SPLINE-Net: Sparse Photometric Stereo through Lighting Interpolation and Normal Estimation Networks

Qian Zheng[1♯*]   Yiming Jia[2♯†]   Boxin Shi[3,4*]   Xudong Jiang[1]   Ling-Yu Duan[3,4]   Alex C. Kot[1]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Department of Precision Instrument, Tsinghua University, Beijing, China
[3]National Engineering Laboratory for Video Technology, Department of CS, Peking University, Beijing, China
[4]Peng Cheng Laboratory, Shenzhen, China
{zhengqian, exdjiang, eackot}@ntu.edu.sg, jiaym15@outlook.com, {shiboxin, lingyu}@pku.edu.cn

Abstract

This paper solves the Sparse Photometric stereo through Lighting Interpolation and Normal Estimation using a generative Network (SPLINE-Net). SPLINE-Net contains a lighting interpolation network to generate dense lighting observations given a sparse set of lights as inputs followed by a normal estimation network to estimate surface normals. Both networks are jointly constrained by the proposed symmetric and asymmetric loss functions to enforce isotropic constrain and perform outlier rejection of global illumination effects. SPLINE-Net is verified to outperform existing methods for photometric stereo of general BRDFs by using only ten images of different lights instead of using nearly one hundred images.

1. Introduction

The problem of photometric stereo [35] inversely solves for the radiometric image formation model to recover surface normals from different appearances of objects under various lighting conditions with a fixed camera view. The classic method [35] assumes an ideal Lambertian image formation model without global illumination effects (such as inter-reflection and shadows), which deviates from the realistic scenario and prevents photometric stereo from being able to handle real-world objects. To make photometric stereo practical, the major difficulties lie in dealing with objects of general reflectance and global illumination effects. These can be achieved by either exploring analytical Bidirectional Reflectance Distribution Function (BRDF) representations (*e.g.*, [14]) and general BRDF properties (*e.g.*, [30]) to model non-Lambertian interactions of lighting and surface normal or suppressing global effects by treating them as outliers (*e.g.*, [37]). Recently, deep learning
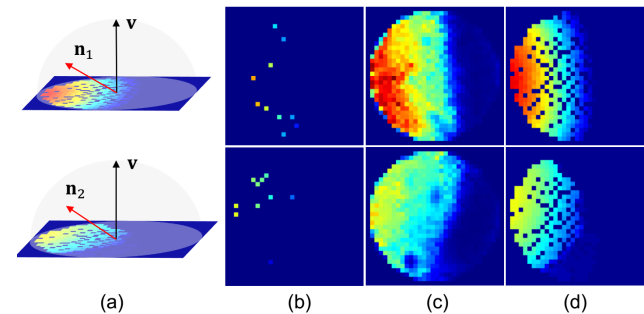

Figure 1. An illustration of observation maps corresponding to two surface normals (a brief introduction of observation maps can be found in Section 3.2 and [18]). (a) Two surface normals and their observation maps with dense lights, (b) sparse observation maps with 10 order-agnostic lights, (c) dense observation maps generated by our SPLINE-Net given sparse observation maps in (b) as inputs, and (d) ground truth of dense observation maps with 1000 lights. We use $\mathbf{v} = (0, 0, 1)^\top$ to represent viewing direction and $\mathbf{n}$ to represent surface normal in this paper.

based approaches are introduced to solve these difficulties by implicitly learning both the image formation process and global illumination effects from training data (*e.g.*, [9, 18]).

According to a comprehensive benchmark evaluation [28] (including quantitative results for representative methods published before 2016) and the additional results reported in most recent works [9, 18, 40], a moderately dense lighting distribution (*e.g.*, around 100 directional lights randomly sampled from the visible hemisphere) is required to achieve reasonably good normal estimation for objects with general materials (*e.g.*, angular error around $10°$ for a shiny plastic). This is because multi-illumination observations under a dense set of lights are required to fit the parameters in analytic BRDF models [14], to analyze general BRDF properties [30], to observe sufficient inliers and outliers [37], and to ensure the convergence of training neural networks [9]. How to achieve high accuracy in normal estimation for objects given general BRDFs with a sparse set of lights (*e.g.*, 10), which we call sparse photometric stereo in this paper, is still an open yet challenging problem [28].

In this paper, we propose to solve Sparse Photometric stereo through Lighting Interpolation and Normal Estimation Networks, namely SPLINE-Net. The SPLINE-Net is composed of two sub-networks: the Lighting Interpolation Network (LI-Net) to generate dense observations given a sparse set of input lights and the Normal Estimation Network (NE-Net) to estimate surface normal from the generated dense observations. LI-Net takes advantage of a learnable representation for dense lighting called *observation map* [18], and we propose to deal with sparse observation maps as damaged paintings and generate dense observation through inpainting (as shown in Figure 1[1]). NE-Net then follows LI-Net to infer surface normal guided by dense observation maps. To accurately guide the lighting interpolation and normal estimation specially under the photometric stereo context, we propose a *symmetric loss* and an *asymmetric loss* to explicitly consider general BRDF properties and outlier rejections. More specifically, the symmetric loss is derived according to the property of isotropy for general reflectance, which constrains pixel values on a generated observation map to be symmetrically distributed w.r.t. an axis determined by the corresponding surface normal. The asymmetric loss is derived from contaminated observation maps with global illumination effects, which constrains the difference between values of symmetrically distributed pixels to be equal to a non-zero amount. SPINE-Net is validated to achieve superior normal estimation accuracy given a small number of input images (*e.g.*, 10) comparing to state-of-the-art methods using a much larger number (*e.g.*, 96), which greatly relieves data capture and lighting calibration labor for photometric stereo with general BRDFs. The contributions of this paper are two-fold:

- We propose the SPLINE-Net to address the problem of photometric stereo with general BRDFs using a small number of images through an integrated learning procedure of lighting interpolation and normal estimation.

- We show how symmetric and asymmetric loss functions can be formulated to facilitate the learning of lighting interpolation and normal estimation with isotropy constraint and outlier rejection of global illumination effects considered.

2. Related Works

In this section, we briefly review traditional methods and deep learning based methods for non-Lambertian photometric stereo with general materials and known lightings. For other generalizations of photometric stereo, we refer readers to survey papers in [15, 1, 28].

---

[1]Note that holes in ground truth of observation maps are produced by the discrete projection from a limited number of lights to a grid with limited resolution (*e.g.*, from 1000 lighting directions to observation maps with size 32×32 in this figure).

