Check for updates

DOI: 10.1111/cgf.70592

COMPUTER GRAPHICS *forum*
Volume 0 (2026), number 0, e70592

Recent Trends in Inverse Rendering

S. Ullah,[1] F. Pellacini[2] and A. Giachetti[1]

[1]University of Verona, Verona, Italy
{shakir.ullah, andrea.giachetti}@univr.it
[2]University of Modena and Reggio Emilia, Modena, Italy
fabio.pellacini@unimore.it

Abstract

Inverse rendering is the problem of recovering scene geometry, material properties, and illumination from one or more observed images. It has experienced rapid progress in recent years, driven by advances in differentiable rendering, neural scene representations, and generative priors. This paper presents a comprehensive review of inverse rendering methods published between 2020 and 2025. In contrast to prior surveys on neural rendering, which review methods primarily focused on novel-view synthesis without explicit material and lighting decomposition, and intrinsic image decomposition surveys, which review approaches for reflectance and shading separation at the image level, we focus on methods that aim to solve the complete inverse rendering problem. We organize the surveyed approaches using a structured taxonomy that categorizes methods by input and output representations, computational strategies, and evaluation protocols. We further analyze commonly used datasets, evaluation strategies, and performance trade-offs, and discuss the strengths and limitations of existing approaches in terms of accuracy, efficiency, and generalization. Finally, we identify open challenges, including the ill-posed nature of material and illumination disentanglement, the lack of standardized benchmarks for joint evaluation, and the limited exploration of downstream applications such as relighting and scene editing.

Keywords: 3D modelling, 3D rendering, artificial intelligence, computer graphics, computer science, image formation, image-based modelling and rendering, inverse problem, visual computing

CCS Concepts: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Rendering—I.2.10 [Artificial Intelligence]: Vision and Scene Understanding—Representations—

1. Introduction

Inverse rendering is a fundamental problem at the intersection of computer vision and computer graphics, where the objective is to recover scene properties such as geometry, material, and illumination from visual data (single-view or multi-view images) [LSR*20, ZLH*22, LZF*24, JLX*23]. Whereas forward rendering synthesizes an image from predefined scene parameters using a deterministic transport model (such as the rendering equation [Kaj86]), inverse rendering must reverse this process: it takes an image (or a sequence of images) as input and estimates intrinsic components that were responsible for producing the observed image [DM23, RH01]. The recovered representation transforms raw pixels into editable 3 D content, enabling applications such as relighting, material editing, virtual object insertion, AR/VR experiences, robotics, film production, and digital content creation [PP03, RH01] (Figure 1).

However, inverse rendering is inherently ill-posed, as many distinct combinations of geometry, materials, and lighting can produce the same image. A bright pixel may correspond to a strongly reflective surface, an intense light source, or a specific viewing geometry; ambiguities arising from inter-reflections, cast shadows, occlusions, and uncontrolled lighting compound this difficulty [BM14, KLK14]. Moreover, the joint estimation of all the intrinsic components is particularly challenging because errors in one component often propagate into the others. This ambiguity has motivated a wide range of approaches, with particularly rapid progress over the past few years driven by three closely connected developments: (i) **Data-driven priors** learned by deep neural networks [EN24, KSN24, ZZLZ25] from large image datasets, enabling efficient decomposition of scenes whose properties are intrinsically underconstrained from pixel evidence alone; (ii) **Differentiable rendering frameworks** such as OpenDR [LB14], Mitsuba 3 [JSRV22], and

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

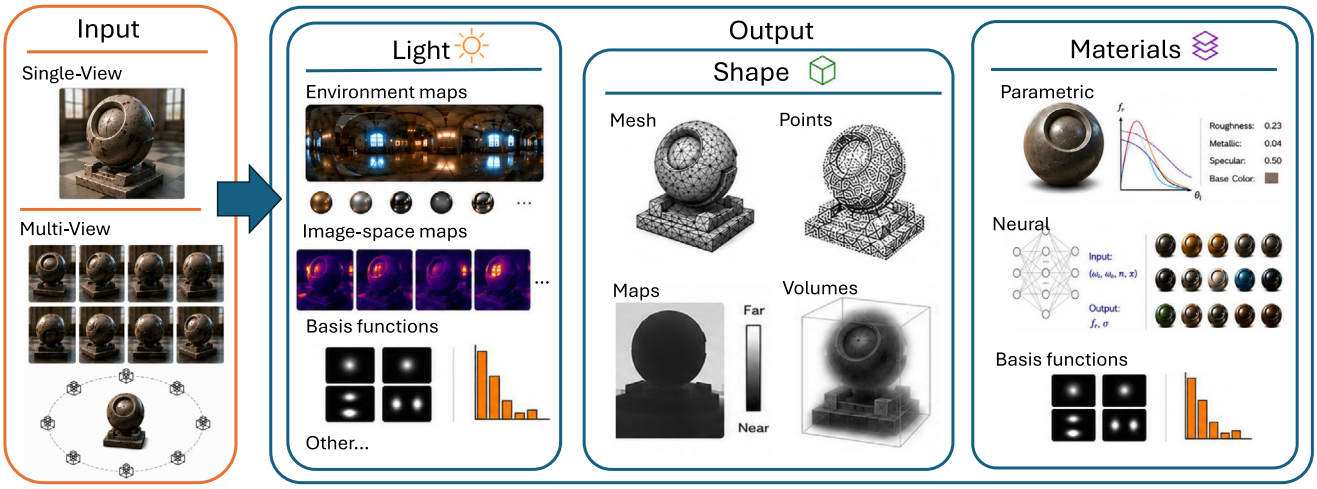


Figure 1: *Our survey provides a comprehensive and up-to-date review of the papers published in the last 6 years in top-rated conferences or journals presenting approaches that attempt to reconstruct the scene description (light, material and shape) from single-view or multi-view image capture.*

nvdiffrast [LHK*20], which provide differentiable formulations of the rendering equation [Kaj86] for gradient-based optimization; (iii) **Neural and Gaussian-based light-field encodings** [MST*21, KKLD23] that provide compact, differentiable scene representations that do not disentangle light, materials, and geometry, but can provide an algorithmic backbone for inverse rendering methods [ZSD*21, RES*22, LZF*24, SWW*25].

1.1. Historical background and scope of the survey

Research on inverse rendering methods has its roots in computer vision techniques like Shape from Shading [ZTCS02] and Photometric Stereo [AG15], where the reflectance behavior of the surfaces is modelled to recover shape, and in the works on intrinsic image reconstruction [BTHR78], aimed at decomposing images into albedo and shading maps. The seminal work of Barrow and Tenenbaum [BTHR78] introduced the basic idea that vision is the inverse of image formation. Image formation involves geometry, reflectance, illumination, and viewpoint; vision consists of recovering these intrinsic properties. It is worth noting that the work was published far before the "forward rendering" formalization currently used in Computer Graphics (the "rendering equation" [Kaj86]). The first explicit use of "inverse rendering" as the estimation of parameters for a rendering algorithm through optimization is attributed to Marschner and Greenberg [Mar98]. While the concept of inverse rendering connects the Computer Vision and Computer Graphics communities and can be considered a founding principle of modern, unified "Visual Computing" science, the fact that inverse rendering methods have been studied separately by the two communities has resulted in non-uniform terminology across the literature. Influential papers on inverse rendering developed in the Computer Vision community, such as the work of Barron and Malik [BM14], describe the same concept, but do not use the term "inverse rendering" to define it. In the first two decades after Marschner and Greenberg's work, the number of papers describing methods to recover the rendering parameters from single or multiple images remained limited. In recent years, however, research efforts have increased significantly thanks to new, successful techniques introduced in the visual computing domain.

The first is the large-scale use of data-driven techniques, e.g., neural networks that can learn to infer scene properties directly from data. The second is the introduction of differentiable rendering frameworks (e.g., OpenDR [LB14], Mitsuba 3[JSRV22]) that further enabled gradient-based optimization over scene parameters, bridging the gap between learning-based models and physically grounded rendering. A third key innovation is the introduction of radiance-field representations like NeRF and Gaussian Splatting, [MST*21, KKLD23], capturing accurate implicit or explicit scene representation, not disentangling shape, material, and illumination, but that can be successfully exploited within inverse rendering frameworks.

A considerable amount of new papers on inverse rendering have been published in the last few years, exploiting these innovations. Current methods span a broad spectrum of input assumptions (e.g., single vs. multi-view), scene types (e.g., indoor vs. outdoor), and representation schemes (e.g., mesh, voxel, neural fields, Gaussian splatting), each with trade-offs in terms of physical realism and computational efficiency. Despite the availability of impressive individual methods, there is a lack of systematic organization and comparison in this diverse set of approaches, making it difficult to assess which techniques are best suited for specific applications or environments. In addition, there are no standard benchmarks, particularly those that jointly evaluate the estimation of geometry, material, and lighting. Synthetic datasets [RRR*21, ZSD*21], offer complete supervision but lack realism, while real-world datasets [LCF*23, UAS*24], offer photorealism but lack complete ground truth, making quantitative comparison difficult. This imbalance is especially evident in generative methods, which prioritize perceptual metrics

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1: *Comparison of related surveys.*

Survey	Primary Focus	G	M	L	NO	DRO	Period
[PP03]	Classical Inverse rendering	√	√	√	✗	√	Pre-2003
[ZSG*18]	RGB-D 3D reconstruction	√	P	✗	✗	✗	2010–2018
[EGH21]	Illumination estimation, relighting	✗	P	√	√	√	2017–2020
[TTM*22]	Neural rendering, view synthesis	√	√	P	✗	√	2020–2021
[GRPCLM22]	Intrinsic image decomposition	P	√	P	√	P	2017–2021
[YLN*24]	Non-rigid 3D reconstruction	√	✗	✗	√	√	2021–2023
[UUE25]	Intrinsic image decomposition	✗	√	√	√	✗	≤ 2024
Our survey	Joint inverse rendering	√	√	√	√	√	2020–2025

Notes: Coverage is marked as √(full), P (partial), or ✗(none).
Abbreviations: G: Geometry; M: Material; L: Lighting; NO: Network Optimization; DRO: Differentiable Rendering Optimization.

(e.g., FID, CLIP similarity) rather than physically grounded evaluations. As a result, both the supervised and self-supervised approaches face challenges in achieving accurate, generalized inverse rendering. To better understand this rapidly growing field and try to address the aforementioned issues and point out the research gaps that need to be filled, we propose a survey of the most recent literature.

We consider papers published at top-tier Computer Vision and Computer Graphics venues from 2020 to 2025 presenting methods that try to solve the fundamental inverse rendering problem recovering disentangled intrinsic scene properties such as geometry, material, and lighting from single-view or multi-view images. In addition to previous work, we exclude the following from our analysis:

1. Methods that focus only on geometry reconstruction without material or lighting estimation (e.g., structure-from-motion, multi-view stereo)[LWZW24, ZSG*18, DHRK24].
2. Image decomposition methods (e.g., intrinsic images or shading–albedo separation) without full physical parameter recovery [UUE25, GRPCLM22].
3. Inverse rendering methods that are specialized for specific object categories (e.g., only human faces, hair, or clothes) [SNA*21, ZCV*25, RBG23, KE19].

To organize our survey, we created a specific taxonomy, described in Section 2, categorizing the methods by scene type, data input, shape representations, material representations, and light representations.

Our work provides useful insights and can both support application developers in selecting the best suited algorithmic choices and researchers in the development of more accurate, efficient, and versatile solutions.

1.2. Related surveys

Several surveys and state-of-the-art reports address topics related to inverse rendering; however, none provide a comprehensive and unified overview of methods that jointly estimate geometry, materials, and illumination, which is the focus of this work. We summarize the principal differences from prior work in Table 1 and discuss the most closely related.

The survey presented by Tewari et al. [TTM*22] focuses on neural rendering approaches that combine classical rendering principles with learnable 3D representations, particularly radiance-field-based methods for novel-view synthesis in both static and dynamic scenes. While these approaches learn 3D-consistent scene representations, they typically encode appearance implicitly and do not explicitly recover a complete set of intrinsic scene properties such as reflectance and illumination. Similarly, the reviews by Ulucan et al. [UUE25] and Garces et al. [GRPCLM22] address the intrinsic image decomposition problem, the 2D task of separating an image into albedo and shading layers. These works do not recover 3D geometry, do not produce relightable scene representations, and treat lighting only implicitly through shading maps. Several further surveys cover specific aspects of the inverse rendering problem. Einabadi et al. [EGH21] focus on illumination estimation and relighting, Zollhöfer et al. [ZSG*18] survey RGB-D scene reconstruction with emphasis on geometry, and Yunus et al. [YLN*24] review non-rigid scene modelling. Each of these works provides valuable insights into its respective domain, but does not consider the full inverse rendering pipeline. Some works address inverse rendering in constrained domains, such as human faces and hair [SNA*21], or indoor environments. Still, these are typically tailored to category-specific priors and do not generalize to arbitrary scenes or objects.

As shown in Table 1, our survey provides a comprehensive and up-to-date overview of single-view and multi-view inverse rendering methods that jointly recover scene geometry and materials using both classical and data-driven approaches.

2. Inverse rendering methods classification

2.1. Problem statement

Before introducing our taxonomy, we first establish a common conceptual and notational framework that will be used throughout the paper.

To define the inverse problem, we first need to introduce the forward one. **Forward rendering** is the process of synthesizing an image from a description of a 3D scene and a camera, and it is the most fundamental technique of computer graphics. The rendering equation gives the basic relationship between scene parameters and the resulting image [Kaj86],

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$L_o(x, \omega_o) = L_e(x, \omega_o) + \int_\Omega f_r(x, \omega_i, \omega_o)\, L_i(x, \omega_i)\, (\omega_i \cdot n)\, d\omega_i \quad (1)$$

Here, x is a surface point, ω_o and ω_i are the outgoing and incident directions, respectively, and n is the surface normal. The term L_e denotes emitted radiance, L_i the incident radiance, and f_r the bidirectional reflectance distribution function (BRDF), which describes how light arriving from direction ω_i is reflected toward direction ω_o. Given the camera geometry and sampling function, the equation can be used to estimate the radiance incident on the virtual image sensor's points and to derive an equation for the pixel intensity. In general, the forward rendering process can be expressed as a function:

$$I = \mathcal{R}(G, M, L; C), \quad (2)$$

where I is the rendered image (grayscale intensity of color channels), G denotes the geometry, M material properties, L the illumination, and C the camera parameters. The operator $\mathcal{R}$ represents the rendering process, typically implemented as a numerical approximation of the rendering equation using techniques such as ray tracing, rasterization, or volumetric integration. In the rendering process, G and C determine the positions where the Equation 1 should be evaluated, M is usually a parametric representation of f_r and L should model the light incident on the surfaces, accounting for interreflections and shadows (global illumination).

Inverse rendering is the recovery of G, M, L from the knowledge of one or more rendered images I with different camera parameters C. It is an ill-posed problem that can be solved only with strong priors, especially when relying on a single image. The solutions proposed to this problem depend on how geometry, materials, and lighting are represented, and on how the rendering equation is discretized. Depending on all these choices, we can have a variety of methods to solve it. Many approaches have been proposed in the literature, and organizing them in a structured taxonomy is therefore useful for analyzing them.

2.2. Inverse rendering methods' taxonomy

Our taxonomy is shown in Table 2. We classify the methods according to the type of input data processed, the type of scene considered, the output scene description parameters, the proposed computing approach, and the evaluation method (including the datasets used).

2.3. Input data

Inverse Rendering methods differ substantially depending on the input images accepted. Methods relying only on **single-view (SV)** acquisitions are particularly challenging due to the inherently ill-posed nature of the problem. A single image may correspond to several combinations of shape, illumination, and materials that could produce it. Due to this ambiguity, SV techniques rely heavily on learned priors. SV methods are suitable for applications like image relighting or material editing.

Most works take as input **multi-view (MV)** images that provide more information about the scene, reducing the problem's inherent ambiguity. Observing a scene from multiple viewpoints allows methods to recover occlusion, shading consistency, and surface appearance viewed at different angles. These cues yield more accurate and robust estimates of geometry, materials, and lighting. These methods process the same input data and can be considered an extension of Structure from Motion/Multi-View stereo pipelines that reconstruct 3D geometry by decoupling shape from materials and lighting. Multi-view stereo methods do not generalize well to complex real-world lighting conditions because they rely on assumptions to make the problem less ill-posed, a problem avoided by Inverse Rendering methods that estimate more scene properties simultaneously. Although most of the proposed methods work with images captured by standard **RGB** cameras, some process enhanced data from depth/LIDAR sensors (**RGB-D**) [CCL*25, YYC*23].

One of the fundamental distinctions in inverse rendering is whether methods target static scenes or dynamic sequences. The **Static** methods, which cover the majority of the current literature, estimate time-invariant geometry, materials, and lighting from one or more images of stationary scenes, in which only the camera viewpoint varies in the captures [ZSD*21, JLX*23]. These methods benefit from simpler problem formulation, more mature material-lighting decomposition techniques, and well-established benchmarks such as [JLX*23] and [JDV*14]. **Dynamic** inverse rendering addresses time-varying scenes captured in video sequences, which require temporal modelling of evolving geometry, deforming surfaces, and/or lighting. Due to this ambiguity, dynamic inverse rendering must move beyond simple novel-view synthesis by disentangling motion from intrinsic properties. This choice defines the Representation Strategy: static scenes typically utilize 3D-only structures (e.g., SDFs or meshes), while dynamic scenes ensure temporal consistency to avoid flickering artifacts, modelling non-rigid deformations, which map moving points to a canonical pose, or 4D Neural Fields (e.g., K-Planes [FKMW*23]) that treat time as a fourth coordinate. However, most dynamic methods focus on novel view synthesis rather than full inverse rendering—they typically bake lighting into appearance representations, limiting relighting and material editing capabilities that define inverse rendering in this survey. Most inverse rendering methods process passive imaging data, while others process active acquisition setups. **Passive methods** operate using only naturally available illumination, without any controlled light sources. These approaches are generally more flexible and easier to deploy, but face inherent ambiguities due to uncontrolled lighting [ZLLS22, ZSH*22]. In contrast, **active methods** employ controlled illumination—such as structured light, depth projection, or time-of-flight sensing—to recover more accurate geometry or reflectance. Although active systems reduce ambiguities and enable more precise material–light decomposition, they require specialized hardware and are less applicable to in-the-wild scenarios [ZLLS22].

