

前言

无人机穿越光照突变的隧道、自动驾驶车辆驶入纹理稀疏的雨夜通路，AR 设备在反光玻璃环绕的空间内进行定位，这些复杂场景正在成为智能时代机器感知系统面临的常态挑战。视觉 SLAM 作为赋予机器“时空认知”能力的关键技术，在传统基于特征点的匹配算法频繁失效的情况下，亟须构建鲁棒性更强的感知范式。本书聚焦于长期被低估但具备显著结构稳定性的边缘特征，力求在高动态、弱纹理环境中探索更可靠的感知路径。

边缘特征天然具备抗光照变化与模糊干扰的优势，但其冗余性、数据关联错误与视觉约束退化问题，成为边缘 SLAM 工程化过程中的三大障碍。本书以“实时性、精确性、鲁棒性”为目标导向，系统提出以下三项关键技术以逐一破解上述难题。

一是提升边缘处理效率。提出基于子模态分区优化的边缘特征优选算法，构建结合运动不确定性的稀疏建图机制，并设计具有 0.5 逼近率的随机分区贪婪算法，实现 CPU 平台下 135FPS 的实时性能。

二是增强抗干扰能力。创新性地引入自适应鲁棒估计框架，将核函数选择与位姿优化耦合建模，并通过 Black-Rangarajan 对偶定理实现外点动态抑制，显著提升系统在高干扰环境下的稳定性与精度。

三是重塑系统稳定性。首次将条件数度量引入信息矩阵优化，结合惯导先验进行约束调控，从根本上缓解视觉退化问题，提升系统在极端场景下的可用性。

本书所提出的方法已在无人机导航、机器人感知等多场景中完成验证，所形成的“特征优选—鲁棒估计—约束增强”技术体系，不仅构筑了边缘 SLAM 的理论框架，也为其实用化进程提供了工程支撑。书中配套详细的数学推导、算法流程图、实验评估，可为研究者深入探索提供路径，也可为工程开发者解决复杂场景下的定位难题提供工具。

谨以此书献给所有致力于机器感知研究的科研与工程人员。边缘是结构的映射，是场景的骨架。相信那些潜藏于图像轮廓中的几何线索，终将成为智能系统理解世界的新语言。

本书的研究成果得到国家自然科学基金委员会、重庆市科学技术局、之江实验室等的资助。本书的出版也得到多位前辈和同行专家的指导、支持和鼓励，在此表示衷心的感谢。

尽管本书内容历经多次讨论、修改，但是由于作者水平有限，书中难免存在不足之处，敬请广大读者批评指正。

古富强

2025 年 12 月 1 日

目 录

前言

第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	3
1.2.1 传统的视觉 SLAM	4
1.2.2 视觉惯导融合的 SLAM	7
1.2.3 基于边缘特征的视觉 SLAM	10
1.3 存在的研究问题与挑战	12
1.3.1 边缘特征冗余	13
1.3.2 数据关联错误	14
1.3.3 视觉约束退化	15
第 2 章 基于边缘特征的位姿估计理论基础	17
2.1 基本术语与坐标系	17
2.1.1 传感器坐标系	17
2.1.2 位姿矩阵表示	18
2.1.3 相机成像模型	20
2.2 边缘特征提取	21
2.2.1 微分算子法	21
2.2.2 最优边缘检测算法	23
2.2.3 深度学习法	24
2.3 边缘特征对齐	25
2.3.1 基于光度误差的对齐	25
2.3.2 基于几何误差的对齐	26
2.3.3 距离变换	27
2.4 本章小结	29
第 3 章 基于边缘特征优选的视觉里程计	30
3.1 问题分析	30
3.2 基于子模态分区优化的边缘特征优选算法	31
3.2.1 边缘特征优选问题构建	32
3.2.2 子模态函数分区优化	33

3.2.3 重观测概率建模·····	35
3.3 基于边缘特征的视觉里程计设计·····	36
3.3.1 相机跟踪·····	37
3.3.2 局部优化·····	39
3.3.3 边缘特征选择·····	41
3.4 实验结果与分析·····	42
3.4.1 实验设置·····	42
3.4.2 系统对比评估·····	42
3.4.3 算法消融实验·····	47
3.5 本章小结·····	50
第 4 章 基于自适应鲁棒估计的视觉 SLAM ·····	52
4.1 问题分析·····	52
4.2 基于广义核函数的自适应鲁棒估计算法·····	53
4.2.1 广义鲁棒核函数·····	54
4.2.2 Black-Rangarajan 对偶上界·····	55
4.2.3 松弛优化及收敛性证明·····	59
4.3 基于边缘特征的视觉 SLAM 设计·····	62
4.3.1 视觉里程计·····	62
4.3.2 闭环修正·····	64
4.4 实验结果与分析·····	65
4.4.1 实验设置·····	66
4.4.2 系统对比评估·····	66
4.4.3 算法消融实验·····	73
4.5 本章小结·····	75
第 5 章 基于信息矩阵条件数优化的视惯融合 SLAM ·····	77
5.1 问题分析·····	77
5.2 基于 ω -条件数度量的信息矩阵优化算法·····	79
5.2.1 ω -条件数度量·····	79
5.2.2 视惯融合模型·····	80
5.2.3 信息矩阵优化·····	82
5.3 基于边缘特征的视惯融合 SLAM 设计·····	86
5.3.1 边缘追踪·····	86
5.3.2 MSCKF 优化·····	89
5.3.3 信息矩阵条件数优化·····	93
5.4 实验结果与分析·····	94