Traditional methods. The classical method [35] for the problem of photometric stereo is to assume a Lambertian surface reflectance and recover surface normals pixelwisely. Such an assumption is too strong to provide an accurate recovery in real-world due to densely observed non-Lambertian reflectance caused by materials with diverse reflectance and global illumination effects. In order to address non-Lambertian reflectance from broad classes of materials, modern algorithms attempt to use a mathematically tractable form to describe BRDFs. Analytic models exploit all available data to fit a nonlinear analytic BRDF, such as Blinn-Phong model [33], Torrance-Sparrow model [13], Ward model and its variations [12, 14, 2], specular spike model [11, 38], and microfacet BRDF model [10]. Empirical models consider general properties of a BRDF, such as isotropy, monotonicity. Some basic derivations for isotropy BRDFs are provided in [4, 31, 7]. Excellent performance has been achieved by methods based on empirical models, including combining isotropy and monotonicity with visibility constraint [16], using isotropic constraint for the estimation of elevation angle [3, 29, 22], and approximating isotropic BRDFs by bivariate functions [25, 19, 30]. However, most of these methods based on analytic models and empirical models are pixel-wise so that they cannot explicitly consider global illumination effects such as interreflection and cast shadows. Outlier rejection based methods are developed to suppress global illumination effects by considering them as outliers. Earlier works select a subset of Lambertian images from inputs for the accurate recovery of surface normals [24, 34, 23, 39, 37]. Recent methods apply robust analysis by assuming non-Lambertian reflectance is sparse [36, 20]. However, these methods still rely on the existence of a dominant Lambertian reflectance component.

Deep learning based methods. Recently, with the great success in both high-level and low-level computer vision tasks achieved by neural networks, researchers have introduced deep learning based methods to solve the problem of photometric stereo. Instead of explicitly modeling image formation process and global illumination effects as in traditional methods, deep learning based methods attempt to learn such information from data. DPSN [26] is the first attempt and it uses a deep fully-connect network to regress surface normals from given observations captured under pre-defined lightings in a supervised manner. However, pre-definition of lightings limits its practicality for photometric stereo where the number of inputs often varies. PS-FCN [9] is proposed to address such a limitation and handle images under various lightings in an order-agnostic manner by aggregating features of inputs using the max-pooling operation. CNN-PS [18] is another work to accept order-agnostic inputs by introducing observation map, which is a fixed shape representation invariant to inputs. Besides neural networks trained in a supervised manner, Taniai and
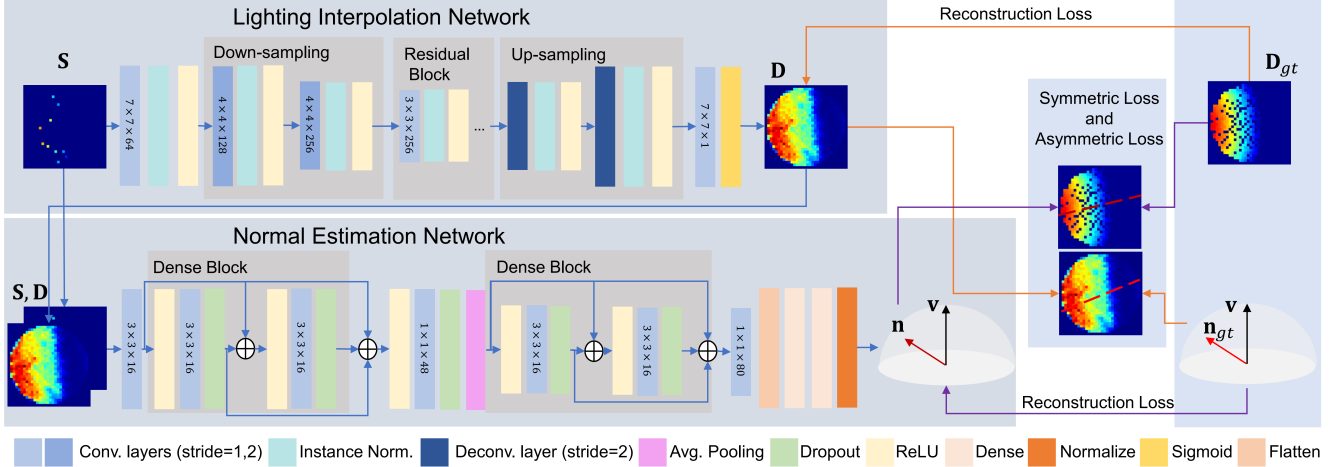
Figure 2. The framework of the proposed SPLINE-Net. The lighting interpolation network generates dense observation maps $\mathbf{D}$ given sparse observation maps $\mathbf{S}$ as inputs. The normal estimation network estimates surface normals $\mathbf{n}$ given $\mathbf{S}$ and $\mathbf{D}$ as inputs. Both networks are trained in a supervised manner where ground truth of observation maps $\mathbf{D}_{gt}$ and surface normals $\mathbf{n}_{gt}$ are known.

Maehara [32] presented an unsupervised learning framework where surface normals and BRDFs are recovered by minimizing a reconstruction loss between inputs and images synthesized based on a rendering equation.

Only a few earlier works address the problem of photometric stereo with general reflectance using a small number of images in the literature (*e.g.*, analytic model based method [14], shadow analysis based method [6]). Our paper revisits this problem due to its low costs for the labor of data capture and lighting calibration.

3. The Proposed SPLINE-Net

In this section, we introduce our solution to the problem of photometric stereo with general reflectance using a small number of images. We first present the framework of our SPLINE-Net in Section 3.1. Then we detail the symmetric loss and the asymmetric loss in Section 3.2.

3.1. Framework

As illustrated in Figure 2, our SPLINE-Net, which consists of a Lighting Interpolation Network (LI-Net) and a Normal Estimation Network (NE-Net), is optimized in a supervised manner. LI-Net (represented as a regression function f) interpolates dense observation maps $\mathbf{D}$ from sparse observation maps $\mathbf{S}$ (*i.e.*, sparse sets of lights),

$$f : \mathbf{S} \rightarrow \mathbf{D}. \tag{1}$$

Such densely interpolated observation maps $\mathbf{D}$ are then concatenated to original inputs $\mathbf{S}$ and help estimate surface normals $\mathbf{n}$ in NE-Net (represented as a regression function g)

$$g : \mathbf{S}, \mathbf{D} \rightarrow \mathbf{n}. \tag{2}$$

LI-Net and NE-Net are trained in an alternating iteratively manner, where fixing one network when optimizing the other. Specifically, we update LI-Net once after updating NE-Net five times. The loss function for each network composes of a reconstruction loss, a symmetric loss, and an asymmetric loss.

