4DCap: Multi-View Dynamic 3D Object Reconstruction System

Wen-Chao Chen, Fu-Jen Hsiao, Chih-Jen Teng, De-Jin Shau

Industrial Technology Research Institute
Chaody@itri.org.tw, Hsiaofujen@itri.org.tw, Dalice@itri.org.tw, Djshau@itri.org.tw

Abstract

The 4DCap system captures dynamic objects from different viewpoints with eight synchronized high quality cameras. Dynamic 3D textured models are reconstructed from multi-view images by advanced image-based 3D modeling algorithm. The 4DCap system allows a user to visualize dynamic objects from free viewpoints by applying augmented reality techniques and to fit all kinds of needs for 3D imaging applications.

1. Introduction

1.1. System architecture

We implemented a multi-view capturing system, which consists of eight synchronized IEEE 1394b color cameras connecting with eight PCs as shown in Figure 1. Those cameras are calibrated in advance and captures video sequence in 1024x768 resolution at 30 fps. For the purpose of background subtraction, the studio was setup as a blue screen-like environment in 3m x 6m space with height of 2m and cameras are mounted on the ceiling around the space. Figure 2 shows captured images in our studio.

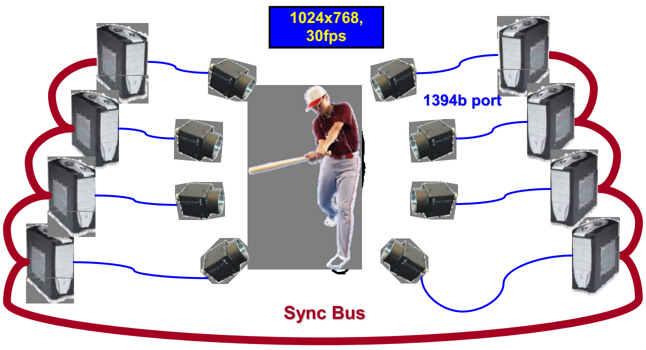


Figure 1. System architecture

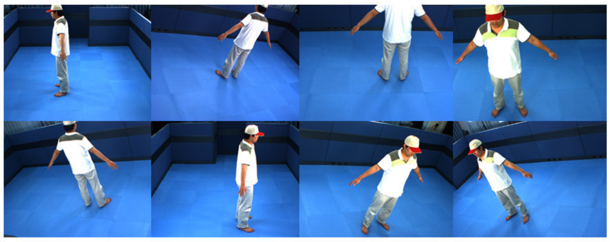


Figure 2. Captured images

1.2. System flow chart

Figure 3 depicts the overall organization of the 4DCap system. First, the preprocessing step comprises color calibration and geometrical calibration to acquire camera parameters. After capturing synchronized video sequence, the system applies post-processing which consists of background subtraction, shape-from-silhouette, Poisson Surface Reconstruction [1] and multi-view texturing, to obtain textured 3D models off-line. Finally, the developed AR authoring tool can adjust the size and position of reconstructed 3D models relative to 2D barcodes to perform on the ARToolKit[2] correctly.

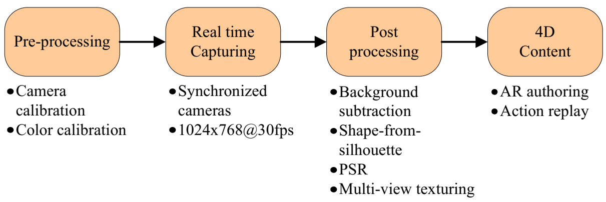


Figure 3. System flow chart

2. Proposed Methods Overview

This section describes the approaches used in 4DCap system to reconstruct textured 3D models and render dynamic models on ARToolKit.

2.1. Background subtraction

Although objects are captured in a blue screen-like environment, foreground extraction is still tough due to shadows casted by the object itself. We implement

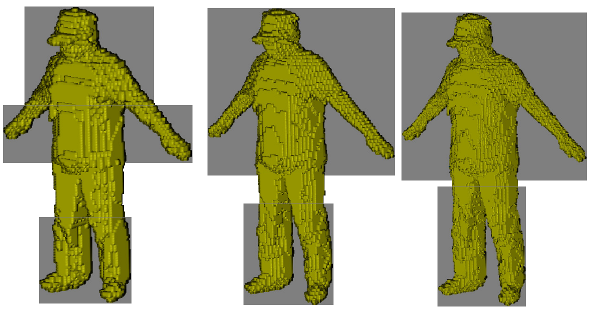
ICAT 2008
Dec. 1-3, Yokohama, Japan
ISSN: 1345-1278

combined method proposed by Zhang[3] and Horprasert[4] with threshold assigned manually. The threshold does not change in the same video clip, in other words, the threshold is only assigned once per sequence to obtain better background subtraction results for shape-from-silhouette algorithm.

2.2. Shape-from-silhouette algorithm

The conventional shape-from-silhouette (SFS) algorithms convert silhouette information into a visual hull by subdividing the initial voxel iteratively at consecutive levels, such as Kim's work[7]. The resultant models at different subdivision levels contain quite different compositions, so the system performance, in terms of model accuracy, memory space and construction time, changes widely with the subdivision level number. Since the big system performance gap is not desirable for the practical application, new construction methods must have a finer control over the system performance.

