

# 室内复杂环境中 LIO-SLAM 算法的改进与优化\*

郝亮, 陈国杰, 胡肖彤, 叶俊杰, 王奇斌

西安电子科技大学机电工程学院, 陕西 西安 710071

**摘要:** 针对传统开源的激光惯性里程计 (LIO, lidar-inertial odometry) 和即时定位与地图构建 (SLAM, simultaneous localization and mapping) 结合的 LIO-SLAM 在室内复杂环境中受激光特征稀疏与动态遮挡影响、定位精度下降等问题, 提出一种融合视觉里程计的改进方法。在保持 LIO-SLAM 激光惯性紧耦合框架的基础上, 引入基于 ORB 特征的三维定位与地图构建算法 (ORB-SLAM) 作为独立的视觉里程计模块, 为系统提供高频率、丰富纹理的视觉约束信息。通过自适应权重融合策略, 实现激光、惯性与视觉观测的多源优化, 增强了在弱几何约束、纹理丰富但结构复杂环境中的鲁棒性。在多种典型室内场景 (走廊、开放大厅及动态人群环境) 中开展了实验验证。结果表明, 相较于原始 LIO-SLAM, 整体轨迹误差降低至原始系统的 70%。研究验证了视觉-激光-惯性多模态融合在室内复杂环境下的可行性与有效性, 为高精度室内自主定位与地图构建提供了新的思路。

**关键词:** 室内自主定位; LIO-SLAM; ORB-SLAM; 视觉里程计; 多传感器融合

**中图分类号:** TP24 **文献标志码:** A **文章编号:** 2097-0137(2026)01-0023-10

## Improvement and optimization of LIO-SLAM algorithm in indoor complex environments

HAO Liang, CHEN Guojie, HU Xiaotong, YE Junjie, WANG Qibin

School of Mechano-Electronic Engineering, Xidian University, Xi'an 710071, China

**Abstract:** This paper proposes an improved method that integrates visual odometry to address the defect of traditional LIO-SLAM being affected by sparse laser features and dynamic occlusion in complex indoor environments, resulting in decreased positioning accuracy. Specifically, while maintaining the LIO-SLAM laser inertial tight coupling framework, ORB-SLAM is introduced as an independent visual odometry module to provide high-frequency and rich texture visual constraint information for the system. By using an adaptive weight fusion strategy, multi-source optimization of laser, inertial, and visual observations is achieved, enhancing robustness in environments with weak geometric constraints and rich textures but complex structures. The experiment was conducted in various typical indoor scenarios (corridors, open halls, and dynamic crowd environments), and the results showed that the overall trajectory error was reduced to 70% of the original system, compared to the original LIO-SLAM. This study validates the feasibility and effectiveness of visual laser inertial multimodal fusion in complex indoor environments, providing new ideas for high-precision indoor autonomous positioning and map construction.

**Key words:** indoor positioning; LIO-SLAM; ORB-SLAM; visual odometry; multi-sensor fusion

\* 收稿日期: 2025-09-28

录用日期: 2025-10-30

网络首发日期: 2026-01-08

基金项目: 中央高校基本业务费 (ZYTS25049)

作者简介: 郝亮 (1985 年生), 男; 研究方向: 机器人和 SLAM 导航; E-mail: haoliang@xidian.edu.cn

通信作者: 王奇斌 (1988 年生), 男; 研究方向: 人工智能算法; E-mail: qbwang@xidian.edu.cn

全文阅读



ZR20250210

即时定位与地图构建 (SLAM, simultaneous localization and mapping) 技术广泛应用于机器人、无人机和智能驾驶等领域 (蔡培培等, 2025; 李旭等, 2025)。SLAM 系统中的激光雷达和视觉传感器, 如相机等都是必不可少的硬件组成设备 (丁健等, 2025)。激光雷达 (LiDAR) 在定位精度和抗光照干扰方面表现突出, 但其建图结果依赖于环境的几何特征, 面对结构单一场景时易出现退化 (冯迎宾等, 2025; 洪志云, 2025)。与之相对, 视觉传感器可捕获丰富的纹理与语义信息, 却对光照变化和运动模糊敏感, 初始化环节也相对复杂, 因而单纯依赖视觉的 SLAM 难以保证稳定性。为弥补二者不足, 融合多种传感器的 SLAM 技术逐渐成为机器人感知领域的重要研究方向 (蒋风洋等, 2025)。已有研究提出面向视觉、视觉-惯性及多地图场景的开源框架, 通过联合相机与惯性测量单元 (IMU) 提升高速运动以及纹理缺失环境下的建图质量, 但系统初始化仍需较严苛的条件。另一类工作发展了紧耦合的激光雷达-惯性平滑建图方法 (LIO-SLAM), 借助 IMU 消除点云畸变并为帧间位姿估计提供初值, 然而在长廊等几何约束不足的区域易发生精度退化。进一步地, 相关研究将视觉模块与 LIO-SLAM 相结合, 形成激光雷达-视觉-惯性的多源耦合体系 (LVI-SLAM), 并利用单目视觉惯性状态估计器 (VINS-Mono) 实现更鲁棒的姿态推断, 但在尺度恢复和长距离回环检测方面仍存在局限。随后, 有学者引入点线特征、动态点云剔除以及特征深度恢复等策略, 以改进 LVI-SLAM 的精度与鲁棒性, 但这类方法普遍需要针对具体场景调节回环参数, 且在大规模室内外环境中的验证仍显不足。

当前, 商用系统多依赖将激光雷达的实时扫描与先验点云地图进行配准, 但在精度与稳定性方面仍面临挑战 (蒋风洋等, 2025)。针对建图精度不高以及环境适应性差的难题, 本文提出一种基于双目视觉的紧耦合激光雷达-视觉-惯性的多传感器建图方案 (LVI-SLAM-Stereo), 通过引入视觉增强的视觉里程计, 减少激光回环误匹配的影响, 从而提升长距离导航场景下的定位可靠性与地图构建质量。

目前, 在多传感器融合的定位与建图系统中, 紧耦合 (tightly coupled) 与松耦合 (loosely coupled) 是 2 种典型的融合策略。如图 1 所示, 松耦合通常将各传感器独立处理, 视觉前端先单独完成位姿估计或里程计计算, IMU 或激光雷达各自输出里程计结果, 然后在后端仅将这些“高层”结果进行加权或

滤波融合 (陆佳嘉等, 2025)。其优点是结构清晰、模块化强, 便于维护与替换, 但由于不同模块间只共享最终结果, 信息利用率有限, 对观测噪声的鲁棒性较差 (罗娟等, 2025)。

