A Neural Height-Map Approach for the Binocular Photometric Stereo Problem Supplementary matterial

Fotios Logothetis
Cambridge Research Laboratory,Toshiba Europe Ltd.
Cambridge, UK
fotios.logothetis@toshiba.eu

Ignas Budvytis
University of Cambridge
Cambridge, UK
ib255@cam.ac.uk

Roberto Cipolla
University of Cambridge
Cambridge, UK
rc10001@cam.ac.uk

Abstract

This document presents supplementary material for our main submission. In Section 1 we provide additional details for our newly introduced Binocular Photometric Stereo LUCES-ST dataset. Section 2 provides some additional details about the learned BRDF renderer.

1. LUCES-Stereo

First, we note that we only reused 7 of the original LUCES [3] objects and their CT scanned GT meshes and all of the stereo capture data are new. We chose a variety of materials: *Bell* is bronze, *Bunny* is shiny plastic, *Cup* is aluminium, *Hippo* is plastic, *Owl* is ceramic, *Queen* is plaster and *Squirrel* is porcelain. In fact, the stereo capture device only contains 15 lights as the data acquisition speed is very important in an industrial inspection setting. This sparse lighting setting makes the photometric stereo problem highly challenging adding to the value of our dataset. Note that as the 2 stereo images per light are captured simultaneously, the effective light directions for each pixel differ for the 2 views (due to parallax) in contrast to turntable setups like DiLiGenT-MV. This makes the application of uncalibrated PS methods (such as [1]) that rely on the lighting vectors being the same at each less practicable. Indeed, calibrated PS SOTA normal estimation on these objects (using [3], see Table 1) achieves a non negligible error (i.e. mean 10.5^o). We hope that our PS data will be useful for future research on sparse PS aiming to minimise this error.

Similarly to competing datasets (original LUCES [3], DiLiGenT [2]) the camera intrinsics and baseline were computed using a the standard checkerboard calibration target procedure and the lighting calibration parameters (using the near lighting model of [5] that includes position, brightness, principal direction and attenuation factor as explained in Section 3 of the main paper) using a diffuse calibration target[1]. We note that the target was captured in 5 distances of 22, 24, 25, 29, 30 cm, thus providing a total of 5x2x15=150 calibration images. The lighting model parameters were fitted with differentiable rendering obtaining a final re-rendering error of ≈ 0.005, thus the expected accuracy of the light calibration should be around 0.5%. We note that despite the fact that the images of the stereo pair are captured simultaneously, the effective brightness of the LEDs are not identical to both views due to different camera sensitivity (which is also channel dependent). This makes uncalibrated PS especially challenging as both light positions and orientations (from the cameras point of view) and brightness different between respective stereo pair cameras.

The ground truth meshes are also aligned with Meshlab[2] and thus ground truth normals and depth are rendered (with Blender[3]) for each view. In addition, we note that segmentation masks were manually edited to exclude some points that the 'GT' may be unreliable such as the wheels at the bottom of Bunny. In addition, for the making a fair evaluation of stereo methods, the segmentation masks were further cropped to only include points on the field of views of both cameras (but not necessarily co-visible due to self occlusions).

Synthetic LUCES-ST. To further investigate synthetic to real gap, we also rendered all 7 objects with Blender using the same 15 stereo lights setup at the real one and with reasonable material guesses. This is shown in Figure 1 along with the results of the normals + rendering variation. Note however that the poses are not identical to the real data.

2. BRDF Renderer

This section provides additional details about the learnt BRDF renderer. First of all, we note that considering RGB images has very little value compared to grayscale ones (de-

[1] https://www.edmundoptics.co.uk/f/white-balance-reflectance-targets/13169/

[2] https://www.meshlab.net/

[3] https:https://www.blender.org/

Object	Bell	Bunny	Cup	Hippo	Owl	Queen	Squirrel	Average
View 1 angular error (degrees)	15.12	6.08	13.80	5.88	10.75	9.65	13.76	10.21
View 2 angular error (degrees)	15.16	5.71	15.98	7.07	11.94	9.21	13.60	10.85

Table 1. Evaluation of the accuracy of the normals predicted by [3] on images of LUCES-stereo dataset. Not surprisingly, this is not negligible as there are only 15 lights available and the object's geometry is challenging (causing shadows, self reflections etc). The normal error maps are shown in Figure 5 of the main paper.

