

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2025)07-2514-14

论文引用格式: Lyu H Y, Liu Y L, Li H, Yuan X and Xing G Y. 2025. Multi-view intrinsic decomposition of indoor scenes under a 3D Gaussian splatting framework. Journal of Image and Graphics, 30(7): 2514-2527 (吕恒烨, 刘艳丽, 李宏, 袁霞, 邢冠宇. 2025. 三维高斯泼溅框架下的室内场景多视图本征分解. 中国图象图形学报, 30(7): 2514-2527) [DOI: 10. 11834/jig. 240505]

三维高斯泼溅框架下的室内场景多视图本征分解

吕恒烨¹, 刘艳丽¹, 李宏², 袁霞³, 邢冠宇^{4*}

1. 四川大学计算机学院, 成都 610065; 2. 四川大学视觉合成图形图像技术国家级重点实验室, 成都 610065;
3. 成都信息工程大学计算机学院, 成都 610225; 4. 四川大学网络空间安全学院, 成都 610065

摘要: 目的 三维场景本征分解尝试将场景分解为反射率与照明的乘积, 其分解结果可用于虚拟物体插入、图像材质编辑、重光照等任务, 受到广泛关注与研究。但是, 分解规模较大且布局复杂的室内场景是一个高度病态的问题, 获得正确的分解结果具有较高的挑战性。本文基于当前最先进的辐射场表示技术三维高斯泼溅, 提出一种针对室内场景的本征分解算法, 大大提高了室内场景本征分解的精度和效率。**方法** 为了更好地解耦本征属性, 本文基于三维高斯泼溅技术设计了一种针对室内场景的本征分解模型, 将场景分解为反射率及偏移、照明和残差项, 并引入新的反射率分段稀疏、照明平滑与色度先验约束, 以减少分解过程中的歧义, 保证分解结果的合理性。同时还利用捕获的深度数据增强场景的几何信息, 有利于反射率和照明能够更好的解耦, 提高合成图像的质量。**结果** 对来自合成数据集 Replica 的 8 个场景和真实数据集 ScanNet++ 的 5 个场景进行实验, 并对分解结果进行可视化; 同时测试了新视角下合成图像的峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity index, SSIM) 和学习感知图像块相似度 (learned perceptual image patch similarity, LPIPS) 指标。结果显示, 本文方法不仅可以得到在视觉上更加合理的分解结果, 并且合成图像的上述指标在 Replica 数据集上平均达到 34.695 5 dB、0.965 4 和 0.086 1, 在 ScanNet++ 数据集上平均达到 27.949 6 dB、0.895 0 和 0.144 4, 优于对比的三维场景本征分解算法。**结论** 与以往工作相比, 本文方法能够快速分解室内场景, 并支持在新视角下推理场景属性和合成高保真的图像, 具有较高的应用价值。

关键词: 本征分解; 三维高斯泼溅 (3DGS); 室内场景分解; Retinex 理论; 辐射场

Multi-view intrinsic decomposition of indoor scenes under a 3D Gaussian splatting framework

Lyu Hengye¹, Liu Yanli¹, Li Hong², Yuan Xia³, Xing Guanyu^{4*}

1. College of Computer Science, Sichuan University, Chengdu 610065, China; 2. National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China; 3. College of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China; 4. School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, China

Abstract: Objective The intrinsic decomposition of 3D scenes involves breaking down a scene into a product of reflectance and shading. Decomposition results can be applied to various tasks, such as virtual object insertion, image material replacement, and relighting, and thus, they have received widespread attention and have been the subject of numerous

收稿日期: 2024-09-24; 修回日期: 2024-11-20; 预印本日期: 2024-11-27

* 通信作者: 邢冠宇 xingguanyu@scu.edu.cn

基金项目: 国家自然科学基金项目 (62172290, 61972271); 四川省重点研发计划资助 (2023YFS0454)

Supported by: National Natural Science Foundation of China (62172290, 61972271); Key R&D Program of Sichuan Province, China (2023YFS0454)

researches. However, decomposing large-scale and intricately structured indoor scenes presents a highly ill-posed problem due to the complexity of indoor environments, which contain various light sources, materials, and geometries. This complexity makes accurately disentangling reflectance and shading difficult because multiple possible combinations of these factors can explain the same observed image. Traditional methods frequently use prior constraints based on physical knowledge, human intuition, and visual perception to facilitate the reasonable separation of intrinsic attributes. Despite such effort, traditional methods still struggle in handling the complex interaction between reflectance and shading, resulting in poor decomposition results and limited applicability to different scenarios. Therefore, to achieve accurate and consistent decomposition results, addressing the inherent ambiguities and uncertainties involved is essential. **Method** To address these challenges, this study proposes an intrinsic decomposition algorithm that is specifically designed for indoor scenes, leveraging the state-of-the-art radiance field representation technique called 3D Gaussian splatting. This approach significantly improves the accuracy and efficiency of intrinsic decomposition for complex indoor environments. At present, many studies have further improved the rendering quality and efficiency of 3D Gaussian splatting and applied it to inverse rendering, dynamic scene modeling, and other tasks. Such development has inspired us to use 3D Gaussian splatting as a proxy for scene representation in this work. Building on this foundation, we develop an intrinsic decomposition model that is tailored for indoor scenes. This model is based on Retinex theory and aims to decompose a scene into reflectance, which represents the albedo of objects, and shading, which represents the interaction between lighting and object geometry. However, decomposing a scene into reflectance and shading is insufficient for representing all types of objects in indoor scenes. We use residual terms to fit specular reflections because indoor scenes have specular materials. To prevent loss of texture in reflectance during the optimization process, we use a reflectance offset to capture these details. The four aforementioned components effectively represent most objects in indoor scenes and facilitate the decomposition and optimization of intrinsic scene attributes. To better separate the above attributes, we introduce new reflectance piecewise sparsity, shading smoothness, and chromaticity prior constraints to reduce ambiguity in the decomposition process and ensure the rationality of the results. In accordance with the assumptions of Retinex theory, changes in gradient within a scene are attributed to illumination-independent reflectance and shading. Reflectance refers to the inherent color of an object and is the primary factor that is responsible for significant changes in appearance. Reflectance exhibits sparsity, because it tends to cluster in the RGB color space, forming distinct groups. By contrast, shading, which is influenced by the slow variation of light intensity in a scene, is generally assumed to be smooth. Directly separating reflectance from shading is challenging, and thus, we adopt methods from other computer vision fields. We segment a scene piecewise, apply sparsity constraints to each reflectance piece, constrain shading to be smoother in flat areas, and use chromaticity priors to ensure the color consistency of reflectance. These constraints effectively ensure the rationality of the decomposition results. In addition to these novel constraints, we enhance the geometric information of a scene by incorporating captured depth data, which provide valuable cues about a scene's structure. This improved understanding allows for a more accurate decoupling of reflectance and shading, ultimately leading to higher-quality synthesized images. **Result** We conducted experiments on eight scenes from the synthetic dataset Replica and five scenes from the real-world dataset ScanNet++. The decomposition results were visualized in the main text, and we evaluated the quality of synthesized images from novel views by using common image quality metrics, including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS). The results demonstrate that our method not only produces more visually plausible decomposition results but also outperforms previous 3D scene intrinsic decomposition algorithms in terms of quantitative metrics. In particular, our method achieved average PSNR, SSIM, and LPIPS values of 34.695 5 dB, 0.965 4, and 0.086 1, respectively, on the Replica dataset. On the ScanNet++ dataset, our method achieved average values of 27.949 6 dB, 0.895 0, and 0.144 4, respectively. These improvements highlight the effectiveness of our approach in handling complex indoor scenes and producing high-fidelity decomposition results. Our experiments also showcase the proposed method's ability to synthesize realistic images from previously unseen views. This capability is particularly important for applications, such as augmented reality and virtual reality. The high quality of the synthesized images further demonstrates our method's robustness in decoupling reflectance and illumination. **Conclusion** Compared with previous methods, our approach excels in rapidly and accurately decomposing indoor scenes. Its strong generalization across diverse indoor environments, ranging from

synthetic setups to real-world datasets, makes it highly valuable for applications, such as virtual object insertion, scene relighting, and material editing. This versatility significantly enhances the practical utility of our method. We believe that this work lays a strong foundation for future research, paving the way for more advanced techniques in the field of intrinsic decomposition and beyond.

