

基于实例分割和多约束优化的动态 SLAM 方法

李金宝, 李凯, 边宇峰, 吴璋琪, 贾文辉, 聂鹏飞

(中北大学 智能探测技术与装备山西省重点实验室, 太原 030051)

摘要: 针对传统视觉 SLAM 方法在动态环境中易受运动目标和光照变化干扰导致定位精度和系统鲁棒性下降的问题, 提出一种基于实例分割与多约束优化的动态视觉 SLAM 方法, 通过改进的轻量化 YOLOv11 模型实现高效实例分割, 减少动态特征对特征提取与数据关联的影响; 并在特征匹配中引入局部亮度归一化和光照不变性约束, 提升特征匹配精度; 同时结合信息论与协方差约束设计一种关键帧选择与边缘化策略, 通过量化候选关键帧对系统可观性的贡献, 结合不确定性度量实现关键帧优化, 在多个动态数据集和实际动态场景上的实验结果表明, 该方法在高动态场景下较 ORB-SLAM3 的绝对轨迹误差的均方根误差和平均误差分别降低了 96.19% 和 95.23%, 显著提升了复杂动态场景下的定位精度与鲁棒性。

关键词: 视觉 SLAM; 动态场景; 实例分割; 光度约束; 信息增益; 关键帧优化

Dynamic SLAM Method Based on Instance Segmentation and Multi-Constraint Optimization

LI Jinbao, LI Kai, BIAN Yufeng, WU Xuanqi, JIA Wenhui, NIE Pengfei

(Shanxi Key Laboratory of Intelligent Detection Technology & Equipment, North University of China, Taiyuan 030051, China)

Abstract: Aiming at the problem that traditional visual simultaneous localization and mapping (SLAM) methods are susceptible to interference with moving objects and illumination changes in dynamic environments, leading to a degradation in positioning accuracy and system robustness, a dynamic visual SLAM method based on instance segmentation and multi-constraint optimization is proposed. An improved lightweight YOLOv11 model is adopted to achieve efficient instance segmentation, which reduces the impact of its dynamic features on feature extraction and data association. In addition, local brightness normalization and illumination invariance constraints are introduced into feature matching to improve the accuracy of feature matching. Meanwhile, a keyframe selection and marginalization strategy is designed by combining information theory with covariance constraints. By quantifying the contribution of candidate key frames to system observability and combining with uncertainty measurement, the keyframe optimization is realized. Experimental results on multiple dynamic datasets and real dynamic scenes show that, compared with the ORB-SLAM3 in high-dynamic scenes, the proposed method reduces the root mean square error and average error of the absolute trajectory error by 96.19% and 95.23%, respectively, which significantly improves the positioning accuracy and robustness in complex dynamic environments.

Keywords: visual SLAM; dynamic scenes; instance segmentation; photometric constraint; information gain; key-frame optimization

收稿日期: 2025-12-16; 修回日期: 2026-01-29。

基金项目: 中央引导地方科技发展资金资助项目(YDZJSX20231A034, YDZJSX20244D032, YDZJSX2025D025); 山西重点研发计划资助项目(202202010101007); 山西省科技成果转化引导专项项目(202204021301044, 202304021301028); 山西省 2024 年度研究生实践创新项目(2024SJ271)。

作者简介: 李金宝(2000-), 男, 硕士研究生。

通讯作者: 李凯(1982-), 男, 博士, 教授, 博士生导师。

引用格式: 李金宝, 李凯, 边宇峰, 等. 基于实例分割和多约束优化的动态 SLAM 方法[J]. 计算机测量与控制, 2026, 34(4): 249-257.

0 引言

近年来,随着人工智能与传感器技术的发展,智能机器人在各领域被广泛应用,其中自主导航依赖于同步定位与建图(SLAM, simultaneous localization and mapping)技术^[1]。SLAM通过环境感知与建模为路径规划提供支撑。主流SLAM系统可分为激光SLAM^[2]与视觉SLAM^[3]两类,后者以信息丰富、成本低等优势在移动机器人与增强现实中备受关注。然而随着视觉SLAM技术的不断发展,ORB-SLAM系列如ORB-SLAM2^[4]与ORB-SLAM3^[5]提升了静态场景下视觉SLAM的精度与鲁棒性,但真实场景中不可避免地存在动态物体,动态目标的存在不仅会导致特征匹配失败和前端跟踪不稳定,还会对后端非线性优化过程造成严重影响^[6]。一方面,动态特征引入的错误约束会导致错误的位姿估计结果,导致轨迹漂移;另一方面,在动态场景下,SLAM系统往往需要引入更多关键帧以提高跟踪稳定性,若缺乏有效的关键帧管理机制,关键帧数量可能会迅速增长,影响系统实时性^[7-8]。

针对上述问题,国内外学者展开了大量的研究。目前应对动态场景的方法主要有基于几何约束的方法和基于深度学习的方法^[9]。基于几何约束的方法通过多视角几何运动一致性约束识别动态特征。文献[10]通过量化深度图实现运动分割和像素的分类以去除动态物体,进而提高了SLAM系统在动态场景下的定位精度。文献[11]提出了一种稀疏运动去除模型,利用相邻帧间的异同性过滤动态特征,进而提高了动态场景下的定位精度;文献[12]通过将地图点构建为图结构,利用点间相关性实现动静点分离。文献[13]提出了DynaVINS算法,引入基于IMU姿态先验的光束平差与关键帧分组策略,有效抑制动态物体的影响。

与几何方法相比,基于深度学习的语义感知方法能够通过场景中的语义信息明显地识别动态物体,从而提升SLAM系统的定位精度,效果更好,应对动态场景的能力更强。典型算法如DynaSLAM^[14]、YOLO-SLAM^[15],它们分别采用Mask R-CNN^[16]、YOLOv3等不同的网络模型提取语义信息进而剔除动态特征。文献[17]将改进的YOLOv8与ORB-SLAM3融合实现了动态场景下的定位与建图。文献[18]等通过贝叶斯滤波、Mask RCNN与深度聚类剔除动态点。上述方法利用深度学习增强SLAM在动态环境下的性能,但仍难兼顾精度与实时性。

