

针对室内动态场景的实时语义 SLAM 可视化算法^{*}

符 强^{1,2} 钟 振^{1,2} 纪元法^{1,2,3} 任风华^{1,2}

(1. 桂林电子科技大学广西精密导航技术与应用重点实验室 桂林 541004; 2. 桂林电子科技大学信息与通信学院 桂林 541004;
3. 时空信息与智能位置服务国际联合实验室 桂林 541004)

摘 要: 针对大部分视觉同步定位与建图(SLAM)在室内动态场景下定位精度及鲁棒性不足的问题,提出了基于 Photo-SLAM 算法框架改进的实时语义 SLAM 可视化系统。首先在原有算法框架中添加轻量级语义分割网络,针对图像中的动态物体特征进行剔除;同时增加掩膜边缘自适应膨胀方法,剔除由于语义分割边缘识别不足导致的动态特征残留点,进一步提高系统在动态环境中的定位精度。其次,在三维高斯溅射过程中引入语义信息,去除动态物体部分的高斯体,构建静态、平滑且连续的 3D 高斯地图。最后,在 Photo-SLAM 算法框架中添加稠密点云建图线程,根据语义信息与关键帧构建静态背景的稠密点云地图,并添加多种滤波方法进一步剔除离群点对点云地图的影响。在 TUM 数据集上进行的实验结果表明,相较于 Photo-SLAM 算法,改进后算法在低动态场景下的定位精度提升 40% 以上,在高动态场景下提升 93% 以上。与其他动态 SLAM 算法相比,算法在大部分场景下拥有更高的定位精度,且更具实时性,同时还可以创建去除动态物体干扰后的 3D 高斯地图以及稠密点云地图,实现静态背景的可视化。

关键词: 同时定位与建图; Photo-SLAM; 语义分割; 三维高斯溅射; 稠密建图

中图分类号: TP391.9; TN98 **文献标识码:** A **国家标准学科分类代码:** 510.99

Real-time semantic SLAM visualization algorithm for indoor dynamic scenes

Fu Qiang^{1,2} Zhong Zhen^{1,2} Ji Yuanfa^{1,2,3} Ren Fenghua^{1,2}

(1. Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin 541004, China; 2. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; 3. International Joint Laboratory of Spatiotemporal Information and Intelligent Location Services, Guilin 541004, China)

Abstract: To address the limitations of most visual simultaneous localization and mapping (SLAM) systems in terms of positioning accuracy and robustness within indoor dynamic environments, this paper proposes a real-time semantic SLAM visualization system based on an enhanced Photo-SLAM algorithm framework. First, a lightweight semantic segmentation network is added to the original algorithm framework to remove dynamic object features from the image. Additionally, an adaptive edge expansion method is introduced to eliminate residual dynamic feature points caused by insufficient edge detection in semantic segmentation, further enhancing the system's localization accuracy in dynamic environments. Second, semantic information is introduced into the 3D Gaussian Splatting process to remove the Gaussian bodies of dynamic objects, constructing a static, smooth, and continuous 3D Gaussian map. Finally, a dense point cloud mapping thread is added to the Photo-SLAM algorithm framework to construct a dense point cloud map of the static background based on semantic information and key frames, and multiple filtering methods are added to further remove the influence of outliers on the point cloud map. Experimental results on the TUM dataset show that the improved algorithm achieves over 40% higher localization accuracy than the Photo-SLAM algorithm in low-dynamic scenes and over 93% higher accuracy in high-dynamic scenes. Compared to other dynamic SLAM algorithms, the proposed algorithm achieves higher localization accuracy in most scenarios and is more real-time. It can also create 3D Gaussian maps and dense point cloud maps after removing the interference of dynamic objects, enabling visualization of the static background.

Keywords: simultaneous localization and mapping; Photo-SLAM; semantic segmentation; 3D Gaussian splatting; dense mapping

0 引 言

随着人工智能、机器人技术以及智能感知系统的迅猛

发展,同步定位与建图(simultaneous localization and mapping, SLAM)已逐步成为实现自主导航与环境理解的核心技术,并受到学术界与工业界的广泛关注。SLAM 系

收稿日期:2025-07-17

^{*} 基金项目:广西科技计划项目(AA23062038, ZY23055048, AB23026120, AA24206025, AA24206043)、国家自然科学基金(U23A20280, 62161007, 62471153)、南宁市科学研究与技术开发计划(20231029, 20231011)、产研计划项目(CYY-HT2023-JSJJ-0023-1, CYY-HT2023-JSJJ-0024-1)、广西科技基地和人才专项(AD25069103)资助

统能够使移动机器人在未知环境中利用不同传感器采集数据构建环境地图的同时实现精准定位,为自动驾驶、无人机、空间测绘等领域提供支持。根据传感器的不同,SLAM 系统大致分为激光、视觉以及多传感器融合。其中,基于相机的视觉 SLAM 凭借成本低、信息丰富等优点已逐步成为研究热点,并涌现出 ORB-SLAM2^[1]、ORB-SLAM3^[2]、LSD-SLAM^[3]、DSM^[4]等极具代表性的开源算法,为该领域的研究奠定了坚实基础。然而传统的 SLAM 系统大多基于静态环境假设,在包含多种动态物体的场景中,鲁棒性和精度会显著下降。

针对动态环境下 SLAM 系统定位精度及鲁棒性不足的问题,近年来许多学者也进行了相关的研究和改进。DynaSLAM^[5]在 ORB-SLAM2 基础上加入 Mask R-CNN^[6]进行语义分割,并结合多视图几何一致性检测,剔除动态目标,提升在动态场景下的系统定位鲁棒性。DS-SLAM^[7]采用 SegNet^[8]语义分割和移动一致性检测联合剔除动态特征点,减少动态物体对定位过程的影响。RDS-SLAM^[9]在 ORB-SLAM3 的基础上添加语义分割,并通过优化语义分割策略与关键帧选择策略,提升了动态场景中特征点筛选的准确性和系统运行效率。SG-SLAM^[10]基于 ORB-SLAM2 框架加入了两个并行线程,用于获取语义信息以及构建语义地图,并结合语义和几何信息进行动态特征剔除。栾添添等^[11]使用傅里叶-梅林方法进行运动补偿后提取运动掩膜,并结合语义分割得到物体掩膜,联合识别运动物体区域进行动态剔除。以上算法通过引入语义分割与动态特征剔除策略,提升了动态环境下定位鲁棒性,但在实时性与地图构建方面仍略有不足。因此,如何在保证系统实时性的前提下,进一步提升语义分割在复杂动态场景中对目标边缘的识别能力,仍是值得关注的问题。已有研究表明,在复杂或多模态视觉场景中引入语义分割并结合多尺度特征融合策略,有助于改善目标边缘模糊与不稳定现象,从而为后续定位与建图任务提供更加可靠的先验信息^[12-13]。

