

基于改进关键帧筛选的多状态约束卡尔曼滤波

修瑾智^{1,2}, 方 针^{1,3}, 彭 慧^{1,2}, 陈燕苹^{1,2}, 邹梦强^{1,2}, 刘 宇^{1,2}, 杨诚霖^{1,2}, 王 森^{1,2}

(1. 重庆邮电大学 自主导航与微系统重庆市重点实验室, 重庆 400065;

2. 重庆邮电大学 智能传感技术与微系统重庆市高校工程研究中心, 重庆 400065;

3. 中国电子科技集团公司 第二十六研究所, 重庆 400060)

摘 要: 基于多状态约束卡尔曼滤波的融合算法仅利用单帧图像进行位姿估计, 若初始化不正确, 会导致视觉位姿估计发散严重; 若将每个视觉特征点都包含进系统状态向量, 则极易增加系统计算负担。针对上述问题, 提出了一种改进的关键帧选择算法, 利用多个视觉关键帧对相同特征点的约束来减小视觉测量误差, 提高定位精度, 同时只将关键帧解算出的相机位姿融入系统状态向量, 有效地降低了系统计算量。实验表明, 改进算法与 EKF 相比, 其定位精度和计算效率分别提升了 29.09% 和 32.2%。与 Orb-slam2 相比, 改进算法的计算效率提升了 35.48%。

关键词: 关键帧; 卡尔曼滤波; 融合算法; 视觉定位; 惯性定位

中图分类号: TN713 **文献标识码:** A

Multi-State Constrained Kalman Filtering Based on Improved Keyframe Filtering

XIU Jinzhi^{1,2}, FANG Zhen^{1,3}, PENG Hui^{1,2}, CHEN Yanping^{1,2},

ZOU Mengqiang^{1,2}, LIU Yu^{1,2}, YANG Chenglin^{1,2}, WANG Sen^{1,2}

(1. Chongqing Key Lab. of Autonomous Navigation and Microsystems, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Chongqing Engineering Research Center of Intelligent Sensing Technology and Microsystem, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 3. The 26th Research Institute of China Electronics Technology Group Corporation, Chongqing 400060, China)

Abstract: The fusion algorithm based on multi-state constrained Kalman filtering solely uses a single frame image for pose estimation. If the initialization is incorrect, then it can cause severe divergence in visual pose estimation. Furthermore, each visual feature point in the system state vector can easily lead to computational burden to the system. Given the aforementioned problems, an improved keyframe selection algorithm is proposed, which uses multiple visual keyframes to constrain the same feature points for reducing visual measurement errors and improving positioning accuracy. Simultaneously, only the camera pose calculated from keyframes is integrated into the system state vector, which can effectively reduce system computation. The experiment shows that the improved algorithm enhances positioning accuracy and computational efficiency by 29.09% and 32.2%, respectively, when compared to EKF. Additionally, the proposed algorithm increases computational efficiency by 35.48% when compared to that of Orb-slam2.

Key words: keyframes; Kalman filter; fusion algorithm; visual positioning; INS

0 引言

视觉与惯导融合定位系统是一种结合了惯性测量单元(IMU)和视觉传感器的定位系统^[1]。IMU对细微的位姿变化敏感,可适应载体快速运动,但其误差随着时间累积而变大^[2],相机采集的图

像信息丰富,但在快速运动或受光照影响时,采集图像的纹理急剧变少,严重时将导致位姿估计失败,故而将相机与IMU进行融合,可互相弥补缺陷,使定位更准确。

对于视觉惯导融合定位中视觉数据与惯性数

收稿日期:2023-12-25

基金项目:国家自然科学基金资助项目(52175531, 62305039);重庆市自然科学基金项目资助(CSTB2023NSCQ-MSX0568, CSTB2022NSCQ-LZX0050, CSTB2023NSCQ-LMX0028, cstc2022ycjh-bgzxm0190)

