

中图法分类号: TP391.4; TP242.5 文献标识码: A 文章编号: 1006-8961(2025)06-1792-16

论文引用格式: Tan Z, Niu Z Y, Zhang J P, Chen X Y L and Hu D W. 2025. New opportunities in SLAM-Gaussian splatting technology. Journal of Image and Graphics, 30(6): 1792-1807 [谭臻, 牛中颜, 张津浦, 陈谢沅澧, 胡德文. 2025. SLAM新机遇—高斯溅射技术. 中国图象图形学报, 30(6): 1792-1807] [DOI: 10.11834/jig.240443]

SLAM新机遇—高斯溅射技术

谭臻, 牛中颜, 张津浦, 陈谢沅澧, 胡德文*

国防科技大学智能科学学院, 长沙 410073

摘要: 同步定位与建图(simultaneous localization and mapping, SLAM)是指在未知环境中同时实现自主移动机器人的定位和环境地图构建,其在机器人技术和自动驾驶等领域有着重要价值。本文首先回顾SLAM技术的发展历程,从早期的人工特征提取方法到现代的深度驱动解决方案。其中,基于神经辐射场(neural radiance fields, NeRF)的SLAM方法利用神经网络进行场景表征,进一步提高了建图的可视化效果。然而,这类方法在渲染速度上仍然面临挑战,限制了其实时应用的可能性。相比之下,基于高斯溅射(Gaussian splatting, GS)的SLAM方法以其实时的渲染速度和照片级的场景渲染效果,为SLAM领域带来新的研究热点和机遇。接着,按照RGB/RGBD、多模态数据以及语义信息3种不同应用类型对基于高斯溅射的SLAM方法进行分类和总结,并针对每种情况讨论相应SLAM方法的优势和局限性。最后,针对当前基于高斯溅射的SLAM方法面临的实时性、基准一致化、大场景的扩展性以及灾难性遗忘等问题进行分析,并对未来研究方向进行展望。通过这些探讨和分析,旨在为SLAM领域的研究人员和工程师提供全面的视角和启发,帮助分析和理解当前SLAM系统面临的关键问题,推动该领域的技术进步和应用拓展。

关键词: 同步定位与建图(SLAM);神经辐射场(NeRF);高斯溅射(GS);RGB-(D);多模态;语义信息

New opportunities in SLAM-Gaussian splatting technology

Tan Zhen, Niu Zhongyan, Zhang Jingpu, Chen Xieyuanli, Hu Dewen*

College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China

Abstract: Simultaneous localization and mapping (SLAM) has undergone a profound evolution, transitioning through various stages and methodologies, each of which has contributed significantly to advancements in accuracy, robustness, and applicability across diverse scenarios. This paper provides a comprehensive exploration of the historical development and current trends in SLAM, with a particular focus on the progression from manual feature extraction to the adoption of modern deep learning and 3D graphics-based approaches. In the early stages of SLAM development, the process relied heavily on manual feature extraction, where visual features were carefully selected and extracted by human operators to facilitate localization and mapping tasks. Although this method proved effective in relatively simple environments, it was highly susceptible to the complexities of more dynamic scenes and variations in illumination. The dependency on human intervention for feature selection not only limited the scalability of these systems but also constrained their robustness in dynamic or unpredictable environments. The introduction of visual SLAM marked a pivotal advancement in the field. By leveraging the rapid progress in computer vision technologies, such as improved feature matching algorithms and visual odometry, visual SLAM systems significantly enhanced both the robustness and accuracy of SLAM. These innovations enabled SLAM systems to

收稿日期: 2024-07-31; 修回日期: 2024-10-08; 预印本日期: 2024-10-15

* 通信作者: 胡德文 dwhu@nudt.edu.cn

perform more reliably across a wide array of real-world environments, thereby paving the way for more automated and efficient approaches to SLAM. The integration of deep learning into SLAM methodologies represents a paradigm shift in scene understanding and reconstruction. One of the most notable advancements in this area is the emergence of neural radiance fields (NeRF)-based SLAM methods. NeRF-based approaches can model dense depth and color information with unprecedented accuracy, providing a more detailed and precise understanding of the environment. However, these methods are not without their challenges, particularly in terms of computational efficiency and real-time performance. Such challenges are especially pertinent in applications where rapid data processing and immediate response are crucial. In response to the limitations of existing SLAM methods, 3D Gaussian splatting (3DGS) technology has emerged as a promising alternative. SLAM methods based on 3DGS offer significant improvements in rendering speed and high-fidelity scene reconstruction. This technology enhances the speed and quality of spatial data processing, making it particularly well-suited for applications such as augmented reality and autonomous navigation systems. The inherent robustness of 3DGS-based SLAM methods makes them especially effective in large-scale environments and scenarios characterized by dynamic changes. The categorization of SLAM methodologies can be further refined by considering the types of sensory inputs and the specific application needs they address. Initially, SLAM methods utilizing simple RGB and RGB-D sensors were primarily focused on capturing visual and depth information, which proved effective in controlled environments, particularly indoors. These methods excel in scenarios with ample lighting, well-defined textures, and minimal occlusions. However, in more challenging conditions — such as outdoor environments, low-texture or textureless surfaces, and areas with significant lighting variations — RGB and RGB-D SLAM methods often face substantial limitations. For instance, in outdoor environments, changes in lighting, shadows, or exposure to direct sunlight can severely degrade the performance of these methods, leading to inaccuracies in localization and mapping. Similarly, in textureless regions like white walls or glass surfaces, RGB-D sensors struggle to capture sufficient visual features, resulting in failed or unreliable reconstructions. These limitations underscore the need for more robust SLAM solutions that can operate effectively across a broader range of conditions. Multimodal SLAM approaches have been developed to address these challenges by integrating data from multiple sensors, such as LiDAR, thermal cameras, and inertial measurement units (IMUs). By combining visual data with other sensory inputs, multimodal SLAM systems can overcome the weaknesses inherent in RGB and RGB-D-based methods. For example, LiDAR can provide accurate depth measurements even in low-light or textureless environments, whereas thermal cameras can detect heat signatures, aiding in environments where visual data are insufficient or unreliable. Moreover, IMUs contribute to maintaining accurate localization in scenarios with rapid motion or poor visual conditions by providing supplementary motion and orientation data. In scenarios where precise geometric reconstruction and higher-level environmental understanding are required, recent advancements have combined semantic information with SLAM methods based on 3DGS technology. By incorporating semantic cues, these methods not only enhance the accuracy of scene reconstruction and localization but also provide crucial perceptual information for downstream tasks such as robotic navigation, augmented reality, and embodied intelligence. This integration allows SLAM systems to interpret and adapt to complex, dynamic environments more effectively, making them suitable for a wide range of real-world applications, from indoor mapping to outdoor navigation in varied lighting and textural conditions. The nuanced understanding of when to deploy RGB or RGB-D SLAM versus multimodal SLAM is critical for optimizing the performance and applicability of these systems. Although RGB and RGB-D methods are efficient and effective in controlled, well-lit indoor environments, multimodal SLAM approaches are indispensable for applications in outdoor, textureless, or dynamically changing environments where robustness and adaptability are paramount. Despite the significant advancements in SLAM technology, current 3DGS-based methods still face several challenges. These include issues related to scalability in large-scale scenes, adaptability to dynamic environments, and the optimization required for real-time performance. Addressing these challenges is a key focus of ongoing research, which aims to integrate deep learning techniques with traditional geometric methods. Such integration is expected to further enhance the overall performance and versatility of SLAM systems. Moreover, the establishment of unified evaluation benchmarks is crucial for standardizing performance metrics across different SLAM methodologies. Such benchmarks will facilitate greater transparency and comparability in research outcomes, thereby driving further innovation in the field. In conclusion, the evolution of SLAM methodologies — from manual feature extraction to deep learning and 3D graphics-based

approaches — has significantly advanced the capabilities of SLAM systems. By examining the historical developments, current methodologies, and future research directions, this paper provides researchers and engineers with comprehensive insights into the complexities and opportunities associated with advancing SLAM technology. Continued innovation and interdisciplinary collaboration will be essential in driving further advancements, enabling SLAM systems to fulfill their potential across a wide range of practical applications.

Key words: simultaneous localization and mapping (SLAM); neural radiance fields (NeRF); Gaussian splatting (GS); RGB-(D); multimodal; semantic information

0 引言

同步定位与建图(simultaneous localization and mapping, SLAM)是指在未知环境中同时实现自主移动机器人的定位(localization)和环境地图构建(mapping),是机器人技术和自动驾驶中的一项关键技术,在过去的几十年中得到了广泛的研究(高翔等, 2017; Chen 等, 2020; 刘瑞军 等, 2020; 龙霄潇等, 2021; 马艳阳 等, 2021; Tosi 等, 2024)。如图 1 所示,SLAM 技术的发展经历了多个重要的阶段。

1) 20 世纪 80 年代中期,随着机器人技术的进步,人们开始关注移动机器人在未知环境中的导航问题。Smith 和 Cheeseman(1986)正式提出 SLAM 问题,并提出使用扩展卡尔曼滤波(extended Kalman filter, EKF)(Kalman 和 Bucy, 1961)来解决 SLAM 问题。

2) 20 世纪 90 年代,SLAM 问题开始受到广泛关注。Hugh Durrant-Whyte 的研究小组展示了 SLAM 问题的解在无限数据极限下是存在的(Leonard 和 Durrant-Whyte, 1991)。Simmons 和 Koenig(1995)引入概率机器人学(probabilistic robotics)概念,使用贝叶斯滤波(Derin 等, 1984)和粒子滤波(particle filter)等方法来处理 SLAM 问题的不确定性。

