Understanding Purposeful Human Motion

Christopher R. Wren, Alex P. Pentland
MIT Media Laboratory Vision and Modeling Group
Cambridge Massachusetts USA

Abstract

Human motion can be understood on several levels. The most basic level is the notion that humans are collections of things that have predictable visual appearance. Next is the notion that humans exist in a physical universe, as a consequence of this, a large part of human motion can be modeled and predicted with the laws of physics. Finally there is the notion that humans utilize muscles to actively shape purposeful motion. We employ a recursive framework for real-time, 3-D tracking of human motion that enables pixel-level, probabilistic processes to take advantage of the contextual knowledge encoded in the higher-level models, including models of dynamic constraints on human motion. We will show that models of purposeful action arise naturally from this framework, and further, that those models can be used to improve the perception of human motion. Results are shown that demonstrate both qualitative and quantitative gains in tracking performance.
***Keywords*— Human Motion Capture, Behaviour Interpretation, Interactive Visual Interfaces, Physics-based Modeling of Human Motion**

I. Introduction

This paper describes a real-time, fully-dynamic, 3-D person tracking system that is able to tolerate full (temporary) occlusions and whose performance is substantially unaffected by the presence of multiple people. The system is driven by 2-D *blob features* observed in two or more cameras [1;2] and by behavior models that estimate control signals. These features and controls are then probabilistically integrated into a fully-dynamic 3-D skeletal model, which in turn drives the 2-D feature tracking process by setting appropriate prior probabilities. The intrinsic state of the skeletal model is also used by the behavior module to choose the appropriate control strategy.

The feedback between 3-D model and 2-D image features is a recursive filter, similar to an extended Kalman. One unusual aspect of our approach is that the filter directly couples raw pixel measurements with an articulated dynamic model of the human skeleton. Previous attempts at person tracking have utilized a generic set of image features (e.g., edges, optical flow) that were computed as a preprocessing step, without consideration of the task to be accomplished. In this aspect our system is similar to that of Dickmanns in automobile control [3], and our previous research shows that we obtain similar advantages in efficiency and stability though this direct coupling.

We will show how this framework can go beyond passive physics of the body by incorporating various patterns of control (which we call 'behaviors') that are *learned* from observing humans while they perform various tasks. Behaviors are defined as those aspects of the motion that cannot be explained by passive physics alone. In the untrained tracker these manifest as significant structure in the innovations process (the sequence of prediction errors). Learned models of this structure can be used to recognize and predict this purposeful aspect of human motion.

This paper will briefly discuss the formulation of our 3-D skeletal model in Section II.A, followed by an explanation of how to drive that model from 2-D probabilistic measurements, and how 2-D observations and feedback relate to that model in Section II.B. Section II.C explains the behavior system and its intimate relationship with the physical model. Finally, we will report on experiments showing an increase in 3-D tracking accuracy, insensitivity to temporary occlusion, and the ability to handle multiple people in Section IV.

A. Related Work

In recent years there has been much interest in tracking the human body using 3-D models with kinematic and dynamic constraints. Perhaps the first efforts at body tracking were by Badler and O'Rourke 1980, followed by Hogg 1988 [4;5]. These early efforts used edge information to drive a kinematic model of the human body. These systems require fairly precise hand initialization, and can not handle the full range of common body motion.

Following this early work using kinematic models, some researchers began using dynamic constraints to track the human body. Pentland and Horowitz 1991 employed non-rigid finite element models driven by optical flow [6], and Metaxas and Terzopolous's 1993 system employing deformable superquadrics [7;8] driven by 3-D point and 2-D edge measurements. Again, these systems required precise initialization and could handle a limited range of body motion.

More recently, several authors have applied variations on the basic kinematic analysis-synthesis approach method to the body tracking problem [9;10;11]. Gavrila and Davis [12] and Rehg and Kanade [13], have demonstrated that this

C. Wren and A. Pentland conducted this research at the MIT Media Laboratory Vision and Modeling Group; Room E15-384; 77 Massachusetts Avenue; Cambridge MA 02139 USA. E-mail: cwren@media.mit.edu, sandy@media.mit.edu. For funding information please refer to http://www.media.mit.edu/Sponsors/

approach has the potential to deal with limited occlusions, and thus to handle a greater range of body motions.