5.4.1 实验设置	94
5.4.2 系统对比评估	94
5.4.3 算法消融实验	98
5.5 本章小结	100
第 6 章 结论与展望	101
6.1 研究工作总结	101
6.2 未来工作展望	102
参考文献	104

第1章 绪 论

1.1 研究背景及意义

位置信息在智能机器人、物联网、智慧城市等领域发挥着重要作用，为智能机器人的自主导航和任务执行、物联网的设备连接和管理、智慧城市的智能化服务和资源优化等提供了基础信息支持^[1,2]。随着社会经济技术的发展，不同领域和应用对定位的要求逐渐提升，迫切需要研究可靠准确的定位系统，为各种应用场景提供准确的位置信息，支持各种关键任务和决策，推动无人系统智能化、自主化和高效化的发展。

通常，移动设备能够使用多种传感器进行定位，例如，通过全球导航卫星系统（global navigation satellite system, GNSS）可以获得较为准确、可靠的目标位置信息^[3,4]，然而该方式严重依赖于卫星信号，在卫星信号拒止环境下则存在定位不准确甚至无法定位等问题^[5]，如城市峡谷或室内环境等；又如，基于惯性导航系统的定位方式，根据加速计和陀螺仪的测量数据，通过积分来估计朝向和位置，该方式不限制运动场景，能够独立工作，但是累积误差十分严重，定位精度较差^[6]；除此之外，基于蓝牙、Wi-Fi、超宽带（ultra wide band, UWB）、射频识别（radio frequency identification, RFID）及地磁等的定位技术也经常被采用，但是这些方法存在部署成本高或者定位精度低的问题^[7]。因此人们需要一套不依赖外源信号，而且可以在未知区域进行定位操作的定位系统。为此，同时定位和建图（simultaneous localization and mapping, SLAM）应运而生，并在机器人、无人机、自动驾驶及增强现实等领域展现出巨大的应用潜力^[8,9]，如图 1.1 所示。

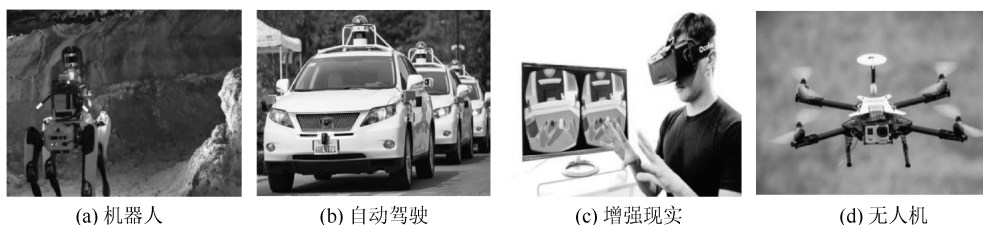


图 1.1 SLAM 应用举例

SLAM 是在运动过程中通过对物理环境的观测实现位姿和地图的同步估计^[10]。该项技术能够在未知环境下仅依赖机器自身的传感器进行稳定工作，属于测绘、计算机、人工智能等领域的交叉方向，是当前主流的定位方法之一。按照传感器观测特性的差异，SLAM 技术通常可划分为激光雷达 SLAM^[11-13]和视觉 SLAM^[14-16]。其中，激光雷达 SLAM 通过激光雷达扫描获取三维点云数据，利用点云配准的方式实现载体的位姿估计和地图构建。视觉 SLAM 通过相机获取环境的图像信息，根据特征关联进行位姿估计，并采用

三角测量的方法构建地图。相比于激光雷达，相机具有低成本和测量信息丰富的优势，这使得视觉 SLAM 能够具有更强的感知能力，灵活地适应不同场景。因此，视觉 SLAM 具有重要的应用和研究价值。

视觉 SLAM 主要包括视觉里程计（visual odometry, VO）和全局优化两个核心模块。其理论研究历史可追溯到 20 世纪 80 年代^[10]，最初的研究工作集中在单目相机的视觉里程计方面，通过分析连续图像序列来估计相机的运动轨迹。随着计算机视觉技术与计算能力的发展，研究者开始探索如何结合图像处理和三维重建等技术来实现视觉 SLAM，自 2000 年以来涌现了一系列优秀的视觉 SLAM 方法，这些方法根据数据关联模式的不同可以划分为：特征法，如 MonoSLAM^[14]、PTAM（parallel tracking and mapping，并行跟踪与建图）^[15]、ORB-SLAM^[16]等；直接法，如 DTAM^[17]、LSD-SLAM^[18]、DSO^[19]等。如图 1.2 所示，特征法通过关键点的提取和基于描述子的匹配，实现相机的位姿估计；而直接法将数据关联和位姿估计两个任务耦合，采用光度对齐的方式实现优化计算。

