Daisuke Miyazaki, Kenji Hara, Katsushi Ikeuchi,
"Photometric Stereo beyond Glass: Active Separation of Transparent Layer and Five-Light
Photometric Stereo with M-Estimator Using Laplace Distribution for a Virtual Museum,"
Proceedings of International Workshop on Photometric Analysis for Computer Vision,
Rio de Janeiro, Brazil, 2007.10

http://www.cvl.iis.u-tokyo.ac.jp/~miyazaki/
Photometric Stereo beyond Glass:
Active Separation of Transparent Layer and Five-Light Photometric Stereo with M-Estimator Using Laplace Distribution for a Virtual Museum

Daisuke Miyazaki†, Kenji Hara††, and Katsushi Ikeuchi†
†The University of Tokyo
Institute of Industrial Science, Komaba 4–6–1, Meguro-ku, Tokyo, 153–8505 Japan
††Kyushu University
Faculty of Design, Shiobaru 4–9–1, Minato-ku, Fukuoka, 815–8540 Japan

Abstract

One of the necessary techniques for constructing a virtual museum is to digitize the artwork arranged inside a glass or acrylic display case without bringing the artwork out of the display case. By using photometric stereo, we estimate the shape (surface normal) and the reflectance (albedo) of the artwork arranged inside a transparent display case. If we illuminate the display case, the light will reflect at its surface; thus, we cannot apply conventional photometric stereo as is. In this paper, we propose a five-light photometric stereo that estimates the shape and the reflectance of an object that has specularities under the circumstances that the light is reflected at the surface of the display case.

1. Introduction

It is a benefit to society to allow people to become familiar with precious artworks in museums by broadcasting these objects through the Internet or mobile phones. We can do this by digitally archiving these works. However, these precious objects are exhibited inside glass display cases and are not allowed to be removed from these cases. We propose a method to estimate the surface normal and the albedo of these objects without removing them from the display case.

If we take a photograph of such objects in a well-lighted museum, we also observe the scene reflected by the display case. Many researchers have proposed methods to separate such reflections [1–9]. To add to the complication of separating the object from its reflection, if we illuminate the object by a lamp, the light is also reflected at the surface of the glass. In this paper, we partially remove the reflection at the glass surface, and estimate the surface normal and the albedo of the object by using photometric stereo.

Our method assumes that the target object has both diffuse reflection and specular reflection.

The flow of the proposed algorithm will go as shown in Fig. 1. First, we take multiple images of the object under a different single light source. The image includes not only the object and the background but also the light and the foreground (Fig. 2). In this paper, we remove the foreground and the background by using a simpler method than previous methods. However, after this process, the output image still includes the reflection of the light as well as the object; thus, we cannot apply conventional photometric stereo. Therefore, we use so-called four-light photometric stereo [10–12].

Some photometric stereo methods use multiple images to enhance the quality of the output. Hayakawa [13] proposed...
2. Active Separation of Transparent Layers

As is shown in Fig. 2, the observed image includes foreground, background, light, and object. The scene in front of the display case will be reflected, and is observed as a foreground. The light is also reflected at the display case. We also observe the scene behind the display case, defined as a background. Also, the object is affected by ambient light. In addition, some noise occurs due to the dark current of the camera. Moreover, some amount of the light reaches to the background. By considering these effects, the observed image can be reformulated as follows:

\[ I_{\text{on}} = I_{\text{foreground}} + I_{\text{background}} + I_{\text{background.light}} + I_{\text{Object,light}} + I_{\text{Object,ambient}} + I_{\text{noise}}. \]  

(1)

We assume the camera noise can be represented as a white noise (Gaussian noise). Therefore, by taking multiple images and calculating the average, the noise can be represented as a constant value.

