

• 中国汽车工程学会年会优秀论文(选登) •

# 基于二维概率网格的语义地图构建

张栋翔 邓一民 李天然

(上海汽车集团股份有限公司, 上海 200041)

**【摘要】**在同步定位与地图构建(SLAM)算法构建语义地图的过程中,为解决使用深度学习技术提取的语义信息中含有噪声的问题,提出通过网格的概率表征语义观测的语义匹配法,结合网格匹配和概率更新降低噪声的影响,无需特征点匹配等辅助手段即可增量地构建清晰的地图。基于实车数据构建的地图所包含的斑马线和车道线均清晰准确,证明了该方法的有效性。

**关键词:**语义匹配 语义噪声 概率网格 同步定位与地图构建

**中图分类号:** TP391.41      **文献标志码:** A      **DOI:** 10.20104/j.cnki.1674-6546.20230350

## Semantic Map Construction Based on 2D Probabilistic Grids

Zhang Dongxiang, Deng Yimin, Li Tianran

(SAIC Motor Corporation Limited, Shanghai 200041)

**【Abstract】**In the semantics mapping construction of Simultaneous Localization And Mapping (SLAM) algorithm, to solve the problem that the semantics information extracted by using deep learning technology contains noise, this paper proposes a representation method for semantic observation by probabilistic grid. The influence of noise is reduced by grid matching and probability updating, in this way, a clear map can be incrementally constructed without auxiliary means like features matching. The effectiveness of the proposed method is verified by experiments in which clear zebra crossings and lane lines are contained in the constructed map.

**Key words:** Semantic matching, Semantic noise, Probabilistic grid, Simultaneous localization and mapping

**【引用格式】**张栋翔, 邓一民, 李天然. 基于二维概率网格的语义地图构建[J]. 汽车工程师, 2023(10): 6-10.

ZHANG D X, DENG Y M, LI T R. Semantic Map Construction Based on 2D Probabilistic Grids[J]. Automotive Engineer, 2023(10): 6-10.

## 1 前言

近年来,随着语义分割、目标检测等领域的快速发展,语义地图因其稳定、可靠且占用内存空间小等优势,在智能导航中受到广泛关注。

通过视觉语义构建地图的方法可以分为2类:点云语义地图构建与概率语义地图构建。秦通等<sup>[1]</sup>基于点云配准(Iterative Closest Point, ICP)算法构建停车场的库位语义地图,由于点云密集易混淆,而未提取斑马线。胡佳欣<sup>[2]</sup>采用航位推算和基于ICP算法的语义点云匹配方法,在鸟瞰视觉(Bird's Eye

View, BEV)顶视图上采用ORB特征点(Oriented FAST and Rotated BRIEF Features, ORB Features)进行回环检测和全局优化,但该方法为构建清晰的斑马线地图,需要载体的行驶路径形成回环,且使用特征点作为辅助手段,增加了系统的复杂性。Zaganidis等<sup>[3]</sup>采用正态分布变换(Normal Distance Transform, NDT)方法验证了语义点云建图的鲁棒性和准确性,利用不同类别点云的法向量构成的NDT直方图估计局部地图之间的相似性,并在此基础上添加语义回环检测,但该研究对象为立体相机捕获的点云。

Yang等<sup>[4]</sup>融合语义立体框与特征点,提出立方体建图(Cube Simultaneous Localization and Mapping, CubeSLAM)方法,结合图像的模式识别技术,通过单目视觉提取物体的三维立体框。该方法可有效降低视觉漂移,提高相机位姿估计的精度,同时证明了语义识别和基于特征点的同步定位与地图构建(Simultaneous Localization And Mapping, SLAM)技术可相互促进,目标识别可提供更大范围的几何和尺度约束,SLAM技术能够进一步提升目标识别的精度。Bowman等<sup>[5]</sup>对识别结果使用期望最大化(Expectation Maximization)进行数据关联,进一步优化了机器人及地标的位姿。与CubeSLAM类似,该方法在采用最大似然估计推导出的最小二乘优化中增加约束,从而提高建图精度。受Bowman的启发,Doherty<sup>[6]</sup>认为数据关联在概率上呈现非高斯分布,提出使用最大混合类(Max-Mixture-Type)模型,将多个可能性数据关联进行先验性建模。在推理过程中,用最大边缘化法(Max-Marginalization)去除数据关联中的变量,同时保持标准的服从高斯分布的后验假设,最终完成基于非高斯概率分布的语义地图的构建。Bowman和Doherty等人的方法主要针对特征点与目标语义框的数据关联,并未阐述图像分割的语义点云匹配问题。

