

A Real-world Display Inverse Rendering Dataset

Seokjun Choi* Hoon-Gyu Chung* Yujin Jeon* Giljoo Nam† Seung-Hwan Baek*
 *POSTECH †Meta

Abstract

Inverse rendering aims to reconstruct geometry and reflectance from captured images. Display-camera imaging systems offer unique advantages for this task: each pixel can easily function as a programmable point light source, and the polarized light emitted by LCD displays facilitates diffuse-specular separation. Despite these benefits, there is currently no public real-world dataset captured using display-camera systems, unlike other setups such as light stages. This absence hinders the development and evaluation of display-based inverse rendering methods. In this paper, we introduce the first real-world dataset for display-based inverse rendering. To achieve this, we construct and calibrate an imaging system comprising an LCD display and stereo polarization cameras. We then capture a diverse set of objects with diverse geometry and reflectance under one-light-at-a-time (OLAT) display patterns. We also provide high-quality ground-truth geometry. Our dataset enables the synthesis of captured images under arbitrary display patterns and different noise levels. Using this dataset, we evaluate the performance of existing photometric stereo and inverse rendering methods, and provide a simple, yet effective baseline for display inverse rendering, outperforming state-of-the-art inverse rendering methods. Code and dataset are available on our project page at <https://michaelcsj.github.io/DIR/>.

1. Introduction

Inverse rendering is a long-standing problem in computer vision and graphics, aiming to recover scene properties such as geometry and reflectance from captured images [36, 54]. Recent progress in inverse rendering methods heavily rely on datasets that provide images of objects under well-characterized multiple lighting conditions [5, 11, 59], allowing for evaluation and training of models that infer geometry and reflectance from images.

Among various inverse rendering setups, display-camera imaging systems offer unique advantages. Unlike conventional light stages [18, 32, 50, 62], displays can serve as high-resolution, programmable light sources, allowing con-

venient control over illumination [1, 78]. Moreover, LCD displays emit polarized light, which facilitates the separation of diffuse and specular reflections [10, 35]. These characteristics make display-camera systems a compelling choice for inverse rendering research. However, despite their potential, the lack of publicly available datasets captured using such systems has hindered progress in this direction. Unlike other setups, such as light stages, which have been extensively used for photometric stereo and reflectance capture, display-camera inverse rendering lacks a standardized benchmark for method development and comparison.

In this paper, we introduce the first real-world dataset for display-based inverse rendering. We construct a display-camera imaging system consisting of a LCD monitor and a stereo polarization camera setup, enabling controlled illumination capture at two views with diffuse and specular separation. Using this system, we capture a diverse set of objects with varying geometries and reflectance properties under one-light-at-a-time (OLAT) display patterns. Each object is accompanied by ground-truth geometry obtained via structured-light scanning, enabling precise evaluation of inverse rendering methods. Our dataset also supports synthetic relighting and noise simulation, allowing researchers to generate novel lighting conditions using linear combinations of captured images. We also introduce a simple baseline method for display inverse rendering that effectively addresses associated challenges, outperforming previous methods. Our specific contributions are as follows:

- We build and calibrate a display-camera imaging system incorporating display backlight, which enables display-based illumination and stereo polarization imaging.
- We acquire the first high-quality real-world dataset for display-camera inverse rendering, featuring objects with diverse reflectance and ground-truth geometry.
- We evaluate existing photometric stereo and inverse rendering methods on our dataset, highlighting the challenges of display inverse rendering.
- We propose a simple yet effective baseline for display inverse rendering, outperforming previous methods.

2. Related Work

Imaging Systems for Inverse Rendering Inverse rendering typically requires observations of a target object under various lighting conditions. In the literature, different hardware configurations to modulate lighting conditions have been proposed. Light stages, a dome structure equipped with numerous high brightness LEDs, offer dense light-view angular samples for high-quality inverse rendering at the cost of large form factors and high instrumentation costs [18, 32, 50, 62]. Flash photography with mobile cameras provides a practical multi-view, multi-light setup, capturing many images from different views [3, 15, 23, 52, 56]. However, this requires moving the cameras and capturing objects multiple times. Using displays as controllable light sources provides a cost-effective and compact alternative, enabling convenient multi-light capture, having a potential for practical and high-quality inverse rendering [1, 10, 35, 78]. Display-camera systems present unique challenges and opportunities due to near-field lighting effects, limited light power, polarization properties of LCDs, and constrained light-view angular sampling. Addressing these challenges is an open problem.

Inverse Rendering Dataset Table 1 summarizes representative publicly available datasets for inverse rendering. While synthetic datasets provide ground truth under ideal scenarios [8, 24, 26], real-world datasets offer environments for realistic evaluation. Existing real-world datasets are captured with various imaging systems such as commodity cameras [19, 33, 57], light probes [34], gantries [20, 55, 66], robots [29, 65], and light stages [7, 47, 72]. Despite the increasing availability of real-world datasets, existing datasets fail to comprehensively evaluate inverse rendering in display-camera settings due to the use of other imaging systems for data acquisition. Recently, Choi et al. [10] employs 3D-printed objects for display photometric stereo. However, the 3D-printed dataset has limited material diversity, unsuitable as an inverse rendering dataset for real-world diverse objects.

Inverse Rendering Methods Learning-based inverse-rendering methods utilize CNN [6, 40, 41, 60, 63, 68, 69, 74], RNN [44], transformers [27, 82], and diffusion models [9, 16, 21, 42, 46, 49, 61] to infer geometry and reflectance in a data-driven manner. In contrast, analysis-by-synthesis methods take a physics-based approach, iteratively optimizing geometry and reflectance, ensuring that rendered images match the input images via differentiable forward rendering. Various differentiable rendering techniques have been explored, including volumetric rendering [48, 67, 73, 75–77, 79, 80], spherical Gaussians [76, 80], tensor-based formulations [31], point-based rendering [11],

Table 1. **Real-world inverse rendering datasets.** We present the first dataset for display inverse rendering with calibrated display and stereo polarization cameras. We also provide high-quality ground-truth geometry.

