ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 2,Mar. 2026,pp. 592－619
DOI:10. 16337/j. 1004-9037. 2026. 02. 015

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

复杂低空环境下无人机自主定位技术研究进展

许悦雷[1,2,3]，王铉彬[1,2,3]，薛尚捷[1,2,3]，徐金海[1,2,3]

(1. 西北工业大学无人系统技术研究院,西安 710072;2. 西北工业大学人工智能学院,西安 710072;3. 无人飞行器技术全国重点实验室,西安 710072)

摘　要: 复杂低空环境通常呈现出多源干扰叠加、感知条件剧烈变化与信息不完备并存等特征,对无人机自主定位的连续性、可靠性与可信性提出了严峻挑战。在此类环境下,全球卫星导航系统(Global navigation satellite system, GNSS)信号易受遮挡与干扰而失效,视觉感知面临弱纹理、强动态与光照突变等退化问题,惯性测量则不可避免地产生长期累积漂移,三者耦合作用显著削弱了定位系统的稳定性与鲁棒性。为此,本文系统梳理了低空典型退化环境类型,重点分析了多源混合干扰场景下视觉特征缺失、IMU误差发散与卫星定位性能退化等关键技术瓶颈。在此基础上,综述了无人机视觉导航定位技术的发展脉络,涵盖基于卫星/先验地图的视觉匹配定位方法以及视觉SLAM的最新研究进展;进一步总结了视觉-惯性系统融合建模与感知增强方法,阐明其在提升定位精度与稳健性方面的技术优势。随后,论述了多源融合导航框架及面向拒止环境的鲁棒融合策略,重点关注视觉、惯性、激光雷达以及卫星等多模态信息的协同建模、退化感知与完好性监测。最后,展望了数据驱动的多模态自适应导航方法以及轻量化、智能化的无人机高可信导航技术发展趋势。旨在为复杂低空环境下无人机高可靠自主定位技术的研究与工程应用提供系统参考。

关键词: 卫星拒止;视觉导航;视觉惯性里程计;激光雷达;多源融合;低空应用

中图分类号: V279　　**文献标志码:** A

引用格式: 许悦雷,王铉彬,薛尚捷,等.复杂低空环境下无人机自主定位技术研究进展[J].数据采集与处理,2026,41(2):592-619. XU Yuelei, WANG Xuanbin, XUE Shangjie, et al. A review of autonomous localization technologies for unmanned aerial vehicles in complex low-altitude environments[J]. Journal of Data Acquisition and Processing, 2026,41(2):592-619.

引　　言

无人机在城市管理、应急救援、基础设施巡检、农业监测以及战术侦察等领域的应用日益广泛,其自主导航定位能力已成为决定任务执行效能与安全性的核心要素。然而,在复杂低空环境下,无人机自主定位面临严峻挑战:全球卫星导航系统(Global navigation satellite system, GNSS)信号易受建筑遮挡、电磁干扰与多径效应影响,在城市峡谷、室内空间、隧道以及强对抗环境中可用性急剧下降;视觉感知系统在弱纹理、低照度、强光照变化、动态干扰以及恶劣气象条件下性能显著退化;惯性测量单元(Inertial measurement unit, IMU)虽可提供高频连续的运动信息,但其固有的累积漂移特性导致长时间运行后定位误差不可接受。这些退化机制相互耦合,使得单一传感器难以在复杂环境中维持可靠的定位输出。

近年来,随着计算机视觉、多传感器融合以及深度学习技术的快速发展,基于视觉的导航定位、视

基金项目: 国家自然科学基金(42504030);中央高校基本科研业务费(科技专项)(D5000250047)。
收稿日期: 2026-01-12;**修订日期:** 2026-02-28

觉惯性等多源异构信息融合定位方法在理论与工程实践中取得显著进展。基于卫星地图的视觉匹配导航通过建立机载图像与遥感影像的跨视角对应关系,实现了GNSS拒止条件下的绝对地理定位;视觉同步定位与建图(Simultaneous localization and mapping,SLAM)技术在未知环境中同步完成相机位姿跟踪与地图构建,为无人机提供了相对导航能力;视觉惯性融合系统将视觉观测的相对约束与IMU的高频运动先验紧密结合,在弱纹理、快速机动等退化场景下展现出更强的鲁棒性;多源融合导航进一步集成GNSS、激光雷达等异构传感器,通过优势互补与冗余设计,显著提升了系统在复杂拒止环境下的连续性与可靠性。

然而,现有技术在面对多源混合干扰、极端视觉退化以及长时间GNSS拒止等挑战性场景时,仍存在可观测性不足、累积漂移难以抑制、系统完好性评估缺失等问题。如何实现在退化场景下的鲁棒感知、一致性状态估计、多模态自适应融合以及可信度量化评估,已成为无人机自主导航领域亟待解决的关键科学与工程问题。

本文围绕复杂低空环境下无人机自主定位的核心挑战,系统性地梳理了相关技术的最新研究进展。如图1所示,首先针对复杂低空应用场景深入剖析了由GNSS信号遮蔽、视觉感知退化及惯性测量漂移等因素引发的定位退化机理,并在此基础上,系统地综述了3条关键技术路线:(1)纯视觉定位技术,涵盖了基于卫星地图的匹配定位与视觉SLAM方法;(2)视觉惯性定位技术,重点阐述了其在提升连续性与鲁棒性方面的优势;(3)多源异构信息融合定位技术,梳理了视觉、惯性、激光雷达等异构信息融合方法。最后对于未来复杂低空环境中无人机高可靠自主定位技术发展趋势进行了展望。本文主要贡献在于构建了一个从传感器退化机理到多源鲁棒融合的完整技术图景,为低空复杂环境下无人机高可靠自主定位技术的研究与工程应用提供了系统性的参考与前瞻性的展望。

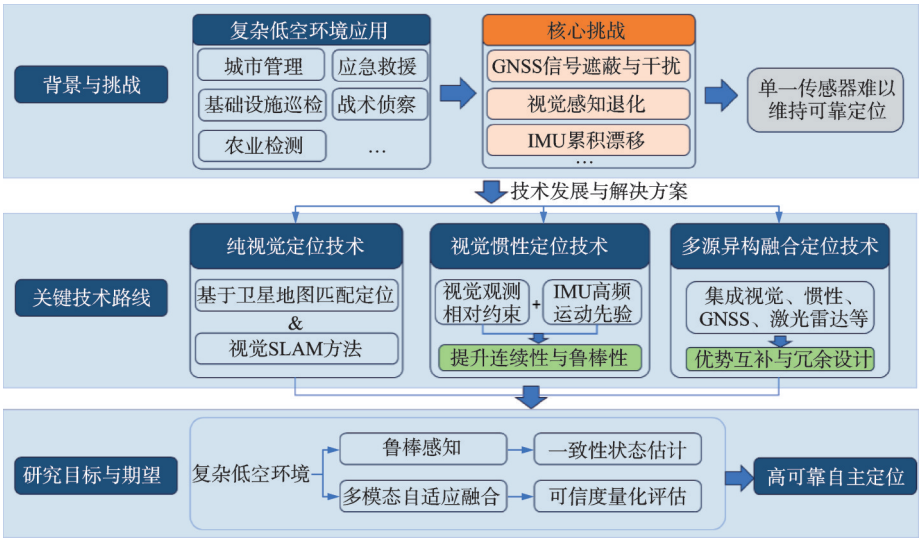

图1 复杂低空环境下无人机自主定位技术的发展脉络与研究框架

Fig.1 Development context and research framework of autonomous localization technology for unmanned aerial vehicles in complex low-altitude environments

1 低空拒止环境类型及其内涵

1.1 GNSS拒止或退化

GNSS可以为全球用户提供全天候、全天时的高精度定位、导航和授时(Positioning,navigation and

timing, PNT)服务,在交通运输、农林渔业、水文监测、气象预报、通信授时、电力调度、救灾减灾及公共安全等领域发挥着重要作用。作为当前应用最为广泛的定位导航技术,GNSS的工作原理主要是基于接收来自多系统(BDS,GPS,GLONASS,Galileo)卫星导航信号,通过测量信号传播时间计算伪距,进而解算出接收机的三维位置。然而,GNSS信号从距地面约20 200 km的卫星传播至地面接收机时,信号功率已极其微弱,典型功率仅为−130～−160 dBm量级[1]。这种固有的信号脆弱性使得GNSS系统极易受到各类环境因素的影响,导致定位性能显著下降甚至完全失效。

根据GNSS信号受影响的程度,卫星拒止环境可分为完全拒止环境与信号退化环境两大类[2],如表1所示。完全拒止环境是指卫星信号完全无法到达接收机的场景,主要包括:(1) 室内环境,如大型建筑内部、地下停车场、商场等;(2) 地下与水下环境,如隧道、矿井、地铁及水下作业场景;(3) 密集城市峡谷,高层建筑形成的深度遮挡区域;(4) 强电磁干扰区域,存在大功率干扰源的特殊场景[3]。信号退化环境则是指GNSS信号虽可部分接收,但由于各类干扰因素导致定位精度与可靠性显著下降的场景。

表1 GNSS拒止与退化环境分类

Table 1 GNSS rejection and degraded environment classification

类型	典型场景	信号特征	定位影响
完全拒止	室内环境(建筑内部、商场)	信号完全阻断	无法定位,需完全依赖替代导航源
	地下环境(隧道、矿井、地铁)	信号无法穿透	
	强干扰区(大功率干扰源)	信号被淹没	
信号退化	城市峡谷(高楼林立街道)	部分遮挡、多径	卫星几何分布退化,定位精度骤降
	树荫遮蔽(森林、林荫道)	信号衰减以及散射	信号间歇失锁
	桥梁/立交下方	短时遮挡	可见星数骤降,定位中断
	电磁干扰区(机场、军事区)	干扰/欺骗	精度下降或输出错误位置

从应用场景角度分析,无人机在执行城市巡检、桥梁检测、室内探测以及战术侦察等任务时,频繁经历GNSS信号状态的动态切换。这种信号可用度的时空变异性对无人机自主导航系统提出了严峻挑战,迫切需要发展能够在拒止/退化动态环境下维持可靠定位的替代技术方案。

1.2 视觉场景退化

视觉感知系统能够为无人机提供丰富的环境纹理与几何结构信息,是实现自主定位与导航的“眼睛”。作为当前应用最为广泛的外部感知手段,视觉导航的工作原理主要是基于相机获取的连续图像序列,通过特征提取、跟踪与匹配来解算载体的三维位姿。然而,视觉传感器的成像质量具有极高的敏感性。这种环境依赖性使得视觉系统极易发生性能退化,导致匹配精度下降甚至定位完全失效。根据环境因素对视觉成像及算法几何约束的影响机理,视觉退化环境主要可归纳为以下两类:

第1类是光照与气象导致的成像退化环境。这主要包括光照剧变、极端光照以及恶劣气象(雨、雪、雾、沙尘)。这些因素直接破坏了基于灰度不变性假设的特征提取基础,导致视觉特征的显著性丧失和描述子的区分能力下降,进而引发前端数据关联的失败。

第2类是纹理缺失与场景动态导致的几何退化环境。其一为弱纹理或重复纹理环境,如开阔水面、雪地、沙漠(弱纹理)以及大面积的建筑窗格、地砖(重复纹理)。前者因缺乏足够的特征点而导致位姿估计的自由度退化,后者则因特征外观的高度相似性而引发匹配歧义[4]。其二为动态与遮挡环境,在复杂的城市低空或作战场景中,频繁出现的移动行人、车辆及障碍物遮挡,破坏了视觉SLAM算法通常依赖的“静态世界”假设,导致特征跟踪不稳定及视觉观测的间歇性中断。

从应用场景角度分析,无人机在执行城市巡检、桥梁检测或战术侦察任务时,不仅面临GNSS信号的拒止,往往还需频繁经历上述视觉环境的动态切换。这种视觉观测条件的时空变异性(如从白天强光转入涵洞暗光)以及多源退化耦合,严重制约了单一视觉导航系统的连续性与可靠性,迫切需要发展能够适应多变环境的鲁棒定位技术。

1.3 惯性导航漂移

惯性导航系统依赖对加速度与角速度的连续积分来推算载体的姿态、速度与位置,其突出优势在于自主性强、短时精度高。然而,由于惯性传感器不可避免地存在零偏、比例因子误差与噪声等不完美因素,这些微小误差在时间积分过程中被持续放大,导致导航解算结果随时间不断累积偏离真实状态,表现为典型的"漂移"现象。尤其在长时间运行或缺乏外部约束的情况下,惯性导航的定位误差将呈发散趋势,严重制约其独立应用能力[5]。惯性误差主要通过两种机制随时间不断放大:一类是随机噪声,使系统不确定性逐步扩散;另一类是零偏误差,尤其通过"姿态-重力耦合"机制引发更为显著的系统性漂移。具体而言,陀螺零偏会导致姿态估计误差随时间近似线性增长,进而引入重力方向的投影误差,形成并不存在的水平加速度,使速度和位置误差被持续放大;加速度零偏则更为直接,在速度和位置的积分过程中不断累积。这使得惯性导航呈现出"短时间内平滑可靠、长时间运行易发生漂移"的典型特性。

在拒止环境中,视觉、地图或GNSS等外部校正信息往往间歇可用或质量下降,系统对IMU递推结果的依赖显著增强,惯导漂移会更加突出,常表现为轨迹逐渐偏离真实路径、姿态缓慢发生偏转以及速度出现系统性偏差,并进一步导致视觉导航回环难以触发、融合残差增大等一系列退化问题。

1.4 多源混叠误差及其影响

前述分析表明,卫星导航拒止、视觉场景退化与惯性导航漂移分别从不同维度对无人机自主定位能力构成威胁。然而,在实际拒止环境中,上述3类退化因素往往并非孤立出现,而是以时空交叠、动态耦合的方式共同作用于导航系统,形成"多源混叠"干扰态势。这种复合退化模式所引发的系统级误差传播与放大效应,是制约当前无人机拒止环境下自主导航性能的核心难题[6]。

低空应用场景因其空间结构的复杂性和电磁环境的多样性,是上述多源混叠干扰最为典型的实践场域。如图2所示,无人机在飞行过程中需频繁穿越多种异构退化环境:在城市峡谷区域,高层建筑群

