

# Reliability-Based 3D Reconstruction in Real Environment

Hansung Kim   Ryuuki Sakamoto   Itaru Kitahara   Tomoji Toriyama   Kiyoshi Kogure  
Knowledge Science Lab, ATR   Univ. of Tsukuba   Knowledge Science Lab, ATR  
Kyoto, Japan   Ibaraki, Japan   Kyoto, Japan  
+81-774-95-1401   +81-29-853-7818   +81-774-95-1401  
{hskim, skmt}@atr.jp   kitahara@computer.org   {toriyama, kogure}@atr.jp

## ABSTRACT

We present a practical 3D reconstruction method that guarantees robust visual hull construction in real environments where segmentation errors and occlusion exist. The proposed method consists of foreground extraction and reliability-based shape-from-silhouette, and they are connected by the intra-/inter-silhouette reliabilities. In foreground extraction, all regions are classified into four categories based on their intra-reliabilities. Then the reliability-based shape-from-silhouette technique reconstructs a visual hull by carving a 3D space based on the intra-/inter-silhouette reliabilities. The proposed method provides a reliable visual hull in real environments without much increment of the system complexity compared with conventional systems.

## Categories and Subject Descriptors

I.4.8 [Scene Analysis]

## General Terms

Algorithms, Reliability

## 1. INTRODUCTION

Since Kanade et al. proposed “Virtualized Reality” as a new visual medium for manipulating and rendering prerecorded scenes in a controlled environment [8], many computer vision-based 3D imaging systems have been developed [4][10]. They reconstruct 3D models from captured video streams and generate realistic free-view video of those objects from virtual cameras. A shape-from-silhouette (SFS) method is the most common way of converting silhouette contours into a visual hull [11][12]. However, these systems have several limitations for use in real environments as shown in Fig. 1.

First, most SFS methods assume perfect silhouette information of objects in advance or use simple background subtraction techniques in restricted environments such as blue screen or simple-colored backgrounds [4, 11, 12]. Chueng et al. and Ishikawa et al. included their own segmentation methods developed for the visual hull in real environment to their systems [1][7], but segmentations errors directly affect the visual hull in

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM07, September 23-29, 2007, Augsburg, Bavaria, Germany.  
Copyright 2007 ACM 978-1-59593-701-8/07/0009...\$5.00.

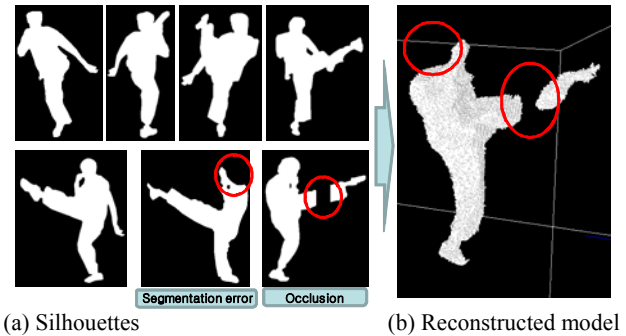


Figure 1. Shape-from-silhouette with outliers

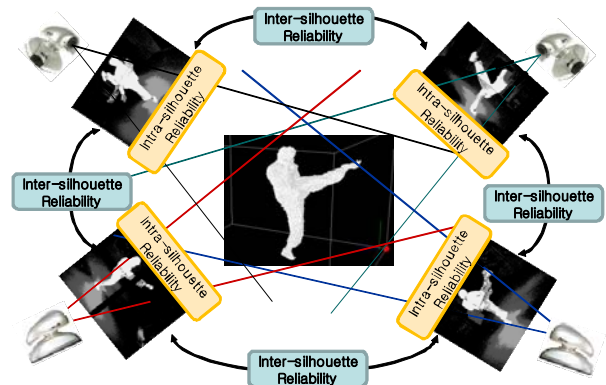


Figure 2. Reliability-based 3D reconstruction

these systems because the segmentation processes are absolutely independent of reconstruction processes. Grauman et al. proposed a Bayesian approach to compensate for modeling errors from false segmentation [3]. They modeled prior density using probabilistic principal components analysis and estimated a maximum-a-posteriori reconstruction of multi-view contours. This approach reconstructs good error-compensated models from erroneous silhouette information, but it needs prior knowledge about the objects and ground-truth training data.

