

# 基于视觉惯性融合的半直接单目视觉里程计

龚赵慧, 张霄力, 彭侠夫, 李 鑫

(厦门大学航空航天学院, 福建 厦门 361005)

**摘要:** 针对半直接单目视觉里程计缺乏尺度信息并且在快速运动中鲁棒性较差的缺点, 设计了一种融合惯性测量信息的半直接单目视觉里程计, 通过 IMU (惯性测量单元) 信息弥补视觉里程计的缺陷, 有效提高跟踪精度与系统鲁棒性. 本文联合惯性测量信息与视觉信息进行初始化, 较准确地恢复了环境尺度信息. 为提高运动跟踪的鲁棒性, 提出一种 IMU 加权的运动先验模型. 通过预积分获取 IMU 的状态估计, 根据 IMU 先验误差调整权重系数, 使用 IMU 先验信息的加权值为前端提供精确的初值. 后端构建了紧耦合的图优化模型, 融合惯性、视觉以及 3 维地图点信息进行联合优化, 同时在滑动窗口中使用强共视关系作为约束, 在消除局部累积误差的同时提高优化效率与优化精度. 实验结果表明, 本文的先验模型优于匀速运动模型与 IMU 先验模型, 单帧先验误差小于 1 cm. 后端优化方法改进后, 计算效率提高为原来的 1.52 倍, 同时轨迹精度与优化稳定性也得到提高. 在 EuRoC 数据集上进行测试, 定位效果优于 OKVIS 算法, 轨迹均方根误差减小为原视觉里程计的 1/3.

**关键词:** 视觉惯性里程计; 初始化; 先验模型; 图优化; 单目视觉; 半直接法

中图分类号: TP242

文献标识码: A

文章编号: 1002-0446(2020)-05-0595-11

## Semi-Direct Monocular Visual Odometry Based on Visual-Inertial Fusion

GONG Zhaohui, ZHANG Xiaoli, PENG Xiafu, LI Xin

(School of Aerospace Engineering, Xiamen University, Xiamen 361005, China)

**Abstract:** Considering the disadvantages of semi-direct monocular visual odometry which lacks scale factor and shows poor robustness in fast motion, a semi-direct monocular visual odometry based on inertial fusion is designed. By using IMU (inertial measurement unit) information to make up for the deficiencies of visual odometry, the tracking accuracy and the system robustness can be effectively improved. Through the joint initialization of visual information and inertial measurement, the environmental scale can be accurately recovered. In order to improve the robustness of motion tracking, an IMU-weighted prior model is proposed. IMU state estimation is obtained by preintegration, the weight coefficient is adjusted according to the IMU prior error, and then the weighted value of IMU state is used to provide accurate initial estimation for the front-end. A tightly coupled graph optimization model is constructed in the back-end, which combines inertia, vision and 3D map point for joint optimization. Using common-view relationships as constraints in sliding window, it can improve the optimization efficiency and accuracy while eliminating the local cumulative error. The experimental results show that the prior model is better than both the uniform motion model and the IMU prior model, and the single frame prior error is less than 1 cm. By improving the back-end method, the calculation efficiency is increased by 0.52 times, and both the trajectory accuracy and the optimization stability are improved at the same time. Experimental results on the public dataset EuRoC demonstrate that the proposed algorithm outperforms the open keyframe-based visual inertial SLAM (OKVIS) algorithm, and the root mean square error of trajectory is reduced to 1/3 compared with the original visual odometry.

**Keywords:** visual-inertial odometry; initialization; prior model; graph optimization; monocular vision; semi-direct method

## 1 引言 (Introduction)

近年来, 无人设备在生产生活中的应用越发普及, 如: 无人车、无人机、扫地机器人等. 小型无人设备要想在复杂环境下实现自主导航<sup>[1]</sup>与精确定位的目标, 不仅要满足实时性和精确性的要求, 还

要符合小型设备轻量级与低功耗的特点. 这对传感器器件以及导航算法提出了巨大考验.

视觉传感器可获取丰富的环境信息, 具有质量轻、价格低、体积小等优势. 基于视觉传感器的同时定位与建图 (SLAM) 技术现已成为机器人导航定位的一项核心技术<sup>[2]</sup>. 视觉 SLAM 通过图像跟踪

实现机体的运动估计,在导航定位的同时可构建环境地图,在机器人导航方面有着广泛应用<sup>[3]</sup>.

根据视觉前端方法的不同,主流的视觉里程计可分成3大类:特征点法<sup>[4-5]</sup>、半直接法<sup>[6-7]</sup>和直接法<sup>[8-9]</sup>.

2015年提出的ORB-SLAM(oriented FAST and rotated BRIEF SLAM)算法<sup>[5]</sup>是一种基于特征点匹配的视觉里程计.该方法首先对图像进行特征提取、描述与匹配,再通过最小化重投影误差进行运动估计.ORB-SLAM具有较高的稳定性和鲁棒性,但丰富的特征信息意味着较大的计算复杂度.DSO(direct sparse odometry)<sup>[9]</sup>是一种具有代表性的稀疏直接视觉里程计.该方法通过最小化光度误差来估计相机运动,相比特征点法极大减小了计算量.但DSO方法对图像质量以及相机初始位姿要求较高,不适用于帧间运动较大的情况.由于直接法中不进行特征点的提取,因此难以进行地图重用.Forster提出的半直接视觉里程计SVO(semi-direct visual odometry)<sup>[6]</sup>结合了特征点法和直接法的优点,在关键帧中提取特征点,在普通帧上进行特征跟踪,具有较快的单帧处理速率以及良好的实时性,但在运动跟踪精度以及鲁棒性上仍有不足之处.

上述方法都基于单一的视觉传感器,易受到环境因素影响,在相机快速运动、图像模糊、弱纹理区域等情况下都容易出现跟踪丢失现象.随着环境复杂性与所执行任务的多样性不断提高,单一的视觉传感器已经难以满足小型无人设备对SLAM实时性和稳定性的需求.

IMU具有体积小、频率高等特点,可提供精确的短期运动估计.同时视觉导航可有效修正IMU测量的漂移.2种传感器功能互补,能够提高导航系统的鲁棒性,视觉与惯性的融合成为近年来SLAM技术研究的一个重要方向<sup>[10]</sup>.文[11]提出IMU预积分技术,通过预积分将惯性数据加入到视觉算法中.文[12]在IMU预积分基础上提出了流形上的预积分理论,并在同一个框架下对视觉与IMU进行优化.文[13]通过4个步骤将视觉与惯性信息对齐,解决了视觉惯性系统的初始化问题.文[14]使用一种新的初始化方法,并在初始化完成后构建紧耦合的视觉惯性导航系统.

本文在SVO方法<sup>[6]</sup>基础上融合惯性数据,提出一种单目视觉惯性里程计(visual-inertial odometry, VIO)算法.联合视觉信息与IMU数据进行初始化,使得单目尺度具有可观性;充分利用IMU在短时间内测量的准确性,提出一种加权的IMU先

验模型,为视觉前端提供准确的初始估计;构建了一种基于共视约束的滑动窗口优化模型,联合IMU状态参数、关键帧位姿、局部地图点位置等信息,对局部共视关键帧以及滑动窗口内关键帧进行非线性优化,获得更精确的载体位姿与3维环境信息.

本文主要贡献如下:

- 1) 在SVO算法的基础上融合惯性数据,构建半直接法的视觉惯性里程计算法框架.
- 2) 提出一种判断尺度收敛的方法作为视觉惯性联合初始化的终止条件,以提前结束初始化计算.
- 3) 提出一种IMU加权先验模型,为帧间运动估计提供精确的初值,提高运动跟踪的鲁棒性.
- 4) 构建一种基于共视约束的滑动窗口优化方法.在滑动窗口中,将添加局部优化后的共视关键帧位姿作为约束项,可提高局部运动精度并减小累计误差.

