

DOI:10.19651/j.cnki.emt.2417368

基于紧耦合 IMU 的辐射场视觉惯性 SLAM 算法^{*}

王风华^{1,2} 徐志城^{1,2} 赵冷锐^{1,2}

(1. 中国石油大学(华东)青岛软件学院计算机科学与技术学院 青岛 266580; 2. 山东省智能油气工业软件重点实验室 青岛 266580)

摘要: 针对辐射场视觉 SLAM 算法的位姿估计误差大以及与 IMU 融合过程中的鲁棒性问题, 本文提出了一种基于紧耦合 IMU 的辐射场视觉惯性 SLAM 算法。该算法使用改进的预积分模块实现紧耦合框架, 改进的初始化策略来处理鲁棒性问题, 结合辐射场损失联合优化位姿和零偏。将本文的改进算法分别应用于 NICE-SLAM 和 MonoGS 的定位模块, 并在 IMU-RGBD 数据集 OpenLORIS 上进行了实验测试, 紧耦合模块对定位精度的提升最大分别为 34.3% 和 14.8%。与 MM3DGS 相比, 本算法具备更高的鲁棒性, 可有效提高定位精度, 对于提升辐射场 SLAM 性能具有较好的泛化能力。

关键词: 视觉惯性 SLAM; 神经辐射场; 三维高斯飞溅; 紧耦合

中图分类号: TN98 **文献标识码:** A **国家标准学科分类代码:** 510.99

Radiance field visual inertial SLAM algorithm based on tightly coupled IMU

Wang Fenghua^{1,2} Xu Zhicheng^{1,2} Zhao Lengrui^{1,2}(1. Qingdao Institute of Software, College of Computer Science and Technology, China University of Petrol, Qingdao 266580, China;
2. Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, Qingdao 266580, China)

Abstract: In order to solve the large pose estimation error of radiant field Visual SLAM algorithm and poor robustness in the process of fusion with inertial measurement unit, this paper proposes a radiance field visual inertial SLAM algorithm based on tightly coupled IMU. The algorithm uses an improved pre-integration module to implement a tightly coupled framework, the improved initialization strategy to deal with the robustness problem, combined with radiation field loss to optimize pose and bias. The proposed algorithm is applied to the positioning modules of NICE-SLAM and MonoGS, and is experimentally tested on the IMU-RGBD dataset OpenLORIS, and the tight-coupled module can improve the positioning accuracy by 34.3% and 14.8% respectively. Compared with MM3DGS, the proposed algorithm has higher robustness, which can effectively improve the positioning accuracy and has a good generalization ability to improve the SLAM performance of the radiance field.

Keywords: visual inertial SLAM; neural radiance field; 3D Gaussian splatting; tight coupling

0 引言

同时定位与建图 (simultaneous localization and mapping, SLAM) 是计算机视觉领域的一个重要研究课题, 它需要在没有先验地图信息的情况下实时地探索环境并估计相机的姿态。

传统 SLAM 采用离散的表面表示 (例如, 点云、体素网格和八叉树)^[1-3] 导致地图重建过程中不可避免产生失真。在要求高保真 3D 重建的应用场景中, 以隐式神经辐射场 (neural radiance field, NeRF)^[4] 和 3D 高斯飞溅 (3D

Gaussian splatting, 3DGS)^[5] 为典型代表的辐射场算法可以建立更为精细的地图。与传统的地图表示相比, 辐射场支持连续曲面建模, 减少内存需求, 提供了高质量的地图重建方案。Sucar 等^[6] 提出的 iMAP 通过减少 NeRF 的参数量和使用 RGB-D 数据, 首次实现了基于多层感知机 (MLP) 的地图构建。Zhu 等^[7] 提出的 NICE-SLAM 采用了多层级的网络结构的显式辐射场, 分层次优化场景表示的参数。Matsuki 等^[8] 提出的 Gaussian Splatting-SLAM (MonoGS) 基于 3DGS 建图, 通过链式求导法则建立位姿相对高斯特征的解析雅可比矩阵, 实现了基于 3DGS 的

收稿日期: 2024-11-14

^{*} 基金项目: 国家自然科学基金(61305008)、中央高校基本科研业务费专项资金项目(19CX02030A)资助

SLAM 系统。

尽管基于辐射场的 SLAM 算法能够生成高质量的地图,但与传统的 SLAM 方法相比,它在定位精度存在不足。为解决此问题, Yan 等^[9]提出 GS-SLAM 从粗到细地优化姿态。 Ha 等^[10]提出 GS-ICP SLAM 算法将广义最近点算法(G-ICP)与 3DGS 相结合,使得系统能直接基于 3D 信息优化相机姿态。然而,与传统 SLAM 技术相比,定位精度仍有差距^[3]。

近期研究表明基于辐射场理论的 SLAM 系统定位精度的提升需要更多信息来收敛位姿。目前已有不少研究工作探索了惯性测量单元(inertial measurement unit, IMU)在辐射场的 SLAM 领域的应用。Lisus 等^[11]提出将运动数据应用在基于 NeRF 的 SLAM 中,但是该算法并未涉及真实的 IMU 数据,实用性不足。Sun 等^[12]提出的 MM3DGS 的多模态辐射场 SLAM 算法,将 IMU 数据与基于 3DGS 的 SLAM 算法定位模块融合,但没有对重力方向解耦,缺乏对零偏的优化,导致算法的鲁棒性受限于 IMU 数据的质量和重力方向先验。

针对 IMU 数据与辐射场 SLAM 算法融合的鲁棒性问题,本研究借鉴了传统 IMU 数据融合算法。传统方法在对 IMU 数据进行融合的过程中,为了提高 IMU 数据的置信度,必须对数据进行降噪处理。在基于运动惯性数据的多模态 SLAM 系统中,通常使用两种策略来降低噪声的影响:基于滤波器的松耦合方法^[13-14]以及基于优化的紧耦合

