This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

Discrete Search Photometric Stereo for Fast and Accurate Shape Estimation

Kenji Enomoto, Michael Waechter, Fumio Okura, Kiriakos N. Kutulakos, and Yasuyuki Matsushita

Abstract—We consider the problem of estimating surface normals of a scene with spatially varying, general bidirectional reflectance distribution functions (BRDFs) observed by a static camera under varying distant illuminations. Unlike previous approaches that rely on continuous optimization of surface normals, we cast the problem as a *discrete* search problem over a set of finely discretized surface normals. In this setting, we show that the expensive processes can be precomputed in a scene-independent manner, resulting in accelerated inference. We discuss two variants of our Discrete Search Photometric Stereo (DSPS), one working with continuous linear combinations of BRDF bases and the other working with discrete BRDFs sampled from a BRDF space. Experiments show that DSPS has comparable accuracy to state-of-the-art exemplar-based photometric stereo methods while achieving 10–100x acceleration.

Index Terms—Photometric stereo, spatially varying BRDFs, discrete search, nearest-neighbor search

1 Introduction

Photometric stereo recovers fine surface details in the form of surface normals from images taken by a static camera under varying lightings. Traditional photometric stereo methods [1] assume Lambertian reflectance, which deviates from real-world reflectances, thus introducing errors in surface normal estimates. Many modern methods use more sophisticated reflectance models [2], [3], [4], [5] for more accurate surface normal recovery; however, they generally encounter the 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 continuous optimization.

This paper presents Discrete Search Photometric Stereo (DSPS), which casts photometric stereo as *discrete* search over a set of finely discretized surface normals. Since a surface normal has only two degrees of freedom (*i.e.*, a unit vector represented by the azimuth and elevation angles over a hemisphere), discretization yields a relatively small number of surface normal candidates; *e.g.*, discretizing in one-degree intervals yields $360 \times 90 = 32,400$ candidates. DSPS searches over the space of finely discretized surface normals while avoiding problems with local minima.

To reduce the computational cost of our discrete search, we show that expensive computations can be packed into a *scene-independent precomputation step* which accelerates our discrete search without sacrificing accuracy. As Fig. 1 illustrates, the precomputed database can be reused for arbitrary scenes that can be well described by a set of BRDF bases.

In this paper, we discuss two variants of our DSPS algorithm suited for two different classes of BRDF models:

1) DSPS-C for continuous BRDF models represented as the linear combination of multiple BRDF bases. This representation is often used in photometric stereo methods for general BRDFs [2], [6], [7]. In DSPS-C, offline precomputation on discretized surface normals greatly speeds up the inference phase.
2) DSPS-D for discrete BRDFs. We turn the inference problem into a nearest-neighbor search problem, for which we benefit from existing fast nearest-neighbor search algorithms.

Our methods are motivated by the success of example-based [6] and virtual exemplar-based [7] photometric stereo. Unlike example-based methods, DSPS does not require placing a reference object in the scene. Also, unlike virtual exemplar-based methods that perform "coarse-to-fine search" without the guarantee of finding the globally optimal surface normal, we treat the problem as a discrete search problem, which enables us to use exhaustive search to find the globally optimal surface normal within the discretized space.

Contribution: We propose a precomputation strategy for photometric stereo with continuous and discrete BRDF models, via the discretization of surface normals. It accelerates the surface normal estimation and enables exhaustive search to find globally optimal surface normals in a reasonable amount of time, which greatly contribute photometric stereo applications such as fast defect inspection [8]. We assess the efficiency and accuracy of our method using synthetic and real-world data and show 10–100x acceleration over state-of-the-art methods while maintaining accuracy.

In an earlier version of this work [9], we discussed the continuous BRDF model (variant 1 above). In this paper, we extend our previous work to achieve more efficient surface normal estimation by treating reference BRDFs in a discrete manner. This simple extension connects photometric stereo to nearest-neighbor search, which results in one to four orders of magnitude faster inference while maintaining accuracy.

- *K. Enomoto, M. Waechter, F. Okura, and Y. Matsushita are with Graduate School of Information Science and Technology, Osaka University. E-mail: {enomoto.kenji, okura, yasumat}@ist.osaka-u.ac.jp*
- *K. N. Kutulakos is with Department of Computer Science, University of Toronto. E-mail: kyros@cs.toronto.edu*

Manuscript received April 19, 2005; revised August 26, 2015.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

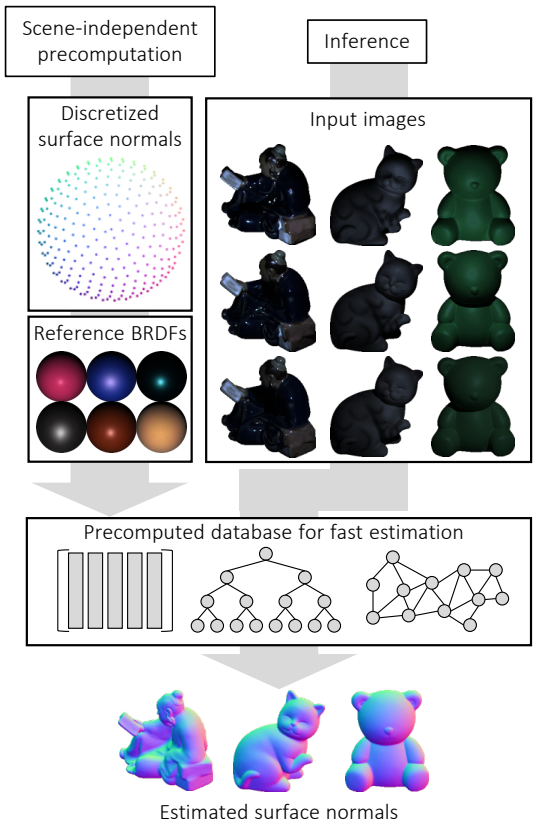


Fig. 1: An overview of our algorithm, Discrete Search Photometric Stereo (DSPS). It precomputes a database from discretized surface normals and reference BRDFs in a scene-independent manner. During inference from input images, surface normals of any scene are efficiently recovered by searching the precomputed database.

2 Related Work

This section describes previous non-Lambertian photometric stereo and their relation to our methods. Modern non-Lambertian photometric stereo can be roughly categorized into model-based, exemplar-based, and learning-based methods.

2.1 Model-based photometric stereo

Model-based methods use parametric expressions for BRDFs. The model parameters, including the surface normal, are typically estimated via optimization. The key for these methods is the choice of a parametric BRDF model. Woodham's original work [1] 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 [10], the Torrance-Sparrow model [11], the Ward model [12], the specular spike model [13], [14], and a microfacet BRDF with ellipsoidal normal distributions [5] have been developed. While these models lead to non-convex optimization for photometric stereo, some recent methods use bivariate functions to avoid the non-convexity. For representing low-frequency reflectances, Shi *et al*. [3] use a bi-polynomial function and Ikehata and Aizawa [4] use a sum of lobes with unknown center directions. Although model-based methods can be used on a wide range of materials, they always involve a trade-off between representative power and optimization complexity.

2.2 Exemplar-based photometric stereo

Early work on exemplar-based (also called example-based) photometric stereo relies on the concept of orientation consistency [6], *i.e.*, two surfaces with the same surface normal and BRDF will have the same appearance under the same illumination. Another work along this direction is found in Horn and Ikeuchi [15]. In these approaches, a reference object with known surface normals is placed in a target scene and the reference object's BRDF is assumed to be the same as the target object's. 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 [6] introduced two reference objects, diffuse and specular spheres, placed in the target scene. They approximate the target BRDF by a non-negative linear combination of the reference BRDFs.

Hui and Sankaranarayanan [7] introduced virtual exemplar-based photometric stereo that performs exemplar-based photometric stereo without actually introducing reference objects in the target scene. They render virtual exemplars of appearances 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 their method, there are time-consuming processes such as rendering virtual exemplars, 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.