The work described in this paper attempts to combine the the dynamic modeling work with the advantages of a recursive approach, by use of a formulation related to the extended Kalman filter that couples a fully dynamic skeletal model with observations of raw pixel values, as modeled by probabilistic 'blob' models.

This system also attempts to incorporate learned patterns of control into the body model. The approach we take is based on the behavior modeling framework introduced in Pentland and Liu 1995 [14]; it is also related to the behavior modeling work of Blake 1996 [15] and Bregler 1997 [16]. However, the controller described here operates on a 3-D non-linear model of human motion that is closer to true body dynamics than 2-D linear models.

II. Mathematical Framework

The human body is a complex dynamic system, whose visual features are time-varying, noisy signals. Accurately tracking the state of such a system requires use of a recursive estimation framework, as illustrated in figure 1. The elements of the framework are the observation model relating noisy pixel-level features to the higher-level skeletal model and *vice versa*, the dynamic skeletal model, and a model of typical behaviors. We will first describe the dynamic and observation models, and then the behavior model.

A. Dynamics

There are a wide variety of ways to model physical systems. The model needs to include parameters that describe the *links* that compose the system, as well as information about the hard *constraints* that connect these links to one another. A model that only includes this information is called a *kinematic* model, and can only describe the static states of a system. The state vector of a kinematic model consists of the model state, $\mathbf{q}$, and the model parameters, $\mathbf{p}$.

A system in motion is more completely modeled when the *dynamics* of the system are modeled as well. A dynamic model describes the state evolution of the system over time. In a dynamic model the state vector includes velocity as well as position: $\mathbf{q}, \dot{\mathbf{q}}; \mathbf{p}$). And state evolves according to Newton's First Law:

$$\ddot{\mathbf{q}} = \mathbf{W} \cdot \mathbf{Q} \tag{1}$$

Where $\mathbf{Q}$ is the vector of external forces applied to the system, and $\mathbf{W}$ is the inverse of the system mass matrix. The mass matrix describes the distribution of mass in the system.

A. Hard Constraints

Hard constraints represent absolute limitations imposed on the system. One example of a kinematic constraint is a skeletal joint. Our model follows the *virtual work* formulation [17]. In a virtual work formulation, all the links in a model have full range of unconstrained motion. Hard kinematic constraints on the system are enforced by a special set of forces $\mathbf{c}$:

$$\ddot{\mathbf{q}} = \mathbf{W} \cdot (\mathbf{Q} + \mathbf{c}(\mathbf{q}, t)) \tag{2}$$

The formulas governing these constraints can be modified at run-time.

It is essential that the constraint forces do not add energy to the system. It can be shown that this requirement is satisfied if they are constructed so they lie in the null space complement of the constraint Jacobian:

$$\mathbf{c}(\mathbf{q}, t) = \boldsymbol{\lambda} \frac{\partial \dot{\mathbf{c}}}{\partial \mathbf{q}} \tag{3}$$

Combining that equation with the definition of the constraints results in a linear system of equations with only the one unknown, $\boldsymbol{\lambda}$:

$$-\left[\frac{\partial \mathbf{c}}{\partial \mathbf{q}}^T \mathbf{W} \frac{\partial \mathbf{c}}{\partial \mathbf{q}}\right] \boldsymbol{\lambda} = \frac{\partial \mathbf{c}}{\partial \mathbf{q}}^T \mathbf{W}\mathbf{Q} + \frac{\partial \dot{\mathbf{c}}}{\partial \mathbf{q}}\dot{\mathbf{q}} + \frac{\partial^2 \mathbf{c}}{\partial t^2} \tag{4}$$

This equation can be rewritten to emphasize its linear nature. $\mathbf{J}$ is the constraint Jacobian, $\boldsymbol{\rho}$ is a known constant vector, and $\boldsymbol{\lambda}$ is the vector of unknown Lagrange multipliers:

$$-\mathbf{J}^T \mathbf{W} \mathbf{J} \boldsymbol{\lambda} = \boldsymbol{\rho} \tag{5}$$

Many fast, stable methods exist for solving equations of this form.