构建语义地图时,图像分割引起的语义误识别、边界像素分割不佳等问题,将导致无法正确进行语义匹配。因此,本文提出二维概率网络语义法,针对图像分割得到的语义点云,构建概率语义地图,对语义观测的时间先后顺序和空间连续性进行建模,无需借助特征点(如ORB Features),不保留短时间内的语义感知错误,从而滤除噪声,构建稳定、清晰的地图。

## 2 系统概述

在智能驾驶的应用场景中,通过地平面上的车道线、斑马线、箭头、停止线等静态地标构建语义地图。本文以上述交通标志为背景,阐述概率网络语义匹配策略。

### 2.1 框架

如图1所示,智能车辆可安装多个相机,能够在360°范围内覆盖车辆周边视野。惯性测量单元(Intertial Measurement Unit, IMU)、GPS、轮速和转向盘角度等传感器信息为视觉语义信息提供先验位姿。语义信息网格化后,基于先验位姿对语义信息

进行匹配,并将匹配结果转换为概率化网格。根据匹配过程中的约束,对地图进行最小二乘优化。其优化和匹配的结果反馈到多源信号的融合端,从而对先验信息进行纠正。最后,将大于给定概率阈值的网格设为地标物体占据的网格,输出语义地图。

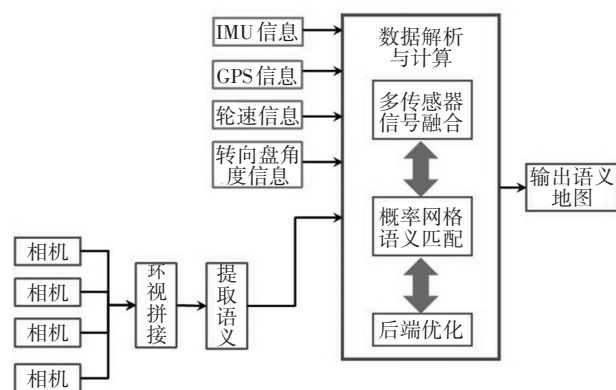


图1 智能车辆语义地图构建框架

### 2.2 环视标定拼接

使用智能车辆的多个单目相机采集图像,结合相机外参标定和图像拼接技术,将载体周边空间的物体的体素投影到与地面平行的平面,空间维度由三维降低到二维,物体的尺度在水平方向上保持不变。通过时序同步将采集的图像实时拼接成覆盖车辆周围360°范围的顶视图。而对于位于载体行驶轨迹一定空间范围内的物体,在一段时间内,智能载体对其覆盖的平面进行多次观测,将平面的各点映射到顶视图图像。

### 2.3 语义像素的分割提取

从2.2节的顶视图中提取所有组成车道线、斑马线、箭头、停止线以及可通行空间边缘线等语义的像素。上述语义信息大多属于静态的背景,对光照变化不敏感,不受天气变化影响,因而可鉴别性强。典型的车辆周边不同类别的语义信息如图2所示。