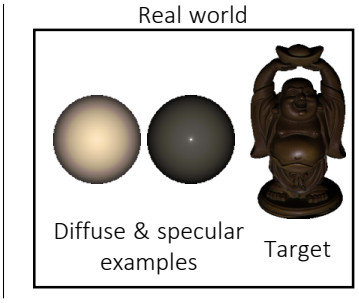
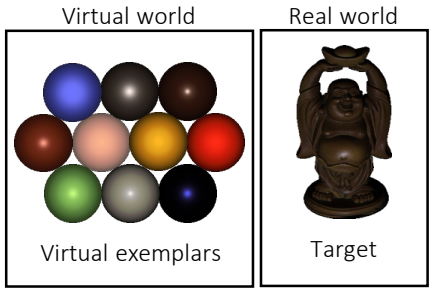
Our DSPS is categorized as an exemplar-based method that does not require reference objects. Unlike virtual exemplar-based methods, our DSPS allows the exhaustive discrete search that guarantees to reach the globally optimal solution within the bounds of the objective function. Our discrete search strategy in DSPS is accelerated through scene-independent precomputation. Moreover, our DSPS-D treats BRDFs as well as surface normals in a discrete space, which makes the surface normal estimation problem similar to classic nearest-neighbor search. This allows using any fast nearest-neighbor search method for efficiency without sacrificing accuracy. The differences among exemplar-based photometric stereo methods, including our methods, are summarized in Table 1.

2.3 Learning-based photometric stereo

Recently, deep learning-based photometric stereo methods have been proposed. They learn a mapping from measured intensities under known illuminations to surface normals using a neural network. These methods show strong results on various scenes due to the network being trained with diverse shapes and materials. In particular, learning-based methods effectively deal with global illumination effects,

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

TABLE 1: Comparison of exemplar-based photometric stereo methods and their properties.

	Hertzmann & Seitz [6]	Hui & Sankaranarayanan [7]	our DSPS-C	our DSPS-D
surface normal representation	discrete (example-based)	discrete to continuous	discrete	
BRDF representation	continuous (non-negative linear combinations)	continuous (non-negative linear combinations)	continuous (linear combination)	discrete
solution method	iterative non-negative least squares	iterative non-negative least squares	closed-form least squares	nearest-neighbor search
setting				

such as cast shadows and inter-reflections, which are difficult for model-based and exemplar-based methods, by including such effects in the training data. Santo *et al*. [17] and Chen *et al*. [18] created a training dataset by rendering the Blobby [19] and Sculpture [20] shape datasets with 100 MERL BRDFs [16]. Ikehata [21] also introduced a training dataset, called CyclesPS dataset, containing several objects rendered with a diverse set of materials from Disney's principled BSDFs [22] with global illumination effects. Logothetis *et al*. [23] proposed a per-pixel data generation strategy considering global illumination effects to simplify and speed up the rendering. Typical learning-based methods suffer from sparse light configurations, which is subsequently addressed by some recent papers [24], [25], [26]. Wang *et al*. [27] also addressed surface normal recovery under sparse lightings using monotonicity of isotropic reflectance and a special lighting setup with a collocated light. Beside learning-based methods in supervised settings, Taniai and Maehara [28] proposed an unsupervised method that minimizes the reconstruction loss between input and re-rendered images.

Our DSPS-D, which uses nearest-neighbor search, can be considered to be a learning-based approach as it is a "lazy learner" that memorizes the entire training dataset. An advantage of nearest-neighbor search is the simplicity of the training compared to deep learning approaches. Much like the growth in datasets in various machine learning tasks such as image classification [29], [30], it is expected that datasets for photometric stereo will also grow. Therefore, we consider that it may raise issues in stable learning for neural networks, such as the issue of training on a biased dataset [31], [32]. In contrast, nearest-neighbor search is less affected by biases in training datasets since it only requires that training datasets contain data similar to an input query.

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 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*. [6], and Hui and Sankaranarayanan [7]. 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

$$\mathbf{m} = \mathbf{E}\mathbf{B}\mathbf{c} \stackrel{\text{def}}{=} \mathbf{D}\mathbf{c} \tag{2}$$

where $\mathbf{D}(= \mathbf{E}\mathbf{B}) \in \mathbb{R}_+^{L' \times M}$.

Problem statement: Our goal is to find the optimal surface normal $\mathbf{n}$ for each surface point, given measurements $\mathbf{m}$ and associated light directions $\{\mathbf{l}_1, \dots, \mathbf{l}_{L'}\}$ based on Eq. (2).

4 Proposed Method

This section presents our DSPS algorithm with continuous and discrete BRDF models.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

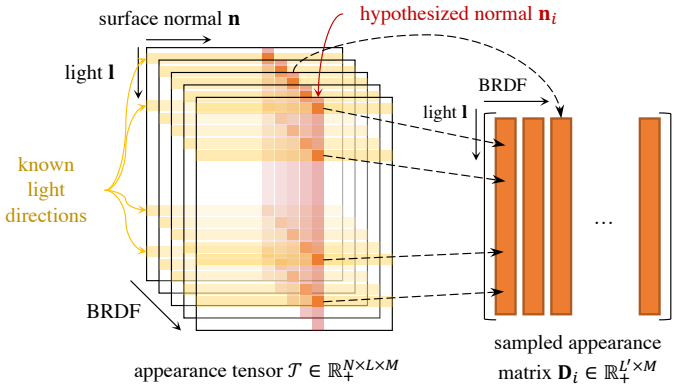


Fig. 2: Starting from the appearance tensor $\mathcal{T}$ that represents appearances for a comprehensive set of light directions, surface normals, and BRDFs, we slice out a sampled appearance matrix $\mathbf{D}_i$ for a set of known light directions and a hypothesized surface normal $\mathbf{n}_i$.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

4.1 Discrete Search Photometric Stereo with a continuous BRDF model

Our Discrete Search Photometric Stereo with a continuous BRDF model (DSPS-C) casts the photometric stereo problem as 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, DSPS-C is able to find the globally optimal surface normal $\mathbf{n}$ that best satisfies Eq. (2).

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 the set of surface normal *candidates*. We prepare a tensor representation for diverse appearances whose axes are (1) surface normals, (2) light directions, and (3) BRDFs. 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 appearance 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 appearance 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 appearance matrix* $\mathbf{D}_i \in \mathbb{R}_+^{L'\times M}$ from the appearance tensor $\mathcal{T}$ along the hypothesized surface normal $\mathbf{n}_i$ and a set of L' known light directions as illustrated in Fig. 2. Using $\mathbf{D}_i$ instead of $\mathbf{D}$, Eq. (2) becomes

$$\mathbf{m} \simeq \mathbf{D}_i \mathbf{c}.$$

For the overdetermined case $L' > M$, the least-squares solution for the BRDF coefficients $\mathbf{c}$ that best explains the measurements is

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

where $\mathbf{D}_i^\dagger$ is the pseudo-inverse of $\mathbf{D}_i$. The estimated BRDF coefficients $\mathbf{c}_i$ are least-squares optimal for the hypothesized normal $\mathbf{n}_i$ and the space of sampled appearances $\mathbf{D}_i$. We can

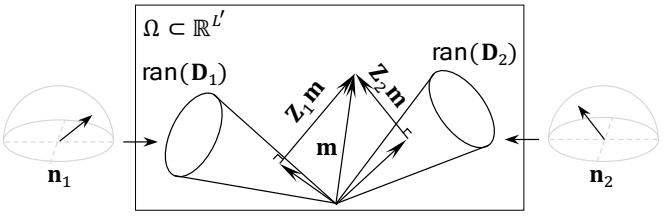


Fig. 3: Geometric interpretation of the measurement reconstruction error. The reconstruction error of measurements $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ can be seen as distance between the measurement vector $\mathbf{m}$ and the subspace spanned by $\mathbf{D}_i$ in the L'-dimensional space Ω.

test the validity of the hypothesized $\mathbf{n}_i$ by evaluating the ℓ_2 measurement reconstruction error as

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

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 computational effort for solving this problem. We thus introduce an efficient scene-independent precomputation strategy.

Scene-independent precomputation: The reconstruction error e_i in Eq. (3) can be further simplified as

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

As long as the lighting and BRDF bases are fixed, $\mathbf{Z}_i (= \mathbf{I} - \mathbf{D}_i\mathbf{D}_i^\dagger) \in \mathbb{R}^{L'\times L'}$ is uniquely determined given a normal hypothesis $\mathbf{n}_i$. We, thus, can precompute a set of $\{\mathbf{Z}_i\}$ for all normal candidates in $\mathcal{N}$. At inference 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 with the same lighting.

