

基于点线特征融合的延迟边缘化视觉惯性 SLAM 方法

齐咏生^{1,2} 宋继鹏¹ 刘利强^{1,2} 苏建强^{1,2} 张丽杰^{1,2}

(1. 内蒙古工业大学电力学院, 呼和浩特 010080;

2. 内蒙古自治区高等学校智慧能源技术与装备工程研究中心, 呼和浩特 010080)

摘要: 针对单一传感器 SLAM 技术在复杂环境中存在精度低、感知退化、可靠性差, 导致无法准确估计摄像机轨迹的问题, 提出一种基于点线特征融合的延迟边缘化视觉惯性 SLAM 算法 (DM-VI-SLAM)。首先, 采用因子图优化模型, 提出以 IMU 为主系统, 视觉为辅系统, 通过引入辅系统观测因子约束 IMU 主系统偏差, 并接收 IMU 里程计因子实现运动预测与融合的全新结构。其次, 在前端加入点线特征, 设计一种基于线段中点的特征匹配方法, 在后端加入滑窗机制实现历史状态信息回溯, 并构建非线性联合优化问题, 提升匹配精度。最后, 为加速求解, 引入一种延迟边缘化策略, 允许重新推进延迟因子图, 进而产生新的和一致性的线性化点更新边缘化。通过与典型 SLAM 算法进行比较, 并在 EuRoC 公共数据集上和真实场景中验证算法有效性, 实验结果表明在复杂高速运动场景和低特征纹理场景中本文算法均具有更高的精度和可靠性。

关键词: 同时定位与建图; 惯性测量单元; 多传感器融合; 非线性优化; 位姿估计

中图分类号: TP242.6 **文献标识码:** A **文章编号:** 1000-1298 (2024) 12-0373-10

OSID:



Delayed Marginalization Visual Inertia SLAM Method Based on Point and Line Feature Fusion

QI Yongsheng^{1,2} SONG Jipeng¹ LIU Liqiang^{1,2} SU Jianqiang^{1,2} ZHANG Lijie^{1,2}

(1. College of Electric Power, Inner Mongolia University of Technology, Hohhot 010080, China

2. Intelligent Energy Technology and Equipment Engineering Research Centre of College and Universities in the Inner Mongolia Autonomous Region, Hohhot 010080, China)

Abstract: A delayed edge based visual inertial SLAM algorithm (DM-VI-SLAM) based on point line feature fusion was proposed to address the issues of low accuracy, perceptual degradation, and poor reliability of single sensor SLAM technology in complex environments, which made it difficult to accurately estimate camera trajectories. Firstly, a factor graph optimization model was employed, proposing a novel structure that took the inertial measurement unit (IMU) as the primary system and vision as the auxiliary system. This structure introduced auxiliary system observation factors to constrain the biases of the IMU primary system and receiving IMU odometer factors to achieve motion prediction and fusion. Secondly, by adding point and line features in the front-end, a feature matching method based on the midpoint of a line segment was designed. A sliding window mechanism was added in the back-end to achieve historical state information backtracking, and a nonlinear joint optimization problem was constructed to improve matching accuracy. Finally, to accelerate the solution, a delayed marginalization strategy was introduced that allowed for the readvancement of the delay factor graph, thereby generating new and consistent linearization points to update the marginalization. By comparing with typical SLAM algorithms and verifying their effectiveness on EuRoC public datasets and real scenes, experimental results showed that the proposed algorithm had higher accuracy and reliability in complex high-speed motion scenes and low feature texture scenes.

Key words: simultaneous localization and mapping; IMU; multi-sensor fusion; nonlinear optimization; pose estimation

收稿日期: 2024-01-13 修回日期: 2024-03-20

基金项目: 国家自然科学基金项目 (62363029)、内蒙古科技计划项目 (2020GG0283、2021GG164) 和内蒙古自然科学基金项目 (2022MS06018、2021MS06018)

作者简介: 齐咏生 (1975—), 男, 教授, 博士, 主要从事机器人协同控制技术研究, E-mail: qys@imut.edu.cn

0 引言

同时定位与建图 (Simultaneous localization and mapping, SLAM^[1])被认为是无人机、移动机器人等自主定位与导航作业的核心技术之一,现在已被广泛应用于规划无人机、无人车自动驾驶和虚拟现实等领域。使用单一传感器会出现初始化、尺度问题以及视野阻碍问题,在系统运行过程中会产生累积误差导致轨迹漂移,这已成为移动机器人运动控制的主要难点。

针对不同的环境,传感器的选择不尽相同,经典的激光雷达 SLAM 方案有 LOAM^[2]、A-LOAM^[3]等,它们所构建的地图精度高,对环境的光照和纹理等因素依赖性较小,但是在长廊或是动态变化大的环境中,单纯依靠激光雷达 SLAM 容易发生定位丢失的情况。经典的视觉 SLAM 方案有 ORB-SLAM2^[4]、SVO^[5]、DSO^[6]等,视觉 SLAM 可以直接进行对象检测和跟踪,提供丰富的环境信息,有助于后续处理复杂动态环境^[7],但在动态环境下或者光线变化较大时,可能会导致定位精度下降。因此,为了提高定位建图算法在不同场景中的适用性和对位姿估计的准确性,研究者开始提出将多种传感器组合,利用不同传感器的优势来互补缺陷,例如视觉惯导、激光惯导、激光视觉惯导等多传感器融合的方法^[8-11]。多传感器融合 SLAM 能够更有效地获取自身位置,提高机器人运动时的稳健性。例如,基于视觉与 IMU 融合 SLAM 系统,经典算法有 VINS-Mono^[12]、VINS-Fusion^[13]、ORB-SLAM3^[14]等。这些方案可以通过 IMU 测量恢复尺度辅助视觉系统初始化,利用 SFM^[15]技术,通过纯视觉方法在滑动窗口内估计所有帧的位姿和特征点的逆深度,并与 IMU 预积分数据进行对齐,从而确定视觉尺度和初始状态量。通过融合 IMU 的度量测量,可以将尺度、滚动角和俯仰角完全恢复,提高视觉系统的性能,解决系统短时间内的特征信息缺失问题,从而提升系统的鲁棒性和精度,提高视觉系统的性能。传统的基于点特征的 SLAM 算法在一些特定环境下,例如室内弱纹理场景,可能由于难以追踪到足够多的有效特征点而导致系统定位精度和鲁棒性较差,甚至完全失效。基于点线特征的视觉 SLAM 算法,如 PL-SLAM^[16]和 PL-VINS^[17]基于点线特征并融合惯性测量单元 (IMU) 的双目视觉惯性 SLAM 算法等,和近年来在传统 SLAM 方案上的优化和改进^[18],是通过结合点特征和线特征,解决在复杂环境中精度低、感知退化、可靠性差等问题。然而,引入线特征虽然可以提高在这些困难环境下的精度与

