Article

Color Photometric Stereo using Multi-band Camera constrained by Median Filter and Occluding Boundary

Daisuke Miyazaki [1*], **Yuka Onishi** [1,†] **and Shinsaku Hiura** [2]

[1] Hiroshima City University
[2] University of Hyogo
* Correspondence: miyazaki@hiroshima-cu.ac.jp
† Current address: gumi Inc.

Version July 1, 2019 submitted to J. Imaging

Abstract: One of the main problems faced by the photometric stereo method is that several measurements are required, as this method needs illumination from light sources from different directions. A solution to this problem is the color photometric stereo method, which conducts one-shot measurements by simultaneously illuminating lights of different wavelengths. However, the classic color photometric stereo method only allows measurements of white objects, while a surface-normal estimation of a multicolored object using this method is theoretically impossible. Therefore, it is necessary to add some constraints to estimate the surface normal of a multicolored object using the framework of the color photometric stereo method. In this study, a median filter is employed as the constraint condition of albedo, and the surface normal of occluding boundary is employed as the constraint condition of surface normal. By employing a median filter as the constraint condition, the smooth distribution of the albedo and normal is calculated while the sharp features at the boundary of different albedos and normals are preserved. The surface normal at occluding boundary is propagated into the inner part of object region, and forms the abstract shape of the object. Such surface normal gives a great clue to be used as an initial guess to the surface normal. To demonstrate the effectiveness of this study, a measurement device that can realize the multispectral photometric stereo method with seven colors is employed instead of the classic color photometric stereo method with three colors.

Keywords: photometric stereo; color photometric stereo; multispectral imaging

1. Introduction

To reproduce a detailed surface shape, normal information is necessary. To obtain this information, the photometric stereo method was proposed, which estimates the normal by transitioning the brightness levels of several pictures by changing the direction of the light source. However, as it requires multiple photoshoots, the photometric stereo method is not suitable for modeling a moving object. To measure the shape of a moving object, the color photometric stereo method, which employs several colored light sources, was developed. This method involves placing light sources of red, green, and blue colors in three different directions, which simultaneously illuminate the target object. This paper proposes a technique that employs some constraints so that it can be applied to colored objects, which is impossible for conventional color photometric stereo. Unlike the common color photometric stereo method, we use seven narrow-band lights with different peak wavelengths while observing the target object with a seven-band multispectral camera.

2. Related work

The photometric stereo method [31,35] estimates the normal of the surface of an object by illuminating the object and analyzing the resulting shadings on the object's surface. In this method, light is illuminated on the object from one white parallel light source (an infinity point light source) to obtain a picture. Then, two more pictures are captured with different light source directions. In other words, it requires capturing three pictures with different light source directions. Therefore, it is impossible to measure a dynamic object. This problem can be resolved using the color photometric stereo method. In this method, lights are simultaneously illuminated from red, green, and blue light sources, and one picture photographed with an RGB color camera is captured. Such one-shot photograph enables the measurement of a dynamic object.

The color photometric stereo method [9,21,36] (also known as shape-from-color) was developed in the 1990s. Since then, various studies [1,4–8,12,14,15,18,20–22,25,29,30,33,34] have been conducted in this regard. However, many problems are inherent in the color photometric stereo method. Many researchers in the past have struggled with this method, and even till recently, it has been an ongoing problem. The principle problem of the color photometric stereo method is the fact that it can only be used with white objects. This is an inevitable problem as long as lights are illuminated from three colored light sources to estimate the surface normal.

Recently, various techniques have been proposed to apply the color photometric stereo method to multicolored objects. Roubtsova et al. [30] applied the color photometric stereo method to objects with arbitrary BRDF (bidirectional reflectance distribution function) by incorporating the Helmholtz Stereo method. However, the principle of this method does not allow for real-time measurement. Therefore, an optical flow is required to measure a dynamic object. Kim et al. [20] and Gotardo et al. [13] also tracked dynamic objects using optical flow, and estimated the surface shape of objects by utilizing several images taken at different times. Fyffe et al. [12] proposed a color photometric stereo method that employs six band cameras and three white color sources. All three light sources used in their method appear white to the human eye. However, all of them possess different spectral distributions. Furthermore, this method pre-measures the reflectance of various objects to prepare a database, and calculated four bases. Using this technique, it is possible to obtain an analytic solution, as there are four unknown numbers in relation to albedo (four base coefficients) and two in relation to the normal (because the three-dimensional vector is normalized), and six equations are obtainable. Anderson et al. [1] estimated the object color using the normal of multi-view stereo. However, owing to the low accuracy of the normal of multi-view stereo, they improved the estimation accuracy of object color based on the hypothesis that an object is composed of a limited number of colors. Their technique incorporates the framework of region segmentation, where the number of the regions is automatically determined based on the Bayesian information criterion. Chakrabarti et al. [5] calculated the candidates of object color by approximating the shape inside the patch of neighboring areas using a polynomial. They calculated the histogram of the object color candidates, chose only the limited number of colors that gained most votes, and evaluated the normal by postulating that the object is composed of these limited number of colors. Jiao et al. [18] divided a picture into super pixel regions and estimated the normal by postulating that the object color inside each region is uniform.

In this paper, the problem faced by the color photometric stereo method is solved using a different approach from those used in previous studies. Our proposed technique employs a median filter as the constraint condition of the albedo and surface normal. We also use occluding boundary constraint for surface normal. Thanks to this constraint, we have a good estimate from the initial state of surface normal, which results in robust estimation.

The techniques of Gotardo et al. [13], Kim et al. [20], and Roubtsova et al. [30] need to employ optical flow to measure a dynamic object, while the technique of Fyffe et al. [12] requires a reflectance database to be prepared prior to the measurement. Our proposed technique does not require a shape obtained from other sensors such as multi-view stereo or laser sensor, unlike the technique of Anderson et al. [1] Moreover, unlike the techniques of Chakrabarti et al. [5] and Jiao et al., [18] our proposed method does not require region segmentation. Previous color photometric stereo methods used three lights with red, green, and blue colors and observed the object with an RGB color camera. In our study, seven lights with different wavelengths are used to illuminate the object, which is then observed by a seven-band multispectral camera. This paper demonstrates that multi-spectral cameras and multi-spectral light sources are also effective for the color photometric stereo method.

3. Multispectral color photometric stereo method

3.1. Image formulation

A photometric stereo method that employs independent colored light is called the color photometric stereo method. A characteristic of this method is that it enables the estimation of the surface normal with one photoshoot. The widespread color photometric stereo method is conducted with three types of colored lights. While the conventional photometric stereo method results in several grayscale images, the color photometric stereo method results in a multi-spectral image.

Although the fundamental theory is given in several number of literatures [23,26], we briefly explain the formulation of the problem. The spectral sensitivity of a camera is denoted as $Q_c(\lambda)$, the spectral distribution of the light source is $E(\lambda)$, and the spectral reflectance of the object is $S(\lambda)$. Moreover, c denotes the channel. In this case, the brightness obtained from each channel of the camera can be attained from Equation (1).

$$I_c = \int_0^\infty Q_c(\lambda)E(\lambda)S(\lambda)d\lambda \quad . \tag{1}$$

Suppose that we use single light $E(\lambda)$ whose spectral distribution can be represented as a delta function $\delta(\cdot)$ whose peak wavelength is λ_c.

$$E(\lambda) = e_c\delta(\lambda - \lambda_c)\,, \tag{2}$$

where e_c represents the brightness of the light. Suppose that the channel c is only sensitive to the wavelength λ_c, and suppose that other channels cannot detect the wavelength λ_c.

$$Q_c(\lambda)E(\lambda) = q_c e_c \delta(\lambda - \lambda_c)\,, \tag{3}$$

where q_c represents the sensitivity at wavelength λ_c. Suppose that we lit a single parallel light source (infinite-far point light source) whose spectral distribution is represented as delta function and its peak wavelength is λ_c, the pixel brightness I_c can be represented as follows using the formulation that the diffuse reflection is represented as $S(\lambda_c) = \tilde{s}_c \max(\mathbf{n} \cdot \mathbf{l}_c, 0)$.

$$I_c = q_c e_c \tilde{s}_c \max(\mathbf{n} \cdot \mathbf{l}_c, 0)\,, \tag{4}$$

where $\tilde{s}_c$ represents the reflectance. $\mathbf{n}$ is a normal vector and $\mathbf{l}_c$ is the light source direction vector of channel c. Denoting as $A_c = q_c e_c \tilde{s}_c$, Equation (4) becomes as follows.

$$I_c = A_c \max(\mathbf{n} \cdot \mathbf{l}_c, 0) \quad . \tag{5}$$

Hereinafter, we call A_c albedo. Note that the camera sensitivity and light source brightness are included in A_c.

As shown in Fig. 1, this study conducts a photoshoot of a multicolored object using seven channels (Fig. 2). Following Equation (5), the brightness is obtained from this photoshoot as follows.

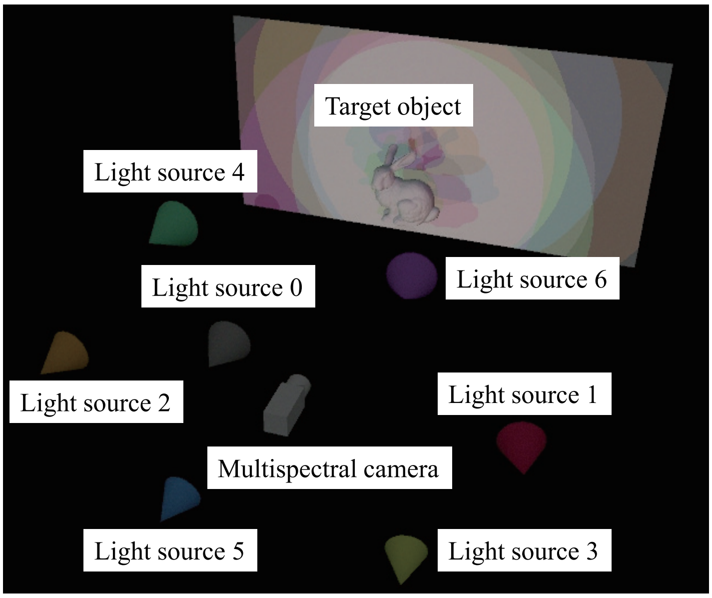

Figure 1. Conceptual explanation of multispectral color photometric stereo. Target object is illuminated by multiple light sources whose wavelengths are different. One image is taken using multispectral camera.

$$\begin{aligned} I_0 &= A_0 \max(\mathbf{n} \cdot \mathbf{l}_0, 0) \quad , \\ I_1 &= A_1 \max(\mathbf{n} \cdot \mathbf{l}_1, 0) \quad , \\ &\vdots \\ I_6 &= A_6 \max(\mathbf{n} \cdot \mathbf{l}_6, 0) \quad . \end{aligned} \tag{6}$$

The surface normal $\mathbf{n}$ is a 3D vector; however, the degree-of-freedom is two because it is constrained to be a unit vector (such constraint reduces one degree-of-freedom). Albedo A_c is represented by seven parameters. There are seven equations, as shown in Equation (6), and nine unknown parameters ($A_0, A_1, \ldots, A_6, n_x, n_y, n_z$, s.t., $n_x^2 + n_y^2 + n_z^2 = 1$, namely seven for albedo and two for surface normal). Therefore, color photometric stereo, without any assumption or constraint, is an ill-posed problem.

The most commonly used assumption is to limit the color of the target objects to white ($A_0 = A_1 = \cdots = A_6$). If we set $\mathbf{s} = A_c\mathbf{n}$ and if we ignore the shadow, the surface normal $\mathbf{s}$ (scaled with albedo) can be directly solved.

$$\begin{pmatrix} \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{l}_0^\top \\ \mathbf{l}_1^\top \\ \vdots \\ \mathbf{l}_6^\top \end{pmatrix}^+ \begin{pmatrix} I_0 \\ I_1 \\ \vdots \\ I_6 \end{pmatrix} . \tag{7}$$

As is shown above, the color photometric stereo for white objects, or in other words, the conventional photometric stereo can directly solve the surface normal, without iterative optimization nor additional constraints such as smoothness constraints. However, this paper analyzes the methods with multi-colored objects. Therefore, we add smoothness constraints and iteratively solved the problem formulated as Equation (6).