Dimensionality reduction of sampled appearance matrix: Eq. (4) is only a necessary condition for correct surface normal solution. When the sampled appearance matrix $\mathbf{D}_i$ has fewer rows than columns or when $\mathbf{m} \in \mathsf{ran}(\mathbf{D}_i) \in \mathbb{R}^{L'\times M}$ ($\mathbf{D}_i$'s range) for all $\mathbf{D}_i$, there exist greater than or equal to one BRDF coefficient vectors $\mathbf{c}_i$ that make all reconstruction errors $\{e_i\}$ zero.

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 $\mathbf{D}_i\mathbf{c}_i = \mathbf{D}_i\mathbf{D}_i^\dagger\mathbf{m}$ reside in this subspace. Thus, the reconstruction error $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ can be seen 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 and the reconstruction error becomes always zero regardless of the correctness of the surface normal hypothesis $\mathbf{n}_i$.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

To avoid this, we replace $\mathbf{D}_i$ with its first $M'(< L')$ left singular vectors $\mathbf{U}'_i \in \mathbb{R}^{L' \times M'}$ obtained through SVD. With this, $\mathbf{Z}_i$ can be precomputed in a simpler form as

$$\mathbf{Z}_i = \mathbf{I} - \mathbf{U}'_i \mathbf{U}'^{\dagger}_i = \mathbf{I} - \mathbf{U}'_i \mathbf{U}'^{\top}_i$$

due to the orthogonality of each singular vector.

The effective value of M' is surprisingly small, around three even for $L' \sim 100$ (see the supplementary material for details). It indicates that the best surface normal can be estimated even if Eq. (2) is not strictly satisfied. While the dimensionality reduction reduces the span of the original BRDF bases, more BRDF bases are preferable to find a better subspace of $\mathbf{D}_i$.

4.2 Discrete Search Photometric Stereo with a discrete BRDF model

We further present Discrete Search Photometric Stereo with a discrete BRDF model (DSPS-D), that treats both surface normals and BRDFs in a discrete manner. DSPS-D *hypothesizes* a surface normal $\mathbf{n}$ and a BRDF $\rho(\cdot)$ and *tests* whether they approximately satisfy an image formation model. Dropping the constraint that the BRDF model needs to be closed under linear combinations may appear to diminish the expressiveness of the model, because it means that our appearance tensor needs to relatively densely cover the space of materials that we want to infer. However, DSPS-D can perform much faster surface normal estimation than DSPS-C using existing nearest-neighbor search techniques and can in practice achieve an accuracy comparable to DSPS-C.

Let $\mathcal{N} = \{\mathbf{n}_i \mid i = 1, \ldots, N\}$ and $\mathcal{B} = \{\rho_j(\cdot) \mid j = 1, \ldots, M\}$ be sets of discretized surface normals and reference BRDFs, which we call surface normal candidates and BRDF candidates, respectively. If we hypothesize a surface normal $\mathbf{n}_i \in \mathcal{N}$ and BRDF $\rho_j(\cdot) \in \mathcal{B}$ for a scene point, using $L' \leq L$ known light directions, we can slice a *sampled appearance vector* $\mathbf{d}_{ij} \in \mathbb{R}^{L'}_{+}$ from the appearance tensor $\mathcal{T}$. With unknown scaling s, the measurement vector $\mathbf{m}$ can be approximated as follows

$$\mathbf{m} \simeq s\mathbf{d}_{ij}. \tag{5}$$

We can transform this into a scaling-free form as

$$\tilde{\mathbf{m}} \simeq \tilde{\mathbf{d}}_{ij}$$

where $\tilde{\mathbf{m}} = \frac{\mathbf{m}}{\|\mathbf{m}\|_2}$ and $\tilde{\mathbf{d}}_{ij} = \frac{\mathbf{d}_{ij}}{\|\mathbf{d}_{ij}\|_2}$. We can test the validity of the hypothesized $\mathbf{n}_i$ and $\rho_j(\cdot)$ by the ℓ_2 measurement reconstruction error, and, therefore, the optimal surface normal $\mathbf{n}^*$ and BRDF $\rho^*(\cdot)$ can be found as

$$\mathbf{n}^* = \mathbf{n}_{i^*},\ \rho^*(\cdot) = \rho_{j^*}(\cdot)$$
$$i^*, j^* = \underset{i,j}{\operatorname{argmin}} \left\| \tilde{\mathbf{m}} - \tilde{\mathbf{d}}_{ij} \right\|_2^2.$$

This objective is equivalent to the one of nearest-neighbor search with Euclidean distance; hence, we can rely on any exact or approximate nearest-neighbor search method to search for the optimal surface normal and BRDF.

Shadow masking: Our nearest neighbor search strategy can naturally take additional sampled appearance vectors. Here, we introduce a shadow masking strategy for augmenting the sample appearance vectors to deal with cast shadows that has been ignored in our image formation.

We simulate cast shadows by masking several elements of the sampled appearance vectors. It is empirically known that cast shadows appear with some regularity instead of randomly [24]; therefore, we apply the method for the occlusion layer proposed by Li *et al*. [24] to generate shadow masks. Importantly, the shadow masks are generated for every sampled appearance vector independently to simulate diverse patterns of shadows. Once the shadow masks are applied to the original sampled appearance vectors, the masked appearance vectors are appended to the original ones. The shadow masking is performed K times to all the original sampled appearance vectors, leading $K + 1$ times larger set of sampled appearance vectors.

5 Experiments

This section describes experiments on DSPS-C's and DSPS-D's accuracy and computational efficiency using synthetic and real-world data. We also show comparisons with recent photometric stereo methods.

5.1 Implementation

DSPS-C: For all experiments in this paper, we set $M' = 3$, which was the most robust to noisy images.[1] Before applying SVD, we normalize each column of $\mathbf{D}_i$ as with the existing work [7].

DSPS-D: Our DSPS-D can benefit from any exact or approximate nearest-neighbor search method based on ℓ_2 distance (*e.g.*, [33], [34], [35], [36], [37]) implemented in modern libraries [36], [38], [39], [40]. In our experiments, we used a simple linear search algorithm implemented in FAISS [39] as an exact method. As an approximate method, we adopted a combination of an inverted file system with asymmetric distance computation (IVFADC) [41] and a hierarchical navigable small worlds (HNSW) indexing structure [42] implemented in FAISS [39]. (See the supplemental material for details on the hyper-parameters.) DSPS-D using FAISS can be performed on either a CPU or a GPU.

In the following, we denote DSPS-D with exact and approximate nearest-neighbor search as **DSPS-DE** and **DSPS-DA**, respectively.

5.2 Preparation

Appearance tensor: The appearance tensor is constructed from three components; BRDFs, surface normals, and light directions. For BRDFs, we used the MERL BRDF database [16] which consists of 100 distinct BRDFs including diffuse, specular, and metallic materials.[2] For surface normal sampling we followed Hui's method [7] and obtained 20001 candidates. In all experiments of this paper, we assume that the appearance tensor contains the known light directions. In Sec. 5.9, we discuss how the surface normal estimation accuracy is affected by the discretization of light directions.

Here we explain the test datasets used for evaluating the proposed method.

1. See the experimental analysis in the supplementary material.
2. See the supplementary material for analysis of our method with appearance tensors constructed from other BRDFs.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

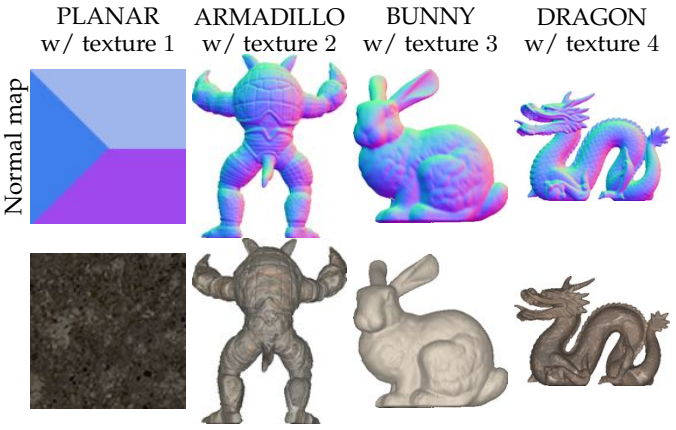


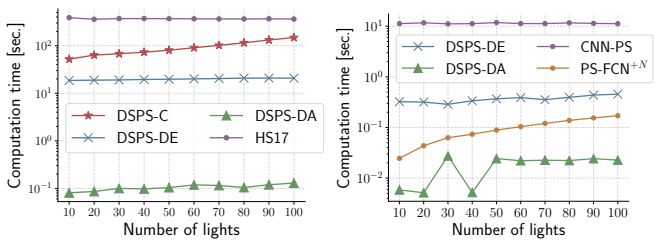
Fig. 4: Ground truth surface normals and example images of PrincipledPS dataset.

