Efficient and Practical Exemplar-based Photometric Stereo

Submitted to
Graduate School of Information Science and Technology
Osaka University

July 2022

Kenji ENOMOTO

List of Publications

Journal

1. Kenji Enomoto, Michael Waechter, Fumio Okura, Kiriakos N. Kutulakos, and Yasuyuki Matsushita: "Discrete Search Photometric Stereo for Fast and Accurate Shape Estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (Minor Revision).

International conference

1. Kenji Enomoto, Michael Waechter, Kiriakos N. Kutulakos, and Yasuyuki Matsushita: "Photometric Stereo via Discrete Hypothesis-and-Test Search," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Jun. 2020).

2. Kenji Enomoto, Ken Sakurada, Weimin Wang, Masashi Matsuoka, Nobuo Kawaguchi, and Ryosuke Nakamura: "Image Translation Between SAR and Optical Imagery with Generative Adversarial Nets," In Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (Jul. 2018).

International workshop

1. Kenji Enomoto, Ken Sakurada, Weimin Wang, Hiroshi Fukui, Masashi Matsuoka, Ryosuke Nakamura, and Nobuo Kawaguchi: "Filmy Cloud Removal on Satellite Imagery with Multispectral Conditional Generative Adversarial Nets," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop (EARTHVISION) (Jul. 2017).

Domestic conference

1. 榎本 憲二，櫻田 健，王 維民，福井 宏，松岡 昌志，中村 良介，河口 信夫: "マルチスペクトル cGANs による衛星画像の薄雲除去," 画像の認識・理解シンポジウム（MIRU）(2017 年 8 月).
2. 櫻田 健，榎本 憲二，Nevrez Imamoglu，中村 良介，Lin Weisi，河口 信夫: "密なオプティカルフローに基づく畳み込みネットを用いたカメラ視点の違いに頑健なシーン変化検出," 画像の認識・理解シンポジウム（MIRU）(2017 年 8 月).

Acknowlegements

First and foremost, I would like to express my greatest gratitude to my supervisor, Prof. Yasuyuki Matsushita, for his continuous support, invaluable advice, and patience during my Ph.D. life. I personally respect him as a top researcher, an educationist, and a man, and learned a lot from his attitude and thoughts. I am truly proud and honored to be a student of Prof. Matsushita for the rest of my life.

I have been incredibly lucky to work with brilliant researchers, and would like to thank them for their invaluable advice and help: Prof. Kiriakos N. Kutulakos and Dr. Michael Waechter for my first and second projects (Chapters 2 and 3); Prof. Fumio Okura for my second and third projects (Chapters 3 and 4); Prof. Hiroaki Santo for my third project (Chapter 4).

I would also like to thank my thesis committee, Prof. Kaname Harumoto and Prof. Naoto Yanai. Their valuable feedback was indispensable to polish this thesis.

I would like to express my thanks to the members of the Computer Vision Laboratory at Osaka University, Dr. Heng Guo, Dr. Xu Cao, Mr. Feiran Li, Mr. Kensuke Uchida, Mr. Xinpeng Liu, Mr. Tatsuya Ishibashi, Mr. Daichi Iwata, Mr. Takao Mizuno, Mr. Hiroki Nakamura, Ms. Tomoka Takemura, Mr. Haruya Sakashita, Mr. Kota Nakamura, Mr. Itsuki Onishi, Mr. Yuki Konishiike, Ms. Lilika Makabe, Mr. Ryo Nakano, Mr. Kazuma Minami, Mr. Takashi Yamauchi, Mr. Yudai Matsuoka, Mr. Reo Izubuchi, Mr. Naohiro Shimada, Mr. Yuuki Takayama, Mr. Toshiaki Tanaka, Mr. Arashi Fukui, Ms. Chan Weng Ian, Mr. Zhuoyu Yang, Mr. Genma Kosaka, Mr. Naoki Asada, Ms. Mai Ido, Mr. Shota Ueno, Mr. Mutsuki Tomigashi, Mr. Takuto Narumoto, Mr. Kaito Fukui, Ms. Jie Tang, Mr. Kohei Ashida, Mr. Shota Ishii, Mr. Shota Ueno, Mr. Akira Uchida,

Mr. Yosuke Kii, Ms. Mone Tanaka, Mr. Ryohei Miyakawa, Mr. Junpei Watanabe, and Mr. Koki Fukai. It was a great pleasure to work together with so many warm-hearted people. I would also like to express my thank to the secretary of our lab, Ms. Mihoko Kaneda, for her support in my daily life at Osaka University.

Last, all my gratitude goes to my parents and family for their patience and unconditional love.

Abstract of thesis entitled

"Efficient and Practical Exemplar-based Photometric Stereo"

Submitted by

Kenji Enomoto

for the degree of Doctor of Philosophy

at Osaka University

in July, 2022

Throughout this dissertation, we consider high-fidelity 3D shape recovery from images, which is a fundamental problem in the computer vision field and required in various applications such as cultural heritage archives, film creation, and virtual reality. Photometric stereo is the most promising candidate for this purpose due to its ability of shape estimation in a per-pixel manner. It takes a set of images taken by a static camera under varying, known distant illuminations as input and recovers a scene's shape in the form of surface normal orientation by disentangling the interplay of a surface normal and reflectance in the image formation. Traditional photometric stereo assumes the Lambertian reflectance and convex surfaces, which deviate from real-world observations, thus introducing errors in surface normal estimates. This dissertation proposes photometric stereo methods for non-Lambertian, general reflectances and convex/non-convex surfaces with simple searching strategies that give a guarantee of reaching the globally optimal solution within the bound of an objective.

First, we address the photometric stereo problem for spatially varying, general reflectances. Unlike previous methods 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 reflectances in a reasonable amount of time.

Second, we propose the first nearest neighbor search-based photometric stereo, named Discrete Search Photometric Stereo (DSPS), for a scene with spatially varying, general reflectances. We show that the photometric stereo problem for general reflectances can be turned into a well-known nearest neighbor search problem over a set of *appearance exemplars*; a set of synthetic appearances generated from all possible pairs of finely discretized surface normals and reflectances. We demonstrate that the proposed method efficiently and accurately estimates both surface normals and reflectances, powered by advanced nearest neighbor search methods.

Third, we address the photometric stereo problem for a general scene with spatially varying, general reflectances and non-convex surfaces. Since the accuracy of our DSPS is determined by the coverage of the appearance exemplars, an augmentation of the appearance exemplars directly improves the surface normal estimation. We, therefore, introduce *general appearance exemplars* that take into account non-convex surfaces and more diverse reflectances than existing appearance exemplars. Our general appearance exemplars can be easily plugged into DSPS and improve the surface normal estimation accuracy, particularly in non-convex regions. Furthermore, our general appearance exemplars allow us to estimate a convexity (convex or non-convex) of a surface and incorporate benefits of different photometric stereo methods using the knowledge of the estimated convexity. We show that our DSPS with general appearance exemplars can accurately estimate surface normals on both convex and non-convex surfaces with diverse reflectances. We also demonstrate that incorporating different photometric stereo methods based on the estimated convexity provides more accurate surface normal estimates than either.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Computer vision techniques aim to derive meaningful information from visual data (images or videos) beyond a set of pixel intensities to understand a scene as humans do. Shape recovery from multiple images is a fundamental computer vision technique and plays a lot of roles in the real world. For example, shape recovery of a product such as a furniture improves the experience of online shopping by putting the product in which we want to place it virtually using augmented reality. In recent years, recovered shape and reflectance of an actor are used for the creation of more realistic films [9–11]. More recently, recovered city-scale shape and appearance are expected to be used for training of self-driving artificial intelligence [12–14]. In the field of agriculture, the recovered shape of crops is being used to analyze the condition of the crops, which enables harvesting at the best timing without human efforts [15, 16].

Shape recovery from multiple images can be roughly categorized into geometric and photometric approaches. The geometric approach first finds corresponding scene points across images captured from different viewpoints, then the 3D positions of the scene points can be estimated by performing triangulation. Representative methods of the geometric approach are stereoscopic photography [17], where the shape of a scene is recovered using two cameras like human eyes, and structure from motion [18, 19],

which assumes two or more viewpoints. The geometric approach works under natural illumination; therefore, it has been widely implemented in common devices such as smartphones. However, the geometric approach only estimates a coarse 3D shape due to the difficulty of correspondence matching, particularly when a target scene has smooth surfaces with less textures.

In contrast, the photometric approach achieves high-fidelity shape recovery. Scene appearances are caused by the interplays between shape (surface normal orientation), reflectance, and lighting. The photometric approach typically restricts lighting conditions and disentangles the interplays to estimate an object's shape. Photometric stereo is a representative method of the photometric approach, which estimates an object's shape in the form of surface normal orientation using dozens or hundreds of images captured from a static camera under known, varying lightings (Fig. 1.1). Since all images are captured from an identical view point, the correspondence matching is unnecessary for photometric stereo. Therefore, photometric stereo is able to produce per-pixel surface normals regardless of the smoothness and texture of a target scene. Lastly, a full shape is recovered by integrating the estimated surface normals.

An important challenge in photometric stereo is a stable surface normal estimation for general reflectances. Since photometric stereo disentangles the interplays between surface normal and reflectance, the accuracy of surface normal estimation depends on scene's reflectance property. Traditional photometric stereo methods [20, 21] assume the Lambertian reflectance; however, it is deviated from most reflectances in the real world, thus, introducing large errors in surface normal estimates. While recent methods use sophisticated reflectance models to handle non-Lambertian reflectances [22, 23], they are necessary to optimize surface normal and reflectance parameters simultaneously and generally encounter an issue of non-convex optimization.

Another challenge in photometric stereo is a surface normal estimation robust to global illumination effects. The global illumination effects such as cast shadows and inter-reflections cannot be described in a per-pixel manner and are difficult to be modeled for general scenes; therefore, they are ignored in most photometric stereo

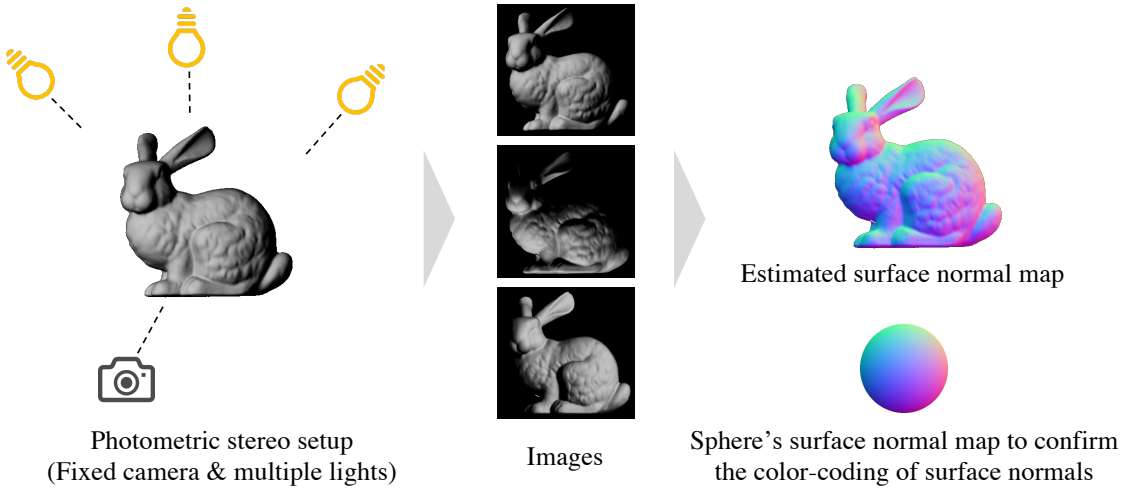


Fig. 1.1 Overview of photometric stereo. Given multiple images of an object taken from a static camera under known, varying lightings, photometric stereo recovers the shape of an object in the form of surface normals. A three dimensional surface normal is often visualized by RGB color coding. A sphere's surface normal map is attached to see the coding.

methods, although they appear in many objects in the real world. In recent years, learning-based methods [4, 24] achieve a robust surface normal estimation on this challenge; however, they interestingly degrade on surfaces without global illumination effects, where classical methods work well.

This dissertation addresses both challenges, general reflectances and global illuminations. The photometric stereo problem is typically formulated as a minimization problem of a loss function with surface normal and reflectance parameters. Previous methods treat the surface normal and reflectance parameters as continuous quantities and optimize them to minimize a loss. However, the loss function is often highly non-convex as shown in the left figure in Fig. 1.2; therefore, they are often trapped in local minima, leading to undesirable surface normal estimates. To overcome this issue, this dissertation proposes to treat surface normal and reflectance parameters as (1) discrete and continuous quantities, respectively, (2) both discrete quantities. In the first proposal, we present that a continuous optimization of reflectance parameters with a fixed, discretized surface normal becomes a well-posed problem; hence, we can

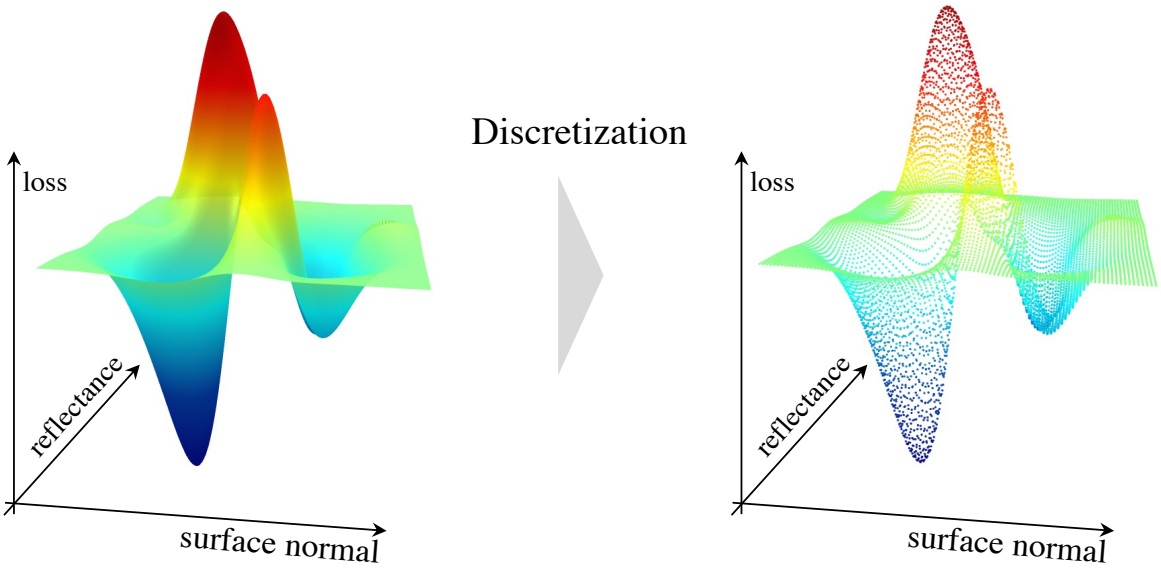


Fig. 1.2 A picture of our approach with a pseudo example. Previous methods that explore the optimal solution by performing a non-convex optimization over a continuous space of loss function (left), which is often trapped in local minima. In contrast, our approach first discretizes a space of loss function and then performs a discrete search over all discretized points (right). With this approach, our method can always find the globally optimal surface normal within the bound of discretized space.

find the globally optimal surface normal and reflectance parameters by performing a discrete search over all possible discretized surface normals. In the second proposal, we can naturally find the globally optimal surface normal and reflectance parameters by performing a discrete search over all possible surface normals and reflectances as shown in the right figure in Fig. 1.2. We present that this discrete search can be turned into the well-known nearest neighbor search problem; therefore, it can be performed in a highly efficient manner using advanced nearest neighbor search methods. While the discrete reflectance representation loses the expressions compared to the continuous one, our method with the discrete reflectance representation exhibits comparable accuracy with our method with the continuous one. Lastly, by extending the loss space considering global illumination effects (*i.e.*, cast shadows and inter-reflections), our method gains robustness to the global illumination effects while maintaining the accuracy on surfaces without global illumination effects.

1.2 Contributions

The main contributions of this dissertation can be summarized as follows:

- a search-based photometric stereo method for general reflectances. Instead of treating surface normals to be estimated as a continuous quantity, we finely discretize the space of surface normals and search for the best surface normal. To alleviate an issue of computing cost in a full search, we developed a precomputation method that performs expensive computations in a scene-independent manner prior to the inference for a new scene.

- the first nearest neighbor search-based photometric stereo method for general reflectances. By treating reflectances as a discrete quantity in addition to surface normals, we formulate the photometric stereo problem as a well known nearest neighbor search problem over a set of *appearance exemplars*; a set of synthetic appearances generated from all possible pairs of finely discretized surface normals and reflectances. Our method achieves the state-of-the-art accuracy on convex surfaces with diverse materials.

- a set of general appearance exemplars to broaden the applicability of our nearest neighbor search-based photometric stereo to more diverse reflectances and non-convex surfaces. We build a new set of appearance exemplars by extending existing ones that only consider a limited number of reflectances and convex shapes. Our general appearance exemplars improve the accuracy of surface normal estimation on general surfaces and allow us to estimate a convexity of a surface. The knowledge of estimated convexity also allows us to apply different photometric stereo methods to convex and non-convex surfaces, respectively, leading to further accuracy.

1.3 Chapter organization

This dissertation introduces two accurate and efficient search-based photometric stereo methods for general reflectances and one dataset to broaden the applicability of nearest neighbor search-based photometric stereo to more diverse reflectances and non-convex surfaces. The remainder of this dissertation is organized as follows.

Chapter 2 This chapter addresses the problem of estimating surface normals of a scene with spatially varying, general reflectances observed by a static camera under varying, known, distant illumination. In this chapter, we propose Hypothesis-and-Test Search Photometric Stereo (HaTS-PS). Unlike previous methods 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 reflectances in a reasonable amount of time.

Chapter 3 This chapter also addresses the photometric stereo problem for a scene with spatially varying, general reflectances. In this chapter, we propose Discrete Search Photometric Stereo (DSPS). While HaTS-PS employ a continuous reflectance model, DSPS treats reflectances as a discrete quantity as well as surface normals. Unlike previous methods that rely on continuous optimization over non-convex objective functions to estimate a shape and reflectance, the proposed method casts the problem as a discrete search over a set of *appearance exemplars*; a set of synthetic appearances generated from all possible pairs of finely discretized surface normals and reflectances. We show that the proposed discrete search approach leads to efficient and accurate estimation of surface normals and reflectances, powered by advanced nearest neighbor

search methods.

Chapter 4 This chapter addresses the photometric stereo problem for a general scene with spatially varying diverse reflectances and non-convex surfaces. Since the accuracy of our DSPS is determined by the coverage of the appearance exemplars, the augmentation of the appearance exemplars directly improves the surface normal estimation. In this chapter, we introduce general appearance exemplars that take into account non-convex surfaces and more diverse reflectances than existing appearance exemplars. Our general appearance exemplars can be easily plugged into DSPS and improve the surface normal estimation accuracy, particularly in non-convex regions. Furthermore, our general appearance exemplars allow us to estimate a convexity (convex or non-convex) of a surface and incorporate the benefits of different photometric stereo using the knowledge of the estimated convexity. We show that our DSPS with general appearance exemplars can accurately estimate surface normals on both convex and non-convex surfaces with diverse reflectances. We also demonstrate that incorporating different photometric stereo methods based on the estimated convexity provides more accurate surface normal estimates than either.

Chapter 5 This chapter concludes this dissertation by summarizing the proposed methods and dataset and discussing potential future research directions.

Chapter 2

Efficient Exemplar-based Photometric Stereo with Scene-independent Precomputation

2.1 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 [20, 21] 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 [25–29] used more sophisticated reflectance models 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 chapter, we cast surface normal estimation as a *discrete hypothesis-and-test search* problem; thus, we call our method Hypothesis-and-Test Search Photometric Stereo (HaTS-PS). Instead of treating surface normals to be estimated as a continuous quantity, our method finely discretizes the space of surface normals and finds the

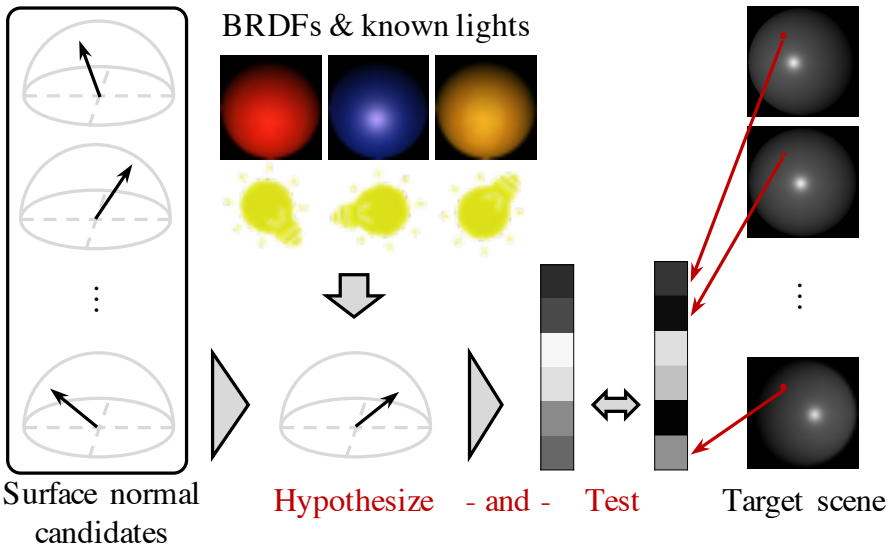


Fig. 2.1 An overview of our Hypothesis-and-Test Search Photometric Stereo (HaTS-PS) proposed in this chapter. We hypothesize a surface normal and test whether it can explain the target measurements. By conducting the hypothesis-and-test for all possible surface normals, our method is able to find a globally optimal surface normal.

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. HaTS-PS uses each surface normal candidate as *hypothesis* and *tests* its suitability to the measured intensities of a targe scene as illustrated in Fig. 2.1. In this manner, HaTS-PS 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 appearances, whose axes are the space of surface normals, light directions, and bidirectional reflectance distribution functions (BRDFs), for a fixed viewing direction. The table of appearances, which we call an appearance tensor, can contain an arbitrary number of BRDFs, and importantly, the

number of reference BRDFs considered in the appearance tensor does not influence the computation time during inference.

Our HaTS-PS is motivated by the success of example-based [30] and virtual exemplar-based [2] methods. The example-based method introduces a reference object having known shape and the identical material with a target object into a target scene. A surface normal is recovered by searching for appearances in the reference object that correspond to the target object's ones. In contrast, our method do not require placing a reference object in the scene. The virtual exemplar-based method and our HaTS-PS shares a basic strategy for surface normal estimation; however, the virtual exemplar-based method performs a continuous local search using a non-convex objective function to reduce their huge computation cost, which eliminates the guarantee of finding the optimal solution. In contrast, our precomputation enables an efficient exhaustive search, which allows us to find a globally optimal surface normal within the bounds of our objective function.

The chief contributions of this chapter 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.

2.2 Related work

Photometric stereo methods for diverse materials can be roughly divided into three categories; model-based, learning-based, and example-based approaches. In the fol-

lowing, we discuss the corresponding related works.

2.2.1 Model-based photometric stereo

A model-based approach uses parametric expressions for BRDFs, and the model parameters including the surface normal are estimated, typically, by optimizing them to well explain measured intensities of a target scene. Key for the model-based approach is the choice of a parametric BRDF model. Woodham's original work [20] 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 [22], the Torrance-Sparrow model [23], the Ward model [31], the specular spike model [32, 33], and a microfacet BRDF with ellipsoidal normal distributions [29] 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.* [27] use a bi-polynomial function and Ikehata and Aizawa [28] use a sum of lobes with unknown center directions. Although these model-based methods can be used in a relatively wide range of materials, there are always problematic materials, especially metallic materials are hard to be modeled by a simple function such as a bivariate function.

2.2.2 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 [4, 24, 34–36]. Santo *et al.* [34] proposed the first learning-based method to estimate a surface normal from a fixed number measured intensities under known lightings. Chen *et al.* [24] and Ikehata [4] introduced network architectures being applicable to arbitrary number of lightings, which inspire many follow-up

works [35, 37, 38]. Their networks are trained with synthetic datasets containing various shapes and materials since it is difficult to collect huge training dataset for the photometric stereo task. For example, Santo *et al.* [39] and Chen *et al.* [40] created their training datasets by rendering the Blobby [41] and Sculpture [42] shape datasets with 100 BRDFs from the MERL dataset [43]. Ikehata [4] also created a dataset by rendering fifteen objects with Disney's principled bidirectional scattering distribution function (BSDF) [5]. While the learning-based methods showed promising results on various scenes owing to the networks being trained with diverse shapes and materials, they surprisingly suffer from simple convex surfaces and diffuse materials that can be well fitted by traditional methods [38].

2.2.3 Example-based photometric stereo

Example-based photometric stereo relies on the concept of orientation-consistency [30], *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 [44]. In the example-based approach, 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 [30] 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 method 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 [2] introduced virtual exemplar-based method 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 [43] and assume that the target BRDF lies in the non-negative span of the MERL BRDFs. In the virtual exemplar-based method, 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 for all surface normal candidates in a reasonable amount of time owing to an efficient precomputation.