Dataset	Illumination system	Illumination type	Ground-truth geometry	Polarization
Alldrin et al. [2]	Light rig	Far-field	✗	✗
Grosse et al. [19]	Light rig	Far-field	✗	✗
Xiong et al. [71]	Light rig	Far-field	✗	✗
Jensen et al. [29]	Light rig	Far-field	✓	✗
Shi et al. [64]	Light rig	Far-field	✓	✗
Li et al. [39]	Light rig	Far-field	✓	✗
Mecca et al. [51]	Light rig	Near-field	✓	✗
Chabert et al. [7]	Light stage	Far-field	✗	✗
Liu et al. [47]	Light stage	Far-field	✗	✗
Yang et al. [72]	Light stage	Far-field	Pseudo	✓
Toschi et al. [65]	Gantry	Far-field	✗	✗
Kuang et al. [33]	In-the-wild	Env. map	✗	✗
Kuang et al. [34]	In-the-wild	Env. map	✓	✗
Ours	LCD display	Near-field	✓	✓

and Gaussian-based representations [5, 12, 17, 30, 43], and image-based neural representations [37]. Inverse rendering for display-camera systems introduces unique challenges and benefits for reconstruction methods due to near-field lighting conditions, display backlight, low signal-to-noise ratios, LCD polarization effects, and non-uniform angular sampling [1, 35, 78]. Developing reconstruction methods for display inverse rendering remains for future research.

3. Display-camera Imaging System

Setup To acquire a real-world dataset for display inverse rendering, we built a display-camera system, shown in Figure 1(a). Our setup consists of an LCD monitor (Samsung Odyssey Ark) and stereo polarization RGB cameras (FLIR BFS-U3-51S5PC-C) equipped with 8 mm focal-length lenses, covering 30° field of view. The LCD monitor emits vertically polarized light based on the principles of LCD [22]. The monitor maximum brightness is 600 cd/m², and each pixel only outputs a maximum intensity of 0.06 mcd, which is too dim to capture even with maximum-exposure imaging. Following [10], we parameterize display pixels using $144 = 16 \times 9$ superpixels, where each superpixel consists of 240×240 display pixels. Thus, we represent the display pattern as $\mathcal{L} = \{L_1, \dots, L_N\}$, where each superpixel has an RGB intensity L_i , and N denotes the total number of superpixels. The polarization RGB cameras capture the linearly-polarized light intensity for the R, G, and B channels at 0°, 45°, 90°, and 135°[4].

Display Backlight and Nonlinearity LCDs often cannot achieve complete darkness even when set to a black value as shown in Figure 1(b). Modeling this backlight is crucial, as backlight from all display pixels becomes visible in the

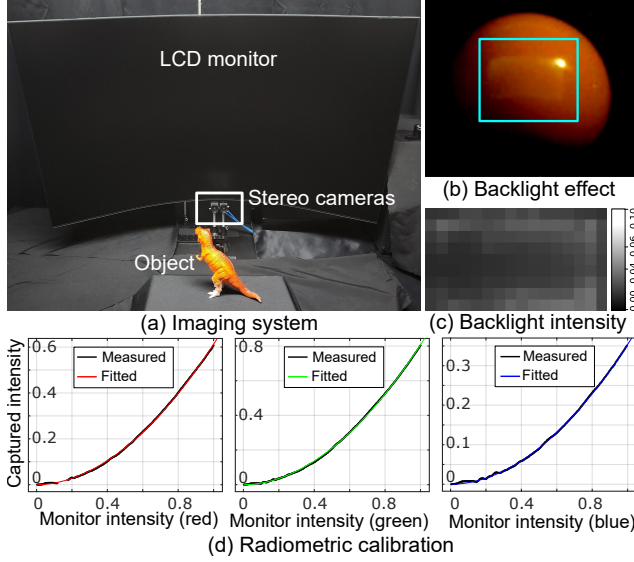


Figure 1. **Display-camera imaging system.** (a) Our imaging system consists of an LCD monitor and stereo polarization cameras. (b) The LCD monitor exhibits spatially-varying backlight as shown in one of the OLAT images, which (c) we calibrate for accurate inverse rendering. (d) We also obtain the non-linearity of the monitor intensity.

captured images. Also, the display intensity is nonlinearly mapped to the value to set, which should be also calibrated. Taking these into account, we model the i -th display superpixel light intensity given the corresponding RGB pattern value we set to display P_i as

$$L_i = s(P_i + B_i)^\gamma, \quad (1)$$

where s is a global scalar, γ is the non-linear mapping exponent, and B_i is the corresponding spatially-varying backlight intensity. To calibrate s , B_i , and γ , we captured a spherical object with known geometry and reflectance under OLAT patterns, and optimize the three parameters with a loss that minimizes the difference between the OLAT captured images and rendered OLAT images. Figure 1(c) shows the calibrated spatially-varying backlight that resembles the visible backlight in Figure 1(b).

Geometric Calibration We calibrate the stereo-camera intrinsic and extrinsic parameters using the checkerboard method [81]. We then estimate the position of each display superpixel relative to the reference left camera using the mirror-based checkerboard method [10].

Image Formation When illuminating a scene point with a display pattern $\mathcal{L}$, the captured intensity by a camera is

modeled as:

$$I = \text{clip} \left(\sum_{i=1}^N (\mathbf{n} \cdot \mathbf{i}) f(\mathbf{i}, \mathbf{o}) \frac{L_i}{d_i^2} + \epsilon \right), \quad (2)$$

where f is the BRDF, $\mathbf{n}$ is the surface normal, $\mathbf{i}$ is the incident light direction from the i -th display superpixel, $\mathbf{o}$ is the outgoing view vector, and d_i is the distance from the i -th display superpixel to the scene point. The function $\text{clip}(\cdot)$ applies clipping to the camera dynamic range, and ϵ is Gaussian noise.

4. Display Inverse Rendering Dataset

Figure 2 shows our real-world dataset for display inverse rendering. Each object has corresponding stereo-polarization RGB images captured under OLAT patterns, ground-truth depth maps, normal maps, and object masks.