作者简介:修瑾智(1998-),男,重庆市人,硕士生。

据的融合问题,一般采用扩展卡尔曼滤波(EKF)来解决系统的融合问题^[3]。在传统 EKF 框架下实现紧耦合的视觉惯性里程计(VIO)中,是将 IMU 的状态和观测到的特征点一起加入状态向量中进行状态估计^[4-6]。而在实际应用场景中,每帧图像的特征点数量都十分丰富,因此,这种方式易给系统带来计算上的负担。特征点相对于相机的位姿仅利用单帧图像进行估计,若初始化不正确,将导致视觉位姿估计严重发散。

针对传统 EKF 算法中计算量大及位姿估计发散严重的问题,本文提出了基于改进关键帧筛选策略的多状态卡尔曼滤波算法。利用多个视觉关键帧对相同特征点进行约束,同时只将关键帧解算出的相机位姿融入系统状态向量,在减少计算量的同时提升定位精度。

1 关键帧筛选

利用多个视觉关键帧对相同特征点的约束来减小视觉测量误差,同时只将关键帧解算出的相机位姿融入系统状态向量,以有效减少系统计算量,故本算法中关键帧的筛选十分重要。筛选关键帧过程中应保证在拥有丰富特征信息的前提下,关键帧的数目尽可能均匀,防止插入关键帧过多,给系统造成计算负担,但也不能插入太少,否则无法提供充足的视觉信息。关键帧筛选策略如下:

1) 相邻关键帧运动变化量判断

假设当前关键帧为 F_{cur} ,前一关键帧为 $F_{\text{cur}-1}$,则相邻帧之间的运动变化量 L 为

$$L = \|\Delta t\| + \min(2\pi - \|\Delta R\|^2) \quad (1)$$

式中 $\Delta t, \Delta R$ 分别为前后两帧之间的位姿变化量。

设定阈值为 L_{\max}, L_{\min} 。若 $L < L_{\min}$,则认为两帧之间位置变化不明显,为了防止关键帧插入过于密集,需要舍弃当前关键帧;若 $L > L_{\max}$,表明两帧之间位置变化太大,则舍弃当前关键帧,重新判断当前关键帧的下一帧;若 $L_{\min} < L < L_{\max}$,则认为前后关键帧运动变化量合适,可插入关键帧。

2) 相邻关键帧匹配度判断

完成上一步判断后,继续判断两帧之间匹配的特征点数目占比 K 为

$$K = \frac{M_{F_{C,k-1}}}{M_{F_k-1}} \quad (2)$$

式中: $M_{F_{C,k-1}}$ 为两帧之间匹配的特征点数; M_{F_k-1} 为上一关键帧中的特征点数。

设定阈值为 K_{\max}, K_{\min} 。当 $K_{\min} < K < K_{\max}$,说明当前关键帧和上一关键帧相似度合适,此时可插入当前关键帧。若 $K > K_{\max}$ 或 $K < K_{\min}$,则丢弃当前关键帧,重新判断下一帧。关键帧筛选算法流程如图 1 所示。

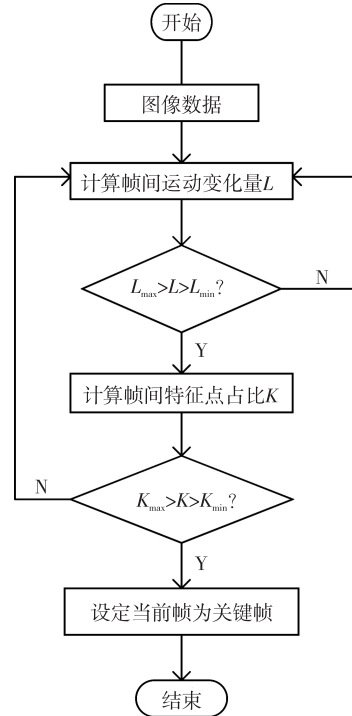


图1 关键帧筛选流程

2 多状态约束和观测模型建立

2.1 多状态约束模型建立

为减少计算量,本文只将关键帧解算得到的位姿信息加入定位算法的状态向量中,若 k 时刻有 N 个关键帧,则系统状态向量:

$$\hat{\mathbf{X}}_k = [\hat{\mathbf{X}}_{\text{IMU}_k} \quad {}^{C_1}\hat{\mathbf{q}}^T \quad {}^G\hat{\mathbf{p}}_{C_1}^T \quad \cdots \quad {}^{C_i}\hat{\mathbf{q}}^T \quad {}^G\hat{\mathbf{p}}_{C_i}^T] \quad (i=1, \dots, N) \quad (3)$$

