

引用格式:成坚炼,韩军,张国强. 基于半直接法视觉里程计的单目视觉惯性系统[J]. 微电子学与计算机,2021,38(4):35-39.
[CHENG J L,HAN J,ZHANG G Q. Monocular visual inertial system based on semi-direct visual odometry [J]. Microelec-
tronics & Computer,2021,38(4):35-39.]

基于半直接法视觉里程计的单目视觉惯性系统

成坚炼^{1,2},韩 军^{1,2},张国强^{1,2}

(1 上海大学 通信与信息工程学院,上海 200444; 2 上海先进通信与数据科学研究院,上海 200444)

摘 要: 针对单目视觉系统姿态估计精度不高的问题,提出了一种基于半直接法视觉里程计(SVO)的单目视觉惯性系统.系统使用误差状态卡尔曼滤波器以松耦合方式对视觉信息和惯性导航信息进行融合.视觉前端采用 SVO 处理图像数据,并修改了 SVO 的关键帧选择策略以及解决了相机前视场景中由于重定位失败导致姿态估计失败的问题.使用 EuRoc 数据集中的 MH 01 子集对所提算法进行验证.实验表明,基于 SVO 的单目视觉惯性系统能够有效的提高姿态估计精度.相比单目 SVO 算法,所提算法的位置误差减少了 13.4%,旋转角的误差减少了 30%.

关键词: 视觉惯性系统;半直接法视觉里程计;误差状态卡尔曼滤波器;姿态估计

中图分类号: TP274

文献标识码: A

文章编号: 1000-7180(2021)04-0035-05

Monocular visual inertial system based on semi-direct visual odometry

CHENG Jian-lian^{1,2},HAN Jun^{1,2},ZHANG Guo-qiang^{1,2}

(1 College of Communication and information Engineering, Shanghai University, Shanghai 200444, China;
2 Shanghai Institute of Advanced Communications and Data Science,Shanghai 200444, China)

Abstract: Aiming at the problem of low attitude estimation accuracy of monocular visual system, a monocular visual inertial system based on semi-direct visual odometry (SVO) is proposed. The system uses an error-state Kalman filter to fuse visual information and inertial navigation information in a loosely coupled manner. SVO is used to process image data in the visual front-end, and the SVO keyframe selection strategy is modified to solve the problem of attitude estimation failure due to relocation failure in the front-view scene of the camera. The proposed algorithm is validated by using the MH 01 subset of the EuRoc dataset. The experiment shows that the monocular visual inertial system based on the SVO can effectively improve the accuracy of attitude estimation. Compared with the monocular SVO algorithm, the position error of the proposed algorithm is reduced by 13.4%, the error of the rotation angle is reduced by 30%.

Key words: Visual inertial system; Semi-direct Visual Odometry; error-state Kalman filter; attitude estimation

1 引言

单目相机由于体积小、成本低,常用来估计位姿^[1].但是单目视觉导航系统估计的三维位置与实际位置存在一个未知的尺度关系.而同样体积小、成本低的惯性测量单元(Inertial Measurement Unit, IMU)估算出的位置信息能够解决上述问题.因此

单目相机与 IMU 融合的姿态估计方法成为了近年来的研究趋势^[2].单目视觉惯性系统利用了单目相机和 IMU 的互补特性,基于融合两种传感器测量数据的方法,具有更好的鲁棒性以及更好的估计精度.

根据处理视觉和 IMU 信息融合方式的不同可以分为紧耦合和松耦合^[3-5].紧耦合是把视觉信息和

IMU 信息合并起来共同估计姿态. 文献[6]提出了一种基于滑动窗口优化的单目视觉惯性系统-VINS-Mono. 系统可以在未知环境下进行初始化, 具备回环检测和重定位模块, 估计精度高. 紧耦合考虑了视觉信息和 IMU 信息之间的相关性, 精度高, 但计算量大. 相反松耦合将视觉算法作为一个独立的模块, 估算结果仅用于更新状态. 本文提出了一种基于 SVO^[7] 的视觉惯性系统的姿态估计方法. SVO 称为半直接法视觉里程计, 它结合了特征点法和直接法的特性. 考虑到相机前视的场景, 设计了 SVO 的关键帧选择策略以及解决由于重定位失败导致姿态估计失败的问题. 使用误差状态卡尔曼滤波器 (Error-State Kalman Filter, ESKF) 对 SVO 的输出结果和惯性导航信息进行松耦合, 状态向量中加入了对尺度因子的估计.

