Online Illumination Planning for Shadow-Robust Photometric Stereo

Hirochika Tanikawa[1], Ryo Kawahara[1][0000−0002−9819−3634], and Takahiro Okabe[1][0000−0002−2183−7112]

Department of Artificial Intelligence,
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
okabe@ai.kyutech.ac.jp

Abstract. Photometric stereo is a technique for estimating normals of an object surface from its images taken under different light source directions. In general, photometric stereo suffers from shadows, because almost no information on surface normals is available from shadowed pixels. In this paper, we propose an illumination planning for shadow-robust Lambertian photometric stereo; it optimizes the light source directions adaptively for an object of interest, because cast shadows depend on the entire shape of the object. More specifically, our proposed method iteratively adds the optimal light source for surface normal estimation by taking the visibility and linear independence of light source directions into consideration on the basis of the previously captured images of the object. We implemented our illumination planning with a programmable light source in an online manner, and achieve shadow-robust surface normal estimation from a small number of images.

Keywords: photometric stereo · illumination planning · attached shadow · cast shadow · visibility

1 Introduction

Photometric stereo is a technique for estimating surface normals of an object of interest from its images taken from a fixed viewpoint but under different light source directions. The classical photometric stereo [18, 15] assumes the Lambert model and known light sources, and then estimates surface normals from at least three images. The extension of the classical photometric stereo to non-Lambertian surfaces, unknown light sources, and so on is still one of the most actively studied topics in computer vision [1, 14].

In this paper, we focus on shadows, *i.e.* another aspect of the extension of the classical photometric stereo, and propose photometric stereo robust to shadows. In general, shadows are classified into *attached shadows* and *cast shadows*; attached shadow is local effect and occurs when the angle between a surface normal and a light source direction is obtuse, but cast shadow is global effect and occurs when a light source is blocked by another part of an object surface. Because

almost no information on surface normals is available from shadowed pixels [1], we need to illuminate an object surface by using appropriate light sources so that each pixel on the surface is sufficiently illuminated.

Optimizing light sources, so-called *illumination planning* [11] is studied also in the context of photometric stereo. Drbohlav and Chantler [6] address the optimal light source configurations for the Lambertian photometric stereo on the basis of the noise propagation analysis. For example, they show that, when the number of light sources is three, the three light sources should be placed so that those directions are orthogonal to each other. Unfortunately, however, their analysis does not take shadows into consideration, although attached shadows occur even on convex objects in practice. More importantly, they find the optimal light source directions *a priori* independent of the shape of an object. Actually, cast shadows occur on non-convex objects depending on the entire shapes of the objects.

Accordingly, we propose an *online* illumination planning for shadow-robust Lambertian photometric stereo. Our proposed method estimates surface normals of an object of interest by iteratively adding a novel image of the object taken under the optimal light source direction. More specifically, we take the visibility and linear independence of light source directions into consideration, and propose a method for optimizing the light source direction adaptively for an object of interest on the basis of the previously captured images of the object. We implemented our illumination planning by using a Liquid Crystal Display (LCD) as a programmable light source, and confirmed the effectiveness of our method through a number of experiments using synthetic and real images.

The main contributions of this study are twofold. First, we propose a novel illumination planning for shadow-robust Lambertian photometric stereo. Our proposed method iteratively adds the optimal light source according to an object of interest on the basis of the previous observations of the object. Second, we implemented our illumination planning with a programmable light source in an online manner, and achieve shadow-robust surface normal estimation from a small number of images.

2 Related Work

Conventional Approach: Outlier Removal

Both attached shadows and cast shadows are considered as outliers, because the pixel values in those shadows deviate from the Lambert model. Therefore, a number of methods are proposed for detecting shadows in the images for photometric stereo and removing them as outliers from surface normal estimation. For example, the shadows in photometric stereo images are detected by selecting the optimal three light sources out of four [2], by using RANdom SAmple Consensus (RANSAC) [13, 10, 16], graph cut [3], and low-rank and sparse decomposition [19, 8]

[1] It is known that surface normals can be recovered from attached shadows if a large number of images taken under varying light source directions are given [12].

Our objective is to achieve shadow-robust photometric stereo too. In contrast to the above methods, however, our focus is not on detecting shadows but on illuminating an object surface in the context of active illumination. Specifically, we illuminate an object surface so that each pixel is sufficiently illuminated for surface normal estimation.

Learning-Based Approach

The learning-based approach is getting more common also in photometric stereo. There are a number of learning-based methods for non-Lambertian photometric stereo and uncalibrated photometric stereo [5, 17, 4, 20]. It is certain that the learning-based approach is suitable for dealing with non-Lambertian BRDFs because the reflectance is local effect. Unfortunately, however, cast shadows depend on the entire shape of an object, and then they are global effects. Therefore, efficiently handling cast shadows is difficult for the learning-based approach; a huge amount of training data is required in general.