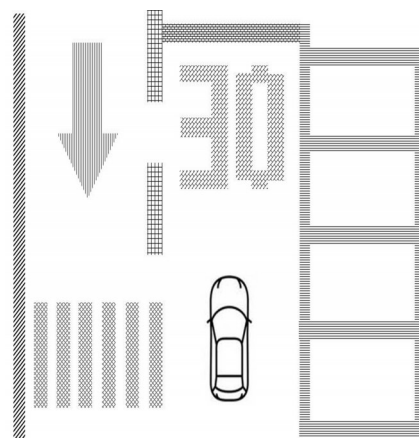


图2 顶视图的语义类别

### 2.4 网格语义预处理

对车辆周边的语义空间进行二维网格化预处理,如图3所示,车辆周边以分辨率 $\lambda$ 进行网格化。对含有语义信息的网格标记为“1”,没有语义信息的网格标记为“0”,同时对不同类别的语义进行标记。

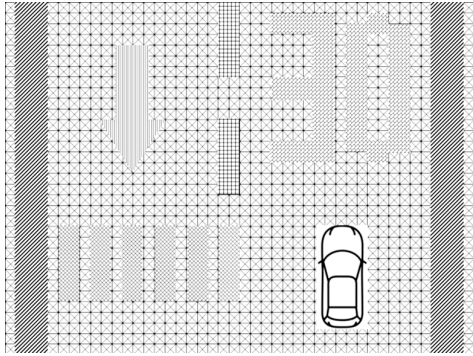


图3 语义网格化

### 2.5 多源信息融合的位姿估计

将多传感器信号进行时间戳对齐,利用卡尔滤波或因子里图(Georgia Tech Smoothing and Mapping, GTSAM)融合车载惯导、轮速计、转向盘角度等信息,为车辆的位姿估计提供初始值 $\chi_a=(P_a, \theta_a)$ ,其中, $P_a$ 为车辆的位置坐标, $\theta_a$ 为车辆的姿态朝向。

### 2.6 概率网格语义匹配

对语义网格进行概率化处理,采用最大化后验概率的方法进行匹配。图4所示为车辆顶视图转换的网格,通过概率网格语义匹配构建概率网格语义地图。其中,交通标志的网格概率大于背景网格。具体步骤如下:

a. 定义地图上某位置的网格(网格A)有语义(Hit,标记为“1”)的概率为 $\alpha$ ,无语义(Miss,标记为“0”)的概率为 $(1-\alpha)$ ,并对网格A的概率 $P$ 进行初始化:

$$f(P) = \frac{P}{1-P}, P = \alpha \quad (1)$$

随着车辆移动,相机对网格A在不同时刻进行观测,同时对网格A的概率进行更新:

$$\begin{cases} P_t = f^{-1}(C_{hit} \times f(P_{t-1})), z = 1 \\ P_t = f^{-1}(C_{miss} \times f(P_{t-1})), z = 0 \end{cases} \quad (2)$$

式中, $P_t$ 为 $t$ 时刻网格A的概率; $z=1$ 、 $z=0$ 分别表示观测结果为有语义、无语义; $C_{hit}$ 、 $C_{miss}$ 分别为观测结果为有语义和无语义时传感器的更新系数,由传感器的属性决定; $f^{-1}()$ 为式(1)的反函数。

b. 考虑到角度对语义匹配结果的影响,以2.5

节的初始位姿朝向 $\theta_a$ 为中心,在窗口 $(-\tau, +\tau)$ 内,以 $\delta$ 为角度间隔,在已有地图上对当前帧进行匹配,并在窗口内找到使式(3)最大的值,即当前车辆的朝向估计值 $\theta_b$ ,保持朝向不变,以2.5节的初始移动车辆位置 $P_a$ 为中心,在窗口内找到使式(3)概率最大的值 $P_b$ ,令 $\chi_b=(P_b, \theta_b)$ :

$$P(Z_t|\chi_t, m) = \prod_j P(Z_t^j|\chi_t, m) \quad (3)$$

建图过程中车辆位姿和地图中地标状态(包括网格的位置估计和网格概率)联合概率的最大似然估计<sup>[7]</sup>为:

$$P(\chi_t|\chi_{t-1}, u_t, m, Z_t) = P(Z_t|\chi_t, m)P(\chi_t|\chi_{t-1}, u_t) \quad (4)$$