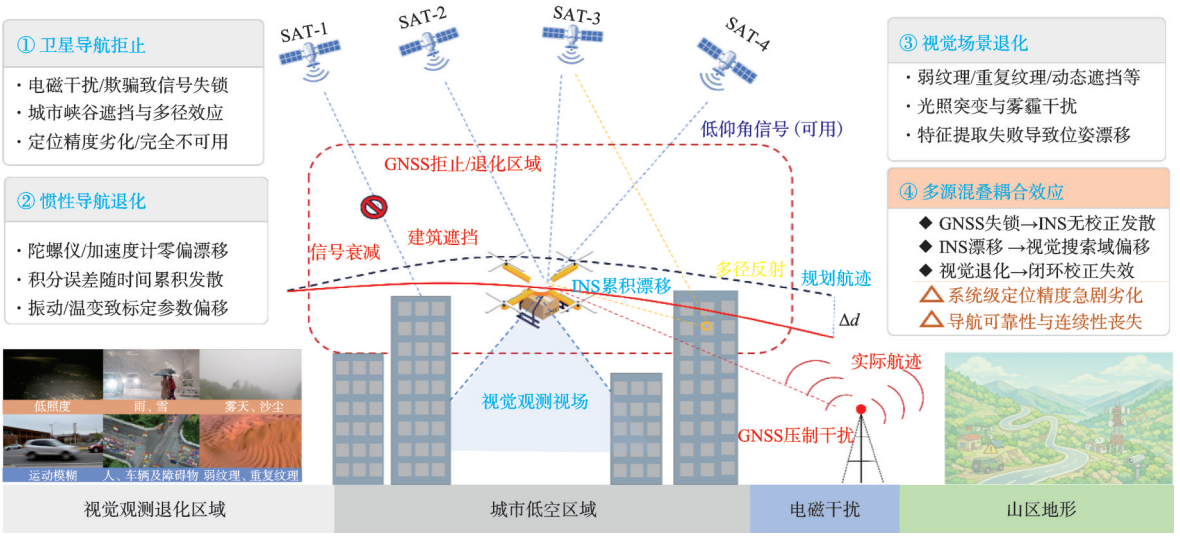

图2 复杂低空区域存在的多源混叠误差及其影响

Fig.2 Multi-source mixed errors in complex low-altitude environments and their impacts

导致卫星信号遮蔽和建筑表面反射引发严重多径效应，几何精度因子(GDOP)显著恶化[7]；在密林覆盖区域，植被冠层的散射衰减使卫星信号载噪比劣化；在电磁对抗或通信基站密集区域，人为干扰或射频噪声直接压制GNSS接收机[8]。上述多种干扰因素在时间和空间上交替出现、相互叠加，导致无人机实际飞行航迹偏离规划航迹，系统级定位可靠性急剧下降。

由此可见，在复杂低空环境下依赖单一导航信息源或简单的双源组合策略均难以保证无人机定位的连续性与可靠性。提升无人机在复杂拒止环境下的自主导航能力，必须立足于多源异构信息融合的技术框架，充分利用GNSS、INS、视觉等多种导航信息源的互补特性，构建冗余度更高的观测体系；同时发展具备退化感知能力的智能融合算法，使系统能够实时识别各信息源的可信度状态，动态调整融合权重与策略，在部分信息源退化甚至完全失效的条件下维持高可靠、高可信的连续定位能力。后续将围绕上述技术方向，系统梳理各类自主定位方法的研究进展与发展趋势。

2 基于视觉信息的自主定位

基于纯视觉信息的导航定位技术是拒止条件下无人机实现高可靠自主导航的核心手段之一，主要依赖机载视觉传感器实时采集的图像信息，通过先进的计算机视觉算法完成位姿估计、环境感知与路径规划。根据参考信息的来源与处理方式，目前主流的视觉导航定位技术可概括为两大类互补的技术路线：一类是基于卫星地图的视觉匹配导航；另一类是视觉SLAM导航技术。

2.1 基于卫星地图的视觉匹配导航

如图3所示，基于卫星地图的视觉匹配导航主要利用无人机机载视觉传感器获取的图像与事先已知地理坐标的卫星遥感影像进行匹配，从而实现无人机自主定位与导航。该类方法不依赖于GNSS信号，在复杂电磁环境、城市峡谷以及GNSS拒止条件下具有显著优势。

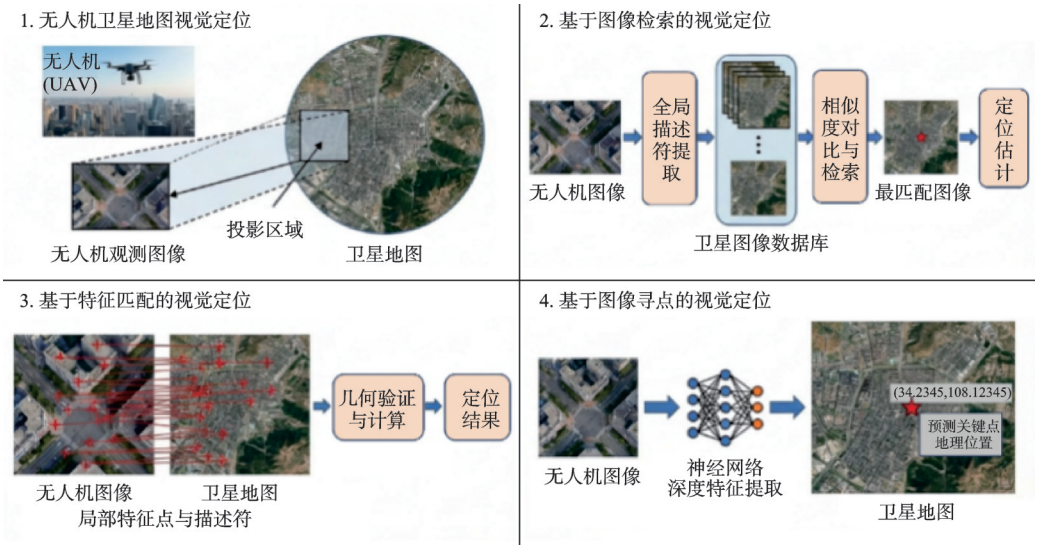

图3 基于卫星地图的视觉匹配导航方法示意图

Fig.3 Schematic diagram of satellite map-based visual matching navigation method

从技术实现角度来看，视觉匹配导航主要通过建立机载视角图像与卫星俯视图像之间的对应关系来实现定位。现有视觉匹配导航方法主要划分为3类：基于图像检索的方法、基于局部特征匹配的方法以及基于图像寻点的方法。以下将分别对这3类方法的研究进展与技术特点进行综述。

2.1.1 图像检索方法

图像检索方法是最早被应用于基于卫星地图视觉定位的一类经典技术路线,其核心思想是将无人机采集的图像作为查询图像,通过提取全局或局部描述特征,在带有地理标注的卫星图像数据库中检索最相似的图像,从而推断无人机的大致地理位置,其主要流程如图4所示。

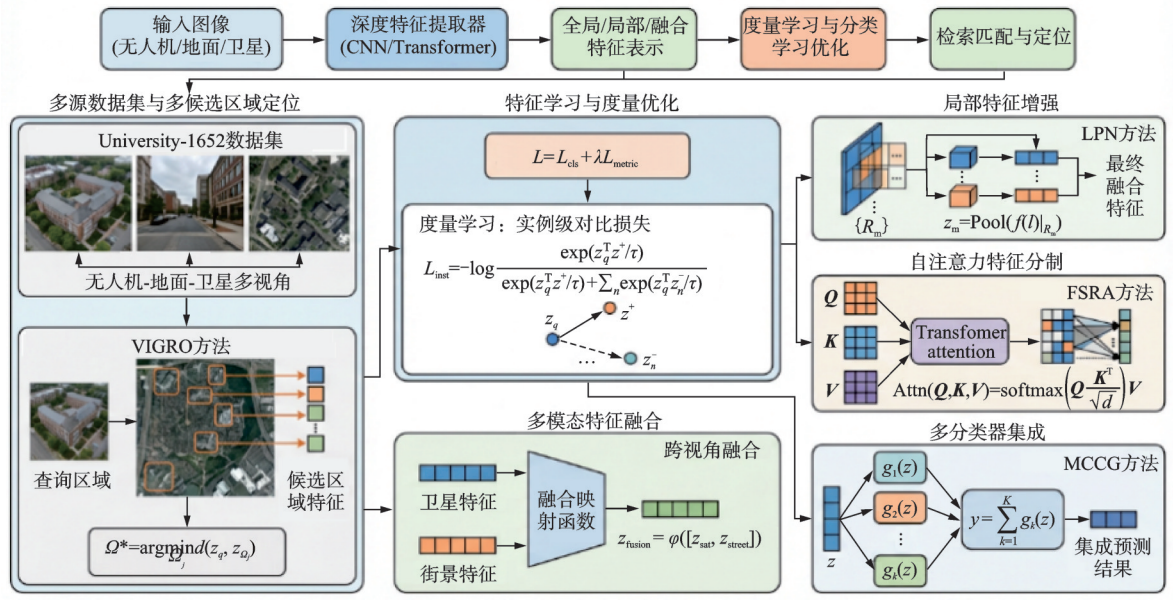

图4 图像检索方法示意图

Fig.4 Schematic diagram of image retrieval methods

对于该类方法,Zheng等[9]提出了University-1652数据集,这是首个大规模、多视角、多源的无人机地理定位基准数据集。该数据集涵盖全球72所大学的1 652栋建筑物,总计超过5万张图像,支持两个全新任务:无人机图像-卫星图像检索定位目标建筑物和卫星图像-无人机图像导航。作者同时通过对比损失有效促进跨视角特征的对齐与判别学习,取得了良好的精度。Dai等[10]提出了一种基于Transformer的特征分割和区域对齐(FSRA)方法,针对多高度、多场景的无人机视角地理定位。Shi等[11]通过融合卫星图像与街道视图混合检索策略,进一步提升了复杂城市场景下鲁棒性。

2.1.2 特征匹配方法

为进一步提升定位精度,研究者提出了基于局部特征匹配的视觉定位方法。该类方法首先从无人机图像和卫星地图中提取局部特征点及其描述子,然后通过特征匹配建立两幅图像间的对应关系,最后结合匹配点的空间分布、相机模型及卫星地图地理信息,利用PnP(Perspective-n-point)或其变体算法求解无人机的精确位姿。

早期的特征匹配方法主要依赖手工设计的特征算子,通过梯度信息实现特征提取。Lowe等[12]提出的SIFT算法通过构建尺度空间以实现尺度不变性。SURF通过Hessian矩阵的行列式进行特征点检测[13]。FAST通过像素环比较机制快速检测角点[14],ORB则结合FAST检测器和旋转BRIEF描述子,实现了高效的二进制特征表示与实时性能[15]。然而,这些方法在光照剧烈变化、大视角差异或尺度剧变等复杂场景下,特征的重复性与匹配稳定性明显下降。

如表2所示,近年来,深度学习极大地推动了局部特征匹配技术的进步[16,22,26]。根据匹配稠密程度的不同,可分为稀疏匹配[16]、半密集匹配[22]和密集匹配[26]3类。稀疏匹配方法基于关键点检测与描述,

通过两阶段学习表征实现图像间的精确映射。Detone等[16]提出SuperPoint算法，通过同时预测关键点置信度图和描述子。Revaud等[18]提出的R2D2，利用可靠性和可重复性联合优化策略，进一步提升了特征在极端条件下的表现。Sarlin等[19]提出SuperGlue，将特征匹配问题转化为可微分的最优传输问题实现联合的对应关系估计与非匹配点剔除。对于半密集匹配方法，Sun等[22]提出LoFTR算法，通过粗到细的Transformer架构实现无检测器的匹配，如图5所示。Chen等[23]提出ASpanFormer，构建分层注意力框架的无检测器匹配器平衡全局与局部上下文。Wang等[25]提出的Efficient LoFTR则通过聚合注意

表2 无人机视觉定位中特征匹配方法对比

Table 2 Comparison of feature matching methods in UAV visual localization

类别		算法	年份	核心原理	作用
手工设计		SIFT[12]	2004	尺度空间极值检测和梯度方向直方图描述	关键点提取
		FAST[14]	2006	像素环比较检测角点	关键点提取
		SURF[13]	2008	积分图像和快速Hessian检测	关键点提取
		ORB[15]	2011	结合FAST检测器和旋转BRIEF描述子	关键点提取
深度学习	稀疏匹配	SuperPoint[16]	2018	自监督和半监督联合训练	关键点提取
	稀疏匹配	D2-Net[17]	2019	端到端联合优化检测与描述	关键点提取＋匹配
	稀疏匹配	R2D2[18]	2019	可靠性和可重复性联合优化	关键点提取＋匹配
	稀疏匹配	SuperGlue[19]	2020	图神经网络与最优传输	关键点匹配
	稀疏匹配	LightGlue[20]	2023	旋转位置编码、自适应剪枝机制和早退策略	关键点匹配
	半密集匹配	LoFTR[22]	2021	自注意力＋交叉注意力	端到端匹配
	半密集匹配	ASpanFormer[23]	2022	分层注意力	端到端匹配
	半密集匹配	DeepMatcher[24]	2024	多尺度卷积和轻量化Transformer	端到端匹配
	半密集匹配	Efficient LoFTR[25]	2024	聚合注意力和动态令牌选择机制	端到端匹配
	半密集匹配	OmniGlue[21]	2024	利用DinoV2提取图像特征	端到端匹配
	密集匹配	DKM[27]	2023	基于高斯过程的全局核回归匹配器	端到端匹配
	密集匹配	RoMa[28]	2024	提出预测锚点概率Transformer解码器	端到端匹配
	密集匹配	HomoMatcher[26]	2025	基于单应估计的稠密匹配	端到端匹配

