

移动机器人视觉惯性 SLAM 研究进展

施俊屹¹, 查富生^{1,2}, 孙立宁¹, 郭伟¹, 王鹏飞¹, 李满天¹

(1. 哈尔滨工业大学机器人技术与系统国家重点实验室, 黑龙江 哈尔滨 150080;

2. 深圳航天科技创新研究院, 广东 深圳 518057)

摘要: 在研究领域, 基于滤波和基于优化是两种实现视觉惯性 SLAM (同时定位与地图创建) 的主导方法. 本文基于这两种方法介绍视觉惯性 SLAM, 说明了视觉惯性 SLAM 的最新研究进展和关键问题, 对比了几种代表性的视觉惯性 SLAM 框架, 并对未来进行了展望.

关键词: 视觉惯性 SLAM; 移动机器人; 位姿估计; 滤波; 优化

中图分类号: TP24

文献标识码: A

文章编号: 1002-0446(2020)-06-0734-15

A Survey of Visual-Inertial SLAM for Mobile Robots

SHI Junyi¹, ZHA Fusheng^{1,2}, SUN Lining¹, GUO Wei¹, WANG Pengfei¹, LI Mantian¹

(1. State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150080, China;

2. Shenzhen Academy of Aerospace Technology, Shenzhen 518057, China)

Abstract: Filtering-based and optimization-based methods are the two leading methods of VI-SLAM (visual-inertial simultaneous localization and mapping) used in the research field. Firstly, VI-SLAM based on these two methods are introduced, and its latest research progress and key issues are illustrated. Furthermore, the representative frameworks of VI-SLAM are compared. Finally, the future of visual-inertial SLAM is discussed.

Keywords: visual-inertial SLAM (simultaneous localization and mapping); mobile robot; state estimation; filtering; optimization

1 引言 (Introduction)

移动机器人本质上研究的是机器人如何在变化环境中有效执行任务的问题, 环境感知与定位导航功能是其基础之一. 当前, SLAM (simultaneous localization and mapping) 方法是移动机器人领域解决环境感知与定位导航问题的有效手段之一. SLAM 包括移动机器人运动状态估计和传感器检测到的环境模型的构建 2 个部分. 其中, 移动机器人的状态可由它的位姿来描述, 而建图则是对移动机器人所处环境的信息如位置、障碍等的描述.

建立环境地图对移动机器人来说是非常必要的. 首先, 地图可以有效限制移动机器人状态估计时的误差. 位置估计会随着时间的推移逐渐产生漂移, 导致移动机器人无法有效地完成任务甚至造成灾难性后果. 因此, 可以通过引入一组可区分的路标点构成的地图并判断图像间的相似性, 使机器人

能够识别曾到达过的场景, 进而得到全局一致的轨迹^[1-3]. 另一方面, SLAM 作为一种较底层的技术, 往往是为上层应用提供信息的. 通过建立全局的地图, 移动机器人可以进行导航, 从地图中任意 2 个点之间寻找路径, 并控制自己到达目标点. 更进一步地, 使用者希望通过地图来获得周围环境的重建效果, 通过语义地图实现人与移动机器人之间的交互, 正因为有上述需求, 因此 SLAM 技术在移动机器人的不同应用中都能找到用武之地.

一个完整的 SLAM 框架包括传感器数据、前端、后端、回环检测与建图, 如图 1 所示. 其中, 前端将传感器的数据抽象成适用于估计的模型, 回环检测判断机器人是否经过已知的位置. 而后端接受不同时刻前端测量的位姿和回环检测的信息并对它们进行优化, 从而得到全局一致的轨迹. 建图则是根据状态估计得到的轨迹建立与任务要求相对应的地图. 通常, 仅含有前端和局部后端的框架被称

基金项目: 国家自然科学基金 (61773139); 国家自然科学基金委员会创新研究群体科学基金 (51521003); 深圳市科技计划 (JCYJ20190813171009236); 深圳市海外高层次人才创新创业专项资金 (KQTD2016112515134654).

通信作者: 孙立宁, lnsun@hit.edu.cn 收稿/录用/修回: 2019-12-25/2020-04-02/2020-04-21

为里程计, 而带有回环检测和全局后端的完整框架被称为 SLAM。

近十年中, SLAM 及其相关技术的研究取得了长足的进展^[4-8], 大量的 SLAM 实现方法涌现, SLAM 的主要开源库也得到了发展^[9-10]。随着研究人员越来越关注将该技术应用于微小型无人机、小型排雷机器人和小型反恐侦察机器人等小型移动机器人平台, SLAM 的研究重点开始从激光雷达转向相机和惯性传感器 (IMU)。得益于芯片和微机电系统 (MEMS) 器件的快速发展^[11], 相机和 IMU 等传感器实现了高精度化、小型化和低成本化。采用相机和 IMU 的 SLAM 方法被称为视觉惯性 SLAM (visual-inertial SLAM, VI-SLAM), 只含有很小的漂移 (小于轨迹长度的 0.5%^[4,12])。

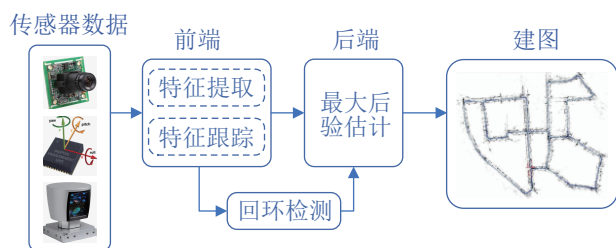


图 1 典型的 SLAM 框架

Fig.1 A typical SLAM framework

在多数情况下, 最准确的状态估计需要有效地融合使用内传感型和外传感型传感器的观测值^[13]。比如, 融合 GPS (全球定位系统) 和 IMU 是估计车辆位置与行驶速度的通用方法, 融合太阳/星敏感器和陀螺仪则是经常用于测定卫星姿态的方法。视觉传感器因其体积小、成本低和易于进行硬件设置而获得了广泛的关注, 大量基于视觉传感器的 SLAM 方法被提出^[14-18]。但纯视觉 SLAM 方法存在无法在图像纹理少的区域工作、快速运动时图像模糊等问题, 限制了其在实际机器人工作场景中的应用。相关生物学研究表明^[19], 人类和一些动物的导航方式取决于运动传感方式和视觉之间的组合。IMU 可测量角速度和加速度, 经常用于与其他传感器协同工作^[20-21]。IMU 被认为与相机具有明显的互补性, 并且在融合之后能够得到更加完善的 SLAM 系统^[22-24]。原因主要有以下几点:

(1) 将相机数据与 IMU 数据融合使用可以提高整体算法的鲁棒性, 改善纯视觉算法因为受某些运动或纹理较简单的场景影响而出现的不准确结果。

(2) IMU 虽然可以测得角速度和加速度, 但是测得的数据存在明显的零位漂移和温度漂移, 积分 2 次后得到的位姿数据非常不可靠。相较于 IMU,

相机采集的数据基本不会有漂移。假设相机是固定的, 那么视觉 SLAM 算法所估计的位姿也是固定的。因此, 相机数据可以有效地修正 IMU 读数中的漂移, 使得机器人在慢速运动中的位姿估计更加准确。

(3) 当机器人运动速度过快时, 2 帧相机照片之间的重叠区域太少导致算法无法进行特征匹配, 因此, 纯视觉 SLAM 系统在机器人快速运动时效果不佳。而添加了 IMU 后, 即使在快速运动时系统仍能保持较准确的位姿估计, 这是纯视觉算法无法做到的。

(4) 除了上述纯视觉 SLAM 算法均存在的问题, 基于单目相机的 SLAM 算法还存在尺度不确定性问题, 即无法仅凭图像确定真实世界尺度, 与 IMU 融合可以有效解决这个问题。

综上所述, 由于 IMU 在快速运动时效果较精确, 而相机又能够弥补 IMU 在慢速运动下的漂移问题, 因而可以看出, 2 个传感器具有很强的互补性。

使用 IMU 测量角速度和加速度需要引入运动学计算, 这增加了数据融合的难度。目前, 已有的 VI-SLAM 数据融合方法分为 2 大类: 紧耦合和松耦合^[25], 如图 2 所示。其中, 紧耦合是指把 IMU 的状态与相机的状态合并在一起进行位姿估计。松耦合是指相机和 IMU 分别进行自身的位姿估计, 然后对它们的估计结果进行融合。

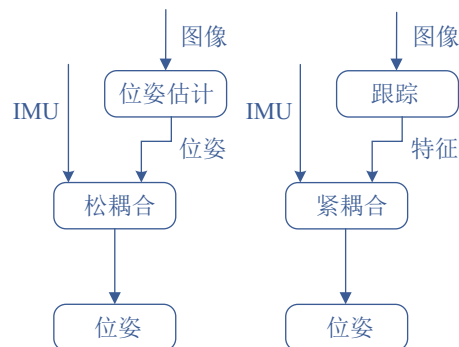


图 2 紧耦合和松耦合的示意图

Fig.2 A diagram of tightly coupled and loosely coupled methods

与纯视觉 SLAM 问题相似的是, VI-SLAM 也可以分为基于滤波和基于优化 2 类方法。在滤波方法中, 传统的 EKF (扩展卡尔曼滤波器)^[26]、UKF (无迹卡尔曼滤波器)^[27] 和改进的 MSCKF (多状态约束卡尔曼滤波器)^[28] 都取得了一定的成果, 而基于优化的方法则慢慢占据了主流。本文将介绍视觉惯性 SLAM 的最新研究进展, 从基于滤波和基

于优化 2 大类不同方法进行组织, 并说明它们的基本模型、代表性的算法和最新成果. 总结了该研究领域的关键问题, 将不同方法进行了对比, 并对未来的发展方向进行了展望.

