NEURAL BRDF PLUGIN FOR UNSUPERVISED PHOTOMETRIC STEREO

Heng Guo[†]   *Boxin Shi*[‡]   *Yasuyuki Matsushita*[†]

[†]Osaka University   [‡]Peking University

ABSTRACT

Recovering surface normal under unknown general reflectance is challenging for photometric stereo. Supervised learning-based methods implicitly handle general reflectances by learning from training data. However, it suffers from a demand for a large number of rendered images as training data and also from a domain gap due to the biased shape distribution in the training set. Unsupervised methods mitigate these issues by assessing image re-rendering losses, but their performance is limited primarily due to the fact that multiple combinations of reflectance parameters can explain the same appearance. We combine the merits of these two approaches by introducing a *BRDF plugin*, a neural network-based BRDF dictionary that is trained using a measured BRDF dataset, offering a data prior to effectively constrain reflectances. Specifically, it maps each measured BRDF to a unique latent code, forming a latent BRDF space that serves as the data prior for reflectances. Building upon the BRDF plugin, we develop a new unsupervised photometric stereo method to recover both surface normal and reflectance, which exhibits superior accuracy than previous methods in experiments on synthetic and real-world datasets.

Key Words— Photometric stereo; Neural BRDF; Unsupervised learning

1. INTRODUCTION

Photometric stereo aims at recovering surface normal and reflectance from a set of images captured with a fixed camera under varying illuminations. One of the key challenges in photometric stereo is to handle general surface reflectances, which are described via bidirectional reflectance distribution functions (BRDFs). Early photometric stereo methods assume an ideal Lambertian model [1], which is unfortunately not flexible enough to represent real-world reflectances.

To deal with general BRDFs, *supervised* learning-based photometric stereo methods, such as DPSN [2] have been developed, which implicitly address general BRDFs by learning data prior from synthetic image datasets rendered with diverse shapes and BRDFs. Subsequent works use altered network structures, including all-pixel based branches (*e.g.*, PS-FCN [3], NormAttentionPSN [4]), per-pixel based branches (*e.g.*, CNN-PS [5], PX-Net [6]), and hybrid approaches (*e.g.*,

GPS-Net [7]). More comprehensive reviews are found in recent surveys [8, 9] on learning-based photometric stereo. Supervised photometric stereo methods require extensive training datasets, which can lead to significant time and memory costs, potential domain gaps, and biases in real-world applications due to specific shape and BRDF distributions. They also limit methods' applicability to specific lighting conditions, *e.g.*, PS-FCN [3] trained on distant-light datasets cannot deal with near-light settings.

To bypass these training dataset-related issues in supervised photometric stereo, *unsupervised* photometric stereo methods based on image re-rendering losses have been introduced [10, 11], which exhibit flexibility in dealing with diverse shapes and lighting conditions. However, current unsupervised photometric stereo methods could suffer from the fact that multiple combinations of parameters can explain the same appearance, as they estimate surface reflectances purely from image cues without the additional constraints offered by training data that are used in supervised methods. Consequently, the estimated surface normals may not align well with the ground-truth (GT) even when low image re-rendering errors are achieved.

We address the above problem by bringing the data prior of supervised photometric stereo to unsupervised methods. We observe that the general BRDFs in photometric stereo are the primary component requiring the data prior for accurate modeling. Motivated by the observation, we propose a *BRDF plugin*, a neural network-based BRDF dictionary trained on real-world measured BRDF datasets, such as MERL [12]. In our BRDF plugin, each measured BRDF is mapped to a unique high-dimensional latent code, creating a latent BRDF space. Once trained, our BRDF plugin can be applied to various tasks associated with general BRDF modeling, including unsupervised photometric stereo.

