A Real-world Display Inverse Rendering Dataset
-Supplemental Document-

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

In this supplemental document, we provide additional results and details in support of our findings in the main manuscript.

Contents

1. Details on Imaging System

Lighting Module For our display module, we employ a commercially available large curved LCD monitor (Samsung Odyssey Ark). This display features a 55-inch liquid-crystal panel with a resolution of 2160×3840 pixels and a peak brightness of 600 cd/m^2. Owing to the polarization-sensitive optical components inherent in LCD technology, each pixel emits horizontally linearly-polarized light spanning the trichromatic RGB spectrum.

Capture Module We further utilize two polarization camera (FLIR BFS-U3-51S5PC-C) that integrates on-sensor linear polarization filters oriented at four distinct angles. Consequently, the camera records four polarized light intensities at $0°$, $45°$, $90°$, and $135°$, denoted as $I_{0°}$, $I_{45°}$, $I_{90°}$, and $I_{135°}$, respectively. The sensor captures a linear raw signal, to which we apply a series of linear image processing steps, including black level subtraction, demosaicing, and undistortion. All images in the dataset were captured using a fixed shutter time of 440 ms under a single-exposure setting. As a cost-effective alternative to a high-end polarization camera, one may adopt a conventional camera augmented with a linear-polarization film. Aligning the film's polarization axis perpendicular to that of the display facilitates the capture of diffuse image components.

Device Control To manage the display outputs and control the polarization camera, we employ the PyGame and PySpin libraries, respectively. The devices are interfaced with a desktop computer via an HDMI cable and a USB3 connection, with software synchronization ensuring coordinated operation between the display and the camera.

Radiometric Calibration The monitor's emitted radiance does not exhibit a linear correlation with the pixel values of the display pattern. To correct for this nonlinearity, we capture images of gray patches on a color checker across various intensity levels. An exponential function is then fitted to the measured intensities as a function of the monitor's pixel values for each individual color channel.

Light attenuation calibration To calibrate the light attenuation effect as a function of the distance between the object and the light source, a color checker was positioned in front of the imaging system and an OLAT image was captured as shown in Figure S1.(a). For each light position, the brightness variation of the color patches is measured, and the parameters a, b, and c in the following equation are fitted:

$$\frac{1}{a + b \times distance^2} + c \tag{1}$$

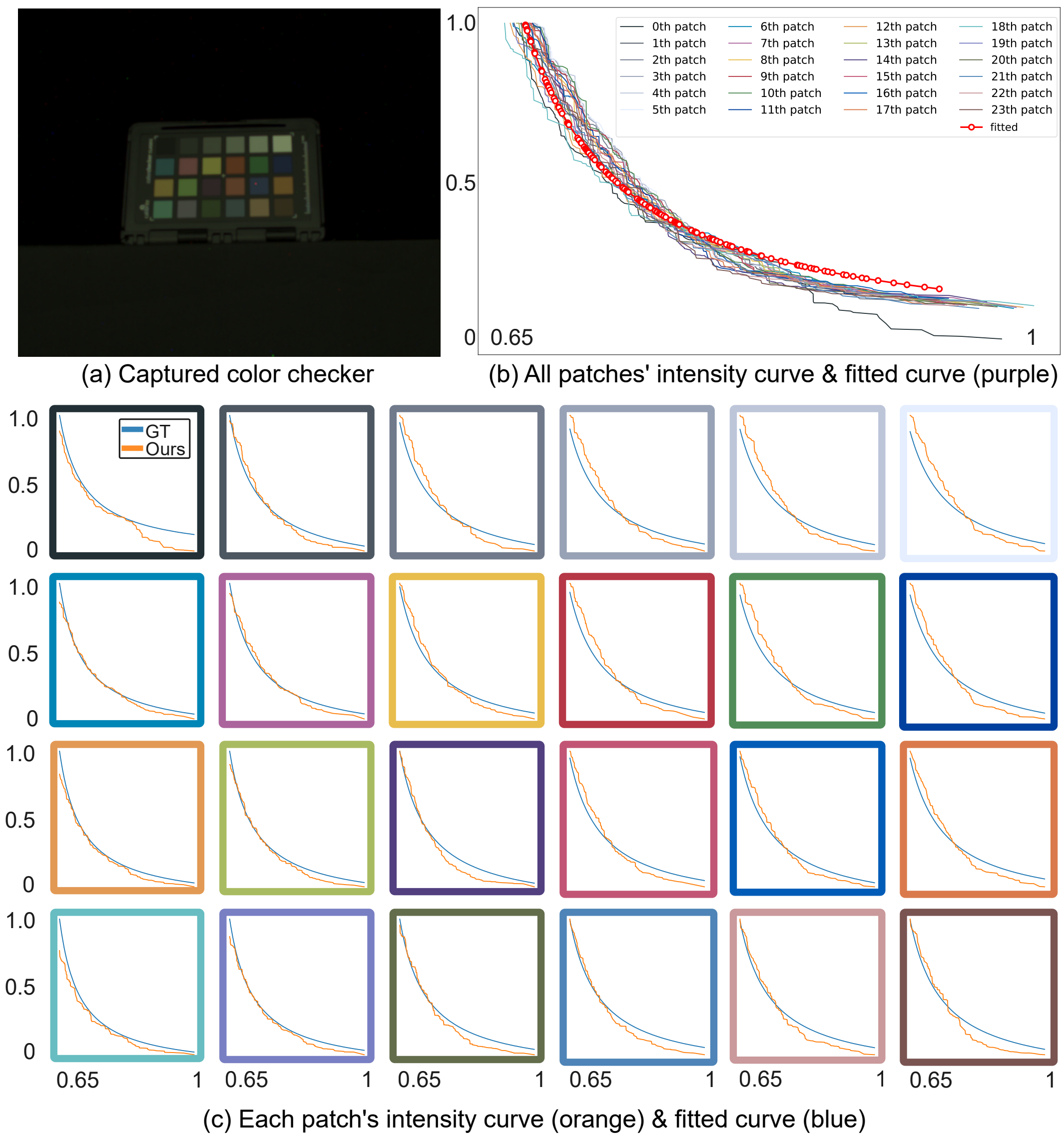

Figure S1. **Light fall-off calibration.** The parameters are fitted by analyzing the variation in pixel intensity as a function of the distance to the light source.

Figure S1. (b) and (c) shows curves fitted to captured intensity variation.

2. Details on Scan System

We employed the EinScan SP V2 scanner to acquire mesh data of three-dimensional objects. The EinScan SP V2 is a structured light active stereo imaging device that scans objects placed on a turntable. Device calibration is performed by positioning a checkerboard on the turntable. During a single rotation, eight images are captured to generate data points, which

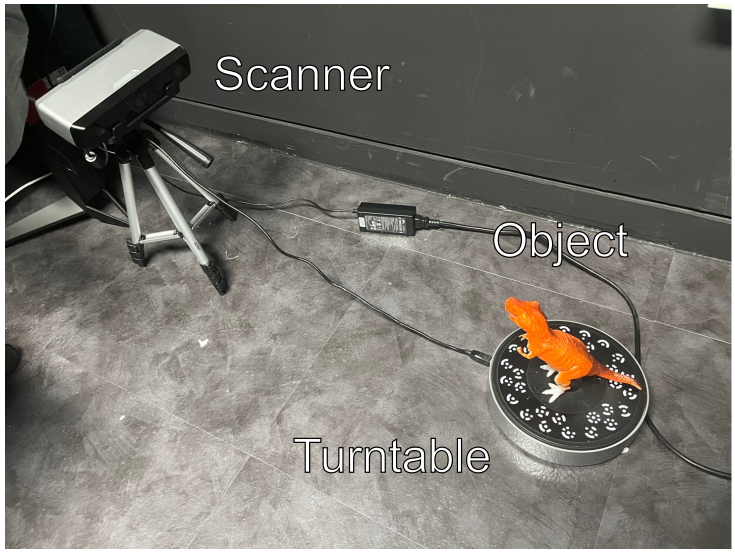

Figure S2. **EinScan SP V2.** An object is placed on turntable, and the scanner is looking at the object.

are then registered to reconstruct the object's geometry. The system maintains an accuracy within a tolerance of 0.05 mm. The scan scenario is shown in Figure S2.

Mesh-raster alignment The scanned mesh is aligned with the captured images to extract the ground truth depth map, normal map, and mask. The registration between the images and the mesh is performed in a semi-automatic and semi-manual manner using the mutual information methods [7] available in MeshLab. The mesh pose obtained through this alignment is then imported into the differentiable renderer, Mitsuba3, to render the depth map, normal map, and mask.