2 基于滤波的方法 (Filtering-based approaches)

2.1 基于滤波的 VI-SLAM 方法原理

不同滤波方法的分类如图 3 所示. 卡尔曼滤波器^[29]作为一种解决状态估计问题的经典模型, 主要存在 2 个问题^[30]. 首先, 它要求时序和测量方程是线性的, 这个问题可以由 EKF^[31] 和 UKF^[32] 解决. 其次, 卡尔曼滤波器假设后验分布是单峰分布的, 而且需要通过均值和协方差来表达, 因此, 它针对物体的位姿只能有一个假设, 并不能保证关于状态的多个假设, 这个问题可以由粒子滤波器 (particle filter, PF)^[33] 解决.

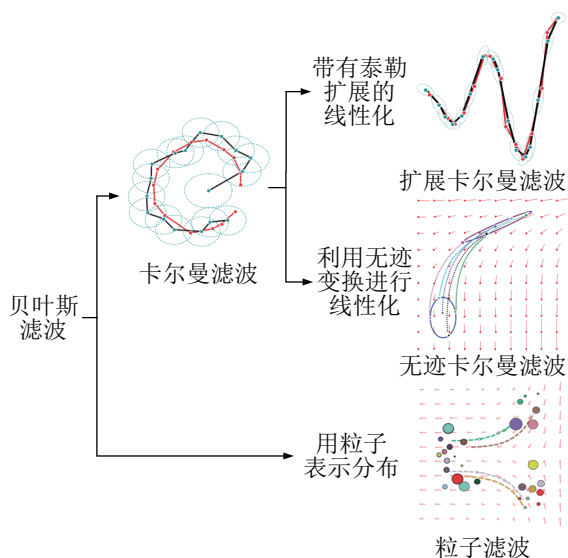


图 3 SLAM 中不同滤波方法的分类

Fig.3 Classification of different filtering methods in SLAM

一个完整的 EKF 框架包括预测步骤和更新步骤. 对于基于滤波的 VI-SLAM 方法, 惯性传感器能够提供 3 轴的加速度和角速度, 用于计算 3 维刚体运动的动态模型并在预测步骤中进行预测. 相机能够提供在特征和移动机器人之间的角度和距离测量结果, 并在更新步骤中更新预测结果.

由于 IMU 数据中存在偏差和噪声, 预测步骤的结果会随着时间的增加而变得越来越不准确. 因此, 通过引入视觉传感器的测量结果, 向系统中增加关键信息以限制误差的积累. 为了在 EKF 框架中使用这种方法, 从图像中提取的关键信息一般被转换为测量方程. 在松耦合方法中, 通过图像配准

直接获得位置和方向的变化, 然后在 EKF 框架中融合两者估计结果^[34]. 而紧耦合方法则使用从图像提取的关键信息. 关键信息可以通过特征检测器从图像特征中提取得到^[14], 也可以由具有深度信息的直接光强度^[35-36] 与半直接光强度得到^[16].

早期的 SLAM 工作主要是基于扩展卡尔曼滤波器, 由 Smith 等^[37] 基于早期的工作^[38-40] 实现. Jones 等^[26] 将 EKF 框架引入 VI-SLAM 系统. 该系统实现了将状态和参数作为在线程序的一部分进行估计, 并用 EKF 框架进行了有效实现. Kelly 等^[27] 将 UKF 框架引入 VI-SLAM 系统. 该方法能够在线、随时校准更新位姿, 比如正在进行的导航或建图任务.

2.2 代表性的基于滤波的 VI-SLAM 系统

MSCKF 是 Mourikis 在 2007 年提出的^[28], MSCKF 2.0 也在随后由 Mourikis 和 Li 提出^[41], 是目前许多 VI-SLAM 系统的基础. MSCKF 是一种基于扩展卡尔曼滤波器的 VI-SLAM 框架. 在传统 EKF 框架中, 特征点信息会加入到特征向量和协方差矩阵里, 即用多个特征点同时约束一个相机位姿, 进行 KF 更新, 如图 4 所示. 这种方法的缺点是特征点的信息会给定一个初始尺度和初始协方差, 如果不正确的话极容易导致后面的结果不收敛. MSCKF 工作的最主要贡献在于推导出一种测量模型, 该模型能够表达从多个相机位姿观察到静态特征时出现的几何约束. 系统中维护一个位姿的滑动窗口, 如果一个特征点在滑动窗口内的几个位姿都被观察到的话就会在这几个位姿间建立约束. 这种方法用一个特征点约束多个相机位姿, 从而进行 KF 的更新, 旧的特征点和滑动窗口之外的相机姿态则被丢弃. 近几年, 许多基于 MSCKF 的工作相继提出^[42-45], 框架整体的精度和鲁棒性得到了不断的提升.

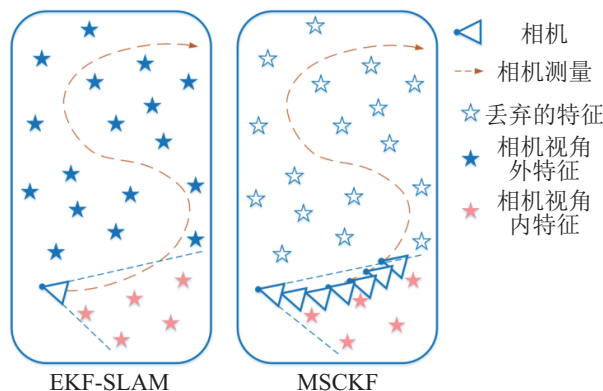


图 4 MSCKF、EKF-SLAM 的示意图

Fig.4 A diagram of MSCKF and EKF-SLAM

苏黎世联邦理工学院无人系统实验室 (Autonomous Systems Lab, ASL) 的 Bloesch 等提出了一种基于迭代扩展卡尔曼滤波器 (IEKF) 的直接法单目视觉惯性里程计, 简称 ROVIO (robust visual inertial odometry) [46-47], 如图 5 所示. 其通过直接使用图像块 (patch) 的光度误差 (photometric error), 实现精确的跟踪性能, 在低纹理场景中也表现出较高的鲁棒性. 对于视觉方面的信息, 系统通过将路标点在图像中对应的点周围的图像块作为路标点的描述子, 从而得到光度误差. 对光度误差进行变换后, 得到 IEKF 中的更新量, 进而对滤波状态进行更新. 其滤波方程是以移动机器人中心 (robocentric) 构造的, 即在一个以 IMU 为原点的坐标系上构造, 从而保证能观状态不受不断增长的全局协方差的影响, 减小了因非线性而造成的误差. 此外, 模型中还使用了一种新颖的路标点参数化方式. 系统中将当前相机坐标系下路标点的空间位置信息拆成了方位向量和逆深度 2 项来表示, 并在 IMU 预测阶段对路标点进行预测, 在视觉更新时对其修正, 不像其他框架一般仅在视觉阶段去计算. 这样的构造方式可以避免不能观状态量, 并且使得初始化没有延迟. 由于使用了基于 QR 分解的测量空间缩减方法, 并执行每个路标点的更新迭代, 因此系统具有较高的效率和准确性.

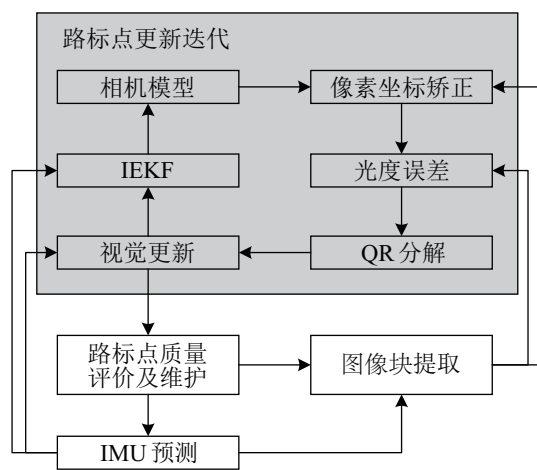


图 5 ROVIO 框架的流程图 [47]

Fig.5 Flow of the ROVIO framework

ARL 实验室随后又推出了 Maplab 框架 [48], 与 ROVIO 不同的是, Maplab 带有一个完整的 VI-SLAM 系统, 具有回环检测和重定位功能. 系统由两部分组成, 一部分是 ROVIOLI (ROVIO with localization integration), 一个在线的视觉惯性全局定位系统. 其接收图像与惯性传感器数据作为输入, 输出全局的位姿估计, 并建立地图. 另一部分是离

线的 Maplab 控制台, 可以让使用者以离线批处理方式在地图上应用各种算法.

在松耦合方法中, Faessler 等 [49] 使用了一种基于 SVO (semi-direct visual odometry) 和 MSF (multi-sensor fusion) 的方案. SVO [15] 是一种计算量较小的稀疏直接算法. 其通过跟踪 FAST (from accelerated segment test) [50] 特征并最小化周围图像块的光度误差来匹配帧间图像, 使用非线性最小二乘法来最小化特征中的重投影误差, 得到仅由相机估计的位姿. MSF [51] 是一种通用的 EKF 框架, 用于在位姿估计中融合来自不同传感器的数据. 系统将纯视觉 SVO 得到的位姿作为通用位姿传感器的输出提供给 MSF, 然后与 IMU 数据融合. 由于是一种松耦合框架, 位姿的尺度需要近似正确, 因而时常要手动初始化.