Our system implements methods proposed by Chen[8, 9] to reconstruct high quality voxel-based visual hulls on CPU only. Figure 4. shows the constructed results at different size of minimum voxel and Table 1. shows the computation time in different minimum voxel size. This result is running at 2.2-GHz with Intel Core 2 processor and 2GB main system memory. It is apparent from Table 1 that construction time required by SFS algorithm can be reduced significantly, especially when high accurate models are required. The GPU-based version of SFS algorithm will be implemented in the near future.



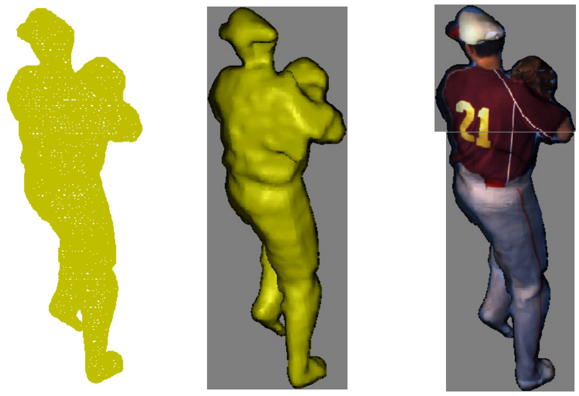
(a) (b) (c)

Figure 4. Reconstructed results with minimum voxel size of (a)1.56cm, (b)0.78cm, (c)0.39cm.

Table 1. Minimum voxel size vs. computation time

Min. voxel size (cm^3)	Computation time (ms)	Number of vertices
1.56	218	75,872
0.78	469	332,592
0.39	532	1,471,096

When the reconstructed visual hull is ready, the Poisson Surface Reconstruction software is utilized to convert all vertices which form the surface of the visual hull into a triangulated mesh model, as shown in Figure 5. (a) and (b).



(a) (b) (c)

Figure 5. (a) triangulated mesh model (b) shaded model (c) Multi-view textured model

2.3. GPU-based multi-view texturing

Before performing multi-view texturing, occlusion problem should be solved first. Our system acquires depth buffer from GPU by rendering triangulated mesh in eight views. The depth buffer is served as visibility information for texture mapping.

2.4. Rendering with ARToolKit

To visualize reconstructed 3D models in a more friendly way, we adopt ARToolKit for building augmented reality as our rendering environment. We also develop an authoring tool to adjust the size, position and orientation of reconstructed 3D models easily. Dynamic motions are performed by changing 3D models reconstructed from different sets of images captured from multiple views, as shown in Figure 6.

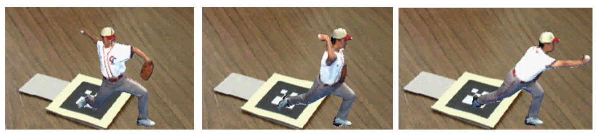


Figure 6. Dynamic ARToolKit rendering results

3. Demonstration Description

The 4DCap system presents a hybrid approach from multi-view capturing to 3D model reconstruction and combines 3D content production with novel augmented reality techniques. The capturing system can extent to marker-less motion capture device for sports education and posture analysis in unlimited viewing angle. This system also shows that dynamic 3D modeling is one of the more practical ways of free-view video.

Figure 7. shows the setup of proposed demonstration. Through the interaction with audience, we would like to exchange ideas with experts in different fields and enable the possibility of 3DTV broadcasting in the near future.

Figure 7. Demonstration setup

Figure 8. Interaction with audience

References

[1] M. Kazhdan, M. Bolitho, H. Hoppe. Poisson Surface Reconstruction. Eurographics Symposium on Geometry Processing, 2006.

[2] H. Kato, "ARToolKit." [Online]. Available: *http://www.hitl.washington.edu/artoolkit/*

[3] R. Zhang, S. Zhang, and S. Yu. Moving Objects Detection Method Based on Brightness Distortion and Chromaticity Distortion. IEEE Transactions on Consumer Electronics, Vol. 53, No. 3, 2007.

[4] T. Horprasert, D. Harwood, and L. S. Davis. A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection. Proc. ICCV 1999.

[5] H. L. Chou and Z. Chen. Fast octree construction endowed with an error bound controlled subdivision scheme. Journal of Information Science and Engineering, vol. 22, no. 3, pp. 641-657, 2006.

[6] Z. Chen, H. L. Chou, and W. C. Chen. A Performance Controllable Octree Construction Method. Proc. ICPR, 2008.

[7] H. Kim, R. Sakamoto, I. Kitahara, N. Orman, T. Toriyama, and K. Kogure. Compensated Visual Hull for defective Segmentation and Occlusion. Proc. ICAT, 2007.

[8] J. Carranza, C. Theobalt, M. Magnor, and H.-P. Seidel. Free-view point video of human actors. ACM Trans. Graph., 22(3):569–577, 2003.

[9] E. Aguiar, C. Theobalt, M. Magnor, H. Theisel, H. Seidel. M^3: Marker-Free Model Reconstruction and Motion Tracking from 3D Voxel Data. Proc. of the Computer Graphics and Applications, 2004.

[10] MPII, "Kung-Fu Girl sequence." [Online]. Available: *http://www.mpi-inf.mpg.de/departments/irg3/kungfu/index.html*

[11] W. Lee. "Camera Models, WONWOO's AUGMENTED REALITY WEB PAGE." [Online]. Available: *http://old.uvr.gist.ac.kr/wlee/web/techReports/ar/Camera%20Models.html*