Differentiable Display Photometric Stereo

Seokjun Choi[†]   Seungwoo Yoon[†]   Giljoo Nam[*]   Seungyong Lee[†]   Seung-Hwan Baek[†]

[†] POSTECH   [*] Meta Reality Labs

Figure 1. We propose differentiable display photometric stereo, a method that facilitates (a) the learning of display patterns, enabling high-quality reconstruction of surface normals using (b) a monitor and a camera. (c) Capturing a scene with the learned patterns allows for estimating (d) high-quality surface normals.

Abstract

Photometric stereo leverages variations in illumination conditions to reconstruct surface normals. Display photometric stereo, which employs a conventional monitor as an illumination source, has the potential to overcome limitations often encountered in bulky and difficult-to-use conventional setups. In this paper, we present differentiable display photometric stereo (DDPS), addressing an often overlooked challenge in display photometric stereo: the design of display patterns. Departing from using heuristic display patterns, DDPS learns the display patterns that yield accurate normal reconstruction for a target system in an end-to-end manner. To this end, we propose a differentiable framework that couples basis-illumination image formation with analytic photometric-stereo reconstruction. The differentiable framework facilitates the effective learning of display patterns via auto-differentiation. Also, for training supervision, we propose to use 3D printing for creating a real-world training dataset, enabling accurate reconstruction on the target real-world setup. Finally, we exploit that conventional LCD monitors emit polarized light, which allows for the optical separation of diffuse and specular reflections when combined with a polarization camera, leading to accurate normal reconstruction. Extensive evaluation of DDPS shows improved normal-reconstruction accuracy compared to heuristic patterns and demonstrates compelling properties such as robustness to pattern initialization, calibration errors, and simplifications in image formation and reconstruction.

1. Introduction

Reconstructing high-quality surface normals is pivotal in computer vision and graphics for 3D reconstruction [32, 40], relighting [36, 39], and inverse rendering [45, 52]. Among various techniques, photometric stereo [50] leverages the intensity variation of a scene point under varied illumination conditions to reconstruct normals. Photometric stereo finds its application in various imaging systems including light stages [29, 35, 49, 56], handheld-flash cameras [3, 10, 37, 52], and display-camera systems [1, 28, 46].

Display photometric stereo uses monitors and cameras as a versatile and accessible system that can be conveniently placed on a desk [1, 28, 46]. Producing diverse illumination conditions can be simply achieved by displaying multiple patterns using pixels on the display as programmable point light sources. This convenient and intricate modulation of illumination conditions significantly enlarges the design space of illumination patterns for display photometric stereo. Nevertheless, existing approaches often rely on heuristic display patterns, resulting in sub-optimal reconstruction quality.

In this paper, to exploit the large design space of illumination patterns in display photometric stereo, we propose differentiable display photometric stereo (DDPS). The key idea is to learn display patterns that lead to improved reconstruction of surface normals for a target system in an end-to-end manner. To this end, we introduce a differentiable framework that combines basis-illumination image formation and an optimization-based photometric stereo method. This enables effective pattern learning by directly optimiz-

ing the display patterns via auto-differentiation. To compute the normal-reconstruction loss for backpropagation, we propose the use of 3D printing for creating a real-world training dataset with known geometry. Combined with the basis-illumination image formation, using the 3D-printed dataset allows for efficient and realistic simulation of relit images during end-to-end optimization. In addition, we leverage that conventional LCD monitors emit polarized light. Thus, using a polarization camera, we can optically remove specular reflection that often deteriorates photometric-stereo reconstruction.

Extensive evaluation of DDPS on diverse objects shows that using the learned patterns significantly improves normal accuracy compared to using heuristic patterns. Moreover, DDPS exhibits robustness to pattern initialization, calibration error, and simplifications in image formation and reconstruction, promising its practical applicability. We will release code and data upon acceptance.

In summary, our contributions are as follows:

- Departing from using heuristic patterns for display photometric stereo, we directly learn display patterns that lead to high-quality normal reconstruction for display photometric stereo in an end-to-end manner.
- For DDPS, we propose the differentiable framework consisting of basis-illumination image formation and analytic photometric-stereo reconstruction, the use of 3D-printed objects for a training dataset, and using a polarized LCD and a polarization camera.
- We perform extensive experiments, demonstrating the effectiveness of learned patterns, which outperforms heuristic patterns, and the robustness of DDPS against various factors including pattern initialization and calibration errors.

2. Related Work

Illumination Patterns for Photometric Stereo One crucial but often overlooked problem in photometric stereo is deciding on illumination patterns, which is a set of intensity distributions of light sources, so that accurate surface normals can be reconstructed. A standard option is the one-light-at-a-time (OLAT) pattern that turns on each light source at its maximum intensity one by one [47, 54]. OLAT is typically employed when the intensity of each light source is sufficient enough to provide light energy to be detected by a camera sensor without significant noise, such as in light stages [13]. Extending OLAT patterns with a group of neighboring light sources increases light energy, reducing measurement noise [8, 48]. Spherical gradient illumination, designed for light stages, enables rapid acquisition of high-fidelity normals by exploiting polarization [32], color [35], or both [16]. Complementary patterns, where half of the lights are turned on and the other half off for each three-dimensional axis, also enable rapid reconstruc-

tion when applied to light stages and monitors [24, 28]. Wenger et al.[48] propose random binary patterns that provide high light efficiency. However, the aforementioned illumination patterns are heuristically designed, which often result in sub-optimal reconstruction accuracy and capture efficiency. For a specific display-camera system, it is challenging to determine which display patterns would provide high-quality photometric stereo. DDPS departs from using heuristic patterns and instead learns display patterns for robust photometric stereo.

Illumination-optimized Systems Recent studies have investigated optimizing illumination designs for inverse rendering [25, 26, 33, 53], active-stereo depth imaging [5], and holographic display [41]. These approaches typically rely on dedicated illumination modules such as LED arrays, diffractive optical elements, and spatial light modulators. In contrast, DDPS exploits ubiquitous LCD devices and their polarization state for display illumination. Also, DDPS directly applies normal reconstruction loss to illumination learning using the 3D-printed dataset, unlike previous method that employ intermediary metrics, such as lumitexel prediction [25, 26, 33]. Zhang et al.[53] optimize a single illumination pattern for inverse rendering, only targeting planar objects. In contrast, DDPS reconstructs surface normals of general objects with complex shapes and capable of optimizing multiple illumination patterns.

