

大作业分数：

指导教师：

# 哈尔滨理工大学大作业

题 目：基于 3D 高斯溅射的 3 维重建技术

课程名称：\_\_\_\_\_数字图像处理\_\_\_\_\_

院（系）：\_\_\_\_\_自动化学院\_\_\_\_\_

班 级：\_\_\_\_\_

姓 名：\_\_\_\_\_

学 号：\_\_\_\_\_

年 月 日

# 基于 3D 高斯溅射的 3 维重建技术

## 1. 国外研究现状

3 维重建指的是给定一个场景的多个视角图像，重建出这个场景的 3 维模型，是机器人感知模块中的关键技术之一。传统的 3 维重建方法发展比较成熟，但传统重建方法在面对复杂场景时效率低且精度差，而神经辐射场（NeRF）<sup>[1]</sup> 的出现，给 3 维重建领域带来了革新和生命力。

NeRF 的成功实质上是深度学习在 3 维重建领域的成功。但是使用深度神经网络来建模一个连续体场景，对场景中的细节进行充分还原，实现逼真的视觉效果，需要付出巨大代价。

在此背景下，3 维高斯溅射（3D GS）技术<sup>[2]</sup> 应运而生。3D GS 技术改变了 NeRF 技术的底层范式，彻底颠覆了 NeRF 的整个架构，从场景表示和渲染方式上作出了革新。另外，3D GS 这种显式场景表达方式提高了场景的可编辑性，在 3 维重建和新视图合成领域达到了新的高度，基于 3D GS 的改进技术和下游应用也层出不穷。

## 2. 3D GS 技术框架概述

本节主要从 3D GS 的核心技术原理出发，凝练出 3D GS 的整体架构，对 3D GS 的各个模块进行必要的解释和分析，并指出 3D GS 技术相较于 NeRF 技术的优势所在<sup>[3]</sup>。

图 1 展示了 3D GS 技术的整体架构，主要分为正向误差计算和反向梯度传播 2 个数据流，包含 6 个模块：SfM 点云初始化、3D

高斯椭球集初始化、3D 椭球参数投影、光栅化图像渲染、损失计算以及自适应密度控制，整体流程就是正向计算和反向优化的过程。

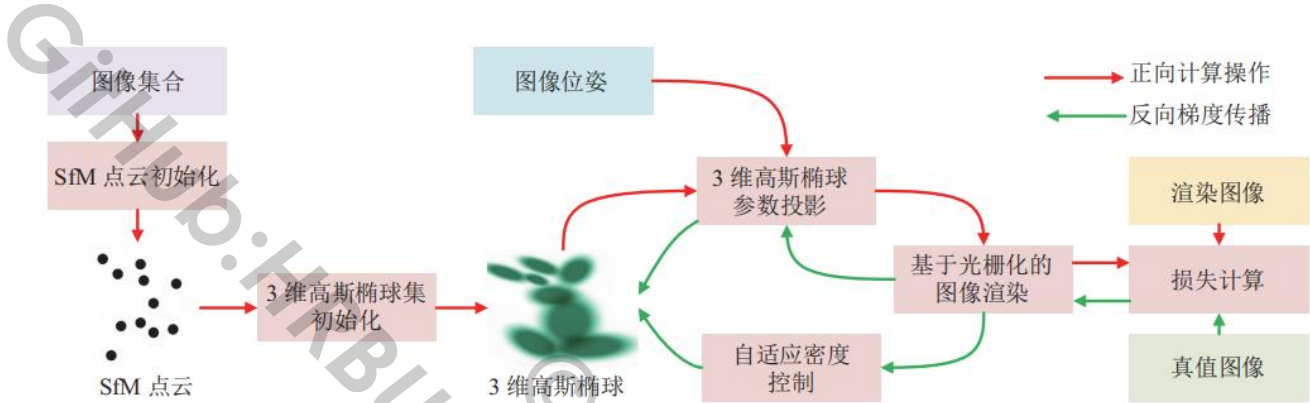


图 1 3D 高斯溅射技术整体架构

NeRF 是一种典型的隐式辐射场技术，主要使用隐式的基于坐标的模型将空间坐标映射到像素值，核心思想是用一个多层感知机（MLP）<sup>[4]</sup> 来表示一个连续的场景，输入为空间点的 3D 坐标和观察方向，输出是空间点的颜色和密度信息。NeRF 的渲染过程是基于体渲染（volume rendering）的，它将每条从相机发出的射线上的点的颜色和密度进行加权求和，得到该射线的最终颜色。NeRF 的训练过程是通过最小化渲染结果和真实图像之间的 L2 损失来优化 MLP 的参数。

