Photometric Stereo via Discrete Hypothesis-and-Test Search

Kenji Enomoto[1] Michael Waechter[1] Kiriakos N. Kutulakos[2] Yasuyuki Matsushita[1]
[1]Osaka University [2]University of Toronto

Abstract

In this paper, we consider the problem of estimating surface normals of a scene with spatially varying, general BRDFs observed by a static camera under varying, known, distant illumination. Unlike previous approaches that are mostly based on continuous local optimization, we cast the problem as a discrete hypothesis-and-test search problem over the discretized space of surface normals. While a naïve search requires a significant amount of time, we show that the expensive computation block can be precomputed in a scene-independent manner, resulting in accelerated inference for new scenes. It allows us to perform a full search over the finely discretized space of surface normals to determine the globally optimal surface normal for each scene point. We show that our method can accurately estimate surface normals of scenes with spatially varying different reflectances in a reasonable amount of time.

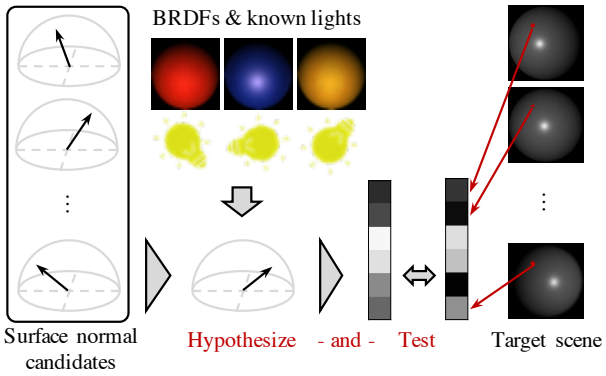

Figure 1: An overview of our approach. We hypothesize a surface normal and test whether it can explain the target measurements. By conducting the hypothesis-and-test for all possible surface normals, our method is able to find a globally optimal surface normal.

1. Introduction

Photometric stereo recovers fine surface details in the form of surface normals from images taken by a static camera under varying lightings. While traditional photometric stereo methods [25] assume Lambertian reflectance or simplified parametric reflectance models, it is understood that their deviation from real-world reflectances introduces errors in surface normal estimates. In the past, other studies used more sophisticated reflectance models [1, 15, 21, 13, 4] for more accurate surface normal recovery; however, they generally encounter an issue of non-convex optimization in determining the surface normals. The problem is rooted in the fact that these methods frame the estimation problem as a continuous optimization problem.

In this paper, we cast surface normal estimation as a *discrete hypothesis-and-test search* problem. Instead of treating surface normals to be estimated as a continuous quantity, our method finely discretizes the space of surface normals and finds the best surface normal by a hypothesis-and-test search. Since a surface normal vector has only two degrees of freedom (a unit 3D vector) represented by its azimuth and elevation angles in a hemisphere, discretization results in a relatively small number of surface normal candidates. For example, even if we discretize the angles in one-degree intervals, it results in $32,400 = 360 \times 90$ normal candidates. Our method uses each surface normal candidate as *hypothesis* and *tests* its suitability to the image formation model; thus, we call it a hypothesis-and-test search strategy. In this manner, our method searches for the globally optimal surface normal from all (discretized) possible ones.

To alleviate the issue of computing cost in our discrete search, we developed a precomputation method that performs expensive computations in a scene-independent manner prior to the inference for a new scene. To deal with a diverse set of reflectances, we use a non-parametric, discrete table of bidirectional reflectance distribution functions (BRDFs), whose axes are the space of surface normals, light directions, and materials, for a fixed viewing direction. The table of BRDFs, which we call a BRDF tensor, can contain an arbitrary number of materials, and importantly, the number of reference materials considered in the BRDF tensor does not influence the computation time during inference.

Our method is motivated by the success of example-based [9] and virtual exemplar-based [11] approaches. However, unlike example-based methods, we do not require

placing a reference object in the scene. Also, unlike the virtual exemplar-based method that performs a continuous local search using a non-convex objective function, we treat the problem as a discrete search problem and perform an exhaustive search over the discretized space. It enables us to find a globally optimal surface normal within the bounds of our objective function.

The chief contributions of this paper are twofold. First, we propose a discrete hypothesis-and-test search strategy for photometric stereo. By finely discretizing the space of surface normals, our method finds the globally optimal surface normal through exhaustive search. Second, we show that expensive computation can be performed prior to the surface normal estimation, allowing the global hypothesis-and-test search to work in a reasonable amount of time. We assess the accuracy of the proposed method using both synthetic and real-world data and show its favorable performance in determining surface normals of a scene. In particular, the proposed method achieves a stable estimate, *i.e*., superior average/variance of mean angular error over a diverse set of materials.

2. Related work

Calibrated photometric stereo methods for diverse materials can be roughly divided into three categories; model-based, learning-based, and example-based approaches. In the following, we discuss the corresponding related works.

Model-based photometric stereo Model-based approaches use parametric expressions for BRDFs and the model parameters including the surface normal are estimated, typically, via optimization. Key for these model-based methods is the choice of a parametric BRDF model. Woodham's original work [25] assumed Lambertian reflectance, which allows using convex least-squares optimization to determine surface normals and albedos. Parametric modeling of non-Lambertian BRDFs is actively studied, particularly in the graphics community. For example, the Blinn-Phong model [23], the Torrance-Sparrow model [7], the Ward model [6], the specular spike model [5, 26], and a microfacet BRDF with ellipsoidal normal distributions [4] have been developed. However, each of these models is limited to a class of materials, and such models are highly nonlinear, resulting in non-convex photometric stereo problems. Thus, some recent methods use a bivariate function instead. For representing low-frequency reflectances, Shi *et al*. [21] use a bi-polynomial function and Ikehata and Aizawa [13] use a sum of lobes with unknown center directions. Although these model-based approaches can be used in a relatively wide range of materials, there are always problematic materials.

