Physics-Based Person Tracking Using Simplified Lower-Body Dynamics

Marcus A. Brubaker, David J. Fleet, Aaron Hertzmann

Department of Computer Science

University of Toronto

{mbrubake,fleet,hertzman}@cs.toronto.edu

Abstract

We introduce a physics-based model for 3D person tracking. Based on a biomechanical characterization of lower-body dynamics, the model captures important physical properties of bipedal locomotion such as balance and ground contact, generalizes naturally to variations in style due to changes in speed, step-length, and mass, and avoids common problems such as footskate that arise with existing trackers. The model dynamics comprises a two degree-of-freedom representation of human locomotion with inelastic ground contact. A stochastic controller generates impulsive forces during the toe-off stage of walking and spring-like forces between the legs. A higher-dimensional kinematic observation model is then conditioned on the underlying dynamics. We use the model for tracking walking people from video, including examples with turning, occlusion, and varying gait.

1. Introduction

Current methods for recovering human pose and motion from video rely on *kinematic* models. Early models were specified by hand (e.g., with joint limits and smoothness constraints), while many recent models, have been learned from motion capture data (e.g., [2, 6, 8, 11, 25, 26, 27, 30, 33]). While such models help reduce ambiguities in tracking, the 3D motions reconstructed with these methods are often physically implausible. The most common artifacts include jerky motions, feet that slide when in contact with the ground (or float above it), and out-of-plane rotations that violate balance. Additionally, kinematic models have difficulty generalizing far beyond the training data. For example, a model trained on walking with a short stride may have difficulty tracking and reconstructing the motion of someone walking with a long stride.

To address these issues, we propose the use of *physics-based models of human motion*. We hypothesize that modeling the underlying dynamics of motion will lead to accurate tracking methods that naturally obey essential physical properties of human motion. In this paper, we consider the important special case of walking. Rather than attempt-
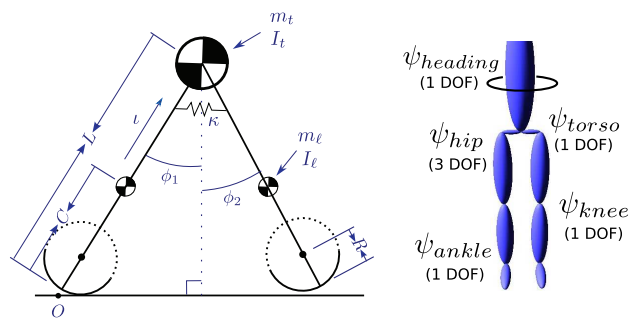

Figure 1. The Anthropomorphic Walker (left) and our Kinematic Model (right).

ing to model full-body dynamics, our approach is inspired by simplified biomechanical models of human locomotion [4, 5, 14, 17]. For example, the *Anthropomorphic Walker* [14, 15], shown in Fig. 1, exhibits human-like walking, including an inverted pendular trajectory and sudden changes in velocity at ground contacts. Further, it has very stable dynamics and naturally simulates walking with human-like speeds and step-lengths.

Our generative model comprises the Anthropomorphic Walker, a stochastic controller that generates forces, and a higher-dimensional kinematic model conditioned on the low-dimensional dynamics. The image likelihood is defined as a function of the kinematic state. Tracking is performed by simulating the model in a particle filter.

We demonstrate that monocular tracking with this model can be stable over long walking sequences. It handles occlusion, varying gait styles, and turning, producing realistic 3D reconstructions. With lower-body occlusions, it still produces realistic reconstructions and infers the time and location of ground contacts. Since the model represents the underlying physics of human motion with a smooth, low-dimensional state space, we gain the advantages of a physic-based model without the complexity of full-body dynamics.

Unlike most state-of-the-art methods, our model achieves high-quality tracking without learning prior pose or motion models. In principle, however, the use of physical models does not preclude the use of training data. In fact, we believe that significant potential for improved tracking lies in learning physics-based models from data.

2. Related Work

Most existing 3D person trackers assume an articulated kinematic model, often with joint limits and temporal smoothness assumptions (e.g., [11, 28, 33]). It has also been common to learn activity-specific models from motion capture data, either by learning a map from images to estimated poses (e.g., [1, 8]), or by learning prior density models with which one infers the most likely motion given image data (e.g., [19, 26, 27, 30]).

In highly-constrained scenarios, kinematic trackers can produce good results. However, such methods generally suffer from two major problems. First, they make unrealistic assumptions; e.g., motions are assumed to be smooth (which is violated at ground contact), and independent of global position and orientation. As a result, tracking algorithms exhibit a number of characteristic errors, including "foot-skate," in which a foot in contact with the ground appears to slide or float in space, and rotations of the body that violate balance. Second, algorithms that learn kinematic models have difficulty generalizing beyond the training data. In essence, such models describe the likelihood of a motion by comparison to training poses; i.e., motions "similar" to the training data are considered likely. This means that, for every motion to be tracked, there must be a similar motion in the training database. In order to build a general tracker using current methods, an enormous database of human motion capture would be necessary.

A major theme of recent tracking methods is dimensionality reduction for learning low-dimensional models and dynamical systems from data [8, 23, 27, 30, 31]. In this paper, we employ a hand-designed low-dimensional representation based on models from the biomechanics literature. These models are known to accurately represent properties of human locomotion (such as gait variation and ground contact) that have not been demonstrated with learned models. We thus gain the advantages of a physics-based model without the complexity of full-body dynamics, and without the need for inference in a high-dimensional state space.

A few authors have employed physical models of motion for tracking. Pentland and Horowitz [20] and Metaxas and Terzopoulos [18] describe elastic solid models for tracking in conjunction with Kalman filtering, and give simple examples of articulated tracking by enforcing constraints. For these tracking problems, the dynamics are relatively smooth but high-dimensional. In contrast, we employ a model that captures the specific features of walking, including the nonlinearities of ground contact, without the complexity of modeling elastic motion.