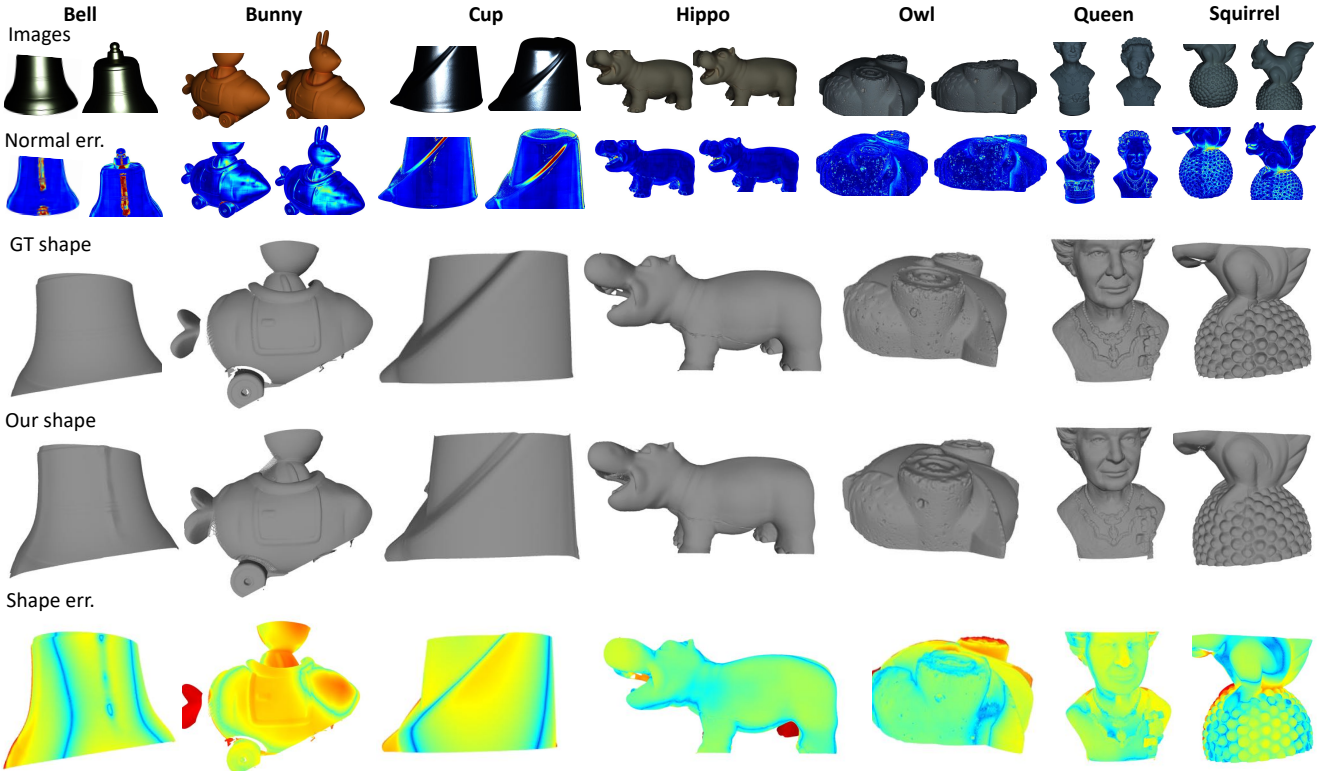

Figure 1. This figure shows the qualitative results of our method (normals+intenisty loss variation) on LUCES-ST-synthetic dataset. This is a synthetic counterpart to Figure 5 of the main text. The first three rows show the cropped images and corresponding error images of normals estimated from PX-Net [3] and the ground truth shape. The final two rows show the shape predicted by our method and corresponding error map (from ground truth to reconstruction). Similarly to Figures 4 and 5 of the main text, any errors larger than 1.5mm are clamped to a dark red color. In contrast to real data results, out method performs relatively well on all objects (performance drop on metallic objects such as Bell and Cup is less significant than in the real data) with the challenging geometry regions (around occlusion boundaries) accumulating most of the error.

spite the added computation and memory overhead), especially since RGB cameras usually have Bayer pattern filters and the RGB color are recovered using demosaicing, we which we optimised to mostly preserve brigthness and not 'true' color. Therefore, we follow standard RGB to gray conversion[4] on input images, and optimise our RGB renderer using scalar albedos and intenisty rendering.