方法^[15-17]。然而,基于滤波器的方法受到线性化误差影响。基于优化的方法采用高斯牛顿法等非线性优化方法更准确的估计状态,但是对计算资源需求增加。Forster 等^[18]提出了基于流形的预积分算法,将 IMU 和重力解耦,并求出预积分项的雅可比矩阵,有效提高了运算效率。

本研究结合基于优化的紧耦合 IMU 降噪策略,提出基于紧耦合的 IMU 的辐射场视觉惯性 SLAM 算法,相较于其他的多模态辐射场 SLAM 算法^[11-12],该方法利用辐射场损失为 IMU 零偏优化提供视觉约束,并采用改进的初始化方法修正重力方向和初速度。

1 关键技术研究

1.1 概述

辐射场 SLAM 算法的基本框架分为建图模块和定位模块,通过场景生成器生成对位姿和场景参数可微的估计图像,使用梯度下降法最小化辐射场损失,在建图模块中联合优化场景参数和关键帧位姿,在定位模块中固定场景参数仅优化位姿。

针对辐射场 SLAM 算法的定位精度问题,本文引入 IMU 输入流,并针对与 IMU 融合过程中的鲁棒性问题,以辐射场 SLAM 的定位模块为主体框架,实现了基于辐射场损失的 IMU 紧耦合框架,该框架可以通用于辐射场 SLAM 算法的定位模块。算法具体流程如图 1 所示,图 1 中红色虚线框内是本文提出的紧耦合算法的框架。

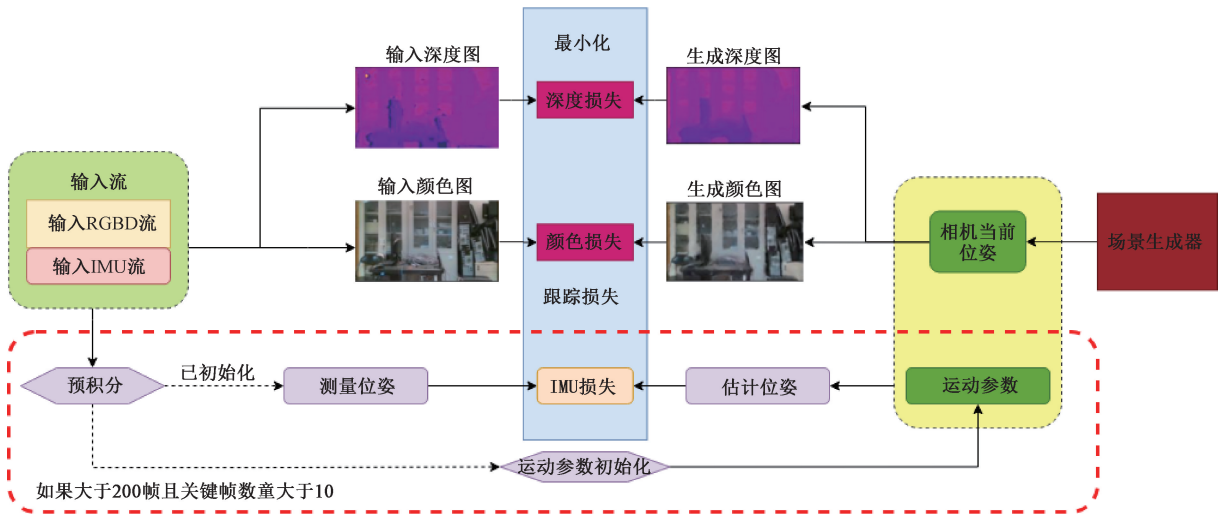


图 1 改进的紧耦合定位模块

Fig. 1 Improved tightly coupled positioning module

图 1 中定位模块的跟踪损失(即辐射场损失)表示为:

$$loss_c = L_1(\sum c_{est} - c_{gt}) + L_1(\sum d_{est} - d_{gt}) \quad (1)$$

其中, c_{est} 和 d_{est} 是估计的颜色图和深度图, c_{gt} 和 d_{gt} 是真实颜色图和深度图, L_1 表示 L1 范数。基于 NeRF 的 SLAM 算法的场景参数由 MLP 解码器表示,而基于 3DGS 的 SLAM 的场景参数由 3D 高斯表示。其中 NeRF 和

3DGS 的建图流程分别如下:

1) 在基于 NeRF 的 SLAM 中,使用 MLP 解码器和特征网格作为场景表示来渲染图像,并且不同的特征网格和解码器可以分别表示颜色和占用值。

$$c_p = d_c(\mathbf{P}, \varphi_c(\mathbf{P})) \quad (2)$$

$$O_p = d_o(\mathbf{P}, \varphi_o(\mathbf{P})) \quad (3)$$

其中, $d(\cdot)$ 指示解码器, \mathbf{P} 表示采样点的位置和射线方向, $\varphi(\cdot)$ 表示特征网格。对采样点的占有率积分, 得到射线终点的占用率:

$$w_{p_i} = O_{p_i} \prod_{k=1}^i (1 - O_{p_k}) \quad (4)$$

对射线进行采样后, 对射线上的颜色和深度积分, 得到估计的颜色和深度值 c_{est} 和 d_{est} :

$$c_{est} = \sum w_i c_i \quad (5)$$

$$d_{est} = \sum w_i d_i \quad (6)$$

2) 在基于 3DGS 的 SLAM 算法中, 场景表示使用 3D 高斯特征作为场景表示:

$$G = \{G_i: (\mathbf{X}_i, \boldsymbol{\Sigma}_i, \mathbf{O}_i, c_i) \mid i = 1, \dots, N\} \quad (7)$$

