

低纹理动态场景下基于特征增强的视觉惯导SLAM方法

解明扬, 徐鑫, 杨晨, 余子锐, 王焱

(南京航空航天大学自动化学院, 江苏 南京 211106)

摘要: 针对低纹理动态场景下SLAM系统难以提取足够多的稳定特征点且易发生错误匹配, 进而导致系统位姿估计精度与鲁棒性不足的难题, 提出一种面向低纹理动态场景的单目视觉惯导SLAM系统, 实现特征提取匹配与动态特征点判别的性能提升。首先, 用基于深度学习的SuperPoint特征提取与LightGlue特征匹配模块代替现有ORB-SLAM3系统前端, 提升低纹理区域特征提取与匹配的鲁棒性; 其次, 结合YOLOv8-seg实现动态区域语义分割, 利用IMU预积分估计相机位姿变化, 通过构建联合的动态点剔除机制实现更细粒度的动态点剔除, 提高动态干扰场景下系统的精度和鲁棒性。最后, 在公开数据集和实际场景中进行对比实验, 验证所提方法的性能。结果表明, 相较于现有基于视觉及视觉惯导的SLAM方法, 所提系统在实际低纹理动态场景中绝对轨迹误差的均方根误差降低88.4%及以上, 标准差降低90%及以上, 具有更高的定位精度和鲁棒性。

关键词: 动态SLAM; 低纹理; 深度学习; 特征提取; 语义分割; IMU预积分

中图分类号: TP242.6

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202543

Feature-enhanced visual-inertial SLAM method for low-texture dynamic environment

XIE Mingyang, XU Xin, YANG Chen, YU Zirui, WANG Yao

College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Abstract: To address the challenge of SLAM systems struggling to extract sufficient stable feature points and being prone to incorrect matching in low-texture dynamic environments, which leads to poor accuracy and robustness in pose estimation, this paper proposes a monocular visual-inertial SLAM system specifically designed for low-texture dynamic scenarios. The proposed method achieves improved performance in feature extraction, matching, and dynamic feature point discrimination. First, a deep learning-based SuperPoint feature extraction and LightGlue feature matching modules are employed to replace the existing ORB-SLAM3 frontend, significantly enhancing the robustness of feature extraction and matching in weak texture areas. Second, by integrating YOLO-seg for dynamic region semantic segmentation and leveraging IMU pre-integration to estimate camera pose changes, a joint dynamic point removal mechanism is constructed to achieve finer-grained dynamic point filtering, thereby enhancing system accuracy and robustness in dynamic interference scenarios. Finally, the performance of the proposed method was validated through comparative experiments on public datasets and real-world scenarios, and the proposed system achieves a reduction of 88.4% or more in the root mean square error of absolute trajectory error, and 90% or more in the standard deviation in real-world low-texture dynamic scenarios, exhibiting supe-

收稿日期: 2025-06-21; 修回日期: 2025-12-04

通信作者: 解明扬, myxie@nuaa.edu.cn

基金项目: 国家重点研发计划项目 (No.2023YFB4704400); 国家自然科学基金项目 (No.62373186); 江苏省自然科学基金项目 (No.BK20231440); 中央高校基本科研业务费 (No.NZ2024-033)

Foundation Items: The National Key Research and Development Program of China (No.2023YFB4704400), The National Natural Science Foundation of China (No.62373186), The Natural Science Foundation of Jiangsu Province (No.BK20231440), The Fundamental Research Funds for the Central Universities (No.NZ2024-033)

rior positioning accuracy and robustness when compared with existing visual and visual-inertial SLAM approaches.

Key words: dynamic SLAM, low-texture, deep learning, feature extraction, semantic segmentation, IMU pre-integration

0 引言

同时定位与地图构建 (simultaneous localization and mapping, SLAM)^[1]是移动机器人、增强现实等许多任务的基础技术。特别是对于移动机器人自主导航任务来说, SLAM能够为其提供精确的定位功能并构建环境地图。目前针对纹理丰富且静态的场景, 视觉SLAM已经得到很好的应用。然而在低纹理场景中, 例如建筑物内的长直走廊, 缺乏明显角点且存在大量重复纹理特征, SLAM系统前端特征提取模块难以提取足够多的特征点, 导致可用于位姿估计的约束不足, 使得后端非线性最小二乘优化问题难以求解, 最终导致相机位姿估计精度下降, SLAM系统常常难以正常运行^[2-4]。此外, 在动态场景下, 大尺度视差极易导致物体特征点与背景点误匹配, 引入大量错误约束, 严重影响位姿估计的精度, 削弱地图的一致性与长期鲁棒性^[5]。

在低纹理环境下, 基于灰度图像的视觉里程计 (visual odometry, VO) 方法难以提取足够多的特征点, 因此, 有学者引入深度信息辅助特征提取方法, 从而提升建图和定位能力。Jiang等^[4]利用特征点的深度差以及惯性测量单元 (inertial measurement unit, IMU) 预积分, 提出了一种全新的深度特征点匹配方法, 但并未考虑动态场景。此外, 国内外研究人员尝试从有限的图像中提取更稳定的直线特征来建立特征关联。Pumarola等^[6]在ORB-SLAM2 (oriented FAST and rotated BRIEF-based simultaneous localization and mapping 2) 的基础上扩展了线特征, 显著提升了系统在低纹理场景下的特征提取能力。He等^[7]提出基于线特征融合的PL-VIO (point-line visual-inertial odometry) 系统, 提升了低纹理场景下系统的精度和鲁棒性。Fu等^[8]提出了基于优化的线特征提取与融合的PL-VINS (point-line visual-inertial navigation system), 减少了跟踪时间。现有特征提取与匹配算法在特征稳定的丰富场景中表现出优良性能, 但是在低纹理、光照变化大的场景中, 难以提取到丰富且稳定的特征点, 从而导致系统漂移或跟踪失败。此外, 这些特征提取和匹配往往缺乏高级语义信息和全局一致性, 难以适应大尺度视差和非刚体变形的情况。

近年来, 随着深度学习的发展, 视觉SLAM中关于特征点提取与匹配的研究取得了显著进展。SuperPoint作为一种自监督特征点提取方法^[9], 因其鲁棒性强和端到端能力强被广泛应用。Tyszkiewicz等^[10]提出DISK (discrete keypoints) 方法, 该方法虽具备高密度特征点检测的能力, 但在SLAM中匹配效率不足; LoFTR^[11]和Efficient LoFTR^[12]方法利用Transformer实现高精度匹配, 但由于其计算开销大, 难以在实时系统中部署。为了提升计算效率, Potje等^[13]提出了基于轻量CNN的XFeat, 但受限于低维描述符, XFeat在闭环检测中性能有限。在特征匹配方面, SuperGlue网络^[14]通过引入图神经网络提升了匹配精度, 但在低纹理场景中性能仍需提升; Lindenberger等^[15]提出的LightGlue则通过简化结构兼顾了匹配精度与实时性, 成为当前轻量化特征匹配的主流方案。此外, Jiang等^[16]提出的OmniGlue提升了匹配网络的泛化能力, 但存在匹配速度低、难以满足实时性要求的问题。

