

# 移动机器人同步定位与建图技术综述

任伟建<sup>1,2</sup>, 高 强<sup>1,2</sup>, 康朝海<sup>1,2</sup>, 霍凤财<sup>1,2</sup>, 张志强<sup>1,2</sup>

(1. 东北石油大学 电气信息工程学院, 黑龙江 大庆 163318;

2. 黑龙江省网络化与智能控制重点实验室, 黑龙江 大庆 163318)

**摘要:** 同步定位与建图技术 (SLAM) 一直是移动机器人领域比较热门的研究方向, 它可以给机器人提供强大的环境感知能力; 传统的依靠外部位置参考来定位的方法如果无法获得时, 移动机器人需要即时定位自身位置来构建增量式地图, 因此 SLAM 技术也就应运而生; 对激光 SLAM 和视觉 SLAM 的研究现状及最新标志性成果进行了介绍, 重点对以相机与激光雷达融合、相机与 IMU 融合、激光雷达与 IMU 融合为代表的多传感器融合 SLAM 技术展开讨论、系统地梳理了几种融合方式的优势与不足, 同时介绍了该领域的研究热点语义 SLAM, 最后讨论了 SLAM 技术在该领域未来的发展方向以及存在的挑战。

**关键词:** 视觉 SLAM; 激光 SLAM; 多传感器融合 SLAM; 语义 SLAM

## Overview of Simultaneous Localization and Mapping Technology of Mobile Robots

REN Weijian<sup>1,2</sup>, GAO Qiang<sup>1,2</sup>, KANG Chaohai<sup>1,2</sup>, HUO Fengcai<sup>1,2</sup>, ZHANG Zhiqiang<sup>1,2</sup>

(1. Department of Electrical Information Engineering, Northeast Petroleum University, Daqing 163318, China;

2. Heilongjiang Provincial Key Laboratory of Networking and Intelligent Control, Daqing 163318, China)

**Abstract:** Simultaneous Localization and Mapping (SLAM) always is the popular research direction in the field of mobile robots. It can provide robots with powerful environmental perception capabilities. If the traditional localization method relying on external location reference is not available, the mobile robot needs to build incremental maps by locating its own position in time. Therefore, the SLAM technology comes into being. Introduces the new research status and the latest landmark achievements of laser SLAM and visual SLAM, and focus on discussing the multi-sensor fusion SLAM technology, which is represented by camera and lidar fusion, camera and Inertial Measurement Unit (IMU) fusion, and lidar and IMU fusion, and systematically sort out the advantages and disadvantages of several fusion methods. At the same time, the research hotspot of semantic SLAM in this field is introduced. Finally, it is discussed that the future development direction of SLAM technology in this field and the existing challenges.

**Keywords:** visual simultaneous localization and mapping (SLAM); lidar SLAM; multi-sensor fusion SLAM; semantic SLAM

收稿日期: 2021-08-17; 修回日期: 2021-09-09。

基金项目: 国家自然科学基金资助项目(61933007, 61873058); 黑龙江省自然科学基金(F2018004, F2018005)。

作者简介: 任伟建(1963-), 女, 黑龙江大庆人, 教授, 博士生导师, 主要从事油气集输过程故障诊断方向的研究。

通讯作者: 康朝海(1976-), 男, 黑龙江望奎人, 副教授, 硕士生导师, 主要从事智能控制、模式识别、复杂系统与控制等方向的研究。

引用格式: 任伟建, 高 强, 康朝海, 等. 移动机器人同步定位与建图技术综述[J]. 计算机测量与控制, 2022, 30(2): 1-10, 37.

## 0 引言

SLAM 技术是当前移动机器人定位技术主流的研究方法之一。随着移动机器人应用领域的复杂性不断增加,高效的定位导航系统是控制系统和路径规划的前提,特别是当无法获得诸如全球定位系统(GPS)等外部位置参考时,需要移动机器人即时定位自身位置来构建增量式地图。SLAM 技术研究如何一致地将传感器收集到的信息集成到数学模型上来实现建图功能,与此同时估测机器人的坐标和姿势形态实现定位功能<sup>[1]</sup>。实际上,在误差和噪音条件下,这两部分如果分开解决,将毫无意义,必须在各自进程中同时求得连续解;同时,不同进程中的相互迭代反馈对连续解的求解起到一定作用。因此,SLAM 问题是一个不断迭代而又复杂的耦合问题。

SLAM 技术起源于 1986 年在旧金山举行的 IEEE Robotics and Automation Conference,这是一个概率学方法刚刚开始被引入到机器人学和人工智能学的年代。随后,R. C. Smith 等将统计贝叶斯方法用在了描述位姿和几何不确定性之间的关系上<sup>[2]</sup>,它表明在机器人移动过程当中对于不同位姿的估计之间必然存在高度的相关性,这种用概率的方法来描述空间的不确定性与位姿的关系为后期 SLAM 技术的快速发展提供了一种理论基础。因此在之后的一段时间里,SLAM 问题的研究进入了所谓的“经典时期”(1986~2004),这一时期研究的主要内容包括基于扩展卡尔曼滤波方法(EKF, extended Kalman filters)<sup>[3]</sup>、粒子滤波方法(RBPF, rao-blackwellized)<sup>[4]</sup>、极大似然估计方法(MLE, maximum likelihood estimate)<sup>[5]</sup>,这些方法在 SLAM 研究上遇到的最大的问题是当样本点较多时,由于计算复杂度比较高,很难高效率地在大规模地图上构建地图;2004~2015 年进入算法改进时期,人们对 SLAM 基本理论作了深入探讨和拓展,Dissanayake 等人对 SLAM 基本性质(收敛性、可观测性和一致性)以及 SLAM 算法的时间复杂度进行深度分析<sup>[6]</sup>,Cadena 等人对 SLAM 算法的鲁棒性、可伸缩性、地图构建形式等问题做了系统性的综述<sup>[7]</sup>,在这一时期,研究者们提出的稀疏特征方法在 SLAM 技术中起到重要作用,开源了主要的 SLAM 库,存在最大的问题是在未知复杂环境中系统的鲁棒性较差;2015

