

文章编号: 2095-2163(2021)03-0205-05

中图分类号: TN929.51

文献标志码: A

# 面向地下停车场的无人车视觉语义建图与定位

曹文冠, 黄孝慈, 舒方林, 孙 昊, 刘景锋

(上海工程技术大学, 上海 201620)

**摘要:** 无人车在狭窄、没有 GPS 信号的地下停车场中行驶, 精确的定位能力非常重要。传统的基于视觉的定位方法由于无纹理的区域、重复的结构和外观的变化而导致跟踪丢失。本文利用视觉语义信息构建停车场地图并对车辆进行定位。语义特征包括经常出现在停车场中的路标、停车线、减速带等, 与传统特征相比, 这些语义特征无纹理区域和外观的变化具有长期稳定和鲁棒性。文中采用单目相机对外界环境进行感知。在惯性测量单元 (IMU) 和车轮编码器的协助下, 提出一种生成全局语义地图的方法。此地图用于厘米级别上车辆定位。分析了该系统的准确性和召回率, 并与现有的方法进行了比较。最后, 通过自动停车的应用验证了该系统的实用性。

**关键词:** 视觉语义信息; 语义地图; 定位; 单目相机

## Visual semantic mapping and positioning of unmanned vehicle facing underground parking lot

CAO Wenguan, HUANG Xiaoci, SHU Fanglin, SUN Hao, LIU Jingfeng

(Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** For unmanned vehicles to drive in narrow underground parking lots without GPS signals, accurate positioning capabilities are very important. Traditional vision-based localization methods lead to loss of tracking due to untextured regions, repetitive structures and changes in appearance. This paper uses visual semantic information to construct a parking lot map and locate vehicles. Semantic features include road signs, parking lines, speed bumps, etc, which often appear in parking lots. Compared with traditional features, these semantic features have long-term stability and robustness for their textureless areas and appearance changes. The research uses monocular cameras to perceive the external environment. With the assistance of IMU (Inertial Measurement Unit) and wheel encoders, a method to generate a global semantic map is proposed. This map is used for vehicle positioning on the centimeter level. The accuracy and recall rate of the system are analyzed and compared with existing methods. Finally, the practicability of the system is verified through the application of automatic parking.

**[Key words]** visual semantic information; semantic map; location; monocular camera

## 0 引言

近年来,城市发展过程中,停车问题已经成为了一个亟待解决的研究热点。作为解决这一问题的有效方法,自动泊车技术逐渐受到多方关注。同时研究可知,该技术也是自动驾驶技术的重要组成部分<sup>[1]</sup>。自动泊车系统的主要功能包括:高精度定位、车位识别、轨迹规划与路径跟踪控制。系统运行时,车载环境感知传感器(如视觉传感器、激光雷达、超声波雷达等)开始工作,感知车辆周边的车位情况,并确定车位的相对位置信息;然后,由系统控制单元判定车位是否可用,且规划出合理可行的泊车路径;接着,车辆的控制权移交给泊车控制器,由

控制单元向车辆执行机构发出控制信号,与此同时则利用车辆传感器信息实时计算车辆的估计位姿,估计位姿与规划轨迹的误差作为跟踪控制的反馈信息实现轨迹准确跟踪,控制车辆按预定的轨迹泊入车位<sup>[2]</sup>。

目前的自动泊车功能难以在无 GPS 信号,环境复杂的地下停车场进行推广<sup>[3]</sup>。针对这一状况,文中采用了一种新的视觉方法特征,即语义特征。语义特征中包括路标、停车线和减速带,通常情况下这些指示标识都会在停车场出现。与传统的几何特征相比,上述语义特征具有长期稳定、鲁棒性强的特点<sup>[3]</sup>。在本次研究中,提出了一个基于语义特征的地图和定位系统,使车辆能够在停车场自动导航。

