3D Shape Reconstruction from Multiple Silhouettes for Objects in Rigid Motion

Masahiro Toyoura

March 2008

Abstract

We discuss 3D shape reconstruction from multiple silhouettes for objects in rigid motion. The 3D shape reconstruction from multiple silhouettes is called *Shape-from-Silhouette*. It has been addressed by many research groups. The Shape-from-Silhouette provides the basis of recent 3D shape reconstruction methods. Recent proposed methods often reconstruct accurate shapes with colors, textures or continuity of the objects, which is based on the reconstructed shapes from silhouettes. Our proposed method in multiple frames enables us to reconstruct accurate shapes only from silhouettes. Moreover, it also provides massive images for the recent reconstruction methods.

A main contribution of our research is the elimination of assumptions on colors and textures of the objects. We have realized the 3D shape reconstruction without colors and textures of the objects. One of the advantages of the Shape-from-Silhouette is the robustness for various environment and objects. Our contribution is important to keep the advantage of the Shape-from-Silhouette.

The visual hull is reconstructed as the intersection of the regions calculated by silhouettes. It is guaranteed that the object is included in the regions. With small number of cameras, the visual hull includes additional regions, which do not represent the region of the object. Decreasing the additional regions means that the reconstructed shapes become more accurate. The additional regions decrease with increasing the number of cameras. However, it is not realistic to install so many cameras around the object. The number of cameras is limited by the conditions on physical size of cameras, space for setting cameras, prices of cameras and so on. To exceed the limitation, a shape reconstruction method that integrates silhouettes of multiple frames has been proposed.

Let us suppose that the object is in a rigid motion. When the object moves rigidly, cameras change their relative positions to the object at every

moment. If the rigid motion of the object can be correctly estimated, images obtained by the cameras at different moments are treated as the images in different positions virtually. With these virtual images, an accurate visual hull is reconstructed without increasing the number of physical cameras. Increasing the number of cameras also contributes the recent shape reconstruction methods with colors, textures or continuity of the object. The methods based on the volume intersection method are expected to reconstruct more accurate shapes by increasing number of cameras.

The shape reconstruction from silhouettes in multiple frames is composed of two parts of techniques. One is silhouette extraction, and the other is silhouette integration. We propose a silhouette integration method which preserves the robustness of the volume intersection method. In previous works, there are problems both in silhouette extraction and silhouette integration.

The silhouette extraction for the shape reconstruction has not been discussed enough. To reconstruct shapes of various objects, the silhouettes should be extracted for the various objects accurately enough. In our research, we propose a silhouette extraction method from images obtained from multiple cameras. Even for objects in unknown color, the method can be adopted. The method is realized with our proposed *random pattern background*. The random pattern has many small regions with randomly-selected colors. By using the random pattern backgrounds, we can keep the rate of missing parts below a specified percentage. Moreover, for refining the silhouettes, we detect and fill in the missing parts by integrating multiple images. From the images captured by multiple cameras used to observe the object, the object colors can be estimated. The missing parts can be detected by comparing the object color with its corresponding background color. In our experiments, we confirmed that this method effectively extracts silhouettes and reconstructs 3D shapes.

Secondly, we have discussed object motion estimation for the silhouette integration in multiple frames. In previous works, colors and textures of the objects have been used for the motion estimation. They have not discussed without colors and textures of the objects. We have proposed *outcrop point*, which is 3D feature points extracted without colors and textures on surfaces of the object. The problem for 3D feature point extraction is the fact that reconstructed visual hulls might include the additional regions. The outcrop points are guaranteed to be included not in the additional regions but in the object region. This characteristic supports robust motion estimation, since the outcrop points are continued to be extracted between different frames.

Moreover, we proposed an intelligent method of integrating incomplete silhouettes and motion where outcrop points play an important role. In the volume intersection method, shapes are reconstructed as an intersection region of all possible regions calculated with silhouettes obtained from all cameras. When object motion is estimated with large error, the integrated shape will have many missing parts. Especially when silhouettes are extracted with many missing parts or additional parts, 3D feature points are difficult to be extracted. When the estimated motion has large error, shapes are reconstructed with missing parts. We cannot prevent silhouettes from being extracted with additional and missing regions in real environments. To solve the missing problem, we have designed a function calculated with outcrop points, visual hulls and estimated motion. The reconstructed shape are preserved can be evaluated referring to how many outcrop points have been included in the reconstructed shape of another frame. The outcrop points tend to be extracted from outstanding parts on the object surface, and the outstanding parts characterize what the object is. Preserving these points in the integrated shapes gives us to reconstruct accurate shapes. Silhouettes in multiple frames can be integrated with fewer missing parts based on this evaluation.

In this paper, the silhouette extraction and the silhouette integration are discussed. The accuracy of the silhouette extraction improves that of the silhouette integration. The accuracy of the silhouette integration also improves that of the silhouette extraction. Accurate 3D shapes can be reconstructed by considering both of the silhouette extraction and silhouette integration.

Acknowledgements

I wish to thank my advisor Professor Michihiko Minoh for his guidance, encouragement, support, inspiration and enthusiasm during my time at Kyoto University. He contributed greatly to writing all papers that relate to this thesis. He has managed our laboratory with great fervor. Without his help, I could never carry out this thesis and never spend such wonderful days.

Associate Professor Koh Kakusho also provided significant support for all researches that relate to this work. He attentively guided how to write technical papers. His enthusiastic guidance conducts me to the accomplishing this work. Dr. Masaaki Iiyama is one of members who made the most significant contributions to this work. He has joined all discussions related to this work. I was guided for writing my master's and bachelor theses with several inspiring suggestions. His 4π observation system is a core element for constructing this work. I would like to thank Dr. Takuya Funatomi. I have been influenced by his studious and dedicated research style. He provided me several useful software tools.

I am deeply grateful to Professor Takashi Matsuyama for his inspiring advices. He has given me several impressive advices since I was in master's course. Professor Yuichi Nakamura also gave me suggestive advices and comments. His advices contributed to improve the quality of this thesis.

I would like to thank Dr. Satoshi Nishiguchi, Dr. Masayuki Murakami, Mr. Takafumi Marutani, Mr. Yoshitaka Morimura and Ms. Makiko Uesugi for discussions and suggestions in daily life. They have supported me in both public and private life. The other members of Model group and the other groups have also joined discussions in daily life. This work was supported by a JSPS Research Fellowship for Young Scientists.

Last but not least, I wish to thank my mother, Chieko for her mental and financial support. Without her dedicated support, none of this could have been completed.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We discuss 3D shape reconstruction from silhouettes in multiple frames. The 3D shape reconstruction from silhouettes is called *Shape-from-Silhouette*. It has been addressed by many research groups. The Shape-from-Silhouette provides the basis of many kinds of recent 3D shape reconstruction methods. Recent proposed methods reconstruct accurate shapes with colors of the object with consistency between reconstructed shape and obtained images or continuity of the shapes, which is based on the reconstructed shapes from silhouettes. Our proposed image integration method in multiple frames enables us to reconstruct accurate shapes. Moreover, it also provides useful information for the recent proposed methods.

In the Shape-from-Silhouettes in multiple frames, our main contribution of this research is the elimination of assumptions on colors and textures of target objects. We have realized 3D shape reconstruction without colors and textures of the objects. One of the advantages of the Shape-from-Silhouette is the robustness for various environment and objects. Our contribution is important to keep the advantage of the Shape-from-Silhouette.

1.1 Shape from Silhouettes of Multiple Frames

With the volume intersection method [30], a shape of a 3D physical object is reconstructed from silhouettes of obtained images. The volume intersection method is a representative method of Shape-from-Silhouette. In the volume intersection method, the reconstructed shapes are called *visual hulls*, or *VHs*. The volume intersection method is intrinsically a method of shape reconstruction only from obtained silhouettes. he method is one of the most robust methods for 3D shape reconstruction. Silhouette extraction is not easily affected by changing optical environment. The shape reconstruction from silhouettes is robust for color changing of object surfaces. Non-Lambertian reflectance can be dealt by the shape reconstruction from silhouettes. Unlike stereo vision and other 3D shape reconstruction methods, the volume intersection method does not require the existence of texture on the surface of the object, since the method can reconstruct 3D shapes only from silhouettes; it is available even in adverse lighting environments. In recent works [41, 47, 10, 14, 25, 43, 19, 48, 37] , visual hulls are refined by any information other than the silhouettes. Together with the volume intersection method, colors of the object, consistency between reconstructed shape and obtained images or continuity of the shapes are used to reconstruct more accurate shapes. The works have been built upon the robustness of the volume intersection methods.

The visual hull is reconstructed as the intersection of the regions calculated by silhouettes. The regions have possibilities of existence of a part of the object. With small number of cameras, the visual hull includes additional regions, which do not represent parts of the object. The additional regions decrease with increasing the number of cameras. Decreasing the additional regions means that the reconstructed shapes become more accurate. However, it is not realistic to install so many cameras around the object. The number of cameras is limited by the conditions on physical size of cameras, space for setting cameras, prices of cameras and so on. To exceed the limitation, a shape reconstruction method that integrates silhouettes obtained in multiple frames have been proposed [4].

It has been proposed in many previous works on the volume intersection method. They employed silhouettes of an object in motion in order to improve accuracy in the reconstructed shape. In some of those works, the object is observed by cameras while being rotated by turntables [49, 56] . Obtained images of the object in rotation can be applied to the volume intersection method. Changing relative positions between the cameras and the object is described by the rotation parameter of a turntable. Although this approach using the turntable is easy and effective for introducing motion of objects to the volume intersection method, it is possible that motion of the object is limited only to the rotation around the axis of the turntable.

Let us suppose that the object is in a rigid motion. When the object moves rigidly, cameras change their relative positions to the object at every moment. If the rigid motion of the object can be correctly estimated, images obtained by the cameras at different moments are treated as the images in different positions virtually. With these virtual images, an accurate VH is reconstructed without increasing the number of cameras. With these virtual cameras, we can improve accuracy of the reconstructed 3D shape without increasing the number of cameras. The increasing the number of cameras also contributes the shape reconstruction with colors, consistency or continuity in the images. The methods based on the volume intersection method are expected to reconstruct more accurate shapes by the increasing the number of cameras in multiple frames.

The shape reconstruction by silhouette integration in multiple frames is composed of two parts of techniques. One is silhouette extraction, and the other is silhouette integration. We have proposed the silhouette integration in multiple frames, which preserves the robustness of the volume intersection method. In previous works, there are problems not only in the silhouette extraction but also in the silhouette integration. Our main contribution the elimination of assumptions on colors and textures of target objects from the methods of previous works.

We address both of the problems of the silhouette extraction and the silhouette integration as shown in Figure 1.1. First, we discuss the silhouette extraction in Chapter 2. In Chapter 2, random pattern, which is special texture, is proposed for the silhouette extraction. The random pattern enables us to extract silhouettes even for unknown color objects. In Chapter 3, we propose feature points which can be extracted by textureless objects. We call the feature points *outcrop points*. With the outcrop points, we realize to estimate a motion of a textureless object. In Chapter 4, we propose a novel silhouette integration method which does not adopt silhouettes of harmful frames. In some frames, silhouettes are registrated with large errors of the motion estimation. The silhouettes of such frames lead to missing parts in the reconstructed shapes. Especially outstanding points on the object surface tend to be missed easily. Since the outstanding points characterize the shapes of the objects, missing outstanding points makes the reconstructed shapes inferior. To avoid integrating the silhouettes of such frames, we realize to reconstruct shapes with preserving the outstanding points and shape characteristics. In Chapter 5, we conclude this paper. We also discuss the relationship between the silhouette extraction and the silhouette integration in the Chapter. We describe how the accuracy of the silhouette extraction affect that of the silhouette integration, and conversely. Future works are discussed in Chapter 6.

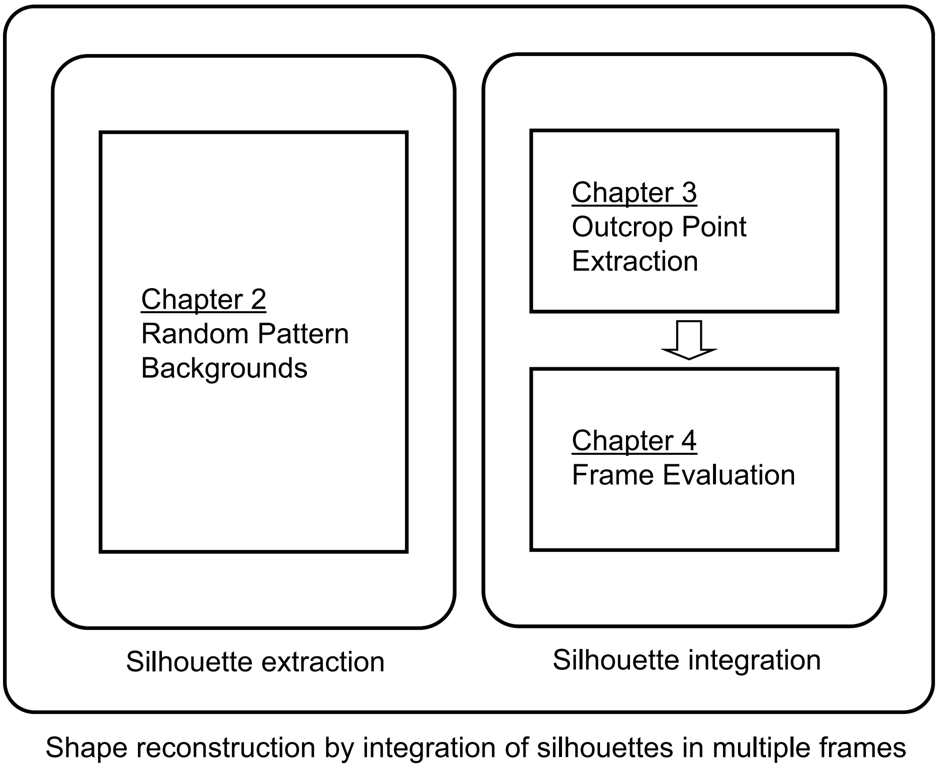


Figure 1.1: The integration of silhouettes in multiple frames.

1.2 Silhouette Extraction with Random Pattern Backgrounds

In Chapter 2, we present a novel approach for extracting silhouettes by using a particular pattern that we call the random pattern.

The volume intersection method reconstructs the shapes of 3D objects from their silhouettes obtained with multiple cameras. With the method, if some parts of the silhouettes are missed, the corresponding parts of the reconstructed shapes are also missed. When colors of the objects and the backgrounds are similar, many parts of the silhouettes are missed. We adopt random pattern backgrounds to extract correct silhouettes. The random pattern has many small regions with randomly-selected colors. By using the random pattern backgrounds, we can keep the rate of missing parts below a specified percentage, even for objects of unknown color. To refine the silhouettes, we detect and fill in the missing parts by integrating multiple images. From the images captured by multiple cameras used to observe the object, the object's colors can be estimated. The missing parts can be detected by comparing the object's color with its corresponding background's color. In our experiments, we confirmed that this method effectively extracts silhouettes and reconstructs 3D shapes.

1.3 Object Motion Estimation from Silhouettes of Multiple Frames

We discuss 3D shape reconstruction of an object in a rigid motion with the volume intersection method in Chapter 3. A reconstructed shape becomes more accurate with an increase number of cameras in the volume intersection method. However, it is not realistic to install so many cameras around the object due to physical limitation on their spatial configurations. When the object is in a rigid motion, the cameras change their positions to the object at every moment. If the rigid motion of the object can be correctly estimated, cameras at different moments are treated as cameras in different positions virtually. With these virtual cameras, we can improve accuracy of the reconstructed 3D shape without increasing their number. Based on this idea, we propose an accurate shape reconstruction method from images of the object in motion. Our method reconstructs the 3D shape while estimating its motion with our new proposed feature points. The feature points are guaranteed to be located on the real surface of the object. As the result, we can acquire an accurate shape from images in multiple frames.

It has been proposed in many previous works on the volume intersection method. They employed silhouettes of an object in motion in order to improve accuracy in the reconstructed shape. Cheung et al. [4] have proposed to extract feature points for estimating the rigid motion of the object from images. However, this method employs color information to extract the feature points. It loses the advantage of the volume intersection method that does not need color information of objects and can be applied even to objects without texture, compared with stereo vision approaches, as discussed above.

In other works, it is proposed to extract feature points of the object for motion tracking. The feature points are located on the surface of visual hulls. They are extracted based on the epipolar geometry [6, 9, 13, 10, 53] . The feature points are called *frontier points* or *epipolar tangencies.* The motion estimation with the feature point from visual hulls does not require color features of the object. A disadvantage is that the feature points fails to be extracted, which caused by the additional regions included in the visual hulls. Since the visual hulls include the additional regions, extracting feature points only from the object region is difficult. From the additional regions, some feature points are extracted at a certain frame, although other some feature points are extracted at a different time. The feature points extracted in different frames are not guaranteed to be same.

To address the problem, we propose a new kind of feature points, which called *outcrop points.* The outcrop points are also extracted from the visual hulls. They are guaranteed to be included in the object region and not included in the additional regions. Stable object motion estimation is realized with the outcrop points.

1.4 Frame Evaluation for Silhouette Integration

In volume intersection method, 3D shapes are reconstructed from silhouettes which obtained by multiple cameras. When more cameras are used, more accurate shapes are reconstructed in the volume intersection method. Since the number of cameras is limited, the accuracy of reconstructed shapes is also limited. In recent works, silhouette integration methods [30, 23] have been proposed for shape reconstruction.

A problem of these methods is that shapes with missing parts are produced from incomplete silhouettes. The incompleteness of silhouettes means missing and over-extracted of silhouettes. It is known that the missing of reconstructed shape causes the error of motion estimation. In previous works, any solution has not proposed to solve the problem. To solve the problem is important, because the incompleteness of silhouettes cannot be avoided in real environment.

In Chapter 4, we discuss how the motion estimation is affected by the incompleteness of extracted silhouettes. Based on the discussion, we propose an integration method for incomplete silhouettes. In the method, outcrop points, which are kinds of feature points for motion estimation, play an important role. By referring outcrop points and reconstructed shapes, estimated motion can be evaluated. Based on the evaluation, silhouettes in multiple frames can be integrated with less missing parts.

Chapter 2

Silhouette Extraction with Random Pattern Backgrounds

2.1 Introduction

In the volume intersection method [30], shapes of 3D objects are reconstructed from the silhouettes of multiple images that are corresponding regions of camera images of those objects. The shapes of the physical objects are often reconstructed with stereo vision approaches [8]. However, these approaches cannot be applied to the objects without rich texture. Laser rangefinders are also used for reconstructing shapes [27, 18]. However, laser rangefinders cannot reconstruct the shapes of the objects that absorb laser light. Unlike the approaches mentioned above, the volume intersection method is not affected by colors or surface characteristics of objects because the method reconstructs the shapes of the objects from their silhouettes only.

However, the volume intersection method has the problem that some parts of reconstructed shapes are missed when the corresponding parts of the silhouettes are missed. The missing parts of silhouettes take the form of *holes* in the reconstructed shapes. For example, with the chroma key systems, which often called blue screen matting systems, many parts of the silhouettes extracted for blue objects are missed. To avoid missing parts of the silhouettes, we need to employ backgrounds in a color different from the colors of object.

In previous works, extracting the silhouettes without depending on objects' colors by switching two backgrounds in different colors has been proposed [44, 31]. In these methods, two different single-color backgrounds are used. At least one of the background colors should be different from each color of the object. The union of the silhouettes obtained when using each of the two backgrounds is guaranteed to have no missing parts. However, in these methods, the object must remain fixed stable while the backgrounds are switched. Another advantage of the volume intersection method is that the method requires little observation time. This advantage is effective for reconstructing moving objects, or to analyze shape transformation [17, 3, 32]. Reduced observation time makes it possible to reconstruct changing shapes in multiple frames. However, due to the requirement that the object remain fixed while the background is switched, the advantage of a short observation time is lost.

In film production, the silhouettes are extracted with special devices. Such silhouette extraction is often called as *matting*. In Z-keying method [20], a camera set as a range sensor is required for each position. In defocus matting [33], a large number of special camera sets are required. These methods are difficult to be applied for the volume intersection method, because many cameras are used for reconstructing 3D shapes. There occurs the problem of physical space and position calibration.

In this article, we propose a real-time silhouette extraction method even for objects of unknown color. We employ *random pattern backgrounds* for silhouette extraction. The random pattern backgrounds have many small regions filled with randomly selected colors. With the random pattern backgrounds, the object's color and the background's color are expected to be different in most regions. As the result, the missing parts of silhouettes are suppressed below a specified percentage.

When the silhouettes have only few missing parts, almost complete shape of the object can be reconstructed. From the shape, we can estimate to which pixels of the obtained images that each part of the object is projected. The color of this part of the object can be estimated from the projected pixels. When the colors of the part and the projected pixels are similar, determining the pixels to be included in the silhouettes is difficult. Correct silhouettes can be obtained by including a large percentage of the pixels in the silhouettes. Using the random pattern backgrounds together with the silhouette correcting method, shapes with fewer missing parts can be obtained even for objects of unknown color.

In the reminder of this article, we first give a brief summary of the volume intersection method, as well as the reason why parts of the reconstructed shapes are missing due to missing parts of the silhouettes in Section 2. In Sections 3 and 4, we propose a method for obtaining accurate shapes of objects of unknown color by using the random pattern backgrounds and a silhouette refining method. Experimental results are presented in Section 5, and future work is discussed in Section 6.

2.2 3D Shape reconstruction from Silhouettes in Multiple Frames

In this Section, we explain the theories of the volume intersection method.

2.2.1 The Volume Intersection Method

The shape of an object is reconstructed from silhouettes obtained by multiple cameras with the volume intersection method [30, 23]. The reconstructed shape is called a *Visual Hull*, or VH.

The volume intersection method at time i is realized as shown in Figure 2.1. Let us denote the cameras that are placed around target object O to capture it by $C_j(j = 1, \cdots, N)$, where N denotes the number of cameras ($N > 1$). All the cameras observe the object synchronously. Time i can be replaced as the i-frame. The 2D region that corresponds to object O is extracted from the images of C_j. The projection matrix of C_j is represented as P_j. This region is called the *silhouette* and denoted by S_{ij}. When the object is in motion, observed silhouettes differ among frames. O is guaranteed to be included in a cone with the apex at the optical center of C_j and the base at S_{ij} is then calculated. This cone is called the *visual cone* of camera C_j, and denoted by

$$V_{ij} = \{v \mid P_j v \in S_{ij}\}, \tag{2.1}$$

where v represents the occupation of a small 3D region, or a *voxel*. *Visual hull* V_i of the i-frame is defined as the intersection of visual cones $V_{i1}, \cdots, V_{iN}$.

$$V_i = \{v \mid \forall j,\ P_j v \in S_{ij}\}. \tag{2.2}$$

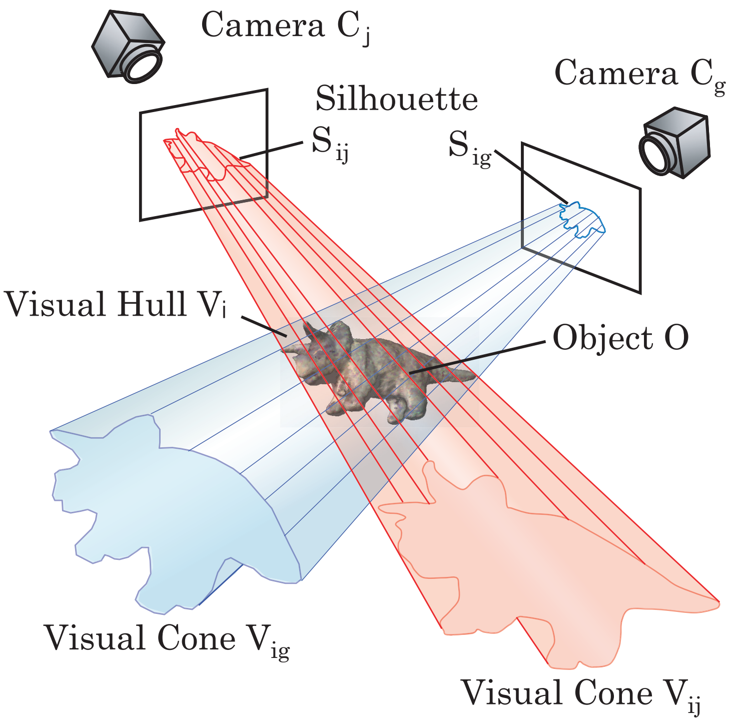


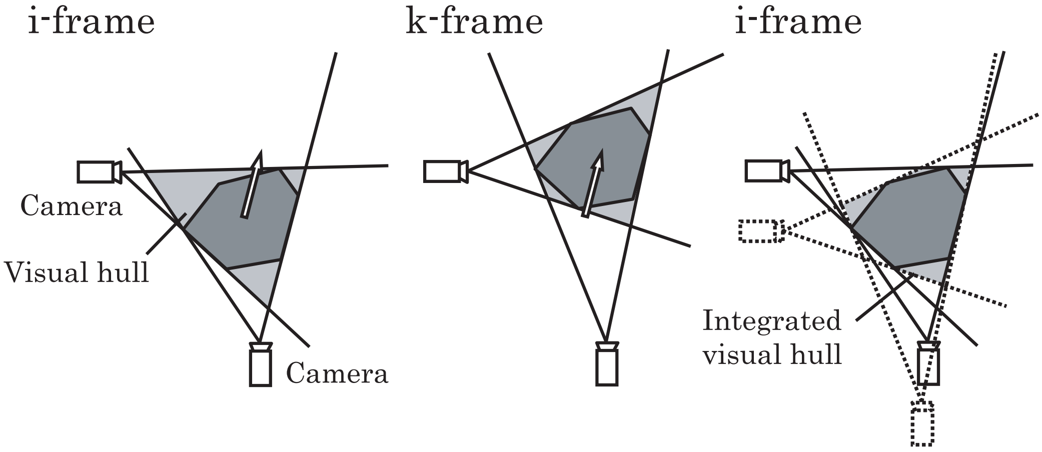
Figure 2.1: Volume intersection method.

In this paper, it is assumed that the visual hull is represented as a set of voxels. Each voxel denotes occupancy of a unit region in the 3D space. Whether the unit region is occupied or not is represented with binary values. The visual hull V_i describes the 3D shape reconstructed with the volume intersection method with cameras C_j $(j = 1, \cdots, N)$, but the visual hull is not exactly an identical shape with that of the target object. Regions corresponding to the target object are called *object regions*. Those regions of the visual hull that are not included in the object regions are called *additional regions*. The additional regions decrease with the increase number of cameras; more accurate shapes can be obtained using more cameras.

2.2.2 Integration of Visual Hulls in Multiple Frames

Although the increase number of cameras improves accuracy of the reconstructed shape with the volume intersection method, it is not realistic to install so many cameras due to physical limitation on placing cameras around the object. Instead of increasing cameras actually, we can virtually realize the situation in which the number of cameras is increased by using images of the object in motion. When the motion of the object is known, cameras at different moments are treated as those at different positions.

As illustrated in Figure 2.2, suppose that a pair of cameras observe the object. The visual hulls before and after the movement of the object is shown in (a) and (b) of Figure 2.2. The images obtained by the pair of cameras after the movement of the object serve those with another pair of cameras at different positions before the movement. Those positions of the cameras can be calculated if the motion of the object is known. By using images obtained from the four different positions of the pair of cameras before and after movement, we can improve accuracy of the reconstructed shape.



(a) VH before movement (b) VH after movement (c) Integrated VH

Figure 2.2: Integration of images at different moment.

The silhouette integration in case of more than 2 cameras is discussed as same as the case of 2 cameras. The silhouette integration of multiple cameras is shown in Fig 2.3.

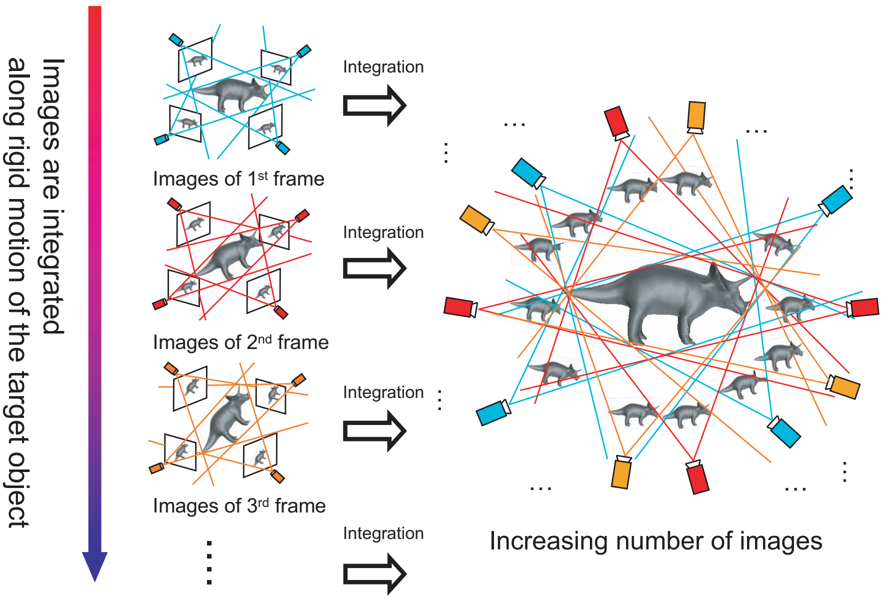


Figure 2.3: Integration of images of multiple cameras in multiple frames.

