

doi:10.19306/j.cnki.2095-8110.2019.03.006

ORB-SLAM 系统优化框架分析概述

邱笑晨¹, 赵晨旭², 张 海¹, 许 辉³

- (1. 北京航空航天大学自动化学院, 北京 100083;
2. 北京航空航天大学中法工程师学院, 北京 100083;
3. 北京公共交通控股(集团)有限公司, 北京 100161)

摘要:首先阐述了 ORB-SLAM 系统的核心思想和基本框架,介绍了 SLAM 系统中光束平差法的应用实现,然后针对视觉里程计部分不同的关键帧位姿优化过程分别进行了详细阐述,并利用 KITTI 数据集的不同序列的实验结果进行比较研究,给出了核心的光束平差法在 ORB-SLAM 中的具体工作方式及性能分析,最后讨论和分析了 ORB-SLAM 的优势和不足。

关键词:视觉同步定位与地图构建;ORB-SLAM;光束平差法;位姿优化

中图分类号:TP391 文献标志码:A 开放科学(资源服务)标识码(OSID):
文章编号:2095-8110(2019)03-0045-07



Overview of ORB-SLAM System Optimization Framework Analysis

QIU Xiao-chen¹, ZHAO Chen-xu², ZHANG Hai¹, XU Hui³

- (1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100083, China;
2. Sino-French Engineer School of Beihang University, Beihang University, Beijing 100083, China;
3. Beijing Public Transport Holdings (Group) Co., Ltd., Beijing 100161, China)

Abstract: This paper firstly expounds the core idea, basic framework of the ORB-SLAM system and the application of bundle adjustment (BA) method in ORB-SLAM. Then, the different processes of keyframe pose optimization for the visual odometry part are elaborated separately, and the experimental results of different sequences of the KITTI dataset are analyzed and compared, the working mode and performance analysis of the core beam adjustment method in ORB-SLAM are presented. Finally, the advantages and disadvantages of ORB-SLAM are discussed and analyzed.

Key words: Visual Simultaneous Localization and Mapping; ORB-SLAM; Bundle adjustment; Pose optimization

0 引言

同步定位与地图构建^[1](Simultaneous Localization and Mapping, SLAM)是指搭载特定传感器

的运动主体,在没有先验环境信息的条件下,在运动过程中建立环境的模型,并估计自己的实时运动状态的方法^[2]。若这里的传感器为相机,则称之为视觉 SLAM。

收稿日期:2018-07-15;修订日期:2018-10-09

基金项目:国家重点研发计划(2016YFB0502004);国家重点研发计划(2017YFC0821102);北京市科技计划项目(Z171100000517006)

作者简介:邱笑晨(1990-),男,博士研究生,主要从事多传感器组合导航方面的研究。E-mail: ares43490@126.com

目前,主流的视觉 SLAM 方法按照估计框架可以分为基于滤波的方法和基于关键帧光束平差法(Bundle Adjustment, BA)^[3]。按照图像信息的利用程度可以分为基于特征点法和基于直接法(直接利用像素信息)。基于滤波的视觉 SLAM 的主要思想是将每一时刻 t 的系统状态用一个高斯概率模型表示,且系统状态由一个滤波器不断更新,从而得到不同的 SLAM 系统,其中最具代表性的是 Mono-SLAM^[2] 和 MSCKF^[4]。并行追踪与制图(Parallel Tracking and Mapping, PTAM)由 Klein 等^[5]于 2007 年提出,是最早期基于关键帧光束平差法的单目视觉 SLAM 系统,利用传统的特征点法实现。ORB-SLAM 是 Mur-Artal 等^[6]于 2015 年提出的基于 PTAM 算法框架改进的视觉 SLAM 系统,它利用了 ORB(Oriented FAST and Rotated, BRIEF)特征,是目前性能最好的单目视觉 SLAM 系统之一,其二代版本加入了双目和 RGB-D 模式,进一步提高了其对不同传感器配置的适应能力。DTAM^[7] 和 LSD-SLAM^[8-9] 是两种基于直接法的具有代表性的单目视觉 SLAM 算法,其主要特点是不依赖特征点的提取与匹配,通常在图像模糊、特征缺失等场景下鲁棒性更好。