2 系统框架

本文提出的视觉惯性系统采用了 ESKF. 整个框架分为预测与更新两步, 如图 1 所示. 利用 IMU 提供的加速度测量值 a_m 和角速度测量值 ω_m 对误差状态和标准状态进行预测; 并使用 SVO 算法算出无人机的位置 p 和姿态 q 作为观测值; 使用观测值和标准状态的预测值计算出观测误差, 其中 K 为卡尔曼增益; 将观测误差插入到标准误差用来更新标准误差; 最后重置误差状态.

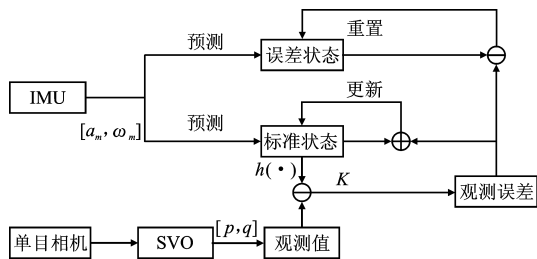


图 1 系统框架

3 视觉里程计设计

SVO 首先提取图像中的 FAST 特征点, 然后使用直接法根据关键点周围的图像信息估计相机的运动. SVO 不需要计算特征点的描述子, 它保留了直接法处理速度快的特点, 在嵌入式系统下也能达到 55 帧每秒^[7].

SVO 算法分为两个线程: 一个线程用于估计相机运动, 另一个线程用于构建地图. 运动估计部分包含了基于稀疏模型的图像对齐、特征对齐、位姿和结构优化 3 个步骤. 建图部分使用深度滤波器对特征

点的深度进行估计, 在特征点的深度收敛后, 将这些特征点添加到地图中. SVO 算法的初始化过程以及关键帧选择策略是针对相机下视场景设计的, 对于相机前视场景, 本文借鉴了文献[8]中 SVO 的初始化过程; 并修改了关键帧选择策略以及解决由于重定位失败导致姿态估计失败的问题.

3.1 初始化过程

SVO 算法主要针对相机朝下的场景设计的. 在无人机视觉传感器朝向地面的场景中, 场景多为平面, 相机距离地面的高度可控. 此时 SVO 初始化过程选取视差大的两帧图像作为初始关键帧用来计算对应的单应矩阵. 但是视觉传感器朝前时, 场景深度分布较为复杂, SVO 初始化质量不高, 影响了姿态估计的精度.

修改后的 SVO 初始化增加对基础矩阵的计算. 在确定好用于初始化的两帧图像之后, 使用光流法跟踪图像中的特征点; 利用随机抽样一致 (RANSAC) 算法计算单应矩阵和基础矩阵, 并统计两个矩阵的内点数量; 最后选择内点数量多的矩阵.

3.2 关键帧选择

在 SVO 中关键帧选择策略为: 当前图像帧距离上一个关键帧的平移量超过场景深度的 12%, 当前帧就被选为关键帧. 这使得关键帧的选择在很大程度上依赖于场景的平均深度. 文献[9]中的关键帧选择策略为: 位姿和结构优化结束后, 当所跟踪的特征点丢失比例超过 15% 时, 将当前帧选取为关键帧. 但是实验发现在上一过程 (即特征对齐步骤) 特征点丢失严重. 因此本文设计的关键帧选择策略与特征点丢失比例无关, 设置为: 当特征点像素坐标的位置移动量 D 超过设定的阈值时, 就将当前帧选为关键帧.

$$D = \frac{1}{n} \sum_{i=1}^n \rho(p1_i, p2_i) \quad (1)$$

式中, $(p1_i, p2_i)$ 是第 i 对特征点, $\rho(\cdot)$ 为求欧式距离的函数.

