

Continuous Material Reflectance Map for Deep Photometric Stereo

NICOLAS PROUTEAU^{1,2}, CLÉMENT JOUBERT¹, BENJAMIN BRINGIER¹, AND MAJDI KHOUDEIR¹

¹ XLIM Institute, UMR CNRS 7252, University of Poitiers, France

² Pprime Institute, UPR CNRS 3346, University of Poitiers, France

* Corresponding author: nicolas.prouteau01@univ-poitiers.fr

Compiled January 12, 2023

Solving calibrated photometric stereo under a sparse set of lights is of great interest for real world applications. Since neural networks are showing advantages in dealing with material appearance, this paper proposes a new BRDF representation based on reflectance maps for a sparse set of lights. We discuss the optimal way to compute these BRDF based photometric stereo maps regarding to its shape, size and resolution and experimentally investigate its contribution on normal map estimation. An analysis is conducted on the training dataset to establish the BRDF data to use between measured and parametric BRDF. The proposed method is compared with state-of-the-art photometric stereo algorithms for different datasets from numerical rendering simulation, DiliGenT and from our two acquisition systems. The results show that our representation outperforms observation maps as BRDF representation for a neural network for various surface appearance on specular and diffuse areas. © 2023 Optica Publishing Group

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. INTRODUCTION

Represent, characterize and manipulate material appearance, i.e. BRDF (Bidirectional Reflectance Distribution Function), on various objects is a complex but necessary task to solve 3D reconstruction problems as calibrated photometric stereo. Indeed, visual appearance of an object is a non-linear combination of its geometry and BRDF and 3D reconstruction methods aims to correctly disentangle these two aspects of a surface.

Calibrated photometric stereo uses multiple images taken from the same viewpoint by modifying the illumination with controlled lights. This method can be considered as an inverse problem on the rendering equation to recover the geometry of an object. Calibrated photometric stereo was first introduced by Woodham in 1980 [1] and operated in a per-pixel manner computing surface normal under diffuse reflectance assumption. To overcome limitations from diffuse reflectance assumption, traditional methods were developed that can be split in three categories. The first one includes robust-estimation based approaches. Non-Lambertian pixels are considered as outliers and the remaining Lambertian pixels are used to compute the normal vector [2–6]. The second category includes reflectance model-based approaches that optimize BRDF by estimating its parameters or using its properties [7–10]. The third category includes example-based methods that use a reference object to estimate the geometry of a surface with the same reflectance [11]. For further information about traditional photometric stereo methods, Ackermann et al. [12] made a survey of these techniques.

Despite their efficiency for various object appearances, traditional methods need long processing times and cumbersome acquisition systems, i.e. high number of lights, to accurately manipulate BRDF and recover the geometry. Indeed, for a sparse set of lights, BRDF is under-sampled which makes traditional methods fail. Neural networks have proven to be effective in function estimation and are applied in photometric stereo to implicitly estimate an inverse BRDF while taking into account the cast shadows to recover the geometry.

A. Related works

Most recent methods are using deep neural networks to compute surface normals. Indeed, after Santo et al. [13] have revived interest in using deep neural networks for photometric stereo, methods were developed with differences in the network input data and in the network architecture.

Concerning the inputs, some methods use global images under different illumination directly as input. Chen et al. [14] developed an approach where a multi-branch siamese network takes photometric stereo images and associated light directions as input and gives surface normal vector. As the light directions are part of the inputs, this method is suitable with various photometric stereo systems. With the same inputs, i.e images and light directions, Cao et al. [15] developed a convolutional neural network (CNN) architecture that allows both inter-frame feature extraction and then intra-frame feature extraction using custom blocks. This approach allows a better recovery of the finest details in the surface normals keeping global information.

Moreover, other methods use pixel-wise approaches where pixel values are preprocessed to a neural network. Indeed, Santo et al. [13] used an approach where the neural network input is a measurement vector of the pixel values under different light directions and the output is the surface normal vector. Their network architecture is based on a fully connected neural network. Ikehata [16] used also a pixel-wise approach and an intermediate representation of the photometric stereo problem called the observation map. If this method is efficient when dealing with dense observation map, i.e. many light sources, the method often fails to recover correct surface normals on sparse observation map, i.e. with few light sources. Based on this approach, Li et al. [17] improved the initial observation map by adding an occlusion layer to improve surface normal estimation in cast shadows and a connection table to select only relevant pixels in an observation map. This connection table allows the network to process photometric stereo with few lights. With the same observation map input Zheng et al. [18] proposes to improve sparse observation map performance with an intermediate dense observation map in their network. The first part of the network estimate dense observation map from a sparse one and the second part learn to retrieve surface normal from the estimated dense observation map. Finally, Logothetis et al. [19] proposed a new strategy to generate observation maps for the training process with respect to real illumination phenomenons.

Therefore, it is desirable to have a deep-learning method that is capable of extracting surface normals with a system equipped with few light sources and with reasonable computing time. Global methods give good results on several objects and reflectances but can also create bias in the estimation for highly detailed object [17]. Recent pixel-wise methods using observation maps [17–19] are finding ways to solve the lack of information in sparse observation maps. Nonetheless, these methods are still limited due to the data representation.

B. Our contributions

As seen in the previous section, having a high number of lights to reconstruct an object needs time for acquisition, data processing and a cumbersome acquisition system. On the other hand, photometric stereo with few lights is an ill-posed problem for complex BRDF i.e. photometric stereo alone does not contain all information needed to disentangle geometry and BRDF. Indeed, the challenge remains to find a way to reduce the number of lights without impacting the normal map estimation. We mainly make the following contributions.

A new data representation to learn photometric stereo

We introduce a modified representation of reflectance map described by Woodham [1] as Continuous Material Reflectance map (CMR map). The main principle of this map is to represent photometric stereo data using the rendering equation. From the best of our knowledge, no recent method uses this representation as an input for a neural network. For real acquisitions and complex BRDF, CMR maps are showing distortion especially in non-linear areas as cast shadows and specular lobes. The idea is to use the information of these distortions with CNN to implicitly estimate the BRDF and then be able to recover the surface normals. Furthermore, our representation permits a predictable representation of these phenomena without discontinuity according to variations of physical parameters.

Physically-based BRDF models analysis for learning

In recent methods, neural networks are trained using data generated with BRDF models that are simulated or measured. In most cases, there is no analysis to conclude whether measured BRDFs or parametric BRDF models should be used instead. In this work, we compare the performance of our approach trained with parametric BRDF models and a database of measured BRDFs. This comparison is quantitative on DiliGenT and qualitative on our acquisition system dataset where plaster copies of real objects covered with 4 different appearances are used. Indeed, we acquired a new photometric stereo dataset with several real world objects using our two photometric stereo acquisition systems. These datasets are used to evaluate photometric stereo methods on objects with different BRDFs and a Lambertian reference with the same geometry and will be available in open access. Furthermore, we present a pixel-wise training dataset generation with randomly distributed gradients and a selected BRDF model.

