

无人机弱纹理与动态场景下的视觉惯性里程计系统优化方法研究

潘星宇, 姚雪莲, 杨 艺

江苏理工学院机械工程学院, 江苏 常州

收稿日期: 2025年7月25日; 录用日期: 2025年9月3日; 发布日期: 2025年9月12日

摘 要

针对无人机在GPS信号丢失或不可用的复杂环境(如室内、夜间、动态场景)中的自主定位导航需求, 提出并实现了一套高性能视觉惯性里程计(VIO)系统的优化方法。系统融合视觉传感器与惯性测量单元(IMU)数据, 获取最终无人机位姿信息, 然而传统视觉惯性里程计方法易受弱纹理、动态光照、非高斯噪声与外观变化等影响。该方法在图像前端采用CLAHE局部自适应直方图均衡化与Gamma矫正算法进行图像增强处理, 有效提升弱光与低纹理场景下的图像质量; 同时, 引入GFTT + BRISK组合特征提取策略, 提高特征稳定性与匹配鲁棒性。在后端优化模块中, 引入神经辐射场(NeRF)几何增强模块, 为每个特征点提供深度先验, 协助优化器在弱纹理与动态场景中恢复更加准确和稳定的三维结构估计。通过公开数据集EuRoc测试, 实验结果表明系统在改进算法过后在更多复杂环境会有更好的定位、建图和实时性方面表现。

关键词

SLAM, 无人机, 视觉惯性里程计, 弱纹理环境, 动态场景, VIO优化

Research on the Optimization Method of Visual Inertial Odometer System in Weak Texture and Dynamic Scene of Unmanned Aerial Vehicle

Xingyu Pan, Xuelian Yao, Yi Yang

School of Mechanical Engineering, Jiangsu University of Technology, Changzhou Jiangsu

Received: Jul. 25th, 2025; accepted: Sep. 3rd, 2025; published: Sep. 12th, 2025

文章引用: 潘星宇, 姚雪莲, 杨艺. 无人机弱纹理与动态场景下的视觉惯性里程计系统优化方法研究[J]. 传感器技术与应用, 2025, 13(5): 750-759. DOI: 10.12677/jsta.2025.135073

Abstract

Aiming at the requirements of autonomous positioning and navigation of UAVs in complex environments (such as indoor, night, and dynamic scenes) where GPS signals are lost or unavailable, an optimization method for high-performance visual inertial odometer (VIO) system is proposed and implemented. The system integrates the data of the vision sensor and the inertial measurement unit (IMU) to obtain the final UAV pose information, but the traditional visual inertial odometer method is susceptible to weak texture, dynamic lighting, non-Gaussian noise and appearance changes. The method uses CLAH local adaptive histogram equalization and Gamma correction algorithm to enhance the image at the front end of the image, which effectively improves the image quality in low-light and low-texture scenes. At the same time, the GFTT BRISK combined feature extraction strategy is introduced to improve feature stability and matching robustness. In the back-end optimization module, the Neural Radiation Field (NeRF) geometry enhancement module is introduced to provide depth priors for each feature point, helping the optimizer to restore more accurate and stable 3D structure estimation in weak textures and dynamic scenes. Through the public dataset EuRoc test, the experimental results show that the system will perform better in positioning, mapping and real-time performance in more complex environments after the improved algorithm.

Keywords

SLAM, Drones, Visual Inertial Odometer, Weak Texture Environment, Dynamic Scene, VIO Optimization

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

无人机在室内、夜间、城市峡谷等 GPS 信号受限或不可用的复杂环境中，自主定位与导航能力成为实现工程应用的核心难题。SLAM 是机器人和无人机领域的核心技术之一，现有的 SLAM 技术根据传感器使用主要分为视觉 SLAM [1]和激光 SLAM [2]，视觉惯性里程计(VIO)在后端方法上主要分为基于滤波的 VIO 和基于优化的 VIO。基于滤波的 VIO 主要有：MSCKF [3] (Multi-State Constraint Kalman Filter)，由 Mourikis 和 Roumeliotis 提出，利用多帧约束提升精度，是早期 VIO 经典算法；OKVIS [3]，使用卡尔曼滤波进行状态估计。基于优化的 VIO 主要有 ORB-SLAM3 [4] (视觉 - 惯性部分)，采用图优化，支持多传感器融合。VINS-Mono [5]/VINS-Fusion [6]：采用滑动窗口非线性优化，将视觉重投影残差、IMU 预积分残差和回环残差统一建模，联合优化每帧的位姿、速度、IMU 偏置等状态变量。

