

# 基于 LVI-SAM-Stereo 的多传感器融合室内外建图定位

蒋风洋<sup>1,2</sup>,程瑶<sup>1,2\*</sup>,韩哲<sup>2</sup>,王怀震<sup>1,2</sup>,周风余<sup>3</sup>,董磊<sup>2</sup>

(1.山东新一代信息产业技术研究院有限公司,山东 济南 250102; 2.浪潮智能终端有限公司,山东 济南 250101; 3.山东大学控制科学与工程学院,山东 济南 250061)

**摘要:**针对机器人在室内外建图定位精度低、场景适应性差的问题,提出一种基于双目相机的紧耦合激光雷达(light detection and ranging, LiDAR)视觉惯性里程计平滑建图定位(tightly-coupled LiDAR-visual-inertial odometry via smoothing, mapping, and relocalization by stereo, LVI-SAM-Stereo)方法。采用点线和点面距离构建激光雷达惯性位姿估计模型;利用多传感器信息交互实现双目惯性里程计快速初始化,基于最小化重投影误差优化里程计位姿;提出融合 Scan-Context 与视觉特征的跨模态回环检测机制,有效减少错误回环;构建重定位双向优化架构,将因子图优化的里程计信息用作视觉跟踪的初始位姿估计,基于多点透视(perspective- $n$ -point, PnP)求解视觉位姿辅助激光雷达点云配准。通过数据集和真实场景的大量试验,相较于紧耦合激光雷达视觉惯性里程计平滑建图(tightly-coupled LiDAR inertial odometry via smoothing and mapping, LIO-SAM)方法和紧耦合激光雷达视觉惯性里程计平滑建图(tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping, LVI-SAM)方法,LVI-SAM-Stereo 方法在室外场景的建图精度分别提升 3.10%和 5.97%,在室内场景的平均漂移分别降低 72.7%和 43.05%,建图精度和场景适应性显著提升。重定位满足机器人自主导航的工程需求。

**关键词:**机器人;多传感器融合;视觉惯性里程计;回环检测;重定位

**中图分类号:**TP391

**文献标志码:**A

**引用格式:**蒋风洋,程瑶,韩哲,等. 基于 LVI-SAM-Stereo 的多传感器融合室内外建图定位[J]. 山东大学学报(工学版),2025,55(4):72-83.

JIANG Fengyang, CHENG Yao, HAN Zhe, et al. The multi-sensor fusion mapping and relocalization based on LVI-SAM-Stereo in indoor and outdoor scenes[J]. Journal of Shandong University (Engineering Science), 2025, 55(4):72-83.

## The multi-sensor fusion mapping and relocalization based on LVI-SAM-Stereo in indoor and outdoor scenes

JIANG Fengyang<sup>1,2</sup>, CHENG Yao<sup>1,2\*</sup>, HAN Zhe<sup>2</sup>, WANG Huaizhen<sup>1,2</sup>, ZHOU Fengyu<sup>3</sup>, DONG Lei<sup>2</sup>

(1. Shandong New Generation Information Industrial Technology Research Institute, Jinan 250102, Shandong, China; 2. Inspur Intelligent Terminal Co., Ltd., Jinan 250101, Shandong, China; 3. School of Control Science and Engineering, Shandong University, Jinan 250061, Shandong, China)

**Abstract:** Aiming at the problems of low mapping and relocalization accuracy, as well as poor scene adaptability, for robots in indoor and outdoor scenes, a tightly-coupled light detection and ranging (LiDAR)-visual-inertial odometry via smoothing, mapping, and relocalization by stereo (LVI-SAM-Stereo) method was proposed. The LiDAR-inertial pose estimation model was constructed by utilizing point-line and point-plane distances. Multi-sensor information interaction enabled rapid initialization of stereo-inertial odometry, with the odometry pose being optimized through reprojection error minimization. A cross-modal loop closure detection mechanism combining Scan-Context with visual features effectively reduced incorrect loop closures. A bidirectional relocalization architecture was developed, where factor graph-optimized odometry provided initial pose estimation for visual tracking, while perspective- $n$ -point (PnP)-derived visual poses assisted LiDAR point cloud registration. A thorough evaluation with both datasets and real-world experiments verified that LVI-SAM-Stereo achieved 3.10% and 5.97% higher outdoor mapping accuracy

**收稿日期:**2025-01-09

**基金项目:**山东省重点研发计划(竞争性创新平台)资助项目(2023CXPT094);山东省重点研发计划(重大科技创新工程)资助项目(2024CXGC010213);济南市第二批市校融合发展战略工程资助项目(JNSX2023012)

**第一作者简介:**蒋风洋(1997—),男,山东济南人,硕士,主要研究方向为机器人建图定位导航。E-mail: jiangfy@inspur.com

**\*通信作者简介:**程瑶(1985—),女,山东济南人,正高级工程师,博士,主要研究方向为机器人建图定位导航。

E-mail: chengyao01@inspur.com

compared to tightly-coupled LiDAR inertial odometry via smoothing and mapping (LIO-SAM) and tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping (LVI-SAM), respectively. Indoor average drift decreased by 72.7% and 43.05% versus these benchmarks. The system significantly improved mapping precision and scene adaptability. The relocalization satisfied the engineering requirements for autonomous navigation of robot products.

**Keywords:** robot; multi-sensor fusion; visual-inertial odometry; loop closure detection; relocalization

## 0 引言

即时定位与地图构建(simultaneous localization and mapping, SLAM)技术广泛应用于机器人、无人机和智能驾驶等领域<sup>[1]</sup>。激光雷达(light detection and ranging, LiDAR)具有精度高、不受光照影响的特点,但过于依赖场景的几何结构。视觉相机能够获取丰富的场景信息,但由于图像模糊、初始化困难、对光照变化具有敏感性,建图效果并不稳健<sup>[2]</sup>。

为克服激光雷达SLAM和视觉SLAM的不足,多传感器融合的SLAM逐步成为机器人领域的研究热点<sup>[3]</sup>。文献[4]提出一种用于视觉、视觉惯性和多地图SLAM的精确开源库,充分发挥视觉相机与惯性测量单元(inertial measurement unit, IMU)的组合优势,有效改善快速转向、纹理缺失场景下的建图效果,但初始化过程较为严苛;文献[5]提出紧耦合激光雷达惯性里程计平滑建图(tightly-coupled LiDAR inertial odometry via smoothing and mapping, LIO-SAM)方法,利用IMU数据去除激光雷达点云畸变,并将IMU数据作为激光雷达帧间位姿变换的初值,但在长走廊等几何结构单一的场景中易发生退化;文献[6]提出紧耦合激光雷达视觉惯性里程计平滑建图(tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping, LVI-SAM)方法,在LIO-SAM方法的基础上融合一种鲁棒通用的单目视觉惯性状态估计器(robust and versatile monocular visual-inertial state estimator, VINS-Mono)<sup>[7]</sup>,构成视觉激光雷达惯性耦合系统,但VINS-Mono在特定场景下的尺度缺失和LIO-SAM易产生错误激光雷达回环的不足,使室外场景下的建图鲁棒性不佳;文献[8]提出使用点和线特征的激光雷达视觉惯性里程计,在LVI-SAM中引入线特征提升位姿估计精度;文献[9]提出一种基于去除动态目标点云的改进LVI-SAM算法;文献[10]基于LVI-SAM提出一种特征点深度恢复模型,实现特征点深度鲁棒估计。上述方法虽在一定程度上提升了机器人的建图精度,但仍局限于特定建图场景,激光雷达回环检测仍需针对不同场景手动调

整回环参数,易产生错误回环,影响建图质量,进行大规模室内外场景建图的试验效果仍有待验证。

机器人基于SLAM完成场景建图后,需再通过重定位求解其在场景中的位姿。这是进行自主导航的前提。当前的商用机器人产品通常采用激光雷达当前帧与点云地图配准的方式获取重定位位姿<sup>[11]</sup>,存在精度低、稳定性差的问题,给机器人产品普及带来挑战。

针对机器人当前建图与重定位精度低、场景适应性差的问题,本研究提出一种基于双目相机的紧耦合激光雷达视觉惯性里程计平滑建图定位(tightly-coupled LiDAR-visual-inertial odometry via smoothing, mapping, and relocalization by stereo, LVI-SAM-Stereo)方法,提升视觉惯性里程计位姿估计的鲁棒性,避免激光雷达点云的错误回环,提高自主导航的重定位精度,增强机器人在室内外场景的适应性。