三维高斯溅射(3D Gaussian splatting, 3DGS)^[14]是一种基于 3D 高斯分布的光栅化技术,主要用于实时渲染高分辨率图像和场景。通过将稀疏点云表示为 3D 高斯分布,结合优化和快速渲染算法,实现高质量的地图渲染。Gaussian Splatting SLAM^[15]首次将 3DGS 应用于单目和 RGB-D 相机,实现高效渲染与场景表达,并引入 SLAM 框架,通过联合优化相机位姿与高斯参数,实现定位与建图。MotionGS^[16]提出了一种融合深度特征、双关键帧选择和 3DGS 的 SLAM 方法,通过特征提取与运动滤波实现跟踪,并在整个建图过程中联合优化相机姿态与高斯参数,从而实现高效精准的定位与紧凑的高斯场景表达。Photo-SLAM^[17]提出了一种基于超基元地图的新型 SLAM 框架,结合显式几何特征用于定位,隐式光度特征用于纹理建模,并引入高斯金字塔训练策略逐步优化多层特征。但以上基

于 3DGS 的 SLAM 系统都基于静态环境假设,在动态场景中容易受到运动物体干扰,导致定位精度与建图质量严重下降。

针对以上问题,本文围绕 SLAM 系统在室内动态环境中鲁棒性不足以及建图质量差的问题,提出了一种基于 Photo-SLAM 的改进算法。首先,该算法在定位过程中引入轻量级 SegFormer^[18]语义分割网络,剔除大部分动态特征点;然后,添加掩膜边缘自适应膨胀方法,进一步剔除由于动态物体边缘识别不足导致的残留特征点;接着,在 3DGS 过程中引入语义信息,构建无动态噪声干扰的 3D 高斯地图;最后,添加稠密建图线程,生成静态环境稠密点云地图,并使用多种滤波方法剔除离群点干扰。

1 系统框架

本文算法在 Photo-SLAM 系统的基础上进行改进,系统框图如图 1 所示。包括跟踪线程、局部建图线程、稠密建图线程、回环检测线程、高斯建图线程和可视化线程。

跟踪线程是整个系统的核心线程,负责接收图像数据并进行相机位姿估计。先对每一帧 RGB 图像进行 ORB 特征提取和匹配,在加入 SegFormer 语义分割模块后,系统会同步获取当前帧的语义掩膜信息,判断动态物体区域。在此基础上,添加掩膜边缘自适应膨胀方法,对语义掩膜边界进行动态扩张,从而更精准地剔除潜在动态区域特征点并保留静态区域特征点,完成相机位姿和地图点的初始化。接着利用地图中的关键帧进行姿态估计,并通过运动模型或重定位机制实现相机位姿的连续跟踪,同时将生成的关键帧传递给局部建图线程。

局部建图线程接收来自跟踪线程的关键帧后,首先插入地图,然后进行局部光束平差(bundle adjustment, BA)优化以提升关键帧和地图点的精度。同时,该线程还会进行地图点的三角化生成、新旧点的筛选和冗余关键帧剔除等操作,从而不断完善局部地图结构。

稠密建图线程在接收到新的关键帧后,结合相应的深度图像和 SegFormer 语义分割掩膜,将每个关键帧中的深度数据投影为稠密点云,并连续融合到全局点云地图中。同时加入点云滤波方法进一步滤除离群点云,最后保存全局点云地图。

回环检测线程负责检测和处理回环,一旦检测到与当前关键帧相似的历史关键帧,就进行 Sim3 相似变换求解和姿态对齐,然后构建闭环图并触发全局图优化,从而有效消除累计漂移,提高全局一致性。

高斯建图线程根据关键帧的图像、深度图、相机位姿以及语义信息,将静态区域的特征点转换到三维空间,逐点融合为全局高斯地图。并通过更新和优化,使得地图结构更加平滑、连续。

