DiLiGenT10^2: A Photometric Stereo Benchmark Dataset with Controlled Shape and Material Variation

Jieji Ren[1†], Feishi Wang[2], Jiahao Zhang[2], Qian Zheng[3], Mingjun Ren[1*], Boxin Shi[2,4,5*]

[1]School of Mechanical Engineering, Shanghai Jiao Tong University

[2]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[3]College of Computer Science and Technology, Zhejiang University

[4]Institute for Artificial Intelligence, Peking University, [5]Beijing Academy of Artificial Intelligence

{jiejiren, renmj}@sjtu.edu.cn, {wangfeishi, 1900013264, shiboxin}@pku.edu.cn, qianzheng@zju.edu.cn

Abstract

Evaluating photometric stereo using real-world dataset is important yet difficult. Existing datasets are insufficient due to their limited scale and random distributions in shape and material. This paper presents a new real-world photometric stereo dataset with "ground truth" normal maps, which is 10 times larger than the widely adopted one. More importantly, we propose to control the shape and material variations by fabricating objects from CAD models with carefully selected materials, covering typical aspects of reflectance properties that are distinctive for evaluating photometric stereo methods. By benchmarking recent photometric stereo methods using these 100 sets of images, with a special focus on recent learning based solutions, a 10×10 *shape-material error distribution matrix is visualized to depict a "portrait" for each evaluated method. From such comprehensive analysis, open problems in this field are discussed. To inspire future research, this dataset is available at* https://photometricstereo.github.io.

1. Introduction

Photometric stereo recovers pixel-wise surface normals of an object by observing it under different lighting conditions [46]. Compared with other 3D modeling approaches, photometric stereo can produce 3D (strictly speaking 2.5D) information with abundant details due to the per-wise recovery. Photometric stereo is an ill-posed problem because of the unknown reflectance and lighting conditions of the target object. There is a long research history to incorporate different assumptions on shapes (*e.g.*, a convex shape to avoid cast shadow) and reflectance (*e.g.*, an isotropic BRDF (Bidirectional Reflectance Distribution Function) to apply reflectance symmetry prior) to make the problem tractable.

To figure out the effectiveness of various assumptions for solving real-world photometric stereo problems, the DiLiGenT [37] benchmark dataset has been built and successfully applied to conduct quantitative evaluations for photometric stereo methods published before the year 2016 [1].

The development of deep learning based photometric stereo methods demonstrates performance boost (over those reported in 2016) according to the constantly refreshing benchmark records reported by papers using DiLiGenT as a benchmark [37]. However, as a dataset originally proposed in the era that non-learning based photometric stereo methods dominate, DiLiGenT [37] shows deficiencies in two aspects for evaluating recent photometric stereo methods. First, as learning based methods are often trained by large-scale data with various shapes and reflectance, the small data scale (10 objects) of DiLiGenT [37] can be easily overfitted, which brings bottleneck for model evaluation and improvement. Second, as learning based methods are less interpretable, the in-depth analysis to understanding a pre-trained neural photometric stereo model can hardly be achieved, as DiLiGenT [37] uses randomly picked up objects (shapes and materials) from daily life.

To address these challenges, this paper builds a new photometric stereo benchmark dataset[2], named **DiLiGenT10^2**, which contains 100 objects of 10 shapes multiplied by 10 materials[3], as shown in Fig. 1. These 100 objects are fabricated with the high-precise computer numerical control

[*]Corresponding authors.

[†]Part of this work was finished as a remote intern at Peking University.

[1]The earlier version of [37] was published in CVPR 2016 [39]

[2]"DiLiGenT" stands for benchmarking photometric stereo under Directional Lighting, General reflectance, with "ground Truth" shape. Since our dataset follows the same setup assumption for evaluating photometric stereo methods, we borrow the abbreviation of DiLiGenT as prefix.

[3]The abbreviations of materials and shapes are listed below: For materials, polyoxymethylene (POM), polypropylene (PP), polyamide (NYLON); polyvinyl chloride (PVC), acrylonitrile butadiene styrene (ABS), phenol formaldehyde resin (BAKELITE), aluminium alloy 6061 (AL), brass H59 (CU), die steel P20 (STEEL), and polymethyl methacrylate (ACRYLIC); for the polyhedron shapes, SQUARE is rhombicuboctahedron, PENTAGON is the upper half of icosahedron, and HEXAGON is truncated icosahedron.