Lighting Interpolation Network. Inspired by a recent work [8] that suggests inferring lighting information as intermediate results can benefit the estimation of the surface normal, we design the LI-Net to generate observation maps regarding dense lighting directions. The basic idea is to in-paint sparse observation maps and obtain dense ones, based on learnable properties (*e.g.*, spatial continuity). LI-Net is designed using an encoder-decoder structure due to its excellent image generation capacity [21, 41]. The loss function of LI-Net $\mathcal{L}_f$ is formulated as,

$$\mathcal{L}_f = \mathcal{L}_f^{rec} + \lambda_s \mathcal{L}_f^s + \lambda_a \mathcal{L}_f^a. \tag{3}$$

The reconstruction loss $\mathcal{L}_f^{rec}$ is defined as[2],

$$\mathcal{L}_f^{rec} = \arccos(\mathbf{n}^\top \mathbf{n}_{gt}) + |\mathbf{D} - \mathbf{D}_{gt}|_1 + |\mathbf{M}_s \circ (\mathbf{D} - \mathbf{D}_{gt})|_1, \tag{4}$$

where $\mathbf{D} = f(\mathbf{S}), \mathbf{n} = g(\mathbf{S}, \mathbf{D})$, $\mathbf{n}_{gt}$ and $\mathbf{D}_{gt}$ are ground truth of a surface normal and its corresponding dense observation map, respectively, $\mathbf{M}_s$ is a binary mask indicating positions of non-zero value of $\mathbf{S}$, $\circ$ represents element-wise multiplication. $\mathcal{L}_f^s$ and $\mathcal{L}_f^a$ are our symmetric and asymmetric loss to be introduced in Section 3.2.

Normal Estimation Net. We use the same architecture as that in [18] (a variation of DenseNet [17]) for NE-Net due to its excellent capacity to model the relation between observation maps and surface normals. The loss function of NE-Net is formulated as,

$$\mathcal{L}_g = \mathcal{L}_g^{rec} + \lambda_s \mathcal{L}_g^s + \lambda_a \mathcal{L}_g^a, \tag{5}$$

---

[2]We experimentally find L1 and L2 distances provide similar results and here we compute reconstruction loss $\mathcal{L}_f^{rec}$ using L1 distance.
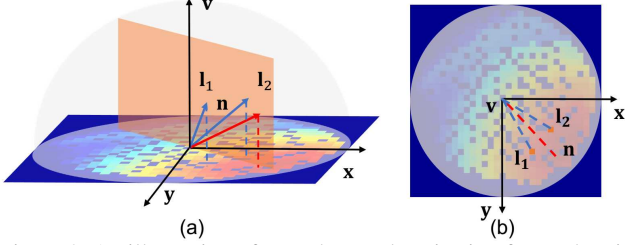
Figure 3. An illustration of an orthogonal projection from a hemisphere surface to its base (gray) and an interpretation of isotropy for a dense observation map. (a) Front view: $\mathbf{l}_1$ and $\mathbf{l}_2$ represent two lighting directions which are symmetric w.r.t. the plane spanned by viewing direction $\mathbf{v}$ and surface normal $\mathbf{n}$ (orange plane). (b) Top view: irradiance values (orange dots), which are projected by $\mathbf{l}_1$ and $\mathbf{l}_2$, are numerically equal due to isotropy.

where $\mathcal{L}_g^s$ and $\mathcal{L}_g^a$ are symmetric loss and asymmetric loss, and reconstruction loss is,

$$\mathcal{L}_g^{rec} = \arccos(\mathbf{n}^\top \mathbf{n}_{gt}). \qquad (6)$$

3.2. Symmetric Loss and Asymmetric Loss

In this section, we first revisit the observation map from [18]. Then, we further investigate its characteristics by considering isotropy and global illumination effects. Finally, we present symmetric and asymmetric loss functions.
Observation maps. As introduced in [18], each point on a surface normal map corresponds to an observation map (as shown in Figure 3 (a)). Elements on such a map describe observed irradiance values under different lighting directions. These lighting directions are mapped to positions of elements, which is an orthogonal projection. As illustrated in Figure 3 (a), a dense observation map can be regarded as generated by projecting a hemisphere surface to its base plane, where each point on the hemisphere surface represents a direction of lighting and its projecting value describes an observed irradiance value under such a light. Such a projecting relation motivates us to introduce isotropy to narrow the solution space of our SPLINE-Net.
Isotropic BRDFs and global illumination effects. Isotropy BRDFs $\rho(\mathbf{n}^\top \boldsymbol{\ell}, \mathbf{n}^\top \mathbf{v}, \mathbf{v}^\top \boldsymbol{\ell})$ for general materials have the property that reflectance values are numerically equal if lighting directions are symmetric about the plane spanned by $\mathbf{v}$ and $\mathbf{n}$ as shown in Figure 3 (a). Considering the relation of the one-to-one mapping between lighting directions and positions of observed irradiance values, these values are numerically equal if their positions are symmetrically distributed regarding to the axis projected by surface normals in observation maps, as shown by Figure 3 (b). However, such symmetric patterns can be destroyed by global illumination effects because observation maps are pixel-wisely generated. Therefore, unpredictable shapes produce cast shadows or inter-reflection which can
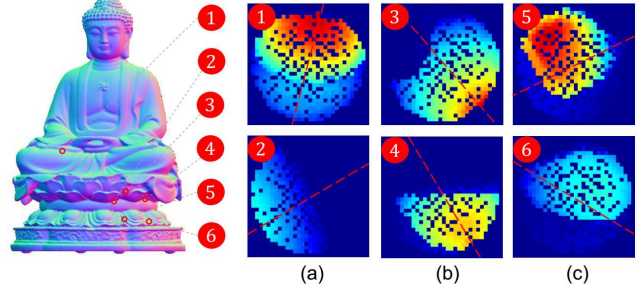

Figure 4. Six dense observation maps of data BUDDHA from CY-CLESPS [18] with (a) isotropic reflectance, (b) cast shadows, and (c) inter-reflection. Red-dotted lines indicate directions of their corresponding surface normals.