Objects We captured 16 objects made of various materials and reflectances from diffuse to specular: resin (FROG, PIG, GNOME, SNOWMAN), ceramic (OWL, OBJET), metallic paint (CAT, ROBOT, NEFERTITI), wood (CHICKEN), clay (GIRL, BOY), plastic (TREX), bronze (HORSE), plaster (PLASTER), and composite (ELEPHANT). In terms of shape, the objects range from those with simple forms (OWL, CAT, PIG, OBJET, CHICKEN) to those featuring tiny parts (NEFERTITI), thin structures (HORSE, SNOWMAN), complex details (ELEPHANT, TREX) and curvature (PLASTER), as well as concave parts (FROG, GIRL, BOY, GNOME, ROBOT). The object sizes range from 8 cm to 25 cm. Objects are placed at 50 cm from the cameras for the capture.

Ground-truth Geometry To obtain ground-truth object shapes, we use structured-light scanning with a high-precision 3D scanner (EinScan SP V2), with a precision tolerance of 0.05 mm. We align the scanned 3D meshes to the captured images using the mutual information method [14]. Subsequently, we render depth maps, normal maps, and object masks for the camera views on Mitsuba3 [28].

Polarimetric Image Processing We first convert the captured polarization images at 0° , 45° , 90° , and 135° as $\{I_\theta\}_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}}$ into linear-polarization Stokes-vector RGB images [13]:

$$s_0 = \frac{\sum_\theta I_\theta}{2}, s_1 = I_{0^\circ} - I_{90^\circ}, s_2 = I_{45^\circ} - I_{135^\circ}. \quad (3)$$

Specular reflection tends to maintain the polarization state of display light whereas diffuse reflection becomes mostly unpolarized [10]. This enables us to obtain specular and diffuse images as $I_{\text{specular}} = \sqrt{(s_1)^2 + (s_2)^2}$ and $I_{\text{diffuse}} = s_0 - I_{\text{specular}}$, respectively, which are shown in Figure 2.



Figure 2. **Display Inverse Rendering Dataset.** We introduce the first display inverse rendering dataset. We obtain (a) combined, (b) diffuse, and (c) specular stereo images captured under (f–h) OLAT patterns. We provide ground-truth (d) normal maps and (e) depth maps.

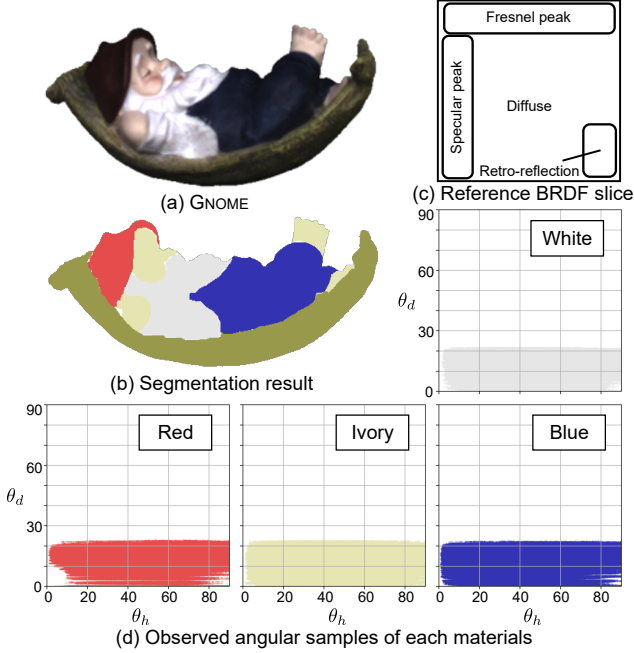


Figure 3. **Light-view angular samples.** Our display-camera system captures limited light-view angular samples. (a)&(b) For a segmented scene, (d) we show the sample plots of four segments in θ_d, θ_h Rusinkiewicz space [58]. (c) The sampled region corresponds to the typical specular, diffuse, and grazing reflections [53], allowing for inverse rendering.

Light-view Angular Samples Display inverse rendering poses challenges due to the limited coverage of light-view angular samples. In Figure 3, we examine the angular distribution of light-view samples for the segmented four material components. While a full BRDF requires sampling across all Rusinkiewicz coordinates [58], the display-camera setup provides only partial coverage, particularly in terms of θ_d , the angle between the half-way vector and the illumination vector. However, it is worth noting that the half-way angle θ_h is well-covered from 0 to $\pi/2$, enabling effective sampling of the specular lobe. Additionally, the sampled region corresponds to both diffuse and specular reflections[53]—a key factor that makes inverse rendering feasible.

Simulation for an Arbitrary Display Pattern Leveraging the linearity of incoherent light transport, we simulate a scene illuminated by an arbitrary display pattern $\mathcal{P} = \{P_1, \dots, P_N\}$, using Equation (2) and Equation (1), as:

$$I(\mathcal{P}) = \text{clip} \left(\sum_{i=1}^N I_i s(P_i + B_i)^\gamma + \epsilon \right), \quad (4)$$

where P_i is the display superpixel RGB value, I_i is the captured image under the i -th OLAT illumination. The standard

deviation of the Gaussian noise ϵ can be adjusted to reflect different noise levels.

5. A Baseline for Display Inverse Rendering

We propose a simple yet effective baseline for display inverse rendering, designed to handle inputs captured under M arbitrary display patterns, $\mathcal{P}_1, \dots, \mathcal{P}_M$. As an initialization step, we estimate the normal map using the analytical RGB photometric stereo method [10], which leverages M captured images. Additionally, we estimate a depth map by using the averaged stereo images across multiple patterns as inputs to RAFT stereo [45]. Given these normal map and depth map, we optimize the normal map and the reflectance (diffuse albedo, specular albedo, and roughness) of the Cook-Torrance BRDF model. To address the limitations of light-view angular sampling in the display-camera system, we adopt the basis BRDF representation, which models spatially varying BRDFs as a weighted sum of basis BRDFs [11, 12, 37]. Specifically, we use the analytic Cook-Torrance model to define each basis BRDF. We then differentially render reference-view images for the display patterns $\mathcal{P}_1, \dots, \mathcal{P}_M$ by implementing Equation 2 in PyTorch and iteratively update the scene representation—comprising normals, basis BRDFs, and their weight maps—by minimizing the RMSE error between the rendered and input images. Despite challenges such as limited light-view angular samples, display backlight, and near-field lighting in the display-camera setup, our baseline approach enables effective inverse rendering in only 150 seconds.

6. Evaluation

We assess previous photometric stereo methods, inverse rendering approaches, and our proposed baseline method (Section 5) using our display-camera dataset.

