Gaussian Process based Illumination Planning for Photometric Stereo

Yuji Oyamada[1]

Tottori University, Japan oyamada@tottori-u.ac.jp

Abstract. Photometric Stereo estimates surface normal from a set of shading images observed by a fixed camera under different lights. Although robust regression techniques and deep neural networks improves accuracy and robustness of normal estimation, all of the literature assumes input shading images are high quality, e.g., the corresponding light positions are spatially distributed and each image has as less shadow as possible. This is not easy to reproduce especially for less-experienced persons. Therefore, we typically increase the number of image acquisition. An interesting method was proposed by Tanikawa et al. that considers shadow and the distribution of light positions. This paper proposes a Gaussian Process based illumination planning for shading image acquisition. The method first chooses three pre-defined light positions and observes images independently and then iterates optimum light selection and shading image acquisition until some condition is satisfied. The experiment compares the proposed method, Tanikawa's method, and random selection to validate the effectiveness of the proposed method.

Keywords: Illumination planning · Photometric Stereo · Gaussian Process.

1 Photometric Stereo

This section introduces the motivation of this paper: what is photometric stereo and why we need illumination planning.

1.1 Potometric Stereo

Photometric stereo is a technique of 3D scene recovery that estimates object surface from a set of shading images observed by a fixed camera under different lightning conditions [11]. The simplest model assumes Lambertian surface and point distance light source [13]. When a point of an object of Lambertian reflectance is lit by point distanct light source, the observed pixel intensity $i \in mathbbR_+$ is written as

$$i = \rho \max(\boldsymbol{n} \cdot \boldsymbol{l}, 0), \tag{1}$$

where $\boldsymbol{n} \in \mathbb{R}^3, \|\boldsymbol{n}\| = 1$ denotes the surface normal vector, $\rho \in \mathbb{R}$ the surface albedo, and $\boldsymbol{l} \in \mathbb{R}^3$ the light vector. As the equation explains, the brightness of

the observation is proportional to the surface albedo ρ and the cosine similarity of the surface normal $\boldsymbol{n}$ and the light vector $\boldsymbol{l}$. Note that the brightness i is 0 when the dot product $\boldsymbol{n} \cdot \boldsymbol{l}$ is negative (when the light does not arrive at the point in other words). Lambertian Photometric Stereo (LPS) recovers the surface normal vector from the observation under $F \geq 3$ non-collinear lighting conditions. By formulating the F observation as a linear equation

$$\underbrace{\begin{bmatrix} i_1 \\ \vdots \\ i_F \end{bmatrix}}_{\boldsymbol{i} \in \mathbb{R}^{F \times 1}} = \underbrace{\begin{bmatrix} \boldsymbol{l}_1^\top \\ \vdots \\ \boldsymbol{l}_F^\top \end{bmatrix}}_{\boldsymbol{L} \in \mathbb{R}^{F \times 3}} (\rho \boldsymbol{n}) \tag{2}$$

$$\rightarrow \boldsymbol{i} = \boldsymbol{L}\boldsymbol{s}, \tag{3}$$

LPS recovers the surface vector $\boldsymbol{s}$ by least squares and then decouples it into ρ and $\boldsymbol{n}$ as

$$\hat{\boldsymbol{s}} = \boldsymbol{L}^\dagger \boldsymbol{i} \tag{4}$$

$$\hat{\rho} = \|\boldsymbol{s}\| \tag{5}$$

$$\hat{\boldsymbol{n}} = \frac{\boldsymbol{s}}{\hat{\rho}}, \tag{6}$$

where $\boldsymbol{L}^\dagger$ denotes the pseudo inverse matrix of the light matrix $\boldsymbol{L}$. Typically, 10-20 images are used for obtaining stable and accurate 3D reconstruction. The limitation of LPS is due to its assumption. When a target object has complex reflection property such as specular reflection and subsurface scattering, Eq. 1 does not hold. Attached and cast shadow also violates the observation model even with Lambertian objects. Furthermore, lights in daily life such as near light requires different models from Eq. 1.

State of the art Photometric Stereo methods are categorized into two types: robust least squares or deep learning methods. Robust Photometric Stereo (RPS) introcues the sense of sparse regression [14, 7, 8]. Regarding observation violating Eq. 1 as outliers, RPS recovers surface normal of non-Lambertian objects even with shadows. RPS relies on a simple assumption that inlier observation is dominant and therefore robust regression can elimiate the negative effect of outlier observations. Deep Photometric Stereo (DPS) solves the Photometric Stereo problem with deep neural network manners [10, 6, 2]. Thanks to the powerful deep neural networks, DPS can handle non-Lambertian objects and even unknown light conditions.