lead to sudden changes of irradiance values on observation maps. Figure 4 illustrates examples of isotropy, cast shadow, inter-reflection.
Symmetric and asymmetric loss functions. In order to further narrow solution spaces for the lighting interpolation of dense observation maps to facilitate accurate estimation of the surface normal, we propose symmetric and asymmetric loss functions to exploit above observations for LI-Net and NE-Net. More specifically, given a dense observation map $\mathbf{D}$ and its corresponding surface normal $\mathbf{n}$, the symmetric loss forces the isotropic properties of general BRDFs which are valid on various real-world reflectance. That is, it constrains irradiance values, which are symmetrically distributed w.r.t. an axis determined by its surface normal (red-dotted lines in Figure 4), to be numerically equal,

$$\mathcal{L}_s = \mathcal{L}_s(\mathbf{D}, \mathbf{n}) = |\mathbf{D} - r(\mathbf{D}, \mathbf{n})|_1, \qquad (7)$$

where function $r(\mathbf{D}, \mathbf{n})$ mirrors the observation map $\mathbf{D}$ w.r.t. the axis determined by $\mathbf{n}$. Different from symmetric loss, the asymmetric loss models the asymmetric pattern brought by outliers such as global illuminations. It constrains the difference between values of symmetrically distributed pixels to be equal to a non-zero amount η,

$$\begin{aligned}
\mathcal{L}_a = \mathcal{L}_a(\mathbf{D}, \mathbf{n}) = & \big||\mathbf{D} - r(\mathbf{D}), \mathbf{n}|_1 - \eta\big|_1 \\
& + \lambda_c \big||p(\mathbf{D}) - r(p(\mathbf{D}), \mathbf{n})|_1 - \eta\big|_1,
\end{aligned} \qquad (8)$$

where λ_c is a weight parameter, function $p(\cdot)$ performs an average pooling operation with stride of 2 to ensure spatial continuity of observation maps. Empirically, we set $\eta = 1, \lambda_c = 50$ for all experiments. Both of $\mathcal{L}_s$ and $\mathcal{L}_a$ aim to better fit observations of symmetric and asymmetric patterns (as illustrated in Figure 4) during training. We integrate symmetric and the asymmetric loss functions to optimize LI-Net by setting,

$$\begin{aligned}
\mathcal{L}_f^s &= \mathcal{L}_s(\mathbf{D}, \mathbf{n}_{gt}), \\
\mathcal{L}_f^a &= \mathcal{L}_a(\mathbf{D}, \mathbf{n}_{gt}),
\end{aligned} \qquad (9)$$

Table 1. Quantitative comparisons in terms of angular error on CYCLESPS-TEST dataset [18]. Results of three shapes (PAPERBOWL, SPHERE, TURTLE) with metallic (M) and specular (S) materials are reported. Note that all results are averaged over 100 random trials. For subset L17 (left) and subset L305 (right), results over 6 different data are averaged (Avg.) for each method.

	PAPERBOWL		SPHERE		TURTLE		Avg.	PAPERBOWL		SPHERE		TURTLE		Avg.
	M	S	M	S	M	S		M	S	M	S	M	S	
LS [35]	41.47	35.09	18.85	10.76	27.74	19.89	25.63	43.09	37.36	20.19	12.79	28.51	21.76	27.28
IW12 [20]	46.68	33.86	16.77	2.23	31.83	12.65	24.00	48.01	37.10	21.93	3.19	34.91	16.32	26.91
ST14 [30]	42.94	35.13	22.58	4.18	34.30	17.01	26.02	44.44	37.35	25.41	4.89	36.01	19.06	27.86
IA14 [19]	48.25	43.51	18.62	11.71	30.59	23.55	29.37	49.01	45.37	21.52	13.63	32.82	26.27	31.44
CNN-PS [18]	37.14	23.40	17.44	6.99	22.86	10.74	19.76	38.45	26.90	18.25	9.04	23.91	14.36	21.82
SPLINE-Net	29.87	18.65	6.59	3.82	15.07	7.85	13.64	33.99	23.15	9.21	6.69	17.35	12.01	17.07

and to optimize NE-Net by setting,

$$\mathcal{L}_g^s = \mathcal{L}_s(\mathbf{D}_{gt}, \mathbf{n}),$$
$$\mathcal{L}_g^a = \mathcal{L}_a(\mathbf{D}_{gt}, \mathbf{n}). \quad (10)$$

We empirically set $\lambda_s = 2 \times 10^{-2}, \lambda_a = 2 \times 10^{-5}$ due to the fact that global illumination effects are always observed for small regions of real-world objects.

4. Experiments

Settings and implementation details. A recent survey work [28] implies that photometric stereo with general BRDFs shows significant performance drop if only around 10 images are provided. Therefore, we define 10 as the number of sparse lights and use 10 randomly sampled lights as inputs for both training and testing. The resolution of our observation map is set to 32×32 and the batch size is set to 128, which are the same as those in [18], for easier comparisons. Adam optimizer is used to optimize our networks with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

Datasets and evaluation. We use CYCLESPS data provided by [18] as our training data. There are 45 training data including 15 shapes with 3 categories of reflectance (diffuse, metallic, and specular). Our testing sets are built based on public evaluation datasets. To cover as many lighting conditions as possible, we construct 100 instances for each testing data from these datasets and each instance contains images illuminated under 10 randomly selected lights. Quantitative results are averaged of mean angular errors and metric used in the paper is the angular error in degrees (unit is omitted in all tables).

Compared methods. We compare with five methods, including the least squares based Lambertian photometric stereo method LS [35], an outlier rejection based method IW12 [20], two state-of-the-art methods exploring general BRDF properties ST14 [30] and IA14 [19], and a deep learning based method CNN-PS [18]. We re-train CNN-PS [18] by taking 10 observed irradiance values as inputs.[3]

---

[3]Considering the overall quantitative results of CNN-PS [18] with default settings (taking 50 observed irradiance values as inputs) on CYCLESPS-TEST [18] (L17: 31.08° and L305: 34.90°) and DILIGENT [28] (14.10°), we report results re-trained by our setting. PS-FCN [9] is not compared due to different requirements of training data.

4.1. Synthetic Data

CYCLESPS-TEST is a testing dataset from [18], which consists of two subsets (denoted as 'L17' and 'L305'). Each of subsets contains three shapes and these shapes are rendered by reflectance of 'metallic' and 'specular'.[4] For all the 12 data, we construct 100 instances (each with 10 randomly selected images) to build our testing set.