Key words: intrinsic decomposition; 3D Gaussian splatting(3DGS); indoor scenes decomposition; Retinex theory; radiance fields

0 引言

依据不同的理论和假设,本征分解可以将图像、视频等目标分解为具有底层物理特征的独立成分,例如反射率(reflectance)和照明(shading)等。由于其分解结果能够反映出场景材质、光照等场景本征属性,因此本征分解可应用于场景材质编辑、重光照等下游任务,具有非常重要的研究价值。

本征分解是一个病态的问题,因为不同的反射率和照明组合均可合成相同的图像,所以必须添加合理的先验约束来获得符合物理规律的分解结果。目前大部分本征图像分解算法所添加的约束均基于经典的Retinex(Land和McCann,1971)理论,该理论假设图像梯度是由光照无关的反射率和光照相关的照明引起的。其中反射率是物体本身的材质颜色,决定了图像最直观的色度,是引起图像显著变化的最主要原因,对应着图像较大的梯度。由于反射率在RGB颜色空间内表现出聚集性,形成一定数量的簇,因此在图像空间表现出一定的稀疏性;照明则是场景内的光照与物体表面几何共同作用后的明暗信息,其变化对应着图像较小的梯度。通常假设场景内光照的强度变化平缓,所以照明的大部分区域具有一定的平滑性。传统的本征分解算法主要针对单幅图像,包括基于优化的算法(Bell等,2014;Bi等,2015;邢冠宇等,2017)和基于学习的算法(Li和Snavely,2018;Luo等,2020;王玉洁等,2022)。基于优化的算法通常将图像的局部与全局信息结合作为分解的先验,而基于学习的算法需要训练模型来捕获场景中的特征。这些算法虽然能够很好地利用单幅图像的空间与色度信息,得到的反射率表现出很好的稀疏性,却没有考虑多视角图像在三维空间的关联性,因此在分解来自同一场景的图像时,可能造成多视图分解结果不一致的问题,不利于扩展到三维任务与应用中。

近年来,神经辐射场(neural radiance field, NeRF)(Mildenhall等,2020)在新视图合成和三维场景重建领域表现出巨大的潜力(肖强等,2024),引发了人们对辐射场表示的广泛关注和兴趣,同时也推动了研究人员将其引入到三维场景本征分解的研究中,IntrinsicNeRF(Ye等,2023)便是其中具有代表性的工作。该工作基于NeRF将空间描述为关于反射率和照明的辐射场;此外,由于场景内存在不满足漫反射假设的材质,IntrinsicNeRF还添加了残差场来拟合方向相关的属性。然而由于NeRF是通过光线投射的方式渲染像素的颜色,因此IntrinsicNeRF只能以像素为单位进行渲染和优化,在优化过程中无法利用图像的全局信息,反射率难以在图像中聚集,只能依赖场景的语义信息来提高稀疏性;同时依赖语义也导致了其泛用性受限,部分纹理可能被错误地分配到残差中。此外NeRF较慢的渲染和优化速度也严重限制了IntrinsicNeRF的分解效率。

三维高斯泼溅(3D Gaussian splatting, 3DGS)(Kerbl等,2023)是目前最先进的新视图合成技术,该技术能够利用快速可微的渲染器进行渲染和优化,极大地提高了训练和推理速度以及渲染结果的保真度。这些优势使得3DGS在计算机视觉和图形学的各项任务中得到了广泛应用。为此,本文基于3DGS设计了一种针对室内场景的本征分解算法,在实现快速分解的同时,还能在新视角下合成高质量的合成图像,克服了单幅图像分解和IntrinsicNeRF等工作存在的问题。

本文使用3DGS表示室内场景,并在高斯椭球上添加反射率、反射率偏移、照明和残差等属性,其中反射率、照明和残差与IntrinsicNeRF的设置相似,而反射率偏移用来拟合不易分离的纹理细节。随后根据本征分解方程计算出每个高斯椭球的颜色,并通过泼溅算法合成图像以及场景的各个属性图,在此基础上进一步进行优化;除了直接监督合成的图像与原始图像接近外,引入反射率分段稀疏和照明

平滑项来约束分解结果符合 Retinex 理论的假设,同时使用色度先验确保反射率的色度合理。在优化过程中还使用深度数据来增强场景属性中的几何信息,并随着优化的进行调整损失函数的各项权重以获得更加合理的分解结果。综上所述,本文创新点包括以下3点:1)提出一种基于3DGS的室内场景本征分解方法,能够将三维场景分解为各个本征属性的辐射场,解决了多视图分解不一致以及分解质量和效率不高的问题;2)设计一种新的本征分解模型,引入反射率偏移来保留更多纹理细节,避免了反射率纹理的错误分配,从而获得更好的分解结果,有利于应用到下游任务;3)利用场景的深度信息,引入了反射率分段稀疏、照明平滑和色度先验等约束,使得反射率表现出更好稀疏性的同时,照明仍然可以保持平滑,并保证分解结果的合理性。

1 相关工作

1.1 基于辐射场的场景表示

基于辐射场的场景表示在新视图合成和三维重建等任务上取得了许多进展与突破,极大地推动了计算机视觉与图形学在相关领域的发展。其中,NeRF基于多层感知机隐式表示场景,并使用体积渲染合成每个像素的颜色。虽然NeRF形式简单,但在训练和推理过程中需要大量重复查询神经网络,效率低下。3DGS是目前最先进的新视图合成和场景重建方法,它结合了显式点云与隐式神经网络的优势,能够在实现快速优化与推理的同时,提供更高质量的渲染结果(Chen等,2024)。目前已有许多工作对3DGS的效率和渲染质量进行了进一步提升和改进,并将3DGS应用于逆渲染(Liang等,2024; Shi等,2024; Gao等,2024)、动态场景建模(Luiten等,2024)等工作中。

1.2 逆渲染

逆渲染指的是从二维图像中推理出场景的几何、材质和光照表示(邢健开和徐昆,2024)。传统的逆渲染包括基于优化的方法和基于训练的方法,其中基于优化的方法需要设计合理的先验约束,这通常依赖于人类对物理规律的总结和经验,在处理复杂场景时效果不佳;而基于学习的方法(Li等,2020, 2022)通过训练神经网络等模型来学习场景的特征,虽然可以取得较好的效果,但需要昂贵的标注以及长

时间的训练。此外,以上面向图像的方法缺乏三维场景的全局信息,因此即使是来自于同一场景的不同图像,也会造成多视图中分解结果不一致的问题。

目前已经有许多工作基于辐射场表示对三维场景进行逆渲染的研究(Wu等,2024b; Fei等,2024)。例如NEILF(neural incident light field)和NEILF++(Yao等,2022; Zhang等,2023)估计了场景的材质场和反射光照场;Wu等人(2023a)通过分解光传输方程联合估计室内场景的材质和光照;Ma等人(2024)基于高斯定向编码对室内场景的镜面反射进行了建模;DE-NeRF(Wu等,2023b)提出一种解耦几何、外观和照明的方法,能够实现视图一致的外观编辑和高频环境光的重光照;此外,Sun等人(2023)对太阳光进行了建模,用于执行室外场景的分解和重光照。然而这些方法都受到NeRF优化和推理速度慢、渲染质量较低的限制,难以取得高质量的分解结果。目前许多研究人员开始基于3DGS对场景进行逆渲染的研究,例如Gao等人(2024)提出一种可重光照的3DGS技术,对场景进行分解和重渲染;Liang等人(2024)则参考实时渲染中的预计算技术,将场景中光照相关的属性存储在体积中,更好地解耦了场景的材质与光照。此外,GaussianShader(Jiang等,2024)基于3DGS对反射表面进行建模,DeferredGS(Wu等,2024a)利用延迟渲染技术实现了场景属性之间的解耦。然而,由于3DGS难以精确表达物体表面,这些方法难以准确恢复场景的法线,从而影响了场景几何属性的恢复以及其他属性的解耦。

结合以上逆渲染存在的问题,本文选择使用本征分解来对三维场景进行解耦。通过本征分解,在避免求解场景几何的同时,仍然能够得到引起视觉感知变化的主要属性,并支持材质替换、重光照等下游任务。