Second, most previous methods assume that the modeling space is free of occlusion. Therefore, if the model is occluded by any background object in a certain view, conventional methods will produce an incomplete model by following “Liebig’s law of the minimum.” Guan et al. pointed out this problem and suggested making occlusion masks by observing boundaries of a moving object [5]. This method produces tight and correct visual hulls in the presence of partial occlusion, but it needs long observation of correct boundaries of objects to make occlusion masks.

The final goal of our research is to develop a practical 3D imaging system that can be used in daily life. Our proposed method's basic idea is to couple reconstructing 3D models with foreground segmentation in real environments. We propose a foreground segmentation technique with multiple thresholds and a reliability-based shape-from-silhouette (RSFS) method considering segmentation errors and partial occlusion. These two processes are connected by the intra-/inter-silhouette reliabilities as shown in Fig. 2. This paper has three main contributions to construct a visual hull in real environments. First, the reliability of the segmentation results is included in the silhouette masks. Though it requires one more bit per pixel in silhouette masks, this one bit compensates for carving errors from false segmentation or partial occlusion. Second, we propose an improved SFS technique to reconstruct the visual hull in the presence of segmentation errors and partial occlusion. Finally, we realize this system with simple reformation of conventional structure so that it does not much increase computational complexity compared with conventional ones.

## 2. SILHOUETTE EXTRACTION

The background subtraction technique is one of the most common approaches for extracting foreground objects from video sequences because it works very quickly and distinguishes semantic object regions from static backgrounds. The background extraction techniques can be classified into two categories: parametric and non-parametric approaches.

Parametric approaches set a form of background distribution in advance and estimate the parameters of the model. The Gaussian mixture model has been a representative approach [14]. Recently, Kim et al. used Laplace model because pixel variance in a static scene over time in indoor scenes taken with the latest camera is closer to Laplace distribution than Gaussian [9].

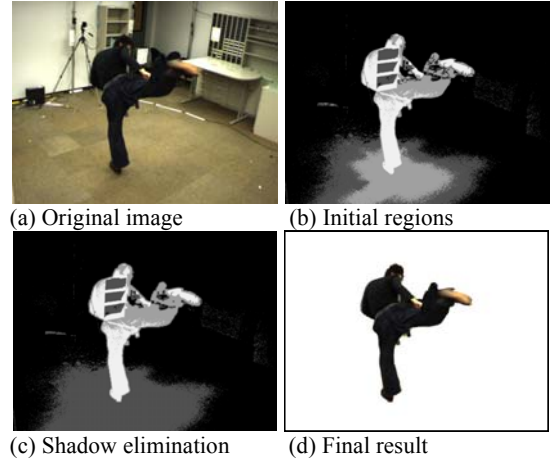
Nonparametric approaches directly estimate density functions from sample data. Elgammal et al. used Kernel Density Estimators (KDE) to quickly adapt to background changes [2], and several advanced approaches using KDE have also been proposed [6][13].

In this paper, we model the background based on the Kim's approach [9] and the silhouettes of foreground objects are extracted with multiple thresholds and morphological processes. Each pixel of the segmented regions is expressed with two bits: the first bit indicates the group of the pixel and the second bit represents the intra-reliability of the result, as shown in Table 1.

**Table 1. Categories of segmented masks**

Region	Code	Group	Intra-reliability
(a)	01	Background	Reliable
(b)	00	Background	Suspicious
(c)	10	Foreground	Suspicious
(d)	11	Foreground	Reliable

At first, initial region classification is performed by subtracting the luminance components of the current frame from the background model. We classify the initial object region into four categories using multiple thresholds based on background subtraction  $BD$ , as in Eq. (1).  $L_I$  and  $L_B$  indicate the luminance components of pixel  $p$  in the current frame and the background



**Figure 3. Segmentation results in each step**

model, respectively, and  $\sigma$  is a standard deviation of the background model.

$$BD(p) = |L_I(p) - L_B(p)| \quad (1)$$

$$\begin{cases} BD(p) < K_1\sigma(p) & \Rightarrow \text{(a) Reliable Background} \\ K_1\sigma(p) \leq BD(p) \leq K_2\sigma(p) & \Rightarrow \text{(b) Suspicious Background} \\ K_2\sigma(p) \leq BD(p) \leq K_3\sigma(p) & \Rightarrow \text{(c) Suspicious Foreground} \\ K_3\sigma(p) \leq BD(p) & \Rightarrow \text{(d) Reliable Foreground} \end{cases}$$