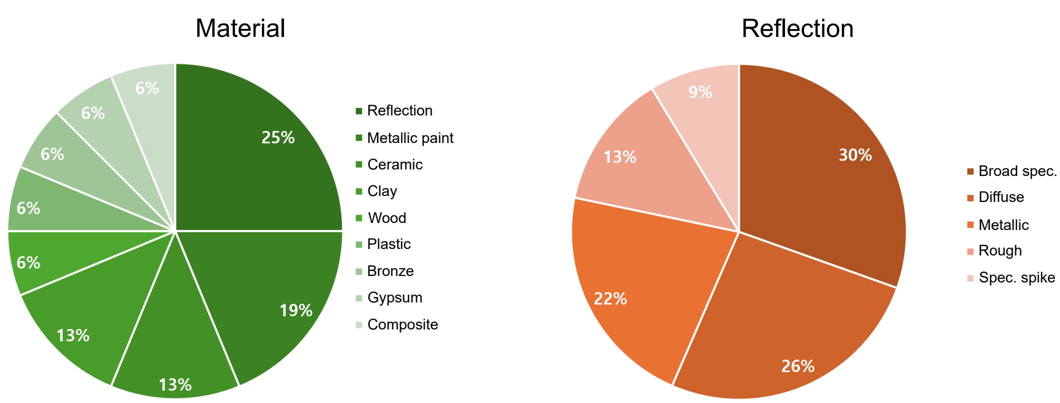

Figure S3. **Data statistics.** We conducted a statistical analysis of 16 objects based on two aspects: their material properties and surface reflectance characteristics.

3. Dataset

Comparison with Existing datasets. Table S1 presents the existing inverse rendering datasets. Some of these datasets were originally developed for photometric stereo, yet they have also been widely adopted for inverse rendering. Each dataset exhibits unique characteristics based on whether ground truth geometry is provided, the type of illumination and imaging system used, the available lighting calibration information, the number of captured objects, the number of views, and the number of light sources employed. These factors determine the suitability of each dataset for different research settings.

Our dataset is specifically designed for inverse rendering in a display-camera system. It provides ground truth geometry obtained via scanning, three-dimensional position information for a total of 144 superpixels from an LCD monitor, and

Dataset	Ground-truth geometry	Imaging system	Lighting calibration	Number of objects	Number of views	Number of lights
Blobby and Sculpture [3]	✓	Synthetic	Direction	18	1	64
CyclePS [11]	✓	Synthetic	Direction	45	1	1300
PS-Wild [12]	✓	Synthetic	✗	410	1	10
Gourd and Apple [1]	✗	Light rig	Direction	3	1	112
Harvard [28]	✗	Light rig	Direction	7	1	20
DTU [14]	✓	Robot	✗	80	64	7
MIT-intrinsic [8]	✗	Commodity camera	✗	20	1	10
NeROIC [15]	✗	Commodity camera	Env. map	3	40	6
NeRF-OSR [24]	✗	Commodity camera	Env. map	8	3240	110
Stanford ORB [16]	✓	Commodity camera	Env. map	14	70	7
ReNe [26]	✗	Robot	Position	20	50	40
Light Stage Data Gallery [2]	✗	Light stage	Direction	9	1	253
Open Illumination [20]	✗	Light stage	Position	64	72	154
Polar-lightstage [29]	Pseudo	Light stage	Direction	26	8	346
LUCES [21]	✓	Light rig	Position	14	1	52
DiLiGenT [25]	✓	Light rig	Position	10	1	96
DiLiGenT102 [23]	✓	Gantry	Direction	100	1	100
DiLiGenT-PI [27]	✓	Gantry	Direction	34	1	100
DiLiGenRT [9]	✓	Gantry	Direction	54	1	100
DiLiGenTMV [18]	✓	Studio/desktop scanner	Direction	5	20	96
Ours	✓	**Display and camera**	**Position**	**16**	**2**	**144**

Table S1. **Inverse rendering datasets.** We classified the datasets related to inverse rendering into their respective categories and organized them into a table.

stereo views of 16 objects. The stereo camera system is comprised of a polarization camera, which enables the capture of polarization information of the objects. Figure S3 shows the statistics of our dataset's material. In dataset described in this paper, objects are placed at 50 cm from the cameras, we will release additional scene with an object which are placed at multiple distances.

4. Additional Details on the Proposed Baseline Model

In this section, we introduce a simple yet effective baseline model for display-based inverse rendering. The proposed model is designed to handle input images captured under M arbitrary display patterns with intrinsic backlighting, while addressing the challenges posed by limited angular sampling and modeling the effects of near-field lighting.

