面向逆渲染的室内场景光源建模

彭思达[1)], 陈希[1)], 周晨旭[1)], 王哲[2)], 周晓巍[1)*]

[1)] (浙江大学计算机科学与技术学院 杭州 310058)
[2)] (阿里巴巴网络技术有限公司 杭州 310058)
(xwzhou@zju.edu.cn)

摘 要: 近年来的神经逆渲染方法使用神经网络表示物体几何形状和表面材质, 通过基于物理的渲染过程从多视图图像学习网络参数. 然而, 这些方法通常假设光源位于无限远处, 在具有复杂光照条件的室内场景中, 这种假设很少成立. 为解决该问题, 提出一种基于点云的光照表示方法, 用于建模室内场景中的随空间坐标变化的高频光照效果; 在输入的多视图图像中检测二维光源掩模, 通过三维重建方法得到三维光源点; 在基于物理的渲染过程中显式引入所提出的光照表示, 提高了渲染模型表达镜面反射效果的能力, 有效地减轻逆向渲染过程中的歧义性. 在真实和合成数据集上的实验结果表明, 所提方法达到了逆渲染的优异性能, 并能够产生逼真的重光照结果.

关键词: 逆向渲染; 重光照; 神经场景表示
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2024-00342

Modeling Emitters in Indoor Scenes for Inverse Rendering

Peng Sida[1)], Chen Xi[1)], Zhou Chenxu[1)], Wang Zhe[2)], and Zhou Xiaowei[1)*]

[1)] (*College of Computer Science and Technology, Zhejiang University, Hangzhou* 310058)
[2)] (*Alibaba Group, Hangzhou* 310058)

Abstract: Recent neural inverse rendering methods represent object geometry and materials with neural networks and learn network parameters from multi-view images through physically based rendering. However, these methods typically assume that the light source is located at an infinite distance, which seldomly holds in indoor scenarios that exhibit complex illumination. To address this issue, we propose a point cloud-based lighting representation method to model the spatially-varying and high-frequency lighting effects in indoor scenes. Our method first detects 2D light source masks in the input multi-view images, and then obtains a set of emitters through 3D reconstruction algorithm. We explicitly incorporate emitters in the Monte Carlo sampling, which improves the ability to model the specular effects, thus effectively alleviating the ambiguity in the inverse rendering process. Experiments on real and synthetic datasets demonstrate that the proposed method achieves the best performance in inverse rendering and can produce realistic relighting results.

Key words: inverse rendering; relighting; neural scene representation

收稿日期: 2024-07-02; 修回日期: 2024-08-21. **基金项目**: 国家自然科学基金(62172364). **彭思达**(1996—), 男, 博士, 研究员, 博士生导师, CCF 会员, 主要研究方向为计算机图形学、三维计算机视觉; **陈希**(2000—), 男, 硕士研究生, 主要研究方向为计算机图形学; **周晨旭**(2002—), 男, 博士研究生, 主要研究方向为三维计算机视觉; **王哲**(1994—), 男, 硕士, 阿里巴巴研究员, 主要研究方向为三维计算机视觉; **周晓巍**(1986—), 男, 博士, 副教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为计算机图形学、三维计算机视觉.

从图像中恢复几何、材质和光照的问题被称为逆渲染, 是计算机视觉和计算机图形学领域长期面临的挑战. 恢复物体的物理属性对于许多应用非常重要, 如虚拟现实/增强现实、电影制作和视频游戏. 然而, 由于真实世界物体材质的固有复杂性以及它们与光的相互作用, 逆渲染仍然是一个欠定问题. 已有的工作通过使用复杂的捕捉系统[1-2]或在黑暗环境中使用共位闪光灯和相机[3-5]来克服这个问题; 但这些方法需要特殊的设备或环境, 限制了它们的应用.

当前, 一些方法[6-8]利用神经网络逆渲染技术从在自然光照下拍摄的图像中恢复物体材质, 它们利用基于物理的渲染过程通过计算渲染方程[9]来渲染图像, 该渲染方程考虑了物体几何形状、物体材质和环境光照条件. 通常, 神经网络逆渲染技术通过神经网络对物体的几何形状和空间变化的材质建模, 并通过一系列球面高斯函数[7-8]或环境贴图[10-11]表示环境光照条件, 最后通过计算渲染图像与输入图像的残差来优化模型参数, 这些方法在重建物体材质方面表现出卓越的效果. 然而, 本文实验显示, 这些方法在处理复杂室内光照下拍摄的物体时往往无法生成合理的结果. 例如, NvDiffrecMC[11]方法估计的材质的镜面反射往往过高, 导致重光照的结果不正确.

