Dynamical Models for People Tracking

Marcus Brubaker

University of Toronto

Supervisor: David Fleet

People Tracking

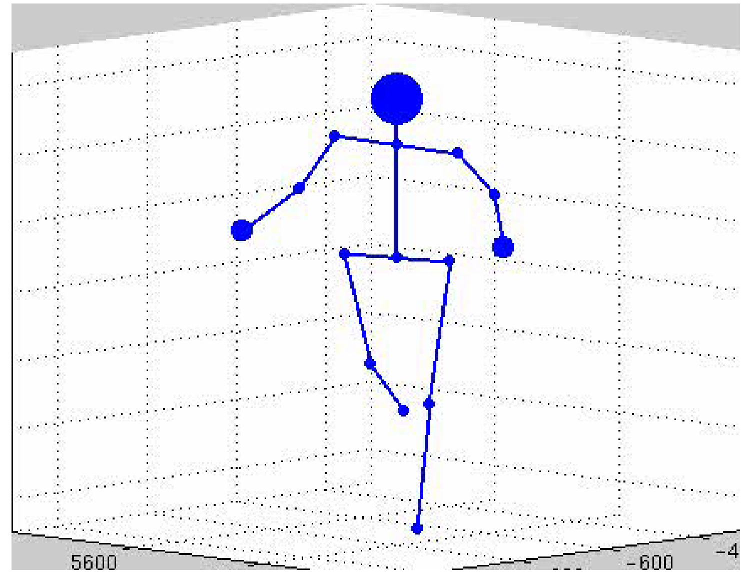
Human pose and motion are ambiguous in video

- Occlusion, reflection, resolution, symmetry
- Priors are needed to help resolve these ambiguities

Kinematic models have been used extensively to constrain tracking and pose estimation

- Model of joint angle limits and of typical poses / motions
- Does not easily model environmental interactions and other physical subtleties of motion, leading to errors in tracking (e.g., out of plane rotation, balance irregularities, footskate, …)

Kinematic models



[Poon & Fleet, 2002]

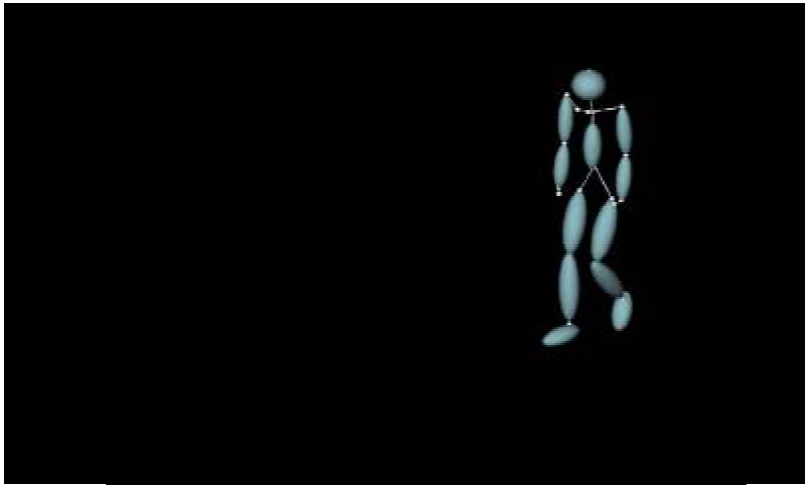
- Kinematics: linear, 2nd-order Markov model with Gaussian process noise and joint angle limits
- Observations: image edge (steerable pyramid)
- Inference: hybrid Monte Carlo particle filter

Kinematic models

[Urtasun, Fleet, Hertzmann & Fua, 2005]

- Kinematics: nonlinear probabilistic model of latent pose manifold with linear 2nd-order Markov model
- Observations: tracker 2D body parts (WSL tracker)
- Inference: MAP estimation (hill-climbing)

Kinematic models



[Urtasun, Fleet & Fua, 2006]

- Kinematics: Gaussian process latent variable dynamical model
- Observations: tracker 2D body parts (WSL tracker)
- Inference: MAP estimation in sliding window (hill climbing)

Learning prior models

Why are kinematic prior models hard to learn?