3.3 解决重定位失败问题

SVO 中触发重定位的条件为: 重定位的触发条件为: 其一, 经过特征对齐步骤优化后的特征点数量 n_{obs} 小于阈值; 其二, 丢失的特征点数量 n_{dif} 大于阈值. 在 Ubuntu 系统下使用 SVO 运行 EuRoc 数据集, 经过多次实验发现: 在特征对齐部分特征丢失严重, 丢失的特征点数量大于阈值触发重定位. 但是在重定位过程中, 由于图像中匹配的特征点数量少或者在特征对齐步骤中丢失大部分特征点, 导致没有足

够的特征点实现重定位,从而使得重定位失败.如图2所示,在特征跟踪过程中丢失42个特征,大于阈值40,触发重定位,又因为没有足够的特征点导致重定位失败,从而使得算法无法估计姿态.重定位失败的因素有特征点数量、特征点对齐、以及重定位触发条件,因此修改方案有:

1) 增加特征点数量.由于投影到当前帧上的特征点数目 $n_{\text{mat}} = \min(n_{\text{mat}}, n_{\text{max}})$,所以参数 n_{max} (重投影到当前帧上的特征点最大数目)限制了重投影到当前帧上特征点的数目,通过增大 n_{max} 数值使得投影到当前帧上点的数目得到增加.

2) 改进特征对齐判断条件.在设定的迭代次数内,当误差小于阈值时,认为找到最佳投影点,退出迭代.由于算法追求高精度,判断特征对齐的条件比较严格,导致在迭代次数内没有找到最佳投影点以致于两帧图像上相近的特征点被丢掉.因此可以通过降低迭代终止条件,减少丢失的特征点数量.

3) 修改重定位触发条件.修改触发重定位的第二个条件,修改为:当丢失的特征点数量 n_{df} 占 n_{mat} 的比例 $rate$ 大于阈值时触发重定位.由于 n_{df} 减小, n_{mat} 增大, n_{obs} 很难小于阈值,并且 $rate$ 远远小于阈值,这大大地减少了触发重定位的次数.

$$n_{\text{df}} = n_{\text{mat}} - n_{\text{obs}}, rate = \frac{n_{\text{df}}}{n_{\text{mat}}} \quad (2)$$

```
WARN] [1590062058.609129967]: Lost 42 features!
WARN] [1590062058.655827565]: Relocalizing frame
WARN] [1590062058.667219278]: Not enough matched features.
WARN] [1590062059.655839596]: Relocalizing frame
WARN] [1590062059.717508059]: Not enough matched features.
WARN] [1590062060.703933974]: Relocalizing frame
WARN] [1590062060.764255079]: Not enough matched features.
WARN] [1590062061.706987855]: Relocalizing frame
WARN] [1590062061.775694758]: Not enough matched features.
WARN] [1590062062.768690780]: Relocalizing frame
WARN] [1590062062.779463750]: Not enough matched features.
WARN] [1590062063.807197585]: Relocalizing frame
WARN] [1590062063.813689670]: Not enough matched features.
WARN] [1590062064.857272152]: Relocalizing frame
```

图2 重定位失败实验图

4 视觉惯性系统设计

在ESKF中有三个状态分别是:真实状态、标准状态、误差状态.真实状态被认为是由标准状态和误差状态组成.因为标准状态不考虑噪声和其他扰动,而真实状态下信号是存在噪声和扰动,所以标准状态和真实状态之间的误差由误差状态收集并通过ESKF估计.