## 2 半直接视觉惯性里程计算法(Semi-direct visual-inertial odometry algorithm)

### 2.1 符号说明

本文定义了3种坐标系:世界坐标系 $w$ 、相机坐标系 $c$ 、IMU坐标系 $b$ .以第1帧图像对应的相机位置为世界坐标系,相机系与IMU系随着机体运动不断变化.IMU系到相机系之间的位姿转换阵 $T_b^c$ 通过提前标定已知.

### 2.2 算法框架

本文算法在半直接单目视觉里程计SVO的基础上,增加了视觉惯性联合初始化模块和后端优化模块,算法结构如图1所示.

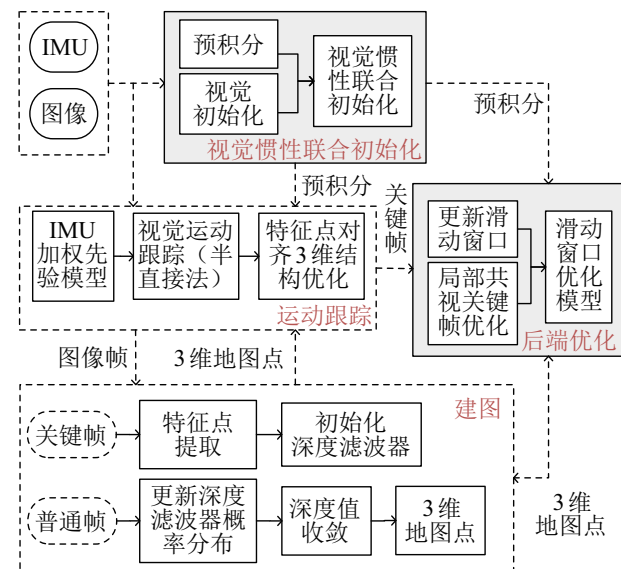


图1 算法流程图

Fig.1 Flow chart of the proposed algorithm

视觉惯性联合初始化模块用于对齐视觉与惯性信息, 可恢复单目视觉尺度, 并对惯性偏置以及重力矢量进行估计. 运动跟踪模块以加权的 IMU 先验位姿作为初值, 使用半直接法进行相邻图像帧的运动估计与位姿优化. 后端采用局部优化模型, 将共视关系作为约束项, 联合视觉、惯性以及地图点信息, 对滑动窗口内的关键帧进行优化, 并对局部轨迹进行调整. 建图模块主要维护深度滤波器的概率分布, 在关键帧上进行特征点提取与深度滤波器的初始化, 在普通帧上进行极线搜索和三角测量以更新特征点的深度分布. 深度值收敛的特征点将作为地图点保存, 并在跟踪线程中参与运动估计与优化.

本文算法中, 惯性测量数据主要以预积分的方式与视觉算法进行融合: 对相邻关键帧之间的 IMU 数据进行预积分, 用于视觉惯性联合初始化, 并在后端优化中作为关键帧间相对运动的测量值. 相邻图像帧之间的预积分用于估计 IMU 先验状态, 联合权重系数作为前端跟踪的先验知识, 为视觉位姿估计提供一个较好的初值.

文 [12] 中给出了 IMU 预积分的详细推导过程. 根据预积分理论, 可由  $i$  到  $j-1$  时刻的角速度、加速度测量数据  $\tilde{\mathbf{w}}^b$ 、 $\tilde{\mathbf{a}}^b$  计算出 IMU 系下的预积分增量:

$$\begin{aligned}\Delta \mathbf{R}_{ij} &= \mathbf{R}_i^T \mathbf{R}_j = \prod_{k=i}^{j-1} \exp((\tilde{\mathbf{w}}_k - \mathbf{b}_k^g - \boldsymbol{\eta}_k^g) \cdot \Delta t) \\ \Delta \mathbf{v}_{ij} &= \mathbf{R}_i^T (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g}^w \cdot \Delta t_{ij}) \\ &= \sum_{k=i}^{j-1} \Delta \mathbf{R}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \boldsymbol{\eta}_k^a) \cdot \Delta t \\ \Delta \mathbf{p}_{ij} &= \mathbf{R}_i^T \left( \mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i \cdot \Delta t_{ij} - \frac{1}{2} \mathbf{g}^w \cdot \Delta t_{ij}^2 \right) \\ &= \sum_{k=i}^{j-1} \left[ \Delta \mathbf{v}_{ik} \cdot \Delta t + \frac{1}{2} \Delta \mathbf{R}_{ik} (\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \boldsymbol{\eta}_k^a) \cdot \Delta t^2 \right] \quad (1)\end{aligned}$$

其中  $\mathbf{w}^b$  为 IMU 系下的角速度理想值,  $\mathbf{a}^w$  为世界系下的加速度理想值,  $\mathbf{b}^g$  与  $\mathbf{b}^a$  分别表示陀螺仪和加速度计的零偏,  $\mathbf{g}^w$  为世界坐标系下的重力矢量,  $\mathbf{R}_b^w$  表示 IMU 系到世界系的旋转矩阵的转置,  $\boldsymbol{\eta}^a$  和  $\boldsymbol{\eta}^g$  为测量噪声, 视为高斯白噪声.  $\Delta t$  为间隔时间.

### 2.3 视觉惯性联合初始化

单一的视觉里程计无法获得绝对尺度<sup>[15]</sup>, 联合惯性数据能够使得整个视觉系统的尺度与重力方向可观测, 这对导航定位和 3 维重建具有重要意义<sup>[16]</sup>.

在联合初始化之前, 需要先进行单目视觉初始化: 首先根据前 2 个关键帧中所有特征点像素位移的方差大小判断初始化场景结构, 继而求解单应性矩阵  $\mathbf{H}$  或是本质矩阵  $\mathbf{E}$ . 然后将前 2 个关键帧的相对平移量作为单位尺度, 通过三角化测量获取特征点深度估计结果, 得到相机的初始位姿.

视觉惯性联合初始化主要通过 4 个子步骤, 逐步估计出尺度、重力矢量、速度以及 IMU 偏置.

#### 2.3.1 陀螺仪偏置估计

通过陀螺仪角速度积分能得到旋转增量, 利用视觉估计出的相邻关键帧位姿也可以计算相对旋转. 将视觉估计量与陀螺仪积分量对齐, 通过最小化相对旋转的误差, 得到陀螺仪零偏的初始估计:

$$\arg \min_{\mathbf{b}^g} \sum_{i=1}^{N-1} \left\| \log((\Delta \mathbf{R}_{i,i+1} \exp(\mathbf{J}_{\Delta \mathbf{R}}^g \mathbf{b}^g))^T \mathbf{R}_w^{b(i+1)} \mathbf{R}_{bi}^w) \right\|^2 \quad (2)$$

其中  $N$  为关键帧数目,  $\Delta \mathbf{R}_{i,i+1}$  为相邻关键帧之间的陀螺仪积分,  $\mathbf{R}_b^w$  为 IMU 坐标系到世界坐标系的旋转矩阵.

由于积分数据在 IMU 系下得到, 因此要将视觉的旋转矩阵  $\mathbf{R}_c^w$  从相机系转换到 IMU 系进行计算:

$$\mathbf{R}_b^w = \mathbf{R}_c^w \mathbf{R}_b^c \quad (3)$$

其中  $\mathbf{R}_b^c$  为外参矩阵, 通过标定得到. 式 (2) 的雅可比矩阵在预积分过程中计算.

#### 2.3.2 尺度与重力矢量估计

在先前的视觉初始化中, 使用前 2 个关键帧的相对位移作为单位尺度. 在联合初始化中, 将具有真实尺度的 IMU 测量数据与视觉信息对齐, 则可获得二者之间的比例差异, 该比例即为待估计的单目视觉尺度  $s$ .

