Self-calibrating Near-light Photometric Stereo under Anisotropic Light Emission

HENG GUO[1,a)] BOXIN SHI[2] MICHAEL WAECHTER[3] YASUYUKI MATSUSHITA[1]

Abstract

In conventional calibrated photometric stereo settings, a light source is assumed to be directional or to have an isotropic radiant intensity distribution (RID). These assumptions are hard to achieve in practice, and the deviation brings inaccurate shape recovery. In this work, we propose a self-calibrating near light photometric stereo method to jointly estimate the object shape and the anisotropic point light RID. Specifically, our method takes image observations, point light positions and sparse depth of the target object as input, and iterates the procedures of the anisotropic RID estimation and shape reconstruction. Experimental results on both synthetic and real data demonstrate the effectiveness of our method.

1. Introduction

Photometric stereo recovers surface orientations and albedos from a set of images taken from a fixed viewpoint under different light conditions. Early photometric stereo approaches [2, 10] assume distant and uniform light source, which is hard to perfectly achieve in reality. Near-light photometric stereo [5, 8] setup relieves such assumptions by taking the effects of light falloff and spatially varying light directions into account.

To achieve accurate shape recovery, precise calibration of the involved point light sources is required, which includes their 3D positions and the radiant intensity distributions (RIDs), *i.e.*, radiant intensity of a light source under different emitted directions. Existing near light photometric stereo methods either assume isotropic point light emission [4, 12], or use a calibration object to estimate the parametric RID models before applying photometric stereo algorithms [7, 8, 11]. However, existing calibration approaches have some disadvantages: 1) the radiance emitted by a point light typically changes over time due to ramp-up time and camera auto-exposure [5], leading to different photometric observations between the RID calibration step and the following image acquisition step in photometric stereo. 2) current RID representation models used in RID calibration [7, 11] and near-light photometric stereo [5, 6, 8] are limited to radially symmetric RIDs, which is not flexible enough to handle diversity of RIDs in the real world. To address these problems, we propose a self-calibrating photometric stereo framework in which object shape reconstruction and RID calibration are conducted simultaneously without the use of reference objects. Our key observation is that the observed RID as shown in Fig. 1 is located in a local region when illuminating an object. We assume this local RID is smooth and can be interpolated by spherical harmonic basis which is widely used for low-frequency illumination approximation. Based on the RID representation, we aim to estimate the shape of a Lambertian object with given calibrated point light positions and an initial estimate of sparse depth map for the target object.

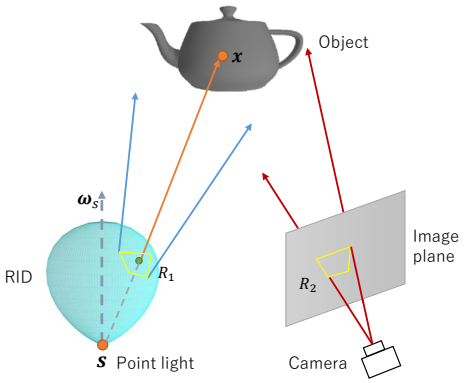

Figure 1: Illustration of a near-light scene under an anisotropic radiant intensity distribution (RID).

This paper makes the following contributions.

- We propose a near-light photometric stereo method with RID auto-calibration without using a reference object.
- We propose an easy-to-fit RID representation in a local region for real-world anisotropic light emission.
- We show our self-calibrated near-light photometric stereo solution achieves higher accuracy on shape recovery than state-of-the-arts.

2. Proposed method

As shown in Fig. 2, our method takes image observations, point light positions, and sparse depth as inputs, and output object shape, albedo and the local RID for each point light source. Our algorithm includes two modules: radiant intensity distribution estimation and dense surface reconstruction, *i.e.*, we first calculate the light RID, then apply it to the surface reconstruction process to optimize dense surface normal and depth alternatively. In the following sections, we will introduce the two modules in detail.