4.1 滤波器预测阶段

系统方案中定义了三个坐标系:IMU坐标系、相机坐标系、世界坐标系分别用下标字母 i, c, w 区分. IMU和相机为固连状态,则认为它们之间的相

对位姿关系恒定.由于单目视觉里程计估计的位移与实际位移存在一个未知比例关系,因此在状态量中引入一个尺度因子 λ .系统的标准状态 x 可设为:

$$x = \{p_i^w, v_i^w, q_i^w, b_a, b_{\omega}, \lambda\}^T \quad (3)$$

式中, p_i^w, v_i^w, q_i^w 分别表示在世界坐标系下的IMU的位置、速度和姿态, b_a 和 b_{ω} 分别表示加速度偏差和角速度偏差. IMU的测量值存在偏差和噪声干扰,真实值与测量值之间的关系有: $a = a_m - b_a - n_a$, $\omega = \omega_m - b_{\omega} - n_{\omega}$,其中 n_a 和 n_{ω} 分别表示加速度计和陀螺仪的测量噪声,并且它们为高斯白噪声.在标准状态不考虑噪声和其他扰动,因此标准状态的差分方程可以写成:

$$p_{n+1} = p_n + v_n \Delta t + \frac{1}{2} (R_n (a_{m_n} - b_{a_n}) - g) \Delta t^2 \quad (4)$$

$$v_{n+1} = v_n + (R_n (a_{m_n} - b_{a_n}) - g) \Delta t \quad (5)$$

$$q_{n+1} = q_n \otimes q \{(\omega_{m_n} - b_{\omega_n}) \Delta t\} \quad (6)$$

公式(4~6)用来预测标准状态,其中 R_n 表示 n 时刻四元数 q 对应的旋转矩阵; g 为世界坐标系下的重力加速度; $q\{\theta\}$ 表示欧拉角对应的四元数.因为标准状态下不考虑噪声和干扰,所以加速度偏移 b_a 、角速度偏移 b_{ω} 、尺度因子 λ 与上一刻的值相同,这对于高频率的IMU数据来说,能够减少标准状态的预测时间. ESKF的核心在于对误差状态进行预测和更新,而经典KF则是始终针对标准状态进行预测和更新.对误差状态进行预测和更新的好处在于误差状态线性化更有效,可以容易忽略泰勒展开的二阶项.误差状态用 δx 表示.使用四元数描述的状态向量不是最小维表征,存在参数冗余.为了避免因此引起的协方差矩阵奇异性问题,引入四元数的半角形式,有:

$$\delta q = \left[\cos\left(\frac{\theta}{2}\right) u \sin\left(\frac{\theta}{2}\right) \right]^T \approx \left[1 \quad \frac{1}{2} \delta \theta \right]^T \quad (7)$$

系统的误差状态可定义为:

$$\delta x = \{\delta p_i^w, \delta v_i^w, \delta \theta_i^w, \delta b_a, \delta b_{\omega}, \delta \lambda\}^T \quad (8)$$

根据误差状态的定义,可以得到离散时间下误差状态的差分方程:

$$\delta p_{n+1} = \delta p_n + \delta v_n \Delta t \quad (9)$$

$$\delta v_{n+1} = \delta v_n - (R_n [a_{m_n} - b_{a_n}] \times \delta \theta_n + R_n \delta b_a) \Delta t + v_i \quad (10)$$

$$\delta \theta_{n+1} = R_n^T \{\omega_{m_n} - b_{\omega_n}\} \Delta t \delta \theta_n - \delta b_{\omega_n} \Delta t + \theta_i \quad (11)$$

$$\delta b_{a_{n+1}} = \delta b_{a_n} + a_i \quad (12)$$

$$\delta b_{\omega_{n+1}} = \delta b_{\omega_n} + \omega_i \quad (13)$$

$$\delta\lambda_{n+1} = \delta\lambda_n \quad (14)$$

式中, Δt 为 IMU 的采样周期, 形如 $[\mathbf{a}]_{\times}$ 表示向量 $\mathbf{a} = [a_1, a_2, a_3]$ 的叉乘矩阵, 定义为:

$$[\mathbf{a}]_{\times} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \quad (15)$$

误差状态方程可以写为:

$$\delta x_{n+1} = F\delta x_n + F_i i \quad (16)$$

式中, F 和 F_i 可以由公式 (9 ~ 14) 得到, $i = [v_i \theta_i a_i \omega_i]^T$, 矩阵内各个分量分别表示速度、角度、加速度和角速度上的高斯白噪声. 状态协方差矩阵为:

$$P_{n|n-1} = F_d P_{n-1|n-1} F_d^T + F_i Q_i F_i^T \quad (17)$$

$$F_i Q_i F_i^T =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_a^2 \Delta t^2 I & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\omega^2 \Delta t^2 I & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{b_a}^2 \Delta t^2 I & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{b_\omega}^2 \Delta t^2 I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (18)$$