Finally, the term "physics-based models" is used in different ways in computer vision. Among these, physics is often used as a metaphor for minimization, by applying virtual "forces" (e.g., [6, 12, 29]); unlike in our work, these forces are not meant to represent forces in the world.

3. Dynamic Model of Human Walking

Building realistic full-body models of human motion, like the control of complex dynamical systems in general, is extremely challenging. Nonetheless, work in biomechanics and robotics suggests that the dynamics of bipedal walking may be well described by relatively simple *passive-dynamic walking* models. Such models exhibit stable, bipedal walking as a natural limit cycle of their dynamics. The earliest such models, introduced by McGeer [16], were entirely passive and could walk downhill solely under the force of gravity. Related models have since been developed, including one with a kneed swing leg, another with an active torso, and one capable of running [17].

More recently, powered walkers based on passive-dynamic principles have been demonstrated to walk stably on level-ground [3, 14, 15]. These models exhibit human-like gaits and energy-efficiency. The energetics of such models have also been shown to accurately predict the preferred relationship between speed step-length in human walking [14]. In contrast, traditional approaches in robotics (e.g., as used by Honda's humanoid robot *Asimo*), employ highly-conservative control strategies that are significantly less energy-efficient and less human-like in appearance, making them a poor basis for modeling human walking [3, 22].

3.1. Dynamics

In this paper we design a stochastic model based on the minimally-powered *Anthropomorphic Walker* of Kuo [14, 15]. As shown in Fig. 1, it is a 2D abstraction with straight legs. It has a torsional spring between the legs and an impulsive "toe-off" by the back leg.

During normal walking, the *stance leg* is in contact with the ground, and the *swing leg* swings freely. As in a Lagrangian formulation, we define generalized coordinates representing the configuration of the walker at a given instant: $\mathbf{q} = (\phi_1, \phi_2)^T$, where ϕ_1 and ϕ_2 are the global orientations of the stance and swing legs, respectively. The equations of motion during normal walking are then written as a function of the current state $(\mathbf{q}, \dot{\mathbf{q}})$, with $\dot{\mathbf{q}} \equiv \frac{d\mathbf{q}}{dt}$:

$$\mathcal{M}(\mathbf{q})\,\ddot{\mathbf{q}} = \mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \kappa)\,, \qquad (1)$$

where $\mathcal{M}(\mathbf{q})$ is known as the generalized mass matrix and $\mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \kappa)$ is a generalized force vector which includes the spring force between the legs that is parameterized by its stiffness κ. This equation is a generalization of Newton's Second Law of Motion. Solving (1) at any instant gives the generalized acceleration $\ddot{\mathbf{q}}$. The details of (1) are given in Appendix A.

An important feature of walking is the collision of the swing leg with the ground. Collisions of the foot with the ground plane are treated as impulsive and perfectly inelastic, resulting in an instantaneous change in velocity. To al-

low for the "toe-off" characteristic of human walking — in which the stance leg gives a small push before swinging — an impulse with magnitude ι is applied at the time of collision. The pre- and post-collision velocities and the impulsive toe-off are related by the following generalized conservation of momentum equation:

$$\mathcal{M}^+(\mathbf{q})\,\dot{\mathbf{q}}^+ = \mathcal{M}^-(\mathbf{q})\,\dot{\mathbf{q}}^- + \mathcal{I}(\mathbf{q}, \iota) \qquad (2)$$

where $\mathcal{M}^-(\mathbf{q})$ and $\mathcal{M}^+(\mathbf{q})$ are the pre- and post-collision generalized mass matrices, $\dot{\mathbf{q}}^-$ and $\dot{\mathbf{q}}^+$ are the pre- and post-collision velocities, and $\mathcal{I}(\mathbf{q}, \iota)$ is the change in momentum due to the toe-off force. The post-collision velocities $\dot{\mathbf{q}}^+$ are found by solving the above system of equations. Details of (2) and collision detection are given in Appendix B.

Given κ and ι, the dynamical model is simulated using a standard ODE solver applied to (1); we use a fourth-order Runge-Kutta method. At contact, the simulation must be restarted after solving (2) for the post-collision velocities.

3.2. Control

The walking model has two control parameters $\theta = (\kappa, \iota)$, where κ is the spring stiffness and ι is the magnitude of the impulsive toe-off. Because these parameters are unknown prior to tracking, they are treated as hidden random variables. For effective tracking we desire a prior distribution over θ which, together with the dynamical model, defines a distribution over motions. A gait may then be generated by sampling θ and simulating the dynamics.

We assume that likely walking motions are characterized by stable, cyclic gaits. To specify this distribution, we first determine control parameters that generate cyclic gaits spanning the natural range of human walking speeds (roughly 2-7 km/h) and step-lengths (roughly 0.5-1.2m). For a given speed and step-length, we specify an initial pose $\mathbf{q}_0$ with both feet on the ground at the desired step-length and a simulation duration T determined by the desired speed and step-length. We then use Newton's method to solve

$$\mathcal{D}(\mathbf{q}_0, \dot{\mathbf{q}}_0, \theta, T) - (\mathbf{q}_0, \dot{\mathbf{q}}_0) = 0 \,, \qquad (3)$$

for $\dot{\mathbf{q}}_0$ and θ where $\mathcal{D}$ is a function that simulates the dynamics for duration T given an initial state $(\mathbf{q}_0, \dot{\mathbf{q}}_0)$ and parameters θ. Solving (3) for a discrete set of speeds and step-lengths yields the results shown in Figure 2. These plots show that the dynamic control parameters depend smoothly on the speed and step-length of the motion.