传统视觉惯性里程计(VIO)系统在弱纹理、动态场景和资源受限平台下易出现定位漂移、鲁棒性下降等问题。本文基于无人机应用需求，提出面向弱纹理和动态场景的 VIO 系统多模块优化方案，使用图像增强算法，对图像进行预处理，增强图像特征纹理，集成 CLAHE 局部自适应直方图均衡化[7]和 Gamma 矫正算法[8]，对输入图像进行预处理，显著提升夜间、弱光环境下的特征数量和质量。同时使用 GFTT + BRISK 特征组合的特征提取算法代替原模型的特征提取算法，保证特征在弱纹理区域也能选取到关键点，BRISK [9]描述子对旋转、尺度、光照变化具有良好不变性，比原 KLT 光流只用灰度信息更加鲁棒，适合动态、弱光、纹理单调等场景。后端部分滑动窗口与非线性优化使用多线程边缘化计算使用多个线程并

行构建 Hessian 矩阵，提升计算效率，并且在优化中添加 NeRF 残差模块，模块提供每个特征点的深度先验，用残差约束方式协助系统在弱纹理或动态场景下恢复更加准确和稳定的三维结构。通过公开数据集验证，系统在 GPS 丢失环境下实现高精度定位与建图。

2. 算法总框架

本文算法基于 VINS-Fusion 视觉惯性里程计的算法框架实现，输入为相机的图像数据和 IMU 数据，算法前端首先使用 CLAHE 局部自适应直方图均衡化和 Gamma 矫正算法对相机输入的图像进行预处理，增强图像特征纹理。然后使用 GFTT + BRISK 特征提取算法对图像进行特征点提取以及跟踪。IMU 数据包括加速度计和陀螺仪测量值，对 IMU 数据进行预积分，把连续高频 IMU 数据转化为视觉帧间的运动约束残差，实现视觉与惯性信息的高效融合，提升系统定位精度和鲁棒性。对 IMU 数据和图像数据通过时间戳对齐的方式进行时间同步操作。最终前端输出的 IMU 因子、视觉因子、特征点云、关键帧信息作为输入进入后端，后端初始化模块首先对视觉结构进行初始化，使用 SFM 算法和三角化，估算这些图像帧的相对运动和部分特征点的三维位置，通过这些特征点的空间分布，初步构建视觉运动轨迹和稀疏点云。然后进行视觉 - 惯性对齐，正确关联视觉特征点与 IMU 运动信息，使两者在同一参考系下工作，实现高精度融合。最后进行偏置估计用于消除 IMU 固有误差和尺度恢复用于让视觉轨迹获得真实物理长度。后端优化模块将对滑动窗口内的视觉和 IMU 的信息融合，进行非线性联合优化，对系统当前及历史状态的高精度估算。滑动窗口管理模块动态维护最近帧及其状态更新，边缘化历史信息，实现实时高效的联合优化，最终输出位姿、速度、偏置等信息。算法总体框架图如图 1 所示。