MERL sphere dataset: The MERL sphere dataset consists of 100 synthetic sphere scenes rendered with the 100 MERL BRDFs [16]. We rendered the images under ten lighting environments consisting of $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ uniformly distributed light sources. (See the supplemental material for illustrations of the light distributions.) Image resolution was set to 100×100, yielding 7860 valid pixels.

PrincipledPS dataset: To quantitatively evaluate our method on varying sets of BRDFs, textures, and shapes, we rendered a synthetic dataset including PLANAR, BUNNY, DRAGON, and ARMADILLO shapes with the Principled BSDFs [22]. We call this dataset as *PrincipledPS*. For each shape, we prepared two materials, Specular and Metallic, as defined by Ikehata [21], four spatially varying textures, and sparse and dense (10 and 100) light configurations, totally, 64 scenes. Figure 4 shows the ground truth surface normal maps and example images of the PrincipledPS dataset.

Real-world dataset: We use an existing real-world dataset, the DiLiGenT dataset [43], which contains 10 real objects of general reflectance illuminated from 96 different known directions. This dataset provides the ground truth surface normal maps for all objects measured by high-precision laser scanning that can be used for quantitative evaluation. For the Bear object we discarded the first 20 images where a part of measurements is corrupted as pointed out by Ikehata [21]. In addition to the original dataset, for testing sparse light cases, we prepared 20 datasets, each containing 10 randomly selected images.

As baselines we used Lambertian photometric stereo (LPS) [1], the model-based method ST14 [3], the virtual exemplar-based method HS17 [7], the unsupervised learning (*i.e.*, neural inverse rendering)-based method NIR-PS [28], the supervised learning methods PX-NET [23], PS-FCN^{+N} [18], WJ20 [27], CNN-PS [21], and SPLINE-Net [25]. For a fair comparison in computation time, we reimplemented HS17 in Python based on the authors' MATLAB implementation. We solve the non-negative least-squares sub-problem in HS17 using `scipy.optimize.nnls` from the SciPy package [38] resulting in the authors' implementation speedup without any accuracy drop. We implemented the coarse-to-fine search they proposed for efficient surface normal estimation following their original implementation.



(a) (b)

Fig. 5: (a) *CPU* computation time of our methods and HS17. (b) *GPU* computation time of DSPS-D and CNN-PS.

Since PS-FCN^{+N} is trained on a dataset with MERL BRDFs, for fear of data leakage we omit PS-FCN^{+N} in the experiments on the MERL sphere dataset. While the published, pre-trained SPLINE-Net model has been trained specifically for 10 lights, it works well for other small numbers of light sources. Therefore, we show SPLINE-Net's scores for cases other than 10 lights for reference in this paper. Further, for testing with the MERL sphere dataset, although PX-NET, PS-FCN^{+N}, and WJ20 include the target material in their pre-trained models, we list their scores for reference.

5.3 Efficiency of surface normal estimation

This section shows comparisons of computation times with the baseline methods running on CPU and/or GPU. We use the MERL sphere dataset with the ten light sets. We measured the computation times of DSPS-C, DSPS-DE, DSPS-DA, and the existing exemplar-based method HS17 on a CPU. We also measured the computation time of DSPS-DE, DSPS-DA, CNN-PS, and PS-FCN^{+N} on a GPU. In this section, we eliminate the results of inefficient iterative methods, ST14 and NIR-PS, and the extension of CNN-PS, *i.e.*, PX-NET and SPLINE-Net, that are always slower than CNN-PS. We used 40 cores of an Intel® Xeon® Gold 6148 CPU @ 2.40 GHz and NVIDIA TITAN X GPU. On the CPU we performed pixel-wise parallelization. Note that our methods are executable on common CPUs and GPUs because the sampled appearance matrix only requires a small amount of memory. For example, sampled appearance matrices stored in 64-bit floating point numbers for a typical setting, where $N = 20001$, $M = 100$, $L = 100$, only require 3.1 GB storage space.

Figure 5a shows the computation time for one scene on the CPU, averaged over all MERL spheres for each light configuration. DSPS-DA is 3-4 orders of magnitude faster than HS17 while DSPS-C and DSPS-DE are around one order of magnitude faster than HS17. Figure 5b shows the computation time for one scene on the GPU. DSPS-DE and DSPS-DA are accelerated one order of magnitude using the GPU. While typical exemplar-based approaches are computationally expensive, our methods achieve comparable or faster inference than the learning-based methods using feed-forward networks.

5.4 Accuracy of surface normal estimation

We estimated surface normals on synthetic and real datasets to confirm that our methods work with diverse scenes.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

TABLE 2: Comparisons on the MERL sphere dataset with light configuration 10 sets. Numbers represent averages and standard deviations of angular errors over all pixels.

	#lights	10	20	30	40	50	60	70	80	90	100
Exemplar-based	DSPS-C	4.2/6.6	2.5/3.6	2.2/3.0	2.1/2.9	2.0/2.8	1.9/2.7	1.9/2.7	1.9/2.6	1.8/2.7	1.8/2.7
	DSPS-DE ($K=0$)	3.0/4.3	2.2/3.1	2.0/2.8	1.9/2.6	1.8/2.5	1.8/2.4	1.7/2.4	1.7/2.4	1.7/2.4	1.7/2.4
	DSPS-DA ($K=0$)	3.0/4.3	2.3/3.1	2.1/2.8	2.1/2.7	2.0/2.5	2.0/2.5	2.0/2.5	2.0/2.5	2.0/2.5	2.0/2.5
	DSPS-DE ($K=1$)	3.2/4.8	2.3/3.5	2.1/3.0	1.9/2.7	1.8/2.6	1.8/2.5	1.7/2.5	1.7/2.4	1.7/2.4	1.7/2.4
	DSPS-DA ($K=1$)	3.2/4.8	2.4/3.5	2.2/3.0	2.1/2.8	2.1/2.6	2.1/2.6	2.1/2.6	2.1/2.6	2.1/2.6	2.1/2.6
	HS17	3.6/5.2	2.2/3.3	1.9/2.8	1.8/2.6	1.7/2.5	1.6/2.4	1.6/2.4	1.7/3.5	1.6/2.4	1.6/2.4
Learning-based	PX-NET[a]	13.4/14.3	11.0/14.3	9.3/12.8	9.7/12.9	3.5/7.2	3.4/7.4	3.4/7.9	3.5/8.2	3.5/8.3	3.5/8.5
	PS-FCN^{+N}[a]	4.5/4.6	2.7/2.6	2.7/2.5	3.0/2.7	3.1/2.7	3.2/2.9	3.4/3.0	3.4/3.0	3.6/3.1	3.7/3.2
	WJ20[a]	3.7/4.2	3.3/3.5	3.2/3.3	3.2/3.4	3.3/3.3	3.2/3.3	3.3/3.3	3.3/3.3	3.3/3.4	3.3/3.2
	SPLINE-Net	13.0/20.0	9.3/16.1	10.2/13.1	15.9/18.4	27.5/28.8	38.8/33.8	45.5/34.8	49.0/33.7	51.4/32.9	50.0/31.5
	CNN-PS[b]	33.6/23.9	6.2/6.4	4.7/5.7	4.0/5.3	3.7/5.2	3.2/4.6	3.0/4.2	2.9/4.3	2.6/3.9	2.5/3.8
	NIR-PS	21.7/44.8	15.6/36.3	18.0/40.8	15.2/37.0	18.9/42.5	16.0/38.5	14.8/35.7	14.4/34.2	13.7/33.5	14.6/34.3
Model-based	ST14	15.5/9.9	11.5/15.6	10.9/13.7	10.9/13.9	9.8/13.4	5.5/8.1	2.7/4.4	1.7/3.1	1.4/2.6	1.2/2.3
	LPS	13.6/9.9	13.0/9.4	12.8/9.4	12.7/9.3	12.7/9.3	12.6/9.3	12.6/9.4	12.6/9.4	12.6/9.4	12.6/9.4

[a] Training dataset of PX-NET, PS-FCN^{+N}, and WJ20 include target materials. [b] CNN-PS is trained with 50-100 lights.