Imaging Systems for Photometric Stereo Many photometric stereo systems have been proposed, including moving a point light source, such as a flashlight on a mobile phone [20, 43], a DSLR camera flash [14, 17]. and installing multiple point light sources in light stage systems [29, 35] and other custom devices [19, 24–26, 33]. Display photometric stereo exploits off-the-shelf displays as cost-effective, versatile active-illumination modules capable of generating spatially-varying trichromatic intensity variation [1, 11, 15, 18, 28, 31, 38]. Lattas et al. [28] demonstrated facial capture using multiple off-the-shelf monitors and multi-view cameras with trichromatic complementary illumination. In our paper, we build on display photometric stereo and propose to learn the display patterns to obtain high-quality normal reconstruction.

Photometric Stereo Dataset Many datasets have been proposed for photometric stereo [2, 30, 34, 42, 51] for evaluation or training photometric stereo methods. Early datasets often relied on synthetic rendering [9, 44]. However, using synthetic datasets for a real-world target system requires highly accurate calibration of the target photometric-stereo system, its replication on the rendering, and physically realistic light-transport simulation. Real-world datasets relax these constraints by capturing real-
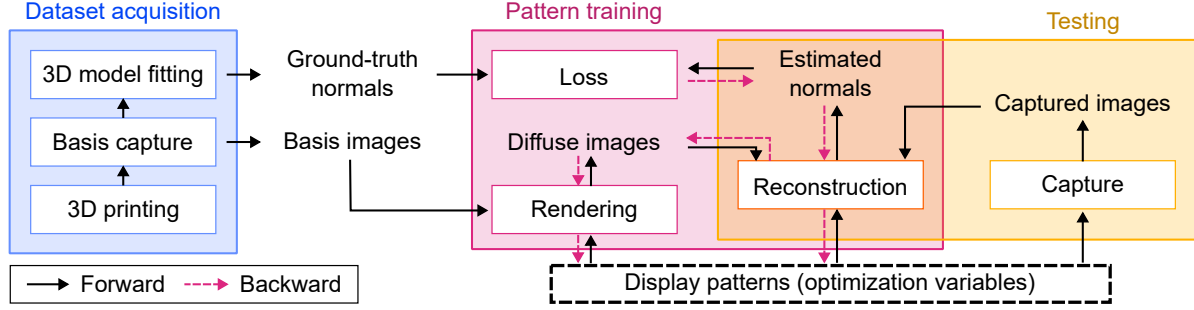
Figure 2. **Overview of DDPS.** DDPS consists of three stages: dataset acquisition, pattern training, and testing.

world objects under multiple point light sources [30, 42]. However, acquiring ground-truth normals of real-world objects often demands using high-quality commercial 3D scanners. In contrast, DDPS uses 3D printing to obtain objects with known geometry. Using the 3D-printed dataset combined with 3D model fitting allows for effectively supervising the pattern learning in an end-to-end manner.

3. Overview

DDPS consists of three stages as shown in Figure 2: dataset acquisition, pattern training, and testing. First, in the dataset-acquisition stage, we 3D-print various objects, capture their basis-illumination images with a target display-camera setup, and obtain ground-truth surface normal maps via 3D model fitting. Then, in the pattern-training stage, we learn the optimal display patterns that lead to high-quality normal reconstruction using the real-world training dataset. To this end, we develop the differentiable framework of basis-illumination image formation and analytic photometric-stereo reconstructor. In the testing phase, we capture diverse real-world objects under the patterns learned on our training dataset and reconstruct surface normals using the photometric-stereo reconstructor.

4. Polarimetric Display-Camera Imaging

Polarimetric Light Transport We first describe our imaging system, shown in Figure 3(a). We use off-the-shelf components: a curved 4K LCD monitor and a polarization camera. Linearly-polarized light is emitted from the LCD monitor, due to the polarization-based working principle of LCDs [12]. The light interacts with a real-world scene, generating both specular and diffuse reflections. The specular reflection tends to maintain the polarization state of light, while diffuse reflection becomes unpolarized [7]. The polarization camera then captures the reflected light at four different linear-polarization angles: $\{I_\theta\}_{\theta \in \{0°, 45°, 90°, 135°\}}$. We then convert the captured raw intensities $\{I_\theta\}$ into the linear Stokes-vector elements [12]:

$$s_0 = \frac{\sum_\theta I_\theta}{2}, \; s_1 = I_{0°} - I_{90°}, \; s_2 = 2I_{45°} - I_{0°}, \quad (1)$$

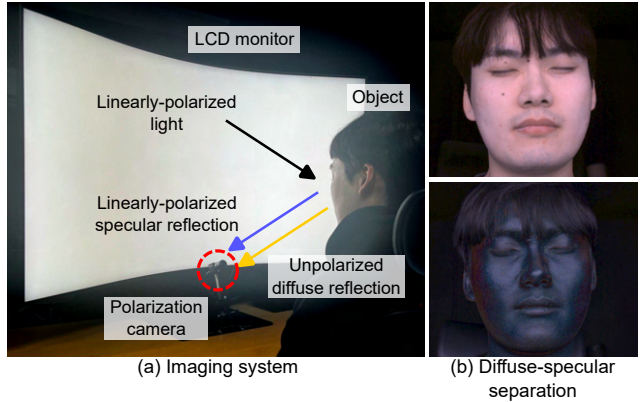(a) Imaging system   (b) Diffuse-specular separation

Figure 3. **Polarimetric imaging system.** (a) Imaging system consisting of an LCD monitor and a polarization camera. Decomposed (b) diffuse image and specular image using linearly-polarized light emitted from the monitor.

and compute the diffuse reflection I_{diffuse} and specular reflection I_{specular}: $I_{\text{specular}} = \sqrt{s_1^2 + s_2^2}$, $I_{\text{diffuse}} = s_0 - I_{\text{specular}}$. Hereafter, we will denote $I \leftarrow I_{\text{diffuse}}$ as the diffuse image obtained by the polarimetric decomposition. Figure 3(b) shows the separated diffuse and specular images. The diffuse image I will be used for photometric stereo. Note that this diffuse-specular separation using polarized illumination and cameras has been often used in other systems [15, 18] such as light stages. DDPS applies the same principle to the display photometric stereo by using a conventional LCD and a polarization camera.