基于滤波的方法形式简洁、应用广泛, 尤其是 EKF 方法, 广泛应用于各类状态估计与非线性控制系统. 基于滤波的方法已经比较成熟, 在硬件系统中取得了一定的成果. 其中 Google Project Tango 借助滤波方法实现了精确的定位效果, 并具有较理想的运动追踪和深度感知功能. 但由于基于滤波的方法具有一定的非线性误差, 目前研究人员已经把关注的重点转向基于优化的方法.

3 基于优化的方法 (Optimization-based approaches)

基于优化的方法主要依靠图像处理技术进行特征提取和图像匹配, 而 IMU 数据则被视为先验项或者正则化项. 自 PTAM (parallel tracking and mapping) [14] 之后, 批量非线性优化方法一般都分为 2 个线程: 跟踪和建图. 在跟踪线程, 通过各种特征检测器从图像中提取 3 维空间中的点、线或其他路标特征. 然后, 针对检测到的所有特征, 在 2 个图像之间定义重投影误差. 之后用该误差建立优化的代价函数, 以便找到特征或地标的坐标, 即光束平差法 (bundle adjustment, BA) [52]. 在建图线程, 地图中特征和路标的坐标被用于定义 2 幅图像之间的重投影误差, 然后再次应用优化算法来发现移动机器人位姿的变化. 将优化问题分成 2 个并行线程的目的是使跟踪部分实时响应图像数据, 以便快速获得跟踪结果. 而地图的优化没有必要实时计算, 可以放在后端缓慢运行. 基于优化的方法通常使用 g^2o [53]、Ceres [54]、GTSAM [55] 等非线性优化库实现优化.

基于优化的方法也称为平滑方法，可以分为固定滞后平滑（fixed-lag smoothing）方法和完全平滑（full smoothing）方法。完全平滑方法估计移动机器人运动过程中的所有位姿，固定滞后平滑方法考虑几个最新位姿集合的窗口，而基于滤波的方法仅估计最新位姿。

3.1 固定滞后平滑方法

固定滞后平滑方法对落入给定时间范围内的状态进行估计，同时边缘化旧的状态^[56-60]以维持用于实时操作的优化窗口。通过引入关键帧的机制，丢弃除选定关键帧^[56,61-63]以外的帧，不必精细地处理每一幅图像。对于非线性问题，平滑方法通常比滤波方法的精度更高，因为它们可以对过去的测量结果重新进行线性化处理^[64]。此外，平滑方法对异常值更具弹性，可以在优化之后将异常值丢弃，也可以通过使用鲁棒的代价函数来处理这些异常值。不利的一面是，估计窗口之外的状态被边缘化导致了稠密的高斯先验，这在一定程度上阻碍了有效的估计。针对这个问题，一些文献^[56-57]中提出了为了稀疏性而放弃某些测量的方法。此外，由于采用了边缘化的方法，固定滞后平滑方法具有与部分滤波方法相似的问题，如一致性，线性化误差的累积等^[60,65-66]。

苏黎世联邦理工学院 ASL 实验室 Leutenegger 等提出了 OKVIS（open keyframe-based visual-

inertial SLAM）^[56-57]，利用基于关键帧的滑动窗口进行批量非线性优化，先于滑动窗口的关键帧被边缘化，不用来进行估计。系统前端使用多尺度 Harris 特征检测器^[67]来提取特征，然后在其基础上计算 BRISK（binary robust invariant scalable keypoint）描述子^[68]，以便在帧与帧之间进行数据关联。如图 6 所示，左边表示纯视觉 SLAM 问题的示意图，右边示意图表示引入惯性测量后，不仅会在连续位姿之间创建时间约束，还会在 IMU 的速度和偏差估计之间创建时间约束，从而增加了移动机器人的状态向量。后端通过结合视觉特征和加权惯性误差项的加权重投影误差制定代价函数来进行非线性优化，并使用谷歌的 Ceres Solver 非线性优化库^[55]来实现后端优化。

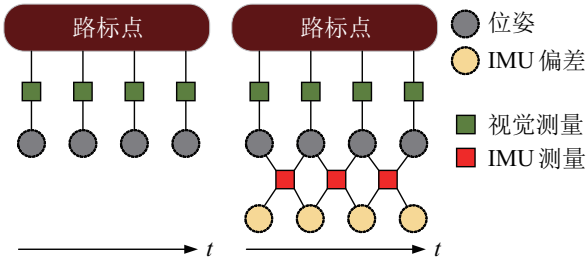


图 6 视觉 SLAM 问题与视觉惯性 SLAM 问题的状态变量和测量值的对比图^[56]

Fig.6 Graphs of the state variables and measurements involved in the visual SLAM versus visual-inertial SLAM problems

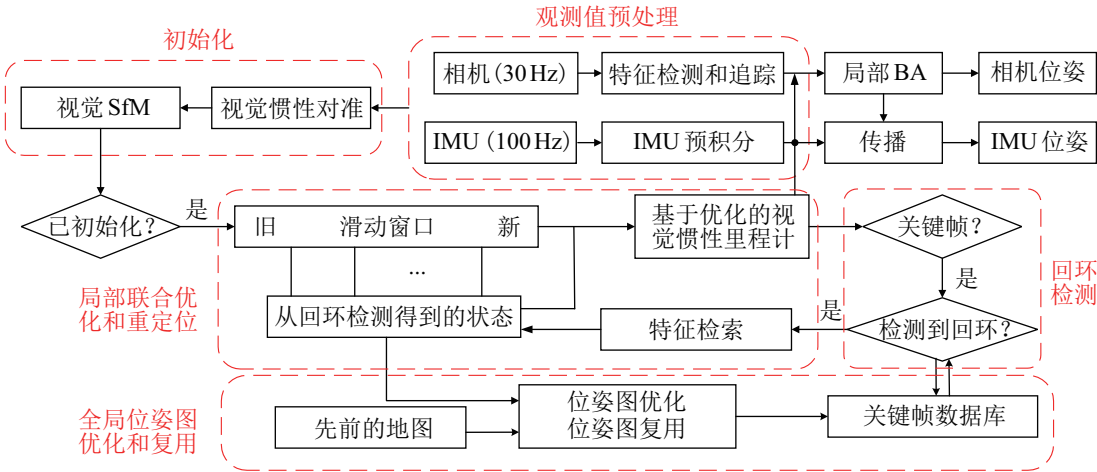


图 7 VINS-Mono 的完整框架^[69]

Fig.7 The full pipeline of the monocular VINS

香港科技大学飞行机器人实验室提出的 VINS-Mono（monocular visual-inertial system）方法^[69]类似于 OKVIS，是一种基于非线性优化的 VI-SLAM 方法。VINS-Mono 方法为这类框架引入了几个全新的功能，其完整系统包括观测值预处理、初始

化、局部视觉惯性联合优化、全局图优化和回环检测 5 个部分，如图 7 所示。前端提取 Harris 特征点^[67]，并采用 LK（Lucas-Kanade）光流（optical flow）法^[70]跟踪相邻帧。光流法是一种描述像素随时间在图像之间运动的方法。LK 光流法计算部分

像素, 在 SLAM 中用于跟踪特征点的位置. VINS-Mono 方法只计算特征点, 不计算描述子, 同时使用光流法跟踪特征点的运动. 这样就减少了计算和匹配描述子的时间和资源, 只需要计算光流. 在初始化部分, 作者使用了一种松耦合的传感器融合初始化程序, 使系统可以在优化之前从任意初始状态引入估计器. 系统采用与 OKVIS^[56] 相似的基于滑动窗口的紧耦合位姿估计方法, 并且加入了基于 DBoW2 (bag of binary words 2)^[2] 的回环检测线程, 使系统具有重定位功能.

香港科技大学飞行机器人实验室随后又推出了 VINS-Mobile^[71-73] 和 VINS-Fusion^[74]. VINS-Mobile 在 iOS 设备上运行, 并为增强现实 (AR) 应用程序提供本地化服务. VINS-Mobile 对 VINS-Mono 框架作了许多轻量化处理, 并采用了一种基于小滑动窗口的联合优化方案. VINS-Fusion 在完善 VINS 框架的同时, 提供了 4 个不同硬件的版本, 包括单目 IMU、双目 IMU、纯双目和双目 IMU 与 GPS 松耦合.

百度 AR 技术小组提出的 ICE-BA (incremental, consistent and efficient bundle adjustment)^[75] 沿用了 OKVIS 以及 VINS-Mono 这一类框架. 前端提取 Harris 特征点^[67] 并采用 LK 光流法^[70] 跟踪相邻帧, 与 VINS-Mono 相似. 后端则是论文中提出的增量式 BA, 主要分为 3 个部分: 局部 BA、全局 BA 以及相对边缘化 (relative-marginalization), 前两者采用增量式方法提升了后端速度, 后者保证了局部 BA 和全局 BA 的一致性. 其中, 全局 BA 在建立增量方程时, 对系统已经计算过且不变的状态向量不进行计算. 因为即使重新计算, 精度也不会有太大的改善. 对于局部 BA 来说, 由于几乎每个点都被滑动窗口里面所有帧看到, 因此即使只重新线性化一部分点, 也会带来较大的改动. 因此在局部 BA 这个部分, 系统中采用了一个更高效的针对局部 BA 的后端设计. 这个方法把长期跟踪的特征点分成很多短期跟踪段, 这样每次重新计算特征点的时候, 重新线性化的区域减小, 从而提高了效率. 对于固定滞后系统, 往往都有边缘化这一步, 把有些本应该出了滑动窗口的信息以另一种形式存储下来. 但是当有全局 BA 存在时, 边缘化的效果未必会有很大的提升, 反而可能会下降. ICE-BA 中改进了边缘化方法, 保持了边缘化和全局 BA 的一致性.

