

A CNN Based Approach for the Near-Field Photometric Stereo Problem

Fotios Logothetis¹
flogothetis@crl.toshiba.co.uk

Ignas Budvytis²
ib255@cam.ac.uk

Roberto Mecca¹
rmecca@crl.toshiba.co.uk

Roberto Cipolla²
rc10001@cam.ac.uk

¹ Cambridge Research Laboratory
Toshiba Europe Ltd.
Cambridge, UK

² University of Cambridge
Cambridge, UK

Abstract

Reconstructing the 3D shape of an object using several images under different light sources is a very challenging task, especially when realistic assumptions such as light propagation and attenuation, perspective viewing geometry and specular light reflection are considered. Many of works tackling Photometric Stereo (PS) problems often relax most of the aforementioned assumptions. Especially they ignore specular reflection and global illumination effects. In this work, we propose the first CNN based approach capable of handling these realistic assumptions in Photometric Stereo. We leverage recent improvements of deep neural networks for far-field Photometric Stereo and adapt them to near field setup. We achieve this by employing an iterative procedure for shape estimation which has two main steps. Firstly we train a per-pixel CNN to predict surface normals from reflectance samples. Secondly, we compute the depth by integrating the normal field in order to iteratively estimate light directions and attenuation which is used to compensate the input images to compute reflectance samples for the next iteration. To the best of our knowledge this is the first near-field framework which is able to accurately predict 3D shape from highly specular objects. Our method outperforms competing state-of-the-art near-field Photometric Stereo approaches on both synthetic and real experiments.

1 Introduction

Retrieving the 3D shape of an object from observations under varying illumination is a very challenging problem in Computer Vision introduced by [BZ] under the name of Photometric Stereo (PS). The mathematical model of the Photometric Stereo problem proposed in this work relied on four main assumptions: orthographic viewing geometry, diffuse light reflection, uniform light propagation and the lack of ambient nuisance. Due to restrictive assumptions, the method of [BZ] was limited to

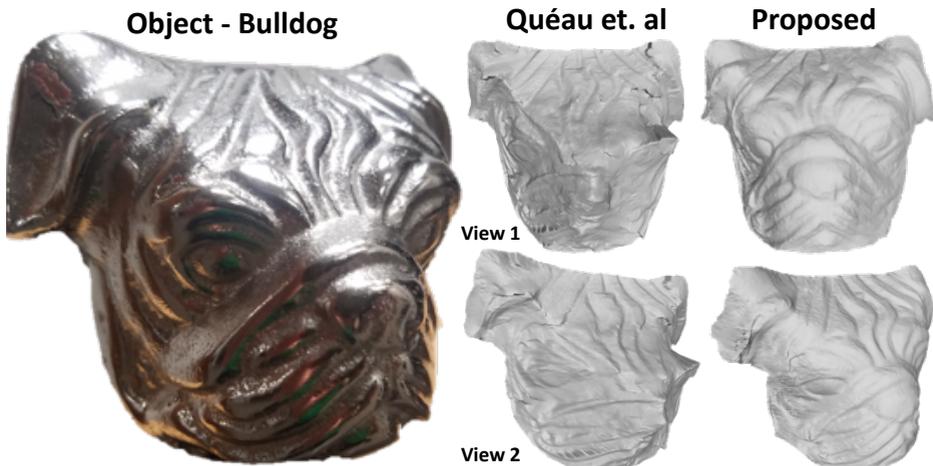


Figure 1: Our proposed approach accurately reconstructs highly specular objects, clearly outperforming the previous state-of-the-art [25].

very narrowly specified scenarios. Since then an extensive research has been carried out to relax these assumptions.

With the aim of solving the PS problem under more realistic conditions, researchers have modelled perspective viewing geometry [22, 23, 30], specular light reflections [24] and point light sources parameterising radial propagation of light [6, 10]. These effects lead to highly non-linear models requiring sophisticated optimisation strategies [25, 31]. As the complexity of the models becomes intractable, especially when dealing with realistic material properties, a lot of recent works have opted to neglect specular highlights and instead rely on robust optimisation techniques [9, 10]. Furthermore, materials that reflect light in a non-diffuse manner are prone to a number of complex physical effects which make the explicit mathematical modeling very hard to invert.

