EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2020)
F.-J. Melero and M. Spagnuolo (Editors)

# SynthPS: a benchmark for evaluation of Photometric Stereo algorithms for Cultural Heritage applications

T. G. Dulecha[1], R. Pintus[2], E. Gobbetti[2], and A. Giachetti[1]

[1] University of Verona, Italy
[2] CRS4, Italy

**Abstract**

*Photometric Stereo (PS) is a technique for estimating surface normals from a collection of images captured from a fixed viewpoint and with variable lighting. Over the years, several methods have been proposed for the task, trying to cope with different materials, lights, and camera calibration issues. An accurate evaluation and selection of the best PS methods for different materials and acquisition setups is a fundamental step for the accurate quantitative reconstruction of objects' shapes. In particular, it would boost quantitative reconstruction in the Cultural Heritage domain, where a large amount of Multi-Light Image Collections are captured with light domes or handheld Reflectance Transformation Imaging protocols. However, the lack of benchmarks specifically designed for this goal makes it difficult to compare the available methods and choose the most suitable technique for practical applications. An ideal benchmark should enable the evaluation of the quality of the reconstructed normals on the kind of surfaces typically captured in real-world applications, possibly evaluating performance variability as a function of material properties, light distribution, and image quality. The evaluation should not depend on light and camera calibration issues. In this paper, we propose a benchmark of this kind, SynthPS, which includes synthetic, physically-based renderings of Cultural Heritage object models with different assigned materials. SynthPS allowed us to evaluate the performance of classical, robust and learning-based Photometric Stereo approaches on different materials with different light distributions, also analyzing their robustness against errors typically arising in practical acquisition settings, including robustness against gamma correction and light calibration errors.*

## 1. Introduction

Photometric Stereo (PS) is a shape reconstruction technique that relies on multi-light image collections (MLIC), i.e., sets of images of a surface captured from a fixed viewpoint with changing illumination direction. Basically, it is a technique for estimating the surface normals of an object given constraints on lights and reflectance properties of materials. The 3D shape can then be fully recovered from dense normal maps through spatial integration.

PS was first introduced by Woodham [Woo80] for pure Lambertian light scattering. Since then, a lot of work has been done to extend the original idea to surfaces with general reflectance properties and/or to improve the estimation precision. In recent years, in addition to the classical techniques, (deep)neural network based photometric stereo approach is also emerging [SSS*17, HGGL18, TM18, CHW18, CHS*19].

As the capture of MLIC data is flexible and affordable, it is quite popular in the Cultural Heritage (CH) domain. In this context, PS could be a really useful tool to acquire high-resolution geometrical detail of the objects. However, it is not very clear which are the algorithmic choices that should be adopted to optimize the reconstruction quality. PS algorithms are typically evaluated on a few benchmarks not particularly representative of typical CH applications, and they do not allow the evaluation of the variability of the methods' performances as a function of material types, camera and light calibration error, and image pre-processing.

In this work, we propose a novel benchmark to test PS methods that is specifically designed to evaluate the variability of algorithms' performance in different contexts. The paper is organized as follows: Sec. 2 presents the related work on PS and PS benchmarking; Sec. 3 presents the proposed benchmark; Sec. 4 discusses the performed tests that use SynthPS to compare state-of-the-art and baseline methods.

## 2. Related Work

### 2.1. PS approaches

Most of the existing methods for photometric stereo [SMW*10, PF14] assume a simplified reflectance model, such as the Lambertian model for its simplicity. However, this assumption doesn't work due to the fact that most of the real-world objects are non-Lambertian. Thus, many photometric stereo algorithms have been developed to deal with non-Lambertian materials. The most popular ones are based on outlier rejection and the use of Lamber-



DOI: 10.2312/gch.20201288

tian model for the remaining inliers. Within this category, various methods have been proposed that rely on different principles, such as RANSAC [MIS07], median values [MHI10], expectation maximization [WT10], sparse Bayesian regression [IWMA12], Least Median of Squares [DHOMH12, PGPG17], Low-rank matrix completion and recovery [WGS*10], or Sparse Regression [IWMA14].

Unlike outlier rejection methods, methods based on a sophisticated (analytical) reflectance model fit a model to all observations. This is achieved by solving complex optimization problems. Many sophisticated analytical reflectance models have been proposed to approximate the behavior non-Lambertian materials, including the Torrance-Sparrow model [Geo03], Ward model [CJ08], Cook-Torrance model [RK09], etc. The downside of this type of methods is that they can only handle limited classes of material. On the other hand, example based methods typically require an example object with known surface normal, shape and reflectance, to be placed in the scene. Usually, this requirement limits its practical use, so other approaches employ a dictionary of BRDFs to render virtual examples that guide the normal estimation problem. To this end, Hui and Sankaranarayanan [HS15] proposed a BRDF dictionary to render virtual spheres without using a real example object.

Learning based methods estimate mappings from measured intensities under known (or unknown) lighting to surface normals by using machine learning tools or deep learning techniques [IA14, CHW18, CHS*19]. Deep learning is a powerful learning method inspired by how the brain works. Convolutional or Fully-Connected Neural Network based methods have recently replaced traditional PS techniques. Santo et al. [SSS*17] proposed a deep fully-connected neural network, called Deep Photometric Stereo Network (DPSN), to learn the mapping between reflectance observations and surface normals in a per-pixel manner given a fixed number of observations captured under a pre-defined set of light directions. In this work, for each image point of the object, all its observations and light directions are concatenated to form a fixed-length vector, which is fed into a fully-connected network to regress a single normal vector. The weakness of this work arises from the assumption that the light directions are pre-defined and remain the same between training and prediction phases, which in turn limits its practical use. Chen et al. proposed two different methods based on convolutional neural networks [CHW18, CHS*19]. The first method [CHW18] proposes a flexible fully convolutional network, called PS-FCN, for estimating a normal map of an object. The network consists in three components, namely a shared-weight feature extractor for extracting feature representations from the input images, a fusion layer for aggregating features from multiple input images, and a normal regression network for inferring the normal map. The second method proposes a two-stage model named Self-calibrating Deep Photometric Stereo Networks (SDPS-Net). The first stage of SDPS-Net, denoted as Lighting Calibration Network (LCNet), takes an arbitrary number of images as input and estimates their corresponding light directions and intensities. The second stage, denoted as Normal Estimation Network (NENet), estimates a surface normal map of a scene based on the lighting conditions estimated by LCNet and the input images. In both cases, to simulate real-world, complex non-Lambertian surfaces they trained their model on synthetic datasets created using shapes from the blobby shape dataset [JA11] and the sculpture shape dataset [WZ17], and BRDFs from the MERL BRDF dataset [MPBM03]. Another Neural Network based method, which have demonstrated the possibility of recovering normals better than traditional methods, is the one proposed by Ikehata [Ike18]. This network accepts an arbitrary number of input images, merge them into the intermediate representation called observation map, which has a fixed shape, and then regresses the normal map. This network can directly learn the relationships between the photometric stereo input and surface normals of a scene. For a detailed and up-to-date survey please refer to Shi et al. [SMW*19].