鲁棒性,但也可能增加特征提取与匹配环节时间消耗,从而降低系统运行效率。

针对上述问题,本文充分考虑多传感器各自的优势,设计一种视觉惯性 SLAM 方法,以与环境无关的 IMU 里程计为主系统,视觉惯性里程计 (VIO) 为辅系统。在 IMU 里程计主因子图中以紧耦合的方式引入子约束 IMU 偏差,通过 VIO 系统 IMU 里程计实现运动预测,有效提升系统的鲁棒性和精度。在前端部分加入点线特征的点线融合 SLAM,使用 LSD^[19]进行线段提取,采用线段中点进行特征匹配,降低特征提取与匹配环节的时间消耗。之后,在后端使用光束平差法 (Bundle adjustment, BA^[20]) 加入滑窗机制实现历史状态信息回溯,将构建一个非线性联合优化问题。为防止保留大量历史状态信息,导致计算矩阵维数增加,计算成本变高,引入一种延迟边缘化策略,优化计算成本,从而提升系统运行效率,保留系统历史约束的同时,进一步提升视觉惯性系统精度。最后在含有视觉信息和惯性信息的数据集上进行仿真实验,并在室内弱纹理和光线照射场景中进行稠密地图构建与性能测试,以验证本文方法的有效性。

1 几何约束与状态估计原理

1.1 极线几何和本质矩阵

在视觉 SLAM 的理论基础中,通过极线几何和本质矩阵,能够建立起多个相机视图之间的几何关系。考虑 2 个相机视图,它们之间的关系可以通过本质矩阵 E 来表示。

2 个相机视图的图像坐标分别为 $\boldsymbol{x}$ 和 $\boldsymbol{x}'$,通过本质矩阵表示的关系为

$$\boldsymbol{x}'^T E \boldsymbol{x} = 0$$

(1)

式(1)表达了极线几何基本原理,对于在一幅图像中找到的特征点,该特征点在另一幅图像上可能位于一条线上。本质矩阵包含相机内部参数信息以及 2 个相机之间的运动关系,即

$$E = [\boldsymbol{t}]_{\times} \boldsymbol{R}$$

(2)

式中 $\boldsymbol{t}$ ——平移向量

$[\boldsymbol{t}]_{\times}$ —— $\boldsymbol{t}$ 反对称矩阵

$\boldsymbol{R}$ ——旋转矩阵

通过分解本质矩阵,可以恢复相机运动姿态。

1.2 贝叶斯滤波与状态估计

在 SLAM 中,需要实时地估计相机姿态以及场景中的地图。这涉及到状态估计,而贝叶斯滤波是这一任务的常用方法之一。考虑一个简化的 SLAM 问题,包含相机位姿 $\boldsymbol{x}$ 和地图状态 $\boldsymbol{m}$,则有

$$P(\mathbf{x}, m | z, u) = \frac{P(z | \mathbf{x}, m) P(\mathbf{x} | u) P(m)}{P(z | u)} \quad (3)$$

式中 z ——观测数据

u ——运动控制数据

$P(\mathbf{x}, m | z, u)$ 是在观测数据 z 和控制输入 u 的条件下,状态 $\mathbf{x}$ 和地图 m 的联合后验概率。它表示在给定观测和控制输入情况下,相机姿态和地图状态的概率分布。 $P(z | \mathbf{x}, m)$ 是观测模型,表示在给定状态 $\mathbf{x}$ 和地图 m 的条件下观测到数据 z 的概率。它描述了状态和地图如何生成观测数据。 $P(\mathbf{x} | u)$ 是运动模型,表示在给定控制输入 u 的条件下状态 $\mathbf{x}$ 的概率。它描述了控制输入如何影响状态的转移。 $P(m)$ 是先验模型,表示地图 m 的先验概率。它提供了在进行任何观测之前对地图状态的初始信念。 $P(z | u)$ 是在给定控制输入 u 的条件下观测到数据 z 的概率。它是通过边缘化状态和地图变量得到的,即 $P(z | u) = \sum_{\mathbf{x}, m} P(z | \mathbf{x}, m) P(\mathbf{x} | u) P(m)$ 。这个项作为归一化常数,确保联合后验概率的总和为 1。

公式(3)右侧的各项分别表示观测模型、运动模型和先验模型。通过贝叶斯滤波,能够更新状态的后验概率,即估计出相机姿态和地图的当前状态 $P(\mathbf{x}, m | z_{1:t}, u_{1:t})$ 。其中, $z_{1:t}$ 表示时间从 1 到 t 的观测序列, $u_{1:t}$ 表示时间 1 到 t 的运动控制序列。

这个后验概率是 SLAM 问题的核心,通过最大

化后验概率,可以获得对相机姿态和地图的最优估计。

2 DM-VI-SLAM 算法

2.1 算法总体结构设计

延迟边缘化视觉惯性 SLAM 算法 (DM-VI-SLAM) 结构设计如图 1 所示,以 IMU 作为主体坐标系,进行龙格库塔预积分^[21],经过前端数据处理与视觉数据对齐,进行特征提取后用光流追踪的方式匹配,以紧耦合的方式构建 VIO,由于视差提供了关键的场景深度信息,从而可能在三维空间中更准确地匹配特征点。特征点匹配和光流法不仅能够提供相邻帧之间的视觉运动信息,有助于构建关键帧之间的时空关系,而且结合 IMU 的先验姿态信息,有效地降低了视觉里程计的累积误差,显著提升了系统的鲁棒性。因此,根据关键帧当前状态和先验信息进行位姿估计,以最大程度地综合利用这些关键信息来提高系统的性能。并采用 PnP 方法通过最小化重投影误差来估计相机的姿态,进而得到更准确的位姿信息;将位姿信息以因子图的形式进行优化,通过滑动窗口和延迟边缘化的方式提升系统运算效率;本算法还将局部优化后的 VIO 里程计联合视觉回环检测进行全局位姿优化,以提高全局地图与运动轨迹之间的精度。完成全部的位姿估计后,最后进行稠密建图。

