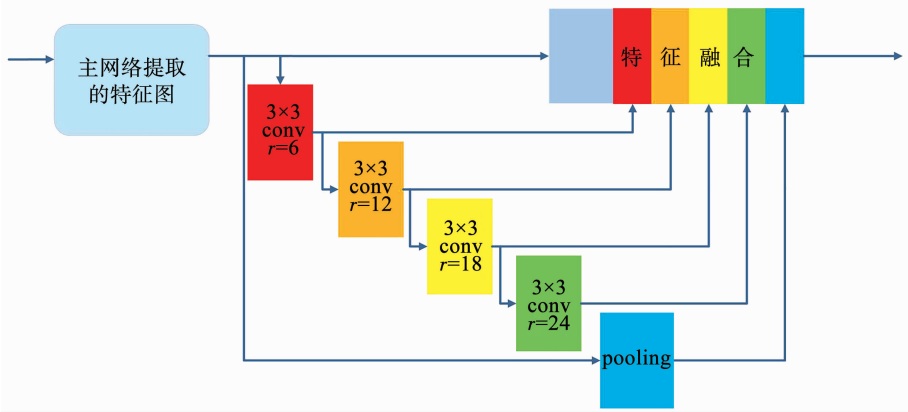


图 3 改进的带孔金字塔池化 (ad-ASPP) 模块
Fig. 3 Advanced atrous spatial pyramid pooling (ad-ASPP) module

3 基于路面结构化特征的位姿计算

若行驶过程中相机拍摄到的运动物体过多,特征点匹配过程中根据语义分割结果筛选掉大部分近处的运动特征点,只剩下远处的匹配特征点,在双目立体匹配过程中这些点可能会由于超出其有效距离范围而无法正确进行准确的立体匹配,这会导致相机位姿计算结果失效,所以本文提出了一种基于路面结构化特征检测与匹配的相机位姿计算方法. 算法整体流程见图 4.

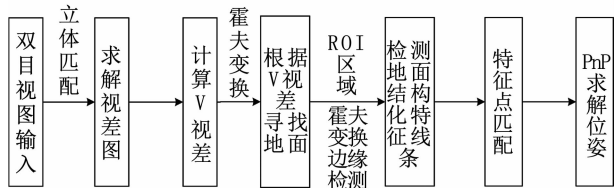


图 4 路面结构化特征位姿求解算法

Fig. 4 Pose solving algorithm based on structured features of road surface

首先,对输入的双目视图进行立体匹配,求解出该帧图像的视差图,然后利用视差图获得 V 视差图;根据世界坐标中某一平面在 V 视差图的表现,对 V 视差图进行霍夫变换,检测出表示地面的直线;根据这条直线能够准确地找到地面在图像中的区域,并将该区域设为感兴趣 (range of interest, ROI) 区域;在 ROI 区域中利用边缘检测和霍夫变换检测路面结构化特征线条,然后对线条上的特征点进行帧间匹配,最后根据匹配结果利用 PnP 算法求解出相机位姿.

3.1 基于 V 视差的地面检测

在世界坐标系下,设某点 P 的坐标为 $P^w = [x^w, y^w, z^w]^T$, 其投影到相机成像平面的像素坐标为 $P^p = [u, v]^T$. 对于世界坐标系中的某一平面,其数学模型为: $y^w = h$, h 代表该平面的水平高度. 则该平面视差值满足

$$\frac{h}{b} \cdot d_{\text{disp}} = f \cdot \sin \theta + v \cdot \cos \theta.$$

(1)

式中: d_{disp} 代表平面某点视差值, f 代表相机焦距, b 代表双目相机的基线长度, θ 为相机主光轴方向与道路平面夹角. 可知, 地面上点的像素纵坐标与其视差值成线性关系.

V 视差^[25]是基于双目立体匹配获得的视差图得到的, 包含 3 个维度的信息. 其图像的纵轴表示原

视差图的行坐标, 横轴表示 0 ~ 255 的视差值, 强度轴表示原视差图中视差值为某个值的像素点数量, 见图 5. 根据上述推论, 水平路面在 V 视差图中表现为一条斜率为负的倾斜直线. 对 V 视差图进行霍夫检测, 可以求出代表地面的直线 (图 5 中的红色斜线), 即地面的视差方程为 $d_{\text{disp}} = k \cdot v + b$.

