

一种高精度紧耦合的双目 VI-SLAM 算法

陈世明^{1,2}, 谭立宇^{1,2}, 高彦丽^{1,2}, 陶 丹^{1,2}

1. 华东交通大学电气与自动化工程学院, 江西 南昌 330013; 2. 江西省先进控制与优化重点实验室, 江西 南昌 330013

基金项目: 国家自然科学基金资助项目(61973118), 江西省自然科学基金资助项目(20202BAB212007), 江西省主要学科学术和技术带头人计划项目(20182BCB22009)

通信作者: 陈世明, c1977318@hotmail.com 收稿/录用/修回: 2020-10-23/2021-02-25/2021-04-15

摘要

针对视觉惯性同时定位与地图构建(visual-inertial simultaneous localization and mapping, VI-SLAM)系统中存在计算效率低和位姿估计精确度低问题,提出了一种同时采用扩展卡尔曼滤波和增量式光束平差法的紧耦合双目 VI-SLAM 算法.前端采用扩展卡尔曼滤波将惯性测量单元与双目相机的测量信息进行耦合估计位姿与速度,在后端通过增量式光束平差法来优化位姿获得全局一致的运动轨迹.与只采用滤波方法或者优化方法的 SLAM 算法相比较,在公开数据集 EuRoC(European robotics challenges)上本文算法能够减小位姿估计的线性化误差提升估计精度,并减少对计算资源的占用.使用移动机器人在真实环境下进行测试,验证了算法的实际可行性.

关键词

同时定位与地图构建(SLAM)
扩展卡尔曼滤波
增量式光束平差法
视觉惯性紧耦合
中图分类号: TP242
文献标识码: A

A High Precision Tightly-coupled Stereo VI-SLAM Algorithm

CHEN Shiming^{1,2}, TAN Liyu^{1,2}, GAO Yanli^{1,2}, TAO Dan^{1,2}

1. School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China;

2. The Jiangxi Province Key Laboratory of Advanced Control and Optimization, Nanchang 330013, China

Abstract

Given the problem of the low computational efficiency and accuracy of pose estimation in the visual-inertial simultaneous localization and mapping (VI-SLAM) system, a tightly-coupled stereo VI-SLAM algorithm combining an extended Kalman filter and incremental bundle adjustment is proposed. In the front end, an extended Kalman filter is employed to couple the measurement information of the inertial measurement unit and the stereo camera to estimate the pose and velocity. In the back end, the pose is optimized by incremental bundle adjustment to obtain a globally consistent motion trajectory. In comparison with the SLAM algorithm that only uses the filtering method or the optimization method, the proposed algorithm can improve the accuracy of pose estimation by reducing the linearization error of pose estimation and the occupation of computing resources on the public dataset EuRoC. A mobile robot is used to test the algorithm in a real environment, and our results verify the practical feasibility of the algorithm.

Keywords

simultaneous localization and mapping (SLAM);
extend Kalman filter;
incremental bundle adjustment;
visual inertial tightly-coupled

0 引言

视觉同时定位与地图构建(visual simultaneous localization and mapping, VSLAM)在机器人技术和计算机视觉领域应用广泛且取得了许多成果^[1-2].近年来,随着多传感器与传统 VSLAM 融合技术发展,SLAM 算法在性能上得到了显著提升^[3].因其具有实时精确的位姿估计功能,越来越多地应用于如小型飞行器(MAV)、自主机器人与增

强现实和虚拟现实技术中.

传统的视觉 SLAM 通过前端视觉里程计来为所识别场景提供良好的运动跟踪以及丰富的地图信息.目前,基于纯视觉的状态估计方法已经十分成熟.例如, Mur-Artal 在 2015 年提出间接法的 ORB-SLAM^[4],利用 ORB(Oriented fast and Rotated Brief)特征对视觉图像信息进行特征匹配与重定位来恢复相机的三维运动,同时加入位姿图优化(pose graph)来闭合回路以消除误差积累.而采用直接法

的 LSD-SLAM (Large Scale Direct monocular SLAM)^[5] 以及 DTAM (dense tracking and mapping)^[6] 则利用最小化光度误差恢复地图。

由于纯视觉 SLAM 受传感器限制, 对遮挡物多、特征不明显、运动模糊、光照变化剧烈的场景十分敏感。因此, 为了获得更加精确与鲁棒的结果, 通常融合多种传感器进行定位。其中, 将惯性测量单元 (inertial measurement unit, IMU) 与相机耦合为视觉惯性里程计的方式在 SLAM 研究中十分流行^[7]。IMU 由加速度计和陀螺仪组成, 它们能够辅助视觉测量信息估算机器人的姿态和三位航向信息, 通常应用于无人机导航与定位中。同时, 针对 IMU 采样频率和相机采样频率不同及 IMU 重复积分问题, Forster^[8] 提出了对 IMU 模型进行预积分方法。