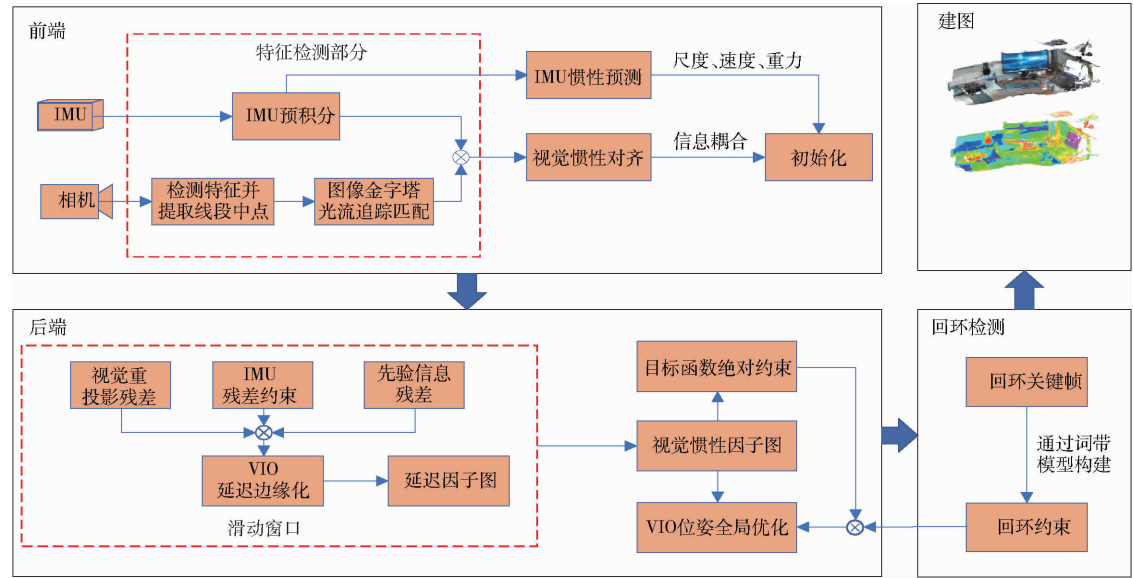


图 1 DM-VI-SLAM 算法整体框架图

Fig. 1 DM-VI-SLAM algorithm framework

2.2 VIO 视觉惯性里程计系统

2.2.1 点线特征提取

前端视觉惯性里程计采用双目相机输入的左右目图像,在图像中进行关键帧的筛取、Shi-Tomasi^[22]角点检测、KLT 稀疏光流法追踪。Shi-

Tomasi 角点作为 Harris^[23]角点的改进,使其对噪声更加鲁棒,并能够更准确地控制检测到的角点数量。为进一步提升精度,特征点根据像素坐标计算出亚像素坐标,再根据亚像素坐标点索引和速度采用双向光流追踪和左右目相机帧双向追踪 2 种模式,并

基于反向光流法追踪结合 RANSAC^[24] (Random sample consensus)方法剔除误差点;线特征使用 LSD 算法进行匹配搜索,根据线段中点坐标与特征点像素坐标位置进行合理取舍,防止特征点与特征线段重叠或同处于某些特征密集区域。RANSAC 方法从数据集中随机选择小部分数据作为样本集,然后使用这些样本来估计模型参数,用于估计模型参数的迭代方法,处理数据中存在离群值的情况。如果图像过暗或过亮,采取均衡化策略,并且使特征点均匀化,防止特征点过于密集。

在运动后若前后两帧角点距离相对较远,容易在此区域内陷入局部最小问题,出现误匹配。为了提升稳定性,采用图像金字塔方法。通过图像金字塔,可以在不同尺度下对图像进行处理,使得在较小尺度下更容易捕捉到细微的运动特征,从而更有效地避免陷入局部最小问题。金字塔结构的应用有助于提高算法对运动场景的鲁棒性,使得在不同情况下都能更准确地进行角点匹配,从而提高整体的匹配精度。再使用光流追踪,根据前一帧特征点坐标判断新一帧坐标和其范围内的一小块区域,最终落实到点对点,将上一层图像金字塔光流追踪的结果作为下一层的初值,保证每层出现的误差不会累加,最后趋于收敛。

线特征采用普吕克线坐标^[25],普吕克线坐标 L 可以由一个方向向量 d 和过坐标原点及直线的平面法向量 n 组成

$$L = [n^T \quad d^T]^T \tag{4}$$

一条直线只有 4 个自由度,为直线绕自身方向的旋转和沿着直线方向的平移。通过变换矩阵 T 进行坐标系变换,可以得到新坐标系下的普吕克线坐标 L' ,这个变换实际上涉及到了对点和直线进行投影、旋转和平移等操作,确保了在不同坐标系下仍能准确描述同一条直线。具体计算式为

$$L' = TL \tag{5}$$

其中
$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{6}$$

$$\begin{cases} SO(3) = \{R \in \mathbf{R}^{3 \times 3} | RR^T = I, \det(R) = 1\} \\ SE(3) = \left\{T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in \mathbf{R}^{4 \times 4} | R \in SO(3), t \in \mathbf{R}^3\right\} \end{cases} \tag{7}$$

式中 $SO(3)$ ——正交旋转矩阵群
 $SE(3)$ ——变换矩阵群
 I ——单位矩阵

2.2.2 视觉惯性约束重投影误差

VIO 视觉惯性里程计中,优化问题中有预积分约束,也有重投影约束。视觉重投影常用于验证或

优化相机姿态估计或三维点云重建的准确性。重投影误差即是原始三维点与反向投影后的点在图像上位置之间的差异,如图 2 所示。

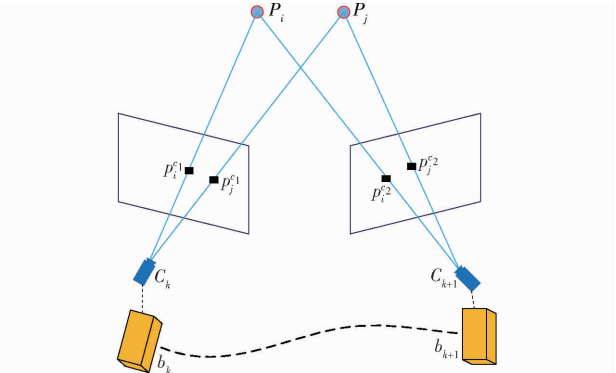


图 2 视觉重投影示意图
Fig.2 Visual reprojection diagram

视觉重投影是计算机视觉中用于三维重建和相机姿态估计的一个重要步骤。首先,通过立体视觉或结构恢复运动(SFM)方法,可以获得场景中三维点的位置以及相机在不同时刻的姿态。接着,这些三维点通过相机的内部参数矩阵被投影到图像平面上,生成对应的二维图像点坐标。然后,为了验证和优化估计的准确性,使用相机内部参数矩阵的逆矩阵将这些二维点反向投影回三维空间,得到新的三维点位置。最后,通过计算原始三维点与反向投影得到的三维点之间的差异,通常用像素坐标的欧氏距离来表示,可以评估三维重建精度和相机姿态的准确性。这个过程确保了三维点在图像平面上的投影与实际观测到的二维图像点尽可能一致,从而提高了整个系统的准确性和鲁棒性。