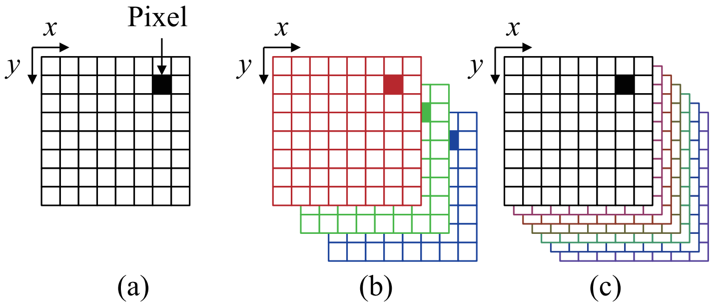

Figure 2. Explanation of multi-channel image: (a) Grayscale image with single channel, (b) RGB color image with 3 channels, and (c) multispectral image with 7 channels.

The proposed technique estimates the surface normal through an iteration process. The cost function that is minimized through the iteration process is explained in Section 3.2. Each term of the cost function is explained in Sections 3.3, 3.4, 3.5, and 3.6. The initial value required in the iteration process is explained in Section 3.6 and Section 3.7, and the update rule for each iteration is shown in Section 3.8. Detection of specular reflection is explained in Section 3.9. A method to integrate the surface normal to obtain the geometrical structure of the object surface is shown in Section 3.10, and Section 3.11 explains how to cancel the channel crosstalk.

3.2. Cost function

The cost function $\iint F dxdy$ is expressed through the following four terms:

$$
\begin{aligned}
F = & \iint_{(x,y)\in\mathcal{P}\backslash\partial\mathcal{P}} F_1(\mathbf{n}(x,y), \mathbf{A}(x,y), \mathbf{I}(x,y), \mathbf{L}) dxdy \\
& + \iint_{(x,y)\in\mathcal{P}\backslash\partial\mathcal{P}} F_2(\mathbf{n}(x,y)) dxdy \\
& + K_2 \iint_{(x,y)\in\mathcal{P}\backslash\partial\mathcal{P}} F_3(\mathbf{A}(x,y)) dxdy \\
& + \iint_{(x,y)\in\partial\mathcal{P}} F_4(\mathbf{n}(x,y)) dxdy \quad .
\end{aligned} \tag{8}
$$

Equation (8) is minimized under the condition that surface normal $\mathbf{n}$ should be an unit vector, $\|\mathbf{n}\| = 1$. Here, $\mathbf{A} = (A_0(x,y), A_1(x,y), \cdots, A_6(x,y))^\top$, $\mathbf{L} = (\mathbf{l}_0, \mathbf{l}_1, \cdots, \mathbf{l}_6)^\top$, and $\mathbf{I} = (I_0(x,y), I_1(x,y), \cdots, I_6(x,y))^\mathrm{T}$. K_2 is the Lagrange multiplier. The area where the target object is observed is denoted as $\mathcal{P}$, and the occluding boundary is denoted as $\partial\mathcal{P}$. The first three terms F_1, F_2, and F_3 are the soft constraints defined inside the object region $\mathcal{P}\backslash\partial\mathcal{P}$, and the fourth term F_4 are the hard constraint defined at the occluding boundary $\partial\mathcal{P}$. Orthographic projection is assumed in this paper for camera model.

Following are the four terms of cost functions, where K_{11} and K_{12} are the Lagrange multipliers.

$$
F_1 = \sum_{c=0}^{6} (I_c(x,y) - A_c(x,y) \max(\mathbf{l}_c^\mathrm{T} \mathbf{n}(x,y), 0))^2 \,, \tag{9}
$$

$$
F_2 = K_{11} \left(\left\| \frac{\partial \mathbf{n}(x,y)}{\partial x} \right\|^2 + \left\| \frac{\partial \mathbf{n}(x,y)}{\partial y} \right\|^2 \right) + K_{12} \left(\left\| \frac{\partial \mathbf{n}(x,y)}{\partial x} \right\| + \left\| \frac{\partial \mathbf{n}(x,y)}{\partial y} \right\| \right) \,, \tag{10}
$$

$$
F_3 = \left\| \frac{\partial \mathbf{A}(x,y)}{\partial x} \right\| + \left\| \frac{\partial \mathbf{A}(x,y)}{\partial y} \right\| \,, \tag{11}
$$

$$
F_4 = ||\mathbf{n}(x,y) - \mathbf{n}_b(x,y)||^2 \,. \tag{12}
$$

Sections 3.3, 3.4, 3.5, and 3.6 explain F_1, F_2, F_3, and F_4, respectively. F_1 expresses the residual of Lambertian reflectance and the input image brightness. I is the input image brightness, A is the albedo, $\mathbf{l}$ is the light source direction, and $\mathbf{n}$ is the surface normal. Here, c is the index that identifies the channel, and $\max(\mathbf{l}^\top \mathbf{n}, 0)$ represents the shading. F_2 is the smoothing term of the surface normal, and indicates that the gradient of the surface normal should be small; F_3 is the smoothing term of albedo, and indicates that the gradient of albedo should be small; and F_4 is the constraint condition of the surface normal at the occluding boundary. The surface normal at the occluding boundary $\mathbf{n}_b$ can be derived from differential geometry. F_4 indicates that the surface normal at the occluding boundary should be equal to $\mathbf{n}_b$. The reason why F_2 use both L1 norm and L2 norm is discussed in Section 3.4.

As we will explain in Sections 3.3, 3.4, 3.5, and 3.6, we do not minimize Equation (8) at once but minimize F_1, F_2, F_3, and F_4 separately. Although we cannot mathematically prove that such piecewise minimization results in global minimum, it is empirically known that piecewise minimization make the cost function smaller through the iteration. Since Equation (8) is a non-linear equation with several number of constraints, convergence speed is low. On the other hand, our approach is robust, stable, and speedy since we can minimize the cost function with closed form solution as is shown in Sections 3.3 and 3.6 (F_1 and F_4) and minimizing it with straightforward filtering as is shown in Sections 3.4 and 3.5.

Section 3.3 explains that F_1 solely cannot solve the problem. In order to solve the problem, we have to add F_2 or F_3 as it will be explained in Section 3.4 and Section 3.5. The surface normal will be smooth if we add F_2, and the albedo will be smooth if we add F_3. If we add both F_2 and F_3, the surface normal and the albedo becomes relatively sharper than adding either F_2 or F_3. Since we want to suppress the surface normal and the albedo to be smooth, we add not only F_2 and F_3 but also F_4.

3.3. Determining surface normal and albedo

If we ignore the influence of the shadow, the first term F_1 shown in Equation (9) can be represented as Equation (13).

$$F_1 = \sum_{c=0}^{6} (I_c(x,y) - A_c(x,y)(\mathbf{l}_c^\top \mathbf{n}(x,y)))^2 \quad . \tag{13}$$

The solution obtained by minimizing Equation (13) is expressed as Equation (14).

$$I_c(x,y) = A_c(x,y)(\mathbf{l}_c^\top \mathbf{n}(x,y)) \quad . \tag{14}$$

When albedos $A_0, A_1, \cdots, A_6$ are known, the surface normal $\mathbf{n}$ can be obtained by calculating the pseudo-inverse matrix $\mathbf{L}^+$ of matrix $\mathbf{L}$, as shown in Equation (15).

$$\begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} l_{x0} & l_{y0} & l_{z0} \\ l_{x1} & l_{y1} & l_{z1} \\ & \vdots & \\ l_{x6} & l_{y6} & l_{z6} \end{pmatrix}^+ \begin{pmatrix} I_0(x,y)/(A_0(x,y)+\varepsilon_1) \\ I_1(x,y)/(A_1(x,y)+\varepsilon_1) \\ \vdots \\ I_6(x,y)/(A_6(x,y)+\varepsilon_1) \end{pmatrix} . \tag{15}$$

Here, ε_1 is a small positive constant introduced to prevent division-by-zero. As the surface normal $\mathbf{n}$ is expressed as a unit vector ($\|\mathbf{n}\| = 1$), it is normalized after calculating Equation (15). The unit vector $\hat{\mathbf{n}}$ of the surface normal $\mathbf{n}$ can be calculated by dividing its length $\|\mathbf{n}\|$ as $\hat{\mathbf{n}} = \mathbf{n}/\|\mathbf{n}\|$.

Shadow has a low brightness, and thus, thresholding the brightness results in detecting the shadow, as is shown in Section 3.9. As for the channel which is detected as a shadow using the procedure shown in Section 3.9, Equation (15) cannot be used for surface normal estimation. To avoid this, $\mathbf{n}$ is calculated by weighting the c'th row of $\mathbf{L}$ by a small value d in relation to channel c, which is a shadow. This situation is expressed as follows.

$$\begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} l_{x0} & l_{y0} & l_{z0} \\ \vdots & \vdots & \vdots \\ l_{x,c-1} & l_{y,c-1} & l_{z,c-1} \\ dl_{x,c} & dl_{y,c} & dl_{z,c} \\ l_{x,c+1} & l_{y,c+1} & l_{z,c+1} \\ \vdots & \vdots & \vdots \\ l_{x6} & l_{y6} & l_{z6} \end{pmatrix}^{+} \begin{pmatrix} I_0(x,y)/(A_0(x,y)+\varepsilon_1) \\ \vdots \\ I_{c-1}(x,y)/(A_{c-1}(x,y)+\varepsilon_1) \\ dI_c(x,y)/(A_c(x,y)+\varepsilon_1) \\ I_{c+1}(x,y)/(A_{c+1}(x,y)+\varepsilon_1) \\ \vdots \\ I_6(x,y)/(A_6(x,y)+\varepsilon_1) \end{pmatrix} . \tag{16}$$

As usual, the surface normal $\mathbf{n}$ is normalized after calculating Equation (16).

When the surface normal is known, albedo can be calculated as shown in Equation (17) derived from Equation (14).

$$A_c = \frac{I_c}{\mathbf{l}_c^\top \mathbf{n}} \quad . \tag{17}$$

To prevent division-by-zero, Equation (17) is calculated when $\mathbf{l}_c^\top \mathbf{n} > \varepsilon_2$ holds, where ε_2 is a small positive constant. In addition, if the pixel is detected as an outlier (Section 3.9), Equation (17) is also not calculated.

There are seven constraint condition equations in Equation (14). The input brightness $I_0, I_1, \cdots, I_6$ and the unit vector that expresses the light source directions $\mathbf{l}_0, \mathbf{l}_1, \cdots, \mathbf{l}_6$ are known. Albedos $A_0, A_1, \cdots, A_6$ and normal vectors n_x, n_y, n_z are unknown parameters. Because the 3D normal vector is conditioned to be the unit vector, its degree-of-freedom is two. Therefore, the total number of unknown parameters is nine, with seven albedos and two surface normal components. At this point, calculations are not possible because the number of the unknown numbers is larger than the number of equations. Thus, the smoothing of the surface normal, smoothing of albedos, and constraint condition of the surface normal at the occluding boundary are introduced to the cost function.

3.4. Smoothness constraint for surface normal

As explained in Section 3.3, surface normal and albedo cannot be calculated because there are too many unknowns. Therefore, the smoothing of the surface normal is conducted as a constraint condition. The second term F_2 of cost function F, which expresses the smoothing term of the normal, is expressed as Equation (10).

The discretization of the first term of Equation (10) results in Equation (18) and that of the second term results in Equation (19).

$$\begin{aligned} \mathbf{n}(x,y) &= \frac{1}{4}\{\mathbf{n}(x+1,y)+\mathbf{n}(x-1,y)+\mathbf{n}(x,y+1)+\mathbf{n}(x,y-1)\}\,, && (18)\\ \mathbf{n}(x,y) &= \mathrm{median}\{\mathbf{n}(x+1,y),\mathbf{n}(x-1,y),\mathbf{n}(x,y+1),\mathbf{n}(x,y-1)\}\,. && (19) \end{aligned}$$

In our software, Equation (18) is implemented as Gaussian filter, and Equation (19) is implemented as median filter. Convolving Equation (18) multiple times can be approximated by Gaussian filter. Therefore, instead of applying Equation (18) multiple times, we applied Gaussian filter once. We first apply median filter before Gaussian filter. After the surface normal is smoothed, it is normalized to be a unit vector.