BRDF parameterisation. Our aim is to learn a single BRDF model (assuming uniform material for the whole object), following the principles describe in the MERL real material database [4]. First, we note that in the case of normal, viewing and lighting vectors having relative angles $> 90^o$, the reflected light is always 0, therefore any real rendering should include a $\text{sign}(\mathbf{n}\cdot\mathbf{l})\text{sign}(\mathbf{n}\cdot\mathbf{v})\text{sign}(\mathbf{v}\cdot\mathbf{l})$ (sign is a binary flag 0 for negatives, 1 for positives and was ommited from equation in Line 396 of main text for clarity). In addition, we note that the BRDF is only a function of the relative angles of these vectors, so they can all be rotated such as $\mathbf{n} = [0,0,1]$ (by rotating around the $\mathbf{n}\times[0,0,1]$ axis with angle $\arccos(\mathbf{n}\cdot[0,0,1])$). Thus $\text{BRDF}(\mathbf{n},\mathbf{l},\mathbf{v}) := \text{BRDF}(\mathbf{l_n},\mathbf{v_n})$ with $\mathbf{l_n},\mathbf{v_n}$ the new rotated vectors. In addition, as most of the specular lobe is

[4] https://docs.opencv.org/3.1.0/de/d25/imgproc_color_conversions.html

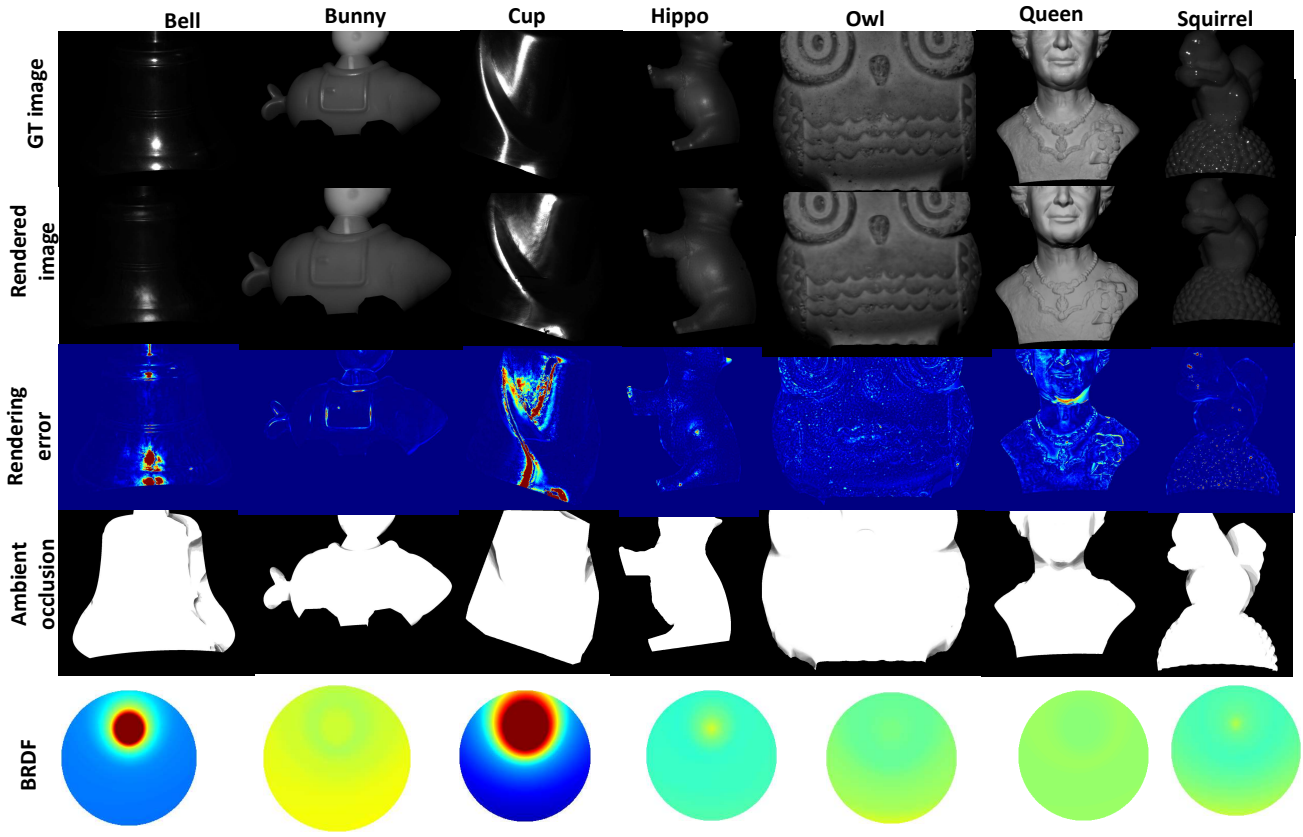

Figure 2. Rendering visualisation for all LUCES-ST objects. From top to bottom, rows include ground truth first image of first camera, respective rendered image, rendering error (with dark red corresponding to 0.25), ambient occlusion (number of shadows/number of images) and recovered BRDF. We note that the square of the ambient occlusion is used to weigh the rendering loss, hence rendering error in these regions does not really affect the training procedure. We also note that the BRDF is visualised excluding the $(\mathbf{n} \cdot \mathbf{l})$ component (hence the values around the edges of the sphere are meaningless). Moreover, this BRDF rendering corresponds to $\mathbf{v} = [0, 0, 1]$ (i.e. 'orthographic viewing') and $\mathbf{l} = [0, \sqrt{2}/2, \sqrt{2}/2]$ and the colormap is chosen with green corresponding the 1 and dark red to > 2.

around the half vector $\mathbf{h} = \frac{\mathbf{l_n}+\mathbf{v_n}}{|\mathbf{l_n}+\mathbf{v_n}|}$, [4] recommends parameterising $\mathbf{h}$, as well as the difference vector $\mathbf{d}$ between $\mathbf{h}$ and $\mathbf{l_n}$. In fact, $\mathbf{d}$ is computed by rotating both $\mathbf{h}$ and $\mathbf{l_n}$ such as $\mathbf{h}$ is aligned with $[0, 0, 1]$. Thus the 4 angles defining the BRDF can now be computed as (using superscripts to denote x, y, z components of 3D vectors):

- $\theta_h = \arccos(h^z) \in [0, \pi/2]$.
- $\phi_h = \text{atan2}(h^y, h^x) \in [-\pi, \pi]$.