通过显式的 3D 高斯表达式直接表示颜色和不透明度。其中 $\mathbf{X}_i \in \mathbf{R}^3$ 表示位置, $\boldsymbol{\Sigma}_i \in \mathbf{R}^{3 \times 3}$ 表示 3D 高斯协方差, $\mathbf{O}_i \in \mathbf{R}$ 表示不透明度, $c_i \in \mathbf{R}^{12}$ 表示球谐函数系数, 用于表示颜色。为了降低 3DGS 的参数数量, 协方差参数表示为:

$$\boldsymbol{\Sigma} = \mathbf{R} \mathbf{S} \mathbf{S}^T \mathbf{R}^T \quad (8)$$

其中, $\mathbf{S} \in \mathbf{R}^3$ 表示缩放变换, $\mathbf{R} \in \mathbf{R}^{3 \times 3}$ 是旋转矩阵, 使用四元数表达。给定视图变换 \mathbf{W} 和协方差矩阵 $\boldsymbol{\Sigma}$, 生成 3D 高斯的 2D 投影平面:

$$\boldsymbol{\Sigma}' = \mathbf{J} \mathbf{W}^{-1} \boldsymbol{\Sigma} (\mathbf{w}^{-1})^T \mathbf{J}^T \quad (9)$$

其中, \mathbf{J} 是射影函数的仿射逼近的雅可比行列式。在投影之后, 按深度对高斯特征排序, 并以与式(5)、(6)类似的积分方式渲染视角。因此辐射场 SLAM 的定位模块存在共通性。

在数据融合阶段中, 本研究采用紧耦合的策略, 提出了一种通用 IMU 损失函数, 以便与定位模块的梯度下降算法的耦合。为了避免重力和速度先验的影响, 本研究提出改进的初始化模块以初修正重力方向和速度。

1.2 IMU 预积分阶段

由于 IMU 测量的处理涉及姿态积分, 因此系统对白噪声以及加速度计和陀螺仪中固有的零偏噪声, 这可能导致姿态和速度误差随时间累积。此外, 加速度计的输出包含重力矢量的分量, 这意味着姿态估计中的误差会直接影响加速度计测量中的重力分量, 进一步加剧整个系统的累积误差, 并可能导致积分结果的不稳定分散。鉴于此, 本研究采用基于流形的预积分算法^[18]来融合 IMU 数据和辐射场损失。 T_i 时刻下相机运动参数被定义为 $\mathbf{C}_i = [\mathbf{R}_i, \mathbf{P}_i, \mathbf{V}_i, \mathbf{b}a_i, \mathbf{b}g_i]$, 其中 $\mathbf{R}_i \in \mathbf{R}^{3 \times 3}$ 表示世界坐标系下的相机方向, $\mathbf{P}_i \in \mathbf{R}^3$ 表示世界坐标系下的摄影机位置, $\mathbf{V}_i \in \mathbf{R}^3$ 表示世界坐标系下的摄影机速度, $\mathbf{R}_i, \mathbf{P}_i, \mathbf{V}_i$ 是摄像机的运动参数项。 $\mathbf{b}a_i, \mathbf{b}g_i \in \mathbf{R}^3$ 是相机的噪声参数项, 分别表示加速度计测量的加速度偏差和陀螺仪测量的角速度零偏。

在预积分过程中, 不仅计算了帧间速度、位置和旋转角的预积分项, 还得到了预积分项关于加速度偏差 $\mathbf{b}a$ 和

角速度偏差 $\mathbf{b}g$ 的雅可比矩阵。预积分项的公式为:

$$\Delta \mathbf{R}_{ij} = \prod_{k=i}^j \text{Exp}((\boldsymbol{\omega}_k - \mathbf{b}g) \Delta t) \quad (10)$$

$$\Delta \mathbf{v}_{ij} = \sum_{k=i}^j \Delta \mathbf{R}_{ik} (\mathbf{a}_k - \mathbf{b}a) \Delta t \quad (11)$$

$$\Delta \mathbf{P}_{ij} = \sum_{k=i}^j \Delta \mathbf{V}_{ik} \Delta t + 0.5 \Delta \mathbf{R}_{ik} (\mathbf{a}_k - \mathbf{b}a) \Delta t^2 \quad (12)$$

其中, Δt 表示 IMU 帧间时间间隔。 $\boldsymbol{\omega}_k$ 表示陀螺仪在第 k 个 IMU 帧中的角速度值, 以及 \mathbf{a}_k 表示第 k 个 IMU 帧中的加速度测量值。假定图像帧之间的零偏恒定。

为了减少预积分更新次数, 预积分阶段迭代计算并保存预积分项对零偏的雅可比矩阵。

当外部噪声被引入到积分过程中时, 预积分项的噪声协方差矩阵(以下称为协方差矩阵)成为预积分结果的位置度的关键度量。初始协方差矩阵为 $\mathbf{0}^{9 \times 9}$ 第 $j-1$ 帧的协方差矩阵定义为 \mathbf{C}_{j-1} 。当接收到第 j 个 IMU 帧时, 递推公式为:

$$\mathbf{A}_{j-1} = \begin{bmatrix} \Delta \mathbf{R}_{j,j-1} & \mathbf{0} & \mathbf{0} \\ -\Delta \mathbf{R}_{j,j-1} (\mathbf{a}_j - \mathbf{b}a)^\Delta \Delta t & \mathbf{I} & \mathbf{0} \\ -0.5 \Delta \mathbf{R}_{j,j-1} (\mathbf{a}_j - \mathbf{b}a)^\Delta \Delta t & -\mathbf{I} \Delta t & \mathbf{I} \end{bmatrix} \quad (13)$$

$$\mathbf{B}_{j-1} = \begin{bmatrix} \mathbf{J}_r^{j-1} \Delta t & \mathbf{0} \\ 0 & \Delta \mathbf{R}_{j,j-1} \Delta t \\ 0 & 0.5 \Delta \mathbf{R}_{j,j-1} \Delta t^2 \end{bmatrix} \quad (14)$$

$$\mathbf{C}_j = \mathbf{A}_{j-1} \mathbf{C}_{j-1} \mathbf{A}_{j-1}^T + \mathbf{B}_{j-1} \boldsymbol{\Sigma}_j \mathbf{B}_{j-1}^T \quad (15)$$