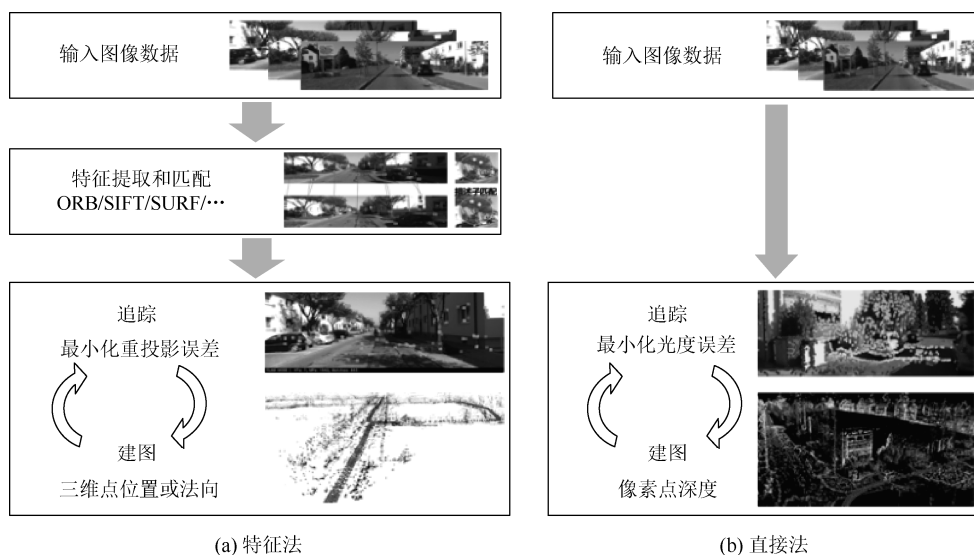


图 1.2 传统视觉 SLAM 方法框架图

尽管传统的视觉 SLAM 方法已经取得较为成熟的结果，但是在复杂观测环境下其精度、效率及鲁棒性方面仍然面临巨大挑战^[20-22]。一方面，特征法中数据关联的计算十分耗时，而直接法由于图像的强非凸性需要多次迭代对齐以保障计算收敛，整个优化过程同样耗费大量计算时间。另一方面，在墙面、地板等弱纹理环境中，特征法难以提取到足够的关键点，或者提取到的关键点的分布较为集中，并且特征描述子的可区分度较差；在光照变化场景下，物体表面的光度值在不同观测时刻呈现较大的差异性，直接法所依赖的光度一致性往往无效。这些因素使得传统视觉 SLAM 难以在复杂观测场景下建立充分且可靠的数据关联关系，导致位姿估计精度下降，甚至出现相机追踪失败的情况^[23]。因此，提高视觉 SLAM 对复杂场景的适应性是当前研究领域的重点。

相较于传统的视觉 SLAM 方法,基于边缘特征的视觉 SLAM 方法在图像信息利用和误差度量两个层面上具有显著优势,其在弱纹理、光照变化等复杂场景中具备更强的稳健性^[24]。在图像信息利用层面,基于边缘特征的视觉 SLAM 方法不仅能提取到传统的关键点和直线段特征,还能够捕捉更加完整的物体轮廓和场景结构信息。这种特征表达方式超越了简单的几何元素,使系统在复杂环境中能够获取更丰富的视觉信息,从而提升了 SLAM 系统对环境的全面感知能力与鲁棒性。在误差度量层面,基于边缘特征的视觉 SLAM 方法采用了基于几何误差的边缘对齐策略,既避免了繁杂的特征匹配过程,同时缓解了对纹理和光度的依赖性,具有较强的优化收敛能力。这些特性使得基于边缘特征的视觉 SLAM 方法可以在复杂观测环境下更加稳定地工作,减少由纹理缺失或光照变化导致的定位漂移。然而,在实际应用过程中,海量的边缘特征会给视觉 SLAM 带来较重的计算负担,这导致系统的实时性较差;数据关联错误较多导致位姿估计面临严重的外点干扰;视觉约束容易产生退化现象,难以保障优化过程的稳定性。在这些因素的共同作用下,基于边缘特征的视觉 SLAM 方法应用和发展受到了严重的制约。因此,如何发挥边缘特征在复杂场景中的潜力,提高视觉 SLAM 系统的实时性、精确性和鲁棒性是一个亟待解决的问题。

本书以基于边缘特征的视觉 SLAM 方法为主要研究对象,以提高系统实时性、精确性和鲁棒性为主要研究目标。本书针对边缘特征冗余、数据关联错误及视觉约束退化问题展开研究,提出了边缘特征优选、自适应鲁棒估计及观测矩阵条件数优化的方法以克服上述问题的制约。

1.2 国内外研究现状

视觉 SLAM 问题的研究主要经历了三个重要阶段^[10, 21]。2006 年以前被称为经典时期,主要聚焦于状态估计理论模型的构建。Smith 等首次提出了 SLAM 问题并对其进行概率建模^[25],将路标点和机器人位置作为状态变量,根据机器人观测的不确定性,采用扩展卡尔曼滤波^[26, 27](extended Kalman filter, EKF)的方法实现状态估计。在随后的近 20 年时间内,基于概率学的建模方法成为了视觉 SLAM 的研究主流^[28, 29],相关学者针对其中的数据关联、状态维度等问题进行了进一步的研究和发展^[30-32],其中诞生了以 MonoSLAM 为代表的工作^[14]。该系统利用稀疏特征点的追踪降低了系统的计算量,实现了在 CPU 上实时无尺度的单目位姿估计。

2006~2016 年被称为算法分析时期,在这一时期,相关学者深入分析了状态估计中的可观性、收敛性和一致性,并且将非线性优化理论引入状态估计中,由此涌现出一系列的优化框架,包括 Ceres^[33]、G2o^[34]及 GTSAM^[35]等。此外,在该时期基本形成了视觉 SLAM 前后端分离的计算架构,前端即视觉里程计,主要负责相机的追踪和局部的优化,后端结合回环检测进行相机轨迹全局一致性的优化,通过前后端的并行计算,大幅提高了视觉 SLAM 的实时性。PTAM^[15]是该时期的代表工作,该系统也是首个基于非线性优化的实时视觉 SLAM 系统。