将 IMU 与相机测量信息融合在一起的方式主要分为松耦合和紧耦合。松耦合是分别独立地利用 IMU 和相机测量信息进行位姿估计, 再将其结果进行融合解算^[9-10]。相比之下, 紧耦合是直接将 IMU 与相机测量信息作为一个状态向量共同构建观测方程, 之后再使用滤波或者优化方法联合估计姿态。对于滤波的方式, 学界通常采用扩展卡尔曼滤波 (extend Kalman filter, EKF)、粒子滤波等。其中粒子滤波主要应用于平面运动的机器人领域, 因为在三维空间中, 系统会随着观测量的增加导致计算复杂度呈指数级别上升, 所以使用基于 EKF 的 SLAM 系统更具有实用性。MSCKF (multi-state constraint Kalman filter)^[11] 采用滑动窗口来对加入不同时刻的相机位姿的状态向量进行维护, 当有特征点被滑动窗口观测到, 则通过特征点与相机位姿之间几何约束来对 EKF 进行更新。同样采用 EKF 框架的 ROVIO (RObust Visual Inertial Odometry)^[12] 使用图像像素光度误差进行跟踪。Michael^[13] 则采取迭代扩展卡尔曼滤波方法, 通过多次迭代减少线性化误差。EKF 只考虑相邻两个时刻的差异, 没有考虑历史信息, 其精度受到线性化误差影响。而基于优化方案的光束平差法 (bundle adjustment, BA) 挑选关键帧对位姿进行优化, 并利用矩阵的稀疏性边缘化旧状态, 以构建更加全局一致的地图。其中

OKVIS (Open Keyframe-based Visual-Inertial SLAM)^[14] 与 VINS (Visual-INertial System)^[15] 都是将视觉和 IMU (inertial measurement unit) 的状态量看成一个整体进行 BA 的联合优化。这种通过建立约束求解最小化残差方案能够获得很高精度的位姿估计。但由于梯度化求解最大后验问题时占用了过多的计算资源, 为了提高运算效率以及算法实时性, 利用 SLAM 测量的增量特性, 通过相对边缘化算法的 BA 求解器, 能显著地提高计算效率, 如 ICE-BA^[16]、iSAM2^[17]、SLAM++^[18] 等后端优化算法。这些算法虽然能提升运算效率, 但本质上仍需要大量的迭代, 效率仍不及滤波的方法。而滤波更新相当于最大后验的迭代, 使用 EKF 跟踪轨迹能更有效地减少 BA 的迭代次数。

因此, 为了减少算法计算资源消耗以及提升算法运动估计精度, 基于滤波和优化方法, 本文提出一种紧耦合的实时双目视觉惯性 SLAM 系统。通过 EKF 耦合 IMU 和相机测量信息估计速度与位姿。然后, 用基于增量式光束平差法 (incremental bundle adjustment, IBA) 的方法在并行线程中构建全局一致的估计, 从而实现能与基于优化方法媲美的运动跟踪估计精度。使用 PCG (preconditioned conjugate gradient) 算法^[19] 高效地求解舒尔补方程, 有效提升了算法的计算效率, 还原了更加精确鲁棒的轨迹。

1 融合 EKF 和 IBA 的双目 VI-SLAM 算法

系统结构如图 1 所示。系统主要分为前端和后端, 前端执行 EKF 跟踪运动, 利用逆深度参数化表示路标点状态, 计算重投影误差来预测并更新观测方程。为了减小非线性误差, 选取关键帧交给后端的 BA 进行优化以获得全局一致的轨迹。由于 BA 是使用高斯牛顿法处理一个最小二乘问题, 采用一种增量式的 BA 利用矩阵稀疏性的特点边缘化老的因子, 同时按时序将所有帧排列在滑动窗口中输入给局部 BA, 将选择的关键帧输入全局 BA 进行优化, 而优化后的关键帧位姿会更新给局部 BA, 来获取更加精确的位姿估计。

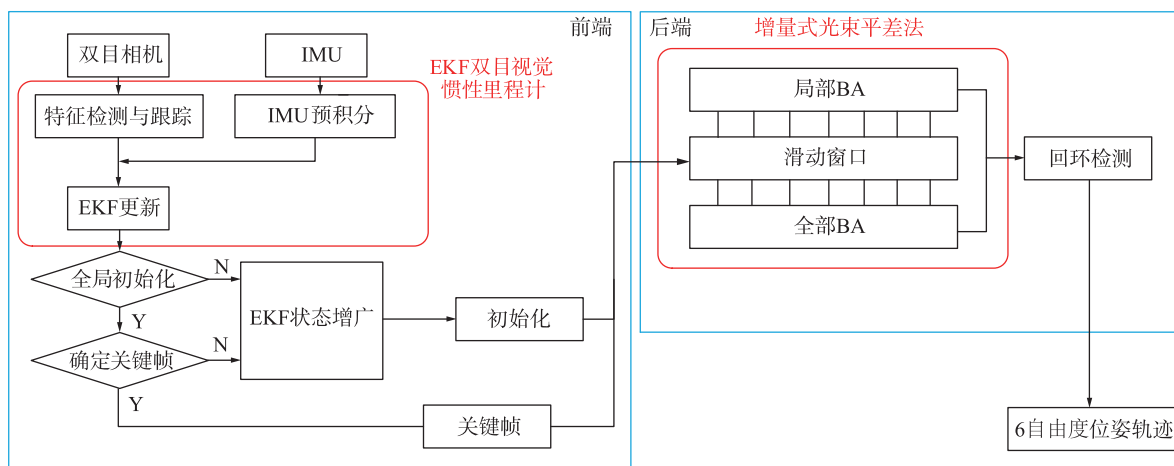


图 1 系统结构图

Fig.1 System structure

1.1 系统前端

在视觉前端利用双目相机获取视觉信息并使用 KLT (Kanade-Lucas-Tomasi) 光流法对图像特征进行跟踪, 结合 IMU 预积分估计当前帧位姿. 如果新的一帧与其他帧之间有足够视差, 那么生成关键帧传递到后端进行优化.

