Robust Depth Map Acquisition against Depth Edges with Silhouette Consistency

Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh

Academic Center for Computing and Media Studies, Kyoto University, Kyoto, 606-8501 Japan

SUMMARY

The authors propose a method to measure the smooth curved surface of an object and the depressed surface of an object which cannot be measured using the visual hull method by using a needle diagram obtained via photometric stereo. Errors occur in the distance map as a result of depth edges when recovering a distance map using a needle diagram. In order to reduce these errors, the silhouettes from other directions are used. The distance map restored without taking into consideration the depth edges often lacks consistency with the silhouettes obtained from other directions. The authors take advantage of this feature, define the energy representing the consistency between the distance map and the silhouette as well as the consistency between the distance map and the needle diagram, and then reduce the errors in the distance map by obtaining a distance map which minimizes this energy. Through experiments using the proposed method on simulation data and a real object, the authors evaluate the effectiveness of the proposed method. © 2007 Wiley Periodicals, Inc. Syst Comp Jpn, 38(14): 29–40, 2007; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/scj.20763

Key words: visual hull method; photometric stereo; depth map.

1. Introduction

Interest in readily measuring the shape of objects is increasing. For instance, this can be used to measure the shape of a particular person and then design clothes made to order or to reduce the development time and costs for the design and development of industrial products by measuring the shape of a mockup and then immediately converting it to CAD data.

In the past, the light-section method and spatial coding method [1] widely used as methods for measuring the shape of objects required slit lasers, pattern projection equipment, or other expensive apparatus. Moreover, the shape obtained from using such apparatus was a scale image seen from only one direction, and if the shape of all sides of the image was to be measured, then either more equipment or a rotating platform was needed. As was described above, when considering measuring the shape of an image in a simple fashion, a method that measures the shape of the image without using such costly equipment would be ideal. In order to measure a precise shape using the light division method, the laser scanning time increases, creating problems, and capturing the shape of an image which does not reflect the laser well is difficult. Even in the spatial coding method measuring a precise shape is a problem because of the need to make the projected pattern precise and to handle blurriness due to scattering of the projected light.

Given this situation, methods to measure the shape of an object using several low-cost cameras have gained attention in recent years. The multiple-baseline stereo method [2], space carving [3], and the visual hull method [4] are all
methods to measure the shape using several cameras. Among these methods, the multiple-baseline stereo method and the space carving method can match pixels among cameras using color information. When the surface of the object to be measured is textureless, however, the shape can no longer be acquired accurately due to errors in this matching. As mentioned in the examples above, use for textureless objects such as people and mockups would not be possible with these methods.

On the other hand, the visual hull method has the advantage of acquiring the shape of the object from silhouettes taken from several cameras, and being able to handle even textureless objects. However, the shape obtained using the visual hull method bulges from its contact with the object [5], and as a result measuring smooth curved surfaces on the object or depressed surfaces on the object is difficult or impossible. In the past, methods [6–9] to increase the number of points virtually by combining visual hulls obtained as a time series and then measure the smooth curved surfaces, or methods [3, 10, 11] to measure the depressed surfaces in the object by creating a correspondence among cameras based on color information were proposed for the problem above. In the method which combines visual hulls as a time series, the smooth surfaces can be measured, but the principal problem in the visual hull method of being unable to measure depressed surfaces in the object cannot be resolved. Moreover, methods that use color information have the problem of losing the advantages of the visual hull method which can be used for textureless objects.

Thus, in this research the authors focus on shade information of the object that can be observed using a camera and propose a method to measure the smooth curved surfaces of an object and the depressed surfaces of an object, which cannot be performed under the visual hull method, by using not only the object silhouette but also shade information. Using this method, the advantages of the visual hull method which can be used for textureless objects are preserved, while the smooth curved surfaces of an object and the depressed surfaces of an object, which cannot be measured under the visual hull method, can be measured.

Methods to measure shape using the shade of an object include the photometric stereo [12]. The photometric stereo acquires as a needle diagram the normal direction of the object surface based on image groups (three or more) obtained by providing illumination from three or more different directions. It can also be used for textureless objects, and can precisely acquire normal lines for object surfaces without having to create correspondences between images. The smooth curved surfaces of an object and the depressed surfaces of an object, which cannot be measured under the visual hull method, can be measured using the needle diagram.

