

DOI:10.11918/202407032

室内环境下融合点线特征的双目 VI-SLAM 方法

郭旭¹, 袁杰², 谢霖伟¹, 鲍慧敏¹, 李世钰¹

(1. 新疆大学 电气工程学院, 乌鲁木齐 830017; 2. 新疆大学 智能科学与技术学院, 乌鲁木齐 830017)

摘要: 为解决室内环境下弱纹理场景中关键点特征稀疏、结构化场景中结构化特征应用不充分以及相机快速移动时关键帧容易跟踪失败的问题,提出了一种基于点线特征融合的双目视觉惯性 SLAM 方法。首先,基于 EDlines 线段提取方法,结合高斯图像金字塔实现多尺度线段的提取,以增强线段匹配的尺度不变性。同时,对不同尺度下的线段端点的不确定性进行建模,并结合平铺技术对线段的二进制描述符进行分块处理,从而加速线段匹配并提高线特征匹配的鲁棒性与效率。其次,优化惯性传感器的预积分模型,融合双目视觉的点特征重投影误差、线特征重投影误差以及惯性传感器的预积分约束,采用滑动窗口的非线性优化方法进行联合优化,以提高系统位姿估计精度。最后,文中在包含弱纹理、结构化以及相机快速移动等复杂环境的 EuRoC 数据集上进行实验。结果表明,VI-SLAM 方法在 EuRoC 数据集上的相机轨迹均方根误差为 0.031 m,平均误差为 0.027 m,拥有更强的鲁棒性和更高的定位精度,尤其在弱纹理和相机快速运动场景中,定位精度优势显著。

关键词: 同步定位与建图(SLAM);视觉惯性;点线特征;双目相机;多尺度;非线性优化

中图分类号: TP242

文献标志码: A

文章编号: 0367-6234(2025)08-0069-10

Stereo VI-SLAM method with fused point and line features in indoor environments

GUO Xu¹, YUAN Jie², XIE Linwei¹, BAO Huimin¹, LI Shiyu¹

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China;

2. School of Intelligence Science and Technology, Xinjiang University, Urumqi 830017, China)

Abstract: To address the issues of sparse keypoint features in weakly textured indoor environments, insufficient utilization of structured features in structured scenes, and keyframe tracking failures during rapid camera movements, a stereo visual-inertial SLAM method based on the fusion of point and line features is proposed. First, the EDlines line segment extraction method, combined with a Gaussian image pyramid, is employed to achieve multi-scale line segment extraction, enhancing the scale invariance of line segment matching. Simultaneously, the uncertainty of line segment endpoints at different scales is modeled, and binary descriptors of line segments are partitioned using tiling technology to accelerate line segment matching, thereby improving the robustness and efficiency of line feature matching. Second, the pre-integration model of the inertial sensor is optimized, and a sliding window nonlinear optimization is performed by fusing the point feature reprojection error from stereo vision, the line feature reprojection error, and the pre-integration constraints of the inertial sensor, thereby improving the system's pose estimation accuracy. Finally, extensive experiments are conducted on the EuRoC dataset which includes complex environments such as low-texture, structured scenes, and rapid camera movements. The experimental results demonstrate that the proposed method achieves a root mean square error of 0.031 m and an average error of 0.027 m on the EuRoC dataset, exhibiting stronger robustness and higher localization accuracy, especially in low-texture and rapid camera movement scenarios where the accuracy advantage is particularly significant.

Keywords: simultaneous localization and mapping (SLAM); visual-inertial; point and line features; stereo camera; multi-scale; nonlinear optimization

同步定位与建图(simultaneous localization and mapping, SLAM)是一种用于估计机器人状态(如位置、方向、速度等)并同时构建环境模型的技术^[1]。

SLAM 广泛应用于服务机器人、自动驾驶、增强现实(augmented reality, AR)和虚拟现实(virtual reality, VR)等领域,是实现机器人自主导航的关键技术^[2]。

收稿日期: 2024-07-10; 录用日期: 2024-08-30 网络首发日期: 2025-06-25

网络首发地址: <https://link.cnki.net/urlid/23.1235.t.20250624.1418.006>

基金项目: 国家自然科学基金(62263031); 新疆维吾尔自治区自然科学基金(2022D01C53)

作者简介: 郭旭(2002-),男,硕士研究生; 袁杰(1975-),男,教授,博士生导师

通信作者: 袁杰, yuanjie@xju.edu.cn

根据传感器类型的不同,SLAM 可分为视觉 SLAM、视觉惯性 SLAM 和激光雷达 SLAM^[3]。其中,基于相机的视觉 SLAM^[4]因其硬件成本低、信息丰富,在机器人姿态估计中得到了广泛使用。惯性测量单元(inertial measurement unit, IMU)通过提供加速度和角速度信息,与视觉传感器结合,进一步提升环境感知能力和系统的鲁棒性^[5]。

视觉 SLAM 方法通常分为特征法和直接法^[6]。直接法如 Gao 等^[7]利用原始图像像素,假设亮度恒定,最小化连续帧之间的光度误差进行位姿估计。基于特征的方法如 Mur-artal 等^[8]、Campos 等^[9]通过提取图像关键点,最小化重投影误差来恢复位姿轨迹和地标。然而,在弱纹理环境中^[10],基于点特征的 SLAM 方法难以提取到稳定且足够的关键点,导致关键帧跟踪失败,影响系统的定位和地图构建,降低位姿估计精度和系统稳定性。此外,在结构化环境下,这些方法未能充分利用如直线、平面等结构化信息,导致建模的环境整体结构不准确,影响系统的鲁棒性。为解决该问题,Pumarola 等^[11]在 ORB-SLAM2 基础上引入了线特征,使用直线段检测法(least significant digit, LSD)检测线段,并通过 LBD(line band descriptor)算法进行线段匹配,从而在低纹理环境中能够提取到更多有效特征,提高系统的鲁棒性和稳定性。Zhou 等^[12]进一步通过优化线端点的逆深度和光度误差提高了位姿估计精度,使系统在有显著线特征的场景中表现更为优异。

尽管基于线特征的 SLAM 算法在室内环境中的定位精度有所提升,但在相机快速移动时仍面临挑战。快速运动引起的运动模糊、视角变化以及动态场景中的干扰,会影响特征提取和匹配的准确性,特别是在关键帧选择过程中,特征匹配不准确可能导致关键帧跟踪失败,从而影响定位和建图的精度。为此,视觉惯性融合的 SLAM 方法^[13]通过结合 IMU 数据,提供相机的姿态和运动速度信息,纠正快速运动引起的误差,显著提高了 SLAM 系统的鲁棒性和精度。Qin 等^[14]通过紧耦合的非线性优化方法,融合 IMU 预积分测量与视觉特征观测数据,校正视觉信息中的误差,确保系统在快速运动和复杂环境中的稳定性和精度。为进一步提高弱纹理环境下的位姿估计精度,He 等^[15]结合点特征和线特征模型,通过最小化投影空间中线端点与观测线特征之间的距离,优化了位姿估计过程,更好地捕捉场景中的几何结构。基于 VINS-Mono 的 PL-VINS 则引入了改进的 LSD 算法,将线特征纳入姿态估计中,提升了在结构化环境中的定位精度和鲁棒性,从而扩展了

