

doi:10.19306/j.cnki.2095-8110.2021.06.001

基于多模态感知与融合的无人车韧性导航系统

仲训昱, 武东杰, 陈登龙, 庄明溪, 吴汶鸿, 彭侠夫

(厦门大学航空航天学院, 厦门 361005)

摘要:针对野外复杂环境下的无人车自主导航需要,建立了一种基于多源融合定位、语义建图与运动规划的智能导航系统。首先,针对 IMU、轮式里程计、视觉 SLAM 与激光雷达 SLAM 等测量子系统,设计了误差状态扩展卡尔曼滤波器进行融合定位。其次,基于改进的 CNN 语义分割网络生成环境的语义图像,与 3D 激光雷达点云融合,并使用最大概率更新算法构建语义 3D 地图。接着,在语义和几何信息投影获得可通行性代价的基础上,提出了一种语义动态窗口的局部路径规划方法。最后,将以上感知、定位与规划方法整合成完整的智能导航系统,在城市与野外典型场景的测试中,相对定位误差小于 $0.4\%D$,具备一定的韧性导航定位和智能感知规划能力。

关键词:多模态感知;语义 3D 地图;融合定位;运动规划;误差状态扩展卡尔曼滤波;地面无人车

中图分类号: TP391

文献标志码: A

文章编号: 2095-8110(2021)06-0001-12

Flexible Navigation System for Unmanned Ground Vehicle Based on Multi-modal Sensing and Fusion

ZHONG Xun-yu, WU Dong-jie, CHEN Deng-long, ZHUANG Ming-xi, WU Wen-hong, PENG Xia-fu

(School of Aerospace Engineering, Xiamen University, Xiamen 361005, China)

Abstract: An intelligent navigation system based on multi-source fusion positioning, semantic mapping and motion planning is established to meet the needs of autonomous navigation of unmanned ground vehicles in complex field environments. Firstly, the measurement subsystems, such as IMU, wheeled odometer, visual SLAM and lidar SLAM, are fused for localization by error state extended Kalman filter. Secondly, the semantic image of the environment is generated based on the improved CNN semantic segmentation network, which is fused with the 3D lidar point cloud, and the semantic 3D map is constructed using the maximum probability update algorithm. Then, based on the passage cost computed by the projection of semantic and geometric information, a local path planning method based on semantic dynamic window is proposed. Finally, the above sensing, positioning and planning methods are integrated into a complete intelligent navigation system. In the tests of typical urban and wild scenes, the relative positioning error is less than $0.4\%D$, and it has a certain ability of tough navigation positioning and intelligent perception planning.

Key words: Multi-modal sensing; Semantic 3D map; Fusion localization; Motion planning; Error State Expanded Kalman Filtering; Unmanned ground vehicle

收稿日期: 2021-07-15; 修订日期: 2021-08-01

基金项目: 装备预先研究基金(61405180205)

作者简介: 仲训昱(1980-), 男, 博士, 主要从事多模态融合环境感知及智能导航方面的研究。

0 引言

在卫星拒止、物理空间变化等复杂环境下,为了使地面无人车具备全天候、全时域的韧性导航能力,需要解决基于视觉、激光雷达、惯性测量单元(Inertial Measurement Unit, IMU)、轮式里程计及磁力计等多模态环境感知与信息获取、基于信息融合的最优位姿估计与导航规划等问题^[1-2]。

针对最优位姿估计问题(基于微机电系统低精度 IMU),国内外学者主要开展了基于卡尔曼滤波的融合定位方法研究。其中扩展卡尔曼滤波(Extended Kalman Filter, EKF)是低延迟、高效状态估计的典型选择,已用于履带机器人的 IMU 和轮式里程计融合定位(精度达 $1.2\%D$)^[3],还用于美国国防部高级研究计划局(Defense Advanced Research Projects Agency, DARPA) LS3 四足机器人的 IMU、相机、足式里程计融合定位(精度达 $1.0\%D$)^[4],以及无人机组合导航^[5]等。近年,误差状态 EKF(Error State EKF, ESEKF)和不变 EKF(Invariant EKF, IEKF)凭借其在提高精度、收敛性及实时性方面的优势,在多源融合定位中不断得到应用^[6-8]。

针对导航规划(路径规划或运动规划)问题,在以往的研究中提出了基于环境的先验 3D 几何信息以选择最佳通行区域^[9],但在越野环境中如果只依赖几何信息则会导致令人失望的结果,例如可通行的草丛或细小植物等都会被认为是不可通过的障碍物,这显然不是类似人类的感知行为。近年来,随着深度学习技术的发展,文献[10]提出了使用语义分割实时构建环境的 2.5D 语义地图,同时考虑

环境的语义和几何信息使得无人车在导航过程中具备类似人的感知能力,例如可以自主选择是否穿越草丛行驶。但是 2.5D 语义地图对环境的表示能力不足,因为无人车在感知最真实的 3D 环境时,仍然需要重建完整的 3D 环境模型;文献[9]使用的局部规划器过于简单,无法满足复杂场景;并且语义分割模型的速度也有待提高,这些缺陷极大地限制了无人车在未知越野环境中的自主导航能力。

为了解决这些问题,本文提出了一种新颖的无人车自主导航系统。首先,为了尽量提高语义分割的速度,对实时网络 LedNet^[11]进行裁剪以加快模型推断速度,这样可以进一步提高后续环境模型构建的实时性。其次,使用了一种相较于传统贝叶斯算法^[12]速度更快的局部语义 3D 地图构建算法,并结合语义和几何信息实时投影生成可通行性代价地图,为规划提供必要的实时性条件。最后,改进了传统动态窗口(Dynamic Window Approach, DWA)局部规划算法^[13],为其增加一个语义约束,使得 DWA 算法无需全局规划器的指导即可具有实时感知前方最低通行难度区域的能力。

1 基于滤波的融合定位

1.1 融合框架与子系统设计

基于滤波的多模态信息融合位姿估计的总体框架如图 1 所示,包括四部分:1)具备多种信号感知能力的传感器集合;2)单一模态位姿估计子系统;3)基于 ESEKF 的位姿估计器;4)用于提高韧性的多模态测量信息管理器。

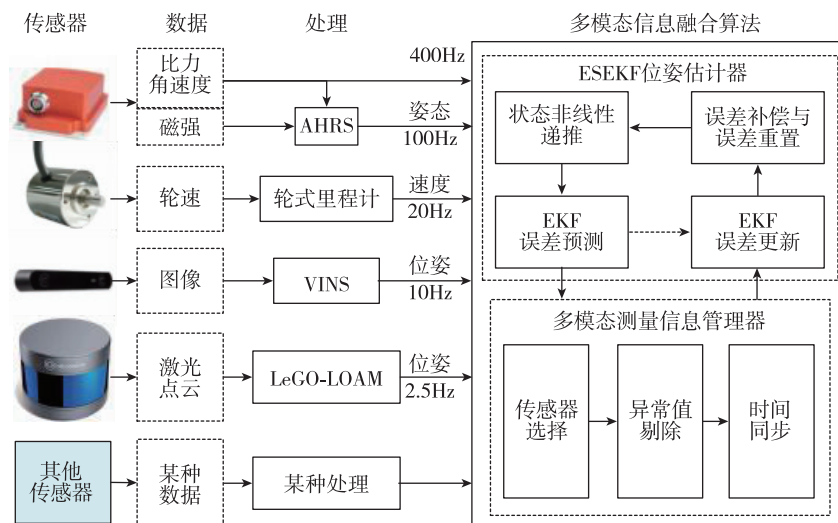


图 1 基于 ESEKF 的多模态信息融合总体方案图

Fig. 1 Overall scheme of multi-modal information fusion based on ESEKF

在各位姿估计子系统中,IMU 和旋转式编码器可感知无人车的运动学或动力学状态(如角速度、加速度、轮速等);在磁场均匀不受干扰的条件下,IMU 中的磁力计可以提供一个绝对的偏航角参考;相机和激光雷达等传感器可以采用同步定位与构图(Simultaneous Localization and Mapping, SLAM)算法(如