2.3 Scene-independent precomputation for exemplar-based photometric stereo

Starting from an image formation and problem statement, this section describes our HaTS-PS that casts the photometric stereo problem as a discrete search where the space of surface normals is discretized. We *hypothesize* a surface normal and *test* whether it satisfies the image formation model introduced in Sec. 2.3.1. By conducting this hypothesis-and-test for all possible surface normals, our method is able to find a globally optimal surface normal.

2.3.1 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

$$m \propto \rho(\mathbf{n}, \mathbf{l}) \max(\mathbf{n}^\top \mathbf{l}, 0), \tag{2.1}$$

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) and $\max(\mathbf{n}^\top \mathbf{l}, 0)$ is a function, which returns the largest value in inputs, representing a shadow caused when a surface normal is not facing a light source.

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. (2.1) can be written in matrix form as

$$\underbrace{\begin{pmatrix} m_1 \\ \vdots \\ m_{L'} \end{pmatrix}}_{\mathbf{m}} \propto \underbrace{\begin{pmatrix} \max(\mathbf{n}^\top \mathbf{l}_1, 0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \max(\mathbf{n}^\top \mathbf{l}_{L'}, 0) \end{pmatrix}}_{\mathbf{E}} \underbrace{\begin{pmatrix} \rho(\mathbf{n}, \mathbf{l}_1) \\ \vdots \\ \rho(\mathbf{n}, \mathbf{l}_{L'}) \end{pmatrix}}_{\boldsymbol{\rho}}, \tag{2.2}$$

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.* [30], and Hui and Sankaranarayanan [2]. By stacking M known BRDF basis vectors in a BRDF basis matrix $\mathbf{B}$, $\boldsymbol{\rho}$ can be written as

$$\boldsymbol{\rho} = \underbrace{\begin{pmatrix} \rho_1(\mathbf{n}, \mathbf{l}_1) & \dots & \rho_M(\mathbf{n}, \mathbf{l}_1) \\ \vdots & \ddots & \vdots \\ \rho_1(\mathbf{n}, \mathbf{l}_{L'}) & \dots & \rho_M(\mathbf{n}, \mathbf{l}_{L'}) \end{pmatrix}}_{\mathbf{B}} \mathbf{c}, \tag{2.3}$$

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

$$\mathbf{m} = \mathbf{EBc} \stackrel{\text{def}}{=} \mathbf{Dc}, \tag{2.4}$$

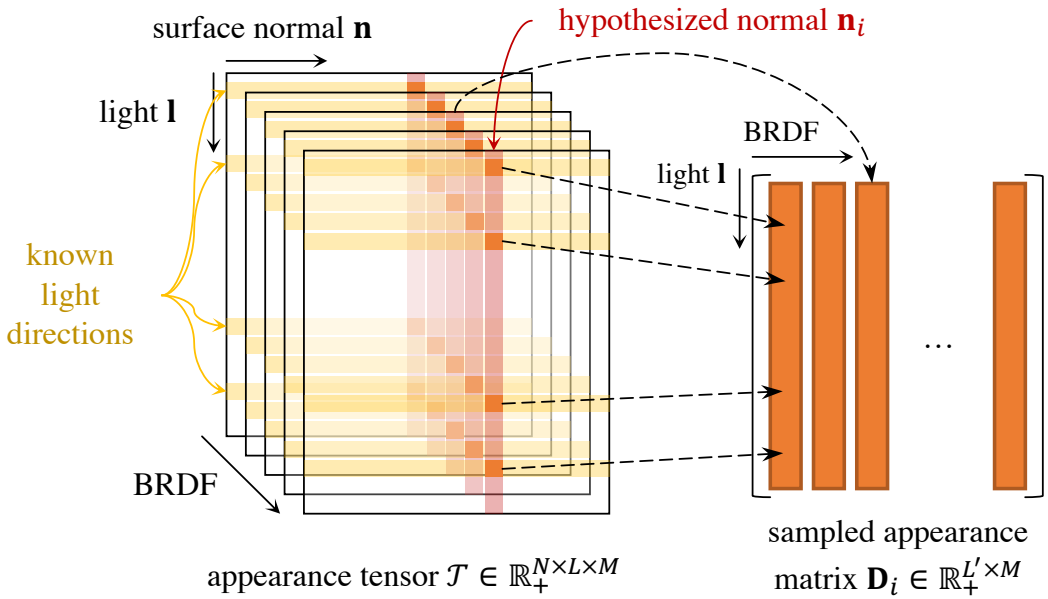


Fig. 2.2 Starting from the appearance tensor $\mathcal{T}$ that represents appearances for a comprehensive set of light directions, surface normals, and BRDFs, we slice out a sampled appearance matrix $\mathbf{D}_i$ for a set of known light directions and a hypothesized surface normal $\mathbf{n}_i$. 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.

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

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. (2.4).

2.3.2 Hypothesis-and-test strategy

We tackle the problem stated above by hypothesis-and-test photometric stereo (HaTS-PS) that hypothesizes a surface normal, tests whether it satisfies Eq. (2.4), and repeats these steps for all possible surface normals to find the optimal surface normal. 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. 2.2).

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 of the observed scene, 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. 2.2. Using $\mathbf{D}_i$ instead of $\mathbf{D}$, Eq. (2.4) becomes

$$\mathbf{m} \simeq \mathbf{D}_i \mathbf{c}. \tag{2.5}$$

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

$$\mathbf{c}_i = \left(\mathbf{D}_i^{\top} \mathbf{D}_i\right)^{-1} \mathbf{D}_i^{\top} \mathbf{m} = \mathbf{D}_i^{\dagger} \mathbf{m}, \tag{2.6}$$

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

$$e_i = \left\| \mathbf{m} - \mathbf{D}_i \mathbf{c}_i \right\|_2^2 . \tag{2.7}$$

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

$$\mathbf{n}^* = \mathbf{n}_{i^*}, \quad i^* = \underset{i \in \{1,\ldots,N\}}{\operatorname{argmin}} \; e_i. \tag{2.8}$$

A naïve implementation may require a significant computational effort for solving this problem. We thus introduce an efficient scene-independent precomputation strategy in the next section.

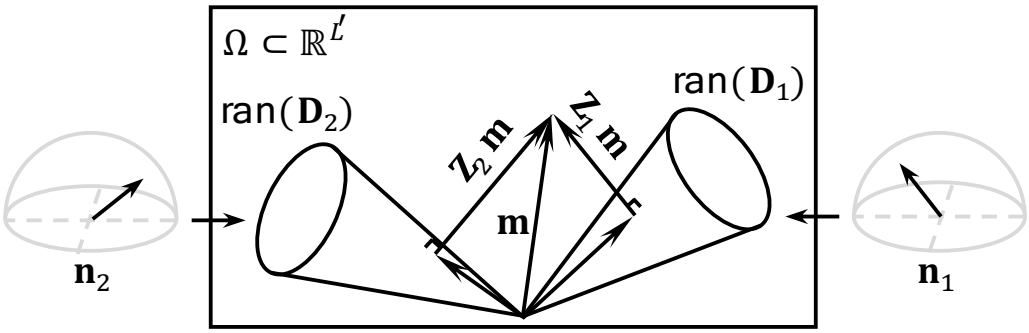


Fig. 2.3 Geometric interpretation of the measurement reconstruction error. The reconstruction error of measurements $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ can be seen as distance between the measurement vector $\mathbf{m}$ and the subspace spanned by $\mathbf{D}_i$ in the L'-dimensional space Ω.

2.3.3 Scene-independent precomputation

The reconstruction error e_i in Eq. (2.7) can be further simplified as

$$e_i = \|\mathbf{m} - \mathbf{D}_i\mathbf{c}_i\|_2^2 = \left\|\mathbf{m} - \mathbf{D}_i\mathbf{D}_i^\dagger\mathbf{m}\right\|_2^2 \tag{2.9}$$

$$= \left\|\left(\mathbf{I} - \mathbf{D}_i\mathbf{D}_i^\dagger\right)\mathbf{m}\right\|_2^2 \stackrel{\text{def}}{=} \|\mathbf{Z}_i\mathbf{m}\|_2^2 . \tag{2.10}$$

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 surface normal hypothesis $\mathbf{n}_i$. We, thus, can precompute a set of $\{\mathbf{Z}_i\}$ for all surface 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.

2.3.4 Dimensionality reduction of sampled appearance matrix

Equation (2.8) is only a necessary condition for finding 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. 2.3, 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 $\mathbf{n}_i$.

To avoid this, we shrink the subspace spanned by each $\mathbf{D}_i$ by reducing the rank of $\mathbf{D}_i$ to $M'(< L')$. Specifically, we replace $\mathbf{D}_i$ with its first M' 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

$$\mathbf{Z}_i = \mathbf{I} - \mathbf{U}_i'\mathbf{U}_i'^{\dagger} = \mathbf{I} - \mathbf{U}_i'\mathbf{U}_i'^{\top} \tag{2.11}$$

due to the orthogonality of each singular vector.

We empirically found that the proper value of M' is related to the noise level in the observations. In Sec. 2.4.3, we examine the accuracy of surface normal estimation with varying M' and discuss the choices for M'.

2.4 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 appearance tensor, the synthetic and real-world datasets, and baseline methods that we use for evaluation.

Appearence tensor: The appearance tensor is constructed from three components; BRDFs, surface normals, and light directions. For BRDFs, we used the MERL BRDF database [43] that consists of 100 distinct BRDFs including diffuse, specular, and

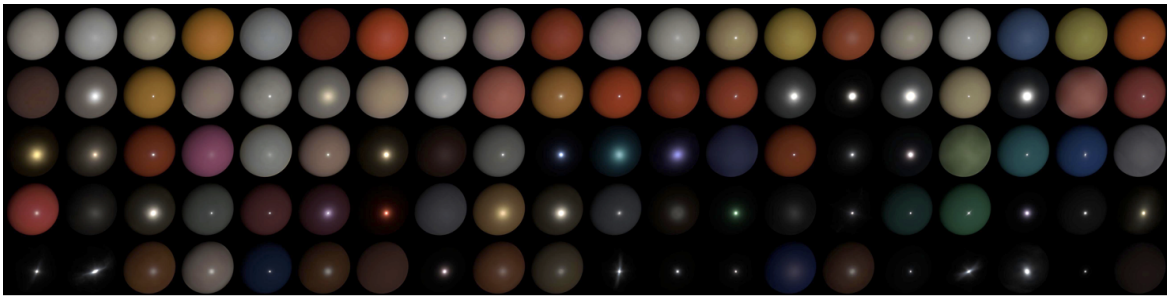


Fig. 2.4 Example images rendered with 100 MERL BRDFs. The MERL BRDFs consists of various materials, from soft diffuse to hard specular materials.

metallic materials as shown in Fig. 2.4. We discretized the surface normal on the unit hemisphere [1] and obtained 20001 surface normal candidates with nearly 0.5° intervals. In all experiments of this chapter, we assume that the appearance tensor contains the known light directions. In Sec. 2.4.5, we discuss how the surface normal estimation accuracy is affected by the discretization of light directions.

MERL sphere dataset: The MERL sphere dataset consists of 100 synthetic sphere scenes rendered with the 100 MERL BRDFs [43]. We rendered the images under ten lighting environments consisting of $L' = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ uniformly distributed light sources shown in Fig. 2.5. Image resolution was set to 100×100, yielding 7860 valid pixels. We also created a noisy MERL sphere dataset by adding signal-independent and signal-dependent noise [45] 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.

Real-world benchmark: We took an existing real-world dataset, the DiLiGenT dataset [3], which contains 10 real objects of general reflectance illuminated from 96 different known directions. Each object data has tens of thousands of valid pixels. This dataset provides ground truth surface normal maps for all objects measured by high-precision laser scanning, enabling a quantitative evaluation. For the BEAR object we discarded the first 20 images where a part of measurements is corrupted as pointed

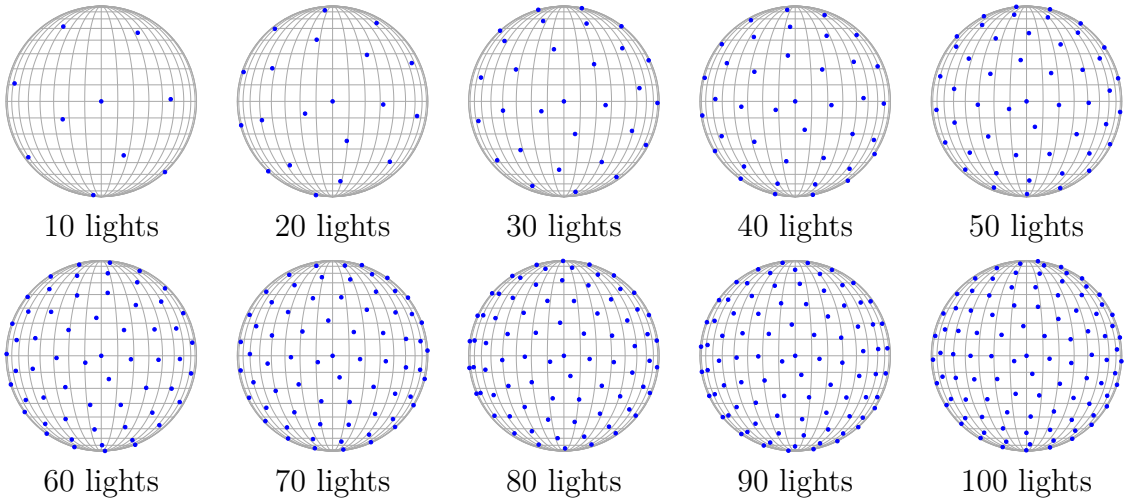


Fig. 2.5 Ten variants of light distributions for the MERL sphere dataset. These light distributions are generated by uniform or equi-angular sampling on the sphere [1].

out by Ikehata [4].

Baselines: As baselines we used Lambertian photometric stereo (LPS) [20], a model-based method ST14 [27], the virtual exemplar-based method HS17 [2], the unsupervised learning (*i.e.*, neural inverse rendering)-based method NIR-PS [46], the supervised learning methods PX-NET [35], PS-FCN^{+N} [24], WJ20 [47], CNN-PS [4], and SPLINE-Net [38]. 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 [48] 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. 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. 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.

2.4.1 Efficiency of surface normal estimation

For inference, our method evaluates the reconstruction error $\|\mathbf{Z}_i\mathbf{m}\|_2^2$ in Eq. (2.10) 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. The computation is highly parallelizable, *e.g.*, by pixel-wise or surface normal candidate-wise parallelization. Note that our method is executable on common CPUs because the matrices $\mathbf{Z}_i$ only require a small amount of memory. For example, matrices $\mathbf{Z}_i$ 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.

This experiment shows a comparison of computation time with the existing exemplar-based method. We use the MERL sphere dataset with the ten light sets. We measured the computation time of our method and the existing exemplar-based method HS17 [2] on an Intel® Xeon® Gold 6148 CPU @ 2.40 GHz with 40 cores. We performed pixel-wise parallelization. Figure 2.6 shows the computation time for a single pixel on the CPU, averaged over all MERL spheres for each light configuration (number and distribution shown in Fig. 2.5). Our method achieves 2–5 times faster surface normal estimation than HS17.

2.4.2 Accuracy of surface normal estimation

We estimated surface normals on synthetic and real datasets to confirm that our method works with diverse scenes. We evaluate the accuracy of surface normal estimation by "mean angular error" that is an average of angular errors of estimated surface normals over all pixels. The angular error is calculated by $\cos^{-1}\left(\mathbf{n}_{\mathrm{gt}}^\top \mathbf{n}_{\mathrm{est}}\right)$,

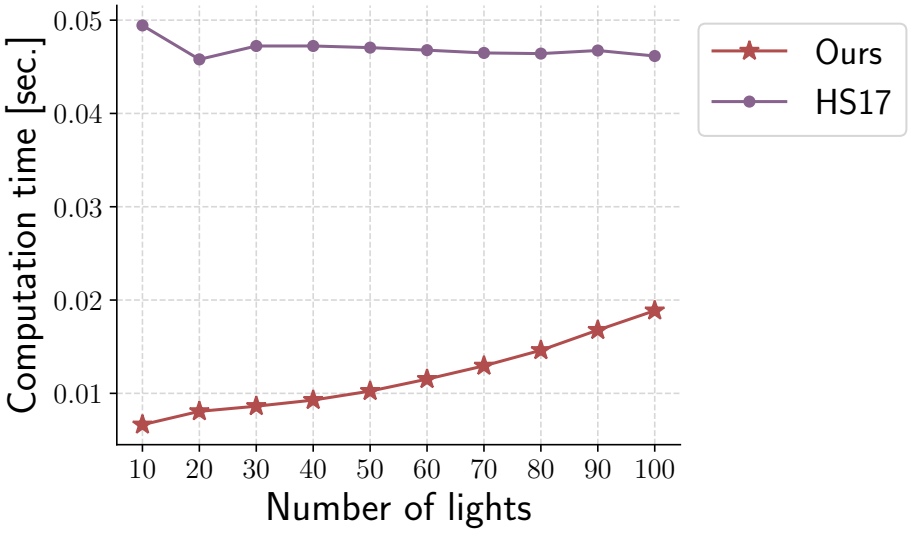


Fig. 2.6 Computation time of our HaTS-PS and HS17 [2] for a single pixel on a CPU. The experiments are performed on the MERL sphere dataset with light configuration 10 sets.

where $\mathbf{n}_{\text{gt}}$ and $\mathbf{n}_{\text{est}}$ are ground truth and estimated surface normals, respectively.

MERL sphere: We compared our method and the baseline methods using the MERL sphere dataset. Since there is no global illumination in the MERL sphere dataset, we can evaluate only the ability of our method to adapt to diverse materials. For the materials in our method 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 2.1 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 method stably yield small errors in all light configurations when compared with the baseline methods. While HS17 also achieves competitive accuracy, it is around 2–5 times slower than our method 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 our method and several baseline

Table 2.1 Comparisons on the MERL sphere dataset with ten light configuration 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	Ours	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
	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.

methods for each material on the MERL sphere datasets with 100 lights in Fig. 2.7.

DiLiGenT: We show quantitative results on the real-world dataset DiLiGenT in Tab. 2.2, where we compare our method with the baseline methods including very recent methods such as PX-NET and WJ20 in terms of mean angular error. Our method demonstrate comparable or better accuracy compared to the exemplar-based methods, although showing a degradation compared to the learning-based methods. This is considered to be due to factors not modeled in our method, namely cast shadows or inter-reflections.

Figures 2.8 and 2.9 show visual comparisons between our method and the baseline methods. 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 learning-based methods and estimates the surface normals well, *e.g.*, the BALL object or the body of BEAR and CAT objects.

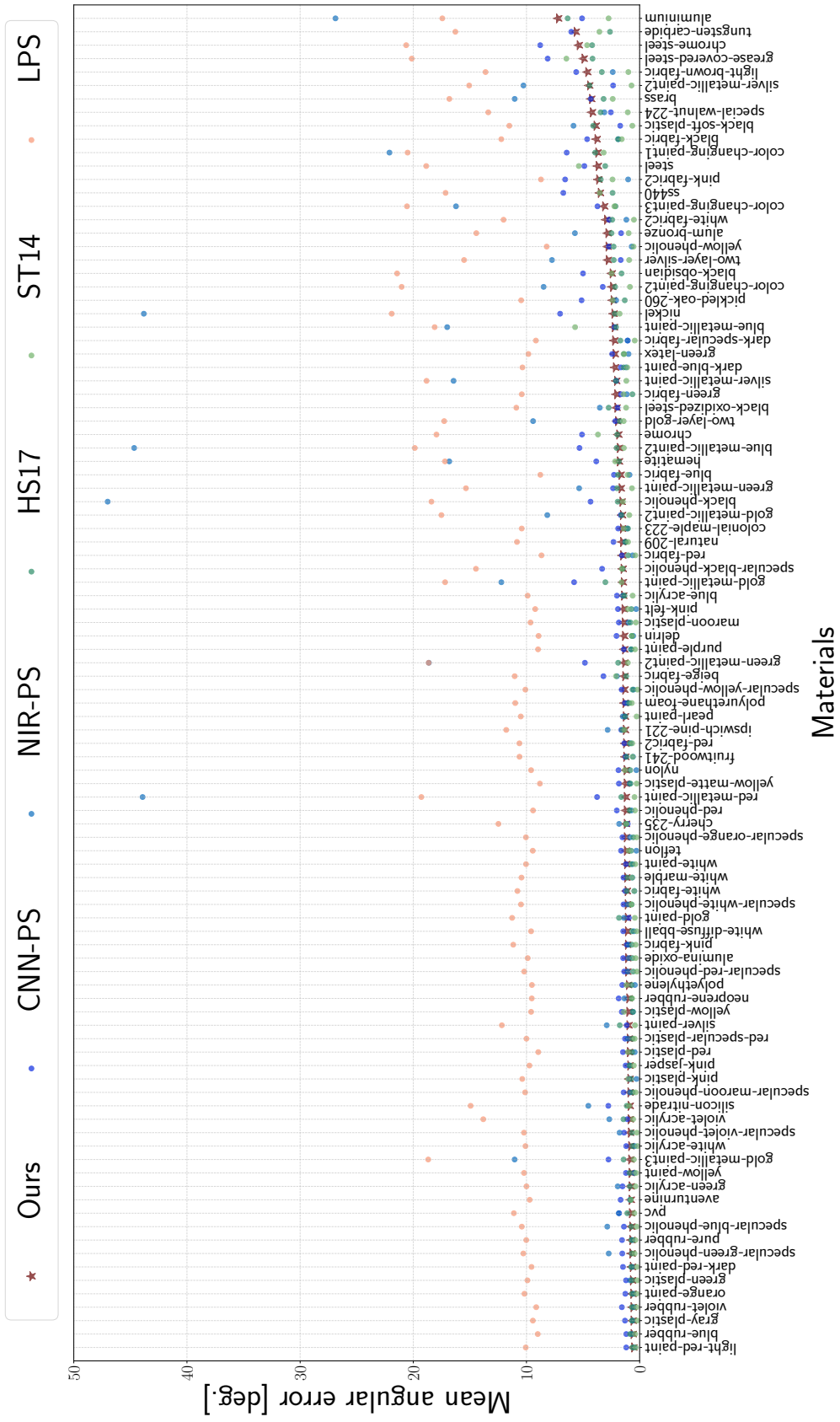


Fig. 2.7 Mean angular errors of our method and the baseline methods for each MERL shere data with 100 lights.

Table 2.2 Comparisons on the DiLiGenT dataset. Numbers in the table represent mean angular errors in degrees.

		BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
Exemplar-based	Ours	1.6	5.9	13.1	6.1	9.2	11.0	18.7	6.6	7.2	15.0	9.4
	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

2.4.3 Choice of dimension M' for noisy data

We empirically observed that M' is related to our method's robustness against noise. Thus, we determine an optimal M' by a validation using the noisy MERL sphere dataset.

We applied a leave-one-out scheme, testing it on one MERL BRDF while constructing the appearance tensor from the remaining 99 BRDFs. We test varying $M' = \{2, 3, 4, 5, 7, 10\}$ and varying noise $\mu/\lambda = \{5/30, 30/5, 30/30\}$ under five light configuration sets, *i.e.*, 20, 40, 60, 80, 100 lights.

Figure 2.10 shows mean angular errors of estimated surface normals in degrees. In most of configurations, $M' = 3$ produces the lowest angular errors among the candidates of M', indicating that $M' = 3$ is the most robust to noise. For this reason, we applied $M' = 3$ in all experiments of this chapter.