年以后 SLAM 算法进入鲁棒感知时期,比较有代表的是基于稀疏直接法的视觉里程计(DSO, direct sparse odometry)<sup>[8]</sup>、ORB-SLAM2<sup>[9]</sup>等算法,这一时期 SLAM 问题最大的挑战是构建一致收敛的轨迹和地图并且控制计算量大小以及如何在保证移动机器人在高精度定位的同时最小化时间复杂度。

随着计算机处理性能的大幅提升,基于激光雷达、惯性测量单元(IMU, inertial measurement unit)和相机的 SLAM 定位导航技术逐步在移动机器人上有了较好的表现。根据移动机器人感知周围环境传感器的分类,SLAM 技术主要分为:激光 SLAM(lidar SLAM)、视觉 SLAM(visual SLAM)以及语义 SLAM(semantic SLAM),但是由于 SLAM 算法应用的场景比较广泛,覆盖室内外、海洋以及航天等复杂环境,激光雷达、惯性测量单元和视觉相机都有一定局限性,比如:IMU 系统累计误差较大;激光 SLAM 重定位能力较差,在追踪丢失后难以回到工作状态,且受雷达探测范围的限制较大;视觉 SLAM 在无纹理或者光照弱的环境下表现较差,易受到动态物体或光源的干扰,且在距离未知的情况下,存在尺度漂移、计算复杂度高等缺点;而语义 SLAM 研究依赖神经网络学习图像的层次化特征,复杂环境下其本身在正确率上存在限制,所以在目前效果往往不尽如人意。

本文第二和第三节分别对激光和视觉 SLAM 技术现状展开综述,并从算法性能等指标上列出了常用的开源视觉 SLAM 方案。第四节综述了当前 SLAM 技术的研究热点:多传感器融合 SLAM、语义 SLAM 的研究现状。第五节指出当前 SLAM 研究的技术难点并为未来的研究提供了几个方向。

## 1 激光 SLAM

激光 SLAM 任务是移动机器人搭载激光雷达传感器来估计自身的位姿,同时建立周围的环境地图,主要应用于室内环境。有许多算法框架用来解决激光 SLAM 问题,通常根据数学优化思想主要分为两大类:基于滤波器(filter-based)以及基于图优化(graph-based)的激光 SLAM。

### 1.1 基于滤波器的激光 SLAM 方案

基于滤波器的激光 SLAM 方案是激光 SLAM 中较为早期的位姿估计与建图方案。在进行机器人位姿

估计时主要估计的是机器人在当前时刻的状态, 通过机器人当前时刻的状态进行机器人的定位和位姿的构图。

20 世纪 90 年代初, R. Smith 等人提出基于扩展卡尔曼滤波 SLAM 方案<sup>[10]</sup>, 使用极大似然估计方法对传感器获取的数据进行关联。该方案的缺点是构建的地图特征不明显, 数据处理能力不足、鲁棒性差。

由 Milios 等人在 1997 年提出的全局一致的范围扫描以进行环境映射思想<sup>[11]</sup>, 利用局部与全局的相对空间关系来制定准则, 组合出最优的空间关系作为约束来解决数据帧姿态, 从而实现一致性。该思想的不足之处是连续扫描将生成大量必须采样的数据, 因此数据量过多问题在当时的计算水平下很难实现。

Chatila 等人在 1999 年提出了循环环境的增量映射方案, 利用了一种局部配准和全局相关的方法, 渐进式地从密集范围数据中重建一致的全局地图, 并以恒定时间运行, 但是不可避免的会出现累计漂移过大的问题<sup>[12]</sup>。

Montemerlo 等人于 2003 年提出了 FastSLAM 方案<sup>[13]</sup>, 该方案是基于马尔可夫聚类算法以及低维卡尔曼滤波来完成定位和地图构建的, 是最早能够实时输出栅格地图的激光 SLAM 方案。该方案的不足之处是对于大尺度环境, 若里程计误差较大则严重消耗内存, 且由于重采样的随机性以及采样次数增多, 粒子耗散问题对地图的构建影响较大。

Blanco 等人在 2010 年提出了基于 RBPF 的 Gmapping 方案<sup>[14]</sup>, 在一定的条件下, 部分状态空间向量可以通过利用解析的方法求得, 在里程计模型传播时选择最优的粒子作为真实传播, 从而减少粒子个数和防止最优粒子退化的情况。该方案不足在于非常依赖于里程计信息, 不适合构建大场景地图, 应选择周围具有较多特征物的点, 这样在进行回环的过程中更容易提高正确粒子所占的权重。Gmapping 方案在较小的环境中能实现较好的建图效果, 是目前应用最为广泛的基于二维激光雷达的 SLAM 算法。

总体而言, 基于滤波器的激光 SLAM 系统计算资源受限、待估计量比较简单, 滤波方法比较有效。而 Gmapping 方案在小场景已经成熟, 后续没有实质性的发展, 随着现代硬件计算能力的不断提升, 人们更多地关注运用图优化的方式来解决激光 SLAM 问题。

## 1.2 基于图优化的激光 SLAM 方案

基于图优化的激光 SLAM 方案是在滤波器方案后发展的新兴优化机器人位姿方案。图优化的激光 SLAM 可以将先前位姿信息和当前的位姿信息同时进行优化, 进而提高定位的精确度。在进行机器人的位姿估计时图优化主要以构建点云的增量地图的顶点为机器人走过的位姿, 对于大场景以及精度要求不高的环境具有重要的价值。

Konolige 等人在 2010 年开源了第一个基于图优化算法框架的方案 Karto SLAM<sup>[15]</sup>, 此方案系统的阐述了稀疏性在 SLAM 中起的作用, 用高度优化和非迭代 cholesky 分解思想进行系统解耦, 在整个激光 SLAM 系统中表现较好, 因此在一定程度上替代了基于滤波器的方案。该方案的不足之处在于: 系统局部地图匹配之前需要构建好子图, 当搜索范围大时初始响应较慢, 容易陷入局部地图更新失败。