可视化线程负责实时渲染,包括相机轨迹、关键帧、地图点以及高斯地图。将跟踪视图、自由视角以及 3DGS 结

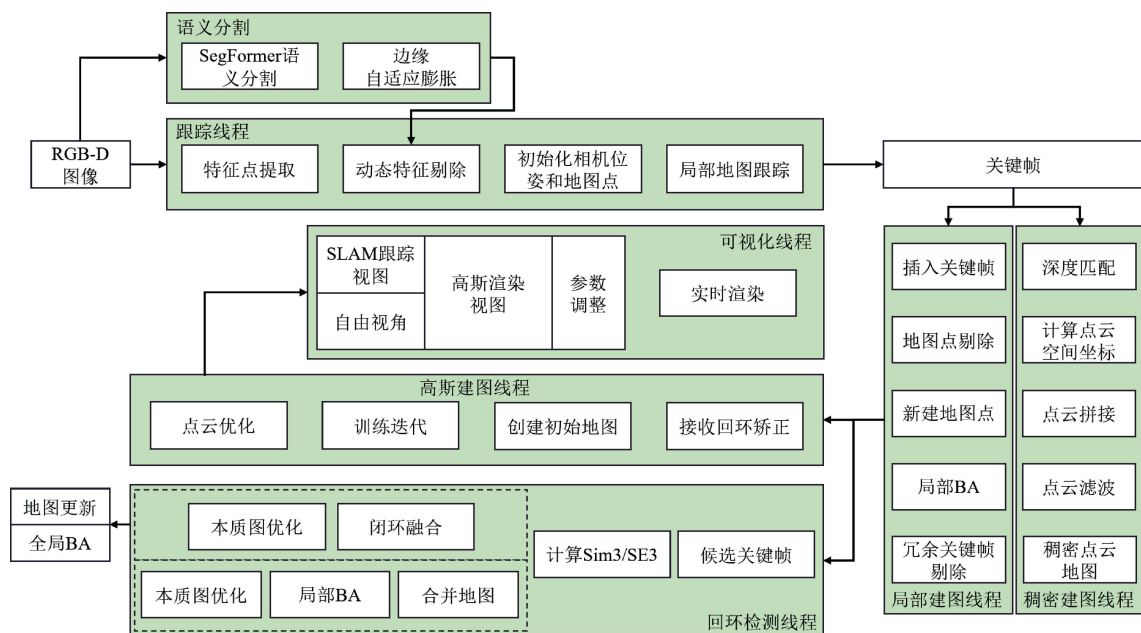


图 1 系统框图

Fig. 1 System flowchart

果可视化,直观展示系统的运行状态。该线程支持交互式查看视角、启用或关闭各类可视元素。

1.1 基于 SegFormer 与边缘自适应膨胀的动态点剔除

SegFormer 是近年提出的一种高效语义分割模型,其结构由一个层级式 Transformer^[19] 编码器和一个轻量级多层感知机 (multilayer perceptron, MLP)^[20] 解码器组成。其编码器采用分层结构,通过多层 Transformer 模块逐步提取图像的多尺度特征,每一层都生成不同分辨率的特征图,从而在保持空间细节的同时获取全局语义信息。与传统视觉 Transformer 不同,SegFormer 的编码器不依赖位置编码,而是通过结构本身保留位置信息,使模型在输入图像大小变化时依然具有良好的适应性。解码器部分则使用 MLP 对来自各层的特征图进行通道对齐和特征融合,所有特征被统一映射到相同的通道数后上采样至相同的分辨率,并进行逐像素融合,最终输出语义分割图。这种结构简洁、高效,能够在保持分割精度的同时,降低模型复杂度,提高推理速度。

为了提升本文算法运行的实时性,采用语义分割网络与 SLAM 系统并行的方式进行协同处理。同时为了降低计算开销,对语义分割模型进行了轻量化设计,采用较小的模型进行语义分割推理,但这在一定程度上导致了分割精度的下降。

通常情况下,SLAM 系统直接采用语义分割掩膜或使用固定大小膨胀核的掩膜进行动态特征剔除。式(1)为掩膜膨胀公式。

$$A \oplus B = \{z \mid (B)_z \cap A \neq \emptyset\} \quad (1)$$

其中, A 表示原始的二值化掩膜图像(前景像素点集

合), B 是结构元素(如 3×3 或 5×5 的卷积核), $(B)_z$ 表示结构元素平移至位置 z , $A \oplus B$ 即为膨胀后的掩膜结果,只要结构元素覆盖区域与原始掩膜有重叠,中心点就被标记为前景,实现掩膜的扩展。

上述这种方式在分割精度较高的条件下能够获得较为理想的剔除效果,但其语义分割所需要的时间较长,且掩膜无法根据动态目标的形态变化进行自适应调整。由于语义分割模型轻量化所导致的边缘分割误差或掩膜覆盖不足的问题较为普遍,固定膨胀核在处理目标边界模糊或形态不规则区域时,容易出现掩膜膨胀不足(图 2(a))导致动态特征剔除不足(图 2(b)),或膨胀过度(图 3(a))影响静态区域的情况(图 3(b)),从而影响系统定位精度与稳定性。

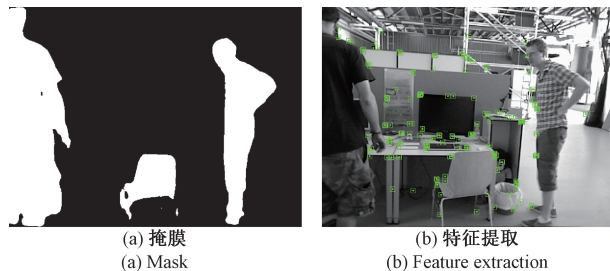


图 2 特征剔除不足

Fig. 2 Insufficient feature removal

为此,本文提出了一种掩膜边缘自适应膨胀方法,提升动态目标剔除的有效性。该方法针对语义掩膜边缘区域,通过动态调整膨胀核的大小和膨胀次数,实现对边缘区域的针对性扩张。

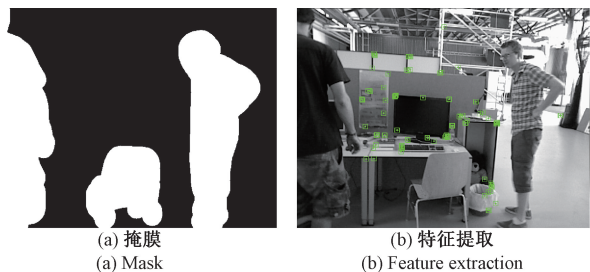


图 3 特征剔除过度

Fig. 3 Feature removal overkill

系统首先根据语义掩膜边缘的梯度强度、目标区域尺寸以及分割结果的边界连续性,调整初始膨胀参数。当边缘梯度较弱或目标区域较大时,系统采用较大的膨胀核以补偿语义分割边缘识别不足;而当边界梯度较高或轮廓清晰时,则采用较小的膨胀核以避免静态区域被误扩张。随后,系统根据膨胀后特征点的空间分布评估掩膜覆盖是否充分。首先,统计掩膜区域与非掩膜区域特征点的数量比例,以此确定掩膜覆盖度。当覆盖度较低时,说明动态区域仍存在未被完全标记的残留特征点,系统将自适应地增大膨胀核执行一次额外的膨胀操作;当覆盖度较高时,则表明掩膜已充分覆盖动态区域,无需进一步膨胀。从而实现自适应双阶段扩张,每次膨胀操作仅执行一次迭代,以控制计算复杂度并保证实时性能。

这种方法在保持系统运行效率的前提下,弥补轻量分割模型在边界细节上的缺失,合理扩展动态区域的掩膜覆盖范围,从而增强对动态特征的剔除能力,避免对静态区域的误剔除。此外,也有研究表明,在动态场景中融合语义信息能够显著提升动态特征区域的检测与剔除效果,文献[21]提出一种基于稀疏光流的语义视觉 SLAM 方法,通过结合光流一致性与语义掩膜有效识别运动区域,提高了系统在复杂动态环境中的鲁棒性。文献[22]则利用全景分割和多视图几何约束联合判断动态目标区域,实现了较高精度的动态特征剔除。上述研究的结论均进一步验证了在动态场景中的引入语义信息进行动态特征剔除的必要性与有效性。