Let the object moves rigidly and the motion be known. An intersection of the visual hulls $V_i (i = 1, \cdots, M)$ means the shape from silhouettes of all frames. When k-th frame is the main frame, the intersection V^k is calculated by the estimated motion D_{ik} between k-frame and i-th frame $(i = 1, \cdots, M)$ by Eq. (2.3). The intersection V^k is called an integrated shape.

$$V^k = \{v \mid \forall i, D_{ik} v \in V_i\}. \tag{2.3}$$

2.2.3 Evaluation for Reconstructed Shapes

To evaluate the accuracy of reconstructed shapes, the numbers of the missing regions and additional regions are available. The reconstructed visual hull might not match to the object region completely as shown in 2.4. The intersection of these regions is *common region*. *Additional region* is defined as the region which is included in the visual hull and not included in the common region. *Missing region* is also defined as the region which is not included in the visual hull and included in the common region. Since the visual hull

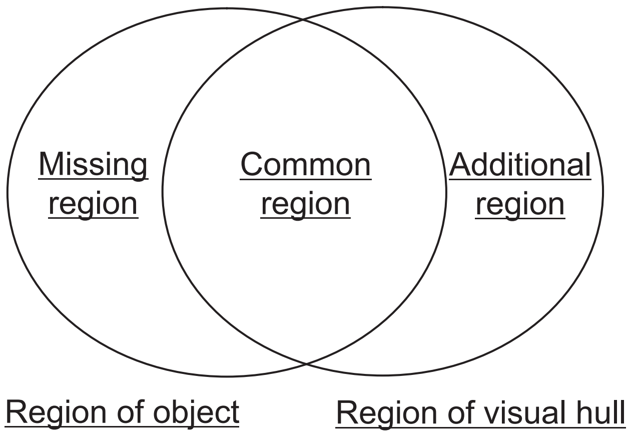


Figure 2.4: A missing region and additional region of a visual hull.

ideally matches to the object region, the additional region and missing region should be decreased. *Error region* is defined as the sum of the additional and missing regions. From the number of voxels included in the error region, we will evaluate the reconstructed shapes.

2.3 Random Pattern Backgrounds

2.3.1 Silhouette Extraction for Objects of Unknown Color

Generally, the silhouettes of an object are extracted based on the difference in color between an object and its background. When the colors of the object and its background are similar, silhouettes extracted would have many missing parts. The missing parts of the silhouettes take the form of *holes*. To decrease the amount of missing parts of the silhouettes, the background colors should be different from those of the object. However, this condition cannot be fulfilled by using a background in a single color when the color of the object is unknown. Although, employing backgrounds of different colors by switching those backgrounds depending on the colors of objects is possible [31, 44], to prepare and change those backgrounds takes time and effort.

We adopt *random pattern backgrounds* to extract correct silhouettes. The random pattern have many small regions with randomly-selected colors, as shown in Figure 2.5. The silhouettes extracted with the random pattern backgrounds have missing parts below a specific percentage, even for objects of unknown color.

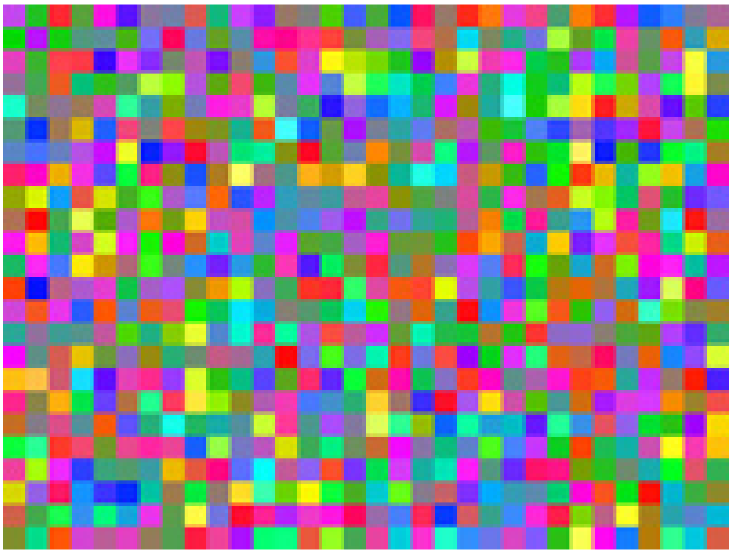


Figure 2.5: Sample of random pattern.

Figure 2.6 illustrates the rate of missing parts of a silhouette for an object, when the object has various colors and the silhouette is extracted under the background in a specific color. In the figure, the horizontal axis denotes the color of the object and the vertical axis the rate of missing silhouette parts for that color. When the object has a color different from its background, the silhouette for the object is correctly extracted as is shown by the solid line (a) in Figure 2.6, i.e., the rate of missing parts is 0.

However, when the object has a color similar to the background, i.e., the difference between the colors is less than a threshold denoted by Δ_1 in Figure 2.6, the rate of missing parts would be drastically high. With random pattern backgrounds, the silhouettes have missing parts below a specific rate Δ_2 shown as the broken line (b) in Figure 2.6. The missing rate Δ_2 does not depend on the object's color.

For objects in multiple colors with single-colored backgrounds, multiple peaks of the missing parts rate, which correspond to the object's colors,

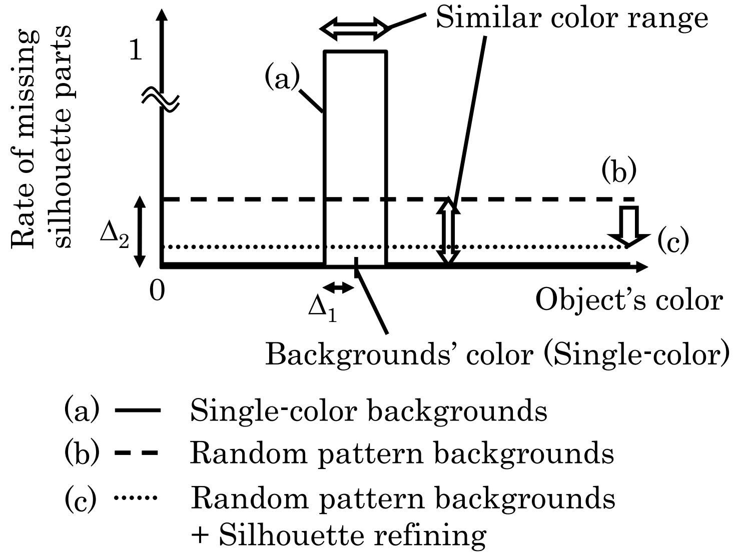

Similar color range
Rate of missing silhouette parts
1
(a)
(b)
(c)
Δ2
0
Δ1
Object's color
Backgrounds' color (Single-color)
(a) — Single-color backgrounds
(b) - - Random pattern backgrounds
(c) ······ Random pattern backgrounds + Silhouette refining

Figure 2.6: Relation between object's color and rate of missing silhouette parts for each kind of background.

appear in the solid line (a). With the random pattern backgrounds, those peaks of the missing parts rate do not appear, even for objects in multiple colors.

Moreover, with the random pattern backgrounds, the object's color and the corresponding background's color are expected to be different in most images, which are provided from the small size of each region of the random pattern. The object's colors are easy to be estimated when the silhouettes are correctly extracted in most images. It leads that the silhouettes are easy to be refined with our method discussed in Section 2.4.

2.3.2 Expected Rate of Silhouette Missing Parts

For extracting silhouettes, many kinds of color spaces have been proposed. In this article, values of U and V of YUV color space are chosen for extracting silhouettes, because we know that silhouette extraction in the U-V space is not affected by shadows on the object or the backgrounds. The values of R,G,B are calculated for those of Y,U,V as in the following equation.

$$\begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1.000 & 0 & 1.402 \\ 1.000 & \text{-}0.344 & \text{-}0.714 \\ 1.000 & 1.772 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{Y} \\ \mathrm{U} \\ \mathrm{V} \end{pmatrix} \tag{2.4}$$

However, our proposed method can be applied to any color space.

Suppose the colors obtained by cameras are represented in the colors of RGB in a color space. Each value of the color is represented in $[0, 255]$. In the U-V space, the obtained colors form a hexagonal region as shown in Figure 2.7(a).

For the object's color, shown as × in Figure 2.7(a), colors in a rectangular region with the width of $2U_{th}$ and the height of $2V_{th}$ are treated as *similar* to the object's color, where U_{th} and V_{th} denote the thresholds for the silhouette extraction. We call this rectangular region the *similar-color region*. The similar-color region occupies an area of $2U_{th} \times 2V_{th}$ at most. When the object's color is around the boundaries of the hexagon region, the similar-color region becomes smaller than $2U_{th} \times 2V_{th}$.

The color for each region of the random pattern is selected with the same probability in the U-V space. A histogram of colors in a random pattern is shown as Figure 2.7(b). If the background's color is incidentally in the similar-color region, the corresponding region of the silhouette is missed. The largest silhouette missing rate p is calculated from the area of the rectangular region as follows:

$$p \leq \frac{4U_{th}V_{th}}{S}, \tag{2.5}$$

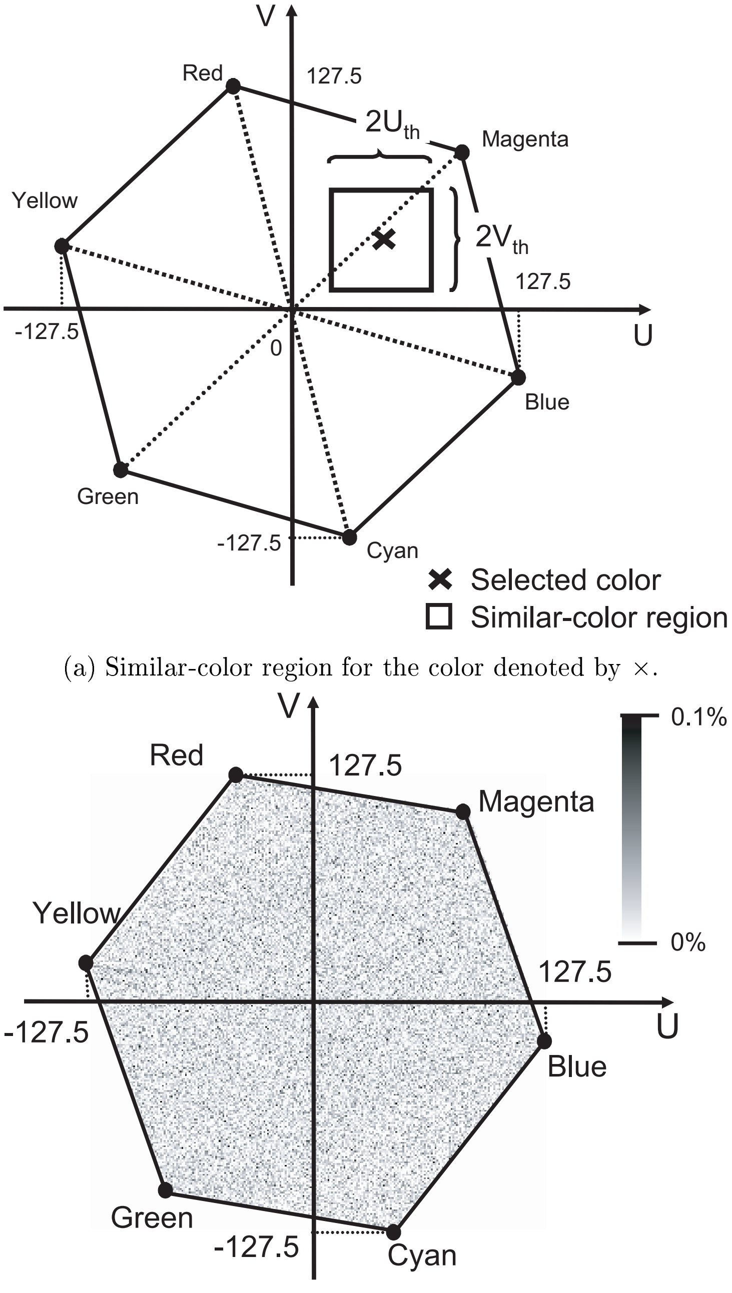
where S denotes the area of the hexagonal region in the U-V space. By using random pattern backgrounds, the rate of silhouette missing parts given is guaranteed to be below p. For example, this value p is calculated to be less than 5.21% when $U_{th} = 10, V_{th} = 10$ as described by Eq. (2.5).

2.3.3 Generation of Random Patterns

In this subsection, we express a method for generating random patterns.

Selection of Color of Each region

The color for each region is selected with the same probability in the U-V space, whereas the value of Y, which controls the brightness of an image, is



(a) Similar-color region for the color denoted by ×.

(b) Histogram of colors in random pattern.

Figure 2.7: When colors are given as values of RGB, which range from 0 to 255, U-V space is described as a hexagonal region. When a color is selected as ×, the similar-color region is described as a rectangular region in the U-V space.

not decided uniquely. If the value of Y is not adequately decided, observed colors are biased in the U-V space. Thus, some regions appear visually as white or black regions, depending on lighting conditions. To address the problem, we set a target value Y_t to fit the lighting conditions. Note that the target value Y_t does not depend on the objects' colors. The value of Y of each region is calculated from given values of U, V and target value Y_t.

Suppose the colors obtained by cameras are represented in RGB color space. Each value of the color is represented in $[0, 255]$. In the YUV space, colors exist within the interior and boundary of a hexahedron as shown in Figure 2.8. The boundaries of the hexahedron are defined by $R = 0$, $G = 0$, $B = 0$, $R = 255$, $G = 255$ and $B = 255$.

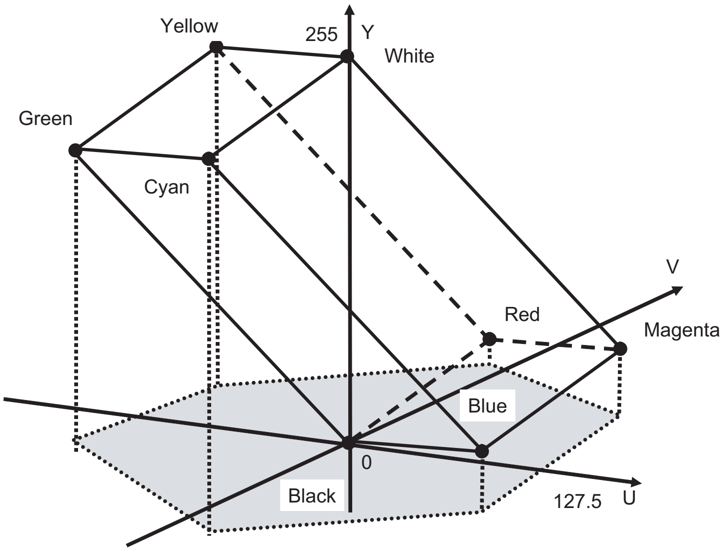


Figure 2.8: Selection of color for each region.

The color of each region is selected with the following procedure.

1. Select a pair of values of U,V at random in $[-127.5, 127.5]$.
2. For the target value Y_t, judge whether the color of (Y_t, U_s, V_s) is in the interior of the hexahedron as shown in Fig 2.8. If the color is in the interior, we adopt the color. If not, go to 3.

3. Search for an intersecting point Y_s between the line $(U, V) = (U_s, V_s)$ and the boundaries of the hexahedron. If there are one or more points, select the nearest Y_s from Y_t, and the color (Y_s, U_s, V_s) is adopted. If no intersection point is found, go to 1 and select a new pair of U_s, V_s.

In Step 2, the intersection points is calculated by the following equations denoted by Eq. (2.4) and $0 \leq R, G, B \leq 255$.

$$
\begin{aligned}
Y_s &= -1.402V_s && (2.6)\\
Y_s &= 0.344U_s + 0.714V_s && (2.7)\\
Y_s &= -1.772U_s && (2.8)\\
Y_s &= 255 - 1.402V_s && (2.9)\\
Y_s &= 255 + 0.344U_s + 0.714V_s && (2.10)\\
Y_s &= 255 - 1.772U_s && (2.11)
\end{aligned}
$$

Selection of Size of Each region

To ensure a randomness of colors of the pattern, each region of the pattern needs to be set small enough. However, when the size is set below a certain level, edge blurring would be a problem [34]. In regions with edge blurring, colors among adjacent colors are observed. Therefore, observed colors are biased in the U-V space.

To set the size of each region, we made many sizes of random pattern backgrounds and set them in our experimental environment. The best size, which best kept the randomness of colors, was selected. The size depends on experimental environments, do not on objects. This means that the size needs not to be changed by objects.

2.4 Recovering Missing Parts of Silhouettes

Let us focus on a part of the object. With the random pattern backgrounds, most parts of the backgrounds for the corresponding part of the object are expected to have different colors in each image as shown in Figure 2.9. This means that the silhouettes of the object can be extracted correctly in the

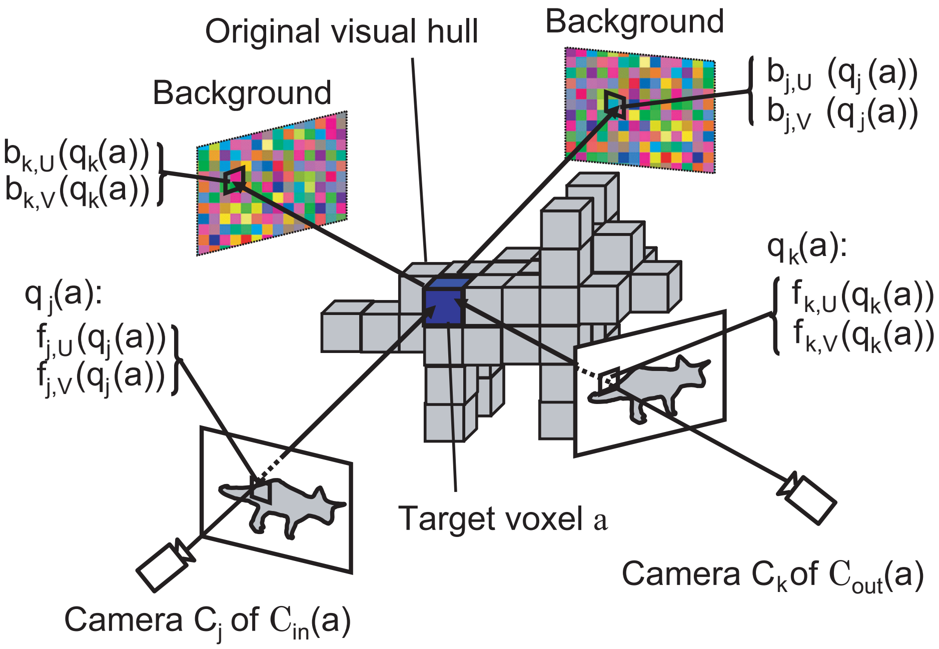


Figure 2.9: Object color estimation.

images from most of cameras. From the colors of the regions of those silhouettes, the color of each part of the object can be estimated. By comparing an estimated color with the background color captured by each camera, we can detect and refine the missing parts of the silhouette as described in detail.

The silhouettes are extracted using conventional background subtraction with the thresholds U_{th} and V_{th}. We call the VH, which is reconstructed from the original silhouettes, the original VH. Let us denote one of the neighboring voxels of the original VH by a. The color of the voxel a can be obtained from the images of multiple cameras observing a. Let $\mathcal{C}_{vis}(a)$ denote the set of cameras observing a, and let $N_{vis}(a)$ denote the number of cameras in $\mathcal{C}_{vis}(a)$. We denote the set of the cameras in $\mathcal{C}_{vis}(a)$ for which a is projected into the inside or the border of the silhouettes by $\mathcal{C}_{in}(a)$. The set of the cameras that are not included in $\mathcal{C}_{in}(a)$ is denoted by $\mathcal{C}_{out}(a)$. The color of voxel a can be estimated from the images in $\mathcal{C}_{in}(a)$ as shown in Figure 2.9. We denote U and V values of a pixel to which a is projected for the image of camera C_i by $f_{i,U}(q_i(a))$ and $f_{i,V}(q_i(a))$. For the background in the image

C_i, we denote U and V values of a pixel to which a is projected by $b_{i,U}(q_i(a))$ and $b_{i,V}(q_i(a))$. The estimated U and V values of a, which are denoted by $Ave_U(a)$ and $Ave_U(a)$, are calculated as follows :

$$
\begin{aligned}
Ave_U(a) &= \frac{\sum_{C_i \in \mathcal{C}_{in}(a)} f_{i,U}(q_i(a))}{N_{vis}(a)} && (2.12)\\
Ave_V(a) &= \frac{\sum_{C_i \in \mathcal{C}_{in}(a)} f_{i,V}(q_i(a))}{N_{vis}(a)}. && (2.13)
\end{aligned}
$$

When the differences between $Ave_U(a)$ and $Ave_V(a)$ with $b_{k,U}(q_k(a))$ and $b_{k,V}(q_k(a))$ for the image of $C_k \in \mathcal{C}_{out}(a)$ are not substantially large, the pixel $q_k(a)$ is regarded to be missed and added to the silhouette of C_k, because the object region cannot be discriminated from the background with the conventional background subtraction under this condition.

$$
\begin{aligned}
&\mid Ave_U(a) - b_{k,U}(q_k(a)) \mid < U_{th}\\
&\qquad or\\
&\mid Ave_V(a) - b_{k,V}(q_k(a)) \mid < V_{th} \qquad (2.14)
\end{aligned}
$$

This procedure is at first applied to the voxels neighboring the voxels of the original VH. The original VH is refined by reconstructing VH from the silhouettes obtained in the step above. The $\mathcal{C}_{in}(a)$, $\mathcal{C}_{out}(a)$ and $\mathcal{C}_{vis}(a)$ are also recalculated in this step. The procedure is repeated by choosing a voxel from the refined VH until no pixel is added to the silhouettes in the previous procedure.

2.5 Experimental Results

In our experiment, the shape of a bumpy triceratops toy and a horse toy were reconstructed with the volume intersection method. We used 19 cameras surrounding them as shown in Figure 2.10 [17]. Positions and colors of all cameras were calibrated in advance. Each region of the random pattern backgrounds was printed in the color randomly chosen from the U-V space. The random pattern backgrounds were secured to plastic boards. Both U_{th} and V_{th} for the silhouette extraction were set to 10 in this experiment. To evaluate the silhouettes, we used silhouettes that were extracted manually as correct silhouettes.

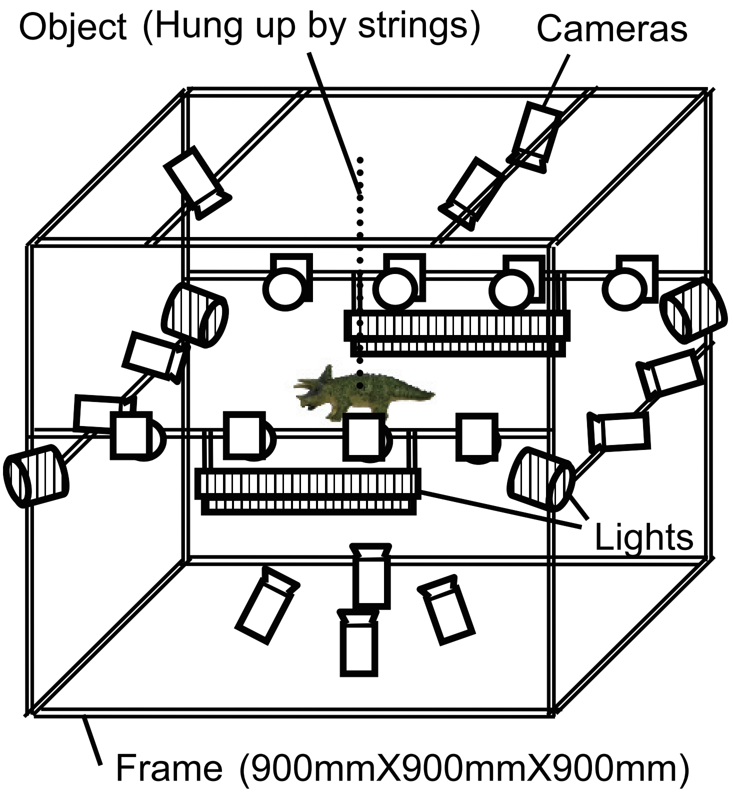


Figure 2.10: 4π measurement system[17].

2.5.1 Silhouette Refining

As described in Section 3, the rate of missing silhouette parts is below a specific percentage due to use of the random pattern backgrounds. In this subsection, we confirmed that the rate of the missing parts indicated as described in Section 2.3. A triceratops toy and a horse toy are selected as the objects. A sample image from a camera is shown as Figure 2.11(a). The silhouette extracted under the random pattern backgrounds is shown as white regions in Figure 2.11(b). When the silhouettes are extracted from the obtained images in YUV color space, the regions which are filled with white are given as the silhouettes as described in Figure 2.11(c) and Figure 2.12(c). We can observe small holes only in the silhouettes. With the random pattern backgrounds, even though some parts of silhouettes are missed, most parts of the silhouettes are correctly extracted. The missing parts are below a certain rate.

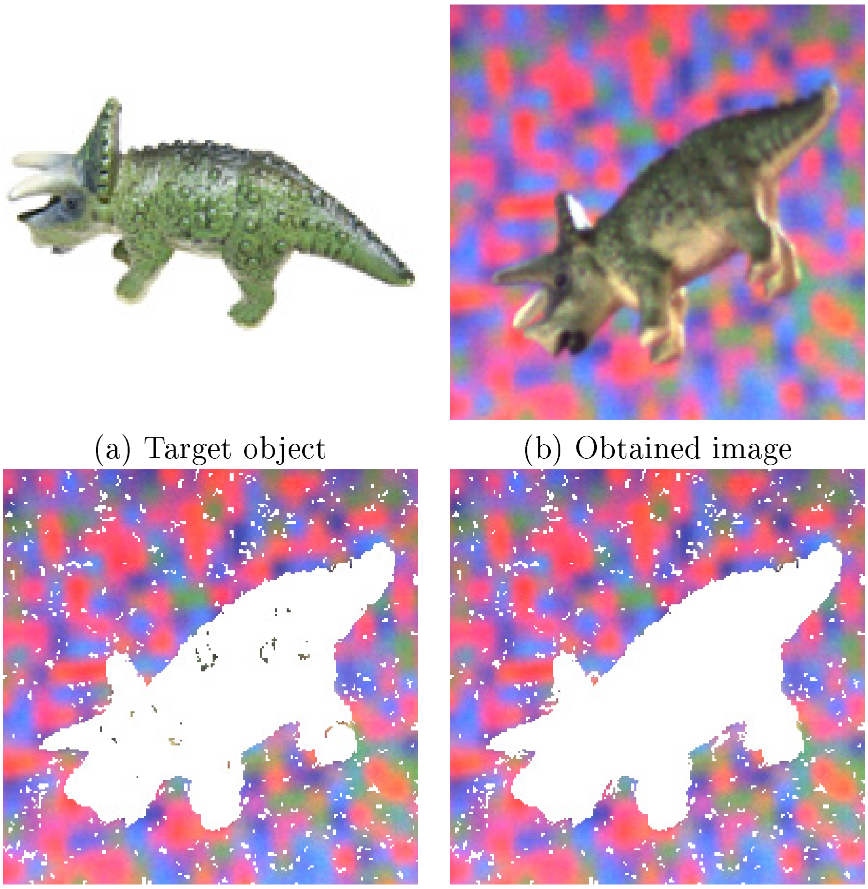
Although a few small regions are missing in the silhouette, no large missing part is found. This means that parts of the silhouette were actually missed, however, the missing rate was suppressed to below a small amount by using the random pattern backgrounds. Unlike the blue screen matting, the missing rate does not depend on colors of objects.

With manually extracted silhouettes as answer silhouettes, the extracted silhouettes with the random pattern backgrounds are evaluated. On average, 3.55% of the whole silhouettes was missed (14.11% at a maximum, 0.87% at a minimum) for the triceratops, and 4.88% of the whole silhouettes was missed (12.57% at a maximum, 2.23% at a minimum) for a horse in Figure 2.15(a).

From Eq. (2.5), the rate is expected to be less than 5.21%. The experimental rates were less than the calculated rate on average, but were more than the calculated rate at the maximum. This was caused by biased color observation in a real environment.

The silhouettes with missing parts recovered with our method are shown in Figure 2.11(c). In this result, the rates of missing parts were reduced to 0.89% of the whole silhouette on average (1.81% at a maximum, 0.09% at a minimum) for the triceratops, and 2.33% of the whole silhouette on average (4.33% at a maximum, 0.97% at a minimum) for the horse. Notably, the maximum rates were markedly decreased.

