This WACV 2023 paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

# Surface normal estimation from optimized and distributed light sources using DNN-based photometric stereo

Takafumi Iwaguchi
Kyushu University, Fukuoka, Japan
iwaguchi@ait.kyushu-u.ac.jp

Hiroshi Kawasaki
Kyushu University, Fukuoka, Japan
kawasaki@ait.kyushu-u.ac.jp

## Abstract

*Photometric stereo (PS) is a major technique to recover surface normal for each pixel. However, since it assumes Lambertian surface and directional light to estimate the value, a large number of images are usually required to avoid the effects of outliers and noise. In this paper, we propose a technique to reduce the number of images by using distributed light sources, where the patterns are optimized by a deep neural network (DNN). In addition, to efficiently realize the distributed light, we use an optical diffuser with a video projector, where the diffuser is illuminated by the projector from behind, the illuminated area on the diffuser works as if an arbitrary-shaped area light. To estimate the surface normal using the distributed light source, we propose a near-light photometric stereo (NLPS) using DNN. Since optimization of the pattern of distributed light is achieved by a differentiable renderer, it is connected with NLPS network, achieving end-to-end learning. The experiments are conducted to show the successful estimation of the surface normal by our method from a small number of images.*

## 1. Introduction

Among the wide variety of active lighting-based 3D acquisition methods, photometric stereo (PS) has attracted researchers for more than 40 years because it has a unique feature in that it is capable of robust recovery of the pixel-wise surface normal just from images where the scene is illuminated from various light directions [13]. To efficiently estimate the surface normal, linear solution assuming Lambertian surfaces and directional light sources have been proposed. By using the technique, it is possible to recover surface normal for each pixel just using three inputs (pixel values for three lighting directions) in theory. However, since there are many types of reflections, such as specularity or subsurface scattering, a large number of images are usually required to suppress those effects as well as outliers and noise. In addition, there are many types of light sources, such as a point light source or area light sources, non-linear techniques are usually applied after the linear solution, *e.g.*, a near light photometric stereo (NLPS) [2, 9] or deep neural network solution [21, 24, 40, 34, 22] are proposed recently.

In this paper, we propose a technique to reduce the number of images by using distributed light sources, where the patterns are optimized by a deep neural network (DNN). To deal with this problem, we propose a method to simultaneously design arbitrary-shaped distributed light sources as well as train DNN for NLPS. In the method, since the design of the distributed light pattern is optimized by using a differentiable renderer, it is connected with NLPS module to achieve end-to-end learning.

In terms of the realization of arbitrary-shaped distributed light sources, one approach is to physically arrange many point light sources, which is a laborious task. In our method, we propose to use diffuse optics for flexible measurements. In the system, an optical diffuser with a static video projector is used, which has not been proposed yet for PS, since it is not equivalent to directional light. If the diffuser is illuminated from behind by the video projector, the illuminated area on the diffuser works as an arbitrary-shaped area light. *i.e.*, a large number of point-light sources are arranged on the diffuser.

The experimental results are shown that the proposed method can estimate the surface normals effectively, not only for simulation data but also for real environments by achieving shape reconstruction with a small number of measurements. We have made the following contributions,

- DNN solution for NLPS, which is designed to use light parameters as the input being allowed arbitrary distribution of light sources without retraining, is proposed.
- Optimal distribution of light patterns is designed by DNN in an end-to-end manner, which allows robust estimation of surface normal using a small number of inputs. Although a rigorous analysis of the number of measurements is practically difficult, it was in part verified by our experiments.

- Efficient capturing system using an optical diffuser with a video projector, which can achieve precise and arbitrary distribution of light sources without actually setting up independent light sources, is proposed.

## 2. Related work

### 2.1. Diffuse optics for active measurement

Diffusers have been used as light sources to perform active measurements, but little for PS. For example, when diffused by an isotropic diffuser, a specular reflection is known to vary its intensity according to the normal direction and this property has been used for processing objects with specular reflections, such as metal objects, as same as Lambertian objects [12, 30, 29]. Similarly, Nayar *et al*. realized structured light on a metallic object by projecting an encoded pattern onto a diffuser [28]. On the other hand, the concept of the use of light projected on a diffuse wall as a light source for PS was presented by Schechner *et al*. [35]. Since the diffuse reflection is weak, it can capture only dark shading images, and thus, to improve its signal-to-noise ratio, they proposed a light multiplexing method, requiring a huge number of images. In addition, to approximate the projected area as an infinite light source, the technique can recover only small objects even in a large acquisition room.

In contrast, since intensity distribution to forward direction is dominant for an optical diffuser, energy efficiency can be much improved and a small data acquisition room is possible. In addition, by introducing NLPS, the room size can be further reduced.

### 2.2. PS for non-Lambertian surface and NLPS

One of the most challenging problems for PS is the material of the non-Lambertian surfaces. As possible methods to remove non-Lambertian reflection, such as specular reflections, a method that uses shading images with four or more light sources [3] and a method that separates the reflection components using the optics have been proposed [27]. As an algorithm for specular reflection removal, a median filter-based method [25], methods based on bidirectional reflectance distribution function (BRDF) models [23, 8], and a method based on low-rank matrix completion [38] have all been proposed. Recently, DNN-based methods have been proposed [39, 32, 11, 7, 33, 34] in which the DNNs are trained on datasets that include a variety of materials to handle the non-Lambertian surfaces. However, their techniques can only recover the normals under the same setting for training. In the paper, we improve the DNN to use lighting information as the input so that the network can estimate normals for arbitrary lighting conditions other than training ones.