为解决SLAM算法在动态场景中定位精度不足、鲁棒性差的问题,本文提出了一种基于ORB-SLAM3框架融合改进的YOLOv11和多约束优化的SLAM方法来应对动态环境。基于YOLOv11架构,替换轻量化卷

积,引入RepNCSPeLAN4模块结合CAA注意力机制提高分割性能;融合改进的特征匹配和基于信息论融合协方差约束的关键帧优化策略,提高位姿估计精度,进而实现动态环境下视觉SLAM系统的鲁棒性增强与定位精度提升。

1 系统总体设计与方法概述

本文所提出SLAM系统的整体框架如图1所示,该系统以ORB-SLAM3为基础架构,通过多个线程的并行协同机制实现动态场景下的高精度定位与建图。首先,系统通过改进的轻量级YOLOv11模型对输入的RGB-D图像进行实例分割,提取出的语义掩膜实时传递至跟踪线程,通过动态特征点剔除机制,在前端预处理阶段即消除了动态物体对数据关联的干扰。在随后的特征匹配过程中,系统通过对局部像素归一化并执行光度不变性约束,降低了复杂光照变化的影响,从而显著提升位姿估计的精度。估计得到的位姿与观测信息进入关键帧优化线程,系统采用基于信息论进行优化,根据候选帧对滑动窗口的信息增益贡献进行筛选,并结合基于协方差变化的边缘化策略,在维持关键帧质量的同时抑制冗余并避免模型退化。随后,局部建图与回环检测线程通过局部及全局光束平差法(BA, bundle adjustment)对相机轨迹和地图点进行联合优化,确保了全局尺度的一致性。最终,密集映射线程利用优化后的关键帧生成高密度点云,并结合语义信息剔除动态目标,构建出精确的静态稠密点云地图。

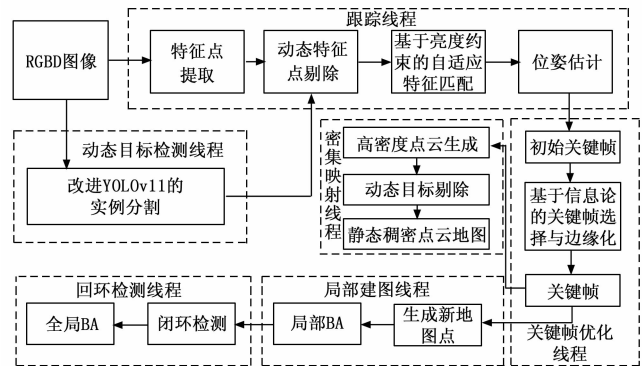


图1 系统整体框图

2 基于改进YOLOv11的动态目标语义分割

在动态环境中,运动目标往往是影响视觉SLAM系统稳定性的主要因素。由动态物体产生的特征点不满足静态假设,若直接参与位姿估计,将显著降低系统精度。因此,在SLAM系统中引入语义分割模块,对动态目标进行准确识别和剔除,是提高系统鲁棒性的关键环节^[19]。

为兼顾检测精度与系统实时性,本文以YOLOv11

模型^[20]为基础进行轻量化改进, 在 YOLOv11 框架中, 采用轻量化幻影卷积 GhostConv^[21]对传统卷积层 Conv 进行替换。对于传统卷积骨干网络, 输入图像需经历空间信息向通道信息的转化, 在此过程中, 空间维度的每一次压缩都会伴随语义信息的损耗。而 GhostConv 将卷积分成两步进行, 通过特征图冗余性分析, 采用线性变换生成“幻影”特征图来替代标准卷积操作。GhostConv 首先通过主卷积核生成 m 通道基础特征图, 随后对基础特征图施加轻量级的深度可分离卷积生成 s 通道补充特征图, 最终将二者通道拼接得到完整输出。具体而言, 基础卷积仅生成少量具有强表征能力的核心特征图, 大幅减少卷积核的数量与计算量, 另一方面, 廉价操作生成的 Ghost 特征, 能够捕捉核心特征中未被覆盖的细粒度细节, 补充分割任务所需的关键特征维度。该过程可形式化表示为:

$$\begin{cases} Y_{\text{base}} = \sigma(\varphi_m * X + b_m) \\ Y_{\text{ghost}} = DWConv(\varphi_s \otimes Y_{\text{base}}) \\ Y_{\text{out}} = Concat(Y_{\text{base}}, Y_{\text{ghost}}) \odot W_{\text{att}} \end{cases} \quad (1)$$

其中: $X \in \mathbb{R}^{H \times W \times C_s}$ 为输入特征图, $\varphi_m \in \mathbb{R}^{k \times k \times C_s \times m}$ 为主卷积核, $\varphi_s \in \mathbb{R}^{d \times d \times s \times m}$ 为幻影卷积核, m 和 s 均为通道数; $*$ 表示标准卷积, \otimes 表示深度可分离卷积, W_{att} 为通道注意力权重。

此外, 为提升特征提取能力与模型的梯度流动效率, 本文基于 YOLOv11 模型引入基于重参数化思想的 RepNCSPPELAN4 模块^[19]。该模块融合重参数化卷积、NCSP 结构与层次聚合结构的思想, 用于实现高效的多尺度特征提取与梯度流动增强。模块内部采用多分支并行卷积路径, 每个分支具有不同感受野深度, 能够在浅层捕获局部纹理信息, 在深层提取语义特征。设输入特征为 $X \in \mathbb{R}^{C \times H \times W}$, 则其输出可表示为:

$$Y = Concat[f_1(X), f_2(X), f_3(X), f_4(X)]$$

其中: $f_i(\cdot)$ 表示不同尺度的卷积分支。在训练阶段保留多分支以增强特征多样性, 在推理阶段则通过重参数化将其折叠为单一等效卷积, 具体如下式所示:

$$Y' = (W_{\text{eff}} * X) + b_{\text{eff}}, W_{\text{eff}} = \sum_{i=1}^n W_i \quad (2)$$

其中: W_{eff} 为等效卷积核权重, W_i 为第 i 个卷积分支的卷积核权重, 此过程可在保持特征表达能力的同时降低计算复杂度, 实现高表达、低延迟的平衡。