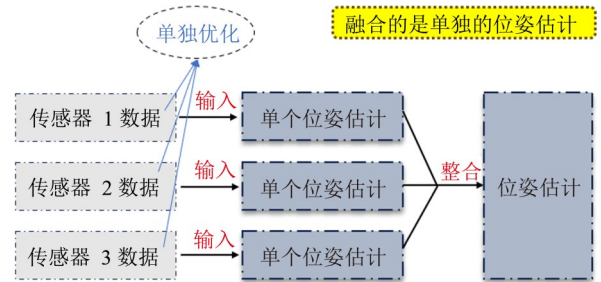


图 1 松耦合方法

Fig. 1 Loosely coupled method

如图 2 所示, 紧耦合策略则在更底层完成信息联合建模: 视觉的特征观测、IMU 的预积分量测、激光的点云约束都直接写入同一优化问题或因子里图中, 通过统一的状态变量 (位姿、速度、偏置等) 同时求解 (柳炳琦等, 2025)。这种方式能够充分利用多源观测之间的时序与空间关联, 提高整体精度与一致性, 并且在单一传感器暂时失效时仍能维持较好的估计效果。不过, 紧耦合系统在实现上更复杂, 对实时性和计算资源提出了更高要求 (王宁等, 2025)。

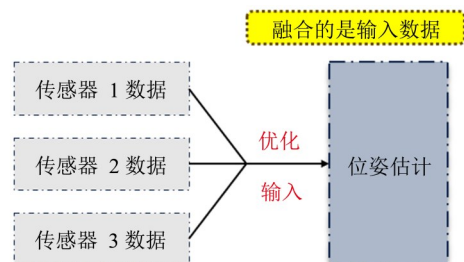


图 2 紧耦合方法

Fig. 2 Tightly coupled method

本文提出的改进方法中, 激光、视觉与 IMU 的观测被统一建模到同一因子图中, 即采用紧耦合方法, 且每一个传感器测量对应一个因子, 与全局状态节点共同优化, 从而充分发挥了多传感器的冗余和互补优势, 实现高精度、鲁棒的位姿估计。

## 1 SLAM 数学模型

假设现在有 1 个机器人, 正携带着某种传感器在未知环境里运动, 由于相机通常是在某些时刻采集数据的, 所以只需要关心这些时刻的位置和地图。这就把一段连续时间的运动变成了离散时刻

$t = 1, 2, \dots, K$ 当中发生的事情(吴蕊等, 2025)。在这些时刻,用 $x$ 表示机器人自身的位置。于是,各时刻的位置就记为 $x_1, \dots, x_k$ ,它们构成了小萝卜的轨迹。在地图方面,我们设地图是由许多个路标点组成的,而每个时刻传感器会测量到一部分路标点,得到路标点的观测数据。不妨设路标点一共有 $N$ 个,用 $y_1, \dots, y_N$ 表示。因此,运动可以表达为:从 $k-1$ 到 $k$ 时刻,机器人的位置 $x$ 是如何变化的;与此同时,观测可以表示为:机器人在 $k$ 时刻,于 $x_k$ 处探测到了某一个路标 $y_i$ 。通常,机器人会携带一个测量自身运动的传感器,比如说码盘或惯性传感器。这个传感器可以测量有关运动的读数,但不一定直接是位置之差,还可能是加速度、角速度等信息。然而,无论是什么传感器,我们都能使用一个通用的、抽象的数学模型(吴永豪等, 2025)。用一般函数 $f$ 来描述这个过程,而不具体指明 $f$ 的作用方式。这使得整个函数可以指代任意的运动传感器,成为一个通用的方程,而不必限定于某个特殊的传感器上。如果把它称为运动方程,运动方程可以表示为

$$x_k = f(x_{k-1}, u_k, w_k),$$

式中 $u_k$ 表示由传感器信息获得的运动测量量,如轮速、角速度或加速度等; $w_k$ 为系统过程噪声,用于描述由于传感器误差、建模不完善等因素引起的不确定性,而且通常假设其服从0均值高斯分布。与运动方程相对应,观测方程描述的是当机器人在 $x_k$ 位置上看到某个路标点 $y_i$ ,产生了一个观测数据 $z(k, j)$ 。同样,用一个抽象函数 $h$ 来描述这个关系。并用 $v_{k,j}$ 表示这个过程中的观测噪声,同样假设其服从0均值高斯分布,则观测方程可以表示为

$$z_{k,j} = h(y_i, x_k, v_{k,j}).$$

## 2 SLAM框架

### 2.1 经典SLAM框架

经典SLAM流程通常包括数据采集、前端处理、回环检测与后端优化以及地图构建几个环节(余浩扬等, 2025),如图3所示。

SLAM系统通过传感器获取环境观测数据,为后续定位与建图提供基础信息。前端处理模块负责根据连续帧观测估计机器人短时位姿,包括特征提取、特征匹配及帧间配准等操作,以实现局部位姿估计(余夏怡等, 2025)。回环检测模块用于识别机器人重访已知区域,为后端提供全局约束。后端优化通过图优化或滤波方法,将前端估计和回环约

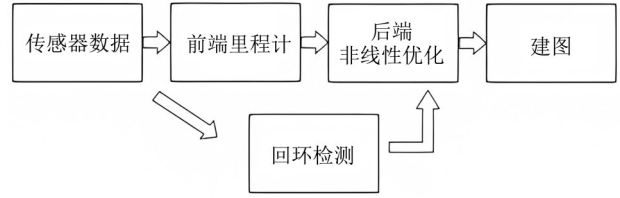


图3 经典SLAM流程图

Fig. 3 Classic SLAM flow chart

束统一建模,以获得全局一致的轨迹与关键帧位姿。最后,基于优化后的位姿信息,系统构建环境地图,通常可采用稠密点云、占据栅格或体素地图表示,从而为导航、路径规划及自主决策提供支持。该流程具有普适性,可根据不同传感器类型或具体应用进行相应扩展(颜金龙, 2025)。

系统的前端主要由里程计模块组成,其核心任务是根据连续帧传感器观测数据估算机器人在短时间范围内的运动,并构建局部地图以反映当前环境的结构特征(曾宪阳等, 2025)。前端里程计通过对相邻帧观测进行配准、特征匹配或扫描匹配,生成局部位姿估计,为后续全局优化提供初值与约束信息(周淑凯等, 2025)。

后端模块则接收来自不同时间点的视觉里程计输出的相机位姿,以及回环检测模块提供的回环约束信息,并通过非线性优化或图优化方法对这些数据进行统一建模和优化处理,从而得到全局一致的轨迹和环境地图。