Cho and Minamitani [13] and Chen and colleagues [14] have shown that the shape of an object can be measured by restoring the distance map based on the needle diagram obtained from the photometric stereo using several cameras, and the shape of an object can be measured by matching the position of the distance map obtained from each camera in three-dimensional space. The restoration of the distance map from the needle diagram is achieved by obtaining a distance map with a high consistency between the needle diagram and the normal line for the object surface calculated based on the distance map. Moreover, positioning the distance map is performed by finding the offset value for the distance map with a high consistency with the silhouette.

The presence of depth edges is a problem when restoring a distance map from a needle diagram. In places where the depth changes irregularly (depth edges), the normal lines cannot be calculated based on the distance map. As a result, consistency with the needle diagram fundamentally cannot be created. At this point, if the normal lines for the object surface are calculated using the continuous changes in the depth even where depth edges occur, but without taking the depth edges into consideration, then errors in the distance map appear near the depth edges.

Thus, in order to reduce the errors in the distance map caused by the depth edges, the authors considered using a silhouette from another direction. A distance map restored without taking the depth edges into consideration often lacks consistency with the silhouette obtained from another direction. In the authors’ method, this feature is exploited. The authors define the energy representing the consistency between the distance map and the needle diagram as well as the consistency between the distance map and the silhouette, and then obtain a shape that minimizes this energy. As a result, the consistency between the silhouette obtained from each camera is maintained, a distance map with a high level of consistency with the needle diagram is acquired, and the number of errors in the distance map due to the depth edges is reduced.

In the methods proposed by Cho and Minamitani and Chen and colleagues, which represent research related to the authors’, the distance map is acquired based on the needle diagram, and the shape is acquired by positioning the distance maps obtained from several cameras in three-dimensional space. In this related research, accurately positioning the distance maps has been evaluated, but the effects on the depth edges generated when obtaining the distance map have not been considered. On the other hand, the authors’ method is characterized by addressing the effects of the depth edges. In the authors’ method, the silhouette consistency is used not only for positioning the distance maps, but also to restore the distance map, which reduces the effect of the depth edges. Improvements in the
precision of shape acquisition for shapes of objects with depth edges can therefore be expected.

This paper is structured as follows. Section 2 describes the visual hull method and the photometric stereo which form the basis for the proposed method. In Section 3, the energy representing the consistency between the distance maps and the needle diagram as well as the consistency between the distance maps and the silhouettes is defined. In Section 4, the authors evaluate the effectiveness of their proposed method through experiments using the proposed method on real objects with eight cameras. Section 5 provides a summary.

2. Visual Hull Method and the Photometric Stereo

In this section the authors describe the visual hull method and the photometric stereo which form the basis for the proposed method.

2.1. The visual hull method

The visual hull method (Fig. 1) is explained here. This method acquires the three-dimensional shape of an object using the silhouettes obtained from several cameras.

First, based on the difference between the images of the object captured by each camera and the background images obtained without the object in position, the region (silhouette) where the object is projected in the images is extracted. Below, the number of cameras is indicated using \(C\), and the silhouette obtained using camera \(c\) \((c = 1, \ldots, C)\) is represented as \(S_c\).

Each camera is represented as a pinhole camera model, and if we assume that the projection matrix is known, then the center of the camera lens can be taken to be the endpoint, and the open space (visual cone) for each cone including the silhouette can be obtained for each camera. The volumetric space for each cone obtained from each camera is called a visual hull, and this becomes the restored shape obtained using the visual hull method.

If even a silhouette can be obtained for a textureless object in the visual hull method, then the shape can be acquired. However, the visual hull obtained using the visual hull method is a protrusion contained within the true shape of the object, and measuring depressed surfaces in the object is in principle impossible. Moreover, if a smooth surface on the object, even without any depressions, is to be measured, a large number of cameras is needed.

2.2. Photometric stereo

Let us consider the photometric stereo as a means to measure depressed surfaces or smooth surfaces on an object which are difficult to acquire using the visual hull method.

The photometric stereo acquires as a needle diagram the normal line for the object surface based on groups of images obtained using illumination from light sources in different positions while turning on and off these sources. If an image illuminated from three different directions can be obtained, then in principle the normal line direction can be estimated.

The reflective characteristics of the object surface and the following conditions for the light source must be set.

- The light source must be a planar source (or a point light source sufficiently distant from the object that it can be seen as a parallel light source).
- The direction of illumination from the light source must be known, and there must be at least three illumination direction vectors which are independent.
- There must not be any cast-shadows.
- The object surface must be a diffusely reflective surface.