The fastest way to calculate Equation (19) is to calculate the median for each element as follows.

$$\begin{aligned} n_x &= \mathrm{median}\{n_x(x+1,y),n_x(x-1,y),n_x(x,y+1),n_x(x,y-1)\}\,,\\ n_y &= \mathrm{median}\{n_y(x+1,y),n_y(x-1,y),n_y(x,y+1),n_y(x,y-1)\}\,,\\ n_z &= \mathrm{median}\{n_z(x+1,y),n_z(x-1,y),n_z(x,y+1),n_z(x,y-1)\}\,. \end{aligned} \tag{20}$$

After that the vector is normalized to be a unit vector. This procedure calculates the median in Euclidean distance, not in Riemannian distance (geodesic distance). However, this difference does not matter in practice since the surface is assumed to be smooth: Namely, since the angle between neighboring pixels is small, the Euclidean distance of two vectors can be approximated as the Riemannian distance.

In order to keep the sharp feature of surface normal, median filter (Equation (19)) is used. The median filter will not change surface normal over neighboring pixels at sharp features. Although median filter is preferable to keep the sharp features, we also use Gaussian filter (Equation (18)) to constrain the surface normal to be smooth. Median filter does not change the surface normal at shapr features, and such pixels may be stuck in local minima. Gaussian filter (Equation (18)) can modify the surface normal even for such edges. We empirically found beneficial to use both median filter and Gaussian filter since these filters can find a good balance between smooth normals and sharp features.

As shown in Equation (6), there are nine unknown parameters and seven equations. Although Equation (18) or Equation (19) comprises three equations, the surface normal should be constrained as a unit vector; thus, Equation (18) or Equation (19) has two degrees-of-freedom. Now, we have nine unknown parameters and nine equations per pixel. The problem is now well-posed, but an over-smoothed surface normal will be obtained if we solely use this constraint. We add another constraint F_3, as shown in Section 3.5, in order to relatively reduce the influence of F_2.

3.5. Smoothness constraint for albedo

As discussed in Section 3.4, smoothing of the surface normal alone is insufficient as a constraint condition. Therefore, albedo smoothing is also conducted. The third term F_3 in the cost function, which expresses the albedo smoothing, is shown in Equation (11). Equation (11) is discretized as Equation (21).

$$\mathbf{A}(x,y) = \mathrm{median}\{\mathbf{A}(x+1,y), \mathbf{A}(x-1,y), \mathbf{A}(x,y+1), \mathbf{A}(x,y-1)\}\,. \tag{21}$$

Namely, we applied median filter to the albedo. As shown in Equation (6), there are nine unknown parameters and seven equations. Equation (21) implies seven equations because there are seven channels. Now, we have 9 unknown parameters and 14 equations per pixel, which results in a well-posed problem. However, an over-smoothed albedo will be obtained if we solely use this constraint. We add another constraint F_2 as shown in Section 3.4 in order to relatively reduce the influence of F_3.

3.6. Occluding boundary constraint and initial value of surface normal

The target objects of this study are smooth and closed surfaces. Here, the occluding boundary is the border region where the surface normal of the object begins to turn toward the rear just before it becomes invisible. The angle between the observation direction vector and the normal vector is 90° since we assume orthographic projection for camera model. It means that it is possible to correctly estimate the surface normal at the occluding boundary, which is orthogonal to the object area contour. This is incorporated into the cost function as F_4. The occluding boundary normal is defined as $\mathbf{n}_b$ (Equation (12)). Now, the solution that minimizes F_4 is $\mathbf{n}(x,y) = \mathbf{n}_b(x,y)$. At the occluding boundary, $\mathbf{n}_b$ is used as the estimation of the surface normal.

Although F_2 or F_3 are enough for solving Equation (8), using also F_4 is beneficial. The function F_2 itself has no boundary condition, and if we minimize F_2 only, the surface normal will be extraordinary smooth. In order to restrict the surface normal to be smooth, we will add F_4 as the boundary condition.

In addition, the pixel brightness close to the occluding boundary is unreliable, since it contains shadow in most of the channels. Since the reliability of the data term F_1 is small at occluding boundary, adding F_4 is reasonable.

To conduct the iteration process using the cost function, initial values are required for the surface normal and albedo. As follows, the initial value of surface normal is calculated from the surface normal

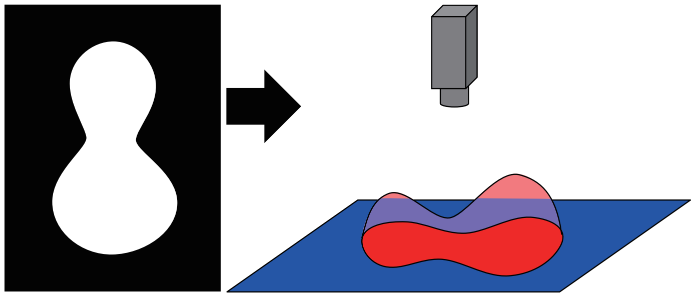

Figure 3. Approximate shape used for initial guess to surface normal. The shape is inflated using the silhouette of the object region.

at occluding boundary (Figure 3). As is done in previous work [24], we also inflated the silhouette to make the approximate shape. Our approach is shown as follows.

First, we calculate the distance from each pixel to the nearest occluding boundary pixel, and next, we sort the distance. As for initial guess, we assume that the probability distribution of the height of the target object is the same as that of the hemisphere. Let us denote the maximum of the distance as D_{max}. The number of the pixels in object region is $|\mathcal{P}|$. The order of the sorted pixel (x, y) is denoted as o. If we assume that the object is hemisphere whose radius is r, then r is calculated from $|\mathcal{P}| = \pi r^2$. The area o whose length from the center of the circle is denoted as l can be represented as $o = \pi l^2$. Therefore, l can be calculated from o. The height h is represented as $r^2 = h^2 + l^2$ where the distance from the center of the circle is l. Therefore, h can be calculated. Height of the unit hemisphere is calculated by dividing r from h. Multiplying D_{max} results in the height of the hemisphere where its maximum height is D_{max}. After that, the height field is slightly smoothed.

The initial height (Figure 3) is obtained by above procedure. Differentiating the height and normalizing it as follows results in the surface normal $\hat{\mathbf{n}}$.

$$n_x = -\frac{\partial h}{\partial x}, \quad n_y = -\frac{\partial h}{\partial y}, \quad n_z = 1\,. \tag{22}$$

$$\hat{\mathbf{n}} = \frac{(n_x, n_y, n_z)}{\sqrt{n_x^2 + n_y^2 + n_z^2}}\,. \tag{23}$$

Finally, the smoothed and normalized surface normal is used as the initial value.

3.7. Initial value of albedo

It is better to use an initial value of albedo which is close to the true albedo as much as possible, in order to speed up the convergence. However, since we do not use additional sensors or data, we have to calculate the initial albedo solely from input image. The input image is a single seven-channel image, whose light source direction is different. We calculate the average of seven channels, and such average image I_{avg} works well for initial albedo.

$$\tilde{I}_{\mathrm{avg}}(x, y) = \frac{1}{7}\left(I_0(x, y) + I_1(x, y) + \cdots + I_6(x, y)\right)\,, \tag{24}$$

$$I_{\mathrm{avg}} = \mathrm{bilateral}(\tilde{I}_{\mathrm{avg}})\,. \tag{25}$$

This is the sole image we can obtain from seven input images closest to the true albedo. If an infinite number of light sources illuminate the object uniformly from the surroundings, the observation of the object becomes the same as that of the albedo with constant scaling. This is the reason why

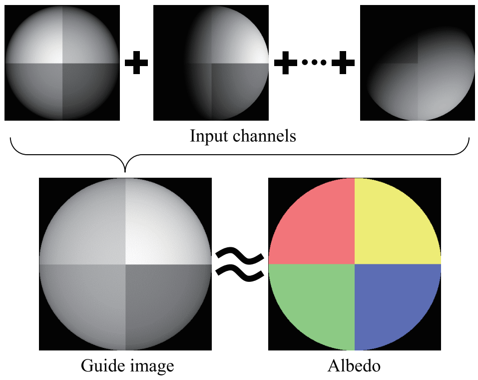

Figure 4. Average image calculated from seven channel images resembles the albedo.

the average image can be a good estimate of albedo. As shown in Fig. 4, the true albedo value and brightness of the average image I_{avg} are similar; therefore, the average image can be used as the initial guess. In order to decrease the effect of the shadow, bilateral filter is applied to the average image.

The albedo $\mathbf{A}$ is highly correlated with the input image brightness $\mathbf{I}$. The initial albedo $A_c(x, y)$ is set to be an image where the brightness of the average image I_{avg} is scaled to be the same as the brightness of each channel.

$$A_c(x, y) = I_{\mathrm{avg}}(x, y) \operatorname*{median}_{(\tilde{x},\tilde{y})\in\mathcal{P}} \left(\frac{I_c(\tilde{x}, \tilde{y})}{I_{\mathrm{avg}}(\tilde{x}, \tilde{y})} \right) . \tag{26}$$

In order to robustly calculate the ratio I_c / I_{avg}, median of the ratio is used.

3.8. Update rule

After the initial values for the normal $\mathbf{n}$ and albedo $\mathbf{A}$ are calculated, as shown in Section 3.6 and Section 3.7, the calculations are iterated several times. First, the surface normal is calculated according to the procedure shown in Section 3.3. The calculated surface normal is denoted as $\mathbf{n}'$, and the surface normal of the previous step is denoted as $\mathbf{n}''$. Instead of using $\mathbf{n}'$, the new surface normal $\mathbf{n}$ for the next step is calculated as Equation (27).

$$\mathbf{n} = (1 - \alpha_n)\mathbf{n}' + \alpha_n \mathbf{n}'' . \tag{27}$$

The constant α_n stabilizes the update of the surface normal, resulting in robust optimization. Actually, instead of Equation (27), we implemented our software as follows.

$$\begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \begin{pmatrix} rl_{x0} & rl_{y0} & rl_{z0} \\ \vdots & \vdots & \vdots \\ rl_{x,c-1} & rl_{y,c-1} & rl_{z,c-1} \\ drl_{x,c} & drl_{y,c} & drl_{z,c} \\ rl_{x,c+1} & rl_{y,c+1} & rl_{z,c+1} \\ \vdots & \vdots & \vdots \\ rl_{x6} & rl_{y6} & rl_{z6} \\ \tilde{\alpha}_n & 0 & 0 \\ 0 & \tilde{\alpha}_n & 0 \\ 0 & 0 & \tilde{\alpha}_n \end{pmatrix}^{+} \begin{pmatrix} rI_0(x,y)/(A_0(x,y)+\varepsilon_1) \\ \vdots \\ rI_{c-1}(x,y)/(A_{c-1}(x,y)+\varepsilon_1) \\ drI_c(x,y)/(A_c(x,y)+\varepsilon_1) \\ rI_{c+1}(x,y)/(A_{c+1}(x,y)+\varepsilon_1) \\ \vdots \\ rI_6(x,y)/(A_6(x,y)+\varepsilon_1) \\ \tilde{\alpha}_n \tilde{n}_x \\ \tilde{\alpha}_n \tilde{n}_y \\ \tilde{\alpha}_n \tilde{n}_z \end{pmatrix} . \tag{28}$$

Here, the surface normal of previous iteration is represented as $(\tilde{n}_x, \tilde{n}_y, \tilde{n}_z)$ and the updated surface normal to be taken over to the next iteration is represented as (n_x, n_y, n_z). After solving this equation, the obtained surface normal is normalized.

Here, we have employed additional weight r. This weight depends on the number of valid channels for each pixel. If there are no shadows and speculars in all seven channels, we set r as large number, so that the surface normal calculated by photometric stereo equation becomes much important than the surface normal of the previous iteration $(\tilde{n}_x, \tilde{n}_y, \tilde{n}_z)$. If there are many invalid channels, the surface normal calculated by photometric stereo equation becomes unreliable, thus we set r small so that surface normal will be unchanged. We define r as follows using the parameter w.

$$r = \left(\frac{\max(v-2, 0)}{7-2} \right)^w . \tag{29}$$

Here, v is the number of valid channels. We found empirically that $w > 1$ is good for stable computation.

Next, albedo is calculated according to the procedure shown in Section 3.3. The calculated albedo is denoted as $\mathbf{A}'$, and the albedo of the previous step is denoted as $\mathbf{A}''$. The update rule for albedo is shown in Equation (30).

$$\mathbf{A} = (1-\alpha_a)\mathbf{A}' + \alpha_a \mathbf{A}'' . \tag{30}$$

The constant value α_a stabilizes the optimization.

