



宁波大学
NINGBO UNIVERSITY

本科毕业设计（论文）

文献综述

题目： 面向轻量化定位与地图构建方法研究

学 院	信息科学与工程学院
专 业	计算机科学与技术
班 级	22 计算机一班
学 号	226002618
学生姓名	李杰
指导教师	彭成斌
开题日期	2025 年 12 月 25 日

面向轻量化定位与地图构建方法研究

摘 要

视觉—惯性同步定位与地图构建（Visual-Inertial Simultaneous Localization and Mapping, VI-SLAM）技术通过融合摄像头与惯性测量单元（IMU）信息，实现对运动载体位姿的高精度估计与环境建图，已广泛应用于移动机器人、无人机、自动驾驶及增强现实等领域。随着嵌入式设备和移动终端的快速发展，VI-SLAM 系统在资源受限平台上的实时性、稳定性与功耗问题逐渐成为研究热点。现有主流 VI-SLAM 系统如 VINS-Fusion [1]、ORB-SLAM3 [2] 和 OpenVINS [3] 等，在精度和鲁棒性方面取得了显著进展，但普遍存在计算复杂度高、内存占用大、对硬件性能依赖强等不足，限制了其在轻量化场景中的部署。本文围绕“面向轻量化定位与地图构建方法研究”这一主题，对 VI-SLAM 领域的研究背景、技术路线和代表性系统进行系统梳理与分析，重点综述了优化型与滤波型 VI-SLAM 方法在系统结构、算法特性及资源消耗方面的差异。在此基础上，进一步总结了近年来面向轻量化的关键研究方向，包括前端特征处理优化、后端状态管理与地图稀疏化策略等。通过文献综述与对比分析，明确了现有研究的不足与改进空间，为后续基于 OpenVINS 的系统轻量化优化与实验验证奠定理论基础。

关键字 视觉惯性 SLAM；轻量化优化；定位与地图构建；嵌入式系统；OpenVINS

1 引言与 VI-SLAM 发展概览

视觉-惯性 SLAM (VI-SLAM) 是机器人、无人机、增强/虚拟现实 (AR/VR) 等领域实现自主定位和地图构建的核心技术。它通过结合摄像头和惯性测量单元 (IMU) 数据，既克服了纯视觉系统在弱纹理或快速运动时鲁棒性不足的问题，又能够在单目情况下恢复真实尺度[1, 9]。例如，香港科技大学的研究组指出，VINS-Fusion 等视觉惯性算法能够在无人机、汽车以及 AR/VR 等应用中实现高精度定位[1]。然而，当前高精度 SLAM 系统（如 VINS-Fusion [1]、ORB-SLAM3 [2]、Kimera [6] 等）通常采用非线性优化策略，通过全局或滑动窗口的束束调整（BA）来获得最优解，因而对计算资源要求极高，难以在资源受限的平台（如 NVIDIA Jetson Nano）上实时运行[4, 5]。Jeon 等人在 Jetson 板上对多种 VIO 算法进行了基准测试，发现尽管 Jetson 平台配备了较强的 CPU 和 GPU，但现有研究很少对这些算法在计算资源消耗（CPU 占用、内存使用）和精度之间的平衡进行系统分析[4]。因此，如何在保证定位精度的前提下，优化算法的计算和内存开销，

是 VI-SLAM 领域亟待解决的关键问题。

1.1 VI-SLAM 的重要性与挑战

VI-SLAM 对移动机器人和 AR/VR 等领域至关重要。一方面,摄像头和 IMU 的融合可以提供密集丰富的位姿信息和即时的运动补偿能力,使系统在各种环境条件下保持鲁棒性。例如, VINS-Fusion 等方法通过与惯性传感器的紧密耦合,在动态和弱纹理环境中也能进行可靠定位。另一方面,传统单目视觉 SLAM 存在尺度模糊问题,而融合惯性测量后可恢复运动的真实尺度,提高定位的准确性。然而,这种跨传感器的集成也带来了新的挑战:包括对惯性测量预积分、传感器初始化、外参标定和非线性优化求解的需求,这些过程计算复杂且易陷入局部最优。此外,环境光照变化、视野遮挡、震动干扰等因素,也给视觉-惯性算法的鲁棒性带来考验。值得注意的是,在资源受限的嵌入式平台上,高精度算法的实时部署受到显著限制。Jeon 等人的研究表明,尽管 Jetson 板载有强大的处理器,但现有研究很少对这些板卡在执行视觉(-惯性)里程计算法时的资源占用进行详细评估。因此,本研究聚焦于在嵌入式平台上实现 VI-SLAM 轻量化运行的关键问题。

