A Survey on 3D Vision Techniques: Inverse Rendering, Relighting, and Material Estimation

www.surveyx.cn

Abstract

This survey paper provides a comprehensive examination of contemporary 3D vision techniques, focusing on inverse rendering, relighting, and material estimation. These methodologies are pivotal in reconstructing and manipulating three-dimensional scenes from two-dimensional images, with applications extending to augmented and virtual reality (AR/VR), gaming, and robotics. Recent advancements in neural rendering have significantly enhanced visual realism and expanded the capabilities of computer vision and graphics. The paper synthesizes methodologies for estimating geometry, reflectance, and illumination from diverse lighting conditions, addressing challenges such as self-shadows and interreflections. It also explores automated methods for material estimation from multi-view captures, improving rendering workflows and providing benchmarks for novel view synthesis. The integration of machine learning with traditional techniques has further advanced the field, with neural approaches offering enhanced realism and accuracy in scene reconstructions. Despite these advancements, challenges remain, particularly in handling complex lighting scenarios and the computational demands of high-quality reconstructions. The survey concludes by highlighting potential future research directions, emphasizing the need for robust and efficient inverse rendering techniques and addressing ethical considerations in visual content manipulation. By continuing to push the boundaries of 3D vision research, the field aims to unlock new possibilities for immersive and interactive digital experiences.

1 Introduction

1.1 Significance of 3D Vision in Computer Vision and Graphics

3D vision is pivotal in computer vision and graphics, enabling the reconstruction and manipulation of three-dimensional structures from two-dimensional images. This capability is essential for accurately recovering scene properties—geometry, reflectance, and illumination—fundamental to advancements in 3D vision technologies [1]. The disentanglement of these properties supports the creation of realistic digital models, particularly in augmented and virtual reality (AR/VR), where reconstructing complex scenes from sparse measurements presents significant computational challenges [2].

The demand for 3D vision techniques is particularly pronounced in AR/VR applications, where novel view synthesis and relighting are crucial for immersive environments [3]. These techniques depend on decoupling illumination components, which enhances visual realism [4]. Recent advances in neural rendering have further revolutionized the field, providing methods that enhance visual realism and expand the capabilities of computer vision and graphics [5].

Reconstructing 3D geometry and reflectance properties from 2D images remains a long-standing challenge with wide-ranging applications in 3D visualization, relighting, and AR/VR [6]. The complexity of these processes necessitates explicit 3D asset reconstruction and costly simulations, highlighting the computational demands involved [7]. Enhancing realism in human relighting scenarios is critical for creating lifelike digital avatars and interactive experiences [8].

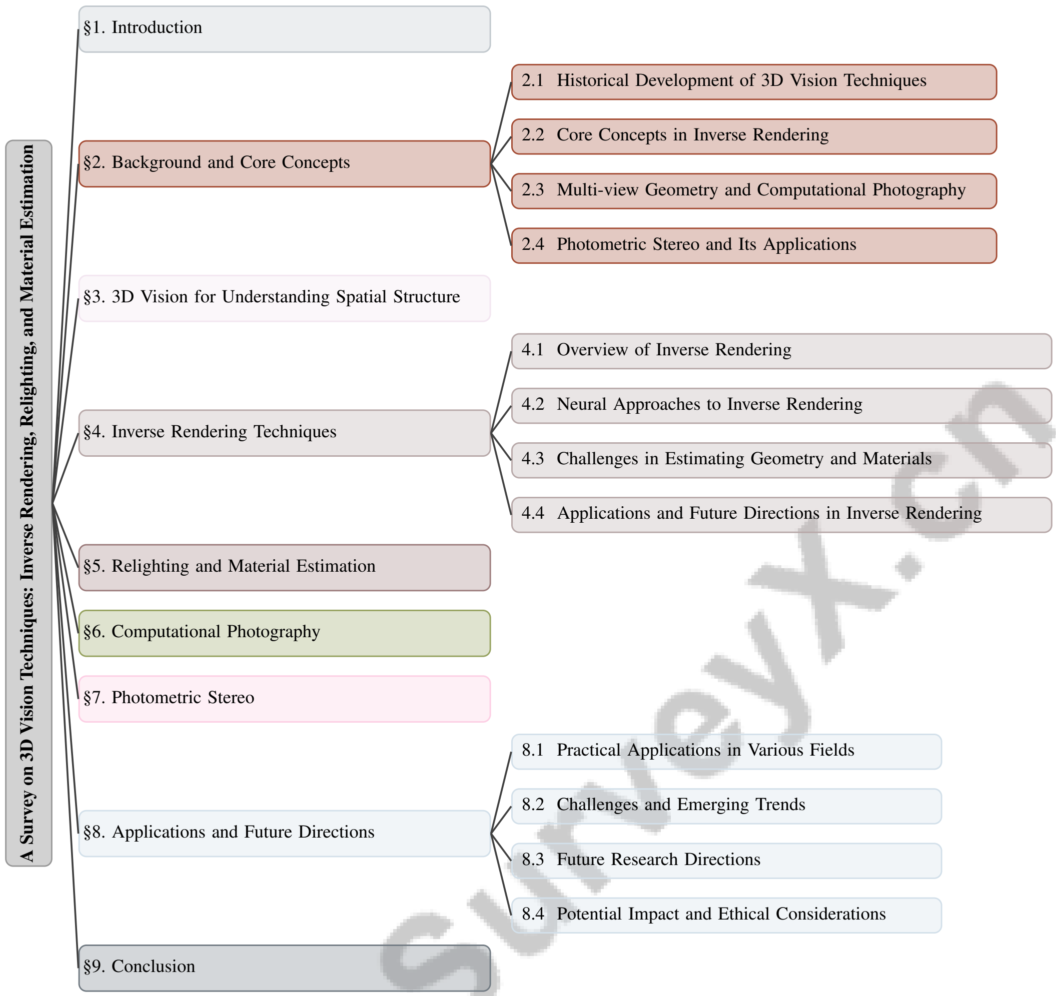

Figure 1: chapter structure

The accurate reconstruction of shape and material properties from 2D images significantly impacts various applications in computer graphics and vision, facilitating the development of detailed and realistic digital models [9]. Addressing challenges such as self-shadows and interreflections, 3D vision techniques are essential for estimating shape, materials, and lighting from images, thereby driving innovation in gaming, robotics, and AR [10]. Through these advancements, 3D vision enhances the creation and manipulation of complex visual environments, solidifying its role as a cornerstone of the field.

1.2 Relevance of Core Keywords

Core keywords in 3D vision—including 3D vision, inverse rendering, relighting, and material estimation—are vital for enhancing the capabilities of computer vision and graphics. These concepts are integral to advancing techniques such as relightable asset creation, where methods like RRM and IllumiNeRF extract physical parameters for accurate scene rendering under diverse lighting conditions. Innovative approaches like Texture-based Lighting (TBL) and advanced material optimization strategies significantly improve rendering efficiency and quality in large-scale indoor environments [11, 12, 13, 14]. These keywords encapsulate the methodologies and challenges in reconstructing and manipulating three-dimensional structures from two-dimensional images. Inverse rendering, for instance, is crucial for deducing physical scene attributes, enhancing the accuracy of 3D shape recovery, especially in scenarios involving complex artifacts and rotationally symmetric objects.

Relighting, a critical technique in digital environments, adjusts lighting conditions to create realistic visual effects, enhancing overall visual realism while addressing challenges such as light source

AI-generated, for reference only.

detection and illumination pattern inference. Recent advancements, including Holo-Relighting and Deep Relighting Networks, have improved the ability to synthesize novel viewpoints and manipulate lighting effects from single images, allowing for complex non-Lambertian lighting effects like specular highlights and shadows [15, 16, 17, 18, 19]. This technique often requires integrating geometric information and depth maps for photorealistic image synthesis, addressing challenges such as occlusion and varying lighting conditions. The manipulation of lighting conditions is further explored through single-image relighting and global illumination decomposition, essential for realistic rendering results.

Material estimation, closely related to inverse rendering, involves determining surface properties of objects, ensuring multi-view consistency and enhancing 3D model realism through intrinsic decomposition and illumination-augmented training. Keywords such as image relighting, illumination adjustment, and material properties are central to understanding the challenges and innovations in 3D vision techniques [20].

Recent advancements in 3D vision techniques underscore their transformative role in computer vision and graphics, particularly through innovations in image restoration, neural rendering, and material extraction. These developments enhance visual quality and improve recognition accuracy in challenging conditions, enabling effective applications in real-world scenarios. For instance, the UG$^2 datasetfacilitatestheevaluationofalgorithmsbridgingcomputationalphotographyandvisualrecognition,whilen$
10, 12, 5]$.Theseelementsarefoundationalforongoingresearchanddevelopment,highlightingtheintegrationofmach$

1.3 Objectives of the Survey Paper

This survey paper aims to comprehensively examine current 3D vision techniques, focusing on inverse rendering, relighting, and material estimation. By synthesizing recent advancements, the survey elucidates methodologies for estimating shape, spatially varying Bidirectional Reflectance Distribution Function (BRDF) material properties, and illumination from images captured under diverse lighting conditions [1]. Addressing the challenges of accurately reconstructing geometry and material properties is crucial for advancing rendering workflows and enhancing the field of 3D vision.

Another objective is to explore automated methods for material estimation from multi-view captures, improving rendering workflows and providing benchmarks for comparing models and approaches in novel view synthesis and relighting [10]. The survey establishes a comprehensive framework for reconstructing geometry, spatially varying surface reflectance, and lighting from a single RGB image of arbitrary indoor scenes, offering holistic solutions for complex environments [4].

Furthermore, the paper emphasizes the necessity of a holistic, data-driven approach to jointly estimate multiple attributes, including physical geometry and material properties from images. The critical role of combining physics-based and learning-based techniques in inverse rendering is examined through advanced methodologies such as inverse transport networks and innovative multi-view inverse rendering approaches. These techniques aim to derive physical scene properties—including shape, material, and illumination—directly from image measurements. By leveraging a differentiable rendering framework, inverse transport networks are trained to predict parameters closely matching ground truth while ensuring accurate image reconstructions when used with physically based graphics renderers. The integration of physics-based models for indirect illumination enhances the accuracy of illumination, geometry, and material predictions, improving the overall effectiveness of the inverse rendering process in applications such as robotics and computer graphics [22, 23, 24].

The survey also addresses the challenge of relighting images under varying illumination conditions from a single input image and develops new frameworks for relighting human figures from a single image while accurately separating albedo, geometry, and lighting. By tackling these objectives, the survey intends to provide a critical resource for researchers and practitioners seeking to advance the field of 3D vision, offering insights into the latest developments and potential future research directions [9].

1.4 Structure of the Survey

The survey is meticulously organized to guide readers through the intricate landscape of 3D vision techniques, starting with a foundational introduction that underscores the significance and relevance of 3D vision, inverse rendering, relighting, and material estimation in computer vision and graphics.

Following the introduction, the paper delves into the historical development and core concepts of 3D vision, providing a comprehensive background necessary for understanding subsequent sections.