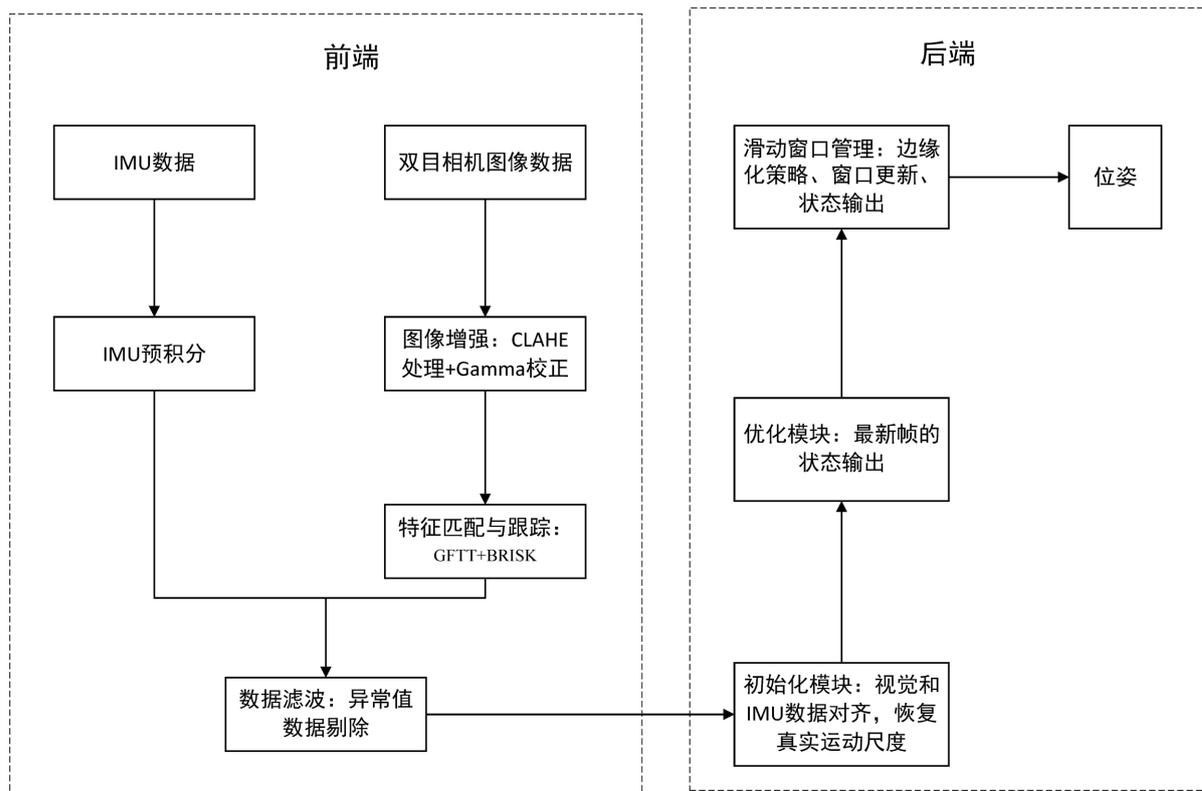


Figure 1. Overall framework diagram of the system
图 1. 系统总体框架图

3. 前端处理

CLAHE(对比度受限的自适应直方图均衡化)是一种常用于图像增强的算法,尤其适用于医学图像、夜间视觉、低光环境等需要提升细节和对比度的场景。将输入的图像分割为多个 $M*N$ 像素大小的小块,对每个小块计算灰度直方图,并设定一个最大高度,超过部分重新分配到其他灰度级,然后对每个小块分别做直方图均衡化。

假设每个像素灰度值范围是 $[0, L-1]$, 其中 L 是灰度级数量(通常为 256), 设输入图像为 $f(x, y)$, 计算计算图像中每个灰度级 r_k 的出现频率 $p_r(r_k)$:

$$p_r(r_k) = \frac{n_k}{MN}$$

其中 n_k 为在 $M*N$ 像素大小的图像中灰度级为 r_k 的像素个数, MN 为图像总像素数, $k = 0, 1, 2, \dots, L-1$ 。

计算累积分布函数(CDF):

$$CDF(r_k) = \sum_{j=0}^k p_r(r_j)$$

变换后的灰度级为

$$s_k = (L-1) \times CDF(r_k)$$

其中 s_k 是像素 r_k 对应的均衡化后新灰度级。

最后将增强后的结果,通过对每个像素的输出灰度值采用其四个邻近小块的映射结果进行双线性插值平滑的拼接,避免块状边界。能有效提升输入图像的局部对比度,增强细节,优化特征提取和跟踪。

Gamma 校正其核心作用是调节图像的亮度非线性响应,使之更加符合人眼视觉特性和设备属性,有效提升图像细节、视觉一致性和处理效果。

对于每一个像素,其灰度值 s_k 的范围是 $[0, L-1]$, 其中 L 是灰度级数量(通常为 256), 首先归一化转换为范围是 $[0, 1]$, 然后最终输出 Gamma 校正的变换图像的灰度值为:

$$s'_k = (s_k + \varepsilon)^\gamma$$

其中 s_k 为输入图像的灰度值, s'_k 为输出图像的灰度值, γ 为 Gamma 校正系数, 决定亮度非线性变换的强

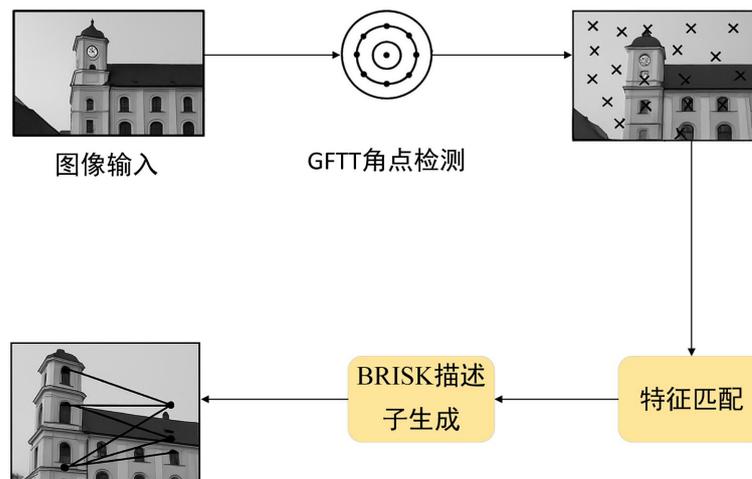


Figure 2. Schematic diagram of GFTT + BRISK feature detection and matching
图 2. GFTT + BRISK 特征检测与匹配示意图