2.4.4 Surface normal discretization

Tables 2.3 and 2.4 show mean angular error and computation time for varying numbers of surface normal candidates on the MERL sphere dataset with 100 lights. Throughout

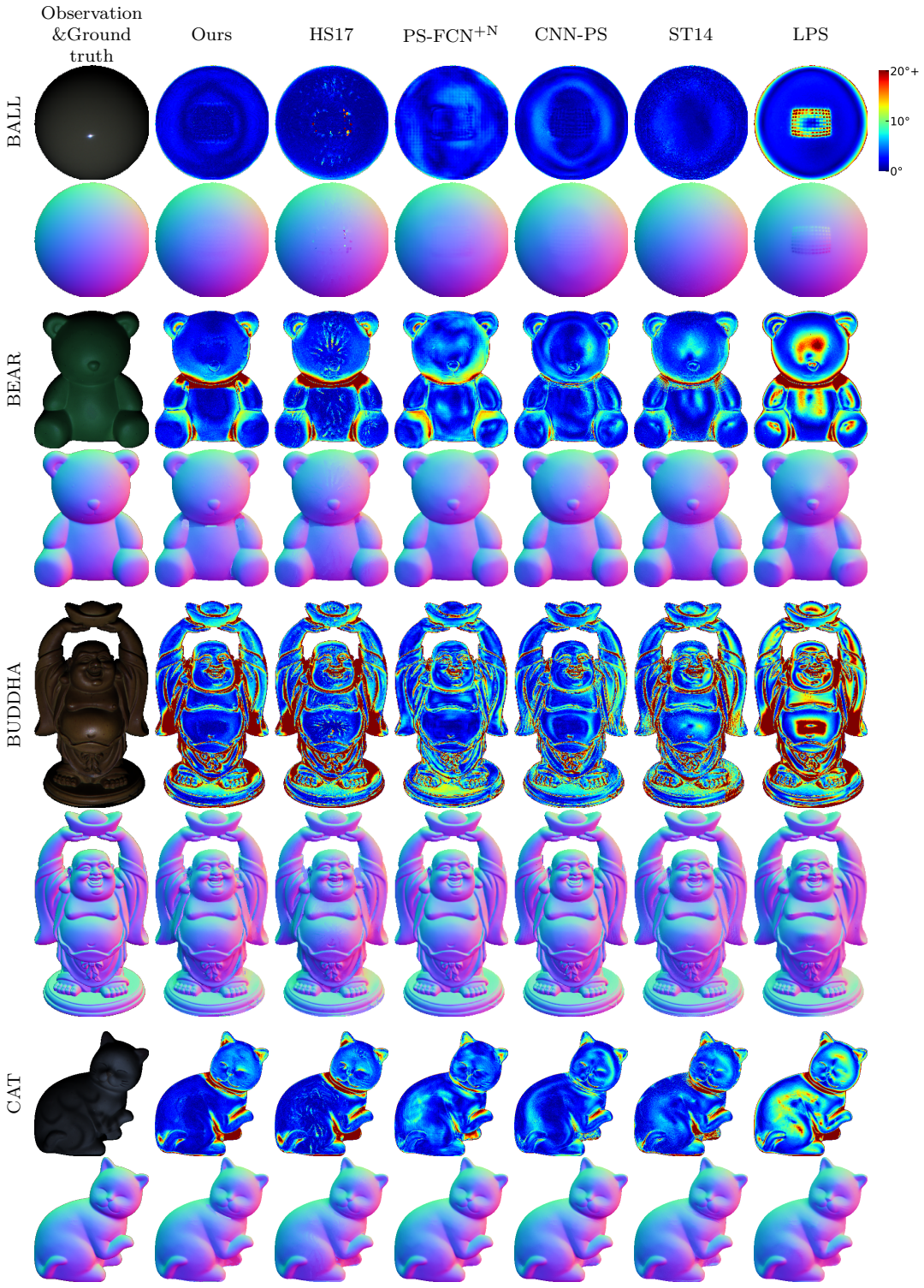


Fig. 2.8 Angular error maps and estimated surface normal maps for BALL, BEAR, BUDDHA, and CAT objects in the DiLiGenT dataset [3].

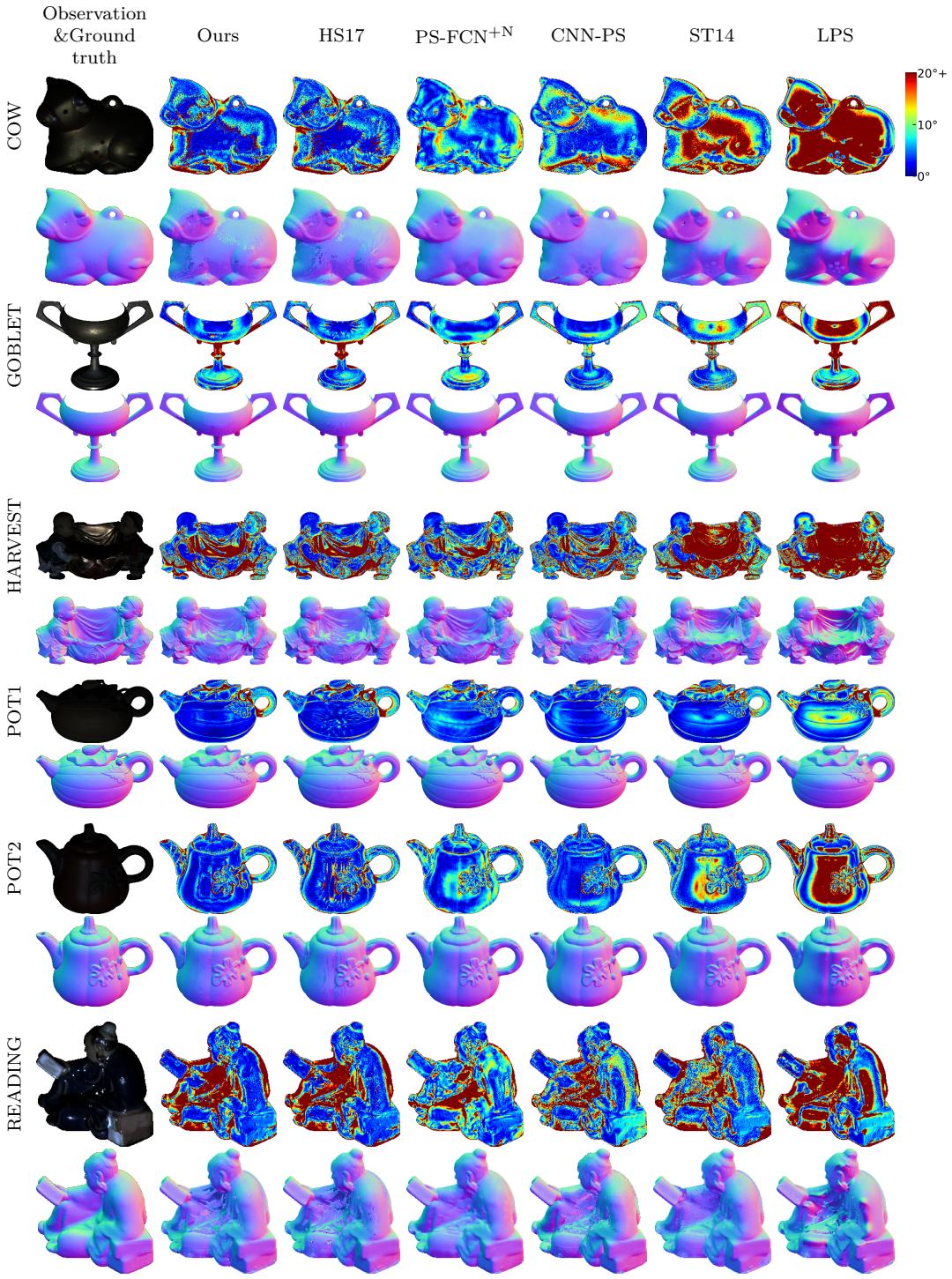


Fig. 2.9 Angular error maps and estimated surface normal maps for COW, GOBLET, HARVEST, POT1, POT2, and READING objects in the DiLiGenT dataset [3].

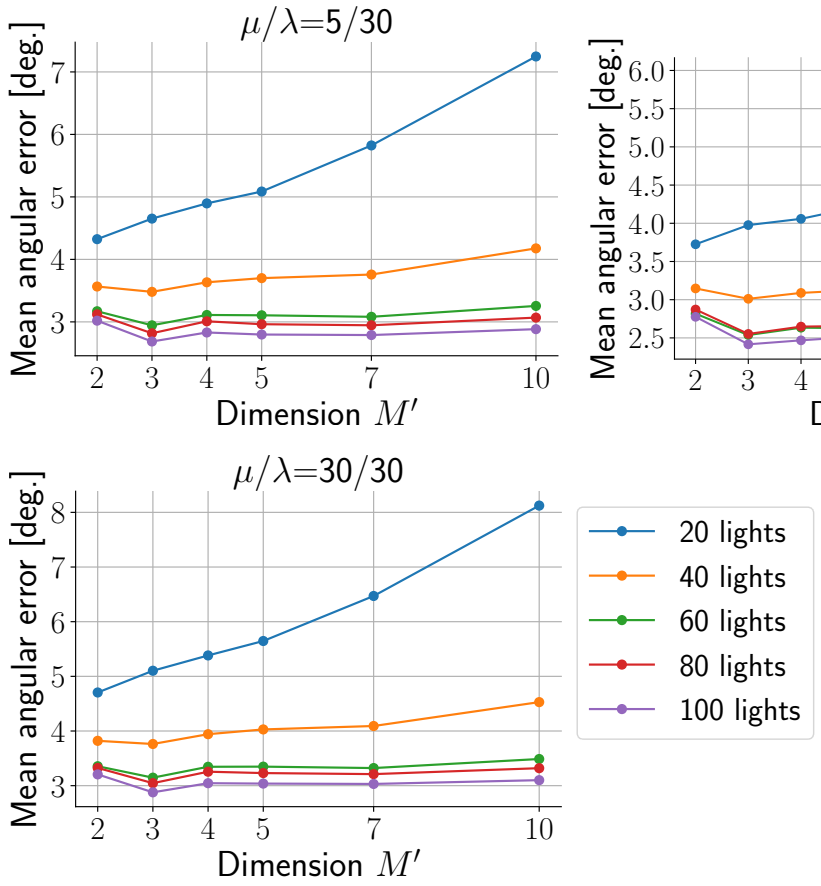


Fig. 2.10 Mean angular error of estimated surface normals with varying M' on noisy MERL sphere dataset under five light configuration sets. μ and λ are parameters for controlling the magnitude of signal-independent and signal-dependent noises.

this chapter 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.

2.4.5 Light direction discretization

In all experiments so far, we assumed that the appearance tensor $\mathcal{T}$ 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

Table 2.3 Mean angular errors for estimated surface normals in degrees for varying numbers of surface normal candidates. The experiment is performed on the MERL sphere dataset with 100 lights.

number of surface normal candidates N							
1001	2001	3001	5001	10001	20001	30001	40001
2.82	2.39	2.21	2.06	1.91	1.83	1.80	1.78

Table 2.4 Computation time of our method in milliseconds for varying numbers of surface normal candidates. The experiment is performed on the MERL sphere dataset with 100 lights, and the computation time is calculated by taking average over all MERL sphere's pixels.

number of surface normal candidates N							
1001	2001	3001	5001	10001	20001	30001	40001
1.09	1.72	2.42	4.46	9.43	18.9	28.0	35.7

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 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 2.5 shows the increases of mean angular errors (*i.e.*, ones shown in Tab. 2.1) due to the light discretization on the MERL sphere dataset. We observe that the increases are generally small ($< 0.1^\circ$), which suggests that it is acceptable to prepare an appearance tensor $\mathcal{T}$ for sufficiently finely discretized light directions and sample appearance matrices $\{\mathbf{D}_i\}$ for light directions closest to target scene's ones. Hence, there is no need to calculate appearance matrices for each light configuration.

Table 2.5 Increases of angular errors due to discretized lights. As pre-defined light directions in the appearance tensor we used 20001 directions created in the same way as the surface normal candidates. 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
0.02	0.00	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01

Table 2.6 Precomputation time in seconds for varying number of lights. These precomputation time is measured in a typical case of 20001 surface normal candidates, 100 BRDF bases, and 100 light directions.

number of lights									
10	20	30	40	50	60	70	80	90	100
10	14	12	14	14	16	18	21	29	31

2.4.6 Precomputation cost

Our method achieves efficient surface normal estimation by precomputation of $\mathbf{Z}_i \in \mathbb{R}^{L' \times L'}$ from an appearance matrix $\mathbf{D}_i \in \mathbb{R}^{L' \times M}$ that is performed only once for a light configuration. Table 2.6 shows the precomputation time of our method for varying number of lights. It shows that our method only requires tens of seconds for the precomputation. We consider that this precomputation cost is worth paying for the efficient surface normal estimation, especially when performing photometric stereo for multiple subjects under an identical light configuration.

2.5 Conclusion

In this chapter, we have presented a photometric stereo method based on discrete hypothesis-and-test search. The proposed method can work with a diverse set of appearances that are represented in an appearance 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. This 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.

Chapter 3

Nearest Neighbor Search-based Photometric Stereo

3.1 Introduction

Photometric stereo aims at recovering surface normals and bidirectional reflectance distribution functions (BRDFs) from image measurements taken by a static camera under varying and known distant lights. Today, Lambertian photometric stereo [20] is already well understood; however, non-Lambertian photometric stereo still remains a difficult problem, and there have been various approaches in the past. Recent search-based (a.k.a. exemplar-based) methods including our HaTS-PS presented in Chapter 2 estimate accurate surface normals on non-Lambertian surfaces at the cost of an exhaustive search over finely discretized surface normals. While HaTS-PS reduces the estimation time by a precomputation strategy, a faster estimation is required in several scenarios, such as on high-resolution images. Unlike existing search-based methods using continuous BRDF models, we treat BRDFs in a discrete manner as well as surface normal. It turns the photometric stereo problem into the well-known nearest neighbor search problem; hence the estimation time is dramatically saved using advanced nearest neighbor search methods. Although the discrete BRDF model only represents less diverse materials than continuous ones, surprisingly, our method exhibits com-

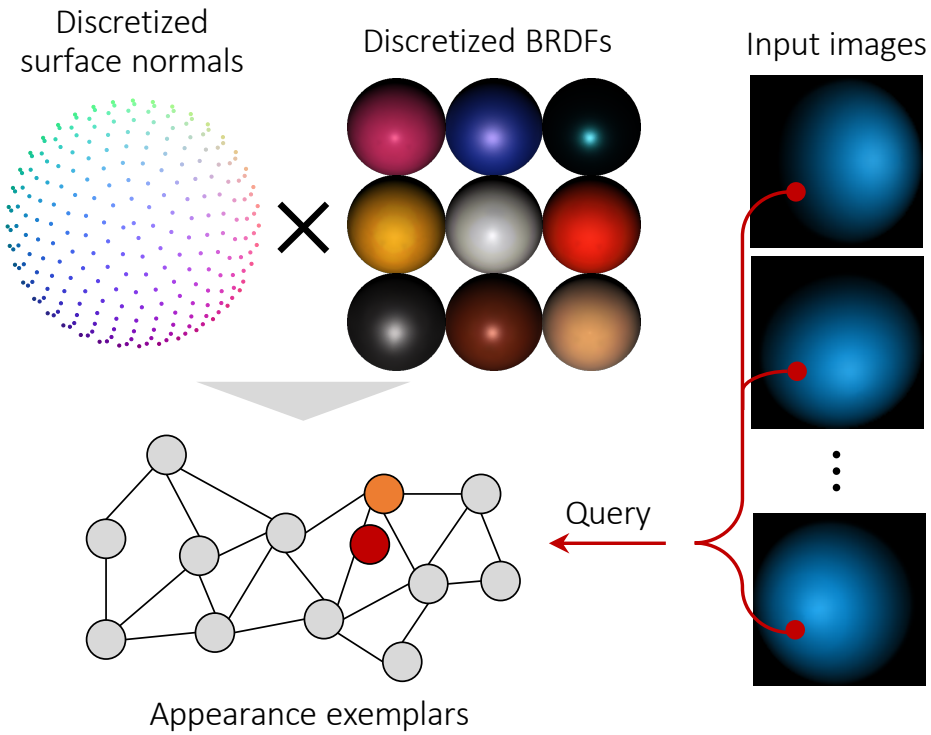


Fig. 3.1 An overview of ours DSPS proposed in this chapter. We estimate a surface normal and BRDF by a discrete search over the discretized space of surface normals and BRDFs. The problem can be solved by any nearest neighbor search method, which reduces an estimation cost dramatically.

parable or high accuracy in determining surface normals to exemplar-based methods with continuous BRDF models.

This chapter presents Discrete Search Photometric Stereo (DSPS), in which the non-Lambertian photometric stereo problem is turned into a discrete search over a finely-discretized space of surface normals and BRDFs. The discretized space is formed by *appearance exemplars*; a set of synthetic appearances corresponding to all possible pairs of discretized surface normals and BRDFs. Given known light directions and the associated image measurements, our method resamples the discretized space and performs a nearest neighbor search over the resampled space to determine surface normal and BRDF in a per-pixel manner as shown in Fig. 3.1. Similar to other search-based photometric stereo methods [2, 30, 49], DSPS is built upon the observation that appearance exemplars having similar surface normals and BRDFs are naturally

similar [30, 44]. Unlike other search-based methods, our DSPS fully discretizes continuous surface normals and BRDFs and performs a discrete search without relying on continuous optimization. This allows us to leverage many methods for efficient exact/approximate nearest neighbor search [50–53].

Naturally, the accuracy of DSPS depends on the granularity of the discretization of surface normals and BRDFs, and also on the number of BRDF samples contained in the space. Although our experiments show that DSPS already yields favorable accuracy for diverse materials, it has the potential of becoming even more powerful as processing power and BRDF datasets grow further.

In summary, the key features of our DSPS are:

Simplicity: Discrete search is conceptually simple and intuitive, and its behavior is well understood.

Efficiency: DSPS benefits from advances in fast nearest neighbor search algorithms.

Accuracy: Discrete search over the finely-discretized space leads to a stable and accurate estimation of both surface normals and BRDFs. Since DSPS operates in a per-pixel manner, it naturally handles spatially-varying BRDFs.

3.2 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, example-based, and learning-based methods. Here, we review the example-based and learning-based methods. See Sec. 2.2 for the model-based methods.

3.2.1 Example-based photometric stereo

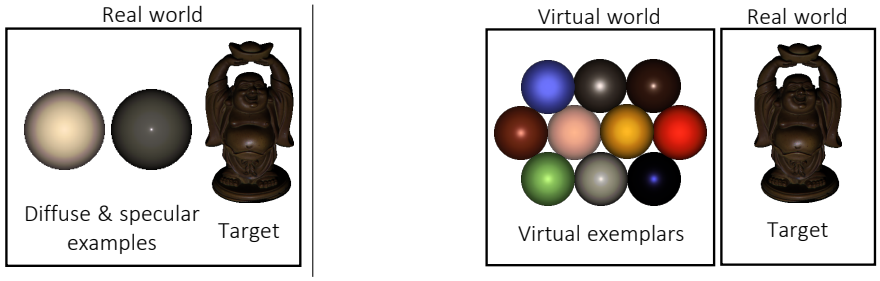
Early work on example-based photometric stereo relies on the concept of orientation consistency [30], *i.e.*, two surfaces with the same surface normal and BRDF will have the same appearance under the same illumination. Another work along this direc-

tion is found in Horn and Ikeuchi [44]. In these methods, 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 [30] 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 [2] introduced virtual exemplar-based photometric stereo that performs example-based photometric stereo without actually introducing reference objects in the target scene. They render virtual exemplars of appearances under the target scene illumination with the MERL BRDFs [43] 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 (or example-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. Moreover, unlike virtual exemplar-based method and hypothesis-and-test search photometric stereo (HaTS-PS) presented in Chapter 2 that treat BRDFs as a continuous quantity, our DSPS treats BRDFs in a discrete manner as well as surface normals, 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 HaTS-PS, are

Table 3.1 Comparison of exemplar-based photometric stereo methods and their properties.

	Hertzmann & Seitz [30]	Hui & Sankaranarayanan [2]	HaTS-PS (Chapter 2)	our DSPS
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				

summarized in Tab. 3.1.

3.2.2 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 [24, 28, 34–36]. 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, 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.* [34] and Chen *et al.* [24] created a training dataset by rendering the Blobby [41] and Sculpture [42] shape datasets with 100 MERL BRDFs [43]. Ikehata [4] also introduced a training dataset, called CyclesPS dataset, containing several objects rendered with a diverse set of materials from Disney's principled BSDFs [5] with global illumination effects. Logothetis *et al.* [35] 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 [37, 38, 54]. Wang *et al.* [47] 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 [46] proposed an unsupervised method that minimizes the reconstruction loss between input and re-rendered images.

Our DSPS, which uses nearest-neighbor search, can be considered to be a learning-based method 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 methods. Much like the growth in datasets in various machine learning tasks such as image classification [55–57], 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 [58, 59]. 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.

3.3 Discrete search photometric stereo

This section describes how the photometric stereo problem is turned into a nearest neighbor search problem. Starting from the image formation definition, we introduce the first nearest neighbor search-based photometric stereo.

3.3.1 Image formation

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

$$m \propto \rho(\mathbf{n}, \mathbf{l}) \max(\mathbf{n}^\top \mathbf{l}, 0), \tag{3.1}$$

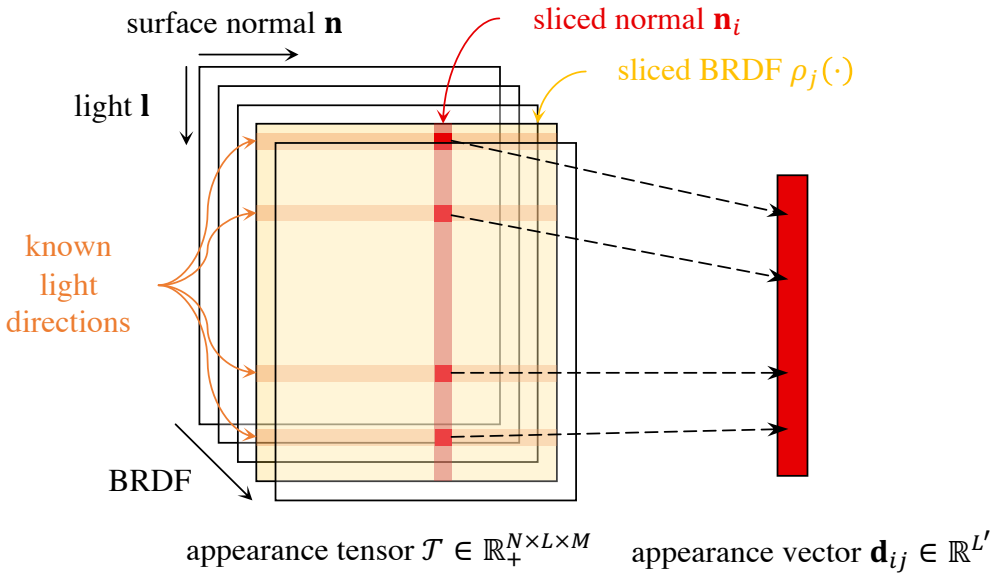


Fig. 3.2 Slice of the appearance tensor $\mathcal{T}$ that represents appearances for a comprehensive set of light directions, surface normals, and BRDFs. Given a set of known light directions, we can slice out all possible appearance vectors $\mathbf{d}_{ij}$.

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

In calibrated photometric stereo, a 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. (3.1) can be written in a vector form as

$$\underbrace{\begin{pmatrix} m_1 \\ \vdots \\ m_{L'} \end{pmatrix}}_{\mathbf{m}} \propto \underbrace{\begin{pmatrix} \rho(\mathbf{n}, \mathbf{l}_1) \max(\mathbf{n}^\top \mathbf{l}_1, 0) \\ \vdots \\ \rho(\mathbf{n}, \mathbf{l}_{L'}) \max(\mathbf{n}^\top \mathbf{l}_{L'}, 0) \end{pmatrix}}_{\mathbf{d}}, \tag{3.2}$$

where $\mathbf{m}$ is a measurement vector, $\mathbf{d}$ is an appearance vector with a fixed scale. Our goal is to find the optimal surface normal $\mathbf{n}$ and BRDF $\rho(\cdot)$ for each surface point, given measurements $\mathbf{m}$ and associated light directions $(\mathbf{l}_1, \dots \mathbf{l}_{L'})$ based on the model of Eq. (3.2).

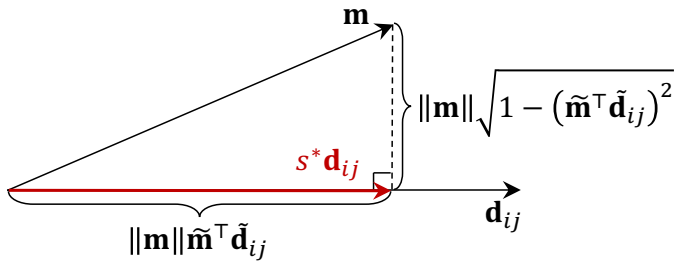


Fig. 3.3 Minimal ℓ_2 distance between $\mathbf{m}$ and $s\mathbf{d}_{ij}$. The optimal scaling parameter s^* scales the vector $\mathbf{d}_{ij}$ to the point closest to $\mathbf{m}$.

3.3.2 From photometric stereo to nearest neighbor search

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 BRDFs, which we call surface normal candidates and BRDF candidates, respectively. The surface normal candidates are generated by discretizing the angular direction over hemisphere, while the BRDF candidates are based on a set of measured BRDFs or discretizing an analytic BRDF model. We use the appearance tensor $\mathcal{T} \in \mathbb{R}_+^{N \times L \times M}$ to represent the appearance for all combinations of the surface normal candidates $\mathcal{N}$, BRDF candidates $\mathcal{B}$, and discretized L incoming light directions.

Given a certain combination of surface normal $\mathbf{n}_i$ and BRDF $\rho_j(\cdot)$ under a set of L' known light directions, we can slice out a synthetic appearance vector $\mathbf{d}_{ij}$ as illustrated in Fig. 3.2. The synthetic appearance vector $\mathbf{d}_{ij}$ can be obtained for all possible pairs of the surface normal candidates $\{\mathbf{n}_i\}$ and BRDF candidates $\{\rho_j\}$; thus, we can form a set of synthetic measurement vectors $\mathcal{A} = \{\mathbf{d}_{ij} \mid i = 1, \ldots, N; j = 1, \ldots, M\}$, which we call *appearance exemplars*.

