摘要

场景光照解析技术，作为在虚拟现实、增强现实等应用中保障真实感体验的关键技术，已成为计算机视觉与图形学交叉领域的核心课题。由于专业光照测量仪器在实际场景应用中面临的诸多限制，基于低成本且广泛可得的图像对场景在成像模型中对应的环境光照、物体材质、物体几何等构成要素进行解析成为主流研究范式，具有较高的应用潜力与价值。然而，由于室外场景的极高动态范围特性，以及由于局部物体遮挡、相互反射等光传播现象引起的环境光照局部可变特性，目前室外场景光照估计的效果亟待提升。针对这一问题，本文研究基于解耦空间光照参数化表示的室外场景局部可变光照估计算法，实现了对室外场景中存在的局部可变光照的准确估计，且光照表示具有高度可编辑性。针对在图像编辑任务中缺少光照解析约束导致编辑结果缺少真实感的问题，本文进一步将单图室外场景光照感知能力引入代表性的图像条件重绘制方法中，能有效将物理真实的光照效果体现在编辑结果中。由于相机投影的信息损失，单视角图像难以摆脱严重的光照解析二义性。针对这一问题，本文研究针对类镜面场景的多视角图像三维重建与逆渲染算法，对存在高频局部可变镜面反射纹理的场景实现了高质量三维重建与逆渲染解析效果。本文的创新点具体包括以下几个方面：

（1）针对室外复杂遮挡场景下光照条件存在的局部可变特性与高动态范围特性，本文研究对室外场景局部可变光照不同成分的分析，依此提出对应的解耦隐空间编码光照建模，并通过端到端神经网络在光照编码空间中进行光照估计，实现了更加准确的室外场景局部可变光照估计。为此，本文首先对室外场景局部可变环境光照进行了分析，将其分解为对应天空穹顶光照的全局一致部分和对应地面局部物体表观的局部可变部分，并进一步解耦为太阳位置、太阳光照、天空光照、局部可变内容等成分，该表示具有参数紧凑性和可编辑性。对应地，本文提出了室外场景局部可变光照估计算法SOLD-Net，其网络框架中全局光照编码—解码器和局部光照编码—渲染器被用于学习不同光照成分对应的神经网络编码隐空间，而光照估计网络被用于在所学习到的光照编码空间表示中进行局部可变光照估计。为了应对高质量训练数据的匮乏，本文通过物体材质增强和光照条件增强等数据增强策略得到了多样化的高仿真合成数据集，并且采集了首个带有完整高动态范围的真值光照标注的真实室外场景局部可变光照数据集。为了验证算法的有效性，本文在合成数据集和真实场景数据上进行了全局光照估计和局部光照估计等实验评估，实验结果显示 SOLD-Net 较对比方法在太阳位置估计、重光照准确性等效果取得明显提升。此外，本文还展示了光照表示的可编辑性、光照估计网络的可泛化性、光照估计结果的虚拟物体插入应用展示。

（2）针对日益增长的图像可控编辑需求与光照解析约束在图像编辑中的应用不足，

本文研究在图像编辑技术中引入单图场景光照感知能力，依此提出对应的光照解析约束的图像编辑任务，通过基于物理原理的方式在图像编辑过程中引入约束，实现了光照真实和谐的室外场景图像编辑。为此，本文以图像条件重绘制为基础，提出了光照感知的图像条件重绘制任务建模和对应神经网络算法 LuminAIRe。本文以轻量化参数建模作为光照表示，以基于学习的方法引入单图室外场景光照感知能力，进而从输入背景图像区域估计场景三维光照条件，并类似地从输入重绘制区域场景解析图估计对应三维几何。为了将三维空间中的光照解析约束应用于二维图像域，本文以基于物理的渲染方程作为成像模型，通过预设标准材质组得到光照候选图，包含重绘制区域可能对应的光照效果。本文通过光照注意力模块从用户指定的材质属性条件和场景解析图得到光照系数图，并计算得到最终用于图像编辑的光照图。为增强对不同颗粒度场景解析图的鲁棒性，本文在网络训练过程中使用了多层级语义标签增强的设计。本文通过在室外场景图像中批量进行仿真车辆模型的虚拟物体插入采集了合成数据集 CarLuminAIRe，在该数据集上进行的实验评估以及用户研究均显示 LuminAIRe 相较于未引入光照解析约束的对比方法在光照真实感和和谐性上取得明显提升。此外，本文还展示了 LuminAIRe 的可泛化性，以及对不同材质条件、不同几何条件、带噪声场景解析图的鲁棒性。