1.1.1 EKF 状态向量

将 k 时刻的 IMU 状态向量定义为

$$\mathbf{X}_l = [\mathbf{p}_{B_k}^{wT} \quad \mathbf{v}_{B_k}^{wT} \quad \mathbf{q}_{B_k}^{wT} \quad \mathbf{b}_{\omega_k}^T \quad \mathbf{b}_{a_k}^T] \quad (1)$$

其中, $\{B_k\}$ 表示 IMU 坐标系, $\{W\}$ 表示世界坐标系, $\mathbf{p}_{B_k}^w \in \mathbb{R}^3$, $\mathbf{v}_{B_k}^w \in \mathbb{R}^3$ 是由 IMU 坐标系 $\{B_k\}$ 旋转到世界坐标系 $\{W\}$ 的三维位姿和速度, $\mathbf{q}_{B_k}^w$ 表示从 $\{B_k\}$ 旋转到 $\{W\}$ 的四元数, $\mathbf{b}_{\omega} \in \mathbb{R}^3$, $\mathbf{b}_a \in \mathbb{R}^3$ 分别是陀螺仪和加速度计的偏差.

将 k 时刻 m 个路标点的状态向量定义为

$$\mathbf{X}_{l_k} = [l_1^{wT} \quad \cdots \quad l_m^{wT}] \quad (2)$$

其中, $l_1^w = [x_l \quad y_l \quad z_l \quad \theta_l \quad \varphi_l \quad \rho_l]^T$, $[x_l \quad y_l \quad z_l]^T$ 是第 l 个路标点观测到的相机的位置坐标, θ_l , φ_l 是方向角和俯仰角, ρ_l 是逆深度.

$$\mathbf{X}_k = [\mathbf{X}_l^T \quad \mathbf{X}_{l_k}^T]^T \quad (3)$$

对于误差量 $\delta \mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, 使用四元数来定义. 假定绕单位轴 u 旋转了一个微小的角度 $\delta \theta$, 用四元数表达为

$$\mathbf{q} = \hat{\mathbf{q}} \oplus \delta \theta \approx \hat{\mathbf{q}} \otimes \left[\frac{1}{2} \delta \theta \right] \quad (4)$$

最终将完整的误差状态向量表示为

$$\delta \mathbf{X}_k = [\delta \mathbf{p}_{B_k}^{wT} \quad \delta \mathbf{v}_{B_k}^{wT} \quad \delta \theta_{B_k}^{wT} \quad \delta \mathbf{b}_{\omega_k}^{wT} \quad \delta \mathbf{b}_{a_k}^{wT} \quad \delta l_1^{wT} \quad \cdots \quad \delta l_m^{wT}]^T \quad (5)$$

1.1.2 IMU 预测模型

通常一个 IMU 由一个三轴的陀螺仪和一个三轴的加速度计组成. IMU 测量参数通常会受到高斯白噪声和零偏移的影响, 其陀螺仪和加速度模型如下:

$$\begin{aligned} \tilde{\boldsymbol{\omega}}_B(t) &= \boldsymbol{\omega}_B(t) + \mathbf{b}_{\omega}(t) + \boldsymbol{\eta}_{\omega}(t) \\ \tilde{\mathbf{a}}_B(t) &= \mathbf{R}_w^B(\mathbf{a}_w(t) - \mathbf{g}_w) + \mathbf{b}_a(t) + \boldsymbol{\eta}_a(t) \end{aligned} \quad (6)$$

其中, $\tilde{\boldsymbol{\omega}}_B(t)$ 表示角速度测量值, $\boldsymbol{\omega}_B(t)$ 表示 IMU 坐标系下角速度的真实值, $\mathbf{b}_{\omega}(t)$ 表示角速度零偏, $\boldsymbol{\eta}_{\omega}(t)$ 表示陀螺仪的高斯噪声; $\tilde{\mathbf{a}}_B(t)$ 表示加速度计测量值, $\mathbf{a}_w(t)$ 表示世界坐标系下加速度真实值, \mathbf{g}_w 表示世界坐标系下的重力加速度矢量, 左乘 \mathbf{R}_w^B 旋转矩阵表示将坐标系旋转到 IMU 坐标系中, $\mathbf{b}_a(t)$ 表示加速度计零偏, $\boldsymbol{\eta}_a(t)$ 表示加速度计的高斯噪声.

对位姿、速度和旋转求导可以得到连续时间 IMU 的状态传播模型, 其中旋转采用四元数来表示:

$$\begin{aligned} \dot{\mathbf{p}}_B^w &= \mathbf{v}_B^w \\ \dot{\mathbf{v}}_B^w &= \mathbf{a}_w \\ \dot{\mathbf{q}}_B^w &= \mathbf{q}_B^w \otimes \left[\frac{0}{\frac{1}{2} \boldsymbol{\omega}_B} \right] \end{aligned} \quad (7)$$