If the set of appearance exemplars $\mathcal{A}$ is large enough, the actual measurement vector $\mathbf{m}$ from a scene point can be well approximated by an element of $\mathcal{A}$ as

$$\mathbf{m} \simeq s\mathbf{d}_{ij}, \tag{3.3}$$

where s is an unknown scaling in Eq. (3.2). Under this assumption, the optimal

indices of the surface normal candidate i^* and BRDF candidate j^* can be found as the minimizer of the following objective:

$$i^*, j^* = \underset{i,j}{\operatorname{argmin}} \|\mathbf{m} - s\mathbf{d}_{ij}\|_2^2 . \tag{3.4}$$

As illustrated in Fig. 3.3, the optimal scaling parameter s^* should scale the $\mathbf{d}_{ij}$ to the point closest to $\mathbf{m}$; therefore, the objective can be written in a parameter-free form by eliminating the (unknown) optimal scaling s^* as

$$\|\mathbf{m} - s^*\mathbf{d}_{ij}\|_2^2 = \|\mathbf{m}\|_2^2 \left(1 - \left(\tilde{\mathbf{m}}^\top \tilde{\mathbf{d}}_{ij}\right)^2\right), \tag{3.5}$$

where $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{d}}_{ij}$ are normalized $\mathbf{m}$ and $\mathbf{d}_{ij}$, respectively. Consequently, our objective is transformed to

$$\underset{i,j}{\operatorname{argmin}} \|\mathbf{m} - s\mathbf{d}_{ij}\|_2^2 \Leftrightarrow \underset{i,j}{\operatorname{argmax}} \tilde{\mathbf{m}}^\top \tilde{\mathbf{d}}_{ij}, \tag{3.6}$$

because $\|\mathbf{m}\|_2^2 = \text{const.}$, and $0 \leq \tilde{\mathbf{m}}^\top \tilde{\mathbf{d}}_{ij} \leq 1$ derived from the non-negativity of both vectors. Lastly, with the fact of $\left\|\tilde{\mathbf{m}} - \tilde{\mathbf{d}}_{ij}\right\|_2^2 = 2 - 2\tilde{\mathbf{m}}^\top \tilde{\mathbf{d}}_{ij}$, our objective becomes

$$\underset{i,j}{\operatorname{argmax}} \tilde{\mathbf{m}}^\top \tilde{\mathbf{d}}_{ij} \Leftrightarrow \underset{i,j}{\operatorname{argmin}} \left\|\tilde{\mathbf{m}} - \tilde{\mathbf{d}}_{ij}\right\|_2^2 . \tag{3.7}$$

Therefore, our final objective can be written concisely as

$$i^*, j^* = \underset{i,j}{\operatorname{argmin}} \left\|\tilde{\mathbf{m}} - \tilde{\mathbf{d}}_{ij}\right\|_2^2 . \tag{3.8}$$

This objective is equivalent to the nearest neighbor search problem with the Euclidean distance; hence, we can rely on any exact or approximate nearest neighbor search method to minimize it. This yields the optimal surface normal $\mathbf{n}^* = \mathbf{n}_{i^*}$ and BRDF $\rho^*(\cdot) = \rho_{j^*}(\cdot)$.

3.4 Experiments

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

3.4.1 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 [43] which consists of 100 distinct BRDFs including diffuse, specular, and metallic materials. We discretized the surface normal with equi-angular sampling from the unit hemisphere [1] and obtained 20001 surface normal candidates with nearly 0.5° intervals. In all experiments of this paper, we assume that the appearance tensor contains the known light directions. In Sec. 3.4.6, we discuss how the surface normal estimation accuracy is affected by the discretization of light directions.

MERL sphere dataset: The MERL sphere dataset consists of 100 synthetic sphere scenes rendered with the 100 MERL BRDFs [43]. We rendered the images under ten lighting environments consisting of $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ uniformly distributed light sources. 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 [5]. We call this dataset as *PrincipledPS*. For each shape, we prepared two materials, Specular and Metallic, as defined by Ikehata [4], four spatially varying textures, and sparse and dense (10 and 100) light configurations, totally, 64 scenes. Figure 3.4 shows the ground truth surface normal maps and example images of the PrincipledPS dataset.

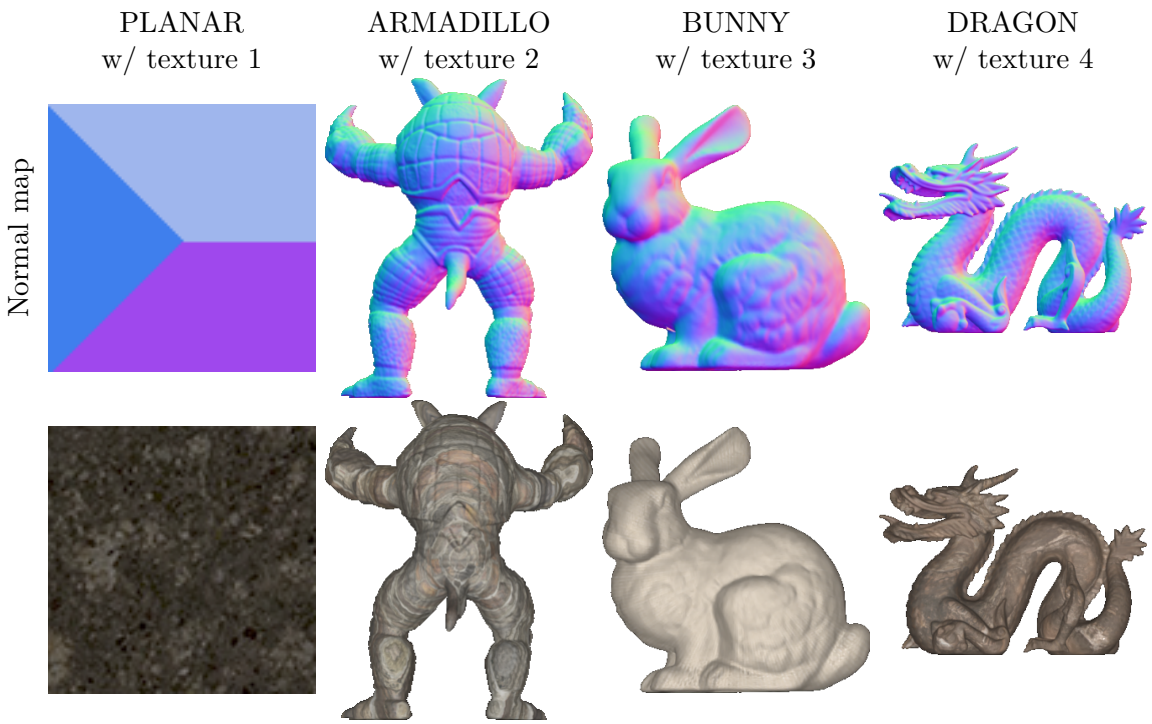


Fig. 3.4 Ground truth surface normals and example images of PrincipledPS dataset.

Real-world dataset: We use an existing real-world dataset, the DiLiGenT dataset [3], 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 [4]. In addition to the original dataset, for testing sparse light cases, we prepared 20 datasets, each containing 10 randomly selected images.

Baselines: As baselines we used Lambertian photometric stereo (LPS) [20], the model-based method ST14 [27], the virtual exemplar-based method HS17 [2], hypothesis-and-test search (HaTS-PS) presented in Chapter 2, the unsupervised learning (*i.e.*, neural inverse rendering)-based method NIR-PS [46], the supervised learning methods PX-NET [35], PS-FCN^{+N} [24], WJ20 [47], CNN-PS [4], and SPLINE-Net [38]. 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 [48] 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. 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. 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. Also for testing with the PrincipledPS dataset, although PX-NET, CNN-PS, and SPLINE-Net include the target material in their pre-trained models, we list their scores for reference.

3.4.2 Implementation

Our DSPS can benefit from any exact or approximate nearest-neighbor search method based on ℓ_2 distance (*e.g.*, [50, 51, 60–64]) implemented in modern libraries [48, 63, 65, 66]. In our experiments, we used a simple linear search algorithm implemented in FAISS [65] as an exact method. As an approximate method, we adopted a combination of an inverted file system with asymmetric distance computation (IVFADC) [52] and a hierarchical navigable small worlds (HNSW) indexing structure [53] implemented in FAISS [65]. The HNSW and IVFADC require to set their hyper-parameters listed in Tab. 3.2. In all the experiments of this chapter, we used 32, 1000, 8, and 8 for HNSW_M, nlist, nbits_per_idx, and nprobe, respectively. For the hyper-parameter M_sub, we have to use a different value in each experiment depending on the number of lights due to its requirements[1]. For all experiments on the MERL sphere and PrincipledPS dataset, we used M_sub= 10. On the DiLiGenT dataset, we used M_sub= $5, 19, 24$ for the 10, 76, and 96 lights, respectively.

In the following, we denote DSPS with exact and approximate nearest neighbor

[1]See the wiki of the FAISS for details.

Table 3.2 Hyper-parameters for HNSW and IVFADC.

HNSW_M	The number of neighbors in HNSW
nlist	The number of cells for space partitioning in IVFADC
M_sub	The number of sub-vector in IVFADC
nbits_per_idx	Bits per sub-vector in IVFADC
nprobe	The number of probes at query time in IVFADC

search as **DSPS-E** and **DSPS-A**, respectively. Both DSPS-E and DSPS-A using FAISS can be performed on either a CPU or a GPU.

3.4.3 Efficiency of surface normal estimation

This section shows comparisons of computation time 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 time of DSPS-E, DSPS-A, HaTS-PS, and HS17 [2] on a CPU. We also measured the computation time of DSPS-E, DSPS-A, CNN-PS [4], 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 3.5a shows the computation time for a single pixel on the CPU, averaged over all MERL spheres for each light configuration. DSPS-A is 3–4 orders of magnitude faster than HS17 while DSPS-E are around one order of magnitude faster than HS17. Figure 3.5b shows the computation time for a single pixel on the GPU. DSPS-E and DSPS-A are accelerated one order of magnitude using the GPU. While typical exemplar-based methods are computationally expensive, our methods achieve compa-

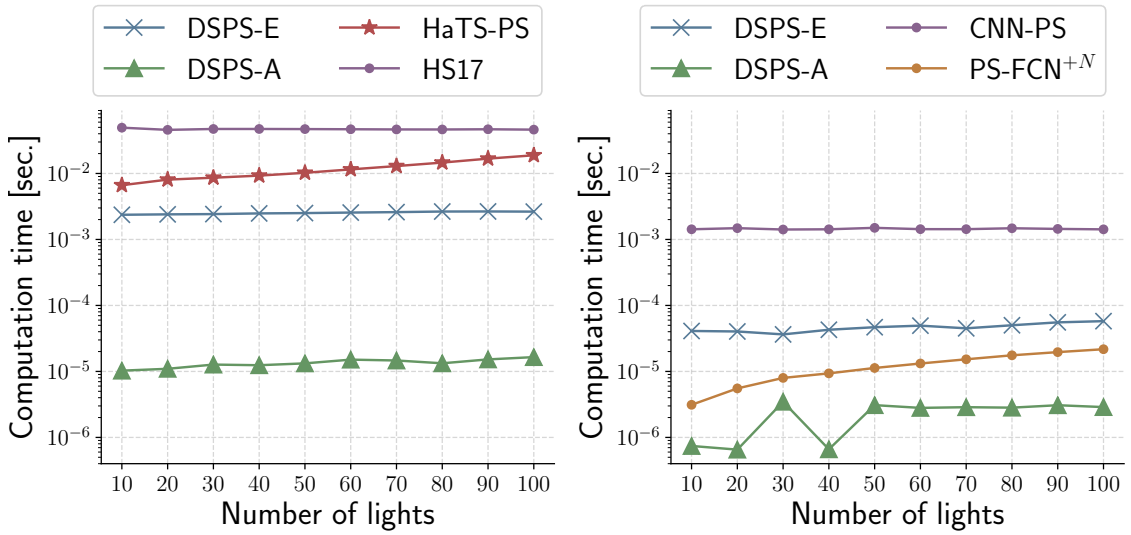


Fig. 3.5 (a) *CPU* computation time of our methods, HaTS-PS, and HS17 [2] for a single pixel. (b) *GPU* computation time of our methods, CNN-PS [4], and PS-FCN^{+N}.

rable or faster inference than the learning-based methods using feed-forward networks.

3.4.4 Accuracy of surface normal estimation

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

MERL sphere: We compared our methods and the baseline methods using the MERL sphere dataset. For the materials in our method, HaTS-PS, 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 3.3 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. The stable and high accuracy for diverse materials of our methods is confirmed in Fig. 3.6, showing

Table 3.3 Comparisons on the MERL sphere dataset with ten light configuration sets. Numbers represent averages and standard deviations of angular errors over 100 MERL spheres.

	#lights	10	20	30	40	50	60	70	80	90	100
Exemplar-based	DSPS-E	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-A	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
	HaTS-PS	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
	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.

mean angular errors of our method and several baseline methods for each material in the 100 lights case. While HS17 also achieves competitive accuracy, it is more than three orders of magnitude slower than DSPS-A as shown previously. Our methods achieve remarkably stable surface normal estimation in the few lights case such as ten lights. It is a benefit of treating BRDFs in a discrete manner instead of a continuous manner that tends to be over-fit to the measurements in a few light cases. Incidentally, NIR-PS yields large angular errors in this experiment. We observed that NIR-PS has extremely large errors for several materials as shown in Fig. 3.6, which affect the averaged scores.

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 3.4 shows averages of angular errors over eight scenes (*i.e.*, all combinations

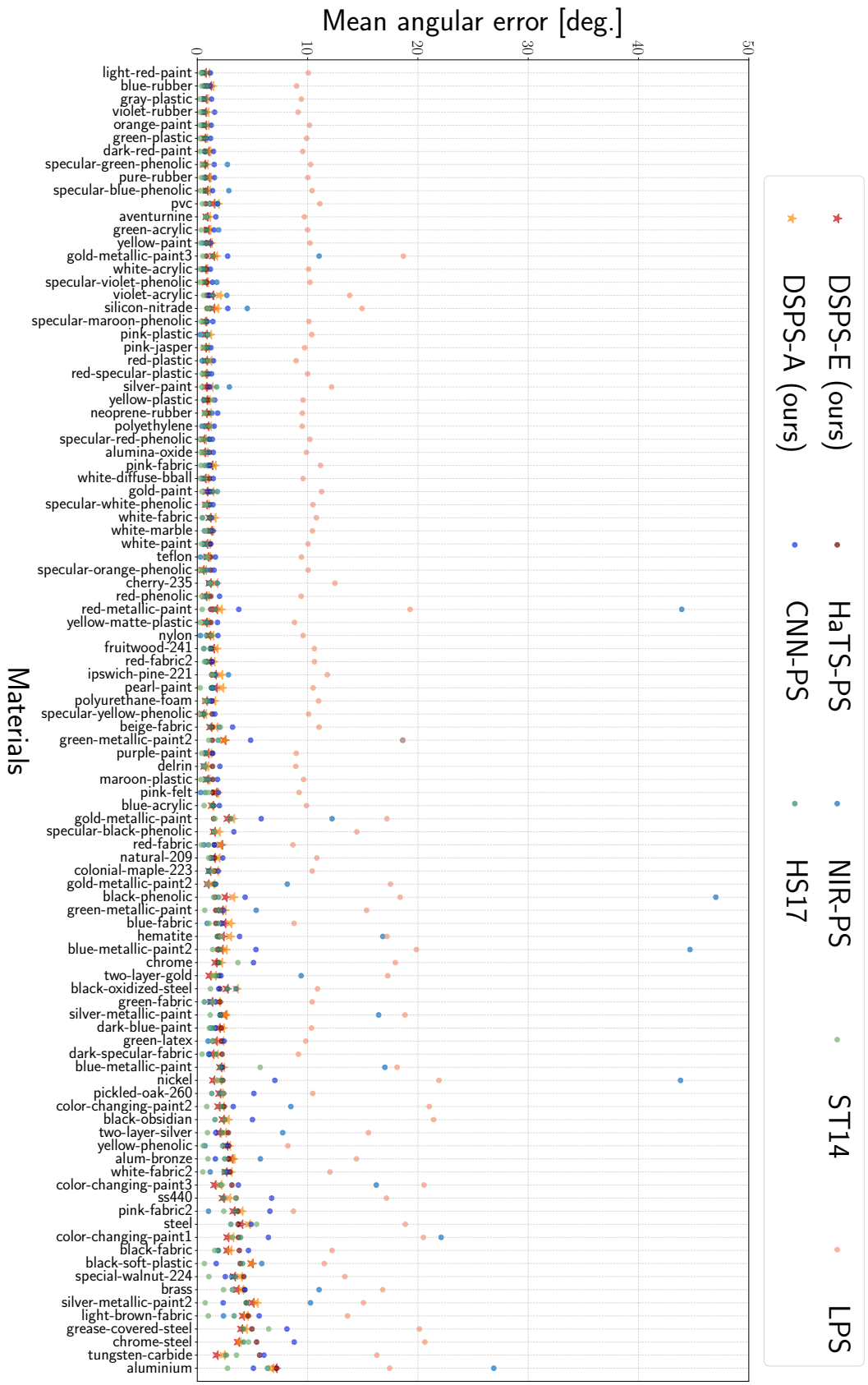


Fig. 3.6 Mean angular errors of our method and the baseline methods for each MERL sphere data with 100 lights.

Table 3.4 Comparisons on the PrincipledPS dataset. Numbers represent averages of angular errors over eight scenes, *i.e.*, two materials and four textures.

		10 lights					100 lights				
		PLANAR	ARMADILLO	BUNNY	DRAGON	Avg.	PLANAR	ARMADILLO	BUNNY	DRAGON	Avg.
Exemplar-based	DSPS-E	1.5	3.6	3.5	3.6	3.0	1.0	3.4	3.1	3.2	2.7
	DSPS-A	1.5	3.6	3.5	3.6	3.0	1.6	3.6	3.5	3.4	3.0
	HaTS-PS	15.9	4.0	4.0	4.1	7.0	1.1	2.7	2.6	2.7	2.3
	HS17	1.6	3.9	3.7	3.8	3.3	1.2	2.4	2.2	2.3	2.0
Learning-based	PX-NET[a]	10.6	4.8	5.3	5.0	6.4	1.2	1.4	1.4	1.5	1.4
	PS-FCN^{+N}	6.9	5.1	4.8	5.4	5.6	2.9	3.6	3.5	3.8	3.4
	WJ20	4.9	3.6	3.5	3.6	3.9	2.3	2.9	2.7	2.9	2.7
	SPLINE-Net[a]	9.2	6.2	6.4	6.4	7.0	33.8	41.3	45.4	42.0	40.6
	CNN-PS[ab]	28.7	30.3	34.2	30.5	30.9	4.9	1.8	2.0	1.9	2.6
	NIR-PS	48.0	2.9	2.5	3.1	14.1	41.0	2.9	2.6	2.8	12.3
Model-based	ST14	15.0	12.2	12.7	11.3	12.8	1.6	7.0	2.5	7.2	4.6
	LPS	13.7	10.2	10.4	9.3	10.9	13.4	8.2	8.3	7.5	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.

of two materials and four textures) for each shape and number of lights. The results on the PrincipledPS dataset also show that our DSPS achieves accurate and stable surface normal estimation for diverse materials in both sparse and dense lighting cases. In the sparse lighting case, DSPS has a higher accuracy than HaTS-PS. This is because DSPS treats BRDFs in a discrete manner, while HaTS-PS treats them in a continuous manner that tends to be over-fit to the measurements, particularly in sparse lighting case. In contrast, HaTS-PS is more accurate than DSPS in the dense lighting case since the continuous BRDF model can represent more diverse materials than the discrete BRDF model of DSPS and the over-fitting of the continuous BRDF model rarely happen if the number of lights is large. 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 3.5 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 datasets with different light distributions.

96 lights

		BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
Exemplar-based	DSPS-E	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
	DSPS-A	1.4	6.4	14.2	6.8	8.0	11.7	17.5	7.4	7.4	15.3	9.6
	HaTS-PS	1.6	5.9	13.1	6.1	9.2	11.0	18.7	6.6	7.2	15.0	9.4
	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-E	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-A	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
	HaTS-PS	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
	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
	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[b]	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] CNN-PS is trained with 50-100 lights.

DiLiGenT: We show quantitative results on the real-world dataset DiLiGenT with 96 and 10 lights in Tab. 3.5, where we compare our methods with the baseline methods in terms of mean angular error. Figures 3.7 and 3.8 show visual comparisons between our methods and the baseline methods in 96 lights case. 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-E 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. 3.8. Although our DSPS accurately estimates surface normals on convex surfaces, HaTS-PS achieves further accuracy at several pixels (*e.g.*, the body of BEAR, BUDDHA, and CAT in Fig. 3.7). This is due to the difference in the BRDF models, namely continuous or discrete model. HaTS-PS employs a continuous BRDF model that can represent more diverse materials than the discrete model employed in DSPS; therefore, HaTS-PS can estimate better surface normals than DSPS.

For the scenes with 10 lights, our methods achieve comparable accuracy to the learning-based methods. The standard deviations of our DSPS tend to be small compared to the baselines, which suggest that DSPS 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.

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).

Reliability of surface normal estimates: In practical applications, it is important to know the reliability of estimated surface normals. When the measurement

Fig. 3.7 Angular error maps and estimated surface normal maps for BALL, BEAR, BUDDHA, and CAT objects in the DiLiGenT dataset [3] with all the 96 lights.

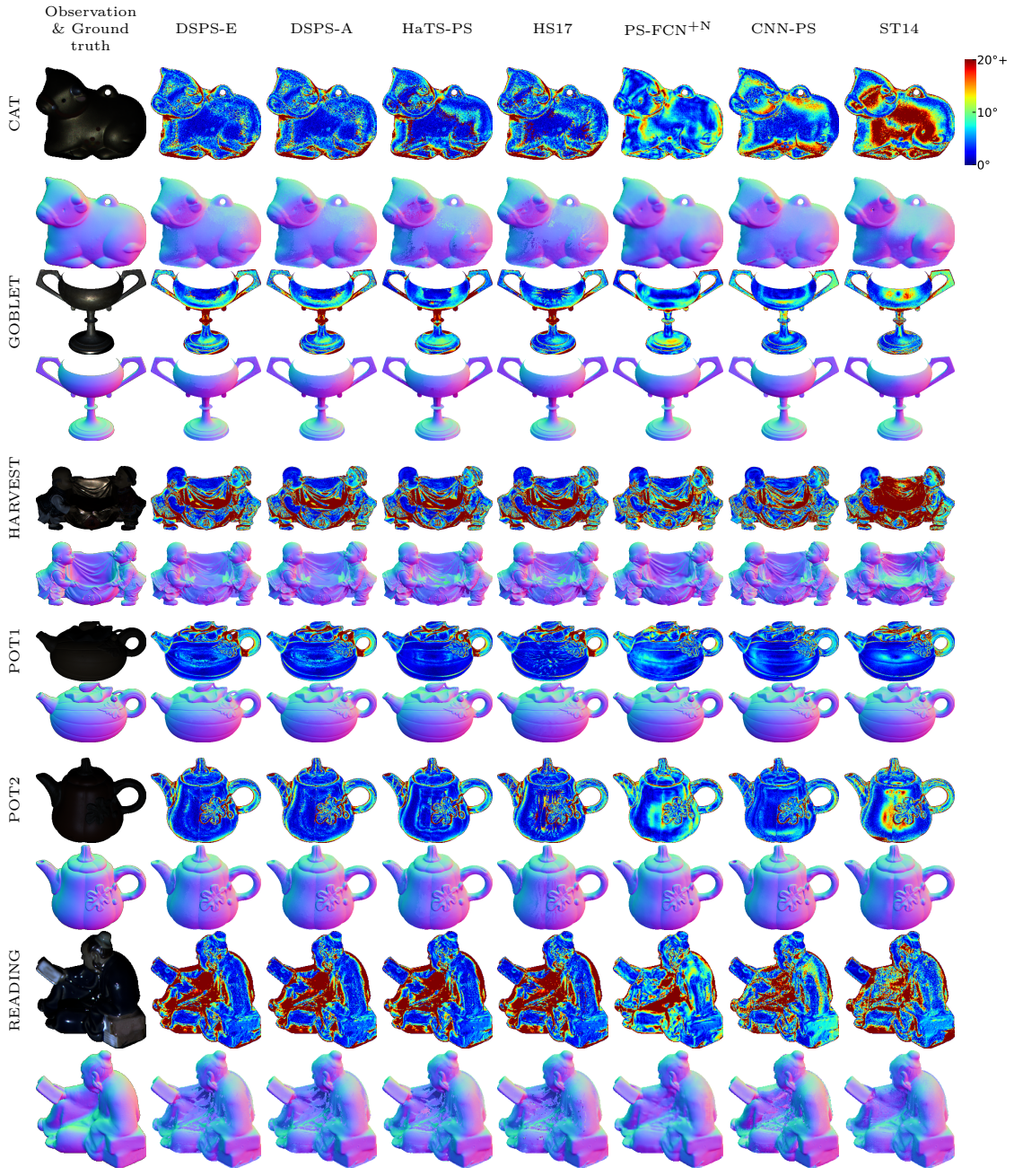


Fig. 3.8 Angular error maps and estimated surface normal maps for COW, GOBLET, HARVEST, POT1, POT2, and READING objects in the DiLiGenT dataset [3] with all the 96 lights.