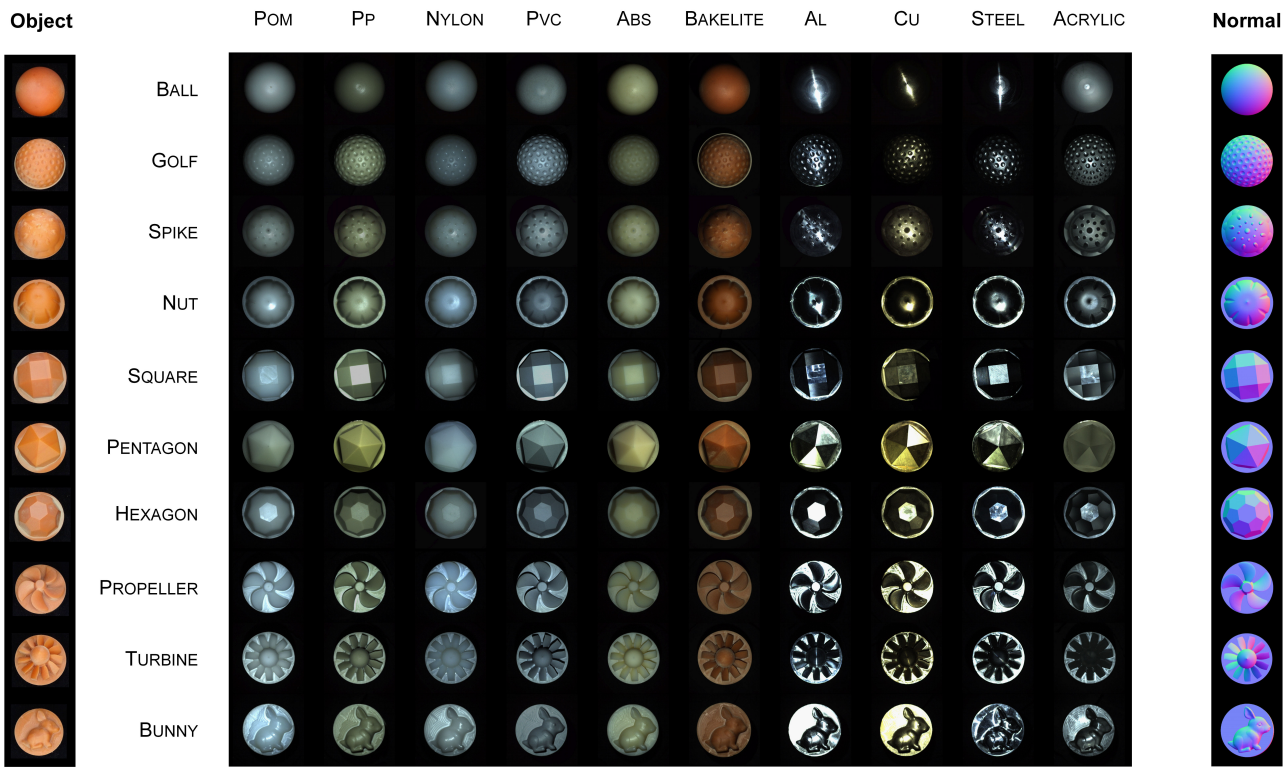

Figure 1. 100 sets of data in the DiLiGenT10^2 dataset represented as a **shape-material** (10×10) matrix. For shape variation (row by row), 10 materials are divided into three groups: **Sphere group** contains smooth shapes of a sphere – BALL and its variants with different types of details – GOLF, SPIKE, and NUT; **Polyhedron group** contains piece-wise shapes of multiple directional faces – SQUARE, PENTAGON, and HEXAGON; **General-shape group** contains objects from industry machining and daily life with discontinuous surfaces – PROPELLER, TURBINE, and BUNNY. For material variation (column by column), 10 materials are also divided into three groups: **Isotropic group** with diffuse and moderately specular materials – POM, PP, NYLON, PVC, ABS, and BAKELITE; **Anisotropic group** contains metallic materials with anisotropic reflectance – AL, CU, and STEEL; **Challenging group** contains a translucent material – ACRYLIC. The leftmost column shows objects made by BAKELITE captured under natural illumination (and a slightly slanted viewpoint) to provide an intuitive reference (not used for computation). The rightmost column shows "ground truth" surface normal map corresponding to the shape in each row.

(CNC) machining process to guarantee the consistency of shape across diverse materials. As a result, we can control the shape variation from simple to complex, and the material variation with different levels of difficulties for photometric stereo. CAD models used for fabricating these objects are registered to the image coordinate for calibrating "ground truth" surface normal maps. To the best of our knowledge, DiLiGenT10^2 is the largest real-world photometric stereo benchmark dataset until now.

We apply DiLiGenT10^2 dataset to conduct a comprehensive benchmark evaluation for up-to-date photometric stereo methods, with a special focus on learning based solutions. With the large-scale size and controlled shape and material variations, DiLiGenT10^2 provides an insightful comparison to distinctively reveal pros and cons for photometric stereo methods in the era that learning based solutions dominate. In summary, this paper contributes the largest real-world photometric stereo benchmark dataset with the following new characteristics:

- $10\times$ larger scale than the DiLiGenT dataset [37];
- with controlled shape (high-precision CAD models as "ground truth") and materials (carefully fabricated materials covering isotropic (diffuse and specular), anisotropic, and translucent reflectance); and
- up-to-date benchmark evaluation results (the most comprehensive one for learning based methods until now) presented as shape-material error matrices with inspiration to future work.

2. Related Work

2.1. Photometric Stereo Methods

Representative non-learning based photometric stereo methods (until the year 2016) have been comprehensively discussed and evaluated in [37]. Despite that there are non-learning based methods being proposed after the survey of [37], *e.g.*, a hypothesis-and-test search solution for general reflectance [16], a patch-wise solution for uncal-

ibrated photometric stereo under natural lighting [18], a majority of recent photometric stereo methods adopt deep learning, so we focus on reviewing learning based solutions to be evaluated using the new dataset. Non-typical photometric stereo image capturing setups such as multi-view photometric stereo [30] and multi-spectral photometric stereo [19, 24, 26] are beyond the scope.

We follow the categorization proposed by Shi *et al.* [36] and Zheng *et al.* [51], which divides learning based methods into all-pixel and per-pixel branches. All-pixel methods focus on estimating normal by aggregating global features from different images [13], while per-pixel methods take the input in the form of intensity profile [35] or observation map [21] and focus on the modeling of reflectance properties (*e.g.*, isotropy [50]) or global illumination effects [21, 29].

Recently, increasing attention has been paid to jointly modeling the reflectance and shape for learning based photometric stereo. GPS-Net [49] adopted graph based feature extraction to model local reflectance, and simultaneously utilized CNN based regression networks for exploring global shape information to balance the advantages between all-pixel and per-pixel solutions. PX-NET [31] augmented observation maps for allowing the networks to learn global rendering effects. Attention-weighted loss was introduced to deal with areas with delicate structures [25], and dual-regression considering consistency between observed images and reconstructed images was introduced to form a closed-loop that provided additional supervision [23].

Deep learning has also been demonstrated to be effective for uncalibrated photometric stereo. SDPS-Net [11] and its extensions [12, 14] utilized a lighting estimation network to calibrate directional lighting intensity and direction first before feeding them into a normal estimation network. Kaya *et al.* [28] proposed an uncalibrated neural inverse rendering approach to estimate lighting directions from images and calculated the surface normal by optimizing an image reconstruction loss.