2. REFLECTANCE REPRESENTATION

In practice, radiance is reflected by a surface element lit by its surrounding environment, depending on the light source and according to object reflectance. Photometric stereo methods aim to estimate a normal for each surface element from several distinct radiance values achieved by moving the light source position. Furthermore, knowing the lighting conditions and assuming an ideal sensor, we consider that object reflectance is acquired rather than radiance. Two representations are generally used to represent reflectance as a map. This section discusses the theoretical background behind these map representations.

A. Light, BRDF, radiance and pixel value

The intensity of a pixel at position (x, y) along a single light direction \mathbf{i} is described by:

$$L_o(x, y, \mathbf{o}, \mathbf{n}) = L_i(x, y, \mathbf{i}) f_r(x, y, \mathbf{i}, \mathbf{o}, \mathbf{n}) \max(\mathbf{i}\mathbf{n}, 0) \quad (1)$$

with L_o the reflected light intensity captured by the sensor, L_i the light intensity from the source, \mathbf{n} the normal of the surface, \mathbf{i} the directional light position vector, \mathbf{o} the sensor direction vector and f_r the BRDF. Modern BRDF models have a glossy and a matte component (rep. f_s and f_b):

$$f_r(x, y, \mathbf{i}, \mathbf{o}, \mathbf{n}) = f_s(x, y, \mathbf{i}, \mathbf{o}, \mathbf{n}) + f_b(x, y, \mathbf{i}, \mathbf{o}, \mathbf{n}) \quad (2)$$

where the specular component f_s generally corresponds to the Cook-Torrance BRDF model [20], while the matte component f_b is an empirically weighted Lambertian coefficient to ensure energy conservation [21] or a more sophisticated approach [22].

If the imaging system is ideal, the irradiance received by a pixel sensor is related to scene radiance L_o and converted into a digital value:

$$I(x, y) = \min \{L_o(x, y, \mathbf{o}, \mathbf{n}) + \eta, I_{max}\} \quad (3)$$

where I_{max} is the maximum numerical value depending on the number of bits for quantizing information and η is the sensor noise. If the acquisition system is calibrated and high dynamic range (HDR) acquisition is used, $I(x, y)$ is a good approximation of L_o .

The Eq. 1 expresses how the material will react to directional light on the position (x, y) on the surface. Within this expression, we focus our attention on the normal \mathbf{n} of the surface on the point (x, y) . This normal can be expressed using the gradients of the surface $S(x, y)$:

$$\mathbf{n} = \frac{1}{\sqrt{p^2 + q^2 + 1}} [p, q, 1] \quad (4)$$

where the surface orientation is expressed in the gradient space with $p = -\frac{\partial S(x,y)}{\partial x}$ and $q = -\frac{\partial S(x,y)}{\partial y}$ the partial derivative of the surface w.r.t. x and y .

Then the scalar product between the light direction \mathbf{i} and the normal of the surface \mathbf{n} which is expressed with light angles in polar coordinates is :

$$\mathbf{in} = \frac{-p\cos(\tau)\sin(\sigma) - q\sin(\tau)\sin(\sigma) + \cos(\sigma)}{\sqrt{p^2 + q^2 + 1}} \quad (5)$$

with τ and σ the light azimuth and zenith as shown in the Fig. 1.

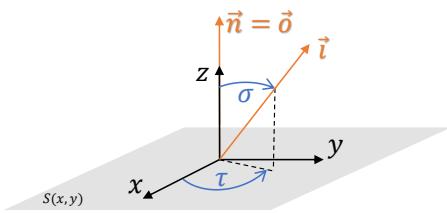


Fig. 1. Diagram of the implementation of the acquisition system on the surface $S(x,y)$ with \mathbf{n} its normal, \mathbf{o} the sensor orientation, \mathbf{i} the directional light orientation and τ and σ the light angles in polar coordinates.

B. Representation of photometric stereo pixel values

From Eq. 5, the easiest method to represent the pixel values $I_k(x,y)$ from k light source positions is the observation map (Fig. 2). For one position (x,y) , the observation map uses the polar coordinate system where each pixel value $I_k(x,y)$ is drawn according to the corresponding light position angles τ_k and σ_k . This type of graph is widely used to represent a BRDF function $f_r(x,y,\mathbf{i},\mathbf{o},\mathbf{n})$ in 2D for one position of \mathbf{i} and all positions of \mathbf{o} . In [16], Ikehata uses this representation as an input data for a per-pixel CNN photometric-stereo method. In this case, the function $f_r(x,y,\mathbf{i},\mathbf{o},\mathbf{n})$ is represented in 2D for one position of \mathbf{o} and all positions of \mathbf{i} . As indicated by [18], this representation is problematic in case of a sparse photometric stereo acquisition (i.e. around ten light positions or less).

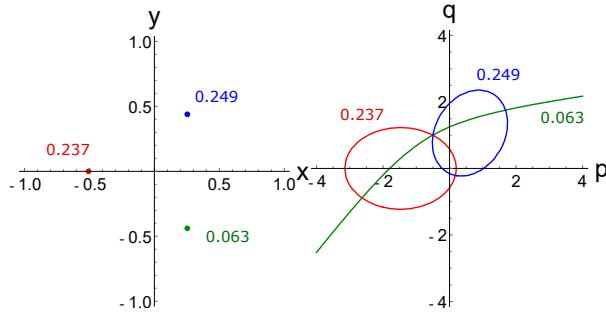


Fig. 2. Representation of photometric stereo pixel values I_k in case of $k = 3$ light positions with $p = -0.5$, $q = 1$ and $\rho = 1$: (a) observation map and (b) reflectance map.

The second possible representation of photometric stereo pixel values is the reflectance map. Once the geometry of the system as well as the physical BRDF function f_r of the object are known, the main idea is to represent the correspondences between a pixel value I_k and all the compatible gradients p and q . The prior method in [1] suggests to use the Lambertian BRDF function which is the easiest way of representing the reflection phenomenon by using a constant term ρ related to the color of the object. The Eq. 1 becomes:

$$L_o(x,y,\mathbf{o},\mathbf{n}) = L_i(x,y,\mathbf{i})\rho(x,y)\max(\mathbf{in}, 0) \quad (6)$$

Fig. 2 shows an example of reflectance map with different light source positions. For each light source k and its corresponding pixel value I_k , an iso-reflectance curve is drawn. In the special case of $L_i(x,y,\mathbf{i})\rho(x,y) = 1$, the distance between the iso-reflectance curve and the light position expressed in derivative space corresponds directly to $\mathbf{in} = \cos(\sigma)$. While an observation map allows a representation of pixel intensities and light source positions, a reflectance map provides in addition the correspondence between pixel intensities and compatible gradients with a selected BRDF model. Thus, a normalization and linearization process is applied to the reflectance maps, resulting in CMR maps as shown in Sec. 3.

3. PROPOSED METHOD

In this section, we present our method that uses reflectance map theory from [1] to generate a CMR map. As our method operates in a per-pixel manner as shown in Fig. 3, the CMR map generation process must be efficient. Indeed, each input pixel vector is processed using its values for different exposure times to extract HDR representation. Those HDR values with their associated light configuration are used to generate a CMR map based on Lambertian assumptions with an iso-reflectance for each light. The CMR map is used as an input for a CNN based on Ikehata et al. [16] to estimate the normal on the object pixel. Thus, this section introduce the CMR map computation and discretization.