为进一步增强特征的空间聚焦能力, 本文在 RepNCSPPELAN4 模块中(见图 2)引入上下文锚定注意力(CAA, context anchor attention)机制^[22], 通过在水平与垂直方向分别执行全局平均池化嵌入位置信息, 并生成方向相关的权重矩阵, 进而实现空间与通道的联合加权, 强化与分割任务强相关的特征通道, 抑制无关背景特征的干扰, 提升特征利用率, 进而提升网络

的特征聚焦能力与语义辨识精度, 该过程可表示为:

$$Y'(c, h, w) = Y(c, h, w) \cdot a_x(c, h) \cdot a_y(c, w) \quad (3)$$

其中: c 为特征通道维度, h 为特征图高度维度, w 为特征图宽度维度, a_x 与 a_y 由 Sigmoid 函数归一化后获得, 能够根据坐标信息动态调整特征响应强度。

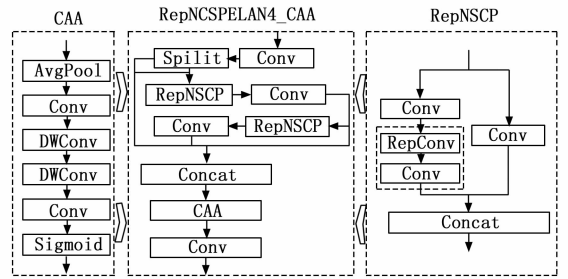


图 2 RepNCSPPELAN4_CAA 结构

综上所述, 改进后的实例分割模型具体结构如图 3 所示, 采用 GhostConv 替换传统卷积, 通过显式建模特征冗余关系, 引入 RepNCSPPELAN4 模块, 并结合 CAA 注意力机制对通道与空间维度进行联合加权, 强化动态目标区域特征, 进而缓解轻量化带来的特征压缩问题, 在降低参数数量和计算复杂度的同时保持了特征表达能力。

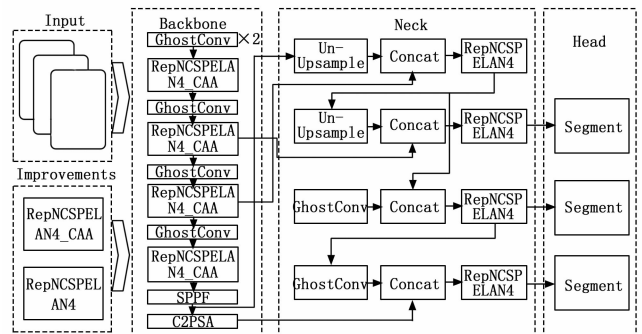


图 3 改进后的实例分割模型

3 多约束优化 SLAM 方法

3.1 特征匹配增强

在动态场景和复杂光照条件下, 传统 ORB 特征的分辨率会下降, 从而导致特征匹配的误差增加和重投影误差的二次方增长, 进而影响前端跟踪的稳定性。针对此问题, 本文提出了一种基于亮度约束的特征匹配增强方法。首先对特征点周围的局部像素块进行亮度归一化处理, 以降低光照的变化对特征描述的影响。并在特征匹配过程中引入光照不变性约束, 对候选匹配点进行进一步筛选, 进而有效抑制由光照变化引起的误匹配, 提高特征匹配结果的稳定性。

对于图像中的一个点 c , 取其 $n \times n$ 邻近窗口, 窗口中像素的平均值 μ 和标准偏差 σ 计算如下:

$$\mu = \frac{1}{n^2} \sum_{i=1}^{n^2} I(c_i) \quad (4)$$

$$\sigma = \sqrt{\frac{1}{n^2} \sum_{i=1}^{n^2} [I(c_i) - \mu]^2} \quad (5)$$

其中： $I(c_i)$ 为该点的本地光度不变性约束函数为。

$$L = \left\| \frac{I_1(c_i) - \mu_1}{\sigma_1 + \epsilon} - \frac{I_2(c'_i) - \mu_2}{\sigma_2 + \epsilon} \right\|_2 \quad (6)$$

其中： $\epsilon = 10^{-6}$ 。考虑到描述符距离 d_{ORB} 和光度成本 L ，最终的特征匹配判定函数为。

$$M = \begin{cases} 1 & \text{if } d_{ORB} < \tau_d \text{ and } L < \tau_l \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

其中： τ_d 为描述符距离阈值； τ_l 为描述符距离阈值。考虑到 ORB 描述符在旋转与尺度变化下具有较强的稳定性，而光照变化具有明显的非平稳特性，光度一致性约束值具体为：

$$\tau_l = \max(\tau_{\min}, \mu_L + \beta \sigma_L)$$

其中： μ_L 为光度代价集合的均值， σ_L 为光度代价集合的标准差， β 为经验系数， τ_{\min} 为最小保护阈值。

如图 4 和表 1 所示，图像块的归一化提高了对光照变化的鲁棒性。引入局部亮度不变性约束后，匹配正确率从 95.28% 提高到 96.76%，提高了 1.48%。匹配率仅略微下降了 0.7%，匹配时间仅增加了 0.725 ms。实验结果表明，引入的条件有效地过滤了模糊或不可靠的匹配对，同时保留了大部分正确的对应关系。



图 4 特征匹配比较

表 1 匹配结果对比

匹配方式	匹配点数	匹配对数	匹配准确率/%	平均匹配率/%	时间/ms
传统方法	1 000	364	95.28	37.39	1.025
本文方法	1 000	357	96.76	36.69	1.75

3.2 基于信息论的关键帧选择与边缘化策略

关键帧管理是保障 SLAM 系统稳定性与定位建图精度的核心环节。当关键帧数量过多时系统会引入大量冗余信息，影响系统实时性；若关键帧数量过少则会导致约束关系退化，降低位姿估计结果的鲁棒性。针对此问题，本文提出一种基于信息论的关键帧优化与融合协方差的边缘化策略，通过量化新帧引入的信息增益及对状态估计的不确定性影响，实现关键帧的优化管理，进