3) 进入 21 世纪,SLAM 技术逐渐成熟。Montemerlo 等人(2002)提出 FastSLAM 算法。Grisetti 等人(2010)提出图优化 SLAM 方法。视觉 SLAM 使用相机作为主要传感器,结合计算机视觉技术。ORB-SLAM (Mur-Artal 等, 2015)和 VINS-Mono (Qin 等, 2018)是其中的代表性工作。现代 SLAM 系统通常融合多种传感器数据,如激光雷达(light detection and ranging, LiDAR)、相机和惯性测量单元(inertial measurement unit, IMU)等,以提高 SLAM 的精度和鲁棒性。例如,Google 的 Cartographer(Hess 等,

2016)。

4) 随着学习方法的兴起,深度学习等方法在图像处理和特征提取方面的强大能力开始被引入到 SLAM 中。神经辐射场(neural radiance fields, NeRF)(Mildenhall 等, 2021)因其具有照片级的可视化场景重建能力,不少工作开始研究如何将 NeRF 与 SLAM 进行结合。基于 NeRF 的 SLAM 方法相较于传统方法在处理复杂场景如复杂光照、弱纹理时表现出色,但其渲染速度和精度仍然具有较大的提升空间。为了能够进一步提高渲染速度和质量,3D 高斯溅射(Gaussian splatting, GS)(Kerbl 等, 2023)迅速被引入到 SLAM 任务中,高斯溅射技术以其实时渲染和显式表征的特点为 SLAM 问题带来了新的热潮。本文对高斯溅射技术在 SLAM 的不同应用类型进行系统梳理,旨在为基于高斯溅射的 SLAM 方法在机器人导航、增强现实和虚拟现实等领域的进一步发展提供参考。

1 SLAM 技术概述

1.1 SLAM 基本原理

SLAM 问题可以被形式化为一个状态估计问题,其目标是在未知环境中,根据机器人的运动轨迹和传感器观测到的数据,估计机器人的当前位置以及环境的地图。其基本原理涉及传感器数据的处理、状态估计和地图更新 3 个主要步骤。

1) 传感器数据的处理。SLAM 系统通常集成多种传感器,如相机、LiDAR 和 IMU,以全面获取环境信息。相机捕捉视觉信息,提供图像数据用于视觉 SLAM。激光雷达通过发射和接收激光脉冲,生成高精度点云数据。IMU 提供加速度和角速度信息,帮助估计机器人的运动状态。所有这些传感器数据通过复杂算法进行融合和分析,以形成时间序列数据。

2)状态估计。在SLAM中,机器人需要同时估计自身的位置(定位)和环境的结构(地图)。这一过程通过融合传感器数据和先验信息动态更新机器人的状态,常用的方法包括粒子滤波器(Thrun,2002)、扩展卡尔曼滤波器(EKF)(Kalman 和 Bucy,1961)和优化方法。粒子滤波器适用于非线性和高维度状态空间,通过一组粒子近似状态分布。EKF通过线性化近似非线性系统的状态。优化方法通过最小化误差函数,全局优化机器人轨迹和地图。

3)地图更新。通过融合新观测数据更新环境地图,确保其精度和一致性。传统的SLAM系统中,地图通常以特征点或地标的形式表示。特征提取与描述是地图更新的基础,通过算法识别环境中的显著特征点。地图更新方法根据不同表示方式而异,特征点地图适用于稀疏特征环境,体素地图将环境划分为三维网格,网格地图用于二维平面。现代SLAM系统开始采用神经网络表示和更新地图,通过新观测数据进行训练和调整。

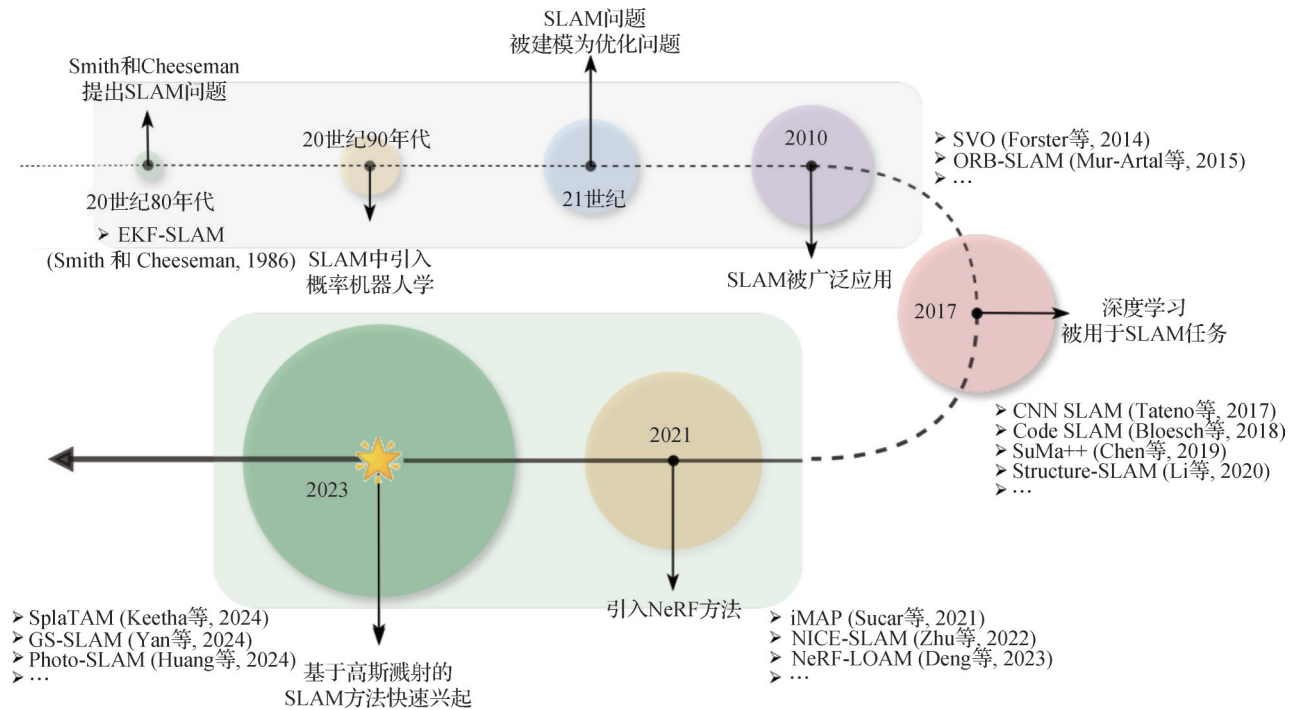


图1 SLAM的发展历程

Fig. 1 The overall development of SLAM

1.2 SLAM方法回顾

1.2.1 基于手工特征的SLAM

基于手工设计的特征提取和匹配技术的SLAM是一个经典且非常重要的研究方向。借助于从图像或点云数据中识别的显著特征,如角点(corners)和边缘(edges),实现环境地图的构建和机器人自身定位。这些特征具有明确的几何和视觉特性,能够在不同视角和光照条件下保持较高的鲁棒性和可辨识度。随着计算机视觉和图像处理技术的进步,视觉SLAM开始受到关注。视觉SLAM利用相机捕获的图像进行定位和地图构建,适用于无全球定位系统(global positioning system, GPS)信号的室内外环境。

ORB-SLAM(Mur-Artal 等,2015)是一种典型的基于视觉特征的SLAM系统,使用ORB特征进行实时的定位和地图构建。它结合了FAST(features from accelerated segment test)特征检测(Rosten 和 Drummond,2006)、BRIEF(binary robust independent elementary features)特征描述符(Calonder 等,2010)和ORB(oriented FAST and rotated BRIEF)描述符(Rublee 等,2011),以及RANSAC(random sample consensus)(Fischler 和 Bolles,1981)等技术来实现实时的视觉SLAM。VINS-Mono(Qin 等,2018)是一种结合单目视觉和惯性测量单元(IMU)数据的紧耦合SLAM系统,通过滑动窗口优化和回环检测,实现高精度的实时位姿估计和环境地图构建,适用于动态

和复杂环境。PIPO-SLAM (Ge 等, 2024) 是一种轻量级的视觉惯性 SLAM 算法, 提出基于流形的预积分合并理论, 可以有效地合并预积分项和噪声协方差, 采用仅位姿描述的多视图几何方法, 降低了算法的计算复杂度, 使其更加轻量级和高效。

激光雷达与 SLAM 问题的结合也为 SLAM 提供了新的发展思路。与相机相比, 激光雷达不受光照变化的影响, 能够在各种光线条件下提供稳定的测距数据。随着激光雷达技术的发展, 基于其高分辨率和精确测距能力的 SLAM 方法越来越流行, 特别适用于无人机和机器人等领域。

除此之外, 还有其他一些使用手工设计特征的 SLAM 系统, 如 PTAM (parallel tracking and mapping) (Klein 和 Murray, 2007)、SVO (semi-direct visual odometry) (Forster 等, 2014) 等。这些系统在一定程度上解决了视觉 SLAM 中的定位和建图问题, 但它们通常对环境的要求比较严格, 对于复杂、动态的场景可能表现不佳。随着深度学习技术的发展, 基于深度学习的 SLAM 系统 (如基于神经网络的特征提取与匹配) 逐渐成为研究的热点, 能够更好地适应复杂环境下的定位和建图任务。