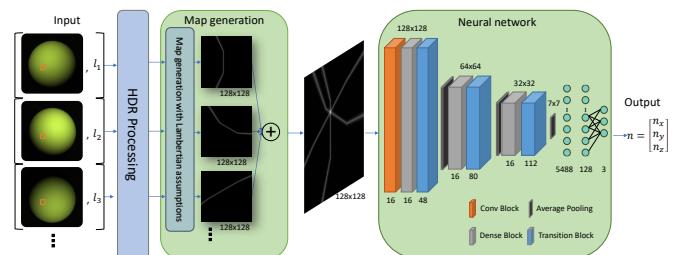


Fig. 3. Global diagram of the end-to-end method from pixels to surface normals.

A. Ratio reflectance map

A well-known problem with traditional reflectance map is that iso-reflectance curves are strongly dependent on the BRDF function f_r . In the Lambertian case, there is a scale problem according to the value of ρ . The most commonly used solution is the photometric ratio [23] to remove dependencies from the albedo. If a light direction \mathbf{i}_k is expressed in the gradient space as:

$$\mathbf{i}_k = \frac{1}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} [p_{sk}, q_{sk}, 1] \quad (7)$$

By neglecting the \max function in Eq. 1, i.e. without self- or cast-shadow, the ratio between two digital values I_1 and I_2 from respectively two different light directions \mathbf{i}_1 and \mathbf{i}_2 becomes:

$$p(\alpha p_{s2} - \beta p_{s1}) + q(\alpha q_{s2} - \beta q_{s1}) + (\alpha - \beta) = 0 \quad (8)$$

with:

$$\begin{aligned} \alpha &= I_1(x, y) f_r(x, y, \mathbf{i}_2, \mathbf{o}, \mathbf{n}) \sqrt{p_{s1}^2 + q_{s1}^2 + 1} \\ \beta &= I_2(x, y) f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n}) \sqrt{p_{s2}^2 + q_{s2}^2 + 1} \end{aligned} \quad (9)$$

This equation describes a line in the gradient space if the function f_r does not depend on the light direction \mathbf{i} , nor \mathbf{n} . Using a single pixel I_2 as a reference would have required to select the same light source for each position (x, y) and ensure that the corresponding pixel is neither in a specular nor a shadow area which cannot be guaranteed.

A more robust approach is therefore to use the mean of pixel values. The line coefficients in Eq. 8 become :

$$Ap + Bq + C = 0 \quad (10)$$

with:

$$\begin{aligned} A &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n}) p_{sk}}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n}) p_{s1}}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \\ B &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n}) q_{sk}}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n}) q_{s1}}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \\ C &= I_1 \sum_k \frac{f_r(x, y, \mathbf{i}_k, \mathbf{o}, \mathbf{n})}{\sqrt{p_{sk}^2 + q_{sk}^2 + 1}} - \frac{f_r(x, y, \mathbf{i}_1, \mathbf{o}, \mathbf{n})}{\sqrt{p_{s1}^2 + q_{s1}^2 + 1}} \sum_k I_k \end{aligned} \quad (11)$$

Equations above are used to draw each curve on a CMR map. The CMR map must then be discretized to feed the CNN.

B. From a chart to an image

A method to analyze these maps consists in converting a CMR map into a small image that can be used as a CNN input. By creating this image, we must keep in mind that each pixel in the acquisition will have to be transformed into an image with an efficient computation. Thus, we need to discretize CMR map curves. Our map is made of segments which are connected in the manner of a piece-wise function. Discontinuities are related to shadow appearance as shown in Fig. 4. The more light sources there are in the system configuration, the more segments and discontinuities there will be on the map.

Another important point to keep in mind is that gradient space is not a linear representation of the surface orientation i.e. according to the angle σ . To control the accuracy on small or high σ value when the reflectance map is drawn, we apply a custom logarithmic scale function to both axes p and q :

$$q_{log} = q_{max} \times \frac{e^{|q|/R} - e^0}{e^{q_{max}/R} - e^0} \times sgn(q) \quad (12)$$

with q_{log} the new value of q according to the limit of the chart q_{max} and a constant R to set the accuracy on low value of σ . Discontinuities and the logarithmic scale make the use of classical drawing algorithms not efficient and will lead to a coarse discretization of the curves of our map. Thus, a scan approach is used to compute the final CMR map pixel intensity.

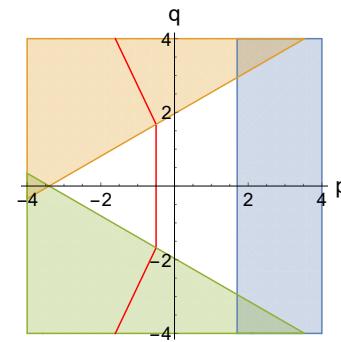


Fig. 4. The discontinuities on the iso-reflectance ratio curve (red line) are due to the presence of self-shadow represented on this map by the orange ($p_l = 0.29, q_l = -0.51$), green ($p_l = 0.29, q_l = 0.51$) and blue ($p_l = -0.58, q_l = 0$) areas. In these areas, coefficient in Eq. 11 are changing as modeled in Eq. 6.

For each pixel, the difference between the theoretical reflectance ratio and the observed reflectance ratio is computed as:

$$\varepsilon_j(p, q) = \frac{Lo_j(p, q)}{\sum_{i=1}^k Lo_i(p, q)} - \frac{kI_j}{\sum_{i=1}^k I_i} \quad (13)$$

The left part of this equation can be precomputed for each light configuration as k theoretical reflectance maps. To control the line thickness and his fading function, a pixel value of the final chart for one light j is equal to:

$$X_j(p, q) = \frac{\left[\left(\frac{1}{a} - \min \left(\frac{1}{a}, |\varepsilon_j(p, q)| \right) \right) \times a \right]^b}{k} \quad (14)$$

where a control the thickness and b the fading. Fig. 5 shows the impact of parameters a and b with our drawing function. Finally, the pixel value $X(p, q)$ of the CNN input image for k lights is:

$$X(p, q) = \sum_{i=1}^k X_i(p, q) \quad (15)$$

Fig. 5 shows an example of our radiance map CNN input in case of linear linear or our custom logarithmic scale function according to p and q .

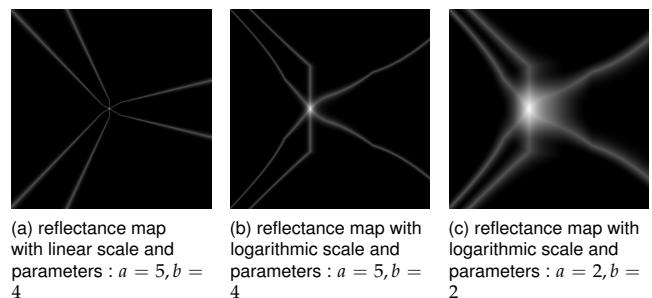


Fig. 5. Three representations of reflectance map for gradient : $p = -0.5, q = 0$ with 3 lights. Reflectance map resolution is 128 and $q_{max} = p_{max} = 20$.