而提升 SLAM 系统的性能。

在稀疏光束平差法 (BA, bundle adjustment) 框架下，系统状态的不确定性可以通过 Fisher 信息矩阵进行表述。本文通过引入信息论思想，对候选关键帧对系统的可观性贡献进行量化，并将信息增益作为关键帧选择的重要判据。当候选关键帧引入的信息增益超出设定阈值时，将其提升为关键帧，进而增强系统约束结构。

在稀疏 BA 框架下，系统的状态向量 \mathbf{x} 由相机位姿 x_p 与地图路标点 x_l 构成，即 $\mathbf{x} = \begin{bmatrix} x_p \\ x_l \end{bmatrix}$ 。观测残差经线性化后，采用高斯-牛顿 (GN, Gauss-Newton) 法近似的最小二乘目标函数可表示为：

$$E(\delta x) = \frac{1}{2} \delta x^T (\mathbf{J}^T \mathbf{W} \mathbf{J}) \delta x \quad (8)$$

其中： \mathbf{J} 为雅可比矩阵， \mathbf{W} 为观测权重矩阵。其 Fisher 信息矩阵为 $\mathbf{I} = \mathbf{J}^T \mathbf{W} \mathbf{J}$ ，其逆 $\mathbf{P} = \mathbf{I}^{-1}$ 表示状态估计的协方差矩阵。当候选关键帧 k 产生新观测时，信息矩阵更新为：

$$\mathbf{I}' = \mathbf{I} + \Delta \mathbf{I}_k, \Delta \mathbf{I}_k = \mathbf{J}_k^T \mathbf{W}_k \mathbf{J}_k \quad (9)$$

定义候选帧 k 的信息增益为其对系统信息量的提升程度，表达式为：

$$\Delta \varphi_k = \log \det(\mathbf{I} + \Delta \mathbf{I}_k) - \log \det(\mathbf{I}) \quad (10)$$

其中： \mathbf{I} 为当前滑动窗口内的系统信息矩阵； $\Delta \mathbf{I}_k$ 为候选帧 k 带来的增量信息矩阵； $\Delta \mathbf{I}_k$ 为关键帧的信息增益得分。 $\Delta \varphi_k$ 为降低计算复杂度，仅保留局部窗口 W 内关键帧信息矩阵 \mathbf{I}_w ，并采用一阶迹近似：

$$\Delta \varphi_k \approx \text{trace}(\mathbf{I}_w^{-1} \Delta \mathbf{I}_k) \quad (11)$$

其中： $\Delta \varphi_k$ 为候选帧对位姿信息矩阵的增量； $\text{trace}(\cdot)$ 为矩阵迹运算，其结果表征候选帧带来的总信息量增幅。若该得分超过阈值 τ_{add} ，即 $\Delta \varphi_k > \tau_{\text{add}}$ 则该帧将被提升为关键帧。通过量化“信息提升量”实现对场景变化的自适应响应。

当滑动窗口内关键帧数量达到预设上限 K_{max} 时，对低贡献关键帧进行边缘化。当删除关键帧 i 时，新的信息矩阵为：

$$\mathbf{I}' = \mathbf{I} - \Delta \mathbf{I}_i$$

其对应协方差近似为：

$$\mathbf{P}' \approx \mathbf{I}^{-1} + \mathbf{I}^{-1} \Delta \mathbf{I}_i \mathbf{I}^{-1} \quad (12)$$

则关键帧的重要程度量由协方差变化量表示：

$$(\Delta P_i = \text{trace}(\mathbf{I}^{-1} \Delta \mathbf{I}_i \mathbf{I}^{-1})) \quad (13)$$

其中： \mathbf{I}^{-1} 为系统当前估计协方差； $\Delta \mathbf{I}_i$ 为关键帧 i 对信息矩阵的贡献； ΔP_i 为边缘化该帧后引起的不确定性增量。若 $\Delta P_i < \tau_{\text{marg}}$ ，则该帧可安全边缘化，否则保留，执行边缘化操作，从而在保持系统精度的同时控制滑动窗口规模。

4 实验与结果分析

4.1 实例分割改进结果分析

经过模型轻量化改进后, 改进后的 YOLOv11 模型的性能量化结果如表 2 所示。改进模型的计算复杂度由 10.4 GFLOPs 降低至 9.4 GFLOPs, 降低了 9.6%, 参数量下降了 10.3%, 而实例分割的精度仅损失 0.8%。此外, 模型计算效率提升, 改进后的模型的平均推理时间从 34.4 ms 缩短至 33.7 ms。由此可见, 在相同硬件环境下, 改进后的模型能更快完成目标检测与实例分割任务。

表 2 改进前后性能对比结果

模型	mAP/%	参数量/GFLOPs	计算量	推理时间/ms
改进前	32.0	2.9	10.4	34.4
改进后	31.2	2.6	9.4	33.7

4.2 关键帧优化实验结果分析

为验证所提出的基于信息论的关键帧选择与边缘化策略在动态场景中的有效性, 对关键帧数量变化与信息增益之间的关系进行了系统性实验分析。

图 5、图 6 展示了进行边缘化前后的信息增益与关键帧数量随时间变化的对照结果。其中, 信息增益曲线反映了候选帧对系统整体可观性的贡献程度, 而关键帧数量曲线则描述了滑动窗口内关键帧规模的动态变化情况。

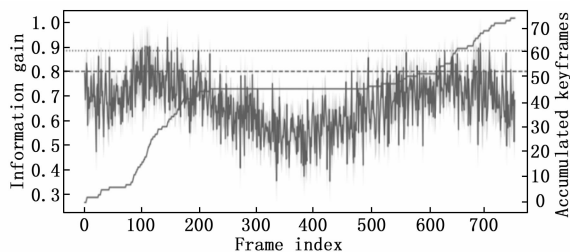


图 5 信息增益与关键帧数量对照结果

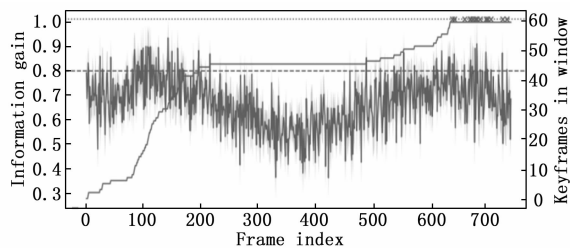


图 6 边缘化效果曲线

实验结果表明, 当相机运动幅度增大或视角变化明显时, 信息增益曲线呈现出明显上升趋势, 对应时刻系统更倾向于将当前帧提升为关键帧, 从而增强对系统状态的约束能力; 而在运动较为平缓或视角变化较小的阶段, 信息增益整体维持在较低水平, 系统能够有效抑制