1.2 评估标准

为了对比不同 VI-SLAM 系统的性能,通常采用绝对轨迹误差 (Absolute Trajectory Error, ATE) 和相对位姿误差 (Relative Pose Error, RPE) 等指标进行定量评估。ATE 衡量估计轨迹与真实轨迹的全局一致性,是评价全局定位精度的常用指标;RPE 则主要反映相邻两帧之间的漂移情况,用于评价算法的局部准确性和漂移速度[3, 4]。除了精度指标外,本研究还强调资源消耗指标的评估:包括帧率 (FPS)、CPU 占用率和内存使用量等,以验证算法的轻量化效果和实时能力[4, 5]。例如,Jeon 等人在评测中明确提出,要在不同 Jetson 平台上考察算法的精度、CPU 使用率和内存使用量,以找出适合嵌入式应用的方案[4]。综上所述,本部分明确了 VI-SLAM 的重要性和现实挑战,并引出了后续研究所关注的精度-资源平衡问题。

1.3 视觉 SLAM 的演进:从单模态到多传感器融合

视觉定位与地图构建的研究最早可追溯至 20 世纪 80 年代的机器人导航研究。在早期阶段,受限于计算能力,研究者多采用扩展卡尔曼滤波 (EKF) 来维护地图点和相机位姿。2007 年, Davison 提出的 MonoSLAM [11] 标志着首个实时单目视觉 SLAM 系统的诞生。然而,由于 EKF 的协方差矩阵随地图规模呈平方级增长,系统难以处理大规模场景。

随后, Klein 等人提出的 PTAM (Parallel Tracking and Mapping) [12]引入了“并行化”思想, 将跟踪与建图分拆为两个线程, 并首次证明了非线性优化(Bundle Adjustment, BA)在实时性上的潜力。这一里程碑式的工作奠定了现代视觉 SLAM 的基本框架。2015 年, Mur-Artal 提出的 ORB-SLAM [13]系列通过高效的 ORB 特征和三线程结构, 解决了系统鲁棒性和回环检测的难题。

然而, 纯视觉系统在实际应用中暴露了三大致命缺陷: 尺度缺失、快速运动失效以及退化环境下的定位丢失。为了弥补上述不足, 融合惯性测量单元(IMU)成为必然选择。这种从“纯视觉”向“视惯融合”的演进, 标志着 SLAM 技术从算法验证阶段迈向了追求鲁棒感知的新时代 [14]。

为了弥补上述不足, 融合惯性测量单元(IMU)成为必然选择。IMU 能够提供高频的角速度与加速度测量, 在短时间内具有极高的相对定位精度, 恰好与视觉传感器形成互补。这种“视惯融合”的演进, 使得系统能够在复杂多变的现实环境中保持亚米级的定位精度。

2 VI-SLAM 核心算法分类与对比分析

现有视觉惯性 SLAM 系统主要可分为两大类别: 一类基于非线性优化(BA)的批量或滑动窗口方法, 另一类基于滤波(EKF/MSCKF)的递归估计方法。下面分别对这两类主流方法的核心原理、代表性系统及其特点做深入分析, 并探讨它们在轻量化方向的潜力。

2.1 基于优化的方法 (Optimization-based SLAM)