假设真实线段 2 个端点在图像上的坐标为 $P_{true}(x_1, y_1)$ 和 $P_{true}(x_2, y_2)$,而算法检测到的线段 2 个端点在图像上的坐标为 $P_{detected}(x'_1, y'_1)$ 和 $P_{detected}(x'_2, y'_2)$,线段重投影误差 E 计算式为

$$E = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_{1,i} - x'_{1,i})^2 + (y_{1,i} - y'_{1,i})^2 + (x_{2,i} - x'_{2,i})^2 + (y_{2,i} - y'_{2,i})^2} \tag{8}$$

式中 N ——线段数量 i ——线段序号

绝对角度误差 E_{AAE} 关注真实线段和检测到线段的方向。对于每条线段,计算它们的方向向量,然后使用反余弦函数计算它们之间的夹角。

$$E_{AAE} = \frac{1}{N} \sum_{i=1}^N \arccos \frac{uu'}{\|u\| \|u'\|} \tag{9}$$

式中 u ——真实线段方向向量
 u' ——检测线段方向向量

绝对角度误差对于角度偏差非常敏感,因为它直接度量了线段方向的差异。

中点投影误差 E_{MPE} 关注线段的中点,而不是整

体方向。对于每条线段,计算真实线段和检测到线段的中点坐标,然后计算这 2 个中点之间的距离。

$$E_{\text{MPE}} = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_{\text{mid},i} - x'_{\text{mid},i})^2 + (y_{\text{mid},i} - y'_{\text{mid},i})^2} \quad (10)$$

式中 $(x_{\text{mid}}, y_{\text{mid}})$ ——真实线段中点坐标

$(x'_{\text{mid}}, y'_{\text{mid}})$ ——检测线段中点坐标

中点投影误差关注了线段中点位置差异。采用重点投影误差和绝对角度误差作为系统优化变量。

采用线段中点进行误差重投影,线段中点重投影误差考虑了线段整体几何信息,因此更为稳健。此外,线段中点相对于端点更为鲁棒,能够在存在噪声和不确定性的情况下提供更稳定的估计。相较于端点重投影误差,线段中点重投影误差在处理非线性畸变、图像噪声以及场景深度不确定性等方面表现更为出色。这种方法提高了对线段结构的建模精度,从而在计算机视觉任务中取得更可靠的结果。

2.3 延迟边缘化

因子图优化是一种在概率图模型中处理不确定

性和推断的方法。它基于因子图表示,其中节点表示变量,边表示这些变量之间的因果关系。每个边都与一个因子相关,表示了这 2 个变量之间的关系。

因子图优化在 SLAM 中先构建因子图,将 SLAM 问题建模为一个因子图,用节点表示机器人的位置、姿态和地图特征,边表示它们之间的关系,包括测量和运动模型。添加因子,将传感器测量和运动模型映射为因子,有效地融合各种传感器信息,与相应的图节点相连接,这些因子捕捉了传感器测量与机器人状态之间的关系。构建误差模型,加入先验因子、视觉重投影约束因子和 IMU 预积分因子,利用非线性最小二乘法,调整误差变量的值来最小化误差的平方和,调整机器人状态和地图点的估计值,以减少所有因子之间的不一致性,使用增量 QR 矩阵(正交三角分解法)分解技术,有效地处理新添加的约束,并且能够保持先前计算的结果,从而减少重复计算,最终得到机器人状态和环境地图的最优估计。因子图如图 3 所示。

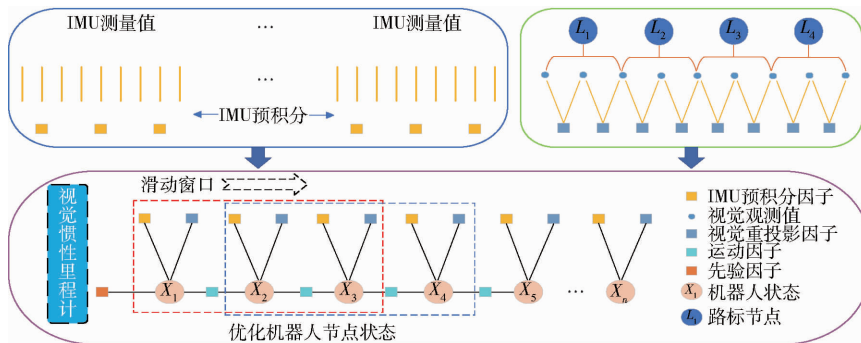


图 3 SLAM 问题中因子图的构建

Fig. 3 Construction of factor graphs in SLAM

为了管理图优化的计算复杂度,使用边缘化方法有选择地剔除信息。在滑动窗口优化中,通常会边缘化最老的帧,即时间上最早的帧,并建立与窗口内其他帧的关系。选取的枢纽帧是滑动窗口内的一个较早的帧,它与窗口内最后一帧的时间间隔尽可能大,以确保覆盖较长的时间跨度。同时,枢纽帧需要有足够的匹配点对,且匹配点对的视差也应足够大,本文选取的匹配点对大于 20 对,视差大于 30 像素,以保证足够的几何约束强度。通过枢纽帧与窗口内最后一帧之间的位姿变换关系,可以计算出滑动窗口中各帧的相对位姿。这种策略不仅减少了优化问题的规模,还确保了优化过程的稳定性和准确性。

需保留的变量和边缘化去除的变量为 δx_1 和 δx_2 ,则有

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (11)$$

式中 H_{11} ——未边缘化变量因子雅可比矩阵

H_{12}, H_{21} ——变量间相互作用雅可比矩阵

H_{22} ——边缘化变量因子雅可比矩阵

b_1 ——未边缘化变量误差向量

b_2 ——边缘化变量误差向量

式(11)两边同时左乘矩阵,得

$$\begin{bmatrix} I & -H_{12}H_{22}^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} I & -H_{12}H_{22}^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad (12)$$

$$\begin{bmatrix} H_{11} - H_{12}H_{22}^{-1}H_{21} & O \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \end{bmatrix} = \begin{bmatrix} b_1 - H_{12}H_{22}^{-1}b_2 \\ b_2 \end{bmatrix} \quad (13)$$

式中 O ——零矩阵

之后将矩阵要保留信息的一行展开,得

$$(\mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21})\delta\mathbf{x}_1 = \mathbf{b}_1 - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{b}_2 \quad (14)$$

每完成一步边缘化后,先前时刻位姿变量将用于下一时刻位姿变量的计算并产生新的关联约束。这表明先前时刻位姿变量的线性化点必须是固定的,且时间上保持连续。若前一帧所含的尺度信息被边缘化掉,整个系统的尺度将无法测定,未来的三维地图点将会有有一个不确定的尺度信息,延迟边缘化可以解决此种问题。