### 2.2. PS benchmarks

The most popular PS benchmark is currently the DiLiGenT dataset [SWM*16]. It is composed by captured images of real objects made of different materials, together with accurate metadata about light and camera calibration. The dataset has been used in a large number of papers, and with calibrated and uncalibrated PS methods. Sablatnig and Wimmer [BZS18] used a dataset of ancient coins to evaluate the accuracy of a single PS method when the light sampling is varied. However, no details on ground truth estimation are provided and those data have not been publicly released. Xiong et al. [XCB*15a] created a PS dataset with seven relatively-diffuse objects, captured with a CanonEOS 40D camera under directional lighting, and calibrated with chrome spheres [XCB*15b]. Alldrin et al. [AZK08] acquired high-dynamic-range images of two known objects in a dark room and calibrated the lights by employing reflective spheres.

While the use of real, calibrated images allows the evaluation of the methods on real materials and illumination, there are still few drawbacks. Light intensity and direction need to be calibrated, and it is not possible to avoid calibration errors. The camera model is not orthographic and lights are not directional but are approximately point lights. Shi et al. [SWM*16] provide both light positions and camera parameters, but often they are not used by many tested algorithms, which assume orthographic camera model and directional lights. This may introduce a bias in the evaluation. Moreover, real lights are typically not uniform, and the intensity calibration can introduce errors as well. The objects and the acquisition setups used to create the benchmarks are not really similar to those employed in the classical handheld RTI or light dome acquisition methods typically used in the Cultural Heritage domain. Furthermore, by using real acquisitions it is not possible to control and modify the local properties of the materials so it is hard to evaluate how the different PS methods behave with varying reflectance functions and spatial variations of them.

A multi-light images dataset with synthetic rendering (CyclesPS) has already been proposed, for example in [Ike18]. However, it only features non-realistic objects with different materials assigned in superpixels, used to train (and test) local CNN-based PS algorithms.

This motivated us to create SynthPS, i.e., a specific dataset with synthetic, physically-based renderings of surfaces that can be considered typical examples of Cultural Heritage objects, which are made of different homogeneous and heterogeneous materials, and have different geometrical complexity.

These features make the proposed dataset also suitable to design specific evaluation tests, testing the robustness of the different methods to a variety of specific factors (materials, material uniformity, depth variations, etc.) that are useful to choose the best approach for practical applications.

We rendered the objects with different configurations of uniform directional lights, by changing their number and their spatial configuration. We use this dataset to evaluate several recent PS approaches, both in terms of their accuracy and robustness against different factors. The resulting contribution consists in several guidelines to choose the proper image capture strategy and processing algorithm, together with a public benchmark that can be used by researchers to evaluate novel PS methods.

## 3. SynthPS: image collections and tasks

SynthPS is a set of multi-light image collections synthetically rendered with the physically-based reflectance models of the Blender Cycles engine. The dataset will be publicly distributed after paper publication. Each collection is rendered with 77 directional lights placed in concentric rings in the $l_x, l_y$ plane at 9 different elevation values (10,20,30,40,50,60,70,80,90 degrees). Figure 4 shows the full set of light directions used. Lights are exactly directional and no ambient light is employed. The camera model is orthographic, the size of the rendered images is $320 \times 320$, and the depth is 16 bits linearly mapped. SynthPS is composed of two subsets: Single-material, i.e., all the items are created using constant materials; Multi-material, i.e., all the items (different from those used in the Single-material case) are textured with captured spatially-varying albedo and subdivided in patches with different roughness.

The Single material dataset has been created with three untextured geometric models of CH items, downloaded from SketchFab (*https://sketchfab.com*) and distributed under the Creative Commons 4.0 license (Figure 1, top). The first is a nearly-flat surface, actually the 3D scan of an oil on canvas painting by W. Turner, performed by R.M. Navarro. The second is the scan of a cuneiform tablet from Colgate University. The third is the scan of relief in marble "The dance of the Muses on Helicon" by G. C. Freund, digitized by G. Marchal. For these renderings, we have set an orthographic camera looking at the object surfaces, removed the ambient illumination, and rendered the set of images with the 77 directional lights. For each model, we created 9 collections of images assigning 9 different uniform materials to the surfaces. The assigned materials simulate matte, plastic and metallic behaviors with varying gray achromatic albedo and roughness and material with subsurface scattering (subsurface parameter set to 0.5 and radii 1.0,0.2,0.1 in the Blender PBR settings). These sets are reported in Table 1. Metal=1 means no diffuse scatter and specular reflection tinted with the base color. Specular=1 means dielectric specular reflection equal to 8%. Roughness is the microfacet roughness of the surface for diffuse and specular reflection. This allows comparing the performances of the PS methods when specific material features are changed, e.g., albedo, roughness, subsurface scattering. It is also possible to understand how the methods are robust to shape variations, locally by plotting error as a function of the normal/view angle, by globally comparing the results on the different objects of different complexity. One way to evaluate the object complexity is to evaluate the amount of shadow created on the different images that can be evaluated on the related rendering pass. For the three untextured objects, the total percentages of shadowed pixels in the 77 images are $0.20\%, 23.88\%, 16.68\%$, respectively. Figure 2 shows example renderings of the same object with the assigned 9 uniform materials, illuminated from the same light direction.

| # | material | albedo | metal | spec | rough | subs |
|---|---|---|---|---|---|---|
| 1 | matte white (MW) | 0.8 | 0 | 0 | 0 | |
| 2 | white plastic smooth (PLA_WS) | 0.8 | 0 | 1 | 0.4 | |
| 3 | white plastic (PLA_W) | 0.8 | 0 | 1 | 0.5 | |
| 4 | white plastic rough (PLA_WR) | 0.8 | 0 | 1 | 0.6 | |
| 5 | gray plastic (PLA_G) | 0.5 | 0 | 1 | 0.5 | |
| 6 | dark plastic (PLA_D) | 0.2 | 0 | 1 | 0.5 | |
| 7 | metal smooth (MET_S) | 0.8 | 1 | 0 | 0.5 | |
| 8 | metal rough (MET_R) | 0.8 | 1 | 0 | 0.7 | |
| 9 | subsurface (SUB) | 0.5 | 0 | 1 | 0.5 | * |

**Table 1:** *Parameters of the Cycles principled BSDF model used to create the 9 materials of the single material dataset.*

**Figure 1:** *Geometrical models used to create the SynthPS dataset. Top left: untextured models used for the SingleMaterial renderings. Bottom right: textured models used for the MultiMaterial renderings.*

The Multi-material dataset has been created with two different textured geometric models of CH items. All two have been digitized by G. Marchal and distributed under the Creative Commons 4.0 license. The first is the reconstruction of The lion of Goddess Ishtar, from Nye Carlsberg Glyptotek downloaded from SketchFab. The second is a totem pole (Giant-Cannibal with eagle and copper plate in his hand) of Kwakiutl Culture (British Columbia) from Musée du Cinquantenaire (Brussels, Belgium). For these models, we kept the original texture as the diffuse texture in the material

