Silhouette Extraction with Random Pattern Backgrounds for the Volume Intersection Method

Masahiro Toyoura
Graduate School of Informatics
Kyoto University

Masaaki Iiyama Koh Kakusho Michihiko Minoh
Academic Center for Computing and Media Studies
Kyoto University

Abstract

In this paper, 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. Introduction

In the volume intersection method [6], 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 [1]. However, these approaches can not be applied to the objects without rich texture. Laser rangefinders are also used for reconstructing shapes [3]. However, laser rangefinders can not 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 chromakey 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 work, extracting the silhouettes without depending on objects' colors by switching two backgrounds in different colors has been proposed [10]. 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 [2], [7]. Reduced observation time makes it possible to reconstruct changing shapes in time sequences. 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 [4], a camera set as a range sensor is required for each position. In defocus matting [8], 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 problems 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.

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE

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, an 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. Missing Parts of Silhouettes and Reconstructed Shapes

Let us denote the cameras that are set around the target object O to capture it by $C_m(m = 1, \cdots, N)$, where N denotes the number of cameras ($N > 1$). From the images of C_m, the 2D region that corresponds to object O is extracted. This region is called the *silhouette* and denoted by R_m. It is guaranteed that O is included in a cone with the apex at the optical center of C_m and the base at R_m is calculated. This cone is called the *visual cone* of camera C_m, and denoted by V_m. *Visual hull (VH)* V is defined as the intersection of the visual cones $V_1, \cdots, V_n$. (See Figure 1.)

$$V = \bigcap_{m=1}^{N} V_m \tag{1}$$

In the volume intersection method, when the number of cameras is increased, additional regions of the VH, which are regions not included in the object's region, are decreased. The additional regions mean those regions that are not included in the objects' regions but are included in the VH. Theoretically, more accurate VH should be obtained

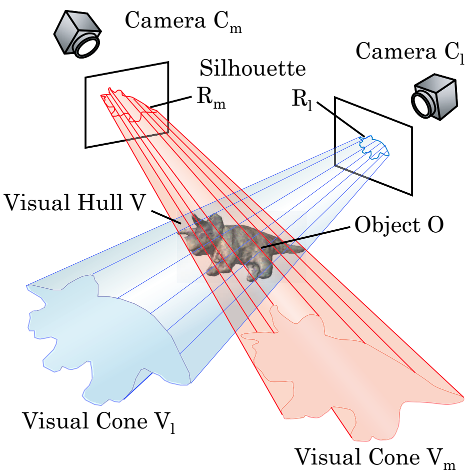


Figure 1. Volume intersection method.

by using more cameras. However, in real environments, increasing the number of cameras often causes an increase in the missing parts of the silhouettes. Since the visual cones are calculated from the silhouettes and the VH is calculated as the intersection of the visual cones, the missing parts of the silhouettes are accumulated in the VH. As the result, often a more accurate VH is not reconstructed in spite of the increase of the cameras.

3. Random Pattern Backgrounds

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 can not 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 [10], 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. The silhouettes extracted with the random pattern backgrounds have missing parts below a specific percentage, even for objects of unknown color.

Figure 3 illustrates the rate of missing parts of a silhou-

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE

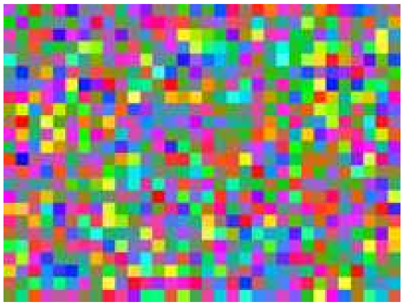


Figure 2. Sample of random pattern.

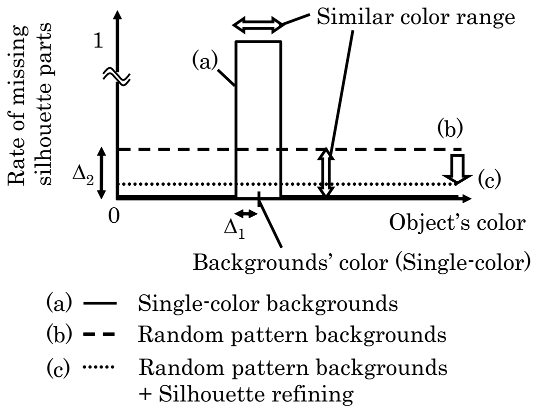


Figure 3. Relation between object's color and rate of missing silhouette parts for each kind of background.

ette 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 3, 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 3, the rate of missing parts would be dramatically high. With random pattern backgrounds, the silhouettes have missing parts below a specific rate Δ_2 shown as the broken line (b) in Figure 3. 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, 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.