1.3 本征分解

本征图像分解的目标是将图像分解为反射率层和照明层。一般情况下,场景中的反射率呈现出一定的分段和稀疏性,而照明在平坦的表面上则显得更加平滑。基于这些先验知识,许多工作设计了不同的分解方法,其目的在于尽可能地将图像中的色度信息保留在反射率层中,以反映物体表面的材质和纹理等属性,而照明层则捕获了场景内的光照与物体的几何特征。Garces等人(2012)使用聚类算法在颜色空间内对反射率进行聚类,Bi等人(2015)引

入了图像平坦化方法来减少阴影对分离反射率的影响,Chen 和 Koltun(2013)则将深度信息添加到本征分解的先验中。此外,一些工作也探索了基于深度学习的本征图像分解算法,例如 Luo 等人(2020)在多个合成与真实数据集上训练神经网络,并结合表面法线来预测场景的属性。近年来,扩散模型在预测任务中表现出巨大的潜力,一些工作探索了基于扩散模型的单视图本征分解(Kocsis 等, 2024; Zeng 等, 2024; Luo 等, 2024),在材质估计方面取得了良好的效果。但是这些面向图像分解的方法缺乏三维场景的几何属性,在分解时容易产生多视图不一致的问题。如图 1 所示,通过这些分解方法得到的同

一区域的反射率,在色度和亮度存在差异,因此不利于运用到三维场景的分解中。

IntrinsicNeRF 是与本文最为接近的工作。该方法基于 NeRF 和 Retinex 理论,将场景分解为不同属性的辐射场,并能够在新视角下进行场景分解与属性合成。然而,IntrinsicNeRF 在处理室内场景时需要依赖场景分割的语义信息,这导致了部分纹理细节的丢失。相比之下,本文工作使用深度信息来增强场景的几何特征,并基于 3DGS 将场景的属性表示为辐射场。相较于 IntrinsicNeRF,本文工作在优化与合成速度上大幅提高,并获得了更为令人信服

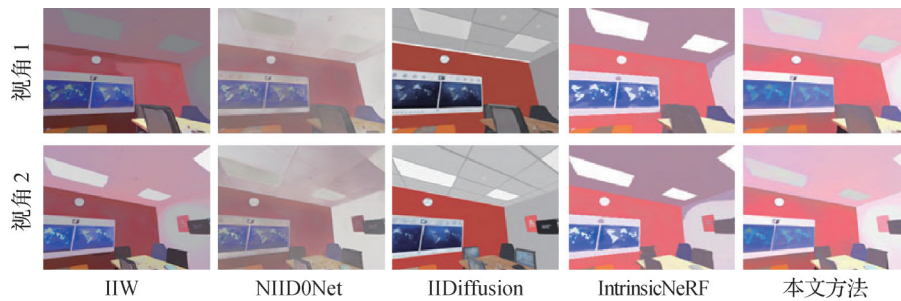


图 1 多视图分解可视化结果

Fig. 1 Visualization of multi-view decomposition results

2 方 法

2.1 三维高斯泼溅

3DGS(Kerbl 等, 2023)从随机生成或运动结构恢复(structure from motion, SfM)获取的稀疏点云开始,将每个点初始化为具有均值以及协方差矩阵的高斯椭球来表示场景,具体为

$$G(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (1)$$

式中,均值 $\boldsymbol{\mu} \in \mathbf{R}^3$ 表示高斯椭球的中心位置,协方差矩阵 $\boldsymbol{\Sigma}$ 控制了高斯椭球的大小以及旋转的角度,并通过旋转四元数以及缩放向量进行重构表示和联合优化。后文使用上标表示像素在二维图像中的索引,下标表示高斯椭球的索引。

3DGS使用可微的泼溅技术进行渲染,能够利用GPU的特性快速高效地渲染图像。在渲染过程中,先将所有的高斯投影到图像空间中,然后将图像分块,并将每个图块上重叠的高斯按照深度进行排序,从前往后依次进行阿尔法混合,最终得到的颜色可

以表示为

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

式中, C^j 表示第 j 个像素最终合成的颜色, c_i 是第 i 个高斯的颜色, α_i 是第 i 个高斯投影到图像空间的不透明度,它通过高斯椭球的不透明度与投影到图像的坐标得到(Chen 和 Wang, 2024),具体为

$$\alpha = o \exp\left(-\frac{1}{2}(\mathbf{x}' - \boldsymbol{\mu}')^T \boldsymbol{\Sigma}'^{-1}(\mathbf{x}' - \boldsymbol{\mu}')\right) \quad (3)$$

式中, o 是高斯椭球的不透明度, $\mathbf{x}'$ 与 $\boldsymbol{\mu}'$ 表示投影到图像空间的坐标, $\boldsymbol{\Sigma}'$ 是二维协方差矩阵,具体为

$$\boldsymbol{\Sigma}' = \mathbf{J} \mathbf{W} \boldsymbol{\Sigma} \mathbf{W}^T \mathbf{J}^T \quad (4)$$

式中, $\mathbf{W}$ 和 $\mathbf{J}$ 分别表示视图变换矩阵和透视投影变换的仿射近似雅可比矩阵(Zwicker 等, 2001)。

2.2 本征分解方程

本文提出的方法流程如图 2 所示。根据 Retinex 理论(Land 和 McCann, 1971),场景中颜色的变化是由反射率和照明引起的,前者是物体表面反射光照的能力,即漫反射系数;后者对应场景内光照与物体几何交互作用的结果。此外,由于室内场景存在不

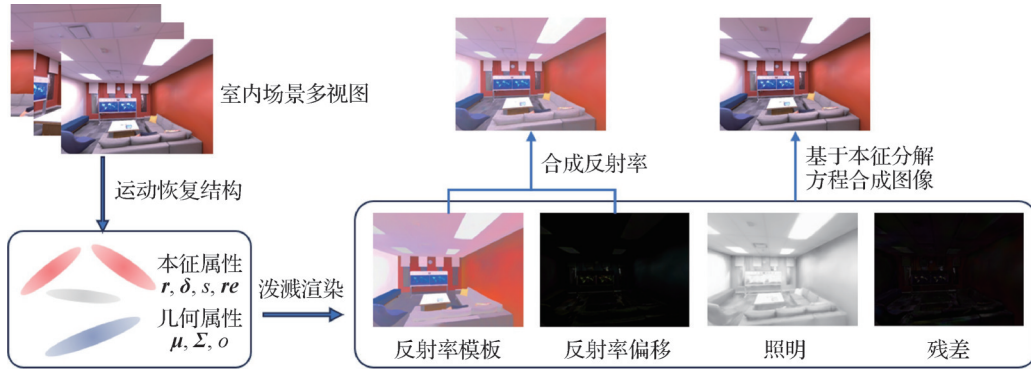


图2 本文提出的室内场景多视图本征分解方法概述

Fig. 2 Overview of the proposed method for multi-view intrinsic decomposition of indoor scenes in this paper

满足朗伯假设的物体,因此需要添加方向相关的残差项来拟合这一部分,以保证优化过程中其他属性可以收敛到合理的结果(Ye等,2023)。因此本征分解方程的形式为

$$c_i = r_i \otimes s_i + e_i \quad (5)$$

式中, i 是高斯椭球的索引, $r_i \in [0, 1]^3$ 表示高斯椭球的反射率, $s_i \in [0, 1]$ 表示照明, $\otimes$ 表示逐通道相乘, e_i 是与方向相关的残差,使用球谐系数表示。

由于三维场景中存在大量不同种类的反射率,分离出所有的反射率是非常困难的。同时本文还注意到 IntrinsicNeRF 在优化过程中可能会将部分方向无关的纹理分配到残差中,从而导致反射率和残差不合理。受其他三维场景分解工作(Kuang等,2023)以及图像软分割方法(Akimoto等,2020;Aksoy等,2017)的启发,本文将反射率视为“模板”,允许在其基础上存在轻微的偏移。如图3所示,这种方法不仅使反射率保持了稀疏性,还有助于让照明更加平滑自然,同时在反射率偏移中保留了更多的纹理细节,使得残差结果合理。因此式(5)可以修正为

$$c_i = (r_i + \delta_i) \otimes s_i + e_i \quad (6)$$

式中, $\delta_i \in \mathbb{R}^3$ 表示反射率偏移。在优化过程中 $(r_i + \delta_i)$ 的范围被剪裁到 $[0, 1]$ 内,以保证反射率的数值合理。注意,由于本文工作基于Retinex理论,并且在接下来的优化过程中不依赖于语义信息,因此对具有大面积的强烈高光和非白光光照的场景分解效果不佳。