## 1 系统整体框架

基于LVI-SAM-Stereo的建图定位方法示意图如图1所示。受LVI-SAM的启发,本研究所提LVI-SAM-Stereo由激光雷达惯性系统和视觉惯性系统两个子系统构成,加强多传感器间的信息交互能力。在激光雷达惯性系统中,激光雷达点云经下采样后分别提取边缘特征点与平面特征点,在视觉惯性系统初始化前,采用IMU预积分后的数据作为初始位姿,去除点云畸变,利用点线和点面距离求解激光雷达惯性里程计位姿。在视觉惯性系统中,提取视觉特征点,采用激光雷达惯性里程计位姿加速基于双目相机与IMU耦合的视觉惯性里程计初始化,提升视觉帧间跟踪精度。通过跟踪构建局部地图,进一步优化里程计位姿,为激光雷达惯性系统的点云配准提供比IMU预积分更可靠的初始位姿估计。本研究提出视觉校验激光雷达回环方法,激光雷达数据经Scan-Context进行初步回环检测后,需再次经视觉特征匹配加以校验,有效减少错误回环,减轻激光雷达回环检测对参数的依赖。分别构建激光雷达里程计因子、回环因子和IMU预积

分因子,进行多传感器因子图优化,获得全局里程计和全局地图。其中,全局里程计为视觉下一帧跟踪局部地图提供初始位姿,提升重定位时视觉地图与全局地图的重合度。重定位时获取当前视觉帧提取带方向的加速分割测试特征与旋转的二进制鲁棒独立基本特征(oriented feature from

accelerated segment test and rotated binary robust independent elementary features, ORB)特征点,与建图所得视觉地图进行特征匹配,通过多点透视(perspective- $n$ -point, PnP)求解视觉位姿<sup>[12]</sup>,以该位姿为初始估计辅助激光雷达当前帧与全局地图进行点云配准,获取机器人当前的精准定位位姿。

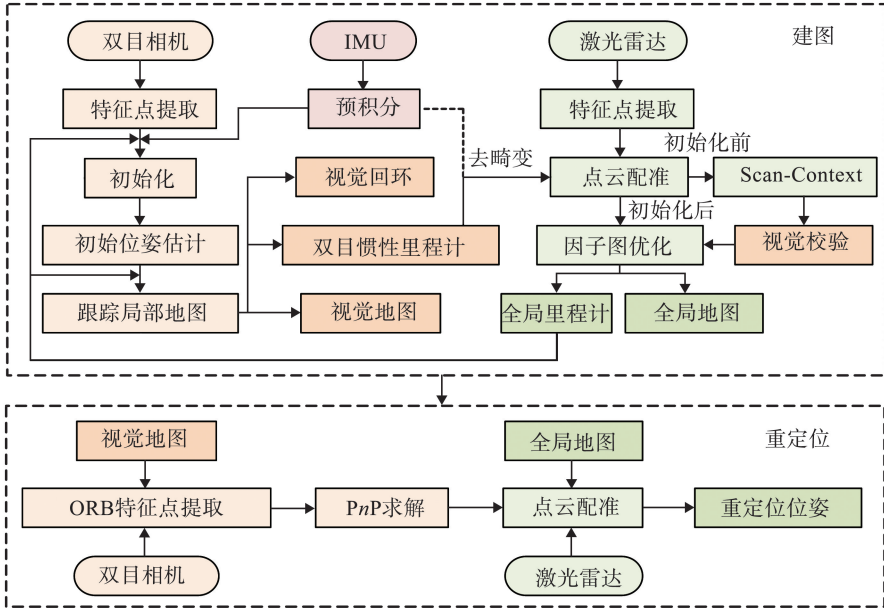


图1 基于LVI-SAM-Stereo的建图定位方法示意图

Fig.1 The system structure of the mapping and relocalization framework based on LVI-SAM-Stereo

## 2 LVI-SAM-Stereo 原理

### 2.1 多传感器标定与时间同步

多传感器联合标定实现了不同传感器间的空间同步,是数据融合的前提。采用开源工程 LiDAR\_IMU\_calib 实现激光雷达与 IMU 的联合标定<sup>[13]</sup>。激光雷达、双目相机与 IMU 数据间的时间同步采用机器人操作系统(robot operating system, ROS)的时间同步特性<sup>[14]</sup>,在话题消息发布和订阅时自动对齐时间戳,确保各传感器间精准的数据同步,为数据融合和处理提供准确的数据支撑。

### 2.2 双目惯性里程计

为避免单目惯性里程计易失去尺度信息或出现漂移的问题,本研究采用双目相机与 IMU 构建双目惯性里程计,为激光雷达惯性系统帧间跟踪提供鲁棒的位姿估计。本研究提出激光雷达辅助双目初始化方法,解决双目惯性里程计在机器人匀速行驶或位于纹理单一场景中初始化困难的问题。在激光雷达惯性系统中,对采集的激光雷达点云数据下采样,每帧点云基于平滑度分别提取边缘特征点和平面特征点。点云平滑度

$$c = \frac{1}{r \|\mathbf{x}_{(h,i)}^L\|} \left\| \sum_{i,j \in \xi, j \neq i} (\mathbf{x}_{(h,i)}^L - \mathbf{x}_{(h,j)}^L) \right\|, \quad (1)$$

式中,  $\mathbf{x}_{(h,i)}^L$  和  $\mathbf{x}_{(h,j)}^L$  分别为激光雷达第  $h$  次扫描点  $i$  和点  $j$  在激光雷达坐标系下的位置向量,  $\xi$  为激光雷达第  $h$  次扫描的点集,  $r$  为点集  $\xi$  中点的数量。根据  $c$  的取值依次对激光雷达扫描点进行排序,选取最大平滑度点为边缘特征点,最小平滑度点为平面特征点。在视觉惯性系统初始化前采用 IMU 预积分数据作为激光雷达惯性系统中帧间匹配的初始位姿,采用点与线和点与面之间的距离关系,建立位姿估计模型。第  $k$  个边缘特征点和第  $k$  个平面特征点的变换距离分别为

$$d_k^e = \frac{\|(\mathbf{p}_{t+1,k}^e - \mathbf{p}_{t,u}^e) \times (\mathbf{p}_{t+1,k}^e - \mathbf{p}_{t,v}^e)\|}{\|\mathbf{p}_{t,u}^e - \mathbf{p}_{t,v}^e\|}, \quad (2)$$

$$d_k^p = \frac{\left\| \begin{array}{c} (\mathbf{p}_{t+1,k}^p - \mathbf{p}_{t,u}^p) \\ (\mathbf{p}_{t,u}^p - \mathbf{p}_{t,v}^p) \times (\mathbf{p}_{t,u}^p - \mathbf{p}_{t,w}^p) \end{array} \right\|}{\|(\mathbf{p}_{t,u}^p - \mathbf{p}_{t,v}^p) \times (\mathbf{p}_{t,u}^p - \mathbf{p}_{t,w}^p)\|}, \quad (3)$$

式中,  $\mathbf{p}_{t+1,k}^e$ 、 $\mathbf{p}_{t+1,k}^p$  分别为  $t+1$  时刻第  $k$  个边缘特征点向量和平面特征点向量,  $\mathbf{p}_{t,u}^e$ 、 $\mathbf{p}_{t,v}^e$  分别为  $t$  时刻第  $u$  个边缘特征点向量和第  $v$  个边缘特征点向量,  $\mathbf{p}_{t,u}^p$ 、 $\mathbf{p}_{t,v}^p$ 、 $\mathbf{p}_{t,w}^p$  分别为  $t$  时刻第  $u$  个平面特征点向量、

第  $v$  个平面特征点向量和第  $w$  个平面特征点向量。采用高斯牛顿法通过  $d_k^e$  和  $d_k^p$  最小化目标误差函数  $f$ , 即

$$f = \min \left\{ \sum_{T_{t+1}} d_k^e + \sum_{p_{t+1,k}^e} d_k^p \right\}, \quad (4)$$

式中  $T_{t+1}$  为  $t+1$  时刻的位姿变换矩阵。计算  $t+1$  时刻相对  $t$  时刻的相对位姿变换矩阵

$$\Delta T_{t,t+1} = T_t^T T_{t+1}, \quad (5)$$

式中  $T_t$  为  $t$  时刻的位姿变换矩阵。通过激光雷达惯性位姿及传感器外参构造视觉初始关键帧和地图点,完成视觉惯性系统的初始化过程。同时,激光雷达惯性位姿还用作视觉惯性系统下一帧帧间跟踪的初始位姿,保证视觉惯性地图与全局地图高度重合,为视觉激光雷达耦合重定位做铺垫。