B. Soft Constraints

Some constraints are probabilistic in nature. Noisy image measurements are a constraint of this sort, they influence the dynamic model but do not impose hard constraints on its behavior.

Soft constraints such as these can be expressed as a potential field acting on the dynamic system. The incorporation of a potential field function that models a probability density pushes the dynamic evolution of the model toward the most likely value, starting from the current model state.

Note that functions that take the model state as input, such as a the controller from Section II.C, can be represented as a time-varying potential field. One relevant example is incorporation of a probability distribution over link position and velocity:

$$\mathbf{Q}_f = f(\mathbf{X}, \mathbf{q}, \dot{\mathbf{q}}) \tag{6}$$

B. The Observation Model

The low-level features extracted from video comprise the final element of our system. Our system tracks regions that are visually similar, and spatially coherent: blobs. We can represent these 2-D regions by their low-order statistics. Clusters of 2-D points have 2-D spatial means and covariance matrices, which we shall denote $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The blob

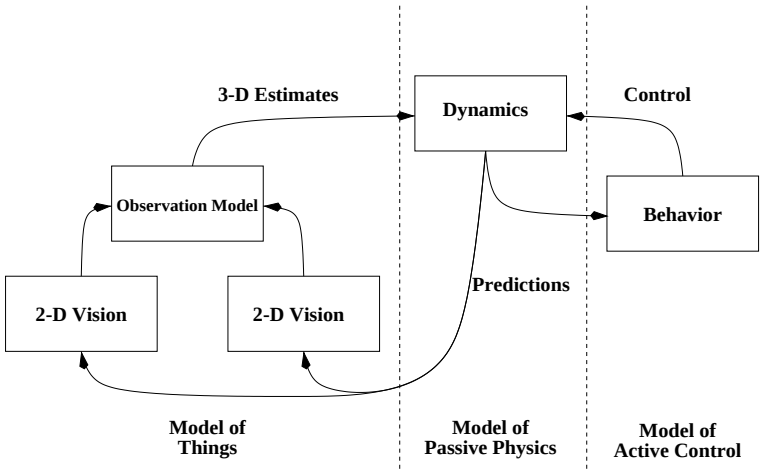

Fig. 1. The flow of information though the system. Predictive feedback from the 3-D dynamic model becomes prior knowledge for the 2-D observations process. Predicted control allows for more sensible predictive feedback.

Fig. 2. **Left:** video and 2-D blobs from one camera in the stereo pair. **Right:** corresponding configurations of the dynamic model.

spatial statistics are described in terms of their second-order properties; for computational convenience we will interpret this as a Gaussian model:

$$\Pr(\mathbf{O}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\exp(-\frac{1}{2}(\mathbf{O}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{O}-\boldsymbol{\mu}_k))}{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \quad (7)$$

The Gaussian interpretation is not terribly significant, because we also keep a pixel-by-pixel *support map* showing the actual occupancy [18].

These 2-D features are the input to the 3-D blob estimation equation used by Azarbayejani and Pentland [1]. This observation equation relates the 2-D distribution of pixel values to a tracked object's 3-D position and orientation.

These observations supply constraints on the underlying 3-D human model. Due to their statistical nature, observations are easily modeled as soft constraints. Observations are integrated into the dynamic evolution of the system by modeling them as descriptions of potential fields, as discussed in Section II.A.2.

A. The Inverse Observation Model

In the open-loop system, the vision system uses a Maximum Likelihood (ML) framework to label individual pixels in the scene:

$$\Gamma_{ij} = \arg\max_k [\Pr(\mathbf{O}_{ij}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)] \quad (8)$$

where Γ_{ij} is the labeling of pixel (i, j), and $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ are the second-order statistics of model k.