SLAM 系统的应用范围。

上述基于点线特征的视觉惯性 SLAM 方法在移动机器人定位与自主导航领域得到了广泛应用,但在实际应用中仍存在一些挑战。线特征提取与匹配过程中,误匹配可能导致系统生成冗余关键帧,影响位姿估计的精度。这些冗余关键帧不仅增加了计算负担,还可能引入错误的位置信息,降低系统的鲁棒性。IMU 在长时间运行中产生的累计误差可能导致位姿漂移,影响全局定位精度。这些问题限制了 SLAM 系统在长期运行和复杂环境中的表现,需要进一步优化以提高其可靠性和精度。

为了解决关键点特征稀疏、结构化特征应用不充分的问题,本文基于 EDlines 线段提取方法,结合高斯图像金字塔提取多尺度线段,并在每层图像中使用特定的噪声方差来模拟线段端点检测的不确定性,从而提高线段匹配的准确性。同时,采用 LBD 方法为提取线段关联二进制描述符,利用平铺技术对线特征进行分块处理,并按区域进行线段匹配,提高线特征匹配效率。针对相机快速移动时关键帧容易跟踪失败的问题,本文通过优化 IMU 预积分模型,将双目视觉的点特征重投影误差、线特征的重投影误差与惯性传感器的预积分约束相结合,进行滑动窗口的非线性优化,以此提高系统的位姿估计精度,纠正快速运动引起的误差。最后本文在 EuRoC 公开数据集进行多次实验,验证了本文方法的可行性和有效性。结果表明,本文算法在室内环境下拥有更强的鲁棒性和更高的定位精度。

1 点线双目视觉惯性 SLAM 方法

本文提出了一种基于点线特征的双目视觉惯性 SLAM 方法,如图 1 所示,包含 3 个主要线程:跟踪、局部优化和回环检测。系统中点线特征的提取与 IMU 信息的处理并行进行,以提高效率和准确性。

双目图像和 IMU 测量数据输入到跟踪线程。跟踪线程对图像进行高斯图像金字塔处理,生成多尺度图像以捕捉不同分辨率下的特征信息,从而增强系统的鲁棒性。在这些图像中,点和线特征会分别进行提取、匹配和筛选,用于构建场景的几何信息,并为后续的视觉惯性对齐提供数据支持。与此同时,系统对时间序列中连续两帧之间的 IMU 信息进行预积分处理,将 IMU 的加速度计和陀螺仪数据整合成紧凑的表示形式减少计算量并提高精度,从而提供对姿态和位置的估计。在此基础上,系统结合视觉和惯性信息生成局部的点线图,实现对场景的初步建模。

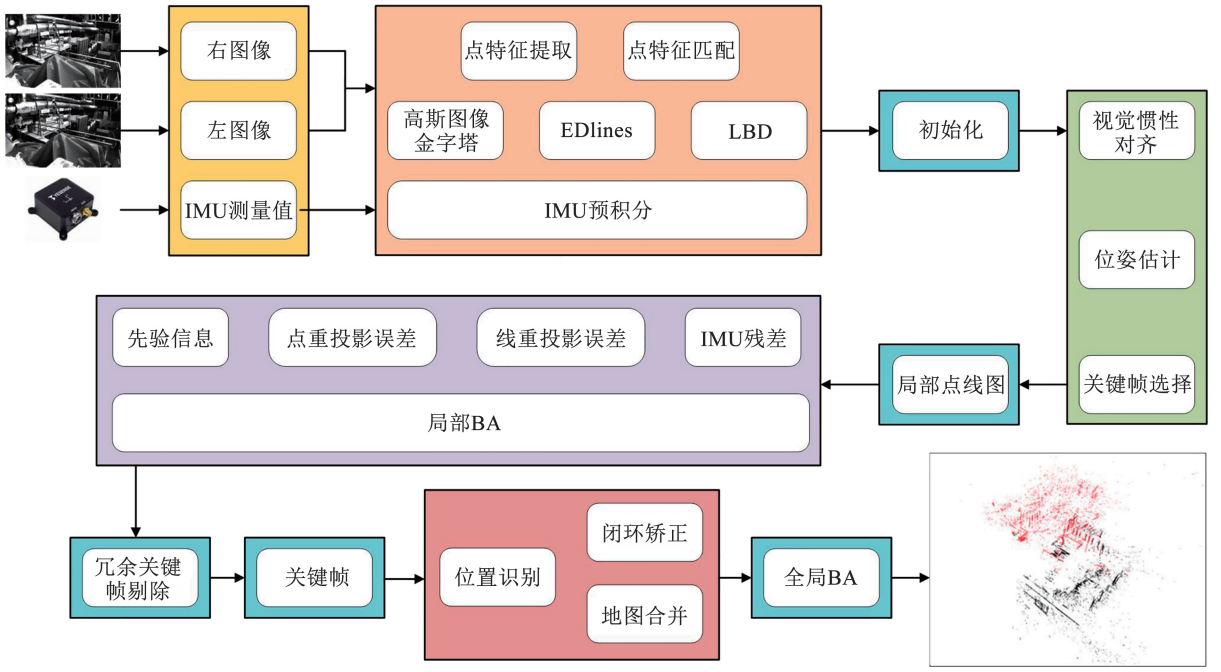


图 1 点线双目视觉惯性 SLAM 方法

Fig. 1 Framework of stereo visual-inertial SLAM method with point and line features

初步建模后,局部优化线程使用 BA (bundle adjustment) 算法对局部地图进行优化,结合边缘先验信息、IMU 残差、点特征重投影误差以及线特征重投影误差进行优化处理,提高系统的定位精度,剔除冗余的关键帧,保持地图的简洁性。最后,系统通过回环检测进行闭环矫正,识别并消除累计误差导致的漂移,增强系统的长期稳定性。通过闭环矫正,系统进行全局优化,更新地图和位姿,使得最终生成的全局地图更加准确和一致。

2 视觉惯性融合中的模型与特征处理

2.1 IMU 预积分模型

针对相机快速移动时关键帧跟踪失败的问题,融合相机和 IMU (惯性测量单元) 信息,可以增强 SLAM 系统的鲁棒性。IMU 提供高频姿态和加速度信息,可以弥补视觉特征不稳定时的不足,帮助维持关键帧的有效跟踪,避免跟踪失败。多传感器融合是提高 SLAM 系统在复杂环境中稳定性和精确性的有效方法。

惯性测量组合作为惯性导航的基本测量单元,由加速度计和陀螺仪组成。IMU 使用陀螺仪和加速度计测量角速度和线性加速度。陀螺仪测量模型可以描述如下:

$$\tilde{\omega}_b(t) = \omega_b(t) + \mathbf{b}_g(t) + \boldsymbol{\eta}_g(t) \quad (1)$$

式中: \mathbf{b}_g 、 $\boldsymbol{\eta}_g$ 分别为陀螺仪的偏置和白噪声。

加速度计测量模型可以描述如下:

$$\tilde{\mathbf{a}}_b(t) = \mathbf{R}_b^w(\mathbf{a}_w - \mathbf{g}_w) + \mathbf{b}_a(t) + \boldsymbol{\eta}_a(t) \quad (2)$$

