

DOI:10.19651/j.cnki.emt.2519407

基于3D高斯溅射的复杂室内环境 SNGO-SLAM 算法*

姜俊超^{1,2} 王永兰^{1,2} 房建东^{1,2} 朱瑾^{1,2}

(1. 内蒙古工业大学信息工程学院 呼和浩特 010080; 2. 内蒙古工业大学内蒙古感知技术与智能系统重点实验室 呼和浩特 010080)

摘要:近年来,3D高斯溅射技术在同步定位与建图系统中的应用使得利用显式三维高斯模型进行高质量图像渲染成为可能,显著提升了环境重建的保真度。然而,现有的基于3DGS的方法在复杂室内环境的三维重建中存在跟踪精度有限、缺乏全局一致性问题。为此,提出了一种基于3D高斯溅射的密集SLAM算法——SNGO-SLAM。该算法结合帧到模型和帧到帧两种跟踪方法的优点,利用表面法线感知获得更丰富的几何信息,显著提升了跟踪精度。为了解决随时间推移产生的跟踪误差,算法设计了环路闭合过程,并优化了3D高斯点表示问题,进一步提高了跟踪精度。此外,该算法还引入了双重高斯修剪策略,优化了内存使用,确保了精确的相机跟踪。在Replica、ScanNet和TUM RGBD数据集上的实验表明,该算法在保持高渲染质量的同时,在Replica数据集上的绝对轨迹均方根误差达到了0.27 cm,与NICE-SLAM、Vox-Fusion、Gaussian-SLAM和SplaTAM相比,跟踪精度分别提高了74.53%、91.26%、12.90%和28.95%,为SLAM技术提供了新的思路。

关键词:密集SLAM;3D高斯溅射;三维重建;环路闭合;复杂室内环境

中图分类号: TP249; TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

SNGO-SLAM algorithm for complex indoor environments
based on 3D Gaussian splattingJiang Junchao^{1,2} Wang Yonglan^{1,2} Fang Jiandong^{1,2} Zhu Jin^{1,2}

(1. School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010080, China;

2. Inner Mongolia Key Laboratory of Perception Technology and Intelligent System, Inner Mongolia University of Technology, Hohhot 010080, China)

Abstract: In recent years, the application of 3D Gaussian splatting technology in simultaneous localization and mapping systems has made it possible to perform high-quality image rendering using explicit 3D Gaussian models, significantly improving the fidelity of environmental reconstruction. However, the existing methods based on 3DGS have problems such as limited tracking accuracy and lack of global consistency in the 3D reconstruction of complex indoor environments. For this purpose, this paper proposes a dense SLAM algorithm based on 3D Gaussian splatting—SNGO-SLAM. This algorithm combines the advantages of both frame-to-model and frame-to-frame tracking methods, and uses surface normal perception to obtain richer geometric information, significantly improving the tracking accuracy. To address the tracking error that occurs over time, the algorithm introduces a loop closure process and optimizes the 3D Gaussian point representation problem, further enhancing the tracking accuracy. In addition, this algorithm also introduces a dual Gaussian pruning strategy, optimizing memory usage and ensuring precise camera tracking. Experiments on the Replica, ScanNet and TUM RGBD datasets show that while maintaining high rendering quality, the absolute root mean square error of the trajectory of this algorithm on the Replica dataset reaches 0.27 cm. Compared with NICE SLAM, Vox-Fusion, Gaussian SLAM and SplaTAM, the tracking accuracy has increased by 74.53%, 91.26%, 12.90% and 28.95% respectively, providing new ideas for SLAM technology.

Keywords: dense SLAM; 3D Gaussian splatting; 3D reconstruction; loop closure; complex indoor environments

0 引言

视觉同步定位与建图(simultaneous localization and

mapping, SLAM)是一项在未知环境中同时进行三维地图重建与相机姿态估计的核心技术,已被广泛应用于机器人、自动驾驶以及增强/虚拟现实(AR/VR)等多个领域^[1]。其

收稿日期:2025-07-18

* 基金项目:内蒙古自治区科技计划项目(2023YFJM0002, 2025KYPT0088)、内蒙古自治区直属高校基本科研业务费基金项目(JY20240076)资助

中,稠密地图构建是视觉 SLAM 系统中的重要模块。一方面,它使机器人能够更全面、细致地感知周围环境;另一方面,也为抓取、操控与人机交互等下游任务提供了基础性的环境信息支持。

在实时密集视觉 SLAM 的研究中,基于 RGB-D 相机的方法已取得显著进展。例如,传统的 ORB-SLAM3^[2] 方法通过稠密跟踪和地图构建实现了高效的三维地图生成。然而,传统的密集视觉 SLAM 仅依赖点云进行场景重建,无法实现高保真的环境重建。为此,BAD SLAM^[3] 结合了稠密信息和光束平差法(bundle adjustment,BA)^[4] 优化,尽管在大规模场景中提高了细节保留能力,但其计算复杂度较高,无法满足资源受限环境中的实时需求。

近年来,神经辐射场(neural radiance field,NeRF)^[5] 作为一种新兴的隐式表示方法,因其能够密集且连续地表示场景的显著能力而受到广泛关注。在基于 NeRF 的 SLAM 研究中,各种神经场方法为解决当前 SLAM 面临的挑战提供了可行的方案,极大地推动了神经 SLAM 的发展。例如,Sucar 等^[6] 首次尝试采用多层感知机(multi-layer perceptron,MLP)动态生成特定场景的隐式 3D 模型,展示了多层感知机在动态创建特定场景隐式 3D 模型方面的巨大潜力。然而,模型容量的限制导致了其细节重建的不足,并面临着灾难性遗忘的问题。Zhu 等^[7] 通过使用体素网格存储神经特征,并结合分层策略和几何先验,显著提高了优化过程的精确度和效率。然而,其预测能力仍局限于粗略的表示尺度,且未能处理回环检测问题。

Yang 等^[8] 将传统的体积融合方法与神经隐式表示相结合,采用基于体素的神经隐式表面表示对每个体素内的场景进行编码和优化,并引入基于八叉树的结构,以实现动态体素分配策略。Johari 等^[9] 通过多尺度轴对齐特征平面,并采用二次缩放优化了内存使用,使用截断符号距离函数(truncated signed distance function,TSDF)作为几何表示,显著提高了收敛速度和重建质量。Liso 等^[10] 采用子地图形式的神经点云进行局部建图与跟踪,有效解决了相机跟踪中的误差累积问题,并避免了重叠区域中的可见接缝。Sandström 等^[11] 引入了一种动态神经点云表示,根据输入数据的信息调整点的密度,动态的对点进行区域分配。

尽管基于 NeRF 的 SLAM 方法在精度和效率方面取得了一定进展,但仍面临诸如过度平滑、易受灾难性遗忘影响等挑战^[12]。此外,神经地图本身不支持刚体变换,这大大限制了地图校正的效率。最近,3D 高斯溅射(3D gaussian splatting,3DGS)^[13] 作为一种有效的场景表示方法出现,采用一组高斯点来表示场景。与依赖光线行进的神神经隐式表示方法不同,3D 高斯溅射使用 3D 高斯点作为基本元素,以显式形式表示空间^[14]。通过添加更多的高斯基元、充分利用每个像素的密集光度损失和直接参数梯度流,3D 高斯溅射显著提高了地图容量的灵活性,从而促进了快速优化^[15]。这使得 SLAM 系统能够受益于显式的体

积表示、快速渲染、丰富的优化机制、直接的梯度流、扩展的地图容量以及显式的空间范围控制。为了将这些优势应用于密集视觉 SLAM,已有多种方法提出利用 3D 高斯溅射进行显示空间表示。

例如,Keetha 等^[16] 将场景表示为一组简化的 3D 高斯分布,从而实现了高质量的彩色和深度图像渲染。Yugay 等^[17] 通过将场景组织为独立优化的子地图,能够高效地对新探索的区域添加新的高斯分布,并在线优化它们,从而实现对更大场景的扩展。此外,Yugay 等^[18] 还利用 3D 高斯作为支持刚体变换的场景表示,通过高效的建图优化和融合策略,减少了所需的磁盘存储和处理时间,并支持任意数量的同时操作代理进行一致的 3D 重建。Xu 等^[1] 通过创新的运动掩模生成算法,将时空一致的深度掩模与语义先验相结合,从而显著提高了运动物体分割的精度。Zhu 等^[19] 提出了一种在基于 3D 高斯溅射的 SLAM 管道中进行环路闭合的方法,该方法能够在无需任何训练或预处理的情况下,直接对 SLAM 前端的相机姿态估计进行操作。