2.4. Scene types

Inverse rendering methods differ substantially in terms of the scene being reconstructed, as each environment presents a unique set of challenges, priors, and constraints. **Object-level scenes (Obj)** include a single object or small collections of objects, captured under controlled environments. These scenes have well-defined boundaries, manageable geometric complexity, and controllable lighting conditions, making them ideal for detailed material analysis and high-quality reconstruction. The controlled nature often allows for

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2: *Taxonomy of inverse rendering methods reviewed in this work.*

Input data	View	Single-View (SV) [HWH*25, ZZLZ25, WTC*25, KSN24]
		Multi-View (MV) [JSL*25, TCZD25, WBB*25, CLP*25]
	Sensor	RGB [HYD*25, GWZ*25, LGL*25]
		RGB+D [LSB*22, YYC*23, CCL*25]
	Time	Static [WLLL25, WTL*24, WCD*20]
		Dynamic [CCL*25, LGL*25]
	Capture setting	Passive [DWZ*24, LYX*24, EN24]
		Active [BJK*20, ZLLS22, CCB24]
Scene type	Indoor (IND) [LSR*20, KSN24, LHL*25, WLLL25]	
	Outdoor (OUT) [WSG*23, RES*22]	
	Object (OBJ) [WTC*25, CXG*25, DLDN25, JLX*23, LZF*24]	
	Generic (GEN) [EN24, LGL*25, CLZ25, SGX*25, JSL*25]	
Scene description	Shape	Image maps (IM) [WPFK21, ZGW*23, YLH*23]
		Point Density (PD) [SWW*25, DLDN25, SGX*25]
		Volume Density (VD) [YYC*23, JLX*23, JTL*24]
		Signed Distance Functions (SDF) [DLLY22, WHL*23, YN23]
		Polygon Meshes (M) [BXS*20b, LZBD21, DWW*24]
	Material	Parametric BRDF (P) [HWH*25, WTC*25, WBB*25]
		Basis (B) [ZSD*21, CCB24, DLDN25]
		Neural (N) [SDZ*21, YYC*23, FTTS24]
	Light	Environment Map (EM) [BXS*20a, JTL*24, WRG*25]
		Spherical Harmonics (SH) [BJK*20, LZF*24, CCL*25]
		Spherical Gaussians (SG) [YLHG*22, ZLW*21, YGY*24]
		Illumination Maps (I) [ZGW*23, LHL*25, CLP*23]
		Global Illumination (GI) [WPFK21, LWZW24, ZHY*23]
		Light Fields (LF) [LOU*24, LYX*24, BCC25]
Computing approach	Network optimization (NO)	Convolutional Neural Networks (CNN) [ZLH*22, WTC*25]
		Vision Transformers (ViT) [ZLM*22, CLP*23, CLP*25]
		Diffusion Models (DM) [EN24, KSN24, HWH*25, CPY*24]
	Differentiable rendering optimization (DRO)	Splatting (S) [HYD*25, GWZ*25, SGX*25, TCZD25]
		Volumetric (V) [ERB*24, FTTS24, SCL*23, WBB*25]
		Path tracing (PT) [YLHG*22, WZY*23, LYX*24, WRG*25]
Evaluation	Weighted Human Disagreement Rate (WHDR) [LSR*20, WPFK21, ZLH*22]	
	Image Differences (IMD) [BBJ*21, BJB*21, WTL*24, LLJ*25]	
	Environment Maps Differences (EMD) [DLLY22, MHS*22, SWW*25, LHG*25]	
	Rendering Error (RE) [RES*22, HHM22, WSG*23, YGY*24]	
	Novel View Synthesis Error (NVSE) [BXS*20b, BJB*21, YZL*22, ZHY*23, YGY*24]	
Applications	Scene Editing (SE) [LSR*20, LGL*25, MAAD*23]	
	Material Editing (ME) [KSN24, ZLM*22, WTC*25, JLX*23]	
	Light Editing (LE) [KSN24, SGX*25, LCL*23, ZLW*21]	

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ground-truth validation through independent measurements of geometry, materials, and lighting. **Indoor scenes (Ind)**, which include interior environments such as rooms and offices, introduce greater complexity than object-level scenes. They are characterized by structured, bounded geometry, diverse materials (such as painted surfaces, wood, metal, glass, and fabrics), and mixed natural and artificial lighting within enclosed spaces. This complex, bounded lighting leads to complex indirect lighting, inter-reflections, and sharp shadows. **Outdoor scenes (Out)** represent the most challenging cases, and include, for example, urban environments, natural landscapes, and large-scale architectural exteriors. These scenes exhibit significant geometric diversity, atmospheric effects (scattering, fog, weather), and uncontrolled natural lighting dominated by sun and sky illumination with strong temporal variations. Material diversity includes building materials, vegetation, natural elements, and weathered surfaces that change over time. **Generic scenes (Gen)** is the category where we classify the methods trying to solve the inverse rendering task, independent of the environment type.

2.5. Scene description

Given the input data representing the selected scene type, the inverse rendering methods should typically reconstruct shape (S), materials (M), and lighting (L) information. However, there are some exceptions: some of the methods focus on reconstructing objects and do not explicitly recover lighting information [CCB24]; others do not recover geometry, as they assume known geometries given as input [YYC*23]. In this case, as in depth-enhanced acquisitions, the geometry data acts as a structural prior and simplifies the estimation of materials and illumination, leading to more accurate estimates of scene properties. In the following, we describe the different ways in which reconstruction systems represent shape, materials, and light as outputs.

2.5.1. *Shape representation*

The shape of objects in the scene can be represented in different ways, depending on the input data and algorithmic choices. It is also important to note that the shape representation used in the algorithms' optimization may not match the desired final representation or the representation used to evaluate the approach's effectiveness. In fact, several methods (e.g., those based on NeRF and Gaussian Splatting) leverage 3D density encodings to represent shapes and optimize volume rendering to match training views. However, these density based methods are not suitable for many applications (e.g., modelling) and cannot be directly evaluated. For this reason, some papers also provide additional 'output' representations derived from the optimized fields, such as meshes or image-space maps.

In this shape classification, we aim to characterize each method based on the primary representation estimated and also the additional outputs, if explicitly evaluated in the framework.

In single-view methods, it is typical to represent shape information with **image-space maps (IM)**: depth maps, which store the per-pixel distance from the camera to the first visible scene surface, and normal maps, which represent, for each pixel, the normal vector of the surface hit by the image ray. Not all methods that provide this kind of reconstruction produce depth maps, as these are not fundamental for rendering, but this limits the amount of information encoded (image-space depth could, in principle, be derived from normal integration, but the procedure is not reliable due to noise amplification). This representation is obviously incomplete and unsuitable for object reconstruction or novel view synthesis applications. However, it supports applications related to relighting, material editing, and scene editing.

Point clouds are typically the output of SfM pipelines, but they cannot represent the continuous boundaries of objects that require further processing to estimate. However, techniques like Gaussian Splatting, which associate density and light-field information with the surroundings of points, make it possible to render shapes encoded as point clouds in a realistic way from different viewpoints, and the differential rendering of GS has been exploited in inverse rendering frameworks. We classify the corresponding shape representations as **Point Density (PD)**. Clearly, this representation loses accuracy in representing objects' surfaces.

Volumetric density functions (VD) stored in grids or continuously represented using MLPs are the choice of many recent approaches (e.g., NeRF-based ones) for storing shape information. These approaches are based on differentiable volume rendering optimization. Although this choice enables volumetric shading and transparency, it loses the ability to reconstruct object shapes and associated surface properties without further processing.

For this reason, some of the methods add (or replace) this volumetric encoding with implicit surface representation as the **Signed Distance Functions (SDF)** that represent the distance of any 3D point from the nearest surface.

3D meshes (M) explicitly representing the scene geometry as collections of polygons are the standard representations used in 3D modelling and provide correct boundary representations, supporting rendering from arbitrary views and geometric editing. Material parameters and normals can be associated with polygon vertices or faces.

2.5.2. *Material representation*

Material properties determine how light interacts with the geometry of the scene, controlling the appearance of objects under different illumination conditions. Most methods estimate materials as **parametric models (P)**. The choice of shading model directly influences inverse rendering optimization. The Lambertian model assumes perfectly diffuse reflectance [BTHR78] and is common in single-view inverse rendering [LSR*20, ZLH*22] because its view-independence removes specular ambiguity from the estimation problem. However, the Lambertian assumption causes specular highlights to be absorbed into the diffuse albedo estimate [BM14], producing albedo maps that conflate illumination variation with surface reflectance. The Phong and Blinn-Phong models introduce a specular component through a power-of-cosine term but violate energy conservation [CT82, WMLT07], which is a critical constraint for physically grounded inverse rendering, and produce poorly conditioned gradients due to their non-physical parameterization. The most recent work adopts microfacet-based models in the

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cook-Torrance framework [CT82], with the GGX normal distribution [WMLT07] as the dominant choice for the specular lobe. These use simplified Disney BRDF models [BS12], including albedo or base color for the diffuse component, and additional parameters such as roughness, metallic, and glossiness. The main limitation of these models in inverse rendering is the diffuse-specular ambiguity. Under single unknown illumination conditions, a bright specular highlight from a rough surface can be difficult to distinguish from a strong diffuse response caused by intense lighting, leading to incorrect estimates of albedo and roughness. The parameters of the models can be stored in image-space maps, on texture maps for 3D meshes, or encoded directly in the geometric representations used to store scene content, such as point clouds [XXP*22], neural functions [LWL*23], or Gaussian [GGL*24].

Another popular choice is the use of **Basis-based representations (B)** describing the reflectance functions as linear combinations of pre-defined parametric functions [CCB25] (e.g., angular functions). The idea has a long history in graphics, dating back to wavelet and spherical-harmonic representations of measured BRDFs [LRR04, Mat03] and to the analytic-basis decomposition used by NeRD [BBJ*21] to model spatially varying reflectance under unknown illumination. These representations are more flexible than fixed parametric models and easier to fit, but they may lack direct physical interpretability. A third family of methods uses **Neural representations (N)**, in which a compact neural network, typically a multilayer perceptron (MLP), encodes the BRDF. The network takes viewing and lighting directions, and sometimes surface position or latent features, as input and predicts either the BRDF value or reflected radiance. The main advantage of this representation is its ability to model complex reflectance behavior that is difficult to express with fixed analytic functions, including measured materials, layered or anisotropic appearance, and view-dependent effects on glossy or near-specular surfaces. NeILF [YZL*22] and NeILF++ [ZYL*23] follow this direction by jointly representing illumination and material properties as neural fields. Neural BRDFs are highly flexible, but the lack of explicit parameters makes editing and physical interpretation more difficult, and the recovered material is often tied to the specific illumination conditions encountered during training.

2.5.3. *Lighting representation*

Illumination is handled in many different ways in the surveyed literature, and the output reconstruction of the methods is heavily dependent on the approximations done in the scene modelling. In particular, object-level reconstruction methods often model environment light only, while others also represent global illumination effects, storing specific parameters to describe light coming from close surfaces.

The most popular solution for modelling far-away illumination (coming from the sky or distant objects) is the use of **Environment maps (EM)**, represented image textures that encode directional illumination, typically with High Dynamic Range. An alternative option to represent directional lights is the use of coefficients of basis functions. **Spherical Harmonics (SH)** [KSS02] form a linear basis on the sphere and are often used to encode low-frequency environment with few coefficients. **Spherical Gaussians (SG)** [WRG*09] form a non-linear basis on the sphere, and are used to succinctly represent all-frequency lighting using gaussian of varying standard deviation.

A different choice is to encode lighting information in pixel-space **Illumination Maps (I)** that represent the per-pixel incident illumination that arrives at a point from all sources. They are used to represent incoming illumination directly without explicitly reconstructing light sources. They can represent direct illumination or other indirect illumination effects, such as Shadow Maps [Wil78] representing object occlusions from a single light, and are used to model shadowing effects independently of the recovered geometry or ambient occlusion.

Some methods estimate the **Indirect/global Illumination(GI)** coming from multi-bounce lighting effects, where light reflects off multiple surfaces before reaching a point. Recovering this separately from direct illumination allows for more accurate estimation, since indirect effects are modelling, and estimated, directly. Indirect illumination can be encoded in different ways and can model different components such as visibility/shadow maps or inter-reflections. Due to the many different ways used to account for these effects, we label all of them as GI. **Neural light fields (LF)** are another option for encoding different kinds of illumination effect, and we explicitly include this category in the taxonomy.

2.6. Computing approaches

To understand the research trends in inverse rendering, it is important to point out which computational approaches are primarily used to solve the problem. For this reason, we introduce a specific field in our taxonomy with two main categories, network optimization-based methods and differentiable rendering optimization methods (Figure 2).

2.6.1. *Network optimization (NO)*

This category includes learning-based methods that train a feed-forward network on a large dataset to directly estimate the model parameters from image data. At test time, a single forward pass through the network predicts the intrinsic scene properties directly from one or more input images (Figure 2, left). Within the Network optimization branch, we further classify methods according to the dominant neural architecture used to estimate scene properties.

Convolutional Neural Networks (CNN) based methods use architectures such as U-Nets or ResNets to infer per-pixel scene properties (geometry, materials, illumination) from a single or multi-view input image data. Early scene-level inverse rendering systems, such as Li et al. [LSR*20] and [ZLH*22] established a common formulation in which a shared encoder-decoder network produces per-pixel albedo maps, surface normals, roughness, and a parametric lighting representation. These methods are good at capturing local texture and appearance patterns, but can struggle with global consistency and physical constraints. Papers are classified in this category if they use predominantly convolutional layers as their primary computational building blocks. **Vision Transformer (ViT) based** methods use attention mechanisms and transformer architectures to capture long-range dependencies within an image. IRISformer

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

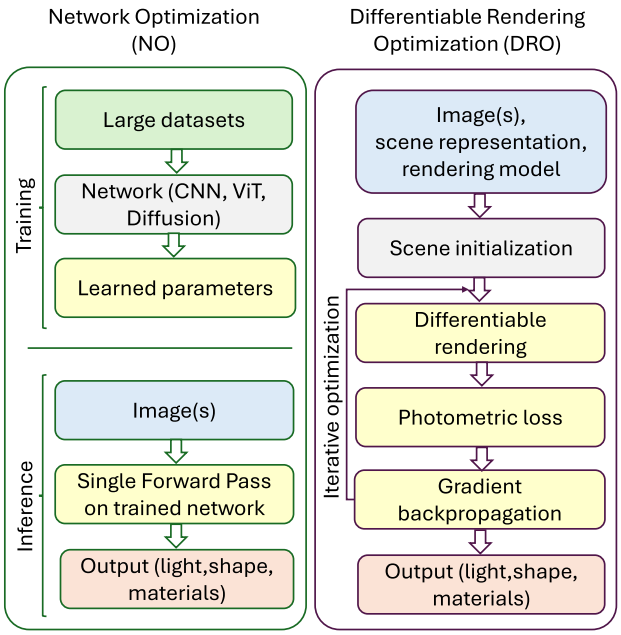


Figure 2: *Comparison of the two dominant computational paradigms for inverse rendering. Network Optimization (NO) learns a mapping from images to scene parameters using a neural network trained on large datasets, enabling fast inference through a single forward pass. Differentiable Rendering Optimization (DRO) estimates scene parameters by iteratively minimizing the discrepancy between rendered and observed images through a differentiable renderer. While NO prioritizes inference efficiency and generalization, DRO emphasizes physical consistency and scene-specific optimization*

[ZLM*22] adapts a dense vision transformer to single-view indoor inverse rendering, and MAIR / MAIR++ [CLP*23, CLP*25] extend the idea to multi-view attention. These architectures help capture global illumination effects, making them suitable for estimating complex materials and illumination. Papers using vision transformers, attention mechanisms, or transformer-based architectures as their core processing units belong to this category. **Diffusion Model (DM) based** methods integrate generative diffusion models as strong priors, enabling high-quality estimation of geometry, material, and illumination. Recent methods include Intrinsic Image Diffusion [KSN24] for indoor material estimation, DiffusionRenderer [LGL*25] for joint forward and inverse rendering with video diffusion, and DNF-Intrinsic [ZZLZ25] for deterministic noise-free indoor decomposition. Diffusion priors are particularly valuable when the input is ambiguous, such as a single low-dynamic-range image of an indoor scene, because the generative prior can fill in plausible material and lighting hypotheses that are consistent with the observation. In this group, we classify the papers that utilize diffusion processes, denoising networks, or score-based generative models for inverse rendering.

2.6.2. *Differentiable rendering optimization (DRO)*

The methods based on differentiable rendering optimization estimate the parameters of a chosen scene representation by minimizing the difference between rendered and observed images through a differentiable renderer (Figure 2, right). Differentiable rendering optimization is slower than NO but can, in principle, adapt to any scene as long as the differentiable renderer and the chosen representation are expressive enough.

DRO-based methods are characterized by the use of various geometric representations and rendering solutions. We have already discussed geometry representations such as polygon meshes, signed distance fields, continuous volume densities, tensor fields, or Gaussian point primitives in Section 2.5.1 and do not duplicate that discussion here. The choice of scene representation is largely independent of the computing approach because the same SDF can be optimized by a volume ray marcher or by sphere tracing, and the same Gaussian primitives can be rendered by splatting or by ray tracing. In our taxonomy, therefore, we subdivide the DRO-based methods according to the rendering algorithm that maps the chosen scene representation back to image pixels. Crucially, it also propagates image space gradients back to the scene parameters during optimization. We classify the rendering approaches found into three categories.