1.2 静态特征 3DGS

近年来,三维场景表示方法快速发展,其中基于隐式场的三维建模技术在稠密点云表达与高精度场景重讲方面展现出优越性能。作为典型代表之一的神经辐射场(neural radiance field, NeRF)通过可微的隐式场模型描述场景几何与外观,在视觉 SLAM 建图中展现出良好效果。文献[23]将 NeRF 与稠密点云计算结合,通过隐式场表示实现了连续、高精度的三维几何复原,有效提升了场景细节表达能力与地图一致性。此类研究表明,基于隐式场或密度场的三维表达在提升 SLAM 建图质量方面具有显著潜力。

在隐式场框架下,除了 NeRF 外,近年来也涌现出一

类基于连续密度场的显式参数化表示,其中 3DGS 是具有代表性的方向。3DGS 是一种基于连续密度场的三维场景建模方法,其核心思想是通过在三维空间中分布大量带有属性的高斯分布体,构建出一个密度连续、表达可微的空间表示。每个高斯体被视为一个渲染单元,具备位置 μ 、不透明度 α 、三维协方差矩阵 Σ 和颜色 c 这 4 种属性,并且反向传播进行优化,从而能够在空间中形成局部有形的概率体积。其中协方差矩阵由缩放变换 S 和旋转变换 R 组合得到,如式(2)所示。

$$\Sigma = RSS^T R^T \quad (2)$$

在构建地图的过程中,当要求区域没有被充分覆盖时(欠重建),克隆相应的高斯体;当要求区域被过度覆盖时(过重建),将高斯体分成两个部分,如图 4 所示。

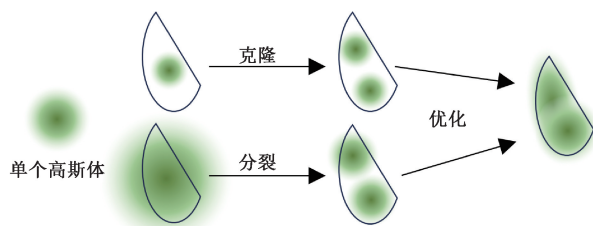


图 4 3DGS 过程示意图

Fig. 4 Schematic diagram of 3DGS process

Photo-SLAM 系统主要由 4 个不同模块构建并维护 1 个超基元地图,包含定位模块、几何映射模块、真实映射模块以及闭环检测模块。系统将超基元定义为一种点云集合,其与以下多种属性相关联,包括 ORB 特征、旋转参数、缩放参数、密度值以及球谐系数。系统根据图像帧中提取的 ORB 特征构建 2D-2D 及 2D-3D 对应关系,并根据相邻帧之间充足的 2D-2D 匹配点估算变换矩阵后,通过三角测量初始化超基元地图并启动位姿跟踪。跟踪阶段中,定位模块通过 2D-3D 对应关系实时解算相机位姿,几何映射模块同步增量式初始化稀疏超基元,真实映射模块则负责优化由几何映射线程逐步创建的超基元,通过基于平铺的渲染器将超基元光栅化,以生成与关键帧姿态对应的图像。渲染公式如式(3)所示。

$$C(R, t) = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (3)$$

其中, N 表示超基元的数量, c_i 表示从球谐系数转换的颜色, α_i 表示密度 σ_i 与 3D 高斯分布^[14] $\mathcal{G}(R, t, P_i, r_i, s_i)$ 的乘积,优化通过最小化渲染图像 I_r 与真实图像 I_{gt} 之间的光度损失 L 来进行,光度损失公式如式(4)所示。

$$L = (1 - \lambda) \|I_r - I_{gt}\|_1 + \lambda (1 - \text{SSIM}(I_r, I_{gt})) \quad (4)$$

其中, $\text{SSIM}(I_r, I_{gt})$ 表示两幅图像之间的结构相似性, λ 是平衡因子。

尽管 Photo-SLAM 系统在室内静态场景中能够实现较为真实的 3DGS 效果,但其在动态环境下的鲁棒性明显

不足。由于未对动态区域进行识别与剔除,图像中的移动物体往往被误认为是静态背景而参与特征匹配与三角测量,进而影响相机位姿估计的准确性,并且在后续 3DGS 过程中引入大量动态特征的干扰。动态特征误匹配不仅会导致超基元密度分布失真,还会在最终渲染图像中形成动态残影,影响建图效果。

为解决上述问题,本文在原有框架基础上提出了一种面向室内动态场景的去动态 3DGS 方法。该方法引入轻量级语义分割网络 SegFormer,对输入图像进行像素级语义分割以识别动态物体,并结合掩膜边缘自适应膨胀策略扩展动态区域边界,确保动态区域在图像处理阶段被充分剔除。在特征提取过程中,系统仅保留静态区域内的特征点,并结合深度图和相机位姿进行三维反投影,生成高斯体的初始位置、颜色和协方差等属性。随后,系统在多关键帧之间融合空间重叠的高斯体,对其属性进行联合优化,构建出连续稠密、无动态干扰的静态高斯集合。最终,渲染阶段仅基于静态高斯进行重投影与光度损失优化,避免动态特征带来的误差积累,提升系统在动态场景下的鲁棒性。改进后的 3DGS 部分流程如图 5 所示。

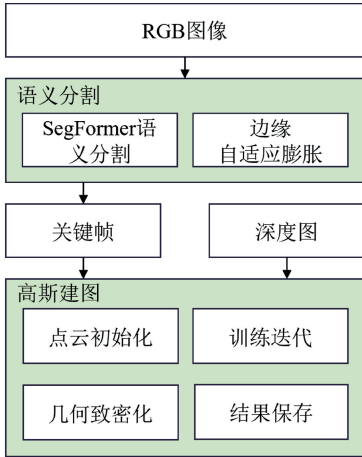


图 5 改进后的 3DGS 流程

Fig. 5 Improved 3DGS process

1.3 静态特征稠密点云建图

除了 3D 高斯地图外,为进一步提升地图结构的完整性,本文还添加了基于静态特征的稠密点云建图模块,作为对 3DGS 的有效补充。

该模块利用关键帧对应的深度图与自适应膨胀后的语义掩膜信息,对动态物体区域进行剔除,确保仅有静态区域参与建图,并通过相机内参和位姿信息将其反投影到三维空间,生成带有颜色信息的稠密点云。点云投影依据针孔成像原理,数学模型如式(5)所示。

$$s \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{K} \left(\mathbf{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbf{t} \right) = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \left(\mathbf{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \mathbf{t} \right) \quad (5)$$

