Measurement 210 (2023) 112543

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

DPPS: A deep-learning based point-light photometric stereo method for 3D reconstruction of metallic surfaces

Ru Yang, Yaoke Wang, Shuheng Liao, Ping Guo *

Department of Mechanical Engineering, Northwestern University, Evanston, IL, USA

ARTICLE INFO

Dataset link: https://github.com/RuYangNU/point-light-photometric-stereo

Keywords:
3D reconstruction
Photometric stereo
Convolutional neural network
Deep learning
Point light

ABSTRACT

Three-dimensional (3D) measurement provides essential geometric information for quality control and process monitoring in many manufacturing applications. Photometric stereo is one of the potential solutions for in-process metrology and active geometry compensation, which takes multiple images of an object under different illuminations as inputs and recovers its surface normal map based on a reflectance model. Deep learning approaches have shown their potential in solving the highly nonlinear problem for photometric stereo, but the main challenge preventing their practical application in process metrology lies in the difficulties in the generation of a comprehensive dataset for training the deep learning model. This paper presents a new **D**eep-learning based **P**oint-light **P**hotometric **S**tereo method, DPPS, which utilizes a multi-channel deep convolutional neural network (CNN) to achieve end-to-end prediction for both the surface normal and height maps in a semi-calibrated fashion. The key contribution is a new dataset generation method combining both physics-based and data-driven approaches, which minimizes the training cost and enables DPPS to handle reflective metal surfaces with unknown surface roughness. Even trained only with fully synthetic and high-fidelity dataset, our DPPS surpasses the state-of-the-art with an accuracy better than 0.15 cm over a 10 cm × 10 cm area and its real-life experimental results are on par with commercial 3D scanners. The demonstrated results provide guidance on improving the generalizability and robustness of deep-learning based computer vision metrology with minimized training cost as well as show the potential for in-process 3D metrology in advanced manufacturing processes.

1. Introduction

Advances in in-situ metrology techniques have upgraded the capability of modern manufacturing systems to a more competitive level. These techniques are critical enablers for in-process performance monitoring and quality control. Beyond traditional dimensional measurement techniques, three-dimensional (3D) measurement can provide comprehensive geometrical information of a physical object, which has gained increasing popularity in academia and industry [1–3]. Many emerging manufacturing processes, such as metal additive manufacturing, 5-axis machining, incremental forming, etc. [4–6], will benefit greatly from in-process 3D measurement techniques for improved geometry accuracy. In addition, in-process 3D metrology will provide a powerful tool enabling the investigation of some unique and complex process mechanics, such as the thermal expansion in additive manufacturing [7] and the geometric deviations in metal forming and forging [8,9].

Non-contact metrology provides the possibility of in-process measurement without interfering with the manufacturing process. Overall, scanning-based 3D reconstruction, such as laser scanning [10], microwave imaging [11], and computed tomography (CT) [12], can offer sub-millimeter level accuracy with moderate efficiency (millions of points in one scan); however, limitations of these methods include system complexity, high cost, and material restrictions, making it difficult to be applied for in-process metrology and monitoring. For example, the most popular scanning technology, laser scanning, faces difficulty in collecting high-quality data when scanning reflective surfaces that are prevalent with metal components. Though the speckle noise and spurious reflections can be removed with cross polarization [13] and image pre-processing [14], the low receiving reflection intensity caused by specular reflection is still a challenge for getting a confident scanning result.

As an alternative to scanning-based metrology, image-based methods are gaining popularity because of their simple and low-cost hardware requirements. With the recent advancement in computer vision techniques, multiple image-based 3D reconstruction methods have been developed with rapidly improving accuracy. Passive methods of 3D

* Corresponding author.
E-mail address: ping.guo@northwestern.edu (P. Guo).

https://doi.org/10.1016/j.measurement.2023.112543
Received 14 September 2022; Received in revised form 31 December 2022; Accepted 24 January 2023
Available online 28 January 2023
0263-2241/© 2023 Elsevier Ltd. All rights reserved.

reconstruction, such as multi-view stereo [15,16] and structure-from-motion [17], recover 3D information by mapping the feature correspondence from a set of 2D overlapping images. These methods work better with objects that have plenty of unique features for matching. For the reconstruction of textureless surfaces, which are commonly encountered in manufacturing applications, the accuracy of these methods is limited. Derived methods are developed to improve the accuracy in the case of textureless surface reconstruction by projecting patterns on the surface [18–20] or powdering the object surface [21]. And recently, NeRF (neural radiance field) [22], a completely new direction is introduced that trains multilayer perceptrons [22] or neural 3D points [23] to represent volumetric radiance field from a large number of images of the object from various views. These methods require multiple camera angles, which are difficult to implement in many manufacturing process setups, and still suffer from lower measurement accuracy.

Active methods for 3D reconstruction, such as deflectometry [24], structured light reconstruction [25], and photometric stereo [26], utilize lighting and projection conditions to extract more detailed shape and appearance information from 2D images. Deflectometry and structured-light reconstruction use almost identical setup, which includes a projector to project pattern on the object and one or two cameras to capture images. Deflectometry projects fringe patterns to a target object and reconstructs the surface normal based on the phase information in reflected fringe patterns. Because of its high sensitivity to surface curvature change, it is good at reconstructing specular (shiny) and nearly flat surface [27]. To be compared, structured light reconstruction recovers 3D geometry from the distorted fringe or chessboard-like patterns using the principle of triangulation [28] and is more suitable for reconstructing less specular or more complex surfaces. It is already adopted by multiple commercial indoor RGB-D cameras as the basic algorithm [29]. Deflectometry and structured light reconstruction capture images from a limited number of viewpoints (commonly one or two), while the projected structured pattern comes from a single source [30,31]. Thus, reconstruction failure occurs for surfaces with large fluctuations when the projected pattern is occluded by the shadow effect imposed by surface terrain.

Photometric stereo [26,32,33] is another active method for 3D reconstruction which utilizes images taken under different illuminations to reconstruct the normal map of a surface. It utilizes reflectance models for estimating surface properties from transformations of image intensities that arise from illumination changes [26,34]. Different from structured light reconstruction and deflectometry, photometric stereo utilizes light sources from multiple directions and takes multiple captures, which lessens the effect of occlusion. Compared with deflectometry, which is better at reconstructing specular and nearly flat surfaces, photometric stereo is more capable of reconstructing 3D objects with more fluctuations. Furthermore, the reconstruction resolution of photometric stereo is not limited by the resolution of fringe patterns, so photometric stereo tends to recover finer details in the surface normal variations.

The surface normal reconstruction in photometric stereo solves the inverse reflection process with prior knowledge of the reflectance property of the target material. The fully diffusive reflection from so-called Lambertian surfaces can be analytically calculated by the photometric stereo, whose reflectance is not influenced by the viewing angle [35]. However, most real-world surfaces are not Lambertian and exhibit both diffusive and specular appearances. The specular appearance introduces nonlinearity to the reflectance model, which adds more complexity when solving the inverse reflection process. Early research tried to handle the specular appearance by developing analytical reflectance models for more general materials [36–39] or excluding non-Lambertian effects as outliers [40–42]. For non-Lambertian surfaces, the reflectance is usually described by a bi-directional reflectance distribution function (BRDF). Shi et al. [37] developed a bi-polynomial BRDF model to account for the low-frequency non-Lambertian effects. More generally, sophisticated BRDF models were developed to approximate reflectance with highly non-Lambertian effects [38]. However, methods in this category require solving complicated optimization problems and thus are computationally inefficient. For outliers-rejection methods, multiple machine learning approaches such as sparse regression [36,40], random sample consensus [41], and rank minimization [42] have been utilized. For example, Wu et al. [42] regarded the problem as a rank minimization problem where Lambertian reflection was treated as a low-rank subspace while non-Lambertian reflectance was excluded as outliers during rank minimization. This group of methods usually requires a large number of image inputs and works only when non-Lambertian (specular) observations occupy a small portion of the whole image.

Based on their success in computer vision tasks, deep learning methods have been introduced to photometric stereo to directly reconstruct non-Lambertian surfaces without the explicit knowledge of an analytical reflectance model and brought the accuracy to the state-of-the-art [43–49]. Santo et al. [43] first attempted a deep learning approach in photometric stereo. The surface normal was estimated in a point-by-point manner with a 6-layer fully connected neural network. Ikehata utilized convolutional neural networks (CNNs) to reconstruct very shiny surfaces with higher robustness which took into account both the global illumination effect (such as illumination variance and shadows) and inter-reflectance [44]. Chen et al. developed two CNNs, PS-FCN [45] and SDPS-Net [46], that supported an arbitrary number of input images taken under random lighting conditions. A combined loss function was defined in [46] that enabled the CNN to learn the surface normal as well as lighting directions of each image input. Li et al. [48] developed a deep learning framework in which a connection table was applied to select the most relevant lighting directions for reconstruction. As a result, the number of input images was reduced while the prediction accuracy was maintained. Ju et al. [49] introduced a dual-regression task to photometric stereo, which synthesizes reconstructed images from the estimated normal map to add additional supervision and uplift the accuracy.