Next, we take an image with the lamp switched off. In this case, the observed light will be

\[ I_{\text{off}} = I_{\text{foreground}} + I_{\text{background}} + I_{\text{Object,ambient}} + I_{\text{noise}}. \]  

(2)

Then, if we calculate the difference between Eq. (1) and Eq. (2), we obtain

\[ I_{\text{on}} - I_{\text{off}} = I_{\text{light}} + I_{\text{Object,light}} + I_{\text{background.light}}. \]  

(3)

The background does not affect the appearance of the object; thus, only the object and the light are obtained at the pixels in object region. We can find out that the image represented by Eq. (3) does not have noise, effect of ambient light, foreground, or background. The light might reach the foreground by reflecting at the transparent surface of the display case; however, this can be ignored because the light is attenuated after traveling a long distance. The result of the separation is shown in Fig. 4.

In the field of shape-from-shading research, calculating the difference between the light-on image and the light-off image is used as a preprocessing for removing the dark current and the ambient light. Effectively using the light-on image and the light-off image is simple but beneficial, and it is also used aggressively in recent research [24–28]. Therefore, we use this method for separation.

The image represented by Eq. (3) contains not only the object but also the light; thus, we cannot apply conventional photometric stereo. The direct reflection of the light on the display surface is too bright, and the ordinary polarizers cannot completely remove the reflection. In our experiment, we avoid saturation of the diffuse reflection component; however, the direct reflection of the light often causes saturation. The separation procedure works successfully even for saturated pixels. Section 3 presents a method to estimate the surface normal and the albedo even if reflection of the light is saturated. If we need to analyze the specular reflection of the object, we carefully avoid its saturation.
3. Photometric Stereo with Five Lamps

3.1. Five-Light Photometric Stereo

Photometric stereo is effectively used for digital archiving of artworks [29, 30]. Conventional photometric stereo cannot be applied to an object that has specular reflection; moreover, it cannot be applied if there is a reflection at the glass case.

In this paper, we follow the idea of the four-light photometric stereo [10–12]. Conventional photometric stereo solves the equation from three images taken under three different light sources. Let us increase the number of the light sources to four. Four combinations can be calculated if we choose three intensities from four intensities. From the chosen three intensities, the conventional photometric stereo can estimate the surface normal. Therefore, there are four candidates for the surface normal at each pixel. If all four intensities are caused by diffuse reflection, the four surface normals will be close together. If one of the intensities includes specular reflection, the four surface normals will be different. In this case, the surface normal can be calculated from the darkest three intensities. This is the basic idea of the four-light photometric stereo proposed by Coleman and Jain [10].

Considering the reflection at the glass case, five lights will be enough for the measurement (Fig. 5). We will explain the detail below by using a Gaussian sphere viewed from above (Fig. 6). Five lights are placed as an equilateral pentagon. Fig. 6(a)–(d) are the Gaussian spheres when the angle between the light and the camera is (a) 30°, (b) 45°, (c) 60°, and (d) 75°. The red circle represents the specular reflection of the object’s surface. The numbers represent the numbers of the lights that illuminate the surface. In case “30°,” all regions are illuminated by three or more lights; thus, the surface normal can be calculated in all regions. The specular reflection of the object’s surface appears in region “5”; thus, we have four input data sets to calculate the surface normal. Even if the specular reflection of the glass display appeared in region “5,” we have three input data sets to calculate the surface normal. Region “2,” which is illuminated only by two lights, will appear in case “45°.” However, the observed area is small, and we can interpolate from neighboring pixels. The area of region “2” becomes larger in case “60°.” Consequently, if the angle between the light source and the viewing direction is less than 45°, we can estimate the surface normal and the albedo from five lights.

The above discussion considers the attached shadow,
while the cast shadow is not considered. In addition, we do not know which region each pixel belongs to, when the surface normal is unknown. Therefore, we propose an iterative algorithm in Section 3.2.

Conventional binocular stereo cannot be applied because the specular reflection interferes with finding the corresponding points. Our method treats the specular reflection and the shadow as outliers. Therefore, our method can estimate the surface normal and the albedo even if the specular pixels are saturated. There are many methods to separate diffuse reflection and specular reflection; however, most of them cannot be applied if the pixel is saturated, and most of them cannot be applied if the object is white. We assume that the direction of the light sources is known, thus, we can estimate the orientation of the glass case and detect the reflection of the display case; however, we do not estimate it. Even if we detect the reflection of the display case, we cannot detect the specular reflection of the object and the shadow, thus, we propose the outlier removal algorithm.