1.2 Good Shading Images

Aside from the aforementioned PS techniques, the quality of shading images takes very important part in PS. As least squares favors observation at distributed positions, PS does shading images with variety of lighting conditions [3].

Specifically, the spatial variance of $\boldsymbol{L}$ is important. One of the simplest formulation of the variance is proposed by Drbohlav and Chantler as

$$\text{trace}\left(\boldsymbol{L}^\top \boldsymbol{L}\right). \tag{7}$$

The variance is strongly related with normal estimation uncertainty. Representing observation noise as δ_f, 1 is formulated as

$$\begin{bmatrix} i_1 \\ \vdots \\ i_F \end{bmatrix} = \begin{bmatrix} \boldsymbol{l}_1^\top \\ \vdots \\ \boldsymbol{l}_F^\top \end{bmatrix} (\rho \boldsymbol{n}) + \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_F. \end{bmatrix} \tag{8}$$

Then, the expected error given $\boldsymbol{L}$ is derived as

$$\epsilon(\boldsymbol{L}) = E\left[(\boldsymbol{n} - \hat{\boldsymbol{n}})^\top (\boldsymbol{n} - \hat{\boldsymbol{n}})\right] \tag{9}$$

$$= \sigma^2 \text{trace}\left(\left(\boldsymbol{L}^\top \boldsymbol{L}\right)^{-1}\right). \tag{10}$$

1.3 How to Acquire Good Shading Images?

The way of shading image acquisition requires either expertise knowledge or computer guidance. The expertise knowledge is from intuition from experts that light positions must be distributed and as less shadow as possible, which strongly connected to the theory by Drbohlav and Chantler [3]. We have a dilemma that some people heavily rely on this kind of expertise knowledge but some avoid such indescribable knowledge.

The computer guidance method (online illumination planning) proposes next best light position to the observer [12]. The method first takes three images under pre-defined light positions and then iterates optimum light position prediction and image acquisition. Each prediction in the second step first selects a single pixel maximizing Eq. 10, namely a pixel with worst estimation uncertainty. The optimum light position is computed from the previous observations that simultaneously avoids shadow and puts the light position further from the previously observed light positions. As far as the authors know, the paper is the first work that introduces the concept of (online) view planning into PS image acquisition.

This method has two main drawbacks. The first one is in the optimum light position estimation. The effect of shadow dominates their evaluation cost and the domination tends to select pixels with same/opposite normal vectors located around object boundaries. This tendency results light positions similar to each other. The second drawback is in the pixel selection step. The method predict estimation uncertainty for all the pixels and selects the one with the largest uncertainty. We should care how many pixels a selected light shed. The worst case scenario is that a selected light sheds the pixel but not the remaining ones.

2 Proposed method

This paper proposes a Gaussian Process based illumination planning for shading image acquisition. The basic idea of the method is same as the online illumina-

Algorithm 1 The proposed method

Input: Pre-defined light positions $(\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3)$, potential light position candidates $\mathcal{L} = \{\boldsymbol{l}\}$, and the maximum number of observations N.

$n = 1$
$\mathcal{D} = \emptyset$
while $n \leq 3$ **do**
  $\boldsymbol{I}_n = \mathrm{CaptureImage}(\boldsymbol{l}_n)$
  $\mathcal{D} = \mathcal{D} \cup \{(\boldsymbol{l}_n, \boldsymbol{I}_n)\}$
  $n+ = 1$
end while
while $n \leq N$ **do**
  $\boldsymbol{l}_n = \arg\max_{\boldsymbol{l} \in \mathcal{L}} \mathrm{ComputeUncertainty}(\mathcal{D}, \boldsymbol{l})$
  $\boldsymbol{I}_n = \mathrm{CaptureImage}(\boldsymbol{l}_n)$
  $\mathcal{D} = \mathcal{D} \cup \{(\boldsymbol{l}_n, \boldsymbol{I}_n)\}$
  $\mathcal{L} = \mathcal{L} \setminus \boldsymbol{l}_n$
  $n+ = 1$
end while

tion planning by Tanikawa et al. [12]. Algorithm 1 shows the procedure of the proposed method. The method (1) first chooses three pre-defined light positions and observes images independently, (2) selects representing pixels by considering their correlation given the observed three images, and (3) iterates optimum light selection and observation until some condition is satisfied.