由于 VI-SLAM 系统实现了优异的效果, 一些著名的视觉 SLAM 系统^[15-18] 相继推出了自己系统

的 VI 版本^[12,76-78], VI-ORB SLAM^[76] 就是其中最具有代表性的系统之一. ORB (oriented FAST and rotated BRIEF)^[79] 是一种新颖的实时图像特征, 它改进了 FAST^[50] 特征点不具有方向性的问题, 并采用改进的二进制描述子 BRIEF (binary robust independent elementary feature)^[80], 加快了图像特征提取的速度. ORB-SLAM^[17,81-84] 首次使用 3 个线程: 实时跟踪特征点的线程、局部建图的优化线程 (co-visibility graph) 和全局回环检测与优化线程 (essential graph), 来完成 SLAM. 跟踪线程对每幅图像提取 ORB 特征点, 并与最近的关键帧比较, 计算特征点位置并估计位姿. 局部 BA 线程求解更精细的相机位姿和特征点空间位置. 全局回环检测线程采用 DBoW2^[2] 对全局的地图与关键帧进行回环检测, 消除累积误差. ORB-SLAM 的三线程结构取得了优异的跟踪和建图效果, 保证了移动机器人轨迹与地图的全局一致性, 被后续许多研究者认可和采用.

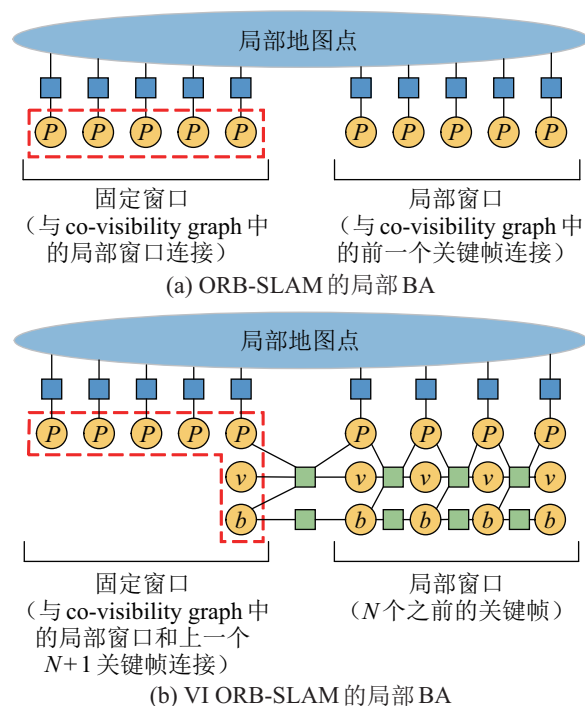


图 8 ORB-SLAM 与 VI ORB-SLAM 的局部 BA 比较^[76]

Fig.8 Comparison of local bundle adjustment between original ORB-SLAM and visual-inertial ORB-SLAM

VI-ORB SLAM^[76] 改进自 ORB-SLAM2^[84], 作者引入 IMU 尝试解决在快速运动时丢失特征点的问题. VI ORB-SLAM 分别对 3 个线程作了修改, 用以融合 IMU 信息. 在跟踪线程, 基于重投影误差和 IMU 预积分, 建立帧与帧之间的约束关系来构造代价函数, 从而得到当前帧位姿的最优估计.

在局部建图线程,有了新的关键帧之后,将会对前 N 个关键帧进行优化,当前的关键帧(第 $N+1$ 帧)将固定不变,提供 IMU 预积分约束.如图 8 所示, P 、 v 、 b 分别为优化的位姿、IMU 速度与偏差.在全局回环检测线程,由于 IMU 提供了尺度信息,因此全局优化将从 7 个自由度下降到 6 个自由度.全局位姿优化将忽略 IMU 信息,因此不再优化速度和偏差,当完成全局位姿优化后,再根据矫正后的位姿对速度进行矫正.国内,天津大学张玉龙^[85-86]、东南大学代维^[87-88]等也在基于 ORB 特征点的 VI-SLAM 算法方面做了许多工作.

3.2 完全平滑方法

完全平滑方法通过解决一个大型的非线性优化问题^[89-93]来估计移动机器人的全部轨迹,包括相机轨迹和 3 维路标点.在目前所有的方法里,完全平滑方法可确保最高的精度.但是,由于轨迹和地图随着时间的增长,实时操作很快变得不可行. iSAM^[94]、iSAM2^[95]等增量平滑技术的发展是完全平滑方法的一项突破,该技术利用因子图(factor graph)^[96-97]来保持稀疏性,并仅识别和更新新度量影响的变量的一小部分.

iSAM2^[95]使用贝叶斯树^[98]数据结构来获得有效的变量排序,从而最大程度地减少了计算时间.由于新的测量结果只对估计值的局部产生影响,通过分析受影响的区域,可以减少一些没必要的计算,加速优化流程.

Forster 等^[12]实现了基于完全平滑方法的视觉惯性系统.其前端采用 SVO^[15],跟踪 FAST^[50]特征点并最小化周围图像块的光度误差来匹配帧间图像.工作中提出了一种新颖的 IMU 预积分理论,将 2 个关键帧之间的许多惯性测量组合成单个相对运动约束,并构建为因子图模型.后端采用 iSAM2^[95]执行因子图优化,并用 GTSAM4.0^[54]实现.

基于优化的方法是目前 VI-SLAM 领域以及 SLAM 相关领域的热点之一,非线性优化方法实现了相较于滤波方法更精确的状态估计效果.但由于计算资源有限,目前的方法往往以牺牲一部分精度和路标点为代价换取系统的实时性.新型的优化器以及并行计算方法的出现有望进一步减少对计算资源的占用,但大范围在线系统的实时性仍然是一个有待解决的问题.

4 视觉惯性 SLAM 的关键问题 (Key issues of visual-inertial SLAM)

视觉惯性 SLAM 方法要解决的问题是根据惯性

和视觉传感器的测量结果,在线恢复全局框架中的移动机器人轨迹.然而,当实际实施时,这个问题变得十分复杂.除了基于滤波或优化方法的复杂性外,一些参数在代表移动机器人轨迹的动态状态估计中也起着至关重要的作用.这些参数包括:(1)相机内部参数:焦距、主点和镜头失真系数.(2)IMU 参数:加速度和陀螺仪偏差.(3)时间参数:IMU 和相机测量之间的时间延迟.(4)空间参数:IMU 和相机之间的变换.

除了 IMU 偏差之外,其余所有参数都是时不变变量,而移动机器人的轨迹由状态这个时变变量表示.仅考虑 IMU 和相机的测量值,可以通过能观性和可识别性分析这些状态和参数是否可以恢复.由于这些参数的存在,一些 SLAM 系统通常通过 IMU 预积分和初始化来增加系统的精度和减少对运算资源的占用.

4.1 状态能观性

系统的能观性在进行状态估计时起着关键性的作用^[99],研究系统的能观性也是 VI-SLAM 问题的热点之一.对于一个系统,能观性决定了这个系统在进行状态估计时,哪些自由度是可以被估计出来的,并且能观性是不受估计方法影响的.相关研究主要集中在传感器或测量值组合的能观性^[100-101]和仅使用 IMU 测量的能观性^[102].

移动机器人整体位置的不能观、围绕重力矢量的旋转(即偏航)的不能观和相对于初始参考位姿的不确定性的不断增加是视觉惯性估计问题的固有因素^[103-104].原因在于相机是纯方位传感器,而 IMU 是位置的双重积分器,因此无法在全局框架中提供绝对位置和偏航信息^[105-106].VI-SLAM 问题中一些参数的不能观给通常不能进行重复线性化的基于滤波的方法带来了挑战.EKF 框架对非线性系统进行了线性化,但线性系统的能观性不同于相应的非线性系统.Huang 等^[107]表明,滤波方法所导致的线性化误差可能会错误地使估计状态的某些部分在数值上能观,即在线性化系统中可观察到偏航.这种可观察到的自由度可能会导致结果不一致:估计的协方差比实际测量结果更准确.