VINS^[14]、LeGO-LOAM^[15]等),获得相对位姿信息及速度测量信息。

1.2 ESEKF 多模态信息融合位姿估计器

设计的 ESEKF 融合算法实现流程如图 2 所示,主要包括全状态预测、误差状态 EKF 预测、误差状态 EKF 更新和误差补偿与重置等模块。

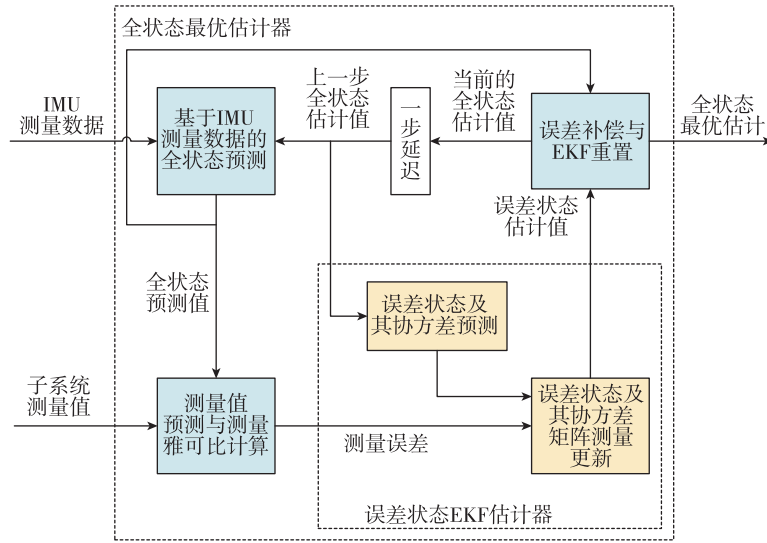


图 2 基于 ESEKF 的多模态融合定位架构

Fig. 2 Multi-modal fusion positioning architecture based on ESEKF

全状态 $x \in \mathbf{R}^{16}$ 定义为

$$x = [p^T \ v^T \ q^T \ a_b^T \ \omega_b^T]^T \quad (1)$$

其中, $p \in \mathbf{R}^3$ 、 $v \in \mathbf{R}^3$ 、 $q \in \mathbf{S}^3$ 、 $a_b \in \mathbf{R}^3$ 和 $\omega_b \in \mathbf{R}^3$ 分别表示无人车的位置、速度、姿态四元数、加速度计的测量偏差和陀螺仪的测量偏差;误差状态 $\delta x \in \mathbf{R}^{15}$ 定义为系统真实状态 x 与其估计值 \hat{x} 之间的差,向量表示为

$$\delta x = [\delta p^T \ \delta v^T \ \delta \theta^T \ \delta a_b^T \ \delta \omega_b^T]^T \quad (2)$$

其中,位置误差 $\delta p \in \mathbf{R}^3$ 、速度误差 $\delta v \in \mathbf{R}^3$ 、加速度计测量偏差误差 $\delta a_b \in \mathbf{R}^3$ 和陀螺仪测量偏差误差 $\delta \omega_b \in \mathbf{R}^3$ 由线性空间减法定义;姿态误差 $\delta \theta \in \mathbf{R}^3$ 由四元数乘法定义^[16]。

ESEKF 多模态信息融合位姿估计器的思路是^[17]:1)利用高频 IMU 数据预测当前时刻的全状态和误差值;2)基于 EKF 算法,利用多模态子系统的测量值获得系统误差的最优估计值;3)将误差估计值和全状态预测值合成,获得系统全状态的最优估计。

(1) 全状态预测估计过程

设 a_m 和 ω_m 分别为加速度计和陀螺仪的测量值,利用数值积分知识可得到全状态的预测方程,

设为

$$\hat{x}_{k+1} \leftarrow f(\hat{x}_k, u_k) \quad (3)$$

其中,下标 k 和 $k+1$ 分别表示上一时刻和当前时刻,输入 $u_k \triangleq [a_m^T \ \omega_m^T]^T$;当采用欧拉数值积分时, f 的具体表达式如下^[16]

$$\begin{aligned} p_{k+1} &\leftarrow p_k + v_k \Delta t \\ v_{k+1} &\leftarrow v_k + (R(a_m - a_b) + g) \Delta t \\ q_{k+1} &\leftarrow q_k \otimes q \{ (\omega_m - \omega_b) \Delta t \} \\ a_{b,k+1} &\leftarrow a_{b,k} \\ \omega_{b,k+1} &\leftarrow \omega_{b,k} \end{aligned} \quad (4)$$

其中, $R \in SO(3)$ 表示姿态的旋转矩阵; g 为重力加速度矢量; $q\{\}$ 表示将旋转向量转换为相应的四元数; \otimes 表示四元数乘法。

(2) 误差状态 EKF 估计过程

根据各变量的误差定义及其小量的特性,在参考点 $x = \hat{x}_k$ 处进行线性近似,得到误差状态运动的差分模型^[16]

$$\delta x_{k+1} = F_k \delta x_k + w_{F,k} \quad (5)$$

其中

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I} & \mathbf{I}\Delta t & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & -[\mathbf{R}(a_m - a_o)\Delta t]_{\times} & -\mathbf{R}\Delta t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & -\mathbf{R}\Delta t \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \Big|_{x=\hat{x}_k}$$

$$\mathbf{w}_{F,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_F)$$

其中, \mathbf{I} 为单位矩阵; $\mathbf{0}$ 为零矩阵; $[\]_{\times}$ 表示将向量转换为对应的反对称矩阵; $\mathbf{w}_{F,k}$ 为过程噪声, 其协方差设为 $\mathbf{Q}_{F,k}$ 。

当前时刻的全状态测量模型设为

$$\mathbf{y}_{k+1} = h(\mathbf{x}_{k+1}) + \mathbf{w}_{h,k+1} \quad (6)$$

其中, 函数 h 的具体形式由子系统的输出性质决定; $\mathbf{w}_{h,k+1}$ 表示测量噪声, 其协方差记为 $\mathbf{R}_{h,k+1}$ 。