Recently, Ikehata [7] and Li *et al.* [9] propose methods for efficiently handling cast shadows in the learning-based approach by incorporating cast shadows into so-called observation map. Similar to the conventional approach, their objective is to remove the impact of cast shadows from surface normal estimation implicitly. On the other hand, our focus is on illuminating an object surface so that sufficient information is available for surface normal estimation.

Illumination Planning

Drbohlav and Chantler [6] address the optimal light source configurations for the Lambertian photometric stereo. Specifically, they reveal the relationship between the noises in pixel values and the expected errors of estimated surface normals on the basis of noise propagation analysis, and derive the optimal light source configurations for given numbers of light sources. Unfortunately, however, their analysis does not take shadows into consideration and finds the optimal light source directions *a priori* independent of the shape of an object.

In contrast to their method, we take account of shadows for optimizing light source directions in the Lambertian photometric stereo. In particular, our method optimizes the light source directions adaptively for an object of interest on the basis of the previously captured images of the object in an online manner, because cast shadows are global effects depending on the entire shapes of the objects.

3 Proposed Method

3.1 Overview

Our proposed method aims to illuminate an object surface from the optimal light source directions so that each pixel on the surface is *sufficiently* illuminated. To this end, our method iteratively finds the optimal light source direction on the basis of the previously captured images of the object, and then capture a novel image of the object under the optimal light source direction. The pipeline of our method is as follows; our proposed illumination planning

1. randomly selects three initial light source directions, and then captures three images of the object under each of those light source directions,
2. finds out **the pixel with the worst accuracy** in surface normal estimation,
3. finds out **the optimal light source direction** for improving the accuracy of surface normal estimation at the worst pixel, and then captures a novel image under the optimal light source direction,
4. repeats 2 and 3 a number of times,

and finally estimates surface normals from all of the captured images.

In Section 3.2, we describe how to estimate surface normals from images with shadows. Then, we explain the way of finding the worst pixel and the optimal light source direction in Sections 3.3 and 3.4 respectively.

3.2 Estimating Surface Normals from Images with Shadows

Here, we explicitly formulate the Lambertian photometric stereo from images with shadows. Let us denote the pixel value at the p-th pixel ($p = 1, 2, 3, ..., P$) on an object surface under the l-th light source direction ($l = 1, 2, 3, ..., L$) by i_{pl}. According to the Lambert model, the pixel value i_{pl} is represented by

$$i_{pl} = v_{pl} \boldsymbol{s}_l'^{\top} \boldsymbol{n}_p'. \tag{1}$$

Here, $\boldsymbol{s}_l'$ is the product of the direction $\boldsymbol{s}_l$ and intensity of the l-th light source, and $\boldsymbol{n}_p'$ is the product of the surface normal $\boldsymbol{n}_p$ and albedo at the p-th pixel. Both $\boldsymbol{s}_l$ and $\boldsymbol{n}_p$ are unit vectors. We denote the visibility of the p-th pixel from the l-th light source direction by v_{pl}; $v_{pl} = 0$ if the p-th pixel is in attached shadow or cast shadow and $v_{pl} = 1$ otherwise[2].

We can rewrite eq.(1) at the p-th surface point under the L different light source directions by using matrices as

$$\begin{pmatrix} i_{p1} \\ i_{p2} \\ \vdots \\ i_{pL} \end{pmatrix} = \begin{pmatrix} v_{p1} \boldsymbol{s}_1'^{\top} \\ v_{p2} \boldsymbol{s}_2'^{\top} \\ \vdots \\ v_{pL} \boldsymbol{s}_L'^{\top} \end{pmatrix} \boldsymbol{n}_p' = \begin{pmatrix} v_{p1} & 0 & \cdots & 0 \\ 0 & v_{p2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & v_{pL} \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1'^{\top} \\ \boldsymbol{s}_2'^{\top} \\ \vdots \\ \boldsymbol{s}_L'^{\top} \end{pmatrix} \boldsymbol{n}_p', \tag{2}$$

$$\boldsymbol{i}_{pL} = \boldsymbol{V}_{pL} \boldsymbol{S}_L \boldsymbol{n}_p'. \tag{3}$$

Therefore, we can compute the scaled surface normal $\boldsymbol{n}_p'$ by using the pseudo-inverse matrix $(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^{\dagger}$ as

$$\boldsymbol{n}_p' = (\boldsymbol{V}_{pL}\boldsymbol{S}_L)^{\dagger} \boldsymbol{i}_{pL} = \{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^{\top}(\boldsymbol{V}_{pL}\boldsymbol{S}_L)\}^{-1}(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^{\top} \boldsymbol{i}_{pL}, \tag{4}$$

if $\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^{\top}(\boldsymbol{V}_{pL}\boldsymbol{S}_L)\}$ is a regular matrix. Note that the surface normal $\boldsymbol{n}_p$ with unit length is given by $\boldsymbol{n}_p = \boldsymbol{n}_p'/|\boldsymbol{n}_p'|$. When $\boldsymbol{V}_{pL}$ is the $L \times L$ identity matrix, the estimation of the surface normal in eq.(4) results in the estimation from images without shadows.