实际上, IMU 是在离散时间提供测量, 获得相邻两个时刻 k 到 $k+1$ 的测量值为

$$\begin{aligned} \mathbf{p}_{B_{k+1}}^w &= \mathbf{p}_{B_k}^w + \mathbf{v}_{B_k}^w \Delta t_k + \iint_{t \in [t_k, t_{k+1}]} (\mathbf{R}_B^w(\tilde{\mathbf{a}}_{B_t} - \mathbf{b}_{a_t} - \boldsymbol{\eta}_{a_t}) - \mathbf{g}_w) dt^2 \\ \mathbf{v}_{B_{k+1}}^w &= \mathbf{v}_{B_k}^w + \int_{t \in [t_k, t_{k+1}]} (\mathbf{R}_B^w(\tilde{\mathbf{a}}_{B_t} - \mathbf{b}_{a_t} - \boldsymbol{\eta}_{a_t}) - \mathbf{g}_w) dt \\ \mathbf{q}_{B_{k+1}}^w &= \int_{t \in [t_k, t_{k+1}]} \mathbf{q}_{B_k}^w \otimes \left[\frac{0}{\frac{1}{2} \boldsymbol{\omega}_B} \right] dt \end{aligned}$$

其中, Δt_k 是时间间隔 $[t_k, t_{k+1}]$ 之间的持续时间.

对于非线性系统 $\mathbf{x}_k = f(\mathbf{x}_{k-1}, u_{k-1})$ 的状态误差的线性递推模型如下:

$$\delta \mathbf{x}_k = \mathbf{F} \delta \mathbf{x}_{k-1} + \mathbf{G} \boldsymbol{\eta}_B \quad (9)$$

其中, $\boldsymbol{\eta}_B = [\boldsymbol{\eta}_a^T, \boldsymbol{\eta}_{\omega}^T, \boldsymbol{\eta}_{b_a}^T, \boldsymbol{\eta}_{b_{\omega}}^T]$ 是系统噪声.

然后可以得到协方差的传播模型为

$$\begin{aligned} \mathbf{P}_{k+11k} &= \begin{bmatrix} \mathbf{P}_{l_{k+11k}} & \mathbf{P}_{l_{C_{k+11k}}} \\ \mathbf{P}_{l_{C_{k+11k}}} & \mathbf{P}_{C_{k+11k}} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{F} \mathbf{P}_{l_k} \mathbf{F}^T + \mathbf{G} \mathbf{Q} \mathbf{G}^T & \mathbf{F} \mathbf{P}_{l_{C_k}} \\ \mathbf{P}_{l_{C_k}}^T \mathbf{F}^T & \mathbf{P}_{C_{k+11k}} \end{bmatrix} \end{aligned} \quad (10)$$

其中, $\mathbf{P}_{l_{k+11k}}$ 是 IMU 状态的协方差矩阵, $\mathbf{P}_{l_{C_{k+11k}}}$ 是相机位姿估计的协方差矩阵, $\mathbf{P}_{l_{C_{k+11k}}}$ 是 IMU 状态误差与相机位姿估计之间的相关, \mathbf{Q} 为噪声项的对角协方差矩阵.

1.1.3 扩展卡尔曼滤波更新

为了更好地利用路标点信息, 选择使用路标点的逆深度参数化^[20]来表示. 可以将第 1 个路标点转换为相应的欧氏距离得到 XYZ 坐标 $\mathbf{X}(l_1^w)$, 如下所示:

$$\mathbf{X}(l_1^w) = \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} + \frac{1}{\rho_l} \begin{bmatrix} \cos \varphi_l \sin \theta_l \\ -\sin \varphi_l \\ \cos \varphi_l \cos \theta_l \end{bmatrix} \quad (11)$$

因此, 在逆深度参数化的坐标中第 l 个路标点投影到第 k 帧图像上的测量模型为

$$\tilde{z}_{kl} = h_{kl}(\mathbf{X}_{B_k}, l_l^w) + \mathbf{e}_{kl} \quad (12)$$

其中, \mathbf{e}_{kl} 是视觉测量协方差 $\boldsymbol{\Sigma}_{kl}$ 的测量噪声.

根据测量模型, 可以得到重投影误差并做线性化近似:

$$\begin{aligned} \mathbf{r}_{kl} &= \tilde{z}_{kl} - h_{kl}(\hat{\mathbf{X}}_{B_{klk-1}}, \hat{l}_l^w) \\ &\approx \mathbf{H}_{B_{kl}} \delta \mathbf{X}_{B_k} + \mathbf{H}_{l_{kl}} \delta l_l^w + \mathbf{e}_{kl} \\ &= \mathbf{H}_{kl} \delta \mathbf{X}_k + \mathbf{e}_{kl} \end{aligned} \quad (13)$$

其中, $\mathbf{H}_{B_{kl}}$ 、 $\mathbf{H}_{l_{kl}}$ 分别是测量方程 $h_{kl}(\mathbf{X}_{B_k}, l_l^w)$ IMU 状态和路标点位置的导数.

将第 k 帧的 m 个路标的测量误差和测量的雅可比矩阵 \mathbf{H}_{kl} 结合起来, 可以得到:

$$\mathbf{r}_k = \mathbf{H}_k \delta \mathbf{X}_k + \mathbf{e}_k \quad (14)$$

所以, EKF 状态和协方差更新为

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \boldsymbol{\Sigma}_{e_k})^{-1} \\ \hat{\mathbf{X}}_{k|k} &= \hat{\mathbf{X}}_{k|k-1} + \mathbf{K}_k \mathbf{r}_k \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \end{aligned} \quad (15)$$

1.2 系统后端

为了减小前端进行位姿估计后产生的非线性误差,同时能高效地处理增量方程计算量问题,本文在后端设计一种增量式的光束平差法的解决方案,利用滑动窗口排列优化帧,使用全局 BA 优化关键帧并将优化后的关键帧更新给局部 BA 继续优化估计位姿. 建立增量方程时对于已经计算过的状态向量不再计算其雅可比矩阵,只计算变化的路标点.

