Real-Time Markerless Human Body Tracking Using Colored Voxels and 3-D Blobs

Fabrice Caillette & Toby Howard
Advanced Interfaces Group, Department of Computer Science
University of Manchester, Manchester M13 9PL, UK
fabrice@cs.man.ac.uk

Abstract

This paper presents a robust method for real-time visual human body tracking. We perform a hierarchical 3-D reconstruction from multiple camera views as a basis for tracking. Blobs attached to a kinematic model are then used to reliably track individual body parts with both volume and color information. We describe how the blob-model is dynamically adjusted to accommodate different body configurations. In tests, our system has proved robust in presence of noisy data and self-occlusions.

1 Introduction

Human body tracking has a wide range of applications in human-computer interfaces and virtual environments. The interactive control of an avatar is particularly important for telepresence and motion capture. Associated with movement recognition it can be used for surveillance or to drive virtual interfaces. In the context of Augmented Reality, displaying information on tracked body parts is useful, for example, for virtual training or re-education applications.

The work presented here addresses the problem of recovering posture parameters from video images of a scene as observed by multiple cameras. The multi-view aspect is justified by an improved accuracy at a very moderate extra cost. In order to be practical, the tracking system has to operate in an unaugmented environment (no markers), in real-time on a single PC. This immediately raises a number of challenges and drives most of our design decisions.

Instead of working directly with 2-D camera images, our system first reconstructs a 3-D voxel-based representation of the person, and then performs the matching of a kinematic model in this 3-D space. This approach brings more consistency to the tracking step which is then working in real-world coordinates, and dismisses most of the noise which is not coherent across multiple camera views.

A common 3-D reconstruction technique is the visual-hull [6] which defines the volume of interest as the intersection of the 3-D projections of its silhouettes. There has been a very recent interest in using the visual-hull as a basis for tracking: Some attempts of matching a kinematic model onto the reconstructed volume gave promising results [4, 5] but remained incapable of handling self-occlusions. Using 2-D feature tracking along with the visual-hull [7] was a step toward the inclusion of color. Hierarchical reconstruction and fitting schemes have also been explored [1], but real-time performance was still out of reach.

Our work extends the statistical scene description found in the Pfinder System [8] into 3-D space, with the addition of a kinematic model replacing the Kalman filter for blob dynamics. Blobs are statistical entities modeling coherent attributes and their variations. Novel aspects of our approach include a statistical reconstruction method which incorporates color information, and the use of blobs with dynamic re-configuration for robust real-time tracking.

2 Description of the System

Our reconstruction process is similar to that used by Cheung *et al.* [3] in the sense that voxels are classified from pixel-samples taken inside their 2-D projections onto the surface planes. However, in order to attain real-time performance, we do not perform a prior binary segmentation on the input images, but instead compute a measure of the distance to the background model for each 2-D sample. Voxels are then classified from statistics on these distances across the available views, being eventually either discarded, subdivided or retained as belonging to the foreground. This 3-D reconstruction process is hierarchical because each voxel can be subdivided recursively, but no static data-structure (such as an octree) is maintained. Instead voxels are created dynamically during the recursive reconstruction process, minimizing both processing time and memory footprint and allowing the zone of tracking to be moved and scaled around the expected position of the subject at no cost.

Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR 2004)
0-7695-2191-6/04 $20.00 © 2004 IEEE

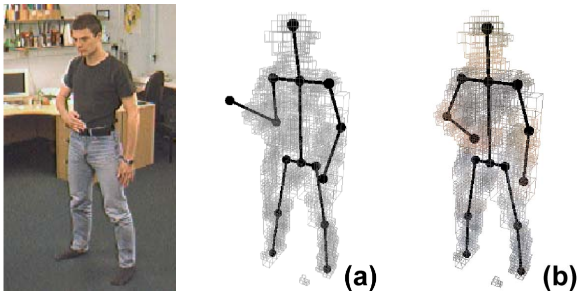

Figure 1. Model fitting on the voxels (a) using only volume information and (b) adding color.