Kohlbrecher 等人在 2011 年提出的 Hector-SLAM 算法的主要过程是将激光数据点转换到地图中<sup>[16]</sup>, 使用高斯牛顿方法通过与地图环境的匹配找到匹配程度最高的机器人位姿, 不需要里程计数据, 只根据激光信息便可构建地图, 可以用于地面不平坦区域及空中飞行器。Hector-SLAM 算法的不足之处是对激光雷达传感器精度要求较高, 且不适用于高速运动 (尤其是强旋转) 的场景, 在大地图, 低特征场景中, Hector-SLAM 的建图误差明显高于 Gmapping。

谷歌公司于 2016 年提出的基于图优化形式的 Cartographer 开源方案是对 Karto SLAM 算法的优化<sup>[17]</sup>, 核心思想是用闭环检测来消除构图过程中产生的累计误差。闭环检测主要采用分支界定的思想: 分支就是进行拓展, 定界就是剪枝, 以此来提高运算效率, 并且 Cartographer 方案支持多传感器数据、2D SLAM 和 3D SLAM 建图。由于算法的实时性相对较高, 比较完善的匹配系统, 包含建图和定位, 一经发布, 就受到对精度要求不是太高市场的青睐。该方案的不足是拿存储空间来换时间效率, 为保证实时需求, 因而闭环检测策略比较落后, 尤其是在一些几何对称的环境中, 易引起不稳定的闭环。

综上, 基于图优化的激光 SLAM 允许大规模场景, 通过非线性优化方法优化累计误差, 允许简单的闭环检测过程, 但没有认识到系统的稀疏性, 实时性

有待提高<sup>[18]</sup>。总体而言,激光 SLAM 算法框架已确立,相对理论已被研究的较为透彻,成为目前室内机器人领域主流的定位方法。

## 2 视觉 SLAM

随着 CPU 和 GPU 计算能力的不断提升以及图形处理能力的快速发展,视觉 SLAM 技术因其包含的信息量大、适用范围广、泛化性强等优点受到研究者广泛关注。视觉 SLAM 系统以视觉传感器获取的序列图像数据为基础,利用图像信息和成像模型来估计位姿。过去的十几年,开发出了一套比较成熟的视觉 SLAM 框架,可分为信息采集、前端、后端、闭环检测以及建图五大部分,视觉 SLAM 框架介绍如图 1 所示。

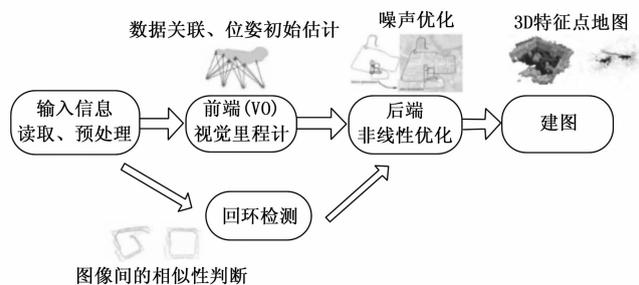


图 1 视觉 SLAM 框架

传感器对采集到的数据进行预处理后交给前端,视觉里程计把相邻时刻的运动“串”起来构成了机器人的运动轨迹,根据每个时刻相机的位置来计算出各像素对应的空间点的位姿,在这个过程中产生的运动畸变以及每次估计的误差,将不可避免地出现累计漂移,为了解决漂移问题,引入回环检测以及后端优化部分,回环检测部分负责把“机器人是否回到原点”的问题检测出来,而后端优化处理在此过程中产生的噪声问题,以此来校正机器人行进的轨迹形状。在视觉 SLAM 系统中总体来说,前端往往和计算机视觉研究领域更为相关,例如图像的特征提取与匹配等,后端则主要是滤波与非线性优化算法。视觉 SLAM 系统按照是否需要提取特征,主要分为基于特征点法的视觉 SLAM 和直接法的视觉 SLAM。

### 2.1 基于特征点法的视觉 SLAM

基于特征点法的视觉 SLAM 指的是对输入的图像进行特征点检测及提取,常采用较稀疏的特征点,比如 SIFT (scale-invariant feature transform, 尺

度不变特征变换)<sup>[19]</sup>和 ORB (oriented FAST and rotated BRIEF, 旋转不变特征)<sup>[20]</sup>,根据提取特征点的信息进行数据关联,从而对运动位姿进行估计<sup>[21]</sup>。为减小噪声对位姿估计的影响,通常使用 RANSAC 算法 (random sample consensus, 随机抽样一致性)来剔除错误匹配的特征点,同时根据 2D、3D 的特征匹配信息来估计位姿及构建地图,在保存图像重要信息的同时能有效减少计算量<sup>[22]</sup>。

2007 年 Davison 等提出的 MonoSLAM 是基于拓展卡尔曼滤波的单目 SLAM 系统<sup>[23]</sup>,第一次成功应用在移动单目相机条件下,实时且无漂移现象,但 EKF 存在着计算复杂性和由于线性化而精度不高的问题。同年 Michal 等人提出 PTAM (parallel tracking and mapping) 方法<sup>[24]</sup>,首次区分出前端和后端,实现了跟踪和建图的并行化,通过非线性优化的方法引入关键帧,同时将它们串联起来优化移动轨迹和特征方位来增强算法的实时性,如图 2 所示,形成了基于最优化的视觉 SLAM 算法基本形式框架,其后该类型的 SLAM 算法大多延续了 PTAM 的思路。