Photometric Stereo using OLAT Patterns Photometric stereo is a subtask of inverse rendering that focuses on normal reconstruction. We evaluate both calibrated [8, 25, 37, 70] and uncalibrated [26, 27, 38] methods on our dataset. As shown in Table 2 and Figure 4, recent uncalibrated photometric stereo techniques—particularly SDM-UniPS [27]—demonstrate highly accurate normal estimation. This indicates that the 144 OLAT images in our display setup provide sufficient information for precise normal reconstruction.

Inverse Rendering using OLAT Patterns Many existing inverse rendering methods cannot be directly applied to the display inverse rendering configuration due to the inherent challenges such as limited light-view angular samples, backlight, and near-field effects. To evaluate performance in this setting, we test four available inverse ren-

	ELEPHANT	OWL	CAT	FROG	ROBOT	PIG	CHICKEN	GIRL	BOY	NEFERTITI	TREX	GNOME	HORSE	SNOWMAN	PLASTER	OBJET
Woodham [70]	27.02	26.60	21.05	21.58	28.18	17.02	18.39	24.86	21.44	37.03	18.98	19.83	19.27	32.21	19.56	17.28
PS-FCN [8]	<u>20.26</u>	15.17	10.61	19.15	16.68	<u>15.80</u>	<u>11.91</u>	25.96	22.27	20.03	18.22	19.33	<u>17.48</u>	18.75	17.25	7.73
PS-Transformer [25]	26.42	36.43	21.11	35.34	27.31	49.10	16.20	38.66	35.91	30.64	29.86	36.53	35.06	54.26	33.97	24.06
SRS [37]	26.21	18.49	16.95	23.42	19.09	32.76	17.88	37.14	31.19	23.97	25.05	27.44	27.70	27.96	26.93	21.87
SCPS-NIR [38]	22.75	7.93	8.97	16.28	17.87	34.89	10.43	45.12	37.18	52.97	21.85	16.64	48.98	<u>15.65</u>	21.30	<u>7.94</u>
UniPS [26]	25.14	17.34	19.69	24.09	22.03	25.77	22.94	26.06	30.00	28.55	21.64	24.32	27.24	18.86	19.70	15.90
UniPS [26] ($M=64$)	24.93	18.33	19.54	24.99	22.18	25.72	23.07	26.38	30.65	28.71	21.86	24.48	26.72	18.89	19.43	16.39
SDM-UniPS [27] ($M=64$)	18.83	14.37	9.70	14.12	14.85	15.33	16.05	14.99	15.22	<u>22.73</u>	14.58	13.46	16.93	15.18	12.55	9.38
SDM-UniPS [27] ($M=10$)	20.53	<u>12.77</u>	<u>9.43</u>	<u>15.23</u>	<u>16.48</u>	16.12	16.10	<u>15.23</u>	<u>17.25</u>	24.32	<u>15.36</u>	<u>15.47</u>	17.62	16.57	<u>13.39</u>	9.58

Table 2. **Photometric-stereo evaluation using OLAT patterns.** Normal reconstruction error in Mean Angular Error (MAE) for calibrated (red) and uncalibrated (blue) photometric stereo. Highest performance in **bold** and the second-best in underline. When M is specified, it means the M number of uniform-sampled OLAT patterns is used for evaluation.

Method Patterns	Ours Multiplexed	Ours OLAT	SRS [37] OLAT	DPIR [11] OLAT	IIR [12] OLAT
PSNR [dB] $\uparrow$	37.27	<u>39.33</u>	41.28	34.30	38.20
SSIM $\uparrow$	0.9766	0.9821	0.9895	0.9790	<u>0.9850</u>
MAE [$^\circ$] $\downarrow$	<u>23.97</u>	20.94	25.25	41.09	38.38

Table 3. **Inverse-rendering evaluation.** Our baseline method enables high-quality relighting accuracy in PSNR and SSIM (first two rows) and normal accuracy in MAE (last row) for both OLAT and multiplexed patterns. While SRS enables effective relighting, the normal accuracy is low and non-trivial to support multiplexed patterns.

	Learned display pattern [10] $\downarrow$			Heuristic display pattern $\downarrow$		
	$M=2$	$M=4$	$M=10$	(a) $M=2$	(b) $M=4$	(c) $M=10$
UniPS [26]	27.7078	25.9408	<u>25.7541</u>	65.7171	63.1694	63.4573
SDM-UniPS [27]	23.5079	19.8946	18.1829	<u>42.3576</u>	29.9320	32.0718
DDPS [10]	<u>24.5678</u>	<u>23.3800</u>	29.3716	32.0480	<u>35.1451</u>	<u>36.5606</u>

Table 4. **Multiplexed patterns with varying numbers.** We evaluate normal reconstruction accuracy of photometric stereo methods using various numbers of heuristic patterns and learned display patterns.

	Learned display pattern [10] $\downarrow$		
	$M=2$	$M=4$	$M=10$
DDPS [10] (Diffuse +Specular)	24.5678	23.3800	29.3716
DDPS [10] (Diffuse)	23.2807	<u>21.2126</u>	<u>27.7281</u>
SDM-UniPS [27] (Diffuse +Specular))	<u>23.5079</u>	19.8946	18.1829
SDM-UniPS [27] (Diffuse)	35.0658	31.2040	30.1058

Table 5. **Photometric stereo with diffuse component and varying number of patterns.** We evaluate the impact of using diffuse images rather than the captured one containing both diffuse and specular components.

dering methods: one single-view approach [37], two multi-view methods [11, 12], and our proposed baseline model. For evaluation, we divide the 144 OLAT images into training and testing sets with a 5:1 ratio. As shown in Table 3 and Figure 5, our proposed baseline model achieves accurate relighting of specular appearances, whereas other methods produce blurry relighting results. This demonstrates that our approach effectively handles the challenges of limited light-view angular samples, backlight, and near-field effects, leading to robust display inverse rendering.