本文将这种不准确性归因于 2 个因素. (1) 光照建模不正确. 现有的方法通常假设所有光源都来自无限远的环境, 但其并不适用于室内场景. 由于大多数室内场景的空间有限, 光照条件与所处的三维位置和方向有关, 因此更应该被随空间坐标变化的光照表示, 而不是被环境贴图所表示. 在室内场景中, 使用环境贴图来照明物体很可能导致材质恢复不正确. (2) 材质优化过程中存在的歧义性. 由于室内环境中光照条件的复杂性, 在优化过程中环境光照和物体材质之间存在歧义, 图像中的高频效果更容易通过材质而不是光照来解释. 镜面光可以通过更具反射性的表面和低频光照来表示, 导致物体呈现出比实际更具反射性的外观.

本文提出一种面向室内物体逆向渲染的方法, 并设计了一个随空间坐标变化的光照模型, 通过提取室内场景中的光源准确地表示高频光照. 从输入图像定位一组环境光源, 并为每个光源定义可学习的光强参数, 用于表示高频光照. 基于本文提出的光照模型进行渲染时, 首先根据输入图像检测和提取场景中的光源; 然后对齐目标物体和场景的坐标系, 将提取的光源融入物体渲染方程; 最后对于每个像素追踪主光线到物体网格上的表面点, 并利用蒙特卡罗积分产生指向每个光源的次光线.

在合成和真实的数据集上验证了本文方法的有效性. 由于现有的数据集重点关注由环境照明的物体, 没有可用于室内物体评估的数据集, 因此本文创建了一个合成的室内数据集, 将合成物体放置在室内模型中, 并使用 Blender 渲染图像. 在合成和真实数据集上的实验结果表明, 本文方法可以恢复出准确的材质, 达到了较对比方法最优的材质估计性能, 实现了高质量逼真的重光照效果和自由视角合成效果. 相比于已有的逆渲染方法, 本文方法具有几个关键优势. 首先, 本文准确地建模了室内场景中的复杂光照条件, 而已有方法大多数局限于假设无限远光照条件的情况; 其次, 通过明确约束光源的位置和形状, 本文有效地消除了光照和材质之间的歧义, 从而提高了材质估计准确性; 第三, 本文方法通过仅对次要光线采样指向光源的方向, 实现了重要性采样, 减少了方差. 因此, 相比于 NvDiffrecMC[11]方法, 本文方法中的优化过程更加稳定, 并产生更加平滑的结果.

1　相关工作

1.1　逆渲染

逆渲染旨在通过估计图像的底层属性, 包括几何、材质和光照条件, 逆向求解图像变换. 逆渲染是一个欠定问题, 在优化过程中存在大量局部最优解, 该问题已经被研究了几十年.

为了使逆渲染问题变得非完全欠定, 许多方法[12-16]借助关于特定场景或物体的先验知识. SunStage[17]利用太阳照射下的人脸更好地估计材质和几何. 传统的逆向渲染流程的关键组成部分是可微分渲染器[18-22], 它将几何、材质和光源组合成输出图像; 然而, 为了应用这种可微分渲染器, 这些方法通常需要在完全或部分已知环境光照的受控环境中拍摄大量输入图像. 本文方法也包括一个可微分渲染器. 为了提高逆向渲染的质量, 本文从图像估计光源, 减少逆渲染的歧义性.

1.2　神经场景表示

文献[23-27]使用隐式神经场表示建模三维场景, 这些方法使用多层感知机(multilayer perceptron, MLP)表示三维场景, 实现了逼真的新视角合成和准确的三维重建, 取得优越的效果. 这些神经场景表示也被最近的逆渲染方法采用, 通过 MLP 建模

物体材质、物体几何和环境光照. 但是, 其中许多方法[5,28-29]仍然需要通过不同的共位手电筒照亮密集的输入图像.

1.3 面向逆渲染的神经场景表示

一些工作利用神经场景表示, 直接从未知光照下的多视图图像中恢复目标物体的材质, 以解决逆渲染中的欠定问题. NeRD[30]和 PhySG[7]将场景分解为由 MLP 编码的材质、几何和直接光照, 而 NeRFactor[6]额外考虑了可见性, 并使用一个双向反射分布函数(bidirectional reflectance distribution function, BRDF)以解耦阴影和材质. 为了实现高效的可见性计算, NeRV[31]利用神经可见性场减少已知环境照明条件下间接光照的计算成本; NeILF[32]使用入射光场表示直接光照和间接光照, 解决了未知照明条件下的光照表示问题. 为了更容易地进行光照优化, InvRender[8]将间接光照参数转化为球面高斯函数, 并基于此拟合目标场景的神经辐射场; NVDIFFREC[10]引入可微分的 Marching Tetrahedrons 算法, 从神经场景表示中提取几何形状, 并使用 Split Sum 算法[10]建模光照表示环境光照的高频细节; 进一步, NvDiffrecMC[11]结合蒙特卡罗光线追踪, 包括所有频率的光照和阴影, 并使用额外的去噪器处理蒙特卡罗积分产生的噪声. 本文方法也是直接使用蒙特卡罗采样解决逆渲染问题, 通过估计室内场景的光源并在光源位置上应用重要性采样, 减少了逆渲染中的歧义性.