根据 EKF 框架, 误差状态最优估计过程为^[16]

$$\begin{aligned} \delta \hat{\mathbf{x}}_{k+1} &\leftarrow \mathbf{F}_k \delta \hat{\mathbf{x}}_k \\ \mathbf{P}_{k+1} &\leftarrow \mathbf{F}_k \mathbf{P}_k \mathbf{F}_k^T + \mathbf{Q}_{F,k} \\ \mathbf{K}_{k+1} &\leftarrow \mathbf{P}_{k+1} \mathbf{H}_{k+1}^T (\mathbf{H}_{k+1} \mathbf{P}_{k+1} \mathbf{H}_{k+1}^T + \mathbf{R}_{h,k+1})^{-1} \\ \delta \hat{\mathbf{x}}_{k+1} &\leftarrow \delta \hat{\mathbf{x}}_{k+1} + \mathbf{K}_{k+1} (\mathbf{y}_{k+1} - h(\hat{\mathbf{x}}_{k+1})) \\ \mathbf{P}_{k+1} &\leftarrow (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \mathbf{P}_{k+1} \end{aligned} \quad (7)$$

其中, \mathbf{P} 为滤波协方差矩阵; \mathbf{H}_{k+1} 为测量雅可比矩阵, 计算公式如下

$$\mathbf{H}_{k+1} \triangleq \left. \frac{\partial h}{\partial \delta \mathbf{x}} \right|_{x=\hat{x}_{k+1}} = \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{x=\hat{x}_{k+1}} \left. \frac{\partial \mathbf{x}}{\partial \delta \mathbf{x}} \right|_{\delta \mathbf{x}=\mathbf{0}} \quad (8)$$

其中

$$\left. \frac{\partial \mathbf{x}}{\partial \delta \mathbf{x}} \right|_{\delta \mathbf{x}=\mathbf{0}} = \begin{bmatrix} \mathbf{I}_6 & & \\ & \mathbf{Q}_{\delta \theta} & \\ & & \mathbf{I}_9 \end{bmatrix}$$

$$\mathbf{Q}_{\delta \theta} = \frac{1}{2} \begin{bmatrix} -q_x & -q_y & -q_z \\ q_w & -q_z & q_y \\ q_z & q_w & -q_x \\ -q_y & q_x & q_w \end{bmatrix}$$

位置测量方程为

$$\mathbf{p}_{m,k+1} = \mathbf{p}_{k+1} + \mathbf{w}_{p,k+1} \quad (9)$$

其中, $\mathbf{p}_{m,k+1}$ 为测量值; $\mathbf{w}_{p,k+1}$ 为位置测量噪声; 测量雅可比为三维单位矩阵。

姿态测量方程为

$$\mathbf{q}_{m,k+1} = \mathbf{q}_{k+1} + \mathbf{w}_{q,k+1} \quad (10)$$

其中, $\mathbf{q}_{m,k+1}$ 为测量值; $\mathbf{w}_{q,k+1}$ 为姿态测量噪声; 测量雅可比为四维单位矩阵。

(3) 全状态更新估计过程

1) 误差补偿

利用误差的 EKF 最优估计值 $\delta \hat{\mathbf{x}}_{k+1}$ 对全状态

的预测值 $\hat{\mathbf{x}}_{k+1}$ 进行补偿, 便可得到全状态的最优估计值, 误差补偿公式为^[16]

$$\begin{aligned} \hat{\mathbf{p}}_{k+1} &\leftarrow \hat{\mathbf{p}}_{k+1} + \delta \hat{\mathbf{p}}_{k+1} \\ \hat{\mathbf{v}}_{k+1} &\leftarrow \hat{\mathbf{v}}_{k+1} + \delta \hat{\mathbf{v}}_{k+1} \\ \hat{\mathbf{q}}_{k+1} &\leftarrow \hat{\mathbf{q}}_{k+1} \otimes \mathbf{q} \{ \delta \hat{\theta}_{k+1} \} \\ \hat{\mathbf{a}}_{b,k+1} &\leftarrow \hat{\mathbf{a}}_{b,k+1} + \delta \hat{\mathbf{a}}_{b,k+1} \\ \hat{\boldsymbol{\omega}}_{b,k+1} &\leftarrow \hat{\boldsymbol{\omega}}_{b,k+1} + \delta \hat{\boldsymbol{\omega}}_{b+1} \end{aligned} \quad (11)$$

2) 误差重置

当完成对全状态预测值的补偿后, 需要对 EKF 误差状态估计器中的变量进行重置, 过程如下^[16]

$$\begin{aligned} \delta \hat{\mathbf{x}}_{k+1} &\leftarrow \mathbf{0} \\ \mathbf{P}_{k+1} &\leftarrow \mathbf{S}_{k+1} \mathbf{P}_{k+1} \mathbf{S}_{k+1}^T \end{aligned} \quad (12)$$

其中

$$\mathbf{S}_{k+1} = \text{diag} \left(\mathbf{I}, \mathbf{I}, \mathbf{I} + \left[\frac{1}{2} \delta \hat{\theta}_{k+1} \right]_{\times}, \mathbf{I}, \mathbf{I} \right)$$

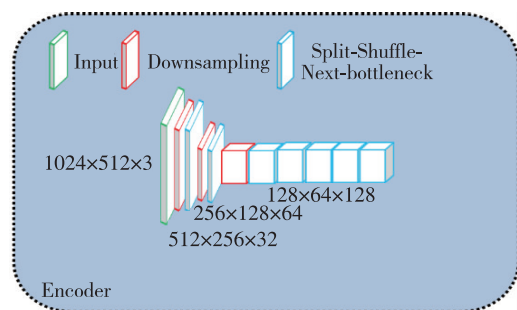
2 环境感知与语义 3D 地图构建

2.1 基于深度语义分割的环境感知

为提高实时性, 对 LedNet 编码器端进行修改, 得到计算速度更高的语义分割网络(如图 3(a)所示), 称为 MiniLedNet。MiniLedNet 在 LedNet 的基础上, 引入瓶颈型深度可分离卷积模块(SS-Next-bt, 如图 3(b)所示), 相较于 LedNet 的 SS-nbt 模块添加了具有扩张率的深度可分离卷积以及 MobileNext 提出来的 SandGlass 模块^[18]。由于 1×1 的点卷积对通道数进行了缩小导致部分信息的丢失, 所以又引入了 MobileNext 的线性瓶颈模块, 再将这三者组合得到 SS-Next-bt, 该模块可在高通量特征层中大大减少运算量, 同时仅小幅降低准确率, 从而对 LedNet 模型进行压缩以提高推断速度。

2.2 融合激光点云的语义 3D 地图

如图 4 所示, 双目相机 RGB 图像经 MiniLedNet



(a) MiniLedNet

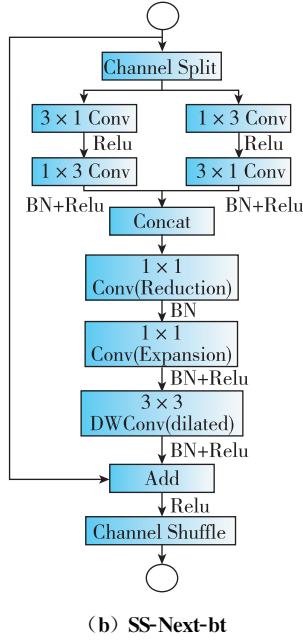


图 3 MiniLedNet 编码器及其 Split-Shuffle-Next-bottleneck 模块
Fig. 3 The encoder and its Split-Shuffle-Next-bottleneck module