动态场景下SLAM解决方法主要分为纯视觉方法与多传感器融合方法两类^[17]。纯视觉方法依赖图像信息本身, 又可以细分为基于学习、传统数学以及两者融合的方法。Yu等^[18]提出面向动态场景的DS-SLAM系统, 采用SegNet对图像进行实时分割, 从而提前消除潜在动态区域对系统精度和鲁棒性的影响。Kundu等^[19]提出基于极线约束和光流向量差异的动态目标检测方法; Zou等^[20]采用三角化方法对特征点的深度信息一致性进行检验以区分动态特征与静态特征; Narayana等^[21]则基于光流方向定义了概率模型以估计独立运动物体的个数。李嘉铭^[22]在ORB-SLAM2的基础上, 利用目标检测得到动态物体掩码框, 并为每个地图点设置运动概率来判断其移动状态。此外, 一些工作尝试融合语义与传统方法来提升动态点判别方法的准确性。Xiao等^[23]提出Dynamic-SLAM系统, 采用基于融合语义与几何约束的联合优化策略来共同检测动态点; Bescos等^[24]提出的DynaSLAM系统通过结合Mask R-CNN的语义信息与多视图几何一致性, 精确剔除了动态物体, 并完成遮挡背景修复。Kim等^[25]利用IMU传感器的角速度信息进行图像旋转补偿, 并结合轮式传感器的速度信息实现静态

特征点识别；Xu等^[26]提出面向动态场景的OD-SLAM系统，利用轮式里程计直接选取静态特征点进行位姿估计；Chavez-Garcia等^[27]采用融合激光雷达、毫米波雷达与摄像头信息实现多模态动态目标检测。

可以看出，目前单一低纹理和动态场景中的SLAM技术研究得到了极大发展，但是面向低纹理动态场景仍然没有很好的解决方案。现有多数SLAM方法受限于局部信息，难以将特征与环境信息建立联系，亟须采用基于深度学习的方法获取高级特征。此外，目前动态SLAM方法多采用直接剔除动态掩码中特征点的策略，极易将动态框内的静态点误删除，严重影响系统在低纹理环境中的鲁棒性。

本文提出一种特征增强的视觉惯导SLAM策略，以解决动态低纹理场景下特征点关联弱和容易误删除的问题。如图1所示，将ORB-SLAM3的单目IMU模式作为基础框架，将原系统中基于ORB（oriented FAST and rotated BRIEF）的特征提取与BF（brute-force）匹配方式强化为基于深度学习的SuperPoint特征提取与LightGlue特征匹配；针对特征点容易被误删除的问题，利用语义分割和IMU预积分构建联合动态特征点剔除策略，实现动态特征点精准剔除。

本文主要贡献如下。

（1）系统地研究低纹理动态场景中的视觉SLAM解决方案，包括前端特征提取与后端动态特征点剔除；同时实验验证了本文提出的系统在低纹理动态场景下的定位精度与鲁棒性。

（2）引入基于SuperPoint和LightGlue的特征提取与匹配模块，有效提升了低纹理动态环境中单

目视觉SLAM系统特征提取与匹配的稳定性 and 准确率。

（3）构建了基于语义分割掩膜与IMU特征点一致性检测结合的联合动态特征点筛选策略，实现了掩膜区域内动态点的高细粒度剔除，最大化保留静态特征点。

1 基于深度学习的特征提取与匹配

1.1 基于SuperPoint网络的特征点提取

SuperPoint是一种全卷积自监督网络模型^[9]，主要由共享编码器、特征点解码器和描述子解码器构成。SuperPoint网络结构如图2所示，输入图像大小为 $W \times H \times 1$ ，其中， W 表示图像宽度， H 表示图像高度，1表示通道个数。共享编码器采用VGG风格的网络结构进行特征提取，包括卷积层、下采样和非线性激活，两个独立的解码器共用同一个编码器。特征点解码器接收来自共享编码器的 $W/8 \times H/8 \times 128$ 维张量，再经过两次卷积之后得到 $W/8 \times H/8 \times 65$ 维张量，张量每一个像素点的65个通道对应原图中局部互不重叠的 8×8 共64个像素块，以及1个额外的用于判断该像素块是否含有特征点的dustbin通道。在对通道进行softmax之后去掉dustbin通道，并将剩下的64个通道重新拼成完整的图像，最终得到所有特征点的位置。描述子的解码器同样接收共享编码器的输出，再进行卷积和双3次线性差值后得到 $W \times H \times D$ 维度的描述子（ D 表示描述子维度），进行L2范数归一化之后，得到最终需要的描述子。

