

文章编号 1004-924X(2025)08-1259-15

基于关键平面的异质特征融合视觉惯性 SLAM 系统

沈晔湖¹, 何一凡^{2*}, 魏季坤¹, 张大庆¹

(1. 苏州科技大学 机械工程学院, 江苏 苏州 215009;

2. 深圳职业技术大学 智能科学与工程研究院, 广东 深圳 518055)

摘要:平面特征作为一种高层几何特征而广泛存在于结构化环境中,对于大多数同时定位与地图构建(Simultaneous Localization and Mapping, SLAM)系统来说是个很好的补充。为了解决特征点与平面特征融合时引入了新的误差并且平面存在着退化的可能,本文提出了一个融合异质特征的单目视觉惯性 SLAM 系统。首先从灰度图像中提取特征点;其次对特征点集合进行三角剖分,并将三角剖分的结果转换到世界坐标系下;接着将初始化过程建模为有约束的优化问题,并用交替方向乘子法分布式求解;然后对相似平面进行聚类,并用所提出的平面碰撞概率模型拟合平面,得到对应的有界平面参数;最后在因子图中引入了平面特征的几何约束,通过误差模型同时优化相机运动以及平面参数。与典型的视觉惯性 SLAM 系统 VINS 相比,本文提出的系统在 EuRoC 数据集的绝对轨迹误差平均值降低了 50%;在 TUM-VI 数据集的绝对轨迹误差平均值降低了 40%。该方法能够在结构化场景中稳定、连续地工作,并且提高了弱纹理区域的定位精度和鲁棒性。

关键词:SLAM;视觉惯性;分布式求解;有界平面提取;非线性优化

中图分类号:TP391 **文献标识码:**A

doi:10.37188/OPE.20253308.1259

CSTR:32169.14.OPE.20253308.1259

A visual inertial SLAM system based on key planes with heterogeneous feature fusion

SHEN Yehu¹, HE Yifan^{2*}, WEI Jikun¹, ZHANG Daqing¹

(1. College of Mechanical Engineering, Suzhou University of Science and Technology,
Suzhou 215009, China;

2. Institute of Intelligent Science and Engineering, Shenzhen Polytechnic University,
Shenzhen 518055, China)

* Corresponding author, E-mail: heyifan@reconova.com

Abstract: Planar feature is widely used in structured environments as a high-level geometric feature and is a good complement to most Simultaneous Localization and Mapping (SLAM) systems. In order to address the fact that new errors were introduced when fusing feature points with planar features and the possi-

收稿日期:2024-12-09;修订日期:2025-02-10.

基金项目:国家自然科学基金(No. 51975394);江苏省研究生科研创新计划(No. SJCX24_1871);江苏省自然科学基金(No. BK20211336)

bility of planar degradation existed, we proposed a monocular visual inertial SLAM system that fuses heterogeneous features in this paper. Firstly, feature points were extracted from grayscale images; secondly, the set of feature points was triangulated and the results of the triangulation were transformed to the world coordinate system. Next, the initialization process was modeled as a constrained optimization problem and solved with the alternating-direction multiplier method in a distributed fashion. Then, the similar planes were clustered and the planes were fitted with the proposed planar collision probability model to get the corresponding bounded-plane parameters. Finally, geometric constraints on the plane features were introduced in the factor graph, and the camera motion as well as the plane parameters were simultaneously optimized by the error model. Compared with the typical visual inertial SLAM system VINS, the mean absolute trajectory error of the system proposed in this paper was reduced by 50% in the EuRoC dataset. The mean absolute trajectory error on the TUM-VI dataset was reduced by 40%. The method works stably and continuously in structured scenes and improves the localization accuracy and robustness in weakly textured regions.

Key words: Simultaneous Localization and Mapping(SLAM); visual inertia; distributed solving; bounded plane extraction; nonlinear optimization

1 引言

随着计算平台与传感器定位技术的不断发展,SLAM技术取得了长足的进步。为了追求SLAM系统的收敛性与鲁棒性,越来越多的传感器应用于SLAM系统中。结构简单的相机与成本低廉的惯性测量单元(Inertial Measurement Unit, IMU)成为当前研究的热点^[1-4]。

然而单目视觉惯性SLAM在移动机器人中的实际应用受到几个因素的限制。首先,特征提取需要适应不同的场景。在无纹理环境中,高维特征(如平面特征,线特征等)与特征点进行异质特征融合往往能取得更高的定位精度。其次,高度非线性系统需要更多的计算资源,而轻型无人设备只能提供有限的计算资源。所以如何实现较好的定位精度,是限制无人设备实际应用的一个不可避免的问题。最后,由于漂移始终存在且无法消除,因此有必要通过回环检测和全局优化来减少系统的总误差。

近年来,为了应对算力不足的情况,传统视觉惯性SLAM大多采用基于卡尔曼滤波^[5]的方法。但是用线性系统对问题建模,求解非线性的优化问题,在迭代时会陷入局部最优解。同时为了缓解估计姿态和路标点导致复杂度随着路标数量呈指数倍增长的问题,Mourikis等人^[6]提出了基于扩展卡尔曼滤波的多传感器融合算法

(Multi-State Constraint Kalman Filter, MSCKF)。随着算力的增加,传统视觉惯性SLAM开始使用基于固定滞后平滑的方法和基于全平滑的方法^[7]。

基于固定滞后平滑的方法对滑动窗口内的测量结果重新进行线性化处理并将其转化为先验信息,使用边缘化剔除历史信息^[8]。Rosinol等人^[9]引入3D动态场景图来缩小机器人和人类感知之间的差距。LV等人^[10]提出了一种多传感器组合的多模态SLAM算法,能够限制连续时间固定滞后平滑器的计算复杂性。

基于全平滑的方法通过求解非线性优化问题来估计整个历史状态,每隔一段时间进行一次全局优化^[11-14]。如:Qin等人^[15]提出了一个基于单目视觉惯性的SLAM系统(A Robust and Versatile Monocular Visual-Inertial State Estimator, VINS-Mono)。它能够实现未知环境下的初始化、在线的外参标定、全局四自由度优化。Liso等人^[16]提出了一种可以全局优化位姿和密集3D模型的方法,通过在线触发回环进行全局位置识别。Zhang等人^[17]提出了一个可全局实时优化位姿和3D重建的SLAM框架,确保3D重建的全局一致性。BASALT^[18]恢复了在更一般的IMU预积分步骤中丢失的关键帧之间的IMU测量值的非线性。

深度学习近年来在计算机视觉领域取得了

突破性进展,其强大的特征提取和表示学习能力使其在SLAM系统中的应用成为一个热点研究方向^[19]。基于深度学习的视觉惯性SLAM模型可以通过学习大量数据中的潜在模式来提高系统的性能。根据后端融合方式,视觉惯性SLAM系统可以分为嵌入深度学习的传统方法以及基于端到端的方法。

嵌入深度学习的传统方法在后端优化上与传统方法相似,只是在前端特征提取时基于深度学习设计提取适应当前场景的鲁棒特征。Ram-bach等人^[20]提出了首个监督学习的视觉惯性SLAM系统,训练的长短期记忆(Long Short Term Memory, LSTM)网络能够学习摄像头位姿和IMU测量值之间的关系。Clark等人^[21]提出了首个端到端的视觉惯性SLAM方法VINet。该方法的前端将视觉特征转化为高维特征,后端基于LSTM网络 and 全连接层将视觉惯性特征融合实现鲁棒定位。Peng等人^[22]提出了一种具有双视觉因子的深度SLAM网络,该网络将光度因子和重投影因子集成到端到端中,能动态学习和调整两视觉因素的置信度图,并可扩展包含IMU元素。

当前部分方法关注密集纹理处的平面,并利用单应性矩阵来跟踪和解决摄像机运动问题。DPPTAM^[23] (Dense Piecewise Planar Tracking And Mapping)通过填充空白区域来重建平面。Shu等人^[24]利用CNN模型提取场景中的平面,解决了地图构建中比例不一致的问题。但是这种方法增加了计算复杂度,所以一些基于视觉惯性的SLAM方法结合了结构化场景中的平面特征。Hua等人^[25]假设平面目标是水平的,以消除对外部姿态估计器的依赖。

本文提出了一种异质特征融合的视觉惯性SLAM系统。该系统利用提取的平面特征来关联点特征,并利用准确模拟平面边界的约束函数避免无限平面扩展的错误假设。然后,根据点和平面之间的几何关系,实现异质特征的数据融合并构建残差,从而增强状态估计过程的鲁棒性。最后,将平面先验整合到全局定位优化中显著减少累积漂移,从而提高系统的整体性能。

本文创新点如下:

(1)针对结构化场景下丰富的平面特征,本

文提出了一种使用单目相机提取场景中有界平面的高效方法,进一步提出了平面碰撞概率模型实现有界平面的边界约束以及平面间的数据融合。

(2)针对视觉惯性SLAM初始化过程较为敏感的问题,本文提出了一种基于交替方向乘子法的初始化方法,通过将初始化过程分解为多个子问题,使其收敛于最优值,既降低了运算复杂度,又提升了位姿估计的精度。