不同于单幅图像的本征分解算法,本文以高斯椭球作为载体,首先通过式(6)在三维空间中计算出每个高斯椭球合成的颜色,然后渲染出图像。在不同视角下渲染图像时,相关高斯椭球的属性可以适配到不同的方向,从而保证了多视图分解的一致性。

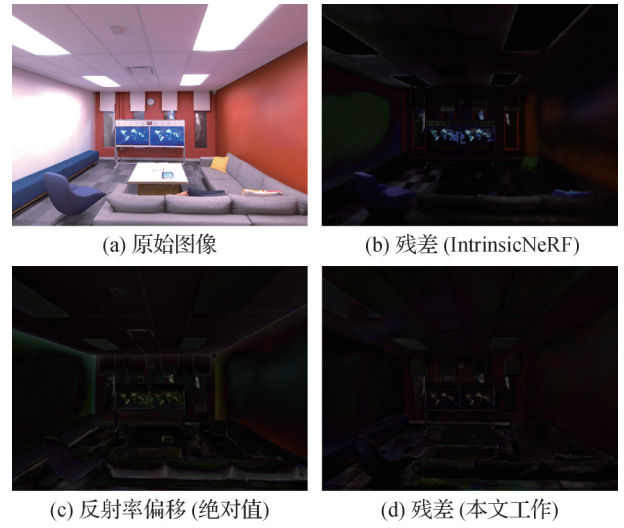


图3 残差与反射率偏移可视化

Fig. 3 Visualization of residuals and reflectance offset
(a) original image; (b) residual (from IntrinsicNeRF);
(c) reflectance offset (absolute value); (d) residual (ours))

2.3 本征分解的监督与约束

2.3.1 颜色监督

合成的图像应当与原图保持一致,因此本文首先使用 Kerbl 等人(2023)提出的损失函数对合成的图像进行监督,具体为

$$L_{\text{color}} = (1 - \lambda)L_1 + \lambda L_{\text{D-SSIM}} \quad (7)$$

由于3DGS采用的是从前向后的渲染管线,一次渲染即可生成整幅图像,不需要像基于NeRF的方法对像素进行逐一采样并渲染,因此可以对图像整体进行优化,从而提高训练速度和效果。但是仅依靠 L_{color} 无法得到符合Retinex理论假设的属性,因此必须添加先验约束来保证分解结果的合理性。本文通过泼溅渲染得到各个属性图,包括反射率图 R 、偏移图 Δ 、照明图 S 、残差图 e 以及深度图 D ,并进一

步对这些属性进行约束,即

$$\{R^i, \Delta^i, S^i, e^i, D^i\} = \{r_i, \delta_i, s_i, e_i, d_i\} \sum_{i \in M} \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (8)$$

式中, d_i 是高斯椭球的深度。

2.3.2 深度监督

在实验中发现分解一定规模的室内场景较为困难,部分区域特别是纹理不足的场景可能会出现高斯椭球缺失的现象,因此需要引入深度真值作为监督,约束高斯椭球的位置合理。对于合成数据集,由于其提供的相机参数与深度数据保持一致,因此本文直接使用深度的真值进行约束,具体为

$$L_{\text{depth}} = \|D - D_{\text{gt}}\|_1 \quad (9)$$

式中, D_{gt} 为场景的真实深度。对于真实场景的数据集,由于 SfM 得到的点云与真实深度存在尺度上的歧义,因此本文使用尺寸不变损失 (Eigen 等, 2014) 对深度进行约束,具体为

$$L_{\text{depth}} = \frac{1}{n} \sum_{i=1}^n \left| \log D_{\text{gt}}^i - \log D^i + \varphi(D_{\text{gt}}, D) \right| \quad (10)$$

$$\varphi(D_{\text{gt}}, D) = \frac{1}{n} \sum_j (\log D_{\text{gt}}^j - \log D^j) \quad (11)$$

2.3.3 反射率分段稀疏

由于室内场景规模较大且反射率种类众多,其中大部分反射率占据了大量比例,而另一些反射率的数量占比很小,因此在优化期间可能会丢失部分反射率。为了解决这一问题,本文没有从一开始就强行解耦所有的反射率,而是借鉴了图像分段平坦化的思想 (Bi 等, 2015), 在优化阶段先对场景的反射率进行分段稀疏,这样可以在不改变场景结构的前提下,去除不必要的细节,并将视觉上较为显著的边缘保留在反射率中。接下来,使用反射率偏移来恢复场景中的纹理细节,特别是数量占比很小的反射率;最后使用残差来拟合与视图相关的部分,确保合成的图像与原图一致。

根据 Retinex 理论,反射率表现出一定的稀疏性,受到 Gong 等人 (2023) 的启发,引入了一种新的稀疏函数来鼓励场景内反射率表示出强烈的稀疏性,具体为

$$H(x) = (1 + \exp(\omega x - \beta))^{-1} \quad (12)$$

这是一个可微的近似阶跃函数,可以通过调整参数 ω 和 β 控制稀疏的程度与范围,在本文的实验

中分别设置为 80 和 16。

CIELab 颜色空间可以很好地反映颜色色度与亮度的关系,有助于反射率和照明的分离。对于 RGB 图像 $I^i \in [0, 1]^3$ 中的像素,本文通过如下变换将其映射到新的色度空间 (Bi 等, 2015), 具体为

$$t(I^i) = [\kappa l, a, b] \quad (13)$$

式中, l, a, b 分别是在 CIELab 颜色空间的三通道分量, κ 是用于削弱亮度的缩放系数,设置为 0.3。本文按照上述规则定义映射 T , 将 RGB 空间转换到新的色度空间,具体为

$$T: I \rightarrow \{t(I^i) | I^i \in I\} \quad (14)$$

这个空间能够更好地反映图像中的反射率属性,减少因照明变化造成的影响。

由于室内场景规模较大且布局复杂,同时缺乏场景内物体的语义信息,因此难以将具有相同反射率的较大面积的区域进行统一处理。但是根据 Bi 等人 (2015) 的见解,虽然人们希望尽可能去除反射率中因照明变化产生的边缘 (例如阴影或室内墙壁之间的边缘),但如果这些边缘在视觉感知中具有重要性,则仍需将其保留在反射率中。因此,本文将这些边缘都保留在反射率中,并分别对边缘两侧的反射率进行分段稀疏

$$L_{\text{sparsity}} = w_{\text{sparsity}} \|\nabla R\|_1 \quad (15)$$

$$w_{\text{sparsity}} = H(\max(\|\nabla T(C)\|_1, \eta_c \|\nabla C\|_1, \eta_N \|\nabla N\|_1)) \quad (16)$$

式中, $\|\nabla T(C)\|_1$ 是图像在色度空间的梯度, $\|\nabla C\|_1$ 为原图像的梯度; N 是法线,可以通过深度和相机的位姿参数计算得到, $\|\nabla N\|_1$ 为法线的梯度; η_c 和 η_N 为缩放系数。 w_{sparsity} 可以使得反射率按边缘分段,并在段内表现出聚集性,有利于反射率和照明的分离。

2.3.4 照明平滑

照明是场景内光照与几何作用的结果。本文假设场景内光照的强度变化平缓,引起场景内较为明显的明暗变化主要是由于场景的几何引起的,所以在照明图中需要保留场景的几何边缘,而平坦的区域则约束为平滑。因此照明平滑约束表达为

$$L_{\text{smooth}} = \sum_{i \in \{i | \|\nabla N^i\|_1 < \varepsilon_N\}} w_{\text{smooth}} \|\nabla S^i\|_2^2 \quad (17)$$

$$w_{\text{smooth}}^i = \|\nabla T^i(C)\|_2^2 \quad (18)$$

式中, ε_N 为判断平坦程度的阈值,设置为 0.1。本文

只对法线梯度大于阈值的像素进行约束,以保证照明在平坦表面平滑的同时仍可以保持比较明显的几何特征。同时 L_{smooth} 可以使得无法保存在反射率的纹理细节优化到偏移 Δ 中。

2.3.5 色度先验

此外,反射率的色度不应当随着图像的亮度变化而改变(Ye等,2023;Meka等,2016),因此本文假设反射率和原始图像的色度应当一致。为了得到更令人信服的反射率,本文使用如下色度先验进行约束,具体为

$$L_{\text{chrom}} = \|T(R + \Delta) - T(C)\|_2^2 \quad (19)$$

式中, $T(R + \Delta)$ 是反射率图与偏移图的组合经过转换后得到的色度, $T(C)$ 是原始图像的色度。 L_{chrom} 可以使得分解得到的反射率表现出合理的色度,减少光照对反射率的影响。