式中, $Z_t$ 为 $t$ 时刻的2D语义网格,包括由顶视图转换的像素网格、以初始位姿 $\chi_a$ 为参考位置的周边语义网格的状态; $Z_t^j$ 为上述2D语义网格中各网格的语义状态; $\chi_t$ 为 $t$ 时刻的位姿; $u_t$ 为 $t$ 时刻车辆的运动状态; $m$ 为地图中的网格状态; $P(\cdot|\cdot)$ 为条件概率。

当式(3)取得最大值时,式(4)为最大概率值。

c. 为了得到精确值,建立误差模型,在初值 $\chi_a$ 附近添加扰动,采用最小二乘优化法得到精确解:

$$\chi^* = \arg \min_{\chi} \sum \left\| \left( 1 - P(Z_{hit}|\chi, m) \right) \right\|^2 \quad (5)$$

式中, $Z_{hit}$ 为有语义的网格,表示此时精确解 $\chi^*$ 使当前顶视图含有语义网格与已有地图中含有语义网格相互匹配。

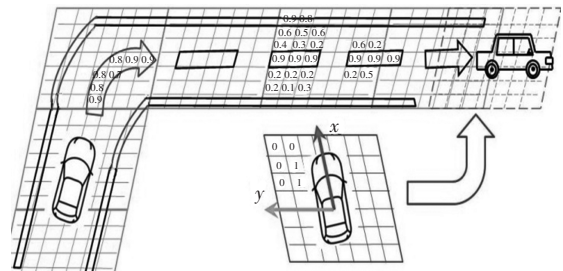


图4 网格匹配过程

确定车辆的位姿 $\chi^*$ 后,通过式(2)更新地图中已有网格的概率,同时创建新网格并赋予初始概率。

随着顶视图转换的网格持续更新,按照上述步骤执行,从而增量地构建地图,见图4(地图中数字表示概率的取值,无实际意义)。基于顶视图的多次观测产生的2D语义网格,对网格A的概率进行迭代更新。同理,地图中各网格的概率可多次更新,且不易受到偶尔的语义错误影响。因此,短时间内的语义识别错误或像素分割错误不会在地图上留下痕迹,从而提高地图的准确性。

### 3 试验结果

#### 3.1 试验算法流程

本文使用卡尔曼滤波融合IMU和轮速信息,为语义匹配提供先验位姿,如图5所示。将各网格的概率转换为图像的灰度值,以灰度图的方式输出语义匹配后的网格地图。

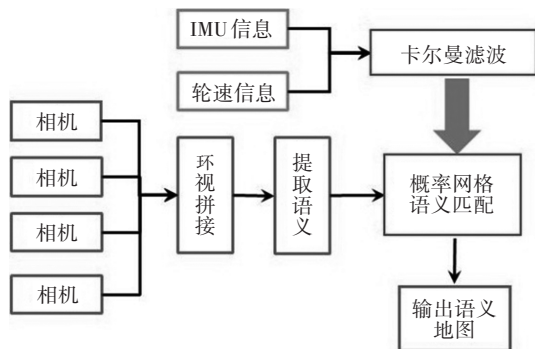


图5 算法实现框架

#### 3.2 试验环境

本文使用乘用车的4个环视相机采集的原始图像数据,在时间戳同步的基础上,将原始图像拼接为顶视图,其覆盖范围为 $14\text{ m}\times 14\text{ m}$ ,网格化分辨率为 $288\times 288$ 。试验运行的处理器为Intel® Core™ i7-4710MQ,主频2.5 GHz,内存8 GB。根据经验, $C_{\text{hit}}$ 和 $C_{\text{miss}}$ 分别设置为0.55和0.49。

使用深度层聚合(Deep Layer Aggregation, DLA)为网络的主干结构,通过U-Net<sup>[8]</sup>进行语义分割,从顶视图中提取不同类型的语义。