As can be observed from Table 1, metallic materials are more challenging as compared with specular materials. Even for the simple shape (SPHERE) containing few global illumination effects, all methods fail to estimate accurate surface normals for metallic materials. The performance advantage of our method is superior, *i.e.*, the overall performance is much better than the second best one. Interestingly, two traditional methods, IW12 [20] and ST14 [30], outperform other methods on data SPHERE with specular materials. However, their performance degrades for complex shapes (PAPERBOWL and TURTLE), while our method consistently achieves the best performance.

Visual comparisons in Figure 5 further validates the effectiveness of our method. It deals with specular reflectance more robustly (SPHERE) and consistently produces the best estimation for most regions on a more complex shape (PAPERBOWL). The superior performance shows the effectiveness of our method to address the problem of photometric stereo with general BRDFs using a small number of images.

4.2. Real Data

The benchmark dataset DILIGENT [28] consists of 10 real data. Similarly, for each of these 10 data, we construct 100 instances, each of which contains 10 randomly selected images, to build the testing set.

As can be found from quantitative results in Table 2, our method demonstrates obvious superiority for most data except for BALL and POT1, which get similar or worse results as compared to two traditional methods LS [35] and IW12 [20]. The reason is these two data are diffuse-dominant so that traditional methods with Lambertian assumption fit well even for a small number of observed irradiance values. However, our data-driven approach considers general reflectance and global illumination effects

---

[4]'Metallic' and 'specular' reflectance are generated by controlling 11 parameters from Disney's principled BSDF [5], which are detailed in [18].
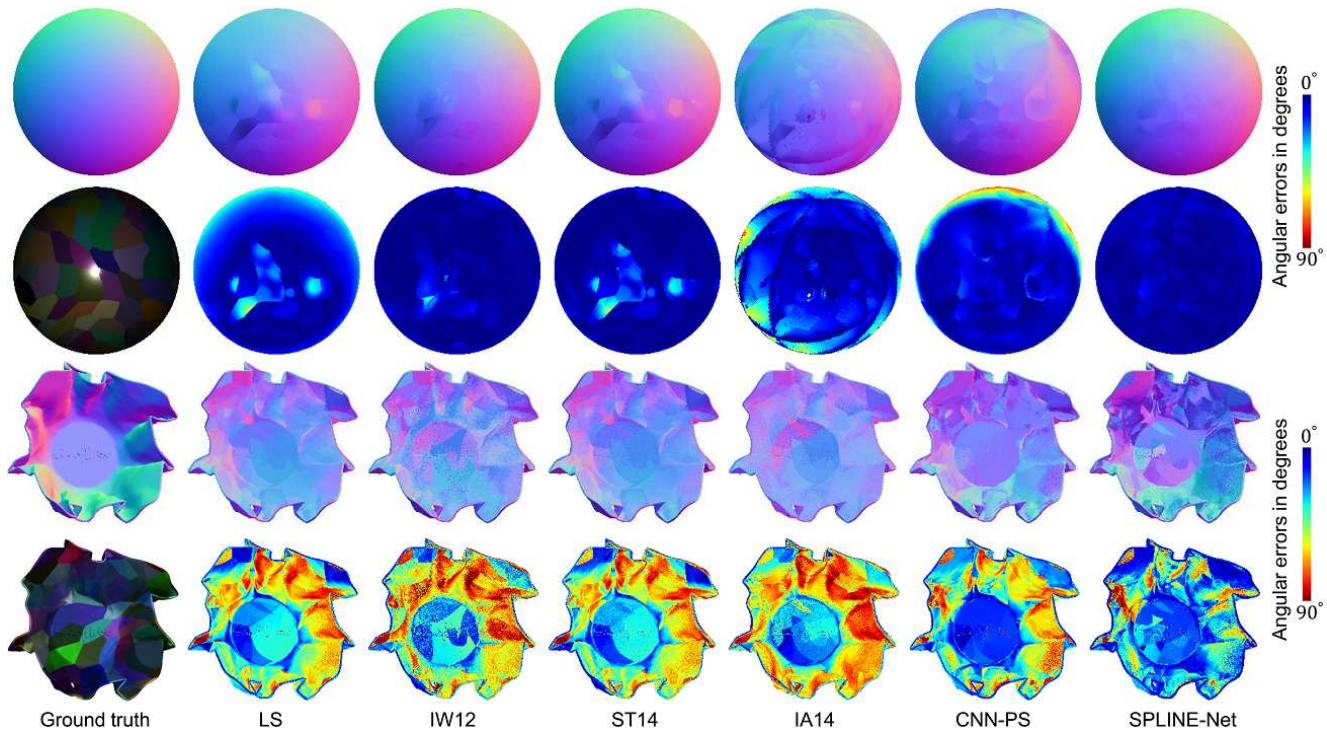
Figure 5. Comparisons of normal maps and angular error maps (in degrees) on SPHERE with specular materials (top) and PAPERBOWL with metallic materials (bottom) from L17, CYCLESPS-TEST [18].

Table 2. Quantitative comparisons in terms of angular error on DiLiGenT dataset [28]. Note that all results are averaged over 100 random trials. Results over 10 different data are averaged (Avg.) for each method.

Methods	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
LS [35]	4.41	9.05	15.62	9.03	26.42	19.59	31.31	9.46	15.37	20.16	16.04
IW12 [20]	**3.33**	7.62	13.36	8.13	25.01	18.01	29.37	8.73	14.60	16.63	14.48
ST14 [30]	5.24	9.39	15.79	9.34	26.08	19.71	30.85	9.76	15.57	20.08	16.18
IA14 [19]	12.94	16.40	20.63	15.53	18.08	18.73	32.50	**6.28**	14.31	24.99	19.04
CNN-PS [18]	17.86	13.08	19.25	15.67	19.28	21.56	21.52	16.95	18.52	21.30	18.50
Nets *w/o* loss	6.06	7.01	10.69	8.38	10.39	11.37	**19.02**	9.42	12.34	16.18	11.09
Nets with $\mathcal{L}^s$	5.04	**5.89**	10.11	7.79	9.38	10.84	19.03	8.91	**11.47**	**15.87**	10.43
SPLINE-Net	4.96	5.99	**10.07**	**7.52**	**8.80**	**10.43**	19.05	8.77	11.79	16.13	**10.35**

evenly during model optimization and hence may underfit Lambertian surfaces with simple shapes. Unlike excellent performance on synthetic data, CNN-PS [18] achieve less accuracy on real data. We think that the reason is mainly due to the problem of overfitting during training, *i.e.*, synthetic data for testing are constructed in a similar manner as training data. Our method achieves the best accuracy for most of data, such as COW (metallic paint materials) and GOBLET, HARVEST (most of regions contain interreflection or cast shadows).

