The Kneed Walker for Human Pose Tracking

Marcus A. Brubaker   David J. Fleet
Department of Computer Science
University of Toronto
{mbrubake,fleet}@cs.toronto.edu

Abstract

The Kneed Walker is a physics-based model derived from a planar biomechanical characterization of human locomotion. By controlling torques at the knees, hips and torso, the model captures a full range of walking motions with foot contact and balance. Constraints are used to properly handle ground collisions and joint limits. A prior density over walking motions is based on dynamics that are optimized for efficient cyclic gaits over a wide range of natural human walking speeds and step lengths, on different slopes. The generative model used for monocular tracking comprises the Kneed Walker prior, a 3D kinematic model constrained to be consistent with the underlying dynamics, and a simple measurement model in terms of appearance and optical flow. The tracker is applied to people walking with varying speeds, on hills, and with occlusion.

1. Introduction

Physics-based models offer a natural way to constrain human pose and motion estimation [2, 4, 17, 31]. They provide principled parameterizations of human motion that generalize naturally to people of varying size and mass, and, for example, to different speeds of walking and running. Not only might physics-based parameterizations be more effective for modeling plausible poses and motions, they also capture the dependence of pose and motion on factors such as gravity, ground contact, and other physical interactions with the environment.

Nevertheless physics-based models for human motion remain largely unexplored in computer vision. Brubaker et al [4] proposed a model based on the *Anthropomorphic Walker* [12], a physics-based planar model with two straight legs, a single torsional spring and an impulsive collision model. While their monocular tracking results are encouraging, the Anthropomorphic Walker is extremely simple, as it only exhibits human-like gaits on level ground.

This paper shows that significantly more complex physics-based models can be designed for tracking a wider range of walking motions with greater accuracy. We propose a new model based on biomechanical characterizations of human walking [16] called the *Kneed Walker*. It has a torso and two legs with knees and ankles. It is capable of exhibiting a wide range of plausible gait styles.

One of the key contributions in this paper is to characterize the space of suitable joint torques. We show that one can optimize a parameterization of the joint torques, as a function of speed, step length and ground slope, to find stable human-like gaits. In doing so, we also address the proper handling of ground collisions and joint limits, both of which produce discontinuous motion. Finally, based on the Kneed Walker, we propose a simple generative model for monocular, video-based people tracking. We demonstrate the approach on several image sequences, showing that the new pose tracker is much more accurate but similarly efficient as that in [4] over a wide range of walking speeds. We also show that the new tracker handles people walking on steep hills. Unlike kinematic-based trackers the model produces physically plausible motions with no footskate.

2. Background

Current approaches to monocular people tracking in machine vision are dominated by models learned from motion capture (mocap) data. This includes generative approaches [13, 24, 27, 25, 29] and discriminative methods [1, 8, 26, 28]. While such approaches produce good results, generalization remains a major issue; for each motion to be tracked, there must be a similar training motion. As an example, tracking people over variable terrain has only been demonstrated with multiple cameras where prior models are less significant [6]. To model walking with all possible combinations of speeds, step-lengths, surface slopes, mass distributions, etc., an enormous mocap database would be necessary. We conjecture that physics-based models will generalize without the need for a large mocap database.

The second issue with kinematic models concerns their tendency to produce unrealistic motions. The most common problems are footskate, in which a foot in contact with the ground appears to slide or float in space, and body rotations that violate balance. These problems can be avoided with the use of physics-based models.

To date there has been very little work in computer vision on physics-based models for human pose and motion estimation. Metaxas and Terzopoulos [17] describe elastic solid models for tracking from 3D data with a Kalman Filter. They provide a simple example of upper-body articulated tracking for smooth motions, without contact. Wren and Pentland [32] used a physics-based formulation to constrain simple upper body tracking using stereo input. Working with 3D mocap data, Bissacco [2] used the underlying physics of human motion to help isolate contacts (modeled as impulsive collisions), characterizing the motion between contacts with linear dynamical systems.

Brubaker et al [4] proposed a generative model of human walking based on the Anthropomorphic Walker [12], a planar abstraction of human locomotion. It produces human-like gaits with realistic ground contact. However, for general locomotion the model is not sufficiently expressive since it has no knees, nor an actively controlled torso. In particular, it only produces human-like gaits on nearly level ground. To walk uphill a bent knee is required when the front foot hits the ground. The Kneed Walker is capable of a human-like gait over a wide range of speeds and step lengths, with realistic knee bend and torso sway. It also produces natural gaits on a wide range of ground slopes.

The Kneed Walker is inspired by planar biomechanical models which exhibit essential physical properties, such as balance and ground contact, while walking and running with human-like gaits and efficiency [5, 11, 12, 15]. Monopode models [19, 20] are even simpler, but do not explicitly model both legs during biped locomotion. There exist more complex anatomical models in biomechanics (e.g., [21]) and computer graphics (e.g., [14]), but with often more than a hundred degrees of freedom, control is a major issue.