From the survey above, we can tell that modeling reflectance characteristics of diverse materials still plays an important role in making photometric stereo accurate for complex shapes. A dataset with controlled shape and material variation is able to decouple shape and material's mutual influence for better understanding what has been learned and what needs to be solved for learning based photometric stereo.

2.2. Photometric Stereo Datasets

To drive the modeling process of learning based photometric stereo methods, various synthetic and real-world datasets have been proposed. Synthetic datasets are created by forwardly rendering the photometric image formation model, *i.e.*, applying BRDF models to 3D shapes (which naturally provide ground truth normals) and generating images under varying directional lightings. The first deep photometric stereo method [35] took 10 blobby shapes [22] and applied measured MERL BRDFs [32] to generate a large-scale synthetic dataset for training. After that, more complex scanned real-world 3D shapes [45] and large-scale shape datasets [10], were combined with more controllable parametric BRDFs [7] to render training datasets in follow-up works [13, 14, 21]. Besides, synthetic photometric stereo datasets were created for cultural heritage restoration [15], coins measurements [5], and panoramic cameras [8, 9]. These data come with truly accurate directional lighting and surface normal maps, and it is cheap to create them at a large scale, which greatly promotes the development of learning based photometric stereo. However, synthetic dataset can never faithfully model complex reflectance from real-world materials and their interactions with shapes during the real imaging pipeline. Real-world datasets are always necessary for evaluating whether a photometric stereo method acts effectively in practice.

Some authors released real-world datasets they captured when evaluating their own methods. For example, the Gourd&Apple dataset [2] provided high-quality HDR images for objects with general isotropic BRDFs but without providing reference normal maps, and the Harvard dataset [48] included several objects with diffuse reflectance and reference normal maps calculated using least squares based photometric stereo. DiLiGenT [37] was the first real-word dataset that was specially designed for benchmarking photometric stereo under directional lighting for objects with general BRDFs; it utilized a high-precision scanner to measure the 3D mesh of target objects and registered them to the photometric images to calculate "ground truth" normal. After that, various datasets were proposed for evaluating different aspects of photometric stereo methods, such as multi-views [30], near-field [33], inter-reflections [28], outdoor environments [1, 20, 27], large-scale 3D reconstruction [43], and for specific applications, such as plants [4] and faces [6, 44]. These datasets provide diverse shapes and materials for quantitatively evaluating photometric stereo methods, however, due to the expensive and laborious procedures for creating real-world photometric stereo datasets (in general, HDR images with highly accurate lighting and geometry calibrations are required), the scale of real-world datasets are still limited, which may lead the networks to over-fitted results and impede the evaluation and improvement of learning based photometric stereo methods. Moreover, the scanned shape sometimes cannot include delicate structures of objects with sufficiently fine details for evaluating photometric stereo as discussed in [37].

We summarize attributes of representative synthetic and real photometric stereo datasets in Table 1. Compared with existing datasets, the unique characteristics of

Table 1. Summary of photometric stereo datasets (ordered by scales). Type: real-world or synthetic dataset. Material: controlled (fabricated with carefully selected materials) or uncontrolled (randomly picked up from daily objects) for real data; BRDF models for synthetic data. Ground truth: from CAD/Scanned models with registration (+Reg), photometric stereo (PS), or synthetic shapes (Mesh). Number of shapes, lights, views, and sets (one set means a sequence of photometric stereo images under varying lighting conditions used for computation).

Dataset	Ground truth	Type	Material	# Shapes	# Lights	# Views	# Sets
DiLiGenT10^2 (Ours)	CAD+Reg	Real	Controlled	10	100	1	100
DiLiGenT-MV [30]	Scan+Reg	Real	Uncontrolled	5	96	20	100
LUCES [33]	Scan+Reg	Real	Uncontrolled	14	52	1	14
DiLiGenT [37]	Scan+Reg	Real	Uncontrolled	10	96	1	10
Harvard [48]	PS	Real	Uncontrolled	7	20	1	7
ETHz [28]	Scan+Reg	Real	Uncontrolled	3	260	1	3
Gourd&Apple [2]	–	Real	Uncontrolled	2	102/112	1	2
SculpturePS [13]	Mesh	Synthetic	MERL	8	64	1,296	59,292
BlobbyPS [35]	Mesh	Synthetic	MERL	8	96	1	800
CyclePS-Train [21]	Mesh	Synthetic	Disney	15	1,280/1,000	1	45

DiLiGenT10^2 are demonstrated. We hope such a large-scale real-world photometric stereo dataset with diverse materials and complex shapes under controllable variations not only increases the scale of real-world photometric stereo dataset, but also inspires the development of future learning based methods through a more systematical benchmark evaluation.

3. DiLiGen10^2 Dataset

DiLiGenT10^2 dataset contains 100 objects (sets), which consist of 10 gradually changed shapes and 10 carefully selected materials. Each set provides 100 observed images under differently calibrated light directions and corresponding "ground truth" normal map, as shown in Fig. 1.

Objects in DiLiGenT10^2 compose a 10×10 shape-material matrix. Along the shape dimension, there are 10 shapes changing from simple to complex with different curvatures and structures, to provide rich multi-scale features causing diverse shading, shadow, inter-reflection effects. Along the material dimension, there are 10 materials changing from diffuse or moderately specular plastic to highly specular metal, even to a complex translucent surface, based on statistics from manufacturing and daily life [3, 41], to cover a broad range of general BRDFs that are discriminative for evaluating photometric stereo methods.

3.1. Characteristics of Shapes and Materials

Shapes in DiLiGenT10^2 dataset vary from a simple ball to a general bunny (please check Fig. 1 row by row), which are divided into three groups.