[1] Osaka University
[2] Peking University
[3] Facebook
[a)] heng.guo@ist.osaka-u.ac.jp

2.1 Radiant intensity distribution estimation

Under lambertian reflectance, a measured irradiance illumiated by the i-th near point light source $\mathbf{s}^i$ can be written as

$$m^i(\mathbf{p}) = \mathbf{b}^\top(\mathbf{p})\mathbf{l}^i(\mathbf{p}), i \in \{1, \cdots, f\} \quad (1)$$

where $\mathbf{b}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p}) \in \mathbb{R}^3$ is an albedo-scaled surface normal vector at image coordinate $\mathbf{p}$. Considering light falloff and anisotropic radiant intensity distribution, the lighting vector at position $\mathbf{p}$ induced by the point light $\mathbf{s}^i$ can be defined as

$$\mathbf{l}^i(\mathbf{p}) = e^i(\mathbf{p})\frac{\mathbf{s}^i - \mathbf{x}(\mathbf{p})}{\|\mathbf{s}^i - \mathbf{x}(\mathbf{p})\|_2^3}, \quad (2)$$

where $\mathbf{x}(\mathbf{p}) \in \mathbb{R}^3$ is the surface point position corresponding to pixel position $\mathbf{p}$, and $e^i(\mathbf{p}) \in \mathbb{R}$ represents the irradiance on this point emitted from i-th point light. We describe the light irradiance by the radiant intensity distribution fucntion (RIDF) $e^i(\cdot)$.

In general, each point light source has its independent RIDF $e^i(\cdot)$. An isotropic point light source emits light equally in all directions. Therefore, the RIDF $e^i(\cdot)$ is a constant function. In real-world, most RIDs are anisotropic with regard to the light directions. Existing methods [6, 8] use an exponential function with cosine as basis to represent the RID, as shown in Eq. (3)

$$e^i(\mathbf{p}) = \cos^{\mu_i}(\boldsymbol{\omega}^i \cdot \frac{\mathbf{s}^i - \mathbf{x}(\mathbf{p})}{\|\mathbf{s}^i - \mathbf{x}(\mathbf{p})\|}), \quad (3)$$

where $\boldsymbol{\omega}^i$ is the principal direction of point light, μ^i is the attenuation coefficient which can be estimated from light source datasheet. This RID representation model is limited to radially symmetric RIDs and also requires light source manufacturing information and additional calibration for the point light principle direction. To handle these problems, we propose a new parametric RID representation model which is able to handle various RIDs and has no need of extra calibration.

Since RIDF can be defined as a function in the spherical coordinate system, it is convenient to represent it with the linear combination of spherical harmonic (SH) basis, as shown below,

$$e^i(\mathbf{p}) = \sum_{k=1}^{h} c_k^i y_k(\frac{\mathbf{s}^i - \mathbf{x}(\mathbf{p})}{\|\mathbf{s}^i - \mathbf{x}(\mathbf{p})\|}), \quad (4)$$

where c_k^i and $y_k(\cdot)$ denote the coefficient and the basis of the k-th spherical harmonics for the i-th point light source. As shown in Fig. 1, the light direction from point light to the object fall into a limited region R_1. For photometric stereo, we mainly care about the radiant intensity distribution in this local region and it is reasonable to assume the RID in this patch is smooth enough to be interpolated by limited order of spherical harmonic (SH) basis.

Experiments shown in Fig. 3 support our observations. Here we use the top-9 coefficients of SH to fit real world RIDs under a local region defined in the range of 50 degrees deviated from the light principal direction. Compared with cosine-power based method used in [8], our SH-based RID representation model achieve higher fitting accuracy in anisotropic RID cases.