Though deep learning methods have improved the performance and flexibility of photometric stereo, there remain several inherent challenges that prevent the further application of photometric stereo in manufacturing applications. The first challenge is the dataset generation for training deep learning models in photometric stereo. Since photometric stereo is an inverse mapping process from the image intensity to the normal map, the reflectance model critically determines the accuracy and robustness of photometric stereo. For deep-learning based photometric stereo, the reflectance model is usually not explicitly defined but embodied in the training dataset. A deep learning model implicitly learns surface reflectance and encodes this hidden knowledge in the neural network to recover the normal map. Therefore the resemblance between the dataset and reality determines the accuracy of the reconstruction result. Previously, the dataset for training deep-learning based photometric stereo was either captured from a limited set of experiments or generated from computer simulation. The experimentally captured datasets give the most realistic reflectance conditions [50,51]; however, they are usually too expensive with limited representations to train a deep learning model. On the other hand, synthetic datasets [45,47] are generated by taking linear combinations of different BRDF models from an experimental-based reflectance library [52]. This approach is unrealistic as it will smooth out the high-frequency Lambertian effects. There still lacks an approach to generate a realistic and comprehensive reflectance model that covers various materials. In addition, simulated datasets are often rendered in an ideal condition without considering uncertainties and variations in illumination and imaging. The performance of these trained models would be unstable and sensitive to inevitable environmental variations. In order to generate high-quality and realistic datasets, both experimental data and the physics of reflectance should be considered.

Another limitation of photometric stereo lies in the ill-poised integration from surface normal maps to height maps. According to

the reflectance principle, the captured image intensity depends on the normal vector for a given material and illumination direction. By solving the inverse reflectance problem, photometric stereo inherently only outputs the surface normal map of an object. Though the height map can be derived by integrating the normal map, the problem is ill-posed when the surface has discontinuity or free shape boundaries [53]. This physical restriction limits the reconstruction to 2.5D instead of full 3D. To alleviate this limitation, Vlasicl et al. [54] described a multi-view photometric stereo that recovered the normal map of an object from different views and matched the correspondence of different views to recover the full 3D surface. Haefner et al. [55] utilized an RGB-D camera to perform photometric stereo so a high-resolution height map can be reconstructed by combining the low-resolution geometric constraint from the depth camera and the details from the integrated normal map calculated with photometric stereo. However, these methods all depend on a separate algorithm that does not directly relate to photometric stereo. There is potential in deep learning to use an end-to-end approach for height map reconstruction by fully utilizing the captured image information from photometric stereo.

To address the above challenges, in this paper, we present DPPS, a novel Deep-learning based Point-light Photometric Stereo for 3D reconstruction of metallic surfaces with unknown surface roughness. The proposed DPPS method has the following major contributions:

- DPPS directly reconstructs both the height map and normal map of a 3D shape from images taken under different point-light conditions utilizing the principle of photometric stereo and the nonlinear mapping abilities of deep CNNs.
- This work is the first attempt to apply deep learning to point-light photometric stereo. By utilizing point light sources, an extra dimension is introduced into the photometric stereo framework, which enables direct prediction of the height map of a 3D object.
- Another major contribution of the work is a new dataset generation method combining physics-based and data-driven approaches that include experimentally calibrating light sources, training a data-driven reflectance model, and rendering realistic image sets. To improve the model robustness in experiments, careful considerations are made to include image capture noises, calibration uncertainty, and light intensity variations.
- Last but not least, we have demonstrated DPPS's performance on both synthetic datasets and real-world experiments. DPPS shows great generalizability to work with metal materials with unknown surface roughness. By only training DPPS with our fully synthetic and high-fidelity dataset, its performance surpasses the state-of-the-art, while its real-life experimental results are on par with commercial 3D scanners. The demonstrated results provide guidance on improving the generalizability and robustness of deep-learning based computer vision metrology with minimized training cost as well as showing the potential for in-process 3D metrology in advanced manufacturing applications.

2. Methods

The overall flowchart of DPPS is illustrated in Fig. 1. A dome-shaped setup is used to capture images of the target object under 96 point-light illumination conditions. A multi-channel CNN is designed to take the captured images as inputs and predict the target object's surface normal and height maps simultaneously. The novel design of the CNN implicitly assigns the ill-posed problem of direct integration to the deep neural network, thus avoiding the integration error from normal to height. The model is trained by a synthetic dataset, which is the other key part of this research work. The dataset generation process incorporates calibrated light conditions (Section 2.1), a data-driven reflectance model considering unknown surface roughness (Section 2.2), and image rendering method considering lighting variation and measurement noises (Section 2.3). The design of the new CNN architecture and training details are given in Section 2.4.

2.1. Point-light photometric stereo and setup calibration

Traditional photometric stereo assumes the incident light to be parallel. This condition holds true when the light source has a far larger diameter than the dimension of the reconstructed object (such as sunlight) or when the light is placed far enough from the object. By replacing the parallel lights with point lights, the practicality of photometric stereo for in-process metrology is largely improved. In this section, we will introduce the setup of the proposed deep-learning based photometric stereo, DPPS, that works under point light illumination and the calibration procedure of the light sources. In point-light photometric stereo, the incident light intensity is both direction- and location-dependent, forming a highly nonlinear problem. For a given point on the target surface with a coordinate vector, $\boldsymbol{X_P}$, and the point-light source location, $\boldsymbol{L_k}$, the light intensity vector, $\boldsymbol{I}$, is given by [56]

$$I = \frac{L_k - X_P}{\|L_k - X_P\|} \cdot \frac{1}{\|L_k - X_P\|^2} \cdot I_0 \tag{1}$$

Eq. (1) indicates that the incident light for each point on the 3D surface directs from the light source to the object point and has an intensity that is proportional to the inverse squared distance between the light source and object point. Unlike the uniform incident light in parallel-light photometric stereo, the location-dependency of point-light illumination brings an additional dimension to the framework of photometric stereo, which gives rise to the direct prediction of height map of the reconstructed surface. An iteration procedure is often involved in solving the height map from the normal map in point-light photometric stereo.

The proposed deep learning approach relies on the nonlinear mapping ability of deep CNNs to solve the surface normal and height in an end-to-end manner. The complex iteration procedures with an analytical approach are accomplished by a single CNN in the proposed DPPS. The light conditions are consistent in the training dataset and physical setup, so the CNN will learn to implicitly determine the light source locations. This is a semi-calibrate approach in which the light source information is not directly input to DPPS networks, but needs to be calibrated for rendering the training dataset. The setup of the proposed method and the calibration of the point-light locations are detailed as follows.

We adopt a point-light photometric stereo configuration in the DPPS framework with 96 individually controlled light-emitting diodes (LEDs), as shown in Fig. 2(a). The 96 point-light sources are mounted on the inner surface of an aluminum dome shell, as shown in Fig. 2(a). The dome shell has a parabolic shape with a focal length of 152.4 mm and a bottom diameter of 609.6 mm. The 96 LEDs are located at five different height levels, while the number of LEDs at each level from top to bottom is 6, 10, 18, 28, and 34, respectively. They are approximately spaced evenly at each level. The white light LEDs have a diameter of 1.4 mm and are parallelly controlled by an Arduino micro-controller. A 2.2 megapixel monochrome CMOS camera (MQ022MG-CM, *Ximea*, Germany) is mounted above the dome for image capture through a center hole drilled at the top of the dome. The working distance of the camera is set to 520 mm. The lens used is a C-mount fixed focal length lens (16 mm) with a maximum aperture of f/1.6. The ground sampling distance (GSD) of the proposed method is:

$$\frac{(flight\ altitude \times sensor\ height)}{(focal\ length \times image\ height)} = \frac{(520\ \text{mm} \times 11.27\ \text{mm})}{(16\ \text{mm} \times (2048\ \text{pixel}/4.461))} = 0.798\ \text{mm} \tag{2}$$

where the image height is divided by 4.461 which is the ratio of captured image size in the reconstruction area and the reconstruction size.

Calibration of the point-light positions is performed in two steps. The first step is to determine the LED locations in the dome coordinate. As shown in Fig. 2(b), the inner surface of the dome is captured

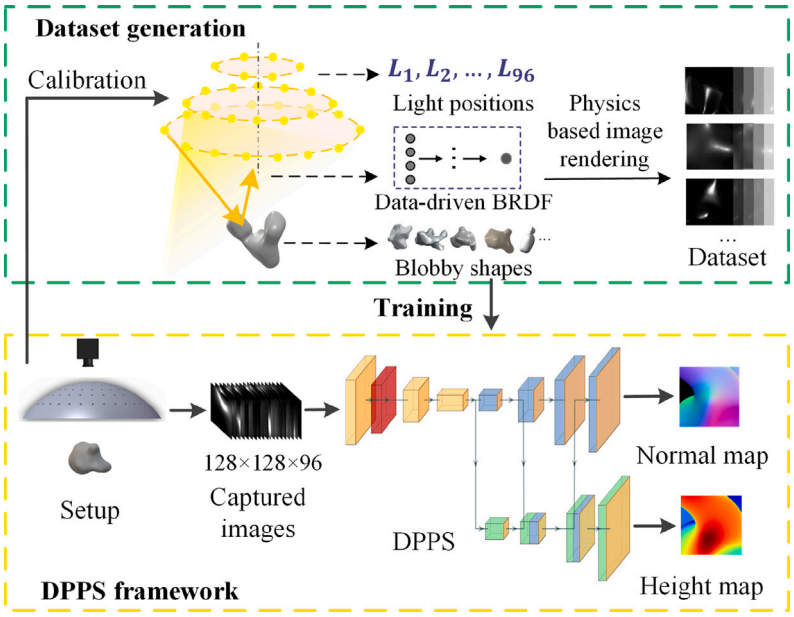

Fig. 1. Flowchart of DPPS.