[2] The visibility is often used for representing cast shadows, but we use it for representing both attached shadows and cast shadows in this paper.

3.3 Finding Pixel with Worst Accuracy

In a similar manner to Drbohlav and Chantler [6], we can estimate the accuracy of the estimated surface normal via noise propagation analysis. Let us denote the estimated and ground truth scaled surface normals by $\boldsymbol{n}'_p$ and $\bar{\boldsymbol{n}}'_p$ respectively. Then, we study the variance-covariance matrix $\boldsymbol{\Sigma}$ of the surface normal defined by

$$\boldsymbol{\Sigma} = E\left[(\boldsymbol{n}'_p - \bar{\boldsymbol{n}}'_p)(\boldsymbol{n}'_p - \bar{\boldsymbol{n}}'_p)^\top\right], \tag{5}$$

where $E[\ ]$ stands for the expectation value.

Suppose that the observed pixel value $\tilde{i}_{pl}$ is contaminated by additive noise δ_{pl} and deviates from the ideal pixel value i_{pl}[3] as

$$\tilde{i}_{pl} = i_{pl} + \delta_{pl}, \tag{6}$$

we can derive

$$\begin{aligned}\boldsymbol{\Sigma} &= E\left[\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\dagger\boldsymbol{\delta}_{pL}\}\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\dagger\boldsymbol{\delta}_{pL})\}^\top\right] \\ &= \sigma^2(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\dagger\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\dagger\}^\top \\ &= \sigma^2\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\top(\boldsymbol{V}_{pL}\boldsymbol{S}_L)\}^{-\top}.\end{aligned} \tag{7}$$

Here, $\boldsymbol{\delta}_{pL} = (\delta_{p1}, \delta_{p2}, \delta_{p3}, \cdots, \delta_{pL})^\top$, and we assume independent and identically distributed noises whose mean and variance are 0 and σ^2 respectively. Since the Mean Square Error (MSE) is proportional to the trace of the variance-covariance matrix, we obtain the MSE of the estimated surface normal as

$$\mathrm{MSE} \propto \mathrm{Tr}\left[\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\top(\boldsymbol{V}_{pL}\boldsymbol{S}_L)\}^{-1}\right]. \tag{8}$$

Therefore, our proposed method finds out the pixel $\hat{p}$ with the worst accuracy in surface normal estimation as

$$\hat{p} = \arg\max_p \mathrm{Tr}\left[\{(\boldsymbol{V}_{pL}\boldsymbol{S}_L)^\top(\boldsymbol{V}_{pL}\boldsymbol{S}_L)\}^{-1}\right]. \tag{9}$$

Note that the MSE depends on the surface points via the visibility matrix $\boldsymbol{V}_{pL}$[4]. Intuitively, the MSE is larger for the pixel that is illuminated by a smaller number of light sources with smaller linear independence. When $\boldsymbol{V}_{pL}$ is the $L \times L$ identity matrix, *i.e.* when we do not take account of shadows, the MSE results in that of Drbohlav and Chantler [6].

3.4 Finding Optimal Light Source Direction

Suppose that the L images of an object are captured and the $\hat{p}$-th pixel is selected as the pixel with the worst accuracy according to eq.(9). Our proposed

[3] Note that not the noise δ_{pl} but the visibility v_{pl} is used for representing shadows as shown in eq.(1).

[4] When multiple pixels have the same visibility matrix and therefore have the same MSE, we randomly select one of the pixels as the worst pixel.

method finds the optimal light source direction, *i.e.* the $(L+1)$-th light source direction under which a novel image of the object is captured for improving the accuracy of the estimated surface normal at the $\hat{p}$-th pixel. Our clue to the optimal light source direction is the visibility and linear independence; the optimal light source direction should be visible from the $\hat{p}$-th pixel and have the largest linear independence from the previous light source directions with $v_{\hat{p}l} = 1$.

As mentioned in Section 1, our implementation makes use of an LCD as a programmable light source. Then, we represent the direction of a light source by the 2-D coordinate system of the display. We assume that our setup is geometrically calibrated in advance, and represent a light source by using the 2-D coordinate (x, y) of the display and the 3-D vector $\boldsymbol{s}(x, y)$ interchangeably. Our proposed method maximizes the cost function $C(x, y)$ defined as

$$C(x, y) = C_{\mathrm{vis}}(x, y) \times C_{\mathrm{lin}}(x, y) \tag{10}$$

with respect to (x, y). Here, $C_{\mathrm{vis}}(x, y)$ and $C_{\mathrm{lin}}(x, y)$ are the cost functions for the visibility and linear independence of the light source $\boldsymbol{s}(x, y)$ respectively.