For the camera \(c\), the number \(N_c\) of light sources providing illumination from the same side as camera \(c\) to the object and the illumination direction (vector from the object surface to the light source) \(l_j^c\) \((j = 1, \ldots, N_c)\) for each light source must be known.

The incident light from the illumination direction \(l_j^c\) is diffusely reflected on the object surface and observed by a camera. At this point, the intensity \(I_{ij}^c\) of the reflected light objected at the pixel \(m_i^c\) \((i = 1, \ldots, M_c)\) among the \(M_c\) pixels included in \(S_c\) can be represented as shown below, where
the intensity of the incident light striking the object surface from the light source is $L_j^c$, the normal line to the object surface observed at the pixel $m_j^c$ is $\mathbf{n}_j^c$, and the coefficient of diffused reflection is $\rho_i^c$:

$$I_{ij}^c = L_j^c \rho_i^c \frac{\mathbf{V}_j^c \cdot \mathbf{n}_j^c}{|\mathbf{V}_j^c|}$$  \hspace{1cm} (1)

Here, when we consider the matrix $\mathbb{I}$ with $M_c$ rows and $N_c$ columns created by arranging $I_{ij}^c$, $\mathbb{I}$ is comprised of the matrix $\mathbb{N}$ with $M_c$ rows and three columns including the normal line for the object surface, and the matrix $\mathbb{L}^T$ with three rows and $N_c$ columns including the illumination direction for the light source, as can be seen in Eq. (2). By breaking down $\mathbb{I}$ into $\mathbb{N}$ and $\mathbb{L}^T$, the normal line $\mathbf{n}_j$ for the object surface observed by the pixel $m_j^c$ can be obtained.

$$\mathbb{I} = \begin{pmatrix}
I_{11}^c & \cdots & I_{1N_c}^c \\
\vdots & \ddots & \vdots \\
I_{M_c1}^c & \cdots & I_{M_cN_c}^c
\end{pmatrix} = \mathbb{NL}^T$$

$$= \begin{pmatrix}
\rho_i^c \frac{\mathbf{V}_1^c \cdot \mathbf{n}_1^c}{|\mathbf{V}_1^c|} \\
\vdots \\
\rho_i^c \frac{\mathbf{V}_{M_c1}^c \cdot \mathbf{n}_{M_c1}^c}{|\mathbf{V}_{M_c1}^c|}
\end{pmatrix} \begin{pmatrix}
L_1^c & \cdots & L_{N_c}^c
\end{pmatrix}$$  \hspace{1cm} (2)

### 2.3. Distance map restoration based on the needle diagram

The photometric stereo has the advantage of obtaining a needle diagram even for textureless objects when the conditions for reflective characteristics for the object surface and light sources described above are satisfied. However, what is obtained using the photometric stereo is a needle diagram representing the normal lines, and not the shape of the object itself. As a result, the distance map must be restored from the needle diagram during postprocessing, and processing to acquire the shape of the object is required.

Note that the restoration of the distance map from the needle diagram is achieved by determining the distance map so that the consistency between the needle diagram and the normal lines for the object surface calculated from the distance map is high. At this point, because the shape is discontinuous in places where depth edges occur, the normal lines fundamentally cannot be calculated, and caution is required because consistency with the needle diagram cannot be created.

For instance, in Fig. 2(a), the part including the mascot’s body and the pole the mascot is holding is discontinuous. As a result, a normal line cannot be calculated where the body and pole are adjacent. Even on the needle diagram two different surfaces are projected to the same pixel where the body and the pole meet, and so the normal line resulting there includes errors. Here, if the body and the pole are taken to be continuous without taking the depth edges into consideration, then a shape in which the body and the pole are smoothly joined results.

### 3. Using the Consistency between the Silhouette and the Needle Diagram

In order to reduce the effects of the depth edges described in the previous section, the authors decided to take advantage of the silhouettes from other directions.

As an example, in Fig. 2(a) the places where the body and pole meet do not meet in an image [Fig. 2(b)] observed from a different direction. If the shape obtained by assuming that the surface that is fundamentally discontinuous is continuous using Fig. 2(a) is observed, then we can expect to be unable to create consistency with Fig. 2(b).

Thus, the consistency between the distance map and the needle diagram as well as the consistency between the distance map and the silhouette are each defined as energy functions. By minimizing the sum of these two energy functions, a distance map that is compatible with the needle diagram and the silhouette can be generated, and the effects of the depth edges can be reduced.