- **Sphere group**: This group includes BALL and its 3 variants, to include smooth surfaces with varying levels of local details. BALL represents the most standard and frequently used shape that covers continuous surface normal changes from all directions. GOLF has multiple concave pits, which may lead to slight inter-reflection and cast shadow effects in many local regions. SPIKE, the counterpart of Golf, has many convex cones, to provide abrupt changes in local shapes and shadow. NUT provides both types of local shape changes in the form of peaks and ravines while maintaining majority of the surface smooth.
- **Polyhedron group**: This group contains a discrete set of multiple directional faces, to verify whether a photometric stereo method consistently predicts normals on planar surfaces with different orientations. SQUARE includes 4 rectangular faces and 4 triangular faces, and a main square face on the top. PENTAGON is a typically regular polyhedron with 5 main faces and 5 minor faces in our design. HEXAGON is a football-like hexagon shape with one main hexagon on top and 6 side faces (3 pentagons and 3 hexagons separate each other) facing to different orientations; it also contains 9 half polygons at the bottom (3 pentagons and 6 hexagons) which are almost vertical to the basis; these faces are challenging to photometric stereo due to that they are observed with severe attached shadows for the majority of lights.
- **General-shape group**: This group contains objects from industry machining and daily life, which provides natural surfaces with moderately complex curvatures and details. PROPELLER is a common industrial part (used in navigation, energy, and irrigation industries; its measurement has strong demands of surface quality in production implementation), which has several curved surfaces. TURBINE, commonly used in electric generator and aeronautic engine, has more blades but with less curved surface compared to PROPELLER. BUNNY is inspired by the famous Stanford Bunny [42] to represent a commonly seen daily life object with irregular geometric variations.

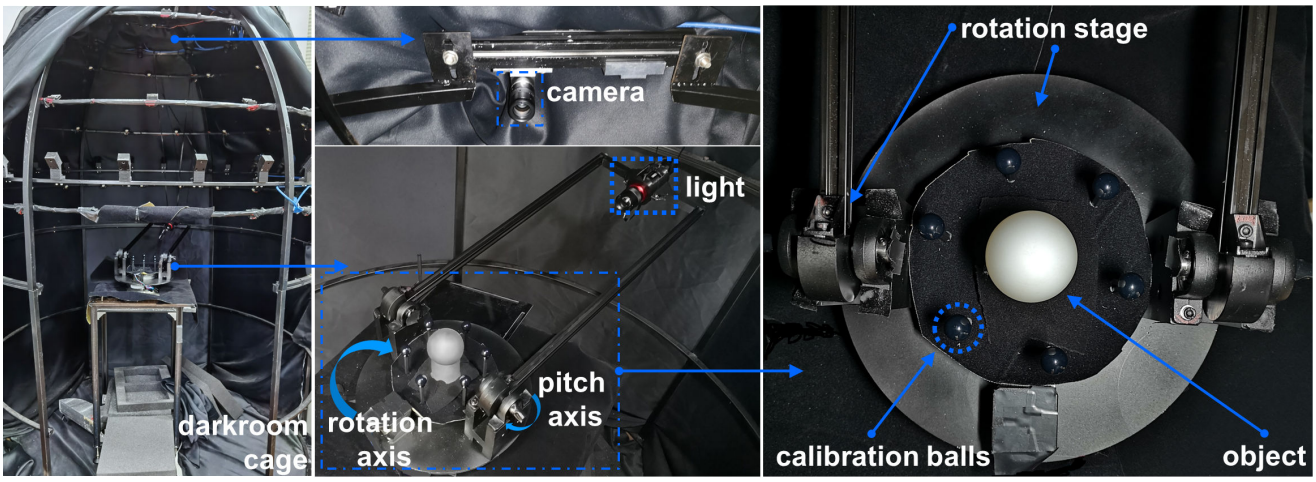

Figure 2. Lighting and imaging setup for building the DiLiGenT10^2 dataset. A cage is built and covered with black-out cloth to act as a darkroom. Camera is mounted on the zenith of the cage. An LED light source is mounted on a two-axes stage, and the target object is placed at the rotation center. Six specular balls are placed around the target object to calculate the lighting direction.

Materials in DiLiGenT10^2 dataset vary from simple (diffuse) to complex (anisotropic, translucent) reflectance (please check Fig. 1 column by column), which are also divided into three groups.

- **Isotropic group**: POM is a representative structural plastic material. PP, NYLON, and PVC are widely used plastics/polymers in daily life, *e.g.*, acting as packing materials for catering and sales industries. ABS is a typical engineering plastic and is commonly adopted as the outer case of products like home appliances and consumer electronics. BAKELITE is made from wood fiber and resin, which is a popular material used for electrical insulation. These materials cover a great diversity from diffuse to moderately specular materials, which are mostly isotropic. They have different roughness and are similar to materials used in existing datasets.
- **Anisotropic group**: The alloy made by aluminium AL, brass CU, and ferrum STEEL is widely used in making and constructing various types of objects in our world. We choose these three representative metallic materials because they also have diverse BRDFs. We adopt turning (for the BALL shape) and milling (for other shapes) to fabricate the objects. It is interesting to note that micro-structures caused by industry machining process make these objects demonstrate strong anisotropic reflectance, which is not commonly included in existing real photometric stereo datasets.
- **Challenging group**: We further include ACRYLIC, a translucent material which is utilized as optical plastics and alternatives of glass, to challenge existing photometric stereo methods (since few methods could model subsurface scattering) and inspire future research.

3.2. Lighting and Imaging Setup

Uniform intensity. A fast and convenient way to capture photometric stereo images under varying lighting conditions is to build a frame to fix LEDs at some pre-defined positions [37] and turn on each light sequentially to illuminate the target object. But such an approach introduces a tedious intensity calibration process, let alone the spatio-temporal inconsistency of each LED. First, the luminance from each LED is usually different in practice, even if we buy them in the same production batch. Second, the luminance of LEDs decreases over time, it could not work at a stable status when turning on/off instantaneously. Besides, the pre-fixed location is inflexible and it is usually not easy to put them near the equator regions due to occlusion *e.g.*, the lights in [37] mainly concentrate at the pole area limited by the rectangular frame adopted. To address these issues, we design a different illumination equipment with omnidirectional concentric equidistant design and a single light source, as shown in the middle part of Fig. 2. The concentric design ensures the light source to be located on a "virtual" hemisphere surface. Two rotation axes provide an omnidirectional angular resolution that can flexibly illuminate the target from any direction. Our setup exempts the intensity calibration process (and avoids corresponding errors) to physically ensure uniform lighting intensity.

