Model-Based Human Body Tracking

Lily Lee

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: To understand human activity in a close-up video monitored environment, we need detailed information about the pose of the body and the angles formed by the joints, and relationship of one human body to the environment and other people in the scene. I will address the problem of tracking the motion of human body segments from a single stationary video source.

Motivation: A computer vision system that could visually track the motion of a human body would be a key enabling technology for high-level interpretation of human motion. For example, in order to monitor human activities, an observer may want a symbolic description of the human actions. Similarly, in physical training contexts, a coach may be interested in the position and orientation of particular body segments. Both situation would benefit from an automatic, visually-based system that can recover the joint angles of a human body.

Previous Work: There have been many approaches to solve the human body tracking problem. They differ mainly by whether the recovered motion description is 2D or 3D, and whether there is an explicit model of the human body.

Polana[4] obtained frequency information from a video sequence and classified periodic activities using the frequency signature. Niyogi and Adelson[6] detected people walking in the plane of the image by looking for the braided pattern formed by the knees over time. Bobick and Davis[2] developed a template based action recognition system, the motion history image(MHI), which records the recency of activity at each pixel, with each action having its own MHI template. Morris and Rehg[7] built a 2D stick figure kinematic model of the body for tracking the motion of the body through single source video. Bregler[1] employed a mathematical result familiar to roboticists, the twist/screw motion of kinematic chain, and related it to image gradients to solve for differential motion of the body joints. Leventon and Freeman[3] collected statistical data of the likelihood of body configurations and used them to recover body joint angles.

Approach: The approach taken to the problem of human body tracking is to fit a 3D articulated human model to a single video source. The human body is well suited for modeling by rigid 3D parts that are connected in a kinematic chain. This representation allows for out of plane rotation, and facilitates texture mapping onto the model for synthesis of new body configurations. Self-occlusion of body parts—which can be difficult to handle in a 2D model—can be easily explained and predicted using the velocity information from joints in a 3D model of a human body.

Allowing for the overall translation, rotation, and scaling of the entire body, I have built a human body model with a total of 28 degrees of freedom(DOF). Each DOF of a joint is represented by a fixed axis, and rotation about each axis is constrained to a certain range. The jointed motion of body segments are represented as twist/screw motion about the fixed axis. Givein this representation, the projection of the 3D model can be used as an image to compre to the video sequence. In addition, continuity constriant and texture mapping of the 3D model can be used in the tracking process.

Difficulty: It is difficult to identify the location and size of a human, and initialize the tracking process from an unknown video source. However, assuming that we know the location of moving objects through background subtraction (see Figure 1). Whether or not a moving object is that of a human can be decided using a system such as[9]. I will assume that the location and size of a person is known. The initial pose and joint angles can be found with a precomputed library of signature based on images of quantized body poses.

Research Support: This research is supported in part by DARPA under ONR grant N00014-97-0363 and N00014-95-1-0600.

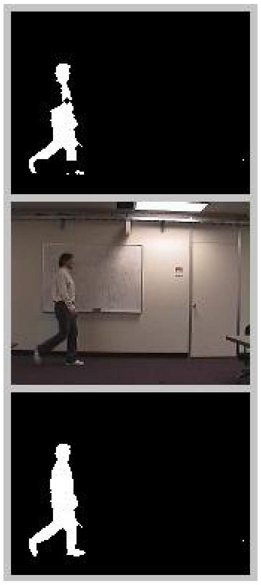

Figure 1: Top: The result of segmenting the human body from the stationary background is not perfect, as this example illustrates. Middle: This is the frame to which the body tracking system will fit the body model. Bottom: This is an example of a projection of the human model fitted to a frame.

References:

[1] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *CVPR*, 1998.

[2] J. Davis and A. Bobick. The representation and recognition of action using temporal templates. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 928–934, 1997.

[3] M. Leventon and W. Freeman. Bayesian estimation of 3-d human motion from an image sequence. Technical Report TR-98-06, Mitsubishi Electric Research Laboratory, Cambridge, MA, July 1998.

[4] R. Polana. *Temporal Texture and Activity Recognition*. PhD thesis, Department pf Computer Science, University of Rochester, 1994.

[5] J.M. Rehg. *Visual Analysis of High DOF Articulated Objects with Application to Hand Tracking*. PhD thesis, School of Computer Science, Carnegie Mellon University, 1995.

[6] S. Niyogi and E. Adelson. Analysing and recognizing walking figures in xyt. In *Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 469–474, 1994.

[7] D. Morris and J Rehg. Singularity analysis for articulated object tracking. In *CVPR*, 1998.

[8] H. Rowley and J. Rehg. Analyzing articulated motion using expectation-maximization. In *CVPR*, pages 935–941, 1997.

[9] C. Papageorgiou and T. Poggio. Trainable Pedestrian Detection. In *Proceedings of International Conference on Image Processing (ICIP'99)*, Kobe, Japan, October 1999.