式中: \mathbf{b}_a 、 $\boldsymbol{\eta}_a$ 分别为加速度计的偏置和白噪声。

为了减少计算开销并避免在每次迭代后重新积分位置、速度和旋转变量,本文基于 IMU 预积分原理进行 IMU 测量的预积分如下:

$$\Delta \tilde{\mathbf{R}}_{ij} \approx \Delta \mathbf{R}_{ij} \exp(\delta \phi_{ij}) = \mathbf{R}_i^T \mathbf{R}_j \exp(\delta \phi_{ij}) \quad (3)$$

$$\Delta \tilde{\mathbf{v}}_{ij} \approx \Delta \mathbf{v}_{ij} + \delta \mathbf{v}_{ij} = \mathbf{R}_i^T (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g} \Delta t_{ij}) + \delta \mathbf{v}_{ij} \quad (4)$$

$$\Delta \tilde{\mathbf{p}}_{ij} \approx \Delta \mathbf{p}_{ij} + \delta \mathbf{p}_{ij} = \mathbf{R}_i^T \left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i \Delta t_{ij} - \frac{1}{2} \mathbf{g} \Delta t_{ij}^2 \right) + \delta \mathbf{p}_{ij} \quad (5)$$

式中: $\Delta \tilde{\mathbf{R}}_{ij}$ 、 $\Delta \tilde{\mathbf{v}}_{ij}$ 、 $\Delta \tilde{\mathbf{p}}_{ij}$ 分别为从第 i 帧到第 j 帧的姿态、速度和位置向量。

2.2 空间直线参数化

在视觉 SLAM 中,空间直线的表示在不同阶段有所不同^[16]。前端使用 Plücker 坐标捕捉 3D 线的几何特性,但在后端优化时,为避免过参数化和奇异性问题,改用正交表示法,这种方式可以简化优化过程,提高算法的稳定性。

2.2.1 Plücker 坐标表示

如图 2 所示,给定空间线段 l 的两个端点为 \mathbf{p} 和 \mathbf{q} ,空间直线 L 的 Plücker 坐标可以表示为

$$L = \begin{bmatrix} \mathbf{p} \times \mathbf{q} \\ \mathbf{p} - \mathbf{q} \end{bmatrix} = \begin{bmatrix} \mathbf{n} \\ \mathbf{v} \end{bmatrix} \in \mathbf{R}^6 \quad (6)$$

式中: $\mathbf{n} \in \mathbf{R}^3$ 是由直线 L 和坐标原点确定的平面 π_w 的法向量, $\mathbf{v} \in \mathbf{R}^3$ 是直线 L 的方向向量,满足 Klein 二次曲面约束 $\mathbf{n}^T \mathbf{v} = 0$ 。

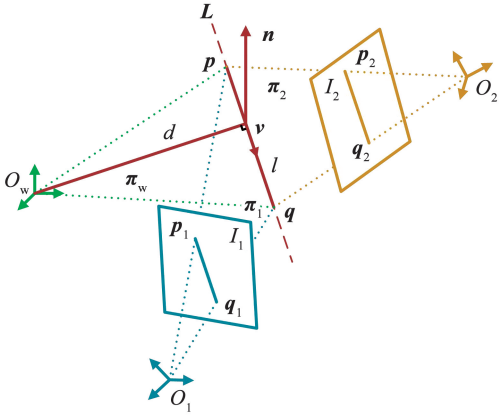


图 2 空间直线的 Plücker 坐标表示

Fig. 2 Plücker coordinate representation of a spatial line

将 3D 线从世界坐标系下的 Plücker 坐标转换为相机坐标系下的 Plücker 坐标满足以下公式:

$$L^c = T^{wc} L = \begin{pmatrix} R^{wc} & [t^{wc}]_{\times} R^{wc} \\ 0 & R^{wc} \end{pmatrix} \begin{pmatrix} n^w \\ v^w \end{pmatrix} \quad (7)$$

式中: T^{wc} 为从世界坐标系转换为相机坐标系的转换矩阵, R^{wc} 为旋转矩阵, t^{wc} 为平移向量。

对于 3D 线段 l , 其在归一化平面 I_i 上的投影为 l_i , 其两个端点的投影分别为 $p_i \in \mathbf{R}^3$ 和 $q_i \in \mathbf{R}^3$ 。相机坐标系的原点坐标为 $O_i^c = (x_i, y_i, z_i)^T$, 平面 $\pi_i = (\pi_1, \pi_2, \pi_3, \pi_4)$ 的坐标可以通过下式获得:

$$\begin{cases} [\pi_1, \pi_2, \pi_3]^T = [p_1]_{\times} q_1 \\ \pi_4 = \pi_1 x_i + \pi_2 y_i + \pi_3 z_i \end{cases} \quad (8)$$

式中 $[\cdot]_{\times}$ 为三维向量的斜对称矩阵。则根据得到的平面坐标 $\pi_i = (\pi_1, \pi_2, \pi_3, \pi_4)$, 可以得到 3D 线段 l 的 Plücker 坐标为

$$L^* = \begin{bmatrix} [n]_{\times} & v \\ -v^T & 0 \end{bmatrix} = \pi_i \pi_k^T - \pi_k \pi_i^T \quad (9)$$

2.2.2 正交表示

正交表示法通过减少参数数目和去除约束, 将 3D 线优化问题转化为一个更为简单、易解且高效的优化问题, 有助于在后端 BA 中实现更好的性能。空间直线 L 的正交表示 $(U, W) \in \text{SO}(3) \times \text{SO}(2)$ 可以通过对矩阵 $[n|v]$ 进行正交-三角分解获得:

$$[n|v] = U \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \\ 0 & 0 \end{bmatrix}, W = \begin{bmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{bmatrix} \quad (10)$$

Plücker 坐标可以由下式转换为正交表示:

$$[n|v] = \begin{bmatrix} n & v & n \times v \\ \|n\| & \|v\| & \|n \times v\| \end{bmatrix} \begin{bmatrix} \|n\| & 0 \\ 0 & \|v\| \\ 0 & 0 \end{bmatrix} \quad (11)$$

可以得到:

$$U = R(\delta) = \begin{bmatrix} n & v & n \times v \\ \|n\| & \|v\| & \|n \times v\| \end{bmatrix} \quad (12)$$

$$W = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{bmatrix} = \frac{1}{\sqrt{\|n\|^2 + \|v\|^2}} \begin{bmatrix} \|n\| & -\|v\| \\ \|v\| & \|n\| \end{bmatrix} \quad (13)$$