平移向量在相机坐标系与 IMU 坐标系之间的转换可以用一个含尺度的关系式表示:

$$\mathbf{p}_b^w = s \cdot \mathbf{p}_c^w + \mathbf{R}_c^w \mathbf{p}_b^c \quad (4)$$

联立 3 个连续关键帧, 可获得 2 组相对位移的关系式, 以尺度  $s$  和重力矢量  $\mathbf{g}^w$  作为待估计量并消去速度变量, 方程可表示为

$$[\lambda(i) \ \beta(i)] \begin{bmatrix} s \\ \mathbf{g}^w \end{bmatrix} = \gamma(i) \quad (5)$$

其中  $i$  为关键帧序号. 使用 SVD (singular value decomposition), 得到尺度与重力矢量的粗略估计.

### 2.3.3 加速度计偏置估计与尺度、重力方向校正

加速度计偏置的量级远小于重力矢量, 若同时进行估计, 会使得加速度偏置不可观测. 因此采用先对重力矢量粗估计, 后细化调整重力方向并估计加速度偏置的方法.

设定一个重力方向为  $\hat{\mathbf{g}}^l = [0 \ 0 \ -1]^T$  的惯性参考系  $l$ , 将上一步得到的重力矢量  $\mathbf{g}^w$  与参考系  $l$  在  $z$  轴上对齐. 通过转换矩阵  $\mathbf{R}_l^w$ , 得到重力方向在  $x$  轴和  $y$  轴上的优化增量  $\delta\theta$ :

$$\delta\theta = [\delta\theta_x \ \delta\theta_y \ 0]^T \quad (6)$$

令  $G = 9.8 \text{ m/s}^2$ , 则重力矢量可近似为

$$\mathbf{g}^w \approx \mathbf{R}_l^w \hat{\mathbf{g}}^l G - \mathbf{R}_l^w (\hat{\mathbf{g}}^l)_\times G \delta\theta \quad (7)$$

将式 (7) 代入位移递推式中, 考虑加速度偏置的影响, 联立连续的 3 个关键帧数据, 消除速度变量, 以尺度、重力方向校准量、加速度计偏置为待估量, 得到关系式 (8), 其中  $i$  为关键帧的序号.

$$[\lambda(i) \ \varphi(i) \ \zeta(i)] \begin{bmatrix} s^* \\ \delta\theta_{xy}^* \\ \mathbf{b}^{a*} \end{bmatrix} = \phi(i) \quad (8)$$

通过 SVD 分解得到精确估计的尺度  $s^*$ 、加速度计偏置  $\mathbf{b}^{a*}$  以及重力方向的优化增量  $\delta\theta_{xy}^*$ , 得到校正后的重力矢量:

$$\mathbf{g}^{w*} = \mathbf{R}_l^w \exp(\delta\theta^*) \hat{\mathbf{g}}^l G \quad (9)$$

在计算出尺度、重力矢量、惯性偏置之后, 可直接求解出关键帧对应的速度.

### 2.3.4 初始化终止条件

文 [13] 在视觉惯性系统运行的前 15 s 进行联合初始化, 以确保所有参数可观, 但固定初始化时长的方法可能存在不确定性, 还容易导致初始化时间延长. 为提高初始化结果的可靠性, 本文以尺度收敛作为联合初始化的终止条件, 考虑尺度估计的变化增量以及连续 10 次尺度估计的标准差, 设定了 2 个阈值进行收敛判断.

每确定新的关键帧之后都进行一次初始化计算, 将视觉惯性初始化求得的尺度加入尺度队列进行收敛性判断, 当尺度满足收敛条件时终止联合初始化计算. 尺度收敛判断的伪代码见算法 1.

为避免尺度估计不满足收敛精度而导致初始化持续进行, 设置 15 s 作为时长上限. 关于联合初始化终止时的尺度以及其他参数的收敛性, 本文在 4.2 节进行详细验证.

### 算法 1 尺度收敛判断方法

联合初始化估计尺度因子  $s_k$

将  $s_k$  放入尺度序列  $\mathbf{s} = [s_1, s_2, \dots, s_k]$

if  $|s_k - s_{k-1}| < \text{阈值 1}$

count<sub>1</sub> ++

计算  $s_{k-9}, s_{k-8}, \dots, s_k$  标准差  $\sigma_k$

if 标准差  $\sigma_k < \text{阈值 2}$

count<sub>2</sub> ++

else count<sub>2</sub> = 0

else count<sub>1</sub> = count<sub>2</sub> = 0

if (count<sub>1</sub> > 10) && (count<sub>2</sub> > 5)

则判断尺度收敛

### 2.4 IMU 加权先验模型

由于图像具有严重的非凸性, 通过匹配与优化方法进行运动估计需要基于一个较好的初始位姿.

IMU 对于短期运动以及快速运动都具有较高的测量精度, 但当机体处于静止状态或缓慢运动时, 惯性器件的漂移与测量噪声使得先验模型产生一定误差. 而在静止条件下, 视觉不容易出现跳变, 表现出较强的稳定性.

结合 IMU 先验与视觉先验的特点, 本文提出一种 IMU 加权的运动先验模型: 首先通过预积分更新 IMU 状态先验; 接着根据视觉静止模型与 IMU 先验模型的误差大小分配权重占比; 最后使用 IMU 加权先验为前端提供精确的初值估计, 提高视觉惯性系统的跟踪性能.

#### 2.4.1 IMU 状态更新

在新一帧图像到来前, 对 2 帧之间的 IMU 测量数据进行预积分处理, 获得位移、速度以及旋转增量  $\Delta\mathbf{p}$ 、 $\Delta\mathbf{v}$ 、 $\Delta\mathbf{R}$ , 并更新 IMU 误差的雅可比矩阵. 结合预积分增量与上一帧时刻的 IMU 状态, 通过下式更新 IMU 状态:

$$\begin{aligned} \mathbf{R}_{k+1} &= \mathbf{R}_k \cdot \Delta\mathbf{R} \\ \mathbf{v}_{k+1} &= \mathbf{v}_k + \mathbf{R}_k \cdot \Delta\mathbf{v} + \mathbf{g}^w \cdot \Delta t \\ \mathbf{p}_{k+1} &= \mathbf{p}_k + \mathbf{v}_k \cdot \Delta t + 0.5 \cdot \mathbf{g}^w \cdot \Delta t^2 + \mathbf{R}_k \cdot \Delta\mathbf{p} \end{aligned} \quad (10)$$

其中下标  $k, k+1$  为相邻图像帧到达的时刻,  $\mathbf{g}^w$  为联合初始化中估计的重力矢量,  $\Delta t$  为 2 帧之间的积分时间.

式 (10) 中的惯性状态量均在 IMU 系中得到, 通过相机坐标系与 IMU 坐标系的外参转换阵  $\mathbf{T}_c^b$ , 将其转换为相机坐标系下的 IMU 先验信息, 使用右上标 <sup>IMU</sup> 注明:  $\mathbf{R}_{k+1}^{\text{IMU}}$ 、 $\mathbf{v}_{k+1}^{\text{IMU}}$ 、 $\mathbf{p}_{k+1}^{\text{IMU}}$ .