改善一致性的一种方法是在由于模型线性化而产生可观察错误的方向施加约束^[108].还有一些研究人员尝试使用连续对称性对能观性进行分析,从而进行状态能观性和参数可识别性的解析推导^[109].其他一些研究^[64,103-105,110]中得出的公式可以选择系统的线性化点,这样使得线性化和非线性系统的能观性相同.目前多数基于非线性优化的系统中^[56]

均采用 FEJ (first estimate Jacobian), 每当将变量线性化时, 为涉及该变量的后续线性化确定线性化点, 从而使线性化的状态点保持一致。

4.2 参数可识别性

在大多数与视觉相关的定位技术中相机的内部参数都是预先确定的, 也可以在线进行校准^[111]。而 IMU 的偏差随时间变化, 在大多数情况下被建模为时变状态。这些参数都是能观的。

相机和 IMU 之间的空间参数是 6 个自由度的变换, 这个结果必须准确才能确保估计结果的精度。对于系统能观性的分析结果表明, 6 自由度相机的 IMU 变换、IMU 偏差、重力矢量和度量场景结构都是能观的。完全的能观性需要足够的激励, 机器人需要至少沿着 IMU 坐标系的 2 个方向进行旋转和加速^[26-27,112]。

IMU 和相机测量之间的时间参数是时间延迟。通过构造涉及时间延迟和其他变量的约束方程来确定时间延迟的局部可识别性, 结果表明时间延迟在轨迹中是局部可识别的。Li 和 Mourikis^[113]描述了导致可识别性丧失的关键轨迹, Kelly 和 Sukhatme^[114]采用一种 ICP (迭代最近点) 算法的变形来估算 IMU 和相机测量之间的时间延迟。

4.3 IMU 预积分

惯性测量的高频率 (通常为 100 Hz 至 1 kHz) 对 VI-SLAM 系统构成了挑战。在基于滤波的方法中, 通常使用 IMU 进行状态传播^[23,40]以解决高频率惯性测量问题。在基于优化的方法中, 最原始方法是在每个 IMU 测量时都添加一个新状态, 系统很快变得非常缓慢^[115]。因此, 通常在 2 帧之间对惯性测量进行积分, 以便形成相对运动约束^[56,116-119]。对于 2 帧之间的 IMU 积分, 初始化的状态往往由第 1 帧的状态估计给出。但是, 在优化的每次迭代中, 状态估计都会发生变化, 这就需要在所有帧之间重复进行 IMU 积分^[56]。

Lupton 和 Sukkarieh 表明^[120], 可以通过重新参数化相对运动约束来避免这种重复积分。这种重新参数化称为 IMU 预积分。Forster 等^[12,121]通过添加后验协方差校正完善了 IMU 预积分理论, 并被 VI ORB-SLAM 等框架沿用^[76-78]。VINS-Mono 中使用了 Shen 等^[122]基于四元数的连续时间推导方法并考虑了 IMU 的偏差, 取得了与 Forster 等相同的数值结果^[69]。

4.4 初始化

对 VI-SLAM 系统进行初始化的目的是将纯视觉 SLAM 无法估计的参数先估计出来, 包括尺度

信息、包含方向和大小的重力向量、在每个位置的速度和 IMU 的初始偏差。如果这些参数的初始估计偏差很大, 那整个系统的精度会很低。基于滤波的方法可能结果会偏差很大, 而基于非线性优化的方法可能会陷入局部最优解。一些框架的初始化还会把相机和 IMU 之间的相对位姿关系估计出来^[69,71,123], 避免了预先对相机和 IMU 进行外参标定。

Yang 等^[71]采用一种类似紧耦合的初始化方式, 求解出尺度、重力、速度和相机与 IMU 之间的内外参。但当其系统处于特征点的深度分布范围比较大的场景时, 估计的结果不够准确。并且由于这种方法忽略了 IMU 偏差, 可能会在系统中引入误差。VI ORB-SLAM^[76]中采用了一种松耦合的初始化方式。首先对纯视觉 SLAM 进行初始化, 得到了 3 维空间中一系列的相机位姿和地图点。在这个过程中, IMU 持续进行预积分, 得到每 2 个关键帧之间的预积分量。之后再将视觉初始化得到的相机位姿和预积分量进行对齐。在 VI ORB-SLAM 的初始化中估计了尺度、重力、速度、陀螺仪和加速度计的偏差, 而没有估计外参。VINS-Mono 的初始化方法与 VI ORB-SLAM 相似, 也是一种松耦合的方式^[69]。其首先对视觉部分进行初始化, 然后与 IMU 预积分的值对齐, 最后求解出重力、尺度、陀螺仪偏差和每一帧对应的速度。VINS-Mono 并没有像 VI ORB-SLAM 一样消去速度项, 而是选择将其作为未知变量估计出来。VINS-Mono 在初始化的时候忽略了加速度计的偏差, 并通过仿真证明如果没有足够的旋转激励, 加速度计的偏差是不能观的^[72]。忽略加速度计的偏差对估计其他初始值的影响不大, 并且系统能通过之后的紧耦合位姿估计将加速度计偏差估计出来。Huang 和 Liu^[123]提出了一种迭代进行外参估计和陀螺仪偏差估计的方式, 以解决旋转外参估计和陀螺仪偏差估计相互依赖的问题。

5 滤波方法与优化方法的联系与对比 (Links and comparison between filtering and optimization based approaches)

利用贝叶斯推断可以建立基于滤波和基于优化这 2 类方法之间的联系。基于滤波的方法, 位姿的先验分布由内感受型传感器的测量构建, 似然分布由外感受型传感器的测量建立, 故可以视为最大后验估计 (maximum a posteriori, MAP) 问题。基于优化的方法, 通过迭代找到测量总概率最高的状

态，故可视为最大似然估计（maximum likelihood, ML）问题。基于优化的方法，可以通过向传感器的测量值中添加正则化项或先验项，使其由 ML 问题转化为 MAP 问题。

如果把 EKF 框架看作是非线性高斯-牛顿法或高斯法的近似，那么它的表现无疑是达不到要求的。主要原因是，EKF 没有迭代至收敛的过程，其雅可比矩阵只计算 1 次，可能会远离所期望的最优估计。但由于 EKF 没有一次性计算所有的雅可比

矩阵，EKF 的结果比单次的高斯-牛顿法迭代更精确，它的缺陷只在于没有迭代这个步骤。从优化的角度来看，这是一个显而易见的问题，因为优化是需要最后迭代至收敛的。EKF 使用了马尔可夫假设来实现其递归形式，由于使用了马尔可夫假设，一旦滤波器建立在该假设上就无法摆脱它。包括 IEKF 在内，虽然 IEKF 一次迭代了一个时间步长，但它仍然依赖于马尔可夫假设，而且仅在一个时刻上进行了迭代，并非在整个轨迹上。

表 1 代表性 VI-SLAM 框架对比
Tab.1 Comparison of the mainstream VI-SLAM frameworks

VI-SLAM 框架	耦合方案	初始化	前端	后端	视觉误差	回环	参考文献
MSCKF ^a	紧耦合	静止	FAST+光流	EKF	重投影	无	[44]
ROVIO ^b	紧耦合	静止	FAST+光度	IEKF	光度	无	[47]
SVO ^c +MSF ^d	松耦合	静止	FAST+光度	EKF	光度	无	[49]
OKVIS ^e	紧耦合	静止	Harris+BRISK	优化	重投影	无	[56]
VINS ^f	紧耦合	动态	Harris+光流	优化	重投影	有	[69]
ICE-BA ^g	紧耦合	静止	Harris+光流	优化	重投影	有	[75]
VI ORB-SLAM ^h	紧耦合	动态	ORB	优化	重投影	有	[76]
SVO+GTSAM ⁱ	紧耦合	静止	FAST+光度	优化	光度	无	[12]

本文比较了几种代表性的 VI-SLAM 框架，如表 1 所示。可以看出，目前主流的 VI-SLAM 实现方法以紧耦合的优化方法为主。相较于松耦合的方法，把 IMU 状态与相机状态合并在一起进行状态估计的紧耦合方法具有更高的精度^[124]。而由于基于滤波的方法具有马尔可夫性，无法考虑到某时刻状态与之前所有时刻状态的关系，目前普遍认为在计算资源足够的情况下，基于优化的方法会得到更精确的效果^[61,124]。但是在计算资源受限，或者移动机器人位姿轨迹比较简单的场合里，基于滤波的方法仍然是一种有效的方式。

6 验证数据集（Validation dataset）

数据集的存在使研究者在没有移动机器人硬件平台的情况下仍然可以测验算法的效果，同样在 VI-SLAM 领域的发展中产生了多种可用的数据集。本文从数据集的发布时间、环境、移动平台、相机的类型、IMU 的频率以及真值的来源等方面介绍几

个 VI-SLAM 系统的验证数据集，如表 2 所示。

这些数据集具有不同的特点，其中 EuRoC^[125]和 KITTI 数据集^[126]是研究者们通常使用的 2 个数据集。EuRoC 数据集^[125]是由小型无人机采集的室内场景数据集，无人机运动比较剧烈，适宜测验系统的鲁棒性。而 KITTI 数据集^[126]包含市区、乡村和高速公路等场景下采集的真实图像数据，用以测验系统在大范围场景中的全局定位精度。

7 未来展望（Future prospects）

7.1 与深度学习结合

构建语义地图、在帧间匹配和回环检测中采用深度学习的方法是目目前 SLAM 的研究热点之一^[132-133]。深度学习方法的引入使移动机器人可以理解周围环境的语义信息，完成一些更复杂的任务。

端到端的 SLAM 方法已经通过视觉里程计^[134-135]和惯性里程计^[136]实现，一些工作也尝

a. https://github.com/daniilidis-group/msckf_mono
b. <https://github.com/ethz-asl/rovio>
c. <http://rpg.ifi.uzh.ch/svo2.html>
d. <https://github.com/ethz-asl/ethzasl-msf>
e. https://github.com/ethz-asl/okvis_ros
f. <https://github.com/HKUST-Aerial-Robotics/VINS-Mono>
g. <https://github.com/baidu/ICE-BA>
h. <https://github.com/jingpang/LearnVIOORB>
i. <https://bitbucket.org/gtborg/gtsam/>

试使用端到端深度神经网络解决视觉惯性里程计的问题。VINet 网络^[137]使用 FlowNet 网络^[138]来建模相邻帧间的光流运动特征，用一个小型 LSTM（长短期记忆）网络^[139]来处理 IMU 的时序数据，从而获得 IMU 数据的运动特征，对视觉运动特征和 IMU 运动特征作特征融合之后，送入另一个 LSTM

