Workshop on Material Appearance Modeling
MANER Conference - Material Appearance Network for Education and Research (2024)
J. Y. Hardeberg and H. Rushmeier (Editors)

The Challenges of Relighting from Multi-View Observations

Daniel Siersleben[1], Rodrigo Ortiz-Cayon[1], Klemen Istenic[1], Yusuke Tomoto[1] and Simone Schaub-Meyer[2]

[1]Fyusion Inc. [2]Technical University of Darmstadt

Abstract

High-quality geometry reconstruction from multi-view images with subsequent appearance decomposition into the physical shading components could enable a seamless integration of neural reconstructions into the modern rendering workflow. While 3D reconstruction techniques have steadily improved, the task of inverse rendering by decomposing an appearance into lighting effects and material properties remains fundamentally ill-posed and highly ambiguous. We show that current state-of-the-art inverse rendering approaches fail to accurately recover material properties, significantly impacting relighting quality. Furthermore, we demonstrate that existing evaluation methods, which rely on image-based metrics, do not adequately capture the reconstruction quality in novel lighting conditions. Our findings illustrate the dependence of current systems on simplified setups with predefined illumination, which are necessary to reliably disentangle light and material contributions and to ultimately achieve convincing relighting.

1. Introduction

Recent advances in multi-view reconstruction, particularly centered around novel-view synthesis, optimize representations that entangle view-dependent lighting, material interactions, and object color [MST*21, KKLD23]. Given multiple observations of the same object and the corresponding camera poses, these systems can synthesize novel views of the object within the *same* static environment. Addressing the broader problem of inverse rendering, fully-fledged differentiable rendering pipelines such as *NvDiffRec* [MHS*22, HHM22] achieve comparable synthesis quality for novel viewpoints. Unlike the implicit representations produced by NeRF-like approaches [MST*21, VHM*22], the inverse rendering framework provides physically interpretable components that can be directly integrated into modern forward rendering workflows. This allows for rapid prototyping for 3D artists and facilitates real-world applications such as scene relighting or appearance editing in 3D. In contrast, entangled material-light estimations negatively affect the usability, as 3D animations require an accurate estimation of the material decomposition to capture the object's true appearance even in different lighting scenarios.

In the following work: *(i)* We investigate the usefulness of cutting-edge inverse rendering pipelines for relighting given multi-view images. *(ii)* We show that state-of-the-art inverse rendering pipelines struggle even in simplified scenarios with accurate material-light decomposition and argue that the ill-posed problem formulation is a fundamental challenge for material estimation. *(iii)* We discuss possible approaches and requirements necessary for convincing relighting on real data.

Figure 1: *Challenges of relighting: After material and light estimation,* (left) *rendering novel view in estimated light looks reasonable. However,* (middle) *erroneous disentanglement of light and material degrades the reconstruction when relighted.* (right) *Reference. Used renderer:* NvDiffRec *[MHS*22].*

2. Inverse Rendering and the Problem of Ambiguity in Light-Material Interaction

Inverse rendering revolves around the reconstruction of 3D shape, object decomposition, and lighting of a scene's content based on a collection of 2D images. Fully differentiable simulations of the rendering equation [Kaj86] are used to describe each pixel based on an explicit representation of the 3D geometry, material decomposition described by a Bidirectional Scattering Distribution Function (BSDF) in addition to some representation of the surrounding illumination. By integrating over the entire hemisphere Ω surrounding

DOI: 10.2312/mam.20241181

the surface normal $\mathbf{n}$,

$$L(\omega_o) = \int_\Omega L(\omega_i) f(\omega_i, \omega_o)(\omega_i \cdot \mathbf{n})\, d\omega_i, \tag{1}$$

the outgoing radiance $L(\omega_o)$ depends on the surface interaction described by the BSDF $f(\omega_i, \omega_o)$ and the amount of incoming light $L(\omega_i)$.