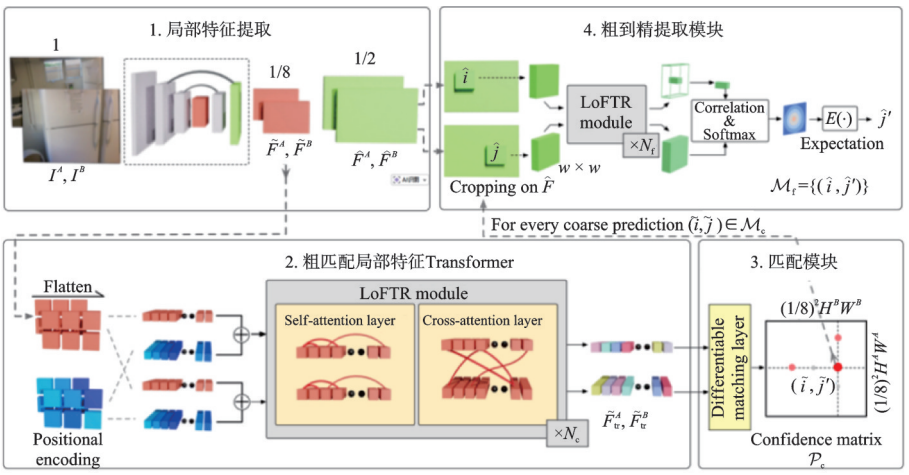

图5 LoFTR算法流程图[22]

Fig.5 LoFTR algorithm flowchart[22]

力和动态令牌选择机制显著优化效率。对于密集匹配,Wang等[26]提出的HomoMatcher基于单应估计的稠密匹配实现高精度对应关系的细化。Edstedt等[28]提出RoMa利用冻结的DINOv2基础模型特征结合专用ConvNet细特征构建特征金字塔,采用回归分类+鲁棒回归的损失设计,大幅提升极端条件下的匹配鲁棒性。

基于上述特征匹配技术,一些研究工作进一步提出了端到端的视觉定位框架[29-32],将特征匹配与位姿估计紧密耦合,实现无人机在卫星地图上的高精度定位。Xu等[29]通过端到端训练框架实现了卫星图像检索与精细特征匹配的有机结合。Shi等[30]通过VIGOR数据集实现了动态候选区域优化与精确导航,Cui等[33]通过分段软边三元组损失函数实现了单阶段图像检索。

2.1.3 寻点定位方法

针对图像检索方法定位精度受数据库分辨率与采样密度限制、特征匹配方法在实时性与极端场景鲁棒性方面存在的瓶颈,近年来涌现出一类新型的基于图像寻点的视觉定位范式(Finding point with image,FPI)。该类方法的核心思想是:摒弃大规模图像检索或显式特征匹配的传统路径,直接通过深度神经网络在卫星参考地图上预测与无人机图像中心对应的关键地理位置点,从而实现端到端、高效且高精度的绝对定位。典型代表包括WAMF-FPI[34]、OS-FPI[35]和DCD-FPI[36]。这类方法通常将跨视角定位问题转化为热力图回归或概率分布预测任务,通过学习无人机图像与卫星地图之间的隐式空间对应关系,在推理阶段显著降低计算复杂度,更适合资源受限的无人机实时应用。

综上所述,现有基于卫星地图的视觉匹配导航方法主要包括图像检索方法、基于局部特征匹配的方法以及基于图像寻点的方法,三者在技术路径、计算复杂度与工程适用性方面各具特点。

对于图像检索方法而言,其在跨视角条件下表现出较强的鲁棒性,但其本质仍属于粗定位范畴。该类方法通常仅能返回与已知地理坐标点对应的最相似卫星图像,无法提供精确的几何位姿信息,因此定位精度高度依赖卫星图像数据库的分辨率与采样密度。此外,深度检索模型参数量大、计算开销高,在资源受限的无人机平台上实现实时部署仍面临挑战。因此,图像检索方法更适合作为视觉导航系统中的候选区域筛选或粗定位模块,而难以独立满足高精度定位需求。

基于局部特征匹配的方法具备较强的几何约束与可解释性,在纹理丰富区域可获得较高定位精度。然而,在跨视角、尺度变化显著或纹理重复区域,传统局部特征易出现匹配退化;同时,大量特征提取与匹配操作带来较高计算复杂度,对嵌入式平台实时运行构成挑战。此外,该方法对光照变化与季节差异较为敏感,泛化能力受限。

基于图像寻点的定位方法为卫星地图辅助的无人机视觉导航开辟了一条全新的技术路径,在定位精度、计算效率与实时性之间取得了良好平衡。与传统图像检索和特征匹配方法相比,该范式避免了大规模数据库依赖与复杂匹配过程,推理阶段计算开销显著降低,更适合嵌入式平台部署。然而,在极端弱纹理、严重遮挡或高度变化剧烈的场景下,热力图峰值预测的稳定性仍有待进一步提升,未来可结合多模态信息或显式几何约束进一步增强系统鲁棒性。

2.2 视觉SLAM导航技术

视觉SLAM导航技术是无人机在未知环境中实现自主定位与地图构建的核心方法之一,其大致流程包括前端(视觉里程计)和后端(优化与地图管理)两大模块。前端负责从连续图像序列中提取特征(如角点、边缘或像素强度),通过特征匹配或直接最小化光度误差来估计相机位姿增量,实现实时跟踪;后端则通过全局优化修正前端积累的漂移,进行回环闭合检测以确保地图一致性,并构建稀疏、半稠密或稠密地图,表3对比了代表性视觉SLAM方法及其关键技术。

表3 视觉SLAM导航技术主要方法对比

Table 3 Comparison of main methods of visual SLAM navigation technology

类别	子类别	算法	年份	关键技术
基于几何的经典SLAM	特征点法	PTAM[37]	2007	并行跟踪与建图,采用FAST角点
	直接法	LSD-SLAM[39]	2014	半稠密直接法,光度误差最小化
	半直接法	SVO[40]	2014	稀疏直接图像对齐,特征匹配辅助
	特征点法	ORB-SLAM[38]	2015	采用ORB特征,支持DBoW2词袋,全局BA以及回环检测
基于深度学习的SLAM	学习增强	CNN-SLAM[41]	2017	融合LSD-SLAM实现稠密重建
	端到端	DeepVO[42]	2017	CNN+RNN端到端序列位姿回归
	神经隐式	NICER-SLAM[43]	2023	采用分层神经隐式SDF编码
	神经隐式	Uni-SLAM[44]	2024	基于几何解耦哈希网格进行实时稠密重建
基于语义感知的SLAM	语义融合	SemanticFusion[45]	2017	采用CNN语义分割,构建语义3D地图
	动态剔除	DynaSLAM[46]	2018	采用Mask R-CNN,可实现背景修复
	动态跟踪	MaskFusion[47]	2018	实现多物体的实例分割,跟踪以及重建
	语义过滤	DS-SLAM[48]	2018	采用SegNet进行语义分割,基于运动一致性检查提升准确率

2.2.1 基于几何的视觉SLAM方法

基于几何的视觉SLAM是视觉定位领域的经典范式,其核心思想是利用多视图几何,通过构建图像观测与三维空间之间的几何约束,实现相机位姿的在线估计与环境地图构建。该类方法依赖明确的成像模型与优化框架,具有良好的可解释性和较高的定位精度,但对动态环境、弱纹理或光照变化较为敏感。

在基于特征点的几何SLAM方法中,核心问题是通过多帧图像中匹配的特征点恢复相机位姿。根据图像信息利用方式的不同,基于几何的视觉SLAM方法可分为基于特征点的方法和基于直接法的方法。基于特征点的方法通过构建稀疏三维地图,并在滑动窗口或全局范围内进行光束法平差(Bundle adjustment, BA),其标准优化形式为

$$\min_{\{\boldsymbol{R}_k,\boldsymbol{t}_k,\boldsymbol{X}_j\}}\sum_{k,j}\boldsymbol{u}_{kj}-\pi(\boldsymbol{K}(\boldsymbol{R}_k\boldsymbol{X}_j+\boldsymbol{t}_k))_2^2 \tag{1}$$

式中:$(\boldsymbol{R}_k,\boldsymbol{t}_k)$表示第$k$帧相机位姿,$\boldsymbol{X}_j$为第$j$个三维地图点,$\boldsymbol{u}_{kj}$为其在第$k$帧中的观测。相比之下,直接法SLAM不显式提取特征点,而是通过最小化像素灰度一致性进行位姿估计,其光度误差目标函数通常表示为

$$\min_{T}\sum_{\boldsymbol{p}\in\Omega}I_1(\boldsymbol{p})-I_2\Big(\pi(T\Pi^{-1}(\boldsymbol{p},d_p))\Big)_2^2 \tag{2}$$

式中:I_1和I_2表示相邻两帧图像,$\boldsymbol{p}$为参考帧中的像素点,d_p为其对应深度,$\Pi^{-1}(\cdot)$表示从像素到三维空间的反投影,$\boldsymbol{T}$为帧间位姿变换。

典型几何SLAM系统如PTAM[37]、ORB-SLAM[38]、LSD-SLAM[39]、SVO[40]等。PTAM是首个将跟踪与建图分离为并行线程的实时单目SLAM系统,采用FAST角点检测与局部匹配实现高效位姿估计,适用于小规模室内场景,但由于缺乏全局优化机制,长期运行中易产生漂移。ORB-SLAM构建了完整的特征点SLAM框架,包括跟踪、局部建图、回环检测与全局BA,并通过DBoW2实现快速重定位SVO通过稀疏图像对齐最小化光度误差,并结合小块匹配实现高速位姿估计,特别适用于无人机等快速运动平台,但地图较为稀疏且缺乏全局一致性优化。

2.2.2 基于学习的视觉SLAM方法

基于深度学习的视觉SLAM方法通过引入神经网络,对视觉特征表示、场景建模或运动估计过程进行数据驱动优化,以提升系统在复杂环境下的鲁棒性。

一类典型方法是学习增强的几何SLAM,通过学习特征点、匹配关系或深度先验,提高系统在弱纹理、强视角变化等场景中的性能。另一类则尝试端到端网络从图像序列直接回归相机位姿,减少显式几何建模依赖。近年来,神经隐式表示方法进一步扩展了学习型视觉SLAM的范式,通过连续、可微的场景表示实现位姿与地图的联合优化。Tateno等[41]提出了CNN-SLAM,该方法首次将卷积神经网络(CNN)用于实时SLAM系统,适用于室内场景但依赖训练数据泛化。Wang等[42]开发了DeepVO,一种端到端深度循环卷积神经网络(RNN-CNN)用于视觉里程计,通过序列学习从连续图像中直接回归位姿。Bloesch等[49]设计了CodeSLAM,通过学习紧凑的隐式场景表示(代码向量)与传统SLAM结合,实现稠密重建和位姿优化。Li等[50]开发了UndeepVO,通过自监督训练从单目序列中联合估计位姿和深度。

近年来,神经隐式表示方法(如基于NeRF或神经SDF的SLAM)进一步扩展了学习型视觉SLAM的研究范式。如图6所示,这些方法利用神经辐射场(NeRF)或符号距离函数(SDF)来隐式编码场景几何和外观,支持高保真重建和鲁棒跟踪,适用于动态或稀疏观测场景,但往往需处理优化效率和泛化问题[43]。

Wang等[44]提出的Uni-SLAM是一种不确定性感知的神经隐式SLAM系统,该方法引入像素级不确定性分析来重新加权损失函数,识别异常值,并使用图像级不确定性指导局部到全局的捆绑调整。Zhu等[43]开发了NICER-SLAM,一种用于RGB SLAM的神经隐式场景编码方法,该框架引入分层神经隐式编码以优化SDF表示,尤其在光照变化场景中提升鲁棒性。Yan等[51]设计了CLID-SLAM,一种耦合LiDAR-惯性神经隐式稠密SLAM系统,在复杂户外环境中表现出色。

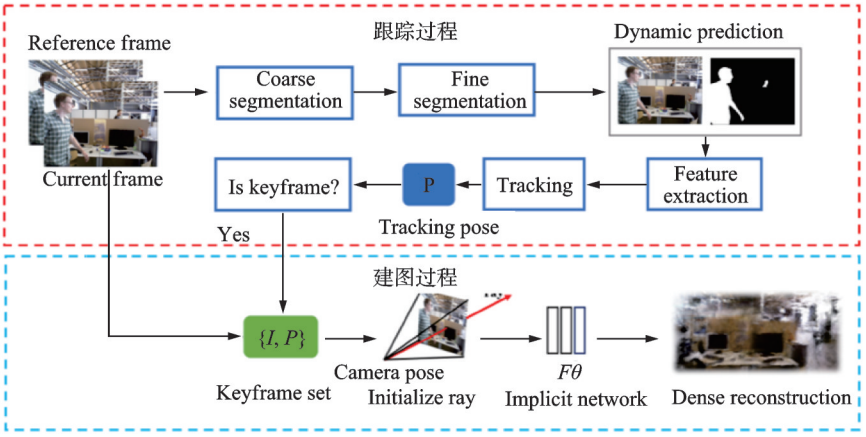

图6 基于NerF的定位方法流程图[43]

Fig.6 Flowchart of NerF-based localization method[43]

基于语义感知的视觉SLAM方法通过引入高层语义信息,使SLAM系统在几何建模基础上具备场景理解能力。该类方法通常结合语义分割、实例分割或目标检测技术,对环境中的物体和区域进行语义标注,从而辅助定位与建图过程。语义信息通过识别并剔除动态物体,提升位姿估计稳定性,还有助于构建结构化和长期一致的环境模型,增强系统在跨时间、跨环境条件下的定位能力。然而,语义感知SLAM往往依赖复杂深度神经网络,其计算开销和语义推理误差仍是需进一步研究的问题。

Mccormac等[45]提出了SemanticFusion，该方法结合CNN语义分割与ElasticFusion稠密SLAM，通过概率融合构建语义标注的3D地图，在室内环境中实现实时语义重建。Bescos等[46]开发了DynaSLAM，通过Mask R-CNN实例分割与多视几何验证检测动态对象。Rosinol等[52]引入Kimera，一个开源库支持实时度量-语义定位与映射，通过GTSAM优化构建语义网格，在多机器人系统中实现全局一致性。Runz等[47]提出MaskFusion，通过Mask R-CNN分割动态对象并跟踪其运动，在动态环境中构建分离的静态与动态地图。Bowman等[53]开发了Semantic SLAM框架，通过概率数据关联链接多视点语义观测，提升动态环境中的鲁棒性。Yu等[48]设计了DS-SLAM，扩展ORB-SLAM2通过SegNet语义分割与运动一致性检查过滤动态特征，提升高动态环境中的位姿估计精度并构建稠密语义地图。

3 基于视觉惯性融合的自主定位

视觉惯性定位系统通常由初始化、状态建模与估计等关键环节构成，其核心目标是在缺乏稳定外部绝对观测的条件下，实现位姿与环境状态的持续、可靠估计。本节首先围绕视觉惯性系统的初始化与状态建模方法展开论述，并进一步介绍视觉惯性融合定位与感知增强相关技术的研究进展。