In fact, these global illumination effects (cast shadows, self reflections, ambient light) are one of the most challenging aspects of PS. [15, 32] tackle the case of fixed ambient light which however is too simple of a model to cover realistic inter-reflections. The global illumination issue is firstly adequately addressed in [8] by employing a Convolutional Neural Network (CNN). This method works by arranging reflectance (BRDF) samples into a fixed size observation map for each pixel. Observational maps are then provided as an input to the CNN which is trained to output a surface normal per pixel. [18] extends this work and shows how a training data augmentation strategy can be used to deal with general reflectance BRDF such as MERL dataset [19] or Disney BSDF [9] and global illumination effects in the far-field setting. However, these approaches are only directly applicable to the far-field photometric stereo since the nonlinear light attenuation of near-field images does not allow to directly compute valid observation maps.

We propose to overcome the aforementioned limitations by using a three step process. Firstly, the effect of the light attenuation is compensated using an estimate of the object geometry, to produce the equivalent far-field reflectance samples. Secondly, a CNN is used to regress pixel normals from these samples. Finally, a

numerical integration is used to update the estimate of the object geometry for the next iteration step. In addition, since during test time the observation map sample space is highly limited due to a fixed LED capture setup, we render our training observational maps from a virtual copy of this near-field image capture device. In this way we significantly reduce the training data variation and allow ourselves for various geometric augmentations such as pixel depth variation.

We evaluate our method on both artificial and real near-field image datasets. We significantly outperform competing approaches [16, 25] on both datasets, improving by more than 6° on the average angular error of the synthetic one.

The rest of this work is divided as follows. Section 2 discusses relevant work in Photometric Stereo. Section 3 provides details of our proposed method. Sections 4 and 5 describe the experiment setup and corresponding results.

2 Related Work

In this section we provide an overview of the relevant latest improvements. For a detailed, fairly recent PS survey, refer to [1].

Near-Field PS. Differently from the classical far-field PS, near-field approaches assume that the illumination is non-linear with respect to the position of the light sources, thus making analytical models a lot more complicated and harder to solve in practice. Some works embedded these non-linearities in a PDE-based formulation so the depth is directly calculated without passing through the computation of the normal field [20]. [13, 22] took advantage of image ratios in order to eliminate the dependence on the surface albedo and thus reduce the number of unknowns. Image ratios were also used in the variational framework of [21] in order to make the approach more robust to specular highlights by unifying diffuse and Blinn-Phong specular [2] reflections into a single mathematical formulation. This general variational framework is also applicable in a weekly calibrated setting [16] or even a volumetric one [17]. Recently, the LED-based approach introduced by Quéau *et al.* [25] presented a convoluted variational approach based on alternating weighted least-square scheme also capable of calibrating the light brightness of the light sources. Finally, [14] exploited a circular LED setup to compute the relative mean distance between the camera and the object.

Deep Learning based approaches for PS. Computer graphics is a well understood topic and many tools capable of rendering highly non-linear irradiance equations are publicly available¹ [19]. This allowed to create reliable datasets for supervised DL approaches. The potential of DL for solving the PS problem can be divided in two main advantages. Firstly, CNNs have the capability of inverting highly non-linear reflectance models comprising of numerous physically based parameters. Secondly, CNNs can be made to deal with the complicated real world imperfection (shadows, self reflections, noise) through the use of data augmentation. So far, several DL approaches have been proposed. A preliminary work by [7, 28] considered diffuse reflection only. [53] proposed a library where set of novel layers can be incorporated into a generic neural network to embed explicit models of photometric image formation. More recently, several approaches have tackled the problem of reconstructing complex objects. [26] proposed a method to find correspondences between simu-