(3)本文提出了一种基于异质特征融合残差的后端优化方法,提升了视觉惯性SLAM系统在弱纹理区域定位精度与鲁棒性。

2 系统概述

2.1 主要流程

图1展示了本文提出的基于关键平面的异质特征融合SLAM系统。本系统由三个模块组成:视觉惯性前端、融合平面特征的视觉惯性里程计以及回环检测。视觉惯性前端分为两个部分:测量值预处理和位姿初始化。测量值预处理部分接收传感器的测量值,检测并跟踪图像中的特征点,对特征点集合进行三角剖分,对IMU数据进行预积分。位姿初始化部分将初始化过程建模为有约束的Lasso^[26]问题,并用交替方向乘子法(Alternating Direction Method of Multipliers, AD-MM)^[27]分布求解。融合平面特征的视觉惯性里程计分为两个部分:平面提取和滑动窗口优化。平面提取部分对场景中的有界平面参数优化并将其中的关键平面参数传入滑动窗口优化。滑动窗口优化部分结合先验信息、IMU残差、视觉残差以及异质特征融合残差进行滑动窗口范围内的状态估计。回环检测模块通过检测发生回环的关键帧,结合平面先验信息,构建基于全平滑的全局四自由度位姿优化。

2.2 符号说明

在这个部分,本文将 $(\bullet)^w, (\bullet)^b, (\bullet)^c$ 分别定义为世界坐标系、IMU坐标系、相机坐标系。本文使用旋转矩阵 R 和四元数 Q 表示旋转, $\otimes$ 表示四元数的乘法运算, $(\bullet)$ 表示带有噪声的测量值或者估计值, H, W 分别表示图像的高和宽。

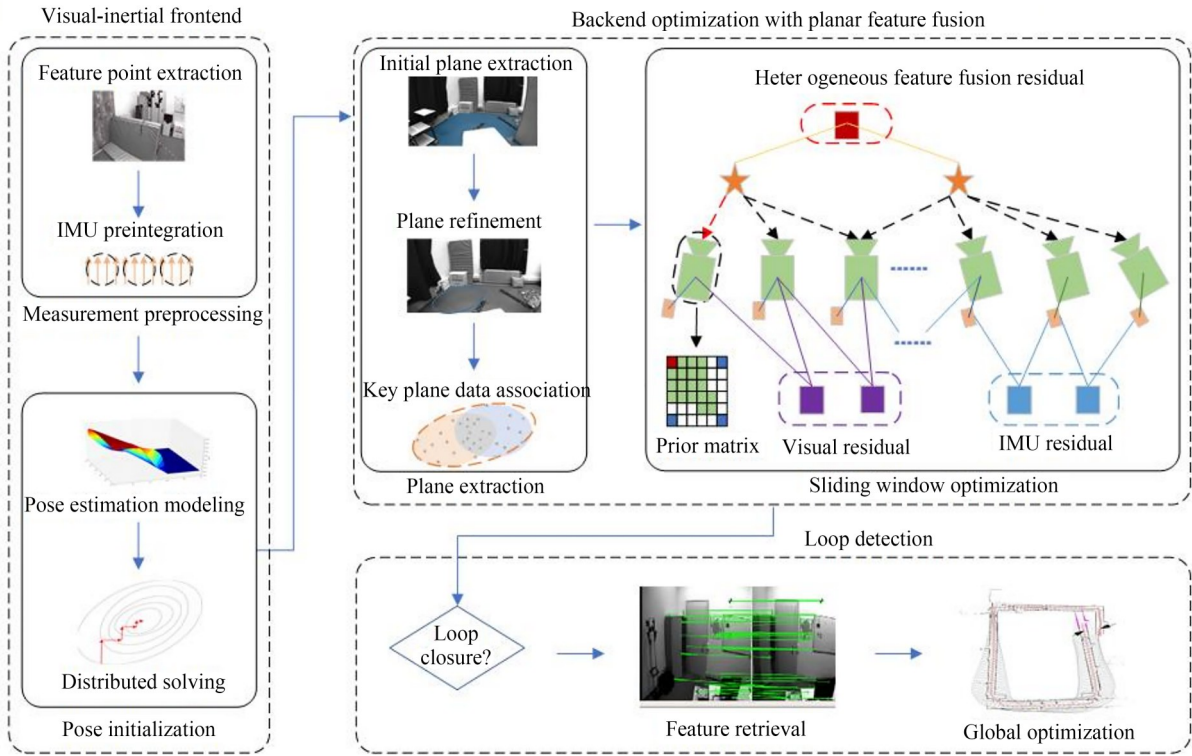


图1 系统框架

Fig. 1 System framework diagram

3 视觉惯性前端

3.1 测量值预处理

本文提取 ORB^[28] (Oriented FAST and Rotated BRIEF) 作为特征点, 并使用 KLT (Kanade-Lucas-Tomasi) 光流^[29] 对特征点进行跟踪。如图 2 所示, 在视觉预处理结束之后, 对光流跟踪得到的特征点集合计算德劳内三角剖分^[30]。

在德劳内三角剖分完成后需要进一步剔除畸形的三角网格如图 2(b), 而图 2(a) 假设图像尺寸为 $H \times W$, 三角形边长分别为: l_1, l_2, l_3 , 角度分

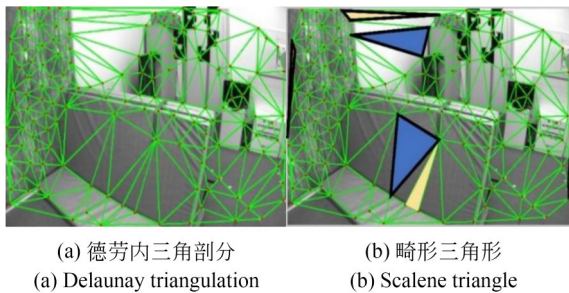
(a) 德劳内三角剖分
(a) Delaunay triangulation(b) 畸形三角形
(b) Scalene triangle

图2 单帧图像德劳内三角剖分结果

Fig. 2 Delaunay triangulation results for single-frame images

别为 $\theta_1, \theta_2, \theta_3$ 。畸形三角形边与角的关系满足:

$$\begin{aligned} \max(\theta_1, \theta_2, \theta_3) &\geq 150^\circ \\ \frac{\max(l_1, l_2, l_3)}{\min(l_1, l_2, l_3)} &\geq 5 \end{aligned} \quad (1)$$

$$\max(l_1, l_2, l_3) \geq 0.1 * \min(H, W)$$

考虑到图像与 IMU 的采样频率不一致, 因此将两者数据对齐是获取两个图像帧之间相对位姿的必要条件。对于时间间隔 $[t_k, t_{k+1}]$ 内的所有 IMU 数据, 通过式 (2) 和式 (3) 可得到第 k 个图像帧到 $k+1$ 个图像帧的相对运动:

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= \iint_{t \in [t_k, t_{k+1}]} R_t^{b_k} (\hat{a}_t - b_{a_t}) dt^2 \\ \beta_{b_{k+1}}^{b_k} &= \int_{t \in [t_k, t_{k+1}]} R_t^{b_k} (\hat{a}_t - b_{a_t}) dt \end{aligned} \quad (2)$$

$$\begin{aligned} \gamma_{b_{k+1}}^{b_k} &= \int_{t \in [t_k, t_{k+1}]} \frac{1}{2} \Omega(\hat{\omega}_t - b_{\omega_t}) \gamma_t^{b_k} dt \\ \Omega(w) &= \begin{bmatrix} -[w]_{\times} & w \\ w^T & 0 \end{bmatrix} \\ [w]_{\times} &= \begin{bmatrix} 0 & -w_z & w_y \\ w_z & 0 & -w_x \\ -w_y & w_x & 0 \end{bmatrix} \end{aligned} \quad (3)$$

其中: $R_t^{b_k}$ 表示 t 时刻的旋转, $\alpha_{b_{k+1}}^{b_k}, \beta_{b_{k+1}}^{b_k}, \gamma_{b_{k+1}}^{b_k}$ 分别表示从 t_k 时刻到 t_{k+1} 时刻的相对平移、速度、旋转, $\Omega(\hat{\omega}_t - b_{w_t})$ 表示角速度的扰动, b_{a_t} 表示 t 时刻的加速度偏差, b_{w_t} 表示 t 时刻的陀螺仪偏差,它们被定义为随机游走偏差,其一阶导数符合高斯白噪声。

3.2 位姿初始化

3.2.1 位姿初始化建模

初始化阶段需要求解的其他状态量包括速度、重力和真实尺度。具体如式(4)所示:

$$X = [v_{b_0}, v_{b_1}, \dots, v_{b_n}, g^{c_0}, s], \quad (4)$$

其中: v_{b_n} 表示IMU坐标系下的速度, g^{c_0} 表示第0帧(参考帧)相机坐标系下的重力向量, s 表示尺度。根据预积分的离散形式,再根据旋转矩阵在不同坐标系下的变换,有如下的关系:

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= R_{c_0}^{b_k} [R_{b_k}^c (p_{b_{k+1}}^w - p_{b_k}^w) - v_{b_k}^{c_0} \Delta t_k + \frac{1}{2} g^{c_0} \Delta t_k^2], \\ \beta_{b_{k+1}}^{b_k} &= R_{c_0}^{b_k} (v_{b_{k+1}}^{c_0} - v_{b_k}^{c_0} + g^{c_0} \Delta t_k) \end{aligned} \quad (5)$$

其中: $\alpha_{b_{k+1}}^{b_k}, \beta_{b_{k+1}}^{b_k}$ 分别表示从 t_k 时刻到 t_{k+1} 时刻的相对平移、速度。因此,预积分量的线性测量模型如式(6)所示:

$$\begin{aligned} \hat{z}_{b_{k+1}}^{b_k} &= \begin{bmatrix} \hat{\alpha}_{b_{k+1}}^{b_k} - p_c^b + R_{c_0}^{b_k} R_{b_{k+1}}^{c_0} p_c^b \\ \hat{\beta}_{b_{k+1}}^{b_k} \end{bmatrix} \\ H_{b_{k+1}}^{b_k} &= \begin{bmatrix} -I \Delta t_k & 0 & \frac{1}{2} R_{c_0}^{b_k} \Delta t_k^2 & R_{c_0}^{b_k} (\bar{p}_{c_{k+1}}^{c_0} - \bar{p}_{c_k}^{c_0}) \\ -I & R_{c_0}^{b_k} R_{b_{k+1}}^{c_0} & R_{c_0}^{b_k} \Delta t_k & 0 \end{bmatrix}. \end{aligned} \quad (6)$$

通过求解如式(7)所示的优化问题,可求出相应的状态量。

$$\min_X \sum \| \hat{z}_{b_{k+1}}^{b_k} - H_{b_{k+1}}^{b_k} X \|. \quad (7)$$

3.2.2 分布式求解

交替方向乘子法是一种解决可分解凸优化问题的高效方法。其基本思想是通过引入拉格朗日乘子来满足约束条件,并将原问题转化为一组无约束条件的子问题。初始化过程包括几十帧,每一帧都可以构建如式(7)所示的约束关系,因此可以分布求解初始化问题。结合式(7),目标函数可以改写为:

$$\begin{aligned} \arg \min_X & \frac{1}{2} \| H_{b_{k+1}}^{b_k} X - \hat{z}_{b_{k+1}}^{b_k} \|_2^2 + \mu \| z \|_1, \\ \text{s.t.} & X - z = 0 \end{aligned} \quad (8)$$

其增广拉格朗日函数为:

$$\begin{aligned} L_\rho(X, z, u) &= \frac{1}{2} \| H_{b_{k+1}}^{b_k} X - \hat{z}_{b_{k+1}}^{b_k} \|_2^2 + \\ & \mu \| z \|_1 - \frac{\rho}{2} \| u \|^2 + \frac{\rho}{2} (r + u)^2. \end{aligned} \quad (9)$$

第 $k+1$ 次迭代可得:

$$\begin{aligned} X^{k+1} &= \arg \min_X \left(\frac{1}{2} \| H_{b_{k+1}}^{b_k} X - \hat{z}_{b_{k+1}}^{b_k} \|_2^2 + \frac{\rho}{2} \| X - z^k + u^k \|_2^2 \right) \\ z^{k+1} &= \arg \min_z \left(\frac{\rho}{2} \| X^{k+1} - z + u^k \|_2^2 + \mu \| z \|_1 \right) \\ u^{k+1} &= u^k + X^{k+1} - z^{k+1}, \end{aligned} \quad (10)$$

其中: X^{k+1} 的更新包含了与 X 的相关项,并且 $\| H_{b_{k+1}}^{b_k} X - \hat{z}_{b_{k+1}}^{b_k} \|_2^2$ 是凸函数,将其对于 X 的导数写成矩阵形式,因此可求得 X 的更新方式:

$$\begin{aligned} X^{k+1} &= \left((H')^T H' + \rho I \right)^{-1} \left((H')^T \hat{z}_{b_{k+1}}^{b_k} + \rho (z^k - u^k) \right), \\ H' &= H_{b_{k+1}}^{b_k} \end{aligned} \quad (11)$$

同理,可求得 z^{k+1} 的更新:

$$\begin{aligned} z^{k+1} &= S_{\mu/\rho}(u^k + X^{k+1}) \\ S_{\mu/\rho}(a) &= \begin{cases} a - \frac{\mu}{\rho}, & a > \frac{\mu}{\rho} \\ 0, & |a| \leq \frac{\mu}{\rho} \\ a + \frac{\mu}{\rho}, & a < -\frac{\mu}{\rho} \end{cases}. \end{aligned} \quad (12)$$

4 融合平面特征的视觉惯性里程计

4.1 平面提取

4.1.1 初始平面提取

初始化过程得到的相机尺度使得德劳内三角形进一步投影为真实世界的三角网格。根据式(13)求得的夹角筛选出近似于水平面或垂直面的三角网格:

$$\theta_g = \arccos \left(\frac{\dot{g} n_\pi}{\| \dot{g} \| \| n_\pi \|} \right), \quad (13)$$

其中: $\dot{g}$ 表示重力方向的单位向量, n_π 表示三角网格的法线, θ_g 表示法线与重力方向的夹角。其次,构建关于三角网格顶点高度的直方图。最后,提取直方图中的局部最大值。

将顶点数量的局部最大值作为新的平面高度,重新计算网格顶点到平面的距离。根据点到平面距离的不同,将三角网格的顶点分为两类:点到平面的距离小于阈值 d_M 的集合 M_1^P 和距离大于阈值 d_M 的集合 M_2^P , d_M 设为0.1m。初始平面 M 参数可由式(14)计算得到:

$$d_\pi = \frac{\sum_{P \in M_1^P} d_P}{|M_1^P|}, n_\pi = \begin{cases} n_\pi, & |M_1^P| < 3 \\ n_{new}, & |M_1^P| \geq 3 \end{cases}, \quad (14)$$

其中: d_P 表示路标点 P 到初始平面 M 的距离, $|M_1^P|$ 表示集合 M_1^P 的势。通过式(15)将平面参数尺度统一,即可得到平面 M 的初始参数。

$$\|\ell\| = \sqrt{n_x^2 + n_y^2 + n_z^2}, n_{new} = \frac{1}{\|\ell\|} (n_x, n_y, n_z). \quad (15)$$

4.1.2 平面细化

上述平面提取方法的缺点在于假设所提取到的平面是无限平面,但结构化场景下的真实平面具有确定边界。为了缓解平面与特征点的错误关联,需要从初始平面 M 中提取有界平面 M_{limit} 。首先重新构建包含平面 M 的直角坐标系 XoY 。并将 M_1^P 中的所有点都投影到坐标系 XoY 中。由于人造结构场景下的平面大多以矩形或圆形为主,因此构造椭圆边界 m 进行有界平面拟合。对于点集 (x_i, y_i) ,多项式 F_E 表示点 (x, y) 到给定圆锥曲线的距离,将其建模为有约束的最优化问题:

$$\min F_E = \min \sum_{(x_i, y_i)} \left(ax_i^2 + bx_i y_i + cy_i^2 + \right. \\ \left. dx_i + ey_i + f \right)^2 \quad (16)$$

s. t. $b^2 - 4ac < 0, (x_i, y_i) \in M_{limit}$.

定义系数向量 $A = [a, b, c, d, e, f]^T$,参数向量 $N_i = [x_i^2, x_i y_i, y_i^2, x_i, y_i, 1]$ 。令 $D = [N_1, N_2, \dots, N_j]$,可改写为:

$$\min \|DA\|^2 = \min A^T D^T DA \text{ s. t. } A^T C' A = 1. \quad (17)$$

构造拉格朗日函数,令 $D^T D = S$,并适当变形得:

$$\min \|DA\|^2 = \min A^T S A = \min A^T \mu C' A = \mu. \quad (18)$$

为了得到最优解,只需找出最大特征值 μ^{-1} 对应的特征向量并将约束条件 $A^T C' A = 1$ 缩放回原尺度。接着计算当前帧上所有路标点在该平

面的距离。对于距离小于 d_M 的路标点,将其放入集合中。通过如式(19)所示的平面碰撞概率模型判断该点是否处于有界平面:

$$P(\tilde{P}_M \in M_{limit}) \propto P(|d_M^P| < d_M) P(\tilde{P}_M \in m), \quad (19)$$

其中: $P(\tilde{P}_M \in M_{limit})$ 表示路标点 P_M 落在有界平面 M_{limit} 的概率。 $P(|d_M^P| < d_M)$ 表示点 P_M 到平面 M 的距离 $|d_M^P|$ 小于 d_M 的概率。 $P(\tilde{P}_M \in m)$ 表示点 P_M 的投影点 $\tilde{P}_M$ 在有界平面 m 内的概率。本方法中当投影点 $\tilde{P}_M$ 处于椭圆边界外并且到椭圆边界的最短距离 d_m^P 小于阈值 d_m ,该投影点仍有概率处于椭圆边界内。