3.1 视觉惯性系统初始化与状态建模

3.1.1 动态初始化与尺度重构

运动状态初始化的核心是在短时间窗内联合确定后端优化可用的初值，包括初始重力、速度、尺度与零偏等状态参数[54]。对单目系统而言，视觉里程计只能得到尺度不确定的轨迹$\bar{\boldsymbol{p}}_i$，真实位置满足$\boldsymbol{p}_i = s\bar{\boldsymbol{p}}_i$；初始化即通过IMU预积分约束在窗口内求解尺度$s$、重力$\boldsymbol{g}$、速度$\boldsymbol{v}_i$与IMU零偏$(\boldsymbol{b}_g, \boldsymbol{b}_a)$，典型的视觉-惯性联合初始化问题可表述为带鲁棒核的最小二乘问题，即

$$\min_{s,\boldsymbol{g},\{\boldsymbol{v}_i\},\boldsymbol{b}_g,\boldsymbol{b}_a}\sum_{(i,j)}\rho\left(\left\|\boldsymbol{r}_{ij}^{\mathrm{imu}}(s,\boldsymbol{g},\boldsymbol{v},\boldsymbol{b})\right\|^2\right)+\sum_{k}\rho\left(\left\|\boldsymbol{r}_{k}^{\mathrm{vis}}(s)\right\|^2\right) \tag{3}$$

式中：$\boldsymbol{r}_{ij}^{\mathrm{imu}}$来自关键帧间预积分相对运动$(\Delta\boldsymbol{R}_{ij}, \Delta\boldsymbol{v}_{ij}, \Delta\boldsymbol{p}_{ij})$并与重力、零偏参数耦合，$\boldsymbol{r}_{k}^{\mathrm{vis}}$则来自视觉重投影约束[54]。如图7所示，经典系统VINS-Mono在滑窗优化框架下将预积分因子与重投影因子紧耦合，并采用视觉结构恢复+惯性对齐的两阶段思路完成尺度与重力一致化，成为单目VINS初始化的代表性基线[54]；ORB-SLAM3则在多地图与重定位能力之上进一步融合IMU原始数据并融入全局图优化以提升整体鲁棒性[55]。相较之下，纯视觉ORB-SLAM2虽具备较强的闭环与跟踪能力，但在快速机动与模糊条件下更易因缺少惯性约束而出现早期失锁与漂移积累，侧面凸显尺度、重力可观测对拒止场景视觉惯性稳健融合的基础作用[56]。

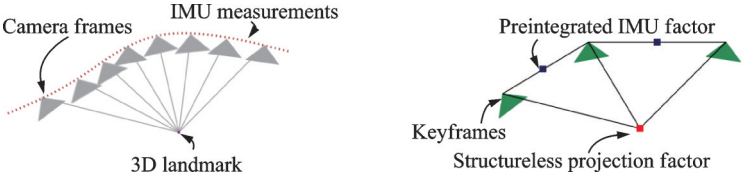

图7 视觉惯性里程计的观测组织与因子图建模示意图[54]

Fig.7 Schematic diagram of observation organization and factor map modeling for visual inertial odometry[54]

如何快速恢复单目视觉尺度信息一直是众多视觉惯性定位系统研究的重点，近年对于视觉尺度恢复的方法可归纳为以下3类：

(1) 学习先验注入尺度。Merrill等[57]将单帧深度网络输出的仿射不变深度表示为全局尺度和偏移

的低维参数,并与IMU预积分联合求解,再结合RANSAC等鲁棒估计策略抑制错误先验,从而在弱激励或小视差下显著提升启动成功率与收敛速度,进一步降低对显式三角化稳定性的依赖,缓解初始化锁死[58]。

(2)几何要素增强约束。Liu等[59]将点、线特征与IMU约束统一到闭式初始化求解中,在低纹理或点特征稀疏时增加可用几何信息,提高初始化稳定性与效率。

(3)度量观测直接确定尺度。在高空、小基线或弱激励导致尺度不可观时,双目或其他测距手段可直接提供度量尺度[60]。Hu等[61]在MSCKF中融合一维测距信息,结合延迟初始化与外参在线标定为视觉三角化特征提供深度锚点,提升尺度稳定性;Alberico等[62]提出结构无关的视惯框架,使用先验尺度持续约束系统抑制尺度漂移。此外,BIT-VIO以高帧率二值特征与融合更新持续校正尺度,减缓早期漂移扩散[63];DOGE先稳定外参旋转与陀螺零偏等慢变量再进入尺度和重力恢复,降低误初始化概率[64]。

3.1.2 系统状态建模与误差传播

状态空间建模核心在于位置、速度、姿态以及零偏等状态如何参数化,以及在离散时间上如何传播均值与不确定度,这些关键因素直接影响VIO在观测稀疏场景下的稳定性与一致性。

为高效融合高频IMU与低频视觉,Forster等[65]在SO(3)上提出流形预积分,将两关键帧间惯性信息压缩为$(\Delta R,\Delta v,\Delta p)$,并给出协方差递推及一阶bias校正。DRI-VINS通过解耦右不变误差传播与特征处理降低不一致与传播开销,使误差演化更稳定[66];FEJ2进一步显式建模,将其不确定度注入更新以抑制信息虚增[67]。在实现层面,R-VIO2采用robocentric状态空间与平方根信息增量求解,并联合在线时空标定以提升数值稳定性与抗漂移能力[68];随机克隆平方根信息滤波器面向边缘设备进一步强调稳定传播与快速更新[69]。为统一解释不同表示带来的差异,DES框架解耦状态/误差表征以缓解线性化一致性问题[70];观测器视角工作则将预积分与PEBO联系起来,为传播机制提供统一解释[71]。

相较于传统视觉惯性系统状态估计建模方法,深度学习技术在VIO中更常以可度量的先验或者量测增强出现,而非端到端替代,表4列举了部分代表性工作。网络负责补足几何信息难以精确刻画的部分,并输出不确定性,以便后端加权,从而在退化观测与传感器漂移时维持可用性与鲁棒性[72]。统一地,可将网络输出写成带协方差的观测(或先验)并注入EKF或者因子图进行优化,即

$$\hat{\boldsymbol{z}}_\theta,\boldsymbol{\Sigma}_\theta=f_\theta(\boldsymbol{I},\boldsymbol{u}),\ \min_{\boldsymbol{x}}\|\boldsymbol{r}_{\text{geom}}\|^2_{\boldsymbol{\Sigma}^{-1}_{\text{geom}}}+\|\boldsymbol{r}_{\text{imu}}\|^2_{\boldsymbol{\Sigma}^{-1}_{\text{imu}}}+\|\boldsymbol{z}(\boldsymbol{x})-\hat{\boldsymbol{z}}_\theta\|^2_{\boldsymbol{\Sigma}^{-1}_\theta} \tag{4}$$

式中$\boldsymbol{\Sigma}_\theta$(或置信度)决定学习信息对状态更新的影响强弱:不确定度大则少信一点,不确定度小则多用一点,从机制上避免网络输出在分布变化或局部退化时把系统拉偏。

在学习输出可写成可加权因子的框架下,相关研究可按注入对象归纳为3类。其一是学习几何紧凑表示并显式给出可信度,例如CodeVIO用轻量CVAE将稠密深度压缩为“深度码”,同时预测不确定度并转为协方差,使学习深度以可加权量测形式进入紧耦合估计,既补充几何约束又便于后端抑制错误先验[74]。其二是学习难建模误差并形成闭环校正,Adaptive VIO在线学习视觉对应置信度与IMU零偏,在可微预积分与因子图中去偏或加权,并用优化结果反向自监督更新以实现跨环境持续适应,如图8所示[73];类似地,深度IMU bias推断将偏置先验注入因子图以增强鲁棒性[72]。其三是学习观测可用性与不确定性来驱动门控或加权,CUAHN-VIO输出单应量测及方差并在EKF中自适应调节权重以提升退化场景稳定性,高效深度VIO用策略网络决定是否启用视觉编码器以在精度与算力间折中;IMO将学习到的相对位移作为量测注入惯性滤波闭环以抑制漂移,而SelfVIO展示了自监督端到端回归的替代路径[75]。

表4 深度学习辅助的系统建模与不确定性估计方法

Table 4 Deep learning-assisted system modeling and uncertainty estimation methods

代表工作	学习输出	不确定性形式	典型适用退化
CodeVIO[74]	深度不确定度图	方差/协方差	弱纹理、小视差、深度不稳定
Adaptive VIO[73]	视觉深度及其不确定度，IMU零偏信息	方差/协方差	观测间歇、光照或场景变化、传感器漂移
Deep IMU bias inference[72]	IMU零偏	先验方差	视觉退化、遮挡、低纹理、长时间纯惯性段
CUAHN-VIO[75]	视觉深度及其不确定度	方差/协方差	动态干扰、运动模糊、非平面场景、弱纹理
自适应视觉模态选择[76]	视觉开关策略（是否启用视觉编码器）	隐式表征	高速/高角速度、算力受限、视觉信息贡献不稳定
IMO[77]	相对位移	方差/协方差	高速竞速、强模糊、极低纹理、视觉失效段
SelfVIO[83]	端到端位姿和视觉深度	隐式表征	标定不准、训练分布覆盖较好时
深度先验[78]；RAFT/RAFT-Stereo[79]；DROID-SLAM[80]；DPVO[81]；SAM[82]	视觉匹配对，语义掩膜以及深度信息等	光流/深度显式，部分隐式表征	弱纹理、重复纹理、动态目标、光照

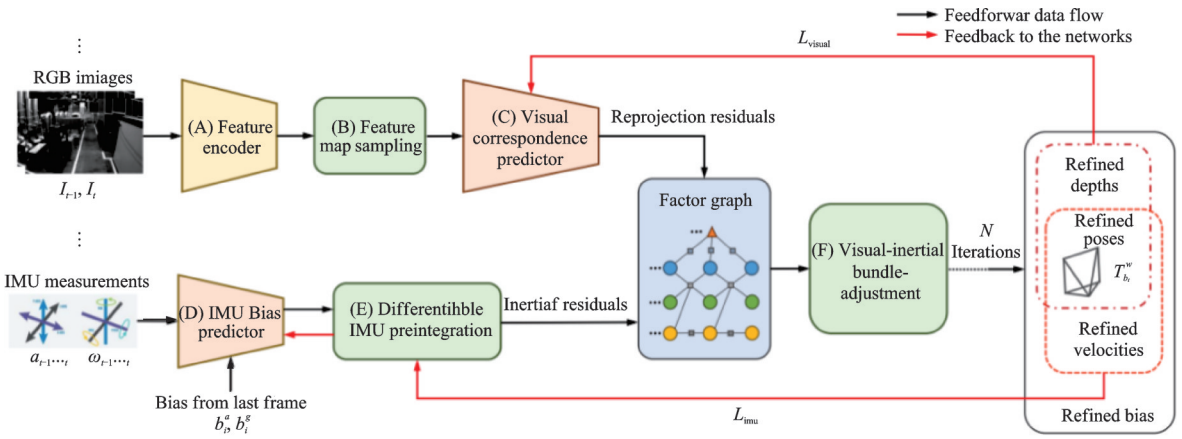

图8 学习辅助视觉惯性里程计的闭环优化框架示意图[73]

Fig.8 Schematic diagram of the closed-loop optimization framework for learning-assisted visual inertial odometry[73]

与此同时，另一条常被并行采用的路径是用学习增强视觉观测质量，再交给几何后端去融合：深度模型为单目几何与尺度提供更强先验[78]；RAFT/RAFT-Stereo、DROID-SLAM与DPVO等方法则提升稠密约束与关联建模能力，在弱纹理/快速运动下提供更稳定的前端输入[79]；语义工具（Segment anything、grounding DINO）可帮助剔除动态或低置信区域，降低错误观测污染[82]。这类模块若能输出可信度，同样更适合以协方差形式接入后端加权融合。

综上所述，深度学习辅助的系统建模与不确定性估计方法这一方向的关键不在于让网络替代VIO，而在于把学习输出变成可解释、可控、可加权的先验或量测，并用不确定性将其稳妥地嵌入状态估计链条，使系统在退化与拒止条件下仍能保持稳定的误差传播与融合更新。

3.2 视觉惯性状态估计与感知增强

3.2.1 视觉退化场景下鲁棒融合与特征增强策略

复杂低空环境中,弱纹理、强光照变化、运动模糊、低照度与动态干扰等退化因素会直接削弱视觉测量的可用性,使得视觉约束变弱、惯导误差累积最终导致估计器发散。与追求单帧最优精度不同,退化场景更关注在可接受误差范围内维持连续输出:一方面尽可能提升观测质量,另一方面在融合层显式建模观测可信度,并通过鲁棒代价或自适应权重抑制异常测量的影响,如图9所示[84]。

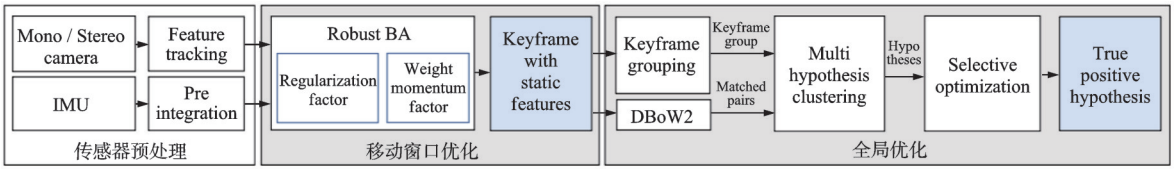

图9 动态环境下视觉-惯性SLAM的鲁棒处理框架[84]

Fig.9 Robust processing framework for vision-inertial SLAM in dynamic environments[84]

在滤波式框架中,可将退化影响统一到测量噪声或权重自适应上。以线性化后的量测更新为例,令残差为r_k、雅可比为H_k、测量噪声为R_k,则更新可写为以下形式

$$\begin{cases} K_k = P_{k|k-1}H_k^{\mathrm{T}}\left(H_k P_{k|k-1}H_k^{\mathrm{T}} + R_k\right)^{-1} \\ \delta x_k = K_k r_k \\ P_{k|k} = (I - K_k H_k)P_{k|k-1}(I - K_k H_k)^{\mathrm{T}} + K_k R_k K_k^{\mathrm{T}} \end{cases} \tag{5}$$