The key idea is from Gaussian Process based sensor location problem [5]. By formulating target observation by a Gaussian Process [9], we can compute the uncertainty of unobserved light positions. Coupling with an approximation algorithm, the quality of light position selection is guaranteed with a constant factor. The proposed method considers the uncertainty of unobserved light positions and selects the one that decreases the uncertainty the most. Considering the uncertainty of unobserved light positions at some pixels [1], the proposed method can select as distributed light positions as possible.

2.1 Observation uncertainty of a pixel

This subsection defines the observation uncertainty of a pixel.

Assuming that shading image acquisition follows a Gaussian Process as

$$\begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_F \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(f(\boldsymbol{l}_1)) \\ \mu(f(\boldsymbol{l}_2)) \\ \vdots \\ \mu(f(\boldsymbol{l}_F)) \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{l}_1, \boldsymbol{l}_1) & k(\boldsymbol{l}_1, \boldsymbol{l}_2) & \dots & k(\boldsymbol{l}_1, \boldsymbol{l}_F) \\ k(\boldsymbol{l}_2, \boldsymbol{l}_1) & k(\boldsymbol{l}_2, \boldsymbol{l}_2) & \dots & k(\boldsymbol{l}_2, \boldsymbol{l}_F) \\ \vdots & \vdots & \vdots & \vdots \\ k(\boldsymbol{l}_F, \boldsymbol{l}_1) & k(\boldsymbol{l}_F, \boldsymbol{l}_2) & \dots & k(\boldsymbol{l}_F, \boldsymbol{l}_F) \end{bmatrix} \right), \tag{11}$$

$$\boldsymbol{i} \sim \mathcal{N}\left(\mu(\boldsymbol{L}), \boldsymbol{K}\right) \tag{12}$$

[1] Note that the proposed method considers the limited number of representing pixels but not all pixels.

where μ denotes the mean function, $f()$ the observation function, and k kernel function represents the similarity between two light positions. Utilizing the property of Gaussian distribution, we can formulate a joint probability of the existing observation and one under a new light position $\boldsymbol{l}_*$ as

$$\begin{bmatrix} \boldsymbol{i} \\ i_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\boldsymbol{L}) \\ \mu(\boldsymbol{l}_*) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{k}_* \\ \boldsymbol{k}_*^\top & \boldsymbol{K}_* \end{bmatrix} \right), \tag{13}$$

$$\boldsymbol{k}_* = (k(\boldsymbol{l}_*, \boldsymbol{l}_1), k(\boldsymbol{l}_*, \boldsymbol{l}_2), \ldots, k(\boldsymbol{l}_*, \boldsymbol{l}_F))^\top \tag{14}$$

$$\boldsymbol{K}_* = k(\boldsymbol{l}_*, \boldsymbol{l}_*). \tag{15}$$

Thus, the mean $\mu(\boldsymbol{l}_*)$ and variance $\sigma(\boldsymbol{l}_*)$ of the new light position $\boldsymbol{l}_*$ is

$$\mu(\boldsymbol{l}_*) = \boldsymbol{k}_*^\top \boldsymbol{K}^{-1} \boldsymbol{i} \tag{16}$$

$$\sigma(\boldsymbol{l}_*) = \boldsymbol{K}_* - \boldsymbol{k}_*^\top \boldsymbol{K}^{-1} \boldsymbol{k}_*. \tag{17}$$

Thus, we can select the next best light position that reduces the observation uncertainty.

The key factor of this model is the definition of the kernel function $k()$. With the linear kernel

$$k_{\text{linear}}(\boldsymbol{l}_i, \boldsymbol{l}_j) = \boldsymbol{l}_i \cdot \boldsymbol{l}_j \tag{18}$$

the uncertainty behaves similar to Eq. 7. The proposed method uses RBF kernel as

$$k_{\text{rbf}}(\boldsymbol{l}_i, \boldsymbol{l}_j) = \exp\left(-\frac{|\boldsymbol{l}_i - \boldsymbol{l}_j|}{\theta} \right) \tag{19}$$

2.2 Observation uncertainty of an image

This subsection extends the observation uncertainty model from a pixel to an image. The proopsed method utilizes a multi-task GP model [1] that computes multiple output values, multiple pixel intensities in other words, for each input light vector. To reduce memory usage, the proposed method computes the observation uncertainty of some pixels, 16 in the experiment.