3. The Kneed Walker

The Kneed Walker is a powered generalization of passive-dynamic planar models [15, 16]. It comprises a torso and two legs, modeled as articulated rigid bodies. It does not have an explicit ankle joint, but rather a rounded foot that rolls along the ground to simulate the effects of ankle articulation. The model's kinematic and inertial parameters are specified in Fig. 1(left). The mass m, center of mass offsets c and w, and the moment of inertia for each part are consistent with Dempster's body segment parameters [21]. Geometric parameters, including segment lengths ℓ, foot radius R, and foot eccentricity e, are based on [15].

In humans, antagonistic muscles between segments tighten and relax to exert forces on the body. These forces are represented using joint torques (see Fig. 1(right)). A parametric model of the joint torques is defined in terms of torsional springs. The swing knee is defined to have a damped spring with a stiffness κ_{K2}, resting length ϕ_{K2}^0 and

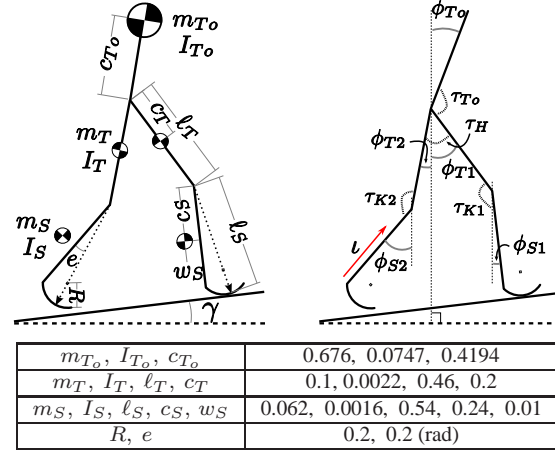| m_{T_o}, I_{T_o}, c_{T_o}	0.676, 0.0747, 0.4194
m_T, I_T, ℓ_T, c_T	0.1, 0.0022, 0.46, 0.2
m_S, I_S, ℓ_S, c_S, w_S	0.062, 0.0016, 0.54, 0.24, 0.01
R, e	0.2, 0.2 (rad)

Figure 1. The Kneed Walker. (Left) The kinematic and inertia parameters. (Right) The joint degrees of freedom and torques. The model variables are defined in the text.

damping constant d_{K2}. This specifies torque as

$$\tau_{K2} = -\kappa_{K2}(\phi_{T2} - \phi_{S2} - \phi_{K2}^0) - d_{K2}(\dot{\phi}_{T2} - \dot{\phi}_{S2}). \quad (1)$$

The stance knee torque τ_{K1} is defined similarly, with a seperate set of parameters, with resting length $\phi_{K1}^0 = 0$. Inspired by [12], the hip spring is undamped with a resting length of $\phi_H^0 = 0$, thereby producing torque

$$\tau_H = -\kappa_H(\phi_{T1} - \phi_{T2}). \quad (2)$$

Finally the torque on the torso is defined as

$$\tau_{To} = -\kappa_{To}(\phi_{To} - \phi_{To}^0) - d_{To}\dot{\phi}_{To}. \quad (3)$$

In addition to the torques applied during simulation, an impulsive force with magnitude ι is added at the time of ground contact (see Fig. 1(right)). This simulates the effects of the ankle during *toe-off*, where the back leg pushes off as support is transfered to the front foot [4].

3.1. Equations of motion

The generalized coordinates for the Kneed Walker comprise the 2D orientation of each rigid part, i.e., $\mathbf{q} = (\phi_{To}, \phi_{T1}, \phi_{S1}, \phi_{T2}, \phi_{S2})$. The pose $\mathbf{q}$ and its velocity $\dot{\mathbf{q}} = d\mathbf{q}/dt$ define the state of the dynamics, denoted $\mathbf{x} = (\mathbf{q}, \dot{\mathbf{q}})$. The equations of motion for the Kneed Walker are second-order ordinary differential equations defining the generalized acceleration $\ddot{\mathbf{q}}$ at each time in terms of $\mathbf{q}$, $\dot{\mathbf{q}}$, and the forces acting on the body:

$$\mathcal{M}(\mathbf{q})\ddot{\mathbf{q}} = \mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \theta, \gamma), \quad (4)$$

where $\mathcal{M}$ is a generalized mass matrix, $\mathcal{F}$ is a generalized force vector that includes gravity and all internal forces, θ specifies the spring parameters defined above, and γ is the ground slope. To derive these equations we use the TMT method [30]. This derivation is not trivial, but we omit it here as it is a well understood process.

3.2. Non-holonomic constraints and simulation

The equations of motion (4) fully characterize the dynamics of the unconstrained model. We also wish to impose joint limits on the knees, and prevent the feet from penetrating the ground. Doing so requires the use of unilateral, non-holonomic constraints, which can be challenging to handle computationally [3]. They can be incorporated using Lagrange multipliers or springs that are active only near constraint boundaries. However, these approaches produce stiff equations of motion that are computationally expensive to simulate with a realistic model of the (discontinuous) motion at constraint boundaries. This is unsuitable for tracking where efficient simulation is critical. Below we outline ways to better handle ground contact and joint limits.