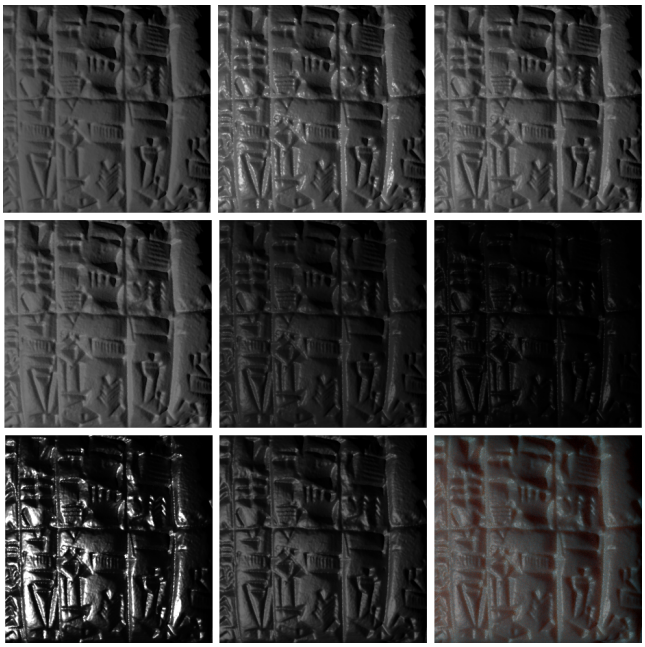

**Figure 2:** *Sample images of the same CH object with the assigned 9 materials and lit by the same illumination direction.*

settings and then used the Voronoi Texture node of Blender to define a spatial pattern used to map different roughness values for the material in different patches, ranging from 0.4 to 0.7 (see Figure 3).

For these two objects, the total percentage of shadowed pixels in the 77 images collections is 8.00%, 58.84%, respectively.

**Figure 3:** *Left: the Voronoi texture associated with the assigned roughness. Center, right: two renderings with different directional lights of the surface with the albedo texture and the regional roughnesses.*

The goal is to understand the limits and the peculiarities of each approach by evaluating the accuracy of different state-of-the-art approaches to recover surface normals as a function of different characteristics of the input images (number and distribution of the lights, and material properties like albedo and roughness) removing biases due to light calibration, and also testing the robustness of the methods against light calibration errors. Benchmarked tasks are therefore the recovery of the surface normals from all the rendered collections (varying single materials and materials mixtures) by using the full set of light directions and selected subsets (see Figure 4). In this way, we test how the methods perform in acquisitions made with dense, intermediate and sparse dome light configurations (77, 49 and 28 lights), and in the case of a single elevation ring (10 lights), as suggested by Sablatnig and Wimmer [BZS18]. We also tested the performances in the case of asymmetric light distributions, which may occur in handheld onsite acquisitions when selected regions around the object of interests are not accessible, and in the case of controlled simulated errors in light direction calibration. A further test has been designed to check the robustness of calibrated PS methods against random error in light direction calibration. We associated the sets of images with both exact and noisy light direction files obtained by deforming each light vector by a fixed angle along a random axis. Original linear 16-bits image files have also been linearly and nonlinearly (standard gamma correction) converted to 8 bits and jpeg-compressed images, to both test the robustness of the methods to different image preprocessing steps, and to derive related guidelines for practitioners.

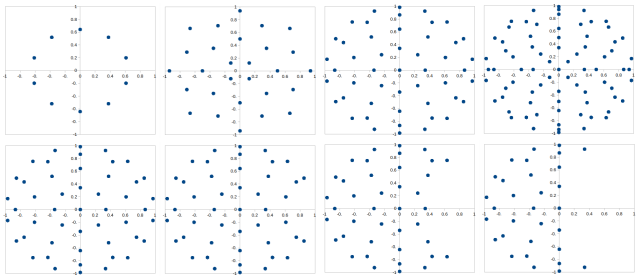

**Figure 4:** *Dome light configuration used for SynthPS benchmarking. Top: symmetric domes, from left, 10 light ring at 50 deg. elevation, 28, 49 and 77 directions domes. Bottom: asymmetric configuration created by removing lights from right in the 49 lights dome. From left: original, 46 lights, 40 lights, 31 lights represented in the $l_x$, $l_y$ plane.*

To wrap up, we have a novel dataset simulating CH objects with different materials and materials distributions, and a set of predefined tasks that allow evaluating calibrated (and uncalibrated) PS methods by measuring their dependency on material properties, shape features, image depth and linearity.

## 4. Evaluation

### 4.1. Tested algorithms

In this work, we have used out benchmark dataset to test several PS methods, ranging from classical ones, to robust approaches, and those based on (deep) neural networks. The main reasons for this selection are related to the popularity of the chosen methods, the fact that they exhibit good results in existing benchmarks, and the public availability of their implementations.

The selected PS algorithms based on model fitting are: standard Least Squares fitting of Lambertian model (LS) [Woo80]; Trimmed Least Squares (Trimmed), i.e., LS fed with pruned data, typically obtained by removing saturated values and a fixed percentage of high, possibly non-Lambertian, measures (in our tests 5%); Least Median of Squares(LMS) [DHOMH12, PGPG17]; Bayesian Regression(BR) [IWMA12]; Low-rank matrix completion and recovery (LMR) [WGS*10]; Sparse Regression [IWMA14].

Among deep learning algorithms, we either choose local or global methods; in the former ones, the normals are obtained on a per-pixel basis, while the latter train the network on the entire image domain. Furthermore, we also considered some uncalibrated methods (i.e., unknown input light directions). The tested deep learning based algorithms are: PS-FCN [CHW18] (calibrated / global); CNN-PS [Ike18] (calibrated / local); SDPS-Net [CHS*19] (uncalibrated / global). PS-FCN is a CNN based architecture that accepts a set of images as input, extracts features for each image, aggregates them via max-pooling, and finally infers the normal map. The method does not require a pre-defined set of light directions during training and testing, and it can handle multiple images and light directions in an order-agnostic manner. CNN-PS accepts as input an observation map, which is a 32x32 grid that encodes light intensity as a function of the light direction. A wide set of synthetic images is used in the training phase. Several different maps are generated in the prediction step, forcing a rotational invariance nature of the normal map estimation, which makes the method more robust. SDPS-net is a two-stage CNN based architecture. The first stage estimates the light directions, while the second stage uses the result of the first stage together with the captured images to compute the final normal map. Despite being uncalibrated, the method exhibits state-of-the-art performances on the DiLiGenT benchmark and performs well on challenging surfaces. However, it assumes surfaces with non-negligible relief and made of homogeneous material.

