Usage of Needle Maps and Shadows to Overcome Depth Edges in Depth Map Reconstruction

Masaaki Iiyama, Koki Hamada, Koh Kakusho, and Michihiko Minoh
Kyoto University, Japan

Abstract

Photometric stereo is a method of recovering surface normals (needle map) from images. The surface integral of surface normals is used to reconstruct a depth map; however, the depth edges, which are discontinuous boundaries of the depth map, pose a problem for photometric stereo. When the surface of objects includes depth edges, the reconstructed depth map may contain errors. To solve this problem, we detect depth edges using shadows and compute a relative depth between two distant points using the widths of the corresponding shadows. We define an error function and reconstruct the depth map by minimizing the error function. Experimental results with synthetic and with real image data demonstrate the effectiveness of our approach.

1 Introduction

Photometric stereo is a method of recovering surface normals (needle map) from a set of images captured by a single camera; each image taken under different lighting conditions. Compared with other shape acquisition methods, this method has several advantages. For example, photometric stereo does not require stereo correspondence, which is a challenging problem when object surfaces are texture-less. Although Shape-from-Silhouette provides the same advantage, it requires multiple cameras. Photometric stereo is also advantageous in that it can recover surface normals and reflectance parameters, that are not recovered by other single-camera methods (e.g., laser scanning techniques).

When determining a depth map from a needle map, depth edges [5][3], which are discontinuous boundaries of the depth map, pose a problem. The depth map is determined using the surface integral of surface normals. Further, the correct location of the depth edges is necessary for correct integration. Without the location, a continuous depth map could be incorrectly generated in areas where the surface is discontinuous. Even if the position of a depth edge is known, an incorrect depth map can be generated when a closed region surrounded by the depth edges exists on the depth map, because a relative depth between the closed region and the rest of the depth map cannot be determined due to the surrounding depth edges. Vogiatzis et al. [7] proposed a method that reconstructs a complex shape regardless of the depth edges, but their approach required multiple cameras.

In this paper, we propose a single-camera method applicable to objects regardless of the depth edges.

We compute the relative depth using the widths of the corresponding shadows. A shadow made by a light ray is projected on an image as a line segment. Given the direction of the light ray and the two endpoints of the line segment, the relative depth between the two endpoints can be computed. If the line segment crosses the depth edges surrounding the closed region, one of the end points belongs to the closed region and the other belongs to the rest of the depth map. Thus, the relative depth between these two separated regions is computed as the relative depth between the endpoints.

Based on this idea, we reconstruct the depth map by minimizing two error functions. One of the error function is minimized when the reconstructed depth map is consistent with the needle map and the other is minimized when the reconstructed depth map is consistent with the relative depth computed from the shadows.

Several shadow-based methods have been proposed. Shadow carving [6] is a space-carving-like [4] method that computes a volume that is consistent with the observed cast shadows, and a method of Feris et al. [2] uses shadows to solve occlusion problem in stereo correspondence; however, these approach requires multiple cameras. A shape acquisition method using shadows instead of structured lights is proposed by Bouguet et al. [1], but this method could acquire neither surface normals nor reflectance parameters.
2 Depth map reconstruction

The camera projection model used for reconstructing the depth map is an orthographic projection model. The image plane exists on \( z = f \), and the pixel size is \( 1 \times 1 \); a pixel \((x, y) \in \mathbb{Z}^2\) exists on \((x, y, f)^T \in \mathbb{R}^3\) in the camera coordinate system.

At least three images captured under different lighting conditions are used for recovering a needle map. We recover the needle map by applying photometric stereo beforehand – i.e., the surface normal vectors at each pixel have already been acquired. Using Rasker’s stereo beforehand – i.e., the depth map is an orthographic projection model. The camera projection model used for reconstructing continuous depth maps when the object has discontinuities and detects the depth edges from the estimated location. The depth edge is expressed as a pair of neighboring pixels (see Figure 1).

![Figure 1. Depth edge](image)

Our goal is to reconstruct the depth map of a scene using the needle map, the depth edges, and the shadows of the scene.