Ground Contact   Following [2, 4, 12, 15] we treat ground collisions as impulsive events that cause an instantaneous change in momentum. For the Kneed Walker we assume that ground contact coincides with the transfer of support from one leg to the other. Contact can therefore be detected by monitoring the height of the swing foot during simulation. Such events are expected to be relatively infrequent. Upon contact the simulation is stopped, the change in momentum is computed, and the simulation is restarted but with the roles of the swing and stance legs reversed.

With this formulation, one can derive a constraint on the velocities immediately before and after the collision to model the change in momentum [4]. Given pre-collision velocity $\dot{\mathbf{q}}^-$ and toe-off impulse magnitude ι, the post-collision velocities, $\dot{\mathbf{q}}^+$, are found by solving

$$\mathcal{M}^+(\mathbf{q})\,\dot{\mathbf{q}}^+ = \mathcal{M}^-(\mathbf{q})\,\dot{\mathbf{q}}^- + \mathcal{I}(\mathbf{q}, \iota)\,. \qquad (5)$$

As above, the specific forms of the generalized mass matrices before and after collision, $\mathcal{M}^-$ and $\mathcal{M}^+$, and the impulsive force $\mathcal{I}$ can be derived using the TMT method [4, 30].

Joint Limits   Unlike ground contact, joint limit collisions are problematic for event-driven strategies. If a joint remains close to its limit, small variations in joint angle can produce large numbers of collisions in a short period of time. When at the joint limit, the equations of motion can be switched to prevent constraint violations (e.g., locking the knee as in [15]). But this yields multiple equations of motion along with the need to detect when to switch among them. Two knees, locked or unlocked, yields 4 seperate equations of motion plus switching conditions.

Instead, we advocate a variant of the approach in [22, 23]. Like event-driven strategies, constraints are monitored, and when violations are detected the constraint boundary is localized and velocities are instantaneously updated. Once on the boundary, the equations of motion (4) are modified to prevent acceleration into the prohibited region.

In detail, let the jth joint limit be written as $\mathbf{a}_j^T \mathbf{q} \geq b_j$ for some vector $\mathbf{a}_j$ and scalar b_j. For instance, the stance knee

joint limit is $\phi_{T1} - \phi_{S1} \geq 0$, so $\mathbf{a}_{K1} = (0, 1, -1, 0, 0)$ and $b_{K1} = 0$. When an event is localized, we need to (instantaneously) remove the component of velocity normal to the constraint boundary. That is, we solve for the new velocity $\dot{\mathbf{q}}^+$, given the old velocity $\dot{\mathbf{q}}^-$ using

$$\begin{bmatrix} \mathcal{M}(\mathbf{q}) & -\mathbf{a}_j \\ \mathbf{a}_j^T & 0 \end{bmatrix} \begin{bmatrix} \dot{\mathbf{q}}^+ \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathcal{M}(\mathbf{q})\,\dot{\mathbf{q}}^- \\ 0 \end{bmatrix}\,. \qquad (6)$$

With this instantaneous change in velocity the system is then located on the constraint boundary.

A constraint is then called *active* when on the constraint boundary and the current forces would otherwise violate the joint limits; i.e., the jth constraint is active when

$$\mathbf{a}_j^T \mathbf{q} = b_j\,, \ \ \mathbf{a}_j^T \dot{\mathbf{q}} = 0\,, \ \ \mathbf{a}_j^T \mathcal{M}(\mathbf{q})^{-1} \mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \theta.\gamma) < 0\,. \ \ (7)$$

To ensure that accelerations do not push the pose $\mathbf{q}$ into the prohibited region of the pose space, we require that $\mathbf{a}_j^T \ddot{\mathbf{q}} = 0$ for all *active* constraints j. This is achieved by adding virtual torques which operate normal to the constraint boundary for each active constraint. For the knee, these forces can be thought of as reactive forces caused by the kneecap to prevent hyperextension.

Let A be a matrix whose columns contain the vectors $\mathbf{a}_j$ for all active constraints. Virtual torques, given by $A\tau_v$ where τ_v is the vector of torque magnitudes, are added to the right side of (4). We solve for the virtual torque magnitudes, τ_v, with the following augmented equations of motion

$$\begin{bmatrix} \mathcal{M}(\mathbf{q}) & -A \\ A^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{q}} \\ \tau_v \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\mathbf{q}, \dot{\mathbf{q}}, \theta, \gamma) \\ \mathbf{0} \end{bmatrix}\,. \qquad (8)$$

In practice, numerical error prevents (6) and (8) from being exactly satisfied, and a final least-squares projection onto the boundary is often necessary. This technique has proved to be a stable, efficient component of the Kneed Walker.

3.3. Efficient, Cyclic Gaits

The *control space* of the Kneed Walker includes the impulsive toe-off and the four joint torques, parameterized as damped springs. However, most points in the control space will not generate plausible human-like gaits. It is therefore useful to formulate a prior distribution over the control space. One could learn a prior by fitting the Kneed Walker to mocap data, and then characterizing the space of forces. Unfortunately, this requires an enormous mocap database to cover the desired range of walking speeds, step-lengths and ground slopes for several subjects.

An alternative approach stems from first principles, with the assumption that human walking motions are fundamentally efficient. We characterize the space of plausible walking motions by searching for efficient, periodic gaits at a dense set of speeds, step-lengths and slopes. We then assume that plausible walking motions lie in the neighbourhood of these optimal gaits.