其中, $(\cdot)^\Delta$ 表示 \mathbf{R}^3 到反对称矩阵的映射, $\Delta \mathbf{R}_{j,j-1}$ 是帧间的旋转增量, \mathbf{J}_r^{j-1} 是旋转增量在李群上的右雅可比矩阵。 $\boldsymbol{\Sigma}_j \in \mathbf{R}^{6 \times 6}$ 是加速度和角速度的噪声协方差矩阵。

对于噪声协方差矩阵的逆矩阵 \mathbf{C}^{-1} , 预积分算法^[18]直接将其作为残差损失的权重。值得注意的是, 预积分算法^[18]是基于高斯-牛顿法的, 但是辐射场 SLAM 的优化方法是基于梯度下降法的, 因此本文在将协方差矩阵应用到权重表示的过程中, 对协方差矩阵作对角化处理, 仅保留对角线元素, 忽略非对角线元素以避免优化项其他维度分量对梯度的额外影响。实验部分验证了本文的改进是更接近仿真结果的。

1.3 运动参数初始化

运动参数包括重力方向和初始速度, 对运动约束损失的收敛结果和姿态估计有重要的影响。本文将运动参数的初始化问题视为线性最小二乘优化问题, 改进了 Vins-Mono^[15]的初始化方法。与 VINS-Mono 不同的是, 本文利用 RGBD 图像数据, 因此不需要优化比例因子, 并且由于零偏的初始化需要对之前的所有帧重新预积分, 为了减少运算消耗, 因此本文不对零偏进行初始化。在初始化完成之前, IMU 数据仅用于预积分阶段, 而不参与位姿优化阶段。只有当关键帧的累积超过 10 帧并且图像帧超过 200 帧时, 才使用已估计的姿态来初始化每个图像帧的速度和

重力方向。

初始化分为两个阶段:1)速度和重力加速度初始化;2)重力方向细化。

第 1 阶段:待优化的变量表示为 $\mathbf{x} = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{g}]$, 其中 \mathbf{v}_i 表示第 i 帧的速度, \mathbf{g} 表示重力。优化目标是速度和姿态的估计值和测量值的残差, 由于未初始化零偏, 因此忽略零偏噪声的影响。将待优化变量分离, 转换为矩阵方程的形式:

$$[-\mathbf{I} \quad \mathbf{I} \quad -\mathbf{I}\Delta t] \times \begin{bmatrix} \mathbf{v}_i \\ \mathbf{v}_{i+1} \\ \mathbf{g} \end{bmatrix} = \mathbf{R}_i \Delta \mathbf{v} \quad (16)$$

$$[-\mathbf{I}\Delta t \quad \mathbf{0} \quad -0.5\mathbf{I}\Delta t^2] \times \begin{bmatrix} \mathbf{v}_i \\ \mathbf{v}_{i+1} \\ \mathbf{g} \end{bmatrix} = \mathbf{R}_i \Delta \mathbf{P} + \mathbf{P}_i - \mathbf{P}_{i+1} \quad (17)$$

组合得到超定方程:

$$\begin{bmatrix} -\mathbf{I} & \mathbf{I} & -\mathbf{I}\Delta t \\ -\mathbf{I}\Delta t & \mathbf{0} & -0.5\mathbf{I}\Delta t^2 \end{bmatrix} \times \begin{bmatrix} \mathbf{v}_i \\ \mathbf{v}_{i+1} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_i \Delta \mathbf{v} \\ \mathbf{R}_i \Delta \mathbf{P} + \mathbf{P}_i - \mathbf{P}_{i+1} \end{bmatrix} \quad (18)$$

矩阵方程形如 $\mathbf{Ax} = \mathbf{B}$, 优化策略等价于求解该线性方程的最小二乘解, 其等价于求解:

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{B} \quad (19)$$

第 2 阶段:重力向量的模长固定为重力大小, 将一对与重力方向正交的单位向量作为待优化量, 则重力加速度可表示为:

$$\mathbf{g} = g\bar{\mathbf{g}} + \mathbf{w}_1 \mathbf{b}_1^{3 \times 1} + \mathbf{w}_2 \mathbf{b}_2^{3 \times 1} = g\bar{\mathbf{g}} + \mathbf{b}^{3 \times 2} \mathbf{w}^{2 \times 1} \quad (20)$$

其中, $\mathbf{w}^{2 \times 1}$ 是待优化量, \mathbf{g} 表示重力模长, $\bar{\mathbf{g}}$ 表示重力的单位方向, $\mathbf{b}^{3 \times 2}$ 表示两组单位正交向量。在每一轮迭代更新之后, 更新的重力方向被归一化以更新 $\bar{\mathbf{g}}$ 。将式(20)带入式(18)得到矩阵形式:

$$\begin{bmatrix} -\mathbf{I} & \mathbf{I} & -\mathbf{b}\Delta t \\ -\mathbf{I}\Delta t & \mathbf{0} & -0.5\mathbf{b}\Delta t^2 \end{bmatrix} \times \begin{bmatrix} \mathbf{v}_i \\ \mathbf{v}_{i+1} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_i \Delta \mathbf{v} + \|\mathbf{g}\| \bar{\mathbf{g}} \Delta t \\ \mathbf{R}_i \Delta \mathbf{P} + \mathbf{P}_i - \mathbf{P}_{i+1} + 0.5 \|\mathbf{g}\| \bar{\mathbf{g}} \Delta t^2 \end{bmatrix} \quad (21)$$