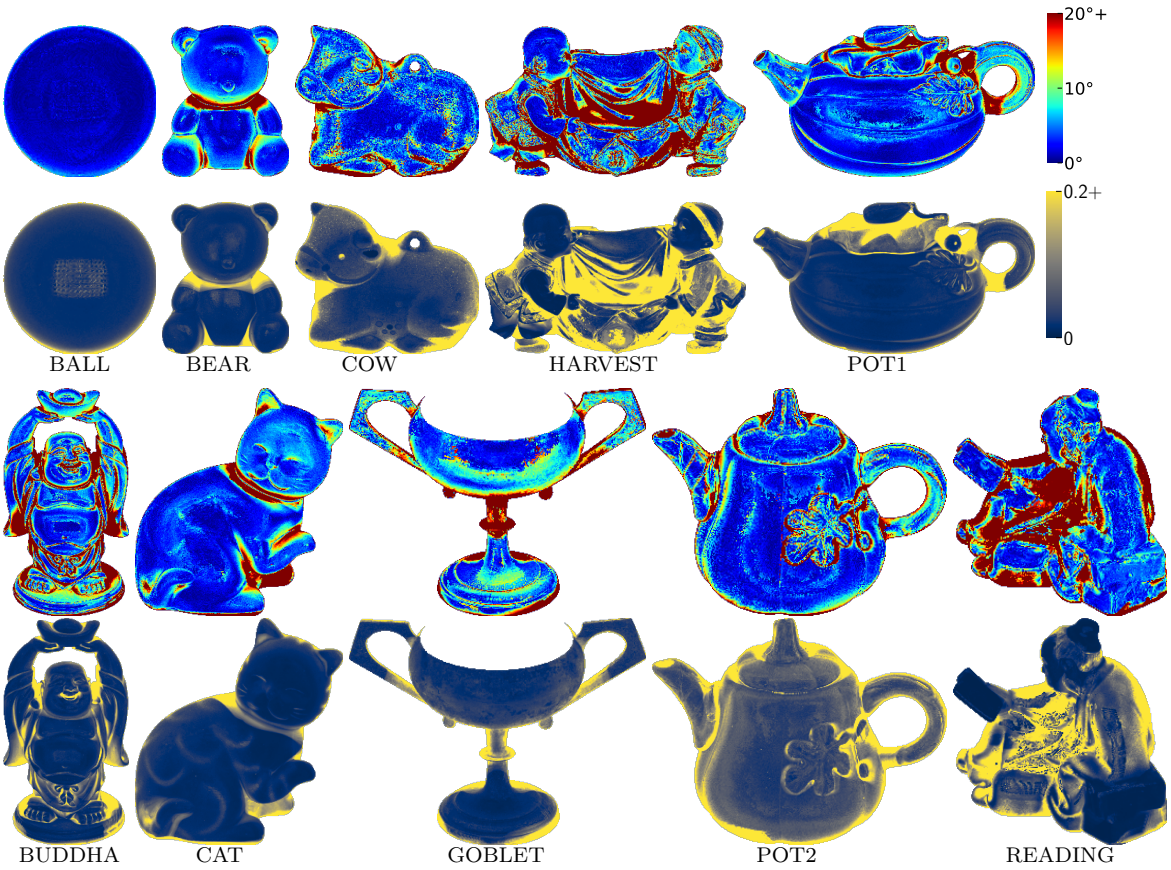


Fig. 3.9 Visual relation between angular errors and image reconstruction errors. For each object in the DiLiGenT dataset, we show angular error maps (above) and image reconstruction error maps (below).

vector is far from any appearance exemplars, it can be considered unstable estimation; therefore, we can assess the reliability via the nearest neighbor search process. Indeed, we can observe that larger image reconstruction errors, which can be calculated with Eq. (3.8), tend to correspond to higher angular errors as shown in Fig. 3.9. In particular, we can observe large angular errors and image reconstruction errors at pixels that can be considered affected by global illuminations (*e.g.*, the neck of BEAR and CAT). Therefore, we can find such pixels with unreliable surface normal estimates by our method and may use other photometric stereo method such as learning-based method for more reliable surface normal estimation.

3.4.5 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. We simulated the camera noise by adding signal-independent and signal-dependent noise [45] to images in the same manner as previous work [27]; $\tilde{m} = m + (\mu + \lambda\sqrt{m})X$, 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. For the ambient light and saturation, we followed dataset generation process of PX-NET [35]. To simulate the ambient light, we added $\mathbf{n}^\top \mathbf{v} X$ to images, where $\mathbf{n}$ is a surface normal, $\mathbf{v}$ is a viewing direction, and X is a $\mathcal{U}(0, 0.001)$-distributed random variable and constant over a single scene. To simulate the saturation of pixel intensity, we clipped the top 5% of pixel intensities with the highest values in the half of the images.

Table 3.6 shows mean angular errors and standard deviations for each corrupted data. The results suggest that exemplar-based methods including ours, HaTS-PS, 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 the saturation since they eliminate large measurement values as outliers.

The robustness of exemplar-based methods can be explained by interpreting the exemplar-based approach 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.

Table 3.6 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-E	1.7/2.4	2.4/4.7	7.7/15.9	2.8/4.1
DSPS-A	2.0/2.5	2.8/4.8	7.9/15.6	3.0/4.1
HaTS-PS	1.8/2.7	2.7/5.2	7.7/16.1	2.6/3.2
HS17	1.6/2.4	2.4/4.7	7.7/15.9	2.6/3.8
CNN-PS	2.5/3.8	6.2/14.0	8.9/18.8	2.6/3.8
ST14	1.2/2.3	22.9/13.3	34.5/33.7	1.2/2.3

3.4.6 Analysis of appearance tensor

The appearance tensor is constructed from three components; BRDFs, surface normals, and light directions. In the experiments so far, we used the appearance tensor with 100 MERL BRDFs, 20001 surface normals, and exact light directions of a target scene.

This section analyzes the effect of varying appearance tensors on the quality of surface normal estimation.

Appearance tensor with non-MERL BRDFs: We investigate whether BRDF bases from synthetic non-MERL BRDFs improve our method. Here, we use Disney's Principled BSDFs [5], Oren-Nayar [6], Blinn-Phong [7], and Cook-Torrance [8] BRDF models for the appearance tensor. We discretize material parameters for each BRDF model and prepare 162 bases from Principled BSDFs, 100 bases from Oren-Nayar BRDF, 11 bases from Blinn-Phong BRDF, and 54 bases from Cook-Torrance BRDF.

Table 3.7 shows mean angular errors of our DSPS-E with different BRDF bases on the DiLiGenT dataset. The results suggest that the additional synthetic BRDF bases do not contribute to the quality of surface normal estimation on the real data. It is visually confirmed by Fig. 3.10, which shows the difference in the angular error maps between DSPS-E with MERL BRDF bases and that with MERL & Principled BSDF bases. This trend is consistent in DSPS-E with other BRDF bases. We consider that

Table 3.7 Our DSPS-E with different BRDF candidates. We use Disney's principled BSDF [5], Oren-Nayar [6], Blinn-Phong [7], and Cook-Torrance [8]. The experiments are performed on the DiLiGenT dataset.

	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
MERL	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
Principled	1.4	6.4	14.8	7.8	7.7	13.4	18.6	8.3	9.6	16.0	10.4
MERL & Principled	1.4	6.4	14.3	7.1	7.3	11.5	17.5	7.8	7.4	15.5	9.6
MERL & Oren-Nayar	1.4	6.5	14.2	6.8	7.8	11.5	17.5	7.4	7.4	15.3	9.6
MERL & Blinn-Phong	1.3	6.6	15.4	7.4	8.5	12.4	17.7	7.7	7.9	17.4	10.2
MERL & Cook-Torrance	1.3	6.4	14.3	7.0	7.8	11.9	17.3	7.4	7.6	15.5	9.7

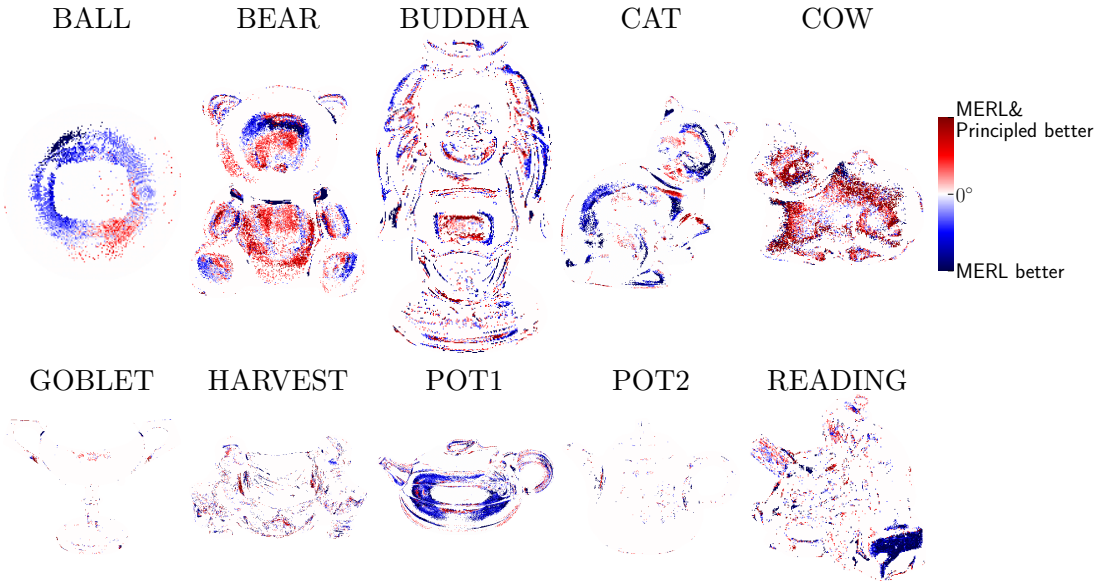


Fig. 3.10 Difference in the angular error maps between DSPS-E with MERL BRDF bases and MERL & Principled BSDF bases. Blue color indicates that the MERL only BRDF bases work better than the MERL & Principled BSDF bases and red color indicates the opposite.

it is because the analytic BRDFs still deviate from real-world BRDFs even though they add more diversity to our appearance tensor. Hence, we conclude to recommend using only the MERL BRDFs for the appearance tensor.

Varying number of BRDFs: The experimental results so far show that our DSPS is consistently comparable or better than HaTS-PS in terms of efficiency and accuracy. However, it is of interest to see how the accuracy of DSPS varies when the number

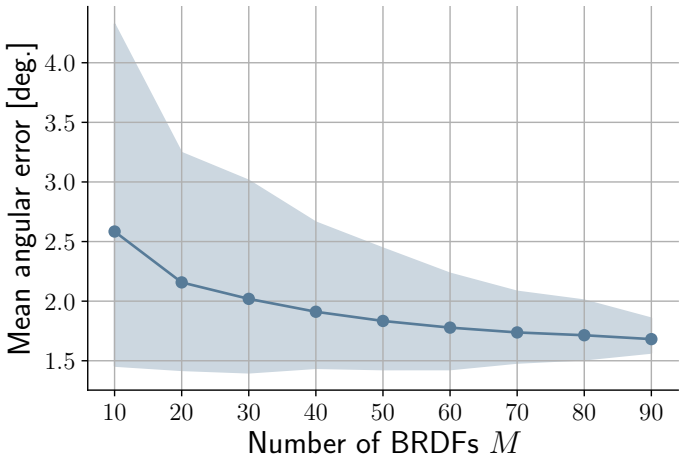
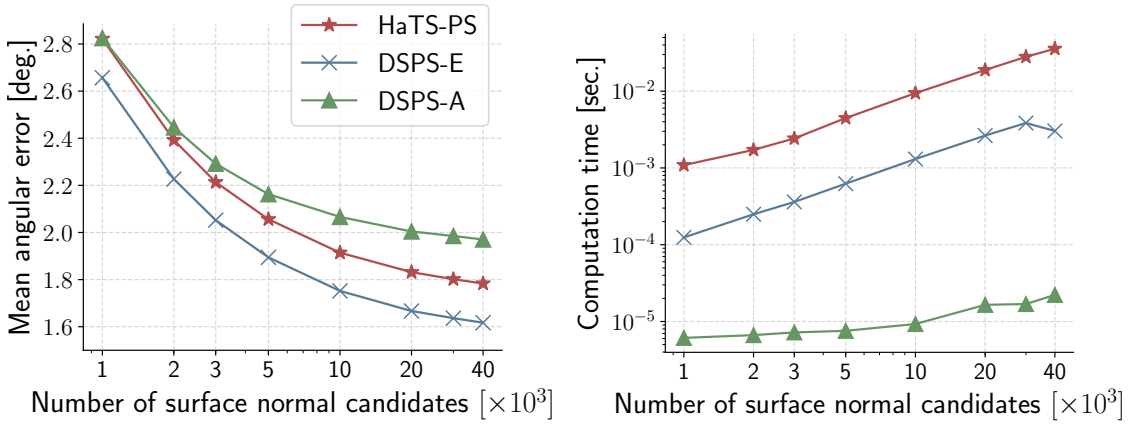


Fig. 3.11 Relationship between the accuracy of surface normal estimation and the number of BRDFs in the appearance tensor in DSPS. This experiment uses the MERL sphere dataset with 100 lights. The solid line shows the mean angular error of the ten trials, and the colored area shows the maximum and minimum angular errors of the trials.

of BRDFs of the appearance tensor is limited since DSPS 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, and repeat them ten times for obtaining the average accuracy.

Figure 3.11 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 30 BRDFs or more give promising surface normal estimation, around 2° in average, around 3° at worst. The reason why DSPS with such small number of BRDFs successfully works is that the Eq. (3.3) 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.

Surface normal discretization: The accuracy and efficiency of our DSPS and HaTS-PS are naturally affected by the granularity of the surface normal discretization.



(a) (b)

Fig. 3.12 (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 3.8 Increases of angular errors due to discretized lights. As pre-defined light directions in the appearance tensor we used 20001 directions created in the same way as the surface normal candidates. 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-E	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
DSPS-A	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Figure 3.12 shows mean angular error and computation time for a single pixel with varying numbers of surface normal candidates. This experiments are performed on the MERL sphere dataset with 100 lights. Throughout this chapter 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.

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 vector for a surface normal, 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 an experiment on the MERL sphere dataset with ten types of light configurations.

Table 3.8 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 prepare an appearance tensor for sufficiently finely discretized light directions and there is no need to calculate a new appearance tensor for each light configuration.

3.4.7 Precomputation cost

Nearest neighbor search methods used in our DSPS need precomputation/pretraining for each light configuration to enable efficient search. This section investigates the costs of precomputation on the CPUs and GPUs used in Sec. 3.4.3.

Figure 3.13 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 once for each light configuration is worth paying for the efficient inference shown in Figs 3.5a and b.

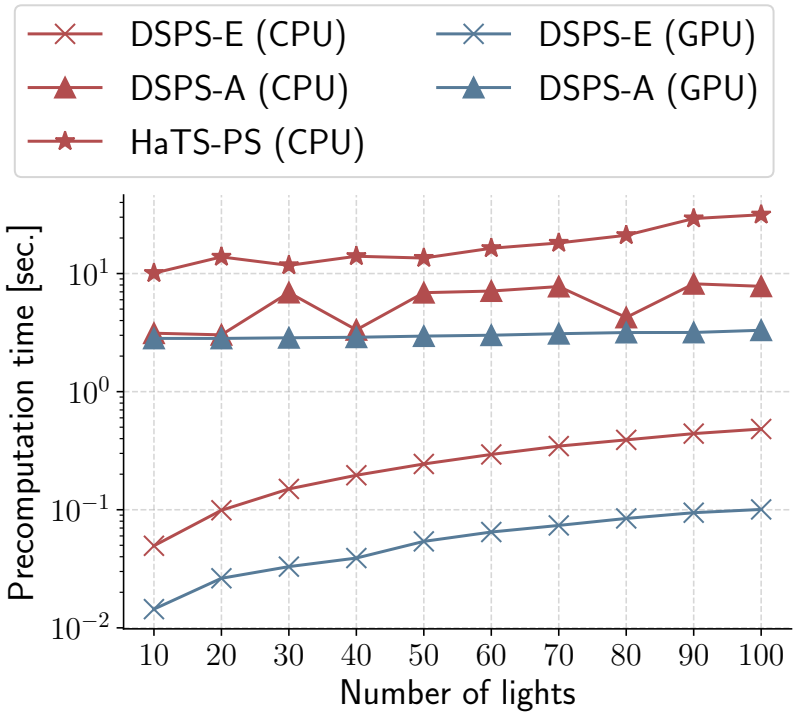


Fig. 3.13 Precomputation time of our methods on a CPU and GPU for varying light configurations.

3.4.8 Relighting quality

Relighting is a practical application that uses per-pixel surface normals and BRDFs recovered by the photometric stereo, which renders a recovered scene with novel illuminations. Also, evaluating relighting quality can assess the accuracy of both surface normal and BRDF estimations. Here, we evaluate the relighting quality by the relighting error e_{relit} defined as

$$e_{\text{relit}} = \left\| \frac{\mathbf{m}_{\text{relit}}}{\|\mathbf{m}_{\text{relit}}\|_2} - \frac{\mathbf{d}_{\text{relit}}}{\|\mathbf{d}_{\text{relit}}\|_2} \right\|_2, \quad \mathbf{d}_{\text{relit}} = \begin{pmatrix} \rho^*(\mathbf{n}^*, \hat{\mathbf{l}}_1) \max(\mathbf{n}^{*\top} \hat{\mathbf{l}}_1, 0) \\ \vdots \\ \rho^*(\mathbf{n}^*, \hat{\mathbf{l}}_{\hat{L}}) \max(\mathbf{n}^{*\top} \hat{\mathbf{l}}_{\hat{L}}, 0) \end{pmatrix}, \tag{3.9}$$

where $\mathbf{m}_{\text{relit}}$ is a ground truth measurement vector under novel illuminations $\hat{\mathcal{L}} = \{\hat{\mathbf{l}}_1, \ldots, \hat{\mathbf{l}}_{\hat{L}}\}$ unused for surface normal and BRDF recovery and $\mathbf{d}_{\text{relit}}$ is relit scene's appearances using estimated surface normal $\mathbf{n}^*$ and BRDF $\rho^*(\cdot)$ under novel illumi-

nations $\hat{\mathcal{L}}$. In addition, we analyze the degree of over-fitting by treating the image reconstruction and relighting error as training and test errors, respectively. For this, we relit the MERL spheres with 251 uniformly distributed lights using the estimated surface normals and BRDFs for 10 and 100 lights cases. We used an existing exemplar-based method (HS17 [2]) as a baseline.

Figure 3.14 shows cumulative histograms of the relighting errors and image reconstruction errors for all pixels in the MERL sphere dataset. Our method produces smaller relighting errors than HS17 in the case of 10 lights. This suggests that our method estimates both the surface normals and the BRDFs more accurately than HS17. The superiority of our method can be also visually confirmed in Fig. 3.15. Moreover, for the case 100 lights, our method achieves relighting errors competitive to HS17.

Compared to our method, HS17 produces small reconstruction errors and large relighting errors, especially in the case of 10 lights. This can be viewed as an evidence of over-fitting at several pixels since HS17 adopts a linear combination of BRDF candidates to explain the target measurements. In contrast, our method avoids over-fitting by using only a single BRDF candidate to approximate the target measurements. This results in a stable estimation of both surface normals and BRDFs.

3.5 Conclusion

In this chapter, we have presented Discrete Search Photometric Stereo (DSPS), which reduces the photometric stereo problem to the well-known nearest neighbor search problem. DSPS can stably recover surface normals of a scene with spatially varying general BRDFs in various light configurations. Using advanced nearest neighbor search methods 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 com-

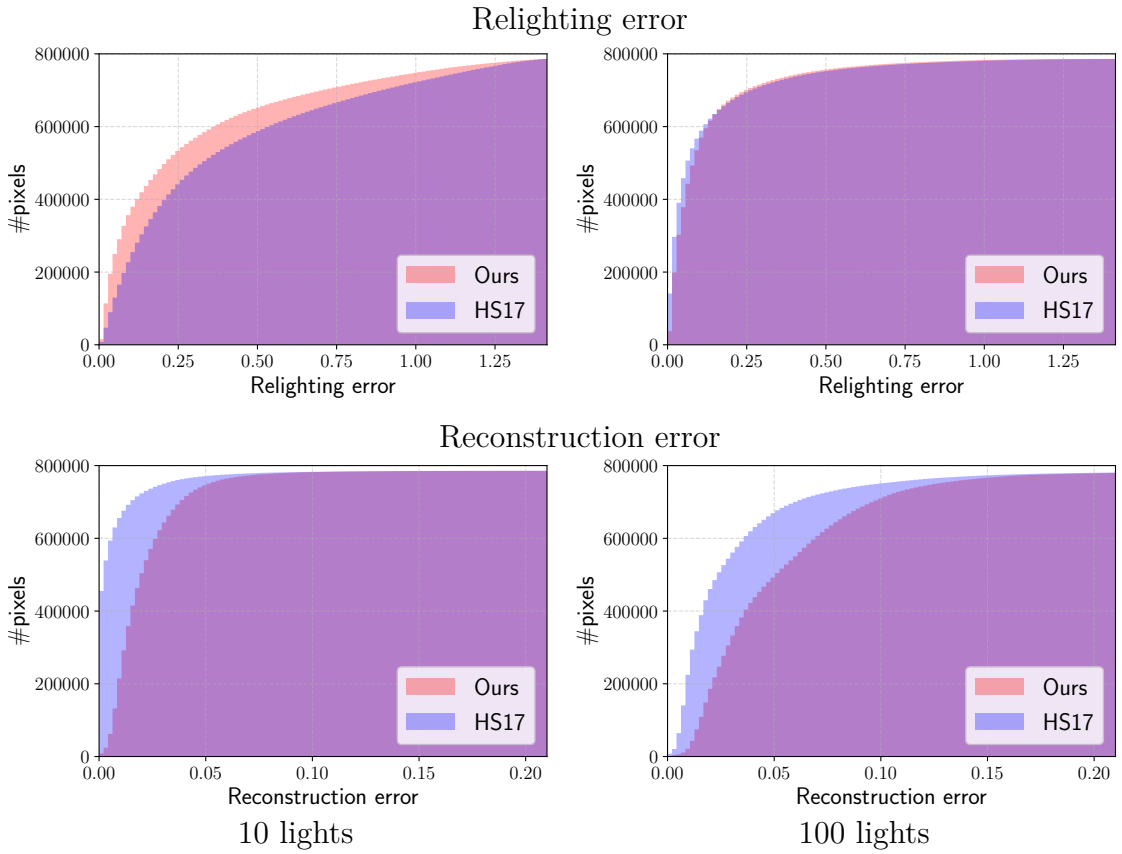
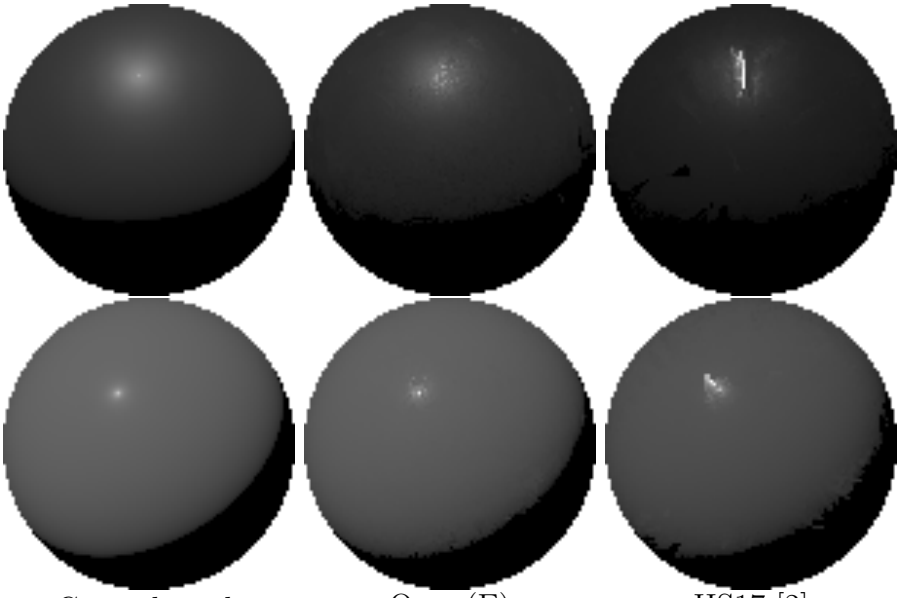


Fig. 3.14 Cumulative histograms of the relighting and reconstruction errors for all the pixels in the MERL sphere dataset. Relighting errors are calculated from the estimated surface normals and BRDFs in 10 and 100 lights cases.

parable accuracy to the state-of-the-art exemplar-based photometric stereo methods while achieving 100–1000× acceleration. In addition, we experimentally observed that our DSPS is robust to imaging noise compared to model-based and learning-based methods. Since it is hard to entirely avoid imaging noise in real-world experiments, DSPS is one of the best choices for stable surface normal estimation.