¹www.blender.org and www.disneyanimation.com/technology/brdf.html

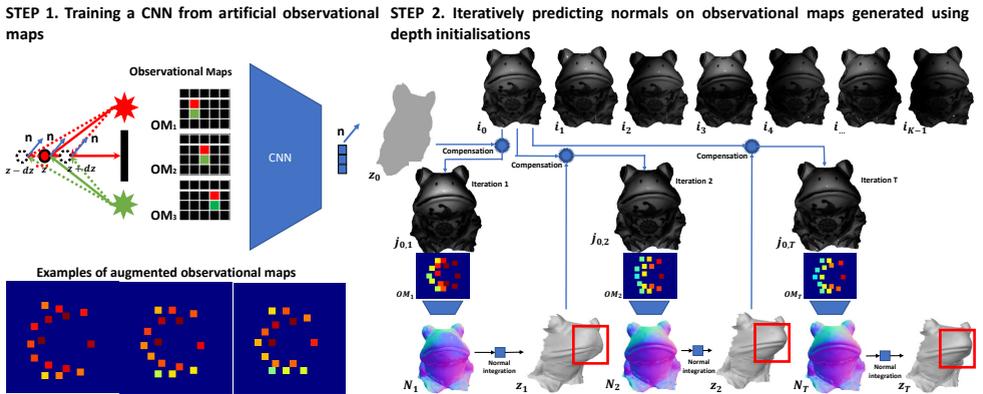


Figure 2: This figure illustrates two key steps of our proposed approach. On the left, the network training is illustrated consisting of sampling points inside the camera’s frustum and rendering the respective observation maps. As the depth will only be approximately known at test time, this is slightly perturbed before mapping resulting to a structured change of the map (shown at bottom left: middle is actual depth (10 cm), left and right correspond to ± 1 cm respectively). On the right, the reconstruction process is shown. Images $i_0 \dots i_{k-1}$ are used with conjunction with previous depth estimate to compensate for light attenuation ($j_0 \dots j_{k-1}$), compute observation maps (shown for pixel at image center here), regress normals and finally update the shape. Note the improvement of the shape of frogs beak (red square) from iteration 1 to iteration 2.

lated observation rendered by the MERL BRDF dataset [19] and the normal map of the target object, handling non-local effects using a dropout strategy. [20] leveraged DL to learn the information from multispectral images to get RGB information. [21] proposed generating training data on the go to minimise the image re-projection error. Although this method is a training data free approach, the whole procedure is slow. Recently, [22] proposed rendering patches of different surface materials in order to get training data. This method is also extended in [6] for solving the uncalibrated PS. Finally, [8] proposed arranging all the reflectance samples of a pixel (i.e. different illumination images in the far-field setting) into a fixed size *observation* map which is used by a CNN to regress pixel normals. The CNN is essentially learning to invert the BRDF with added robustness to global illumination effects, as training data are made with physics based renderer. In [18], this method was extended by simplifying the training procedure. However, none of these DL approaches is directly applicable to the NFPS problem and non-linear attenuation from point light sources together with the viewing direction dependency drastically increase the problem space exploding the training data requirements. To address this issue, the NF problem is re-arranged appropriately into a far-field one allowing the use of the observation maps. Finally, to maximise performance, only the appropriate subset of maps is sampled at train time.

3 Method

In this section we describe our method for tackling the near field Photometric Stereo problem. In particular, we provide both the details of the assumed image formation model and how normals can be predicted for near-field images by using CNN’s trained on far-field reflectance samples (also see Figure 2).

Near field Modeling. Similar to [20], we assume calibrated point light sources at positions \mathbf{P}_m (w.r.t the camera center at $\mathbf{0}$) resulting in variable lighting vectors $\mathbf{L}_m = \mathbf{P}_m - \mathbf{X}$. Here $\mathbf{X} = [x, y, z]^\top$ is the 3D surface point coordinates. We also model the light attenuation considering the following non-linear radial model of dissipation

$$a_m(\mathbf{X}) = \phi_m \frac{(\hat{\mathbf{L}}_m(\mathbf{X}) \cdot \hat{\mathbf{S}}_m)^{\mu_m}}{\|\mathbf{L}_m(\mathbf{X})\|^2}, \quad (1)$$

where $\hat{\mathbf{L}}_m = \frac{\mathbf{L}_m}{\|\mathbf{L}_m\|}$ is the lighting direction, ϕ_m is the intrinsic brightness of the light source, $\hat{\mathbf{S}}_m$ is the principal direction (i.e. the orientation of the LED point light source) and μ_m is an angular dissipation factor. Defining $\hat{\mathbf{V}} = -\frac{\mathbf{X}}{\|\mathbf{X}\|}$ as the viewing vector, the general image irradiance equation becomes:

$$i_m = a_m \mathbf{B}(\mathbf{N}, \hat{\mathbf{L}}_m, \hat{\mathbf{V}}, \rho) \quad (2)$$