尽管 3DGS 凭借其高效的渲染、显式表示和丰富的优化能力成为了一个强有力的替代方案,但现有方法仍存在一些局限性。这些问题包括在大规模场景或扩展的视频序列中容易出现灾难性遗忘,无法精确编码细致的几何结构,且性能受限于模型容量和计算资源。为解决上述问题,本文提出了一种高效的 3DGS-SLAM 算法——SNGO-SLAM,本文的主要贡献为:

1) 提出了一种混合表面法线跟踪方法,结合帧到模型和帧到帧两种跟踪方法的优点,通过表面法线感知获取更丰富的几何信息,以辅助进行最小化跟踪损失,避免优化过程陷入局部最优解,提高相机的跟踪精度。

2) 采用曝光调整方法修正关键帧间因曝光度差异引起的误差,并通过法线一致性损失约束 3D 高斯的表面法线,在提高渲染质量的同时保持高精度的空间细节。

3) 引入了一种双重高斯点修剪策略,通过通过不透明度、最大尺度和椭球形状进行初筛,利用滑动窗口检测关键帧上的观测值进行二次筛选,确保了几何完整性,并有助于精确的相机跟踪。

4) 设计了一个环路闭合,并引入了高斯初始几何重叠比对环路进一步筛选,避免了闭环的误检、漏检问题,确保精确的轨迹估计并维持连贯的 3D 地图表示。

5) 提出了一种刚性高斯点云配准方法,通过结合利用刚性变换进行的高斯配准和点云配准,解决了 3D 高斯中心点无法准确表示场景表面的问题。实验结果表明,SNGO-SLAM 在多个公开数据集上的表现优于现有经典方法,在重建质量、跟踪精度和运行效率上均取得了显著提升。

1 SNGO-SLAM 总体设计及工作原理

1.1 子地图构建

现有的方法通常表现出灾难性遗忘的倾向,虽然一些

3D 高斯风格的显式表示通过稀疏帧采样来缓解灾难性遗忘,但这同样导致内存需求增加和处理速度变慢,尤其在复杂室内场景中。近期,一些策略^[10,17]建议将环境划分为子图,并将局部 SLAM 任务分配给不同的代理。实践证明,这种子地图构建策略在解决灾难性遗忘问题上无疑是一种非常有效的手段。

因此,本文提出一种 SNGO-SLAM 算法,如图 1 所示。该算法基于 Yugay 等^[17]提出的子地图构建的视觉 SLAM 框架(虚线部分)进行改进(实线部分),主要包含以下步骤:首先,对多视角二维图像序列进行处理,采用运动恢复结构(structure from motion, SfM)技术,得到初始的稀疏点云以及相机位姿参数;其次,针对恒速运动假设难以准确推断初始优化姿态值,容易导致优化过程陷入局部最优解的问题,提出混合表面法线跟踪方法,并通过表面法线感知获取

更丰富的几何信息进行最小化跟踪损失,提升跟踪精度;然后,在子地图初始化后,针对不同关键帧之间产生的曝光度差异问题,引入曝光调整方法进行修正,并利用法线一致性损失约束 3D 高斯的表面法线,解决 3DGS 无法精确表征细节几何结构的问题。针对渲染图像中存在的伪影现象和高斯点冗余问题,引入双重高斯点修剪策略对高斯点进行更高效的修剪。再然后,针对缺乏全局优化技术,导致随着时间的推移会产生跟踪误差的问题,设计了一个环路闭合用于相机姿态和 3D 重建的实时全局优化,引入高斯初始几何重叠比提高环路检测的准确性,解决单纯依赖视觉相似性容易导致错误的环路边生成的问题,并提出刚性高斯点云配准方法,解决 3D 高斯中心点无法准确表示场景表面的问题。最后,将所有子地图合并成一个全局地图,并进行最终优化,从而实现高质量的室内场景重建。

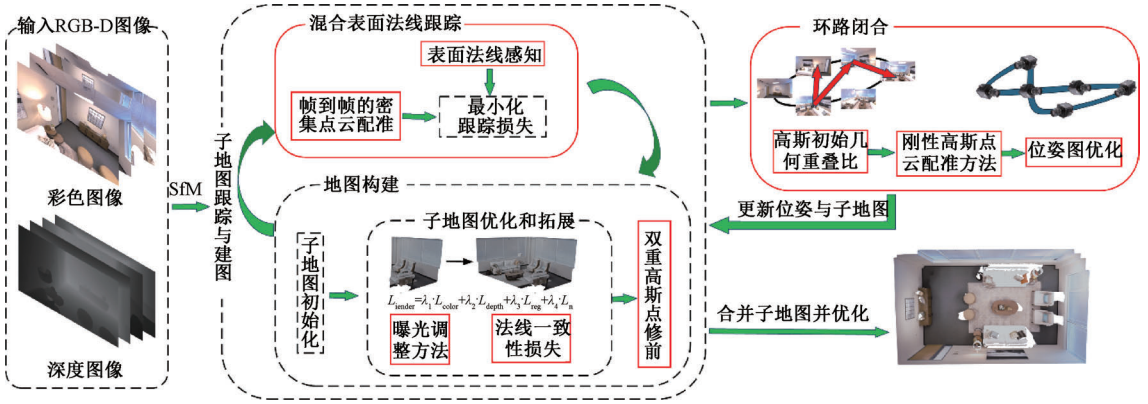


图 1 SNGO-SLAM 算法框架

Fig. 1 SNGO-SLAM algorithm framework

一个单独的 3D 高斯分布由以下参数构成:均值 $\boldsymbol{\mu} \in \mathbb{R}^3$ 、不透明度 $o \in \mathbb{R}$ 、协方差矩阵 $\boldsymbol{\Sigma} \in \mathbb{R}^{3 \times 3}$ 和 RGB 颜色 $\mathbf{C} \in \mathbb{R}^3$ 。本文使用 N 个单独的 3D 高斯分布来构成每个子地图的高斯点云 P^s :

$$P^s = \{G_i(\boldsymbol{\mu}_i^s, \boldsymbol{\Sigma}_i^s, o_i^s, \mathbf{C}_i^s) \mid i = 1, \dots, N\} \quad (1)$$

1.2 子地图优化

遵循 Gaussian-SLAM^[17]的子地图优化方法,通过选取固定的关键帧将新高斯分布添加至子地图后,在活跃子地图对应的所有关键帧上计算渲染损失,并通过最小化渲染损失对所有高斯分布进行联合优化。

由于相机拍摄视角的不同可能引起光照的反射和折射变化,从而导致不同关键帧之间曝光度的差异。这种曝光度的变化将直接影响 3D 高斯分布中渲染颜色 \mathbf{C} 的表现,进而影响子地图的渲染质量。

为解决上述问题,本文在计算颜色监督时引入了一种曝光调整方法。该方法根据所选视点的曝光系数对图像的亮度和对比度进行调整,并通过加权组合 L_1 损失和 SSIM 损失来实现颜色监督的计算:

$$L_{\text{color}} = (1 - \lambda_1) \times \|\hat{\mathbf{I}} \times \exp(\gamma) + \delta - \mathbf{I}\|_1 + \lambda_1 (1 - \text{SSIM}(\hat{\mathbf{I}} \times \exp(\gamma) + \delta, \mathbf{I})) \quad (2)$$

其中, $\lambda_1 \in [0, 1]$ 是损失函数权重值, γ 是对比度调整因子, δ 是亮度偏移量, $\hat{\mathbf{I}}$ 和 \mathbf{I} 分别是渲染后图像和原始图像。深度监督的计算公式为:

$$L_{\text{depth}} = \|\hat{\mathbf{D}} - \mathbf{D}\|_1 \quad (3)$$

其中, $\hat{\mathbf{D}}$ 和 \mathbf{D} 分别是重建深度图和真实深度图。为解决 3D 高斯风格 SLAM 方法中普遍存在的无法精确编码细节几何结构的问题,本文引入法线一致性损失,以约束 3D 高斯分布的法线方向:

$$L_n = \sum_i \omega_i (1 - \mathbf{n}_i^\top \mathbf{N}_{\text{GT}}) \quad (4)$$

$$\mathbf{N}_{\text{GT}} = \frac{\nabla_x \mathbf{x}_{\text{GT}} \times \nabla_y \mathbf{x}_{\text{GT}}}{\|\nabla_x \mathbf{x}_{\text{GT}} \times \nabla_y \mathbf{x}_{\text{GT}}\|} \quad (5)$$