1.2.2 基于学习的 SLAM

基于学习的 SLAM 技术利用机器学习、深度学习的强大学习能力, 实现对复杂场景的高精度定位和高效建图, 具有自动化特征提取、端到端优化和多模态数据融合的优势。其中, CNN-SLAM (Tateno 等, 2017) 将深度学习与传统 SLAM 系统结合, 利用卷积神经网络 (convolutional neural network, CNN) 提取图像特征, 学习环境语义信息和机器人状态, 显著提升地图构建和定位精度。同时, Code-SLAM (Bloesch 等, 2018) 结合编码器解码器架构与 SLAM 系统, 编码器将输入数据转换为潜在空间表示, 解码器将其转换为所需输出, 有效地将传感器数据转换为高维特征表示, 实现精确的地图构建和位置估计。而 Structure-SLAM (Li 等, 2020) 利用环境结构信息增强 SLAM 性能, 包括几何形状、特征和语义信息 (如墙壁、地板和家具), 提高地图准确性和机器人定位, 进而实现更可靠的室内导航和建图能力。SuMa++ (Chen 等, 2019) 则利用快速激光雷达范围图像语义分割结果 (Li 等, 2022) 实现了室外自动驾驶场景中的实时激光雷达语义 SLAM 以及准确鲁棒的大范围长时间地点识别和闭环检测 (Chen 等, 2022)。

近年来, NeRF 技术 (Mildenhall 等, 2021) 的出现推动了基于 NeRF 的 SLAM 系统的发展。NeRF 是一种基于学习的隐式场景表征方法, 用于场景重建和新视角合成任务。NeRF 通过训练一个神经网络, 将输入的图像坐标和视线方向映射到场景中的颜色和密度。通过这种方式, NeRF 能够生成高质量的三维场景重建结果。NeRF 技术相对于之前的深度学习三维重建方法在场景建模上具有更高的精度和逼真度, 能够以全局一致的方式捕捉空间中每个点的光线密度和颜色信息。

目前已有大量的研究者聚焦于基于 NeRF 的 SLAM 方法上。iMAP (Sucar 等, 2021) 首先提出利用单一的多层感知机 (multilayer perceptron, MLP) 表示场景, 采用了关键帧结构、多线程计算流和动态信息引导的像素采样技术。与传统的密集 SLAM 技术相比, iMAP 的基于隐式 MLP 的方法具有高效的几何表示、自动细节控制, 以及对未观测区域如物体背面的平滑、合理的填充优势。NICE-SLAM (Zhu 等, 2022) 则进一步采用分层策略和智能特征更新, 优化可见网格特征以解决重建挑战。DIM-SLAM (Li 等, 2023) 引入神经隐式地图表示和多视图光度扭曲损失, 实现深度和颜色预测及场景优化。FMapping (Hua 等, 2023) 则提出利用神经场制图框架和互补滑动窗口策略优化场景重建, 提升了系统鲁棒性和稳定性。Hi-Map (Hua 等, 2024) 结合多层次特征网格和双路径编码增强了对无纹理区域的建模能力。NeRF-LOAM (Deng 等, 2023) 整合了 NeRF 和三维点云与建图技术, 通过神经隐式方法实现室外环境的稠密 3D 表示。LONER (Isaacson 等, 2023) 通过并行处理激光雷达扫描和 MLP 表示场景, 优化了地图质量和准确性。NeU-NBV (Jin 等, 2023) 利用 NeRF 渲染的不确定性来规划下一步相机采样的位置, 从而提升三维重建整体的精度。PIN-SLAM (Pan 等, 2024) 利用基于点的隐式神经地图表示实现了全局一致性的地图构建和闭环检测。NeuV-SLAM (Guo 等, 2024a) 利用神经符号距离函数 (sign distance function, SDF) 体素和颜色特征优化隐式场景表示, 提升了地图质量和准确性。Plenoxel-SLAM (Teigen 等, 2024) 通过 Plenoxel 辐射场模型和三线性插值实现实时跟踪, 未采用明确的学习表征的 3D 网格重建。

此外, 还有一些基于 NeRF 的联合位姿优化和场景重建方法。NeRFmm (Wang 等, 2022) 能够

在只有一组 2D 图像的情况下, 无需预先计算相机参数即可进行新视角合成和场景重建。NoPe-NeRF (Bian 等, 2023) 则通过进一步引入单目深度先验, 解决了重建中的部分伪影和联合优化不收敛问题。为了能够提高对深度噪声的抗干扰能力, TD-NeRF (Tan 等, 2024) 提出一种基于截断正态分布的光线采样策略, 引入了一种由粗到细的训练策略, 不仅可以提高相机姿态估计的收敛速度和精度, 还能够逐步提高深度几何的精度, 避免陷入局部最小值。TD-NeRF 在联合优化相机姿态和 NeRF 方面取得了优于之前方法的性能, 并生成了更加精确的深度几何信息。这些工作为 NeRF 在 3D 重建和 SLAM 任务中的应用提供了重要的技术支持。

然而, 该系列系统也面临着一些挑战, 如计算成本高、数据要求大等。从 2023 年高斯溅射技术出现之后, 显著降低了计算成本, 大大加快了渲染速度, 逐渐得到了广泛的应用, 基于高斯溅射的 SLAM 方法逐渐增多。

2 高斯溅射技术概述

高斯溅射技术 (Kerbl 等, 2023) 是一种用于三维场景表示和渲染的有效方法。这种方法通过将场景中的几何信息编码为一组高斯分布, 能够以一种紧凑且连续的方式表示三维空间中的显式点云数据。

空间中的物体通常是稀疏离散存在的, 而在 NeRF 中将空间使用 MLP 建模成一种连续的隐式表征, 这实际上存在较大的浪费, 并且场景的泛化能力和再编辑能力受限于这种隐式表征; 此外, 射线采样的方法也显著降低了渲染速度。而高斯溅射使用 3D 高斯球作为基元去拟合有界闭合平滑的曲面, 并且使用自适应策略去学习空间中物体的分布。每个基元由一个三维协方差矩阵 $\Sigma_i \in \mathbf{R}^{3 \times 3}$ 及均值 $\mu_i \in \mathbf{R}^3$ 表示, 具体为

$$g_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i)\right) \quad (1)$$

式中, $\mathbf{x}$ 表示空间坐标, $\Sigma = RSS^T R^T$, $R \in \mathbf{R}^{3 \times 3}$ 表示旋转矩阵, $S \in \mathbf{R}^3$ 表示空间尺度矩阵。将 3D 高斯球投影为 2D 高斯椭圆的过程称为溅射 (splating), 投影后的 2D 协方差矩阵和均值分别为 $\Sigma' = J\mathbf{W}\Sigma\mathbf{W}^T J^T$ 和 $\mu' = J\mathbf{W}\mu$, 其中 $\mathbf{W}$ 表示世界坐标系到相机坐标系

的变换, J 表示投影变换的仿射近似的雅可比矩阵。与 NeRF 类似, 高斯溅射使用离散的体渲染公式来计算每个像素上重叠的 3D 高斯投影混合颜色 C , 不同高斯的次序由其深度决定, 表达式为

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (2)$$

式中, c_i 表示第 i 个高斯的颜色。 α_i 表示最终的不透明度, 即不透明度 o_i 和 2D 高斯密度的乘积, 计算式为

$$\alpha_i = o_i \exp\left(-\frac{1}{2}(\mathbf{x}' - \mu'_i)^T \Sigma'_i{}^{-1}(\mathbf{x}' - \mu'_i)\right) \quad (3)$$

式中, $\mathbf{x}'$ 和 μ'_i 表示投影空间中的坐标。与 NeRF 类似, 深度 D 被渲染为

$$D = \sum_{i \in N} d_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (4)$$

式中, d_i 表示第 i 个 3D 高斯的中心的 z 轴深度。

高斯溅射技术在计算机视觉领域有着广泛的应用, 例如, 稀疏建图 (Li 等, 2024a)、渲染新视图 (Hamdi 等, 2024; Song 等, 2024)、人体姿态建模 (Hu 等, 2024b)、压缩渲染场 (Niedermayr 等, 2024)、动态场与变形场 (Yang 等, 2024; Li 等, 2024d) 等方向。最近, 高斯溅射技术也被应用于 SLAM 中。

3 基于高斯溅射的 SLAM 方法

在 SLAM 任务中引入 3D 高斯溅射技术, 通过实时可微的 3D 高斯渲染管线, 实现了对场景的高质量和高效率渲染, 高斯溅射 SLAM 的相关算法如表 1 所示。与传统的 SLAM 方法相比, 3D 高斯渲染能够提供更精细的几何结构和更逼真的视觉效果, 同时保持了实时渲染速度和优化效率。这种表征方式还为相机位姿估计和场景优化提供了一种可微的方法, 使得相机位姿优化和场景重建变得更加平滑和一致。这种可微特性极大地增强了 SLAM 系统的稳定性和可控性, 使得在复杂场景中的应用更加可靠和精确。此外, 3D 高斯的显式表示为 SLAM 系统的后续应用, 如动态重建、几何编辑和物理模拟等, 提供了更多可能性。本文将从不同的应用属性详细介绍高斯溅射与 SLAM 结合的新方法。

3.1 基于高斯溅射的 RGB/RGB-D SLAM

本小节将重点介绍基于高斯溅射的 RGB 和 RGB-D SLAM 方法及其方法描述。如图 2 所示, 在

表 1 高斯溅射 SLAM 相关算法总结
Table 1 The summary of different SLAM methods with Gaussian splatting