The cost function for the visibility has larger value if the light source direction (x, y) is likely visible from the $\hat{p}$-th pixel and vice versa. More specifically, if the direction (x_l, y_l) of the l-th light source is visible from the $\hat{p}$-th pixel, its neighboring directions are also considered to be visible. On the other hand, if the direction of the l-th light source is invisible, its neighboring directions are also considered to be invisible. Therefore, we define the cost function for the visibility by the linear combination of the Gaussian distributions as

$$C_{\mathrm{vis}}(x, y) = \frac{1}{2\pi w^2}\left[e^{-\frac{(x-x_c)^2+(y-y_c)^2}{2w^2}} + \sum_{l=1}^{L}(2v_{\hat{p}l} - 1)e^{-\frac{(x-x_l)^2+(y-y_l)^2}{2w^2}}\right]. \tag{11}$$

Here, the first term corresponds to the direction of a camera (x_c, y_c), and the second term corresponds to the previous L light sources. The coefficient $(2v_{\hat{p}l} - 1) = 1$ for visible light sources from the $\hat{p}$-th pixel, and $(2v_{\hat{p}l} - 1) = -1$ for invisible light sources. Because we consider an object surface visible from the camera, the coefficient of the first term is 1. Since the cost function for the visibility is considered as a likelihood, we adopt the upper and lower limits; we set $C_{\mathrm{vis}}(x, y) = 1$ if $C_{\mathrm{vis}}(x, y) > 1$, and $C_{\mathrm{vis}}(x, y) = -1$ if $C_{\mathrm{vis}}(x, y) < -1$[5].

The cost function for the linear independence has larger value if the light source direction (x, y) has larger linear independence from the previous light source directions with $v_{\hat{p}l} = 1$. Therefore, we compute the principal directions of those previous light sources, and define the cost function for the linear independence as

$$C_{\mathrm{lin}}(x, y) = \max[\boldsymbol{s}_{\mathrm{min}}^{\top}\boldsymbol{s}(x, y), -\boldsymbol{s}_{\mathrm{min}}^{\top}\boldsymbol{s}(x, y)]. \tag{12}$$

Here, we denote the 3-D eigenvector corresponding to the smallest eigenvalue of $(\boldsymbol{V}_{\hat{p}L}\boldsymbol{S}_L)$ by $\boldsymbol{s}_{\mathrm{min}}$, and $\max[\ ]$ returns the maximum argument.

[5] We set the lower limit not to 0 but to 1 so that the gradient of the cost function $C(x, y)$ does not vanish for gradient-based optimization.

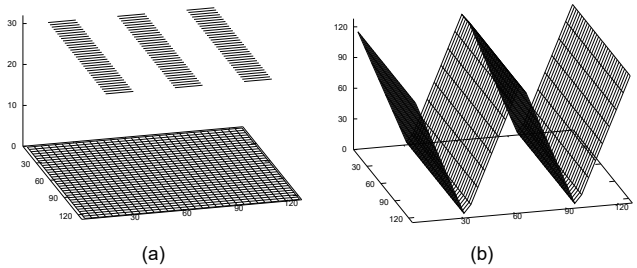

Fig. 1. Two objects used in the experiments with synthetic images: (a) a slit and (b) a wave. The camera and display are located above the objects.

Since the cost function in eq.(10) is nonlinear with respect to (x, y), the optimization depends on the initial conditions. In our current implementation, we used the interior point algorithm (fmincon in MATLAB), and tested multiple random initial conditions and then selected the best solution that maximizes the cost function.

4 Experiments

To confirm the effectiveness of our proposed method, we conducted a number of experiments using both synthetic and real images. In the experiments using synthetic images, we assumed a similar setup to our prototype system for the experiments using real images shown in Fig. 8. We assumed that the visibility v_{pl} is 0 if the pixel value i_{pl} is smaller than a threshold depending on the noise level of images, and v_{pl} is 1 otherwise.

We changed the standard deviation w of the Gaussian distributions in eq.(11) as $w = w_0/\sqrt{L}$ so that the cost function for the visibility can represent finer details as the number of light sources L increases. As shown in Fig. 8, we used an ultra wide display and normalized the scale of the coordinate system as the height of the display is 1, *i.e.* the range of x is $[0, 1]$, and we empirically set $w_0 = 0.7$.