277 4. IMPLEMENTATION

278 To train our network, we create a dataset of 1 million photo-
 279 metric stereo pixels for a fixed lighting configuration. Pixels
 280 are generated with Eq. 1 independently of each other using
 281 randomly parameterised BRDFs and gradients. In particular
 282 conditions as specular and shadow, we dedicate a part of the
 283 training dataset to improve our algorithm on those areas.

284 A. BRDF

285 BRDF representation and manipulation is a complex task to rep-
 286 resent appearance of materials in real world applications. Using
 287 pixel-wise simulation, there are two ways to represent it. We
 288 can use a measured dataset of BRDF or compute an analytical
 289 function using different parameters. Measured BRDFs allow to
 290 represent real materials but in limited quantity. On the other
 291 hand, a parametric BRDF model can be used to represent an infi-
 292 nite diversity of materials within the limits of the chosen model.
 293 Then, we implemented 3 different approaches to compute the
 294 BRDF f_r and compare it in section 5:

- 295 • Mitsuba 2 integrated "roughplastic" BRDF model with a
 296 Beckmann distribution [24] with randomly generated pa-
 297 rameters $K_d \in [0, 0.8]$; $\alpha \in [0.0001, 0.5]$; $N_i \in [1.0, 1.8]$.
- 298 • Simonot with a Beckmann distribution [25] with randomly
 299 generated parameters $K_d \in [0, 0.8]$; $\alpha \in [0.0001, 0.5]$; $N_i \in$
 300 $[1.0, 1.8]$ implemented by [22]. Unlike Mitsuba 2 integrated
 301 "roughplastic" BRDF model, this model describes surfaces
 302 as flat interfaces on a Lambertian background.
- 303 • MERL BRDF database [26] that contains measured real re-
 304 flectance functions of 100 different materials.

305 B. Gradient selection

306 To generate gradients we use a random uniform selection of
 307 p and q with a limit set to 6. Knowing that specular effect is
 308 a complex phenomenon for the neural network, we chose to
 309 dedicate 30% of the database to represent specular pixels by
 310 using importance sampling [27] i.e. gradients are chosen in the
 311 specular lobe of the BRDF.

312 C. Shadow simulation

313 Furthermore, as our pixel values are not determined from an
 314 homogeneous rendered surface, 40% of the dataset is dedicated
 315 to represent cast shadows. For this purpose, we use the occlusion
 316 layer principle presented in [17] to replace pixels with a noise
 317 value to simulate cast shadow. This random additive noise has
 318 a standard deviation of 1e-4 to simulate camera noise as in [19].

319 D. Proposed Network

320 As our method works pixel-wise on the full image for each
 321 light but takes feature images as input, we use a CNN based
 322 on DenseNet [28] as the one used in Ikehata et al. [16]. In our
 323 case, we use images of bigger size than observation maps. Thus,
 324 we add one dense block in the architecture to limit the number
 325 of parameters of the network, increase its abstraction level and
 326 intuitively being more suited for our input images. This CNN is
 327 able to recognize patterns in an image that correspond to high-
 328 lights or shadows, further explanations are given in Supplement
 329 1. A description of this architecture is shown in Fig. 3. The loss
 330 function \mathcal{L} used for training is the Mean Squared Error (MSE)
 331 and its optimizer is Adam with a learning rate set to 1e-4. The
 332 CNN for a map size of 128 has 773K parameters.

333 E. Experimental setup

334 Experiments were performed on a PC with GeForce GTX 1080
 335 and 32 Go of RAM. Training dataset is generated with a C++
 336 framework and stored for training. For training and prediction
 337 we use PyTorch [29] library with Pytorch Lightning overlay. The
 338 CNN was trained with 40 epochs and batches of 256 CMR maps.

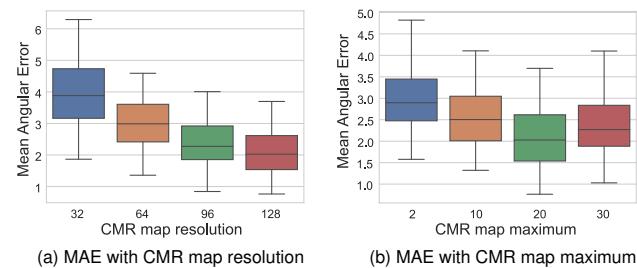
339 5. EXPERIMENTS

340 In this section, we evaluate our method to find optimal par-
 341 ameters, i.e. CMR map configuration and BRDF model, allowing
 342 the best accuracy and execution time. During this process,
 343 we compare our method trained with CMR maps and observa-
 344 tion maps with state-of-the-art photometric stereo techniques
 345 as Least Square [1], PS-FCN [14] and CNN-PS [16]. As metric
 346 for the evaluation, we compute the Mean Angular Error (MAE)
 347 in degree between the ground truth and estimated normal map
 348 and then generate its associated error map. First, we use path
 349 tracing rendering with Mitsuba 2 [24] to form a database with
 350 ground truths. It will serve to find the optimal reflectance map
 351 configuration as resolution and limit of the map and evaluate the
 352 effectiveness of shadow and specular integration in the training
 353 dataset through an ablation study. Second, we use DiLiGenT
 354 dataset to find the best BRDF model for real surface normal map
 355 estimation. Finally, we introduce a new dataset with two acqui-
 356 sition systems using surfaces with the same geometry but with
 357 different appearances. We also evaluate the representation be-
 358 tween observation map and CMR map using these 3 databases.

360 A. Synthetic experiments

361 Synthetic experiments will allow us to establish optimal resolu-
 362 tion and limits for p and q axes of the reflectance map. We also
 363 compare the efficiency of CMR maps and observation maps by
 364 training them on the same network architecture and evaluate on
 365 this data. The synthetic data is rendered using Mitsuba 2 Ren-
 366 derer [24] with its integrated "roughplastic" Beckmann BRDF
 367 on 9 objects and 9 BRDF configurations ($\alpha \in \{0.01, 0.05, 0.1, 0.2\}$
 368 and $N_i \in \{1.0, 1.2, 1.5\}$). Those configurations were chosen to
 369 have a large variation of parameters from matte to glossy sur-
 370 faces representing most real surfaces.

371 A.1. CMR map optimization



372 **Fig. 6.** Mean angular error (MAE) for 11 simulated object
 373 normal maps estimated with our method trained with Si-
 374 monot BRDF model : (a) map resolution analysis $W, H \in$
 375 $\{32, 64, 96, 128\}$ for a map maximum set to 20 and (b) maxi-
 376 mum analysis $p_{max}, q_{max} \in \{2, 10, 20, 30\}$ for a map resolution
 377 set to 128 on synthetic data.

Table 1. Evaluation of methods on synthetic dataset with a model trained with Simonot BRDF [22] : the result for an object is an average of the results obtained under different 9 BRDF configurations with $\alpha \in \{0.01, 0.05, 0.1, 0.2\}$ and $N_i \in \{1.2, 1.5\}$ and Lambertian configuration ($\alpha = 0.2$ and $N_i = 1.0$). We compare our method trained with CMR maps (CMR) and observation maps (OM).