While our DSPS showed promising surface normal estimation, it was limited to convex surfaces since the appearance exemplars only consider convex surfaces and do not consider global illumination effects that are likely to occur in non-convex surfaces. We leave the extension of DSPS to non-convex surfaces as future work, which is addressed in Chapter 4.



Ground truth Ours (E) HS17 [2]

Fig. 3.15 Visual comparison of relighting results for our method and HS17 [2]. We performed the relighting with 251 novel lights using the surface normals and BRDFs estimated from just 10 lighting directions.

Chapter 4

General Appearance Exemplars for Nearest Neighbor Search-based Photometric Stereo

4.1 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 [20] assume Lambertian reflectance, which deviates from real-world reflectances, thus introducing errors in surface normal estimates. Discrete search photometric stereo (DSPS) proposed in Chapter 3 achieves accurate surface normal estimation for diverse reflectances by a discrete search for the appearances closest to target scene's ones over a set of appearance exemplars. However, the applicability is limited by the coverage of the appearance exemplars; namely, if target scene's appearances are distant from any appearance exemplar, the estimation should be unreliable. Indeed, the accuracy of DSPS is degraded at non-convex surfaces due to global illumination effects such as cast shadows and inter-reflections that are not considered in the set of appearance exemplars used in Chapter 3. We, therefore, extend the applicability of DSPS by augmenting the set of appearance exemplars with more reflectances and

global illumination effects.

This chapter introduces a set of general appearance exemplars to broaden the applicability of the DSPS to non-convex surfaces and more diverse materials. We design the general appearance exemplars to have real reflectances and not be restricted to a specific class of shapes. For real reflectances, we use actually measured bidirectional reflectance distribution function (BRDF) database. A limited number of BRDFs are only available, specifically 100 in the MERL BRDF database [43]; therefore, we augment the measured BRDFs while maintaining its realistic reflectance property. For non-convex surfaces not restricted to a specific class of shapes, we use a large scale shape dataset containing randomly corrupted primitive shapes (*e.g.*, cubes, ellipsoids, cylinders).

To maintain the favorable accuracy of the DSPS on convex surfaces, we build convex and non-convex appearance exemplars in a respective manner. The convex appearance exemplars are rendered with finely discretized surface normals and augmented BRDFs. While the non-convex appearance exemplars containing global illumination effects can be obtained by rendering non-convex shapes, it also includes convex appearance exemplars, resulting in a redundant dataset if convex appearance exemplars are combined. Therefore, we extract only purely non-convex appearance exemplars with our metric. Our general appearance exemplars constructed from pure convex and non-convex appearance exemplars allow us to estimate a convexity (convex or non-convex) of a surface in addition to a surface normal and BRDF. The knowledge of convexity further allows us to apply different photometric stereo methods to convex and non-convex surfaces.

Our general appearance exemplars are motivated by the success of training datasets for learning-based photometric stereo methods [4, 24]. Recent learning-based methods achieve robust surface normal estimation for non-convex surfaces and diverse reflectances using neural networks being trained with datasets containing diverse convex/non-convex shapes and materials. However, the CyclesPS dataset proposed by Ikehata [4] is constructed from only fifteen 3D models and Disney's Principled

BSDFs [5] that is unsuitable for photometric stereo as shown in Chapter 3. The training dataset proposed by Chen *et al.* [24] is constructed from ten shapes of Blobby dataset [41] and eight shapes of Sculpture dataset [42] and 100 measured BRDFs [43]. In contrast, we construct the general appearance exemplars using more than a hundred randomly generated shapes and augmented measured BRDFs.

We assess the validity of our general appearance exemplars on the DSPS using synthetic and real-world datasets. We also analyze the effects of measured BRDF and shape augmentation, respectively. Lastly, we present that combining different photometric stereo methods using the knowledge of estimated convexity improves the accuracy from both methods.

4.2 Related work

We first review representative BRDF representations, analytic BRDF and measured BRDF. We then describe datasets for training learning-based photometric stereo methods and their relation to our work.

4.2.1 Analytic BRDF models

Analytic BRDF models aim to reproduce real-world reflectances by analytical formulas. The Lambertian model and more generalized Oren-Nayar model [6] are early analytic models for diffuse reflectances. The specular reflectance is much more complicated to describe, and a lot of analytic models are proposed. Early specular models are derived based on empirical observations, *e.g.*, Phong [67], Blinn-Phong [7], Ward [68], and Lafortune [69] models. More recently, physically-based microfacet models [8, 70–73] are introduced to better represent the roughness at a fine scale. Generally, real-world reflectances are approximated by combining diffuse and specular models. For example, Disney's principled BSDF [5] incorporates several analytic models to represent diffuse, specular, and metallic reflectances with a single model [5]. These analytic models are successful in rendering realistic scenes; however, they are still deviated

from real reflectance behaviors and introduce errors in photometry vision applications such as photometric stereo.

4.2.2 Measured BRDF datasets

In contrast to analytic BRDF models, measured BRDFs are always real. The MERL BRDF dataset [43] is the first large-scale measured BRDF dataset, which contains densely sampled 100 real-world isotropic materials, from diffuse materials to hard specular materials. UTIA dataset [74] consists of 150 anisotropic BRDFs. Recently, Dupuy and Jakob [75] proposed an adaptive BRDF parameterization and efficient sampling technique and measured 51 isotropic and 11 anisotropic BRDFs. However, the UTIA dataset and adaptively sampled BRDF dataset only contain sparse measurements, which do not fit the photometric stereo application. Therefore, we use the MERL BRDF dataset and augment it for more diversity.

4.2.3 Dataset for learning photometric stereo

Recent learning-based photometric stereo using neural networks are trained with synthetic datasets. Santo *et al.* [34] and Li *et al.* [37] built their training datasets by rendering ten shapes in the Blobby shape dataset [41] with the 100 MERL BRDFs. Chen *et al.* [24] employed ten shapes in the Blobby shape dataset and eight shapes in the Sculpture dataset [42] and the MERL BRDFs to create their training dataset, which is also used in several following learning-based methods [47, 54]. Ikehata [4] proposed the CyclesPS dataset [28] to train their network, which is constructed from fifteen scenes rendered with Disney's principled BSDF [5]. The CyclesPS dataset is also used in SPLINE-Net [38]. Logothetis *et al.* [35] first generate a set of direct reflectance components with the Disney's principled BSDF and MERL BRDFs. To increase the realism of their dataset, they manually simulate the effects of cast shadows, inter-reflections, surface discontinuities, ambient lights, noises, and pixel saturations in their proposed manner. While these datasets enable learning neural networks for

the photometric stereo, they rely on unrealistic analytic BRDFs or a limited class of shapes. We, therefore, design our general appearance exemplars to have more diverse, real BRDFs and various shapes not limited to a specific class.

4.3 General appearance exemplars

This section presents our new set of appearance exemplars, called general appearance exemplars. Our DSPS presented in Chapter 3 demonstrates promising surface normal and BRDF estimation, especially on convex surfaces. However, existing appearance exemplars used in Chapter 3 only consider convex surfaces and 100 BRDFs, thus, introducing errors on non-convex surfaces or some problematic materials (see the experiment on the DiLiGenT dataset in Sec. 3.4.4). Therefore, we extend the appearance exemplars with respect to BRDF and shape in the following sections and build general appearance exemplars.

4.3.1 BRDF augmentation

The appearance exemplars used in DSPS are rendered with the MERL BRDFs [43], which are representative measured BRDFs and well-known as it successfully works on the photometric stereo task. However, the MERL BRDFs only contain 100 materials that are a limited set for addressing diverse reflectances. To overcome this issue, we augment the MERL BRDFs while maintaining its realistic reflectance property.

Let $\mathcal{B} = \{\rho_i(\cdot) \mid i = 1, \ldots, 100\}$ be a set of the MERL BRDFs. We simply generate a new BRDF by linearly combining randomly selected two MERL BRDFs as $\rho'_j(\cdot) = \rho_p(\cdot) + \rho_q(\cdot)$, where $p, q \in [1, 100], p \neq q$. Repeating this step $M' \leq {}_{100}\mathrm{C}_2$ times and generate an additional set of BRDFs $\mathcal{B}' = \{\rho'_j(\cdot) \mid j = 1, \ldots, M'\}$. Lastly, a set of $M' + 100$ BRDFs can be obtained as $\mathcal{B} \cup \mathcal{B}'$.

This BRDF augmentation is motivated by a success of a BRDF model that linearly combines multiple BRDFs [2, 25, 30, 49]. They often impose *non-negativity* and *sparsity* on the coefficients of the linear combination to avoid generating unrealistic

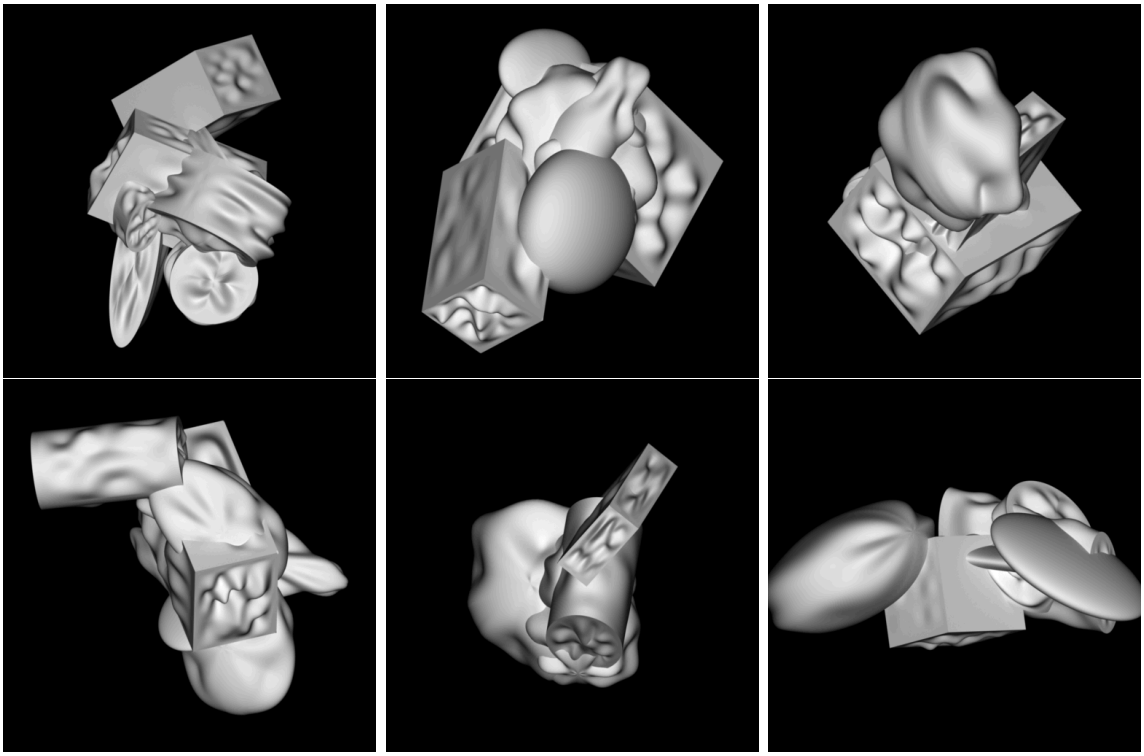


Fig. 4.1 Examples of randomly generated non-convex shapes combined with multiple corrupted primitive shapes.

BRDFs. These constraints are naturally satisfied in our BRDF augmentation. While a linear combination with random coefficients can generate more diverse BRDFs than a simple additive model, it runs a risk of increasing the noise of measured BRDFs, and the simple additive model can generate huge number of new BRDFs, ${}_{100}\mathrm{C}_2 = 4950$ at most. Therefore, we conclude to use the simple additive model of two BRDFs.

4.3.2 Non-convex appearance exemplars

The appearance exemplars used in DSPS only consider convex surfaces without global illumination effects such as cast shadows and inter-reflections. It largely degrades the accuracy of DSPS on non-convex surfaces, where global illuminations are likely caused. We then aim to create an additional set of appearance exemplars considering global illumination effects, which we call non-convex appearance exemplars.

We first generate diverse and random non-convex shapes. For this purpose, we

generate primitive shapes (*e.g.*, cubes, ellipsoids, and cylinders) with random parameters and apply random corruptions by following Xu's method [76] originally used for relighting task. Then, we construct a large set of non-convex shapes by combining multiple corrupted primitive shapes after random translations and rotations. The examples of the generated non-convex shapes are illustrated in Fig. 4.1. Non-convex appearance exemplars can be obtained by rendering the generated non-convex shapes with diverse BRDFs; however, it also contains appearance exemplars not affected by global illumination effects, resulting in a redundant set of appearance exemplars if convex appearance exemplars are combined. To reduce the redundancy of the appearance exemplars, we extract only appearance exemplars affected by global illuminations with scale invariant thresholding.

We render non-convex shapes with and without global illuminations, which provides appearance vectors $\mathbf{d}_w$ and $\mathbf{d}_{wo}$ for an identical scene point. If the scene point in the rendered images is affected by global illuminations, $\mathbf{d}_w$ must be different from $\mathbf{d}_{wo}$. We then find an appearance vector affected by global illuminations using the following thresholding:

$$\max\left(|(\mathbf{d}_w - \mathbf{d}_{wo}) \oslash (\mathbf{d}_{wo} + \epsilon)|\right) > \tau, \tag{4.1}$$

where τ and ϵ is a threshold and a small value to prevent zero-division, $\oslash$ indicates an element-wise division, $\max(\cdot)$ is a function taking the maximum value of a vector, and $|\cdot|$ is a function taking absolute values of vector's elements. This thresholding is invariant to a scale of an appearance vector that depends on a setting of rendering software and material, and a constant τ works well on every scene. With this thresholding, we extract only appearance vectors affected by global illuminations from rendered images as shown in Fig. 4.2 and define them as non-convex appearance exemplars. Throughout the experiments in this chapter, we use $\tau = 0.1$ and $\epsilon = 0.001$. We manually found that reasonable appearance exemplars can be extracted with $\tau = 0.1$ as shown in Fig. 4.2. τ less than 0.1 always leads to a more accurate surface normal estimation

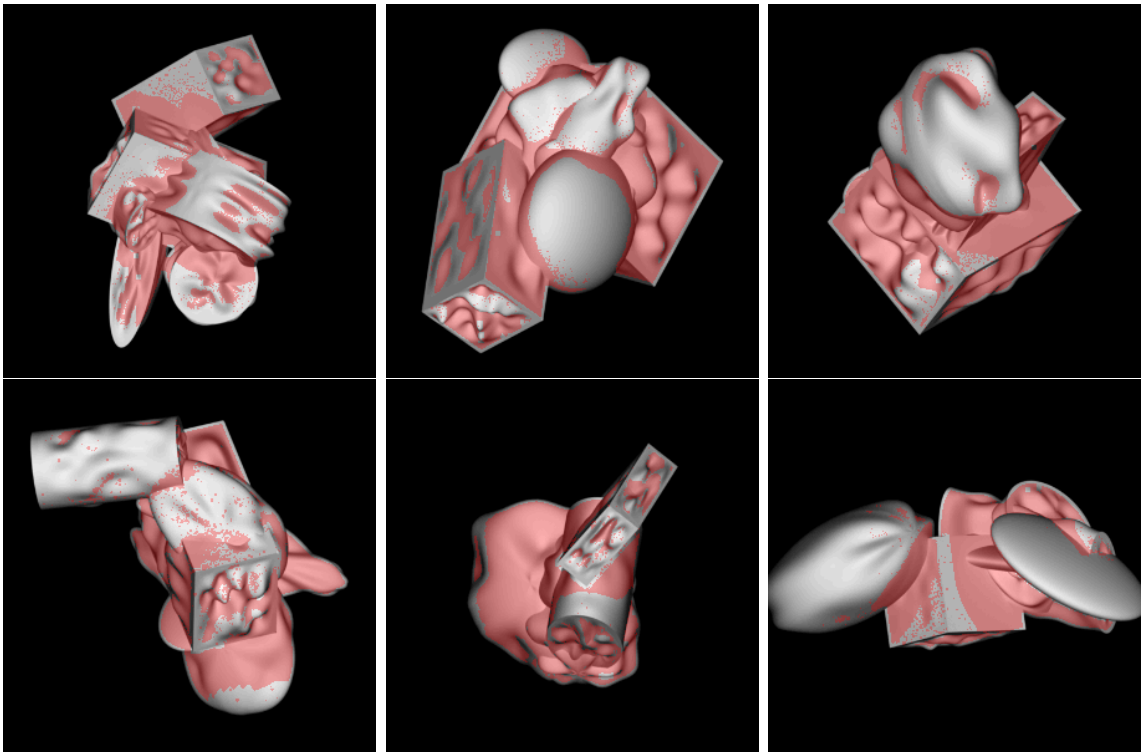


Fig. 4.2 Non-convex appearance exemplar extraction. The red masked pixels show non-convex surfaces (*i.e.*, affected by cast shadows or inter-reflections) extracted by our thresholding under the DiLiGenT's 96 lightings.

at the cost of additional computation.

4.3.3 General appearance exemplar

We build a set of general appearance exemplars by combining convex and non-convex appearance exemplars. The appearance exemplars are constructed from three components; surface normals, BRDFs, and light directions. We create convex appearance exemplars with finely discretized 20001 surface normals and 500 BRDFs, where 100 BRDFs are the MERL BRDFs and the remaining 400 BRDFs are augmented ones as described in Sec. 4.3.1. For non-convex appearance exemplars, we render 500 non-convex shapes with the same 500 BRDFs as the convex appearance exemplars and extract pure non-convex appearance exemplars. Both convex and non-convex appearance exemplars contain uniformly sampled 5000 light directions. Given a target scene,

we use appearance exemplars corresponding to the pre-defined light directions closest to the target scene's ones. Note that we only pull non-convex appearance exemplars affected by global illuminations using binary flags indicating whether appearances are affected by global illuminations or not.

Convexity estimation: A set of general appearance exemplars allows us to estimate convexity of a surface using convexity labels indicating whether an appearance exemplar is affected by global illuminations or not. Since our sets of convex and non-convex appearance exemplars only contains appearances for convex and non-convex surfaces, respectively, the convexity label can be obtained from which set the appearance exemplar belongs to. Thus, DSPS with our general appearance exemplars provides surface normal, BRDF, and convexity estimates. In the following experiments, we show that the surface normal estimation can be improved by applying different photometric stereo methods to convex and non-convex surfaces.

4.4 Experiments

This section describes improvements of DSPS by our general appearance exemplars and comparison to recent photometric stereo methods using synthetic and real-world datasets. We further discuss combining different photometric stereo methods using the knowledge of estimated convexity.

4.4.1 Preparation

CyclesPSTest dataset: CyclesPSTest dataset [4] is a synthetic dataset consisting of three objects (SPHERE, TURTLE, and PAPERBOWL), two types of materials (specular and metallic), and two types of illuminations (17 and 305 lights), yielding 12 scenes. In the following experiment, we use SPHERE and TURTLE scenes for the evaluation since the PAPERBOWL shape is too extreme. Each scene is rendered with spatially varying principled BSDFs [5].

Real-world dataset: We use an existing real-world dataset, the DiLiGenT dataset [3], 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 [4]. In addition to the original dataset, for testing sparse light cases, we prepared 20 datasets, each containing 10 randomly selected images.

Baselines: As baselines we used Lambertian photometric stereo (LPS) [20], model-based method ST14 [27], virtual exemplar-based method HS17 [2], hypothesis-and-test search method HaTS-PS presented in Chapter 2, discrete search photometric stereo (DSPS) with exact and approximated nearest neighbor search DSPS-E and DSPS-A, unsupervised learning (*i.e.*, neural inverse rendering)-based method NIR-PS [46], supervised learning-based methods PX-NET [35], PS-FCN^{+N} [24], WJ20 [47], CNN-PS [4], and SPLINE-Net [38]. 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 chapter. Further, for testing with the CyclesPSTest dataset, although training dataset of PX-NET includes the target material, we list their scores for reference.

Throughout the experiments, we denote DSPS-E and DSPS-A with our general appearance exemplars as DSPS-E+ and DSPS-A+, respectively.

4.4.2 Accuracy of surface normal estimation

We estimated surface normals on synthetic and real datasets containing diverse reflectances and non-convex shapes to confirm that our general appearance exemplars work on diverse scenes.

CyclesPSTest: We conducted quantitative evaluation on the CyclesPSTest dataset. While training dataset of PX-NET is also rendered with the principled BSDFs as with the CyclesPSTest and therefore may include the target materials, their scores are shown as reference.

Table 4.1 shows averages and standard deviations of angular errors on the eight scenes (*i.e.*, two objects, two materials, two lightings). Figures 4.3 and 4.4 illustrate the visual results of angular error maps and estimated surface normal maps. For specular material, our general appearance exemplars remarkably improve the accuracy of DSPS on non-convex surfaces (TURTLE scenes) while maintaining the accuracy on convex surfaces (SPHERE scenes). For metallic material, using our general appearance exemplars slightly degrades the accuracy of DSPS on convex surfaces. Here, we show the estimated convexity (convex or non-convex) and difference in angular errors between DSPS-E and DSPS-E+ for each scene in Fig. 4.5. This visualization indicates that surfaces whose accuracy are degraded by general appearance exemplars tend to be incorrectly estimated as "non-convex", and this trend appears on metallic surfaces more than on specular surfaces. It suggests that a possibility of extremely similar appearances in the sense of ℓ_2 distance can be generated from different surface normals when considering diverse materials and global illuminations, which is a possible reason of the slight degradation on metallic convex surfaces. We further discuss this issue even in the following experiments.

DiLiGenT: We show quantitative results on the real-world dataset DiLiGenT in Tab. 4.2, where we compare our methods with the baseline methods in terms of mean angular error. Figures 4.6 and 4.7 shows visual comparisons between our methods and the baseline methods.

For the scenes with 96 lights, our general appearance exemplars improve the accuracy of DSPS on convex surfaces such as the BALL, the body of CAT, and POT1 owing to the BRDF augmentation and achieve the best score on the BALL object. The accuracy on non-convex surfaces is also largely improved using the general appearance

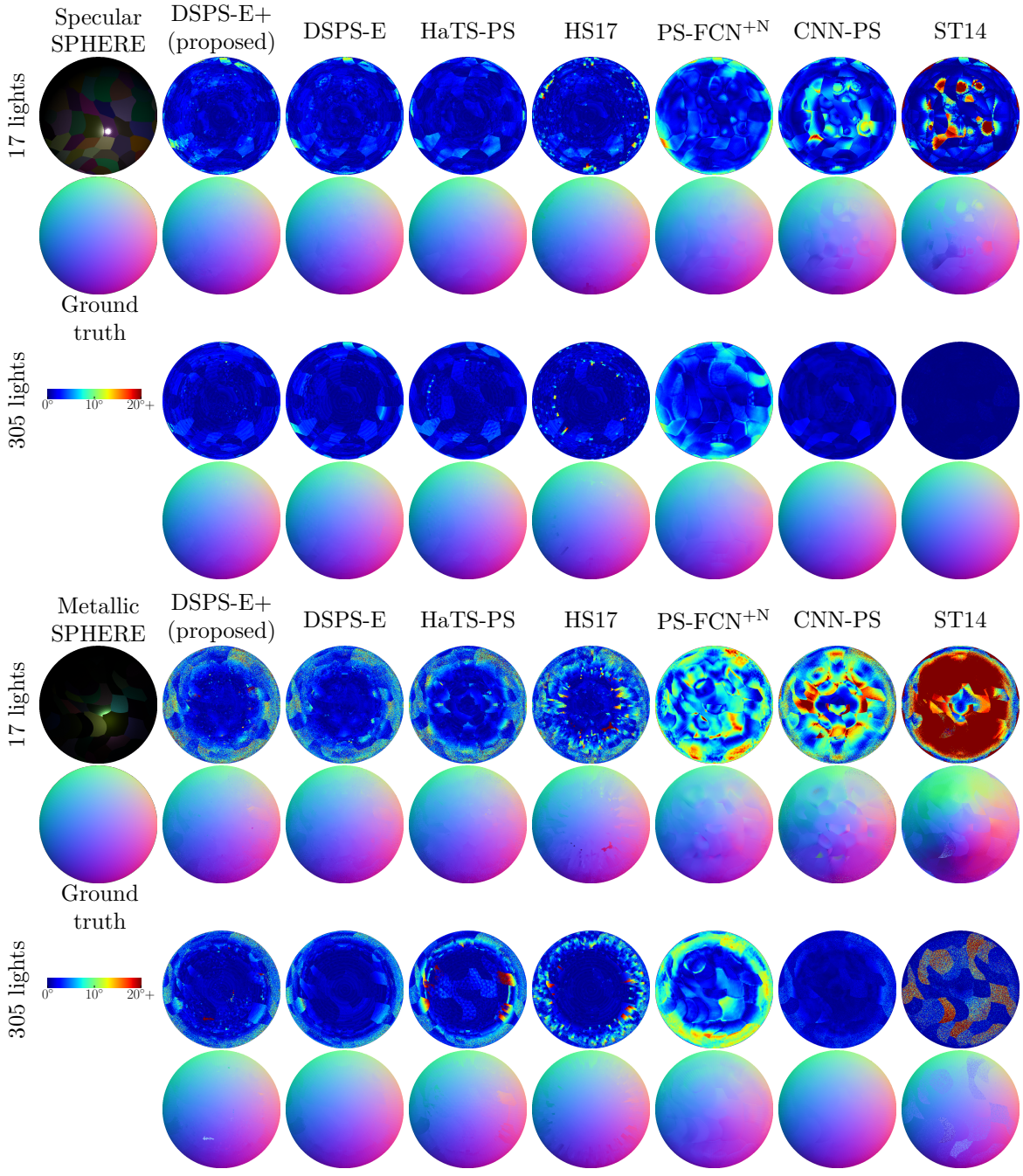


Fig. 4.3 Angular error maps and estimated surface normal maps for the specular and metallic SPHERE scenes in the CyclesPSTest dataset.