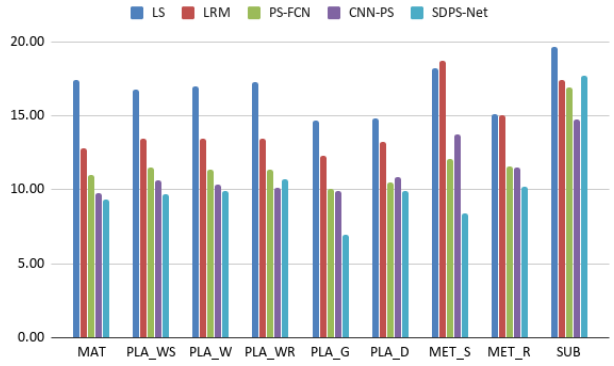


**Figure 5:** *Average errors for selected techniques: basic Lambertian (LS), a robust fitting (LRM), calibrated global network-based method (PS-FCN), calibrated local network-based method (CNN-PS) and uncalibrated network-based method (SDPS-Net) on the 49-lights normal recovery for the bas-relief models as a function of the different uniform materials.*

## 4.2. Results

The comparison of the methods is based on the statistics of angular error. For each pixel, the angular error is calculated as $\arccos(n_o^T n)$ in degrees, where $n_0$ and n are 'ground truth' and estimated normals respectively.

| | LS | Trim. | LMS | BR | LRM | SR | PSFCN | CNNPS | SDPS |
|---|---|---|---|---|---|---|---|---|---|
| MAT | 0.35 | 0.56 | 0.83 | 0.28 | 0.26 | **0.24** | 1.58 | 0.26 | 17.91 |
| PLA_WS | 0.33 | 0.59 | 0.39 | 0.40 | 0.26 | **0.24** | 1.72 | 0.33 | 14.44 |
| PLA_W | 0.32 | 0.50 | 0.40 | 0.25 | **0.23** | 0.24 | 1.10 | **0.23** | 17.38 |
| PLA_WR | 0.32 | 0.40 | 0.41 | 0.37 | 0.40 | **0.23** | 1.83 | 0.25 | 23.91 |
| PLA_G | 0.37 | 0.77 | 0.40 | 0.73 | 0.71 | 0.29 | 0.67 | **0.25** | 42.08 |
| PLA_D | 0.59 | 1.69 | 0.56 | 2.52 | 2.58 | 0.50 | 3.11 | **0.24** | 33.60 |
| MET_S | 0.79 | 3.45 | 0.77 | 4.25 | 4.35 | **0.68** | 1.87 | 0.76 | 7.13 |
| MET_R | 0.54 | 0.67 | 0.62 | 1.62 | 1.69 | 0.46 | 2.24 | **0.23** | 42.27 |
| SUB | 0.42 | 0.74 | 0.57 | 0.43 | 0.45 | 0.35 | 1.34 | **0.33** | 23.03 |
| **Average** | 0.45 | 1.04 | 0.55 | 1.21 | 1.21 | 0.36 | 1.72 | **0.32** | 24.64 |

(a)

| | LS | Trim. | LMS | BR | LRM | SR | PSFCN | CNNPS | SDPS |
|---|---|---|---|---|---|---|---|---|---|
| MAT | 17.40 | 16.57 | 12.86 | 13.59 | 12.82 | 18.36 | 10.99 | 9.78 | **9.35** |
| PLA_WS | 16.75 | 16.90 | 13.34 | 14.10 | 13.44 | 18.54 | 11.55 | 10.68 | **9.73** |
| PLA_W | 17.01 | 16.83 | 13.29 | 14.09 | 13.44 | 18.47 | 11.36 | 10.36 | **9.90** |
| PLA_WR | 17.31 | 16.80 | 13.25 | 14.11 | 13.45 | 18.48 | 11.40 | **10.17** | 10.69 |
| PLA_G | 14.71 | 15.03 | 12.57 | 12.87 | 12.30 | 16.89 | 10.08 | 9.93 | **7.00** |
| PLA_D | 14.85 | 15.15 | 13.70 | 13.64 | 13.27 | 15.42 | 10.53 | 10.84 | **9.95** |
| MET_S | 18.23 | 19.52 | 22.07 | 18.60 | 18.71 | 18.14 | 12.12 | 13.76 | **8.42** |
| MET_R | 15.12 | 16.17 | 16.42 | 15.20 | 15.06 | 16.55 | 11.59 | 11.52 | **10.21** |
| SUB | 19.67 | 20.05 | 17.29 | 17.89 | 17.46 | 20.85 | 16.91 | **14.77** | 17.71 |
| **Average** | 16.78 | 17.00 | 14.98 | 14.90 | 14.44 | 17.97 | 11.84 | 11.31 | **10.33** |

(b)

| | LS | Trim. | LMS | BR | LRM | SR | PSFCN | CNNPS | SDPS |
|---|---|---|---|---|---|---|---|---|---|
| MAT | 11.98 | 9.26 | 6.76 | 7.84 | 7.18 | 12.06 | 7.05 | **6.61** | 9.59 |
| PLA_WS | 10.67 | 9.55 | 7.18 | 8.11 | 7.58 | 11.84 | 7.26 | 7.14 | **6.68** |
| PLA_W | 10.98 | 9.52 | 7.19 | 8.14 | 7.61 | 11.77 | 7.05 | **6.69** | 7.45 |
| PLA_WR | 11.37 | 9.42 | 7.21 | 8.20 | 7.67 | 11.79 | 7.07 | **6.47** | 7.84 |
| PLA_G | 9.46 | 9.29 | 6.49 | 7.49 | 6.98 | 10.67 | 6.61 | 6.79 | **4.29** |
| PLA_D | 12.79 | 14.01 | 7.96 | 9.16 | 8.43 | 10.18 | 6.88 | 7.46 | **4.15** |
| MET_S | 16.83 | 18.59 | 14.03 | 13.43 | 12.77 | 12.41 | 7.65 | 9.15 | **5.34** |
| MET_R | 11.11 | 13.54 | 10.42 | 10.64 | 10.50 | 10.75 | 8.20 | 7.36 | **5.19** |
| SUB | 13.60 | 13.63 | 11.59 | 11.93 | 11.58 | 14.47 | 10.96 | **9.63** | 10.84 |
| **Average** | 12.09 | 11.87 | 8.76 | 9.44 | 8.92 | 11.77 | 7.64 | 7.48 | **6.82** |

(c)

**Table 2:** *Mean angular error (deg.) for the normal reconstructions on the three objects of the SingleMaterial dataset with simulated 49 light dome and all the assigned materials. (a) nearly planar canvas, (b) bas-relief, (c) tablet. Bold fonts indicate the best results.*

### 4.2.1. Uniform materials

For the three uniform material objects, the normal reconstruction errors for the typical 49-light dome configuration are reported in Table 2. It is possible to see that the accuracy of "global" neural methods is low for the nearly planar objects, while the local method (CNN-PS) provides good results like robust methods. CNN-based methods are far better than model-fitting techniques on normal-varying objects (bas-relief and tablet); although it explicitly needs shadows and relevant shading to solve for the light directions, the uncalibrated approach (SDPS) provides the best results.

Looking at the performances on different materials (Figure 5), it is possible to see that the performances of the best methods for each category are almost the same for matte and plastic materials, independently of roughness and are in some cases improved with smaller albedo values. Plots are done for the bas-relief results but are similar for the tablet. The performances of robust fitting meth-

ods are poor in the case of metallic surfaces. Neural methods are far better in this case, with SPDS-net (uncalibrated/global) generally providing the best results. Subsurface scattering is instead deteriorating the performances of all the methods in the same way.