得到语义图像,与 3D 激光雷达的点云融合得到语义点云,再结合融合定位信息,通过八叉树存储结构的占有更新、语义更新建立局部和全局语义 3D 地图^[19]。此外,为了获得更高精度的语义 3D 地图,通过语义标签分布修正和运动目标去除算法对地图进行纠错和修正,为地面无人车平台执行复杂任务提供准确的导航特征地图。

(1) 单帧语义融合

经外参标定对齐激光雷达和相机的坐标后,利用语义图像,将语义标签及其类别概率和 3D 点云进行关联^[20]。语义分割网络首先在一帧 RGB 图像 I 上对每个像素 (i, j) 分别计算其属于类别 k 的置信度 $C_{i,j}^k$, 然后使用 softmax 正则化函数将置信度转换为归一化的标签概率

$$p(k | I_{i,j}) = \frac{e^{C_{i,j}^k}}{\sum_{k \in \text{class}} e^{C_{i,j}^k}} \quad (13)$$

为了得到每个像素最可能的标签概率,取所有

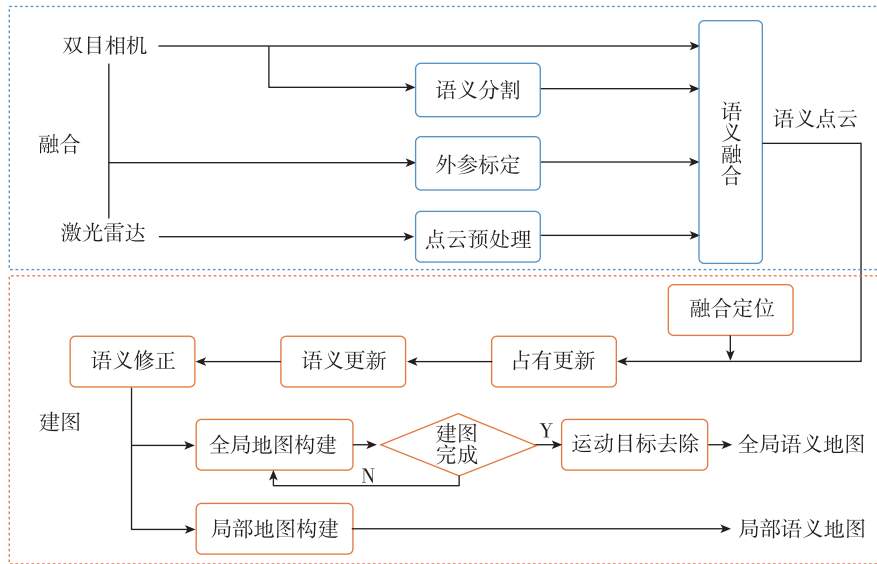


图 4 语义 3D 地图构建总流程图
Fig. 4 The building process of semantic 3D map

类别标签概率的最大值

$$P_{i,j} = \max[p(1 | I_{i,j}), p(2 | I_{i,j}), \dots, p(k | I_{i,j})] \quad (14)$$

在得到每个像素的语义标签和标签概率后,通过外参矩阵将其与对应的点云进行融合得到语义点云。

(2) 体素占有概率更新

当新的单帧融合后的语义点云插入到八叉树存储结构的地图时,八叉树内每个体素占有概率的更新公式如下^[21]

$$P(n | z_{1:t}) = \left[1 + \frac{1 - P(n | z_t)}{P(n | z_t)} \frac{1 - P(n | z_{1:t-1})}{P(n | z_{1:t-1})} \frac{P(n)}{1 - P(n)} \right]^{-1} \quad (15)$$

其中, $z_{1:t}$ 为第 1 ~ t 时刻所有的观测数据; $P(n | z_{1:t})$ 为第 n 个节点基于第 1 ~ t 时刻观测数据的后验概率; $P(n | z_t)$ 为第 n 个节点基于当前观测数据的占有概率; $P(n)$ 为节点 n 的先验概率(取 0.5)。

根据 logit 变换,将式(15)简化成如下形式

$$L(n | z_{1:t}) = L(n | z_{1:t-1}) + L(n | z_t) \quad (16)$$

(3) 体素语义信息更新

采用八叉树结构存储语义地图,每个语义点云用立方体栅格表示,其中每一个体素在多帧图像中都可能被观察到并可能被语义分割网络预测一个不同的语义类别,因此需要通过融合多个时刻观察到的语义来确定一个体素的最终类别。

为保证实时性(尤其当语义类别较多时),采用最大概率更新方法。体素 v 在 t 时刻的语义类别为 $L(v_t)$, 标签概率 $P(c | z_t)$ 依赖于当前体素测量值 z_t 和 $t-1$ 时刻的状态量 $P(c | z_{t-1})$, 则最大概率融合后的当前体素 v 的语义类别 $L(v_t)$ 的计算公式如下

$$L(v_t) = \begin{cases} L(v_{t-1}), & \text{if: } P(c | z_{t-1}) > P(c | z_t) \\ L(v_t), & \text{if: } P(c | z_{t-1}) \leq P(c | z_t) \end{cases} \quad (17)$$

最大概率融合后的当前体素 v 的标签概率 $P(c | z_t)$ 的计算如下

$$P(c | z_t) = \begin{cases} \frac{P(c | z_{t-1}) + P(c | z_t)}{2}, & \text{if: } L(v_{t-1}) = L(v_t) \\ \max\{P(c | z_{t-1}), P(c | z_t)\} * \epsilon, & \text{if: } L(v_{t-1}) \neq L(v_t) \end{cases} \quad (18)$$

其中, $\epsilon \in (0, 1)$ 表示 t 时刻体素 v 的语义标签概率对 $t-1$ 时刻体素 v 语义标签概率的修正因子, 取 $\epsilon = 0.9$ 。

(4) 体素语义修正

在更新完语义信息后,由于外参标定和语义分割存在误差,所以导致个别体素语义类别分配错误,因此采用空间推理法减少这些体素。首先,检测体素的 K 近邻体素,并计算这些体素和体素之间的平均距离。随后,如果平均距离小于某个阈值,则它对应于分割良好的体素,否则它被认为是噪声体素。表达式如下

$$P^v = \begin{cases} 1, & \text{if: } \text{Avg} | P^v - P_{mn}^v | < \delta_d \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

其中, 1 和 0 分别表示好的分割和错误的分割。另外, P_{mn}^v 对应于 P^v 的 K 近邻体素, δ_d 是相应的阈值, K 取 10, δ_d 取 1m。