Initialization In the initialization step, we estimate surface normals $\mathbf{n}$ using analytical photometric stereo [4], which operates on M multiplexed images. A depth map is then estimated using RAFT-Stereo [19], applied to the average of stereo image pairs captured under multiple patterns. Given this initial geometry, we proceed to optimize the per-pixel reflectance and normal map.

Image Formation When a scene point is illuminated by a display pattern $\mathcal{L}$, the observed image intensity is modeled as:

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

where L_i denotes the RGB intensity of the i-th superpixel composing the display pattern $\mathcal{L} = \{L_1, \ldots, L_N\}$, f is the BRDF, $\mathbf{n}$ is the surface normal, $\mathbf{i}$ is the incident direction from the i-th superpixel, $\mathbf{o}$ is the outgoing view direction, and d_i is the distance between the i-th superpixel and the scene point. The function $\text{clip}(\cdot)$ accounts for camera dynamic range clipping, and ϵ models Gaussian noise.

Simulation for Arbitrary Display Patterns To simulate the image under an arbitrary display pattern $\mathcal{P} = \{P_1, \cdots, P_N\}$ based on the equation 2, we model the i-th display superpixel intensity given the corresponding RGB pattern value we set to display P_i as

$$L_i = s(P_i + B_i)^\gamma, \tag{3}$$

where s is a global scalar, γ is the non-linear mapping exponent, and B_i is the corresponding spatially-varying backlight intensity. Then, a scene illuminated by an arbitrary display patten can be described, using Equation (2) and Equation (3), as:

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

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. We represent the captured or rendered image under the i-th OLAT illumination, I_i as:

$$I_i = f(\mathbf{i}, \mathbf{o})(\mathbf{n} \cdot \mathbf{i})\frac{1}{d_i^2} \tag{5}$$

For modeling near-field effects, spatially-varying lighting direction $\mathbf{i}$ and spatially-varying intensity falloff scalar $\frac{1}{d_i^2}$ are computed by using calibrated superpixel positions and a depth map estimated in initialization step.

Reflectance Model Limited angular sampling in display-camera systems presents challenges for accurate reflectance estimation. Following previous approaches [5, 17, 22], we adopt a basis BRDF representation to regularize the underdetermined nature of the problem. Specifically, we model the SVBRDF as a weighted sum of J basis BRDFs.

Let w_j denote the coefficient for the j-th basis BRDF. The overall SVBRDF f is expressed as:

$$f(\mathbf{i}, \mathbf{o}) = \sum_{j=1}^{J} w_j f_j(\mathbf{i}, \mathbf{o}). \tag{6}$$

Each basis BRDF f_j is parameterized by a diffuse albedo $\rho_j^d \in \mathbb{R}^3$, roughness $\sigma_j \in \mathbb{R}^3$, and specular albedo $\rho_j^s \in \mathbb{R}^3$, using the Cook-Torrance reflectance model [6]:

$$f_j(\mathbf{i}, \mathbf{o}) = \rho_j^d + \rho_j^s \frac{D(\mathbf{h}; \sigma_j) F(\mathbf{o}, \mathbf{h}) G(\mathbf{i}, \mathbf{o}, \mathbf{n}; \sigma_j)}{4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})}, \tag{7}$$

where $\mathbf{h}$ is the half-vector between $\mathbf{i}$ and $\mathbf{o}$, D is the normal distribution function, F is the Fresnel term, and G is the geometric attenuation factor.

Optimization At the initialization stage, we performed photometric stereo to estimate surface normals and obtain a pseudo diffuse image. In the HSV color space, we apply K-means clustering on the hue and saturation channels of this pseudo diffuse image to obtain J clusters. Each cluster is converted into a one-hot encoded representation, which serves as an initial estimate of J weight maps. The centroid color of each cluster is assigned as the diffuse albedo of a basis BRDF, initializing the corresponding coefficient. Both roughness and specular albedo are uniformly initialized to 0.5.

We then iteratively optimize the per-pixel surface normals, basis coefficients, and basis BRDFs by minimizing the RMSE loss between the M captured images and the corresponding rendered images. To mitigate noisy per-pixel updates, we incorporate a total-variation (TV) regularization term during optimization.