Instead of using Equation (30), we implemented this process as follows.

$$A_c = \frac{(1-\tilde{\alpha}_a) I_c + \tilde{\alpha}_a \tilde{A}_c}{(1-\tilde{\alpha}_a)(\mathbf{l}_c^\top \mathbf{n}) + \tilde{\alpha}_a} . \tag{31}$$

This is a weighted sum of Equation (17) and the previously calculated albedo $\tilde{A}_c$ with the weight $\tilde{\alpha}_a$. Note that Equation (31) is calculated if channel c is marked as valid through the process shown in Section 3.9, and $A_c = \tilde{A}_c$ is used if it is invalid.

3.9. Outlier detection

Detecting specular reflection in color photometric stereo problems is difficult. One of the common approaches for detecting specular reflection is to use color. The colors of diffuse reflection and specular reflection are usually different; thus, the diffuse reflection and specular reflection can be separated when the scene is illuminated by a nearly white light source. However, the color photometric stereo illuminates the object with three different colors, and thus, the color-based approach cannot solve the problem. Another approach is to use principal component analysis or singular value decomposition, which represents the image with three orthonormal bases. However, the color of each light is different in color photometric stereo approach, and thus, the images cannot be represented by a linear sum of three bases. As a result, the remaining approach is to use the strong brightness change caused at specular reflection.

Therefore, we have no choice but to use thresholding approach for outlier (specular / shadow) detection. Suppose that the maximum brightness of the object for all channels is $I_{\max}$ and the minimum is $I_{\min}$. We use $T_{\max} = I_{\max} - t_{\max}$ and $T_{\min} = I_{\min} + t_{\min}$ as thresholds, where $t_{\max}$ and $t_{\min}$ are small positive constants. Outlier map N, which is 1 for outlier and 0 for valid pixel, is designed as follows.

$$\tilde{N}_c(x,y) = \begin{cases} 1 & \textit{if } I_c(x,y) > T_{\max} \textit{ or } I_c(x,y) < T_{\min} , \\ 0 & \text{otherwise} , \end{cases} \tag{32}$$

$$N_c = \mathrm{dilation}(\tilde{N}_c) . \tag{33}$$

Here, "dilation" is an operator which dilates the "1" pixels, which is a well-known operator in binary image processing, which we skip to explain.

3.10. Calculating height from surface normal

In this section, we briefly introduce the procedure to calculate the height from surface normal. Here, we assume orthographic projection, and the perspective projection case is shown in the literature [28]. More details are given in the literature [16,17,28].

The shape is represented as the height H set for each pixel. The partial derivatives of the heights with respect to x and y are called gradient, and represented as p and q, respectively.

$$p = H_x = \frac{\partial H}{\partial x}, \quad q = H_y = \frac{\partial H}{\partial y} \quad . \tag{34}$$

The surface normal $\mathbf{n}$ is represented by these gradients as shown below.

$$\mathbf{n} = \frac{(-p, -q, 1)^\top}{\sqrt{p^2 + q^2 + 1}} . \tag{35}$$

The cost function that relates the surface normal to the height is shown below.

$$\iint (H_x - p)^2 + (H_y - q)^2 \, dxdy \quad . \tag{36}$$

The Euler equation (Euler-Lagrange differential equation) that minimizes the equation

$$\iint F(u, u_x, u_y) dxdy \quad , \tag{37}$$

can be expressed as

$$F_u - \frac{\partial F_{u_x}}{\partial x} - \frac{\partial F_{u_y}}{\partial y} = 0 \quad . \tag{38}$$

As for H, the Euler equation that minimizes Equation (36) is derived as follows:

$$H_{xx} + H_{yy} - p_x - q_y = 0 \quad . \tag{39}$$

Here, H_{xx} and H_{yy} can be discretized as follows:

$$H_{xx} = H(x+1, y) + H(x-1, y) - 2H(x, y) \tag{40}$$

$$H_{yy} = H(x, y+1) + H(x, y-1) - 2H(x, y) \quad . \tag{41}$$

Thus, substituting Equations (40) and (41) into Equation (39) yields the following equation.

$$H(x,y) = \frac{1}{4}(H(x+1,y) + H(x-1,y) + H(x,y+1) + H(x,y-1)) - \frac{1}{4}(p_x(x,y) + q_y(x,y)) . \tag{42}$$

As is shown in Equation (35), the gradients p and q are calculated from the surface normal $\mathbf{n}$. The partial differentiation of gradients used for Equation (42) is discretized as follows.

$$\begin{aligned} p_x(x,y) &= p(x+1,y) - p(x-1,y) , \\ q_y(x,y) &= q(x,y+1) - q(x,y-1) . \end{aligned} \tag{43}$$

After computing Equation (43), we solve Equation (42) to determine the height H. In this paper, we solve Equation (42) using the successive over-relaxation method, but any other methods are also applicable, such as Fourier transform [11] or preconditioned conjugate gradient [2].

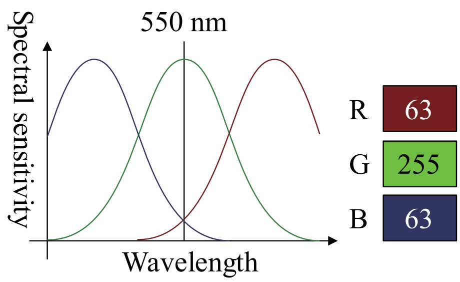

Figure 5. Example of camera spectral sensitivity which has channel crosstalk.

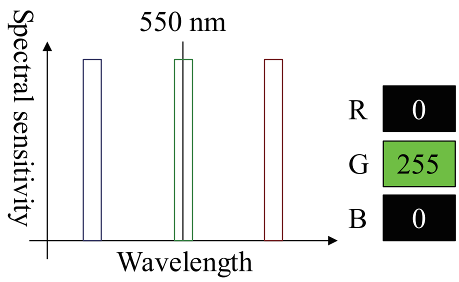

Figure 6. Example of camera spectral sensitivity which does not have channel crosstalk.

3.11. Channel crosstalk

In an instrument that independently uses signals of two or more channels, signal leaking from one channel to another is called crosstalk. Our experiment uses a multi-band camera that has seven channels and detects undesired colors of other channels. The undesired effect of a color camera is called channel crosstalk [3,8,10,19].

Figure 5 is an example of a three-band RGB camera that detects 550 nm green light as $(R, G, B) = (63, 255, 63)$. This signal should be $(R, G, B) = (0, 255, 0)$ since the observed green light wavelength is 550 nm. As shown in Fig. 5, the bandwidth of each spectral sensitivity is wide, and thus, has some overlaps; therefore, the R and B channels also detect the color of green light. Color photometric stereo assumes that the sensor has no channel crosstalk, as shown in Fig. 6; thus, we must remove channel crosstalk.

To detect the channel crosstalk, we use a diffuse white reflectance standard, which has flat spectral reflectance for each wavelength. The seven-band camera captures seven images of the diffuse white reflectance standard illuminated by one of the seven light sources, which are lit one-by-one. A single channel is sensitive to each light; thus, the signals of other channels are caused by the crosstalk.

Channel crosstalk can be represented by a color mixing matrix $\mathbf{X}$. Since we use a seven-band camera, the size of matrix $\mathbf{X}$ is 7×7. Let us denote the ideal signal without channel crosstalk as $\mathbf{d}_i$. This seven-dimensional column vector $\mathbf{d}_i$ becomes $\mathbf{d}_o$ because it is affected by channel crosstalk. The relation between these signals and the color mixing matrix is as follows.

$$\mathbf{d}_o = \mathbf{X}\mathbf{d}_i \quad . \tag{44}$$

The original signal $\mathbf{d}_i$ can be recovered from the captured signal $\mathbf{d}_o$ as follows.

$$\mathbf{d}_i = \mathbf{X}^{-1}\mathbf{d}_o \quad . \tag{45}$$

The color mixing matrix $\mathbf{X}$ should be obtained prior to the measurement, and the input image should be converted by the inverse of the color mixing matrix $\mathbf{X}^{-1}$ before applying the proposed algorithm.

Suppose that we look at the 0th channel of the diffuse white reflectance standard illuminated by the 0th light with narrow-band wavelength. Ideally, the signal should be zero for each channel, except the 0th channel. We define the value of the 0th channel as 1. Namely, the ideal signal $\mathbf{d}_i = (1, 0, 0, 0, 0, 0, 0)^{\mathrm{T}}$ becomes $\mathbf{d}_o = (w_{0,0}, w_{1,0}, \cdots, w_{6,0})^{\mathrm{T}}$ after observation.

$$\begin{pmatrix} w_{0,0} & w_{1,0} & w_{2,0} & \cdots & w_{6,0} \end{pmatrix}^{\top} = \mathbf{X} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix}^{\top} . \tag{46}$$

Similarly, the diffuse white reflectance standard illuminated by the 1st light is expressed as follows.

$$\begin{pmatrix} w_{0,1} & w_{1,1} & w_{2,1} & \cdots & w_{6,1} \end{pmatrix}^{\top} = \mathbf{X} \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^{\top} . \tag{47}$$

This procedure is repeated until the 6th light. The following equation expresses the whole measurement, which is conducted seven times.

$$\begin{pmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,6} \\ w_{1,0} & w_{1,1} & \dots & w_{1,6} \\ \vdots & \vdots & \ddots & \vdots \\ w_{6,0} & w_{6,1} & \dots & w_{6,6} \end{pmatrix} = \mathbf{X} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} . \tag{48}$$

As a result, the color mixing matrix $\mathbf{X}$ is obtained as follows.

$$\mathbf{X} = \begin{pmatrix} w_{0,0} & w_{0,1} & \dots & w_{0,6} \\ w_{1,0} & w_{1,1} & \dots & w_{1,6} \\ \vdots & \vdots & \ddots & \vdots \\ w_{6,0} & w_{6,1} & \dots & w_{6,6} \end{pmatrix} . \tag{49}$$

The inverse of the color mixing matrix $\mathbf{X}^{-1}$ can cancel the channel crosstalk of the observed signal. The output ideal signal d_i is calibrated such that the signal of the diffuse white reflectance standard would be $(1, 1, \cdots, 1)$.

4. Experiment

4.1. Experimental setup

The camera used for this experiment is an FD-1665 3CCD multi-spectral camera by FluxData, Inc., USA, as shown in Fig. 7. It comprises two color sensors and a near-infrared (NIR) sensor. Each sensor is sensitive to its respective wavelength; i.e., each color sensor can record the components from three channels, and the NIR sensor can record the components from one channel. Figure 8 shows the spectral sensitivity of the camera. As shown in Fig. 8, channel crosstalk occurred among all camera channels. Therefore, the method shown in Section 3.11 is used to remove the channel crosstalk in the photographed input image. The diffuse white reflectance standard is used to obtain the color mixing matrix shown in Figure 9, where the row denotes the channel number and the column denotes the light number. The color mixing matrix is created using the average value of the diffuse white reflectance standard.

Table 1 shows the full width at half maximum (FWHM) for each light source used in this experiment.

The light source directions were determined prior to the experiment by photographing a mirrored ball. The locations of the light sources and the camera were then left unchanged.

Figure 7. Multispectral camera "FluxData FD-1665 (USA)."

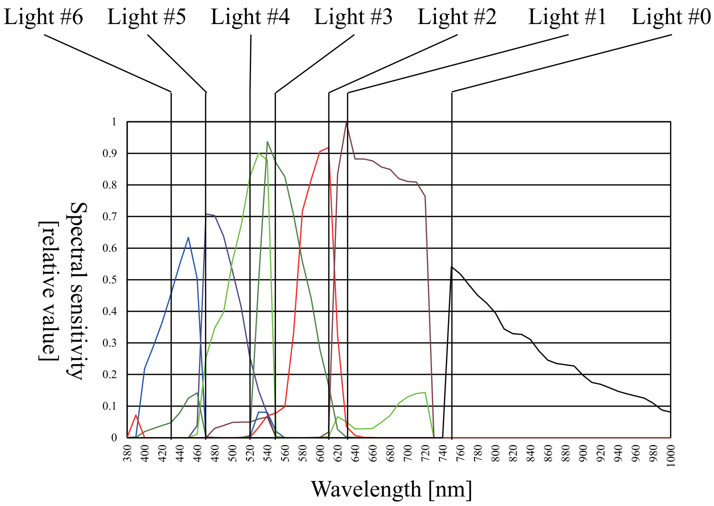

Figure 8. Spectral sensitivity of multispectral camera and peak wavelength of each light sources.