延迟边缘化的思路是:边缘化不能消除,但可以推迟,如图 4 所示。延迟边缘化并非立即执行边缘

化过程,而是通过引入第 2 个延迟边缘化先验,将帧的边缘化在时间上延迟了一定量。这种延迟执行的策略为系统提供了更灵活的处理能力。延迟边缘化保持了与原始图像相同边缘化顺序,确保在延迟因子图始终具有相同尺寸。这有助于维持系统结构和一致性。延迟图中的每个线性化因子都与生成该因子时活跃的关键帧相连接,从而保留了位姿路标等信息的一致性,这对于视觉的准确建模问题至关重要。延迟边缘化通过引入第 2 个延迟边缘化先验和相应的因子图,实现了部分边缘化撤销。

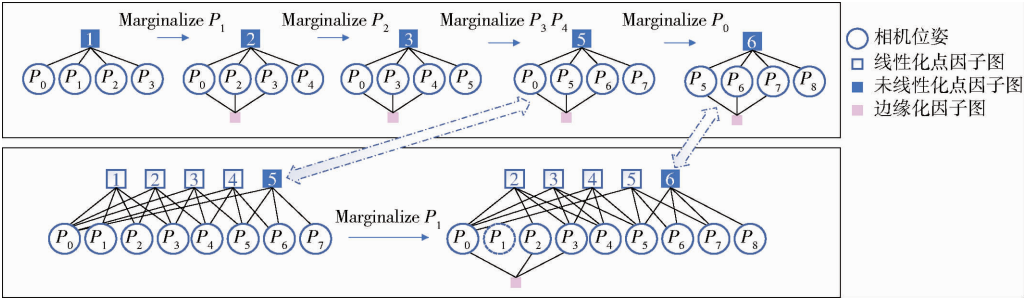


图 4 延迟边缘化示意图
Fig.4 Diagram of delayed marginalization

虽然边缘化本身无法完全撤销,但延迟边缘化允许在一定程度上回退边缘化过程^[26],从而更灵活地处理系统状态的变化。通过保持与原始图相同的边缘化顺序,延迟边缘化保持了系统的一致性。这对于维护边缘化先验的一致性至关重要,尤其是在视觉惯性里程计等系统中,可以有效处理尺度变化等问题。延迟边缘化的运行时开销非常小,即使对于任意大的延迟,它仅相当于每个延迟图的额外边缘化过程。通过引入延迟边缘化,系统能够在保持高效性的同时更灵活地处理复杂的状态变化,特别是在涉及到尺度变化等挑战性问题的场景中。这能够确保系统的高效性,同时维持与普通边缘化相近的运算复杂度。

3 实验验证

本文进行 3 个实验来评估所提出的算法。第 1 个实验为使用不同的特征点匹配方式进行特征匹配,分别为暴力匹配、Shi - Tomasi 角点匹配与 RANSAC 方法外点剔除后的匹配。第 2 个实验为本文算法在公共数据集上与其他先进的算法进行比较,用数值分析来证明该系统准确性。第 3 个实验为室内环境测试系统,以评估在重复场景中的性能。然后进行小规模实验建图,验证该系统稳定性。

3.1 特征点线提取与匹配实验

点特征通过预设 3 个最大角点数量和 2 个最小角点距离参数,分别计算特征点提取率、亚像素坐标

与实际坐标距离总和和系统运行时间。尝试不同的特征匹配方法,包括暴力匹配、金字塔光流匹配以及金字塔光流匹配结合 RANSAC 方法。

从表 1 可得出,当最小角点距离为 20 像素时,无论选取哪个预设最大角点数量,角点提取率均能达到 100%,满足预设角点数量期望,但随着最大角点数量增加,亚像素坐标与实际坐标距离总和会提高,系统运行时间会增加,表明无论是从精度还是效率上都有所下降。

表 1 不同特征点参数的实验数据
Tab.1 Experimental data for selecting parameters of different feature points

最大角点数量	最小角点距离/像素	角点提取率/%	亚像素坐标与实际坐标距离总和/像素	时间/ μs
150	20	100	247.119	1 229 468
200	20	100	319.197	1 349 923
250	20	100	390.679	1 351 588
150	25	100	240.831	1 347 559
200	25	93.0	285.621	1 316 210
250	25	74.4	285.621	1 297 711
150	30	96.7	221.165	1 224 786
200	30	72.5	221.165	1 211 614
250	30	58.0	221.165	1 200 192

当最小角点距离为 25 像素时,随着最大角点数量增加,角点提取率逐渐下降,亚像素坐标与实际坐标距离总和先增加后不变,系统运行时间减少,表明

在该种角点距离下,实际提取的角点数量会达到阈值后不再增加,设置过多的角点数量会影响精度。当最小角点距离为 30 像素时,随着最大角点数量增加,角点提取率相较于前 2 次实验降低且降低速率加快,亚像素坐标与实际坐标距离总和保持不变,系统运行时间减少,表明在该种角点距离下,实际角点数量提取优先,设置更大的角点数量是冗余操作,对精度无影响,反而降低角点提取率。综合考虑以上 3 种性能指标,从精度和效率可得最佳数据为最大角点数量 150 和最小角点距离 30 像素,若考虑角点提取率,最佳数据为角点数量 150 和最小角点距离 20 像素。

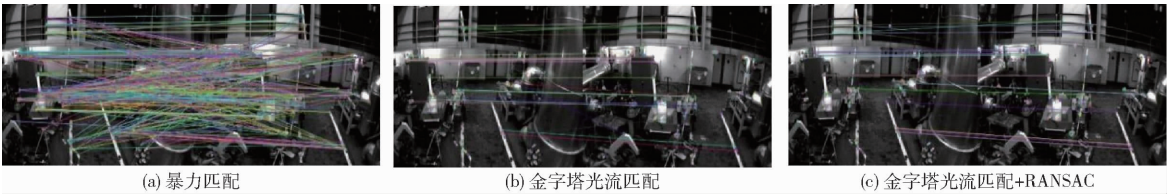


图 5 不同算法匹配结果

Fig. 5 Results of matching algorithms

在运动情况下相机会出现模糊、散焦情况,在像素运动情况下,可使用光流进行追踪匹配。在动态场景和运动相机条件下匹配结果如图 6 所示,光流技术能够有效地跟踪图像中像素的运动,对于相邻帧之间的特征点匹配具有更好的适应性。特别是在存在视角变化和快速运动的情况下,光流方法显示出卓越性能,能够准确捕捉像素级别的运动轨迹,这使得光流技术成为处理具有复杂运动情况的图像序列时的首选方法。在加入 RANSAC 方法后的光流匹配法能够有效地应对图像中存在的非一致性和异常值,图像中基本没有误匹配点,并且特征点的分布更加均匀。RANSAC 方法作为一种鲁棒的模型拟合方法,在特征点匹配中展现了明显优势。