式中: U 为线坐标与相机坐标系之间的旋转矩阵, $\delta = (\delta_1, \delta_2, \delta_3)$ 为线坐标相对于相机坐标系的旋转角度, $\theta \in (0, \pi/2)$, 原点 O_w 到直线 L 的距离为 $d = \|n\|/\|v\| = w_1/w_2$ 。将正交表示定义为四参数向量:

$$O = (\tau^T, \theta)^T \quad (14)$$

给定正交表示 (U, W) , 空间线段 l 的 Plücker 坐标可以表示为

$$L = \begin{bmatrix} w_1 \mu_1 \\ w_2 \mu_2 \end{bmatrix} \quad (15)$$

式中 μ_i 为 U 的第 i 列。

2.3 线特征提取与匹配

针对弱纹理场景中, 关键点特征稀疏、结构化场景中的特征应用不充分的问题。引入线特征可以捕捉更多几何信息, 增强 SLAM 系统的鲁棒性和准确性, 尤其在纹理少或结构化明显的环境中。结合点特征和线特征能更全面地利用场景信息, 改善定位与建图效果。传统线特征提取与匹配方法直接从图像中提取线段, 并通过简单的描述符进行匹配, 存在对尺度变化适应性差、匹配效率低的问题。此外, 端点检测过程中存在较大的不确定性, 导致匹配过程的鲁棒性不足, 进而影响了系统性能。

本文提出了一种改进的线段提取和匹配方法。图像经过高斯金字塔处理, 生成了多层次的图像表示。在每一层图像上, 采用 EDlines 方法提取多尺度线段, 以提高线段匹配的尺度不变性。为处理端点检测中的不确定性, 本文模拟了线段端点的噪声方差, 使用 LBD 方法为每个提取的线段生成二进制描述符。通过平铺技术加速线段匹配, 从而提升线特征匹配的鲁棒性和效率。

2.3.1 高斯图像金字塔

对每个处理的图像, 计算其高斯图像金字塔。高斯图像金字塔通过迭代应用高斯模糊和平滑处理, 生成多层次的图像表示, 使得特征提取可以在多个尺度上进行, 增强后续步骤的鲁棒性和效率。

2.3.2 线段提取

在图像金字塔的每一层上, 使用 EDlines 方法进行线段提取。EDlines 是一种高效的线段检测算

法,基于边缘检测方法(edge drawing, ED)技术,从边缘点出发,沿梯度方向追踪,最终形成线段。EDlines方法具有高重复性、精度和计算效率,其线性时间复杂度使其在处理大规模图像数据时具有显著优势,同时具备误检控制,能够在不同图像条件下稳定提取线段。

在图像金字塔的每一层上,本文使用特定的噪声方差来模拟端点检测过程中的不确定性。具体来说,每个线段的端点 p_i 和 q_i 都存在一定的检测误差,用 $\sigma_{u_i}^2 \in \mathbf{R}$ 表示图像端点 p_i, q_i 每个坐标上的噪声方差。这种不确定性建模能够更好地适应实际应用中的检测误差,提升匹配的准确性。

2.3.3 二进制描述符关联

为了实现相邻帧间的线段匹配,本文使用LBD方法为每个提取的线段关联一个二进制描述符。LBD通过对线段附近的图像纹理进行编码,生成一个独特的二进制特征向量。具体来说,LBD在每条线段周围定义一个带状区域,并对该区域内的像素梯度方向和幅值进行统计,生成描述符。这个过程有效地捕捉了线段周围的局部结构信息,使描述符对旋转、尺度变化以及光照变化具有较好的鲁棒性。

2.3.4 平铺技术加速匹配

在线段提取时,本文使用一种平铺技术来加速线段匹配。具体来说,在提取阶段,为每条线段计算其法线方向和带符号的距离。在图像平面上,采用如下形式的2D线方程:

$$\boldsymbol{\eta}^T \mathbf{u} - h = 0 \quad (16)$$

式中: $\boldsymbol{\eta} = [\eta_x, \eta_y] \in \mathbf{R}^2$ 为单位法向量, $h \in \mathbf{R}$ 为所考虑线与图像平面原点之间的有符号距离。给定两个点 $\mathbf{u}_1, \mathbf{u}_2 \in \Omega$,单位法向量 $\boldsymbol{\eta}$ 和有符号距离 h 可由下式得出:

$$\begin{cases} \boldsymbol{\eta} = \frac{[\Delta v, -\Delta u]^T}{\sqrt{(\Delta^2 u + \Delta^2 v)}} \\ h = \boldsymbol{\eta}^T \mathbf{u}_i \end{cases} \quad (17)$$

式中:法线方向 $\boldsymbol{\eta} = \arctan 2(\eta_x, \eta_y)$ 为线段的朝向,带符号的距离 h 表示线段到图像原点的距离。本文将2D参数流形 $(\boldsymbol{\eta}, h)$ 划分为多个平铺区域,将具有相似参数的线段归为一组。对于每张图像 I_i ,按其对应的平铺区域对提取的线段进行分组,并将结果存储在 I_i 中。这种方法不仅减少了匹配时的计算量,还提高了匹配的精度。

2.3.5 线段匹配

在进行线段匹配时,将一个3D线段 $l \in M_s$ 投影到目标图像 I_i 上,计算其投影线的表示 $(\boldsymbol{\eta}_i, h_i)$ 。然后,识别出在 I_i 中与 l 的参数 $(\boldsymbol{\eta}, h)$ 落在同一平铺

区域内的所有线段作为候选匹配,根据描述符之间的汉明距离对这些候选匹配进行评分。

2.3.6 剔除误匹配

为了排除可能的错误匹配,本文采用以下两项验证条件:

1) 汉明距离比值。匹配线段的汉明距离与第2近邻的比值需低于某个阈值,确保最近邻匹配优于次优匹配,从而提高匹配的可靠性。

2) 线段距离阈值。匹配线段在图像平面上的投影距离需低于某个阈值,以确保匹配线段在图像中的位置接近,进一步提高匹配的精度。

这种多层次的匹配验证机制显著减少了误匹配的概率,提升了匹配的可靠性和精度。

3 后端非线性优化

本文对点、线特征的投影误差进行了理论分析和公式推导,包括空间点的重投影误差、空间线的重投影误差以及局部BA优化函数。

3.1 点特征重投影误差计算

对于特征点 $p \in P$,观测点为 p_i ,世界坐标系的坐标为 $\mathbf{P}^w = [X^w, Y^w, Z^w]^T$,相机坐标系下为 $\mathbf{P}^c = [X^c, Y^c, Z^c]^T$,投影坐标为 $\mathbf{P} = [u, v]^T$ 。 \mathbf{K} 为相机内参矩阵。将一个点从世界坐标系变换到相机坐标系的相机变换矩阵表示为

$$\mathbf{T}^{wc} = \begin{bmatrix} \mathbf{R}^{wc} & \mathbf{t}^{wc} \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (18)$$

式中: \mathbf{R}^{wc} 为旋转矩阵, \mathbf{t}^{wc} 为平移矩阵。对于3D点,本文使用上标 w 和 c 分别表示在世界坐标系和相机坐标系。则投影坐标、世界坐标系下的坐标以及相机坐标系下的坐标满足以下关系:

$$\mathbf{P}^c = (\mathbf{R}^{wc} \mathbf{P}^w + \mathbf{t}^{wc}) \quad (19)$$

$$\mathbf{P} = \pi(\mathbf{K} \mathbf{P}^c) = \left[f_x \frac{X}{Z} + c_x, f_y \frac{Y}{Z} + c_y \right]^T \quad (20)$$