The survey explores the role of 3D vision in understanding spatial structures, highlighting techniques for spatial structure recovery and high-resolution 3D reconstructions. It delves into learning-based approaches within multi-view geometry, emphasizing how incorporating machine learning techniques, such as differentiable rendering frameworks and geometry-aware neural networks, enhances traditional methods for multi-view relighting and 3D object prediction [25, 26].

Inverse rendering techniques are scrutinized, offering an overview of methodologies used to deduce scene properties from images. This section discusses neural approaches, challenges in estimating geometry and materials, and applications and future directions. The inclusion of neural network applications in inverse rendering reflects recent advancements in the field [27].

The paper then discusses relighting and material estimation, examining methods for handling illumination changes and integrating physics-based and learning-based methods. The innovative use of a pretrained 3D GAN (EG3D) to reconstruct geometry and appearance, enabling complex non-Lambertian lighting effects, is also explored [28].

Subsequent sections focus on computational photography and photometric stereo, detailing advanced imaging techniques and the integration of these methods with 3D vision. The role of deep learning in enhancing photometric stereo techniques and recent advancements in uncalibrated methods are also covered.

The survey provides an in-depth analysis of practical applications of 3D vision technologies, highlighting specific challenges faced in real-world scenarios, such as medical imaging and visual recognition tasks. It explores emerging trends, including advancements in neural rendering and the integration of complex scene representations, while concluding with insightful recommendations for future research directions and addressing ethical considerations related to these technologies [21, 10, 29, 30]. This structured approach ensures a comprehensive understanding of the field, providing valuable insights into current developments and future possibilities.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Historical Development of 3D Vision Techniques

The evolution of 3D vision techniques encompasses significant theoretical and practical advancements. Initial methods, grounded in geometric principles, utilized projective geometry and photogrammetry for reconstructing 3D structures from 2D images, emphasizing the necessity of precise camera calibration and feature extraction [31]. The advent of neural rendering marked a paradigm shift, integrating deep generative models to enhance visual data representation and animation synthesis [32]. This transition addressed historical challenges in animation realism from near-duplicate photos, expanding 3D vision's application scope [5].

Simultaneously, non-photorealistic rendering methods introduced novel visual data processing strategies, refining traditional global processing through content-aware approaches to enhance context-sensitive visual representations [33]. Advances in projection techniques exploiting signal sparsity have driven innovations in computational methods, improving data processing efficiency and visual signal capture [34]. The creation of comprehensive datasets, such as those featuring omnidirectional images, has been instrumental in training computer vision algorithms, highlighting the historical reliance on empirical data for model validation [35].

The historical trajectory of 3D vision techniques reflects a dynamic synergy between theoretical developments, like rendering algorithms and photometric methods, and practical innovations, including neural rendering and differentiable frameworks. These advancements have significantly enhanced the accuracy and realism of 3D reconstructions from 2D images [21, 36, 26, 29, 25]. By building on foundational geometric methods and embracing cutting-edge neural rendering approaches, the field has progressed substantially, offering new avenues for realistic scene reconstruction and visualization.

2.2 Core Concepts in Inverse Rendering

Inverse rendering is pivotal in computer vision and graphics, deducing intrinsic scene properties—geometry, material characteristics, and lighting—from image data. This technique enables the estimation of spatial configurations, materials, and illumination sources solely from visible image information. Recent advancements in implicit neural representations and differentiable rendering have enhanced this field, allowing simultaneous recovery of geometry and materials from multi-view RGB images under unknown lighting conditions. These techniques efficiently model indirect illumination, traditionally challenging due to the computational demands of recursive path tracing, facilitating applications in robotics, computer graphics, and image editing [37, 24]. This process involves disentangling complex interactions among elements, often represented through Bidirectional Reflectance Distribution Functions (BRDFs), to enable realistic scene reconstructions.

Recent methodologies have improved scene reconstruction accuracy and computational efficiency. The GenLit framework, for instance, employs video diffusion models for single-image relighting, simulating dynamic lighting changes while maintaining static scene structure, offering insights into managing illumination variations and enhancing geometry and material predictions in complex lighting scenarios [38, 37, 16, 22, 24].

The Factored-NeuS method advances the reconstruction of surfaces, materials, and illumination from posed multi-view images, particularly for glossy objects. It employs a three-stage progressive inverse rendering process to reconstruct scene radiance and signed distance functions (SDF), effectively managing specular reflections. This method outperforms existing techniques without additional data, demonstrating its effectiveness in handling complex view-dependent lighting effects and delivering high-quality surface reconstructions [1, 39, 40].

In physically-based rendering, frameworks like MIRReS optimize geometry, material, and lighting using explicit triangle meshes, emphasizing the integration of geometric and photometric information for robust scene understanding, particularly in complex applications like medical endoscopy [41, 42, 6, 29, 25].

Neural rendering techniques, exemplified by Neural Reflectance Decomposition (NeRD), facilitate simultaneous optimization of shape, BRDF, and illumination properties, effectively decomposing a scene into its geometric and material components under varying lighting conditions. NeRD supports fast real-time rendering and high-quality relighting of 3D assets, addressing challenges faced by traditional methods in complex lighting scenarios [1, 43, 44, 39]. The DiPIR method further exemplifies the fusion of personalized diffusion models with rendering pipelines for effective lighting recovery, emphasizing machine learning's role in advancing inverse rendering capabilities.

Core concepts in inverse rendering revolve around balancing geometric precision, material characterization, and lighting estimation. By integrating traditional physically-based models and cutting-edge neural techniques, inverse rendering continues to evolve, offering robust frameworks for realistic and flexible scene reconstruction in computer vision and graphics. The development of methods like 3D vision-language Gaussian splatting, integrating visual and semantic modalities through a novel cross-modal rasterization approach, further illustrates the potential for advancing inverse rendering techniques [9].

2.3 Multi-view Geometry and Computational Photography

Multi-view geometry, a foundational concept in computer vision, focuses on analyzing and reconstructing 3D scenes from multiple 2D images, leveraging geometric relationships between views to infer depth and spatial structure. Computational photography enhances traditional imaging techniques by employing advanced algorithms to process and integrate visual data, resulting in improved image quality, innovative visual effects, and enhanced interpretability for manual analysis and automated visual recognition tasks. Recent developments include sophisticated methods for image restoration, content-aware non-photorealistic rendering, and generating photorealistic videos from near-duplicate photos, leveraging unique data characteristics to create superior visual outputs. Ongoing research aims to bridge computational photography and visual recognition, facilitating the deployment of visual recognition tools in real-world applications [34, 10, 5, 33]. The synergy between multi-view geometry and computational photography has led to significant advancements in capturing and processing visual signals.

A key challenge in multi-view geometry is accurately estimating scene components, such as depth and silhouette information, from multiple images. Techniques proposed by Lin et al. [45] utilize optimization sequences to minimize differences between rendered and observed images, effectively enhancing depth and silhouette estimation. This approach exemplifies the integration of geometric principles with computational methods to improve scene fidelity.

Advanced neural networks have further improved the accuracy of scene component estimation and rendering in multi-view images. Methods developed by Choi et al. [46] leverage multi-view attention mechanisms to process images, accurately estimating scene components and rendering them realistically, highlighting machine learning's role in refining multi-view geometry techniques for more precise and realistic scene reconstructions.

In computational photography, evaluating projection techniques is crucial for capturing high-dimensional visual signals. Pandharkar et al. [34] provide benchmarking for empirically evaluating the effectiveness of projection techniques, comparing progressive and randomized sampling methods, essential for understanding the strengths and limitations of various approaches in visual data capture and processing.

The integration of multi-view geometry with computational photography is exemplified by methods optimizing resource allocation based on data characteristics. Zhu et al. [47] introduce an adaptive mechanism that optimizes resource allocation, demonstrating computational photography's potential to enhance multi-view geometry techniques' efficiency and effectiveness.

Robust bundle adjustment and affine correspondences, as proposed by Eichhardt et al. [48], improve surface point and normal estimation accuracy, underscoring the importance of combining geometric and computational approaches for high-quality scene reconstructions.

The principles of multi-view geometry and computational photography are deeply interconnected, as advancements in computational photography—such as image restoration and enhancement algorithms—significantly improve visual recognition tasks, especially in challenging conditions. This synergy emphasizes the necessity for developing algorithms that simultaneously enhance visual quality and facilitate accurate scene classification, enabling more effective deployment of visual recognition tools in real-world applications. Recent initiatives, such as the UG^2 dataset and geometry-aware neural networks, exemplify ongoing efforts to bridge these fields, presenting numerous opportunities for further innovation and collaboration [48, 25, 10]. By leveraging geometric insights and computational techniques, researchers continue to push the boundaries of 3D scene reconstruction and visualization.

2.4 Photometric Stereo and Its Applications

Photometric stereo is a fundamental technique in computer vision for estimating surface normals by analyzing intensity variations across images captured under different lighting conditions. This method is essential for accurately recovering surface geometry, crucial for applications in 3D modeling, material analysis, and facial recognition. Traditional photometric stereo techniques operate under the assumption of direct illumination, which can lead to inaccuracies in 3D shape reconstruction due to neglecting indirect lighting effects, such as inter-reflections and global illumination. These effects are particularly problematic when surface materials deviate from the ideal Lambertian reflectance model or when light sources are positioned at finite distances, common in real-world scenarios. Thus, achieving reliable photometric stereo results from actual objects remains challenging, necessitating more sophisticated models that account for complex lighting conditions and surface reflectance behaviors [41, 29, 49]. To address these limitations, advanced models have been developed to effectively manage complex lighting interactions.

Recent advancements in photometric stereo have introduced innovative methods to enhance accuracy and applicability. The integration of neural networks has significantly expanded its capabilities. For example, the Universal Photometric Stereo Network (UPSN) tackles the challenge of reconstructing 3D shapes from multiple images under varying light sources without assuming specific physical lighting models [50]. This approach highlights the potential of machine learning to improve the robustness of surface normal estimation across diverse shapes and materials.

In scenarios involving uncalibrated lighting, methods utilizing flash photography to separate contributions of different illuminants have demonstrated enhanced surface normal recovery [6]. This advancement makes photometric stereo more versatile in practical applications.

Methods like 3D Moments, which combine camera and scene motion to create photorealistic space-time videos, illustrate photometric stereo's role in recovering surface normals by integrating temporal dynamics to enhance the detail and accuracy of reconstructed surfaces [5].

Photometric stereo's application in recovering surface normals extends to handling realistic assumptions such as light propagation, attenuation, perspective viewing geometry, and specular reflection [51]. This capability effectively enhances the fidelity of 3D reconstructions in complex environments.

Photometric stereo remains a cornerstone technique in computer vision, facilitating high-fidelity 3D reconstructions by leveraging multiple images of a static scene captured from a single camera position under varying lighting conditions. While traditional approaches predominantly utilize orthographic projection and the Lambertian reflectance model, recent advancements have integrated the complete Blinn-Phong reflectance model with perspective projection, enhancing 3D reconstruction accuracy even in complex real-world scenarios, such as medical endoscopy, where specular highlights are prevalent [42, 29]. By leveraging advanced computational methods and integrating spatial and photometric information, researchers are enhancing the accuracy and applicability of photometric stereo, solidifying its role in recovering surface normals in complex environments.