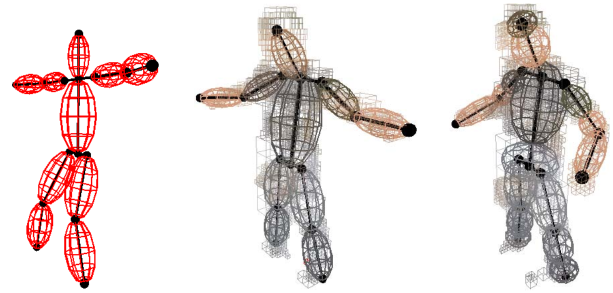

Figure 2. Automatic blobs reconfiguration from a naive model (left) to a better one (right).

Another important aspect of our reconstruction scheme is that color information is kept in each voxel, allowing more robust tracking especially in cases of self occlusions and ambiguous poses. For example, in Figure 1 the pose of the right forearm cannot be recovered from the reconstructed volume only, but using colour helps in finding the correct pose. Voxels contain the mean color of the samples for each camera-view: We do not perform any visibility test at this point to decide the real color of voxels since the subsequent tracking scheme automatically selects the color nearest to the model. More details on the reconstruction scheme can be found in an earlier work [2].

The tracking step fits a kinematic model on the 3-D voxels using Expectation-Maximization (EM) with blobs attached to the bones of the model. The voxels are first assigned to the nearest blob using the Mahalanobis distance on both position and color, and the blobs attributes are reevaluated during the maximization. The new positions of the blobs trackers are then used to define goal positions for the joints of the model. Because of incomplete data and occlusions, some goals can be missing or be misplaced. We use inverse kinematics to robustly recover the model pose from the available goals. If a body part is totally occluded (no voxels assigned to it), the corresponding blobs are temporarily deactivated until it becomes visible again.

The initial distribution of the blobs onto the model bones is rarely optimal due to different body shapes and clothing. Between each iteration of the tracking loop, the blobs are regenerated to match best both the model and the voxel data. Depending on their mixed shape/color variance, blobs are split or merged so as to find the best compromise between specialization and processing speed, resulting in a final distribution best matching the clothing of the subject. Figure 2 (left to right) illustrates this process over 20 frames.

3 Results and Conclusion

The system was tested on a single $2\,\mathrm{GHz}$ PC, with four IEEE 1394 low-resolution webcams. The combined acquisition, reconstruction and tracking ran at $15\,\mathrm{fps}$, which is sufficient for most interactive applications. The tracking process proved to be robust to noisy data and self occlusions thanks to the EM statistical framework and the use of color (Figure 1). With a maximal precision of $1.5\,\mathrm{cm}$, our system can capture a broad range of body movements, and the small jitter due to the camera noise can easily be filtered.

The main contributions of this paper are the real-time aspect of the reconstruction process, and the use of color for robust tracking in presence of self-occlusions. The automatic reconfiguration of the blobs also constitutes a decisive advantage over other methods. Future work will focus motion prediction using autoregressive processes, and the inclusion of learned soft constraints to infer plausible poses, even in presence of severe occlusions.

References

[1] E. Borovikov and L. Davis. 3D shape estimation on density driven model fitting. In *3DPVT'02*, page 116.

[2] F. Caillette and T. Howard. Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction. In *Proceedings of BMVC*, Sept. 2004.

[3] K. M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3D voxel reconstruction of human motions. In *CVPR'00, V2, P714*.

[4] J. P. Luck, D. E. Small, and C. Q. Little. Real-time tracking of articulated human models using a 3D shape-from-silhouette method. In *CVPRIP'2002*, 2002.

[5] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *IJCV*, 53(3), July/August 2003.

[6] R. Szeliski. Real-time octree generation from rotating objects. Technical Report 90/12, Digital Equipment Corporation, Cambridge Research Lab, 1990.

[7] C. Theobalt, M. Magnor, P. Schueler, and H. P. Seidel. Combining 2D feature tracking and volume reconstruction for online video-based human motion capture. In *Proc. of Pacific Graphics, Beijing*, pages 96–103, 2002.

[8] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, 1997.

0-7695-2191-6/04 $20.00 © 2004 **IEEE**