网络进行位姿估计。VIO Learner 网络^[140]学习整合 IMU 测量结果并生成估计轨迹，根据相对于像素坐标的空间网格的缩放图像误差进行在线校正。Chen 等^[141]提出了一种针对单目相机的端到端的多传感器融合策略，Lee 等^[142]针对传统方法需要进行标定的问题，提出了一种不需要提前标定的网络。

表 2 包含视觉和 IMU 数据的数据集对比
Tab.2 Comparison of the datasets with vision and IMU data

数据集	发布时间	环境	移动平台	相机	IMU 频率	真值来源	参考文献
EuRoC MAV	2016	室内	小型无人机	1×双目	1×200 Hz	激光跟踪仪/ 运动捕捉系统	[125]
KITTI Odometry	2013	室外	车辆	2×双目	1×10 Hz	GPS	[126]
TUM VI	2018	室内/室外	手持	1×双目	1×200 Hz	运动捕捉系统/ 基准标记	[127]
Zurich Urban MAV	2017	室外	小型无人机	1×单目	1×10 Hz	合成数据	[128]
Malaga Urban	2014	室外	车辆	1×双目	1×100 Hz	GPS	[129]
UMich NCLT	2015	室内/室外	Segway 机器人	6×单目（全向）	1×100 Hz	合成数据	[130]
PennCOSYVIO	2017	室内/室外	手持	4×单目	1×200 Hz	基准标记	[131]
				1×双目 1×鱼眼	1×128 Hz 1×100 Hz		

传统的 VI-SLAM 框架中，相机和 IMU 数据帧的同步与外参的标定对于算法精度的影响无疑是巨大的。当传感器的标定参数存在误差的时候，基于数据驱动的深度学习方法较常规方法而言具有更高的鲁棒性。但深度学习方法对于计算资源的需求巨大，在硬件条件较差的嵌入式场景难以使用，需要研究人员进一步地改进网络。

7.2 轻量化

SLAM 本身是为了给上层应用提供自身位姿估计，在实际应用中，研究人员并不希望算法占用太多运算资源。使用者希望 SLAM 框架能够轻量化，不影响移动机器人所要完成的其他工作。相比于激光 SLAM 等成本高昂的方法，由于 VI-SLAM 系统只使用相机和惯性传感器，在无人机或手持移动设备上良好运行 SLAM 程序成为可能^[143]。通过继续改进算法，今后有望在低端硬件上实现更理想的位姿估计效果。

7.3 稠密化

在实际场景应用中，使用者往往对移动机器人有导航、避障和场景重建等其他需求。稀疏路标点地图只能实现定位功能，要想实现其余的功能，需要建立稠密地图^[144]。相比于激光 SLAM 的建图效果，目前的 VI-SLAM 方法的建图效果仍然比较稀疏。在建图的方法中，特征点法只能构建稀疏特征

点地图，而直接法有恢复稠密或半稠密场景的能力。目前的 VI-SLAM 框架以稀疏特征点法和稀疏光流法为主，采用直接法构建稠密地图将会是未来 VI-SLAM 的一个重要发展方向。

7.4 与更多传感器融合

视觉与 IMU 的融合已经在不同的领域取得了较理想的效果，但是当场景特别大时，视觉和 IMU 传感器的效果不太理想，这时就需要用到其他传感器。在自动驾驶领域，将激光雷达、GPS/GNSS（全球卫星导航系统）、相机以及 IMU 等多传感器进行融合已经成为一种趋势^[145]，引入激光雷达等传感器会为系统提供更准确的深度信息。但是更多的传感器会带来更复杂的标定与融合方面的问题，有待研究人员加以解决。

7.5 嵌入式系统

移动机器人通常使用嵌入式设备作为平台的计算硬件，而随着计算能力的大力发展，嵌入式设备的使用频率逐渐增多。新型 GPU（图形处理器）的出现使得嵌入式系统的性能更加接近 PC 机，但是计算能力仍然有限。一些嵌入式硬件，例如 Jetson Nano、TX1/2^[146] 的出现使得研究人员可以在嵌入式系统中使用 GPU，便于移动机器人在恶劣的条件下应用 VI-SLAM 算法。随着一些代表性产品 Google Project Tango、Microsoft Hololens、Magic

Leap 等的应用, VI-SLAM 技术会逐步融入到平常生活之中。

8 结论 (Conclusion)

视觉惯性 SLAM 方法基于视觉和惯性传感器, 通过数据耦合的方式提升了位姿估计轨迹的精度和鲁棒性。由于 2 个传感器具有体积小、重量轻、便携性好等特点, 视觉惯性 SLAM 系统正在越来越多地应用于移动机器人、自动驾驶汽车和移动设备等各类现实环境中。视觉惯性 SLAM 方法会在一些仍需要提供可靠指标定位的特殊场合, 特别是 GPS 不可用或不能提供足够准确性的情况下使用。然而, 现有的系统仍然有鲁棒性不够、对光照敏感、无法解决快速运动等问题, 需要进一步探索与研究。

参考文献 (References)

- [1] Lowry S, Sünderhauf N, Newman P, et al. Visual place recognition: A survey[J]. IEEE Transactions on Robotics, 2015, 32(1): 1-19.
- [2] Gálvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
- [3] 刘强, 段富海, 桑勇, 等. 复杂环境下视觉 SLAM 闭环检测方法综述 [J]. 机器人, 2019, 41(1): 112-123,136.
Liu Q, Duan F H, Sang Y, et al. A survey of loop-closure detection method of visual SLAM in complex environments[J]. Robot, 2019, 41(1): 112-123,136.
- [4] 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.
- [5] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: Part I [J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
- [6] Bailey T, Durrant-Whyte H. Simultaneous localization and mapping (SLAM): Part II [J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.
- [7] Thrun S, Burgard W, Fox D. Probabilistic robotics[M]. Cambridge, USA: MIT press, 2005.
- [8] Strasdat H, Davison A J, Montiel J M M, et al. Double window optimisation for constant time visual SLAM[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2352-2359.
- [9] Dissanayake G, Huang S D, Wang Z, et al. A review of recent developments in simultaneous localization and mapping[C]//6th International Conference on Industrial and Information Systems. Piscataway, USA: IEEE, 2011: 477-482.
- [10] Kelly A. Mobile robotics: Mathematics, models, and methods [M]. Cambridge, UK: Cambridge University Press, 2013.
- [11] 孙立宁, 周兆英, 龚振邦. MEMS 国内外发展状况及我国 MEMS 发展战略的思考 [J]. 机器人技术与应用, 2002(2): 2-4.
Sun L N, Zhou Z Y, Gong Z B. Thoughts on the development status of MEMS and MEMS development strategy[J]. Robot Technique and Application, 2002(2): 2-4.
- [12] 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.
- [13] Barfoot T D. State estimation for robotics[M]. Cambridge, UK: Cambridge University Press, 2017.
- [14] 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. DOI: 10.1109/ISMAR.2007.4538852.
- [15] Forster C, Zhang Z C, Gassner M, et al. SVO: Semidirect visual odometry for monocular and multicamera systems[J]. IEEE Transactions on Robotics, 2017, 33(2): 249-265.
- [16] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]//European Conference on Computer Vision. Cham, Switzerland: Springer, 2014: 834-849.
- [17] Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
- [18] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.
- [19] Wolf H. Odometry and insect navigation[J]. Journal of Experimental Biology, 2011, 214(10): 1629-1641.
- [20] 陈诺. 基于 UWB 与惯导融合的室内导航系统研究 [D]. 哈尔滨: 哈尔滨工业大学, 2018.
Chen N. Research on indoor navigation system based on UWB and INS fusion[D]. Harbin: Harbin Institute of Technology, 2018.
- [21] Wang P F, Chen N, Zha F S, et al. Research on adaptive Monte Carlo location algorithm aided by ultra-wideband array [C]//13th World Congress on Intelligent Control and Automation. Piscataway, USA: IEEE, 2018: 566-571.
- [22] Gui J J, Gu D B, Wang S, et al. A review of visual inertial odometry from filtering and optimisation perspectives[J]. Advanced Robotics, 2015, 29(20): 1289-1301.
- [23] 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 and Automation. Piscataway, USA: IEEE, 2012: 957-964.
- [24] Li M, Mourikis A I. Improving the accuracy of EKF-based visual-inertial odometry[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2012: 828-835.
- [25] Martinelli A. Closed-form solution of visual-inertial structure from motion[J]. International Journal of Computer Vision, 2014, 106(2): 138-152.
- [26] 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.
- [27] Kelly J, Sukhatme G S. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration[J]. International Journal of Robotics Research, 2011, 30(1): 56-79.
- [28] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2007: 3565-3572.