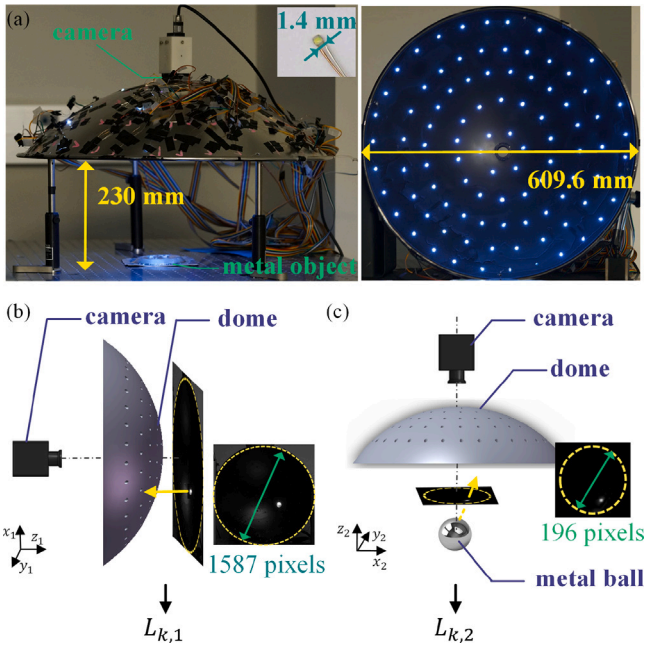

Fig. 2. (a) Point-light photometric stereo setup and the inner surface of the dome setup mounted with 96 LEDs; schematics showing (b) light position calibration and (c) alignment calibration.

with each LED individually turned on. In the captured images, the dome has a diameter of 1,587 pixels, which is large enough to give precise positioning of each point-light. The dome center is determined through the best fit of a circle, while the X, Y, and Z positions of the LED can be calculated from the analytical expression of the dome geometry. The resultant coordinates are denoted as a 96 × 3 matrix $\boldsymbol{L_1}$. The second step is to align the dome coordinate to the camera coordinate, so the accurate light positions can be determined with respect to the measurement frame. We place a 1.5-inch shiny metal ball (G25 precision chrome steel ball, *PGN Bearings*, USA) below the dome and along its center axis, which has a diameter of 196 pixels in the camera view, as shown in Fig. 2(c). We then take 96 pictures of the light reflection from each light source. In each captured image, the light spot location indicates the illumination direction by considering the geometrical relationship of the metal ball and the dome geometry. The light locations are then back-calculated, considering the intersection between the illumination direction and the dome surface. The results are denoted as a 96 × 3 matrix $\boldsymbol{L_2}$.

We use a singular value decomposition (SVD) method to decompose the inner product of $\boldsymbol{L_1}$ and $\boldsymbol{L_2}$ to two unitary matrices: $\boldsymbol{U}$ and $\boldsymbol{V}$, and a rectangular diagonal matrix $\boldsymbol{S}$, as defined by

$$[\boldsymbol{U}, \boldsymbol{S}, \boldsymbol{V}] = SVD(\boldsymbol{L_2^T} \cdot \boldsymbol{L_1}) \tag{3}$$

The orthogonal matrix $\boldsymbol{V}$ is a rotation matrix that maps $\boldsymbol{L_1}$ from the dome coordinate to the camera coordinate. The final calibrated point-light positions $\boldsymbol{L}$ (96 × 3) are derived as

$$\boldsymbol{L} = \boldsymbol{L_1} \cdot \boldsymbol{V} \tag{4}$$

2.2. Reflectance model for metals with different roughness

In order to render realistic reflection images with a range of different surface roughness, we build a data-driven and physics-based reflectance model for the dataset generation. In this section, a modified bidirectional reflection function (BRDF) is established to take into account the surface roughness of metal surfaces. Different metal materials exhibit similar 'metallic' reflection characteristics, thus we use aluminum to represent all metal materials. An experimental procedure is designed to collect metal reflection properties using aluminum sphere artifacts. The collected data are used to train a fully-connected neural network, enabling the interpolation and extrapolation of arbitrary $\theta_i, \theta_r, \phi_i - \phi_r$ and surface roughness. Though this data-driven reflectance model is only used for dataset generation, not directly used in the final DPPS, it is an essential component to ensure the quality of the training dataset, and thus significantly influencing the final DPPS performance.

The reflectance properties of an object can be described by a BRDF, which assumes that the reflection intensity is a function of four input angles consisting of the incident lighting direction (θ_i, ϕ_i) and the viewing direction (θ_r, ϕ_r), as shown in Fig. 3(a). For an isotropic material, the in-plane rotations will not affect the reflectance, so the input angles can be reduced to $\theta_i, \theta_r, \phi_i - \phi_r$. In our modified BRDF model, the surface roughness is added as an additional input parameter to the function. The data-driven model can perform the interpolation not only on the

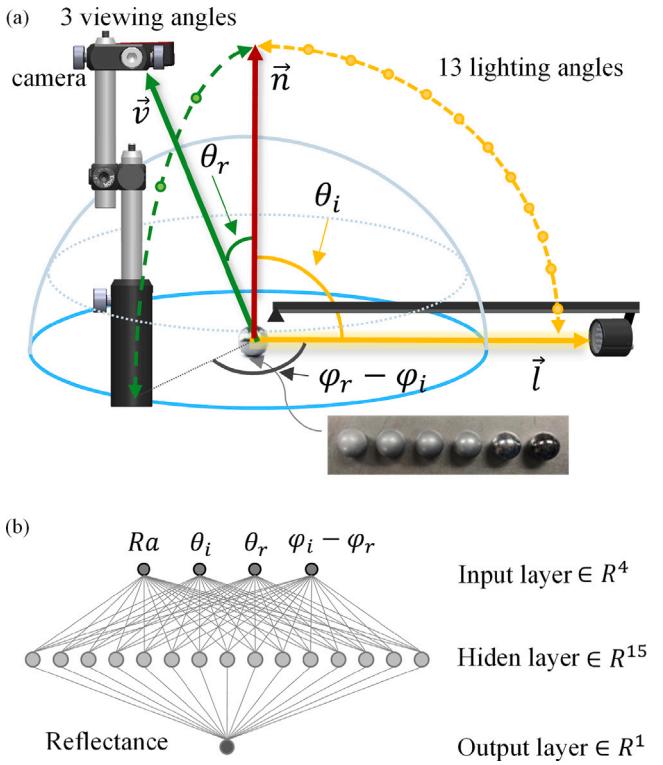

Fig. 3. (a) Experiment setup for BRDF measurement and 6 metal spheres with different surface roughness as the target artifacts; (b) structure of fully connected neural network for a data-driven BRDF model.

viewing and illumination angles, but also on the surface roughness. The modified BRDF model and the relationship between three input angles and three direction vectors can be described by

$$\begin{aligned} BRDF: Reflectance &= f(\theta_i, \theta_r, \phi_i - \phi_r, Ra) \\ \theta_i &= \left\langle \vec{l}, \vec{n} \right\rangle \\ \theta_r &= \langle \vec{v}, \vec{n} \rangle \\ \phi_i - \phi_r &= \left\langle \vec{l} - (\vec{l} \cdot \vec{n})\vec{n}, \vec{v} - (\vec{v} \cdot \vec{n})\vec{n} \right\rangle \end{aligned} \tag{5}$$

where $\vec{l}$ indicates the lighting direction vector; $\vec{n}$ is the surface normal vector; and $\vec{v}$ is the viewing direction vector. Operator $\langle , \rangle$ calculates the angle between two vectors, that is:

$$\left\langle \vec{a}, \vec{b} \right\rangle = arccos\left(\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}\right) \tag{6}$$

The data collection for training the modified BRDF model is performed with the following procedures. We use six aluminum spheres with a diameter of 3/4 inch as the artifacts for BRDF data collection, which are polished to different surface roughness with a Ra value of 0.52 μm, 0.68 μm, 0.78 μm, 1.07 μm, 2.13 μm, and 4.07 μm, respectively. The spherical shape of the artifact can provide a wide range of surface normal directions in a single shot, whose directions can be determined from its projected lateral positions.

The reflection intensities of six artifacts are recorded at different viewing and lighting angles. In the experiment, we set the camera to three different viewing angles (the zenith angle, θ_r, equal to 0°, 22°, 54°; and the azimuth angle, $\phi_i - \phi_r$, equal to 90°). The metal sphere is placed at the center of the camera view. A narrow-angle LED bulb (rated at 12 W, 1200 Lumen, *Lusta LED*, China) is fixed on the end of an adjustable arm beam as the light source, which can be rotated to adjust the zenith angle, θ_i. For each viewing angle, the lighting angle is set to 19 levels from horizontal to vertical, at an interval of 5°. For each captured image with a set of illumination and viewing angles, all pixels on the metal sphere are utilized as the training data for BRDF modeling. For each pixel on the sphere, the surface normal vector is calculated based on its relative position to the sphere center. Based on the illumination, viewing, and surface normal directions, the corresponding input angles are derived for $\theta_i, \theta_r, \phi_i - \phi_r$, following Eq. (5).

The BRDF model adopts a shallow neural network with one hidden layer to map the three geometrical angles and the surface roughness to the reflection intensity, as shown in Fig. 3(b). It is trained with 271,728 data points collected from the experimental procedures described above. After 100 epochs of learning, the coefficient of determination (R-value) reaches 0.98, and the mean squared error (MSE) is reduced to the level of 10^{-6}.

The training results of our modified BRDF model are plotted in Fig. 4. In Fig. 4(a), the surface roughness Ra is fixed at 0.55 μm, while $\phi_i - \phi_r$ is set to 0°. The contour plot shows the output of normalized reflection intensity under different zenith angles. As θ_i and θ_r increase, the intensity drops rapidly, which is consistent with the specular surface reflectance effect. ϕ_i and ϕ_r have almost the same influence on the output intensity, as the material is isotropic regarding the reflectance property. Fig. 4(b) shows the BRDF output with the change of θ_i and Ra when $\phi_i - \phi_r$ and ϕ_r are fixed to 0° and 30°, respectively. When Ra is small, the surface tends to be specular, so the output intensity drops more rapidly with the increase in θ_i. As Ra increases, the surface resembles more of a matte finish, so the output intensity is less sensitive to the change of illumination and viewing angles.