基于优化的方法通常采用最大后验估计和束调整技术, 将摄像头观测和 IMU 预积分后的测量误差以非线性最小二乘的方式联合优化。这类方法的优点在于精度最高、鲁棒性强, 特别是在回环检测 (Loop Closing) 后可以通过全局优化消除累计误差, 实现全局一致性。例如, Campos 等人提出的 ORB-SLAM3 系统依托 MAP 估计策略, 在各种传感器配置下都能实时运行, 其精度是之前方法的两到十倍[2]。ORB-SLAM3 创新地引入了多地图机制: 当系统在视觉信息不足时断链, 它会建立新地图并在重新识别之前区域时自动将其与旧地图无缝融合。这种跨会话的地图合并策略允许利用之前构建的所有地图信息来增强束调整, 即使相机经过了长时间且具有高视差的路径, 也能在全局优化中纳入这些共视关键帧, 从而显著提升最终定位的准确性[2]。类似地, P. Geneva 等人开发的 Kimera 框架则进一步在此基础上集成了语义标签和网格重建模块[6], 使系统可

在实时时间下生成稠密的三维网格和语义地图。可以看出，这些优化式系统在性能上处于领先地位，但其缺点是不容忽视的高计算开销[2, 4]。非线性优化过程需要对大量变量迭代求解，计算量随地图大小和关键帧数目呈指数增长。因此，这类方法的实时性高度依赖于平台的计算能力，并且往往需要借助 GPU 加速或离线处理在大环境中保持可行性。

2.2 基于滤波的方法 (Filtering-based VIO)

另一大类方法基于扩展卡尔曼滤波 (EKF) 或 MSCKF 等滤波器进行状态估计。其核心思想是采用递归预测-更新结构，将系统状态向量 (包含位姿、速度、IMU 偏置等) 和协方差矩阵一同维护，通过融合新观测来更新状态估计。与优化式方法不同，滤波方法并不显式求解批量优化问题，而是不断利用线性化的滤波过程来逼近非线性估计问题，因此通常计算效率更高、延迟更低，天生更适合资源受限的平台。代表性系统包括 MSCKF (Mourikis 等提出，2007 年 IEEE ICRA) [7]，OKVIS (Leutenegger 等人提出，2015 年)、R-VIO (紧凑惯性滤波器) 和最新的 OpenVINS 等[3]。例如 MSCKF 的核心贡献在于引入多状态约束模型，使得静态特征经过多个相机位姿观测的几何约束能够在滤波中得到表达，但无需将每个地图点状态加入滤波器中，从而使计算复杂度仅与特征数量线性增长[7]。OpenVINS 是由日内瓦等人开发的一个开源研究平台[3, 5]，它基于改进的 EKF/MSCKF 架构，为研究者提供了灵活的模块化框架。OpenVINS 强调可重复性和可移植性，其实现的 VIO 系统经实验证明能与最先进的公开算法竞争。与优化法相比，滤波法固有地更轻量，适合在嵌入式板卡上运行，其瓶颈主要在于状态向量的维度和协方差矩阵规模会随时间累积增长。滤波系统需要不断通过增加 IMU 克隆或关键帧来提高状态估计精度，而每新增一个状态变量都会导致协方差矩阵增大，进而增加计算负担。因此，滤波方法通常采用边缘化 (marginalization) 策略将过时状态转化为先验约束，以控制问题维度。例如，当状态超过滑动窗口范围时，MSCKF 会将最老的相机克隆边缘化掉，将其约束以先验形式保留下来，从而使状态维度保持可控。总体来看，优化法与滤波法各有千秋：前者精度高且善于全局一致性，但计算消耗大；后者效率高、天生轻量，但累积误差需要通过姿态图等方法修正。