如图3所示,浅绿色点表示落在椭圆 m 内的情况;蓝色点表示椭圆外的点到椭圆边界的最短距离小于阈值;黄色点表示椭圆外的点到椭圆的最短距离大于阈值(彩图见期刊电子版)。 $P(\tilde{P}_M \in m)$ 定义如式(20)所示:

$$P(\tilde{P}_M \in m) = \begin{cases} 1, & \tilde{P}_M \in m \\ P(|d_m^P| < d_m), & \tilde{P}_M \notin m \end{cases}. \quad (20)$$

平面由多个特征点关联,特征点的位置也受噪声的影响,因此, d_M^P, d_m^P 定义如下:

$$d_M^P = \hat{d}_M^P + n_M, d_m^P = \hat{d}_m^P + n_m. \quad (21)$$

$\hat{d}_M^P$ 表示距离的测量值,结合式(20)和式(21)可得:

$$P(|d_m^P| < d_m) = \int_{|d_m^P| - d_m}^{|d_m^P| + d_m} f(x) dx \\ f(x) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-\frac{(x - d_m^P)^2}{2\sigma_m^2}\right). \quad (22)$$

同理,可求得 $P(|d_M^P| < d_M)$ 。

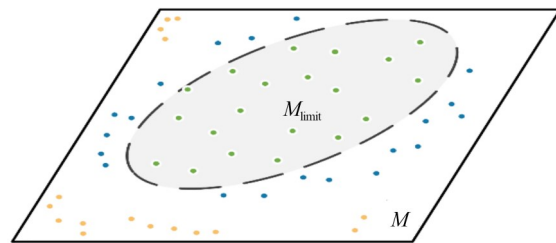


图3 投影点与有界平面关系

Fig. 3 Relationship between projected points and bounded planes

4.1.3 关键平面数据关联

平面数据关联包括两个阶段。第一阶段比较两平面中相同路标点的数量,以实现单帧内的数据融合。式(23)通过比较平面之间的

角度和距离,避免平面融合时假阳性结果的出现:

$$\arccos\|\mathbf{n}_2 - \mathbf{n}_1\| < \theta_{th}, |d_2 - d_1| < d_M, \quad (23)$$

其中: $\theta_{th}=0.98^\circ$ 。 $[\mathbf{n}_1, d_1]$ 、 $[\mathbf{n}_2, d_2]$ 分别为待融合平面1和待融合平面2的平面参数。在满足式(23)之后,如图4所示,将两个平面的边界以及它们对应的关联点投影到二维坐标系。

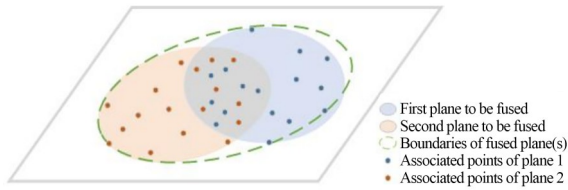


图4 平面融合示意图

Fig. 4 Planar fusion schematic

第二阶段多帧平面的数据关联。首先统计处于一个滑动窗口内的所有平面,并利用式(23)对数据进行关联与融合;每次数据关联之后仍需要检查这些平面是否已经存在于优化问题中,以避免冗余的优化问题。

为了保证后端优化的实时性,将滑动窗口内关联的路标点最多的两个水平面以及两个垂直面设为关键平面,后端优化时只构建关键平面的几何约束而忽略那些关联路标点较少的平面。

4.2 滑动窗口优化

本文结合边缘化的先验残差、IMU测量值残差、视觉重投影残差以及异质特征融合残差构建因子图模型并优化求解,因子图模型如图5所示。

滑动窗口中待优化的状态向量定义如式(24):

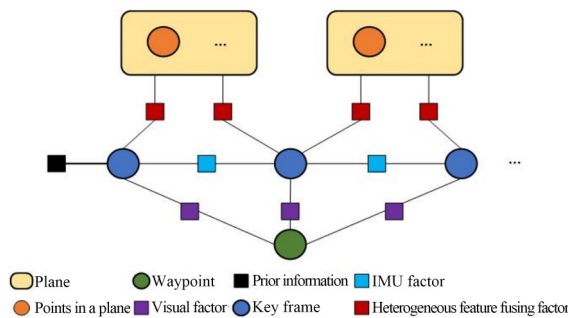


图5 融合平面特征的因子图模型

Fig. 5 Factor graph models incorporating planar features

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}_0, \dots, \mathbf{x}_n, \mathbf{x}_c^b, \lambda_0, \dots, \lambda_m, \pi_0, \dots, \pi_\varpi] \\ \mathbf{x}_k &= [\mathbf{p}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^w, b_a, b_g] \\ \mathbf{x}_c^b &= [\mathbf{p}_c^b, \mathbf{q}_c^b], \pi_\eta = [\mathbf{n}_\eta, d_\eta] \end{aligned} \quad (24)$$

其中: $\mathbf{x}_k$ 表示第 k 帧图像出现时对应的IMU状态,即 t_k 时刻世界坐标系下的位置 $\mathbf{p}_{b_k}^w$ 、速度 $\mathbf{v}_{b_k}^w$ 、旋转 $\mathbf{q}_{b_k}^w$,以及在机体坐标中的加速度偏差 b_a 与陀螺仪偏差 b_g 。 $\mathbf{x}_c^b$ 表示相机与IMU的外参。 n 表示滑动窗口中关键帧的数量, m 表示滑动窗口中路标点数量, λ 表示路标首次被观测到的图像帧对应的逆深度, ϖ 表示滑动窗口中平面特征的数量。

最大后验估计的求解可以转换为最小二乘优化问题,代价函数为:

$$\begin{aligned} r(\mathbf{X}) &= \|\mathbf{r}_{PF} - \mathbf{H}_{PF}\mathbf{X}\|^2 + \sum_{k \in B} \|\mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}, \mathbf{X})\|^2 + \\ &\quad \sum_{(l,j) \in C} \|\mathbf{r}_C(\hat{\mathbf{z}}_l^{C_j}, \mathbf{X})\|^2 + \sum_{(i,\eta) \in E} \|\mathbf{r}_E(\hat{\mathbf{z}}_\eta^{C_i}, \mathbf{X})\|^2, \end{aligned} \quad (25)$$

其中: $\{\mathbf{r}_{PF}, \mathbf{H}_{PF}\}$ 表示先验残差、 $\mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}, \mathbf{X})$ 表示IMU测量值残差、 $\mathbf{r}_C(\hat{\mathbf{z}}_l^{C_j}, \mathbf{X})$ 表示视觉重投影残差、 $\mathbf{r}_E(\hat{\mathbf{z}}_\eta^{C_i}, \mathbf{X})$ 表示异质特征融合残差。

为了适应如鱼眼相机等更多种类的相机,相机测量残差定义在如图6所示的单位球面上。考虑第一次在第 i 张图像中观察到的第 l 个特征,第 j 张图像中特征观察的残差定义为:

$$\begin{aligned} \mathbf{r}_C(\hat{\mathbf{z}}_l^{C_j}, \mathbf{X}) &= [b_1, b_2]^T \times \left(\bar{\mathbf{P}}_l^{C_j} - \frac{\mathbf{P}_l^{C_j}}{\|\rho_l^{C_j}\|} \right) \\ \mathbf{P}_l^{C_j} &= \mathbf{R}_b^c (\mathbf{R}_w^c (\mathbf{R}_{b_i}^w \left(\frac{\mathbf{R}_b^b \bar{\mathbf{P}}_l^{C_j}}{\lambda_l} + \mathbf{t}_c^b \right) + \mathbf{t}_{b_i}^w) - \mathbf{t}_w^b) - \mathbf{t}_b^c, \end{aligned} \quad (26)$$

其中: $\bar{\mathbf{P}}_l^{C_j}$ 表示第 l 个路标点在第 j 个归一化相机坐标系中观察到的坐标, $\mathbf{P}_l^{C_j}$ 表示第 l 个路标点在第 j 个归一化相机坐标系中的估计值, $[b_1, b_2]^T$ 为

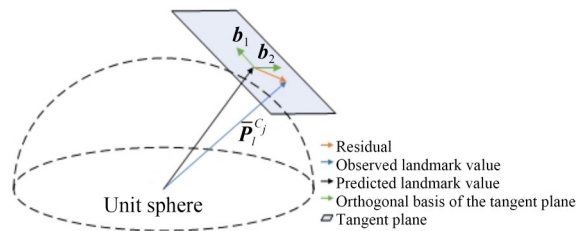


图6 单位球体视觉重投影残差

Fig. 6 Visual reprojection residual on a unit sphere

切平面上的任一正交基。

视觉重投影残差依赖于路标点的特征匹配,容易受到光照变化和纹理的影响。平面特征为 SLAM 提供了额外的几何约束,有助于在复杂环境中提高位姿估计的鲁棒性和准确性。为此,本文增加了如图 7 所示的异质特征融合残差,通过比较实际观测到的平面参数与估计出的平面参数之间的差异,提高弱纹理场景下平面估计的准确性和运动估计的稳定性。