2.3. Rendering of training datasets

Three components are necessary for our DPPS framework to generate the training dataset. The point-light position calibration is described in Section 2.1. A data-driven and physics-based BRDF model is detailed in Section 2.2. In this section, the detailed algorithms are described for rendering realistic reflection images with quality variances.

The dataset generation module uses 10 random 3D objects from the Blobby library [57]. They are randomly rotated to 500 orientations and cropped with a size 128 × 128 associated with a physical dimension of 10 cm by 10 cm. The orientation angles (α_y, α_x) are defined in the range of (0, 360°). Each shape is then randomly assigned with two values of surface roughness in the range from 0.52 μm to 4.07 μm, resulting in a total of 10×500×2 = 10,000 sets of data. The corresponding variation of the height is in the range from −5cm to 10 cm. The rendering module then generates 96 reflection images for each object shape under 96 point-light illumination conditions.

For each Blobby shape in the set, we derive the surface normal and surface height information for each point from the geometry of the shape. Then for a given calibrated point-light position $\boldsymbol{L_k}$, a point on the target surface $\boldsymbol{X_P}$, and the fixed camera position $\boldsymbol{X_c}$, the viewing vector $\vec{v}$ and the illumination vector $\vec{l}$ can be directly calculated by normalizing vector $\boldsymbol{X_c} - \boldsymbol{X_P}$ and $\boldsymbol{L_k} - \boldsymbol{X_P}$. The three geometric angles $\theta_i, \theta_r, \phi_i - \phi_r$ are then calculated according to Eq. (5). The trained BRDF model takes the three input angles and the assigned surface roughness to output the reflectance intensity at each surface point. To account for the light attenuation, the distance between the light position $\boldsymbol{L_k}$ and the surface point $\boldsymbol{X_P}$ is calculated according to Eq. (7). The attenuation factor q is taken as 3, since the LED can be approximated as an ideal point-light source [58]. The reflection intensity I at point $\boldsymbol{X_P}$ on the surface under the illumination of a point-light at $\boldsymbol{L_k}$ can be calculated by

$$I_{X_P} = \frac{I_0}{\|\boldsymbol{L_k} - \boldsymbol{X_P}\|^2} BRDF(\theta_i, \theta_r, \phi_i - \phi_r, Ra) \tag{7}$$

For each rendered image, we normalize the intensity by rescaling the largest pixel intensity to 255. This approach yields a better training

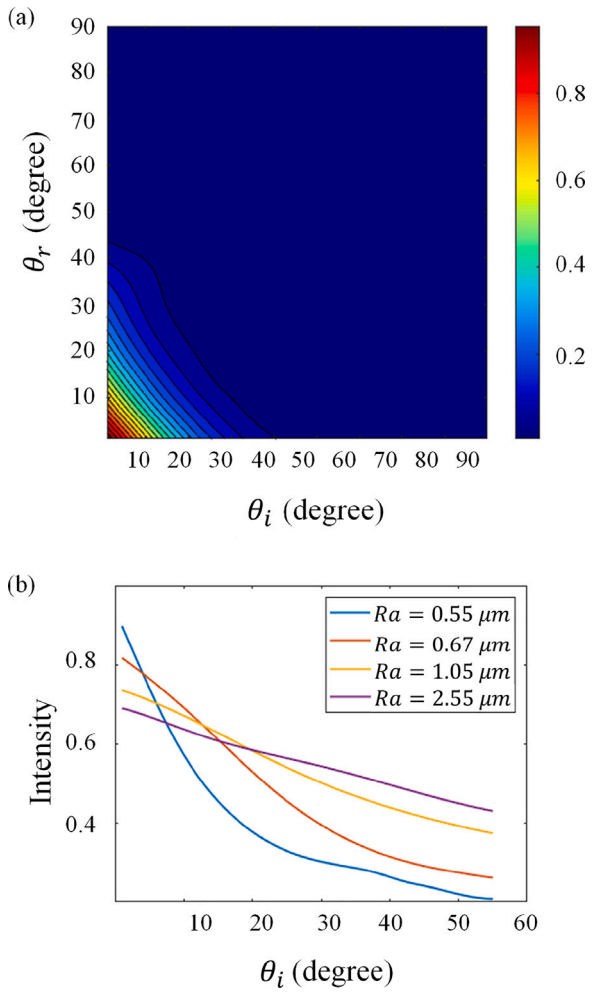

Fig. 4. (a) Contour plot of the BRDF model for different zenith angles; (b) plot of the BRDF model for different surface roughness.

performance of DPPS. After scaling, Gaussian noises with a zero mean and a standard deviation from 10^{-4} to 10^{-2} are applied to each rendered image to simulate the image capture noises.

To further increase the robustness and adaptivity of DPPS, we deliberately generate data with quality variances, including over-exposed image sets with a larger intensity factor I_0 (12.5% of the total samples), image sets with varying light intensities among the 96 light sources (6.25% of the total samples); image sets rendered under varying point-light positions to account for the calibration uncertainties (6.25% of the total samples). We summarize our dataset generation and rendering algorithms in Table 1.

The rendering procedures and examples of rendered image sets are shown in Fig. 5. Two rendered image sets are plotted for the same object under different point-light illumination and with two assigned surface roughness of Ra = 0.55 μm and Ra = 0.84 μm. It should be noted that these two surface roughness values are not the same as the calibrated artifacts but interpolated by the trained BRDF.

2.4. Architecture of DPPS and training details

In the proposed DPPS, we develop a two-channel CNN to directly extract the height map and surface normal map in an end-to-end manner, so the normal-map-targeted 2.5D reconstruction is extended to full 3D reconstruction. The structure of the CNN preserves the intrinsic character of photometric stereo while avoiding the ill-posed surface normal integration problem.

Table 1
Image rendering procedures.

1. For 10 Blobby objects, rotate each object with 25×20 rotation angles (α_y, α_x), and randomly crop to a size of 128×128 with the surface normal and height maps as the ground truth.
2. **for** k in 96 point-light sources
   for P in all points on the 128×128 grid
   Calculate:

$$\theta_i = \left\langle \vec{l}, \vec{n} \right\rangle$$

$$\theta_r = \langle \vec{v}, \vec{n} \rangle$$

$$\phi_i - \phi_r = \left\langle \vec{l} - (\vec{l} \cdot \vec{n})\vec{n}, \vec{v} - (\vec{v} \cdot \vec{n})\vec{n} \right\rangle$$

   Calculate: image intensity

$$I_{X_P} = \frac{I_0}{||\boldsymbol{L_k} - \boldsymbol{X_P}||^2} BRDF(\theta_i, \theta_r, \phi_i - \phi_r, Ra)$$

   end for
   end for
3. **Apply** Gaussian blur to the images.
4. **Normalize** the image intensity.
5. **Add** quality variations:
   Over exposure: $I_0 \in (1200, 1600)$
   Varying light intensity:

$$I_k \to \beta I_k, \ where\ \beta \in (0.95, 1.05)$$

   Varying point-light position:

$$\boldsymbol{L_k} = \boldsymbol{L_k} + [a, b, c]\ where\ a, b, c \in (-1\ cm, 1\ cm)$$

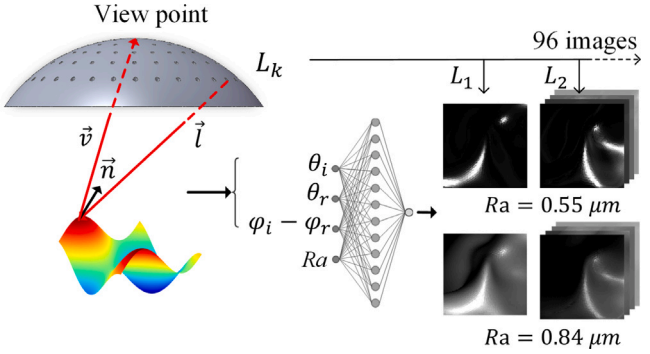

Fig. 5. Schematics and examples of image data rendering.

The two-channel encoder–decoder CNN structure is illustrated in Fig. 6. The CNN takes 96 images with a frame size of 128 × 128 and then concatenates them into a 96 × 128 × 128 tensor. The input tensor will go through a series of convolutional layers, which serve as an encoder in the widely adopted encoder–decoder CNN structure [59,60]. In the encoder stage, we stack multiple ResNet BasicBlocks [61] with increasing depth and decreasing feature size. The encoder stage allows CNNs to extract high-dimensional features from the sparse information in the input images. The resultant feature map of the encoder will be used by two parallel decoder channels as the input. The first output channel predicts a 3 × 128 × 128 normal map, which contains three channels of surface normal directions, while the other output channel predicts the height map with a size of 1 × 128 × 128. In the decoder stage, the depth of the feature map will decrease while its size will increase until it reaches 256 × 256. A final convolutional layer will output the tensor with the desired size of 128 × 128. In this part, the height and surface normal information is recovered from the high-dimensional feature map. All modules in the decoder stage include a convolutional or deconvolutional layer, a batch normalization layer,