图 6 动态场景下光流匹配结果

Fig. 6 Optical flow matching results in dynamic scenes

如图 7 所示,在不同场景下多特征可以协同帧

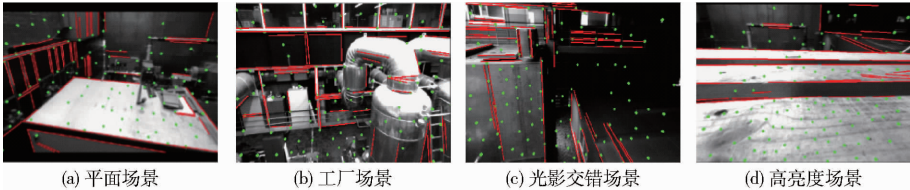


图 7 不同场景下点线特征

Fig. 7 Point and line features in different scenarios

当最大角点数量 150 和最小角点距离 20 像素时,在此基础上进行不同特征匹配方法的比较,暴力匹配、金字塔光流匹配和加入 RANSAC 方法外点剔除的金字塔光流匹配所用时间分别为 1 939 385、1 880 859、1 923 950 μ s。

结果表明:金字塔光流匹配方法相对于暴力匹配方法表现更好,具有较短的执行时间。图 5 为不同算法匹配的实验结果。由图 5 可得,特征点匹配阶段采用暴力匹配方法相较于金字塔光流匹配表现较差。暴力匹配在提取特征点和进行匹配时受到环境复杂性、视角变化、图像噪声、运动模糊的影响,导致匹配的准确性和稳健性相对较低。

间追踪。首先,在漆黑表面上,图 7 的 4 个场景中出现黑色而缺少纹理的区域,无法检测特征点的情况下,可以通过使用线特征来进行替代。由于漆黑表面缺乏明显的纹理或颜色变化,传统的点特征追踪方法可能无法准确地定位目标物体。然而,通过利用线特征,可以利用边缘信息来追踪目标物体的位置。线特征可以捕捉到物体的边缘轮廓,即使没有明显的点特征,也能够提供相对稳定的追踪结果。在光度变化迅速的场景中,线特征聚集在物体边缘。

在这种情况下,点特征可能会变得稀少,因为光线的变化导致物体表面的纹理和颜色变化较大。然而,根据该点与线段的位置关系,可以有效地避免误匹配。通过分析点特征与周围线特征的相对位置,可以判断该点是否属于目标物体,从而减少误匹配的可能性。

进行算法测试和相应的结果分析,匹配结果如图 8 所示。对于特征点和线段多的情况,在图中密集的特征点和线段中,本文提出的方法表现出更强大的特征提取能力。在这种情况下算法能够准确地捕捉到场景中的细微结构和纹理,从而提高了位姿

估计的精度。这对于需要更多几何信息的场景,如室内桌面,尤其有利。即使在纹理相对较少或简单的情况下,本文算法仍然能够有效地提取到关键的场景特征。这表明本文方法对于处理不同类型和密度的特征点和线段都具有鲁棒性,使其更具广泛适用性。多特征结合在不同场景下具有协作追踪的优势。通过利用线特征来替代漆黑表面上的特征点以及在光度变化迅速的场景中利用点特征与线特征的位置关系来避免误匹配,可以提高追踪的准确性和鲁棒性。这种优势使得多特征结合成为一种有效的追踪方法,适用于各种复杂的实际应用场景。

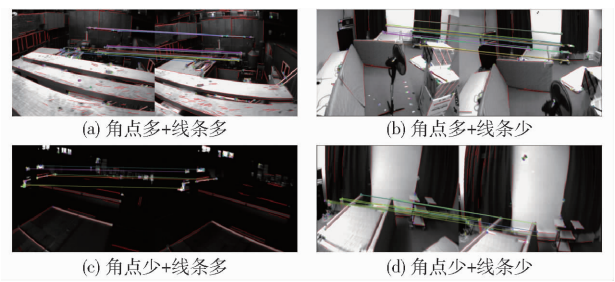


图 8 不同角点和线条的匹配结果
Fig. 8 Matching results of different corners and lines

3.2 EuRoC 数据集轨迹误差对比

EuRoC 数据集是由瑞士苏黎世联邦理工学院机器人与感知组创建的一系列用于视觉定位和 SLAM 研究的数据集。该数据集是专门为移动机器人和无人机领域设计的,并提供了高质量的传感器数据以及精确的运动轨迹地面真实值。EuRoC 数据集包含了来自微型飞行器的多模态传感器数据,包括立体相机图像、惯性测量单元测量以及一些数据集中的其他传感器数据。图像数据用于视觉 SLAM,IMU 数据用于测量运动和姿态。数据集中的相机图像通常采用全局快门相机,提供高帧率和高分辨率的图像,有助于研究在复杂场景中的运动物体或在快速移动时的视觉定位;IMU 数据提供了同步的惯性测量单元数据,包括角速度和线性加速度。数据集提供 MH(机械大厅)和 V(Vicon 房间)序列, MH 序列以工业环境为特色,包括办公室、工厂和户外环境,具有动态场景和复杂光照,其中 MH01 和 MH02 是在较为空旷的大厅中飞行,而 MH03 ~ MH05 则包含更多的动态场景;V 序列在装备有 Vicon 运动捕捉系统的房间内收集的,提供了高精度地面真实轨迹作为参考,其中 V101 和 V102 是在较为简单的室内环境中飞行,而 V103 ~ V203 则包含更加复杂的场景,如无人机在家具和障碍物之间飞行。

在验证算法前对传感器进行校准,包含 IMU 标定和 IMU 与相机的外参标定过程,标定完成后对

IMU 加速度计与陀螺仪测量值的方差进行分析,其结果如图 9 所示,图中 a_x 、 a_y 、 a_z 、 g_x 、 g_y 和 g_z 分别表示从惯性测量单元获取的加速度和陀螺仪数据,分别对应于 x 、 y 、 z 轴加速度和角速度分量。加速度计与陀螺仪在三轴上的波动平滑,波动不大,标准差逐渐降低,表明在一定时间尺度下,误差呈现较为稳定的特征,随着时间推移,系统不确定性逐渐减小。