3.2. Robust Iterative Computation

3.2.1 M-Estimation Using Laplace Distribution

We obtain \( L \) intensities per pixel from \( L \) input images. \( L \) intensities sometimes include the specular reflection or the shadow. Therefore, we adopt a strategy to calculate the optimal normal and albedo through iterative computation.

First, we explain the smoothness constraint of the albedo. By minimizing
\[
\varepsilon_{\rho,\text{avg}} = (\rho_x(x, y))^2 + (\rho_y(x, y))^2 ,
\]
we obtain the following albedo
\[
\rho(x, y) = \text{average}\left(\rho(x + 1, y), \rho(x - 1, y), \rho(x, y + 1), \rho(x, y - 1)\right) .
\]
Eq. (5) is derived by differentiating Eq. (4) with \( \rho \) at pixel \((x, y)\) fixing other pixels, and setting “=0.”

Eq. (4) uses L2-norm, which means that it uses Gaussian distribution as M-estimator [31]. This M-estimator is not robust to outliers. We use Laplace distribution (double exponential distribution) as M-estimator, which is more robust than Gaussian distribution [31]. In this case, the smoothness constraint of the albedo is represented as the following L1-norm.
\[
\varepsilon_{\rho,\text{med}} = |\rho_x(x, y)| + |\rho_y(x, y)| .
\]
The albedo \( \rho \) which minimizes this equation would be
\[
\rho(x, y) = \text{median}\left(\rho(x + 1, y), \rho(x - 1, y), \rho(x, y + 1), \rho(x, y - 1)\right) .
\]
Eq. (7) is derived by differentiating Eq. (6) with \( \rho \) at pixel \((x, y)\) fixing other pixels, and setting “=0”. We present the proof in the Appendix.

Next, we explain the optimization of the albedo. We represent the \( L \) input images with the subscript \( i = \{1, \ldots, L\} \).

Since there are some outliers like shadows, we use Laplace distribution as the M-estimator.
\[
\varepsilon_{\rho,\text{opt}} = \sum_{i=1}^{L} \left| \rho - \frac{I_i}{\cos L_i \cdot N} \right| .
\]
By fixing the surface normal \( N \), the albedo \( \rho \) which minimizes Eq. (8) will be given as follows.
\[
\rho = \text{median}\left(\frac{I_i}{\cos L_i \cdot N} \mid i = 1, \ldots, L\right) .
\]

Finally, we explain the optimization of the surface normal. In order to calculate the surface normal by photometric stereo, we need at least three data sets. The number of all combinations for selecting the three data sets from \( L \) data is \( L_{C_3} = \binom{L}{3} \). We represent each surface normal estimated from each combination as \( N_m(m = 1, \ldots, L_{C_3}) \). Each surface normal is calculated by conventional photometric stereo [32]. We estimate the surface normal by minimizing the following formula.
\[
\varepsilon_{N,\text{opt}} = \sum_{m=1}^{L_{C_3}} |N - N_m| .
\]
By minimizing Eq. (10), we obtain the following solution.
\[
N = \text{median}\{N_m \mid m = 1, \ldots, L_{C_3}\} .
\]

3.2.2 Proposed Algorithm

Let us organize the algorithm. First, we set the appropriate initial values for the surface normal and the albedo. Next, we calculate all \( L_{C_3} \) candidates of the surface normal for all pixels. Then, we iterate the following process until convergence.