Table 4.1 Comparisons on the SPHERE and TURTLE scenes from the CyclesPSTest dataset. Numbers represent averages and standard deviations of angular errors.

		SPHERE				TURTLE			
		17 lights		305 lights		17 lights		305 lights	
		Specular	Metallic	Specular	Metallic	Specular	Metallic	Specular	Metallic
Exemplar-based	DSPS-E+	1.8/2.2	4.1/5.0	1.4/1.8	2.5/3.9	7.9/11.4	15.5/23.0	8.5/12.4	12.1/21.6
	DSPS-A+	1.8/2.2	4.2/5.1	1.5/1.9	2.8/4.3	8.0/11.5	15.5/22.9	8.6/12.5	12.3/21.7
	DSPS-E	1.8/2.1	3.6/4.2	1.5/2.0	2.2/2.8	17.8/21.1	15.7/22.4	13.9/18.7	12.4/21.5
	DSPS-A	1.8/2.1	3.7/4.3	1.5/2.0	2.4/3.0	17.4/20.6	15.8/22.3	13.9/18.5	12.7/21.7
	HaTS-PS	1.6/1.9	4.0/5.1	1.4/1.9	3.2/4.2	17.3/20.0	16.2/22.4	14.1/19.4	13.6/22.4
	HS17	1.7/2.4	4.3/5.6	1.3/2.0	2.6/3.4	17.8/20.4	16.1/22.0	14.1/18.6	12.8/21.4
Learning-based	PX-NET[a]	2.6/2.6	9.5/11.5	0.5/1.5	6.6/13.0	7.4/9.5	16.6/19.5	3.3/6.5	11.4/18.7
	PS-FCN^{+N}	3.1/2.7	6.9/3.8	3.4/2.5	5.7/3.9	10.7/11.6	14.5/16.1	10.4/10.8	12.9/13.9
	WJ20	2.4/2.6	5.6/4.1	3.0/2.7	5.1/3.9	7.8/10.2	13.1/16.2	7.1/8.8	11.5/14.1
	SPLINE-Net	2.7/1.8	4.1/4.2	24.1/16.5	29.3/19.8	6.1/8.1	11.7/19.3	24.8/16.1	33.0/19.8
	CNN-PS[b]	3.3/2.8	9.0/7.7	0.9/1.0	1.4/1.4	9.9/11.7	17.8/18.1	3.2/4.9	5.7/11.9
	NIR-PS[c]	1.6/2.1	10.3/10.8	-	-	13.4/14.8	24.2/18.5	-	-
Model-based	ST14	4.4/8.3	23.5/13.7	0.2/1.5	6.1/10.2	17.7/17.9	32.0/18.4	29.5/26.3	28.5/25.7
	LPS	10.1/7.5	17.4/9.7	10.1/8.3	16.5/9.3	19.0/16.4	26.0/17.7	18.8/16.1	24.3/17.2

[a] Training data of PX-NET include target materials.

[b] CNN-PS is trained with 50-100 lights.

[c] We could not execute NIR-PS with 305 lights as it exceeded the memory of an NVIDIA Quadro RTX 8000 with 48 GB.

exemplars on the most of objects without sacrificing the accuracy on convex surfaces.

Meanwhile, using the general appearance exemplars degrades the accuracy on the COW object. In Fig. 4.8, we illustrate the difference in angular errors between DSPS-E and DSPS-E+ and estimated convexity map for each object in the DiLiGenT dataset. As shown in Fig. 4.8, the angular errors at several pixels on the COW object are increased by using the general appearance exemplars, particularly at pixels that are considered to be convex surfaces but tend to be estimated as "non-convex". Considering the COW object has metallic material at most pixels, this observation is consistent with the result on the CyclesPSTest dataset. It suggests a possibility of ambiguity of photometric stereo problem on metallic surfaces when it considers global illuminations.

For the scenes with 10 lights, the general appearance exemplars promote DSPS to the comparable level with the learning-based methods. Our general appearance

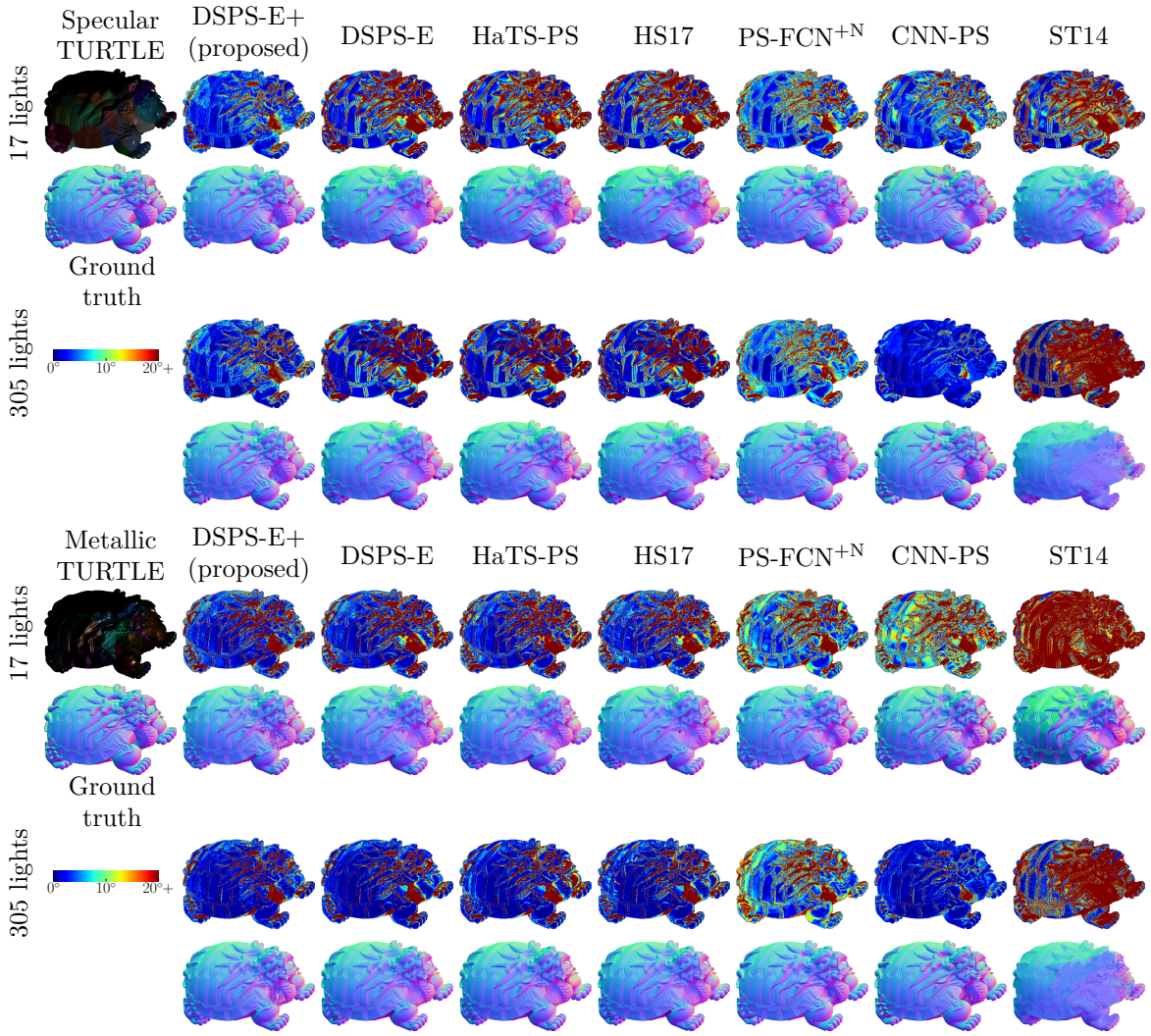


Fig. 4.4 Angular error maps and estimated surface normal maps for the specular and metallic TURTLE scenes in the CyclesPSTest dataset.

exemplars successfully improve the accuracy on the non-convex objects from DSPS even in the few lights case except for the COW object.

4.4.3 Combining photometric stereo methods based on the knowledge of estimated convexity

We demonstrate that surface normal estimation can be improved by taking advantages of different photometric stereo methods based on the knowledge of estimated convexity. Our DSPS-E+ achieves highly accurate surface normal estimation on convex surfaces

Table 4.2 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 datasets with different light distributions.

96 lights

		BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
Exemplar-based	DSPS-E+	1.2	4.6	8.4	4.9	9.2	8.7	16.2	5.6	6.2	12.6	7.8
	DSPS-A+	1.3	4.6	8.5	5.0	9.2	8.9	16.3	5.6	6.4	12.9	7.9
	DSPS-E	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
	DSPS-A	1.4	6.4	14.2	6.8	8.0	11.7	17.5	7.4	7.4	15.3	9.6
	HaTS-PS	1.6	5.9	13.1	6.1	9.2	11.0	18.7	6.6	7.2	15.0	9.4
	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-E+	3.0/0.8	5.3/0.2	10.1/0.6	6.6/0.6	12.5/0.6	11.2/0.7	19.9/0.7	7.2/0.4	9.2/0.9	15.1/0.9	10.0
	DSPS-A+	2.9/0.7	5.4/0.2	10.2/0.6	6.5/0.5	12.4/0.6	11.3/0.7	19.9/0.6	7.1/0.4	9.2/0.9	15.1/0.9	10.0
	DSPS-E	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-A	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
	HaTS-PS	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
	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
	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[b]	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] CNN-PS is trained with 50-100 lights.

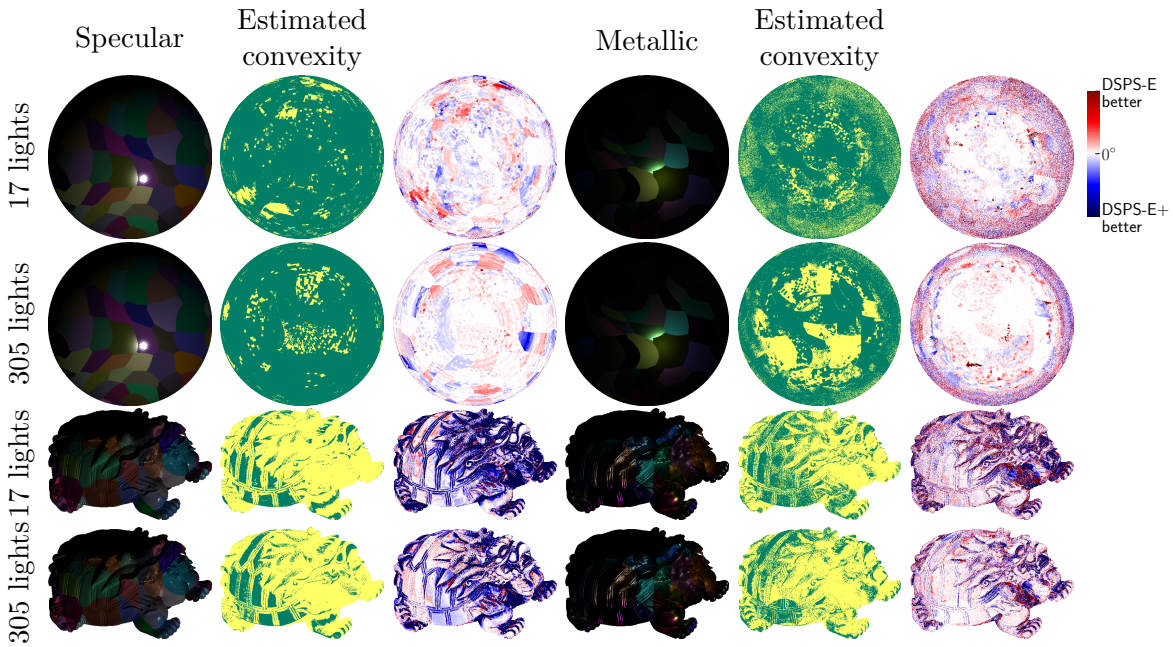


Fig. 4.5 From left to right, an example image, estimated convexity, and difference in angular errors between DSPS-E and DSPS-E+ for each scene. In the estimated convexity maps, green indicates pixels estimated as convex surfaces, and yellow indicates pixels estimated as non-convex surfaces.

with diverse materials. In contrast, the accuracy on non-convex surfaces is slightly inferior to learning-based methods, while it is largely improved by introducing the general appearance exemplars. Therefore, we adopt estimated surface normals of DSPS-E+ and a learning-based method for pixels estimated as "convex" and "non-convex," respectively.

Table 4.3 shows quantitative results on the DiLiGenT dataset with 96 and 10 lights. For 96 and 10 lights cases, we select CNN-PS and PS-FCN^{+N} as learning-based methods. At most objects, the combined results are more accurate than either. The combined results achieve both high accuracy at convex surfaces (*e.g.*, BALL) like DSPS and robustness to global illumination effects like learning-based methods. This result may motivate us to develop a photometric stereo method specific to non-convex surfaces and combine with DSPS in the future.

Fig. 4.6 Angular error maps and estimated surface normal maps for BALL, BEAR, BUDDHA, and CAT objects in the DiLiGenT dataset [3] with all the 96 lights.

Fig. 4.7 Angular error maps and estimated surface normal maps for COW, GOBLET, HARVEST, POT1, POT2, and READING objects in the DiLiGenT dataset [3] with all the 96 lights.

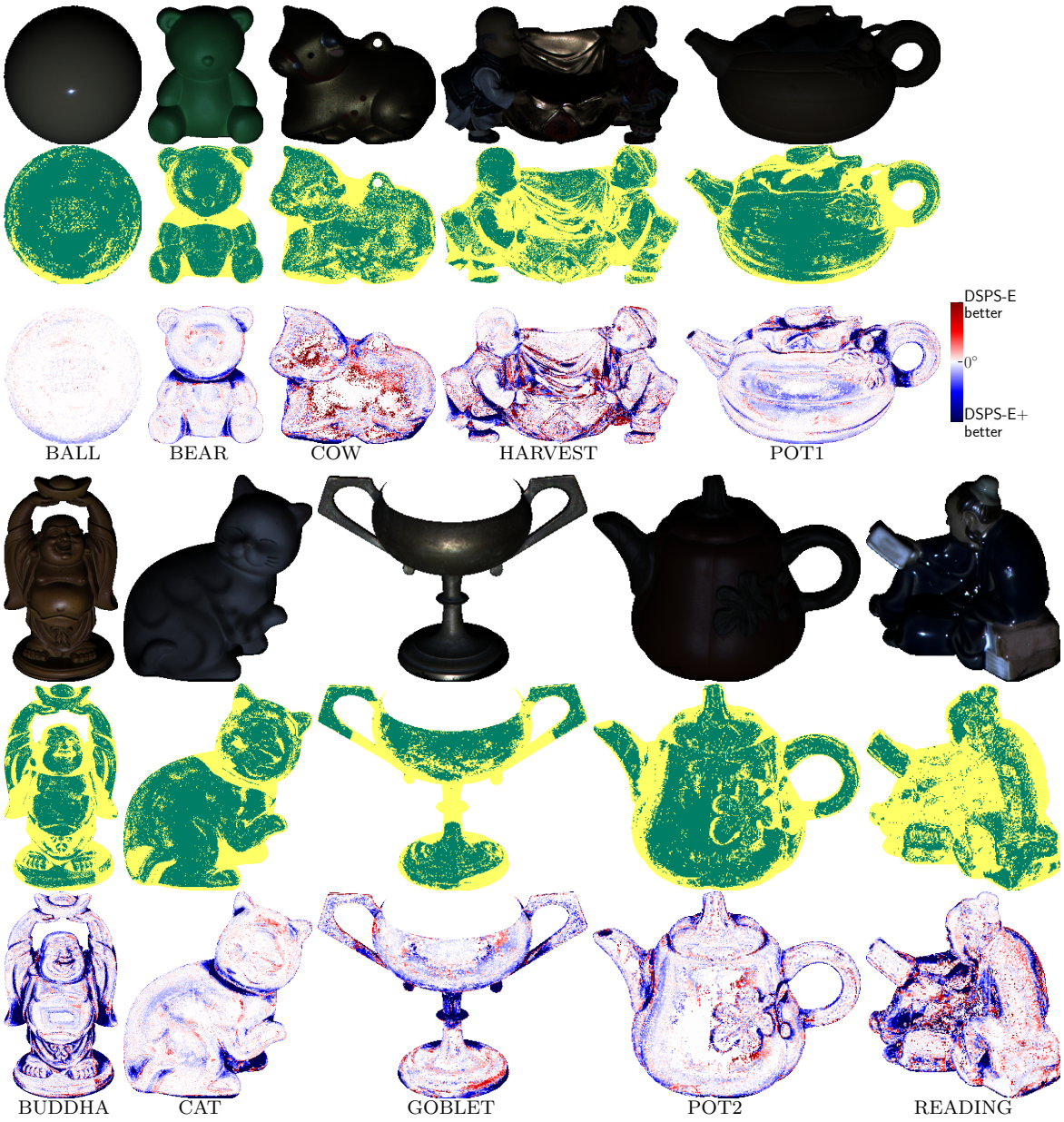


Fig. 4.8 An example image, estimated convexity map, and difference in angular errors between DSPS-E and DSPS-E+ for each object in the DiLiGenT dataset. In the estimated convexity maps, green indicates pixels estimated as convex surfaces, and yellow indicates pixels estimated as non-convex surfaces.

4.4.4 Ablation study of shape and BRDF augmentation

We verify the effectiveness of the additional components for appearance exemplars, augmented BRDFs and non-convex shapes, in the surface normal estimation. We use

Table 4.3 Evaluation of combining different photometric stereo methods using the knowledge of estimated convexity on the DiLiGent dataset. We show the results in 96 lights and 10 lights cases. For the 96 lights case, we adopt estimated surface normals of DSPS-E+ and CNN-PS for pixels estimated as "convex" and "non-convex," respectively. For the 10 lights case, we adopt estimated surface normals of DSPS-E+ and PS-FCN^{+N} for pixels estimated as "convex" and "non-convex," respectively.

96 lights											
	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
DSPS-E+	1.2	4.6	8.4	4.9	9.2	8.7	16.2	5.6	6.2	12.6	7.8
CNN-PS	2.1	4.2	8.1	4.4	7.9	7.4	13.8	5.4	6.4	12.1	7.2
Combined	1.2	4.2	8.1	4.2	6.5	7.4	14.0	5.0	5.9	11.7	6.8

10 lights											
	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
DSPS-E+	3.0	5.3	10.1	6.6	12.5	11.2	19.9	7.2	9.2	15.1	10.0
PS-FCN^{+N}	4.3	6.8	9.7	6.3	12.2	10.5	17.5	7.7	10.0	13.0	9.8
Combined	2.9	5.4	9.3	5.6	9.8	9.8	16.7	6.5	8.4	12.7	8.7

DSPS-E with appearance exemplars constructed from 100 MERL BRDFs and convex shapes as the baseline and show the effect of each BRDF and shape augmentation, respectively.

Table 4.4 shows the ablation study of the BRDF and shape augmentation for appearance exemplars on the DiLiGenT dataset. This study indicates that the BRDF augmentation improves the accuracy on all the objects. While the improvements by the BRDF augmentation are limited in the averaged score since it is effective only for convex surfaces, the effect is more remarkable when combining with the shape augmentation. The shape augmentation largely improves the surface normal estimation on the almost all objects since the original appearance exemplars do not consider global illumination effects at all. However, the shape augmentation introduces a degradation on the COW object. This implies a possibility that considering global illumination effects incurs an ambiguity in surface normal estimation at a certain material under the DiLiGenT lightings. While it is difficult to be concluded here, this observation should be analyzed more in the future.

Table 4.4 Ablation study of the BRDF and shape augmentation for appearance exemplars. Numbers represent mean angular errors on the DiLiGenT dataset. The baseline is DSPS-E with appearance exemplars constructed from 100 MERL BRDFs and convex shapes. We observe accuracies of DSPS-E's surface normal estimation when introducing augmented BRDFs and non-convex shapes to the original appearance exemplars, respectively.

	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
DSPS-E	1.3	6.3	14.0	6.8	7.8	11.5	17.4	7.3	7.4	15.2	9.5
+BRDF aug.	1.3	6.2	13.8	6.6	7.7	11.0	17.2	6.9	7.0	15.1	9.3
+shape aug.	1.3	4.8	8.9	5.2	8.6	9.3	17.3	6.3	6.3	13.0	8.1
DSPS-E+	1.2	4.6	8.4	4.9	9.2	8.7	16.2	5.6	6.2	12.6	7.8

4.4.5 Computation cost

This section examines the computation costs of our DSPS-E+ and DSPS-A+ for surface normal estimation and precomputation (*i.e.*, training for nearest neighbor search method) on CPU and GPU. We measure the computation cost on the DiLiGenT dataset with varying number of lights. For each number of lights, we prepared 10 datasets, each containing randomly selected images from all the 96 images. The computation cost is calculated by taking average over the 10 datasets. 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.

Figure 4.9 shows computation time of surface normal estimation for a single pixel on the CPU and GPU. The estimation costs of DSPS-A and DSPS-A+ are surprisingly comparable owing to the efficient nearest neighbor search algorithm. DSPS-E+ requires around one order of magnitude larger estimation cost than DSPS-E due to the additional appearance exemplars. We consider that this additional cost is worth paying for the improvements of surface normal estimation shown in Tab. 4.2 or 4.3, while DSPS-E is still a strong option to use if most regions of a target scene are convex.

Figure 4.10 shows precomputation time on the CPU and GPU. Both DSPS-E+ and DSPS-A+ naturally require additional precomputation cost to DSPS-E and DSPS-A. However, the precomputation is required only once for a light configuration and takes

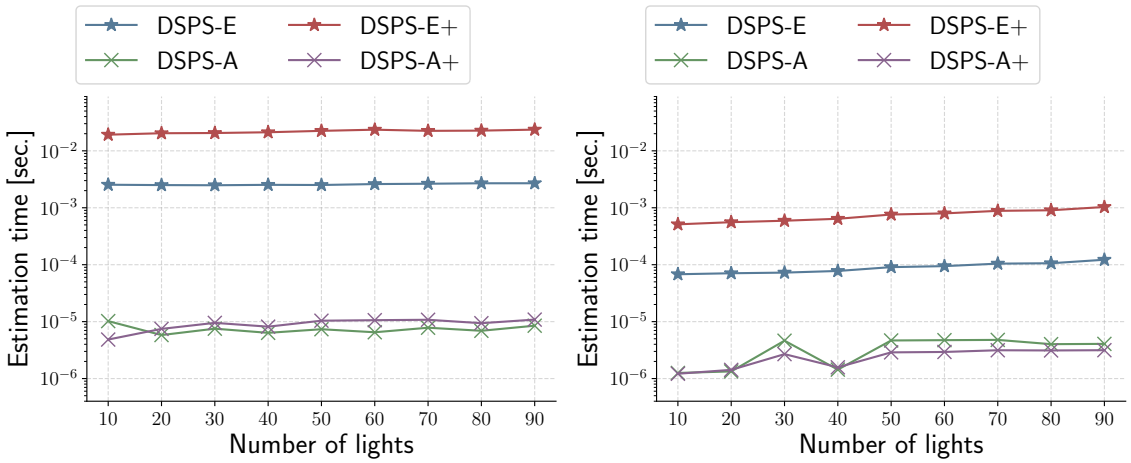


Fig. 4.9 (a) *CPU* estimation time of our methods for a single pixel. (b) *GPU* computation time of our methods for a single pixel.