视觉惯性系统初始化后,双目图像提取 ORB 特征点,若机器人行驶过程中因快速运动或遮挡等导致跟踪到的特征点数量不足时,通过插入关键帧或匹配当前帧和参考关键帧之间的特征点恢复更多地图点,以提升视觉帧间跟踪的稳定性。采用基于随机采样一致(random sample consensus, RANSAC)算法去除异常匹配点<sup>[15]</sup>,以提升跟踪精度。基于 PnP 求解双目惯性里程计初始位姿,通过最小化地图点在当前帧中的重投影误差进一步优化位姿。将待优化的相机位姿向量  $\zeta_o$  ( $o=1, 2, \dots, m$ ) 和地图点位置向量  $p_g$  ( $g=1, 2, \dots, n$ ) 组成优化矩阵

$$a = [\zeta_1 \ \zeta_2 \ \dots \ \zeta_m \ p_1 \ p_2 \ \dots \ p_n]^T. \quad (6)$$

以待优化的李代数表示的相机位姿向量

$$\zeta^* = \arg \min_{\zeta} \frac{1}{2} \sum_{g=1}^n \left\| u_g - \frac{1}{s_g} K \exp(\zeta^\wedge) p_g \right\|_2^2, \quad (7)$$

式中,  $u_g$  为特征点的像素坐标向量,  $s_g$  为尺度信息,  $K$  为相机内参矩阵,  $\zeta$  为相机位姿向量,  $\zeta^\wedge$  为相机位姿向量的反对称矩阵。

将双目惯性里程计位姿用作激光雷达惯性系统帧间跟踪的初始位姿估计,避免在单一几何结构场景中因长期依赖 IMU 预积分而产生较大累积误差,造成里程计退化或漂移,以此提升激光雷达惯性系统的建图鲁棒性。双目惯性里程计位姿需进行如下坐标系变换:

$$T_{W-C} = T_{W-Worb} \cdot T_{C-Worb}^{-1}, \quad (8)$$

$$T_{W-L} = T_{W-C} \cdot T_{C-L} \cdot T_{L-I}, \quad (9)$$

式中,  $T_{W-C}$  为相机左目坐标系到世界坐标系的变换矩阵,  $T_{W-Worb}$  为视觉惯性坐标系到世界坐标系的变换矩阵,  $T_{C-Worb}$  为视觉惯性坐标系到相机左目坐标系的变换矩阵,  $T_{W-L}$  为激光雷达坐标系到世界坐标系的变换矩阵,  $T_{C-L}$  为 IMU 坐标系到相机左目坐标

系的变换矩阵,  $T_{L-I}$  为激光雷达坐标系到 IMU 坐标系的变换矩阵。

随着行驶距离增加,里程计的累积误差逐步增大,回环检测是实现校正里程计漂移的有效方式。但在回环时,双目惯性里程计会产生较大的位姿突变,使传入激光雷达惯性系统的位姿严重失真,造成建图畸变。因此,本研究提出双目惯性里程计位姿矫正方法。通过监听视觉回环信号,计算因回环产生的位姿变换矩阵

$$T_{loop} = P_{after} \cdot P_{before}^{-1}, \quad (10)$$

式中,  $P_{before}$  和  $P_{after}$  分别为回环前后的位姿矩阵。回环后新的位姿矩阵

$$P_{new} = T_{loop} \cdot P_{all}, \quad (11)$$

式中  $P_{all}$  为双目惯性里程计中所有位姿矩阵。

上述矫正方法能够避免双目惯性里程计位姿失真,为激光雷达惯性系统的帧间跟踪提供鲁棒的初始位姿估计。

### 2.3 视觉校验激光雷达回环

在室外空旷区域或室内长走廊等几何结构单一的场景中,激光雷达点云易产生错误回环,降低建图质量。因此,本研究提出视觉校验激光雷达回环的跨模态回环检测机制,提升建图性能。构建 Scan-Context 描述子计算激光雷达关键帧间相似度,相似度函数

$$d(I^q, I^c) = \frac{1}{N_s} \sum_{\alpha=1}^{N_s} \left( 1 - \frac{c_\alpha^q \cdot c_\alpha^c}{\|c_\alpha^q\| \cdot \|c_\alpha^c\|} \right), \quad (12)$$

式中:  $I^q$  和  $I^c$  为两个 Scan-Context 描述子矩阵;  $c_\alpha^q$  和  $c_\alpha^c$  分别为  $I^q$  和  $I^c$  的第  $\alpha$  列向量,  $\alpha=1, 2, \dots, N_s$ 。将  $c_\alpha^q$  和  $c_\alpha^c$  的余弦距离作为第  $\alpha$  列的对比值,每列对比值之和作为两个描述子的相似度,满足相似度函数的激光雷达关键帧需进行视觉特征点校验。激光雷达闭环帧的回环校验采用汉明距离  $h(d_p, d_q)$  比较视觉特征描述子向量,计算式为

$$h(d_p, d_q) = \sum_{\beta=1}^n (d_{p\beta} \oplus d_{q\beta}), \quad (13)$$

式中,  $d_p$  和  $d_q$  分别为两个视觉帧对应特征点  $p$  和  $q$  的二进制特征描述子向量,  $d_{p\beta}$  和  $d_{q\beta}$  分别为  $d_p$  和  $d_q$  中的元素,  $n$  为特征描述子向量的维度。对每个描述子向量  $d_p$  找到视觉回环帧的最近邻点描述子向量  $d_q$  和次近邻点描述子向量  $d_s$ , 分别计算汉明距离  $h(d_p, d_q)$  和  $h(d_p, d_s)$ , 二者之比若小于阈值,则接受该特征匹配。不同时间和视角下视觉特征点匹配情况如图 2 所示。在室内外场景中,依据视觉特征,当前帧能够正确匹配处于不同光线和视角下机器人先前经过的位置点。

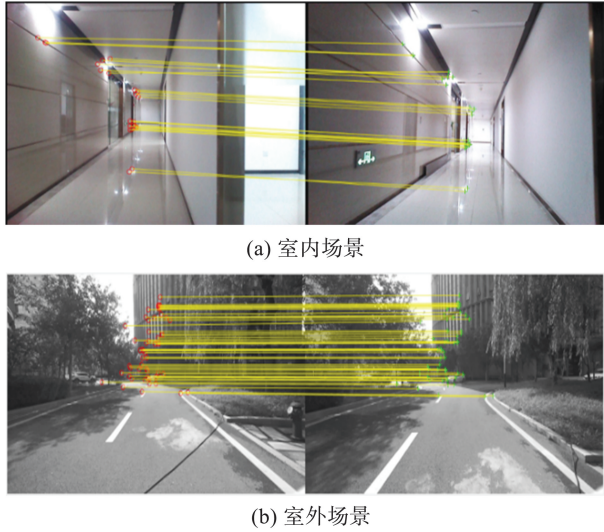


图2 不同时间和视角下视觉特征点匹配情况  
Fig.2 The matching of visual feature points at different times and from various perspectives

当执行激光雷达回环检测时,视觉当前帧与回环帧特征点匹配数目是决定激光雷达回环正确与否的关键。在室内环形走廊场景建图一圈后,视觉当前帧与历史关键帧 ORB 特征点的匹配情况如图 3 所示。由图 3 可以看出,视觉当前帧与历史关键帧序号 20 附近(即同一地点)的特征点匹配数量显著高于和其他历史关键帧(即不同地点)特征点的匹配数量,充分体现以视觉特征点匹配作为校验激光雷达回环的合理性。视觉校验激光雷达回环流程如算法 1 所示。

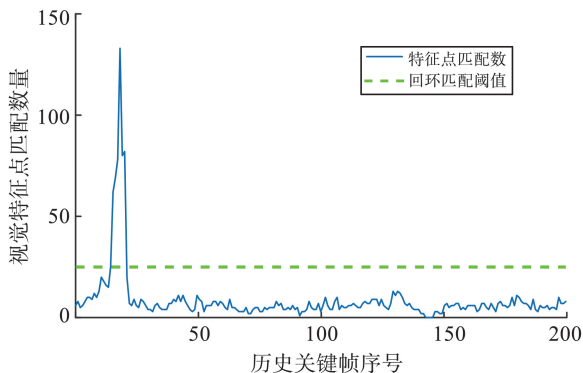


图3 视觉当前帧与历史关键帧 ORB 特征点匹配情况  
Fig.3 Matching status of ORB feature points between the current visual frame and historical key frames

#### 算法 1 视觉校验激光雷达回环

输入:激光雷达当前帧点云  $C_{now}$ 、激光雷达回环帧点云  $C_{loop}$ 、激光雷达回环阈值  $t_{lidar}$ 、视觉特征点匹配阈值  $t_{visual}$ 。

输出:检验状态  $s$ 。

- (1)  $d_{loop} \leftarrow C_{now}$  和  $C_{loop}$  的相似度;
- (2) if  $d_{loop} > t_{lidar}$  then;
- (3)  $V_{now} \leftarrow$  与  $C_{now}$  时间戳对应的视觉帧;