Display Superpixels For the computational efficiency of our end-to-end optimization, we parameterize the display with $P = 16 \times 9$ superpixels, where each superpixel is a group of 240×240 pixels. Ablation on the superpixel resolution can be found in the Supplemental Document.

Calibration We estimate the location of each superpixel with respect to the camera. To this end, we develop a mirror-based calibration method that estimates superpixel locations by using display patterns reflected on a mirror. We refer to the Supplemental Document for the details on

the mirror-based calibration. We also calibrate the intrinsic parameters of the camera and the non-linearity of display intensity using standard methods [55]. Figure 7 shows the calibrated superpixel locations.

5. Dataset Creation using 3D Printing

We describe our strategy for creating a training dataset using 3D printing. This allows for easily creating a real-world dataset with known geometry that can be used for DDPS. Figure 4(a)&(b) show the 3D printed objects and their ground-truth 3D models. For each training scene, we capture raw basis images $\mathcal{B} = \{B_j\}_{j=1}^P$, where j is the index of the basis illumination of which only j-th superpixel is turned on with its full intensity as white color. We then extract the silhouette mask S using the average image of the basis images I_{avg} that present well-lit appearance for most of the object scene points as shown in Figure 4(c). Given the silhouette mask S, we align the ground-truth geometry of the 3D-printed object in the scene by optimizing the pose of the ground-truth mesh with a silhouette rendering loss:

$$\underset{\mathbf{t},\mathbf{r}}{\text{minimize}} \| f_s(\pi; \mathbf{t}, \mathbf{r}) - S \|_2^2, \tag{2}$$

where π is the known 3D model, $\mathbf{t}$ and $\mathbf{r}$ are the translation and rotation of the model. $f_s(\cdot)$ is the differentiable silhouette rendering function. We solve Equation (2) using gradient descent in Mitsuba3 [23]. Once the pose parameters are obtained, we render the normal map with the 3D model at the optimized pose, which serves as the ground-truth normal map N_{GT}, shown in Figure 4. We create 40 training scenes and 4 test scenes with ground-truth normals. Note that even trained on the 3D-printed objects, DDPS enables effective reconstruction for diverse real-world objects as demonstrated in the results.

6. Learning Display Patterns

We learn display patterns using the 3D-printed training dataset consisting of ground-truth normal maps N_{GT} and basis images $\mathcal{B} = \{B_j\}_{j=1}^P$. We denote K different display patterns as $\mathcal{M} = \{\mathcal{M}_i\}_{i=1}^K$, where the i-th display pattern $\mathcal{M}_i$ is modeled as an RGB intensity pattern of P superpixels: $\mathcal{M}_i \in \mathbb{R}^{P \times 3}$, which is our optimization variable.

For end-to-end training of the display RGB intensity patterns $\mathcal{M}$, we develop a differentiable image formation function $f_I(\cdot)$ and a differentiable photometric-stereo method $f_n(\cdot)$, which are chained together via auto-differentiation. The differentiable image formation $f_I(\cdot)$ takes a display pattern $\mathcal{M}_i$ and the basis images $\mathcal{B}$ of a training scene, and simulates the captured images $\mathcal{I} = \{I_i\}_{i=1}^K$ for the display patterns being optimized. The photometric-stereo method $f_n(\cdot)$ then processes the simulated captured images $\mathcal{I}$ to estimate surface normal N. Below, we describe each component in details.

(a) 3D-printed objects (b) Rendered objects

(c) Average image overlayed with the ground-truth silhouette (d) Fitted silhouette (red) overlayed with the average image (e) Ground-truth normal map from the fitted 3D model
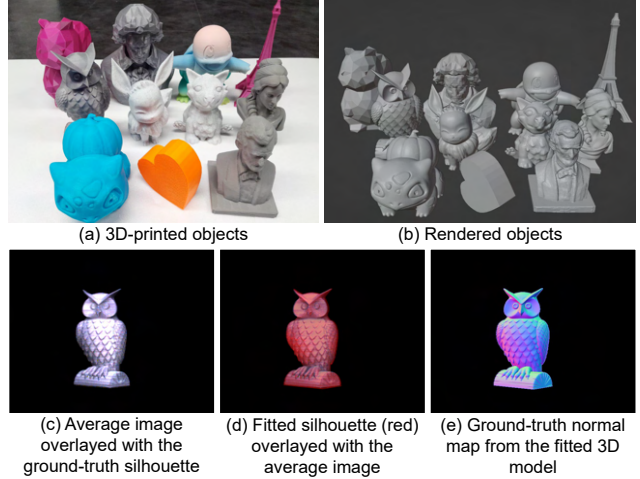
Figure 4. **Training dataset creation with 3D printing.** To learn display patterns, we propose to use (a) 3D-printed objects that have corresponding (b) known ground-truth 3D models. (c) We extract the silhouette S from the averaged basis images and (d) align the ground-truth 3D models with the captured image as depicted with the fitted silhouette in red on top of the average image. (e) We obtain a ground-truth normal map from the fitted 3D model.

6.1. Differentiable Image Formation

For the basis images $\mathcal{B}$ of a training sample, we simulate a raw image captured under a display pattern $\mathcal{M}_i$ as

$$I_i = f_I(\mathcal{M}_i, \mathcal{B}) = \sum_{j=1}^P B_j \mathcal{M}_{i,j}, \tag{3}$$

where $\mathcal{M}_{i,j}$ is the j-th superpixel RGB intensity in the display pattern $\mathcal{M}_i$. For K display patterns, we synthesize each image as

$$\mathcal{I} = \{f_I(\mathcal{M}_i, \mathcal{B})\}_{i=1}^K. \tag{4}$$

Figure 5 shows the overview of our image formation.