Another challenge for PS is a non-directional light source, such as a point light source or an area light source. Because normal estimation using point light sources is dependent on depth and cannot be computed in a linear system, several algorithms for NLPS have been studied [21, 24, 40, 34, 22]. Since they assumed isotropic light distribution, precise and complicated calibration is required, which is usually a difficult task. There is another NLPS that used an area light source using a computer display [36], where rectangular patterns were projected onto a display, and the normal and depth were then reconstructed. Since a typical liquid crystal display (LCD) has complex emission characteristics, including angular dependence, the reconstructed shapes had limited quality. In our method, DNN is first trained by CG for arbitrary light distributions followed by a fine-tune using real data sets to achieve sufficient quality.

### 2.3. Optimal illumination for the measurement

Li *et al*. proposed a method to improve the normal estimation accuracy, even when a small number of measurements is used [19]. In the technique, they assumed directional light and decided the optimal directions of the light by considering the presence of cast shadows caused by self-occlusion. In contrast, our method estimates the distribution of multiple point light sources, which can not only compensate for cast shadows but also improve the accuracy for low light conditions.

There is a technique to find optimal light distribution for acquiring BRDF [17], but not for PS. They used a multiplexing approach to reduce the number of measurements using a device with a high-density LED array. A DNN was used to learn how to multiplex light and decode BRDFs from observed images when the number of measurements performed was limited. We also apply a similar approach to estimate the distribution of light sources to reduce the necessary number of measurements for PS.

## 3. PS using a spatially distributed light source

### 3.1. PS with an optical diffuser

Conventional PS requires the light source information, such as the direction of each light source, to be known and requires calibration. In our method, rather than using physical light sources, we use an optical diffuser lit by a video projector that projects light patterns as shown in Fig. 1. The diffuser lit by a light pattern works as an area light source, which is equivalent to numerous point light sources distributed on the diffuser. The normals of the scene are estimated by extended NLPS using DNN.

This setup offers the following advantages. 1. The optical diffuser has strong intensity to forward direction and can achieve better energy efficiency than reflection ones [35]. 2. By simultaneously illuminating multiple positions of the diffuser by the video projector, it outputs more energy than

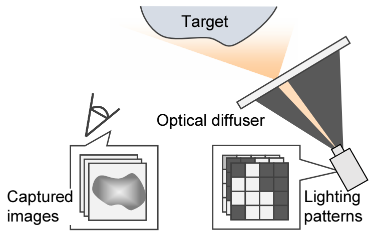



Figure 1: Illustration of the measurement setup. A spatially distributed light source is realized on the diffuser.

illuminating a single position, achieving a high SN ratio. 3. Multiplexed illuminations soften the shadows, reducing the effects of shadows near object boundaries with complex shapes and making normal estimations easier. Furthermore, it allows a flexible set-up, allowing the camera to be placed in front of the diffuser.

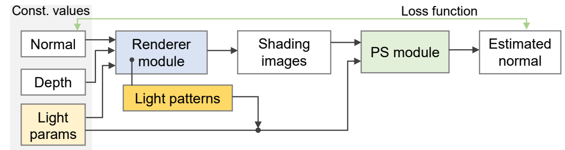
### 3.2. Optimal distribution of light source for NLPS

Multiple shading images under different lighting conditions are required to estimate surface normals, but the optimal distributions of light sources for NLPS are not obvious. In our approach, to determine a suitable distribution of light sources, we design the pattern for the video projector such that the accuracy of the estimated surface normals is high. One of the difficulties involved in this approach is that no general method is available to estimate the normals from scene images shaded under the set of arbitrarily distributed light sources. Therefore, in the proposed method, we design the patterns and the normal estimation algorithm simultaneously. For this purpose, we use a DNN-based PS-module for estimating normals and a differentiable renderer for synthesizing shading images from the light patterns that are optimized. Details are described in Sec. 4.

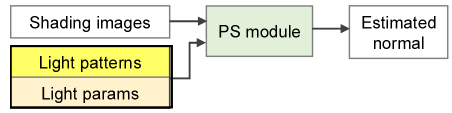
### 3.3. Algorithm overview

An overview of our proposed algorithm is shown in Fig. 2. During training (Fig. 2a), we use the differentiable renderer to synthesize shading images under arbitrary multiplexed lighting conditions. The inputs to the renderer are surface normals and depth maps, which are computed by rasterizing meshes in advance. Light parameters, including the center, pose, and size of the area light source, is also input. The renderer module then generates shading images based on the light parameters and weights representing the lighting patterns. The PS module then takes these shading images as inputs. To deal with a variety of lighting conditions, light information (*i.e.*, parameters and patterns) is also input. An end-to-end architecture is used that combines the renderer and PS modules to learn the area-lighting patterns and perform PS estimation simultaneously.