To close the loop, we need to incorporate information from the 3-D model. Given the current state of the model $\mathbf{q}$, it is possible to compute the state of an individual link that matches a specific tracked feature (say the hand), and call it $\mathbf{v}$. Then, given a model of the camera, it is possible to calculate the perspective projection of that state into 2-D and call it $\mathbf{v}^*$.

Since the vision system uses a stochastic framework, it is necessary to represent this link projection as a statistical model: $\Pr(\mathbf{O}_{ij}|\mathbf{v}_k^*)$. Integrating this information into the 2-D statistical decision framework results in a Maximum *A Posteriori* decision rule:

$$\Gamma_{ij} = \arg\max_k [\Pr(\mathbf{O}_{ij}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \cdot \Pr(\mathbf{O}_{ij}|\mathbf{v}_k^*)] \quad (9)$$

C. Models of Purposeful Motion

Observations of the human body reveal an interplay between the passive evolution of a physical system (the human body) and the influences of a an active, complex controller (the nervous system). Section II.A explains how, with a bit of work, it is possible to model the physical aspects of the system. However, it is *very* difficult to explicitly model the human nervous system, so the approach of using observed data to estimate probability distributions over control space is very appealing.

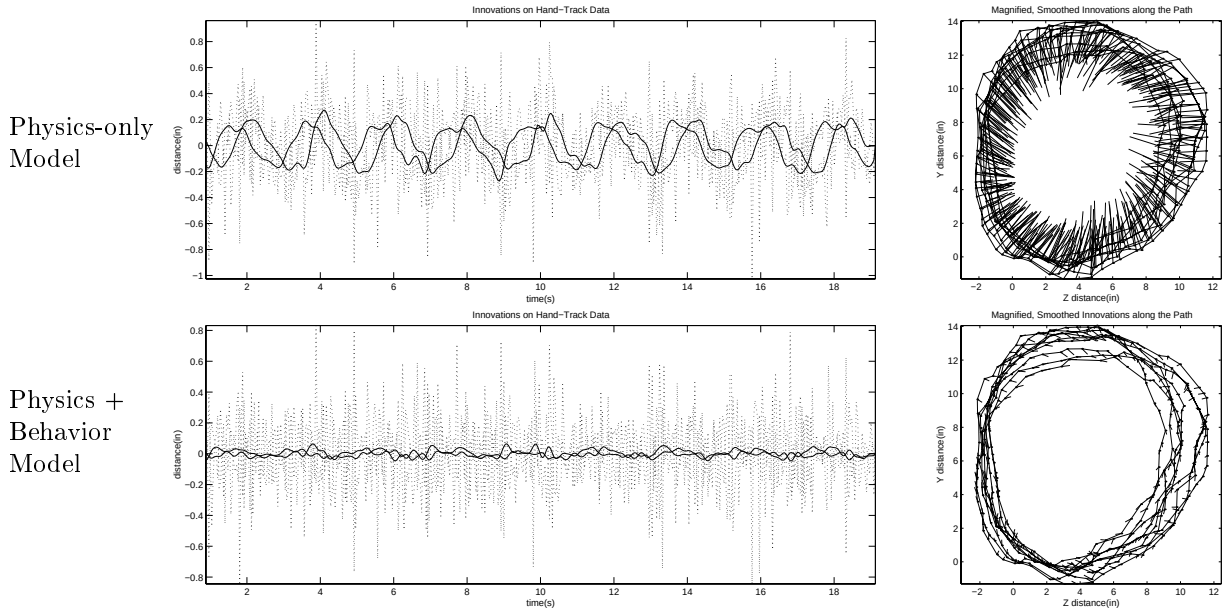

Fig. 3. Modeling tracking data of circular hand motion. Passive physics alone leaves significant structure in the innovations process. **Top Left:** Smoothing the innovations reveals unexplained structure. **Top Right:** Plotting the Innovations along the path makes the purposeful aspect of the action clear. **Bottom:** In this example, using a learned control model to improve predictions leaves only white process noise in the innovations process. The smoothed innovations stay near zero.