This weighted-sum formulation exploits the basis images acquired for real-world 3D printed objects, based on light-transport linearity in the regime of ray optics. Compared to using variants of rendering equations as differentiable image formations [5, 6], the image formation with basis images synthesizes realistic images in a computationally efficient manner, comprising only a single weighted summation, *serving as a memory-efficient and realistic image formation suitable for end-to-end pattern learning.*

6.2. Differentiable Photometric Stereo

We reconstruct surface normal N from the images $\mathcal{I}$ captured or simulated under the display patterns $\mathcal{M}$:

$$N = f_n(\mathcal{I}, \mathcal{M}). \tag{5}$$

Note that the images $\mathcal{I}$ mostly contain diffuse-reflection components as a result of the polarimetric diffuse-specular

Differentiable image formation

Element-wise product

Basis images of 3D-printed objects

Display patterns

Simulated captured images

Differentiable photometric stereo
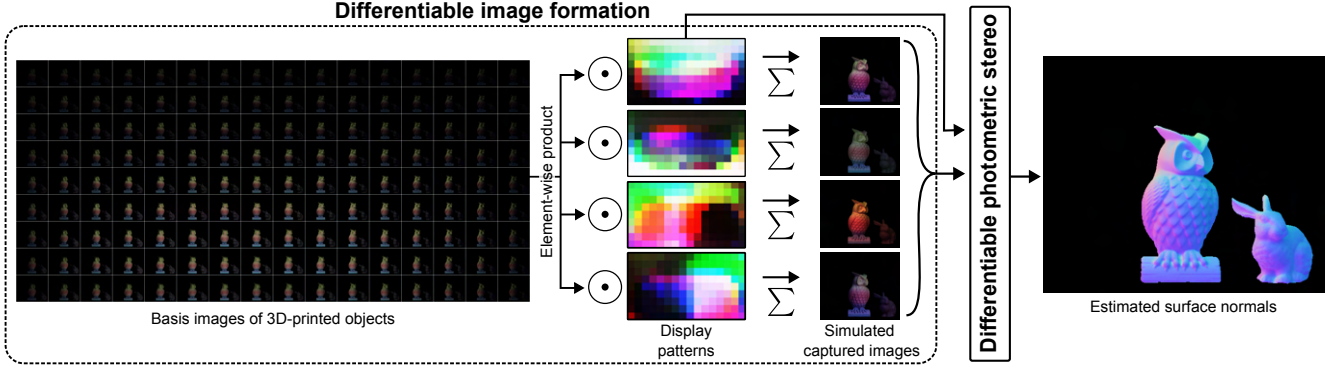
Estimated surface normals

Figure 5. **Differentiable framework.** Using 3D-printed objects as a dataset allows for simulating real-world captured images with a differentiable image formation. We reconstruct high-fidelity surface normals using differentiable photometric stereo from the simulated captured images.

separation described in Section 4. Using the optically-separated diffuse image $\mathcal{I}$, we develop an analytic trinocular photometric-stereo method that is independent of the training dataset and has no training parameters. This enables *effective end-to-end learning of display patterns by sorely focusing on optimizing display patterns without any other learning variables such as neural networks.*

We start by denoting the captured diffuse RGB intensity of a camera pixel under the i-th display pattern as I_i^c, where c is the color channel $c \in \{R, G, B\}$. Note that dependency on the pixel is omitted in the notation of I_i^c for simplicity. We denote the spatially-varying per-pixel illumination vector coming from the center of j-th superpixel on the monitor to a scene point corresponding to the pixel as l_j. Note that the illumination vectors are computed considering the different locations of the scene points. The scene points are assumed to lie on a plane which is a fixed distance (50 cm in our experiment) away from the camera. We then formulate a linear equation as

$$\mathbf{I} = \boldsymbol{\rho} \odot \mathbf{M}\mathbf{l}\mathbf{N}, \qquad (6)$$

where $\mathbf{I} \in \mathbb{R}^{3K \times 1}$, $\boldsymbol{\rho} \in \mathbb{R}^{3K \times 1}$, and $\mathbf{N} \in \mathbb{R}^{3 \times 1}$ are the vectorized intensity, albedo, and surface normals. $\odot$ is Hadarmard product. $\mathbf{M} \in \mathbb{R}^{3K \times P}$, $\mathbf{l} \in \mathbb{R}^{P \times 3}$ are the matrices for the pattern intensity and illumination directions. Note that the only unknown variables are the surface normal $\mathbf{N}$ and the albedo $\boldsymbol{\rho}$. Refer to the Supplemental Document for the formulation details.

We set the albedo $\boldsymbol{\rho}$ as the max intensities among captures to for numerical stability and solve for the surface normal $\mathbf{N}$ using the pseudo-inverse method: $\mathbf{N} \leftarrow (\boldsymbol{\rho} \odot \mathbf{M}\mathbf{l})^\dagger \mathbf{I}$, where $\dagger$ is the pseudo-inverse operator. Figure 5 shows the reconstructed surface normals. We exploit the differentiability of our analytic reconstructor for effective end-to-end optimization of display patterns.

6.3. Training

Equipped with the image formation and the reconstructor, we learn the display patterns $\mathcal{M}$ by solving an optimization problem:

$$\underset{\mathcal{M}}{\text{minimize}} \sum_{\mathcal{B}, N_{\text{GT}}} \text{loss}\left(f_n\left(\{f_I(\mathcal{M}_i, \mathcal{B})\}_{i=1}^K, \mathcal{M}\right), N_{\text{GT}}\right),$$
$$(7)$$

where $\text{loss}(\cdot) = (1 - N \cdot N_{\text{GT}})/2$, which is the normalized cosine distance, penalizes the angular difference between the estimated and the ground-truth normals from the 3D-printed dataset, meaning that the patterns are learned on the entire training dataset. To ensure the physically-valid intensity range from zero to one of the display pattern $\mathcal{M}$, we apply a sigmoid function to the optimization variable: $\mathcal{M} \leftarrow \text{sigmoid}(\mathcal{M})$. We use Adam optimizer [27].