If we know surface point position $\mathbf{x}(\mathbf{p})$ and the light position

*1 https://www.nichia.co.jp/en/product/led.html

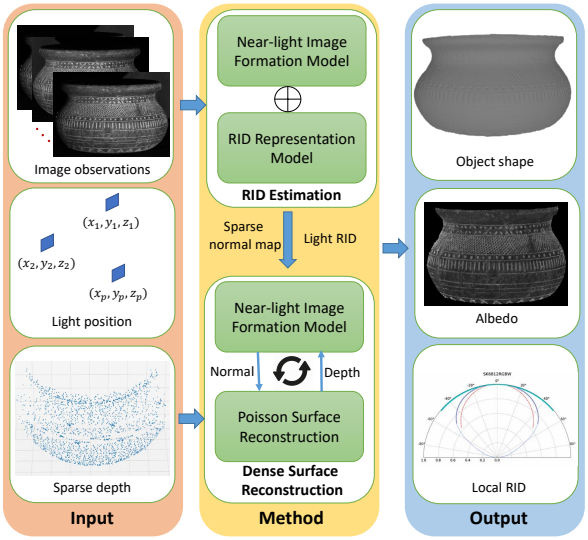

Figure 2: The pipeline of our method.

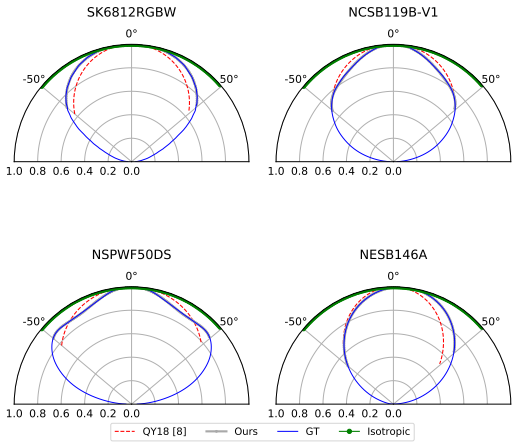

Figure 3: The representation power comparison of different approaches on four real world RIDs extracted from NACHIA Corp datasheet *1. The product name of the point lights are shown on the top of each RID plot. Notice that the point light "NESB146A" is not radially symmetric.

$\mathbf{s}^i$, we formulate the image formation constraint about albedo-scaled normal vector $\mathbf{b}(\mathbf{p})$ and inverse RID $f^i(\mathbf{p})$ following Eq. (1) and Eq. (2).

$$E_f\{\mathbf{b}(\mathbf{p}), f^i(\mathbf{p})\} = \{m^i(\mathbf{p})\|\mathbf{s}^i - \mathbf{x}(\mathbf{p})\|_2^3 f^i(\mathbf{p}) - \mathbf{b}^\top(\mathbf{p})(\mathbf{s}^i - \mathbf{x}(\mathbf{p}))\}^2 \quad (5)$$

where the inverse RID $f^i(\mathbf{p}) = 1/e^i(\mathbf{p})$ is also a function defined in the spherical coordinate, which can be interpolated by SH basis. Similarly, we define the inverse RID constraint about the inverse RID $f^i(\mathbf{p})$ and the corresponding SH coefficient of d_k^i from Eq. (4)

$$E_r\{f^i(\mathbf{p}), d_k^i\} = \{f^i(\mathbf{p}) - \sum_{k=1}^{h} d_k^i y_k(\frac{\mathbf{s}^i - \mathbf{x}(\mathbf{p})}{\|\mathbf{s}^i - \mathbf{x}(\mathbf{p})\|})\}^2 \quad (6)$$

Since we have the input depth z at sparse positions $\hat{\mathbf{p}}$, we can recover the corresponding 3D sparse surface point position based on perspective camera model as shown in Eq. (7).

$$\mathbf{x}(\hat{\mathbf{p}}) = z(\hat{\mathbf{p}})\mathbf{K}^{-1}\hat{\mathbf{p}}_h, \tag{7}$$

where $\mathbf{K}$ and $\hat{\mathbf{p}}_h$ are camera intrinsic matrix and homogeneous image coordinates.