Real-time rendering involves frequent evaluations of Equation 1 which requires efficient approximations and simplifications such as Image Based Lighting (IBL) where direct illumination is integrated by sampling from a cube map texture, capturing surrounding light information. In the inverse setting, the forward process has to be fully differentiable to allow for gradient-based optimization. To handle the extensive sampling needed for parameter optimization, inverse rendering pipelines often utilize similar approximations as those used in real-time rendering [MHS*22].

Fitting a novel scene with the inverse rendering pipeline is done by minimizing a simple image-space loss $\mathcal{L}\left(\hat{I}_{G,M,L}(p), I_{\text{ref}}(p)\right)$ between the reference image I_{ref} and the image $\hat{I}_{G,M,L}(p)$ obtained through sampling the rendering equation based on the provided camera position p, where the geometry G, material properties M, and global illumination L are jointly optimized.
Due to the complex interactions described by the rendering equation, resulting pixel values can be traced back to a multitude of effects, making the inverse rendering problem highly ill-posed. Specifically, given a single view or a limited number of observations, separating an object's color from chromatic effects caused by environmental lighting becomes a challenging task. For example, one might perceive an object as a certain color because *(a)* the object's diffuse albedo exhibits that color, or *(b)* the environment emits light in a specific tone, and the interaction with the object's properties makes it appear in the observed color. This ambiguity ultimately leads to entangled representations in inverse rendering engines, where material properties leak into the estimated environment and vice versa. For novel views within the same static scene – where material and lighting conditions remain consistent – the estimated components might produce decent and plausible results but degrade significantly as soon as one component changes, for example the lighting, as depicted in Figure 1.

To address the inherent ambiguity in reconstructing illumination from single image observations, [BM15] treats the problem as a case of statistical inference, aiming to optimize the most likely scenario of shape and composition prevalent in natural images. With multiple observations from different views and the corresponding camera positions, the problem, especially shape reconstruction, becomes less ambiguous but still remains challenging.

However, for realistic relighting of the reconstructed object, high accuracy in decomposing the images into contributions made by *(i)* the environmental illumination and *(ii)* the object color is essential. Metrics considered in most works on multi-view reconstruction, even those offering relighting capabilities, place great importance on image-based comparisons within the same static scene, which hardly capture the estimation quality of individual shading components. As a direct result, common inverse rendering applications excel at shape reconstruction, which is necessary for all downstream rendering tasks given the importance of accurate normal estimation in Equation 1. However, the emphasis set by image-based metrics within the static scene tends to overlook deficiencies in the material-light decomposition, which is crucial for convincing relighting and remains an unresolved challenge in state-of-the-art approaches to inverse rendering.

3. Material Decomposition from Multi-View Images

As a leading model for neural inverse rendering, we investigate the material decomposition quality of *NvDiffRec* [MHS*22] and its successor *NvDiffRecMC* [HHM22], which incorporates more sophisticated light interaction using Monte Carlo importance sampling. Both closely resemble the common PBR pipeline and directly optimize textures for material and lighting, allowing for straightforward comparison and assessment of material estimation. To evaluate the quality of the separation between material and illumination, we use simplified synthetic scenes based on the Shiny-Blender dataset [VHM*22]. To accurately measure material estimation capabilities, we assume perfect information about the geometry and only estimate texture maps for diffuse and specular material components, as well as the environment lighting. Furthermore, the material properties for each reference scene are strongly simplified, and each image is rendered using the exact same rendering process that *NvDiffRec* optimizes. This setup ensures that the reconstruction approach can reproduce the same visual effects present in the reference and allows us to quantitatively compare the estimated material components to the ground truth, as otherwise material components might not directly correspond between different rendering implementations.