Multiplexed Display Patterns for Photometric Stereo

While OLAT images provide sufficient information for inverse rendering, capturing all 144 OLAT patterns is

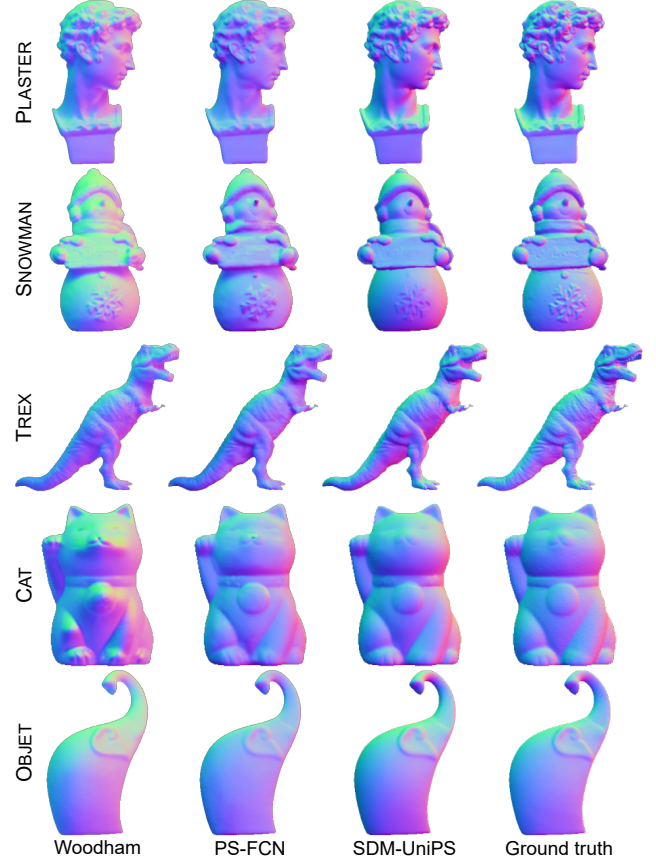


Figure 4. **Photometric stereo with OLAT patterns.** SDM-UniPS [27] demonstrates highly accurate normal reconstruction results, outperforming other methods.

time-consuming. A more efficient approach in display-camera systems is to use M multiplexed display patterns, formed as linear combinations of the OLAT patterns. We evaluated two multiplexed display pattern strategies: manually-designed and computationally learned patterns from DDPS [10]. As shown in Table 4 and Figure 6, even with just two multiplexed patterns, accurate normal reconstruction is achievable. Additionally, Table 4 presents results for the learned “Tri-random ($M=2$)” [10] and “Mono-gradient ($M=4$)” [50] patterns from DDPS, along with a concatenated pattern ($M=10$) that integrates these

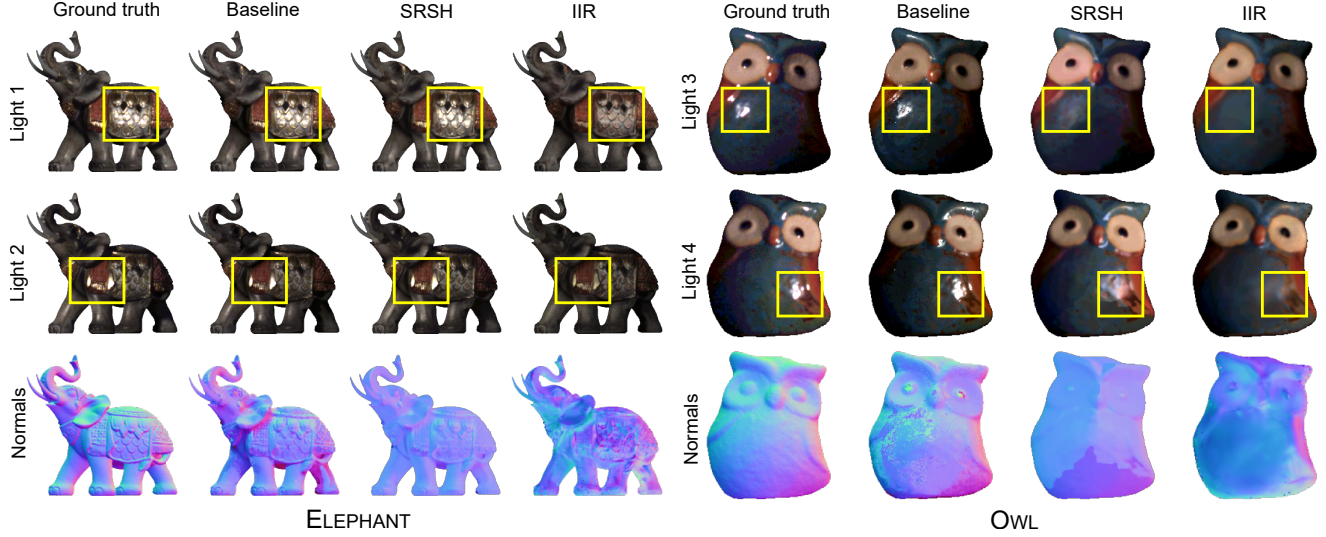


Figure 5. **Inverse rendering with OLAT patterns.** Our proposed baseline method (second column) achieves qualitatively more accurate relighting and normal reconstruction, outperforming other inverse rendering methods.

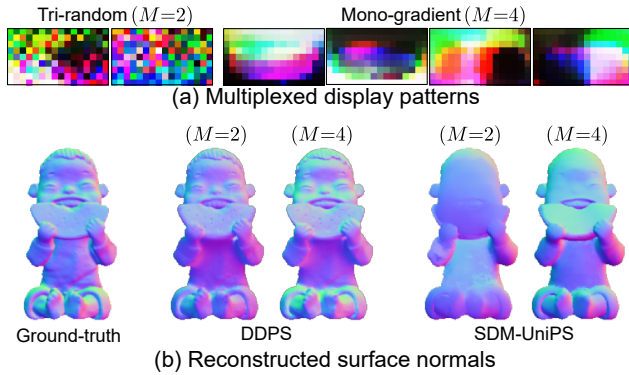


Figure 6. **Multiplexed display patterns for photometric stereo.** We found that analytical photometric stereo such as DDPS [10] is more robust to small number of display patterns than the learning-based photometric stereo such as SDM-UniPS.

with the “Mono-complementary” pattern [32]. For heuristic patterns, we tested the “Tri-complementary ($M=2$)” [35] and “Mono-gradient ($M=4$)” patterns [50], as well as a concatenated ($M=10$) pattern combining them with the “Mono-complementary” pattern [32]. Our results indicate that learned patterns consistently outperform heuristic patterns when using the same number of patterns. However, simply increasing the number of learned patterns does not always lead to further improvements in performance.