Distant lighting approximation. The distant lighting assumption is still the most widely adopted for photometric stereo by far. Previous datasets like [37] approximate such an assumption by putting the light source at a large distance to the target object. However, the lighting distance could hardly be far enough due to the size limitation of the data capture environment. To better satisfy the distant lighting

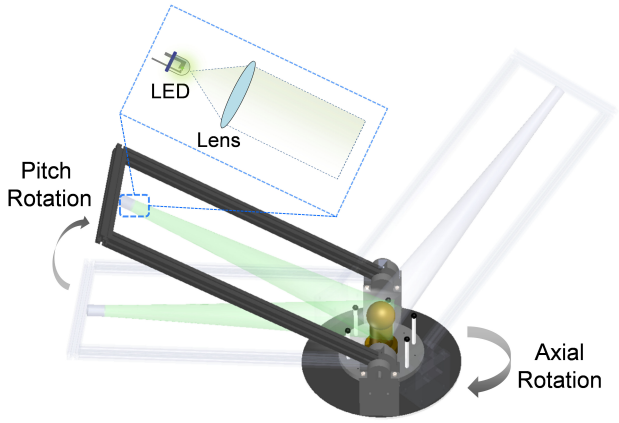

Figure 3. Illustration of the two-axes illumination system. It has two degrees of freedom to move the light at any position on a hemisphere while keeping a constant distance to the center for casting uniform lighting intensity. A convex lens is put in front of the light source to provide parallel illumination.

assumption, we place a focus lens in front of the LED to converge more light and increase the intensity for achieving better SNR, as shown in Fig. 3.

Image capture. Since the lighting direction is controlled using the two-axes system, we need to capture both images of the target object and lighting calibration objects simultaneously and conduct an "on the fly" lighting direction calibration process. The LED light source is constantly placed at $50cm$ away from the object and it faces toward the object to cast uniform lighting intensity. Images are captured by the DaHeng Image MER-503-36U3C camera with a $50mm$ lens and raw resolution at 2448×2048. The central region is cropped to 1001×1001 resolution as images for photometric stereo computation. All objects have a diameter of $5cm$ for the basis and are placed about $1.5m$ under the camera to approximate the orthographic projection assumption. Depending on different reflectivity of each material, we adjust camera exposure time in the range of $1ms$ to $4ms$ for avoiding saturation. The camera has a linear radiometric response, and PNG images (without compression) are stored.

Lighting direction calibration. We begin the capture process when LED warms up and luminance becomes stable. We manually control the two-axes illumination system to obtain a dense (more than 300) sampling on the hemisphere. To obtain an approximately uniform lighting direction distribution for consistently evaluating all objects, we adopt the Fibonacci sphere sampling method [17], as shown in Fig. 4. Specifically, we first utilize this method to generate 100 uniformly distributed sample positions on the hemisphere; then we search the nearest direction for each calibrated lighting direction based on the cosine distance. This re-sampling process selects 100 lighting directions (out of the dense observations) that are closest to the uniform distribution.

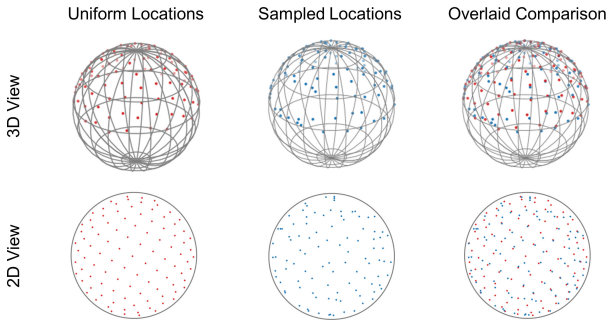

Figure 4. Uniform re-sampling process for lighting directions.

The lighting direction calibration is conducted using the same way as [37], *i.e.*, by relating the coordinates of specular spots (lighting direction), viewpoint, and surface normal of that point according to the law of reflection[4]. After illuminating, capturing, and calibrating lighting directions for each object, we obtain 100 sets of objects under 100 different lighting conditions.

3.3. Calibrating the Normal Map

Previous datasets [33, 37] use 3D scanners to acquire the shapes of objects first, align the scanned shapes with images, and finally render the surface normal map as "ground truth". The accuracy of such an approach may rely on the precision of the scanner. Moreover, some metallic objects are rather difficult to scan, and using painted coat may change the shape of the original object. We therefore choose to render the precise CAD model as "ground truth".

We first utilize CNC to fabricate objects based on their corresponding CAD models. The machining error of advanced CNC can be controlled to the order of about $10\mu m$, which is smaller than the measurement error of our camera system. Moreover, the repetitiveness from CNC procedure guarantees the consistency of objects across different materials, especially for those are difficult or even impossible to scan.

To align the CAD model with the captured images, we employ the same operation steps as [37] (please refer to their original paper and released toolbox for details). Now, we have the complete set of photometric stereo images with lighting directions (intensities fixed to 1) and normal map calibrated, which is ready for benchmark evaluation.

4. Benchmark Results

This section presents the up-to-date benchmark results for photometric stereo methods using the DiLiGenT10[2] dataset. Based on the benchmark results for non-learning based methods reported in [37], we only re-evaluate the representative ones for those have been tested: The baseline

[4]Details are in the supplementary material.