传统基于ORB的方法在低纹理区域往往难以提取足够多的特征点，导致约束不足；尺度不变特

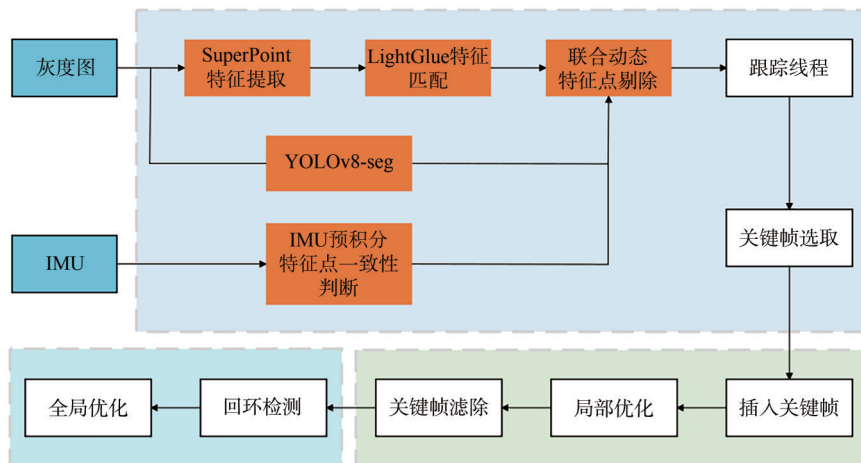


图1 基于特征增强的视觉惯导SLAM系统整体架构

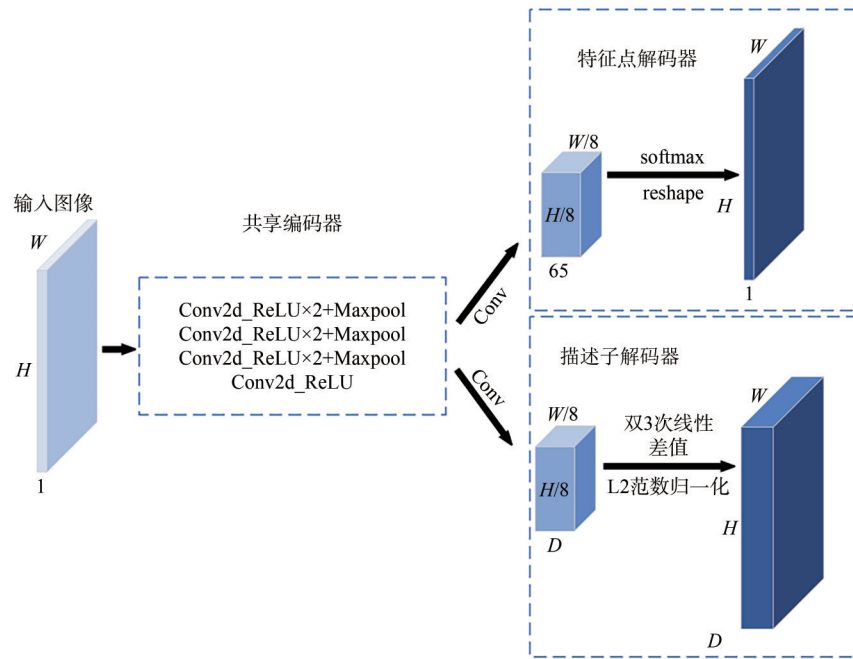


图2 SuperPoint网络结构示意图

征转换（scale-invariant feature transform, SIFT）等方法虽具备更强的尺度与光照鲁棒性，但计算开销较大；DISK等深度学习特征提取方法能够生成更多的特征点，但实时性不足。综合考虑特征提取的稳定性、实时性与部署成本，本文选取SuperPoint作为特征点提取模块。为了实现SuperPoint在SLAM系统中实时高效运行，本文将原始模型导出为ONNX格式，并基于C++ ONNX Runtime推理框架完成部署；对输入张量进行归一化预处理与输出解码，确保该模块能够在本文系统中实时稳定运行，为后续LightGlue特征匹配提供可靠基础。

1.2 基于LightGlue网络的特征匹配

LightGlue网络^[15]在SuperGlue网络的基础上进行改进，是一种基于Transformer的图匹配网络，其核心思想是通过自注意力与交叉注意力机制，对图像间的关键点进行上下文建模，从而获得一致性

更强、空间几何更合理的匹配结果。其网络架构如图3所示，主要包括特征编码、上下文增强与动态匹配决策。首先，将提取到的SuperPoint关键点位置与其对应的256维描述子进行融合，通过线性映射使增强之后的描述子具备空间感知能力。然后，分别对两张图的描述子进行多层自注意力和交叉注意力编码，增强每个点对自身邻域以及跨图匹配关系的理解，使网络能够综合双图特征进行精确匹配。最后，进行匹配置信度预测与剪枝融合，使网络在匹配精度与计算效率之间取得良好平衡，并具备在特征稀疏或不完整时提前终止推理的能力，从而提升整体匹配效率与鲁棒性。

与传统的暴力匹配或FLANN匹配器相比，LightGlue网络基于Transformer架构，能够利用全局上下文信息来提升匹配精度。在深度学习方法中，SuperGlue具备较高精度，但计算复杂度高；

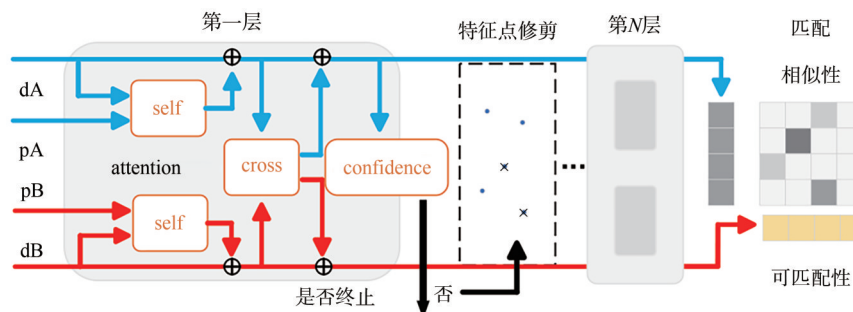


图3 LightGlue网络结构示意图

LoFTR作为端到端检测器可以直接输出稠密匹配，在图像配准任务上表现优异，但计算开销大、实时性差。相比之下，LightGlue通过逐层早停机制，在保证匹配精度的同时显著降低计算开销，更适合低纹理动态场景下的实时SLAM系统。LightGlue模块同样被导出为ONNX格式，并通过C++集成至SLAM前端，模块接收SuperPoint输出的特征点位置和对应的描述子，通过ONNX Runtime实现高效推理。

2 联合动态特征点剔除策略

虽然SuperPoint与LightGlue能够在低纹理区域提取并且匹配到更多高质量的特征点，但是在动态场景中，动态物体的移动极易造成特征点误匹配，严重影响系统精度与鲁棒性。本节首先引入轻量级语义分割模型YOLOv8-seg^[28]对图像中潜在的动态目标进行像素级分割；接着利用IMU预积分对特征点进行一致性判断，解决语义分割带来的静态特征点误删除问题；最后结合语义分割与特征点一致性判断，实现动态特征点的高效剔除。

2.1 基于YOLOv8-seg的动态区域语义分割

由于不受深度值的影响，相较于目标检测与深度聚类结合的分割方法，语义分割方法具备端到端的检测能力，并且具有更高的检测精度与边界表达能力。如图4、图5所示，图4(c)出现漏检，基于语义分割的动态区域检测方法相比于目标检测与深度聚类结合的策略，具有更高的空间分辨率与更精细的边界刻画能力，能够有效覆盖目标的全部动态区域，降低漏检风险，从而提升动态场景中的检测精度和鲁棒性。

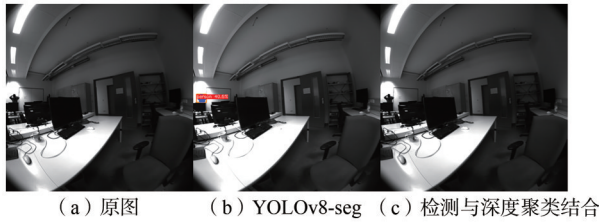


图4 TUM-VI数据集掩码检测结果

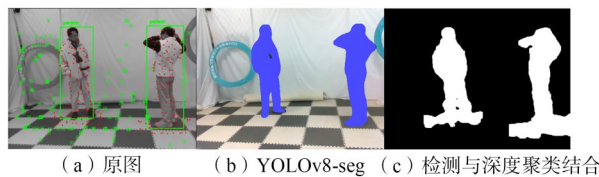


图5 D4数据集掩码检测结果

值得注意的是，相较于目前最新版本的YOLO算法，YOLOv8-seg是当前YOLO系列中分割性能最稳定的模型，而且具有实时性强的优点。在兼顾精度、实时性和边界质量的情况下，本文选用与系统硬件更适配的YOLOv8-seg算法。此外，选择成熟稳定的版本是当前诸多优秀动态SLAM系统的普遍做法，例如，Huang等^[29]在ADM-SLAM中仍使用Deeplabv3^[30]，Liu等^[31]采用YOLOv3^[32]进行目标检测，Cong^[33]等使用YOLOv5s进行语义分割。