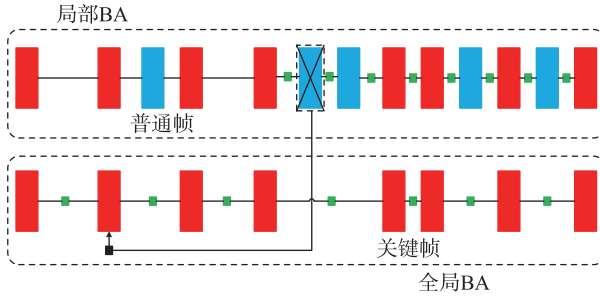


图2 局部 BA 与全局 BA

Fig.2 Local BA and global BA

1.2.1 局部与全局 BA 约束函数

被优化的变量 χ 包含在滑窗内所有 N 个相机状态和 m 个路标点的逆深度,即:

$$\chi = [x_{b1}, \dots, x_{bN}, \rho_1, \dots, \rho_m] \quad (16)$$

局部 BA 的目标函数包含 3 个约束,分别是:先验约束、IMU 约束、视觉约束.

$$\chi^* = \underset{\chi}{\operatorname{argmin}} \|r_p\|_{\Sigma_p}^2 + \sum_{(i,j) \in \kappa} \|r_{I_{ij}}\|_{\Sigma_{I_{ij}}}^2 + \sum_{i \in \kappa, l \in C} \|r_{C_{il}}\|_{\Sigma_{C_{il}}}^2 \quad (17)$$

$r_{C_{il}}$ 是视觉重投影误差,定义为

$$r_{C_{il}} = \pi((R_{B_k}^W R_C^B)^T (\rho_l - p_C)) - \tilde{z}_{il} \quad (18)$$

根据 i 和 j 两帧间 IMU 的预积分作为测量值,对两个时刻的状态进行约束,可得 IMU 残差为 $r_{I_{ij}} = [r_p^T \quad r_v^T \quad r_q^T \quad r_{b_\omega}^T \quad r_{b_a}^T]$, 其中:

$$\begin{aligned} r_p &= R_{B_k}^{WT} \left(p_j - p_i - v_i \Delta t_{ij} - \frac{1}{2} g \Delta t_{ij}^2 \right) - (\Delta \hat{p}_{ij} + J_{b_\omega}^{\Delta v_{ij}} \delta b_\omega + J_{b_a}^{\Delta v_{ij}} \delta b_a) \\ r_v &= R_{B_k}^{WT} (v_j - v_i - g \Delta t_{ij}) - (\Delta \hat{v}_{ij} + J_{b_\omega}^{\Delta v_{ij}} \delta b_\omega + J_{b_a}^{\Delta v_{ij}} \delta b_a) \\ r_q &= \log((\Delta \hat{R}_{ij} \exp(J_{b_\omega}^{\Delta R_{ij}} \delta b_\omega))^T R_i^T R_j) \\ r_{b_\omega} &= b_{\omega_j} - b_{\omega_i} \\ r_{b_a} &= b_{a_j} - b_{a_i} \end{aligned} \quad (19)$$

全局 BA 在局部 BA 的基础上加入相对边缘化约束,可以得到目标函数为

$$\chi^* = \underset{\chi}{\operatorname{argmin}} \|r_p\|_{\Sigma_p}^2 + \sum_{(i,j) \in \kappa} \|r_{I_{ij}}\|_{\Sigma_{I_{ij}}}^2 + \sum_{i \in \kappa, l \in C} \|r_{C_{il}}\|_{\Sigma_{C_{il}}}^2 + \sum_{i \in \kappa, k \in I_i} \|r_{q_k}\|_{\Sigma_{q_k}}^2 \quad (20)$$

其中, I_i 是包含在第 i 个相对姿态约束中得关键帧集. 相对边缘化约束的参考坐标系不再是世界坐标系,而是与当前帧共视且距离最近的关键帧的相机系.

1.2.2 闭环检测

本文采用 DBoW2 词袋模型来检测系统的回环. 将图

像中检测到的特征点进行 Kmean++ 聚类,建立由多个单词组成的词袋,再将待检索的图像中的特征与词袋中的单词进行匹配,来确定各帧间是否为闭环. 一旦检测到闭环,就执行基于 Sim3 的位姿图优化算法,优化整个 BA 以消除累积飘逸误差.

2 实验结果与分析

2.1 公开数据集实验结果

首先使用公开数据集 EuRoC^[21] 对提出算法进行测试,并与几种典型的 VI-SLAM 算法进行比较. 该数据集由 11 个数据序列组成,包含两个 30 m² 的室内房间和一个 300 m² 的工厂中小型无人机的飞行数据,根据照明、纹理和运动速度等,将数据序列分为:容易、中等、困难三个级别. 记录的数据表包括了 30 Hz 的双目相机数据与 200 Hz 的 IMU 数据及外部传感器提供的真实轨迹. 实验环境为: 2.20 GHz 和 16 G RAM, 显卡 GTX1060 6 G 的 Intel Core i7-8750H CPU 的笔记本电脑. 进行实验时,对每一帧设置提取 1 000 个特征点,前端视觉里程计特征跟踪状态如图 3 所示,能够有效且均匀地识别特征点.