而 3D GS 是一种典型的显式辐射场技术<sup>[5]</sup>，通过数以万计的 3D 高斯基元这样的显式场景表示，填充整个需要重建的场景，再通过并行的分块光栅化算法，实现高效的图像渲染效果。3D GS 技术的训练过程是通过最小化图像之间 L1 损失和 SSIM 损失来优化 3D 高斯基元的参数。比起 NeRF，3D GS 技术不仅可以实现场景的实时渲染，还提高了场景的可编辑性。

### 3. 基于 3D GS 的视觉 SLAM

视觉 SLAM 是机器人利用视觉传感器（单目相机、双目相机、RGB-D 相机）在未知环境中进行同步定位和地图构建的一种技术，是机器人在未知环境中进行探索、导航和交互的必要技术<sup>[6]</sup>。视觉 SLAM 技术可以感知场景中丰富的纹理、结构和语义信息，凭借其鲁棒的定位能力和准确的建图效果，在室内服务机器人上应用广泛。视觉 SLAM 的核心在于创建环境地图，因此，完整准确地表达环境、构建高精度的稠密地图是视觉 SLAM 的关键问题。

传统视觉 SLAM 技术根据使用图像信息的不同，可以分为直接法和间接法。直接法指的是利用图像上每个像素的信息来估计相机位姿，基本原理是光度一致性，通过最小化光度误差来实现跟踪<sup>[7]</sup>。直接法在纹理较差的部分比间接法更鲁棒，但当场景中的光照变化后，直接法容易失效。间接法常用于稀疏视觉 SLAM 算法，只使用图像中显著的特征进行位姿估计，首先检测角点特征并计算包含方向和尺度信息的特征描述向量，然后通过特征匹配和最小化帧间重投影误差进行跟踪。

随着 NeRF 的提出，场景的 3D 重建实现了非常高的精度和逼真的视觉效果<sup>[8]</sup>。因此，许多研究人员尝试将 NeRF 方法应用到 SLAM 算法中。基于 NeRF 的 SLAM 算法可以分为 3 种主要类型：基于 MLP 的方法、混合表示方法和显式方法。基于 MLP 的 iMAP<sup>[9]</sup> 算法提供了可扩展且内存负担小的地图表示法，但在较大场景中面临灾难性遗忘的挑战。

3D GS 技术的出现解决了 NeRF 技术训练和渲染速度慢的缺

点，它使用 3D 高斯基元作为实时神经渲染的基元，可在不牺牲图像质量的情况下显著提高渲染速度。因此，很多研究人员转向基于 3D GS 的视觉 SLAM 算法研究。

### 4. 基于 3D GS 的动态场景重建

动态场景重建是一种利用多个视频帧数据重建出场景 3D 动态模型的技术，也是支撑机器人众多应用的关键技术。动态场景是机器人进行导航规划和交互的主要挑战，动态场景的感知和建模对于机器人是至关重要的，有助于提高机器人对周围环境的理解和反应能力。

3D GS 技术的出现重构了 3D 重建和新视图合成领域的底层架构，显式表达和可微分渲染的优势使得 3D GS 技术可以在不降低重建质量的情况下实现实时渲染。但是 3D GS 技术专注于静态场景重建，直接将其应用到动态场景时效果很差。因此，4D GS 技术应运而生，它将 3D GS 技术在时序上进行扩展和改进，为 3D 高斯基元增加了适应场景变形和运动的属性。

动态场景重建和一般的静态场景重建的区别在于动态场景中动态物体在图像中占据较大的画面且其运动变化复杂，而且运动和变形在时序上难以捕捉。因此，动态场景重建技术的核心在于对动态场景中的运动和变形规律进行建模。NeRF 技术凭借其隐式辐射场高质量的渲染能力，在一定程度上推动了动态场景重建技术的发展，也出现了一系列优秀的基于 NeRF 的动态场景重建方法：DNeRF<sup>[10]</sup>、HyperNeRF、Nerfies、TiNeuVoxB、FDNeRF、K-Planes、