To verify the effectiveness of our BRDF plugin, we develop an unsupervised photometric stereo method built upon our BRDF plugin to predict surface normal and reflectance simultaneously, in which the unknown BRDF of the target surfaces is assumed to reside within the latent BRDF space. In other words, the latent BRDF code of the reflectance, when mapped via our BRDF plugin, can be described by a linear combination of the latent codes of the measured BRDFs. This strategy allows the data prior by the BRDF plugin to enhance unsupervised photometric stereo by providing additional con-

straints on reflectance estimation. In addition, as our BRDF plugin is supervised only on BRDFs but not shapes, we eliminate the need for time and memory-intensive rendering processes associated with building large-scale image datasets and avoid shape-related biases. By evaluating our unsupervised photometric stereo method on both synthetic and real-world datasets, we showcase the superior performance of our approach over previous methods.

2. PROPOSED METHOD

We begin with modeling general BRDFs with the proposed BRDF plugin and then introduce our unsupervised photometric stereo method powered by the BRDF plugin.

2.1. BRDF plugin

A BRDF R is a scalar function of surface normal $\mathbf{n}$ and incident-outgoing light directions $(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$, *i.e.*, $R(\mathbf{n}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$. As shown in Fig. 1, an isotropic BRDF can be represented in a more compact manner using the Rusinkiewicz coordinate [13] $\mathbf{x} = [\theta_h, \theta_d, \phi_d]^\top$, where $\theta_h \in [0°, 90°]$ is the angle between surface normal and a halfway vector $\mathbf{h} = (\boldsymbol{\omega}_i + \boldsymbol{\omega}_o)/2$, $\theta_d \in [0°, 90°]$ and $\phi_d \in [0°, 180°]$ are the spherical coordinates of incident light direction $\boldsymbol{\omega}_i$ in the coordinate system where $\mathbf{h}$ is at the north pole. Based on Rusinkiewicz coordinate, we represent BRDF as $R(\mathbf{x}) = R(\theta_h, \theta_d, \phi_d)$. There are two challenges in compactly representing BRDFs: 1) BRDFs greatly vary across different materials; 2) the domain of BRDF functions is continuous. Therefore, we use a conditional coordinated-based multilayer perceptron (MLP) for enabling the BRDF plugin.

As shown in Fig. 1, our BRDF plugin consists of three modules: a projection network, a mapping network, and a sampling network, with their network parameters labeled as $\boldsymbol{\theta}_p$, $\boldsymbol{\theta}_m$, and $\boldsymbol{\theta}_s$, respectively. The projection network f_p maps a BRDF to its latent code $\mathbf{z}$ serving as an identifier for diverse BRDFs, *i.e.*,

$$\mathbf{z} = f_p(\mathbf{R}; \boldsymbol{\theta}_p), \tag{1}$$

where $\mathbf{R} \in \mathbb{R}^{90 \times 180 \times 180}$ is a 3D measured BRDF tensor from MERL dataset [12], indexed by the three variables of the Rusinkiewicz coordinates. Thus, we use a 3D CNN [14] in the projection network.

The mapping network and sampling network in our BRDF plugin produce a BRDF value r sampled at a specific Rusinkiewicz coordinate $\mathbf{x}$ controlled by a material latent code $\mathbf{z}$. To achieve this, we design a conditional coordinate-based MLP following π-GAN [15]. Specifically, the mapping network f_m takes latent code $\mathbf{z}$ as input and outputs latent feature $\mathbf{t}$ for the sampling network, *i.e.*,

$$\mathbf{t} = [\boldsymbol{\gamma}, \boldsymbol{\beta}] = f_m(\mathbf{z}; \boldsymbol{\theta}_m), \tag{2}$$
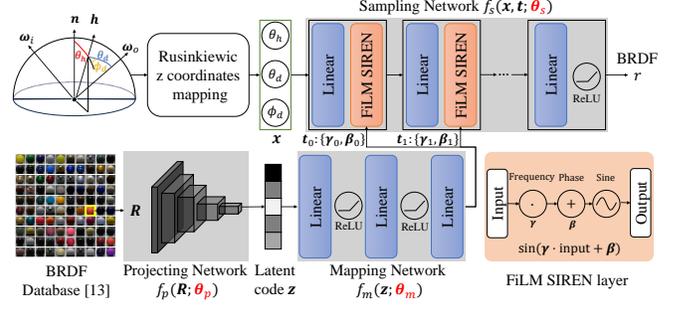
Fig. 1: Network structure of our BRDF plugin.