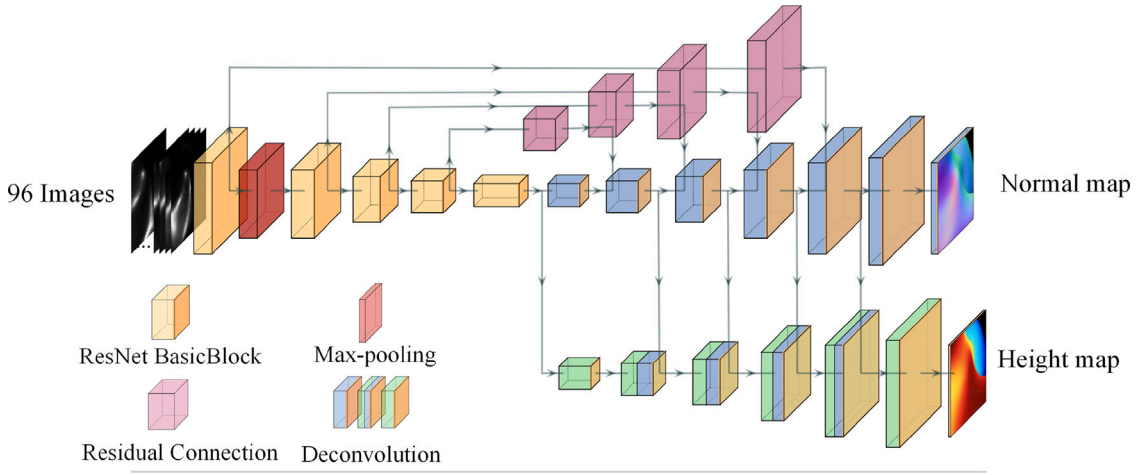

Fig. 6. DPPS Architecture: multi-channel CNN for end-to-end surface normal and height map reconstruction.

and a LeakyReLU layer [62] with a slope of 0.1 for negative values. Since the captured intensity of an arbitrary pixel depends more on surface normals than locations, we concatenate feature maps from the normal map output channel to the height map channel to facilitate the height map prediction. We also apply multiple residual connection layers by adding the feature maps of ResNet BasicBlocks to the feature maps in the deconvolutional layers. The residual connection layers allow the preservation of detailed features between the encoder and decoder stages. Instead of concatenating the feature maps to the deconvolutional layers, we use element-wise addition for residual connection to increase the depth of CNN without enlarging the CNN dimensions with extra parameters [63]. The loss function for DPPS is defined as the mean squared error (MSE) between the prediction and the ground truth. To train the CNN, we choose Adam [64] as the optimization method, which can adaptively change the learning rate according to the current gradient, thus being computationally efficient. The two momentum parameters for Adam are set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

A total of 10,000 sets of data, including clean data and quality variants, are divided into three parts: 7,000 in the training set, 1,500 in the validation set to monitor the overfitting, and 1,500 in the test dataset to evaluate the model performance. The CNNs are built on Pytorch (version 1.6.0) [65]. The training is first performed on clean data for 50 epochs with a learning rate of 0.001. Then we add the data with quality variations into both the training and validation sets and train the models for another 50 epochs. The learning rate is then reduced to 10–5 after 50 epochs of training. This multi-step training method speeds up the training process compared with directly training on the whole dataset and can effectively avoid overfitting.

The model size of DPPS is 100.5 megabytes. The average run time is 0.056 s for one reconstruction on a graphics processing unit (GeForce RTX 2080 Ti, *NVIDIA*, USA). In practice, the run time for reconstruction is nearly negligible compared with the image capturing and processing time.

3. Results and discussion

In this section, we present the performance of DPPS tested on both rendered datasets and in real experiments. The results of both the surface normal map and height map for the test dataset are first analyzed. To demonstrate the advantage of end-to-end prediction of height maps, we compare the height map reconstruction results with the integrated height maps from the surface normal map that is conventionally adopted in current photometric stereo practice. Then, the performance of DPPS on a standard spherical shape is compared with the state-of-the-art, including those based on deep learning approaches. Furthermore, we test the performance of DPPS on real metal objects and compare the results with those obtained from commercial 3D scanners.

Table 2
Error comparison of the validation set, test dataset and two reconstruction samples.

	Mean error for normal map	Mean angular error	Mean error for height map
Validation dataset	0.0089	1.394°	0.0415 cm
Test dataset	0.0211	2.961°	0.1155 cm
Example 1	0.0052	1.131°	0.0514 cm
Example 2	0.0060	1.291°	0.0640 cm

3.1. Results on the rendered dataset

The quantitative analysis of DPPS of the prediction errors for the validation and test datasets is presented in Table 2. We show two sample reconstructions from the test dataset in Fig. 7 with their mean prediction errors listed in Table 2. We use the mean surface normal difference and the mean height difference to evaluate the prediction accuracy of the surface normal and height maps, respectively. The corresponding mean angular error is calculated based on the surface normal error for reference. For the test dataset results, the mean angular error is 2.961°, while the average error of the height map is 0.1155 cm, or below 1% of the overall height dimension of the object.

We notice that the prediction errors in the test dataset are two times larger than those in the validation set for the angular error and more than ten-fold for the height map prediction. This is largely due to the quality variations included in the dataset, which randomly varies the illumination light intensities and positions in a small range to represent the measurement uncertainties. When we remove the quality variants from the dataset, the test dataset performance converges to the validation set accuracy nicely. Though there are some discrepancies between the validation and test dataset performance, we believe the results presented are more realistic performance indices.

Two reconstruction examples from the test dataset are plotted in Fig. 7. The reference object shapes, rendered image samples (2 out of 96 image inputs), surface normal maps, height maps, and the corresponding error maps are all illustrated. The error map for the surface normal prediction is plotted as an RGB image, where the color information indicates both the direction and magnitude. The largest errors for both the surface normal and height maps come from the surface discontinuity. The corresponding quantitative analysis is summarized in Table 2.

One advantage of DPPS is its ability to directly output the height map from image inputs to avoid the direct surface normal integration problem. In Fig. 7(b), we compare the height map predictions from DPPS and from the traditional surface normal integration method. We used the Frankot–Chellappa method [66] as the baseline comparison, which is a path-independent integration method to achieve high

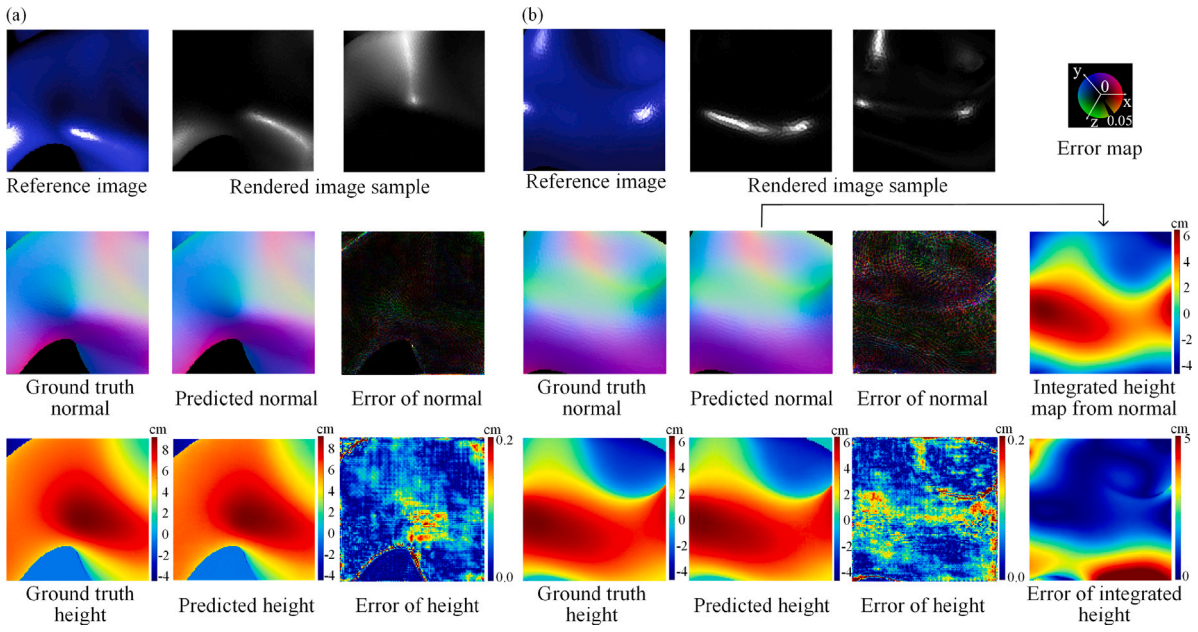

Fig. 7. Two examples of DPPS results on the test dataset: (a) a convex shape; (b) a concave shape and the comparison with surface normal integration.

Table 3
Comparison of reconstruction results of a sphere.

Model	Mean angular error
L2	4.10°
ST14	1.74°
IA14	3.34°
DPSN	2.02°
PS-FCN	2.82°
DPPS	1.80°

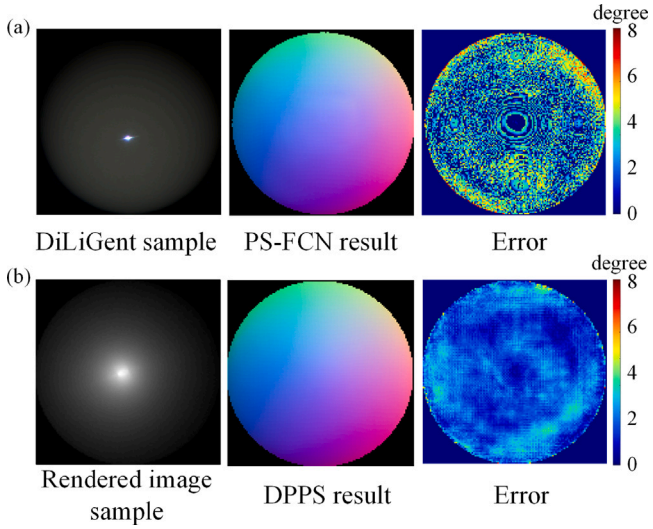

Fig. 8. (a) An example image of DiLigenT sphere shape data, reconstructed normal map by PS-FCN [34] and error map; (b) an example image of rendered data for the same sphere shape with point-light sources, reconstructed normal map by DPPS and error map.

computational efficiency while preserving the details of surface discontinuities. The calculated height from the ground truth normal map exhibits severe global shape deformation. The largest error is 5.7 cm, several orders of magnitude larger than the direct prediction from the second channel of DPPS. If there are additional estimation errors in the surface normal, the prediction error will be accumulated even further to deteriorate the height map prediction. In comparison, DPPS not only saves computation of normal integration but also achieves higher accuracy for the height map reconstruction.