- $\theta_d = \arccos(d^z) \in [0, \pi/2]$
- $\phi_d = \text{atan2}(d^y, d^x) \in [-\pi, \pi]$

In addition, any real BRDF must follow the Helmholtz reciprocity constraint which enforces symmetry between $\mathbf{l_n}$ and $\mathbf{v_n}$ vectors; in the 4 angle parameterisation that translates to periodicty wrt to ϕ_d with period π (instead of 2π), therefore from a learning perspective $\phi_d \in [0, \pi]$ is sufficient. Moreover, since we aim to learn these kind of BRDFs from limited image data, it is preferred to only only consider isotropic materials (such as the 100 ones in [4]) to minimise the overfitting chances. Thus, this corresponds to ignoring ϕ_h. Finally, to simplify the learning procedure, we explicitly factor out the the incident light component $(\mathbf{n} \cdot \mathbf{l})$ and thus learn BRDF $:= (\mathbf{n} \cdot \mathbf{l_m})\text{MLP}(\theta_h, \theta_d, \phi_d)$ (in fact depending on the literature sources, the BRDF is usually defined as the remaining component after $(\mathbf{n} \cdot \mathbf{l})$ is factored out).

For the MLP component, we use three fully connected layers containing 16 units each with relu activation followed by one fully connected with a single value output and exponential activation function. That forces the values to be always non-zero and encourages them to be around 1 which

corresponds to Lambertian reflectace[5] and it is a reasonable mean value.

Visualisation of the renderings and recovered BRDFs for all LUCES-ST objects are shown in Figure 2. It is observed that for the metallic objects (Bell, Cup) very narrow specular lobes are recovered; specular dielectric materials (Bunny, Hippo and Squirrel) have less peaky specular lobes and for mostly diffuse objects (Owl,Queen) a mostly diffuse BRDF is recovered.

References

[1] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Sdps-net: Self-calibrating deep photometric stereo networks. In *CVPR*, 2019. 1

[2] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Trans. Image Process.*, 2020. 1

[3] Fotios Logothetis, Roberto Mecca, Ignas Budvytis, and Roberto Cipolla. A cnn based approach for the point-light photometric stereo problem. *IJCV*, 2022. 1, 2

[4] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM TOG*, 2003. 2, 3

[5] Roberto Mecca, A. Wetzler, A. Bruckstein, and R. Kimmel. Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 2014. 1

[5] Advanced graphics models computing multiple light bounches include a $1/\pi$ factor for energy conservation but that is beyond the scope of out work since we do not compute self reflections and thus any scaling factor is meangless.