Learning-based photometric stereo Recently, deep learning-based photometric stereo methods have been proposed. They learn a mapping from measured intensities under known lightings to surface normals using a neural network [19, 3, 13]. These methods showed promising results on various scenes owing to the network being trained with diverse shapes and materials. Santo *et al*. [19] and Chen *et al*. [3] created a training dataset by rendering the Blobby [14] and Sculpture [24] shape datasets with 100 BRDFs from the MERL dataset [16]. Ikehata [12] also introduced a training dataset for calibrated photometric stereo, called CyclesPS dataset, containing several objects with a diverse set of materials rendered using Disney's principled BSDF [2]. Learning-based methods require a large number of shapes & BRDFs for training. When they are applied to new scenes with shapes & BRDFs that are very different from the training data, it is necessary to do costly re-training of the network.

Example-based photometric stereo Example-based photometric stereo relies on the concept of orientation-consistency [9], *i.e*., two surfaces with the same surface normal and BRDF will have the same appearance under the same illumination. An early work along this direction is found in Horn and Ikeuchi [10]. In example-based approaches, a reference object with known surface normals is placed in a target scene. Further, the BRDF of the reference object is assumed to be the same as that of the target object. Then, a surface normal is recovered for each point of the target object by searching the corresponding pixel intensity of the reference object that best matches the target's appearance. To relax the assumption of identical BRDF between reference and target, Hertzmann and Seitz [9] introduced two reference objects, a diffuse and a specular sphere, placed in the target scene, and approximate the target BRDF by a non-negative linear combination of the reference BRDFs. Although this approach makes example-based photometric stereo applicable to more diverse materials, it is still inaccurate to approximate a diverse set of materials by a linear combination of two BRDFs. In addition, in many practical applications it is undesirable to place reference objects in a target scene.

Hui and Sankaranarayanan [11] introduced virtual exemplar-based photometric stereo that performs example-based photometric stereo without actually introducing reference objects into a target scene. They render virtual reference spheres under the target scene illumination with MERL BRDFs [16] and assume that the target BRDF lies in the non-negative span of the MERL BRDFs. In the virtual exemplar-based approach, however, there are many time-consuming processes such as rendering virtual spheres, an iterative optimization for solving a non-negative least squares problem, and searching over all possible surface

normals. To reduce the computation cost, they proposed an efficient search algorithm which however eliminates the guarantee of finding the optimal solution.

Our method shares the assumption that the target BRDF can be represented by a combination of several reference BRDFs. However, we cast the problem as a discrete hypothesis-and-test search problem, which gives a guarantee of reaching the globally optimal solution within the bound of the objective function. Additionally, our method enables search all surface normal candidates in reasonable time owing to an efficient precomputation.

3. Image formation and problem statement

Suppose a surface point with a unit surface normal $\mathbf{n} \in \mathcal{S}^2 \subset \mathbb{R}^3$ is illuminated by an incoming directional light $\mathbf{l} \in \mathcal{S}^2$, without ambient lighting or global illumination effects such as cast shadows or inter-reflections. When this surface point is observed by a camera with linear response, the measured intensity $m \in \mathbb{R}_+$ can be written as

$$m \propto \rho(\mathbf{n}, \mathbf{l}) \max(\mathbf{n}^\top \mathbf{l}, 0), \tag{1}$$

where $\rho(\mathbf{n}, \mathbf{l}) \colon \mathcal{S}^2 \times \mathcal{S}^2 \to \mathbb{R}_+$ is a general isotropic bidirectional reflectance distribution function (BRDF).

In calibrated photometric stereo, a static camera records multiple, say L', measurements $\{m_1, \dots m_{L'}\}$ for each surface point under various light directions $\{\mathbf{l}_1, \dots \mathbf{l}_{L'}\}$. Then, Eq. (1) can be written in matrix form as

$$\underbrace{\begin{pmatrix} m_1 \\ \vdots \\ m_{L'} \end{pmatrix}}_{\mathbf{m}} \propto \underbrace{\begin{pmatrix} \max(\mathbf{n}^\top \mathbf{l}_1, 0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \max(\mathbf{n}^\top \mathbf{l}_{L'}, 0) \end{pmatrix}}_{\mathbf{E}} \underbrace{\begin{pmatrix} \rho(\mathbf{n}, \mathbf{l}_1) \\ \vdots \\ \rho(\mathbf{n}, \mathbf{l}_{L'}) \end{pmatrix}}_{\boldsymbol{\rho}},$$

where $\mathbf{m}$ is a measurement vector, $\mathbf{E}$ is a diagonal irradiance matrix, and $\boldsymbol{\rho}$ is a reflectance vector. We model the reflectance $\boldsymbol{\rho}$ by a linear combination of BRDF basis vectors in a similar manner to Hertzmann *et al.* [9], and Hui and Sankaranarayanan [11]. By stacking M known BRDF basis vectors in a BRDF basis matrix $\mathbf{B}$, $\boldsymbol{\rho}$ can be written as