To find efficient gaits, we first define an objective function that penalizes large joint torques and large impulsive toe-off forces. That is, for simulation duration T we use

$$E(\mathbf{x}_0, \theta; \gamma) = \alpha_\iota \iota^\rho + \sum_{j \in \text{joints}} \frac{\alpha_j}{T} \int_0^T \tau_j^2 dt, \qquad (9)$$

where the torques τ_j depend on the initial state $\mathbf{x}_0$, the spring parameters θ, and the slope γ. The weights α_j balance the costs of joint torques, and $\rho = 1.5$ is set based on the energy function in [11]. The optimizations were robust to choices of α; we used $\alpha_{K1} = 0.3$, $\alpha_H = \alpha_{K2} = 0.007$, $\alpha_{To} = 0.034$ and $\alpha_\iota = 0.62$, placing the greatest penalties on the stance knee torque and the impulse magnitude.

To find optimal cyclic gaits, we minimize $E(\mathbf{x}_0, \theta; \gamma)$ over control parameters θ and the initial state $\mathbf{x}_0$ such that the simulated motion has the target speed and step length for slope γ. That is, let $\mathbf{S}(\mathbf{x}_0, \theta; \gamma)$ be the *stride function* that simulates the Kneed Walker from the initial state until the first ground contact; fixed points $\mathbf{S}(\mathbf{x}_0, \theta; \gamma) = \mathbf{x}_0$ are cyclic gaits. Also, let $V(\mathbf{x}_0, \theta; \gamma)$ and $L(\mathbf{x}_0, \theta; \gamma)$ be the speed and step length after simulation to the first ground contact. Thus, given target speed v, step length ℓ and ground slope γ, we minimize (9) subject to

$$\mathbf{S}(\mathbf{x}_0, \theta; \gamma) = \mathbf{x}_0, \ V(\mathbf{x}_0, \theta; \gamma) = v, \ L(\mathbf{x}_0, \theta; \gamma) = \ell. \quad (10)$$

This is solved using constrained optimization [18], with gradients approximated using finite differences. We do this for speeds between 3 and 7 km/h, step lengths from 0.5 to 1 meters, and ground slopes from $-4.3°$ to $4.3°$.

We find that optimal gaits, like those in Fig. 2, exhibit many important characteristics of natural human walking. For instance, a natural bend of the swing knee is clearly evident throughout the entire motion. Also, the stereotypical lean of the upper body can be seen, including a forward lean when climbing up a hill and a slight backwards lean when walking down. In a validation of passive dynamic models [5], the optimal parameters for the swing knee spring were small, suggesting a damped but otherwise passive joint.

3.4. Stochastic Prior Model

Our prior walking model based on the Kneed Walker assumes that plausible motions lie in the vicinity of the optimal gaits. First, for optimal gaits we find that the torque for most joints is well modeled with a subset of the spring parameters. This is significant as it reduces the number of hidden random variables. In particular, we fix the damping constant for the knee springs to be the median of the optimized damping constants for both legs, $d_{K1} = d_{K2} = 0.05$. Further, given the nearly passive nature of the swing knee in the optimal motions, we set $\kappa_{K2} = 0$. We also simplify the torso spring model by setting its resting length relative to the ground slope to $\phi_{To}^0 = -\gamma/3$. Finally, the torso damping constant is fixed at $d_{To} = 1.5$. This is much larger than
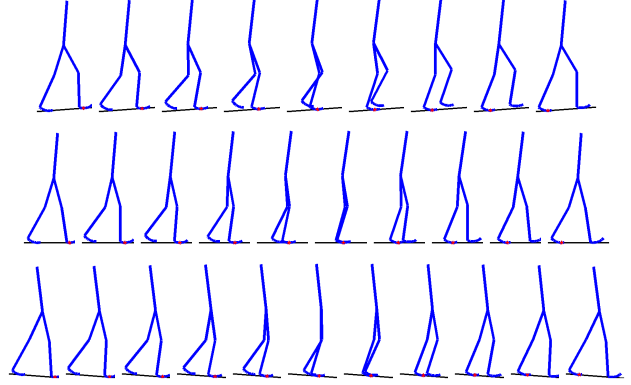

Figure 2. Three optimal gaits for the Kneed Walker; (Top) walking uphill $(4.3°)$, (Middle) on level ground, and (Bottom) downhill $(-4.3°)$. Note the knee bend at contact while walking up hill. There is also a subtle but important bend in the knee just after contact while walking downhill.

that found by the optimizations, to account for noise during tracking and other dynamic phenomenon not captured in the optimizations, such as speed changes which require the rapid dissipation of momentum.

The remaining stiffnesses are modeled as follows. For a joint j within a stride s there is an unknown mean stiffness $\bar{\kappa}_j(s)$. The prior over $\bar{\kappa}_j(s)$ is Gaussian with a mean and variance set roughly according to the optimizations with the exception that the means for the torso and stance knee spring stiffnesses are higher to account for differences between tracked motions and optimal cyclic gaits.[1] Within a stride, the stiffness at time t, $\kappa_j(t)$, is Guassian with mean $\bar{\kappa}_j(s)$ and variance σ_j^2.