We also compare the performance of DPPS to the state-of-the-art, including both traditional, machine learning, and deep learning approaches. We use a popular photometric stereo benchmark 'sphere shape' to compare the surface normal prediction accuracy, as shown in Table 3. L2 [26] belongs to the traditional photometric stereo category. ST14 [37] and IA14 [39] are based on the conventional machine learning approaches. DPSN [43] and PS-FCN [45] are deep-learning-based approaches. The mean angular error is calculated for the reconstruction of DPPS and other calibrated or semi-calibrated photometric stereo methods. DPPS's prediction accuracy is comparable to or even better than the reference methods in this limited case.

It is worth noting that: (1) All other reference photometric stereo methods do not give height map results for comparison, so we only compare the accuracy of the mean angular error of the surface normal. (2) Only DPPS assumes point-light illumination, which not only can be better implemented in a manufacturing setting but also provide potentially higher prediction accuracy due to the calibrated light positions (a semi-calibrated configuration in DPPS to be more precise).

To further visualize the performance comparison, we plot the 3D reconstruction results from PS-FCN [45] on a spherical object from the DiLiGent benchmark [50] and the DPPS results on the same object rendered with our generation method in Fig. 8. Both PS-FCN and DPPS work in a semi-calibrated manner with 96 input images for reconstruction. Due to the different assumptions of illumination conditions (parallel vs. point-light), we cannot directly use DiLiGent images but render our own image sets based on the calibrated light conditions as described in Section 2. DPPS result exhibits smaller angular errors throughout the surface normal map.

3.2. Performance analysis of experiments

We further conduct experimental validation to evaluate the real-life performance of DPPS with simple and complex shapes. The chosen object shapes are very different from the training, validation, and test dataset to test the generalization capabilities of DPPS. It should be

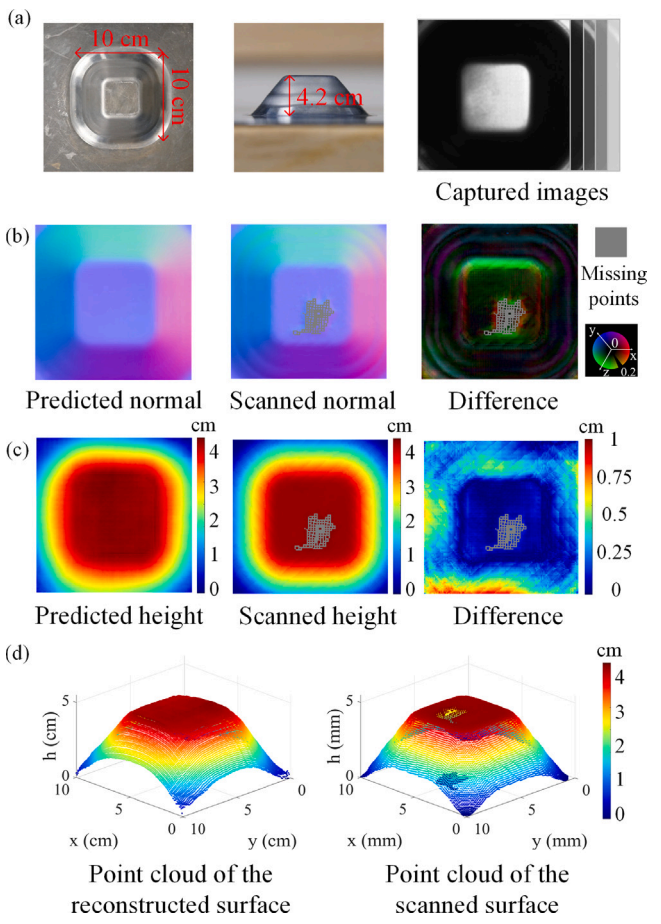

Fig. 9. (a) Pictures of the metal forming part for reconstruction and examples of experimentally captured images; comparison of the predicted and scanned results: (b) surface normal map and (c) height map; (d) comparison of 3D point cloud of the reconstructed surface by DPPS and the scanning result.

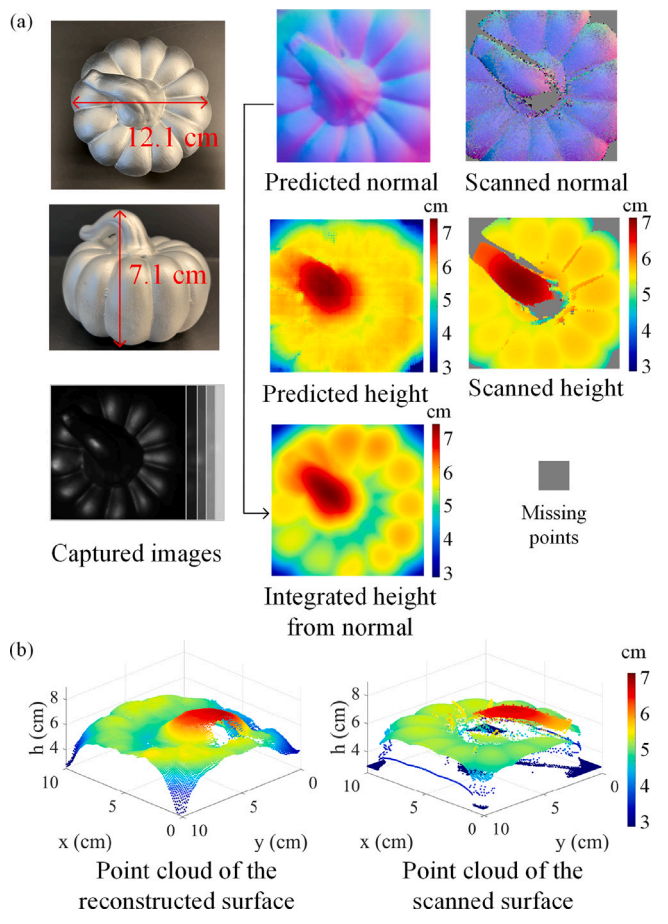

Fig. 10. (a) 3D reconstruction results of a pumpkin sculpture with DPPS and a 3D camera; (b) comparison of point cloud reconstructed by DPPS and a 3D camera.

noted that all previous CNN-based photometric stereo methods have only been tested on rendered or standard datasets. Their performance in real scenarios is unknown. This is the first work to evaluate the real-life performance of deep-learning based photometric stereo and to compare the results, particularly the height maps, with commercial 3D scanners.

The first example is an aluminum sheet metal part made from incremental forming [67]. The formed part has a truncated pyramid shape with an approximate size of 35 cm×35 cm×5 cm, as shown in Fig. 9(a). The measurement region is located at the center with an area size of 10 cm×10 cm. The surface roughness Ra is measured to be 0.72 μm, which is within the calibrated range of 0.52 μm to 4.07 μm. To be noted, the measured Ra value is unknown to our DPPS model and has not been used in the 3D reconstruction. Ninety-six images are captured with each LED lit on in sequence. The images are cropped and fed into DPPS that outputs the surface normal map and height map, as shown in Fig. 9(b) and (c).

We use a laser scanner (AS1, *Hexagon*, USA) to provide the reference surface normal and height maps (Fig. 9(b),(c)) to evaluate the performance of DPPS. The scanner is equipped with an encoded arm system (RA-7535 Si, *Hexagon*, USA), so the scanning is performed at multiple angles to ensure that the projected laser is perpendicular to the surface. However, one limitation of the laser scanner is its ability to deal with highly reflective surfaces. We directly scan the parts without applying an additional matte spray, which results in some missing data points where the scanner fails to register a point. It also demonstrates the advantage of DPPS in handling specular reflection and its potential for in-process metrology of metallic parts.

The comparison of predicted and scanned results is shown in Fig. 9(b) and 9(c). The predicted normal map from DPPS shows a good match with the scanned result except for the missing data points. The largest error comes from the edge region. This is mainly because the sharp edges were over-exposed in various lighting conditions, which produced a discontinuous intensity field in the captured image. As for the height map, the predicted height map is able to recover the object's overall shape with some local distortion. The average difference between the predicted and scanned height maps is 0.317 cm. Part of the error is suspected to come from surface textures of the formed part, which results in anisotropic and uneven reflectance on the surface, as shown in Fig. 9(a). We also compare the 3D reconstruction result and scanned result in 3D point cloud format, as shown in Fig. 9(d), which indicates that DPPS reconstructs a more complete 3D surface than direct laser scanning result.

The second example part is a pumpkin-shaped sculpture painted with a metallic coating, as shown in Fig. 10. The sculpture size is 12.1 cm×12.1 cm×7.1 cm. The reconstruction region is the center area with an area size of 10 cm×10 cm. The sculpture has both complex convex and concave shapes, while the pumpkin stem creates discontinuous height profiles and occlusion. The reference results are measured using an industrial 3D camera (PhoXi 3D Scanner XS, *Photoneo*, USA), which utilizes structured light projection. The depth of view of the 3D camera is 161 205 mm. It is positioned 160 mm above the sculpture to ensure the top surface is within the focal region.

For the 3D camera, the structured light is projected from a single projector position, so the shadow effect caused by occlusion and concavity will cause missing data points in the scanned result. In addition, since the surface normal map is not directly provided by the point cloud

output, the reference surface normal is manually calculated using local fitting from the point cloud data.