Volume rendering–based methods (V) estimate scene properties by integrating radiance along camera rays through volumetric or implicit scene representations such as NeRFs or signed distance fields. Recent work has produced unbiased gradient formulations for both grid-based SDFs [VSJ22] and neural SDFs [PBGL*22], which together extend the volume-ray-marching machinery to support physically meaningful surface optimization. Volume ray marching naturally handles soft visibility, uncertain geometry, and topology changes during optimization, which is one reason it has become the dominant approach for joint geometry and material recovery from multi-view images. However, they typically require hundreds of samples per ray and often involve expensive optimization.

Splatting-based methods (S) are conceptually similar to rasterization but operate on point or Gaussian primitives rather than triangles. Each primitive is projected onto the image plane and accumulated using a differentiable alpha-blending rule. Recent inverse rendering approaches have extended Gaussian Splatting to physically based inverse rendering by combining it with shading models such as GaussianShader [JTL*24] and Jiang et al. [JSL*25]. Similar to rasterization-based approaches, splatting methods handle primary visibility natively. Still, accurate modelling of secondary effects, such as indirect illumination and inter-reflection, typically requires additional light-transport computations or auxiliary rendering modules.

Path tracing-based methods (PT) model full light transport, including multi-bounce indirect illumination, shadows, and other global illumination effects by sampling light paths stochastically through the scene. These methods utilize differentiable Monte Carlo formulations to obtain unbiased gradients despite discontinuities introduced by visibility [LADL18, ZYZ21, LHJ19]. These methods are computationally expensive and often suffer from noisy gradients during optimization. However, these methods provide the highest physical accuracy, and they are increasingly common in scene-level inverse rendering, where it is not possible to ignore lighting effects.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.7. Evaluation

Evaluation of inverse rendering methods remains challenging due to their diverse outputs and lighting model and to the fact that ground truth data on materials are obviously not available for real images and may not be consistent with the method outcomes for the synthetic ones. For this reason, several approaches have been proposed to evaluate, typically in indirect ways, the outputs or the methods, each of which presents clear limitations.

A popular method to evaluate albedo maps in single-view applications coming from the intrinsic image decomposition domain is the use of a metric called **Weighted Human Disagreement Rate (WHDR)**, based on human judgments about relative reflectance between pixel pairs that are sparsely annotated on benchmark images. Wu et al. [WCS*23] recently pointed out the limitations of this approach, proposing additional metrics.

If the methods provide an estimation of depth, normals or shading parameters and ground truth values are available for the input images, the evaluation can be based on **image-space maps differences (IMD)**. This approach requires well-designed calibrated or synthetic data and still presents some issues related to the scale ambiguity of reflectance parameters. Scalar maps can be evaluated with mean average differences, and specific metrics can be applied to evaluate other maps, for example mean angular error (AE) for normal maps or Absolute Mean Relative Error (AMRE) for monocular depths. Recent papers also use the Fréchet Inception Distance (FID) to evaluate the realism and quality of the generated texture maps. FID compares feature representations extracted from a pre-trained Inception-v3 network and evaluates similarity in terms of content and style. Image-space differences are used both in the evaluation of SV and MV methods.

Error in geometry reconstructed with 3D meshes can be estimated with **mesh error (ME)** metrics including accuracy, precision, recall, and F-score. Methods that reconstruct illumination as directional lights or environment maps can evaluate the accuracy of the estimation using environment maps differences **EMD** if ground truth values for the parameters are provided. However, this is possible only with specifically designed datasets that use accurate calibration or synthetic data.

Another way to evaluate the methods is to compute the **rendering errors (RE)** with respect to input images or ground truth relighted images. The metrics for image comparison can be different, from simple MSE to Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [WBSS04], and Learned Perceptual Image Patch Similarity (LPIPS) [ZIE*18].

Most multi-view inverse rendering approaches are evaluated indirectly by estimating the quality of **novel-view synthesis errors (NVSE)**. The evaluation based on rendered images is clearly the easiest, as it is independent of the method's output representations and can be done on a dataset with real or synthetic images, where only the camera pose and intrinsics need to be known. This kind of evaluation can exploit the large number of datasets used for benchmarking light field encoding techniques (e.g., NeRF or Gaussian Splatting), but cannot assess how well the methods actually separate the information on shape, materials and lighting.

The evaluation of inverse rendering methods relies on a diverse collection of synthetic and real-world datasets, each designed to evaluate a specific aspect of scene description. NeRF-based methods, learning an implicit volumetric radiance field, require a dense multi-view image collection with precise camera poses, such as synthetic datasets [SDZ*21, MST*21], and real-world datasets [JDV*14, YLL*20] recover volumetric density and view-dependent color. These methods require laboratory setup and synthetic scenes to properly train neural radiance fields, but are challenging to train on sparse real-world data. Methods like [RES*22] extend this to unbounded scenes using [SDZ*21] and a custom outdoor scene dataset. Gaussian Splatting methods require similar multi-view datasets (e.g., [VHM*22, SDZ*21]), but often require fewer views due to their explicit, Gaussian/point nature. These methods demonstrate better efficiency with clear geometric structures such as [KPZK17] and synthetic object-centric scenes [JDV*14]. Data-driven approaches that utilize learned priors require large-scale datasets with complete, pixel-wise ground-truth decomposition. Single-view CNN/ViT approaches are trained on large-scale datasets such as [ZLH*22, LYS*21], and [SHKF12], making them more practical for real-world scenes. Multi-view approaches often train on diverse image collections and evaluate on both synthetic datasets [LSM*18, JLX*23] and real-world images ([BBS14], custom captures).

Finally, Differentiable Rendering Optimization (DRO) approaches are mostly validated with smaller, specialized object-scanning benchmarks such as as [JLX*23, KKO*23, KZY*23, ZYL*23]. These methods are less dependent on training data, but require good initialization and multiple views with known camera poses, making them suitable for controlled capture scenarios rather than in-the-wild applications

2.8. Applications

The main practical applications of inverse rendering algorithms demonstrated in research works or commercial tools are related to the scene manipulation allowed by the reconstructed 3D representations with disentangled geometry, materials, and lighting.

2.8.1. *Light editing*

Light editing applications can change direction, intensity, and light source while maintaining physically consistent shadows and reflections across a scene. For example, in film and VFX production [Aut18], lighting conditions can be changed after the shot is captured, such as converting a daytime scene into a sunset mood or adjusting the key light direction to match CGI assets without reshooting.

[KSN24] (see Figure 3a) demonstrates relighting using environment maps and adjusting lighting intensity for captured objects, while SVG-IR [SGX*25] (see Figure 3b) demonstrates local illumination editing, where only the lamp light is modified, affecting the surrounding environment. The quality of light editing depends on how accurately the light and material are decomposed across a scene. Single-view methods often struggle to relight a scene due to illumination-material ambiguity. Similarly, dynamic scene inverse

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(a) Envmap-based relighting [KSN24]

(b) Local light editing [SGX*25]

Figure 3: *Light editing examples: (a) Captured scenes are illuminated using novel HDR environment maps. (b) Local change of the lamp illumination. Images from: Sun et al., 2025, https://learner-shx.github.io/project_pages/SVG-IR/index, CC BY-SA 4.0; Kocsis et al., 2024, https://openreview.net/forum?id=1VnQ7AYgNW, CC BY-SA 4.0.*

rendering methods rarely support relighting, as most methods cannot decompose materials and lighting across time-varying scenes. Other examples include ENVIDR [LCL*23], which targets explicitly manipulating controllable environment lighting via learned incident light fields. Recent neural rendering approaches, including NeILF [YZL*22] and PhySG [ZLW*21], demonstrate physically-based relighting through spherical Gaussian lighting parameterization and neural reflectance modelling.

2.8.2. *Scene editing*

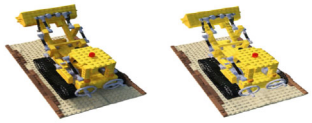
A classical scene-editing application is virtual object insertion. For example, in e-commerce and AR applications [Ado24], it is common to edit the scene, adding virtual objects or furniture. Accurate lighting estimation ensures realistic shading and reflections on the inserted object. At the same time, the reconstructed scene geometry enables virtual objects to cast shadows onto real surfaces, receive reflections from real materials, and participate in global illumination with proper inter-reflections and occlusions. Object insertion results are demonstrated in research papers proposing specific AR tools. Figure 4a shows an example of virtual object insertion in a captured scene from [LGL*25]. The table is inserted into the image while maintaining consistent illumination in the scene. Figure 4b shows an example of object removal while the output image looks realistic (from [MAAD*23]). Figure 4c demonstrates the removal of shadows from an image to produce a clean and coherent result (from [YGY*24]). However, several challenges remain, including the diversity of products, difficulties with transparent or reflective materials, and the need for a visually convincing real-time rendering.

2.8.3. *Material editing*

Material editing changes the reflectance parameters at the image or object level before rendering. This capability allows creative professionals to change the surface appearance of a scene, such as its

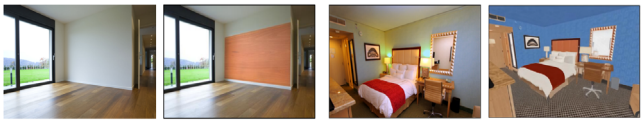
(a) Object insertion [LGL*25]

(b) Object removal [MAAD*23]



(c) De-shadowing [YGY*24]

Figure 4: *Scene editing examples. (a) The table is inserted into the input scene and correctly illuminated. Object removal (b) and de-shadowing (c) applications produce clean, coherent results modifying the original scene. Images from: Liang et al., 2025, https://arxiv.org/abs/2501.18590, CC BY 4.0; Mirzaei et al., 2023, https://openreview.net/forum?id=iTgS4irmbf, CC BY-SA 4.0; Yang et al., 2024, https://openreview.net/forum?id=iTgS4irmb CC BY-SA 4.0.*



(a) Texture edit [ZLM*22] (b) Albedo edit [KSN24]

(c) Transparency edit [WTC*25] (d) Metal/specular edit [JLX*23]

Figure 5: *Examples of material properties' editing. (a) Wall texture modification and (b) albedo (base color) editing demonstrate changes to surface appearance while preserving scene geometry. Transparency editing (c) and metalness/specularity adjustments alter light transmission and reflectance behavior. Images from: Zhu et al., 2022, https://arxiv.org/abs/2206.08423 CC BY-SA 4.0; Zheng et al., 2025 https://openreview.net/forum?id=m6IisFwo2t CC BY 4.0; Jin et al. 2023 https://openreview.net/forum?id=4ULck9r3mY CC BY-SA 4.0; Wang et al., 2026, https://link.springer.com/article/10.1007/s11263-026-02833-z CC BY-SA 4.0.*

color, roughness, or metallic properties. For example, an interior designer [RHS24] can change wall colors, floor textures, or furniture finishes in a virtual room to explore different design options before making a final choice. Similarly, in e-commerce [Ado24], customers can preview products such as furniture or appliances in various colors or surface finishes. Inverse rendering methods that estimate explicit material parameters, especially spatially varying BRDFs (SVBRDFs), make these edits possible by providing parameter maps that can be modified directly at the image level or by allowing to replace entire materials at the object level. Several recent methods support material editing in different ways. Techniques that recover detailed spatially varying BRDF maps (e.g., [KSN24, ZLM*22]) allow users to adjust albedo, normals, or roughness in texture space (see Figure 5a, Figure 5b). Materialist [WTC*25]

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3: *Classification of single-view inverse rendering methods reviewed in this survey.*

Reference	Year	Input				Output		Method		Scene description			Evaluation	Apps.
		View	Dyn.	Dep.	Act.	S/M/L	Scene	Approach	Sub-Appr.	Shape	Mat.	Light		
[LSR*20]	2020	SV	-	-	-	SML	Ind	NO	CNN	IM	P	SG	WHDR, IMD	SE, ME
[WCD*20]	2020	SV	-	✓	-	SML	Ind	NO	CNN	IM	P	EM	EMD, RE	-
[SC20]	2020	SV	-	-	-	SML	Obj	NO	CNN	IM	P, N	EM, SH	IMD, RE	-
[BJK*20]	2020	SV	-	-	✓	SML	Obj	NO	CNN	IM	P	SH	IMD	-
[WPFK21]	2021	SV	-	-	-	SML	Ind	NO	CNN	IM	P	GI	WHDR, IMD, RE	-
[ZLM*22]	2022	SV	-	-	-	SML	Ind	NO	ViT	IM	P	I	IMD, RE	SE, ME
[ZLH*22]	2022	SV	-	-	-	SML	Ind	NO	CNN	IM	P	EM, GI	WHDR, IMD	SE, ME
[YLHG*22]	2022	SV	-	-	-	ML	Ind	DRO	PT	M	P	SG	IMD, NVSE	ME, LE, SE
[LSB*22]	2022	SV	-	✓	-	ML	Ind	NO	CNN	IM	P	SG	EMD, NVSE	ME, LE, SE
[ZGW*23]	2023	SV	-	-	✓	SML	Obj	NO	CNN	IM	P	I	IMD, RE	-
[YLH*23]	2023	SV	-	-	-	SML	Ind	DRO	PT	M	P	EM	IMD, NVSE	ME, LE, SE
[EN24]	2024	SV	-	-	-	ML	Gen	NO	DM	-	P	EM	-	-
[KSN24]	2024	SV	-	-	-	SML	Ind	NO	DM	IM	P	SG	WHDR, IMD, RE	ME, LE
[WTC*25]	2025	SV	-	-	-	SML	Obj	NO	CNN	IM	P	EM	IMD, EMD	SE, ME, LE
[LGL*25]	2025	SV	✓	-	-	SML	Gen	NO	DM	IM	P	EM	IMD, RE	SE, ME, LE
[ZZLZ25]	2025	SV	-	-	-	SM	Ind	NO	DM	IM	P	-	IMD	LE, ME, SE
[HWH*25]	2025	SV	-	-	-	SML	Obj	NO	DM	IM	P	-	IMD, RE	-
[CXG*25]	2025	SV	-	-	-	SML	Obj	NO	DM	IM	P	-	IMD, RE	-

Notes: Apps: applications demonstrated (Please see Table 2 for symbol legend).
A dash (–) indicates that the feature is not applicable.
Abbreviations: Dyn. = dynamic data, Dep. = depth enhanced input, Act = active capture setup.

(Figure 5c) modifies material transparency by decomposing the image into physically-based material properties, allowing for the direct manipulation of refractive parameters. Similarly, TensoIR [JLX*23] (see Figure 5d) enables changes of per-surface materials through its factorized representation. Other approaches use learned priors to support semantic material replacement, such as converting wood to metal while preserving the object's geometry. The primary challenge in material editing is to maintain spatial consistency; edits to one surface should not affect neighboring areas, which is easier when materials are stored in explicit textures but harder with implicit neural representations, where surface boundaries do not cleanly separate material. Additionally, the material model used further limits the user control: methods that use simplified parametric BRDFs can adjust only a limited set of parameters and cannot address complex effects such as anisotropy, multilayer materials, or subsurface scattering.

3. Reviewed methods

In this section, we summarize the characteristics of the methods proposed in the selected papers, grouped by the technical sub-problem addressed, to better highlight the core architectural design choices and trade-offs that define the current state of the art.

3.1. Single-view inverse rendering

Table 3 summarizes the main features of the single-view inverse rendering approaches we found in the surveyed papers.

3.1.1. *Indoor scenes*

In an indoor setting, the primary challenges come from spatially varying illumination, cast shadows, inter-reflections, and the wide range of materials present in these environments.

Most approaches follow a template first introduced by Li et al. [LSR*20]: a cascade of CNN decoders predicts per-pixel albedo, normals, roughness, and a parametric lighting field. Wei et al. [WCD*20] add rendering-aware constraints for scene-consistent illumination from RGB-D observations, and Wang et al. [WPFK21] replace the global environment map with a Volumetric Spherical Gaussian representation that captures spatially varying illumination on a voxel grid. More recent works improved global scene reasoning through transformer architectures. IRISformer [ZLM*22] replaces the U-Net backbone with a dense vision transformer that captures the long-range light transport indoor shading requires, while Zhu et al. [ZLH*22] augment a CNN with screen-space ray tracing and out-of-view lighting prediction. The Physically-Based Editing system of Li et al. [LSB*22] extends this work with explicit 3D lamp predictions for downstream light insertion and removal.

Another direction exploits differentiable rendering at test time. PhotoScene [YLHG*22] and PSDR-Room [YLH*23] fit a procedural material graph and parametric light to a coarse retrieved CAD model through a physics-based differentiable renderer, taking geometry as input rather than recovering it. These methods sit under DRO/P in our taxonomy and produce parameters directly usable by existing 3D content tools. The most recent shift comes from generative diffusion priors. Intrinsic Image Diffusion [KSN24] fine-tunes Stable Diffusion to sample plausible albedo and roughness maps from a single indoor image, and DNF-Intrinsic [ZZLZ25] replaces stochastic denoising with a deterministic flow formulation for more stable decompositions across runs.

Overall, these scene-level approaches provide functional pipelines for single-view indoor editing. The remaining limitation, shown in Table 5, is that even diffusion methods reach only mid-level albedo quality on the InteriorVerse dataset [ZLH*22].

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Recent progress has been driven primarily by improvements in illumination modelling and learned priors rather than fundamental changes in geometric reconstruction.

3.1.2. *Object-level methods*

When considering input images that contain a single object or a small collection of objects, the inverse rendering problem shifts in priority. In this case, spatially varying illumination matters less, but specular reflectance, complex material behavior, and ambiguity between lighting and surface appearance matter more.

The authors of [SC20] predict diffuse and specular maps with an environment map and spherical harmonics lighting decomposition from one image. Boss et al. [BJK*20] use a two-shot flash and ambient capture that jointly refines coarse shape and reflectance with residual learning between the two illumination conditions. Zhang et al. [ZGW*23] replace fixed flash lighting with a learned planar pattern optimized jointly with the network. Recent works increasingly adopt generative priors. Materialist [WTC*25] layers progressive differentiable rendering on top of learned priors to recover a full PBR material and environment map suitable for physically based editing. Neural LightRig [HWH*25] synthesizes consistent multi-light images from one input, then estimates normals and PBR parameters from the synthesized stack, effectively turning a single image into a synthetic photometric stereo problem. UniRenderer [CXG*25] formulates forward and inverse rendering as two halves of a dual-stream diffusion model with cycle consistency between them.