事件-帧联合VIO进一步将对齐不确定性显式转化为滤波权重,自适应调整更新强度,使高速机动与强光照变化下仍能稳定输出位姿约束[85]。在后端一致性层面,通过更一致的线性化与误差传播抑制退化运动下的伪信息增益,也能降低发散风险。

在优化式滑窗中,退化常表现为外点增多与约束稀疏,可将鲁棒性归结为鲁棒核和结构/先验补约束。典型目标函数可概括为

$$\min_{\mathcal{X}} \left\| r_{\mathrm{imu}}(\mathcal{X}) \right\|^2_{\Sigma^{-1}_{\mathrm{imu}}} + \sum_i \rho\left(\left\| r_{\mathrm{vis},i}(\mathcal{X}) \right\|^2_{\Sigma^{-1}_{\mathrm{vis},i}} \right) \tag{6}$$

式中:$\rho(\cdot)$为Huber或Cauchy等鲁棒核,$\Sigma_{\mathrm{vis},i}$亦可由特征质量或退化检测结果自适应调整。动态干扰是最常见的系统性外点来源之一,通过语义与运动一致性识别动态区域并在优化中降低其贡献,可以显著减少错误几何约束对轨迹的污染,从而提升动态场景稳定性[84]。当纹理不足或视差退化导致几何约束稀疏时,引入可重复的结构先验也可增强可观测性;例如显式利用平面正则将平面-特征-位姿的一致性纳入滑窗,可在室内走廊、地下空间等弱纹理环境中抑制尺度与姿态漂移[85]。在资源受限且图像质量波动的场景下,直接在传感器/特征层面降低不稳定性同样有效,例如焦平面二值特征与IMU紧耦合能以更低计算开销维持可重复观测,从而改善低照度、模糊下的实时性与连续性[63,86]。

当可见光严重退化甚至接近不可用时,多模态观测提供了更物理一致的兜底路径:事件相机在高动态与HDR条件下具备优势,点线联合的事件VIO、纯事件双目VIO及其体素地图管理等工作,分别从结构约束增强、双目事件关联与抗噪地图更新等角度提升退化场景鲁棒性;热成像在弱光/烟雾等条件下成像更稳定,将其与事件VIO融合可在极端退化中补齐观测约束[87];进一步地,毫米波雷达等非视觉几何传感器可在视觉几乎失效的边界场景中与IMU紧耦合,提供速度/匹配残差约束以限制漂移扩散[88]。

在特征增强层面,学习型匹配与稠密运动估计为退化条件下的观测供给提供了新选择。Transformer匹配在弱纹理或重复纹理中更易获得稳定对应[89],自监督表征可提升跨光照与外观变化下的特征鲁棒性[50],稠密光流/双目估计可在局部特征不足时补充连续运动约束,而深度SLAM/VO在弱纹理

与快速运动下展示了更强的前端稳健性，具备作为VIO前端替换或增强模块的潜力[80]（表5）。总体而言，视觉退化下的可靠融合更像质量感知的约束管理：先把退化显式转化为可调权重，再用鲁棒核、结构先验与多模态观测补足约束密度，并辅以学习特征提升可用测量的下限，从而把能运行转化为可控地运行。

表5 典型视觉退化类型、增强策略与代表性工作汇总

Table 5 Summary of typical visual degradation types, enhancement strategies, and representative works

退化或失效因素	直接后果	典型策略	代表工作
动态干扰（人车、临时静止物体）	错误重投影约束污染后端	动态特征抑制＋鲁棒BA优化	DynaVINS[84]
弱纹理、低视差、结构单一	约束稀疏、尺度和姿态漂移	引入平面、线等结构先验，提升可观测性	平面正则VIO[85]
高动态、运动模糊	帧特征匹配不稳、外点增多	事件观测补偿、不确定性驱动的自适应调权	自适应事件－帧VIO[90]
低照、烟雾等可见光严重退化	视觉观测接近不可用	热成像或事件多模态互补增强	TEVIO[87]
视觉几乎不可用（尘雾雨雪、纯暗）	纯VIO崩溃、漂移快速累积	非视觉传感器兜底（雷达等）＋惯导	4D雷达辅助惯导[88]
算力受限和前端特征退化	提取匹配成本高且不稳定	低功耗、二值特征与传感器内计算	BIT-VIO[63]

3.2.2 基于语义与几何先验约束的鲁棒定位

在拒止环境中，视觉前端常因弱纹理、动态干扰或视差不足而难以持续提供稳定几何约束，此时“语义信息＋结构化几何先验”的价值在于为状态估计引入更稳定、可解释的锚点。语义用于筛除不可信观测并强调与任务相关的结构，几何先验（深度、平面、测距、激光几何主干等）用于补足尺度与结构约束，从而把可用信息显式转化为后端可优化的约束项，如图10所示[90]，其关键并非简单叠加模块，而是将先验以统一残差形式写入后端，并对其不确定性进行建模，避免在弱观测条件下被系统过度相信而诱发漂移或不一致。

从统一建模角度，融合导航可写成带先验项的非线性最小二乘问题，即

$$\begin{aligned}&\min_{\mathcal{X}} \mathcal{J}(\mathcal{X}) = \mathcal{J}_{\mathrm{VIO}}(\mathcal{X}) + \lambda_g \left\| r_{\mathrm{geo}}(\mathcal{X}) \right\|^2 + \lambda_m \left\| r_{\mathrm{map}}(\mathcal{X}) \right\|^2 \\ &\mathcal{J}_{\mathrm{VIO}}(\mathcal{X}) = \left\| r_{\mathrm{imu}}(\mathcal{X}) \right\|^2_{\Sigma^{-1}_{\mathrm{imu}}} + \sum_i \rho\left(\left\| r_{\mathrm{vis},i}(\mathcal{X}) \right\|^2_{\Sigma_{\mathrm{vis},i}} \right) + \lambda_s \sum_i w_i \left\| r_{\mathrm{vis},i}(\mathcal{X}) \right\|^2\end{aligned} \tag{7}$$

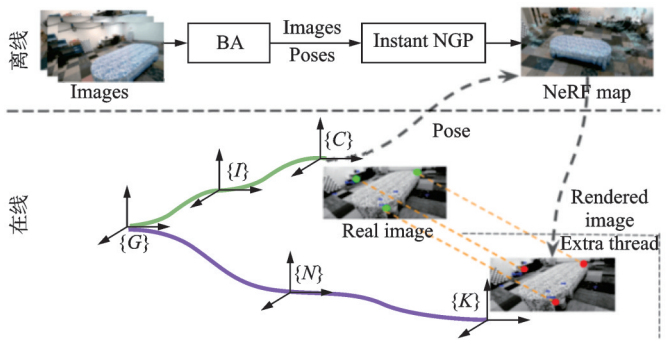

图10 基于NeRF地图的视觉惯性导航框架[90]

Fig.10 A visual-inertial navigation framework based on NeRF maps[90]

式中:$\mathcal{X}$为窗口内位姿、速度、偏置等状态;$\rho(\cdot)$为鲁棒核;r_{geo}表示结构化几何先验残差(深度/平面/测距/激光等);w_i表示语义与质量评估得到的视觉残差权重;r_{map}表示地图/地理参照或隐式地图带来的重定位与漂移约束。该形式的直接好处是可以将不同来源的信息以"残差-权重-协方差"的方式进入同一后端,便于分析先验对可观测性与稳定性的贡献。

结构化几何先验常用于补偿单目尺度与弱视差退化。随着稠密预测Transformer推动深度质量提升,具备更强泛化与度量尺度能力的单目深度模型可作为结构先验来源,并已被用于加速单目视惯初始化、在短轨迹与弱纹理条件下增强尺度可观测性[88]。相比学习深度这类软先验,少量可靠测距量测往往更接近硬几何锚点。

语义信息更适合承担观测选择与可信度调度的角色,其通过分割/检测将动态物体、反光区域或低可信区域从几何约束中剔除或降权,避免伪约束污染后端。提示式分割可快速给出可控的前景/背景掩膜,用于隔离动态对象与低可信区域[82];开放词汇检测有利于在复杂拒止环境中选择更稳定的结构性目标(如建筑轮廓、道路边界、地标区域等)[89]。这些语义结果可用于对观测值的精确筛选与标记从而提高弱纹理与强外观变化下的有效匹配数量与稳定性[92]。

地图与重定位先验提供的是更长期的漂移约束:地理参照信息可作为全局位置/航向约束写入后端,面向大尺度航迹的漂移可控需求,更进一步的思路是将隐式神经地图作为可优化的观测来源:NeRF用连续隐式场表达外观与几何[93],并借助Fourier特征增强高频表示能力;在已知NeRF地图的前提下,可通过反向优化渲染误差进行位姿估计与重定位,并在视惯框架中与IMU运动约束结合,将隐式地图一致性转化为可用观测以抑制长航程漂移[91]。

4 基于多源信息融合的自主定位

4.1 多源融合定位框架设计

多源融合定位技术通过集成视觉、惯性、激光雷达以及卫星等异构信息源实现优势互补,提升复杂环境下的定位精度与可靠性。本节从融合框架设计与具体算法进展两个层面,系统梳理多源融合自主导航技术的发展脉络。

4.1.1 基于滤波/优化的多源融合定位

根据状态估计方法的不同,多源融合框架可分为基于滤波与基于优化两大类,二者在计算效率、估计精度与系统复杂度方面各有权衡,如图11所示。

基于滤波的方法以EKF及其变体为核心,采用递归贝叶斯估计框架,在每个时刻融合当前观测更新状态估计。其优势在于计算复杂度低、实时性好,适用于资源受限平台。经典工作包括:Mourikis等[94]提出的多状态约束卡尔曼滤波(MSCKF),通过在状态向量中维护滑动窗口内的相机位姿,实现了视觉惯性里程计的高效求解;Bloesch等[95]提出的ROVIO采用直接法光度误差与EKF结合,避免了特征提取开销。然而,滤波方法受马尔可夫假设限制,难以充分利用历史观测的约束关系,且线性化误差随时间累积。基于优化的方法将状态估计建模为非线性最小二乘问题,通过联合优化滑动窗口内的所有状态与观测残差,获得全局一致的最优估计。因子图(Factor graph)为此类方法提供了统一的数学框架[96],将先验、IMU预积分、视觉重投影、LiDAR配准等约束表示为因子节点,通过高斯-牛顿或列文伯格-马夸尔特算法迭代求解。代表性工作如OKVIS[97]首次将关键帧优化引入视觉惯性系统;VINS-Mono[54]采用滑动窗口优化与边缘化策略,结合回环检测实现全局一致建图。优化方法精度更高,但计算负担较重,需借助增量求解器(如iSAM2[98])提升实时性。

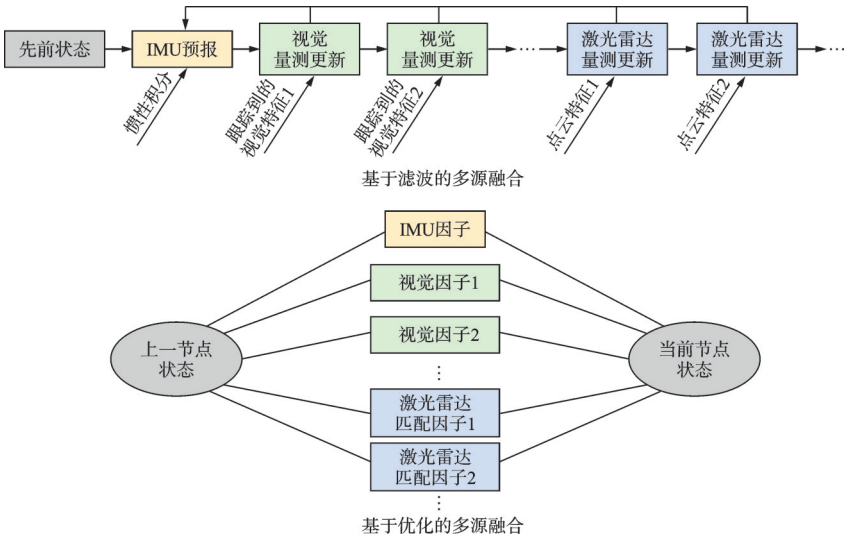

图11 滤波和因子图优化方法对比

Fig.11 Comparison of filtering and factor graph optimization methods

4.1.2 激光雷达/惯性/视觉融合方法

在复杂环境与GNSS不可用或受限条件下，单一传感器难以同时满足无人机自主导航对精度、鲁棒性与连续性的综合需求[99-101]。视觉传感器具有信息密度高、易获取环境结构与语义特征等优势，但在弱纹理、强光照变化及高速运动条件下易发生退化；惯性测量单元(IMU)具备高频、短时精度高的特点，可为姿态与运动提供连续约束，但其误差随时间累积不可避免；激光雷达则能够直接获取尺度一致的三维几何信息，对光照变化不敏感，在结构化或半结构化场景中表现出较高的测距精度与稳定性。将视觉、惯性与激光雷达进行深度融合，可在时域、空间域与信息层面形成优势互补，通过高频运动约束、稠密几何感知与丰富外观信息的协同作用，实现在复杂拒止环境下更高精度、更强鲁棒性的自主导航定位。

激光/惯性SLAM方法融合LiDAR与IMU，利用LiDAR的高精度测距与IMU的运动先验实现鲁棒定位。早期工作如LOAM[102]提出了边缘-平面特征提取与两阶段优化策略，成为后续研究的重要基线。在紧耦合LIO方面，LIO-SAM[103]采用因子图框架融合IMU预积分、LiDAR里程计、GPS及回环约束，实现了高效的增量平滑与建图。FAST-LIO[104]提出了基于迭代卡尔曼滤波(IEKF)的紧耦合方案，通过增量KD树实现高效点云配准；FAST-LIO2[105]引入增量体素地图(iKD-Tree)进一步提升效率，支持稀疏特征环境下的稳定运行。Faster-LIO[106]采用增量体素化(iVox)替代KD树，在保持精度的同时显著降低计算开销。

激光/惯性/视觉SLAM进一步融合相机、LiDAR与IMU三类传感器，综合利用视觉的纹理信息、LiDAR的几何精度与IMU的动态响应[107-110]。LVI-SAM[111]在LIO-SAM基础上引入视觉惯性子系统，通过因子图实现多传感器紧耦合。王铉彬等[107]在原始观测值层面将点云特征、IMU量测以及视觉特征信息进行紧耦合进一步提升了室外大尺度场景中定位建图的精度与一致性。R3LIVE[108]采用IEKF框架融合LiDAR、视觉与IMU，并构建带颜色的稠密点云地图。FAST-LIVO[112]将直接法视觉里程计与FAST-LIO2结合，通过光度误差与几何误差联合优化提升鲁棒性。