Below, the regular region in the image in which the pixel $m_j^c$ included in the silhouette $S_j$ predominates is defined as $((x_j^c, x_j^c + 1], [y_j^c, y_j^c + 1))$, and the point $(x_j^c, y_j^c)$ is referred to as the representative point for $m_j^c$. The distance map can be obtained by estimating the depth $Z(x_j^c, y_j^c)$ from the point $X_j^c$ on the object surface projected onto the representative point for the pixel $m_j^c$ to the focal point of the camera.

#### 3.1. Needle diagram consistency

First, let us consider the consistency between the distance map and the needle diagram.
We will assume that the depth $Z(x_i^c, y_i^c)$ for the pixel $m_i^c$ in the distance map and the normal line $n_i^c = (p_i^c, q_i^c - 1)^T$ for the object surface observed by that pixel are given (Fig. 3).

If the surface of the object projected onto $m_i^c$ is a flat surface at the normal line $n_i^c$, then the depth $Z_{\pm0}, Z_{\pm1}, Z_{-1}$ at the three points $(x_i^c + 1, y_i^c), (x_i^c, y_i^c + 1)$, and $(x_i^c + 1, y_i^c + 1)$ on the distance map can be expressed as follows using the depth $Z(x_i^c, y_i^c)$ at the representative point $(x_i^c, y_i^c)$ for the pixel, the normal line $n_i^c$, and the focal distance $f_c$ for the camera $c$:

\[
Z_{-1}(x_i^c + 1, y_i^c) = Z(x_i^c, y_i^c) \left(1 + \frac{p_i^c}{f_c}\right)
\]

\[
Z_{\pm0}(x_i^c, y_i^c + 1) = Z(x_i^c, y_i^c) \left(1 + \frac{q_i^c}{f_c}\right)
\]

\[
Z_{-1}(x_i^c + 1, y_i^c + 1) = Z(x_i^c, y_i^c) \left(1 + \frac{p_i^c + q_i^c}{f_c}\right)
\]

Given this, the depth corresponding to the point $(x_i^c, y_i^c)$ on the distance map can be obtained using four calculation methods: $Z(x_i^c, y_i^c)$, the depth calculated based on $(x_i^c - 1, y_i^c)$ using Eq. (3), the depth calculated from $(x_i^c, y_i^c + 1)$ using Eq. (4), and the depth calculated from $(x_i^c - 1, y_i^c - 1)$ using Eq. (5). Thus, the more the depths obtained using these four methods match, the better. Based on a high consistency between the distance map and the needle diagram, the energy representing the consistency between the distance map and the needle diagram can be defined using the following equation. Note that the greater the consistency, the lower the energy.

\[
E_N = \sum_{m_i^c \in S_c} \left( \left| Z(x_i^c, y_i^c) - Z_{\pm1}(x_i^c, y_i^c) \right|^2 + \left| Z(x_i^c, y_i^c) - Z_{\pm0}(x_i^c, y_i^c) \right|^2 \right)
\]

3.2. Silhouette consistency

Next let us consider the consistency between the distance map and the silhouette. The visual hull method described in Section 2.1 is a method for acquiring shapes that are compatible with the silhouette. The visual hull is present in the shape tangential to the object. Laurentini [5] pointed out the feature of it in which the silhouette and the projected image obtained by projecting the visual hull onto a camera match. A method [7] to create a correspondence between cameras using this feature and a method [15] to handle extraction errors for silhouettes necessary to create a visual hull have been proposed. The authors also use this feature of a visual hull and evaluate the consistency between the distance map and the silhouette in order to eliminate the restoration errors in the distance map caused by the depth edges.

3.2.1. Visual hull line constraints

Let us consider the lines that pass through the representative point in the boundary pixels with the focal point of the camera being an end point with respect to the pixels (boundary pixels) positioned at the boundary of the silhouette and the background among the pixels included in the silhouette (Fig. 4). Note that here, as the number of boundary pixels rises, the number of visual lines used as a constraint rises. Based on the idea that a more accurate shape can be acquired, the pixels in the silhouette which are not
among the eight pixels near the silhouette are taken to be boundary pixels.

If this visual line is projected to a different camera, then the projected image (the projected image is also a straight line) intersects with the silhouette in that camera. Here, let us consider the interval in which the projected image is included in the silhouettes for all other cameras among the visual lines that pass through the representative point in the boundary pixels. This is called the visual hull line. The visual hull line represents a portion of the surface of the visual hull (though there may be an error as large as one pixel).