6.4. Testing

Once the display patterns are learned, we perform testing on real-world objects. Specifically, we capture images under the learned K display patterns, perform diffuse-specular separation, and obtain diffuse image I_i for the i-th display pattern. We then estimate surface normals using our photometric stereo method:

$$N = f_n(\mathcal{I}). \qquad (8)$$

7. Assessments

We assess DDPS on diverse objects. Refer to the Supplemental Document for complete results.

Learned Patterns Figure 6 shows the patterns learned with DDPS. The learned patterns exhibit distinctively-colored regions and adjusted brightness for robust normal reconstruction. We evaluate the learned patterns regarding normal-reconstruction accuracy with common heuristic patterns: OLAT [47], group OLAT [8], monochromatic

Flat gray initialization | Mono-gradient initialization

Iteration 0 | Iteration 4 | Iteration 450 | Iteration 0 | Iteration 4 | Iteration 450

(a) Patterns

0   0.5   1   0   0.5   1

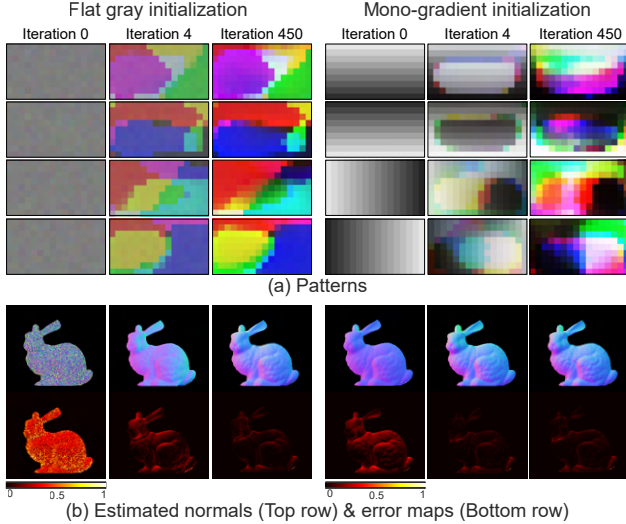(b) Estimated normals (Top row) & error maps (Bottom row)

Figure 6. **Learning process.** DDPS allows the learning of display patterns for high-quality normal reconstruction, not only from sub-optimal heuristically-designed patterns but also from flat gray noise that does not require any prior knowledge of the imaging system.

Illumination patterns	Number of patterns	Reconstruction error ↓	
		Initial	Learned
OLAT	4	0.1707	0.0486
Group OLAT	4	0.0805	0.0475
Mono-gradient	4	0.0913	0.0443
Mono-complementary	4	0.1044	0.0453
Tri-gradient	2	0.0933	0.0512
Tri-complementary	2	0.0923	0.0478
Flat gray	4	0.3930	0.0466
Mono-random	4	0.2533	0.0484
Tri-random	2	0.1461	0.0476

Table 1. Comparison of display patterns without and with our end-to-end optimization.

Number of patterns	Reconstruction error ↓	
	Initial	Learned
2	0.1461	0.0476
3	0.1415	0.0467
4	0.1096	0.0463
5	0.1001	0.0467

Table 2. Quantitative results of reconstructed surface normals with varying number of patterns for the trichromatic random patterns.

gradient [32], monochromatic complementary [24], trichromatic gradient [35], trichromatic complementary [28]. Table 1 show that the learned patterns outperform all existing heuristic patterns on the test dataset. We measured the average reconstruction error loss(·) on the 3D-printed test dataset. It is worth noting that *DDPS allows using initial patterns that do not require any prior knowledge of the imaging system*. That is, initialization with monochromatic random, trichromatic random, and flat gray noise also results in competitive results.

Real-world Objects   For the experiments, we used 40 training scenes containing various 3D-printed objects. Even though increasing the number of training samples is feasible, we found that DDPS already allows for high-quality reconstruction for in-the-wild real-world objects in this configuration, as shown in Figure 10. We speculate that this capability originates from effective rendering and analytical reconstruction without any additional training parameters as well as supervision with the 3D-printed dataset captured by a real setup.

Number of Patterns   Since photometric stereo solves for five unknowns (RGB diffuse albedo and surface normals), the minimum number of patterns is set to two, providing six measurements with the RGB channel for each. Table 2 shows that using two patterns learned by DDPS already outperforms any tested heuristic design using four patterns, demonstrating improved capture efficiency. Moreover, using two learned patterns is often sufficient, as shown by the

converged reconstruction errors.

Robustness to Simplifications   For efficient end-to-end pattern learning, DDPS has made assumptions including light source modeling and intensity falloff in its image formation and reconstruction. While the validity of these assumptions is often critical for conventional approaches that use synthetic training data, DDPS exhibits robustness against such simplifications, as demonstrated in all the qualitative and quantitative results. This is because the learned display patterns are optimized to achieve accurate normal reconstruction on a real-world 3D-printed dataset, taking into account such assumptions.

Here, we conduct additional experiments to test the robustness of DDPS. First, we evaluate DDPS under inaccurate superpixel locations. Instead of using our mirror-based calibration (Section 4), we manually place superpixels to lie at grid locations on a 3D plane, which deviates from the ground-truth locations. See Figure 7. DDPS with the inaccurate superpixel locations still provides accurate normal reconstruction with the error 0.0456 comparable to 0.0453 corresponding to using accurate superpixel locations. Second, we evaluate the assumption of consistent intensity with respect to distance. DDPS with and without intensity fall-off show comparable reconstruction errors of 0.0429 and 0.0453, indicating the robustness of DDPS against light fall-off modeling. Third, we test test DDPS for an object at varying depths: 40/50/80/100 cm. Even though we assume planar scene geometry at a fixed distance of 50 cm in our image formation, DDPS enables accurate normal reconstruction with the corresponding er-
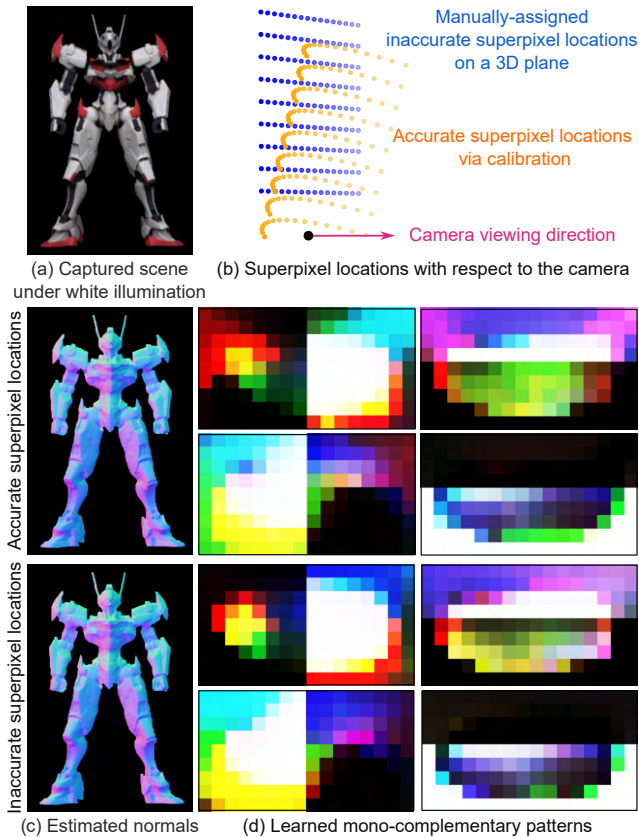
(a) Captured scene under white illumination