MERL sphere: We compared our methods and the baseline methods using the MERL sphere dataset. For the materials in our methods and HS17, we applied a leave-one-out scheme, testing them on one MERL BRDF while constructing the appearance tensor from the remaining 99 BRDFs so that the appearance tensor does not contain the target BRDF.

Table 2 shows the averages and standard deviations of angular errors over all pixels in the MERL sphere dataset for the ten light configuration sets. The small averages and standard deviations show that our methods stably yield small errors in all light configurations when compared with the baseline methods. While HS17 also achieves competitive accuracy, it is more than three orders of magnitude slower than DSPS-DA as shown previously. Incidentally, NIR-PS yields large angular errors in this experiment. We observed that NIR-PS has extremely large errors for several materials, which affect the averaged scores. We show mean angular errors of the baseline methods for each material in the supplemental material.

PrincipledPS: We conducted quantitative evaluation on the PrincipledPS dataset. While training datasets of PX-NET, CNN-PS, and SPLINE-Net are also rendered with the Principled BSDFs and therefore may include the target materials, their scores are shown as reference.

Table 3 shows averages of angular errors over four scenes, *i.e.*, four textures, for each shape, material, and number of lights. The results on the PrincipledPS dataset also show that our DSPS-D achieves accurate surface normal estimation for every scene with both sparse and dense lightings. PS-FCN^{+N} and WJ20 also yield promising results; however, the different behavior than ours is observed especially when few lights on the PLANAR, which is an extreme shape but often appears in the real-world. One possible reason for the difference is that PS-FCN^{+N} and WJ20 use patch-based processing, *i.e.*, their surface normal estimates depend on not only local appearances but also global appearances. Therefore, the accuracy of patch-based methods slightly degrades on scenes with non-informative global appearances.

Table 3 exhibits that DSPS-C causes a large error on the metallic PLANAR scene. The main reason of the large error is due to that the choice of $M' = 3$ was too large for the 10 lights case, making DSPS-C unstable in the few lights case. The second reason is that DSPS-C exhibits a higher standard deviation for metallic objects, and the surface normals contained in the PLANAR scene were particularly hard for DSPS-C by chance. In the supplementary material, we verified these with an additional experiment using SPHERE scenes having the identical materials and textures with the PLANAR scenes and confirmed that the mean angular error of DSPS-C converges to the comparable one with other exemplar-based methods if M' is suitable and a target scene has diverse surface normals.

DiLiGenT: We show quantitative results on the real-world dataset DiLiGenT in Table 4, where we compare our methods with the baseline methods in terms of mean angular error. Figure 6 shows visual comparisons between our methods and the baseline methods. Our DSPS methods demonstrate comparable or better accuracy compared to the exemplar-based methods, although showing a slight degradation compared to the learning-based methods.

For the scenes with 96 lights, DSPS-DE achieves the best score on the BALL object having fully convex surfaces. The same trend of high accuracy on convex regions can be observed in other scenes, *e.g.*, the body of the COW object and the arm of the READING object in Fig. 6. Even for non-convex regions, our DSPS-DE with the shadow masking yields robust surface normal estimates as shown by the improvement on DSPS-DE owing to the shadow masking in Table 4. Still, the learning-based methods are superior to our methods particularly on BUDDHA, GOBLET, HARVEST, or READING, where strong inter-reflections are observed.

For the scenes with 10 lights, our methods, especially with shadow masking, achieve comparable accuracy to the learning-based methods. The standard deviations of our DSPS-D tend to be small compared to the baselines, which suggest that DSPS-D is insusceptible to the light distributions. This robustness is preferable since it is hard to know which light distribution is the best for each method in practice.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

TABLE 3: Comparisons on the PrincipledPS dataset. Numbers represent averages of angular errors over four scenes, *i.e.*, four textures. S and M indicate specular and metallic, respectively.

		10 lights									100 lights								
		PLANAR		ARMADILLO		BUNNY		DRAGON		Avg.	PLANAR		ARMADILLO		BUNNY		DRAGON		Avg.
		S	M	S	M	S	M	S	M		S	M	S	M	S	M	S	M	
Exemplar-based	DSPS-C	3.6	28.2	3.4	4.6	3.4	4.5	3.4	4.8	7.0	1.2	1.1	1.7	3.7	1.8	3.4	1.7	3.6	2.3
	DSPS-DE ($K=0$)	1.2	1.9	2.4	4.7	2.4	4.6	2.4	4.9	3.0	1.0	1.1	1.9	4.9	1.8	4.5	1.8	4.6	2.7
	DSPS-DA ($K=0$)	1.1	1.8	2.4	4.7	2.4	4.6	2.4	4.9	3.0	1.4	1.8	2.1	5.2	2.1	4.9	2.0	4.8	3.0
	DSPS-DE ($K=1$)	1.2	1.9	2.4	4.7	2.4	4.6	2.4	4.9	3.1	1.0	1.1	1.9	4.9	1.8	4.5	1.8	4.6	2.7
	DSPS-DA ($K=1$)	1.1	2.0	2.4	4.7	2.5	4.6	2.4	4.9	3.1	1.2	1.3	2.2	5.1	2.2	4.8	2.1	4.8	3.0
	HS17	1.5	1.6	3.1	4.7	2.9	4.6	2.9	4.7	3.3	1.3	1.0	1.8	2.9	1.6	2.7	1.8	2.7	2.0
Learning-based	PX-NET[a]	5.8	15.5	4.4	5.2	5.1	5.5	4.4	5.5	6.4	1.2	1.1	1.5	1.3	1.5	1.2	1.6	1.3	1.4
	PS-FCN^{+N}	2.2	11.6	3.1	7.1	2.5	7.0	3.2	7.6	5.6	1.7	4.1	2.6	4.6	2.5	4.5	2.6	4.9	3.4
	WJ20	2.4	7.5	2.8	4.4	2.9	4.1	2.7	4.6	3.9	2.0	2.7	2.4	3.5	2.5	3.0	2.3	3.5	2.7
	SPLINE-Net[a]	5.9	12.4	5.9	6.5	5.8	7.0	5.7	7.2	7.0	49.6	17.9	36.3	46.2	43.4	47.4	36.0	48.0	40.6
	CNN-PS[ab]	29.9	27.6	30.9	29.7	36.4	32.0	32.0	29.0	30.9	2.2	7.7	1.5	2.0	2.0	2.0	1.6	2.1	2.6
	NIR-PS	1.5	94.6	2.2	3.6	1.3	3.6	2.1	4.0	14.1	0.4	81.7	2.1	3.8	1.5	3.6	2.1	3.4	12.3
Model-based	ST14	6.8	23.1	13.2	11.2	14.1	11.2	12.6	10.1	12.8	0.5	2.7	7.7	6.4	2.7	2.2	7.7	6.6	4.6
	LPS	4.2	23.2	10.3	10.2	10.8	9.9	9.5	9.0	10.9	3.3	23.4	9.0	7.5	8.9	7.6	7.9	7.2	9.4

[a] Training dataset of PX-NET, CNN-PS, and SPLINE-Net may include target materials. [b] CNN-PS is trained with 50-100 lights.

TABLE 4: Comparisons on the DiLiGenT dataset with 96 and 10 lights. Numbers in the table above are mean angular errors in degrees. Numbers in the table below are averages and standard deviations of mean angular errors over 20 trials.