Let us now look at the effect of dome light density. Figure 6 shows the average angular errors on the bas-relief dataset with varying density. It is possible to see that the average errors are nearly the same by reducing the light density to 28 lights in a reasonably regular distribution, even if not all the methods are robust against the replacement of the dome configuration with the 50 degrees ring proposed by Sablatnig and Wimmer [BZS18]. Neural methods are the best, but CNN-PS is not effective in the ring configuration (10 lights). This is due to the encoding of the pixel information as a 2D "observation map" that requires a reasonable density. The uncalibrated method performances are also decreased in the ring configuration as a larger number of lights are required for calibration.

Neural methods are also more robust against asymmetry of the light direction configuration as shown in Figure 7.

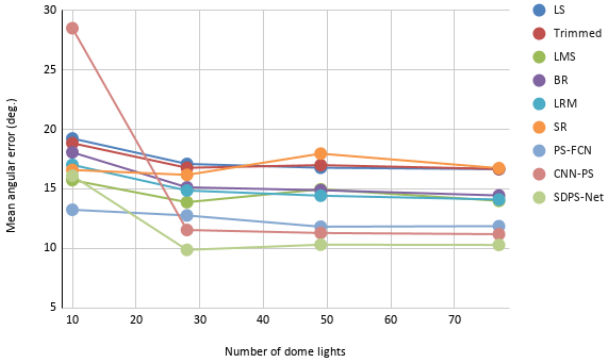


**Figure 6:** *Average mean angular errors (on all the different material) for the bas-relief normal estimation vs number of lights in radially symmetric configurations.*

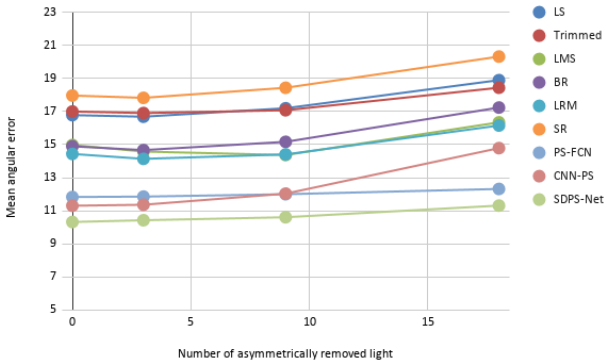


**Figure 7:** *Average errors on bas-relief normals estimation vs number of lights removed from a side of the 49-lights dome.*

It might be surprising to note that the uncalibrated method performs well despite the error in the light direction evaluation. However, it is possible to see that the addition of a randomly directed fixed-angle noise to the input light directions does not alter too much the accuracy of the methods as shown in Figure 8. The only method that seems to require accurate light directions is CNN-PS.

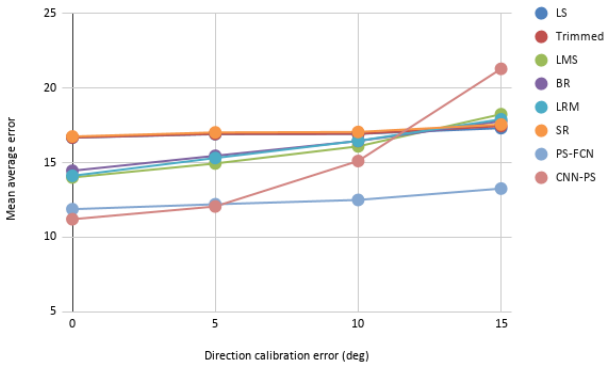


**Figure 8:** *Average errors on bas-relief normals estimation vs simulated error in light direction calibration.*

| | Object1 | Object2 | Object3 |
|---|---|---|---|
| MAT | 57.60 | 6.35 | 9.22 |
| PLA_WS | 61.71 | 8.33 | 8.39 |
| PLA_W | 58.44 | 8.06 | 8.51 |
| PLA_WR | 57.97 | 7.31 | 8.77 |
| PLA_G | 60.47 | 7.68 | 8.11 |
| PLA_D | 65.98 | 9.14 | 8.97 |
| MET_S | 59.76 | 12.03 | 10.20 |
| MET_R | 60.20 | 9.51 | 8.67 |
| SUB | 60.14 | 16.79 | 10.56 |
| Average | 60.25 | 9.47 | 9.04 |

**Table 3:** *Light direction estimation errors obtained on the 49-lights image collections with the self-calibration module of SDPS-Net on the three uniform surfaces with different materials assigned*

This fact also explains the success of the uncalibrated SPDS-Net method: as shown in Table 3 the average accuracy of the light direction estimated by its LCNet subnet estimating the light directions is not too accurate: fails on Object 1 and has a MAE close to 10 degrees on the depth varying surfaces, but the normal estimation subnet is robust against the error in the input light direction when it is lower than 10 degrees.

The average errors on the entire images, however, do not show how the performances of the methods change in critical regions, for example where the normals create large angles with the view directions and where nonlocal effects like shadows are relevant.

To analyze the performance of the methods in challenging regions we can look at error maps, like those shown in Figure 9, and plot methods accuracy against local properties, e.g., the z-component of the ground truth normals, or the percentage of shadowed light directions, which can be obtained from the synthetic rendering output. We report in Figure 9 the error maps for the bas-relief 49-lights dome test, material 3 (white plastic). It is possible

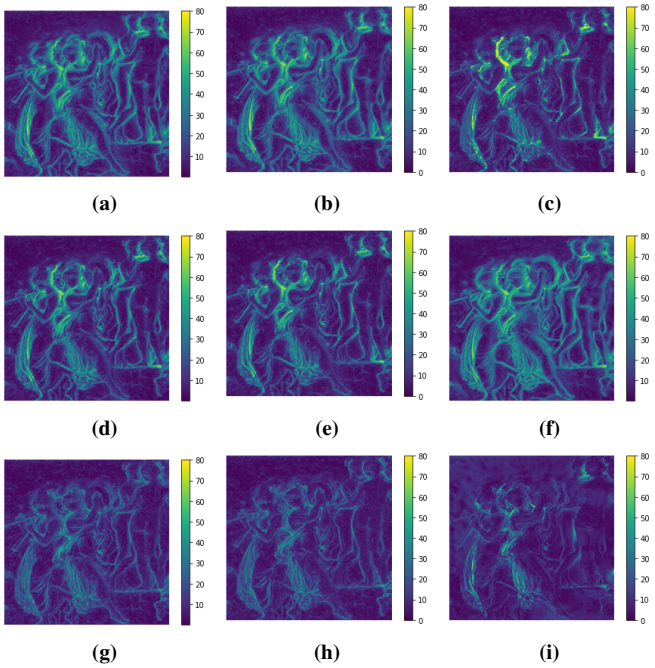


**Figure 9:** *Error maps encoding local errors in normals for the white plastic bas-relief captured with the simulated 49-lights dome and reconstructed with: (a) Least Squares; (b) Trimmed LS; (c) Least Median of Squares; (d) Bayesian Regression; (e) Low-Rank Matrix; (f) Sparse Regression; (g) PS-FCN; (h) CNN-PS; (i) SDPS-Net.*