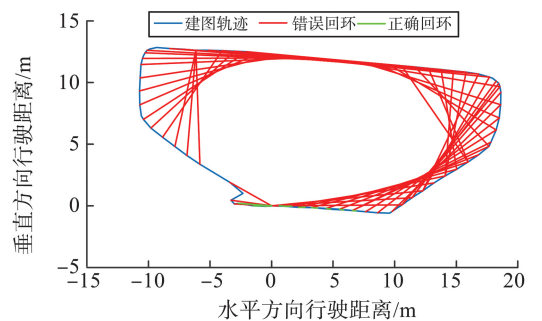
- (4)  $V_{loop} \leftarrow$  与  $C_{loop}$  时间戳对应的视觉帧;
- (5)  $n_{orb} \leftarrow V_{now}$  与  $V_{loop}$  的 ORB 特征点匹配数;
- (6) if  $n_{orb} > t_{visual}$  then;
- (7)  $s \leftarrow true$ ;
- (8) else;
- (9)  $s \leftarrow false$ ;
- (10) end if;
- (11) else;
- (12)  $s \leftarrow false$ ;
- (13) end if;
- (14) end。

经视觉校验激光雷达回环后的回环信息构成回环因子,与激光雷达里程计因子和 IMU 预积分因子一同进行多传感器因子图优化。当机器人新的状态节点进入因子图时,通过调整状态节点取值最小化因子图中所有因子的误差和。机器人状态节点集合

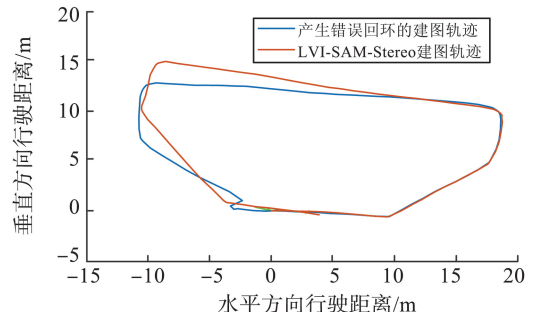
$$X = \{R^T, p^T, v^T, b^T\}, \quad (14)$$

式中,  $R$  和  $p$  分别为机器人本体坐标系到世界坐标系的旋转矩阵和位置向量,  $v$  为速度向量,  $b$  为 IMU 偏差向量。

激光雷达回环对建图轨迹的影响如图 4 所示。由图 4(a) 可以看出,在室内长走廊场景中, LVI-SAM 采用单纯的激光雷达回环检测,大量错误回环导致建图轨迹畸变;由图 4(b) 可以看出, LVI-SAM-Stereo 采用视觉校验激光雷达回环,有效避免错误回环,建图轨迹更加精准,回环参数对建图性能的影响降低,场景适应性极大提高。



(a) LVI-SAM 中激光雷达错误回环对建图轨迹的影响



(b) 激光雷达错误回环存在与否对建图轨迹的影响

图4 激光雷达回环对建图轨迹的影响

Fig.4 The impact of LiDAR loop closures on the mapping trajectory

### 2.4 视觉激光雷达耦合重定位

对机器人所处场景构建完整地图后,通过重定位求解机器人在当前场景中的位姿,为实现自主导航提供精准定位支持。当前以激光雷达配准为代表的机器人重定位算法存在精度低、稳定性差的问题,因此,本研究提出视觉激光雷达耦合重定位双向优化架构,如图5所示。

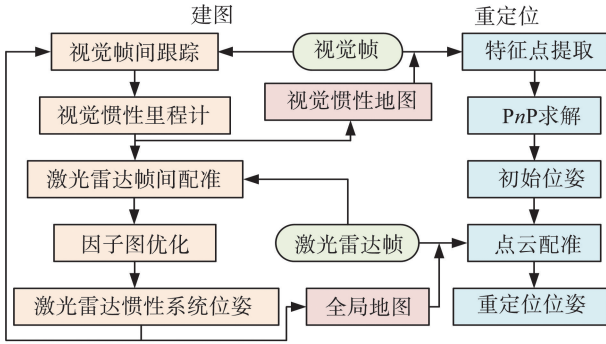


图5 视觉激光雷达耦合重定位流程图

Fig.5 Flow chart of visual-LiDAR fusion relocalization

如2.2节所述,建图时,将激光雷达当前帧位姿传递给视觉惯性系统,用作双目惯性里程计下一帧进行帧间跟踪的初始位姿,保证视觉惯性地图与全局地图的高度重合。重定位时,在视觉惯性系统中提取视觉当前帧ORB特征点。ORB算法通过构建图像金字塔,检测图像金字塔每层角点,当相机运动时,匹配不同层的图像实现尺度不变性,基于灰度质心法实现旋转不变性。灰度质心法计算图像块的矩

$$m_{\gamma\delta} = \sum_{x,y} x^\gamma y^\delta I(x,y), \quad (15)$$

式中: $I(x,y)$ 为像素坐标 $(x,y)$ 的图像灰度; $\gamma$ 和 $\delta$ 为矩的阶数, $\gamma,\delta \in \{0,1\}$ 。通过图像块的不同矩计算得出图像块的质心坐标

$$c = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right), \quad (16)$$

式中, $m_{00}$ 为图像块的零阶矩, $m_{10}$ 和 $m_{01}$ 分别为相对于 $x$ 和 $y$ 的一阶矩。特征点的主方向为圆心到图像

块质心 $c$ 的方向,特征点的旋转角

$$\theta = \arctan(m_{01}, m_{10}). \quad (17)$$

通过上述过程完成ORB特征点提取,然后通过预先训练的词袋模型(bag of words, BoW)计算基于词袋的相似度得分,实现快速检索候选关键帧。采用ORB描述子的汉明距离进行候选关键帧的特征匹配,结合RANSAC算法剔除异常匹配,基于PnP求解当前帧位姿矩阵 $P_{emp}$ ,作为视觉激光雷达耦合重定位的初始位姿估计。当激光雷达惯性系统获取新的状态 $X_{next}$ 时,基于视觉先验位姿搜索因子图,获取最接近 $X_{next}$ 的先验状态 $X_{emp}$ ,尝试将激光雷达当前帧与子关键帧组进行点云匹配,获取相对位姿变换矩阵 $\Delta T_{next,emp}$ ,求取最终重定位的精准位姿矩阵

$$P_{next} = \Delta T_{next,emp} \cdot P_{emp} \circ \quad (18)$$

### 3 试验与评估

在室内外不同场景中,采用Hilti SLAM数据集<sup>[16]</sup>及机器人测试平台验证所提LVI-SAM-Stereo相较于LVI-SAM和LIO-SAM的场景适应性及建图定位精度。所有试验采用的硬件环境为Intel i7-1165G7 2.80 GHz处理器、16 GB内存、NVIDIA GeForce RTX 2060显卡,软件环境为Ubuntu18.04操作系统。

#### 3.1 数据集试验

Hilti SLAM数据集为面向复杂室内外场景建图定位研究的高精度多模态基准数据集,其中典型建图场景可视化结果如图6所示。本研究所用数据均为手持录制,具有较强的抖动和角度变化,对算法的建图鲁棒性提出挑战。绝对轨迹误差是评估求解位姿与真值位姿间差异的度量。采用绝对轨迹误差的均方根误差 $E_{RMS}$ 对比分析各算法的建图效果。

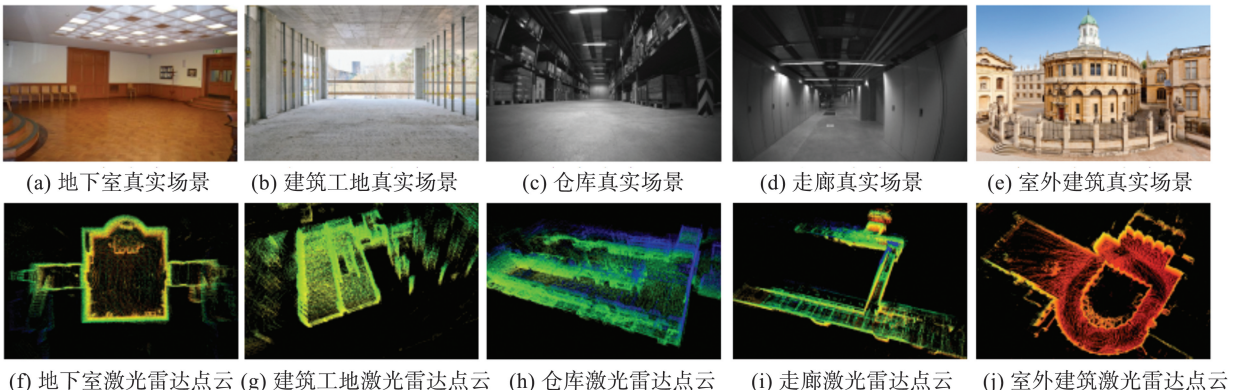
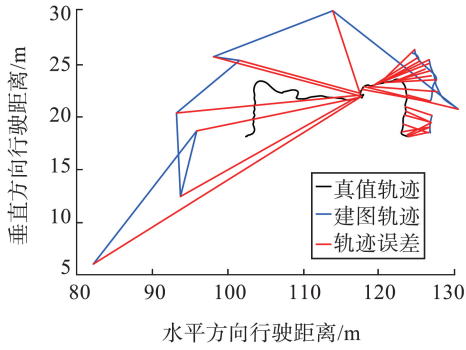