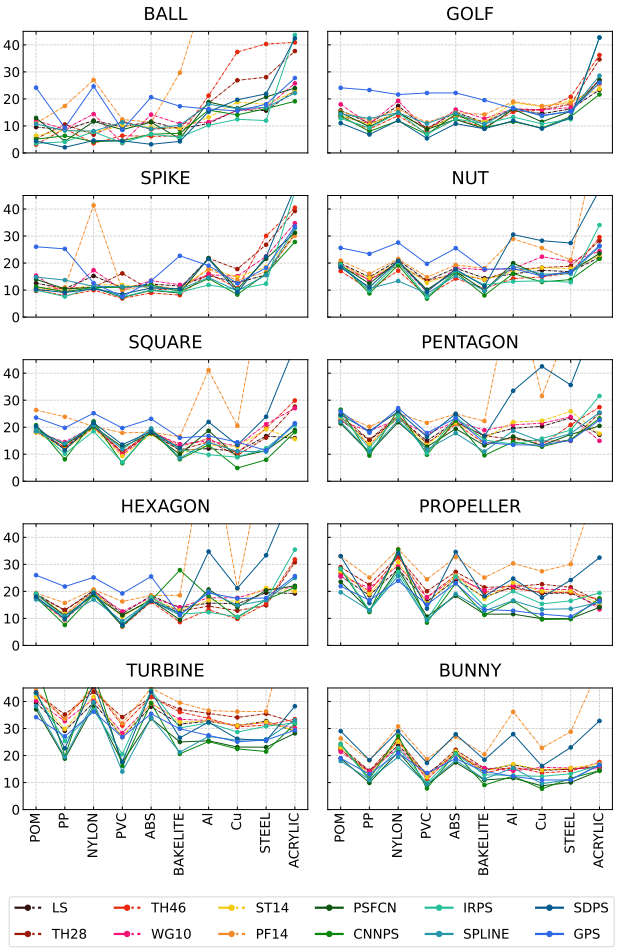

Figure 5. Shape (per sub-figure)-material (X-axis) error (Y-axis) plot for all methods. Best viewed in color and electronic version for magnified details.

(least squares based Lambertian method, LS), LS with position threshold [37] (TH28 and TH46 refers to using threshold as $[20\%, 80\%]$ and $[40\%, 60\%]$), WG10 [47] (a representative outlier rejection method), ST14 [38] (the best-performing general reflectance method reported in [37]), and PF14 [34] (the best-performing uncalibrated method reported in [37]). For learning based methods, we choose to evaluate PS-FCN [13], CNN-PS [21], IRPS [40], SDPS-Net [11], SPLINE-Net [50], and GPS-Net [49], which were published at CVPR/ICCV/ECCV/ICML/NeurIPS within the past three years. For non-learning based methods, we use the authors' released code to calculate the surface normal; for learning based methods, we adopt their pre-trained models to directly test on our dataset, which is fair to all methods since none of them have "seen" this dataset before. We adopt the most widely used mean angular error (across all valid pixels in each set) in degrees as error metric. For angular errors $\geq 90°$, they are truncated to $90°$ for easy visualization.

4.1. Shape-Material Error Plot for All Methods

In Fig. 5, we show the shape-material error plot for all methods. Each sub-figure is about the results of a fixed shape, and shows how the errors vary with materials for all evaluated methods (warm colors for non-learning while cold colors for learning based methods).

For the sphere group, *i.e.*, BALL, GOLF, SPIKE, NUT, the increasing trend of errors from left to right is quite clear. BALL is the easiest case due to its smoothness, while NUT is challenging due to different types of details on the surface. This group of shapes are generally convex, which contain few regions causing cast shadow and inter-reflection, therefore the variation of material type (reflectance property) plays a key role in affecting the results. Learning based methods generally outperform non-learning based ones. But non-learning based methods outperform learning based ones when their reflectance assumptions (such as isotropic constraint) are well satisfied for some materials like POM and NYLON for BALL and NUT.

For the polyhedron group, *i.e.*, SQUARE, PENTAGON, HEXAGON, the increasing trend for errors is less obvious, because this shape group contains less normal variation to involve complex reflectance phenomenon, so that the variation caused by different materials seem to put smaller weights than the sphere group case. The overall errors are larger than the sphere group, probably due to some nearly vertical faces to the basis causing strong attached shadows, like the contour region of PENTAGON and HEXAGON. It is interesting to note for complex materials in the anisotropic and challenging groups, the errors here are lower than the sphere group, and some non-learning based methods perform best.

For the general-shape group *i.e.*, PROPELLER, TURBINE, BUNNY, the errors also fluctuate with materials. The errors of PROPELLER is similar to those from polyhedron group, but TURBINE causes difficulties to a majority of methods since it contains much closer blades to introduce strong inter-reflections and cast shadows (especially when the lights are near the equator). BUNNY seems to be a relatively easier shape for many methods, partially due to its convex shape[5].

As a comparison between non-learning and learning based methods, the former group displays error plots with strong correlation (similar shapes) due to their similar reliance on reflectance properties, while the latter group shows weak correlation, since learning based methods have diverse strategies to predict surface normals.

[5]Another reason that BUNNY shows lower errors for anisotropic and challenging material is because its milling machining procedure is the most complicated (since its shape contains more diverse curvatures than other shapes), which requires more "sculpt" and introduces more surface microstructures that attenuate the anisotropic effect and transparency.