(5) 运动目标去除

当语义地图完成时,许多地图点是由移动的物体产生的,这使得地图变得混乱,必须消除这些地图点。采用一种具有低参数特征的基于密度的噪声应用空间聚类(Density-Based Spatial Clustering

of Applications with Noise, DBSCAN)算法,基于密度对体素进行分组。然而,其缺点是生成的映射的大小,存储效率显著降低,因此需要长的计算时间。为了解决这个问题,将三维体素表示转换为二维栅格表示。通过将所有三维体素投影到相同的 Z 来实现,定义如下

$$M_{2D}^v(x, y) = M_{3D}^v(x, y, z = 0) \quad (20)$$

其中, M_{2D}^v 和 M_{3D}^v 分别代表 2D 栅格图和 3D 体素图。随后,在 2D 栅格上进行 DBSCAN,基于空间分布创建聚类。给定聚类群,应用几个约束以区分移动和静态物体。对应于移动物体的聚类包含较多体素,并且与静止同类物体相比,长度相对较长。因此,基于体素数量和长度设置约束对运动物体和静止物体进行分类。如下所示

$$C_i = \begin{cases} \text{stationary}, & \text{if: } (D_i < \eta_d) \wedge (L_i < \eta_l) \\ \text{moving}, & \text{otherwise} \end{cases} \quad (21)$$

其中, D_i 为体素数量; L_i 为 i_{th} 聚类的长度; η_d 和 η_l 为相应的阈值。

3 基于语义 3D 地图的运动规划

如图 5 所示,基于以上方法构建以无人车为中心的局部语义 3D 地图后,通过语义和高度投影得到可通行性代价地图;基于投影得到 2D 地图后,运动规划器能够感知可通过区域(道路、草丛等)、不可通过区域(树木和其他障碍等)以及高度信息,从而选择一条可通行路径。

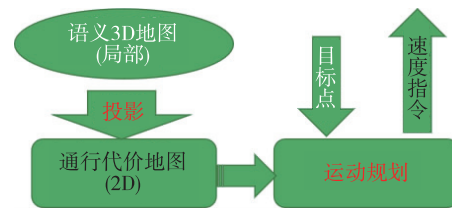


图 5 基于语义地图的运动规划

Fig. 5 Motion planning based on semantic map

3.1 语义和高度投影

如图 5 所示,通过将局部语义 3D 地图投影成 2D 代价地图的方式,为无人车导航提供运动规划的代价地图。地图中可通过区域如道路、草丛等通行成本分别设为 1 和 10,高的绿植的成本设为 50,不可通过的障碍物(岩石等)成本设为 200;另外,为了弥补语义分割的不足(误分类),同时对 3D 语义栅格的高度也进行最大值投影。语义和高度投影如

图 6 所示, $Z_0 < \{z_0, z_1, \dots, z_n\} < Z_0 + H$ 为 (x_i, y_i) 垂直方向上各栅格的高度信息 (Z_0 为无人车中心点对应栅格的高度, H 为无人车的车身高度), 其对应的语义成本为 $\{s_0, s_1, \dots, s_n\}$, 投影到 2D 代价地图后, 对应栅格 (x_i, y_i) 的语义成本 $s = s_{\max}$ (取 $\{s_0, s_1, \dots, s_n\}$ 中最大值), 高度取 $z = z_{\max}$ (取 $\{z_0, z_1, \dots, z_n\}$ 中最大值)。

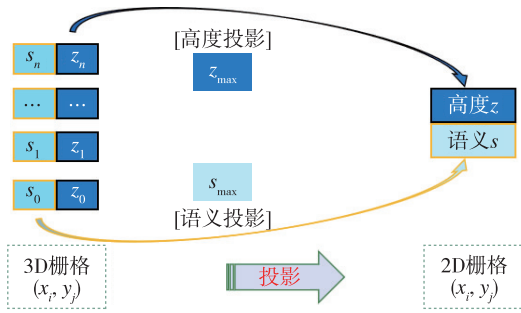


图 6 语义和高度投影示意图

Fig. 6 Projection of semantic and height information

3.2 单条路径通行代价计算

在投影得到局部 2D 代价地图和某一路径后, 先在该地图上计算一条采样路径中每个路径点在语义代价地图中对应的语义成本, 并结合高度值进行修正得到总的通行代价; 再将一条中间路径按照无人车的宽度复制 6 条平移轨迹, 分别计算每条平移轨迹的通行代价, 并将 7 条轨迹的通行代价累加起来作为该路径的总通行代价。单条路径通行代价的计算原理如图 7 所示。

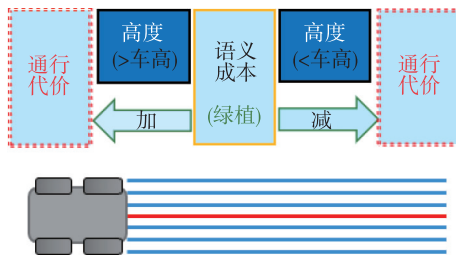


图 7 通行代价计算

Fig. 7 Passage cost calculation principle

3.3 基于 DWA 的运动规划

采用 DWA 实现局部路径规划, 主要流程如下:

- 1) 在无人车运动空间进行速度离散采样 ($\Delta x, \Delta y, \Delta \theta$);
- 2) 对每一个采样速度执行行进模拟, 检测使用该采样速度移动一段时间会发生什么;

3) 评价行进模拟中每个前行轨迹, 评价准则有: 靠近障碍物、靠近目标、贴近全局路径、丢弃非法轨迹 (如靠近障碍物的轨迹);

4) 挑出得分最高的轨迹 (路径) 并发送对应的速度给运动控制器;

5) 重复上面步骤。

在每个采样周期内的线速度 v 和角速度 ω 受到电机最大最小速度和加减速度性能的限制, 即

$$v_{\min_max} = \{v \in [v_{\min}, v_{\max}], \omega \in [\omega_{\min}, \omega_{\max}]\}$$

$$v_{acc} = \{v \in [v_c - \dot{v}_d t_\Delta, v_c + \dot{v}_a t_\Delta], \omega \in [\omega_c - \dot{\omega}_d t_\Delta, \omega_c + \dot{\omega}_a t_\Delta]\} \quad (22)$$

$$v_{limit} = \{(v, \omega) \in v_{\min_max} \cap (v, \omega) \in v_{acc}\}$$

每个时刻无人车在全局坐标系下的平面位置 (x, y) 分别等于上一时刻的位置加上当前 Δt 时间内的位移。DWA 算法在每个周期内会根据运动学模型和速度、加速度等限制进行路径采样。为了在投影的 2D 代价地图中计算通行成本, 修改了 DWA 算法, 不同之处主要为: 增加了语义代价函数; 在固定窗口内采样等长度路径。固定窗口参数: 固定车辆速度为 v , 角速度范围为 $[-\omega, \omega]$, 角速度增量为 Δyaw , 在每次采样周期内采样 n 条长度为 L 的轨迹, 如图 8 所示。

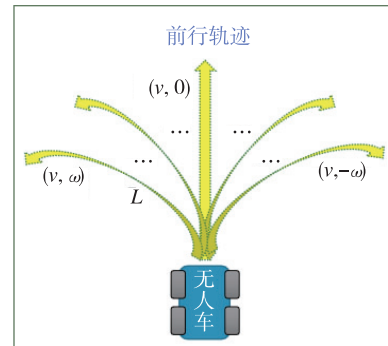


图 8 局部语义规划器采样空间

Fig. 8 The sampling space of the local semantic planner

4 试验验证

验证平台如图 9 所示, 为四轮驱动的无人地面车辆, 自主导航系统的主要传感器有安装在车辆顶部的 3D 激光雷达 (Robosense-16E)、安装在车辆前方的双目相机 (ZED2)、安装在车辆中心的 IMU (Xsens MTi-30)、安装在驱动电机后的编码器, 以及 GNSS 模块 (司南 M600 mini GNSS)。