式中: $\mathbf{R}$ 为相机旋转向量, $\mathbf{t}$ 为相机平移向量, $\mathbf{K}$ 是相机内参矩阵, s 是深度尺度取固定值。利用深度值 d 和像素位置 (u, v) , 即可计算出点云空间坐标 (x, y, z) , 其转换关系如式(6)所示。

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} (u - c_x) \cdot \frac{z}{f_x} \\ (v - c_y) \cdot \frac{z}{f_y} \\ \frac{d}{s} \end{pmatrix} \quad (6)$$

随后采用逐帧融合策略,将每帧的静态点云持续累积形成完整的全局稠密点云地图。在融合过程中,首先采用体素滤波,将空间划分为三维网格每个体素保留 1 个代表点,在减少点云数量降低计算量的同时,起到一定的平滑和噪声抑制作用。再使用统计离群点剔除,即每个点与其近邻点之间的距离显著大于全局平均,则为离群点进行剔除。先计算每个点的平均距离 d_i 如式(7)所示。

$$d_i = \frac{1}{k} \sum_{j=1}^k \| \mathbf{p}_i - \mathbf{p}_{ij} \|_2 \quad (7)$$

其中, $\mathbf{p}_i$ 表示当前点, $\mathbf{p}_{ij}$ 表示第 j 个邻近点, k 表示邻近点数量, $\| \mathbf{p}_i - \mathbf{p}_{ij} \|_2$ 表示两点之间的欧氏距离。

然后计算所有点的均值(式 8)与标准差(式 9)。

$$\mu_d = \frac{1}{N} \sum_{i=1}^N d_i \quad (8)$$

$$\sigma_d = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - \mu_d)^2} \quad (9)$$

其中, N 为点云总数,如果 $d_i >$ 全局平均值则视为离群点进行剔除。最后使用半径离群点剔除方法进行进一步剔除,如式(10)所示。

$$\mathcal{N}_i = \{ \mathbf{p}_j \in P \mid \| \mathbf{p}_i - \mathbf{p}_{ij} \|_2 < r \} \quad (10)$$

其中, r 半径, $\mathcal{N}_i$ 为邻近点集合,如果一点在半径内的邻近点数低于阈值,则为离群点进行剔除。使用多种滤波方法结合,进一步提高点云地图的一致性与稳定性。

在完成滤波算法的设计之后,本文进一步对滤波算法的关键参数进行了实验与优化,以平衡去噪能力和结构保真度。首先,在体素滤波阶段,体素叶片大小设为 0.02 m。该参数依据深度相机在 1~6 m 内的测距精度与空间分辨率确定,可在保持几何细节的同时显著降低点云规模与计算负担。其次,在统计离群点剔除阶段,邻域点数量为 50、标准差倍数阈值为 0.5。该设置基于点云局部密度分布统计结果,在多数室内场景中 50 个邻域点能稳定反映局部结构特征,而 0.5 倍标准差阈值可有效抑制局部孤立点的同时避免边缘过度平滑。最后,在半径离群点剔除阶段,搜索半径设为 0.06 m,最小邻居点数阈值为 6。该半径根据相机内参与深度分辨率推算,约对应相机中等距离处 3~4 pixel 的空间覆盖范围,可有效剔除稀疏边缘点与误匹配点。3 种滤波算法依次融合实现由全局到局部的多层

次滤波优化,体素滤波控制点云整体密度,统计滤波抑制局部噪声,半径滤波强化空间连续性。既保证了全局点云的平滑一致性,又保持了静态结构的精细几何特征。

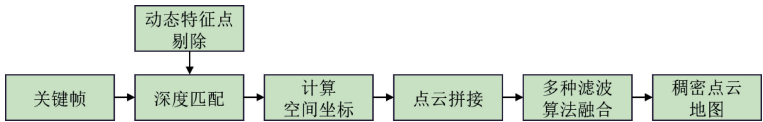


图 6 稠密点云构建流程

Fig. 6 Dense point cloud construction process

2 实验测试

本文算法运行的硬件配置为 Intel (R) Core (TM) i5-13600KF CPU,32 G 运行内存,GPU 为 NVIDIA RXT 4070 Ti,操作系统为 Ubuntu 20. 04,CUDA 11. 8。

2.1 系统定位精度对比

本文系统定位精度实验采用 TUM RGB-D 数据集,其包含低动态序列 fr3_s,主要是人物坐姿状态下的小幅度动作,以及高动态序列 fr3_w,主要是人物行走状态下的大幅度动作。序列不同后缀 static、xyz、rpy、half 表示相机不同运动状态,分别对应相机静止不移动,相机沿 xyz 轴平移运动,相机沿滚转、俯仰、偏航三轴角度变化的旋转运动以及相机沿半球面轨迹运动。

系统定位精度评价指标采用绝对轨迹误差 (absolute

最终生成的静态稠密点云地图不仅可以为 3D 场景的可视化提供直观的结构信息,也可作为后续地图评估与几何精度对比的重要依据。稠密点云建图部分流程如图 6 所示。

trajectory error, ATE),使用 evo 评估工具计算其均方根误差 (root mean squared error, RMSE) 和标准差 (standard deviation, STD),衡量系统误差的平均水平以及波动。

不同序列下消融实验中绝对轨迹误差的 RMSE 与 STD 如表 1 所示。其中 SegFormer 部分表示算法仅添加了初始掩膜图像进行动态剔除的效果,本文算法部分则表示使用边缘自适应膨胀方法处理掩膜图像后进行动态剔除的效果,提升部分表示本文算法与 Photo-SLAM 算法相比得出的提升率。从实验数据中看出,在高动态序列中本文算法相较于改进前原算法的定位精度有显著提升,绝对轨迹误差的 RMSE 降低 93% 以上。同时在人物移动范围较小且相机基本保持静止的低动态序列下,本文算法仍能将绝对轨迹误差的 RMSE 减小 40% 以上,进一步提升系统精度,表现出更好的稳定性和适应性。