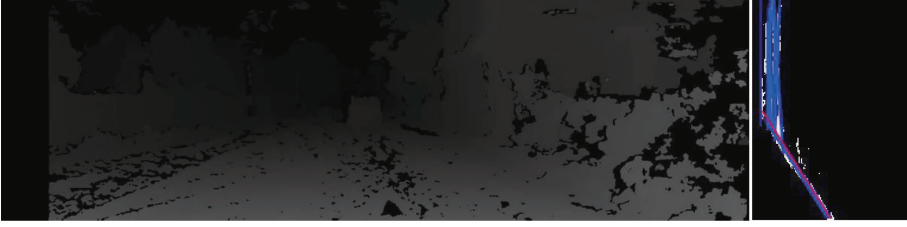


图 5 通过 V 视差图获取地面点视差

Fig. 5 V-disparity image generated from original disparity map

3.2 道路结构化特征检测

找到地面在图像中的位置后, 在属于地平面邻近平面的图像范围内检测道路上的结构化特征. 首先对图像进行二值化和 Canny 边缘检测, 结果见图 7(a); 利用本文求解得到的地面视差方程将属于地面的区域保留, 结果见图 6(b), 基本滤除了非地面的部分; 利用霍夫变换检测出路面上的结构化特征线条, 检测结果见图 6(c).



图 6 道路结构化特征检测

Fig. 6 Road structured features detection

3.3 PnP 位姿求解

检测到道路结构特征线条后, 对前后两帧线条上的特征点进行描述子计算和匹配, 得到用于位姿计算的点对, 用于位姿求解. 本文采取 RANSAC-PnP^[26]方法中的 3D-2D 模型求解相机位

姿. 该求解模型需要输入控制点, 即匹配点对的 3D 世界坐标和待求帧该点的 2D 像素坐标. 该特征点在世界坐标系中的坐标求解过程如下.

1) 像素坐标系 - 相机坐标系. 由相机成像原理可知, 相机成像过程中一共包含 4 个坐标系. 设图像中存在 n 个像素点, 在世界坐标系中的坐标为 $P_i^w = [x_i^w, y_i^w, z_i^w]^T$, 在相机坐标系中的坐标为 $P_i^c = [x_i^c, y_i^c, z_i^c]^T$, 这些点在图像坐标系中的坐标为 $p_i = [x_i, y_i]^T$, 像素坐标为 $P^p = [u_i, v_i]^T$. 从像素坐标系到世界坐标系的转换关系为

$$\begin{cases} x_i = (u_i - c_x) \cdot \frac{z_i}{f_x}, \\ y_i = (v_i - c_y) \cdot \frac{z_i}{f_y}, \\ z_i = d. \end{cases} \quad (2)$$

式中: f_x, f_y, c_x, c_y 为相机内参; (u, v) 为图像坐标; (x, y, z) 为相机坐标系坐标, 其中 z 为该点深度值. 根据本文中对地面视差的求解结果, 通过其映射关系可以得到地面上的像素点较为准确的视差值 d_{disp} , 进而通过下式求出该点深度值

$$d_{\text{depth}} = \frac{bf}{d_{\text{disp}}}. \quad (3)$$

2) 相机坐标系 - 世界坐标系. 相机坐标系到世界坐标系的转换关系为

$$P_i^w = T_i^{wc} P_i^c, \quad (4)$$

式中, T_i^{wc} 为该关键帧的位姿, 由 R, t 组成. R, t 称作相机的外参, R 为相机的旋转矩阵, $R \in SO(3)$, 代表相机当前的姿态; t 为相机的平移向量, $t \in R^{3 \times 1}$, 代表相机当前所处的位置.

基于上述公式, 根据特征点的像素坐标及深度值求出其相机坐标系坐标, 然后再根据前一帧的相机位姿计算得到特征点的 3D 世界坐标, 结合这些

特征点在当前帧上的 2D 像素坐标,就可以利用 PnP 方法求解出当前帧的位姿。

3)稠密地图的构建是基于 SLAM 算法得到的关键帧位姿,利用关键帧对应的语义地图以及视差图进行三维重建的过程. 设当前参与稠密点云地图构建的关键帧图像中所有的像素点群为 p_i , 已知其像素坐标 $p_i = [x_i, y_i]^T$, 则构建点云地图需要求解出其相应的世界坐标系中的坐标 $P_i^w = [x_i^w, y_i^w, z_i^w]^T$.