For P pixels, Eq. 12 is extended to

$$\begin{bmatrix} \boldsymbol{i}_1 \\ \vdots \\ \boldsymbol{i}_P \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1(\boldsymbol{L}) \\ \vdots \\ \mu_P(\boldsymbol{L}) \end{bmatrix}, \boldsymbol{K}_l \otimes \boldsymbol{K}_p \right), \tag{20}$$

$$\boldsymbol{K}_l = \begin{bmatrix} k_l(\boldsymbol{l}_1, \boldsymbol{l}_1) & k_l(\boldsymbol{l}_1, \boldsymbol{l}_2) & \dots & k_l(\boldsymbol{l}_1, \boldsymbol{l}_F) \\ k_l(\boldsymbol{l}_2, \boldsymbol{l}_1) & k_l(\boldsymbol{l}_2, \boldsymbol{l}_2) & \dots & k_l(\boldsymbol{l}_2, \boldsymbol{l}_F) \\ \vdots & \vdots & \vdots & \vdots \\ k_l(\boldsymbol{l}_F, \boldsymbol{l}_1) & k_l(\boldsymbol{l}_F, \boldsymbol{l}_2) & \dots & k_l(\boldsymbol{l}_F, \boldsymbol{l}_F) \end{bmatrix} \tag{21}$$

$$\boldsymbol{K}_p = \begin{bmatrix} k_p(\boldsymbol{x}_1, \boldsymbol{x}_1) & k_p(\boldsymbol{x}_1, \boldsymbol{x}_2) & \dots & k_p(\boldsymbol{x}_1, \boldsymbol{x}_P) \\ k_p(\boldsymbol{x}_2, \boldsymbol{x}_1) & k_p(\boldsymbol{x}_2, \boldsymbol{x}_2) & \dots & k_p(\boldsymbol{x}_2, \boldsymbol{x}_P) \\ \vdots & \vdots & \vdots & \vdots \\ k_p(\boldsymbol{x}_P, \boldsymbol{x}_1) & k_p(\boldsymbol{x}_P, \boldsymbol{x}_2) & \dots & k_p(\boldsymbol{x}_P, \boldsymbol{x}_P) \end{bmatrix} \tag{22}$$

where $\otimes$ denotes Hadamardt product of two matrices, $\boldsymbol{K}_l$ and $\boldsymbol{K}_p$ denote the covariance matrices of light vector and pixel locations, $k_l()$ and $k_p()$ denote the kernel function for light vectors and pixel locations, and $\boldsymbol{x}_p$ denotes the pixel location of a pixel p. Similar to the single pixel case, we can derive the uncertainty of an unobserved light vector $\boldsymbol{l}_*$ as

$$\Sigma(\boldsymbol{l}_*) = \boldsymbol{K}_* - (\boldsymbol{K}_l \otimes \boldsymbol{k}_*)^\top (\boldsymbol{K}_l \otimes \boldsymbol{K}_p)^{-1} (\boldsymbol{K}_l \otimes \boldsymbol{k}_*)^\top. \tag{23}$$

As mentioned at the beginning of this section, the proposed method considers the observation uncertainty of some pixels but not entire image. One critical drawback of the proposed method is its memory usage. GP computes the inverse matrix whose computational cost is proportional to $(PF)^3$. It is redundant to consider all the pixels because neighboring pixels have similar information. Therefore, the proposed method selects representing pixels from all the pixels and uses the observation uncertainty obtained from the selected pixels using Eq. 23. In this paper, we randomly select 16 pixels from entire image.

2.3 Optimum light position from observation uncertainty

Here, we describe a method selecting the next best light position from the observation uncertainty. For finite light position candidates $\mathcal{L} = \{\boldsymbol{l}_1, \dots, \boldsymbol{l}_N\}$, we select a light vector that satisfies the following condition

$$\hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{l} \in \mathcal{L}} \Sigma(\boldsymbol{l}). \tag{24}$$

Note that the proposed method solves illumination planning as a discrete optimization. Under the selected light vector $\hat{\boldsymbol{l}}$, we put the light at the corresponding position and acquire a shading image. Finally, we recover surface normal by applying a Photometric Stereo method.