4.1 Synthetic Images

Visibility

First, we show how the cost function for the visibility behaves actually. We used a slit shown in Fig. 1(a) as an object of interest. For the sake of simplicity, we investigated the visibility at a fixed point on the lower plane of the object, and the light source directions are randomly selected instead of using our illumination planning.

Fig. 2 visualizes the cost function for the visibility in the display domain (x, y): (a) the ground truth and the visibilities when the numbers of light sources

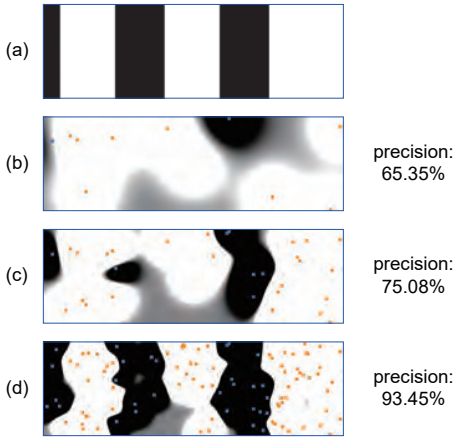

Fig. 2. The cost function for the visibility in the display domain (x, y): (a) the ground truth and the visibilities when the numbers of light sources L are (b) 10, (c) 30, and (d) 100 respectively.

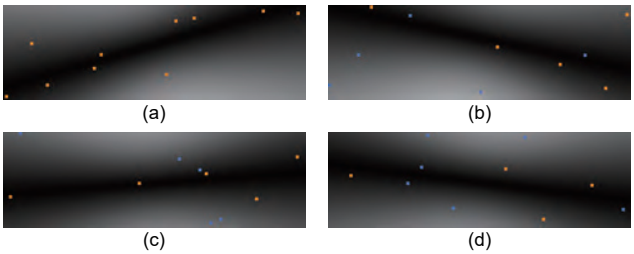

Fig. 3. The cost function for the linear independence for different sets of 10 light source directions.

L are (b) 10, (c) 30, and (d) 100 respectively. Here, the visibility v_{pl} is 1/0 at white/black pixels in (a) the ground truth, and C_{vis} is 1/-1 at white/black pixels from (b) to (d). The selected light source directions are superimposed in those images: orange/light blue points correspond to $v_{pl} = 1/0$ respectively. We can see that the cost function for the visibility gets closer to the ground truth as the number of light sources increases. In addition, we binarize the cost function for the visibility to 1/0 when $C_{\text{vis}} \geq 0/ < 0$, and computed the precision of the visibility, *i.e.* the fraction of the light source directions with $v_{pl} = 1$ among those with $C_{\text{vis}} \geq 0$, when L =10, 30, and 100. We can see quantitatively that our cost function for the visibility works well; the precision gets closer to 100% as the number of light sources increases.

Linear Independence

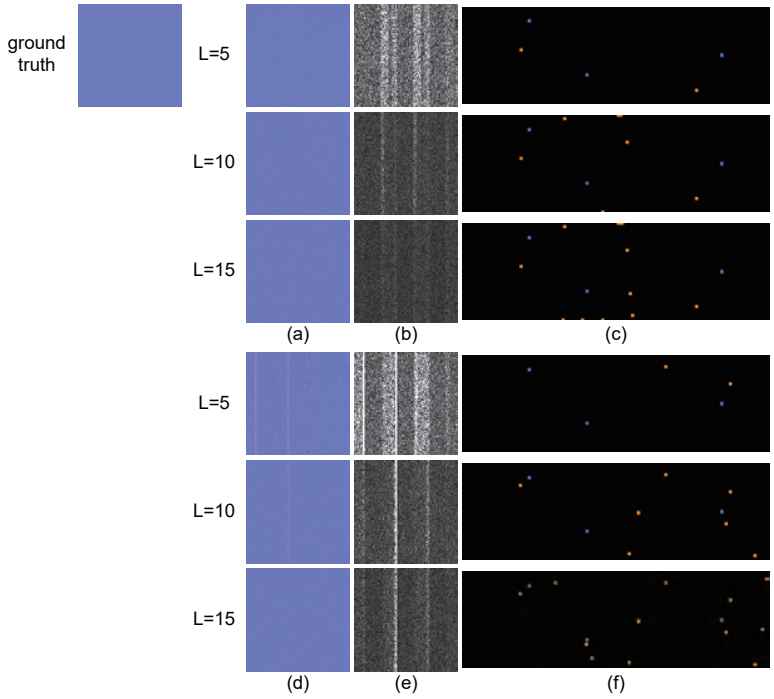

Fig. 4. The qualitative results of the slit when $\sigma = 0.02$: the ground truth surface normals and (a) the estimated surface normals, (b) the errors of the surface normals, and (c) the selected light sources by using our proposed illumination planning, and (d) (e) (f) those of the random selection when $L = 5$, 10, and 15.