3 3D Vision for Understanding Spatial Structure

3D vision plays a crucial role in deciphering spatial structures, vital for applications in robotics and virtual reality. This section delves into the integration of multi-view geometry and photometric methods to enhance 3D reconstruction accuracy, addressing challenges in complex visual environments. As illustrated in Figure 2, the hierarchical structure of key concepts in 3D vision is essential for understanding spatial structure. This figure highlights the roles, techniques, and advancements in spatial structure recovery, high-resolution reconstructions, and learning-based methods in multi-view geometry, thereby providing a comprehensive overview of the critical elements that contribute to the field's progress.

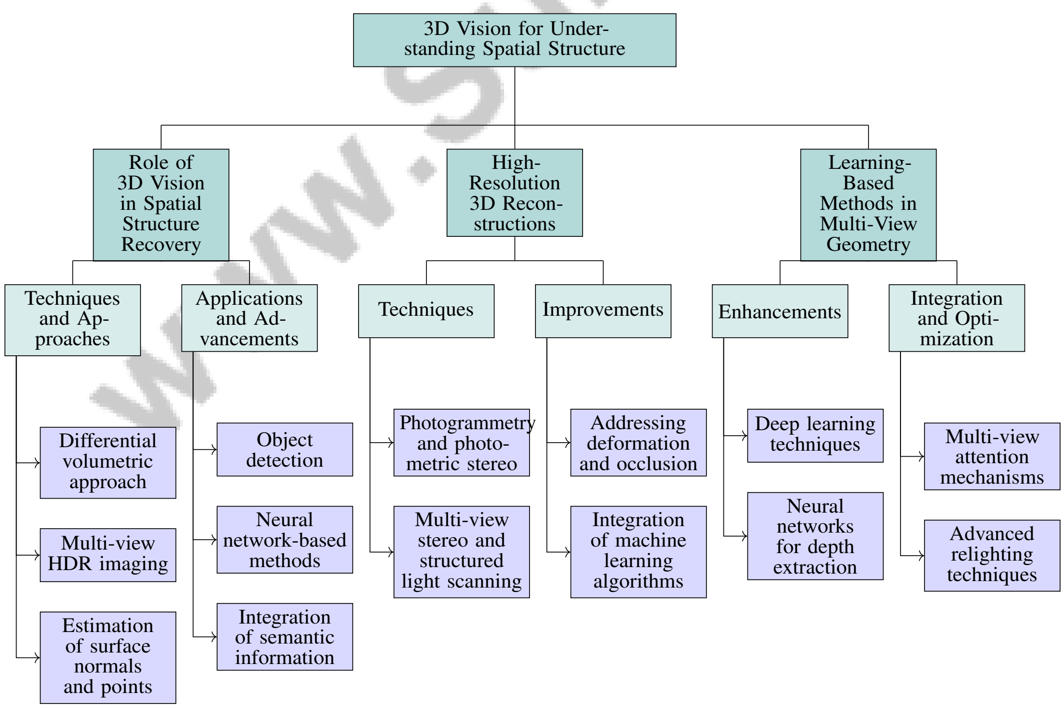

Figure 2: This figure illustrates the hierarchical structure of key concepts in 3D vision for understanding spatial structure, highlighting the roles, techniques, and advancements in spatial structure recovery, high-resolution reconstructions, and learning-based methods in multi-view geometry.

AI-generated, for reference only.

3.1 Role of 3D Vision in Spatial Structure Recovery

Spatial structure recovery in 3D vision involves synthesizing computational techniques for precise reconstructions, especially under challenging lighting and occlusions. The differential volumetric approach exemplifies this by merging multi-view reconstructions with photometric stereo shading, effectively resolving visibility issues [52]. Multi-view high dynamic range (HDR) imaging aids in accurately estimating shadows and materials in complex indoor environments [53]. Estimating surface normals and points is critical for applications like object detection, where precise geometry is essential [48]. Neural network-based methods advance this field by effectively handling near-field images under realistic conditions, including complex material properties [51]. Integrating semantic information with visual data, particularly in handling translucent and reflective objects, enhances 3D scene understanding [9].

3.2 High-Resolution 3D Reconstructions

Achieving high-resolution 3D reconstructions necessitates advanced techniques to capture intricate details and complex geometries. Combining photogrammetry and photometric stereo yields accurate, detailed reconstructions of non-collaborative surfaces [54]. Multi-view reconstruction and photometric stereo address issues like deformation and occlusion, improving modeling accuracy [48, 29]. Techniques like multi-view stereo and structured light scanning enhance resolution and accuracy, while integrating machine learning algorithms improves processing efficiency and reconstruction quality [10, 29].

3.3 Learning-Based Methods in Multi-View Geometry

Learning-based methods have significantly advanced 3D vision by enhancing traditional geometric approaches. Deep learning techniques improve accuracy and efficiency in estimating scene components from multiple viewpoints, addressing challenges of non-Lambertian surfaces and intricate lighting conditions [21, 26, 55, 56]. Neural networks facilitate high-precision extraction of depth and spatial information, with multi-view attention mechanisms refining scene component estimation [46]. Integrating machine learning with geometric principles optimizes resource allocation based on data characteristics, enhancing multi-view geometry efficiency [47]. Advanced relighting techniques leverage deep learning to decompose images into intrinsic components, facilitating realistic object transfers and accurate lighting reflection [15, 57]. These methods enhance 3D reconstruction robustness and accuracy, marking a transformative shift in 3D vision [21, 25, 26, 55].

4 Inverse Rendering Techniques

Category	Feature	Method
Neural Approaches to Inverse Rendering	Lighting Manipulation and Synthesis	GL[58], GSHR[59], SRIE[57]
	Geometry and Surface Analysis	DPSN[60], UNIR[61]
	Perspective Generation	NRTF[62]
Challenges in Estimating Geometry and Materials	Scene Property Deduction	NeRD[1], VHS[8], PX-NET[7], DPIR[2], NF-PS-CNN[51], UniPS-GLC[50], DiPIR[4], 3D-VLGS[9]
Applications and Future Directions in Inverse Rendering	Scene Adaptation Techniques	GaNI[63], PS-UL[64], NIR-UPS[65], NS[66], NIRF[67]
	Lighting and Reflection Control	HR3D[68], RISE-SDF[69], HR[28], GI-GS[70], F-NuS[40]
	Efficiency and Performance	ASSM[71], PRIR[72]
	Material and Lighting Representation	DIR-IBR[73]

Table 1: Table ef provides a comprehensive summary of various methods and approaches in the field of inverse rendering, categorized into neural approaches, challenges in geometry and materials estimation, and applications and future directions. Each category is further detailed with specific features and corresponding methods, highlighting recent advancements and ongoing research in the domain.

Advancements in computer vision and graphics underscore the importance of inverse rendering techniques, which aim to recover intrinsic scene properties from image data. Table 1 presents a categorized overview of current methods in inverse rendering, illustrating the significant advancements and challenges in the field, and outlining potential future directions for research and application. Additionally, Table 5 offers a structured comparison of different inverse rendering techniques,

AI-generated, for reference only.

elucidating their optimization methods, inherent challenges, and practical applications in the context of advancing computer vision and graphics. This section delves into the core principles, methodologies, and challenges of inverse rendering, with a detailed examination in the subsequent subsection titled *Overview of Inverse Rendering*.

4.1 Overview of Inverse Rendering

Inverse rendering is crucial for extracting intrinsic properties such as geometry, material characteristics, and lighting from images, essential for realistic digital representations in virtual and augmented reality, as well as photorealistic rendering. Traditional methods face challenges from complex lighting scenarios and material interactions, necessitating innovative approaches for accurate scene attribute recovery [62]. Differentiable rendering has emerged as a powerful technique, enabling high-quality 3D scene reconstruction by decomposing them into shape, materials, and lighting [1]. Techniques like DiPIR recover scene lighting and tone-mapping parameters from a single image to facilitate photorealistic virtual object insertion [4].

The integration of neural networks with traditional rendering techniques has significantly advanced the field. For instance, NeRD decomposes images into geometry, reflectance, and illumination components, allowing for fast real-time rendering with novel illuminations [1]. Moreover, enhancing representation learning of language modalities while maintaining effective visual representation is crucial for addressing scene reconstruction complexities [9]. Inverse rendering also addresses the challenges of slow and ill-posed high-quality relighting, particularly in modeling complex light transport. The UG2 benchmark introduces a psychophysics-based evaluation regime for assessing image quality improvements and recognition performance, providing insights into current methods [10].

The ongoing evolution of inverse rendering through machine learning, differentiable rendering, and advanced computational models is vital for achieving high-fidelity scene reconstructions and realistic visualizations. Integrating image restoration algorithms with visual recognition tasks can enhance imagery quality captured under challenging conditions, facilitating more accurate automatic object recognition and novel viewpoint synthesis in both static and dynamic scenes [21, 10]. The integration of structured data, scalable frameworks, and advanced rendering techniques ensures that inverse rendering remains a critical area of research and application in computer vision and graphics.

4.2 Neural Approaches to Inverse Rendering

Method Name	Methodology Type	Application Focus	Innovative Features
DPSN[60]	Deep Learning	Surface Normal Estimation	Learning Diverse Brdfs
UNIR[61]	Neural Networks	Surface Normal Estimation	Modeling OF Interreflections
GL[58]	Video Diffusion Models	Single-image Relighting	Fine-grained Control
GSHR[59]	Neural Rendering	Human Relighting	Geometry-aware Framework
SRIE[57]	Encoder-decoder Architecture	Image Relighting	Automatic Illumination Extraction
NF-PS-CNN[51]	Convolutional Neural Networks	3D Shape Reconstruction	Adapting Cnn Techniques
PX-NET[7]	Convolutional Neural Network	Surface Normal Estimation	Per-pixel Observation Map
UniPS-GLC[50]	Data-driven Method	Surface Normal Estimation	Global Lighting Contexts
NRTF[62]	Neural Network-based	Relighting And Synthesis	Complex Indirect Illumination

Table 2: Overview of various neural approaches to inverse rendering, detailing the methodology type, application focus, and innovative features of each method. The table highlights the diversity and specificity of techniques used in enhancing surface normal estimation, relighting, and 3D shape reconstruction through neural networks.

Neural approaches have revolutionized inverse rendering by employing deep learning to enhance realism, efficiency, and accuracy in scene reconstructions. Table 2 presents a comprehensive comparison of neural approaches to inverse rendering, showcasing the diverse methodologies and their specific applications in improving scene reconstruction and relighting. Neural Radiance Fields (NeRF) exemplify this shift, synthesizing novel views from 3D representations using volume rendering to capture intricate lighting interactions and provide high-quality reconstructions [21]. The Deep Photometric Stereo Network (DPSN) utilizes deep learning to infer surface normals from reflectance observations, showcasing neural networks' potential for improving surface normal estimation accuracy [60]. In uncalibrated light conditions, neural networks estimate light directions and optimize image reconstruction losses, enhancing predictions of surface normals and scene parameters [61].