由于环境具有复杂性,图像分割时均会存在不同程度的噪声。图6所示为地下车库典型场景中不同类型的语义噪声。此外,图中均存在斑马线界限不清晰以及部分条纹线未识别的问题。图6b中道路间的地面由于白炽灯光反射被误识别为车道线。

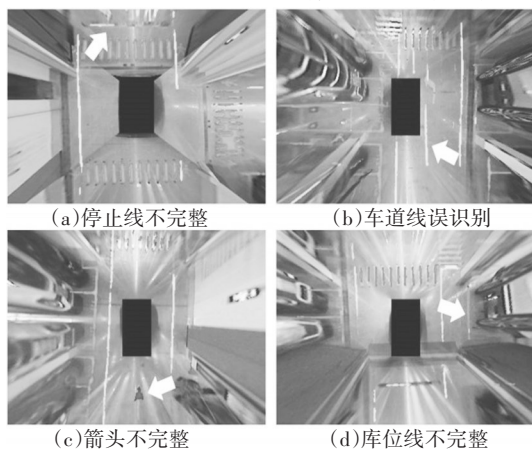


图6 语义分割的噪声

#### 3.3 试验结果

比较本文提出的概率网格匹配法与NDT语义点云匹配法所构建的语义地图。如图7所示,NDT法不能抵抗语义分割产生的噪声,其构建的语义地图存在语义边界不清晰,甚至无法区分纹理的现象,在U形转弯处语义混淆现象尤其明显。

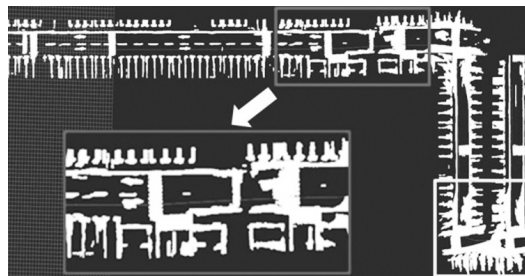


图7 NDT法构建的语义点云地图

本文提出的概率网格匹配法构建的语义地图如图8所示。与图7相比,概率网格匹配法能有效抵抗语义噪声,U形转弯区域所包含的斑马线条纹清晰可见。从细节上看,NDT法构建的点云地图语义轮廓粗糙、混叠不清晰,完全无法区分与条纹状斑马线相似度高且容易匹配错误的语义,而本文提出的概率网格匹配法构造的斑马线纹理清晰且正确。



图8 概率网格匹配法构建的语义点云地图

图9所示为卡尔曼滤波融合的里程计轨迹(简称里程计轨迹)、基于NDT语义点云匹配的建图轨迹和基于概率网格匹配的建图轨迹。

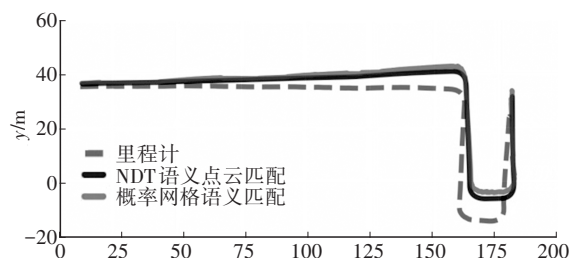


图9 里程计、NDT及概率网格建图轨迹

以里程计轨迹为基准,比较其他2种方法的建图轨迹与里程计轨迹之间的相对位姿误差(Relative Pose Error, RPE),如图10所示。

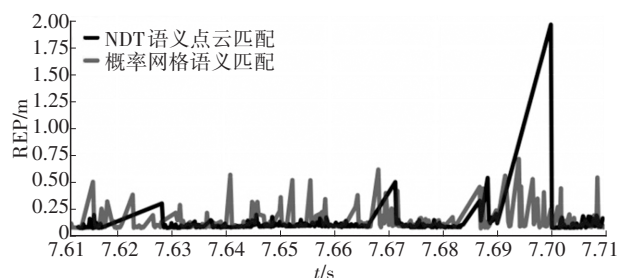


图 10 NDT与概率网格建图RPE曲线

表1列出了基于NDT语义点云匹配及基于概率网格语义匹配的RPE参数。由表1可知,概率网格匹配法与NDT点云语义匹配法多项指标接近,但概率网格匹配法最大RPE远小于NDT点云匹配法的RPE。RPE比较的是不同轨迹上相邻两帧之间的位姿变换,作为计算RPE基准的车辆里程计轨迹是相对精确的。RPE越大,表明建图轨迹与车辆里程计轨迹之间的误差越大,NDT法建图轨迹在部分区域与实际路径相差甚远。