度, ε 为补偿系数, 根据图像亮度动态调节。

GFTT 是特征点检测器, 用于选择角点或质心明显的区域作为跟踪点, 通过计算图像块的梯度协方差矩阵, 选出响应值最大的点。对光照变化具有一定鲁棒性, 计算快速, 适合弱纹理环境, 在夜间、室内、重复图案较少的环境中仍能检测到有用特征点。BRISK 是特征描述子, 用于表达每个特征点周围局部图像结构, 支持特征匹配, 匹配时采用 Hamming 距离, 速度快, 鲁棒性强, 具备尺度和旋转不变性, 对模糊、轻微变形具备较好鲁棒性, 适合动态环境。GFTT + BRISK 的特征检测与匹配示意图如图 2 所示。

4. 后端处理

后端优化模块通过滑动窗口优化, 将视觉测量(特征点观测)、IMU 预积分约束等多源信息联合起来, 精确估算每一时刻的相机/IMU 的位姿、速度、偏置等状态。从帧 i 到帧 j , 使用预积分项构建运动预测残差:

位置残差:

$$r_p = \mathbf{p}_j^w - \mathbf{p}_i^w - \mathbf{v}_i \Delta t - \frac{1}{2} \mathbf{g} (\Delta t)^2 - \mathbf{R}_w^i \boldsymbol{\alpha}_{ij}$$

速度残差:

$$r_v = \mathbf{v}_j^w - \mathbf{v}_i^w - \mathbf{g} \Delta t - \mathbf{R}_w^i \boldsymbol{\beta}_{ij}$$

姿态残差:

$$r_q = \log \left(\gamma_{ij}^{-1} \cdot (\mathbf{q}_i^{w-1} \cdot \mathbf{q}_j^w) \right)$$

其中: $\boldsymbol{\alpha}_{ij}$ 、 $\boldsymbol{\beta}_{ij}$ 、 γ_{ij} 是 IMU 预积分项, $\log(\cdot)$ 表示 SO(3) 李群上的误差对数映射, \mathbf{R}_w^i 表示第 k 帧在世界坐标系下的旋转矩阵。

将估算得到的三维空间点通过相机模型和估算的相机位姿投影到图像平面, 与实际观测到的像素点位置相比, 两者的距离就是重投影误差。对于视觉重投影残差项, 对第 j 个特征点在图像第 k 帧中的观测构造投影误差:

$$r_{\text{proj}} = u_k^j - \pi \left(\mathbf{R}_w^k \cdot (\lambda_j \cdot \mathbf{d}_j) + \mathbf{p}_k^w \right)$$

其中 u_k^j 代表第 k 帧图像中特征点 j 的图像坐标观测, \mathbf{R}_w^k 第 k 帧的相机旋转矩阵, λ_j 为特征点逆深度, 用于三维点构建, 在首次进入滑动窗口块, $\lambda_j = \frac{1}{\|\mathbf{X}_k\|}$, \mathbf{d}_j 为单位方向向量, $\lambda_j \cdot \mathbf{d}_j$ 表示第 j 个特征点三维位置, \mathbf{p}_k^w 第 k 帧相机在世界系下的位置, $\pi(\cdot)$ 为相机投影函数。

NeRF 几何增强残差项, NeRF 模块为每个特征点提供深度先验 d_j^{nerf} , NeRF(神经辐射场)模型推理得到的第 j 个特征点对应的空间深度, 与视觉估计值 $d_j^{\text{opt}} = \frac{1}{\lambda_j}$, 即优化器当前估计出的特征点三维深度构造几何一致性残差:

$$r_j^{\text{nerf}} = \frac{1}{\lambda_j} - d_j^{\text{nerf}}$$

该残差项作为软约束加入优化器, 选用 Huber Loss 赋予稳健权重。当特征点弱纹理或被遮挡时, 该项能有效引导优化器收敛至正确三维结构。

边缘化先验项，当首次进入滑动窗口块，不加入边缘化先验项，即 $r_{\text{prior}} = 0$ 。在滑窗中，被剔除帧的信息被保留为先验，用于构造边缘化残差：

$$r_{\text{prior}} = H \cdot X - b$$

其中 X 为当前优化变量向量， X 为信息矩阵，表示边缘化后对当前状态变量施加的先验约束权重， b 为边缘化误差项，表示历史观测残差对当前变量的影响，通过 Schur 消元与信息矩阵更新实现。

总体优化目标函数：

$$\min_X \left[\sum_{(i,j) \in \text{IMU}} \|r_p\|^2 + \|r_v\|^2 + \|r_q\|^2 + \sum_{(k,j) \in \text{Visual}} \|r_{\text{proj}}\|^2 + \sum_j \|r_j^{\text{nerf}}\|^2 + \|r_{\text{prior}}\|^2 \right]$$