AI-generated, for reference only.

Innovative methods like GenLit reformulate single-image relighting using a point light source with fine-grained control over position and intensity, leveraging the physical understanding capabilities of video foundation models [58]. Geometry-aware techniques improve albedo estimation and shadow realism in relit images, demonstrating neural networks' effectiveness in refining scene attributes [59]. Additionally, neural networks separate scene geometry and lighting elements for relighting tasks, emphasizing their role in disentangling complex interactions [57].

Near-Field Photometric Stereo with CNN (NF-PS-CNN) employs a CNN to predict surface normals from reflectance samples while iteratively estimating depth, compensating for light attenuation [51]. PX-NET predicts surface normals from pixel-wise observation maps, capturing the effects of varying illumination [7]. The Universal Photometric Stereo Network (UPSN) eliminates the need for physical lighting parameter recovery, effectively handling complex spatially-varying lighting effects [50]. Moreover, neural precomputed radiance transfer (PRT) functions enable effective scene relighting and novel view synthesis, showcasing the versatility of neural networks in inverse rendering [62].

Neural approaches continue to push the boundaries of computer vision and graphics. By integrating deep learning with traditional rendering methods, these advanced techniques significantly enhance the extraction of intrinsic properties such as reflectance and shading from complex environments, facilitating accurate modeling of real-world scenes and innovative applications like novel viewpoint synthesis and image editing [21, 74, 44].

4.3 Challenges in Estimating Geometry and Materials

Method Name	Inverse Rendering Challenges	Computational Complexity	Lighting and Material Representation
NeRD[1]	Inherent Ambiguity Present	Multiple Sampling Need	Lack Photometric Calibration
DiPIR[4]	Ill-posed Nature	Multiple Sampling Need	Spherical Gaussian Limitations
DPIR[2]	Ambiguities IN Brdf	Excessive Computational Costs	Lacks Photometric Calibration
VHS[8]	Photometric Calibration	-	Photometric Calibration
PX-NET[7]	Global Illumination Effects	Computational Inefficiency Rendering	Global Illumination Effects
NF-PS-CNN[51]	Non-linear Attenuation	High Computational Cost	Complex Material Properties
UniPS-GLC[50]	Unknown Lighting Conditions	High Computational Cost	Arbitrary Lighting Variations
3D-VLGS[9]	-	-	Varying Lighting Conditions

Table 3: This table presents a comparative analysis of various inverse rendering methods, highlighting their challenges, computational complexity, and approaches to lighting and material representation. The methods evaluated include NeRD, DiPIR, DPIR, VHS, PX-NET, NF-PS-CNN, UniPS-GLC, and 3D-VLGS, each exhibiting unique limitations such as inherent ambiguity, computational inefficiency, and photometric calibration issues. The table serves as a comprehensive reference for understanding the current limitations and computational demands in the field of inverse rendering.

Estimating geometry and materials from images presents numerous challenges, primarily due to the complex interplay between surface characteristics and lighting conditions. A key difficulty lies in the inherent ambiguity of inverse rendering, where estimating shape, illumination, and reflectance is particularly challenging under variable lighting [1]. This ambiguity is exacerbated by the ill-posed nature of inverse rendering, with methods often struggling due to limited input data, leading to inaccuracies in lighting effects and material representation [4]. Table 3 provides a detailed comparison of different inverse rendering methods, focusing on their specific challenges, computational complexity, and techniques for representing lighting and materials.

The computational cost of volumetric rendering complicates reconstruction, as multiple sampling for each ray is necessary to accurately deduce Bidirectional Reflectance Distribution Functions (BRDFs) and surface normals. Limited light-view angular samples hinder the capture of detailed reflectance properties [2]. Additionally, separating illuminants from single photographs poses significant challenges, particularly when target illumination must be extracted from a single image.

Moreover, the lack of photometric calibration in existing methods often results in unrealistic renderings, especially in indoor scenes where modifications can lead to discrepancies in lighting and material representation [8]. The inefficiency of rendering full images with realistic global illumination effects limits the diversity and quantity of training data for developing robust inverse rendering techniques [7].

Handling non-linear light attenuation, specular reflections, and global illumination effects further complicates existing methods, which often lack the capacity to address these complexities effectively

[51]. Many photometric stereo algorithms struggle without prior knowledge of lighting conditions, complicating accurate surface normal estimation [50].

The naive application of color rasterization functions to semantic representations presents a core obstacle, necessitating innovative approaches that integrate visual and semantic modalities, as demonstrated by advancements in 3D vision techniques [9].

Overcoming these challenges is essential for advancing computer vision and graphics. By integrating advanced reflectance models like the Blinn–Phong model and employing perspective projection, researchers can significantly improve the accuracy and realism of 3D models, enhancing fidelity and expanding applicability in fields such as medical imaging, augmented reality, and visual recognition systems [21, 10, 29, 55].

4.4 Applications and Future Directions in Inverse Rendering

Method Name	Application Areas	Technological Advancements	Future Directions
RISE-SDF[69]	Virtual Reality	Rise-SDF	Neural Rendering
HR[28]	Portrait Photography	3D Gan	Real-time Applications
NS[66]	Outdoor Scenes	Neusky	Reflective Surfaces
PS-UL[64]	Virtual Reality	Rise-SDF	Neural Rendering
GaNI[63]	Inverse Rendering Techniques	Gani	Full Room-scale
HR3D[68]	Security Document Verification	Recurrent Optimization Network	Optimization Techniques
PRIR[72]	Various Domains	Path Recycling	Path Sorting Algorithms
GI-GS[70]	Virtual Reality	Gi-GS	Specular Lighting
DIR-IBR[73]	Scene Editing	Dynamic Adjustment	Global Illumination Effects
NIR-UPS[65]	Virtual Reality	Rise-SDF	Neural Rendering
ASSM[71]	Photorealistic Content Creation	Rise-SDF	Neural Rendering
F-NuS[40]	Real-world Applications	Factored-NeuS	Dynamic Objects
NIRF[67]	Virtual Reality	Differentiable Rendering Layer	Reflective Surfaces

Table 4: This table presents a comparative analysis of various inverse rendering methods, highlighting their respective application areas, technological advancements, and future research directions. The methods detailed include RISE-SDF, HR, NS, and others, illustrating their roles in fields such as virtual reality, portrait photography, and scene editing. Key technological innovations and anticipated future developments are also summarized, providing insights into the evolving landscape of inverse rendering techniques.

Inverse rendering techniques have transformed computer vision and graphics, enabling realistic reconstruction and rendering of three-dimensional scenes from two-dimensional images. These techniques are crucial for applications in virtual reality, augmented reality, and photorealistic content creation, where accurate material and lighting depiction is essential. Recent advancements, such as the RISE-SDF method, demonstrate state-of-the-art performance in inverse rendering for glossy objects, achieving high-quality geometry, material reconstruction, and effective relighting [69].

The integration of methods like Holo-Relighting, which synthesizes high-quality relit portraits with control over lighting, head pose, and viewpoint, exemplifies the capability to achieve state-of-the-art results in photorealism and view consistency [28]. NeuSky's novel approach to outdoor scene inverse rendering effectively incorporates sky pixel observations and a differentiable visibility model, showcasing potential advancements in handling outdoor and complex lighting scenarios [66].

Methods that do not require prior knowledge of light source positions enhance inverse rendering practicality in real-world scenarios, particularly for dynamic lighting conditions in live events and interactive media [64]. The GaNI framework shows significant improvements in geometry reconstruction and reflectance recovery over existing techniques [63].

Emerging methods, such as those by Khanian et al., demonstrate significant improvements in 3D shape reconstruction accuracy and detail using consumer-level equipment, paving the way for accessible solutions in 3D modeling [68]. The PRIR method accelerates inverse rendering tasks, achieving faster convergence and reduced computational demands compared to existing methods [72].

Future research directions include extending current methods to encompass specular lighting within frameworks like GI-GS, exploring additional geometric constraints for improved accuracy [70], and enhancing robustness to diverse lighting conditions and global illumination effects [73]. The exploration of self-calibrating photometric stereo methods that resolve generalized bas-relief ambiguity indicates promising avenues for improving lighting estimation and shape recovery in challenging

AI-generated, for reference only.

datasets [65]. The development of adaptive screen-space meshing approaches offers significant speed-up in normal integration, despite challenges with depth discontinuities [71].

Experiments with methods like Factored-NeuS demonstrate improved reconstruction of surfaces, materials, and lighting for glossy objects, indicating promising research directions [40]. The OpenIllumination dataset provides a valuable resource for quantitatively evaluating inverse rendering and material decomposition techniques, highlighting the performance of various state-of-the-art methods [75]. Future research could also enhance methods for reflective surfaces and improve consistency in novel views captured by moving flashlights [67].

The future of inverse rendering research is poised to explore the integration of advanced neural rendering techniques with traditional methods, aiming to overcome limitations in handling diffuse materials and unknown geometry. Anticipated growth into multi-modal and sensor-based domains, along with the creation of extensive datasets, is expected to catalyze significant advancements in inverse rendering technology. These developments will enhance realism and applicability across various fields, evidenced by innovations like Texture-based Lighting (TBL) for large-scale indoor scenes, hybrid differentiable rendering methods, and algorithms for physically grounded image editing. Such innovations facilitate accurate material editing and relighting in mixed-reality applications and improve scene reconstruction efficiency from multi-view images, expanding inverse rendering potential in robotics and computer graphics [36, 13, 24]. Table 4 provides a comprehensive overview of current inverse rendering methods, their application areas, and potential future directions, serving as a critical reference for understanding advancements and trends in this domain.

Feature	Overview of Inverse Rendering	Neural Approaches to Inverse Rendering	Challenges in Estimating Geometry and Materials
Optimization Method	Differentiable Rendering	Deep Learning	Volumetric Rendering
Challenges	Complex Lighting Scenarios	Ill-posed Conditions	Ambiguity, Limited Data
Applications	Virtual/augmented Reality	Scene Reconstruction	Material Representation

Table 5: This table provides a comparative analysis of various inverse rendering methods, focusing on their optimization techniques, challenges, and applications. It highlights the distinctions between traditional differentiable rendering, neural approaches leveraging deep learning, and volumetric rendering, emphasizing their respective advantages and limitations in handling complex lighting, ill-posed conditions, and data scarcity. This comparison serves as a foundation for understanding the current landscape and future research directions in inverse rendering.

5 Relighting and Material Estimation

The manipulation of lighting conditions and estimation of material properties are crucial for achieving photorealistic rendering in computer vision and graphics. This section explores the challenges of illumination changes and the methodologies developed to tackle these issues, focusing on the integration of traditional physics-based techniques with innovative learning-based methods to enhance relighting capabilities.