$$\boldsymbol{\rho} = \underbrace{\begin{pmatrix} \rho_1(\mathbf{n}, \mathbf{l}_1) & \dots & \rho_M(\mathbf{n}, \mathbf{l}_1) \\ \vdots & \ddots & \vdots \\ \rho_1(\mathbf{n}, \mathbf{l}_{L'}) & \dots & \rho_M(\mathbf{n}, \mathbf{l}_{L'}) \end{pmatrix}}_{\mathbf{B}} \mathbf{c},$$

where $\mathbf{c} = [c_1, \dots, c_M]^\top$ is a BRDF coefficient vector. With this, the image formation model can be simplified to

$$\mathbf{m} = \mathbf{EBc}. \tag{2}$$

Problem statement Our goal is to find the optimal surface normal $\mathbf{n}$ and BRDF coefficients $\mathbf{c}$ for each surface point, given observations $\mathbf{m}$ and associated light directions $\{\mathbf{l}_1, \dots \mathbf{l}_{L'}\}$ based on the model of Eq. (2). The irradiance matrix $\mathbf{E}$ and BRDF matrix $\mathbf{B}$ are functions of the surface normal $\mathbf{n}$.

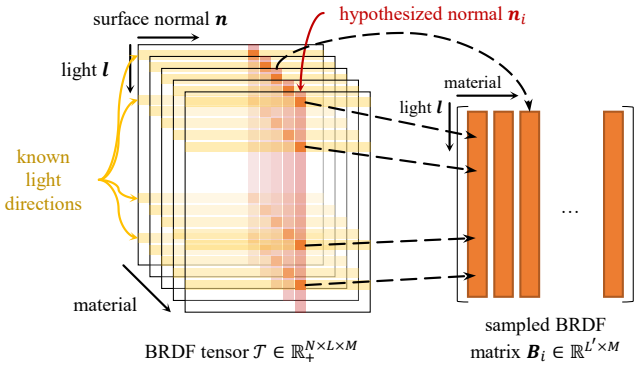

Figure 2: Starting from the BRDF tensor $\mathcal{T}$ that represents reflectances for a comprehensive set of light directions, surface normals and materials (BRDFs), we slice out a sampled BRDF matrix $\mathbf{B}_i$ for a set of known light directions and a hypothesized surface normal $\mathbf{n}_i$. The column space of $\mathbf{B}_i$ is the space of reflectances over all possible materials for the hypothesized normal under the known light directions.

4. Proposed method

Our method casts the photometric stereo problem as a discrete search where the space of surface normals is discretized. We *hypothesize* a surface normal $\mathbf{n}$ and *test* whether it (approximately) satisfies the image formation model of Eq. (2). By conducting this hypothesis-and-test for all possible surface normals, our method is able to find a globally optimal surface normal $\mathbf{n}$ and the associated BRDF coefficients $\mathbf{c}$ that best satisfy Eq. (2).

4.1. Hypothesis-and-test strategy

Let $\mathcal{N} = \{\mathbf{n}_i \mid i = 1, \dots, N\}$ be the discretized space of surface normals, which we call a set of surface normal *candidates*. We prepare a tensor representation for diverse BRDFs whose axes are (1) surface normals, (2) light directions, and (3) materials. Suppose the spaces of surface normals and light directions are discretized into N and L bins, respectively, and there are M distinct BRDFs. Then, the BRDF tensor $\mathcal{T}$ can be defined as $\mathcal{T} \in \mathbb{R}_+^{N \times L \times M}$ (see the left of Fig. 2).

For simplicity, let us assume that the BRDF tensor contains the actual light directions of the observed scene. If we hypothesize a certain surface normal $\mathbf{n}_i \in \mathcal{N}$ for a scene point, using $L' \leq L$ known light directions, we can slice a *sampled BRDF matrix* $\mathbf{B}_i \in \mathbb{R}_+^{L' \times M}$ from the BRDF tensor $\mathcal{T}$ along the hypothesized surface normal $\mathbf{n}_i$ and a set of L' known light directions as illustrated in Fig. 2. We can also

form an irradiance matrix $\mathbf{E}_i$ for the hypothesized surface normal $\mathbf{n}_i$. Using $\mathbf{B}_i$ and $\mathbf{E}_i$ instead of $\mathbf{B}$ and $\mathbf{E}$, Eq. (2) becomes

$$\mathbf{m} \simeq \mathbf{E}_i\mathbf{B}_i\mathbf{c} \stackrel{\text{def}}{=} \mathbf{D}_i\mathbf{c},$$

where $\mathbf{D}_i(=\mathbf{E}_i\mathbf{B}_i) \in \mathbb{R}^{L'\times M}$. For the overdetermined case $L' > M$, the least-squares solution for the BRDF coefficients $\mathbf{c}$ that best explains the measurements, is

$$\mathbf{c}^* = \left(\mathbf{D}^\top\mathbf{D}\right)^{-1}\mathbf{D}^\top\mathbf{m} = \mathbf{D}_i^\dagger\mathbf{m},$$

where $\mathbf{D}_i^\dagger$ is $\mathbf{D}_i$'s pseudo-inverse. The estimated BRDF coefficients $\mathbf{c}^*$ are least-squares optimal for the hypothesized normal $\mathbf{n}_i$ and the space of sampled BRDFs $\mathbf{B}_i$. We can test the validity of the hypothesized $\mathbf{n}_i$ by evaluating the ℓ_2 reconstruction error as

$$e_i = \|\mathbf{m} - \mathbf{D}_i\mathbf{c}^*\|_2^2. \tag{3}$$

Therefore, the optimal surface normal $\mathbf{n}^*$ can be found as the minimizer of the following objective

$$\mathbf{n}^* = \mathbf{n}_{i^*}, \quad i^* = \underset{i\in\{1,\dots,N\}}{\operatorname{argmin}}\ e_i. \tag{4}$$

A naïve implementation may require a significant computation effort for solving this problem. We therefore introduce an efficient precomputation strategy now.