对于该非线性最小二乘的优化问题，采用 Ceres 优化器进行迭代求解，最后输出更加精确的状态估计： p_k^w , q_k^w , v_k^w , b_a^i , b_g^i , λ_j 。

窗口内包含最近 N 帧的状态(位姿、速度、IMU 偏置等)以及被这些帧观测到的部分地图点。每当新帧进入，最旧的一帧被移出，同时进行一次批量优化。每帧的状态为 $x = [p_i^w, v_i^w, q_i^w, b_a^i, b_g^i]$ ，地图点： $X_k = \lambda_j \cdot d_j$ ，当新帧进入窗口，新增状态变量，并将最旧帧移出窗口，移除变量，但用边缘化保留对全局估计的约束信息。设边缘化变量 X_m ，保留变量为 X_k ，联合高斯误差项 $E(x_m, x_k)$ ，通过 Schur 消元与信息矩阵更新得到 H 和 b ，并作为先验项继续参与后续窗口的优化。

5. 实验验证

使用公开数据集 EuRoc 进行评估，与 VINS-Fusion 算法进行比较评估，验证该算法的效率以及精度。所有测试在 intel i5-13500H、2.6GHzCPU 频率、Ubuntu20.04 操作系统、32G 运行内存的硬件条件下进行测试。

通过该算法和 VINS-Fusion 在数据集 EuRoc 下进行对比实验。通过对比在每组数据集的估计轨迹和真实轨迹之间的均方根平移误差和均方根旋转误差用于评估算法的估计精度。

Table 1. Comparison of absolute trajectory root mean square error of algorithm

表 1. 算法绝对轨迹均方根误差对比

序列	VINS-Fusion		改进算法		改进	
	平移误差/%	旋转误差/[°]/m	平移误差/%	旋转误差/[°]/m	平移误差	旋转误差
MH_01	0.255499	4.325622	0.248064	3.961001	2.91%	8.43%
MH_02	0.205331	3.654905	0.204469	3.796471	-19.06%	-3.87%
MH_04	0.407661	4.152358	0.401037	4.478853	1.63%	-7.86%
V1_01	0.245503	8.050611	0.239942	8.393355	2.27%	-4.26%
V1_02	0.517003	13.234568	0.499081	13.548226	3.47%	-2.37%
V1_03	0.429837	12.87	0.397423	13.299286	7.54%	-3.34%
V2_01	0.19561	5.315078	0.20462	5.691085	-4.61%	-7.07%
V2_02	0.383327	10.687355	0.363096	10.299223	5.28%	3.63%
V2_03	0.532300	11.530759	0.451518	11.189953	15.18%	2.96%

表 1 为该算法和 VINS-Fusion 算法在运行 EuRoc 数据集的绝对轨迹的均方根平移误差和均方根旋转

误差, 实验结果表明在一些简单场景如 MH_01、V1_01、V1_02, 特征纹理较明显的场景下, 改进算法效

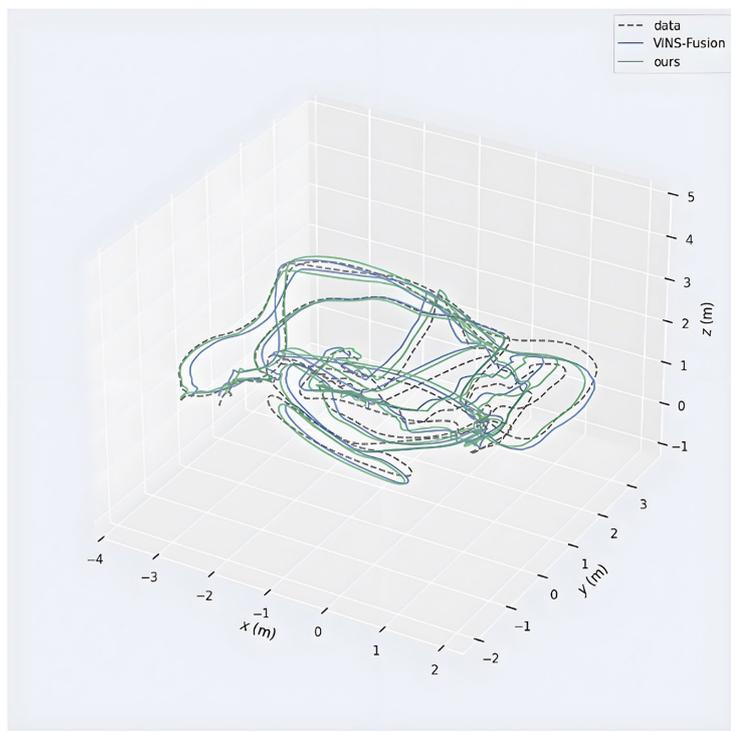


Figure 3. Comparison of 3D trajectories on V2_03 datasets
图 3. 在 V2_03 数据集上的三维轨迹对比图