在本系统中，YOLOv8-seg被转为ONNX格式，并使用英伟达推出的TensorRT进行部署，在RTX-3080 Laptop GPU上以大于100 FPS的帧率稳定运行，满足SLAM系统的实时性要求。

2.2 基于IMU预积分的特征点一致性判断

语义分割模块可以精确分割出动态区域，但存在误删除动态区域中静态特征点的问题，尤其在低纹理动态场景中更易造成约束不足、位姿优化精度降低的问题。本节利用帧间惯性测量预测相机的相对位姿，并将上一帧中特征点在当前帧的预测位置与相机实际得到的观测进行对比，进而识别出不满足一致性的真正的动态特征点。其原理如图6所示。

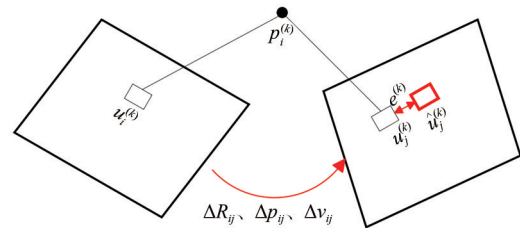


图6 基于IMU预积分的特征点一致性检测示意图

在视觉惯导SLAM中，为了避免重复频繁积分，常使用IMU预积分（IMU pre-integration）技术，预积分将两帧之间的IMU数据累加起来得到帧间的位姿变换。

$$\begin{cases} \Delta R_{ij} \doteq R_i^\top R_j = \prod_{k=i}^{j-1} \exp\left(\left(\tilde{\omega}_k - b_{g,k} - \eta_{gd,k}\right) \Delta t\right) \\ \Delta v_{ij} \doteq R_i^\top \left(v_j - v_i - g \Delta t_{ij}\right) = \sum_{k=i}^{j-1} \Delta R_{ij} \left(\tilde{a}_k - b_{a,k} - \eta_{ad,k}\right) \Delta t \\ \Delta p_{ij} \doteq R_i^\top \left(p_j - p_i - v_i \Delta t_{ij} - \frac{1}{2} \sum_{k=i}^{j-1} g \Delta t^2\right) \end{cases} \quad (1)$$

其中， ΔR_{ij} 表示从时刻*i*到时刻*j*的相对旋转预积分量，定义为当前帧旋转 R_j 相对于起始帧旋转 R_i 的

变换; $\tilde{\omega}_k$ 是第 k 时刻的原始陀螺仪角速度观测值; $b_{g,k}$ 为对应时刻的陀螺仪偏置; $\eta_{gd,k}$ 为陀螺仪观测噪声; Δt 为采样间隔; $\Delta \mathbf{v}_{ij}$ 表示速度预积分量, 定义为两帧之间相对速度变化; $\Delta \mathbf{R}_{ik}$ 表示从 i 到 j 中任意 k 时刻到 i 帧的旋转预积分增量; $\tilde{\mathbf{a}}_k$ 是加速度观测; $b_{a,k}$ 和 $\eta_{ad,k}$ 分别为加速度计偏置与噪声; $\Delta \mathbf{p}_{ij}$ 表示两帧之间的位置变化。

定义预积分避免了优化过程中的重复积分, 极大降低了计算的复杂度。一旦新的数据到来, 在更新当前帧的预积分之后, 可以粗略恢复出当前帧的位姿:

$$\begin{cases} \hat{\mathbf{R}}_j = \mathbf{R}_i \cdot \Delta \mathbf{R}_{ij} \\ \hat{\mathbf{t}}_j = \mathbf{p}_i + \mathbf{v}_i \cdot \Delta t_{ij} + \frac{1}{2} \mathbf{g} \cdot \Delta t_{ij}^2 + \mathbf{R}_i \cdot \Delta \mathbf{p}_{ij} \end{cases} \quad (2)$$

如图6所示, 上一帧第 k 个像素坐标 $u_i^{(k)}$ 对应世界坐标系下的三维点为 $p_i^{(k)}$, 将 $p_i^{(k)}$ 投影到当前帧的预测位置为:

$$\hat{u}_j^{(k)} = \pi \left(\hat{\mathbf{R}}_j^T \left(p_i^{(k)} - \hat{\mathbf{t}}_j \right) \right) \quad (3)$$

其中, $\pi(\cdot)$ 表示相机的投影函数, 定义为:

$$\pi(\mathbf{p}) = \begin{bmatrix} f_x \frac{X}{Z} + c_x \\ f_y \frac{Y}{Z} + c_y \end{bmatrix} \quad (4)$$

将预测得到的像素坐标 $\hat{u}_j^{(k)}$ 和相机观测匹配到的像素位置 $u_j^{(k)}$ 作差, 得到第 k 个点的一致性误差:

$$e^{(k)} = \left\| u_j^{(k)} - \hat{u}_j^{(k)} \right\|^2 \quad (5)$$

2.3 联合动态特征点剔除策略

尽管图像分割模型能够有效检测图像中的潜在动态区域, 但是并不能准确区分出掩码内真正的动态特征点。而基于 IMU 预积分的运动一致性判断能够根据传感器惯性测量信息检测出与整体相机运动不一致的特征点。为充分发挥视觉以及 IMU 在动态识别中的互补优势, 本节提出了融合语义分割与运动一致性判断的联合动态特征点剔除机制, 其原理架构如图7所示。

在特征筛选方面, 将每帧图像的掩膜输出与第1.1节 SuperPoint 提取的特征点进行空间重合判断, 以筛选潜在的动态特征点。

将当前帧中提取到的特征点集合记为:

$$\mathcal{P}_k = \{p_1^{(k)}, p_2^{(k)}, \dots, p_{N_k}^{(k)}\} \quad (6)$$

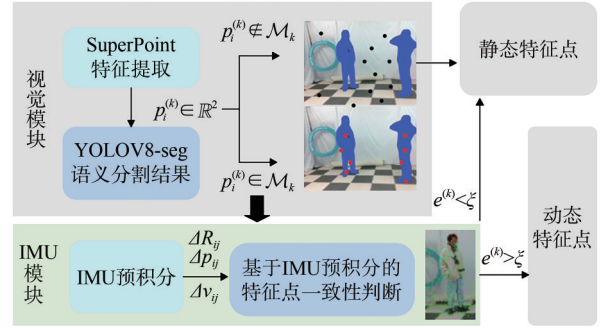


图7 联合动态特征点剔除策略示意图

其中, $p_i^{(k)} \in \mathbb{R}^2$ 表示第 i 帧图像中第 k 个特征点, N_k 为特征点的数量。

YOLOv8-seg 输出的二值掩码记作 $\mathcal{M}_k \subset \mathbb{R}^2$, 表示潜在动态区域对应的像素集合, 定义判别函数:

$$\delta_i^{(k)} = \begin{cases} 1, & p_i^{(k)} \in \mathcal{M}_k \\ 0, & p_i^{(k)} \notin \mathcal{M}_k \end{cases} \quad (7)$$