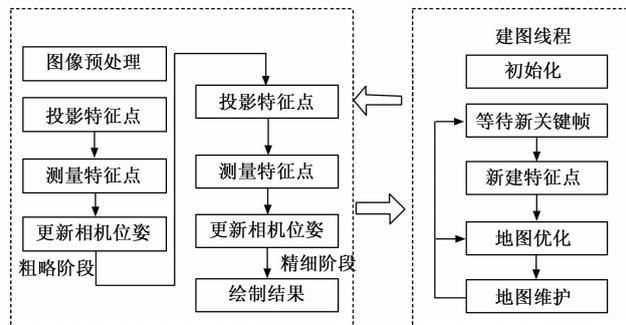


图 2 PTAM 基本框架

随着 RGB-D 传感器 (深度相机) 的出现,基于 PTAM 框架, Mur-Artal 等人在 2015 年提出了比较完善的基于关键帧的单目 ORB-SLAM 算法<sup>[25]</sup>,该系统摒弃之前使用 SURF 特征点只对 3D-3D 对应点进行位姿估计、没有充分利用 2D-2D、2D-3D 对应点信息以及 SURF 特征点的提取计算复杂度高影响系统实时性的缺点,改用更加稳定的 ORB 特征,增加了回环检测功能,该算法首次使用 3 个平行线程执行视觉里程计、后端优化和回环检测,通过从图像中提取 ORB 特征来进行场景的稀疏重构,并进行相机位姿的估计,同时对局部和全局范围内的位姿使用 BA (bundle adjustment) 算法进行

优化<sup>[26]</sup>, 具备较高的精确度, 可以实时运行。2017年ETH Zurich开源的基于优化的ORB-SLAM2方案, 在ORB-SLAM算法基础上新增了对双目相机、RGB-D相机的原生支持, 是目前市场上主流的具有实时性好、地图重用和动态管理关键帧等优点的完整VSLAM系统。2020年最新开源的基于ORB-SLAM2和ORB-SLAM-VI构建的ORB-SLAM3算法<sup>[27]</sup>, 是第一个能够用单目、立体和RGB-D相机, 利用针孔和鱼镜头模型进行视觉、视觉惯性和多地图(multi-map)的VSLAM系统, 在视觉里程计部分甚至IMU初始化阶段利用最大后验概率MAP进行估计, 可长时间在恶劣视觉信息下生存, 当位姿丢失时, 它会重用所有先验信息, 利用BA调整共视关键帧来提高系统的精度。

## 2.2 基于直接法的视觉SLAM

特征点法存在提取特征点与计算描述子耗时较长、忽略可能有用的图像信息以及没有明显纹理信息、找不到足够的匹配点来计算相机运动等缺点, 因此有研究者基于像素灰度不变性的假设, 提出基于直接法的视觉SLAM系统: 不用计算关键点和描述子, 根据像素灰度信息来最小化光度误差(photometric error)估计出相机运动以及点的投影, 把该问题当作能量最小化问题, 位姿估计则通过能量函数进行迭代优化, 只要场景中存在明暗变化, 直接法就可以工作。

2011年提出的基于单目相机的DTAM是直接法的鼻祖<sup>[28]</sup>, 对每个像素点进行概率的深度测量并通过优化的方式构建稠密深度地图, 该方法通过整幅图像的对准来估计相机运动, 有效降低了位姿的不确定性。基于同样的原理, Engel等人于2013年提出了基于直接跟踪的视觉里程计(VO, semi-dense visual odometry)系统, 随后将地图优化融入该VO系统并扩展为LSD-SLAM<sup>[29]</sup>, 将图像中的像素以随机的深度初始化, 并利用新产生的数据不断迭代优化直至收敛应用于半稠密单目SLAM系统, 是单目直接法在VSLAM上成功应用的标志性方法, 但是它对相机内参和曝光非常敏感。Forster等人于2013年提出的SVO(semi-direct visual odometry)采用“半(稀疏)直接法”最小化图像块重投影残差迭代优化更新位姿, 不必计算描述子以及稠密和半稠密信息, 时间复杂度低, 有较好的实时性, 可达到每秒100多

帧的速度。在随后的SVO2.0中<sup>[30]</sup>, 速度最高可到每秒400多帧, 目前来看在实际应用中SVO系统在平视相机中容易丢失, 而且丢失后重定位能力差以及累计误差过大。2016年慕尼黑工业大学的Engel等人公布的基于视觉里程计的DSO方法<sup>[31]</sup>, 它将光度误差模型和所有模型参数融入到优化函数中进行联合优化, 提出了完整的光度标定方法, 并在多个数据集上进行了测试, 达到了很好的精度和速度, 但由于DSO方法依赖于使用梯度下降来优化问题, 目标函数从初始值到最优值在一定场景下不单调, 带来的实际误差相对较大。2020年Liang等人提出了一种基于半直接法的单目视觉SLAM算法, 将帧间信息融合到视觉词袋(BoVW)模型中, 并采用了改进的金字塔词频模型来评分匹配方案, 以解决复杂环境下的低纹理、运动物体以及感知混叠等问题<sup>[32-33]</sup>。

综上, 直接法以更整体的方式处理了数据关联问题, 需要有不错的初始估计和较好的图像, 擅长求解连续图像的定位。特征点法依赖重复性较强的特征工程及正确的特征匹配, 在环境纹理较好, 角点较多时能较好地计算相机运动, 更适合全局匹配与回环检测。

## 3 多传感器融合SLAM

在复杂场景下, 使用单一传感器已无法独自支撑一个完整的实时定位应用。多传感器融合SLAM是一项提高移动机器人定位精度和鲁棒性的技术, 不同的传感器, 意味着不同的信息输入, 将不同传感器获取的信息综合处理, 就可弥补单一传感器的缺陷, 使其对环境的感知更加丰富, 提高SLAM系统对环境的扩展能力。比较典型的多传感器融合方式有: 相机融合激光雷达、相机融合IMU、激光雷达融合IMU等。

### 3.1 相机与激光雷达融合

相机与激光雷达相结合的多传感器结合方式, 通过把生成的点云平面转化为相机坐标系, 从而获取周围环境信息和更加精确的位姿信息, 因此周围环境与激光雷达相结合可以为VSLAM提供点云, 视觉也可以为雷达获得更好的深度信息。