自 2016 年至今被称为鲁棒感知时期，主要关注算法在复杂场景中的适应性，以及计算资源受限情况下算法的实时性等问题。研究者进行了多方面的创新，首先在多传感器融合方面，通过集成相机、惯性测量单元（inertial measurement unit, IMU）及激光雷达等设备，显著增强了系统在复杂环境下的稳定性^[36]。其次，在多种感知范式的结合上，从单一类型的特征到点、线、面及物体的多特征融合^[37]，从单一的几何信息测量到高层次语义信息的结合^[38]，大幅提高了视觉 SLAM 的精度和环境感知能力。此外，为了应对 SLAM 的高计算需求，研究者正在探索利用多种计算设备和架构，包括图形处理单元（graphics processing unit, GPU）加速^[39]、云边端协同计算^[40]及现场可编程门阵列（field programmable gate array, FPGA）硬件加速^[41]等，不仅提高了 SLAM 系统的整体性能和效率，还使得在资源受限的移动平台上部署复杂 SLAM 算法成为可能。

总之，视觉 SLAM 领域的研究如火如荼，本章主要对其中所涉及的部分内容进行详细介绍，主要包括传统的视觉 SLAM、视觉惯导融合的 SLAM，以及基于边缘特征的视觉 SLAM。

1.2.1 传统的视觉 SLAM

传统的视觉 SLAM 的关键在于数据关联，按照关联的模式不同，可将其分为特征法和直接法，具体分类如图 1.3 所示。

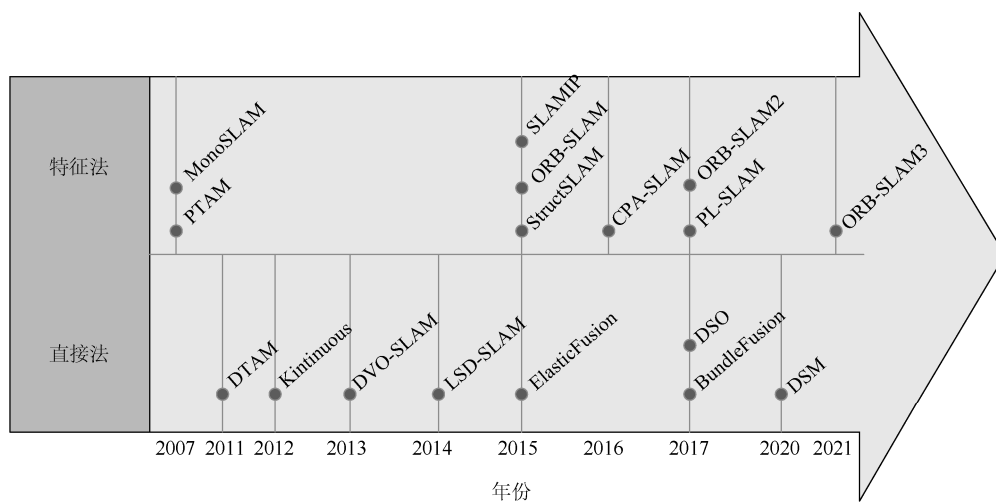


图 1.3 传统的视觉 SLAM 方法分类

1. 特征法

特征法的核心思想是提取每一帧图像中的显著特征，通过使用描述子或者光流进行匹配，然后基于匹配关系建立重投影误差，实现位姿估计和地图构建^[42]。MonoSLAM^[14]是特征法的开创性工作，首次验证了基于单目视觉的实时 SLAM 系统的可行性。该方法

将空间点作为环境表征,通过连续帧间的特征匹配来估计深度信息。虽然该方法在小规模环境中表现出色,但计算复杂度随地图规模呈平方增长,限制了其在大规模场景中的应用。这一局限性促使研究者探索更高效的算法结构,为后续的技术突破奠定了基础。随后,Klein 等提出的 PTAM^[15]算法标志着特征法 SLAM 进入了新的发展阶段。PTAM 的创新之处在于将 SLAM 问题分解为并行的跟踪和建图两个过程,这种结构设计不仅提高了系统的实时性能,还为处理大规模场景问题提供了可能。更重要的是,PTAM 首次引入了基于非线性优化的后端处理方法,即光束法平差^[43] (bundle adjustment, BA),大幅提高了系统的精度和一致性。这种前端跟踪与后端优化相分离的架构为后续的 SLAM 系统设计提供了范式。

在 PTAM 奠定的基础上,ORB-SLAM^[16]的出现是特征法 SLAM 的又一重要里程碑。该系统进一步细化了 SLAM 的功能模块,将其划分为跟踪、局部优化和全局优化三个并行线程。这种设计不仅提高了系统的模块化程度,也使得各个功能模块能够更加高效地利用计算资源。ORB-SLAM 的另一个重要贡献是引入了基于词袋模型^[44]的回环检测机制,使得系统具备在长时间、大尺度环境中保持全局一致性的能力。如图 1.4 (a) 所示,ORB-SLAM2^[45]支持单目、双目和 RGB-D 相机,直接获取特征点深度信息,解决了单目系统的尺度不确定性问题。ORB-SLAM3^[46]则引入了多地图策略和 Atlas 数据结构^[47],有效处理了视觉 SLAM 中短期、中期及长期的数据关联问题,使得系统能够应对相机快速运动、遮挡和临时跟踪丢失等挑战性场景,同时,整合了视觉-惯性里程计 (visual inertial odometry, VIO),以提高系统在快速运动和弱纹理环境中的鲁棒性。