式中: Z 为相机坐标系下的深度, f 为焦距, c 为图像中点。则单目观测点的点重投影误差为

$$\mathbf{r}^m = \begin{pmatrix} u_L \\ v_L \end{pmatrix} - \begin{pmatrix} u_L^i \\ v_L^i \end{pmatrix} \quad (21)$$

本文中,第 i 个双目图像 (I_{iL}, I_{iR}) 由一个矫正后的左侧彩色图像 I_{iL} 和右侧彩色图像 I_{iR} 组成。对于双目相机,三维坐标 \mathbf{P}^w 在左、右相机下的投影坐标分别为 $\mathbf{P}_L = [u_L, v_L]^T$ 和 $\mathbf{P}_R = [u_R, v_R]^T$,由视差估计可以得到深度:

$$Z = \frac{bf}{(u_L - u_R)} \quad (22)$$

式中 b 为立体基线。

右相机与左相机的不同之处在于其位置相对于左相机的基线偏移 b 。因此,对于同一个三维点 P^w ,它在右相机坐标系中的 X 坐标会变为 $X - b$,基于立体视觉中左右相机的几何关系可以得出:

$$u_R = f \frac{(X - b)}{Z} + c \quad (23)$$

则立体点重投影误差定义为

$$\mathbf{r}^s = \begin{pmatrix} u_L \\ v_L \end{pmatrix} - \begin{pmatrix} u_L^i \\ v_L^i \end{pmatrix} \quad (24)$$

对于双目相机,既包含左相机坐标系下的单目点重投影误差又包含立体点重投影误差,则点特征重投影误差为

$$\mathbf{r}_p = \mathbf{r}^m + \mathbf{r}^s \quad (25)$$

3.2 线特征重投影误差计算

对于特征线,重投影误差可以定义为观察的线段端点与同一图像平面中重投影线之间的距离,如图 3 所示,给定一个在世界坐标系中使用 Plücker 坐标表示的空间直线 l ,其线段端点在图像 I_i 中的观测点为 (p_i, q_i) ,世界坐标系中的端点为 (P, Q) 以及平面 I_i 中的线段投影端点 (p, q) 。

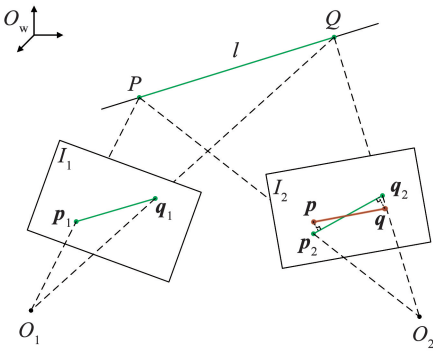


图 3 线投影误差

Fig. 3 Line projection errors

由式(6)可得直线 l 在相机坐标系的 Plücker 坐标为 $L^c = [n^c, v^c]^T$ 。则直线 l 在平面 I_k 上的投影为

$$L^k = K_l n^c = \begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix} \quad (26)$$

其中

$$K_l = \begin{bmatrix} f_v & 0 & 0 \\ 0 & f_u & 0 \\ -f_v c_u & -f_u c_v & f_u f_v \end{bmatrix}$$

式中 K_l 为线投影矩阵。

则线特征重投影误差为

$$\mathbf{r}_L = \begin{bmatrix} d(\mathbf{p}^T, L^k) \\ d(\mathbf{q}^T, L^k) \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{p}^T L^k}{\sqrt{L_1^2 + L_2^2}} \\ \frac{\mathbf{q}^T L^k}{\sqrt{L_1^2 + L_2^2}} \end{bmatrix} \quad (27)$$

3.3 基于滑动窗口的非线性优化

基于滑动窗口因子图模型,本文采用了视觉惯性非线性紧耦合优化方法,见图 4。

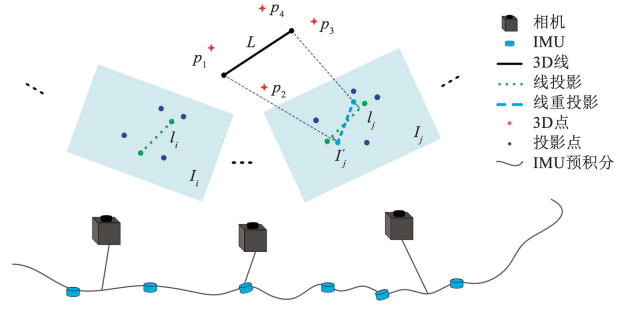


图 4 滑动窗口的非线性优化

Fig. 4 Nonlinear optimization of the sliding window

需要优化的变量如下:

$$\boldsymbol{\chi} = [\lambda_1, \lambda_2, \dots, \lambda_n, \mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_m, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_h] \quad (28)$$

$$\mathbf{x}_k = [p_{b_k}^w, q_{b_k}^w, v_{b_k}^w, b_a^k, b_g^k] \quad (29)$$

式中: $k \in [0, h]$, $\lambda_i (i \in [1, n])$ 为三维点的逆深度, $\mathbf{O}_j (j \in ([1, m]))$ 为线的正交表示, $\mathbf{x}_k (k \in [0, h])$ 为滑动窗口中第 k 个关键帧在世界坐标系中的位置、旋转、速度以及加速度计和陀螺仪的偏置。

基于需要优化的变量,可以建立非线性优化函数:

$$F(\boldsymbol{\chi}) = \min_{\boldsymbol{\chi}} \left\{ \begin{aligned} & \left\| \mathbf{r}_p - \mathbf{H}_p \boldsymbol{\chi} \right\|_{\Sigma_p}^2 + \\ & \sum_{k \in B} \left\| \mathbf{r}_B(\hat{z}_{b_{k+1}}^k, \boldsymbol{\chi}) \right\|_{\Sigma_{b_{k+1}}}^2 + \\ & \sum_K \left(\sum_{i \in N} \rho(\| \mathbf{r}_p^m(\hat{z}_j^{c_i}, \boldsymbol{\chi}) \|_{\Sigma_p}) + \right. \\ & \left. \sum_{i \in N} \rho(\| \mathbf{r}_p^s(\hat{z}_j^{c_i}, \boldsymbol{\chi}) \|_{\Sigma_p}) + \right. \\ & \left. \sum_{i \in M} \rho(\| \mathbf{r}_L(\hat{z}_j^{c_i}, \boldsymbol{\chi}) \|_{\Sigma_L}) \right) \end{aligned} \right\} \quad (30)$$

式中: $\mathbf{r}_p - \mathbf{H}_p \boldsymbol{\chi}$ 为边缘化的先验信息残差, $\mathbf{r}_B(\hat{z}_{b_{k+1}}^k, \boldsymbol{\chi})$ 为 IMU 的测量残差, $\mathbf{r}_p(\hat{z}_j^{c_i}, \boldsymbol{\chi})$ 、 $\mathbf{r}_L(\hat{z}_j^{c_i}, \boldsymbol{\chi})$ 分别为特征点和特征线的重投影误差, Σ 为协方差矩阵, ρ 为鲁棒的 Huber 代价函数。定义为 $\rho(\| \mathbf{r} \|_{\Sigma}) = \mathbf{r}^T \Sigma \mathbf{r}$, K 为关键帧集合, N, M 分别为当前帧中的点和线的数量。此优化函数的目标是最小化先验信息残差、IMU 测量残差、特征点重投影误差、特征线重投影误差的加权和。