2013年Levinson等人介绍了一种概率监测算法和一个连续校准优化器, 使相机和激光雷达的校准能够在线、自动地进行<sup>[34]</sup>。2015年Singh等人提出了

视觉里程计和激光雷达里程计相结合的总体框架 V-LOAM<sup>[35]</sup>, 这个方法利用视觉里程计估计机器的运动, 基于激光雷达的 scan-match 来优化运动估和点云对准, 从视觉里程计和基于扫描匹配的激光雷达里程计两个方面入手, 同时改进了实时的运动估计和点云配准算法性能。Graeter 等人在 2018 年提出了一种基于 LIDAR 测量的深度提取算法 LIMO<sup>[36]</sup>, 用于单目相机特征轨迹的提取和运动估计, 并利用基于鲁棒关键帧的 BA 算法来估计运动, 该传感器融合的能力在具有竞争力的 KITTI 数据集上得到了较好的验证。2019 年 Shao 等人提出的 VIL-SLAM 方法将紧耦合的双目视觉惯性里程计 (VIO) 与激光雷达映射和激光雷达增强的视觉环路闭合结合在一起, LiDAR 扫描匹配辅助估计相对运动, 利用累计高保真度的 3D 点云来形成一个准确的地图<sup>[37]</sup>。

### 3.2 相机与 IMU 融合

相机与 IMU 结合的融合是当前较好的融合方式, 其具有互补性, 而且尺寸小, 价格低廉, 所以对视觉-惯性导航系统 (VINS, visual inertial navigation system) 探索研究近年来引起了极大的关注, 在移动端的虚拟现实 (VR, virtual reality)、增强现实 (AR, augmented reality) 以及无人系统的自主导航任务中发挥重要作用, 具有重要的理论研究价值和实际应用需求。这种融合方式的主要思想是利用 IMU 为视觉 SLAM 技术提供精准的尺度信息较为准确的运动观测, 从而提升整体定位结果的精确性与鲁棒性。VINS 的基本框架如图 3 所示。

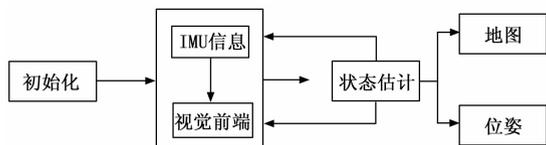


图 3 VINS 基本框架

VINS 系统的状态估计是整个 SLAM 系统的核心问题。主要分为两类: 一类是基于滤波的方案, 另一类是基于优化的方案。

#### 3.2.1 基于滤波的 VINS

2007 年 Mourikis 提出以多状态约束卡尔曼滤波 (MSCKF, multi-state constraint Kalman filter) 的方案<sup>[38]</sup>, 融合了 IMU 信息和视觉信息, 由于相机位姿的个数会远小于特征点的个数, 状态向量的维度相

较 EKF-SLAM 大大降低, 能适应剧烈的运动、一定时间的纹理缺失等场景, 适合在 CPU 资源有限的嵌入式设备上运行, 后来关于滤波方法的 VINS 很多都是基于 MSCKF 的框架实现的, 如在 2013 年 Li 等人提出的一种实时的、能够实现一致估计的 EKF-VIO 算法 (MSCKF2.0)<sup>[39]</sup>, 确保了建立的线性系统模型正确的观测性; 在 2020 年 Geneva 等人提出基于 MSCKF 框架的 Open-VINS 开放性平台<sup>[40]</sup>, 加入了 SLAM 的局部地图提升前端 VINS 定位精度、可扩展的视觉惯性系统模拟器、丰富的算法评估工具箱值得大家学习。

#### 3.2.2 基于优化的 VINS

2015 年 Stefan 等人提出的 OKVIS 算法是紧耦合中比较经典的, 通过 IMU 测量值对当前状态做预测, 从而进行特征提取和特征匹配, 对累计的误差放在一起做优化, 但是它不支持重定位, 也没有闭环校正方案<sup>[41]</sup>。同年 Bloesch 等人提出的 ROVIO 是基于稀疏图像块的 EKF 滤波实现的紧耦合单目 VIO 系统, 使用 FAST 角点特征作为路标点, 对于每一帧图像进行提取图像金字塔, 并在 EKF 框架中的 IMU 预测阶段对路标点进行预测以及特征的匹配跟踪<sup>[42]</sup>。2017 年香港科技大学 Shen Shaojie 团队开源的 VINS-Mono 是一个基于紧耦合滑动窗口非线性优化方法的单目 VIO 系统<sup>[43]</sup>, 提出了鲁棒的初始化步骤, 可以对系统从未知的初始化状态加速以及在线的回环检测和紧耦合重定位, 适合实时无人机导航、大尺度定位和手机 AR 应用。2018 年由百度联合浙江大学共同发提出的 ICE-BA 系统利用 SLAM 问题的增量特性, 高效融合优化视觉特征、惯性传感器数据、回环约束等信息, 保证优化结果的全局一致性同时大幅提升了系统的跟踪能力<sup>[44]</sup>。2019 年浙大开源的基于多平面先验的高效 VIO 提出了一个高效鲁棒的轻量 PVIO 系统<sup>[45]</sup>, 利用视觉-惯性-平面 PnP 算法来实现快速的定位, 同时基于重投影一致性来可靠地扩展平面, 增强在深度高度不确定情况下的跟踪鲁棒性。同年 Ye 等人提出的 PL-VINS 是一种基于实时优化的具有点和线特征的单目视觉惯性定位系统, 使用 LSD 算法来提取线特征, 这种策略类似于流行的基于点的帧到模型策略<sup>[46]</sup>。文献[27]介绍的基于优化的 ORB-SLAM3 方法, 在每次迭代更新时可以通过重线性化的方式来减小线性化误差, 容易获得更高

的精度。

目前来看, 基于滤波的 VINS 方案计算效率高, 实时性好, 但是线性化近似误差可能会导致滤波精度降低, 基于优化的 VINS 方法容易获得更高的精度, 但代价是计算资源的增加。图 4 按照时间节点列举了一些具有代表性的 VINS 算法。