A. A Model for Control

Kalman filtering includes the concept of an *innovations process*. This is the difference between the actual observation and the predicted observation transformed by the Kalman gain:

$$\boldsymbol{\nu}_t = \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\Phi}_t\hat{\mathbf{x}}_{t-1}) \tag{10}$$

The innovations process $\boldsymbol{\nu}$ is the sequence of information in the observations that was not adequately predicted by the model. According to control theory, if we have a sufficient model of the dynamic process and the observation model, and white, zero-mean Gaussian noise is added to the system, both in the observation stream and into the real dynamic system itself, then the innovations process will be zero-mean and white. Inadequate models will manifest as correlations in the innovations process.

As described above, we have significant models of the observed human in terms of appearance, perspective, and passive physics. The most significant unmodeled aspect of human motion is active control. Purposeful human motion includes significant structure due to the active control of nerves and muscles that is not well modeled by the passive physics, perspective or appearance modules. We should expect this significant, unmodeled structure to color the innovations process.

A simple example is helpful for illustrating this idea. If we track the hand moving in a circular motion, then we have a sequence of observations of hand position. This sequence is the result of a physical thing being measured by a noisy observation process. Assuming that the hand moves according to a linear, constant dynamic model, it is possible to estimate the true state of the hand, and predict future states and observations. If this model is sufficient, then the errors in the predictions should be solely due to the noise in the system.

The upper plots in Figure 3 show that model is not sufficient. Smoothing $\boldsymbol{\nu}$ reveals significant structure (top left). Plotting the innovations along the path of observations make the relationship between the observations and the innovations clear: there is some unmodeled process acting to keep the hand moving in a circular motion (top right). The only significant unmodeled process is the purposeful control signal that being applied to the hand by the muscles.

In this example, there is one active, cyclo-stationary control behavior, and it's relationship to the state of the physical system is straightforward. There is a one-to-one mapping between the state and the phase offset into the cyclic control, and a one-to-one mapping between the offset and the control to be applied. If we use the smoothed innovations as our model and assume a linear control model of identity, then the linear prediction becomes:

$$\hat{\mathbf{x}}_t = \boldsymbol{\Phi}_t\hat{\mathbf{x}}_{t-1} + \mathbf{I}\mathbf{u}_{t-1} \tag{11}$$

where $\mathbf{u}_{t-1}$ is the control signal applied to the system. The lower plots in Figure 3 show the result of modeling the hand motion with a model of passive physics and a model of the active control. The smoothed innovations are basically zero: there is no part of the signal that deviates from our model except for the observation noise.

In this simple, linear example the system state, and thus the innovations, are represented the same coordinate system as the observations. With more complex dynamic and observations models, such as described in Section II.A, they could be represented in any arbitrary system, including spaces related to observation space in non-linear ways, for example as joint angles or other appropriate models.

The next section examines a more powerful form of model for control.

B. Multiple Behavior Models

Human behavior, in all but the simplest tasks, is not as simple as a single control model. The next most complex model of human behavior is to have *several* alternative models of control, one for each class of response. Since the innovations process is the part of the observation data that is unexplained by the dynamic model, the behavior model that explains the largest portion of the observations is, of course, the model most likely to be correct. This is known as the *multiple model* or *generalized likelihood* approach, and produces a generalized maximum likelihood estimate of the current and future values of the state variables [19].

At each time step, we calculate the probability $Pr^{(i)}$ of the m-dimensional innovations $\boldsymbol{\nu}$ given the i^{th} model and choose the model with the largest probability. The chosen model is then used in Equation 11 to estimate the current value of the state variables, to predict their future values, and to choose among alternative responses. The cost of these calculations is sufficiently small to make the approach quite practical for small control vocabularies.