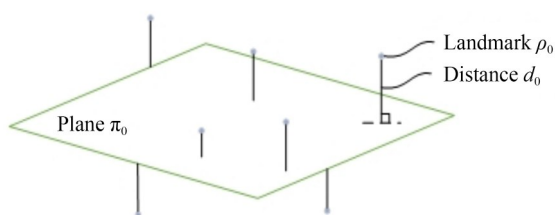


图 7 异质特征融合残差

Fig. 7 Heterogeneous feature fusion residuals

异质特征融合残差包含从每个关键平面的关联点到其余关键平面的距离之和。单个平面关联点 ρ 到关键平面的距离 $d_{\perp}$ 为:

$$d_{\perp} = \mathbf{n}^T \cdot \mathbf{P} + d. \quad (27)$$

$r_E(\hat{z}_{F_i}^{C_i}, \mathbf{X})$ 定义为:

$$r_E(\hat{z}_{F_i}^{C_i}, \mathbf{X}) = n_{\eta} [R_{b_i}^w \left(R_c^b \frac{1}{\lambda_i} \bar{P}_e^{C_i} + t_c^b \right) + t_{b_i}^w] + d_{\eta}. \quad (28)$$

5 回环检测

在单目视觉惯性里程计中使用滑动窗口和边缘化可以降低优化的计算复杂度,但也会导致系统误差的累积。为了有效消除累积误差,本文采用了一种与视觉惯性里程计模块紧密耦合的回环检测模块。该方法基于 BRIEF (Binary Robust Independent Elementary Features) 描述子^[31]的词袋模型^[32]检测回环。如图 8 所示,从图像中检测出的特征点和描述子被输入到具有树形结构的词袋模型中,得到词频直方图后计算当前帧与词袋的相似度分数。然后从关键帧数据库中匹配具有相似单词权重和描述子的候选帧。最后,进行回环一致性检测,以确定潜在的循环帧。回环一致性检测包含两个阶段。

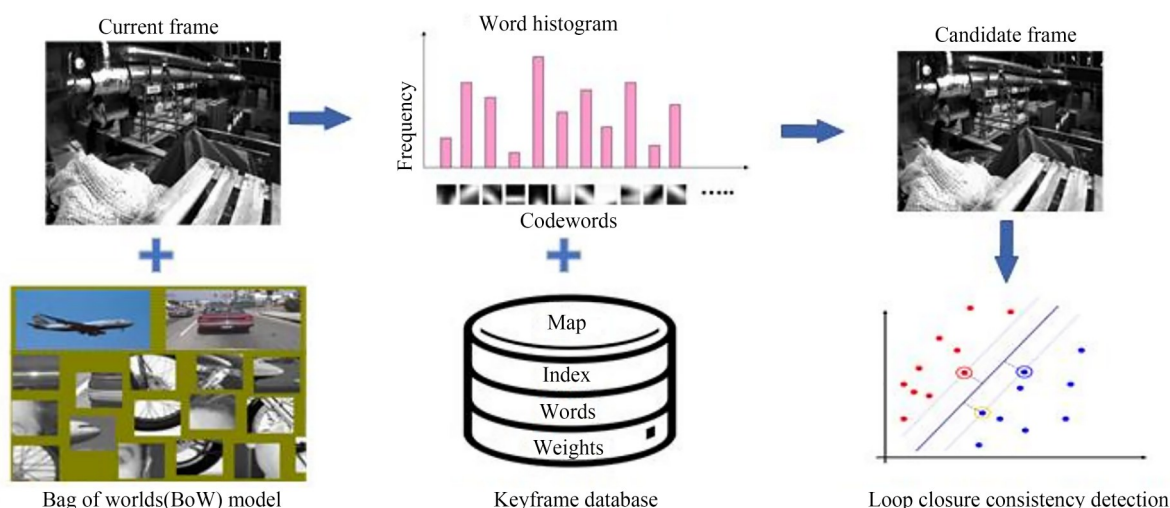


图 8 回环检测流程

Fig. 8 Loop closure detection Process

第一阶段通过基于基础矩阵的随机样本一致算法 (Random Sample Consensus, RANSAC) 消除当前帧与候选帧之间的特征匹配误差。

第二阶段结合滑动窗口优化获得的特征点三维坐标以及回环候选帧中特征点的二维坐标,

使用基于 PnP 的 RANSAC 算法进一步消除外点。在回环一致性检测后,候选帧与当前帧还存在足够多的匹配点,则被认为发生了回环,并通过高斯-牛顿法优化 (29) 式所示的全局位姿残差。

$$\min_{\mathbf{p}, \psi, \mathbf{n}} \sum_{(i,j) \in \varsigma} \|\mathbf{r}_{PL}(\mathbf{p}_{ij}^w, \psi_{ij})\|^2 + \sum_{(l,k) \in \tau} \|\mathbf{r}_P(\mathbf{n}_{l,k})\|^2 + \sum_{(l,k) \in \tau} h(\|\mathbf{r}_{PL}(\mathbf{p}_{lk}^w, \psi_{lk})\|)^2$$

$$\mathbf{r}_{PL}(\mathbf{p}_{ij}^w, \psi_{ij}) = \begin{bmatrix} R(\hat{\phi}_i, \hat{\theta}_i, \psi_i)^{-1}(\mathbf{p}_j^w - \mathbf{p}_i^w) - \hat{\mathbf{p}}_{ij}^i \\ \psi_j - \psi_i - \hat{\psi}_{ij} \end{bmatrix}, \quad (29)$$

其中: $\hat{\mathbf{p}}_{ij}^i$ 表示两帧之间的相对位姿, $h[\cdot]$ 表示 Huber 鲁棒损失函数, ς 表示关键帧数据库中的关键帧, τ 表示回环帧, $\hat{\psi}_{ij}$ 表示两帧之间的偏航角之差, $\mathbf{r}_p$ 表示回环帧与候选帧之间的平面位置关系, 其定义如下:

$$\mathbf{r}_p(\mathbf{n}_{l,k}) = \begin{cases} \|\mathbf{n}_l - \mathbf{n}_k\|, \pi_l \parallel \pi_k \\ \mathbf{n}_l \mathbf{n}_k, \pi_l \perp \pi_k \end{cases}, \quad (30)$$

其中, π_l 与 π_k 分别为回环帧和候选帧内的水平平面与垂直平面。

6 实 验

本文在 EuRoC 数据集^[33]以及 TUM-VI^[34]数据集验证本文提出方法的有效性, 此外还进行了消融实验测试不同模块的速率与效果。实验平台为 64 位 Linux 操作系统 Ubuntu18.04, CPU 为 Intel(R) Core TM i5-12600HQ@3.70 GHz, 内存为 16 GB。

6.1 初始化性能测试

视觉惯性 SLAM 初始化的性能指标通常包括以下几个方面: (1) 时间; (2) 精度; (3) 鲁棒性。常见的评估指标包括绝对轨迹误差 (Absolute

lute Trajectory Error, ATE) 与相对位姿误差 (Relative Pose Error, RPE)^[35]。理想的初始化方法应具备在不同场景和环境下的稳定性, 都能提供可靠的初始化结果。本节通过对比基线方法 Vins-Mono^[15]来测试本文提出方法的初始化性能。

具体实验设置如下, 将 Vins-Mono 的初始化求解修改为本文提出方法, 其他设置均保持与原方法相同, 记录从移动机器人开始运动到初始化完成所需的时间; 其次在初始化完成后, 挑选中间相同的 50 帧, 计算 ATE; 最后, 计算全局 ATE, 评估本文提出的初始化方法的有效性。除此之外, 为了进一步测试本文提出方法的鲁棒性, 在具有不同特性的数据集上进行了实验, 包括不同光照、不同纹理等复杂场景, 并计算了在不同条件下的 ATE。以此来评估本文提出方法的鲁棒性。

实验结果如表 1 所示, 本文提出的初始化方法能够在大部分序列中降低初始化所需的时间, 在局部绝对轨迹误差低于基线方法的序列, 全局轨迹误差也都有所降低。在少量序列, 尽管局部绝对轨迹误差较低, 但由于该序列本身绝对轨迹

表 1 EuRoC 数据集初始化性能对比结果

Tab. 1 Comparison results of initialization performance of EuRoC dataset

序列	本文提出方法			Vins-Mono ^[15]		
	时间	局部 ATE	全局 ATE	时间	局部 ATE	全局 ATE
MH01	10.67	0.13	0.12	9.44	0.15	0.12
MH02	11.32	0.09	0.11	11.58	0.11	0.12
MH03	10.75	0.11	0.12	12.16	0.11	0.13
MH04	9.88	0.16	0.18	10.81	0.16	0.18
MH05	13.23	0.21	0.19	14.52	0.24	0.21
V1_01	8.76	0.07	0.07	8.58	0.07	0.07
V1_02	9.78	0.09	0.08	9.47	0.11	0.08
V1_03	11.24	0.12	0.17	12.46	0.14	0.19
V2_01	9.06	0.06	0.08	9.24	0.06	0.08
V2_02	9.52	0.11	0.14	10.79	0.11	0.16
V2_03	12.31	0.14	0.22	13.23	0.15	0.22