Second, we show how the cost function for the linear independence behaves actually. For the sake of simplicity, we investigated the linear independence by using dummy data. More specifically, we randomly selected 10 light source directions and assumed that a surface point of attention is shadowed under some of them.

Fig. 3 visualizes the cost function for the linear independence for different sets of 10 light source directions. Here, the linear independence C_{lin} is 1/0 at white/black pixels in the images. The selected light source directions are superimposed in those images: orange/light blue points correspond to $v_{pl} = 1/0$ respectively. We can see that the cost function for the linear independence takes larger values at the orthogonal directions to the visible light source directions illuminating the surface point of attention as expected.

Surface Normal Estimation

Third, we compare the performance of our proposed illumination planning with the random selection of light source directions for surface normal estimation. We tested two objects: a slit and a wave in Fig. 1. The camera and display are

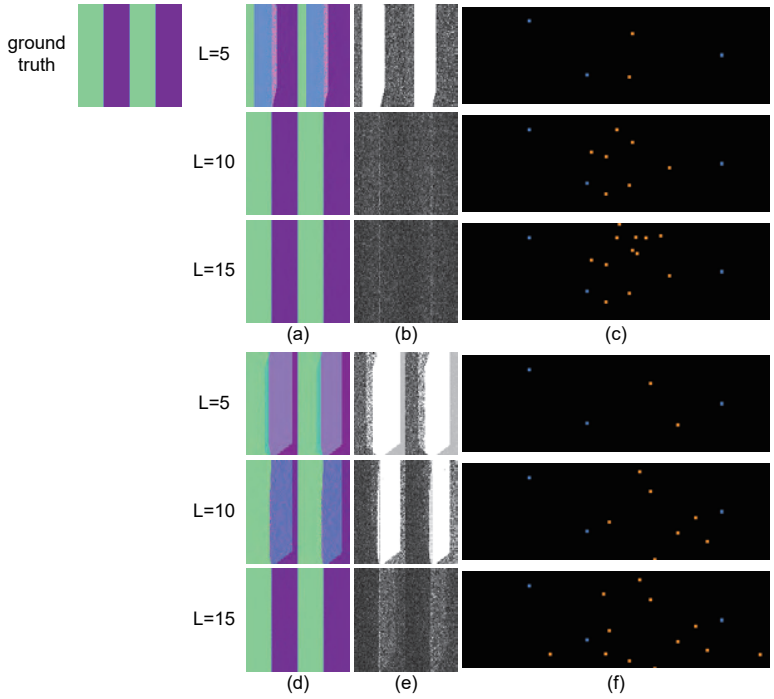

Fig. 5. The qualitative results of the wave when $\sigma = 0.02$: the ground truth surface normals and (a) the estimated surface normals, (b) the errors of the surface normals, and (c) the selected light sources by using our proposed illumination planning, and (d) (e) (f) those of the random selection when $L = 5$, 10, and 15.

located above the objects. In order to investigate the robustness to noises in pixel values, we added zero-mean Gaussian noises, whose standard deviations σ are 0.01, 0.02, and 0.04 for pixel values normalized to $[0, 1]$, to the synthetic images.

Fig. 4 shows the qualitative results of the slit when $\sigma = 0.02$: the ground truth surface normals and (a) the estimated surface normals, (b) the errors of the surface normals, and (c) the selected light sources by using our proposed illumination planning, and (d) (e) (f) those of the random selection when $L = 5$, 10, and 15. The whiter the pixels in (b) and (e) are, the larger the errors in surface normal estimation are. We can see that our illumination planning works better than the random selection; our illumination planning achieves better accuracy by using smaller number of light sources than the random selection. We can also see that the light source directions selected by our illumination planning are different from those of the random selection. Here, the light blue points stand for the initial three light source directions, and the orange points stand for the iteratively added light source directions.

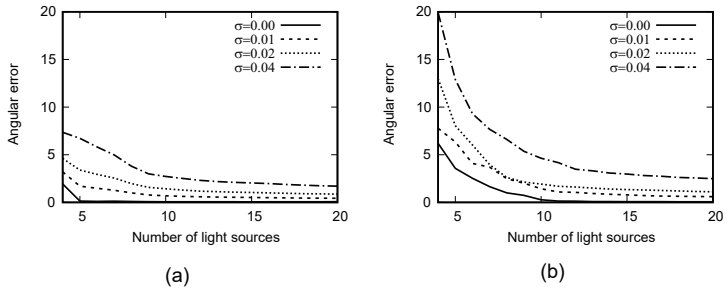

Fig. 6. The quantitative results of the slit; the angular error in degree vs. the number of light sources under various noise levels: (a) our proposed illumination planing and (b) the random selection of light source directions.