- [29] Kalman R E. A new approach to linear filtering and prediction problems[J]. *Journal of Basic Engineering*, 1960, 82(1): 35-45.
- [30] Prince S J D. *Computer vision: Models, learning, and inference*[M]. Cambridge, UK: Cambridge University Press, 2012.
- [31] McGee L A, Schmidt S F. *Discovery of the Kalman filter as a practical tool for aerospace and industry*[R]. Washington, USA: NASA, 1985.
- [32] Julier S, Uhlmann J K. *A general method for approximating nonlinear transformations of probability distributions*[R]. Oxford, UK: University of Oxford, 1996.
- [33] Thrun S, Fox D, Burgard W, et al. Robust Monte Carlo localization for mobile robots[J]. *Artificial Intelligence*, 2001, 128(1-2): 99-141.
- [34] Weiss S, Achtelik M W, Lynen S, et al. Monocular vision for long-term micro aerial vehicle state estimation: A compendium[J]. *Journal of Field Robotics*, 2013, 30(5): 803-831.
- [35] Newcombe R A, Lovegrove S J, Davison A J. DTAM: Dense tracking and mapping in real-time[C]//*IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE, 2011: 2320-2327.
- [36] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. *International Journal of Robotics Research*, 2012, 31(5): 647-663.
- [37] Smith R, Self M, Cheeseman P. *Estimating uncertain spatial relationships in robotics*[M]//*Autonomous Robot Vehicles*. New York, USA: Springer, 1990: 167-193.
- [38] Durrant-Whyte H F. Uncertain geometry in robotics[J]. *IEEE Journal on Robotics and Automation*, 1988, 4(1): 23-31.
- [39] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty[J]. *International Journal of Robotics Research*, 1986, 5(4): 56-68.
- [40] Smith R, Self M, Cheeseman P. A stochastic map for uncertain spatial relationships[C]//*4th International Symposium of Robotics Research*. Cambridge, USA: MIT Press, 1988: 467-474.
- [41] Li M Y, Mourikis A I. High-precision, consistent EKF-based visual-inertial odometry[J]. *International Journal of Robotics Research*, 2013, 32(6): 690-711.
- [42] Zhu A Z, Atanasov N, Daniilidis K. Event-based visual inertial odometry[C]//*30th IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2017: 5816-5824.
- [43] Zheng X, Moratto Z, Li M Y, et al. Photometric patch-based visual-inertial odometry[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2017: 3264-3271.
- [44] Sun K, Mohta K, Pfrommer B, et al. Robust stereo visual inertial odometry for fast autonomous flight[J]. *IEEE Robotics and Automation Letters*, 2018, 3(2): 965-972.
- [45] Huai Z, Huang G Q. Robocentric visual-inertial odometry[C]//*IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2018: 6319-6326.
- [46] Bloesch M, Omani S, Hutter M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]//*IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2015: 298-304.
- [47] Bloesch M, Burri M, Omari S, et al. Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback[J]. *International Journal of Robotics Research*, 2017, 36(10): 1053-1072.
- [48] Schneider T, Dymczyk M, Fehr M, et al. Maplab: An open framework for research in visual-inertial mapping and localization[J]. *IEEE Robotics and Automation Letters*, 2018, 3(3): 1418-1425.
- [49] Faessler M, Fontana F, Forster C, et al. Autonomous, vision-based flight and live dense 3D mapping with a quadrotor micro aerial vehicle[J]. *Journal of Field Robotics*, 2016, 33(4): 431-450.
- [50] Rosten E, Drummond T. *Machine learning for high-speed corner detection*[C]//*9th European Conference on Computer Vision*. Berlin, Germany: Springer, 2006: 430-443.
- [51] Lynen S, Achtelik M W, Weiss S, et al. A robust and modular multi-sensor fusion approach applied to MAV navigation [C]//*IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, USA: IEEE, 2013: 3923-3929.
- [52] Triggs B, McLauchlan P F, Hartley R I, et al. *Bundle adjustment – A modern synthesis*[C]//*International Workshop on Vision Algorithms*. Berlin, Germany: Springer, 2000: 298-372.
- [53] Kümmerle R, Grisetti G, Strasdat H, et al. g²o: A general framework for graph optimization[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2011: 3607-3613.
- [54] Agarwal S, Mierle K. Ceres solver[DB/OL]. (2012-05-01) [2019-06-15]. <http://ceres-solver.org>.
- [55] Dellaert F. *Factor graphs and GTSAM: A hands-on introduction*[R]. Atlanta, USA: Georgia Institute of Technology, 2012.
- [56] 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.
- [57] Leutenegger S, Furgale P, Rabaud V, et al. Keyframe-based visual-inertial SLAM using nonlinear optimization[C]//*Robotics: Science and Systems*. 2013. DOI: 10.15607/RSS.2013.IX.037.
- [58] Mourikis A I, Roumeliotis S I. A dual-layer estimator architecture for long-term localization[C]//*IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2008: 1310-1317.
- [59] Sibley G, Matthies L, Sukhatme G. Sliding window filter with application to planetary landing[J]. *Journal of Field Robotics*, 2010, 27(5): 587-608.
- [60] Dong-Si T C, Mourikis A I. Motion tracking with fixed-lag smoothing: Algorithm and consistency analysis[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2011: 5655-5662.
- [61] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM: Why filter?[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2010: 2657-2664.
- [62] Nerurkar E D, Wu K J, Roumeliotis S I. C-KLAM: Constrained keyframe-based localization and mapping[C]//*IEEE International Conference on Robotics and Automation*. Piscataway, USA: IEEE, 2014: 3638-3643.
- [63] Klein G, Murray D. Parallel tracking and mapping on a camera phone[C]//*8th IEEE International Symposium on Mixed and Augmented Reality*. Piscataway, USA: IEEE, 2009: 83-86.

- [64] Maybeck P S. Stochastic models, estimation, and control[M]. New York, USA: Academic Press, 1982.
- [65] Huang G Q, Mourikis A I, Roumeliotis S I. An observability-constrained sliding window filter for SLAM[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2011: 65-72.
- [66] Hesch J A, Kottas D G, Bowman S L, et al. Camera-IMU-based localization: Observability analysis and consistency improvement[J]. International Journal of Robotics Research, 2014, 33(1): 182-201.
- [67] Harris C, Stephens M. A combined corner and edge detector[C]//4th Alvey Vision Conference. 1988: 147-151.
- [68] Leutenegger S, Chli M, Siegwart R. BRISK: Binary robust invariant scalable keypoints[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2548-2555.
- [69] 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.
- [70] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//7th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers Inc, 1981: 674-679.
- [71] 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.
- [72] 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.
- [73] Li P L, Qin T, Hu B T, et al. Monocular visual-inertial state estimation for mobile augmented reality[C]//IEEE International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2017: 11-21.
- [74] Qin T, Shen S J. Online temporal calibration for monocular visual-inertial systems[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2018: 3662-3669.
- [75] Liu H M, Chen M Y, Zhang G, et al. ICE-BA: Incremental, consistent and efficient bundle adjustment for visual-inertial SLAM[C]//31st IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 1974-1982.
- [76] Mur-Artal R, Tardós J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.
- [77] 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.
- [78] Usenko V, Engel J, Stückler J, et al. Direct visual-inertial odometry with stereo cameras[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2016: 1885-1892.
- [79] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2564-2571.
- [80] Calonder M, Lepetit V, Strecha C, et al. BRIEF: Binary robust independent elementary features[C]//11th European Conference on Computer Vision. Berlin, Germany: Springer, 2010: 778-792.
- [81] Mur-Artal R, Tardós J D. Fast relocalisation and loop closing in keyframe-based SLAM[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 846-853.
- [82] Mur-Artal R, Tardós J D. ORB-SLAM: Tracking and mapping recognizable features[C]//Robotics: Science and Systems (RSS) Workshop on Multi View Geometry. 2014.
- [83] Mur-Artal R, Tardós J D. Probabilistic semi-dense mapping from highly accurate feature-based monocular SLAM[C]//Robotics: Science and Systems. 2015. DOI: 10.15607/RSS.2015.XI.041.
- [84] 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.
- [85] 张玉龙. 基于关键帧的视觉惯性 SLAM 算法 [D]. 天津: 天津大学, 2017.
- Zhang Y L. Keyframe-based visual-inertial SLAM algorithm [D]. Tianjin: Tianjin University, 2017.
- [86] 张玉龙, 张国山. 基于关键帧的视觉惯性 SLAM 闭环检测算法 [J]. 计算机科学与探索, 2018, 12(11): 1777-1787.
- Zhang Y L, Zhang G S. Loop-closing detection algorithm of keyframe-based visual-inertial SLAM[J]. Journal of Frontiers of Computer Science and Technology, 2018, 12(11): 1777-1787.
- [87] 代维. 室内环境下基于视觉/惯性/里程计的自主定位技术研究 [D]. 南京: 东南大学, 2018.
- Dai W. Research on autonomous positioning technology based on visual inertial mileage meter in indoor environment[D]. Nanjing: Southeast University, 2018.
- [88] 徐晓苏, 代维, 杨博, 等. 室内环境下基于图优化的视觉惯性 SLAM 方法 [J]. 中国惯性技术学报, 2017, 25(3): 313-319.
- Xu X S, Dai W, Yang B, et al. Visual-aid inertial SLAM method based on graph optimization in indoor[J]. Journal of Chinese Inertial Technology, 2017, 25(3): 313-319.
- [89] Jung S H, Taylor C J. Camera trajectory estimation using inertial sensor measurements and structure from motion results [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2001: 732-737.
- [90] Strelow D, Singh S. Motion estimation from image and inertial measurements[J]. International Journal of Robotics Research, 2004, 23(12): 1157-1195.
- [91] Bryson M, Johnson-Roberson M, Sukkarieh S. Airborne smoothing and mapping using vision and inertial sensors[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2009: 3143-3148.
- [92] Indelman V, Williams S, Kaess M, et al. Information fusion in navigation systems via factor graph based incremental smoothing[J]. Robotics and Autonomous Systems, 2013, 61(8): 721-738.
- [93] Patron-Perez A, Lovegrove S, Sibley G. A spline-based trajectory representation for sensor fusion and rolling shutter cameras[J]. International Journal of Computer Vision, 2015, 113(3): 208-219.