In the method, the PS module is trained independently by fixing the weights used for the arbitrary lighting conditions (Fig. 2b). The PS module is first pre-trained as part of the end-to-end architecture, with subsequent fine-tuning being performed using photo-realistic rendered images or real-world observations to compensate for global illumination, which is not rendered by the renderer module. For real-world data, the shading images of objects, where ground truth is known such as a sphere or a box, are captured using the learned patterns. The surface normal is then estimated using the PS module based on the shading images, the pre-calibrated light parameters, and ground truth.




(a) Training architecture.




(b) Inference architecture.

Figure 2: Algorithm overview.

## 4. Implementation

### 4.1. Renderer module

The renderer module generates a shading image under an arbitrary distributed light source. The shading image is generated by summing the basic shading images corresponding to each light source. We use two approaches to rendering basic shading images, and the algorithms are described in detail in Sec. 4.4. The module's input is a base shading image $B_i$ of each source $i$ in the distributed light source turned on at unit intensity, and a vector $E_i$ represents the intensity of each light source. The generation of the basic shading image is described in the dataset chapter. The intensity of the observed shading image $I$ is calculated as follows:

$$I = \min\left(\sum_i E_i \odot B_i + \epsilon, 1\right), \qquad (1)$$

where $\odot$ denotes element-wise multiplication. Here, to reproduce the observation in the real world, noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is added to the whole image according to a normal distribution, and pixel values exceed 1 is clipped to simulate saturation.

The intensity vector of the distributed light source is the parameter to be optimized. In the case of a projector, it corresponds to each pixel value of the projected pattern and technically may have any value within the range of output bit depths. However, the network may learn to rely on small changes in light source intensity, which is too sensitive to noise and leads to estimation failure for real-world projectors with nonlinear intensity response curves. To prevent this, we apply a Gumbel softmax function [15] $GS$ to the network parameters to learn the binary intensities:

$$E_i = GS(c_i, \tau), \tag{2}$$

where $c_i$ denotes a raw weight value for intensity of the i-th source and $\tau$ denotes temperature parameter of Gumbel softmax.

## 4.2. PS module

The PS module estimates the surface normal from the shading images. The amount of light received by an object's surface is given by the integral of the contributions made by point-light sources defined by the pattern of the video projector. Because each contribution is dependent on the distance between the point light source and the object surface, it is difficult to determine the object's normal analytically from the shadings when the object shape is unknown, which is typical chicken and egg problem. We, therefore, use DNN-based machine learning to estimate the surface normals from the shading.

The shading image depends on light source positions. Thus, a model trained on a dataset obtained under a specific light source configuration will not estimate the normals correctly for a different lighting configuration. One solution is uncalibrated PS [4, 26, 10], which recovers the normals without knowing lighting information, however, normals and lighting configurations are not determined uniquely only from shadings; this is known as bas-relief ambiguity [5]. Therefore, to estimate the normals under arbitrary light source configurations other than the trained one, we input both the shading images and the light source parameters to the PS module.

As shown in Fig. 3a, the network consists of a shading branch for processing the shading images, a light branch for processing the light source parameters, and a decoder. The shading branch and the decoder form a U-Net convolutional neural network structure to estimate the normal while also maintaining the geometry features. The input to this network is a stack of shading images, and its output is the normal. The input to the light branch is a two-dimensional image that describes the spatial distribution of the light source and the position and intensity of each source. Each pixel of the image is a four-dimensional vector which is given by concatenating 3D coordinates of the center of the discrete area of $\mathbf{P}$ and its intensity $\mathbf{E}_i$. This representation allows us to handle spatially distributed light with an arbitrary shape in any position and with any orientation. The light source information is encoded into a feature vector through a light branch. The features that are extracted by the encoder for the shading branch and the features of the light source information are then concatenated to form the decoder input.

The shading branch consists of three downsampled layers and the decoder consists of three upsampled layers that are connected via skip connections. The light branch consists of three downsampled layers. Batch normalization (BN), convolution, and activation (rectified linear unit or ReLU) processes are performed twice in each layer.

## 4.3. Training details

During training, the two modules are connected and are then used as an end-to-end architecture to learn the area lighting patterns and perform PS estimation simultaneously.

Two learning strategies are used to achieve the two goals of this work: learning efficient patterns and obtaining the required generalization performance for the light source locations. When learning efficient patterns, we use an augmentation approach with a small range of light source parameter variations to accelerate the training of the network weights that represent the lighting patterns by allowing the PS module to converge quickly. To obtain the generalization performance required for light source placement, we fix the network weights that represent the patterns and then train the PS module only. The generalization performance can be improved by increasing the number of variations in the light source configuration. The OR dataset, which will be presented in Sec. 4.4, is rendered online during training to allow augmentation with respect to light source configuration.

To deal with real-world images, which include global illumination, we fine-tuned the PS module using either photo-realistic computer graphics or real-world images. During the fine-tuning process, only the weights of the encoder layer in the shading branch are updated by the additional training.