图6 Hilti SLAM数据集测试场景

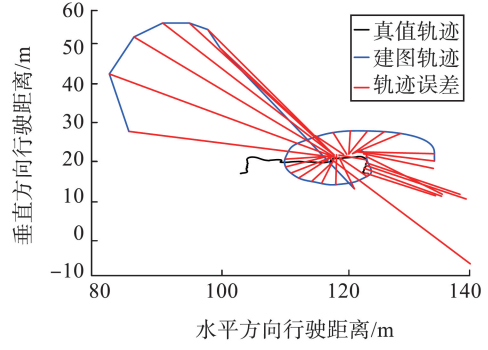
Fig.6 Experimental scenarios with the Hilti SLAM datasets

地下室与建筑工地的建图对比效果如图7和表1所示。其中,地下室场景建图是机器人由一楼开始经楼梯行进至地下室中,具有较大的垂直高度变化。由表1可知:在地下室场景中,LVI-SAM-Stereo的建图轨迹与真值轨迹的 $E_{RMS}$ 为

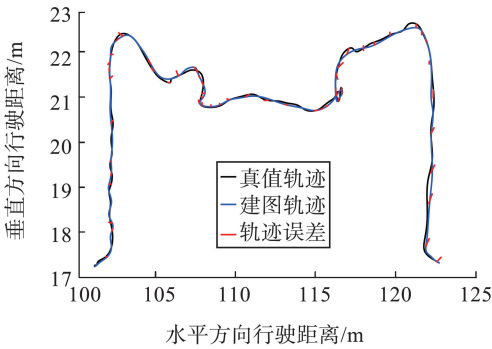
0.181 3 m,而LIO-SAM与LVI-SAM均产生显著漂移,未表现良好的建图效果;在建筑工地场景中,LVI-SAM-Stereo的建图轨迹与真值轨迹间的 $E_{RMS}$ 最小,建图精度相较LIO-SAM提升3.8%,相较LVI-SAM提升6.1%。



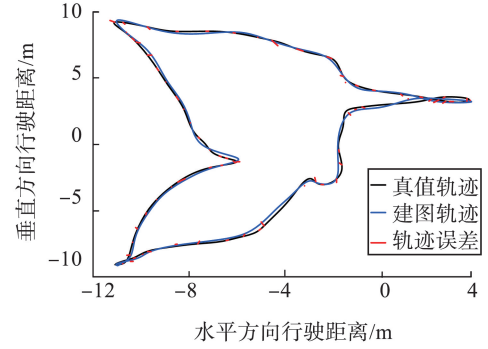
(a) LIO-SAM在地下室的建图轨迹



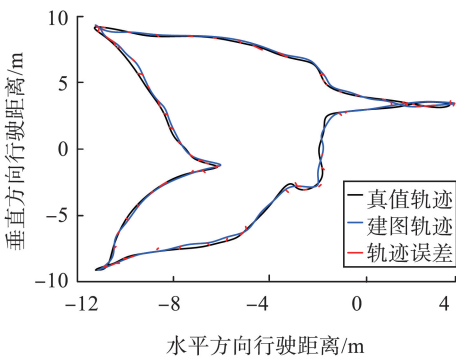
(b) LVI-SAM在地下室的建图轨迹



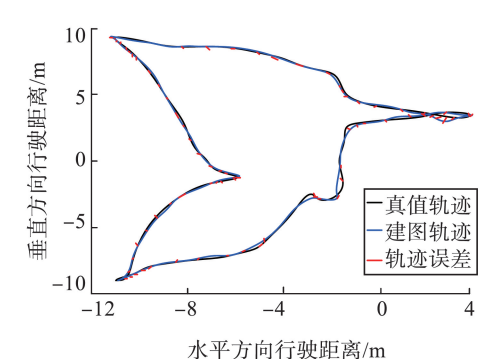
(c) LVI-SAM-Stereo在地下室的建图轨迹



(d) LIO-SAM在建筑工地的建图轨迹



(e) LVI-SAM在建筑工地的建图轨迹



(f) LVI-SAM-Stereo在建筑工地的建图轨迹

图7 Hilti SLAM数据集建图轨迹与真值轨迹对比

Fig.7 Mapping and ground truth trajectories with the Hilti SLAM datasets

表1 Hilti SLAM数据集不同场景的 $E_{RMS}$ Table 1 The  $E_{RMS}$  for different scenes with the Hilti SLAM datasets

场景	$E_{RMS}/m$		
	LIO-SAM	LVI-SAM	LVI-SAM-Stereo
地下室	—	—	0.181 3
建筑工地	0.199 2	0.204 0	0.191 6

注:“—”表示未完成建图,无法获得 $E_{RMS}$ 。

在仓库、走廊及室外建筑场景中,Hilti SLAM数据集尚未公开真值轨迹,因此仅对3种算法的建图轨迹进行对比,如图8所示。在仓库场景中,机器

人行进在狭长的货架中,原地转向后折返,LIO-SAM与LVI-SAM在建图过程中均出现漂移,仅LVI-SAM-Stereo完成建图。在走廊场景中,机器人行进至电梯附近,当电梯门打开瞬间,只有LVI-SAM-Stereo完成建图,体现LVI-SAM-Stereo对场景中动态物体建图的鲁棒性。室外建筑场景涵盖室外与室内场景,在室外部分的建图3种算法表现均良好,但当机器人进入室内走上楼梯时,只有LVI-SAM-Stereo建图效果最佳。

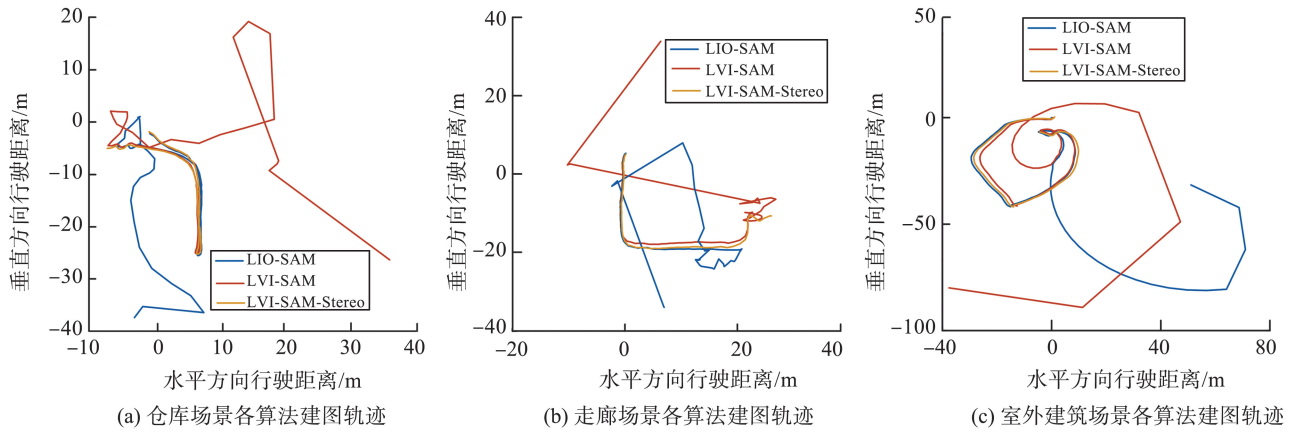


图 8 Hilti SLAM 数据集建图轨迹对比  
Fig.8 Mapping trajectories with the Hilti SLAM datasets

### 3.2 真实场景试验

为进一步验证所提 LVI-SAM-Stereo 建图及重定位性能,本研究搭建机器人测试平台,分别在室内外场景中验证分析。机器人测试平台搭载双目相机惯导模组、16 线激光雷达、实时动态载波相位差分技术(real-time kinematic, RTK)设备及计算单元,如图 9 所示。

#### 3.2.1 建图试验

机器人测试平台进行建图验证的室内外真实场景图和激光雷达点云图如图 10 所示,其中激光雷达点云图为 LVI-SAM-Stereo 的建图结果。图 10 中的场景用于全方面对比验证不同算法间的建图优越性和室内外场景的适应性。

