3D Reconstruction from Multiple Videos in Real Environments

H. Kim R. Sakamoto I. Kitahara T. Toriyama K. Kogure *

Summary. We propose an advanced visual hull technique to compensate for outliers using the reliabilities of the silhouettes. The proposed method consists of a background subtraction technique using multiple thresholds and a reliability-based shape-from-silhouette algorithm. They are connected by the intra-/inter-silhouette reliabilities to compensate for carving errors from defective segmentation or partial occlusion which may occur in real environments. In final view rendering, we use a cinematographic camera control system and an ARToolkit to control virtual cameras.

1 Introduction

We have introduced " Cinematized Reality, " which aimed to record unexpected moments and create movie-like footage by a virtual camera generated from multiple cameras [2]. However, this system still has problems for applications to real environments because it is based on a visual hull technique that carves 3D space with silhouette information. In real scenes, the silhouette includes outliers from defective segmentation and occlusions, and these outliers directly affect the visual hull as shown in Fig. 1. Therefore we propose an advanced visual hull technique to compensate for outliers using reliabilities of the silhouettes.

2 3D Reconstruction

We extract silhouette information from video sequences using the background subtraction technique, which is one of the most common approaches. In the proposed approach, we include the reliability of the segmentation results in the silhouette masks. We classify all regions 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 model, respectively, and σ is standard deviation of the background model. The final silhouette is extracted from the categorized regions using morphological processes [3]:

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

$$\begin{cases} BD(p) < K_1\sigma(p) \\ \Rightarrow (a) Reliable Background \\ K_1\sigma(p) \leq BD(p) \leq k_2\sigma(p) \\ \Rightarrow (b) Suspicious Background \\ K_2\sigma(p) \leq BD(p) \leq k_3\sigma(p) \\ \Rightarrow (c) Suspicious Foreground \\ K_3\sigma(p) \leq BD(p) \\ \Rightarrow (d) Reliable Foreground \end{cases}$$

Then the visual hull is constructed by projecting silhouettes from each view into the space using projection matrixes obtained by camera calibration. To avoid damages from outliers, we consider both the intra- and inter-reliabilities of the silhouettes that represent the reliability of the segmentation itself and the reliability from the relationship to the other silhouettes, respectively. We establish a set of rules to distinguish outliers in silhouette masks:

Segmentation error: if the projected point of a certain view is in a suspicious background region and all other projected points are in the foreground regions of each image plane, then the

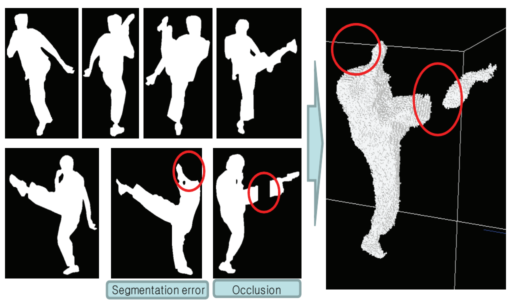

Figure 1. Visual hull with outliers

* Hansung Kim, Ryuuki Sakamoto, Tomoji Toriyama and Kiyoshi Kogure, Knowledge Science Lab, ATR, Itaru Kitahara, Dept. of Intelligent Interaction Technologies, Univ. of Tsukuba

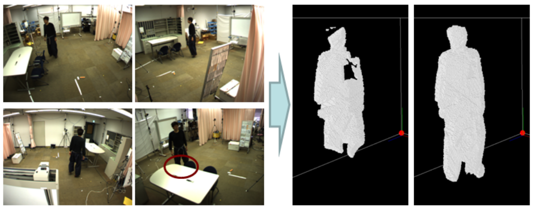

Figure 2. 3D reconstruction with occlusions

point is an outlier due to segmentation error.

Partial occlusion: if the projected point of a certain view is in a reliable background region and both projected points in the images of neighbor cameras are in the reliable foreground regions, then the point is an outlier due to partial occlusion.

If the projected point of a voxel in the silhouette is judged to be segmentation error or a partially occluded point, the point is excluded in computing the visual hull.

In Fig.2, a foreground model is partially occluded by a table in certain views so that the legs of the model were cut off in the results generated from a conventional method. In the test set, the result was affected even by segmentation errors. However, the proposed method constructed more natural models by compensating for the outliers.

3 Virtual View Rendering

In the proposed 3D video system, we provide two methods to generate a final video according to the purpose of rendering. ARToolkit provides very easy real-time interface for free-view rendering [1]. Viewers can control a fiducial marker and a small USB camera as if they are a miniature set of the 3D space and a real camera to shoot the scene, respectively. In order to change a voxel-based 3D model to mesh model, we use a marching cube algorithm [4]. Figure 3 shows scenes that a user is controlling a USB camera and the generated video from the designated camera position. Although we can generate free-viewpoint video by controlling a real camera, the sway of the camera leads to viewers becoming confused and the final video will shows unstable and naive scene. The second method is off-line rendering using a cinematographic virtual camera control system that helps the user to generate final videos from the 3D model by referring to the database of experts ’ camera works [2]. Many kinds of camera shots and principal camera actions are stored in the system as expertise. Therefore, even non-experts can easily convert the 3D video to attractive movies with the help of the system ’s expertise. Figure 4 shows an example of a generated 3D video with camera controls in the Cinematized Reality system.

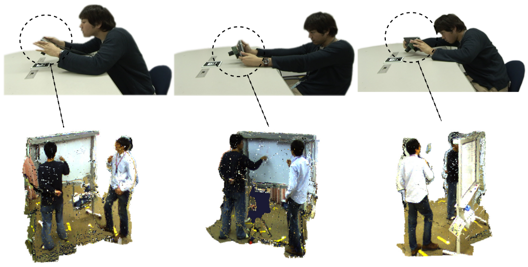

Figure 3. Real-time rendering

Figure 4. Rendering with camera works

4 Conclusion

We proposed a 3D construction technique using multiple videos in a real environment where segmentation errors and occlusion exist. Then, we provided two methods for generating final 3D videos by user ’s interaction. Future work will include improving reconstruction speed. We are now implementing the algorithm on GPU for acceleration.

Acknowledgement

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

References

[1] http://www.hitl.washington.edu/artoolkit/.

[2] H. Kim, R. Sakamoto, I. Kitahara, and K. Kogure. Cinematized Reality: Cinematographic 3D Video System for Daily Life Using Multiple Outer/Inner Cameras. In *Proc. 3DCINE workshop in CVPR 2006*, p. 168, 2006.

[3] H. Kim, R. Sakamoto, I. Kitahara, T. Toriyama, and K. Kogure. Robust Silhouette Extraction Technique Using Background Subtraction. In *Proc. MIRU 2007*, pp. 552–557, 2007.

[4] W. E. Lorensen and H. E. Cline. Marching Cubes: A high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.