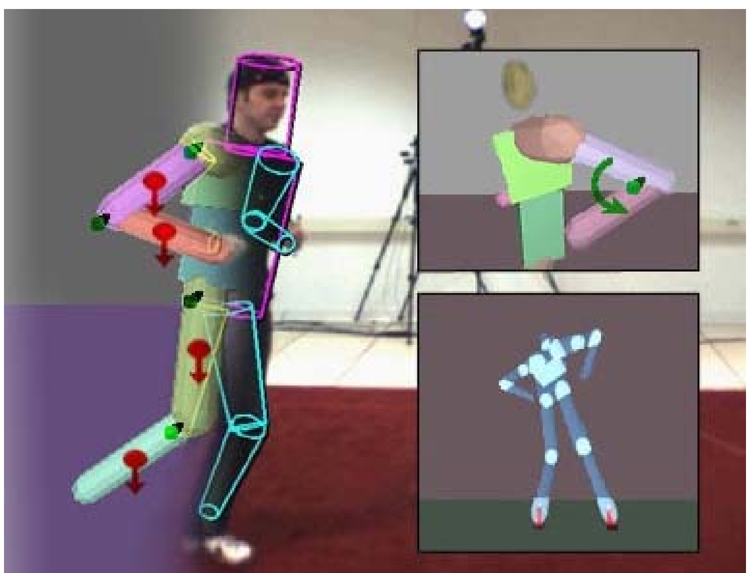
Physical Simulation for Probabilistic Motion Tracking

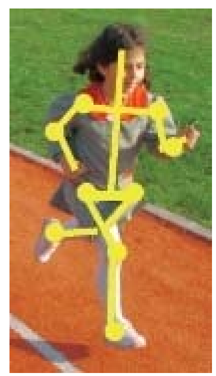
Marek Vondrak[*] **Leonid Sigal**[‡] **Chad Jenkins**[*]



[*]Brown University
[‡]University of Toronto

Articulated Tracking

- **Given:** observed image sequence $\{\mathbf{I}_\mathbf{f}\}$
- **Infer:** kinematic poses $\{\mathbf{q}_\mathbf{f}\}$ over time



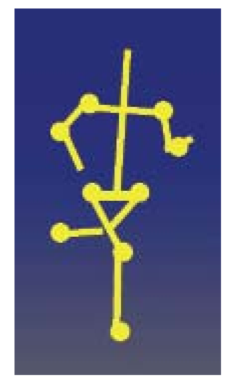
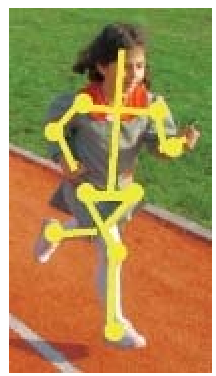


Image $\mathbf{I}_\mathbf{f}$ Frame **f** Pose $\mathbf{q}_\mathbf{f}$

Articulated Tracking

- **Given:** observed image sequence $\{\mathbf{I_f}\}$
- **Infer:** kinematic poses $\{\mathbf{q_f}\}$ over time



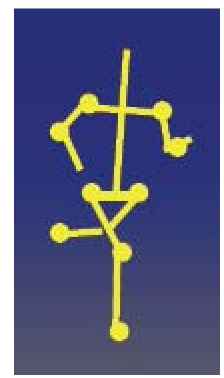


Image $\mathbf{I_f}$ Frame **f** Pose $\mathbf{q_f}$

- **Such that: $\{\mathbf{q_f}\}$** are physically plausible

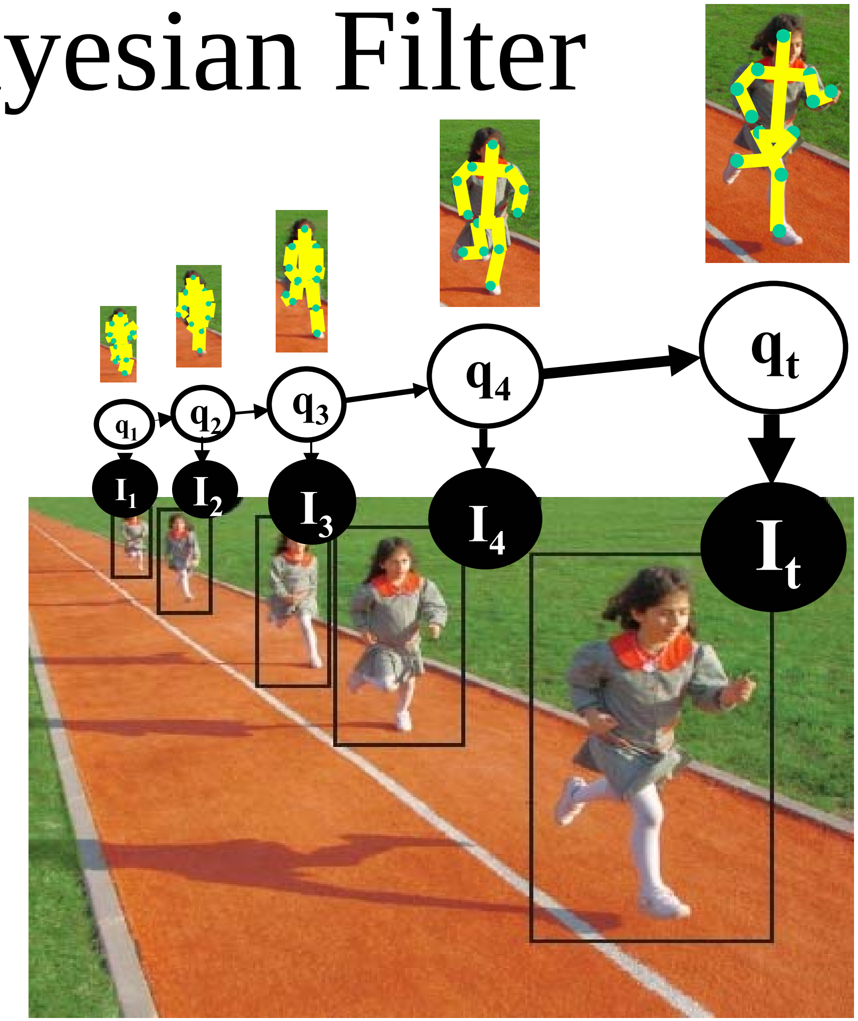
Kinematic Bayesian Filter

[A. Blake, M. Isard, ’98]

[Deutscher et. at., ’00]

– Motion model
  - $p(q_f \mid q_{f-1})$

– Likelihood model
  - $p(I_f \mid q_f)$



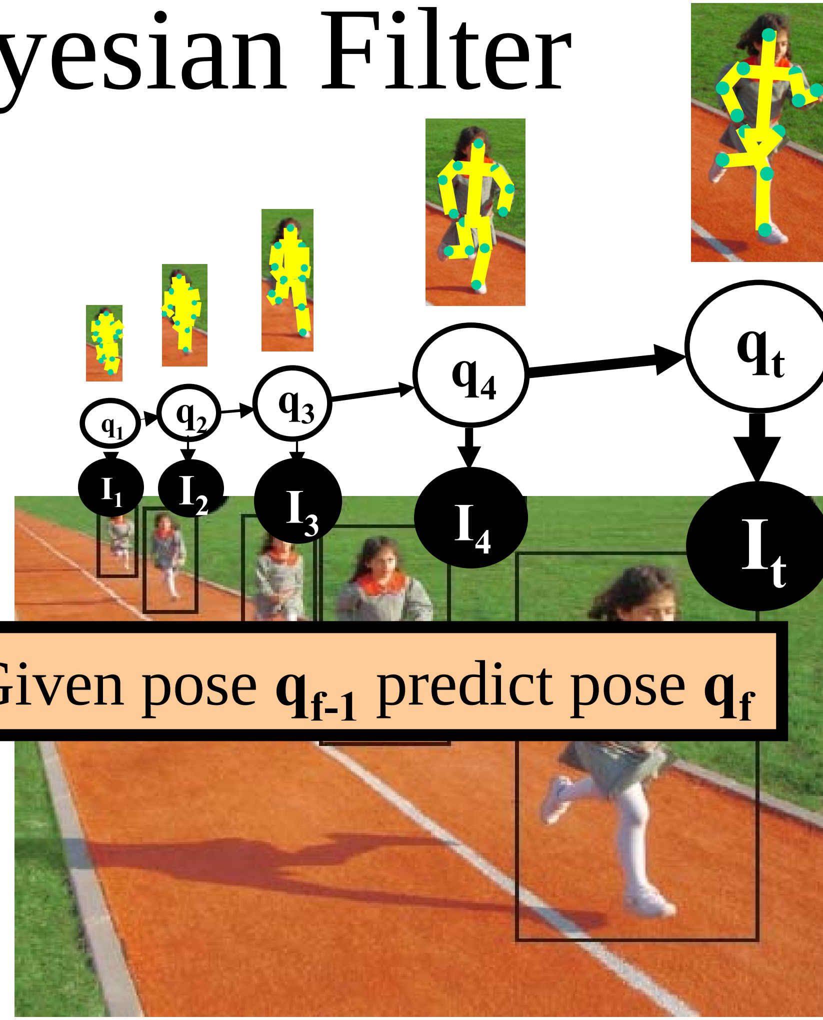
$$p(q_f \mid I_f) = p(I_f \mid q_f) \int_{q_{f-1}} p(q_f \mid q_{f-1})\, p(q_{f-1} \mid I_{f-1})$$

Kinematic Bayesian Filter

[A. Blake, M. Isard, ’98]

[Deutscher et. at., ’00]

- Motion model
  - $p(q_f \mid q_{f-1})$
- Likelihood model
  - $p(I_f \mid q_f)$



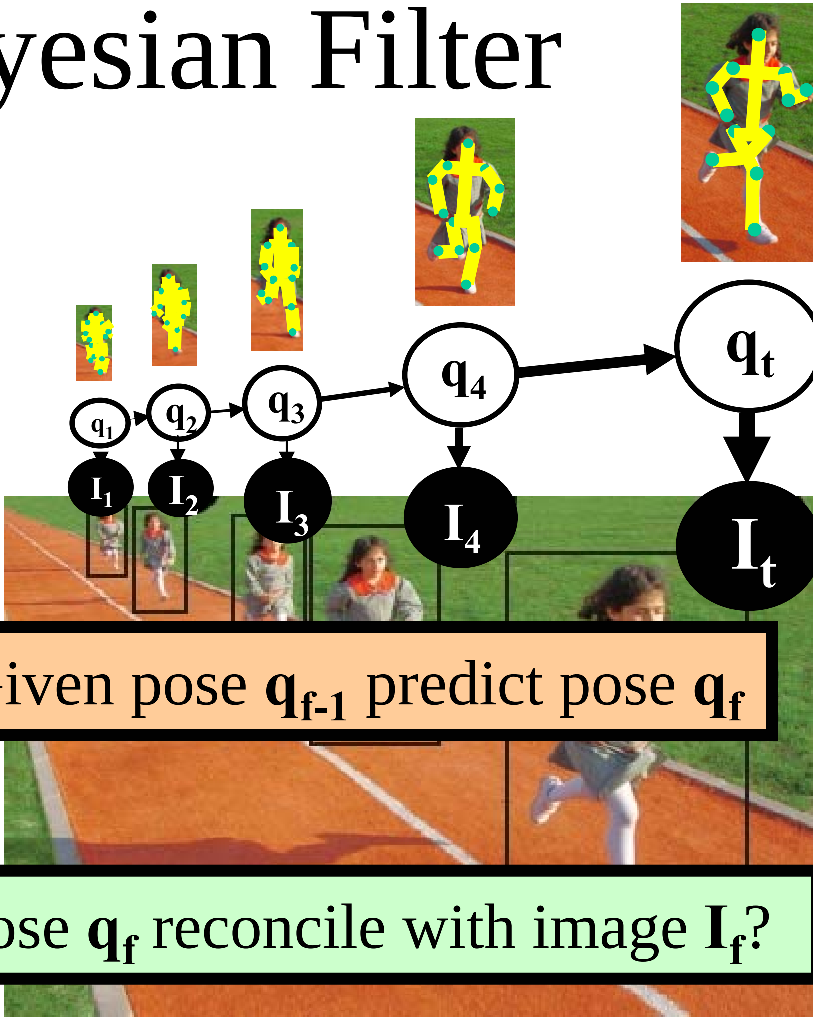
$$p(q_f \mid I_f) = p(I_f \mid q_f) \int_{q_{f-1}} p(q_f \mid q_{f-1})\, p(q_{f-1} \mid I_{f-1})$$

Kinematic Bayesian Filter

[A. Blake, M. Isard, ’98]

[Deutscher et. at., ’00]

- Motion model
  - $\mathbf{p(q_f \mid q_{f-1})}$
- Likelihood model
  - $\mathbf{p(I_f \mid q_f)}$



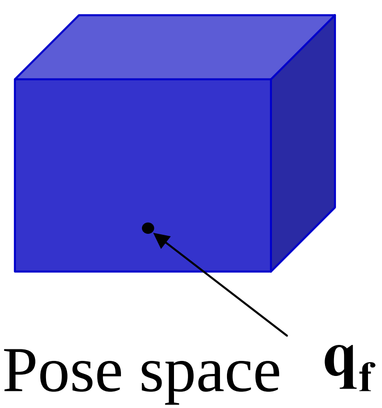
Given pose $\mathbf{q_{f-1}}$ predict pose $\mathbf{q_f}$

Does pose $\mathbf{q_f}$ reconcile with image $\mathbf{I_f}$?

$$\mathbf{p(q_f \mid I_f) = p(I_f \mid q_f) \int_{q_{f-1}} p(q_f \mid q_{f-1})\, p(q_{f-1} \mid I_{f-1})}$$

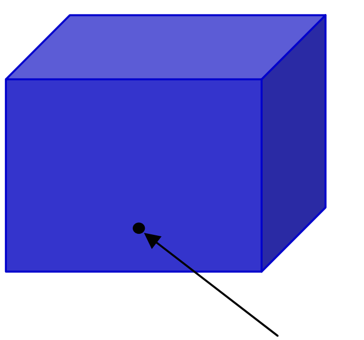
Why Is Tracking Hard?

- High dimensionality (> 30 degrees of freedom)
- Variability in imaging conditions
- Variability in appearance and clothing
- Physically realistic motion priors are hard (and often expensive) to model and characterize
  - Foot skate, out-of-plane rotations, jerky motion, etc.



Why Is Tracking Hard?