Now, taking into consideration the relationship between the visual hull line and the shape of the object, the following two constraints (visual hull line constraints) obtain based on the feature that the projected image for the visual hull and the silhouette match.

- The object does not intersect with any visual hull lines.
- The object is in contact with at least one point on each visual hull line.

When the first constraint is not satisfied, the visual hull line represents the surface of the visual hull. Given this, the constraint that the visual hull be in contact with the object is not satisfied. Furthermore, when the second constraint is not satisfied, there is no object on the visual hull lines. In other words, the boundary pixels which comprise this line are not included in the silhouette, and there is no consistency between the visual hull and the silhouette.

Consistency between the distance map and the silhouette can be evaluated by using the extent to which the visual hull line constraints are satisfied.

### 3.2.2. Using the visual hull line constraints

Here, let us consider which visual hull lines impose constraints on the distance map. Among the visual hull lines, if some can be observed from a particular camera, then there are also others that are hidden by other visual hull lines, and are not observed by that camera. Whether or not constraints are imposed on the distance map obtained from that camera depends on to what extent the visual hull lines can be observed from that camera.

For a particular visual hull line, the visual hull line imposes both visual hull line constraints on the distance map in camera A when the lines overall can be observed from camera A [Fig. 5(a)]. On the other hand, when the visual hull lines cannot be observed from the camera [Fig. 5(c)], there are no constraints on the distance map. Moreover, when only a portion of the visual hull lines can be observed [Fig. 5(b)], the first constraint among the visual hull line constraints (the object does not intersect with any visual hull lines) is imposed on the distance map, but the second constraint (the object is in contact with at least one point on each visual hull line) allows for the object to be in contact at any place not observed by the camera among the visual hull lines. As a result, this cannot be used as a constraint for the distance map. Note that whether or not a visual hull line can be observed can be determined by creating a Z buffer by projecting all of the visual hull lines onto the image at once, and then performing a comparison with the Z buffer.

Based on the above, the authors used a comparison at each intersection of the depth to the object surface from the camera obtained from the distance map and the depth to the visual hull line from the camera and then evaluated whether or not the visual hull line constraints were satisfied and to what extent after projecting the visual hull lines onto a distance map and extracting the intersections (Fig. 5) with the pixel lattice in the distance map.

Let the intersections with the pixel lattice and the projected image for the portion that can be observed among the visual hull lines be \((v_{x1}, v_{y1}), \ldots, (v_{xK}, v_{yK})\). The depth \(Z(v_{xk}, v_{yk})\) from the camera at a particular intersection point \((v_{xk}, v_{yk})\) to the surface of the object can be calculated with the equation below using the depth \(Z([v_{xk}, v_{yk}])\) having a pixel for which \([v_{xk}, v_{yk}])\) is the representative point, and the normal line \(n = (p_k, q_k, -1)^T\) corresponding to that pixel:

\[
Z(v_{xk}, v_{yk}) = Z([v_{xk}, v_{yk}])
\cdot \left(1 + \frac{v_{xk} - \frac{[v_{xk}]}{f_c} p_k + \frac{v_{yk} - \frac{[v_{yk}]}{f_c}} q_k \right)
\]  \(7\)

As a result of the first constraint on the visual hull lines, the surface of the object immediately in front of the visual hull line for the part that can be observed from among the visual hull lines does not exist. Consequently, the depth of the visual hull lines is smaller than the depth to the surface of the object for all intersection points.

Moreover, as a result of the second constraint on visual hull lines, if the entirety of a visual hull line can be observed, then the depth to the object surface and the depth of the visual hull line match at at least one point among the intersection points.
If the difference in the depth from the camera to the surface of the object obtained using Eq. (7) and the depth from the camera at the intersection point \((v_{ik}, v_{ik})\) to the visual hull line is represented as \(\delta Z_{ab}\), and the minimum among \(\delta Z_{ab}\) is represented as \(\delta Z_{min}\), then based on the first constraint on the visual hull lines, \(\delta Z_{ab} \geq 0\) for all \(k\). Moreover, based on the second constraint on the visual hull lines, if the entire visual hull line can be observed, then \(\delta Z_{min} = 0\). Given the above, the energy based on the constraint that the visual hull lines impose on the distance map can be defined using the equation below:

\[
E_{VL} = \begin{cases} 
0 & \text{if } \Delta Z_{\text{min}} \geq 0 \text{ or a portion of the visual hull line cannot be observed} \\
\Delta Z_{\text{min}}^2 & \text{otherwise} 
\end{cases}
\]

The total for all the visual hull lines represents the energy based on the consistency between the distance map and the silhouette.