2 本文方法

2.1 概 述

首先介绍本文方法中使用的 BRDF 模型; 然后对现有光照表示进行深入分析, 并讨论它们的歧义性; 为了解决这些歧义, 通过在室内场景中检测一组环境光源显式地建模高频的环境光, 以在渲染方程中表示阴影和反射, 从而实现物体材质参数的准确恢复. 给定一组多视图图像, 采用本文方法可恢复物体的材质属性和场景光照参数.

2.2 基于可微分渲染的材质获取

本文将物体的几何形状定义为三角网格, 并将物体的材质定义为 UV 映射到该网格上的纹理贴图. 将物体的材质属性定义为反射率 k_{d}、粗糙度 ρ 和切线空间法线 $\boldsymbol{t}$. 为了渲染像素颜色, 采用基于物理的渲染方法[9], 其定义为

$$L_o(\omega_o;x)=\int_{\Omega}L_i(\omega_i)f_r(x;\omega_o,\omega_i)(\omega_i\cdot\boldsymbol{n})d\omega_i.$$

其中, x 表示表面位置; ω_{o} 表示出射方向, 即观察方向的反方向. $L_{\mathrm{o}}(\omega_{\mathrm{o}};x)$ 通过对表面法线 $\boldsymbol{n}$ 周围的半球 Ω 的入射辐射度 $L_i(\omega_i)$ 沿方向 ω_i 进行积分得到.

本文采用 Cook-Torrance 模型[33]作为渲染公式中的 BRDF 函数 f_{r}, 其包含漫反射分量 f_{d} 和镜面反射分量 f_{s}, 定义为

$$f_{\mathrm{r}}(x;\omega_{\mathrm{o}},\omega_i)=f_{\mathrm{d}}+f_{\mathrm{s}}(x;\omega_{\mathrm{o}},\omega_i)=\frac{k_{\mathrm{d}}}{\pi}+\frac{D\cdot F\cdot G}{4(\boldsymbol{n}\cdot\omega_{\mathrm{o}})(\boldsymbol{n}\cdot\omega_i)}.$$

其中, 漫反射分量由漫反射反射率 k_{d} 控制. 本文采用微平面 GGX 作为镜面反射分量 f_{s}, 包含法线分布函数 D、菲涅尔项 F 和几何项 G, D 和 G 受表面粗糙度 ρ 控制. 更多细节详见文献[33].

为了使用这个渲染模型估计材质参数, 本文遵循 InvRender 方法[8]重建物体的几何形状, 然后优化材质和光照参数. 物体的几何形状采用三维重建方法 NeuS[26]获得, 首先通过优化符号距离场(signed distance field, SDF)并使用 Marching Cubes 算法提取三角网格; 然后使用 Blender[34]对重建的物体几何形状进行 UV 展开, 将表面映射到纹理贴图上. 包括反射率、粗糙度和切线在内的材质参数以纹理图像的形式存储, 并通过 UV 映射将其映射到物体网格上, 本文通过渲染损失函数优化这些材质参数. 在介绍如何在本文方法中优化这些材质参数之前, 首先分析现有神经逆渲染方法面临的歧义性问题, 以及无法准确地估计材质的原因.

2.3 已有方法面临的挑战

基于可微分渲染从一组多视图图像中优化材质参数是一个众所周知的病态问题. 在只有物体的观测情况下, 逆渲染需要将其分解为光照和材质参数, 因此在这个逆渲染过程中, 准确地建模光照具有重要的意义. 为了计算该方程, 现有方法大多采用遥远光源的假设, 因此受到光照-材质歧义的困扰.

遥远光源假设. 已有的一些工作[8,11,33]通常将光照表示为环境贴图, 假设光源与目标物体之间的距离是无限远的. 然而, 在室内场景中, 由于目标物体与光源之间的距离较近, 该假设是不成立的. 在室内场景中, 同一个光源在物体表面的不同位置可能具有不同的方向, 此时若直接使用环境贴图会使得光照恢复不准确, 进而导致材质估计的错误.