Thresholds  $K_1 \sim K_3$  used in Eq. (1) are determined by training data with the following condition, where  $\beta$  was empirically set to 3 because false positive errors are generally more critical than false negative errors in foreground extraction:

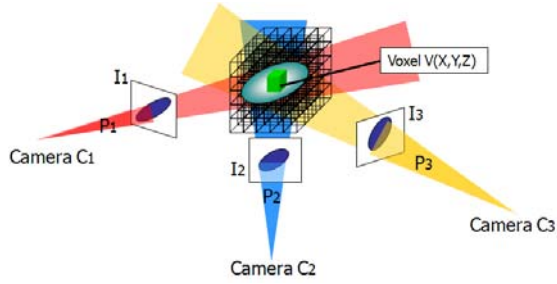
$$(K_1, K_2, K_3) = \arg \min_{K_1, K_2, K_3} \left( \begin{array}{l} \beta \times \text{False Positive Error} \\ + \text{False Negative Error} \end{array} \right) \quad (2)$$

However, a large amount of background can be assimilated in the suspicious foreground region in the results, caused by an object's shadow that changes background brightness. We eliminate the shadows from the initial object by using a color component because the shadow changes the color property of the background much less than the luminance. With Eq. (3), shadows in *Suspicious foreground* region (c) are merged into *Suspicious background* region (b).  $H$  indicates the color components of the images, and  $\sigma_H$  is a standard deviation of the color component in the background model.

$$\text{if } (p \in \text{region (c)} \ \& \ |H_I(p) - H_B(p)| < K_1\sigma_H(p)) \quad (3) \\ \text{then } p \Rightarrow \text{region (b)}$$

Finally, small regions whose sizes are smaller than a threshold  $Th_{RG}$  are eliminated, and the final silhouette is extracted using a silhouette extraction technique [9] to smooth the foreground boundaries and eliminate holes inside the regions. This technique denotes the final foreground region by wrapping the object with four drapes with elasticity. Newly covered background regions by silhouette extraction are changed to *Suspicious foreground* regions.

Figure 3 shows a test image and the results of silhouette extraction in each step. In the results, the black, dark gray, light



**Figure 4. 3D reconstruction by shape-from-silhouette**

gray and white regions indicate *Reliable background*, *Suspicious background*, *Suspicious foreground* and *Reliable foreground*, respectively.

### 3. 3D RECONSTRUCTION

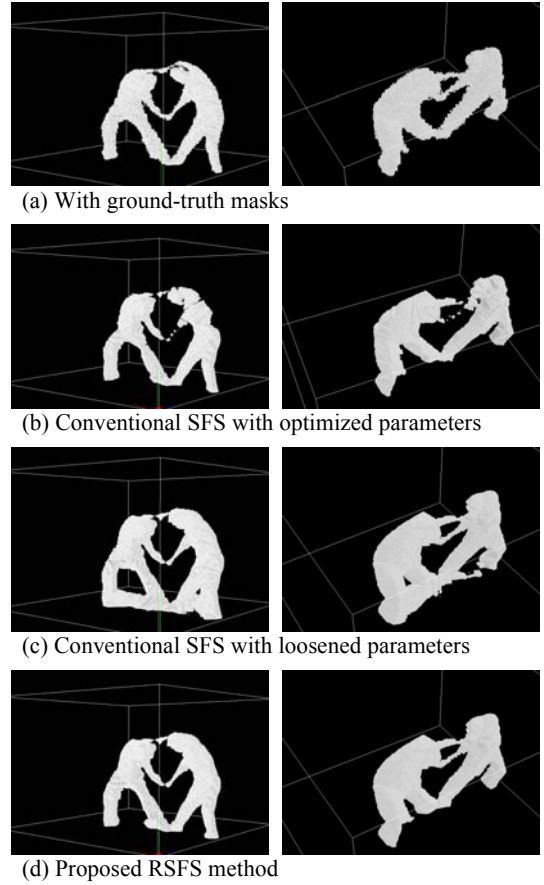
The SFS technique carves 3D space by projecting silhouettes from each view into the space using projection matrixes obtained by camera calibration. The visual hull constructed using SFS provides an upper bound on the object shape. Figure 4 illustrates the relation of the multiple cameras and the voxels that are set up in the 3D space. Each camera is labeled  $C_n$ , with projection matrix  $P_n$ , a silhouette with  $C_n$  as  $I_n$  ( $n=1, \dots, N$ ), and each voxel as  $V(X, Y, Z)$ .