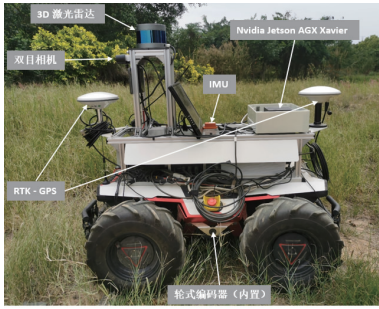


图9 地面无人平台 SCOUT

Fig. 9 Unmanned ground platform SCOUT

4.1 融合定位试验结果

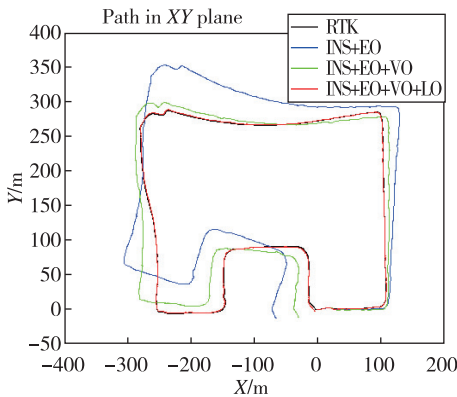
图1融合定位系统中,用INS表示惯性定位,EO为轮式里程计(编码器)定位,VO为视觉定位(VINS),LO为激光雷达定位(LeGO-LOAM)。

(1)街道场景(校园)

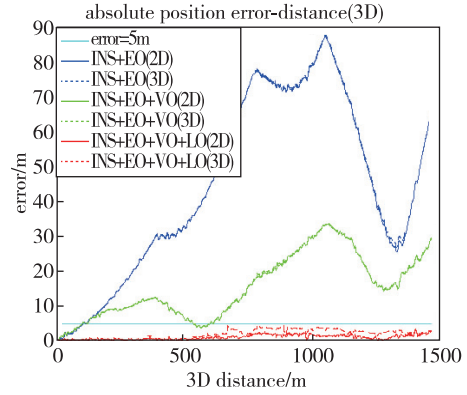
路程 $D \approx 1460\text{m}$,定位结果如图10所示。采用INS/EO/VO/LO多传感器融合定位,其平均误差(2D/3D)为 $2.210\text{m}/2.916\text{m}$,相对定位误差 $< 0.2\% D$ 。



(a)实景图



(b) 定位轨迹/路径(平面)



(c) 全局定位误差曲线

图10 街道场景(校园)定位结果(参考值:RTK)

Fig. 10 Positioning results in street scene (campus)

(2)树林场景(夜晚)

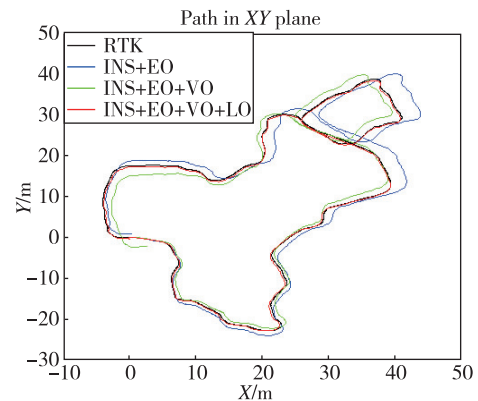
路程 $D \approx 242\text{m}$,定位结果如图11所示。采用INS/EO/VO/LO多传感器融合定位,其平均误差(2D/3D)为 $0.289\text{m}/0.357\text{m}$,相对定位误差 $< 0.2\% D$ 。

(3)野外场景

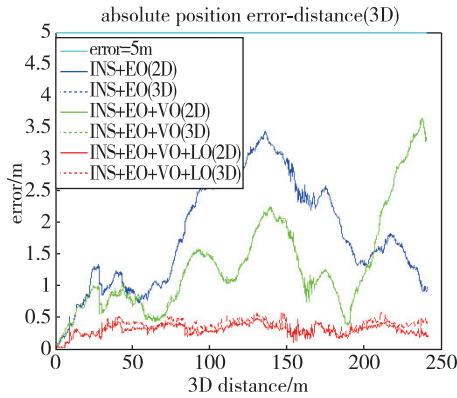
路程 $D \approx 175\text{m}$,定位结果如图12所示。采用



(a)实景图



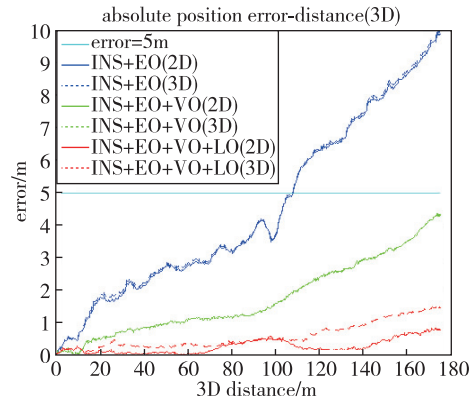
(b) 定位轨迹/路径(平面)



(c) 全局定位误差曲线

图 11 树林场景(夜晚)定位结果(参考值:RTK)

Fig. 11 Positioning results in forest scene (night)



(c) 全局定位误差曲线

图 12 野外场景(山区)定位结果(参考值:RTK)

Fig. 12 Positioning results in wild scene (mountainous area)

INS/EO/VO/LO 多传感器融合定位,其平均误差(2D/3D)为 0.304m/0.650m,相对定位误差 $<0.4\%D$ 。

4.2 环境感知试验结果

(1)语义分割

真实环境测试(厦大翔安校区及荒地)的语义分割结果示例如图 13 所示,第 1 列为原始图像,第 2 列为 MiniLedNet 语义分割结果;其中第 1 行为校

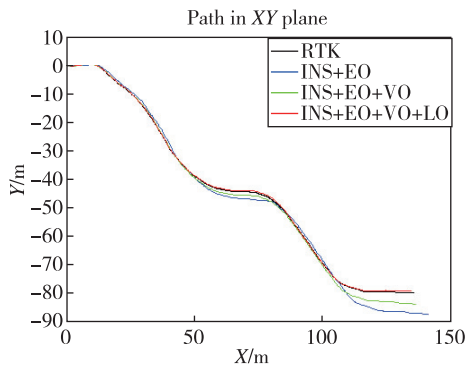


图 13 真实环境语义分割测试结果(■道路 ■人行道 ■建筑 ■路灯(杆类) ■交通标志 ■天空 ■高植 ■草丛 ■未知)

Fig. 13 Real environment semantic segmentation results



(a)实景图



(b) 定位轨迹/路径(平面)

园环境,第 2 行为未开发的荒地。MiniLedNet 能对道路、人行道、建筑物、路灯/杆类、交通标志、行人、汽车、天空、高植和草丛等物体进行分割。

在 Cityscapes 数据集测试中,MiniLedNet 相较于 LedNet 速度提高了 31.2%,达到了 20Hz 的处理速度,精度损失只有 3.1%;在 Freiburg 森林数据集测试中,MiniLedNet 模型对 12 类的平均交并比(Mean Intersection over Union, MIoU)达到 79.32%,高于 CNNs-FCN 和 Dark-Fcn。