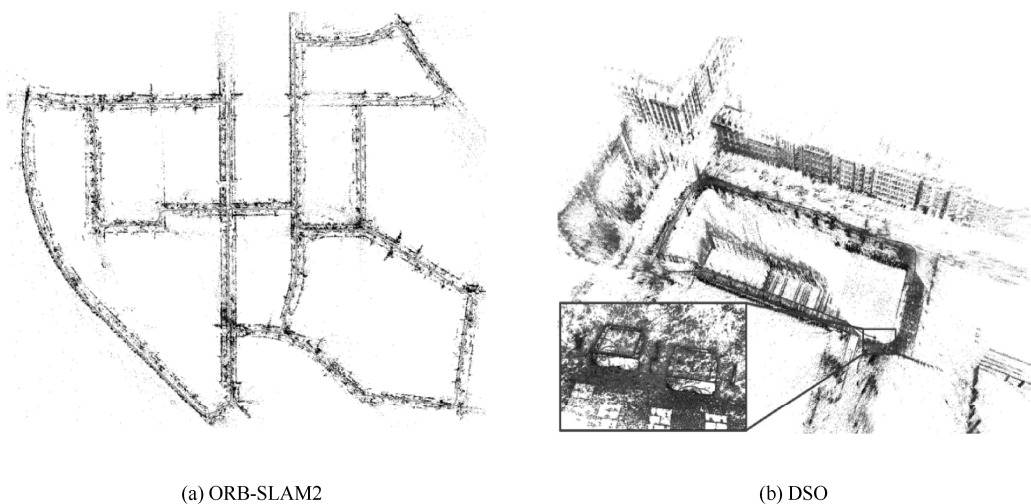


图 1.4 传统的视觉 SLAM 运行效果图

上述方法要求特征点数量充足、空间分布均匀和匹配关系可靠以保障准确和鲁棒的姿态估计^[22],这些条件在弱纹理环境中很难实现,因此相关学者探索利用多种特征来构建视觉 SLAM 系统^[48]。StructSLAM (结构化线特征 SLAM)^[49]以建筑结构线作为主要特征并在连续帧之间进行匹配。数据关联的过程是基于线段的几何位置和方向,利用线段

之间的拓扑关系,确保在不同视角下的特征匹配一致性。PL-SLAM^[50]利用 LSD^[51]算法检测空间中的结构线,通过基于描述子的点线特征匹配构建位姿优化约束,采用非线性优化算法联合估计位姿,并在全局地图中分别维护点和线特征。Kaess^[52]将场景中的结构表示为无限延伸的平面,并利用平面间的几何约束条件来优化相机的位姿估计,而 CPA-SLAM^[53]则通过平面模型的对齐优化位姿参数。此外,Proença 等^[54]和 Shu 等^[55]则利用点、线、面等多种几何特征进行数据关联来估计相机位姿和更新地图。

特征法通过提取图像中显著的特征,能够在不同光照条件下保持稳定的特征描述,实现可靠的帧间匹配。在处理快速运动问题时,该方法能够通过跟踪高分度度的特征点,有效估计相机的大幅运动。然而,在纹理匮乏的场景下,如白墙、均匀地面等,系统难以提取足够数量的稳健特征点,这导致特征匹配稀疏或不可靠。在这种情况下,位姿估计的精度会显著下降,甚至可能导致跟踪完全失败。虽然多特征融合(如结合点、线、面特征)能在一定程度上缓解该问题,但也增加了算法的复杂度和计算负担。

2. 直接法

直接法是直接比较图像像素的强度值,即通过最小化像素之间的光度误差来估计相机位姿及地图点位置^[56]。相较于特征法而言,直接法不需要耗时的特征提取和匹配,同时能够利用图像所有信息以缓解弱纹理环境下视觉约束不足的问题。DTAM^[17]是首个直接法 SLAM 系统,其利用图像中所有像素点,在选定的关键帧中通过变分优化方法求解深度图,引入正则化项处理弱纹理区域和遮挡问题。在跟踪阶段,DTAM 通过最小化当前帧与重建稠密模型之间的光度误差来估计相机位姿。这种全像素处理方法虽然信息利用充分,但计算量大,通常需要 GPU 加速才能实现实时性能。为了降低计算复杂度,LSD-SLAM^[18]仅利用图像中高梯度区域的像素点,采用逆向组合的策略最小化光度误差,实现图像的直接配准,这种方式大幅提高了前端的追踪效率。此外,深度估计通过滤波方法不断进行更新,同时使用姿态图优化以保持全局一致性。LSD-SLAM 还引入了尺度感知的相似变换优化,有效处理了单目 SLAM 中的尺度漂移问题。