a)首先,通过对当前关键帧的左右视图进行立体匹配,可以获取某像素点的视差值 d_{disp} , 根据式(3)可求得其深度值 d_i . 再根据式(2)可以求得这些像素点的相机坐标系下坐标 $P_i^c = [x_i^c, y_i^c, z_i^c]^T$.

b)求解出这些像素点的相机坐标后,根据式(4)可以计算得到其世界坐标,其中 T_i^{wc} 即为 SLAM 位姿求解模块所求得当前关键帧位姿。

c)得到这些点群的世界坐标后,结合本文的语义分割结果,得到包含语义信息的点群. 对每一个关键帧进行上述操作,建立稠密语义点云地图。

4 实验分析

4.1 语义分割网络

4.1.1 数据集

1)针对改进的语义分割网络进行训练,训练的框架基于 TensorFlow 软件平台. 选择 Cityscapes^[27]数据集作为语义分割网络的研究对象。

2)数据集进入神经网络之前需进行预处理. 首先,为了减小训练时长和降低计算机内存消耗,将图像尺寸转化为 256 × 512;其次,本文使用 pixel-mean 方法取均值,使得神经网络在不会产生梯度过大的基础上节省反复试验最佳学习率过程的时间。

3)由于数据集的图片数量较小,为防止网络训练过程中过拟合现象的发生,对图片进行数据增强. 主要操作为对图片进行一定的旋转、平移、缩放、镜像,这种样本的增加可以提高网络训练结果的泛化性。

4.1.2 测试结果及分析

在 Cityscapes 数据集上,将本文设计的神经网络与使用同样基础网络的 PSPNet 进行比较. 训练网络所用的计算机软、硬件配置及深度学习框架见表 2。

表 2 训练软、硬件条件

Tab.2 Training software and hardware conditions	
项目	内容
CPU	Intel Xeon E5-2620
GPU	GeForce GTX TITAN X
RAM	32 GB
Cuda	Cuda9.0 with Cudnn 7.0
深度学习框架	TensorFlow

根据硬件的计算能力,将图片输入的大小设为 512 × 1 024, batch_size 的大小设为 3, 当损失值稳定收敛后停止训练,共计 16 万次,得到表 3 的结果. 可知,本文所设计的网络结构在运行时间基本相同的条件下,精度上相较于 PSPNet 有明显的提高,证明了语义分割网络改进的有效性。

图 7(b)、(c)分别为表中两个网络的语义分割结果. 可以看出,图(c)与图(b)相比,分割的噪声更少,道路与车辆边界的区域分割更加准确. 所以,本文采用的改进的金字塔模块能够更好地提取图像上下文的信息,实现更加准确的图像像素级语义分割。

表 3 网络性能定量比较

Tab.3 Quantitative comparison of network performances			
网络名称	mIoU/%	测试时间/ms	模型大小/M
ResNet50 + ad-ASPP	71.86	332.9	306
PSPNet50	70.28	333.5	374

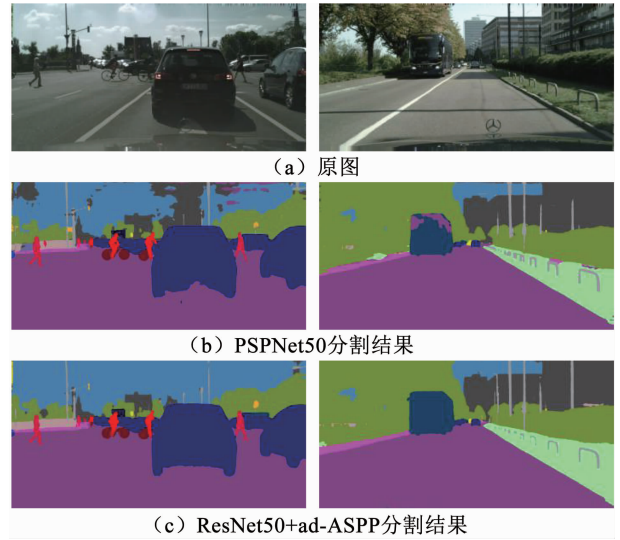


图 7 基于 RGB 图像的分割效果对比

Fig.7 Comparison of segmentation effects based on RGB image

4.2 道路结构化特征位姿求解

1)数据集. 本文选取 Kitti_odometry 数据集^[28]作为位姿计算模块的数据集. 该数据集包含 22 个不同交通场景的双目图像数据,配套提供了 11 个带有地面真实轨迹,便于 SLAM 结果的评估. 本文所设计的 SLAM 方法需要使用到图像的语义信息,所以在输入 SLAM 系统之前需要对图像进行语义分割。