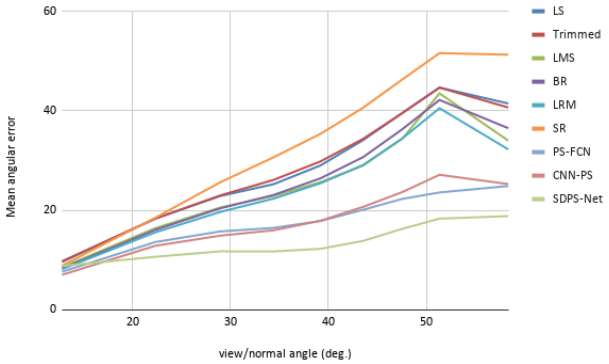


**Figure 10:** *Average MAE (all materials) obtained with all the tested methods on the pixels of the bas-relief 49-lights image collections vs angle between normal and view direction.*

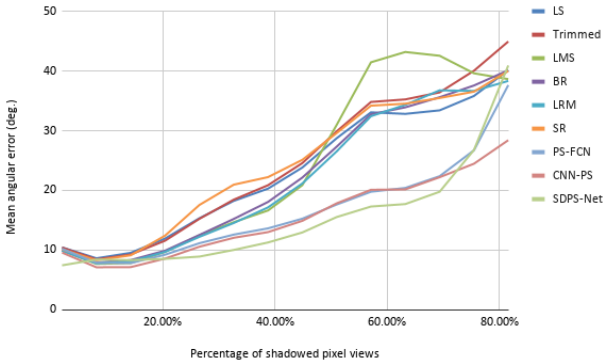


**Figure 11:** *Average MAE (all materials) obtained with all the tested methods on the pixels of the bas-relief 49-lights image collections vs percentage of shadowed directions.*

to see that, as expected, the error is higher where the surface is not flat and the effect of shadows and inter-reflections is higher; this is more evident for Lambertian model-fitting methods. This fact can be quantitatively shown by plotting the errors as a function of local surface properties. Figure 10 shows the average pixel errors on the 49-lights capture of the bas-relief model (average on all the materials) for pixels as a function of the angle between the actual surface normal and the view direction. It is possible to see that, while for surfaces nearly perpendicular to the view direction all the methods are similarly accurate, the errors grow with the angle at different ratios, with neural methods far better at large angles. Figure 11 shows the average pixel errors on the 49-lights capture of the bas-relief model (average on all the materials) as a function of the percentage of locally shadowed lights (known from the rendering step. The error is similar for all the methods and low when shadowed directions are less than 20%, then there is a rapid growth with neural methods far better. CNN-PS seems more robust when the percentage exceeds 70%. SDPS-Net, despite not using input light directions, provides the best results, even if the largest errors seem more relevant than those of other neural methods in Figure 9. As reflected in Figure 11, these points correspond to the ones with a large percentage of shadowed pixels.

### 4.2.2. Non-uniform materials

When the materials are not uniform, the ranking of the methods is completely different. As expected, due to the fact that the hypothesis of uniform material is exploited in the direction calibration network, SPDS-Net fails, but also PS-FCN is no longer performing well. CNN-PS is clearly the best one, probably due to its local nature, even if robust fitting methods are not too far, as shown in Table 4.

It is interesting to see in Figure 12 that the accuracy of CNN-PS is practically constant when the number of the directional dome

| | LS | Trim. | LMS | BR | LRM | SR | PSFCN | CNNPS | SDPS |
|---|---|---|---|---|---|---|---|---|---|
| Object1 | 5.52 | 5.21 | 4.38 | 4.81 | 4.67 | 5.19 | 6.99 | **4.04** | 12.10 |
| Object2 | 12.23 | 11.80 | 9.28 | 9.73 | 9.17 | 12.62 | 9.74 | **7.92** | 19.87 |
| **Average** | 8.88 | 8.51 | 6.83 | 7.27 | 6.92 | 8.92 | 8.37 | **5.98** | 15.99 |

**Table 4:** *Average of the MAE obtained by the different methods on the two multi-material objects. Bold fonts indicate the best results.*

light is reduced from 77 to 28 lights uniformly distributed, but this method completely fails if the dome is replaced with a single-elevation ring. This is due to the impossibility of estimating a dense observation map exploiting rotational symmetry. Least Median of Squared robust fitting, on the other hand, seems sufficiently accurate also in this case, even if not as in the uniform light dome configuration.

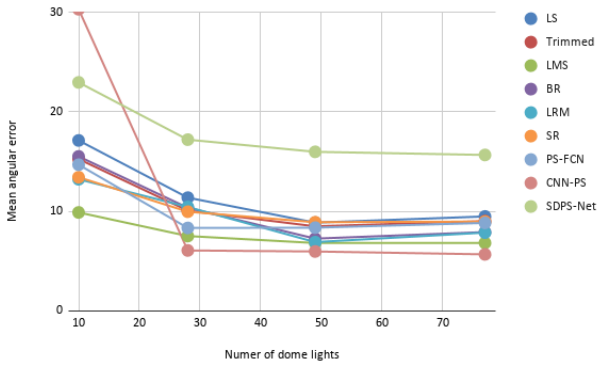


**Figure 12:** *Average MAE on the two multimaterial objects vs number of lights in a symmetric dome configuration.*

Another limit of the CNN-PS method, at least with the training provided by the authors, is related to the symmetry of the input light set. In onsite handheld acquisitions typical of the Cultural Heritage domain it is rather usual that lights cannot be placed on a side of the surface of interest. However, CNN-PS is sensitive to the asymmetrical removal of lights as shown in Figure 13

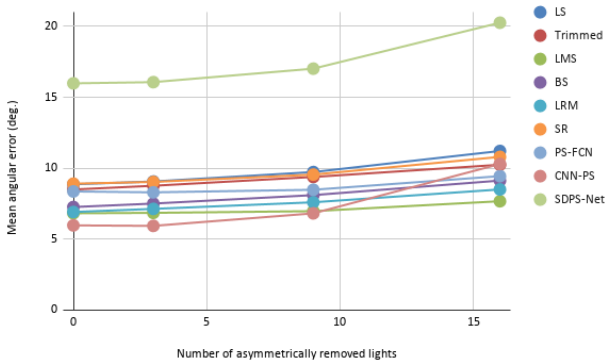


**Figure 13:** *Average MAE on the two multimaterial objects vs number of lights removed from a side of the simulated 49-lights dome.*

Looking at the error maps for the Lion 49-light capture (Figure 14) we can see that the error of SDPS-Net is strongly influenced by the background roughness (patches are visible in the error map). CNN-PS is clearly the most effective technique, even if robust fitting methods are effective as well.