其中, $\delta_i^{(k)} = 1$ 表示该特征点位于掩码区域内, 被视为潜在动态特征点, 需要进入后续 IMU 一致性模块进一步判断; 而 $\delta_i^{(k)} = 0$ 的特征点被视为静态特征点, 直接保留。虽然这部分特征点被保留, 但是为了保证地图的全局一致性, 这部分特征点只参与位姿跟踪与优化而不参与地图点生成与地图构建。此外, 第 i 帧需要根据当前帧的跟踪方式是运动模型跟踪还是关键帧跟踪来决定。

得益于高质量的前端匹配, 后端能够有效清除潜在动态干扰, 避免错误匹配污染优化过程。二者协同作用, 使系统在低纹理动态场景下能够同时保持较高的精度与鲁棒性。

3 实验结果及分析

为了全面评估本文方法在不同场景下的定位精度、鲁棒性和重复试验的稳定性, 本文采用绝对轨迹误差 (ATE) 作为系统评价指标, 通过计算其均方根误差 (RMSE) 来评价系统精度和鲁棒性, 通过计算标准差 (S.D.) 来衡量系统运行的稳定性。本文在两个公开数据集 (EuRoC^[34] 和 TUM-VI^[35]) 上进行了测试。同时, 利用 Nokov 动态捕捉系统构建了真实低纹理动态场景进行实机测试。所有实验均在同一硬件平台上进行测试, 并对每组实验数据重复测试 5 次, 求取平均值以保证性能评估的稳定性。

3.1 公开数据集实验

EuRoC是一个广泛使用的视觉惯性SLAM数据集，拍摄场景主要为室内工业厂房与结构化办公空间。该数据集通过搭载双目相机与IMU的微型无人机采集，具有以下特点：轨迹速度变化剧烈、显著的光照变化与运动模糊，以及存在部分低纹理区域。

表1给出了在EUROC数据集上测试得到的绝对轨迹误差（ATE）的均方根误差（RMSE）对比结果。可以看出，本文方法在多个典型序列中取得了优异的性能。在MH01、MH02和MH03等低速飞行序列中，本文方法与ORB-SLAM3^[36]性能相当；在MH04和MH05等高难度序列中，由于存在剧烈相机运动与强光照变化，传统基于稀疏特征的系统（如OKVIS^[37]、ROVIO^[38]）普遍误差较高（见表1，MH05中OKVIS方法为0.470 m，ROVIO方法为0.520 m），而本文方法绝对轨迹误差的均方根误差为0.068 m，显著低于BASALT^[39]和VINS-Fusion^[40]方法，表现出良好的精度与鲁棒性。在V101~V203中，尽管这些序列整体纹理条件较好，但部分场景存在光照不均、快速穿越等问题，本文方法依然展现出优良的精度与稳定性，在V201中绝对轨迹误差的均方根误差为0.021 m，优于所有对比方法。

表1 EUROC数据集绝对轨迹误差（ATE RMSE/m）
实验结果对比

序列	ORB-SLAM3	VINS-Fusion	OKVIS	ROVIO	BASALT	OURS
MH01	0.016	0.540	0.160	0.210	0.080	0.018
MH02	0.027	0.460	0.220	0.250	0.060	0.027
MH03	0.028	0.330	0.240	0.250	0.050	0.026
MH04	0.138	0.780	0.340	0.490	0.100	0.127
MH05	0.072	0.500	0.470	0.520	0.080	0.068
V101	0.033	0.550	0.090	0.100	0.040	0.035
V102	0.015	0.230	0.200	0.100	0.020	0.016
V103	0.033	—	0.240	0.140	0.030	0.032
V201	0.023	0.230	0.130	0.120	0.030	0.021
V202	0.029	0.200	0.160	0.140	0.020	0.026
V203	—	—	0.290	0.140	—	—

TUM-VI是一个高质量的视觉惯性SLAM数据集，专为应对低纹理、光照变化和真实惯性干扰环境下的系统鲁棒性挑战而设计。该数据集使用双目相机与IMU采集，并提供精确的时间同步和地面真值。其中的Corridor序列具有典型的长走廊、重

复结构和低纹理墙面等特征，是测试系统在低纹理场景下特征提取能力和跟踪稳定性的重要基准。

表2给出了在TUM-VI数据集上测试得到的性能对比结果。可以看出，本文方法在低纹理区域仍能稳定提取高质量特征点并完成精确匹配，在Corridor1与Corridor2中绝对轨迹误差的均方根误差分别达到0.05 m和0.03 m，显著优于ORB-SLAM3、BASALT等系统，在低纹理场景下表现出较强的精度与鲁棒性。由于存在少量的动态物体，传统基于稀疏点的系统（如OKVIS、ROVIO、VINS-Fusion）在Corridor3和Corridor5等序列中精度显著降低，其中ROVIO在Corridor5序列中绝对轨迹误差的均方根误差高达2.09 m，VINS-Fusion在Corridor3序列中绝对轨迹误差的均方根误差达到了1.56 m，表明这些系统在动态场景下易出现漂移。

表2 TUM-VI数据集绝对轨迹误差（ATE RMSE/m）
实验结果对比

序列	ORB-SLAM3	VINS-Fusion	OKVIS	ROVIO	BASALT	OURS
Corridor1	0.08	0.63	0.33	0.47	0.34	0.05
Corridor2	0.06	0.95	0.47	0.75	0.42	0.03
Corridor3	0.05	1.56	0.57	0.85	0.35	0.40
Corridor4	0.03	0.25	0.26	0.13	0.21	0.23
Corridor5	0.06	0.77	0.39	2.09	0.37	0.05

3.2 实机实验

为了进一步验证本文提出的系统在真实低纹理动态环境中的实用性与鲁棒性，在实验室中搭建了测试平台，使用白色背景墙纸和行人等方式构造低纹理动态场景。实验相机采用RealSense D435i，轨迹真值由高精度Nokov动作捕捉系统提供，以确保评估的准确性。共采集4个序列，序列场景如图8所示。