**基金项目:** 上海工程技术大学研究生科研创新项目(19KY0107)。

**作者简介:** 曹文冠(1995-),男,硕士研究生,主要研究方向:无人驾驶、机器视觉;黄孝慈(1980-),男,博士,副教授,主要研究方向:智能汽车驾驶技术;舒方林(1995-),男,硕士研究生,主要研究方向:计算机视觉;孙 昊(1995-),男,硕士研究生,主要研究方向:深度学习;刘景锋(1996-),男,硕士研究生,主要研究方向:重型车辆智能控制与网联技术。

收稿日期: 2020-10-28

## 1 语义地图构建与定位

系统采用相机和惯导提高定位精度。这是无人驾驶常见配置。IMU 和车轮编码器构成测程法,提供相对位姿,但存在累积误差。该框架由 2 部分组成,如图 1 所示。图 1 中,第一部分是建图,可为环境构建一个全局语义导图。神经网络检测语义特

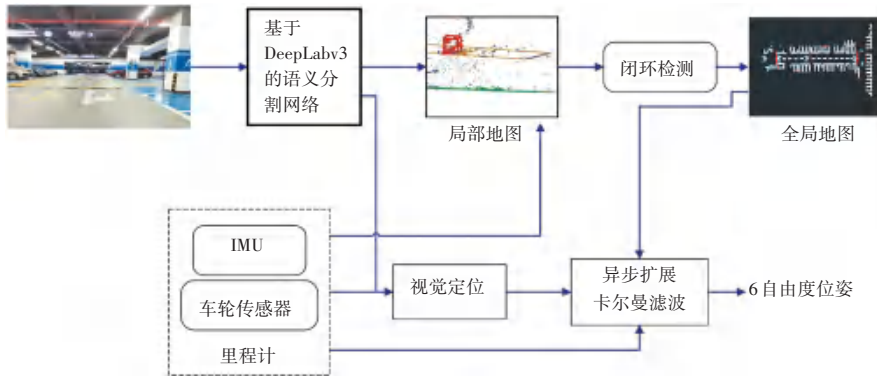


图 1 总体框架图

Fig. 1 General frame diagram

### 1.1 语义分割

本文采用卷积神经网络方法进行语义特征检测。有很多分割网络可以用于特征检测,参见文献[4-6]。在本文中,研究对 DeeLab3<sup>[7]</sup>进行了改进,在 DeepLabv3 的基础上增加了一个解码器模型用来增强物体边缘的分割效果。此外,本文引用 Xception 网络<sup>[8]</sup>中的深度可分卷积,应用在空间金字塔(Atrous Spatial Pyramid Pooling, ASPP)<sup>[9]</sup>和解码器模型上,提高了编码器-解码器网络的训练速度和鲁棒性。网络结构如图 2 所示。

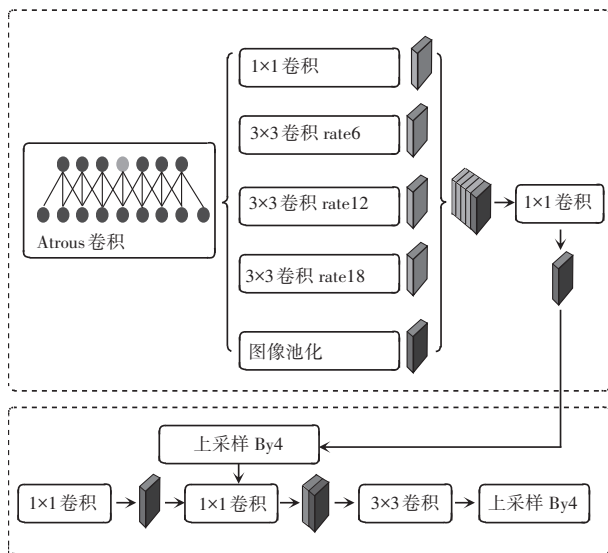


图 2 基于 DeeLab3 语义分割网络

Fig. 2 Semantic segmentation network based on DeeLab3

征,包括:车道、停车线、路标和减速带。第二部分是定位。与建图部分相同,从图像中提取语义特征,与前期构建的地图进行匹配,可以对车辆进行定位。在此基础上,为保证无纹理区域定位与建图的精度,利用(扩展卡尔曼滤波器(EKF)将视觉定位结果与里程计融合,保证了系统的平滑输出和无纹理区域的鲁棒性。