We then define the distribution over θ as follows. The impulsive force $\iota > 0$ is assumed to have a Gamma density, conditioned on the precollision velocity $\dot{\mathbf{q}}^-$, and is only sampled when a collision is detected. The Gamma distribution parameters are chosen such that $E[\iota] = \mu_\iota(\dot{\mathbf{q}}^-)$, where $\mu_\iota(\dot{\mathbf{q}}^-)$ is a bilinear function fit to the solutions of equation 3 and the variance is a constant.
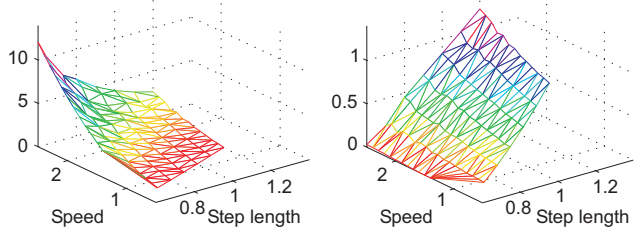

Figure 2. Optimal stiffness κ (left) and impulse magnitude ι (right) as functions of speed and step length are shown. These plots illustrate the flexibility and expressiveness of the model's control parameters.

For the ith stride, a mean value of κ is selected, $\bar{\kappa}^{(i)}$. Then, at each time-step, $\kappa_t \sim \mathcal{N}(\bar{\kappa}^{(i)}, \sigma_\kappa^2)$ where $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2. The mean $\bar{\kappa}^{(i)}$ is conditioned on $\bar{\kappa}^{(i-1)}$ such that $\bar{\kappa}^{(i)} \sim \mathcal{N}(\alpha\mu_\kappa + (1-\alpha)\bar{\kappa}^{(i-1)}, \sigma_{\bar{\kappa}}^2)$, where μ_κ is a global mean spring stiffness and α determines how close $\bar{\kappa}^{(i)}$ remains to μ_κ over time.

Because the walking model is very stable, the model is relatively robust to the choice of stochastic control. Other controllers may work just as well or better.

3.3. Conditional Kinematics

The model above is low-dimensional, easy to control, and produces human-like gaits. Nevertheless, it is purely 2D and does not represent all pose parameters of interest, such as the torso, knees and feet. We therefore add a higher-dimensional 3D kinematic model, conditioned on the underlying dynamics. The coupling of a simple physics-based model with a detailed kinematic model is similar to Popović and Witkin's physics-based motion editing system [21].

The kinematic model, depicted in Fig. 1(right), has legs, knees, feet and a torso. It has ball-and-socket joints at the hips, and hinge joints for knees and ankles. Although the upper body is not used in the physics model, it provides useful features for tracking. The upper body in the kinematic model comprises a single rigid body attached to the legs by a hinge joint.

The kinematic model is constrained to match the dynamic model at every instant: the upper-leg orientations (in the sagittal plane) and which foot is on the ground must be the same. Each remaining kinematic DOF $\psi_{j,t}$ is modeled as a smooth, 2nd-order Markov process:

$$\psi_{j,t} = \psi_{j,t-1} + \Delta t(\alpha_j\dot{\psi}_{j,t-1} + k_j(\bar{\psi} - \psi_{j,t-1})) + \eta_j \quad (4)$$

where Δt is the size of the timestep, $\dot{\psi}_{j,t-1} = (\psi_{j,t-1} - \psi_{j,t-2})/\Delta t$ is the joint angle velocity, and η_j is IID Gaussian noise with mean zero and variance σ_j^2. This model is analogous to a damped spring model with noisy accelerations where k_j is the spring constant, $\bar{\psi}_j$ is the rest position, α_j is related to the damping constant and η_j is noisy acceleration. Joint limits which require that $\psi_j^{\min} \leq \psi_j \leq \psi_j^{\max}$
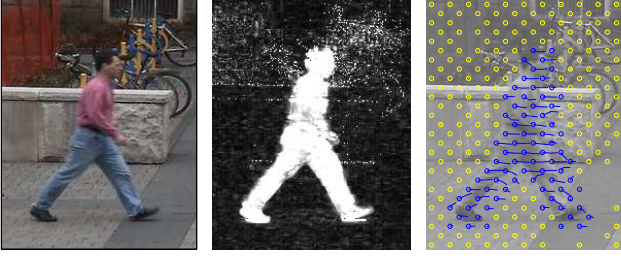
Figure 3. A cropped image (left) is shown with a example of the background negative log likelihood (middle), and a grid of motion trajectories (blue/yellow depict large/small speeds).

are imposed where appropriate and η_j is truncated [24] to satisfy the joint limits.

The joint evolution constants are fixed, with the exception of the knee rest position of the swing leg. Due to the sharp bend in the knee immediately after toe-off, a simple smoothness prior has difficulty modelling this joint. To account for this, we define $\bar{\psi}_{swing\ knee}$ as a linear function of the sagittal hip angle.

4. Sequential Monte Carlo Tracking

Tracking is formulated with a state-space representation. The state $\mathbf{s}_t$ at time t comprises dynamics parameters, $\mathbf{d}_t$, and the kinematic DOFs, $\mathbf{k}_t$; i.e., $\mathbf{s}_t = (\mathbf{d}_t, \mathbf{k}_t)$. With the Markov properties of the generative model above, and conditional independence of the measurements, one can write the posterior recursively;

$$p(\mathbf{s}_{1:t} \,|\, \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t \,|\, \mathbf{s}_t)\, p(\mathbf{s}_t \,|\, \mathbf{s}_{t-1})\, p(\mathbf{s}_{1:t-1} \,|\, \mathbf{z}_{1:t-1}) \quad (5)$$

where $\mathbf{s}_{1:t} \equiv [\mathbf{s}_1, ..., \mathbf{s}_t]$ denotes a state sequence, $\mathbf{z}_{1:t} \equiv [\mathbf{z}_1, ..., \mathbf{z}_t]$ denotes the observation history, $p(\mathbf{z}_t \,|\, \mathbf{s}_t)$ is the observation likelihood described below, and $p(\mathbf{s}_t \,|\, \mathbf{s}_{t-1})$ is the dynamic model described above.