3.2. Silhouette Missing Parts Rate

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.

$$\begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1.000 & 0 & 1.402 \\ 1.000 & -0.344 & -0.714 \\ 1.000 & 1.772 & 0 \end{pmatrix} \begin{pmatrix} \mathrm{Y} \\ \mathrm{U} \\ \mathrm{V} \end{pmatrix} \quad (2)$$

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 4(a).

For the object's color, shown as $\times$ in Figure 4(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 4(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:

$$p \leq \frac{4U_{th}V_{th}}{S}, \quad (3)$$

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. (3).

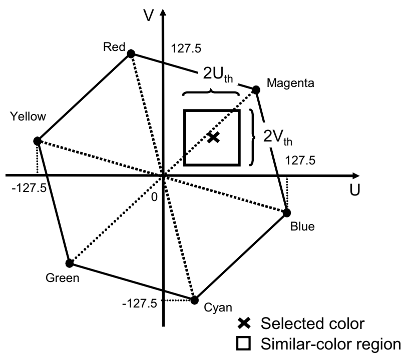
3.3. Generation of Random Patterns

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

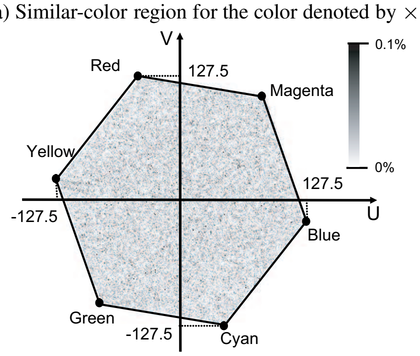
3.3.1 Selection of Color of Each region

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

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE



(a) Similar-color region for the color denoted by $\times$.



(b) Histgram of colors in random pattern.

Figure 4. 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 $\times$, the similar-color region is described as a rectangular region in the U-V space.

the brightness of an image, is 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 an 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 5. The boundaries of the hexahedron are defined by $R = 0$, $G = 0$, $B = 0$, $R = 255$, $G = 255$ and $B = 255$.

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 5. If the color is in the interior, we adopt the color. If not, go to 3.

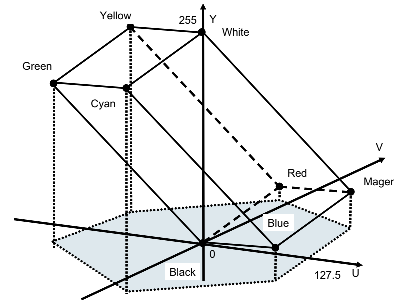


Figure 5. Selection of color for each region.

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) and $0 \leq R, G, B \leq 255$.

$$
\begin{aligned}
Y_s &= -1.402V_s && (4)\\
Y_s &= 0.344U_s + 0.714V_s && (5)\\
Y_s &= -1.772U_s && (6)\\
Y_s &= 255 - 1.402V_s && (7)\\
Y_s &= 255 + 0.344U_s + 0.714V_s && (8)\\
Y_s &= 255 - 1.772U_s && (9)
\end{aligned}
$$

3.3.2 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 [9]. 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.

4. Recovering Missing Silhouette Pixels

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 6. This means

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE

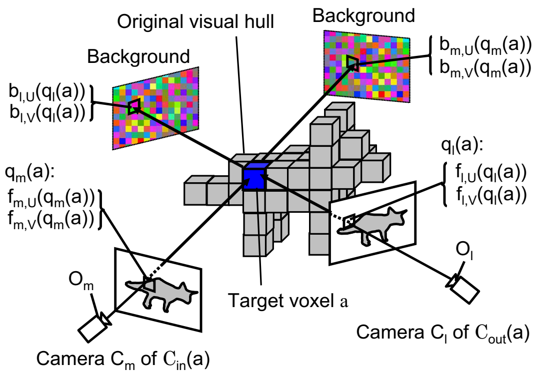


Figure 6. Object's color estimation.