为了解决 DTAM 与 LSD-SLAM 计算复杂度高的问题,Engel 等提出了 DSO^[19]算法,采用稀疏的策略在计算效率和精度之间寻求平衡。如图 1.4 (b) 所示,该方案在图像梯度较大的区域仅选择部分像素点进行计算,通过由粗到精的最小化光度误差实现高效的相机位姿追踪。DSO 采用滑动窗口的框架作为局部优化,对窗口内的相机内参、外参、逆深度和光度参数进行联合优化。此外,DSO 引入了光度校正因子来处理曝光变化,提高了系统在不同光照条件下的鲁棒性。DSO 算法的提出意味着直接法进入了一个新的发展阶段,可以实时有效地处理大规模的 SLAM 问题。为了缓解临时地图的局限性,DSM^[57]算法引入了一个局部地图共视窗口,将地图点的重观测纳入优化过程,通过时间和共视性标准选择关键帧,确保在时间上相距较远的关键帧也能包含在优化过程中。该算法通过持续重用地图信息保持系统的一致性,相较于之前的直接法在轨迹和地图重建的精度上有了显著提升。

为了进一步提高位姿估计和场景重建的精度和鲁棒性,相关学者结合 RGB-D 相机来构建 SLAM 系统,例如,DVO-SLAM^[58]和 Kintinuous^[59]等方法将光度对齐优化和迭代最

近点^[60] (iterative closest point, ICP) 优化结合起来, 综合利用相机观测的图像信息和深度信息实现位姿估计。在此基础上, ElasticFusion^[61]引入了基于曲面元素的环境表示方法, 每个面元 (surfel) 包含位置、法线、颜色和置信度等信息, 采用局部和全局两种图像配准策略, 其中在全局配准中利用非刚性形变提高数据关联的准确性, 由此实现高精度的相机位姿估计及三维地图构建。BundleFusion^[62]额外添加了特征匹配模块, 结合多种数据关联机制, 采用分层优化策略, 先进行粗配准, 再进行精配准, 以提高位姿估计的精度。此外, 该方法采用体素哈希技术, 提高了关键帧及地图检索的效率。

直接法的核心优势在于其对图像信息的充分利用, 不同于特征法仅依赖稀疏特征点, 直接法利用图像中的全部或大部分像素信息, 在弱纹理环境中仍能有效工作, 展现出较强的适应性和鲁棒性。直接法避开了耗时的特征提取和匹配步骤, 可以实现亚像素级别的数据关联, 这在高精度定位任务中具有显著优势。然而, 直接法需要建立在光度一致性的假设上, 在光照变化场景下, 该假设通常难以成立, 这导致定位精度较低。由于图像的强非凸性和非线性, 直接法收敛困难, 通常需要较好的初始位姿估计, 否则可能陷入局部最优解。此外, 直接法需要处理大量的像素点以保障位姿估计的精度和鲁棒性, 这大幅增加了系统的计算负担。

1.2.2 视觉惯导融合的 SLAM

视觉惯导融合的 SLAM 系统通过融合视觉传感器和惯性测量单元 (IMU) 的数据, 有效克服了单一传感器方法固有的局限性^[63]。视觉信息能够反映环境的结构和外观特征, 但是在弱纹理及光照变化情况下存在观测信息少、外观不一致的问题, IMU 数据则提供了高频率的运动估计和重力方向信息, 但是存在长周期运行时累积误差较大的问题。融合两种传感器模态在测量方面存在一定的互补性, 当视觉系统因弱纹理或光照变化而失效时, IMU 可以提供短期的稳定运动估计, 维持系统的连续性, 而 IMU 信息中的累积误差可以通过视觉特征跟踪或回环检测进行抑制, 确保长期定位精度。因此, 有效融合视觉和惯导信息可以增强系统在复杂环境下的鲁棒性。

根据传感器耦合模式的不同可以将视觉惯导融合的 SLAM 划分为松耦合^[64, 65]和紧耦合^[66]两种方式。其中, 松耦合方式将位姿估计划分为两个独立的子系统, 视觉系统通过提取图像中的特征点并进行匹配和跟踪, 从而估计相机的运动轨迹; 惯导系统则根据 IMU 测量的加速度和角速度数据, 通过连续积分得到平台的相对位姿变化。针对两者提供的运动估计结果, 采用卡尔曼滤波器或非线性优化方法进行融合。该类型的代表性方法为 Weiss 等^[64, 67]提出的相关工作, 他们将 PTAM 作为独立的视觉子系统进行位姿估计输出, 然后利用 EKF 对 IMU 的积分位姿进行矫正, 该方法在无人机导航中进行了验证, 取得了良好的结果。松耦合方式的优势在于系统设计的简洁性和模块化, 其适合处理不同频率的传感器信息, 此外, 松耦合方式在计算资源上相对较为经济, 整体计算的复杂度较低^[68]。然而, 视觉和惯导数据在早期阶段没有进行充分的交互融合, 系统可能无法充分利用二者之间的互补性, 因此在实际应用中, 松耦合方式的精度和鲁棒性通常较差。

紧耦合方式直接将视觉和惯性传感器的原始数据进行融合, 从而在复杂环境下提高

了系统的准确性和鲁棒性，因此紧耦合方式成为目前较为主流的融合方式^[69, 70]。按照计算过程的不同，紧耦合方式又可以分为基于非线性优化的方法和基于滤波优化的方法，如图 1.5 所示。