从性能上,ORB-SLAM 具有良好的泛用性,支持单目、双目、RGB-D 等多种模式^[10];其回环检测算法保证了 ORB-SLAM 能够有效地抑制累计误差,并且利用相同手段实现的重定位功能使得系统在失效后能够迅速重定位;ORB-SLAM 使用 3 个线程完成系统,实现了快速跟踪和建图,能够保证轨迹和建图的一致性^[11]。

从功能上,可以将 ORB-SLAM 分为视觉里程计(Visual Odometry, VO)前端部分和回环(Loop Detection)后端部分。本文主要针对单目 ORB-SLAM 的系统架构进行简要介绍,以视觉导航中关键视觉里程计为对象,对影响系统精度的不同环节进行分析,结论中提出可能的改进措施。

1 ORB-SLAM 系统概述

ORB-SLAM 系统的算法部分主要由 3 个并行的线程组成,分别是:跟踪(Tracking)、局部建图(Local Mapping)和回环检测(Loop Closing)。此外,在初始化构建地图时会临时发起 2 个线程分别求解单应矩阵和本质矩阵,执行完毕后自行停止;在每次执行完回环位姿图优化后,会在回环检测线

程中临时发起 1 个线程,执行全局 BA 优化,该线程在执行完毕后自行停止,或者被另一次全局 BA 优化打断。本节主要介绍单目 ORB-SLAM 的具体架构及优化框架。

1.1 系统架构分析

ORB-SLAM 系统架构如图 1^[6]所示,结合图示分析 3 个主要线程的基本功能如下所示。

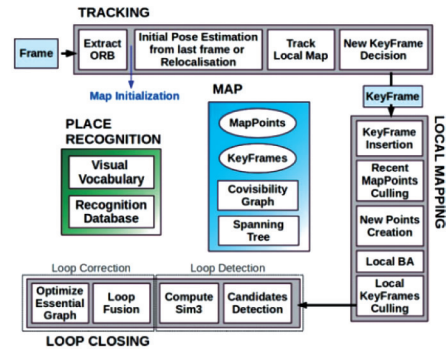


图 1 ORB-SLAM 系统架构

Fig. 1 Overview of ORB-SLAM system structure

1)跟踪:线程输入为每一帧图像,在未初始化时,尝试利用两帧图像进行初始化。初始化完成后,对获得的每一帧图像,通过特征描述子实现图像特征点与局部地图点的有效匹配,运用光束平差法来最小化重投影误差,从而优化当前帧相机位姿,实现每帧图像采集时刻相机的跟踪和定位。满足特定条件时,将当前帧确定为关键帧。

2)局部建图:线程的输入为跟踪线程插入的关键帧。基于新添加的关键帧,维护和拓展新的局部地图点,并运用光束平差法优化局部地图中所有关键帧的位姿及局部地图中的所有地图点。同时对关键帧进行筛选,剔除冗余关键帧。

3)回环检测:线程的输入为经过局部建图筛选过的关键帧。将当前关键帧的词袋(Bag of Words, BoW)向量存入全局词袋数据库中,从而加速后续帧的匹配。同时检测是否存在回环,若存在则通过执行位姿图(Pose Graph)优化来优化全体关键帧位姿,抑制累计漂移误差。在位姿图优化完成之后,会临时发起一个独立线程执行全局光束平差法,来得到整个系统最优结构(地图点)和运动(关键帧位姿)的结果^[12]。

1.2 单目 ORB-SLAM 中的光束平差法

ORB-SLAM 系统在若干个部分采用光束平差

法^[13]进行优化,包括在跟踪过程中优化当前帧的位姿(纯运动 BA),在局部建图线程中优化局部地图中的关键帧位姿和地图点位置(局部 BA),以及在回环检测后优化所有的关键帧位姿和地图点位置(全局 BA)。这三部分的主要内容分别如下:

1) 纯运动 BA: 针对通过特征描述子匹配的地图点 $\mathbf{X}^i \in \mathbf{R}^3$ 和图像特征点 $\mathbf{x}_m^i \in \mathbf{R}^2$, 通过最小化重投影误差, 优化相机当前帧相对于世界坐标系的旋转矩阵 $\mathbf{R} \in SO(3)$ 和平移矢量 $\mathbf{t} \in \mathbf{R}^3$ (两者组成了当前帧的位姿 $\mathbf{T} = [\mathbf{R} | \mathbf{t}]$), 上述过程由公式描述如下

$$\{\mathbf{R}, \mathbf{t}\} = \operatorname{argmin}_{\mathbf{R}, \mathbf{t}} \sum_i (\rho(\|\mathbf{x}_m^i - \pi_m(\mathbf{R}\mathbf{X}^i + \mathbf{t})\|_{\Sigma}^2)) \quad (1)$$

式中, ρ 是 Huber 代价函数; Σ 是协方差矩阵, 与特征点的尺度相关; π_m 是针孔相机投影函数, 定义如下

$$\pi_m([X \ Y \ Z]^T) = \begin{bmatrix} f_x \times X/Z + c_x \\ f_y \times Y/Z + c_y \end{bmatrix} \quad (2)$$

其中, (f_x, f_y) 是焦距, (c_x, c_y) 是主点坐标。

由于上述过程不对地图点位置进行优化, 相当于进行 PnP, 但 ORB-SLAM 作者称之为纯运动 BA。

2) 局部 BA: 同样通过最小化重投影误差, 同时优化当前关键帧位姿、与其存在共视关系的局部关键帧的位姿 \mathbf{T}_l 以及局部关键帧所能观测到的所有地图点的三维坐标 \mathbf{P}_l , 并将与局部关键帧形成共视但不与当前关键帧构成共视的关键帧的位姿 \mathbf{T}_F 作为约束项参与优化。优化的过程中不改变 \mathbf{T}_F , 从而使得优化结果关于全局地图存在一致性。上述过程由公式描述如下

$$\begin{aligned} \{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l | i \in \mathbf{P}_l, l \in \mathbf{T}_l\} = \\ \operatorname{argmin}_{\mathbf{X}^i, \mathbf{R}_l, \mathbf{t}_l} \sum_{j \in \mathbf{T}_l} \sum_{k \in \mathcal{X}_j} \rho(E(j, k)) \quad (3) \\ E(j, k) = \|\mathbf{x}_m^k - \pi_m(\mathbf{R}_j \mathbf{X}^k + \mathbf{t}_j)\|_{\Sigma}^2 \end{aligned}$$

其中, \mathcal{X}_j 为第 j 个局部关键帧匹配到地图点的特征点集合。

3) 全局 BA: 具体过程与局部 BA 类似, 其优化对象包括除了第一帧的其余所有关键帧位姿及全局地图中所有的地图点。初始帧位置被固定是为了保持全局一致性, 消除随机化^[12]。

2 系统精度分析

ORB-SLAM 作为基于关键帧 BA 的视觉 SLAM 系统, 在其运行过程中光束平差法的效果对于系统精

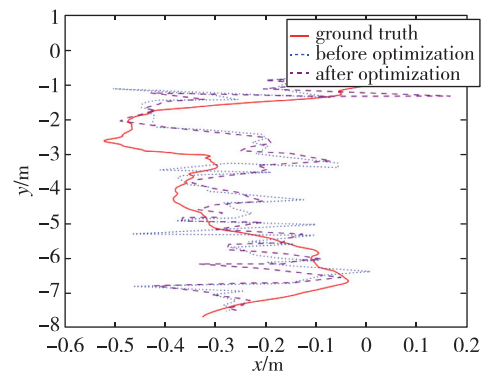
度的影响至关重要。在单目 ORB-SLAM 视觉里程计部分(不包括回环和全局 BA 后端), 针对关键帧位姿的优化共有三种情况: 第一种是当前帧成功匹配足够多的已有地图点后进行的第一次纯运动 BA, 优化得到当前帧的初始位姿; 第二种是成功更新当前帧共视的局部关键帧和局部地图点后进行的第二次纯运动 BA, 进一步优化当前帧的位姿; 第三种是当前帧被确定为关键帧后, 针对当前关键帧及其所对应的局部关键帧集合和局部地图点集合的局部 BA 优化, 实现局部地图的整体优化, 三种 BA 的优化作用分析如下。