5. Additional Experiments

Comparison with Feed-forward Inverse Rendering Models Figure S4 shows additional results for transformer-based SDM-UniPS [13] and diffusion-based Neural LightRig [10]. These learned methods show lower-quality results compared to our baseline.

Environmental Relighting The optimized scene representation enables relighting under a novel environment map as shown in Fig. S5.

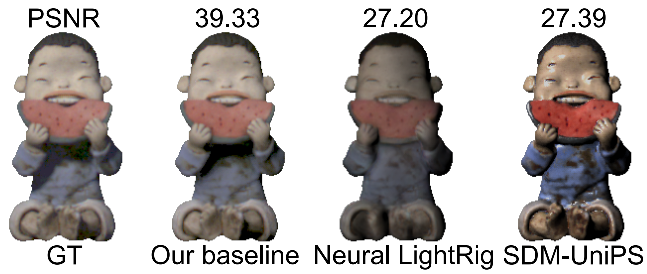

Figure S4. Evaluation of learned methods.

Figure S5. Relighting results with an environment map.

Robustness of Camera Positional Errors Our method is robust to camera positional error, obtaining relighting PSNR 38.74 and normal MAE 28.26 for the 5 cm-displaced camera position.

Robustness without Stereo Imaging Even without stereo imaging, our baseline, using uniform depth, outperforms previous methods, with relighting PSNR 38.8 and normal MAE 28.29, as shown in Table 3. We clarify this explicitly for a fair comparison.

Analysis of Display Backlight Figure S6 (a) and (b) show images of the display when the superpixel intensity is set to 1 (maximum) and 0.5 (half), respectively. The difference image shows that the backlight remains constant with different intensities in signal inputs.

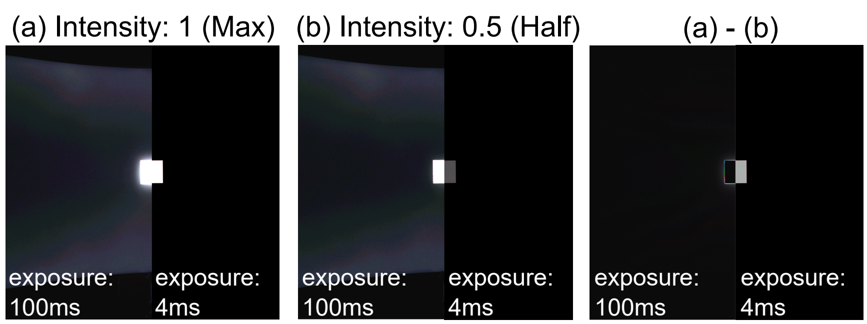

Figure S6. Backlight is invariant to super-pixel intensity: (a) 1 and (b) 0.5. The difference image shows consistent backlight. There is lens flare around the saturated areas.

Analysis of Point-light Assumption as a Superpixel Assuming a superpixel as a point light source introduces a trade-off between blurred reflectance and noise caused by limited exposure. Although the design choice in the baseline deviates from the ideal point light assumption, Figure S7 demonstrates that the size of the superpixel has a minimal impact on the specular lobe.

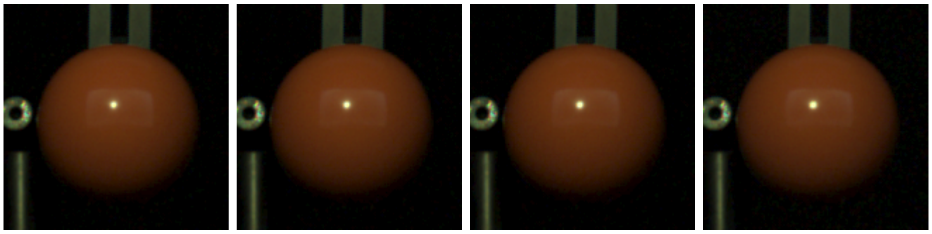

Figure S7. Images of a glossy sphere with superpixels of different sizes. Numbers below indicate pixels per superpixel.

6. Photometric Stereo Results

The following are the surface normal reconstruction results for all the photometric stereo methods we evaluated.