2.3.6 反射率偏移与残差正则化

为了防止在优化过程中出现反射率都聚集到反射率偏移中,或漫反射相关项都变成0的极端情况,本文需要对反射率偏移与残差进行约束,使得反射率模板和照明在分解中占据主导位置,具体为

$$L_{\text{offset}} = \|\Delta\|_2^2 \quad (20)$$

$$L_{\text{residual}} = \|e\|_2^2 \quad (21)$$

这两项损失的权重在迭代早期设置较高,并随着优化的进行逐步降低。

2.4 实现细节

本文基于Gao等人(2024)实现的渲染器进行实验,共进行30 000次迭代,并使用如下损失联合优化,具体为

$$L = \lambda_{\text{color}} L_{\text{color}} + \lambda_{\text{depth}} L_{\text{depth}} + \lambda_{\text{sparsity}} L_{\text{sparsity}} + \lambda_{\text{smooth}} L_{\text{smooth}} + \lambda_{\text{chrom}} L_{\text{chrom}} + \lambda_{\text{offset}} L_{\text{offset}} + \lambda_{\text{residual}} L_{\text{residual}} \quad (22)$$

在本文的实验中, λ_{color} 设置为1, λ_{depth} 设置为0.05, $\lambda_{\text{sparsity}}$ 设置为0.2, λ_{smooth} 设置为1, λ_{chrom} 设置为1, λ_{offset} 在前5 000次迭代中设置为1,随后按照二次函数衰减到0.02; $\lambda_{\text{residual}}$ 在前15 000次迭代中设置为1,并在随后的优化中设置为0.02。

3 实验

3.1 对比实验

为了更好地与现有针对三维室内场景进行本征

分解的工作进行对比,本文选择在合成数据集Replica(Straub等,2019)与真实数据集ScanNet++(Yeshwanth等,2023)上进行测试与评估,对比的实验包括IIW(intrinsic images in the wild)(Bell等,2014)、NIID-Net(Luo等,2020)和IntrinsicNeRF(Ye等,2023)。其中IIW是基于优化的经典本征图像分解算法,NIID-Net则通过训练神经网络来预测场景的反射率和照明,而IntrinsicNeRF与本文都是基于辐射场表示的三维场景本征分解算法。所有实验都在单张Nvidia RTX 3090显卡上进行。

3.1.1 Replica数据集

Replica数据集包含多个高质量的合成室内场景。为了公平对比,选择与IntrinsicNeRF相同的场景,并在此基础上挑选一致的训练和测试图像。为了获得效果更好的分解结果,使用COLMAP(Schönbberger和Frahm,2016)得到场景的稀疏点云并以此作为初始化条件。由于缺乏场景内反射率和照明等属性的真值,将这些属性可视化以提供对比。此外,与IntrinsicNeRF对比了在新视图下合成图像的相关指标,包括峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity index, SSIM)和学习感知图像块相似度(learned perceptual image patch similarity, LPIPS);与IntrinsicNeRF的训练时间进行了比较。

在合成数据集Replica上测试的结果如图4所示。可以看出,IIW和NIID-Net都无法很好地去除反射率中的阴影,这是因为这些分解算法无法结合三维场景的全局和多视角信息;IIW的照明中仍然存在较为明显的纹理信息无法去除,如地板;NIID-Net得到的反射率部分区域色度偏暗,与原始图像的色度偏差较大;IntrinsicNeRF受限于NeRF只能以像素为单位进行渲染的缺陷,无法对渲染得到的图像整体进行优化,导致了反射率不合理的稀疏;此外IntrinsicNeRF依赖于语义信息去除场景内的阴影和高光,这会导致部分纹理分配到残差中。本文使用反射率偏移来捕获这些细节,使得残差项更好地匹配场景中方向相关的属性,并保证了反射率的稀疏性以及照明的平滑。

本文与IntrinsicNeRF在新视图合成的相关指标如表1所示。数据显示本文方法还可以在新视角下合成更高质量的图像。此外,IntrinsicNeRF在Replica数据集的8个场景平均优化时间为17.5 h,而

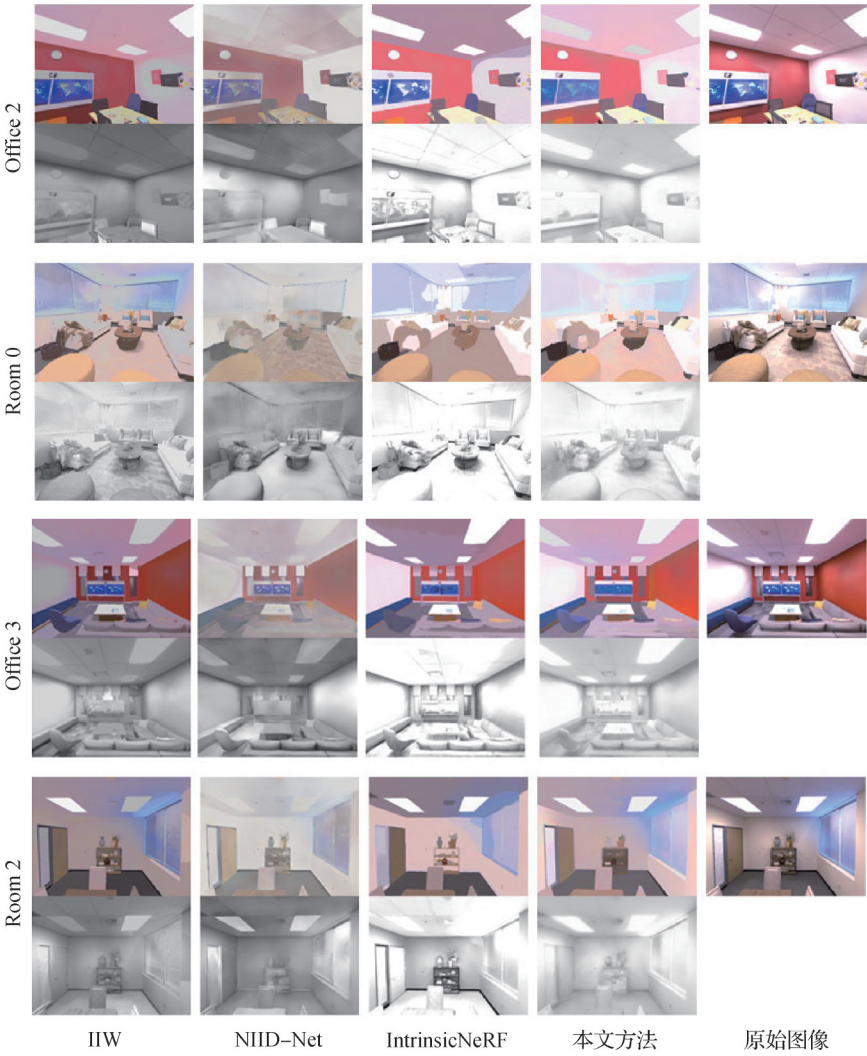


图4 Replica数据集部分场景的分解结果可视化

Fig. 4 Visualization of decomposition results for selected scenes on the Replica dataset

表1 Replica数据集合成图像指标对比

Table 1 Comparison of metrics for synthesized images on the Replica dataset

方法	Office 0			Office 1			Office 2		
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓	PSNR/dB ↑	SSIM ↑	LPIPS ↓
IntrinsicNeRF	33.973 4	0.929 2	0.066 6	35.450 0	0.953 2	0.068 0	30.282 7	0.923 1	0.084 3
本文	36.934 1	0.978 6	0.050 2	37.947 8	0.959 2	0.087 9	34.259 8	0.972 8	0.091 8
方法	Office 3			Office 4			Room 0		
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓
IntrinsicNeRF	29.955 3	0.917 9	0.074 1	30.920 1	0.910 6	0.109 8	27.081 2	0.806 3	0.169 8
本文	34.085 2	0.969 4	0.066 6	32.898 4	0.961 8	0.090 8	32.933 2	0.952 1	0.100 3
方法	Room 1			Room 2			平均指标		
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓	PSNR ↑ /dB	SSIM ↑	LPIPS ↓
IntrinsicNeRF	28.185 2	0.804 8	0.205 6	29.787 3	0.880 9	0.134 3	30.704 4	0.890 8	0.114 0
本文	34.042 7	0.956 8	0.110 6	34.462 0	0.972 9	0.090 3	34.695 5	0.965 4	0.086 1