2)算法结果比较. 如图 8 所示,在该场景下,相机视野左侧存在运动车辆,正以与相机所在车辆近乎相对静止的速度行进. 在特征点检测与匹配的过程中,特征点法 SLAM 特征点检测以及匹配过程会将其列为特征点并参与匹配,匹配结果见图 8。

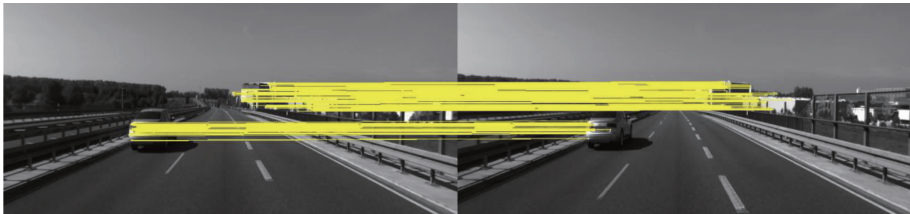


图 8 未去除运动物体特征点的匹配结果

Fig.8 Matching results without eliminating moving objects

经过计算,未去除动态特征点的匹配得出的位姿结果中,图 8 两帧之间在车辆前进方向上的相对位移为 0.294 8 m,但其真实位移值为 2.14 m,可见动态特征点参与匹配会使相机位姿计算的准确率降低.并且在此类场景中,若去除属于车辆的特征点和过远处的特征点,则用于计算位姿的特征点数量是不足的.

本实验节选了 4 段包含动态物体且有效近景特征不足连续帧场景(见图 9),分别利用本文提出

的算法和纯特征点匹配求解相机位姿,比较结果见表 4.从表中可知,分别从误差最大值、平均值、中值、最小值和均方根误差值分析,在场景中存在运动物体并且其他特征点质量不满足要求的情况下,本文提出的结合道路结构化特征的位姿计算方法具有更高的精度.此外,表 4 中还给出了 2 种算法的单帧平均运行时间.由于本文增加了道路结构化特征点检测与匹配的过程,算法的平均耗时约为未增加该过程算法的 1.5~1.7 倍.



图 9 包含运动物体的场景

Fig.9 Scenes with moving objects

表 4 绝对轨迹误差结果比较

Tab.4 Comparison of absolute trajectory error results

评价参数	场景 a		场景 b		场景 c		场景 d	
	道路结构化	特征点匹配	道路结构化	特征点匹配	道路结构化	特征点匹配	道路结构化	特征点匹配
	特征位姿求解	位姿求解	特征位姿求解	位姿求解	特征位姿求解	位姿求解	特征位姿求解	位姿求解
误差最大值/m	1.729 052	2.801 711	2.220 739	2.826 005	1.213 037	1.658 743	0.925 077	2.825 505
误差平均/m	1.699 147	2.800 682	2.109 046	2.825 913	1.205 763	1.644 442	0.905 964	2.824 851
误差中值/m	1.700 763	2.800 611	2.128 177	2.825 916	1.203 548	1.643 564	0.915 736	2.825 104
误差最小/m	1.674 567	2.799 458	1.951 456	2.825 829	1.200 626	1.634 796	0.800 799	2.823 842
均方根误差/m	1.699 262	2.800 682	2.111 262	2.825 913	1.205 769	1.644 458	0.906 372	2.824 851
单帧平均运行时间/ms	559.671	328.141	468.964	298.180	559.603	395.187	551.290	310.764

5.3 3D 稠密语义点云地图构建

场景整体的地图见图 10.其中,图 10(a)为原 ORB-SLAM 构建出的整体场景稀疏点云地图,可见虽然该地图包含了环境的整体架构信息,但是缺少很多局部场景细节.而本文加入语义重建模块后,构

建出的 3D 稠密语义点云地图如图 10(b)所示,实现了场景细节的补充.图 10(c)是其中某一场景的稠密语义地图局部展示.场景中包含的车辆、树木、道路、建筑物都以点云的形式构建在地图中,点的颜色代表了它们所属的目标类别.可见,该地图在包含环

境几何信息的基础上,还加入了局部场景的细节与环境的语义信息. 本文构建的 3D 稠密语义地图模

拟了驾驶员驾驶车辆时对环境信息的感知,可以为智能驾驶提供更丰富的环境信息.