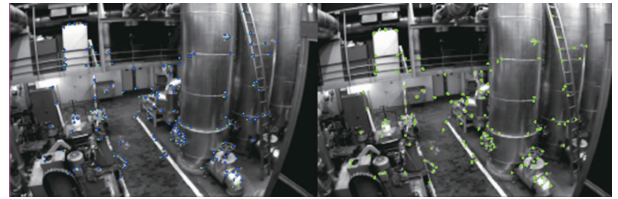


图3 双目相机特征点跟踪

Fig.3 Stereo camera feature tracking

对 V1_01_easy, V2_02_medium, V2_03_difficult 和 MH_03_medium 四个序列做定性的轨迹分析,将本文的算法和基于优化和基于滤波的双目视觉惯性 SLAM 系统、VINS 和 S-MSCKF^[22] 做比较. 如图 4 所示,在 4 个序列上本文算法的估计轨迹更加接近真实轨迹.

本文采用绝对轨迹误差 (ATE) 下的均方根误差 (RMSE)^[23] 来定量分析算法性能,结果如表 1 所示. 对于 V2 序列由于场景明暗变化不均匀,在 V2_01_easy 序列准确度低于 VINS,但随着飞行距离的增加,本文的方法能获得更高的准确度. 在 MH(machine hall)工厂场景中因为无人机飞行速度过快,场景复杂,光照较为剧烈,在复杂程度更高的 MH_04 与 MH_05 序列上,表现不如 VINS 和 S-MSCKF. 这是由于场景的低纹理与相机高速旋转运动使得特征缺失共视关系导致 BA 中的约束较少,所以估计精度没有得到很大的提升. 但在整体上,对于本文方法在 V1、V2、MH 序列中平均 RMSE 是 0.060 m, 0.071 m, 0.216 m, 相比于表现较好的 VINS,在 V1, V2 序列上的平均 RMSE 是 0.130 m, 0.194 m, 准确性约提升了 53%、63%. 因此,本文方法除在个别场景下,相比于其他的 VI-SLAM 算法,估计轨迹要更加接近真实轨迹.