Here \mathbf{N} is the surface normal. \mathbf{B} is assumed to be a general BRDF and ρ is the surface albedo (allowing for the most general case, images and ρ are RGB and the reflectance is different per channel). In addition, we allow for the possibility of global illumination effects (shadows, self reflections) which are incorporated into \mathbf{B} . This can be re-arranged into a BRDF inversion problem as:

$$j_m = \frac{i_m}{a_m} = \mathbf{B}(\mathbf{N}, \hat{\mathbf{L}}_m, \hat{\mathbf{V}}, \rho). \quad (3)$$

We note that $\hat{\mathbf{V}}$ is known but \mathbf{L}_m and a_m are unknowns due to the nonlinear dependence on z . Our objective is to recover the surface normals \mathbf{N} and depth z .

Normal Prediction. The first step of our method includes training a CNN for per-pixel normal prediction using BRDF samples. This is done through the the observational map parameterisation introduced by [8] in order to tackle the far-field photometric stereo problem. Note that this is equivalent to BRDF inversion under the special case of $\hat{\mathbf{V}} = [0, 0, 1]$. As described in [8] an observational map records relative pixel intensities (BRDF samples) on a 2D grid (e.g. 32×32) of discretised light directions. Such a representation is highly convenient for use with classical CNN architectures as it provides a 2D input and is of fixed shape despite a potentially varying number of lights. While [8] proposes to train CNNs on rendered images of objects, it is shown in [18] that simpler per-pixel renderers can be used instead, making the training procedure much faster and simpler. We use the latter approach in this work. In addition, the known viewing direction can be incorporated by augmenting the map with another two channels, set to $\hat{\mathbf{V}}_x$ and $\hat{\mathbf{V}}_y$ respectively.

Adapting to the near-field setup. In order to solve the near-field PS problem for a specific capture setup, we adapt the training procedure to only sample observation maps which are plausible at test time. We note that our capture hardware has a fixed number of 15 lights rigidly attached to a circuit board around the camera looking at

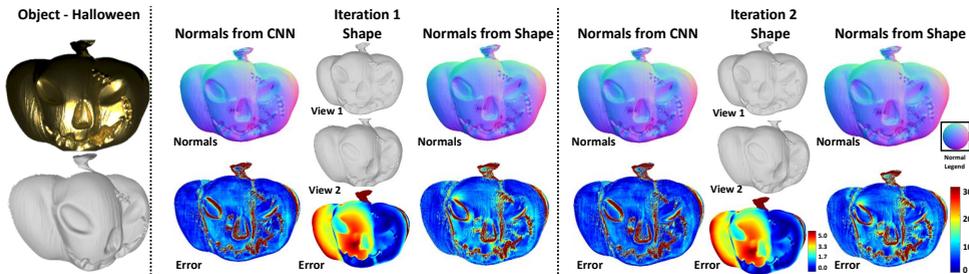


Figure 3: Iterative refinement of the geometry for the Halloween synthetic object. On the left, a sample image and GT shape are shown. The other 2 sections show 2 steps of the iterative refinement with the respective normals (both raw network predictions and differentiated ones), normal error maps and depth error maps. As the difference is minimal between steps 1 and 2, the process is converged.

objects at a limited depth range of around $10\text{-}20\text{ cm}^2$. Therefore, instead of sampling a random set of light directions as in [18], we sample 3D points inside the camera frustum. Surface normal and material parameters are sampled independently from the point position so as to be able to deal with general surfaces. After the point parameters are sampled, the 15 images are rendered using the renderer of [18] with the addition of near-field light attenuation (Eq. 1). Additional global illumination augmentations for shadows, reflections and ambient light are also applied as in [18]. In order to get robustness to imprecise depth initialisation at test time, the training procedure involves perturbing the ground truth depth value z by $\delta z \sim \mathcal{N}(0, 4\text{mm})^3$ so that the light directions $\hat{\mathbf{L}}$ (determining the sparsity pattern of the map) and attenuation (determining the values-Eq. 3) are slightly perturbed. See Step 1 in Figure 2 for an illustration.