All these object-level methods follow the same direction: how much of the diffuse-specular ambiguity a learned prior can resolve from a single view. Early work relied on controlled hardware-acquisition tricks to gather missing appearance data; recent work replaces hardware constraints with generative priors strong enough to hallucinate the missing observations.

3.1.3. *Generic in-the-wild methods*

Recent research attempts to develop generic inverse rendering methods independent of the environments. Diffusion-based methods such as [EN24] cast inverse rendering as a stochastic problem, training a diffusion model to sample reflectance and illumination for unknown scenes under complex natural lighting. DiffusionRenderer [LGL*25] extends the idea to video with a forward-and-inverse diffusion model that recovers temporally coherent material and lighting across a clip. These methodologies indicate a growing trend toward large generative priors trained on diverse image collections, enabling inverse rendering beyond the controlled environments typically represented in existing benchmarks.

3.2. Multi-view inverse rendering

Multiple viewpoints provide the method with observations of the same surface point under different angular conditions, thereby reducing much of the ambiguity that defeats single-view methods. Table 4 summarizes the main features of the multi-view inverse rendering approaches we found in the surveyed papers.

3.2.1. *Indoor scenes*

Indoor multi-view inverse rendering is challenging due to complex indirect illumination, sharp shadows, and diverse material properties, not captured by single BRDF models. I^2-SDF [ZHY*23] combines neural signed distance fields with differentiable Monte Carlo rendering to enable relighting and editing of indoor scenes. MILO [YYC*23] handles indoor scenes that contain light-emitting objects, with geometry given as input. Li et al. [LWC*23] optimize material and lighting on a recovered mesh in a three-stage process guided by semantic segmentation priors to regularize the decomposition of scene parameters. Modelling physically accurate light transport in an indoor environment introduces significant algorithmic complexity, limiting the optimization process's efficiency. To address this, Factorized Inverse Path Tracing [WZY*23] introduces a mathematical factorization of the path-tracing integral that renders global illumination calculations tractable over large volumes. While most architectures assume raw inputs have a high dynamic range, IRIS [LHL*25] addresses a common real-world limitation by jointly optimizing the non-linear camera response function and material properties from standard low-dynamic-range captures. Addressing localized illumination artifacts, SIR [WLLL25] implements a decomposable shadow modelling technique to prevent ambient occlusion features from bleeding into the estimated diffuse albedo.

These methods demonstrate consistent progress in indoor inverse rendering. However, most approaches currently require either input geometry, reconstructed meshes, or strong scene priors. None of these methods jointly recovers geometry, material, and lighting from raw multi-view captures at the quality achieved by object-level methods.

3.2.2. *Object-level methods*

Object-level reconstruction under unknown illumination is the most studied setting in multi-view inverse rendering. The central design choices are the geometric representation, the lighting parameterization, and the specific differentiable rendering algorithm used to minimize the reconstruction loss.

A first popular solution for this task is based on **volume rendering**. Neural radiance fields (NeRF) [MST*21] revolutionized novel view synthesis by implicitly representing a scene as a continuous volumetric function, and were later extended to inverse rendering by explicitly decomposing network outputs into reflectance, illumination, and shape. NeRD [BBJ*21] decomposes the radiance field into a Disney-principled BRDF and a per-image illumination model encoded as a learned latent vector, enabling relighting under novel illumination conditions without committing to a fixed lighting geometry. NeRFactor [ZSD*21] distills the volumetric geometry of a pre-trained NeRF into a surface representation, recovering per-point surface normals and light visibility, and then jointly refines this geometry while solving for spatially varying BRDFs and environment lighting under a data-driven BRDF prior learned from real-world reflectance measurements. By explicitly modelling light visibility, NeRFactor separates cast shadows from albedo, a capability absent in both NeRD and PhySG [ZLW*21], which do not account for visibility. Neural-PIL [BJB*21] replaces the costly per-direction

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4: *Classification of multi-view inverse rendering methods reviewed in this survey.*

Reference	Year	Input				Output		Method		Scene description			Evaluation	Apps.
		View	Dyn.	Dep.	Act.	S/M/L	Scene	Approach	Sub-Ap.	Shape	Mat.	Light		
[BXS*20b]	2020	MV	-	-	✓	SM	Obj	NO	CNN	M, IM	P	-	IMD, NVSE	SE, LE
[BXS*20a]	2020	MV	-	-	✓	SML	Obj	DRO	PT	VD	P	EM	NVSE, EMD	LE, ME, SE
[BBJ*21]	2021	MV	-	-	-	SML	Obj	DRO	V	VD, M	P	SG	IMD, NVSE	-
[ZSD*21]	2021	MV	-	-	-	SML	Obj	DRO	V	VD	B, N	EM	IMD, RE	ME, LE
[SDZ*21]	2021	MV	-	-	-	SML	Obj	DRO	V	VD, IM	P, N	GI, LF	RE, NVSE	-
[BJB*21]	2021	MV	-	-	-	SML	Obj	DRO	V	VD, IM	P	EM	IMD, RE, NVSE	-
[ZLW*21]	2021	MV	-	-	-	SML	Obj	DRO	V	SDF, M	P, N	SG	IMD, MD, NVSE	-
[LZBD21]	2021	MV	-	-	✓	SM	Obj	DRO	PT	M	P	-	IMD, RE, NVSE	ME, SE
[ZLLS22]	2022	MV	-	-	✓	SML	Obj	DRO	V	VD, IM	P	-	IMD, ME	SE
[ZSH*22]	2022	MV	-	-	-	SML	Obj	DRO	V	SDF, IM	P	EM, GI	IMD, NVSE	-
[RES*22]	2022	MV	-	-	-	SML	Out	DRO	V	VD, IM	P, N	SH	RE	-
[DLLY22]	2022	MV	-	-	-	SML	Obj	DRO	V	SDF	P	GI	IMD, NVSE, EMD	-
[BEK*22]	2022	MV	-	-	-	SML	Obj	DRO	V	VD, IM	P	I	NVSE	-
[HHM22]	2022	MV	-	-	-	SML	Obj	DRO	PT	SDF, IM	P	EM	RE, NVSE	-
[MHS*22]	2022	MV	-	-	-	SML	Obj	DRO	V	SDF, IM	P, N	EM	IMD, RE, NVSE, EMD	SE
[YZL*22]	2022	MV	-	-	-	ML	Obj	DRO	PT	-	P	LF, GI	NVSE	-
[LTH*23]	2023	MV	-	-	-	ML	Gen	NO	DM	-	P	EM	IMD, NVSE	-
[ZHY*23]	2023	MV	-	-	-	SML	Ind	DRO	V	VD, SDF	P, N	GI	IMD, ME, NVSE	SE, LE
[CLP*23]	2023	MV	-	-	-	SML	Gen	NO	ViT	IM	P, N	I	IMD, RE	SE
[LWC*23]	2023	MV	-	-	-	SML	Ind	DRO	PT	M	P	EM, GI	IMD, RE	ME, LE
[YN23]	2023	MV	-	-	-	SML	Obj	DRO	PT	SDF	P	LF	-	-
[WHL*23]	2023	MV	-	-	-	SML	Obj	DRO	V	SDF	P, N	EM, GI	IMD, NVSE	-
[ZXY*23]	2023	MV	-	-	-	SML	Obj	DRO	V	VD, IM	N	SG	IMD, NVSE, RE	-
[LWL*23]	2023	MV	-	-	-	SML	Obj	DRO	V	SDF, IM	P	LF	ME, RE	-
[WSG*23]	2023	MV	-	-	-	SML	Out	DRO	V	SDF, M	P	EM	RE	-
[GGL*24]	2023	MV	-	-	-	SML	Obj	DRO	S	PD	P	EM, GI	IMD, NVSE	-
[YYC*23]	2023	MV	-	✓	-	ML	Ind	DRO	V	-	P, N	EM	RE	SE, ME
[JLX*23]	2023	MV	-	-	-	SML	Obj	DRO	V	VD	P	EM, GI	IMD, RE, NVSE	-
[WZY*23]	2023	MV	-	-	-	ML	Ind	DRO	PT	M	P	EM, GI	IMD, NVSE	ME, LE, SE
[LOU*24]	2024	MV	-	-	-	SML	Obj	DRO	V	VD	P	LF	-	-
[YGY*24]	2024	MV	-	-	-	SML	Obj	DRO	V	SDF, IM	P, N	SG, GI	IMD, RE	-
[JTL*24]	2024	MV	-	-	-	SML	Gen	DRO	S	VD	P	EM	NVSE, RE	-
[CCB24]	2024	MV	-	-	✓	SM	Obj	DRO	S	PD, SDF, IM	B, P	-	IMD, NVSE, RE	SE, LE
[LZF*24]	2024	MV	-	-	-	SML	Obj	DRO	S	PD, IM	P	SH, GI	IMD, RE, NVSE	LE
[CPY*24]	2024	MV	-	-	-	ML	Obj	NO	DM	SDF, M	P	EM, GI	IMD, RE	-
[DWZ*24]	2024	MV	-	-	-	SML	Obj	DRO	V	VD, M	P, N	EM, GI	IMD, ME, RE, NVSE	-
[LYX*24]	2024	MV	-	-	-	SML	Obj	DRO	PT	SDF, M	P	LF	NVSE, RE	-
[SCL*23]	2024	MV	-	-	-	SML	Obj	DRO	V	VD	P	EM	IMD, RE, NVSE, ME	-
[ERB*24]	2024	MV	-	-	-	SML	Obj	DRO	V	VD	P	I	NVSE, ME	-
[LWZW24]	2024	MV	-	-	-	SML	Obj	DRO	V	SDF, M	P, N	GI	IMD, NVSE, RE	-
[WTL*24]	2024	MV	-	-	-	SML	Gen	DRO	V	VD	P	SG, EM	IMD, RE, NVSE	-
[FTTS24]	2024	MV	-	-	✓	SML	Gen	DRO	V	SDF, M	P, N	SG, SH	IMD, RE, NVSE	-
[BZZ*24]	2024	MV	-	-	✓	SML	Obj	DRO	S	PD	P, B, N	GI	RE	-
[DWW*24]	2024	MV	-	-	✓	SM	Obj	DRO	PT	M	P	-	IMD, RE	ME
[CCB25]	2025	MV	-	-	✓	SM	Obj	DRO	S	PD, M	B, P	-	IMD	SE
[SWW*25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD	P	EM, GI	RE, NVSE, EMD	-
[LHG*25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD, IM	P, N	EM	IMD, RE, NVSE, EMD	-
[LLJ*25]	2025	MV	-	-	-	SML	Gen	DRO	S	PD, IM	B, P	EM, GI	IMD, NVSE	-
[LOU*25]	2025	MV	-	-	-	SM	Obj	DRO	V	SDF	P	-	IMD	-
[YGL*25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD, M	P	EM	IMD, RE, NVSE	-
[CLZ25]	2025	MV	-	-	-	SML	Gen	DRO	S	PD	P	EM, GI	IMD, RE, NVSE	LE
[DLDN25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD	B, P	EM	NVSE	-
[CCL*25]	2025	MV	✓	✓	-	SML	Gen	DRO	S	PD	P	SH, GI	IMD, RE, NVSE	-
[GWZ*25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD, IM	P	EM, GI	IMD, NVSE	-
[LHL*25]	2025	MV	-	-	-	SML	Ind	DRO	PT	SDF, M	P	I	IMD, NVSE	SE
[CLP*25]	2025	MV	-	-	-	SML	Gen	NO	ViT	IM	P, N	I, GI	IMD, RE	SE, ME
[LPD*25]	2025	MV	-	-	-	SML	Obj	NO	DM	IM	P	EM	IMD, NVSE	-
[WBB*25]	2025	MV	-	-	-	SML	Obj	DRO	V	SDF	P	EM	EMD, NVSE	-
[BCC25]	2025	MV	-	-	-	SML	Obj	DRO	V	SDF, M	P	LF	IMD, NVSE	-
[WLLL25]	2025	MV	-	-	-	SML	Ind	DRO	V	SDF	P, N	EM, GI	IMD, NVSE	-
[SGX*25]	2025	MV	-	-	-	SML	Gen	DRO	S	PD	P	GI, EM	RE, NVSE	-
[GWZZ25]	2025	MV	-	-	-	SML	Obj	DRO	V	SDF, M	P	EM, GI	IMD, RE, NVSE	-
[TCZD25]	2025	MV	-	-	✓	SML	Obj	DRO	V	SDF, M	P, B	GI	NVSE, ME	-
[HYD*25]	2025	MV	-	-	-	SML	Obj	DRO	S	PD, SDF, M	P	SG	IMD, NVSE	-
[JSL*25]	2025	MV	-	-	-	SML	Gen	DRO	S	PD	P	EM, GI	IMD, NVSE, RE	ME, LE, SE
[WRG*25]	2025	MV	-	-	✓	SM	Obj	DRO	PT	M	P	EM	IMD, RE	ME, LE

Notes: Apps: applications demonstrated (Please see Table 2 for symbol legend).
A dash (–) indicates that the feature is not applicable.
Abbreviations: Dyn. = dynamic data, Dep. = depth enhanced input, Act = active capture setup.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

illumination integral with a learned low-dimensional approximation of the BRDF-environment convolution, and NeRV [SDZ*21] adds explicit visibility modelling for accurate shadow handling. PhySG [ZLW*21] establishes the pipeline followed by most subsequent volumetric methods, combining a neural SDF with spherical Gaussian illumination and analytical microfacet BRDFs optimized end-to-end via differentiable sphere tracing.

Subsequent works expand volumetric representation to handle complex indirect lighting and unknown camera poses. IRON [ZLLS22] combines volumetric and surface-based rendering to produce clean meshes with spatially varying BRDFs. Zhang et al. [ZSH*22] derive indirect lighting directly from the pre-learned radiance field, whereas Munkberg et al. [MHS*22] extract triangle meshes optimized for downstream rendering applications. SAMURAI [BEK*22] and SHINOBI [ERB*24] address the more challenging setting of in-the-wild image collections that are both unposed and unordered, images of the same object gathered from the internet without controlled capture, by jointly optimizing camera parameters, geometry, material, and illumination simultaneously. This setting is fundamentally more challenging than posed multi-view capture because the pose optimization introduces an additional source of gradient interference with the material and illumination estimates. Deng et al. [DLLY22] formulate physics-based indirect illumination as a separate stage that helps convergence under strong inter-reflections.

Latest volumetric methods diversify the core SDF with volume rendering. NeRO [LWL*23] integrate the neural BRDF with implicit surfaces for strongly reflective objects, while RobIR [YGY*24] adds shadow handling for high-contrast illumination scenes. Near-field indirect illumination was modelled in NeFII [WHL*23] via neural radiance caching, whereas NeMF [ZXY*23] replaces the surface representation with a neural microflake fields to capture translucent and intricate geometries. Neural-PBIR [SCL*23] combines neural-field initialization with physics-based optimization refinement. [DWZ*24] implemented explicit path tracing with reservoir sampling, and NeISF [LOU*24] adds polarization as an extra cue for difficult materials. Recent work address residual estimation errors, where PBR-NeRF [WBB*25] adds physics-based BRDF priors to the neural-field optimization, PIR [BCC25] jointly optimises light-source position and indirect illumination through importance sampling, NeISF++ [LOU*25] extends polarized inverse rendering to both conductors and dielectrics, and TIRPL [TCZD25] combines volume rendering with path tracing for point-light capture configurations.

Tensorial representations replace dense multi-layer perceptron networks with low-rank tensor decomposition. TensoIR [JLX*23] demonstrated this and extended TensoRF to inverse rendering, jointly estimating scene parameters from multi-view images under unknown lighting, and this approach has been extended by TensoSDF [LWZW24] with roughness awareness and by TensoFlow [GWZZ25] with a learned importance sampler.

3D Gaussian Splatting (3DGS) [KKLD23] introduces an explicit point-based representation with real-time differentiable rendering. Extending Gaussian splatting to inverse rendering require per-Gaussian shading parameters and indirect light handling without ray tracing. GS-IR [LZF*24] attaches depth-derived normals to each Gaussian and bakes indirect light into a precomputed occlusion model. GIR [SWW*25] adds self regularised normals and voxel based indirect illumination tracing. Relightable 3D Gaussians [GGL*24] attach explicit BRDF parameters with ray traced shadows.

GlossyGS [LHG*25] adds microfacet material priors, GlossGau [DLDN25] integrates anisotropic spherical Gaussians into each primitive, TransparentGS [HYD*25] adds transparent primitives with deferred refraction, and GeoSplatting [YGL*25] combines point-based splatting with an explicit mesh that guides surface normals.

Recent splatting architectures focuses on comprehensive global illumination modelling. GS3 [BZZ*24] uses triple Gaussian formulation (spatial, angular, residual) to capture self-shadowing effects, while IRGS [GWZ*25] traces inter-reflections through 2D Gaussian ray tracing. The basis BRDF methods [CCB24, CCB25] completed this branch with point-based splatting using an interpretable BRDF basis instead of a parametric one.