The behavior of the methods is different with respect to the surface normal or the percentage of shadowed directions. Figure 15 shows that many robust fitting methods (but not LMS) present a

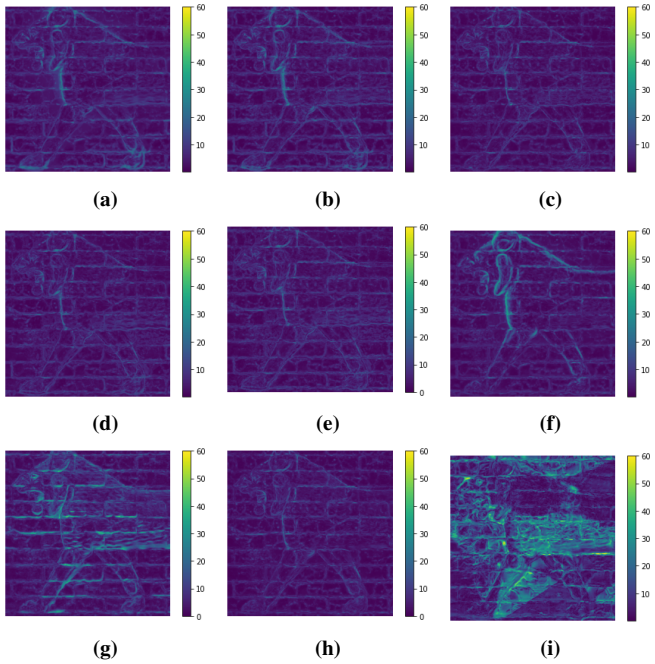


**Figure 14:** *Error maps encoding local errors in normals for a multimaterial object captured with the simulated 49-lights dome. (a) Least Squares. (b) Trimmed LS. (c) Least Median of Squares. (d) Bayesian Regression. (e) Low-Rank Matrix.(f) Sparse Regression.(g). PS-FCN (h) CNN-PS. (i) SDPS-Net.*

strong growth of the error with the angle between normal and view direction, while the error of SDPS-Net is large but mostly unaffected by normal direction.

If we look at the effect of shadowed directions, CNN-PS is constantly the best method independently of the amount of shadowed directions. LMS starts to fail when the number of shadow outliers exceeds 40%, while LRM behaves well and is quite close to CNN-PS at large values (see Figure 16).

### 4.2.3. Effect of different image encoding

In our synthetic image encoding, we did not apply gamma correction and we recorded a linear signal, encoded with 16 bits per channel, simulating raw images captured with a sensor with a high dynamic range. However, multi-light image captures often are performed with low-cost hardware and may be encoded as 8 bits images, possibly with gamma correction or unknown nonlinear mapping. We performed some tests also to verify the effects of this on the accuracy of the reconstructed normals performed with PS algorithms. Figure 17 shows the results obtained with the tested methods on the original 16 bits linear images (blue bars), on 8 bits linear images (red bars), 8 bits nonlinear images ($\gamma = 2.2$) remapped linearly before normal estimation, and 8 bits nonlinear images not corrected (purple bars). The results show that for simulated acquisitions the 8-bits discretization does not affect the results. With real acquisitions and non-ideal sensors, results can be different, but in

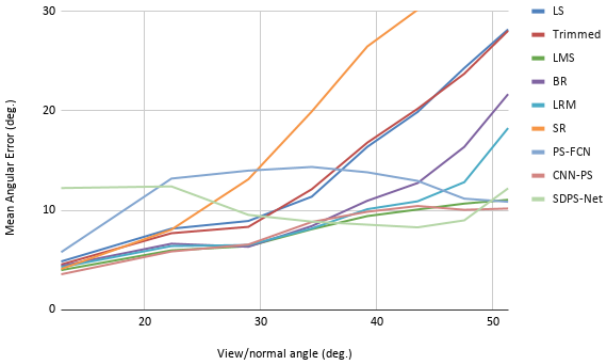


**Figure 15:** *Average MAE obtained on the pixels of the Lion 49-lights capture by the different methods vs angle between normal and view direction.*

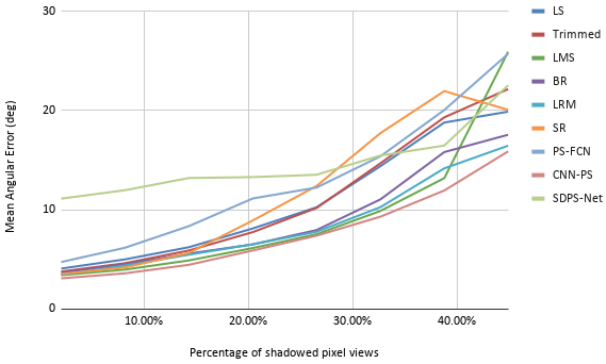


**Figure 16:** *Average MAE obtained on the pixels of the Lion 49-lights capture by the different methods vs percentage of shadowed directions.*

this case, there is no need for keeping the full dynamic range. Furthermore, it is interesting to note that the best method (CNN-PS) is rather robust with respect to the lack of correct linearization of the input illumination.

### 4.3. Discussion

In this paper we propose a novel benchmark (SynthPS) to evaluate Photometric Stereo algorithms to reconstruct surfaces of objects typically captured in multi-light acquisitions performed in the Cultural Heritage domain. The benchmark is composed of synthetic images simulating acquisitions of surfaces with different geometrical complexity, homogeneous and heterogeneous materials and assuming perfectly directional lights and orthographic view, thus avoiding inaccuracy in light or camera calibration. The image sets can be used to simulate light domes configurations with different density and asymmetric layout simulating the effect of obstacles preventing the positioning of the lights from one side of the object.

Exploiting the features of SynthPS we performed a set of tests on state of the art PS techniques, that can be used to derive suggestions

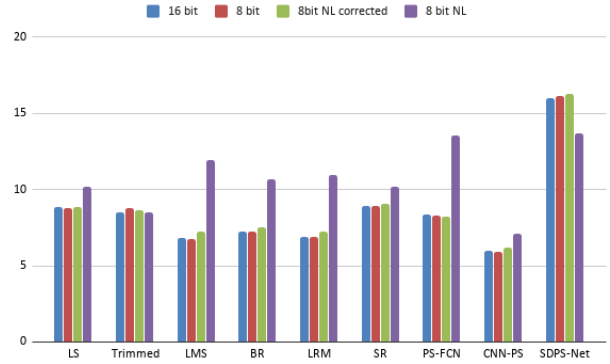


**Figure 17:** *Average MAE on the two multimaterial objects obtained with the tested methods on differently encoded images: 16 bits linear, 8 bits linear, 8 bits with standard gamma correction and inverse correction before fitting, 8 bits with gamma correction without linearization.*

or even guidelines for the practical reconstruction of normal maps and surfaces from MLIC data. Here are the main outcomes of our tests:

- Neural methods are quite promising, providing the best results in most of the comparisons. However, due to intrinsic limits or training on limited surface types, no single method can cope with all the possible acquisition settings.
- SDPS-Net is quite effective for uniform materials and does not require light calibration. However, as clearly stated by the authors, it does not work on nearly-planar objects and multi-material surfaces.
- PS-FCN even if calibrated is not accurate on nearly-planar objects and does not perform well on multimaterial objects. This may be due to the use of global information
- The methods based on robust Lambertian fitting works sufficiently well on matte and plastic materials, but less on metallic ones.
- CNN-PS seems the most reasonable choice in general, as it works well on nearly flat surfaces and it is the more effective in the case of non-uniform materials. However, it is more sensitive than the other techniques to the accuracy of light direction calibration and the asymmetry of the light layout. This fact is an intrinsic limitation of the method using "observation maps" as input but can be reduced using other learning based method to densify sparse input maps.
- Performances of the training-based methods depend on the training data used. Good performances of CNN-PS may be biased by the fact that it is trained with synthetic data rendered with Cycles as SynthPS. Performances of these methods can however be improved on different sets of data by enlarging the training set including different kinds of data.
- On well-exposed images it seems to be not crucial to keep 12-16 bits depth in input images as there is no apparent decrease in performance with a simple linear mapping onto 8 bits.

