

doi: 10.19562/j.chinasae.qcgc.2024.11.009

融合语义信息与物体级几何特征的 实时动态激光SLAM算法*

兰凤崇^{1,2}, 田小强^{1,2}, 陈吉清^{1,2}, 车宇翔^{1,2}, 周云郊^{1,2}

(1. 华南理工大学机械与汽车工程学院, 广州 510640; 2. 华南理工大学, 广东省汽车工程重点实验室, 广州 510640)

[摘要] 针对现有的激光SLAM算法在动态场景下鲁棒性差、定位与建图精度易受动态物体干扰的问题, 提出了一种融合语义信息与物体级几何特征的实时动态激光SLAM算法 Object-SuMa。首先通过地面滤波、物体分割、方向包围盒解算等过程, 生成物体级几何特征并表示为纹理, 用于并行计算和修正物体内部错误的语义分割结果; 然后在配准过程中分解计算方向包围盒间的IOU, 并基于包围盒IOU和语义分割结果引入物体级几何加权和语义加权, 减少误匹配和动态点匹配; 借助图形渲染管线构建并行计算过程, 进行地面点配准和非地面点配准两步优化, 降低计算复杂度和耗时; 最后在KITTI里程计数据集测试表明, Object-SuMa算法相比SuMa++算法的相对位姿估计精度提升15%, ICP平均耗时下降17%, 改善了动态场景下的激光SLAM定位精度和鲁棒性。

关键词: 激光SLAM; 动态场景; 物体级几何特征; 语义信息; 并行计算

Real-Time Dynamic Laser SLAM Algorithm Combining Object-Level Geometric Features and Semantic Information

Lan Fengchong^{1,2}, Tian Xiaoqiang^{1,2}, Chen Jiqing^{1,2}, Che Yuxiang^{1,2} & Zhou Yunjiao^{1,2}

1. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640;

2. South China University of Technology, Guangdong Provincial Key Laboratory of Automotive Engineering, Guangzhou 510640

[Abstract] In view of the problems of the existing laser SLAM algorithm in dynamic scenes, which has poor robustness and the positioning and mapping accuracy is easily disturbed by dynamic objects, a real-time dynamic laser SLAM algorithm called Object-SuMa that combines object-level geometric feature and semantic information is proposed. Firstly, through processes such as ground filtering, object segmentation and pose size calculation, object-level geometric features are generated and represented as texture and used to correct semantic segmentation errors within the object. Then, in the odometry stage, the IOU calculation of the oriented bounding box is decomposed, and object-level geometric weighting and semantic weighting are introduced based on the bounding box IOU and semantic segmentation results to reduce mismatching and dynamic point matching. In addition, the graphics rendering pipeline is used to build a parallel computing process, and the computational complexity and time consuming are reduced by two-step optimization of ground point registration and non-ground point registration. Finally, tests on the KITTI odometry data set show that compared with SuMa++, the Object-SuMa algorithm has improved the relative pose accuracy by 15% and reduced the average time of ICP by 17%, which improves the positioning accuracy and robustness of laser SLAM in dynamic scenarios.

Keywords: laser SLAM; dynamic environment; object-level geometric feature; semantic information; parallel computing

* 国家自然科学基金(52175267)、广东省自然科学基金(2021A1515010912)、国家车辆事故深度调查体系(NAIS)和新能源汽车事故调查协作网资助。

原稿收到日期为2024年02月26日, 修改稿收到日期为2024年05月30日。

通信作者: 陈吉清, 教授, 博士, E-mail: chjq@scut.edu.cn。

前言

获取准确实时的位姿信息是实现自主导航的必备条件之一,同时定位与建图(SLAM)技术是解决这个问题的常用方法。相比于相机,激光雷达能直接获取更稳定、准确的测量,且随着激光雷达成本的降低,激光雷达已在众多量产车辆上获得应用。然而现有激光SLAM算法大多基于静态场假设^[1-2],在高速公路、城市街区等动态交通场景,由于动态物体的影响,其性能变差。一方面,动态点造成点云发生误配准,导致当前帧与参考帧的位姿估计发生偏差,定位精度下降;另一方面低精度的位姿估计导致构建的点云地图出现大量动态物体的“鬼影”,不利于后续基于地图的规划定位。

针对动态场景,现有研究主要提出了基于无模型变化检测和基于深度学习模型两种方法。基于无模型变化检测的方法将当前帧与参考帧比较,判断场景中点云是否为动态并滤除动态点,主要有视点可见法和光线投影法^[3]。视点可见法^[4-5]将当前帧和同一视点附近的地图投影成深度图,比对两个深度图同一位置像素的深度,若后者深度更深,则认为当前帧该像素位置为动态点, Kim等^[6]基于多分辨率深度图恢复被“误杀”的静态点,提升了视点可见法的精度。然而在非常开阔环境下,由于周围环境缺少远点,基于视点可见性的方法无法滤除运动点;光线投影法^[7-9]基于栅格被激光点击中和穿过的情况计算栅格的占据概率,概率低于阈值则判定栅格为动态,将栅格内点云作为动态点云滤除,然而这种方法须遍历每条光路上的所有栅格,非常消耗计算资源,且“误杀”及“漏杀”动态点受栅格尺寸影响较大。基于深度学习模型^[10-11]的方法直接用深度学习训练出一个end-to-end网络,在预处理环节识别动态点,滤波后的静态点云用于配准。Pfreundschuh等^[12]基于3D-MiniNet网络进行实时3D动态物体检测,将滤除动态物体的点云进行常规的激光SLAM,然而基于深度学习的方法只能识别训练过的动态物体,无法过滤所有动态物体。

基于面元(Surfel)的SLAM算法,如Behley等^[13]提出的SuMa采用面元表示地图,相较于栅格地图和点云地图,面元地图具有内存消耗更小,资源占用更低的优点。SuMa采用二维栅格地图中的滤波方法,维护静态面元地图,降低了动态点的影响。Chen等^[14]提出的SuMa++在面元ICP和面元地图更新中

引入语义约束,提高了动态场景下的定位精度和全局地图一致性。Li等^[15]在LOAM基础上增加语义辅助ICP和语义辅助回环检测,也改善了里程计精度和建图质量。可见,引入语义信息能更加鲁棒地进行位姿估计和环境建模。然而,基于深度学习的语义分割网络都存在语义分割错误的问题,导致基于语义的ICP配准结果受语义分割网络的影响较大。当前SLAM算法对语义信息的应用也多停留在点特征层面,在点云配准时引入语义权重是最常见的方式,缺乏语义和物体级特征信息的结合。

综上,本文沿用SuMa++基于面元地图的技术路线,研究了一种融合语义信息与物体级几何特征的实时动态激光SLAM算法Object-SuMa。主要内容包括:(1)分割点云获得物体并解算物体的方向包围盒,作为点云物体级几何特征,并表示为纹理用于并行计算和修正物体内部错误的语义分割;(2)分解计算方向包围盒间的IOU,并在配准时基于包围盒IOU和语义分割结果引入物体级几何加权和语义加权,增加鲁棒性;(3)借助图形渲染管线构建并行计算过程,分为地面点和非地面点两步优化,降低计算复杂度;(4)通过仿真实验验证了算法的性能并分析了实验结果。