Methods	Alligator	Bunny	Cesar	Mask coin	Mask	Pine cone	Rock	Shield	Sphere	Avg.
LS [1]	6.53	4.34	5.26	4.87	6.1	7.6	5.77	7.01	3.29	5.64
PS-FCN [14]	3.49	3.03	3.32	2.96	3.81	3.28	3.94	3.64	3.95	3.49
CNN-PS [16]	5.57	6.24	5.52	4.87	5.74	6.23	5.52	5.18	6.1	5.66
Ours (OM)	2.63	2.14	2.5	1.95	2.95	3.22	2.64	2.12	1.83	2.44
Ours (CMR)	2.41	1.69	2.13	1.4	2.62	3.48	2.42	1.49	1.26	2.1

The graph on Fig. 6 presents the MAE between estimated normal maps with our method trained with Simonot BRDF model and its associated ground truth, according to resolution Fig. 6 (a) and p and q axis limits of the CMR map Fig. 6 (b). This result allows us to conclude that the larger the map, the smaller the error. The gain brought by an increase in the size of the map is global and applies equally to specular and diffuse areas as shown in Fig. 7. This phenomenon can be explained by the resolution of the maps which gives more information about the exact gradient. However, the size of the map greatly influences the processing time of the method increasing time to generate the CMR map and to process by the neural network. Thus, selecting a map size of 128 allows a trade-off between estimation accuracy and computation time. The conclusion concerning the limit of p and q axes is different since better results are obtained with a limit set at 20 for a logarithmic distribution of gradients as shown in Fig 8. This limit allows a better global precision on the selected objects which presents various geometries. We present the results from the configuration with the size set to 128 and the limit to 20 for the rest of the article.

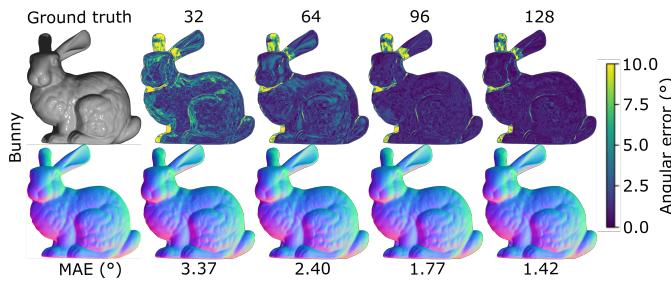


Fig. 7. Comparison of results with map resolutions $W, H \in \{32, 64, 96, 128\}$ on synthetic object Stanford bunny with mitsuba "roughplastic" BRDF $\alpha = 0.05$ and $N_i = 1.2$ and with map limit $p_{max} = q_{max} = 20$.

A.2. Ablation study

For photometric stereo tasks, the challenge remains to estimate normals on specular and cast-shadow areas. We present an ablation study to prove that our emphasis parameters on specular and shadow components improve the results and the understanding of the neural network on various simulated objects. On Tab. 2, we show the evolution of the MAE for the same network trained with and without those parameters. Importance sampling permits to guide the attention of the network on the non-linear areas, i.e. specular areas, which gives a better understanding and representation of the BRDF. Furthermore,

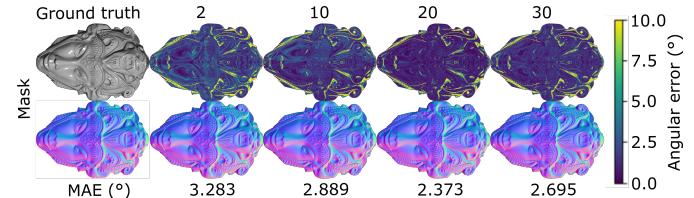


Fig. 8. Comparison of results with map limits $p_{max}, q_{max} \in \{2, 10, 20, 30\}$ on synthetic object mask with mitsuba "roughplastic" BRDF $\alpha = 0.05$ and $N_i = 1.2$ and map resolution equal to 128.

Effects	Average	Std. Deviation
Baseline	4.06	2.04
+Specular	3.88	2.1
+Shadow	2.1	0.75

Table 2. Ablation study on synthetic dataset mean angular error with Average and Standard Deviation using CMR maps with size 128 and maximum set to 20. Baseline is our network trained with a dataset generated without importance sampling (Specular) nor cast-shadow (Shadow).

cast-shadow handling drastically improve results because it is the only way to simulate this effect on an independent pixel-wise data generation.

A.3. Results

We can establish a first comparison with the methods of the literature [1, 14, 16] on the synthetic dataset. For all the synthetic database, Tab. 1 shows that using CMR maps increases the accuracy of normal estimation compared to observation maps. If we take a close look on Fig. 9, we can notice a clear improvement on specular and diffuse areas but a less accurate estimation on cast shadow areas. One interpretation of these results could be that CMR map curves suffer from inter-correlation with shadow pixels while diffuse and specular reflection reveal a good BRDF representation which therefore brings better results. Furthermore, for pixel-wise methods, cast-shadow is a complex problem since these artifacts are generated from global geometry phenomenons.

B. Experiments based on benchmark DiliGenT

DiliGenT is a testing dataset [30] which is composed of 10 different objects illuminated with 96 lights and associated to their

Table 3. Evaluation of methods on DiLiGenT EXR dataset using light configuration as seen in 10 (c) with best EXR results in bold. We compare our method trained with CMR maps (CMR), observation maps (OM) and 3 BRDF configurations (MERL, Mitsuba and Simonot) as explained in Sec. 4-A.

Methods	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	Avg.
LS [1]	4.12	9.47	15.36	8.92	25.95	19.34	31.25	9.38	14.98	18.06	15.68
PS-FCN [14]	5.41	8.43	10.91	7.9	13.98	12.63	20.16	8.22	10.64	14.02	11.23
CNN-PS [16]	13.56	17.69	21.97	16.8	18.53	19.78	25.84	15.86	19.08	20.42	18.95
Ours (Simonot-CMR)	2.52	11.42	13.43	7.71	14.82	14.66	23	7.57	11.1	17.72	12.4
Ours (Mitsuba-CMR)	1.79	11.29	13.16	7.28	14.87	14.61	24.69	7.45	10.64	17.14	12.29
Ours (MERL-OM)	4.55	9.58	14.26	8.1	13.52	13.09	22.62	8.23	10.26	16.98	12.12
Ours (MERL-CMR)	3.42	9.63	13.27	7.61	12.77	13.09	22.67	7.48	9.23	16.55	11.57

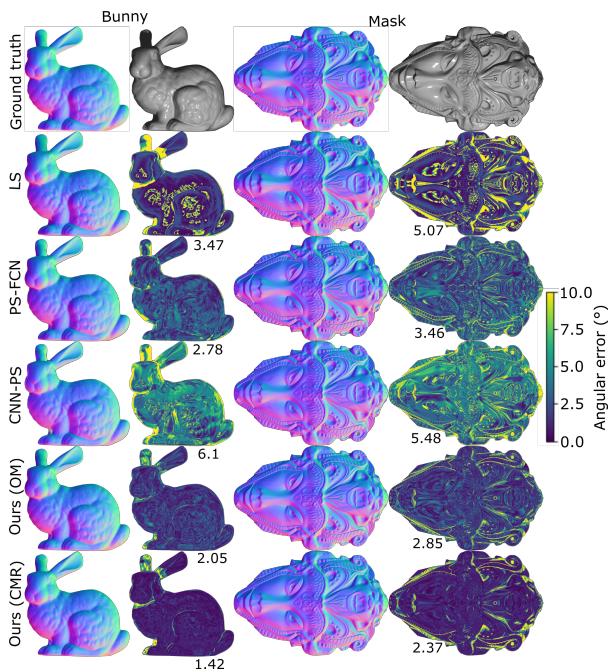


Fig. 9. Comparison on 2 synthetic objects (Stanford bunny and mask) with mitsuba "roughplastic" BRDF $\alpha = 0.05$ and $N_i = 1.2$, map limit is 20 and map resolution is 128. Each method is associated with its mean angular error for each object.