4 结果与分析

本文在公共数据集 EuRoC^[17] 上验证所提出的方法。这些公共数据集都包含立体图像、IMU 测量和相应的真实轨迹。本文所有实验均在系统为 Ubuntu 20.04、ROS 版本为 Noetic、配备 Intel Core i7-12700H(14 核 20 线程)处理器和 16 GB 内存的计算机上进行。

4.1 基于 EuRoC 数据集的算法评估

EuRoC 数据集包含由微型飞行器收集的立体图像和相应的 IMU 测量数据,并广泛用于视觉惯性系统的评估。数据集共有 11 个序列,分为 3 个难度

级别:“容易”(MH01、MH02、V101、V201)、“中等”(MH03、MH04、V102、V202)和“困难”(MH05、V103、V203)。MH 序列是在工厂环境中采集的,VR 序列则通过手持相机在室内采集。容易序列的场景为宽敞的实验室或工业环境,特征丰富;中等序列的场景为狭窄的走廊或有障碍物的区域,特征较少且不均匀,飞行器运动轨迹复杂;困难序列的场景则包括光照极端的区域,特征稀少,飞行器运动快速且不规则。数据集的地面真实值由 Vicon 运动捕捉系统收集,提供了每一时刻的姿态和 IMU 偏置数据。本文算法在 EuRoC 数据集上的评估结果见图 5。

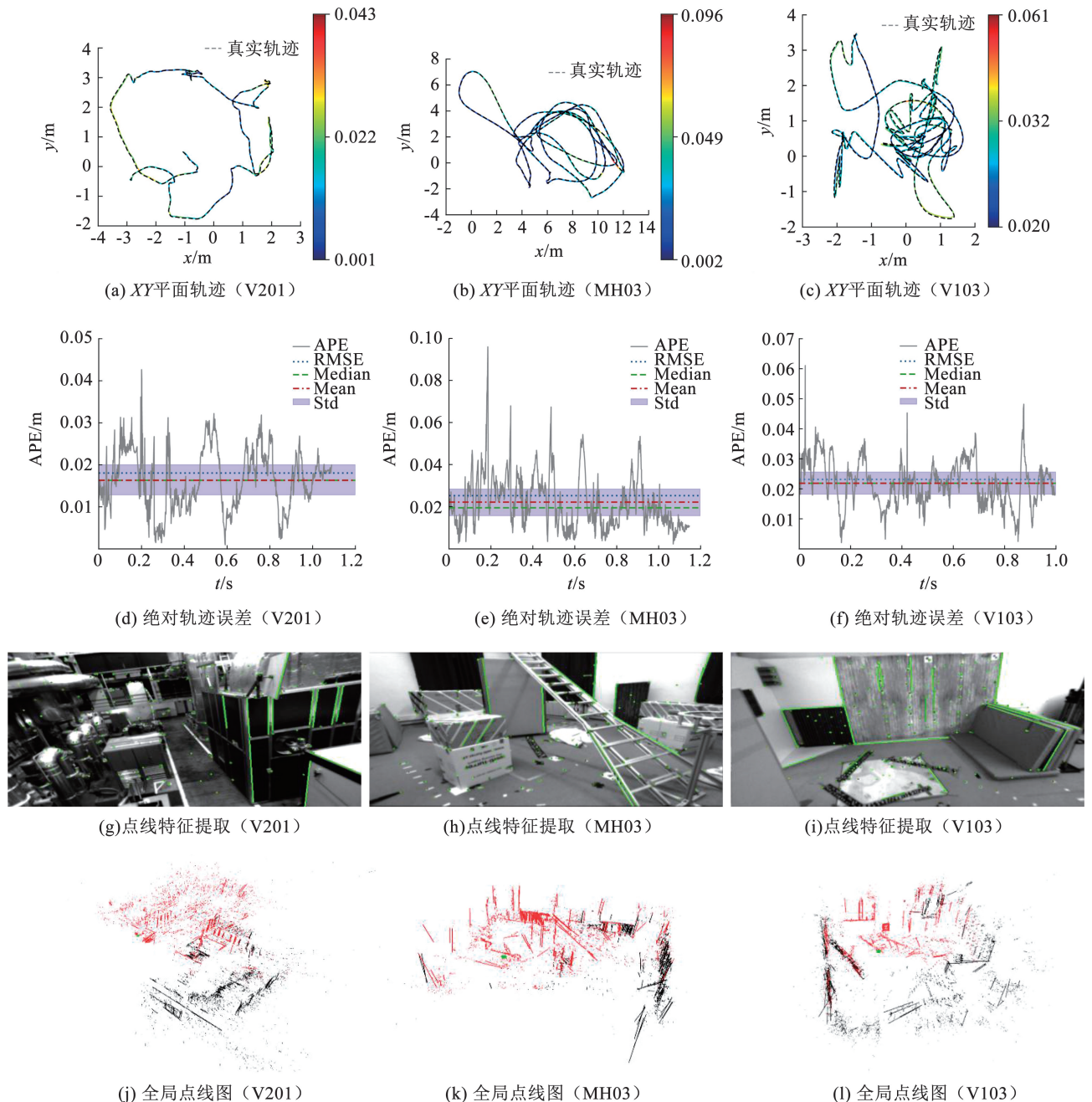


图5 本文算法在 EuRoC 数据集的评估

Fig. 5 Evaluation of the proposed algorithm on the EuRoC dataset

图 5(a) ~ (c) 显示了算法在 EuRoC 数据集不同难度序列(容易的 V201、中等的 MH03、困难的 V103)上的 XY 平面轨迹误差。图 5(d) ~ (f) 展示了轨迹误差随时间变化的具体数值。图 5(g) ~ (i) 反映了在对应数据集序列进行跟踪时的点线特征。图 5(j) ~ (l) 显示了使用 EuRoC 数据集重建地标和轨迹的三维地图。从图 5(a) ~ (f) 可以看出,本文算法在各个难度序列上都展现了与地面真实情况相符的良好精度性能。此外,图 5(g) ~ (i) 中的三维地图重建结果显示,算法通过丰富的特征点和线,成功描述了环境的几何形状,验证了其在复杂场景下的有效性。

表 1 本文算法与点线、视觉惯性 SLAM 的均方根误差

Tab. 1 RMSE of our algorithm and point-line, visual-inertial SLAM

序号	MH01	MH02	MH03	MH04	MH05	V101	V102	V103	V201	V202	V203
ORB-SLAM3	0.036	0.033	0.035	0.051	0.082	0.038	0.014	0.024	0.032	0.014	0.024
PL-SLAM ^[18]	0.042	0.052	0.040	0.064	0.070	0.042	0.046	0.068	0.061	0.057	0.126
PL-VIO	0.126	0.232	0.112	0.139	0.256	0.059	0.076	0.130	0.082	0.094	0.260
PL-VINS	0.093	0.084	0.088	0.202	0.252	0.072	0.092	0.152	0.067	0.113	0.182
OKVIS	0.160	0.220	0.240	0.340	0.470	0.090	0.200	0.240	0.130	0.160	0.290
VINS-Mono	0.084	0.105	0.074	0.122	0.147	0.047	0.066	0.180	0.056	0.090	0.144
Kimera ^[19]	0.080	0.090	0.110	0.150	0.240	0.050	0.110	0.120	0.070	0.100	0.190
VINS-Fusion	0.166	0.152	0.125	0.280	0.284	0.076	0.069	0.114	0.066	0.091	0.096
本文算法	0.035	0.033	0.028	0.047	0.068	0.037	0.013	0.023	0.016	0.014	0.023