注: 粗体表示效果最好

误差较低,所以会出现效果提升不明显或者是没有提升的情况。关于鲁棒性的测试如图 9 所示,在 EuRoC 的 11 个序列中分别运行了 6 次本文提出方法和基线方法 Vins-Mono。其中在 11 个序列中本文方法在 7 个序列上的 ATE 平均值要优于 Vins-Mono,4 个序列上的 ATE 平均值大致相等。其中在 MH05 序列上表现最好,本文方法的 ATE 平均值低 0.3 m。并且在 11 个序列中本文方法在 8 个序列上的 ATE 方差要优于 Vins-Mono,3 个序列上的 ATE 方差大致相等。总体来说,本文提出的初始化方法精度更高,所需时间更短,鲁棒性更好。

6.2 整体性能测试

为了保证实验的客观全面,将对比方法分为两类:基于视觉惯性的 VINS-Mono^[15],BASALT^[18]以及基于平面特征的 Kimera^[9],structure PLP^[24]。

表 2 为本文提出方法与其他 SLAM 方法在 EuRoC 数据集上 ATE 的对比结果。在所有方法都开启回环检测的前提下,基于视觉惯性的方法中,本文提出方法在全部 11 个序列中有 8 个序列表现最好,3 个序列表现第二好,误差平均值也最小。MH04 与 MH05 序列所在场景为工厂环境,

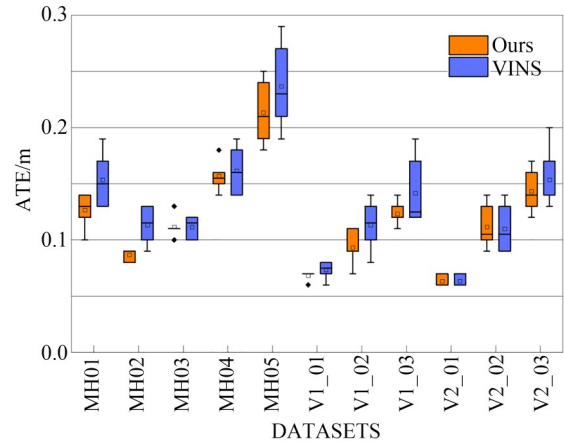


图 9 EuRoC 数据集鲁棒性对比结果

Fig. 9 Robustness comparison results of the EuRoC dataset

导致在 MH05 序列精度处于第二位。序列 V1_01, V1_02 中场景纹理丰富,搭载相机的无人机没有剧烈旋转,列出的所有方法均能实现较好的位姿估计。在序列 V2_02 中,由于相机运动剧烈,导致该约束并不够鲁棒,但精度仍与在该序列表现最好的方法相差不大。BASALT 在序列 V2_03 中,有超过 400 个缺失帧,因此无法提供定位误差。本文提出的基于平面特征的优化方法能够降低累积误差。在基于平面的方法中,本文

表 2 不同 SLAM 方法在 EuRoC 下的 ATE(m) 对比结果

Tab. 2 Comparison results of ATE(m) under EuRoC for different SLAM methods

序列	VINS-Mono ^[15]		BASALT ^[18]		Kimera ^[9]		Structure PLP ^[24]		本文提出方法	
	w/	w/o	w/	w/o	w/	w/o	w/	w/o	w/	w/o
MH01	0.12	0.15	<u>0.07</u>	<u>0.08</u>	<u>0.07</u>	<u>0.11</u>	N/A	N/A	0.06	0.07
MH02	0.12	0.15	<u>0.06</u>	0.08	0.07	0.10	N/A	N/A	0.05	0.08
MH03	0.13	0.22	<u>0.07</u>	<u>0.09</u>	0.11	0.16	N/A	N/A	0.06	0.07
MH04	0.18	0.32	<u>0.13</u>	<u>0.16</u>	0.15	0.24	N/A	N/A	0.09	0.12
MH05	0.21	0.30	0.11	0.15	0.24	0.35	N/A	N/A	<u>0.13</u>	<u>0.17</u>
V1_01	0.07	0.08	0.04	<u>0.06</u>	<u>0.05</u>	0.05	0.10	0.17	0.04	0.07
V1_02	0.08	0.11	<u>0.05</u>	0.07	0.11	0.07	0.10	0.15	0.04	<u>0.08</u>
V1_03	0.19	0.18	<u>0.10</u>	<u>0.13</u>	0.12	0.07	0.14	0.23	0.08	<u>0.13</u>
V2_01	0.08	<u>0.08</u>	0.04	0.10	0.07	<u>0.08</u>	0.12	0.19	<u>0.05</u>	0.07
V2_02	0.16	0.16	0.05	0.07	0.10	<u>0.10</u>	0.14	0.24	<u>0.07</u>	<u>0.10</u>
V2_03	0.22	0.27	N/A	N/A	<u>0.19</u>	0.21	0.14	0.29	0.14	<u>0.23</u>
平均值	0.14	0.18	N/A	N/A	<u>0.13</u>	<u>0.14</u>	N/A	N/A	0.07	0.11

注:粗体表示效果最好,下划线表示效果第二好。w 表示开启回环检测, w/o 表示关闭回环检测。

提出方法在全部 11 个序列中均为最好,误差平均值也最小。structure PLP 是基于 CNN 提取平面的双目 SLAM 系统,由于 CNN 并没有在序列 MH01-MH05 提取到平面,因而运行失败。本文

提出方法加入了有界平面拟合,有效降低了平面与特征点关联时的误匹配,提升了在无明显几何特征场景下的效果。尤其是在 MH01-MH04 序列中,误差比其他方案低超过 50%。

表 3 不同方法在 EuRoC 的 RPE 对比结果

Tab. 3 Comparative RPE results of different methods in EuRoC

序列	VINS-Mono ^[15]		BASALT ^[18]		本文提出方法	
	RPE _t /m	RPE _t /([°])·100 ⁻¹	RPE _t /m	RPE _t /([°])·100 ⁻¹	RPE _t /m	RPE _t /([°])·100 ⁻¹
MH01	<u>0.04</u>	1.5	0.03	<u>1.2</u>	0.03	0.8
MH02	0.06	0.9	<u>0.03</u>	<u>0.9</u>	0.02	0.7
MH03	0.04	0.9	0.04	<u>0.8</u>	0.04	0.7
MH04	0.05	1.3	<u>0.03</u>	<u>0.8</u>	0.02	0.5
MH05	0.05	<u>1.3</u>	0.03	0.7	<u>0.05</u>	0.7
V1_01	<u>0.1</u>	4	0.06	3.1	0.06	<u>3.4</u>
V1_02	<u>0.05</u>	0.9	0.03	<u>0.7</u>	0.03	0.5
V1_03	0.05	1.7	<u>0.04</u>	<u>0.8</u>	0.03	0.7
V2_01	<u>0.04</u>	1.7	0.02	1.0	<u>0.04</u>	<u>1.3</u>
V2_02	0.04	1.3	0.02	1.0	<u>0.03</u>	<u>1.2</u>
V2_03	0.05	<u>2.3</u>	N/A	N/A	0.05	1.7
平均值	<u>0.05</u>	<u>1.6</u>	N/A	N/A	0.04	1.1

注:粗体表示效果最好,下划线表示效果第二好。所有方法默认开启回环检测。表 3 不同方法在 EuRoC 的 RPE 对比结果

表 4 不同方法在 TUM-VI 下的 ATE(m) 对比结果

Tab. 4 ATE(m) TUM-VI comparison results

序列	VINS-Mono ^[15]	BASALT ^[18]	本文方法
Room1	<u>0.07</u>	0.09	0.04
Room2	<u>0.07</u>	<u>0.07</u>	0.04
Room3	0.11	<u>0.13</u>	<u>0.13</u>
Room4	<u>0.04</u>	0.05	0.02
Room5	0.2	<u>0.13</u>	0.09
Room6	0.08	0.02	<u>0.06</u>
平均值	0.10	<u>0.08</u>	0.06

注:粗体表示效果最好,下划线表示效果第二好。

表 3 进一步比较了表 2 中精度最高的两种方法在 EuRoC 数据集上的相对位姿误差,RPE 的评估取决于时间增量的设置,在实验中固定为 1 s。如表 3 所示,本文提出方法在

22 项统计数据中 16 项表现最好,6 项表现第二好。

表 4 为本文提出方法在 TUM-VI 数据集下的表现。在 Room1 到 R. 021/oom6 序列中,本文提出方法的误差都比较好。总体看来,由于测试的 6 个序列时间较短、运动并不剧烈,导致所有方法在减少累积误差的同时获得较好的估计值。

图 10 展示了在 EuRoC 中绝对轨迹误差随时间的变化情况。其中每个序列的图像左侧为全局轨迹图,右侧图像为本文提出方法与 VINS-Mono 在不同时刻的轨迹误差比较。通过图 10 可以发现,本文所构建的视觉惯性 SLAM 系统估计的移动机器人轨迹和数据集提供的真实轨迹大致相同。因此证明本文构建的 SLAM 系统在该数据集上整体表现良好,能够准确估计移动机器人运动轨迹。

