

DOI: 10.19650/j.cnki.cjsi.J2514497

融合纹理感知与自适应特征的视觉 SLAM 算法*

栾添添, 梁洪杰, 李传龙, 孙明晓, 孙骁骏
(哈尔滨理工大学自动化学院 哈尔滨 150080)

摘要:针对传统视觉惯性同时定位与建图(SLAM)算法在昏暗与低纹理场景中提取的特征点质量差,导致视觉惯性 SLAM 算法的跟踪稳定性差、跟踪易丢失、定位精度低的问题。提出了一种融合纹理感知与自适应特征跟踪的无人机双目视觉惯性 SLAM 算法。首先,引入噪声抑制因子,以抑制图像增强过程中噪声被放大,设计了一种纹理感知控制权重,根据图像局部区域的纹理丰富程度,自适应地调节增强强度,在纹理丰富区域适当增强,在纹理稀疏区域则保持较低的增强幅度,从而在提升整体图像对比度的同时,有效避免噪声干扰并保留更多有效细节。其次,采用 XFeat 特征提取网络代替人工设计的图像特征,同时为应对特征点分布不均的问题,提出一种基于密度反馈的检测阈值调整机制,该机制根据当前帧中特征点的空间分布密度,自适应地调整特征点检测的响应阈值,从而在特征稀疏区域检测更多特征点,在密集区域则保持较高阈值以筛选高质量特征,实现更为稳定和均匀的特征跟踪。最后,与视觉惯性同时定位与建图(VINS)系统的后端非线性优化结合,构建一个完整的 SLAM 系统。在 EuRoc 数据集上的实验结果表明,在光线昏暗与纹理稀疏的复合挑战场景中,相较于多传感器状态估计器(VINS-Fusion)与 SuperVINS 算法,所提算法的定位精度分别提升了 29% 与 13%。在真实场景实验中,展现出更优的定位精度,实现了 0.486 m 的闭合误差。

关键词:视觉惯性 SLAM;深度学习;位姿估计;图像增强;无人机

中图分类号: TH85 TP242 **文献标识码:** A **国家标准学科分类代码:** 510.80

A visual SLAM algorithm integrating texture awareness and adaptive features

Luan Tiantian, Liang Hongjie, Li Chuanlong, Sun Mingxiao, Sun Xiaojun
(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: Conventional visual-inertial SLAM algorithms often suffer from poor feature quality in dim and low-texture environments, leading to unstable tracking, frequent tracking loss, and low localization accuracy. To address this, this paper proposes a stereo visual-inertial SLAM algorithm for UAVs that integrates texture awareness and adaptive feature tracking. First, a noise suppression factor is introduced to prevent noise amplification during image enhancement, and a texture-aware control weight is designed to adaptively adjust the enhancement strength based on local texture richness, applying stronger enhancement in texture-rich regions while keeping the enhancement low in texture-sparse areas, thereby improving overall image contrast while effectively suppressing noise and preserving useful details. Second, the XFeat network is used to replace hand-crafted features, and a density-feedback threshold adjustment mechanism is introduced to address uneven feature distribution by adaptively adjusting the detection threshold according to local feature density, enabling more features to be detected in sparse regions while maintaining a higher threshold in dense regions to retain high-quality features, thereby improving the stability and uniformity of feature tracking. Finally, these improvements are integrated with the VINS back-end optimization to build a complete SLAM system. Evaluations on the EuRoc dataset show our method achieves 29% and 13% higher accuracy than VINS-Fusion and SuperVINS, respectively, in challenging dim and low-texture scenarios. Real-world experiments demonstrate a closure error of 0.486 meters, confirming its superior performance.

Keywords: visual-inertial SLAM; deep learning; pose estimation; image enhancement; unmanned aerial vehicles

0 引 言

同时定位与建图(simultaneous location and mapping,

SLAM)技术在无人车运动、无人机巡检、无人机定位与导航领域得到了广泛的应用^[1-3],视觉惯性 SLAM 系统(visual-inertial slam, VINS)通过融合视觉特征观测与惯性测量单元(inertial measurement unit, IMU)数据,实现对

收稿日期:2025-09-13 Received Date: 2025-09-13

* 基金项目:黑龙江省自然科学基金项目(YQ2024E047)、黑龙江省优秀青年教师基础研究支持计划项目(YQJH2024067)资助

相机运动轨迹及姿态的联合估计,有效利用了视觉与惯性传感器在感知特性上的互补性,显著提升了位姿估计的鲁棒性与精度^[4]。

主流视觉惯性 SLAM 方法主要分为传统方法与深度学习两类。其中,传统方法依赖于几何特征跟踪,通过耦合视觉重投影误差与 IMU 预积分约束,构建并优化联合代价函数,以实现精确的相机姿态估计^[5]。Qin 等^[6]于 2018 年提出的 VINS-mono 是一个紧耦合的视觉惯性 SLAM 框架,前端采用稀疏光流法进行特征跟踪,显著降低了视觉前端的计算开销,提升了系统的实时性。随后引入双目相机与全球定位系统传感器提出了多传感器状态估计器 (visual-inertial state estimator-fusion, VINS-Fusion)。但在光照发生显著变化的场景下,其特征点跟踪的稳定性较差,产生严重的跟踪丢失与跟踪漂移。2021 年 Campos 等^[7]以 ORB(oriented fast and brief)特征点为基础的 SLAM 系统 ORB-SLAM2^[8]作为基础框架,引入 IMU 数据,开发了 ORB-SLAM3 系统,该系统继承了上一代的核心架构与先进特性,引入多地图系统,实现了高效的地图复用,显著提升了回环检测召回率。但在弱纹理或重复纹理场景下,ORB 特征面临着提取困难和描述子区分度不足的问题^[9],导致特征匹配阶段产生极高的误匹配率,对系统的鲁棒性构成挑战^[10]。