4.2. Precomputation

The reconstruction error e_i for the hypothesized surface normal $\mathbf{n}_i$ in Eq. (3) can be further simplified as

$$\begin{aligned} e_i &= \|\mathbf{m} - \mathbf{D}_i\mathbf{c}^*\|_2^2 = \left\|\mathbf{m} - \mathbf{D}_i\mathbf{D}_i^\dagger\mathbf{m}\right\|_2^2 \\ &= \left\|\left(\mathbf{I} - \mathbf{D}_i\mathbf{D}_i^\dagger\right)\mathbf{m}\right\|_2^2 \stackrel{\text{def}}{=} \|\mathbf{Z}_i\mathbf{m}\|_2^2, \end{aligned} \tag{5}$$

where $\mathbf{Z}_i(=\mathbf{I} - \mathbf{D}_i\mathbf{D}_i^\dagger) \in \mathbb{R}^{L'\times L'}$ is uniquely determined given a hypothesized normal $\mathbf{n}_i$. We can precompute a set of $\{\mathbf{Z}_i\}$ for all normal candidates in $\mathcal{N}$ and at test time we simply need to assess the magnitude of $\mathbf{Z}_i\mathbf{m}$ for all i.

This precomputation happens only once and the result can be used for any new scene as long as the light configuration is unchanged.

4.3. Dimensionality reduction of $\mathbf{D}_i$

In principle, minimizing Eq. (4) gives us a correct solution for the surface normal $\mathbf{n}$. In practice, however, we need to pay attention to the dimension and range of matrix $\mathbf{D}_i \in \mathbb{R}^{L'\times M}$. Specifically, when $L' < M$ or $\mathbf{m} \in \mathsf{ran}(\mathbf{D}_i)$ (the range of $\mathbf{D}_i$) for all $\mathbf{D}_i$, there exists one or more BRDF coefficient vectors $\mathbf{c}^*$ that makes all reconstruction errors $\{e_i\}$ zero.

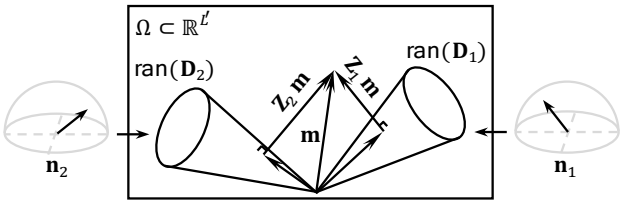

Figure 3: Geometric interpretation of the reconstruction error. The reconstruction error of measurements $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ can be seen as distance between the measurement vector $\mathbf{m}$ and the subspace spanned by $\mathbf{D}_i$ in the L'-dimensional space Ω.

As illustrated in Fig. 3, a measurement vector $\mathbf{m}$ exists in an L'-dimensional space Ω. The column vectors of $\mathbf{D}_i$ span a $\mathsf{rank}(\mathbf{D}_i)$-dimensional subspace in Ω, and the measurement reconstructions $\mathbf{D}_i\mathbf{c}^* = \mathbf{D}_i\mathbf{D}_i^\dagger\mathbf{m}$ reside in this subspace. Thus, geometrically, the reconstruction error $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ can be considered as the distance between the measurement vector $\mathbf{m}$ and the subspace spanned by $\mathbf{D}_i$. From this perspective, if $\mathsf{rank}(\mathbf{D}_i) = L'$, the columns of $\mathbf{D}_i$ span the entire Ω; therefore, the reconstruction error becomes always zero regardless of the correctness of the hypothesized surface normal $\mathbf{n}$.

To avoid such a situation, we reduce the column dimension of $\mathbf{D}_i$ from M to $M' < L'$ if $L' < M$. Specifically, we apply singular value decomposition (SVD) to $\mathbf{D}_i$ as

$$\mathbf{D}_i = \mathbf{U}\mathbf{S}\mathbf{V}^\top,$$

and reduce the dimensionality of $\mathbf{D}_i$ using the first M' singular vectors / values as

$$\mathbf{D}_i \leftarrow \mathbf{U}_{M'}\mathbf{S}_{M'}\mathbf{V}_{M'}^\top,$$

where $\mathbf{U}_{M'}$, $\mathbf{S}_{M'}$, and $\mathbf{V}_{M'}^\top$ are the truncated singular vectors / values.

We empirically found that the proper value of M' is related to the noise level in the observations. In Sec. 5.4, we examine the accuracy of surface normal estimation with varying M' and discuss the choices for M'.

5. Experiments

This section describes the results of experiments with synthetic and real-world data. We further discuss the computation time, the effect of dimensionality reduction and the discretization of the space of light directions. We begin with describing the construction of the BRDF tensor and the synthetic and real-world datasets that we use for evaluation.