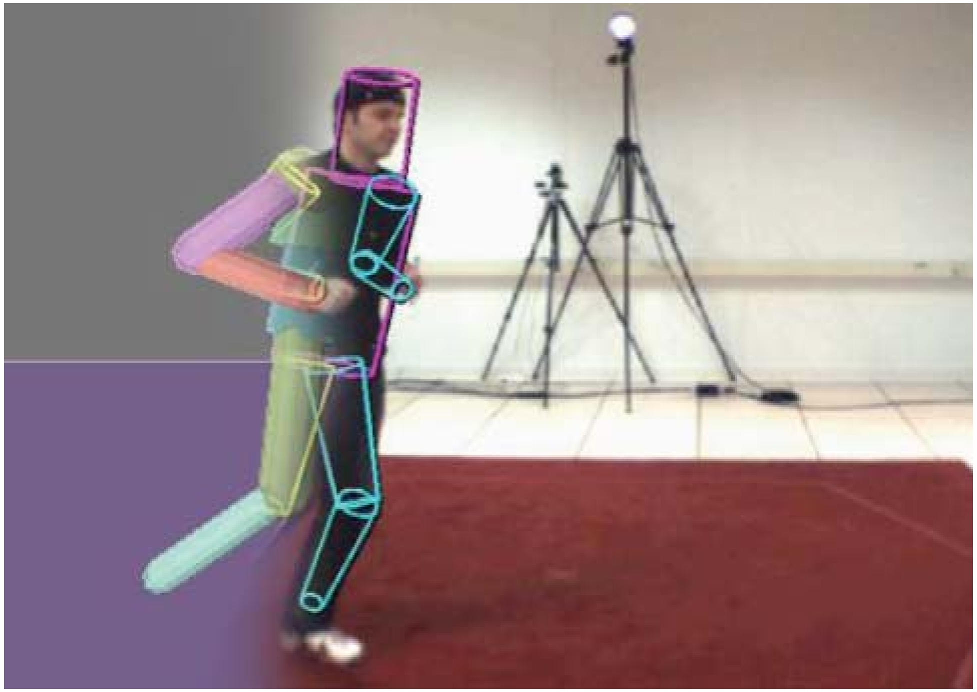
- High dimensionality (> 30 degrees of freedom)
- Variability in imaging conditions
- Variability in appearance and clothing
- Physically realistic motion priors are hard (and often expensive) to model and characterize
  - Foot skate, out-of-plane rotations, jerky motion, etc.



Pose space $\mathbf{q}_f$

Physics-based Tracking

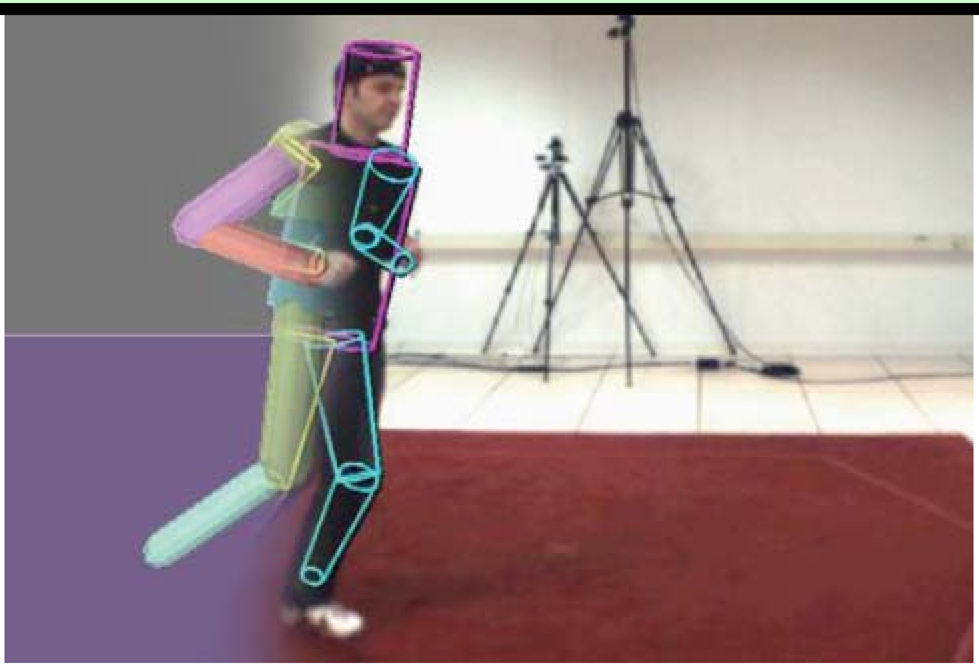
- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation



Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

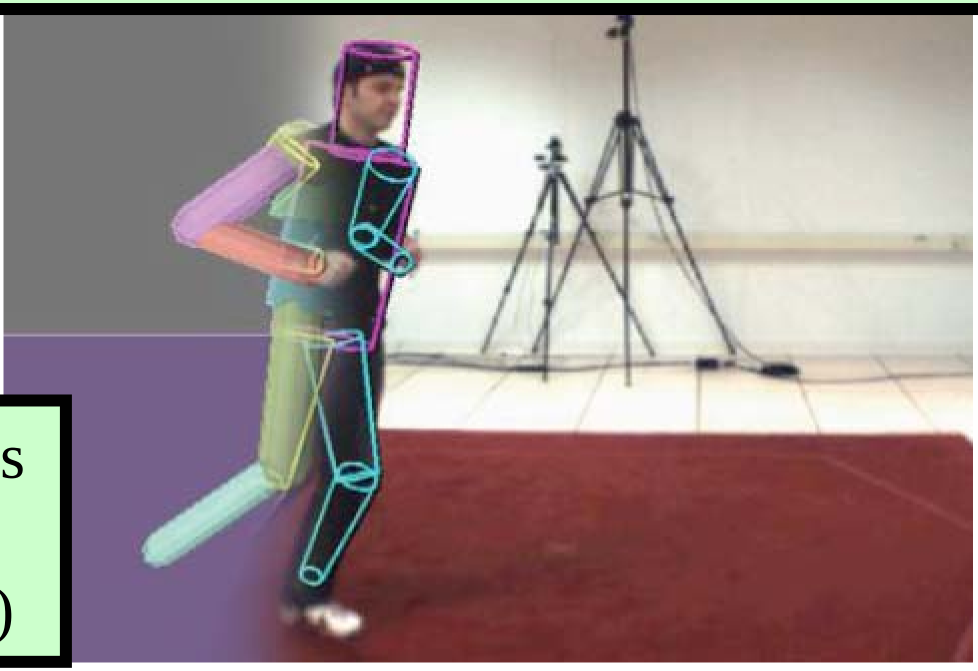
Rigid bodies under influence of forces



Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

Rigid bodies under influence of forces

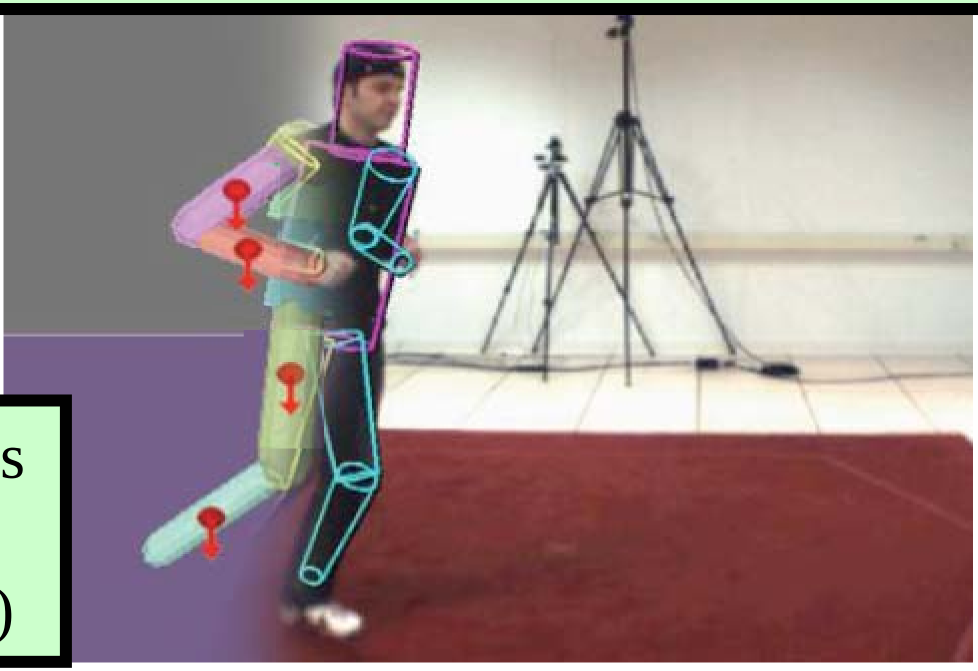


External forces (e.g. gravity, contact forces)

Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

Rigid bodies under influence of forces

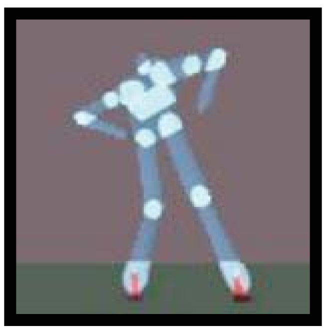


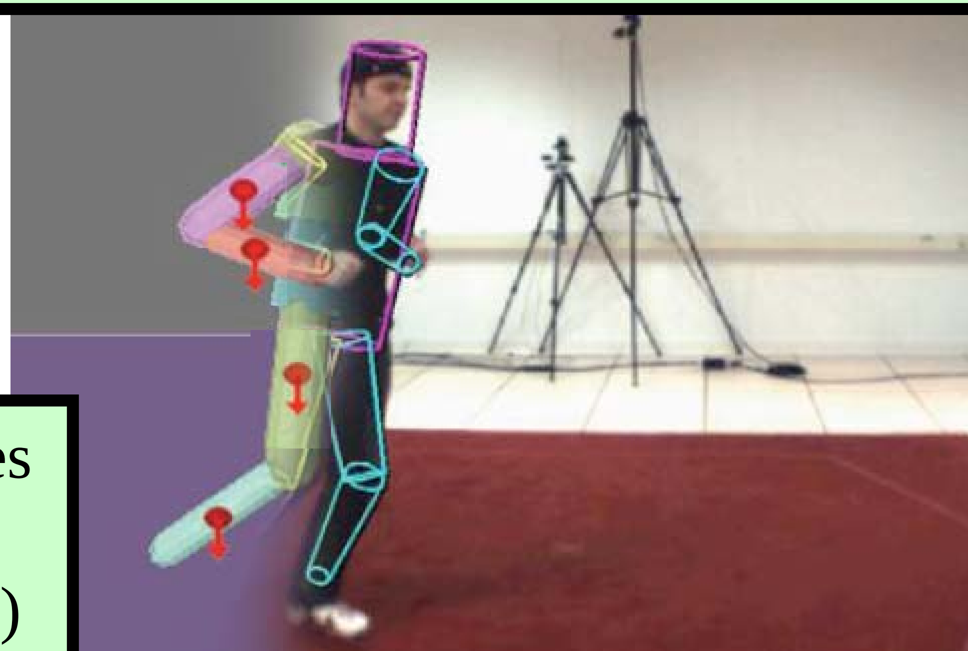
External forces (e.g. gravity, contact forces)

Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

Rigid bodies under influence of forces

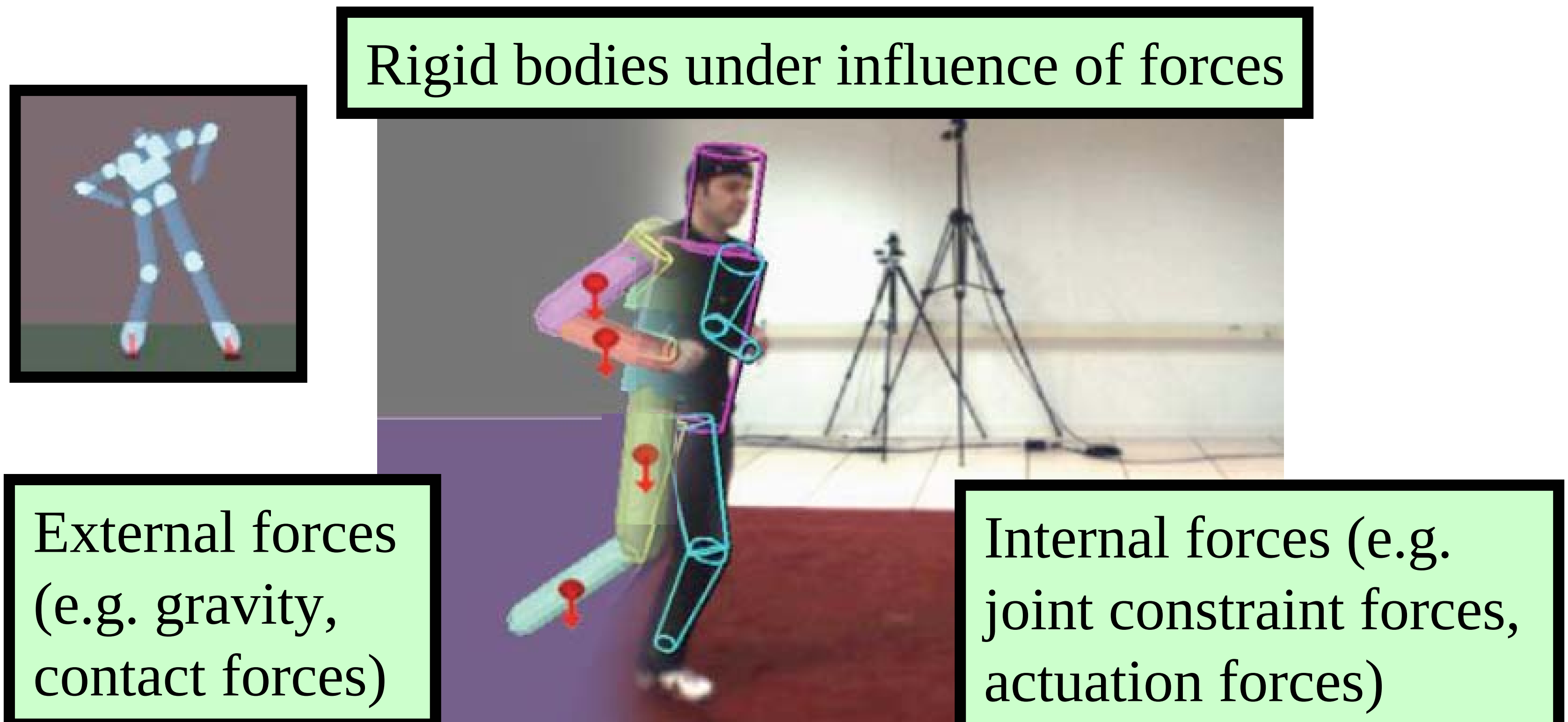




External forces (e.g. gravity, contact forces)