- Huge space of possible independent motions
- Environmental interactions make the space much larger
- Changing physical parameters can significantly change the motion (e.g., mass, stiffness, …)

Collecting enough mocap data appears impossible

Physics-based dynamical models

Why dynamics?

- Contact (action / reaction)
- Forces
- Changing physical parameters

Even silly walks obey basic physical properties.

Physics-based dynamical models

However, dynamics of complex physical models are hard to control

- Two possible solutions: engineering and abstraction

Active control strategies in robotics typically use ZMP-based stability criteria

- Highly inefficient
- Characteristically inhuman motion
- Complex to implement

Kawada Industries HRP-2, Robodex 2003

Models of human locomotion

Passive dynamics

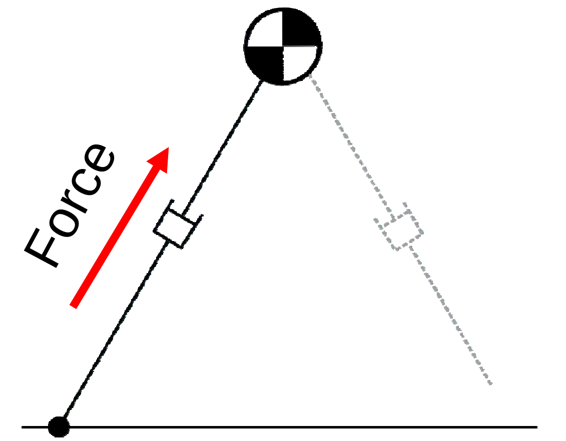
- Efficient, human like walking can be obtained with simple models
- Based on simple, abstracted models of human locomotion
- Expresses many salient characteristics of human walking

Cornell powered biped July 2003. Steve Collins & Andy Ruina. 11 watts total, 3 watts mechanical.

Models of human locomotion

The Monopode

- Very general, widely applicable model
- Capable of exhibiting bipedal walking, running, standing and jumping
- Also used to model cockroaches, quadrupeds, kangaroos, etc



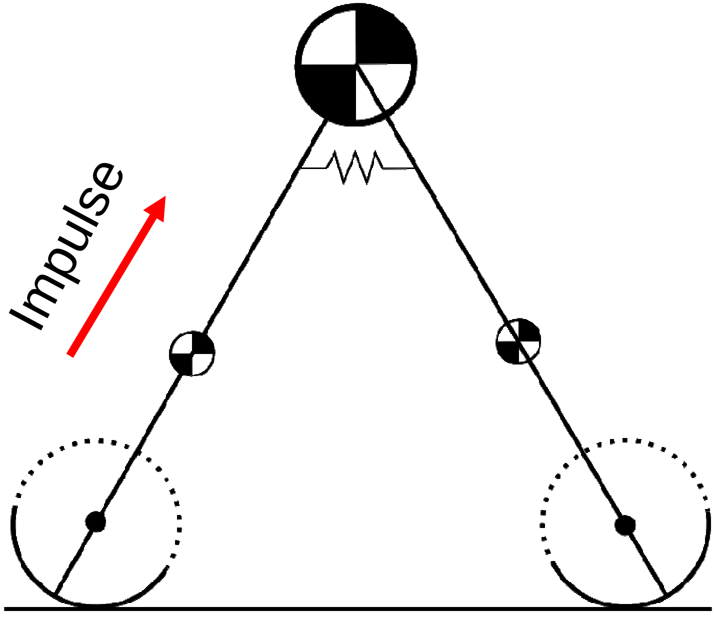
[Srinivasan and Ruina 2006; Blickhan and Full 1993]

Limitations of the monopode

- 2D model with a point-mass for the body
- Properties of support transfer not modeled
- No meaningful model of the swing leg

- Legs are mass-less, prismatic joints
- Moves (roughly) like an inverted pendulum between support transfers
- Impulsive forces act on the mass at support transfer

Models of human locomotion



[McGeer 1990; Kuo 2001, 2002]