关键帧的无序增长, 调节关键帧的添加时机, 避免冗余关键帧的产生。当关键帧数量达到窗口上限后, 基于协方差变化的边缘化策略能够及时移除低贡献的关键帧, 使关键帧数量保持稳定, 保证了系统约束结构的连续性。

关键帧优化前后的关键帧分布情况如图 7、图 8 所示, 由图中可知, 在动态程度较高或场景结构复杂的情况下, 关键帧之间的位姿不确定性与信息增益变化更为剧烈, 信息论度量会倾向于保留具有较高约束价值的关键帧, 从而形成相对更密集的关键帧分布, 避免了动态目标对系统约束结构的破坏。而在静态或结构简单场景中, 冗余关键帧的信息增益有限, 系统将自动提高边缘化概率, 降低关键帧密度, 抑制冗余关键帧的生成, 关键帧主要分布于视角变化显著且几何约束贡献较高的位置, 从而在保证定位精度的同时降低计算开销。

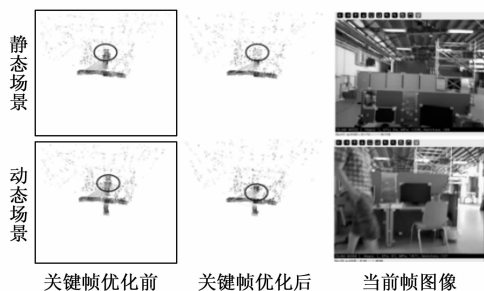


图 7 系统运行时关键帧优化效果

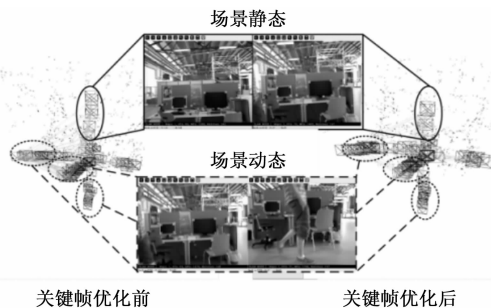


图 8 关键帧优化前后的关键帧分布

综合分析可得, 所提出的结合信息论的关键帧管理策略能够在动态场景中实现关键帧的优化处理, 在保证系统约束充分性的同时有效抑制关键帧冗余。

4.3 动态数据集实验结果分析

4.3.1 实验数据集说明

为了验证本文方法的定位性能, 采用 TUM_RGBD 数据集和 Bonn_RGBD 数据集进行定位实验, 由 Microsoft Kinect 传感器在多种室内场景下采集, 提供同步的 RGB 图像、深度图像及高精度真实轨迹, 广泛用于 SLAM 算法性能评估^[23]。TUM_RGBD 数据集的动态物体类型和运动模式相对有限, Bonn_RGBD 数据集^[24]包含更复杂的动态干扰与运动模式, 可与 TUM_RGBD 数据集形成互补, 从而实现对所提方法定位性

能的更全面评估。故将本文算法应用于以上动态序列以验证本文算法的性能。

4.3.2 实验环境与评价指标

本文算法的实验环境如表 3 所示。为准确评估本文提出算法的性能，采用绝对位姿误差 (ATE, absolute trajectory error) 和相对位姿误差 (RPE, relative pose error) 两个指标对 SLAM 算法性能进行量化并比较。

表 3 具体实验环境

硬件	具体配置
操作系统	Ubuntu18.04
CPU	Intel® Core™ I5-12600KF
GPU	NVIDIA GeForce RTX4060
训练环境	CUDA12.0
内存	64 G
输入	640 × 480

ATE 是基于时间戳的两个位姿 $P_{1:n}$ 和 $Q_{1:n}$ 之间的绝对位姿误差，用来评价算法的精度和轨迹的全局一致性。

$$ATE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n \|\text{trans}(Q_i^{-1}P_i)\|^2\right)} \quad (15)$$

其中： $P_{1:n}$ 表示估计的姿态， $Q_{1:n}$ 表示真实的姿态。

RPE 是基于时间戳的两个位姿 $P_{1:n}$ 和 $Q_{1:n}$ 之间的相对位姿误差，用来评价轨迹中相隔固定时间的局部准

确度。

$$RPE = \sqrt{\frac{1}{M} \sum_{i=1}^n \|\text{trans}[(Q_i^{-1}Q_i + \Delta_i)(P_i^{-1}P_i + \Delta_i)]\|^2} \quad (16)$$

4.3.3 实验结果与分析

本文采用 EVO 工具对本文算法的定位性能进行评估。在相对静态的场景下，本文算法的定位精度和估计轨迹的全局一致性与 ORB-SLAM3 相似。而在动态场景下，本文算法的定位精度、局部准确度和估计轨迹的全局一致性均明显优于 ORB-SLAM3，估计轨迹贴合于真实轨迹。

表 4 为本文算法与 ORB-SLAM3 的绝对轨迹误差实验对比结果，通过分析表 4 中的实验数据可知，在相对静态的环境下，本文方法的绝对轨迹误差的均方根误差值相较 ORB-SLAM3 降低 20.30%，平均绝对轨迹误差比 ORB-SLAM3 降低了 17.37%；在低动态的场景中，相较于 ORB-SLAM3，本文方法的绝对位姿误差的均方根误差值降低了 54.75%，平均绝对轨迹误差降低了 41.69%；在 fr3_w_rpy 等高动态环境中，本文方法的绝对位姿误差的均方根误差较 ORB-SLAM3 均降低了 95% 以上，平均绝对轨迹误差则降低了 95% 左右。