深度学习的迅速发展为视觉惯性 SLAM 开辟了新的途径,据此产生的方法主要可分为两类。一种是利用深度学习技术增强或替代 SLAM 系统中的特定模块^[11],另一种是利用端到端的深度学习模型直接回归出相机的位姿。余洪山等^[12]通过融合改进 SuperPoint 特征提取网络与 VINS-mono 框架构建的单目视觉惯性 SLAM,提高了在挑战性环境下的鲁棒性。纪泽源等^[13]将图神经网络融入 VINS-Fusion 系统进行特征点匹配,提高了 SLAM 前端特征点的跟踪性能,在室内低照度条件下的定位精度显著提升。Zhang 等^[14]使用深度学习

网络进行特征提取与特征匹配,构建了一个混合视觉 SLAM 框架,支持多种模式,在动态光照条件、弱纹理区域和相机抖动等恶劣条件下显著提高 SLAM 的定位精度。Wang 等^[15]提出了基于特征提取网络的单目视觉-惯性 SLAM 系统(monocular visual-inertial SLAM based on feature extraction network, VINS-FEN)框架,将基于卷积神经网络的特征提取网络整合到 VINS 基础架构中,显著提升了系统性能。但其特征提取环节需进行离线处理,难以满足实时处理的需求。Luo 等^[16]提出的 SuperVINS 框架集成 SuperPoint 特征提取网络用于特征点的提取与回环检测,同时采用 LightGlue 匹配网络进行特征匹配,在挑战性场景取得了较高的精度。但其采用的 SuperPoint 网络模型过于复杂,难以在无人机等资源受限的场景中实时运行。Guo 等^[17]提出了一种集成卡尔曼滤波器的端到端神经网络方法。但由于端到端的视觉惯性 SLAM 系统需要大量不同场景数据进行训练,在不同场景下的泛化性较差,而且,端到端深度学习往往存在可解释性不足的问题。2024 年 Pojte 等^[18]提出了 XFeat,这是一种面向资源受限设备的轻量化视觉特征提取网络,该方法基于卷积神经网络重新设计局部特征的检测与提取,通过约束网格通道数同时最大化输入分辨率实现精度与效率的平衡。

尽管上述视觉惯性 SLAM 在复杂环境中展现出超越传统方法的精度与鲁棒性^[19],但其在光照昏暗且纹理稀疏等复合场景仍存在问题,据此,提出一种融合纹理感知增强与自适应特征跟踪的视觉惯性 SLAM 系统。该系统旨在通过增强弱纹理信息并采用轻量级网络,有效提升无人机在严苛视觉条件下的定位精度与鲁棒性。

1 整体算法流程

提出的系统架构如图 1 所示。

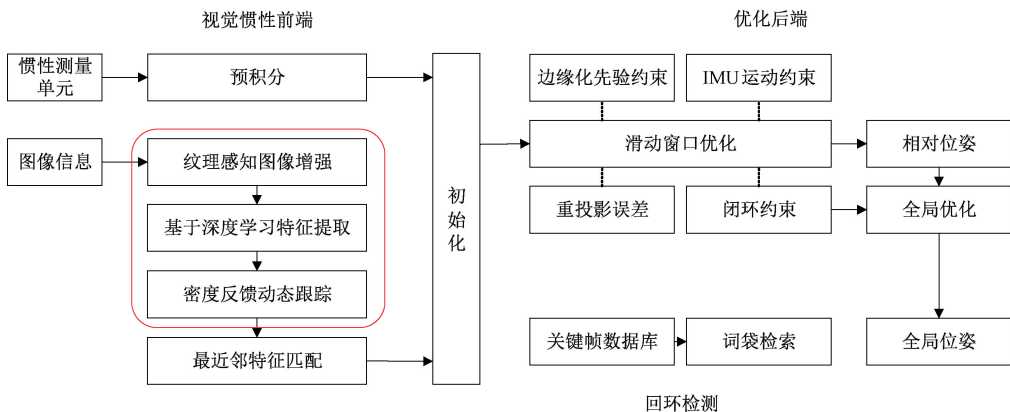


图 1 算法总体框架

Fig. 1 Algorithm overall

本系统基于 VINS-Fusion 框架,包含视觉惯性前端、非线性优化后端和回环检测 3 个核心模块。前端采用 XFeat 网络实现实时的特征检测与跟踪。后端启动双线程:首先基于时间戳对齐视觉与 IMU 数据,随后依次执行滑动窗口优化与回环全局优化,以消除累积漂移,最终输出高精度位姿。过程包括:

1) 视觉惯性前端。首先对输入的图像进行预处理操作,在传统自适应伽马校正算法^[20] (adaptive Gamma correction with weighted distribution, AGCWD) 的基础上引入纹理控制函数与噪声引导项,以增强图像并抑制噪声。接着使用 XFeat 特征提取网络进行图像特征点提取,采用一种基于密度反馈的动态阈值调整机制以维持每帧图像的特征点数量均衡,再通过最近邻匹配与随机抽样一致(random sample consensus, RANSAC) 算法获取可靠的特征跟踪结果。然后,实现图像特征与 IMU 数据的时空对齐,对 IMU 测量数据进行预积分处理,计算 IMU 约束项对应的误差、协方差矩阵及雅可比矩阵。最终,处理后的视觉特征观测值与 IMU 测量误差项被传输至后端模块。