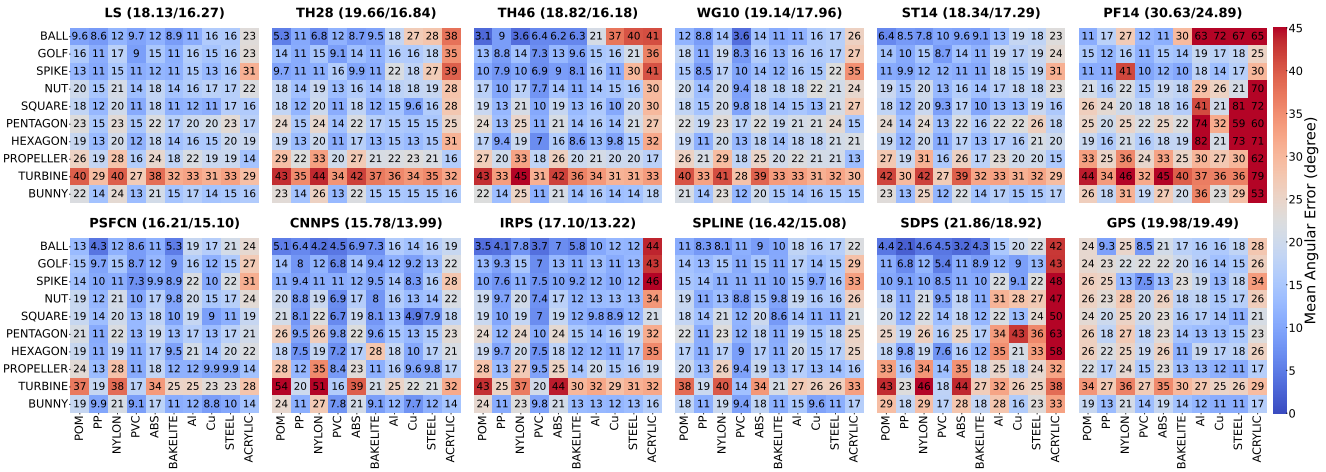

Figure 6. Shape-material error matrix for each method (mean/median). Best viewed in color and electronic version for magnified details.

4.2. Shape-Material Error Matrix for Each Method

In Fig. 6, we show shape-material error matrix (heat map) for each evaluated method. A number in each element of the matrix indicates a mean angular error according to a shape and material index. Such a matrix depicts a "portrait" for each photometric stereo method in the two-dimensional space, which intuitively and comprehensively demonstrates how a method performs on the DiLiGenT10^2 dataset. This type of visualization can never be supported by other real datasets, without controlled shape and material variation like ours.

The upper left corner of each matrix is a "safe zone" for the majority of methods, since they contain smooth shapes with isotropic reflectance. There are also two "red lines" in the row of TURBINE and column of ACRYLIC, which indicate the most challenging shape (due to strong global illumination effects) and material (due to translucency) in our dataset.

Among all non-learning based methods, TH46 [37] shows the lowest median error. Such stability may come from its simple way of computation, which is particularly effective when there is a dominant Lambertian component for materials in the isotropic group, but for low-frequency non-Lambertian materials (*e.g.*, ABS) ST14 [38] produces better results. But its modeling capability to highly specular BRDF is less effective, when compared with learning based methods like PSFCN [13] or CNNPS [21].

As a comparison between calibrated and uncalibrated methods, the former group generally has smaller errors than the latter group, especially for non-learning based methods. This means uncalibrated photometric stereo is still more challenging than the calibrated one. But deep learning helps a lot in suppressing errors from unknowing lighting conditions (SPDS [11] vs. PF14 [34]).

5. Open Problem Discussion

We conclude this paper by discussing open problems revealed from the benchmark results of DiLiGenT10^2 dataset.

An interesting phenomenon DiLiGenT10^2 dataset demonstrates is that increasing the material complexity (like anisotropic) brings less (than expected) distortions to normal estimation, unless the reflectance deviates really far from what a BRDF can describe (the translucent material), while the challenges to normal estimation more strongly relate to global illumination effects such as cast shadows and inter-reflection (*e.g.*, TURBINE). Materials with rough surfaces (*e.g.*, POM and ABS) could be more challenging than highly specular ones for learning based methods, due to that existing training datasets cover less about such BRDFs. Learning an effective photometric stereo solution needs pay attention to global illumination, a balanced shape distribution, and unseen material types.

One of the limitations for the DiLiGenT10^2 dataset is that we still cannot remove the quotation marks from "ground truth" normal, since the shape-to-image registration procedure following [37] unavoidably involves errors, and the shapes we made across different materials are not perfectly consistent with each other as well as their CAD counterparts. Developing a photometric stereo dataset with real ground truth of surface normal is still an open problem.

Acknowledgement

This work is supported by National Natural Science Foundation of China under Grant No. 62136001, 62088102, 61872012, 61925603, 61972119, 52175477. Authors want to thank Jinfa Yang, Bohan Yu, and Xi Wang for technical assistance, and *openbayes.com* for providing computing resource.

References

[1] Jens Ackermann, Fabian Langguth, Simon Fuhrmann, and Michael Goesele. Photometric stereo for outdoor webcams. In *Proc. CVPR*, 2012. 3

[2] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. CVPR*, 2008. 3, 4

[3] Anthony L Andrady and Mike A Neal. Applications and societal benefits of plastics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1526):1977–1984, 2009. 4

[4] Gytis Bernotas, Livia C T Scorza, Mark F Hansen, Ian J Hales, Karen J Halliday, Lyndon N Smith, Melvyn L Smith, and Alistair J Mccormick. A photometric stereo-based 3D imaging system using computer vision and deep learning for tracking plant growth. *GigaScience*, 8(5), 2019. 3

[5] Simon Brenner, Sebastian Zambanini, and Robert Sablatnig. An investigation of optimal light source setups for photometric stereo reconstruction of historical coins. In *Proc. GCH*, 2018. 3

[6] Laurence Broadbent, Khemraj Emrith, Abdul R. Farooq, Melvyn L. Smith, and Lyndon N. Smith. 2.5D facial expression recognition using photometric stereo and the area weighted histogram of shape index. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2012. 3

[7] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *ACM SIGGRAPH*, 2012. 3

[8] Jordan Caracotte, Fabio Morbidi, and El Mustapha Mouaddib. Photometric stereo with central panoramic cameras. *Computer Vision and Image Understanding*, 201:103080, 2020. 3

[9] Jordan Caracotte, Fabio Morbidi, and El Mustapha Mouaddib. Photometric stereo with twin-fisheye cameras. In *Proc. ICPR*, 2021. 3