### 2.4.2 权重计算

在视觉算法完成当前帧位姿优化后, 以优化后的位姿作为真值计算权重系数, 并使用右上标  $v$  表示优化后的位移矢量:  $\mathbf{p}_{k+1}^v$ . 视觉静止模型的帧位姿变化为 0, 因此可将  $k$  时刻的相机位姿作为  $k+1$  时刻先验位姿, 并用右上标  $last$  表示:  $\mathbf{R}_{k+1}^{last}$ 、 $\mathbf{p}_{k+1}^{last}$ . 分别计算 IMU 先验位移  $\mathbf{p}_{k+1}^{IMU}$  与静止模型先验位移  $\mathbf{p}_{k+1}^{last}$  相对于真值  $\mathbf{p}_{k+1}^v$  误差的 2 范数, 根据先验误差的比例关系进行权重分配, 计算方法如下:

$$\begin{aligned} e_{k+1}^{IMU} &= \|\mathbf{p}_{k+1}^v - \mathbf{p}_{k+1}^{IMU}\|_2 \\ e_{k+1}^{last} &= \|\mathbf{p}_{k+1}^v - \mathbf{p}_{k+1}^{last}\|_2 \\ h_{k+2} &= \frac{e_{k+1}^{last}}{e_{k+1}^{IMU} + e_{k+1}^{last}} \end{aligned} \quad (11)$$

其中  $e_{k+1}^{IMU}$  为 IMU 先验位移误差的 2 范数,  $e_{k+1}^{last}$  为静止模型先验位移误差的 2 范数,  $h_{k+2}$  为 IMU 先验的权重系数, 用于  $k+2$  时刻图像帧的先验估计.

### 2.4.3 运动先验估计

使用右上标  $prior$  表示先验估计, 则图像在  $k+2$  时刻的运动先验可表示为

$$\begin{aligned} \mathbf{p}_{k+2}^{prior} &= h_{k+2} \cdot \mathbf{p}_{k+2}^{IMU} + (1 - h_{k+2}) \cdot \mathbf{p}_{k+2}^{last} \\ \mathbf{R}_{k+2}^{prior} &= h_{k+2} \cdot \mathbf{R}_{k+2}^{IMU} + (1 - h_{k+2}) \cdot \mathbf{R}_{k+2}^{last} \end{aligned} \quad (12)$$

通过  $k+2$  时刻相机的先验位移  $\mathbf{p}_{k+2}^{prior}$  与旋转  $\mathbf{R}_{k+2}^{prior}$  可以得到相机坐标系  $c$  相对于世界坐标系  $w$  的转移矩阵  $\mathbf{T}_{c,k+2}^w$ , 将其作为初始值代入相邻帧特征点的稀疏对齐算法中, 通过最小化光度误差进行位姿优化.

在视觉惯性联合初始化完成之前, 本文仅使用视觉静止模型进行先验估计, 当初始化完成之后, 采用 IMU 加权的运动先验模型.

## 2.5 基于共视约束的滑动窗口优化

视觉 SLAM 的后端优化常使用基于关键帧的滑动窗口优化<sup>[17]</sup>方法, 保留邻近的  $N$  个关键帧, 将待优化变量限制在一个滑动窗口内, 实现控制计算规模的目的. 移动滑动窗口时, 移除最旧的关键帧会损失一些有效的关联信息. 文 [9,18] 对移出窗口的关键帧进行比较复杂的边缘化处理, 以保持 BA (bundle adjustment) 优化计算的稀疏性. 文 [13] 在滑动窗口优化算法中将窗口关键帧的所有共视关键帧都作为固定项, 以充分发挥先验信息的约束作用. 但该方法会增加优化环节的计算复杂度; 同时弱共视帧中包含一些不确定信息, 也会对优化精度产生一定影响.

为解决上述问题, 本文结合共视图<sup>[5]</sup>的概念, 在后端优化环节提出一种基于共视约束的滑动窗口

优化模型. 该模型仅使用少数强共视性关键帧作为滑动窗口约束项, 通过优化局部地图以提高共视关联信息的准确性.

算法分为 2 个步骤: (1) 优化局部地图点与共视关键帧; (2) 选取共视性强的非窗口关键帧作为约束项, 结合相机位姿、惯性测量以及 3 维地图点, 构建滑动窗口的图优化模型.

### 2.5.1 局部地图点与共视关键帧优化

在滑动窗口优化之前需要先对局部地图进行初步优化. 局部地图由当前关键帧与共视关键帧以及可视地图点构成. 将关键帧位姿与地图点的优化问题表现为图的形式, 构建如图 2 所示的模型.

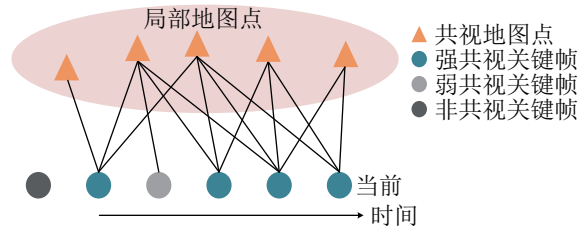


图 2 局部优化模型

Fig.2 Local optimization model

局部地图中共视关系较弱的关键帧不参与优化, 仅作为固定项提供约束. 优化变量包括所有强共视关键帧的位姿  $\mathbf{T}_1, \dots, \mathbf{T}_m$  以及共视地图点 3 维坐标  $\mathbf{p}_1, \dots, \mathbf{p}_n$ :

$$\mathbf{x} = [\mathbf{T}_1, \dots, \mathbf{T}_m, \mathbf{p}_1, \dots, \mathbf{p}_n]^T \quad (13)$$

相机在位姿  $\mathbf{T}_i$  处对地图点  $\mathbf{p}_l$  的观测误差可表示为

$$\mathbf{e}_{il} = \mathbf{z}_{il} - \mathbf{h}(\mathbf{T}_i, \mathbf{p}_l) \quad (14)$$

其中  $\mathbf{z}_{il}$  为地图点在图像上的像素坐标,  $\mathbf{h}(\mathbf{T}, \mathbf{p})$  表示地图点的观测方程: 将世界坐标系下的地图点投影到像素坐标系下.

联合  $n$  个地图点以及  $m$  个关键帧位姿构建重投影误差项, 局部优化的目标函数  $E_{proj}$  可定义为

$$E_{proj} = \sum_{i=1}^m \sum_{l=1}^n \frac{1}{2} \|(\mathbf{z}_{il} - \mathbf{h}(\mathbf{T}_i, \mathbf{p}_l))^T \boldsymbol{\Omega}_l (\mathbf{z}_{il} - \mathbf{h}(\mathbf{T}_i, \mathbf{p}_l))\|^2 \quad (15)$$

其中  $\boldsymbol{\Omega}_l$  为信息矩阵, 代表地图点观测的可信程度.

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} E_{proj} \quad (16)$$

采用 L-M (Levenberg-Marquardt) 算法对上式进行求解, 最小化目标函数以获得优化的关键帧位姿和地图点坐标.

图3为一段局部运动轨迹的图模型。通过连续的帧间位姿估计得到红色虚线所示轨迹，由于累积误差随时间不断增大，因此在后半部分出现漂移。黑色实线为使用局部优化后的轨迹，通过固定时间上靠前的共视关键帧信息，为地图点和当前帧位姿的优化提供关联先验，可将轨迹校正回正确位置。

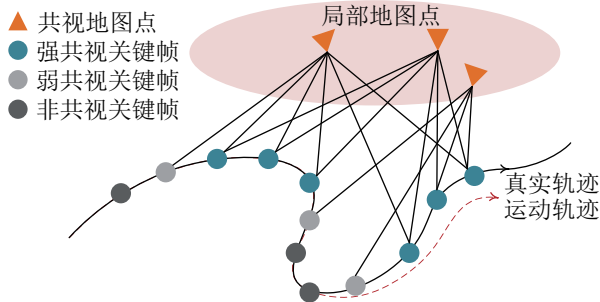


图3 局部运动轨迹的图模型

Fig.3 Graph model of the local trajectory

在滑动窗口优化之前增加该步骤可起到以下四点作用：(1) 剔除误差较大的观测信息，消除局部累积误差对后续优化的影响；(2) 提高局部地图精度，为滑动窗口优化提供良好的初值；(3) 以强共视、高精度的关键帧作为滑动窗口的约束项，可提高优化效率与精度。