其中, i 是沿着光线相交的点上的索引, ω_i 是第 i 个交点的权重, \mathbf{n} 是相对于相机的法线, \mathbf{N}_{GT} 是从邻近点取有限差分得到的真实法线, ∇_x 和 ∇_y 分别是 x 轴和 y 轴的梯度。本文遵循 Gaussian-SLAM^[17]中使用的各向同性正则

化项,以解决在稀疏覆盖区域中3D高斯分布可能存在的过度延伸问题:

$$L_{\text{reg}} = \frac{1}{|K|} \sum_{k \in K} |s_k - \bar{s}_k| \quad (6)$$

其中, $s_k \in \mathbb{R}^3$ 是子地图 K 的3D高斯分布尺度, \bar{s}_k 是平均尺度, $|K|$ 是高斯数量。最后,总渲染损失的计算公式通过联合优化4个损失函数得到:

$$L_{\text{render}} = \lambda_1 \cdot L_{\text{color}} + \lambda_2 \cdot L_{\text{depth}} + \lambda_3 \cdot L_{\text{reg}} + \lambda_4 \cdot L_n \quad (7)$$

以往的3D高斯风格SLAM方法^[17,20]通过删除一定范围内不透明度值低于指定阈值的高斯点来进行高斯点的修剪,以改善重建效果。然而,这些方法由于仅将不透明度值作为修剪高斯点的唯一标准,在减少渲染图像中的伪影数量和处理高斯点冗余问题方面效果有限。

为此,本文引入了一种双重高斯修剪方法以替代原有的修剪方法。具体而言,双重高斯修剪方法综合考虑不透明度值、最大尺度和高斯椭球的形状3个标准进行第一重优化,删除所有不满足这些标准的高斯点。其数学表示为:

$$\begin{cases} \alpha < \tau_{s1} \\ \max(S) > \tau_{s2} \\ \max(S)/\min(S) > \tau_{s3} \end{cases} \quad (8)$$

其中, α 是高斯点的不透明度值, $\max(S)$ 和 $\min(S)$ 分别是3D高斯椭球的最大和最小尺度, τ_{s1} , τ_{s2} 和 τ_{s3} 分别是不透明度值、最大尺度和高斯椭球形状的标准参数。第二重优化如图2所示,图中圆形虚线部分代表被删除的高斯点。通过第一重优化的高斯点被投影到滑动窗口中的关键帧上,并将这些关键帧上高斯点的观测值与观测阈值 τ_{s1} 进行对比,从当前高斯图中删除观测值低于阈值的高斯点。

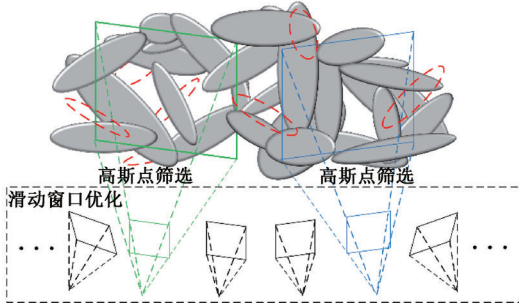


图2 第二重高斯点优化方法

Fig. 2 The second Gaussian point optimization method

1.3 轨迹跟踪

以往的3D高斯风格SLAM方法在跟踪过程中通常采用帧到模型的策略,通过计算相机姿态的梯度^[20]或分析得到近似结果^[21-22]来估计相机姿态。在初始化相机姿态时,大多数高斯SLAM方法假设恒定速度模型。然而,恒定运动假设难以准确推断出合理的初始优化姿态值,不准

确的初始姿态容易导致优化过程陷入局部最优解。相比之下,NeRF风格的SLAM方法采用帧到帧的跟踪方法,利用外部跟踪器如ORB-SLAM3^[23-24]、DROID^[25-26]或ICP配准^[27-28]来估计相机位姿。这种帧到帧的跟踪方法虽然在跟踪精度和稳定性方面不如帧到模型的跟踪方法,但该方法具有较低的计算复杂度,对计算资源和存储需求较少,展现出极高的适应性。

本文提出了一种混合表面法线跟踪方法,该方法结合了帧到帧跟踪和帧到模型跟踪方法的优点,通过表面法线感知方法获取更丰富的几何信息,以提升复杂室内环境下的跟踪精度。具体来说,采用帧到帧的密集点云配准方法进行相机姿态初始化,然后通过结合表面法线感知的帧到模型的方法,进一步最小化跟踪损失,以优化初始相机位姿。

在相机姿态初始化阶段,通过迭代方式在每个尺度上对源点云 P_i 与目标点云 P_{i-1} 进行粗到细的密集点云配准,以此估计相机姿态 $\mathbf{T}_{i-1,i} \in \text{SE}(3)$ 。损失函数通过高斯-牛顿法进行优化,直至收敛。每个尺度的损失函数的计算方式为:

$$E(\mathbf{T}_{i-1,i}) = (1 - \sigma) \sum_{\langle p, q \rangle \in K} (r_C^{\langle p, q \rangle}(\mathbf{T}_{i-1,i}))^2 + \sigma \sum_{\langle p, q \rangle \in K} (r_G^{\langle p, q \rangle}(\mathbf{T}_{i-1,i}))^2 \quad (9)$$

$$r_C^{\langle p, q \rangle} = C_p(f_p(\mathbf{T}_{i-1,i}(\mathbf{q})) - \mathbf{p}) - C(\mathbf{q}) \quad (10)$$

$$r_G^{\langle p, q \rangle} = ((\mathbf{T}_{i-1,i}(\mathbf{q}) - \mathbf{p}) \cdot \mathbf{n}_p)^2 \quad (11)$$

其中, $\sigma \in [0, 1]$ 是损失函数权重值, $K = \{\langle p, q \rangle\}$ 是通过ICP计算的 P_i 和 P_{i-1} 之间的对应集合, $r_C^{\langle p, q \rangle}$ 是颜色残差, $r_G^{\langle p, q \rangle}$ 是几何残差, f_p 函数将点投影到 \mathbf{p} 切平面, C_p 是在点云 P_i 上的连续强度函数, $C(\mathbf{q})$ 是点 \mathbf{q} 的强度值, \mathbf{n}_p 是点 \mathbf{p} 的法向量。

在最小化跟踪损失阶段,采用从3D高斯渲染的alpha图的多项式 M_{alpha} 和误差布尔掩码 M_{inlier} 去除颜色和深度误差大于帧相对误差阈值的像素。令 \mathbf{x} 为从真实世界点投影得到的点, \mathbf{x}_{GT} 为从真实深度获取的点。随后,计算投影点与真实点之间的距离 $d_{\text{point2point}}(\mathbf{x})$ 以及投影点到平面之间的距离 $d_{\text{point2plane}}(\mathbf{x})$:

$$d_{\text{point2point}}(\mathbf{x}) = |z(\mathbf{x}_{\text{GT}}) - z(\mathbf{x})|_1 = |D - \hat{D}|_1 \quad (12)$$

$$d_{\text{point2plane}}(\mathbf{x}) = |(\mathbf{x}_{\text{GT}} - \mathbf{x})^T \mathbf{N}_{\text{GT}}|_1 \quad (13)$$

其中, D 和 \hat{D} 分别为真实深度和渲染深度。最后,对所有有效像素进行求和得到最终的跟踪损失:

$$L_{\text{tracking}} = \sum M_{\text{alpha}} \cdot M_{\text{inlier}} \cdot (\lambda_c \cdot |\hat{I} - I|_1 + (1 - \lambda_c) d_{\text{point2point}}(\mathbf{x}) + d_{\text{point2plane}}(\mathbf{x})) \quad (14)$$

其中,损失函数权重 $\lambda_c \in [0, 1]$ 确保了优化过程主要由那些累积的alpha值接近1且渲染质量高的重建区域来引导。

1.4 环路闭合

在环路检测中,单纯依赖视觉相似性可能导致错误的

环路边生成,从而降低全局姿态优化的性能。

为了解决这一问题,本文引入了高斯初始几何重叠比来进一步筛选环路。具体而言,在使用预训练的 NetVLAD^[29]提取全局描述符后,通过计算包含在两个子地图内的所有关键帧之间的余弦相似度,确定对应于地图的第 a 百分位的自相似度分数和交叉相似度分数。若交叉相似度分数大于两个自相似度分数中的最小值,并且两个子地图之间的高斯初始几何重叠比 $r > 0.2$,则认为检测到一个新的闭环。