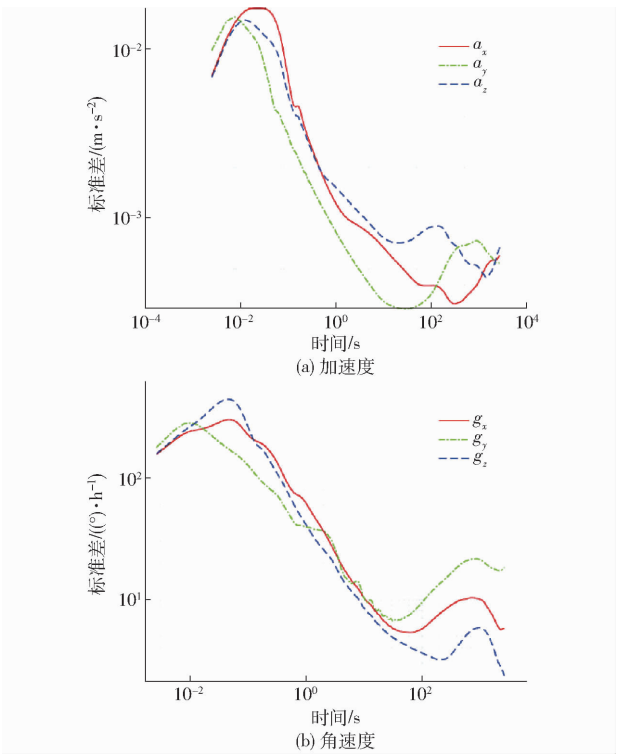


图 9 IMU 加速度计与陀螺仪测量值的方差分析
Fig. 9 Variance analysis of IMU accelerometer and gyroscope

在 SLAM 中,ATE 和 RPE 是 2 种常用的精度评价指标。ATE 全称为绝对轨迹误差 (Absolute trajectory error),它衡量估计轨迹与真实轨迹之间的全局误差,而 RPE 则是相对位姿误差 (Relative pose error),主要描述相隔固定时间间隔内的局部精度。

为更好地验证本文算法优越性,将本文算法与主流 VI-SLAM 算法 ORB-SLAM2^[4]、VINS-Mono^[12] 和 PL-VINS^[17] 进行对比,使用 EVO 进行数据评测,改进算法在 ATE 和 RPE 等方面的全方位改进,结果如表 2 所示。

其中,ORB-SLAM2、VINS-Mono、PL-VINS、DM-VI-SLAM 算法在 EuRoC 序列数据集中部分轨迹估计值与轨迹真值的一致性如图 10 所示,左侧图为算法轨迹对比,右侧图为误差密度可视化图。ORB-SLAM2 算法使用单一传感器,在数据集上均出现漂移,其整体的相对位姿误差远高于其余 3 种算法,在不同序列中均存在严重漂移现象,而本文算法在估计轨迹与真实轨迹基本一致。分析可知,当

表 2 各算法绝对轨迹误差对比

Tab. 2 Comparison of absolute trajectory errors of various algorithms

视频序列	ORB-SLAM2 算法		VINS-Mono 算法		PL-VINS 算法		DM-VI-SLAM 算法	
	ATE	RPE	ATE	RPE	ATE	RPE	ATE	RTE
MH01	0.054 0	0.331 1	0.057 5	0.033 1	0.133 7	0.032 0	0.043 6	0.021 4
MH02	0.043 4	0.288 0	0.044 9	0.026 2	0.130 8	0.015 5	0.031 6	0.026 0
MH03	0.065 4	0.388 0	0.069 4	0.083 4	0.170 0	0.080 5	0.033 0	0.082 9
MH04	0.138 7	0.490 5	0.114 6	0.121 4	0.362 6	0.122 8	0.054 9	0.120 7
MH05	0.154 7	0.382 0	0.182 8	0.105 9	0.308 2	0.106 5	0.047 7	0.055 7
V101	0.093 5	0.202 3	0.051 2	0.024 1	0.077 5	0.038 0	0.041 6	0.012 9
V102	0.063 5	0.286 0	0.057 3	0.138 0	0.101 7	0.175 7	0.057 4	0.153 4
V103	0.062 1	0.187 8	0.177 7	0.035 2	0.166 8	0.030 5	0.178 4	0.035 1
V201	0.053 0	0.118 6	0.051 3	0.033 9	0.069 7	0.035 1	0.080 6	0.030 7
V202	0.054 1	0.184 8	0.103 7	0.132 2	0.129 4	0.151 1	0.105 5	0.129 5
V203	0.229 0	0.139 5	0.159 6	0.025 1	0.229 0	0.030 6	0.224 2	0.026 2
平均值	0.091 9	0.272 6	0.097 5	0.068 9	0.170 8	0.074 4	0.081 7	0.063 1

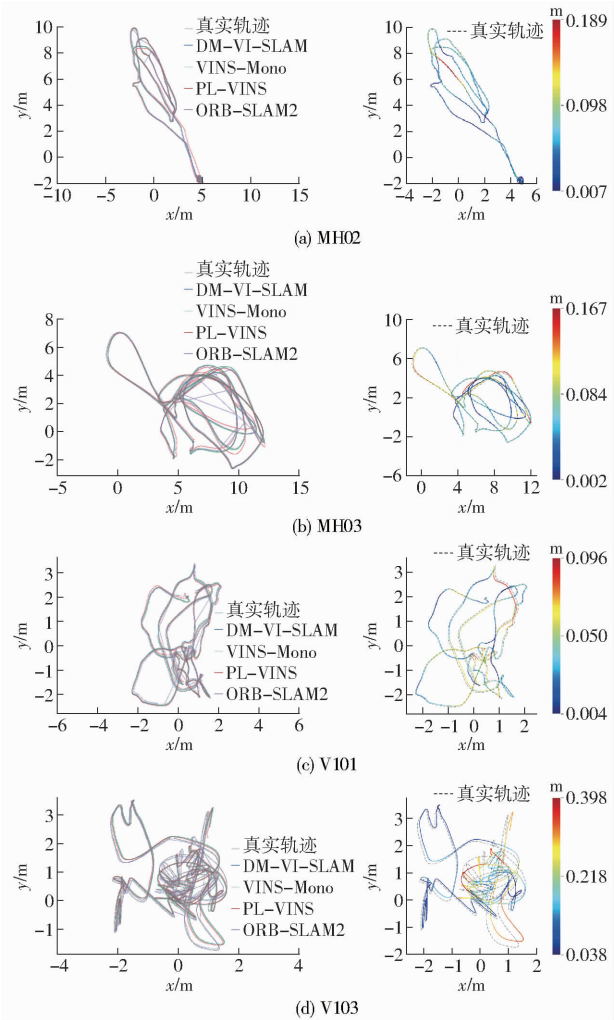


图 10 各算法绝对轨迹误差

Fig. 10 Absolute trajectory error diagrams of each algorithm

机器人长期处于黑暗、高速运动的环境中,单一传感器受到环境干扰,丢失大量信息,系统发生故障。

本文 DM-VI-SLAM 算法在整体上由于常规的视觉惯性算法,且加入线特征,以不受环境影响的 IMU 里程计为主系统,表明该系统精度更高,鲁棒性更强,在点特征消失与高速运动的场景下表现良好。

利用延迟滑窗进行边缘化,保留了大量的历史信息,随时间推移绝对轨迹误差 ATE 的变化,可看出绝大部分时间 DM-VI-SLAM 的 ATE 均为最小,运动结束后累计漂移最低,说明在复杂大场景中 DM-VI-SLAM 算法轨迹全局一致性均优于其他算法,更加准确。