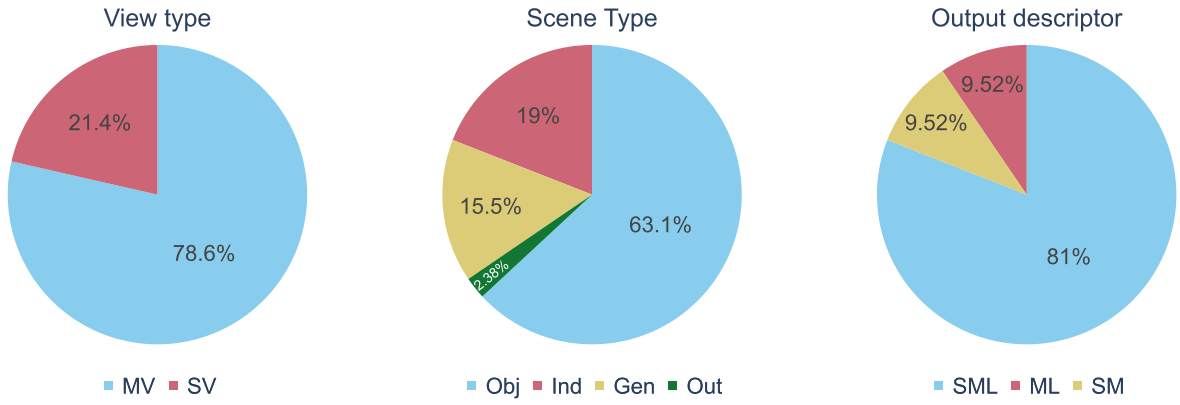
Path tracing based methods solve the inverse-rendering problem with differentiable Monte Carlo path tracing or photometric acquisition. These works [BXS*20b, BXS*20a] capture geometry and reflectance from sparse multi-view photometric images using collocated point lights, first utilizing a convolutional network backbone and then with a relightable volumetric extension. Luan et al. [LZBD21] unify shape and SVBRDF recovery in a single differentiable Monte Carlo optimization. Hasselgren et al. [HHM22] integrates multiple importance sampled Monte Carlo into the inner loop with Monte-Carlo denoising networks. NDJIR [YN23] uses voxel-grid features with Bayesian priors to regularise material uncertainty under joint optimization. NeILF [YZL*22] parameterizes the full incident light field as a neural field, fitting it through differentiable rendering. Ling et al. [LYX*24] treat the neural field as a non-distant environment emitter inside a physics-based pipeline. Deng et al. [DWW*24] recover translucent thin objects (leaves, paper) with a layered position-free volumetric BSDF, and Weier et al. [WRG*25] extend this to textured translucent appearance, recovering scattering parameters and surface texture together.

A more recent approach replaces the per-scene optimization loop with a **learned generative prior** capable of predicting unobserved scene factors. IntrinsicAnything [CPY*24] learns separate diffusion priors for diffuse albedo and specular reflectance, then uses them as conditioning during multi-view material recovery under unknown illumination. MaterialFusion [LPD*25] couples a Stable Diffusion texture prior with a differentiable rendering loop, regularizing the material estimate against the prior at every optimization step. These methods demonstrate that data-driven learned priors can successfully compensate for the ambiguities that challenge purely per-scene optimization.

Most of the inverse rendering field demonstrates convergence toward parametric material models [DLLY22, WRG*25], while learned basis decompositions [ZSD*21] and neural BRDFs [ZXY*23] remain minority choices. The rendering algorithm correlates directly with the specific use case. Volume rendering provides topological flexibility, splatting ensures inference efficiency. Path tracing can provide physical accuracy, while

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) (b) (c)

Figure 6: *Distributions of characteristics of the papers surveyed (a) View types (SV=single View, MV=multi-view); (b) Scene types (Obj=Object-level, Ind=Indoor, Out=Outdoor, Gen=generic); (c) Output description (SML=shape, materials and light; SM=shape and materials; ML=materials and light)*

diffusion priors can solve the issues due to ambiguous observations. Given the state-of-the art of existing benchmarks, it is difficult to evaluate comparatively the four paradigm with a single evaluation (see Section 4.4.5).

3.2.3. *Outdoor scenes*

Outdoor inverse rendering is particularly more challenging due to large spatial extents, uncontrolled illumination, atmospheric effects, and temporal variation.

Two methods address outdoor reconstruction explicitly. Rudnev et al. [RES*22] extended radiance fields to outdoor relighting tasks using spherical harmonics to capture the dominant sky and sun illumination model, supporting joint appearance and viewpoint editing. Wang et al. [WSG*23] optimized for urban scenes by combining neural fields for primary rays with explicit meshes for the secondary rays that compute cast shadows, an architectural choice that scales to street-level captures. InvRGB+L [CCL*25], which is also listed under generic methods, operates on autonomous-driving captures and uses LiDAR intensity cues alongside RGB to provide complementary spectral data in large-scale outdoor scenes.

Outdoor inverse rendering remains less explored and is more challenging due to the modelling large-scale, uncontrolled scenes. Consequently, robust recovery of geometry, materials, and illumination under real-world environmental variation remains an open research problem.

3.2.4. *Generic scenes*

Recent methods increasingly seek to generalize across scene categories rather than specializing to a particular environment. Generic inverse rendering methods focus on multi-view captures without a specific or controlled scene type. MAIR [CLP*23] and MAIR++ [CLP*25] use vision-transformer attention across views to fuse multi-view features into per-pixel geometry, material, and lighting predictions. Diffusion Posterior Illumination [LTH*23] utilizes a diffusion prior over illumination in multi-view material recovery, compensating for the inherent ambiguity. VMINer [FTTS24] evaluates on both object and scene captures with near- and far-field illumination. InvRGB+L [CCL*25] jointly recovers material and lighting from RGB and LiDAR data on autonomous-driving captures, an underexplored modality combination. Several splatting based methods already discussed in section 3.2.2, (GaussianShader [JTL*24], GI-GS [CLZ25], GUS-IR [LLJ*25], SVG-IR [SGX*25], [JSL*25]), and the volumetric based UniVoxel framework [WTL*24] work on generic scenes.

These generic frameworks establish a baseline of methods that do not require scene-type-specific tuning. However, they regularly underperform specialized ones on specific scene types. No single method reported in the literature we surveyed consistently matches the performance of specialized methods across object, indoor, and outdoor settings.

4. Discussion

The large number of surveyed papers demonstrates the growing interest in inverse rendering in the Visual Computing community over the last 6 years. By analyzing the literature in detail, we can identify current research trends and reveal existing gaps that should be addressed in future work. The total number of papers considered is 84. Some of their characteristics are summarized in Figure 6. Most of them work on multi-view captures (a), most of them use passive imaging, and are mostly developed for object-level reconstruction, not for indoor/outdoor or generic scenes (Figure 6b). The number of papers that present non-complete parameter reconstruction, working on known object shapes (ML) or not recovering lighting (SM), is small relative to those that reconstruct all from image data (SML) (Figure 6c). Only 4 papers address depth-enhanced sensor data, and 13 are designed to work with active capture settings.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

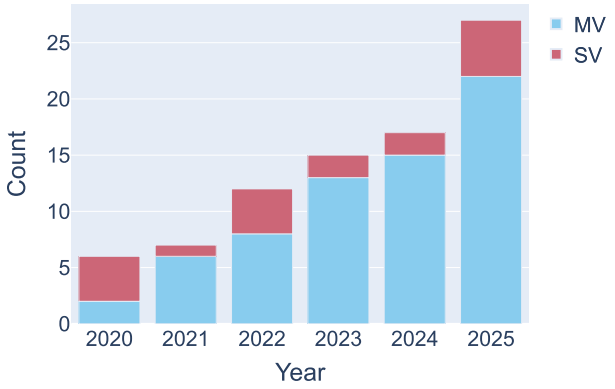


Figure 7: *Stacked bar showing the number of single-view and multi-view inverse rendering papers surveyed, published in the last 6 years.*

4.1. Research trends

While the number of papers published appears to increase linearly, it is worth noting that this trend is mainly driven by the enormous impact of NeRF and Gaussian splatting in the domain of Visual Computing. Figure 7 shows a stacked bar chart showing that the number of papers has increased rapidly in recent years (data from 2025 are clearly incomplete). However, it should be noted that the trends are not uniform across the different method categories. Figure 8a shows the evolution of the number of articles on single-view methods for the different computing approaches. In this case, we see that after an initial hype around CNN-based methods in 2020, we have very few papers in 2021–2023, followed by recent growth due to the adoption of Diffusion Models. For the multi-view methods (Figure 8b), it is clear that the rapidly growing number of papers stems from the exploitation of the novel light-field encoding structures recently introduced. We did not find any papers of this kind for 2020. Inverse rendering methods for processing classical multi-view stereo input appeared in 2021, following the introduction of NeRFs, and the growth received another boost with the introduction of Gaussian Splatting, which has been used in most of the papers that appeared in 2025.

It is also interesting to see which are the most popular choices for representing shape, materials, and light. Figure 9a shows that the output representations provided for shape are quite varied, and in many cases the frameworks provide multiple representations, integrating SDFs with meshes and also point-based representations.

Figure 9b shows the output representations provided for the materials. Most methods recover parametric BRDFs (P), with a smaller subset using neural BRDF (N) [ZXY*23, ZXY*23] or basis representations (B) [CCB25, CCB24].

Figure 9c shows the output representations provided for light. Here, the situation is more complex, as different types of illumination are modelled and techniques are often combined. Most methods focus on estimating distant illumination, while some recover global illumination effects as well. More importantly, the physical lighting model is often simplified, both to decrease computation and since the problem is ill-defined. We did not classify methods based on these approximations since they vary in detail in different methods. These simplifications have an impact on the accuracy of the final estimation. As an example, simplifying the shadowing effects might induce errors in the estimation of material properties.

Figure 10a shows the use of the different evaluation methods discussed in our taxonomy. Most methods are based on image-space metrics related to maps (with possible related biases). Relighting and novel-view synthesis are popularly used in multi-view settings. Only a few methods evaluate the quality of the reconstructed models in terms of the complete shape and the properties of the material. Figure 10b shows the number of papers showing examples of material (ME), Light (LE) and shape (SE) editing. They are only a limited part of the total number of papers. We will discuss the limitations in the application focus of the papers.

4.2. Comparison of different approaches

As described in Section 2.7, datasets and evaluation metrics used to compare the results of the different methods are quite heterogeneous. However, it is possible to provide a quantitative and qualitative comparison of representative inverse rendering methods on selected benchmarks, enabling direct assessment of relative

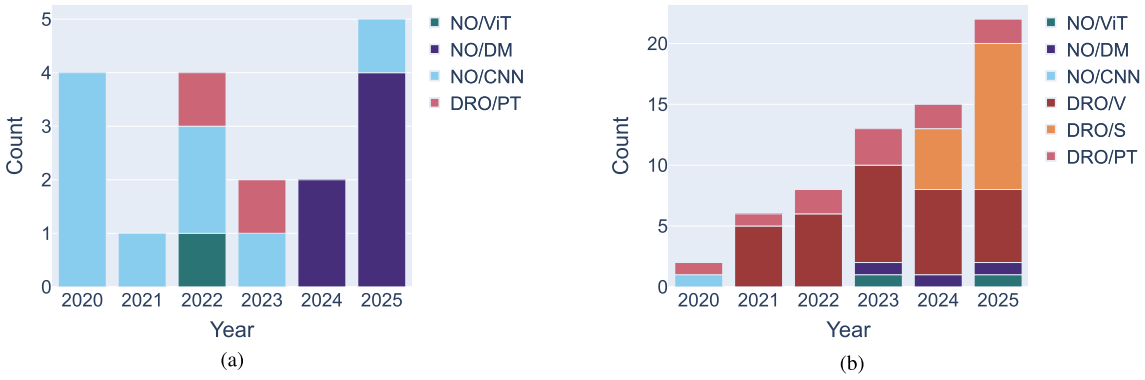
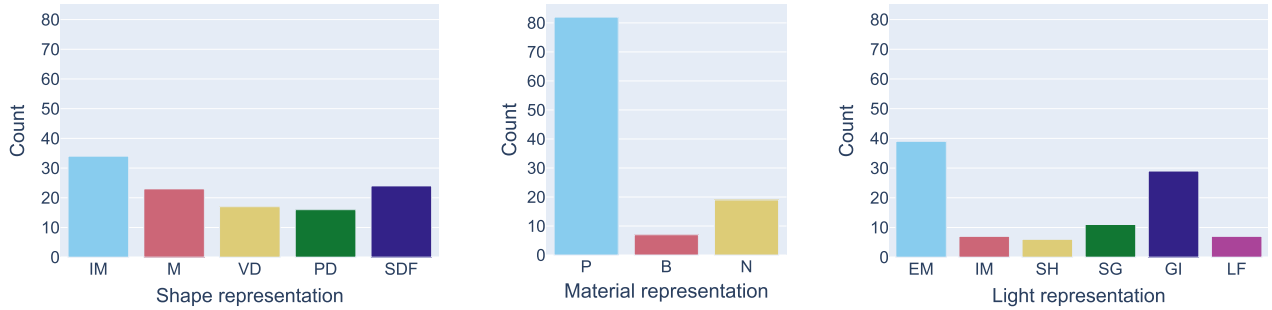


Figure 8: *Stacked bars showing the number of papers surveyed classified by different computing approaches for the single-view input (a) and the multi-view input (b)*

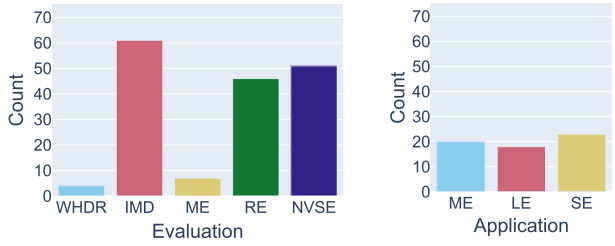
© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) (b) (c)

Figure 9: *Occurrences of the different shape output encodings in the surveyed papers (a); of the different output material encodings (b), of the different light representation choices (c)*



(a) (b)

Figure 10: *Occurrences of the different evaluation methods adopted (a) and of the applications demonstrated in the papers (b).*

performance. We organize comparisons by input configuration (single-view vs. multi-view), as this factor fundamentally affects reconstruction challenges and quality. All metrics are taken from reported results in the original papers on shared evaluation scenes, ensuring fair comparison. We focus on methods spanning various representations, including NeRF-based volumetric, factorized tensorial, and Gaussian splatting, to demonstrate how representation choices affect reconstruction quality across scenes. In Table 5, we compare five **single-view inverse rendering methods** on the Synthetic InteriorVerse benchmark [ZLH*22], which features complex indoor scenes with ground truth for albedo, metallic, roughness, normal, and depth estimation from single images. Similarly, in Table 6, we compare state-of-the-art **multi-view inverse rendering methods** on TensoIR Synthetic Dataset [JLX*23], which is the most commonly used benchmark for evaluating multi-view inverse rendering methods. The dataset contains four complex synthetic objects (ficus, lego, armadillo, and hotdog), with 100 training views and 200 test views under multiple environment maps.

Figures 11 and 12 provide qualitative comparisons of representative single-view and multi-view methods on shared benchmark scenes, complementing the quantitative evaluation summarized in Tables 5 and 6.

Figure 11 shows a qualitative comparison of three representative single-view methods, [LSR*20], [ZLH*22], and [ZZLZ25], on the same indoor scene from the InteriorVerse Dataset [ZLH*22]. The albedo row validate shadow removal quality, As [LSR*20] produces a washed-out, overexposed result in which illumination gradients are visibly absorbed into the estimated reflectance, while [ZLH*22] partially resolves this but still has specular highlights on the wall with surface color. [ZZLZ25] achieves the closest agreement with the ground truth albedo, recovering clean material boundaries and suppressing cast shadows more effectively. Similarly roughness, normal and depth estimation follow a similar trend, with [ZZLZ25] achieve a fair accuracy, outperforms other by approximately 9.6 dB

Table 5: *Quantitative comparison of single-view inverse rendering methods on the Synthetic InteriorVerse benchmark [ZLH*22]. Results include albedo accuracy (PSNR/SSIM/LPIPS) and intrinsic decomposition quality measured via IIW dataset [BBS14] Weighted Human Disagreement Rate (WHDR). Material estimation accuracy is assessed using reconstruction errors for metallic and roughness components. For the surface geometry, we report the mean angular error for the normals and the Absolute Mean Relative Error (AMRE) for the depth. These comparisons highlight how different approaches handle the highly ill-posed nature of single-view inverse rendering and the variations in material consistency, geometric detail, and overall intrinsic scene estimation.*

	Albedo			Albedo	Metallic			Roughness			Normal	Depth
Method	PSNR ↑	SSIM ↑	LPIPS ↓	WHDR ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Angular Error ↓	AMRE ↓
[ZZLZ25]	**21.95**	**0.87**	0.12	20.90	17.64	0.63	**0.22**	**16.72**	**0.71**	**0.24**	12.23°	0.08
[WTC*25]	19.53	0.81	**0.11**	**19.70**	**18.66**	**0.70**	0.24	16.63	0.66	0.29	**11.36°**	**0.07**
[KSN24]	17.42	0.80	0.22	22.02	16.46	0.50	0.29	13.33	0.57	0.34	—	—
[ZLH*22]	15.92	0.78	0.34	22.90	16.72	0.60	0.27	16.13	0.68	0.29	15.41°	0.13
[LSR*20]	12.31	0.68	0.52	21.99	—	—	—	11.76	0.53	0.45	26.07°	0.25

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6: *Quantitative comparison of multi-view inverse rendering methods on the TensoIR Synthetic dataset [JLX*23], which contains four synthetic objects with 100 training views and 200 test views under multiple environment maps. We report novel view synthesis (NVS), relighting under unseen environment maps, albedo reconstruction, and surface normal accuracy (mean angular error, in degrees). The* Run Time *reports approximate per-scene training time. The* Run Time *are reported from the comparison tables of the papers that use different hardware, and are therefore not directly comparable. The training time for [JLX*23, JTL*24, SGX*25] are reported from [SGX*25](NVIDIA RTX 4090); for [ZSD*21, ZSH*22, MHS*22, LZF*24, CLZ25] from [CLZ25] (NVIDIA A5000); and for [YGL*25] from the original paper (NVIDIA RTX 4090). Together, these metrics show the trade-offs between rendering Quality, material accuracy, and computational efficiency across diverse representative cases.*

	NVS			Relighting			Albedo			Normal	Computing
Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Angular Error ↓	Run Time ↓
[ZLW*21]	23.42	**0.99**	0.07	22.63	**0.97**	0.07	21.80	**0.97**	0.18	—	—
[ZSD*21]	24.68	0.92	0.12	23.38	0.91	0.13	25.12	0.94	0.11	6.31	>100 hrs
[ZSH*22]	27.37	0.93	0.09	23.97	0.90	0.10	27.34	0.93	0.10	5.07	14 hrs
[MHS*22]	30.70	0.96	0.05	19.88	0.88	0.10	29.17	0.91	0.12	6.08	< 1 hr
[JLX*23]	35.09	0.98	0.04	28.58	0.94	0.08	29.27	0.95	0.09	4.10	5 hrs
[LZF*24]	35.33	0.97	0.04	24.37	0.89	0.10	30.29	0.94	0.08	4.94	33 min
[JTL*24]	37.57	0.98	**0.02**	23.37	0.91	0.08	—	—	—	6.52	0.5 hr
[YGL*25]	31.14	0.95	0.08	29.95	0.94	**0.05**	29.41	0.94	**0.07**	**4.08**	**14 min**
[SWW*25]	37.63	0.98	0.03	28.88	0.94	0.06	—	—	—	—	—
[CLZ25]	36.75	0.97	0.04	24.70	0.89	0.11	**31.97**	0.94	0.09	**4.08**	28 min
[SGX*25]	36.71	0.97	0.03	**31.09**	0.95	0.06	—	—	—	—	1 hr
[LLJ*25]	**37.61**	0.98	**0.02**	27.53	0.92	0.09	—	—	—	4.49	—

in albedo PSNR and reduces the mean angular normal error from 26.07° to 12.23°.