### 2.5.2 滑动窗口优化模型

在上一步骤局部优化完成后，将共视点数量超过一定阈值的强共视关键帧作为约束项加入滑动窗口，共视关键帧与窗口关键帧之间通过局部地图点建立联系。滑动窗口内，由IMU预积分提供相邻关键帧之间的运动测量值，约束关键帧的位姿优化。结合了惯性数据、视觉位姿以及地图点信息的滑动窗口图优化模型如图4所示。

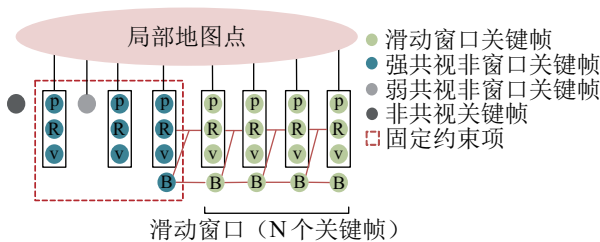


图4 滑动窗口图优化模型

Fig.4 Graph optimization model of the sliding window

设定窗口关键帧数为  $N$ ，以窗口内所有关键帧的位移、旋转、速度为待优化变量。结合每一关键帧时刻对应的IMU状态作为运动约束，并将陀螺仪和加速度的偏置也加入待优化变量中共同优化。

窗口内相邻关键帧  $i, j$  之间，使用IMU预积分作为测量值（图4中红色实线表示）。视觉位姿估计

与预积分测量构成了IMU测量误差项  $E_{\text{IMU}}(i, j)$ ：

$$E_{\text{IMU}}(i, j) = \rho([\mathbf{e}_R^T \ \mathbf{e}_v^T \ \mathbf{e}_p^T] \boldsymbol{\Sigma}_I [\mathbf{e}_R^T \ \mathbf{e}_v^T \ \mathbf{e}_p^T]^T) + \rho(\mathbf{e}_b^T \boldsymbol{\Sigma}_R \mathbf{e}_b) \quad (17)$$

IMU测量误差包含两部分  $\mathbf{R}, \mathbf{v}, \mathbf{p}$  以及偏置  $\mathbf{b}^g, \mathbf{b}^a$ ，将2种偏置合并为6维误差向量  $\mathbf{b}$  进行计算。各误差分量由以下公式给出：

$$\begin{aligned} \mathbf{e}_R &= \log((\Delta \mathbf{R}_{ij} \cdot \exp(\mathbf{J}_{\Delta R}^g \mathbf{b}_j^g))^T \mathbf{R}_w^{bi} \mathbf{R}_{bj}^w) \\ \mathbf{e}_v &= \mathbf{R}_w^{bi} (\mathbf{v}_{bj}^w - \mathbf{v}_{bi}^w - \mathbf{g}^w \Delta t_{ij}) - (\Delta \mathbf{v}_{ij} + \mathbf{J}_{\Delta v}^g \mathbf{b}_j^g + \mathbf{J}_{\Delta v}^a \mathbf{b}_j^a) \\ \mathbf{e}_p &= \mathbf{R}_w^{bi} \left( \mathbf{p}_{bj}^w - \mathbf{p}_{bi}^w - \mathbf{v}_{bi}^w \cdot \Delta t_{ij} - \frac{1}{2} \mathbf{g}^w \cdot \Delta t_{ij}^2 \right) - (\Delta \mathbf{p}_{ij} + \mathbf{J}_{\Delta p}^g \mathbf{b}_j^g + \mathbf{J}_{\Delta p}^a \mathbf{b}_j^a) \\ \mathbf{e}_b &= \mathbf{b}_j - \mathbf{b}_i \end{aligned} \quad (18)$$

其中  $\boldsymbol{\Sigma}_I$  为预积分的信息矩阵， $\boldsymbol{\Sigma}_R$  为测量噪声的信息矩阵。残差的雅可比矩阵  $\mathbf{J}_{\Delta R}^g, \mathbf{J}_{\Delta v}^g, \mathbf{J}_{\Delta v}^a, \mathbf{J}_{\Delta p}^g, \mathbf{J}_{\Delta p}^a$  在IMU预积分过程中计算。设置相应的鲁棒核函数  $\rho(\cdot)$ ，以减少大误差项对目标函数的影响。

如图4所示，在计算中将滑动窗口前的一个关键帧  $\mathbf{R}, \mathbf{v}, \mathbf{p}, \mathbf{b}$  固定，用于限制窗口关键帧优化。它与窗口内第1个关键帧可通过预积分得到一个误差项  $E_{\text{IMU}}(0, 1)$ ，包含在目标函数中进行计算。其余强共视非窗口关键帧仅使用  $\mathbf{R}, \mathbf{v}, \mathbf{p}$  作为约束，为优化提供先验信息。

除IMU误差项之外，窗口内每个关键帧位姿与其观测的地图点构成一组视觉重投影误差，由式(14)求得关键帧  $j$  的视觉误差为

$$E_{\text{proj}}(j) = \sum_{l=1}^n \frac{1}{2} \|\mathbf{e}_{jl}^T \boldsymbol{\Omega}_l \mathbf{e}_{jl}\|^2 \quad (19)$$

其中  $n$  为关键帧  $j$  可观测的地图点数量。

联合  $N$  个关键帧的视觉误差项与IMU误差项，得到滑动窗口优化的目标函数为

$$E_{\text{total}} = \sum_{j=1}^N E_{\text{proj}}(j) + \sum_{j=1}^N E_{\text{IMU}}(j-1, j) \quad (20)$$

本文使用  $g^2o$  (general graph optimization)<sup>[19]</sup> 方法构建后端优化问题的图模型，利用L-M方法求解状态增量使得目标函数达到最小。在优化算法完成后对状态参数进行更新，并用于后续的运动跟踪与位姿估计。

### 2.6 关键帧选取策略

本文采用半直接法的视觉前端，特征点的提取、地图点的更新以及后端优化都基于关键帧进行。关键帧的质量影响整个视觉惯性系统的运算周

期与算法精度. 选取合适的帧不仅可以提高视觉跟踪的准确度, 还有助于提高计算效率.

基于视觉前端的特点, 本文采用空间上靠近的关键帧选取方法, 选取策略为: (1) 当上一个关键帧投影到当前帧上的特征点个数少于阈值时; (2) 当上一个关键帧与当前帧共视点的像素距离的中值大于一定阈值时; (3) 当上一关键帧到当前帧的相机位置平移量超过一定阈值时; (4) 当上一关键帧到当前帧的旋转量超过一定阈值时.

对于视觉惯性融合的系统, 文 [13] 将相邻关键帧的间隔时间限制在 0.5 s 以内, 以防惯性测量漂移带来较大误差. 本文方法中, 机体在静止时不产生新的关键帧, 这势必造成惯性漂移误差的累积. 而本文提出的 IMU 加权运动先验方法可有效减少静止时 IMU 漂移误差对运动估计的影响. 本文算法在完成关键帧位姿优化后, 使用视觉信息校正 IMU 状态, 为后续运动估计提供更准确的状态参数.

### 3 实验结果与分析 (Experiment results and analysis)

本文在半直接视觉里程计 (SVO) 算法上融合惯性数据以提高运动估计精度及算法鲁棒性. 为验证视觉惯性融合效果, 进行了以下实验: 1) 验证本文提出的 IMU 加权运动先验模型的准确性; 2) 验证尺度收敛条件下联合初始化中各参数的收敛性; 3) 验证本文提出的共视约束滑动窗口优化方法的性能; 4) 验证本文算法的精度并与同类型 VIO 算法进行对比.