where $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are frequency and phase terms in the feature-wise linear modulation (FiLM) layer [15], which is embedded in the sampling network for controlling the BRDF output for a specific material. The sampling network adopts the SIREN network structure [16] with a periodic sine activation function. Given latent feature $\mathbf{t}$, the sampling network f_s outputs a BRDF value r at a sampled Rusinkiewicz coordinate $\mathbf{x}$ as

$$r = f_s(\mathbf{x}, \mathbf{t}; \boldsymbol{\theta}_s). \tag{3}$$

Combining Eqs. (1) to (3), we optimize the network parameters of the BRDF plugin by minimizing L1 loss between predicted and measured BRDF values of MERL dataset [12], which contains $k = 100$ diverse BRDF tensors and $p = 1458000$ Rusinkiewicz coordinate samples, *i.e.*,

$$\boldsymbol{\theta}_p^*, \boldsymbol{\theta}_m^*, \boldsymbol{\theta}_s^* = \underset{\boldsymbol{\theta}_p, \boldsymbol{\theta}_m, \boldsymbol{\theta}_s}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{j=1}^{p} \| r_{ij} - f_s(\mathbf{x}_j, \mathbf{t}_i; \boldsymbol{\theta}_s) \|_1$$

$$= \underset{\boldsymbol{\theta}_p, \boldsymbol{\theta}_m, \boldsymbol{\theta}_s}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{j=1}^{p} \| r_{ij} - f_s\left(\mathbf{x}_j, f_m\left(f_p\left(\mathbf{R}_i; \boldsymbol{\theta}_p\right); \boldsymbol{\theta}_m\right); \boldsymbol{\theta}_s\right) \|_1. \tag{4}$$
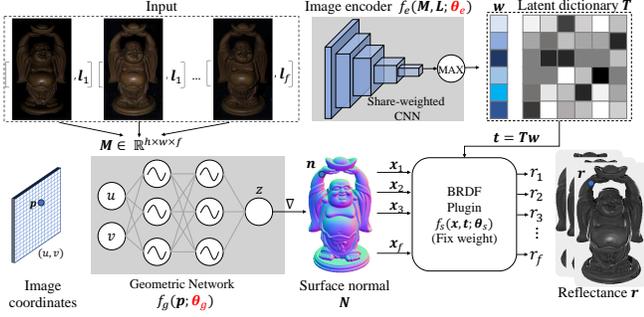
where r_{ij} is the BRDF value of the i-th MERL BRDF tensor $\mathbf{R}_i$ sampled at the j-th coordinate $\mathbf{x}_j$.

To better fit specular highlights, we compute the loss in the logarithmic domain using μ-law [1], where BRDF values are mapped as $\hat{r}_{ij} = \log(1 + \mu r_{ij})/\log(1 + \mu)$. After training, we save the optimized network parameters. The k latent codes $\mathbf{z}$ along with their corresponding latent features $\mathbf{t}$ trained from MERL dataset [12] construct a latent code dictionary $\mathbf{Z} \in \mathbb{R}^{k \times d_z}$ and a latent feature dictionary $\mathbf{T} \in \mathbb{R}^{k \times d_t}$, where t_d and t_z are the dimensions of the latent code and feature, respectively. The BRDF plugin network and the latent dictionaries build a data prior from real-world measured BRDFs, which can be easily embedded into our unsupervised photometric stereo method.

2.2. Unsupervised photometric stereo with BRDF plugin

As shown in Fig. 2, our unsupervised photometric stereo takes image observations and the corresponding light directions as

---

[1]μ-law: https://www.cisco.com/c/en/us/support/docs/voice/h323/8123-waveform-coding.html

Fig. 2: Network structure of unsupervised photometric stereo.

Method	LL22 [11]	PS-FCN [3]	Ours
Avg. MAE	5.18	4.55	**2.41**

Fig. 3: Evaluation on the *MERLSphere* dataset [3] measured by MAE in degree, with material names labeled in x-axis.

input and produces surface normal map $\mathbf{N}$ and per-pixel reflectance vector $\mathbf{r}$. These outputs allow us to reconstruct the image observations and calculate an image re-rendering loss with the input images. Specifically, our unsupervised photometric stereo contains three modules: a geometric network, an image encoder, and the embedded BRDF plugin.