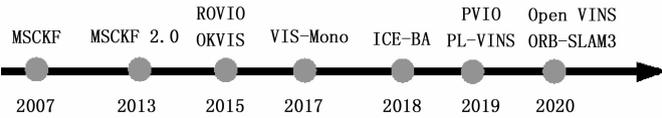


图 4 具有代表性的开源 VINS 算法

### 3.3 激光雷达与 IMU 融合

这种融合方式主要的算法代表为 LOAM、LIO- Mapping 等。通过激光 SLAM 技术和 IMU 技术相融合, 在移动机器人运动的过程中提供当前位姿和估计位姿, 以及点云去畸变过后的状态信息, 进一步对位姿估计和地图构建进行优化。

Singh 等人发表于 2014 年的 LOAM 是基于激光雷达而搭建的在 ROS 平台下的 SLAM 系统<sup>[47]</sup>, 通过融合 IMU 信息来矫正加速度带来的影响, 实现了高频率低精度的处理雷达的速度估计以及低频率高精度的标记点云, 利用边缘点和平面点的特征提取方式在保证高精度建图同时还能保证实时性, 但是 LOAM 系统没有后端优化, 也不能处理大规模的旋转变换。Englot 等人发表于 2018 年的 LeGO- LOAM 是以 LOAM 为框架而衍生出来的新的算法框架<sup>[48]</sup>, 在 LOAM 系统基础之上改进了特征点的提取形式, 添加了后端优化以及基于欧式距离的闭环检测, 因此, 构建出来的地图更加完善。Chen 等人在 2019 年发表于 ICRA 上的基于紧耦合的 LIO- Mapping 系统<sup>[49]</sup>, 总体是借鉴 LOAM 和 VINS- Mono 的思想对激光雷达和 IMU 测量值的代价进行的联合最小化问题, 在快速运动或特征点不充足的情况下, 能够以较高的 IMU 更新频率输出精确的状态估计值, 但没有解决关键帧之间的部分特征使得信息损失较多问题。除了上面介绍的几类融合方式外, 港科大在 2019 年发布的 VINS- Fusion 是基于 VINS- Mono 的一项拓展性工作<sup>[50]</sup>, 这类融合 SLAM 技术不局限于基于某一类硬件配置的定位结果优化, 而是着重于为融合更多的传感器提供一种融合范式 (单目、双目、IMU、GPS), 可以在多样的环境中保持

更高的鲁棒性。

## 4 语义 SLAM

语义 SLAM 将基于神经网络的语义分割、目标检测、实例分割等技术用于 SLAM 系统中, 多用于特征点选取、相机位姿估计, 简而言之语义 SLAM 使用神经网络的方法端到端的图像到位姿、从分割结果来建立标记点云、场景识别。随着数据量和计算能力的不断增长, 基于深度学习的语义 SLAM 技术受到了广泛关注, 它提供了一种以数据驱动的方式解决 SLAM 问题的替代方案, 在 SLAM 中加入语义信息提供高层次的理解、高鲁棒性、资源感知和任务驱动感知。从本质上说, 深度神经网络的输入是一对连续的图像, 输出是两帧图像之间的估计位姿。

Konda 等人在 2015 年提出将视觉里程计作为一个分类问题, 使用卷积神经网络 (ConvNet) 来预测输入图像方向和速度的离散变化。同年 Costante 等人提出了使用卷积神经网络从密集光流中提取视觉特征来输出帧到帧的运动估计<sup>[51]</sup>, 但是它没有实现从图像到运动估计的端到端学习。2017 年 Bowman 等人提出在同个 3D 点重投影之后, 应保持语义一致, 其量化重投影误差的方法在于: 使用概率模型计算出来的物体中心, 重投影到图像上, 应该接近检测框的中心, 数据关联由一组权重决定, BA 和权重重新通过 EM (expectation-maximization) 算法交替优化。Clark 等人提出了基于深度循环卷积神经网络的端到端视觉里程计 DeepVO<sup>[52]</sup>, 直接从一系列原始的 RGB 图像中解算姿态, 可以隐式地从大量的图像中学习和保持全局尺度。Wen 等人提出的 VINet 是第一个将视觉惯导里程计定义为顺序学习问题的工作, 使用长短期记忆 (LSTM) 网络从 IMU 数据序列中提取惯导特征, 以端到端方式实现 VIO。2018 年 Li 等人提出的一种基于非监督深度学习的单目视觉里程计 UnDeepVo 系统<sup>[53]</sup>, 旨在利用立体影像对获取的尺度来训练, 从而得到绝对的尺度, 在单目方法中是最先进的。2020 年由 Stumberg 等人<sup>[54]</sup>提出的 D3VO 框架利用深度学习的深度估计、位姿估计、不确定度估计紧密结合到视觉里程计方法中, 来提升前端追踪以及后端非线性优化性能, 在 KITTI、EuRoC 数据集上都取得了不错的效果。

在语义 SLAM 中, 各个物体在图像中位置变化

可以被预测出来,产生大量的新数据为语义任务提供更多优化条件,且节省人工标定的成本,有利于机器人自主理解和人机交互。尽管目前的深度学习解决方案仍处于初级阶段,现有的模型还不够成熟,但深度学习为 SLAM 的研究带来了新鲜血液,随着深度学习的不断发展,语义 SLAM 势必会成为一个研究热点。

## 5 目前难点和未来研究方向

在移动机器人导航系统中,SLAM 技术起着十分重要的作用,当前,SLAM 技术大部分专注于对噪声的优化管理以及从位姿匹配中获得的更快、更可靠的地标信息解决方案。但是,大多数 SLAM 技术基于高精度定位与实时性建图两种并行方法中获得模糊方法,在实时性以及准确性两个方面需要找到平衡。SLAM 技术包含资源的分配、地图优化、系统关键参数及变量的不确定性、位姿跟踪的准确性、多线程并发执行等,这导致了 SLAM 技术在计算成本和内存成本上仍然有很多亟待解决的问题,这些问题也同时是移动机器人导航领域未来主要的研究方向。