2.1 首次纯运动 BA 优化

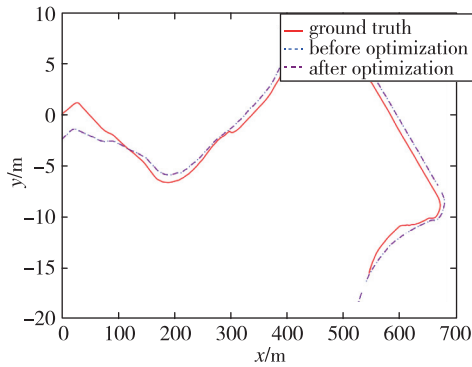
系统在成功初始化后, 会对新来的图像帧进行 ORB 特征点的提取和对应描述子的计算, 然后通过基于速度的模型或基于参考关键帧的模型进行图像特征点与已有地图点的匹配, 并获得当前帧的初始位姿。其中基于速度的模型主要适用于载体匀速运动的情况, 而基于参考关键帧的模型则适用于其他情况。在成功获得足够多的匹配点对后, 通过纯运动 BA 优化(相当于进行 PnP)当前帧位姿。此次纯运动 BA 的输入量是当前帧已得到的特征点与地图点的匹配关系, 输出量为优化后的当前帧位姿。

图 2 展示了 2 组来自 KITTI 数据集^[14]在不同场景下的数据, 在首次纯运动 BA 优化前后的关键帧位姿与真值进行尺度与坐标系对齐^[15]后, 得到的运动轨迹示意图。

通过对图 2(b) 中的轨迹曲线分析可知: 若 SLAM 系统在场景中的运动较为平滑, 则在纯运动 BA 优化前基于速度模型或者参考关键帧模型得到的位姿初值相比于真值的误差已经不是很大, 此时此次 BA 优化的效果并不显著。通过轨迹曲线图也可以看出, 优化前后的关键帧位置几乎重合。



(a) 基于 KITTI 数据集 004 序列得到的轨迹曲线图



(b) 基于 010 序列得到的轨迹曲线图

图 2 首次纯运动 BA 优化后的 KITTI 数据集轨迹曲线

Fig. 2 Trajectory curves of sequences of KITTI dataset after first motion—only BA optimization

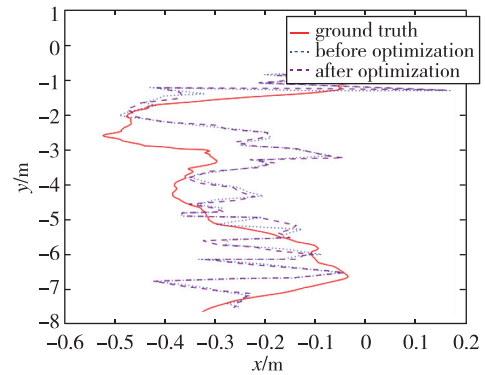
而从图 2(a)中的曲线图可以看出,由于实际场景中存在载体或传感器抖动,即便是在直路行驶的车辆,其位置在微观上也会存在抖动(注意横轴坐标间距相对纵轴而言很小);又由于传感器测量误差和算法等因素影响,其估计结果存在更加明显的抖动。但从图中可以看出,相比较而言纯运动 BA 优化前的抖动更加剧烈。纯运动 BA 能够有效地平滑估计结果,利用已有匹配地图点的位置约束修正当前帧的位姿,从而在一定程度上抑制积累误差,避免将过大的误差带入到后续的优化过程中。

2.2 第二次纯运动 BA 优化

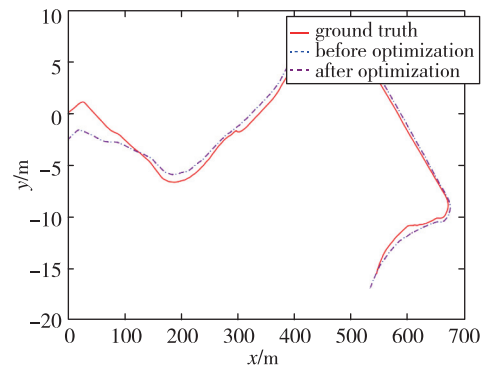
若当前帧的初始位姿被纯运动 BA 修正并剔除错误的匹配点对后,仍有足够多的正确匹配,系统则判定跟踪成功。此时系统会基于修正后的当前帧位姿尝试匹配更多的地图点,并更新当前帧的共视局部关键帧集合和局部地图点集合,然后基于这些局部信息进行第二次纯运动 BA 优化,进一步修正当前位姿。此次纯运动 BA 的输入量为更新后的当前帧所有特征点与地图点的匹配关系,输出量为进一步修正的当前帧位姿。

