Physics-Based Human Pose Tracking

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

1 Introduction

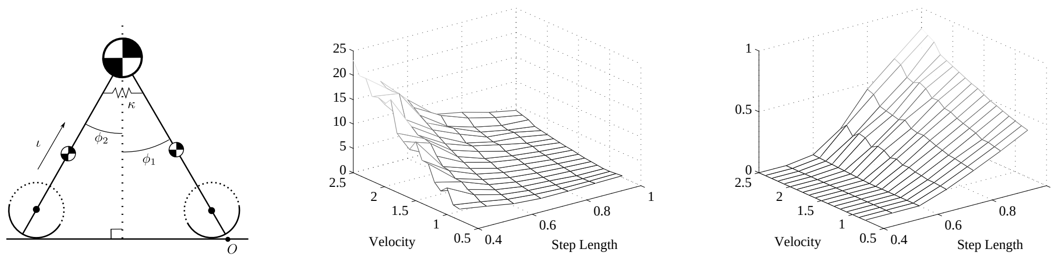
Human motion tracking from single-view 2D measurements contains inherent ambiguities, due to depth ambiguities and the indirect relation of pixel intensities to 3D poses. Prior models of human poses and motion show great promise in resolving these difficulties, by biasing inference toward the most probable configurations. The main challenge is to describe models that accurately describe how people move: in general, we expect that better prior models will give better tracking results.

Current visual tracking algorithms have employed exclusively *kinematic* prior models: models that describe likelihood directly in pose space. The simplest such models include the use of articulated models, joint limits, and temporal smoothness assumptions (e.g., [4, 8, 14, 15, 17]), which provide useful contraints on motion, but do not sufficiently capture the nonlinearities of human motion. A more advanced approach is to learn an activity-specific models from motion capture data (e.g., [3, 7, 11, 13, 16]). In essence, such models describe motion likelihood by comparison to the training poses: motions similar to the training data are considered "more likely" by the model. In highly-constrained scenarios, these models typically yield results with approximately correct body configurations, but are nonetheless physically implausible. Two of the most common errors are "footskate," in which a foot in contact with the ground appears to float in space, and out-of-plane rotations of the body that violate balance. Furthermore, we expect purely kinematic models to have difficulty generalizing beyond the training dataset: for every motion being tracked, there must be a very similar motion included in the training database.

In this abstract, we introduce and evaluate a physics-based model of human motion. The long-term goal of our work is to build accurate prior models by incorporating domain knowledge of physics. To this end, we employ a simplified (low-dimensional) representation of human dynamics (based on models from biomechanics and computer animation) that constrains the motion of a full-body kinematic model. This generative model of human walking accurately represents physical properties of motion such as balance, ground contact, and variations in style due to changes in speed, step-length, and mass-distributions. Our model achieves these goals without requiring any training data. We describe experiments in on-line particle-based tracking of 3D human pose and motion using the HumanEva dataset.

2 A Dynamic Prior Model of Human Walking

Building such prior models, like the control of complex dynamical systems in general, is extremely challenging. Nonetheless, work in robotics and biomechanics suggests that the dynamics of bipedal walking may be well described with relatively simple passive dynamic models [1]. Such abstract models, like that explored in [6], exhibit stable, bipedal walking as a natural limit cycle of their dynamics. Compared to traditional approaches in robotics, such as used by Honda's humanoid robot *Asimo*, these passive-dynamic models are significantly more efficient and exhibit a more human-like gait [10]. One such model, called the *Anthropomorphic Walker* [5], provides the basis for the dynamic model used in this work.



(a) The hatched circles represent centers of mass.

(b) κ as a function of speed and step length

(c) ι as a function of speed and step length

Figure 1: The Anthropomorphic Walker. Figures (b) and (c) illustrate the flexibility and expressivity of the models control parameters. Noise in the plots is due to numerical errors.

2.1 Anthroporphic Walker

The anthropomorphic walker, shown in Figure 1, is a powered variant of McGeer's passive dynamic walker [6]. It is capable of walking downhill with only gravity acting upon it. With simple control strategies and modest forces it also exhibits a wide range of walking motions on level ground, with human-like gaits and energy consumption. Models of this sort have also been developed with knees, with running gaits, and several 3D robotic variants of the basic model have been successfully built.

The generative walking model developed here is based on dynamics simulation and stochastic forces. To simulate the anthropomorphic walker one needs the equations of motion of the model. They can be generally expressed as

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

where $\mathbf{M}$ is the generalized mass matrix, $\mathbf{F}$ is the generalized force vector, $\ddot{q}$ is the acceleration of the generalized coordinates $q = [\phi_1, \phi_2]^T$ and ϕ_1 and ϕ_2 are the 2D global orientations of the stance and swing legs. $\mathbf{F}$ constitutes both force due to gravity and applied forces which are parameterized by a spring constant κ as described in [5].