光照-材质歧义性. 非朗伯的粗糙物体的外观在不同视角下变化缓慢. 为了解释这种缓慢变化

的视角相关外观，一种材质参数估计情况会得到物体大的粗糙度；另一种情况是赋予物体小的粗糙度，并让环境光在不同方向上缓慢变化：这 2 种可能的情况会导致粗糙度估计中的歧义，因为粗糙的外观可以同时由粗糙的表面或低频光照来解释．本文通过实验观察到，一些方法[11,32]通常倾向于给物体赋予小的粗糙度，并根据光照环境调整颜色进行补偿．

2.4　基于光源的新光照表示

为了解决上述歧义，本文提出一种光照表示，明确地模拟室内场景中的高频光照，即光源．本文发现，室内环境中的关键光照组件就是光源，如图 1 所示．镜面视觉效果主要由这些光源引起，光源与物体之间的距离是有限的，因此违反了遥远光源的假设；同时，这些光源负责解释镜面效果，因此它们在确定这些物体的粗糙度方面起着至关重要的作用．基于此观察，本文提出一种空间变化的光照表示，显式地考虑室内场景的光源．



图 1　室内场景中的光源

通过蒙特卡罗积分求解基于物理的渲染方程中的积分．最终的着色颜色由 2 部分组成：光源的着色结果 $L_{\mathrm{o}}^{e}(\omega_{\mathrm{o}};x)$ 和环境贴图的着色结果 $L_{\mathrm{o}}^{a}(\omega_{\mathrm{o}};x)$，表示为

$$L_{\mathrm{o}}(\omega_{\mathrm{o}};x)=L_{\mathrm{o}}^{e}(\omega_{\mathrm{o}};x)+L_{\mathrm{o}}^{a}(\omega_{\mathrm{o}};x) \tag{1}$$

为了计算 $L_{\mathrm{o}}^{e}(\omega_{\mathrm{o}};x)$，本文在每个光源上均匀采样 N_{e} 个点，并对所有采样进行积分，以获得环境光源光照的渲染结果，公式为

$$\begin{cases} L_{\mathrm{o}}^{e_i}(\omega_{\mathrm{o}};x)=\sum\limits_{j=1}^{N_{\mathrm{e}}}\dfrac{L_{e_i}}{\left|x-x_{e_i}^{j}\right|^2}V(x;x_{e_i}^{j})\,f_{\mathrm{r}}(x;\omega_{\mathrm{o}},\omega_i)(\omega_j\cdot\boldsymbol{n}) \\ L_{\mathrm{o}}^{e}(\omega_{\mathrm{o}};x)=\sum\limits_{e_i\in E}L_{\mathrm{o}}^{e_i}(\omega_{\mathrm{o}},x) \end{cases} \tag{2}$$

其中，L_{e_i} 表示光源 e_i 的辐射强度，$x_{e_i}^{j}$ 表示光源 e_i 的随机采样点，可见性项 $V(x;\omega_j)$ 表示点 x 是否被点 ω_j 方向的光线遮挡，光源辐射强度 L_{e_i} 在训练过程中与材质参数一起进行优化．

为了计算 $L_{\mathrm{o}}^{a}(\omega_{\mathrm{o}};x)$，本文额外定义一个经纬度格式的低分辨率环境贴图来表示低频光照．对环境贴图进行重要性采样，通过积分半球面计算出射辐射度

$$L_{\mathrm{o}}^{a}(\omega_{\mathrm{o}};x)=\sum_{i=1}^{N_a}\frac{L_i^{a}(\omega_i)}{p(\omega_i)}V(x;\omega_i)\,f_{\mathrm{r}}(x;\omega_{\mathrm{o}},\omega_i)(\omega_i\cdot\boldsymbol{n}).$$

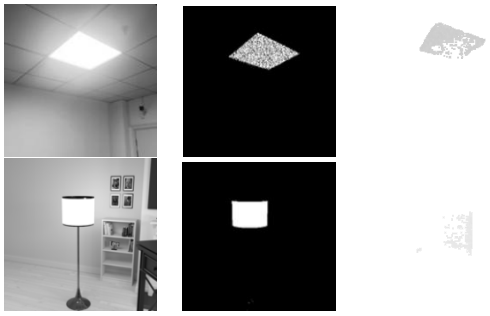
其中，$p(\omega_i)$ 和 $L_i^{a}(\omega_i)$ 分别表示环境贴图在 ω_i 处的概率分布函数和光照．

本文提出的光照表示明确考虑了光源和环境贴图的着色结果．通过追踪光线到光源，可以更准确地建模阴影和镜面反射，放宽了遥远光源的假设，并消除光照和材质之间的歧义性．

2.5　光源的定位

由于计算光源的方程需要对所有光源区域进行采样，因此本文方法需要定位场景中的所有光源．基于输入图像和重建场景提取光源后，再将物体图像匹配到场景中，以对齐光源和物体坐标系．