Physics-based Tracking

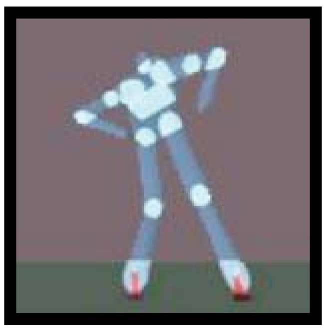
- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

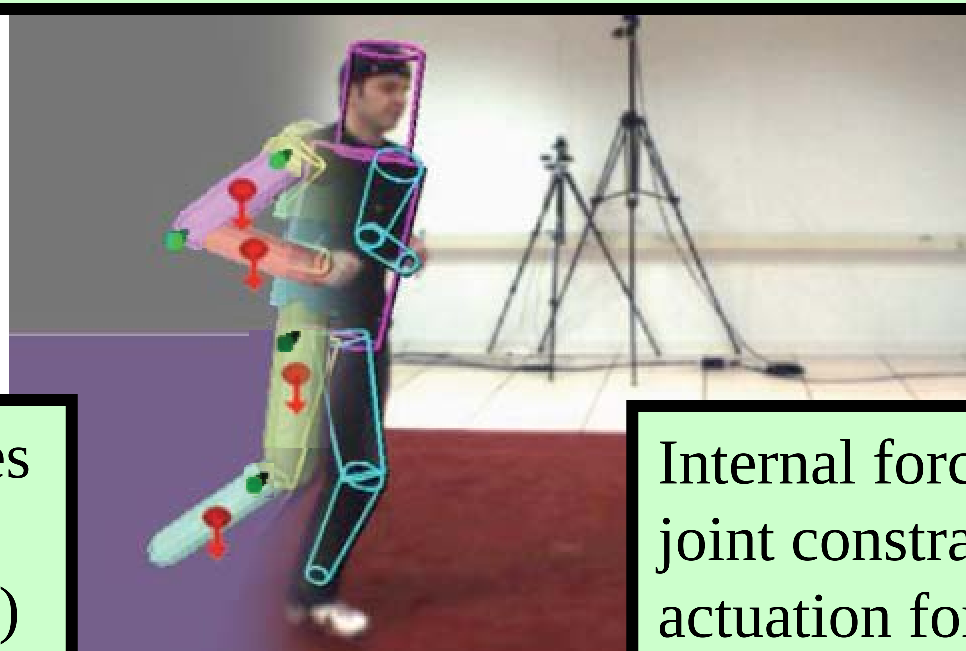


Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation

Rigid bodies under influence of forces



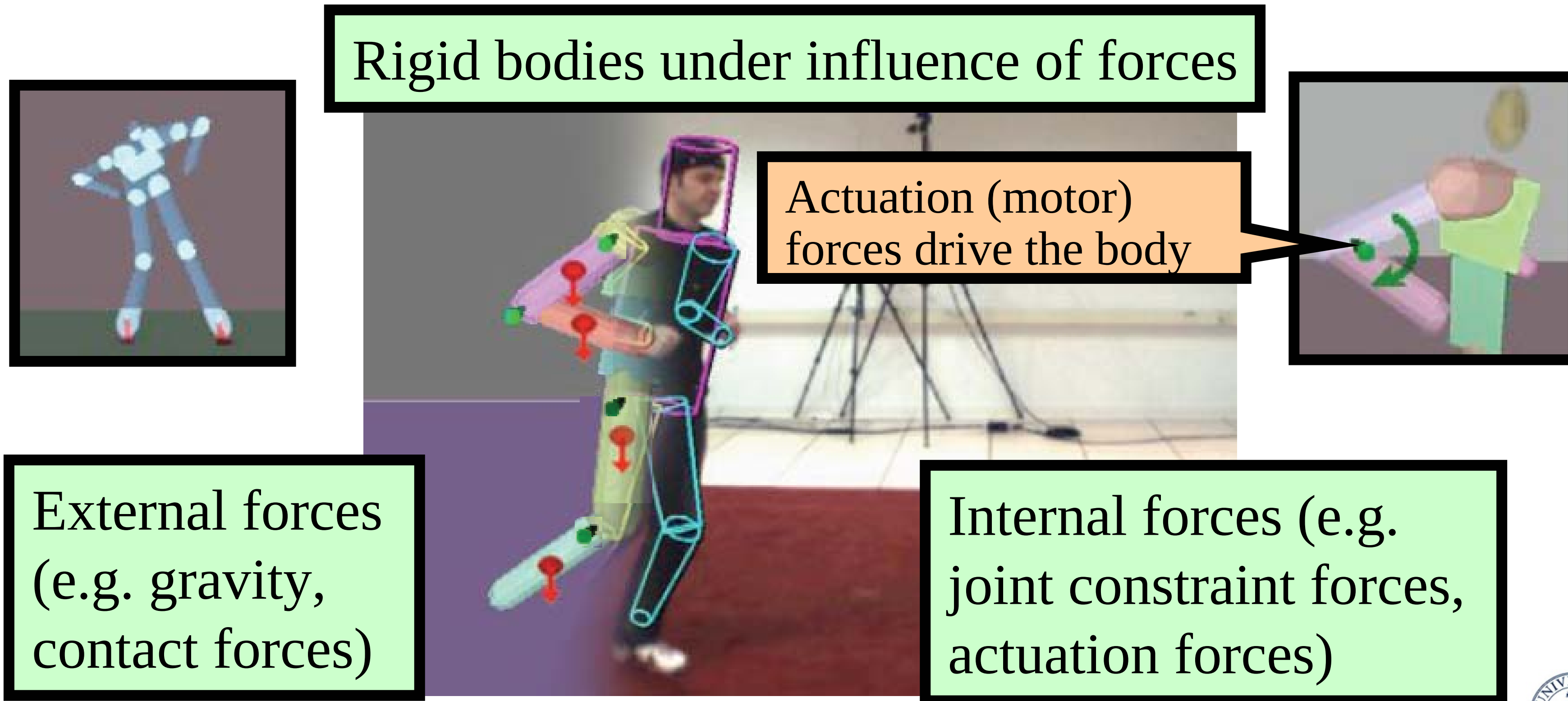


External forces (e.g. gravity, contact forces)

Internal forces (e.g. joint constraint forces, actuation forces)

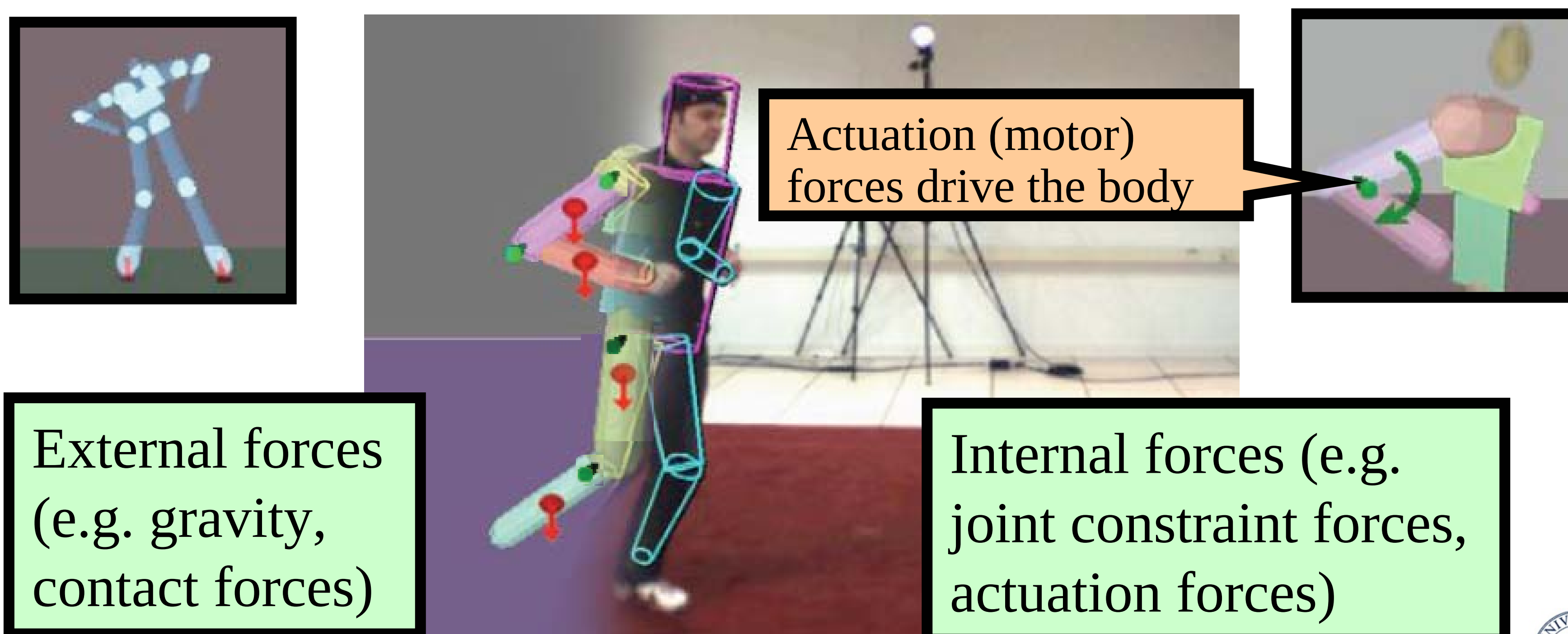
Physics-based Tracking

- Incorporate physics-based predictions into Bayesian Filtering
- Motion model based on full body 3D physical simulation



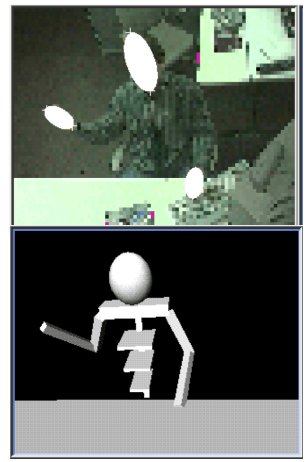
Benefits of modeling dynamics

- Ensures physical realism
- Some constraints are easier to specify (e.g. force limits, balance)
- Should generalize to new environments and dynamic interactions within the environment (e.g. motion adaptation [A. Witkin, M. Kass, ’88])



Related Work

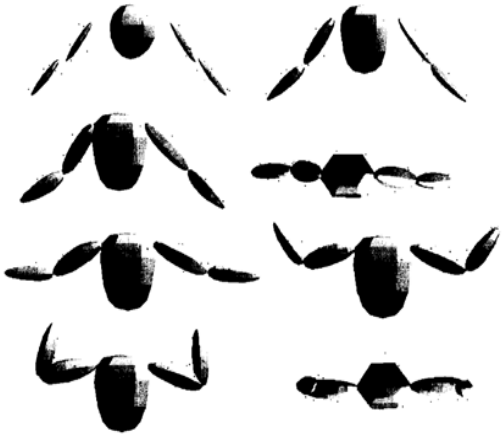
D. Metaxas, D. Terzopoulos, PAMI 1993



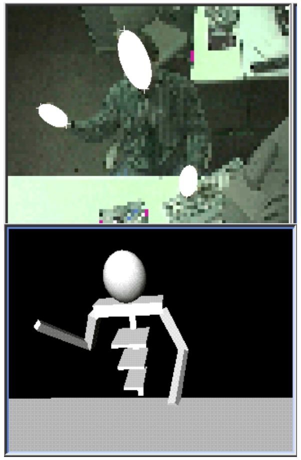
C. Wren, A. Pentland, FG 1998

- Models of dynamics
  - upper body (limited physical interactions)
- Models of observations
  - 3D marker, stereo
- Unimodal model of posterior
  - Kalman filtering

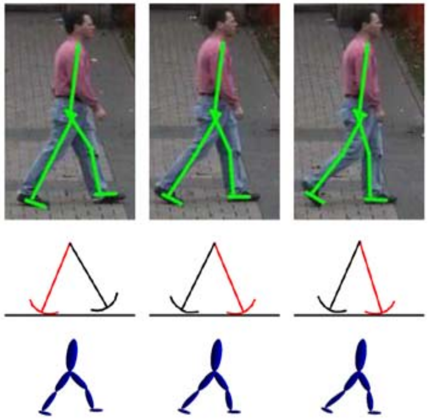
Related Work



D. Metaxas, D. Terzopoulos, PAMI 1993



C. Wren, A. Pentland, FG 1998



M. Brubaker, A. Hertzmann, D. Fleet, CVPR 2007

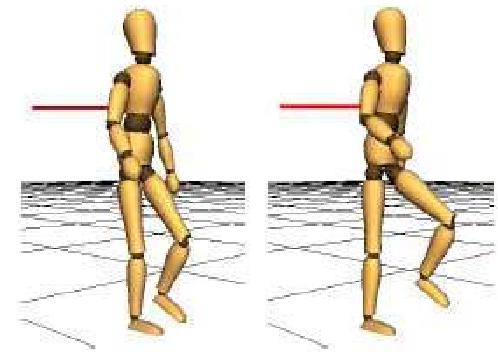
M. Brubaker, D. Fleet, CVPR 2008

- Models of dynamics
  - lower body (biomechanically inspired)
- Models of observations
  - monocular
- Multi-modal model of posterior
  - particle filtering

Related Work (Graphics/Robotics)