We use the L-1 norm for the ground truth and the predicted normal as the loss function for the training as follows

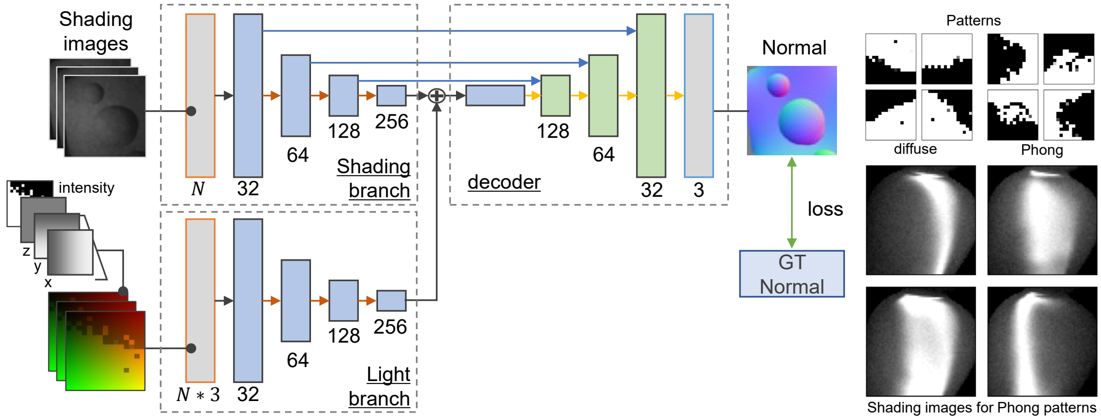
$$\mathcal{L} = \|\mathbf{n} - \tilde{\mathbf{n}}\|_1. \tag{3}$$

The adaptive moment estimation (ADAM) algorithm [18] is used for the optimization process. The learning rate is 0.001 and the number of epochs is 300.

## 4.4. Training dataset

The existing PS dataset cannot be used to train our network because it uses directional light. We create two datasets with different purposes for training.

The first is **Online Rendering (OR) dataset** that covers many variations of light source placement by augmenting them online during training. To achieve this, normal and depth maps are rasterized before training, and shading images are generated online from these maps during training according to the light source placement. The shading images are computed using linear matrix calculations based on the formulations of diffuse and Phong materials. We select 8 meshes from the blobby shape dataset [16] as the shape,




(a) (b)

Figure 3: (a) Network architecture of PS module. (b) Four patterns are learned for diffuse and Phong material, respectively. Each pattern is different for the same material, and patterns are different from the other material. In the example shading images of Phong material, it is confirmed that specular component does not always stay in the same place.

the translation of the object center is randomly set from (-0.5, -0.5, -0.5) to (0.5, 0.5, 0.5), and the rotation is from -180 degrees to 180 degrees about x, y, z axis, and the size is from 0.1 to 1.5. The dataset contains 4000 for diffuse and 4000 for Phong materials. Shading images are generated from the remaining 2 meshes in the same way and used for test.

The second one is **Eevee Prerender (EP) dataset**, which is generated by the Eevee renderer [1] implemented in Blender and contains a variety of materials. This rendering cannot be done during training, so we use pre-rendering. As a GPU-based renderer, Eevee is fast and can handle many types of materials and light sources, although it is not as realistic as path-tracing. To cover real-world materials, a composite image of the material is generated with various parameters. Following [11], we set different parameter sets for the three classes (diffuse, specular, metallic) in Disney's principled BSDF. We set the following parameters with slight modification from the original work: *baseColor, metallic, specular, roughness, sheen, sheen tint, IOR*. Please see supplemental material for the range of each parameter. Shapes are generated using the random shape generator plugin which can generate shapes with edges and smooth shapes using subdivisions. The dataset contains 2000 for diffuse, 1000 for specular, and 1000 for metallic. Images are rendered at a resolution of 256x256 and resized upon loading.

In both datasets, the translation of the center of the diffuser is randomly set from (-0.5, -0.5, 4.5) to (0.5, 0.5, 5.5), and the rotation is set from -15 to 15 degrees for the x, y, and z axes, and the size from 0.5 to 2.0. In the OR dataset, these parameters change at each iteration, whereas in the EP dataset parameters are fixed during training.

## 5. Experimental results

### 5.1. Evaluation with synthesized data

**Comparison with other NLPS methods:** We evaluate the accuracy of the proposed method's normal estimation in a simulation environment. The estimation accuracy of several state-of-the-art NLPS techniques [31, 20, 33] are compared for objects 1–5 made of different materials as shown in the Fig. 4a. [31] is an iterative calibrated PS method. It takes 8 shading images as input, with the light source position and camera parameters as inputs. [20] is a weakly calibrated method, and takes as input 6 shading images taken with the light source positioned in a roughly specified direction. [33] is a calibrated PS method based on DNN, so the camera and light source positions are given as input. The test dataset is generated by rendering the same scene as the setup shown in [33] with the EEVEE renderer since the provided pre-trained model is dependent on their training data. The light sources are placed on a 16x16 flat array. The resolution in this comparison is 64x64. The proposed method uses 16 optimized pattern images, and the model is fine-tuned to the setup. Since all methods use the light source position as an input also leads us to expect that all methods will perform well on this test data.