方法	输入	策略方向	针对问题
EndoGSLAM(Wang等,2024)	RGB	引入高斯溅射和可微分光栅化	内窥镜手术
NGM-SLAM(Li等,2024b)	RGB	引入神经子图	冗余高斯去除,渲染速度提升
Splat-SLAM(Sandström等,2024)	RGB	多视图深度与单目深度结合	重建速度
TAMBRIDGE(Jiang等,2024a)	RGB	ORB和高斯溅射结合(引入传统视觉里程计)	提升速度(>5帧/s),解决运动模糊、传感器噪声问题
MGS-SLAM(Zhu等,2024a)	RGB	稀疏视觉里程计和高斯溅射结合,引入预训练多视图立体网络估计深度	解决单目输入的尺度问题
Monocular Gaussian SLAM(Lan等,2024)	RGB	基于CLIP特征的闭环模块(场景的高层理解)	闭环检测
GS-SLAM(Yan等,2024)	RGBD	初步尝试	解决相机跟踪鲁棒性问题,建图精度提升
SplaTAM(Keetha等,2024)	RGBD	引入剪影掩码筛选有效像素	高斯地图更新
Gaussian-SLAM(Yugay等,2024)	RGBD	独立子图	解决建图时的灾难性遗忘
RGBD GS-ICP SLAM(Ha等,2024)	RGBD	ICP方法作为前端(引入点云匹配策略)	提升速度(系统速度可达107帧/s)
Compact-3DGS SLAM(Deng等,2024)	RGBD	滑动窗口,码书(codebook)	冗余高斯去除,渲染速度提升
High-Fidelity SLAM(Sun等,2024b)	RGBD	引入额外的正则化	渲染速度提升,跟踪鲁棒性
RTG-SLAM(Peng等,2024)	RGBD	高斯属性正则化,优化策略修改	提升速度(>17帧/s)
Object-centric 3DGS(Barad等,2024)	RGBD	以物体为中心进行增量重建和跟踪	物体级重建,相机跟踪鲁棒性问题
CG-SLAM(Hu等,2024a)	RGBD	引入不确定建模	相机跟踪鲁棒性问题,建图精度提升
Photo-SLAM(Huang等,2024)	RGB/RGBD	显示和隐式表征结合,超基元地图	相机跟踪鲁棒性问题,建图精度提升
Gaussian splatting SLAM(Matsuki等,2024)	RGB/RGBD	推导相机位姿相对于三维高斯的解析雅克比,新的高斯正则项	相机跟踪鲁棒性问题
MotionGS(Guo等,2024b)	RGB/RGBD	双关键帧选择,新的损失函数	相机跟踪鲁棒性问题,建图精度提升
SemGauss-SLAM(Zhu等,2024b)	RGBD+语义	语义特征嵌入高斯表征,语义光束法平差	相机跟踪鲁棒性问题,建图一致性问题
SGS-SLAM(Li等,2024c)	RGBD+语义	语义引导的关键帧选择策略	过平滑问题,累积误差
NEDS-SLAM(Ji等,2024)	RGBD+语义	虚拟相机视图修剪策略	冗余高斯去除
MM3DGS SLAM(Sun等,2024a)	多模态	发布数据集(RGB、RGB-D、LiDAR、IMU)	数据集缺失问题
3DGS-ReLOc(Jiang等,2024b)	多模态	2D体素地图和KD树结合	视觉定位问题
Gaussian-LIC(Lang等,2024)	多模态	激光雷达惯性相机融合建图	针对恶劣条件(高动态、低照明等)
MM-Gaussian(Wu等,2024)	多模态	重新设计定位模块	无界场景

RGB SLAM 中,高斯溅射技术通过处理单目或双目相机捕获的图像数据,构建三维环境模型。使用高斯分布来表示每个像素点的几何和颜色信息,可以在不增加计算复杂度的前提下,显著提升场景的细节和视觉效果。而在基于高斯溅射的 RGB-D SLAM 中,进一步结合了深度信息。如图 3 所示,通过 RGB-D 相机同时获取图像和深度数据,构建更加精细和准确的三维环境模型。

1)GS-SLAM(Yan 等,2024)首次采用了 3D 高斯表示和实时可微分 splat 渲染管线的方法。与传统的 SLAM 系统相比,GS-SLAM 能够在效率和准确性之间取得更好的平衡。GS-SLAM 提出一种自适应扩展策略,可以有效地重建新观察到的场景几何形状,并改善之前观察区域的建图。这种策略通过动态管理 3D 高斯元素,优化当前观察并最小化密集地

图和图像中的错误,从而提高了场景重建的性能。此外,GS-SLAM 还引入了一种从粗到细的技术,先使用低分辨率图像进行初始姿态估计,然后使用选择的可靠 3D 高斯进行高分辨率渲染来优化相机姿态。这种方法不仅能够减少运行时间,还能提高系统的鲁棒性。

2)SplaTAM (Keetha 等,2024)的核心创新在于采用 3D 高斯分布来表示环境地图,这种概率性的地图表示方式能够更好地捕捉环境的不确定性。同时,SplaTAM 引入了一种基于轮廓掩码的方法来处理深度数据中的噪声和缺失,进一步增强了系统在复杂环境下的性能。与传统的 SLAM 系统相比,SplaTAM 能够在实时性能的基础上生成密集的 3D 高斯地图,为后续的感知和导航任务提供更丰富的环境信息。

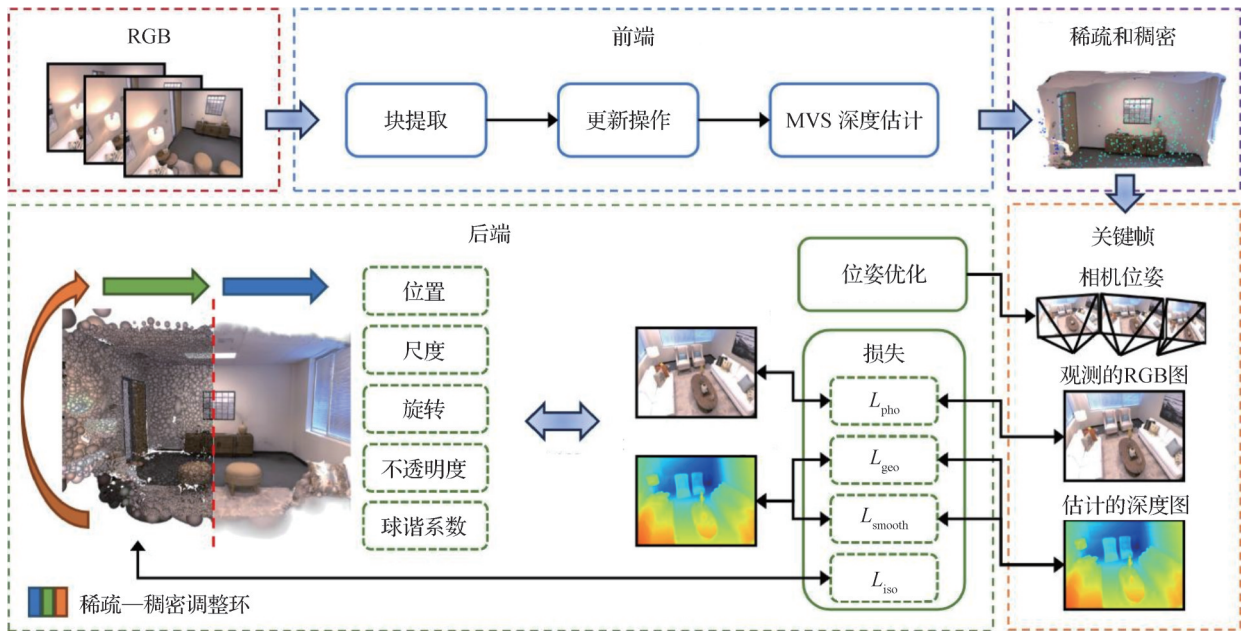


图2 基于高斯溅射的 RGB SLAM 框架(Zhu 等,2024a)

Fig. 2 The framework of the RGB SLAM with Gaussian splatting(Zhu et al. , 2024a)

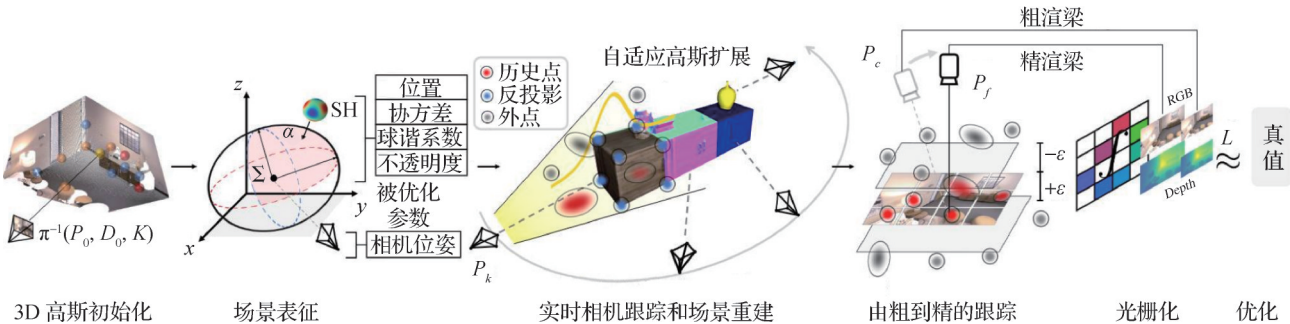


图3 基于高斯溅射的 RGB-D SLAM 框架(Yan 等,2024)

Fig. 3 The framework of the RGB-D SLAM with Gaussian splatting(Yan et al. , 2024)

3) RTG-SLAM (Peng 等, 2024) 是一个实时 3D 重建系统, 它采用高效的高斯表示和优化方案来实现大规模环境的实时重建。与传统的基于 NeRF 的 RGBD SLAM 相比, RTG-SLAM 在重建质量、相机跟踪精度和渲染逼真度方面都有显著优势。RTG-SLAM 使用紧凑的高斯表示来建模场景, 将高斯分布分为不透明和半透明两种。不透明的高斯拟合表面和主要颜色, 半透明的高斯拟合残余颜色。这种表示方式大幅减少了所需的内存和计算开销。同时, RTG-SLAM 只针对新观测到的像素、颜色误差大的像素和深度误差大的像素添加新的高斯分布进行优化, 而不是对所有像素进行优化。它还将高斯分布划分为稳定和 unstable 两类, 只优化不稳定的高斯分布, 大幅降低了优化开销。此外, RTG-SLAM 采用不同的深度和颜色渲染方式, 使单个不透明高斯就可以很好地拟合局部表面区域, 避免了使用多个重叠高斯的需求, 进一步降低了计算和内存开销。这些创新性的设计使得 RTG-SLAM 能够在大规模环境中实现实时高质量的 3D 重建, 在实时性和内存效率方面都有显著优势。