Iterative refinement of depth and normals. Assuming an estimate of normals, the depth can be estimated with numerical integration. This is performed using the ℓ_1 method of [24]. The variational optimisation includes a Tikhonov regulariser $z = z_0$ (weight $\lambda = 10^{-6}$) and is solved in a ADMM scheme⁴. As the BRDF samples j (see Equation 3) depend on the unknown depth, they cannot be directly computed to be input to the network. To overcome this issue, we employ an iterative scheme where the previous estimate of the geometry is used. The procedure involves computing the near to far conversion as described, obtaining a new normal map estimate through the CNN and finally numerical integration. See Figure 3 for an example of intermediate results of our iterative procedure. As it is the case in competing classical methods [17, 25], this iterative procedure is initialised with a flat plane at the approximate mean distance.

²When objects are too close to the camera, their images are out of focus; while when being too far, surface details cannot be resolved.

³The Gaussian distribution encourages the network to be more accurate when δz is small and thus get an improvement in the iterative setting.

⁴Code ported from https://github.com/yqueau/normal_integration.

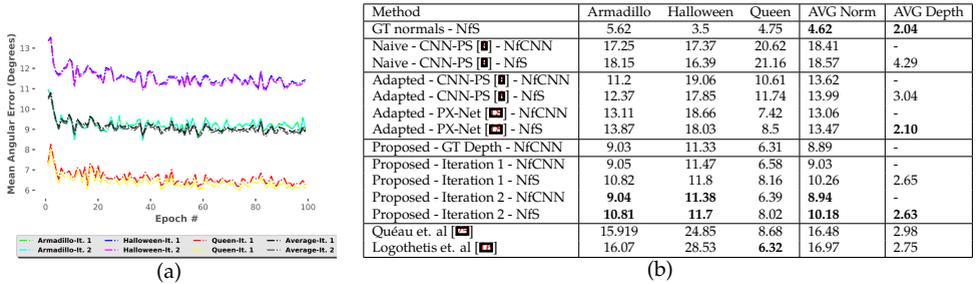


Figure 4: (a) MAE evolution (during training) curves illustrate the performance of our network (PX-NET) on predicting normals (NfCNN) at first and second iteration . (b) Full quantitative comparison on synthetic data. For our method, we report raw normal prediction error as NfCNN and numerically differentiated normals as NfS. Also report NfCNN error when the GT depth is used as initialisation and also NfS error for integrating the GT normals. We compare against state of the art NF methods [15], [16] and naive usage of far field method of [9].

4 Experimental Setup

In this section we provide various experimental setup details related to CNN training and datasets used for evaluation.

CNN training. We use the exact architecture of PX-NET [15] which is heavily inspired by CNN-PS [9]. More specifically, the network includes 7 convolutional, 4 dropout (20%) and 3 fully connected layers as well as a special log layer. See [15] for more detail. The total number of parameters is ~ 4.5 million. The network was implemented in Keras of Tensorow 2.0 The training batch size was set at 1600 with 10000 batches per epoch (16 million maps). We trained for 100 epochs which took around one day on 3x NVIDIA GeForce 1080. We trained the network using the mean squared error loss function on the normals. The Adam optimiser with basic LR of 10^{-3} and LR decay of 0.1% per epoch after epoch 10 was used. The model test performance evolution over time (epochs) can be seen in Figure 4(a). The reconstruction computation time was 2-3 minutes with the bottleneck being the python implementation of the numerical integration.

Datasets. We evaluate our method on real data captured with a custom made setup. This consists of a printed circuit board with 15 white bright LEDs rigidly attached to a camera FLEA3 3.2 MegaPixel with a 8mm lens. The LEDs are evenly spaced around the camera at a maximum distance of 6.5cm and are placed to be coplanar with the image plane. We captured four image sequences namely a metallic-silver Bulldog statue (Figure 1), a metallic-gold Bell, a porcelain Frog as well as a mutli-object scene featuring a shiny wooden elephant statue in front of a porcelain Squirrel (Figure 6). The objects are placed around 15cm away from the camera and the initial depth estimate is approximately measured with a ruler. As there is no ground truth available for these objects, the evaluation on this dataset is only qualitative. Quantitative evaluation is performed using a synthetic dataset rendered with Blender. This artificial light source configuration closely mimics our real capture setup. The Cycles render engine with the Disney BRDF is used to generate realistic