The qualitative evaluation is shown in the Fig. 4a. Objects #1–3 are dominated by diffuse components. The normal of #1 is consistently estimated by all the techniques, while [31] shows a normal shift in the upper part of the object. Object #2 is composed of planes and has an occlusion boundary; the DL-based techniques [20, 32] and proposed method succeed in estimating the boundary, while [31] fails. The fact that the plane is estimated as a curved surface also indicates that the simultaneous estimation of normal, depth, and albedo is an unstable computation. The proposed method is also accurate in the upper occlusion re-

Table 1: Error comparison with the state-of-the-art near light PS methods.

| | [31] | [20] | [33] | Proposed |
|---|---|---|---|---|
| MSE | 0.621 | 0.430 | 0.347 | 0.193 |
| MAE | 38.10 | 26.46 | 21.57 | 13.68 |

gion, showing the effect of diffuse light sources. #3 is a glossy material, and the DL-based methods [20, 32] show an overall normal shift, indicating that the trained model cannot adequately deal with the shading of this material. Object #4 has a specular component, and [32] has an artifact in the center of the object. On the other hand, [20] and the proposed method shows some artifacts, but the overall normal quality is high. #5 is a metallic material, and it can be seen that [20] cannot deal with this material given its low normal accuracy, while [31] fails to estimate the normal because the convergence condition of the optimization is not satisfied. The proposed method can deal with different shadings of the materials, and the estimation accuracy is high even in occlusion regions, confirming that there is no problem even in the presence of occlusion boundaries. The errors are shown in table 1 in Mean Squared Error (MSE) and Mean Angular Error (MAE). The results show the proposed technique outperforms other techniques. The high-resolution results (256x256) for the proposed method are shown in Fig. 4b. The most part of the estimation is highly accurate, but there is an error in the specular scene due to the cast shadows that are constantly existing.

**Evaluation of optimized patterns:** The estimated patterns for diffuse and Phong materials, where four patterns were assumed to be projected onto the scene to estimate surface normal, are shown in Fig. 3. In the patterns for the diffuse material, it is found that the patterns are divided into two regions where the boundaries of the regions are clean and smooth. In contrast, in the patterns for the Phong material, although the patterns are mostly divided into two parts, it is found that the shapes of the boundaries are complicated with small structures. We thought this is because such high-frequency patterns will make distinctive highlights at different locations as illustrated in the same figure(bottom right), which helps to estimate surface normals.

We compare the learned pattern to a random pattern to confirm that the learned pattern improves the accuracy of the normal estimation. The random pattern is binarized and approximately 50% of the total pixels are on. The test is performed on 100 diffuse and 100 specular object shapes generated by the simulation. The results are shown in Fig. 5. The diffuse material can estimate the normals from a small number of observations in different light source directions, and the proposed method achieves high accuracy from four images, while the accuracy of the random pattern does not increase as the number of images increases because the shading changes little depending on the pattern, and the accuracy is affected by the distribution of the images. For specular materials, the overlap between light sources leads to incorrect estimation. The proposed method automatically optimizes the combination of patterns with less overlap as the number of patterns increases, and the highest accuracy is obtained when the number of patterns is 16. On the other hand, the accuracy of random patterns does not change as the number of patterns increases, indicating that the shading is not effectively used.

Table 2: Error comparison of the proposed method with sparse PS method [41]. Errors are shown in MAE.

| $\sigma$ | Proposed | SPLINE-Net [41] |
|---|---|---|
| 0.000 | 13.01 | 10.00 |
| 0.050 | 15.03 | 26.13 |
| 0.100 | 13.78 | 35.36 |
| 0.400 | 36.60 | N/A |

Table 3: MAE on different material dataset.

| Model | diffuse | specular | metallic |
|---|---|---|---|
| D | 2.52 | 6.40 | 10.78 |
| DSM | 4.81 | 6.55 | 7.78 |

**Comparison with sparse PS:** We compared our technique with sparse PS technique [41] which estimates high quality normal from a small number of the measurement. We used a pre-trained model for inference with [41] which is provided by the authors trained for DiLiGenT dataset [37]. The test data are generated by physics-based rendering [14] in the same setup as the DiliGenT dataset. We rendered 100 realistic shading images of blobby shapes with strong specular effects. Note that the models of the proposed method are generated by different renderers for a fair comparison. To make the simulation realistic, we also added photon noise, where the noise becomes large if the exposure becomes low. MAE of the estimated normals of both methods is shown in table 2. If there is no noise, both [41] and our technique can estimate the surface normal with high accuracy. However, if the noise ratio increases, the input of our method does not change its appearance because distributed light keeps enough energy at each pixel by global illumination, whereas signal noise ratio (SNR) drastically decreases for [41] because directional light can provide limited energy at a point, and thus, it is confirmed that accuracy of the estimated surface normal of our method is better than [41].