整个网络会对停车场获得的图像进行训练,将像素分为车道、停车线、路标、减速带、自由空间、障碍和墙壁。结果示例如图 3 所示。在这些类别中,停车线、路标和减速带是用于定位的独特而稳定的特征。停车线也用于停车位检测。自由空间和障碍物用于规划。



图 3 语义分割效果图

Fig. 3 Semantic segmentation renderings

### 1.2 局部地图构建

经过图像分割,将有用的特征(停车线、路标、减速带等)提升到三维空间,如式(1)所示:

$$\begin{bmatrix} \hat{e}_x^v \\ \hat{e}_y^v \\ \hat{e}_1 \end{bmatrix} \hat{u} = K_{imp}^{-1} \begin{bmatrix} \hat{e}_x^{imp} \\ \hat{e}_y^{imp} \\ \hat{e}_1 \end{bmatrix} \hat{u}, \quad (1)$$

转换为世界坐标:

$$\begin{bmatrix} \hat{e}_x^w \\ \hat{e}_y^w \\ \hat{e}_z^w \end{bmatrix} \hat{u} = R_o \begin{bmatrix} \hat{e}_x^v \\ \hat{e}_y^v \\ \hat{e}_1 \end{bmatrix} \hat{u} + t_o. \quad (2)$$

其中,  $[R_o, t_o]$  是里程计中的位姿。这些点聚合成一个局部地图,每隔 25 m 就保留一份局部地

图。局部地图的样本如图4所示。2个局部地图之间有重合的部分,可以进行合并生成全局地图。

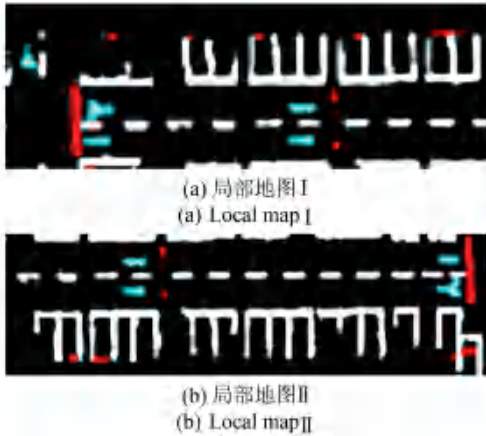


图4 局部地图  
Fig. 4 Local maps

### 1.3 闭环检测与全局优化

由于测程法存在长时间的漂移,本文通过检测环路闭合来消除漂移误差。对于最新的局部地图,文中将其与其他的局部地图进行比较。采用迭代最近点方法对2个局部地图进行匹配。如果2个局部地图匹配成功,则得到2个局部地图之间的相对位姿。该相对位姿将用于全局位姿图的优化以纠正漂移误差。

在循环检测后,对姿态图进行全局优化,消除累积漂移,保持整个地图的一致性。在该位姿图中,节点为每个局部地图的位姿,其中包含3个旋转轴,  $\mathbf{r} = [r_x \ r_y \ r_z]^T$ , 和平移矩阵  $\mathbf{t} = [t_x \ t_y \ t_z]^T$ 。局部地图存在2类边缘。一个是测程边缘,通过测程测量来约束2个连续的局部地图;另一个是闭环检测边缘,可约束局部地图闭环。位姿图优化可表示为以下代价函数:

$$\chi^* = \operatorname{argmin} \sum_t \|f(r_{t+1}, t_{t+1}, r_t, t_t) - z^o_{t,t+1}\|^2 + \sum_{i,j \in \zeta} \|f(r_i, t_i, r_j, t_j) - z^l_{i,j}\|^2. \quad (3)$$