Using a particle filter, we approximate the posterior by a weighted set of N samples $\mathcal{S}_t = \{\mathbf{s}_t^{(j)}, w_t^{(j)}\}_{j=1}^N$. Given the recursive form of (5), the posterior $\mathcal{S}_t$, given $\mathcal{S}_{t-1}$, can be computed in two steps: 1) draw samples $\mathbf{s}_t^{(j)} \sim p(\mathbf{s}_t \,|\, \mathbf{s}_{t-1}^{(j)})$; and 2) update weights $w_t^{(j)} = c\, w_{t-1}^{(j)}\, p(\mathbf{z}_t \,|\, \mathbf{s}_t^{(j)})$ where c is used to ensure the weights sum to 1.

This approach often works well until particle depletion becomes a problem, i.e., where only a small number of weights are significantly non-zero. Rather than re-sampling according to the current weights so that samples with low weights are discarded, as is often done, we use future data to predict how well current samples are likely to fare in the future. To this end we maintain an approximate posterior, $p(\mathbf{s}_{t:t+\tau} \,|\, \mathbf{z}_{1:t+\tau})$, for state sequences in a small window of $\tau + 1$ frames, denoted $\mathcal{S}_{t:t+\tau} = \{\mathbf{s}_{t:t+\tau}^{(j)}, w_{t:t+\tau}^{(j)}\}_{j=1}^N$. The sample set is obtained by simulating the model for $\tau+1$ time steps, given $\mathcal{S}_{t-1}$, and then evaluating the likelihood of each trajectory. Following [7, 13], when the effective number of

samples, $N_{eff} \approx (\sum_j (w_{t:t+\tau}^{(j)})^2)^{-1}$ becomes too small we re-sample $\mathcal{S}_{t-1}$ using importance sampling; i.e.,

1. Draw samples $\mathbf{s}_{t-1}^{(k)}$ from the weights $\{\hat{w}_{t-1}^{(j)}\}_{j=1}^N$ where $\hat{w}_{t-1}^{(j)} = (1 - \gamma)w_{t-1}^{(j)} + \gamma w_{t:t+\tau}^{(j)}$ and γ represents our trust in our approximation $\mathcal{S}_{t:t+\tau}$;

2. Set the new weights to be $w_{t-1}^{(k)}/\hat{w}_{t-1}^{(k)}$ and then normalize the weights so they sum to 1.

The importance re-weighting (step 2) is needed to maintain a properly weighted approximation to the posterior (5).

Looking at future data gives us a better indication of which samples in $\mathcal{S}_{t-1}$ are likely to survive. This provides better proposals since the quality of a sample is not always immediately evident. With this importance sampling we re-sample less frequently, and the tracker is more efficient. We also obtain better marginal posterior approximations at time t. That is, if the weighted sample set is $\mathcal{S}_{t:t+\tau} = \{\mathbf{s}_{t:t+\tau}^{(j)}, w_{t:t+\tau}^{(j)}\}_{j=1}^N$, then the marginal, $p(\mathbf{s}_t \,|\, \mathbf{z}_{1:t+\tau})$, is just $\{\mathbf{s}_t^{(j)}, w_{t:t+\tau}^{(j)}\}_{j=1}^N$. Below we use $\tau = 3$ and $\gamma = 0.75$.

4.1. Likelihood

The 3D articulated body model comprises a torso and limbs, each of which is modeled as a tapered ellipsoidal cylinder. The size of each part is set by hand, as is the pose of the model in the first frame of each sequence. The likelihood is based on foreground/background appearance models and on optical flow measurements [9].

A background model, learned from a small subset of images, comprises a mean background image and a 3×3 covariance matrix (e.g., see Fig. 3). The foreground model assumes that pixels are IID in each part (i.e., foot, legs, torso, head), with densities given by Gaussian mixtures. The mixtures are learned from the initial pose in the first frame.

Optical flow is estimated at grid locations in each frame (e.g., see Fig. 3), using a robust M-estimator with non-overlapping support. The eigenvalues/vectors of the local gradient tensor in each region of support provide a crude approximation to the estimator covariance Σ. For the likelihood of a flow estimate, $\mathbf{v}$, given the 2D motion specified by the state, $\mathbf{u}$, we use a heavy-tailed Student's t distribution (for robustness). The log-likelihood is

$$\log p(\mathbf{v}|\mathbf{u}) = -\frac{\log |\Sigma|}{2} - \frac{n+2}{2} \log(1+e^2) + c \quad (6)$$

where $e^2 = \frac{1}{2}(\mathbf{v} - \mathbf{u})^T \Sigma^{-1}(\mathbf{v} - \mathbf{u})$ and $n = 2$ is the degrees of freedom, and c is a constant. Because the camera is not moving in our image sequences, we define the log-likelihood of a flow measurement on the background as given by (6) with $\mathbf{u} = \mathbf{0}$.