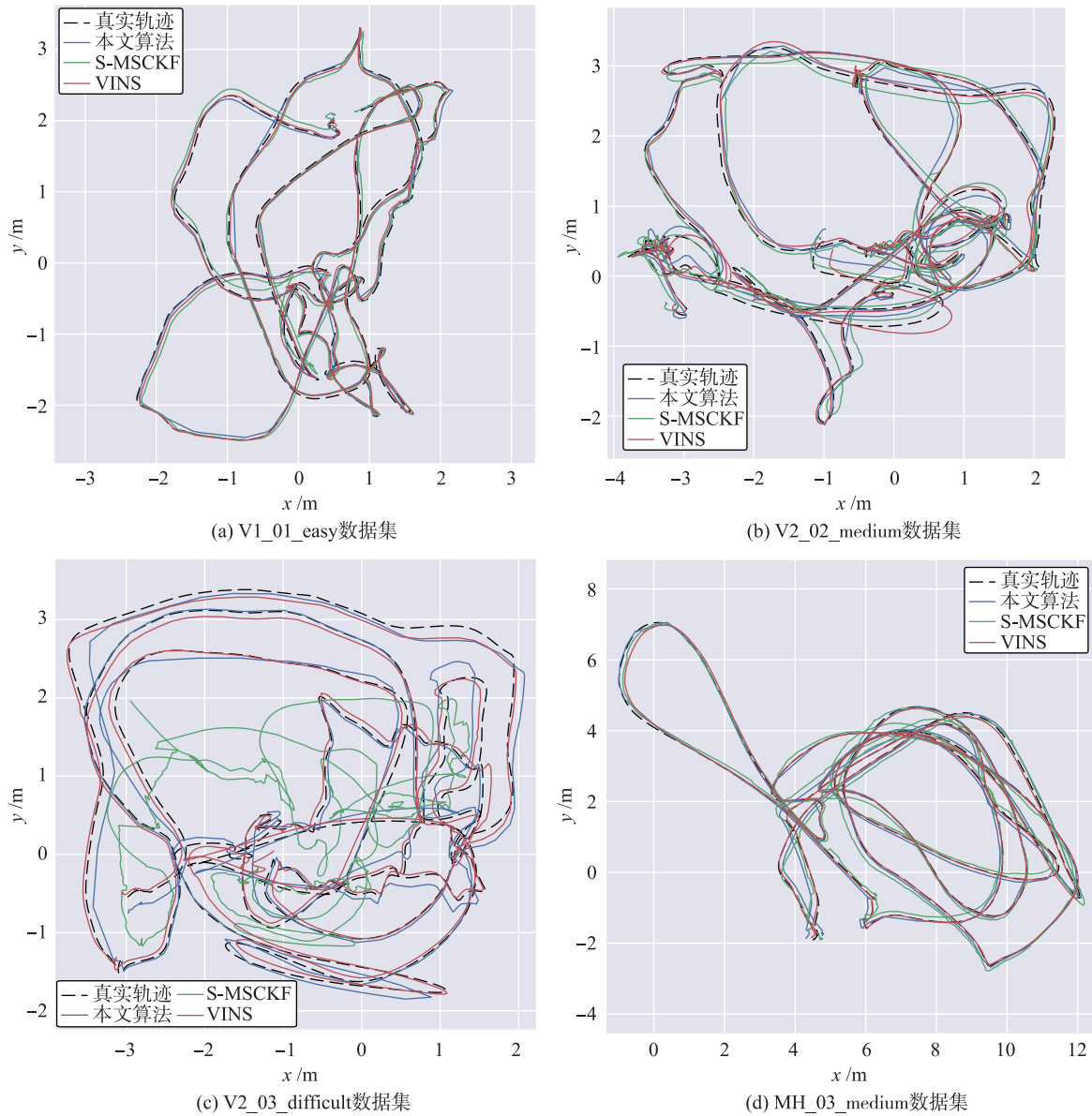


图4 轨迹估计对比图

Fig.4 The comparison graph of estimated trajectory

表1 不同方法在 EuRoC 数据集上的 RMSE

Tab.1 RMSE of different approaches on the EuRoC dataset

| 序列 | ROVIO | OKVIS | S-MSCKF | VINS | 本文算法 |
|-----------------|-------|--------------|--------------|--------------|--------------|
| V1_01_easy | 0.157 | 0.118 | 0.114 | 0.083 | 0.046 |
| V1_02_medium | 0.195 | 0.182 | 0.095 | 0.125 | 0.043 |
| V1_03_difficilt | 0.170 | 0.237 | 0.172 | 0.183 | 0.092 |
| V2_01_easy | 0.287 | 0.138 | 0.080 | 0.068 | 0.070 |
| V2_02_medium | 0.607 | 0.220 | 0.145 | 0.238 | 0.023 |
| V2_03_difficult | 0.189 | 0.315 | 1.204 | 0.276 | 0.121 |
| MH_01_easy | 0.308 | 0.173 | - | 0.180 | 0.178 |
| MH_02_easy | 0.567 | 0.246 | 0.191 | 0.151 | 0.053 |
| MH_03_medium | 0.407 | 0.283 | 0.210 | 0.186 | 0.161 |
| MH_04_difficult | 0.919 | 0.403 | 0.275 | 0.225 | 0.269 |
| MH_05_difficult | 1.266 | 0.483 | 0.271 | 0.287 | 0.334 |

算法处理每帧图片平均耗时的数据如表 2 所示,从表中可以看出,虽然在 V2_03_difficult 序列上提升不明显,但总体上本文算法处理数据耗时更短,更具有实时性.这主要是得益于本文使用了增量式 BA 求解的方案,同时 EKF 为 BA 提供了更加良好的运动跟踪,减小了最大后验的迭代压力,能有效地减少计算资源的消耗.

表 2 算法耗时
Tab.2 Time consumption of methods 单位:ms

| 序列 | VINS | OKVIS | 本文算法 |
|-----------------|------|-------|------|
| V1_01_easy | 45 | 36 | 20 |
| V1_02_medium | 38 | 32 | 21 |
| V1_03_difficult | 30 | 31 | 21 |
| V2_01_easy | 43 | 34 | 22 |
| V2_02_medium | 35 | 33 | 21 |
| V2_03_difficult | 25 | 23 | 22 |
| MH_01_easy | 47 | 45 | 23 |
| MH_02_easy | 44 | 42 | 20 |
| MH_03_medium | 42 | 43 | 20 |
| MH_04_difficult | 41 | 36 | 21 |
| MH_05_difficult | 43 | 37 | 23 |

2.2 机器人运行结果与分析

为了验证系统在现实场景下应用的可行性,本文采用搭载了 Intel Core i5-6198 CPU 2.3 GHz, 3.8 G 内存的移动机器人,以 ZED stereo 双目相机和 200 Hz 的 9DOF Razor IMU 作为传感器,如图 5. 在 Ubuntu16.04 系统上,通过机器人操作系统(ROS)控制机器人在 20 m² 的室内环境,以 0.13 m/s 的速度按照 2.0 m×3.0 m 的矩形运动路线移动三圈,如图 6 为实机 ROS 操作界面.

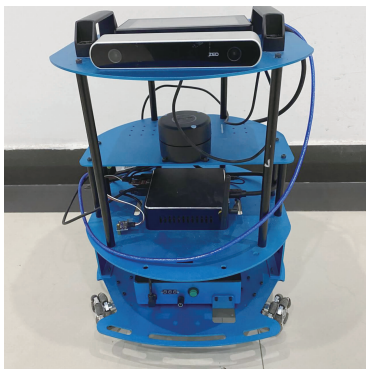


图 5 移动机器人
Fig.5 Motion robot

图 7 为机器人实时估计的移动轨迹,对比机器人真实运动轨迹(ground truth)可知,轨迹总长度为 25.85 m,共 3 圈.与 S-MSCKF (0.32m) 和 VINS (0.14 m) 的 ATE 下的 RMSE 值相比,本文方法最小为 0.12 m,体现了本文算法在精度提升方面的优越性.其中 S-MSCKF 由于没有闭环检测导致在多圈运动后出现较大误差.实验验证了本文算法的实用性与在实际环境下的鲁棒性.