ground truth normal maps. This dataset is commonly used to benchmark photometric stereo methods. DiLiGenT is provided with PNG images of size 612×512 with 16-bit quantization and we would like to thank authors for providing us EXR high resolution data (2448×2048 with floating value). We have selected 6 lights out of 96 available (corresponding to light number 1, 26, 31, 56, 74 and 79) on DiLiGenT EXR dataset to replicate a photometric stereo system with a few homogeneously placed lights as shown in Fig. 10 (c). This number of lights is the limit for photometric stereo techniques to obtain acceptable results due to specular overlapping and shadows. Furthermore, this configuration allows to compare performances between state-of-the-art methods and our method under nearly identical conditions as our own system with ground truth normal maps.

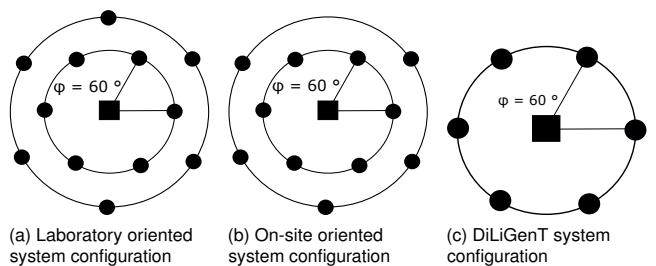


Fig. 10. Diagram of the light configuration for all systems. The camera is represented by the square while the light sources are represented by the circles. In (a), the radius of the small and large rings are 470 mm and 1675 mm. Their heights are 1675 mm and 1575 mm. In (b), the radius of the small and large rings are 140 mm and 235 mm. Their heights are 395 mm and 345 mm. Finally, in (c), we chose 6 lights in DiLiGenT to form a ring with about the same properties as the first rings on our own systems.

First, we used DiLiGenT dataset to choose the best BRDF model among the models, i.e Mitsuba 2 integrated "roughplastic" beckmann BRDF model [24], Simonot BRDF model [25] implemented by [22] and MERL BRDF database [26]. The results in Tab. 3 allows us to conclude that our technique is robust concerning its training BRDF model but gives slightly better results with MERL BRDF database. Therefore, we chose a reflectance map configuration with MERL BRDF database training and a resolution of 128 pixels for the next experiments.

Second, a comparison with other methods is done in Tab. 3. As shown in Fig. 11, our method is outperforming other methods when cast shadows are not present. The estimation of surface normals in the presence of specular pixels is particularly accurate compared to other methods as shown in Fig. 11. Moreover, with sparse set of lights, CMR maps allow better results than observation maps which confirms a better representation of the BRDF. Our method performs about the same as global method PS-FCN [14] which is able to take into account spatial information.

C. Our real world experiments

To confirm the conclusions of the previous parts on real data, we use two photometric stereo acquisition systems. The first

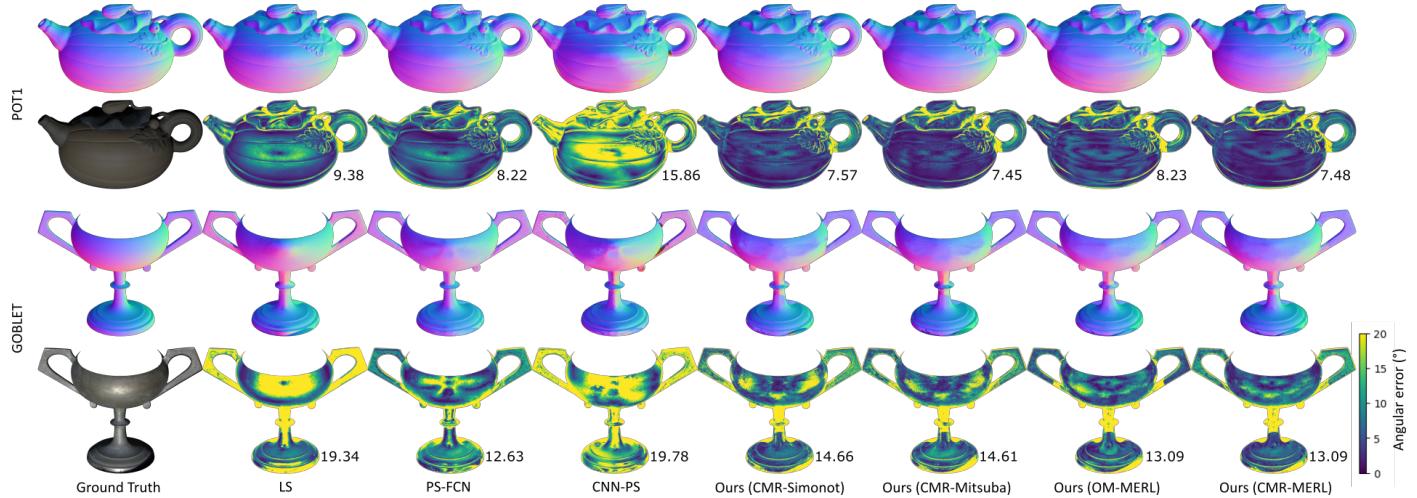


Fig. 11. Comparison on 2 DiLiGenT objects (POT1 and GOBLET) on EXR data for state-of-the-art methods (LS [1], CNN-PS [16], PS-FCN [14]) and our method trained with CMR maps (CMR), observation maps (OM) and 3 BRDF configurations (MERL, Mitsuba and Simonot) as explained in Sec. 4-A.

459 system is composed of 12 lights distributed on two rings with
 460 a 5.3 m^3 volume with a laboratory oriented configuration, no
 461 ambient lighting and lens with a 300 mm focal length. The light
 462 sources and camera are far from the objects which allows to be
 463 close from orthographic projection with directional and uniform
 464 lighting as shown in Fig. 10 (a). The second system is composed
 465 of 11 lights distributed on two rings with a 0.088 m^3 volume and
 466 camera lens with a 75 mm focal length. This system is designed
 467 to be moved and used in challenging environments as shown in
 468 Fig. 10 (b). All the acquisitions are taken with different exposure
 469 times to allow HDR reconstruction that is realized according to
 470 [31] without neighbor processing. This method assures that our
 471 HDR processing is linear and limits the noise on HDR images.