HyperReel、NeRF-DS、HexPlane、NeRFPlayer 等。上述基于 NeRF 的方法虽然取得了一定的进展,但是 NeRF 技术的训练时间和渲染速度使得动态场景实时重建仍然存在困难和挑战。

### 5. 基于 3D GS 的生成式 AI

AIGC (人工智能生成内容) 指的是使用人工智能技术来自动创建内容,包括文本、图像、视频、音乐等多种形式<sup>[11]</sup>。随着大模型的发展,AIGC 在机器人领域的应用日益广泛,例如增强机器人交互能力、提升机器人自主性、优化机器人操作效率等。其中比较热门的应用是文本到 3D 场景的生成模型,用户描述一个场景,AI 则可以根据其描述生成相应的 3D 模型。

当前文本到 3D 模型的生成工作大多是基于分数蒸馏采样的方法和基于扩散模型的方法。但是这些方法通常存在渲染过程缓慢、训练和优化时间过长的的问题,而且其生成的模型细节不足,整体性较差。3D GS 技术凭借显式表达法提升了场景的可编辑性,同时保持了实时渲染的优势,因此,基于 3DGS 的文本到 3D 模型的生成应用随之出现,为场景生成和场景编辑领域注入了新的活力。

许多研究者着力于研究基于 3DGS 的场景生成<sup>[12]</sup>。文[13]是对传统的 SDS (score distillation sampling) 方法的改进。例如,文[13]提出了一种新颖的文本到 3D 场景的生成方法,通过结合 3D GS 技术和分阶段优化策略,有效生成具有精细细节和准确几何形状的 3D 模型。图 2 展示了 GSGEN 算法的场景生成效果。文[14]专注于人体生成,提出了一种新颖的结构感知 SDS 算法,可以生成具有细粒

度几何形状和逼真外观的高质量 3D 人体。



图 2 GSGEN 的场景生成效果

## 6. 基于 3D GS 的自动驾驶

自动驾驶又称无人驾驶，是指通过计算机系统、传感器和人工智能技术，使车辆能够在没有人类驾驶员干预的情况下自动行驶。当前自动驾驶的核心功能之一是识别和感知周围的环境，包括实时准确地重建周围场景（静态背景和动态物体），这对于车辆的导航、决策、控制以及安全都是至关重要的。

NeRF 技术虽然可以实现逼真的视图合成和 3D 重建效果，但训练和渲染速度太慢，与自动驾驶所需的实时性相悖。因此，3D GS 技术的出现可以解决实时渲染的问题，但是由于自动驾驶场景非常复杂，规模也非常大，对其进行实时重建很有挑战，需要大量计算资源，因此当前非常少的研究使用 3D GS 进行自动驾驶场景的建模。

## 7. 未来发展趋势

### 7.1 受限视角下的 3D GS 研究

当前 3D GS 在训练前需要高质量的场景初始 3D 点云先验信息，没有这些信息会出现训练时间变长和效果变差的问题。但是在实

际应用中，视角受限的场景是很常见的，无法为 3D GS 提供多视图先验 3D 点云。因此，研究这种视角受限甚至是单视图条件下的 3D GS 技术是必要的。

### 7.2 内存紧张条件下的 3D GS 研究

3D GS 采用数以万计的 3D 高斯基元填充场景，是显式的表征，比起 NeRF 使用的深度神经网络，参数量比较大并且有很多参数冗余，从而导致内存的占用量也比较大。因此，在内存紧张的环境下优化高斯基元的表征和渲染的密度，在保证精度的同时紧凑地表达场景是未来的一个研究方向。

### 7.3 3D GS 和语义分割结合

3D GS 的可编辑性为其进行场景编辑提供了可能。精准地操控编辑区域、实现比较好的场景一致性效果是机器人对场景进行编辑所需要的。语义分割是机器人感知模块中的一项重要技术，它的目标是将图像分割成若干区域，并为每个区域赋予一个类别标签。因此，使用语义分割作为先验信息，融合 3D GS，可以为机器人场景语义建图和编辑提供助力。