在系统运行过程中,回环检测模块负责判断机器人是否曾经访问过先前位置,一旦检测到回环,它会将回环匹配信息发送给后端,以校正累积漂移,提升轨迹与地图的一致性和精度。最终,基于优化后的全局轨迹,系统能够生成满足任务需求的环境地图,该地图不仅能够精确反映空间结构,还可为路径规划、导航控制及高层任务执行提供可靠的空间参考。

### 2.2 前端里程计

以视觉SLAM为例,视觉里程计关心相邻图像之间的相机运动,最简单的情况当然是2张图像之间的运动关系(周永光等, 2025)。如图4所示,人眼能够感觉到发生在2张图像之间的相对运动,而这就是里程计的工作内容,即旨在通过连续图像序列估计相机的相对运动。里程计是视觉SLAM系统中的核心模块,其基本原理是计算相邻帧图像之间的位姿变化,从而获得短时运动估计和局部场景结构(周云水等, 2025)。



图4 人眼可感知运动的图像

Fig. 4 Images that the human eye can perceive as moving

视觉里程计的核心是估计相邻帧  $I_t, I_{t+1}$  之间的相机位姿变化, 通常表示为一个刚体变换矩阵  $T_i^{t+1} \in \text{SE}(3)$ 。

$$T_i^{t+1} = \begin{bmatrix} R_i^{t+1} & t_i^{t+1} \\ \mathbf{0}^T & 1 \end{bmatrix},$$

其中  $R_i^{t+1} \in \text{SO}(3)$  表示旋转矩阵,  $t_i^{t+1} \in \mathbb{R}^3$  表示平移向量。  $\mathbf{0}^T$  表示  $1 \times 3$  的零向量转置形式, 用于保证齐次变换矩阵的维度一致 (Liu et al., 2025)。

视觉里程计的原理其实可理解为对空间几何关系的量化测量: 相机在连续帧间的旋转与平移会导致图像中像素位置的变化, 因此可以通过分析这些变化来还原相机自身的运动及场景的三维结构。

假设在帧  $t$  中某三维点  $P_i = [X_i, Y_i, Z_i]^T$  对应像素  $p_i^t = [u_i, v_i]^T$ , 在帧  $t+1$  中的投影为

$$p_i^{t+1} = \pi(R_i^{t+1} P_i + t_i^{t+1}),$$

相机投影函数(针孔模型)

$$\pi([X, Y, Z]^T) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix},$$

其中  $[X, Y, Z]^T$  表示空间点在相机坐标系下的三维坐标;  $f_x$  和  $f_y$  分别为相机在水平方向与垂直方向上的等效焦距, 单位为像素;  $c_x$  和  $c_y$  表示主点在图像坐标系中的像素位置。上述参数统称为相机内参 (Chang et al., 2026), 通过相机标定过程获取。则前端通过式(1)的最小化重投影误差方程求解位姿:

$$T_i^{t+1} = \arg \min_{T \in \text{SE}(3)} \sum_i [p_i^{t+1} - \pi(TP_i)]^2. \quad (1)$$

视觉里程计仅关注相邻时刻的运动, 其估计结果不依赖于历史帧的全局信息, 因此在长期运行中不可避免地出现累计误差 (Muslim et al., 2025)。

最开始视觉里程计通过累积帧间位姿得到全局轨迹, 即

$$T_0^{t+1} = T_0^t \cdot T_i^{t+1}.$$

但由于每次帧间估计均存在一定误差, 这些误差会随时间累积, 导致轨迹偏离真实位置。例如, 当机器人依次向左旋转  $90^\circ$  再向右旋转  $90^\circ$  时, 即便后续估计精确, 由于前期误差, 整体轨迹仍可能偏离原点, 从而影响地图一致性 (Xiao et al., 2025)。

由于每次估计存在误差, 这将导致累计漂移

$$\hat{T}_0^{t+1} = T_0^t \cdot T_i^{t+1} + \Delta T_i^{t+1},$$

其中  $\Delta T_i^{t+1}$  表示每次位姿估计的误差。

为了克服这一问题, 回环检测和后端优化成为了现代 SLAM 系统中必不可少的环节 (Ma et al., 2024)。回环检测用于识别机器人是否回到先前访问过的位置, 并生成全局约束信息; 后端优化则基于这些约束对轨迹进行非线性优化, 修正累计漂移, 实现全局一致的定位和地图构建。通过这一流程, 前端里程计不仅提供了短时的相对运动估计, 也为 SLAM 系统的全局精度和环境建图奠定基础。

### 2.3 后端优化

SLAM 体系中, 后端优化的任务是对含噪观测进行一致化处理, 以获得全局精确的状态估计。理想情况下, 所有传感器测量都应当完全可靠。但在真实环境中, 无论传感器精度如何, 测量结果都不可避免地受到噪声干扰。价格较低的传感器往往具有更大的测量误差, 而高精度传感器虽然性能优越, 却也仍会受到磁场、温度等外部因素的影响。因此, 除了“如何从观测中推断运动”这一基本问题外, 更重要的是量化这些观测所携带的不确定性, 包括噪声在时间序列中的传播方式以及当前估计的可信度。

后端优化正是针对上述问题提出的系统化解决方案, 其目标是从带有噪声的多源数据中联合推断整个系统的状态并评估其不确定性 (Xu et al., 2022)。该问题可形式化为最大后验概率估计 (MAP, maximum a posteriori) 问题:

$$\hat{X} = \arg \max_X p(X|Z),$$

其中  $\mathbf{X}$  是待估计的整条轨迹与地图参数,  $\mathbf{Z}$  表示全部观测数据。根据贝叶斯公式, 后验概率可写为

$$p(\mathbf{X}|\mathbf{Z}) \propto p(\mathbf{Z}|\mathbf{X})p(\mathbf{X}).$$

于是, MAP 问题等价于最小化负对数后验:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left[ -\log p(\mathbf{Z}|\mathbf{X}) - \log p(\mathbf{X}) \right].$$

若假设测量噪声为高斯分布,  $\mathbf{Z} = \{z_i\}$  且每个观测独立, 则似然项可进一步展开为非线性最小二乘问题:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \sum_i |r_i(\mathbf{X})|_{\Sigma_i}^2,$$

其中  $r_i(\mathbf{X})$  是第  $i$  个因子的残差函数, 观测协方差矩阵  $\Sigma_i$  满足  $|\mathbf{e}|_{\Sigma_i}^2 = \mathbf{e}^T \Sigma_i^{-1} \mathbf{e}$ .