V. Zoran, A. Majkowska, B. Chiu, M. Fast, SIGGRAPH '05

P. Wrotek, C. Jenkins, M. McGuire, SIGGRAPH '06



K. Yin, K. Loken, M. van de Panne, SIGGRAPH '07

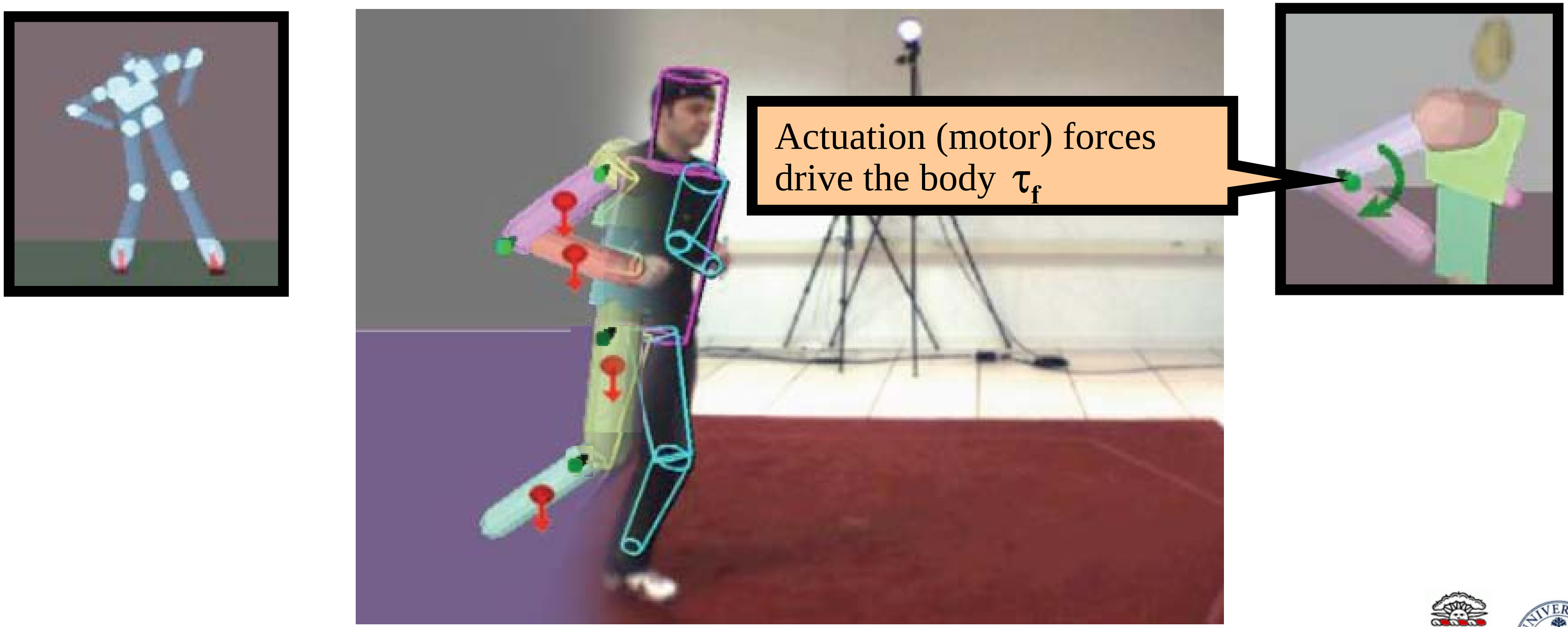
- Generic physics engines allow for complex models of the dynamics
  - Static/dynamic friction
  - Joint limit constraints
  - Active/static balance
  - Muscle models

J. Hodgins, W. Wooten, D. Brogan, J. O'Brien, SIGGRAPH '95

Physics-based Particle Filtering

- We need to let the state $\mathbf{x_f} = [\ldots \quad \tau_\mathbf{f} \quad \ldots]$

$$\mathbf{p}(\mathbf{x_f} \mid \mathbf{I_f}) = \mathbf{p}(\mathbf{I_f} \mid \mathbf{x_f}) \int_{\mathbf{x_{f-1}}} \mathbf{p}(\mathbf{x_f} \mid \mathbf{x_{f-1}})\,\mathbf{p}(\mathbf{x_{f-1}} \mid \mathbf{I_{f-1}})$$

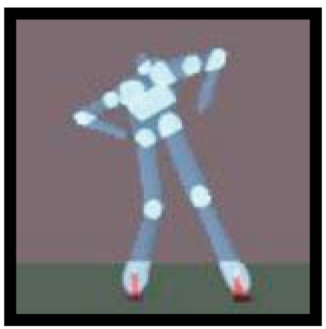


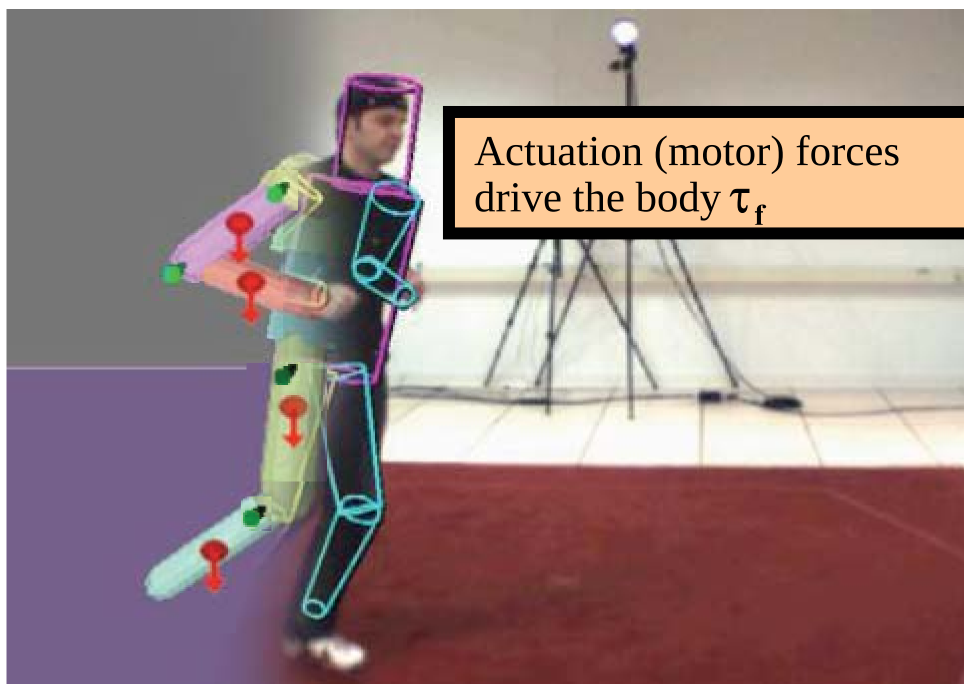
Physics-based Particle Filtering

- **Problem:**
  - Priors over valid motor force (torque) trajectories are hard to characterize
- **Solution(s):**
  - Use simplified physical models for which priors are easy to model

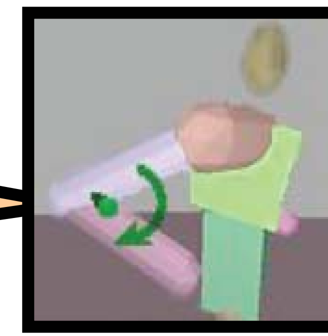
Brubaker et. al., '07

  - **Assume that we have the model for kinematics and solve for the forces implicitly (i.e. use a controller)**



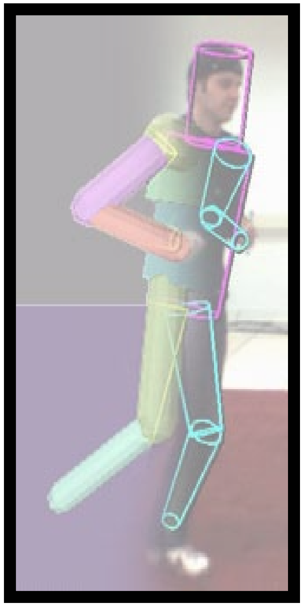


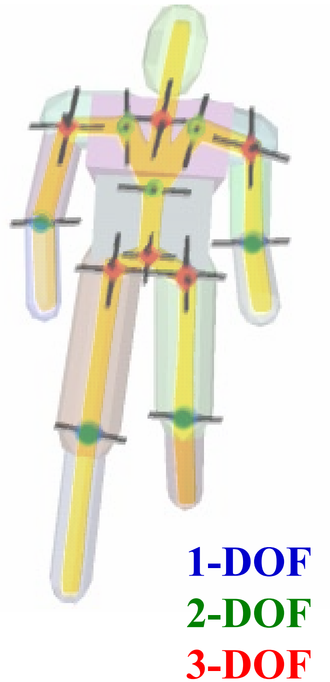
Actuation (motor) forces drive the body τ_f



Body Model and State Space

- World model
  - Known static environment
  - Loop-free articulated figure
    - 31 degrees of freedom (DOFs)
    - 13 rigid bodies
    - Known physical properties (geometry, mass, inertial)





Kinematic State

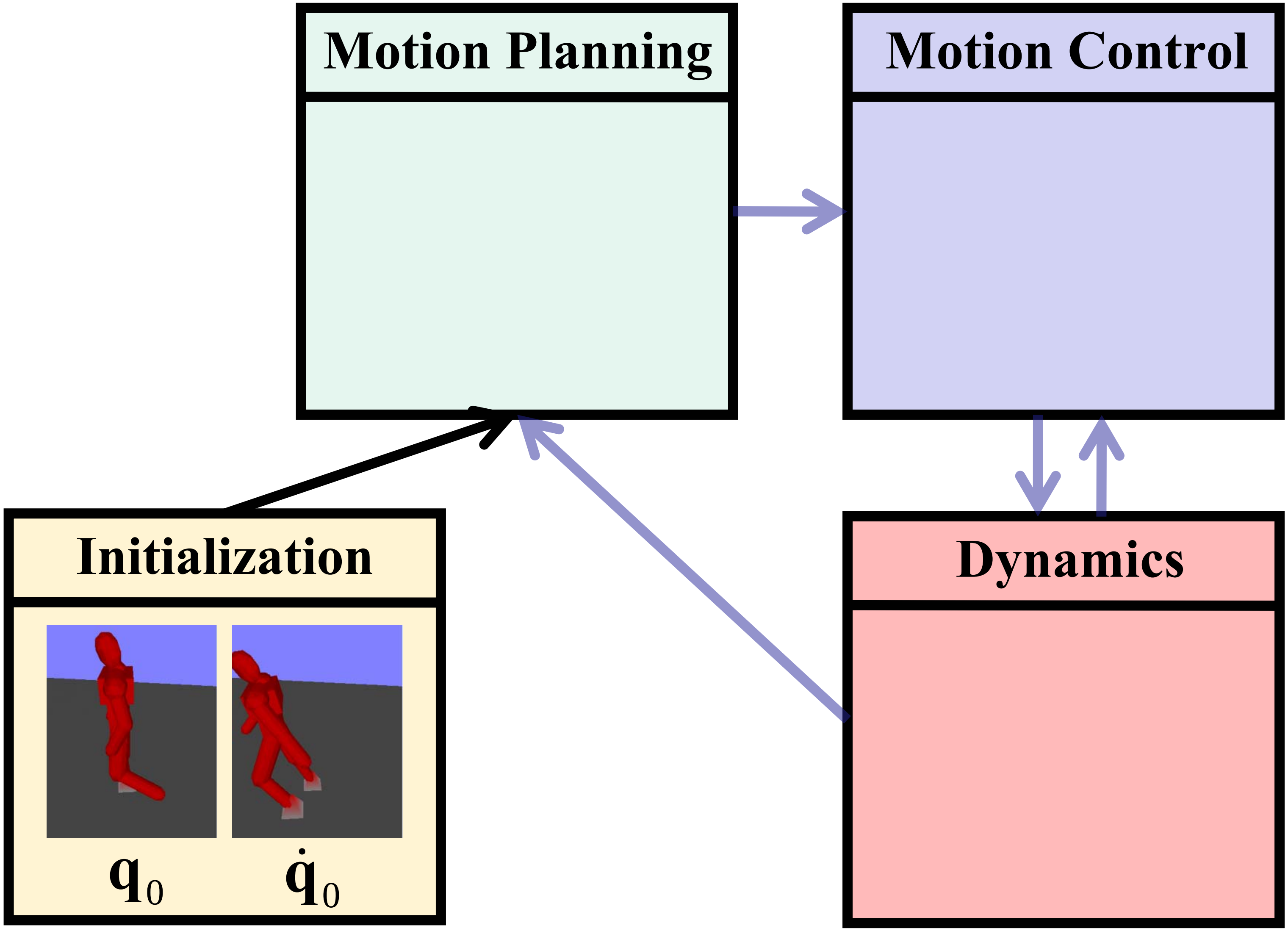
State vector: $\mathbf{x} = [\underbrace{\mathbf{q} \quad \dot{\mathbf{q}}} \quad \pi]$

Control strategy

Dynamic State

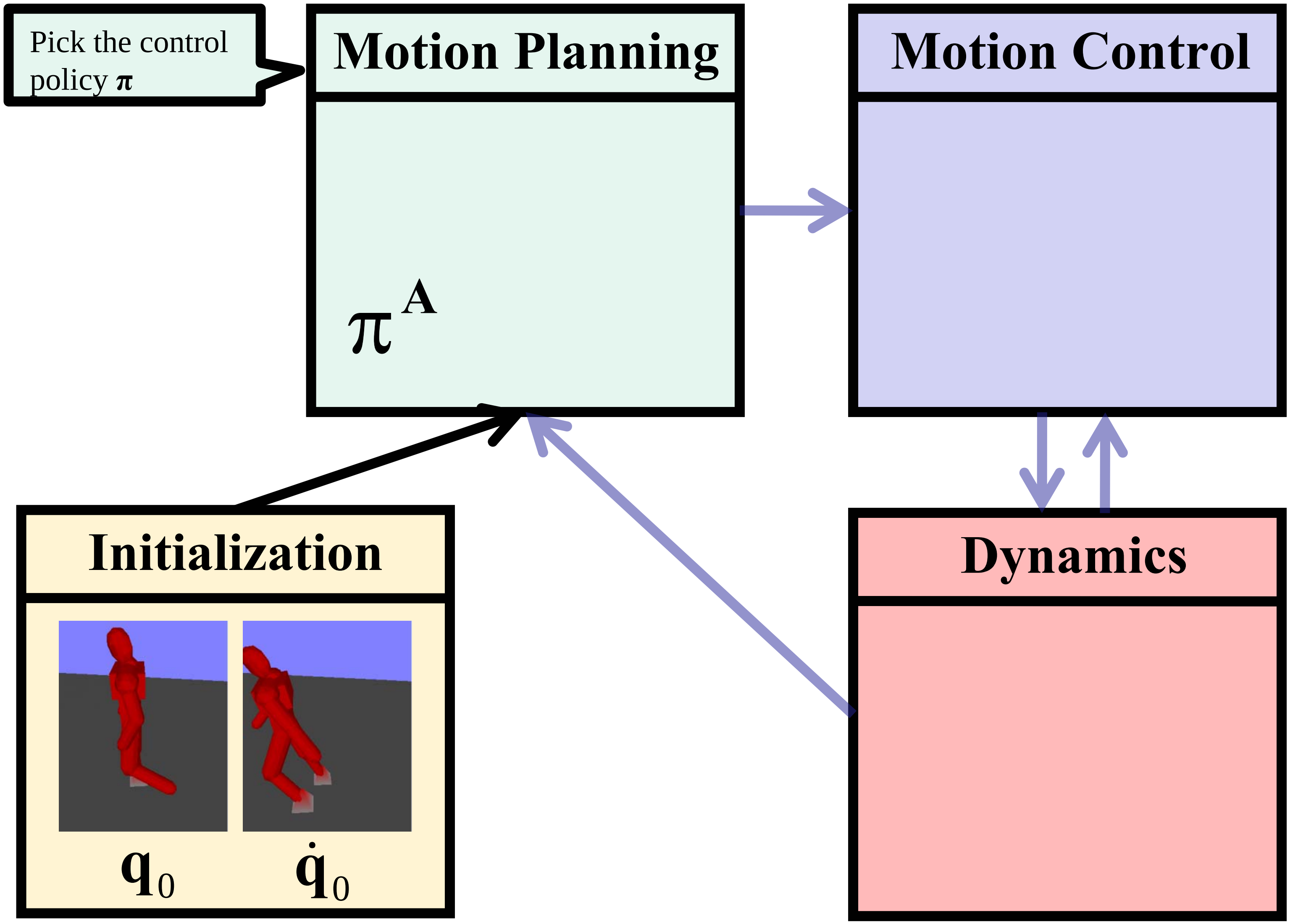
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