式中: ${}^{C_i}\hat{\mathbf{q}}^T$ 为相机的姿态; ${}^G\hat{\mathbf{p}}_{C_i}^T$ 为相机的位置。

故系统误差状态向量:

$$\tilde{\mathbf{X}}_k = [\tilde{\mathbf{X}}_{\text{IMU}_k} \quad \delta\boldsymbol{\theta}_{C_1}^T \quad {}^G\tilde{\mathbf{p}}_{C_1}^T \quad \cdots \quad \delta\boldsymbol{\theta}_{C_i}^T \quad {}^G\tilde{\mathbf{p}}_{C_i}^T] \quad (i=1, \dots, N) \quad (4)$$

式中: $\delta\boldsymbol{\theta}_C^T$ 为旋转误差; ${}^G\tilde{\mathbf{p}}_C^T$ 为位置误差。

协方差矩阵:

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{P}_{\text{II}_{k|k}} & \mathbf{P}_{\text{IC}_{k|k}} \\ \mathbf{P}_{\text{IC}_{k|k}}^T & \mathbf{P}_{\text{CC}_{k|k}} \end{bmatrix} \quad (5)$$

式中: $\mathbf{P}_{\text{I}_{k|k}}$ 为 IMU 误差状态协方差矩阵; $\mathbf{P}_{\text{IC}_{k|k}}$ 为 IMU 误差状态与相机误差状态的关联协方差矩阵; $\mathbf{P}_{\text{CC}_{k|k}}$ 为相机误差状态协方差矩阵。

多状态约束模型中需要利用多个历史关键帧解算出的相机位姿与特征点进行约束。当该特征点跟踪丢失后,或历史关键帧数量达到滑动窗口阈值时,即可进行状态更新。

由于 IMU 与相机存在固定的外参,故可由 IMU 估计得到当前相机位姿:

$$\begin{cases} {}^c\hat{\mathbf{q}} = {}^c\bar{\mathbf{q}} \otimes {}^1\hat{\mathbf{q}} \\ {}^G\hat{\mathbf{q}}_c = {}^G\hat{\mathbf{p}}_I + \mathbf{C}_q^T \mathbf{p}_c \end{cases} \quad (6)$$

式中: ${}^c\bar{\mathbf{q}}$ 为惯性坐标系和相机坐标系之间的旋转; ${}^1\hat{\mathbf{p}}_c$ 为相机和 IMU 之间的平移。

将相机位姿加入状态向量中,使对应的协方差矩阵增广,则有:

$$\mathbf{P}_{k|k} = \begin{bmatrix} \mathbf{I}_{6N+15} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{k|k} \begin{bmatrix} \mathbf{I}_{6N+15} \\ \mathbf{J} \end{bmatrix}^T \quad (7)$$

2.2 观测模型建立与滤波更新

相机观测到的特征点 3D 坐标是由多个相机关键帧进行三角化估计所得,故可以用特征点与各关键帧之间的重投影误差来建立观测模型。

为了适应同时观测多个特征点的情况,建立了多个特征点的残差模型:

$$\mathbf{r}_o = \mathbf{H}_X \tilde{\mathbf{X}} + \mathbf{n}_o \quad (8)$$

使用正交三角分解法可将 \mathbf{H}_X 维数变小,则有:

$$\mathbf{H}_X = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{T}_H \\ 0 \end{bmatrix} \quad (9)$$