BRDF tensor: The BRDF tensor is constructed from three components; materials, surface normal candidates, and light directions. As materials, we used the MERL BRDF dataset [16] which consists of 100 distinct BRDFs

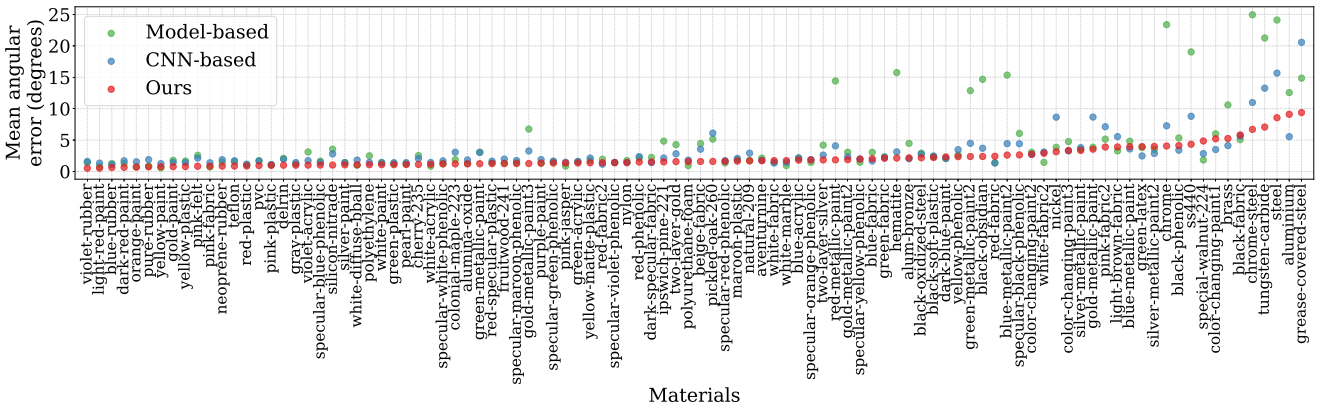

Figure 4: Evaluation on the MERL sphere dataset. We compared our method against a model-based method [13] and a CNN-based method [12]. The average/variance of the mean angular errors over all materials are as follows: Ours: $2.23/3.24$, model-based method [13]: $4.31/29.43$, CNN-based method [12]: $3.15/9.19$.

including diffuse, specular, and metallic materials. For surface normal candidate sampling we followed Hui's method [11] and obtained 20001 candidates by $0.5°$ equiangular sampling over the hemisphere [8]. In all experiments of this paper, we assume that the BRDF tensor contains the known light directions.

MERL sphere: We created a MERL sphere dataset, which is synthetic data rendered with 100 isotropic BRDFs from the MERL BRDF database [16]. We generated images of a sphere object illuminated from $20, 40, 60, 80, 100$ different known directions[1], respectively. We also created a noisy MERL sphere dataset by adding signal-independent and signal-dependent noise [17] to the MERL sphere dataset. The noise model is $\tilde{m} = m + (\mu + \lambda\sqrt{m})X$ where $\tilde{m}$ and m are image signals with and without noise, μ and λ are weighting factors for signal-independent and signal-dependent noise, respectively, and X is a $\mathcal{N}(0, 1)$-distributed random variable.

MERL bunny: We rendered the Stanford bunny with spatially varying materials using 11 distinct MERL BRDFs as test BRDFs and 80 different known light directions.

Real-world benchmark: We took an existing real-world dataset, the DiLiGenT dataset [20], which contains 10 real objects of general reflectance illuminated from 96 different known directions. This dataset provides ground truth surface normal maps for all objects measured by high-precision laser scanning, enabling quantitative evaluation.

[1]We used the generalized spiral points algorithm for uniformly distributing lights on the hemisphere [18]. See the supplementary material for details of the light distributions.

5.1. Evaluation on synthetic data

We performed experiments on synthetic data to confirm that our method works with diverse materials. Since there is no global illumination in the MERL sphere & bunny dataset, we can evaluate only the ability of our method to adapt to diverse materials. To support that our method is applicable to a more diverse set of materials than existing methods, we compared it against a model-based [13], a CNN-based [12], and a dictionary-based method [11]. Additionally, we investigated the accuracy of our method with varying surface normal candidates and lights, and compared it to the dictionary-based method [11]. All following experiments were performed in the 80 lights setting unless otherwise specified.

MERL sphere: We performed a comparative experiment between our method, the model-based method [13], and the CNN-based method [12], and verified that our method is more accurate and stable on diverse material data. For the materials we used a leave-one-out scheme, testing it on one MERL BRDF while constructing the BRDF tensor from the remaining 99 BRDFs.[2]

The results are shown in Fig. 4. The average/variance of the mean angular errors over all materials are $2.23/3.24$ for our method, $4.31/29.43$ for the model-based method [13], and $3.15/9.19$ for the CNN-based method [12]. This result suggests that our method is more accurate and stable than existing methods on most materials in the MERL database.

[2]Due to the high computational costs of the dictionary-based method [11], a leave-one-out scheme on it with the author's implementation cannot be performed in a reasonable amount of time. We do a comparison with it in the following experiments.