ORB-SLAM3 和本文提出算法在 TUM-RGBD 动态序列上的平移旋转误差对比结果如表 5 所示。在相对静态

表 4 本文算法与 ORB-SLAM3 的绝对轨迹误差实验结果对比

序列	ORB-SLAM3				本文算法			
	fr3_w_xyz	fr3_w_stt	fr3_w_rpy	fr3_s_stt	fr3_w_xyz	fr3_w_stt	fr3_w_rpy	fr3_s_stt
Max	1.177 7	0.155 6	1.977 1	0.040 8	0.112 2	0.031 9	0.136 6	0.029 4
Mean	0.485 1	0.016 0	0.680 7	0.008 5	0.022 2	0.009 3	0.025 9	0.007 1
Median	0.409 4	0.012 2	0.680 7	0.007 7	0.017 6	0.008 5	0.022 0	0.006 6
Min	0.111 1	0.001 0	0.709 5	0.000 5	0.001 0	0.000 7	0.001 2	0.000 6
RMSE	0.542 2	0.023 2	0.819 0	0.009 8	0.028 3	0.010 5	0.030 8	0.007 8
SSE	242.8037	0.385 1	579.486 5	0.064 9	0.660 1	0.078 8	0.551 8	0.041 3
Std	0.242 1	0.016 9	0.455 4	0.004 8	0.017 5	0.004 9	0.016 7	0.003 3

表 5 本文算法与 ORB-SLAM3 的相对轨迹误差实验对比结果

序列	ORB-SLAM3						本文算法					
	fr3_w_stt		fr3_w_xyz		fr3_w_rpy		fr3_w_stt		fr3_w_xyz		fr3_w_rpy	
	TP/m	RP/(°)	TP/m	RP/(°)	TP/m	RP/(°)	TP/m	RP/(°)	TP/m	RP/(°)	TP/m	RP/(°)
Max	0.173 1	0.070 5	0.311 2	0.142 6	0.159 3	0.090 1	0.037 5	0.018 8	0.082 1	0.048 9	0.065 8	0.126 0
Mean	0.009 8	0.005 5	0.019 4	0.011 2	0.021 9	0.013 2	0.007 8	0.004 8	0.010 1	0.006 5	0.010 2	0.007 0
Median	0.006 2	0.004 2	0.013 4	0.008 4	0.016 4	0.010 8	0.006 5	0.004 2	0.008 2	0.005 8	0.008 4	0.005 9
Min	0.000 6	0.000 4	0.000 4	0.000 4	0.001 0	0.000 5	0.001 0	0.000 3	0.000 0	0.000 0	0.001 1	0.000 4
RMSE	0.015 7	0.007 5	0.031 3	0.016 7	0.028 7	0.016 6	0.009 2	0.005 5	0.014 0	0.008 4	0.012 4	0.009 8
SSE	0.176 5	0.040 4	0.808 9	0.229 6	0.712 3	0.237 8	0.061 0	0.021 8	0.160 9	0.058 0	0.127 5	0.792 7
Std	0.012 3	0.005 2	0.024 6	0.012 4	0.018 6	0.010 0	0.005 0	0.002 8	0.009 7	0.005 3	0.007 1	0.006 8

场景中, 两者平移与旋转误差相近; 但在动态环境中, 本文算法优势凸显, 本文算法在低动态场景的平移、旋转平均误差降低 20.62%、13.26%、RMSE 分别降低 41.21%、26.51%; 在高动态序列 (如 fr3_w_xyz) 下, 算法平移、旋转误差的 RMSE 分别下降 55.40%、49.75%, 误差均值降幅分别达 48.07%、42.23%。总体而言, 本文算法在动态环境中相较于 ORB-SLAM3 均表现出更高精度与稳定性, 尤其在高动态场景下优势显著。

为进一步验证本文算法在不同室内动态场景和运动模式下的定位性能, 本文在选用了 Bonn_RGBD 数据集集中的 7 个序列进行实验, 以绝对轨迹误差的均方根误差为评价指标, 并与 ORB-SLAM3 和 YOLO-SLAM 的实验结果进行量化对比, 对比结果如表 6 所示。由表中数据可知, 本文算法的定位性能明显优于 ORB-SLAM3, 在多数动态场景下定位精度优于 YOLO-SLAM, 验证了本文算法在动态场景下的定位优越性。

表 6 Bonn 数据集上绝对轨迹误差对比

序列	ORB-SLAM3	YOLO-SLAM	本文算法
crowd1	0.250 6	0.033	0.020 5
crowd2	0.565 1	0.423	0.180 7
crowd3	0.292 5	0.069	0.024 5
person_tracking	0.663 3	0.157	0.051 2
person_tracking2	0.752 3	0.037	0.056 2
synchronous1	0.800 8	0.014	0.136 5
synchronous2	1.303 0	0.007	0.007 5

ORB-SLAM3 在动态目标干扰下轨迹误差较大, 误差沿轨迹呈现明显累积, 局部区域出现较高的误差段, 轨迹明显偏离于参考轨迹, 而本文算法的误差分布普遍

更为集中, 整体以低误差为主, 轨迹更加平滑稳定。由此可见, 本文方法能够有效抑制不同运动模式的多种室内动态场景, 体现了本文算法的优越性。

然而, 目前使用深度学习去除动态点进而应对动态环境的 SLAM 方法很多。本文以 DynaSLAM、DS-SLAM、Detect-SLAM、RDS-SLAM 等经典主流算法, 以及 ORB-SLAM2 和 ORB-SLAM3 为对比基准, 开展对比实验。对比实验结果如表 7、8 所示, 实验结果证明, 在多组动态场景序列中, 本文算法在 ATE 与 RPE 指标上均取得了更优或更具稳定性的表现, 验证了所提方法在室内动态环境中定位精度方面的优越性。

为评估本文在动态环境下各优化模块的有效性, 本文在 TUM_RGBD 数据集的 4 组动态序列上进行了消融实验, 对比了仅引入轻量级实例分割的系统 (系统 A)、在系统 (A) 基础上进一步加入光度约束的系统 (系统 B) 以及包含关键帧优化策略的完整系统。如表 9 所示, SLAM 系统 (A) 通过过滤已知动态物体并减少动态特征干扰, 降低动态特征对匹配过程的干扰, 实现定位精度的初步提升; 在引入光度约束后, 系统 B 的特征匹配精度进一步提升, 位姿估计结果更为精确, 定位精度进一步提高; 而完整系统在高动态序列中表现出更好的误差收敛趋势, 例如 fr3_w_xyz 序列的平均误差与均方根误差分别降至 0.022 2、0.028 3, 同时该系统在低动态序列中的定位精度与前述系统基本持平, 充分验证了多模块协同优化机制对提升室内动态环境下 SLAM 系统定位精度的有效性。