用与上述相同的方法求解该线性方程的最小二乘解。

1.4 IMU 与辐射场定位模块的紧耦合算法

综上所述, 在预积分阶段计算预积分项、雅可比矩阵和协方差矩阵, 并且仅在下次预积分阶段时被更新。

在位姿优化阶段中, 将预测值(由前一帧的运动参数项提供)与测量值(预积分项)之间的残差作为运动约束损失, 通过最小化该损失来优化当前帧和前一帧的运动参数。

残差公式为:

$$\mathbf{r}_R = \log \left[\Delta \mathbf{R}_{ij} \exp \left(\frac{\partial \Delta \mathbf{R}_{ij}}{\partial \mathbf{b} \mathbf{g}} \Delta \mathbf{b} \mathbf{g} \right) \mathbf{R}_i^T \mathbf{R}_j \right] \quad (22)$$

$$\mathbf{r}_v = \mathbf{R}_i^T (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g} \Delta t) - \left[\Delta \mathbf{v}_{ij} + \frac{\partial \Delta \mathbf{V}_{ij}}{\partial \mathbf{b} \mathbf{g}} \Delta \mathbf{b} \mathbf{g} + \frac{\partial \Delta \mathbf{V}_{ij}}{\partial \mathbf{b} \mathbf{a}} \Delta \mathbf{b} \right] \quad (23)$$

$$\mathbf{r}_P = \mathbf{R}_i^T (\mathbf{P}_j - \mathbf{P}_i - \mathbf{v}_i \Delta t - 0.5 \mathbf{g} \Delta t^2) - \left[\Delta \mathbf{P}_{ij} + \frac{\partial \Delta \mathbf{P}_{ij}}{\partial \mathbf{b} \mathbf{g}} \Delta \mathbf{b} \mathbf{g} + \frac{\partial \Delta \mathbf{P}_{ij}}{\partial \mathbf{b} \mathbf{a}} \Delta \mathbf{b} \mathbf{a} \right] \quad (24)$$

运动约束损失如下:

$$\text{loss}_{IMU} = [\mathbf{r}_R, \mathbf{r}_v, \mathbf{r}_P]^T \boldsymbol{\Sigma}^{-1} [\mathbf{r}_R, \mathbf{r}_v, \mathbf{r}_P] \quad (25)$$

其中, $\boldsymbol{\Sigma}^{-1}$ 表示协方差矩阵对角化处理后的逆矩阵。最后, 在定位模块使用辐射场损失与运动约束损失联合优化位姿和零偏:

$$\text{loss}_{track} = \text{loss}_c + \lambda \text{loss}_{IMU} \quad (26)$$

其中, λ 是运动约束损失的权重参数。

为了加速损失的收敛, 通过基于前一帧的运动参数、预积分项和重力对当前帧的速度和姿态进行恒速假设作为位姿和速度的先验:

$$\mathbf{v}_j = \mathbf{v}_i + \mathbf{g} \Delta t + \mathbf{R}_i \Delta \mathbf{v} \quad (27)$$

$$\mathbf{P}_j = \mathbf{P}_i + 0.5 \mathbf{g} \Delta t^2 + \mathbf{v}_i \Delta t + \mathbf{R}_i \Delta \mathbf{p} \quad (28)$$

每次更新零偏时, 为了避免重新预积分, 通过预积分增量对零偏的雅可比矩阵, 对预积分项进行线性更新。

2 实验结果与分析

本文提出的基于紧耦合的 IMU 的辐射场视觉惯性 SLAM 算法可以分别应用在基于 NeRF 的和基于 3DGS 的视觉 SLAM 系统的定位模块, 实现定位性能的提升。为验证算法的有效性, 本文选取了 NeRF SLAM 和 3DGS SLAM 中具有代表性的两个算法 NICE-SLAM 和 MonoGS 作为基线。实验部分分别将本文的紧耦合模块与基线 NICE-SLAM 算法和 MonoGS 算法的定位模块结合并比较改进前后的位姿估计精度, 验证改进算法在辐射场 SLAM 系统中的泛化能力, 与同类的多模态辐射场 SLAM 算法 MM3DGS 对比, 验证改进部分的有效性。最后, 验证了本文针对预积分算法^[18]提出的改进方法的有效性。

2.1 数据集和评估标准

1) 数据集简介

由于 NICE-SLAM 和 Gaussian Splatting SLAM 使用的数据集不包含运动信息, 因此本文在开源数据集 OpenLORIS 上进行了测试。OpenLoris 数据集^[19]是 IROS2019 组织的终身机器人视觉竞赛的官方数据集。安装有 Intel Real Sense D435i 和 T265 摄像头的机器人用于数据收集。D435i 摄像头提供对齐的彩色和深度图像以及 IMU。图像帧的帧频为 30 Hz, IMU 陀螺仪的帧频为 400 Hz, 加速度计的帧频为 250 Hz。机器人的地面真实轨迹是从 OptiTrack 运动捕捉系统获得的。数据集中摄像头和 IMU 的坐标系以及数据集的示例如图 2、3 所示。

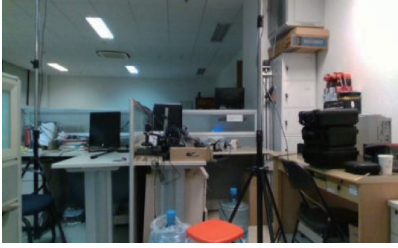


图 2 数据集的示例

Fig. 2 Example of the dataset

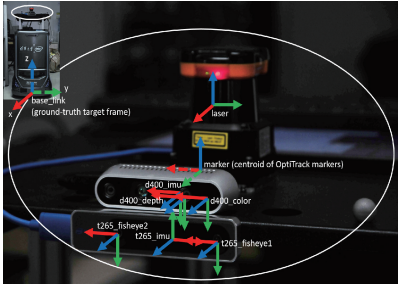


图 3 Openloris 使用的采集机器人

Fig. 3 The collection robot used by Openloris

它的主要特点包括提供高精度的 IMU 传感器数据和 RGBD 数据,适用于使用 RGBD 图像序列和紧密耦合的 IMU 测量来评估本文方法的定位精度。在本文中,对数据集的 office 序列进行了实验。