1 算法概述

如图1所示,以物体方向包围盒的位姿尺寸作为物体级几何特征并借鉴SuMa++算法,设计融合语义信息与物体级几何特征的Object-SuMa算法框架。Object-SuMa系统主要包括预处理、里程计、回环检测、位姿图优化和地图更新5个模块。

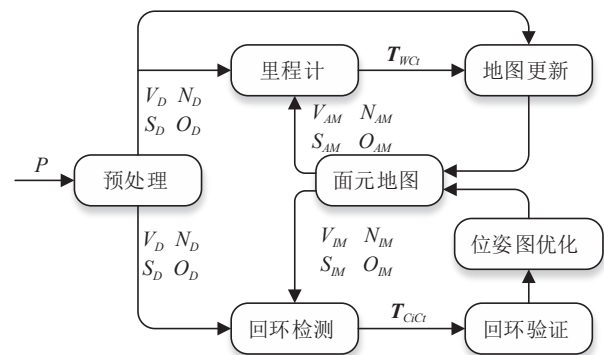


图1 Object-SuMa系统总体架构

(1)在预处理模块,获取点云的语义、法向量和方向包围盒信息,并投影到距离图像上用矩形纹理表征, V_D 、 N_D 、 S_D 、 O_D 分别为坐标图像、法向量图像、语义图像和方向包围盒图像;同时,基于面元地图渲染

出上帧坐标系下的活跃参考帧和非活跃参考帧,分别用作里程计和回环检测,活跃参考帧包含近期创建或被观测更新的活跃面元,非活跃参考帧则相反,两者均含点云坐标(V_{AM} 、 V_{IM})、法向量(N_{AM} 、 N_{IM})、语义(S_{AM} 、 S_{IM})和方向包围盒(O_{AM} 、 O_{IM})信息。

(2)在里程计模块,点云ICP中引入语义信息和物体级几何特征,采用地面点配准和非地面点配准两步优化的策略,获得当前帧 C_i 到世界坐标系 W 的最优位姿估计 T_{WC_i} 。点云物体级几何特征提取和里程计模块在第2和第3节展开阐述。

(3)在回环检测和位姿图优化模块采用SuMa++相同的策略,基于“合成渲染帧”与当前帧执行ICP匹配检测回环, $T_{C_i C_j}$ 表示当前帧 C_i 与某历史帧 C_j 的位姿变换,确认回环后增加回环约束,然后基于gtsam因子图框架优化轨迹位姿。

(4)在地图更新模块,从当前帧提取新面元并由里程计获得的最优位姿 T_{WC_i} 将新面元更新到面元地图中。下面介绍面元概念和面元地图更新方法。

1.1 面元

面元(surfel)^[16]是空间中有方向和大小的圆形平面,面元构成要素包括:中心点 $v_s \in \mathbb{R}^3$ 、法向量 $n_s \in \mathbb{R}^3$ 、半径 $r_s \in \mathbb{R}$ 、语义标签 $s_s \in \mathbb{R}$ 与稳定概率 $l_s \in \mathbb{R}$ 、方向包围盒 $o_s \in \mathbb{R}^8$ 、创建时间 $t_c \in \mathbb{R}$ 与更新时间 $t_u \in \mathbb{R}$ 。如图2所示,基于面元的配准残差为点面距离残差,如当前帧点 v_d 与参考帧对应面元所在平面的距离 d ,通过向量内积求得。

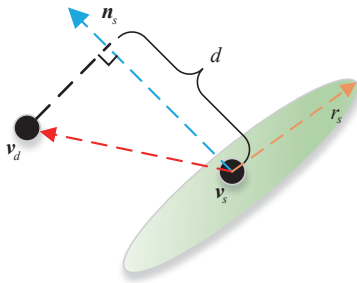


图2 面元残差

1.2 面元地图更新

从当前帧提取新面元并与面元地图更新融合,面元融合一方面扩充了周围环境信息,另一方面降低了面元地图的内存消耗,减少同一位置点云误差带来的重影问题。

定义新面元半径 r'_s 为

$$r'_s = \frac{\sqrt{2} \|v'_s\|_2 \cdot p}{\text{clamp}(-v'_s{}^T n'_s \cdot \|v'_s\|_2^{-1}, 0.5, 1.0)} \quad (1)$$

式中: p 为纹理像素尺寸; v'_s 为面元中心坐标; n'_s 为面元的法向量。待更新面元为 s ,若 $\|n'_s{}^T(v_s - v'_s)\| < \sigma_m$ 且 $\|n'_s \times n_s\| < \sin(\theta_m)$,则认为 s' 与 s 一致,然后采用指数滑动平均方法对面元 s 更新:

$$\begin{cases} v_s^{(t)} = (1 - \gamma) \cdot v'_s + \gamma \cdot v_s^{(t-1)} \\ n_s^{(t)} = (1 - \gamma) \cdot n'_s + \gamma \cdot n_s^{(t-1)} \\ r_s^{(t)} = r'_s, \text{ if } r'_s < r_s \end{cases} \quad (2)$$

式中 γ 为权重。否则,在地图创建新面元 s' 。

稳定概率 l_s 近似于二维栅格地图中的占据概率^[17], l_s 随时间更新,当前时刻概率是上一时刻概率与当前观测的叠加,而稳定的面元会被用于匹配。采用二值贝叶斯滤波更新概率:

$$l_s^{(t)} = l_s^{(t-1)} - \text{odds}(p_{\text{prior}}) - \text{odds}(p_{\text{penalty}}) + \text{odds}\left(p_{\text{stable}} \cdot \exp\left(-\frac{\alpha^2}{\sigma_d^2}\right) \exp\left(-\frac{d^2}{\sigma_d^2}\right)\right) \quad (3)$$

式中: p_{stable} 、 p_{prior} 分别为观测面元的稳定概率和先验概率; σ^2 为测量噪声方差; α 为观测面元的法向量与地图中对应待更新面元的法向量的夹角; d 为观测面元与对应地图中待更新面元的距离; p_{penalty} 表示语义类型惩罚项,若观测面元与地图中对应面元语义类别不一致,则面元稳定概率 l_s 会降低。

2 物体级几何特征与语义修正

图3为点云预处理示意图, Object-SuMa采用RangeNet++^[18]对点云进行语义分割,获取语义标签和概率信息;通过计算距离图像上某点 A 与其两相邻像素点 B 、 C (B 、 C 不同行不同列)间向量 AB 和 AC

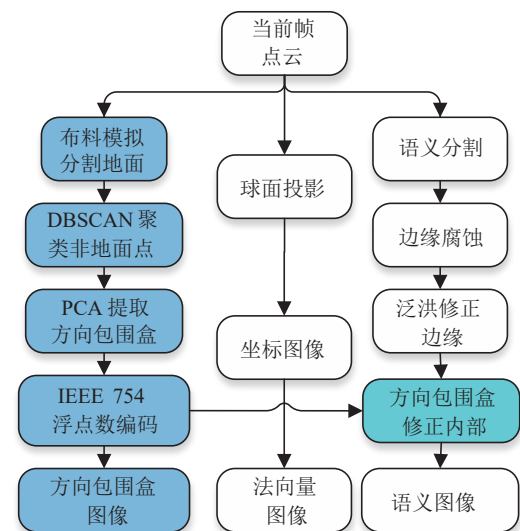


图3 点云预处理过程

的外积,获取点A附近的法向量。图4和图5分别为KITTI数据集00序列00帧点云对应的语义图像和法向量图像。

如图3中非透明圆角矩形框所示,在预处理部

分Object-SuMa相比于SuMa++增加物体分割与方向包围盒求解部分,将方向包围盒的位姿尺寸作为物体级几何特征,并编码为矩形纹理修正物体内部的点云语义分割错误。