为直观描述环境、提升算法实用价值, 本系统进一步引入稠密点云映射。通过在跟踪线程添加基于关键帧选择的密集映射线程, 将跟踪模块点云图输入并联合关

表 7 本文算法与其他算法的绝对轨迹误差实验对比结果

序列	ORB-SLAM2		ORB-SLAM3		DynaSLAM		DS-SLAM		本文方法	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
fr3_w_xyz	0.663 5	0.822 8	0.485 1	0.542 2	0.039 7	0.059 2	0.023 8	0.033 3	0.022 1	0.028 3
fr3_w_stt	0.054 6	0.061 8	0.016 0	0.023 2	0.009 3	0.010 6	0.009 1	0.010 2	0.009 3	0.010 5
fr3_w_rpy	0.730 8	0.849 8	0.680 7	0.819 0	0.109 0	0.161 6	0.094 2	0.150 3	0.025 9	0.030 8
fr3_s_stt	0.008 0	0.008 8	0.008 5	0.009 8	0.007 3	0.008 4	0.006 8	0.007 8	0.007 1	0.007 8

表 8 本文算法与其他算法的相对位姿误差实验对比结果

序列	ORB-SLAM2		ORB-SLAM3		DynaSLAM		DS-SLAM		本文方法	
	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
fr3_w_xyz	0.022 1	0.027 1	0.019 4	0.031 3	0.017 9	0.029 0	0.018 6	0.024 7	0.010 1	0.014 0
fr3_w_stt	0.019 2	0.024 7	0.009 8	0.015 7	0.006 2	0.007 5	0.007 3	0.008 1	0.007 8	0.009 2
fr3_w_rpy	0.026 3	0.084 7	0.021 9	0.028 7	0.022 1	0.045 0	0.376 8	0.444 2	0.020 2	0.027 2
fr3_s_stt	0.004 7	0.005 4	0.004 6	0.005 3	0.005 0	0.006 1	0.005 5	0.006 5	0.006 2	0.007 0

表 9 消融实验结果 m

序列	系统 A		系统 B		完整系统	
	Mean	RMSE	Mean	RMSE	Mean	RMSE
fr3_s_stt	0.007 4	0.009 1	0.006 8	0.008 0	0.007 1	0.007 8
fr3_w_stt	0.008 6	0.009 4	0.008 5	0.009 2	0.009 3	0.010 5
fr3_w_xyz	0.030 6	0.034 2	0.025 4	0.030 2	0.022 2	0.028 3
fr3_w_rpy	0.029 8	0.036 2	0.028 5	0.034 3	0.025 9	0.030 8

键帧位置信息组合，得到密集点云图。对比发现，ORB-SLAM3 获得的密集点云图受动态重影影响严重，重建效果较差。如图 9 所示，结合改进的语义分割模型来识别并消除动态干扰，动态环境下的重建质量和算法的实用价值都得到了显著提高。

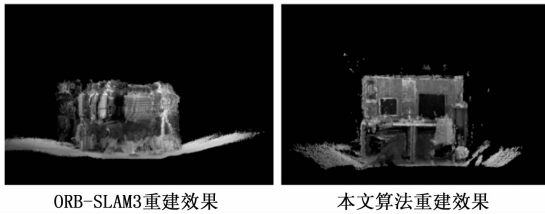


图 9 密集重建效果的比较

此外，本文在相同硬件条件下测试了本文算法的平均处理时间，并与其他算法进行了对比。实验结果如表 10 所示，由于实例分割引入了计算开销，导致本文算法的运行速度较慢于 ORB-SLAM3。然而，由于本文改进的网络模型实现了轻量化，并剔除了冗余的关键帧，本文算法的运行表现显著优于 DynaSLAM。

表 10 三种算法的平均追踪时间对比

系统	时间/ms
ORB-SLAM3	15.62
DynaSLAM	247.33
本文算法	47.84

4.4 真实场景中的实验结果分析

为了进一步验证该算法在动态环境中的性能，我们利用固定的 ASTRA-S 摄像机收集真实世界的动态场景数据，创建了一个包含 640×480 彩色图像和深度图像的自制数据集。通过该数据集来评估本文算法与 ORB-SLAM3 在真实世界动态场景中的对比。由于相机保持静止，因此实际轨迹是一个固定点。

实验结果如图 10 所示。相较于 ORB-SLAM3，本文算法能有效消除动态目标上的所有特征点。且本文算法的轨迹误差显著优于 ORB-SLAM3。如图 10 中的估计的相机轨迹图所示，当实际相机位置为单点时，ORB-SLAM3 受动态特征严重干扰，因动态物体的规律运动导致估计轨迹近似直线，完全偏离真实轨迹。而本文提出的算法成功剔除了动态特征，其估计位置始终精

确保持在原点附近，误差极小。

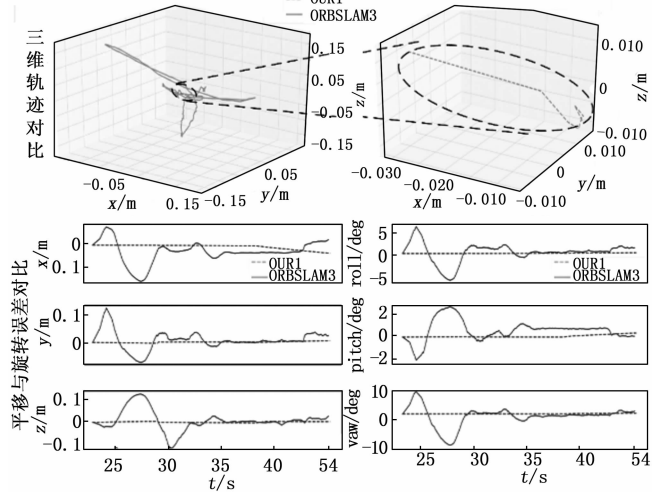


图 10 实际场景实验结果对比

此外，由图中的旋转平移误差曲线图可知，本文算法的平移误差达厘米级，三轴最高平移误差最高约为 2.5 cm，而 ORB-SLAM3 在三轴方向均存在分米级误差。本文算法的三轴旋转误差控制在 0.3 deg 以内，而 ORB-SLAM3 的三轴旋转误差则至少在 0.3 deg 以内，其中偏航角误差更是达到 10 deg。由此可见，本文算法在平移与旋转精度方面均实现显著提升。