综上所述，目前在视觉/惯性/激光雷达融合导航方面形成两条代表性技术路线，一条是基于因子图优化的LVI-SAM框架(图12[111])，一条是基于迭代卡尔曼滤波的FAST-LIVO框架(图13[106])。其中LVI-SAM框架代表了从几何约束向多模态协同演进的技术迭代，其技术发展脉络可追溯至

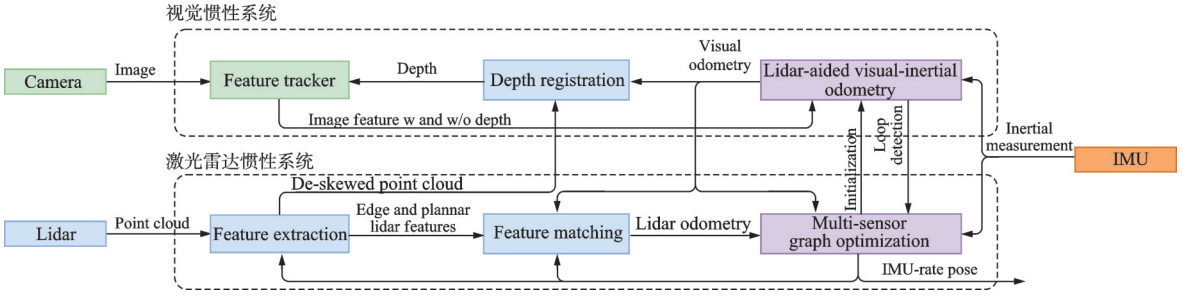

图12 LVI-SAM算法框架[111]

Fig.12 LVI-SAM algorithm framework[111]

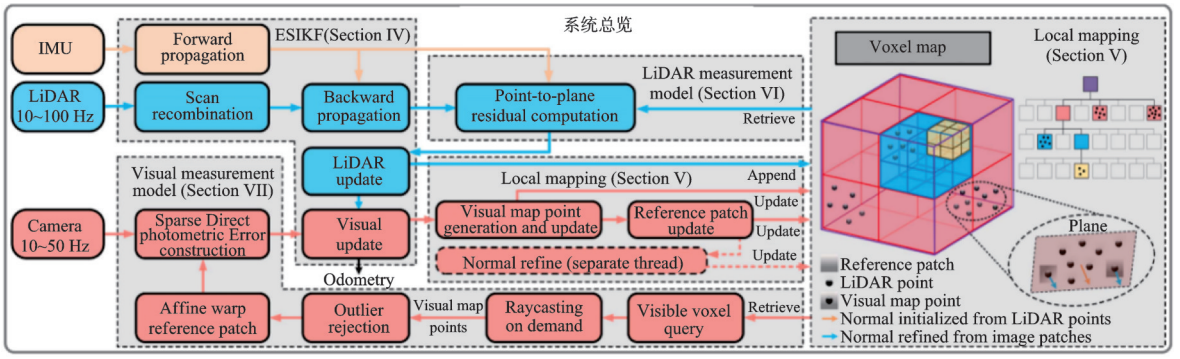

图13 Fast-LIO算法框架[106]

Fig.13 Fast-LIO algorithm framework[106]

LOAM[14]对激光点云边缘与平面特征的提取大幅提升激光雷达SLAM算法效率,历经LeGO-LOAM对地面信息的轻量化处理,最终在LIO-SAM[15]中确立了以iSAM2为核心的因子图后端。LVI-SAM[85]在此基础上通过视觉惯性(VINS)与激光惯性(LIO)双子系统的深度耦合,利用激光提供的厘米级深度图解决视觉尺度漂移,并通过因子图将IMU预积分、视觉残差、激光配准以及回环约束统一于全局一致性优化框架下,展现了退化场景的鲁棒定位建图能力。而基于迭代卡尔曼滤波(IEKF)的FAST-LIVO框架,其演化逻辑更强调算法的极致效率与动态响应。该路线始于FAST-LIO[106]对高效状态估计的探索,在FAST-LIO2[17]中通过引入增量式空间数据结构iKD-Tree,实现了对原始点云的直接处理并消除了特征提取的计算开销。作为该系列的最新演进版本,FAST-LIVO[112]进一步将直接法视觉里程计与流形上的IEKF深度整合,通过在滤波更新步骤中同时最小化视觉光度误差与激光几何误差,使得系统在激光特征退化或视觉纹理匮乏的极端场景下,仍能凭借多源残差的互补约束维持高频、鲁棒的状态输出。此外,诸多工作进一步将激光雷达、惯性、视觉信息与卫星观测融合,通过利用卫星机会信号提升复杂低空场景中多源融合系统的无漂定位建图能力[113-116],有效拓展了智能移动载体在多元化应用场景中的连续、无缝导航定位能力。

4.2 拒止环境下的多源鲁棒定位

尽管多源融合框架在常规环境下已取得优异性能,但在高动态、光照剧变、几何退化等挑战性场景中,系统鲁棒性仍面临严峻考验。提升多源融合系统鲁棒性可从两个层面入手,一方面通过学习增强方法提升特征提取与匹配的鲁棒性,改善视觉、激光雷达等传感器退化场景下的定位稳健性;另一方面可以通过完好性监测与故障隔离保障导航输出可信性。

4.2.1 基于学习的感知增强

传统几何方法(如ORB、SIFT特征、ICP配准)在理想条件下表现优异,但在光照剧变、弱纹理、动态遮挡等极端场景下性能急剧下降。深度学习的引入为感知层鲁棒性提升开辟了新途径。

在特征检测方面,传统手工设计特征对光照、视角变化敏感。学习方法通过数据驱动的方式获得更鲁棒的特征表示。Detone等[16]提出的SuperPoint采用自监督学习训练特征检测与描述联合网络。Sarlin等[19]提出的SuperGlue采用图神经网络(GNN)与注意力机制建模特征间的上下文关系。Sun等[22]提出的LoFTR采用Transformer架构实现无检测器的稠密匹配,在弱纹理区域表现尤为突出。尺度不确定是单目视觉SLAM算法中存在的固有局限,在现有算法中往往通过增加IMU传感器等方式为视觉SLAM提供尺度信息,但基于学习的方法学习方法可从单张图像预测稠密深度,为单目系统提供尺度先验。Yang等[110]提出的Depth Anything利用大规模无标签数据进一步提升泛化能力。

在LiDAR感知领域,通过将语义分割技术引入SLAM框架,识别场景中的动态物体(车辆、行人)并将其从配准中排除,提升里程计鲁棒性。Milioto等[117]提出的RangeNet++将点云投影为距离图像,采用2D卷积进行高效分割;Zhu等[118]提出的Cylinder3D设计圆柱形体素表示更好保留点云几何特性;Aygun等[119]提出的4D-PLS利用时序信息区分静态与动态物体。Chen等[120]提出的SuMa++将语义分割集成到SLAM系统中。在学习点云配准方面,Wang等[121]提出的DCP采用注意力机制学习点对应关系;Yew等[122]提出的RPM-Net通过深度学习预测软对应权重。

综上所述,基于学习的感知增强已在多个任务上取得显著进展,但仍面临以下挑战:(1)泛化性不足,如训练域与部署域差异导致性能下降,需要域自适应或持续学习策略;(2)实时计算效率低,Transformer等架构计算开销大,需轻量化设计或专用硬件加速;(3)可解释性问题,深度模型的黑盒特性不利于安全关键应用的认证。

4.2.2 完好性监测与故障隔离

在无人机低空飞行等安全关键应用中,仅提升定位精度并不足以保障系统安全性[123]。相比误差有多小,完好性(Integrity)更关注定位结果是否可信、何时不可信,其本质是对导航解可靠程度的概率化度量。典型完好性框架通过告警限(Alert limit,AL)、保护级(Protection level,PL)、漏检率(Probability of missed detection,PMD)与虚警率(Probability of false alarm,PFA)对系统性能进行约束。其中,PL刻画在给定统计假设下定位误差的上界,当PL < AL时系统被认为是可用的;反之则需触发告警并进入降级或重构模式,以防止错误导航信息被继续使用。

在卫星定位领域,接收机自主完好性监测(Receiver autonomous integrity monitoring,RAIM)构成了完好性理论与工程实现的基础[124]。RAIM利用冗余卫星观测对定位解内部一致性进行检验。在最小二乘框架下,设观测模型为

$$\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}+\boldsymbol{v} \tag{8}$$

其估计解为

$$\hat{\boldsymbol{x}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y} \tag{9}$$

残差向量为

$$\boldsymbol{r}=\boldsymbol{y}-\boldsymbol{H}\hat{\boldsymbol{x}} \tag{10}$$

反映了观测之间的一致性。基于残差构造的加权残差平方和(Weighted sum of squared errors,WSSE)为

$$\mathrm{WSSE}=\boldsymbol{r}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{r} \tag{11}$$

在无故障假设下服从自由度为$n-4$的卡方分布。通过设定虚警率P_{FA}可确定检测阈值T_{D},当$\mathrm{WSSE} > T_{\mathrm{D}}$时判定存在异常。

随着导航系统向多传感器融合演进,完好性监测需同时应对传感器异构性与误差时间相关性问

题[125]。在多传感器融合定位系统中,状态误差随时间传播,这使得保护级不再仅由单历元观测决定,而需综合考虑历史信息对当前状态不确定性的累积影响。相关研究在此基础上构建了时间相关的保护级计算方法,从而避免在高动态或弱观测条件下对系统安全性产生过度乐观的评估。

多传感器交叉验证是融合系统中提升完好性的重要途径,其思想是对来自不同传感器的独立估计进行一致性检验。

在系统架构层面,联邦滤波为多源系统的完好性管理提供了清晰的工程实现路径。各局部滤波器独立估计状态$\hat{\boldsymbol{x}}_i$及协方差$\boldsymbol{P}_i$,主滤波器在信息域内进行融合,其等效融合结果可表示为

$$\boldsymbol{P}_m^{-1}=\sum_i \boldsymbol{P}_i^{-1},\ \hat{\boldsymbol{x}}_m=\boldsymbol{P}_m\sum_i \boldsymbol{P}_i^{-1}\hat{\boldsymbol{x}}_i \tag{12}$$

当某一局部滤波器通过残差检验或一致性检验被判定为异常时,可将其信息从融合过程中剔除,由其余子系统维持导航解输出,从而实现具备故障隔离能力的鲁棒运行。

完好性监测的有效实施依赖于对传感器退化状态的准确识别。以激光雷达为例,点云配准问题中构造的海森矩阵$\boldsymbol{H}=\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}$反映了各方向上的约束强度,其特征值分布可用于定量分析可观性退化。当满足阈值条件时,表明系统在某一方向上约束不足,易引发状态估计不稳定。将此类退化检测结果反馈至完好性监测与融合策略中,可实现对复杂环境下定位风险的前瞻性评估与主动规避。

综合上述方法,现代多源鲁棒定位系统正逐步形成“感知-估计-完好性”协同设计范式,使系统不仅能够在复杂、拒止环境中保持连续定位能力,更能够对自身可信度进行实时量化评估,为高安全等级自主系统提供坚实的理论与工程基础。

5 发展趋势与展望

面向复杂低空应用环境,现有无人机自主定位技术正经历从单一算法性能驱动向系统级能力构建的范式转变。与地面机器人不同,无人机在三维空间中高速运动、缺乏运动约束,且对时延、载荷与能耗高度敏感,其发展重点将从单纯追求定位精度,转向飞行可用性、任务连续性与系统可信性并重的综合能力构建。

在感知层面,高空俯视、地表纹理重复与快速姿态变化等典型飞行工况,使传统稀疏特征方法易出现匹配退化与尺度不稳定问题。未来研究将重点引入具备跨高度、跨视角泛化能力的视觉基础模型[126],结合单目深度与语义感知为可飞区域判定提供先验支撑。此外,面向低纹理、弱光照及烟雾遮挡等极端环境,多模态感知(可见光、红外、事件相机、激光雷达)协同将成为重要方向。如何在保证轻量化的前提下实现多模态信息互补,是未来感知层的重要研究问题。在融合层面,无人机在起飞、巡航、急转机动与悬停阶段呈现显著不同的动力学特性与观测可观测性。传统固定权重融合策略难以适应快速变化的飞行状态。未来融合框架需显式建模飞行工况变化,引入状态相关的动态权重重构机制[127],实现多源信息的自适应调度与异常观测抑制。进一步地,融合框架将从传感器级融合向任务级融合拓展,将任务约束、飞行意图与环境语义纳入状态估计过程,实现估计结果与任务目标之间的协同优化,从而提升系统整体稳定性与安全性。

在状态估计层面,无人机任务持续时间长、回环触发条件受限,单纯依赖局部优化难以保证全程稳定。基于地形匹配与高程梯度等先验地理信息进行漂移校正的方法已在公里级飞行中展现良好效果[128],未来将进一步探索“数据驱动+物理约束”的混合范式,在保持可解释性的前提下提升长航时可信度[129]。此外,针对GNSS间歇可用场景,研究如何在“可用-拒止”状态频繁切换条件下实现平滑过渡与误差界约束,将成为提升系统可信度的重要方向。

在系统层面,受载荷与功耗限制,面向飞行应用的导航算法将更加注重计算-感知-控制闭环的整体优化,通过固定规模估计、事件驱动感知、模型压缩与异构加速等手段实现轻量化部署[130]。更重要的是,随着任务复杂度提升,无人机导航将不再局限于为控制器提供位姿,而是向感知-导航-决策一体化

演进。大语言模型与视觉语言模型与无人机系统的深度整合已成为新兴热点，为无人机提供语义推理、任务分解与自适应规划等高层认知能力[131]。这一趋势的深化正推动无人机从飞行传感平台向具身智能体演进。具身智能强调通过与物理世界的持续交互实现感知-认知-行动闭环，这与无人机在三维动态环境中的核心挑战高度契合。多模态大语言模型与世界模型的联合架构已被认为是增强无人机自主性的关键路径[132]，其核心在于使无人机具备超越几何建图的语义感知、基于世界模型的风险预判以及面向任务的自适应行为生成能力。

总体而言，未来无人机自主导航将呈现飞行工况强相关、系统级深度融合与可信性优先的显著特征，导航系统将从单一算法提升迈向面向实际任务的工程化、智能化与可验证发展阶段。而随着具身智能应用的兴起，无人机有望成为继自动驾驶车辆和人形机器人之后，具身AI在三维开放环境中最具应用前景的物理载体之一，为低空经济、应急救援与智慧城市等领域的深层次智能化变革提供关键使能技术[133]。

6 结束语

复杂低空环境下，无人机自主定位面临GNSS不可用、视觉退化与惯性漂移叠加引发的持续性与可靠性挑战。本文围绕低空典型退化场景，系统综述了基于卫星地图的视觉匹配定位、视觉SLAM、视觉惯性融合以及多源异构融合导航等关键技术进展，分析了各类方法在不同退化场景下的优势与局限。综述表明，单一传感器已难以满足复杂环境下长期稳定定位需求，多模态信息深度融合、退化感知与一致性估计将成为提升系统可靠性的核心方向。未来的研究有望在数据驱动的自适应融合、语义与结构化先验约束、可信度评估以及集成化轻量化系统设计等方面取得突破，从而为复杂低空环境下无人机高可靠自主定位提供更完善的理论支撑与工程路径。

参考文献：

[1] KAPLAN E D, HEGARTY C J. Understanding GPS/GNSS: Principles and applications[M]. 3rd ed. Boston: Artech House, 2017.