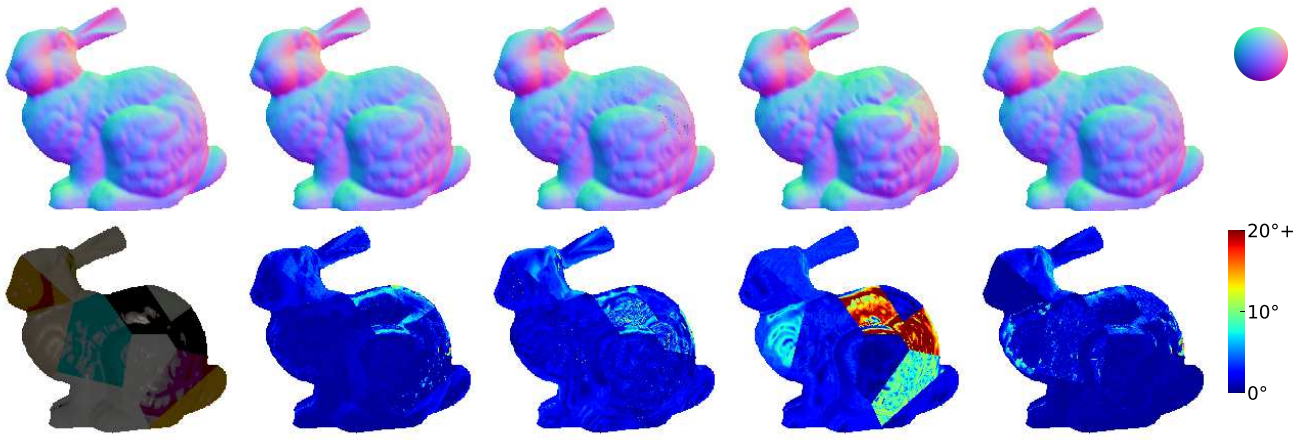

Figure 5: Visual comparison for the Stanford bunny with spatially varying materials. We show estimated normals and angular error maps of our and 3 comparison methods. Values under the error maps indicate the mean/variance of angular errors.

MERL bunny: We evaluated our method, the model-based [13], the CNN-based [12], and the dictionary-based method [11] on the Stanford bunny with spatially varying materials. Our BRDF tensor and the BRDF dictionary of the dictionary-based method are constructed from 80 MERL BRDFs that do not include the 11 test BRDFs mentioned above.

In Fig. 5 we show each method's estimated surface normals and angular error map. The values under the error maps indicate the mean/variance of their angular errors. This result suggests that our method can be applied to materials that are difficult for the CNN-based and the model-based method. Additionally, from the low variance of angular errors we can see that our method is more stable than the existing methods on diverse materials.

Varying number of surface normal candidates and lights: We investigate the angular errors of the estimated surface normals for varying numbers of surface normal candidates and lights, and compare our method to the existing search-based method [11]. In the experiment with a varying number of surface normal candidates, we use $20001, 10001, 1501, 251$ surface normal candidates yielded by $0.5°, 1°, 3°, 5°$ equiangular sampling, respectively. In the experiment with a varying number of lights, we use the MERL sphere dataset with $20, 40, 60, 80, 100$ lights. Due to the high computational costs of the existing dictionary-based method [11], we perform a 5-fold cross-validation, using 80 BRDFs for computing our method's BRDF tensor and the dictionary method's material dictionary [11], and testing on the MERL sphere dataset with the remaining 20 BRDFs.

Table 1: Mean angular errors for estimated surface normals in degrees for varying numbers of normal candidates.

	number of normal candidates N			
	20001	10001	1501	251
Ours	**2.25**	**2.31**	**2.86**	**4.62**
HS17 [11]	2.34	2.41	3.00	4.68

Table 2: Mean angular errors for estimated surface normals in degrees for varying numbers of lights.

	number of lights L'				
	100	80	60	40	20
Ours	**2.22**	**2.25**	**2.28**	**2.35**	**2.69**
HS17 [11]	2.30	2.34	2.39	2.49	2.81

Tables 1 and 2 show the experimental results for varying numbers of surface normal candidates and lights, respectively. Both results suggest that our method consistently estimates surface normals more accurately than the existing method. One of the reasons is a difference in the search strategies. Hui and Sankaranarayanan [11] proposed a search strategy that avoids searching parts of the surface normal space since their method is computationally expensive due to iterative optimization. In contrast, our method searches over all surface normal candidates in reasonable computation time facilitated by the precomputation.

5.2. Evaluation on real-world data

We show quantitative comparisons on the DiLiGenT dataset in Tab. 3 where we compare our method with existing methods in terms of mean angular error. Here, our

Table 3: Comparisons on the DiLiGenT dataset. We apply a least-squares method as baseline. The values represent mean angular error (MAE).

	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	Avg.
Ours	1.58	6.38	13.69	6.30	7.80	11.42	18.74	6.67	7.26	15.49	9.53
SI18 [12]	2.20	**4.10**	**7.90**	**4.60**	7.90	**7.30**	**13.90**	5.40	**6.00**	12.60	**7.20**
HS18 [11]	**1.33**	5.58	8.48	4.88	8.23	7.57	15.80	**5.16**	6.41	12.08	7.55
CH18 [3]	2.80	7.60	**7.90**	6.20	7.30	8.60	15.90	7.10	7.30	13.30	8.40
TM18 [22]	1.47	5.79	10.36	5.44	**6.32**	11.47	22.59	6.09	7.76	**11.03**	8.83
ST14 [21]	1.74	6.12	10.60	6.21	13.90	10.10	25.40	6.51	8.78	13.60	10.30
IA14 [13]	3.34	7.11	10.50	6.74	13.10	9.71	26.00	6.64	8.77	14.20	10.60
Baseline	4.10	8.39	14.90	8.41	25.60	18.50	30.60	8.89	14.70	19.80	15.40