图 3 所示为基于与 2.1 节相同的数据集得到的第二次纯运动 BA 优化前后当前帧位姿与真值的比较结果。

由图 3(b)中的轨迹曲线可知,从宏观角度看,若当前帧的位姿误差已经小于系统修正误差的最小范围,即当前帧的位姿已经足够准确,则此次纯运动 BA 不再起到很明显的作



(a) 基于 KITTI 数据集 004 序列得到的轨迹曲线图



(b) 基于 010 序列得到的轨迹曲线图

图 3 第二次纯运动 BA 优化后的 KITTI 数据集轨迹曲线

Fig. 3 Trajectory curves of sequences of KITTI dataset after second motion—only BA optimization

而由图 3(a)的比较结果可以看出,每次优化前的当前帧位姿相比于真值已经没有特别大的偏差,但是相对优化后的结果而言,部分关键帧位置仍然存在着一些不合理的跳跃现象,主要是因为首次纯运动 BA 优化当前帧位姿时并未充分地利用所有潜在的匹配地图点的约束信息。因此,第二次纯运动 BA 优化当前帧位姿后,不仅进一步修正了误差较大的帧位姿,同时在有限的地图点位置约束下使不同时刻的当前帧的位姿的变化尽可能的光滑。此时较大的位姿误差已经得到尽可能的修正,接着会通过局部 BA 的优化调整所有局部关键帧的位姿及局部地图点的世界坐标系坐标,从而使得位姿变化更加光滑。

此处要求位姿变化是光滑的,是因为每次的当前关键帧位姿都是基于已被修正过的局部关键帧位姿得到的,在当前关键帧引入到系统的位姿误差被约束到合理范围的前提下,不同时刻的当前关键帧的位姿变化应当是平滑的。

2.3 局部 BA 优化

若第二次纯运动 BA 修正当前帧位姿后仍能得到足够多的有效匹配地图点对,且当前帧满足被设置成关键帧的条件,即当前帧与其局部关键帧的相同的地图匹配点数量不超过 90%,且满足局部建图线程空闲或距离上一次插入关键帧的时间超过阈值,当前帧便被设置成关键帧。接着基于当前关键帧三角化新的局部地图点并剔除不合理的局部地图点,然后针对局部地图中所有的关键帧及地图点进行局部 BA 优化。此次局部 BA 优化的输入量是当前关键帧所在的局部地图中所有的局部关键帧位姿和所有局部关键帧中特征点与地图点的匹配关系,同时利用所有与局部关键帧形成共视关系的帧位姿来约束优化结果,这些位姿在优化过程中保持不变,输出量是所有局部关键帧优化后的位姿和局部地图点优化后的三维坐标。

图 4 的三幅曲线图是基于同一数据集局部 BA 优化得到的轨迹曲线图的不同表现形式。其中(a)和(b)图的含义是:在一次共有 83 个关键帧生成的系统运行过程中,第 74 个和第 75 个关键帧分别插入时,当前所有的局部关键帧在局部 BA 优化前后的轨迹曲线图以及全部轨迹真值,而(c)图表示所有不同时刻的当前关键帧在局部 BA 优化前后的轨迹曲线以及真值。第一种方式侧重于横向展示多次局部 BA 对局部关键帧位姿的连续性影响,而第二种方式则侧重于纵向展示每次 BA 对当前新插入关键帧的直接性影响。