Collisions When the swing foot hits the ground a collision occurs and support is transfered between legs. Ground collisions are modelled as impulsive and perfectly inelastic, which results in an instantaneous change in velocity. To allow for the "toe-off" characteristic of human walking, an impulse with magnitude ι is applied at the time of collision.

The dynamical consequence of the collision is determined by a system of equations relating the instantaneous motions immediately before and after the collision, along with the impulsive toe-off:

$$\mathbf{M}(q)\dot{q}^{+} = \mathbf{C}(q)\dot{q}^{-} + \iota I(q)$$

where $\mathbf{C}$ is the collision matrix, $\mathbf{M}$ is the post-collision generalized mass matrix, I is the generalized applied impulse vector and $\dot{q}^{-}$ and $\dot{q}^{+}$ are the pre- and post-impact velocities.

Stochastic Control The equations of motion given above are deterministic; i.e., one can simulate the motion by integrating the equations of motion, given the initial state, the spring constant κ and the magnitude of the impulsive toe-off ι. Figures 1(b) and 1(c) illustrate how these control parameters relate to the speed and step length of the gait. In the context of tracking, these parameters are unknown and are treated as hidden random variables.

2.2 Kinematic Model

While the anthropomorphic walker captures some of the key physical properties of interest, namely, balance and contact, a more complex kinematic model is also needed. The dynamic model, for example, has no knees and it specifies dynamics only in the instantaneous (2D) plane of motion. In order to track people with a richer skeletal model having more degrees of freedom, and in 3D, we use a kinematic model that is conditioned on the underlying dynamics. This approach is similar to the model simplification technique used in [9] for physically realistic motion capture editing.

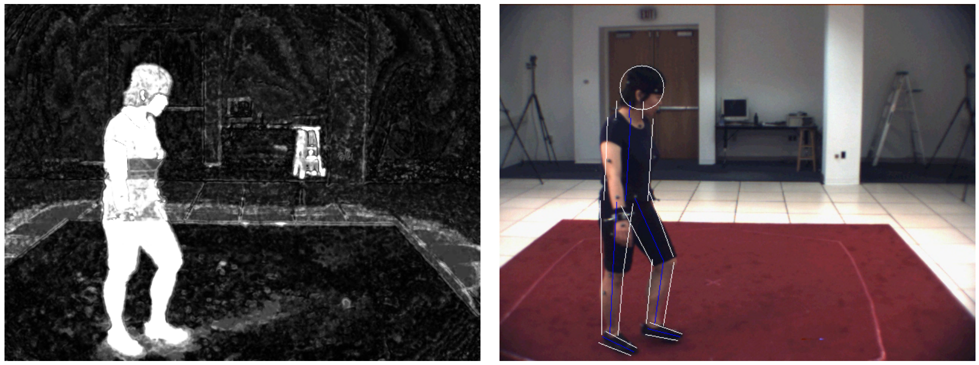


Figure 2: A frame of the tracking results. *Left:* The log of the foreground probability over the background probability is shown to illustrate the observation model. *Right:* The pose of the MAP trajectory is displayed.

The higher degree of freedom kinematic model has a ball and socket joint at the hips, and hinge joints at the knees and ankles. The upper body, though not explicitly modeled in the dynamics, is a crucial feature for tracking. As such the torso, neck and head are modeled as a single rigid object which is attached to the legs by a hinge joint. The dynamic model specifies the orientation of the hip and the point of contact between the foot and the ground. However, the remaining degrees of freedom are unspecified and modeled with a joint angle limited, second-order temporal smoothness prior.

2.3 Observation Model

The geometry of the individual segments are tapered elliptical cylinders except for the head which is modeled as a sphere. The size of the cylinders are fixed and fit by hand to match the subject. The likelihood of an observed image given a kinematic pose consists of the edge-based likelihood used in [8], a uniform foreground likelihood and a stochastic background likelihood. The log of the ratio of the foreground to background probability can be seen in Figure 2(a).

2.4 Inference

3D tracking is performed online, one frame at a time, with a simple particle filter [2]. At each time step we first sample from the random dynamics variables and then simulate the model. We then sample from the kinematic random variables given the dynamics and the previous kinematic state. The likelihood of the each particle is evaluated and used to properly weight the particle set. Inference is hand initialized based on several views of the starting frame.

3 Results and Conclusions

Online inference is performed using sequential Monte Carlo using 5000 particles and resampling when the effective sample size drops below 1250. Let $P(p_{1:t}|o_{1:t})$ be the estimated posterior of pose up to time t given the observations up to time t. Then the expected error at time t given observations up to time t is

$$E[D(X_t, \hat{X}(p_t))|o_{1:t}] = \int D(X_t, \hat{X}(p_t))P(p_{1:t}|o_{1:t})dp_{1:t}$$

where X_t are the ground truth virtual markers at time t, $\hat{X}(p)$ are the virtual markers of pose p and $D(X, \hat{X})$ is the proposed evaluation metric in [12].