4) Photo-SLAM (Huang 等, 2024) 是一种基于高斯点云的 SLAM 系统, 可以实现单目、双目和 RGB-D 相机的实时定位和高保真建图。它利用 3D 高斯点云作为地图表示, 可以有效地融合几何特征和纹理信息, 实现高保真的场景重建。Photo-SLAM 提出一种基于高斯点云的优化方法, 可以动态地适应关键帧的位姿和深度更新, 从而保证地图的一致性。它结合了基于特征的跟踪和基于密集表示的优化, 在保证实时性的同时也能够实现高精度的定位和建图。该系统还引入了语义信息和全局优化策略, 进一步提高了 SLAM 系统的性能和鲁棒性。

5) Gaussian splatting SLAM (Matsuki 等, 2024) 能够实现在 RGB/RGB-D 设置下持续高保真地重建 3D 场景, 该方法首先提出基于直接优化 3D 高斯分布的相机跟踪方法, 实现了快速稳健的跟踪, 并具有较强的收敛域。其次, 利用高斯分布的显式性质, 引入了几何验证和正则化技术来处理增量式 3D 密集重建中出现的歧义问题。最后, 该方法构建了一个完整的 SLAM 系统, 不仅在新视图合成和轨迹估计方面达到了最先进的水平, 还能重建微小甚至透明物体。

除此之外, 还有一些基于高斯溅射的 RGB/RGB-D SLAM 方法。Gaussian-SLAM (Yugay 等,

2024) 扩展了高斯溅射技术以更好地编码几何形状, 在单目相机设置中超出辐射场的重建, 并提出一种在线学习方法。Compact-3DGS SLAM (Deng 等, 2024) 提出基于滑动窗口的在线掩蔽方法, 消除冗余高斯椭圆, 同时提出基于 codebook 的方法恢复高斯的几何。High-Fidelity SLAM (Sun 等, 2024b) 基于渲染损失的高斯稠密化策略映射未观测区域和细化再观测区域, 引入正则化参数缓解连续映射中的“遗忘”问题。Monocular Gaussian SLAM (Lan 等, 2024) 将全局地图表示为 3D 高斯, 利用基于 CLIP (contrastive language-image pre-training) 特征设计的语言扩展闭环模块进行全局优化, 以校正系统运行中的漂移误差。MotionGS (Guo 等, 2024b) 结合深度视觉特征、双关键帧选择和 3D 高斯, 通过选择性跟踪和联合优化实现精确的位姿估计和紧凑的高斯场景表示。NGM-SLAM (Li 等, 2024b) 首次提出基于神经子图的高斯溅射 SLAM 系统, 通过局部到全局的重建策略和全局循环校正, 实现高保真映射和地图校正。Object-centric 3DGS (Barad 等, 2024) 开发以物体为中心的高效在线框架, 用于增量重建和跟踪未知物体的高保真重建和 6 DoF (degree of freedom) 位姿跟踪。RGBD GS-ICP SLAM (Ha 等, 2024) 融合 G-ICP 和 3D 高斯, 利用比例对齐技术最大限度地减少了冗余计算, 实现了高效系统。Splat-SLAM (Sandström 等, 2024) 利用主动变形 3D 高斯地图动态适应关键帧位姿和深度更新, 结合代理深度和学习到的单目深度, 提升了 3D 重建精度。TAMBRIDGE (Jiang 等, 2024a) 提出结合基于跟踪的 ORB 视觉里程计和基于映射的在线 3D 高斯的 SLAM 框架, 通过 Fusion Bridge 模块在跟踪和映射之间进行联合优化, 以提高大规模场景中的渲染收敛性。CG-SLAM (Hu 等, 2024a) 提出深度不确定性模型以提高跟踪效率和精度, 设计多个损失项构建一致且稳定的 3D 高斯场。MGS-SLAM (Zhu 等, 2024a) 结合稀疏视觉 VO (visual odometry) 和 3D 高斯, 通过稀疏—密集调整环保持稀疏视觉测距和密集高斯地图的尺度一致性, 消除对深度图的依赖。EndoGS SLAM (Wang 等, 2024) 通过集成精简高斯表示法和可微分光栅化技术, 在内窥镜手术中实现超过 100 帧/s 的在线相机跟踪和组织重建渲染速度。

3.2 基于高斯溅射的多模态 SLAM

在前面介绍的方法中, 使用 RGB/RGB-D 作为

SLAM 的输入在室内场景中得到了较好的应用,但是在室外复杂场景中仍然会存在深度不准导致几何变差的问题,特别是随着与相机距离的增加,跟踪精度显著下降,最终导致现有的大部分基于高斯溅射的 SLAM 方法无法应用到落地场景中。为了解决这些缺点,如图 4 所示,基于高斯溅射的多模态 SLAM 通过综合利用多种传感器数据,来实现对复杂环境的精确表征和渲染。

这类方法的核心在于其能够处理并融合来自不同类型传感器的数据(王金科等,2022),如 RGB 图像、深度图像、激光雷达(LiDAR)点云等,从而优化场景的几何结构和语义信息的理解,并提高系统对复杂环境的感知能力。

该技术的关键优势在于其强大的环境感知能力。通过融合多种传感器数据,基于高斯溅射的多模态 SLAM 系统能够更全面地理解环境的几何结构,从而在复杂动态场景中实现更鲁棒的定位和建图。此外,高斯溅射技术的灵活数据融合策略允许系统根据实际应用需求,动态调整不同传感器数据

的权重和处理方式,进一步提升了系统的适应性和灵活性。

在已有的基于高斯溅射的多模态 SLAM 方法中,MM3DGS SLAM (Sun 等,2024a)利用多个摄像头实时进行 3D 重建,通过像素级深度信息和多模态数据提高了重建精度和鲁棒性,与基于深度相机的方法相比降低了硬件成本。Gaussian-LIC (Lang 等,2024)利用 LiDAR 和相机数据,通过 3D 高斯分布建模优化 LiDAR 点云,采用可微分渲染技术实现照片级别的 SLAM 效果,在精度和视觉效果方面超越了基于辐射场的系统。3DGS-ReLoc (Jiang 等,2024b)结合激光雷达和相机数据,利用 2D 体素地图和 KD (K-dimensinal)树优化空间查询,通过归一化交叉相关识别图像对应关系并优化相机姿态,展示了在视觉定位任务上的高效性和准确性。MM-Gaussian (Wu 等,2024)利用固态 LiDAR 解决无边界场景中深度不准确问题,采用 3D 高斯点云和像素级梯度下降技术提高渲染效果和系统鲁棒性,设计了重定位模块确保轨迹恢复。

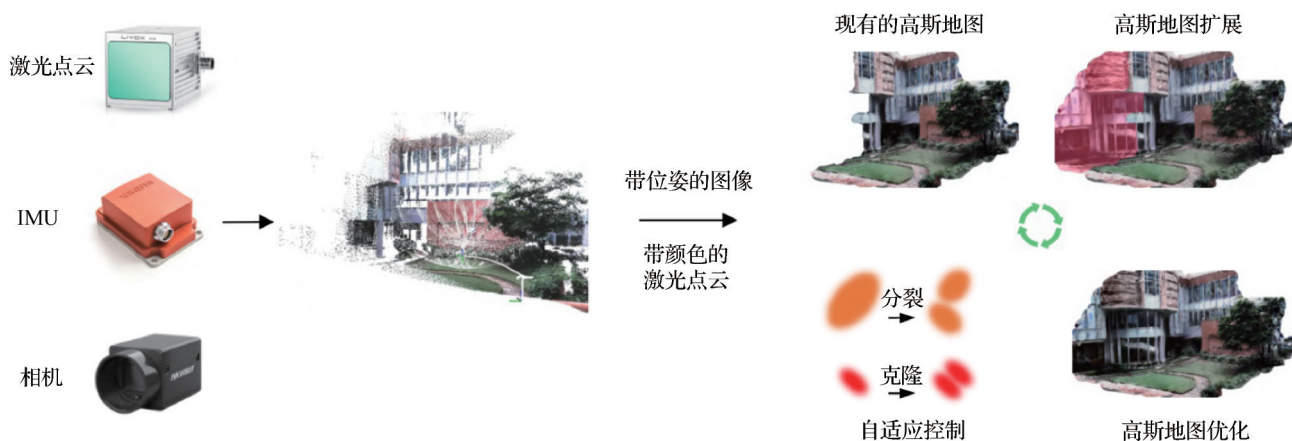


图4 基于高斯溅射的多模态 SLAM 框架(Lang 等,2024)

Fig. 4 The framework of the multimodal SLAM with Gaussian splatting (Lang et al. , 2024)

3.3 基于高斯溅射的语义 SLAM

如图 5 所示,基于高斯溅射的语义 SLAM 技术通过将高斯泼溅与语义信息相结合,不仅实现了场景的精确几何重建和实时渲染,还赋予了系统对环境的深入理解。这种方法利用 3D 高斯来表征场景的几何结构,同时引入语义标签来增强场景的语义信息,使得 SLAM 系统不仅能定位和建图,还能识别和理解环境中的物体和结构。其主要特点包括高效的渲染速度、精确的几何重建、丰富的语义信息,以及对动态环境的适应能力(潘小鸥等,2023),极大地

提升了 SLAM 系统在复杂场景中的应用潜力和实用性。

目前基于高斯溅射的语义 SLAM 方法相对较少,其中 SemGauss-SLAM (Zhu 等,2024b)结合了 3D 高斯表示和语义特征嵌入,提供精确的语义场景重建和定位,在 Replica (Straub 等,2019)和 ScanNet (Dai 等,2017)两个数据集上均表现出色,并展示出了优秀的语义分割和密集语义建图能力,适用于机器人导航和增强现实领域。SGS-SLAM (Li 等,2024c)通过多通道优化整合外观、几何和语义特征,解决了传