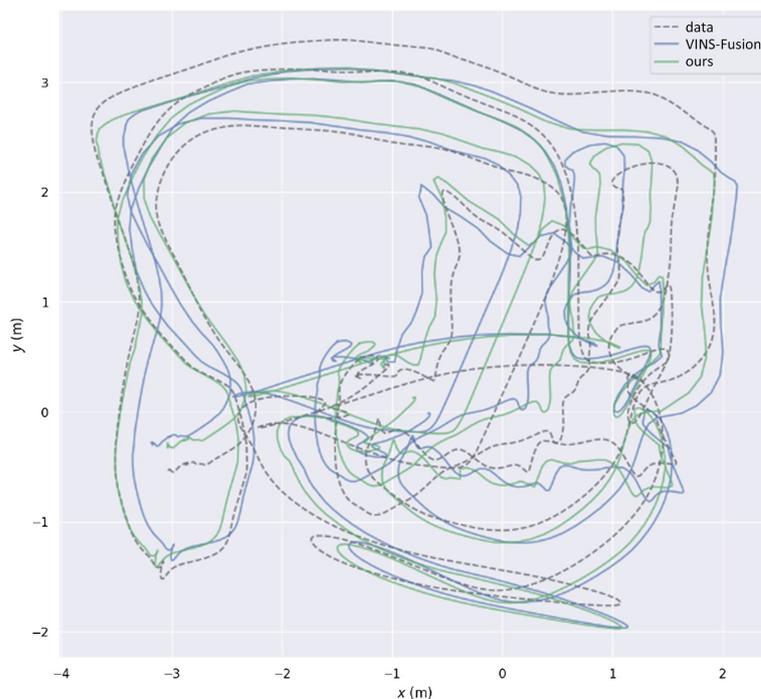


Figure 4. Comparison of 2D trajectories on V2_03 datasets
图 4. 在 V2_03 数据集上的二维轨迹对比图

果并没有太大的改善或者变得不如原算法，在一些复杂环境，弱纹理低光照的环境下，如 V2_02、V2_03 在精度上有一定的提升，在均方根平移误差最高提升了 15.18%。实验表明，通过改进过后的算法，在特征提取上取得了一定的改善，在弱纹理低光照的动态场景下有一定的优势。

图 3 展示了原版算法和改进算法在 V2_03 数据集上的三维轨迹对比图，图 4 展示了原版算法和改进算法在 V2_03 数据集上的在二维平面的轨迹对比图，图 5 展示了原版算法和改进算法在 V2_03 数据集上每个平移分量和每个旋转分量的跟踪情况，图 6 展示了原版算法和改进算法在 V2_03 数据集上平移误差的折线对比图，图 7 展示了原版算法和改进算法在 V2_03 数据集上旋转误差的折线对比图。直观的展示了改进算法相较于 VINS-Fusion 算法有更好的精度表现，尤其是在 V2_03 数据集上。