[2] 史殿习, 刘聪, 佘馥江, 等. GPS拒止环境下基于定位置信度的多无人机协同定位方法[J]. 计算机科学, 2022, 49(4): 302-311.
SHI Dianxi, LIU Cong, SHE Fujiang, et al. Cooperative localization method for multiple UAVs based on positioning confidence in GPS-denied environments[J]. Computer Science, 2022, 49(4): 302-311.

[3] HOFMANN-WELLENHOF B, LICHTENEGGER H, WASLE E. GNSS-global navigation satellite systems: GPS, GLONASS, Galileo, and more[M]. Vienna: Springer, 2008.

[4] ZHANG Z, XU Y, CUI Q, et al. Unsupervised SAR and optical image matching using Siamese domain adaptation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5227116.

[5] TITTERTON D H. Strapdown inertial navigation technology[C]//Proceedings of Institution of Electrical Engineers. [S.l.]: [s.n.], 2004.

[6] 朱徐东. 不依赖卫星的集群无人机协同定位方法研究[D]. 南京:南京航空航天大学, 2023.
ZHU Xudong. Research on cooperative localization method for UAV swarm without satellite support[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2023.

[7] BENSON D O. A comparison of two approaches to pure-inertial and Doppler-inertial error analysis[J]. IEEE Transactions on Aerospace and Electronic Systems, 2007(4): 447-455.

[8] LUO H, LI G, ZOU D, et al. UAV navigation with monocular visual inertial odometry under GNSS-denied environment[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1001615.

[9] ZHENG Z, WEI Y, YANG Y. University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization[C]//Proceedings of the 28th ACM International Conference on Multimedia. [S.l.]: ACM, 2020: 1395-1403.

[10] DAI M, HU J, ZHUANG J, et al. A transformer-based feature segmentation and region alignment method for UAV-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(7): 389.

[11] SHI Y, LIU L, YU X, et al. Where am I? A novel benchmark for UAV localization using satellite imagery and deep convolutional neural networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. [S.l.]: IEEE, 2019: 1-10.

[12] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60 (2): 91-110.

[13] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

[14] ROSTEN E, DRUMMOND T. Machine learning for high-Speed corner detection[C]//Proceedings of the European Conference on Computer Vision. [S.l.]: [s.n.], 2006: 430-443.

[15] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT [C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l]: [s.n.], 2011: 2564-2571.

[16] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: Self-supervised interest point detection and description [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. [S.l.]: IEEE, 224-236.

[17] DUSMANU M, ROCCO I, PAJDLA T, et al. D2-Net: A trainable CNN for joint description and detection of local features [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2019: 8092-8101.

[18] REVAUD J, WEINZAEPFEL P, REZENDE C, et al. R2D2: Reliable and repeatable detector and descriptor [C]// Proceedings of Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2019: 12405-12415.

[19] SARLIN P-E, DETONE D, MALISIEWICZ T, et al. SuperGlue: Learning feature matching with graph neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 4938-4947.

[20] LINDENBERGER P, SARLIN P-E, POLLEFEYS M, LightGlue: Local feature matching at light speed [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2023: 17627-17638.

[21] ZHU H, KARPUR A, CAO B, et al. OmniGlue: Generalizable feature matching with foundation model guidance[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2024: 10658-10668.

[22] SUN J, SHEN Z, WANG Y, et al. LoFTR: Detector-free local feature matching with transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 8922-8931.

[23] CHEN H, LUO Z, ZHOU L, TIAN Y, et al. ASpanFormer: Detector-free image matching with adaptive span transformer [C]//Proceedings of the European Conference on Computer Vision. [S.l]: [s.n.], 2022: 20-36.

[24] XIE T, DAI K, LI K, et al. DeepMatcher: A deep transformer-based network for robust and accurate local feature matching[J]. Expert Systems with Applications, 2024, 237: 121361.

[25] WANG Y, HE X, PENG S, et al. Efficient LoFTR: Semi-dense local feature matching with sparse-like speed[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2024: 20012-20022.

[26] WANG X, LI H, GAO S, et al. HomoMatcher: Dense feature matching results with semi-dense efficiency by homography estimation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l]: AAAI, 2025: 1-9.

[27] EDSTEDT J, ATHANASIADIS I, WADENBÄCK M, et al. DKM: Dense kernelized feature matching for geometry estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l]: IEEE, 2023: 17765-17775.

[28] EDSTEDT J, SUN Q, BÖKMAN G, et al. RoMa: Robust dense feature matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2024: 19790-19800.

[29] XU W, YAO Y, CAO J, et al. UAV-VisLoc: A large-scale dataset for UAV visual localization[J]. arXiv preprint arXiv: 2405.11936, 2024.

[30] SHI Y, LIU L, YU X, et al. VIGOR: Cross-view image geo-localization beyond one-to-one retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 3640-3649.

[31] HUI T, XU Y, ZHOU Q, et al. Cross-viewpoint template matching based on heterogeneous feature alignment and pixel-wise consensus for air- and space-based platforms[J]. Remote Sensing, 2023, 15(9): 2426.

[32] MUGHAL S, KANWAL M, ZAFAR M S, et al. A complete end-to-end trainable architecture for UAV localization[C]// Proceedings of the International Conference on Frontiers of Information Technology. [S.l.]: [s.n.]:, 2022: 1-6.

[33] CUI J, ZHANG J, LI X. A single-stage image retrieval method for UAV localization[J]. IEEE Transactions on Geoscience and

Remote Sensing, 2023, 61: 5603112.

[34] WANG G, CHEN J, DAI M, et al. WAMF-FPI: A weight-adaptive multi-feature fusion network for UAV localization[J]. Remote Sensing, 2023, 15(4): 910.

[35] CHEN J, ZHENG E, DAI M, et al. OS-FPI: A coarse-to-fine one-stream network for UAV geolocalization[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 7852-7866.

[36] HE Y, CHEN F, CHEN J, et al. DCD-FPI: A deformable convolution-based fusion network for unmanned aerial vehicle localization[J]. IEEE Access, 2024, 12: 129308-129318.

[37] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. [S.l.]: IEEE, 2007: 225-234.

[38] 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.

[39] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM[C]//Proceedings of the European Conference on Computer Vision. [S.l.]: [s.n.], 2014: 834-849.

[40] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(3): 611-625.

[41] TATENO K, TOMBARI F, LAINA I, et al. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 6243-6252.

[42] WANG S, CLARK R, WEN H, et al. DeepVO: Towards end-to-end visual odometry with deep recurrent convolutional neural networks[C]//Proceedings of the IEEE International Conference on Robotics and Automation. [S.l]: IEEE, 2017: 2043-2050.

[43] ZHU Z, PENG S, LARSSON V, et al. NICER-SLAM: Neural implicit scene encoding for RGB SLAM[C]//Proceedings of the International Conference on 3D Vision. [S.l.]: [s.n.]: 1-10.

[44] WANG S, XIE Y, CHANG C-P, et al. Uni-SLAM: Uncertainty-aware neural implicit SLAM for real-time dense indoor scene reconstruction[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2025: 2228-2237.

[45] MCCORMAC J, HANDA A, DAVISON A, et al. SemanticFusion: Dense 3D semantic mapping with convolutional neural networks[C]//Proceedings of the IEEE International Conference on Robotics and Automation. [S.l.]: IEEE, 2017: 4628-4635.

[46] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.

[47] RUNZ M, BUFFIER M, AGAPITO L. MaskFusion: Real-time recognition, tracking and reconstruction of multiple moving objects[C]//Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. [S.l.]: IEEE, 2018: 28-39.

[48] YU C, LIU Z, LIU X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. [S.l]: IEEE, 2018: 1168-1174.

[49] BLOESCH M, CZARNOWSKI J, CLARK R, et al. CodeSLAM—Learning a compact, optimisable representation for dense visual SLAM[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 2560-2568.

[50] LI N, KONG Y, GONG Y, et al. UnDeepVO: Monocular visual odometry through unsupervised deep learning[C]//Proceedings of the IEEE International Conference on Robotics and Automation. [S.l]: IEEE, 2018: 728-734.

[51] YAN H, LI Y, ZHANG Z, CLID-SLAM: A coupled LiDAR-inertial neural implicit dense SLAM with region-specific SDF estimation[J]. arXiv preprint arXiv: 2508.05000, 2025.

[52] ROSINOL A, ABATE M, CHANG Y, et al. Kimera: An open-source library for real-time metric-semantic localization and mapping[C]//Proceedings of the IEEE International Conference on Robotics and Automation. [S.l]: IEEE, 2020: 1689-1696.

[53] BOWMAN S L, ATANASOV N, DANIILIDIS K, et al. Probabilistic data association for semantic SLAM[C]//Proceedings of the IEEE International Conference on Robotics and Automation. [S.l]: IEEE, 2017: 1722-1729.

[54] QIN Tong, LI Peiliang, SHEN Shaojie. VINS-Mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics,2018, 34(4): 1004-1020.

[55] CAMPOS C, ELVIRA R, JUAN J, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and

multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
[56] MUR-ARTAL R, TARDOS 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.
[57] MERRILL N, GENEVA P, KATRAGADDA S, et al. Fast monocular visual-inertial initialization leveraging learned single-view depth[J]. Robotics: Science and Systems (RSS) 2023. DOI: 10.15607/rss.2023.xix.72.
[58] SONG J, RICHARD A, OLIVARES-MENDEZ M. Improving monocular visual-inertial initialization with structureless visual-inertial bundle adjustment[J]. arXiv preprint arXiv: 2502.16598, 2025.
[59] LIU H, QIU J, HUANG W. Integrating point and line features for visual-inertial initialization[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2022: 1-12.
[60] SUN Jianjing. Field-VIO: Stereo visual-inertial odometry based on quantitative windows in agricultural open fields[C]//Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2024.
[61] HU Jiaxin. 1D-LRF aided visual-inertial odometry for high-altitude MAV flight[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2022.
[62] ALBERICO I. Structure-invariant range-visual-inertial odometry[C]//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2024.
[63] MATTHEW L. Visual inertial odometry using focal plane binary features (BIT-VIO) [C]//Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2024.
[64] XU Zewen. DOGE: An extrinsic orientation and gyroscope bias estimation for visual-inertial odometry initialization[C]//Proceedings of 2025 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2025.
[65] FORSRER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics. 2016, 33(1): 1-21.
[66] YANG Yulin. Decoupled right invariant error states for consistent visual-inertial navigation[C]//Proceedings of IEEE Robotics and Automation Letters. [S.l.]: IEEE, 2022: 1627-1634.
[67] CHEN Chuchu. FEJ2: A consistent visual-inertial state estimator design[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2022.
[68] ZHENG Huai, HUANG Guoquan. Square-root robocentric visual-inertial odometry with online spatiotemporal calibration[C]//Proceedings of IEEE Robotics and Automation Letters. [S.l.]: IEEE, 2022: 9961-9968.
[69] HU Deshun. A stochastic cloning square-root information filter with accurate feature tracking for visual-inertial odometry[C]//Proceedings of 2025 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2025.
[70] CHEN Chuchu, PENG Yuxiang, HUANG Guoquan. Visual-inertial state estimation with decoupled error and state representations[C]//Proceedings of International Workshop on the Algorithmic Foundations of Robotics. Chicago, USA: [s.n.], 2024.
[71] YI Bowen, MANCHESTER I R. On IMU preintegration: A nonlinear observer viewpoint and its applications[J]. Systems & Control Letters, 2024, 193: 105933.
[72] RUSSELL B. Deep IMU bias inference for robust visual-inertial odometry with factor graphs[J]. IEEE Robotics and Automation Letters, 2022, 8(1): 41-48.
[73] PAN Youqi. Adaptive VIO: Deep visual-inertial odometry with online continual learning[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2024.
[74] ZUO Xingxing. CodeVIO: Visual-inertial odometry with learned optimizable dense depth[C]//Proceedings of 2021 IEEE International Conference on Robotics and Automation. [S.l.]: IEEE, 2021.
[75] XU Y, CROON G C. CUAHN-VIO: Content-and-uncertainty-aware homography network for visual-inertial odometry[J]. Robotics and Autonomous Systems, 2025, 185: 104866.
[76] YANG M, YU C, KIM H S. Efficient deep visual and inertial odometry with adaptive visual modality selection[C]//Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
[77] GIOVANNI C. Learned inertial odometry for autonomous drone racing[J]. IEEE Robotics and Automation Letters. 2023, 8(5): 2684-2691.
[78] RENÉ R, BOCHKOVSKIY A, KOLTUN V. Vision transformers for dense prediction[C]//Proceedings of the IEEE/CVF

International Conference on Computer Vision. [S.l.]: IEEE, 2021.

[79] LAHAV L, TEED Z, JIA D. Raft-stereo: Multilevel recurrent field transforms for stereo matching[C]//Proceedings of 2021 International Conference on 3D Vision (3DV). [S.l.]: IEEE, 2021.