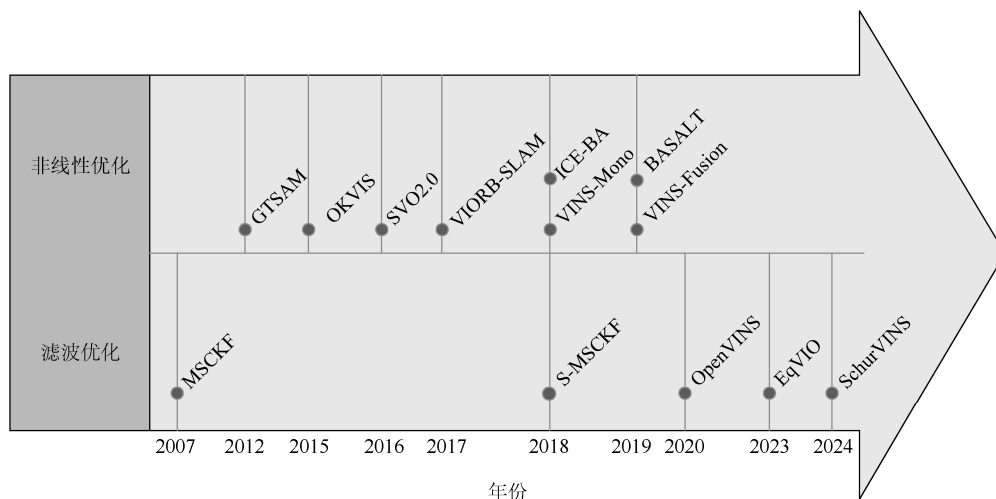


图 1.5 视觉惯导融合 SLAM 方法分类

1. 基于非线性优化的方法

基于非线性优化的方法是将传感器的融合构建为一个基于最小二乘的非线性问题，其中 IMU 测量通过预积分技术生成相邻图像帧间的运动约束，视觉信息则通过特征点的匹配或光流追踪构建重投影误差以形成观测约束，利用两类共同约束构建一个复合代价函数，通过迭代优化实现状态的估计^[71-73]。OKVIS^[74]算法采用滑动窗口优化策略，将 IMU 误差项与视觉路标投影误差集成在同一概率框架内，通过非线性最小二乘优化对窗口内的位姿和地图点进行联合估计，逐步更新系统状态，从而实现对相机运动轨迹和场景结构的精准估计。VIORB-SLAM^[75]则在 ORB-SLAM 的基础上引入了 IMU 预积分模型^[76, 77]，这使得惯导数据和视觉数据在同一优化框架内进行联合估计，大幅提高了信息融合的紧密程度。此外，该方法利用 IMU 数据解决单目 SLAM 初始化中尺度模糊等问题，同时针对重力方向及偏置进行估计，提高 IMU 数据的准确性。同样地，SVO2.0^[78]则是在 SVO^[79]的基础上集成了 IMU 的测量信息，该方法将直接法中像素的光度误差与运动约束进行有效结合，利用惯导信息来提高图像对齐过程中初始值的准确性，利用大量的视觉关联改善 IMU 的漂移，从而在复杂场景下实现了高精度的位姿估计。

为了进一步提高视觉惯导初始化的精度和鲁棒性，有学者^[80]提出了一种松耦合的初始化策略，即 VINS-Mono，其根据视觉数据采用 SFM^[81]算法构建相机的短期运动轨迹，并采用预积分的方法计算 IMU 的运动估计，通过将两种轨迹对齐，联合估计初始的尺度、重力方向和速度，从而为后续的视惯融合提供准确的初始值。该方法采用滑动窗口的融合优化方法，并通过双阶段的边缘化策略保障计算的精度和效率。此外，VINS-Mono 集

成了回环检测与优化模块，通过识别历史关键帧的回环并执行全局优化，有效抑制了累积误差，提升了长时间运行时的全局一致性和地图的准确性，如图 1.6 (a) 所示。随后，VINS-Fusion^[82]在此基础上提供了一个多传感器的统一融合框架，不仅支持单目相机和 IMU，还扩展支持双目相机、GNSS 等多种传感器。其核心特点是通过一个统一的滑动窗口优化框架，将多种传感器数据的观测误差统一建模，并进行联合优化。这一方法允许 VINS-Fusion 在复杂场景中灵活选择和组合传感器，从而在不同应用场景中都能进行精确的定位和建图。