Visual comparisons on data COW and POT1 are shown in Figure 6. Our method provides much more accurate results for metallic materials (COW) which is consistent with the results of synthetic data (Table 1). Most of the compared methods and our method achieve accurate estimation for center regions of POT1, however, our method gets a significant advantage for boundaries (*e.g.*, regions of spout and kettle-holder) containing inter-reflection or cast shadows.

Table 3. The performance of SPLINE-Net on DiLiGenT [28] against different standard deviations (σ) of Gaussian noise ($\mu = 0$) added to lighting directions of inputs.

σ (in degrees)	0 (no noise)	2	4	6	8
Performance	10.35	11.21	12.14	13.76	15.94

4.3. Robustness against Noise

Lights. It is inevitable to introduce errors during lighting calibration in practice. To show the robustness of our method against errors in light source directions, we manually add noise to lights in DiLiGenT [28] and test the quantitative performance regarding different levels of noise. As can be observed in Table 3, our method is able to produce accurate results when the noise is small ($\sigma = 2$). However, its performance drops with higher levels of noise. The robustness of our method to small noise is mainly due to our training data, *i.e.*, small random noise is added to each light when creating training dataset as introduced in [18].
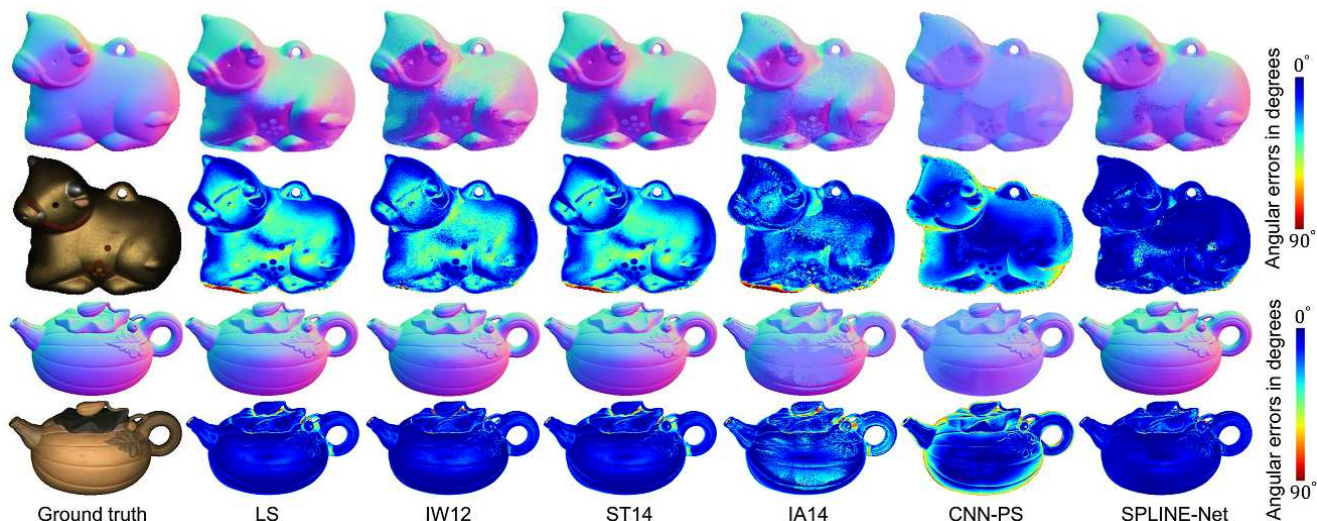
Figure 6. Comparisons of normal maps and angular error maps (in degrees) on COW (top) and POT1 (bottom) from DiLiGenT [28].
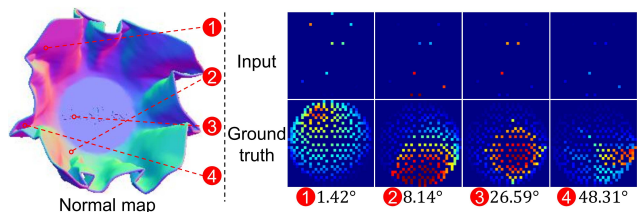

Figure 7. Left: a normal map and its four points. Right: four pairs of observation maps (inputs and their ground truth) corresponding to points on the left; numbers below are errors of estimated surface normal given corresponding inputs. Zoom in for a better view.

Inputs. Because of the sparsity of inputs, the performance of our method is affected by inputs due to outliers resulting from inter-reflection and shadows. Figure 7 (right) illustrates this consideration. As can be found, when the number of outliers of inputs increases (the number of dark dots in inputs increases), the error becomes larger accordingly.

4.4. Ablation Studies

In this section, we perform ablation studies to further investigate the contribution of important components in SPLINE-Net. Considering the same network structure of NE-Net and that in CNN-PS [18] and the fact that our SPLINE-Net is composed of LI-Net and NE-Net, we compare our SPLINE-Net without symmetric loss or asymmetric loss (denoted as 'Nets *w/o* loss') with CNN-PS [18] to validate the effectiveness of LI-Net. By comparing the performance of Nets *w/o* loss with that with an additional symmetric loss (denoted as 'Nets with $\mathcal{L}^s$'), we validate the effectiveness of enforcing isotropy property. The comparison performed between Nets with $\mathcal{L}^s$ and SPLINE-Net is to verify the effectiveness of the consideration of global illumination effects. The same settings and the same testing set as those in Section 4.2 is used for evaluation in this section.

The quantitative performance are reported in Table 2. As can be observed, Nets *w/o* loss significantly outperforms
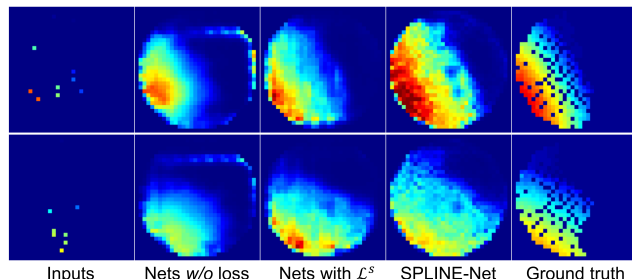

Figure 8. An illustration of observation maps. From left to right columns: inputs (10 order-agnostic lights), maps generated by SPLINE-Net *w/o* our loss, by SPLINE-Net with symmetric loss $\mathcal{L}^s$, by SPLINE-Net, and ground truth (1000 lights).