We follow the neural surface model proposed in [17] to represent scene geometry, where depth is represented as a neural function of image coordinates, and surface normal can be obtained simultaneously based on the analytical derivatives of the neural depth function *w.r.t* coordinate $\mathbf{p}$, *i.e*,

$$z_{\mathbf{p}} = f_g(\mathbf{p}; \boldsymbol{\theta}_g), \qquad (5)$$

$$\mathbf{n}_{\mathbf{p}} = \epsilon([\nabla z, -1])^{\top} = \epsilon([\nabla f_g(\mathbf{p}; \boldsymbol{\theta}_g), -1])^{\top}, \qquad (6)$$

where $\epsilon(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ denotes unit normalization. For a scene point at $\mathbf{p}$, given its surface normal $\mathbf{n}_{\mathbf{p}}$ and f input light directions, we compute its Rusinkiewicz coordinates $\mathbf{X}_{\mathbf{p}} = \{\mathbf{x}_1, \ldots, \mathbf{x}_f\}$ for the f incident-outgoing light ray pairs. These coordinates are fed into the sampling network of our BRDF plugin to predict the corresponding reflectance $\mathbf{r}_{\mathbf{p}}$.
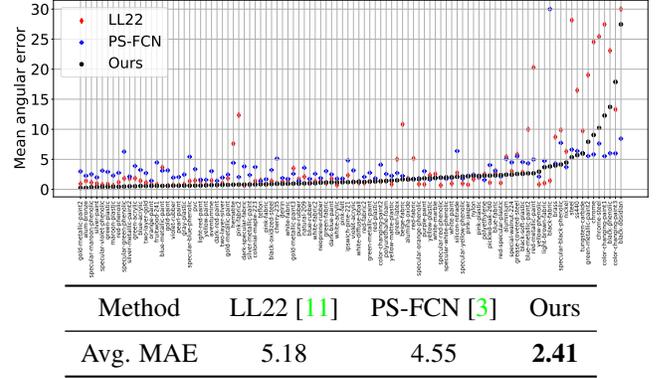
To import data prior from real-world measured BRDFs, we fix the weight of the sampling network $f_s(\cdot)$ in our BRDF plugin and assume the latent BRDF code of the target object can be linearly represented by our latent dictionaries $\mathbf{Z}$ or $\mathbf{T}$. Following the practice of π-GAN [15], we optimize the latent BRDF of the target object in the feature space (using latent feature dictionary $\mathbf{T}$ instead of latent code dictionary $\mathbf{Z}$) to avoid the complex mapping network $f_m(\cdot)$, *i.e.*,

$$\mathbf{t} = \mathbf{Tw}, \qquad (7)$$

where $\mathbf{w}$ represents a latent weight vector that remains uniform across all scene points. As $\mathbf{w}$ is reflectance dependent, we design an image encoder following the shared-weighted structure of PS-FCN [3] to predict it from image observations and the corresponding light directions, *i.e.*,

$$\mathbf{w} = f_e(\mathbf{M}, \mathbf{L}; \boldsymbol{\theta}_e), \qquad (8)$$

where $\mathbf{M} \in \mathbb{R}^{h \times w \times f}$ contains f input images whose resolution is $h \times w$, $\mathbf{L} \in \mathbb{R}^{f \times 3}$ stacks the f light directions. Com-

bining Eqs. (1), (7) and (8), we predict the reflectance

$$\mathbf{r}_{\mathbf{p}} = f_s(\mathbf{X}_{\mathbf{p}}, \mathbf{t}; \boldsymbol{\theta}_s) = f_s(\mathbf{X}_{\mathbf{p}}, \mathbf{T} f_e(\mathbf{M}, \mathbf{L}; \boldsymbol{\theta}_e); \boldsymbol{\theta}_s). \qquad (9)$$