The measurement results from DPPS and the 3D camera are compared in Fig. 10. The surface normal map predicted by DPPS shows finer details of the reconstructed pumpkin, even for the steep slope on the stem region where the 3D camera failed to register data points. This is more obvious in the 3D point cloud comparison, as shown in Fig. 10(b). The height map predicted by DPPS shows high consistency with the scanned result except for the concave regions on the pumpkin surface, where the camera outputs abnormal values. Though shadow also exists due to occlusion by the stem, DPPS results are not affected by the shadows. The height map from surface normal integration is also plotted for comparison. The integrated height map is able to recover the overall shape but cannot preserve the height jump in the stem region.

The above experimental validation suggests that DPPS is able to reconstruct complex metal surfaces even when the shape is very different from those in the training dataset. The reconstruction results are in high consistency with the commercial 3D scanners, while DPPS shows much greater potential for in-process metrology capabilities.

4. Conclusion

In this paper, we present DPPS, a deep-learning based point-light photometric stereo for 3D reconstruction of metal surfaces with unknown surface roughness. DPPS takes 96 image inputs captured under 96 point-light LED illumination and reconstructs both the surface normal and height maps in an end-to-end manner.

The major contributions of the work can be summarized as follows:

1. A new dataset generation procedure combining both physics-based and data-driven approaches is developed. It includes three novel modules: (a) point-light source calibration, (b) a physics-based and data-driven reflectance model, and (c) realistic point-light reflection rendering. The proposed method is able to render comprehensive image sets under location-dependent illumination, a range of different surface roughness, and quality variations.
2. An end-to-end approach is proposed to directly predict the height map from input images from a multi-channel CNN, significantly reducing distortions when calculating the height map from surface normals. The ill-posed surface integration problem is avoided by introducing the position as an extra dimension to photometric stereo, enabling the CNN to learn the height map directly
3. Point-light photometric stereo is introduced in a deep learning framework, which utilizes the nonlinear mapping ability of deep CNNs to implicitly determine the reflection properties determined by the surface roughness and point-light source locations.
4. Though only trained with synthetic datasets, DPPS shows good performance on both rendered test datasets and experimental conditions. DPPS outperforms the state-of-the-art in a sphere reconstruction test with a mean angular error of 1.80°. In real-world tests, DPPS is able to reconstruct a metal forming part and a complex pumpkin sculpture with millimeter level accuracy on height prediction and less than 10° angular errors in surface normal prediction. The results are directly compared with commercial 3D scanners (laser scan and structured light projection).
5. DPPS is a promising computer vision 3D reconstruction method for future in-process 3D metrology of various manufacturing processes, such as metal additive manufacturing. By utilizing our data-driven and physics-based data generation approach, DPPS can be extended to different scales, lighting conditions, and materials.

The proposed DPPS provides insight into utilizing deep learning to overcome the intrinsic limitation of 3D reconstruction methods. It establishes a new solution for dealing with reflectance property with a neural network to quickly render comprehensive datasets for training deep learning models. While applying this method in 3D reconstruction, there are multiple factors that need to be considered. (1) The current model is sensitive to environmental lighting and has to be used in a dark setting. (2) Since the normal map and height map are reconstructed at the same time, the reconstruction range is fixed into a 10 cm × 10 cm region which is not directly scalable. Our future work includes designing a lighting setup with much brighter illumination to eliminate the influence of environmental light and designing a compact system for reconstructing details of metal surfaces, which is applicable for the in-process monitoring of powder bed fusion processes. We are also working on replacing the dome with a robot arm holding and moving a point light to provide different lighting conditions, which largely reduces the system complexity of DPPS.

CRediT authorship contribution statement

Ru Yang: Conceptualization, Software, Validation, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Yaoke Wang:** Investigation, Validation, Visualization, Writing – review & editing. **Shuheng Liao:** Conceptualization, Methodology. **Ping Guo:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Model and data are released on GitHub: https://github.com/RuYangNU/point-light-photometric-stereo.

Acknowledgments

This research was supported by the start-up fund from McCormick School of Engineering, Northwestern University, USA; and the National Science Foundation, USA under Grant number EEC-2133630 and CNS-2229170. The authors would like to thank Dohyun Leem for providing the metal forming part and Guangze Li for providing the 3D camera.

Appendix

A.1. *How we get 271,728 data points for training the brdf function?*

When the viewing angles are 0°, 22° and 54°, the metal ball in the captured image has diameters of 159 pixels, 151 pixels and 153 pixels. Since most parts of the captured image are dark under the single direction lighting, we utilize the symmetric property of the BRDF model to reduce a large amount of redundant data by taking the upper-left quarter of the pixels. This leads to 5,097, 4,606, 4,727 data points for each image taken under 0°, 22° and 54°. And when the viewing angle is 0°, which is right on top of the metal ball and θ_i equals 0°, the incident light is partially occluded by the camera, so we discard a rectangular region (37 × 66) in that image. So the total amount of data points equals $(5,097 + 4,606 + 4,727) \times 19 - 37 \times 66 = 271,728$.

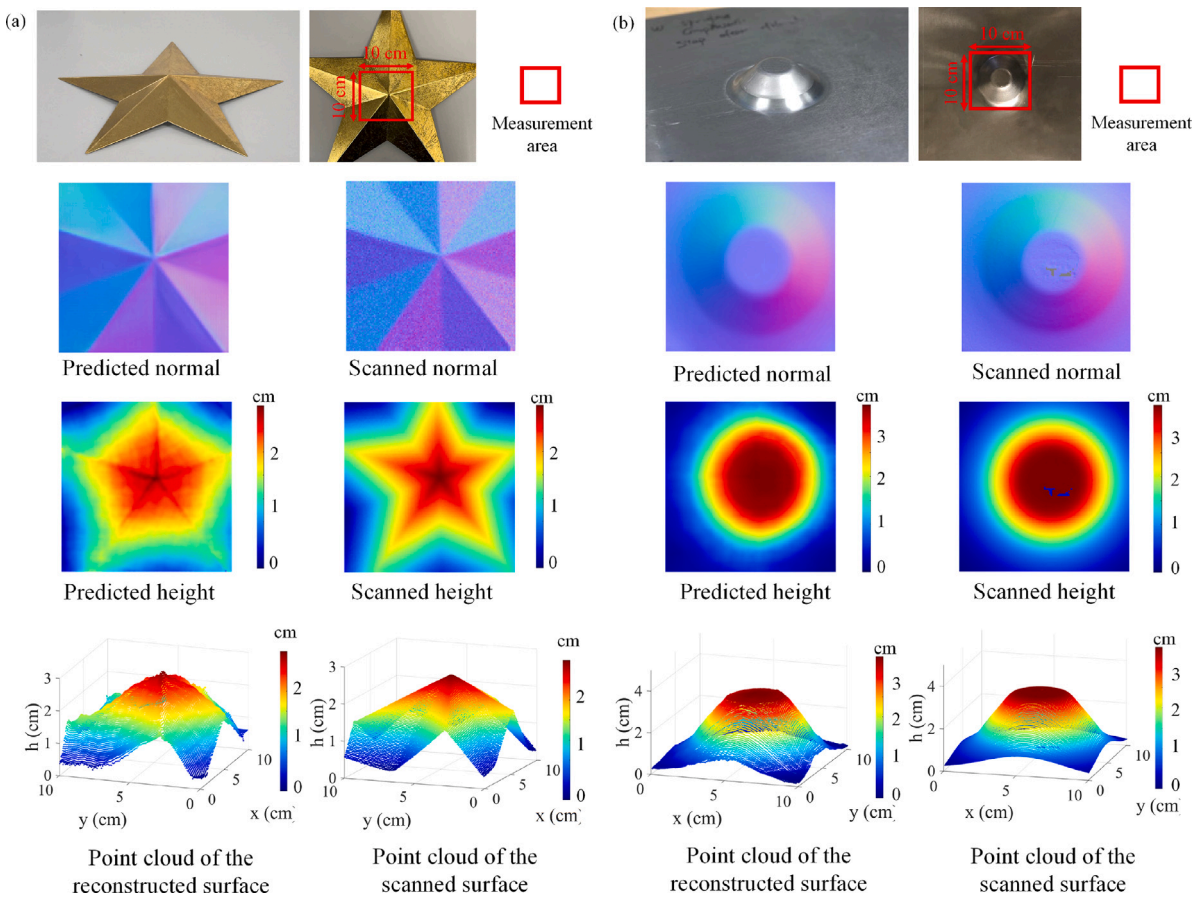

Fig. A.1. (a) Comparison of DPPS reconstruction and scanning results on a 3D bronze star; (b) comparison of DPPS reconstruction and scanning results on a cone-shaped aluminum forming part.

A.2. Additional experimental verification

In this section, we show two experiment results to analyze the performance of DPPS in addition to the two experiments in Section 3.2. The first example is a 3D start made of bronze material and the second example is a cone-shaped aluminum forming part. Both objects are larger than 10 cm × 10 cm in x-y dimension, so the reconstruction area is the maximum scanning area of DPPS, that is 10 cm × 10 cm. In Fig. A.1 we demonstrate the normal map, height map and the point cloud comparison of DPPS reconstruction and scanning results.

References

[1] J.-J. Aguilar, F. Torres, M. Lope, Stereo vision for 3D measurement: Accuracy analysis, calibration and industrial applications, Measurement 18 (4) (1996) 193–200.

[2] S.M. Snyder, E. Schneider, Estimation of mechanical properties of cortical bone by computed tomography, J. Orthop. Res. 9 (3) (1991) 422–431.

[3] R.J. Hocken, P.H. Pereira, et al., Coordinate Measuring Machines and Systems, Vol. 2, CRC press Boca Raton, 2012.