统神经隐式SLAM系统在场景理解和物体级几何方面的问题,引入了独特的语义特征损失函数和语义引导的关键帧选择策略,提高了重建质量和稳定性。NEDS-SLAM (Ji 等, 2024)利用3D高斯表示和空间

一致性特征融合模型,减少了预训练分割模型带来的错误估计,提升了重建质量和稳健性,同时采用轻量级编码—解码器网络和3D高斯散射技术,实现了高效新视图渲染和虚拟相机视图修剪方法。

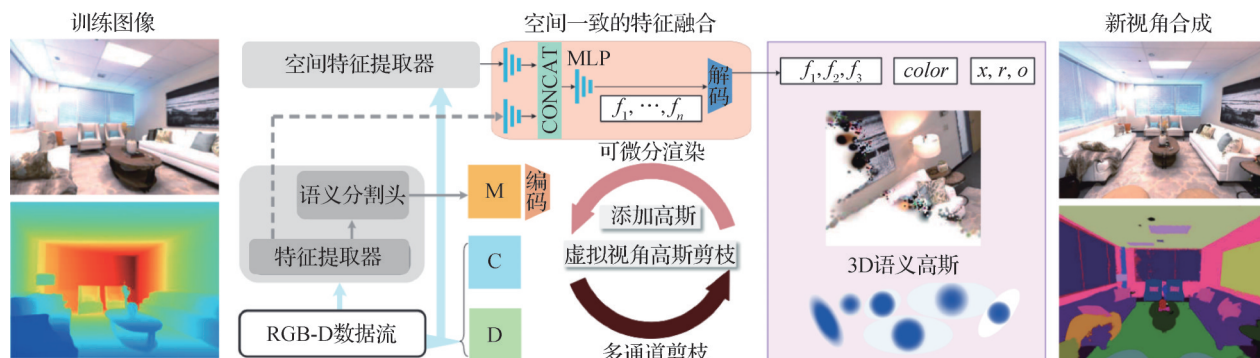


图5 基于高斯溅射的语义SLAM框架(Ji 等, 2024)

Fig. 5 The framework of the semantic SLAM with Gaussian splatting(Ji et al. , 2024)

4 问题与展望

4.1 技术挑战

1)实时性有待提高。许多现代技术在处理实时数据时面临诸多挑战,往往不能达到传感器的帧速率要求。这些挑战主要由于选择的地图数据结构或计算密集型的基于光线的渲染优化所导致,在NeRF风格的SLAM方法中尤为显著。具体而言,采用分层网格的混合方法可以有效减少GPU内存需求,但这种优化往往会牺牲运行时的性能。相反,哈希网格或稀疏体素等高级表示法虽然能够提高计算速度,但需要更多的内存资源支持。尽管这些方法在快速图像渲染方面表现突出,但它们仍然面临着在实时应用场景中同时进行跟踪和映射的挑战。

2)基准一致化。缺乏在线服务器上标准化的基准或明确定义的评估协议,导致评估方法不一致,很难公平比较和确保不同研究论文中方法的一致性。例如,像ScanNet(Dai 等, 2017)数据集面临的挑战显示地面真值姿态来源不一致,引发了对评估结果可靠性和泛化性的关注。此外,使用训练视图来评估渲染性能可能会增加过度拟合特定图像的风险。本文强调需要在SLAM背景下探索评估新视图合成的替代方法,以解决这些问题并提高研究成果的可靠性。

3)大场景的扩展性较差。尽管高斯溅射技术显

著提高了渲染速度,但其在处理室外大场景时,重建质量仍不尽如人意,存在几个关键挑战。首先,由于室外场景的复杂性和多样性,高斯溅射可能难以准确捕捉细节和复杂结构,导致重建结果的精度不高。其次,面对大量数据和复杂几何形状,系统往往需要大量内存以存储和处理信息,这可能限制其在实际应用中的可扩展性和效率。此外,室外环境中的光照变化和遮挡现象也可能影响重建的准确性和稳定性,这需要进一步的研究和技术优化来提升其在复杂环境中的应用性能和实用性。

4)灾难性遗忘。在处理大型场景时灾难性遗忘问题更为显著,现有方法往往表现出一种倾向,即容易忘记先前学习的信息。这主要是因为这些方法通常依赖于单个神经网络或具有固定容量的全局模型,这些模型在优化过程中受到全局变化的影响。对于基于网络的方法而言,常见的解决策略包括使用当前观测值的稀疏射线采样来训练网络,并且通过重播历史数据中的关键帧来保持先前学习内容的记忆。然而,这种策略在大规模增量映射中可能会导致数据的累积增加,需要复杂的重采样过程以提高内存效率。基于网格的方法也面临类似的挑战,尽管已经进行了努力来解决灾难性遗忘问题,但由于空间复杂性的限制,这些方法的可扩展性仍然存在挑战。

4.2 未来研究方向

1)提高实时性渲染速度。高斯溅射直接使用了

CUDA(compute unified device architecture)核函数进行加速渲染,可以通过改进3D高斯的参数化方法和优化其在渲染过程中的处理策略提高渲染速度。同时,可以着眼于优化硬件方面从而提升上层算法的执行速度,通过并行化算法,针对特定硬件的优化,如使用专用的深度学习加速器,可以进一步提高实时性。

2)建立统一的评价基准。包括定义综合性能指标、开发标准化测试集、设计可重复的实验协议、引入实时性和可扩展性评估以及建立开放的评估平台等。

3)提升高斯溅射技术在大场景中的适用性。可从多个方面考虑:开发多尺度和层次化的3D高斯表示方法,允许系统在不同层次上处理场景数据。利用分布式计算和并行处理技术,将大场景的数据处理任务分配到多个计算节点上。开发动态数据管理策略,允许系统根据当前的观察和定位需求,动态地加载和卸载3D高斯数据。

4)引入大模型神经网络。采用多任务学习的方法,训练统一的大模型神经网络同时处理定位、建图、语义分割和物体识别等任务,通过学习场景的全局结构和局部细节,实现高质量的视觉效果,提高资源利用效率和系统的整体性能。

5 结 语

本文立足于SLAM领域,首先,根据SLAM的发展历程,简要梳理了手工特征SLAM和基于学习方法的SLAM。针对学习方法中流行的基于NeRF的SLAM渲染速度慢的痛点,高斯溅射技术的出现在大大加快渲染速度的同时也提高了场景重建的质量,使得SLAM问题迎来了新的热潮。本文从应用类型的角度出发,分别介绍了基于RGB、RGB-D、多模态和语义信息的基于高斯溅射的不同SLAM方法,并深入分析了这些方法的优势与不足。最后,总结了当前基于高斯溅射技术的SLAM方法所面临的挑战,提出未来可能的研究方向。这些研究方向将有助于进一步提升SLAM系统的性能和应用广度,为SLAM技术的发展注入新的动力。

参考文献(References)

Barad K R, Richard A, Dentler J, Olivares-Mendez M and Martinez C.

2024. Object-centric reconstruction and tracking of dynamic unknown objects using 3D Gaussian splatting//Proceedings of 2024 International Conference on Space Robotics. Luxembourg, Luxembourg: IEEE: 202-209 [DOI: 10.1109/iSpaRo60631.2024.10688304]

Bian W J, Wang Z R, Li K J and Bian J W. 2023. NoPe-NeRF: optimising neural radiance field with no pose prior//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 4160-4169 [DOI: 10.1109/CVPR52729.2023.00405]

Bloesch M, Czarnowski J, Clark R, Leutenegger S and Davison A J. 2018. CodeSLAM — learning a compact, optimisable representation for dense visual SLAM//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2560-2568 [DOI: 10.1109/CVPR.2018.00271]

Calonder M, Lepetit V, Strecha C and Fua P. 2010. BRIEF: binary robust independent elementary features//Proceedings of the 11th European Conference on Computer Vision. Crete, Greece: Springer: 778-792 [DOI: 10.1007/978-3-642-15561-1_56]

Chen C H, Wang B, Lu C X, Trigoni N and Markham A. 2020. A survey on deep learning for localization and mapping: towards the age of spatial machine intelligence [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2006.12567.pdf>

Chen X Y L, Läbe T, Milioto A, Röhling T, Behley J and Stachniss C. 2022. OverlapNet: a siamese network for computing LiDAR scan similarity with applications to loop closing and localization. *Autonomous Robots*, 46(1): 61-81 [DOI: 10.1007/s10514-021-09999-0]

Chen X Y L, Milioto A, Palazzolo E, Giguere P, Behley J and Stachniss C. 2019. SuMa++: efficient LiDAR-based semantic SLAM//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China: IEEE: 4530-4537 [DOI: 10.1109/IROS40897.2019.8967704]

Dai A G L, Chang A X, Savva M, Halber M, Funkhouser T and Nießner M. 2017. ScanNet: richly-annotated 3D reconstructions of indoor scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2432-2443 [DOI: 10.1109/CVPR.2017.261]

Deng J Y, Wu Q, Chen X Y L, Xia S P C, Sun Z, Liu G Q, Yu W X and Pei L. 2023. NeRF-LOAM: neural implicit representation for large-scale incremental LiDAR odometry and mapping//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 8184-8193 [DOI: 10.1109/ICCV51070.2023.00755]

Deng T C, Chen Y H, Zhang L Y, Yang J F, Yuan S H, Wang D W, Wang H S and Chen W D. 2024. Compact 3D Gaussian splatting for dense visual SLAM [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2403.11247.pdf>

Derin H, Elliott H, Cristi R and Geman D. 1984. Bayes smoothing algorithms for segmentation of binary images modeled by Markov ran-