Together with the surface normal from the geometric network, we can re-render the image observations at $\mathbf{p}$ as $\hat{\mathbf{m}}_{\mathbf{p}}$, *i.e.*,

$$\begin{aligned}
\hat{\mathbf{m}}_{\mathbf{p}} &= \mathbf{r}_{\mathbf{p}} \odot \max(\mathbf{L} \mathbf{n}_{\mathbf{p}}, \mathbf{0}), \\
&= f_s(\mathbf{X}_{\mathbf{p}}, \mathbf{T} f_e(\mathbf{M}, \mathbf{L}; \boldsymbol{\theta}_e); \boldsymbol{\theta}_s) \odot \max(\mathbf{L} \epsilon([\nabla f_g(\mathbf{p}; \boldsymbol{\theta}_g), -1])^{\top}, \mathbf{0}),
\end{aligned} \qquad (10)$$

where $\odot$ denotes element-wise product. The unknown network parameters $\boldsymbol{\theta}_e$ and $\boldsymbol{\theta}_g$ are optimized by minimizing the L1 loss between observed and re-rendered images $\mathbf{m}$ and $\hat{\mathbf{m}}$:

$$\boldsymbol{\theta}_e^*, \boldsymbol{\theta}_g^* = \operatorname*{argmin}_{\boldsymbol{\theta}_e, \boldsymbol{\theta}_g} \sum_{\mathbf{p}} \|\mathbf{m}_{\mathbf{p}} - \hat{\mathbf{m}}_{\mathbf{p}}\|_1. \qquad (11)$$

With optimized network parameters, we can recover surface normal and depth map based on Eq. (6), and surface reflectance based on Eq. (9).

3. EXPERIMENT

We assess our BRDF plugin-embedded unsupervised photometric stereo method using synthetic and real-world datasets. We select two notable supervised photometric stereo methods, PS-FCN [3] and CNN-PS [5], and a cutting-edge unsupervised method, LL22 [11], as baselines. We evaluate their surface normal recoveries using mean angular error (MAE) between the ground truth and the estimated surface normals.

3.1. Synthetic experiment

As shown in Fig. 3, we evaluate our methods and baselines on a synthetic photometric stereo dataset *MERLSphere* [3], which includes 100 spheres covered by MERL BRDFs [12]. Benefited from learned data prior in our BRDF plugin, our surface normal estimates achieve less than half of the MAE from LL22 [11] on the average of 100 scenes, especially on

CNN-PS (16.66/14.74)   PS-FCN (16.91/15.91)

TM18 (17.93/14.07)   Ours (15.70/13.00)
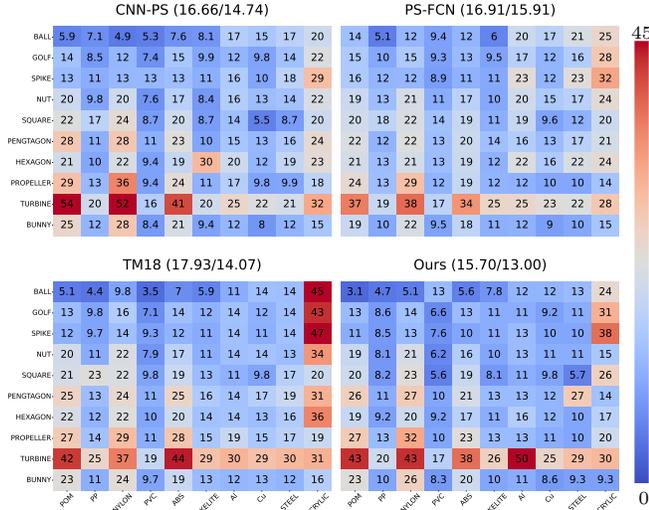
Fig. 4: Evaluation of surface normal by MAE on DiLiGenT10^2 [19] indexed by shape and material, with (mean/median) MAE over 100 scenes shown on the top.

metallic paints and metals such as 'red-metallic-paint' and 'steel'. Our method also outperforms the supervised photometric stereo method PS-FCN [3]. One possible reason is PS-FCN [3] is influenced by the shape distribution in its photometric stereo training image dataset, while our method avoids such a problem by only training on a measured BRDF dataset.