Joint torques due to spring forces remain deterministic functions of stiffness parameters. When the swing leg hits the ground and support is transfered, the impulse ι_s is drawn from an Exponential distribution with a scale of 0.015.

To account for stylistic variations we also allow additive process noise in each joint torque, independent of the spring. The process noise for the jth torque at time t is

$$\eta_j(t) = \beta_j \eta_j(t-1) + s_j \zeta \qquad (11)$$

where $0 \leq \beta_j \leq 1$ is used to bias the torque process noise towards zero, s_j determines the scale of variation over time, and ζ is white with a Cauchy distribution. Beyond the joints specified in Fig. 1, we also allow a torque against the ground by the stance leg, which is also modelled using (11).

Finally, while the Kneed Walker is a 2D abstraction, we wish to perform human pose tracking with a 3D model. We therefore define a 3D kinematic model that is conditioned to be consistent with the Kneed Walker in the sagittal plane. Following [4], the kinematic model has 15 DOFs, comprising 3 DOF hip joints, hinge joints for knees and ankles,

---

[1]For instance, real motions tend to have more bend in the stance knee than seen in the optimizations which requires a stiffer spring to prevent the model from collapsing.

and the 3 DOF global position and 2 DOF for the orientation of the body. Pose variables that are not constrained by the Kneed Walker are modelled using (damped) 2nd-order Markov processes with zero-mean Gaussian acceleration.

To summarize, the model state at time t is given by $\mathbf{s}_t = (\theta_t, \bar{\eta}_t, \mathbf{x}_t, \mathbf{k}_t)$ where θ_t are the spring parameters, $\bar{\eta}_t$ is the process noise, $\mathbf{x}_t = (\mathbf{q}_t, \dot{\mathbf{q}}_t)$ is the dynamics state, and $\mathbf{k}_t$ denotes the kinematic DOFs. The model also defines a state transition density $p(\mathbf{s}_t | \mathbf{s}_{t-1})$ from which one can draw samples. That is, after sampling the dynamics parameters, $(\theta_t, \bar{\eta}_t)$, we deterministically simulate the dynamics to find $\mathbf{x}_t$. Then, we sample $\mathbf{k}_t$ given $\mathbf{x}_t$.

4. Tracking

Tracking is formulated as a filtering problem. With the Markov properties of the generative model above, and conditional independence of the measurements, one can write the posterior recursively, i.e.,

$$p(\mathbf{s}_{1:t} | \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t | \mathbf{s}_t) p(\mathbf{s}_t | \mathbf{s}_{t-1}) p(\mathbf{s}_{1:t-1} | \mathbf{z}_{1:t-1}) \quad (12)$$

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, and $p(\mathbf{s}_t | \mathbf{s}_{t-1})$ is the temporal model described above.

Likelihood: The 3D articulated body model comprises tapered ellipsoidal cylinders for the torso and limbs, the sizes of which are set manually. The likelihood is based on an appearance model and optical flow measurements.

The background model, learned from a small subset of images, includes the mean color (RGB) and intensity gradient at each pixel, with a 5×5 covariance matrix to capture typical color and gradient variations. Foreground pixels are assumed to be IID in each body part (i.e., foot, legs, torso, head). The observation density for each part is a Gaussian mixture, learned from the initial pose in the first frame.

Optical flow [9] is estimated at locations $\mathbf{p}$ on a coarse grid in each frame (e.g., see Fig. 3, row 2), using a robust M-estimator with non-overlapping support. The eigenvalues/vectors of the local 2×2 gradient tensor in the neighbourhood of each grid point give an approximate estimator covariance Σ. The observation density for a flow measurement, $\mathbf{v}(\mathbf{p})$, given the 2D motion predicted by the state , $\mathbf{u}(\mathbf{k}_t, \mathbf{p})$, is a heavy-tailed Student's t distribution; i.e.,

$$\log p(\mathbf{v}(\mathbf{p})|\mathbf{u}(\mathbf{k}_t, \mathbf{p})) = -\frac{\log |\Sigma|}{2} - \frac{n+2}{2} \log(1+e^2) + c \quad (13)$$

where $e^2 = \frac{1}{2}(\mathbf{v} - \mathbf{u})^T \Sigma^{-1}(\mathbf{v} - \mathbf{u})$, $n = 2$ is the degrees of freedom, and c is a constant. The camera is stationary for the experiments below, so the flow log-likelihood for measurements on the background is merely (13) 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 visible measurements for each part. To

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 yields the log-likelihood, $\log p(\mathbf{z}_t | \mathbf{s}_t)$, up to an additive constant.

Inference: We approximate the posterior by a weighted sample set $\mathcal{S}_t = \{\mathbf{s}_{1:t}^{(j)}, w_t^{(j)}\}_{j=1}^N$, where $w_t^{(j)}$ denotes the weight associated with the state sequence $\mathbf{s}_{1:t}^{(j)}$. Given the recursive form of (12), 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. To avoid severe particle depletion, following [7, 10], when the effective number of samples, $N_{eff,t} \approx (\sum_j (w_t^{(j)})^2)^{-1}$ becomes too small we re-sample the particle set using importance sampling.