### 7.4 3D GS 和机械臂抓取结合

机械臂抓取任务是机器人应用中的一种动态任务，需要在抓取过程中不断动态获取局部的场景信息。因此，如何快速更新局部场景表达，从而执行准确的操作是机械臂抓取中重要的一环。而 3D GS 凭借其密度控制和协方差优化的性质，对场景重建有很强的适应能力，通过 3D 高斯基元的扩展和优化，可以快速有效地对场景进行重新



渲染。

### 8.总结

3D GS 技术的出现,颠覆了 NeRF 技术在新视图合成和 3D 重建方面的底层架构,3D GS 技术凭借其显式辐射场的场景表征以及分块并行的高效光栅化,可以在不降低重建质量的前提下实现场景的实时渲染。此外,3D GS 技术的训练时间非常短,3D 高斯基元表征也提高了场景的可编辑性。

然而,3D GS 技术本身在视角受限和内存消耗方面还有待改进,并且当前基于 3D GS 的应用中,一部分展示意义大于实际价值,一部分的扩展性较差,距离实际应用落地还有一段距离,仍然存在很大挑战,需要后续研究者们进一步解决。另外,机器人与人的协作共融是机器人领域的热点方向,旨在使得机器人理解和预测人类行为,从而更好地与人协作,这对机器人场景感知和场景建模的精度和实时性要求比较高。而 3D GS 技术兼顾了精度和实时性,因此,3D GS 技术在机器人与人的协作共融方面具有发掘潜力。通过本文的梳理和展望,希望可以为研究者们进一步改进 3D GS 技术,为 3D GS 在机器人领域的应用和创新提供帮助。

### 参 考 文 献

- [1] Mildenhall B, Srinivasan P P, Tancik M, et al. Representing scenes as neural radiance fields for view synthesis., 2021, 65[J]. DOI: <https://doi.org/10.1145/3503250>, 99-106.
- [2] Kerbl B, Kopanas G, Leimkühler T, et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering[J]. ACM Trans. Graph., 2023, 42(4): 139:1-139:14.
- [3] 李梦雨.基于能量动态反馈的三维多焦点调控方法研究[D].湖北省: 湖北工业大学,2023.
- [4] Olmedo M C, Paegelow M, Mas J F, et al. Geomatic approaches for modeling land change scenarios. An introduction[J]. Geomatic Approaches for Modeling Land Change Scenarios, 2018: 1-8.
- [5] 齐帅.基于场景几何结构的三维光场内容生成关键技术研究[D].北京市:北京邮电大学,2022.
- [6] KAZEROUNI I A, FITZGERALD L, DOOLY G, et al. A survey of state-of-the-art on visual SLAM[J]. Expert Systems with Applications, 2022, 205. DOI: 10.1016/j.eswa.2022.117734.
- [7] 高志奇,王雪香,黄平平,等.一种改进的稀疏恢复直接数据域 STAP 方法[J].信号处理,2023,39(06)
- [8] 郑晓华,耿鑫雷,邓浩坤.基于深度学习的室内动态场景下视觉 SLAM 技术研究[J].测绘地理信息,2024,49(02):51-55.

- [9] 陈壮,俞一彪.噪声自适应拟合补偿的鲁棒性声纹识别算法[J].声学学报,2022,47(01):151-160.
- [10] PUMAROLA A, CORONA E, PONS-MOLL G, et al.  
D-NeRF:Neural radiance fields for dynamic scenes[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition.  
Piscataway, USA: IEEE, 2021: 10313-10322.
- [11] 贺异欣,张新长,吴福成,等.实景三维下的 AIGC 变形监测算法分析[J].测绘科学,2023,48(11):211-217.
- [12] 侯礼杰,沈寅松,刘晓晨,等.三维高斯溅射超分辨视觉场景构建算法[J].中国测试,2024,50(09):13-20.
- [13] TANG J X, REN J W, ZHOU H, et al. DreamGaussian: Generative Gaussian splatting for efficient 3D content creation[DB/OL].  
[2024-04-01].
- [14] 谭巨全,王然.特征融合下田径录像 3D 人体动作 DTW 捕捉算法[J].计算机工程,2024,50(07):71-78.