3.2. Real-world experiment

Besides the synthetic evaluations, we also test our method on real-world datasets DiLiGenT [18] and DiLiGenT10^2 [19]. **DiLiGenT**10^2. As shown in Fig. 4, we show surface normal estimation errors from existing methods and ours indexed by the shapes and materials in DiLiGenT10^2 [19], where CNN-PS [5] and TM18 [10] are the state-of-the-art supervised and unsupervised photometric stereo method correspondingly, based on the benchmark evaluation on DiLiGenT10^2 [19]. Compared to these baselines, our method achieves the smallest mean and median MAE errors on the 100 scenes covering diverse shapes and materials. Similar to previous methods, our approaches also cannot handle the shape 'TURBINE' which introduces severe cast shadows and inter-reflections that are not explicitly modeled in our method. The translucent material 'ACRYLIC' is also challenging to us as our BRDF plugin is trained on the MERL dataset [12], which contains no translucent BRDFs. Our approach could handle this material given additional translucent BRDFs for training. **DiLiGenT.** Photometric stereo under sparse lights is of great interest for efficient capture. We investigate the normal estimation error of existing methods and ours on DiLiGenT [18] under 10 sparse lights. As shown in Table 1, LMPS [20] and SPLINE-Net [21] are designed for sparse-light photomet-

Table 1: Evaluation on DiLiGenT [18] dataset under 10 sparse lights, where the smallest and the second smallest MAE are highlighted in bold and underlined, respectively. The top and bottom parts of the table group supervised and unsupervised photometric stereo methods.

Method	Ball	Cow	Bear	Cat	Pot1	Pot2	Buddha	Goblet	Reading	Harvest	Avg.
LS [1]	4.58	26.48	9.84	8.9	9.59	15.65	16.02	19.23	19.37	31.32	16.1
CNN-PS [5]	8.21	13.83	11.89	9	12.79	15.04	13.39	15.74	16.07	19.36	13.53
PS-FCN [3]	4.35	9.97	5.7	8.24	8.38	10.37	10.54	11.21	_14.34_	18.82	10.19
GPS-Net [7]	4.33	9.34	6.34	6.81	7.5	8.38	**8.87**	10.79	15	**16.92**	**9.43**
LMPS [20]	3.97	10.19	8.73	6.69	7.3	9.74	11.36	10.46	14.37	_17.33_	10.01
SPLINE-Net [21]	4.96	_8.8_	5.99	7.52	8.77	11.79	_10.07_	_10.43_	16.13	19.05	10.35
TM18 [10]	_2.12_	8.87	6.92	6.58	_7.14_	9.61	11.41	14.99	**13.7**	26.55	10.79
LL22 [11]	2.16	**6.26**	_5.69_	_5.83_	8.13	_7.17_	12.65	10.93	31.64	24.05	11.45
Ours	**1.35**	9.79	**5.67**	**5.56**	**6.18**	**6.41**	10.11	**9.67**	14.77	28.63	_9.814_

Sphere   GT. N   Pot1   GT. N

Est. N

Err. map

LL22 [11]   Ours   LL22 [11]   Ours
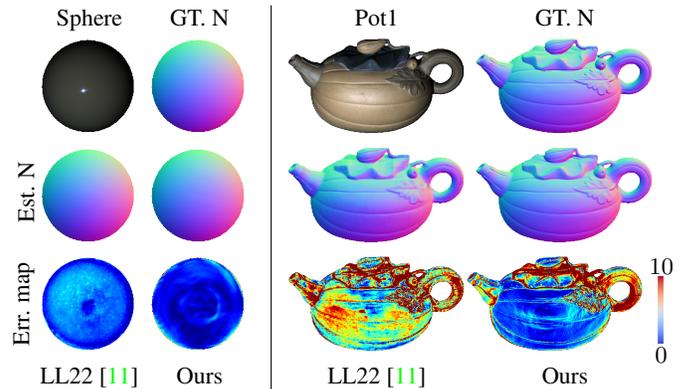
Fig. 5: Comparison with LL22 [11] under 10 sparse lights on two objects from the DiLiGenT [18] dataset.

ric stereo based on supervised learning. Compared with existing methods, our method achieves the smallest MAEs on 6 out of 10 objects of DiLiGenT [18], the smallest average MAE among unsupervised photometric stereo methods, and the second smallest average MAE among existing photometric stereo methods. Figure 5 further shows the estimated surface normals of LL22 [11] and ours on two DiLiGenT [18] objects. Benefiting from the data prior via our BRDF plugin, our normal estimates are closer to the GT. The angular error maps primarily display errors in areas with cast shadows, which our method does not explicitly address.