Intuitively, this solution breaks the person's overall behavior down into several "prototypical" behaviors. We then classify the behavior by determining which model best fits the observations. This idea is similar to the multiple model approach of Friedmann 1993, and Isard 1996[20;15].

C. Hidden Markov Models of Control

Since human motion evolves over time, in a complex way, it is advantageous to explicitly model temporal dependence and internal states in the control process. A Hidden Markov Model (HMM) is one way to do this, and has been shown to perform quite well recognizing human motion[21].

The probability that the model is in a certain state, S_j given a sequence of observations, $\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_N$, is defined recursively. For two observations, the density is:

$$\Pr(\mathbf{O}_1, \mathbf{O}_2, \mathbf{q}_2 = S_j) = \left[\sum_{i=1}^{N} \pi_i b_i(\mathbf{O}_1)\mathbf{a}_{ij}\right] b_j(\mathbf{O}_2) \qquad (12)$$

Where π_i is the prior probability of being in a state i, and $b_i(\mathbf{O})$ is the probability of making the observation $\mathbf{O}$ while in state i. This is the Forward algorithm for HMM models.

Estimation of the control signal proceeds by identifying the most likely state given the current observation and the last state, and then using the observation density of that state as described above. We restrict the observation densities to be either a Gaussian or a mixture of Gaussians. There are well understood techniques for estimating the parameters of the HMM from data.

III. Implementation Details

The dynamic skeleton model currently includes the upper body and arms. The body element is rooted at the bottom to the world with a ball joint. The arms are composed of two elements each connected at the shoulder with a ball joint. The arm element are joined together with a ball joint at the elbow. All joints are also spanned by servos that apply damping torques to the joint. The ball joints are implemented as described in Section II.A.1.

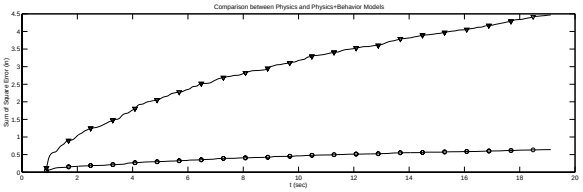

Fig. 4. Sum Square Error of a Physics-only tracker (triangles) vs. error from a Physics+Behavior Tracker

The top of the body element and the ends of the lower arm elements are connected to the perceptual input with soft constraints as described in Section II.A.2. The soft constraints are implemented as bounded proportional-derivative servos. The servos connecting the hands to the hand observations have a stiffness of $50\frac{N}{m}$ and a damping coefficient of $100\frac{Ns}{m}$ with a bound of $100N$. The servo connecting the head to the head observation has a stiffness of $250\frac{N}{m}$ and a damping coefficient of $200\frac{Ns}{m}$ with a bound of $100N$. If a blob is undetected in the current frame, the force bound is set to $0N$ for that body part until the next observation, and the model propagates forward in time without any influence from that observation.

The model performs integration steps at $500Hz$. The observations are updated at video frame rate: $30Hz$. Predictions are sent to the 2-D vision module at $120Hz$ to insure that timely prediction information is available at the start of each processing phase.

IV. Results

Figure 2 shows the real-time response of the model to a set of 2-D observations. The corresponding observations from the stereo pair are not shown. The model interpolates those portions of the body State that are not measured directly, such as the upper body and elbow orientation, by use of the model's intrinsic dynamics and the behavior (control) model. The model also rejects noise that is inconsistent with the dynamic model. Table 4 compares noise in the physics+behavior tracker with the physics-only tracker noise. It can be seen that there is a significant increase in performance.

Figure 5 illustrates another advantage of feedback from higher-level models to the low-level vision system. The solid plots show Y and Z coordinates of left and right hand tracks over time. The dotted plots are projections onto the $Y = 0$ and $Z = 5$ planes to aid readers in interpreting the 3-D structure of the data. The image in the upper left shows a diagram of the motion in relation to the coordinate system. The image sequence is taken from the input video. The results were recorded from the system during real-time operation.