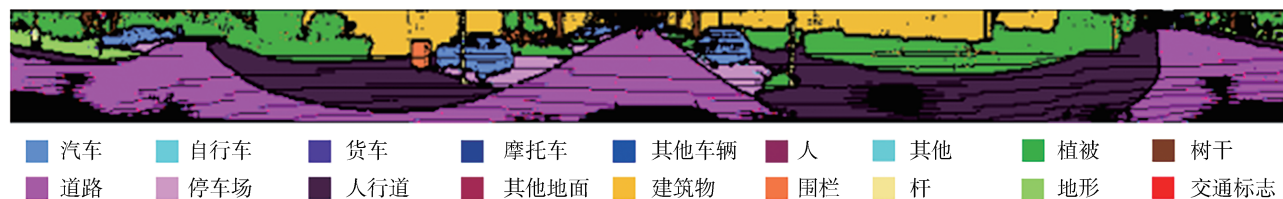


图4 语义图像



图5 法向量图像

2.1 物体分割

采用地面滤波、非地面点聚类的方法分割出物体点云。首先,采用布料模拟滤波算法^[19]分割地面。布料模拟将点云视作具有一定刚度的布料,模拟出地面形状。相较于RANSAC等平面拟合的方法,更适用于有起伏的路面提取。布料模型的计算量随网格分辨率增大而增大,布料模型能够模拟的路面起伏随刚度增大而减小。多次实验发现将布料模型分辨率设置为0.5,刚度设置为2,能够有效快速地分割地面。图6所示为KITTI数据集00序列00帧点云对应的地面分割结果,红色点代表地面点,绿色点代表非地面点。然后采用DBSCAN算法从非地面点聚类出物体。DBSCAN具有聚类速度快、对噪声容忍度高的优点。增大DBSCAN算法的聚类半径 ϵ 和样本点成为核心对象所需要的邻域样本数MinPts会导致本不属于同一簇的点划分到同一簇中,而减小 ϵ 和MinPts会导致更多的小簇形成。多次实验发现将聚类半径设置为0.5 m,样本点成为核心对象所需要的邻域样本数设置为5,能有效对非地面点进行聚类分割。图7所示为KITTI数据集00序列00帧点云的聚类结果及物体方向包围盒求解结果,文中舍弃部分包围盒,物体的包围盒求解在下一节阐述。

2.2 物体级几何特征求解与编码

获取点云聚类簇后,求解物体方向包围盒的位姿尺寸作为物体几何特征,并编码为纹理用于后续运算。首先,简化方向包围盒,使其底面与XOY平面平行;然后采用主成分分析求解方向包围盒的姿

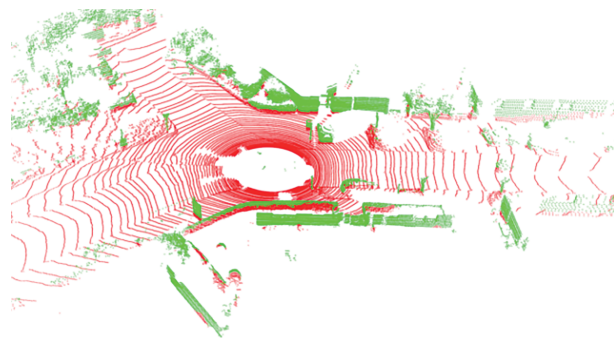


图6 地面分割

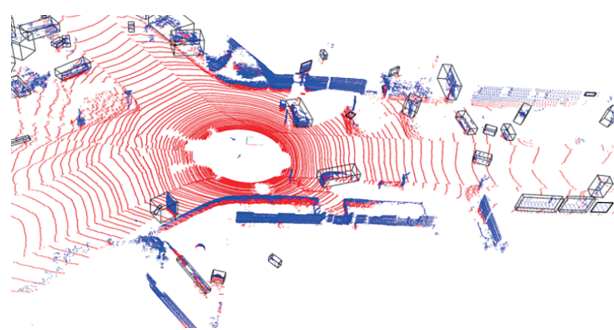


图7 聚类及物体方向包围盒

态,计算聚类簇的协方差矩阵 A 。

$$A = \begin{bmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(x, y) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(x, z) & \text{cov}(y, z) & \text{cov}(z, z) \end{bmatrix}$$

对 A 进行特征值分解和特征向量求解,以特征向量为列元素组成旋转矩阵 B 即为方向包围盒的姿

态;然后求解方向包围盒的位置和尺寸,通过将点云簇左乘旋转矩阵 B 的逆将其转换到方向包围盒姿态下,统计各轴坐标最大值最小值,得到在方向包围盒姿态下包围盒几何中心 C 和尺寸;最后,将几何中心 C 左乘旋转矩阵 B 将其转换到当前帧激光雷达坐标系下,获得方向包围盒的位置,物体的方向包围盒如图7所示。

OpenGL的RGBA纹理像素可存储4个单精度浮点数。经实验发现,若多个纹理共同来存储包围盒位姿尺寸信息,多个纹理在渲染过程中难以保证映射相同的纹理坐标,导致包围盒信息出错。因此在不降低数据可靠性的前提下采用半精度浮点数 half (16位)表示方向包围盒的位姿尺寸,并通过数据重新编码的方式,实现用单个RGBA像素表示方向包围盒的位姿尺寸信息。方向包围盒的位姿矩阵 T 为

$$T = \begin{bmatrix} x_x & y_x & 0 & t_x \\ x_y & y_z & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

式中: $[x_x, x_y, 0]$ 、 $[y_x, y_z, 0]$ 表示包围盒局部坐标系的 x 轴单位向量、 y 轴单位向量在当前帧坐标系的投影; $[t_x, t_y, t_z]$ 表示包围盒几何中心。结合包围盒的尺寸 $[x_{len}, y_{len}, z_{len}]$ 整理得矩阵 T_1 为

$$T_1 = \begin{bmatrix} x_x \cdot x_{len} & x_y \cdot x_{len} & 0 & t_x \\ x_y \cdot x_{len} & y_z \cdot y_{len} & 0 & t_y \\ 0 & 0 & z_{len} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

再由IEEE754规范将 T_1 内各非0元素两两组合编码,如 $x_x \cdot x_{len}$ 和 $x_y \cdot x_{len}$ 共同编码成一个单精度浮点数(32位),作为像素通道A。图8(a)和图8(b)分别为KITTI数据集00序列07帧点云的语义图像和方向包围盒图像,语义图像中同颜色像素对应同一类别点云,如蓝色表示汽车,方向包围盒图像中相同颜色且相邻的像素对应同一物体内的点云。图8(c)为汽车点云的方向包围盒纹理图像。由此可见,数据重编码的方式能够在单个RGBA纹理像素内存储的方向包围盒的位姿尺寸信息,采用纹理存储点云表示物体级几何特征,便于后续点云配准时引入物体级几何加权。