		96 lights										
		BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
Exemplar-based	DSPS-C	1.6	5.9	13.1	6.1	9.2	11.0	18.7	6.6	7.2	15.0	9.4
	DSPS-DE ($K=0$)	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
	DSPS-DA ($K=0$)	1.4	6.4	14.2	6.8	8.0	11.7	17.5	7.4	7.4	15.3	9.6
	DSPS-DE ($K=1$)	1.3	5.3	11.7	5.9	7.3	10.2	16.7	6.6	7.0	14.0	8.6
	DSPS-DA ($K=1$)	1.4	5.3	11.8	5.9	7.5	10.5	16.8	6.6	7.1	14.0	8.7
	HS17	1.5	6.2	13.9	6.4	9.2	10.8	18.8	7.0	7.9	15.3	9.7
Learning-based	PX-NET	2.0	3.5	7.6	4.3	4.7	6.7	13.3	4.9	5.0	9.8	6.2
	PS-FCN^{+N}	2.6	5.4	7.5	4.7	6.7	7.8	12.4	5.9	7.2	10.9	7.1
	WJ20	1.8	4.1	6.1	4.7	6.3	7.2	13.3	6.5	6.4	10.0	6.6
	CNN-PS	2.1	4.2	8.1	4.4	7.9	7.4	13.8	5.4	6.4	12.1	7.2
	NIR-PS	1.6	6.1	11.0	5.6	5.8	11.2	22.0	6.5	8.5	11.3	9.0
Model-based	ST14	1.8	5.1	10.7	6.1	13.8	10.2	25.6	6.5	8.7	13.0	10.2
	LPS	4.2	8.5	14.9	8.4	25.6	18.5	30.6	8.9	14.6	20.0	15.4

		10 lights										
		BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
Exemplar-based	DSPS-C	4.2/1.2	7.9/0.7	16.7/1.4	8.5/1.0	12.4/1.2	15.2/1.3	24.2/0.8	9.3/0.8	11.7/1.7	21.1/1.6	13.1
	DSPS-DE ($K=0$)	2.4/0.5	7.7/0.7	16.1/0.8	8.0/0.4	10.5/0.6	14.0/0.6	20.9/0.6	8.8/0.4	9.8/0.7	18.1/1.1	11.6
	DSPS-DA ($K=0$)	2.5/0.5	7.7/0.7	16.0/0.8	8.0/0.4	10.6/0.6	14.0/0.6	20.9/0.6	8.8/0.4	9.9/0.7	18.0/1.1	11.6
	DSPS-DE ($K=1$)	2.4/0.5	6.1/0.4	12.3/0.7	6.9/0.4	9.7/0.6	12.0/0.7	19.4/0.4	7.6/0.5	9.1/0.7	15.6/1.1	10.1
	DSPS-DA ($K=1$)	2.4/0.5	6.1/0.4	12.3/0.7	6.9/0.4	9.7/0.6	11.9/0.6	19.4/0.4	7.6/0.5	9.1/0.7	15.6/1.1	10.1
	HS17	3.8/0.9	8.1/0.8	16.3/1.0	8.5/0.6	12.9/1.1	14.1/0.7	22.0/0.7	9.2/0.6	11.1/1.0	18.2/1.3	12.4
Learning-based	PX-NET[a]	2.3/0.4	4.7/0.3	9.6/0.5	6.3/0.4	7.3/0.6	9.6/0.9	16.2/0.7	7.0/0.4	7.8/1.1	13.5/0.8	8.4
	PS-FCN^{+N}	4.3/1.0	6.8/0.8	9.7/0.8	6.3/0.6	12.2/1.3	10.5/0.8	17.5/1.0	7.7/0.6	10.0/1.2	13.0/1.1	9.8
	WJ20[b]	2.2/0.4	5.0/0.2	7.0/0.3	5.5/0.2	7.2/0.6	8.7/0.7	15.1/0.7	7.0/0.4	8.1/0.9	10.9/0.8	7.7
	SPLINE-Net	5.1/1.0	5.9/0.6	10.7/1.0	7.9/0.9	9.0/1.1	10.7/1.2	19.2/1.0	9.4/0.8	12.5/1.4	15.3/0.8	10.6
	CNN-PS[c]	10.2/5.5	14.2/4.8	15.0/4.3	12.4/5.8	13.9/1.8	15.5/2.8	20.3/2.6	12.9/4.8	14.9/3.6	16.4/3.5	14.6
	NIR-PS	1.6/0.2	5.9/0.6	10.9/0.8	6.2/0.4	13.3/6.5	16.8/10.0	28.5/4.1	8.0/4.6	8.9/1.0	15.3/4.7	11.5
Model-based	ST14	5.7/0.6	10.0/0.4	16.4/0.7	9.6/0.5	26.3/0.8	20.0/0.9	31.0/0.7	10.2/0.4	16.2/1.0	19.7/1.3	16.5
	LPS	4.6/0.5	9.0/0.4	15.9/0.7	9.2/0.4	26.6/0.7	19.7/0.9	31.4/0.6	9.6/0.4	15.6/1.0	20.2/1.4	16.2

[a] A model specific to few lights is used. [b] $10+1$ lights are used, where 1 light is nearly collocated with the camera.
[c] CNN-PS is trained with 50-100 lights.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

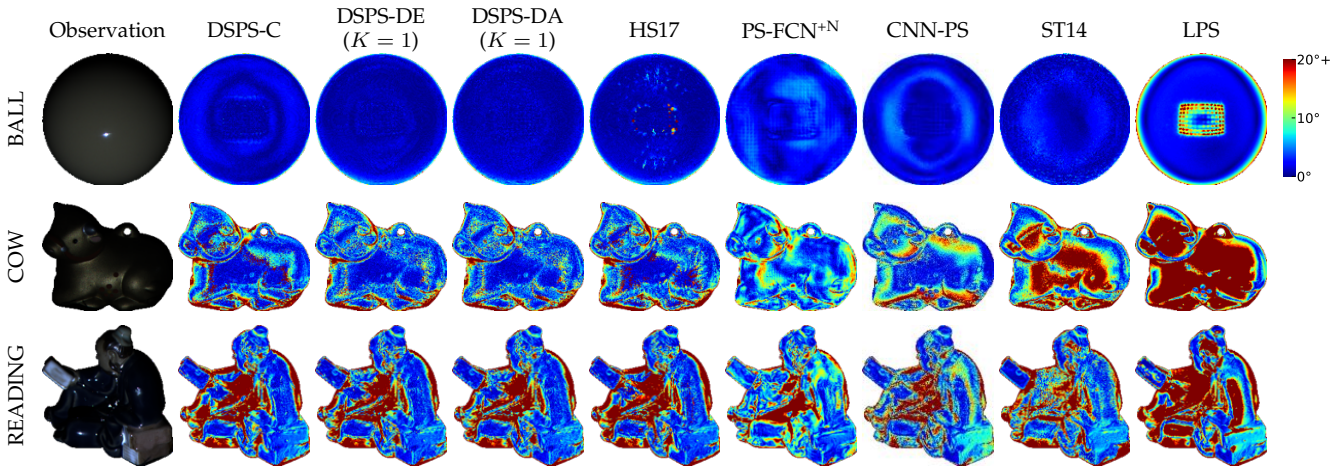


Fig. 6: Angular error maps for BALL, COW, and READING objects from the DiLiGenT dataset [43] with all the 96 lights. See the supplementary materials for more objects.

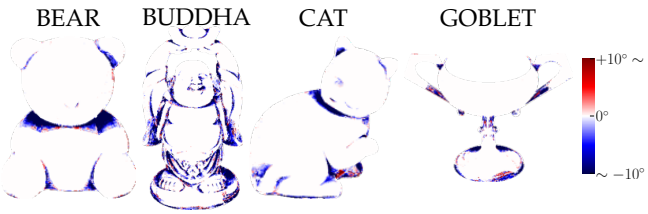


Fig. 7: Difference in the angular error maps between our DSPS-DE with and without shadow masking. Blue indicates pixels, where surface normal estimation is improved by shadow masking, and red indicates the opposite.

TABLE 5: Mean angular errors of our DSPS-DE with K times shadow masking on the DiLiGenT dataset. (The first and second row correspond to the second and fourth row of the top table in Table 4.)

K	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	AVG.
0	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
1	1.3	5.3	11.7	5.9	7.3	10.2	16.7	6.6	7.0	14.0	8.6
3	1.3	5.2	11.7	5.9	7.3	10.2	16.7	6.6	7.0	14.0	8.6

Overall, we observe our DSPS shows comparable or better accuracies compared to the existing exemplar-based methods. For convex shapes, where the global illumination effects can be mostly negligible, the accuracy by our method can further be better than the learning-based methods; this tendency is especially pronounced when few lights (*e.g.*, 10 lights). Even for non-convex shapes, the accuracy of our DSPS-D is drastically improved by introducing shadow masking without degrading the accuracy on convex shapes.