实验采用 EuRoC 数据集<sup>[20]</sup>进行测试, 使用数据集的左相机图像序列以及 IMU 数据作为输入. 得到的轨迹数据通过 EVO (evaluation of odometry and SLAM) 测评工具进行精度评估, 以轨迹的均方根误差 (RMSE) 作为评估算法精度的指标. 实验 PC 配置: CPU 为 Intel i5-8400 处理器, 主频 2.8 GHz, 在 Ubuntu 16.04 系统下运行.

#### 3.1 先验模型误差对比

分别使用 4 种先验模型作为运动先验, 在 3 个数据集上进行测试, 与数据集提供的机体位移矢量与姿态四元数的真值进行比较, 得到 4 种先验模型的误差. 先验模型为: (1) 视觉静止先验模型; (2) 匀速运动先验模型; (3) IMU 先验模型; (4) 本文的 IMU 加权先验模型. 为避免视觉匹配误差以及位姿估计误差对实验的影响, 本文采用单帧先验的绝对误差作为比较量. 在世界系下计算 4 种先验模型估计的先验位置与真实运动增量下的先验位置的绝对

误差, 得到 4 组数据, 将其绘制为误差曲线进行比较.

图 5 显示了 MH\_01 数据集第 600~1000 帧的单帧先验误差对比曲线, 其中包含一段静止状态与一段运动状态. 由图 5 可以看出, IMU 先验误差曲线 (图中红线表示) 与匀速模型误差曲线 (图中黑色曲线) 整体上类似. 由于 IMU 对短期运动具有较高测量精度, 因此在机体运动时, IMU 对应的权重较大, IMU 加权先验模型得到的误差曲线 (图中蓝线表示) 更接近 IMU 先验误差曲线, 并且通过加权减小了运动突变时的波动. 当机体处在静止状态时 (图 5 中虚线框所示), 惯性传感器的漂移使得 IMU 先验估计与真值之间存在一定偏差, 此时 IMU 加权先验模型的权重更倾向于静止模型 (图中黄色曲线), 静止状态时的蓝色曲线与黄色曲线几乎贴合, 先验误差在 0.2 mm 上下浮动.

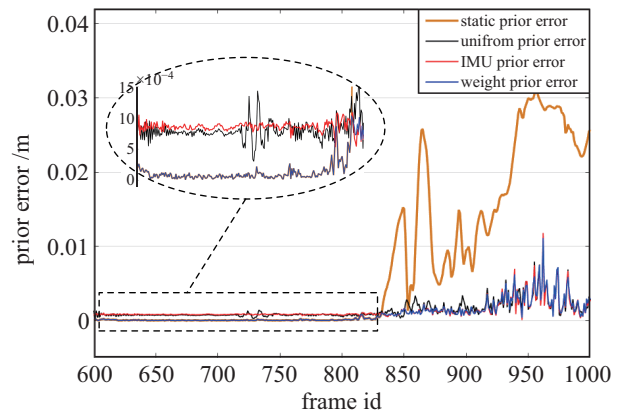


图 5 单帧先验误差对比曲线  
Fig.5 Comparison curve of the prior error per frame

在 3 个数据集中分别对 4 种模型的单帧先验误差求平均值, 得到先验平均误差, 如表 1 所示. 视觉静止模型的先验误差本质上为厘米级的单帧运动增量, 其余 3 种运动模型的先验误差均在毫米级. 本文使用的 IMU 加权先验模型取得最小的平均误差, 优于 IMU 先验模型与匀速模型, 在机体运动或静止状态下都能较好地进行先验估计.

#### 3.2 联合初始化参数收敛性验证

在视觉惯性联合初始化环节对待估计参数进行收敛性验证. 文 [13] 将联合初始化时间设为 15 s, 而本文以尺度收敛作为初始化终止条件. 为对比 2 种条件下初始化参数的收敛情况, 在 MH\_01 数据集上选取 10 个不同时刻进行初始化实验, 得到 2 种条件下估计的尺度、重力矢量、加速度计偏置、陀螺仪偏置, 绘制成散点图进行比较, 如图 6 所示.

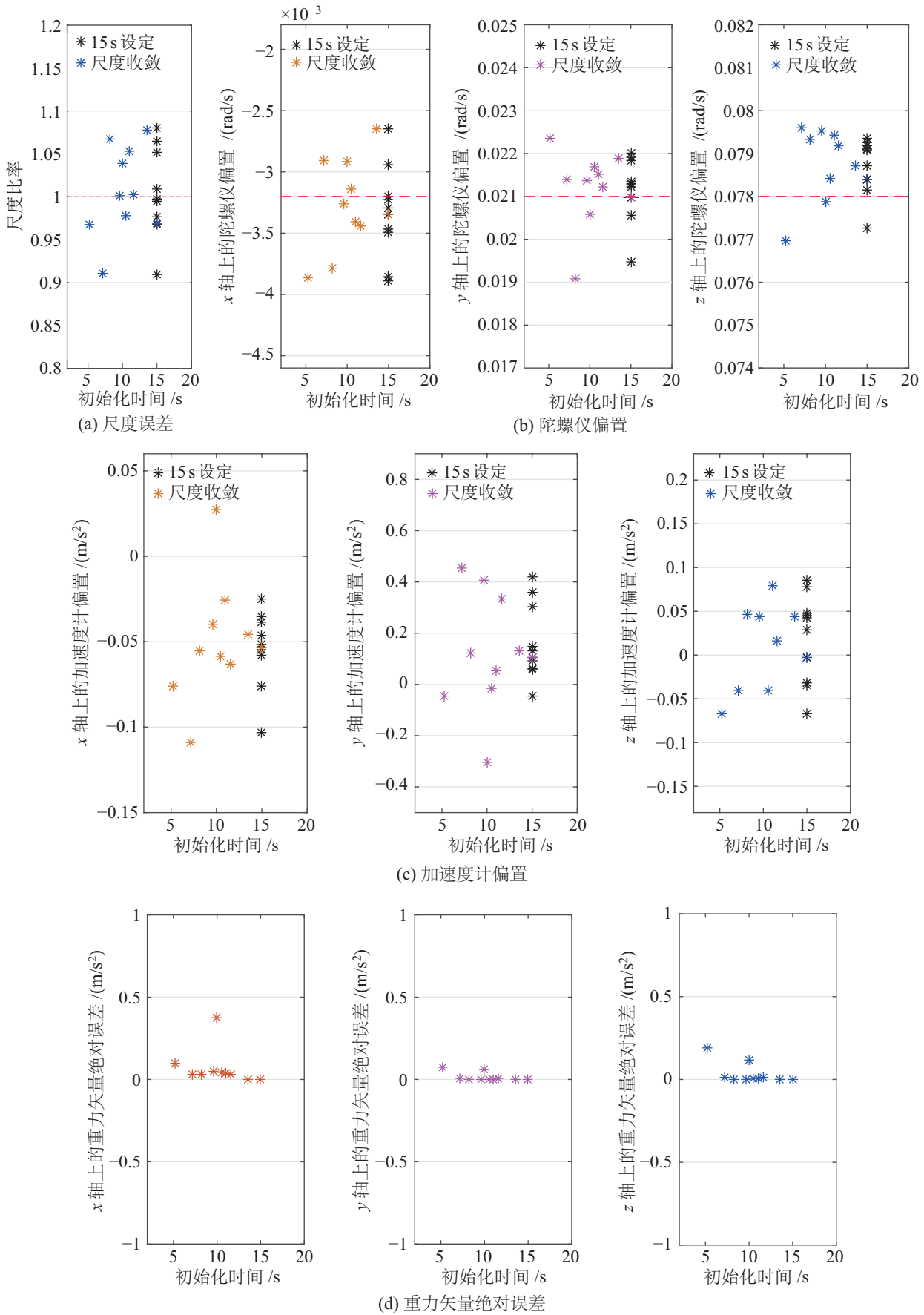


图 6 尺度收敛条件与 15 s 设定条件下的初始化参数比较

Fig.6 Comparison of initialization parameters under scale convergence condition and 15 s setting condition

图 6(a) 对 2 种条件下估计的尺度进行比较, 横坐标为初始化运行时间, 纵坐标为尺度估计值与拟合值的比例. 图 6(b)(c) 分别为 2 种条件下的陀螺仪偏置以及加速度计偏置在时间线与数值上的分布. 图中红色虚线表示参数真值, 陀螺仪偏置真值参考文献 [21] 使用的  $[-0.0032, 0.021, 0.078]$  rad/s.