(b) Superpixel locations with respect to the camera

Manually-assigned inaccurate superpixel locations on a 3D plane

Accurate superpixel locations via calibration

Camera viewing direction

Accurate superpixel locations

Inaccurate superpixel locations

(c) Estimated normals

(d) Learned mono-complementary patterns

Figure 7. **Robustness against inaccurate superpixel locations.** We test DDPS for our calibrated curved-monitor superpixel locations, shown as (b) orange dots, and for the manually placed inaccurate plane superpixel locations, shown as blue dots, respectively. DDPS automatically compensates for the location error of the superpixels by (d) learning display patterns for such configuration, resulting in (c) high-quality normal maps.

rors of 0.0494/0.0417/0.0428/0.0561 for the varying depths. That is, in that depth range, we achieve reconstruction errors lower than 0.0805, which is the error using the best-performing heuristic pattern, group OLAT for the 50cm-distant objects. These experiments further demonstrate the robustness of DDPS against various simplifications.

Impact of Diffuse-specular Separation In order to acquire diffuse-dominant images, DDPS exploits linearly-polarized light emitted from the monitor and the polarization camera. Figure 8 shows that the reconstructed surface normals from the diffuse-dominant images obtained by DDPS provide more accurate reconstruction than using the images containing both diffuse and specular reflections.

Comparison with Learning-based Photometric Stereo We compare the reconstructed normals using the learned patterns to state-of-the-art normal-reconstruction methods that leverage neural networks and support area light sources compatible with our learned patterns: UniPS [21], SDM-
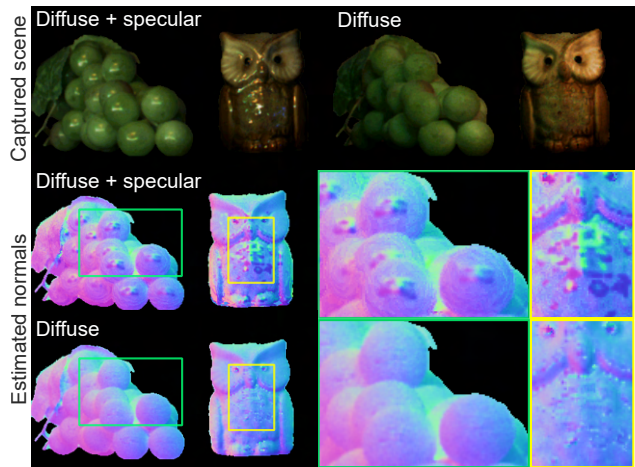

Figure 8. **Impact of diffuse-specular separation.** DDPS exploits polarization for optical diffuse-specular separation, leading to accurate normal reconstruction.
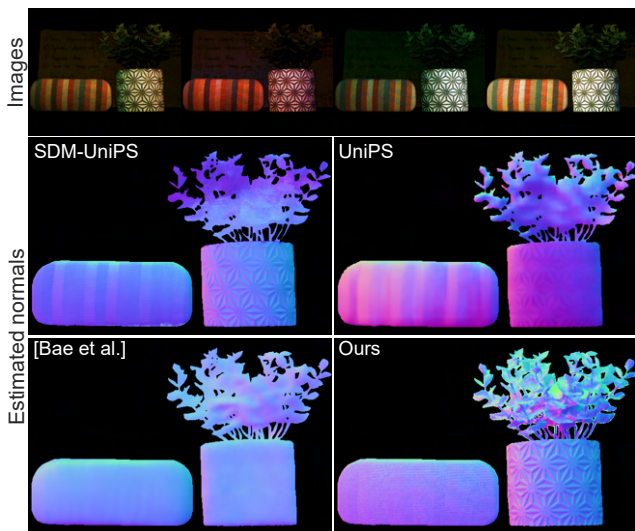

Figure 9. **Comparison to learning-based methods.** DDPS with the analytic reconstructor shows fine geometric details on the leafs, vase, and textile, outperforming the other methods.

UniPS [22], and Bae et al. [4]. UniPS and SDM-UniPS use multiple images under diverse unknown illumination conditions. Bae et al. [4] reconstruct the normal map from a single image. Figure 9 shows that DDPS outperforms the other methods. In particular, uncalibrated learned methods often fail to handle out-of-distribution examples such as the leaves in the scene. In contrast, DDPS exploits shading cue for physically-valid and accurate normal reconstruction.