1. We estimate \( L \) numbers of albedo from the current surface normal. We also add the albedos of four neighborhoods. We calculate the median \( \rho_{\text{opt}} \) of these albedos.
2. We calculate the four-neighbor average of the albedo, \( \rho_{\text{avg}} \). We calculate the new albedo from the weighted average of \( \rho_{\text{opt}} \) and \( \rho_{\text{avg}} \).
3. We calculate the median \( N_{\text{opt}} \) of the \( L_{C_3} \) candidates of surface normal and the surface normals of four neighborhoods.
4. We calculate the four-neighbor average of the surface normal, \( N_{\text{avg}} \), and calculate the new surface normal from the weighted average of \( N_{\text{opt}} \) and \( N_{\text{avg}} \). Finally, the surface normal is integrated to the height map using the natural boundary condition [33]. The surface normal and the albedo are updated using the four-neighbor information in order to avoid falling into a local minimum. RANSAC is a famous technique to remove the outliers; however, if the input images are small (i.e., if \( L \) is small), we can calculate all \( L_{C_3} \) candidates and remove the outliers more robustly. If the number of the images are large, calculating all candidates takes a long time.
$C_3 = \frac{(L-1)(L-2)}{6} \approx O(L^3)$; thus, RANSAC is more effective than our approach in such a case.

4. Experimental Result

4.1. Experimental Setup

The experimental setup is shown in Fig. 7. The target object is covered with a glass or acrylic display case, and is observed with one camera and one lamp. The camera is fixed, and we take the image with three-band RGB. The light source direction and its intensity are assumed to be known. We move the light source, and take four or more images for a diffuse-only object, and take five or more images for an object that includes specularities. For each shot, we take both the image with the light switched on and the image with the light switched off. Each face of the display case should be transparent, and its thickness and orientation are unknown. However, we assume that the display case is a box-type object structured by transparent planes with uniform thickness.

4.2. Experimental Result

Fig. 8 is the input image and Fig. 9 is the separated result. Fig. 10(a)–(c) are the estimated albedo, height, and shape, respectively. Fig. 11 is an example image rendered by using estimated parameters with a different viewpoint. In Fig. 10(c), we did not remove the background; however, it can be easily removed by simply thresholding the dark intensity of the background, or using background cut techniques such as Lazy Snapping [34], GrabCut [35], or Adobe Photoshop CS3. Both results in Fig. 10 are obtained from five input images.

Fig. 12 shows the result of applying conventional photometric stereo using only three images under three different lights. The conventional photometric stereo cannot handle the reflection at the display case or the specular reflection of the object surface; thus, it is impossible to estimate the true surface normal and albedo. The conventional photometric stereo is not robust; thus, it is affected by the error in the intensity of the light source. Calibrated photometric stereo assumes the light source direction and the intensity are known. Although we compensated for these factors before the experiment, some amount of the error of the compensation still remain as a noise.

Fig. 13 shows the result of light-stripe range scanning (active stereo). We scan from exactly the front of the acrylic
Figure 12. Result of conventional photometric stereo: (a) Albedo, (b) shape.

Figure 13. The result of light-stripe range scanning: (a) Setup, (b) estimated shape.

box with a Konica Minolta VIVID 910. Owing to the reflection from the acrylic surface, the sensor cannot obtain the true shape. However, the sensor can possibly determine the shape if it is observed on a slant. The proposed method can estimate the shape by observing the object not only from the front of the case but also from a slanting direction.

Fig. 14(b) shows the result of our method applied to an object 15cm tall, from six input images. Fig. 14(a) is a ground truth obtained by scanning the object brought out from the acrylic case by VIVID910. Fig. 14(c) is the result of conventional photometric stereo from three certain input images. Fig. 14(d) is the same as Fig. 13. The comparison in Table 1 shows that the proposed method is more effective than the other methods. Fig. 14(e)(3) represents the height error, and we can find out that the estimated shape is bent. The distortion is caused when the height is calculated from the surface normal. A possible solution would be to combine the proposed method with a stereo method. Fig. 14(e)(4) represents the surface normal error, and we find out that the error is caused at the concave part. This error is caused by the cast shadow, the interreflection, and the smoothness constraint. A possible solution would be to remove the cast shadow from many light sources or to apply interreflection-removing algorithms.