3 轻量化与资源效率优化策略研究

针对上述计算和内存瓶颈，近年来研究者提出了多种轻量化优化策略，旨在降低前端和后端的负担，为在资源受限平台上稳定运行奠定理论基础。下面从前端特征处理、

后端状态管理和地图稀疏化三个方面进行分析。

3.1 前端特征处理的效率权衡

在视觉惯性 SLAM 中，前端通常负责提取和跟踪图像特征，为后端优化或滤波提供观测。针对资源受限的需求，很多系统放弃了 ORB/SIFT 等计算代价高昂的描述子匹配，而采用 Shi-Tomasi 角点检测与 KLT 光流跟踪的组合。Shi-Tomasi 算法保证了特征点在图像中的均匀分布，而金字塔 KLT 光流能应对较快的相机运动和亮度变化[10]。这种做法避免了特征描述子和暴力匹配的开销，从而显著降低了前端计算负担。例如，在 OpenVINS 和 VINS-Fusion 等系统中，前端仅通过角点检测和光流跟踪来维护特征，与传统 ORB 特征相比更轻量高效。同时，为了进一步优化资源占用，需要动态调节跟踪特征的数量与分布。在嵌入式平台上，如果图像纹理过密而始终使用固定数量特征，会导致冗余计算和实时性波动。为此，可以采用“网格均匀化”策略，即将图像划分网格，并在每个网格内限制特征数量，从而避免某些纹理丰富区域占用过多计算资源。结合动态特征控制策略，使系统根据当前算力和运动状态自动调整跟踪特征数目，有助于在保证跟踪稳定性的前提下，实现前端计算量的自适应可控。这一方向的研究尚处于起步阶段，但对于整体轻量化具有重要意义。

3.2 后端状态管理与维度控制

基于滤波的 VIO 方法的核心瓶颈在于状态向量的维度和协方差矩阵随时间线性增长。滤波器需要通过不断增广状态（如新增的 IMU 克隆和特征）来保持估计精度，但无序增长会导致计算和存储瞬间爆炸。为应对这一问题，主流方法都采用边缘化机制：当状态数目达到阈值时，将最老的状态变量（如最早的相机姿态克隆）边缘化出滤波器，将其历史信息以先验高斯约束的形式融入协方差，以控制问题规模。然而，这种 FIFO（先进先出）式的边缘化策略没有考虑各状态对整体信息的贡献度，因此可能保留一些对当前定位价值不高的状态。同时，一些高信息量状态被无差别地边缘化掉，导致滤波性能下降。为此，研究者提出基于协方差显著性或信息量贡献度的剪枝策略，即在边缘化时评估每个状态对估计精度的增益，将对精度贡献小的状态优先剔除。这一思想类似于信息导向的状态选择，旨在用更少的状态维度维持估计质量，从而加快滤波运算并降低内存需求。在具体实现中，可以通过计算状态协方差的特征值、信息量指标或利用稀疏信息滤波等方法确定边缘化顺序。这类基于信息量的状态修剪方法尚未广泛应用于主流 VIO 系统，但被认为是提升滤波效率和资源利用率的关键。

3.3 大规模 SLAM 的内存瓶颈与稀疏化

随着运行时间增加，优化式和滤波式 SLAM 系统都会面临地图规模膨胀带来的内存问题。尤其在无闭环约束下，累积的大量地图点（尤其是非局部的冗余点）和关键帧会持续累积，占用大量 RAM。即便加入回环检测，地图也会不断增长。例如，如果在 OpenVINS 中集成基于视觉的回环闭合，系统会不断保存新的关键帧和回环产生的图优化数据，最终内存占用可能难以接受。为了解决这一问题，Xiaoyu Zhang 等人提出了滑动窗口地图稀疏化（Sliding Window Map Sparsification, SWMS）技术，用于并行删除非局部的冗余地图点[8]。其核心思想是：在每次优化迭代时，保留当前滑动窗口内的局部地图点以保证跟踪鲁棒性，同时周期性地检查全局地图中被长期遗忘或贡献不大的点，将其从地图中剔除。经实验证明，在保持回环检测和定位精度的前提下，该方法可以将内存消耗减少 70% 以上[8]。具体来说，原文描述只保留约 10% 的地图点，其余冗余地图点通过稀疏化线程并行删除。在公开数据集和实际场景测试中，MS-SLAM 系统实现了与最先进 SLAM 相当的定位准确度，同时显著降低了内存占用，使得视觉 SLAM 在大规模环境中的可扩展性大大增强。这一研究表明，通过合理的稀疏化策略，可以在长时间运行下保持地图规模可控，对嵌入式视觉 SLAM 的可持续运行具有重要意义。