To cope with large correlations between nearby measurement errors, we define the appearance and flow log-likelihood for each body part to be the average log-likelihood over all visible measurements for each part. To
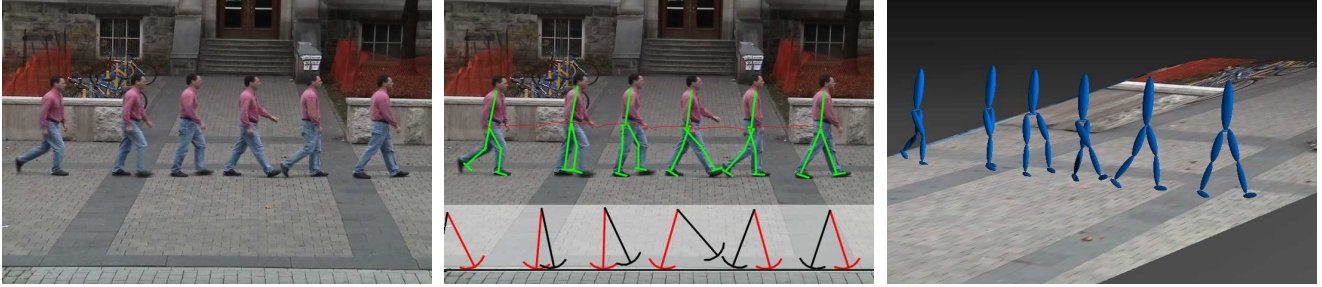
Figure 4. Composite images show the subject at several frames, depicting the motion over the 130 frame sequence: (left) the original images; (middle) the inferred poses of the MAP kinematics overlayed on the images, with the MAP dynamics state depicted along the bottm (the stance leg in red); (right) A 3D rendering of MAP poses from another viewpoint.
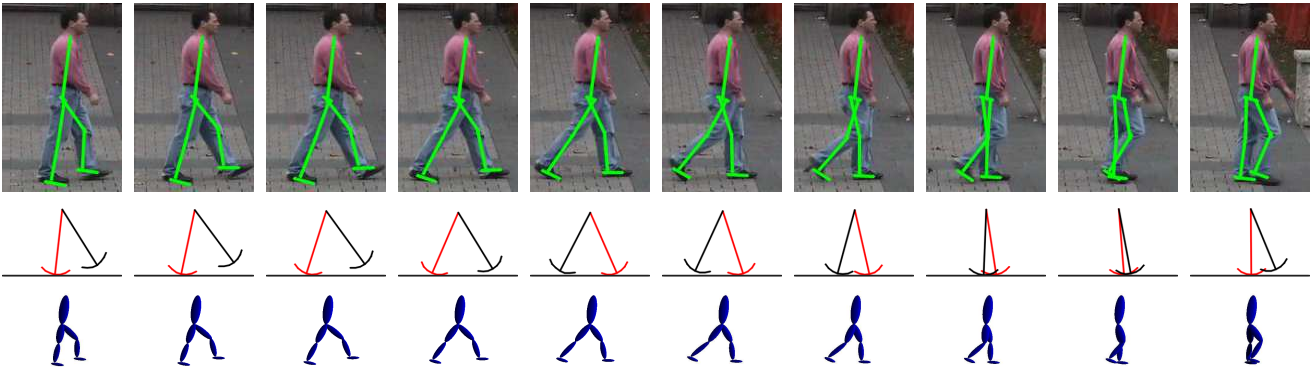

Figure 5. Cropped images showing every 4th frame of the MAP trajectory in Experiment 1 for 1.5 strides during change of speed.
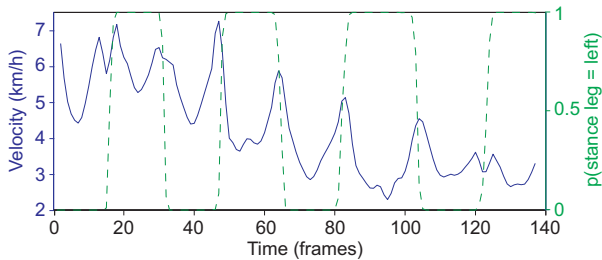

Figure 6. Inferred velocity of the MAP trajectory over time in Experiment 1 (blue). The dashed green line indicates the probability of the left leg being the stance leg given the data up to that frame.

avoid computing the log-likelihood over the entire image, we only compute log-likelihood ratios over regions of the image to which the 3D body geometry projects. Then, the total log-likelihood-ratio is the sum of the appearance and flow log-likelihood-ratios of the parts. This provides the desired log-likelihood, $\log p(\mathbf{z}_t \,|\, \mathbf{s}_t)$, up to an additive constant.

5. Results

Here we show results from three experiments, each using the same parameters for the likelihood, kinematic evolution and dynamics. All experiments used 5000 particles, with resampling when $N_{eff} < 500$. The initial state in each case was hand-initialized.

Experiment 1: Changes in Speed. Figure 4 shows a composite image of a walking sequence in which the sub-

ject's speed decreases from almost 7 to 3 km/hr. Such speed changes are handled naturally by the physics-based model. Fig. 5 shows cropped versions of tracking results for a short subsequence, demonstrating the consistency of the tracker. Weakness in the conditional kinematic model at high speeds leads to subtle anomalies, especially around the knees, which can be seen in the early frames of this subsequence.

Experiment 2: Occlusion. We simulate occlusion by blanking out an image region as shown in Fig. 7. The silhouette of the lower body is therefore lost, and we discard all flow measurements that encroach upon the occluder. Nevertheless, the subtle motion of the torso is enough to track the person, infer foot positions, and recover 3D pose. It is particularly interesting to examine the posterior distribution within and at the edge of the occluder (see Fig. 8). While there is increased posterior uncertainty during the occlusion, it does not diffuse monotonically. Rather, motion of the upper body allows the tracker to infer which leg is the stance leg and the contact location. Notice that soon after ground contact the marginal posterior over the stance foot position shrinks. Finally, during occlusion, leg-switching can occur but is unlikely. This is visible in the posterior distribution and the ambiguity is quickly resolved after the occlusion as would be expected.