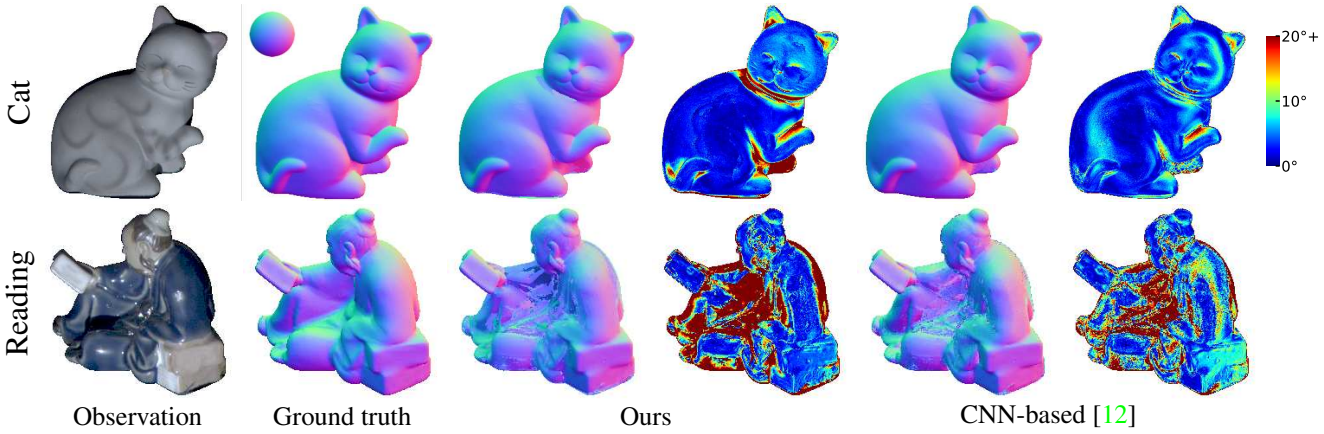

Figure 6: Visual comparison for the non-convex objects Cat and Reading. We show estimated normals and angular error maps of our method and a state-of-the-art CNN-based method [12].

method does not achieve the best results. This is considered to be due to factors not modeled in our method, namely cast shadows or inter-reflections. Figure 6 shows visual comparisons between our method and Ikehata's CNN-based method [12] for two non-convex objects with cast shadows and inter-reflections. Our method causes a large angular error in pixels where cast shadows or inter-reflections are likely to occur. However, in convex parts our method outperforms the CNN-based state-of-the-art method and estimates the surface normals well for the Reading object's complex specular material. We show more visual comparisons in the supplementary material.

5.3. Computation cost

For inference, our method evaluates the reconstruction error $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ in Eq. (5) for each surface normal candidate $\mathbf{n}_i \in \mathcal{N}$. All matrices $\mathbf{Z}_i$ are precomputed; therefore, at inference time we only need to evaluate the reconstruction error of each $\mathbf{n}_i$ and find the minimizer. The dimension of matrix $\mathbf{Z}_i \in \mathbb{R}^{L' \times L'}$ only depends on the number of lights L', but not the number of materials. Table 4 summarizes the computation time per pixel for 20001 surface normal candidates and a varying number of lights L'. The computation is highly parallelizable, *e.g.*, by pixel-wise or normal candidate-wise parallelization.

Table 4: Computation time in seconds for surface normal estimation of one pixel with 20001 surface normal candidates, measured on one core of an Intel Xeon CPU E5-2680 v4 @ 2.40GHz.

	number of lights L'				
	100	80	60	40	20
Time [sec.]	0.31	0.21	0.15	0.062	0.024

5.4. Choice of dimension M' for noisy data

It is difficult to avoid imaging noise in real-world data and it must therefore always be considered. We empirically observed that M' is related to our method's robustness towards noise. Thus, we determine an optimal M' by a validation on the noisy MERL sphere dataset.

Table 5: Surface normal estimation with varying M' on the noisy MERL sphere dataset under three sets of lights, (a) 60 lights, (b) 80 lights, and (c) DiLiGenT's 96 lights. Values indicate mean angular errors in degrees.

(a)

M'	μ/λ 0/0	5/30	30/5	30/30
3	2.21	**3.99**	3.65	**4.37**
5	1.83	4.10	**3.52**	4.50
10	**1.71**	4.29	3.53	4.71
20	1.75	4.95	3.84	5.33

(b)

M'	μ/λ 0/0	5/30	30/5	30/30
3	2.18	**3.99**	3.83	**4.43**
5	1.76	4.12	**3.71**	4.57
10	**1.62**	4.30	3.72	4.77
20	1.62	4.68	3.88	5.12

(c)

M'	μ/λ 0/0	5/30	30/5	30/30
3	2.59	**5.60**	**5.15**	**6.22**
5	2.38	6.19	5.20	6.75
10	**2.33**	7.08	5.39	7.62
20	2.37	8.07	5.70	8.65

Setup: We apply 5-fold cross-validation on the MERL BRDFs; the 100 MERL BRDFs are divided into 80 BRDFs for the BRDF tensor and 20 BRDFs for testing. We test varying $M' = \{3, 5, 10, 20\}$ and varying noise $\mu/\lambda = \{0/0, 5/30, 30/5, 30/30\}$ under three sets of lights: 60 lights, 80 lights, and DiLiGenT's 96 lights.

Results: Table 5 shows mean angular errors of estimated surface normals in degrees for varying M'. In the noiseless cases ($\mu = \lambda = 0$), a larger M' performs better. Contrarily, in most of the noisy cases, $M' = 3$ produces lower angular errors, indicating that $M' = 3$ is most robust to noise. For this reason, we apply $M' = 3$ in all following experiments.

5.5. Discretized light directions

In all experiments so far, we assumed that the BRDF tensor contains the known light directions. In practice, the BRDF tensor rarely contains all of the known light directions and we should use pre-defined light directions closest to the known light directions instead. Here, we examine how the surface normal estimation accuracy is affected by the discretization of light directions.