**Evaluation of generalization ability across materials:** In this experiment, the network is trained and tested by using three different materials to evaluate its generalization ability across materials. The EP dataset [1] is divided into two and the network is trained using them: **D model** is trained on dataset consisting of 1000 diffuse objects and **DSM model**

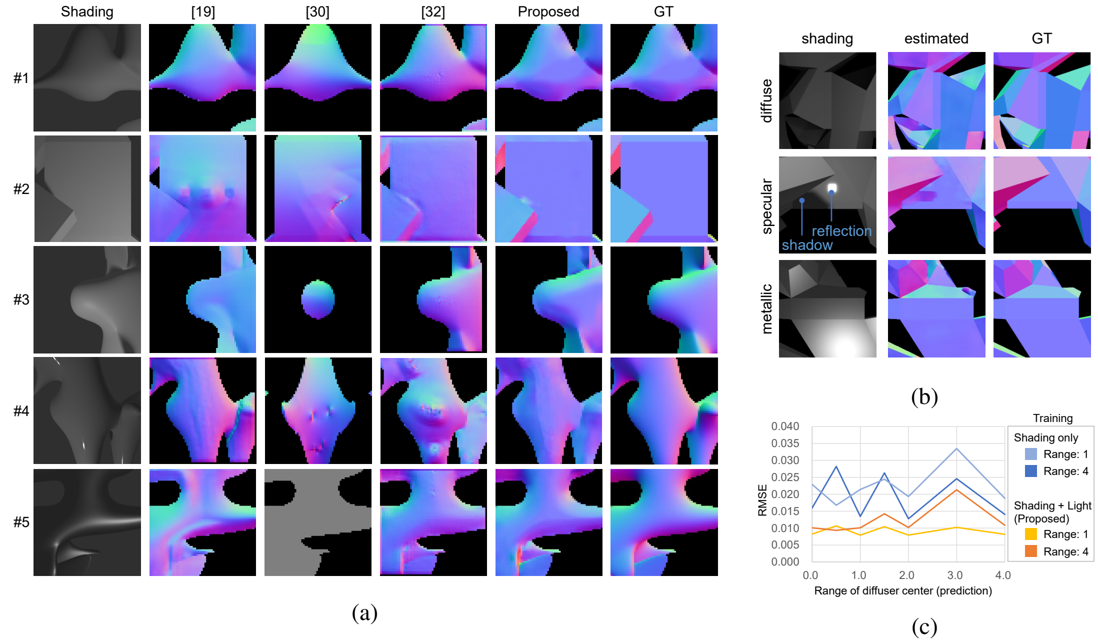



Figure 4: (a) Comparison with state-of-the-art techniques. #1, #2 are diffuse, #3 is glossy, #4 is specular, and #5 is a metallic material. The proposed technique stably estimates the surface normal of different materials, even in the presence of the occluding area and boundaries. (b) Additional results for the proposed method. (c) Evaluation of generalization performance for the configuration.

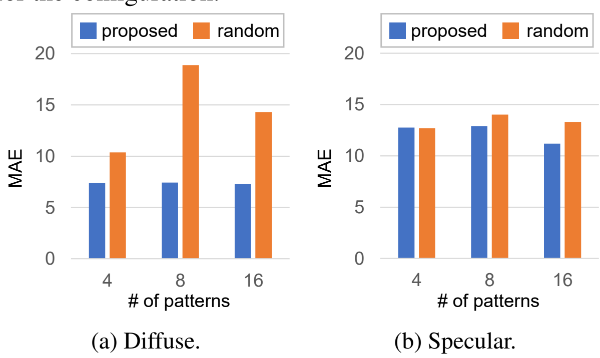



Figure 5: Evaluation of optimized patterns. (a) The proposed method achieves high accuracy from a small number of measurements for diffuse material. (b) For specular materials, the accuracy increases with the number of optimal patterns because it is easier to prevent overlapping reflections of light sources in different patterns.

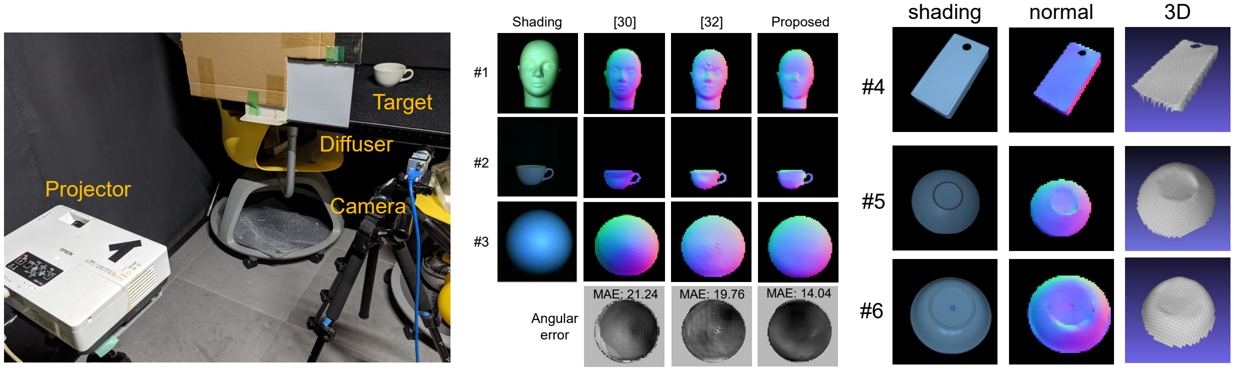
is trained on dataset consisting of 600 diffuse, 200 specular, and 200 metallic objects. The estimated normal errors for each network model on 200 test data for three materials are shown in table 3. The D model has the highest accuracy when test with diffuse objects, but has low accuracy for two other materials. The DSM model is more accurate than the D model for metallic materials, keeping certain accuracy on diffuse and specular objects.