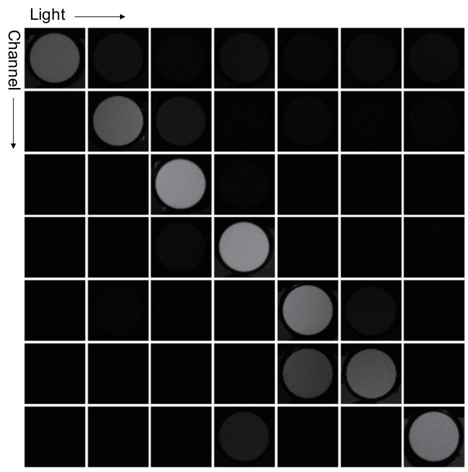

Figure 9. The obtained color mixing matrix for canceling channel crosstalk. The average brightness of white reflectance stardard becomes the color mixing matrix. The matrix will be diagonal matrix if there are no channel crosstalk, however non-diagonal element is slightly bright due to the channel crosstalk.

Table 1. Peak wavelength and FWHM for each light source.

Light	0	1	2	3	4	5	6
Peak	750nm	632nm	610nm	550nm	520nm	470nm	430nm
FWHM	10nm	10nm	10nm	10nm	10nm	10nm	10nm

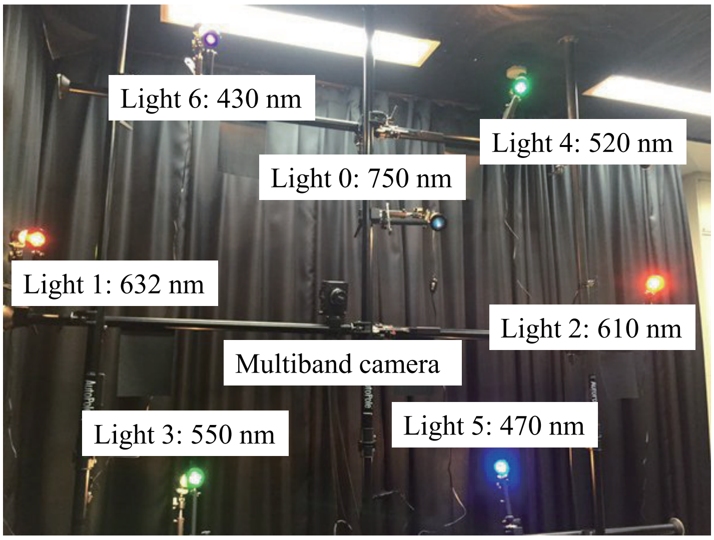

Figure 10. Experimental setup with 7 light sources with different wavelengths and a single 7-band multispectral camera.

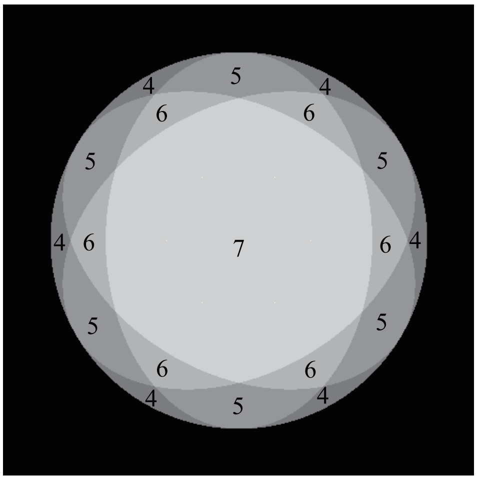

Figure 11. Gaussian sphere representation of surface normal where the north pole is the center of this picture. The number indicates how many light sources are lit for each direction of surface normal.

The experiment was conducted in a darkroom. To increase the amount of supplementary information obtained for objects with narrow-wavelength regions, light sources of close wavelength were positioned opposite to each other. The NIR light source was placed next to the camera. Figure 10 shows a diagram of the experiment.

Each point on the object's surface must always be illuminated by more than three light sources for the photometric stereo method. If there are six light sources, any point on the surface can be illuminated by at least three light sources [32]. Additionally, when specular reflection occurs, one picture that can be used for the photometric stereo method is eliminated. Therefore, the NIR light source is placed next to the camera so that each point is illuminated by at least four light sources. Figure 11 is Gaussian sphere representation of the surface normal, where the number of each region represents the number of light sources illuminated.

In the photometric stereo method, precision increases when the angle between the light sources is widened, i.e., the baseline is lengthened, because it increases the shading contrast. However, when the baseline is lengthened, the shadow area increases. The locations of the light sources must, therefore, be limited to a certain solid angle. When seven points are placed within a fixed circle, the placement of

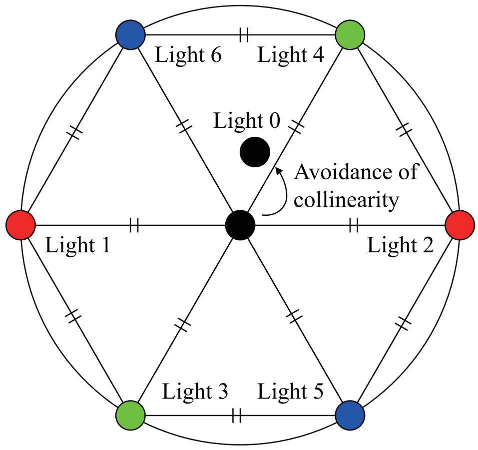

Figure 12. Schematic illustration of the geometrical location of seven light sources. Six lights are placed at each apexes of regular hexagon. Multispectral camera is placed at the center of the hexagon. Infrared light is placed near the camera.

the points must be as far from each other as possible to comprise the vertices of a regular hexagon and its center, as shown in Figure 12. Therefore, when placing seven light sources within a limited area for the photometric stereo method, it is optimal to place them at the vertices of a regular hexagon and its center.

However, when three of the light sources selected from these seven lights are placed on the same straight line, or more precisely, when the three light source vectors are coplanar, the surface normal cannot be estimated by combining the three light sources. This is because combining these three light sources causes the light source matrix to degenerate. Suppose that the surface normal $\mathbf{n}$ is illuminated by light sources $\mathbf{l}_0$, $\mathbf{l}_1$, and $\mathbf{l}_2$, and is observed as the pixel brightnesses I_0, I_1, and I_2, respectively, while ignoring the shadow. If the light source directions are known, the surface normal can be obtained from following equation if there is an inverse of 3×3 light source matrix $(\mathbf{l}_0, \mathbf{l}_1, \mathbf{l}_2)^\top$.

$$\begin{pmatrix} \mathbf{l}_0^\top \\ \mathbf{l}_1^\top \\ \mathbf{l}_2^\top \end{pmatrix} \begin{pmatrix} \mathbf{n} \end{pmatrix} = \begin{pmatrix} I_0 \\ I_1 \\ I_2 \end{pmatrix} \tag{50}$$

The determinant of $(\mathbf{l}_0, \mathbf{l}_1, \mathbf{l}_2)^\top$ is the scalar triple product $\mathbf{l}_0 \cdot (\mathbf{l}_1 \times \mathbf{l}_2)$. If $\mathbf{l}_0$, $\mathbf{l}_1$, and $\mathbf{l}_2$ are coplanar, the vector $\mathbf{l}_1 \times \mathbf{l}_2$ becomes orthogonal to the vector $\mathbf{l}_0$, thus the determinant becomes zero. Although two-light photometric stereo exists [27], it is better to avoid three lights to be coplanar if we have more than two lights. Therefore, the NIR light source is placed at a small distance from the center of the regular hexagon so that no three light sources are on the same straight line. The camera is placed at the center of the regular hexagon.

4.2. Experimental result

The computation time of the main part of the algorithm (*i.e.*, excluding the computation time of calculating the initial value) is about ten seconds for ordinary object and ordinary computer with single thread and without any fine tuning to the source code.

As for all experimental results shown in this section, we used $\tilde{\alpha}_n = 0.1$ and $\tilde{\alpha}_a = 0.99$. These two parameters are the most important parameters which affect the final result, and other parameters are relatively less influential in comparison to these parameters. We used 4 for the standard deviation of Gaussian filter for smoothing the surface normal, and 15×15 and 11×11 for the window size of median filter of surface normal and albedo, respectively. The iteration number was set to be 2. We used

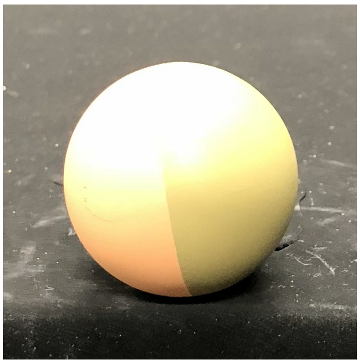

Figure 13. Spherical object with two different colors is used for evaluation since we know the mathematically true surface normal of the sphere.

$w = 16$, $d = 0.0001$, $\epsilon_1 = 0.001$, and $\epsilon_2 = 0.1$. The abovementioned parameters are the all parameters used in the main process.

As for calculating the initial albedo, we used 2 for the standard deviation of spatial parameter, and 20 for the standard deviation of intensity parameter for the bilateral filter. When applying the bilateral filter, the pixel brightness of outlier is scaled by 0.1 when calculating the weighted sum. The iteration number of the bilateral filter is set to be 10. As for calculating the scale, in order to avoid division-by-zero error, $I_{\text{avg}} \leq 0.1$ is not used for calculating Equation (26). As for calculating the initial normal, smoothing filter is applied twice: First it is applied to the height data and next it is applied to the surface normal. As for smoothing, 3×3 box filter is used, and the iteration number was set to be 100, for both the height and the normal. As for outlier detection, $t_{\max} = 15$ and $t_{\min} = 5$ are used. The number of dilation is set to be 1. The abovementioned parameters are the all parameters used in calculating the initial values.

First, we measured a plastic sphere to evaluate our system. The spherical object shown in Fig. 13 consists of two types of albedos. Figure 14 shows the error map with pseudo-color representation. The error is evaluated as an angle between the estimated surface normal and the true surface normal. We measured a sphere because its true surface normal can be obtained from the mathematical expression of the sphere. We compared our method with the so-called "naive color photometric stereo." In this paper, we define the color photometric stereo that assumes white objects as target as naive color photometric stereo. The generalized color photometric stereo problem shown in Equation (6) has nine unknown parameters; however, naive color photometric stereo has three unknown parameters: single albedo value (one parameter) and 3D surface normal (two parameters since it is constrained to be a unit vector). Therefore, naive color photometric stereo directly solves the linear equation even if the image is captured by a three-band color camera. Naive color photometric stereo robustly estimates the surface normal of white shirts, white dresses, and so on. The mean error of naive color photometric stereo (Fig. 14 (a)) were 0.343 [rad]. Our method overwhelms the previous approach, and our mean error (Fig. 14 (b)) was 0.148 [rad].

We used an owl figurine (Fig. 16 (a)) as the measurement object. Fig. 15 shows the seven-channel image obtained from the experiment. The captured image shown in Fig. 15 (a) is contaminated by channel crosstalk, and thus, we cancelled it, which resulted in Fig. 15 (b). The surface normal estimated by naive color photometric stereo is shown in Fig. 16 (b) and that estimated by our method is shown in Fig. 16 (c). As usual, the x, y, and z axes of the surface normal are linearly converted to R, G, and B for the pseudo-color representation of the surface normal. The estimated albedo is shown in Fig. 17. The shapes obtained by naive color photometric stereo and by our method are shown in Fig. 18 (a) and 18 (b), respectively.

The same experiment was also conducted with another multicolored object. The results with the doll and Buddha figurines are shown in Figs. 19–21 and 22–24, respectively.

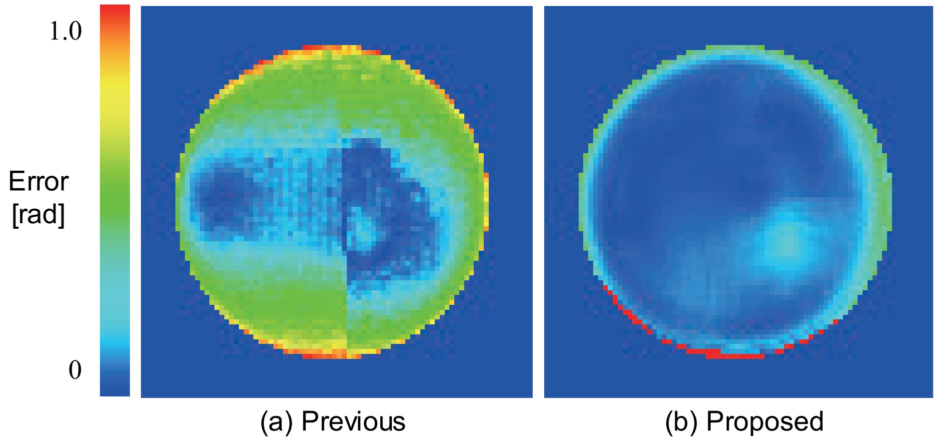

Figure 14. The error map of the sphere where the error is represented as angular difference between estimated value and ground truth (red: large, blue: small): (a) Naive color photometric stereo and (b) our method.