表 1 不同序列消融实验绝对轨迹误差对比

Table 1 Comparison of absolute trajectory errors in ablation experiments with different sequences

序列	Photo-SLAM		SegFormer		本文算法		提升	
	RMSE/m	STD/m	RMSE/m	STD/m	RMSE/m	STD/m	RMSE/%	STD/%
fr3_s_static	0.010 1	0.004 8	0.007 7	0.003 4	0.006 0	0.003 0	40.59	37.50
fr3_w_static	0.091 7	0.061 4	0.013 5	0.006 6	0.006 0	0.002 8	93.46	95.44
fr3_w_xyz	0.888 9	0.459 1	0.016 7	0.008 3	0.014 0	0.007 0	98.43	98.48
fr3_w_rpy	0.978 0	0.538 2	0.062 6	0.050 9	0.042 0	0.028 9	95.71	94.63
fr3_w_half	0.372 5	0.134 4	0.027 9	0.013 9	0.023 6	0.011 9	93.66	91.15

本文算法与其他动态 SLAM 算法在不同序列中绝对轨迹误差的 RMSE 对比如表 2 所示。包含 Dyna-SLAM、RDS-SLAM、SG-SLAM,表中加粗部分表示每个序列下定位精度最高的方案。由实验结果可以看出,相较于其他 3 种动态 SLAM 算法,本文算法在低动态序列中达到了与其他先进算法相同的定位精度,在 fr3_w_static、fr3_w_xyz、fr3_w_half 三组高动态序列中的表现优于其他算法。

本文算法主要依赖语义分割掩膜对动态特征进行剔除,以减少动态物体对定位精度的影响。在 fr3_w_rpy 序列中,相机以旋转运动为主,帧间平移视差小,导致深度与几何一致性难以有效利用。当语义分割在边缘区域存在轻微误差,或前景物体短暂静止时,系统可能将部分静态特征误判为动态特征并予以剔除,从而造成可用匹配点数量下降,

位姿估计精度降低。相比之下,SG-SLAM 在极线约束下结合动态权值判断,DynaSLAM 融合多视图几何一致性与深度检测机制,即使在纯旋转或低视差场景下仍能维持动态与静态区域的可靠区分,因此在该序列上表现更优。

此外,图 7 所示为 Photo-SLAM 算法与本文算法在高动态序列下的绝对轨迹误差,实线代表真实轨迹,点画线代表估计轨迹,虚线部分代表两者之间存在的误差。从图中可以看出,在动态场景下本文算法估计轨迹如图 7(b)、(d)、(f)、(h) 所示,相较于 Photo-SLAM 算法的估计轨迹如图 7(a)、(c)、(e)、(f) 所示,误差更小,与真实轨迹更贴合,在定位精度和鲁棒性上均有明显提升。

2.2 3DGS 效果对比

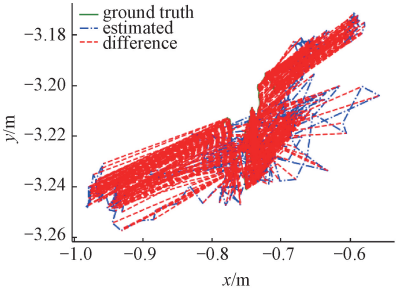
本文还进行了 3DGS 实验,并在高动态场景下与 Photo-

表 2 不同算法绝对轨迹误差对比

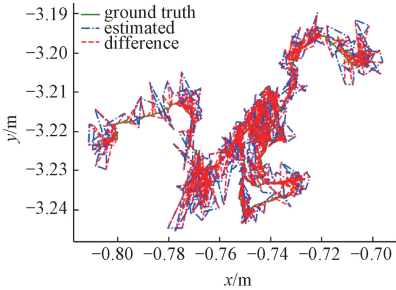
Table 2 Comparison of ATE between different algorithms

(m)

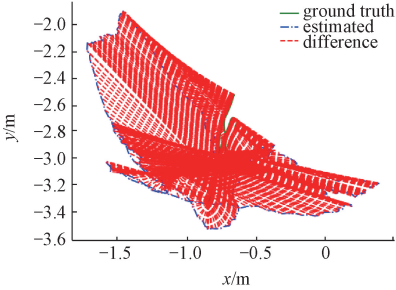
序列	DynaSLAM	RDS-SLAM(Mask R-CNN)	SG-SLAM	本文算法
fr3_s_static	0.006 0	0.008 8	0.006 0	0.006 0
fr3_w_static	0.009 0	0.081 5	0.007 3	0.006 0
fr3_w_xyz	0.015 0	0.021 3	0.015 2	0.014 0
fr3_w_rpy	0.035 0	0.146 8	0.032 4	0.042 0
fr3_w_half	0.025 0	0.025 9	0.026 8	0.023 6



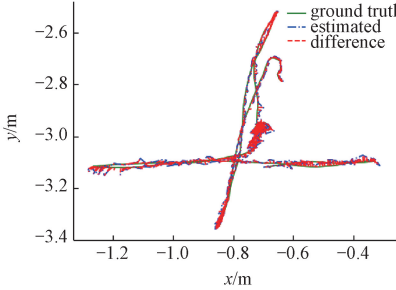
(a) Photo-SLAM/fr3_w_static



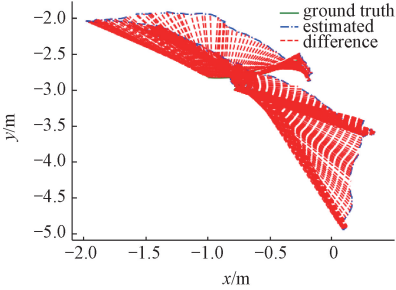
(b) 本文算法/fr3_w_static
(b) Ours algorithms/fr3_w_static