[4] S. Catalucci, N. Senin, D. Sims-Waterhouse, S. Ziegelmeier, S. Piano, R. Leach, Measurement of complex freeform additively manufactured parts by structured light and photogrammetry, Measurement 164 (2020) 108081.

[5] C. Wu, J. Fan, Q. Wang, D. Chen, Machining accuracy improvement of non-orthogonal five-axis machine tools by a new iterative compensation methodology based on the relative motion constraint equation, Int. J. Mach. Tools Manuf. 124 (2018) 80–98.

[6] H. Mostafanezhad, H.G. Menghari, S. Esmaeili, E.M. Shirkharkolaee, Optimization of two-point incremental forming process of AA1050 through response surface methodology, Measurement 127 (2018) 21–28.

[7] H. Attar, S. Ehtemam-Haghighi, D. Kent, M.S. Dargusch, Recent developments and opportunities in additive manufacturing of titanium-based matrix composites: A review, Int. J. Mach. Tools Manuf. 133 (2018) 85–102.

[8] F. Micari, G. Ambrogio, L. Filice, Shape and dimensional accuracy in single point incremental forming: state of the art and future trends, J. Mater Process. Technol. 191 (1–3) (2007) 390–395.

[9] B.K. Belur, R.V. Grandhi, Geometric deviations in forging and cooling operations due to process uncertainties, J. Mater Process. Technol. 152 (2) (2004) 204–214.

[10] E.P. Baltsavias, A comparison between photogrammetry and laser scanning, ISPRS J. Photogramm. Remote Sens. 54 (2–3) (1999) 83–94.

[11] C. Tang, G.Y. Tian, J. Wu, Segmentation-oriented compressed sensing for efficient impact damage detection on CFRP materials, IEEE/ASME Trans. Mechatronics 26 (5) (2020) 2528–2537.

[12] M. Sokac, I. Budak, M. Katic, Z. Jakovljevic, Z. Santosi, D. Vukelic, Improved surface extraction of multi-material components for single-source industrial X-ray computed tomography, Measurement 153 (2020) 107438.

[13] C. Nicolae, E. Nocerino, F. Menna, F. Remondino, Photogrammetry applied to problematic artefacts, Int. Arch. Photogram., Remote Sens. Spat. Inf. Sci. 40 (5) (2014) 451.

[14] Y.M. Amir, B. Thörnberg, High precision laser scanning of metallic surfaces, Int. J. Opt. 2017 (2017).

[15] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, CVPR'06, IEEE, 2006, pp. 519–528.

[16] Y. Furukawa, B. Curless, S.M. Seitz, R. Szeliski, Towards internet-scale multi-view stereo, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 1434–1441.

[17] J.J. Koenderink, A.J. Van Doorn, Affine structure from motion, J. Opt. Soc. Amer. A 8 (2) (1991) 377–385.

[18] Ž. Santoši, I. Budak, V. Stojaković, M. Šokac, Đ. Vukelić, Evaluation of synthetically generated patterns for image-based 3D reconstruction of texture-less objects, Measurement 147 (2019) 106883.

[19] A.H. Ahmadabadian, A. Karami, R. Yazdan, An automatic 3D reconstruction system for texture-less objects, Robot. Auton. Syst. 117 (2019) 29–39.

[20] F. Menna, E. Nocerino, D. Morabito, E. Farella, M. Perini, F. Remondino, An open source low-cost automatic system for image-based 3D digitization, Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci. 42 (2017) 155.

[21] A. Hosseininaveh, R. Yazdan, A. Karami, M. Moradi, F. Ghorbani, A low-cost and portable system for 3D reconstruction of texture-less objects, Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci. 40 (1) (2015) 327.

[22] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, Nerf: Representing scenes as neural radiance fields for view synthesis, Commun. ACM 65 (1) (2021) 99–106.

[23] Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, U. Neumann, Point-nerf: Point-based neural radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5438–5448.

[24] F. Willomitzer, G. Häusler, High resolution single-shot 3D imaging with the "3D movie camera", in: 3D Image Acquisition and Display: Technology, Perception and Applications, Optica Publishing Group, 2018, 3M3G–2.

[25] J.W. Horbach, T. Dang, 3D reconstruction of specular surfaces using a calibrated projector–camera setup, Mach. Vis. Appl. 21 (3) (2010) 331–340.

[26] R.J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.

[27] P. Zhao, N. Gao, Z. Zhang, F. Gao, X. Jiang, Performance analysis and evaluation of direct phase measuring deflectometry, Opt. Lasers Eng. 103 (2018) 24–33.

[28] Z. He, P. Li, X. Zhao, L. Kang, S. Zhang, J. Tan, Chessboard-like high-frequency patterns for 3D measurement of reflective surface, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.

[29] M.B. Shaikh, D. Chai, RGB-D data-based action recognition: A review, Sensors 21 (12) (2021) 4246.

[30] Y. Xu, F. Gao, X. Jiang, A brief review of the technological advancements of phase measuring deflectometry, PhotoniX 1 (1) (2020) 1–10.

[31] S. Zhang, High-speed 3D shape measurement with structured light methods: A review, Opt. Lasers Eng. 106 (2018) 119–131.

[32] J.J. Clark, Active photometric stereo, in: CVPR, Vol. 92, 1992, pp. 29–34.

[33] T. Yoda, H. Nagahara, R.-i. Taniguchi, K. Kagawa, K. Yasutomi, S. Kawahito, The dynamic photometric stereo method using a multi-tap CMOS image sensor, Sensors 18 (3) (2018) 786.

[34] W.M. Silver, Determining Shape and Reflectance Using Multiple Images (Ph.D. thesis), Massachusetts Institute of Technology, 1980.

[35] S.J. Koppal, Lambertian reflectance, in: Computer Vision: A Reference Guide, Springer, 2020, pp. 1–3.

[36] A.S. Georghiades, Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo, in: Computer Vision, IEEE International Conference on, Vol. 3, IEEE Computer Society, 2003, p. 816.

[37] B. Shi, P. Tan, Y. Matsushita, K. Ikeuchi, Bi-polynomial modeling of low-frequency reflectances, IEEE Trans. Pattern Anal. Mach. Intell. 36 (6) (2013) 1078–1091.

[38] H.-S. Chung, J. Jia, Efficient photometric stereo on glossy surfaces with wide specular lobes, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[39] S. Ikehata, K. Aizawa, Photometric stereo using constrained bivariate regression for general isotropic surfaces, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2179–2186.

[40] S. Ikehata, D. Wipf, Y. Matsushita, K. Aizawa, Robust photometric stereo using sparse regression, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 318–325.

[41] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.

[42] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, Y. Ma, Robust photometric stereo via low-rank matrix completion and recovery, in: Asian Conference on Computer Vision, Springer, 2010, pp. 703–717.

[43] H. Santo, M. Samejima, Y. Sugano, B. Shi, Y. Matsushita, Deep photometric stereo network, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 501–509.

[44] S. Ikehata, CNN-PS: CNN-based photometric stereo for general non-convex surfaces, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–18.

[45] G. Chen, K. Han, K.-Y.K. Wong, PS-FCN: A flexible learning framework for photometric stereo, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–18.

[46] G. Chen, K. Han, B. Shi, Y. Matsushita, K.-Y.K. Wong, Self-calibrating deep photometric stereo networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8739–8747.

[47] G. Chen, M. Waechter, B. Shi, K.-Y.K. Wong, Y. Matsushita, What is learned in deep uncalibrated photometric stereo? in: European Conference on Computer Vision, Springer, 2020, pp. 745–762.

[48] J. Li, A. Robles-Kelly, S. You, Y. Matsushita, Learning to minify photometric stereo, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7568–7576.

[49] Y. Ju, J. Dong, S. Chen, Recovering surface normal and arbitrary images: A dual regression network for photometric stereo, IEEE Trans. Image Process. 30 (2021) 3676–3690.

[50] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, P. Tan, A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3707–3716.

[51] R. Mecca, F. Logothetis, I. Budvytis, R. Cipolla, Luces: A dataset for near-field point light source photometric stereo, 2021, arXiv preprint arXiv:2104.13135.

[52] W. Matusik, A Data-Driven Reflectance Model (Ph.D. thesis), Massachusetts Institute of Technology, 2003.

[53] Y. Quéau, J.-D. Durou, J.-F. Aujol, Normal integration: A survey, J. Math. Imaging Vision 60 (4) (2018) 576–593.

[54] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, W. Matusik, Dynamic shape capture using multi-view photometric stereo, in: ACM SIGGRAPH Asia 2009 Papers, 2009, pp. 1–11.

[55] B. Haefner, S. Peng, A. Verma, Y. Quéau, D. Cremers, Photometric depth super-resolution, IEEE Trans. Pattern Anal. Mach. Intell. 42 (10) (2019) 2453–2464.

[56] A. Ryer, U. Light, V. Light, Light measurement handbook, Citeseer, 1997.

[57] M.K. Johnson, E.H. Adelson, Shape estimation in natural illumination, in: CVPR 2011, IEEE, 2011, pp. 2553–2560.

[58] B.B. Baker, E.T. Copson, The Mathematical Theory of Huygens' Principle, Vol. 329, American Mathematical Soc., 2003.

[59] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[60] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[61] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[62] A.L. Maas, A.Y. Hannun, A.Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. Icml, vol. 30, Citeseer, 2013, p. no. 3.

[63] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[64] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[65] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., Pytorch: An imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).

[66] R.T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 439–451.

[67] D. Leem, N. Moser, H. Ren, M. Mozaffar, K.F. Ehmann, J. Cao, Improving the accuracy of double-sided incremental forming simulations by considering kinematic hardening and machine compliance, Procedia Manuf. 29 (2019) 88–95.