The histograms in UV space are calculated by the obtained colors from random pattern images in a computer, as shown in Figure 2.13(a). They also calculated by the obtained colors from the random pattern backgrounds in



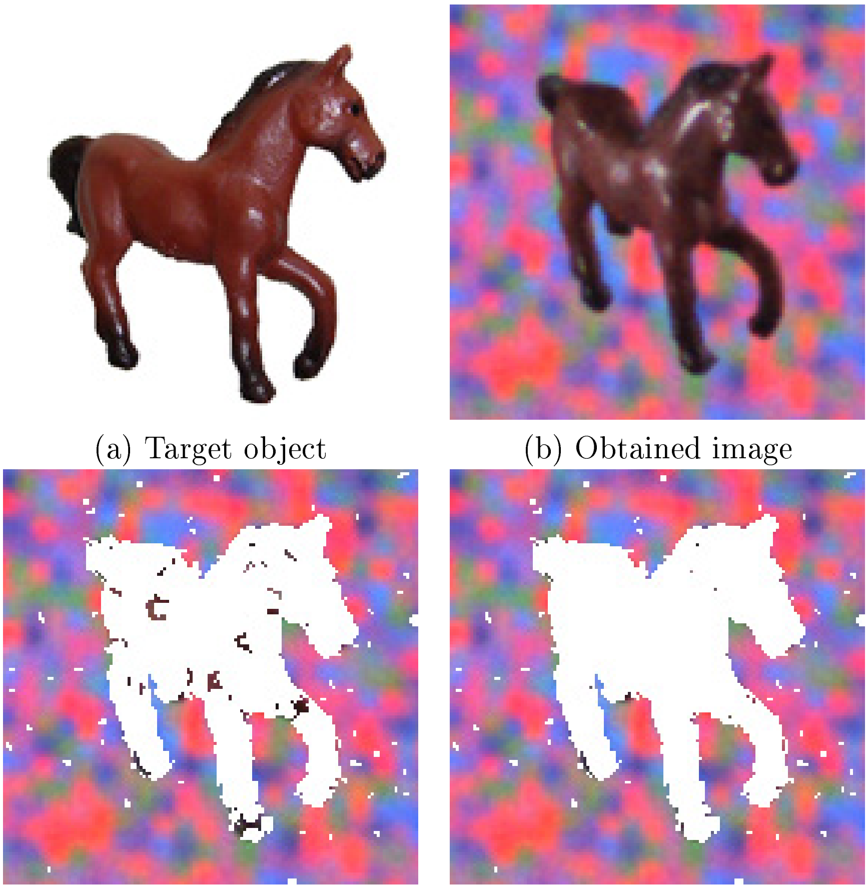
(a) Target object

(b) Obtained image

(c) Silhouette without refining

(d) Silhouette with refining

Figure 2.11: An example of silhouette refining with random pattern backgrounds. (Triceratops)



(a) Target object

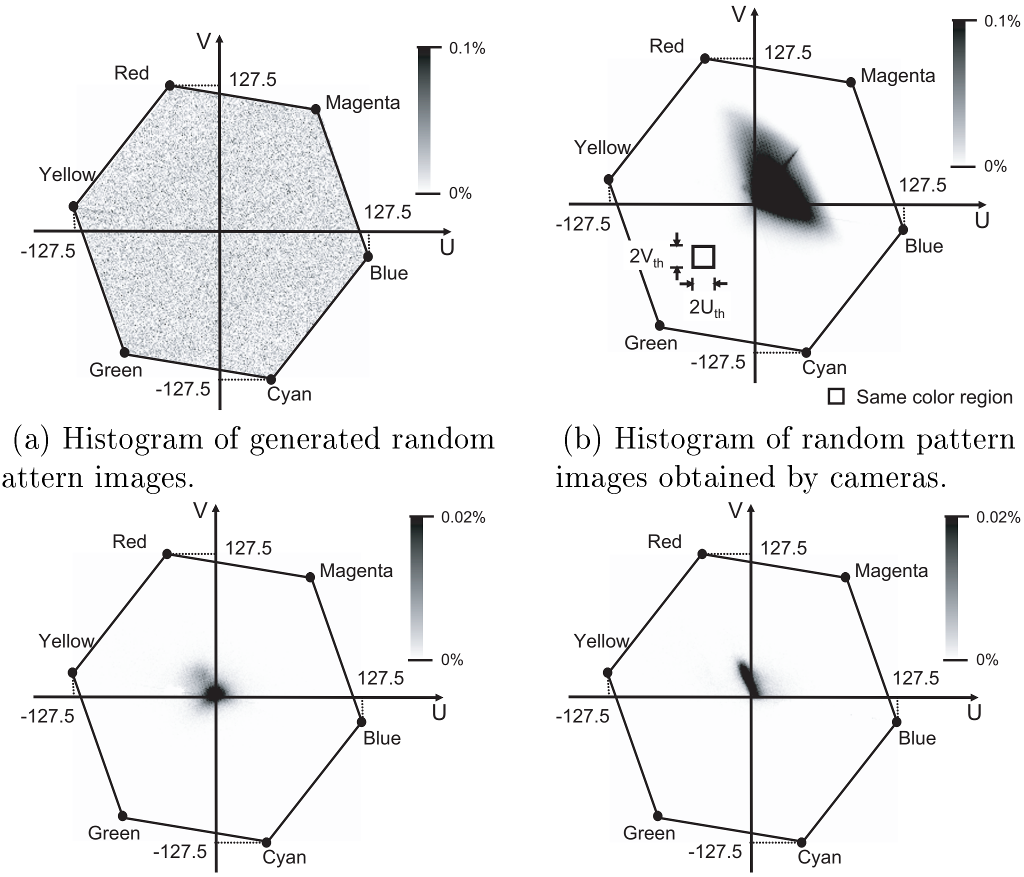
(b) Obtained image

(c) Silhouette without refining

(d) Silhouette with refining

Figure 2.12: An example of silhouette refining with random pattern backgrounds. (Horse)

real environment, as shown in Figure 2.13(b). Whereas, the histograms from regions of the triceratops and horse are given as shown in Figure 2.13(c) and Figure 2.13(d). The rate of each point is represented with thickness of each pixel.



(a) Histogram of generated random pattern images.

(b) Histogram of random pattern images obtained by cameras.

(c) Histogram of regions of Triceratops.

(d) Histogram of regions of Horse.

Figure 2.13: Histograms of random pattern images and target objects.

The histogram for random pattern images in a computer is uniformly distributed as shown in Figure 2.13(a). The pattern is an ideal pattern which guarantees to extract silhouettes below a certain rate for all color objects. The histogram shown in Figure 2.13(b) is distributed to the center.

It is considered that lighting environment makes the distribution of observed colors by inappropriate settings of a printer or cameras. The observed colors, however, cover the similar color region as shown in Figure 2.13(b). Then, the random pattern background is expected to satisfy the requirement described in Section 2.3. The background should guarantee the certain rate of the missing parts. Although the selected objects have similar colors with those of observed random pattern backgrounds, the missing rate is below the expected rate even for the objects. It means that the requirement which the rate of the missing parts is below the expected rate will be satisfied, even if the observed random pattern backgrounds have sort of color distribution.

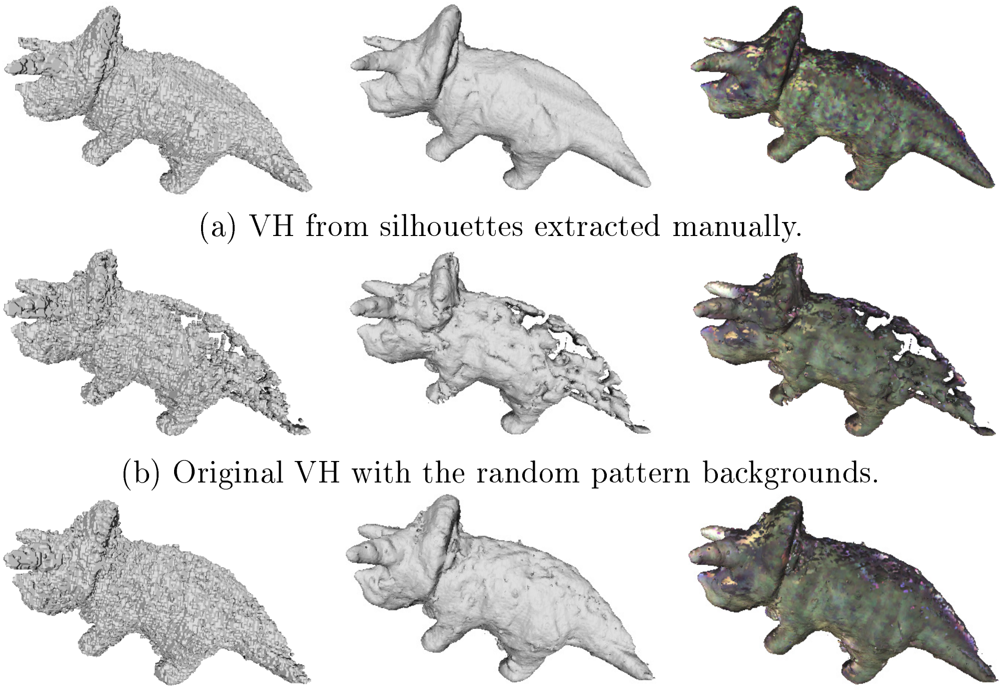
2.5.2 Shape after Refining

The correct shapes reconstructed from a set of the correct silhouettes, which were extracted manually, are shown in Figure 2.14(a). Compared with the VHs reconstructed from the original silhouettes (Figure 2.14(b) and each figure of (c) and (e) in 2.15, 2.16, 2.17 and 2.18, the VHs recovered with our proposed method (Figures 2.14(c) and each right figure of (d) and (f)) have fewer missing parts. The original VH of the triceratops (Figure 2.14(b)) has large parts missing from the back and the tail. The missing parts of the VHs are formed from the missing parts of the silhouettes accumulated in the resultant VHs in the process of calculating the intersection of the visual cones associated with the silhouettes. The missing parts form large holes in the VHs. The holes of the reconstructed VHs are filled using our proposed method.

Tables 2.1 and 2.2 explain the comparison of the original VH and the VH refined with our proposed method. As shown in these tables, the numbers of missing voxels were drastically decreased. Although the number of voxels were increased, the sum of the missing voxels and the additional voxels, or “error voxels”, were decreased as a whole.

2.6 Discussion and Conclusions

We proposed a method of using the volume intersection method to reconstruct correct shapes even for objects of unknown color by using random pattern backgrounds. Using random pattern backgrounds keeps the amount of missing parts of the silhouettes below a specific percentage, even for ob-

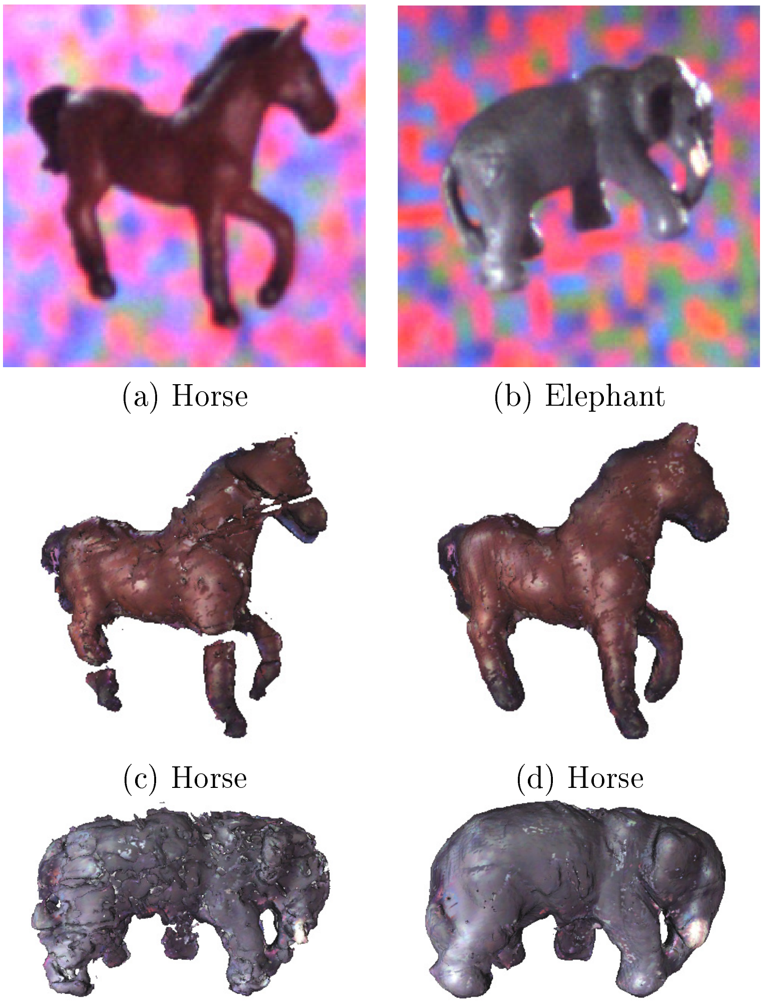


(a) VH from silhouettes extracted manually.

(b) Original VH with the random pattern backgrounds.

(c) Refined VH with out proposed method.

Figure 2.14: VH of a triceratops toy with recovering for silhouette missing parts. Even if reconstructed shapes are applied to smoothing and coloring processes, missing parts of shapes cannot be recovered in appearance. (Left) Each shape drawn with surface patches obtained by marching cube algorithm [28] for the resultant VH. (Center) Shapes of the left ones with surface smoothing. (Right) Colored Shapes of the center ones with a Naive algorithm which is a viewpoint independent patch-based method[5].



(a) Horse

(b) Elephant

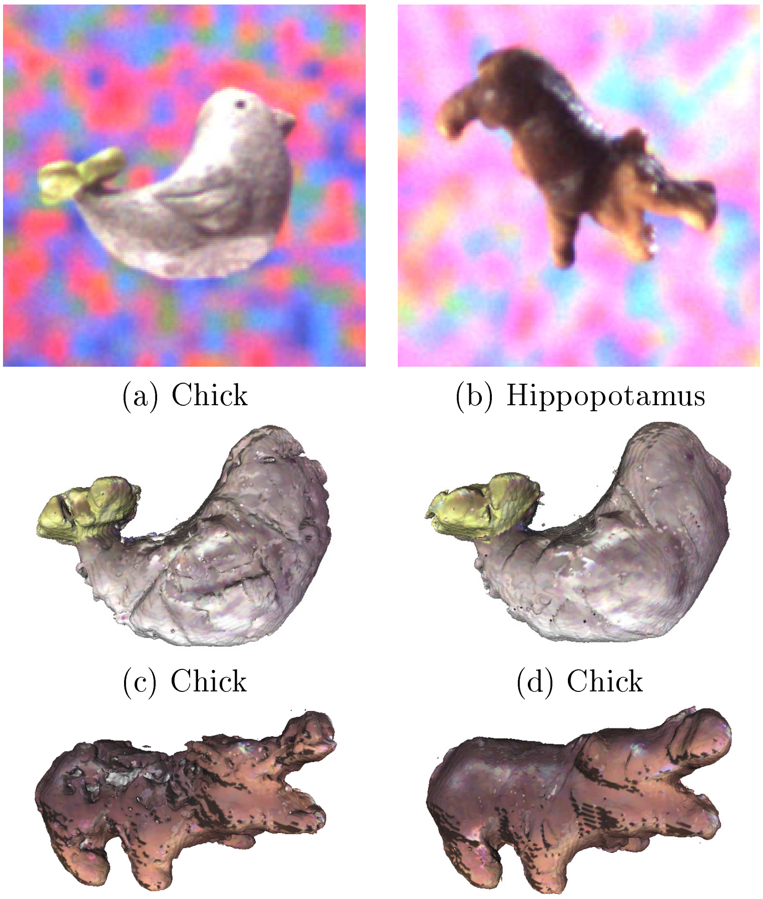
(c) Horse

(d) Horse

(e) Elephant

(f) Elephant

Figure 2.15: Colored Shapes of several toys. (Horse and elephant.)



(a) Chick

(b) Hippopotamus

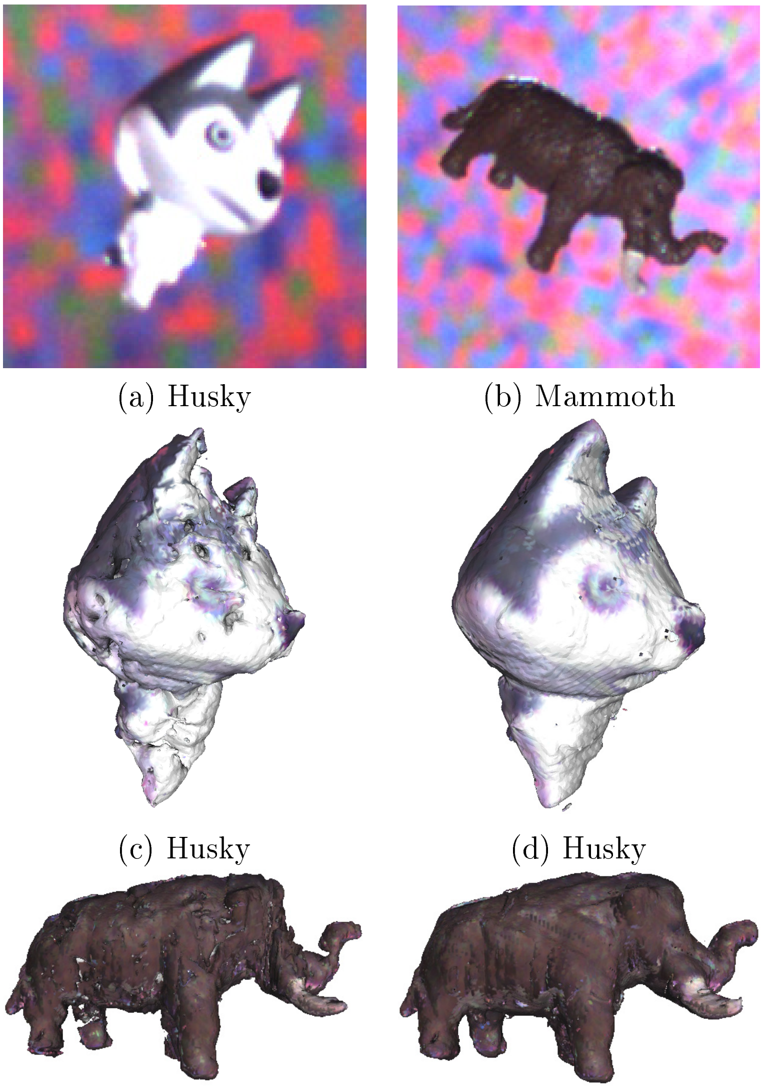
(c) Chick

(d) Chick

(e) Hippopotamus

(f) Hippopotamus

Figure 2.16: Colored Shapes of several toys. (Chick and hippopotamus.)



(a) Husky

(b) Mammoth

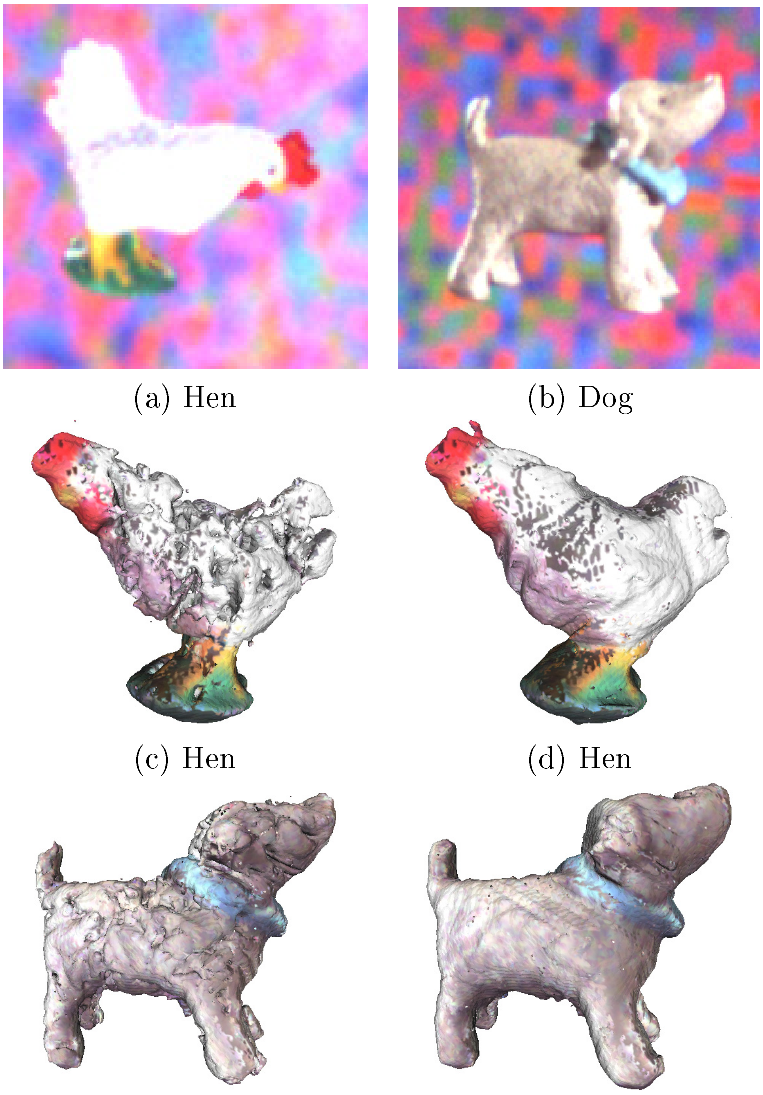
(c) Husky

(d) Husky

(e) Mammoth

(f) Mammoth

Figure 2.17: Colored Shapes of several toys. (Husky and mammoth.)


(a) Hen

(b) Dog

(c) Hen

(d) Hen

(e) Dog

(f) Dog

Figure 2.18: Colored Shapes of several toys. (Hen and dog.)

Table 2.1: Error voxels when using our proposed method. (Triceratops)

	Additional voxels	Missing voxels	Error voxels
Original VH	8297 (3.1%)	153300 (57.5%)	161597 (60.6%)
Refined VH	30068 (11.3%)	7252 (2.7%)	37320 (14.0%)

Table 2.2: Error voxels when using our proposed method. (Horse)

	Additional voxels	Missing voxels	Error voxels
Original VH	5720 (2.5%)	153129 (67.2%)	158849 (69.8%)
Refined VH	25297 (11.1%)	18051 (7.9%)	43348 (19.0%)

jects of unknown color. Correct silhouettes are obtained by adding the missing parts detected from the inconsistency of pixel colors from the random pattern backgrounds. By using the random pattern backgrounds and the missing parts of a silhouette missing recovered as described above, a shape with fewer missing parts can be obtained. In our experiment, we confirmed that a correct silhouette can be obtained by comparing shape reconstructed using the proposed method and the shapes of silhouettes manually extracted.

As feature work, we plan to adjust the size of each region of the random pattern backgrounds so that the region is sufficiently small for a randomness of background colors that does not cause edge blurring by setting the positions of the cameras and the objects, as well as camera parameters, appropriately.

Chapter 3

Outcrop Points for Motion Estimation

3.1 Introduction

In the volume intersection method, the 3D shape of the object is reconstructed from obtained silhouettes. Compared with approaches based on stereo vision or laser range finders, the method needs neither preprocess for extracting detailed image features nor laser light source. It can be employed even for reconstructing shapes of objects that have no texture or absorb laser light. This advantage of the method makes itself applicable to shape reconstruction for natural history, archeology objects, etc. In the volume intersection method, additional regions in the reconstructed shape decrease with an increase number of cameras. More accurate shape is obtained with more number of cameras. However, it is not realistic to install so many cameras around the object due to physical limitation on their spatial configurations. In this Chapter and next Chapter, we discuss reconstructing accurate shapes from silhouettes obtained from small number of cameras.

We assume that the object is in a rigid motion. When the object is in a rigid motion, the cameras change their positions to the object at every moment. If the rigid motion of the object can be correctly estimated, cameras at different time are treated as cameras in different positions virtually. With these virtual cameras, we can improve accuracy of the reconstructed 3D shape without increasing the number of cameras.

It has been proposed in many previous works on the volume intersection method to improve accuracy in the reconstructed shape. In some of those works, the object is observed by cameras while being rotated by turntables [49, 56] . The cameras observing a rotated object can be treated as many cameras observing a stable object. The relative position between each camera and the object is calculated from the rotation of the object. Although this approach using the turntable is easy and effective for introducing motion of objects to the volume intersection method, it is possible that motion of the object is limited only to the rotation around the axis of the turntable. Moreover, the turntable needs to be observed together in images. It means that the images of cameras which are occluded by the turntable are not utilized. In the volume intersection method, many images of many viewpoints are required for accurate shape reconstruction.

Cheung et al. [4] have proposed to extract new feature points for tracking the moving object from images. The relative positions between the object and cameras are changing in multiple frames. If the rigid motion is estimated, images from other positions are virtually obtained. The feature point extraction method by Cheung et al. employs color information of the object. It loses the advantage of the volume intersection method that does not need color information of objects and can be applied even to objects without texture, as discussed above.

In other works, it is also proposed to extract feature points from the visual hulls based on the epipolar geometry [6, 9, 13, 10, 53] . These feature points are called *frontier points*, or *epipolar tangencies* because they are derived from epipolar constraints. It is efficient to extract some feature points for estimating the object motion. However, these feature points are not guaranteed to be included in the object region completely. When the object shape is complicated, the feature points tend to be included in the additional region. It is difficult to extract feature points only from the object region, not from the additional region. The visual hulls in multiple frames have the additional regions which changes their shapes. The point on the surface of the visual hull at a frame might be occluded in the additional region at the other frames.

To address the problem, we propose a method to extract feature points by projecting surface voxels of the visual hull to the silhouettes. With the method, we can detect the voxels which are guaranteed to be included in the object region. The extracted voxels are called *outcrop points*.

In this Chapter, we propose a new kind of feature points called *outcrop points*, The outcrop points are useful for the object motion estimation from the visual hulls in multiple frames. The outcrop points are extracted based on the silhouettes without colors.

In the remainder of this Chapter, we will propose a method for object motion estimation with the outcrop points. In Section 3.2, we give a brief summary of the accuracy limit of reconstructed shapes by integrating visual hulls in multiple frames. In Sections 3.3 and Section 3.4, we propose the procedure for extracting the outcrop points from visual hulls and that for estimating rigid motions of objects based on the outcrop points. Experimental results are given in Section 3.5, and we conclude this Chapter in Section 3.6.

3.2 Reconstruction of Accurate Shapes

Silhouettes are extracted from sampling images. The silhouette sampling restricts the accuracy of reconstructed shapes with the volume intersection method. The reconstructed shapes are also affected by the sampling. This restriction gives a difficulty for our method, which is the volume intersection method in multiple frames. We discuss the sampling effect for setting a requirement value. Niem et al. [36] have discussed the sampling effect for reconstructed shapes. The sampling error was theoretically estimated in their discussion. The estimated error is not practical, since they estimated the error as a theoretical maximum. In this Section, we discuss the practical sampling error of the reconstructed shapes with several simulation data.

Every voxel in an observation region is projected to the plane of projection of each camera. The projection image of the voxel has a width in the plane. Let q at the maximum of the width for all voxels in the observation region. q is given by settings, positions and resolutions of cameras. The size of voxels also affects q. When q is set to a large value, the sampling error for the visual hulls becomes small. Here, we discuss the accuracy limit of reconstructed shapes. The limit is defined by the sampling error.

The color value of a pixel determines whether the pixel is included in the silhouette region. The color value is defined by the color value of the sampling point of the pixel. Both of the regions in the silhouettes and out of the silhouettes can be included in a pixel. The pixel includes parts of continuous-valued contours of the silhouettes. Consider such pixel. We discuss the distance between the sampling point and the contours of the silhouettes. Let k be the maximum width of a pixel in images. The shape error on the surface is k/q at a maximum. A contour of the silhouette is included in the pixel as shown in Figure 3.2.

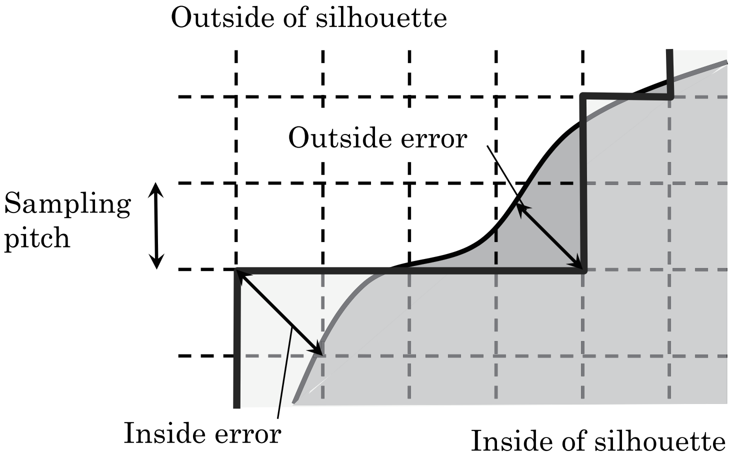


Figure 3.1: Error on surface of visual hull by sampling.

In the pixel of the image, a point out of the silhouette can be located $\sqrt{2}$ pixels in distance from the contour. In a voxel of the observation region, a point out of the shape can be located $\sqrt{2}/q$ in distance from the surface of the shape. Then, the sampling error in images is $\sqrt{2}$ near the contour of the silhouettes. The sampling error in the observation region is $\sqrt{2}/q$ near the surface of the shapes. The sampling errors provide the accuracy limit of the shape reconstruction.