In simple particle filters one re-samples states at time t in proportion to the weights (treating weights as the probabilities of a multinomial distribution); the new weights are then set to $1/N$. Here, following [4], rather than re-sample at the current time, we resample from a previous time $t - \tau_s$. This aims to re-sample before the onset of particle depletion. It also allows the proposal distribution to depend on future observations (i.e., those between $t - \tau_s$ and t), since the quality of a sample is not always immediately evident.

As a proposal distribution we use a mixture of two multinomials, one based on the weights at t, and one based on weights at $t - \tau_s$, with mixing probabilities γ and $1 - \gamma$. Importance re-weighting is then needed to maintain a properly weighted sample set. So the new weights are given by $w_{t-\tau_s}^{(j)}/(\gamma w_t^{(j)} + (1-\gamma)w_{t-\tau_s}^{(j)})$ (up to a constant so they sum to unity). Thus, most of the samples will correspond to probable states based on all information up to time t. The remaining samples are probable states according to the posterior at time $t - \tau_s$. With this form of importance sampling we re-sample less frequently, and the tracker is more efficient. In practice we use $\tau_s = 3$ and $\gamma = 0.95$.

5. Experimental Results

We now describe experimental results with the Kneed Walker on several image sequences of people walking on level ground, with occlusion and changes in speed, and on hills. In all experiments, we have roughly calibrated the camera parameters and the location of the ground plane. We use 5000 particles with a resampling threshold of 500. The initial state is specified coarsely in the first frame, but with a large covariance. One could also initialize the tracker with discriminative methods (e.g., [1, 26]).
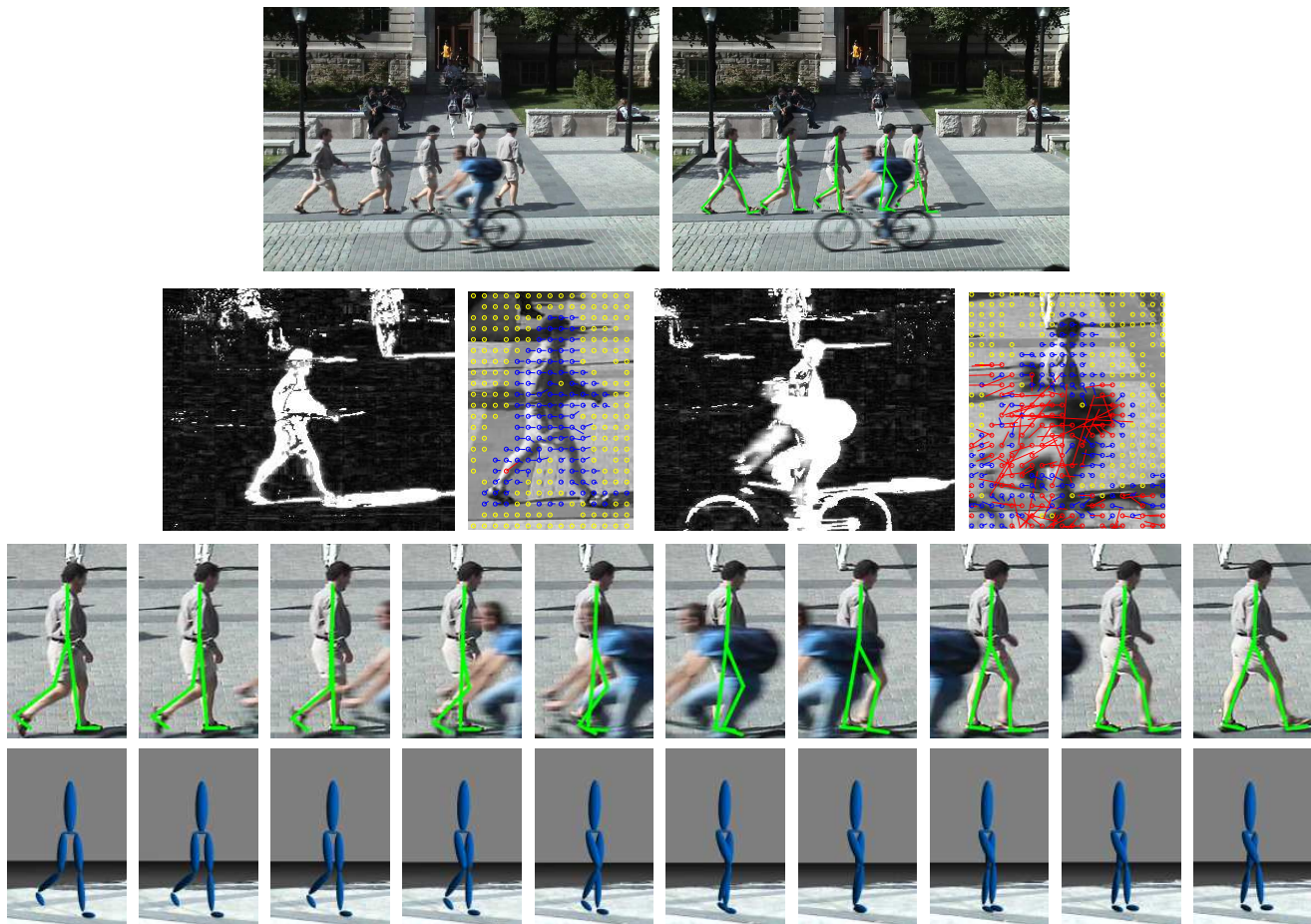
Figure 3. (Top row) Composite of image sequence showing a walking subject and an occluding cyclist. The green stick figure in the right composite depicts on the MAP estimate of the pose on selected frames. (Second row) Examples of the background likelihood and optical flow measurements (yellow, blue, and red flow measurements correspond to slow, moderate and fast speeds). (Bottom two rows) Cropped frames around occlusion. The green skeleton and blue 3D rendering are the recovered MAP trajectory for 10 consecutive frames.