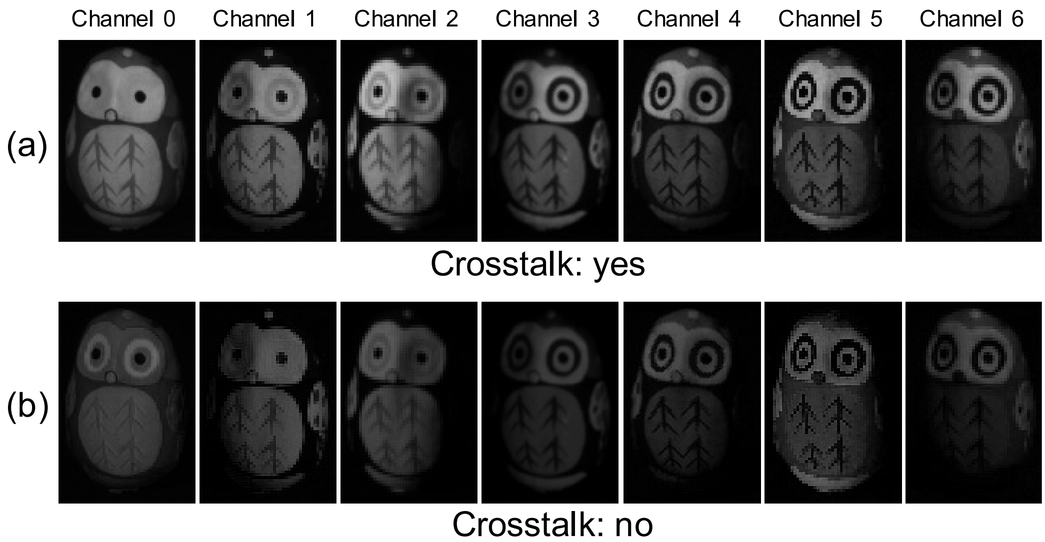

Figure 15. Obtained multi-band image [owl]: (a) Captured image and (b) image after cancelling channel crosstalk. If you look carefully, you may know that the channel crosstalk is removed. However, the difference is difficult to recognize since the crosstalk is small as is shown in Figure 9.

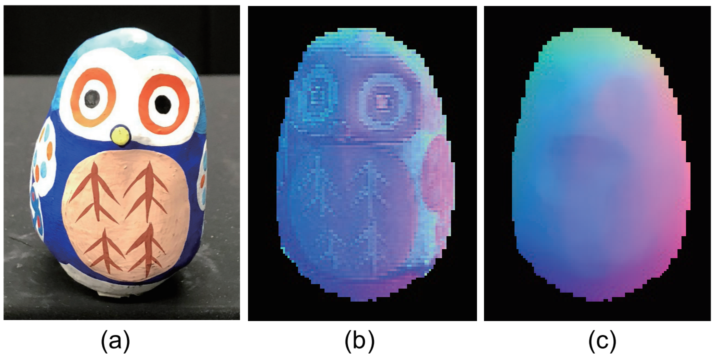

Figure 16. The result of owl object, which only causes diffuse reflection. Estimated surface normal [owl]: (a) Target object, (b) surface normal of naive color photometric stereo, and (c) surface normal of our method. The proposed method is not affected by the albedo difference.

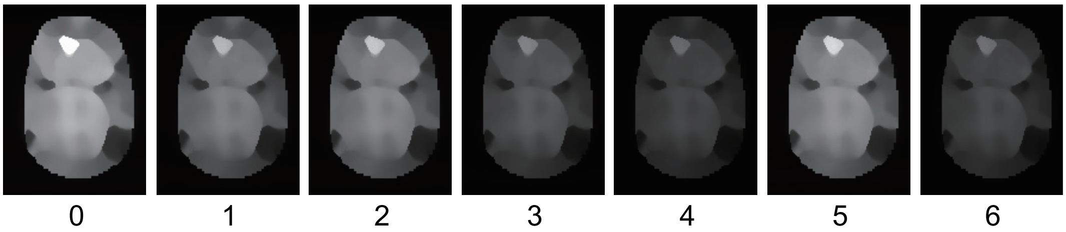

Figure 17. The result of owl object, which only causes diffuse reflection. Estimated albedo is shown, which is smooth enough.

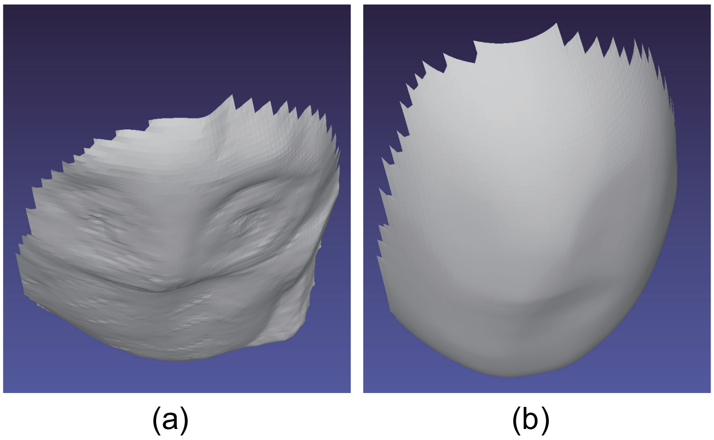

Figure 18. The result of owl object, which only causes diffuse reflection. Estimated geometry [owl]: (a) Naive color photometric stereo and (b) our method. The proposed method is not affected by the albedo difference.

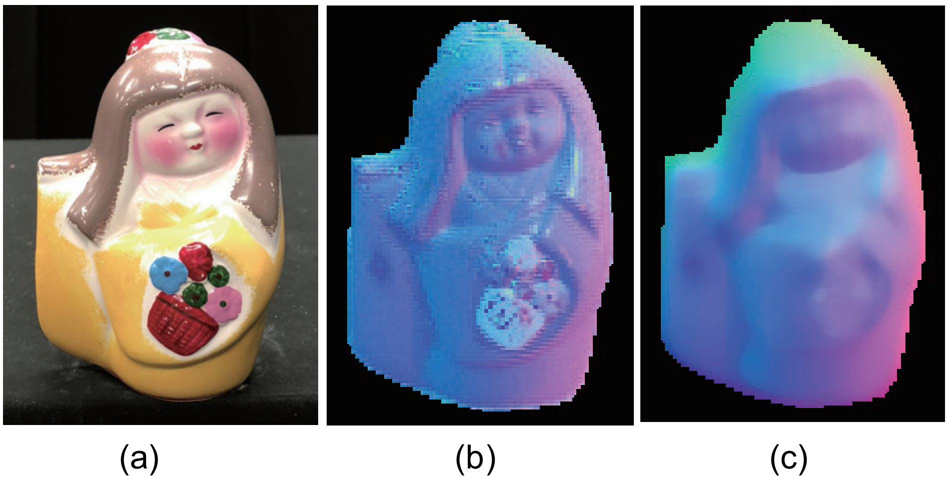

Figure 19. The result of doll object, which causes strong specular reflection. Estimated surface normal [doll]: (a) Target object, (b) surface normal of naive color photometric stereo, and (c) surface normal of our method. The proposed method is not affected by the albedo difference appears at the flower basket.

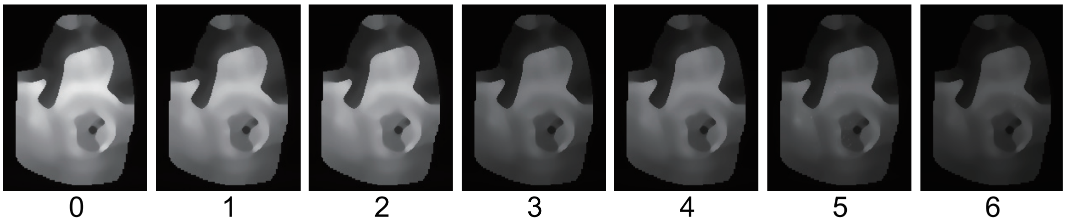

Figure 20. The result of doll object, which causes strong specular reflection. Estimated albedo is shown, which is smooth enough.

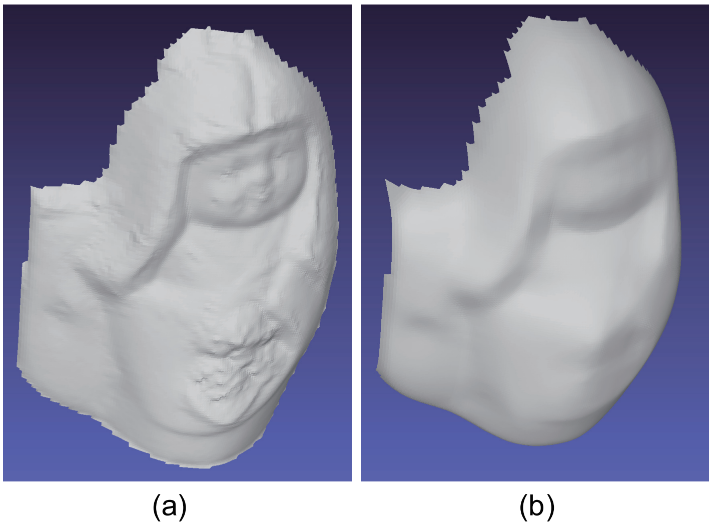

Figure 21. The result of doll object, which causes strong specular reflection. Estimated geometry [doll]: (a) Naive color photometric stereo and (b) our method. The proposed method is not affected by the albedo difference appears at the flower basket.

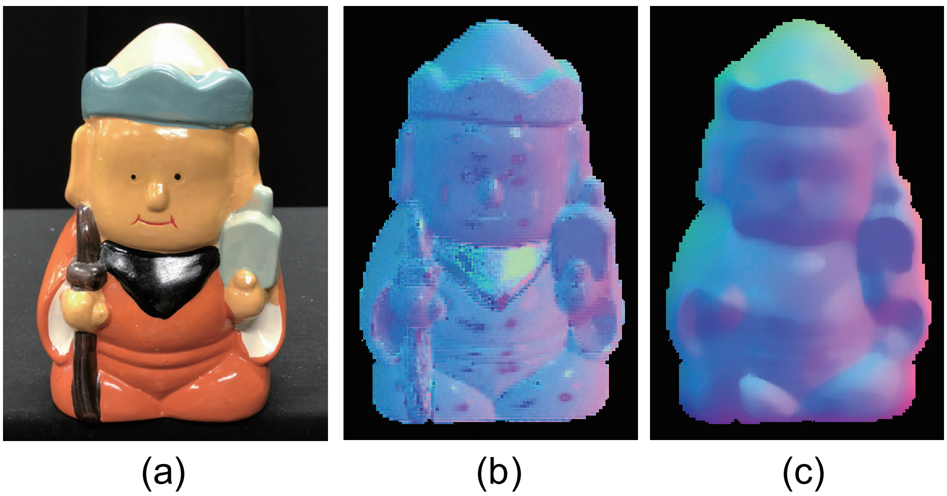

Figure 22. The result of buddha object, which causes strong specular reflection. Estimated surface normal [Buddha]: (a) Target object, (b) surface normal of naive color photometric stereo, and (c) surface normal of our method. The proposed method can smooth the surface normal of the scarf whose surface normal is unreliable due to black paint.

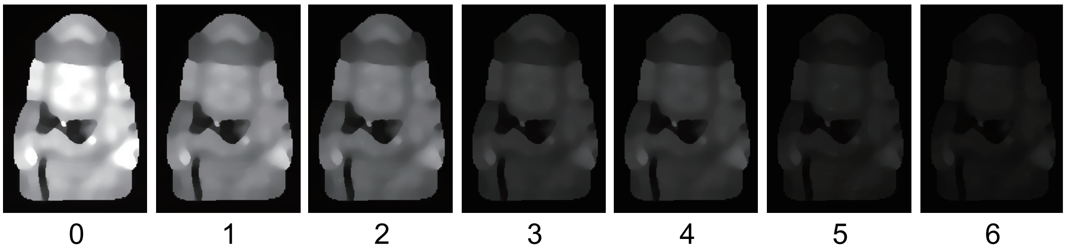

Figure 23. The result of buddha object, which only causes strong specular reflection. Estimated albedo is shown, which is smooth enough.