新的姿态图在检测到新的环路时创建,并通过环路边和初始化为单位矩阵的里程计边将所有子地图的相对姿态校正关联。然而,由于子地图具有动态分辨率,导致不同区域的点云采样密度存在差异,因此无法保证点云的均匀分布。此外,3D 高斯中心点可能无法准确地表示场景表面。这些问题使得直接通过点云配准来初始化环路边的方法难以实现。

为解决上述问题,本文提出一种刚性高斯点云配准方法,采用自粗到细的逐步策略。首先将子地图及其视点视为一个刚体并将高斯配准问题转化成关键帧定位问题,通过刚性变换进行高斯配准,然后通过点云配准进行进一步优化。具体来说,在粗配准阶段,首先并行估计子地图 P 和子地图 Q 之间的两个视点 v_i^p, v_i^q 的刚性变换:

$$\begin{cases} \mathbf{T}_{P \rightarrow Q} = \frac{\mathbf{T}_i^q}{\mathbf{T}_i^p} \\ \mathbf{T}_{Q \rightarrow P} = \frac{\mathbf{T}_i^p}{\mathbf{T}_i^q} \end{cases} \quad (15)$$

其中, \mathbf{T}_i^q 是视点 v_i^q 在子地图 Q 中的相机姿态。然后,在保持 Q 参数不变的同时使用 L_1 损失 L_{color} 和 L_{depth} 优化刚性变换 $\mathbf{T}_{P \rightarrow Q}$, 得到一组变换 $\{(\mathbf{T}_{P \rightarrow Q}, \epsilon)_i\}_{i=1}^{2k}$ 。其中, ϵ 是优化后的渲染残差, k 是在环路检测进程中计算余弦相似度时保留的前 k 对关键帧。最后,将渲染残差 ϵ 的倒数作为每个估计的权重并通过对各个估计的加权平均计算全局旋转:

$$\bar{\mathbf{R}} = \underset{\mathbf{R} \in \text{SO}(3)}{\text{argmin}} \sum_{i=1}^k \frac{1}{\epsilon_i} \|\mathbf{R} - \mathbf{R}_i\|_F + \sum_{i=k+1}^{2k} \frac{1}{\epsilon_i} \|\mathbf{R} - \mathbf{R}_i^{-1}\|_F \quad (16)$$

其中, $\|\cdot\|_F$ 表示 Frobenius 范数。在细配准阶段,通过 ICP 算法对结果进行进一步优化,并使用约束平移量的方法对环边进行筛选。具体来说,通过选择一个合适的百分位数,确保剩余环边的标准差低于设定的阈值 σ_{\min} , 并根据所有环边约束的统计信息计算阈值 t_{\min} 。随后,删除平移量大于动态计算阈值 t_{\min} 的环边。

通过省略球谐函数(SH),能有效的减少高斯地图的大小和计算复杂度,从而提高姿态估计的精度^[30]。本文直接通过姿态校正信息对每个子地图的相机姿态、高斯均值和协方差进行了更新:

$$\begin{cases} \mathbf{T}_j \leftarrow \mathbf{T}_{c_i} \cdot \mathbf{T}_j \\ \boldsymbol{\mu}_i \leftarrow \mathbf{R}_{c_i} \cdot \boldsymbol{\mu}_{S_i} + \mathbf{t}_{c_i} \\ \boldsymbol{\Sigma}_i \leftarrow \mathbf{R}_{c_i} \cdot \boldsymbol{\Sigma}_{S_i} \cdot \mathbf{R}_{c_i}^T \end{cases} \quad (17)$$

其中, \mathbf{T}_{c_i} 是从全局姿态优化的输出中获得姿态校正信息, \mathbf{R}_{c_i} 是第 i 个子地图的旋转矩阵修正, \mathbf{t}_{c_i} 是第 i 个子地图的平移向量修正,索引 j 用于遍历子地图中的关键帧, $\boldsymbol{\mu}_i$ 是子地图 S_i 中高斯分布的均值, $\boldsymbol{\Sigma}_i$ 是子地图 S_i 中协方差矩阵的集合。

2 实验结果及分析

2.1 实验设置

数据集:为了验证本文提出方法的有效性,本文在合成数据集和真实场景数据集上进行了评估。Replica 数据集^[31]是一个包含 8 个高保真室内场景的合成数据集,提供了毫米级精度的 3D 网格和材质真值,能够在静态环境下有效区分算法的几何重建能力和位姿估计误差,具有较高的精确度。ScanNet 数据集^[32]包含 6 个大规模室内场景,如办公室和公寓等,利用 Bundle-Fusion^[33]提供的相机姿态,该数据集评估了算法在复杂环境布局中的重建一致性和在线学习效率,特别是在优化关键帧数量方面的表现。TUM-RGBD 数据集提供了由 Microsoft Kinect 传感器捕获的颜色和深度图像,分辨率为 $640 \text{ pixel} \times 480 \text{ pixel}$,帧率为 30 Hz。该数据集采用运动捕获系统获取的地面真实轨迹,主要针对静态环境中的挑战,如弱纹理、遮挡和传感器噪声等,对算法的性能进行了全面评估,考察其在复杂环境下的鲁棒性。

评估指标:实验部分的性能指标评估采用了多种方法来全面衡量本文算法的表现。为了评估跟踪精度,本文使用了绝对轨迹均方根误差(ATE RMSE),该指标衡量的是估计轨迹与真实轨迹之间的平均平方误差,能够有效反映算法的跟踪精度。为了评估渲染质量,本文通过比较全分辨率渲染图像与输入训练视图,采用了峰值信噪比(PSNR)、结构相似性(SSIM)和感知图像对比度损失(LPIPS)作为标准。PSNR 用于评估图像的重建质量,数值越高表示图像质量越好;SSIM 用于衡量两幅图像之间的相似度,数值越高表示结构相似度越高;LPIPS 用于评估感知质量,数值越低表示感知质量越高。在 Replica 数据集上,本文通过 L1 深度误差(Depth L1)以及 F1 分数来评估重建质量。L1 深度误差衡量估计深度图与真实深度图之间的平均绝对误差,数值越低表示重建质量越好;F1 分数通过综合精确率和召回率来评估重建的几何精度和完整性。此外,本文还评估了平均每帧跟踪和地图优化的运行时间,具体计算方法是通过总优化时间除以序列长度,从而得到每帧的平均处理时间。通过上述多种评估指标,本文算法在跟踪精度、渲染质量和重建质量方面的性能得到了全面的验证。

基线:本文选择了近年来几种基于 NeRF 和 3DGS 的代表性算法进行比较,包括以下几种:基于 NeRF 的经典算法 NICE-SLAM^[7],使用实时全局优化的 NeRF 算法 GO-SLAM^[34],采用体素网格表示三维空间以增强几何结构的 NeRF 算法 Vox-Fusion^[8],使用子地图建图的 3DGS 算法 Gaussian-SLAM^[17],设计了高效回环检测和闭环优化机制的 3DGS 算法 SplaTAM^[16]。所有对比方法在复现时的参数配置均参考原始论文中给出的最佳配置。

实施细节:本次实验平台配置为 5.60 GHz Intel Core i7-14700KF CPU,显卡为 NVIDIA RTX 3090,操作系统为 Ubuntu22.04。在 Replica 数据集中,本文将新的子地图点数 M_u 设置为 600 000,子地图第 1 帧的建图迭代次数设置为 1 000,跟踪损失中的参数 λ_c 设置为 0.95。在 TUM-RGBD 和 ScanNet 数据集中,新的子地图点数设置为 100 000,子地图第 1 帧的建图迭代次数设置为 100, λ_c 设置为 0.6。在所有数据集中,新的子地图梯度点数 M_c 设置为 50 000,后续关键帧的迭代次数设置为 100。在子地图构建时,位移阈值 d_{thre} 设置为 0.5 m,旋转阈值 θ_{thre} 设置为 50° 。在子地图优化时,建图损失函数中的权重 λ_1 、 λ_2 和 λ_3

均设置为 1, λ_4 设置为 0.05,双重高斯修剪中的标准 τ_{s1} 、 τ_{s2} 和 τ_{s3} 分别设置为 0.005、0.4 和 36。在环路检测时,确定自相似性分数的子地图百分点 a 设置为 50%。