that the silhouettes of the object can be extracted correctly in the 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 6. We denote U and V values of a pixel to which a is projected for the image of camera C_m by $f_{m,U}(q_m(a))$ and $f_{m,V}(q_m(a))$. For the background in the image C_m, we denote U and V values of a pixel to which a is projected by $b_{m,U}(q_m(a))$ and $b_{m,V}(q_m(a))$. The estimated U and V values of a, which are denoted by $Ave_U(a)$ and $Ave_U(a)$, are calculated as follows :

$$Ave_U(a) = \frac{\sum_{C_m \in \mathcal{C}_{in}(a)} f_{m,U}(q_m(a))}{N_{vis}(a)} \quad (10)$$

$$Ave_V(a) = \frac{\sum_{C_m \in \mathcal{C}_{in}(a)} f_{m,V}(q_m(a))}{N_{vis}(a)}. \quad (11)$$

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

$$\begin{aligned} &\mid Ave_U(a) - b_{l,U}(q_l(a)) \mid < U_{th} \\ &\quad or \\ &\mid Ave_V(a) - b_{l,V}(q_l(a)) \mid < V_{th} \end{aligned} \quad (12)$$

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.

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. 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.

5.1. Silhouette Refining

A sample image from a camera is shown as Figure 7(a). The silhouette extracted under the random pattern backgrounds is shown as white regions in Figure 7(b). As described in Section 3, the rate of missing silhouette parts is below a specific percentage due to our use of the random pattern backgrounds.

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.

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 9(a). From Eq. (2), 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.

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE

The silhouettes with missing parts recovered with our method are shown in Figure 7(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.

5.2. Shape after Refining

The correct shapes reconstructed from a set of the correct silhouettes, which were extracted manually, are shown in Figure 8(a). Compared with the VHs reconstructed from the original silhouettes (Figure 8(b) and each left figure of 9), the VHs recovered with our proposed method (Figures 8(c) and each right figure of 9) have fewer missing parts. The original VH of the triceratops (Figure 8(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 1 and 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 dramatically 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.

Table 1. Error voxels when using our proposed method. (Triceratops)

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

Table 2. Error voxels when using our proposed method. (Horse)

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

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 objects 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.

Acknowledgement

This work was partially supported by a JSPS Research Fellowship for Young Scientists (No.19-10459) and a Grant-in-Aid for Scientific Research (B) (No.16300055) of the JSPS.

References

[1] I. J. Cox, S. L. Hingorani, and S. B. Rao. A maximum likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.

[2] M. Iiyama, Y. Kameda, and M. Minoh. 4pi measurement system: A complete volume reonstruction system for freely-moving objects. In *MFI2003*, pages 119–124, 2003.

[3] K. Ikeuchi, A. Nakazawa, K. Hasegawa, and T. Ohishi. The great buddha project: Modeling cultural heritage for vr systems through observation. In *ISMAR03*, page 7, Nov 2003.

[4] T. Kanade, K. Oda, A. Yoshida, M. Tanaka, and H. 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.

[5] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *Proc. SIGGRAPH*, 21(4):163–169, 1987.

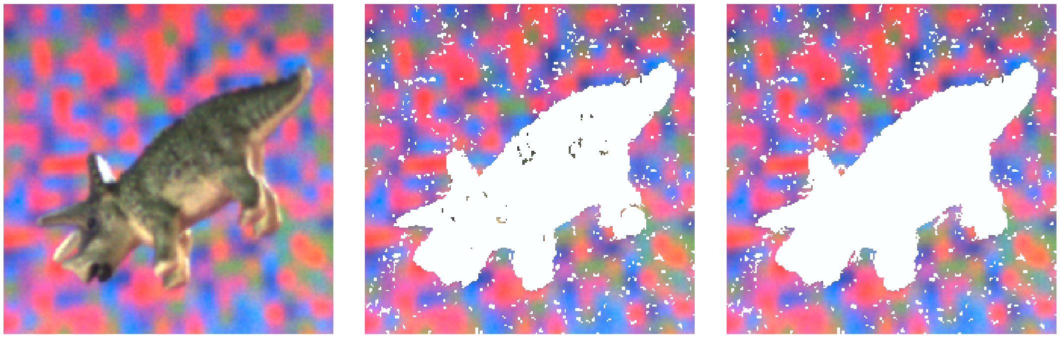
[6] W. N. Martin and J. K. Aggarwal. Volumetric description of objects from multiple views. *PAMI*, 5(2):150–158, 1983.

[7] T. Matsuyama, X. Wu, T. Takai, and S. Nobuhara. Real-time 3dshape reconstruction, dynamic 3d mesh deformation, and high fidelity visualization for 3d video. *CVIU*, 96(4):393–434, 2004.