Setup: As pre-defined light directions in the BRDF tensor we used 20001 discretized directions from a $0.5°$ equiangular sampling over the hemisphere. When a set of known light directions is given, as illustrated in Fig. 2, we can slice out a sampled BRDF matrix for a hypothesized surface normal and the set of light directions that are closest to the known light direction in terms of cosine distance. We can then follow the same estimation process used so far. We performed this experiments with $20, 40, 60, 80, 100$ known lights. We used the MERL sphere dataset and performed 5-fold cross-validation over the BRDFs.

Results: Table 6 shows surface normal estimation results for exactly known and for discretized light directions in the BRDF tensor. We observe that there are only slight differences between both cases ($< 0.1°$ in the mean angular errors), which suggests that it is sufficient to precompute a BRDF tensor for sufficiently finely discretized light directions and there is no need to recalculate the BRDF tensor with known light directions each time one works with a new light setup.

Table 6: The effect of discretizing the light directions and picking the directions closest to a set of known directions. The values indicate mean angular error in degrees.

	number of lights L' 100	80	60	40	20
Using known lights	2.15	2.18	2.21	2.31	2.69
Using discretized lights	2.16	2.19	2.22	2.33	2.73

6. Discussion

In this paper, we have presented a photometric stereo method based on discrete hypothesis-and-test search. The proposed method can work with a diverse set of BRDFs that are represented in a BRDF tensor and can determine surface normals of a scene with spatially varying general BRDFs. By putting most of the computation into a precomputation step, we enabled a full search over all surface normal candidates, leading to a solution guaranteed to be optimal within the bounds of the objective function and the discretization. The approach is also supported by the fact that with the continuing increase of computation power, memory size, and the availability of many-core processors, the applicability of the full search strategy is expanding. We are interested in seeing more applications along the direction.

One interesting question to investigate in the future would be how many BRDFs are needed to fully represent real-world materials. Our BRDF tensor can hold an arbitrary number of BRDFs; therefore, we are interested in increasing the number of BRDFs in the tensor to study this question, while it may only be empirically understood.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19H01123. Michael Waechter was supported through a postdoctoral fellowship by the Japan Society for the Promotion of Science (JP17F17350). Kiriakos N. Kutulakos was supported by the Natural Sciences and Engineering Research Council of Canada under the RGPIN program.

References

[1] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition (CVPR)*, 2008. 1

[2] Brent Burley. Physically-based shading at Disney. In *SIGGRAPH*, 2012. 2

[3] Guanying Chen, Kai Han, and Kwan-Yee Kenneth Wong. PS-FCN: A flexible learning framework for photometric stereo. In *European Conference on Computer Vision (ECCV)*, 2018. 2, 7

[4] 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 *Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 2

[5] Tongbo Chen, Michael Goesele, and Hans-Peter Seidel. Mesostructure from specularity. In *Computer Vision and Pattern Recognition (CVPR)*, 2006. 2

[6] Hin-Shun Chung and Jiaya Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Computer Vision and Pattern Recognition (CVPR)*, 2008. 2

[7] Athinodoros S. Georghiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *International Conference on Computer Vision (ICCV)*, 2003. 2

[8] Radoslav Harman and Vladimír Lacko. On decompositional algorithms for uniform sampling from n-spheres and n-balls. *Journal of Multivariate Analysis*, 101(10):2297–2304, 2010. 5

[9] Aaron Hertzmann and Steven M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(8):1254–1264, 2005. 1, 2, 3

[10] Berthold KP Horn and Katsushi Ikeuchi. The mechanical manipulation of randomly oriented parts. *Scientific American*, 251(2):100–113, 1984. 2

[11] Zhuo Hui and Aswin C. Sankaranarayanan. Shape and spatially-varying reflectance estimation from virtual exemplars. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 39(10):2060–2073, 2017. 1, 2, 3, 5, 6, 7

[12] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *European Conference on Computer Vision (ECCV)*, 2018. 2, 5, 6, 7

[13] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 2, 5, 6, 7

[14] Micah K. Johnson and Edward H. Adelson. Shape estimation in natural illumination. In *Computer Vision and Pattern Recognition (CVPR)*, 2011. 2

[15] Feng Lu, Xiaowu Chen, Imari Sato, and Yoichi Sato. SymPS: BRDF symmetry guided photometric stereo for shape and light source estimation. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 40(1):221–234, 2018. 1

[16] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *Transactions on Graphics (TOG)*, 22(3):759–769, 2003. 2, 4, 5

[17] Netanel Ratner and Yoav Y. Schechner. Illumination multiplexing within fundamental limits. In *Computer Vision and Pattern Recognition (CVPR)*, 2007. 5

[18] Edward B. Saff and Arno Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, 19(1):5–11, 1997. 5

[19] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Workshops of the International Conference on Computer Vision (ICCVW)*, 2017. 2

[20] 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. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 41(2):271–284, 2019. 5

[21] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 36(6):1078–1091, 2014. 1, 2, 7

[22] Tatsunori Taniai and Takanori Maehara. Neural inverse rendering for general reflectance photometric stereo. In *International Conference on Machine Learning (ICML)*, 2018. 7

[23] Silvia Tozza, Roberto Mecca, Marti Duocastella, and Alessio 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

[24] Olivia Wiles and Andrew Zisserman. SilNet: Single- and multi-view reconstruction by learning from silhouettes. *British Machine Vision Conference (BMVC)*, 2017. 2

[25] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1, 2

[26] 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. *Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(4):890–897, 2014. 2