总而言之, 前端模块的职责是为后端提供待优化的数据及初始值。后端则专注于基于滤波或非线性优化的全局推断, 前端确保观测的时序一致与初步姿态估计, 后端则通过全局优化消除累积误差, 保证轨迹与地图的一致性 (Wang et al., 2022)。

从历史角度来看, 后端优化所处理的正是 SLAM 研究的起点。早期文献中, SLAM 问题被描述为“空间状态不确定性的估计”, 强调的正是对运动主体及其周围环境的不确定性建模与推断。解决这一问题需要状态估计理论的支持, 通过非线性优化推导状态分布的均值与方差。例如, 在扩展卡尔曼滤波 (EKF) 框架下, 将系统的非线性状态转移模型设为函数  $f$ , 则预测方程分别为:

$$\mathbf{x}_{k|k-1} = f(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1},$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1},$$

其中  $\mathbf{x}_{k|k-1}$  表示在上一时刻已知的条件下, 对系统状态在时刻  $k$  的预测值,  $\mathbf{x}_{k-1}$  则表示上一时刻的状态估计结果,  $\mathbf{w}_{k-1}$  表示过程噪声, 用于刻画系统建模误差及外界扰动,  $\mathbf{P}_{k|k-1}$  为预测状态的协方差矩阵, 反映不确定性大小,  $\mathbf{F}_{k-1} = \frac{\partial f}{\partial \mathbf{x}}|_{\mathbf{x}_{k-1}}$  为状态转移函数对状态变量的一阶雅可比矩阵,  $\mathbf{Q}_{k-1}$  表示过程噪声的协方差矩阵。

## 2.4 回环检测

如图 5 所示, 回环检测旨在抑制机器人位姿估计在长时间运行中产生的累积漂移。当机器人在真实环境中经过一段运动后重新回到起始位置时, 由于传感器噪声与估计误差, 其状态估计往往无法回到原点。若系统能够识别出机器人已再次到达先前访问的场所, 并将这一约束反馈给后端优化模块, 即可有效“拉回”漂移的轨迹, 实现全局一致的状态估计。

回环检测与定位和建图环节紧密相关, 其核心目

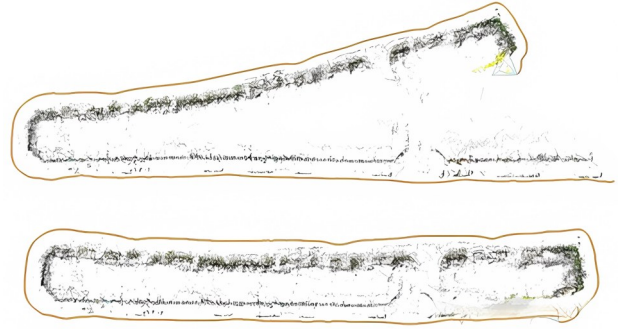


图5 回环检测效果示意图

Fig. 5 Schematic diagram of loop detection effect

标是判断当前观测是否与历史观测来自同一场景。地图在此过程中具有关键意义, 其价值不仅在于环境重构, 更在于为机器人提供“已访问区域”的全局参考。实现回环检测的途径多样, 例如在环境中布设人工标志 (如二维码), 当机器人再次观测到该标志即可确定重访位置。然而, 此类依赖外部设施的方案限制了系统的适用性。相比之下, 更具普适性的策略是依靠机器人自带的传感器, 尤其是视觉传感器, 通过图像相似性评估来判断场景是否重复。

设当前关键帧图像的全局描述子为  $d_i \in R^k$ , 历史关键帧  $i$  的描述子为  $d_i$ 。定义相似度度量函数

$$s_{i,i} = \exp\left(-\frac{|d_i - d_i|^2}{2\sigma^2}\right).$$

若存在某帧  $j$  使得  $s_{i,j} > \tau$  ( $\tau$  为设定阈值)。则认为当前帧与关键帧  $j$  构成一次回环候选, 当系统成功检测到回环后, 将生成“关键帧  $A$  与关键帧  $B$  表示同一物理位置”的约束信息, 并传递至后端优化。后端通常基于因子图进行全局调整, 其目标是最小化包含回环边的整体误差函数

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left[ \sum_{(i,j) \in \mathcal{E}_{\text{odom}}} |r_{ij}^{\text{odom}}(\mathbf{x}_i, \mathbf{x}_j)|_{\Sigma_{ij}^{\text{odom}}}^2 + \sum_{(p,q) \in \mathcal{E}_{\text{loop}}} |r_{pq}^{\text{loop}}(\mathbf{x}_p, \mathbf{x}_q)|_{\Sigma_{pq}^{\text{loop}}}^2 \right],$$

其中  $\mathbf{X} = \{x_0, \dots, x_T\}$  为所有位姿变量,  $\mathcal{E}_{\text{odom}}$  表示里程计 (前端测量) 的因子集合,  $\mathcal{E}_{\text{loop}}$  表示由回环检测生成的闭环边集合,  $r_{ij}^{(\cdot)}$  为对应因子的残差函数,  $\Sigma_{ij}^{(\cdot)}$  为测量协方差矩阵, 同样满足  $|\mathbf{e}|_{\Sigma_{ij}^{(\cdot)}}^2 = \mathbf{e}^T \Sigma_{ij}^{(\cdot)-1} \mathbf{e}$ .

该优化问题可通过 Levenberg-Marquardt 或 Gauss-Newton 等非线性最小二乘法求解, 从而在加入闭环约束后显著削弱累积误差, 实现轨迹与地图的一致性 (Lee et al., 2025)。

### 3 融合算法

#### 3.1 多传感器融合框架

为提升激光雷达惯性导航同步定位与建图 (LIO-SLAM) 在长期、复杂环境下的鲁棒性与精度, 本文在其原有架构中引入了基于视觉的 ORB-SLAM 模块, 构建了 1 套视觉 - 激光 - 惯性多传感器融合框架。该框架旨在充分发挥各传感器的互补优势: 激光雷达提供高精度三维结构信息, 惯性测量单元提供高频动态约束, 视觉系统则通过关键帧特征与回环检测抑制累积漂移, 实现全局一致的状态估计。

**3.1.1 ORB 特征** 在视觉 SLAM 系统中, 特征点的选择直接影响到前端的匹配效果与整体系统的鲁棒性。如图 6 所示, 在相近的两张图像之间寻找和匹配相同的特征点即可还原自身的运动, 本文采用 ORB (oriented FAST and rotated BRIEF) 特征作为系统的视觉前端, 是因为它在设计中引入了方向信息, 解决了传统 BRIEF (二进制的鲁棒独立基本特征) 描述子缺乏旋转不变性的问题。