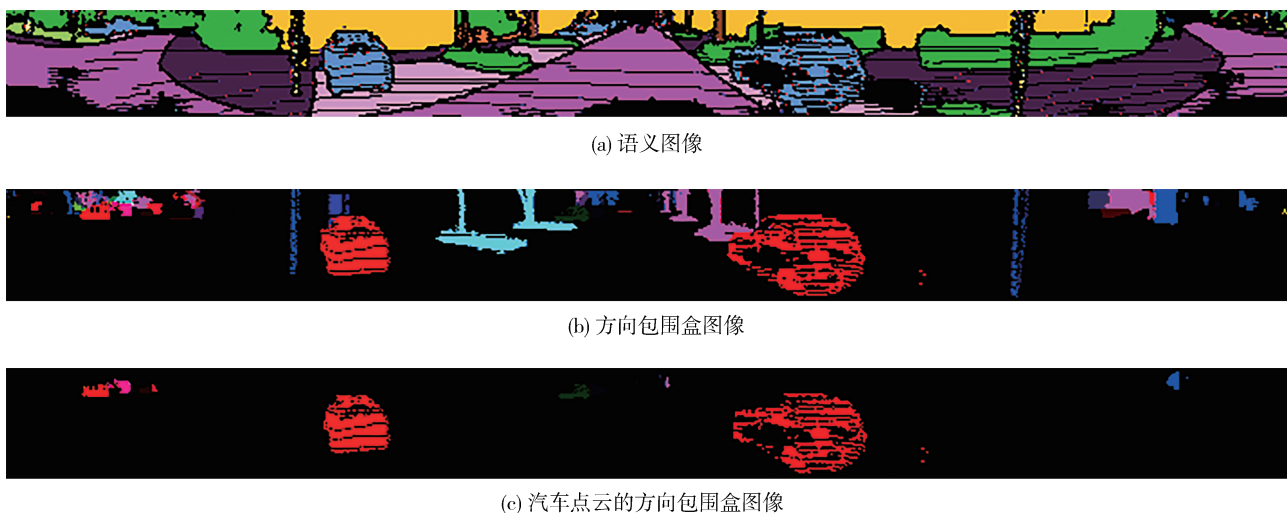


图8 语义图像和方向包围盒图像

2.3 物体内部语义修正

SuMa++采用侵蚀和泛洪算法修正不同类别边缘处点云的错误语义,却无法解决物体内部错误的语义分割。针对这一问题,采用物体的方向包围盒辅助修正物体内部的点云语义。

图形渲染管线在GPU端进行渲染运算,无法构建如kdtree、octree等点云结构。在获得点云的语义图像和方向包围盒图像后,采用在像素坐标上逐层

扩散的方式聚类具有相同包围盒信息的像素为同一物体。如图9所示,与当前像素曼哈顿距离相同的为同一层,逐层向外扩散,最多扩散约40层。扩散时统计各语义类别的像素个数,将像素个数最多的语义类别确定为物体类别,并将其他像素的语义类别置为物体类别。

图10所示为KITTI里程计数据集00序列07帧活跃参考帧的点云对比图。图10(a)为未修正语义

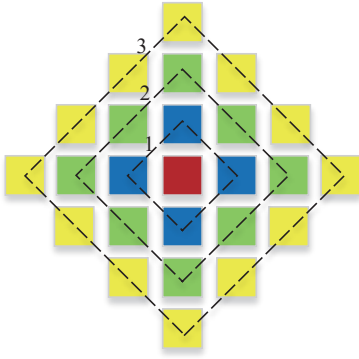
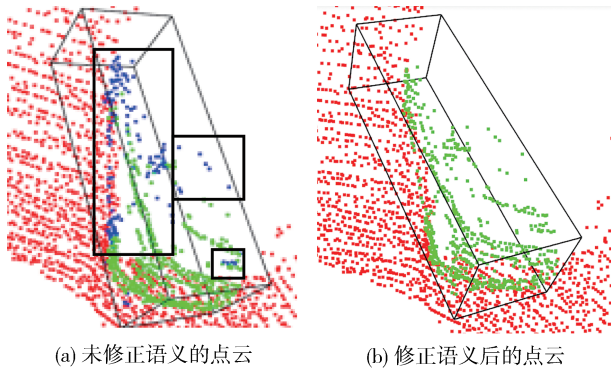


图9 纹理像素扩散过程

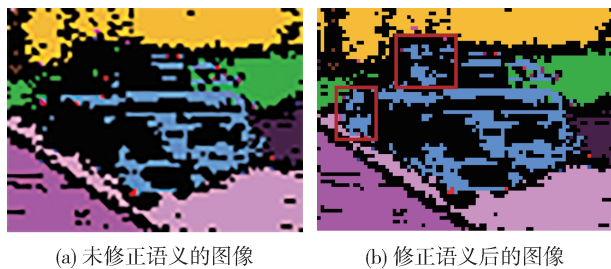
的点云,黑色框内的蓝色点为语义分割错误的点;图10(b)为修正语义的点云,可看出同一辆车的点具有相同颜色,均为车辆语义标签。图11所示为对应的语义图像,图11(a)为未修正的,图11(b)为修正的,红色框内为被修正语义的点,可看出右侧同车车上部分点被修正为车辆语义标签。



(a) 未修正语义的点云

(b) 修正语义后的点云

图10 物体内部语义修正点云对比



(a) 未修正语义的图像

(b) 修正语义后的图像

图11 物体内部语义修正图像对比

3 融合物体级几何特征的两步优化

当前帧与活跃参考帧配准,估计当前帧最优位姿。根据预测位姿变换 $T_{C_{i-1},C_i}^{(k-1)}$ 将当前帧坐标图像

中的像素映射到上一帧坐标系下,由像素坐标直接与活跃参考帧的坐标图像的像素匹配。基于面元的最优位姿估计问题,本质上是最小化点面距离残差问题,即最小化当前帧中所有点到活跃参考帧中对应面元所在平面距离的和:

$$E(\mathcal{V}_D, \mathcal{V}_{AM}, \mathcal{N}_{AM}) = \sum_{v_d \in \mathcal{V}_D} \mathbf{n}_{am}^T \cdot \left(T_{C_{i-1},C_i}^{(k)} v_d - v_{am} \right)^2 \quad (4)$$

式中: v_d 为当前帧中的观测点; v_{am} 为活跃参考帧中对应面元的中心坐标; \mathbf{n}_{am} 为对应面元的法向量; $T_{C_{i-1},C_i}^{(k)}$ 表示第 k 次迭代时当前帧到活跃参考帧的估计位姿。综合考虑地面具有的平面特征和非地面物体的位姿尺寸,提出融合物体级几何特征和语义信息两步优化匹配。