[8] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand. Defocus video matting. *ACM Trans. Graph.*, 24(3):567–576, 2005.

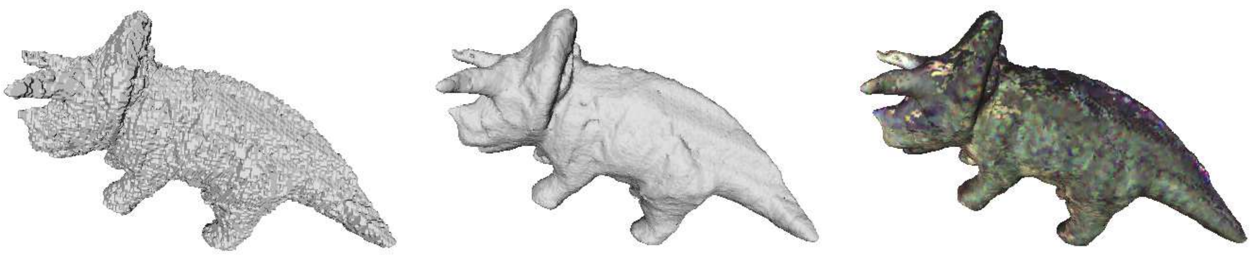
[9] J. D. Mulder and R. van Liere. Fast perception-based depth of field rendering. In *Proc. VRST*, pages 129–133, 2000.

[10] A. R. Smith and J. F. Blinn. Blue screen matting. In *Proc. SIGGRAPH*, pages 259–268, 1996.

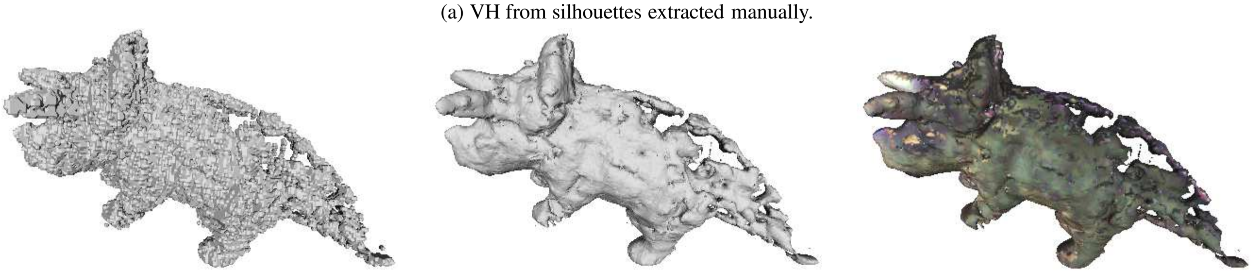


(a) Obtained image (b) Silhouette without refining (c) Silhouette with refining

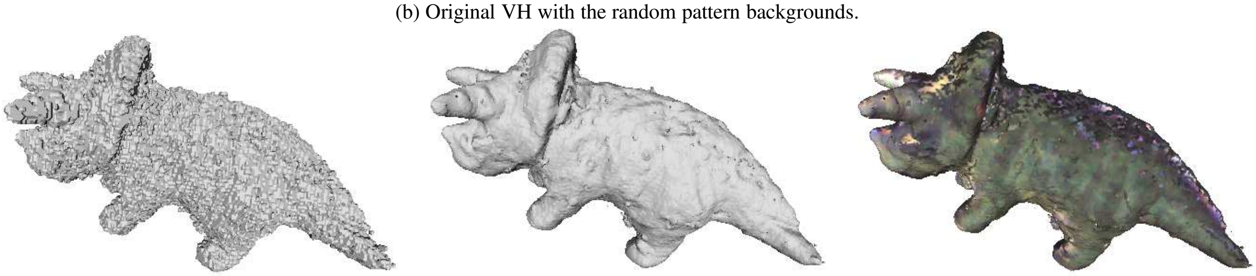
Figure 7. An example of silhouette refining with random pattern backgrounds. (Triceratops)



(a) VH from silhouettes extracted manually.