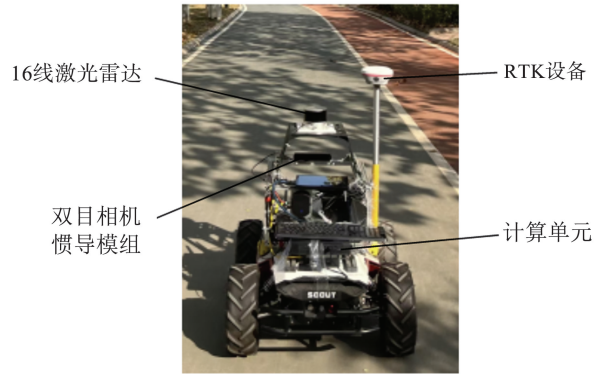


图 9 机器人测试平台  
Fig.9 Robot testing platform

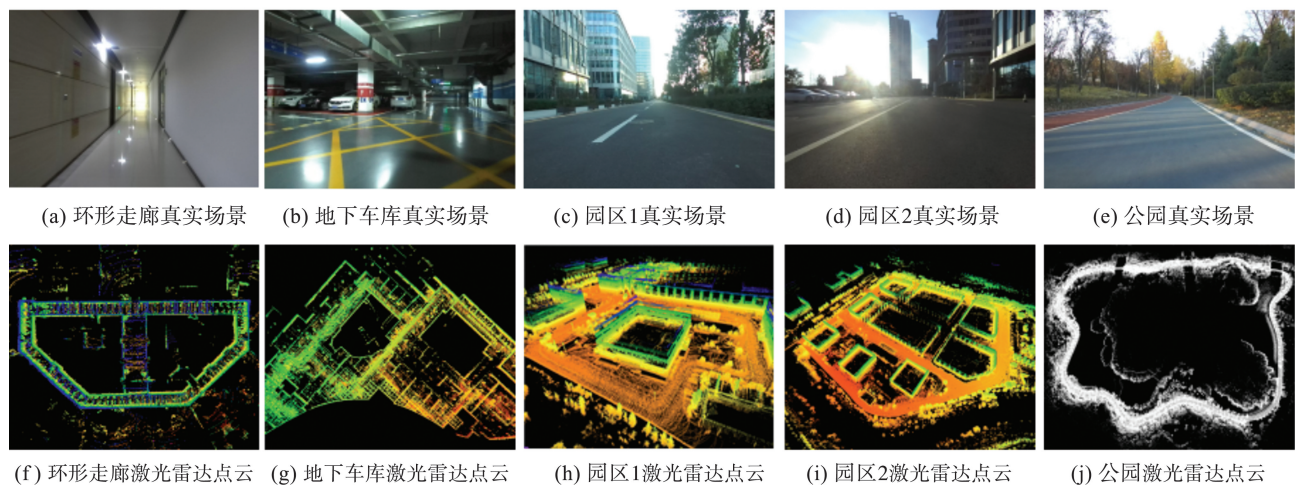


图 10 机器人测试平台建图场景  
Fig.10 Mapping scenarios with the robot testing platform

本研究考虑 RTK 设备主要用于获取室外无遮挡场景的真值轨迹,因此在室内场景中,RTK 设备无法获得真值轨迹,不同算法间的对比验证主要通过建图轨迹与真实环境的相似性和垂直漂移两项指标评估。室内环形走廊场景和地下车库场景的

建图轨迹及垂直偏差如图 11 所示,建图的垂直漂移如表 2 所示。在室内环形走廊场景中,长度为 36 m 的直走廊、对称的场景结构、纹理特征单一的白墙及两个直角转弯处的剧烈角度变化均给建图带来较大挑战。

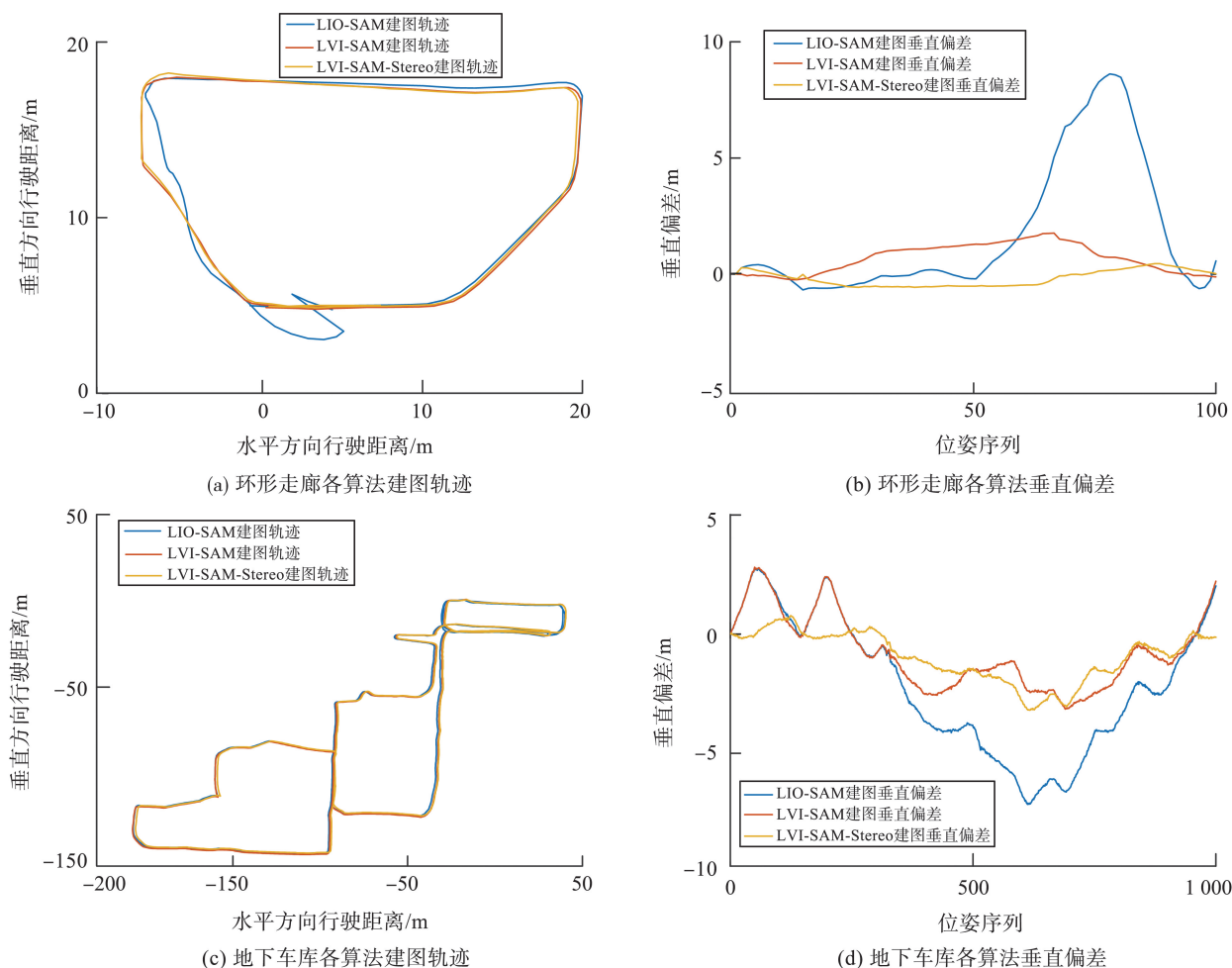


图11 室内建图轨迹及垂直偏差

Fig.11 The mapping trajectories and vertical deviations in indoor scenes

表2 室内不同场景建图的垂直漂移

Table 2 Vertical drift in mapping of various indoor scenes

单位:m

场景	LIO-SAM		LVI-SAM		LVI-SAM-Stereo	
	最大漂移	平均漂移	最大漂移	平均漂移	最大漂移	平均漂移
环形走廊	8.626 6	1.994 3	1.768 1	0.738 3	0.559 7	0.333 1
地下车库	7.136 9	2.774 5	3.141 9	1.527 6	3.192 2	1.051 3

由图11(a)、(b)和表2可知,LIO-SAM在缺乏几何特征的直走廊部分里程计退化严重,LVI-SAM-Stereo的垂直漂移最小,平均漂移相较LIO-SAM减小83.3%,相较LVI-SAM减小54.9%。在地下车库场景中,机器人测试平台进行约4万 $m^2$ 的大范围车库建图,昏暗的灯光和场景的高度相似性对建图有较高难度。由图11(c)、(d)和表2可知,LVI-SAM-Stereo建图效果最佳,平均漂移相较LIO-SAM减小62.1%,相较LVI-SAM减小31.2%。