2) 非线性优化后端。非线性优化后端采用紧耦合策略,融合视觉与 IMU 观测信息,基于滑动窗口机制构建包含视觉重投影误差、IMU 预积分误差、边缘化先验及回环位姿约束的联合优化目标函数,在控制计算复杂度的同时实现系统状态的高精度估计。

3) 回环检测。基于视觉词袋模型进行回环关键帧检索,通过特征描述符匹配建立特征点对应关系,并使用 RANSAC 算法剔除异常匹配对以提升几何一致性,进而通过滑动窗口优化与四自由度位姿图优化求解精确相对位姿并校正累积漂移,实现高精度全局轨迹估计。

2 纹理感知与自适应特征的视觉 SLAM 系统

2.1 融合纹理感知与噪声抑制的图像增强算法

在低光照与弱纹理场景下,传统图像增强算法在提升亮度时,易在细节区域产生过度增强,从而在噪声敏感区引发虚假边缘与结构畸变,严重损害后续特征提取的鲁棒性。为此,引入纹理感知控制与噪声引导项,以精准调控增强力度。

首先构建像素灰度分布的统计信息。设输入灰度图像为 $I(x, y)$, 其灰度值范围为 $[0, 255]$, 统计像素灰度直方图 $p(i)$, 即:

$$p(i) = \frac{n_i}{N}, i \in [0, 255] \quad (1)$$

式中: n_i 表示灰度值为 i 的像素数量; N 为图像总像素数量。为提升图像暗部区域的增强力度,同时抑制高亮区域的过增强,采用如下指数加权机制构建加权直方图,即:

$$P_w(i) = P_{\max} \cdot \left(\frac{P(i) - P_{\min}}{P_{\max} - P_{\min}} \right)^a \quad (2)$$

式中: P_{\max} 和 P_{\min} 为最大和最小水平的 $p(i)$; a 为平滑系数。图像的加权累积分布函数为:

$$C_w(i) = \sum_{l=0}^i \frac{P_w(i)}{\sum_{i=0}^{255} P_w(i)} \quad (3)$$

该函数反映了每一灰度值在图像中的累积分布情况,将在后续用于伽马值计算。

在弱光场景下图像易受传感器噪声干扰,若直接增强将放大这些高频噪声,破坏结构稳定性。为解决该问题,引入像素级噪声抑制因子 $\beta(x, y)$, 在噪声区域下调增强强度,为防止伪边缘扩大,采用图像的平滑残差估计噪声水平。对图像进行高斯滤波,二维高斯函数的概率密度函数为:

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (4)$$

式中: σ 为正态分布的标准差; x, y 为像素的横纵坐标。高斯滤波器的数学模型为:

$$f(x, y) = h(x, y) * I(x, y) \quad (5)$$

式中: $h(x, y)$ 为高斯卷积核; $*$ 表示卷积操作; $I(x, y)$ 为原始图像灰度值。

计算像素残差图,即:

$$R(x, y) = |I(x, y) - f(x, y)| \quad (6)$$

对残差进行归一化操作,即:

$$R_{\text{norm}}(x, y) = \frac{R(x, y) - R_{\min}}{R_{\max} - R_{\min}} \quad (7)$$

式中: R_{\max} 与 R_{\min} 为残差的最大值与最小值,将其映射为噪声抑制权重,即:

$$\beta(x, y) = 1 - (1 - \beta_{\min}) \cdot R_{\text{norm}}(x, y) \quad (8)$$

式中: β_{\min} 是最小置信值,用于限制最低增强幅度。进行伽马校正的参数为:

$$\gamma(x, y) = \beta(x, y) \cdot (1 - C_w(I(x, y))) \quad (9)$$

将经过噪声抑制的伽马参数应用于图像,得到伽马增强后的图像,即:

$$I_{\text{AGCWD}}(x, y) = 255 \cdot \left(\frac{I(x, y)}{255} \right)^{\gamma(x, y)} \quad (10)$$

该方法保留了原始 AGCWD 算法优势,同时引入了对噪声区域的鲁棒性。

尽管伽马校正能改善图像亮度,但在低纹理区域,如地面、墙壁等,直接增强可能造成结构失真。为实现区域自适应增强,引入纹理感知权重,首先获得输入图像的梯度幅值图,采用 Sobel 算子计算梯度向量,即:

$$G_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * I(x, y) \quad (11)$$

$$\mathbf{G}_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * \mathbf{I}(x, y) \quad (12)$$

梯度幅值为梯度向量的欧式范数,即:

$$\mathbf{T}(x, y) = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2} \quad (13)$$

通过归一化梯度图得到融合系数,即:

$$\mathbf{a}(x, y) = \frac{\mathbf{T}(x, y) - \mathbf{T}_{\min}}{\mathbf{T}_{\max} - \mathbf{T}_{\min}}, \alpha \in [0, 1] \quad (14)$$

式中: \mathbf{T}_{\max} 与 \mathbf{T}_{\min} 为梯度幅值图的最大值与最小值, $\mathbf{a}(x, y)$ 越大,表示该像素处纹理越明显,增强占比越高;对于低纹理区域,增强幅度自动减小,降低伪结构产生的风险。融合伽马校正图像与原始图像,得到最终输出:

$$\mathbf{I}_{\text{final}}(x, y) = \mathbf{a}(x, y) \cdot \mathbf{I}_{\text{AGCWD}}(x, y) + (1 - \mathbf{a}(x, y)) \cdot \mathbf{I}(x, y) \quad (15)$$