Figure 12 presents a qualitative comparison of six representative multi-view methods on the *armadillo* and *hotdog* scenes of the TensoIR Synthetic dataset [JLX*23]. The most prominent differences showed in the albedo and normal columns. GI-GS [CLZ25] produces dark albedo on the armadillo, a direct consequence of poor indirect illumination separation causing shading to be absorbed into the material estimate. GS-IR [LZF*24] recovers a more plausible albedo for the same scene, though at the cost of weaker relighting performance as compared to TensoIR [JLX*23], consistent with Table 6. On the hotdog scene, NeRFactor [ZSD*21] produces flat, low-contrast albedo owing to its Lambertian reflectance assumption, which cannot model specular highlights and causes them to bleed into the recovered material. GeoSplatting [YGL*25] and TensoIR [JLX*23] recover visually cleaner albedo maps, with TensoIR [JLX*23] producing the most consistent relighting result owing to its explicit second-bounce ray marching for indirect illumination. SVG-IR [SGX*25] does not report roughness on this benchmark (N/A), and its normal column exhibits visible artifacts in the highlighted region (red box) on the plate surface. In essence, the figure illustrates that high NVS performance does not imply accurate material decomposition, as GI-GS [CLZ25] achieves competitive NVS PSNR in Table 6 yet produces the most severely degraded albedo in the qualitative comparison, underscoring the need to evaluate inverse rendering methods on all output modalities jointly.

Tables 5 and 6 show relevant differences in the metrics estimated across inverse rendering methods. Compared with single-view methods, multi-view methods seem to achieve far better performance, as expected. In particular, Gaussian Splatting-based methods achieve 37–38 dB novel-view PSNR, surface normal errors around 5°, and stable relighting above 28–31 dB, demonstrating the benefit of explicit point-primitives representations. While the earlier NeRF-based volumetric representation methods achieve about 23–27 dB NVS PSNR and require hours to days of training time, they show the limitations of MLP-based volumetric fields for material-aware inverse rendering. Factorized volumetric methods TensoIR [JLX*23] reduce memory cost and improve stability, through a more compact scene representation, but remain behind Gaussian approaches both in quality and speed.

Relighting comparisons show that methods that explicitly model spatially varying BRDFs or global illumination [CLZ25, SGX*25] achieve the most consistent material–lighting decomposition. Approaches primarily designed for novel view synthesis [MHS*22, JTL*24] typically do not yield good results in relighting and exhibit incomplete material lighting separation. This fact suggests that view synthesis quality alone is insufficient for reliable inverse rendering; accurate material decomposition requires modelling both local shading and indirect light transport. Multi-view methods [YGL*25] successfully address this by combining explicit geometry with per-primitive materials and physically-based shading formulations.

Single-view inverse rendering methods have made significant progress but remain far behind multi-view performance. Even the recent diffusion-based approaches achieve only 17–22 dB albedo PSNR and normal errors about 11–12°, reflecting the inherent ambiguity of decomposing shape, materials, and lighting from a single input image. While generative priors improve perceptual realism, they tend to generalize poorly outside the training data, limiting their use for physically consistent relighting. As a result, single-view approaches are currently more effective for material editing and appearance transfer than for physically accurate reconstruction.

Efficiency comparisons show that multi-view training times have decreased from hours or days [ZSD*21, ZLW*21] to minutes

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

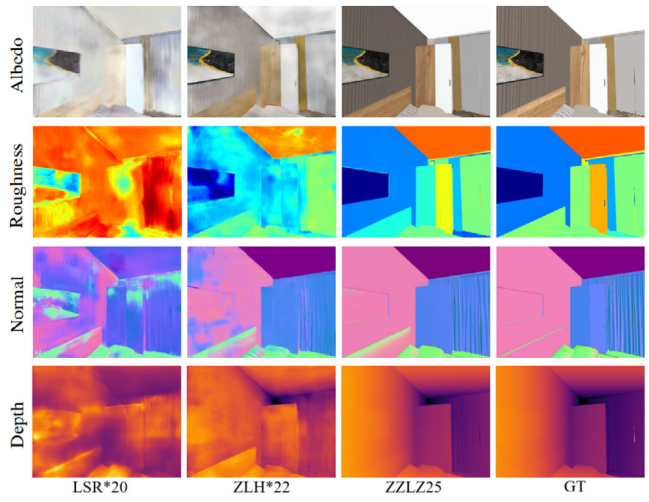


Figure 11: *Qualitative comparison of representative single-view inverse rendering methods on an indoor scene from the InteriorVerse Dataset [ZLH*22]. Rows show, from top to bottom, the estimated albedo, roughness, surface normal, and depth maps. Columns correspond to [LSR*20], [ZLH*22], [ZZLZ25], and the ground truth (GT). The comparison highlights qualitative differences in shadow removal from albedo, material roughness estimation, surface normal coherence, and depth accuracy across methods, complementing the quantitative metrics reported in Table 5. Image from: Zheng et al., 2025 https://openreview.net/forum?id=m6IisFwo2t CC BY 4.0*

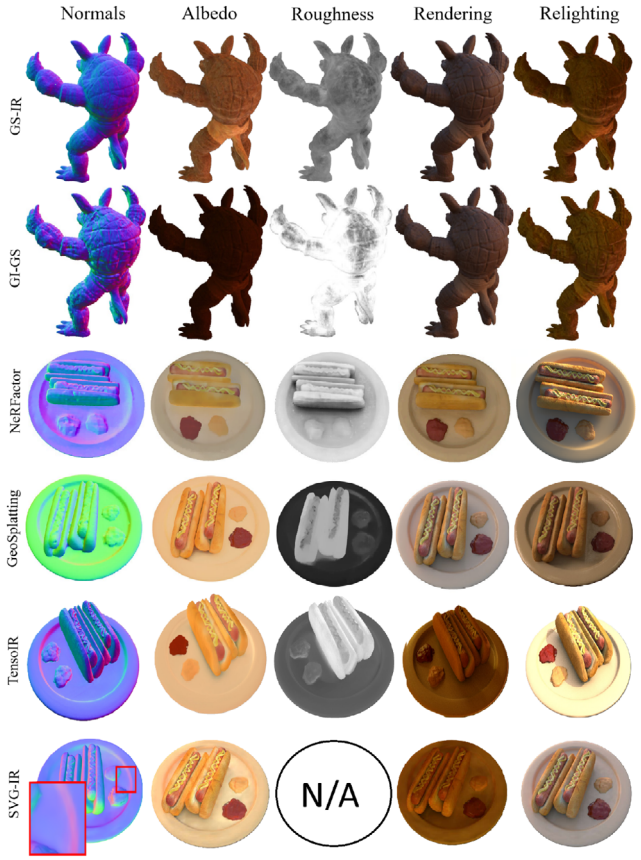


Figure 12: *Qualitative comparison of representative multi-view inverse rendering methods on the* armadillo *(rows 1–2) and* hotdog *(rows 3–6) scenes from the TensoIR Synthetic dataset [JLX*23]. Columns show, from left to right, the estimated surface normals, albedo, roughness, PBR rendering, and relighting result under a novel environment map. Images are sourced from the original publications. GS-IR [LZF*24] and GI-GS [CLZ25] are shown on the armadillo scene, while NeRFactor [ZSD*21], GeoSplatting [YGL*25], TensoIR [JLX*23], and SVG-IR [SGX*25] are shown on the hotdog scene. N/A indicates that the corresponding output is not reported by the original authors for this benchmark. The comparison illustrates qualitative differences in normal accuracy, albedo-lighting disentanglement, and relighting consistency across volumetric, splatting-based, and hybrid representations, complementing the quantitative metrics reported in Table 6. Images from: Liang et al., 2024, https://openreview.net/forum?id=nJlBRkYS7q, CC BY-SA 4.0; Chen et al., 2025 https://openreview.net/forum?id=hJIEtJlvhL CC BY 4.0; Zhang et al., 2021, https://dl.acm.org/doi/10.1145/3478513.3480496, CC BY 4.0; Ye et al., 2025 https://openreview.net/forum?id=DsonaFDlXO CC BY 4.0; Jin et al. 2023 https://openreview.net/forum?id=4ULck9r3mY CC BY-SA 4.0; Sun et al., 2025, https://learner-shx.github.io/project_pages/SVG-IR/index, CC BY-SA 4.0.*

[YGL*25, CLZ25, LLJ*25]. At the same time, reconstruction quality has simultaneously improved by more than 10 dB in some benchmarks. Single-view methods offer near-instant inference but require extensive offline training and still deliver significantly lower physical accuracy. Overall, the comparison demonstrates that representation choice remains the key factor influencing both performance and computational behavior across inverse rendering methods.

4.3. Pros and cons of computing approaches

Apart from the numerical results, it is also important to discuss the specific characteristics of the different approaches that may condition the choice of the best-suited approach for a particular task. Within the **Network Optimization** category, CNN-based methods [SC20, WTC*25] offer fast feed-forward inference, strong generalization within training distributions, and the ability to learn robust priors. They perform well on dense, labeled datasets (e.g. [ZLH*22, LYS*21]), making them ideal for single-view material and lighting estimation. However, they are heavily dependent on supervised data, have poor generalization to unseen materials, and cannot effectively model complex global illumination or inter-reflections due to their local receptive field [WPFK21]. Transformer-based methods [CLP*25] extend CNNs by introducing a global attention mechanism to capture long-range dependencies and to provide consistent estimates of depth, albedo, and lighting. Despite their expressiveness, they are computationally expensive, require massive training datasets, and still suffer from ambiguity in physically complex

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

lighting conditions [ZLM*22]. Diffusion-based methods [LGL*25, CPY*24] introduce probabilistic modelling into inverse rendering, generating diverse plausible decompositions and capturing uncertainty in lighting and material estimation. They use robust generative priors from large image datasets, making them valuable when ground-truth supervision is limited. Generated materials and lighting often achieve exceptional perceptual quality. However, they are computationally expensive to train and infer due to iterative denoising steps and may bias predictions toward their training distribution rather than true scene physics [ZZLZ25].

Within the **Differentiable Rendering Optimization** category, methods differ primarily by the rendering algorithm employed. NeRF-based volumetric methods [LWL*23, SDZ*21] learn continuous volumetric radiance fields from posed multi-view images and achieve high-quality geometry and appearance reconstruction with implicit 3D consistency. Their strength lies in physically grounded rendering via volume integration and view-dependent effects, but they require dense multi-view capture (50–100+ images) with precise calibration [ZSD*21]. These methods are slow to train, often taking hours to days to model a single scene at moderate resolution on a single GPU [DLZR22], and often assume distant lighting, struggling with spatially-varying illumination and indirect lighting.

Splatting-based methods [LZF*24, SWW*25] replace implicit volumetric fields with explicit point primitives, allowing real-time rendering (30–100+ FPS) and faster optimization (10–30 min) [SWW*25]. These methods handle geometry efficiently and achieve superior novel-view synthesis while providing better gradient flow and memory efficiency than volumetric methods. However, point primitives cannot naturally model volumetric effects such as fog and subsurface scattering, which can cause visible artifacts on smooth surfaces. Millions of generated Gaussians also create storage challenges, and individual Gaussians can introduce surface inconsistencies when multiple overlapping primitives represent the same surface point, producing artifacts visible on smooth curved surfaces [KKLD23, LZF*24]. Finally, Path Tracing-based methods [YN23, LHL*25] estimate geometry, BRDFs, and illumination by differentiably simulating full light transport, including multi-bounce indirect illumination and shadows, offering the highest level of physical accuracy while remaining compatible with standard rendering pipelines. Despite their accuracy, these methods are computationally expensive, sensitive to initialization, and require well-calibrated multi-view or HDR datasets, making them less practical for uncontrolled real-world environments [ZLLS22].

4.4. Issues and research gaps

Despite the results obtained with recent techniques, this literature review reveals many aspects that warrant further investigation. The largest concern is that inverse rendering is an ill-posed problem, intrinsically ambiguous in its definition. Also, the reconstruction performed depends on the rendering models used, but most adopted models are significant simplifications of image formation physics and are quite heterogeneous in terms of shape, materials, and lighting models. Furthermore, the representations of shape, materials, and lighting used during the optimization differ from the output representation (classified in our taxonomy), potentially introducing additional errors. In the following sections, we summarize some of the critical points we identified in our analysis.

4.4.1. *Dynamic scene inverse rendering*

Dynamic inverse rendering, the estimation of time-varying geometry, materials, and lighting from video sequences, remains one of the least explored and most impactful research directions in the field of inverse rendering. While static inverse rendering has matured enough to separate reflectance from illumination, the temporal dimension in dynamic scenarios introduces severe ambiguities, as the appearance changes between frames may depend on factors such as object motion revealing new surfaces, material property changes, lighting variations, or view-dependent effects from camera motion. Addressing these challenges while maintaining temporal consistency is significantly harder than static scene inverse rendering. Consequently, most dynamic scene methods focus on novel view synthesis, baking light into appearance representation rather than achieving a full inverse rendering pipeline [PCPMMN21, PSH*21]. Despite recent efficiency improvements through dynamic Gaussian splatting [LKLR24, WYF*24] and factorized space-time representations [CJ23], Computational cost remains a primary barrier for real-time inverse rendering applications, as video sequences containing hundreds of frames require substantial processing time. Potential novel research directions in dynamic scene inverse rendering include: **(1)** extending successful static inverse rendering techniques to canonical space representations in which materials remain time-invariant and only geometry deforms. **(2)** developing methods that jointly estimate dynamic geometry and time-varying lighting rather than assuming static illumination, **(3)** creating comprehensive benchmarks with ground truth dynamic geometry, materials, and lighting to enable rigorous evaluation beyond current novel view synthesis metrics, and **(4)** achieving real-time performance through efficient representations for AR and robotics applications. As these challenges are addressed, dynamic inverse rendering will unlock transformative capabilities in applications including relightable performance capture in film production, free-viewpoint video for immersive media, augmented reality with dynamic scene understanding, and autonomous systems requiring material-aware scene prediction.

4.4.2. *Diversity of tasks and methods*

A first major issue in the analysis of the literature concerns the diversity of tasks and approaches presented in the papers.

Single-view solutions are quite different from multi-view ones, with the former typically handled with Network Optimization (NO) approaches such as CNNs, Vision Transformers, and Diffusion Models, while the latter more frequently employ Differentiable Rendering Optimization (DRO) frameworks built on Structure from Motion and physically-based rendering pipelines. Multi-view tasks are also varied, as there are methods that optimize all parameters simultaneously, others that reconstruct shape first and then recover the other parameters [LHL*25], and others that require known geometry. While most methods are designed for generic images or multi-view captures typical of photogrammetry, some process different input data typical of specific applications (e.g., active/flash

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

illumination for scanning, Lidar for automotive, etc.). The reconstructed lighting models are also quite varied, requiring different lighting model approximations for each task. Not all the characteristics of specific methods are explicitly stated in the papers, which can make it difficult to compare methods and select the best one for a given application.

4.4.3. *Volumetric versus surface rendering*

Most recent rendering methods used in reconstruction rely on volumetric representations (NeRFs or 3DGS), while others rely on surface-based rendering (SDF, Polygon Mesh, Point Cloud). Volumetric approaches reconstruct complex geometry and view-dependent effects directly from images without requiring an explicit geometry prior, making it easy to handle non-smooth geometry, such as plants. However, volumetric methods often suffer from high computational costs and struggle to reconstruct precise geometry. In contrast, surface-based explicit representations enable cleaner separation between geometry and appearance, making material and lighting estimation more tractable. However, surface methods struggle with small or thin surfaces, fuzzy boundaries, and semi-transparent objects. They also require careful handling of topology during optimization, as surface representations lack the flexibility of volumetric approaches to handle topological changes.

4.4.4. *Lack of common representations*

One of the main challenges in inverse rendering research is the lack of common representations for geometry, materials, and lighting. Unlike other computer vision fields that cover the same data formats, such as COCO for object detection and ImageNet for classification, inverse rendering methods produce highly heterogeneous outputs, making them difficult to compare fairly. For example, different methods yield different output shapes (e.g., polygon meshes, SDFs, point clouds, depth maps). For this reason, the geometric quality of the reconstruction cannot be easily compared across different representations, leaving only rendered images as a proxy for evaluating reconstruction quality. Methods such as [MHS*22] aim to bridge this gap by converting implicit representations into explicit meshes, but such conversions introduce artifacts and require careful tuning of extraction parameters.