3.1 两步优化

δ 表示第 k 次迭代相比于 $k-1$ 次迭代时的线性位姿增量, $T_{C_{i-1},C_i}^{(k)} = \exp(\delta) T_{C_{i-1},C_i}^{(k-1)}$ 。用上一帧的估计位姿来初始化当前帧的位姿,即 $T_{C_{i-1},C_i}^{(0)} = T_{C_{i-2},C_{i-1}}$ 。采用高斯牛顿迭代求解 δ 来最小化式(4)中的目标函数:

$$\delta = (J^T W J)^{-1} J^T W r \quad (5)$$

式中: $W \in \mathbb{R}^{n \times n}$ 为权重对角矩阵, n 为观测点个数; w_{u_i} 表示某个残差 r_{u_i} 的权重; $r \in \mathbb{R}^{n \times 1}$ 表示残差向量; $J \in \mathbb{R}^{n \times 6}$ 表示残差 r 关于位姿增量 δ 的雅可比矩阵。对于某个观测点 v_u , 其对应 J 的第 i 行^[20]为

$$J_i = [n_m^T, (v_m \times n_m)^T] \quad (6)$$

式中 n_m 、 v_m 分别表示 v_u 在参考帧中对应面元的法向量和中心坐标。

下面推导两阶段的高斯牛顿计算过程。首先,将线性位姿增量 δ 拆分为 δ_{ground} 和 $\delta_{\text{offground}}$, 两者均 $\in \mathbb{R}^3$, 前者包含俯仰、翻滚和垂向平移3部分位姿增量,后者包含横摆、 X 轴方向平移、 Y 轴方向平移3部分位姿增量;然后将残差 r 拆分为所有地面点配准的残差 r_{ground} 和非地面点配准的残差 $r_{\text{offground}}$;最后,推导 r_{ground} 关于 δ_{ground} 的雅可比矩阵 J_{ground} , 推导 $r_{\text{offground}}$ 关于 $\delta_{\text{offground}}$ 的雅可比矩阵 $J_{\text{offground}}$, 两者均 $\in \mathbb{R}^{n \times 3}$ 。可看出,对于地面观测点 v_{g_i} , 其对应 J_{ground} 的第 i 行 J_{g_i} , 对于非地面观测点 v_{offg_i} , 其对应 $J_{\text{offground}}$ 的第 i 行 J_{offg_i} :

$$J_{g_i} = [n_{m_z}, ((v_m \times n_m)^T)_x, ((v_m \times n_m)^T)_y] \quad (7)$$

$$J_{offg_i} = [n_{m_x}, n_{m_y}, ((v_m \times n_m)^T)_z] \quad (8)$$

3.2 物体包围盒交并比计算

非地面观测点 v_u 配准时,须计算其方向包围盒 Box_u 与参考点方向包围盒 Box_ref 的交并比 IOU。首先,将 Box_u 由 $\exp((\delta_{\text{offground}}^{(k-1)})^\wedge)$ 转换到参考帧坐

标系下获得 $\text{Box_u}'$, 并忽略 $T_{c_{i-1}, c_i}^{(k-1)}$ 和 $T_{c_{i-1}, c_i}^{(0)}$ 之间的俯仰、翻滚和垂向平移运动, 则 $\text{Box_u}'$ 和 Box_ref 底面均平行于 XOY 平面。 $\text{Box_u}'$ 和 Box_ref 的 IOU 等价于鸟瞰图上两俯视矩形框的 IOU, 与正视图上两包围盒高程交并比的乘积, 如图 12 所示。

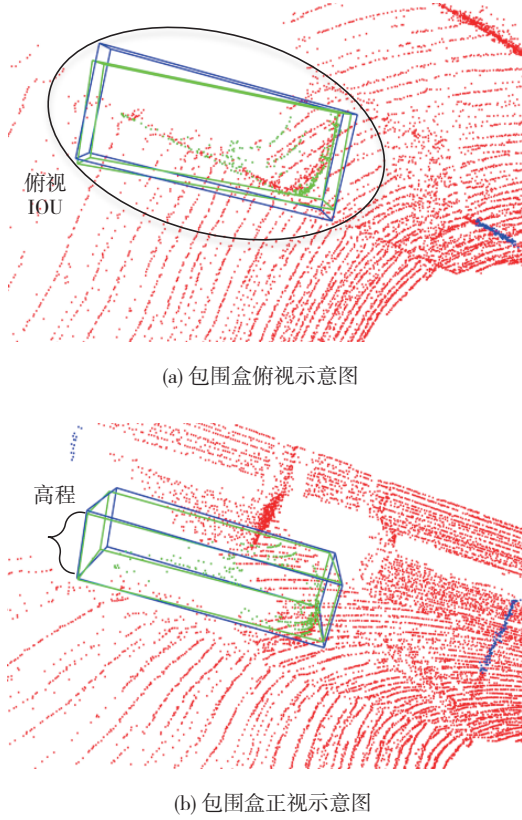


图 12 包围盒 IOU 计算分解图

然后采用简单高效的 Sutherland-Hodgman^[21] 算法和高斯面积公式求解两俯视矩形框的相交部分面积, 用于计算 IOU。图 13 所示为 Sutherland-Hodgman 算法求解两矩形框的相交多边形的原理, 通过轮流延长参考框的边线, 同时遍历观测框的所有边, 添加可见侧的顶点或交点, 形成相交多边形。高斯面积公式又名鞋带公式, 计算过程很像鞋带一样缠绕。鞋带公式计算多边形面积 S 如式(9)所示。

$$S = \sum_{i=1}^{n-1} (x_i \cdot y_{i+1} - y_i \cdot x_{i+1}) + x_n \cdot y_1 - y_n \cdot x_1 \quad (9)$$

3.3 物体级几何加权

点云配准时考虑配准残差的物体级几何权重 C_{geometry} 、语义权重 C_{semantic} 和核函数权重 C_{huber} 。

首先, 将观测点与参考点所在物体的包围盒

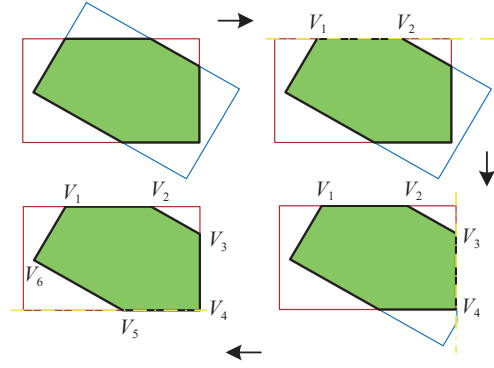


图 13 Sutherland-Hodgman 算法原理示意图

IOU 作为物体级几何权重 C_{geometry} 。包围盒间的 IOU 越大, 观测点为静态点的可能性越大, 如静止车辆点云的配准; 包围盒间的 IOU 越小, 观测点为动态点的可能性越大, 如运动车辆点云的配准。物体级几何权重 C_{geometry} 由式(10)计算获得:

$$C_{\text{geometry}}(O_D(i), O_{AM}(i)) = \begin{cases} 0, & \text{IOU}_{\text{bev}} \cdot \text{IOU}_{\text{height}} < 0.2 \\ 1, & \text{IOU}_{\text{bev}} \cdot \text{IOU}_{\text{height}} > 0.8 \\ \text{IOU}_{\text{bev}} \cdot \text{IOU}_{\text{height}}, & \text{其他} \end{cases} \quad (10)$$

式中: IOU_{bev} 表示包围盒俯视图上两矩形框的 IOU; $\text{IOU}_{\text{height}}$ 表示包围盒高程的交并比。考虑到包围盒的不确定性, 将 $\text{IOU} < 0.2$ 的匹配对滤除, $\text{IOU} > 0.8$ 的 C_{geometry} 置为 1。