With point light positions $\mathbf{s}^i$, sparse surface points $\mathbf{x}(\hat{\mathbf{p}})$ and image observations $m^i(\mathbf{p})$, we solve sparse albedo scaled surface normal, inverse RID and its correspnding SH coefficients for each point light by addressing both image formation constraint Eq. (5) and RID constraint Eq. (6).

$$\{\mathbf{b}(\hat{\mathbf{p}}), f^i(\hat{\mathbf{p}}), d_k^i\} = \sum_{\hat{\mathbf{p}}\in\Omega}\sum_{i=1}^{f}(E_f\{\mathbf{b}(\hat{\mathbf{p}}), f^i(\hat{\mathbf{p}})\} + \lambda E_r\{f^i(\hat{\mathbf{p}}), d_k^i\}) \tag{8}$$

λ is the weight to balance the two constraints. The above energy function can be formulated as a homogeneous linear equation, which can be easily solved by singular value decomposition (SVD) up to a scale ambiguity. The surface normal and albedo can be then calculated from $\mathbf{b}$ by applying per-pixel normalization.

2.2 Dense surface reconstruction

From last section we obtain sparse depth $z(\hat{\mathbf{p}})$ and $\mathbf{n}(\hat{\mathbf{p}})$. Following [3], the complete surface normal field can be calculated on the distinct patches containing such sparse oriented points. We obtained dense surface normal from [3] and refer it as initial surface normal $\mathbf{n}^0(\mathbf{p})$. Next, we refine the surface normal and depth iteratively guided by the image formation constraint and Poisson equation.

Let's denote depth in logarithmic scale as $z_l(\mathbf{p}) = \ln z(\mathbf{p})$. As stated in [9], the relationship between depth $z_l(\mathbf{p})$ and surface normal $\mathbf{n}(\mathbf{p})$ under the perspective camera model is,

$$\begin{cases} \nabla \mathbf{z}_l(\mathbf{p}) = [\boldsymbol{\psi}(\mathbf{p}), \boldsymbol{\gamma}(\mathbf{p})]^\top, \\ \boldsymbol{\psi}(\mathbf{p}) = -\dfrac{n_1(\mathbf{p})}{u(\mathbf{p})n_1(\mathbf{p}) + v(\mathbf{p})n_2(\mathbf{p}) + f_l n_3(\mathbf{p})}, \\ \boldsymbol{\gamma}(\mathbf{p}) = -\dfrac{n_2(\mathbf{p})}{u(\mathbf{p})n_1(\mathbf{p}) + v(\mathbf{p})n_2(\mathbf{p}) + f_l n_3(\mathbf{p})}, \end{cases} \tag{9}$$

where f_l is the camera focal length, $u(\mathbf{p})$ and $v(\mathbf{p})$ represent the horizontal and vertical image coordinate at pixel position $\mathbf{p}$, $\mathbf{n}(\mathbf{p}) = [n_1(\mathbf{p}), n_2(\mathbf{p}), n_3(\mathbf{p})]^\top$.

Applying the divergence operator, we can transform the above equation into a Poisson equation. By solving the following linear system, we can recover the complete depth map based on the initial surface normal map $\mathbf{n}^0(\mathbf{p})$ and sparse depth $z(\hat{\mathbf{p}})$.

$$\begin{pmatrix} \eta\mathbf{I} \\ \Delta \end{pmatrix} z_l(\mathbf{p}) = \begin{pmatrix} \eta z_l(\hat{\mathbf{p}}) \\ \frac{\partial \boldsymbol{\psi}}{\partial u} + \frac{\partial \boldsymbol{\gamma}}{\partial v} \end{pmatrix}. \tag{10}$$

where η defines the weight for the sparse depth constraint, experimentally we set it to 1. After obtaining log-scaled $z_l(\mathbf{p})$, the real complete depth map can be obtained by $z(\mathbf{p}) = e^{z_l(\mathbf{p})}$.

We refer the complete depth map computed from Eq. (10) as $z^\xi(\mathbf{p})$. Fixing the RIDF, we then update the albedo scaled surface normal $\mathbf{b}^{(\xi+1)}(\mathbf{p})$ by addressing the image formation constraint shown in Eq. (8). After that, we use Eq. (10) to calculate the depth $z^{\xi+1}(\mathbf{p})$ with updated surface normal map. We conduct this alternative optimization process until convergence or the iteration times reaching the predefined maximum number.

3. Experiment

We evaluate our method on both synthetic and real data in the setting of anisotropic near-light illumination.