(1) 光源检测．为了通过光源模拟高频光照，本文使用输入的多视图图像和重建的场景几何信息检测光源．每幅 RGB 图像通过对其 R,G,B 通道求平均值转换为照明图像，采用蒙特卡罗采样生成二维光源掩模．通过随机采样 N_{c} 个像素并选择具有最高照明度的像素，获得一个光源像素；这个过程重复 K 次，生成一个二维光源掩模，如图 2 所示．通过渲染场景几何，获得输入二维图像的深度图．基于二维光源掩模和深度图，通过深度融合获得光源的点云；选择值大于 δ_{e} 的点，从融合点云中提取出三维光源点；利用 DBSCAN 算法[6]对光源点进行聚类，选择前 k 个聚类作为检测到的光源．



a. 场景图像　　b. 光源掩模　　c. 光源点云

图 2　定位场景光源

(2) 图像匹配. 由于检测到的光源位于场景坐标系中, 与物体坐标系不同, 因此本文需要对齐 2 个坐标系以在渲染过程中建模光源. 给定目标物体的三维重建模型, 首先将其投影到输入图像中生成物体掩模, 用于获得二维图像中的物体区域; 然后选择物体掩模面积较小的图像进行匹配, 因为这些图像拥有更多场景像素, 使得场景坐标系与物体坐标系的匹配过程获得更好的性能. 本文通过对场景图像执行 SfM 算法[35], 获得具有点描述符的稀疏点云, 并从物体图像中提取 SIFT 描述符[36]; 然后进行图像像素与三维点云的匹配; 最后通过 P*n*P 算法[37]计算每幅图像在场景坐标系中的相机姿态. 因为这些图像在物体和场景坐标系中都有相机姿态, 所以本文方法可以使用最小二乘法计算姿态对之间的全局变换.

2.6 模型训练

本文通过最小化渲染损失函数优化模型参数, 该函数定义为渲染像素颜色与观察到的像素颜色之间的 L_1 损失, 公式为

$$L_{\mathrm{RGB}}=\left|\gamma(\hat{I})-\gamma(I)\right|.$$

其中, γ 表示伽马校正算子, 损失函数在线性颜色空间中计算 2 个像素之间的残差.

(1) 平滑性损失. 本文采用平滑性先验规范反射率 k_{s}、粗糙度 ρ 和切线空间法线 $\boldsymbol{t}$ 的学习[7,11], 平滑性损失为

$$L_{\mathrm{s}}=\sum_{m\in\{k_{\mathrm{s}},\rho,\boldsymbol{t}\}}\lambda_m\left|m(x)-m(x+\varepsilon)\right|;$$

其中, x 表示表面位置, ε 表示一个小扰动, λ_m 是每种材质的权重.

(2) 环境损失. 本文方法约束环境贴图 M^{e} 的强度, 以鼓励光照由光源解释, 环境损失为

$$L_{\mathrm{e}}=\left\|M^{\mathrm{e}}\right\|^2.$$

最终的训练损失定义为

$$L=\lambda_{\mathrm{RGB}}L_{\mathrm{RGB}}+\lambda_{\mathrm{s}}L_{\mathrm{s}}+\lambda_{\mathrm{e}}L_{\mathrm{e}}.$$

其中, λ_{RGB}, λ_{s} 和 λ_{e} 表示每个损失的系数权重值.

3 实验结果及分析

3.1 实现细节

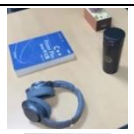
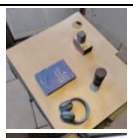
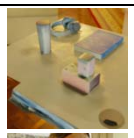
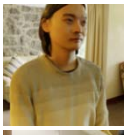
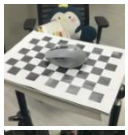
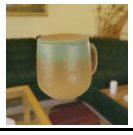
本文使用 Nvdiffrast[38]作为可微栅格化器, 从纹理图像中获取每个像素的物体材质, 并使用上文的基于物理的渲染公式渲染这些材质. 通过光线追踪引擎 optix[39]实现物体网格与光源的可见性测试. 本文方法对每个光源采样 56 个点, 对环境光采样 144 个样本; 每次迭代中, 在一幅图像中随机采样 2×10^5 个像素.

3.2 真实数据集实验

本文在室内光照下采集 4 个真实对象的多视角数据集, 包括 1 个人体、1 张桌子、1 个鼠标和 1 个杯子. 使用手持相机围绕每个目标物体进行 360°旋转拍摄. 每个物体在 960×540 像素分辨率下拍摄了 200 幅图像; 通过 COLMAP 恢复每幅图像的相机位姿; 通过 MonoSDF[40]获得目标室内场景的几何.