5. Conclusion

In this paper, we propose a method for digitizing artworks to be included in a virtual museum. The proposed method can digitize artwork that is kept inside a glass display case as well as artwork that is taken out of the case. In this paper, we improve photometric stereo so that it will be suited for this application.

Figure 14. Comparison: (a) True value, (b) the result of the proposed method, (c) the result of conventional photometric stereo, (d) the result of light-stripe range scanning, (e) the error of the proposed method, (f) the error of conventional photometric stereo, (g) the error of light-stripe range scanning; (1) Azimuth angle (red: upper direction, blue: left-bottom direction, green: right-bottom direction) (2) zenith angle (blue: 0°, red: 90°) (3) the difference of the height (the brighter the noisier), (4) the difference of the surface normal (the brighter the noisier).

<table>
<thead>
<tr>
<th></th>
<th>Height error (RMSE)</th>
<th>Normal error (RMSE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light-stripe range scanning</td>
<td>2.43cm</td>
<td>36.0°</td>
</tr>
<tr>
<td>Conventional photometric stereo</td>
<td>1.32cm</td>
<td>28.7°</td>
</tr>
<tr>
<td>Proposed method</td>
<td>0.86cm</td>
<td>10.1°</td>
</tr>
</tbody>
</table>

Table 1. Evaluation.

One of our future goals is to estimate specular reflection parameters. We can render the diffuse reflection component as is shown in Fig. 15(b). We can obtain the specular image (Fig. 15(c)) by subtracting the rendered image (Fig. 15(b)) from the input image (Fig. 15(a)). The specular pixels were saturated in this experiment; however, we can estimate the specular reflection parameter if we adjust the exposure adequately.

Binocular stereo cannot be applied if there is a reflec-
References


A. Maximum Likelihood Estimation of Laplace Distribution

Eq. (6) can be discretized as follows:

$$
\varepsilon_{p, \text{med}} = |\rho(x+1, y) - \rho(x, y)| + |\rho(x, y) - \rho(x-1, y)| + |\rho(x, y+1) - \rho(x, y)| + |\rho(x, y) - \rho(x, y-1)| .
$$

(12)

By applying the following theorem in minimizing Eq. (12), we obtain Eq. (7).

**Theorem A.1** Suppose that we want to estimate the parameter $p$ from data $d_i$ by maximizing the following likelihood:

$$
Pr = \prod_{i=1}^{n} \frac{1}{2b} \exp \left( -\frac{|p - d_i|}{b} \right) .
$$

(13)

This probability density function is called Laplace distribution, and $b$ is called the scale parameter. The parameter $p$ of this solution will be the following:

$$
p = \text{median}(d_i | i = 1, \ldots, n) .
$$

(14)

**Proof.** By calculating the logarithm of Eq. (13), we obtain the following formula:

$$
pr = -n \log 2b - \frac{1}{b} \sum_{i=1}^{n} |p - d_i| .
$$

(15)

Therefore, maximizing Eq. (13) results in minimizing the following formula:

$$
\varepsilon = \sum_{i=1}^{n} |p - d_i| .
$$

(16)

Let us express the number of the data $d_i$ which is smaller than $p$ as $j$, and the number of the data $d_i$ which is greater than $p$ as $k$, where $j + k = n$. Therefore, Eq. (16) can be expressed as follows:

$$
\varepsilon = \sum_{i=1}^{j} (p - d_i) + \sum_{i=j+1}^{j+k} (d_i - p) ,
$$

(17)

by representing the subscript properly. By differentiating Eq. (17) by $p$, we obtain

$$
\frac{\partial \varepsilon}{\partial p} = \sum_{i=1}^{j} - \sum_{i=j+1}^{j+k} = j - k .
$$

(18)

By letting Eq. (18) be zero, we obtain $j = k$, namely, the number of the data $d_i$ smaller than $p$ and the number of the data $d_i$ greater than $p$ is the same. Therefore, Eq. (14) is obtained.