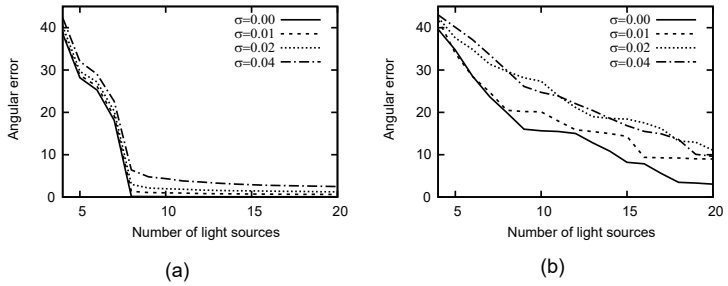

Fig. 7. The quantitative results of the wave; the angular error in degree vs. the number of light sources under various noise levels: (a) our proposed illumination planing and (b) the random selection of light source directions.

Fig. 6 shows the quantitative results of the slit; the angular error in degree vs. the number of light sources under various noise levels: (a) our proposed illumination planing and (b) the random selection of light source directions. We can see that the angular errors of our illumination planning converges to smaller values for smaller number of light sources than the random selection. Those results quantitatively show that our illumination planning works better than the random selection.

The qualitative and quantitative results of the wave are shown in Fig. 5 and Fig. 7 respectively. We can see that we obtain the similar results to those of the slit. Those results for the wave also support that our proposed illumination planning works better than the random selection. It is interesting that the light source directions selected by our illumination planning for the slit and the wave are different from each other. This is because our method finds out the optimal light source directions depending on the shape of an object of interest. Actually,

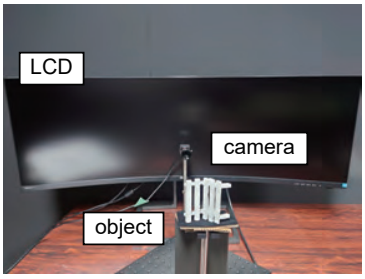

Fig. 8. Our prototype system for the experiments using real images.

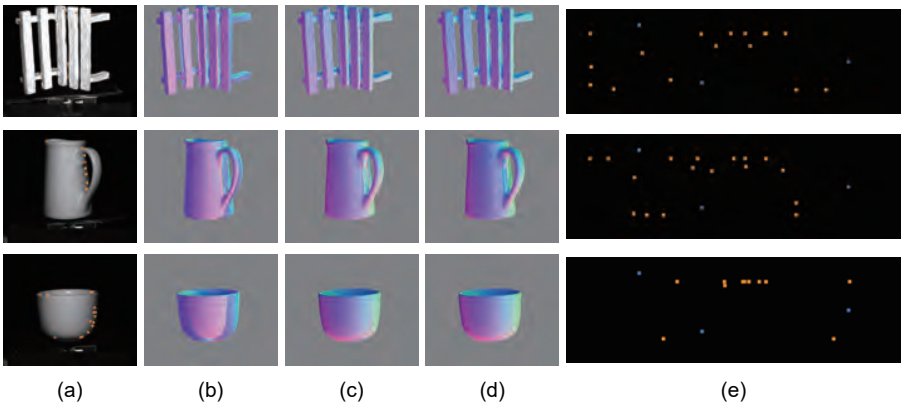

Fig. 9. The qualitative results using real images: (a) the images of target objects under ambient light, the estimated surface normals when (b) $L = 3$, (c) 10, and (d) 20, and (e) the selected optimal light source directions: light blue points ($L = 1$ to 3) and orange points ($L = 4$ to 20).

the optimal light source directions for the wave mainly distribute around the center column of the display in order to illuminate the deep valleys of the wave.

4.2 Real Images

Fig. 8 shows our prototype system for the experiments using real images. It consists of an ultra wide LCD 439P9H1/11 from PHILIPS and a polarimetric camera BFS-U3-51S5P-C from FLIR. We assume sufficiently smaller objects than the distance between the display and the objects, and consider the small white spots displayed on the LCD as distant light sources. The directions and intensities of those light sources are calibrated in advance by using the images of a mirror sphere and a diffuse reflector.

Our proposed online illumination planning, consisting of controlling the LCD and camera, separating reflection components, finding the worst pixel, and finding the optimal light source, is implemented by using MATLAB. We separated the diffuse and specular reflection components in the captured images on the basis of polarization, and used the former for the Lambertian photometric stereo. The optimization of the cost function in eq.(10) takes about 0.05 sec for a single initial condition. Therefore, we can test 80 initial conditions during typical exposure time of 4 sec.