分别采用本文方法与 NvDiffrecMC 方法[11]进行实验, 结果如表 1 所示. 可以看出: (1) 本文方法在所有场景中实现了最逼真的结果; NvDiffrecMC 方法可能会使用法线建模反射率细节, 使其在训练

表 1 在真实数据集上 2 种方法的重光照结果

数据集	输入	本文			NvDiffrecMC[11]		
		重光照 1	重光照 2	重光照 3	重光照 1	重光照 2	重光照 3
桌面							
人像							
鼠标							
杯子							

光照上过拟合. 鼠标的标志在多个视角下不可见, 桌子上的盒子丢失了许多细节. (2) 本文方法可以恢复反射率的细节; NvDiffrecMC 方法恢复的物体粗糙度过低, 导致镜面光更强, 在人脸和桌子上最明显. (3) 本文方法的另一个优点是在保持凹凸贴图的同时产生更加平滑的结果, 如表 1 中的杯子所示.

3.3　合成数据集实验

为了在数值上评估本文方法的有效性, 额外创建了一个合成数据集. 首先加载一个三维室内模型, 创建一个合成的三维室内场景; 然后在该场景中放置 3 个不同结构和材质的物体. 对于每个物体, 使用 Blender Cycles 渲染器渲染 144 幅图像; 在这些渲染视图中, 均匀地采样 20 个测试视图, 并将其余视图用作训练视图来估计物体的材质.

实验中, 除了评估待估计的材质质量外, 还评估新视角合成质量和光照效果质量. 为此, 在每个物体上, 本文方法使用不同的光照条件渲染 20 幅图像, 以评估重光照效果的质量. 图像分辨率为 1 000×1 000 像素.

3.4　基准方法与评价指标

将本文方法与当前的神经逆渲染方法 NvDiffrecMC[11]和 InvRender[8]进行比较, 这 2 种方法都考虑了着色模型中的直接光照和间接光照.

由于反射率和光照之间存在不可避免的尺度模糊性, 因此实验中遵循之前的工作[6-8,11], 重光照之前, 在第 1 个测试视图上对预测结果和真实反射率进行对齐. 对于本文方法和 NvDiffrecMC 方法, 使用 Blender Cycles 渲染器以每个像素 1 024 个样本来渲染重光照图像; InvRender 方法遵循原文献实现, 使用 128 个球面高斯函数拟合目标环境贴图, 生成重光照图像.

评估指标. 本文方法使用峰值信噪比(peak signal to noise ratio, PSNR)、结构相似性指数(structural similarity index, SSIM)和基于学习的感知图像块相似性(learned perceptual image patch similarity, LPIPS)[41], 作为评估对齐反射率、重光照图像和新视角合成的图像质量的指标. 粗糙度使用均方误差(mean squared error, MSE)进行评估.

3.5　对比实验

表 2 和表 3 所示分别为在合成数据集上的定量和定性结果. 可以看出, 本文方法在估计的材质(粗糙度/反射率)、新视角合成质量和光照效果方面取得了显著优于基准方法的性能, 证明了其有效性.

需要注意的是, 表 2 中所有方法在新视角合成质量上都达到了合理的水平, 这是因为这些方法都能够匹配训练图像的颜色. 然而在本实验中, 基准方法难以准确地估计出反射率/粗糙度, 第 2.3 节

表 2　在合成数据集上 3 种方法的定量结果对比

方法	粗糙度	反射率			视角合成			重光照		
	MSE	PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS
InvRender[8]	0.063	20.49	0.938	0.332	23.24	0.969	0.085	20.49	0.938	0.332
NvDiffrecMC[11]	0.087	23.36	0.928	0.084	21.99	0.930	0.118	19.98	0.913	0.125
本文	**0.008**	24.87	**0.930**	**0.053**	**25.27**	0.969	0.056	**26.06**	**0.967**	**0.049**
本文 w/o Env	0.012	21.09	0.881	0.085	23.06	0.948	0.075	22.91	0.957	0.071
本文 w/o Emitter	0.043	24.09	0.929	0.067	22.88	0.965	0.072	22.42	0.952	0.081
本文 w/o Env loss	0.029	**25.49**	0.890	0.100	25.22	**0.969**	**0.055**	25.62	0.927	0.096

注. 加粗表示最优值.

表 3　在合成数据集上 3 种方法的重光照结果

真值	本文	NvDiffrecMC[11]	InvRender[8]	本文 w/o Emitter	本文 w/o Env loss
					

中讨论的材质-光照歧义性是可能造成该结果的原因. 在这种情况下, 这些方法倾向于预测较低的粗糙度, 并通过环境光照图解释物体的视角相关颜色. 图 3 所示为本文方法在合成数据集上的可视化结果. 可以看出, 本文方法明确地估计了光源的空间位置, 能够准确地恢复出正确的粗糙度和反射率.



a. 木杯 b. 黑鼠标 c. 棕色沙发 d. 场景环境

图 3 合成数据集上的可视化结果

3.6 消融实验

在合成数据集上对本文方法进行消融实验, 包括光源的检测、低分辨率环境贴图的利用以及新的环境损失, 结果分别如表 2 和表 4 所示.