In Figure 1, we illustrate a rendered image of a simple coffee cup in the novel-view synthesis setting, a novel viewpoint shaded in the same lighting the scene was optimized in, compared to a relighting setting, where we render the image with a novel environment light. As a comparison, we also relight the reference object in the same novel environment. One can directly observe that the appearance significantly differs between the view synthesis in the previously estimated environment and the novel illumination. The reference relighting provides insight into how the object should appear under different lighting conditions. To highlight the cause of the distorted relighting appearance, we visualize the material decomposition obtained from *NvDiffRec* and *NvDiffRecMC* on two simple shapes in Figure 2. Both approaches fail to faithfully capture the material properties of the object, resulting in environment artifacts baked into the object's material properties and color leaking from the object into the estimated light probe. The entanglement between light and object leads to bright colored estimates of the diffuse albedo since the estimated light already contributes part of the dark red color the object appears in. This entanglement leads to the undesired appearance changes visible once the scene is rendered in novel lighting conditions.

Inside the recovered material textures, we observe two distinct directions in which the ambiguity of light-material contribution becomes evident:

- **Material correctness:** Whether the estimated material values converge towards the correct properties used to render the scene.
- **Material attribution:** Whether objects or regions belonging to the same material are assigned similar material properties.

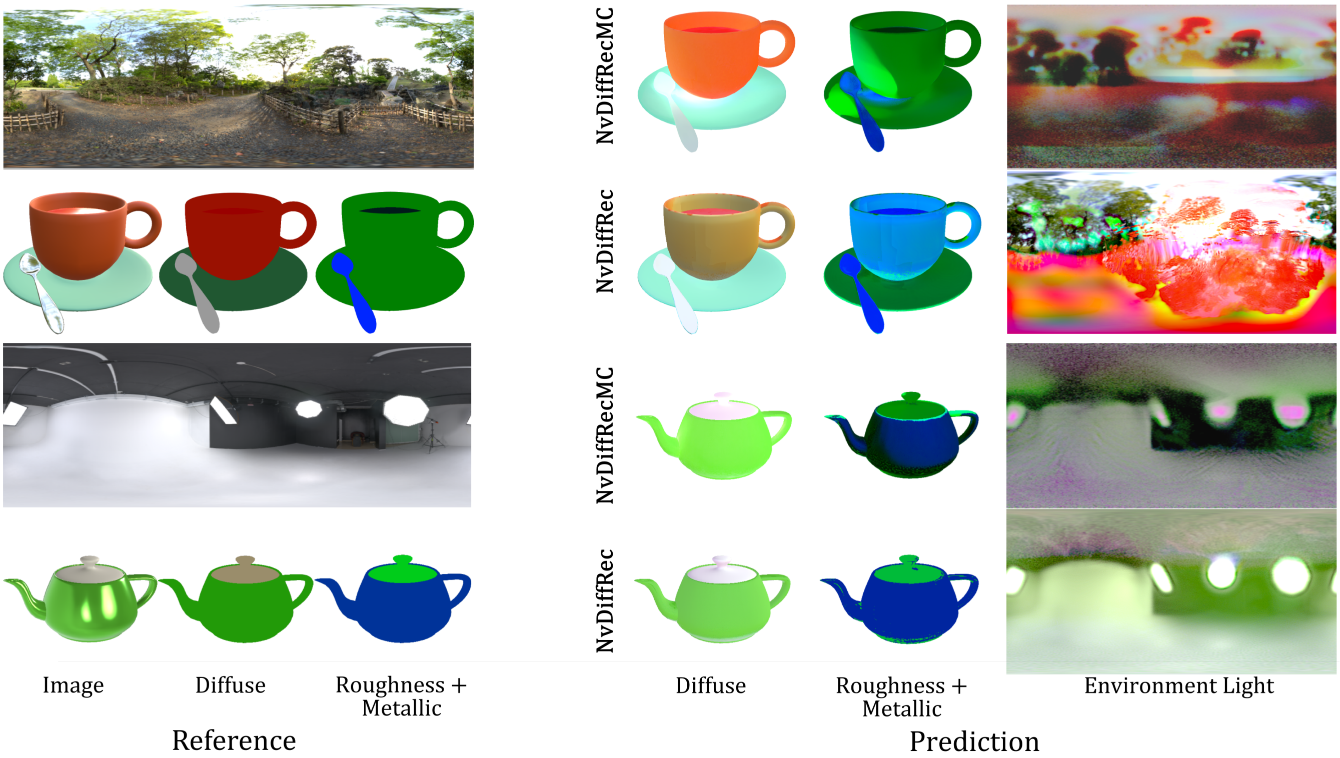

Figure 2: *Material reconstruction obtained with [MHS*22, HHM22] for a simple synthetic scene based on the ShinyBlender dataset [VHM*22]. We render the reference scene with the same rendering pipeline used by NvDiffRec to obtain the corresponding ground truth material values and provide the reference geometry for reconstruction. The material components* $(\emptyset, \mathbf{r}, \mathbf{m})$ *with roughness* $\mathbf{r}$ *and metallic* $\mathbf{m}$ *are visualized as RGB image.*

For the coffee cup, *NvDiffRec* fails to recognize the correct composition of the cup: no contribution of the metallic component and a high roughness. Instead, it attempts to explain the effect with a metallic component and coloration baked into the estimated environment light. While *NvDiffRecMC* performs slightly better on this particular scene, we still observe regions likely perturbed by the light, leading to incorrect material assessment. Regarding material association, both approaches struggle to maintain consistent material estimations in uniform regions, especially visible in the coffee cup sample towards the center of the saucer and the bottom of the cup itself.

(a) *Vanilla* **(b)** *Provided light* **(c)** *Shared light*

Figure 3: *Recovered lighting and specular texture,* $(\emptyset, \mathbf{r}, \mathbf{m})$ *triplet visualized as RGB image, in three scenarios: (a) Vanilla NvDiffRec reconstruction; (b) Material reconstruction with ground truth lighting known a priori; (c) Joint reconstruction of two objects sharing the same environment light.*