2) 评价指标

本文采用绝对轨迹误差(ATE)来评价位姿估计精度,计算对齐后的估计轨迹与地面真实轨迹之间的均方根误差(RMSE),ATE 的 RMSE 越小,SLAM 系统的轨迹越接近真实轨迹,定位性能越好。其中,ATE 的 RMSE 计算如下:

$$ATE_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \log(\mathbf{T}_{gt}^{-1} \mathbf{T}_{est})} \quad (29)$$

其中, \mathbf{T}_{gt} 是姿态真值, \mathbf{T}_{est} 为姿态估计值。

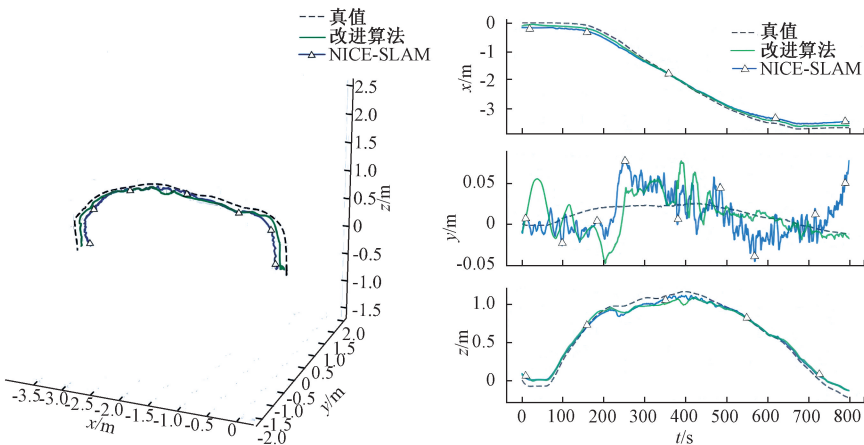


图 4 office1-1 序列上的轨迹和三轴分量

Fig. 4 Trajectory and triaxial components on the office1-1 sequence

2.2 算法性能实验结果与分析

1) 与基于 NeRF 的 SLAM 对比的有效性实验

实验使用的是 12 G 显存 GPU 的 RTX3080Ti。在基于 NeRF 的 SLAM 的算法比较中,以代表性算法 NICE-SLAM 为基线,对基线 NICE-SLAM 额外输入 IMU 数据流并在定位模块中结合本文的紧耦合模块以作为对比算法(NICE+IMU),验证本文提出的紧耦合模块在基于 NeRF 的 SLAM 中的有效性。

其中 NICE-SLAM 的位姿学习率采用原始学习率,将 IMU 的速度、加速度偏差和角速度偏差的学习率分别设置为 0.003、0.000 5 和 0.000 5,运动惯性损失的权重设置为 0.01。定位模块迭代 50 次。如表 1 所示,其中对比算法在定位精度上相比仅使用 RGBD 数据的 NICE-SLAM 均有大幅提高。

表 1 与 NICE-SLAM 绝对轨迹误差(ATE)对比
Table 1 Absolute trajectory error (ATE) compared with NICE-SLAM

序列	NICE-SLAM	NICE+IMU	提升比/%
	RMSE/m	RMSE/m	
office1-1	0.169	0.111	34.3
office1-2	0.231	0.210	9.1
office1-3	0.343	0.290	15.4
office1-4	0.365	0.319	12.6
office1-5	0.459	0.406	11.5
office1-6	0.153	0.137	10.4
office1-7	0.596	0.520	12.7

以 NICE-SLAM 和对比算法在 office1-1 序列的通过 evo 获得的轨迹对比为例,如图 4 所示。NICE-SLAM 的定位轨迹出现了大量的小幅度漂移现象,这在 y 轴上更加明显。通过融合 IMU 数据,这些问题得到了显著改善。此外,在 x 轴和 y 轴上,NICE-SLAM 与真实值的具有一定偏差,通过引入 IMU 数据有效地校正了该偏差。实验表明,本文所提出的改进算法对各轴上的偏移具有约束作用,有效避免了定位过程中的非物理性漂移现象。

2) 与基于 3DGS 的 SLAM 对比的有效性实验

在基于 3DGS 的 SLAM 算法比较中,以 MonoGS 为基线,在基线 MonoGS 中额外输入 IMU 数据流,并在定位模块中结合本文的紧耦合模块作为对比算法(MonoGS+IMU),定位模块迭代 100 次,将 IMU 的速度参数、加速度偏差和角速度偏差的学习率分别设置为 0.003、0.000 5 和 0.000 5,运动惯性损失的权重设置为 0.2。对比结果如表 2 所示,紧耦合模块使得 MonoGS 的定位精度在 office 序列上最高提高了 14.8%,平均提高了 5.6%。尽管对定位性能的提升效果不如基于 NeRF 的 SLAM 算法显著,依然有明显提升。

表 2 与 MonoGS 绝对轨迹误差 (ATE) 对比
Table 2 Absolute trajectory error (ATE) compared with MonoGS

序列	MonoGS RMSE/m	MonoGS+IMU RMSE/m	提升比/%
office1-1	0.081	0.069	14.8
office1-2	0.132	0.122	7.5
office1-3	0.175	0.161	8.0
office1-4	0.199	0.189	5.0
office1-5	0.259	0.246	5.1
office1-6	0.120	0.113	5.8
office1-6	0.334	0.315	5.6