[80] ZACHARY T, JIA D. DROID-SLAM: Deep visual SLAM for monocular, stereo, and RGB-D cameras[J]. Advances in Neural Information Processing Systems, 2021, 34: 16558-16569.

[81] ZACHARY T, LIPSON L, JIA D. Deep patch visual odometry[J]. Advances in Neural Information Processing Systems, 2023, 36: 39033-39051.

[82] ALEXANDER K. Segment anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2023.

[83] YASIN A. SelfVIO: Self-supervised deep monocular visual-inertial odometry and depth estimation[J]. Neural Networks, 2022, 150: 119-136.

[84] SONG Seungwon. DynaVINS: A visual-inertial SLAM for dynamic environments[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 11523-11530.

[85] CHEN Chuchu. Monocular visual-inertial odometry with planar regularities[C]//Proceedings of ICRA. [S.l.]: [s.n.], 2023.

[86] GU G. TEVIO: Thermal-aided event-based visual inertial odometry for robust state estimation in challenging environments[J]. IEEE Transactions on Instrumentation and Measurement, 2025.

[87] ZHU Jinwen. Robust 4D radar-aided inertial navigation for aerial vehicles[J]. arXiv preprint arXiv: 2502.15452, 2025.

[88] MAXIME O. DINOv2: Learning robust visual features without supervision[J]. arXiv preprint arXiv: 2304.07193, 2023.

[89] MERRILL N, GENEVA P, KATRAGADDA S, et al. Fast monocular visual-inertial initialization leveraging learned single-view depth[C]//Proceedings of Robotics: Science and Systems (RSS). [S.l.]: [s.n.], 2023: 1-10.

[90] SEOK M L. Event- and frame-based visual-inertial odometry with adaptive filtering based on 8-DOF warping uncertainty[J]. IEEE Robotics and Automation Letters, 2023, 9(2): 1003-1010.

[91] SAIMOULI K. NERF-VINS: A real-time neural radiance field map-based visual-inertial navigation system[C]//Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2024.

[92] LIU Shilong. Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection[C]//Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.

[93] BEN M. NeRF: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106.

[94] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]//Proceedings of 2007 IEEE International Conference on Robotics and Automation. [S.l.]: IEEE, 2007: 3565-3572.

[95] BLOESCH M, OMARI S, HUTTER M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]//Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2015: 298-304.

[96] DELLAERT F, KAESS M. Factor graphs for robot perception[J]. Foundations and Trends in Robotics, 2017, 6(1/2): 1-139.

[97] 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.

[98] 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.

[99] QIN T, CAO S, PAN J, et al. A general optimization-based framework for global pose estimation with multiple sensors[J]. arXiv preprint arXiv: 1901.03642, 2019.

[100] GENEVA P, ECKENHOFF K, LEE W, et al. Openvins: A research platform for visual-inertial estimation[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2020: 4666-4672.

[101] VON STUMBERG L, USENKO V, CREMERS D. Direct sparse visual-inertial odometry using dynamic marginalization [C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2018: 2510-2517.

[102] ZHANG J, SINGH S. LOAM: Lidar odometry and mapping in real-time[C]//Proceedings of Robotics: Science and Systems. [S.l.]: [s.n.], 2014: 1-9.

[103] SHAN T, ENGLOT B, MEYERS D, et al. LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping [C]//Proceedings of 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2020:

5135-5142.
[104] XU W, ZHANG F. FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3317-3324.
[105] XU W, CAI Y, HE D, et al. FAST-LIO2: Fast direct LiDAR-inertial odometry[J]. IEEE Transactions on Robotics, 2022, 38(4): 2053-2073.
[106] BAI C, XIAO T, CHEN Y, et al. Faster-LIO: Lightweight tightly coupled lidar-inertial odometry using parallel sparse incremental voxels[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 4861-4868.
[107] 王铉彬, 李星星, 廖健驰, 等. 基于图优化的紧耦合双目视觉/惯性/激光雷达SLAM方法[J]. 测绘学报, 2022, 51(8): 1744-1756.
WANG Xuanbin, LI Xingxing, LIAO Jianchi, et al. Tightly-coupled stereo visual-inertial-LiDAR SLAM based on graph optimization[J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(8): 1744-1756.
[108] LIN J, ZHANG F. R3LIVE: A robust, real-time, RGB-colored, LiDAR-inertial-visual tightly-coupled state estimation and mapping package[C]//Proceedings of 2022 International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2022: 10672-10678.
[109] SHAN T, ENGLOT B. LeGO-LOAM: Lightweight and ground-optimized lidar odometry and mapping on variable terrain [C]//Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2018: 4758-4765.
[110] YANG L, KANG B, HUANG Z, et al. Depth anything: Unleashing the power of large-scale unlabeled data[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2024: 10371-10381.
[111] SHAN Tixiao. LVI-SAM: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping[C]//Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]: IEEE, 2021.
[112] ZHENG Chunran. FAST-LIVO: Fast and tightly-coupled sparse-direct lidar-inertial-visual odometry[C]//Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2022.
[113] WANG X, LI X, YU H, et al. GIVL-SLAM: A robust and high-precision SLAM system by tightly coupled GNSS RTK, inertial, vision, and LiDAR[J]. IEEE/ASME Transactions on Mechatronics, 2024, 30(2): 1212-1223.
[114] WANG X, LI X, CHANG H, et al. GIVE: A tightly coupled RTK-inertial-visual state estimator for robust and precise positioning[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1005615.
[115] LI T, PEI L, XIANG Y, et al. P3-LINS: Tightly coupled PPP-GNSS/INS/LiDAR navigation system with effective initialization[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 8501813.
[116] XIA C, LI X, LI S, et al. Invariant-EKF-based GNSS/INS/vision integration with high convergence and accuracy[J]. IEEE/ASME Transactions on Mechatronics, 2024.
[117] MILIOTO A, VIZZO I, BEHLEY J, et al. RangeNet++: Fast and accurate LiDAR semantic segmentation[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2019: 4213-4220.
[118] ZHU X, ZHOU H, WANG T, et al. Cylindrical and asymmetrical 3D convolution networks for lidar segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 9939-9948.
[119] AYGUN M, OSEP A, WEBER M, et al. 4D panoptic LiDAR segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2021: 5527-5537.
[120] CHEN X, MILIOTO A, PALAZZOLO E, et al. SuMa++: Efficient LiDAR-based semantic SLAM[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2019: 4530-4537.
[121] WANG Y, SOLOMON J M. Deep closest point: Learning representations for point cloud registration[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 3523-3532.
[122] YEW Z J, LEE G H. RPM-Net: Robust point matching using learned features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 11824-11833.
[123] 孙淑光, 王添光, 刘瑞华. 基于单滤波器AIME的GNSS/INS紧组合完好性监测[J]. 航空科学技术, 2025, 36(7): 49-56.
SUN Shuguang, WANG Tianguang, LIU Ruihua. Integrity monitoring of tightly coupled GNSS/INS based on single-filter AIME[J]. Aeronautical Science & Technology, 2025, 36(7): 49-56.
[124] 王柳淇, 李亮, 陈雷, 等. 实时精密卫星钟轨完好性服务初步性能评估[J/OL]. 武汉大学学报(信息科学版): 1-18[2026-02-26]. https://doi.org/10.13203/j.whugis20240396.
WANG Liuqi, LI Liang, CHEN Lei, et al. Preliminary performance evaluation of real-time precise satellite clock and orbit in-

tegrity service[J/OL]. Geomatics and Information Science of Wuhan University: 1-18[2026-02-26]. https://doi.org/10.13203/j.whugis20240396.

[125] 赵靖, 宋丹. 无人机GNSS/IMU组合导航系统完好性监测方法[J]. 航空学报, 2024, 45(7): 247-260.
ZHAO Jing, SONG Dan. Integrity monitoring method for UAV GNSS/IMU integrated navigation system[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(7): 247-260.

[126] 刘悦, 李化义, 张世杰, 等. 面向视觉惯导的导航系统初始化技术综述[J]. 计算机工程与应用, 2025, 61(2): 1-18.
LIU Yue, LI Huayi, ZHANG Shijie, et al. Review of initialization techniques for visual-inertial navigation systems[J]. Computer Engineering and Applications, 2025, 61(2): 1-18.

[127] WANG A C, GUO Y Y, CHEN H. Sensor data fusion optimization in UAV integrated navigation based on matrix factorization [J]. Array, 2025, 28: 100524.

[128] 张路, 邵正途, 翁呈祥, 等. 美军有/无人机协同作战运用及关键技术研究[J]. 战术导弹技术, 2022(6): 128-137.
ZHANG Lu, SHAO Zhengtu, WENG Chengxiang, et al. Operational application and key technologies of manned/unmanned teaming in US military[J]. Tactical Missile Technology, 2022(6): 128-137.

[129] GEORGIOS S. Terrain-aided navigation for long-endurance and deep-rated autonomous underwater vehicles[J]. Journal of Field Robotics, 2019, 36(2): 447-474.

[130] 董晶, 胡权富, 刘海桥, 等. 基于图像匹配的机载惯导位置和航向修正方法[J/OL]. 航空学报: 1-16[2026-02-26]. https://link.cnki.net/urlid/11.1929.V.20250911.1323.026.
DONG Jing, HU Quanfu, LIU Haiqiao, et al. Image matching-based position and heading correction method for airborne inertial navigation[J/OL]. Acta Aeronautica et Astronautica Sinica: 1-16[2026-02-26]. https://link. cnki. net/urlid/11.1929. V.20250911.1323.026

[131] PRANAV S, RAGHUVANSHI N, GOVEAS N. Uav-vln: End-to-end vision language guided navigation for UAVs[C]// Proceddings of 2025 European Conference on Mobile Robots (ECMR). [S.l.]: IEEE, 2025.

[132] 罗子岩, 陈帅, 王国栋, 等. 多源融合导航系统的因子图算法综述[J]. 导航与控制, 2021, 20(3): 9-16.
LUO Ziyan, CHEN Shuai, WANG Guodong, et al. Review of factor graph algorithms for multi-source integrated navigation systems[J]. Navigation and Control, 2021, 20(3): 9-16.

[133] 谷美颖, 李航, 张家伟, 等. 基于视觉的无人机定位与导航方法研究综述[J]. 电子学报, 2025, 53(3): 651-685.
GU Meiying, LI Hang, ZHANG Jiawei, et al. Review of vision-based UAV localization and navigation methods[J]. Acta Electronica Sinica, 2025, 53(3): 651-685.

作者简介:

许悦雷(1975-),男,教授,博士生导师,研究方向:多模态感知、自主导航定位、具身飞行器等,E-mail: xuyuelei@nwpu.edu.cn。

王铉彬(1996-),**通信作者**,男,副教授,硕士生导师,研究方向:多源融合导航、高精度定位与建图,E-mail: wangxuanbin@nwpu.edu.cn。

薛尚捷(1998-),男,博士研究生,研究方向:图像匹配、视觉导航与多源融合导航。

徐金海(2003-),男,硕士研究生,研究方向:视觉-惯性融合与自主导航。

(编辑:夏道家)

A Review of Autonomous Localization Technologies for Unmanned Aerial Vehicles in Complex Low-Altitude Environments

XU Yuelei[1,2,3], WANG Xuanbin[1,2,3*], XUE Shangjie[1,2,3], XU Jinhai[1,2,3]

(1. Unmanned Systems Research Institute, Northwestern Polytechnical University, Xi'an 710072, China; 2. School of Artificial Intelligence, Northwestern Polytechnical University, Xi'an 710072, China; 3. National Key Laboratory of Unmanned Aerial Vehicle Technology, Xi'an 710072, China)

Abstract: Complex low-altitude environments are typically characterized by the superposition of multi-source interference, drastic variations in sensing conditions, and incomplete environmental information, which collectively pose significant challenges to the continuity, reliability, and integrity of autonomous localization for unmanned aerial vehicles (UAVs). In such scenarios, Global Navigation Satellite System (GNSS) signals are prone to blockage and interference, visual perception suffers from weak textures, dynamic disturbances, and abrupt illumination changes, and inertial measurements inevitably accumulate long-term drift. The coupled degradation of these sensing modalities substantially undermines the stability and robustness of localization systems. To address these challenges, this paper systematically reviews representative types of degraded low-altitude environments and analyzes key technical bottlenecks under multi-source hybrid interference, including visual feature loss, inertial error divergence, and satellite positioning performance deterioration. Building upon this analysis, the developmental trajectory of vision-based navigation and localization techniques for UAVs is comprehensively surveyed, covering visual matching methods based on satellite signals or prior maps as well as recent advances in visual simultaneous localization and mapping (SLAM). Furthermore, visual-inertial fusion modeling and perception enhancement strategies are summarized, highlighting their technical advantages in improving localization accuracy and robustness. Subsequently, multi-sensor fusion navigation frameworks and robust fusion strategies tailored for GNSS-denied or degraded environments are discussed, with particular emphasis on collaborative modeling, degradation awareness, and integrity monitoring across heterogeneous modalities, including vision, inertial sensors, LiDAR, and satellite positioning. Finally, the paper outlines future directions for data-driven multimodal adaptive navigation methods, as well as the development trends of lightweight and intelligent high-integrity navigation technologies for unmanned aerial vehicles. This survey aims to provide a systematic reference for the research and engineering implementation of highly reliable autonomous localization technologies for UAVs operating in complex low-altitude environments.

Highlights:

1. This work systematically reviews the mechanisms of UAV localization degradation induced by multiple factors, including GNSS signal constraints, visual perception degradation, and inertial measurement drift, and clarifies the sources and impacts of multi-source coupled errors in complex low-altitude environments.
2. It surveys the latest advances in UAV autonomous localization along three key technical routes: pure visual localization, visual-inertial localization, and multi-source heterogeneous fusion localization.
3. It establishes a comprehensive technical framework spanning sensor degradation mechanisms to robust multi-source fusion, providing systematic references and forward-looking insights for the research and engineering application of highly reliable UAV autonomous localization in complex low-altitude environments.

Key words: GNSS denial; visual navigation; visual-inertial odometry; LiDAR; multi-sensor fusion; low-altitude applications

Foundation items: National Natural Science Foundation of China (No.42504030); Fundamental Research Fund for the Central Universities (Science and Technology Program) (No.D5000250047).

Received: 2026-01-12; **Revised**: 2026-02-28

***Corresponding author, E-mail:** wangxuanbin@nwpu.edu.cn.