Experiment 1. Figure 3(top-left) shows composite images of a walking subject on nearly level ground. The scene has harsh shadows, background clutter, and a cyclistic that occludes the subject. Fig. 3(2nd row) shows cropped examples of image measurements, including optical flow estimates and the negative log likelihood of the background, early and then later in the sequence during the occlusion. They are particularly noisy during the occlusion.

Despite the occlusion and noisy measurements, the estimated motion with the Kneed Walker model agrees with the subject's gait. The green stick figure in Fig. 3(top-right) depicts the projection of the 3D kinematic model for the MAP state sequence obtained by the particle filter. More detail can be seen in the cropped images in the bottom two rows of Fig. 3. These cropped images show the recovered MAP estimates for 10 consecutive frames through the occlusion. The last row shows a 3D rendering of the model from a different camera viewpoint to illustrate the 3D pose in each frame. The video in the supplemental material demonstrates

that the recovered motion not only matches the image data, but is also natural in its appearance.

Experiment 2. With the richer dynamics of the Kneed Walker, we find that the knees and torso are estimated more accurately than with the Anthropomorphic Walker. For example, Fig. 4 shows results on a sequence used in [4] in which the subject slows down from roughly 7 km/hr to 3 km/hr. The cropped images in the middle and bottom rows of Fig. 4 show MAP estimates every two frames for the Kneed Walker and the Anthropomorphic Walker. The same likelihood and number of particles were used in both cases.

The Kneed Walker estimates the knee pose more accurately. Interestingly this is the result of a simpler prior model. That is, where Brubaker et al. [4] use a second-order kinematic smoothness model with an ad hoc dependance on the angle between the legs, our model uses a passive knee with a small amount of noise. The knee bend at the beginning of a stride and the straightening towards the end is a fundamental property of the physics of the Kneed Walker.
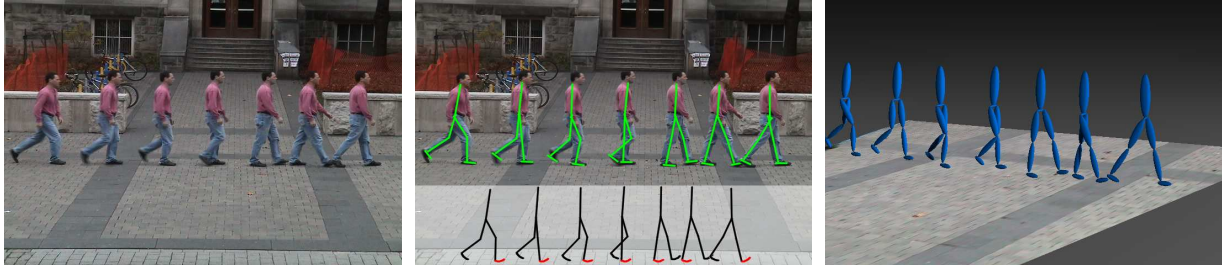
Figure 4. (Top) Composite images showing, for selected frames, the original sequence, the MAP kinematics (green stick figure) and dynamics (superimposed black stick figure), and a 3D rendering of the model pose froma different camera viewpoint. (Middle) Tracking results using the Kneed Walker. (Botttom) Tracking results with the Anthropomorphic Walker [4].

Figure 5. Tracking up hill. (Top) Composites of the input sequence, recovered MAP motion in 2D and 3D. (Bottom) Zoomed in views of every other frame of a subsequence. The hill has a $10.0°$ grade, and the subject is walking $40°$ out of the image plane.

Experiment 3. The last experiment involves a subject walking up an incline of approximately $10°$ which is close to the steepest grade up which cars drive. Because our optimizations only included slopes up to $4.3°$, we adjusted the damping constant on the swing knee and torso and set the mean stiffnesses for the stance knee and torso to be larger to account for the larger slope. All other parameters were identical to those in other experiments.

The results in Fig. 5 show that the tracker faithfully recovers the 3D motion in this sequence despite the large dif-

ference in the gait. In particular, Fig. 5 (top-right) shows the recovered motion from a very different viewpoint looking uphill at the walking figure. One can clearly see the significant knee bend at contact that is characteristic of people walking up hills. Also evident is a slight lean of the torso into the slope. Because the camera is about $40°$ away from a sagittal view, both the knee bend and the torso lean would be difficult to recover purely from the image evidence.