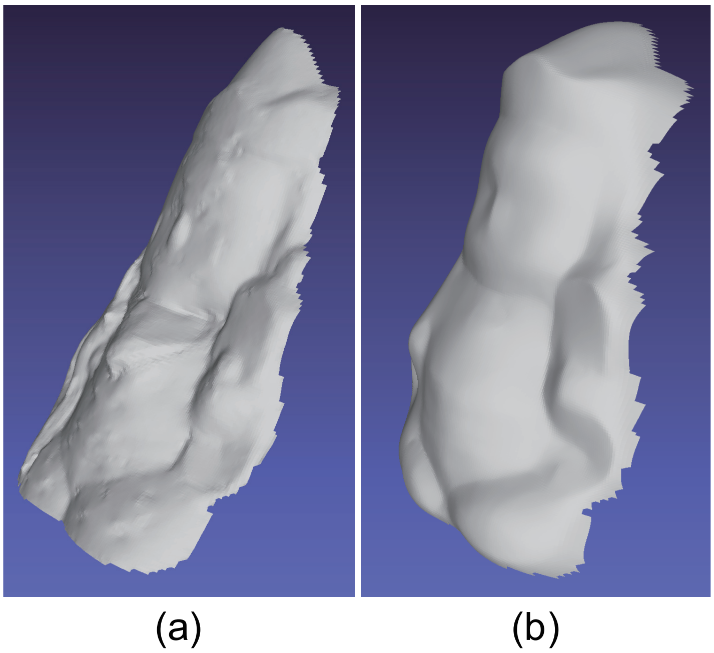

Figure 24. The result of buddha object, which causes strong specular reflection. Estimated geometry [Buddha]: (a) Naive color photometric stereo and (b) our method. The proposed method can smooth the surface normal of the scarf whose surface normal is unreliable due to black paint.

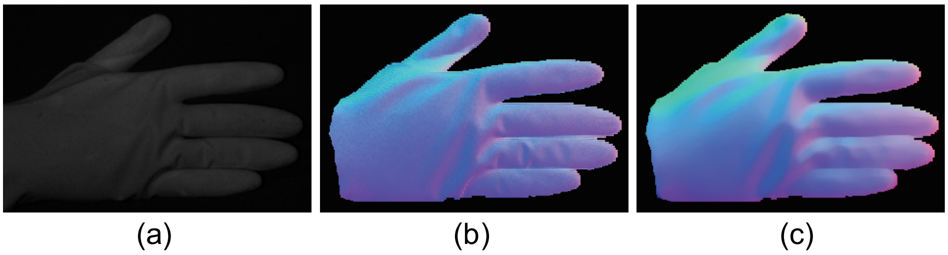

Figure 25. The result of hand with glove. Estimated surface normal [pose 1]: (a) One of the seven channel images, (b) estimated surface normal [naive color photometric stereo], and (c) estimated surface normal [our method]. Both color photometric stereos can estimate the surface normal of the dynamically deforming object.

Figure 26. The result of hand with glove. Estimated albedo [pose 1] is shown, which is smooth enough.

The advantage of color photometric stereo is that the surface normal of dynamic objects can be obtained. Most existing color photometric stereo methods measure white shirts, white dresses, etc., to verify that these methods can be applied to dynamically deforming objects. Due to the small size of the darkroom, we measured a glove instead of clothes. Figures 25–27 show the measurement results, and Figs. 28–30 show the results of the same object but differently deformed.

4.3. Discussion

Figure 31 (a) shows the result of Microsoft Kinect sensor. For comparison, our result is shown in Figure 31 (b). Kinect measures the depth and photometric stereo measures the surface normal. These two sensors are fundamentally different, however, since Kinect is a well-known commercial product of shape measurement, we think beneficial to show Figure 31 for the readers.

Figure 32 shows how the surface normal is affected by the parameters (Equation (28) and Equation (31)). Figures 32 (a) and (b) are the results when $\tilde{\alpha}_a = 0.1$, while Figures 32 (c) and (d) are the results when $\tilde{\alpha}_a = 0.99$. Figures 32 (a) and (c) are the results when $\tilde{\alpha}_n = 0.1$, while Figures 32 (b) and (d) are the results when $\tilde{\alpha}_n = 0.99$. Figure 32 (b) is smoother than Figure 32 (a), and Figure 32 (d) is smoother than

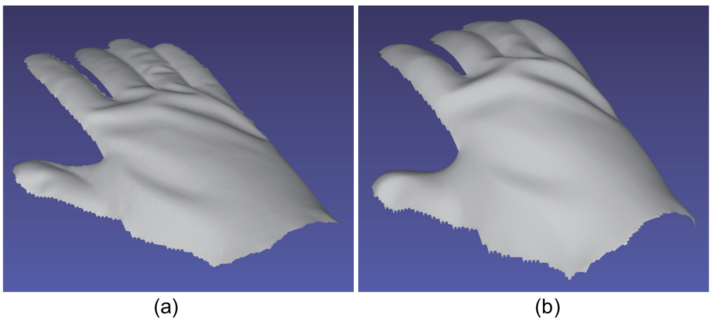

Figure 27. The result of hand with glove. Estimated geometry [pose 1]: (a) Estimated geometry [naive color photometric stereo] and (b) estimated geometry [our method]. Both color photometric stereos can estimate the surface normal of the dynamically deforming object.

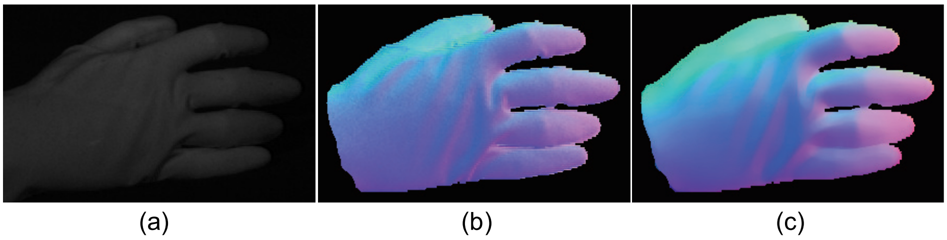

Figure 28. The result of hand with glove. Estimated surface normal [pose 2]: (a) One of the seven channel images, (b) estimated surface normal [naive color photometric stereo], and (c) estimated surface normal [our method]. Both color photometric stereos can estimate the surface normal of the dynamically deforming object.

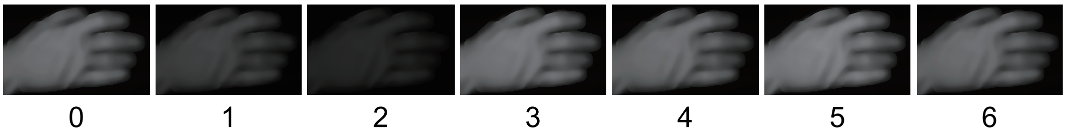

Figure 29. The result of hand with glove. Estimated albedo [pose 2] is shown, which is smooth enough.

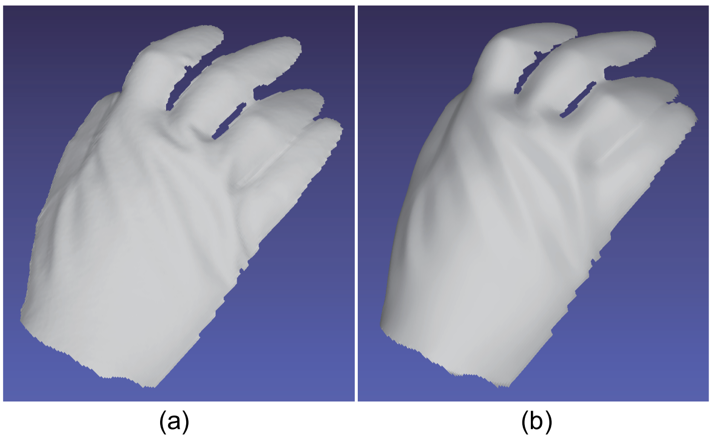

Figure 30. The result of hand with glove. Estimated geometry [pose 2]: (a) Estimated geometry [naive color photometric stereo] and (b) estimated geometry [our method]. Both color photometric stereos can estimate the surface normal of the dynamically deforming object.

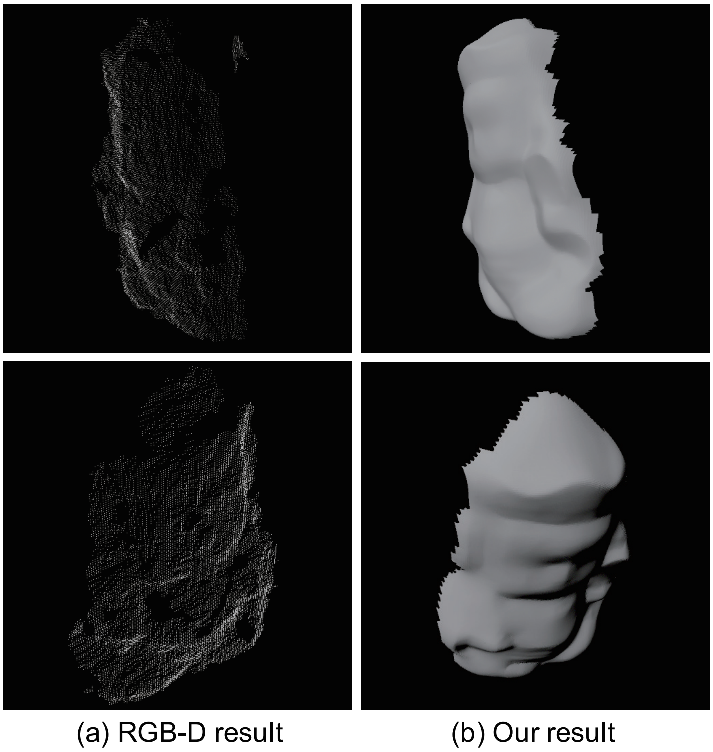

Figure 31. Comparison with off-the-shelf depth sensor: (a) Result of off-the-shelf depth sensor and (b) result of our method. The depth sensor can estimate the 3D coordinate of vertices successfully and the photometric stereo can estimate the surface normal successfully.

Figure 32 (c), since the smoothness constraint of surface normal is stronger. Figure 32 (a) is smoother than Figure 32 (c), and Figure 32 (b) is smoother than Figure 32 (d), since the albedo is not smooth, which means that the surface normal becomes relatively smooth. Although Figures 17, 20, 23, 26, and 29 show over-smoothed result of albedo, it is an adequate way to smooth the albedo in order to obtain sharp features of surface normal.

Figure 33(a) shows the initial value of the surface normal, and Figure 33 (b)–(c) shows how the surface normal is updated. This figure proves that our algorithm is stable since it converges quickly.

As shown in Figure 14, our method is robust to multiple types of albedos. On the other hand, as shown in Figures 16–30, our method over-smoothens the detailed surface structure. The generalized color photometric stereo problem shown in Equation (6) has nine unknown parameters; however, naive color photometric stereo has three unknown parameters, as stated in Section 4.2. Naive color photometric stereo robustly estimates the surface normal of white shirts, white dresses, etc. For multiple albedos, we have to tackle the ill-posed problem shown in Equation (6). Before starting this project, we had planned to use other constraints such as a so-called “integrability constraint.” However, we have chosen the smoothness constraint for constraining the problem since the integrability constraint solely cannot solve the problem. Surface normal $\mathbf{n}$ can be expressed as the gradients p and q (Equation (35)). Equation (6) can be rewritten as follows.

$$\begin{aligned} I_0(x,y) &= f\left(A_0(x,y), p(x,y), q(x,y)\right) , \\ &\vdots \\ I_6(x,y) &= f\left(A_6(x,y), p(x,y), q(x,y)\right) . \end{aligned} \tag{51}$$

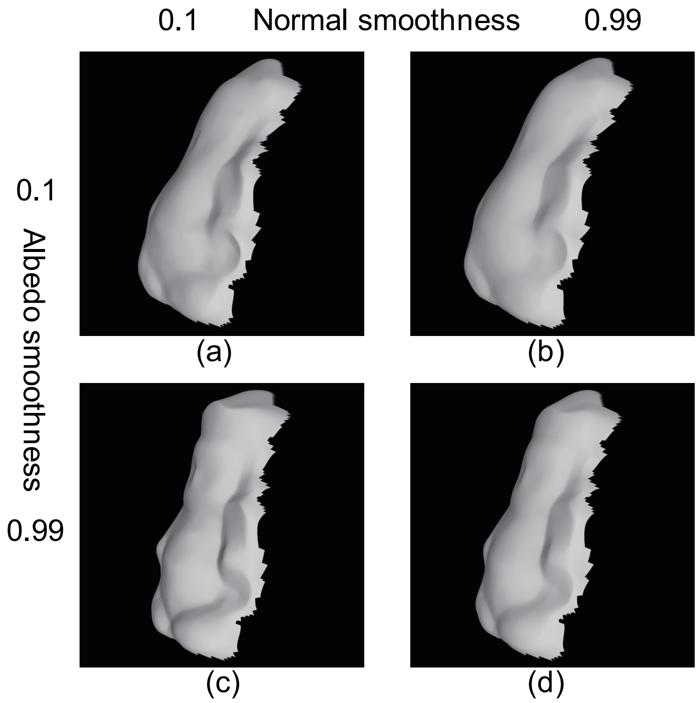

Figure 32. How the weight of smoothness term affects the results: (a) Sharp normal and sharp albedo, (b) smooth normal and sharp albedo, (c) sharp normal and smooth albedo, and (d) smooth normal and smooth albedo.