4.4.5. *Limited evaluation*

A relevant problem that is evident from the papers is that the evaluation of the methods is typically limited to a few datasets and to heterogeneous metrics. The evaluation of single-view methods is more homogeneous, and several papers use, for example, the IIW benchmark to estimate the WHDR metric. However, this metric has recently been criticized [WCS*23], and better alternatives have been suggested.

Evaluation of material reconstruction is sometimes done by directly measuring differences in image-space maps on synthetic data with ground truth. However, it is worth noting that different approaches might reconstruct parameter maps that do not have the same meaning across shading models, making direct comparison unreliable.

The evaluation of re-rendered and relighted images can provide a less biased estimate of the reconstruction's quality, even if it cannot separate the effects of the components. Comparison of relighted images may also be biased by different ways of representing light across methods. In any case, to evaluate inverse rendering based on relighting quality, it would be better to use large datasets that feature different shapes and material characteristics, as well as varied lighting conditions. However, the evaluations in many of the surveyed papers are limited to a few examples.

Only a few methods estimate meshes and quantitatively evaluate the quality of the reconstructed models. In these cases, the comparison is typically limited to a few shapes. Specific benchmarks designed to evaluate methods featuring models with different characteristics, possibly based on real images, are important for supporting further research and properly comparing different techniques.

A potential way to evaluate material reconstruction and test the suitability of inverse rendering methods for applications, not found in the surveyed papers, would be to test material segmentation over complex shapes. Also, in this case, specifically designed benchmarks are needed, and are still not available.

Multi-view methods are often evaluated only for novel view synthesis, using benchmarks derived from those used in the domain of simple light-field encoding (not explicitly reconstructing materials and shape). Many real benchmarks for this purpose feature test views that are not too different from those used to generate the model. For this reason, the reconstruction may not generalize well across different viewing angles. This limitation of many benchmarks used for novel-view synthesis quality assessment is similarly critical for light-field encoding methods that do not disentangle shapes and materials, and should be addressed by creating specifically designed new benchmarks.

4.4.6. *Lack of focus on applications*

Despite the results shown and some examples of potential applications, few papers have actually focused on demonstrating the practical use of the methods across different application contexts, and many potentially practical uses of the proposed inverse rendering frameworks remain uninvestigated. It would be important to test the proposed approaches on user-friendly scene/light editing interfaces to improve interaction quality in virtual/augmented environments or to solve complex material segmentation tasks, for example, in Cultural Heritage or in industrial quality control. This observation is closely related to the previous ones regarding the limited evaluation. A better evaluation of inverse rendering approaches could rely on estimating how well they support applications such as the development of user-friendly image/scene editing tools, 3D scanning, and surface semantic segmentation. End users interested in these applications could provide datasets to train and test the method and develop specific benchmarking protocols.

5. Conclusion

We presented a survey of the recent literature on Inverse Rendering, collecting 84 papers published in top-ranked conferences or

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

journals over the last 6 years, classifying them according to a specific taxonomy, and analyzing research trends and challenges.

Our analysis points out two significant challenges for future research, closely linked. The first is to optimize the promising approaches that are emerging in the literature for the many practical applications that could benefit from inverse rendering techniques. The second is the necessity of creating specific benchmarks, not only by collecting synthetic or real images for testing methods as currently done, but also by developing optimized metrics to provide an indirect estimate of reconstruction quality. Possible ideas include testing novel view synthesis across a wide set of test directions and evaluating material segmentation tasks.

Regarding the methods, an approach that has not been thoroughly investigated is the integration of data-driven priors from foundation models (for monocular depth [YKH*24] or semantic segmentation [KMR*23]) into the inverse rendering framework. This approach is currently widely applied in the shape reconstruction domain, but could be better integrated into a complete inverse rendering method to improve material reconstruction results. Also, in this case, it would be important to collect large sets of annotated (and unannotated) data specific to the end-user application envisioned to fine-tune the model and optimize performance.

Acknowledgements

Shakir Ullah's Ph.D. scholarship is supported by the Italian Ministry of University and Research (MUR) under the Ministerial Decree no. 630 of 24 April 2024, in the framework of the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 3.3, funded by the European Union – NextGenerationEU. Project CUP: B31I24000370004, in collaboration with Visione Emme Srl. This work was partially supported by grant PRD 2026 (Progetto di Ricerca di Dipartimento) of the University of Modena and Reggio Emilia.

Open access publishing facilitated by Universita degli Studi di Verona, as part of the Wiley - CRUI-CARE agreement.

Ethical Statement

This work does not involve human participants, personal data, or animal research, and no ethical approval was therefore required. The authors declare no conflicts of interest. Figures reproduced from prior publications are used in accordance with their respective licenses (Creative Commons). Generative AI tools were used to create elements of the teaser image.

References

[Ado24] Adobe: *Adobe Substance 3D Sampler*, 2024. Version 4.x. URL: https://www.adobe.com/products/substance3d-sampler.html.

[AG15] Ackermann J., Goesele M.: A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision 9*, 3-4 (2015), 149–254.

[Aut18] Autodesk: Production rendering with arnold — hdri lighting workflows. https://docs.arnoldrenderer.com/, 2018.

[BBJ*21] Boss M., Braun R., Jampani V., Barron J. T., Liu C., Lensch H.: Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 12684–12694.

[BBS14] Bell S., Bala K., Snavely N.: Intrinsic images in the wild. *ACM Transactions on Graphics (TOG) 33*, 4 (2014), 1–12.

[BCC25] Bao J., Chen G., Cui S.: PIR: Photometric inverse rendering with shading cues modeling and surface reflectance regularization. In *2025 International Conference on 3D Vision (3DV)* (2025), IEEE, pp. 544–555.

[BEK*22] Boss M., Engelhardt A., Kar A., Li Y., Sun D., Barron J., Lensch H., Jampani V.: SAMURAI: Shape and material from unconstrained real-world arbitrary image collections. *Advances in Neural Information Processing Systems 35* (2022), 26389–26403.

[BJB*21] Boss M., Jampani V., Braun R., Liu C., Barron J., Lensch H.: Neural-PIL: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems 34* (2021), 10691–10704.

[BJK*20] Boss M., Jampani V., Kim K., Lensch H., Kautz J.: Two-shot spatially-varying BRDF and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 3982–3991.

[BM14] Barron J. T., Malik J.: Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 8 (2014), 1670–1687.

[BS12] Burley B., Studios W. D. A.: Physically-based shading at disney. *Acm Siggraph 12* (2012): 1–7.

[BTHR78] Barrow H., Tenenbaum J., Hanson A., Riseman E.: Recovering intrinsic scene characteristics. *Comput. vis. syst2*, *3-26* (1978), 2.

[BXS*20a] Bi S., Xu Z., Sunkavalli K., Hašan M., Hold-Geoffroy Y., Kriegman D., Ramamoorthi R.: Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *European Conference on Computer Vision* (2020), Springer, pp. 294–311.

[BXS*20b] Bi S., Xu Z., Sunkavalli K., Kriegman D., Ramamoorthi R.: Deep 3d capture: Geometry and reflectance from sparse multi-view images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 5960–5969.

[BZZ*24] Bi Z., Zeng Y., Zeng C., Pei F., Feng X., Zhou K., Wu H.: Gs3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers* (2024), pp. 1–12.

[CCB24] Chung H.-G., Choi S., Baek S.-H.: Differentiable point-based inverse rendering. In *Proceedings of the IEEE/CVF*

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conference on Computer Vision and Pattern Recognition (2024), pp. 4399–4409.

[CCB25] Chung H.-G., Choi S., Baek S.-H.: Differentiable inverse rendering with interpretable basis BRDFs. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 475–484.

[CCL*25] Chen X., Chandaka B., Lin C.-H., Zhang Y.-Q., Forsyth D., Zhao H., Wang S.: INVRGB+ l: Inverse rendering of complex scenes with unified color and lidar reflectance modeling. *arXiv preprint arXiv:2507.17613* (2025).

[CJ23] Cao A., Johnson J.: Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 130–141.

[CLP*23] Choi J., Lee S., Park H., Jung S.-W., Kim I.-J., Cho J.: MAIR: multi-view attention inverse rendering with 3d spatially-varying lighting estimation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), IEEE, pp. 8392–8401.

[CLP*25] Choi J., Lee S., Park H., Jung S.-W., Kim I.-J., Cho J.: MAIR++: Improving multi-view attention inverse rendering with implicit lighting representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).

[CLZ25] Chen H., Lin Z., Zhang J.: GI-GS: Global illumination decomposition on gaussian splatting for inverse rendering. In *ICLR* (2025).

[CPY*24] Chen X., Peng S., Yang D., Liu Y., Pan B., Lv C., Zhou X.: Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. In *European Conference on Computer Vision* (2024), Springer, pp. 450–467.

[CT82] Cook R. L., Torrance K. E.: A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG) 1*, 1 (1982), 7–24.

[CXG*25] Chen Z., Xu T., Ge W., Wu L., Yan D., He J., Wang L., Zeng L., Zhang S., Chen Y.-C.: Uni-Renderer: Unifying rendering and inverse rendering via dual stream diffusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 26504–26513.

[DHRK24] Dalal A., Hagen D., Robbersmyr K. G., Knausgård K. M.: Gaussian splatting: 3d reconstruction and novel view synthesis, a review. *IEEE Access* (2024).

[DLDN25] Du B., Li R. B., Du C., Nguyen T.: Glossgau: efficient inverse rendering for glossy surface with anisotropic spherical gaussian. *arXiv preprint arXiv:2502.14129* (2025).

[DLLY22] Deng Y., Li X., Liu S., Yang M.-H.: Physics-based indirect illumination for inverse rendering. *arXiv preprint arXiv:2212.04705* (2022).

[DLZR22] Deng K., Liu A., Zhu J.-Y., Ramanan D.: Depth-supervised NERF: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 12882–12891.

[DM23] Debevec P. E., Malik J.: Recovering high dynamic range radiance maps from photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. Association for Computing Machinery, New York, NY, USA, 2023, pp. 643–652. https://doi.org/10.1145/3588432.3591514.

[DWW*24] Deng X., Wu L., Walter B., Ramamoorthi R., d'Eon E., Marschner S., Weidlich A.: Reconstructing translucent thin objects from photos. In *SIGGRAPH Asia 2024 Conference Papers* (2024), pp. 1–11.

[DWZ*24] Dai Y., Wang Q., Zhu J., Xi D., Huo Y., Qian C., He Y.: Inverse rendering using multi-bounce path tracing and reservoir sampling. *arXiv preprint arXiv:2406.16360* (2024).

[EGH21] Einabadi F., Guillemaut J.-Y., Hilton A.: Deep neural models for illumination estimation and relighting: A survey. In *Computer Graphics Forum* (2021), vol. *40*, Wiley Online Library, pp. 315–331.

[EN24] Enyo Y., Nishino K.: Diffusion reflectance map: Single-image stochastic inverse rendering of illumination and reflectance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 11873–11883.

[ERB*24] Engelhardt A., Raj A., Boss M., Zhang Y., Kar A., Li Y., Sun D., Brualla R. M., Barron J. T., Lensch H., et al.: Shinobi: Shape and illumination using neural object decomposition via brdf optimization in-the-wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 19636–19646.

[FKMW*23] Fridovich-Keil S., Meanti G., Warburg F. R., Recht B., Kanazawa A.: K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 12479–12488.

[FTTS24] Fei F., Tang J., Tan P., Shi B.: Vminer: Versatile multi-view inverse rendering with near-and far-field light sources. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 11800–11809.

[GGL*24] Gao J., Gu C., Lin Y., Li Z., Zhu H., Cao X., Zhang L., Yao Y.: Relightable 3 d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision* (2024), Springer, pp. 73–89.

[GRPCLM22] Garces E., Rodriguez-Pardo C., Casas D., Lopez-Moreno J.: A survey on intrinsic images: Delving deep into lambert and beyond. *International Journal of Computer Vision 130*, 3 (2022), 836–868.

[GWZ*25] Gu C., Wei X., Zeng Z., Yao Y., Zhang L.: Irgs: Inter-reflective gaussian splatting with 2d gaussian ray tracing. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 10943–10952.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[GWZZ25] Gu C., Wei X., Zhang L., Zhu X.: Tensoflow: Tensorial flow-based sampler for inverse rendering. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 495–504.

[HHM22] Hasselgren J., Hofmann N., Munkberg J.: Shape, light, and material decomposition from images using monte carlo rendering and denoising. *Advances in Neural Information Processing Systems 35* (2022), 22856–22869.

[HWH*25] He Z., Wang T., Huang X., Pan X., Liu Z.: Neural lightrig: Unlocking accurate object normal and material estimation with multi-light diffusion. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 26514–26524.

[HYD*25] Huang L., Ye D., Dan J., Tao C., Liu H., Zhou K., Ren B., Li Y., Guo Y., Guo J.: Transparentgs: Fast inverse rendering of transparent objects with gaussians. *ACM Transactions on Graphics (TOG) 44*, 4 (2025), 1–17.

[JDV*14] Jensen R., Dahl A., Vogiatzis G., Tola E., Aanæs H.: Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2014), pp. 406–413.

[JLX*23] Jin H., Liu I., Xu P., Zhang X., Han S., Bi S., Zhou X., Xu Z., Su H.: Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 165–174.

[JSL*25] Jiang K., Sun J.-M., Li Z., Wang D., Li T.-M., Ramamoorthi R.: Differentiable light transport with gaussian surfels via adapted radiosity for efficient relighting and geometry reconstruction. *ACM Transactions on Graphics (TOG) 44*, 6 (2025), 1–25.

[JSRV22] Jakob W., Speierer S., Roussel N., Vicini D.: Dr.jit: A just-in-time compiler for differentiable rendering. *Transactions on Graphics (Proceedings of SIGGRAPH) 41*, 4 (July 2022). https://doi.org/10.1145/3528223.3530099.

[JTL*24] Jiang Y., Tu J., Liu Y., Gao X., Long X., Wang W., Ma Y.: Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 5322–5332.

[Kaj86] Kajiya J. T.: The rendering equation. In *Proceedings of the 13th annual conference on Computer graphics and interactive techniques* (1986), pp. 143–150.

[KE19] Kanamori Y., Endo Y.: Relighting humans: occlusion-aware inverse rendering for full-body human images. *arXiv preprint arXiv:1908.02714* (2019).

[KKLD23] Kerbl B., Kopanas G., Leimkühler T., Drettakis G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph. 42*, 4 (2023), 139–1.

[KKO*23] Kaya B., Kumar S., Oliveira C., Ferrari V., Van Gool L.: Multi-view photometric stereo revisited. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2023), pp. 3126–3135.

[KLK14] Karsch K., Liu C., Kang S. B.: Depth transfer: Depth extraction from video using non-parametric sampling. *IEEE transactions on pattern analysis and machine intelligence 36*, 11 (2014), 2144–2158.

[KMR*23] Kirillov A., Mintun E., Ravi N., Mao H., Rolland C., Gustafson L., Xiao T., Whitehead S., Berg A. C., Lo W.-Y., et al.: Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision* (2023), pp. 4015–4026.

[KPZK17] Knapitsch A., Park J., Zhou Q.-Y., Koltun V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG) 36*, 4 (2017), 1–13.

[KSN24] Kocsis P., Sitzmann V., Nießner M.: Intrinsic image diffusion for indoor single-view material estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 5198–5208.

[KSS02] Kautz J., Snyder J., Sloan P.-P. J.: Fast arbitrary brdf shading for low-frequency lighting using spherical harmonics. *Rendering Techniques 2*, 291-296 (2002), 1.

[KZY*23] Kuang Z., Zhang Y., Yu H.-X., Agarwala S., Wu E., Wu J., et al.: Stanford-orb: A real-world 3d object inverse rendering benchmark. *Advances in Neural Information Processing Systems 36* (2023), 46938–46957.

[LADL18] Li T.-M., Aittala M., Durand F., Lehtinen J.: Differentiable Monte Carlo ray tracing through edge sampling. *ACM Transactions on Graphics (TOG) 37*, 6 (2018), 1–11.

[LB14] Loper M. M., Black M. J.: Opendr: An approximate differentiable renderer. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13* (2014), Springer, pp. 154–169.

[LCF*23] Liu I., Chen L., Fu Z., Wu L., Jin H., Li Z., Wong C. M. R., Xu Y., Ramamoorthi R., Xu Z., et al.: Openillumination: A multi-illumination dataset for inverse rendering evaluation on real objects. *Advances in Neural Information Processing Systems 36* (2023), 36951–36962.

[LCL*23] Liang R., Chen H., Li C., Chen F., Panneer S., Vijaykumar N.: Envidr: Implicit differentiable renderer with neural environment lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 79–89.

[LGL*25] Liang R., Gojcic Z., Ling H., Munkberg J., Hasselgren J., Lin Z.-H., Gao J., Keller A., Vijaykumar N., Fidler S., et al.: DiffusionRenderer: Neural inverse and forward rendering with video diffusion models. *arXiv preprint arXiv:2501.18590* (2025).

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[LHG*25] Lai S., Huang L., Guo J., Cheng K., Pan B., Long X., Lyu J., Lv C., Guo Y.: Glossygs: Inverse rendering of glossy objects with 3d gaussian splatting. *IEEE Transactions on Visualization and Computer Graphics* (2025).

[LHJ19] Loubet G., Holzschuch N., Jakob W.: Reparameterizing discontinuous integrands for differentiable rendering. *ACM Transactions on Graphics (TOG) 38*, 6 (2019), 1–14.

[LHK*20] Laine S., Hellsten J., Karras T., Seol Y., Lehtinen J., Aila T.: Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics 39*, 6 (2020), 1–14.

[LHL*25] Lin C.-H., Huang J.-B., Li Z., Dong Z., Richardt C., Li T., Zollhöfer M., Kopf J., Wang S., Kim C.: Iris: Inverse rendering of indoor scenes from low dynamic range images. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 465–474.