5.1 Handling Illumination and Relighting

Addressing illumination changes and achieving realistic relighting are critical challenges in computer vision and graphics. Recent advancements have introduced methods that allow substantial lighting alterations without relying on geometric information, enabling flexible relighting solutions where geometric data is scarce or unavailable [76]. Separating albedo from lighting information is vital for accurate rendering, yet traditional methods often struggle with this task, resulting in inaccuracies. Recent techniques have improved the ability to disentangle these components, enhancing the fidelity of relighting outcomes [77]. In scenes with multiple light sources, effective separation of illuminants is crucial for precise control over lighting conditions [6].

Innovative approaches such as RRM (Relightable assets using Radiance guided Material extraction) and perceptually-inspired shading models have demonstrated significant progress by extracting physically-based parameters and accommodating complex lighting scenarios, including highly reflective surfaces. RRM excels in parameter retrieval for high-fidelity relighting and novel view synthesis, while the object relighting system facilitates seamless integration of inserted objects into target scenes, enhancing compositional flexibility beyond traditional image-based methods [12, 16].

AI-generated, for reference only.

5.2 Integration of Physics-Based and Learning-Based Methods

The integration of physics-based and learning-based methods has advanced relighting in computer vision, enhancing rendering realism and adaptability. Physics-based approaches provide a solid framework for modeling light interactions with surfaces, enabling precise albedo and reflectance property modeling. The ReCap method exemplifies this by optimizing multiple lighting representations that share common material attributes, effectively addressing the albedo-lighting ambiguity [77]. Learning-based methods, such as conditional GAN frameworks, allow direct image relighting without geometric inputs, enabling significant lighting alterations while preserving scene integrity [76]. Datasets like OpenIllumination support this integration by offering benchmarks for evaluating inverse rendering algorithms and material decomposition methods under varied illumination conditions [75].

The synergy between physics-based and learning-based methods is exemplified by convolutional neural networks (CNNs), which effectively manage complex lighting scenarios. This two-stage process involves de-lighting to recover intrinsic reflectance, geometry, and lighting properties, followed by relighting to match target illumination. Recent advancements, including end-to-end deep learning architectures that account for non-diffuse effects, have significantly improved relighting accuracy and adaptability, setting the stage for further innovations in rendering and visualization within Augmented Reality (AR) [15, 12].

5.3 Applications in Augmented and Virtual Reality

The integration of relighting and material estimation techniques in augmented reality (AR) and virtual reality (VR) has markedly enhanced the realism and interactivity of these environments. Recent advancements have led to sophisticated AR systems capable of dynamically adjusting lighting conditions and material properties to mimic real-world scenarios. Techniques such as neural radiance fields (NeRFs) facilitate realistic relighting and object insertion, allowing virtual objects to blend seamlessly into their surroundings by accurately reflecting actual lighting. Methods incorporating physics-based image formation models enable precise face relighting, while new scene editing approaches from single images capture complex effects like soft shadows and interreflections, significantly improving user engagement in AR experiences [15, 12, 47, 78].

Key applications of relighting in AR and VR include generating photorealistic renderings under varied lighting and viewpoint conditions. Techniques such as Neural Light Transport (NLT) have shown substantial improvements in relighting and view synthesis, essential for creating realistic virtual environments that mimic real-world physics and lighting [79]. The NeRF-OSR method further expands AR and VR capabilities by enabling high-quality, semantically meaningful editing of scene illumination and camera viewpoints in outdoor settings [80]. Material estimation is also crucial for enhancing realism in AR and VR. The ReCap method's integration of multiple lighting representations around shared material attributes enables robust material estimation and physically sound lighting reconstruction [77]. The FEGR framework combines neural fields with explicit meshes to support high-quality inverse rendering of urban scenes, facilitating relighting and virtual object insertion [81].

Moreover, the Deep Relighting Network (DRN) allows for flexible manipulation of light direction and color temperature without geometric information, enhancing visual quality in AR and VR applications [19]. This flexibility enriches the immersive experience by customizing lighting effects to suit user preferences or environmental contexts. Continued advancements in relighting and material estimation techniques drive innovation in AR and VR, enabling applications such as interactive relighting and drone video relighting [25]. Future directions may explore relighting and albedo editing, underscoring the relevance of these techniques in creating more realistic and engaging AR and VR experiences [82].

6 Computational Photography

6.1 Advanced Imaging Techniques

Advanced imaging techniques in computational photography have transformed visual data capture and processing, enhancing image quality and enabling novel visual effects. These techniques leverage computational algorithms to overcome traditional photography limitations, facilitating innovative image manipulation and enhancement. Methods exploiting signal sparsity for efficient data capture, such as progressive and random projections, underscore the importance of projection techniques in the field [34].

The integration of machine learning algorithms with traditional imaging methods has further propelled advancements, enhancing visual data analysis and refinement. Neural networks improve scene component estimation and rendering accuracy, demonstrating the potential of learning-based methods to elevate conventional imaging techniques [46]. This synergy fosters innovative imaging solutions that enhance the realism of captured images, particularly valuable for applications in augmented reality (AR) and virtual reality (VR).

Moreover, combining structured light scanning and multi-view stereo techniques has significantly enhanced the resolution and accuracy of 3D reconstructions. By integrating photometric stereo methods, which excel in capturing high-frequency surface details, with multi-view stereo approaches ensuring global consistency, these techniques effectively model complex materials and improve the fidelity of 3D representations in challenging real-world conditions [83, 84, 29, 55]. The fusion of complementary data from these methods enhances the robustness and detail of reconstructed models, while their integration with machine learning algorithms promises improvements in processing efficiency and quality.

The evolution of advanced imaging techniques in computational photography continues to expand the boundaries of image capture and processing capabilities. By merging computational algorithms with machine learning and traditional methods, researchers significantly enhance visual quality and facilitate automatic visual recognition, addressing critical challenges in real-world applications. The introduction of datasets like UG$^2 andthedevelopmentofinnovativealgorithmsforimagerestorationandenhancementarepavingthewayforbreakth$ 29, 33].

6.2 Projection Techniques and Evaluation

Benchmark	Size	Domain	Task Format	Metric
PRP[34]	1,000,000	Image Compression	Reconstruction Quality Assessment	SNR, Compression Factor
OWL[85]	72	Computer Vision	Relighting	PSNR, SSIM
NeRF-OSR[80]	3,240	Outdoor Scene Relighting	Relighting	PSNR, SSIM
OWL[86]	72	Inverse Rendering	Relighting	PSNR, SSIM
ReNe[87]	40,000	Relighting	Novel View Synthesis	PSNR, SSIM
OmniSCV[35]	1,000,000	Computer Vision	Image Generation	IoU, Acc
OpenIllumination[75]	108,000	Inverse Rendering	Material Decomposition	PSNR
UG2[10]	3,535,382	Object Recognition	Image Restoration And Enhancement	M1, M2

Table 6: This table provides a comprehensive summary of various benchmarks utilized in the evaluation of projection techniques within computational photography. It includes details on the size, domain, task format, and metrics used for assessment, highlighting the diversity and scope of data sets applied in this field.

Projection techniques are fundamental to computational photography, profoundly impacting the acquisition and processing of high-dimensional visual signals, including videos, multi-spectral data, and lightfields. These techniques leverage the sparsity of underlying signals in transformed domains to optimize measurement efficiency and enhance reconstruction quality. Recent advancements encompass both progressive methods, which iteratively capture detail using simple projection bases like DCT or wavelets, and randomized projection methods employing L0 minimization for reconstruction. The integration of these projection techniques with enhanced algorithms for image restoration and recognition is pivotal for improving visual quality and automatic object classification in challenging conditions, bridging the gap between computational photography and visual recognition applications [34, 10, 29, 88].

A key challenge in projection techniques lies in balancing computational efficiency and image quality. Empirical evaluations comparing progressive and random projections provide insights into their effectiveness in capturing high-dimensional visual signals, highlighting the strengths and limitations of various methods [34]. Understanding these dynamics is crucial for optimizing projection techniques to enhance the realism of captured images. Table 6 presents a detailed overview of representative benchmarks that are instrumental in evaluating projection techniques within the domain of computational photography.

The integration of advanced computational methods, such as machine learning algorithms, with traditional projection techniques has further augmented their capabilities. By employing neural networks, researchers improve scene component estimation and rendering accuracy, refining projected image quality [46]. This integration exemplifies the potential of combining learning-based approaches with traditional methods for achieving precise and realistic image reconstructions.

Additionally, the development of adaptive projection techniques that optimize resource allocation based on data characteristics exemplifies advancements in this field. Such techniques dynamically allocate computational resources in accordance with scene complexity, enhancing the efficiency and effectiveness of projection methods [47]. This adaptive approach is particularly advantageous in scenarios involving complex lighting conditions and occlusions, where traditional methods may struggle to maintain image quality.

The evaluation of projection techniques in computational photography is a dynamic and evolving field, driven by the need for efficient capture and processing of high-quality visual data. By continuously refining projection techniques and integrating advanced computational methods, researchers significantly enhance computational photography capabilities. This progress not only improves visual quality and recognition accuracy in challenging conditions but also facilitates the development of innovative algorithms tailored for aesthetic enhancement and automatic visual recognition tasks, enabling broader applications in medical imaging and visual recognition [34, 10, 29, 33].

6.3 Integration with 3D Vision

The integration of computational photography with 3D vision techniques has significantly enhanced image capture and processing capabilities, enabling more detailed and accurate scene reconstructions. This synergy leverages the complementary strengths of computer graphics and machine learning to improve the fidelity and realism of visual data, creating new opportunities for advancements in both domains. By merging traditional rendering techniques with neural scene representations and developing algorithms that enhance image quality while supporting visual recognition tasks, this collaboration facilitates effective applications in real-world scenarios, such as novel viewpoint synthesis and automated object recognition under challenging conditions [21, 10].

A critical area where computational photography complements 3D vision is through the application of advanced imaging techniques for capturing high-dimensional visual signals. Techniques such as structured light scanning and multi-view stereo provide complementary data that enhance the robustness and detail of 3D reconstructions. These methods are particularly effective when combined with machine learning algorithms, which improve processing efficiency and reconstruction quality by effectively managing complex surfaces and materials with varying reflectance properties [46].

Furthermore, the integration of projection techniques in computational photography with 3D vision exemplifies this synergy. Empirical evaluations of projection methods, such as progressive versus random projections, have yielded insights into optimizing these techniques for enhanced image quality [34]. Refining these methods is crucial for achieving precise and realistic image reconstructions, particularly for applications in augmented reality (AR) and virtual reality (VR).

Moreover, the development of adaptive projection techniques that optimize resource allocation based on data characteristics underscores advancements in the integration of computational photography with 3D vision. These techniques dynamically allocate computational resources according to scene complexity, improving the efficiency and effectiveness of both fields [47]. This adaptive approach proves particularly beneficial in scenarios with complex lighting conditions and occlusions, where traditional methods may falter in maintaining image quality.