该机制实现了区域自适应增强,在高纹理区提升边缘对比度,在低纹理区则抑制噪声、保留原貌。这使得图像序列的灰度分布更合理,暗区噪声得到抑制,进而通过提升特征匹配的一致性,显著提高了SLAM的定位精度。

2.2 基于XFeat网络的视觉动态特征跟踪

原始VINS-Fusion系统采用Shi-Tomsai角点检测器和稀疏光流法^[21](lucas kanade, LK)进行特征跟踪,然而,在实际低光照场景中,由人工光源引起的局部光照动态变化会显著降低特征跟踪的鲁棒性。采用一种对低光照鲁棒的深度卷积网络来提取特征点与描述子,通过自适应调整检测阈值以维持跟踪稳定性,并利用最近邻搜索完成特征匹配。

XFeat是一种面向资源受限设备的轻量级特征提取网络,由编码层、检测分支与描述子解码分支构成,能够同步输出特征点位置与描述符。

在图像经过网络前向传播后,网络会生成3个关键输出:关键点热图、描述子特征图和可靠性热图。首先采用最近邻插值从关键点热图中采样关键点置信度分数 P_k ,随后利用双线性插值从可靠性热图中采样匹配可靠性分数 P_r ,最终,通过融合上述两类分数,计算每个关键点的综合分数 P_s :

$$P_s = P_k \times P_r \quad (16)$$

然后依据综合分数对关键点进行降序排列,并筛选前 K 个具有高区分度的关键点,最后,在对应关键点位置处,采用双三次插值从描述子特征图中提取描述子向量,并对描述子执行L2归一化处理使其具有单位长度,从而增强后续特征匹配的鲁棒性与精度。

在低光照及弱纹理环境下,为增强特征点的空间分布均匀性并抑制过度聚集,引入非极大值抑制(non-

maximum suppression, NMS)机制,以约束特征点之间的最小间隔。该方法通过在局部邻域保留最高响应特征点,有效提升了特征的空间分布均匀性、代表性及可跟踪性。经系统实验验证,当抑制半径设置为5 pixels时,能够在特征数量与分布均匀性之间达到最佳平衡,此时特征提取的综合性能最优,显著改善了在挑战性场景下的视觉跟踪稳定性与鲁棒性。

传统XFeat网络依赖固定阈值进行特征点检测,在光照剧变的低光照环境中,这会导致特征点数量锐减,甚至引发跟踪失败。为此,提出一种基于密度反馈的动态阈值调节机制,通过稳定特征点数量来提升系统在极端环境下的鲁棒性。动态特征跟踪整体流程如图2所示。

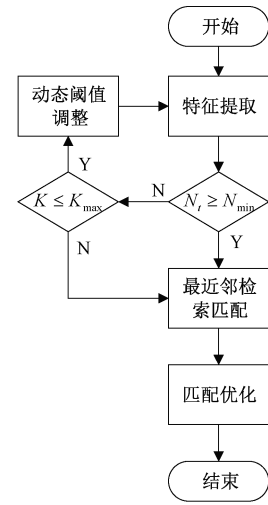


图2 特征动态跟踪流程

Fig. 2 Feature dynamic tracking flowchart

在VINS-Fusion系统中,特征点跟踪数量通常维持在100~300,为确保跟踪的鲁棒性,设定特征点数量阈值 $N_{\min} = 100$ 作为触发条件,当检测到当前帧特征点数量 $N_t < N_{\min}$ 时,触发动态阈值调整机制,即:

$$\tau = \left(1 - \lambda \frac{N_{\min} - N_t}{N_{\min}} \right) \tau_{\text{base}} \quad (17)$$

式中: τ_{base} 为预设基础阈值; λ 为调整幅度,取值为0.7。调整过程循环迭代执行,直至满足 $N_t \geq N_{\min}$ 或达到最大迭代次数 $K_{\max} = 5$,该机制有效避免了因单一阈值不适应图像内容变化而导致的特征丢失问题,从而保障特征提取的稳定性。在特征匹配阶段,采用最近邻检索建立当前帧与参考帧的特征点对应关系,通过计算特征描述符间的欧氏距离,为当前帧各特征点选取参考帧中距离最小的候选点,形成初始匹配集合,随后基于RANSAC进行鲁棒性优化剔除不符合几何约束的异常匹配对。

3 实验与结果

本研究采用公开数据集 EuRoC^[22]与真实场景数据,将所提出的视觉惯性 SLAM 系统与当前主流方案进行对比评估。本次实验主要包含 3 部分:1) 针对不同场景下的图像序列数据,评估所提方法在低光照与低纹理条件下的图像增强性能;2) 在公开数据集上开展绝对轨迹误差的定量评估;3) 在真实场景中系统输出的运动轨迹进行闭合误差分析。本研究实现的 SLAM 系统采用模块化架构:视觉前端基于 Xfeat 网络,使用 C++语言并结合 ONNX Runtime 推理引擎进行高效计算;非线性优化后端与回环检测模块则基于原有 VINS-Fusion 系统构建。系统各模块间通过机器人操作系统(robot operating system, ROS)节点通信机制实现数据交互与集成。本实验的硬件平台是由 Intel Core i9-13900HX CPU、NVIDIA RTX 4060 显卡、16 GB 内存构成,使用 Ubuntu20.04 操作系统。