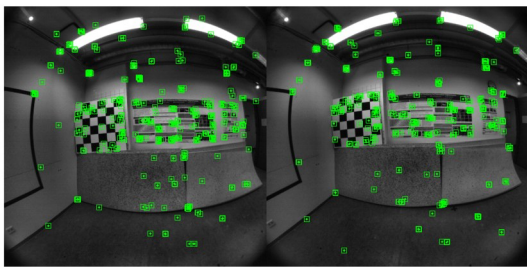


图 6 视觉特征点

Fig. 6 Visual feature points

ORB 特征由关键点与描述子组成。它的关键点称为“Oriented FAST”, 是一种改进的基于图像特征提取 (FAST, features from accelerated segment test) 角点。其关键点选取策略如下:

(1) 如图 7 所示, 有一个像素点  $p$ , 快速将  $p$  像素

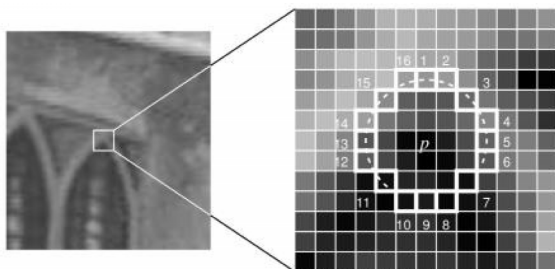


图 7 FAST 关键点

Fig. 7 FAST key points

点的亮度与其周围 16 个像素点的亮度进行比较。这 16 个像素点将被划分为 3 类 (比  $p$  点亮、比  $p$  点暗、与  $p$  点亮度相似)。如果超过 8 个像素点比  $p$  点更亮或更暗, 则把  $p$  点作为一个关键点。

(2) FAST 角点没有方向信息, 且由于它取固定半径为 3 的圆, 因此也存在尺度问题 (远看是角点, 近看不是角点)。所以, ORB 算法使用了图像金字塔来解决尺度问题。如图 8 所示, 图像金字塔每往上一层就对图像进行一个固定倍率的缩放, 较小的图像便可以当作远景。这样就解决了当距离变化时无法在同一个位置检测到特征点的问题。

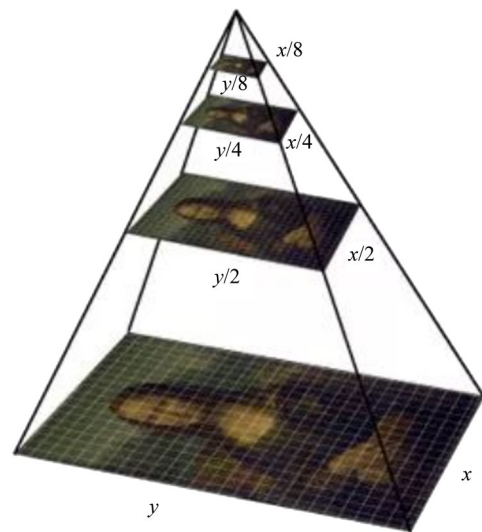


图 8 图像金字塔

Fig. 8 Image Pyramid

通过为每个特征点分配主方向, 并在构造描述子时进行旋转补偿, ORB 特征在存在相机姿态旋转或视角变化时依然能够保持较高的匹配稳定性。为了计算图像块的主方向, ORB 在计算 FAST 关键点的时候额外计算了图像块的矩:

$$m_{pq} = \sum_{x,y} x^p y^q I(x, y).$$

通过矩找到图像块的质心  $C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$ , 可以由图像块的中心  $O$  与质心  $C$  构建一个方向向量  $OC$ , 所以特征点的方向可以定义为

$$\theta = \arctan 2(m_{01}, m_{10}).$$

通过以上的方法, FAST 角点便具有了尺度与旋转的描述, 当图像发生旋转的时候, 此方法保证了特征点的准确识别, 有效提升了其在不同图像中的鲁棒性。除此之外, ORB 还兼具以下优点: 其检测部分基于 FAST 算法, 保证了特征点提取的高速

性;而描述子部分在BRIEF的基础上改进,使用二进制字符串表示,不仅计算效率高,而且在匹配时可以通过汉明距离进行快速比较,显著提升实时性。与传统的尺度不变特征变换(SIFT, scale-invariant feature transform)和加速稳健特征(SURF, speeded-up robust features)相比,ORB在保证一定精度和鲁棒性的同时,计算开销更小,特别适合实时SLAM的应用场景。

**3.1.2 系统总体结构** 整个融合系统主要改进的部分为前端感知层(前端里程计),本系统的激光前端仍基于LIO-SLAM的点云特征提取与里程计实现,输出连续的激光里程计约束。新加入的视觉前端沿用ORB-SLAM的特征提取与跟踪机制,生成相机位姿初值与关键帧数据库,并支持基于词袋模型(bag-of-words)的回环检测。后端基于LIO-SLAM原有的因子图结构进行改进,将视觉里程计因子加入图结构,采用统一的因子图优化器,将激光、视觉、IMU 3类测量共同纳入非线性最小二乘框架。本系统的状态向量定义为

$$\mathbf{X} = \{x_0, \dots, x_T, b_0, \dots, b_T\},$$

其中 $x_k$ 表示第 $k$ 时刻的位姿, $b_k$ 为IMU偏置。优化目标是最大后验估计的方程,可表述为

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left( \left\| r_{ij}^{\text{lidar}} \right\|_{\Sigma_{ij}^{\text{lidar}}}^2 + \left\| r_{ij}^{\text{imu}} \right\|_{\Sigma_{ij}^{\text{imu}}}^2 + \left\| r_{pq}^{\text{vision}} \right\|_{\Sigma_{pq}^{\text{vision}}}^2 \right),$$

其中3类残差项分别对应激光约束、IMU预积分约束与视觉关键帧(回环)约束。此外,不同传感器的时间戳差异会影响融合效果。系统首先利用硬件触发与软件插值完成时间对齐;随后在前端通过最近邻策略或基于运动模型的外插,对激光帧、图像帧与IMU数据进行严格匹配。为确保尺度一致性,在初始阶段引入外参标定矩阵 $\mathbf{T}_{\text{cam}}^{\text{lidar}} \in \text{SE}(3)$ ,其值由离线标定获得并在优化中作为固定量或缓慢更新的参数。当系统启动时,IMU以最高频率运行并提供运动初值;激光前端输出里程计因子,视觉前端并行执行特征提取和关键帧匹配;一旦检测到回环,立即向后端插入闭环因子。此框架在不改变LIO-SLAM高效性的前提下,充分引入了ORB-SLAM的视觉信息与回环能力,为长时间、长距离、多场景的自主导航提供了更高的精度与鲁棒性。后续的章节会进一步给出在实验设置以及在真实数据集上的运行情况,以验证该多传感器融合框架的可行性。