- dom fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6 (6) : 707-720 [DOI: 10.1109/tpami.1984.4767595]
- Fischler M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) : 381-395 [DOI: 10.1145/358669.358692]
- Forster C, Pizzoli M and Scaramuzza D. 2014. SVO: fast semi-direct monocular visual odometry//*Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA)*. Hong Kong, China; IEEE: 15-22 [DOI: 10.1109/ICRA.2014.6906584]
- Gao X, Zhang T, Liu Y and Yan Q R. 2017. 14 Lectures on Visual SLAM: from Theory to Practice. Publishing House of Electronics Industry (高翔, 张涛, 刘毅, 颜沁睿. 2017. 视觉SLAM十四讲: 从理论到实践. 北京: 电子工业出版社) [ISBN: 978-7-121-31104-8]
- Ge Y B, Zhang L L, Wu Y X and Hu D W. 2024. PIPO-SLAM: light-weight visual-inertial SLAM with preintegration merging theory and pose-only descriptions of multiple view geometry. *IEEE Transactions on Robotics*, 40: 2046-2059 [DOI: 10.1109/tro. 2024. 3366815]
- Grisetti G, Kümmerle R, Stachniss C and Burgard W. 2010. A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4) : 31-43 [DOI: 10.1109/MITS.2010.939925]
- Guo W Z, Wang B and Chen L J. 2024a. NeuV-SLAM: fast neural multiresolution voxel optimization for RGBD dense SLAM [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2402.02020.pdf>
- Guo X L, Zhang W D, Liu R N, Han P and Chen H T. 2024b. MotionGS: compact Gaussian splatting SLAM by motion filter//*Proceedings of the 7th International Conference on Robotics, Control and Automation Engineering*. Wuhan, China; IEEE: 685-692. [DOI: 10.1109/RCAE62637.2024.10834156]
- Ha S, Yeon J and Yu H. 2024. RGBD GS-ICP SLAM//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 180-197. [DOI: 10.1007/978-3-031-72764-1_11]
- Hamdi A, Melas-Kyriazi L, Mai J J, Qian G C, Liu R S, Vondrick C, Ghanem B and Vedaldi A. 2024. GES: generalized exponential splatting for efficient radiance field rendering//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 19812-19822 [DOI: 10.1109/CVPR52733.2024.01873]
- Hess W, Kohler D, Rapp H and Andor D. 2016. Real-time loop closure in 2D LIDAR SLAM//*Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Sweden; IEEE: 1271-1278 [DOI: 10.1109/ICRA.2016.7487258]
- Hu J R, Chen X H, Feng B Y, Li G L, Yang L J, Bao H J, Zhang G F and Cui Z P. 2024a. CG-SLAM: efficient dense RGB-D SLAM in a consistent uncertainty-aware 3D Gaussian field//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 93-112. [DOI: 10.1007/978-3-031-72698-9_6]
- Hu S K, Hu T and Liu Z W. 2024b. GauHuman: articulated Gaussian splatting from monocular human videos//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 20418-20431 [DOI: 10.1109/CVPR52733. 2024.01930]
- Hua T Y, Bai H T, Cao Z D, Liu M, Tao D C and Wang L. 2024. Hi-MAP: hierarchical factorized radiance field for high-fidelity monocular dense mapping [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2401.03203.pdf>
- Hua T Y, Bai H T, Cao Z D and Wang L. 2023. FMapping: factorized efficient neural field mapping for real-time dense RGB SLAM [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2306.00579.pdf>
- Huang H J, Li L W, Cheng H and Yeung S K. 2024. Photo-SLAM: real-time simultaneous localization and photorealistic mapping for monocular, stereo, and RGB-D cameras//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 21584-21593 [doi: 10.1109/CVPR52733. 2024.02039]
- Isaacson S, Kung P C, Ramanagopal M, Vasudevan R and Skinner K A. 2023. LONER: LiDAR only neural representations for real-time SLAM. *IEEE Robotics and Automation Letters*, 8(12) : 8042-8049 [DOI: 10.1109/LRA.2023.3324521]
- Ji Y M, Liu Y, Xie G H, Ma B Y and Xie Z W. 2024. NEDS-SLAM: a neural explicit dense semantic SLAM framework using 3D Gaussian splatting [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2403.11679.pdf>
- Jiang P, Pandey G and Saripalli S. 2024b. 3DGS-ReLoc: 3D Gaussian splatting for map representation and visual relocalization [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2403.11367.pdf>
- Jiang P F, Liu H, Li X, Wang T, Zhang F B and Buhmann J M. 2024a. TAMBRIDGE: bridging frame-centered tracking and 3D Gaussian splatting for enhanced SLAM [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2405.19614.pdf>
- Jin L R, Chen X Y L, Rückin J and Popović M. 2023. NeU-NBV: next best view planning using uncertainty estimation in image-based neural rendering//*Proceedings of 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Detroit, USA; IEEE: 11305-11312 [DOI: 10.1109/IROS55552.2023.10342226]
- Kalman R E and Bucy R S. 1961. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83 (1) : 95-108 [DOI: 10.1115/1.3658902]
- Keetha N, Karhade J, Jatavallabhula K M, Yang G S, Scherer S, Ramanan D and Luiten J. 2024. SplatAM: splat, track and map 3D Gaussians for dense RGB-D SLAM//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA; IEEE: 21357-21366 [DOI: 10.1109/CVPR52733. 2024.02018]
- Kerbl B, Kopanas G, Leimkühler T and Drettakis G. 2023. 3D Gaussian

- splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4): #139 [DOI: 10.1145/3592433]
- Klein G and Murray D. 2007. Parallel tracking and mapping for small AR workspaces//*Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. Nara, Japan: IEEE: 225-234 [DOI: 10.1109/ISMAR.2007.4538852]
- Lan T, Lin Q W and Wang H Q. 2024. Monocular Gaussian SLAM with language extended loop closure [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2405.13748.pdf>
- Lang X L, Li L J, Wu C M, Zhao C, Liu L, Liu Y, Lyu J J and Zuo X X. 2024. Gaussian-LIC: real-time photo-realistic SLAM with Gaussian splatting and LiDAR-Inertial-Camera fusion [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2404.06926.pdf>
- Leonard J J and Durrant-Whyte H F. 1991. Simultaneous map building and localization for an autonomous mobile robot//*Proceedings of 1991 IEEE/RSJ International Workshop on Intelligent Robots and Systems*. Osaka, Japan: IEEE: 1442-1447 [DOI: 10.1109/IROS.1991.174711]
- Li H, Gu X D, Yuan W H, Yang L W, Dong Z L and Tan P. 2023. Dense RGB SLAM with neural implicit maps [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2301.08930.pdf>
- Li J H, Zhang J W, Bai X, Zheng J, Ning X, Zhou J and Gu L. 2024a. DNGaussian: optimizing sparse-view 3D Gaussian radiance fields with global-local depth normalization//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 20775-20785 [DOI: 10.1109/CVPR52733.2024.01963]
- Li M R, Huang J W, Sun L, Tian A X, Deng T C and Wang H Y. 2024b. NGM-SLAM: Gaussian splatting SLAM with radiance field submap [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2405.05702.pdf>
- Li M R, Liu S H, Zhou H, Zhu G H, Cheng N, Deng T C and Wang H Y. 2024c. SGS-SLAM: semantic Gaussian splatting for neural dense SLAM//*Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 163-179. [DOI: 10.1007/978-3-031-72751-1_10]
- Li S J, Chen X Y L, Liu Y, Dai D X, Stachniss C and Gall J. 2022. Multi-scale interaction for real-time LiDAR data segmentation on an embedded platform. *IEEE Robotics and Automation Letters*, 7(2): 738-745 [DOI: 10.1109/LRA.2021.3132059]
- Li Y Y, Brasch N, Wang Y D, Navab N and Tombari F. 2020. Structure-SLAM: low-drift monocular SLAM in indoor environments. *IEEE Robotics and Automation Letters*, 5(4): 6583-6590 [DOI: 10.1109/LRA.2020.3015456]
- Li Z, Chen Z, Li Z and Xu Y. 2024d. Spacetime Gaussian feature splatting for real-time dynamic view synthesis//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 8508-8520 [DOI: 10.1109/CVPR52733.2024.00813]
- Liu R J, Wang X S, Zhang C and Zhang B H. 2020. A survey on visual SLAM based on deep learning. *Journal of System Simulation*, 32(7): 1244-1256 (刘瑞军, 王向上, 张晨, 章博华. 2020. 基于深度学习的视觉SLAM综述. *系统仿真学报*, 32(7): 1244-1256) [DOI: 10.16182/j.issn1004731x.joss.19-VR0466]
- Long X X, Cheng X J, Zhu H, Zhang P J, Liu H M, Li J, Zheng L T, Hu Q Y, Liu H, Cao X, Yang R G, Wu Y H, Zhang G F, Liu Y B, Xu K, Guo Y L and Chen B Q. 2021. Recent progress in 3D vision. *Journal of Image and Graphics*, 26(6): 1389-1428 (龙霄潇, 程新景, 朱昊, 张朋举, 刘浩敏, 李俊, 郑林涛, 胡庆拥, 刘浩, 曹汛, 杨睿刚, 吴毅红, 章国锋, 刘焯斌, 徐凯, 郭裕兰, 陈宝权. 2021. 三维视觉前沿进展. *中国图象图形学报*, 26(6): 1389-1428) [DOI: 10.11834/jig.210043]
- Matsuki H, Murai R, Kelly P H J and Davison A J. 2024. Gaussian splatting SLAM//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 18039-18048 [DOI: 10.1109/CVPR52733.2024.01708]
- Ma Y Y, Ye Z H, Liu K H and Chen L. 2021. Event-based visual localization and mapping algorithms: a survey. *Acta Automatica Sinica*, 47(7): 1484-1494 (马艳阳、叶梓豪、刘坤华、陈龙. 2021. 基于事件相机的定位与建图算法: 综述. *自动化学报*, 47(7): 1484-1494) [DOI: 10.16383/j.aas.c190550]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99-106 [DOI: 10.1145/3503250]
- Montemerlo M, Thrun S, Koller D and Wegbreit B. 2002. FastSLAM: a factored solution to the simultaneous localization and mapping problem//*Proceedings of the 18th National Conference on Artificial Intelligence*. Edmonton, Canada: AAAI: 593-598 [DOI: 10.1007/s00244-005-7058-x]
- Mur-Artal R, Montiel J M M and Tardos J D. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5): 1147-1163 [DOI: 10.1109/TRO.2015.2463671]
- Niedermayr S, Stumpfegger J and Westermann R. 2024. Compressed 3D Gaussian splatting for accelerated novel view synthesis//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 10349-10358 [DOI: 10.1109/CVPR52733.2024.00985]
- Pan X K, Liu H M, Fang M, Wang Z, Zhang Y and Zhang G F. 2023. Dynamic 3D scenario-oriented monocular SLAM based on semantic probability prediction. *Journal of Image and Graphics*, 28(7): 2151-2166 (潘小鹏, 刘浩敏, 方铭, 王政, 张涌, 章国锋. 2023. 基于语义概率预测的动态场景单目视觉SLAM. *中国图象图形学报*, 28(7): 2151-2166) [DOI: 10.11834/jig.210632]
- Pan Y, Zhong X G, Wiesmann L, Posewsky T, Behley J and Stachniss C. 2024. PIN-SLAM: LiDAR SLAM using a point-based implicit neural representation for achieving global map consistency. *IEEE Transactions on Robotics*, 40: 4045-4064 [DOI: 10.1109/TRO.