1) 硬件的适配。SLAM 是对传感器获取的数据进行处理,如多传感器融合 SLAM 技术中,各个模块传感器的精度对机器人定位导航的效果影响很大,甚至在整个系统中起到决定性的作用。多个模块单独校准和互相校准,最优化在线调整系统参数以及多传感器集成化等方面仍是研究者们未来需要解决的问题。

2) 多传感器数据关联。多种传感器融合时,数据处理非常耗时,而且各个传感器有些观测相互耦合,有些信息有一定冗余,与期望的实现轻量级、快速响的 SLAM 系统相矛盾。目前,基于优化的 SLAM 方法已经证明比传统的基于滤波器的方法更强大,但是由于鲁棒性较好的 SLAM 算法很难部署在嵌入式设备上,从机器人应用领域的实用性出发,视觉惯性里程计是一个重要的方法,其中对多传感器数据的融合有相当一部分是基于滤波器的方案。基于滤波器的视觉惯性里程计要么为其开发更强大的自适应模型,要么建立更快的干扰模型或重新配置这些滤波器中的参数,在多个传感器之间的有效耦合、在精度、鲁棒性和有限资源之间取得平衡,以实现更可靠的数据关联是一个重要的研究方向。

3) SLAM 的鲁棒性。如视觉 SLAM 在光照变化大,弱纹理、低特征条件下跟踪丢失;激光 SLAM 在高速场景以及自旋下的运动畸变。未来的 SLAM 系统可能会朝着泛化性、鲁棒性发展,增加失效保护机制,利用自动调参的能力调整系统参数来适应不同的场景,使其能在任意环境下以低失效率保证长期运行。

4) 多机 SLAM。在某些复杂的大场景环境中,单机 SLAM 存在重定位能力差、易漂移等诸多问题。随着云端技术的快速发展,为提升复杂环境中 SLAM 技术的速度与精度,局部异常不会导致整个系统的瘫痪,多机器人协同 SLAM 方案成为一个新的研究方向。目前,多机 SLAM 主要有两类:一类为中心化结构,从机在小块区域进行局部建图,并将地图信息实时传输给主机进行数据关联、优化等处理;另一类为去中心化结构,多个机器人同步局部建图,同时多机间进行局部数据交换,得到全局一致性地图以及高精度定位信息。多机器人协作的问题主要是在通信、计算复杂度、初始相对位姿估计、传感器异构、地图融合等方面。解决数据在多机间的通信、共享、分发与处理问题是多机 SLAM 领域未来主要的研究方向。

5) 在线和长期学习。深度学习思想在诸多领域发挥出较好的作用,不少研究者试图利用深度学习中云和端紧耦合的思想重新对 SLAM 技术进行整合以求得较好的效果。目前有研究者将视觉 SLAM 技术中某些环节用深度学习的方法替代,不过在现实场景中并没有体现出其绝对优势,通常会遇到新的物体和场景。比如深度网络遇到的一个重大挑战就是在极少新的类别的训练样本来支持一个动态的 SLAM 系统进行长期的学习。随着资源优化以及深度学习的不断发展,SLAM 技术的一些环节会整体被深度学习的某些方法所取代,精度和鲁棒性显著提升的 SLAM 框架成为目前移动机器人领域研究的一个热点。

## 6 结束语

近年来 SLAM 技术的研究有了较快的发展,其创建地图的好坏对移动机器人之后的自主定位、路径规划以及避障的性能起到一个关键性的作用。本文介绍了 SLAM 技术的基本研究现状,对典型开源的 SLAM 框架进行分析,讨论了 SLAM 技术主要发展

趋势以及存在的挑战, 为了能够满足机器人实际应用中稳定的工作要求, 未来仍然需要进行算法的改进, 提出更加鲁棒性的 SLAM 系统。

#### 参考文献:

- [1] BROSSARD M, BONNABEL S, BARRAU A. Unscented Kalman filter on lie groups for visual inertial odometry [C] //2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018: 649 - 655.
- [2] 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.
- [3] LEE H, CHUN J, JEON K, et al. Efficient ekf—slam algorithm based on measurement clustering and real data simulations [C] //2018 IEEE 88th Vehicular Technology Conference (VTC—Fall), IEEE, 2018: 1 - 5.
- [4] JO H G, CHO H M, JO S, et al. Efficient grid—based rao—blackwellized particle filter slam with interparticle map sharing [J]. IEEE/ASME Transactions on Mechatronics, 2018, 23 (2): 714 - 724.
- [5] TIAN Y, SUWOYO H, WANG W, et al. An AEKF—SLAM algorithm with recursive noise statistic based on MLE and EM [J]. Journal of Intelligent & Robotic Systems, 2020, 97 (2): 339 - 355.
- [6] HUANG S, DISSANAYAKE G. Convergence and consistency analysis for extended Kalman filter based SLAM [J]. IEEE Transactions on Robotics, 2007, 23 (5): 1036 - 1049.
- [7] FAN W, ZHENG Z, ZENG W, et al. Robotically surgical vessel localization using robust hybrid video motion magnification [J]. IEEE Robotics and Automation Letters, 2021 (99): 1.
- [8] 谢 嵩. 基于单目 DSO 的地下空间三维点云重建 [D]. 成都: 电子科技大学, 2018.
- [9] 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.
- [10] SOLA J, VIDAL—CALLEJA T, CIVERA J, et al. Impact of landmark parametrization on monocular EKF—SLAM with points and lines [J]. International Journal of Computer Vision, 2012, 97 (3): 339 - 368.
- [11] BOWMAN S L, ATANASOV N, DANILIDIS K, et al. Probabilistic data association for semantic slam [C] //2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017: 1722 - 1729.
- [12] TANG M, CHEN Z, YIN F. Robot tracking in SLAM with Masreliez—Martin unscented Kalman filter [J]. International Journal of Control, Automation and Systems, 2020, 18: 2315 - 2325.
- [13] LIN M, YANG C, LI D. An improved transformed unscented FastSLAM with adaptive genetic resampling [J]. IEEE Transactions on Industrial Electronics, 2018, 66 (5): 3583 - 3594.
- [14] SU Z, ZHOU J, DAI J, et al. Optimization design and experimental study of gmapping algorithm [C] //2020 Chinese Control and Decision Conference (CCDC), 2020: 4894 - 4898.
- [15] GAO H, ZHANG X, WEN J, et al. Autonomous indoor exploration via polygon map construction and graph—based SLAM using directional endpoint features [J]. IEEE Transactions on Automation Science and Engineering, 2018, 16 (4): 1531 - 1542.
- [16] NAGLA S. 2D Hector SLAM of indoor mobile robot using 2D lidar [C] //2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), 2020: 1 - 4.
- [17] GREBNER D L, BETTINGER P, SIRY J P, et al. Forestry and natural resource management careers [J]. Introduction to Forestry and Natural Resource (2 Edition), 2022: 429 - 444.
- [18] 崔 志, 石秀敏, 李 勇. Cartographer 2D SLAM 算法室内建图分析 [J]. 电子世界, 2020 (11): 27 - 29.
- [19] WANG T, LV G, WANG S, et al. SIFT based monocular SLAM with GPU accelerated [C] //International Conference on Communications and Networking in China, Springer, Cham, 2017: 13 - 22.
- [20] FU Q, YU H, WANG X, et al. FastORB—SLAM: a fast ORB—SLAM method with coarse—to—fine descriptor independent keypoint matching [D]. arXiv preprint arXiv: 2008. 09870, 2020.
- [21] CHO H, KIM E K, KIM S. Indoor SLAM application