### 3.3. Minimizing the energy

The sum \(E_{all}\) of the energy based on the consistency between the distance map and the needle diagram defined in Section 3.1 and the energy based on the consistency between the distance map and the silhouette defined in Section 3.2 is minimized using the annealing method [16], and then the distance map is obtained:

\[
E_{all} = E_N + \lambda E_{VL}
\]

Here, \(\lambda\) represents the weight for the visual hull line constraint. By making \(\lambda\) a sufficiently large value, a distance map with as much consistency as possible with the needle diagram within the range in which the visual hull constraints are satisfied can be obtained. As a result, in areas in which depth edges occur, consistency with the silhouette is emphasized, and the depth value can be estimated, and in areas in which depth edges do not occur, the depth value can be estimated based on consistency with the needle diagram.

In the annealing method, the pixels included in \(S_c\) are selected at random, and then the change in the energy \(E_{all}\) is evaluated for when the depth value for the selected pixels is varied. When the energy decreases or increases to a value below the value (temperature) determined by the annealing method, the depth value for the selected pixels is updated. \(E_{all}\) can be minimized by repeating this process while slowly reducing the temperature.

### 3.4. Joining the distance maps

A three-dimensional shape is obtained by joining the distance maps obtained from each camera. The three-dimensional position \(X_i\) for the surface of the object corresponding to the representative point \((x_i^j, y_i^j)\) for the pixels is represented by the following equation using the focal distance \(f_i\) for the depth \(Z(x_i^j, y_i^j)\) and the camera \(c\), and the lens center \((e_i^c, e_j^c)\) for the camera:

\[
X_i^c = \begin{pmatrix} 
\frac{Z(x_i^c, y_i^c)}{f_i} (x_i - e_x^c) \\
\frac{Z(x_i^c, y_i^c)}{f_i} (y_i - e_y^c) \\
Z(x_i^c, y_i^c) 
\end{pmatrix}
\]

The three-dimensional position for the surface of the object calculated from the distance map using Eq. (10) is on the camera coordinate system where the focal point of the camera is the point of origin. As a result, the three-dimensional position \(\mathbf{X}_i^c\) for the object shape in the world coordinate system can be obtained by setting \(\mathbf{X}_i^c = R_c \mathbf{X}_i + T_c\), using the rotational matrix \(R_c\) and the parallel transition vector \(T_c\) which represent the conversion from the camera coordinate system for the camera \(c\) to the world coordinate system.

### 4. Experiments

In order to evaluate the effectiveness of the proposed method, the authors performed an experiment using simulation data in which the shape was known as well as an experiment using a real object.

The images captured using nine virtual cameras (640 × 480 pixels) set up to surround an object illuminated from 20 directions for simulation data of the dinosaur shown in Figs. 6(a), 6(b), and 6(c) were used as input.

Figures 6(e) through 6(j) represent the object shape restored using the distance maps obtained by minimizing for the camera [Fig. 6(d)] only the energy \(E_{VL}\) for consistency with the silhouettes [Figs. 6(e) and 6(h)], only the energy \(E_N\) for consistency with the needle diagram [Figs. 6(f) and 6(i)], and the sum \(E_{all}\) for the energy for consistency with the silhouette and the energy for consistency with the needle diagram [Figs. 6(g) and 6(j)]. Moreover, Figs. 6(e) through 6(g) are images generated using the same angle as in Fig. 6(a), and Figs. 6(h) through 6(j) are images generated using the same angle as Fig. 6(c).

Based on Figs. 6(e) and 6(h) as well as Figs. 6(g) and 6(j), the authors were able to confirm that consistency with the needle diagram was able to restore smoother surfaces compared to when only consistency with the silhouette was taken into consideration. Moreover, depth edges occurred at the boundary between the left and right legs in the image [Fig. 6(d)]. Consequently, when consistency with the silhouette is not taken into consideration, a shape is created in which the site where the left and right legs are adjacent becomes smooth, and ends up being different from the original shape. Figures 6(f) and 6(i) represent the results of restoration when silhouette consistency was not used. The
left and right legs, which do not touch in the original image, do end up touching in Figs. 6(f) and 6(i), which can be confirmed by comparing them to Figs. 6(a) and 6(c). In contrast, when silhouette consistency is taken into consideration, the majority of the depth edges are detected by using the consistency with the silhouette from the other cameras. The left and right legs, which are touching in Figs. 6(f) and 6(i), are separated in Figs. 6(g) and 6(j), which were restored using silhouette consistency. This confirms the effectiveness of the authors’ method for an object with depth edges.