3 Experiment

We validated the effectiveness of the proposed method. In this experiment, we used an image acquisition simulator that renders a shading image given a light vector. We assumed that the target objectt has Lambertian reflection and uniform albedo object and is lit by a distant light source and no cast shadow appears. The rendering engine adds white Gaussian noise with standard deviation of 0.05 for pixel intensities ranging 0.00 to 1.00. As target object, we used Stanford bunny and Lucy. We set the number of image acquisition to 14 and ran the proposed method, Tanikawa's method, and random selection. In the proposed method, both $k_l()$ and $k_p()$ were RBF kernels with different hyper-parameters. The hyper-parameter is optimized by Adam optimizer with learning rate 0.1 and all the implementation is done by GPyTorch [4]. In Tanikawa's method, our implementation was the discrete optimization version to avoid local minimum. For fair comparison, first three light positions are same for all the methods. As Photometric Stereo solver, we used both L1 residual minimization and Sparse Bayesian learning method of RPS [7] [2].

3.1 Selected light position

Fig. 1 compares the selected light positions of all the methods, we call illumination planning map hereafter for Standard bunny case. A big gray circle represents all the potential light positions and position (x, y) in the circle represents light vector $(x, y, 1-\sqrt{x^2+y^2})$. Each selected light is depicted as a colored circle and the color of the circle represents the selected order.

Comparing the planning maps, our method selected light positions uniformly distributed on a sphere while Tanikawa's method selected less distributed light positions. On the other hand, Random selection selected half distributed and half concentrated light positions. From the result, our method favors spatial distribution of light positions while Tanikawa's method seems to put too much weight on shadow.

3.2 Photometric Stereo accuracy

We conducted quantitative evaluation among the methods. For each object, we ran each method with the maximum number of observations N 15. For each observation, we computed the mean value of the angular difference in degree between the estimated normal map and its ground truth. For Random selection, we ran 20 times and computed their mean and standard deviation. Figure 2 shows the angular difference w.r.t. the number of observations for Stanford bunny and Lucy. In general, all the methods obtain smaller error for more observations. For some cases, we obtain larger error with more observations that is due to observed shadow. The reason why both the illumination planning results worse

[2] Note that RPS works worse for pixels dominated by shadows while one used in the evaluation of Tanikawa's paper [12] avoid such violation from shadow.

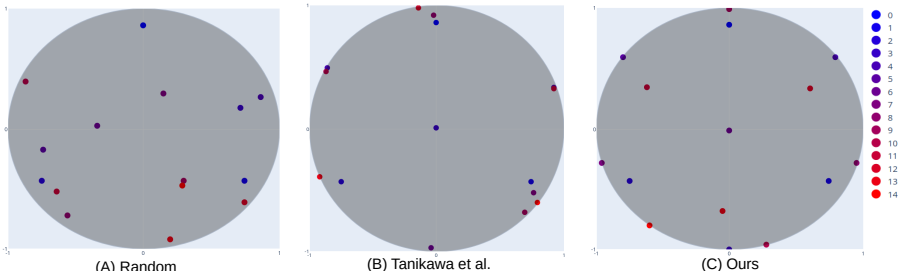


Fig. 1. Experimental result (illumination planning map): (A) Random selection, (B) Tanikawa, and (C) Ours.

error with more images is that they select light position is further to camera direction, which is the border of illumination map. When light directs further from the camera direction, attached shadow appears on more pixels. Those shadow worsen normal estimation quality.

Fig. 3 shows the ground truth and the estimated normal maps with each selected 14 lights. The similarity to the ground truth gives same impression that Tanikawa's method is less similar to the ground truth while Random selection and our method estimates closer normal map to the ground truth.

4 Conclusion

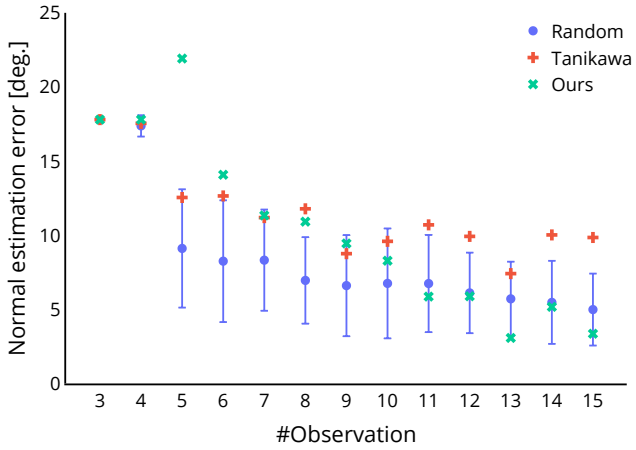
This paper tackles the illumination planning problem for Photometric Stereo. Our idea is to select light position that has the largest observation uncertainty given current observations. The positive effect of the idea is that the selected light positions are distributed more than the existing method. However, the proposed method do not outperform Random selection, Random selection outperforms the proposed method for less number of observations. This unwilling result might happen because of the following factors. First of all, we consider the observation uncertainty of very limited number of pixels to reduce memory usage and the selection is randomly done. To follow this manner, we must consider dependency among the selected pixels that well-represent entire image w.r.t. surface normal. Another factor is to consider camera direction for light position selection. As mentioned above, light position further from the camera direction worsen normal estimation quality. This expertise must be considered.