CNN-PS [18], which verifies the effectiveness of using LI-Net to help the estimation of surface normals and prevent overfitting of directly regressing sparse observation maps to surface normals. The proposed symmetric loss and asymmetric loss help improve overall performance.

An illustration of observation maps generated by our method using different settings is displayed in Figure 8. Our methods (Nets with $\mathcal{L}^s$ and SPLINE-Net) successfully inpaint regions damaged by global illumination effects and it tends to produce smooth ones even if the ground truth is not smooth. Visual comparisons in Figure 9 intuitively shows the advantages of each component of our method to facilitate accurate estimation of surface normals.

5. Conclusion

This paper proposes SPLINE-Net to address the problem of photometric stereo with general reflectance and global illumination effects using a small number of images. The basic idea of SPLINE-Net is to generate dense lighting observations from a sparse set of lights to guide the estimation of surface normals. The proposed SPLINE-Net is further constrained by the proposed symmetric and asymmetric loss
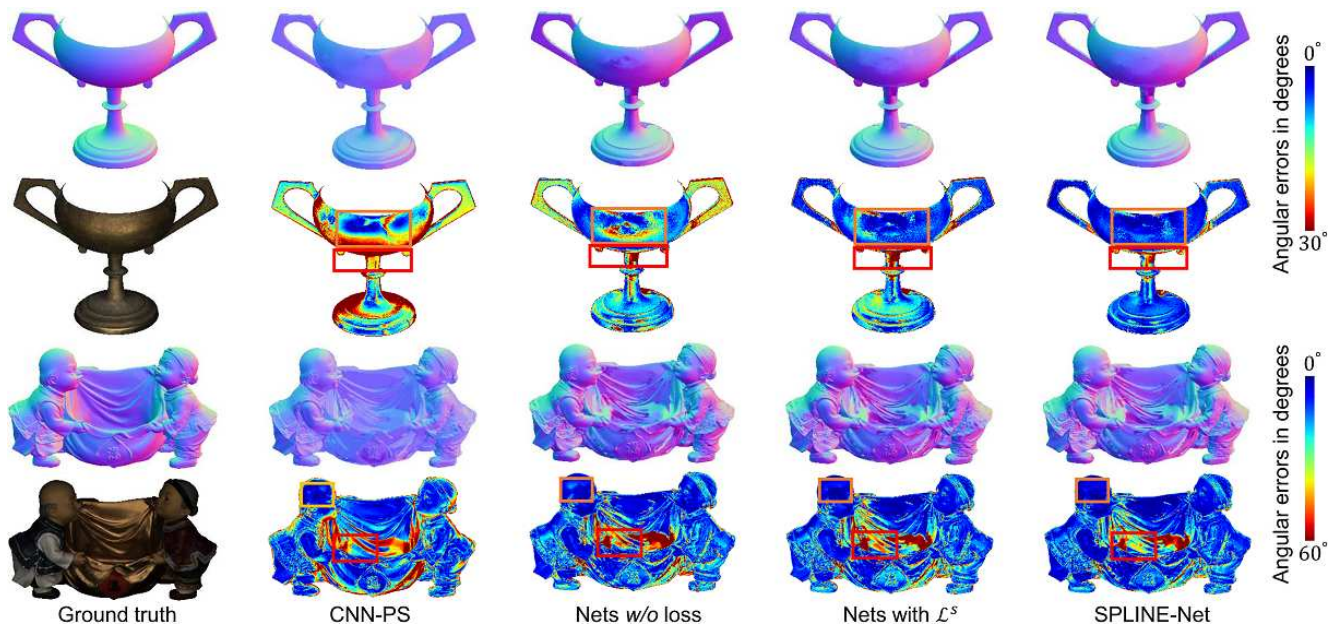
Figure 9. Comparisons of normal maps and angular error maps (in degrees) on GOBLET (top) and READING (bottom) from DiLiGenT [28].
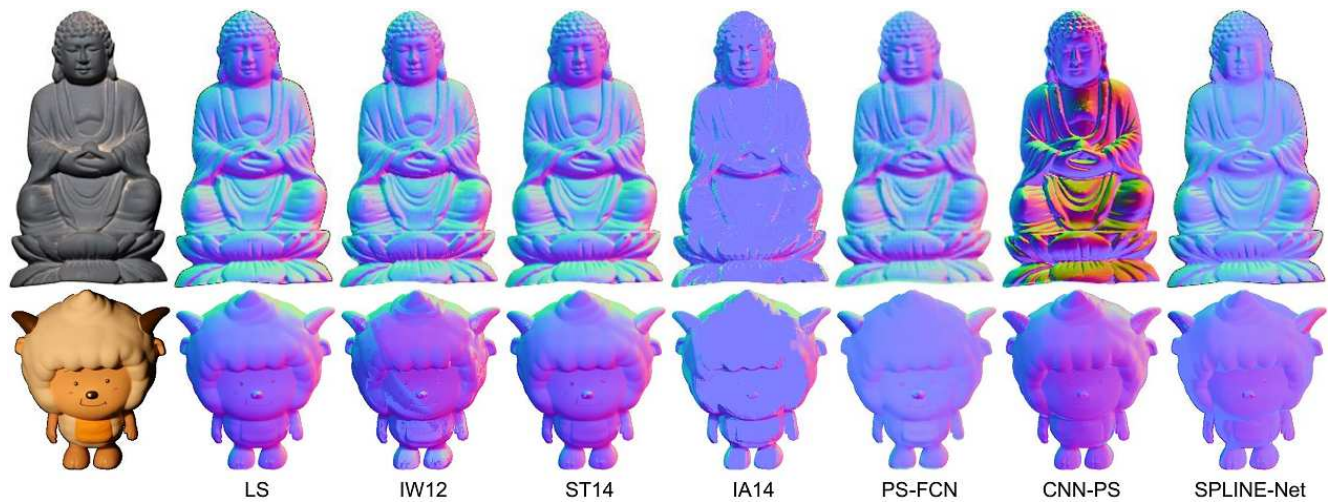

Figure 10. Comparisons of normal maps on diffuse-dominant data BUDDHA (top) and SHEEP (bottom).

functions to enforce isotropic constrain and perform outlier rejection of global illumination effects.