- [94] Kaess M, Ranganathan A, Dellaert F. iSAM: Incremental smoothing and mapping[J]. *IEEE Transactions on Robotics*, 2008, 24(6): 1365-1378.
- [95] Kaess M, Johannsson H, Roberts R, et al. iSAM2: Incremental smoothing and mapping using the Bayes tree[J]. *International Journal of Robotics Research*, 2012, 31(2): 216-235.
- [96] Dellaert F, Kaess M. Factor graphs for robot perception[J]. *Foundations and Trends in Robotics*, 2017, 6(1-2): 1-139.
- [97] Koller D, Friedman N. Probabilistic graphical models: Principles and techniques[M]. Cambridge, USA: MIT Press, 2009.
- [98] Kaess M, Dellaert F, Roberts R, et al. The Bayes tree: Enabling incremental reordering and fluid relinearization for online mapping[R]. Cambridge, USA: MIT Press, 2010.
- [99] Huang G Q. Improving the consistency of nonlinear estimators: Analysis, algorithms, and applications[D]. Minneapolis, USA: University of Minnesota, 2012.
- [100] Martinelli A. State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration[J]. *IEEE Transactions on Robotics*, 2011, 27(2): 239-255.
- [101] Weiss S M. Vision based navigation for micro helicopters[D]. Zurich, Switzerland: ETH, 2012.
- [102] Martinelli A. Visual-inertial structure from motion: Observability vs minimum number of sensors[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 1020-1027.
- [103] Hesch J A, Kottas D G, Bowman S L, et al. Towards consistent vision-aided inertial navigation[C]//10th Workshop on the Algorithmic Foundations of Robotics. Berlin, Germany: Springer, 2013: 559-574.
- [104] Huang G Q, Kaess M, Leonard J J. Towards consistent visual-inertial navigation[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 4926-4933.
- [105] Kottas D G, Hesch J A, Bowman S L, et al. On the consistency of vision-aided inertial navigation[C]//13th International Symposium on Experimental Robotics. Berlin, Germany: Springer, 2013: 303-317.
- [106] Martinelli A. Observability properties and deterministic algorithms in visual-inertial structure from motion[J]. *Foundations and Trends in Robotics*, 2013, 3(3): 139-209.
- [107] Huang G Q, Mourikis A I, Roumeliotis S I. A first-estimates Jacobian EKF for improving SLAM consistency[C]//11th International Symposium on Experimental Robotics. Berlin, Germany: Springer, 2009: 373-382.
- [108] Guo C X, Roumeliotis S I. IMU-RGBD camera 3D pose estimation and extrinsic calibration: Observability analysis and consistency improvement[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2013: 2935-2942.
- [109] Martinelli A. Visual-inertial structure from motion: Observability and resolvability[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2013: 4235-4242.
- [110] Hesch J A, Kottas D G, Bowman S L, et al. Observability-constrained vision-aided inertial navigation[R]. Minneapolis, USA: University of Minnesota, 2012.
- [111] Civera J, Bueno D R, Davison A J, et al. Camera self-calibration for sequential Bayesian structure from motion[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2009: 403-408.
- [112] Mirzaei F M, Roumeliotis S I. A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1143-1156.
- [113] Li M Y, Mourikis A I. Online temporal calibration for camera-IMU systems: Theory and algorithms[J]. *International Journal of Robotics Research*, 2014, 33(7): 947-964.
- [114] Kelly J, Sukhatme G S. A general framework for temporal calibration of multiple proprioceptive and exteroceptive sensors[C]//12th International Symposium on Experimental Robotics. Berlin, Germany: Springer, 2014: 195-209.
- [115] Indelman V, Williams S, Kaess M, et al. Factor graph based incremental smoothing in inertial navigation systems[C]//15th International Conference on Information Fusion. Piscataway, USA: IEEE, 2012: 2154-2161.
- [116] Indelman V, Williams S, Kaess M, et al. Information fusion in navigation systems via factor graph based incremental smoothing[J]. *Robotics and Autonomous Systems*, 2013, 61(8): 721-738.
- [117] Indelman V, Melim A, Dellaert F. Incremental light bundle adjustment for robotics navigation[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2013: 1952-1959.
- [118] Shen S J. Autonomous navigation in complex indoor and outdoor environments with micro aerial vehicles[D]. Philadelphia, USA: University of Pennsylvania, 2014.
- [119] Keivan N, Patron-Perez A, Sibley G. Asynchronous adaptive conditioning for visual-inertial SLAM[C]//14th International Symposium on Experimental Robotics. Cham, Switzerland: Springer, 2016: 309-321.
- [120] Lupton T, Sukkarieh 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.
- [121] Forster C, Carlone L, Dellaert F, et al. IMU preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation[C]//Robotics Science and Systems. 2015. DOI: 10.15607/rss.2015.xi.006.
- [122] Shen S J, Michael N, Kumar V. Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs [C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2015: 5303-5310.
- [123] Huang W B, Liu H. Online initialization and automatic camera-IMU extrinsic calibration for monocular visual-inertial SLAM[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 5182-5189.
- [124] Delmerico J, Scaramuzza D. A benchmark comparison of monocular visual-inertial odometry algorithms for flying robots [C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 2502-2509.
- [125] 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.
- [126] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. *International Journal of Robotics Research*, 2013, 32(11): 1231-1237.

- [127] Schubert D, Goll T, Demmel N, et al. The TUM VI benchmark for evaluating visual-inertial odometry[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2018: 1680-1687.
- [128] Majdik A L, Till C, Scaramuzza D. The Zurich urban micro aerial vehicle dataset[J]. International Journal of Robotics Research, 2017, 36(3): 269-273.
- [129] Blanco-Claraco J L, Moreno-Dueñas F Á, González-Jiménez J. The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario[J]. International Journal of Robotics Research, 2014, 33(2): 207-214.
- [130] Carlevaris-Bianco N, Ushani A K, Eustice R M. University of Michigan North Campus long-term vision and lidar dataset[J]. International Journal of Robotics Research, 2016, 35(9): 1023-1035.
- [131] Pfrommer B, Sanket N, Daniilidis K, et al. PennCOSYVIO: A challenging visual inertial odometry benchmark[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2017: 3847-3854.
- [132] 赵洋, 刘国良, 田国会, 等. 基于深度学习的视觉 SLAM 综述 [J]. 机器人, 2017, 39(6): 889-896.
Zhao Y, Liu G L, Tian G H, et al. A survey of visual SLAM based on deep learning[J]. Robot, 2017, 39(6): 889-896.
- [133] 伍锡如, 黄国明, 孙立宁. 基于深度学习的工业分拣机器人快速视觉识别与定位算法 [J]. 机器人, 2016, 38(6): 711-719.
Wu X R, Huang G M, Sun L N. Fast visual identification and location algorithm for industrial sorting robots based on deep learning[J]. Robot, 2016, 38(6): 711-719.
- [134] Wang S, Clark R, Wen H, et al. End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks[J]. International Journal of Robotics Research, 2018, 37(4-5): 513-542.
- [135] Wang S, Clark R, Wen H, et al. DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2017: 2043-2050.
- [136] Abolfazli Esfahani, M, Wang H, Wu K, et al. AbolDeepIO: A novel deep inertial odometry network for autonomous vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(5): 1941-1950.
- [137] Clark R, Wang S, Wen H, et al. VINet: Visual-inertial odometry as a sequence-to-sequence learning problem[C]//31st AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI, 2017: 3995-4001.
- [138] Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: Evolution of optical flow estimation with deep networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017. DOI: 10.1109/CVPR.2017.179.
- [139] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [140] Shamwell E J, Lindgren K, Leung S, et al. Unsupervised deep visual-inertial odometry with online error correction for RGB-D imagery[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. DOI:10.1109/TPAMI.2019.2909895.
- [141] Chen C, Rosa S, Miao Y, et al. Selective sensor fusion for neural visual-inertial odometry[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2019: 10534-10543.
- [142] Lee H, McCrink M H, Gregory J W. Visual-inertial odometry for unmanned aerial vehicle using deep learning[C]//AIAA Scitech 2019 Forum. Reston, USA: AIAA, 2019. DOI: 10.2514/6.2019-1410.
- [143] Wan G W, Yang X L, Cai R L, et al. Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 4670-4677.
- [144] Li P L, Qin T, Hu B T, et al. Monocular visual-inertial state estimation for mobile augmented reality[C]//IEEE International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2017: 11-21.
- [145] 林辉灿, 吕强, 张洋, 等. 稀疏和稠密的 VSLAM 的研究进展 [J]. 机器人, 2016, 38(5): 621-631.
Lin H C, Lü Q, Zhang Y, et al. The sparse and dense VSLAM: A survey[J]. Robot, 2016, 38(5): 621-631.
- [146] Nikolskiy V P, Stegailov V V, Vecher V S. Efficiency of the Tegra K1 and X1 systems-on-chip for classical molecular dynamics[C]//International Conference on High Performance Computing & Simulation. Piscataway, USA: IEEE, 2016: 682-689.

作者简介:

施俊屹 (1995-), 男, 硕士生. 研究领域: SLAM, 3 维重建, 场景识别.

查富生 (1974-), 男, 博士, 副教授. 研究领域: 足式机器人, 人工智能与自生长网络, 仿生信息处理方法, 神经信息编解码方法.

孙立宁 (1964-), 男, 博士, 教授. 研究领域: 纳米级微驱动及微操作机器人, 工业机器人技术, 医疗机器人, 仿人手臂, 机器人机构与控制.