由星型点的分布情况可知, 采用尺度收敛判断的方法在减少时耗上表现出较大优势. 2 种条件下尺度误差、陀螺仪偏置、加速度计偏置的收敛情况整体上较为相近, 仅个别惯性偏置的估计值存在误差. 初始化估计的尺度与陀螺仪偏置均分布在真值附近, 说明算法可以较好地反映真实环境信息.

表 1 先验模型平均误差  
Tab.1 Average error of the prior models

数据集	静止模型 /cm	匀速模型 /cm	IMU 先验 /cm	IMU 加权先验 /cm
MH.01	2.241 1	0.254 5	0.259 4	0.248 3
MH.02	2.522 3	0.343 4	0.350 1	0.333 1
MH.03	5.575 2	0.828 7	0.847 1	0.814 4

图 6(d) 中的数据为尺度收敛条件与 15 s 设定条件下的重力矢量绝对误差, 3 轴上的重力分量的绝对误差均分布在 0 值附近.

综上可得, 当尺度达到收敛时, 其他初始化参数也基本达到收敛. 对于个别惯性参数估计存在误

差的情况, 本文在后端非线性优化中对惯性偏置的误差量进行估计, 若在初始化环节中无法得到收敛的惯性偏差, 则可通过后端进一步优化. 因此, 本文将尺度收敛作为初始化终止条件可基本满足待估参数收敛的要求. 与 15 s 设定条件相比, 在精度相似的情况下可有效节约初始化时间, 加快视觉与惯性的运动跟踪及位姿优化进程.

除算法层面外, 初始化性能的好坏还与初始时刻的运动状态具有较大关联. 文 [22] 指出当机体处于静止或匀速运动状态时, VI-SfM (visual-inertial structure from motion) 问题将存在无穷解, 导致联合初始化无法运行. 而初始化不佳会严重影响运动估计的性能 [23], 因此对于视觉惯性初始化方法还有待进一步研究与改进.

### 3.3 共视约束滑动窗口优化方法的性能验证

本文在后端使用了共视约束的滑动窗口优化方法对局部轨迹以及 3 维点进行校正. 为验证方法改进后的效果, 在 MH\_01、MH\_02、MH\_03、V2\_01 四个数据集上对本文方法和改进前的滑动窗口优化方法进行测试. 将各数据集的实验分为 2 组: a 组采用改进前的优化方法; b 组采用改进后的优化方法, 并设置相同的实验参数. 为排除其他因素的干扰, 在轨迹评估时使用拟合的尺度, 每组实验测试 13 次. 统计其轨迹 RMSE 以及单帧处理平均时间, 对比结果见图 7.

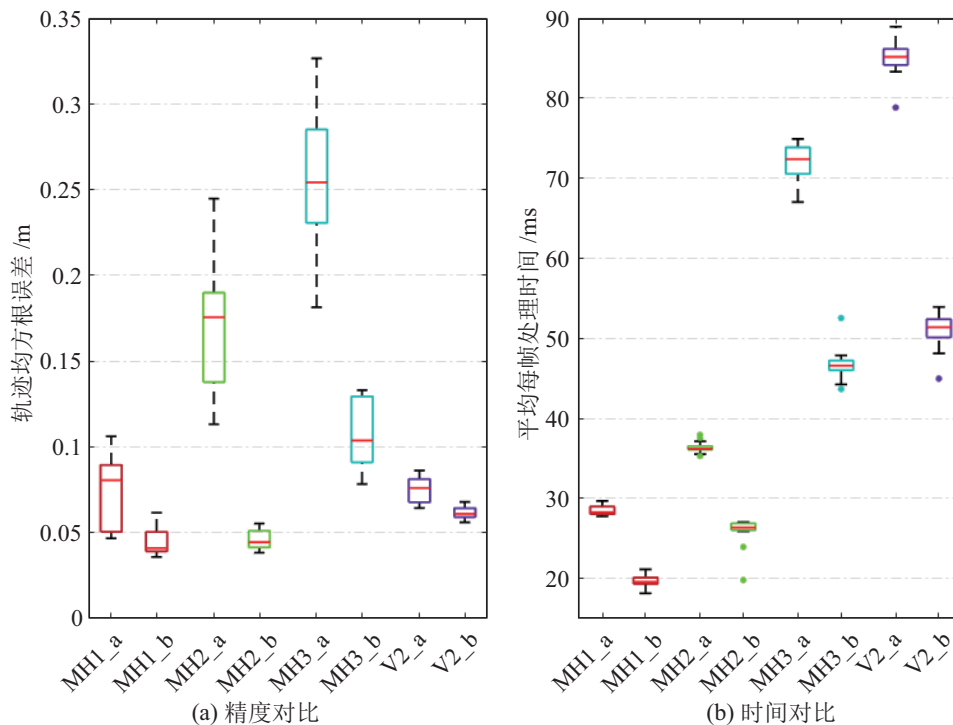


图 7 各数据集上得到的轨迹精度与平均处理时间  
Fig.7 Trajectory accuracy and mean-time on each dataset

图 7(a) 展示了各数据集在 2 种后端优化下的轨迹均方根误差及其分散程度, 每个图形包含 13 个数据, 并通过最小值、第 1 四分位点、中值、第 3 四分位点、最大值对数据进行描述。

对实验数据集采用 b 组 (改进后) 方法后, 其平均轨迹精度相较于 a 组 (改进前) 分别提高了: 0.028 2 m、0.125 4 m、0.154 7 m、0.013 8 m; 在多次重复实验下, b 组数据的波动幅度较小, 表现出更强的稳定性, 而 a 组数据离散程度更大。

图 7(b) 为后端方法改进前后的单帧处理平均时间的比较。使用改进方法后, 各组数据集上的平均耗时有不同程度的减小, b 组方法在计算效率上约为 a 组的 1.52 倍。这说明后端优化方法的改进对提高优化效率具有明显效果。

半直接法的前端在跟踪精度及稳定性上不及特征点法, 因此后端优化环节容易因前端位姿估计不准确导致优化算法耗时增加甚至优化失败。而图 7 的对比结果表明, 本文改进后的后端优化方法可有效提高运动估计精度, 增强算法稳定性, 极大减少了计算耗时。

### 3.4 算法精度对比实验

本文在 SVO 算法基础上设计了单目视觉惯性里程计, 为进一步验证其整体性能, 在多个数据集上进行了测试并与当前主流的 VIO 方法对比, 包括 OKVIS<sup>[18]</sup>、VI-ORB (visual-inertial ORB)<sup>[13]</sup>、VINS-Mono (visual-inertial navigation systems-monocular)<sup>[14]</sup>、VI-DSO (visual-inertial DSO)<sup>[24]</sup>。

OKVIS 前端采用特征点法, 后端采用基于关键帧的滑动窗口优化法; VI-ORB 在 ORB SLAM<sup>[5,25]</sup> 基础上实现惯性融合; VINS-Mono 以光流作为前端, 构建视觉惯性紧耦合系统; VI-DSO 在算法上将 DSO 方法<sup>[9]</sup> 与惯性测量单元进行融合并使用了动态边缘化方法, VI-DSO 对比数据来源于文 [24]。