For each pixel $m_{ij}$ in the acquired distance map, the estimated error for the depth value was calculated. Figure 6(k) shows a logarithmic histogram of the results, and Table 1 shows the results of dividing the total number of pixels for which the mean value for the estimated error and the error were below 0.01, 0.02, and 0.05 by the total number of pixels in the entire silhouette. Note that the magnitude of the estimated error for the depth values is represented by a length for which the total length of the object is set to 1. For instance, an error of 0.01 indicates that a depth error equivalent to 1% of the total length of the object occurred.

The authors confirmed that when using only consistency with the silhouette, only under half of the pixels were kept within a range of an estimated error of 0.01, while using needle diagram consistency led to 85% of the pixels being kept within that range. This confirms an improvement in precision due to the use of the needle diagram. Moreover,
a comparison of the results of using only needle diagram consistency and the results of using the proposed method clearly showed, according to Table 1, that the proposed method could produce a distance map with better precision overall. Moreover, Fig. 6(k) shows that the number of pixels with an error less than 0.05 is far lower compared to the results of using only consistency with the needle diagram. This is thought to be a result of the proposed method improving the shape created with a smooth connection at the places where the left and right legs are adjacent.

Next the authors performed an experiment using a real object. The object to be measured was approximately 4 cm wide, and was set up in the center of a square frame 90 cm on a side. Then, four color CCD cameras (PointGray Corp. Dragonflys) with a resolution of 640 by 480 were set up in front of the object, and another four were set up behind the object for a total of eight cameras. A total of 24 incandescent lamps (40 W) were set up, with 12 in front of the object and 12 behind it. The incandescent lamps are point light sources, but there is an approximately three degree blur in the direction of illumination, so they can be taken to be almost parallel light sources.

The 24 illumination sources were turned on one after the other, and the image series necessary to create an

Table 1. Depth error

<table>
<thead>
<tr>
<th></th>
<th>Only silhouette consistency</th>
<th>Only needle map consistency</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average error</td>
<td>0.019</td>
<td>0.011</td>
<td>0.0063</td>
</tr>
<tr>
<td>Error less than 0.01</td>
<td>44.6%</td>
<td>78.5%</td>
<td>85.0%</td>
</tr>
<tr>
<td>Error less than 0.02</td>
<td>68.1%</td>
<td>85.2%</td>
<td>91.6%</td>
</tr>
<tr>
<td>Error less than 0.05</td>
<td>91.1%</td>
<td>93.0%</td>
<td>97.6%</td>
</tr>
</tbody>
</table>

Fig. 7. Partial shape from a depth map (real data). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
illuminated differential stereo were obtained by observing the result with the eight cameras. In addition, the silhouettes necessary for the visual hull method were extracted using the background differential method.

The experiment was performed with the mascot as a target for measurement. Figure 7 shows the results of object shape restoration based on the distance map obtained by minimizing only $E_{VL}$ [Figs. 7(d) and 7(g)], only $E_{N}$ [Figs. 7(e) and 7(h)], and $E_{all}$ [Figs. 7(f) and 7(i)]. Figures 7(b) and 7(c) show the mascot, and Figs. 7(d) through 7(f) as well as 7(g) through 7(i) are images captured from virtually the same angle. Note that the texture observed by the camera in Fig. 7(a) was applied to Figs. 7(d) through 7(i) for the restored distance maps.

Based on Figs. 7(d) and 7(g) as well as Figs. 7(f) and 7(i), the authors were able to confirm that in comparison to taking only the silhouette consistency into consideration, using the needle diagram consistency resulted in smoother surface restoration, as was the case in the experimental results using simulation data. Moreover, depth edges occurred in the places where the body of the mascot and the pole the mascot was holding were adjacent in the image. By comparing panels (e) and (h), panels (f) and (i), and panels (b) and (c), the authors were able to confirm that the effect due to these depth edges could be handled by using the silhouette consistency.

Figure 8(c) shows the results of joining the distance maps obtained from each camera. The authors were able to confirm that surfaces that were smooth on the object and surfaces with depressions, difficult to acquire using the visual hull method [Fig. 8(a)] without using silhouette consistency, could be measured using the proposed method.