其次, 考虑观测点与参考点的语义标签, 引入语义权重 C_{semantic} 。 C_{semantic} 为观测点与参考点类别相同的概率。若观测点的最大概率的预测类别与参考点相同, 则 C_{semantic} 等于观测点的最大概率预测类别的语义分割概率, 否则为 $1 - \text{语义分割概率}$ 。语义权重 C_{semantic} 由式(11)计算获得:

$$C_{\text{semantic}}(S_D(i), S_{AM}(i)) = \begin{cases} P(l_{S_D(i)} | i), & l_{S_D(i)} = l_{S_{AM}(i)} \\ 1 - P(l_{S_D(i)} | i), & \text{其他} \end{cases} \quad (11)$$

式中: $l_{S_D(i)}$ 表示第 i 个观测点的语义标签; $l_{S_{AM}(i)}$ 表示对应参考点的语义标签; $P(l_{S_D(i)})$ 表示语义分割的概率。

最后, 考虑点面残差大小, 引入核函数权重 C_{huber} 。观测点与参考点的距离超过距离阈值 σ , 则减小核函数权重 C_{huber} , 即减小这对匹配点对位姿的影响。核函数权重 C_{geometry} 由式(12)计算获得:

$$C_{\text{huber}}(r_i) = \begin{cases} 1, & |r_i| < \sigma \\ 0.5|r_i|^{-1}, & \text{其他} \end{cases} \quad (12)$$

式中: r_i 表示第 i 个观测点的残差; 文中设 σ 为 0.5。

对于树干、灯杆和汽车等物体,聚类分割获得的点云簇和提取的有向包围盒参数相对稳定,考虑物体级几何特征约束、语义特征约束和核函数约束;对于其他非地面点,如灌木丛、建筑物,聚类分割获得的点云簇变化较大,考虑引入语义特征约束和核函数约束;对于地面点只考虑核函数加权。

4 Object-SuMa 算法检验

算法检验实验平台参数配置如下:软件环境 Ubuntu20.04;处理器(CPU)Intel Core i7-8700,12核主频 3.2 GHz;内存 16 GB;显卡 GTX1080, 8 GB; CUDA11.0, TensorRT8.2.0,用于点云语义分割网络;OpenGL4.6,基于其图形渲染管线构建SLAM算法并行计算过程;第三方库为 PCL1.10, gtsam4.0。采用KITTI里程计数据集进行实验,里程计数据集有00~10号的基准序列,包含城市主干路、城市快速路等典型大尺度动态场景,环境中包含运动的车辆和行人等动态物体,能很好地评估 Object-SuMa 算法的性能。

4.1 定位精度对比分析

采用基准框架 SuMa++ 与 Object-SuMa 在室外大

场景 KITTI 数据集上进行实验,使用 evo 工具对比分析两种算法。以 KITTI09 序列为例,图 14 和图 15 分别为 Object-SuMa 和 SuMa++ 的俯视投影轨迹和位姿对比结果,可看出 Object-SuMa 得到的轨迹更接近真实轨迹。

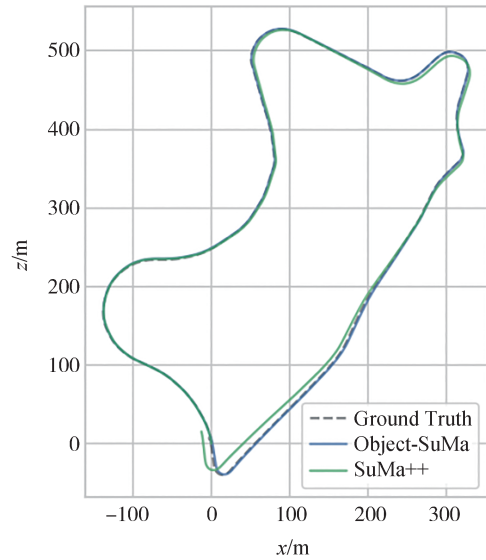


图 14 俯视投影轨迹对比

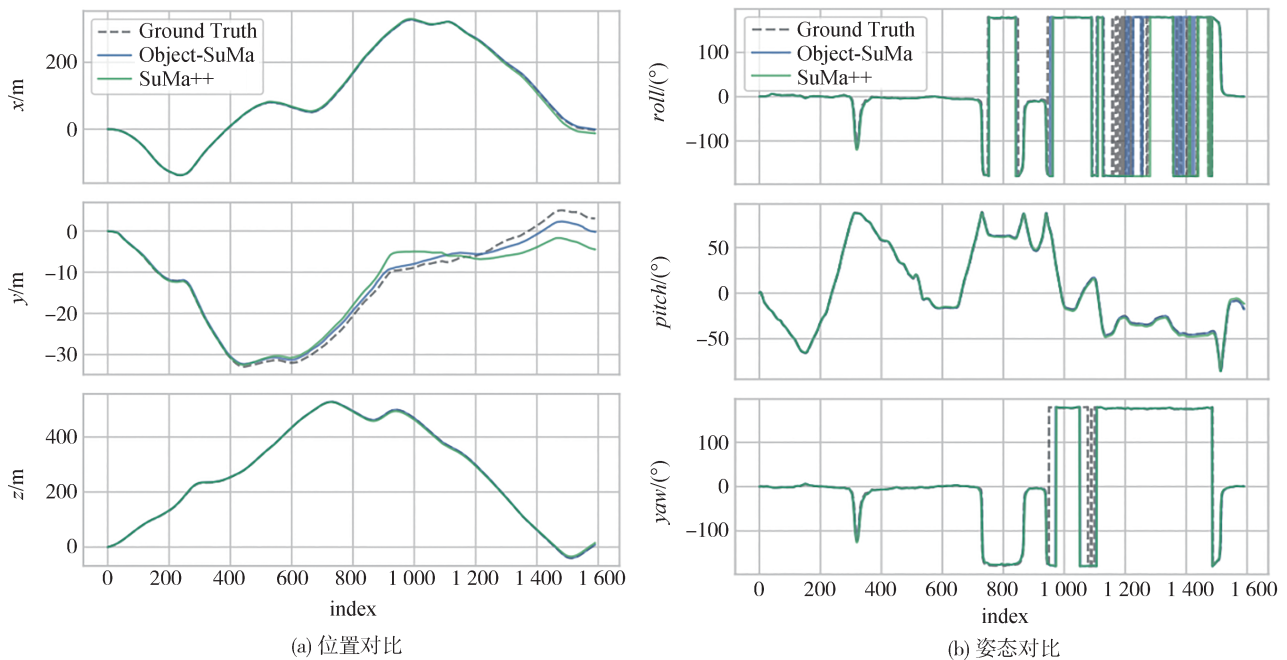


图 15 Object-SuMa 与 SuMa++ 位姿对比

图 16 (a) 和图 16 (b) 分别为 Object-SuMa 与 SuMa++ 在 KITTI09 上绝对位姿误差 APE (平移部分) 的随帧变化情况和整体分布情况。SuMa++ 整体均