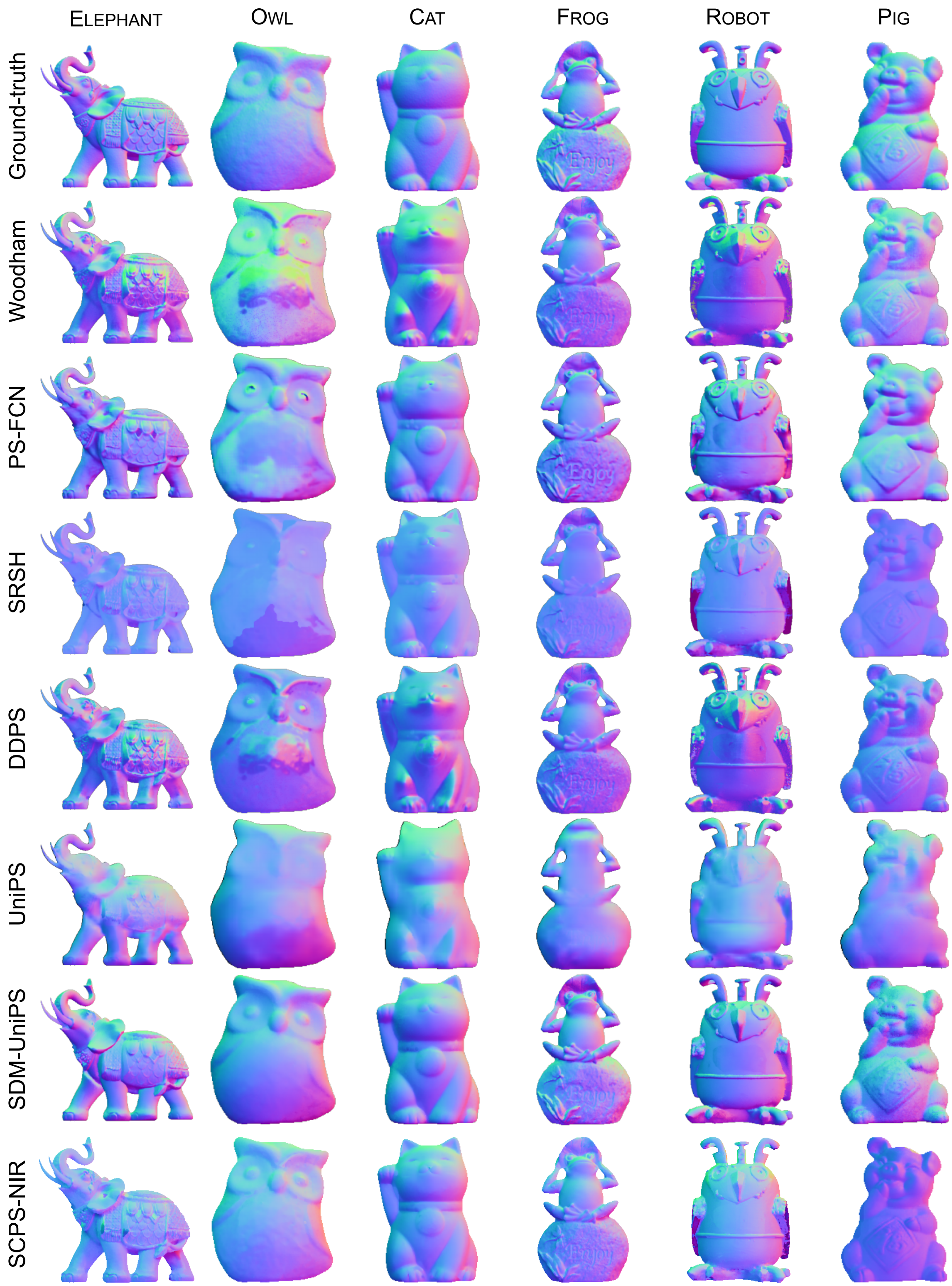

Figure S8. **Reconstructed normal of evaluated methods.** We visualize reconstructed normals of ELEPHANT, OWL, CAT, FROG, ROBOT, PIG

Figure S9. **Reconstructed normal of evaluated methods.** We visualize reconstructed normals of CHICKEN, GIRL BOY, NEFERTITI, TREX