### 3.2 图优化算法改进

本文提出的改进方法使用统一的因子图优化

框架,将激光雷达、视觉以及惯性测量单元(IMU)信息共同约束于同一图结构。整条轨迹被离散为一系列状态节点 $\mathbf{X} = \{x_0, x_1, \dots, x_N\}$ 。

如图9所示,每个节点 $x_i$ 包含机器人在 $i$ 时刻的位姿与相关状态量。传感器观测被建模为连接这些节点的因子(边),用以表达测量之间的概率关系。多源传感器的观测则被表示为条件概率因子,此时联合后验分布写作:

$$p(\mathbf{X}|\mathbf{Z}) \propto \prod_{(i,j) \in \mathcal{E}_{\text{lidar}}} p(z_{ij}^{\text{lidar}} | x_i, x_j) \times \prod_{(i,j) \in \mathcal{E}_{\text{imu}}} p(z_{ij}^{\text{imu}} | x_i, x_j) \prod_{(p,q) \in \mathcal{E}_{\text{vision}}} p(z_{pq}^{\text{vision}} | x_p, x_q).$$

激光雷达因子提供高精度的几何约束,能够在长距离运动时抑制漂移;视觉因子则利用图像特征匹配实现回环检测和关键帧关联,为系统带来全局一致性的约束;IMU因子则以高频率的惯性数据补偿运动过程中的快速动态,保证在激光或视觉短暂失效时仍能维持连续估计。通过最大化联合后验概率,系统在同一优化目标下同时考虑3类传感器的观测与噪声协方差,实现时空上一致的状态估计。

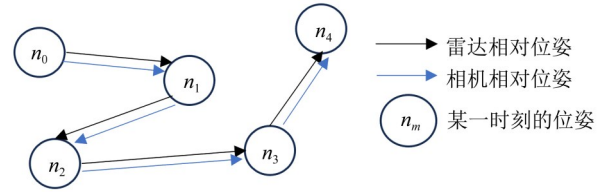


图9 因子图结构

Fig. 9 Factor graph structure

此外,在因子图优化框架中,回环检测不仅仅是一个独立的感知模块,它在数学建模上同样可以抽象为一条特殊的约束因子即回环边。当系统识别出当前时刻的关键帧与历史某一关键帧对应着同一物理位置时,就会在表示这两个关键帧的节点之间新增一条因子边,用来表达“这2个位姿应当一致”的约束关系。换言之,回环检测提供的是一种全局的闭合约束,它与里程计边、IMU边、激光雷达边等一起,统一嵌入到同一个因子图结构中。

在图10中,可以直观地看到:原本顺序连接的位姿节点通过一条横跨时间的“回环边”直接相连,这条边的存在意味着后端优化器在求解整体轨迹时,必须同时满足该闭环测量所带来的残差约束。这样做的结果是,局部累计的漂移会被全局一致的回环观测“拉回”到真实位置,从而显著降低长时间运行产生的累积误差,为了获得最优轨迹与地图,因子图作为后端时的优化问题就是求解一个最大

后验估计

$$\hat{X} = \arg \max_X p(X | \mathcal{Z}).$$

这个问题也可以等价于最小化负对数似然,即带加权残差的非线性最小二乘问题:

$$\arg \min_X \left[ \left\| r_{ij}^{\text{lidar}}(x_i, x_j) \right\|_{\Sigma_{ij}^{\text{lidar}}}^2 + \sum_{(i,j) \in \mathcal{E}_{\text{imu}}} \left\| r_{ij}^{\text{imu}}(x_i, x_j) \right\|_{\Sigma_{ij}^{\text{imu}}}^2 + \sum_{(p,q) \in \mathcal{E}_{\text{vision}}} \left\| r_{pq}^{\text{vision}}(x_p, x_q) \right\|_{\Sigma_{pq}^{\text{vision}}}^2 \right],$$

其中  $r(\cdot)$  为各因子的残差函数,  $\Sigma$  表示相应观测噪声的协方差。

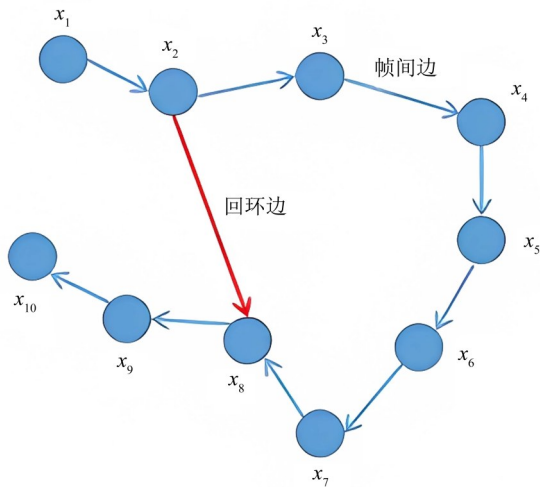


图 10 回环边与帧间边

Fig. 10 Loop edges and inter-frame edges

在程序和算法实现上,本研究使用ROS进行节点间通信,ROS算法架构如图11所示:

本文的改进方法通过在统一目标函数下联合雷达的几何约束、视觉的回环约束与IMU的高频动态约束,系统能够在时空上一致地推断全局状态,实现对环境变化及单一传感器失效的鲁棒定位与建图。如此,雷达的几何精度、视觉的全局识别能力与IMU的短时稳定性在同一因子图内形成互补,最终得到对环境变化与传感器缺失都具备鲁棒性的全局轨迹和地图。

## 4 数据集验证

### 4.1 仿真环境测试

为了验证多传感器融合算法的可行性与鲁棒性,本文在Gazebo与ROS框架下构建了一个完整的仿真与测试平台,如图12所示。

Gazebo作为高保真三维物理仿真器,不仅能精确模拟动力学约束与传感器特性,还能生成符合实际噪声模型的观测数据。ROS则承担数据传输与模块调度的中枢角色,通过其话题(topic)与服务(service)机制,实现激光雷达、视觉相机及IMU节点之间的实时同步与管理。在此平台上,建立了一个包含道路、建筑物与动态障碍物的物理仿真环境,以接近真实室外场景的尺度与纹理为蓝本,生成高密度激光点云、惯性测量以及高清图像序列。