方根误差 RMSE 为 6.19, 而 Object-SuMa 为 2.35, Object-SuMa 相比于 SuMa++ 显著提升了系统总体精度。

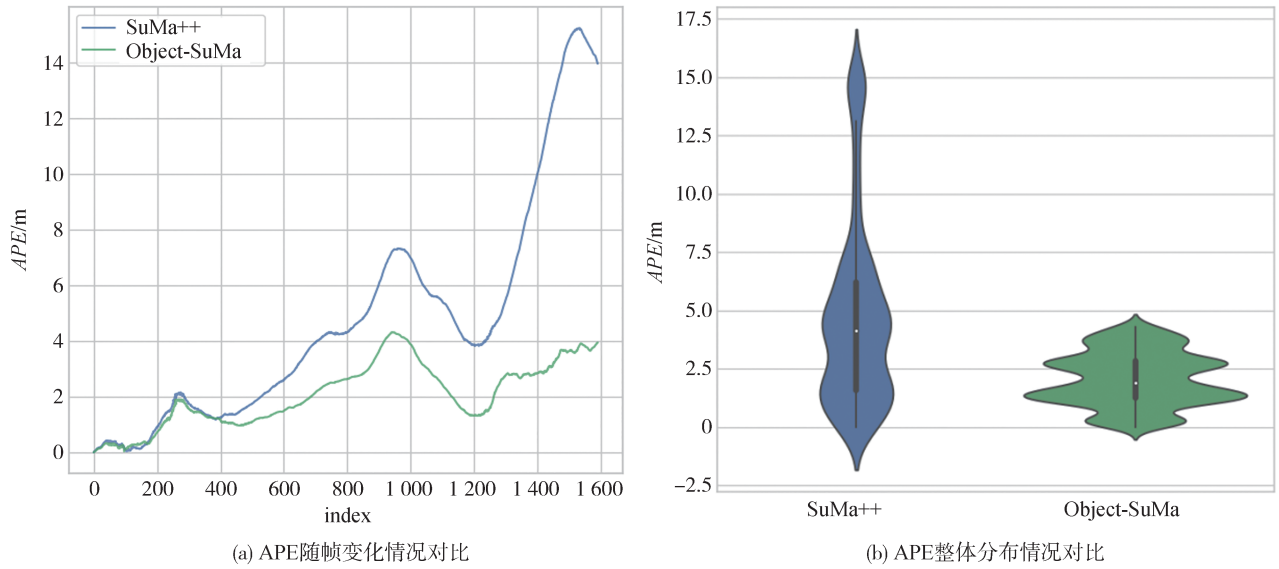


图16 Object-SuMa与SuMa++ APE对比

图17(a)和图17(b)分别为Object-SuMa和SuMa++在KITTI09上相对位姿误差RPE随轨迹的变化图。不同颜色代表不同RPE值,越接近红色,表明此处RPE较大;越接近蓝色,表明RPE越小。SuMa++算法的最大RPE为0.243,整体均方根误差RMSE为0.033;Object-SuMa算法的最大RPE为

0.160,整体均方根误差RMSE为0.028, RMSE减小了15.2%;可以看出, Object-SuMa的相对位姿估计精度高于SuMa++。Object-SuMa在配准过程中引入物体级几何特征,并引入先非地面点配准后地面点配准的两步配准策略,起到了改善里程计位姿估计精度的作用。

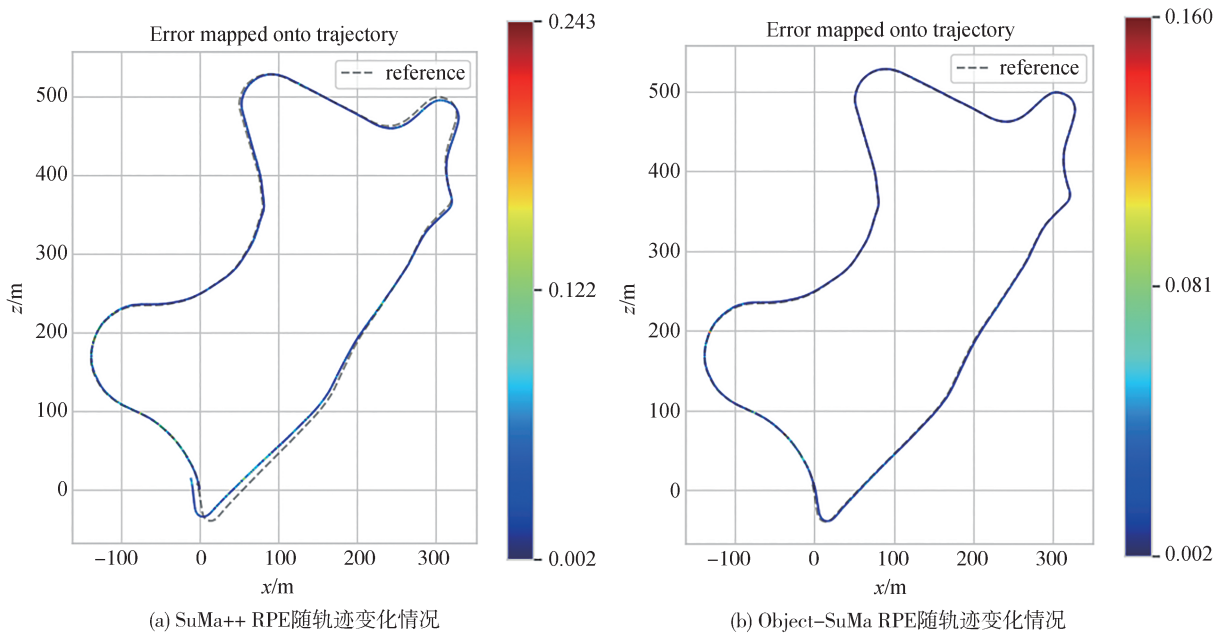


图17 Object-SuMa与SuMa++ RPE对比

表1对比了KITTI 00~10序列Object-SuMa与SuMa、SuMa++和SALTC^[22]算法的相对位姿估计精度。从表中可以看出,在01、02、03、04、09和10这些

含动态物体的序列中, Object-SuMa算法相较于SuMa++的精度提升明显。原因是动态场景下Object-SuMa对于汽车类点云配准时,考虑点云所在

物体方向包围盒的IOU作为物体级几何加权,减小了动态点误匹配静态点的可能。而在00、05、06、07、08这些含动态物体较少或不含动态物体的序列, Object-SuMa相较于SuMa++的精度有较小提升。原因是静态场景下 Object-SuMa的物体方向包围盒能够修正物体内部错误的语义分割,且两步匹配优

化的策略分割非地面点和地面点,减少了静态点误匹配。另外,从表中可以看出, Object-SuMa算法的相对旋转误差的平均值小于SALTC算法,而相对平移误差的平均值大于SALTC算法。SALTC算法在引入语义特征的基础上紧耦合了惯性测量单元的信息,因此平移部分估计精度高于 Object-SuMa算法。