Practically, the sampling errors are not so much. Assume that the silhouettes are defined by the values of pixels. When more than a half of a pixel is the silhouette region, the color value would be the object color, and the pixel would be included in the silhouette. When more than a half of a pixel is out of the silhouette region, the color value would be the background color, and the pixel would not be included in the silhouette. In this case, the sampling error in images is $\sqrt{2}/2$ at a maximum. For a sphere as the target object, the error region on the surface of the sphere is estimated to α of the whole region as follows:

$$\alpha = \frac{\frac{k}{q} \cdot 4\pi r^2}{\frac{4}{3}\pi r^3} = \frac{3k}{qr}, \tag{3.1}$$

where r is the radius of the sphere, and r satisfies $r \gg k/q$.

From eq. 3.1, when $r = 50$, $q = 0.75$ and $k = \sqrt{2}/2$, the sampling error on the sphere surface α is calculated to 5.67%. For the objects other than spheres, the surface area of the objects is larger than that of the sphere. α of the objects becomes larger than that of the sphere, which is represented by eq. 3.1. Let N_v be the number of voxels included in the object region, N_v is same to the number of voxels of the sphere which radius is $(\frac{3N_v}{4\pi})^{\frac{1}{3}}$. The rate of the error voxels on the surface is more than $\alpha = \frac{3k}{q}\left(\frac{3N_v}{4\pi}\right)^{-\frac{1}{3}}$. For some simulation objects, the numbers of cameras with which reconstruct shapes with the accuracy limit are estimated based on the rate α. New cameras are located at a random position on the surface on a unit sphere. With increasing the number of cameras, we examine the transition of the number of the additional voxels. Target objects are a sphere, a cube, a torus and a triceratops used in Section 3.5. q, N_v and α of each object are shown in table 3.1. The transition of the number of the additional voxels is drawn in Figure 3.2. The longitudinal axis is represented by coefficients of α.

For the sphere, the rate of the additional voxels to the voxels of the visual hull is less than α, when images are obtained by 49 cameras. For the objects other than the sphere, the rates are less than α, when more than 49 cameras are used. For the objects with outstanding points, more cameras are required to decrease the rate.

Table 3.1: Parameters of target objects.

	q	N_v	α	Num. of cameras
Sphere	1.679	1265791	1.88%	49
Cube	1.679	2424294	1.52%	406
Torus	1.679	268545	3.16%	103
Triceratops	1.679	438310	2.68%	242

The error voxels represented by α are caused by the sampling error. The error voxels would not be refined by integrating silhouettes in multiple frames. We set the accuracy limit of shapes from the limit of the sampling error. The accuracy limit is the goal for shapes reconstructed by integrating silhouettes in multiple frames, If the summation of the additional and missing voxels is below α, the reconstructed shapes reach to the goal.

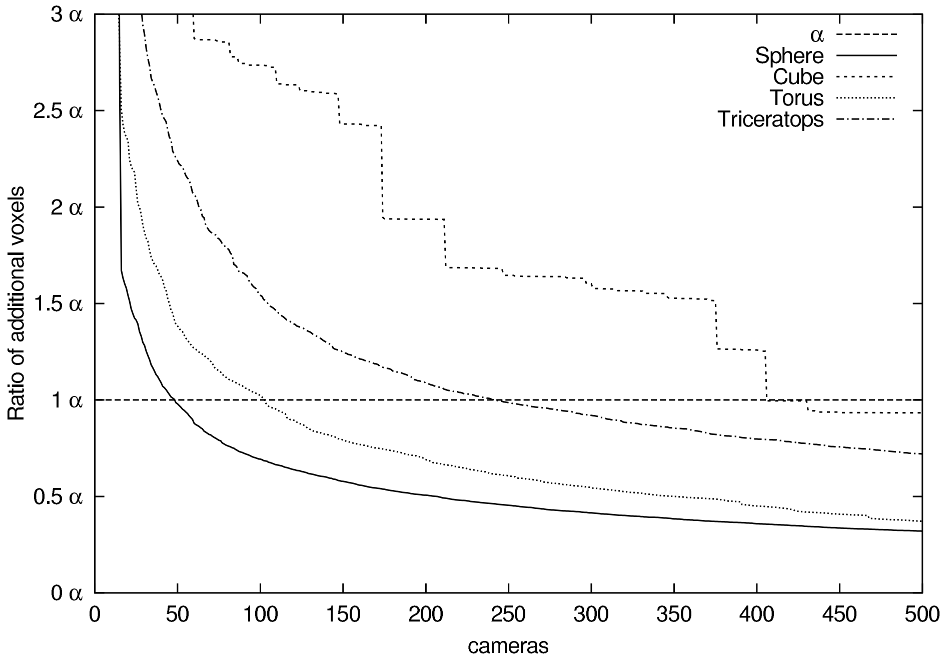


Figure 3.2: Relation between camera number and additional regions.

3.3 Extracting Feature Points

3.3.1 Required Conditions for Feature Points

In order to estimate the motion of the target object, we extract some feature points from the visual hull at each frame. The following conditions need to be satisfied by the feature points:

(1) Feature points are included in the visual hull at each moment.

(2) Feature points continue to be tracked from the object in motion.

Whereas the object region is included in the visual hull for all the frames, the additional regions included in the visual hull changes with the motion of the object. It depends on the relative position between the object and the cameras. From this fact, any voxel in the additional regions of the visual hull does not satisfy condition (1). Such voxel in the additional regions might not be included in visual hulls in some frames. In order to satisfy condition (1), the feature points have to be included in the object region of the visual hull.

However, it is not easy to extract the voxels that are guaranteed to be in the object region from the visual hull. Only visual hull and silhouettes are given to extract such feature points. We cannot detect where the object region of visual hull is, since the additional regions around the object region in the visual hull changes at every frame. Any voxel of the object region could be occluded from the additional regions of the visual hull.

The condition (2) is also difficult to be satisfied. To satisfy the condition (2), same voxels in two different frames have to be extracted from visual hulls in the two frames. The color feature is useful for identifying each point on the surface of the object. If we can identity each point on the surface at every frame, it is easy to extract same points at different frames. However, we do not assume that the object has sufficient color feature as discussed in Chapter 1.

We will propose a method to extract the voxels called *outcrop points*. The outcrop points are guaranteed to be in the object region in the visual hull. It satisfies the condition (1). They are extracted only from the silhouettes and the visual hull.

Since the outcrop points do not satisfy condition (2) completely, we further narrow them down available voxels for object motion estimation. It is realized to apply to robust estimation approaches. This process will be described in Section 3.4.

3.3.2 Frontier Points

In the previous work, there are some proposals for extracting feature points from silhouettes. These feature points are called *frontier points* or *epipolar tangencies* [6, 9, 13, 10, 53]. In this subsection, we discuss the problem of these feature points.

When an object is observed by a pair of cameras, a plane is composed of a 3D point on the surface and the optical centers of the pair of cameras. The plane is called *epipolar plane*. Based on this epipolar geometry, the point included in the object region has been considered to be extracted. As illustrated in Figure 3.3, when a epipolar plane tangents to points on the surface, the points are considered to be included in the object region of visual hull. The points are defined the frontier points.

The frontier point seems to satisfy the condition (1). However, the frontier

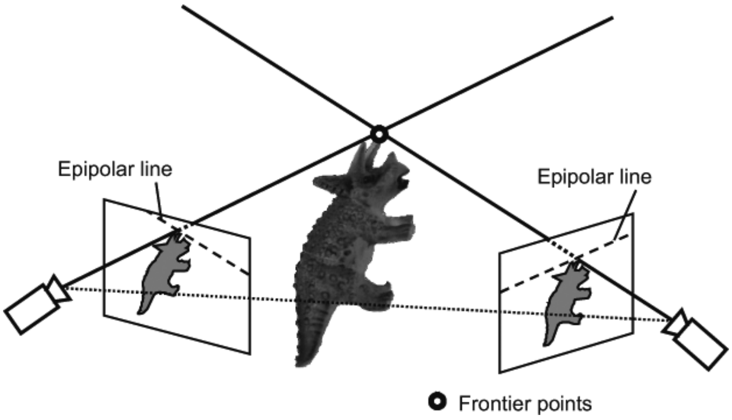


Figure 3.3: Frontier point extraction.

point is not guaranteed to be included in the object region of the visual hull. If an epipolar plane tangents to more than one point on the surface of the object, some false frontier points might be extracted, as shown in Figure 3.4. In the Figure, the point represented by an open circle is extracted as a frontier point, unless the point is not included in visual hull. The point is regarded as the frontier point in spite that it is not actually included in the object region.

This problem occurs when an epipolar plane has more than one tangent point on the object region, When the object has a complicated surface, the problem often happens. Due to the problem, the frontier points cannot be used for the feature points satisfying the condition (1). In the next subsection, we propose a method to extract the new feature points that are guaranteed to satisfy the condition (1).

3.3.3 Outcrop Points

When voxel v in the visual hull satisfies the following conditions as illustrated in Figure 3.5, we call v *outcrop point*:

1. When v is projected onto an image plane of each camera, the projected pixel of v is in contour of the silhouette for at least one camera.

2. For each camera satisfying the condition above, any other voxel of

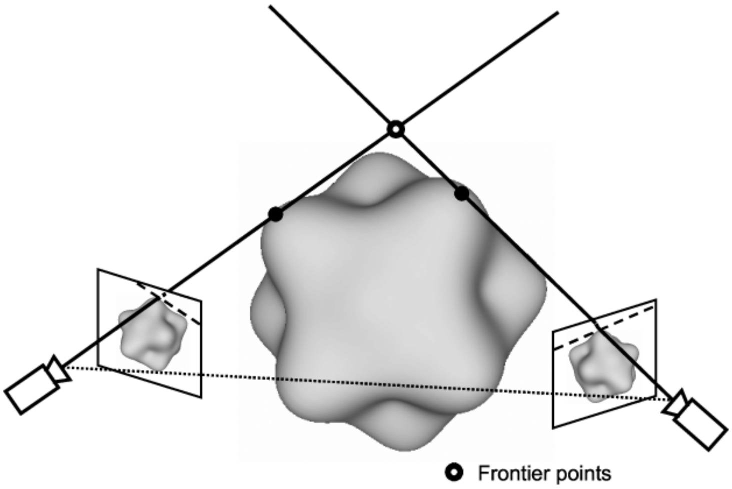


Figure 3.4: Missing points of frontier points.

visual hull is not projected to the pixel.

In Figure 3.5, $contour_j$ is a set of pixels on the contour of the silhouette for camera C_j, and V is the visual hull reconstructed with the silhouettes of all the cameras.

In principle, the outcrop points are guaranteed to be included in the object region. If the outcrop point v satisfies the condition 1 in spite that v is not actually included in the object region, any other voxels are required to be projected to the pixel to which v is projected. Due to the condition 2, any other voxels are not projected to the pixel. A pixel to which no voxels are projected is not a element of the silhouette. If any other voxels are not projected to the pixel, v is a voxel included in the object region.

Outstanding points on the surface of the object tend to be extracted as the outcrop points as shown in Figure 3.6, since the contour of the silhouette for each camera is used to extract the outcrop points. The outstanding points tend to be projected to the contour pixels, even when a relative position between the object and cameras changes.

When the motion of the object is not so large, the same set of outcrop points tends to be extracted during the motion. This tendency is appropriate for satisfying the condition (2) for the feature points. However, all outcrop points are not on the outstanding parts of the surface of object. Some out-

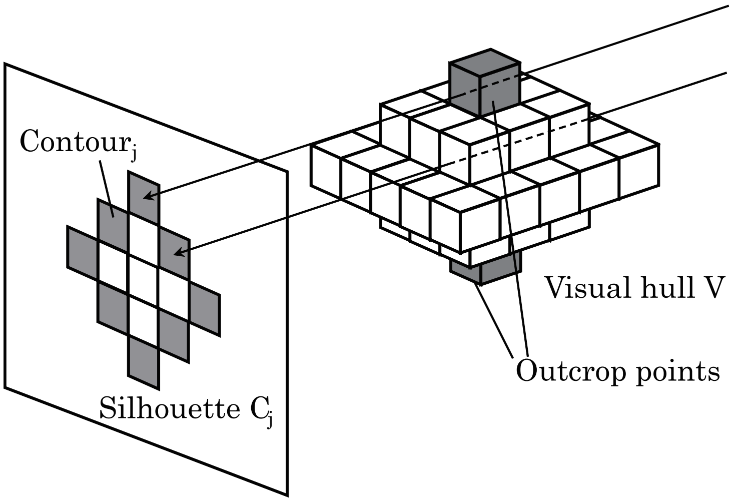


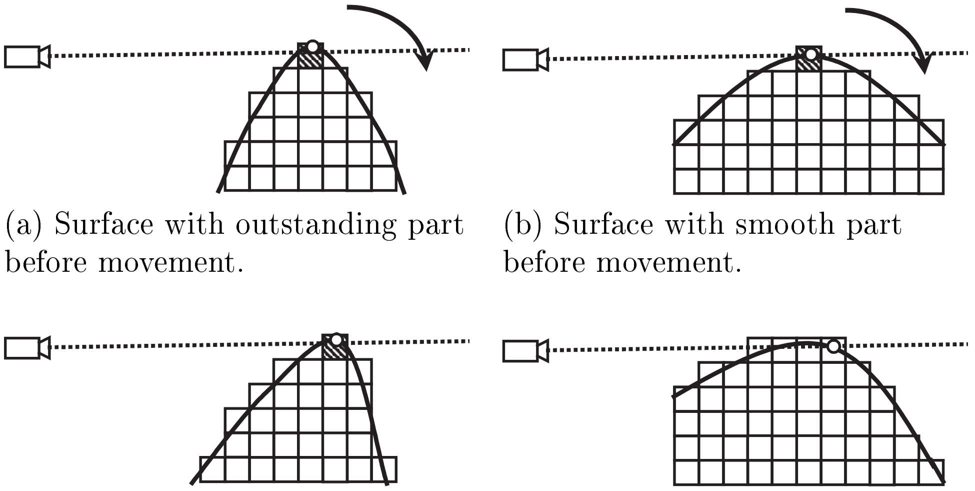
Figure 3.5: Project voxels to a silhouette.

crop points on a smooth part does not continue to be extracted during the motion. In order to cope with this problem, we introduce a robust estimation method. In the robust motion estimation, the outcrop points which have no corresponding points during the motion are not considered for estimating the motion. The detailed procedure will be given in the next Section.

3.3.4 Outcrop Point Extraction and Sampling Grids in 2D Images and 3D Space

When sampling grids in 2D images and 3D space are not appropriately set, outcrop points may not be extracted. Let us denote the 2D sampling grid by d_{2D} and the 3D sampling grid by d_{3D}. d_{2D} is defined by the size of images and the number of pixels in the images. d_{3D} is defined by a system designer. d_{2D} and d_{3D} are independent to each other.

If $d_{2D} \gg d_{3D}$, there are fewer pixels to which only one voxel is projected. This means that there are less extracted outcrop points. Whereas, if $d_{2D} \ll d_{3D}$, almost every pixel in an image has one voxel that is projected to the pixel. This means that most of voxels on the object surface are extracted as outcrop points. However, not so many voxels are located on the object surface when d_{3D} is large.



(a) Surface with outstanding part before movement.

(b) Surface with smooth part before movement.

(c) Surface with outstanding part after movement.

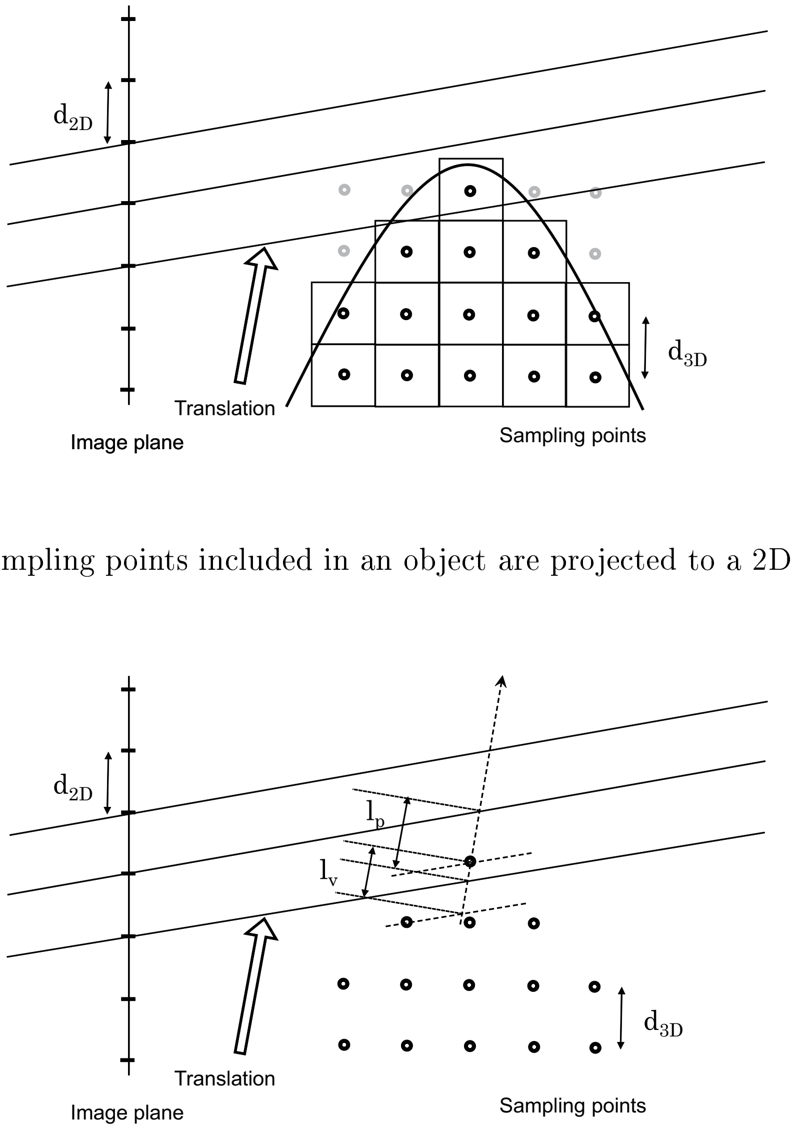
(d) Surface with smooth part after movement.

Figure 3.6: Changing extracted points by related position change between the object and a camera.

The object motion is correctly estimated with many outcrop points. A smaller space sampling grid serves more accurate motion estimation, because the minimum error of the motion estimation is determined with the space sampling grid. To estimate the object motion correctly, the space sampling grid should be set small under the condition that outcrop points can be extracted.

The outcrop points are extracted as shown in Figure 3.7. The object with an outstanding part shown in 3.7(a) is observed with cameras. The space sampling points included in the object are projected to the image planes of multiple cameras. With the condition that the outcrop points are extracted, the space sampling grids which are projected to the pixels on silhouette contours and the corresponding pixel have only one projected space sampling point are extracted as the outcrop points.

In Figure 3.7(b), we consider l_v and l_p for a space sampling point which is possible to be extracted as an outcrop point. l_v is the length along the direction of translation. l_v is calculated based on the distance for the neighbor space sampling points. When there are a few the neighbor sampling grids



(a) 3D sampling points included in an object are projected to a 2D plane.

(b) l_p and l_v determine how many outcrop points are extracted.

Figure 3.7: Projection of 3D sampling points to an 2D image plane.

included in the object, l_v becomes large. When the direction of the translation and the ray of the camera are at right angles to each other and the direction of the translation and the space sampling grid are same, l_v is d_{3D} as a maximum value. d_{3D} is the length of the space sampling grid. Let us denote l_v as $\alpha_{3D}d_{3D}$, which $0 < \alpha_{3D} \leq 1$.

l_p is also the length along the direction of the translation. l_p is calculated based on the distance for the neighbor image sampling points. Consider that $Z \gg 0$ denotes the distance between the center of the camera and the focused space sampling grid, focal length of the camera is f and the image sampling grid of the camera d_{2D}. l_p is $d_{2D} \cdot Z/f$ at a minimum when the direction of the translation and the ray of the camera are at right angles to each other. Let us denote l_v as $\alpha_{3D}d_{3D}$, which $0 < \alpha_{3D} \leq 1$.

l_p and l_v determine the rate that the focused space sampling point is extracted as an outcrop points as shown in Figure 3.7(b).

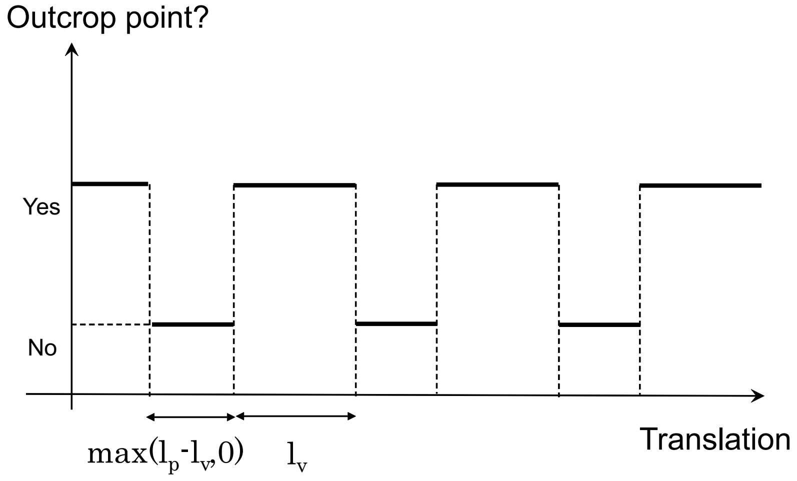


Figure 3.8: The rate that a voxel is extracted as an outcrop point.

The rate that the focused space sampling point is extracted as an outcrop point is $l_v/(l_v + max(l_p - l_v, 0))$. The value of $l_v/(l_v + max(l_p - l_v, 0))$ is 1 at a maximum in case that $\alpha_{2D}d_{2D} \cdot Z/f \leq \alpha_{3D}d_{3D}$ is satisfied. While, the minimum value is $+0$ in case that $d_{3D} = +0$. When d_{3D} is small, outcrop points are not extracted at all.

If the object is rotated, the value of l_v is changed. l_p is not changed with the rotation of the object. $l_v/(l_v + max(l_p - l_v, 0))$ is rotated with the rotation of the object. It means that the rate that the focused space sampling point is extracted as an outcrop point is changed. The rate is changed under the condition that $\alpha_{2D} \geq 1$ and $0 < \alpha_{3D} \leq 1$.

When projected space sampling grid matches the image sampling grid, $\alpha_{2D}d_{2D} \cdot Z/f \leq \alpha_{3D}d_{3D}$ is satisfied. For 3D shape reconstruction, such value is often adopted as the space sampling grid. Considering q discussed in Section 3.2, the space sampling grid which gives $q = \sqrt{3}$ is the best configuration for 3D shape reconstruction. In a special case that the direction of the translation and the ray of the camera are at right angles to each other and the direction of the translation and the space sampling grid are same, $\alpha_{2D}d_{2D} \cdot Z/f \leq \alpha_{3D}d_{3D}$ is satisfied with $d_{3D} = d_{2D} \cdot Z/f$.

To estimate the object motion correctly, the space sampling grid should be set small under the condition that some outcrop points can be extracted. Once the object motion is estimated, the space sampling grid can be changed for reconstructing shapes in multiple frames. The space sampling grid for shape reconstruction is not required to be the same with that for motion estimation. When we estimate the object motion, we should set the optimal space sampling grid. Note that subsampling motion might be estimated when the motion is estimated with many outcrop points. In discussions on super-resolution [39], it is said that the motion can be estimated in subsampling precision.

When d_{3D} is set to $d_{2D} \cdot Z/f$ or such value, some outcrop points that are required for motion estimation are adequately extracted. The outcrop points tend to be extracted from outstanding parts on the object surface. The optimal configuration is such that projected space sampling grid matches the image sampling grid as described above.

3.4 Object Motion Estimation

3.4.1 Transformation Matrix of Rigid Motion

We represent the rigid object motion between i-frame and k-frame by transform matrix D^{ik} which is a matrix of the homogeneous coordinates system. This matrix is represented with a 3×3 rotational matrix denoted by R^{ik} and a 1×3 translational vector denoted by $\mathbf{t}^{ik}$ as follows:

$$D^{ik} = \begin{pmatrix} R^{ik} & \mathbf{t}^{ik} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix}, \tag{3.2}$$

where $\mathbf{0}_{1\times3}$ is a zero vector with the size of 1×3.

R^{ik} is represented with quaternion. The quaternion representation gives R^{ik} linearity to the object motion. It is also useful for avoiding local minima in the process of optimization for estimating rotation. R^{ik} is represented with the quaternion $\mathbf{q} = [\lambda_0, \lambda_1, \lambda_2, \lambda_3]^T$ as follows:

$$R^{ik} = \begin{pmatrix} \lambda_0^2+\lambda_1^2-\lambda_2^2-\lambda_3^2 & 2(\lambda_1\lambda_2-\lambda_0\lambda_3) & 2(\lambda_1\lambda_3+\lambda_0\lambda_2) \\ 2(\lambda_1\lambda_2+\lambda_0\lambda_3) & \lambda_0^2-\lambda_1^2+\lambda_2^2-\lambda_3^2 & 2(\lambda_2\lambda_3-\lambda_0\lambda_1) \\ 2(\lambda_1\lambda_3-\lambda_0\lambda_2) & 2(\lambda_2\lambda_3+\lambda_0\lambda_1) & \lambda_0^2-\lambda_1^2-\lambda_2^2+\lambda_3^2 \end{pmatrix}$$

With estimated D^{ik}, visual hulls in multiple frames can be integrated. Each voxel of the observation region is projected to a silhouette. Symbols for explanation of the visual hull integration are defined as described in Figure 3.4.1.

The object O occupies the region O_i at i-frame. o_i is a voxel included in O_i. Let P_j be a projection matrix of camera C_j. P_j does not change in multiple frames. r_{ij} is a projection point of o_i in an image of C_j. S_{ij} is a set of r_{ij}.

$(D^{ik})^{-1}o_k$ is projected in S_{ij} with P_j:

$$o_k \in O_k, \quad r_i = P_j((D^{ik})^{-1} \cdot o_k) \quad \in S_{ij}.$$

Similarly, $(D^{ki})^{-1}o_i$ is projected in S_{kj} with P_j:

$$\begin{aligned} o_i \in O_i, \quad r_{ij} &= P_j((D^{ki})^{-1} \cdot o_i) \\ &= P_j(D^{ik} \cdot o_i) \quad \in S_{kj}. \end{aligned}$$

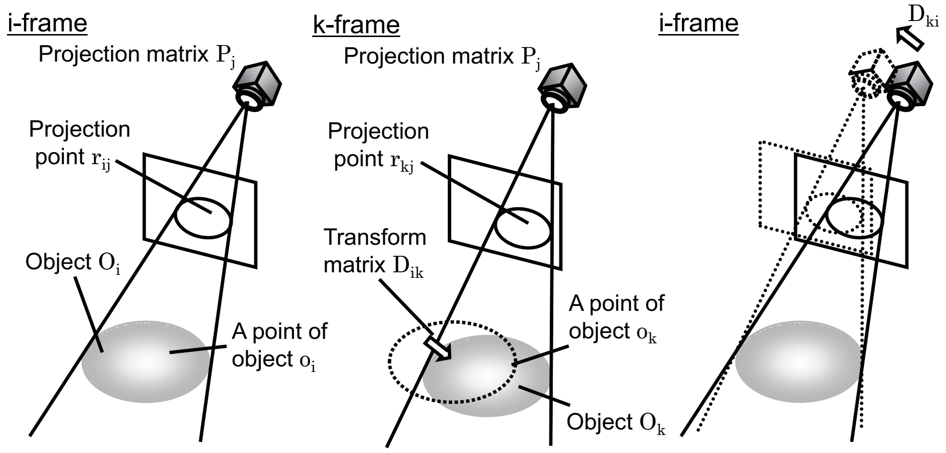


Figure 3.9: Integration visual hulls of multiple frames.

These equations means that O_i is calculated by O_k, D^{ik} and S_{kj} at k-frame. O_i given by the information at k-frame represents the object region calculated by set of cameras in other positions. The virtual position of cameras shown in Figure 3.4.1 are calculated with D^{ki}. The integration of visual hulls at i-frame and j-frame is realized to calculate an intersection of O_i and O_k with D^{ik}.

3.4.2 Robust Estimation Method

Let us denote the position of the s-th outcrop point at i-frame by p^i_s. Consider that the s-th outcrop point at i-frame and the u-th outcrop point at k-frame corresponds to each other. The points are actually the same points in the rigid motion represented by D^{ik}. p^k_u should be equal to $D^{ik}p^i_s$.

Based on this relationship, we estimate D^{ik} from p^i_s and p^k_u. D^{ik} can be estimated by minimizing the difference between p^k_u and $D^{ik}p^i_s$. The corresponding pairs of outcrop points p^i_s and p^k_u are detected by calculating the difference. The corresponding pairs make the difference minimized. For minimizing the difference, we apply Powell's method [38]. The method can realize rapid minimization of functions with many parameters while avoiding local minima.

The problem is that the corresponding pairs of outcrop points might not be extracted in different frames. The outcrop point extracted from smooth parts on the object's surface tends not to be extracted in multiple frames. The outcrop point which is incidentally extracted from the smooth part at a frame might not be extracted at any other frames.

In those case, all the outcrop points at i-frame might not have their corresponding points in the outcrop points at k-frame, and vice versa. We introduce the idea of robust estimation to estimate D^{ik}. We can estimate the motion correctly even if some outcrop points do not have their corresponding points.