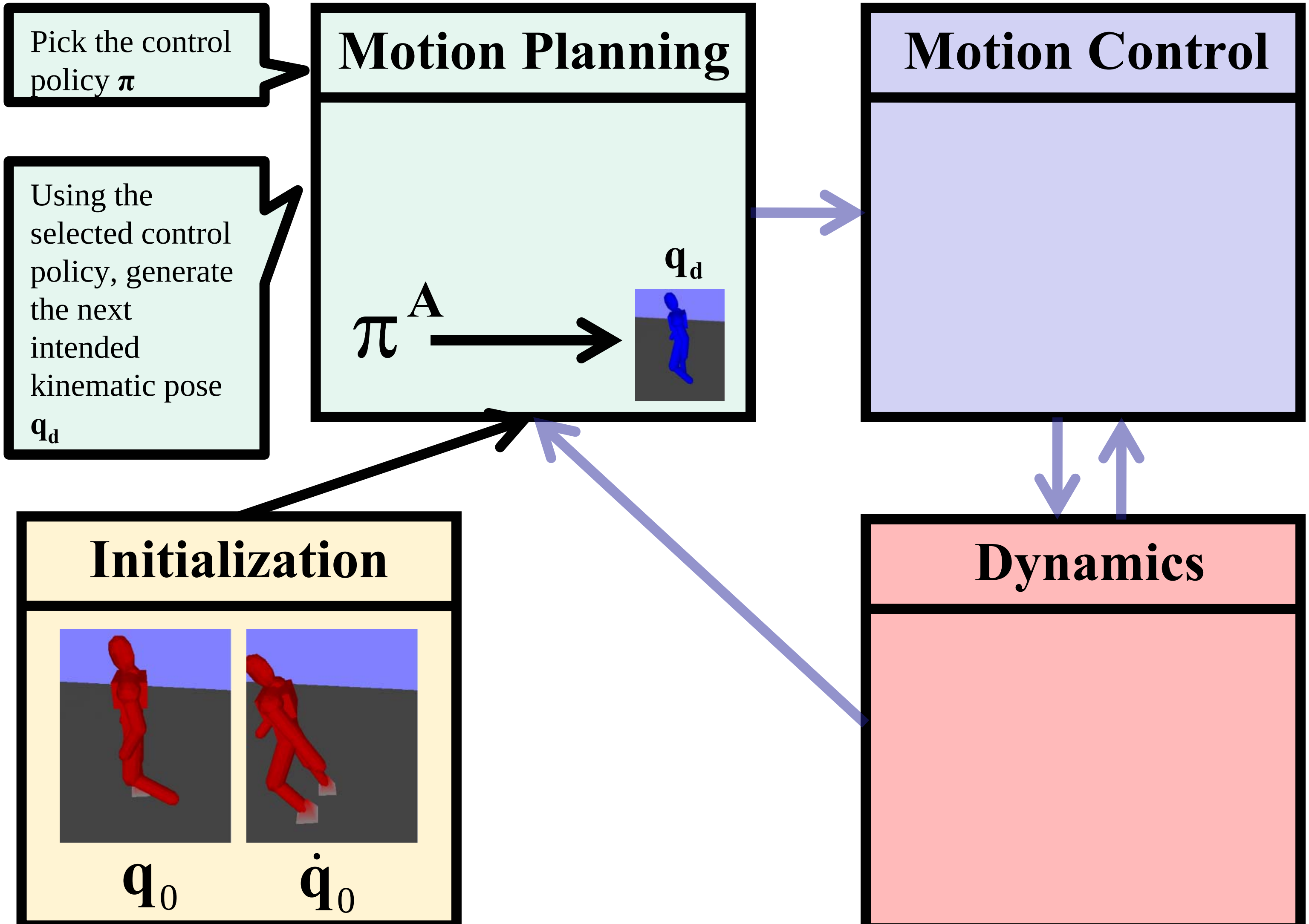
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



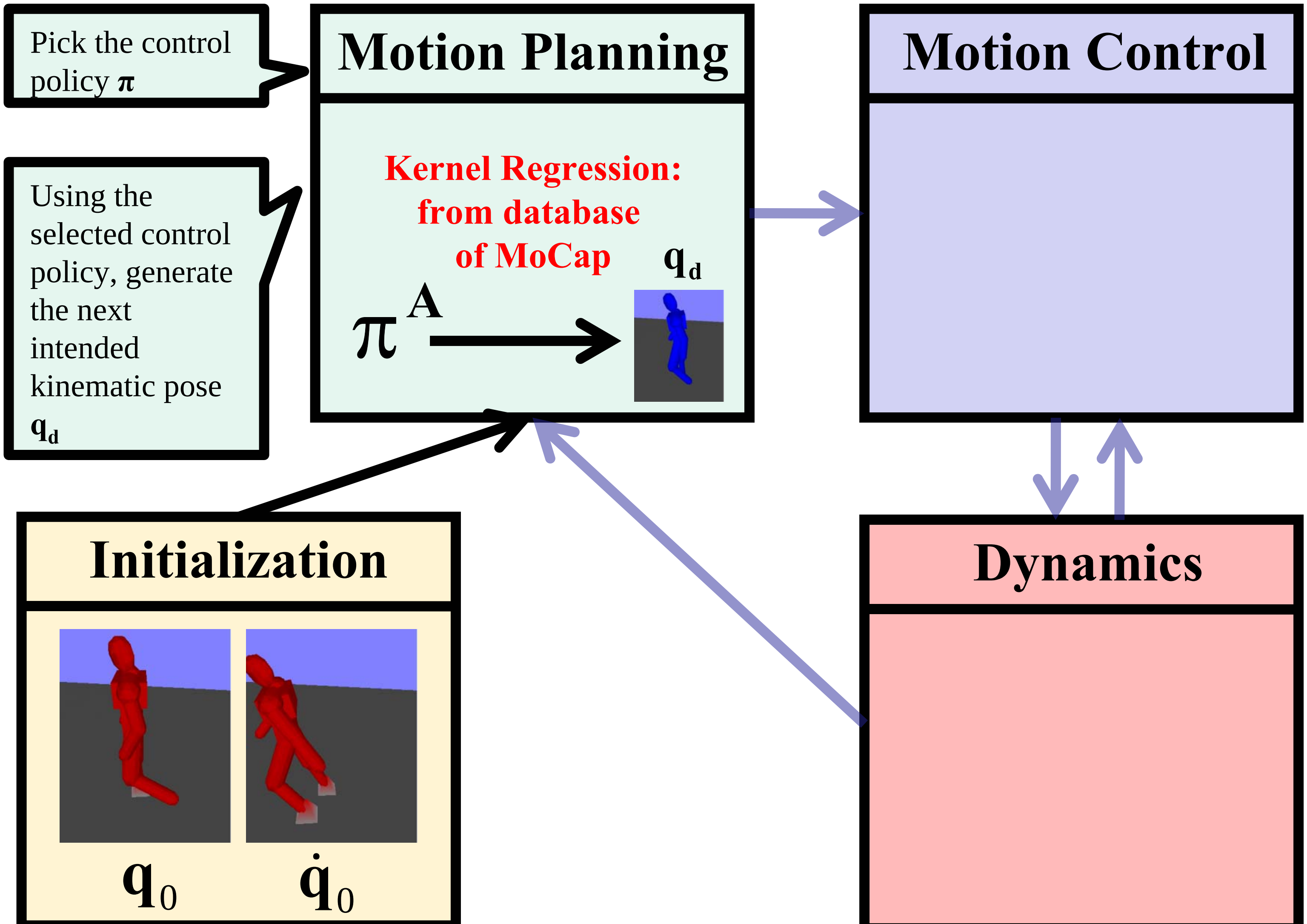
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



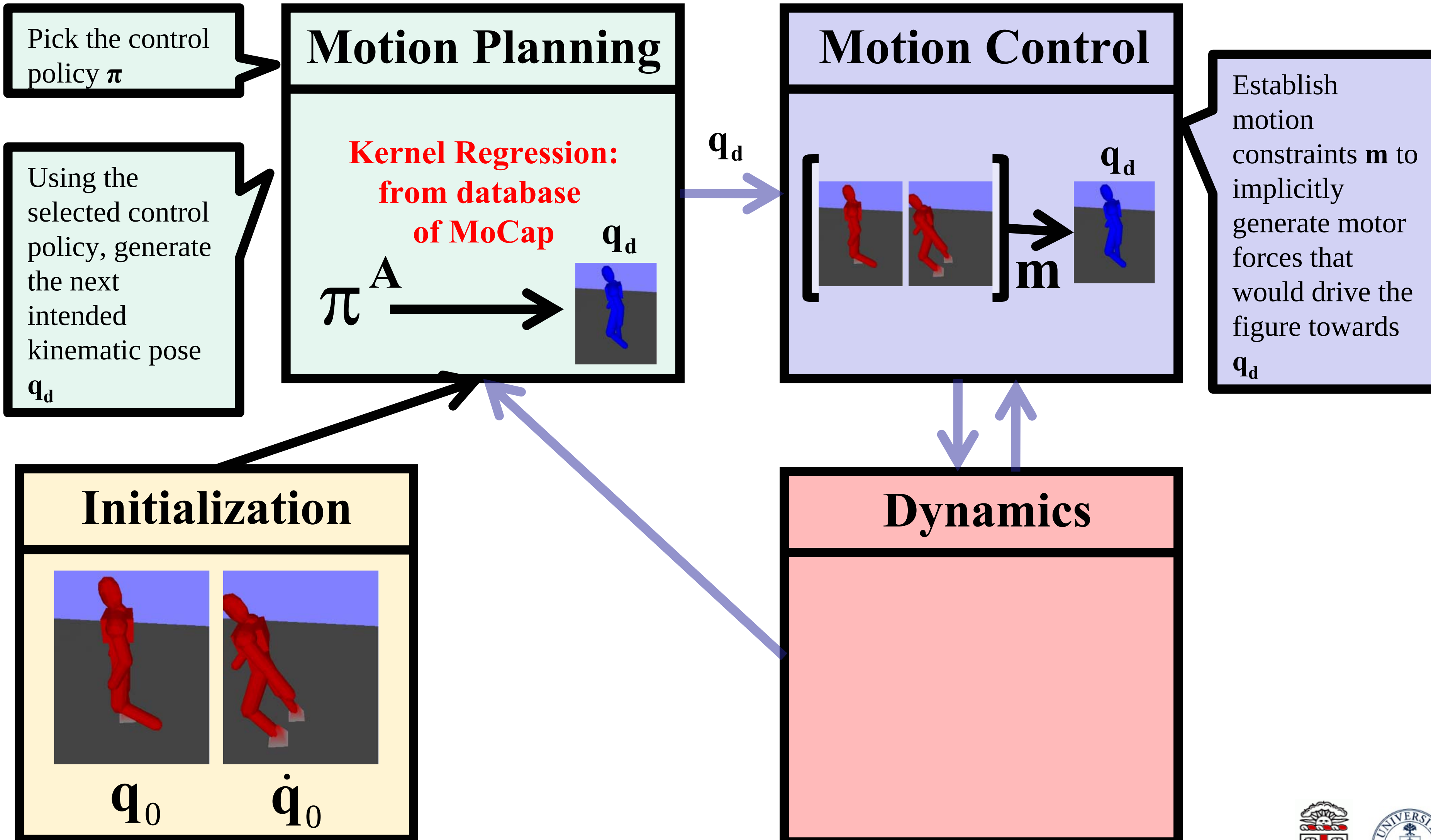
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



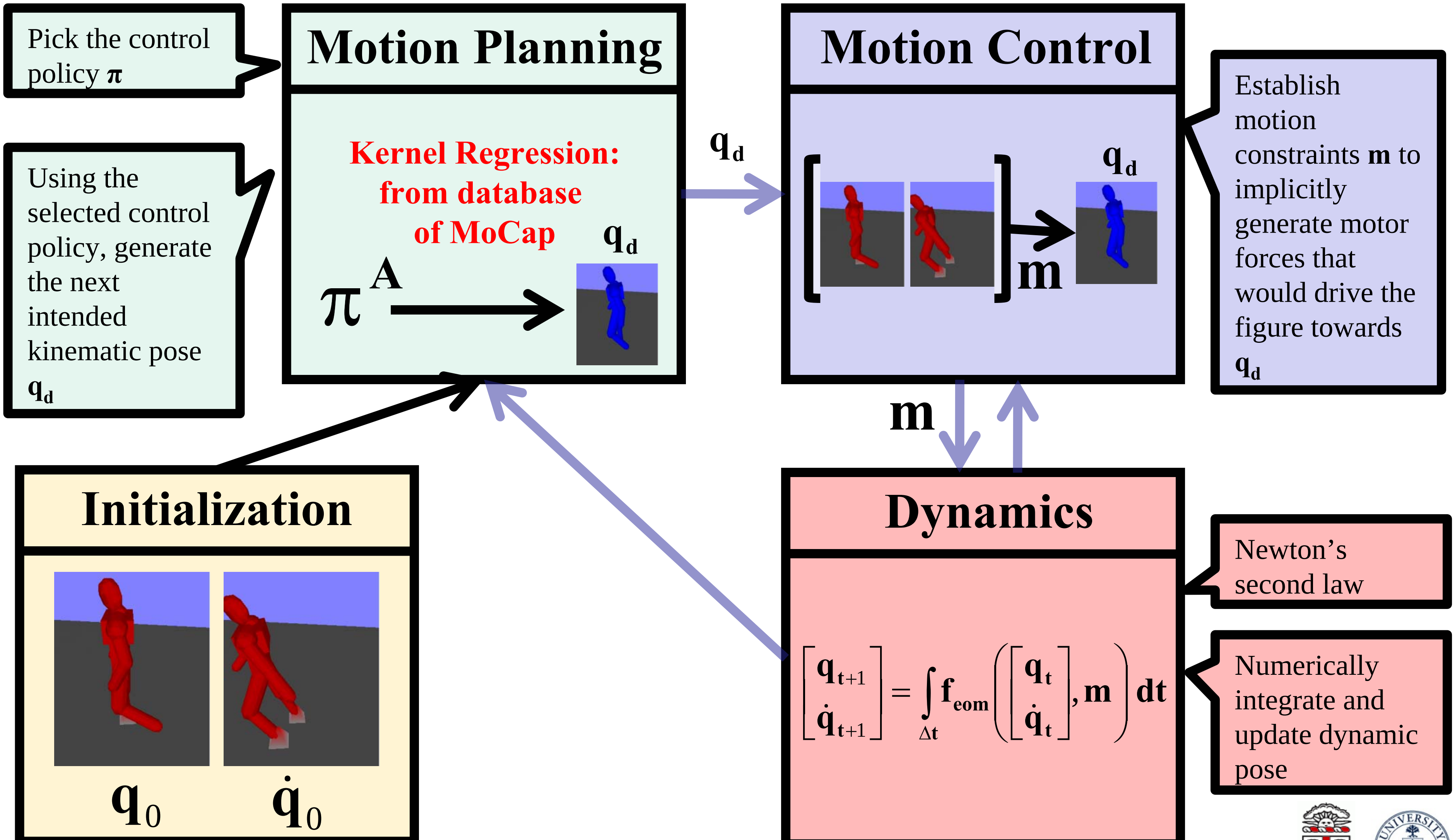
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