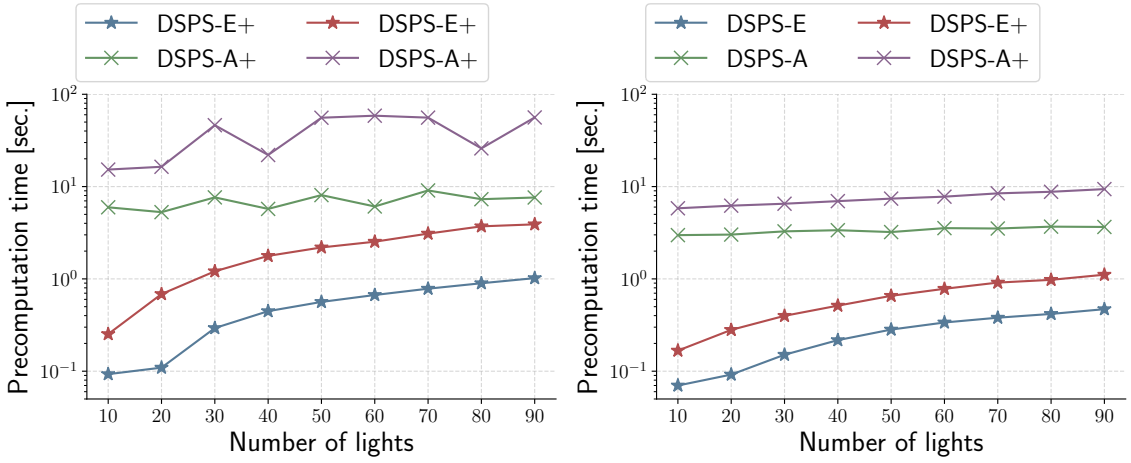


Fig. 4.10 (a) *CPU* precomputation time of our methods. (b) *GPU* precomputation time of our methods.

only several tens seconds at most; therefore, we consider that the precomputation costs are acceptable in most scenarios.

4.5 Conclusion

In this chapter, we have presented general appearance exemplars, which cover both convex and non-convex surfaces and diverse materials beyond available measured BRDFs. Incorporating the general appearance exemplars improves the surface normal estimation of DSPS on both convex and non-convex surfaces. Furthermore, the general appearance exemplars also allow us to estimate a convexity of a surface (convex or non-convex), which leads to further improvement in surface normal estimation by applying different photometric stereo methods to convex and non-convex surfaces.

Experiments on the synthetic and real-world datasets showed that DSPS with our general appearance exemplars achieves the state-of-the-art accuracy for convex surfaces and comparable accuracy to the learning-based methods even for non-convex surfaces. Combining DSPS and a learning-based method based on the estimated convexity successfully takes advantage of each method and produces better surface normal estimates than either.

The experimental results raise a possibility of ambiguity in the photometric stereo problem for general reflectances and shapes. Specifically, in the experiments on the CyclesPS and DiLiGenT dataset, we observe the possibility of ambiguity on metallic-like materials when it considers global illuminations. A theoretical or experimental analysis of the ambiguity is a potential future work, and we believe that it connects to a discussion of the optimal and minimal light configuration for a general photometric stereo problem.

Chapter 5

Conclusion

5.1 Summary

Photometric stereo is a computer vision technique for shape recovery from images, which is able to estimate high-fidelity shape in the form of surface normals. While a traditional photometric stereo method only considers the Lambertian reflectance and direct illumination, in these years, it is time to address the photometric stereo problem for general BRDFs with global illumination effects.

This dissertation focused on the photometric stereo problem for a surface with general BRDFs and global illumination effects. To tackle this problem, we have proposed two novel solutions that perform a discrete search over finely discretized surface normals (and BRDFs). Our discrete search approach successfully achieved an accurate, stable, and efficient surface normal estimation for general BRDFs. Moreover, we have proposed to extend the search space by augmenting BRDFs and introducing global illumination effects, which provide further accuracy on surfaces with more diverse BRDFs and global illuminations. All these efforts greatly improve the quality of photometric stereo in terms of accuracy and stability and provide analyzability owing to their simple and intuitive behaviors.

5.1.1 Photometric stereo for general reflectances by hypothesis-and-test search with scene-independent precomputation

In Chapter 2, we proposed a photometric stereo method based on a hypothesis-and-test strategy, which we call HaTS-PS. HaTS-PS introduces a concept of an appearance tensor that represents a diverse set of appearances constructed from comprehensive surface normals, light directions, and BRDFs. HaTS-PS hypothesizes a surface normal, tests the hypothesized surface normal whether it can explain a target measurements, and repeats these steps for all possible surface normal candidates to find the globally optimal surface normal within the bounds of our objective and discretization. While a naïve hypothesis-and-test search requires a large amount of time, we enabled it in a reasonable amount of time by putting the expensive computation into a scene-independent precomputation step. Experiments on both synthetic and real-world datasets showed that HaTS-PS can accurately and stably estimate surface normals on convex surfaces with diverse materials in a reasonable amount of time.

5.1.2 Photometric stereo for general reflectances by nearest neighbor search over appearance exemplars

In Chapter 3, we proposed the first nearest neighbor search-based photometric stereo, which we call Discrete Search Photometric Stereo (DSPS). In contrast to the HaTS-PS that treats BRDFs in a continuous manner, DSPS treats BRDFs in a discrete manner. Owing to this, the photometric stereo problem can be turned into a well-known nearest neighbor search problem. As a result, DSPS can benefit from advances in fast nearest neighbor search algorithms, leading to highly efficient surface normal estimation with the guarantee of finding the optimal solution within the bounds of the objective function.

Experiments on both synthetic and real-world datasets showed that our DSPS achieves state-of-the-art accuracy on convex surfaces and 100–10000$\times$ acceleration

from existing exemplar-based methods. In addition, we experimentally observed that our DSPS is robust to imaging noise compared to model-based and learning-based methods. Since it is hard to entirely avoid imaging noise in real-world experiments, DSPS is one of the best choices for stable surface normal estimation.

5.1.3 General appearance exemplars for nearest neighbor search-based photometric stereo

In Chapter 4, we proposed a set of general appearance exemplars, which is extended from the existing appearance exemplars that only consider a limited number of BRDFs and convex surfaces. We introduced an additional set of appearance exemplars with augmented BRDFs based on real BRDFs and global illumination effects caused at non-convex surfaces. Incorporating the general appearance exemplars with DSPS greatly improves the surface normal estimation on both convex and non-convex surfaces from DSPS with appearance exemplars only considering a limited number of BRDFs and convex surfaces presented in Chapter 3. The general appearance exemplars also allow us to estimate the convexity of surfaces (convex or non-convex), which enables applying different methods to convex and non-convex surfaces for further accuracy.

Experiments on synthetic and real-world datasets showed that DSPS with our general appearance exemplars achieves the state-of-the-art accuracy on convex surfaces and comparable accuracy to learning-based methods on non-convex surfaces. Combining DSPS and a learning-based method using the knowledge of estimated convexity successfully takes advantage of each method and produces higher accuracy in surface normal estimation than either.

5.2 Future directions

Throughout this dissertation, we have conducted a lot of experiments and observed potential issues as well as advantages of our methods. This dissertation is concluded by discussing several open problems and potential future directions.

5.2.1 Enrichment of measured BRDF

This dissertation presented novel exemplar-based methods to the photometric stereo problem, which require a set of real BRDF data. The photometric stereo task is highly sensitive to corruptions in BRDF data and requires highly accurate and dense BRDF measuring. Indeed, most BRDF databases [74, 75] aim to be used for rendering good-looking scenes under natural environment lightings (*i.e.*, much more light sources than the photometric stereo setup). Therefore, most BRDF databases tolerate noisy and sparse BRDF measurements. In our survey, a BRDF database that satisfies the requirements is only the MERL BRDF database [43] containing 100 BRDFs that is the densest BRDFs and measured with care to avoid noises (*e.g.*, they take 330 high dynamic range pictures and remove lowest and highest 25% of the values to reduce the noise in the measurements). While our methods with the MERL BRDFs can produce promising surface normal estimates for diverse materials, more BRDFs, especially specular or metallic materials, are always preferable and should contribute to further accuracy. A recent measured BRDF database [75] is acquired by an efficient sampling, resulting in sparser BRDF samples compared to the MERL BRDFs, and it did not work well for the photometric stereo task. To sum up, measuring much more BRDF samples by densely sampling or an adaptive sampling that does not degrade the accuracy of photometric stereo is a potential future work.

5.2.2 Nearest neighbor search specific to photometric stereo

Our DSPS uses nearest neighbor search methods developed for a general purpose. Appearance exemplars in the context of photometric stereo have several features such as non-negativity, labels of surface normal and BRDF, and corrupted appearances due to global illumination effects that are preferred to be neglected. A potential future direction is developing a new nearest neighbor search method to achieve more efficient and accurate surface normal estimation by explicitly considering such photometric stereo specific features.

5.2.3 Analysis of optimal light configuration

In recent years, photometric stereo methods including ours are getting more accurate, for instance, on the DiLiGenT benchmark dataset captured under 96 lightings. Although more lightings must lead to further accuracy in surface normal estimation, much more lights are not practical due to physical constraints. We then need to consider the optimal lighting distribution for a fixed number of lights, *e.g.*, 96 lights. A previous work [77] analyzes the optimal configuration in Lambertian photometric stereo; however, there is no theoretical or experimental analysis of the optimal light configuration for general photometric stereo. Therefore, a methodology of analyzing the optimal light configuration in a general setting is still an open problem and should be important in practice.

5.2.4 Extension to multi-view photometric stereo

While this dissertation focuses on single-view photometric stereo, multi-view photometric stereo [78–81] is also actively studied. Compared to the single-view photometric stereo, the multi-view photometric stereo can recover the entire shape of objects, which is desired in several scenarios. Theoretically speaking, the exemplar-based approach can be extended into the multi-view photometric stereo by adding a viewing direction axis to the appearance tensor if the correspondence matching across images taken by different viewing directions can be supposed to be nearly perfect. In practice, erroneous correspondence matching is inevitable in real-world scenarios; therefore, a potential future work is developing an exemplar-based method for the multi-view photometric stereo that is robust to inaccurate correspondence matching.

5.2.5 Photometric stereo in more practical scenarios

Throughout this dissertation, we assume distant lightings and an orthographic projection camera to make the problem tractable. Since these assumptions cannot be strictly held in practice, a relaxation of these assumptions is an important future

work. In fact, photometric stereo methods with nearby lights [82–86] or a perspective camera [87–90] are proposed. However, they generally encounter a non-convex optimization problem when considering non-Lambertian surfaces; therefore, they assume Lambertian or a limited class of BRDF models. While a full-search strategy, as proposed in this dissertation, is effective on such a non-convex optimization problem, the full-search strategy for the near-light or perspective camera setup is expected to require a tremendous amount of computation time.

We also assume a direction and intensity of each light are calibrated in this dissertation. While an accurate light calibration can be performed by sophisticated calibration methods [91–96], it is desired to reduce the calibration efforts in more practical scenarios. Uncalibrated photometric stereo [97–99] realizes surface normal estimation without knowing light directions and even intensities; however, it is challenging to extend the exemplar-based approach to an uncalibrated manner due to its large degree of freedom. Instead, we believe that the exemplar-based approach can be extended to a photometric stereo with special light placements such as symmetric-light and ring-light [100–104] that only requires a prior of relative light placements. This setup only introduces a few degrees of freedom to the exemplar-based approach in addition.

References

[1] R. Harman and V. Lacko, "On decompositional algorithms for uniform sampling from n-spheres and n-balls," *Journal of Multivariate Analysis*, vol. 101, no. 10, pp. 2297–2304, 2010.

[2] Z. Hui and A. C. Sankaranarayanan, "Shape and spatially-varying reflectance estimation from virtual exemplars," *TPAMI*, vol. 39, no. 10, pp. 2060–2073, 2017.

[3] 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," *TPAMI*, vol. 41, no. 2, pp. 271–284, 2019.

[4] S. Ikehata, "CNN-PS: CNN-based photometric stereo for general non-convex surfaces," in *ECCV*, 2018.

[5] B. Burley, "Physically-based shading at Disney," in *SIGGRAPH*, 2012.

[6] M. Oren and S. K. Nayar, "Generalization of the lambertian model and implications for machine vision," *IJCV*, vol. 14, no. 3, pp. 227–251, 1995.

[7] J. F. Blinn, "Models of light reflection for computer synthesized pictures," *ACM SIGGRAPH Computer Graphics*, vol. 11, no. 2, pp. 192–198, 1977.

[8] R. L. Cook and K. E. Torrance, "A reflectance model for computer graphics," *TOG*, vol. 1, no. 1, pp. 7–24, 1982.

[9] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *SIGGRAPH*, 2000.

[10] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, P. E. Debevec *et al.*, "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination." *Rendering Techniques*, vol. 2007, no. 9, p. 10, 2007.

[11] C.-F. Chabert, P. Einarsson, A. Jones, B. Lamond, W.-C. Ma, S. Sylwan, T. Hawkins, and P. Debevec, "Relighting human locomotion with flowed reflectance fields," in *SIGGRAPH Sketches*, 2006.

[12] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011.

[13] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar, "Block-nerf: Scalable large scene neural view synthesis," in *CVPR*, 2022.

[14] Z. Yang, Y. Chai, D. Anguelov, Y. Zhou, P. Sun, D. Erhan, S. Rafferty, and H. Kretzschmar, "Surfelgan: Synthesizing realistic sensor data for autonomous driving," in *CVPR*, 2020.

[15] L. Quan, P. Tan, G. Zeng, L. Yuan, J. Wang, and S. B. Kang, "Image-based plant modeling," in *SIGGRAPH*, 2006.

[16] T. Isokane, F. Okura, A. Ide, Y. Matsushita, and Y. Yagi, "Probabilistic plant modeling via multi-view image-to-image translation," in *CVPR*, 2018.

[17] D. Marr and T. Poggio, "A computational theory of human stereo vision," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 204, no. 1156, pp. 301–328, 1979.

[18] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d," in *SIGGRAPH*, 2006.

[19] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *CVPR*, 2016.

[20] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[21] W. M. Silver, "Determining shape and reflectance using multiple images," Master's thesis, Massachusetts Institute of Technology, 1980.

[22] 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.

[23] A. S. Georghiades, "Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo." in *ICCV*, 2003.

[24] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong, "Deep photometric stereo for non-Lambertian surfaces," *TPAMI*, vol. 44, no. 1, pp. 129–142, 2020.

[25] N. Alldrin, T. Zickler, and D. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *CVPR*, 2008.

[26] F. Lu, X. Chen, I. Sato, and Y. Sato, "SymPS: BRDF symmetry guided photometric stereo for shape and light source estimation," *TPAMI*, vol. 40, no. 1, pp. 221–234, 2018.

[27] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi, "Bi-polynomial modeling of low-frequency reflectances," *TPAMI*, vol. 36, no. 6, pp. 1078–1091, 2014.

[28] S. Ikehata and K. Aizawa, "Photometric stereo using constrained bivariate regression for general isotropic surfaces," in *CVPR*, 2014.

[29] 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 *ICCV*, 2017.

[30] A. Hertzmann and S. M. Seitz, "Example-based photometric stereo: Shape reconstruction with general, varying BRDFs," *TPAMI*, vol. 27, no. 8, pp. 1254–1264, 2005.

[31] H.-S. Chung and J. Jia, "Efficient photometric stereo on glossy surfaces with wide specular lobes," in *CVPR*, 2008.

[32] T. Chen, M. Goesele, and H.-P. Seidel, "Mesostructure from specularity," in *CVPR*, 2006.

[33] 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," *TPAMI*, vol. 37, no. 4, pp. 890–897, 2014.

[34] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita, "Deep photometric stereo networks for determining surface normal and reflectances," *TPAMI*, 2020.

[35] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla, "PX-NET: Simple and efficient pixel-wise training of photometric stereo networks," in *ICCV*, 2021.

[36] S. Ikehata, "PS-Transformer: Learning sparse photometric stereo network using self-attention mechanism," in *BMVC*, 2021.

[37] J. Li, A. Robles-Kelly, S. You, and Y. Matsushita, "Learning to minify photometric stereo," in *CVPR*, 2019.

[38] 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 *ICCV*, 2019.

[39] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita, "Deep photometric stereo network," in *ICCV Workshops*, 2017.

[40] G. Chen, K. Han, and K.-Y. K. Wong, "PS-FCN: A flexible learning framework for photometric stereo," in *ECCV*, 2018.

[41] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *CVPR*, 2011.

[42] O. Wiles and A. Zisserman, "SilNet: Single-and multi-view reconstruction by learning from silhouettes," *BMVC*, 2017.

[43] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *TOG*, vol. 22, no. 3, pp. 759–769, 2003.

[44] B. K. Horn and K. Ikeuchi, "The mechanical manipulation of randomly oriented parts," *Scientific American*, vol. 251, no. 2, pp. 100–113, 1984.

[45] N. Ratner and Y. Y. Schechner, "Illumination multiplexing within fundamental limits," in *CVPR*, 2007.

[46] T. Taniai and T. Maehara, "Neural inverse rendering for general reflectance photometric stereo," in *ICML*, 2018.

[47] X. Wang, Z. Jian, and M. Ren, "Non-lambertian photometric stereo network based on inverse reflectance model with collocated light," *TIP*, vol. 29, pp. 6032–6042, 2020.

[48] 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.

[49] K. Enomoto, M. Waechter, K. N. Kutulakos, and Y. Matsushita, "Photometric stereo via discrete hypothesis-and-test search," in *CVPR*, 2020.

[50] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[51] A. Beygelzimer, S. Kakade, and J. Langford, "Cover trees for nearest neighbor," in *ICML*, 2006.

[52] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *TPAMI*, vol. 33, no. 1, pp. 117–128, 2010.

[53] Y. A. Malkov and D. A. Yashunin, "Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs," *TPAMI*, vol. 42, no. 4, pp. 824–836, 2020.

[54] Z. Yao, K. Li, Y. Fu, H. Hu, and B. Shi, "GPS-Net: Graph-based photometric stereo network," *NIPS*, 2020.

[55] Y. LeCun, C. Cortes, and C. Burges, "MNIST handwritten digit database," *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, vol. 2, 2010.

[56] A. Krizhevsky, "Learning multiple layers of features from tiny images," Tech. Rep., 2009.

[57] 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," *IJCV*, vol. 115, no. 3, pp. 211–252, 2015.

[58] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *CVPR*, 2019.

[59] Y. Wang, W. Gan, J. Yang, W. Wu, and J. Yan, "Dynamic curriculum learning for imbalanced data classification," in *ICCV*, 2019.

[60] S. M. Omohundro, *Five balltree construction algorithms*. International Computer Science Institute Berkeley, 1989.

[61] A. Guttman, "R-trees: A dynamic index structure for spatial searching," in *International Conference on Management of Data*. ACM, 1984.

[62] P. Ciaccia, M. Patella, and P. Zezula, "M-tree: An efficient access method for similarity search in metric spaces," in *International Conference on Very Large Data Bases*, 1997.

[63] 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.

[64] Y. Hwang, B. Han, and H.-K. Ahn, "A fast nearest neighbor search algorithm by nonlinear embedding," in *CVPR*, 2012.

[65] J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," *arXiv preprint arXiv:1702.08734*, 2017.

[66] A. Andoni, P. Indyk, T. Laarhoven, I. Razenshteyn, and L. Schmidt, "Practical and optimal LSH for angular distance," *arXiv preprint arXiv:1509.02897*, 2015.

[67] B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, no. 6, pp. 311–317, 1975.

[68] G. J. Ward, "Measuring and modeling anisotropic reflection," *ACM SIGGRAPH Computer Graphics*, vol. 26, no. 2, pp. 265–272, 1992.

[69] E. P. Lafortune, S.-C. Foo, K. E. Torrance, and D. P. Greenberg, "Non-linear approximation of reflectance functions," in *Proceedings of conference on Computer graphics and interactive techniques*, 1997, pp. 117–126.

[70] W. Jakob, M. Hašan, L.-Q. Yan, J. Lawrence, R. Ramamoorthi, and S. Marschner, "Discrete stochastic microfacet models," *TOG*, vol. 33, no. 4, pp. 1–10, 2014.

[71] L.-Q. Yan, M. Hašan, W. Jakob, J. Lawrence, S. Marschner, and R. Ramamoorthi, "Rendering glints on high-resolution normal-mapped specular surfaces," *TOG*, vol. 33, no. 4, pp. 1–9, 2014.

[72] M. Ashikmin, S. Premože, and P. Shirley, "A microfacet-based brdf generator," in *SIGGRAPH*, 2000.

[73] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, "Microfacet models for refraction through rough surfaces," in *Proceedings of Eurographics conference on Rendering Techniques*, 2007.

[74] J. Filip and R. Vávra, "Template-based sampling of anisotropic brdfs," in *Computer Graphics Forum*, 2014.

[75] J. Dupuy and W. Jakob, "An adaptive parameterization for efficient material acquisition and rendering," *TOG*, vol. 37, no. 6, pp. 1–14, 2018.

[76] Z. Xu, K. Sunkavalli, S. Hadap, and R. Ramamoorthi, "Deep image-based relighting from optimal sparse samples," *TOG*, vol. 37, no. 4, p. 126, 2018.

[77] O. Drbohlav and M. Chantler, "On optimal light configurations in photometric stereo," in *ICCV*, 2005.

[78] C. H. Esteban, G. Vogiatzis, and R. Cipolla, "Multiview photometric stereo," *TPAMI*, vol. 30, no. 3, pp. 548–554, 2008.

[79] M. Li, Z. Zhou, Z. Wu, B. Shi, C. Diao, and P. Tan, "Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials," *TIP*, vol. 29, pp. 4159–4173, 2020.

[80] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. S. Kweon, "Robust multiview photometric stereo using planar mesh parameterization," *TPAMI*, vol. 39, no. 8, pp. 1591–1604, 2016.

[81] B. Kaya, S. Kumar, F. Sarno, V. Ferrari, and L. Van Gool, "Neural radiance fields approach to deep multi-view photometric stereo," in *WACV*, 2022.

[82] Y. Iwahori, H. Sugie, and N. Ishii, "Reconstructing shape from shading images under point light source illumination," in *ICPR*, 1990.

[83] R. Mecca, A. Wetzler, A. M. Bruckstein, and R. Kimmel, "Near field photometric stereo with point light sources," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2732–2770, 2014.

[84] Y. Quéau, T. Wu, F. Lauze, J.-D. Durou, and D. Cremers, "A non-convex variational approach to photometric stereo under inaccurate lighting," in *CVPR*, 2017.

[85] H. Santo, M. Waechter, and Y. Matsushita, "Deep near-light photometric stereo for spatially varying reflectances," in *ECCV*, 2020.

[86] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla, "A cnn based approach for the near-field photometric stereo problem," in *BMVC*, 2020.

[87] A. Tankus and N. Kiryati, "Photometric stereo under perspective projection," in *ICCV*, 2005.

[88] T. Papadhimitri and P. Favaro, "A new perspective on uncalibrated photometric stereo," in *CVPR*, 2013.

[89] R. Mecca, A. Tankus, A. Wetzler, and A. M. Bruckstein, "A direct differential approach to photometric stereo with perspective viewing," *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 579–612, 2014.

[90] M. Li, C.-y. Diao, D.-q. Xu, W. Xing, and D.-m. Lu, "A non-lambertian photometric stereo under perspective projection," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 8, pp. 1191–1205, 2020.

[91] H.-L. Shen and Y. Cheng, "Calibrating light sources by using a planar mirror," *Journal of Electronic Imaging*, vol. 20, no. 1, p. 013002, 2011.

[92] J. Ackermann, S. Fuhrmann, and M. Goesele, "Geometric point light source calibration," in *Vision, Modeling and Visualization*, 2013.

[93] M. W. Powell, S. Sarkar, and D. Goldgof, "A simple strategy for calibrating the geometry of light sources," *TPAMI*, vol. 23, no. 9, pp. 1022–1027, 2001.

[94] W. Zhou and C. Kambhamettu, "Estimation of illuminant direction and intensity of multiple light sources," in *ECCV*, 2002.

[95] T. Takai, A. Maki, K. Niinuma, and T. Matsuyama, "Difference sphere: An approach to near light source estimation," *Computer Vision and Image Understanding*, vol. 113, no. 9, pp. 966–978, 2009.

[96] H. Santo, M. Waechter, W.-Y. Lin, Y. Sugano, and Y. Matsushita, "Light structure from pin motion: geometric point light source calibration," *IJCV*, vol. 128, no. 7, pp. 1889–1912, 2020.

[97] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "Uncalibrated photometric stereo for unknown isotropic reflectances," in *CVPR*, 2013.

[98] B. Kaya, S. Kumar, C. Oliveira, V. Ferrari, and L. Van Gool, "Uncalibrated neural inverse rendering for photometric stereo of general surfaces," in *CVPR*, 2021.

[99] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong, "Self-calibrating deep photometric stereo networks," in *CVPR*, 2019.

[100] K. Minami, H. Santo, F. Okura, and Y. Matsushita, "Symmetric-light photometric stereo," in *WACV*, 2022.

[101] M. Chandraker, J. Bai, and R. Ramamoorthi, "On differential photometric reconstruction for unknown, isotropic brdfs," *TPAMI*, vol. 35, no. 12, pp. 2941–2955, 2012.

[102] N. G. Alldrin and D. J. Kriegman, "Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach," in *ICCV*. IEEE, 2007.

[103] P. Tan and T. Zickler, "A projective framework for radiometric image analysis," in *CVPR*, 2009.

[104] Z. Zhou and P. Tan, "Ring-light photometric stereo," in *ECCV*, 2010.