Limitations. Interestingly, even though deep learning based methods achieve superior performance for non-Lambertian reflectance, their performance drops for diffuse-dominant surfaces that can be well fitted by traditional methods with Lambertian assumption. Figure 10 illustrates results of four traditional methods and three deep learning based methods (including PS-FCN [9]) on two real data[5] with diffuse surfaces. Such results, which are consistent with those of BALL and POT1 in Table 2, indicate the limitation of deep learning methods for diffuse surfaces. To ex-

plicitly consider diffuse surface at the same time maintain the performance advantage on non-Lambertian surfaces for deep learning based methods can be one of further works.

Acknowledgment

---

[5]BUDDHA is courtesy of Dan Goldman and Steven Seitz (found from http://www.cs.washington.edu/education/courses/csep576/05wi//projects/project3). SHEEP is from [27].

References

[1] Jens Ackermann and Michael Goesele. A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision*, 9(3-4):149–254, 2015. 2

[2] Jens Ackermann, Fabian Langguth, Simon Fuhrmann, and Michael Goesele. Photometric stereo for outdoor webcams. In *Proc. of Computer Vision and Pattern Recognition*, pages 262–269, 2012. 2

[3] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[4] Neil G Alldrin and David J Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Proc. of Internatoinal Conference on Computer Vision*, pages 1–8, 2007. 2

[5] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney, part of practical physically based shading in film and game production. SIGGRAPH, 2012. 5

[6] Manmohan Chandraker, Sameer Agarwal, and David Kriegman. Shadowcuts: Photometric stereo with shadows. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2007. 3

[7] Manmohan Chandraker, Jiamin Bai, and Ravi Ramamoorthi. On differential photometric reconstruction for unknown, isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2941–2955, 2013. 2

[8] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *Proc. of Computer Vision and Pattern Recognition*, pages 8739–8747, 2019. 3

[9] Guanying Chen, Kai Han, and Kwan-Yee K Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proc. of European Conference on Computer Vision*, pages 3–19, 2018. 1, 2, 5, 8

[10] Lixiong Chen, Yinqiang Zheng, Boxin Shi, Art Subpa-Asa, and Imari Sato. A microfacet-based reflectance model for photometric stereo with highly specular surfaces. In *Proc. of Internatoinal Conference on Computer Vision*, pages 3162–3170, 2017. 2

[11] Tongbo Chen, Michael Goesele, and H-P Seidel. Mesostructure from specularity. In *Proc. of Computer Vision and Pattern Recognition*, volume 2, pages 1825–1832, 2006. 2

[12] Hin-Shun Chung and Jiaya Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[13] Athinodoros S Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. of Internatoinal Conference on Computer Vision*, volume 3, page 816, 2003. 2

[14] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010. 1, 2, 3

[15] Steffen Herbort and Christian Wöhler. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research*, 2(3):1–17, 2011. 2

[16] Tomoaki Higo, Yasuyuki Matsushita, and Katsushi Ikeuchi. Consensus photometric stereo. In *Proc. of Computer Vision and Pattern Recognition*, pages 1157–1164, 2010. 2

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proc. of Computer Vision and Pattern Recognition*, pages 4700–4708, 2017. 3

[18] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. of European Conference on Computer Vision*, pages 3–18, 2018. 1, 2, 3, 4, 5, 6, 7

[19] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proc. of Computer Vision and Pattern Recognition*, pages 2179–2186, 2014. 2, 5, 6

[20] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *Proc. of Computer Vision and Pattern Recognition*, pages 318–325, 2012. 2, 5, 6

[21] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. of European Conference on Computer Vision*, pages 694–711. Springer, 2016. 3

[22] Si Li and Boxin Shi. Photometric stereo for general isotropic reflectances by spherical linear interpolation. *Optical Engineering*, 54(8):083104, 2015. 2

[23] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2-3):229, 2010. 2

[24] Yasuhiro Mukaigawa, Yasunori Ishii, and Takeshi Shakunaga. Analysis of photometric factors based on photometric linearization. *JOSA A*, 24(10):3326–3334, 2007. 2

[25] Fabiano Romeiro, Yuriy Vasilyev, and Todd Zickler. Passive reflectometry. In *Proc. of European Conference on Computer Vision*, pages 859–872, 2008. 2

[26] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Computer Vision Workshop (ICCVW), IEEE International Conference on*, pages 501–509, 2017. 2

[27] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *Proc. of Computer Vision and Pattern Recognition*, pages 1118–1125, 2010. 8

[28] 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. 1, 2, 5, 6, 7, 8

[29] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Elevation angle from reflectance monotonicity: Photometric stereo for general isotropic reflectances. In *Proc. of European Conference on Computer Vision*, pages 455–468. Springer, 2012. 2

[30] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1078–1091, 2014. 1, 2, 5, 6

[31] Ping Tan, Long Quan, and Todd Zickler. The geometry of reflectance symmetries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2506–2520, 2011. 2

[32] Tatsunori Taniai and Takanori Maehara. Neural inverse rendering for general reflectance photometric stereo. In *International Conference on Machine Learning*, pages 4864–4873, 2018. 3

[33] Silvia Tozza, Roberto Mecca, M Duocastella, and A Del Bue. Direct differential photometric stereo shape recovery of diffuse and specular surfaces. *Journal of Mathematical Imaging and Vision*, 56(1):57–76, 2016. 2

[34] Frank Verbiest and Luc Van Gool. Photometric stereo with coherent outlier handling and confidence estimation. In *Proc. of Computer Vision and Pattern Recognition*, pages 1–8, 2008. 2

[35] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):191139–191139, 1980. 1, 2, 5, 6

[36] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. of Asian Conference on Computer Vision*, pages 703–717. 2010. 2

[37] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):546–560, 2010. 1, 2

[38] Sai-Kit Yeung, Tai-Pang Wu, Chi-Keung Tang, Tony F Chan, and Stanley J Osher. Normal estimation of a transparent object using a video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):890–897, 2015. 2

[39] Chanki Yu, Yongduek Seo, and Sang Wook Lee. Photometric stereo from maximum feasible Lambertian reflections. In *Proc. of European Conference on Computer Vision*, pages 115–126, 2010. 2

[40] Qian Zheng, Ajay Kumar, Boxin Shi, and Gang Pan. Numerical reflectance compensation for non-lambertian photometric stereo. *IEEE Transactions on Image Processing*, 2019. 1

[41] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. of Internatoinal Conference on Computer Vision*, pages 2223–2232, 2017. 3