## 5. Conclusion

We proposed a novel benchmark for Photometric Stereo algorithms specifically tailored to verify robustness against different factors that are typically varied in cultural heritage acquisitions, e.g. material types, shape complexity, lights number and symmetry. Despite the limitations related to the synthetic renderings, that, on the other hand, ensure perfect control of calibration and imaging parameters as well as local material properties, we believe that it can be extremely useful to determine the best method to be used for each experimental setup. We plan to evaluate now other methods and to extend the dataset with other synthetic objects and acquisitions of real objects with both hires 3D scanning and multi-light capture.

**Acknowledgments** This work was supported by the DSURF (PRIN 2015) project funded by the Italian Ministry of University and the MIUR Excellence Departments 2018-2022. The project received funding Sardinian Regional Authorities under project VIGECLAB (POR FESR 2014-2020 Action 1.2.2).

## References

[AZK08] ALLDRIN N., ZICKLER T., KRIEGMAN D.: Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. CVPR* (2008), pp. 1–8. 2

[BZS18] BRENNER S., ZAMBANINI S., SABLATNIG R.: An Investigation of Optimal Light Source Setups for Photometric Stereo Reconstruction of Historical Coins. In *Eurographics Workshop on Graphics and Cultural Heritage* (2018), Sablatnig R., Wimmer M., (Eds.). 2, 4, 6

[CHS*19] CHEN G., HAN K., SHI B., MATSUSHITA Y., WONG K.-Y. K.: Sdps-net: Self-calibrating deep photometric stereo networks. In *CVPR* (2019). 1, 2, 5

[CHW18] CHEN G., HAN K., WONG K.-Y. K.: Ps-fcn: A flexible learning framework for photometric stereo. In *ECCV* (2018). 1, 2, 5

[CJ08] CHUNG H.-S., JIA J.: Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proc. CVPR* (2008), pp. 1–8. 2

[DHOMH12] DREW M. S., HEL-OR Y., MALZBENDER T., HAJARI N.: Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing 30*, 4-5 (2012), 317–331. 2, 4

[Geo03] GEORGHIADES A. S.: Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *null* (2003), Ieee, p. 816. 2

[HGGL18] HOLD-GEOFFROY Y., GOTARDO P. F., LALONDE J.-F.: Deep photometric stereo on a sunny day. *arXiv preprint arXiv:1803.10850* (2018). 1

[HS15] HUI Z., SANKARANARAYANAN A. C.: A dictionary-based approach for estimating shape and spatially-varying reflectance. In *Proc. ICCP* (2015), pp. 1–9. 2

[IA14] IKEHATA S., AIZAWA K.: Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proc. CVPR* (2014), pp. 2179–2186. 2

[Ike18] IKEHATA S.: Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *Proc. ECCV* (2018), pp. 3–18. 2, 5

[IWMA12] IKEHATA S., WIPF D., MATSUSHITA Y., AIZAWA K.: Robust photometric stereo using sparse regression. In *Proc. CVPR* (2012), IEEE, pp. 318–325. 2, 4

[IWMA14] IKEHATA S., WIPF D., MATSUSHITA Y., AIZAWA K.: Photometric stereo using sparse bayesian regression for general diffuse surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 36*, 9 (2014), 1078–1091. 2, 4

[JA11] JOHNSON M. K., ADELSON E. H.: Shape estimation in natural illumination. In *Proc. CVPR* (2011), IEEE, pp. 2553–2560. 2

[MHI10] MIYAZAKI D., HARA K., IKEUCHI K.: Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision 86*, 2-3 (2010), 229. 2

[MIS07] MUKAIGAWA Y., ISHII Y., SHAKUNAGA T.: Analysis of photometric factors based on photometric linearization. *JOSA A 24*, 10 (2007), 3326–3334. 2

[MPBM03] MATUSIK W., PFISTER H., BRAND M., MCMILLAN L.: A data-driven reflectance model. *ACM Trans. Graph. 22*, 3 (July 2003). 2

[PF14] PAPADHIMITRI T., FAVARO P.: A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International journal of computer vision 107*, 2 (2014), 139–154. 1

[PGPG17] PINTUS R., GIACHETTI A., PINTORE G., GOBBETTI E.: Guided robust matte-model fitting for accelerating multi-light reflectance processing techniques. In *Proc. BMVC* (2017). 2, 4

[RK09] RUITERS R., KLEIN R.: Heightfield and spatially varying brdf reconstruction for materials with interreflections. *Computer Graphics Forum (Proc. of Eurographics) 28*, 2 (Apr. 2009), 513–522. 2

[SMW*10] SHI B., MATSUSHITA Y., WEI Y., XU C., TAN P.: Self-calibrating photometric stereo. In *Proc. CVPR* (2010), IEEE, pp. 1118–1125. 1

[SMW*19] SHI B., MO Z., WU Z., DUAN D., YEUNG S., TAN P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence 41*, 2 (2019), 271–284. 2

[SSS*17] SANTO H., SAMEJIMA M., SUGANO Y., SHI B., MATSUSHITA Y.: Deep photometric stereo network. In *Proc. ICCV* (2017), pp. 501–509. 1, 2

[SWM*16] SHI B., WU Z., MO Z., DUAN D., YEUNG S.-K., TAN P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proc. CVPR* (2016), pp. 3707–3716. 2

[TM18] TANIAI T., MAEHARA T.: Neural inverse rendering for general reflectance photometric stereo. In *International Conference on Machine Learning* (2018), pp. 4864–4873. 1

[WGS*10] WU L., GANESH A., SHI B., MATSUSHITA Y., WANG Y., MA Y.: Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision* (2010), Springer, pp. 703–717. 2, 4

[Woo80] WOODHAM R. J.: Photometric method for determining surface orientation from multiple images. *Optical engineering 19*, 1 (1980), 191139. 1, 4

[WT10] WU T.-P., TANG C.-K.: Photometric stereo via expectation maximization. *IEEE transactions on pattern analysis and machine intelligence 32*, 3 (2010), 546–560. 2

[WZ17] WILES O., ZISSERMAN A.: Silnet: Single-and multi-view reconstruction by learning from silhouettes. *arXiv preprint arXiv:1711.07888* (2017). 2

[XCB*15a] XIONG Y., CHAKRABARTI A., BASRI R., GORTLER S. J., JACOBS D. W., ZICKLER T.: From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 1 (Jan 2015), 67–79. 2

[XCB*15b] XIONG Y., CHAKRABARTI A., BASRI R., GORTLER S. J., JACOBS D. W., ZICKLER T.: From shading to local shape - dataset, 2015. [Online; accessed 25-July-2020]. URL: http://vision.seas.harvard.edu/qsfs/. 2