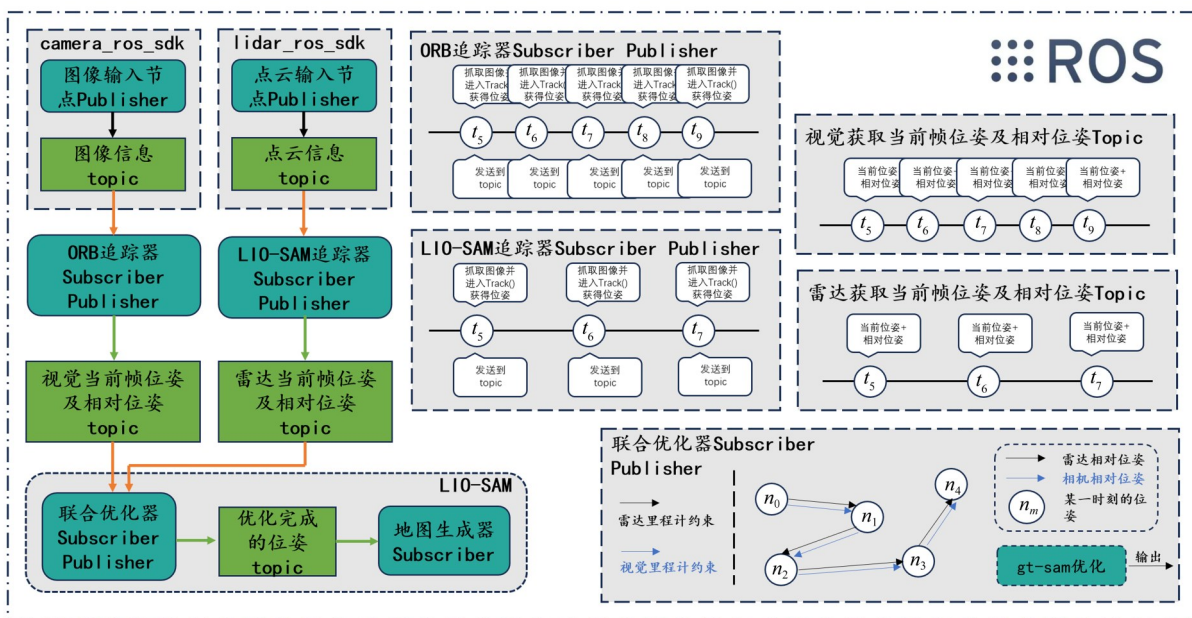


图 11 算法架构

Fig. 11 Algorithm architecture

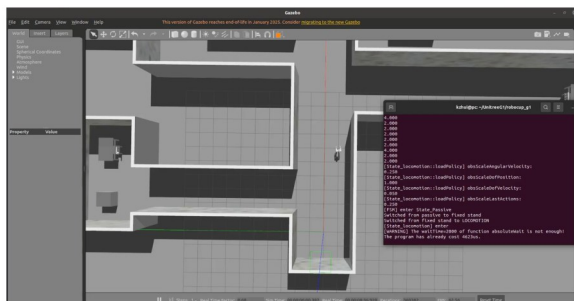


图12 Gazebo 仿真环境

Fig. 12 Gazebo simulation environment

如图13所示,机器人在该环境中按照预设轨迹运行,系统实时接收仿真传感器数据并执行SLAM算法,实现在线的模拟建图与定位过程。通过这一流程,可直观检验地图构建的完整性、回环检测的可靠性以及轨迹估计的精度。

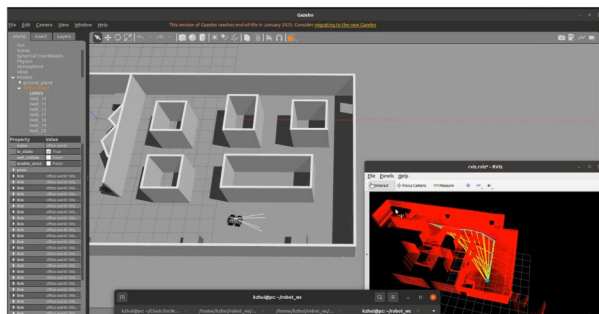


图13 仿真与建图效果

Fig. 13 Simulation and mapping effect

#### 4.2 校园数据集测试

为了在真实场景中验证所提出多传感器融合SLAM系统的有效性与稳健性,本文在校园环境中自建数据集开展实验。实验场景覆盖教学楼、绿化区及道路交叉口等多样化区域,具有建筑密集、光照变化明显、动态行人和车辆干扰等典型特征,可充分考察算法在复杂户外环境下的适应能力。

数据采集平台搭载双目相机、三维激光雷达和高频IMU,通过ROS进行时间同步与话题记录。按照预设路径,机器人在校园内连续运行完成多圈采集,获得了包含图像序列、点云数据和惯性测量的完整数据包。随后,利用所提出的紧耦合激光雷达-视觉-惯性因子图优化框架对数据集进行离线处理与建图。

如图14~15所示,在道路环境和植被繁杂的环境中,系统在长时间运行过程中保持了稳定的轨迹估计,生成的三维稠密地图结构完整、几何细节清

晰,与实际校园布局高度一致。整个建图过程未出现明显的漂移或回环误匹配,验证了本文方法在大尺度、光照多变及存在动态干扰的真实场景中的可行性与鲁棒性,也进一步证明所设计多传感器融合框架能够满足实际机器人自主导航的需求。