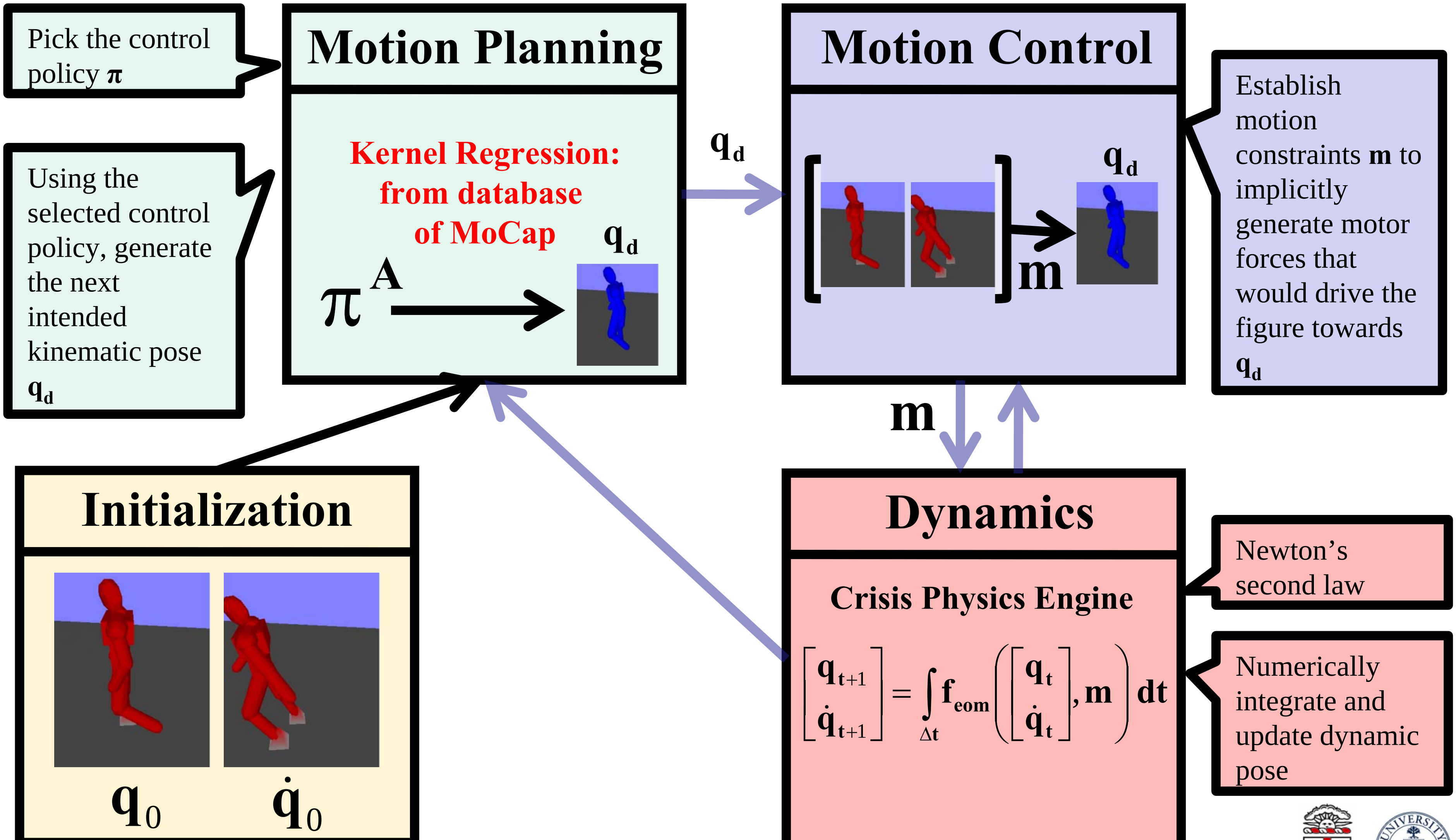
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



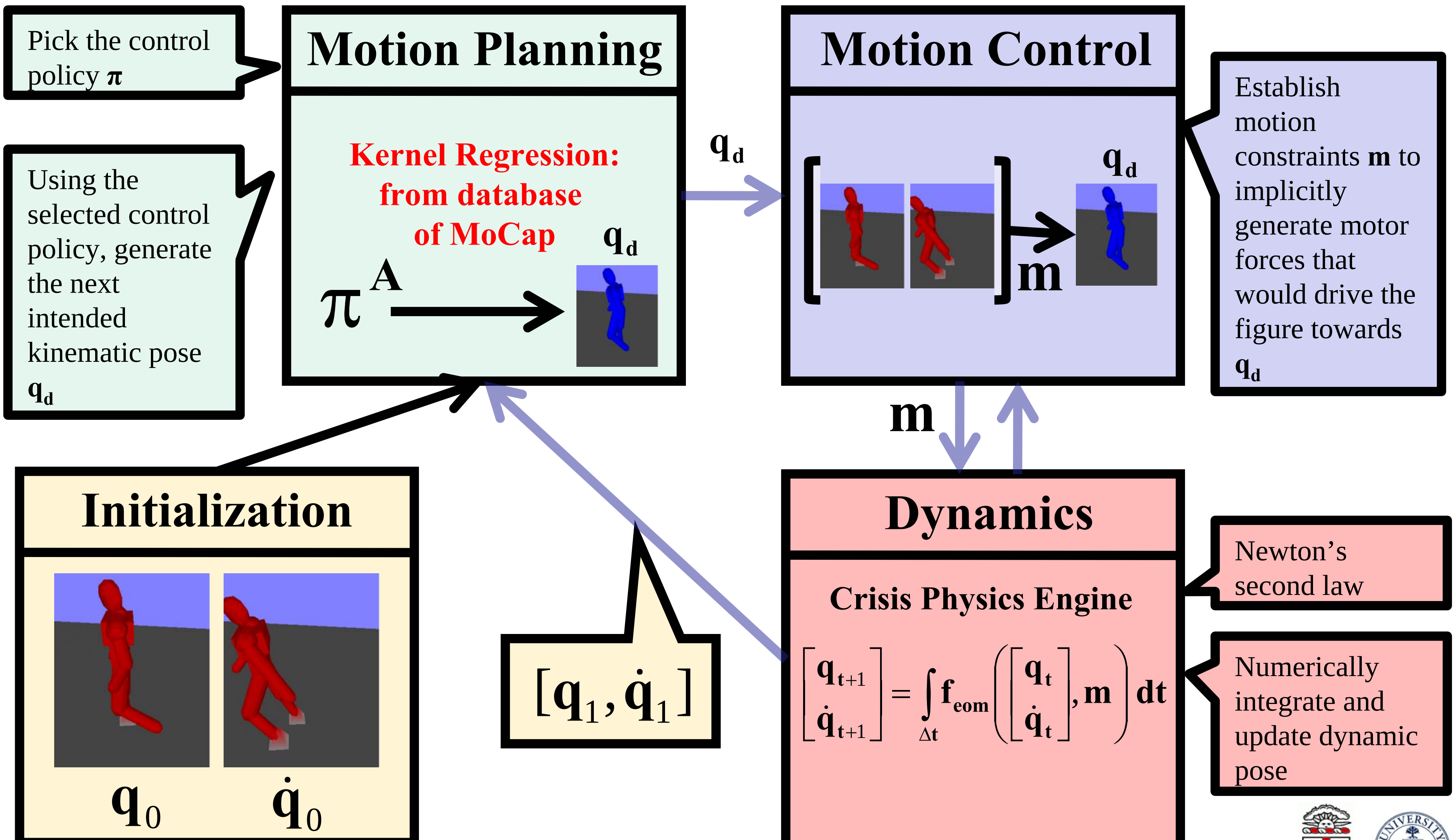
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



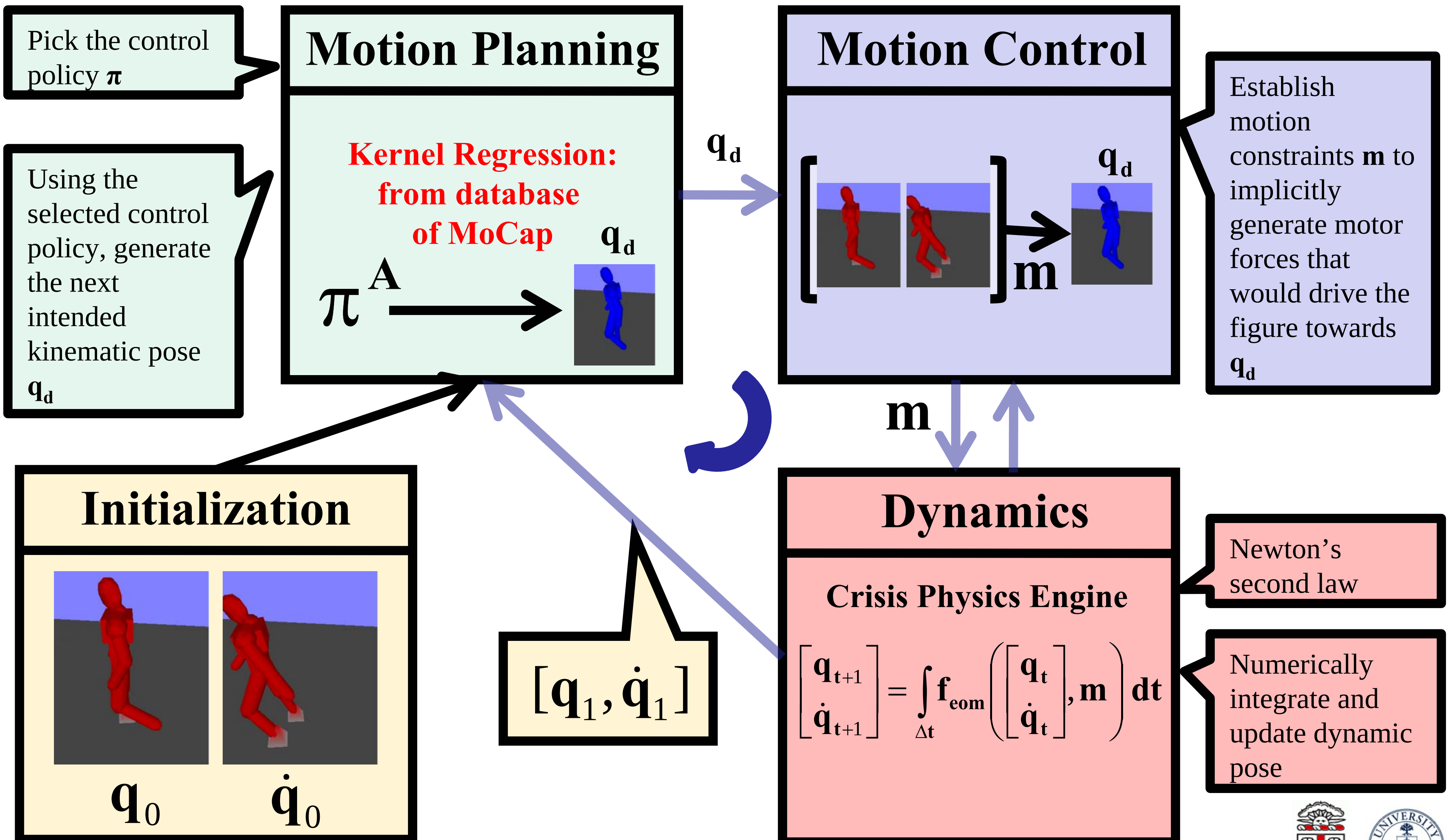
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