为了进一步验证本文算法的可行性,将本文算法与 ORB-SLAM3、VINS-Fusion 以及 PL-VINS 在相同配置环境下进行了比较。实验中,所有算法均采用双目加 IMU 的多传感器组合,且未对源码参数进行任何修改。每个序列进行了 10 组实验,并对数据结果取平均值。通过绝对轨迹误差的均方根误差和平均误差进行对比,定量分析了本文算法的性能。

由图 6 和表 2 可以看出,本文算法在 EuRoC 数据集上的均方根误差和平均误差均取得了最佳效果。EuRoC 数据集分为 MH 序列和 VR 序列,MH 序列是在光照条件较差、包含低纹理场景和大量结构化信息的工厂环境中采集的,而 VR 序列是通过手持相机在室内采集的,具有大量的无规则高速晃动。实验结果显示,本文算法在均方根误差和平均误差方面相较于 VINS-Fusion 平均提升了 77.56% 和 77.55%;相较于 PL-VINS 提升了 76.95% 和 76.97%;相较于 ORB-SLAM3 提升了 27.37% 和 27.47%。上述实验结果表明,本文算法在室内环境下能够有效解决弱纹理场景中关键点特征稀疏、结构化场景中特征应用不充分,以及相机快速移动时关键帧容易跟踪失败的问题,本文算法在这些复杂场景中的定位精度和鲁棒性具有显著优势。

4.2 定位精度评估

为了评估本文算法在 EuRoC 数据集上的定位精度,本文将其与 ORB-SLAM3、PL-SLAM^[18]、PL-VIO、PL-VINS、OKVIS、VINS-Mono、Kimera^[19]、VINS-Fusion 等算法进行了对比实验,见表 1。在表 1 中,本文使用绝对轨迹误差(absolute trajectory error, ATE)的均方根误差(root mean square error, RMSE)来评价定位性能。为了减少实验的偶然性并保证实验的可靠性,对本文算法在每个序列上进行了 10 组实验,并对结果取平均值。对比算法的数据结果则来自相关文献。结果显示,在 11 个序列中,本文算法的精度最高,其次为 ORB-SLAM3。

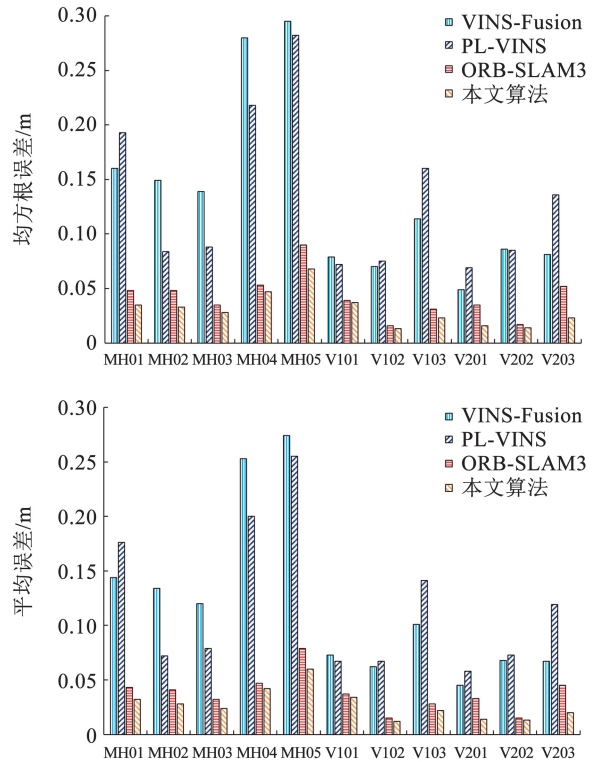


图 6 各序列绝对轨迹误差

Fig. 6 Absolute trajectory error for each sequence

表 2 本文算法与双目点线、视觉惯性 SLAM 的绝对轨迹误差

Tab.2 ATE of our algorithm compared with stereo point-line and visual-inertial SLAM

m

序列	VINS-Fusion		PL-VINS		ORB-SLAM3		本文算法	
	均方根误差	平均误差	均方根误差	平均误差	均方根误差	平均误差	均方根误差	平均误差
MH01	0.160	0.144	0.193	0.176	0.048	0.043	0.035	0.032
MH02	0.149	0.134	0.084	0.072	0.048	0.041	0.033	0.028
MH03	0.139	0.120	0.088	0.079	0.035	0.032	0.028	0.024
MH04	0.280	0.253	0.218	0.200	0.053	0.047	0.047	0.042
MH05	0.295	0.274	0.282	0.255	0.090	0.079	0.068	0.060
V101	0.079	0.073	0.072	0.067	0.039	0.037	0.037	0.034
V102	0.070	0.062	0.075	0.067	0.016	0.015	0.013	0.012
V103	0.114	0.101	0.160	0.141	0.031	0.028	0.023	0.022
V201	0.049	0.045	0.069	0.058	0.035	0.033	0.016	0.014
V202	0.086	0.068	0.085	0.073	0.017	0.015	0.014	0.013
V203	0.081	0.067	0.136	0.119	0.052	0.045	0.023	0.020

为了更直观地展示结果,本文还提供了 PL-VINS、VINS-Fusion、ORB-SLAM3 和本文算法在 V201(容易序列)、MH03(中等序列)和 V203(困难序列)上的 XY 平面估计轨迹。这 3 种序列涵盖了 EuRoC 数据集的所有场景和难易程度,包含无纹理区域、显著的光照变化以及较少的环路闭合等多种挑战性场景。如图 7 所示,第 1 行展示了 3 个序列(V201、MH03

和 V203)的整体轨迹和局部放大图,第 2 行展示了轨迹误差随时间的变化。从整体轨迹图和局部放大图可以看出,本文算法生成的轨迹与地面真实轨迹的距离最近,尤其是在转弯和加速的情况下表现更为出色。轨迹误差图进一步表明,本文算法在整个过程中基本保持了最小的误差。这些结果直观地展示了本文算法的可行性和在复杂环境中的优势。