注:加粗字体表示各列最优结果。“↑”和“↓”分别表示值越大越好、值越小越好。

本文方法的平均优化时间为 32.1 min。这也反映出本文方法能够更加快速地进行场景分解,具有更多的应用前景。

3.1.2 ScanNet++数据集

ScanNet++数据集由多个真实的室内场景组成,本文选择了其中的 5 个场景进行实验。本文使用 Luo 等人(2020)的方法来计算场景法线,然后使用 COLMAP 计算场景的相机参数与稀疏点云。

新视图合成的相关指标如表 2 所示。相比于 IntrinsicNeRF,本文方法在真实场景上仍然能够得

到质量更高的合成结果,这有利于将分解结果应用于场景编辑、重光照等扩展任务。

反射率分解结果如图 5 所示。可以看出,由于真实场景存在许多具有不规则几何的物体,并且部分区域光照强烈复杂,受限于 Retinex 理论的局限性,IIW 和本文方法无法很好地去除高光;相比于其他工作的分解效果,NIID-Net 的反射率色度偏暗;而 IntrinsicNeRF 依靠语义层次聚类去除了墙壁上的高光,但同时也导致反射率的部分纹理细节丢失,一些反射率应当相同的区域(如天花板和墙壁)因语义不

表 2 ScanNet++数据集合成图像指标对比
Table 2 Comparison of metrics for synthesized images on the ScanNet++ dataset

方法	56a0ec536c			b20a261fdf			f8f12e4e6b		
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PSNR ↑/dB	SSIM ↑	LPIPS ↓	PSNR ↑/dB	SSIM ↑	LPIPS ↓
IntrinsicNeRF	24.694 9	0.809 6	0.239 4	27.697 7	0.837 5	0.199 2	24.646 9	0.817 7	0.298 3
本文	26.497 8	0.879 4	0.152 4	30.735 8	0.919 6	0.122 3	26.277 8	0.881 9	0.162 6

方法	8b5caf3398			0b031f3119			平均指标		
	PSNR/dB ↑	SSIM ↑	LPIPS ↓	PSNR ↑/dB	SSIM ↑	LPIPS ↓	PSNR ↑/dB	SSIM ↑	LPIPS ↓
IntrinsicNeRF	27.157 2	0.855 6	0.198 3	25.707 1	0.840 3	0.267 9	25.980 8	0.832 1	0.240 6
本文	29.643 5	0.915 6	0.112 8	26.593 2	0.878 2	0.171 7	27.949 6	0.895 0	0.144 4

注:加粗字体表示各列最优结果。“↑”和“↓”分别表示值越大越好、值越小越好。

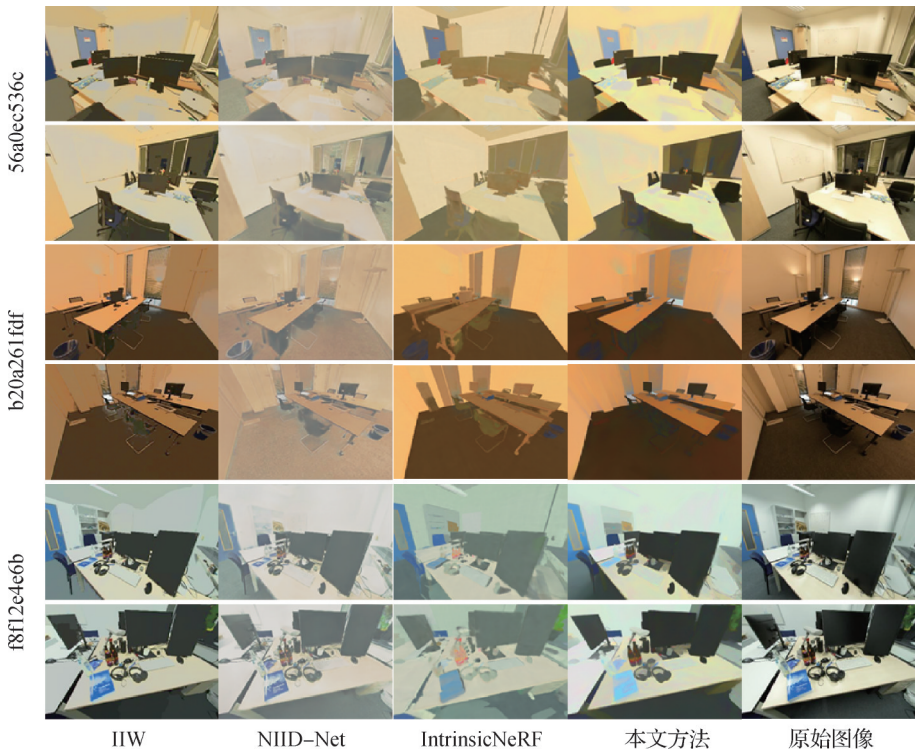


图 5 ScanNet++数据集部分场景的反射率可视化

Fig. 5 Visualization of reflectance for selected scenes on the ScanNet++ dataset

同而导致反射率不同。这体现出基于语义的本征分解算法存在的优势与不足。未来有望结合语义与本文提出的本征分解模型进一步研究,以保证保留更多纹理的前提下更加彻底地去除反射率中的光照。

3.2 消融实验

本文通过消融实验验证深度监督以及在上一节提出的反射率分段稀疏、照明平滑以及色度先验的作用,基准模型是仅使用 L_{color} 、 L_{offset} 和 $L_{residual}$ 监督的模型。从图6可以看出,缺乏以上约束的基准模型无法正确解耦场景属性;缺少深度监督会导致反射率

中部分深度区域色度不合理,同时对应的照明不能很好地反映出光照的变化;缺少照明反射率分段稀疏约束导致场景没有产生明显的分段,无法排除光照因素,从而导致不正确的分离;缺少照明平滑项会导致照明中应当平滑的区域出现明显的突变,例如窗户边缘,此区域的颜色变化应当是由反射率引起的,这不符合照明的特点;而缺少色度先验项则会导致反射率的色度过亮,在感知上不合理。这些现象表明反射率稀疏、照明平滑以及反射率色度先验对得到合理分解结果起到重要作用。

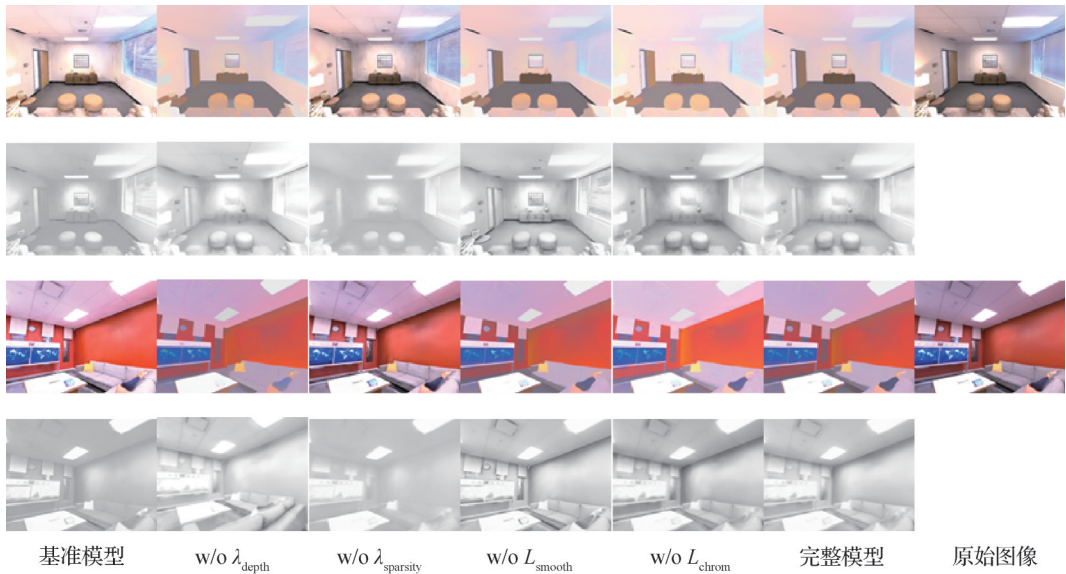


图6 消融实验结果可视化

Fig. 6 Visualization of ablation study

为了显示使用深度来增强场景几何信息的重要性,如图7所示,本文在反射率分段稀疏与照明平滑项中删除了几何对属性图优化的影响。从图中可以看出,由于缺乏几何信息,场景的照明平滑约束难以达到收敛。