表 1 相对位姿误差比较

RPE 参数	基于 NDT 法构建的地图	基于概率网格法构建的地图
最大误差	1.983 701	0.676 627
平均误差	0.039 666	0.071 272 2
中值误差	0.021 386	0.031 323 5
最小误差	0.002 756	0.002 144 71
均方根误差	0.120 422	0.127 398
和方差	5.061 011	5.35 598
标准差	0.113 702	0.105 596

#### 4 结束语

本文提出二维概率网络语义法进行概率语义地图构建,基于多源信息融合提供位姿先验信息,在联合概率的最大似然估计框架下,先确定角度,后确定位置,通过窗口搜索得到初始匹配位姿,再使用最小二乘法优化网格概率更新,得到智能载体的最终位姿。利用语义观测在空间和时间上的连续性,对各网格的概率进行多次更新,从而消除噪声的影响,获得了清晰的语义地图。试验结果证明了所提出方法的有效性。

此外,本文提出的方法无需多次使用相机采集原始图像信息(例如从原始图像上提取特征点),在计算过程中避免使用和存储原始图像,可节约计算资源和存储资源,降低成本。

未来,将构建大范围的三维语义地图,如在多层地库的场景中尝试利用本文提出的网格匹配和概率更新法构建语义地图。本文提出的语义匹配方法还可用于开发代客泊车定位算法。需要指出的是,本文所述的方法不限于智能驾驶,其他类似可获取行驶轨迹的顶视图场景均适用。

#### 参考文献

- [1] QIN T, CHEN T Q, CHEN Y L, et al. AVP-SLAM: Semantic Visual Mapping and Localization for Autonomous Vehicles in the Parking Lot[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, NV, USA: IEEE, 2020: 5939-5945.
- [2] 胡佳欣. 基于视觉语义信息的室内停车场建图与定位方法研究[D]. 上海: 上海交通大学, 2020. DOI:10.27307/d.cnki.gsjtu.2020.001581.  
HU J X. Mapping and Localization Method Based on Visual Semantic Information in InDoor Parking Lots[D]. Shanghai: Shanghai Jiao Tong University, 2020. DOI:10.27307/d.cnki.gsjtu.2020.001581.
- [3] ZAGANIDIS A, MAGNUSSON M, DUCKETT T, et al. Semantic-Assisted 3D Normal Distributions Transform for Scan Registration in Environments with Limited Structure [C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, BC, Canada: IEEE, 2017: 4064-4069.
- [4] YANG S C, SCHERER S. CubeSLAM: Monocular 3D Object SLAM[J]. IEEE Transactions on Robotics, 2019, 35(4): 925-938.
- [5] BOWMAN S L, ATANASOV N, DANILIDIS K, et al. Probabilistic Data Association for Semantic SLAM[C]// 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 1722-1729.
- [6] DOHERTY K J, BAXTER D P, SCHNEEWEISS E, et al. Probabilistic Data Association via Mixture Models for Robust Semantic SLAM[C]// IEEE International Conference on Robotics and Automation ICRA. Paris, France: IEEE, 2020: 1098-1104.
- [7] THRUN S, BURGARD W, FOX D. Probabilistic Robotics[M]. Cambridge, MA, USA: The MIT Press, 2005: 309-312.
- [8] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]// Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015). Munich, Germany: Springer, Cham: 234-241.

(责任编辑 瑞 秋)

修改稿收到日期为2023年8月12日。