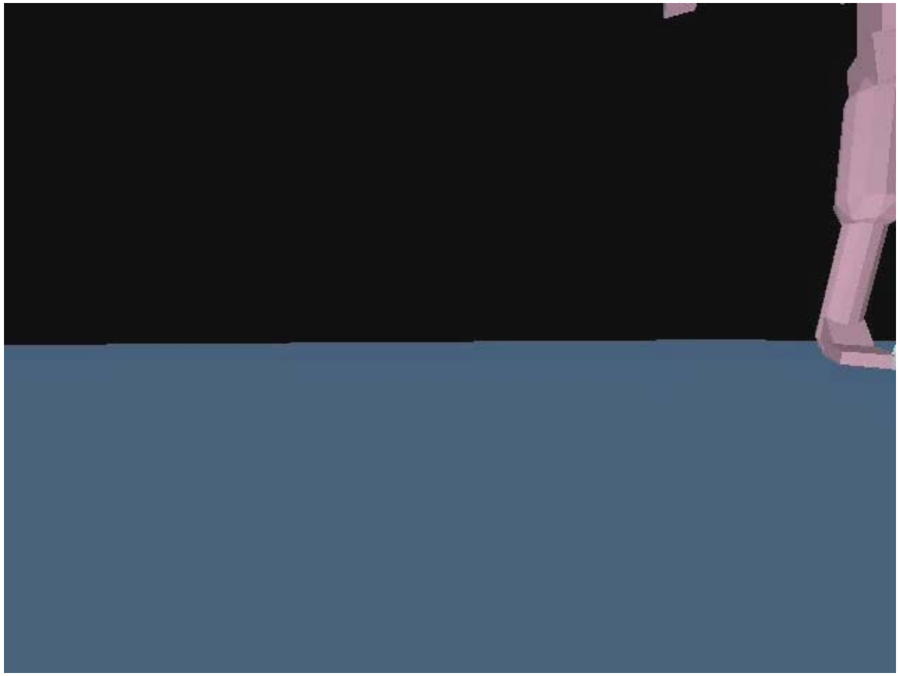
Motion Model: Control Loop

- Executed for every hypothesis in our multi-hypothesis tracking framework



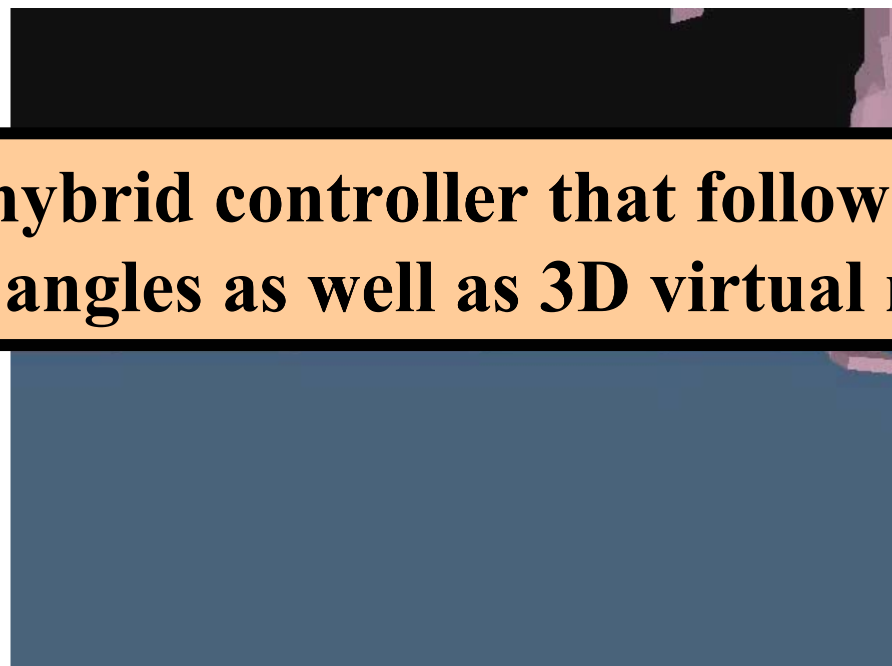
Motion Model: Motion Control

- Control that follows joint angles alone is problematic
- Locomotion results only from interactions with ground



Motion Model: Motion Control

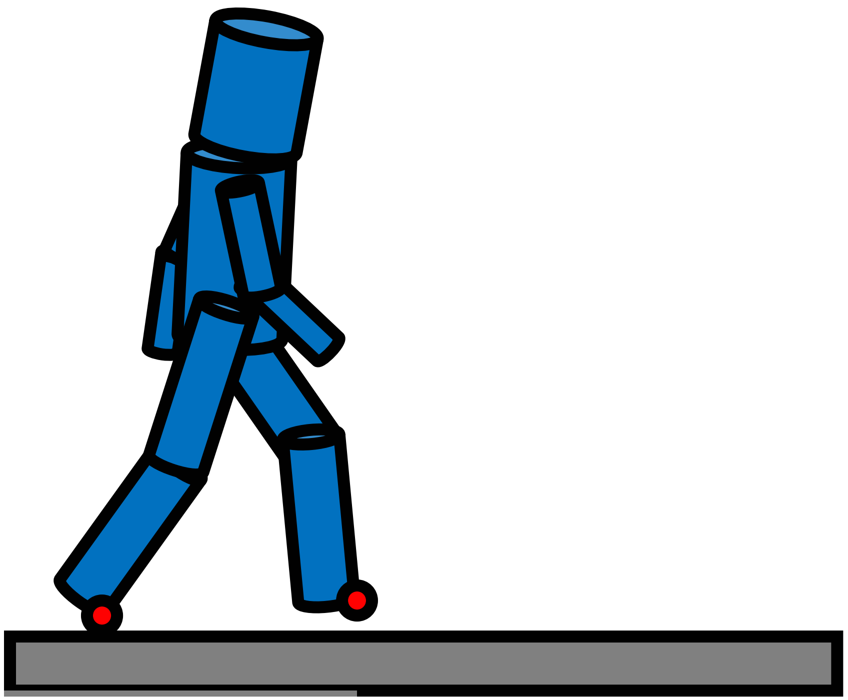
- Control that follows joint angles alone is problematic
- Locomotion results only from interactions with ground



New hybrid controller that follows desired joint angles as well as 3D virtual markers

Motion Model: Motion Control

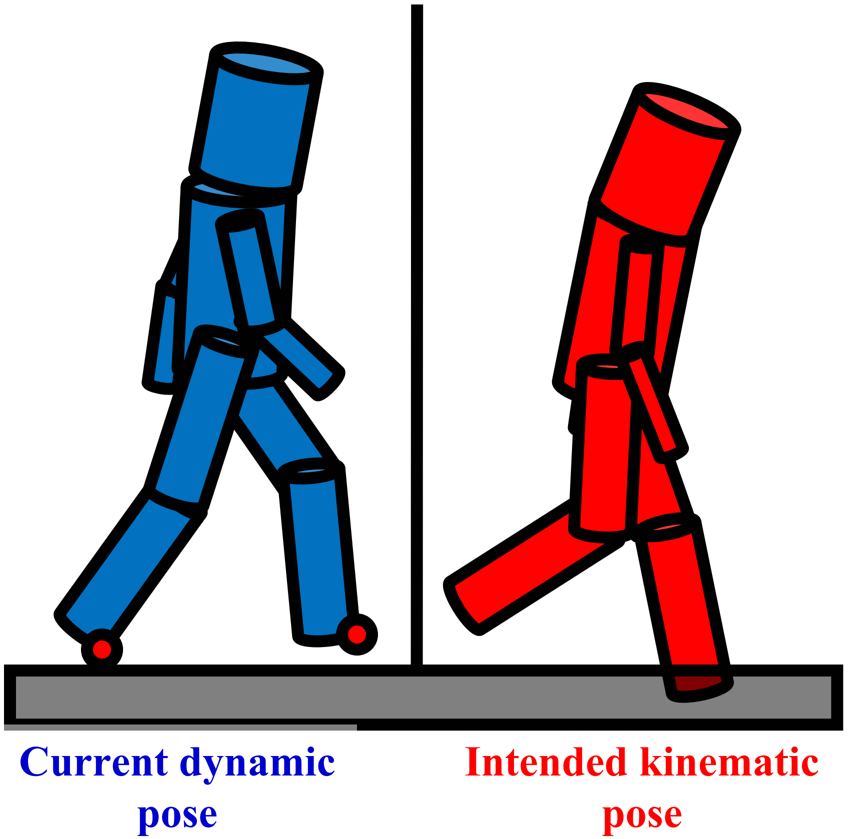
- Given the intended kinematic pose $\mathbf{q_d}$



Current dynamic pose

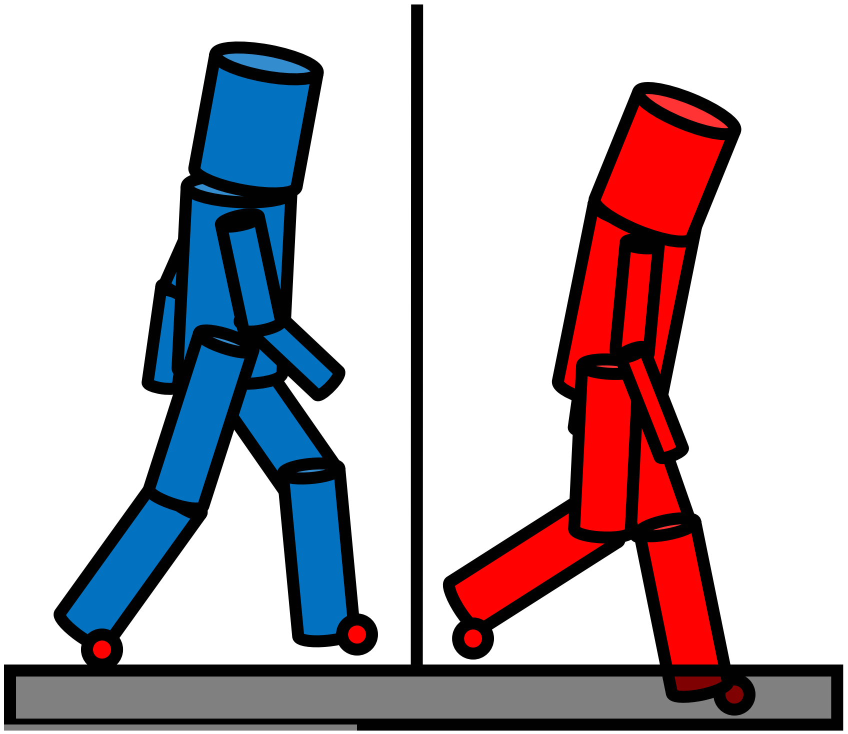
Motion Model: Motion Control

- Given the intended kinematic pose $\mathbf{q}_d$



Motion Model: Motion Control

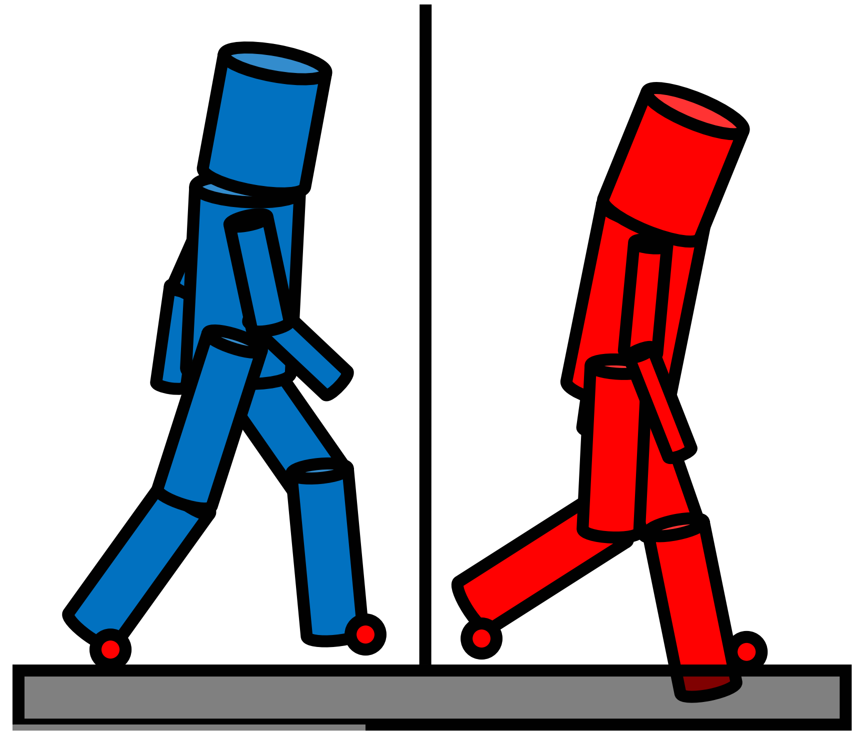
- Given the intended kinematic pose $\mathbf{q_d}$
  - the controller computes intended positions of markers



Current dynamic pose **Intended kinematic pose**

Motion Model: Motion Control

- Given the intended kinematic pose $\mathbf{q_d}$
  - the controller computes intended positions of markers
  - adjusts the positions so they do not penetrate the environment

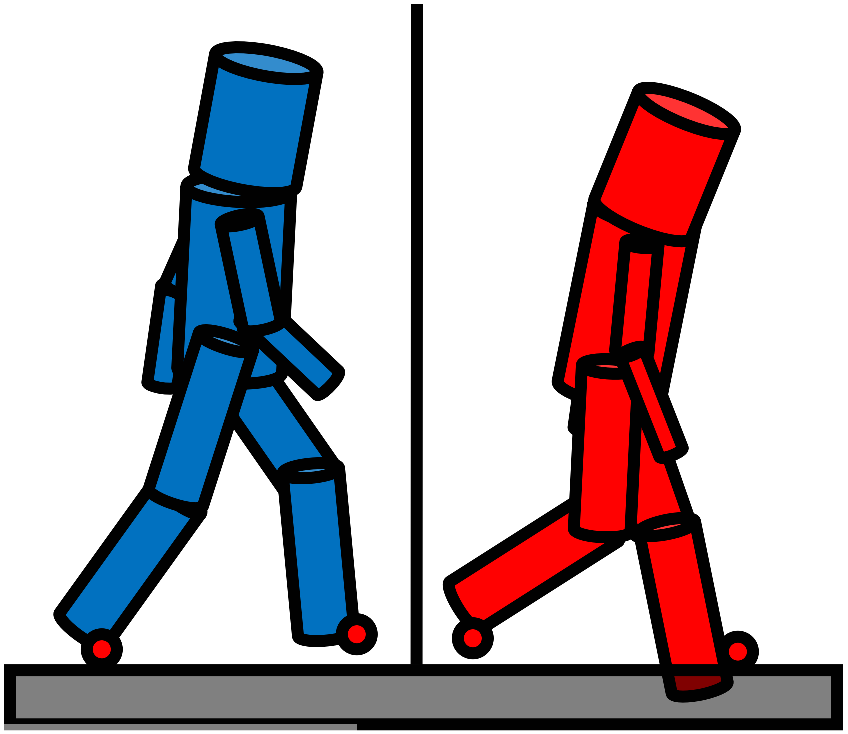


Current dynamic pose

Intended kinematic pose

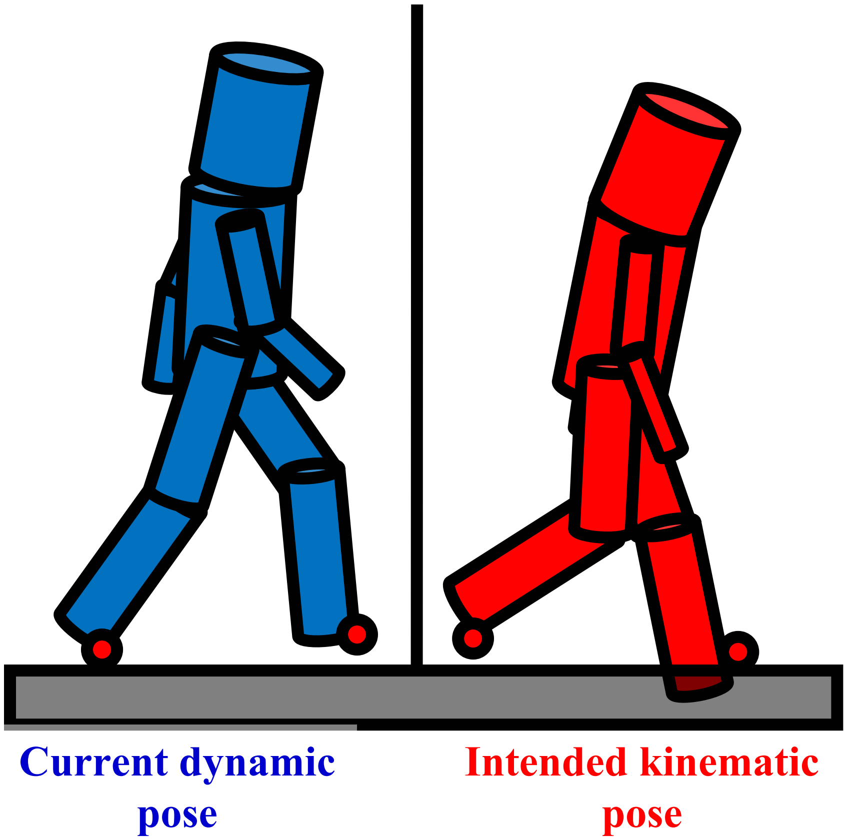
Motion Model: Motion Control

- Given the intended kinematic pose $\mathbf{q_d}$
  - the controller computes intended positions of markers
  - adjusts the positions so they do not penetrate the environment
  - solves for additional constraints that would drive the current **dynamic pose** towards the **adjusted intended kinematic pose**