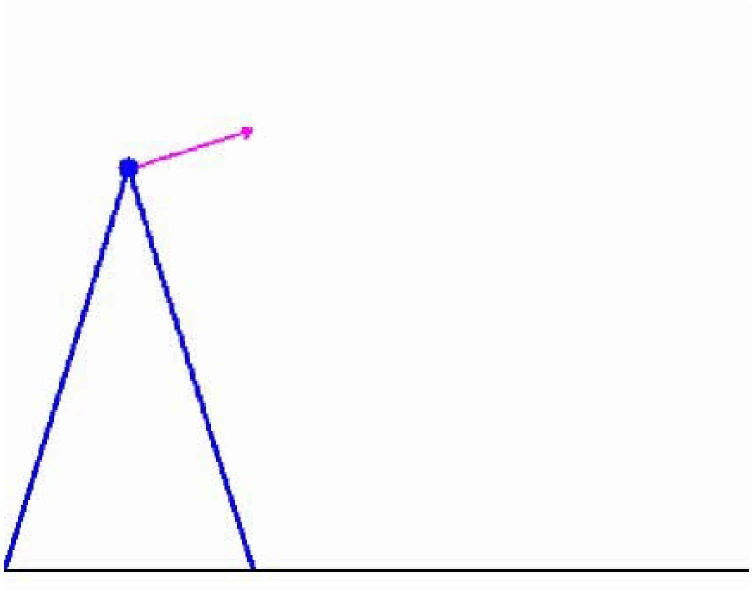
The Anthropomorphic Walker

- 2D model with a point-mass at the hip for the torso
- Small masses for the legs and rounded feet
- Torsional spring between the legs
- Can walk completely passively

Models of human locomotion

How do we use these models?

- The model parameters (i.e., leg length and mass distribution) define a set of equations of motion
- For a set of applied forces and an initial condition, the equations of motion are integrated to find the motion of the model



Components of a stochastic dynamical model

Need a way to express motion diversity (style, speed, …) through a stochastic model

- Can't change the physics, but we can let the forces be stochastic

Use biomechanics to suggest sensible ways to do this

- For the monopode, can apply noisy force during support and random impulses at support transfer
- For the anthropomorphic walker, can use a noisy spring constant model and random impulses at support transfer

Models of human locomotion

Dynamics (partially) constrain pose parameters:

- Stance Position
- Global leg orientation (at least for the stance leg)

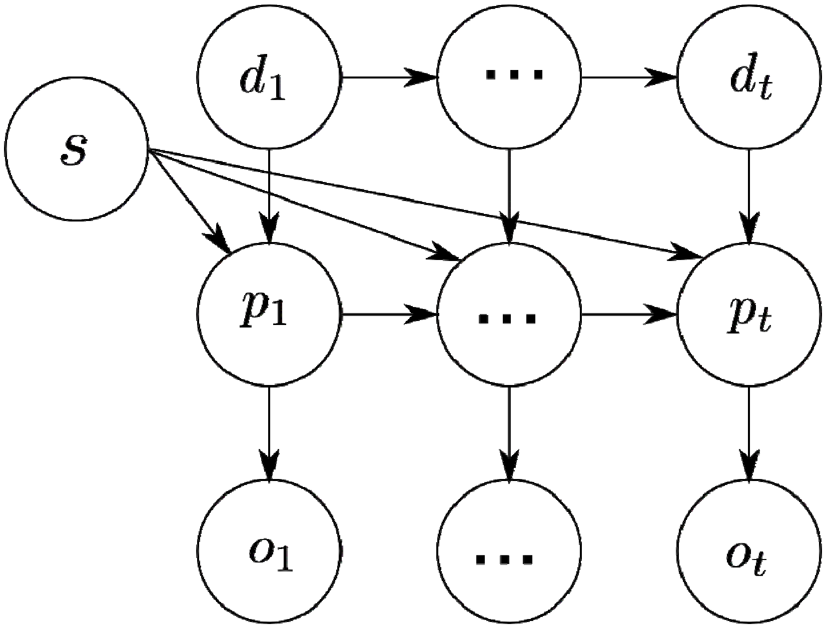
Model parameters:

- Per-person scale parameters used to model density over segment lengths
- Assume fixed mass distribution

These models have no hips, knees, ankles, upper body, etc

- Unconstrained pose variables modeled as 2^{nd}-order Markov

Generative model for people tracking



d_t - Abstracted dynamics, including the leg angles and velocities, forces, stance leg and positions, etc.