2.2 跟踪性能评估

本文在 3 个不同的数据集上评估了相机跟踪精度。表 1 展示了在 Replica 数据集上的相机跟踪精度,结果表明,本文方法在 8 个场景中的表现均优于其他经典 SLAM 方法,且相比于基于 NeRF 的 GO-SLAM 和同样基于 3DGS 的 Gaussian-SLAM,准确率分别提高了 22% 和 12%。表 2 和 3 展示了在真实场景数据集 ScanNet 和 TUM-RGBD 上的评估结果,表明本文算法优于大多数经典 SLAM 方法。尽管本文算法的跟踪性能略低于 GO-SLAM 算法,但它有效缩小了 3D 高斯溅射方法与神经网络方法之间的跟踪差距。同时,在重建精度和地图完整性方面,本文算法相较于 GO-SLAM 算法,L1 深度误差减少了 88.7%,F1 分数提高了 63.6%(表 4)。在网格重建结果对比中(图 3),本文算法在几何细节、重建的视觉质量和完整性方面均表现出明显优势,特别是在复杂室内环境中的应用,展现出更强的鲁棒性和适应性。

表 1 Replica 数据集上的 ATE RMSE 结果

Table 1 The ATE RMSE results on the Replica dataset

方法	room0	room1	room2	office0	office1	office2	office3	office4	平均
NICE-SLAM ^[7]	0.97	1.31	1.07	0.88	1.00	1.06	1.10	1.13	1.06
GO-SLAM ^[34]	0.34	0.29	0.29	0.32	0.30	0.39	0.39	0.46	0.35
Vox-Fusion ^[8]	1.37	4.70	1.47	8.48	2.04	2.58	1.11	2.94	3.09
Gaussian-SLAM ^[17]	0.29	0.29	0.22	0.37	0.23	0.41	0.30	0.35	0.31
SplaTAM ^[16]	0.31	0.40	0.29	0.47	0.27	0.29	0.32	0.72	0.38
MonoGS ^[30]	0.33	0.22	0.29	0.36	0.19	0.25	0.12	0.81	0.32
本文	0.28	0.24	0.19	0.31	0.17	0.43	0.20	0.32	0.27

表 2 ScanNet 数据集上的 ATE RMSE 结果

Table 2 The ATE RMSE results on the ScanNet dataset

方法	Scene0000	Scene0059	Scene0106	Scene0169	Scene0181	Scene0207	平均
NICE-SLAM ^[7]	12.0	14.0	7.9	10.9	13.4	6.2	10.7
GO-SLAM ^[34]	5.4	7.5	7.0	7.7	6.8	6.9	6.8
Vox-Fusion ^[8]	16.6	24.2	8.4	27.3	23.3	9.4	18.2
Gaussian-SLAM ^[17]	21.2	12.8	13.5	16.3	21.0	14.3	16.5
SplaTAM ^[16]	12.8	10.1	17.7	12.1	11.1	7.5	11.9
MonoGS ^[30]	9.8	32.1	8.9	10.7	21.8	7.9	15.2
本文	8.4	9.6	9.8	12.6	10.7	8.3	9.9

2.3 重建质量评估

Replica 数据集是唯一提供高精度地面真实网络的公开数据集,因而非常适合用于评估不同 SLAM 方法的网格重建质量。表 4 展示了在 Replica 数据集上的重建质量对比结果,显示本文算法在网格重建方面的表现优于所有经

典 SLAM 方法。与排名第 2 和第 3 的算法相比,本文方法的 L1 深度误差分别减少了 22% 和 26%,F1 分数分别提升了 1% 和 3.8%。

为更直观地展示重建质量的差异,图 3 对比了本文算法,基于 NeRF 的 GO-SLAM 以及基于 3DGS 的 Gaussian-

表 3 TUM-RGBD 数据集上的 ATE RMSE 结果

Table 3 The ATE RMSE results on the TUM-RGBD dataset

方法	cm			
	fr1/desk	fr2/xyz	fr3/office	平均
NICE-SLAM ^[7]	4.26	6.19	3.87	4.77
GO-SLAM ^[34]	1.50	0.60	1.30	1.13
Vox-Fusion ^[8]	3.52	1.49	26.01	10.34
Gaussian-SLAM ^[17]	2.73	1.39	5.31	3.14
SplaTAM ^[16]	3.35	1.24	5.16	7.14
本文	2.46	1.42	3.59	2.49

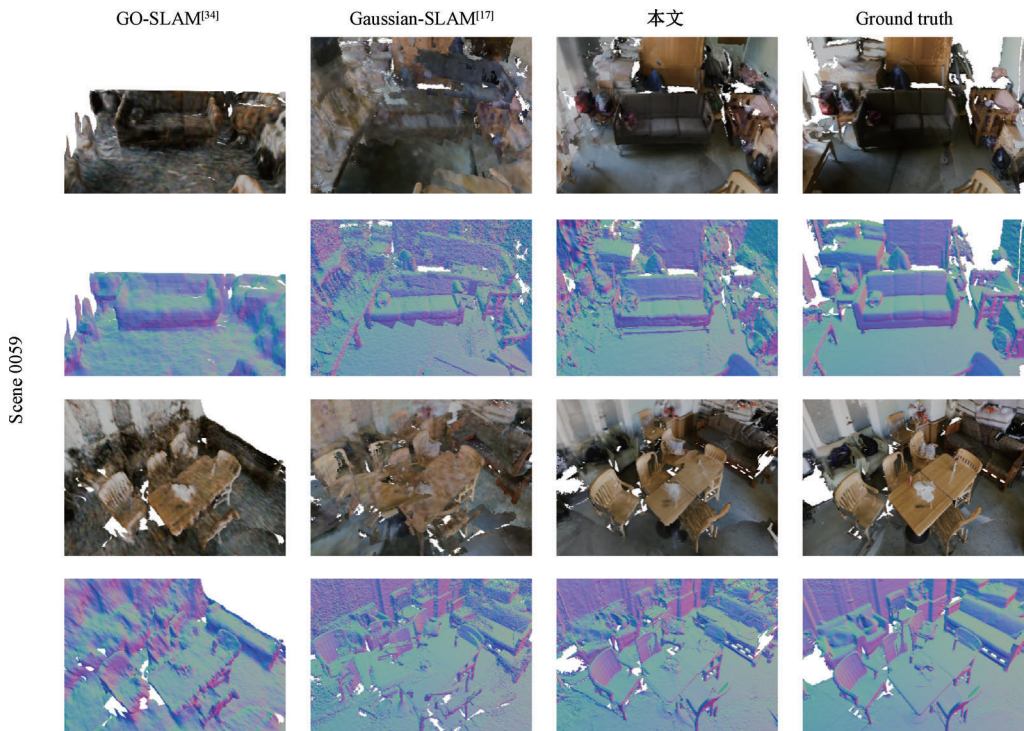
SLAM 在 ScanNet 数据集的 scene0059 和 scene0169 场景, 以及 Replica 数据集的 room1 场景中的网格重建效果。每个场景的上半部分图像采用彩色网格可视化, 用于展现视觉质量和场景完整性; 下半部分则采用法线贴图着色器渲染, 以突出几何形状的细节。

结果表明, 本文算法在各类结构 (如椅子、桌子等) 的场景重建与补全方面均表现出更高的准确性和质量。在法线图中, 本文算法在桌面、椅面及墙体等区域呈现出比其他 SLAM 方法更为清晰的几何细节, 进一步验证了其在几何重建精度方面的优势。

表 4 Replica 数据集上重建质量对比

Table 4 Comparison of reconstruction quality on the Replica dataset

方法	指标	room0	room1	room2	office0	office1	office2	office3	office4	平均
NICE-SLAM ^[7]	Depth L1/cm ↓	1.81	1.44	2.04	1.39	1.76	8.33	4.99	2.01	2.97
	F1/% ↑	45.0	44.8	43.6	50.0	51.9	39.2	39.9	36.5	43.9
GO-SLAM ^[34]	Depth L1/cm ↓	4.56	1.97	3.43	2.47	3.03	10.3	7.31	4.34	4.68
	F1/% ↑	17.3	33.4	24.0	43.0	31.8	21.8	17.3	22.0	26.3
Vox-Fusion ^[8]	Depth L1/cm ↓	1.09	1.90	2.21	2.32	3.40	4.19	2.96	1.61	2.46
	F1/% ↑	69.9	34.4	59.7	46.5	40.8	51.0	64.6	50.7	52.2
Gaussian-SLAM ^[17]	Depth L1/cm ↓	0.61	0.25	0.54	0.50	0.52	0.98	1.63	0.42	0.68
	F1/% ↑	88.8	91.4	90.5	91.7	90.1	87.3	84.2	87.4	88.9
SplaTAM ^[16]	Depth L1/cm ↓	0.43	0.38	0.54	0.44	0.66	1.05	1.60	0.68	0.72
	F1/% ↑	89.3	88.2	88.0	91.7	90.0	85.1	77.1	80.1	86.1
本文	Depth L1/cm ↓	0.41	0.24	0.55	0.33	0.53	0.64	1.13	0.44	0.53
	F1/% ↑	90.3	91.6	90.7	92.4	90.1	88.5	88.1	87.8	89.9