For the motion estimation in this paper, we do not need to consider large amount of motions between i-frame and k-frame. The purpose of the motion estimation is to obtain relative positions between cameras and the object. With the estimated motion, we can improve accuracy of the reconstructed shape instead of increasing cameras. For this purpose, images from slightly different viewpoints should be obtained. The images are virtually considered to the images obtained from close-set cameras. Following the discussion above, we define the error function E to be minimized for estimating D^{ik} as the follows:

$$E = \sum_u f\left(\min_s (p_u^k - D^{ik} p_s^i)^2\right) \tag{3.3}$$

$$f(x) = \begin{cases} x & x \leq M_{th} \\ M_{th} & x > M_{th} \end{cases}$$

where M_{th} is a threshold which means the upper limit value allowed as the object motion. Due to function f, the pair p_s^i and p_u^k is disregarded when the motion estimated from the pair is sufficiently different from other pairs. Since we consider that the object motion is small as discussed above, the motion of the object between adjacent frames can be assumed to be small. When the motion is small, the number of the corresponding outcrop points is sufficiently large. We estimate the motion of only frames between adjacent frames. The motion between any pairs of frames D^{ik} is estimated as follows:

$$D^{ik} = D^{k-1,k} D^{k-2,k-1} \cdots D^{i+1,i+2} D^{t,t+1}. \tag{3.4}$$

3.5 Experimental Results

In our proposed method, when the object has sufficient number of outstanding points, it is expected that the accuracy of the reconstructed shapes is improved. When the object has a smooth surface, outstanding points fail to be extracted. In order to confirm the validation of our method for objects in various kinds of shapes, we applied our method to spherical objects with a simulation shape of triceratops, different numbers of outstanding points, commonplace objects with smooth surface and real object with sufficient outstanding points. The experimental results are given below.

3.5.1 Simulation Object

At first, we applied to simulation data of a triceratops to evaluate the validation of our proposed method. The simulation object is translated by 1 voxel between adjacent frames, and rotated by 2 degrees between adjacent frames. 20 cameras located on the vertexes of a regular dodecahedron observe the object. From the cameras, silhouettes are produced. In the settings, about the same set of silhouettes is obtained by the cameras every 18 frames. All the silhouette possible to be observed by the cameras during the motion of the object can be obtained with $18(=72 \div 2 \div 2)$ frames. In experimental results, shapes are reconstructed with all the silhouettes of 18 frames.

Motion of the objects is estimated as described in table and table . The motion is estimated by tracking outcrop points. In table 3.2, each rotation parameter is described, and each translation parameter is described in table 3.3. The values of q and N_v, which are described in Section 3.2, are 1.679 pixels and 438310 voxels.

The estimation error of the rotation parameter is 0.30 degrees on an average, and 0.93 degrees at a maximum. The estimation error of the translation parameter is 0.38 voxels on an average, and 1.24 voxels at a maximum. The outcrop voxels extracted on the smooth surface make the errors. Focusing on the translation error, almost all the errors are less than a voxel, which is within the sampling error. The rotation error cannot be discussed to compare with the sampling error. From the error voxels, we discuss whether the error is permissible or not.

With Figure 3.5.1, we discuss the difference of the accuracy between the visual hull reconstructed at 1 frame and the integrated shape. The integrated shape is virtually reconstructed 360 silhouettes in multiple frames. The cor-

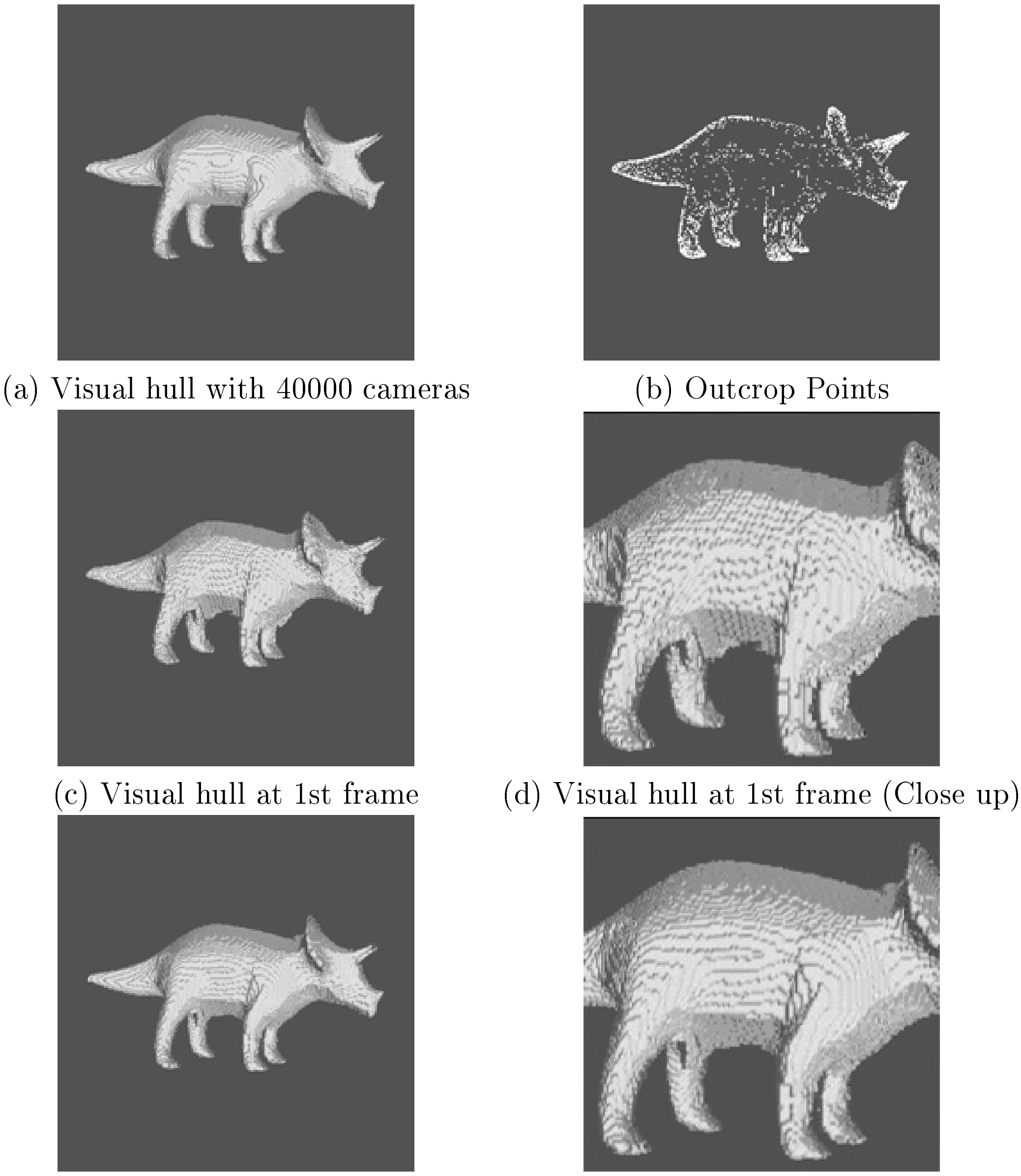
Table 3.2: Rotational parameters.

	x (degrees)		y (degrees)		z (degrees)	
frame	Truth	Diff	Truth	Diff	Truth	Diff
1	0.00	+0.02	0.00	-0.01	2.00	-0.01
2	0.00	+0.02	0.00	-0.02	4.00	-0.06
3	0.00	+0.03	0.00	-0.03	6.00	-0.11
4	0.00	-0.09	0.00	0.01	8.00	-0.14
5	0.00	-0.14	0.00	0.03	10.00	-0.19
6	0.00	-0.16	0.00	0.02	12.00	-0.24
7	0.00	-0.30	0.00	0.04	14.00	-0.24
8	0.00	-0.30	0.00	0.11	16.00	-0.27
9	0.00	-0.36	0.00	0.14	18.00	-0.32
10	0.00	-0.31	0.00	0.15	20.00	-0.39
11	0.00	-0.37	0.00	0.19	22.00	-0.46
12	0.00	-0.43	0.00	0.20	24.00	-0.52
13	0.00	-0.51	0.00	0.23	26.00	-0.56
14	0.00	-0.52	0.00	0.26	28.00	-0.64
15	0.00	-0.66	0.00	0.27	30.00	-0.74
16	0.00	-0.72	0.00	0.25	32.00	-0.79
17	0.00	-0.76	0.00	**0.28**	34.00	-0.83
18	0.00	**-0.77**	0.00	**0.28**	36.00	**-0.93**

rect shape from 40000 silhouettes (in Figure 3.5.1(a)), the visual hull at 1 frame (in Figure 3.5.1(c)), the outcrop points (in Figure 3.5.1(c) and Figure 3.5.1(d)) and the integrated shape (in Figure 3.5.1(e) and Figure 3.5.1(f)) are drawn.

The outcrop points (in Figure 3.5.1(b)) are extracted on the outstanding parts on the object surface as described in Section 3.3. By integrating visual hulls, The additional regions on the abdomen of the triceratops are decreased as shown in Figure 3.5.1(e) and Figure 3.5.1(f). Compared with the original visual hulls (in Figure 3.5.1(c) and 3.5.1(d)), the abdomen has a smooth surface.

We also examined the integrated shapes from a numeric aspect. We discuss on the difference of the number of voxels between the correct shape and the integrated shape.



(a) Visual hull with 40000 cameras

(b) Outcrop Points

(c) Visual hull at 1st frame

(d) Visual hull at 1st frame (Close up)

(e) Integrated visual hull

(f) Integrated visual hull (Close up)

Figure 3.10: Result of integration of 18 visual hulls.

Table 3.3: Translational parameters.

	x (voxels)		y (voxels)		z (voxels)	
frame	Truth	Diff	Truth	Diff	Truth	Diff
1	1.00	-0.04	1.00	-0.02	1.00	-0.06
2	2.00	+0.08	2.00	+0.03	2.00	-0.10
3	3.00	+0.20	3.00	+0.13	3.00	-0.04
4	4.00	+0.35	4.00	+0.14	4.00	-0.04
5	5.00	+0.51	5.00	+0.13	5.00	-0.04
6	6.00	+0.56	6.00	+0.22	6.00	-0.08
7	7.00	+0.60	7.00	+0.24	7.00	-0.09
8	8.00	+0.62	8.00	+0.39	8.00	-0.05
9	9.00	+0.58	9.00	+0.45	9.00	-0.08
10	10.00	+0.60	10.00	+0.52	10.00	-0.09
11	11.00	+0.49	11.00	+0.50	11.00	-0.09
12	12.00	+0.56	12.00	+0.56	12.00	0.00
13	13.00	+0.62	13.00	+0.60	13.00	-0.01
14	14.00	+0.79	14.00	+0.79	14.00	-0.04
15	15.00	+0.88	15.00	+0.86	15.00	+0.06
16	16.00	+1.01	16.00	+0.90	16.00	+0.13
17	17.00	+1.02	17.00	+1.03	17.00	+0.19
18	18.00	**+1.23**	18.00	**+1.24**	18.00	**+0.28**

The reconstructed shape with the volume intersection method is not the object shapes, but the concavities of the object shape, even using infinite number of cameras. We employ the 3D shape reconstructed for the object using 40000 cameras by simulation as the theoretical limitation of the volume intersection method in order to evaluate the error of each experimental result. We call the shape *correct shape* of the object.

Each experimental result is compared with the correct shape based on the three types of voxels: missing voxels, additional voxels and error voxels. The missing voxels are those included in the correct shape and not included in the reconstructed shapes, whereas the additional voxels are those included in the reconstructed shapes and not in the correct shape. The error voxels are summation of the missing voxels and the additional voxels.

In Figure 3.5.1, the transition of the three types of voxel is drawn. The longitudinal axis is scaling based on α.

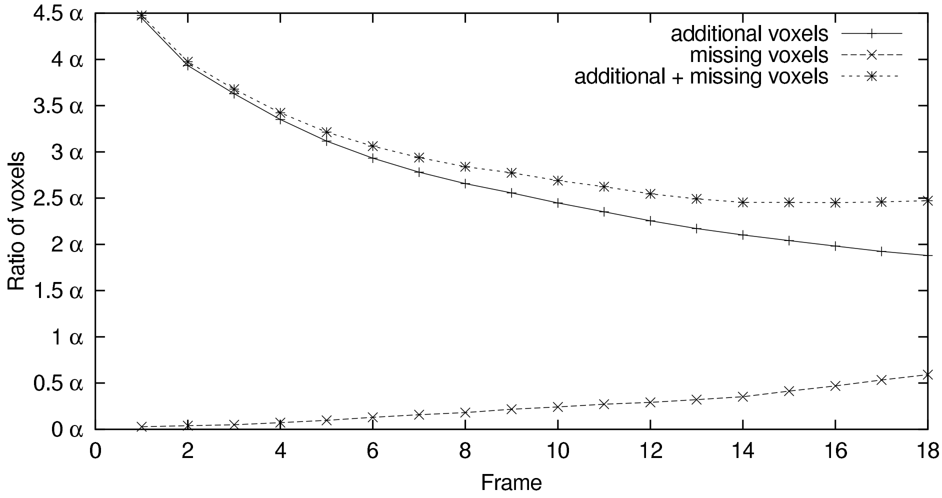


Figure 3.11: Relation between frames and voxels. (q=1.679)

The correct shape is represented with 438310 voxels. q of the shape is 1.679. It means α is 2.68 for the setting of $k = \sqrt{2}/2$. The objective value of the error voxels is 11747 voxels.

By integrating frames, the error voxels are monotonically decreasing. The number of the additional voxels of the visual hull at 1 frame is 4.62α(54358 voxels, 12.40%). Whereas the number of the integrated shape is 1.95α(22975 voxels, 5.24%). The number of the integrated shape is close to the accuracy limit by the sampling error. The missing voxels are caused by the error of the motion estimation. The phenomenon is a specific problem of the shape integration in multiple frames. However, the missing voxels of the integrated shape for 18 frames (0.62α, 7237 voxels, 1.65%)) are less than α. The advantage of visual hull integration is greater than the disadvantage.

To examine the relationship between q and the number of the error voxels, N_v, α and the error voxels for various q are described in table 3.4. In Figure 3.5.1, the transition of the number of the error voxels for various values of q are drawn.

For all values of q, although the integrated shapes have different numbers of the error voxels with the difference of q, all the integrated shapes become more accurate. To conclude the subsection, our proposed outcrop points are

Table 3.4: Relation between frames and voxels with various q.

q	N_v	α	Add'l($\times\alpha$)	Miss($\times\alpha$)	Error($\times\alpha$)
1.926	363118	2.49%	1.83	0.94	2.77
1.679	438310	2.68%	1.95	0.62	2.57
1.564	541697	2.68%	1.60	0.32	1.92
1.398	677086	2.78%	1.54	0.60	2.14
1.291	860893	2.78%	1.55	0.37	1.92

valid to integrate visual hulls in multiple frames. The outcrop points enable to estimate the object motion for various values of q. By integrating them, the reconstructed shapes become more accurate.

3.5.2 Spherical Objects

Our method requires several outstanding points with sufficient amount of length for extraction and tracking of outstanding points. In order to examine minimum number and length required for the outstanding points, we applied our method to objects in spherical shapes with different number of outstanding points with various amount of length, by adding fluctuation to the surface using a sinusoidal function. The position of each point on the surface of the object (x, y, z) is simulated with parameters θ, ϕ as follows:

$$\begin{cases} x' = cos\theta cos\phi \\ y' = sin\theta cos\phi \\ z' = sin\phi \end{cases}$$

$$(0 \leq \theta \leq 2\pi, -\pi/2 \leq \phi \leq \pi/2)$$

$$r' = A_l cos F_l \pi x' \cdot cos F_l \pi y' \cdot cos F_l \pi z'$$

$$\begin{cases} x = (r + r')cos\theta cos\phi \\ y = (r + r')sin\theta cos\phi \\ z = (r + r')sin\phi \end{cases}$$

where A_l denotes the amplitude of the fluctuation, and F_l denotes its frequency. The radius of the sphere r is set to 50. The objects and their outcrop points are drawn in Figure 3.13. Since shapes of the outstanding

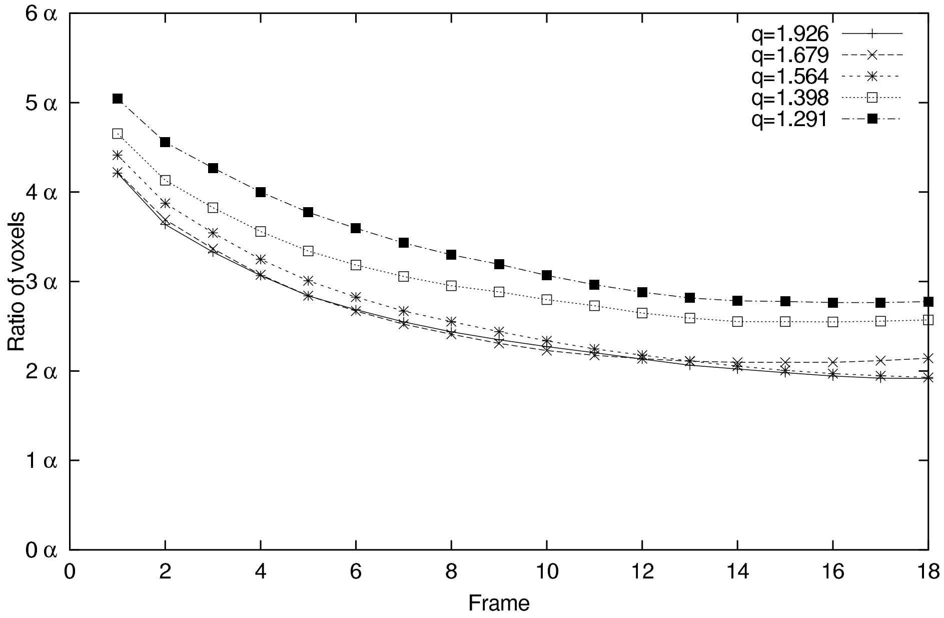


Figure 3.12: Relation between frames and voxels with different 'q's.

points created by the fluctuation is controlled by A_ls and F_ls, the 3D shape reconstructed with our method is expected to be more improved with larger A_l and larger F_l. The results for $F_l = 4, 6$ and $A_l = 2, 4$ are shown below.

In these experiments, objects are translated by 1 voxel along X,Y,Z-axes and rotated by 2 degrees around Z-axis for each frame. 20 cameras are set on the vertexes of a dodecahedron surrounding the object. In the settings, about the same set of silhouettes is obtained by the cameras every 18 frames. All the silhouette possible to be observed by the cameras during the motion of the object can be obtained with 18(=72÷2÷2) frames. Thus we employ 18 sequential frames of all the cameras for shape reconstruction by our method as shown in Figure 3.14.

Figure 3.14(a), (b), (c) and (d) illustrate the transition of three types of voxels for each objects. In the experimental results, except for the object with $F_l = 4, A_l = 2$, the error voxels monotonically decrease with the increase number of the images up to 18 frames. The error voxels for the object with $F_l = 4, A_l = 2$ are increasing after the number of frames exceeds nine. The surface of the object does not have sufficient number of outstanding parts.

When the outstanding parts on the surface do not have enough length, the parts should not be extracted as the outcrop points. To conclude the result, F_l should be larger than 6 or A_l should be larger than 4 in order to improve accuracy of the 3D shape by our method. Since r is set to be 50, $A_l = 4$ corresponds to 8% of the length of the whole object, and 6.24 pixels in the silhouettes. In the shape with $F_l = 6$, outstanding points exist every 30 degrees on the surface. These are the conditions required for the shapes of the target objects for improving accuracy with our method.

3.5.3 Common Objects with Smooth Surfaces

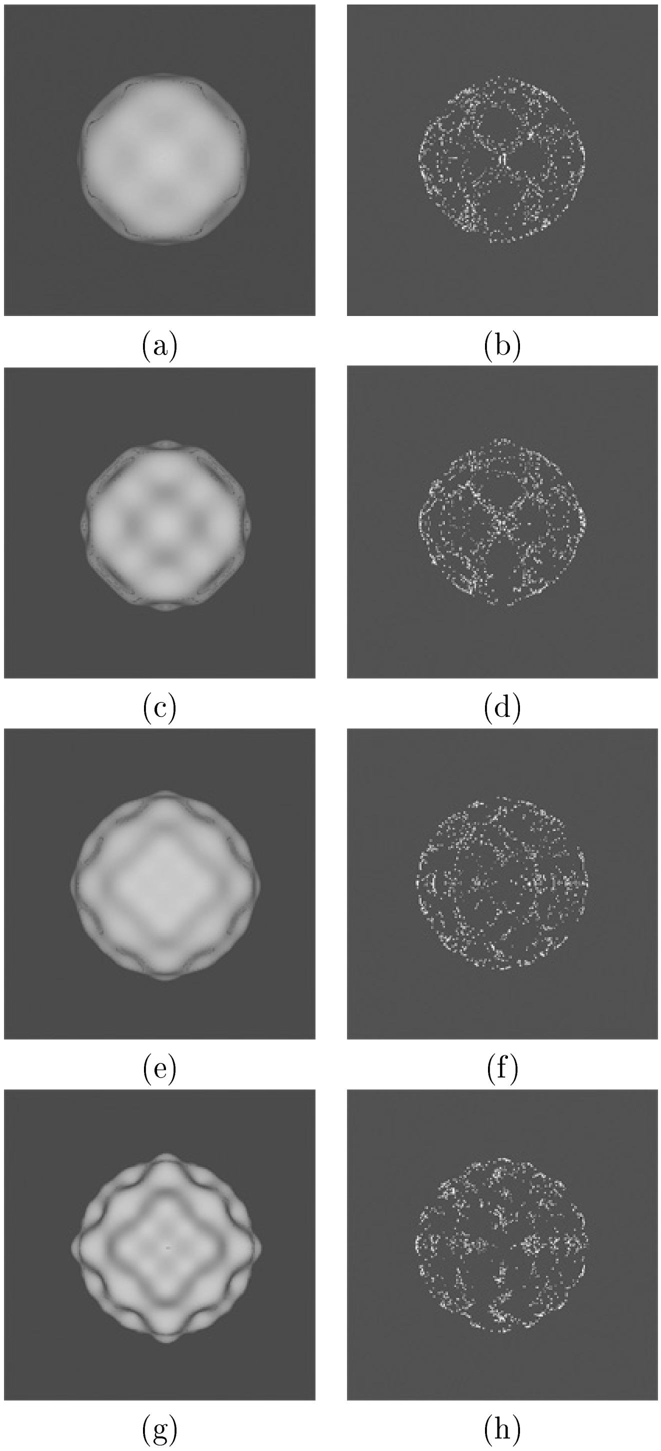
In order to verify the applicability of method to the objects in the real world, we applied our method to the common objects that seem to have difficult, for extraction and tracking, the outstanding points due to the smoothness of the surface. We simulated the shape reconstruction with our method for objects in shapes of a banana, a teapot and a queen of chess. Their results are evaluated by the error voxels similar to the experiment in Subsection 3.5.1.

The outcrop voxels for a banana are shown by dots in Figure 3.15(d). As shown by the graph in Figure 3.16(a), the error voxels increase after the number of frames used for shape reconstruction exceeds seven. It means that the outstanding voxels of the banana cannot be effective for improving accuracy of the reconstructed shape for this object. It is caused by small number and limited length of outstanding points.

Whereas, as shown in Figure 3.16(b) and 3.15(c), the number of the error voxels for the other objects decreases until 18th frames. These results mean that our method based on the outstanding voxels can be effective for the objects in more various shapes than expected.

3.5.4 Real Object

We also applied our method to a real object. We employed a toy triceratops secured by a thread and we took their images with 19 cameras surrounding the object. Small motion was simulated by swinging the toy triceratops slightly. The silhouette from each camera is extracted based on the difference between an observed image and its background image in YUV color space. As illustrated by the dots in Figure 3.17(b), the outcrop voxels are extracted from the real object. However, the shape reconstructed using all



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3.13: Integration for images of 18 frames : (a) An object shape of F_l=4, A_l=2. (c) F_l=4, A_l=4. (e) F_l=6, A_l=2. (g) F_l=6, A_l=4. (b), (d), (f) and (h) are outcrop points for each shape.

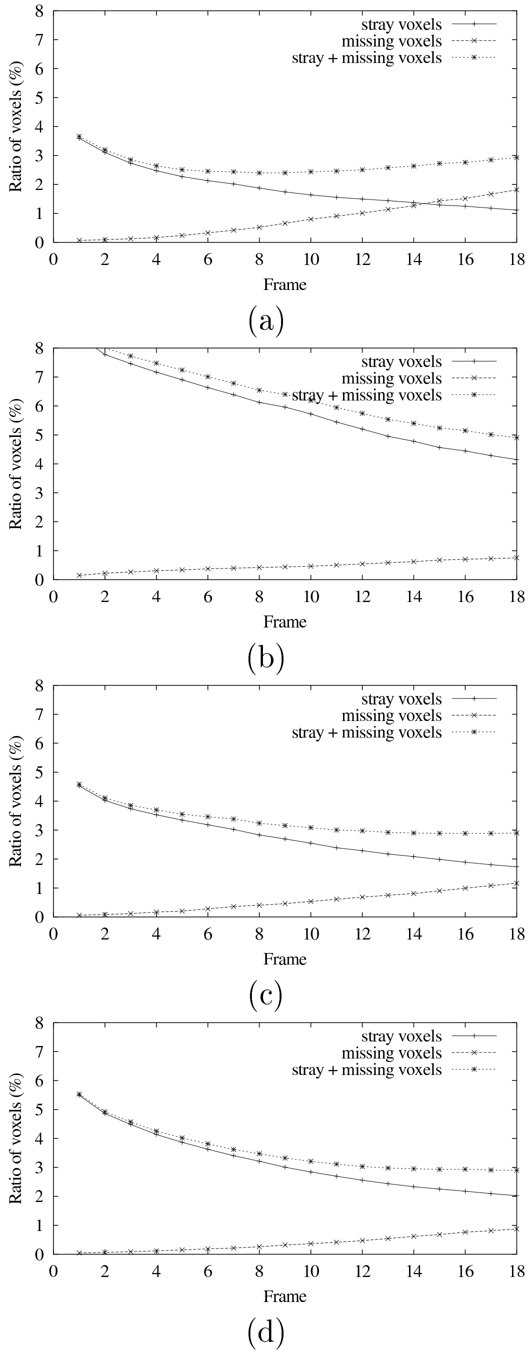
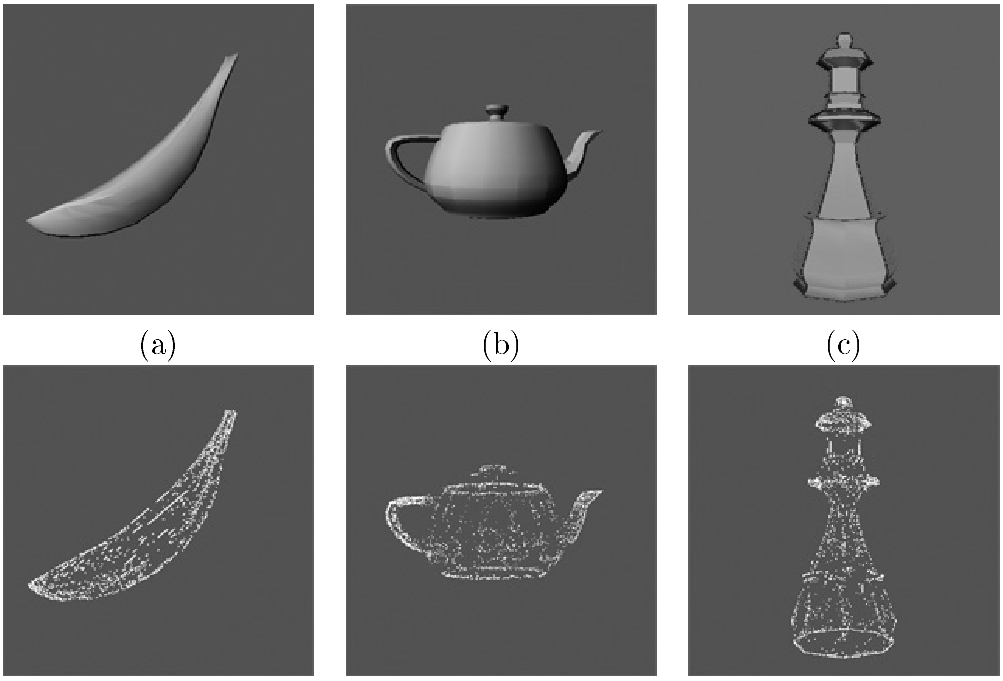


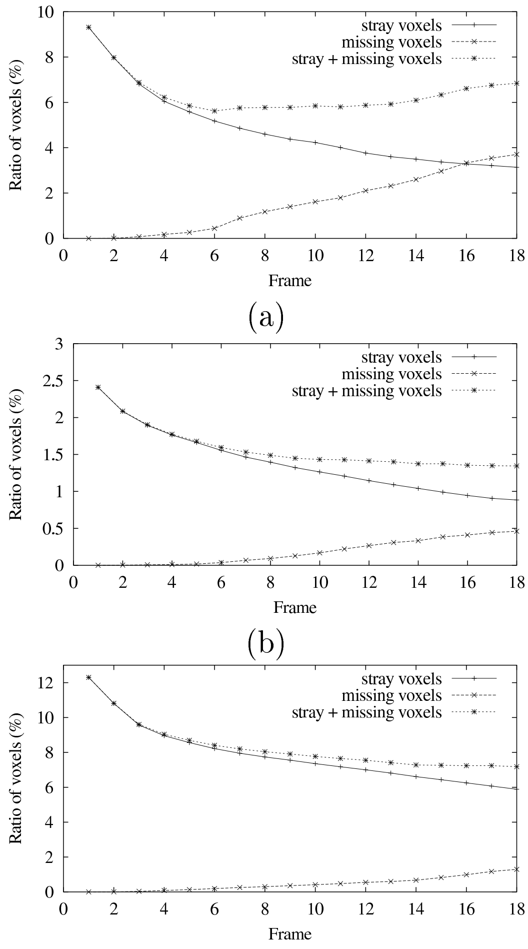
Figure 3.14: The ratios of error voxels of shapes from 18 frames : (a) F_l=4, A_l=2. (b) F_l=4, A_l=4. (c) F_l=6, A_l=2. (d) F_l=6, A_l=4. In (a), summations of additional voxels and missing voxels are increasing after 9th frame, even if the frontier points (FPs) are used. In (b), (c) and (d), the summations are decreasing until 18th frame with both kinds of feature points.