（3）针对三维场景完整重建与逆渲染解析的现实需求与单视角图像观测下存在严重场景光照材质几何二义性的问题，本文研究利用多视角图像进行更完整的场景光照解析，依此提出类镜面场景表面着色模型，并通过在可导渲染优化框架中引入场景间接光照建模，实现了类镜面场景的三维重建与解析。为此，本文以三维高斯泼溅框架作为可导渲染工具，应对类镜面场景中临近物体产生的高频镜面反射纹理重建与解析提出了 SpecTRe-GS 算法。本文采用三维高斯点云作为场景表达，使用可优化的环境光图作为直接光照分量建模，通过光栅化渲染从三维高斯单元中的几何属性与着色属性累积得到相机视角下的场景几何与表面材质。对于场景中的类镜面表面，本文分别建模漫反射分量和镜面反射分量，其中视角无关的漫反射分量和基准镜面反射率直接存储在高斯单元中。本文实现了高效的三维高斯点云光线追踪渲染器，用于查询依赖视点位置和角度的场景入射光照，查询结果在图像域上进行完整的表观渲染以计算重建误差并通过可导渲染框架优化场景表示。本文使用了先验引导的渐进式高斯点云训练优化策略，在训练优化初期使用先验快速得到初步几何初始化，并在训练后期通过光照几何联合优化得到更加精细的重建与解析结果。本文在合成数据集与真实数据上进行了新视角合成和光照解析实验，结果表明 SpecTRe-GS 能取得更好的几何重建质量与高频镜面反射纹理的重建。此外，本文还展示了 SpecTRe-GS 对不同粗糙度物体的重建效果、镜面反射效果的多视角连续性、用于三维场景编辑的应用效果。

关键词：光照估计，逆渲染，成像模型，室外场景

Research on Image-based Outdoor Scene Illumination Decomposition and Analysis

Jiajun Tang (Computer Application Technology)
Supervised by Prof. Boxin Shi

ABSTRACT

Scene illumination decomposition and analysis, as a key technology for ensuring realistic experiences in virtual reality (VR) and augmented reality (AR) applications, has become a core research topic in the interdisciplinary field of computer vision and computer graphics. Due to the limitations of professional illumination measurement instruments in practical scenarios, image-based methods for decomposing scene illumination as environmental lighting, object materials, and geometry for analysis within the physics-based imaging formation model have emerged as a mainstream research paradigm, offering significant application potential and value. However, current outdoor scene lighting estimation faces critical challenges: the extremely high dynamic range (HDR) characteristics of outdoor scenes, as well as spatially-varying lighting caused by local object occlusions and inter-reflections in light transport phenomena, lead to suboptimal estimation accuracy. In response to this problem, this thesis proposes a spatially-varying lighting estimation algorithm with a disentangled parametric representation of lighting, achieving accurate estimation of spatially-varying lighting in outdoor scenes while maintaining high editability. Furthermore, in response to the lack of photorealism in image editing tasks due to insufficient illumination-aware constraints, this thesis integrates single-image outdoor scene illumination decomposition and analysis into representative conditional image repainting methods, effectively embedding physically plausible illumination effects into edited results. Finally, in response to the severe ambiguity in illumination decomposition and analysis caused by information loss in single-view camera projections, this thesis develops a multi-view image-based illumination decomposition and analysis algorithm for highly specular scene reconstruction, enabling high-quality 3D reconstruction and inverse rendering for scenes with high-frequency spatially-varying specular reflection content. The main contributions of this thesis include:

(1) In response to the spatially varying characteristics and extremely high dynamic range of illumination conditions in outdoor scenes with complex occlusions, this thesis explores

the decomposition and analysis of different components in spatially varying outdoor lighting. Based on this analysis, a disentangled latent space encoding framework is proposed for lighting modeling, and an end-to-end neural network is designed to perform lighting estimation within the learned latent space, achieving more accurate spatially varying lighting estimation for outdoor scenes. To this end, this thesis first analyzes the spatially varying environmental lighting in outdoor scenes, decomposing it into a spatially-uniform component corresponding to sky dome and a spatially-varying component corresponding to local scene appearances. These components are further disentangled into interpretable parameters, including sun position, sun light, sky light, and spatially-varying local content, forming a compact and editable representation. Correspondingly, this thesis proposes the SOLD-Net for spatially-varying outdoor lighting estimation with disentangled representation. The network architecture includes a global lighting encoder-decoder and a local content encoder-renderer to learn latent spaces for different lighting components, and a Spatially-varying lighting estimator that predicts spatially-varying lighting within the learned latent spaces. To address the scarcity of high-quality training data, this thesis employs data augmentation strategies, such as material diversity augmentation and lighting condition augmentation, to generate diverse synthetic datasets with high photorealism. Additionally, the first real-world HDR dataset with comprehensive ground-truth annotations for spatially-varying outdoor lighting is collected. Experimental evaluations on both synthetic and real-world datasets demonstrate the effectiveness of the proposed method. Results show that SOLD-Net significantly improves over baseline methods in sun position estimation and object relighting accuracy. Furthermore, this thesis showcases the editability of the disentangled lighting representation, the generalizability of the lighting estimation network, and practical applications such as virtual object insertion under estimated lighting conditions.

(2) In response to the growing demand for controllable image editing and the insufficient integration of illumination-aware constraints in existing methods, this thesis explores the incorporation of scene illumination decomposition and analysis into image editing workflows, aiming to achieve photorealistic and illumination-harmonized editing in outdoor scenes by introducing physics-based constraints derived from illumination decomposition and analysis. Building upon conditional image repainting, this thesis proposes the LuminAIRe framework for illumination-aware conditional image repainting. LuminAIRe employs a lightweight parametric illumination modeling for estimating 3D scene lighting conditions from background image regions with learning-based priors and analogously estimate 3D scene geometry from user-specified scene parsing masks. To bridge 3D lighting and geometry constraints with 2D image editing, LuminAIRe implements a physics-based rendering pipeline that generates il-

lumination candidate images using predefined standard materials, and an illumination attention module that dynamically weights these candidates and gives the final illumination image based on user-specified material properties and scene parsing masks. For enhanced robustness, the framework incorporates hierarchical labeling enhancement during training to handle varying granularities of input scene parsing masks. A synthetic dataset, CAR-LUMINAIRE, is constructed by inserting virtual vehicle models into real outdoor scenes, providing paired data for training and evaluation. Experimental results and user studies demonstrate LuminAIRe's superior performance in illumination realism and visual harmony compared to constraint-free editing methods. Additional analyses verify LuminAIRe's generalization capability across diverse material/geometric conditions and its robustness to noisy input scene parsing masks.

(3) In response to the practical requirements for complete 3D scene reconstruction and inverse rendering, as well as the severe ambiguities in lighting, material, and geometry caused by single-view observations, this thesis explores the use of multi-view images for more comprehensive illumination decomposition and analysis. To achieve 3D reconstruction and decomposition of highly specular scenes, this thesis proposes a shading model for highly specular surfaces and integrates indirect lighting into a differentiable rendering optimization framework. To this end, this thesis proposes SpecTRe-GS to model highly Specular surfaces with reflected nearby objects through tracing rays in the 3D space within the Gaussian splatting framework. SpecTRe-GS represents scenes using 3D Gaussian point clouds, models direct illumination components with optimizable environment maps, and accumulates geometric and shading attributes from 3D Gaussian units rasterization to derive normal maps and material maps in camera views. For highly specular surfaces, diffuse and specular reflections are separately modeled: the view-independent diffuse color is directly stored as attributes in Gaussian units. This thesis implements an efficient ray tracer for querying view-dependent incident lighting in 3D Gaussian point clouds, with final rendered images used to compute reconstruction errors and optimize scene representations through the differentiable framework. A prior-guided progressive training strategy is employed, where geometric initialization is rapidly achieved using monocular normal priors during early steps, followed by joint lighting-geometry optimization in later steps for refined reconstruction and decomposition. Experiments on synthetic and real-world datasets demonstrate that SpecTRe-GS achieves superior geometric reconstruction quality and accurate high-frequency specular reflection recovery compared to baseline methods. This thesis also showcases SpecTRe-GS's effectiveness in multi-view consistency of specular effects and practical applications in 3D scene editing.

KEY WORDS: Lighting estimation, Inverse rendering, Image formation model, Outdoor scene