3.3 扩展应用

室内场景通过本征分解可以得到场景中表示物体纹理的漫反射材质以及光照相关的照明,这有助于进一步的场景编辑应用。如图8所示,本文通过分割模型指定物体并修改其反射率,随后与照明和残差重新合成图像,可以实现场景内的材质替换;同时还可以通过修改照明的数值来实现光照的编辑。

3.4 缺陷与改进

在合成数据集与真实数据集上的结果表明,本工作在具有简单几何的场景下表现更好,这是因为

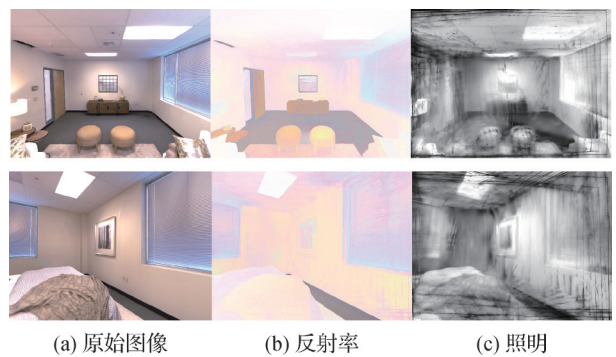


图7 缺失几何信息导致分解质量较差

Fig. 7 The lack of geometric information leads to poor decomposition quality ((a) original images; (b) reflectance; (c) shading)

这样的场景更有利于使用本文中引入的约束;在反射率分段稀疏中,由于缺乏语义信息,对于具有大面积阴影或强烈高光的某一段反射率,本文算法无法

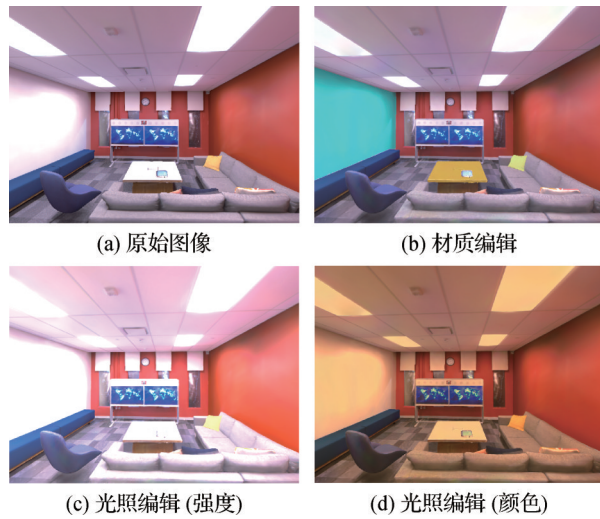


图8 扩展应用示例

Fig. 8 Extended application examples ((a) original image;
(b) material edit; (c) illumination edit (intensity);
(d) illumination edit (color))

正确识别,而是会将其“模糊”到反射率中;由于本文算法基于Retinex理论进行设计,没有考虑光照的颜色,因此在处理彩色光照时无法将照明色度与反射率解耦,不易获得合理的反射率。以上缺陷在Replica数据集集中的场景Room 0以及真实场景的高光区域表现最为明显。目前已经有许多工作基于物理模型对图像进行分解,如NIID-Net(Luo等,2020)显式对场景法线进行了建模,Xing等人(2016)对光照色度进行了建模。未来有望基于语义信息与物理模型在3DGS的基础上进行改进,完善Retinex理论的缺陷和不足,从而获得更高质量的分解属性以及合成图像。

4 结 论

本文提出一种基于3DGS技术的室内场景本征分解方法。为了能够更加准确地分离场景中光照相关和光照无关的属性,本文设计了一种新的本征分解模型,同时利用3DGS的特性将室内场景分解为反射率、偏移、照明以及残差的辐射场,在实现快速优化与渲染的同时,还能够在新视角下合成图像与场景属性。此外,本文引入了新的反射率分段稀疏、照明平滑以及色度先验等约束,并利用深度信息增强场景的几何,用来保证分解得到的属性更加合理。本文的实验结果显示,相比于先前的三维场景分解工作,本文方法不仅能够快速分解场景,还能够生成

更高质量的渲染图像,并得到更加合理的本征属性。未来将结合场景的法线等物理属性对分解模型进行进一步设计,用于处理更加复杂的室内场景,并有望应用于增强现实等领域。

参考文献 (References)

- Akimoto N, Zhu H C, Jin Y H and Aoki Y. 2020. Fast soft color segmentation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 8274-8283 [DOI: 10.1109/CVPR42600.2020.00830]
- Aksoy Y, Aydin T O, Smolić A and Pollefeys M. 2017. Unmixing-based soft color segmentation for image manipulation. *ACM Transactions on Graphics*, 36(2): #19 [DOI: 10.1145/3002176]
- Bell S, Bala K and Snavely N. 2014. Intrinsic images in the wild. *ACM Transactions on Graphics*, 33(4): #159 [DOI: 10.1145/2601097.2601206]
- Bi S, Han X G and Yu Y Z. 2015. An L_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Transactions on Graphics*, 34(4): #78 [DOI: 10.1145/2766946]
- Chen G K and Wang W G. 2024. A survey on 3D Gaussian splatting [EB/OL]. [2024-01-22]. <https://arxiv.org/pdf/2401.03890.pdf>
- Chen Q F and Koltun V. 2013. A simple model for intrinsic image decomposition with depth cues//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 241-248 [DOI: 10.1109/ICCV.2013.37]
- Eigen D, Puhrsch C and Fergus R. 2014. Depth map prediction from a single image using a multi-scale deep network//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2366-2374
- Fei B, Xu J Y, Zhang R, Zhou Q Y, Yang W D and He Y. 2024. 3D Gaussian as a new era: a survey [EB/OL]. [2024-09-24]. <https://arxiv.org/pdf/2402.07181.pdf>
- Gao J, Gu C, Lin Y T, Li Z H, Zhu H, Cao X, Zhang L and Yao Y. 2024. Relightable 3D Gaussians: realistic point cloud relighting with BRDF decomposition and ray tracing//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 73-89 [DOI: 10.1007/978-3-031-72995-9_5]
- Garces E, Munoz A, Lopez-Moreno J and Gutierrez D. 2012. Intrinsic images by clustering. *Computer Graphics Forum*, 31(4): 1415-1424 [DOI: 10.1111/j.1467-8659.2012.03137.x]
- Gong B C, Wang Y H, Han X G and Dou Q. 2023. RecolorNeRF: layer decomposed radiance fields for efficient color editing of 3D scenes//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 8004-8015 [DOI: 10.1145/3581783.3611957]
- Jiang Y W Q, Tu J D, Liu Y, Gao X F, Long X X, Wang W P and Ma Y X. 2024. GaussianShader: 3D Gaussian splatting with shading