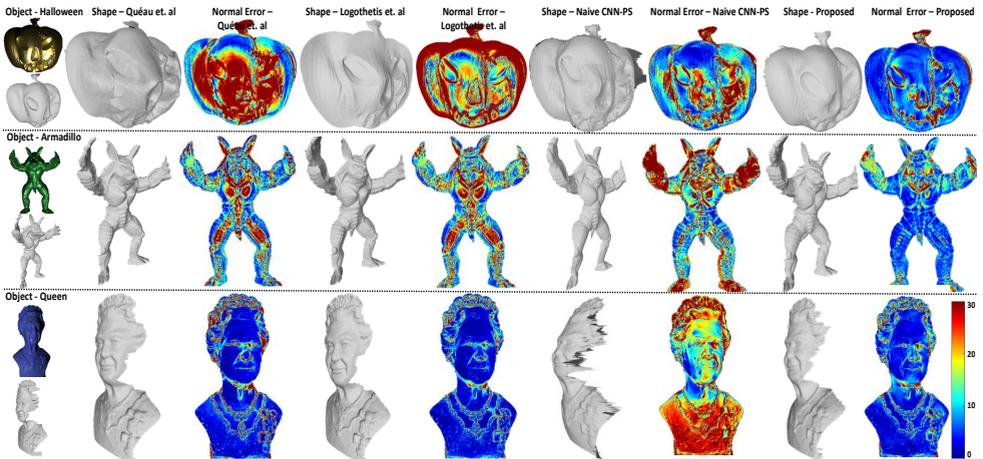


Figure 5: Visualisation of the results of Figure 4(b). Optimisation based methods of Logothetis et. al [16] and Quéau et. al [23] fail on the reflectance of the metallic Halloween and so they exhibit very high MAE. In contrast, the dielectric specular Queen is more consistent with their assumed reflectance model and thus exhibit smaller MAE. The naive far-field method CNN-PS [8] also has significant global deformation as it does not model the light attenuation effect. The proposed approach is able to cope with all, highly different, materials and achieves the state-of-the-art performance.

global illumination effects such as cast shadows and self reflections. We present three objects here namely Queen, which is has a dielectric specular material, Halloween is purely metallic and Armadillo which has an intermediate material. The materials are chosen to match the far-field dataset of [18].

Evaluation Protocol. We compare our method against the far-field CNN approaches of [8] and [18], and the near field methods of [16]⁵ and [23]⁶ (disabling the light source calibration parameter). For the synthetic experiments, the evaluation metric is the mean angular error (MAE) of the normal map (degrees) as well as the average depth error (mm). We note that the depth error is less meaningful as all these methods assume a continuous surfaces and thus a systematic error is introduced in the occlusion boundaries.

5 Experiments

In this section we present various experiments on the synthetic and real datasets introduced in the previews section. These include an ablation study of the numerical integration method, a naive usage of far-field state-of-the-art networks as well as qualitative and quantitative comparison with NF competitors.