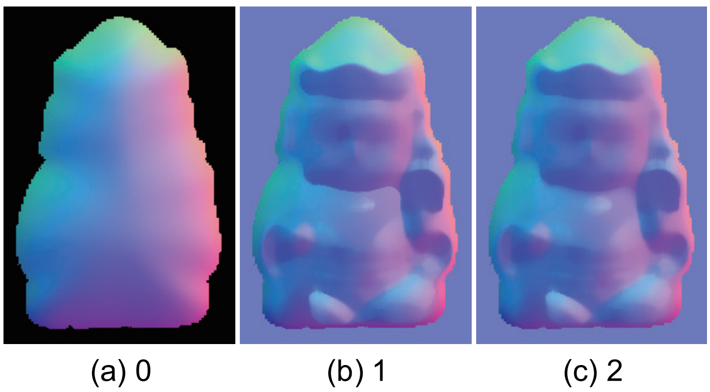

Figure 33. Intermediate state of surface normal through the proposed method: (a) Initial value of the surface normal, (b) the surface normal after 1 iteration, and (c) surface normal after 2 iterations. The proposed method is stable and coverges fast.

Namely, we have 9 unknowns ($A_0, \ldots, A_6$, p, and q) and 7 equations per pixel. Smoothness constraint for p and q can be represented as follows.

$$\begin{aligned} p(x,y) &= \frac{1}{4}\left(p(x,y-1)+p(x-1,y)+p(x+1,y)+p(x,y+1)\right) , \\ q(x,y) &= \frac{1}{4}\left(q(x,y-1)+q(x-1,y)+q(x+1,y)+q(x,y+1)\right) . \end{aligned} \quad (52)$$

Since there are additional two constraints per pixel which results in 9 equations per pixel, we can solve the problem. Integrability constraint can be represented as follows.

$$p(x,y+1) - p(x,y) = q(x+1,y) - q(x,y) . \quad (53)$$

Since only one constraint is added per pixel, we cannot determine 9 parameters from 8 equations. This is the reason why we use smoothness constraint rather than integrability constraint.

The over-smoothing problem is an unavoidable effect in the current approach, which relies on Equation (6). Our future work is to drastically change our approach such that it does not depend on Equation (6). We have to fundamentally consider the basic theory in order to improve the performance of color photometric stereo.

5. Conclusion

In this study, surface normal estimation of multicolored objects was conducted by the multi-spectral color photometric stereo method using median filter and occluding boundary. Note that the conventional color photometric stereo method is an ill-posed problem. Constraining the surface normal and albedo using median filter sucessfully solved this problem. In addition, we used the approximate shape calculated from the occluding boundary as the initial guess to the surface normal. Finally, we assembled measurement hardware that illuminates the object with seven different spectra and captured the image by a seven-band multispectral camera.

As discussed in Section 4.3, our method faces several problems in terms of both hardware and software. These problems cannot be solved with a minor update, so we need a drastic change for further improvement. In the future, we will disassemble the current measurement hardware and create a more useful system. For example, in order to make the hardware robust to shadow, it is better to add more lights and observe the scene with multispectral camera with more than 7 channels. The current method used one point light per channel, however, using area light is one choice for improvement in order to avoid the shadows. Polarization filter is also useful to remove the specular reflection. Additional future work is to reconsider the basic theory and fundamentally reorganize the approach of the algorithm. In order to apply the method to non-Lambertian BRDF, it is useful to measure the database of actual object with proposed system and train them using deep learning or other machine learnings. Database of spectral reflectance of various object decreases the number of unknowns which can make the problem well-posed. Using addtional sensors such as RGB-D camera is also interesting.

Author Contributions: Conceptualization, D.M.; methodology, D.M.; software, D.M. and Y.O.; validation, D.M., Y.O. and S.H.; formal analysis, D.M., Y.O. and S.H.; investigation, D.M. and S.H.; resources, D.M. and S.H.; data curation, D.M. and Y.O.; writing–original draft preparation, D.M. and Y.O.; writing–review and editing, D.M. and S.H.; visualization, D.M. and Y.O.; supervision, D.M. and S.H.; project administration, D.M. and S.H.; funding acquisition, D.M. and S.H.

Funding: This research was supported in part by MEXT KAKENHI Grant Number 15H05925, in part by MEXT KAKENHI Grant Number 15H02727, and in part by JSPS KAKENHI Grant Number 24700176.

Acknowledgments: The authors thank Karin Tsuda and Kenji Ishiguro for engaging in the previous version of this work. The authors thank Ryo Furukawa, Masashi Baba, and Yu Kawamoto for useful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. R. Anderson, B. Stenger, and R. Cipolla, “Color photometric stereo for multicolored surfaces,” In *International Conference on Computer Vision*, pp. 2182-2189, 2011.
2. M. B ahr, M. Breuß, Y. Quéu, A. S. Boroujerdi, and J.-D. Durou, “Fast and accurate surface normal integration on non-rectangular domains,” *Computational Visual Media*, vol. 3, no. 2, pp. 107–129, 2017.
3. K. Barnard, F. Ciurea, and B. Funt, “Sensor sharpening for computational color constancey,” *J. Opt. Soc. Am. A*, vol. 18, no. 11, pp. 2728–2743, 2001.
4. G. J. Brostow, B. Stenger, G. Vogiatzis, C. Hernández, and R. Cipolla, “Video normals from colored lights,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, no. 10, pp. 2104–2114, 2011.
5. A. Chakrabarti and K. Sunkavalli, “Single-image RGB photometric stereo with spatially-varying albedo,” In *International Conference on 3D Vision*, pp. 258–266, 2016.
6. M. S. Drew, “Reduction of rank-reduced orientation-from-color problem with many unknown lights to two-image known-illuminant photometric stereo,” In *Proceedings of International Symposium on Computer Vision*, pp. 419–424, 1995.
7. M. S. Drew, “Direct solution of orientation-from-color problem using a modification of Pentland’s light source direction estimator,” *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 286–299, 1996.
8. M. S. Drew and M. H. Brill, “Color from shape from color: a simple formalism with known light sources,” *Journal of the Optical Society of America A*, vol. 17, no. 8, pp. 1371–1381, 2000.
9. M. Drew and L. Kontsevich, “Closed-form attitude determination under spectrally varying illumination,” In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 985–990, 1994.
10. G. D. Finlayson, “Spectral sharpening: sensor transformations for improved color constancey,” *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1553–1563, 1994.
11. R. T. Frankot and R. Chellappa, “A method for enforcing integrability in shape from shading algorithms,” In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, 1988.
12. G. Fyffe, X. Yu, and P. Debevec, “Single-shot photometric stereo by spectral multiplexing,” *IEEE International Conference on Computational Photography*, pp. 1–6, 2011.
13. P. F. U. Gotardo, T. Simon, Y. Sheikh, and I. Mathews, “Photogeometric scene flow for high-detail dynamic 3D reconstruction,” In *IEEE International Conference on Computer Vision*, pp. 846–854, 2015.
14. C. Hernandez, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla, “Non-rigid photometric stereo with colored lights,” In *IEEE International Conference on Computer Vision*, pp. 8, 2007.
15. C. Hernández, G. Vogiatzis, and R. Cipolla, “Shadows in three-source photometric stereo,” In *ECCV 2008*, pp. 290–303, 2008.
16. B. K. P. Horn and M. J. Brooks, “The variational approach to shape from shading,” *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174–208, 1986.
17. K. Ikeuchi and B. K. P. Horn, “Numerical shape from shading and occluding boundaries,” *Artificial Intelligence*, vol. 17, no. 1–3, pp. 141–184, 1981.
18. H. Jiao, Y. Luo, N. Wang, L. Qi, J. Dong, and H. Lei, “Underwater multi-spectral photometric stereo reconstruction from a single RGBD image,” In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 1–4, 2016.
19. R. Kawakami and K. Ikeuchi, “Color estimation from a single surface color,” In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 635–642, 2009.
20. H. Kim, B. Wilburn, and M. Ben-Ezra, “Photometric stereo for dynamic surface orientations,” In *European Conference on Computer Vision*, LNCS 6311, pp. 59–72, 2010.
21. L. Kontsevich, A. Petrov, and I. Vergelskaya, “Reconstruction of shape from shading in color images,” *Journal of Optical Society of America A*, vol. 11, pp. 1047–1052, 1994.
22. A. Landstrom, M. J. Thurley, and H. Jonsson, “Sub-millimeter crack detection in casted steel using color photometric stereo,” In *International Conference on Digital Image Computing: Techniques and Applications*, pp. 1–7, 2013.
23. F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, “Geometrical considerations and nomenclature of reflectance,” In *Radiometry*; L. B. Wolff, S. A. Shafer, and G. Healey; Jones and Bartlett Publishers, Inc.: USA, 1992; pp.940–145.

24. M. R. Oswald, E. T oppe, and D. Cremers, "Fast and globally optimal single view reconstruction of curved objects," In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 534–541, 2012.
25. A. P. Petrov and L. L. Kontsevich, "Properties of color images of surfaces under multiple illuminants," *J. Opt. Soc. Am. A*, vol. 11, no. 10, pp. 2745–2749, 1994.
26. Y. Quéau, R. Mecca, J.-D. Durou, "Unbiased photometric stereo for colored surfaces: A variational approach," In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4359–4368, 2016.
27. Y. Quéau, R. Mecca, J.-D. Durou, and X. Descombes, "Photometric stereo with only two image: A theoretical study and numerical solution," vol. 57, pp. 175–191, 2017.
28. Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Normal integration: A survey," *Journal of Mathematical Imaging and Vision*, vol. 60, no. 4, pp. 576–593, 2018.
29. S. Rahman, A. Lam, I. Sato, and A. Robles-Kelly, "Color photometric stereo using a rainbow light for non-Lambertian multicolored surfaces," In *Asian Conference on Computer Vision*, pp. 335–350, 2015.
30. N. Roubtsova and J. Y. Guillemaut, "Colour Helmholtz stereopsis for reconstruction of complex dynamic scenes," In *International Conference on 3D Vision*, pp. 251–258, 2014.
31. W. M. Silver, "Determining shape and reflectance using multiple images," *Master's thesis*, Massachusetts Institute of Technology, 1980.
32. J. Sun, M. Smith, L. Smith, S. Midha, J. Bamber, "Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities," *Image and Vision Computing*, vol. 25, pp. 1050–1057, 2007.
33. G. Vogiatzis and C. Hernández, "Practical 3d reconstruction based on photometric stereo," pp. 313–345. In: R. Cipolla, S. Battiato, and G. M. Farinella, "Computer Vision: Detection, Recognition and Reconstruction," *Springer Berlin Heidelberg*, Berlin, Heidelberg, 2010.
34. G. Vogiatzis and C. Hernandez, "Self-calibrated, multi-spectral photometric stereo for 3D face capture," *Int. J. Comput. Vis.*, vol. 97, pp. 91–103, 2012.
35. R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.
36. R. J. Woodham, "Gradient and curvature from photometric stereo including local confidence estimation," *Journal of the Optical Society of America*, vol. 11, pp. 3050–3068, 1994.