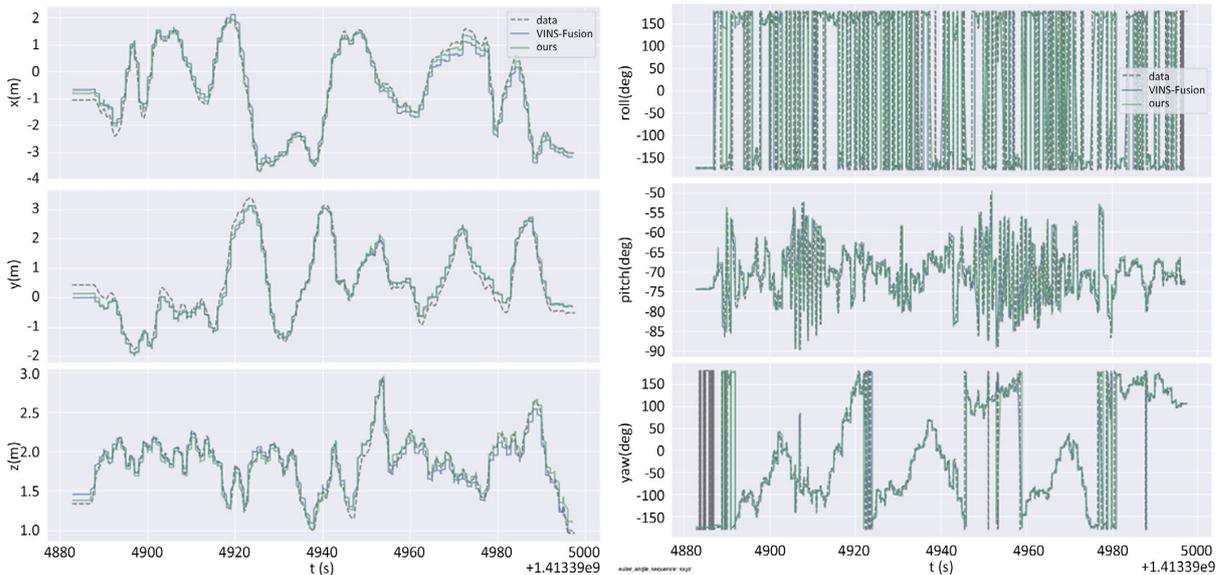


Figure 5. Comparison of the tracking of translation and rotation components on V2_03 dataset