利用传统的特征提取方法（如ORB）对图像进行特征提取，得到的特征点数量普遍小于30，将特征点提取平均数量为20~30个的场景归属于低纹理，特征点提取平均数量小于20个的场景归于极端低纹理场景。图8中，S1与S2为静态低纹理场景，S1含有少量结构纹理，S2为极端低纹理场景；D3与D4则在低纹理基础上加入行人动态干扰。

表3给出在自制数据集上测试得到的本文所提系统与ORB-SLAM3系统的对比结果。可以看出，对于静态序列S1和S2，本文所提系统精度分别提

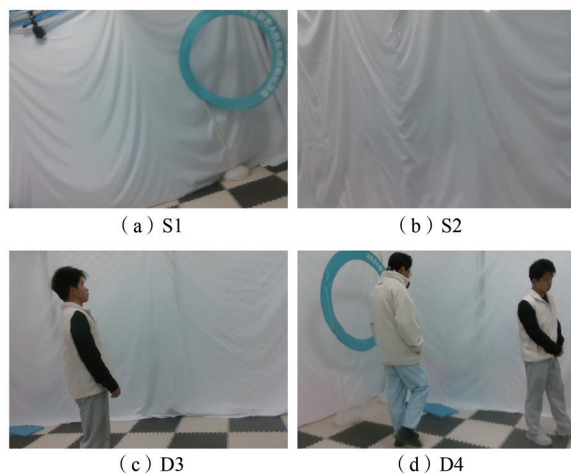


图8 自制数据集

升40.0%和65.4%，验证了本文系统在低纹理场景下具备稳定和优秀的特征提取与匹配性能。在动态场景D3与D4中，ORB-SLAM3受动态干扰影响严重，均方根误差分别达到了1.102 9 m和1.372 3 m，标准差分别达到了1.101 3 m与1.163 9 m。而本文系统的均方根误差减少88.4%及以上，标准差减少90%及以上，充分验证了本文系统在真实动态场景下的高精度与强鲁棒性。

图9为4个自制数据集绝对轨迹误差的可视化结果。从图9可以看出，对于S1数据集，虽然原系统能跟踪得上，但是由于特征点质量较差，发生了大范围的轨迹漂移。对于特征点更严苛的S2数据集，由于提取到的特征点难以维持系统的正常运行，原系统发生了跟丢行为，而本文提出的系统在

表3 自制数据集绝对轨迹误差实验结果对比

Seq	ORB-SLAM3		OURS		Improvement(Y)	
	RMSE/m	S.D./m	RMSE/m	S.D./m	RMSE	S.D.
S1	0.102 7	0.079 8	0.061 7	0.031 1	40.0%	61.0%
S2	0.134 0	0.090 2	0.046 3	0.041 0	65.4%	54.5%
D3	1.102 9	1.101 3	0.127 5	0.110 1	88.4%	90.0%
D4	1.372 3	1.163 9	0.120 3	0.073 5	91.2%	93.7%

保持鲁棒跟踪的同时仍有较好的精度。对于D3和D4数据集，原系统虽然都能跟踪上，但是由于包含大范围的动态物体，轨迹发生了非常严重的漂移。

3.3 消融实验

(1) 深度学习特征提取与匹配模块消融实验

为了验证 SuperPoint 特征提取和 LightGlue 匹配模块在低纹理环境中的改进性能，分别在数据集 EuRoC 的 MH01、MH02，TUM-VI 的 Corridor1、Corridor2 以及 S1 和 S2 中进行测试，对比不同特征点提取与匹配单帧平均点数量及系统精度，结果见表4。

可以看出，在纹理较丰富的 EuRoC MH01 与 MH02 序列中，ORB 与本文提出的 SuperPoint 提取的特征点数量相近，RMSE 也基本持平，表明在纹理丰富的场景下两者性能相当，均可提供稳定支撑。然而，在 Corridor1 和 Corridor2 等低纹理走廊序列中，本文方法提取到的有效特征点数量分别提升了7.0%和15.7%，系统精度分别提升了37.5%和33.3%，表明在低纹理区域，SuperPoint 模块仍能稳定提取高质量点对，且 LightGlue 匹配模块能有效抑制误

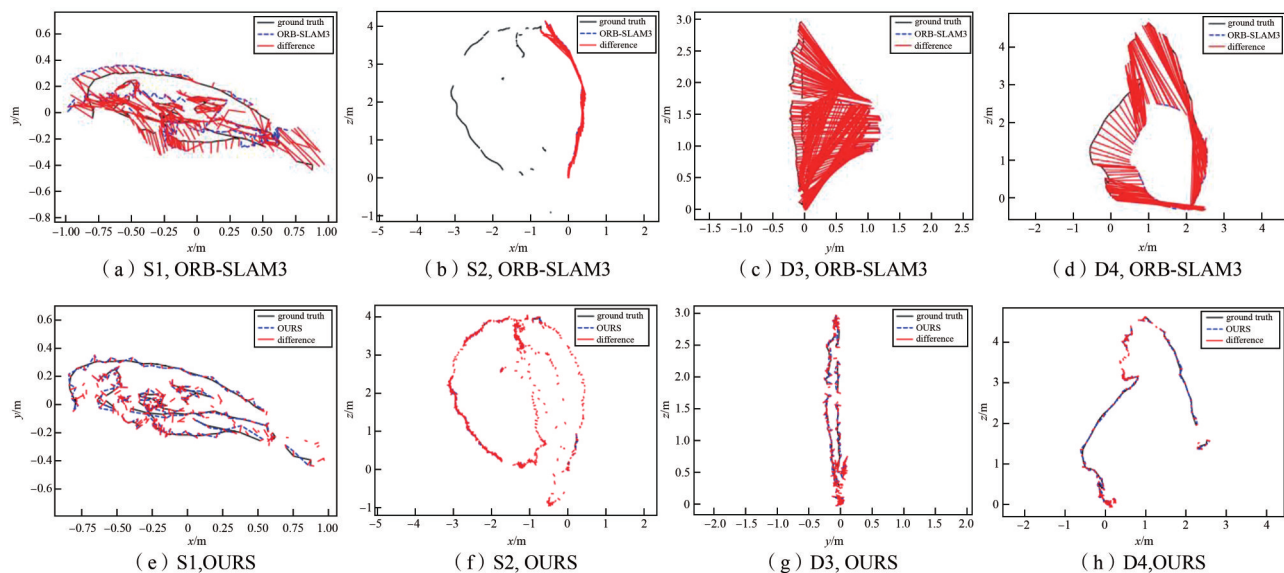


图9 自制数据集轨迹对比结果

表4 特征提取与匹配消融实验对比

序列	ORB-SLAM3: ORB+BF		SP+LG		特征点数量提升率	系统精度提升
	RMSE/m	特征点数量/个	RMSE/m	特征点数量/个		
MH01	0.016 0	271	0.018 0	267	—	—
MH02	0.027 0	237	0.027 0	241	1.7%	—
Corridor1	0.080 0	128	0.050 0	137	7.0%	37.5%
Corridor2	0.060 0	102	0.040 0	118	15.7%	33.3%
S1	0.102 7	27	0.060 7	116	330%	41.0%
S2	0.134 0	18	0.044 3	103	472%	66.9%