在室外场景中,本研究分别在约12万 $m^2$ 的工业园区和约24万 $m^2$ 的空旷公园中进行建图试验,其中将工业园区场景分为园区1和园区2,使用EVO工具评估3种算法的绝对轨迹误差,试

验结果如图12和表3所示。在工业园区场景中,建图时刻意增加剧烈的角度变化和光照亮度变化,以验证室外场景下的建图鲁棒性。在园区1中,单目惯性里程计产生的较大漂移使LVI-SAM未完成建图,而LIO-SAM与LVI-SAM-Stereo均顺利完成。由表3可知:在园区1中,LVI-SAM-Stereo的建图精度更高,相比LIO-SAM,LVI-SAM-Stereo的 $E_{RMS}$ 降低2.0%;在园区2中,LVI-SAM-Stereo的建图精度均高于LIO-SAM和LVI-SAM, $E_{RMS}$ 分别降低2.4%、9.8%;在公园场景中,由于公园包含丰富的物体信息,对3种算法没有产生较大挑战,LVI-SAM-Stereo的建图精度略低于LIO-SAM。

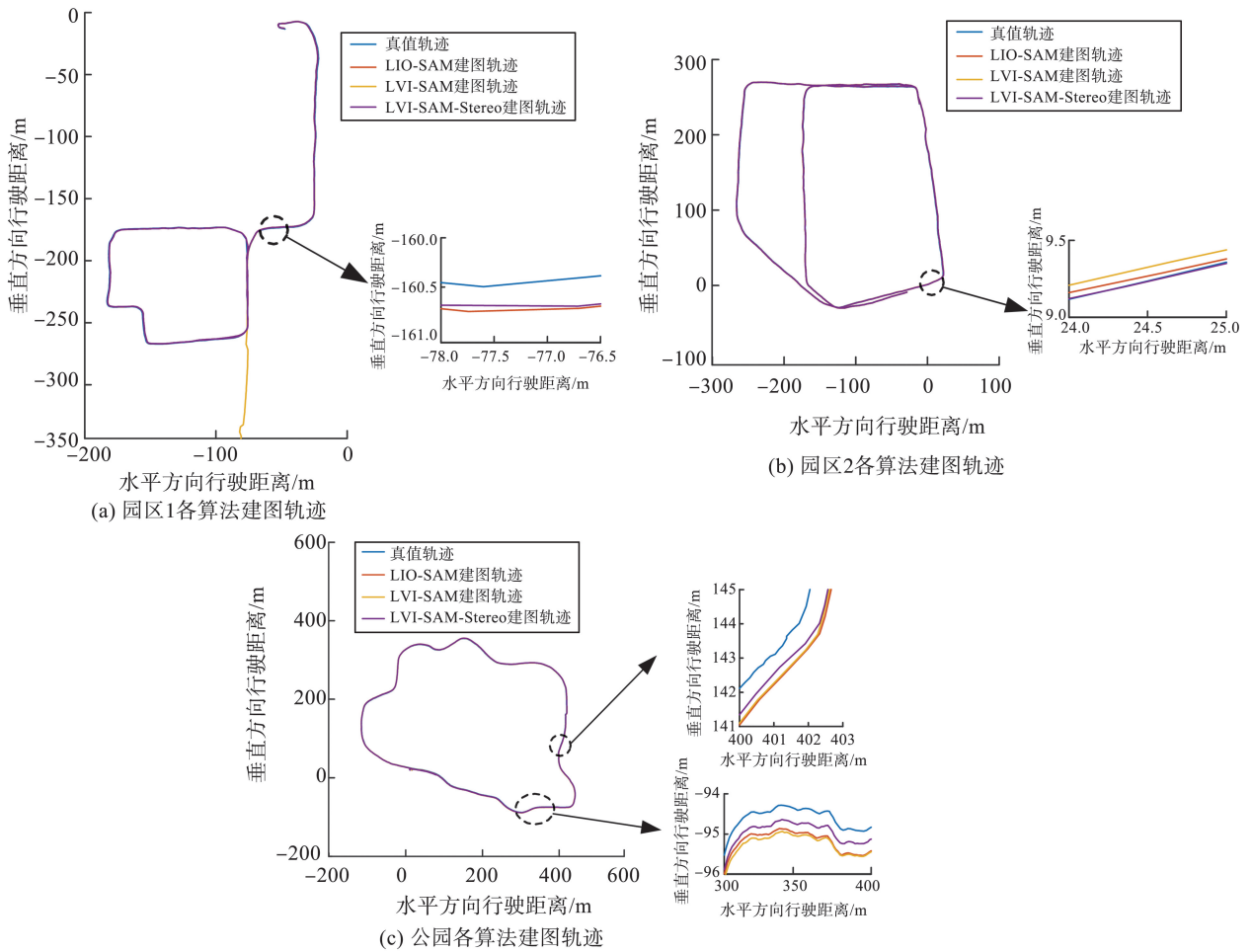


图12 机器人测试平台室外建图轨迹曲线

Fig.12 The mapping trajectories in outdoor experiments with the robot testing platform

表3 机器人测试平台室外不同场景建图的  $E_{RMS}$   
Table 3 The  $E_{RMS}$  for mapping in the outdoor scenes with the robot testing platform

场景	$E_{RMS}/m$		
	LIO-SAM	LVI-SAM	LVI-SAM-Stereo
园区1	0.716 7	—	0.702 5
园区2	1.279 7	1.385 5	1.249 4
公园	2.683 3	2.785 9	2.696 5

注:“—”表示 LVI-SAM 未完成建图,无法获得  $E_{RMS}$ 。

表4 不同试验场景中的建图适应性

Table 4 Mapping adaptability in different experimental scenes

试验环境	场景	LIO-SAM	LVI-SAM	LVI-SAM-Stereo	
Hilti SLAM 数据集	地下室	失败	失败	精度高	
	室内	建筑工地	精度低	精度低	精度高
		仓库	失败	失败	精度高
		走廊	失败	失败	精度高
	室外	室外建筑	失败	失败	精度高
机器人测试平台	室内	失败	精度低	精度高	
	地下车库	精度低	精度低	精度高	
		精度低	精度低	精度高	
	室外	园区1	精度低	失败	精度高
	园区2	精度低	精度低	精度高	
	公园	精度高	精度低	精度低	

在涵盖室内外的不同场景中,3种算法的建图适应性如表4所示。在不同场景特点、不同光照强度、不同行驶轨迹的各个场景中,LVI-SAM-Stereo均能够成功建图,具备较优的建图稳定性。同时,LVI-SAM-Stereo的建图精度也显著优于LIO-SAM和LVI-SAM。因此,本研究所提LVI-SAM-Stereo具有更优的场景适应性,能够满足不同场景的机器人建图需求。

通过上述 Hilti SLAM 数据集和机器人测试平台的建图试验得出, LVI-SAM-Stereo 在室外场景的建图精度相较于 LIO-SAM 和 LVI-SAM 分别提升 3.10% 和 5.97%, 在室内场景的平均漂移相较于 LIO-SAM 和 LVI-SAM 分别降低 72.70% 和 43.05%。因此, 相比 LIO-SAM 和 LVI-SAM, 本研究所提 LVI-SAM-Stereo 具有更优的建图精度和场景适应性。

### 3.2.2 重定位试验

为验证视觉激光雷达耦合重定位的性能, 在如

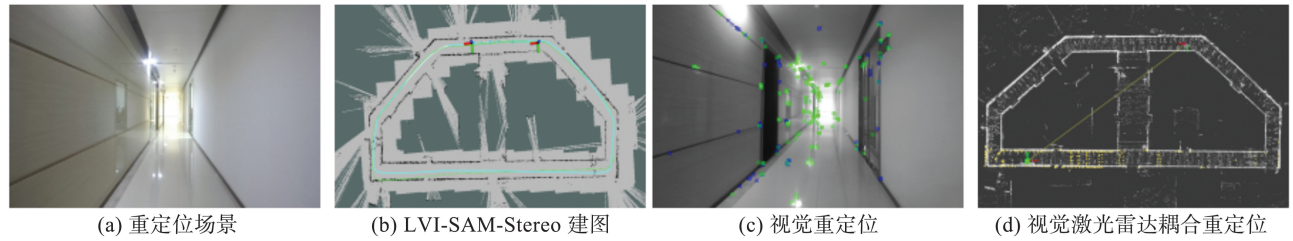


图 13 重定位效果图

Fig. 13 The visualization of relocalization effect

本研究考虑室内场景中无法使用 RTK 获取机器人行驶的真值轨迹, 因此采用同一数据包重定位 10 次, 以第 1 次重定位轨迹为基准, 验证其他重定位轨迹的定位精度和稳定性, 试验结果如图 14、15 所示。由图 14、15 可知, 在几何结构单一的长走廊场景中, 采用传统激光雷达配准重定位方法的各条