3.3 稠密建图

本文算法仅使用 Intel Realsense D435i 型深度相机,仅使用双目加 IMU 在 Ubuntu 18.04 的 Linux 系统下进行建图,使用 CPU 为 2.60 GHz 的 Intel(R) Core(TM) i7-9750H。每个点云数据包含三维坐标信息,通过特征匹配、迭代最近点算法等实现对从不同视角或时间采集的点云进行配准,利用从深度图像中提取的超像素进行建模,将它们融合成一个全局一致的点云,使地图能够实时实现全局一致性。最后进行稠密建图,本文算法建图过程完全依赖于 CPU 计算,且整个过程中维持图像采集频率 25 Hz,确保了在不同环境下的良好扩展性,通过稀疏 SLAM 系统有效地估计了相机位姿,实现对强度图像和深度图像的全局一致融合,同时优化了运行时效率和内存使用,三维点云稠密建图如图 11 所示。

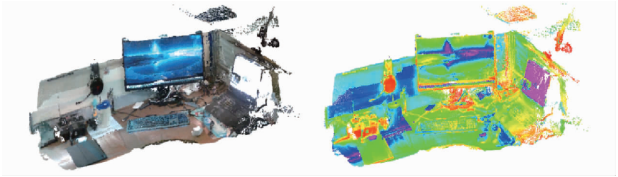


图 11 三维点云稠密建图

Fig. 11 3D point cloud dense mapping

4 结束语

采用视觉惯性融合的视觉里程计系统,前端采用金字塔光流追踪的形式为后端提供初始值,为防止点特征跟丢或产生较大误差加入了线特征,以优化视觉和惯性残差等,建立视觉惯性等约束,为提高

系统运算效率引入滑动窗口和边缘化,防止先验信息丢弃而选择延迟边缘化,因而提出一种新的视觉惯性算法 DM-VI-SLAM。在 VIO 中通过 PVQ 位姿辅助预测系统运动位置,最后与多个经典 SLAM 算法进行了全局轨迹一致性和绝对轨迹误差的对比

实验。实验结果表明,尤其是在快速运动、运动光照变化或几何结构退化的挑战性真实场景中,本文算法在里程计精度和适应能力方面均表现出色。三维点云稠密建图表明本文提出的 DM-VI-SLAM 算法在实际应用中具有显著的实用性。

参 考 文 献

- [1] BAILET T, DURRANT W H. Simultaneous localization and mapping (SLAM): part II [J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.
- [2] ZHANG J, SINGH S. LOAM: LiDAR odometry and map in real-time [C]//Robotics: Science and Systems, 2014: 1-9.
- [3] ZHANG J, SINGH S. Low-drift and real-time LiDAR odometry and mapping [J]. Autonomous Robots, 2017, 41(2): 401-416.
- [4] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [5] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: fast semi-direct monocular visual odometry [C]//Proceedings of IEEE International Conference on Robotics and Automation. IEEE, 2014: 15-22.
- [6] WANG R, SCHWORE M, CREMERS D. Stereo DSO: large-scale direct sparse visual odometry with stereo cameras [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 3903-3911.
- [7] 齐咏生, 陈培亮, 刘利强, 等. 基于单目视觉的动态环境同步定位与多地图构建算法 [J]. 农业机械学报, 2022, 53(4): 280-292.
QI Yongsheng, CHEN Peiliang, LIU Liqiang, et al. Simultaneous localization and multi-mapping algorithm in dynamic environment based on monocular vision [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 280-292. (in Chinese)
- [8] TUBMAN R, POTGIETER J, ARIF K M. Efficient robotic SLAM by fusion of RatSLAM and RGBD SLAM [C]//2016 IEEE International Conference on Mechatronics and Machine Vision in Practice. IEEE, 2016: 47-53.
- [9] LIU Z X, XIE C X, XIE M, et al. Mobile robot positioning method based on multi-sensor information fusion laser SLAM [J]. Cluster Computing, 2019, 22(Suppl. 2): 55-61.
- [10] LIN J, ZHENG C, XU W, et al. R2LIVE: a robust, real-time, LiDAR-inertial-visual tightly-coupled state estimator and mapping [J]. IEEE Robotics and Automation Letters, 2021, 6(4): 7465-7476.
- [11] LIN J, ZHANG F. R3LIVE: a robust, real-time, RGB-colored, LiDAR-inertial-visual tightly-coupled state estimation and mapping package [C]//2022 IEEE International Conference on Robotics and Automation. IEEE, 2022: 10672-10678.
- [12] 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.
- [13] QIN T, PAN J, CAO S Z, et al. A general optimization based framework for local odometry estimation with multiple sensors [J]. arXiv Preprint, arXiv:1901.03638, 2019.
- [14] CAMPOS C, ELVIRA R, RODRIGUEZ 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.
- [15] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4104-4113.
- [16] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: real-time monocular visual SLAM with points and lines [C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 4503-4508.
- [17] FU Q, WANG J, YU H, et al. PL-VINS: real-time monocular visual-inertial SLAM with point and line features [J]. arXiv Preprint, arXiv:2009.07462, 2020.
- [18] 董蕊芳, 王宇鹏, 阚江明. 基于改进 ORB-SLAM2 的机器人视觉导航方法 [J]. 农业机械学报, 2022, 53(10): 306-317.
DONG Ruifang, WANG Yupeng, KAN Jiangming. Visual navigation method for robot based on improved ORB-SLAM2 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(10): 306-317. (in Chinese)
- [19] VON GIOI R G, JAKUBOWICZ J, MOREL J M, et al. LSD: a line segment detector [J]. Image Processing on Line, 2012, 2: 35-55.
- [20] TRIGGS B, MCLAUCHLAN P F, HARTLEY R I, et al. Bundle adjustment-a modern synthesis [C]//Vision Algorithms: Theory and Practice; International Workshop on Vision Algorithms Corfu, 2000: 298-372.
- [21] FORSTER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial odometry [J]. IEEE Transactions on Robotics, 2017, 33(1): 1-21.
- [22] SHI J. Good features to track [C]//1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 1994: 593-600.
- [23] HARRIS C G, STEPHENS M. A combined corner and edge detector [C]//Alvey Vision Conference, 1988: 147-151.
- [24] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of the ACM, 1981, 24(6): 381-395.
- [25] JOSWIG M, THEOBALD T, JOSWIG M, et al. Plücker coordinates and lines in space [J/OL]. Polyhedral and Algebraic Methods in Computational Geometry, 2013: 193-207. https://link.springer.com/chapter/10.1007/978-1-4471-4817-3_12.
- [26] VON STUMBERG L, CREMERS D. Dm-vio: delayed marginalization visual-inertial odometry [J]. IEEE Robotics and Automation Letters, 2022, 7(2): 1408-1415.