**Depth map reconstruction using depth edges** Let \( \{u, v\} \in \mathbb{R}^3 \) represent a pair of neighboring pixels, \( w \in \mathbb{R}^3 \) be the midpoint between \( u \) and \( v \), \( \{U, V, W\} \in \mathbb{R}^3 \) be 3-dimensional surface points projected onto \( u, v \) and \( w \), and \( \{Z_U, Z_V, Z_W\} \in \mathbb{R} \) be depth values of \( U, V \) and \( W \) (see Figure 2). Supposing that \( U \) lies on a plane with normal vector \( n_U \in \mathbb{R}^3 \) and \( V \) on a plane with normal vector \( n_V \in \mathbb{R}^3 \), we obtain the following constraints; \( (W - U) \cdot n_U = 0 \), \( (W - V) \cdot n_V = 0 \). Applying the orthographic projection model, we obtain; \( U = u + (Z_U - f)e_z \), \( V = v + (Z_V - f)e_z \), \( W = w + (Z_W - f)e_z \), where \( e_z = (0, 0, 1)^T \). Combining the above five equations and \( w = (u + v)/2 \), we derive the following equation:

\[
Z_U - Z_V = \frac{(w - u) \cdot n_U}{e_z \cdot n_U} - \frac{(w - v) \cdot n_V}{e_z \cdot n_V}
\]  

(1)

Equation (1) is used to derive the relative depth, \( Z_U - Z_V \), from \( u, v, n_U \) and \( n_V \), and provides the error function that returns 0 when estimated \( \{Z_U, Z_V\} \) values are consistent with \( \{n_U, n_V\} \):

\[
T_{u,v} = (Z_U - Z_V - d_{ndl}(u, v))^2
\]

(2)

where \( d_{ndl}(u, v) \) is the right-hand term of Equation (1). Summing \( T_{u,v} \) over all pairs of neighboring pixels, we define the error function \( E_{ndl} \), which returns 0 when the reconstructed depth map is consistent with the needle map:

\[
E_{ndl} = \sum_{(u,v) \in J} T_{u,v}
\]

(3)

where \( J \) is a set of all neighboring pixel pairs.

![Figure 2. Depth from surface normal](image)

Minimizing \( E_{ndl} \), we obtain a depth map that is consistent with the needle map. However, \( E_{ndl} \) ignores the depth edges on the needle map, resulting in incorrect continuous depth maps when the object has discontinuous surfaces. Summing \( T_{u,v} \) over all pairs in \( J \) except the depth edges, a new error function \( E_{edg} \) is calculated as follows:

\[
E_{edg} = \sum_{(u,v) \in J \setminus C} T_{u,v}
\]

(4)

where \( C \) is a set of pixel pairs detected as the depth edges.

Minimizing \( E_{edg} \), we can properly reconstruct a discontinuous depth map; however, when closed regions surrounded by the depth edges exist in the depth map, ambiguity in determining relative depths between the closed regions and the rest of the depth map remains. Figure 4(a) and (b) shows a typical example of depth edges surrounding a square region. In this case, the ambiguity in determining relative depths between the square region surrounded by the depth edges and the rest of the depth map remains.

**Depth map reconstruction using shadows** To remove the ambiguity in determining the relative depth, we calculate the relative depth using shadow widths.
Suppose a ray whose direction is given by \( l \in \mathbb{R}^3 \) is interrupted at \( P \in \mathbb{R}^3 \) and casts a shadow line on the object’s surface, and the shadow line is terminated at \( Q \in \mathbb{R}^3 \) (see Figure 3). Let \( \{p, q\} \in \mathbb{R}^3 \) be the pixels that project \( P \) and \( Q \) onto themselves, and let \( \{Z_P, Z_Q\} \in \mathbb{R} \) be the depth values of \( P \) and \( Q \).

The vector \( Q - P \) is parallel to \( l \); that is, using \( \gamma \in \mathbb{R} \), the following constraint is derived:

\[
Q - P = \gamma l
\]  

Under the orthographic projection model, we obtain

\[
P = p + (Z_P - f)e_z, \quad Q = q + (Z_Q - f)e_z
\]

where \( e_z = (0, 0, 1)^T \). All the four points, \( P, Q, p, \) and \( Q \), are on the same plane, and the plane is parallel to the vectors, \( l \) and \( e_Z \). In addition, \( e_Z \) and the vector \( q - p \) are linearly independent, because \( p \) and \( q \) are on the same image plane, and \( e_Z \) is orthogonal to the image plane. Thus, \( l \) can be decomposed with two parameters \( \alpha \) and \( \beta \) as follows:

\[
l = \alpha(q - p) + \beta e_z
\]

Thus, given \( p, q, \) and \( l \), we can determine \( \alpha \) and \( \beta \).