图 2(a)、13(a) 所示的激光雷达易退化的环形长走廊中进行机器人测试平台的重定位试验。基于 LVI-SAM-Stereo 构建的可直接应用于导航的栅格地图如图 13(b) 所示。当机器人进行重定位时, 视觉 ORB 特征点的匹配情况如图 13(c) 所示, 其中蓝色特征点为帧间跟踪到的 ORB 特征点, 绿色特征点为当前帧与视觉地图对应关键帧中匹配的 ORB 特征点。在环形走廊场景中, 视觉激光雷达耦合重定位实现了实时精准重定位, 如图 13(d) 所示。

轨迹均出现不同程度的漂移与退化, 重定位精度和稳定性降低, 而采用本研究所提视觉激光雷达耦合重定位方法的各条轨迹精准重叠, 其他重定位轨迹与基准轨迹的  $E_{RMS}$  小于 0.07 m。因此, 视觉激光雷达耦合重定位可提供稳定的重定位位姿, 定位精度较高, 能够满足机器人自主导航的工程化需求。

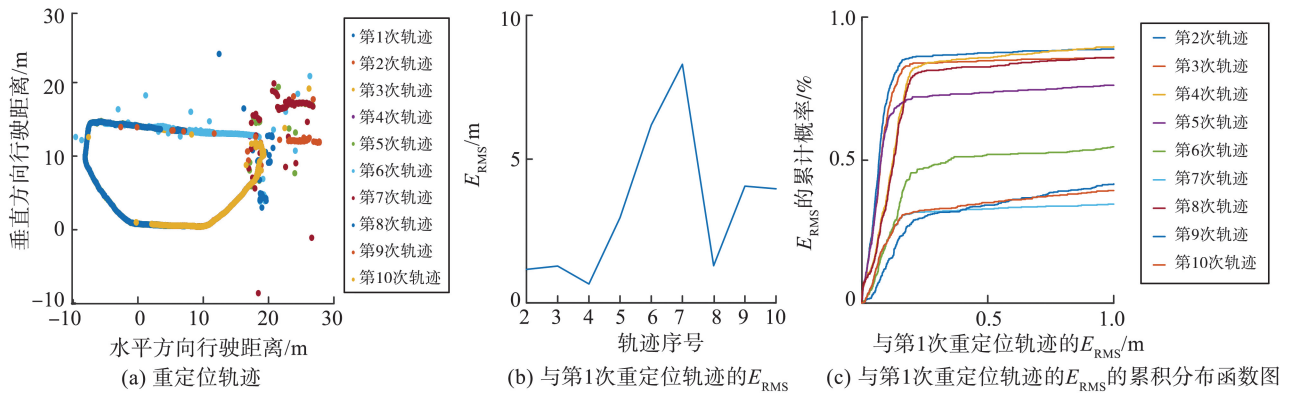


图 14 激光雷达配准重定位轨迹

Fig. 14 The trajectories of LiDAR registration relocalization

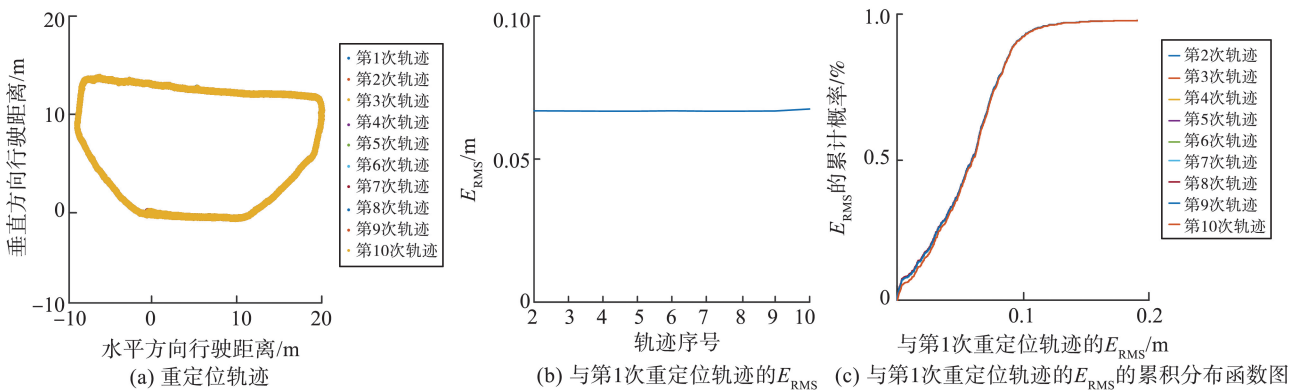


图 15 视觉激光雷达耦合重定位轨迹

Fig. 15 The trajectories of visual-LiDAR fusion relocalization

## 4 结论

本研究提出基于LVI-SAM-Stereo的多传感器融合室内外建图定位方法。激光雷达惯性位姿加速双目惯性里程计的初始化,结合Scan-Context与视觉特征匹配进行激光雷达回环检测,经因子图优化的全局里程计为视觉帧间跟踪提供初始位姿估计。通过Hilti SLAM数据集和真实场景的大量室内外试验得出,相较LIO-SAM和LVI-SAM,LVI-SAM-Stereo具有更优的建图精度和场景适应性,重定位满足机器人产品自主导航需求,为机器人建图定位领域的工程应用提供一定参考和借鉴意义。

### 参考文献:

- [1] YIN H S, LI S M, TAO Y, et al. Dynam-SLAM: an accurate, robust stereo visual-inertial SLAM method in dynamic environments[J]. IEEE Transactions on Robotics, 2023, 39(1): 289-308.
- [2] YU Z L, ZHU L D, LU G Y. Tightly-coupled fusion of VINS and motion constraint for autonomous vehicle[J]. IEEE Transactions on Vehicular Technology, 2022, 71(6): 5799-5810.
- [3] ZHONG X L, LI Y H, ZHU S Q, et al. LVIO-SAM: a multi-sensor fusion odometry via smoothing and mapping [C]//2021 IEEE International Conference on Robotics and Biomimetics (ROBIO). Sanya, China: IEEE, 2021: 440-445.
- [4] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: an accurate open-source library for visual, visual-inertial, and multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [5] SHAN T X, ENGLLOT B, MEYERS D, et al. LIO-SAM: tightly-coupled LiDAR inertial odometry via smoothing and mapping [C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, USA: IEEE, 2020: 5135-5142.
- [6] SHAN T X, ENGLLOT B, RATTI C, et al. LVI-SAM: tightly-coupled LiDAR-visual-inertial odometry via smoothing and mapping[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE, 2021: 5692-5698.
- [7] LIN Y, GAO F, QIN T, et al. Autonomous aerial navigation using monocular visual-inertial fusion[J]. Journal of Field Robotics, 2018, 35(1): 23-51.
- [8] HUANG J, ZHANG Y D, LI X. LiDAR-visual-inertial odometry using point and line features [C]//2022 4th International Conference on Robotics and Computer Vision (ICRCV). Wuhan, China: IEEE, 2022: 215-222.
- [9] JIA Y X, NI Z K, NI X, et al. A multi-sensor fusion localization algorithm via dynamic target removal [C]//2023 15th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Hangzhou, China: IEEE, 2023: 138-142.
- [10] LIU Z B, LI Z K, LIU A, et al. LVI-Fusion: a robust LiDAR-visual-inertial SLAM scheme[J]. Remote Sensing, 2024, 16(9): 1524.
- [11] SEGAL A, HAEHNEL D, THRUN S. Generalized-ICP [C]//Robotics: Science and Systems. Seattle, USA: MIT, 2009: 435.
- [12] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: an accurate  $O(n)$  solution to the PnP problem [J]. International Journal of Computer Vision, 2009, 81(2): 155-166.
- [13] LV J J, XU J H, HU K W, et al. Targetless calibration of LiDAR-IMU system based on continuous-time batch estimation[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Las Vegas, USA: IEEE, 2020: 9968-9975.
- [14] QUIGLEY M, CONLEY K, GERKEY B, et al. ROS: an open-source robot operating system[C]// ICRA Workshop on Open Source Software. Kobe, Japan: IEEE, 2009: 3-5.
- [15] CHUM O, MATAS J, KITTLER J. Locally optimized RANSAC [C]//Joint Pattern Recognition Symposium. Heidelberg, Germany: Springer, 2003: 236-243.
- [16] HELMBERGER M, MORIN K, BERNER B, et al. The Hilti SLAM challenge dataset [J]. IEEE Robotics and Automation Letters, 2022, 7(3): 7518-7525.

(编辑:孙亚彤)