本研究的真实场景实验采用四旋翼无人机作为验证平台,搭载 Intel Realsense D435I 相机作为视觉惯性传感器,用于同步采集图像和 IMU 数据。实验环境选定为一处光线较暗的环形楼道,该场景中包含大量白墙等低纹理区域。通过 ROS 系统录制整个实验过程中的图像与 IMU 数据,图像分辨率为 640 pixels×480 pixels,IMU 频率为 200 Hz,实验平台配置与实验环境如图 3 所示。

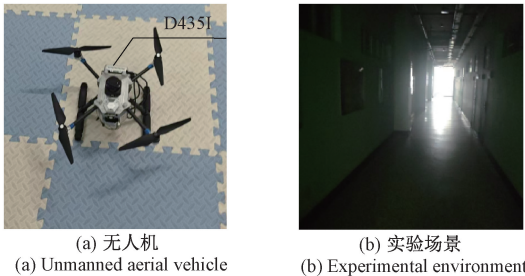


图 3 实验验证平台和实验场景

Fig. 3 Experimental verification platform and experimental scenarios

3.1 图像增强效果对比实验

为评估提出的融合纹理感知与噪声抑制的自适应图像增强算法在低照度、低纹理环境下对特征提取与匹配性能的提升效果,构建了一个基于特征质量与几何一致性的综合评价体系,并将所提方法与对比度受限自适应直方图均衡^[23](contrast limited adaptive histogram equalization, CLAHE)、Retinex^[24]及原始 AGCWD^[20]等典型算法进行对比。评价指标包括特征点平均得分,用于评估特征提取的稳定性;以及经 RANSAC 验证后的特征点匹配对数,用于衡量图像间的几何一致性。此外,针对视觉-惯性 SLAM

系统对实时性的要求,还以单帧图像处理耗时作为效率指标,对不同增强算法的计算开销进行定量分析。

在 EuRoC 数据集中选取具有代表性的 40 张图像序列进行图像增强处理,设置 XFeat 特征点提取数量为 300,根据特征点响应值与可靠性热图计算特征点得分,随后在相邻图像序列中执行特征匹配,并采用 RANSAC 算法以剔除异常匹配,在此基础上评估正确匹配的特征点对数,并进一步统计整个图像序列的平均特征点分数与处理耗时。对图像增强算法的评估结果如表 1 所示。

表 1 图像增强评估结果

Table 1 Image enhancement evaluation results

评价指标	AGCWD	CLAHE	Retinex	本文算法
特征点得分	0.186 3	0.177 6	0.179 1	0.222 3
平均匹配数量/对	103.25	97.50	89.25	132.50
平均耗时/s	0.007 5	0.003 5	0.017 5	0.008 8

从特征点质量来看,本文算法取得的特征点得分为 0.222 3,显著高于其他对比算法,较次优的 AGCWD 算法的特征点平均得分 0.186 3 提升约 19.3%,表明其提取的特征点具有更高的置信度与稳定性。在平均匹配数量上,本文算法正确匹配 132.50 对,远超其他算法,比 AGCWD 正确匹配的 103.25 对高出 28.3%,证明其增强后的图像更有利于生成正确、稳定的特征匹配,直接有利于后续 SLAM 中的跟踪鲁棒性与位姿估计精度。本文算法单帧平均耗时为 0.008 8 s,虽略高于 AGCWD 算法与 CLAHE 算法,但在计算开销增幅极小的情况下,换来了特征质量与匹配数量的显著提升。

综合而言,在低光照、低纹理场景下,本文算法在特征点质量与匹配数量上均显著优于对比方法。本文算法在提升图像亮度的同时,有效保持了细节并抑制了噪声,从而在保证系统实时性的前提下,为 SLAM 在挑战性环境下的稳定运行提供了关键支持。

3.2 VINS 系统定位精度实验

在定位精度评估实验中,本文选用公开数据集 EuRoC,将所提出的系统与 VINS-Fusion 系统以及基于深度学习特征提取的 SuperVINS 系统进行对比。3 种系统除前端采用的方法不同外,其余模块均保持一致,本实验采用 VINS-Fusion 系统中的回环检测部分以保持一致性。所有对比系统均采用其默认配置参数,并基于双目相机与 IMU 的组合传感器模式进行实验。本次实验选取 EuRoC 数据集中的 11 个序列进行测试,采用 evo 工具评估各系统的全局轨迹精度。以绝对轨迹误差(absolute pose error, APE)作为性能评估指标,通过计算轨迹估计值与真实轨迹之间绝对位姿误差的均方根误差(root mean square error, RMSE),以衡量定位算法的准确性。

在 EuRoC 数据集的简单场景序列实验结果如表 2 所示,表中以黑色粗体标示各序列中性能最优的结果。

表 2 EuRoC 简单序列均方根误差

Table 2 The RMSE results on the easy EuRoC dataset (m)

序列	VINS-Fusion	SuperVINS	本文
MH_01	0.056 599	0.057 569	0.056 931
MH_02	0.049 251	0.051 626	0.055 010
V1_01	0.043 284	0.056 954	0.055 274
V2_01	0.065 986	0.052 675	0.046 783
均值	0.053 780	0.054 706	0.053 500

由于环境光照充足、纹理特征丰富,各类算法均能获得可靠的视觉观测信息。因此,本文所提出的算法在定位精度方面与原始 VINS-Fusion 算法及 SuperVINS 算法表现相当,均具有较高的精度。本文算法的绝对轨迹误差在 V2_01 序列中取得了最优的性能,且本文算法的轨迹均方根误差平均值最低,达到 0.053 5 m。在 EuRoC 数据集中等和困难序列上的实验结果如表 3 所示。