对于基于 NeRF 的和基于 3DGS 的辐射场 SLAM 算

表 3 改进算法与 MM3DGS 的绝对轨迹误差 (ATE) 的 RMSE(m) 对比

Table 3 Comparison of RMSE (m) of absolute trajectory error (ATE) between the improved algorithm and the MM3DGS

对比算法	office1-1	office1-2	office1-3	office1-4	office1-5	office1-6	office1-7
MM3DGS 纯视觉	0.077 4	0.128 3	0.174 2	0.191 8	0.252 3	0.116 1	0.332 1
MM3DGS 多模态	0.077 6	0.130 5	0.174 6	0.193 7	0.253 0	0.118 2	0.332 5
MM3DGS 多模态+初始化	0.077 3	0.127 2	0.173 8	0.191 0	0.251 2	0.115 4	0.331 7
MonGS	0.0812	0.132 1	0.175 8	0.199 1	0.259 9	0.120 2	0.334 6
MonoGS+IMU	0.069 5	0.122 7	0.161 3	0.189 5	0.246 1	0.113 3	0.315 1

同样是基于 3DGS 的 SLAM 算法, MonoGS 定位精度相比 MM3DGS 纯视觉算法和 MM3DGS 多模态算法均有所不如。在 MonoGS 在额外使用 IMU 数据并结合本文改进的紧耦合模块后,定位精度超越了同样引入初始化模块作为重力补偿的 MM3DGS 多模态算法,这表明了在同样进行初始化的情况下,本文提出的紧耦合模块相比 MM3DGS 直接将 IMU 数据作为定位先验具备更优越的性能。

4) 实时性影响分析

为了分析 IMU 优化帧的加入对系统整体实时性的影响,本文对相关模块独立测试其运算速度。在实验中,根

据每轮 50 个梯度下降来执行迭代。IMU 模块的帧平均优化所需的时间为 34 ms。表 4 表明 IMU 数据集成所带来的额外计算需求不会严重影响系统的性能。

3) 基于 MM3DGS 对比的有效性实验

在于同类型算法比较中,以同样基于 3DGS 建图的多模态 SLAM 算法 MM3DGS 为基线,均使用 OpenLoris 的 office 的序列作为测试集。MM3DGS 作为多模态 IMU 融合算法并未采用紧耦合的方法,而是追求实时性能,直接将 IMU 数据作为定位先验,而且没有初始化重力和速度,依赖重力方向先验。这导致 MM3DGS 的鲁棒性较差,面对重力先验失真的场景,存在与 IMU 数据融合后反而导致定位精度下降的情况。

为了验证本研究相较 MM3DGS 在 IMU 融合算法上的鲁棒性提升,实验比较了 MM3DGS 未融合 IMU 的纯视觉算法(MM3DGS 纯视觉)、MM3DGS 融合 IMU 的多模态算法(MM3DGS 多模态)、基于本文 1.3 节初始化算法对 MM3DGS 多模态算法的改进算法(MM3DGS 多模态+初始化)三者的定位精度对比,以及与本文改进的紧耦合模块(MonoGS 和 MonoGS+IMU)的效果对比。

如表 3 所示,MM3DGS 多模态算法使用 IMU 数据作为定位先验后,定位性能反而变差。这是由于 MM3DGS 缺少初始化模块,算法的鲁棒性受重力方向先验影响。对比实验在 MM3DGS 多模态算法的 IMU 融合模块中引入运动参数初始化模块动态修正重力方向,实现了定位精度的提升。这验证了本研究中初始化模块在重力失真场景下的补偿作用。

据每轮 50 个梯度下降来执行迭代。IMU 模块的帧平均优化所需的时间为 34 ms。表 4 表明 IMU 数据集成所带来的额外计算需求不会严重影响系统的性能。

2.3 协方差改进方案实验结果与分析

为了验证本文针对预积分算法^[18]的协方差矩阵在权重应用方面所提出改进的有效性,在 OpenLoris 数据集的 officiel-1 序列上设计了消融实验。考虑到速度的估计对位姿估计的影响是线性的,并且速度在优化过程中是直接受协方差矩阵影响的,因此速度分量可以更直观地作为对比指标。实验通过对比改进前后速度与真实值的拟合情况,验证改进部分的有效性。考虑到实际数据中不存在速

表 4 运行时间对比
Table 4 Runtime comparison

对比算法	每帧定位平均 运行时间/ms	影响比/%
NICE-SLAM	127	—
NICE-SLAM+IMU	1 321	3.8
MonoGS	843	—
MonoGS+IMU	885	4.9