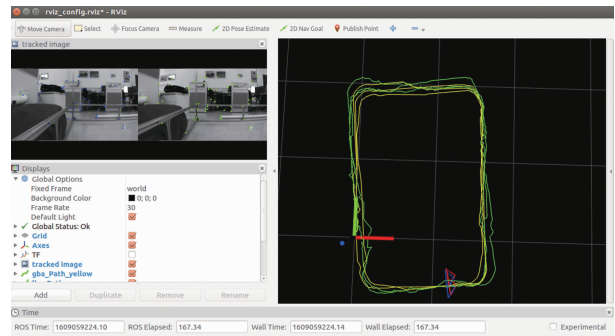


图 6 ROS 操作界面
Fig.6 ROS operation interface

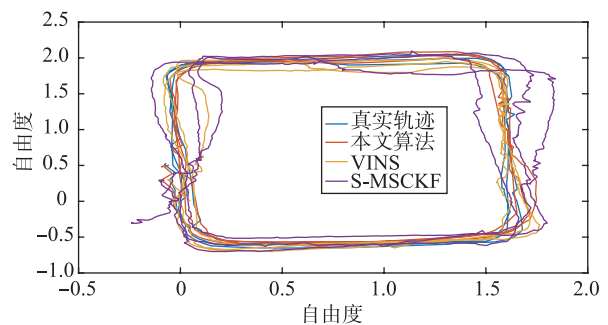


图 7 机器人实时运行轨迹
Fig.7 Real-time move trajectory of the robot

3 结论

本文提出一种基于滤波的双目视觉惯性里程计与非线性优化方法相结合的 SLAM 算法.首先通过 EKF 将 IMU 与相机测量信息紧耦合,并利用测量信息估计位姿.通过使用增量式的 BA,对老的路标点信息进行边缘化,提高了运算效率和估计精确度.基于 EuRoc 数据集与最新的开源算法相比较,验证了所提出方法的有效性及其在实际环境下的可应用性.今后将针对复杂环境下(如光照变化剧烈)位姿估计精度进行研究.

参考文献

- [1] Liu H M, Zhang G F, Bei H J. A survey of monocular simultaneous localization and mapping[J]. Journal of Computer-Aided Design and Computer Graphics, 2018, 23(6): 855-868.
- [2] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age [J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [3] Zhang F, Shen C, Ren X. The application of multi-sensor information fusion by improved trust degree on SLAM[C]//International Conference on Intelligent Human-machine Systems & Cybernetics. Piscataway, USA: IEEE, 2013: 360-364.

- [4] Mur-Artal R, Montiel J M M, Tardos J D, et al. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147–1163.
- [5] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]//*European Conference on Computer Vision*. Berlin, Germany: Springer, 2014: 834–849.
- [6] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C]//*International Conference on Computer Vision(ICCV)*. Piscataway, USA: IEEE, 2011: 2320–2327.
- [7] Quan M, Piao S, Tan M, et al. Accurate monocular visual-inertial slam using a map-assisted EKF approach[J]. *IEEE Access*, 2019, 7: 34289–34300.
- [8] 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.
- [9] Weiss S, Achtelik M W, Lynen S, et al. Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments[C]//*IEEE International Conference on Robotics & Automation*. Piscataway, USA: IEEE, 2012: 957–964.
- [10] Konolige K, Agrawal M, Sola J. Large-scale visual odometry for rough terrain[M]//*Robotics Research*. Berlin, Germany: Springer, 2010: 201–212.
- [11] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//*IEEE International Conference on Robotics & Automation*. Piscataway, USA: IEEE, 2007: 3565–3572.
- [12] Bloesch M, Omari S, Hutter M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]//*IEEE/RSJ International Conference on Intelligent Robots & Systems*. Piscataway, USA: IEEE, 2015: 298–304.
- [13] Bloesch M, Burri M, Omari S, et al. Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback[J]. *The International Journal of Robotics Research*, 2017, 36(10): 1053–1072.
- [14] Leutenegger S, Lynen S, Bosse M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. *International Journal of Robotics Research*, 2014, 34(3): 314–334.
- [15] 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.
- [16] Liu H, Chen M, Zhang G, et al. ICE-BA: Incremental, consistent and efficient bundle adjustment for visual-inertial SLAM[C]//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2018: 1974–1982.
- [17] Kaess M, Johannsson H, Roberts R, et al. iSAM2: Incremental smoothing and mapping using the Bayes tree[J]. *The International Journal of Robotics Research*, 2012, 31(2): 216–35.
- [18] Viorela I, Lukas P, Marak S, et al. Fast incremental bundle adjustment with covariance recovery[C]//*International Conference on 3D Vision*. Piscataway, USA: IEEE, 2017: 4321–4330.
- [19] Jeong Y. Pushing the envelope of modern methods for bundle adjustment[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(8): 1605–1617.
- [20] Civera J, Davison A J, Martínez M. Inverse depth parametrization for monocular SLAM[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 932–945.
- [21] Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets[J]. *International Journal of Robotics Research*, 2016, 35(10): 1157–1163.
- [22] Sun K, Mohta K, Pfrommer B, et al. Robust stereo visual inertial odometry for fast autonomous flight[J]. *IEEE Robotics & Automation Letters*, 2017, 3(2): 965–972.
- [23] Zhang Z, Scaramuzza D. A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry[C]//*IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2019: 7244–7251.

作者简介

陈世明(1977–), 男, 博士, 教授. 研究领域为复杂网络理论及应用, 多智能体系统协调控制, PSO 优化算法.

谭立宇(1995–), 男, 硕士生. 研究领域为视觉 SLAM 优化算法, 多传感器融合.

高彦丽(1978–), 女, 博士生, 副教授. 研究领域为工业过程参数检测, 复杂网络安全及控制.