Now, for example, assume that  $I_p$  is a subset of  $I_n$  and  $v_p$  is a projected point of  $V$  on  $I_p$  by  $P_p$ . If the 3D position of  $V$  is inside the 3D object,  $V$  must be projected onto the foreground regions of all images  $I_p$ . Therefore, if any single projected point is located in the background region in images  $I_p$ , voxel  $V$  is carved from the 3D shape model. As a result, we can estimate the entire 3D object's shape by examining every possible position of voxels  $V(X, Y, Z)$ .

However, this SFS method has a serious limitation when used in a real environment because this technique is too sensitive against segmentation errors and partial occlusion, as pointed out in Section 1. Therefore, we propose a RSFS algorithm that reduces the effect of outliers based on the reliability of the silhouette. The key concept of the RSFS algorithm is that we have to consider both intra- and inter-reliabilities of the silhouettes which represent the reliability of the segmentation itself and the reliability from the relationship to the other silhouettes, respectively. In the SFS process, a single segmentation error classified into the background with low intra-reliability can damage the visual hull, disregarding all correct information in other images. On the other hand, an occluder in front of the target object at the view of a certain camera has a high possibility to be a *Reliable background* region because it is unaffected by the foreground object. However, we must exercise caution when determining if it is a partial occluder or a real hole in the space. Therefore, we establish a set of rules to distinguish outliers in silhouette masks.

**Segmentation error:** if  $v_p$  is in a *Suspicious background* region and all other  $N-1$  projected points  $v_n$  ( $n \neq p$ ) are in the foreground regions of each image plane, then  $v_p$  is an outlier due to segmentation error.

**Partial occlusion:** if  $v_p$  is in a *Reliable background* region and both projected points in the images of neighbor cameras are in the *Reliable foreground* regions, then  $v_p$  is an outlier due to partial occlusion.



**Figure 5. Reconstructed visual hulls**

If the projected point of a voxel in the silhouette is judged as a segmentation error or a partially occluded point, the point is excluded in computing the visual hull. Guan et al. proved that the constructed visual hull after discarding outliers still satisfies the constraint of conservation of the visual hull [5].

In realization of the RSFS algorithm, we adopted an octree data structure [15] because testing all the points in a modeling space is very time-consuming and create excessive data. This structure dramatically reduces the carving speed and the amount of data.

### 4. EXPERIMENTAL RESULTS

We implemented a distributed system using eight calibrated IEEE 1394 cameras. The 3D space was modeled at a resolution of  $550 \times 550 \times 250$  on a  $1\text{cm} \times 1\text{cm} \times 1\text{cm}$  voxel grid, and the cameras were mounted on the walls and ceiling to surround the space. We constructed the visual hull from the video streams using the proposed method and compared it with the ground-truth data and the results from a conventional method. Figure 5 (a) shows snapshots of visual hull with manually segmented ground-truth masks, and Figs. 5 (b)–(d) show the same results generated by a conventional SFS method without reliability and by the proposed RSFS algorithm. Figure 5 (b) is the results with optimized segmentation parameters, but the visual hull was damaged by segmentation errors in certain views. Therefore, we loosened the segmentation parameters to compensate for cracks, but it made

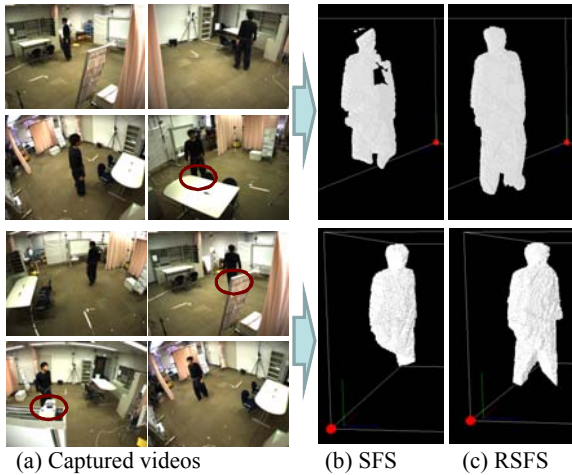


Figure 6. Results with partial occlusion



Figure 7. Snapshots of generated free-viewpoint videos

the visual hull coarser as we can see in Fig. 5 (c). On the other hand, the visual hull generated using the RSFS algorithm based on reliability looks much closer to the real 3D model.