- using geometric and ICP matching methods based on line features [J]. *Robotics and Autonomous Systems*, 2018, 100: 206–224.
- [22] GONULTAS F, ATIK M E, DURAN Z. Extraction of roof planes from different point clouds using RANSAC algorithm [J]. *International Journal of Environment and Geoinformatics*, 2020, 7 (2): 165–171.
- [23] XU Z, CHEN S, BAI J, et al. A localization system for autonomous driving: Global and local location matching based on mono-SLAM [R]. *SAE Technical Paper*, 2018.
- [24] PIRE T, CORTI J, GRINBLAT G. Online object detection and localization on stereo visual SLAM system [J]. *Journal of Intelligent & Robotic Systems*, 2020, 98 (2): 377–386.
- [25] LIU R, YANG J, CHEN Y, et al. eslam: An energy-efficient accelerator for real-time orb-slam on fpga platform [C] // *Proceedings of the 56th Annual Design Automation Conference*, 2019: 1–6.
- [26] LIU H, CHEN M, ZHANG G, et al. Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 1974–1982.
- [27] CAMPOS C, ELVIRA R, RODRIGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM [D]. *arXiv preprint arXiv: 2007. 11898*, 2020.
- [28] SERVIERES M, RENAUDIN V, DUPUIS A, et al. Visual and visual-inertial SLAM: state of the art, classification, and experimental benchmarking [J]. *Journal of Sensors*, 2021.
- [29] ENDO Y, SATO K, YAMASHITA A, et al. Indoor positioning and obstacle detection for visually impaired navigation system based on LSD-SLAM [C] // *2017 International Conference on Biometrics and Kansei Engineering (ICBAKE)*, IEEE, 2017: 158–162.
- [30] WANG S, YUE J, DONG Y, et al. A synthetic dataset for Visual SLAM evaluation [J]. *Robotics and Autonomous Systems*, 2020, 124: 103336.
- [31] WANG R, SCHWORER M, CREMERS D. Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras [C] // *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 3903–3911.
- [32] LI S, ZHANG T, GAO X, et al. Semi-direct monocular visual and visual-inertial SLAM with loop closure detection [J]. *Robotics and Autonomous Systems*, 2019, 112: 201–210.
- [33] LI J, PEI L, ZOU D, et al. Attention-slam: a visual monocular SLAM learning from human gaze [J]. *IEEE Sensors Journal*, 2021, 21 (5): 6408–6420.
- [34] SINGH R, NAGLA K S. Development of an efficient laser grid mapping technique: P-SLAM [J]. *International Journal of Image and Data Fusion*, 2019, 10 (3): 177–198.
- [35] ZOU Q, SUN Q, CHEN L, et al. A comparative analysis of LiDAR SLAM-based indoor navigation for autonomous vehicles [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021 (99): 1–15.
- [36] JIAN R, SU W, LI R, et al. A semantic segmentation based lidar slam system towards dynamic environments [C] // *International Conference on Intelligent Robotics and Applications*. Springer, Cham, 2019: 582–590.
- [37] SHAO W, VIJAYARANGAN S, LI C, et al. Stereo visual inertial lidar simultaneous localization and mapping [J]. *arXiv preprint arXiv: 1902. 10741*, 2019.
- [38] GENEVA P, MALEY J, HUANG G. An efficient schmidt-ekf for 3D visual-inertial SLAM [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 12105–12115.
- [39] MEI Y, WANG Z, CAI B. An IMU-aided visual odometry with loop-closure optimization [C] // *2018 IEEE International Conference on Information and Automation (ICIA)*, IEEE, 2018: 1444–1449.
- [40] ZHU J, YAO C, JANSCHKE K. Stereo visual-inertial fusion for UAV state estimation [J]. *IFAC-Papers OnLine*, 2020, 53 (2): 9420–9425.
- [41] EVANGELIDIS G, MICUSIK B. Revisiting visual-inertial structure-from-motion for odometry and SLAM initialization [J]. *IEEE Robotics and Automation Letters*, 2021 (99): 1.
- [42] ZHAO X, WANG C, ANG M H. Real-time visual-inertial localization using semantic segmentation towards dynamic environments [J]. *IEEE Access*, 2020 (99): 1.

(下转第 37 页)