表 3 EuRoC 中等/困难序列均方根误差

Table 3 The RMSE result on the medium/difficult EuRoC dataset (m)

序列	VINS-Fusion	SuperVINS	本文
MH_03	0.065 072	0.078 330	0.071 356
MH_04	0.080 440	0.092 626	0.077 457
MH_05	0.095 633	0.099 946	0.092 478
V1_02	0.064 066	0.029 676	0.038 518
V1_03	0.179 891	0.124 124	0.044 850
V2_02	0.055 124	0.040 629	0.040 459
V2_03	0.118 488	0.071 485	0.100 787
均值	0.094 102	0.076 688	0.066 558

在 EuRoC 数据集的复杂场景中,采用深度学习作为前端特征提取的 VINS 系统具有较高的精度,原始 VINS-Fusion 系统仅在 MH_03 中等难度的序列中取得了最佳精度。从整体性能来看,本文算法的平均误差为 0.066 558 m,低于 VINS-Fusion 算法的 0.094 102 m 和 SuperVINS 的 0.076 688 m,相较于 VINS-Fusion 算法与 SuperVINS 算法,本文算法定位精度分别提高了 29% 与 13%,这主要归因于所提出的纹理感知增强方法与动态特征点跟踪策略,普遍存在光照昏暗与纹理稀疏的挑战性环境,而且能够自适应地增强图像纹理并维持稳定的特征跟踪,从而显著提升了位姿估计的精度与鲁棒性。实验结果充分表明,本文算法在复杂环境下仍可保持较高的定位性能。

具体到不同序列,本文算法在多数场景中均表现良好,尤其在序列 V1_03 上取得显著优势,误差仅为 0.044 850 m,远低于对比算法。在 MH_04、MH_05 和 V2_02 序列中也保持了最优的性能。VINS-Fusion、SuperVINS 以及本文算法在 V1_03 序列中的轨迹与真实值的对比结果如图 4 所示。

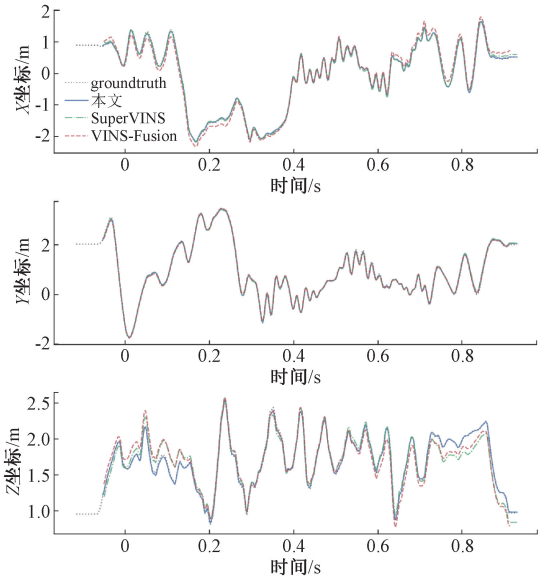


图 4 V1_03 序列轨迹对比

Fig. 4 Trajectory comparison for sequence V1_03

在充满低纹理区域和动态光照的 V1_03 序列中,本文算法展现出显著优势。其估计轨迹在各方向上均最接近真实值,平滑性优异,并能紧密跟踪转弯与高度变化等运动。相比之下,VINS-Fusion 与 SuperVINS 则出现了明显漂移。这充分证明了本文算法在精度与稳健性上的优越性。

3 种算法在 V1_03 序列 3 个方向的误差曲线如图 5 所示。

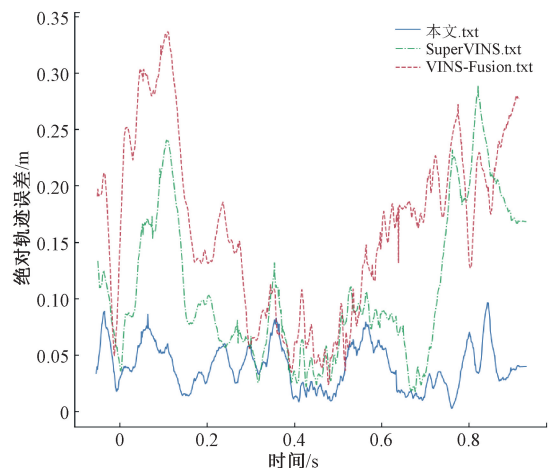


图 5 V1_03 序列误差曲线

Fig. 5 Error curve for sequence V1_03

轨迹对比表明,本文算法的估计误差在整个时间序列上均低于 VINS-Fusion 与 SuperVINS,且后期收敛平稳,鲁棒性更佳。相比之下,两种对比方法的误差曲线存在显著波动与多个峰值,稳定性不足。因此,本文算法在复杂环境下性能最优。

3.3 真实场景实验

实验平台为搭载 Intel Realsense D435I 相机的四旋翼无人机,用于采集双目图像与 IMU 数据。无人机沿一条长约 120 m 的预设环形轨迹飞行,速度保持稳定,并确保起降点基本重合以实现路径闭合。为评估位姿估计的整体一致性与累积误差,本实验在对比中禁用了回环检测模块,转而采用轨迹闭合误差作为评价指标,其定义为轨迹起点与终点间的欧氏距离。

各方向上轨迹对比图如图 6 所示。从图 6 的闭合精度对比可知:在 x 、 y 方向上,本文算法与对比算法精度相当;而在 z 方向上,本文算法的优势尤为突出,误差最小,有效抑制了高度漂移。综合而言,得益于更鲁棒的特征跟踪,本文算法取得了最佳的整体三维闭合精度,闭合误差仅为 0.486 m。

