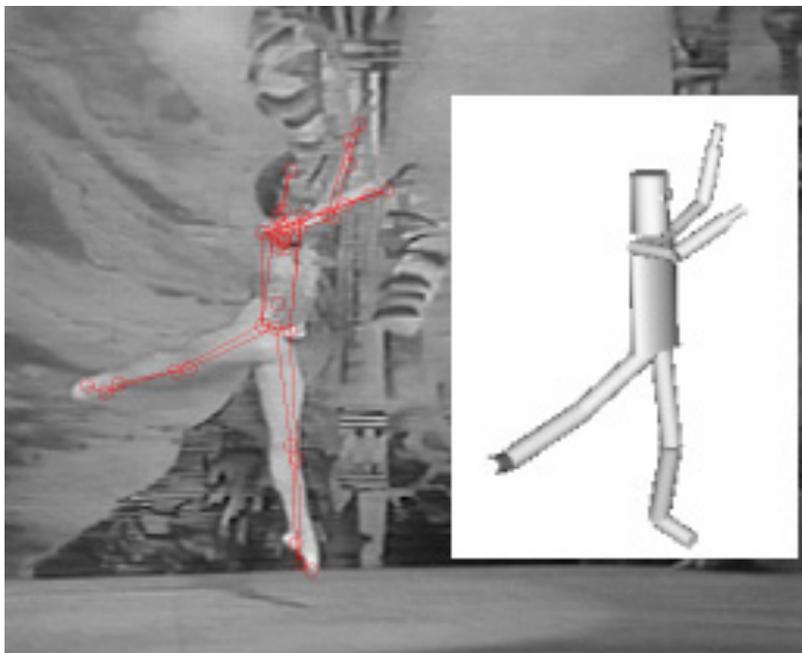


# Body Tracking from Single-Camera Video



We have developed a method to reconstruct the 3-d positions of a moving human figure from observing the figure's motions over time, recorded from a single video camera. This may have application to human-computer interaction, computer graphics, or interactive virtual environments.

We built an interactive tracking system to process real video sequences, and can achieve reasonable 3-d reconstructions of the human figure motion for various sequences.

Reconstructing 3-d from 2-d (image) information is an under-determined problem, and we must rely of prior knowledge about how people tend to move in order to resolve ambiguities. We learned a model of people move, using examples of 3-d human motion data obtained from a multiple-camera "motion capture" system. We used Bayesian methods to

**Background and objectives:** As one watches a film or video of a person moving, one can easily estimate the 3-dimensional motions of the moving person from watching the 2-d projected images over time. A dancer could repeat the motions depicted in the film. Yet such 3-d motion is hard for a computer to estimate. Such estimation is the goal of this work.

**Technical discussion:** Our approach is to use strong prior knowledge about how humans move. We show that this prior knowledge dramatically improves the 3d reconstructions. We learn our prior model from examples of 3-d human motion.

We first studied the 3-d reconstruction in a simplified image rendering domain where a Bayesian analysis provides analytic solutions to fundamental questions about estimating figural motion from image data. Using insights from the simplified domain, we applied our Bayesian method to real images and reconstruct human figure motions from archival video. Our system accomodates interactive correction of automated 2-d tracking errors, which allows reconstruction even from difficult film sequences.

We represent human figure motion as a linear combination of short snippets of the training examples. We use Singular Value Decomposition to obtain the optimal 50 dimensional linear model, given the training data.

Bayes rule provides two terms which multiply together to obtain the optimal 3-d reconstruction. The "likelihood" term constrains the 2-d projection of the 3-d model to match the image data. We require that the image data under the projection of a given limb part have an approximately constant intensity over time.

The "prior" term constrains the reconstructed 3-d model to be a probably 3-d human

motion. We model human motions as a high-dimensional Gaussian distribution.

We use standard optimization techniques to find the 3-d reconstruction which makes the optimal trade-off between fidelity to the image data, and high prior probability of occurring, given the training data.

Those results show how to estimate the 3-d figure motion if we can place a 2-d stick figure over the image of the moving person. We developed such a tracker, allowing interactive correction of tracking mistakes, to test our 3-d recovery method. We show good recovery of 3-d motion for a difficult dance sequence, viewed from a single camera. These results show the power of adding prior knowledge about human motions, in a Bayesian framework, to the problem of interpreting images of people.

*Bayesian reconstruction of 3D human motion from single-camera video*

Nicholas R. Howe, M. E. Leventon and W. T. Freeman,

in Adv. in Neural Information Processing Systems 12 (NIPS), edited by S. A. Solla, T. K. Leen, and K-R. Muller.

Available as [MERL-TR99-37](#)

*Bayesian estimation of 3-d human motion*

M. E. Leventon and W. T. Freeman

Available as [MERL-TR98-06](#)