Multiplexed Display Patterns for Inverse Rendering

We evaluate the impact of multiplexed display patterns on our proposed baseline method for inverse rendering. Table 3 shows the quantitative results and Figure 7 presents the inverse rendering results using two patterns, each consisting of four images: a monochromatic gradient pattern [50] and a

	low res. ($M=32$)	32-inch ($M=50$)	Default
Woodham [70]	55.973	29.175	23.144
PS-FCN [8]	44.516	40.327	17.286
SDM-UniPS [27]	14.838	15.716	14.896

Table 6. **Impact of display configuration.** We found that normal reconstruction error (MAE) of SDM-UniPS [27] is low for different display configurations: our original display setup, low-resolution superpixels, and a 32-inch display size.

learned display pattern [10]. While the relighting results do not achieve the same accuracy as OLAT’s results, they still exhibit reasonable performance, with a relighting PSNR of 38.07 and 37.77 dB respectively. These findings suggest that designing display patterns that enable efficient capture while enhancing inverse rendering performance remains an open research challenge.

Impact of using Diffuse Images We evaluate the effect of incorporating polarization-separated diffuse images under the same set of display patterns in Table 5. As shown in Table 5, using diffuse images can improve normal reconstruction accuracy and efficiency in capture by reducing the number of required input images. However, this improvement is not consistent across all methods, suggesting that developing reconstruction methods that better use optically-separated diffuse and specular images is a future direction.

Impact of Display Specifications

We evaluate how different display specifications impact inverse rendering performance. Table 6 summarizes normal reconstruction results under various conditions, including lower-resolution superpixels and a simulated 32-inch monitor. When using superpixels smaller than 240×240 pixels to enhance reso-

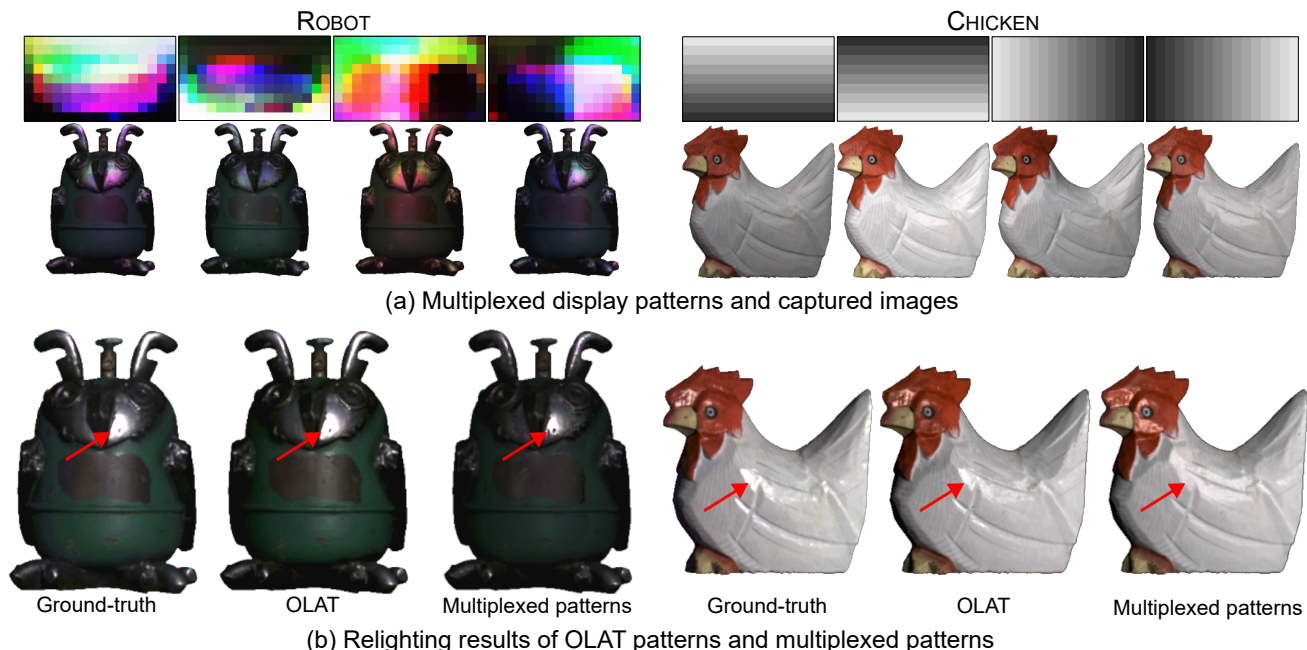


Figure 7. **Multiplexed display patterns for inverse rendering.** Inverse rendering performed with 144 OLAT patterns achieves relighting results that closely approximate the ground truth. Although inverse rendering can be performed using only four heuristic or learned patterns [10], relighting accuracy remains less accurate than that achieved with OLAT patterns.

lution, the captured images remain too dark even at maximum camera exposure, and this is unsuitable for inverse rendering. Conversely, with 480×480 -pixel superpixels arranged in an 8×4 resolution, the display behaves like an area light source, causing both the conventional method [70] and PS-FCN methods to fail in normal reconstruction. However, SDM-UniPS, which accounts for this type of lighting model, maintains relatively stable performance, with errors comparable to those observed when using 32 patterns. Additionally, when sampling only 10×5 superpixels—corresponding to the physical area of a 32-inch display—the Woodham’s method exhibits predictable performance degradation due to a reduced range of incident light angles, while PS-FCN fails to provide reliable estimates under this configuration. A notable observation in inverse rendering is the impact of removing distant light sources. In a 32-inch display setting, these sources are removed and improves the surface normal MAE of SRSR [37] from 25.25 to 17.68, highlighting the significant role of light attenuation in display-based setups. Furthermore, when the baseline model does not account for light attenuation, the PSNR drops from 39.78 to 37.43, confirming the importance of modeling near-field effects.