4 总结与本文研究的创新点

综上所述，当前的研究虽已提出多种高精度和高效的 VI-SLAM/VIO 算法，但它们往往只关注精度或计算效率，而缺乏一个同时考虑计算成本、内存消耗和动态资源管理的集成解决方案。尤其针对资源受限的嵌入式平台（如 Jetson Nano）进行的轻量化优化研究更为稀缺。现有方法多聚焦于单一方向（如加速后端优化或加速前端处理），但如何在保持定位精度的同时系统性地优化全局性能尚缺乏深入研究。为填补这一空白，本文的主要贡献和创新点包括：

- 1. 系统级基线量化：**以经典的 VINS-Fusion (优化法) 和原生 OpenVINS (滤波法) 作为基线系统，在目标嵌入式平台上对比分析它们的“精度-资源”性能曲线。通过实测评估两者在不同配置下的 ATE、RPE 以及帧率、CPU、内存占用等指标，为轻量化优化提供参考基准。
- 2. 前端特征处理优化：**在原生 OpenVINS (KLT) 前端基础上，引入 **动态特征数量控制** 和 **网格均匀化** 策略。动态控制根据当前场景和计算负载实时调整追踪特征点数目，避免过密区域带来冗余计算；网格均匀化则保证特征在图像平面上分布均

匀,从而稳定前端处理量。该方案旨在实现资源可控的自适应前端,使系统能够在精度需求和计算限制之间进行权衡。

3. **后端状态管理革新:** 提出基于信息量贡献度的状态向量 **剪枝策略**。通过分析状态向量各分量对全局精度的贡献,优先边缘化信息量低的历史状态,替代传统的时间先后(FIFO)边缘化方法。这样可智能地控制滤波器状态维度,将计算资源聚焦于高信息量状态,从而显著提高滤波后端的运行效率。
4. **内存控制与地图稀疏化:** 集成滑动窗口地图稀疏化(SWMS)策略,实现全球地图增长的可控化管理。通过并行删除冗余的非局部地图点,在不影响回环检测能力的前提下,将系统内存消耗降低至原来的 20%–30%。该方法解决了 VI-SLAM 长期运行时的内存压力,使系统在大规模环境中更稳定可靠。

综上所述,本文通过在系统层面对传统 VI-SLAM 算法进行精细优化,针对前端、后端和地图管理多个环节提出协同改进措施,构建了一个兼顾精度和资源利用的轻量化定位与地图构建系统。本研究将为在嵌入式平台上实现高精度、高效率的自主导航提供理论和实践依据。

参考文献

- [1] HKUST Aerial Robotics Group. VINS-Fusion: A versatile and extensible multi-sensor visual-inertial odometry system[OL]. (2019-01-11)[2025-12-23]. <https://uav.hkust.edu.hk/>
- [2] Campos C, Elvira R, Rodríguez J J G, et al. Orb-slam3: An accurate open-source library for visual, visual-inertial and multi-map SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [3] Geneva P, Eckenhoff K, Lee W, et al. Opencvins: A research platform for visual-inertial estimation[C]. IEEE International Conference on Robotics and Automation (ICRA), 2020: 4666-4672.
- [4] Jeon J, Jung S, Lee E, et al. Run your visual-inertial odometry on NVIDIA jetson: Benchmark tests on a micro aerial vehicle[J]. arXiv:2103.01655, 2021, 2021: 1-8.
- [5] Deng Z, Luo H, Gao X. Fusion-based localization system integrating UWB, IMU, and vision[J]. Applied Sciences, 2025, 15(12): 6501.
- [6] Rosinol A, Abate M, Chang Y, et al. Kimera: An open-source library for real-time metric-semantic localization and mapping[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 713-720.
- [7] Qin T, Li P, Shen S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.
- [8] 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 (ICRA), 2007: 3565-3572.
- [9] Song X, Zhao S, Sui J, et al. Enhancing VINS with smart feature grading[C]. ISPRS Geospatial Week, 2025: 47-54.
- [10] 张晓宇, 董晋虎, 张寅, 等. Ms-slam: Memory-efficient visual SLAM with sliding window map sparsification[J]. Journal of Field Robotics, 2025, 42(4): 935-951.
- [11] Davison A J, Reid I D, Molton N D, et al. Monoslam: Real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [12] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]. IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007: 225-234.
- [13] 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.
- [14] 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.