We qualitatively show how our proposed illumination planning behaves for real images, since the ground truth surface normals are unknown for real objects. We tested three objects; a toy bench made of plastics, a milk pitcher and a cup made of ceramic. Fig. 9 shows (a) the images of those objects under ambient light, the estimated surface normals when (b) $L = 3$, (c) 10, and (d) 20, and (e) the selected optimal light source directions: light blue points ($L = 1$ to 3) and orange points ($L = 4$ to 20). Here, the orange points in (a) the images of the objects stand for the selected worst pixels from $L = 4$ to 20.

The estimated surface normals qualitatively show that the accuracy significantly improves as the number of light sources increases from (b) $L = 3$ to (d) 20. Interestingly, we can see that (a) the worst pixels are often selected from non-convex areas, and (e) the optimal light source directions are selected so that they can illuminate those pixels. Those results demonstrate that our proposed illumination planning behaves as expected, and estimates surface normals from a small number of images.

5 Conclusion and Future Work

We proposed an illumination planning for shadow-robust Lambertian photometric stereo. Our proposed illumination planning takes the visibility and linear independence of light source directions into consideration, and optimizes the light source directions adaptively for an object of interest. We implemented our illumination planning with a programmable light source in an online manner, and achieve shadow-robust surface normal estimation from a small number of images.

One direction of our future study is to incorporate interreflections into our proposed illumination planning, because they are often observed on non-convex surfaces but our illumination planning does not take account of them. The other direction is the integration with view planning; it optimizes both the light source directions and the viewing directions for modeling an object of interest.

Acknowledgement: This work was supported by JSPS KAKENHI Grant Number JP20H00612.

References

1. Ackermann, J., Michael, G.: A Survey of Photometric Stereo Techniques. Now Foundations and Trends (2015)

2. Barsky, S., Petrou, M.: The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. IEEE Trans. PAMI **25**(10), 1239–1252 (2003)
3. Chandraker, M., Agarwal, S., Kriegman, D.: ShadowCuts: Photometric stereo with shadows. In: Proc. IEEE CVPR2007. pp. 1–8 (2007)
4. Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.Y.K.: Self-calibrating deep photometric stereo networks. In: Proc. IEEE CVPR2019. pp. 8739–8747 (2019)
5. Chen, G., Han, K., Wong, K.Y.K.: PS-FCN: A flexible learning framework for photometric stereo. In: Proc. ECCV2018. pp. 3–19 (2018)
6. Drbohlav, O., Chantler, M.: On optimal light configurations in photometric stereo. In: Proc. IEEE ICCV2005. vol. 2, pp. 1707–1712 (2005)
7. Ikehata, S.: CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In: Proc. ECCV2018. pp. 3–18 (2018)
8. Ikehata, S., Wipf, D., Matsushita, Y., Aizawa, K.: Robust photometric stereo using sparse regression. In: Proc. IEEE CVPR2012. pp. 318–325 (2012)
9. Li, J., Robles-Kelly, A., You, S., Matsushita, Y.: Learning to minify photometric stereo. In: IEEE CVPR2019. pp. 7568–7576 (2019)
10. Mukaigawa, Y., Ishii, Y., Shakunaga, T.: Analysis of photometric factors based on photometric linearization. JOSA A **24**(10), 3326–3334 (2007)
11. Murase, H., Nayar, S.K.: Illumination planning for object recognition using parametric eigenspaces. IEEE Trans. PAMI **16**(12), 1219–1227 (1994)
12. Okabe, T., Sato, I., Sato, Y.: Attached shadow coding: Estimating surface normals from shadows under unknown reflectance and lighting conditions. In: Proc. IEEE ICCV2009. pp. 1693–1700 (2009)
13. Okabe, T., Sato, Y.: Object recognition based on photometric alignment using RANSAC. In: Proc. IEEE CVPR2003. pp. 221–228 (2003)
14. Shi, B., Mo, Z., Wu, Z., Duan, D., Yeung, S.K., Tan, P.: A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. In: Proc. IEEE CVPR2016. pp. 3707–3716 (2016)
15. Silver, W.M.: Determining shape and reflectance using multiple images. Master's thesis, MIT (1980)
16. Sunkavalli, K., Zickler, T., Pfister, H.: Visibility subspaces: uncalibrated photometric stereo with shadows. In: Proc. ECCV2010. pp. 251–264 (2010)
17. Taniai, T., Maehara, T.: Neural inverse rendering for general reflectance photometric stereo. In: Proc. ICML2018. pp. 4857–4866 (2018)
18. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering **19**(1), 139–144 (1980)
19. Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., Ma, Y.: Robust photometric stereo via low-rank matrix completion and recovery. In: Proc. ACCV2010. pp. 703–717 (2010)
20. Zheng, Q., Jia, Y., Shi, B., Jiang, X., Duan, L.Y., Kot, A.C.: SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In: Proc. IEEE ICCV2019. pp. 8549–8558 (2019)