7. Conclusion

In this paper, we introduced the first real-world dataset for display inverse rendering. To construct this dataset, we developed a display-camera imaging system and carefully cal-

ibrated the display and camera parameters relevant to inverse rendering. Using our dataset, we conducted a comprehensive evaluation of existing photometric stereo and inverse rendering methods within the display-camera configuration. Our analysis revealed that current methods require further advancements, particularly in adapting to diverse display patterns, achieving robust reflectance reconstruction under limited light-view angular samples, and leveraging polarization properties inherent to display-camera setups. We hope that our dataset will serve as a resource, driving future developments and evaluations of inverse rendering methods for display-camera systems.

Future Directions Future work could explore advanced methods for effectively exploiting separated diffuse-specular components, as well as methods to handle the challenges posed by limited light-view angular samples. In addition, investigating optimized multiplexed display patterns and their corresponding reconstruction methods presents a promising avenue for further research. We believe that the dataset we have proposed will serve as a valuable resource, accelerating developments in these area.

Acknowledgments Seung-Hwan Baek was partly supported by Korea NRF grants (RS-2023-00211658, RS-2024-00438532), an IITP-ITRC grant (RS-2024-00437866), and a KEIT grant (RS-2024-0045788), funded by the Korea government (MSIT, MOTIE).

References

- [1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. Practical svbrdf capture in the frequency domain. *ACM Trans. Graph.*, 32(4):110–1, 2013. [1](#), [2](#)
- [2] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. [2](#)
- [3] Dejan Azinović, Olivier Maury, Christophe Hery, Matthias Nießner, and Justus Thies. High-res facial appearance capture from polarized smartphone images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16836–16846, 2023. [2](#)
- [4] Seung-Hwan Baek and Felix Heide. Polarimetric spatiotemporal light transport probing. *ACM Transactions on Graphics (TOG)*, 40(6):1–18, 2021. [2](#)
- [5] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–12, 2024. [1](#), [2](#)
- [6] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3982–3991, 2020. [2](#)
- [7] Charles-Félix Chabert, Per Einarsson, Andrew Jones, Bruce Lamond, Wan-Chun Ma, Sebastian Sylwan, Tim Hawkins, and Paul Debevec. Relighting human locomotion with flowed reflectance fields. In *ACM SIGGRAPH 2006 Sketches*, pages 76–es. 2006. [2](#)
- [8] Guanying Chen, Kai Han, and Kwan-Yee K Wong. Ps-fcn: A flexible learning framework for photometric stereo. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–18, 2018. [2](#), [5](#), [6](#), [7](#)
- [9] Xi Chen, Sida Peng, Dongchen Yang, Yuan Liu, Bowen Pan, Chengfei Lv, and Xiaowei Zhou. Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. In *European Conference on Computer Vision*, pages 450–467. Springer, 2025. [2](#)
- [10] Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, and Seung-Hwan Baek. Differentiable display photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11831–11840, 2024. [1](#), [2](#), [3](#), [5](#), [6](#), [7](#), [8](#)
- [11] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable point-based inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. [1](#), [2](#), [5](#), [6](#)
- [12] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable inverse rendering with interpretable basis brdfs. *arXiv preprint arXiv:2411.17994*, 2024. [2](#), [5](#), [6](#)
- [13] Edward Collett. Field guide to polarization. Spie Bellingham, WA, 2005. [3](#)
- [14] Massimiliano Corsini, Matteo Dellepiane, Federico Ponchio, and Roberto Scopigno. Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. In *Computer Graphics Forum*, pages 1755–1764. Wiley Online Library, 2009. [3](#)
- [15] Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. Deep polarization imaging for 3d shape and svbrdf acquisition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15567–15576, 2021. [2](#)
- [16] Yuto Enyo and Ko Nishino. Diffusion reflectance map: Single-image stochastic inverse rendering of illumination and reflectance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11873–11883, 2024. [2](#)
- [17] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. *arXiv preprint arXiv:2311.16043*, 2023. [2](#)
- [18] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Computer Graphics Forum*, pages 1161–1170. Wiley Online Library, 2009. [1](#), [2](#)
- [19] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *2009 IEEE 12th International Conference on Computer Vision*, pages 2335–2342. IEEE, 2009. [2](#)
- [20] Heng Guo, Jieji Ren, Feishi Wang, Boxin Shi, Mingjun Ren, and Yasuyuki Matsushita. Diligenrt: A photometric stereo dataset with quantified roughness and translucency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11810–11820, 2024. [2](#)
- [21] Zexin He, Tengfei Wang, Xin Huang, Xingang Pan, and Ziwei Liu. Neural lightrig: Unlocking accurate object normal and material estimation with multi-light diffusion. *arXiv preprint arXiv:2412.09593*, 2024. [2](#)
- [22] George H Heilmeyer, Louis A Zanon, and Lucian A Barton. Dynamic scattering: A new electrooptic effect in certain classes of nematic liquid crystals. *Proceedings of the IEEE*, 56(7):1162–1171, 1968. [2](#)
- [23] Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C Sankaranarayanan. Reflectance capture using univariate sampling of brdfs. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5362–5370, 2017. [2](#)
- [24] 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. [2](#)
- [25] Satoshi Ikehata. Ps-transformer: Learning sparse photometric stereo network using self-attention mechanism. *arXiv preprint arXiv:2211.11386*, 2022. [5](#), [6](#)
- [26] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12591–12600, 2022. [2](#), [5](#), [6](#)
- [27] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proceedings of the IEEE/CVF Con-*