6. Conclusions and Future Work

This paper introduced the Kneed Walker, a complex physics-based model of bipedal locomotion. As part of this model, we introduced a method for handling joint limits in an efficient but physically realistic manner. We demonstrated that a wide range of realistic walking motions on sloped surfaces and level ground could be found through the constrained optimization of energy. When used in a tracker with a simple control strategy, the Kneed Walker was able to recover subtle aspects of motion such as knee bend and torso lean, even when these were not strongly indicated by the image evidence. Further, as seen in the 3D animations in the supplemental video, the recovered motions exhibit a high degree of realism.

In the future we hope to combine the wide range of motions found through optimization with motion capture data to learn better stochastic control strategies while retaining the diversity of motions found in optimizations. Finally, we expect that a direct generalization of the model and techniques presented here will allow us to capture running and other, more energetic, human motions.

References

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

[2] A. Bissacco. Modeling and learning contact dynamics in human motion. *CVPR*, v. 1, pp. 421–428, 2005.

[3] B. Brogliato, A. ten Dam, L. Paoli, F. Genot, and M. Abadie. Numerical simulation of finite dimensional multibody non-smooth mechanical systems. *Appl. Mech. Eng. Reviews*, 55(2):107–150, 2002.

[4] M. A. Brubaker, D. J. Fleet, and A. Hertzmann. Physics-based person tracking using simplified lower-body dynamics. *CVPR*, pp. 1–8, 2007.

[5] S. Collins, A. Ruina, R. Tedrake, and M. Wisse. Efficient Bipedal Robots Based on Passive-Dynamic Walkers. *Science*, 307(5712):1082–1085, 2005.

[6] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealled particle filtering. *CVPR*, V. II, pp. 126–133, 2000.

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

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

[9] D. Fleet and Y. Weiss. Optical flow estimation. In *Mathematical Models of Computer Vision: The Handbook*, pp. 239–258, Springer, 2005.

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

[11] A. D. Kuo. A Simple Model of Bipedal Walking Predicts the Preferred Speed–Step Length Relationship. *J. Biomech. Eng.*, 123(3):264–269, 2001.

[12] A. D. Kuo. Energetics of Actively Powered Locomotion Using the Simplest Walking Model. *J. Biomech. Eng.*, 124:113–120, 2002.

[13] R. Li, T.-P. Tian, and S. Sclaroff. Simultaneous learning of non-linear manifold and dynamical models for high-dimensional time series. *ICCV*, 2007.

[14] C. K. Liu, A. Hertzmann, and Z. Popović. Learning physics-based motion style with nonlinear inverse optimization. *ACM Trans. Graphics*, 24(3):1071–1081, 2005.

[15] T. McGeer. Passive Walking with Knees. *ICRA*, V. 3, pp. 1640–1645, 1990.

[16] T. McGeer. Dynamics and Control of Bipedal Locomotion. *J. Theoretical Biol.*, 163:277–314, 1993.

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

[18] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Res., 2nd ed., 2006.

[19] M. Raibert. *Legged Robots That Balance*. MIT Press, Cambridge, 1986.

[20] M. H. Raibert and J. K. Hodgins. Animation of dynamic legged locomotion. *SIGGRAPH*, pp. 349–358, 1991.

[21] G. E. Robertson, G. Caldwell, J. Hamill, G. Kamen, and S. Whittlesey. *Research Methods in Biomechanics*. Human Kinetics, 2004.

[22] L. Shampine. Conservation laws and the numerical solutions of ODEs, II. *Comp. Math. Applic.*, 38:61–72, 1999.

[23] L. Shampine, S. Thompson, J. A. Kierzenka, and G. D. Byrne. Non-negative solutions of ODEs. *Appl. Math. Comp.*, 170:56–569, 2005.

[24] H. Sidenbladh, M. Black, and D. J. Fleet. Stochastic Tracking of 3D Human Figures Using 2D Image Motion. *ECCV*, V. 2, pp. 702–718, 2000.

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

[26] C. Sminchisescu, A. Kanaujia, and D. Metaxas. BM^{3}E : Discriminative density propagation for visual tracking. *PAMI*, 29(11):2030–2044, 2007.

[27] C. Sminchisescu and B. Triggs. Kinematic jump processes for monocular 3D human tracking. *CVPR*, pp 69–76, 2003

[28] A. Thayananthan, R. Navaratnam, B. Stenger, P. Torr, and R. Cipolla. Multivariate relevance vector machines for tracking. *ECCV 2006*, V. 3, pp. 124–138. Springer, 2006.

[29] R. Urtasun, D. J. Fleet, and P. Fua. 3D People Tracking with Gaussian Process Dynamical Models. *CVPR*, V 1, pp. 238–245, 2006.

[30] R. Q. van der Linde and A. L. Schwab. Lecture Notes Multibody Dynamics B, wb1413. Lab. for Eng. Mech., Delft Univ. of Technology, 2002.

[31] A. Witkin and M. Kass. Spacetime Constraints. *SIGGRAPH*, V. 22, pp. 159–168, 1988.

[32] C. R. Wren and A. Pentland. Dynamic models of human motion. *IEEE Face and Gesture*, pp. 22-27, 1998.