表1 KITTI序列相对位姿精度对比

算法	误差	00	01	02	03	04	05	06	07	08	09	10	均值
SuMa	$RPE_R / ((^\circ) \cdot 100 \text{ m}^{-1})$	0.23	0.54	0.48	0.50	0.27	0.20	0.30	0.54	0.38	0.22	0.32	0.36
	$RPE_t / \%$	0.68	1.70	1.20	0.74	0.44	0.43	0.54	0.74	1.20	0.62	0.72	0.83
SuMa++	$RPE_R / ((^\circ) \cdot 100 \text{ m}^{-1})$	0.22	0.46	0.37	0.46	0.26	0.20	0.21	0.19	0.35	0.23	0.28	0.29
	$RPE_t / \%$	0.64	1.60	1.00	0.67	0.37	0.40	0.46	0.34	1.10	0.47	0.66	0.70
SALTC	$RPE_R / ((^\circ) \cdot 100 \text{ m}^{-1})$	0.20	0.39	0.31	0.46	0.29	0.22	0.21	0.26	0.29	0.26	0.26	0.28
	$RPE_t / \%$	0.49	0.73	0.52	0.59	0.42	0.33	0.24	0.31	0.84	0.53	0.41	0.54
Object-SuMa	$RPE_R / ((^\circ) \cdot 100 \text{ m}^{-1})$	0.21	0.39	0.31	0.40	0.21	0.19	0.18	0.17	0.32	0.19	0.23	0.25
	$RPE_t / \%$	0.64	1.44	0.92	0.61	0.31	0.38	0.42	0.31	1.07	0.41	0.59	0.64

4.2 配准时长对比分析

图18为 Object-SuMa与SuMa++在KITTI 09序列上的ICP配准时长随帧变化对比图。SuMa++的平均时长为13.8 ms, Object-SuMa为11.4 ms,下降17.3%。另外, SuMa++算法的曲线波动较大,个别帧的配准时长甚至超过40 ms,而 Object-SuMa曲线波动较小,未出现明显的波峰,配准时长表现更稳定。先非地面点配准后地面点配准的两步配准策略,使得计算复杂度下降,ICP配准速度更快,耗时也更稳定。

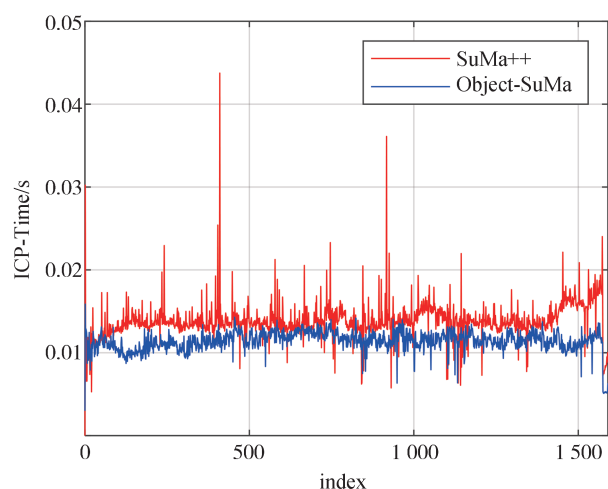


图18 ICP配准时长随帧变化对比图

5 结论

(1)在预处理过程,分割物体、求解物体的包围

盒位姿尺寸作为点云物体级几何特征并编码为纹理,为基于图形渲染管线的并行SLAM算法引入物体级位姿尺寸信息;同时基于包围盒修正物体内部错误的点云语义分割,提升了语义在点云配准时的置信度。

(2)在ICP配准阶段,分解计算方向包围盒的IOU,为汽车和杆状物类点云的配准残差引入物体级几何加权,动态场景下的相对位姿估计精度提升超15%;分为地面点和非地面点两步优化,降低计算复杂度,ICP平均耗时下降超17%。

(3) Object-SuMa考虑了树干、汽车类物体的位姿尺寸信息,在未来的工作中可以深入研究多类别物体的在线精确分割重建与跟踪,并引入到SLAM配准和优化过程。

参考文献

- [1] SHAN T, ENGLÓ B. LeGO-LOAM: lightweight and ground optimized lidar odometry and mapping on variable terrain [C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018:4758-4765.
- [2] XU W, ZHANG F. FAST-LIO: a fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter [J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3317-3324.
- [3] PENG H, ZHAO Z, WANG L. A review of dynamic object filtering in SLAM based on 3D LiDAR [J]. Sensors, 2024, 24(2):645.
- [4] QIAN C, ZHAO X, et al. RF-LIO: removal-first tightly-coupled lidar inertial odometry in high dynamic environments [C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021: 4421-4428.

- [5] 陈耀华, 何丽, 王宏伟, 等. 基于可视点法剔除动态目标的激光-惯导SLAM [J]. 仪器仪表学报, 2023, 44 (9): 248-256.
CHEN Yaohua, HE Li, WANG Hongwei, et al. LiDAR-inertial SLAM based on visible point method to remove dynamic objects [J]. Chinese Journal of Scientific Instrument, 2023, 44 (9): 248-256.
- [6] KIM G, KIM A. Remove, then revert: static point cloud map construction using multiresolution range images [C]. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 10758-10765.
- [7] SCHAUER J, NUECHTER A. The peopleremover-removing dynamic objects from 3-D point cloud data by traversing a voxel occupancy grid [J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1679-1686.
- [8] JOACHIM Gehrungab, HEBELA M, ARENSA M, et al. An approach to extract moving objects from MLS data using a volumetric back-ground representation [J]. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2017, 4.
- [9] FANG Y, QIAN K, ZHANG Y, et al. Segmented curved-voxel occupancy descriptor for dynamic-aware LiDAR odometry and mapping [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-16.
- [10] RUCHTI P, BURGARD W. Mapping with dynamic-object probabilities calculated from single 3D range scans [C]. 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018: 6331-6336.
- [11] 李继文. 面向城市环境智能车辆视觉位姿估计方法研究 [D]. 广州: 华南理工大学, 2023.
LI Jiwen. Research on the visual pose estimation method of intelligent vehicles in urban environment [D]. Guangzhou: South China University of Technology, 2023.
- [12] PFREUNDSCHUH P, HENDRIKX H F C, REIJGWART V, et al. Dynamic object aware LiDAR SLAM based on automatic generation of training data [C]. 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021: 11641-11647.
- [13] BEHLEY J, STACHNISS C. Efficient surfel-based SLAM using 3D laser range data in urban environments [C]. 2018 Robotics: Science and System, 2018.
- [14] CHEN X, MILIOTO A, PALAZZOLO E, et al. SuMa++: efficient LiDAR-based semantic SLAM [C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021.
- [15] LI L, KONG X, ZHAO X, et al. SA-LOAM: semantic-aided LiDAR SLAM with loop closure [C]. 2021 IEEE International Conference on Robotics and Automation (ICRA) 2021. IEEE, 2021: 7627-7634.
- [16] ZHANG J. Surfels: surface elements as rendering primitives [J]. ACM Special Interest Group on Computer Graphics (ACM SIGGRAPH), 2000, 7(1): 335-342.
- [17] HESS W, KOHLER D, RAPP H, et al. Real-time loop closure in 2D LIDAR SLAM [C]. 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2016: 1271-1278.
- [18] MILIOTO A, VIZZO I, BEHLEY J, et al. RangeNet++: fast and accurate LiDAR semantic segmentation [C]. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019.
- [19] ZHANG W, QI J, WAN P, et al. An easy-to-use airborne LiDAR data filtering method based on cloth simulation [J]. Remote Sensing, 2016, 8(6): 501.
- [20] HANDA A. Simplified jacobians in 6-DoF camera tracking [J]. 2014.
- [21] SUTHERLAND I E, HODGMAN G W. Reentrant polygon clipping [J]. Communications of the ACM, 1974.
- [22] LIU Peng, BI Yuxuan, SHI Jialin, et al. Semantic-assisted LiDAR tightly coupled SLAM for dynamic environments [J]. IEEE Access, 2024, 12: 34042-34053.