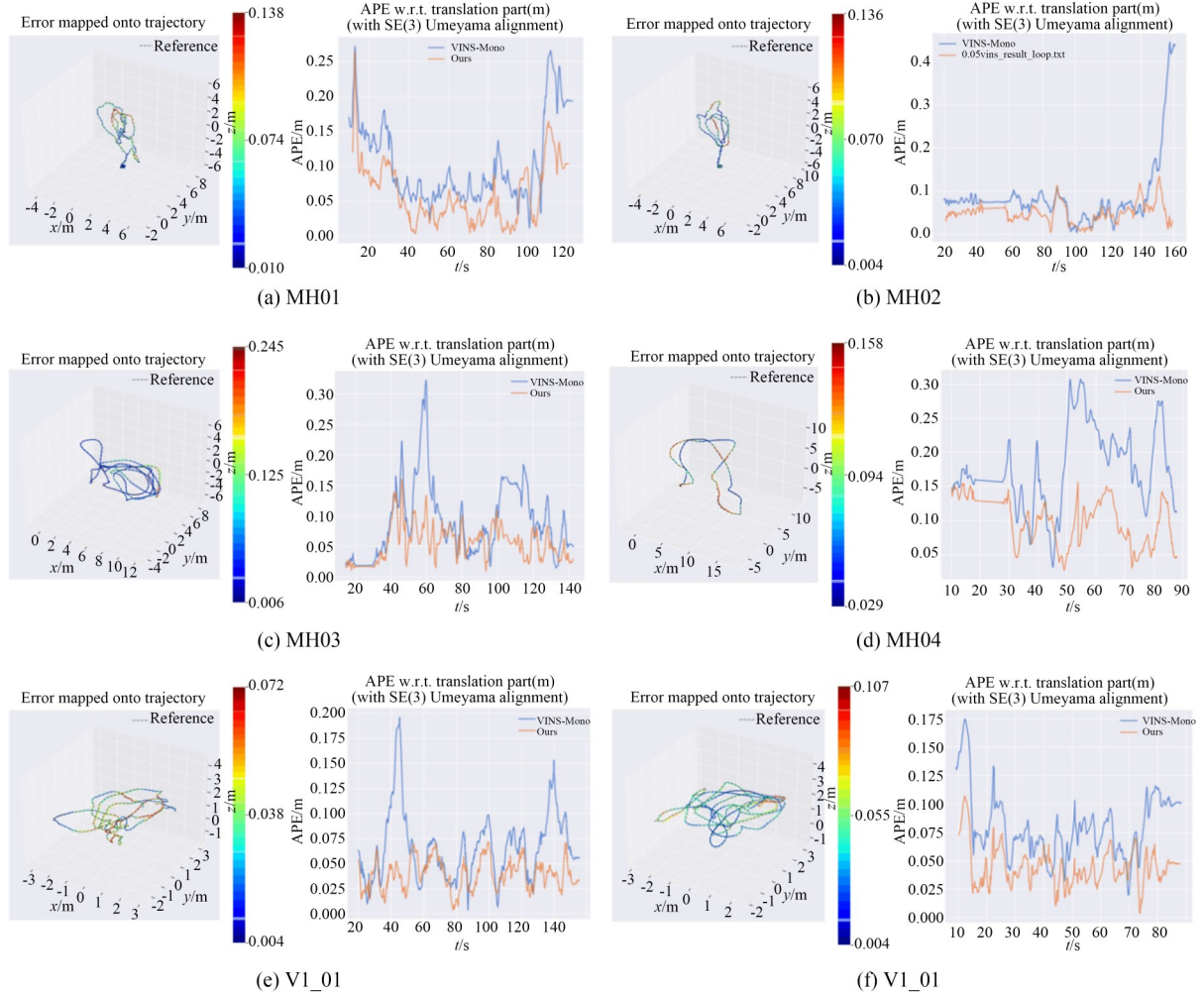


图 10 EuRoC 中本文提出方法的部分轨迹(左列)与轨迹误差图(右列)

Fig. 10 Partial trajectories of the proposed method in this paper (left column) with trajectory error plots (right column)

6.3 消融实验

本节将基线方法 VINS-Mono 定义为 VINS, 平面约束定义为 P, 具体结果如表 5 所示。表中 Ours 即为本文提出方法。通过比较 VINS 与 VINS+P 可知, 将异质特征融合残差代入优化有效降低了整体系统的误差累积, 但 VINS+P 在序列 MH05 中的结果表明, 当场景中的结构特征不够明显时, 直接提取平面特征会导致精度降低。本文提出方法的结果表明, 有界平面拟合的方法能够有效减少路标点与平面错误关联造成的误差, 因此在每一个序列中都达到了最优的结果。

6.4 运算复杂度测试

表 6 显示了本文提出方法在每个模块的平均运行时间。其中, 移动机器人位姿估计模块和滑

表 5 ATE 消融实验

Tab. 5 Ablation experiments on ATE

序列	VINS	VINS+P	Ours
MH01	0.12	<u>0.09</u>	0.06
MH02	0.12	<u>0.09</u>	0.05
MH03	0.13	<u>0.07</u>	0.06
MH04	0.18	<u>0.14</u>	0.09
MH05	<u>0.21</u>	0.28	0.13
V1_01	<u>0.07</u>	0.04	0.04
V1_02	0.08	<u>0.06</u>	0.04
V1_03	0.19	<u>0.18</u>	0.09
V2_01	0.08	<u>0.06</u>	0.05
V2_02	0.16	<u>0.1</u>	0.07
V2_03	<u>0.22</u>	0.25	0.14
平均值	0.14	<u>0.11</u>	0.07

注: 粗体表示效果最好, 下划线表示效果第二好。

表6 本文提出方法各模块的运行时间

Tab. 6 Running time of each module of the method proposed

模块	时间/ms
视觉惯性前端	33.1
融合平面特征的后端优化	20.5
回环检测	48.7

动窗口优化模块每个关键帧执行一次,全局位姿优化模块只在发生回环后执行,例如在V2_01序列,共发生8次回环,平均每次时间约为48.7 ms。在引入并行化线程之后,即为每个模块分配独立线程。本文提出方法的平均进行单帧定位所需的运行时间为35.7 ms,能达到较好的实时性。

为了进一步比较后端优化的实时性,每隔10个关键帧记录一次VINS, VINS+P, 本文提出方法在滑动窗口优化模块的运行时间,分别记作 t_{VINS} , t_{VINS}^P , t_{Ours} 。

如图11所示,随着关键帧数量的增长, t_{VINS} 在三类方法中总体是最低的。这是因为引入平面约束在提升精度同时,需要消耗额外的计算资源。由于有界平面拟合限制了平面与路标点的关联规模,使得总体上 $t_{VINS}^P > t_{Ours}$,而且 t_{Ours} 仅比最快的 t_{VINS} 高约5 ms。

7 结 论

本文提出了一种基于关键平面的异质特征融合SLAM系统。该系统能够在结构化场景中稳定、连续地工作。首先将传感器传入的数据进行预处理,通过ADMM算法分布式求解初始化过程。其次提取结构化场景下的平面参数,为了使该方法能够真实反映场景下的平面,利用约束条件提取有界平面参数,并将其纳入紧密耦合的

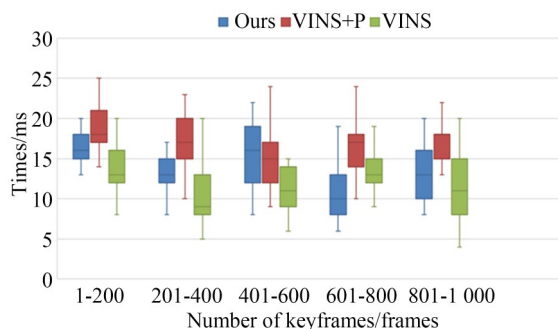


图11 滑动窗口优化模块运行时间比较

Fig. 11 Sliding window optimization module runtime comparison

非线性优化。最后,评估了本文所提方法在EuRoC数据集以及TUM-VI数据集中的精度,与基于视觉惯性的SLAM、基于平面的SLAM相比,本文所提方法在EuRoC数据集11个序列中,有8个序列精度最高,3个序列精度第二高。

由于默认的环境是静态场景,在动态环境下系统将无法判断是场景中的物体在运动还是自身在运动。如果不能有效滤除动态因子的影响,将会对整体系统的精度造成较大影响。此外在户外、大规模环境下的实验效果尚不理想,这也是本文提出方法面临的一个挑战。因此,在接下来的工作中,将针对更为复杂的室外动态环境,积极探索结构化场景下的语义地图构建方法。

作者贡献说明:

沈晔湖:提供资源,论文构思,验证,审核与编辑写作;

何一凡:获取资助,方法论,项目管理,研究调查;

魏季坤:验证,可视化呈现,数据管理;

张大庆:初稿写作,形式分析。

参考文献:

- [1] WEN T C, FANG Y C, LU B, *et al.* LIVER: a tightly coupled LiDAR-inertial-visual state estimator with high robustness for underground environments [J]. *IEEE Robotics and Automation Letters*, 2024, 9(3): 2399-2406.
- [2] LEE J, KOMATSU R, SHINOZAKI M, *et al.* Switch-SLAM: switching-based LiDAR-inertial-vi-

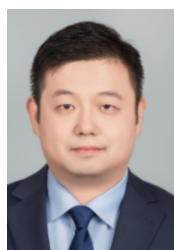
sual SLAM for degenerate environments[J]. *IEEE Robotics and Automation Letters*, 2024, 9(8): 7270-7277.