图 5. 在 V2_03 数据集上的平移分量和旋转分量的跟踪情况对比

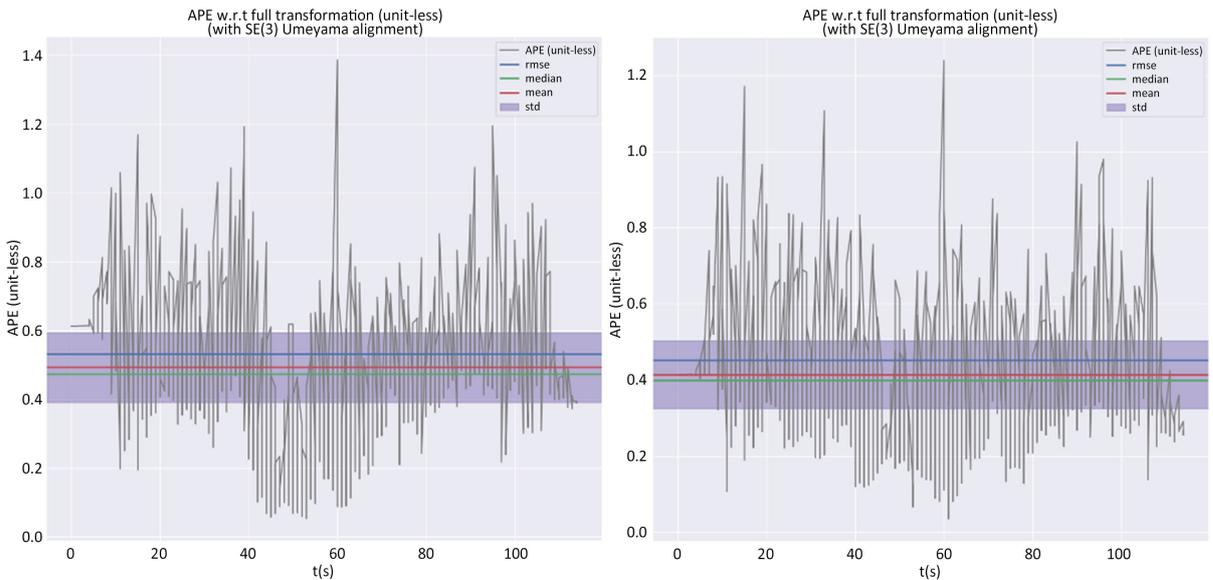


Figure 6. Comparison of the line of translation error on the V2_03 dataset

图 6. 在 V2_03 数据集上平移误差的折线对比

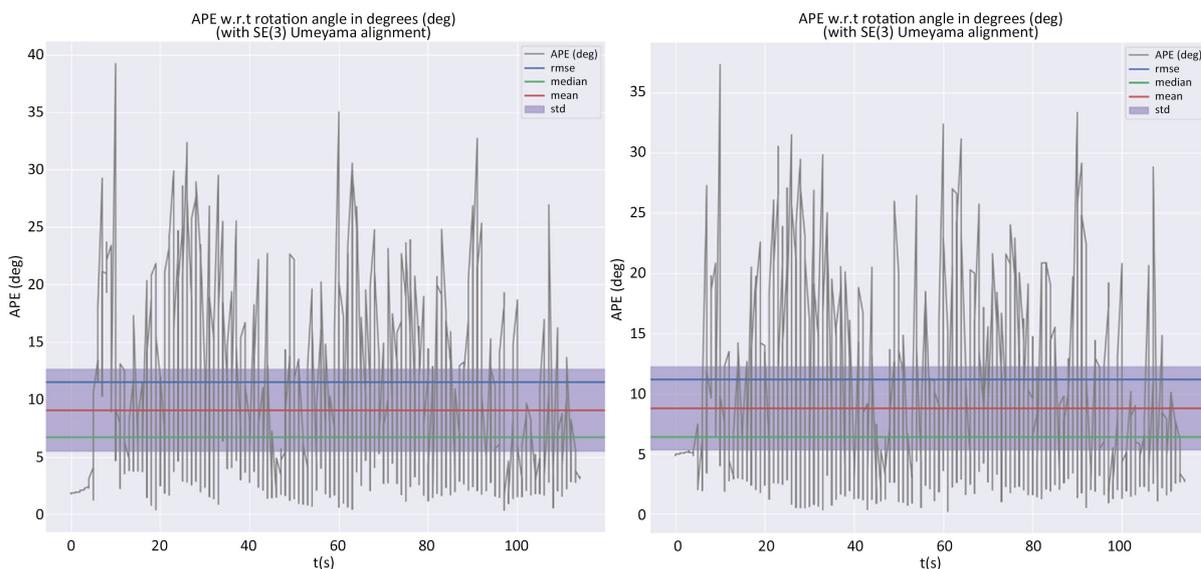


Figure 7. Comparison of the lines of rotation errors on V2_03 datasets

图 7. 在 V2_03 数据集上旋转误差的折线对比

6. 实验结论

对于面向无人机应用的视觉惯性里程计系统, 针对无人机在 GPS 信号丢失或不可用的复杂环境(如室内、夜间、动态场景)中获取准确的定位信息的问题, 系统在算法、结构和工程实现上进行了改进, 有效提升了复杂环境下的定位精度和鲁棒性。实验结果表明, 系统在弱纹理、动态场景、资源受限平台上均具有显著优势, 各技术模块协同提升了整体定位与导航能力。主要贡献如下:

1) 在视觉前端中, 在进行特征匹配跟踪之前, 对获取图像进行图像增强处理, 增强图像的纹理特征以及图像的亮度, 提高获取特征点的数量, 为后续计算位姿信息提高更多的准确数据。

2) 使用 GFTT + BRISK 特征组合的特征提取算法代替原模型的特征提取算法, 解决视觉惯性里程计在弱纹理与动态场景下出现的特征难以提取, 导致位姿信息估计不准确的问题, 保证特征在弱纹理区域也能选取到关键点。

3) 在优化中加入 NeRF(神经辐射场)模型, 作用是视觉惯性系统中的三维点提供结构先验——也就是每个特征点的空间深度参考, 当无人机处于弱纹理或遮挡场景时, 传统视觉重投影残差容易退化, 导致三角化精度下降。此时 NeRF 可以从历史图像生成一个真实场景的几何估计, 为特征点提供深度值, 通过优化过后收敛到符合真实结构的值, 使滑窗中三维点更准确, 系统整体更鲁棒。

参考文献

- [1] 张彧, 檀祖冰, 曹东璞, 等. 基于视觉和惯性测量单元的里程计关键技术研究综述[J]. 机械工程学报, 2024, 60(10): 3-21.
- [2] 刘铭哲, 徐光辉, 唐堂, 等. 激光雷达 SLAM 算法综述[J]. 计算机工程与应用, 2024, 60(1): 1-14.
- [3] Mourikis, A.I. and Roumeliotis, S.I. (2007) A Multi-State Constraint Kalman Filter for Vision-Aided Inertial Navigation. *The 24th IEEE International Conference on Robotics and Automation (ICRA)*, Rome, 10-14 April 2007, 3565-3572. <https://doi.org/10.1109/robot.2007.364024>
- [4] Campos, C., Elvira, R., Rodriguez, J.J.G., M. Montiel, J.M. and D. Tardos, J. (2021) ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-inertial, and Multimap Slam. *IEEE Transactions on Robotics*, 37, 1874-1890. <https://doi.org/10.1109/tro.2021.3075644>
- [5] Qin, T., Li, P. and Shen, S. (2018) VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE*

Transactions on Robotics, **34**, 1004-1020. <https://doi.org/10.1109/tro.2018.2853729>

- [6] Qin, T., Pan, J., Cao, S., *et al.* (2019) A General Optimization-Based Framework for Local Odometry Estimation with Multiple Sensors. arXiv: 1901.03638.
- [7] Wang, W. and Yang, Y. (2023) A Histogram Equalization Model for Color Image Contrast Enhancement. *Signal, Image and Video Processing*, **18**, 1725-1732. <https://doi.org/10.1007/s11760-023-02881-9>
- [8] 马鑫, 喻春雨, 陈刚, 等. 分区自适应伽马校正的非均匀光照图像增强[J]. *激光与光电子学进展*, 2024, 61(14): 383-390.
- [9] 杨永奎. BRISK 特征点检测匹配算法的探究[J]. *建设机械技术与管理*, 2016, 29(5): 77-80.