式中: $\mathbf{Q}_1, \mathbf{Q}_2$ 为正交矩阵; \mathbf{T}_H 为上三角矩阵。

对应的残差模型为

$$\mathbf{r}_o = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \begin{bmatrix} \mathbf{T}_H \\ 0 \end{bmatrix} \tilde{\mathbf{X}} + \mathbf{n}_o \quad (10)$$

将式(10)两边都乘以 $\mathbf{Q}^{-1} = \mathbf{Q}^T$ 后可得:

$$\begin{bmatrix} \mathbf{Q}_1^T \mathbf{r}_o \\ \mathbf{Q}_2^T \mathbf{r}_o \end{bmatrix} = \begin{bmatrix} \mathbf{T}_H \\ 0 \end{bmatrix} \tilde{\mathbf{X}} + \begin{bmatrix} \mathbf{Q}_1^T \mathbf{n}_o \\ \mathbf{Q}_2^T \mathbf{n}_o \end{bmatrix} \quad (11)$$

去除噪声项 $\mathbf{Q}_2^T \mathbf{r}_o$, 得到最终的残差模型:

$$\mathbf{r}_n = \mathbf{Q}_1^T \mathbf{r}_o = \mathbf{T}_H \tilde{\mathbf{X}} + \mathbf{n}_n \quad (12)$$

计算可得卡尔曼增益:

$$\mathbf{K} = \mathbf{T}_H^T (\mathbf{T}_H \mathbf{P} \mathbf{T}_H + \mathbf{R}_n)^{-1} \quad (13)$$

结合式(12)、(13)得到最终观测误差状态:

$$\Delta \mathbf{X} = \mathbf{K} \mathbf{r}_n \quad (14)$$

协方差矩阵:

$$\mathbf{P}_{k+1,k+1} = (\mathbf{I} - \mathbf{K} \mathbf{T}_H) \mathbf{P}_{k+1,k} (\mathbf{I} - \mathbf{K} \mathbf{T}_H)^T + \mathbf{K} \mathbf{R}_n \mathbf{K}^T \quad (15)$$

将误差状态与系统状态合成以修正累积误差:

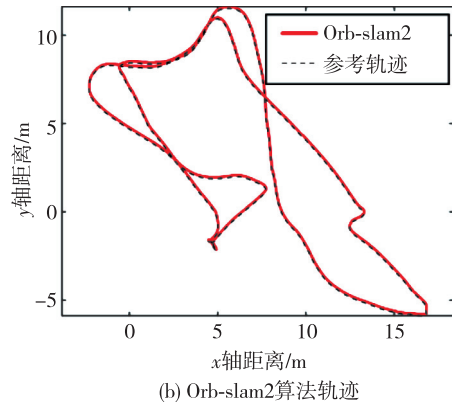
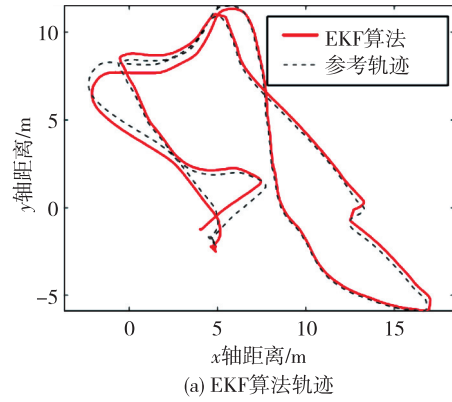
$$\hat{\mathbf{X}} = \hat{\mathbf{X}} \oplus \tilde{\mathbf{X}} \quad (16)$$

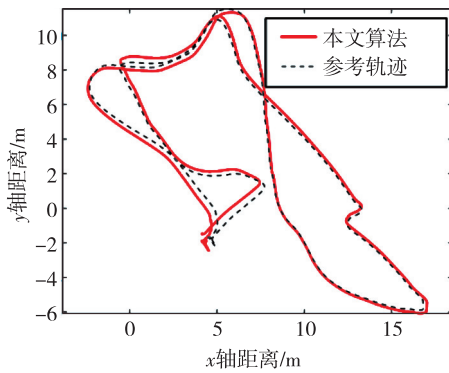
至此完成了滤波器更新的步骤,使用视觉观测对整个系统的误差进行了修正,并更新协方差矩阵将状态不断递推下去,获得最优的系统输出。