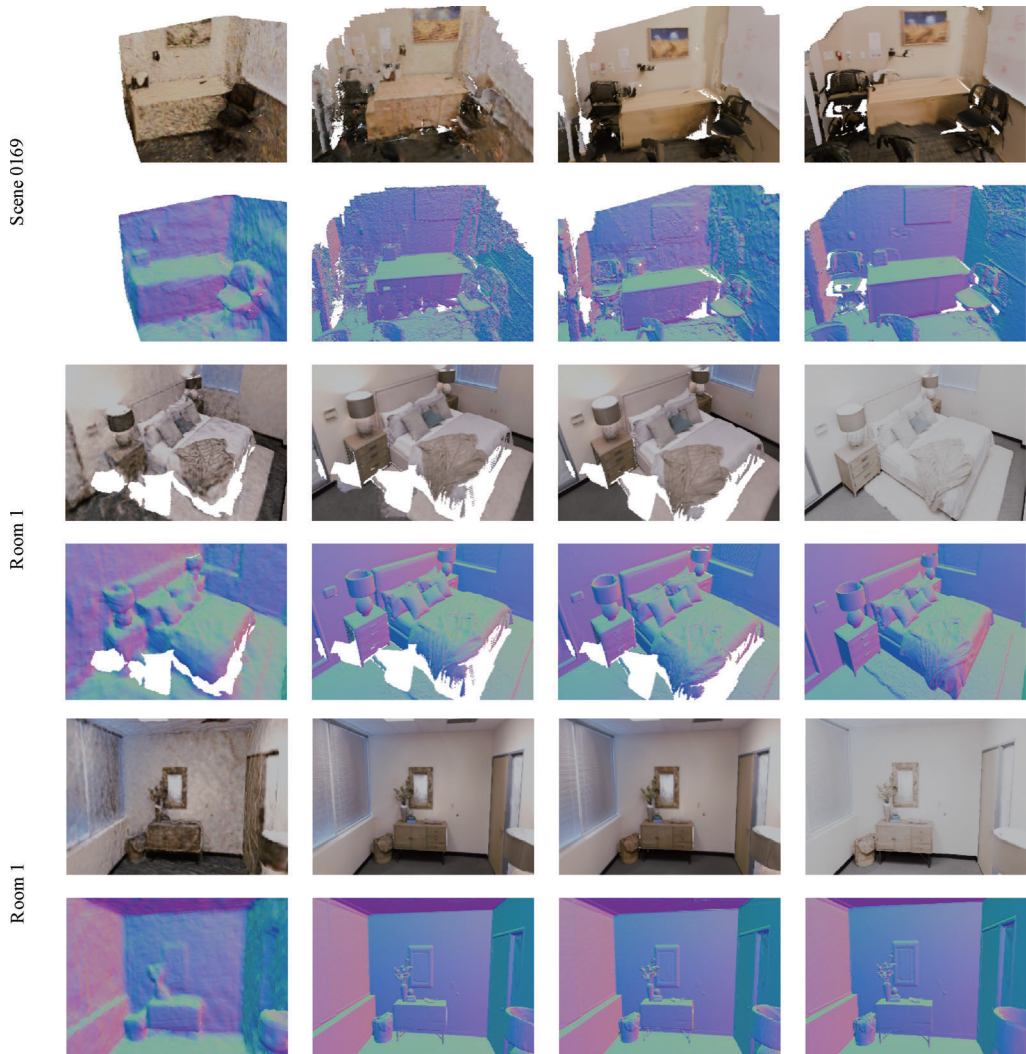


图3 ScanNet 和 Replica 数据集上的网格重建可视化结果对比

Fig. 3 Comparison of visualization results of mesh reconstruction on the ScanNet and Replica dataset

2.4 渲染质量评估

为了确保公平比较,本文将所有子地图合并为全局地图,并使用估计的相机姿态对全局地图进行优化,以避免子地图上的局部过拟合。表5展示了在合成数据集 Replica 上的渲染质量对比,表6和7则展示了在真实世界数据集 ScanNet 和 TUM-RGBD 上的评估结果。结果表明,本文算法在所有 SLAM 算法中渲染质量排名第2。与 Gaussian-SLAM 相比,本文算法的渲染质量略低,主要原因在于,本文在使用环路闭合优化跟踪位姿时,为提高运行速度,采用了省略球谐函数的方式以减少高斯地图的大小和计算复杂度。具体而言,算法通过姿态校正信息直接更新每个子地图的相机姿态、高斯均值和协方差。随着所有子地图合并成全局地图,这些误差逐渐累积,从而影响了全局地图的渲染质量。

2.5 内存和运行速度评估

表8展示了本文算法在运行速度与内存使用方面的

对比结果。实验表明,本文算法在逐帧跟踪和地图优化的时间上仅略低于 Gaussian-SLAM 方法。这是因为本文提出的混合表面法线跟踪方法以及在子地图优化部分使用的多种改进方法,虽然牺牲了一些运行速度,但获得了更高的跟踪精度和重建质量。得益于双重高斯点修剪策略对子地图高斯点数量的有效控制,本文算法的地图总内存容量比 Gaussian-SLAM 减少了 8.2%。同时,在处理单个房间规模的场景时,本文算法对 GPU 内存的占用最少。相比之下, Vox-Fusion 和 SplaTAM 等方法的 GPU 内存需求均超过 15 GB,资源消耗显著更高。

2.6 消融实验

为验证算法改进的有效性,本文在保持相同参数配置和训练环境的前提下,使用 ScanNet 数据集中的 Scene0059 场景进行消融实验。实验通过逐步将每项改进加入原始 Gaussian-SLAM 算法中进行消融实验,以评估其对整体性能的影响。具体而言,实验先后引入了曝光调

表 5 Replica 数据集上渲染质量对比

Table 5 Comparison of rendering quality on the Replica dataset

方法	指标	room0	room1	room2	office0	office1	office2	office3	office4	平均
NICE-SLAM ^[7]	PSNR ↑	22.12	22.47	24.52	29.07	30.34	19.66	22.23	24.94	24.42
	SSIM ↑	0.689	0.757	0.814	0.874	0.886	0.797	0.801	0.856	0.809
	LPIPS ↓	0.330	0.271	0.208	0.229	0.181	0.235	0.209	0.198	0.233
Vox-Fusion ^[8]	PSNR ↑	22.39	22.36	23.92	27.79	29.83	20.33	23.47	25.21	24.41
	SSIM ↑	0.683	0.751	0.798	0.857	0.876	0.794	0.803	0.847	0.801
	LPIPS ↓	0.303	0.269	0.234	0.241	0.184	0.243	0.213	0.199	0.236
Gaussian-SLAM ^[17]	PSNR ↑	38.88	41.80	42.44	46.40	45.29	40.10	39.06	42.65	42.08
	SSIM ↑	0.993	0.996	0.996	0.998	0.997	0.997	0.997	0.997	0.996
	LPIPS ↓	0.017	0.018	0.019	0.015	0.016	0.020	0.020	0.020	0.018
SplaTAM ^[16]	PSNR ↑	32.86	33.89	35.25	38.26	39.17	31.97	29.70	31.81	34.11
	SSIM ↑	0.987	0.976	0.983	0.989	0.985	0.971	0.954	0.958	0.975
	LPIPS ↓	0.073	0.104	0.082	0.096	0.095	0.104	0.128	0.153	0.104
ESLAM ^[9]	PSNR ↑	25.25	27.39	28.09	30.33	27.04	27.99	29.27	29.15	28.06
	SSIM ↑	0.874	0.890	0.935	0.934	0.910	0.942	0.953	0.948	0.923
	LPIPS ↓	0.315	0.296	0.245	0.213	0.254	0.238	0.186	0.210	0.245
本文	PSNR ↑	34.84	37.30	38.48	43.92	42.83	37.44	36.34	39.99	38.89
	SSIM ↑	0.977	0.981	0.982	0.987	0.987	0.986	0.986	0.986	0.984
	LPIPS ↓	0.068	0.055	0.064	0.043	0.056	0.061	0.062	0.059	0.058