从图 4(c)的比较结果可以看出,当前关键帧的位姿在局部 BA 中得到了进一步的修正,因此轨迹曲线变得更加平滑。考虑到局部 BA 优化是将所有局部关键帧中局部地图点的重投影误差作为代价函数,即在有效修正位姿误差较大的关键帧的同时,不断将系统模型带来的误差平均到所有局部关键帧位姿上。而从图 4(a)和(b)的轨迹曲线可以看出,当前关键帧会在之后有限个关键帧插入时再次参与其局部 BA 的优化,从而不断平均系统位姿解算误差,使得最终系统输出的关键帧位姿曲线变得十分平滑,且每个关键帧位姿的误差均在一定的范围内,如图 5 所示。

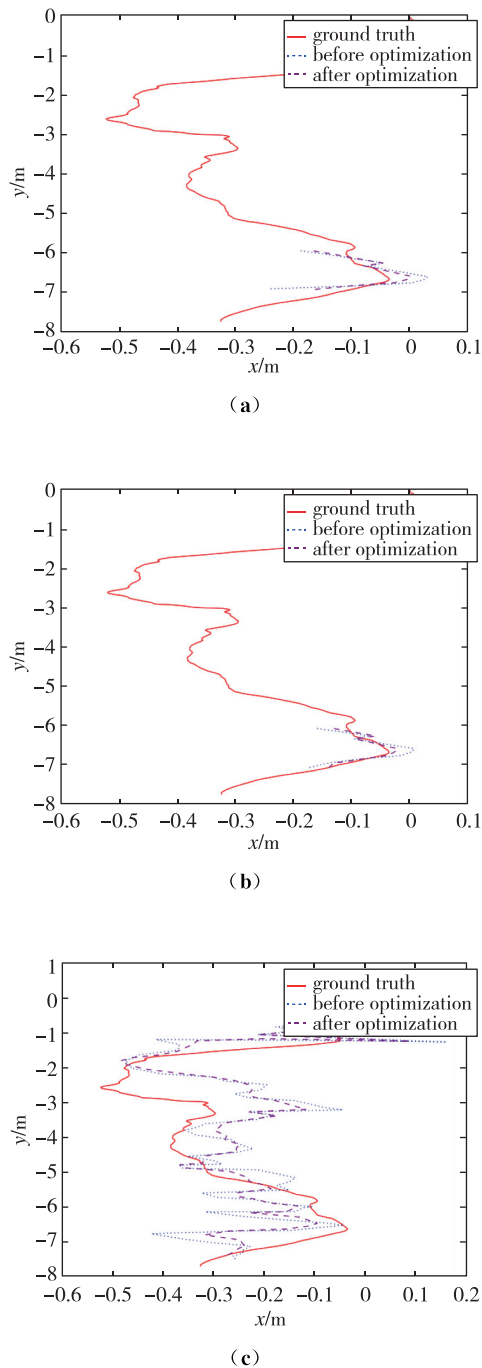


图 4 (a)和(b)表示基于 KITTI 数据集 004 序列得到的,第 74 帧关键帧和第 75 帧关键帧插入时所有局部关键帧优化前后的轨迹曲线示意图;(c)表示不同时刻的当前关键帧在局部 BA 优化前后的轨迹曲线图

Fig. 4 (a) and (b) respectively show the trajectory curves of all local keyframes before and after optimization when the 74th keyframe and the 75th keyframe insert based on the sequence 004 of KITTI dataset; (c) indicates the trajectory curve of the current keyframe before and after local BA optimization at different times

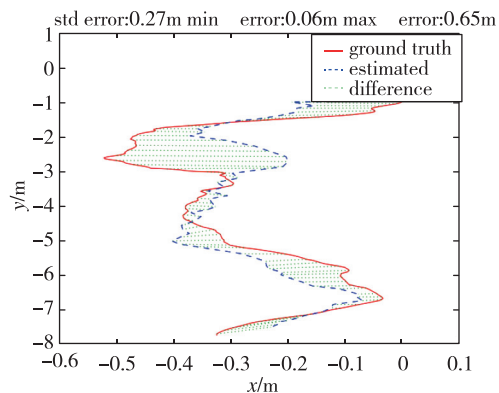


图5 基于KITTI数据集004序列, Visual Odometry部分得到的最终优化的全局关键帧轨迹示意图

Fig. 5 Final optimized keyframe trajectory graph obtained by the visual odometry part based on the 004 sequence of KITTI dataset