The left plot shows that that without feedback, the 2-D

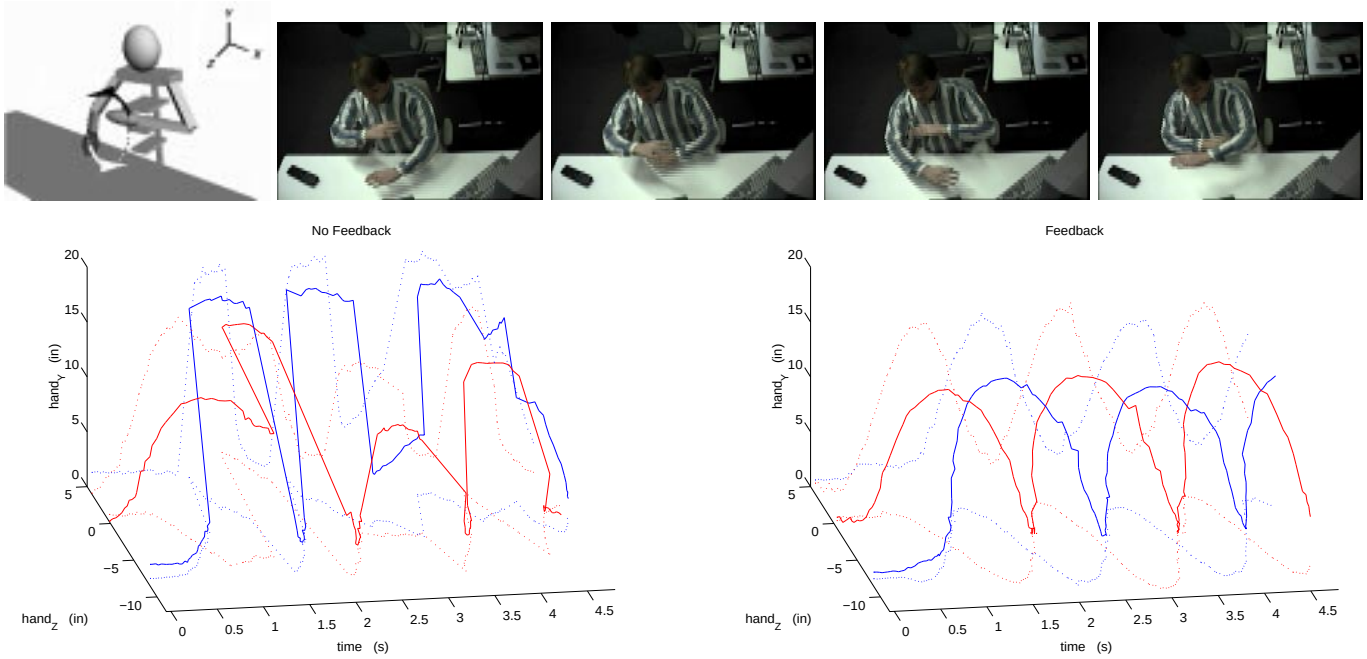

Fig. 5. Tracking performance on a sequence with significant occlusion. **Top:** A diagram of the sequence and a single camera's view of the motion. **Left:** A graph of tracking results without feedback. **Right:** Correct tracking when feedback is enabled.

tracker fails if there is even partial self-occlusion, or occlusion of an object with similar appearance (such as another person), from a single camera's perspective. These failures propagate through the system and result in these erroneous 3-D tracking results. These failures manifest in Figure 5 as discontinuities in the left plot. With feedback, information from the dynamic model and the perspective constraints can be used to resolve ambiguity during 2-D tracking. The improved results are shown in the right plot. With models of behavior, the systems should be able to tolerate longer occlusions.