度真值,采用 1.3 节中的最小二乘法计算所有图像帧的仿

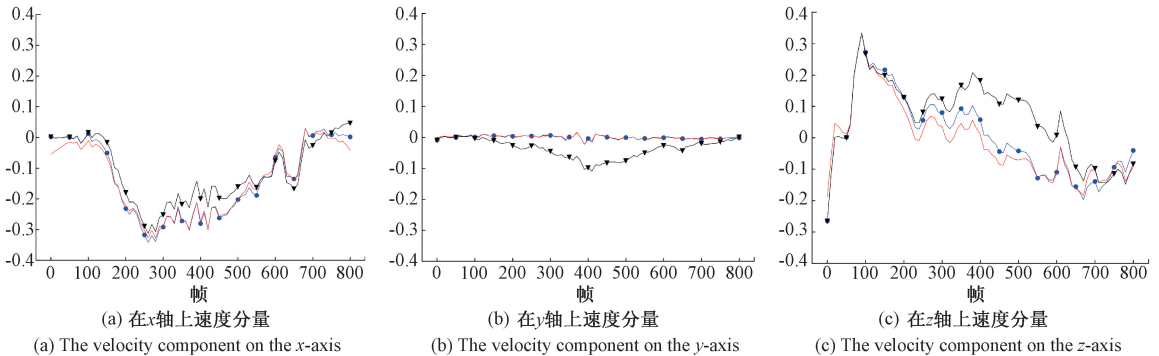


图 5 与仿真速度三轴分量的对比

Fig. 5 Comparison with the three-axis component of the simulated speed

3 结 论

针对辐射场 SLAM 算法的定位精度问题,本文提出了一种基于紧耦合的 IMU 的辐射场视觉惯性 SLAM 算法,该算法相比其他多模态辐射场 SLAM 是更加鲁棒的。该方法提出了适用于辐射场 SLAM 的运动惯性损失,改进基于最小二乘的初始化方法用以初始化重力方向和速度,有效提高了定位精度。最后,在 OpenLORIS 数据集上的实验结果表明,该融合算法可以广泛的应用于基于辐射场 SLAM 的定位模块。与 MM3DGS 的对比实验证明,本文提出的 IMU 数据融合算法具备更优越的鲁棒性和定位精度。

参考文献

- [1] 余联想,郑明魁,欧文君,等.多传感器融合的移动机器人室外激光 SLAM 算法优化与系统实现[J].电子测量与仪器学报,2023,37(2):48-55.
YU L X, ZHENG M K, OU W J, et al. Optimization and system implementation of outdoor laser SLAM for mobile robots based on multi-sensor fusion [J]. Journal of Electronic Measurement and Instrumentation, 2023,37(2):48-55.
- [2] 叶俊强,尤睿,于明鑫,等.基于稀疏直接法的特征增强视觉 SLAM 算法[J].仪器仪表学报,2023,44(6):205-212.
YE J Q, YOU R, YU M X, et al. Feature enhanced visual

真速度。

图 5 是速度的对比图,实线表示由最小二乘法拟合得到的各帧速度的仿真值,实心标记的实线表示改进后的速度分量,三角形标记的实线表示改进前直接使用协方差矩阵的逆矩阵作为权重所得的速度分量。与改进前相比,改进后的速度分量在各轴上明显更接近仿真值,在 x 轴和 y 轴上几乎拟合仿真值。而在 y 轴和 z 轴上,未经处理的协方差矩阵导致了优化项的其他维度的分量对速度分量梯度的额外影响,使得速度分量大幅偏离仿真值。实验表明了本文对噪声协方差在权重应用方面改进的有效性。

SLAM algorithm based on sparse direct method [J]. Chinese Journal of Scientific Instrument, 2023, 44(6): 205-212.

- [3] RAUL M A, JUAN D T. ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras[J]. IEEE Transactions on Robotics, 2018, 32(4):874-888.
- [4] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis[C]. European Conference on Computer Vision(ECCV), 2020.
- [5] KERBL B, KOPANAS G, LEIMKÜHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics(TOG), 2023,42(4):1-14.
- [6] SUCAR E, LIU S, ORTIZ J, et al. IMAP: Implicit mapping and positioning in real-time[C]. IEEE/CVF International Conference on Computer Vision(ICCV), 2021:6229-6238.
- [7] ZHU Z, PENG S, LARSSON V, et al. NICE-SLAM: Neural implicit scalable encoding for SLAM[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:12786-12796.
- [8] MATSUKI H, MURAI R, KELLY P H, et al. Gaussian splatting SLAM[C]. IEEE/CVF Conference

- on Computer Vision and Pattern Recognition, 2024: 18039-18048.
- [9] YAN C, QU D, XU D, et al. GS-SLAM: Dense visual slam with 3D Gaussian splatting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024.
- [10] HA S, YEON J, YU H, et al. RGBD GS-ICP SLAM[C]. European Conference on Computer Vision. Springer, Cham, 2025.
- [11] LISUS D, HOLMES C, WASLANDER S, et al. Towards open world nerf-based SLAM[C]. 2023 20th Conference on Robots and Vision(CRV), IEEE, 2023.
- [12] SUN L C, BHATT N P, LIU J, et al. MM3DGS SLAM: Multi-modal 3D gaussian splatting for SLAM using vision, depth, and inertial Measurements[C]. ArXiv preprint arXiv:2404.00923, 2024.
- [13] PAUL M K, WU K, HESCH J A, et al. A comparative analysis of tightly-coupled monocular, binocular, and stereo VINS [C]. 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017:165-172.
- [14] HUANG G, ECKENHOFF K, LEONARD J, et al. Optimal-state-constraint EKF for visual-inertial navigation[C]. Robotics Research, 2018.
- [15] QIN T, LI P, SHEN S, et al. VINS-Mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE transactions on robotics, 2018, 34(4):1004-1020.
- [16] 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.
- [17] 邓鹏, 罗静. 复杂环境下机器人多传感器融合定位方法[J]. 电子测量与仪器学报, 2023, 37(12):48-57.
- DENG P, LUO J. Robot multi-sensor fusion positioning method in complex environment[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12):48-57.
- [18] FORSTER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics, 2016, 33(1):1-21.
- [19] SHE Q, FENG F, HAO X, et al. OpenLORIS-object: A robotic vision dataset and benchmark for lifelong deep learning [C]. 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020:4767-4773.

作者简介

王风华(通信作者), 博士, 讲师, 主要研究方向为计算机视觉、边缘 AI、机器人定位技术。

E-mail: fenghuawang@upc.edu.cn

徐志城, 硕士研究生, 主要研究方向为视觉 SLAM 和 3D 重建。

E-mail: z22070096@s.upc.edu.cn

赵冷锐, 硕士研究生, 主要研究方向为计算机视觉、动态 SLAM。

E-mail: 310218706@qq.com