- 2024.3422055]
- Peng Z X, Shao T J, Liu Y, Zhou J K, Yang Y, Wang J D and Zhou K. 2024. RTG-SLAM: real-time 3D reconstruction at scale using Gaussian Splatting//Proceedings of 2024 ACM SIGGRAPH Conference Papers. Colorado, USA: ACM: #30 [DOI: 10.1145/3641519.3657455]
- Qin T, Li P L and Shen S J. 2018. VINS-Mono: a robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4): 1004-1020 [DOI: 10.1109/TRO.2018.2853729]
- Rosten E and Drummond T. 2006. Machine learning for high-speed corner detection//Proceedings of the 9th European Conference on Computer Vision. Graz, Austria: Springer: 430-443 [DOI: 10.1007/11744023_34]
- Rublee E, Rabaud V, Konolige K and Bradski G. 2011. ORB: an efficient alternative to SIFT or SURF//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 2564-2571 [DOI: 10.1109/ICCV.2011.6126544]
- Sandström E, Tateno K, Oechsle M, Niemeyer M, Van Gool L, Oswald M R and Tombari F. 2024. Splat-SLAM: globally optimized RGB-only SLAM with 3D Gaussians [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2405.16544.pdf>
- Simmons R and Koenig S. 1995. Probabilistic robot navigation in partially observable environments//Proceedings of the 14th International Joint Conference on Artificial Intelligence — Volume 2. Montreal, Canada: Morgan Kaufmann Publishers Inc.: 1080-1087
- Smith R C and Cheeseman P. 1986. On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research*, 5(4): 56-68 [DOI: 10.1177/027836498600500404]
- Song X W, Zheng J, Yuan S R, Gao H A, Zhao J W, He X, Gu W H and Zhao H. 2024. SA-GS: scale-adaptive Gaussian splatting for training-free anti-aliasing [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2403.19615.pdf>
- Straub J, Whelan T, Ma L N, Chen Y F, Wijmans E, Green S, Engel J J, Mur-Artal R, Ren C, Verma S, Clarkson A, Yan M F, Budge B, Yan Y J, Pan X Q, Yon J, Zou Y Y, Leon K, Carter N, Briales J, Gillingham T, Mueggler E, Pesqueira L, Savva M, Batra D, Strasdat H M, De Nardi R, Goesele M, Lovegrove S and Newcombe R. 2019. The replica dataset: a digital replica of indoor spaces [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/1906.05797.pdf>
- Sucar E, Liu S K, Ortiz J and Davison A J. 2021. iMAP: implicit mapping and positioning in real-time//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 6209-6218 [DOI: 10.1109/ICCV48922.2021.00617]
- Sun L C, Bhatt N P, Liu J C, Fan Z W, Wang Z Y, Humphreys T E and Topcu U. 2024a. MM3DGS SLAM: multi-modal 3D Gaussian splatting for SLAM using vision, depth, and inertial measurements//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE: 10159-10166. [DOI: 10.1109/iros58592.2024.10802389]
- Sun S, Mielle M, Lilienthal A J and Magnusson M. 2024b. High-fidelity SLAM using Gaussian splatting with rendering-guided densification and regularized optimization//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE: 10476-10482. [DOI: 10.1109/iros58592.2024.10802373]
- Tan Z, Zhou Z T, Ge Y B, Wang Z, Chen X Y L and Hu D W. 2024. TD-NeRF: novel truncated depth prior for joint camera pose and neural radiance field optimization//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE: 372-379 [DOI: 10.1109/iros58592.2024.10802634]
- Tateno K, Tombari F, Laina I and Navab N. 2017. CNN-SLAM: real-time dense monocular SLAM with learned depth prediction//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6565-6574 [DOI: 10.1109/CVPR.2017.695]
- Teigen A L, Park Y, Stahl A and Mester R. 2024. RGB-D mapping and tracking in a plenoxel radiance field//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 3330-3339 [DOI: 10.1109/WACV57701.2024.00331]
- Thrun S. 2002. Particle filters in robotics//Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence. Alberta, Canada: Morgan Kaufmann Publishers Inc.: 511-518
- Tosi F, Zhang Y M, Gong Z R, Sandström E, Mattoccia S, Oswald M R and Poggi M. 2024. How NeRFs and 3D Gaussian splatting are reshaping SLAM: a survey [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2402.13255.pdf>
- Wang J K, Zuo X X, Zhao X R, Lü J J and Liu Y. 2022. Review of multi-source fusion SLAM: current status and challenges. *Journal of Image and Graphics*, 27(2): 368-389 (王金科, 左星星, 赵祥瑞, 吕佳俊, 刘勇. 2022. 多源融合SLAM的现状与挑战. 中国图象图形学报, 27(2): 368-389) [DOI: 10.11834/jig.210547]
- Wang K L, Yang C, Wang Y H, Li S K, Wang Y, Dou Q, Yang X K and Shen W. 2024. EndoGSLAM: real-time dense reconstruction and tracking in endoscopic surgeries using Gaussian splatting//Proceedings of the 27th International Conference on Medical Image Computing and Computer-Assisted Intervention. Marrakesh, Morocco: Springer: 219-229. [DOI: 10.1007/978-3-031-72089-5_21]
- Wang Z R, Wu S Z, Xie W D, Chen M and Prisacariu V A. 2022. NeRF--: neural radiance fields without known camera parameters [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2102.07064.pdf>
- Wu C Y, Duan Y F, Zhang X R, Sheng Y, Ji J M and Zhang Y Y. 2024. MM-Gaussian: 3D Gaussian-based multi-modal fusion for localization and reconstruction in unbounded scenes//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE: 12287-12293. [DOI: DOI: 10.1109/IROS58592.2024.10801605]

- Yan C, Qu D L, Xu D, Zhao B, Wang Z G, Wang D and Li X L. 2024. GS-SLAM: dense visual SLAM with 3D Gaussian splatting//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 19595-19604 [doi: 10.1109/CVPR52733.2024.01853]
- Yang Z Y, Gao X Y, Zhou W, Jiao S H, Zhang Y Q and Jin X G. 2024. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 20331-20341 [DOI: 10.1109/CVPR52733.2024.01922]
- Yugay V, Li Y, Gevers T and Oswald M R. 2024. Gaussian-SLAM: photo-realistic dense SLAM with Gaussian splatting [EB/OL]. [2024-07-18]. <https://arxiv.org/pdf/2312.10070.pdf>
- Zhu P C, Zhuang Y M, Chen B Q, Li L, Wu C D and Liu Z L. 2024a. MGS-SLAM: monocular sparse tracking and Gaussian mapping with depth smooth regularization. IEEE Robotics and Automation Letters, 9: 9486-9493. [DOI:10.1109/LRA.2024.3460431]
- Zhu S T, Qin R J, Wang G M, Liu J M and Wang H S. 2024b. SemGauss-SLAM: dense semantic Gaussian splatting SLAM [EB/

OL]. [2024-07-18]. <https://arxiv.org/pdf/2403.07494.pdf>

- Zhu Z H, Peng S Y, Larsson V, Xu W W, Bao H J, Cui Z P, Oswald M R and Pollefeys M. 2022. NICE-SLAM: neural implicit scalable encoding for SLAM//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12776-12786 [DOI: 10.1109/CVPR52688.2022.01245]

作者简介

谭臻,男,博士研究生,主要研究方向为三维重建和SLAM。

E-mail: tanzhen1996@nudt.edu.cn

胡德文,通信作者,男,教授,主要研究方向为模式识别、类脑智能和视觉认知。E-mail: dwhu@nudt.edu.cn

牛中颜,男,硕士研究生,主要研究方向为三维重建和视觉定位。E-mail: niuzhongyan@nudt.edu.cn

张津浦,男,讲师,主要研究方向为计算机视觉和模式识别。E-mail: zhangjinpu@nudt.edu.cn

陈谢沅澧,男,副教授,主要研究方向为视觉和激光SLAM。

E-mail: xieyuanli.chen@nudt.edu.cn