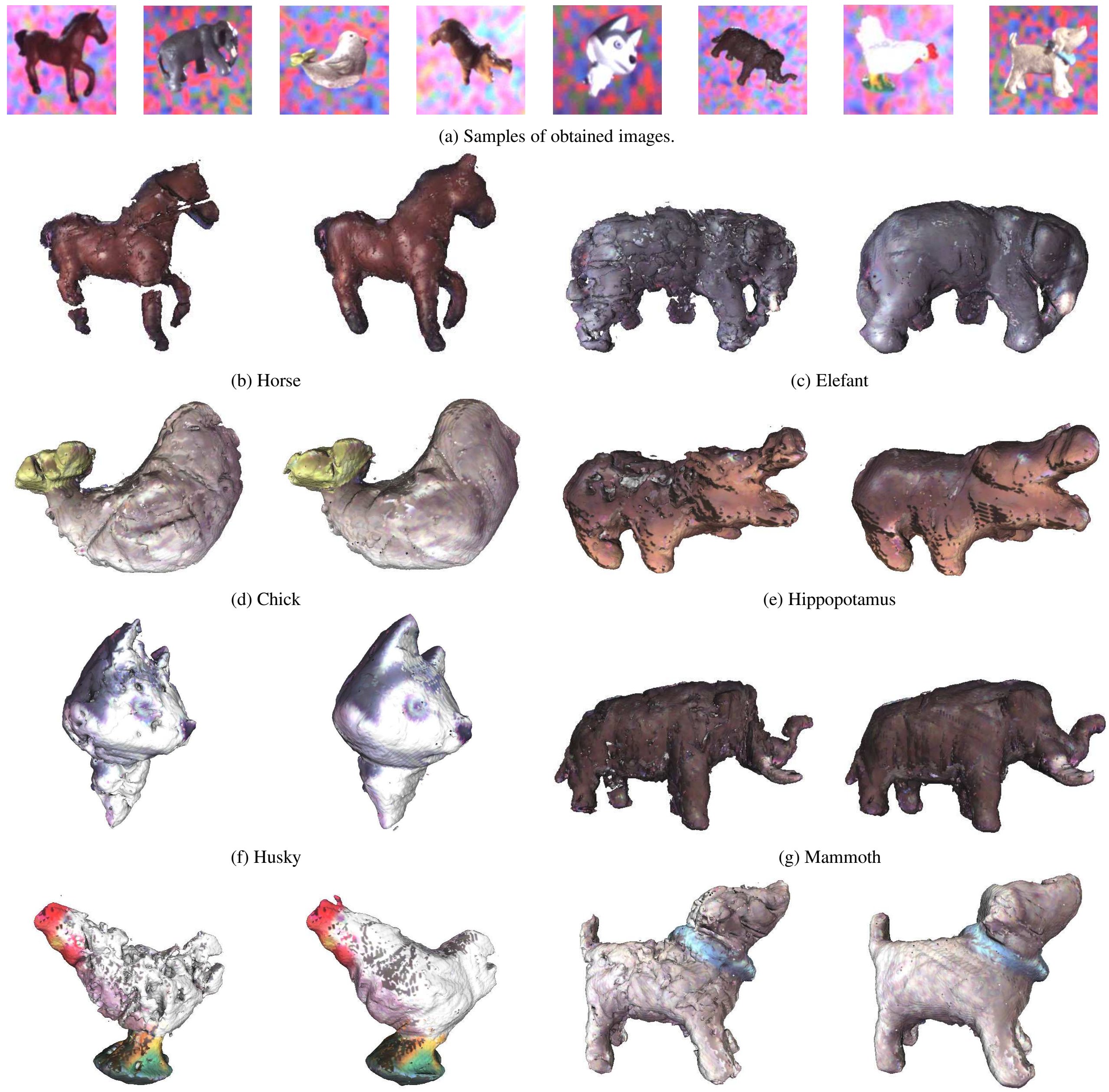
(b) Original VH with the random pattern backgrounds.



(c) Refined VH with out proposed method.

Figure 8. 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 can not be recovered in appearance. (Left) Each shape drawn with surface patches obtained by marching cube algorithm [5] for the resultant VH. (Center) Shapes of the left ones with surface smooting. (Right) Colored Shapes of the center ones with a Naive algorithm(a viewpoint independent patch-based method, [7]).

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE



(a) Samples of obtained images.

(b) Horse

(c) Elefant

(d) Chick

(e) Hippopotamus

(f) Husky

(g) Mammoth

(h) Hen

(i) Dog

Figure 9. Colored Shapes of several toys. (Left) Original VH with the random pattern backgrounds. (Right) Refined VH with our proposed silhoutte refining method.

Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)
0-7695-2939-4/07 $25.00 © 2007 IEEE