表 6 ScanNet 数据集上渲染质量对比

Table 6 Comparison of rendering quality on the ScanNet dataset

方法	指标	Scene0000	Scene0059	Scene0106	Scene0169	Scene0181	Scene0207	平均
NICE-SLAM ^[7]	PSNR ↑	18.71	16.55	17.29	18.75	15.56	18.38	17.54
	SSIM ↑	0.641	0.605	0.646	0.629	0.562	0.646	0.621
	LPIPS ↓	0.561	0.534	0.510	0.534	0.602	0.552	0.548
Vox-Fusion ^[8]	PSNR ↑	19.06	16.38	18.46	18.69	16.75	19.66	18.17
	SSIM ↑	0.662	0.615	0.753	0.650	0.666	0.696	0.673
	LPIPS ↓	0.515	0.528	0.439	0.513	0.532	0.500	0.504
Gaussian-SLAM ^[17]	PSNR ↑	28.54	26.21	26.26	28.60	27.79	28.63	27.67
	SSIM ↑	0.926	0.934	0.926	0.917	0.922	0.914	0.923
	LPIPS ↓	0.271	0.211	0.217	0.226	0.277	0.288	0.248
SplaTAM ^[16]	PSNR ↑	19.33	19.27	17.73	21.97	16.76	19.83	19.14
	SSIM ↑	0.660	0.792	0.690	0.776	0.683	0.696	0.716
	LPIPS ↓	0.438	0.289	0.376	0.281	0.420	0.341	0.358
ESLAM ^[9]	PSNR ↑	15.70	14.48	15.44	14.56	14.22	17.32	15.29
	SSIM ↑	0.687	0.632	0.628	0.656	0.696	0.653	0.658
	LPIPS ↓	0.515	0.528	0.439	0.513	0.532	0.500	0.504
本文	PSNR ↑	26.17	25.38	25.42	27.74	26.18	27.46	26.39
	SSIM ↑	0.863	0.852	0.869	0.884	0.868	0.874	0.868
	LPIPS ↓	0.339	0.387	0.368	0.312	0.335	0.327	0.345

整、法线一致性损失、双重高斯修剪策略、混合法线跟踪方法、高斯初始几何重叠比和刚性高斯点云配准方法。通过这种逐步加入改进的方法,能够清晰地展示每项改进对算

法性能的提升效果。实验结果如表 9 所示。

表 9 的结果表明,所有提出的改进方法在提升相机跟踪精度和渲染质量方面均表现出显著效果。具体来说,在

表7 TUM-RGBD数据集上渲染质量对比

Table 7 Comparison of rendering quality on the TUM-RGBD dataset

方法	指标	fr1/desk	fr2/xyz	fr3/office	平均
NICE-SLAM ^[7]	PSNR ↑	13.83	17.87	12.890	14.86
	SSIM ↑	0.569	0.718	0.554	0.614
	LPIPS ↓	0.482	0.344	0.498	0.441
Vox-Fusion ^[8]	PSNR ↑	15.79	16.32	17.27	16.46
	SSIM ↑	0.647	0.706	0.677	0.677
	LPIPS ↓	0.523	0.433	0.456	0.471
Gaussian-SLAM ^[17]	PSNR ↑	24.01	25.02	26.13	25.05
	SSIM ↑	0.924	0.924	0.939	0.929
	LPIPS ↓	0.178	0.186	0.141	0.168
SplaTAM ^[16]	PSNR ↑	22.00	24.50	21.90	22.80
	SSIM ↑	0.857	0.947	0.876	0.893
	LPIPS ↓	0.232	0.100	0.202	0.178
ESLAM ^[9]	PSNR ↑	11.29	17.46	17.02	15.26
	SSIM ↑	0.666	0.310	0.457	0.478
	LPIPS ↓	0.358	0.698	0.652	0.569
本文	PSNR ↑	23.57	23.96	24.85	24.12
	SSIM ↑	0.886	0.892	0.907	0.895
	LPIPS ↓	0.226	0.207	0.193	0.209

表8 Replica数据集上运行速度和内存使用对比

Table 8 Comparison of running speed and memory usage on the Replica dataset

方法	跟踪每次	跟踪每帧	建图每次	建图每帧	渲染速度/	地图总内存	峰值 GPU
	迭代时间/ms	运行时间/s	迭代时间/ms	运行时间/s	fps	容量/MiB	使用量/GiB
NICE-SLAM ^[7]	27	1.06	89	1.15	2.64	95.9	12.0
Vox-Fusion ^[8]	64	1.92	98	1.47	1.63	0.15	17.6
Point-SLAM ^[11]	27	1.11	57	3.52	2.96	27.2	7.7
SplaTAM ^[16]	67	2.70	81	4.89	2.175	404.5	18.5
Gaussian-SLAM ^[17]	14	0.83	24	0.93	2.175	54.8	7.6
本文	16	0.85	27	0.96	2.175	50.3	7.4

表9 场景 Scene0059 上消融实验结果

Table 9 The results of ablation study in Scene0059

曝光调整	法线一致性损失	双重高斯修剪	混合法线跟踪	环路闭合		ATE RMSE/ cm ↓	PSNR/ dB ↑
				高斯初始几何重叠比	环路边配准		
✓						12.84	26.21
✓						12.87	26.66
✓	✓					12.78	27.31
✓	✓	✓				12.32	27.46
✓	✓	✓	✓			11.45	27.49
✓	✓	✓	✓	✓	PPFH+ICP	14.76	25.22
✓	✓	✓	✓		刚性高斯点云配准	9.93	25.34
✓	✓	✓	✓	✓	刚性高斯点云配准	9.68	25.38

子地图优化阶段,曝光调整策略有效改善了不同关键帧之间曝光度的差异所产生的误差,峰值信噪比提高了 1.7%。法线一致性损失同样有效约束了 3DGS 表面法线,导致峰值信噪比提高了 2.4%,显著改善了渲染图像的质量。采用双重高斯修剪策略替换原有的高斯修剪后,提高了 3D 高斯地图的重建质量,从而间接的提升了跟踪性能,绝对轨迹均方根误差减少了 3.6%,证明了双重高斯修剪方法相比原有方法的优越性。

在跟踪阶段,将基于恒速假设的帧到模型跟踪方法替换为本文提出的混合法线跟踪方法后,模型的绝对轨迹均方根误差减少了 7.1%,显著提高了跟踪精度。在环路闭合阶段,采用高斯初始几何重叠比进行环路检测后,绝对轨迹均方根误差减少了 2.5%,表明高斯初始几何重叠比能够有效排除生成的错误环路边。由于作为基准的 Gaussian-SLAM 算法未使用环路闭合,因此采用对比直接使用点云配准的 FPFH+ICP 方法和本文提出的刚性高斯点云配准方法进行消融实验。结果表明,直接将 FPFH+ICP 应用于 3D 高斯中心点会导致环路边缘精度下降,进而影响环路闭合效果。而将 FPFH+ICP 替换为刚性高斯点云配准方法后,绝对轨迹均方根误差从原来的 14.76 cm 降低至 9.68 cm,减少了 34.4%。这表明,刚性高斯点云配准方法不仅有效解决了在环路边直接使用点云配准时出现的问题,显著提升了跟踪精度和性能,还有效减少了随时间推移的跟踪误差,确保了精确的轨迹估计,并维持了连贯的 3D 地图表示,为复杂室内场景的处理提供了高效的解决方案。

3 结 论

本研究针对现有的基于 3DGS 的方法在复杂室内环境的三维重建中存在跟踪精度有限、缺乏全局一致性等难题,提出一种基于 3D 高斯溅射的密集 SLAM 算法——SNGO-SLAM。在子地图优化中,通过曝光调整和法线一致性损失,确保了全局一致性,同时引入双重高斯点修剪策略,以提升重建质量。在跟踪过程,提出了一种混合表面法线跟踪方法,显著提高了估计姿态和重建地图之间的一致性和准确性。在环路闭合过程中,提出一种刚性高斯点云配准方法,解决 3D 高斯中心点不能准确地表示场景表面的问题。