表 4 对环境贴图影响的消融实验结果

方法	反射率	渲染结果	对齐后的反射率	重光照
本文 w/o Env				
本文				

3.6.1 光源的影响

光源的利用使得本文方法能够减少光照-材质的模糊性. 计算渲染结果时, 在式(2)中舍弃来自光源的光线. 从表 2 可以看出, 由于光照-材质的模糊性, 估计的反射率和粗糙度质量严重降低.

3.6.2 环境贴图的影响

低分辨率的环境贴图在解释低频环境光方面起着重要作用. 本文方法在式(1)中仅使用光源光照进行渲染. 从表 4 可以看出, 低分辨率环境贴图对于解释未直接受到光源照射的区域的颜色是必不可少的, 如果没有低分辨率环境贴图, 这些区域的反射率将过亮, 无法正确地呈现这些区域的颜色.

3.6.3 环境损失的影响

对本文提出的环境损失进行消融实验, 结果如表 5 所示. 可以看出, 如果没有环境损失, 粗糙度估计的质量会显著降低, 原因是环境损失抑制了低分辨率环境贴图上的高频光照, 从而强制将镜面光解释为光源光, 如果不使用环境损失, 高频光照可能出现在环境贴图上, 物体将变得更加平滑且反射率偏高.

表 5 合成数据集上的材质估计结果

数据	方法	粗糙度	反射率	渲染图像
1	真值			
	本文			
	NvDiffrecMC[11]			
2	真值			
	本文			
	NvDiffrecMC[11]			
3	真值			
	本文			
	NvDiffrecMC[11]			

4 结 语

本文提出一种逆向渲染方法, 用于从多视角图像中估计室内环境中物体的材质参数. 通过对现有的逆渲染方法进行深入分析, 提出一种光照表示方式, 明确考虑室内场景中的光源, 极大地提高了本文方法解释镜面视觉效果和正确估计材质的能力, 并展示了远光照假设和光照-材质歧义性对逆渲染方法性能的退化影响结果. 在合成和真实数据集上的实验结果表明, 本文方法能够消除歧义性, 准确地恢复物体材质, 实现物体的高质量重光照效果.

本文方法中采用的 BRDF 模型是一个 Cook-

Torrance 模型，不能处理各向异性反射等高级效果．虽然该方法能够准确地估计人体皮肤和衣物的材质，但难以对头发进行重光照，准确地建模头发材质可能需要额外的 BRDF 设计，这将是未来需要深入研究的问题．

参考文献(References):

[1] Debevec P, Hawkins T, Tchou C, *et al*. Acquiring the reflectance field of a human face[C] //Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2000: 145-156

[2] Guo K W, Lincoln P, Davidson P, *et al*. The relightables: volumetric performance capture of humans with realistic relighting[J]. ACM Transactions on Graphics, 2019, 38(6): Article No.217

[3] Nam G, Lee J H, Gutierrez D, *et al*. Practical SVBRDF acquisition of 3D objects with unstructured flash photography[J]. ACM Transactions on Graphics, 2018, 37(6): Article No.267

[4] Bi S, Lombardi S, Saito S, *et al*. Deep relightable appearance models for animatable faces[J]. ACM Transactions on Graphics, 2021, 40(4): Article No.89

[5] Zhang K, Luan F J, Li Z Q, *et al*. IRON: inverse rendering by optimizing neural SDFs and materials from photometric images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 5555-5564

[6] Zhang X M, Srinivasan P P, Deng B Y, *et al*. NeRFactor: neural factorization of shape and reflectance under an unknown illumination[J]. ACM Transactions on Graphics, 2021, 40(6): Article No.237

[7] Zhang K, Luan F J, Wang Q Q, *et al*. PhySG: inverse rendering with spherical Gaussians for physics-based material editing and relighting[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 5449-5458

[8] Zhang Y Q, Sun J M, He X Y, *et al*. Modeling indirect illumination for inverse rendering[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 18622-18631

[9] Kajiya J T. The rendering equation[C] //Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1986: 143-150

[10] Munkberg J, Chen W Z, Hasselgren J, *et al*. Extracting triangular 3D models, materials, and lighting from images[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 8270-8280

[11] Hasselgren J, Hofmann N, Munkberg J. Shape, light, and material decomposition from images using Monte Carlo rendering and denoising[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2022: Article No.1661