[10] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[11] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *Proc. CVPR*, 2019. 3, 7, 8

[12] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee Kenneth Wong. Deep photometric stereo for non-lambertian surfaces. *IEEE TPAMI*, 2020. 3

[13] Guanying Chen, Kai Han, and Kwan-Yee K Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proc. ECCV*, 2018. 3, 4, 7, 8

[14] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *Proc. ECCV*, 2020. 3

[15] Tinsae Gebrechristos Dulecha, Ruggero Pintus, Enrico Gobbetti, and Andrea Giachetti. SynthPS: A benchmark for evaluation of photometric stereo algorithms for cultural heritage applications. In *The Eurographics Workshop on Graphics and Cultural Heritage*, 2020. 3

[16] Kenji Enomoto, Michael Waechter, Kiriakos N Kutulakos, and Yasuyuki Matsushita. Photometric stereo via discrete hypothesis-and-test search. In *Proc. CVPR*, 2020. 2

[17] Álvaro González. Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, 42(1):49–64, 2010. 6

[18] Heng Guo, Zhipeng Mo, Boxin Shi, Feng Lu, Sai Kit Yeung, Ping Tan, and Yasuyuki Matsushita. Patch-based uncalibrated photometric stereo under natural illumination. *IEEE TPAMI*, 2021. 3

[19] Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita. Multispectral photometric stereo for spatially-varying spectral reflectances: A well posed problem? In *Proc. CVPR*, 2021. 3

[20] Yannick Hold-Geoffroy, Paulo F. U. Gotardo, and Jean-François Lalonde. Single day outdoor photometric stereo. *IEEE TPAMI*, 43:2062–2074, 2021. 3

[21] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. ECCV*, 2018. 3, 4, 7, 8

[22] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *Proc. CVPR*, 2011. 3

[23] Yakun Ju, Junyu Dong, and Sheng Chen. Recovering surface normal and arbitrary images: A dual regression network for photometric stereo. *IEEE TIP*, 30:3676 – 3690, 2021. 3

[24] Yakun Ju, Xinghui Dong, Yingyu Wang, Lin Qi, and Junyu Dong. A dual-cue network for multispectral photometric stereo. *Pattern Recognition*, 100:107162, 2020. 3

[25] Yakun Ju, Kin-Man Lam, Yang Chen, Lin Qi, and Junyu Dong. Pay attention to devils: A photometric stereo network for better details. In *IJCAI*, 2020. 3

[26] Yakun Ju, Lin Qi, Jichao He, Xinghui Dong, Feng Gao, and Junyu Dong. MPS-Net: Learning to recover surface normal for multispectral photometric stereo. *Neurocomputing*, 375:62–70, 2020. 3

[27] Jiyoung Jung, Joon-Young Lee, and In So Kweon. One-day outdoor photometric stereo via skylight estimation. In *Proc. CVPR*, 2015. 3

[28] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncalibrated neural inverse rendering for photometric stereo of general surfaces. In *Proc. CVPR*, 2021. 3, 4

[29] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *Proc. CVPR*, 2019. 3

[30] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE TIP*, 29:4159–4173, 2020. 3, 4

[31] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. PX-NET: Simple and efficient pixel-wise training of photometric stereo networks. In *Proc. ICCV*, 2021. 3

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

[33] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. LUCES: A dataset for near-field point light source photometric stereo. *arXiv preprint arXiv:2104.13135*, 2021. 3, 4, 6

[34] Thoma Papadhimitri and Paolo Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *IJCV*, 107(2):139–154, 2014. 7, 8

[35] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *ICCV Workshops*, 2017. 3, 4

[36] Boxin Shi. Data-driven photometric 3D modeling. In *SIGGRAPH Asia Courses*, 2019. 3

[37] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE TPAMI*, 41(2):271–284, 2019. 1, 2, 3, 4, 5, 6, 7, 8

[38] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE TPAMI*, 36(6):1078–1091, 2013. 7, 8

[39] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proc. CVPR*, 2016. 1

[40] Tatsunori Taniai and Takanori Maehara. Neural inverse rendering for general reflectance photometric stereo. In *Proc. ICML*, 2018. 7

[41] George E Totten, Lin Xie, and Kiyoshi Funatani. *Handbook of mechanical alloy design*, volume 164. CRC press, 2003. 4

[42] Greg Turk and Marc Levoy. Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, 1994. 4

[43] Igor Vasiljevic, Nick Kolkin, Shanyi Zhang, Ruotian Luo, Haochen Wang, Falcon Z Dai, Andrea F Daniele, Mohammadreza Mostajabi, Steven Basart, Matthew R Walter, and Gregory Shakhnarovich. DIODE: A Dense Indoor and Outdoor DEpth Dataset. *arXiv preprint arXiv:1908.00463*, 2019. 3

[44] Xueying Wang, Yudong Guo, Bailin Deng, and Juyong Zhang. Lightweight photometric stereo for facial details recovery. In *Proc. CVPR*, 2020. 3

[45] Olivia Wiles and Andrew Zisserman. SilNet: Single-and multi-view reconstruction by learning from silhouettes. In *Proc. BMVC*, 2017. 3

[46] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139, 1980. 1

[47] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. ACCV*, 2010. 7

[48] Ying Xiong, Ayan Chakrabarti, Ronen Basri, Steven J. Gortler, David W. Jacobs, and Todd Zickler. From shading to local shape. *IEEE TPAMI*, 37(1):67–79, 2015. 3, 4

[49] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi. GPS-Net: Graph-based photometric stereo network. In *Proc. NeurIPS*, 2020. 3, 7

[50] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling-Yu Duan, and Alex C Kot. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In *Proc. ICCV*, 2019. 3, 7

[51] Qian Zheng, Boxin Shi, and Gang Pan. Summary study of data-driven photometric stereo methods. *Virtual Reality & Intelligent Hardware*, 2(3):213–221, 2020. 3