以上 VIO 算法的前端、后端方法有所差异, 因此在参数设置上均采用相应开源算法的默认值, 实验中使用无闭环的 VI-ORB、VINS-Mono 算法。在 5 个数据集上进行测试, 重复 10 次实验并对轨迹均方根误差求取平均值, 对比结果如表 2 所示。

由表 2 的数据可得, 本文方法在上述数据集上的精度与 VI-DSO 接近, 优于无闭环情况下的 VI-ORB 算法、无闭环的 VINS-Mono 算法和 OKVIS 算法。在 MH.01 数据集上, 本文方法得到的轨迹均方根误差为 0.046 m, 具有明显优势。本文算法轨迹与 VINS-Mono (无闭环)、VI-ORB (无闭环) 轨迹以及真实轨迹在 MH.01 数据集上的比较结果如

图 8 所示。

表 2 VIO 算法轨迹精度对比 (单位: m)

Tab.2 Comparison of trajectory accuracy of VIO algorithms

数据集	本文算法	VI-ORB (no loop)	VINS (no loop)	OKVIS (mono)	VI-DSO
MH.01	<b>0.046</b>	0.109	0.161	0.332	0.062
MH.02	0.056	0.083	0.178	0.361	<b>0.044</b>
MH.03	0.114	<b>0.094</b>	0.210	0.296	0.117
V1.01	0.085	0.096	0.086	0.108	<b>0.059</b>
V2.01	<b>0.061</b>	0.079	0.093	0.125	0.067

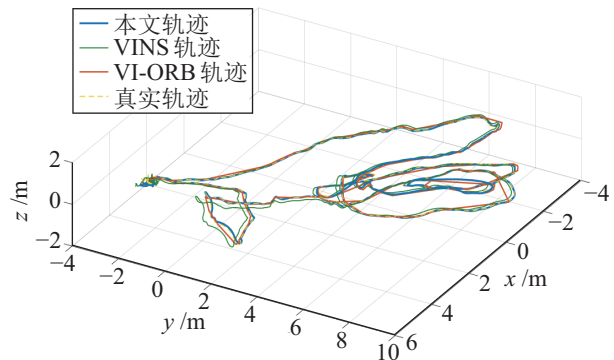


图 8 MH.01 数据集上的轨迹对比

Fig.8 Trajectory comparison on MH.01 dataset

相比于 SVO 算法<sup>[6-7]</sup>, 本文融合了 IMU 观测结果后, 轨迹精度提高了 3 倍并且运动跟踪更具鲁棒性。在与目前主流的 VIO 算法 (无闭环) 的比较中, 本文方法同样表现出一定优势。

## 4 结论与展望 (Conclusion and future work)

本文在 SVO 视觉前端上融合惯性测量结果, 提出一种半直接视觉惯性里程计算法。解决了单目视觉尺度缺失问题, 并设计一种 IMU 加权先验模型, 可在运动跟踪过程中提供精确的初值估计, 使得静止或者快速运动中均能保持鲁棒性。提出共视约束的滑动窗口优化方法以提高后端精度与优化效率并消除局部轨迹误差。在通用数据集上验证本文方法的可行性, 实验结果表明本文方法不仅有效地提高了运动估计精度与稳定性, 而且在与同类型 VIO 算法的比较中也具有较大优势。

本程序采用多线程计算, 将后端优化合并到运动跟踪线程中, 若改为多线程的计算方式, 则可进一步提高计算效率。对于联合初始化受到运动状态影响的问题, 需在现有方法基础上进一步考虑初始化起始条件的影响, 以提高参数估计的收敛性

能. 对于单目相机在运动过程中可能出现的尺度漂移问题, 可采用非线性优化的方法对尺度进行实时估计并校正, 以获取全局一致的运动轨迹与地图. 本文会进一步完善半直接视觉惯性里程计算法, 并将视觉与 IMU 的融合运用到重定位、回环检测与优化等模块中, 构建性能完备的半直接视觉惯性系统.

### 参考文献 (References)

- [1] Royer E, Lhuillier M, Dhome M, et al. Monocular vision for mobile robot localization and autonomous navigation[J]. *International Journal of Computer Vision*, 2007, 74(3): 237-260.
- [2] Moreno-Armendariz M A, Calvo H. Visual SLAM and obstacle avoidance in real time for mobile robots navigation[C]//*IEEE International Conference on Mechatronics, Electronics and Automotive Engineering*. Piscataway, USA: IEEE, 2014: 44-49.
- [3] Taketomi T, Uchiyama H, Ikeda S. Visual SLAM algorithms: A survey from 2010 to 2016[J]. *IPSA Transactions on Computer Vision and Applications*, 2017, 9(1): 16-26.
- [4] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//*6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. Piscataway, USA: IEEE, 2007: 225-234.
- [5] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163.
- [6] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2014: 15-22.
- [7] Forster C, Zhang Z, Gassner M, et al. SVO: Semidirect visual odometry for monocular and multicamera systems[J]. *IEEE Transactions on Robotics*, 2017, 33(2): 249-265.
- [8] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[C]//*13th European Conference on Computer Vision*. Berlin, Germany: Springer, 2014: 834-849.
- [9] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(3): 611-625.
- [10] Kelly J, Sukhatme G S. Visual-inertial simultaneous localization, mapping and sensor-to-sensor self-calibration[C]//*IEEE International Symposium on Computational Intelligence in Robotics and Automation*. Piscataway, USA: IEEE, 2009: 360-368.
- [11] Lupton T, Sukkariéh S. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions[J]. *IEEE Transactions on Robotics*, 2012, 28(1): 61-76.
- [12] Forster C, Carlone L, Dellaert F, et al. On-manifold preintegration for real-time visual-inertial odometry[J]. *IEEE Transactions on Robotics*, 2017, 30(1): 1-21.
- [13] Mur-Artal R, Tardos J D. Visual-inertial monocular SLAM with map reuse[J]. *IEEE Robotics and Automation Letters*, 2017, 2(2): 796-803.
- [14] 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.
- [15] Liu Y, Chen Z, Zheng W J, et al. Monocular visual-inertial SLAM: Continuous preintegration and reliable initialization[J]. *Sensors*, 2017, 17(11): 2613-2637.
- [16] Jones E S, Soatto S. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach[J]. *International Journal of Robotics Research*, 2011, 30(4): 407-430.
- [17] 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.
- [18] Leutenegger S, Lynen S, Bosse M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. *International Journal of Robotics Research*, 2015, 34(3): 314-334.
- [19] Kummerle R, Grisetti G, Strasdat H, et al. g<sup>2</sup>o: A general framework for graph optimization[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2011: 3607-3613.
- [20] 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.
- [21] Qin T, Shen S J. Robust initialization of monocular visual-inertial estimation on aerial robots[C]//*IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2017: 4225-4232.
- [22] Martinelli A. Closed-form solution of visual-inertial structure from motion[J]. *International Journal of Computer Vision*, 2014, 106(2): 138-152.
- [23] Kaiser J, Martinelli A, Fontana F, et al. Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation[J]. *IEEE Robotics and Automation Letters*, 2017, 2(1): 18-25.
- [24] von Stumberg L, Usenko V, Cremers D. Direct sparse visual-inertial odometry using dynamic marginalization[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2018: 2510-2517.
- [25] Mur-Artal R, Tardos J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255-1262.

### 作者简介:

龚赵慧 (1996 -), 女, 硕士生. 研究领域: 多源融合导航, 视觉 SLAM.

张霄力 (1970 -), 男, 博士, 副教授. 研究领域: 组合导航技术, 非线性系统和鲁棒控制.

彭侠夫 (1963 -), 男, 博士, 教授. 研究领域: 控制与导航技术, 运动体智能装置与系统.