4. Towards Convincing Relighting for Real World Captures

In light of the apparent ambiguity, recent works addressing relighting for real-world captures consider specific image domains [VGD*24], simplified representations of the rendering equation [VHM*22, TDMS*23], or require more complex capturing setups [RES*22, PEL*21] that include information about the environmental lighting conditions. Addressing material correctness requires disentangling light information from the inherent properties of the object. By introducing various priors on the learnable shading components and implicitly treating the problem as an instance of statistical inference, *NvDiffRec* provides a toolbox that achieves reasonable reconstruction results for material estimation on multi-view images in the general case without any further information about the scene. For regularization on lighting conditions, *NvDiffRec* prioritizes solutions with white, colorless light, while *NvDiffRecMC* employs a monochromatic loss term that favors structural information over color information from the estimated light [MHS*22,HHM22]. While introducing more priors can further stabilize the approach and improve material estimation, devising general priors that are suitable for most 3D scenes is challenging. Priors addressing light and material decomposition tend to favor specific configurations, which can easily fail in scenarios where these assumptions do not hold. Although numerous, possibly scene-dependent, priors are necessary to achieve good separation of materials, this approach is akin to guessing the specific conditions under which the scene was captured.

Instead of using priors, an alternative to disentangle the light-material interaction could be achieved for example by capturing additional information about global illumination, thereby loosening the ill-posed nature of the problem. In Figure 3 we visualize the recovered textures for different settings. While the vanilla reconstruction (Figure 3a) converges to entangled representations, the ill-posed problem of reconstruction becomes well defined upon providing the correct lighting information (Figure 3b) upfront. In practice this could be achieved for example by using an additional 360° capture of the environment or some form of color lookup to calibrate the estimated light contribution. Further, we observe that having captures of multiple objects exposed to the same environment could also be helpful. With similar reconstruction quality, we show in (Figure 3c) that the shared information between two scenes, teapot and coffee cup, captured under the same environment map is sufficient for *NvDiffRec* to converge to the correct texture values. Here, we have a unique set of material textures for each object but a single, shared lighting environment is jointly optimized and used for the rendering of both scenes. Outlined in [HHM22], more accurate estimates during rendering can also improve material reconstruction quality as a more sophisticated sampling approach for the light interaction can narrow the range of possible materials that result in a visibly similar outcome. However, switching to Monte Carlo estimates with *NvDiffRecMC* has only shown insignificant quality improvements in our experiments with slightly more visible noise inside lighting estimates, *e. g.*, noticeable within the coffee cup in Figure 2.