Experiment 3: Turning. While the physics-based dynamics is two dimensional we are still able to successfully
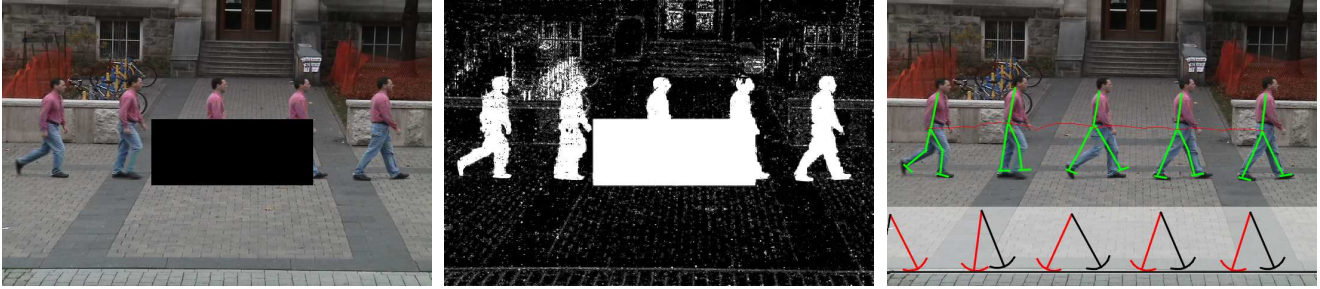
Figure 7. Composite images show the MAP trajectory at several frames for Experiment 2.
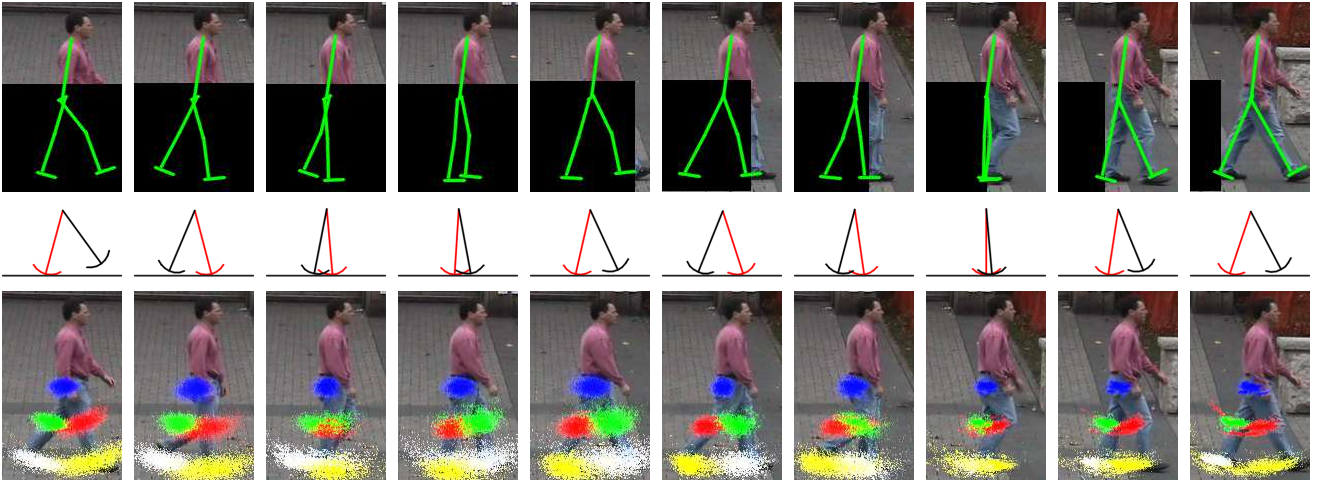

Figure 8. Cropped images showing every 4th frame of the MAP trajectory (top) and posterior distribution (bottom) in Experiment 2. In the posterior images, the feet, knee and hip points are plotted for each particle with intensity proportional to their log weight.
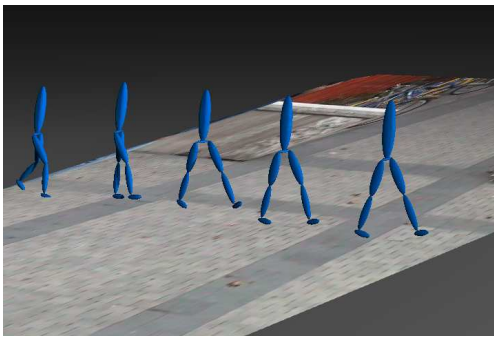

Figure 9. 3D rendering of the MAP trajectory in Experiment 2.

Figure 10. MAP trajectory velocity (blue) and stance leg posterior (dashed green) for the times shown in Figure 11.

6. Discussion and Future Work

In this paper we showed that physics-based models offer significant benefits in terms of accuracy, stability, and generality for person tracking. Here we used a simple powered walking model, but we are currently exploring more sophisticated physical models (e.g., [17]) which may yield even more general trackers for other types of motion. Nevertheless, there will be a trade-off between model generality and the difficulty of designing a controller. Although our approach employs online Bayesian inference, it should also be possible to incorporate physical laws within other tracking frameworks such as discriminative methods. Models similar to this may also be used for modelling and tracking

track 3D walking motions because of the conditional kinematics. As can been seen in Figure 11, the model successfully tracks the person through a sharp turn in a sequence of more than 400 frames. Despite the limitations of the physical model, it is able to accurately represent the dynamics of the motion in 2D while the conditional kinematic model represents the turning motion. Looking at Figure 10, it can be seen that during the turn there is significant uncertainty about the stance leg. This is expected given the visual ambiguitity of the individual frames in Figure 11.
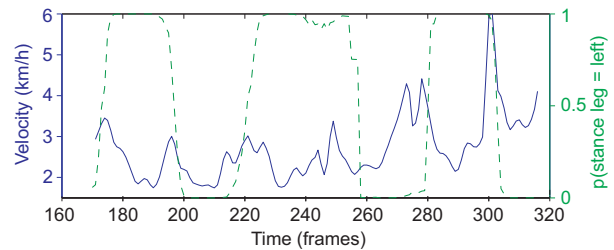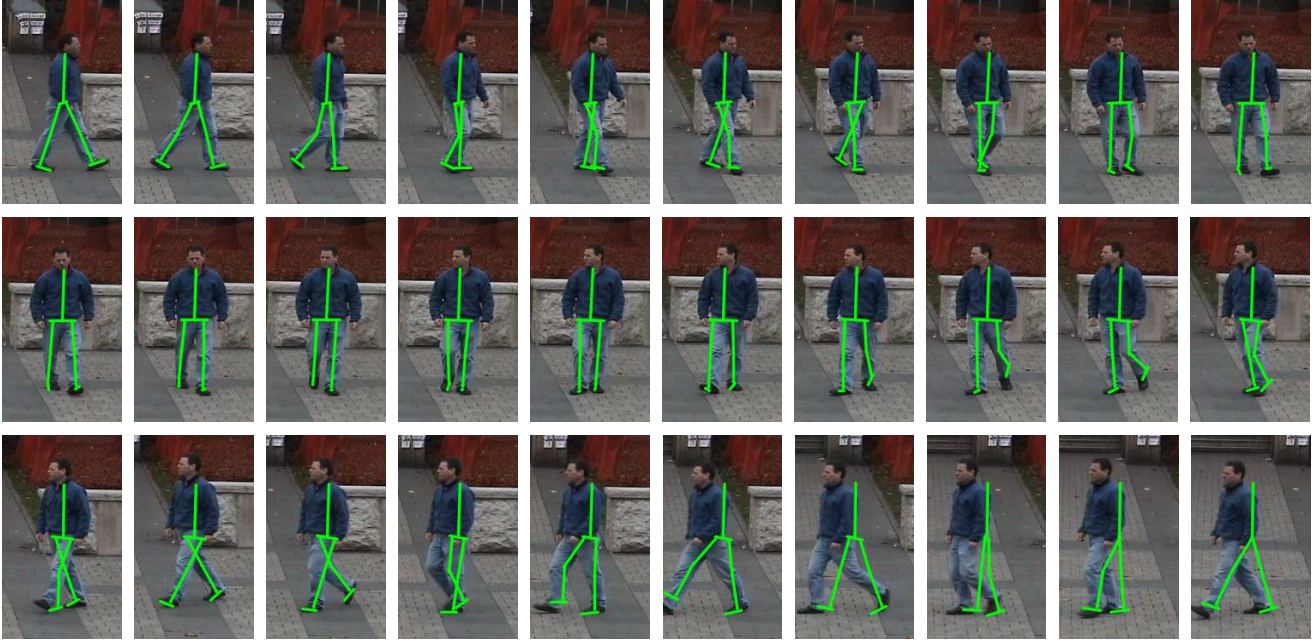
Figure 11. Cropped images showing every 5th frame of the MAP trajectory through an acceleration and sharp turn, starting at frame 170.