5 结束语

SLAM 技术是智能机器人领域的研究重点之一。本文针对动态环境下视觉 SLAM 系统易受动态目标干扰、关键帧冗余和优化退化的问题提出了一种基于实例分割和多约束优化的动态稠密 SLAM 方法。通过采用改进的实例分割网络提取先验语义信息，进而剔除动态物体；并在特征提取阶段引入光照不变约束，结合基于信息论的关键帧选择和基于协方差约束的边缘管理策略，提高位姿估计精度，进而提高了系统定位性能。实验结果表明，改进方案的性能相较于基准系统 ORB-SLAM3 取得明显提高，在高动态的环境下绝对位姿误差的 RMSE 降低了 95% 以上、平均误差约降低了 95%，算法平移、旋转误差的 RMSE 分别下降 55.40%、49.75%，误差均值降幅分别达 48.07%、42.23%。由此可见，在动态环境中，相较于 ORB-SLAM3 和其他语义 SLAM 系统，本文算法的定位精度更高、综合性能相对

更好。尽管本文方法取得了较好的定位效果,但由于实例分割模型增加了系统计算开销,实时性仍有优化空间。此外,由于本文算法主要针对室内动态场景,所以当前验证主要集中于室内动态场景,对于室外复杂交通环境及超大规模动态场景的系统性验证仍待完善。未来工作将聚焦于模型轻量化策略,并进一步提升系统在更大规模复杂场景下的计算效率、鲁棒性与泛化性能。

参考文献:

- [1] 谷新建. 室外环境智能驾驶技术综述 [J]. 计算机测量与控制, 2024, 32 (10): 1-8.
- [2] 杨 鸥, 章文誉, 汪步云, 等. 基于全局位姿优化的移动机器人3D激光融合定位与建图 [J]. 计算机测量与控制, 2025, 33 (4): 209-216.
- [3] 蒋祥龙, 邓文亮, 何胜喜. 单目视觉驱动的机器人实时高精度稠密场景重建算法 [J]. 测绘通报, 2025 (10): 71-75.
- [4] MURARTAL 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.
- [5] CAMPOS C, ELVIRA R, RODRIGUEZ G J 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.
- [6] 徐淑萍, 杨定哲, 房嘉翔, 等. 动态场景优化 ORB-SLAM3 算法 [J]. 中国惯性技术学报, 2025, 33 (10): 998-1007.
- [7] 徐 武, 高 寒, 王欣达, 等. 改进 ORB-SLAM2 算法的关键帧选取及地图构建研究 [J]. 电子测量技术, 2022, 45 (20): 143-150.
- [8] 王 伟, 汤琴琴, 汪先伟. 基于特征点改进的视觉 SLAM 定位研究 [J]. 计算机测量与控制, 2024, 32 (2): 219-226.
- [9] 柳炳琦, 张峻山, 唐 辉, 等. 激光与视觉 SLAM 的算法综述 [J/OL]. 激光与光电子学进展, (2025-01-23) [2025-08-06]. <https://link.cnki.net/urlid/31.1690.TN.20250714.1516.054>.
- [10] SUN Y X, LIU M, MENG Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach [J]. Robotics and Autonomous Systems, 2016, 89: 110-122.
- [11] SCONA R, JAIMEZ M, PETILLOT Y R, et al. StaticFusion: Background reconstruction for dense RGB-D SLAM in dynamic environments [C] //Brisbane: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018: 3849-3856.
- [12] DAI W C, ZHANG Y, LI P, et al. RGB-D SLAM in dynamic environments using point correlations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44: 373-389.
- [13] SONG S, LIM H, LEE A J, et al. DynaVINS: a visual-Inertial SLAM for dynamic environments [J]. IEEE Robotics and Automation Letters, 2022, 7 (4): 11523-11530.
- [14] BESCOS B, FACIL 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.
- [15] WU W X, GUO L, GAO H L, et al. YOLO-SLAM: A semantic SLAM system towards dynamic environment with geometric constraint [J]. Neural Computing and Applications, 2022, 34: 6011-6026.
- [16] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C] //2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980-2988.
- [17] LI Y K, SHEN H B, FU Y P, et al. A method of dense point cloud SLAM based on improved YOLOv8 and fused with ORB-SLAM3 to cope with dynamic environments [J]. Expert Systems with Applications, 2025, 255: 124918.
- [18] 严瀚宇, 孙 博, 马天力, 等. 动态场景下融合深度信息的实时语义 SLAM 方法 [J]. 传感器与微系统, 2025, 44 (2): 139-142.
- [19] WANG L L, LETCHMUNAN S, XIAO R H. GelanSE: squeeze and stimulus attention based target detection network for Gelan architecture [J]. IEEE Access, 2024, 12: 182259-182273.
- [20] ZHOU S C, YANG L, LIU H T, et al. A lightweight drone detection method integrated into a linear attention mechanism based on improved YOLOv11 [J]. Remote Sensing, 2025, 17 (4): 705.
- [21] HAN K, WANG Y H, TIAN Q, et al. Ghostnet: more features from cheap operations [C] //Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 1577-1586.
- [22] HUANG Y, JIA W J, HE X J, et al. CAA: Channelized axial attention for semantic segmentation [J]. 2021. 10.48550/arXiv.2101.07434.
- [23] STURM J, ENGELHARD N, ENDRES F, et al. Creemers. a benchmark for the evaluation of RGB-D SLAM systems [C] // Vilamoura: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012: 573-580.
- [24] PALAZZOLO E, BEHLEY J, LOTTES P, et al. ReFusion: 3D reconstruction in dynamic environments for RGB-D cameras exploiting residuals [C] //Macau: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019: 7855-7862.