p_t - Pose, including segment sizes and joint angles

o_t - Observations

s - Person-specific scale parameters for segment lengths

Bayesian people tracking

Image Observations: $o_{1:t} \equiv (o_1, ..., o_t)$

State: $\phi_t = [d_t, p_t]$

dynamics pose

Posterior (Filtering) Distribution:

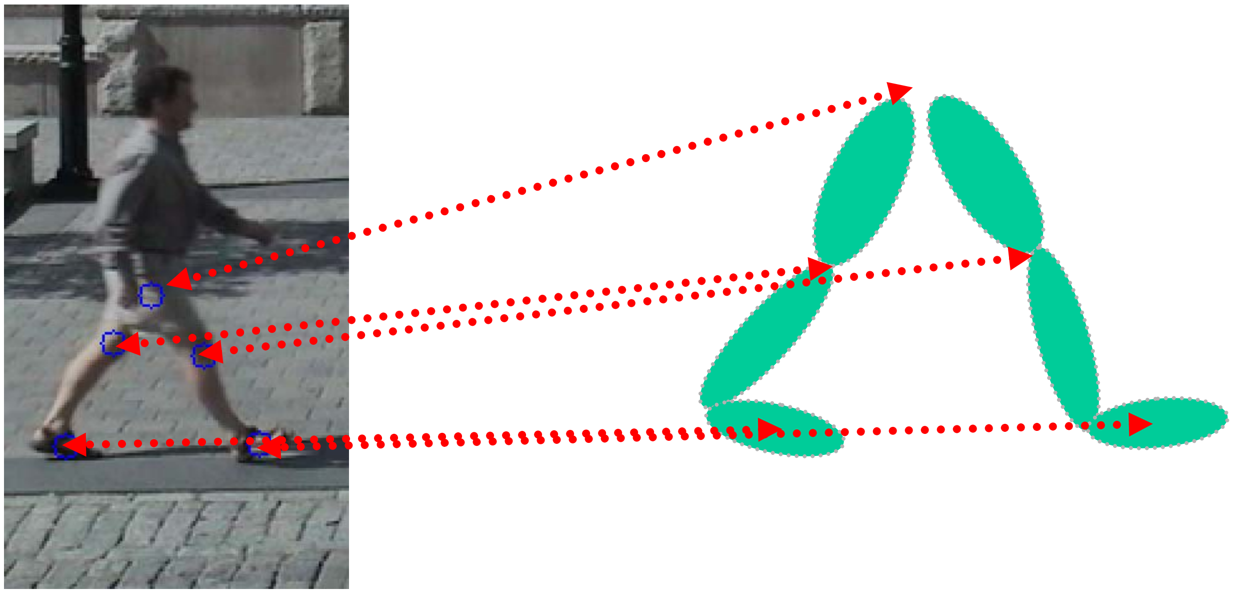
$$p(\phi_t \,|\, \mathbf{o}_{1:t}) \;\propto\; p(\mathbf{o}_t \,|\, \phi_t)\; p(\phi_t \,|\, \mathbf{o}_{1:t-1})$$

likelihood prediction

Online Inference:

- Particle filter with the prediction density as the proposal distribution when re-sampling
- Re-sampling occurs only when the effective number of particles drops below a threshold

Observation likelihood



2D positions of J points are tracked (up to IID Gaussian noise):

$$-\ln p(\mathbf{o}_t \mid \phi_t) \;=\; \sum_{j=1}^{J} \frac{||o_t^j - T(\phi_t^j)\,||^2}{2\sigma_j^2}$$

$T(\phi_t^j)$ is the perspective projection of point j at time t.

o_t^j is the associated image measurement

Experimental results

Calibrated video with known ground plane

Hand labeled data

- Could use tracks from WSL or other image trackers

Manual initialization at first frame

Experiment 1: Walking

# of particles: 500 (~30fps)

Resampling Threshold: 50

Experiment 2: Changing Direction

# of particles: 5000 (~5fps)

Resampling Threshold: 300

Experiment 3: Occlusion

Missing data: 30 sequential frames missing points on both legs

# of particles: 500 (~30fps)

Resampling threshold: 50

The End