[LKLR24] Luiten J., Kopanas G., Leibe B., Ramanan D.: Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *2024 International Conference on 3D Vision (3DV)* (2024), IEEE, pp. 800–809.

[LLJ*25] Liang Z., Li H., Jia K., Guo K., Zhang Q.: Gus-ir: Gaussian splatting with unified shading for inverse rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).

[LOU*24] Li C., Ono T., Uemori T., Mihara H., Gatto A., Nagahara H., Moriuchi Y.: Neisf: Neural incident stokes field for geometry and material estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 21434–21445.

[LOU*25] Li C., Ono T., Uemori T., Nitta S., Mihara H., Gatto A., Nagahara H., Moriuchi Y.: Neisf++: Neural incident stokes field for polarized inverse rendering of conductors and dielectrics. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 26493–26503.

[LPD*25] Litman Y., Patashnik O., Deng K., Agrawal A., Zawar R., De la Torre F., Tulsiani S.: Materialfusion: Enhancing inverse rendering with material diffusion priors. In *2025 International Conference on 3D Vision (3DV)* (2025), IEEE, pp. 802–812.

[LRR04] Lawrence J., Rusinkiewicz S., Ramamoorthi R.: Efficient brdf importance sampling using a factored representation. *ACM Transactions on Graphics (ToG) 23*, 3 (2004), 496–505.

[LSB*22] Li Z., Shi J., Bi S., Zhu R., Sunkavalli K., Hašan M., Xu Z., Ramamoorthi R., Chandraker M.: Physically-based editing of indoor scene lighting from a single image. In *European Conference on Computer Vision* (2022), Springer, pp. 555–572.

[LSM*18] Li W., Saeedi S., McCormac J., Clark R., Tzoumanikas D., Ye Q., Huang Y., Tang R., Leutenegger S.: Interiornet: Mega-scale multi-sensor photo-realistic indoor scenes dataset. *arXiv preprint arXiv:1809.00716* (2018).

[LSR*20] Li Z., Shafiei M., Ramamoorthi R., Sunkavalli K., Chandraker M.: Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 2475–2484.

[LTH*23] Lyu L., Tewari A., Habermann M., Saito S., Zollhöfer M., Leimkühler T., Theobalt C.: Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Transactions on Graphics (TOG) 42*, 6 (2023), 1–14.

[LWC*23] Li Z., Wang L., Cheng M., Pan C., Yang J.: Multi-view inverse rendering for large-scale real-world indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 12499–12509.

[LWL*23] Liu Y., Wang P., Lin C., Long X., Wang J., Liu L., Komura T., Wang W.: Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. *ACM Transactions on Graphics (ToG) 42*, 4 (2023), 1–22.

[LWZW24] Li J., Wang L., Zhang L., Wang B.: Tensosdf: Roughness-aware tensorial representation for robust geometry and material reconstruction. *ACM Transactions on Graphics (TOG) 43*, 4 (2024), 1–13.

[LYS*21] Li Z., Yu T.-W., Sang S., Wang S., Song M., Liu Y., Yeh Y.-Y., Zhu R., Gundavarapu N., Shi J., et al.: Openrooms: An open framework for photorealistic indoor scene datasets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 7190–7199.

[LYX*24] Ling J., Yu R., Xu F., Du C., Zhao S.: Nerf as a non-distant environment emitter in physics-based inverse rendering. In *ACM SIGGRAPH 2024 Conference Papers* (2024), pp. 1–12.

[LZBD21] Luan F., Zhao S., Bala K., Dong Z.: Unified shape and svbrdf recovery using differentiable monte carlo rendering. In *Computer Graphics Forum* (2021), vol. *40*, Wiley Online Library, pp. 101–113.

[LZF*24] Liang Z., Zhang Q., Feng Y., Shan Y., Jia K.: Gs-ir: 3d gaussian splatting for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 21644–21653.

[MAAD*23] Mirzaei A., Aumentado-Armstrong T., Derpanis K. G., Kelly J., Brubaker M. A., Gilitschenski I., Levinshtein A.: Spin-nerf: Multiview segmentation and perceptual inpainting with neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 20669–20679.

[Mar98] Marschner S. R.: *Inverse rendering for computer graphics*. Cornell University, 1998.

[Mat03] Matusik W.: *A data-driven reflectance model*. PhD thesis, Massachusetts Institute of Technology, 2003.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 8280–8290.

[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. *Communications of the ACM 65*, 1 (2021), 99–106.

[PBGL*22] Praveen Bangaru S., Gharbi M., Li T.-M., Luan F., Sunkavalli K., Hašan M., Bi S., Xu Z., Bernstein G., Durand F.: Differentiable rendering of neural sdfs through reparameterization. *arXiv e-prints* (2022), arXiv–2206.

[PCPMMN21] Pumarola A., Corona E., Pons-Moll G., Moreno-Noguer F.: D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 10318–10327.

[PP03] Patow G., Pueyo X.: A survey of inverse rendering problems. In *Computer graphics forum* (2003), vol. *22*, Wiley Online Library, pp. 663–687.

[PSH*21] Park K., Sinha U., Hedman P., Barron J. T., Bouaziz S., Goldman D. B., Martin-Brualla R., Seitz S. M.: Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228* (2021).

[RBG23] Rainer G., Bridgeman L., Ghosh A.: Neural shading fields for efficient facial inverse rendering. In *Computer Graphics Forum* (2023), vol. *42*, Wiley Online Library, p. e14943.

[RES*22] Rudnev V., Elgharib M., Smith W., Liu L., Golyanik V., Theobalt C.: Nerf for outdoor scene relighting. In *European Conference on Computer Vision* (2022), Springer, pp. 615–631.

[RH01] Ramamoorthi R., Hanrahan P.: A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), pp. 117–128.

[RHS24] Revathy S., Harini A., Sruthika S.: Augmented reality in interior design. *Journal of Innovative Image Processing 6*, 3 (2024), 305–313.

[RRR*21] Roberts M., Ramapuram J., Ranjan A., Kumar A., Bautista M. A., Paczan N., Webb R., Susskind J. M.: Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In *Proceedings of the IEEE/CVF international conference on computer vision* (2021), pp. 10912–10922.

[SC20] Sang S., Chandraker M.: Single-shot neural relighting and svbrdf estimation. In *European Conference on Computer Vision* (2020), Springer, pp. 85–101.

[SCL*23] Sun C., Cai G., Li Z., Yan K., Zhang C., Marshall C., Huang J.-B., Zhao S., Dong Z.: Neural-pbir reconstruction of shape, material, and illumination. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 18046–18056.

[SDZ*21] Srinivasan P. P., Deng B., Zhang X., Tancik M., Mildenhall B., Barron J. T.: NeRV: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 7495–7504.

[SGX*25] Sun H., Gao Y., Xie J., Yang J., Wang B.: SVG-IR: Spatially-varying gaussian splatting for inverse rendering. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 16143–16152.

[SHKF12] Silberman N., Hoiem D., Kohli P., Fergus R.: Indoor segmentation and support inference from RGBD images. In *European conference on computer vision* (2012), Springer, pp. 746–760.

[SNA*21] Sun T., Nam G., Aliaga C., Hery C., Ramamoorthi R.: Human Hair Inverse Rendering using Multi-View Photometric data. In *Eurographics Symposium on Rendering - DL-only Track* (2021), Bousseau A., McGuire M., (Eds.), The Eurographics Association. https://doi.org/10.2312/sr.20211301.

[SWW*25] Shi Y., Wu Y., Wu C., Liu X., Zhao C., Feng H., Zhang J., Zhou B., Ding E., Wang J.: GIR: 3d gaussian inverse rendering for relightable scene factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).

[TCZD25] Tian Z., Cheng S., Zhao J., Duan F.: TIRPL: Tailored-made inverse rendering for point-light scenes. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2025), IEEE, pp. 1–5.

[TTM*22] Tewari A., Thies J., Mildenhall B., Srinivasan P., Tretschk E., Yifan W., Lassner C., Sitzmann V., Martin-Brualla R., Lombardi S., et al.: Advances in neural rendering. In *Computer Graphics Forum* (2022), vol. *41*, Wiley Online Library, pp. 703–735.

[UAS*24] Ummenhofer B., Agrawal S., Sepulveda R., Lao Y., Zhang K., Cheng T., Richter S., Wang S., Ros G.: Objects with lighting: A real-world dataset for evaluating reconstruction and rendering for object relighting. In *2024 International Conference on 3D Vision (3DV)* (2024), IEEE, pp. 137–147.

[UUE25] Ulucan D., Ulucan O., Ebner M.: Challenges and applications of intrinsic image decomposition: A short review. *SN Computer Science 6*, 2 (2025), 1–13.

[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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), IEEE, pp. 5481–5490.

[VSJ22] Vicini D., Speierer S., Jakob W.: Differentiable signed distance function rendering. *ACM Transactions on Graphics (ToG) 41*, 4 (2022), 1–18.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[WBB*25] Wu S., Basu S., Broedermann T., Van Gool L., Sakaridis C.: PBR-NERF: Inverse rendering with physics-based neural fields. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025), pp. 10974–10984.

[WBSS04] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612.

[WCD*20] Wei X., Chen G., Dong Y., Lin S., Tong X.: Object-based illumination estimation with rendering-aware neural networks. In *European Conference on Computer Vision* (2020), Springer, pp. 380–396.

[WCS*23] Wu J., Chowdhury S., Shanmugaraja H., Jacobs D., Sengupta S.: Measured albedo in the wild: Filling the gap in intrinsics evaluation. In *2023 IEEE International Conference on Computational Photography (ICCP)* (2023), IEEE, pp. 1–12.

[WHL*23] Wu H., Hu Z., Li L., Zhang Y., Fan C., Yu X.: Nefii: Inverse rendering for reflectance decomposition with near-field indirect illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 4295–4304.

[Wil78] Williams L.: Casting curved shadows on curved surfaces. In *Proceedings of the 5th annual conference on Computer graphics and interactive techniques* (1978), pp. 270–274.

[WLLL25] Wei X., Liu Z., Li P., Luximon Y.: SIR: Multi-view inverse rendering with decomposable shadow under indoor intense lighting. In *2025 IEEE International Conference on Multimedia and Expo (ICME)* (2025), IEEE, pp. 1–6.

[WMLT07] Walter B., Marschner S. R., Li H., Torrance K. E.: Microfacet models for refraction through rough surfaces. *Rendering techniques 2007* (2007), 18th.

[WPFK21] Wang Z., Philion J., Fidler S., Kautz J.: Learning indoor inverse rendering with 3d spatially-varying lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 12538–12547.

[WRG*09] Wang J., Ren P., Gong M., Snyder J., Guo B.: All-frequency rendering of dynamic, spatially-varying reflectance. In *ACM SIGGRAPH Asia 2009 papers*. 2009, pp. 1–10.

[WRG*25] Weier P., Riviere J., Guseinov R., Garbin S., Slusallek P., Bickel B., Beeler T., Vicini D.: Practical inverse rendering of textured and translucent appearance. *ACM Transactions on Graphics (TOG) 44*, 4 (2025), 1–16.

[WSG*23] Wang Z., Shen T., Gao J., Huang S., Munkberg J., Hasselgren J., Gojcic Z., Chen W., Fidler S.: Neural fields meet explicit geometric representations for inverse rendering of urban scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2023), pp. 8370–8380.

[WTC*25] Wang L., Tran D. M., Cui R., TG T., Dahl A. B., Bigdeli S. A., Frisvad J. R., Chandraker M.: Materialist: Physically based editing using single-image inverse rendering. *arXiv preprint arXiv:2501.03717* (2025).

[WTL*24] Wu S., Tang S., Lu G., Liu J., Pei W.: Univoxel: Fast inverse rendering by unified voxelization of scene representation. In *European Conference on Computer Vision* (2024), Springer, pp. 360–376.

[WYF*24] Wu G., Yi T., Fang J., Xie L., Zhang X., Wei W., Liu W., Tian Q., Wang X.: 4d Gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2024), pp. 20310–20320.

[WZY*23] Wu L., Zhu R., Yaldiz M. B., Zhu Y., Cai H., Matai J., Porikli F., Li T.-M., Chandraker M., Ramamoorthi R.: Factorized inverse path tracing for efficient and accurate material-lighting estimation. In *Proceedings of the IEEE/CVF international conference on computer vision* (2023), pp. 3848–3858.

[XXP*22] Xu Q., Xu Z., Philip J., Bi S., Shu Z., Sunkavalli K., Neumann U.: Point-NeRF: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 5438–5448.

[YGL*25] Ye K., Gao C., Li G., Chen W., Chen B.: GeoSplatting: Towards geometry guided gaussian splatting for physically-based inverse rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2025), pp. 28991–29000.

[YGY*24] Yang Z., Gao X., Yu W., Zhou X., Jin X.: Robir: Robust inverse rendering for high-illumination scenes. *Advances in Neural Information Processing Systems 37* (2024), 124114–124136.

[YKH*24] Yang L., Kang B., Huang Z., Xu X., Feng J., Zhao H.: Depth anything: Unleashing the power of large-scale unlabeled data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2024), pp. 10371–10381.

[YLH*23] Yan K., Luan F., Hašan M., Groueix T., Deschaintre V., Zhao S.: PSDR-Room: Single photo to scene using differentiable rendering. In *SIGGRAPH Asia 2023 conference papers* (2023), pp. 1–11.

[YLHG*22] Yeh Y.-Y., Li Z., Hold-Geoffroy Y., Zhu R., Xu Z., Hašan M., Sunkavalli K., Chandraker M.: Photoscene: Photorealistic material and lighting transfer for indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 18562–18571.

[YLL*20] Yao Y., Luo Z., Li S., Zhang J., Ren Y., Zhou L., Fang T., Quan L.: Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 1790–1799.

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[YLN*24] Yunus R., Lenssen J. E., Niemeyer M., Liao Y., Rupprecht C., Theobalt C., Pons-Moll G., Huang J.-B., Golyanik V., Ilg E.: Recent trends in 3d reconstruction of general non-rigid scenes. In *Computer Graphics Forum* (2024), vol. *43*, Wiley Online Library, p. e15062.

[YN23] Yoshiyama K., Narihira T.: Ndjir: Neural direct and joint inverse rendering for geometry, lights, and materials of real object. *arXiv preprint arXiv:2302.00675* (2023).

[YYC*23] Yu B., Yang S., Cui X., Dong S., Chen B., Shi B.: Milo: Multi-bounce inverse rendering for indoor scene with light-emitting objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 45*, 8 (2023), 10129–10142.

[YZL*22] Yao Y., Zhang J., Liu J., Qu Y., Fang T., McKinnon D., Tsin Y., Quan L.: NeILF: Neural incident light field for physically-based material estimation. In *European conference on computer vision* (2022), Springer, pp. 700–716.

[ZCV*25] Zheng Y., Chai M., Vicini D., Zhou Y., Xu Y., Guibas L., Wetzstein G., Beeler T.: GroomLight: Hybrid inverse rendering for relightable human hair appearance modeling. *arXiv preprint arXiv:2503.10597* (2025).

[ZGW*23] Zhang L., Gao F., Wang L., Yu M., Cheng J., Zhang J.: Deep SVBRDF estimation from single image under learned planar lighting. In *ACM SIGGRAPH 2023 conference proceedings* (2023), pp. 1–11.

[ZHY*23] Zhu J., Huo Y., Ye Q., Luan F., Li J., Xi D., Wang L., Tang R., Hua W., Bao H., et al.: I2-sdf: Intrinsic indoor scene reconstruction and editing via raytracing in neural sdfs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2023), pp. 12489–12498.

[ZIE*18] Zhang R., Isola P., Efros A. A., Shechtman E., Wang O.: The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 586–595.

[ZLH*22] Zhu J., Luan F., Huo Y., Lin Z., Zhong Z., Xi D., Wang R., Bao H., Zheng J., Tang R.: Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *SIGGRAPH Asia 2022 Conference Papers* (2022), pp. 1–8.

[ZLLS22] Zhang K., Luan F., Li Z., Snavely N.: IRON: Inverse rendering by optimizing neural SDFs and materials from photometric images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022), pp. 5565–5574.

[ZLM*22] Zhu R., Li Z., Matai J., Porikli F., Chandraker M.: Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 2822–2831.

[ZLW*21] Zhang K., Luan F., Wang Q., Bala K., Snavely N.: Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 5453–5462.

[ZSD*21] Zhang X., Srinivasan P. P., Deng B., Debevec P., Freeman W. T., Barron J. T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG) 40*, 6 (2021), 1–18.

[ZSG*18] Zollhöfer M., Stotko P., Görlitz A., Theobalt C., Nießner M., Klein R., Kolb A.: State of the art on 3d reconstruction with RGB-D cameras. *Computer Graphics Forum 37* (2018): 625–652.

[ZSH*22] Zhang Y., Sun J., He X., Fu H., Jia R., Zhou X.: Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 18643–18652.

[ZTCS02] Zhang R., Tsai P.-S., Cryer J. E., Shah M.: Shape-from-shading: A survey. *IEEE transactions on pattern analysis and machine intelligence 21*, 8 (2002), 690–706.

[ZXY*23] Zhang Y., Xu T., Yu J., Ye Y., Jing Y., Wang J., Yu J., Yang W.: NEMF: Inverse volume rendering with neural microflake field. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 22919–22929.

[ZYL*23] Zhang J., Yao Y., Li S., Liu J., Fang T., McKinnon D., Tsin Y., Quan L.: NeILF++: Inter-reflectable light fields for geometry and material estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2023), pp. 3601–3610.

[ZYZ21] Zhang C., Yu Z., Zhao S.: Path-space differentiable rendering of participating media. *ACM Transactions on Graphics (TOG) 40*, 4 (2021), 1–15.

[ZZLZ25] Zheng R., Zhang Q., Long C., Zheng W.-S.: Dnf-intrinsic: Deterministic noise-free diffusion for indoor inverse rendering. *arXiv preprint arXiv:2507.03924* (2025).

© 2026 The Author(s). *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.70592 by CochraneItalia, Wiley Online Library on [23/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License