- functions for reflective surfaces//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5322-5332 [DOI: 10.1109/cvpr52733.2024.00509]
- 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]
- Kocsis P, Sitzmann V and Nießner M. 2024. Intrinsic image diffusion for indoor single-view material estimation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5198-5208 [DOI: 10.1109/CVPR52733.2024.00497]
- Kuang Z F, Luan F J, Bi S, Shu Z X, Wetzstein G and Sunkavalli K. 2023. PaletteNeRF: palette-based appearance editing of neural radiance fields//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 20691-20700 [DOI: 10.1109/CVPR52729.2023.01982]
- Land E H and McCann J J. 1971. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61(1): 1-11 [DOI: 10.1364/JOSA.61.000001]
- Li Z Q, Shafiei M, Ramamoorthi R, Sunkavalli K and Chandraker M. 2020. Inverse rendering for complex indoor scenes: shape, spatially-varying lighting and SVBRDF from a single image//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2472-2481 [DOI: 10.1109/CVPR42600.2020.00255]
- Li Z Q, Shi J, Bi S, Zhu R, Sunkavalli K, Hasan M, Xu Z X, Ramamoorthi R and Chandraker M. 2022. Physically-based editing of indoor scene lighting from a single image//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 555-572 [DOI: 10.1007/978-3-031-20068-7_32]
- Li Z Q and Snavely N. 2018. CGIntrinsics: better intrinsic image decomposition through physically-based rendering//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 381-399 [DOI: 10.1007/978-3-030-01219-9_23]
- Liang Z H, Zhang Q, Feng Y, Shan Y and Jia K. 2024. GS-IR: 3D Gaussian splatting for inverse rendering//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 21644-21653 [DOI: 10.1109/CVPR52733.2024.02045]
- Luiten J, Kopanas G, Leibe B and Ramanan D. 2024. Dynamic 3D Gaussians: tracking by persistent dynamic view synthesis//Proceedings of 2024 International Conference on 3D Vision. Davos, Switzerland: IEEE: 800-809 [DOI: 10.1109/3DV62453.2024.00044]
- Luo J D, Ceylan D, Yoon J S, Zhao N X, Philip J, Frühstück A, Li W B, Richardt C and Wang T F. 2024. IntrinsicDiffusion: joint intrinsic layers from latent diffusion models//Proceedings of 2024 ACM SIGGRAPH Conference Papers. Denver, USA: ACM: 74 [DOI: 10.1145/3641519.3657472]
- Luo J D, Huang Z Y, Li Y J, Zhou X W, Zhang G F and Bao H J. 2020. NIID-Net: adapting surface normal knowledge for intrinsic image decomposition in indoor scenes. *IEEE Transactions on Visualization and Computer Graphics*, 26(12): 3434-3445 [DOI: 10.1109/TVCG.2020.3023565]
- Ma L, Agrawal V, Turki H, Kim C, Gao C, Sander P, Zollhöfer M and Richardt C. 2024. SpecNeRF: Gaussian directional encoding for specular reflections//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 21188-21198 [DOI: 10.1109/CVPR52733.2024.02002]
- Meka A, Zollhöfer M, Richardt C and Theobalt C. 2016. Live intrinsic video. *ACM Transactions on Graphics*, 35(4): #109 [DOI: 10.1145/2897824.2925907]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]
- Schonberger J L and Frahm J M. 2016. Structure-from-motion revisited//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Shi Y H, Wu Y M, Wu C M, Liu X, Zhao C, Feng H C, Zhang J, Zhou B, Ding E and Wang J D. 2024. GIR: 3D Gaussian inverse rendering for relightable scene factorization [EB/OL]. [2024-08-15]. <https://arxiv.org/pdf/2312.05133.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]. [2024-09-24]. <https://arxiv.org/pdf/1906.05797.pdf>
- Sun J M, Wu T, Yang Y L, Lai Y K and Gao L. 2023. SOL-NeRF: sun-light modeling for outdoor scene decomposition and relighting//Proceedings of 2023 SIGGRAPH Asia Conference Papers. Sydney, Australia: ACM: #31 [DOI: 10.1145/3610548.3618143]
- Wang Y J, Fan Q N, Li K, Chen D D, Yang J Y, Lu J Z, Lischinski D and Chen B Q. 2022. High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition. *Journal of Image and Graphics*, 27(2): 404-420 (王玉洁, 樊庆楠, 李坤, 陈冬冬, 杨敬钰, 卢健智, Lischinski D, 陈宝权. 2022. 面向本征图像分解的高质量渲染数据集与非局部卷积网络. *中国图象图形学报*, 27(2): 404-420) [DOI: 10.11834/jig.210705]
- Wu L W, Zhu R, Yaldiz M B, Zhu Y H, Cai H, Matai J, Porikli F, Li T M, Chandraker M and Ramamoorthi R. 2023a. Factorized inverse path tracing for efficient and accurate material-lighting estimation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3825-3835 [DOI: 10.1109/

- ICCV51070.2023.00356]
- Wu T, Sun J M, Lai Y K and Gao L. 2023b. DE-NeRF: decoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting//Proceedings of 2023 ACM SIGGRAPH Conference. Los Angeles, USA: ACM: #74 [DOI: 10.1145/3588432.3591483]
- Wu T, Sun J M, Lai Y K, Ma Y W, Kobbelt L and Gao L. 2024a. DeferredGS: decoupled and editable Gaussian splatting with deferred shading [EB/OL]. [2024-09-24]. <https://arxiv.org/pdf/2404.09412.pdf>
- Wu T, Yuan Y J, Zhang L X, Yang J, Cao Y P, Yan L Q and Gao L. 2024b. Recent advances in 3D Gaussian splatting. Computational Visual Media, 10 (4): 613-642 [DOI: 10.1007/s41095-024-0436-y]
- Xiao Q, Chen M L, Zhang Y and Huang X H. 2024. Neural radiance field reconstruction for sparse indoor panoramas. Journal of Image and Graphics, 29(9): 2596-2609 (肖强, 陈铭林, 张晔, 黄小红. 2024. 室内稀疏全景图的神经辐射场重建. 中国图象图形学报, 29(9): 2596-2609) [DOI: 10.11834/jig.230643]
- Xing G Y, Liu Y L, Zhang W F and Ling H B. 2016. Light mixture intrinsic image decomposition based on a single RGB-D image. The Visual Computer, 32(6): 1013-1023 [DOI: 10.1007/s00371-016-1238-8]
- Xing G Y, Yuan X, Li L and Liu Y L. 2017. User-assisted intrinsic image decomposition based on a single RGB-D image. Journal of Computer-Aided Design and Computer Graphics, 29(7): 1283-1291 (邢冠宇, 袁霞, 李龙, 刘艳丽. 2017. 单幅RGB-D图像本征图像交互分解算法. 计算机辅助设计与图形学学报, 29(7): 1283-1291)
- Xing J K and Xu K. 2024. Physically based differentiable rendering: a survey. Journal of Image and Graphics, 29(10): 2912-2925 (邢健开, 徐昆. 2024. 基于物理的可微渲染综述. 中国图象图形学报, 29(10): 2912-2925) [DOI: 10.11834/jig.230715]
- Yao Y, Zhang J Y, Liu J B, Qu Y H, Fang T, McKinnon D, Tsin Y and Quan L. 2022. NeILF: neural incident light field for physically-based material estimation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 700-716 [DOI: 10.1007/978-3-031-19821-2_40]
- Ye W C, Chen S, Bao C, Bao H J, Pollefeys M, Cui Z P and Zhang G F. 2023. IntrinsicNeRF: learning intrinsic neural radiance fields for editable novel view synthesis//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 339-351 [DOI: 10.1109/ICCV51070.2023.00038]
- Yeshwanth C, Liu Y C, Nießner M and Dai A. 2023. ScanNet++: a high-fidelity dataset of 3D indoor scenes//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12-22 [DOI: 10.1109/ICCV51070.2023.00008]
- Zeng Z, Deschaintre V, Georgiev I, Hold-Geoffroy Y, Hu Y W, Luan F J, Yan L Q and Hašan M. 2024. RGB $\leftrightarrow$ X: image decomposition and synthesis using material- and lighting-aware diffusion models//Proceedings of 2024 ACM SIGGRAPH Conference Papers. Denver, USA: ACM: #75 [DOI: 10.1145/3641519.3657445]
- Zhang J Y, Yao Y, Li S W, Liu J B, Fang T, McKinnon D, Tsin Y and Quan L. 2023. NeILF++: inter-reflectable light fields for geometry and material estimation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 3578-3587 [DOI: 10.1109/ICCV51070.2023.00333]
- Zwicker M, Pfister H, Van Baar J and Gross M. 2001. EWA volume splatting//Proceedings Visualization, 2001. San Diego, USA: IEEE: 29-36 [DOI: 10.1109/VISUAL.2001.964490]

作者简介

吕恒烨,男,硕士研究生,主要研究方向为本征分解与逆渲染。E-mail:2022223040107@stu.scu.edu.cn

邢冠宇,通信作者,男,副教授,主要研究方向为计算机图形学、增强现实和逆向渲染。E-mail: xingguanyu@scu.edu.cn

刘艳丽,女,教授,主要研究方向为计算机图形学、增强现实、图像/视频处理。E-mail: yanlilui@scu.edu.cn

李宏,男,硕士研究生,主要研究方向为三维重建与神经辐射场。E-mail: leehong@stu.scu.edu.cn

袁霞,女,副教授,主要研究方向为计算机图形学、计算机视觉和AIGC。E-mail: yuanxia@cuit.edu.cn