Then, we tested the proposed algorithm in an environment where several partial occluders exist. In Fig. 6, a foreground model is partially occluded by a table, a rack, or a whiteboard in certain views so that the legs of the model were cut off in the results by a conventional method. In the first test set, the result was even affected by segmentation errors. However, the RSFS method constructed more natural models by compensating for the outliers, as we can see in Fig. 6 (c). In the second test set, the legs are not restored perfectly because their lower part is occluded by two different occluders. We restricted the maximum number of partial occluders to one per voxel in our system because the voxels recovered from more than two occluders may become too coarse due to a lack of information; moreover it makes the other part of the visual hull rough and increases computational complexity.

Figure 7 shows snapshots of the free-viewpoint videos rendered from the visual hulls. Natural scenes from any point of view can be rendered from the proposed algorithm.

## 5. CONCLUSIONS

We presented a practical method to construct the visual hull from multiple videos in a real environment. The main idea of the proposed method couples reconstructing 3D models with object

segmentation to compensate for outliers from defective segmentation and occlusions. We provided an improved space carving technique to construct a visual hull by considering the intra-/inter-silhouette reliabilities. We proved that the proposed RSFS technique constructs a compact visual hull even in the presence of segmentation errors and occlusions.

## 6. ACKNOWLEDGMENTS

This research was supported by the National Institute of Information and Communications Technology.

## 7. REFERENCES

- [1] Cheung, G., Kanade, T., Bouguet, J.Y., and Holler, M. A real time system for robust 3d voxel reconstruction of human motions. *Proc. CVPR* (2000), 714-720.
- [2] Elgammal, A. et al.. Non-parametric model for background subtraction. *Proc. ECCV* (2000), 751-767.
- [3] Grauman, K., Shakhnarovich, G., and Darrell, T. A Bayesian Approach to Image-Based Visual Hull Reconstruction. *Proc. CVPR* (2003), 187-194.
- [4] Gross, M. et al. Blue-c: A spatially immersive display and 3D video portal for telepresence. *Proc. SIGGRAPH* (2003), 819-827.
- [5] Guan, L., Sinha, S., Franco, J.S., and Pollefeys, M. Visual Hull Construction in the Presence of Partial Occlusion. *Proc. 3DPVT* (2006).
- [6] Han, B., Comaniciu, D., and Davis, L. Sequential kernel density approximation through mode propagation: applications to background modeling. *Proc. ACCV* (2004).
- [7] Ishikawa, T., Yamazawa, K., and Yokoya, N. Real-time generation of novel views of a dynamic scene using morphing and visual hull. *Proc. ICIP* (2005), 1013-1016.
- [8] Kanade, T., Rander, P. W., and Narayanan, P. J. Virtualized Reality: Constructing Virtual Worlds from Real Scenes. *IEEE Multimedia*, 4, 1 (1997), 34-47.
- [9] Kim, H., Sakamoto, R., Kitahara, I., Toriyama, T., and Kogure, K. Robust Foreground Segmentation from Color Video Sequences Using Background Subtraction with Multiple Thresholds. *Proc. KJPR* (2006), 188-193.
- [10] Magnor, M.A. *Video-Based Rendering*, A K Peters, 2005.
- [11] Matsuyama, T., Wu, X., Takai, T., and Wada, T. Real-Time Dynamic 3-D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3-D Video. *IEEE Trans. CSVT*, 14, 3 (2004), 357-369.
- [12] Matusik, W., Buehler, C., Raskar, R., Gortler, S., and McMillan, L. Image-based visual hulls. *Proc. SIGGRAPH* (2000), 369-374.
- [13] Mittal, A., and Paragios, N. Motion-based background subtraction using adaptive kernel density estimation. *Proc. CVPR* (2004), 302-309.
- [14] Stauffer, C., and Grimson, W.E.L. Adaptive background mixture models for real-time tracking. *Proc. CVPR* (1999), 246-252.
- [15] Szeliski, R. Rapid octree construction from image sequences. *CVGIP: Image Understanding*, 58, 1 (1993), 23-32.