The integration of computational photography with 3D vision is advancing image capture and processing capabilities by enabling innovative algorithms that enhance visual quality and improve automatic visual recognition. Recent developments, including the introduction of the UG^2 dataset and novel evaluation metrics, reveal the potential of these technologies to address challenges in image restoration and enhancement, especially under suboptimal conditions. Additionally, new methods such as content-aware non-photorealistic rendering and the creation of 3D Moments illustrate how manipulating image features and interpolating scene motion can yield impressive results in visual representation, expanding possibilities for both artistic rendering and practical applications [10, 5, 33]. By leveraging the strengths of both fields and integrating advanced computational methods, researchers are enhancing the capabilities of computer vision and graphics, paving the way for further innovations and applications in the field.

7 Photometric Stereo

Recent advancements in photometric stereo have markedly improved the accuracy and robustness of surface normal estimation. A significant breakthrough is the incorporation of deep learning, which has redefined traditional methodologies and addressed numerous challenges in the field. The following subsections delve into the impact of deep learning on photometric stereo, particularly its role in enhancing estimation processes under diverse lighting conditions.

7.1 Deep Learning and Photometric Stereo

Deep learning has profoundly transformed photometric stereo by facilitating accurate surface normal estimation across varied lighting conditions. Convolutional neural networks (CNNs) enable direct estimation from unstructured data, as demonstrated by methods using fixed-size observation maps [89]. This approach overcomes the limitations of traditional techniques reliant on structured lighting and predefined models. The well-posedness of uncalibrated photometric stereo under general illumination is established, with integrability reducing ambiguities in both orthographic and perspective cases [90]. Variational methods further enhance robustness by jointly recovering shape, reflectance, and illumination, ensuring integrability through direct depth estimation [91, 92].

Innovative deep learning strategies, such as CTPS, optimize illumination direction selection, achieving accurate surface normal estimation with minimal images [93]. This is particularly useful when traditional assumptions are violated, allowing for the identification of image subsets that better align with model assumptions, thereby improving reconstruction quality [41]. The development of differentiable display photometric stereo (DDPS) exemplifies deep learning's role in optimizing display patterns for enhanced reconstruction quality, demonstrating resilience to calibration errors [94]. Moreover, methods eliminating the need for light source calibration while maintaining accurate estimates underscore deep learning's efficacy in addressing conventional challenges [95].

Recent advances in deep learning have been particularly successful in handling non-Lambertian surfaces, enabling precise recovery from fewer images under varying conditions, as evidenced by benchmark evaluations [56, 93]. These innovations streamline data acquisition and enhance performance, pushing the boundaries of photometric stereo and facilitating further innovations in computer vision and graphics.

7.2 Advancements in Uncalibrated Photometric Stereo

Uncalibrated photometric stereo has advanced significantly, improving surface normal estimation without prior lighting knowledge. Integrability reduces solution ambiguity, enhancing accuracy in both orthographic and perspective scenarios [90]. Variational approaches effectively address the nonconvex nature of optimization problems by ensuring integrability through direct depth estimation, robustly recovering shape, reflectance, and illumination [92]. Spherical harmonic expansions of the Lambertian model further enhance these methods' capabilities under general illumination.

These advancements demonstrate uncalibrated photometric stereo's potential to surpass traditional limitations in 3D shape reconstruction. By integrating realistic assumptions like the Blinn–Phong model and perspective projection, these methods achieve improved accuracy across diverse lighting conditions, including complex scenarios like medical endoscopy [41, 29]. By focusing on integrability and advanced optimization, researchers expand surface normal estimation capabilities, paving the way for further innovations in computer vision and graphics.

7.3 Handling Non-Lambertian Surfaces and Complex Reflectance

Addressing non-Lambertian surfaces and complex reflectance in photometric stereo is challenging due to intricate light interactions with surfaces exhibiting specularities. Traditional methods often assume Lambertian reflectance, simplifying the problem by treating surfaces as perfectly diffuse, which is inadequate for real-world surfaces with mixed reflection characteristics. This limitation arises from idealized conditions like orthographic projection and distant light sources, neglecting complexities encountered in applications such as medical endoscopy. Comprehensive reflectance models, like the Blinn–Phong model with perspective projection, better capture light interactions, enhancing 3D reconstruction accuracy [41, 29].

Recent techniques leveraging deep learning and variational approaches tackle these challenges. The DeepPS2 method highlights deep learning's potential to enhance robustness in data-limited scenarios, resolving surface normal estimation ambiguities with minimal images [96]. The MT-PS-CNN model achieves high accuracy for non-Lambertian objects with fewer parameters, illustrating learning-based techniques' effectiveness in capturing intricate interactions [97]. Variational approaches contribute by estimating shape and reflectance from images under varying lighting without calibration [91].

Despite progress, challenges persist, particularly with shiny materials or narrow specularities poorly represented in training datasets [89]. Current studies often rely on calibrated conditions and extensive training data, limiting generalizability [56]. Overcoming these limitations is crucial for improving accuracy and applicability in handling non-Lambertian surfaces and complex reflectance.

Integrating deep learning with variational methods advances photometric stereo, offering innovative solutions for accurate surface normal estimation in challenging scenarios. Recent studies highlight deep learning's effectiveness in addressing real-world material complexities, enabling robust estimation under varied illumination. These advancements enhance recovery accuracy while reducing required input images, improving photometric stereo's practicality across diverse environments. Techniques like CNNs leverage spatial and photometric contexts, refining estimation processes and outperforming traditional methods [56, 97, 93, 60, 98]. These developments pave the way for further innovations in computer vision and graphics, enhancing 3D reconstructions' realism and fidelity.

8 Applications and Future Directions

The exploration of 3D vision techniques showcases their transformative potential across various fields, enhancing functionality and user experience in domains such as robotics, digital media, and medical imaging. This section highlights specific applications and insights into the broader implications of these technologies.

8.1 Practical Applications in Various Fields

3D vision techniques are pivotal in diverse applications, notably improving capabilities in robotics, autonomous vehicles, digital media, and real estate. In digital media, advancements in inverse rendering enable photorealistic object insertion and material editing, crucial for augmented reality (AR) and interior design, allowing users to visualize new decor in their spaces [8]. The generation of high-quality animations from near-duplicate photos further illustrates 3D vision's potential in enhancing visual storytelling and immersive experiences [5].

In robotics and autonomous vehicles, accurate 3D environment reconstruction is essential for navigation and interaction. Techniques like multi-view normal estimation and bundle adjustment provide precise surface normal and point estimations, vital for reverse engineering [48]. These advancements contribute to developing robust autonomous systems capable of operating in complex environments.

Photography and cinematography benefit from 3D vision through realistic lighting effects and scene relighting. Dherse et al. [57] demonstrate practical applications that enhance visual realism and creative expression. Additionally, capturing spectral power distributions without expensive equipment broadens accessibility for artists, enabling high-fidelity digital models [6].

In medical imaging, 3D vision techniques are applied for tissue rendering and in entertainment for visual effects and virtual/augmented reality experiences. These techniques enhance scene understanding through improved semantic representations and address challenges in accurately reconstructing complex geometries and reflectance properties from 2D images [26, 55, 9, 29, 25].

Comprehensive benchmarks for generating large, photorealistic datasets are vital for advancing computer vision model training, significantly improving models' accuracy and robustness in real-world applications. Datasets like UG² provide challenging video imagery under varying conditions and incorporate tasks to assess the impact of algorithms on visual quality and object recognition

[21, 10, 86]. This capability is crucial for developing intelligent systems capable of performing complex tasks in dynamic environments.

The extensive applications of 3D vision techniques significantly advance innovation and capabilities in fields such as medical imaging, robotics, and virtual/augmented reality. For example, photometric stereo methods enable high-accuracy 3D shape reconstruction from images under varying lighting conditions, while learning-based algorithms enhance image relighting and scene understanding by integrating geometry cues and semantic information. These advancements facilitate complex tasks like medical endoscopy and enhance the realism and interactivity of virtual environments [9, 25, 55, 29]. As research progresses, these techniques are expected to play an increasingly pivotal role in shaping technology and its applications.

8.2 Challenges and Emerging Trends

3D vision research is characterized by significant advancements and persistent challenges, with emerging trends shaping its future trajectory. A primary challenge is the computational demand of methods like light attenuation map recovery, which are often time-consuming and resource-intensive. Many approaches also rely on depth information, limiting their applicability in scenarios lacking such data [6]. The dependence on synthetic data for training further complicates matters, as it may not capture the variability of real-world lighting conditions, posing challenges in 3D vision research [75].

Handling complex lighting scenarios remains a significant challenge, particularly for accurate relighting and shadow effects. Methods that do not enforce physical accuracy in lighting direction can lead to artifacts, including incorrect shadow representations. Existing methods often struggle with large shadow areas or significant inter-reflections, as small spatial contexts may not adequately capture these effects [77]. The challenges in human relighting, such as assumptions of Lambertian materials and inaccuracies in geometry reconstruction, underscore the need for more robust approaches [4].

Emerging trends in 3D vision focus on data-driven approaches that learn intrinsic properties from data, simplifying relighting and enhancing generalization. The ability to learn without ground truth data and adapt to various test scenes without pre-training offers significant advantages, as demonstrated by neural methods that address complex lighting scenarios and provide sharper shadows through visibility approximations [10].

The integration of neural rendering frameworks that support free viewpoint rendering and address challenges in refining geometry and adaptively sampling light directions exemplifies advancements in this field. However, challenges remain in optimizing GPU memory requirements and reducing training times, particularly for relightable outdoor scenes. Certain methods stand out in relighting due to their advanced capability to model indirect lighting and produce high-quality results, as shown by perceptually-inspired shading models that decompose shading into components, efficient techniques for recovering spatially-varying indirect illumination using neural radiance fields, and volumetric relighting systems that synthesize complex lighting effects from single images without explicit physical lighting priors [37, 16, 99, 18].

Despite advancements, challenges persist in handling materials with strong mirror reflections and curved reflective surfaces, which may introduce inaccuracies in the separation process [4]. Incorrect normals in the presence of specular inter-reflections remain a critical issue.

Addressing challenges in 3D vision while leveraging emerging trends, such as integrating neural rendering techniques and advancements in computational photography, is essential for pushing the field's boundaries. This approach enhances visual quality and interpretability of images and facilitates the development of algorithms that improve image restoration and visual recognition. The introduction of the UG^2 dataset and ongoing exploration of deep generative models will pave the way for innovative applications across computer vision and graphics, including novel viewpoint synthesis, semantic photo manipulation, and realistic avatar creation for virtual and augmented reality experiences [21, 10, 32]. Key advantages include the ability to handle large scenes, joint optimization of geometry and materials, and simplified optimization processes through a single objective function.

8.3 Future Research Directions

Future research in 3D vision aims to advance the robustness, accuracy, and versatility of inverse rendering techniques and other applications. A critical focus area is enhancing the handling of complex lighting effects and improving texture prediction robustness, which could significantly enhance the realism of 3D reconstructions [26]. Additionally, refining models for better indoor scene handling and exploring more general reflectance models could lead to advancements in inverse rendering, allowing for more accurate scene reconstructions [100].