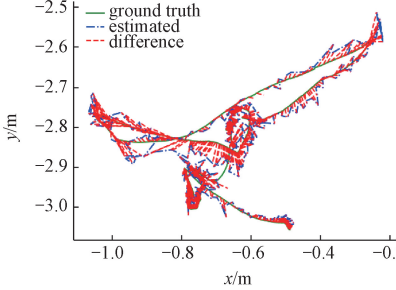
(c) Photo-SLAM/fr3_w_xyz



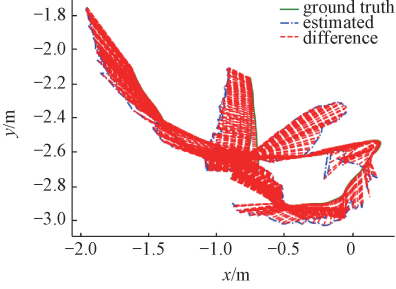
(d) 本文算法/fr3_w_xyz
(d) Ours algorithms/fr3_w_xyz



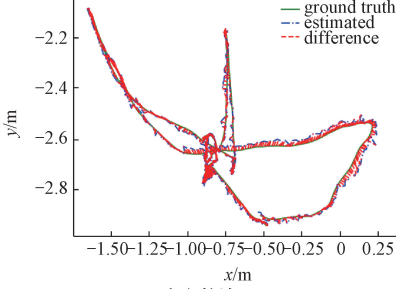
(e) Photo-SLAM/fr3_w_rpy



(f) 本文算法/fr3_w_rpy
(f) Ours algorithms/fr3_w_rpy



(g) Photo-SLAM/fr3_w_half



(h) 本文算法/fr3_w_half
(h) Ours algorithms/fr3_w_half

图 7 轨迹误差对比

Fig. 7 Trajectory error comparison

SLAM 算法生成的高斯地图效果进行对比。图 8 所示为实际场景和两种算法的建图结果,其中图 8(a)为真实场景,图 8(b)为 Photo-SLAM 算法构建的高斯地图,图 8(c)为本文算法的 3DGS 效果。



图 8 高斯建图效果对比
Fig. 8 Comparison of Gaussian mapping effects

从图 8 可以观察到,Photo-SLAM 在 3DGS 过程中未考虑动态目标剔除,导致图像中的动态人物被误认为是静态背景参与建图,从而在高斯体分布中引入大量动态残影,人物位置处的高斯密度异常聚集,严重影响了地图的整体结构与渲染效果。

相比之下,本文结合语义信息添加的掩膜边缘自适应膨胀策略能够针对性扩展动态区域边界,更全面地覆盖动态物体轮廓,在 3DGS 过程中剔除了大部分动态区域的高斯体,使最终生成的高斯地图结构更加清晰且无动态残留。尤其在人物移动轨迹密集的区域,本文算法有效避免了动态物体高斯体叠加的现象,地图纹理表现更为平滑,整体密度分布更合理,视觉一致性得到提升。

实验结果说明,本文提出的静态特征 3DGS 策略在动态环境下具备更强的鲁棒性和更优的建图效果,能够有效

抑制动态物体干扰,对传统 3DGS 进行了进一步优化。

2.3 稠密点云建图效果对比

此外,本文算法还针对稠密点云地图构建功能做了改进。图 9(a)为去动态之前的稠密点云地图,图 9(b)为仅使用掩膜去除动态特征构建的稠密点云地图,图 9(c)为使用掩膜去除动态特征后再使用点云滤波方法构建的稠密地图。由对应图像可以看出,在未使用语义分割方法之前图像中存在大量动态人物移动残影,严重影响建图效果。在添加了语义分割方法后,虽然去除了地图中大部分的人物残影,但仍然存在部分离群点云。在加入语义分割的基础上添加点云滤波后,生成的稠密点云噪声明显减小,建图效果得到明显提升。该实验结果表明,本文算法能够在动态环境中构建无动态特征影响的静态稠密点云地图。

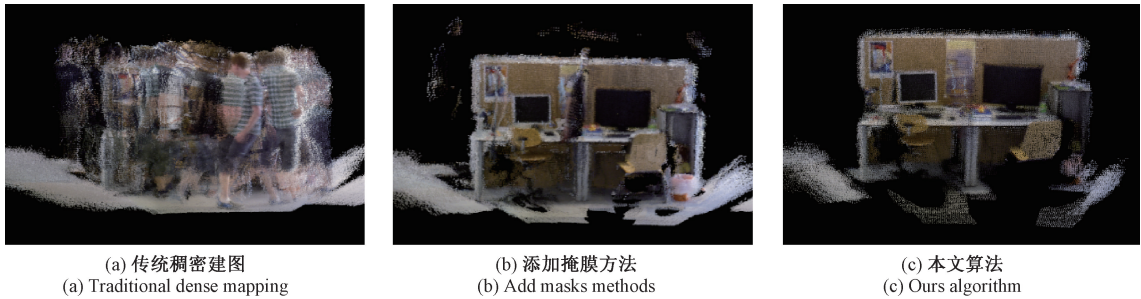


图 9 稠密建图效果对比
Fig. 9 Comparison of dense mapping effects

2.4 实时性评估

为全面评估算法的实时性,本文首先比较了系统总体运行时间与帧率。以 fr3_w_xyz 序列为例,在完整运行条件下,本文算法平均帧率约为 18 fps,略低于原 ORB-SLAM3 的 22 fps 和改进前原算法 Photo-SLAM 的 21.94 fps,如表 3 所示。在此基础上,本文还测试了每帧平均语义分割时间以及平均跟踪时间,并与其他算法进行对比,结果如表 4 所示。本文采用 SegFormer 语义分割模型处理图片平均每帧耗时 48.3 ms,相比于 DynaSLAM 和 RDS-SLAM

表 3 运行效率对比

Table 3 Operational efficiency comparison		
算法	运行时间/s	帧率/fps
ORB-SLAM3	37.11	22.28
Photo-SLAM	37.69	21.94
本文算法	45.38	18.22

算法使用的 Mask R-CNN 分割速度得到显著提升,虽然耗时仍略多于 DS-SLAM 和 RDS-SLAM 算法使用的 SegNet

表 4 不同算法实时性对比

Table 4 Real-time comparison of different algorithms

算法	分割网络	(ms)	
		分割 时间	跟踪 时间
ORB-SLAM3	—	—	9.6
DynaSLAM	Mask R-CNN	195.0	>300.0
DS-SLAM	SegNet	37.5	>65.0
RDS-SLAM	Mask R-CNN	200.0	50.0~65.0
RDS-SLAM	SegNet	30.0	50.0~65.0
本文算法	SegFormer	48.3	19.4