其中,  $\mathbf{X} = [r_0, t_0, \dots, r_t, t_t]^T$  表示所有局部地图的位姿;  $z^o_{t,t+1}$  表示测程法中局部地图  $t$  与  $t+1$  的相对位姿;  $\zeta$  表示所有闭环点对的集合;  $z^l_{i,j}$  表示闭环帧  $i$  与  $j$  之间的相对位姿。函数  $f(\cdot)$  计算2个局部地图之间的相对位姿,并采用高斯-牛顿法进行优化。

在对全局位姿图进行优化后,利用更新后的位姿将局部地图叠加在一起。这样,就生成了一致性的全局地图。处理后的地图效果见图5。

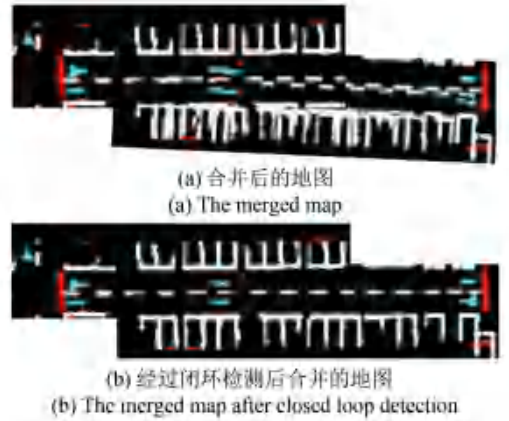


图5 处理后的地图效果

Fig. 5 Map effect after processing

### 1.4 定位

基于语义地图,当车辆再次到达该停车场时,可以对车辆进行定位。与建图过程类似,在图像上检测语义特征并将其提升到车辆坐标中,然后通过地图匹配当前特征点来估计车辆位姿,其对应的数学公式可写为:

$$\mathbf{r}^*, \mathbf{t}^* = \operatorname{argmin} \sum_{k \in S} \left\| \begin{matrix} \hat{\mathbf{e}}_k^v \\ \hat{\mathbf{e}}_k^u \\ \mathbf{0} \end{matrix} R(\mathbf{r}) + \mathbf{t} - \begin{matrix} \hat{\mathbf{e}}_k^w \\ \hat{\mathbf{e}}_k^w \\ \hat{\mathbf{e}}_k^w \end{matrix} \right\|^2. \quad (4)$$

其中,  $\mathbf{r}, \mathbf{t}$  分别表示当前坐标系的三维旋转平移向量;  $S$  表示当前特征点的集合;  $[\ x_k^v \ y_k^v \ 0 \ ]^T$  表示车辆坐标下的当前特征;  $[\ x_k^w \ y_k^w \ z_k^w \ ]^T$  表示特征在全球坐标下的估计点。

## 2 实验与结果分析

### 2.1 系统硬件介绍

本文实验结果均在 NVIDIA GeForce GTX1050Ti 显卡, Intel Core i5-8300H 处理器、Ubuntu16.04 操作系统上。实验装置设备如图6所示。

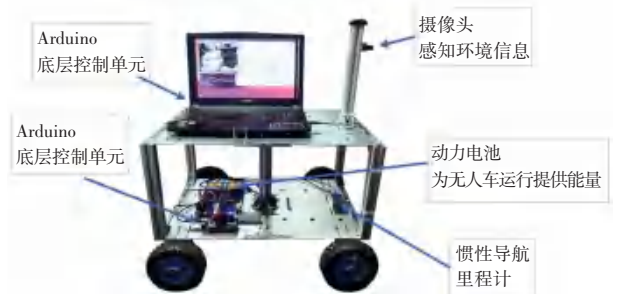


图6 无人车实验平台

Fig. 6 Unmanned vehicle experimental platform

由于只采用单目相机进行环境感知,因此选择视角较大的摄像头作为实验传感器,摄像头的型号

为林柏视 F.MONCAN-S970。该摄像头支持 3 种不同的分辨率和帧率,分辨率可达 1080 P,最高帧率可达 60 帧/s,满足无人车自动泊车场景下的需求。