实验结果表明,SNGO-SLAM 在多个公开数据集上的表现超越了现有的主流方法,在跟踪精度、重建质量、运行速度和内存使用等方面均取得了显著进步。这表明 SNGO-SLAM 在需要高效、精确三维建图的复杂室内环境中具有极大的应用潜力。

尽管如此,仍需注意的是,本研究算法在全局姿态更新时通过选择省略球谐函数的方式,以牺牲渲染质量为代价保证实时性。此外,本研究算法未能充分利用 3D 高斯分布在承载语义特征方面的优势。这些问题为未来的改

进提供了明确的方向。未来研究将聚焦于将语义特征集成到 3D 高斯分布中并联合优化相机姿态、几何、外观和语义,以更好的满足复杂室内场景的实时应用需求。

参考文献

- [1] XU Y M, JIANG H CH, XIAO ZH Y, et al. DG-SLAM: Robust dynamic Gaussian splatting SLAM with hybrid pose optimization[J]. *Advances in Neural Information Processing Systems*, 2024, 37:51577-51596.
- [2] 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.
- [3] SCHÖPS T, SATTTLER T, POLLEFEYS M. BAD SLAM: Bundle adjusted direct RGB-D SLAM[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019:134-144.
- [4] 刘钰嵩,何丽,袁亮,等. 动态场景下基于光流的语义 RGBD-SLAM 算法[J]. *仪器仪表学报*, 2022, 43(12):139-148.
- [5] LIU Y S, HE L, YUAN L, et al. Semantic RGBD-SLAM algorithm based on optical flow in dynamic scenes[J]. *Chinese Journal of Scientific Instrument*, 2022, 43(12):139-148.
- [6] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis[J]. *Communications of the ACM*, 2022, 65(1):99-106.
- [7] SUCAR E, LIU S K, ORTIZ J, et al. iMAP: Implicit mapping and positioning in real-time[C]. *IEEE/CVF International Conference on Computer Vision*, 2021:6209-6218.
- [8] ZHU Z H, PENG S Y, LARSSON V, et al. NICE-SLAM: Neural implicit scalable encoding for SLAM[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022:12776-12786.
- [9] YANG X R, LI H, ZHAI H J, et al. Vox-Fusion: Dense tracking and mapping with voxel-based neural implicit representation [C]. *IEEE International Symposium on Mixed and Augmented Reality*, 2022: 499-507.
- [9] JOHARI M M, CARTA C, FLEURET F. ESLAM: Efficient dense SLAM system based on hybrid representation of signed distance fields[C]. *IEEE/CVF Conference on Computer Vision and Pattern*

- Recognition, 2023;17408-17419.
- [10] LISO L, SANDSTROM E, YUGAY V, et al. Loopy-SLAM: Dense neural SLAM with loop closures[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024;20363-20373.
- [11] SANDSTRÖM E, LI Y, VAN GOOL L, et al. Point-SLAM: Dense neural point cloud-based SLAM[C]. IEEE/CVF International Conference on Computer Vision, 2023;18387-18398.
- [12] TOSI F, ZHANG Y M, GONG Z R, et al. How NeRFs and 3D Gaussian splatting are reshaping SLAM: A survey[J]. ArXiv preprint arXiv:2402.13255, 2024.
- [13] KERBL B, KOPANAS G, LEIMKÜHLER T, et al. 3D Gaussian splatting for real-time radiance field rendering[J]. ACM Transactions on Graphics, 2023, 42(4):139.
- [14] 李斌, 张文慧, 项颖, 等. 基于3D高斯溅射的人物化身重建算法[J]. 电子测量技术, 2025, 48(10):177-185.
- LI B, ZHANG W H, XIANG Y, et al. A 3D Gaussian sputtering-based avatar reconstruction algorithm[J]. Electronic Measurement Technology, 2025, 48(10):177-185.
- [15] HUANG Y M, CUI B L, BAI L, et al. Advancing dense endoscopic reconstruction with Gaussian splatting-driven surface normal-aware tracking and mapping [C]. IEEE International Conference on Robotics and Automation, 2025;4084-4091.
- [16] KEETHA N, KARHADE J, JATAVALLABHULA K M, et al. SplatAM: Splat, track & map 3D Gaussians for dense RGB-D SLAM[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024;21357-21366.
- [17] YUGAY V, LI Y, GEVERS T, et al. Gaussian-SLAM: Photo-realistic dense SLAM with Gaussian splatting [J]. ArXiv preprint arXiv: 2312.10070, 2023.
- [18] YUGAY V, GEVERS T, OSWALD M R, MAGiC-SLAM: Multi-Agent Gaussian globally consistent SLAM[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2025;6741-6750.
- [19] ZHU L Y, LI Y, SANDSTRÖM E, et al. LoopSplat: Loop closure by registering 3D Gaussian splats[C]. International Conference on 3D Vision, 2025;156-167.
- [20] MATSUKI H, TATENO K, NIEMEYER M, et al. NEWTON: Neural view-centric mapping for on-the-fly large-scale SLAM [J]. IEEE Robotics and Automation Letters, 2024, 9(4):3704-3711.
- [21] HU J R, CHEN X H, FENG B Y, et al. CG-SLAM: Efficient dense RGB-D SLAM in a consistent uncertainty-aware 3D Gaussian field [C]. European Conference on Computer Vision, 2025;93-112.
- [22] YAN CH, QU D L, XU D, et al. GS-SLAM: Dense visual SLAM with 3D Gaussian splatting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024;19595-19604.
- [23] LI M R, GUO ZH T, DENG T C, et al. DDN-SLAM: Real time dense dynamic neural implicit SLAM[J]. IEEE Robotics and Automation Letters, 2025, 10(5):4300-4307.
- [24] MAO Y X, YU X, ZHANG ZH Q, et al. NGEL-SLAM: Neural implicit representation-based global consistent low-latency SLAM system [C]. IEEE International Conference on Robotics and Automation, 2024;6952-6958.
- [25] ZHOU H, GUO Z T, REN Y X, et al. MoD-SLAM: Monocular dense mapping for unbounded 3D scene reconstruction [J]. IEEE Robotics and Automation Letters, 2025, 10(1):484-491.
- [26] ZHANG G L, SANDSTRÖM E, ZHANG Y M, et al. GIORIE-SLAM: Globally optimized RGB-only implicit encoding point cloud SLAM [J]. ArXiv preprint arXiv:2403.19549, 2024.
- [27] KRUKHKOVA E, SAVINYKH A, KARPYSHEV P, et al. MeSLAM: Memory efficient SLAM based on neural fields [C]. IEEE International Conference on Systems, Man and Cybernetics, 2022;430-435.
- [28] ISAACSON S, KUNG P C, RAMANAGOPAL M, et al. LONER: LiDAR only neural representations for real-time SLAM[J]. IEEE Robotics and Automation Letters, 2023, 8(12):8042-8049.
- [29] ARANDJELOVIC R, GRONAT P, TORII A, et al. NetVLAD: CNN architecture for weakly supervised place recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(6):1437-1451.

- [30] MATSUKI H, MURAI R, KELLY P H J, et al. Gaussian splatting SLAM[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 18039-18048.
- [31] STRAUB J, WHELAN T, MA L N, et al. The replica dataset: A digital replica of indoor spaces[J]. ArXiv preprint arXiv:1906.05797, 2019.
- [32] DAI A, CHANG A X, SAVVA M, et al. ScanNet: Richly-annotated 3D reconstructions of indoor scenes[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:2432-2443.
- [33] DAI A, NIEßNER M, ZOLLHÖFER M, et al. BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface reintegration[J]. ACM Transactions on Graphics, 2017, 36(3):1-12.
- [34] ZHANG Y M, TOSI F, MATTOCCIA S, et al. GO-

SLAM: Global optimization for consistent 3D instant reconstruction[C]. IEEE/CVF International Conference on Computer Vision, 2023:3704-3714.

作者简介

姜俊超, 硕士研究生, 主要研究方向为视觉 SLAM、三维重建。

E-mail:jiangjunchao0013@163.com

王永兰(通信作者), 副教授, 硕士生导师, 主要研究方向为 SLAM、机器学习、目标检测。

E-mail:wangyl4901@163.com

房建东, 教授, 硕士生导师, 主要研究方向为机器人控制、机器人自主决策。

E-mail:z1435562262@126.com

朱瑾, 硕士研究生, 主要研究方向为自动驾驶、目标检测与跟踪。

E-mail:20231800089@imut.edu.cn