472 To evaluate photometric stereo methods and systems, ground
 473 truths are complicated to obtain (i.e. same resolution, surface
 474 registration errors, etc.). Thus, we have chosen 4 real objects to
 475 make painted plaster copies to qualitatively compare the same
 476 geometry with different materials. They have been mold with
 477 Silastic RTV-3481 silicone and duplicated with ALPHA 3500
 478 plaster. They have been duplicated four times and three of them
 479 have been colored with various types of paints (matte, satin
 480 like and glossy) to evaluate different shapes of specular lobes.
 481 Within these objects, the shape and appearance can be controlled,
 482 and the results are thus compared. The acquisitions made with
 483 these systems will be available in open access.

484 As shown in Fig. 12 and 13, our method is resilient to a wide
 485 range of BRDF especially when the estimation is done in the
 486 presence of specular pixels. As we have seen in the previous
 487 sections, CNN-PS and PS-FCN struggle to estimate homogeneous
 488 normals in presence of glossy artefacts and especially on thin
 489 specular lobes. Furthermore, observation maps qualitatively
 490 create more artifacts on specular areas than CMR maps. Concerning
 491 the BRDF used for training, the qualitative analysis on real data
 492 agrees with the quantitative analysis made on DiLiGenT. Certain
 493 areas are misinterpreted when the training is based on a para-
 494 metric BRDF model (Mitsuba and Simonot BRDF models). These
 495 areas are well reconstructed with a training based on the MERL
 496 database with an example on Fig. 14. These results suggest that
 497 MERL BRDF dataset represents a wider range of BRDF than a
 498 parametric BRDF model implying better network generalization.

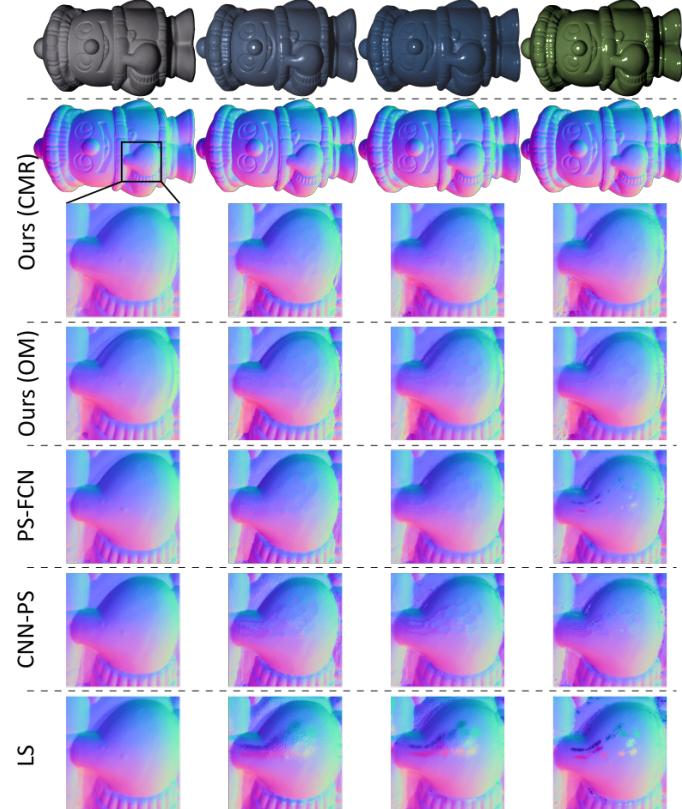


Fig. 12. Comparison between LS [1], CNN-PS [16], PS-FCN [14] and the proposed method normal maps trained with CMR maps (CMR) and observation maps (OM) on the on-site system using 4 objects with the same geometry but with different appearances.

6. CONCLUSION AND LIMITATIONS

500 In this paper, we have presented a new approach for pixel-wise
 501 calibrated photometric stereo with few lights keeping high-
 502 quality 3D reconstruction for surfaces. We introduced a new

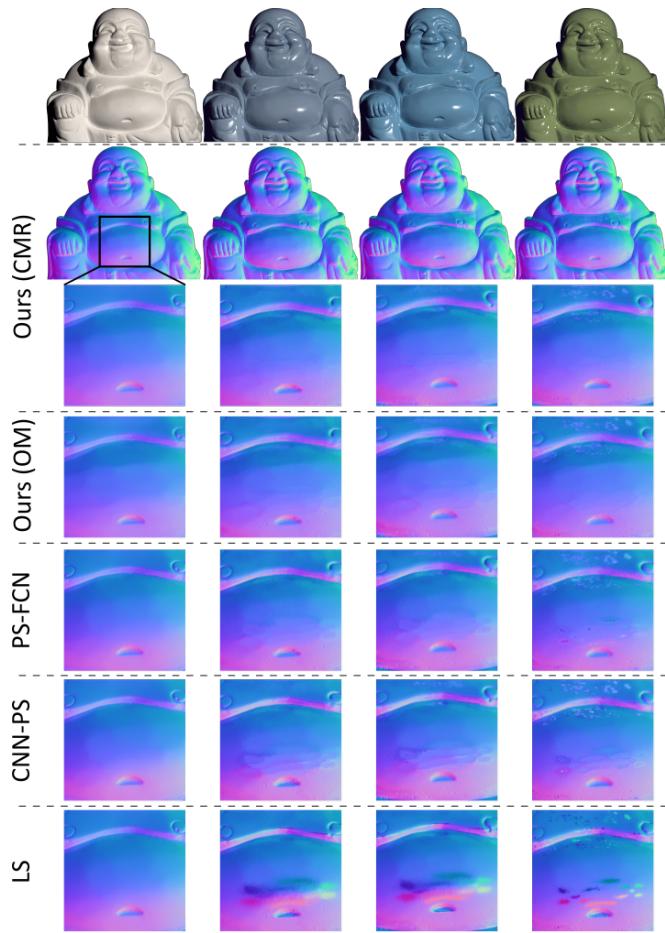


Fig. 13. Comparison between LS [1], CNN-PS [16], PS-FCN [14] and the proposed method normal maps trained with CMR maps (CMR) and observation maps (OM) on the laboratory-oriented system using 4 objects with the same geometry but with different appearances.

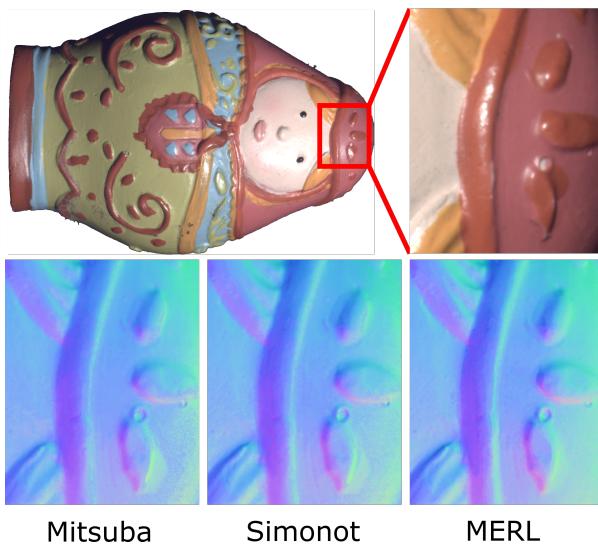


Fig. 14. Comparison of results with our method with training data based on Simonot BRDF model, Mitsuba 2 integrated "roughplastic" BRDF model, MERL BRDF Dataset.

feature to represent photometric stereo information for neural network analysis by using CMR maps instead of observation maps. Our method's accuracy is similar to or outperforms state-of-the-art methods when using a system with few light positions and further improvement can be obtained by integrating the training dataset contributions of PX-NET [19] in our method. This conclusion is true on three different types of data : synthetic, DiLiGenT and real world dataset which was acquired by our own systems. Photometric stereo is used to extract details on surfaces and global information add bias in the measurement chain by increasing the dimensionality of the problem which allows a better estimation of the normals in cast shadow regions but creates smoothing patch on reconstructed objects. Applying pixel-wise photometric stereo to these systems requires a suitable data representation. With our data representation, self shadow, diffuse and specular pixels lead to a particularly accurate estimation of the normal for various BRDF.