**Evaluation of generalization performance for the configuration:** Next, we evaluate the effectiveness of the proposed network structure in terms of its generalization performance for the configuration of a camera, a diffuser, and a projector position. To show the generalization ability of our method in terms of system configuration, we compared our method with/without the light branch. We evaluated the error of estimated surface normals by changing the diffuser center position randomly in a bounding box. We tested with two bounding boxes sized, such as small (1) and large (4) and the results are shown in Fig. 4c. From the graph, it is observed that the models which are trained with light branches are better than those without. In addition, the small bounding box is better than the large one, probably because more data are required to improve the accuracy when there are large variations for training, such as the large bounding box. These results show that the proposed network achieves a sufficient generalization performance, even if the number of variations for training is small.

## 5.2. Real-world measurement

We verify the effectiveness of the proposed method in a real environment. The measurement setup consisted of a diffuser (plastic), an LCD projector, and a camera as shown in Fig. 6a. The area lighting on the diffuser is calibrated using the highlight positions for two metal spheres. The diffuser center is located 47.3 degrees left and 405 $\mathrm{mm}$ from the object center. As the area light sources, patterns are projected onto a square area of $387\mathrm{mm} \times 387\mathrm{mm}$.

**Surface normal estimation of various materials:** We




(a) (b) (c)

Figure 6: Real-world experiment and results. (a) Setup of the system. It is shown that the distance between the optical diffuser and the target object is close so that the system becomes compact. (b) Comparison with other NLPS techniques. #1 and #2 are diffuse, and #3 is specular material. The normals restored by each method agree closely, and the details are well reconstructed. (c) Normal estimation for real objects of different materials. Each object has a different material; diffuse plastic, glossy ceramic, and shiny ceramic from top to bottom. 3D reconstructions are shown on the rightmost column.

measure objects made of various materials. We use 16 patterns trained by the EP dataset. The setup of this experiment is slightly modified: the diffuser is placed in front of the object, the distance of the diffuser from the object is 400 mm, and the projected area of the pattern is 650mm x 650mm.

Figure 6b shows a comparison with other NLPS techniques. #1 is a sphere with a diffuse surface. The normals estimated by each technique are smooth and nearly identical. Material #2 shows weak sub-surface scattering. Each method correctly estimates the global shape. The proposed method estimates slightly smaller bumps on the nose and eyes, but this is due to the lack of scattering material in the training data. #3 is a specular material. [31] shows an overall shift of the normals, due to the presence of an occluding boundary on the top of the object, which negatively affected the optimization. [33] and the proposed method's normals are DL-based and the training data contains specular material, resulting in the correct normals. The estimation accuracy has been evaluated for #1 whose ground-truth normals is calculated by assuming it to be spheres. MAE of [31], [33], and the proposed method are 21.44, 19.76, and 14.04, respectively, which quantitatively confirms that the proposed method is the most accurate.

In the Fig. 6c, another object is shown with its shading image, the normals obtained by the proposed method, and the 3D shape estimated from the normals. For the 3D reconstruction, we use Poisson normal integration as implemented in [6]. Overall, the normals are correct and consistent, and there is no clear degradation in the accuracy of normal estimation, even if materials have glossy reflection, which is observed in #4 and #5. Because the shape variations are covered by the dataset, the bottom edges of the box are reproduced, as the two planes of the box are correctly reproduced with almost orthogonal to each other. Also, the result of #5 and #6 shows the ability the reproduction of detailed shapes. This is because the EP dataset used to train the model includes more shape variation. On the other hand, #5 shows distortions in the 3D reconstruction. This is because the 3D reconstruction assumes an orthographic projection model although the camera is a perspective projection in reality, and the normals are inconsistent at the boundaries where the normals are discontinuous.

## 6. Conclusion

In this paper, we proposed a method to estimate surface normals by the NLPS using distributed light patterns projected by a video projector onto the optical diffuser. The network, which consists of a differentiable renderer and a PS module, is designed to generate optimal distributed light patterns to estimate surface normals from a small number of measurements. We also construct a network where the PS module can learn surface normals from not only shading images, but also independent of light configuration parameters as inputs, which greatly generalizes the method. In the experiments, we demonstrated the effectiveness of the proposed method in both simulation and real data and confirmed that the proposed method is effective for various materials under low exposure conditions. Construction of optimal patterns for textured objects as well as high-frequency shape reconstruction NSPS is our future work.

## Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP20K19825, JP20H00611, JP18H04119 and JP21H01457.

## References


[1] Eevee renderer. https://docs.blender.org/manual/en/latest/render/eevee/. [accessed 1-March-2022].

[2] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*, 2008.

[3] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *PAMI*, 25(10), 2003.

[4] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72:239–257, 05 2007.