Integrating techniques such as hash encoding for acceleration and utilizing fine-detailed 3D priors for enhanced decomposition accuracy represents another promising research avenue. These approaches could improve the efficiency and precision of geometry and reflectance disentanglement, facilitating more detailed and accurate 3D models [101]. Furthermore, extending models to account for ambient light and global illumination effects, as well as improving the expressiveness of the Bidirectional Reflectance Distribution Function (BRDF) model, could address complex material properties and enhance the realism of rendered scenes [102].

In inverse rendering, future research should enhance geometry optimization capabilities and explore efficient gradient computation methods to improve reconstruction accuracy [3]. Additionally, extending GenLit to incorporate HDRI maps for comprehensive ambient environment manipulation presents a promising direction [58]. Improving regularization techniques and exploring alternative differentiable mesh generation methods are essential for enhancing reconstruction quality [103].

For Factored-NeuS, future research should focus on capturing fine details in material reconstruction and extending the approach to dynamic objects and additional modalities [40]. In photometric stereo, extending the PX-NET approach to multi-view settings and exploring complex material interactions could lead to significant advancements [7]. Developing a real dataset with ground truth for quantitative evaluations would provide valuable insights into photometric stereo methods' performance [50].

Future work should also refine network architectures and experiment with different loss functions to enhance the realism of generated images in scene relighting [57]. Improving photo pair selection for processing and enhancing robustness to complex scenes are potential future research directions in 3D vision [5]. Additionally, enhancing robustness in scenes with fewer pure pixels and exploring dynamic lighting environments are critical areas for future exploration [6].

By addressing these research directions, 3D vision is poised for significant advancements, paving the way for more accurate, efficient, and versatile applications across various domains. Future work may focus on improving indirect illumination handling and optimizing the performance of methods [77]. Exploring alternative environment representations for better high-frequency lighting effects handling and enhancing shadowing models could significantly impact future research efforts [1]. Future directions also include improving runtime efficiency and exploring joint optimization of geometry, material, and lighting [62]. Additionally, investigating complex lighting representations and techniques to mitigate model personalization overhead remains crucial for ongoing evolution in 3D vision techniques [4]. Developing more robust algorithms to handle a wider range of imaging artifacts while enhancing both visual quality and recognition accuracy is also essential [10]. Lastly, exploring additional modalities for enriching scene representations and extending frameworks to dynamic scenes is vital for the continued evolution of 3D vision techniques [9].

8.4 Potential Impact and Ethical Considerations

Advancements in 3D vision techniques have the potential to transform industries by enhancing the realism and interactivity of digital environments. The ability to synthesize realistic lighting effects, as demonstrated by methods like GSPhongMeta [104], can significantly improve visual media quality, offering immersive experiences in entertainment, virtual reality, and digital marketing. However, these advancements raise ethical challenges, particularly regarding misleading content creation. High-fidelity manipulation of lighting and material properties may lead to deceptive images or videos, raising concerns about authenticity and trust in digital media.

The application of 3D vision techniques to images containing humans or commercial intellectual properties, as seen in models like RelitLRM [105], introduces additional ethical considerations. The potential misuse in altering or fabricating visual content without consent could infringe on privacy rights and intellectual property laws, necessitating ethical guidelines and regulations for responsible use.

The development of face manipulation technologies, such as methods that add self-shadows [106], highlights the dual-use nature of these technologies. While enhancing visual realism, there is a risk of misuse in creating deepfakes or impacting surveillance systems. The limited scope of malicious applications due to the specific nature of these enhancements does not diminish the need for vigilance in monitoring and mitigating potential abuses.

9 Conclusion

The exploration of 3D vision techniques within this survey underscores both the remarkable progress and the ongoing challenges in the field. Central themes include inverse rendering, relighting, and material estimation, each playing a crucial role in enhancing computational photography and photometric stereo. The integration of dense metric depth emerges as a critical factor for improving the precision of 3D reconstructions and the effectiveness of relighting processes, underscoring the importance of accurate depth information in achieving high-fidelity visual outputs.

Innovative developments, such as the Illumination-Aware Network (IAN), demonstrate significant advancements in relighting tasks, achieving notable improvements in performance while maintaining efficiency. These advancements highlight the continuous efforts to balance computational resource optimization with the enhancement of realism and flexibility in relighting techniques. Additionally, systems like MulayCap illustrate the transformative potential of 3D vision in the entertainment and digital media sectors, capturing dynamic human performances with exceptional geometric and textural detail, thus facilitating more adaptable workflows.

The survey further emphasizes the pivotal role of neural rendering and learning-based methodologies in advancing 3D vision, enhancing the extraction of intrinsic scene properties and elevating the realism of digital environments. Despite these advancements, challenges remain, particularly in managing complex lighting scenarios, non-Lambertian surfaces, and the computational intensity required for high-quality reconstructions.

Looking ahead, future research is expected to prioritize the enhancement of robustness and precision in inverse rendering techniques, alongside the development of more efficient computational models. Addressing ethical considerations related to visual content manipulation will also be a key focus. By advancing the current capabilities, researchers aim to unlock new applications for 3D vision across diverse fields, fostering the creation of more immersive and interactive digital experiences.

References

[1] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021.

[2] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable point-based inverse rendering, 2024.

[3] Yuxin Dai, Qi Wang, Jingsen Zhu, Dianbing Xi, Yuchi Huo, Chen Qian, and Ying He. Inverse rendering using multi-bounce path tracing and reservoir sampling, 2025.

[4] Ruofan Liang, Zan Gojcic, Merlin Nimier-David, David Acuna, Nandita Vijaykumar, Sanja Fidler, and Zian Wang. Photorealistic object insertion with diffusion-guided inverse rendering, 2024.

[5] Qianqian Wang, Zhengqi Li, David Salesin, Noah Snavely, Brian Curless, and Janne Kontkanen. 3d moments from near-duplicate photos, 2022.

[6] Zhuo Hui, Kalyan Sunkavalli, Sunil Hadap, and Aswin C. Sankaranarayanan. Illuminant spectra-based source separation using flash photography, 2017.

[7] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. Px-net: Simple and efficient pixel-wise training of photometric stereo networks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12757–12766, 2021.

[8] Guanzhou Ji, Azadeh O. Sawyer, and Srinivasa G. Narasimhan. Virtual home staging: Inverse rendering and editing an indoor panorama under natural illumination, 2024.

[9] Qucheng Peng, Benjamin Planche, Zhongpai Gao, Meng Zheng, Anwesa Choudhuri, Terrence Chen, Chen Chen, and Ziyan Wu. 3d vision-language gaussian splatting, 2024.

[10] Rosaura G. VidalMata, Sreya Banerjee, Brandon RichardWebster, Michael Albright, Pedro Davalos, Scott McCloskey, Ben Miller, Asong Tambo, Sushobhan Ghosh, Sudarshan Nagesh, Ye Yuan, Yueyu Hu, Junru Wu, Wenhan Yang, Xiaoshuai Zhang, Jiaying Liu, Zhangyang Wang, Hwann-Tzong Chen, Tzu-Wei Huang, Wen-Chi Chin, Yi-Chun Li, Mahmoud Lababidi, Charles Otto, and Walter J. Scheirer. Bridging the gap between computational photography and visual recognition, 2020.

[11] Xiaoming Zhao, Pratul P. Srinivasan, Dor Verbin, Keunhong Park, Ricardo Martin Brualla, and Philipp Henzler. Illuminerf: 3d relighting without inverse rendering, 2024.

[12] Diego Gomez, Julien Philip, Adrien Kaiser, and Élie Michel. Rrm: Relightable assets using radiance guided material extraction, 2024.

[13] Zhen Li, Lingli Wang, Mofang Cheng, Cihui Pan, and Jiaqi Yang. Multi-view inverse rendering for large-scale real-world indoor scenes, 2023.

[14] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. Shape, light, and material decomposition from images using monte carlo rendering and denoising, 2022.

[15] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5124–5133, 2020.

[16] Zicheng Liao, Kevin Karsch, Hongyi Zhang, and David Forsyth. An approximate shading model with detail decomposition for object relighting, 2018.

[17] Audrey Cui, Ali Jahanian, Agata Lapedriza, Antonio Torralba, Shahin Mahdizadehaghdam, Rohit Kumar, and David Bau. Local relighting of real scenes, 2022.

[18] This cvpr paper is the open acce.

[19] Li-Wen Wang, Wan-Chi Siu, Zhi-Song Liu, Chu-Tak Li, and Daniel PK Lun. Deep relighting networks for image light source manipulation. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 550–567. Springer, 2020.

[20] Xin Jin, Yannan Li, Ningning Liu, Xiaodong Li, Xianggang Jiang, Chaoen Xiao, and Shiming Ge. Single reference image based scene relighting via material guided filtering, 2017.

[21] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Yifan Wang, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, Tomas Simon, Christian Theobalt, Matthias Niessner, Jonathan T. Barron, Gordon Wetzstein, Michael Zollhoefer, and Vladislav Golyanik. Advances in neural rendering, 2022.

[22] Youming Deng, Xueting Li, Sifei Liu, and Ming-Hsuan Yang. Physics-based indirect illumination for inverse rendering, 2023.

[23] Chengqian Che, Fujun Luan, Shuang Zhao, Kavita Bala, and Ioannis Gkioulekas. Inverse transport networks, 2018.

[24] Kevin Karsch. Inverse rendering techniques for physically grounded image editing, 2019.

[25] Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A Efros, and George Drettakis. Multi-view relighting using a geometry-aware network. *ACM Trans. Graph.*, 38(4):78–1, 2019.

[26] Learning to predict 3d objects w.

[27] Ye Yu, Abhimitra Meka, Mohamed Elgharib, Hans-Peter Seidel, Christian Theobalt, and William AP Smith. Self-supervised outdoor scene relighting. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 84–101. Springer, 2020.

[28] Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, and Vishal M. Patel. Holo-relighting: Controllable volumetric portrait relighting from a single image, 2024.

[29] Research article.

[30] Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Nerf: Neural radiance field in 3d vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*, 2022.

[31] Zezhun Shi. Feature extraction reimagined: Achieving superior accuracy in camera calibration, 2024.

[32] Ayush Tewari, Ohad Fried, Justus Thies, Vincent Sitzmann, Stephen Lombardi, Kalyan Sunkavalli, Ricardo Martin-Brualla, Tomas Simon, Jason Saragih, Matthias Nießner, Rohit Pandey, Sean Fanello, Gordon Wetzstein, Jun-Yan Zhu, Christian Theobalt, Maneesh Agrawala, Eli Shechtman, Dan B Goldman, and Michael Zollhöfer. State of the art on neural rendering, 2020.

[33] Akshay Gadi Patil and Shanmuganathan Raman. Automatic content-aware non-photorealistic rendering of images, 2016.

[34] Rohit Pandharkar, Ashok Veeraraghavan, and Ramesh Raskar. Progressive versus random projections for compressive capture of images, lightfields and higher dimensional visual signals, 2011.

[35] Bruno Berenguel-Baeta, Jesus Bermudez-Cameo, and Jose J. Guerrero. Omniscv: An omnidirectional synthetic image generator for computer vision, 2024.

[36] Xiangyang Zhu, Yiling Pan, Bailin Deng, and Bin Wang. Efficient multi-view inverse rendering using a hybrid differentiable rendering method, 2023.