(2)地图构建

无人车在校区街道场景下行驶,建立的局部地图如图 14 和图 15 所示。通过点云和语义信息的融合、更新,实现了路面、草丛、高植、建筑、汽车准确感知与 3D 建模,建图速度达到 15Hz。

无人车在荒地和树林环境下行驶,建立的局部地图如图 16 和图 17 所示,实现了路面、草丛、高植、岩石的准确感知与 3D 建模,建图速度达到 15Hz。其中过高的草丛在语义分割中被识别为高植。

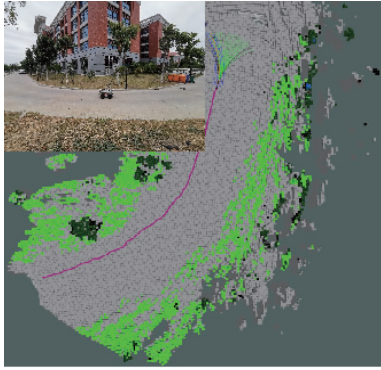


图 14 城区环境局部语义 3D 地图

Fig. 14 Local semantic 3D map of urban environment I

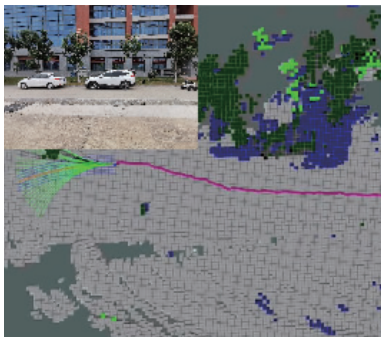


图 15 城区环境局部语义 3D 地图

Fig. 15 Local semantic 3D map of urban environment II

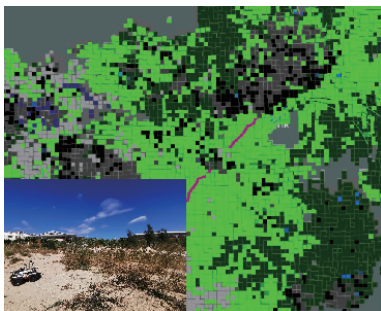


图 16 荒地局部语义 3D 地图

Fig. 16 Local semantic 3D map of wasteland

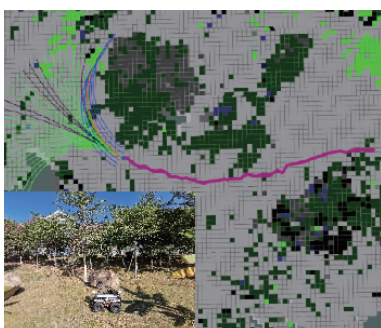


图 17 树林局部语义 3D 地图

Fig. 17 Local semantic 3D map of forest

4.3 运动规划试验结果

(1) 路沿避让测试

本测试中路沿区域也被误分类为地面,路径规划对于高度大于指定高度的体素(非草丛或高植),认其为障碍物。本试验高度设置为 0.24m,测试结果如图 18 所示,无人车能够绕行 90°到达目标点,没有与高的路沿发生碰撞(如果不考虑高度则将发生碰撞)。

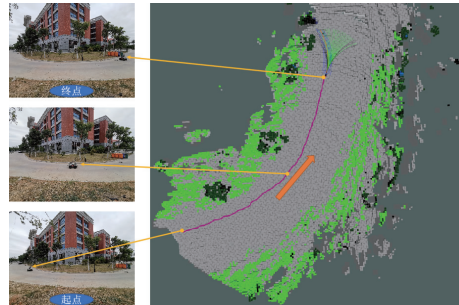


图 18 路沿避让测试结果(轨迹)

Fig. 18 Roadside avoidance test results (track)

(2) 岩石避让测试

测试中岩石被当作未知类别(默认为障碍物),语义地图中岩石可能是黑色或深绿色。如图 19 所示,无人车遇到大的石头时,能够成功绕行。

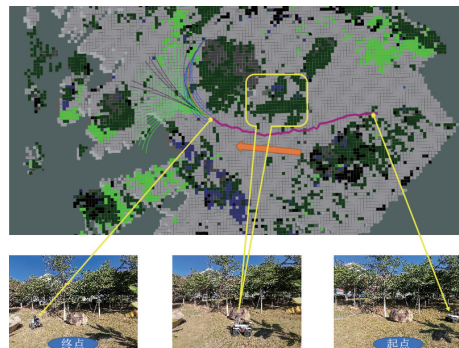


图 19 岩石避让路径规划结果

Fig. 19 Rock avoidance path planning results

(3) 草丛可通行测试

如图 20 所示,无人车能够在有路面和草地的环境中选择最优路径行驶达到目标点,草丛尽管有一定高度但不会造成绕行行为。

(4) 高植通行对比测试

对于高度比车高低的绿植,路径规划器会根据高度信息对高植的语义代价进行修正,得到行驶轨迹如图 21 所示,可以看出高度较低的绿植(主要是在语义分割中误认为高植的草丛)并没有影响通行性。

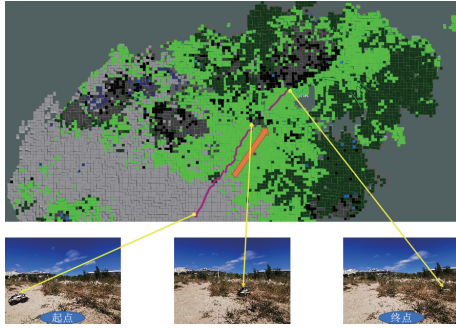


图 20 穿越草丛的通行测试(轨迹)

Fig. 20 Passage test through grass (track)

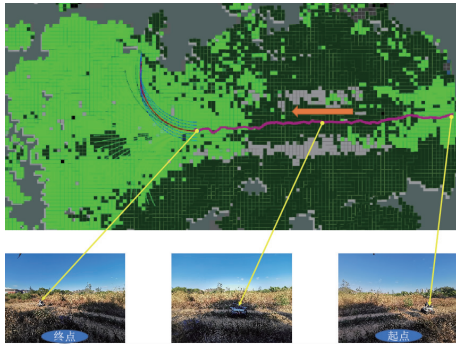
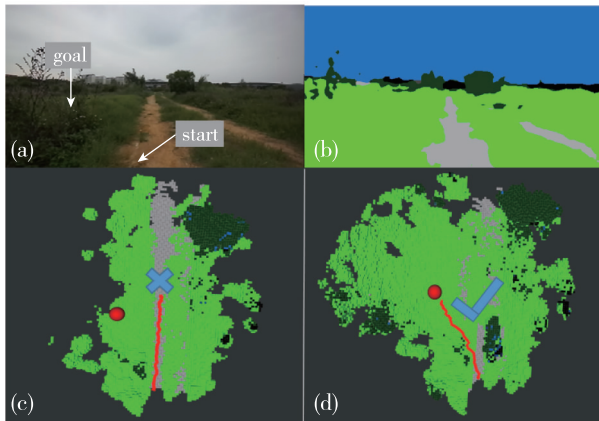