(a) (b) (c)

(d) (e) (f)

Figure 3.15: Result of integration for images of 18 frames : (a), (b) and (c) are original shapes of a banana, a teapot and a queen of chess respectively. (d), (e) and (f) are emerged voxels for each shape. (g), (h) and (i) show error voxels of integrated visual hulls. In (h) and (i), summations of stray voxels and missing voxels are decreasing until 18th frame.



(a)

(b)

(c)

Figure 3.16: Transition of the error voxels : (a), (b) and (c) show error voxels of integrated visual hulls. In (b) and (c), summations of stray voxels and missing voxels are decreasing until 18th frame.

the images of 5 frames has missing parts. It is caused by missing of the silhouettes extracted by background subtraction. When colors of the object in the observed images are similar to those of the background, the silhouette is missing. Since the 3D shape is reconstructed as the intersections of all visual cones, the reconstructed shape has missing parts when one of the silhouettes has missing regions. This problem often happens for the volume intersection method using a large number of cameras as well as this situation.

In order to cope with this problem, it is proposed in the previous work to ease the condition for extracting the visual hull from visual cones [46] ; instead of calculating pure intersection of visual cones for all the cameras, voxels included in the visual cones for at least $n - N_{allow}$ cameras are allowed to be included in the visual hull. Referring to this solution, N_{allow} was set as 1 (Figure 3.17(e)) and 2 (Figure 3.17(f)) in our experiment. In the case of $N_{allow} = 1$, more accurate shape is reconstructed with all the images of all the frames, compared with the shape reconstructed from images of a single frame. But even if we make N_{allow} larger than 1 for 19 cameras, the result is not improved any more than the shape of $N_{allow} = 1$.

Then, how can we set the value of N_{allow}? Although several researchers have referred to the configuration problem for N_{allow} in previous works [2], deeper discussion is required for an adequate configuration of N_{allow}.

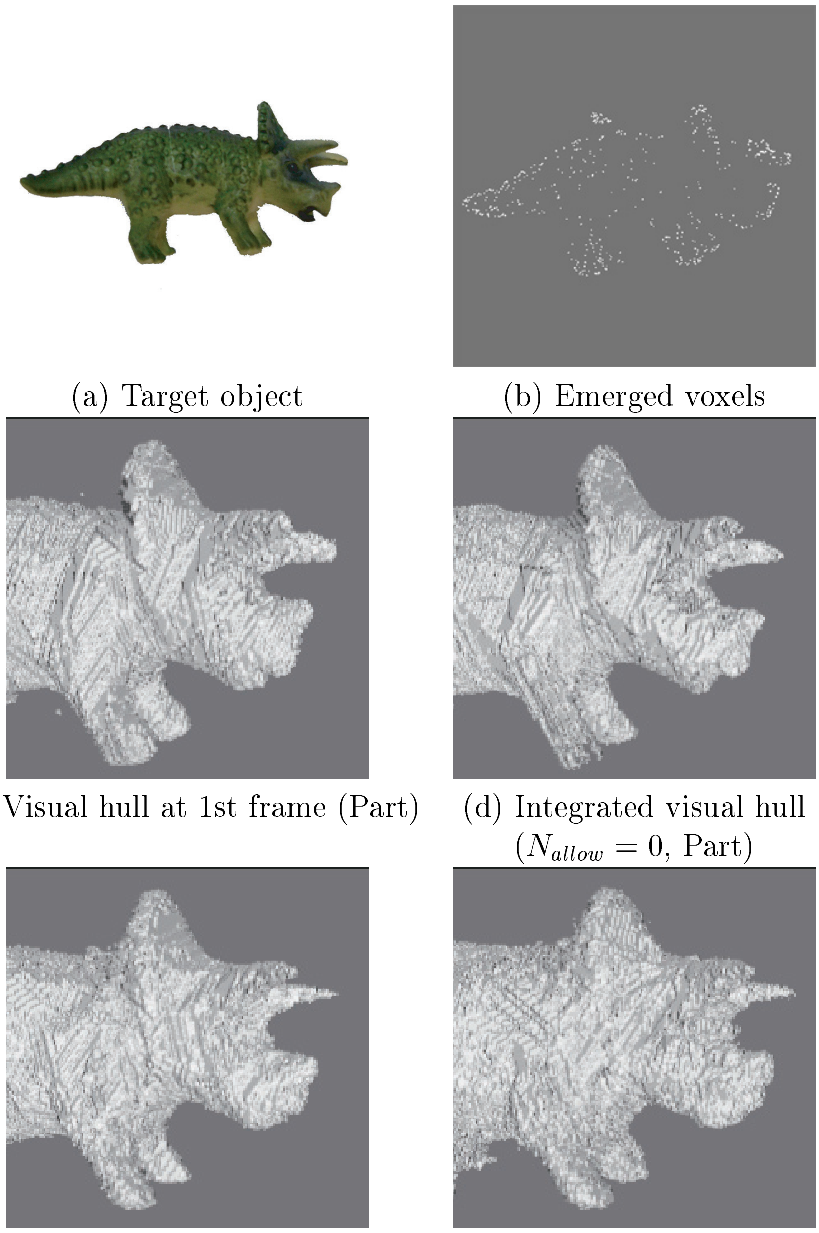
Let us define that silhouette missing rate is p and the number of cameras is N. False Rejection (FR) means an outside voxel is misclassified as inside. False Acceptance (FA) means an inside voxel is misclassified as outside. For an voxel, corresponding pixels in the k-silhouettes are missed at the rate of ${}_NC_k p^k (1-p)^{N-k}$. In case N_{allow} is set to N_{th}, the focused voxels is included in the visual hull if $k < N_{th}$ is satisfied. The rate $P(FR)$ that an inside voxel is *not* included in the visual hull is given by :

$$P(FR) \quad = \quad 1 - \sum_{k=0}^{N_{th}} {}_NC_k p^k (1-p)^{N-k}. \tag{3.5}$$

The change of $P(FR)$ for various values of N and p is shown in Table 3.18.

How much missing parts are allowed is determined by users with the tables. Generally, high silhouette missing rates and big numbers of cameras require that N_{th} is set to a larger value.

However FA is also discussed in [2], only outside voxels which are not included in all silhouettes are focused. How many silhouettes classify the



(a) Target object

(b) Emerged voxels

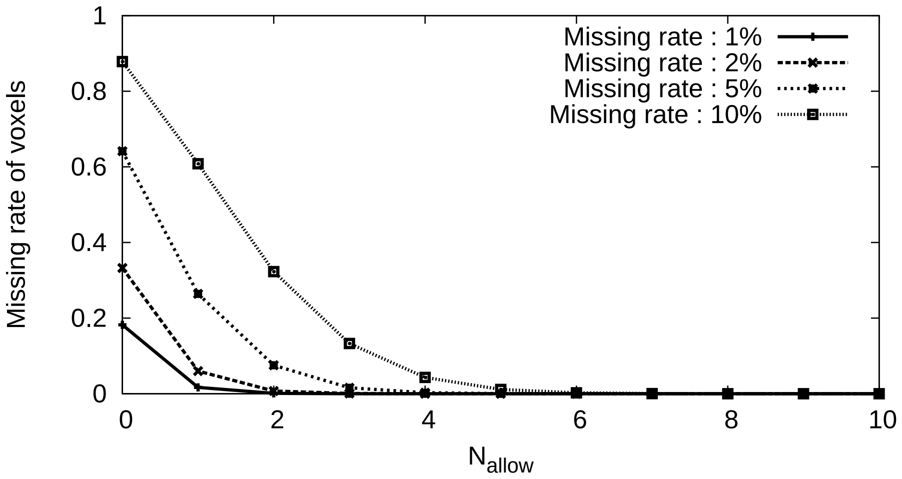
(c) Visual hull at 1st frame (Part)

(d) Integrated visual hull ($N_{allow} = 0$, Part)

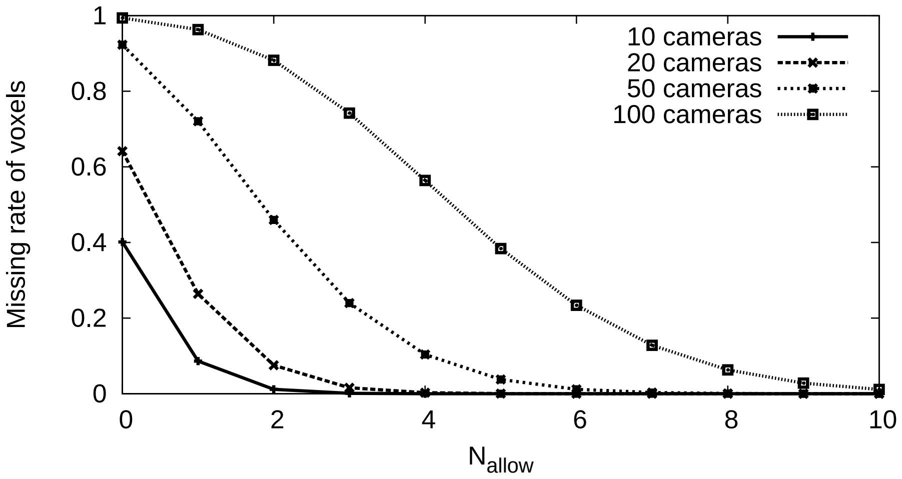
(e) Integrated visual hull ($N_{allow} = 1$, Part)

(f) Integrated visual hull ($N_{allow} = 2$, Part)

Figure 3.17: Result of integration of images of 5 sequencial frames.



(a) Relationship between values of N_{allow} and voxel missing rate in case that the number of cameras is 20.



(b) Relationship between values of N_{allow} and voxel missing rate in case that silhouette missing rate is 5%.

Figure 3.18: Relationship between values of N_{allow} and voxel missing rate in simulation.

focused voxel as inside is important element to calculate the value of $P(FA)$. The number how many silhouettes classify the focused voxel as inside depends of the complexity of the object and the arrangement of the cameras. It differs according to voxels. $P(FA)$ is difficult to be calculated. How much additional parts are allowed is determined by users from the reconstructed shapes with various values of N_{th}.

Several values of N_{th} is set for reconstructing shapes of Figure 3.17. The number of the missing and additional voxels are shown in Table 3.5. The correct visual hull in Table 3.5 is reconstructed from the manually extracted silhouettes.

Table 3.5: Relationship between values of N_{allow} and voxel error rate in real environment.

N_{allow}	Visual hull	Additional Voxels	Missing Voxels
(Correct VH)	277496	0	0
0	265472	10502(3.78%)	22526(8.12%)
1	309424	34652(12.49%)	2724(0.98%)
2	343754	66789(24.07%)	531(0.19%)

Even when N_{allow} is set to 1, 12.49% additional voxels of the correct shape are included in the reconstructed shape, whereas the missing rate is less than 1%. When N_{allow} is set to 2 or more, much more additional voxels is included in the reconstructed shape, whereas the missing rate does not decrease so much. From the Table 3.5, the best configuration of the N_{allow} is 1 for the data.

In order to obtain better results, we need to improve the image processing for extracting the silhouettes. One solution for the problem is the silhouette refinement with the random pattern backgrounds as described in Chapter 2. Another solution is an improved shape integration method in multiple frames as described in Chapter 4. In the method, only silhouettes that give good result are used to reconstruct shapes, since we can obtain many silhouettes in multiple frames. The method enables to reconstruct accurate shapes from the silhouettes with missing parts.

3.6 Conclusions

In this Chapter, we proposed a method for improving accuracy of the 3D shapes reconstruction with the volume intersection method by using the rigid motion of the target object. In order to estimate the motion of the object from the visual hull composed at each moment, we proposed to use feature points called outcrop points. In the experiments both with simulated objects and a real object, it is confirmed that the outcrop points are effective for estimating the motion of the objects when the object has sufficient outstanding points on its surface. To conclude the Chapter, more accurate shapes can be reconstructed, compared with those from images of a single frame. Furthermore, our method is applicable even for the objects that seem to have smooth surface.

Similar to the conventional volume intersection method, the reconstructed shape is sensitive to the error of image processing for extracting the silhouettes. One solution for the problem is the silhouette refinement with the random pattern backgrounds as described in Chapter 2. Another solution is an improved shape integration method in multiple frames as described in Chapter 4. Together with the refined silhouette extraction and the refined shape integration, the visual hull integration method with the outcrop point extraction should improve the integrated shapes much more.

Chapter 4

Frame Evaluation for Silhouette Integration

4.1 Introduction

Shapes of objects are reconstructed from silhouette with the volume intersection method [23, 30], The silhouettes are extracted from images obtained by multiple cameras. Each silhouette defines the region in which the object is possible to exist. The object region is calculated as the intersection region of the regions from all silhouettes. In the volume intersection method, the calculated region is called *visual hull*.

The visual hull is calculated more accurately when more cameras capture the object. However, it is not realistic to set the infinite number of cameras in real environments. To make images captured from a large number of cameras, we proposed outcrop point extraction method for visual hulls and estimated object motion, as described in Chapter 3. The silhouette integration in multiple frames provides us accurate reconstructed shapes. Outstanding points on the object surface tend to be extracted as the outcrop points.

A problem of the volume intersection method in multiple frames is to reconstruct shapes with large missing parts, when object motion is estimated with a large error. The reconstructed shapes in multiple frames are calculated as the intersection of visual hulls of all frames. Only a set of frame with the large error causes large missing parts in the reconstructed shape in multiple frames. The large error in object motion estimation is caused by failure of the outcrop point extraction. Missing parts and additional parts of the silhouettes adversely affect the outcrop point extraction. The outcrops are extracted under the assumption that the extracted silhouettes are complete. In case the extracted silhouettes include the missing and additional parts, corresponding outcrop points between different frames become difficult to be extracted.

In this Chapter, we proposed a method to suppress the missing parts in the reconstructed shapes from the silhouettes in multiple frames.

The error in object motion is the reason of the missing parts in the reconstructed shape. However, the motion cannot be estimated completely from the outcrop points which may not have the corresponding points. Since the answer motion is not given, the error in motion cannot be estimated. When the outcrop points may not have the corresponding points, the residual error of the motion estimation cannot represent the error in the object motion. Even when the outcrop points are translated with the answer motion, the residual error of the motion estimation does not become 0. It is caused by the outcrop points with no corresponding points.

We focused on the fact that outstanding points on the object surface tend to be extracted as the outcrop points. The outstanding points characterize the object shape. Using this fact, we could select frames that retain the outstanding points and process those into the reconstructed shape in multiple frames. Even in previous works, the fact is focused for surface smoothing [9, 11, 24, 1] . On the surface on the reconstructed shape, a mesh covering the shape consists of frontier points, which is the origin of the CSPs. The positions of the frontier points are fixed when the surface is applied to the smoothing process. The reconstructed shape retain outstanding points, which characterize the object shape. We define a function for measuring how the outstanding points are kept in the reconstructed shape by integrating a visual hull of a frame. By integrating visual hulls of only frames with high score, the reconstructed shape in multiple frames is guaranteed to include the outstanding points.

In Section 4.2, we will propose the function for measuring how the outstanding points are kept in the reconstructed shape by integrating a visual hull of a frame. The procedure by which the shape is reconstructed using the function in multiple frames is explained. Section 4.3 discusses how the function is affected by missing and additional parts of extracted silhouettes. The experimental results are presented in Section 4.4. The validity of the proposed function is verified on the basis of the experimental results. At the end of the Chapter, Section 4.5 concludes this Chapter.

4.2 Evaluation Function Based on Preserving Outcrop Points

When the rigid object motion, D_{ik}, between the i-frame and k-frame is correctly estimated, the outcrop points, OP_i, which are extracted in the i-frame, are included in V_k by translating D_{ki}. Similarly, the outcrop points, OP_k, which are extracted in the k-frame, are included in V_i by translating D_{ik}.

$$p_i \in OP_i, D_{ik}p_i \in V_k, \tag{4.1}$$

$$p_k \in OP_k, D_{ki}p_k \in V_i. \tag{4.2}$$

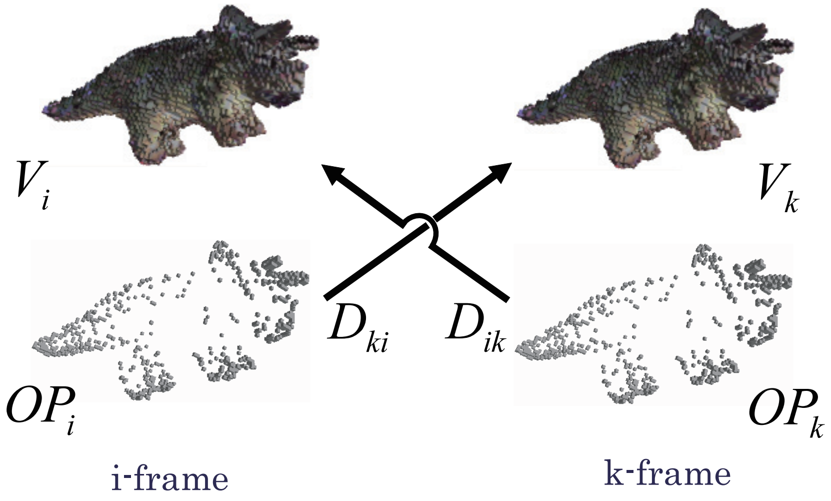


Figure 4.1: Frame evaluation from visual hulls and outcrop points.

If Eqs. (4.1) and (4.2) are completely satisfied, all outstanding parts of the object shape will be included in the reconstructed shape in multiple frames. They are not completely satisfied in real environments, because there are missing parts of visual hulls or errors in the estimated motion. To evaluate how many outstanding parts are included in the reconstructed shape, we can utilize the rate of outcrop points that satisfies Eqs. (4.1) and (4.2). The rate, $E_m(i,k)$, is defined by

$$E_m(i,k) = \frac{n_{ik} + n_{ki}}{2}, \tag{4.3}$$

where

$$
\begin{aligned}
n_{ik} &= \frac{n\{p_i | p_i \in OP_i, D_{ik}p_i \in V_k\}}{n\{p_i | p_i \in OP_i\}}, \\
n_{ki} &= \frac{n\{p_k | p_k \in OP_k, D_{ki}p_k \in V_i\}}{n\{p_k | p_k \in OP_k\}}.
\end{aligned}
$$

Here, $n\{\cdot\}$ is the number of voxels included in a set. OP_i and OP_k are sets of outcrop points in the i-frame and k-frame. V_i and V_k are sets of voxels included in the visual hulls of the i-frame and k-frame.

The evaluation function, $E_m(i,k)$, ranges from 0 to 1. When $E_m(i,k)$ indicates a large value, the integrated shape preserves many outcrop points within it. Conversely, a small value for $E_m(i,k)$ means that the integrated shape has lost the outcrop points. By only using frames where $E_m(i,k)$ has large values, the integrated shape can preserve outstanding parts. If threshold E_m^{th} is given, appropriate frames can be selected by $E_m(i,k) < E_m^{th}$. When the 0-th frame is chosen as the base frame, the frames that satisfy $E_m(0,i) < E_m^{th}$ are selected. Relabeling the frames as $i'(i' = 1, \cdots, M')$, the integrated shape is calculated as an intersection of the visual hulls of $V_{i'}$.

4.3 Outcrop Point Extraction from Incomplete Silhouettes

The missing and additional parts of the silhouettes might lead to the missing of outcrop points. The evaluation function $E_m(i,k)$ is designed on the assumption that the outcrop points are included in the object region. When many outcrop points are missed, $E_m(i,k)$ is not calculated correctly.

In this Section, first, we discuss the silhouette refinement to extract more outcrop points. We also examine that how the missing and additional parts of the silhouettes affect the value of $E_m(i,k)$.

4.3.1 Silhouette Refinement

In recent works, many methods have been proposed to extract accurate silhouettes for the shape reconstruction, although the methods suppose the silhouette extraction not in multiple frames but in only 1 frame. Under the assumption of the coexistence of neighborhoods in the voxel space, graph cut theory is used for the silhouette extraction [45]. The assumption for characteristics of the background objects is also used for the silhouette extraction [51]. The shapes can be reconstructed without the silhouette extraction from images [54]. In the method, the obtained images are divided into small regions in advance. The shape is reconstructed to keep the consistency between colors of the divided small regions and the shape. When many cameras are used for the shape reconstruction, the color consistency in images is possible to be a cue of the silhouette extraction. Based on space carving [21] or voxel coloring [42], the extracted silhouettes can be refined by the consistency of obtained images as described in Chapter 2. In other methods, the shape is reconstructed under the assumption that the silhouettes must be missing. With SPOT(Sparse Pixel Occupancy Test) [5, 46] or SfIS(Shape from Inconsistent Silhouettes) [22], how many times a voxel is projected into the silhouettes is counted. The reconstructed shape is a set of the voxels which are projected into the silhouette more times than a threshold. The method ignores that some voxels are projected to out of silhouettes in a few images. The missing parts of a few silhouettes do not affect the shape reconstruction.

Although each method has some assumptions, a certain level of the silhouette and shape refinement is given by the methods. We adopt the silhouette refinement method by the consistency of colors in Chapter 2. In addition to the method, the shape is reconstructed with SPOT [5] to avoid that the missing parts of silhouette make the missing parts of the reconstructed shape. However, any method cannot refine the silhouettes and the shapes completely.

The percentages of missing and additional regions in silhouettes are described as 4.3% and 2.1% in [5]. We have also described the missing percentage as less than 5.21% and refined to 2.33% in Chapter 2. In this Chapter, we assume that the percentages are less than 10%. Under the assumption, how many the outcrop points are extracted correctly is discussed.

4.3.2 Silhouette Incompleteness and Outcrop Extraction

We reconstruct a simulation shape of a triceratops from obtained silhouettes with random noise. The random noise is stochastically generated in the silhouettes. From the silhouettes with random noise, the outcrop points are extracted. We examine how many outcrop points are included in the original

Table 4.1: Relation between silhouette missing/additional percentages and outcrop point extraction. (6 cameras)

		Additional percentage				
		1%	2%	5%	10%	20%
Missing percentage	1%	88.6% (88)	87.2% (94)	82.8% (99)	61.0% (159)	19.8% (582)
	2%	88.0% (125)	86.6% (119)	85.2% (142)	65.7% (198)	24.3% (668)
	5%	**90.0%** (279)	85.5% (297)	83.1% (337)	75.5% (433)	44.2% (978)
	10%	89.3% (653)	88.9% (682)	86.9% (800)	79.3% (929)	58.9% (1434)
	20%	**90.6%** (1072)	**91.9%** (1142)	**90.6%** (1215)	87.2% (1222)	69.7% (1302)

Table 4.2: Relation between silhouette missing/additional percentages and outcrop point extraction. (12 cameras)

		Additional percentage				
		1%	2%	5%	10%	20%
Missing percentage	1%	**92.0%** (1129)	**93.1%** (1245)	**93.2%** (1673)	**93.2%** (2311)	**94.0%** (4873)
	2%	**95.1%** (2976)	**95.6%** (3206)	**96.0%** (4083)	**95.6%** (5357)	**96.0%** (6967)
	5%	**97.0%** (7100)	**97.4%** (7006)	**97.0%** (6965)	**96.8%** (6103)	**96.1%** (2336)
	10%	**97.4%** (5003)	**97.8%** (4602)	**97.0%** (3270)	**96.4%** (1867)	**94.8%** (135)
	20%	**96.3%** (1055)	**97.4%** (923)	**97.8%** (447)	**98.6%** (73)	– (0)

Table 4.3: Relation between silhouette missing/additional percentages and outcrop point extraction. (20 cameras)

		Additional percentage				
		1%	2%	5%	10%	20%
Missing percentage	1%	**97.5%** (3480)	**97.4%** (4006)	**97.7%** (5569)	**98.1%** (7807)	**98.4%** (2766)
	2%	**98.4%** (7002)	**98.5%** (7361)	**98.4%** (7873)	**98.5%** (5618)	**98.5%** (200)
	5%	**98.9%** (5935)	**98.8%** (5230)	**98.7%** (3391)	**98.5%** (867)	– (0)
	10%	**98.8%** (1887)	**98.2%** (1413)	**97.8%** (459)	**100.0%** (16)	– (0)
	20%	**96.9%** (131)	**100.0%** (66)	**100.0%** (1)	– (0)	– (0)

region of the object. When the number of cameras is changed to 6, 12 and 20, the percentage that the outcrop points are included in the original shape is changed as shown in table 4.1, table 4.2 and table 4.3. The elements over 90% are written in bold characters and the elements over 95% are written with underlines. The elements in parentheses are values of the number of extracted outcrop points. Under the conditions that the number of cameras is small, there are many missing parts of the silhouettes and less additional parts of the silhouettes, false outcrop points not included in the original regions are extracted as shown in table 4.1. The false outcrop points are included in the additional regions of the reconstructed shape. When the number of cameras is greater than or equal to 12, most of extracted outcrop points are included in the original regions, even if there are many missing parts and many additional parts in the silhouettes as shown in table 4.2 and table 4.3. In conclusion, most of extracted outcrop points are included in the original regions under the conditions that the number of cameras is greater than or equal to 12 and the percentages of missing parts and additional parts of silhouettes are less than 10%.

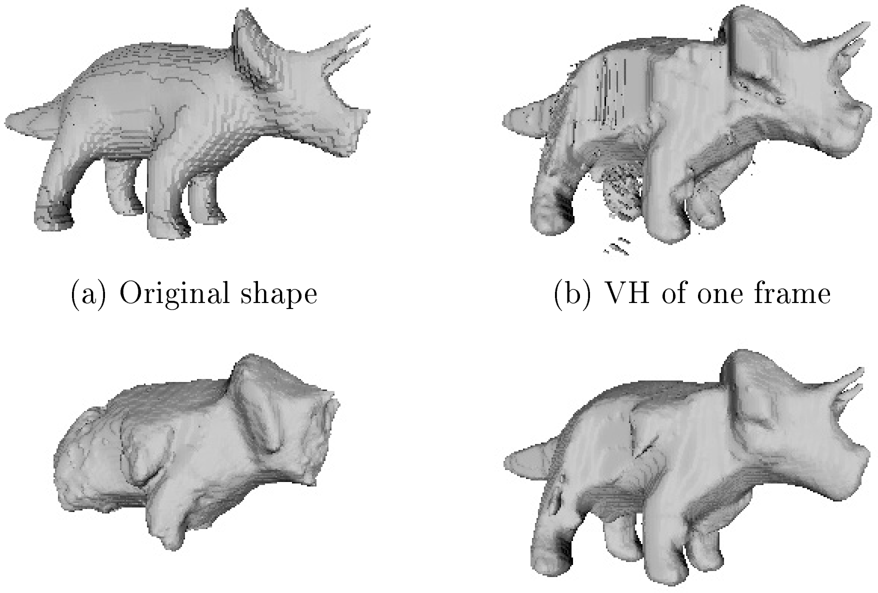
4.4 Experimental Results

The experimental results for simulated and real objects were used to evaluate how valid our proposed function, $E_m(i,k)$, was. We examined whether the outstanding parts on the object's surface were preserved.

4.4.1 Simulation Data

We obtained silhouettes from the simulation data of a triceratops toy. We arranged 12 cameras to observe the toy. We adopted our proposed method of volume integration for the silhouettes. The percentages of the missing and additional parts of silhouettes were set to A: 1%, B: 2%, C: 5%, D: 7.5% and E: 10%. The settings of the number of cameras and the percentages of the missing and additional parts correspond to Table 4.2. The objects in these experiments were translated by +0.5 voxels along the X,Y, and Z-axes and rotated by 1 degree around the axes for each frame. Experimental results for A,B,C,D and E are shown in Figure 4.2, 4.3, 4.4, 4.5 and 4.6. (a) is a original shape, (b) is a visual hull reconstructed in 1 frame, (c) is an integrated shape in 50 frames, and (d) is an integrated shape in 50 frames with our proposed method. For our method, E_m^{th} is set to 0.97. All shapes are depicted with surface patches obtained using the marching cube algorithm [28] and a smoothing process.