3 实验结果及分析

为验证所提算法的可行性,利用 EuRoc 数据集进行仿真实验。EuRoc 数据集包含了在真实环境中采集到的视觉和惯导数据,同时提供了载体的真实轨迹数据,以此作为判定标准来评估本文算法的性能^[7]。机房(MH)是 EuRoc 中最常用的数据,按照载体运行的抖动程度和幅度分为 easy、medium 和 difficult 3 种不同运动场景,其中 difficult 分为 MH_04 和 MH_05 两种场景。

为了评估本文所提算法的定位性能,将传统的 EKF 视觉惯导融合定位算法、Orb-slam2 算法和本文算法在 MH_05 场景下进行对比试验,3 种算法的轨迹对比如图 2 所示。





(c) 本文算法轨迹

图2 3种算法轨迹对比

为了评估3种算法的整体定位效果,对定位数据进行更加细致的分析,计算出3种算法在MH_05场景下的轨迹偏差最大值、最小值、平均值、标准差以及均方根误差,如表1所示。

表1 3种算法轨迹数据对比

算法	轨迹偏差/m				均方根 误差	CPU/%
	最大值	最小值	平均值	标准差		
EKF	0.986 5	0.159 8	0.397 3	0.211 7	0.474 6	59
Orb-slam2	0.168 0	0.003 5	0.042 9	0.027 6	0.051 0	62
本文算法	0.355 1	0.087 2	0.248 1	0.150 1	0.204 8	40

由表1可看出,本文所提定位算法的标准差和均方根误差比EKF融合定位算法分别低0.061 6 m和0.269 8 m,与EKF相比定位精度提升了29.09%,说明本文算法轨迹离散度更低,且更接近真实值。EKF算法仅利用当前图像帧进行位姿约束,不能很好地应对载体剧烈抖动,且特征点变多也会对系统造成计算负担,最终本文算法计算效率比EKF算法提升了32.2%。在定位精度上比添加了回环检测和地图点机制的Orb-slam2算法差,但计算效率较Orb-slam2算法提升了35.48%,降低了对硬件的性能要求。

4 结束语

本文提出了一种基于改进关键帧筛选的多状

态卡尔曼滤波算法,利用多个视觉关键帧对相同特征点的约束来减小视觉测量误差,同时只将关键帧解算出的相机位姿融入状态向量。筛选关键帧过程中保证了在拥有丰富特征信息的前提下,关键帧的数目尽可能均匀,防止插入关键帧过多,给系统造成计算负担,同时也不会因为插入太少,而不能提供充足的视觉信息。实验结果表明,与现有算法对比,本文算法在计算效率和定位精度上均有提升。

参考文献:

- [1] FENG G, WU W. Algorithm for monocular visual odometry/sins integrated navigation[J]. Journal of Chinese Inertial Technology, 2017, 19(3): 11-16.
- [2] 方针,曹加昇,崔巍,等. Mesh自组网与惯导组合的班组协同定位算法研究[J]. 压电与声光, 2021, 43(3): 385-390.
FANG Zhen, CAO Jiasheng, CUI Wei, et al. Research on team cooperative positioning algorithm based on Mesh ad hoc network and inertial navigation[J]. Piezoelectrics & Acoustooptics, 2021, 43(3): 385-390.
- [3] 杨梦佳. 惯导视觉SLAM技术综述[J]. 信息技术与信息化, 2019, 232(7): 213-215.
- [4] 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.
- [5] DING C, DAI Y, FENG X, et al. Stereo vision SLAM-based 3D reconstruction on UAV development platforms[J]. Journal of Electronic Imaging, 2023, 32(1): 013041.
- [6] 廖涛涛, 敬忠良, 李旻哲, 等. 基于李群的视觉/惯性自适应组合导航算法[J]. 信息与控制, 2020, 49(5): 585-590.
- [7] MUR-ARTAL R, TARDOS J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.