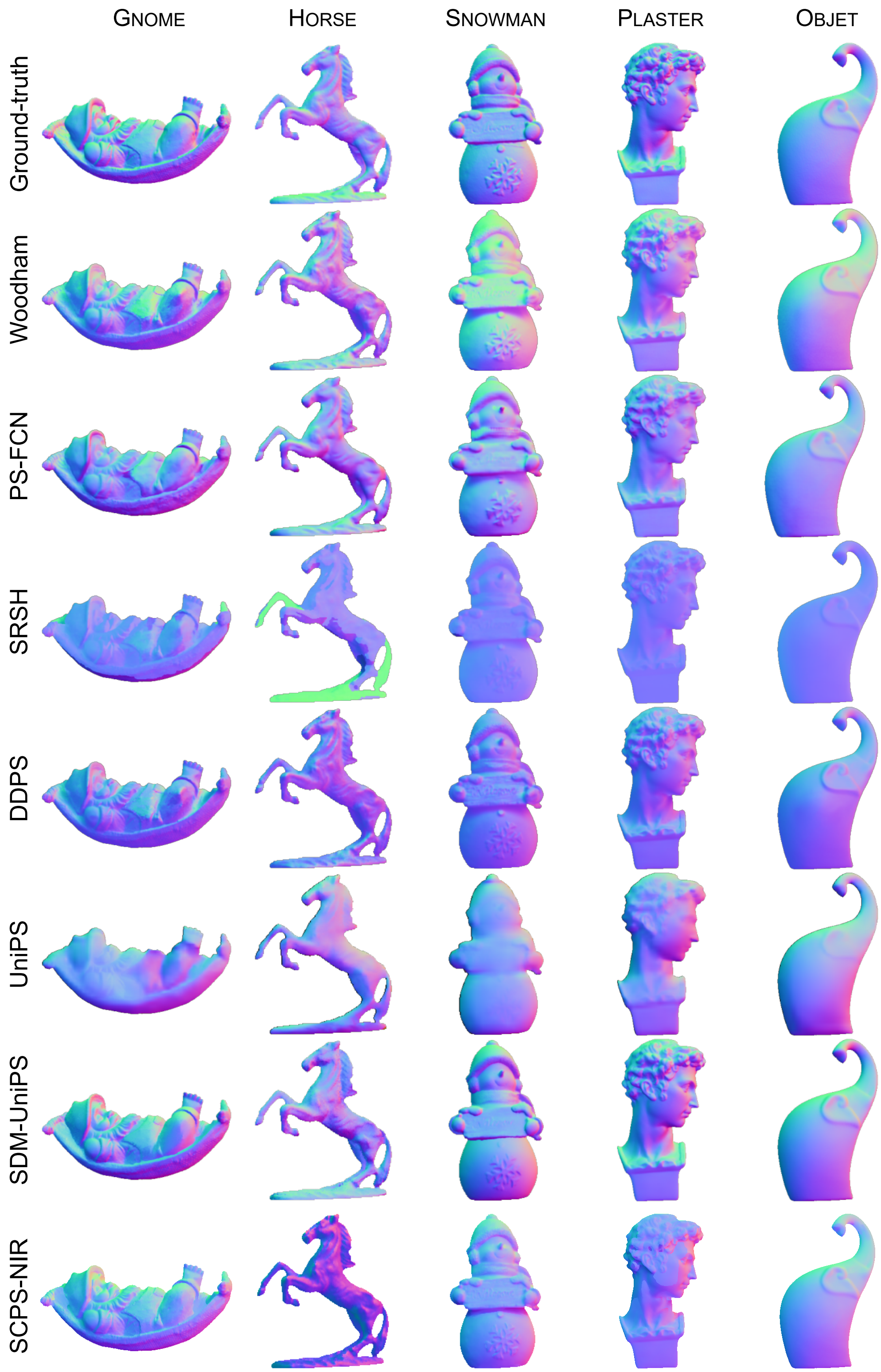

Figure S10. **Reconstructed normal of evaluated methods.** We visualize reconstructed normals of GNOME HORSE, SNOWMAN, PLASTER, OBJET

References

[1] 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. 4

[2] 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. 4

[3] 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. 4

[4] 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. 4

[5] 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. 5

[6] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)*, 1(1): 7–24, 1982. 5

[7] 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

[8] 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. 4

[9] 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. 4

[10] 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. 5

[11] 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. 4

[12] 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. 4

[13] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13198–13207, 2023. 5

[14] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 406–413, 2014. 4

[15] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: Neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 4

[16] 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. 4

[17] 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. 5

[18] 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. 4

[19] 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. 4

[20] 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. 4

[21] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. Luces: A dataset for near-field point light source photometric stereo. *arXiv preprint arXiv:2104.13135*, 2021. 4

[22] 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. 5

[23] 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. 4

[24] 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. 4

[25] 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. 4

[26] 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. 4

[27] 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. 4

[28] 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. 4

[29] 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. 4