4.2 滤波器更新阶段

单目视觉里程计得到的位置和姿态用于滤波更新, 与 IMU 的数据融合得到更加精确的位姿. 定义:

$$\tilde{\mathbf{z}} = \begin{bmatrix} \tilde{z}_p \\ \tilde{z}_q \end{bmatrix} = \mathbf{z} - h(\mathbf{x}) \quad (19)$$

式中, $\tilde{z}_p$ 和 $\tilde{z}_q$ 分别表示位置和四元数的测量误差, 则有:

$$\tilde{z}_p = z_p - \hat{z}_p = (p_i^w + R(q_i^w) p_c^i) \lambda - (\hat{p}_i^w + \hat{R}(q_i^w) \hat{p}_c^i) \hat{\lambda} \quad (20)$$

$$\tilde{z}_q = (\hat{q}_i^w \otimes \hat{q}_c^i)^{-1} \otimes (q_i^w \otimes q_c^i) \quad (21)$$

矩阵 $\mathbf{H}$ 与误差状态相关, 可以使用链式法则获得^[10]. 接下来计算卡尔曼增益 K 、更新标准状态 $x_{n|n}$ 和误差状态协方差矩阵 $P_{n|n}$:

$$K = P_{n|n-1} \mathbf{H}^T (\mathbf{H} P_{n|n-1} \mathbf{H}^T + \mathbf{R}_m)^{-1} \quad (22)$$

$$x_{n|n} = x_{n|n-1} + K \tilde{\mathbf{z}} \quad (23)$$

$$P_{n|n} = (I - K \mathbf{H}) P_{n|n-1} \quad (24)$$

5 实验结果

使用 EuRoc 数据集中的 Machine Hall01 子集对提出的算法进行有效性验证, 并与 VINS-Mono 算法进行比较. EuRoc 数据集包含 IMU、LEICA、双目摄像头等传感器数据. Ground Truth 数据是由

Leica Nova MS50 提供, 其精度为毫米级. 本文仅使用左目摄像机 (cam0) 拍摄的图像和 IMU 数据. 所提算法修改了 SVO 的关键帧选择策略以及解决了重定位失败的问题, 最终使得 SVO 可以正常完整的运行 EuRoc, 由此验证修改方案的有效性.

实验之前给定初始值. 位置初始值 p_i^w 、 p_c^i 、 q_c^i 以及 IMU 传感器中的噪声信息由数据集提供, 姿态初始值设置为单位四元数. 速度、加速度偏差和角速度偏差的初始值设置为 0. 尺度设置为 1. 世界坐标系的 z 轴是垂直地面, 设置为 $g = [0, 0, 9.8]^T$, $\Delta t = 0.005$ s. 在实验开始, 播放 MH_01 的 ROS 数据包 MH_01_easy.bag, 发布左目相机图像序列的 ROS 话题 /cam0/image_raw, SVO 订阅 /cam0/image_raw.