[37] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering, 2022.

[38] Haoyuan Wang, Wenbo Hu, Lei Zhu, and Rynson W. H. Lau. Inverse rendering of glossy objects via the neural plenoptic function and radiance fields, 2024.

[39] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P. A. Lensch. Nerd: Neural reflectance decomposition from image collections, 2021.

[40] Yue Fan, Ivan Skorokhodov, Oleg Voynov, Savva Ignatyev, Evgeny Burnaev, Peter Wonka, and Yiqun Wang. Factored-neus: Reconstructing surfaces, illumination, and materials of possibly glossy objects, 2023.

[41] Elisa Crabu, Federica Pes, Giuseppe Rodriguez, and Giuseppa Tanda. Ascertaining the ideality of photometric stereo datasets under unknown lighting, 2023.

[42] Maryam Khanian, Ali Sharifi Boroujerdi, and Michael Breuß. Photometric stereo for strong specular highlights, 2017.

[43] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis, 2020.

[44] Julian Knodt, Joe Bartusek, Seung-Hwan Baek, and Felix Heide. Neural ray-tracing: Learning surfaces and reflectance for relighting and view synthesis, 2021.

[45] Lixiang Lin, Jianke Zhu, and Yisu Zhang. Multiview textured mesh recovery by differentiable rendering, 2022.

[46] JunYong Choi, SeokYeong Lee, Haesol Park, Seung-Won Jung, Ig-Jae Kim, and Junghyun Cho. Mair++: Improving multi-view attention inverse rendering with implicit lighting representation, 2024.

[47] Xuening Zhu, Renjiao Yi, Xin Wen, Chenyang Zhu, and Kai Xu. Relighting scenes with object insertions in neural radiance fields, 2024.

[48] Ivan Eichhardt and Levente Hajder. Computer vision meets geometric modeling: Multi-view reconstruction of surface points and normals using affine correspondences. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2427–2435, 2017.

[49] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, François Lauze, and Jean-Denis Durou. Led-based photometric stereo: Modeling, calibration and numerical solution, 2017.

[50] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts, 2022.

[51] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. A cnn based approach for the near-field photometric stereo problem. *arXiv preprint arXiv:2009.05792*, 2020.

[52] Fotios Logothetis, Roberto Mecca, and Roberto Cipolla. A differential volumetric approach to multi-view photometric stereo, 2019.

[53] Xiaokang Wei, Zhuoman Liu, and Yan Luximon. Sir: Multi-view inverse rendering with decomposable shadow for indoor scenes, 2024.

[54] A Karami, F Menna, and F Remondino. Investigating 3d reconstruction of non-collaborative surfaces through photogrammetry and photometric stereo. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:519–526, 2021.

[55] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, David Kriegman, and Ravi Ramamoorthi. Deep 3d capture: Geometry and reflectance from sparse multi-view images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5960–5969, 2020.

[56] Yakun Ju, Kin-Man Lam, Wuyuan Xie, Huiyu Zhou, Junyu Dong, and Boxin Shi. Deep learning methods for calibrated photometric stereo and beyond, 2024.

[57] Alexandre Pierre Dherse, Martin Nicolas Everaert, and Jakub Jan Gwizdała. Scene relighting with illumination estimation in the latent space on an encoder-decoder scheme, 2020.

[58] Shrisha Bharadwaj, Haiwen Feng, Victoria Abrevaya, and Michael J. Black. Genlit: Reformulating single-image relighting as video generation, 2024.

[59] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. Geometry-aware single-image full-body human relighting. In *European Conference on Computer Vision*, pages 388–405. Springer, 2022.

[60] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 501–509, 2017.

[61] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncalibrated neural inverse rendering for photometric stereo of general surfaces, 2021.

[62] Linjie Lyu, Ayush Tewari, Thomas Leimkuehler, Marc Habermann, and Christian Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination, 2022.

[63] Jiaye Wu, Saeed Hadadan, Geng Lin, Matthias Zwicker, David Jacobs, and Roni Sengupta. Gani: Global and near field illumination aware neural inverse rendering, 2024.

[64] A. Concas, R. Dessì, C. Fenu, G. Rodriguez, and M. Vanzi. Identifying the lights position in photometric stereo under unknown lighting, 2021.

[65] Junxuan Li and Hongdong Li. Self-calibrating photometric stereo by neural inverse rendering, 2022.

[66] James A. D. Gardner, Evgenii Kashin, Bernhard Egger, and William A. P. Smith. The sky's the limit: Re-lightable outdoor scenes via a sky-pixel constrained illumination prior and outside-in visibility, 2024.

[67] Jingzhi Bao, Guanying Chen, and Shuguang Cui. Pir: Photometric inverse rendering with shading cues modeling and surface reflectance regularization, 2025.

[68] Maryam Khanian, Ali Sharifi Boroujerdi, and Michael Breuss. Real-time 3d shape of micro-details, 2018.

[69] Deheng Zhang, Jingyu Wang, Shaofei Wang, Marko Mihajlovic, Sergey Prokudin, Hendrik P. A. Lensch, and Siyu Tang. Rise-sdf: a relightable information-shared signed distance field for glossy object inverse rendering, 2024.

[70] Hongze Chen, Zehong Lin, and Jun Zhang. Gi-gs: Global illumination decomposition on gaussian splatting for inverse rendering, 2024.

[71] Moritz Heep and Eduard Zell. An adaptive screen-space meshing approach for normal integration, 2024.

[72] Ido Czerninski and Yoav Y. Schechner. Accelerating inverse rendering by using a gpu and reuse of light paths, 2021.

[73] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable inverse rendering with interpretable basis brdfs, 2024.

[74] Yixiong Yang, Shilin Hu, Haoyu Wu, Ramon Baldrich, Dimitris Samaras, and Maria Vanrell. Learning relighting and intrinsic decomposition in neural radiance fields, 2024.

[75] Isabella Liu, Linghao Chen, Ziyang Fu, Liwen Wu, Haian Jin, Zhong Li, Chin Ming Ryan Wong, Yi Xu, Ravi Ramamoorthi, Zexiang Xu, and Hao Su. Openillumination: A multi-illumination dataset for inverse rendering evaluation on real objects, 2024.

[76] Paul Gafton and Erick Maraz. 2d image relighting with image-to-image translation, 2020.

[77] Jingzhi Li, Zongwei Wu, Eduard Zamfir, and Radu Timofte. Recap: Better gaussian relighting with cross-environment captures, 2024.

[78] Zhengqin Li, Jia Shi, Sai Bi, Rui Zhu, Kalyan Sunkavalli, Miloš Hašan, Zexiang Xu, Ravi Ramamoorthi, and Manmohan Chandraker. Physically-based editing of indoor scene lighting from a single image, 2022.

[79] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, Jonathan T. Barron, Ravi Ramamoorthi, and William T. Freeman. Neural light transport for relighting and view synthesis, 2021.

[80] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting, 2022.

[81] Zian Wang, Tianchang Shen, Jun Gao, Shengyu Huang, Jacob Munkberg, Jon Hasselgren, Zan Gojcic, Wenzheng Chen, and Sanja Fidler. Neural fields meet explicit geometric representation for inverse rendering of urban scenes, 2023.

[82] Learning inverse rendering of fa.

[83] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials, 2020.

[84] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Multi-view photometric stereo revisited. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3126–3135, 2023.

[85] Benjamin Ummenhofer, Sanskar Agrawal, Rene Sepulveda, Yixing Lao, Kai Zhang, Tianhang Cheng, Stephan Richter, Shenlong Wang, and German Ros. Objects with lighting: A real-world dataset for evaluating reconstruction and rendering for object relighting, 2024.

[86] Benjamin Ummenhofer, Sanskar Agrawal, Rene Sepulveda, Yixing Lao, Kai Zhang, Tianhang Cheng, Stephan Richter, Shenlong Wang, and German Ros. Objects with lighting: A real-world dataset for evaluating reconstruction and rendering for object relighting. In *2024 International Conference on 3D Vision (3DV)*, pages 137–147. IEEE, 2024.

[87] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20762–20772, 2023.

[88] Bingyao Huang and Haibin Ling. Deprocams: Simultaneous relighting, compensation and shape reconstruction for projector-camera systems, 2021.

[89] Satoshi Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–18, 2018.

[90] Mohammed Brahimi, Yvain Quéau, Bjoern Haefner, and Daniel Cremers. On the well-posedness of uncalibrated photometric stereo under general lighting, 2020.

[91] Bjoern Haefner, Zhenzhang Ye, Maolin Gao, Tao Wu, Yvain Quéau, and Daniel Cremers. Variational uncalibrated photometric stereo under general lighting, 2019.

[92] This iccv paper is the open acce.

[93] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7568–7576, 2019.

[94] Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, and Seung-Hwan Baek. Differentiable display photometric stereo, 2024.

[95] Kazuma Minami, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Symmetric-light photometric stereo. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2706–2714, 2022.

[96] Ashish Tiwari and Shanmuganathan Raman. Deepps2: Revisiting photometric stereo using two differently illuminated images. In *European Conference on Computer Vision*, pages 129–145. Springer, 2022.

[97] Yanlong Cao, Binjie Ding, Zewei He, Jiangxin Yang, Jingxi Chen, Yanpeng Cao, and Xin Li. Learning inter- and intraframe representations for non-lambertian photometric stereo, 2021.

[98] David Honzátko, Engin Türetken, Pascal Fua, and L. Andrea Dunbar. Leveraging spatial and photometric context for calibrated non-lambertian photometric stereo, 2021.

[99] Parisha Joshi and Daljit Singh J. Dhillon. Practical and accurate reconstruction of an illuminant's spectral power distribution for inverse rendering pipelines, 2024.

[100] Ye Yu and William A. P. Smith. Outdoor inverse rendering from a single image using multiview self-supervision, 2021.

[101] Daisheng Jin, Jiangbei Hu, Baixin Xu, Yuxin Dai, Chen Qian, and Ying He. Robust geometry and reflectance disentanglement for 3d face reconstruction from sparse-view images, 2023.

[102] Carolin Schmitt, Božidar Antić, Andrei Neculai, Joo Ho Lee, and Andreas Geiger. Towards scalable multi-view reconstruction of geometry and materials, 2023.

[103] Guangyan Cai, Kai Yan, Zhao Dong, Ioannis Gkioulekas, and Shuang Zhao. Physics-based inverse rendering using combined implicit and explicit geometries, 2022.

[104] Yumeng He, Yunbo Wang, and Xiaokang Yang. Gs-phong: Meta-learned 3d gaussians for relightable novel view synthesis, 2024.

[105] Tianyuan Zhang, Zhengfei Kuang, Haian Jin, Zexiang Xu, Sai Bi, Hao Tan, He Zhang, Yiwei Hu, Milos Hasan, William T. Freeman, Kai Zhang, and Fujun Luan. Relitlrm: Generative relightable radiance for large reconstruction models, 2024.

[106] Andrew Hou, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Face relighting with geometrically consistent shadows. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4217–4226, 2022.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

AI-generated, for reference only.