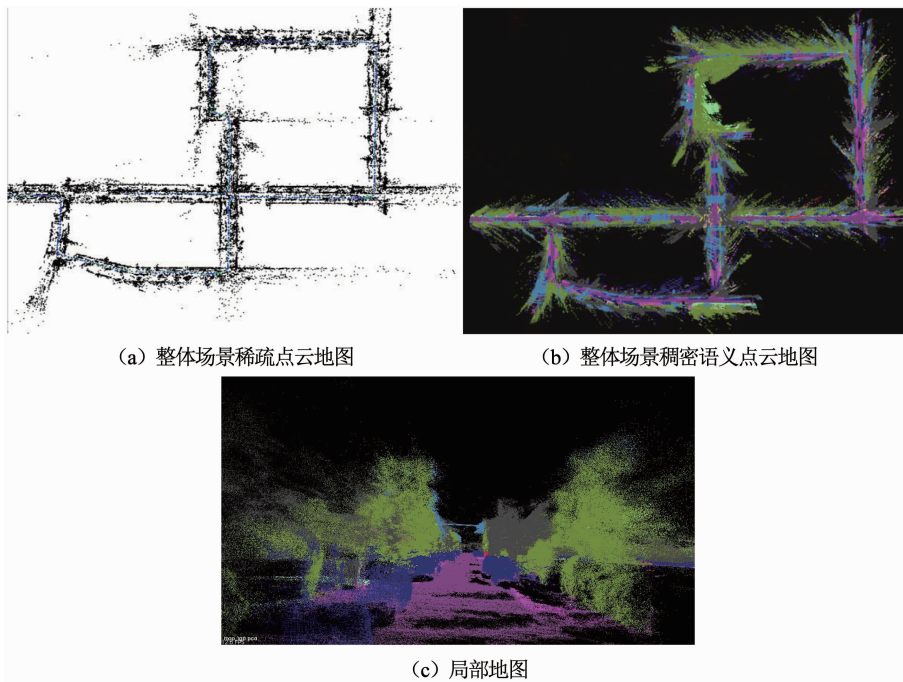


图 10 3D 稠密语义点云地图

Fig. 10 3D dense semantic point cloud maps

5 结 论

本文设计了一套在动态环境下对 SLAM 算法位姿计算模块的改进方法. 提出了一种带有级联金字塔空洞卷积池化模块的语义分割网络,提取了图像中更加丰富的上下文信息,提高了交通场景下像素级语义分割的精度. 利用语义分割的结果,本文在 SLAM 特征点检测过程中剔除了属于包含动态可能性类别的物体,防止运动物体使相机位姿计算产生误差. 同时,针对剔除掉动态物体特征点后有效近景特征不足的场景,提出了一种通过检测道路结构化特征并进行帧间匹配进行相机位姿计算的方法. 经过实验证明,该方法能够有效提高动态场景下相机位姿计算的准确性. 最后,将 SLAM 位姿计算结果与语义分割结果进行融合,建立了包含语义信息的稠密点云地图,为自动驾驶任务提供更加丰富的环境信息,对实现更加复杂的驾驶任务具有重要意义.

参考文献

- [1] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309. DOI: 10.1109/TRO.2016.2624754
- [2] SHEN Dong, HUANG Yakun, WANG Yangxi, et al. Research and implementation of SLAM based on LIDAR for four-wheeled mobile robot [C]//2018 IEEE International Conference of Intelligent

- Robotic and Control Engineering (IRCE). Lanzhou: IEEE, 2018; 19. DOI: 10.1109/IRCE.2018.8492968
- [3] ZHANG Fukai, RUI Ting, YANG Chengsong, et al. LAP-SLAM: a line-assisted point-based monocular VSLAM[J]. Electronics, 2019, 8(2): 243. DOI: 10.3390/electronics8020243
- [4] ZHOU Yijun, HAN Kang, LUO Chen, et al. SVO-PL: stereo visual odometry with fusion of points and line segments[C]//2018 IEEE International Conference on Mechatronics and Automation (ICMA). Changchun: IEEE, 2018; 900. DOI: 10.1109/ICMA.2018.8484479
- [5] ENGEL J, SCHPS T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM [C]// Computer Vision - ECCV. Cham: Springer, 2014; 834. DOI: 10.1007/978-3-319-10605-2_54
- [6] CVISIC I, CESIC J, MARKOVIC I, et al. SOFT-SLAM: computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles[J]. Journal of Field Robotics, 2018, 35(4): 578. DOI: 10.1002/rob.21762
- [7] MUR-ARTAI R, TARDÓS 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. DOI: 10.1109/TRO.2017.2705103
- [8] AGUDO A, MORENO-NOGUER F, CALVO B, et al. Real-time 3D reconstruction of non-rigid shapes with a single moving camera[J]. Computer Vision and Image Understanding, 2016, 153(12): 37. DOI: 10.1016/j.cviu.2016.05.004
- [9] AHUJA N A, SUBEDAR M, TICKOO O, et al. A factorization approach for enabling structure-from-motion/SLAM using integer arithmetic[C]//2017 IEEE International Conference on Computer Vision Workshops. Venice: IEEE, 2017; 554. DOI: 10.1109/ICCVW.2017.72
- [10] YU Chao, LIU Zuxin, LIU Xinjun, et al. DS-SLAM: a semantic

- visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 1168. DOI: 10.1109/IROS.2018.8593691
- [11] LI Xuanpeng, AO Huanxuan, BELAROUSSI R, et al. Fast semi-dense 3D semantic mapping with monocular visual SLAM [C]//2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Yokohama: IEEE, 2017: 385. DOI: 10.1109/ITSC.2017.8317942
- [12] CHEN Yongbin, HE Hanwu, CHEN Heen, et al. Improving registration of augmented reality by incorporating DCNNs into visual SLAM [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2018, 32(12): e1855022. DOI: 10.1142/S0218001418550224
- [13] CHEN L C, ZHU Yukun, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Computer Vision - ECCV. Cham: Springer, 2018: 833. DOI: 10.1007/978-3-030-01234-2_49
- [14] RAMÖA J G, ALEXANDRE L A, MOGO S. Real-time 3D door detection and classification on a low-power device[C]//2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). Ponta Delgada: IEEE, 2020: 96. DOI: 10.1109/ICARSC49921.2020.9096155
- [15] ZHU Yi, SAPRA K, REDA F A, et al. Improving semantic segmentation via video propagation and label relaxation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 8848. DOI: 10.1109/CVPR.2019.00906
- [16] ZHAO Hengshuang, SHI Jianping, QI Xiaojuan, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6230. DOI: 10.1109/CVPR.2017.660
- [17] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: real-time monocular visual SLAM with points and lines[C]//2017 IEEE International Conference on Robotics and Automation. Singapore: IEEE, 2017: 4503. DOI: 10.1109/ICRA.2017.7989522
- [18] LIN Weiyang, HU Jianjun, XU Hong, et al. Graph-based SLAM in indoor environment using corner feature from laser sensor [C]//2017 32nd Youth Academic Conference of Chinese Association of Automation. Hefei: IEEE, 2017: 1211. DOI: 10.1109/YAC.2017.7967597
- [19] JEONG J, CHO Y, KIM A. Road-SLAM: road marking based SLAM with lane-level accuracy [C]//2017 IEEE Intelligent Vehicles Symposium (IV). Los Angeles: IEEE, 2017: 1736. DOI: 10.1109/IVS.2017.7995958
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770. DOI: 10.1109/CVPR.2016.90
- [21] CHEN L C, ZHU Yukun, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Computer Vision - ECCV. Cham: Springer, 2018: 833. DOI: 10.1007/978-3-030-01234-2_49
- [22] ZHANG Qiao, CUI Zhipeng, NIU Xiaoguang, et al. Image segmentation with pyramid dilated convolution based on ResNet and U-Net [C]// Neural Information Processing. Cham: Springer, 2017: 364. DOI: 10.1007/978-3-319-70096-0_38
- [23] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(4): 834. DOI: 10.1109/TPAMI.2017.2699184
- [24] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [C]//2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). San Diego: IEEE, 2019: 409
- [25] SOQUET N, AUBERT D, HAUTIERE N. Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation [C]//2017 IEEE Intelligent Vehicles Symposium. Istanbul: IEEE, 2007: 160. DOI: 10.1109/IVS.2007.4290108
- [26] ZHOU Haoyin, ZHANG Tao, JAGADEESAN J. Re-weighting and 1-point RANSAC-based PnP solution to handle outliers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(12): 3022. DOI: 10.1109/TPAMI.2018.2871832
- [27] CORDTS M, OMRAN M, RAMOS S, et al. The Cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society. Las Vegas: IEEE, 2016: 3213. DOI: 10.1109/CVPR.2016.350
- [28] 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. DOI: 10.1177/0278364913491297

(编辑 苗秀芝)