Learning-based Reconstructor DDPS uses analytic photometric stereo as a training-free and dataset-independent module for normal reconstruction. When we simply replace the analytic photometric stereo with a learning-based photometric stereo, UniPS [21], the average

(a) Learned patterns and captured scenes
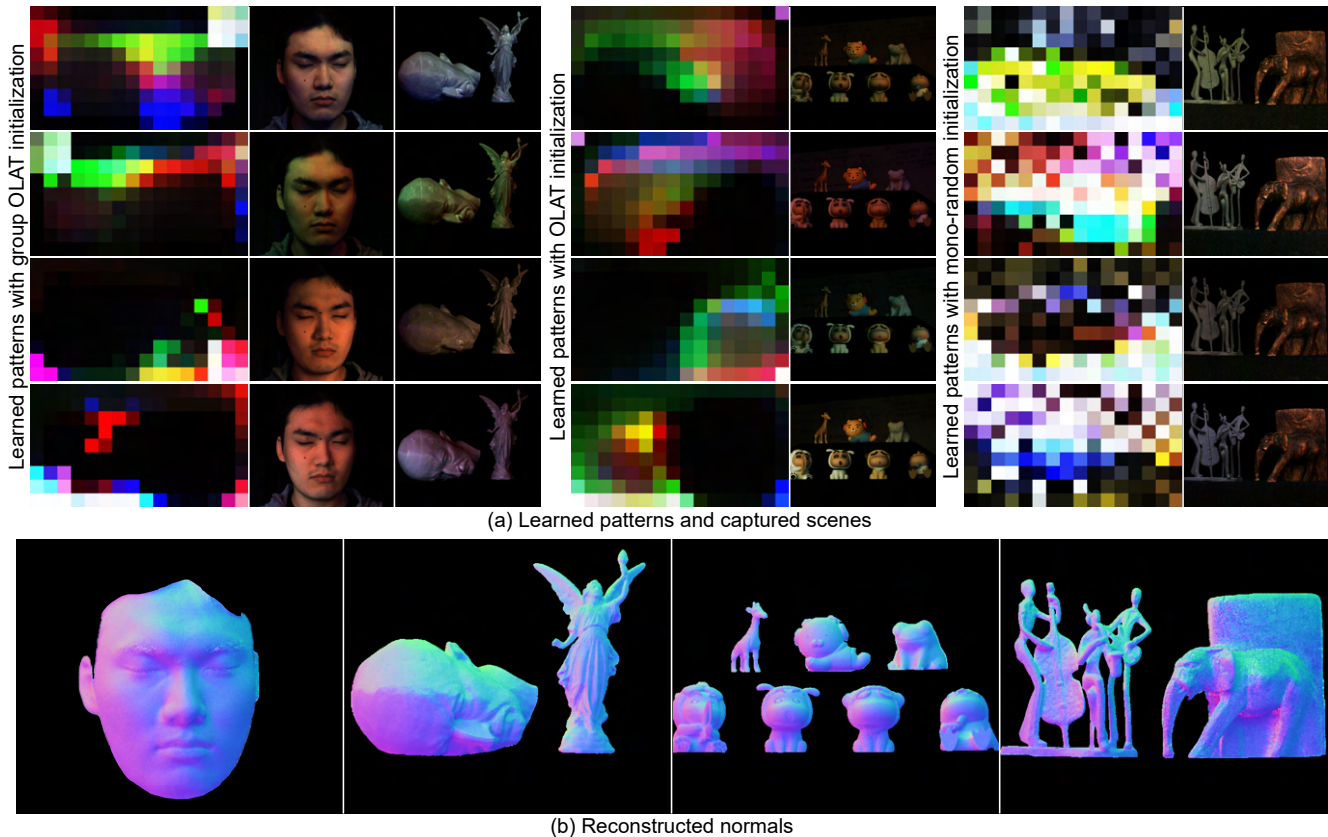

(b) Reconstructed normals

Figure 10. **Reconstruction results.** We reconstruct normals of diverse objects with the learned patterns using DDPS. Note that the patterns are learned on our 3D-printed training dataset.

reconstruction error increases from 0.0475 to 0.0951, using the group-OLAT initialization. This degradation can be attributed to that the backward gradient for the display patterns does not flow as effectively due to the complex network structure of UniPS. Also, the network is not designed to effectively utilize complex display patterns. Developing a learning-based photometric stereo suitable for DDPS would be an interesting future work.

8. Discussion

First, DDPS focuses on estimating normals, leaving depth reconstruction as a future work. Using multi-view cameras could resolve the problem and prompt research into optimizing patterns for multi-view cameras. Second, we encountered challenges in achieving high-speed synchronization between the display and the camera. This could potentially be circumvented with external hardware triggering, which would facilitate the reconstruction of surface normals for dynamic objects. Third, it would be interesting to apply DDPS for various types of display-camera systems such as a mobile phone. Lastly, our image formation model does not consider shadow and global illumination, which we further analyze in our Supplemental Document.

9. Conclusion

In this paper, we presented DDPS, a method for learning display patterns for robust display photometric stereo departing from using heuristic patterns. Our differentiable framework consisting of basis-illumination image formation and analytic photometric stereo, the use of 3D printing for real-dataset creation, and display polarimetric separation allow for learning display patterns that leads to high-quality normal reconstruction for diverse objects. Also, DDPS demonstrates robustness against various simplifications in image formation, reconstruction, and calibration. We believe that DDPS takes a step towards practical high-quality 3D reconstruction.Beyond display photometric stereo, the principles underpinning DDPS would be applied to a range of illumination-camera systems, including light stages, mobile phones, and large-scale displays.

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 *IEEE Conf. Comput. Vis. Pattern Recog.*, June 2023. 1

[4] Gwangbin Bae, Ignas Budvytis, and Roberto Cipolla. Estimating and exploiting the aleatoric uncertainty in surface normal estimation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13137–13146, 2021. 7

[5] Seung-Hwan Baek and Felix Heide. Polka lines: Learning structured illumination and reconstruction for active stereo. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5757–5767, 2021. 2, 4

[6] Seung-Hwan Baek and Felix Heide. All-photon polarimetric time-of-flight imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17876–17885, 2022. 4

[7] Seung-Hwan Baek, Tizian Zeltner, Hyunjin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM Trans. Graph.*, 39(4):139, 2020. 3

[8] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM Trans. Graph.*, 40(4):1–15, 2021. 2, 5

[9] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *Eur. Conf. Comput. Vis.*, pages 745–762. Springer, 2020. 2

[10] Ziang Cheng, Junxuan Li, and Hongdong Li. Wildlight: In-the-wild inverse rendering with a flashlight. In *IEEE Conf. Comput. Vis. Pattern Recog.*, June 2023. 1