5.5 Analysis of shadow masking

As shown in Table 4, the shadow masking improves the accuracy of our DSPS-D on non-convex objects. This section shows the details of this improvement and discuss the effective value of K for shadow masking.

Figure 7 shows the difference in the angular error maps between our DSPS-DE with and without the shadow masking. Blue color indicates pixels where surface normal estimation is improved by the shadow masking, while the red color indicates the pixels that are degraded. This figure shows that the shadow masking improves the surface normal estimates at regions, where cast shadows are likely observed. As seen in the figure, the shadow masking does not much affect the surface normal estimates at convex regions.

Table 5 shows mean angular errors of our DSPS-DE with K times shadow masking on the DiLiGenT dataset. The results indicate that a greater K only gives the slight improvement. Considering that the K times shadow masking leads to a $K + 1$ times larger set of sampled appearance vectors, we conclude that a single time shadow masking has the best balance between accuracy and efficiency.

5.6 Robustness to image corruptions

We examine the robustness of our methods and baseline methods against common corruptions of photometric stereo images, camera noise, ambient light, and saturation. We prepared evaluation datasets by applying such corruptions to the MERL sphere dataset with 100 lights. See the supplementary material for details of simulating the corruptions.

Table 6 shows mean angular errors and standard deviations for each corrupted data. The results suggest that exemplar-based methods including ours and HS17 are robust to uniform and small perturbations of measurements (*i.e.*, camera noise and ambient light) compared to learning-based and model-based methods. For partial and relatively large corruption (*i.e.*, saturation), every method is generally robust. In particular, ST14 is almost unaffected by saturation since they eliminate large measurement values as outliers.

The robustness of exemplar-based methods can be explained by interpreting exemplar-based approaches as space partitioning along the surface normal candidates: They can be considered as separating the whole L'-dimensional measurement vector space to N subspaces, each of which corresponds to one of the surface normal candidates. Here, each subspace has a spatial margin to its neighboring subspaces, which yields robustness to measurement perturbations caused by corruptions.

5.7 Effect of varying number of BRDFs in the appearance tensor to DSPS-D

The experimental results so far show that our DSPS-D is consistently comparable or better than DSPS-C in terms of efficiency and accuracy. However, it is of interest to see

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

TABLE 6: Mean angular errors and standard deviations (mean angular error/standard deviation) on the corrupted MERL sphere datasets with 100 lights. Numbers are in degrees obtained from 100 MERL spheres.

	No noise	Camera noise	Ambient light	Saturation
DSPS-C	1.8/2.7	2.7/5.2	7.7/16.1	2.6/3.2
DSPS-DE	1.7/2.4	2.4/4.7	7.7/15.9	2.8/4.1
DSPS-DA	2.0/2.5	2.8/4.8	7.9/15.6	3.0/4.1
CNN-PS	2.5/3.8	6.2/14.0	8.9/18.8	2.6/3.8
HS17	1.6/2.4	2.4/4.7	7.7/15.9	2.6/3.8
ST14	1.2/2.3	22.9/13.3	34.5/33.7	1.2/2.3

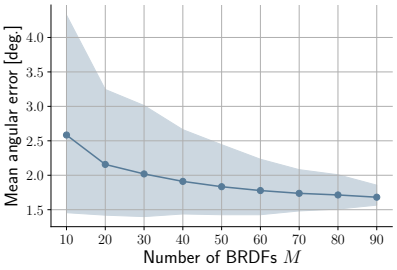


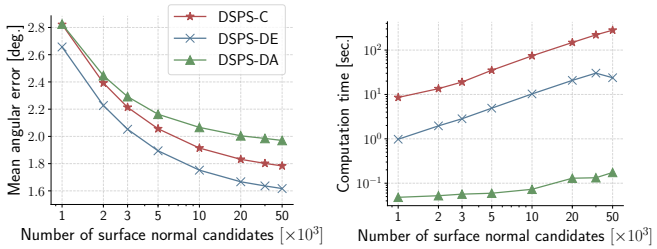
Fig. 8: Relationship between the accuracy of surface normal estimation and the number of BRDFs in the appearance tensor in DSPS-D. The solid line shows the mean angular error of the ten trials, and the colored area shows the maximum and minimum angular errors of the trials.

how the accuracy of DSPS-D varies when the number of BRDFs of the appearance tensor is limited since DSPS-D treats BRDFs in a discrete manner. Therefore, we validate this using the MERL sphere dataset with 100 lights. For each BRDF of the test data, we randomly sample BRDFs from the remaining 99 MERL BRDFs, run DSPS-D, and repeat them ten times for obtaining the average accuracy.

Figure 8 shows the relationship between the accuracy of surface normal estimation and the number of BRDFs in the appearance tensor. Naturally, the angular error of estimated surface normals becomes smaller as the number of BRDFs increases. The result suggests that 20 BRDFs or more give promising surface normal estimation, around 2° in average, around 3° at worst. The reason why DSPS-D with such small number of BRDFs successfully works is that the Eq. (5) only needs to be approximately satisfied for a good surface normal estimation, and that is sufficient as long as the nearest exemplar has a surface normal close to the true one.

5.8 Surface normal discretization

Figures 9a and b show mean angular error and computation time for varying numbers of surface normal candidates on the MERL sphere dataset with 100 lights. Throughout the paper we chose 20001 surface normal candidates because it balances accuracy and computation time well. For accurate surface normal estimation, 20001 or denser surface normal candidates are recommended. However, the choice of surface normal candidate discretization coarseness depends on the use case and a coarser discretization may be acceptable when fast inference is required.



(a) (b)

Fig. 9: (a) Mean angular errors and (b) Computation time of our methods with varying number of surface normal candidates. This experiment is performed on the MERL sphere dataset with 100 lights.

TABLE 7: Increases of angular errors due to discretized lights. The numbers represent the increase of mean angular error in degrees on the MERL sphere dataset.

	Number of lights									
	10	20	30	40	50	60	70	80	90	100
DSPS-C	0.02	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
DSPS-DE	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
DSPS-DA	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

5.9 Light direction discretization

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

As pre-defined light directions in the appearance tensor, we used 20001 discretized directions created in the same manner with the surface normal candidates. When a set of known light directions is given, we can slice out a sampled appearance matrix/vector for a hypothesized surface normal (and BRDF) 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 such an experiment on the MERL sphere dataset with ten types of light configurations.

Table 7 shows the increases of angular errors due to discretized lights on the MERL sphere dataset. We observe that the increases are generally small ($< 0.1°$), which suggests that it is acceptable to precompute an appearance tensor for sufficiently finely discretized light directions and there is no need to calculate a new appearance tensor for each lighting setup.

5.10 Precomputation cost

Our DSPS-C and DSPS-D pay costs on precomputation for each light configuration to enable efficient surface normal estimation. This section investigates the costs of precomputation for the CPUs and GPUs used in Sec. 5.3.

Figure 10 shows the precomputation times of our methods on a CPU and GPU for varying light configurations. This result shows that our methods only require tens of seconds or less. We consider that this cost that is only paid

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

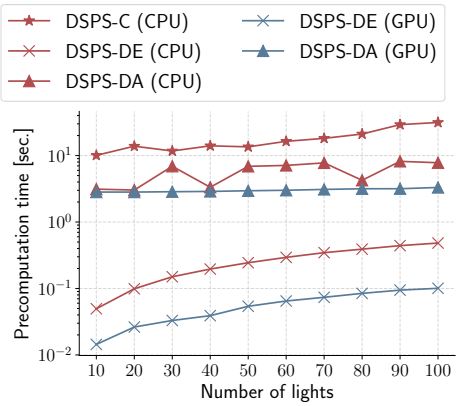


Fig. 10: Precomputation time of our three methods on a CPU and GPU for varying light configurations.

once for each light configuration is worth paying for the efficient inference shown in Figures 5a and b.

6 Conclusion

In this paper, we have presented Discrete Search Photometric Stereo (DSPS), where the photometric stereo problem is cast as a discrete search problem over a set of finely discretized surface normals. DSPS can stably recover surface normals of a scene with spatially varying general BRDFs in various light configurations. By putting most of the computation into a precomputation step, we enabled 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.