图 21 高植语义代价修正后的路径规划结果

Fig. 21 Path planning results after high plant semantic cost modification

(5)有无语义代价的 DWA 测试

为验证语义信息的存在与否对 DWA 算法产生的作用,如图 22(c)所示,传统的基于几何的 DWA 规划不会进入草丛,因为它总是把有一定高度的点云当作障碍物;如图 22(d)所示,当考虑草丛的语义



(a)测试场景;(b)语义分割结果;(c)不使用草丛语义信息的规划路径(失败);(d)使用草丛语义信息的规划路径(成功)

图 22 有无语义信息下的路径规划对比实验

Fig. 22 A comparative experiment of path planning with or without semantic information

信息后,因为草丛是可以通过的(代价低),因此会选择直接进入草丛到达目标点goal,说明增加语义信息可以让 DWA 规划器在实际可通行的区域(草丛)表现出类似人的驾驶行为。

以上实时路径规划的频率(1/周期)达到 10Hz。图 14~图 22 中,不同颜色代表的含义为:■道路 ■草丛 ■高植 ■建筑 ■车辆 ■天空 ■未知障碍;存在误分割的情况。

5 结论

本文针对复杂环境下地面无人平台的连续导航和智能感知规划问题,提出了卫星拒止下 IMU/磁力计/轮式里程计/双目视觉/3D 激光雷达的融合定位、基于多模态感知的语义 3D 地图构建与路径规划方案。算法分析与试验结果表明:

1)基于序贯式 ESEKF 的融合定位算法,扩展性好、场景适应性强,为后续全源导航的传感器即插即用及在线配置问题提供了解决思路。

2)改进的深度学习(图像)语义分割算法——MiniLedNet,在保持精度的同时具有快速推断能力;语义与点云的融合为自主导航提供了类人的智能感知规划能力,克服了基于几何地图方法的不足。

3)在真实环境中进行了整个系统的试验验证,表明该系统能够使无人车实时完成导航任务。为了进一步验证和完善所建立的导航定位、感知、规划方法,今后将在雨天、灰雾、复杂地形、电磁干扰环境下进行更充分的测试和完善。

参考文献

[1] Morales J, Vázquez-Martín R, Mandow A, et al. The UMA-SAR dataset: multimodal data collection from a ground vehicle during outdoor disaster response training exercises[J]. The International Journal of Robotics Research, 2021, 40(6-7): 835-847.

[2] 赖际舟, 袁诚, 吕品, 等. 不依赖于卫星的无人系统视觉/激光雷达感知与自主导航技术[J]. 导航定位与授时, 2021, 8(3): 1-14.

Lai Jizhou, Yuan Cheng, Lyu Pin, et al. Unmanned system visual/LiDAR perception and navigation technology independent of GPS[J]. Navigation Positioning and Timing, 2021, 8(3): 1-14(in Chinese).

[3] Simanek J, Reinstein M, Kubelka V. Evaluation of the EKF-based estimation architectures for data fusion in mobile robots[J]. IEEE/ASME Transactions on Mechatronics, 2015, 20(2): 985-990.

- [4] Ma J, Bajracharya M, Susca S, et al. Real-time pose estimation of a dynamic quadruped in GPS-denied environments for 24-hour operation[J]. *The International Journal of Robotics Research*, 2015, 35(6): 631-653.
- [5] Chambers A, Scherer S, Yoder L, et al. Robust multi-sensor fusion for micro aerial vehicle navigation in GPS-degraded/denied environments[C]// *Proceedings of American Control Conference (ACC)*. IEEE, 2014: 1892-1899.
- [6] Kubelka V, Oswald L, Pomerleau F, et al. Robust data fusion of multimodal sensory information for mobile robots[J]. *Journal of Field Robotics*, 2015, 32(4): 447-473.
- [7] Heo S, Park C G. Consistent EKF-based visual-inertial odometry on matrix lie group[J]. *IEEE Sensors Journal*, 2018, 18(9): 3780-3788.
- [8] 黄伟杰, 张国山. 基于快速不变卡尔曼滤波的视觉惯性里程计[J]. *控制与决策*, 2019, 34(12): 2585-2593.
Huang Weijie, Zhang Guoshan. Visual-inertial odometry based on fast invariant Kalman filter[J]. *Control and Decision*, 2019, 34(12): 2585-2593(in Chinese).
- [9] Kolvenbach H, Wisth D, Buchanan R, et al. Towards autonomous inspection of concrete deterioration in sewers with legged robots[J]. *Journal of Field Robotics*, 2020, 37(8): 21964.
- [10] Maturana D, Chou P W, Uenoyama M, et al. Real-time semantic mapping for autonomous off-road navigation [M]. *Field and Service Robotics*, 2018: 335-350.
- [11] Wang Y, Zhou Q, Liu J, et al. LEDNet: a lightweight encoder-decoder network for real-time semantic segmentation[C]// *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2019: 1860-1864.
- [12] Berrio J S, Zhou W, Ward J, et al. Octree map based on sparse point cloud and heuristic probability distribution for labeled images[C]// *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018: 3174-3181.
- [13] Chang L, Shan L, Jiang C, et al. Reinforcement based mobile robot path planning with improved dynamic window approach in unknown environment[J]. *Autonomous Robots*, 2021, 45(2): 51-76.
- [14] Qin T, Li P, Shen S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator[J]. *IEEE Transactions on Robotics*, 2018, 34(4): 1004-1020.
- [15] Shan T, Englot B. LeGO-LOAM: lightweight and ground-optimized lidar odometry and mapping on variable terrain[C]// *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018: 4758-4765.
- [16] Sola J. Quaternion kinematics for the error-state Kalman filter[J]. *arXiv e-prints arXiv: 1711.02508*, 2017.
- [17] 张雪涛, 方勇纯, 张雪波, 等. 基于误差状态卡尔曼滤波估计的旋翼无人机输入饱和控制[J]. *机器人*, 2020, 42(4): 394-405.
Zhang Xuetao, Fang Yongchun, Zhang Xuebo, et al. Error state Kalman filter estimator based input saturated control for rotorcraft unmanned aerial vehicle [J]. *Robot*, 2020, 42(4): 394-405(in Chinese).
- [18] Zhou D, Hou Q, Chen Y, et al. Rethinking bottleneck structure for efficient mobile network design[C]// *Proceedings of European Conference on Computer Vision*, 2020: 680-697.
- [19] Hornung A, Wurm K M, Bennewitz M, et al. OctoMap: an efficient probabilistic 3D mapping framework based on octrees[J]. *Autonomous Robots*, 2013, 34(3): 189-206.
- [20] Berrio J S, Ward J, Wei Z, et al. Fusing lidar and semantic image information in octree maps[C]// *Proceedings of Australasian Conference on Robotics and Automation*, 2017: 1-7.
- [21] 李扬宇. 基于深度传感器的移动机器人视觉 SLAM 研究[D]. 重庆: 重庆大学, 2017: 69-70.
Li Yangyu. Research on visual simultaneous localization and mapping of mobile robot based on depth sensor[D]. Chongqing: Chongqing University, 2017: 69-70(in Chinese).

(编辑:孟彬)