匹配与漂移。在极端低纹理序列S1和S2中，传统ORB特征提取效果显著下降，见表4，仅能提取27个和18个特征点，导致系统精度严重下降；而本文方法提取的特征点数分别为116和103个，提升幅度高达330%和472%，系统精度分别提升41.0%和66.9%。

(2) 联合动态特征点剔除策略消融实验

为了验证本文提出的联合动态点剔除策略的有效性，在真实动态场景下分别测试无剔除、仅掩码剔除和联合剔除时的系统性能差异。表5给出D3和D4序列中联合动态特征点剔除策略消融实验结果。可以看出，无动态剔除策略时，系统受动态特征干扰严重，系统均方根误差分别高达0.989 2 m和1.123 7 m，提取的特征点数量仅为27个和18个，出现漂移甚至失败；仅采用掩码剔除策略提取的特征点数量显著提升至93个和87个，系统的均方根误差降低至0.182 3 m和0.191 7 m；加入联合动态剔除策略之后，本文所提系统提取的特征点数量进一步提升至116个和103个，系统定位精度进一步提升，均方根误差降至0.127 5 m和0.120 3 m。图10给出了在D3和D4数据集上基于联合动态特征点剔除策略的特征点提取结果。该结果说明联合剔除策略能够缓解因盲目剔除导致的特征稀疏问题，从而

表5 联合动态特征点剔除消融实验对比

序列	无动态剔除策略		仅掩码剔除		联合动态剔除	
	RMSE/m	特征点数量	RMSE/m	特征点数量	RMSE/m	特征点数量
D3	0.989 2	27	0.182 3	93	0.127 5	116
D4	1.123 7	18	0.191 7	87	0.120 3	103

显著提升系统精度与稳定性。

3.4 系统运行效率对比实验

为了验证本文所提方法的运行效率，在相同的硬件平台（AMD R9-5900HX+RTX3080）上对本文不同模块的运行效率进行了分析。对自制数据集的两个动态场景D3和D4分别运行求取单帧平均运行时间，结果见表6。原ORB系统前端耗时约为14.2 ms，采用SuperPoint作为特征点提取方式后，在GPU的帮助下特征提取和匹配的时间缩短至11.6 ms。采用SuperPoint和LightGlue作为特征提取和匹配之后，前端运行时间明显上升至22.4 ms。对于联合动态特征点提取模块，单帧耗时稳定在7~8 ms。总的来说，改进后系统单帧处理耗时有所上升，但依旧能满足SLAM系统对实时性的要求；更重要的是，相比原始ORB-SLAM3，本文提出的系统在低纹理动态场景下具有更强的定位精度和鲁棒性。

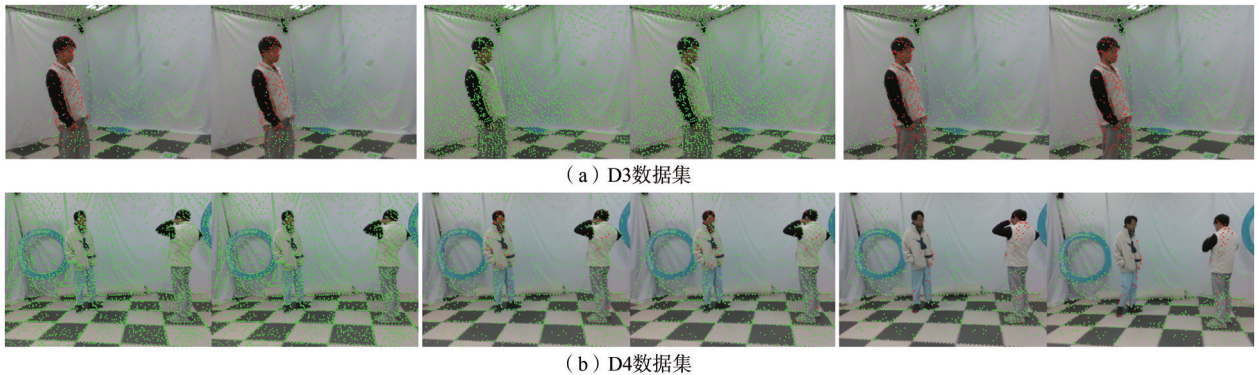


图10 D3和D4数据集上联合动态特征点剔除策略结果

表6 系统运行效率对比

选用方法	特征提取与匹配/ms	语义分割/ms	特征点一致性检测/ms	改进后系统单帧耗时/ms	原系统单帧耗时/ms
ORB+BF	14.2	6.6	0.8	28.5	21.6
SP+BF	11.6	7.1	0.8	25.8	22.7
SP+LG	22.4	7.4	0.8	36.9	21.2

4 结束语

本文提出了一种基于深度学习的特征提取匹配与语义-IMU联合动态点剔除策略的单目视觉惯导SLAM系统,提升了低纹理动态场景下的系统定位精度与鲁棒性,通过引入SuperPoint与LightGlue,实现了在低纹理环境中的稳定特征提取与匹配。此外,本文提出了融合语义分割与IMU预积分的动态点细粒度筛除机制,有效解决了特征误匹配与稀疏问题。最后,在EuRoC、TUM-VI公开数据集及实验室场景中验证了本文方法的性能,相较于传统ORB-SLAM3和语义剔除策略在定位精度与鲁棒性上得到显著提升。未来工作将聚焦于进一步提升前端轻量化与动态建模能力,满足更复杂的动态真实环境需求。

参考文献:

- [1] 王霞, 左一凡. 视觉SLAM研究进展[J]. 智能系统学报, 2020, 15(5): 825-834.
WANG X, ZUO Y F. Advances in visual SLAM research[J]. CAAI Transactions on Intelligent Systems, 2020, 15(5): 825-834.
- [2] 蔡显奇, 王晓松, 李玮. 一种室内弱纹理环境下的视觉SLAM算法[J]. 机器人, 2024, 46(3): 284-293, 304.
CAI X Q, WANG X S, LI W. A visual SLAM algorithm in indoor weak texture environment[J]. Robot, 2024, 46(3): 284-293, 304.
- [3] HUANG G P, MOURIKIS A I, ROUMELIOTIS S I. Observability-based rules for designing consistent EKF SLAM estimators[J]. International Journal of Robotics Research, 2010, 29(5): 502-528.
- [4] JIANG X Y, LI H, CHEN C Q, et al. DDIO-mapping: a fast and robust visual-inertial odometry for low-texture environment challenge[J]. IEEE Transactions on Industrial Informatics, 2024, 20(3): 4418-4428.
- [5] WEN S H, LI X F, LIU X, et al. Dynamic SLAM: a visual SLAM in outdoor dynamic scenes[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 5028911.
- [6] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: real-time monocular visual SLAM with points and lines[C]//Proceedings of the 2017 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 4503-4508.
- [7] HE Y J, ZHAO J, GUO Y, et al. PL-VIO: tightly-coupled monocular visual-inertial odometry using point and line features[J]. Sensors, 2018, 18(4): 1159.
- [8] FU Q, WANG J, YU H, et al. PL-VINS: real-time monocular visual-inertial SLAM with point and line features[EB]. arXiv preprint, 2020. arXiv:2009.07462.
- [9] DETONE D, MALISIEWICZ T, RABINOVICH A. SuperPoint: self-supervised interest point detection and description[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2018: 337-33712.
- [10] TYSZKIEWICZ M, FUA P, TRULLS E. DISK: learning local features with policy gradient[J]. Advances in Neural Information Processing Systems, 2020, 33: 14254-14265.
- [11] SUN J M, SHEN Z H, WANG Y A, et al. LoFTR: detector-free local feature matching with transformers[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 8918-8927.
- [12] WANG Y F, HE X Y, PENG S D, et al. Efficient LoFTR: semi-dense local feature matching with sparse-like speed[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 21666-21675.
- [13] POTJE G, CADAR F, ARAUJO A, et al. XFeat: accelerated features for lightweight image matching[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 2682-2691.
- [14] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: learning feature matching with graph neural networks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 4937-4946.
- [15] LINDENBERGER P, SARLIN P E, POLLEFEYS M. LightGlue: local feature matching at light speed[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2023: 17581-17592.
- [16] JIANG H W, KARPUR A, CAO B Y, et al. OmniGlue: generalizable feature matching with foundation model guidance[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2024: 19865-19875.
- [17] 王柯赛, 姚锡凡, 黄宇, 等. 动态环境下的视觉SLAM研究评述[J]. 机器人, 2021, 43(6): 715-732.
WANG K S, YAO X F, HUANG Y, et al. Review of visual SLAM in dynamic environment[J]. Robot, 2021, 43(6): 715-732.
- [18] YU C, LIU Z X, LIU X J, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//Proceedings of the 2018 IEEE/RISJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 1168-1174.
- [19] KUNDU A, KRISHNA K M, SIVASWAMY J. Moving object detection by multi-view geometric techniques from a single camera mounted robot[C]//Proceedings of the 2009 IEEE/RISJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2009: 4306-4312.
- [20] ZOU D P, TAN P. CoSLAM: collaborative visual SLAM in dynamic environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(2): 354-366.
- [21] NARAYANA M, HANSON A, LEARNED-MILLER E. Coherent motion segmentation in moving camera videos using optical flow orientations[C]//Proceedings of the 2013 IEEE International Conference on

- Computer Vision. Piscataway: IEEE Press, 2013: 1577-1584.
- [22] 李嘉铭, 解明扬, 张民, 等. 动态环境下基于语义信息与几何约束的视觉SLAM系统[J]. 智能科学与技术学报, 2023, 5(4): 477-485.
- LI J M, XIE M Y, ZHANG M, et al. Visual SLAM based on semantic information and geometric constraints in dynamic environment[J]. Chinese Journal of Intelligent Science and Technology, 2023, 5(4): 477-485.
- [23] XIAO L H, WANG J G, QIU X S, et al. Dynamic-SLAM: semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. Robotics and Autonomous Systems, 2019, 117: 1-16.
- [24] 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.
- [25] KIM D H, HAN S B, KIM J H. Visual odometry algorithm using an RGB-D sensor and IMU in a highly dynamic environment[C]//Robot Intelligence Technology and Applications 3. Cham: Springer, 2015: 11-26.
- [26] XU H, YANG C G, LI Z J. OD-SLAM: real-time localization and mapping in dynamic environment through multi-sensor fusion[C]//Proceedings of the 2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM). Piscataway: IEEE Press, 2020: 172-177.
- [27] CHAVEZ-GARCIA R O, AYCARD O. Multiple sensor fusion and classification for moving object detection and tracking[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(2): 525-534.
- [28] VARGHESE R, M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]//Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems. Piscataway: IEEE Press, 2024: 1-6.
- [29] HUANG X T, CHEN X B, ZHANG N, et al. ADM-SLAM: accurate and fast dynamic visual SLAM with adaptive feature point extraction, Deeplabv3pro, and multi-view geometry[J]. Sensors, 2024, 24(11): 3578.
- [30] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[EB]. arXiv preprint, 2017, arXiv:1706.05587.
- [31] LIU J H, LI X F, LIU Y Q, et al. RGB-D inertial odometry for a resource-restricted robot in dynamic environments[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 9573-9580.
- [32] REDMON J, FARHADI A. Yolov3: an incremental improvement[EB]. arXiv preprint, 2018, arXiv:1804.02767.
- [33] CONG P C, LI J X, LIU J J, et al. SEG-SLAM: dynamic indoor RGB-D visual SLAM integrating geometric and YOLOv5-based semantic information[J]. Sensors, 2024, 24(7): 2102.
- [34] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets[J]. International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [35] SCHUBERT D, GOLL T, DEMMEL N, et al. The TUM VI benchmark for evaluating visual-inertial odometry[C]//Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 1680-1687.
- [36] CAMPOS C, ELVIRA R, RODRÍGUEZ 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.
- [37] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334.
- [38] HUI Z, HUANG G Q. Robocentric visual-inertial odometry[C]//Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 6319-6326.
- [39] USENKO V, DEMMEL N, SCHUBERT D, et al. Visual-inertial mapping with non-linear factor recovery[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 422-429.
- [40] QIN T, CAO S, PAN J, et al. A general optimization-based framework for global pose estimation with multiple sensors[EB]. arXiv preprint, 2019, arXiv: 1901.03642.

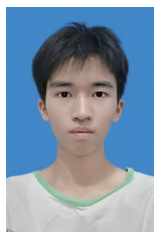
[作者简介]



解明扬 (1988-), 男, 博士, 南京航空航天大学自动化学院副研究员, 主要研究方向为智能机器人技术。



徐鑫 (2001-), 男, 南京航空航天大学自动化学院硕士生, 主要研究方向为移动机器人自主导航。



杨晨 (2004-), 男, 南京航空航天大学自动化学院研究助理, 主要研究方向为机器人路径规划技术。



余子锐 (2005-), 男, 南京航空航天大学自动化学院研究助理, 主要研究方向为机器人路径规划技术。



王垚 (2005-), 女, 南京航空航天大学自动化学院研究助理, 主要研究方向为机器人路径规划技术。