- [3] HU C, ZHU S Q, LIANG Y M, *et al.* Tightly-coupled visual-inertial-pressure fusion using forward and backward IMU preintegration[J]. *IEEE Robotics and Automation Letters*, 2022, 7(3): 6790-6797.

- [4] JUNG J H, CHOE Y, PARK C G. Photometric visual-inertial navigation with uncertainty-aware ensembles [J]. *IEEE Transactions on Robotics*, 2022, 38(4): 2039-2052.
- [5] KALMAN R E. A new approach to linear filtering and prediction problems [J]. *Journal of Basic Engineering*, 1960, 82(1): 35-45.
- [6] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint kalman filter for vision-aided inertial navigation [C]. *Proceedings 2007 IEEE International Conference on Robotics and Automation. April 10-14, 2007, Rome, Italy*. IEEE, 2007: 3565-3572.
- [7] ZHANG Z C, SCARAMUZZA D. A tutorial on quantitative trajectory evaluation for visual (inertial) odometry [C]. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). October 1-5, 2018, Madrid, Spain*. IEEE, 2018: 7244-7251.
- [8] LIU T B, SHEN S J. High altitude monocular visual-inertial state estimation: initialization and sensor fusion [C]. *2017 IEEE International Conference on Robotics and Automation (ICRA). May 29 - June 3, 2017, Singapore*. IEEE, 2017: 4544-4551.
- [9] ROSINOL A, VIOLETTE A, ABATE M, et al. Kimera: from SLAM to spatial perception with 3D dynamic scene graphs [J]. *International Journal of Robotics Research*, 2021, 40(12): 1510-1546.
- [10] LV J J, LANG X L, XU J H, et al. Continuous-time fixed-lag smoothing for LiDAR-inertial-camera SLAM [J]. *IEEE/ASME Transactions on Mechatronics*, 2023, 28(4): 2259-2270.
- [11] LEUTENEGGER S, FURGALE P, RABAUD V, et al. Keyframe-based visual-inertial SLAM using nonlinear optimization [C]. *Robotics: Science and Systems IX. Robotics: Science and Systems Foundation*, 2013.
- [12] PATRON-PEREZ A, LOVEGROVE S, SIBLEY G. A spline-based trajectory representation for sensor fusion and rolling shutter cameras [J]. *International Journal of Computer Vision*, 2015, 113(3): 208-219.
- [13] VON STUMBERG L, CREMERS D. DM-VIO: delayed marginalization visual-inertial odometry [J]. *IEEE Robotics and Automation Letters*, 2022, 7(2): 1408-1415.
- [14] XU L, YIN H S, SHI T, et al. EPLF-VINS: real-time monocular visual-inertial SLAM with efficient point-line flow features [J]. *IEEE Robotics and Automation Letters*, 2023, 8(2): 752-759.
- [15] QIN T, LI P L, SHEN S J. VINS-mono: a robust and versatile monocular visual-inertial state estimator [J]. *IEEE Transactions on Robotics*, 2018, 34(4): 1004-1020.
- [16] LISO L, SANDSTRÖM E, YUGAY V, et al. Loopy-SLAM: dense neural SLAM with loop closures [C]. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 16-22, 2024, Seattle, WA, USA*. IEEE, 2024: 20363-20373.
- [17] ZHANG Y M, TOSI F, MATTOCCIA S, et al. GO-SLAM: Global optimization for consistent 3D instant reconstruction [C]. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023: 3704-3714.
- [18] USENKO V, DEMMEL N, SCHUBERT D, et al. Visual-inertial mapping with non-linear factor recovery [J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 422-429.
- [19] CHEN C H, WANG B, LU C X, et al. Deep learning for visual localization and mapping: a survey [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(12): 17000-17020.
- [20] RAMBACH J R, TEWARI A, PAGANI A, et al. Learning to fuse: a deep learning approach to visual-inertial camera pose estimation [C]. *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). September 19-23, 2016, Merida, Mexico*. IEEE, 2016: 71-76.
- [21] CLARK R, WANG S, WEN H K, et al. VINet: visual-inertial odometry as a sequence-to-sequence learning problem [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, 31(1): 3995-4001.
- [22] PENG X F, LIU Z H, LI W M, et al. DVI-SLAM: A dual visual inertial SLAM network [C]. *2024 IEEE International Conference on Robotics and Automation (ICRA). May 13-17, 2024, Yokohama, Japan*. IEEE, 2024: 12020-12026.
- [23] CONCHA A, CIVERA J. DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence [C]. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*

- (IROS). September 28-October 2, 2015, Hamburg, Germany. IEEE, 2015: 5686-5693.
- [24] SHU F W, WANG J X, PAGANI A, *et al.* Structure PLP-SLAM: efficient sparse mapping and localization using point, line and plane for monocular, RGB-D and stereo cameras [C]. 2023 *IEEE International Conference on Robotics and Automation (ICRA)*. May 29-June 2, 2023, London, United Kingdom. IEEE, 2023: 2105-2112.
- [25] HUA M D, MANERIKAR N, HAMEL T, *et al.* Attitude, linear velocity and depth estimation of a camera observing a planar target using continuous homography and inertial data[C]. 2018 *IEEE International Conference on Robotics and Automation (ICRA)*. May 21-25, 2018, Brisbane, QLD, Australia. IEEE, 2018: 1429-1435.
- [26] 付漫侠, 周水生. 基于融合Lasso的非参数加性分位数回归模型[J]. 模式识别与人工智能, 2024, 37(1): 58-72.
- FU M X, ZHOU S S. Nonparametric additive quantile regression model based on fused lasso[J]. *Pattern Recognition and Artificial Intelligence*, 2024, 37(1): 58-72. (in Chinese)
- [27] 雷大江. 分布式机器学习: 交替方向乘法在机器学习中的应用[M]. 北京: 清华大学出版社, 2021: 12-23.
- LEI D J. *Distributed Machine Learning*[M]. Beijing: Tsinghua University Press, 2021: 12-23. (in Chinese)
- [28] RUBLEE E, RABAU D V, KONOLIGE K, *et al.* ORB: an efficient alternative to SIFT or SURF [C]. 2011 *International Conference on Computer Vision*. November 6-13, 2011, Barcelona, Spain. IEEE, 2011: 2564-2571.
- [29] 高翔, 张涛, 刘毅, 等. 视觉SLAM十四讲: 从理论到实践[M]. 2版. 北京: 电子工业出版社, 2019: 37-81.
- GAO X, ZHANG T, LIU Y, *et al.* *Fourteen Lectures on Visual Slam: from Theory to Practice* [M]. 2nd ed. Beijing: Publishing House of Electronics Industry, 2019: 37-81. (in Chinese)
- [30] LEE D T, SCHACHTER B J. Two algorithms for constructing a delaunay triangulation[J]. *International Journal of Computer & Information Sciences*, 1980, 9(3): 219-242.
- [31] CALONDER M, LEPETIT V, STRECHA C, *et al.* BRIEF: Binary Robust Independent Elementary Features[M]. Computer Vision-ECCV 2010. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 778-792.
- [32] GALVEZ-LÓPEZ D, TARDOS J D. Bags of binary words for fast place recognition in image sequences [J]. *IEEE Transactions on Robotics*, 2012, 28(5): 1188-1197.
- [33] BURRI M, NIKOLIC J, GOHL P, *et al.* The EuRoC micro aerial vehicle datasets[J]. *The International Journal of Robotics Research*, 2016, 35(10): 1157-1163.
- [34] SCHUBERT D, GOLL T, DEMMEL N, *et al.* The TUM VI benchmark for evaluating visual-inertial odometry [C]. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. October 1-5, 2018, Madrid, Spain. IEEE, 2018: 1680-1687.
- [35] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C]. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*. June 16-21, 2012, Providence, RI, USA. IEEE, 2012: 3354-3361.

作者简介:



沈晔湖(1982—),男,江苏苏州人,博士,副教授,硕士生导师,2004年于浙江大学获得学士学位,2009年于浙江大学获得博士学位,主要从事机器人视觉导航、人工智能和机器视觉技术。E-mail: yehushen@mail.usts.edu.cn

通讯作者:



何一凡(1981—),男,浙江杭州人,博士,高级工程师,福建省A类人才,2004、2006年分别于浙江大学获得学士学位和硕士学位,2008年、2013年于荷兰埃因霍芬理工大学分别获得第二硕士学位和博士学位。主要从事嵌入式系统、人工智能算法优化、处理器架构设计等方面的研究。E-mail: heyifan@reconova.com