Regarding material attribution, we find that optimizing fully spatially varying BRDF components can degrade the quality of material separation. Similar to [SMM20], who observe entanglement between material components partly caused by the nature of gradient descent, we notice similar tendencies in material estimation due to changing appearances. Changes in appearance affect gradients for all components involved in shading. Therefore, even with simplified synthetic data in uniform regions, where appearance changes can be fully explained by illumination, the inverse renderer converges to solutions with continuous changes across all material components, observable in both scenes depicted in Figure 2. This issue is further exacerbated with real-world data. To address this, we propose a stage of material clustering or texture smoothing to simplify the solution and to constrain the material variance during optimization.

5. Conclusions

We investigate the shortcomings of current inverse rendering approaches in relighting scenarios. We identify the ambiguity between light and material as a major challenge for accurate material decomposition while emphasizing the necessity of high-quality material separation for effective relighting. Current evaluations of relighting systems predominantly use image-based metrics that capture the quality within the same scene. When relighting is assessed, it is typically done using the same metrics in novel environments. We argue that this approach is insufficient to fully capture relighting quality because it does not individually assess material separation, which may work well in the captured environment but fail in others. Therefore, we suggest evaluating material-light separation quality as an additional and independent metric. We also recognizing that current systems require simplified setups with upfront information about illumination to achieve convincing relighting due to the fundamental ambiguity persistent in the inverse rendering formulation.

References

[BM15] Barron J. T., Malik J.: Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell. 37*, 8 (2015), 1670–1687. 2

[HHM22] Hasselgren J., Hofmann N., Munkberg J.: Shape, light, and material decomposition from images using Monte Carlo rendering and denoising. In *NeurIPS* (2022). 1, 2, 3, 4

[Kaj86] Kajiya J. T.: The rendering equation. *SIGGRAPH Comput. Graph. 20*, 4 (1986), 143–150. 1

[KKLD23] Kerbl B., Kopanas G., Leimkühler T., Drettakis G.: 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph. 42*, 4 (2023). 1

[MHS*22] Munkberg J., Hasselgren J., Shen T., Gao J., Chen W., Evans A., Müller T., Fidler S.: Extracting triangular 3D models, materials, and lighting from images. In *CVPR* (2022), pp. 8280–8290. 1, 2, 3

[MST*21] Mildenhall B., Srinivasan P. P., Tancik M., Barron J. T., Ramamoorthi R., Ng R.: NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM 65*, 1 (2021), 99–106. 1

[PEL*21] Pandey R., Escolano S. O., Legendre C., Häne C., Bouaziz S., Rhemann C., Debevec P., Fanello S.: Total relighting: Learning to relight portraits for background replacement. *ACM Trans. Graph. 40*, 4 (2021). 3

[RES*22] Rudnev V., Elgharib M., Smith W., Liu L., Golyanik V., Theobalt C.: NeRF for outdoor scene relighting. In *ECCV* (2022). 3

[SMM20] Saryazdi S., Murphy C., Mudur S. P.: The problem of entangled material properties in SVBRDF recovery. In *MAM@ EGSR* (2020), pp. 5–8. 4

[TDMS*23] Toschi M., De Matteo R., Spezialetti R., De Gregorio D., Di Stefano L., Salti S.: ReLight My NeRF: A dataset for novel view synthesis and relighting of real world objects. In *CVPR* (2023), pp. 20762–20772. 3

[VGD*24] Violante N., Gauthier A., Diolatzis S., Leimkühler T., Drettakis G.: Physically-based lighting for 3D generative models of cars. In *Comput. Graph. Forum* (2024). 3

[VHM*22] Verbin D., Hedman P., Mildenhall B., Zickler T., Barron J. T., Srinivasan P. P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *CVPR* (2022), pp. 5481–5490. 1, 2, 3