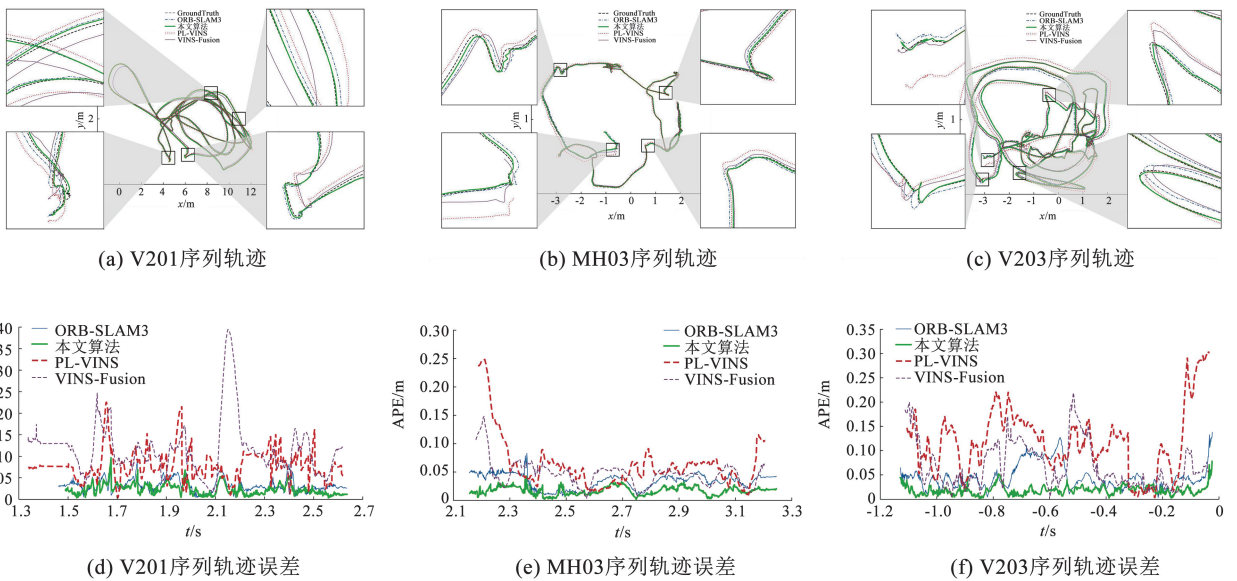


图 7 双目视觉惯性 SLAM 的轨迹对比

Fig.7 Trajectory comparison of stereo visual-inertial SLAM

4.3 消融实验

为了验证每个模块的有效性,本文分别对各个模块进行了启用和禁用测试。具体来说,“Line”表示启用了线特征提取与匹配模块,“IMU”表示加入

了惯性传感器,“B”表示不包含这些模块的基础算法。当绝对轨迹误差的均方根误差超过 1 m 或无法完成某些序列的所有跟踪时,不记录。比较结果见表 3。

表 3 不同模块的均方根误差

Tab. 3 RMSE results comparison for different modules

m

序号	MH01	MH02	MH03	MH04	MH05	V101	V102	V103	V201	V202	V203
B	0.040	0.049	0.049	0.093	0.060	0.087	0.066	0.097	0.062	0.057	
B + Line	0.039	0.048	0.044	0.088	0.054	0.088	0.066	0.087	0.060	0.054	
B + IMU	0.039	0.035	0.032	0.048	0.072	0.038	0.015	0.026	0.029	0.015	0.026
本文算法	0.035	0.033	0.028	0.047	0.068	0.037	0.013	0.023	0.016	0.014	0.023

从整体结果来看,每个模块的启用都对系统性能有积极影响,显著提升了基础算法的效果。在MH序列中,尤其是MH05序列中,场景包含弱纹理场景和结构化特征,仅启用线特征提取与匹配模块时,定位精度达到最高。这证明了线特征能够有效表示结构化信息,并提升算法的定位精度和鲁棒性。在VR序列中,场景是通过手持相机在室内采集的,包含大量无规则的高速晃动。表3结果显示,加入惯性传感器模块后,在剧烈运动时能显著改善特征提取的难度,提高系统的稳定性。整体来看,本文算法通过有效融合各个模块,实现了良好的定位精度和鲁棒性。结果表明,完整的系统在各种复杂环境下都具有出色的性能。

5 结 论

1) 本文提出了一种基于点线特征融合的双目视觉惯性SLAM方法,针对弱纹理场景中关键点特征稀疏、结构化场景中结构化特征应用不足以及相机快速移动导致的关键帧跟踪失败等问题,采用高斯图像金字塔结合EDlines线段提取方法,实现了多尺度线段的提取。通过模拟不同尺度下线段端点检测过程中的不确定性,并使用LBD方法为线段关联二进制描述符,结合平铺技术加速线段匹配,设定距离阈值剔除误匹配,从而显著提高了线特征匹配的鲁棒性和效率。

2) 在系统优化过程中,将双目视觉的点特征重投影误差、线特征重投影误差以及惯性传感器的预积分约束相结合,采用滑动窗口的非线性优化方法,有效提升了系统的位姿估计精度。通过多传感器信息融合,充分发挥了点线特征和惯性信息的互补优势,增强了系统在复杂环境下的鲁棒性和精确度。

3) 在EuRoC数据集的多种环境条件下进行的实验表明,本文方法在处理弱纹理场景、结构化场景以及相机快速运动等复杂条件时,具有明显的性能

优势。实验结果显示,本文方法的轨迹均方根误差达到0.031 m,平均误差为0.027 m,相较于现有算法,展示出更高的定位精度和更强的鲁棒性,验证了各模块的可行性和有效性。

参考文献

- [1] ZHU Jun, LI Hongyi, ZHANG Tao. Camera, LiDAR, and IMU based multi-sensor fusion SLAM: a survey [J]. Tsinghua Science and Technology, 2024, 29(2): 415. DOI:10.26599/TST.2023.9010010
- [2] LIU Xin, WEN Shuhuan, ZHANG Hong. A real-time stereo visual-inertial SLAM system based on point-and-line features [J]. IEEE Transactions on Vehicular Technology, 2023, 72(5): 5747. DOI: 10.1109/TVT.2022.3233721
- [3] 陈明方, 黄良恩, 王森, 等. 移动机器人视觉里程计技术研究综述 [J]. 农业机械学报, 2024, 55(3): 1
CHEN Mingfang, HUANG Liang'en, WANG Sen, et al. Survey of research on visual odometry technology for mobile robots [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 1. DOI:10.6041/j.issn.1000-1298.2024.03.001
- [4] 樊铭瑞, 张世栋, 李运, 等. 融合轨道动力学的小行星探测器自主视觉定位 [J]. 哈尔滨工业大学学报, 2024, 56(5): 19
FAN Mingrui, ZHANG Shidong, LI Yun, et al. Autonomous visual localization for asteroid probe fusion orbital dynamics [J]. Journal of Harbin Institute of Technology, 2024, 56(5): 19. DOI:10.11918/202202032
- [5] GE Yangbing, ZHANG Lilian, WU Yuanxin, et al. PIPO-SLAM: lightweight visual-inertial SLAM with preintegration merging theory and pose-only descriptions of multiple view geometry [J]. IEEE Transactions on Robotics, 2024, 40: 2046. DOI: 10.1109/TRO.2024.3366815
- [6] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611. DOI:10.1109/TPAMI.2017.2658577
- [7] GAO Xiang, WANG Rui, DEMMEL N, et al. LDSO: direct sparse odometry with loop closure [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2018: 2198. DOI:10.1109/IROS.2018.8593376
- [8] MUR-ARTAL R, TARDÓS 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. DOI: 10.1109/TRO.2017.2705103