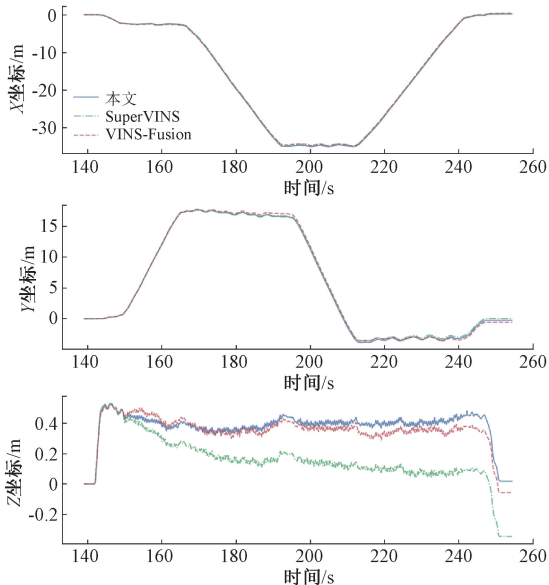


图 6 真实场景轨迹误差曲线

Fig. 6 Real world trajectory error curve

系统各阶段运行时间如表 4 所示,本文系统前端平均耗时仅为 35.2 ms(约 28 Hz),显著高于传感器 20 Hz 的数据输入速率,也高于 VINS-Fusion 后端 10 Hz 的优化速率。因此,该系统完全满足无人机高速飞行的实时性要求。

表 4 系统平均运行时间

Table 4 Average system operation time (ms)

序列	图像增强	特征提取	特征匹配	位姿估计
1	8.6	7.2	16.7	3.1
2	8.8	7.3	15.8	3.3
3	8.5	7.0	16.2	3.1
均值	8.6	7.2	16.2	3.2

4 结 论

针对无人机在昏暗及低纹理复合场景下视觉惯性 SLAM 定位精度骤降的挑战,提出融合纹理感知增强与轻量化动态特征提取网络的解决方案。1)通过设计噪声可控的自适应纹理增强模型,强化特征区域响应并稳定提升弱纹理信息,有效提高了特征点得分与匹配质量;2)同时将轻量化深度学习特征提取网络集成至 VINS-Fusion 前端,通过动态阈值调控,保障了特征点分布的连续稳定性,显著提升视觉观测的精度与鲁棒性;3)在 EuRoC 数据集上的实验结果表明,所提算法在纹理丰富且光照充足场景下的定位精度与 VINS-Fusion 基本相当,绝对轨迹均方根误差约为 0.053 5 m,而在光线昏暗与纹理稀疏的复合挑战场景中,相较于 VINS-Fusion 与 SuperVINS 算法,本文算法的定位精度分别提升了 29% 与 13%。在真实场景实验中,本文算法展现出更优的回环性能,实现了 0.486 m 的闭合误差。

未来工作中,计划引入基于深度学习的特征匹配网络,以提升视觉惯性 SLAM 系统中特征点之间的信息关联质量,从而缓解前端匹配可靠性不足的问题。同时,采用深度学习的特征描述符构建视觉词袋模型,并结合前述匹配网络,共同增强回环检测的召回率与鲁棒性,进一步提高 SLAM 系统的全局一致性精度。

参考文献

- [1] 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 M Y, LI H, ZHANG J W, et al. A review of vision-based UAV localization and navigation methods[J]. Acta Electronica Sinica, 2025, 53(3): 651-685.
- [2] 崔玉明,刘送永,吕振礼,等. 基于级联优化和强度特征的地下退化环境机器人自主精准定位[J]. 仪器仪表学报, 2023, 44(12): 208-216.
CUI Y M, LIU S Y, LYU ZH L, et al. Autonomous and accurate robot positioning in degraded underground