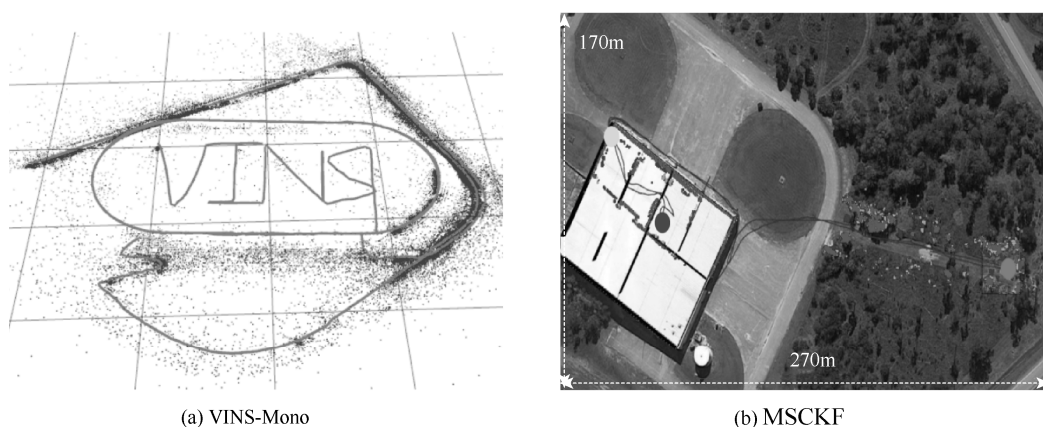


图 1.6 视惯融合算法运行示意图

为了提高后端融合优化的计算效率，相关学者对 BA 方法进行了改进。其中具有代表性的工作为 GTSAM^[35]，该方法将复杂的非线性优化问题表示为因子图，其中变量节点表示待估计的状态，因子节点表示各类约束。GTSAM 采用 iSAM2^[73]算法实现增量式求解，通过贝叶斯树数据结构高效组织因子图，支持快速的边缘化和条件概率计算。该框架的显著特征包括具有灵活的前端接口，支持多种传感器模型和约束类型的表达，以及智能线性化策略，仅在必要时重新线性化变量以提高计算效率。GTSAM 的设计使其适用于广泛的 SLAM 应用场景，能够高效处理大规模优化问题。在此基础上，ICE-BA^[83]引入一种增量式的更新机制，在每次新的传感器数据到达时，仅更新与新数据直接相关的子图，而非重新优化整个图结构，从而显著降低了计算复杂度。该方法使用了一种基于高斯-牛顿 (Gauss-Newton) 法的高效求解器，并结合了稀疏矩阵技术和先进的线性代数库，使其能够在有限的计算资源下实现近乎实时的优化。此外，ICE-BA 还引入了全局一致性的保障机制，通过在增量更新过程中保持全局地图的约束一致性，确保在长期运行中的精度和稳定性。另外，Basalt^[84]算法避免了计算过程中的显式矩阵求逆操作，直接对平方根信息矩阵进行操作，这样不仅减小了计算量，还提高了数值稳定性。平方根信息矩阵的引入使得 Basalt 算法在执行滑动窗口联合优化时具有更高的数值稳定性和计算效率，即使在状态空间维度较高的场景下，仍能保持良好的实时性能。

基于非线性优化的融合方法虽然能够获得较优的估计精度，但其迭代优化过程会显著增加系统的计算开销。尽管现有算法在计算效率方面进行了诸多改进，但是优化求解

过程中所需的多次线性化与迭代计算仍然带来了较大的计算负载。因此，在计算资源受限的嵌入式平台上，该类方法难以实现理想的实时性能。

2. 基于滤波优化的方法

滤波法通过递归方式来实现对系统状态的估计，其基本原理是将机器人位姿和环境特征建模为一个状态向量，并利用滤波器（如 EKF）对这一状态向量进行连续更新^[85]。滤波方法通过将当前传感器的观测值与前一时刻的状态预测相结合，计算出最优的状态估计。这一过程通常包括两个步骤：预测和更新。首先，利用运动模型对当前状态进行预测，然后通过观测模型结合新的观测数据，对预测的状态进行更新。相较于基于非线性优化的方法，滤波方法的优势在于其能够在线实时处理数据，同时在弱非线性的高斯噪声系统内实现较优的估计精度。其中代表性工作为 MSCKF^[86, 87]（multi-state constraint Kalman filter，多状态约束卡尔曼滤波器），如图 1.6（b）所示。该方法的核心思想是采用 EKF 框架维护一个滑动窗口优化，优化变量包含当前 IMU 以及 N 个历史帧的姿态，通过 IMU 数据对系统状态进行预测，通过相机对丢失特征点的多帧观测获得约束条件，并采用 EKF 对预测状态进行更新。该方法在计算中将特征点三维位置从状态变量中消除，有效降低了卡尔曼滤波的计算维度，因此实现了高效运行。

由于缺少对地图点位置的联合优化使得地图点位置不准，所以 MSCKF 算法的精度和一致性通常较差。S-MSCKF^[88]在此基础上扩展至双目相机，利用立体匹配的方法提高特征点深度计算的精度，既降低了尺度模糊性又增强了位姿估计的精度。OpenVINS^[89]延续了这种方法的核心思想，设计并实现了模块化的融合求解框架，以便于根据实际工程应用自由地选择和替换各个模块。此外，SchurVINS^[90]采用对地图点和位姿的联合估计，利用舒尔补技术^[91]将联合估计问题分解为独立的子问题，实现了对地图点和位姿优化的解耦合，在降低联合优化的计算复杂度的基础上，提高了状态估计的精度和鲁棒性。EqVIO^[92]算法基于等变性理论，通过将相机位姿、速度和 IMU 偏差等状态变量表示在李群空间中，采用等变误差反馈机制，在状态估计过程中动态调整权重，确保状态估计过程中的一致性和稳定性。

基于滤波优化的融合方法的核心优势在于采用递归的方式估计系统状态，能够在低延迟下提供连续的位姿估计，适用于资源受限的嵌入式系统。此外，滤波方法具有内在的稳健性，能够有效融合视觉和惯性数据，降低单一传感器的不确定性。然而，这类方法的主要缺陷在于其对系统模型和噪声统计特性的强依赖，通常需要假设系统噪声符合线性高斯模型，对高度非线性的运动场景适应性较差。此外，相比非线性优化方法，滤波方法难以充分利用全局信息进行精确的状态估计。

1.2.3 基于边缘特征的视觉 SLAM

边缘特征是指图像中物体的轮廓、边界或者纹理的显著变化区域，在这些区域，像素值发生明显的变化，反映出不同物体或区域之间的分界^[93]。这种特征在室内空间或工业场景等人造环境中广泛存在，而且较容易被相机所捕获，即使在弱纹理环境下，利用