3.1 Result on synthetic data

We render three synthetic objects named "Bunny", "Fattony" and "Jolie" with the size of 14 cm, and illuminate the scene with 49 virtual point lights fixed on a planer board placed 30 cm away from the object. The same real-world RID are assigned to all the point lights on the board. We evaluate the depth surface normal and RID estimated by our method and the semi-calibrated photometric stereo algorithm proposed in [8], in which the cosine-based model is used for RID representation. To get a fair comparison, we interpolate the input sparse depth to a complete depth map as the initial depth in [8]. To calculate the accuracy of depth, normal and RID, we apply the mean absolute error, mean angular error and mean squared error between the estimation and the ground truth, respectively. From Table 1 to Table 3 we can see, our self-calibrating photometric stereo method outperforms [8] on both shape and RID estimation.

Table 1: Comparison on depth estimation evaluated in mm

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**0.031**	**0.034**	**0.025**
	NSPWF50DS	**0.016**	**0.017**	**0.015**
	SK6812RGBW	**0.016**	**0.017**	**0.016**
QY18 [8]	NCSB119B-V1	0.767	1.171	2.104
	NSPWF50DS	1.381	0.530	1.154
	SK6812RGBW	1.309	1.301	1.376

Table 2: Comparison on normal estimation evaluated in degree.

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**0.022**	**0.039**	**0.096**
	NSPWF50DS	**0.238**	**0.013**	**0.283**
	SK6812RGBW	**0.097**	**0.191**	**0.319**
QY18 [8]	NCSB119B-V1	1.495	3.509	5.194
	NSPWF50DS	2.249	2.017	3.008
	SK6812RGBW	2.505	3.031	3.454

Table 3: Comparison on RID estimation.

Method	RID	Bunny	FatTony	Julia
Ours	NCSB119B-V1	**1.023e-6**	**1.227e-6**	**7.963e-6**
	NSPWF50DS	**2.018e-5**	**1.285e-7**	**1.407e-5**
	SK6812RGBW	**1.325e-5**	**3.675e-5**	**3.467e-5**
QY18 [8]	NCSB119B-V1	4.717e-5	3.383e-4	3.135e-4
	NSPWF50DS	1.559e-4	6.125e-5	8.416e-5
	SK6812RGBW	1.565e-4	2.590e-4	1.603e-4

3.2 Result on real data

Figure 4 shows our capture setup. It includes two Canon EOS 5D Mark IV cameras with 85mm lens and 256 point lights with their positions calibrated. The baseline between two cameras is 165mm. We apply semi-global stereo matching algorithm [1] to

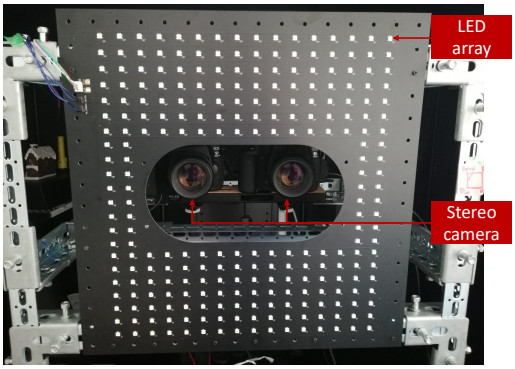

Figure 4: Our capture setup.

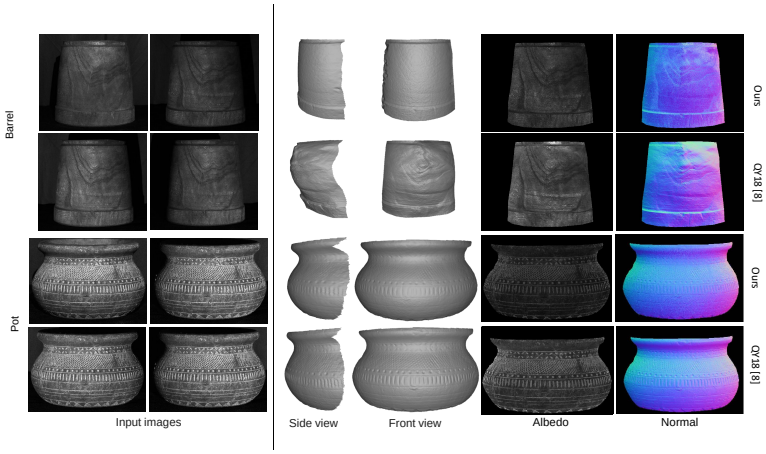

Figure 5: Shape and albedo comparison between our method and Quéau *et al*. [8] on real data