The visual hull in one frame includes many additional regions on its surface, which angulate visual hull as shown in each Figure (d) of A,B,C,D and E. Some additional regions are floating away from the visual hull. The shape integrated without our method has many missing parts as shown in each Figure (c). This is caused by frames that have large error in estimating motion. Especially, E of Figure 4.6(c) illustrates no regions since there are some frames which do not preserve the outcrop points on the object surface. All integrated shapes except for D and E have accurate shapes with our new method. The integrated shape includes the original regions. The integrated shape in Figure 4.5(d) illustrates the reconstructed shape with missing parts in the triceratops' horn. The shape is reconstructed from silhouettes with many missing parts. The missing parts of the visual hull have been accumulated into the integrated shape. The shape retains the outstanding parts unlike that with the conventional method of integration. SPOT [5] [46] with optimal parameters described in Section 4.1 will solve the missing partially. In Figure 4.6(d), the integrated shape is the same shape to the shape in Fig-



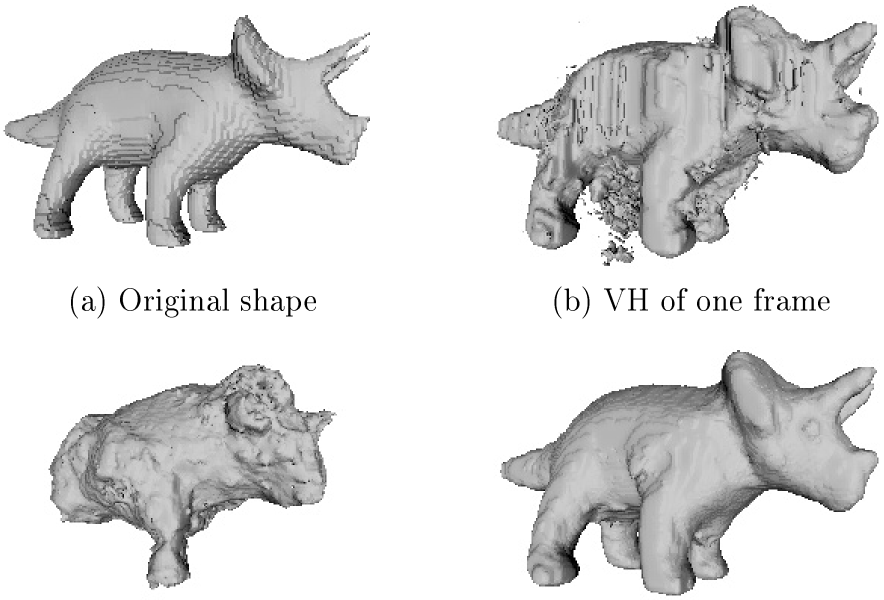
(a) Original shape

(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

Figure 4.2: Reconstructed shape of triceratops toy. (A : (Percentage missing, Percentage added)=(1%, 1%))



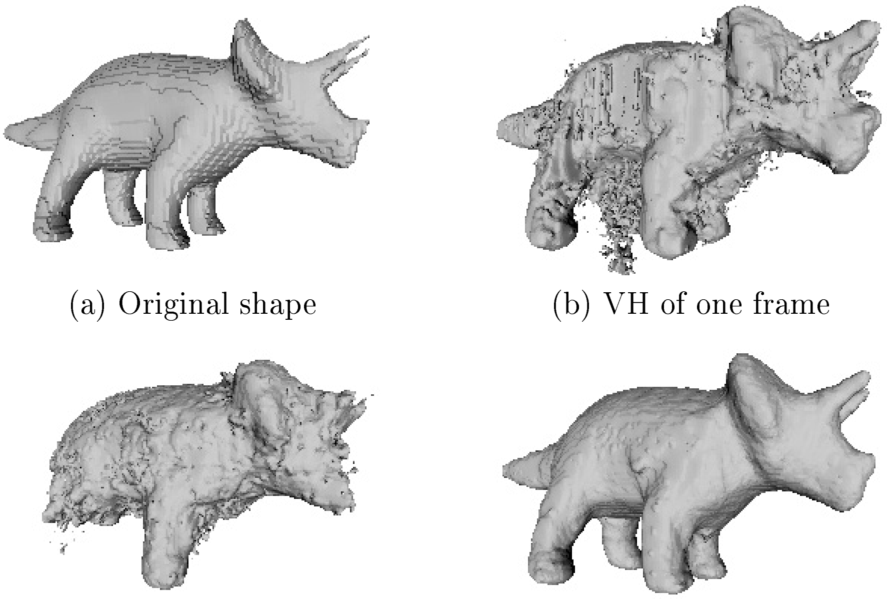
(a) Original shape

(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

Figure 4.3: Reconstructed shape of triceratops toy. (B : (Percentage missing, Percentage added)=(2%, 2%))



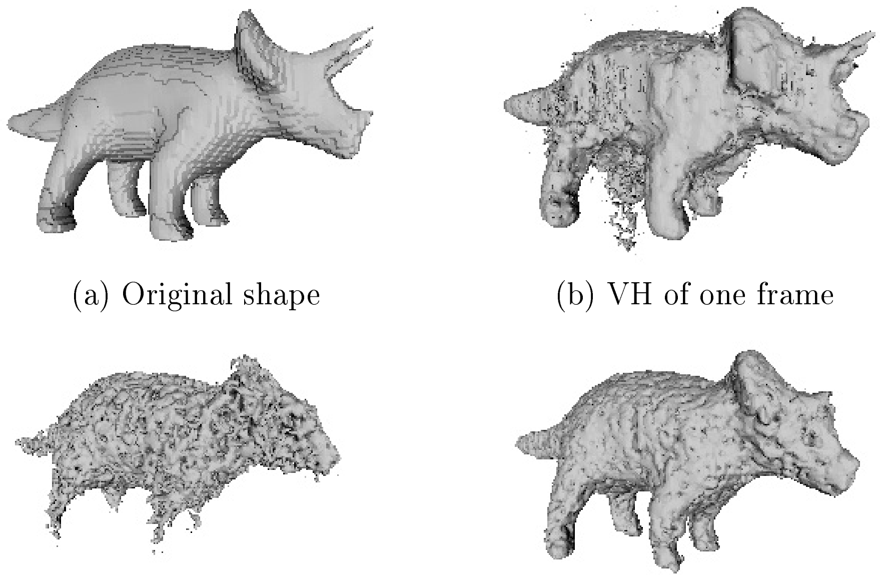
(a) Original shape

(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

Figure 4.4: Reconstructed shape of triceratops toy. (C : (Percentage missing, Percentage added)=(5%, 5%))



(a) Original shape

(b) VH of one frame

(c) Conventional VH

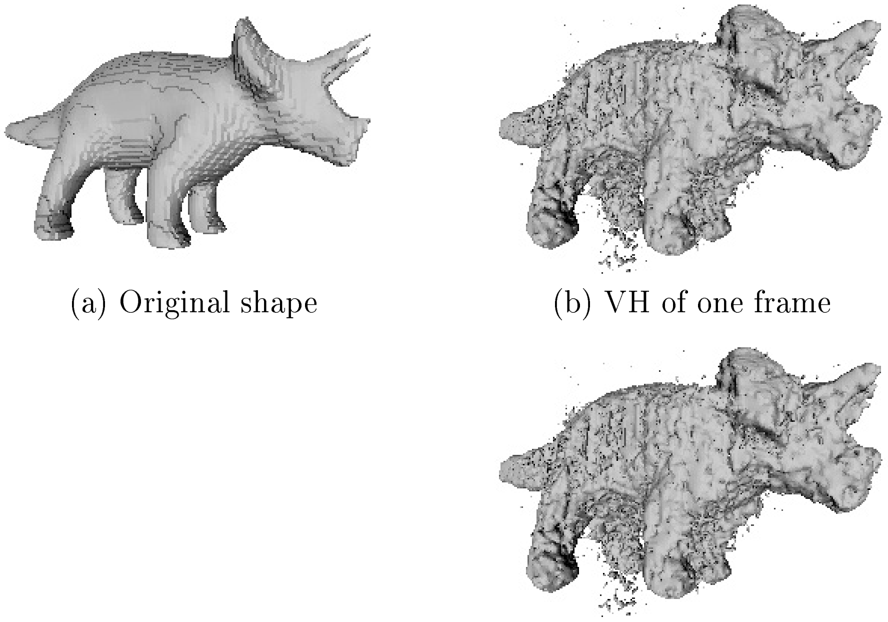
(d) Proposed VH

Figure 4.5: Reconstructed shape of triceratops toy. (D : (Percentage missing, Percentage added)=(7.5%, 7.5%))

ure 4.6(b), since any frame does not satisfy $E_m(0, i) > E_m^{th}$ under the missing and additional percentages.

In Figure (d) of A, B and C, there are not many additional regions on the object surface unlike the visual hulls from the silhouettes of one frame; there are not many missing parts unlike the shapes with the conventional method of integration. Compared with the original visual hulls, the shapes with our new method have smooth surfaces while preserving the outstanding parts on their surfaces. They are similar to the original shapes. To conclude, our proposed evaluation function, $E_m(i, k)$, accurately preserves the outstanding parts on the object surface when the percentages of the missing and additional parts are below 10%.

We also evaluate the integrated shapes with the evaluation method described in Subsection 2.2.1. The percentages of the additional regions, the missing regions and the error regions for the original regions are shown in



(a) Original shape

(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

Figure 4.6: Reconstructed shape of triceratops toy. (E : (Percentage missing, Percentage added)=(10%, 10%))

Table 4.4. Compared with the visual hulls of (b), the shapes of (d) reconstructed with our proposed method have much less error regions. Note that the missing regions are increasing by integrating visual hulls since we also adopt SPOT to reconstruct the integrated shapes. A voxel not included in the visual hull in one frame can be included in the integrated shape when we use SPOT.

Table 4.4: Silhouette integration and volume error regions. (A : (Percentage missing, Percentage added)=(1%, 1%), B : (2%, 2%), C : (5%, 5%), D : (7.5%, 7.5%), E : (10%, 10%), with 12 cameras.)

		Missing percentage	Additional percentage	Error percentage
A	(b)	1.3%	28.8%	30.2%
	(c)	17.4%	15.6%	33.0%
	(d)	0.9%	25.6%	26.5%
B	(b)	2.9%	46.3%	49.2%
	(c)	25.4%	21.0%	46.5%
	(d)	0.2%	35.1%	35.2%
C	(b)	19.3%	39.1%	58.4%
	(c)	35.1%	17.0%	52.1%
	(d)	2.5%	27.6%	30.1%
D	(b)	43.5%	10.0%	53.5%
	(c)	87.6%	0.2%	87.9%
	(d)	73.5%	1.3%	74.8%
E	(b)	66.0%	16.5%	82.5%
	(c)	100.0%	0.0%	100.0%
	(d)	66.0%	16.5%	82.5%

4.4.2 Real Environment

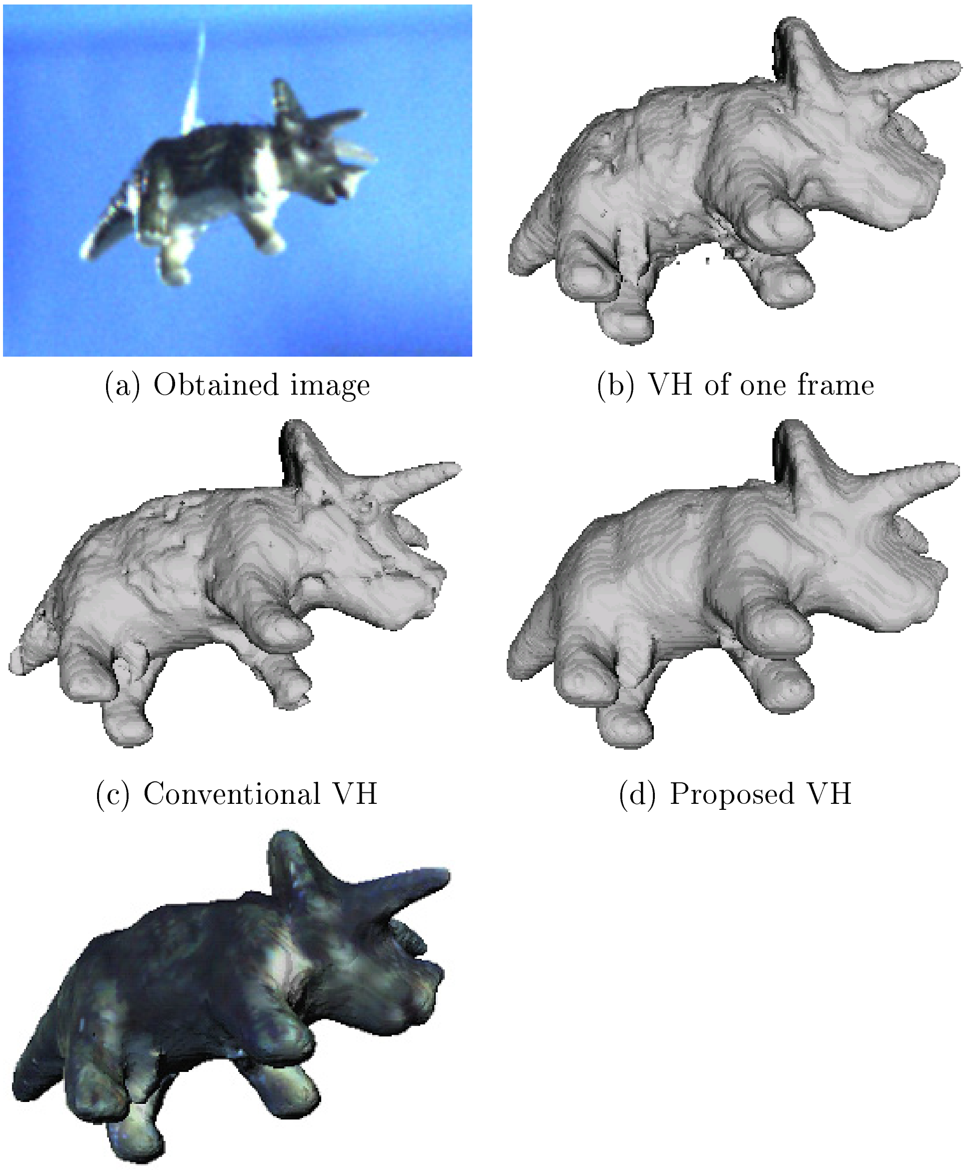
We captured a triceratops toy with multiple cameras. Its shape was reconstructed from the silhouettes in multiple frames. The integrated shapes with and without our proposal method are shown in Figure 4.7. An average of 8.54% of the silhouettes was missing. Additional silhouettes were 5.27%. The silhouettes were refined with a silhouette refining method [52]. The threshold E_m^{th} was set to 0.95. There were some small missing parts in the integrated shape obtained with the conventional method of integration as shown in Figure 4.7. The feet of the triceratops had gaps. The integrated shape with our new method did not have such missing parts. Outstanding points on the surface were preserved.

The shape of a mammoth toy was similarly reconstructed from silhouettes in multiple frames. The integrated shapes with and without our new method are presented in Figure 4.8. An average of 3.22% of the missing silhouettes was extracted. Additional silhouettes were 4.29%. A large part of the head in the integrated shape obtained with the conventional method of integration is missing. The shape integrated with our proposed method does not have such missing parts. Outstanding points on the surface are preserved. Compared with the shape reconstructed from silhouettes from one frame, the areas of additional regions on the surface of the shape are decreased.

We also extract silhouettes with the random pattern backgrounds described in Chapter 2. The silhouettes are extracted from obtained images of a horse toy. The integrated shapes are reconstructed as shown in 4.9. An average of 7.9% of the missing silhouettes was extracted. Additional silhouettes were 11.0%. We assumed that the missing and additional percentages are below 10% in Section 4.3. The percentage added is over 10% in this case. We cannot guarantee that our proposed function $E_m(i,k)$ is valid in this case since the outcrop point are not guaranteed to be included in the object region under the missing percentage. If the threshold E_m^{th} is set to 0.90, only 5 frames are adopted to be integrated. It caused by the large missing rate. The integrated shape from the silhouettes in that 5 frames (in Figure 4.9(d)), however, is more accurate shape than the visual hull of 1 frame (in Figure 4.9(b)). The integrated shape preserves the outstanding parts.

4.5 Discussions and Conclusions

In this Chapter, we proposed an intelligent method of integrating silhouettes in multiple frames, which enabled us to reconstruct accurate shapes even if there were missing or additional parts in the silhouettes. We could reconstruct a more accurate shape than the visual hull of one frame by integrating the silhouettes in multiple frames. We discussed how the missing and additional regions affect the extraction of outcrop points. Based on the discussion, we designed an evaluation function, $E_m(i,k)$, which indicates how many outcrop points are preserved in the integrated shape. Some integrated



(a) Obtained image

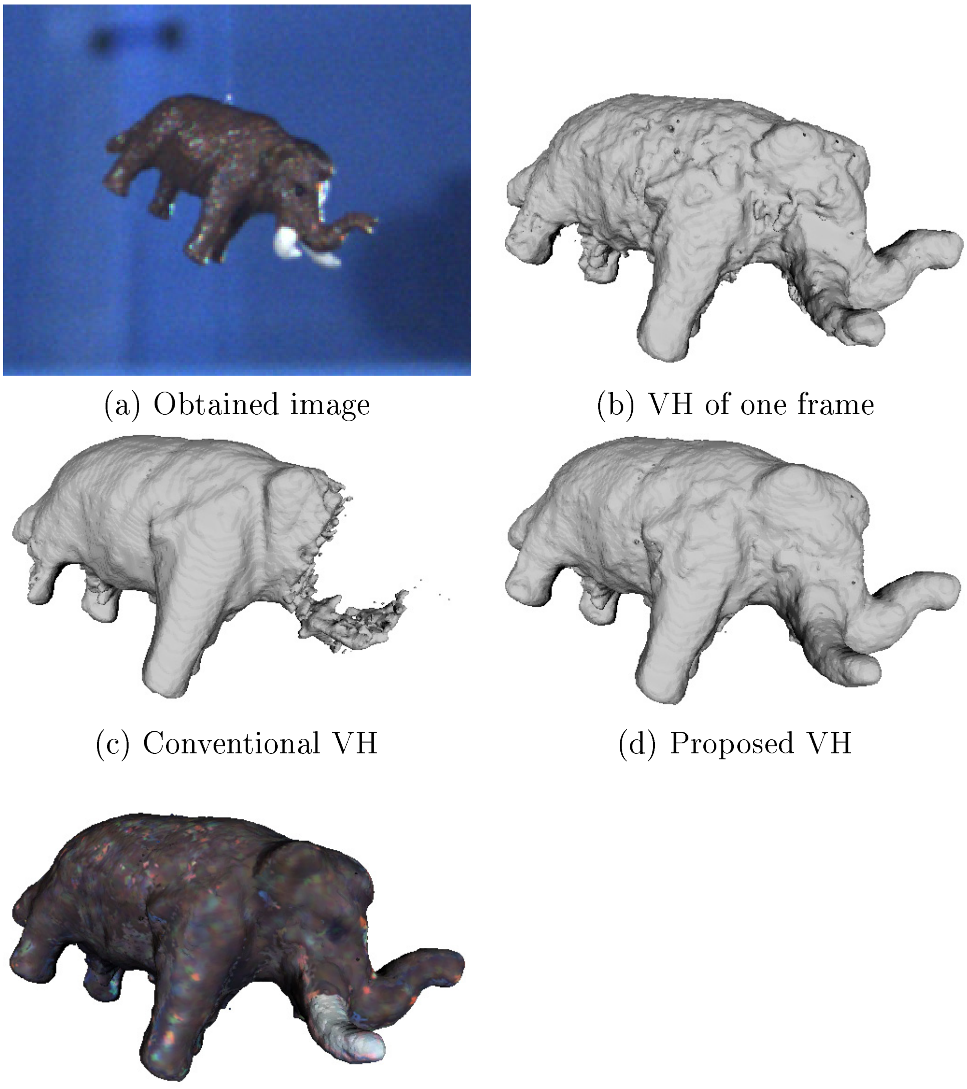
(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

(e) Proposed VH with color

Figure 4.7: Reconstructed shape of triceratops toy.



(a) Obtained image

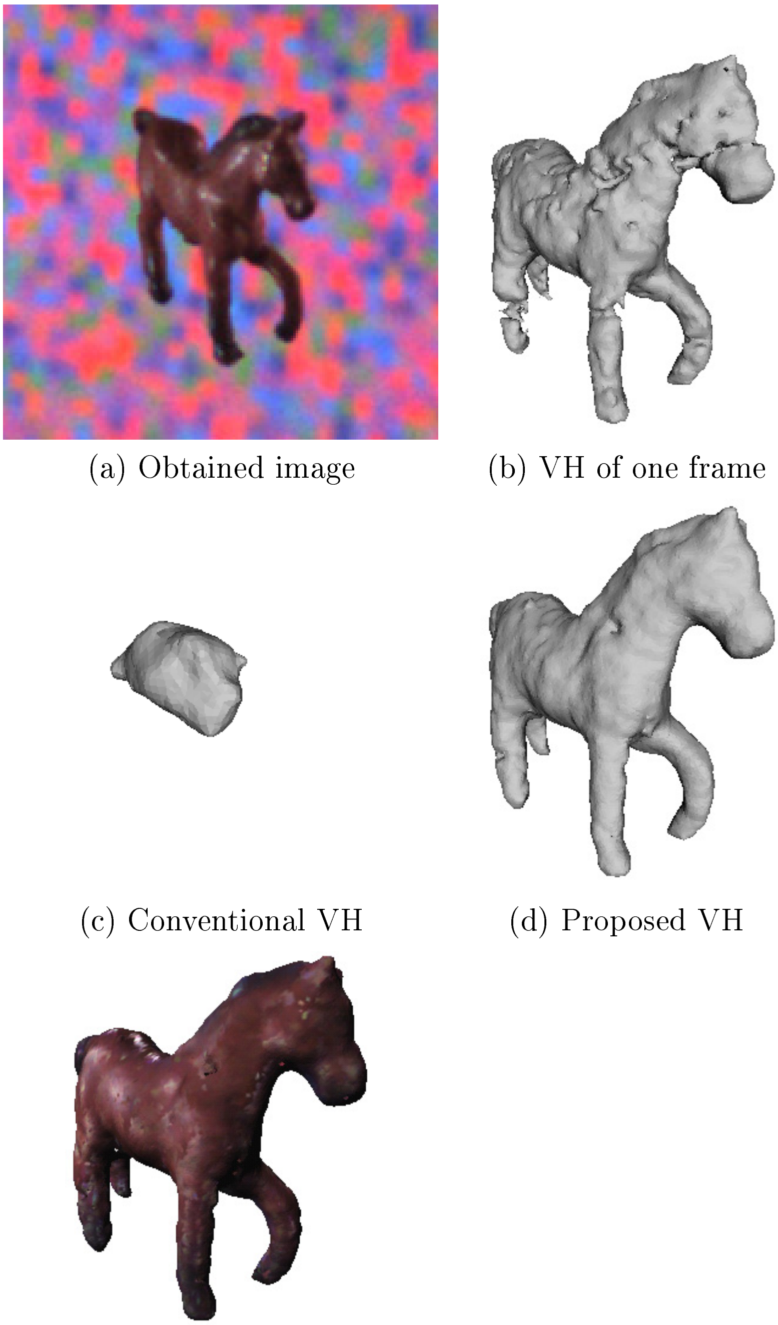
(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

(e) Proposed VH with color

Figure 4.8: Reconstructed shape of mammoth toy.



(a) Obtained image

(b) VH of one frame

(c) Conventional VH

(d) Proposed VH

(e) Proposed VH with color

Figure 4.9: Reconstructed shape for a toy of horse with random pattern backgrounds.

shapes with $E_m(i, k)$ were presented as experimental results. These shapes included outstanding parts of the object. We solved the problem where integrated shapes were missing when motion was incompletely estimated. The shapes were also more accurate than when the visual hull was calculated in one frame.

In future work, we intend to set threshold E_m^{th} automatically, which will be set by the percentages of missing and additional parts of silhouettes, or the numbers of available frames.

Chapter 5

Conclusions and Discussions

5.1 Conclusions

We discussed 3D shape reconstruction from silhouettes of multiple frames. Our main contribution of our research is the elimination of assumptions on colors and textures of target objects.

First, accurate silhouette extraction is required to reconstruct shapes from silhouettes. We proposed the random pattern background for extraction silhouettes of objects in unknown colors, as described in Chapter 2. The random pattern has separated small regions which are filled with randomly selected colors. With the random pattern backgrounds, silhouettes with less missing parts can be extracted even for the objects in unknown colors. Considering the color consistency between multiple cameras, the less missing parts are also refined.

Next, the object motion estimation is required to integrate silhouettes of multiple frames. The rigid object motion is considered as the changing positions of cameras virtually. The difficulty of the motion estimation is that the object shape is reconstructed only from silhouettes. The reconstructed shape is called the visual hull. Since the visual hull includes the additional regions, the corresponding points between different frames are difficult to be extracted. From the visual hull and silhouettes, we extracted new kind of 3D feature points, as described in Chapter 3. We named the feature points as outcrop points. The outcrop points are guaranteed to be included in the object region of the visual hull, when the target object has a surface with outstanding points. We confirmed that only few and small outstanding points are required for the condition. Most of general objects in the real world have such outstanding points.

Whereas, even when the error of the estimated motion is small, the integrated shape of multiple frames has many missing parts. The error of the motion estimation is caused by failure of feature point extraction. To estimate the object motion correctly, the completely corresponding feature points between frames are required. By the incompleteness of the silhouette extraction, the corresponding points are difficult to be extracted. Eventually, the motion estimation has an accuracy limit. We focused on the fact that the outcrop points tend to be extracted on the outstanding parts on the object surface, as described in Chapter 4. We proposed the visual hull integration method which preserves the outstanding parts of the object shape. The evaluation function represents how much outstanding parts are preserved by integrating a visual hull of a frame. Based on the evaluation, integrated visual hulls are selected. With our proposed method, the integrated shape of multiple frames can preserve the outstanding points in it.

5.2 Discussions

In this paper, the silhouette extraction and the silhouette integration are discussed. In this Section, the relationship between them is discussed. We describe how the silhouette extraction is affected by the silhouette integration, and conversely.

5.2.1 Silhouette Extraction for Silhouette Integration

The accuracy of the silhouette extraction improves that of the silhouette integration. The accuracy of the outcrop extraction depends on that of the silhouette extraction as described in Chapter 4. The missing parts and additional parts of the extracted silhouettes prevent the outcrop point extraction. Extracted outcrop points are not guaranteed to correspond to outstanding points on the object surface, when the silhouettes have the missing and additional parts. The error of the outcrop point extraction leads to the motion estimation error for the object. In other words, the accuracy of the silhouette extraction leads to the accuracy of the object motion estimation, and the accuracy of the silhouette extraction.

Then, is the accuracy of the silhouette extraction enough for integrating silhouettes? We have already discussed on the required accuracy of the silhouette extraction for the silhouette integration in Chapter 4. The experimental result of the silhouette integration with random pattern background is shown in Figure 4.9. In Chapter 4, our proposed silhouette integration method assumes the missing and additional regions are less than 10%. With the random pattern backgrounds, the assumption is almost satisfied. In the experimental result, it is confirmed that the integration method can be applied to the silhouettes extracted with the random pattern backgrounds.

The availability of the silhouette integration depends not only on the accuracy of the silhouette extraction, but also on the complexity of the object shapes. We have discussed about how much complexity is required for the object motion estimation in Chapter 3. To extract sufficient outcrop points, the objects should have outstanding parts with sufficient length and frequency on their surfaces. The length is more than 8% of whole object shape, and the frequency is an outstanding part per 30 degrees. The condition is satisfied with general objects in the real world. When the object motion between any frames can be estimated, silhouettes can be integrated with our proposed method based on frame evaluation as described in Chapter 4.

5.2.2 Silhouette Integration for Silhouette Extraction

The accuracy of the silhouette integration also improves that of the silhouette extraction. More number of obtained images helps us to refine the extracted silhouettes. For the silhouette refining, colors of the object are estimated from the obtained images.

The estimated colors are compared with their corresponding colors of backgrounds in case that our silhouette refining method is used as described in Chapter 2. When the colors are similar, we can judge which parts are included in the object. More number of obtained images leads to the accurate estimation of the object colors. We estimate the object colors under the assumption that the colors are diffuse colors. When large number of images is given, obtained specular colors can be regarded, since the specular colors are clearly obtained only from limited angles. The accuracy of estimated diffuse colors supports the accuracy of the silhouette refining. With SPOT(Sparse Pixel Occupancy Test) [5, 46] or SfIS(Shape from Inconsistent Silhouettes) [22], the missing parts of the visual hulls can be increased, even if the extracted silhouettes have the missing parts.

Then, is the accuracy of the silhouette integration enough for integrating extraction? In Chapter 4, we concluded it is required that the silhouette are extracted with less than 10 % missing and additional parts and the number of images is more than or equals to 12 for the silhouette integration. From the images obtained from 12 cameras in n frames, $12n$ images are given in the silhouette integration method. Even if only 1/10 cameras can observe a part of the object surface, $1.2n$ cameras expect to observe the part in our method. When $1.2n$ is more than or equal to 3, the specular element can be specified. By integrating silhouettes in 3 frames or more, the part is guaranteed to be observed with 3 cameras or more. The selected frames should not be only adjacent frames, since cameras located sparsely are required to observe whole surface.