Equations (5), (6), and (7) lead to the following equation:

\[
\gamma \beta = Z_Q - Z_P, \quad \gamma \alpha = 1
\]

Eliminating \( \gamma \), we obtain

\[
Z_Q - Z_P = \frac{\beta}{\alpha}
\]

Given \( p, q, \) and \( l \), equation (9) determines the relative depth \( Z_Q - Z_P \) and provides the non-negative error function that returns 0 when \( \{Z_P, Z_Q\} \) are consistent with the shadow length \( q - p \):

\[
S_{p,q} = Z_Q - Z_P - \frac{\beta}{\alpha}
\]

Summing \( S_{p,q} \) over all shadow lines, we define the error function \( E_{shd} \) as follows:

\[
E_{shd} = \sum_{(p,q) \in G} \left( (Z_Q - Z_P) - \frac{\beta}{\alpha} \right)^2
\]

where \( G \) is a set of pixel pairs extracted as shadow lines. Summing \( E_{edg} \) and \( E_{shd} \) with non-negative weight value \( \lambda \), we define the error function \( E_{all} \) as

\[
E_{all} = \lambda E_{edg} + E_{shd}
\]

Minimizing non-negative function \( E_{all} \), we reconstruct a depth map that is consistent with both the needle map and the shadows.

### 3 Experimental Results

Experimental results obtained using synthetic and real data are described in this section.

We first generate seven images of a torus and a plate (see Figure 5(a)) under seven different lighting conditions and determine the needle map, depth edges, and shadows. One of the seven input images is shown in Figure 5(b). Image resolution for all the image is 800 \( \times \) 600 pixels. We reconstruct the depth maps by minimizing \( E_{ndl}, E_{edg}, \) and \( E_{all}, \) and convert them into surface data. Figure 5(e), (e), and (f) show the reconstructed surfaces. The ground truth taken from the same viewpoint as in Figure 5(e), (e), and (f) is shown in Figure 5(c). As is evident from the figures, our methods cannot reconstruct occluded surfaces; that is, the bottom of the plate occluded by the torus could not be reconstructed. Figure 5(g), (h), and (i) show magnified images of Figure 5(e), (e), and (f) respectively.

Comparing Figure 5(e) and (e) reveals that minimizing \( E_{edg} \) reconstructs a more accurate depth map than minimizing \( E_{ndl} \). Minimizing \( E_{edg} \), however, reconstructs an incorrect surface as pointed by a white arrow in Figure 5(h), which is a closed region surrounded by depth edges. Figure 5(i) shows that our method (minimizing \( E_{all} \)) reconstructs the correct depth map regardless of the closed region.

Table 1 provides the mean error and error distribution values for \( E_{ndl}, E_{edg}, \) and \( E_{all} \). The errors are normalized by the overall depth of the object. For example, an error of 0.01 indicates that the difference between the estimated and correct depths is 1% of the overall depth of the object. Approximately 100% (99.64%) of pixels reconstructed by our method are within an error rate of 0.05, while 66.53% of pixels reconstructed by minimizing \( E_{edg} \) are within an error rate of 0.05.

Considering a real object, the depth map of a flower pot shown in Figure 6(b) is placed in front of a blue background (see Figure 6(a)) and reconstructed using our approach. Images, whose resolution is 800 \( \times \) 600 pixels, are captured under six different lighting conditions and are used for the reconstruction. Figure 6(e), (e), and (f) show the converted surfaces, and Figure
Table 1. Error distribution.

<table>
<thead>
<tr>
<th>Error ≤</th>
<th>$E_{ndl}$</th>
<th>$E_{edg}$</th>
<th>our method ($E_{all}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.184</td>
<td>0.0392</td>
<td>0.0112</td>
</tr>
<tr>
<td>0.02</td>
<td>5.550%</td>
<td>48.25%</td>
<td>55.67%</td>
</tr>
<tr>
<td>0.05</td>
<td>11.53%</td>
<td>63.65%</td>
<td>88.15%</td>
</tr>
<tr>
<td>0.05</td>
<td>37.77%</td>
<td>66.53%</td>
<td>99.64%</td>
</tr>
</tbody>
</table>

Figure 5. Results with synthetic data.

Figure 6. Reconstruction of a flower pot.

6(h), (h), and (i) show upper-side views that correspond to Figure 6(e), (e), and (f), respectively.

Results with $E_{ndl}$ and $E_{edg}$ show that the flower, which is located in front of the blue background, is reconstructed behind the background due to the depth edges on the boundary of objects. On the other hand, our method properly reconstructs the flower in front of the background regardless of the depth edges.

4 Conclusions

This paper proposed a depth map reconstruction method that uses shadows and surface normals. In our method, the relative depth between the two endpoints of a shadow line is calculated from the width of the shadow and is used to overcome the depth edge problem.

While accurate extraction of a shadow area is necessary for our method, we employ a simple image subtraction for shadow extraction. In future work, we will incorporate methods for robust and accurate shadow extraction.

References