4. CONCLUSION

In this paper, we enhance unsupervised photometric stereo by introducing a BRDF plugin that imposes data prior to constraint surface reflectance. Our BRDF plugin is designed as a conditional coordinate-based MLP trained on real-world measured BRDFs, where BRDF data is projected into a learned latent space. Real-world and synthetic experiments show that by leveraging this plugin, our proposed method achieves more accurate surface normal estimates compared to other unsupervised photometric stereo approaches, bypassing the limitations associated with the large-scale training datasets required in supervised photometric stereo methods.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Robert J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical engineering*, 1980. 1, 4

[2] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita, "Deep photometric stereo network," in *Proc. of International Conference on Computer Vision Workshops (ICCVW)*, 2017. 1

[3] Guanying Chen, Kai Han, and Kwan-Yee K. Wong, "PS-FCN: A flexible learning framework for photometric stereo," in *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 1, 3, 4

[4] Yakun Ju, Boxin Shi, Muwei Jian, Lin Qi, Junyu Dong, and Kin-Man Lam, "NormAttention-PSN: A high-frequency region enhanced photometric stereo network with normalized attention," *International Journal of Computer Vision*, 2022. 1

[5] Satoshi Ikehata, "CNN-PS: CNN-based photometric stereo for general non-convex surfaces," in *Proc. of European Conference on Computer Vision (ECCV)*, 2018. 1, 3, 4

[6] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla, "PX-NET: simple and efficient pixel-wise training of photometric stereo networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. 1

[7] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi, "GPS-Net: Graph-based photometric stereo network," *Advances in Neural Information Processing Systems*, 2020. 1, 4

[8] Yakun Ju, Kin-Man Lam, Wuyuan Xie, Huiyu Zhou, Junyu Dong, and Boxin Shi, "Deep learning methods for calibrated photometric stereo and beyond: A survey," *arXiv preprint arXiv:2212.08414*, 2022. 1

[9] Qian Zheng, Boxin Shi, and Gang Pan, "Summary study of data-driven photometric stereo methods," *Virtual Reality & Intelligent Hardware*, 2020. 1

[10] Tatsunori Taniai and Takanori Maehara, "Neural inverse rendering for general reflectance photometric stereo," in *Proc. of the International Conference on Machine Learning (ICML)*. PMLR, 2018, pp. 4857–4866. 1, 4

[11] Junxuan Li and Hongdong Li, "Neural reflectance for shape recovery with shadow handling," *arXiv preprint arXiv:2203.12909*, 2022. 1, 3, 4

[12] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan, "A data-driven reflectance model," in *Proc. of SIGGRAPH*, 2003. 1, 2, 3, 4

[13] Szymon M Rusinkiewicz, "A new change of variables for efficient brdf representation," *Rendering techniques*, vol. 98, pp. 11–22, 1998. 2

[14] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2012. 2

[15] Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein, "pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 5799–5809. 2, 3

[16] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," *Proc. of Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2020. 2

[17] Heng Guo, Hiroaki Santo, Boxin Shi, and Yasuyuki Matsushita, "Edge-preserving near-light photometric stereo with neural surfaces," *arXiv preprint arXiv:2207.04622*, 2022. 3

[18] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan, "A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 4

[19] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi, "DiLiGenT10^2: A photometric stereo benchmark dataset with controlled shape and material variation," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4

[20] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita, "Learning to minify photometric stereo," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 4

[21] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling-Yu Duan, and Alex C. Kot, "SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks," in *Proc. of International Conference on Computer Vision (ICCV)*, 2019. 4