## 2.2 建图精度评估

在地图度量评价中,文中选择了一个形状规则的停车场,便于仿真中测量停车场的真实值。无人车沿着一条 T 形路径运行,然后返回到起点。在这种情况下会发生循环闭包。总长度为 34 m。本文提出的方案与 IMU 和车轮编码器组成的测程法,以及 ORB-SLAM3<sup>[10]</sup>进行了对比。

绝对轨迹误差见表 1。由表 1 可以看出,使用视觉方法的精度明显优于纯测程法。由于测量噪声的存在,测程法不可避免地会产生漂移。本文的方案与 ORB-SLAM3 都通过使用视觉特征的精确环路检测来补偿累积误差,但不同之处则在于,ORB-SLAM3 使用几何特征,本文提出的方案使用语义特征。分析可知,本文提出的方法明显优于 ORB-SLAM3。

表 1 建图精度评估结果

Tab. 1 Evaluation results of mapping accuracy

方法	RMSE	Max	NEES/ %
测程法	9.26	19.23	2.52
ORB-SLAM3	6.73	9.45	1.86
本文的方法	4.22	8.34	1.53

表 1 中,  $RMSE$  是均方根误差。  $NEES$  是归一化估计误差的平方,等于  $RMSE / \text{总长度}$ 。

## 2.3 实车验证

本文实车实验是在上海工程技术大学现代交通工程中心地下停车场进行的。为了验证本文提出的位姿估计方法的有效性,在泊车试验区域建立全局坐标,通过全站仪测得车辆的实际位姿。

首先对停车场部分区域进行地图构建,然后使用研究构建的地图进行定位。本次研究中构建的地图如图 7 所示。



图 7 实车构建的全局地图

Fig. 7 Global map constructed by real vehicles

接下来,对本文提出的语义特征位姿估计方法进行验证,分别对车辆的平移误差和旋转误差进行

试验。世界坐标系原点与车辆初始位置的后轴中点重合,  $x$  轴沿车辆纵轴线向前,  $y$  轴指向车辆的正左方向,试验车速控制在 10 km/h 以内,其结果见表 2。

表 2 实车位姿误差对比试验结果

Tab. 2 Comparison test results of actual vehicle position and attitude error

方法	平移误差/m			旋转误差/rad		
	$x$	$y$	$z$	$yaw$	$pitch$	$roll$
Orb-slam3	-0.106	0.085	0.042	4.39	1.73	6.22
本文的方案	-0.072	0.056	0.036	3.25	1.21	4.03

文中将这 2 种估计方法得到的相机运动轨迹位姿与本文测得的实轨迹位姿进行比较,从旋转角度误差及平移角度误差两个维度,直观地比较这些方法的定位精度。分析后可以发现,本文提出的方法在平移和旋转两方面的能力皆优于  $orb - slam3$  的方法。可见在后续的车辆轨迹跟踪控制中可以使用本文提出的方法进行地图构建与估计车辆位姿。

## 3 结束语

在本文中,研究提出了一个基于视觉语义信息的建图与定位方案,利用鲁棒的语义特征来帮助车辆在停车场导航。本文提出的算法仅在前后端引入了机器学习技术,然而并没有将所有环节融入到机器学习中,例如回环检测和位姿估计都是 SLAM 算法的重要一环,在这 2 方面加入机器学习策略,可能会进一步提高建图与定位的精确度,并且能够降低 SLAM 算法对精密传感器的依赖。同时由于本次研究使用的语义特征的特殊性,文中提出的系统目前只适用于停车场。在未来,则将专注于通用的语义特征的研究,并为更多的场景推广本文的方案。

## 参考文献

- [1] 林泽. 面向自动泊车环境的视觉 SLAM 技术研究[D]. 长春:吉林大学,2020.
- [2] 王成雨. 基于视觉环境感知技术的自动泊车系统研究[D]. 镇江:江苏大学,2019.
- [3] 王鹏. 语义地图及其关键技术研究[D]. 深圳:中国科学院大学(中国科学院深圳先进技术研究院),2020.
- [4] LI Xia, ZHONG Zhisheng, WU Jianlong, et al. Expectation-maximization attention networks for semantic segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 9167-9176.