Moreover, a comparison of the object shape [Fig. 8(b)] acquired without taking silhouette consistency into consideration and the results [Fig. 8(c)] of acquiring the shape using the proposed method confirmed that the shape acquired near places where depth edges occurred was different, and demonstrated that the proposed method is effective for objects with depth edges.

5. Conclusion

In this research, the authors focused on shape information for an object observed using a camera, in addition to the silhouette of the object, and proposed a method to measure the smooth curving surfaces of an object and the depressed surfaces of an object which cannot be measured using the visual hull method using shape information. When generating a distance map using a needle diagram obtained using the photometric stereo, if the depth edges are not taken into consideration, then errors in the distance map occur near places where the depth edges are. The authors used the fact that often a distance map restored without taking the depth edges into consideration lacks consistency with the silhouette obtained from other directions, defined the energy representing the consistency between the distance map and the silhouette, and the consistency between the distance map and the needle diagram, and then reduced the error in the distance map by obtaining the shape which minimizes this energy.

Future topics include the problem of overlapping distance maps obtained from different cameras. Estimation errors in the normal line, estimation errors in the camera position, and extraction errors for the silhouette can all be considered as causes for shifts in the overlapping distance maps. Future research to further reduce such errors is needed. Moreover, research on superimposing distance maps measured from different directions has already been performed, and experiments to determine to what extent these methods can be applied to the authors’ method are needed.

REFERENCES


**APPENDIX**

**Supplement for Eqs. (3), (4), (5), and (7)**

We assume that the camera field of view is narrow, and that the camera and object are sufficiently far apart. The projection model for the camera is approximated using a weak central projection. On a flat plane separated by $Z(x_i', y_i')$ from the focal point of the camera, the length equivalent to one pixel’s width on the image is equivalent to $Z(x_i', y_i')/f_c$. When the point $X_i'$ separated by $Z(x_i', y_i')$ from the focal point of the camera and projected onto $(x_i', y_i')$ and the point $X_i$ projected onto $(x_i' + \alpha, y_i' + \beta)$ are present on the same plane represented by the normal line vector $n_i = (p_i', q_i', -1)^T$, then $(X_i' - X_i)n = 0$ holds.

If the fact that the length equivalent to one pixel’s width on the image is equivalent to $Z(x_i', y_i')/f_c$ is used, then the $X$ coordinates and $Y$ coordinates for the point $X_i'$ are the coordinates separated from $X_i'$ by $(Z(x_i', y_i')/f_c)\alpha$ and $(Z(x_i', y_i')/f_c)\beta$. This result and the depth of the point projected onto $(x_i' + \alpha, y_i' + \beta)$ can be thought of as $Z(x_i', y_i')(1 + (p_i'/f_c)\alpha + (q_i'/f_c)\beta)$ based on $(X_i' - X_i)n = 0$.

---

*When the field of view is wide, this approximation does not hold near the edges of the image. However, this approximation does hold in a situation in which imaging is performed with a narrow field of view and at a distance from the object, as is assumed in the authors’ method.*
Masaaki Iiyama (member) withdrew with distinction from his doctoral studies in the Informatics Research Department at Kyoto University in 2003 and became a lecturer at the Academic Center for Computing and Media Studies there. He is pursuing research on computer vision and three-dimensional data processing. He is a member of IEEE and the Information Processing Society of Japan.

Koh Kakusho (member) completed his doctorate at Osaka University in 1993. He was a visiting researcher at Stanford University from 1993 to 1994. He became a lecturer at Osaka University in 1994. He was appointed an assistant processor at the Center for Computing and Media Studies at Kyoto University in 1997, and moved to the Academic Center for Computing and Media Studies there in 1992. He is pursuing research on visual media processing and communications. He holds a D.Eng. degree.

Michihiko Minoh (member) graduated from the Department of Informatics at Kyoto University in 1978 and became a lecturer on the Faculty of Engineering. From 1987 to 1988 he was a visiting researcher at Massachusetts Institute of Technology. In 1989 he was appointed an assistant processor with the Precision Information Development and Experimental Group in the Faculty of Engineering at Kyoto University, a professor with the Center for Computing and Media Studies there in 1997, and a professor with the Academic Center for Computing and Media Studies in 2002. He is pursuing research related to image processing, artificial intelligence, and intelligent communications. He holds a D.Eng. degree, and is a member of IEEE, ACM, the Information Processing Society of Japan, the Institute of Image Electronics Engineers of Japan, and the Robotics Society of Japan.