语义分割网络,但本文算法在平均跟踪时间上,相比于 ORB-SLAM3 仅增加了 10 ms 左右,耗时明显少于其他算法。综合来看,本文算法满足视觉 SLAM 系统运行的实时性要求。

3 结 论

本研究提出了一种基于室内动态环境下的实时语义视觉 SLAM 算法,该算法在 Photo-SLAM 的基础上进行改进,增加了语义分割部分和稠密建图线程,并加入边缘自适应膨胀方法针对掩膜边缘残留的动态特征点进行进一步剔除,利用语义信息消除了定位过程中动态物体对系统的影响。同时,在 3DGS 的过程中利用语义掩膜去除了动态物体的残影,消除了累计误差,使系统能够构建连续且平滑的 3D 高斯地图。最后,利用语义信息和滤波方法消除动态物体和离群点的影响,提升稠密建图效果。在 TUM 数据集上将本文算法与 DynaSLAM、RDS-SLAM 以及 SG-SLAM 3 种动态 SLAM 算法进行对比,结果表明本文算法相较于其他先进 SLAM 算法有较高的定位精度,同时算法的鲁棒性和实时性也得到一定的提升。但本文算法仅在室内环境中针对 RGB-D 相机进行优化,在未来工作中可以考虑加入 IMU、激光雷达等多种传感器进一步增强算法鲁棒性,并扩展到室外场景。

参考文献

[1] MUR-ARTAL R, TARDÓS 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.

[2] CAMPOS C, ELVIRA R, GÓMEZ RODRÍGUEZ 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.

[3] ENGEL J, SCHOPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM [C]//13th European Conference on Computer Vision, 2014: 834-849.

[4] ZUBIZARRETA J, AGUINAGA I, MONTIEL J M M. Direct sparse mapping [J]. IEEE Transactions on Robotics, 2020, 36(4): 1363-1370.

[5] 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.

[6] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision, 2017: 2980-2988.

[7] YU C, LIU Z X, LIU X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), 2018: 1168-1174.

[8] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12): 2481-2495.

[9] LIU Y B, MIURA J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods [J]. IEEE Access, 2021, 9: 23772-23785.

[10] CHENG S H, SUN C H, ZHANG S J, et al. SG-SLAM: A real-time RGB-D visual SLAM toward dynamic scenes with semantic and geometric information[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12.

[11] 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.

[12] 陈广秋, 任天蓉, 段锦, 等. 基于双分支多尺度特征融合的跨模态语义分割算法[J]. 电子测量与仪器学报, 2025, 39(5): 144-154.

CHEN G Q, REN T R, DUAN J, et al. Cross-modal semantic segmentation algorithm based on dual-branch multi-scale feature fusion[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 144-154.

[13] 李天放, 孙一宸, 于明鑫, 等. 结合语义分割与跨模态差分特征补偿的红外与可见光图像融合方法[J]. 电子测量与仪器学报, 2024, 38(7): 34-45.

LI T F, SUN Y CH, YU M X, et al. A fusion method for infrared and visible light images combining semantic segmentation with cross-modal differential

feature compensation [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (7): 34-45.

[14] KERBL B, KOPANAS G, LEIMKUHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4): 1-14.

[15] MATSUKI H, MURAI R, KELLY P H J, et al. Gaussian splatting SLAM [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 18039-18048.

[16] GUO X L, ZHANG W D, LIU R N, et al. Motiongs: Compact Gaussian splatting slam by motion filter[C]// 2024 7th International Conference on Robotics, Control and Automation Engineering(RCAE), 2024: 685-692.

[17] HUANG H J, LI L W, CHENG H, et al. Photo-SLAM: Real-time simultaneous localization and photorealistic mapping for monocular stereo and rgb-d cameras[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 21584-21593.

[18] XIE EN Z, WANG W H, YU ZH D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. ArXiv preprint arXiv:2105.15203, 2021.

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.

[20] TOLSTIKHIN I O, HOULSBY N, KOLESNIKOV A, et al. MLP-Mixer: An all-MLP architecture for vision[J]. 35th International Conference on Neural Information Processing Systems, 2021, 34: 24261-24272.

[21] 侯余鑫, 介婧, 侯北平, 等. 动态场景下基于稀疏光流的语义视觉 SLAM 算法[J]. 电子测量技术, 2025,

48(16):60-69.

HOU Y X, JIE J, HOU B P, et al. Semantic visual SLAM algorithm based on sparse optical flow in dynamic scenes[J]. Electronic Measurement Technology, 2025, 48(16):60-69.

[22] 王爽, 刘云平, 张柄棋, 等. 基于全景分割与多视图几何的动态 SLAM 方法[J]. 电子测量技术, 2024, 47(24):149-159.

WANG SH, LIU Y P, ZHANG B Q, et al. A dynamic SLAM method based on panoramic segmentation and multi-view geometry[J]. Electronic Measurement Technology, 2024, 47(24):149-159.

[23] 陈久朋, 陈治帆, 伞红军, 等. 基于稠密点云的神经辐射场 NeRF 在视觉 SLAM 建图任务中的应用研究[J]. 仪器仪表学报, 2024, 45(7):110-120.

CHEN J P, CHEN ZH F, SAN H J, et al. Application research of neural radiation field NeRF based on dense point clouds in visual SLAM mapping tasks[J]. Chinese Journal of Scientific Instrument, 2024, 45(7):110-120.

作者简介

符强, 硕士, 正高级实验师, 硕士生导师, 主要研究方向为图像处理、卫星导航和定位。
E-mail: 2416686947@qq.com

钟振, 硕士研究生, 主要研究方向为视觉同时定位与建图。
E-mail: zhongzhen7513@163.com

纪元法, 博士, 教授, 博士生导师, 主要研究方向卫星导航。
E-mail: a1827870@163.com

任风华(通信作者), 硕士, 副教授, 硕士生导师, 主要研究方向为图像处理。
E-mail: 643749271@qq.com