The increase of the number of the cameras provides the accuracy of the color estimation for the object. A color of each region on the object surface is calculated from images of cameras that observe the region. In proposed method, which camera observes the region is determined from the positions of the cameras [26, 40] . Diffuse colors are correctly extracted with the increase of the number of cameras, since specular colors are specified as an outlier. When the diffuse colors are correctly estimated, silhouettes are also correctly refined with our proposed method discussed in Section 2.4.

It is true under the condition that illuminance intensity for each part of the object surface does not change in multiple frames. The condition is difficult to be satisfied in general. In a previous work, geotensity restriction has been proposed to reconstruct shapes with the illuminance change [29]. To regard the illuminance change, the object should be equally illuminated from all directions. Instead of the illumination settings, we can adopt an intelligent method for silhouette extraction. In this paper, we extract the silhouettes focus on values of U and V in YUV color space. Y indicated the intensity. With the silhouette extraction method, some illuminance change can be regarded. Alternatively we also adopt the geotensity restriction for shape reconstruction in case that 20 cameras are provided to observe the object, since extracted silhouettes can be refined enough with 20 cameras as described in Chapter 2.

5.2.3 Positions of Cameras Given by Multi-frame Image Integration

Outcrop parts on an object surface can be observed by cameras at various directions. Many cameras are not required to observe the outcrop parts. The outcrop parts characterize a shape of an object. The outcrop parts play an important role in aesthetic aspect.

The range of directions for cameras which observe smoothing parts on the object surface is narrower than that for cameras which observe outcrop parts. For reconstructing smoothing surfaces, more cameras are required. As shown in Fig 5.3, it is confirmed that our proposed silhouette integration method provided silhouettes obtained from cameras located at sequential positions. The sequential position change is efficient for reconstructing smooth surfaces.

In Section 4.4, 12 cameras are set shown in Figure 5.1. The cameras are homogeneously located on surface of a sphere. In 50 frames, the cameras provide many silhouettes shown in Figure 5.2, which is discussed in Chapter 3 and Chapter 4. For adopting the frame selection method proposed in Chapter 3 to the cameras, only cameras shown in Figure 5.3 are used to reconstruct 3D shapes. Since a few frames guarantee to preserve outstanding parts in the reconstructed shapes, only a few cameras are selected. However, cameras at sequential positions remain to be adopted, since neighboring frames to the first frame tend to be selected. When the motion of the objects is small, outcrop points likely have their corresponding points between multiple frames. The cameras at sequential positions are valid for reconstructing smooth surfaces of the object.

In proposed researches, it is required that the positions of cameras are expected to change sequentially. *Contour generators*, which are extracted from silhouettes, are used to reconstruct 3D shapes in many researchers [7, 25, 11, 14, 35]. Whereas, with turntables, cameras at discrete and planar positions only are realized [41]. Moreover, the cameras are located on the surface of an upper hemisphere. In our proposed silhouette integration method for multiple frames, cameras at locally-sequential and omnidirectional positions can be used to reconstruct 3D shapes.

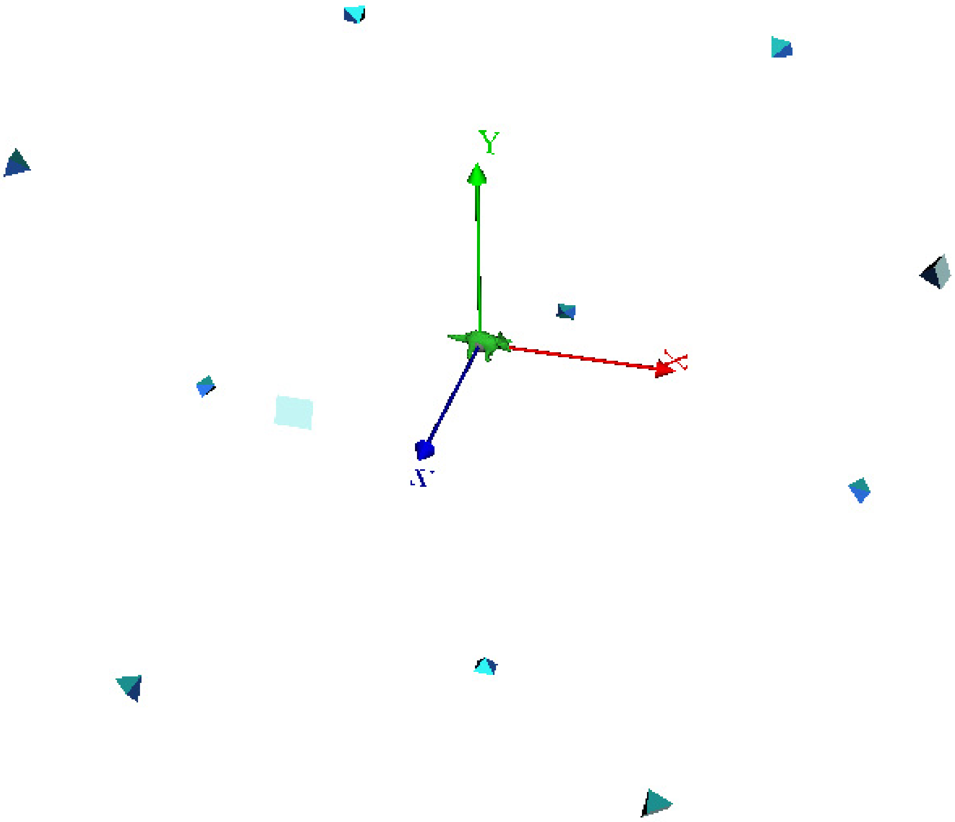


Figure 5.1: Positions of physical 12 cameras.

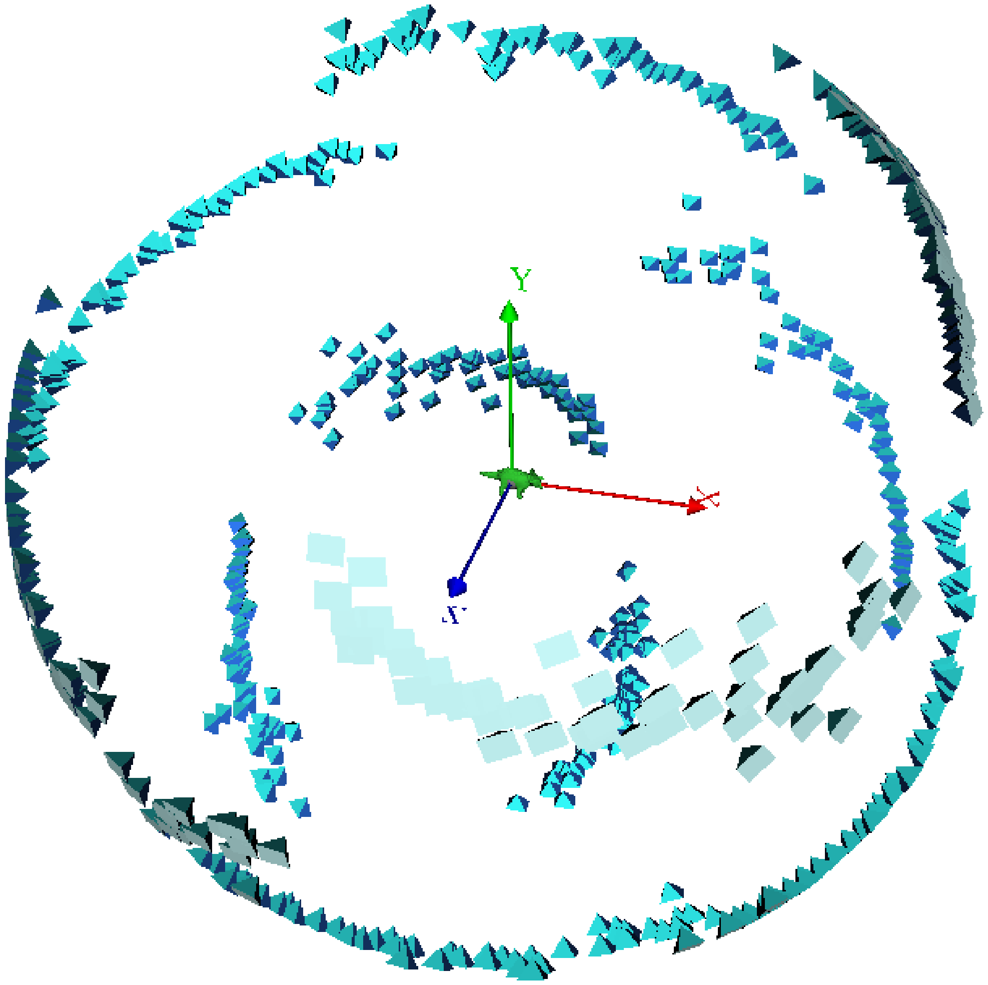


Figure 5.2: Positions of 12 cameras in 50 frames.

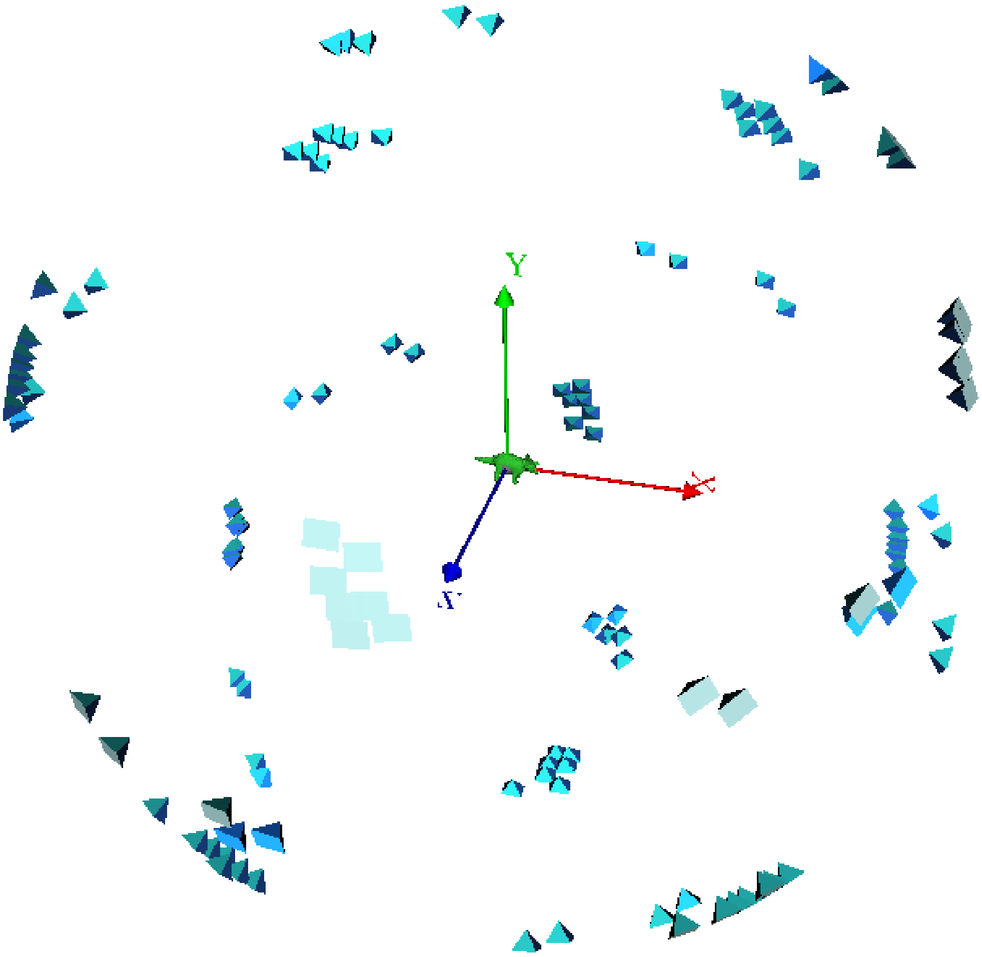


Figure 5.3: Positions of selected 12 cameras in 50 frames.

Chapter 6

Future Works

Accuracy Limit of Shapes and Placement/Number of Cameras

The relationship between the accuracy limit based on the sampling error and the number of cameras are described in Chapter 3. For any objects, the accuracy limit can be satisfied with a realistic number of cameras. For instance, 242 cameras are required to acheive the accuracy for a triceratops shape. However, the discussion was limited to the number of cameras only for sample objects in this paper. It was not confirmed that what placement of cameras could acheive the accuracy. The discussion on the placement and number of cameras is one of our future works.

The complexity of shapes is related to the placement and number of cameras. To describe the relation between the placement and number of cameras and the accuracy, a definition of the complexity of shapes would be required. Based on the definition, the placement and the number of cameras are determined. In previous researches of 3D shape recognition, *viewsphere* [55] or *aspect graph* have been proposed. To describe the relationship between the placement and number of cameras and the accuracy, the concept of the viewsphere might be useful. Consider the appearance of an object when the object is set on the center of a unit sphere and the object is observed from each point on the sphere surface. Based on the appearance difference, the sphere surface is divided. The division of the surface can be considered to represent a characteristic of the object. The complexity of the object can be measured with the division of the surface. The most suitable placement could be determined with the division. A problem is that the division is unknown before observing the object. To solve the problem, some assumption or some condition will be required.

Texture Mapping for Reconstructed Models

We have discussed the accuracy of shapes in this paper. For presentation of the reconstructed shapes, good appearance is also required. We have not discussed the appearance enough. In the experimental results shown in Figure 4.7 and Figure 4.9, the models are not good in appearance. The problem is caused by luck of the resolution of texture images. The integration method of textures is also not enough.

Recent consumer single-lens reflex cameras are available for the purpose. The resolution obtained by them is more than 4000×3000. The cameras are not capable for synchronous observation. To integrate images of moving objects in multiple frames, the synchronous observation is required. Dragonfly, Flea, Scorpion and Grasshopper of PointGray Inc. can observe the objects synchronously. For our current system, the Dragonflies which resolutions are VGA or XGA are adopted. For the special cameras, the resolution is seriously restricted.

We would use the consumer single-lens reflex cameras in conjunction with the synchronous cameras. The shapes are obtained with the synchronous cameras, and the fine textures are obtained with the consumer single-lens reflex cameras. A problem is registration between shapes and textures.

Elimination of Rigid Object Assumption for Objects

We have discussed the shape reconstruction in multiple frames with assumption that the object is rigid. In case when the object is observed in time sequences, one of advantages would be the availability of motion description for the object. In this paper, we have not considered the advantage. Our proposed methods can be also applied to non-rigid objects with some extension. The extension will be done with continuity in space or time. Colors or textures of the objects will be also utilized.

In our laboratory, researches on shape reconstruction for articulated objects have been done. They have been also reconstructed shapes from images in multiple frames. Iiyama [16] reconstructed shapes of the articulated objects from silhouettes in multiple frames, His idea is based on segmentation and integration for visual hulls. Funatomi [12] also reconstructed shapes from images in multiple frames. His target is to reconstruct shapes of human bodies. These researches have been discussed on the assumption that the objects are articulated objects. The outcrop point extraction method described in Chapter 3 and the silhouette extraction method described in Chapter 2 can be applied to the articulated objects with no extension. With some extension for segmentation, the frame evaluation method can be also applied to the articulated objects. Focusing on human bodies, we would like to extend our proposed methods.

Interfaces for 3D Shapes

We have discussed the accurate shape reconstruction in this paper. Interfaces for the reconstructed object shapes, however, have rarely been proposed. The object shapes are ordinarily displayed on a LCD display and manipulated with a mouse.

Recently, signs [50] and digitalized cultural objects [15] are displayed in response to the hand posture. These systems realize *Augmented Reality (AR)*. One of the major problems of these systems is the accuracy of the estimated posture of the hand. When an accurate posture of the hand is estimated, the augmentation of reality is well realized.

We have proven *the matrix pattern glove*. The matrix pattern has many small square regions with different colors. Each region plays the role of an independent marker. The matrix pattern can be considered as many identifiable markers on the hand. The many markers enable us to acquire the posture of the hand.

Bibliography

[1] M. Brand, K. Kang, and D. Cooper. Algebraic solution for the visual hull. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 30–35, 2004.

[2] German Kong Man Cheung. *Visual Hull Construction, Alignment and Refinement for Human Kinematic Modeling, Motion Tracking and Rendering.* PhD thesis, Technical Report CMU-RI-TR-03-44, Robotics Institute, Carnegie Mellon University, October 2003.

[3] German Kong Man Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, volume 1, pages 77–84, 2003.

[4] German Kong Man Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette across time part i: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, May 2005.

[5] German Kong Man Cheung, Takeo Kanade, Jean-Yves Bouguet, and Mark Holler. A real time system for robust 3d voxel reconstruction of human motions. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 714–720, June 2000.

[6] R. Cipolla, K.E. Astrom, and P.J Giblin. Motion from the frontier of curved surfaces. In *IEEE International Conference on Computer Vision (ICCV)*, pages 269–275, 1995.

[7] R. Cipolla and A. Black. The dynamic analysis of apparent contours. In *IEEE International Conference on Computer Vision (ICCV)*, pages 616–623, 1990.

[8] Ingemar J. Cox, Sunita L. Hingorani, Satish B. Rao, and Bruce M. Maggs. A maximum likelihood stereo algorithm. *International Journal on Computer Vision and Image Understanding*, 63(3):542–567, 1996.

[9] G. Cross and A. Zisserman. Surface reconstruction from multiple views using apparent contours and surface texture. In *NATO Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics*, 2000.

[10] Carlos Hernàndez Esteban and Francis Schmitt. Silhouette and stereo fusion for 3d object modeling. *International Journal on Computer Vision and Image Understanding*, 96(3):367–392, 2004.

[11] Jean-Sébastien Franco and Edmond Boyer. Exact polyhedral visual hulls. In *Proceedings of the Fourteenth British Machine Vision Conference*, pages 329–338, September 2003. Norwich, UK.

[12] Takuya Funatomi. *Three dimensional shape modeling of human body in various postures by light stripe triangulation.* PhD thesis, Graduate School of Informatics, Kyoto University, 2007.

[13] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman. Structure and motion from images of smooth textureless objects. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2:287–298, 2004.

[14] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2007.

[15] Chun-Rong Huang, Chu-Song Chen, and Pau-Choo Chung. Tangible photorealistic virtual museum. *IEEE Computer Graphics and Applications*, 25(1):15–17, 2005.

[16] Masaaki Iiyama. *3D Object Model Acquisition from Silhouettes.* PhD thesis, Graduate School of Informatics, Kyoto University, 2006.

[17] Masaaki Iiyama, Yoshinari Kameda, and Michihiko Minoh. 4pi measurement system: A complete volume reconstruction system for freely-moving objects. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 119–124, 2003.

[18] Katsushi Ikeuchi, Atsushi Nakazawa, Kazuhide Hasegawa, and Takeshi Ohishi. The great buddha project: Modeling cultural heritage for vr systems through observation. In *IEEE and ACM International Symposium on Mixed and Augmented Reality(ISMAR)*, page 7, Nov 2003.

[19] John Isidoro and Stan Sclaroff. Stochastic refinement of the visual hull to satisfy photometric and silhouette consistency constraints. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1335–1342, 2003.

[20] Takeo Kanade, Kazou Oda, Atsushi Yoshida, Masaya Tanaka, and Hiroshi Kano. Video-rate z keying: A new method for merging images. Technical Report CMU-RI-TR-95-38, Robotics Institute, Carnegie Mellon University, December 1995.

[21] Kiriakos N. Kutulakos and Steven M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[22] Jose Luis Landabaso, Montse Pardas, and Josep Ramon Casas. Reconstruction of 3d shapes considering inconsistent 2d silhouettes. In *IEEE International Conference on Image Processing (ICIP)*, pages 2209–2212, 2006.

[23] Aldo Laurentini. How far 3d shapes can be understood from 2d silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 17(2):188–195, 1995.

[24] S. Lazebnik, E. Boyer, and J. Ponce. On computing exact visual hulls of solids bounded by smooth surfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 156–161, 2001.

[25] Svetlana Lazebnik, Yasutaka Furukawa, and Jean Ponce. Projective visual hulls. *International Journal of Computer Vision*, 74(2):137–165, August 2007.

[26] Victor Lempitsky, Yuri Boykov, and Denis Ivanov. Oriented visibility for multiview reconstruction. In *European Conference on Computer Vision (ECCV)*, volume 3, pages 225–237, May 2006.

[27] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis,

Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the ACM SIGGRAPH*, pages 131–144, July 2000.

[28] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceeding of the ACM SIGGRAPH*, volume 21, pages 163–169, 1987.

[29] Atsuto Maki, Mutsumi Watanabe, , and Charles Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. *International Jornal of Computer Vision*, 48(2):75–90, September 2002.

[30] W.N. Martin and J.K. Aggarwal. Volumetric description of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 5(2):150–158, 1983.

[31] Hiroto Matsuoka, Tatsuto Takeuchi, Hitoshi Kitazawa, and Akira Onozawa. Representation of pseudo inter-reflection and transparency by considering characteristics of human vision. In *Proceedings of the Computer Graphics Forum (Eurographics2002)*, volume 21, pages 503–510, 2002.

[32] Takashi Matsuyama, Xiaojun Wu, Takeshi Takai, and Shohei Nobuhara. Real-time 3d shape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. *International Journal on Computer Vision and Image Understanding*, 96(4):393–434, 2004.

[33] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, John F. Hughes, and Fredo Durand. Defocus video matting. *ACM Transactions on Graphics*, 24(3):567–576, 2005.

[34] Jurriaan D. Mulder and Robert van Liere. Fast perception-based depth of field rendering. In *ACM Symposium in Virtual Reality Software and Technology (VRST)*, pages 129–133, 2000.

[35] Chen Liang nd Kwan-Yee K. Wong. Robust recovery of shapes with unknown topology from the dual space. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(12):2205–2216, December 2007.

[36] Wolfgang Niem. Error analysis for silhouette-based 3d shape estimation from multiple views. In *Proceedings on International Workshop on Synthetic - Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'97)*, pages 6–9, September 1997.

[37] Shohei Nobuhara and Takashi Matsuyama. Deformable mesh model for complex multi-object 3d motion estimation from multi-viewpoint video. In *The Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, pages 264–271, June 2006.

[38] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flanney. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1992.

[39] Gilles Rochefort, Frédéric Champagnat, Guy Le Besnerais, and Jean-François Giovannelli. An improved observation model for super-resolution under affine motion. *IEEE Transactions on Image Processing*, 15(11):3325–3337, November 2006.

[40] Myint Myint Sein, Masaaki Iiyama, and Minoh Minoh. Reconstructing the arbitrary view of an object using the multiple camera system. In *IEEE International Symposium on Micromechatronics and Human Science (MHS 2003)*, pages 83–88, 2003.

[41] Steve Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Rick Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, volume 1, pages 519–526, 2006.

[42] Steven M. Seitz and Charles R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 1067–1073, 1997.

[43] Sudipta N. Sinha and Marc Pollefeys. Multi-view reconstruction using photo-consistency and exact silhouette constraints: A maximum-flow formulation. In *IEEE International Conference on Computer Vision (ICCV)*, pages 349–356, 2005.

[44] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proceedings of the ACM SIGGRAPH*, pages 259–268, 1996.

[45] Dan Snow, Paul Viola, and Ramin Zabih. Exact voxel occupancy with graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, volume 1, pages 345–352, 2000.

[46] Dan Snow, Paul Viola, and Ramin Zabih. Exact voxel occupancy with graph cuts. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 345–352, 2000.

[47] Jonathan Starck and Adrian Hilton. Surface capture for performance-based animation. *IEEE Computer Graphics and Applications*, 27:21–31, 2007.

[48] Christoph Strecha, Rik Fransens, and Luc Van Gool. Combined depth and outlier estimation in multi-view stereo. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2394–2401, 2006.

[49] S. Sullivan and J. Ponce. Automatic model construction and pose estimation from photographs using triangular splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 20(10):1091–1096, October 1998.

[50] Ryuhei Tenmoku, Masayuki Kanbara, and Naokazu Yokoya. Intuitive annotation of user-viewed objects for wearable ar systems. In *Proceedings of the IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 200–201, 2005.

[51] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. Wallflower: Principles and practice of background maintenance. In *IEEE International Conference on Computer Vision (ICCV)*, pages 255–261, 1999.

[52] Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh. Silhouette extraction with random pattern backgrounds for the volume intersection method. In *The 6th International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, pages 225–232, August 2007.

[53] Sundar Vedula, Simon Baker, Steven Seitz, and Takeo Kanade. Shape and motion carving in 6d. In *IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 592–598, 2000.

[54] Gang Zeng and Long Quan. Silhouette extraction from multiple images of an unknown background. In *Asian Computer on Computer Vision(ACCV)*, pages 628–633, 2004.

[55] S. Zhang, G.D. Sullivan, and K.D. Baker. The automatic construction of a view-independent relational model for 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):531–544, June 1993.

[56] Jiang Yu Zheng. Acquiring 3-d models from a sequence of contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 16(2):163–178, February 1994.

List of Publications

Journal Articles

1. Masahiro Toyoura, Masaaki Iiyama, Takuya Funatomi, Koh Kakusho, and Michihiko Minoh, “3D Shape Reconstruction with Incomplete Silhouettes in Time Sequences,” submitted to IEICE Transactions on Information and Systems. (In Japanese)

2. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Silhouette Refining for the Volume Intersection Method with Random Pattern Backgrounds, ” IEICE Transactions on Information and Systems, pp.2413–2424, Vol.J89-D, No.11, 2006. (In Japanese)

3. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “An Accurate Shape Reconstruction Method by Integrating Visual Hulls in Time Sequences, ” IEICE Transactions on Information and Systems, pp.1549–1563, Vol.J88-D-II, No.8, 2005. (In Japanese)

Refereed Conference Presentations

1. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Silhouette Extraction with Random Pattern Backgrounds for the Volume Intersection Method, ” The 6th International Conference on 3-D Digital Imaging and Modeling (3DIM), p.225–232, 2007.

2. Masahiro Toyoura, “3D Shape Reconstruction with Deliberate Textures, ” The 8th Annual Association of Pacific Rim Universities Doctoral Students Conference (APRU DSC), 2007.

3. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Extraction of Outcrop Points from Visual Hulls for Motion Estimation, ” IEEE International Conference on Multimedia & Expo (ICME), pp.217–220, 2006.

4. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Silhouette Refining for the Volume Intersection Method with Random Pattern Background, ” Meeting on Image Recognition and Understanding 2005 (MIRU 2005), pp.1247–1254, 2005. (In Japanese)

5. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “An Accurate Shape Reconstruction Method by Integrating Visual Hulls in Time Sequences, ” Meeting on Image Recognition and Understanding 2004 (MIRU 2004), Vol.2, pp.139–144, 2004. (In Japanese)

Conference Presentations

1. Masahiro Toyoura, Masaaki Iiyama, Takuya Funatomi, Koh Kakusho, Michihiko Minoh, “3D Shape Reconstruction from Incomplete Silhouettes in Time Sequences, ” Technical Report of IEICE PRMU, PRMU 2007-168, Vol.107, No.427, pp.69–74, 2008. (In Japanese)

2. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, Michihiko Minoh, “3D Shape Reconstruction with Random Pattern Backgrounds,” The 6th International Symposium of the Academic Center for Computing and Media Studies, in Cooperation with International Multimedia Modeling Conference (MMM), p.3–16, 2008.

3. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, Michihiko Minoh, “Reconstruction and Manipulation of 3D Shapes with Deliberate Textures, ” The 6th International Symposium of the Academic Center for Computing and Media Studies, in Cooperation with International Multimedia Modeling Conference (MMM), 2008.

4. Michihiko Minoh, Hideto Obara, Takuya Funatomi, Masahiro Toyoura, and Koh Kakusho, “Direct Manipulation of 3D Virtual Objects by Actors for Recording Live Video Content, ” Second International Conference on Informatics Research for Development of Knowledge Society Infrastructure (ICKS’07), pp.11–18, 2007.

5. Masahiro Toyoura, Masayuki Murakami, Satoshi Nishiguchi, Koh Kakusho, and Michihiko Minoh, “Wiki Access Log Analysis for Supporting Research Activities, ” 2006 IEICE General Conference, D-15-39, 2006. (In Japanese)

6. Pitchayagan Temniranrat, Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Model Based Refinement for Thin Part of Visual Hull,” 2006 IEICE General Conference, D-12-55, 2006. (In Japanese)

7. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “Silhouette Refinement for Visual Hulls with Random Pattern Background,” 2005 IEICE General Conference, D-12-133, 2005. (In Japanese)

8. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “An Accurate Shape Reconstruction Method by Motion Tracking,” The 9th General Conference of the Virtual Reality Society of Japan, 2004. (In Japanese)

9. Masahiro Toyoura, Masaaki Iiyama, Koh Kakusho, and Michihiko Minoh, “An Accurate Shape Reconstruction Method by Integrating Visual Hulls in Time Sequences,” The 31st General Conference of the Institute of Image Electronics Engineers of Japan, 2003. (In Japanese)

Patents

1. Masahiro Toyoura, Masaaki Iiyama, Takuya Funatomi, Koh Kakusho, and Michihiko Minoh, “Silhouette Refining for the Volume Intersection Method with Random Pattern Backgrounds,” JP 2005-201292/2007-017364, 2005.