V. Conclusion

We have presented a framework for human motion understanding, defined as estimation of the physical state of the body combined with interpretation of that part of the motion that cannot be predicted by passive physics alone. The behavior system operates in conjunction with a real-time, fully-dynamic, 3-D person tracking system that provides a mathematically concise formulation for incorporating a wide variety of physical constraints and probabilistic influences. The framework takes the form of a non-linear recursive filter that enables even pixel-level processes to take advantage of the contextual knowledge encoded in the higher-level models. Some of the demonstrated benefits of this approach include: increase in 3-D tracking accuracy, insensitivity to temporary occlusion, and the ability to handle multiple people.

References

[1] Ali Azarbayejani and Alex Pentland, "Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features," in *Proceedings of 13th ICPR*, Vienna, Austria, August 1996, IEEE Computer Society Press.

[2] Christopher R. Wren and Alex P. Pentland, "Dynamic models of human motion," in *Proceedings of FG'98*, Nara, Japan, April 1998, IEEE.

[3] Ernst D. Dickmanns and Birger D. Mysliwetz, "Recursive 3-d road and relative ego-state recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 199–213, February 1992.

[4] J. O'Rourke and N.I. Badler, "Model-based image analysis of human motion using constraint propagation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, pp. 522–536, November 1980.

[5] K. Oatley, G. D. Sullivan, and D. Hogg, "Drawing visual conclusions from analogy: preprocessing, cues and schemata in the perception of three dimensional objects," *Journal of Intelligent Systems*, vol. 1, no. 2, pp. 97–133, 1988.

[6] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 730–742, July 1991.

[7] I. Kakadiaris, D. Metaxas, and R. Bajcsy, "Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach," in *CVPR94*, 1994, pp. 980–984.

[8] Dimitris Metaxas and Dimitris Terzopoulos, "Shape and non-rigid motion estimation through physics-based synthesis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 580–591, 1993.

[9] K. Rohr, "Cvgip: Image understanding," *"Towards Model-Based Recognition of Human Movements in Image Sequences*, vol. 1, no. 59, pp. 94–115, 1994.

[10] A. Baumberg and D. Hogg, "An efficient method for contour tracking using active shape models," in *Proceeding of the Workshop on Motion of Nonrigid and Articulated Objects*. IEEE Computer Society, 1994.

[11] Luis Goncalves, Enrico Di Bernardo, Enrico Ursella, and Pietro Perona, "Monocular tracking of the human arm in 3d," in *International Conference on Computer Vision*, Cambridge, MA, June 1995.

[12] D. M. Gavrila and L. S. Davis, "Towards 3-d model-based tracking and recognition of human movement: a multi-view approach," in *International Workshop on Automatic Face- and Gesture-Recognition*. IEEE Computer Society, 1995, Zurich.

[13] J.M. Rehg and T. Kanade, "Visual tracking of high dof articulated structures: An application to human hand tracking," in *European Conference on Computer Vision*, 1994, pp. B:35–46.

[14] Alex Pentland and Andrew Liu, "Modeling and predicition of human behavior," in *IEEE Intelligent Vehicles 95*, September 1995.

[15] Michael Isard and Andrew Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. European Conference on Computer Vision*, Cambridge, UK, 1996, pp. 343–356.

[16] Christoph Bregler, "Learning and recognizing human dynamics in video sequences," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, June 1997.

[17] Andrew Witkin, Michael Gleicher, and William Welch, "Interactive dynamics," in *ACM SIGGraph, Computer Graphics*. March 1990, vol. 24:2, pp. 11–21, ACM SIGgraph.

[18] Christopher Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, July 1997.

[19] A. S. Willsky, "Detection of abrupt changes in dynamic systems," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, Eds., number 77 in Lecture Notes in Control and Information Sciences, pp. 27–49. Springer-Verlag, 1986.

[20] Martin Friedmann, Thad Starner, and Alex Pentland, "Device synchronization using an optimal linear filter," in *Virtual Reality Systems*, H. Jones, Ed. Academic Press, 1993.

[21] Thad Starner and Alex Pentland, "Real-time american sign language recognition from video using hidden markov models," in *Proceedings of International Symposium on Computer Vision*, Coral Gables, FL, USA, 1995, IEEE Computer Society Press.