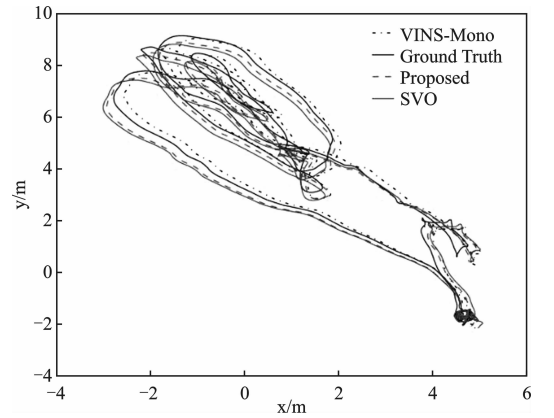


图3 飞行轨迹

将估算的结果转换到 Ground Truth 所在的参考系下进行比较. 无人机的飞行轨迹如图 3 所示. 基于改进 SVO 的视觉惯性系统、VINS-Mono 以及 SVO 的平均位置误差如表 1, 三种算法在 x 轴和 y 轴上平均误差稍大, 但在 z 轴上的平均误差小, 这是因为无人机在 z 轴的运动小于其他两个轴. 本文算法在 x 轴和 y 轴上的位置估计误差比 VINS-Mono 小, 在三个轴上的位置估计误差都小于 SVO. 总之, 本文算法、VINS-Mono 和 SVO 的位置误差分别为 0.207 m, 0.222 m, 0.239 m, 验证了所提算法的有效性.

表1 平均位置误差

位置误差(m)	本文算法	VINS-Mono	SVO
x	0.095	0.124	0.126
y	0.176	0.180	0.187
z	0.052	0.037	0.069

姿态估计的结果为四元数, 将其转换为绕三个

坐标轴旋转的角度进行比较,结果如表 2. 本文算法的姿态估计误差与 VINS-Mono 相近,与 SVO 相比,角度误差分别减少了 33. 3%,30. 8%,35. 0%. 总之,基于 SVO 的视觉惯性系统比仅依靠单目相机的 SVO,位姿估计精度得到整体提升.

表 2 平均姿态误差			
角度误差(°)	本文算法	VINS-Mono	SVO
x	0. 917	1. 018	1. 375
y	0. 507	0. 347	0. 733
z	0. 764	0. 755	1. 176

6 结束语

本文提出了一种基于 SVO 的视觉惯性系统,并修改了 SVO 的关键帧选择策略以及解决了重定位失败的问题以适应相机前视场景. 融合了视觉信息和惯性信息的视觉惯性系统相比单目视觉系统具有更高的位置和姿态估计精度. 但是所提算法并没有改变 SVO 的重定位策略以及 SVO 没有回环检测这个问题,未来可以对上述问题做进一步研究.

参考文献:

[1] CLEMENT L, KELLY J, BARFOOT T D. Monocular visual teach and repeat aided by local ground planarity[M]//WETTERGREEN D S, BARFOOT T D. Field and Service Robotics. Cham: Springer, 2016: 547-561. DOI: 10. 1007/978-3-319-27702-8_36.

[2] YE B. Monocular visual inertial odometry for quadrotor systems[D]. Hanzhou: Zhejiang University, 2017.

[3] MUR-ARTAL R, TARDO S J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803. DOI: 10. 1109/LRA. 2017. 2653359.

[4] YANG Z F, SHEN S J. Monocular visual-inertial

state estimation with online initialization and camera-IMU extrinsic calibration[J]. IEEE Transactions on Automation Science and Engineering, 2017, 14(1): 39-51. DOI: 10. 1109/TASE. 2016. 2550621.

[5] 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. DOI: 10. 1109/TRO. 2016. 2597321.

[6] 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. DOI: 10. 1109/TRO. 2018. 2853729.

[7] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry [C]//2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014: 15-22. DOI: 10. 1109/ICRA. 2014. 6906584.

[8] TANG L. Design and implementation of semi-direct based monocular visual odometry [D]. Chongqing: Chongqing University, 2018.

[9] XIAO Y, RUAN X G, ZHU X Q, et al. Integrated navigation method of monocular SVO/INS for micro aerial vehicles[J]. Journal of Chinese Inertial Technology, 2019, 27(2): 211-219. DOI: 10. 13695/j. cnki. 12-1222/o3. 2019. 02. 011.

[10] SOLÀ J. Quaternion kinematics for the error-state Kalman filter[J]. arXiv:1711. 02508, 2017.

作者简介:

成坚炼 男,(1994-),硕士研究生.研究方向为视觉导航技术. E-mail:jlchan@shu. edu. cn

韩 军 男,(1965-),博士,副教授.研究方向为图像处理和无人机电力线巡检等方向的研究.

张国强 男,(1995-),硕士研究生.研究方向为 SLAM 技术.