⁵Code from https://github.com/fotlogo/semi_calib_ps_cvpr2017

⁶Code from https://github.com/yqueau/near_ps

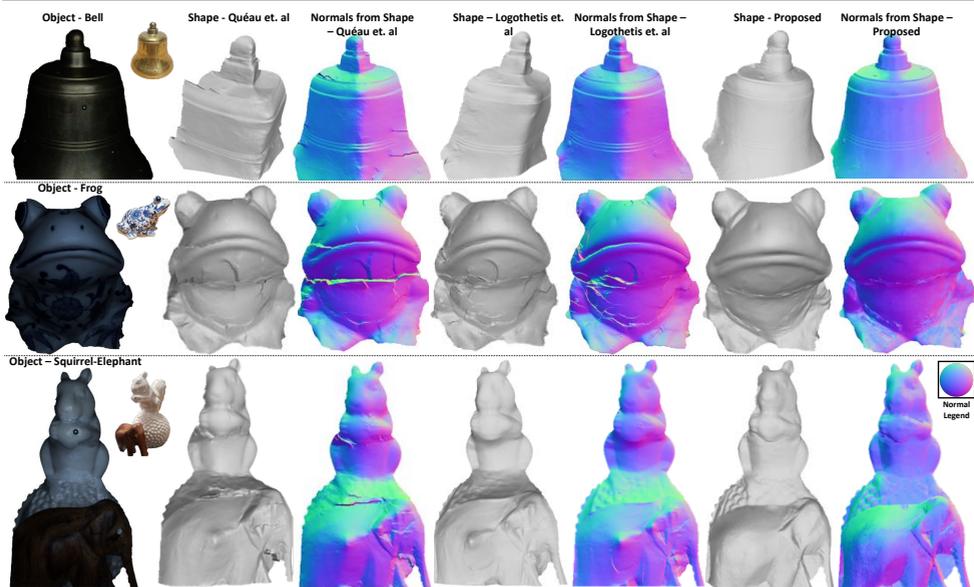


Figure 6: Qualitative comparison of the proposed method to Quéau et. al [25] and Logothetis et. al [16]. The first column shows the average Photometric Stereo image. In contrast to competition, the proposed approach has no visible deformation on the metallic object or the specular highlight in the middle of the elephant. In addition, there is a smooth recovery of belly of the Frog despite the shadows, as well as the bottom of Squirrel despite self reflection.

Shape integration. The first experiment we conducted aimed at calibrating the quality of the numerical integration of the normal map. As no realistic depth map is C^2 continuous, GT normals are not compatible with the GT depth. Indeed, integrating the GT normals and then re-calculating them with numerical differentiation introduces 4.62° (Figure 4(b) top) MAE on average which is by no means negligible. Therefore, for the rest of the experiments we present two figures namely MAE on the raw network predictions (NfCNN) and MAE after differentiation of the surface (NfS). Since the network is trained to regress normals, the first figure tends to be slightly lower.

Naive usage of far-field networks. The next experiment consists of naively using the far-field methods [9] [13] with near field images without any attenuation compensation and by using the average light direction for each LED. The predicted normals are also integrated using our method in order to have a qualitative shape comparison (Figure 5). As expected, the networks are able to cope with the complicated reflectance and global illumination effects (no bumps at specular highlights); however, not accounting for the variable lighting direction and attenuation results to severe distortion, which is also depicted in larger MAE. To add to this point, using these network as part of our iterative process drastically improves the result and in fact they outperform the classical optimisation methods of [16] and [25]. See Table 4(b).

Near field competitors. Then, we present comparison against the state of the art of

NF methods and [16] [25]. They both consist of complex variational optimisations which are assuming simplified reflectance models (mostly diffuse for [25] using a robust Cauchy estimator) and Blinn-Phong specular for [16] using an ℓ_1 loss function. As expected, they are both severely outperformed on the Halloween and Armadillo objects and compare favorably on the dielectric specular Queen. In fact, [16] achieve the minimum MAE for the Queen which is not surprising as this material is mostly consistent with the assumed Blinn-Phong model.

Real Data evaluation. Finally, we present qualitative comparison with [16] and [25] on our challenging real data in Figure 1 and 6. As expected, we outperform the competition and our reconstructions do not exhibit any visible deformation in specular highlights (middle of bell and elephant) or cast shadow regions (belly of frog and bottom of squirrel). The superiority comes from the fact that the network was trained using realistic BRDF samples (Disney and MERL) and a data augmentation strategy was also used to account for shadows and self reflections.

6 Conclusion

In this work we presented the first CNN approach tackling the near-field Photometric Stereo problem. We leveraged the capability of CNNs to learn to predict surface normals from reflectance samples for a wide variety of materials. In order to deal with global effect such as shadows and interreflection, we used an augmentation strategy which allowed our method to handle these physical effects. Finally, to address near-field effects such as radial light propagation of point light sources, light attenuation and viewing dependency of the light reflection, we iteratively integrate the normal map so to let the problem depend on the depth. The main limitation of the proposed approach is the lack of any explicit surface optimisation hence the accuracy is limited by the network predictions. An interesting future work would be to avoid the integration step and let the network directly predict surface depth. Additionally, a future direction of research would be to relax the assumption of training on specific light positions and employ weaker light calibration setting.