- ference on Computer Vision and Pattern Recognition, pages 13198–13207, 2023. 2, 5, 6, 7
- [28] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. <https://mitsuba-renderer.org>. 3
- [29] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanaes. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 406–413, 2014. 2
- [30] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. 2024. 2
- [31] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensor: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2023. 2
- [32] Christos Kampouris, Stefanos Zafeiriou, and Abhijeet Ghosh. Diffuse-specular separation using binary spherical gradient illumination. *EGSR (EI&I)*, 1(10), 2018. 1, 2, 7
- [33] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Nereo: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 2
- [34] Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott Wu, Jiajun Wu, et al. Stanford-orb: a real-world 3d object inverse rendering benchmark. 2023. 2
- [35] Alexandros Lattas, Yiming Lin, Jayanth Kannan, Ekin Ozturk, Luca Filipi, Giuseppe Claudio Guarnera, Gaurav Chawla, and Abhijeet Ghosh. Practical and scalable desktop-based high-quality facial capture. In *European Conference on Computer Vision*, pages 522–537. Springer, 2022. 1, 2, 7
- [36] Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)*, 22(2):234–257, 2003. 1
- [37] Junxuan Li and Hongdong Li. Neural reflectance for shape recovery with shadow handling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16221–16230, 2022. 2, 5, 6, 8
- [38] Junxuan Li and Hongdong Li. Self-calibrating photometric stereo by neural inverse rendering. In *European Conference on Computer Vision*, pages 166–183. Springer, 2022. 5, 6
- [39] 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. In *IEEE Transactions on Image Processing*, pages 29:4159–4173, 2020. 2
- [40] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (TOG)*, 37(6):1–11, 2018. 2
- [41] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2475–2484, 2020. 2
- [42] Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Zhi-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, et al. Diffusionrenderer: Neural inverse and forward rendering with video diffusion models. *arXiv preprint arXiv:2501.18590*, 2025. 2
- [43] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. 2024. 2
- [44] Daniel Lichy, Jiaye Wu, Soumyadip Sengupta, and David W Jacobs. Shape and material capture at home. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6123–6133, 2021. 2
- [45] Lahav Lipson, Zachary Teed, and Jia Deng. Raft-stereo: Multilevel recurrent field transforms for stereo matching. In *2021 International Conference on 3D Vision (3DV)*, pages 218–227. IEEE, 2021. 5
- [46] Yehonathan Litman, Or Patashnik, Kangle Deng, Aviral Agrawal, Rushikesh Zawat, Fernando De la Torre, and Shubham Tulsiani. Materialfusion: Enhancing inverse rendering with material diffusion priors. *3DV 2025*, 2024. 2
- [47] 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. 2
- [48] Yuan Liu, Peng Wang, Cheng Lin, Xiaoxiao Long, Jiepeng Wang, Lingjie Liu, Taku Komura, and Wenping Wang. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images. 2023. 2
- [49] Linjie Lyu, Ayush Tewari, Marc Habermann, Shunsuke Saito, Michael Zollhöfer, Thomas Leimkühler, and Christian Theobalt. Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 2
- [50] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques*, 9(10):2, 2007. 1, 2, 6, 7
- [51] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. Lucas: A dataset for near-field point light source photometric stereo. *arXiv preprint arXiv:2104.13135*, 2021. 2
- [52] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (TOG)*, 37(6):1–12, 2018. 2
- [53] Jannik Boll Nielsen, Henrik Wann Jensen, and Ravi Ramamoorthi. On optimal, minimal brdf sampling for reflectance acquisition. *ACM Transactions on Graphics (TOG)*, 34(6):1–11, 2015. 5
- [54] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 117–128, 2001. 1

- [55] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. Diligent102: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12581–12590, 2022. 2
- [56] J  r  my Riviere, Pieter Peers, and Abhijeet Ghosh. Mobile surface reflectometry. In *ACM SIGGRAPH 2014 Posters*, pages 1–1. 2014. 2
- [57] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Nerf for outdoor scene relighting. In *European Conference on Computer Vision*, pages 615–631. Springer, 2022. 2
- [58] Szymon M Rusinkiewicz. A new change of variables for efficient brdf representation. In *Rendering Techniques’ 98: Proceedings of the Eurographics Workshop in Vienna, Austria, June 29–July 1, 1998*, pages 11–22. Springer, 1998. 5
- [59] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 130–141, 2024. 1
- [60] Shen Sang and Manmohan Chandraker. Single-shot neural relighting and svbrdf estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX* 16, pages 85–101. Springer, 2020. 2
- [61] Sam Sartor and Pieter Peers. Matfusion: a generative diffusion model for svbrdf capture. In *SIGGRAPH Asia 2023 Conference Papers*, pages 1–10, 2023. 2
- [62] Imari Sato, Takahiro Okabe, Yoichi Sato, and Katsushi Ikeuchi. Appearance sampling for obtaining a set of basis images for variable illumination. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 800–807. IEEE, 2003. 1, 2
- [63] Soumyadip Sengupta, Jinwei Gu, Kihwan Kim, Guilin Liu, David W Jacobs, and Jan Kautz. Neural inverse rendering of an indoor scene from a single image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8598–8607, 2019. 2
- [64] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3707–3716, 2016. 2
- [65] 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. 2
- [66] Feishi Wang, Jieji Ren, Heng Guo, Mingjun Ren, and Boxin Shi. Diligent-pi: Photometric stereo for planar surfaces with rich details-benchmark dataset and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9477–9487, 2023. 2
- [67] Haoyuan Wang, Wenbo Hu, Lei Zhu, and Rynson W.H. Lau. Inverse rendering of glossy objects via the neural plenoptic function and radiance fields. In *CVPR*, 2024. 2
- [68] Zian Wang, Jonah Philion, Sanja Fidler, and Jan Kautz. Learning indoor inverse rendering with 3d spatially-varying lighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12538–12547, 2021. 2
- [69] Xin Wei, Guojun Chen, Yue Dong, Stephen Lin, and Xin Tong. Object-based illumination estimation with rendering-aware neural networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV* 16, pages 380–396. Springer, 2020. 2
- [70] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 5, 6, 7, 8
- [71] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *IEEE transactions on pattern analysis and machine intelligence*, 37(1):67–79, 2014. 2
- [72] Jing Yang, Pratusha Bhuvana Prasad, Qing Zhang, and Yajie Zhao. Acquisition of spatially-varying reflectance and surface normals via polarized reflectance fields. *arXiv preprint arXiv:2412.09772*, 2024. 2
- [73] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *European Conference on Computer Vision*, pages 266–284. Springer, 2022. 2
- [74] Ye Yu and William AP Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2019. 2
- [75] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 2
- [76] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021. 2
- [77] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5565–5574, 2022. 2
- [78] Lianghao Zhang, Fangzhou Gao, Li Wang, Minjing Yu, Jiamin Cheng, and Jiawan Zhang. Deep svbrdf estimation from single image under learned planar lighting. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–11, 2023. 1, 2
- [79] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 2

- [80] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022. [2](#)
- [81] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11): 1330–1334, 2000. [3](#)
- [82] Rui Zhu, Zhengqin Li, Janarbek Matai, Fatih Porikli, and Manmohan Chandraker. Irisformer: Dense vision transformers for single-image inverse rendering in indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2822–2831, 2022. [2](#)