other animals [10].

References

[1] A. Agarwal and B. Triggs. Recoving 3D human pose from monocular images. *PAMI*, 28(1):44–58, 2006.

[2] K. Choo and D. Fleet. People tracking using hybrid Monte Carlo filtering. *ICCV*, V. 2, pp. 321–328, 2001.

[3] S. Collins, A. Ruina, R. Tedrake, and M. Wisse. Efficient bipedal robots based on passive-dynamic walkers. *Science*, 307(5712):1082–1085, 2005.

[4] S. H. Collins and A. Ruina. A bipedal walking robot with efficient and human-like gait. *ICRA*, 2005.

[5] S. H. Collins, M. Wisse, and A. Ruina. A three-dimensional passive-dynamic walking robot with two legs and knees. *Int. J. Robotics Research*, 20(7):607–615, 2001.

[6] Q. Delamarre and O. Faugeras. 3D articulated models and multiview tracking with physical forces. *CVIU*, 81(3):328–357, 2001.

[7] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[8] A. Elgammal and C.-S. Lee. Inferring 3D body pose from silhouettes using activity manifold learning. In *CVPR*, V. 2, pp. 681–688, 2004.

[9] D. Fleet and Y. Weiss. Optical flow estimation. In *Mathematical Models of Computer Vision: The Handbook*, N. Paragios et al. (eds). Springer, Ch. 15, pp. 239-258, 2005

[10] R. Full and D. Koditschek. Templates and anchors: Neuromechanical hypotheses of legged locomotion on land. *J. Exper. Biology*, 202:3325–3332, 1999.

[11] L. Herda, R. Urtasun, and P. Fua. Hierarchical implicit surface joint limits for human body tracking. *CVIU*, 99(2):189–209, 2005.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321–331, 1987.

[13] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and bayesian missing data problems. *J. American Stat. Assoc.*, 89(425):278–288, 1994.

[14] A. D. Kuo. A simple model of bipedal walking predicts the preferred speed–step length relationship. *J. Biomech. Eng.*, 123(3):264–269, 2001.

[15] A. D. Kuo. Energetics of actively powered locomotion using the simplest walking Model. *J. Biomech. Eng.*, 124:113–120, 2002.

[16] T. McGeer. Passive dynamic walking. *Int. J. Robotics Research*, 9(2):62–82, 1990.

[17] T. McGeer. Principles of walking and running. *Advances in Comparative and Env. Phys.*, V. 11, ch. 4, pp.113–139. Springer-Verlag, 1992.

[18] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *PAMI*, 15(6):580–591, 1993.

[19] V. Pavlović, J. Rehg, T.-J. Cham, and K. Murphy. A dynamic Bayesian network approach to figure tracking using learned dynamic models. *ICCV*, pp. 94–101, 1999.

[20] A. Pentland and B. Horowitz. Recovery of nonrigid motion and structure. *PAMI*, 13(7):730–742, 1991.

[21] Z. Popović and A. Witkin. Physically based motion transformation. *SIGGRAPH*, pp. 11–20, 1999.

[22] G. Pratt. Legged robots at MIT: What's new since Raibert? *IEEE Rob. & Aut.*, 7(3):15–19, 2000.

[23] A. Rahimi, B. Recht, and T. Darrell. Learning appearance manifolds from video. *CVPR*, 2005.

[24] C. Robert. Simulation of truncated normal variables. *Statistics and Computing*, 5(2):121–125, 1995.

[25] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter-sensitive hashing. *ICCV*, pp. 750–757, 2003.

[26] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3D human figures using 2D image motion. *ECCV*, V. 2, pp. 702–718, 2000.

[27] C. Sminchisescu and A. Jepson. Generative modeling for continuous non-linearly embedded visual inference. *Proc. ICML*, pp. 96–103, 2004.

[28] C. Sminchisescu and B. Triggs. Kinematic jump processes for monocular 3D human tracking. *CVPR*, 2003.

[29] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: deformable superquadrics. *ICCV*, pp. 606–615, 1990.

[30] R. Urtasun, D. Fleet, and P. Fua. 3D people tracking with Gaussian Process dynamical models. *CVPR*, V. 1, pp.238–245, 2006.

[31] R. Urtasun, D. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. *ICCV*, V. 1, pp. 403–410, 2005.

[32] R. Q. Linde and A. L. Schwab. Lecture Notes Multibody Dynamics B, wb1413, course 1997/1998. Lab. for Eng. Mech. , Delft Univ. of Tech. , 2002.