Experiments on synthetic and real-world datasets showed that our DSPS has comparable or better accuracy to state-of-the-art methods, particularly on convex surfaces and in few lights case, while achieving 10–100× acceleration from existing exemplar-based method. The precomputation, which significantly contributes to the acceleration, is also efficient and even when using the appearance tensor with pre-defined light directions that slightly deviate from the actual light directions, this incurs only negligible errors. In addition, we experimentally observed that our DSPS is robust to image corruptions compared to model-based and learning-based methods. Thus, our DSPS is one of the best choices for scenes with mostly convex surfaces and/or few light sources. In addition, because of its robustness against various types of image corruptions, it is recommended when the observed images are anticipated to contain image corruptions.

The approach's validity 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 full search strategies is expanding. We are interested in seeing more applications along this direction.

Acknowledgments

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] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.
[2] N. Alldrin, T. Zickler, and D. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
[3] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "Bi-polynomial modeling of low-frequency reflectances," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 36, no. 6, pp. 1078–1091, 2014.
[4] S. Ikehata and K. Aizawa, "Photometric stereo using constrained bivariate regression for general isotropic surfaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
[5] L. Chen, Y. Zheng, B. Shi, A. Subpa-Asa, and I. Sato, "A microfacet-based reflectance model for photometric stereo with highly specular surfaces," in *International Conference on Computer Vision (ICCV)*, 2017.
[6] A. Hertzmann and S. M. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying BRDFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 8, pp. 1254–1264, 2005.
[7] Z. Hui and A. C. Sankaranarayanan, "Shape and spatially-varying reflectance estimation from virtual exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 39, no. 10, pp. 2060–2073, 2017.
[8] M. Ren, X. Wang, G. Xiao, M. Chen, and L. Fu, "Fast defect inspection based on data-driven photometric stereo," *Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 1148–1156, 2018.
[9] K. Enomoto, M. Waechter, K. N. Kutulakos, and Y. Matsushita, "Photometric stereo via discrete hypothesis-and-test search," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
[10] S. Tozza, R. Mecca, M. Duocastella, and A. Del Bue, "Direct differential photometric stereo shape recovery of diffuse and specular surfaces," *Journal of Mathematical Imaging and Vision*, vol. 56, no. 1, pp. 57–76, 2016.
[11] A. S. Georghiades, "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo." in *International Conference on Computer Vision (ICCV)*, 2003.
[12] H.-S. Chung and J. Jia, "Efficient photometric stereo on glossy surfaces with wide specular lobes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.
[13] T. Chen, M. Goesele, and H.-P. Seidel, "Mesostructure from specularity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.
[14] S.-K. Yeung, T.-P. Wu, C.-K. Tang, T. F. Chan, and S. J. Osher, "Normal estimation of a transparent object using a video," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 37, no. 4, pp. 890–897, 2014.
[15] B. K. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts," *Scientific American*, vol. 251, no. 2, pp. 100–113, 1984.
[16] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, pp. 759–769, 2003.
[17] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita, "Deep photometric stereo networks for determining surface normal and reflectances," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2020.
[18] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong, "Deep photometric stereo for non-Lambertian surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2020.
[19] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.
[20] O. Wiles and A. Zisserman, "SilNet: Single-and multi-view reconstruction by learning from silhouettes," *British Machine Vision Conference (BMVC)*, 2017.
[21] S. Ikehata, "CNN-PS: CNN-based photometric stereo for general non-convex surfaces," in *European Conference on Computer Vision (ECCV)*, 2018.

This is the authors' version of the work. It is posted here by permission of IEEE for personal use. Not for redistribution.
The final authenticated publication is available at DOI: 10.1109/TPAMI.2022.3198729

[22] B. Burley, "Physically-based shading at Disney," in *SIGGRAPH*, 2012.
[23] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla, "PX-NET: Simple and efficient pixel-wise training of photometric stereo networks," in *International Conference on Computer Vision (ICCV)*, 2021.
[24] J. Li, A. Robles-Kelly, S. You, and Y. Matsushita, "Learning to minify photometric stereo," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
[25] Q. Zheng, Y. Jia, B. Shi, X. Jiang, L.-Y. Duan, and A. C. Kot, "SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks," in *International Conference on Computer Vision (ICCV)*, 2019.
[26] Z. Yao, K. Li, Y. Fu, H. Hu, and B. Shi, "GPS-Net: Graph-based photometric stereo network," *Adv. Neural Inform. Process. Syst.*, 2020.
[27] X. Wang, Z. Jian, and M. Ren, "Non-lambertian photometric stereo network based on inverse reflectance model with collocated light," *IEEE Trans. Image Process.*, vol. 29, pp. 6032–6042, 2020.
[28] T. Taniai and T. Maehara, "Neural inverse rendering for general reflectance photometric stereo," in *International Conference on Machine Learning (ICML)*, 2018.
[29] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.
[30] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.
[31] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
[32] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," in *International Conference on Computer Vision (ICCV)*, 2019.
[33] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.
[34] S. M. Omohundro, *Five balltree construction algorithms*. International Computer Science Institute Berkeley, 1989.
[35] A. Beygelzimer, S. Kakade, and J. Langford, "Cover trees for nearest neighbor," in *International Conference on Machine Learning (ICML)*, 2006.
[36] X. Wang, "A fast exact k-nearest neighbors algorithm for high dimensional search using k-means clustering and triangle inequality," in *International Joint Conference on Neural Networks*, 2011.
[37] Y. Hwang, B. Han, and H.-K. Ahn, "A fast nearest neighbor search algorithm by nonlinear embedding," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
[38] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.
[39] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *arXiv preprint arXiv:1702.08734*, 2017.
[40] A. Andoni, P. Indyk, T. Laarhoven, I. Razenshteyn, and L. Schmidt, "Practical and optimal LSH for angular distance," *arXiv preprint arXiv:1509.02897*, 2015.
[41] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 1, pp. 117–128, 2010.
[42] Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 42, no. 4, pp. 824–836, 2020.
[43] B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 41, no. 2, pp. 271–284, 2019.

Kenji Enomoto is a Ph.D. student at Osaka University, Japan. He received his B.S. and M.S. from Nagoya University in 2017 and 2019, respectively. His primary research interests span physics-based vision and machine learning.

Michael Waechter received his M.S. and Ph.D. from the Technical University of Darmstadt, Germany, in 2011 and 2017, respectively. He was a JSPS Postdoctoral Fellow at Osaka University, Japan from 2017 to 2019. His research interests include most areas related to 3D reconstruction.

Fumio Okura received the M.S. and Ph.D. degrees in engineering from the Nara Institute of Science and Technology, in 2011 and 2014, respectively. He has been an Assistant Professor with the Institute of Scientific and Industrial Research, Osaka University, until 2020. He is now an Associate Professor with the Graduate School of Information Science and Technology, Osaka University. His research interest includes the boundary domain between computer vision and computer graphics.

Kiriakos N. Kutulakos is a Professor of Computer Science at the University of Toronto. He received his PhD degree from the University of Wisconsin-Madison in 1994 and his BS degree from the University of Crete in 1988, both in Computer Science. Kyros has been a pioneer in the area of computational light transport, developing theoretical tools and computational cameras to analyze light propagation in real-world environments. He is the recipient of an Alfred P. Sloan Fellowship, a Marr Prize in 1999, a Marr Prize Honorable Mention in 2005 and five more paper awards (CVPR 2019, CVPR 2017, CVPR 2014, ECCV 2006 and CVPR 1994). He was Program Co-Chair of CVPR 2003 and ICCV 2013, and also served as Program Co-Chair of the Second International Conference on Computational Photography in 2010.

Yasuyuki Matsushita received his B.S., M.S. and Ph.D. degrees in EECS from the University of Tokyo in 1998, 2000, and 2003, respectively. From April 2003 to March 2015, he was with Visual Computing group at Microsoft Research Asia. In April 2015, he joined Osaka University as a professor. His research area includes computer vision, machine learning and optimization. He is an Editor-in-Chief of International Journal of Computer Vision (IJCV) and is/was on editorial board of IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), The Visual Computer journal, IPSJ Transactions on Computer Vision Applications (CVA), and Encyclopedia of Computer Vision. He served/is serving as a Program Co-Chair of PSIVT 2010, 3DIMPVT 2011, ACCV 2012, ICCV 2017, and a General Co-Chair for ACCV 2014 and ICCV 2021. He is a senior member of IEEE.