[11] James J Clark. Photometric stereo using lcd displays. *Image and Vision Computing*, 28(4):704–714, 2010. 2

[12] Edward Collett. Field guide to polarization. Spie Bellingham, WA, 2005. 3

[13] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156, 2000. 2

[14] Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. Deep polarization imaging for 3d shape and svbrdf acquisition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 15567–15576, 2021. 2

[15] Yannick Francken, Tom Cuypers, Tom Mertens, Jo Gielis, and Philippe Bekaert. High quality mesostructure acquisition using specularities. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1–7. IEEE, 2008. 2, 3

[16] Graham Fyffe and Paul Debevec. Single-shot reflectance measurement from polarized color gradient illumination. In *IEEE Int. Conf. Comput. Photo.*, pages 1–10. IEEE, 2015. 2

[17] Graham Fyffe, Paul Graham, Borom Tunwattanapong, Abhijeet Ghosh, and Paul Debevec. Near-instant capture of high-resolution facial geometry and reflectance. In *Comput. Graph. Forum*, volume 35, pages 353–363. Wiley Online Library, 2016. 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 *Comput. Graph. Forum*, volume 28, pages 1161–1170. Wiley Online Library, 2009. 2, 3

[19] Vlastimil Havran, Jan Hošek, Šárka Němcová, Jiří Čáp, and Jiří Bittner. Lightdrum—portable light stage for accurate btf measurement on site. *Sensors*, 17(3):423, 2017. 2

[20] Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C Sankaranarayanan. Reflectance capture using univariate sampling of brdfs. In *Int. Conf. Comput. Vis.*, pages 5362–5370, 2017. 2

[21] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12591–12600, 2022. 7

[22] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. *arXiv preprint arXiv:2303.15724*, 2023. 7

[23] 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. 4

[24] Christos Kampouris, Stefanos Zafeiriou, and Abhijeet Ghosh. Diffuse-specular separation using binary spherical gradient illumination. In *EGSR (EI&I)*, pages 1–10, 2018. 2, 6

[25] Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. Efficient reflectance capture using an autoencoder. *ACM Trans. Graph.*, 37(4):1–10, 2018. 2

[26] Kaizhang Kang, Cihui Xie, Chengan He, Mingqi Yi, Minyi Gu, Zimin Chen, Kun Zhou, and Hongzhi Wu. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Trans. Graph.*, 38(6):1–12, 2019. 2

[27] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 5

[28] 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 *Eur. Conf. Comput. Vis.*, pages 522–537. Springer, 2022. 1, 2, 6

[29] Chloe LeGendre, Xueming Yu, Dai Liu, Jay Busch, Andrew Jones, Sumanta Pattanaik, and Paul Debevec. Practical multispectral lighting reproduction. *ACM Trans. Graph.*, 35(4):1–11, 2016. 1, 2

[30] 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. *IEEE Trans. Image Process.*, 29:4159–4173, 2020. 2, 3

[31] Chao Liu, Srinivasa G Narasimhan, and Artur W Dubrawski. Near-light photometric stereo using circularly placed point

light sources. In *IEEE Int. Conf. Comput. Photo.*, pages 1–10. IEEE, 2018. 2

[32] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eur. Conf. Render. Tech.*, pages 183–194, 2007. 1, 2, 6

[33] Xiaohe Ma, Kaizhang Kang, Ruisheng Zhu, Hongzhi Wu, and Kun Zhou. Free-form scanning of non-planar appearance with neural trace photography. *ACM Trans. Graph.*, 40(4):1–13, 2021. 2

[34] 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. 2

[35] Abhimitra Meka, Christian Haene, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, et al. Deep reflectance fields: high-quality facial reflectance field inference from color gradient illumination. *ACM Trans. Graph.*, 38(4):1–12, 2019. 1, 2, 6

[36] Abhimitra Meka, Rohit Pandey, Christian Haene, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. Deep relightable textures: volumetric performance capture with neural rendering. *ACM Trans. Graph.*, 39(6):1–21, 2020. 1

[37] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Trans. Graph.*, 37(6):1–12, 2018. 1

[38] Emilie Nogue, Yiming Lin, and Abhijeet Ghosh. Polarization-imaging surface reflectometry using near-field display. In *Eurographics Symposium on Rendering. The Eurographics Association*, volume 2, 2022. 2

[39] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Trans. Graph.*, 40(4):1–21, 2021. 1

[40] Jaesik Park, Sudipta N Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Robust multiview photometric stereo using planar mesh parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(8):1591–1604, 2016. 1

[41] Yifan Peng, Suyeon Choi, Nitish Padmanaban, and Gordon Wetzstein. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.*, 39(6):1–14, 2020. 2

[42] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12581–12590, 2022. 2, 3

[43] J Riviere, P Peers, and A Ghosh. Mobile surface reflectometry. In *Comput. Graph. Forum*, volume 1, pages 191–202, 2016. 2

[44] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Int. Conf. Comput. Vis. Worksh.*, pages 501–509, 2017. 2

[45] Carolin Schmitt, Simon Donne, Gernot Riegler, Vladlen Koltun, and Andreas Geiger. On joint estimation of pose, geometry and svbrdf from a handheld scanner. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3493–3503, 2020. 1

[46] Soumyadip Sengupta, Brian Curless, Ira Kemelmacher-Shlizerman, and Steven M Seitz. A light stage on every desk. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2420–2429, 2021. 1

[47] Tiancheng Sun, Zexiang Xu, Xiuming Zhang, Sean Fanello, Christoph Rhemann, Paul Debevec, Yun-Ta Tsai, Jonathan T Barron, and Ravi Ramamoorthi. Light stage super-resolution: continuous high-frequency relighting. *ACM Trans. Graph.*, 39(6):1–12, 2020. 2, 5

[48] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph.*, 24(3):756–764, 2005. 2

[49] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, et al. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, 2006. 1

[50] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 1

[51] 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

[52] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5565–5574, 2022. 1

[53] 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. 2

[54] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Trans. Graph.*, 40(1):1–17, 2021. 2

[55] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1330–1334, 2000. 4

[56] Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yi. Relightable neural human assets from multi-view gradient illuminations. 2023. 1