Acknowledgments. Clément Joubert and Nicolas Prouteau are co first authors.

Disclosures. The authors declare no conflicts of interest.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. R. J. Woodham, "Photometric Method For Determining Surface Orientation From Multiple Images," *Opt. Eng.* **19** (1980).
2. E. N. Coleman and R. Jain, "Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry," *Comput. Graph. Image Process.* **18**, 309–328 (1982).
3. S. Barsky and M. Petrou, "The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows," *IEEE Trans. on Pattern Anal. Mach. Intell.* **25**, 1239–1252 (2003).
4. Y. Mukaigawa, Y. Ishii, and T. Shakunaga, "Analysis of photometric factors based on photometric linearization," *J. Opt. Soc. Am. A* **24**, 3326–3334 (2007).
5. L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," *Comput. Vis. – ACCV 2010* pp. 703–717 (2011).
6. B. Bringier, A. Bony, and M. Khoudeir, "Specularity and shadow detection for the multisource photometric reconstruction of a textured surface," *JOSA A* **29**, 11–21 (2012).
7. S. Nayar, K. Ikeuchi, and T. Kanade, "Determining shape and reflectance of hybrid surfaces by photometric sampling," *IEEE Trans. on Robotics Autom.* **6**, 418–431 (1990).
8. A. S. Georgiades, "Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo," *Proc. Ninth IEEE Int. Conf. on Comput. Vis.* pp. 816–823 vol.2 (2003).
9. D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and Spatially-Varying BRDFs from Photometric Stereo," *IEEE Trans. on Pattern Anal. Mach. Intell.* **32**, 1060–1071 (2010).
10. T. Higo, Y. Matsushita, and K. Ikeuchi, "Consensus photometric stereo," *IEEE Comput. Soc. Conf. on Comput. Vis. Pattern Recognit.* pp. 1157–1164 (2010).
11. A. Hertzmann and S. Seitz, "Shape and materials by example: a photometric stereo approach," *IEEE Comput. Soc. Conf. on Comput. Vis. Pattern Recognition, 2003. Proceedings.* **1** (2003).
12. J. Ackermann, M. Goesele *et al.*, "A survey of photometric stereo techniques," *Found. Trends Comput. Graph. Vis.* **9**, 149–254 (2015).
13. H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita, "Deep photometric stereo network," *IEEE Int. Conf. on Comput. Vis. Work. (ICCVW)* pp. 501–509 (2017).
14. G. Chen, K. Han, and K.-Y. K. Wong, "PS-FCN: a flexible learning framework for photometric stereo," *Comput. Vis. – ECCV 2018* pp. 3–19 (2018).

- 565 15. Y. Cao, B. Ding, Z. He, J. Yang, J. Chen, Y. Cao, and X. Li, "Learning
566 inter- and intraframe representations for non-lambertian photometric
567 stereo," *Opt. Lasers Eng.* **150**, 106838 (2022).
- 568 16. S. Ikehata, "CNN-PS: cnn-based photometric stereo for general non-
569 convex surfaces," 15th Eur. Conf. on Comput. Vis. ECCV, Munich, Ger.
570 **11219**, 3–19 (2018).
- 571 17. J. Li, A. Robles-Kelly, S. You, and Y. Matsushita, "Learning to minify
572 photometric stereo," IEEE/CVF Conf. on Comput. Vis. Pattern Recognit.
573 (CVPR) pp. 7560–7568 (2019).
- 574 18. Q. Zheng, Y. Jia, B. Shi, X. Jiang, L. Duan, and A. Kot, "SPLINE-Net:
575 sparse photometric stereo through lighting interpolation and normal
576 estimation networks," IEEE/CVF Int. Conf. on Comput. Vis. (ICCV) pp.
577 8548–8557 (2019).
- 578 19. F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla, "Px-net: Simple
579 and efficient pixel-wise training of photometric stereo networks," in
580 *Proceedings of the IEEE/CVF International Conference on Computer
581 Vision*, (2021), pp. 12757–12766.
- 582 20. R. L. Cook and K. E. Torrance, "A reflectance model for computer
583 graphics," *ACM Trans. Graph.* **1**, 7–24 (1982).
- 584 21. C. Kelemen and L. Szirmay-Kalos, "A microfacet based coupled
585 specular-matte brdf model with importance sampling," *Eurographics
586 Assoc.* (2001).
- 587 22. D. Meneveaux, B. Bringier, E. Tauzia, M. Ribardi  re, and L. Simonot,
588 "Rendering rough opaque materials with interfaced lambertian micro-
589 facets," *IEEE Trans. on Vis. Comput. Graph.* **24**, 1368–1380 (2018).
- 590 23. Y. Qu  au, R. Mecca, and J.-D. Durou, "Unbiased photometric stereo
591 for colored surfaces: A variational approach," *IEEE Conf. on Comput.
592 Vis. Pattern Recognit. (CVPR)* pp. 4359–4368 (2016).
- 593 24. M. Nimier-David, D. Vicini, T. Zeltner, and W. Jakob, "Mitsuba 2: A retar-
594 getable forward and inverse renderer," *Trans. on Graph. (Proceedings
595 SIGGRAPH Asia)* **38** (2019).
- 596 25. L. Simonot, "Photometric model of diffuse surfaces described as a
597 distribution of interfaced Lambertian facets," *Appl. Opt.* **48**, 5793 (2009).
- 598 26. W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven
599 reflectance model," *ACM Trans. on Graph.* **22**, 759–769 (2003).
- 600 27. E. Heitz and E. D'Eon, "Importance Sampling Microfacet-Based BSDFs
601 using the Distribution of Visible Normals," *Comput. Graph. Forum* **33**,
602 103–112 (2014).
- 603 28. G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely
604 connected convolutional networks," *IEEE Conf. on Comput. Vis. Pattern
605 Recognit. (CVPR)* pp. 2261–2269 (2017).
- 606 29. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan,
607 T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative
608 style, high-performance deep learning library," *Adv. neural information
609 processing systems* **32**, 8026–8037 (2019).
- 610 30. B. Shi, Z. Mo, Z. Wu, D. Duan, S.-K. Yeung, and P. Tan, "A benchmark
611 dataset and evaluation for non-lambertian and uncalibrated photometric
612 stereo," *IEEE Trans. on Pattern Anal. Mach. Intell.* **41**, 271–284 (2019).
- 613 31. M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video
614 production system," *ACM Trans. Graph.* **30** (2011).