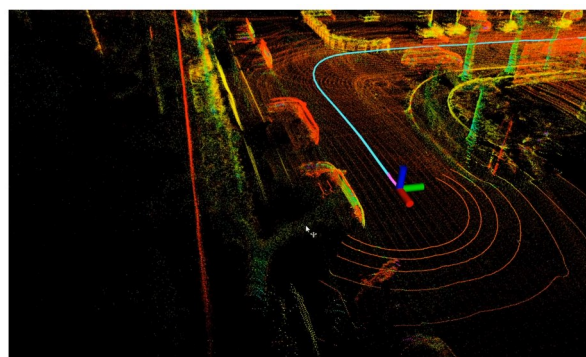


图14 道路环境建图

Fig. 14 Road environment mapping

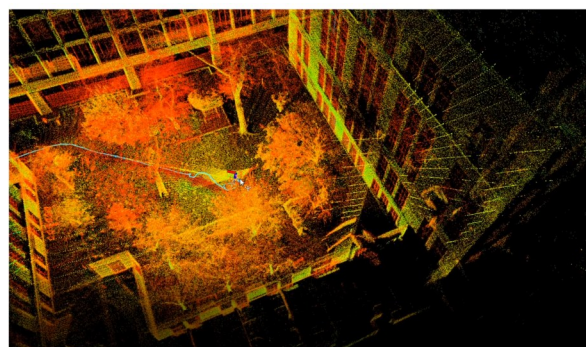


图15 植被环境建图

Fig. 15 Vegetation environment mapping

## 5 结语

本文提出了一种基于激光雷达、视觉与惯性测量单元的紧耦合因子图SLAM系统。通过仿真算法验证实验以及在校园环境中自建数据集并完成建图的实验,可以发现系统能够在光照变化大、结构复杂且存在动态干扰的真实场景中稳定运行,生成的三维地图与实际环境高度一致,验证了所提出方法的可行性与鲁棒性。

未来可从如下两个方面开展研究:一是针对大规模室内外混合场景,研究更高效的回环检测与全局优化策略,以提升实时性与长期稳定性;二是可将当前的离线实验扩展至在线自主导航系统,并评估其在真实移动机器人平台上的长期运行性能。结合上述改进,有望进一步提升本文所设计的系统在复杂环境中的适应性与应用价值。

## 参考文献:

- 蔡培培,刘磊,董凯,等,2025.基于图像稀疏表示的移动机器人视觉SLAM回环检测方法研究[J].自动化与仪器仪表,8:118-122.
- 丁健,唐泽安,吕奕新,等,2025.基于ORB-SLAM3的实时跟踪与稠密建图[J].中国高新科技,13:121-123.
- 冯迎宾,李亚玮,王天龙,等,2025.动态场景下基于激光雷达的SLAM算法[J].沈阳理工大学学报,44(5):1-7.
- 洪志云,2025.多传感器融合系统在家用扫地机器人中的应用[J].产业创新研究,16:83-85.
- 蒋风洋,程瑶,韩哲,等,2025.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),55(4):72-83.
- 李旭,阳奥凯,刘青,等,2025.基于改进ORB-SLAM2算法的温室机器人定位与稠密建图方法[J].农业机械学报,56(8):427-437.
- 柳炳琦,张峻山,唐辉,等,2025.激光与视觉SLAM的算法综述[J/OL].激光与光电子学进展.[2025-07-15].<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD&filename=JGDJ2025071400F>.
- 陆佳嘉,惠彦凯,杭琦琳,等,2025.基于改进ORB的视觉里程计设计[J].电子设计工程,33(16):1-6.
- 罗娟,赵瑞祺,张传伟,等,2025.基于语义信息的无人车激光-惯性SLAM技术研究[J/OL].激光与光电子学进展,1-29.[2025-09-24].<https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFD&filename=JGDJ2025092300P>.
- 王宁,王琪龙,汤奥斐,等,2025.基于特征单一隧道环境的LIDAR/IMU紧耦合建图方法[J/OL].激光与光电子学进展.<https://doi.org/10.3788/LOP251669>.
- 吴蕊,李振华,蔡宇鑫,2025.基于ROS的视觉导航机器人设计[J].物联网技术,15(12):74-77+82.
- 吴永豪,李胜,邹文成,2025.动态场景中的多传感器融合SLAM算法研究[J/OL].电子测量与仪器学报.[2025-08-14].<https://kns.cnki.net/kcms/detail/11.2488.tn.20250814.1403.004.html>.
- 余浩扬,李艳生,肖凌励,等,2025.面向动态环境的巡检机器人轻量级语义视觉SLAM框架[J].电子与信息学报,47(10):3979-3992.
- 余夏怡,郑恩辉,孙海迪,2025.基于目标检测和光流法的动态SLAM算法[J].现代电子技术,48(17):48-54.
- 颜金龙,2025.基于ORB-SLAM3的水下机器人定位方法研究[J].机电技术,48(3):1-6.
- 曾宪阳,于浩,梁远生,等,2025.LIO-SAM改进:自适应降采样与特征筛选优化[J/OL].仪器仪表学报,1-9.<https://doi.org/10.19650/j.cnki.cjsi.J2514244>.
- 周淑凯,张俊杰,田恬,等,2025.四足机器人室内导航的多模块优化Fast-SLAM算法研究[J].现代信息科技,9(18):169-173.
- 周永光,李晓彬,梅海青,等,2025.基于多传感器融合SLAM的AGV自动装车导航系统设计[J].烟草科技,58(9):107-112.
- 周云水,高澄宇,黄圣杰,等,2025.一种高动态场景下视觉激光融合SLAM系统[J].汽车工程,47(9):1712-1720.
- CHANG H, CHOI J B, YEUM C M, 2026. 3D reconstruction by looking: Instantaneous blind spot detection for indoor SLAM through mixed reality [J]. Adv Eng Inform, 69: 04065.
- LEE J, CHUNG W, KIM J, 2025. LiDAR-based long-term mapping in snow-covered environments [J]. Sensors, 25(21):6805.
- LIU Y F, KUANG Y J, LIU C K, et al, 2026. Map-aided visual SLAM factor graph optimization algorithm for vehicle positioning in GNSS-denied areas [J]. Measurement, 259:119687.
- MA Z W, CHENG W S, 2024. Visual-inertial RGB-D SLAM with encoder integration of ORB triangulation and depth measurement uncertainties [J]. Sensors, 24(18):5964.
- MUSLIM M T, SELAMAT H Z, ABURAYA A, 2025. YOSO-SLAM: A real-time object visual SLAM for dynamic scenes with semantic three-dimensional mapping [J/OL]. Arab J Sci Eng. <https://doi.org/10.1007/s13369-025-10840-410.1007/s13369-025-10840-4>.
- WANG J W, WAN W P, 2022. Research on the effectiveness of monocular visual SLAM depth estimation base on improved ORB algorithm [J]. Proc Int Conf Artif Life Robot, 27: 318-322.
- XIAO F, ZHANG Y H, FANG J, et al, 2025. Improving the navigation optimization of hospital logistics robots under complex lighting changes by using improved ORB-SLAM3 and deep learning visual SLAM algorithm [J]. Discov Appl Sci, 7(4): 291.
- XU J, QI H, XU M, et al, 2022. Tracking and mapping strategy for indoor UAV based on entropy theory: An ORB-SLAM3 extension [C]// The 41st Chinese Control Conference. Hefei, China: 3267-3272.