Acknowledgments

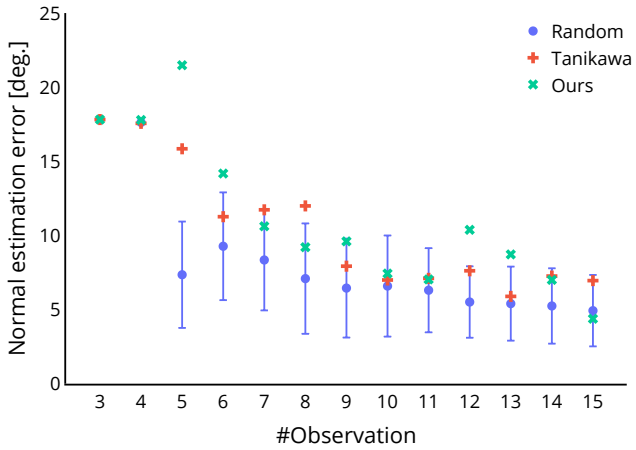
The author would like to thank the anonymous reviewers for giving advises for updating the experiments. This work was partially supported by JSPS KAKENHI Grant Number 19K20296.

References

1. Bonilla, E.V., Chai, K.M.A., Williams, C.K.I.: Multi-task gaussian process prediction. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). pp. 153—-160 (2007)
2. Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.Y.K.: Deep photometric stereo for non-lambertian surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **44**(01), 129–142 (2022)
3. Drbohlav, O., Chantler, M.: On optimal light configurations in photometric stereo. In: IEEE International Conference on Computer Vision (ICCV). vol. 2, pp. 1707–1712 (2005)
4. Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G.: Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In: Proceedings of International Conference on Neural Information Processing Systems (NIPS). pp. 7587–7597 (2018)
5. Guestrin, C., Krause, A., Singh, A.P.: Near-optimal sensor placements in gaussian processes. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 265—-272 (2005)
6. Ikehata, S.: Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Proceedings of European Conference on Computer Vision (ECCV). pp. 3–19 (2018)
7. Ikehata, S., Wipf, D., Matsushita, Y., Aizawa, K.: Robust photometric stereo using sparse regression. In: Proceedings of Computer Vision and Pattern Recognition (CVPR) (2012)
8. Ikehata, S., Wipf, D.P., 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), 1078–1091 (2014)
9. Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning. MIT Press (2006)
10. Santo, H., Samejima, M., Sugano, Y., Shi, B., Matsushita, Y.: Deep photometric stereo network. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). pp. 501–509 (2017)
11. Szeliski, R.: Computer Vision: Algorithms and Applications, chap. 3D Reconstruction. Springer (2022)
12. Tanikawa, H., Kawahara, R., Okabe, T.: Online illumination planning for shadow-robust photometric stereo. In: International Workshop on Frontiers of Computer Vision (IW-FCV). pp. 94–107 (2022)
13. Woodham, R.J.: Photometric Method For Determining Surface Orientation From Multiple Images. Optical Engineering **19**(1), 139–144 (1980)
14. Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., Ma, Y.: Robust photometric stereo via low-rank matrix completion and recovery. In: Proceedings of Asian Conference on Computer Vision (ACCV) (2010)



(a) Stanford bunny



(b) Lucy

Fig. 2. The angular error in degree w.r.t. the number of observations. Blue circle with error bar represents Random selection, red cross Tanikawa's method, and green x ours.

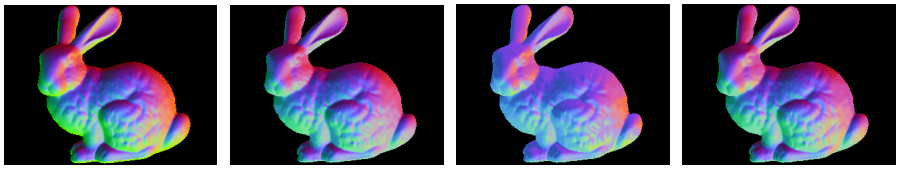


Fig. 3. Experimental result (estimated normal map): (A) Ground truth, (B) Random selection, (C) Tanikawa, and (D) Ours.