3 结论

作为基于关键帧光束平差法的代表方案, ORB-SLAM系统在有效修正误差较大的全局关键帧位姿的同时, 通过不断调整局部关键帧位姿和局部地图点在世界坐标系中的三维坐标, 在系统误差范围内保证位姿变化的连贯性和平滑性。此外, ORB-SLAM还通过各种约束条件和加速手段保证了系统的实时性和鲁棒性, 并通过回环检测的方法进一步精细化修正全局地图和关键帧位姿, 有效地提高了系统的性能和精度。

然而, 由于ORB-SLAM严重依赖特征点的提取与匹配效果, 因此在如纹理缺失的环境下运行会无法得到足够多的稳定匹配点对, 此时光束平差法缺乏足够的输入信息, 无法有效地修正位姿偏差。这种情况下, 可以引入外界的观测如惯性测量单元(Inertial Measurement Unit, IMU)等, 来约束和修正特征缺失带来的偏差, 针对单目情况, 还能得到绝对尺度。近年来也有越来越多的研究偏向于视觉/惯性视觉里程计(Visual/Inertial Odometry)方向^[16-18], 以期望解决纯视觉SLAM存在的问题。

参考文献

- [1] Gu Z, Liu H. A survey of monocular simultaneous localization and mapping[J]. CAAI Transactions on Intelligent Systems, 2015, 10(4): 499-507.
- [2] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007(6): 1052-1067.
- [3] 张海, 吴克强, 张晓鸥. 视觉导航技术的发展[J]. 导航定位与授时, 2017, 4(2): 1-8.
Zhang Hai, Wu Keqiang, Zhang Xiaou. The development of visual navigation [J]. Navigation Positioning and Timing, 2017, 4(2): 1-8 (in Chinese).
- [4] 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. IEEE, 2007: 3565-3572.
- [5] Klein G, Murray D. Parallel tracking and mapping on a camera phone[C]// Proceedings of IEEE International Symposium on Mixed and Augmented Reality, 2009.
- [6] 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.
- [7] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time [C]// 2011 International Conference on Computer Vision. IEEE, 2011: 2320-2327.
- [8] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]// European Conference on Computer Vision. Springer, Cham, 2014: 834-849.
- [9] Engel J, Sturm J, Cremers D. Semi-dense visual odometry for a monocular camera[C]// Proceedings of the IEEE International Conference on Computer Vision, 2013: 1449-1456.
- [10] Mur-Artal R, Tardós 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.
- [11] 高翔, 张涛. 视觉SLAM十四讲[M]. 北京: 电子工业出版社, 2017: 400.
Gao Xiang, Zhang Tao. Visual SLAM fourteen[M]. Beijing: Publishing House of Electronics Industry, 2017: 400(in Chinese).
- [12] 韩昊雯. ORB-SLAM2: 一种开源的VSLAM方案. Han Haowen. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. (in Chinese).
- [13] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment—a modern synthesis [C]// International Workshop on Vision Algorithms. Springer, Berlin, Heidelberg, 1999: 298-372.
- [14] Karlsruhe Institute of Technology. Object Detection Evaluation 2012 [DB/OL]. www.cvlibs.net/

- datasets/kitti/eval_object.php.
- [15] Umeyama S. Least-squares estimation of transformation parameters between two point patterns[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1991(4): 376-380.
- [16] 逯建军, 任晓军, 孙伟, 等. 惯性/双目视觉里程计深组合导航方法[J]. *导航定位与授时*, 2016, 3(3): 37-43.
Lu Jianjun, Ren Xiaojun, Sun Wei, et al. INS/stereo visual odometry deeply integrated navigation method [J]. *Navigation Positioning and Timing*, 2016, 3(3): 37-43(in Chinese).
- [17] 刘畅, 杨良军, 刘峰, 等. 一种基于惯性/视觉信息融合的无人机自主着陆导航算法[J]. *导航定位与授时*, 2016, 3(6): 6-11.
Liu Chang, Yang Liangjun, Liu Feng, et al. Navigation algorithm based on inertial/vision information fusion of UAV autonomous landing[J]. *Navigation Positioning and Timing*, 2016, 3(6): 6-11(in Chinese).
- [18] Yang Z, Shen S. 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.