[12] Li Z Q, Shafiei M, Ramamoorthi R, *et al*. Inverse rendering for complex indoor scenes: shape, spatially-varying lighting and SVBRDF from a single image[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 2472-2481

[13] Gao D, Li X, Dong Y, *et al*. Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images[J]. ACM Transactions on Graphics, 2019, 38(4): Article No.134

[14] Li Z Q, Xu Z X, Ramamoorthi R, *et al*. Learning to reconstruct shape and spatially-varying reflectance from a single image[J]. ACM Transactions on Graphics, 2018, 37(6): Article No.269

[15] Guo Y, Smith C, Hašan M, *et al*. MaterialGAN: reflectance capture using a generative SVBRDF model[J]. ACM Transactions on Graphics, 2020, 39(6): Article No.254

[16] Ye W C, Chen S, Bao C, *et al*. IntrinsicNeRF: learning intrinsic neural radiance fields for editable novel view synthesis[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 339-151

[17] Wang Y F, Holynski A, Zhang X M, *et al*. SunStage: portrait reconstruction and Relighting using the Sun as a Light Stage[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 20792-20802

[18] Nimier-David M, Vicini D, Zeltner T, *et al*. Mitsuba 2: a retargetable forward and inverse renderer[J]. ACM Transactions on Graphics, 2019, 38(6): Article No.203

[19] Liu S C, Chen W K, Li T Y, *et al*. Soft rasterizer: a differentiable renderer for image-based 3D reasoning[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 7707-7716

[20] Chen W Z, Gao J, Ling H, *et al*. Learning to predict 3D objects with an interpolation-based differentiable renderer[C] //Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2019: Article No.862

[21] Chen W Z, Litalien J, Gao J, *et al*. DIB-R++: learning to predict lighting and material with a hybrid differentiable renderer[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.1749

[22] Kato H, Ushiku Y, Harada T. Neural 3D mesh renderer[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 3907-3916

[23] Park J J, Florence P, Straub J, *et al*. DeepSDF: learning continuous signed distance functions for shape representation[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 165-174

[24] Mildenhall B, Srinivasan P P, Tancik M, *et al*. NeRF: representing scenes as neural radiance fields for view synthesis[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 405-421

[25] Barron J T, Mildenhall B, Tancik M, *et al*. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press,

2021: 5835-5844

[26] Wang P, Liu L J, Liu Y, *et al*. NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction[C]// Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.2081

[27] Yariv L, Gu J T, Kasten Y, *et al*. Volume rendering of neural implicit surfaces[C] //Proceedings of the 35th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2021: Article No.367

[28] Bi S, Xu Z X, Sunkavalli K, *et al*. Deep reflectance volumes: relightable reconstructions from multi-view photometric images[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 294-311

[29] Schmitt C, Donné S, Riegler G, *et al*. On joint estimation of pose, geometry and svBRDF from a handheld scanner[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3490-3500

[30] Boss M, Braun R, Jampani V, *et al*. NeRD: neural reflectance decomposition from image collections[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 12664-12674

[31] Srinivasan P P, Deng B Y, Zhang X M, *et al*. NeRV: neural reflectance and visibility fields for relighting and view synthesis[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 7491-7500

[32] Yao Y, Zhang J Y, Liu J B, *et al*. NeILF: neural incident light field for physically-based material estimation[C] //Proceedings of the 17th European Conference on Computer Vision. Heidelberg: Springer, 2022: 700-716

[33] Cook R L, Torrance K E. A reflectance model for computer graphics[J]. ACM Transactions on Graphics, 1982, 1(1): 7-24

[34] Blender. Blender 4.2 LTS[EB/OL]. [2024-07-02]. https://www.blender.org

[35] Schönberger J L, Frahm J M. Structure-from-motion revisited[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 4104-4113

[36] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110

[37] Lepetit V, Moreno-Noguer F, Fua P. EP*n*P: an accurate $O(n)$ solution to the PnP problem[J]. International Journal of Computer Vision, 2009, 81(2): 155-166

[38] Laine S, Hellsten J, Karras T, *et al*. Modular primitives for high-performance differentiable rendering[J]. ACM Transactions on Graphics, 2020, 39(6): Article No.194

[39] Parker S G, Bigler J, Dietrich A, *et al*. OptiX: a general purpose ray tracing engine[J]. ACM Transactions on Graphics, 2010, 29(4): Article No.66

[40] Yu Z H, Peng S Y, Niemeyer M, *et al*. MonoSDF: exploring monocular geometric cues for neural implicit surface reconstruction[C] //Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc, 2022: Article No.1814

[41] Zhang R, Isola P, Efros A A, *et al*. The unreasonable effectiveness of deep features as a perceptual metric[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 586-595