[33] S. Wachter and H. H. Nagel. Tracking persons in monocular image sequences. *CVIU*, 74(3):174–192, 1999.

A. Equations of motion

As shown in Fig. 1, the walker has two legs of length L and a rigid torso with mass m_t and moment of inertia I_t, attached at the hip. The "feet" are circles of radius R which roll along the ground as the model moves. Each leg has mass m_ℓ and moment of inertia I_ℓ, centered at distance C from the foot. The legs are connected by a linear spring with stiffness κ to simulate muscle torques at the hips. The parameters of the dynamic model are similar to those in [15].

To derive the equations of motion for the walking model, we employ the TMT method [32], a convenient recipe for constrained dynamics. The TMT formulation is equivalent to Lagrange's equations of motion, but requires fewer steps to derive. We begin by defining the kinematic transformation, which maps from the generalized coordinates $\mathbf{q} = (\phi_1, \phi_2)$ to a 6×1 vector containing the linear and angular coordinates of each rigid body. The torso is treated as rigidly connected to the stance leg, and thus the kinematic transformation is

$$\mathbf{k}(\mathbf{q}) = \begin{bmatrix} -R\phi_1 - (C_1 - R)\sin\phi_1 \\ R + (C_1 - R)\cos\phi_1 \\ \phi_1 \\ -R\phi_1 - (L - R)\sin\phi_1 + (L - C)\sin\phi_2 \\ R + (L - R)\cos\phi_1 - (L - C)\cos\phi_2 \\ \phi_2 \end{bmatrix}$$

where $C_1 = \frac{(Cm_\ell + Lm_t)}{m_\ell + m_t}$ is the location along the stance leg of the combined center rigid body. Dependence of angles on

time is omitted for brevity. The origin, O, of the coordinate system is at the base of the stance foot as shown in Fig. 1.

The equations of motion are summarized as

$$\mathbf{T}^T \mathbf{M} \mathbf{T} \ddot{\mathbf{q}} = \mathbf{f} + \mathbf{T}^T \mathbf{M} (\mathbf{a} - \mathbf{g}) \qquad (7)$$

where the matrix $\mathbf{T}$ is the 6×2 Jacobian of $\mathbf{k}$, i.e., $\mathbf{T} = \partial \mathbf{k}/\partial \mathbf{q}$. The reduced mass matrix is $\mathbf{M} = \mathrm{diag}(m_1, m_1, I_1, m_\ell, m_\ell, I_\ell)$, where $m_1 = m_\ell + m_t$ is the combined mass of the stance leg, and $I_1 = I_\ell + I_t + (C_1 - C)^2 m_\ell + (L - C_1)^2 m_t$ is the combined moment of inertia of the stance leg. The convective acceleration is

$$\mathbf{g} = \frac{\partial}{\partial \mathbf{q}} \left(\frac{\partial \mathbf{k}}{\partial \mathbf{q}} \dot{\mathbf{q}} \right) \dot{\mathbf{q}} \qquad (8)$$

and $\mathbf{a} = g[0, -1, 0, 0, -1, 0]^T$ is the generalized acceleration vector due to gravity ($g = 9.8 m/s^2$). The generalized spring force is $\mathbf{f} = \kappa[\phi_2 - \phi_1, \phi_1 - \phi_2]^T$. By substitution of variables, it can be seen that (7) is equivalent to (1): $\mathcal{M}(\mathbf{q}) = \mathbf{T}^T \mathbf{M} \mathbf{T}$ and $\mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \kappa) = \mathbf{f} + \mathbf{T}^T \mathbf{M} (\mathbf{a} - \mathbf{g})$.

B. Collision and support transfer

Since the end of the swing leg is even with the ground when $\phi_1 = -\phi_2$, collisions are found by detecting zero-crossings of $\mathcal{C}(\phi_1, \phi_2) = \phi_1 + \phi_2$. However, our model also allows the swing foot to move below the ground[1], and thus a zero-crossing can occur when the foot passes above the ground. Hence, we detect collisions by detecting zero-crossings of $\mathcal{C}$ when $\phi_1 < 0$ and $\dot{\mathcal{C}} < 0$.

The dynamical consequence of collision is determined by a system of equations relating the instantaneous velocities immediately before and after the collision. In particular, the post-collision velocities $\dot{q}^+$ can be solved for using

$$\mathbf{T}^{+T} \mathbf{M} \mathbf{T}^+ \dot{\mathbf{q}}^+ = \mathbf{T}^{+T} (\mathbf{v} + \mathbf{M} \mathbf{T} \dot{\mathbf{q}}^-) \qquad (9)$$

where $\dot{\mathbf{q}}^-$ are the pre-collision velocities, $\mathbf{T}$ is the pre-collision kinematic transfer matrix specified above,

$$\mathbf{k}^+(\mathbf{q}^-) = \begin{bmatrix} -R\phi_2 - (L - R)\sin\phi_2 + (L - C)\sin\phi_1 \\ R + (L - R)\cos\phi_2 - (L - C)\cos\phi_1 \\ \phi_1 \\ -R\phi_2 - (C_1 - R)\sin\phi_2 \\ R + (C_1 - R)\cos\phi_2 \\ \phi_2 \end{bmatrix}$$

is the post-collision kinematic transformation function, $\mathbf{T}^+ = \partial \mathbf{k}^+/\partial \mathbf{q}$, is the post-collision kinematic transfer matrix, $\mathbf{M}$ is the mass matrix as above and $\mathbf{v} = \iota[-\sin\phi_1, \cos\phi_1, 0, 0, 0, 0]^T$ is the impulse vector with magnitude ι. At collision, the origin of the coordinate system shifts forward by $2(R\phi_2 + (L - R)\sin\phi_2)$. The swing and stance leg switch roles; i.e., ϕ_1 and ϕ_2 and their velocities are swapped. Simulation then continues as before.

---

[1]Because the Anthropomorphic Walker does not have knees, it can walk only by passing a foot through the ground.