[5] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International journal of computer vision*, 35(1):33–44, 1999.

[6] Xu Cao, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Normal integration via inverse plane fitting with minimum point-to-plane distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2382–2391, 2021.

[7] Guanying Chen, Kai Han, and Kwan Yee K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *ECCV*, 2018.

[8] Lixiong Chen, Yinqiang Zheng, Boxin Shi, Art Subpa-asa, and Imari Sato. A Microfacet-based Model for Photometric Stereo with General Isotropic Reflectance. *PAMI*, 2019.

[9] Donghyeon Cho, Yasuyuki Matsushita, Yu-Wing Tai, and Inso Kweon. Photometric stereo under non-uniform light intensities and exposures. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 170–186, Cham, 2016. Springer International Publishing.

[10] Shuhei Hashimoto, Daisuke Miyazaki, and Shinsaku Hiura. Uncalibrated photometric stereo constrained by intrinsic reflectance image and shape from silhoutte. In *2019 16th International Conference on Machine Vision Applications (MVA)*, pages 1–6, 2019.

[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.

[12] Katsushi Ikeuchi. Determining Surface Orientations of Specular Surfaces by Using the Photometric Stereo Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(6), 1981.

[13] Katsushi Ikeuchi, Yasuyuki Matsushita, Ryusuke Sagawa, Daisuke Miyazaki, Yasuhiro Mukaigawa, Ryo Furukawa, and Hiroshi Kawasaki. *Active Lighting and Its Application for Computer Vision –40 Years of History of Active Lighting Techniques–*. Springer, 2020.

[14] Wenzel Jakob. Mitsuba renderer, 2010. http://www.mitsuba-renderer.org.

[15] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *ICLR*, 2017.

[16] Micah K. Johnson and Edward H. Adelson. Shape estimation in natural illumination. In *CVPR*, pages 2553–2560, 2011.

[17] Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. Efficient reflectance capture using an autoencoder. *TOG*, 37(4), 2018.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.

[19] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *CVPR*, 2019.

[20] 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*, 2021.

[21] Chao Liu, Srinivasa G. Narasimhan, and Artur W. Dubrawski. Near-light photometric stereo using circularly placed point light sources. In *IEEE International Conference on Computational Photography, ICCP 2018*, 2018.

[22] Fotios Logothetis, Roberto Mecca, Yvain Quéau, and Roberto Cipolla. Near-field photometric stereo in ambient light. 01 2016.

[23] Feng Lu, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, and Yoichi Sato. Uncalibrated photometric stereo for unknown isotropic reflectances. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2013.

[24] Roberto Mecca, Aaron Wetzler, Alfred M. Bruckstein, and Ron Kimmel. Near field photometric stereo with point light sources. *SIAM Journal on Imaging Sciences*, 7(4), 2014.

[25] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. In *IJCV*, volume 86, 2010.

[26] Zhipeng Mo, Boxin Shi, Feng Lu, Sai-Kit Yeung, and Yasuyuki Matsushita. Uncalibrated photometric stereo under natural illumination. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2936–2945, 2018.

[27] Yasuhiro Mukaigawa, Yasunori Ishii, and Takeshi Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America A*, 24(10), 2007.

[28] Shree K. Nayar and Mohit Gupta. Diffuse structured light. In *2012 IEEE International Conference on Computational Photography, ICCP 2012*, 2012.

[29] Shree K. Nayar, Katsushi Ikeuchi, and Takeo Kanade. Determining shape and reflectance of Lambertian, specular, and hybrid surfaces using extended sources. In *International Workshop on Industrial Applications of Machine Intelligence and Vision*, 1989.

[30] Shree K. Nayar and Arthur C. Sanderson. Determining Surface Orientations Of Specular Surfaces By Intensity Encoded Illumination. In *Optics, Illumination, and Image Sensing for Machine Vision II*, volume 0850, 1988.

[31] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, Francois Lauze, and Jean-Denis Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision*, 60, 03 2018.

[32] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep Photometric Stereo Network. In *Proceedings - 2017 IEEE International Conference*

*on Computer Vision Workshops, ICCVW 2017*, volume 2018-Janua, 2017.

[33] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo networks for determining surface normal and reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[34] Hiroaki Santo, Michael Waechter, and Yasuyuki Matsushita. Deep near-light photometric stereo for spatially varying reflectances. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 137–152, Cham, 2020. Springer International Publishing.

[35] Yoav Y. Schechner, Shree K. Nayar, and Peter N. Belhumeur. Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1339–1354, 2007.

[36] Grant Schindler. Photometric stereo via computer screen lighting for real-time surface reconstruction. In *4th International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2008 - Proceedings*, 2008.

[37] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):271–284, 2019.

[38] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *ECCV*, 2011.

[39] Tai Pang Wu, Kam Lun Tang, Chi Keung Tang, and Tien Tsin Wong. Dense photometric stereo: A Markov random field approach. *PAMI*, 28(11), 2006.

[40] Wuyuan Xie, Chengkai Dai, and Charlie C.L. Wang. Photometric stereo with near point lighting: A solution by mesh deformation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, 2015.

[41] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling Yu Duan, and Alex Kot. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In *ICCV*, 2019.