References

- [1] J. Ackermann and M. Goesele. A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision*, 2015.
- [2] J. F. Blinn. Models of light reflection for computer synthesized pictures. In *SIGGRAPH*, 1977.
- [3] B. Burley. Physically-based shading at disney.
- [4] G. Chen, K. Han, and K.-Y. K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *ECCV*, 2018.
- [5] G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong. Self-calibrating deep photometric stereo networks. In *CVPR*, 2019.
- [6] J. J. Clark. Active photometric stereo. In *CVPR*, 1992.

-
- [7] G. E. Hinton. Deep belief networks. *Scholarpedia*, 2009.
- [8] S. Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *ECCV*, 2018.
- [9] S. Ikehata and K. Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *CVPR*, 2014.
- [10] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa. Robust photometric stereo using sparse regression. In *CVPR*, 2012.
- [11] Y. Iwahori, H. Sugie, and N. Ishii. Reconstructing shape from shading images under point light source illumination. In *ICPR*, 1990.
- [12] Y. Ju, L. Qi, H. Zhou, J. Dong, and L. Lu. Demultiplexing colored images for multispectral photometric stereo via deep neural networks. *IEEE Access*, 2018.
- [13] S. Lee and M. Brady. Integrating stereo and photometric stereo to monitor the development of glaucoma. *Image and Vision Computing*, 1991.
- [14] C. Liu, S. G Narasimhan, and A. W. Dubrawski. Near-light photometric stereo using circularly placed point light sources. In *ICCV*, 2018.
- [15] F. Logothetis, R. Mecca, Y. Quéau, and R. Cipolla. Near-field photometric stereo in ambient light. In *BMVC*, 2016.
- [16] F. Logothetis, R Mecca, and R. Cipolla. Semi-calibrated near field photometric stereo. In *CVPR*, 2017.
- [17] F. Logothetis, R. Mecca, and R. Cipolla. A differential volumetric approach to multi-view photometric stereo. In *ICCV*, 2019.
- [18] F. Logothetis, I. Budvytis, R. Mecca, and R. Cipolla. PX-NET: Simple, Efficient Pixel-Wise Training of Photometric Stereo Networks. In *arXiv preprint arXiv:2008.04933*, 2020.
- [19] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 2003.
- [20] R. Mecca, A. Wetzler, A. Bruckstein, and R. Kimmel. Near Field Photometric Stereo with Point Light Sources. *SIAM Journal on Imaging Sciences*, 2014.
- [21] R. Mecca, Y. Quéau, F. Logothetis, and R. Cipolla. A single lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 2016.
- [22] R. Onn and A. Bruckstein. Integrability disambiguates surface recovery in two-image photometric stereo. *IJCV*, 1990.
- [23] E. Prados and O. Faugeras. Perspective shape from shading and viscosity solutions. In *ICCV*, 2003.
- [24] Y. Quéau and J.-D. Durou. Edge-preserving integration of a normal field: Weighted least squares, TV and L1 approaches. In *SSVM*, 2015.

- [25] Y. Quéau, B. Durix, Tao Wu, D. Cremers, F. Lauze, and J.-D. Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *JMIV*, 2018.
- [26] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita. Deep photometric stereo network. In *ICCV Workshops*, 2017.
- [27] W.A.P. Smith and Fufu F. Height from photometric ratio with model-based light source selection. *CVIU*, 2016.
- [28] Y. Tang, R. Salakhutdinov, and G. E. Hinton. Deep lambertian networks. In *ICML*, 2012.
- [29] T. Taniai and T. Maehara. Neural inverse rendering for general reflectance photometric stereo. In *ICML*, 2018.
- [30] A. Tankus and N. Kiryati. Photometric stereo under perspective projection. In *ICCV*, 2005.
- [31] A. Wetzler, R. Mecca, A. M. Bruckstein, and R. Kimmel. Close-range photometric stereo with point light sources. In *3DV*, 2014.
- [32] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980.
- [33] Y. Yu and W. A. P. Smith. Pvnnet: A neural network library for photometric vision. In *ICCV Workshop*, 2017.
- [34] A. L. Yuille, D. Snow, R. Epstein, and P. N. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability. *IJCV*, 1999.