- environment based on cascade optimization and intensity feature[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(12): 208-216.
- [3] 孙明晓, 王鑫源, 栾添添, 等. 动态环境下基于深度图的无人车激光 SLAM 算法[J]. *仪器仪表学报*, 2025, 46(3): 101-109.
- SUN M X, WANG X Y, LUAN T T, et al. A laser SLAM algorithm for unmanned vehicles based on depth map in dynamic environment [J]. *Chinese Journal of Scientific Instrument*, 2025, 46(3): 101-109.
- [4] 曾庆化, 罗怡雪, 孙克诚, 等. 视觉及其融合惯性的 SLAM 技术发展综述[J]. *南京航空航天大学学报*, 2022, 54(6): 1007-1020.
- ZENG Q H, LUO Y X, SUN K CH, et al. Review on SLAM technology development for vision and its fusion of inertial information[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2022, 54(6): 1007-1020.
- [5] 刘哲, 史殿习, 杨绍武, 等. 视觉惯性导航系统初始化方法综述[J]. *国防科技大学学报*, 2023, 45(2): 15-26.
- LIU ZH, SHI D X, YANG SH W, et al. Review of visual-inertial navigation system initialization method[J]. *Journal of National University of Defense Technology*, 2023, 45(2): 15-26.
- [6] QIN T, LI P L, SHEN SH J. VINS-mono: A robust and versatile monocular visual-inertial state estimator [J]. *IEEE Transactions on Robotics*, 2018, 34(4): 1004-1020.
- [7] CAMPOS C, ELVIRA R, RODRIGUEZ J J G, 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.
- [8] 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.
- [9] 叶俊强, 尤睿, 于明鑫, 等. 基于稀疏直接法的特征增强视觉 SLAM 算法[J]. *仪器仪表学报*, 2023, 44(6): 205-212.
- YE J Q, YOU R, YU M X, et al. Feature-enhanced visual SLAM algorithm base on the sparse direct method[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(6): 205-212.
- [10] 栾添添, 吕奉坤, 班喜程, 等. 高动态环境下的傅里叶梅林变换视觉 SLAM 算法[J]. *仪器仪表学报*, 2023, 44(7): 242-251.
- LUAN T T, LYU F K, BAN X CH, et al. Fourier Merlin transform visual SLAM algorithm in highly dynamic environments [J]. *Chinese Journal of Scientific Instrument*, 2023, 44(7): 242-251.
- [11] 张耀, 吴一全, 陈慧娴. 基于深度学习的视觉同时定位与建图研究进展[J]. *仪器仪表学报*, 2023, 44(7): 214-241.
- ZHANG Y, WU Y Q, CHEN H X. Research progress of visual simultaneous localization and mapping based on deep learning [J]. *Chinese Journal of Scientific Instrument*, 2023, 44(7): 214-241.
- [12] 余洪山, 郭丰, 郭林峰, 等. 融合改进 SuperPoint 网络的鲁棒单目视觉惯性 SLAM[J]. *仪器仪表学报*, 2021, 42(1): 116-126.
- YU H SH, GUO F, GUO L F, et al. Robust monocular visual-inertial SLAM based on the improved SuperPoint network[J]. *Chinese Journal of Scientific Instrument*, 2021, 42(1): 116-126.
- [13] 纪泽源, 于潇颖, 付文兴. 基于图神经网络特征点匹配的视觉 SLAM 算法[J]. *仪器仪表学报*, 2024, 45(9): 34-43.
- JI Z Y, YU X Y, FU W X. Visual SLAM algorithm based on graph neural network feature point matching[J]. *Chinese Journal of Scientific Instrument*, 2024, 45(9): 34-43.
- [14] ZHANG X, DONG H B, ZHANG H X, et al. A real-time, robust, and versatile visual-slam framework based on deep learning networks [J]. *IEEE Transactions on Instrumentation and Measurement*, 2025, 74: 1-13.
- [15] WANG K, ZHANG CH, SU D, et al. VINS-FEN: Monocular visual-inertial SLAM based on feature extraction network[C]. 2023 7th International Conference on Machine Vision and Information Technology, 2023: 86-91.
- [16] LUO H K, LIU Y, GUO CH, et al. SuperVINS: A real-time visual-inertial SLAM framework for challenging imaging conditions [J]. *IEEE Sensors Journal*, 2025, 25(13): 26042-26050.
- [17] GUO CH, HU J L, LUO Y R. Learning-based monocular visual-inertial odometry with SE₂(3)-EKF[J]. *Journal*

of Field Robotics, 2024, 41(6): 1780-1796.

- [18] POTJE G, CADAR F, ARAUJO A, et al. XFeat: Accelerated features for lightweight image matching[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 2682-2691.
- [19] DAI ZH, HUANG X H, CHEN W N, et al. A comparison of cnn-based and hand-crafted keypoint descriptors[C]. International Conference on Robotics and Automation (ICRA), 2019: 2399-2404.
- [20] YU ZH Y, LI J, WANG H CH. Image enhancement based on improved AGCWD algorithm[C]. 2025 IEEE 7th International Conference on Communications, Information System and Computer Engineering, 2025: 755-758.
- [21] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision [C]. International Joint Conference on Artificial intelligence, 1981: 674-679.
- [22] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets [J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [23] MUSA P, AL RAFI F, LAMSANI M. A review: Contrast-limited adaptive histogram equalization (CLAHE) methods to help the application of face recognition[C]. International Conference on Informatics and Computing (ICIC), 2018: 1-6.
- [24] CAI Y H, BIAN H, LIN J, et al. Retinexformer: One-stage retinex-based transformer for low-light image enhancement[C]. IEEE/CVF International Conference on Computer Vision, 2023: 12504-12513.

作者简介



栾添添, 2011年、2018年于哈尔滨工程大学分别获得工学学士学位、博士学位, 现为哈尔滨理工大学副教授, 主要研究方向为激光 SLAM、路径规划。

E-mail: luantiantian@hrbust.edu.cn

Luan Tiantian received her B. Eng. and Ph. D. degrees both from Harbin Engineering University in 2011 and 2018, respectively. She is currently an associate professor at Harbin University of Science and Technology. Her main research interests include LiDAR-based SLAM and path planning.



梁洪杰, 2022年于哈尔滨理工大学获得学士学位, 现于哈尔滨理工大学攻读硕士研究生, 主要研究方向为无人机视觉定位、无人机轨迹规划。

E-mail: 2941323844@qq.com

Liang Hongjie received his B. Sc. degree from Harbin University of Science and Technology in 2022. He is currently pursuing his M. Sc. degree at Harbin University of Science and Technology. His main research interests include UAV visual localization and trajectory planning.



孙明晓(通信作者), 2010年、2018年于哈尔滨工程大学分别获得工学学士学位、博士学位, 现为哈尔滨理工大学副教授, 主要研究方向为激光 SLAM、路径规划。

E-mail: sunmingxiao@hrbust.edu.cn

Sun Mingxiao (Corresponding author) received his B. Eng. and Ph. D. degrees both from Harbin Engineering University in 2010 and 2018, respectively. He is currently an associate professor at Harbin University of Science and Technology. His main research interests include LiDAR-based and path planning.