calculate the disparity between the stereo image pairs and further obtain the initial sparse depth of the target. Our system is equipped with point light "SK6812RGBW" from OPSCO Optoelectronics Corp[*2] with its theoretical RID shown in Fig. 3. The distance between the object and light sources is roughly 0.8m.

As shown in Fig. 5, we capture two objects "Barrel" and "Pot" illuminated by 40 of 256 point lights, and then compare with [8] on shape and albedo estimation. From the sideview in Fig. 5 we can see that our method with SH-based RID representation model outputs more accurate surface shape in "Barrel" case and achieves comparable results on "Pot" object.

4. Conclusion

In this paper, we propose a self-calibrating near-light photometric stereo method to handle shape recovery under anisotropic point lights. Our key observation is that in the photometric stereo task the interest RID is located in a local region. Based on this observation, we propose a RID representation model based on spherical harmonic basis which is flexible to various type of RIDs and without extra calibration. Given the sparse depth of the target object as input, our recovered RID and complete object shape achieve higher accuracy compared to the existing methods.

5. Acknowledgement

This work is supported by JSPS KAKENHI Grant Number JP19H01123.

References

[1] Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information, *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 30, No. 2, pp. 328–341 (2007). 3

[2] Ikehata, S., Wipf, D., Matsushita, Y. and Aizawa, K.: Robust photometric stereo using sparse regression, *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 318–325 (2012). 1

[3] Kazhdan, M., Bolitho, M. and Hoppe, H.: Poisson surface reconstruction, *Proceedings of the Eurographics symposium on Geometry processing*, Vol. 7 (2006). 3

[4] Liu, C., Narasimhan, S. G. and Dubrawski, A. W.: Near-light photometric stereo using circularly placed point light sources, *2018 IEEE International Conference on Computational Photography (ICCP)*, IEEE, pp. 1–10 (2018). 1

[5] Logothetis, F., Mecca, R. and Cipolla, R.: Semi-calibrated near field photometric stereo, *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 941–950 (2017). 1

[6] Logothetis, F., Mecca, R., Quéau, Y. and Cipolla, R.: Near-field photometric stereo in ambient light, *Proceedings of the British Machine Vision Conference (BMVC)* (2016). 1, 2

[7] Park, J., Sinha, S. N., Matsushita, Y., Tai, Y.-W. and So Kweon, I.: Calibrating a non-isotropic near point light source using a plane, *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2259–2266 (2014). 1

[8] Quéau, Y., Durix, B., Wu, T., Cremers, D., Lauze, F. and Durou, J.-D.: Led-based photometric stereo: modeling, calibration and numerical solution, *Journal of Mathematical Imaging and Vision*, Vol. 60, No. 3, pp. 313–340 (2018). 1, 2, 3, 4

[9] Quéau, Y., Durou, J.-D. and Aujol, J.-F.: Normal integration: a survey, *Journal of Mathematical Imaging and Vision*, Vol. 60, No. 4, pp. 576–593 (2018). 3

[10] Shi, B., Matsushita, Y., Wei, Y., Xu, C. and Tan, P.: Self-calibrating photometric stereo, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 1118–1125 (2010). 1

[11] Visentini-Scarzanella, M. and Kawasaki, H.: Simultaneous camera, light position and radiant intensity distribution calibration, *Image and Video Technology*, Springer, pp. 557–571 (2015). 1

[12] Xie, L., Xu, Y., Zhang, X., Bao, W., Tong, C. and Shi, B.: A self-calibrated photo-geometric depth camera, *The Visual Computer*, Vol. 35, No. 1, pp. 99–108 (2019). 1

[*2] https://www.opscoled.com/en/about/index.html