We can also compute an estimate of the error if tracking were performed in batch. Let $\tilde{p}_{1:t} = \arg\max_{p_{1:t}} P(p_{1:t}|o_{1:t})$ be the MAP trajectory up to time t and $\tilde{p}_s$ be the pose at time $s \in [1, t]$ in trajectory $\tilde{p}_{1:t}$. Then

$$\tilde{e}_s = D(X_s, \hat{X}(\tilde{p}_s))$$

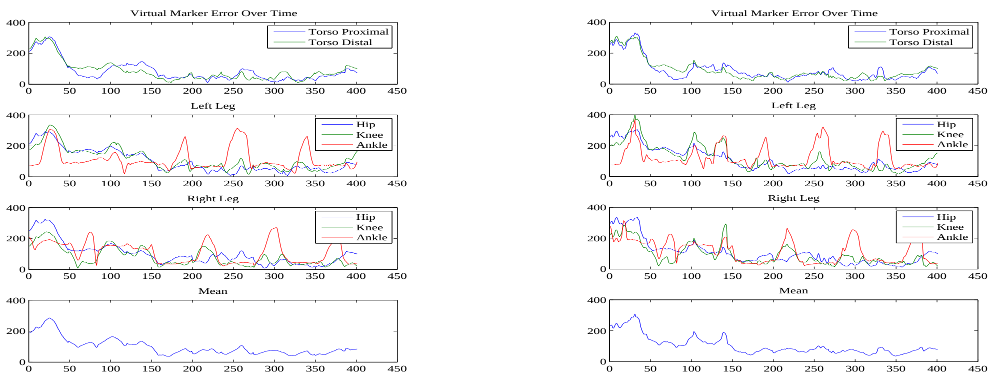


Figure 3: Evaluation Criteria over 400 frames. Units are in millimeters. *Left:* MAP trajectory error. The overall mean error was 99 ± 55.7. *Right:* Expected error. The overall mean error was 104 ± 59.6.

is an estimate of the error at time s given all the observations up to time t. This ensures that the error is computed over a dynamically consistent motion.

We performed inference on the first walking sequence of subject 1 from the HumanEvaI dataset [12] using color camera 1 as the only view. The results can be seen in Figure 2 and evaluation criteria can be seen in Figure 3. Of particular note with the evaluation criteria is the out-of-phase nature of the errors in the right and left ankle. The spike in ankle error occurs in the swing leg and can be directly related to the lack of a strong dynamic model for the knee. In contrast, the errors on the stance leg are nearly constant over time which can be seen from the troughs in the leg error plots.

References

[1] Steve Collins, Andy Ruina, Russ Tedrake, and Martijn Wisse. Efficient Bipedal Robots Based on Passive-Dynamic Walkers. *Science*, 307(5712):1082–1085, 2005.

[2] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208, July 2000.

[3] A. Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Proc. CVPR*, volume 2, pages 681–688, 2004.

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

[5] Arthur D Kuo. Energetics of actively powered locomotion using the simplest walking model. *Journal of Biomechanical Engineering*, 124:113–120, February 2002.

[6] Tad McGeer. Passive dynamic walking. *International Journal of Robotics Research*, 9(2):62–82, 1990.

[7] V. Pavlovic, J.M. Rehg, Tat-Jen Cham, and K.P. Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *Proc. ICCV*, volume 1, pages 94–101, 1999.

[8] E. Poon and D.J. Fleet. Hybrid monte carlo filtering: edge-based people tracking. In *Proc. Workshop on Motion and Video Computing*, pages 151–158, 2002.

[9] Zoran Popović and Andrew Witkin. Physically based motion transformation. In *Proc. SIGGRAPH*, pages 11–20, 1999.

[10] Gill A. Pratt. Legged robots at MIT: what's new since Raibert? *Robotics & Automation Magazine, IEEE*, 7(3):15–19, 2000.

[11] Hedvig Sidenbladh, Michael J. Black, and David J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *Proc. ECCV*, volume 2, pages 702–718, 2000.

[12] Leonid Sigal and Michael J. Black. HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion. Technical Report CS-06-08, Brown University, 2006.

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

[14] C. Sminchisescu and A. Jepson. Variational mixture smoothing for non-linear dynamical systems. In *Proc. CVPR*, volume 2, pages 608–615, 2004.

[15] C. Sminchisescu and B. Triggs. Kinematic jump processes for monocular 3d human tracking. In *Proc. CVPR*, volume 1, pages 69–76, 2003.

[16] Raquel Urtasun, David J. Fleet, Aaron Hertzmann, and Pascal Fua. Priors for people tracking from small training sets. In *Proc. ICCV*, volume 1, pages 403–410, 2005.

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