Motion Model: Motion Control

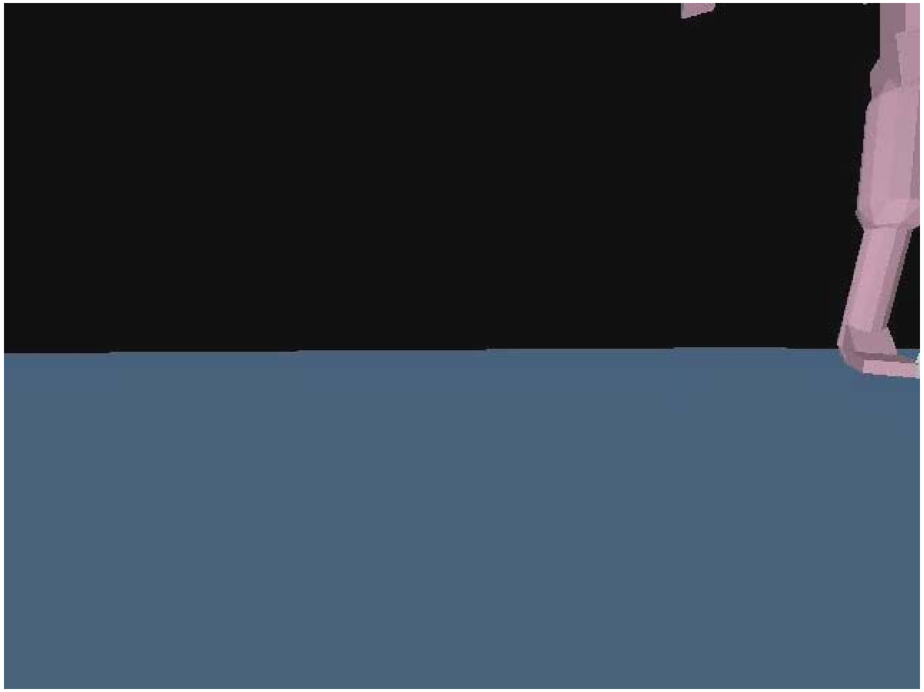
- Given the intended kinematic pose $\mathbf{q_d}$
  - the controller computes intended positions of markers
  - adjusts the positions so they do not penetrate the environment
  - solves for additional constraints that would drive the current **dynamic pose** towards the **adjusted intended kinematic pose**



Inverse Dynamics

Motion Model: Motion Control

Motion simulation using Crisis physics engine



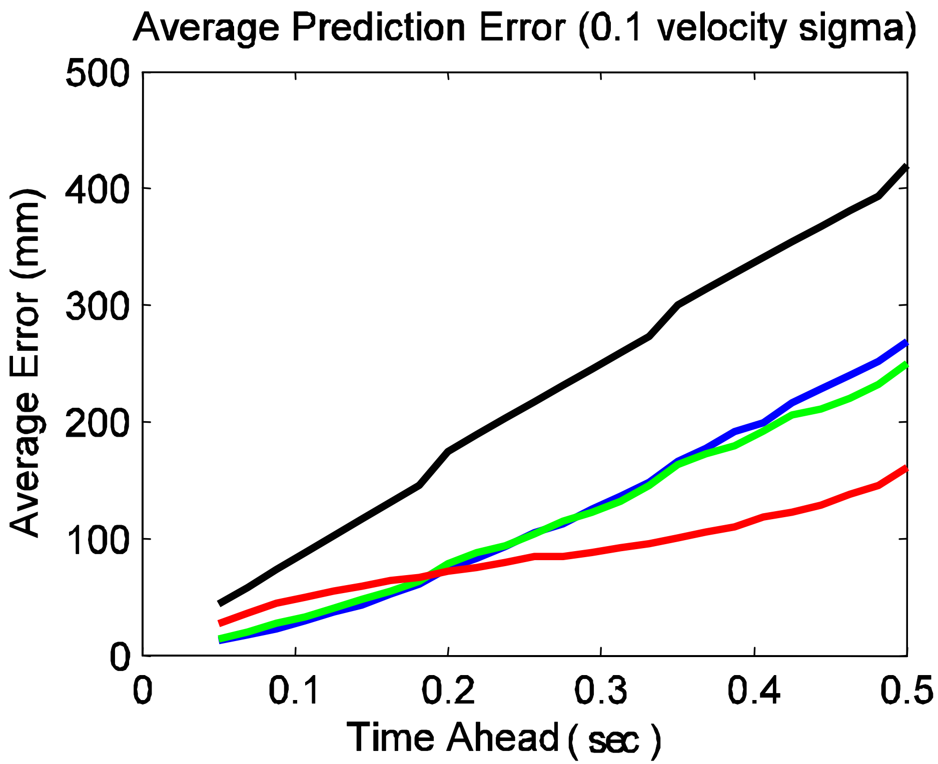
Traditional Controller

Our Controller

How Good is our Motion Model?

Given: a pose (position and velocity) of the body from walking mocap data

Goal: predict the state some time **Δt** into the future



No prior:

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{q}_{\mathbf{f}} + \mathbf{N}(0, \Sigma)$$

Constant velocity prior:

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{q}_{\mathbf{f}} + \Delta\mathbf{t} \cdot \dot{\mathbf{q}}_{\mathbf{f}} + \mathbf{N}(0, \Sigma)$$

Physics (passive rag-doll):

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{f}(\mathbf{q}_{\mathbf{f}}, \dot{\mathbf{q}}_{\mathbf{f}}, \pi^{\mathbf{P}}) + \mathbf{N}(0, \Sigma)$$

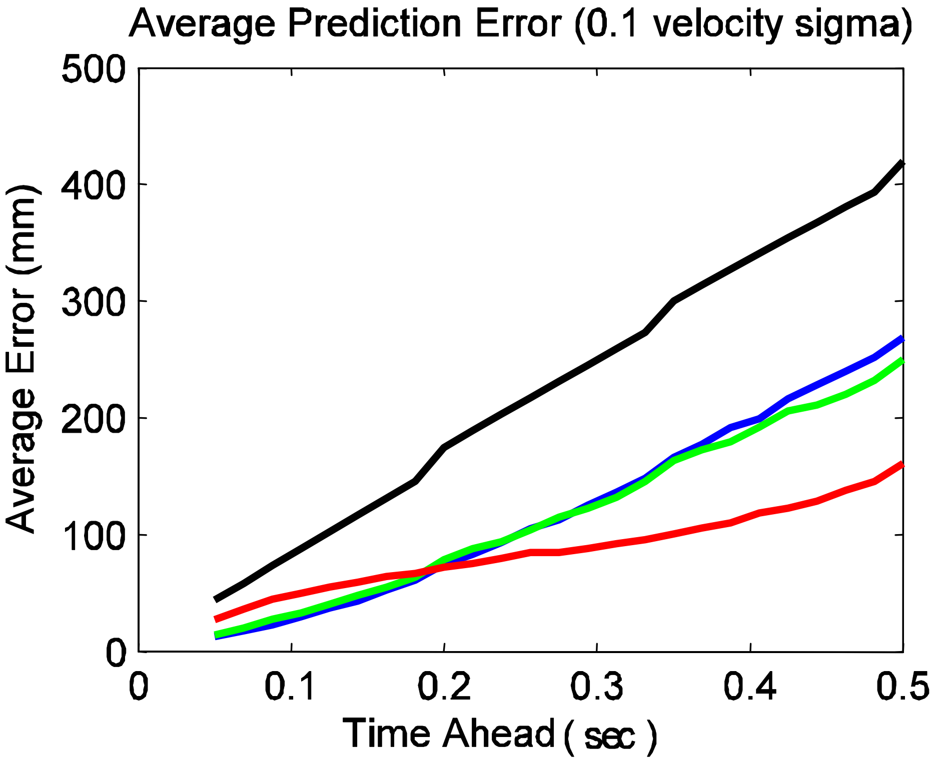
Physics (active mocap control):

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{f}(\mathbf{q}_{\mathbf{f}}, \dot{\mathbf{q}}_{\mathbf{f}}, \pi^{\mathbf{A}}) + \mathbf{N}(0, \Sigma)$$

How Good is our Motion Model?

Given: a pose (position and velocity) of the body
from walking mocap data

Goal: predict the state some time **Δt** into the future

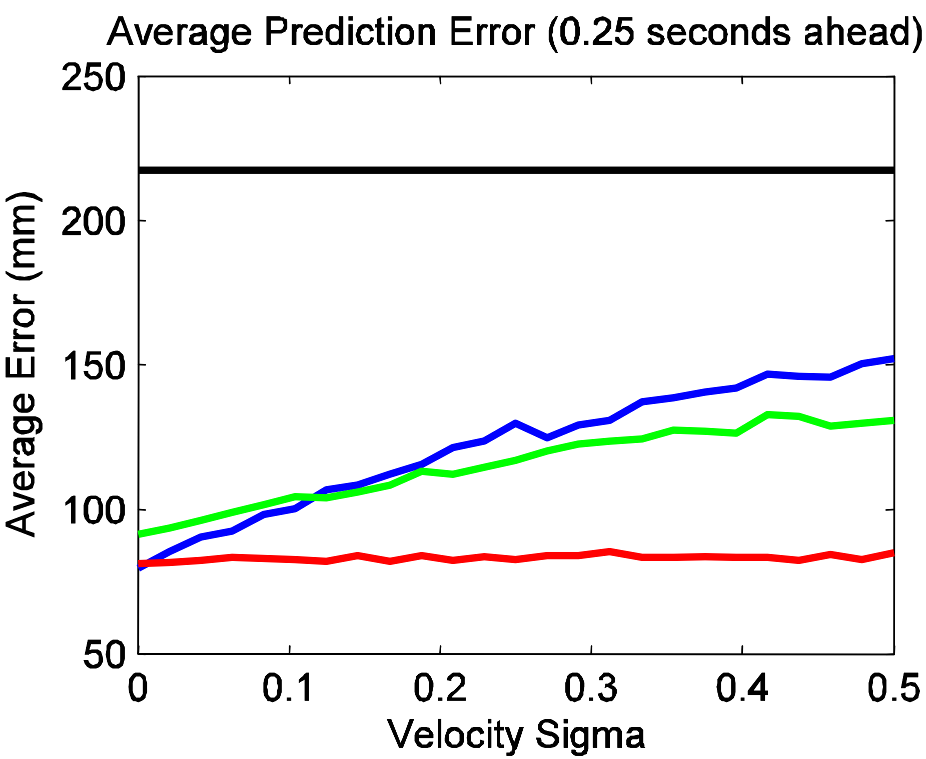


Long-term predictions are more accurate with actively controlled physics-based model

How Good is our Motion Model?

Given: a pose (position and ***corrupted*** velocity) of the body from walking mocap data

Goal: predict the state some time $\mathbf{\Delta t}$ into the future



No prior:

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{q}_{\mathbf{f}} + \mathbf{N}(0, \Sigma)$$

Constant velocity prior:

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{q}_{\mathbf{f}} + \mathbf{\Delta t} \cdot \dot{\mathbf{q}}_{\mathbf{f}} + \mathbf{N}(0, \Sigma)$$

Physics (passive rag-doll):

$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{f}(\mathbf{q}_{\mathbf{f}}, \dot{\mathbf{q}}_{\mathbf{f}}, \pi^{\mathbf{P}}) + \mathbf{N}(0, \Sigma)$$

Physics (active mocap control):

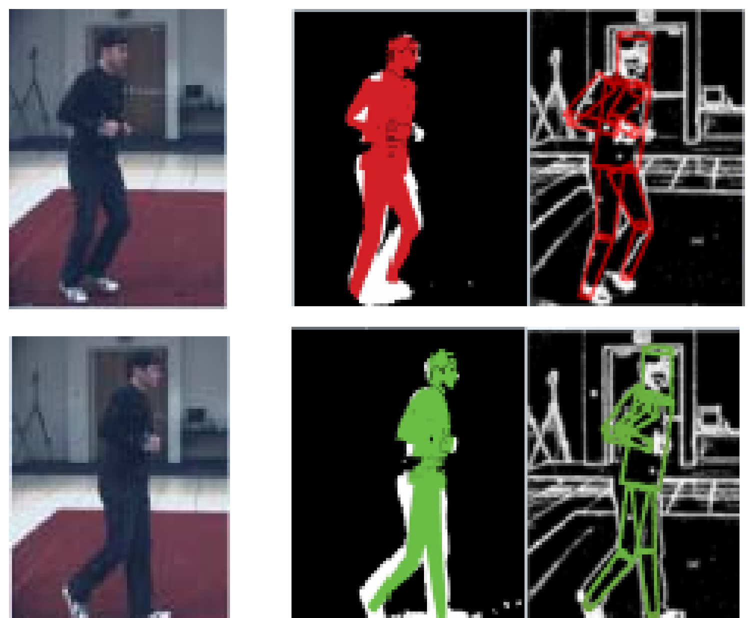
$$\mathbf{q}_{\mathbf{f}+1} = \mathbf{f}(\mathbf{q}_{\mathbf{f}}, \dot{\mathbf{q}}_{\mathbf{f}}, \pi^{\mathbf{A}}) + \mathbf{N}(0, \Sigma)$$

Likelihood Model

- Measures how well a state hypothesis explains image observations **[Balan et. al., '05] [Deutscher et. at., '00]**
  - We use a generic likelihood model based on edges and silhouettes
  - Combine observations from different sources of information and across camera views assuming independence
- States carry both position and velocity information

f

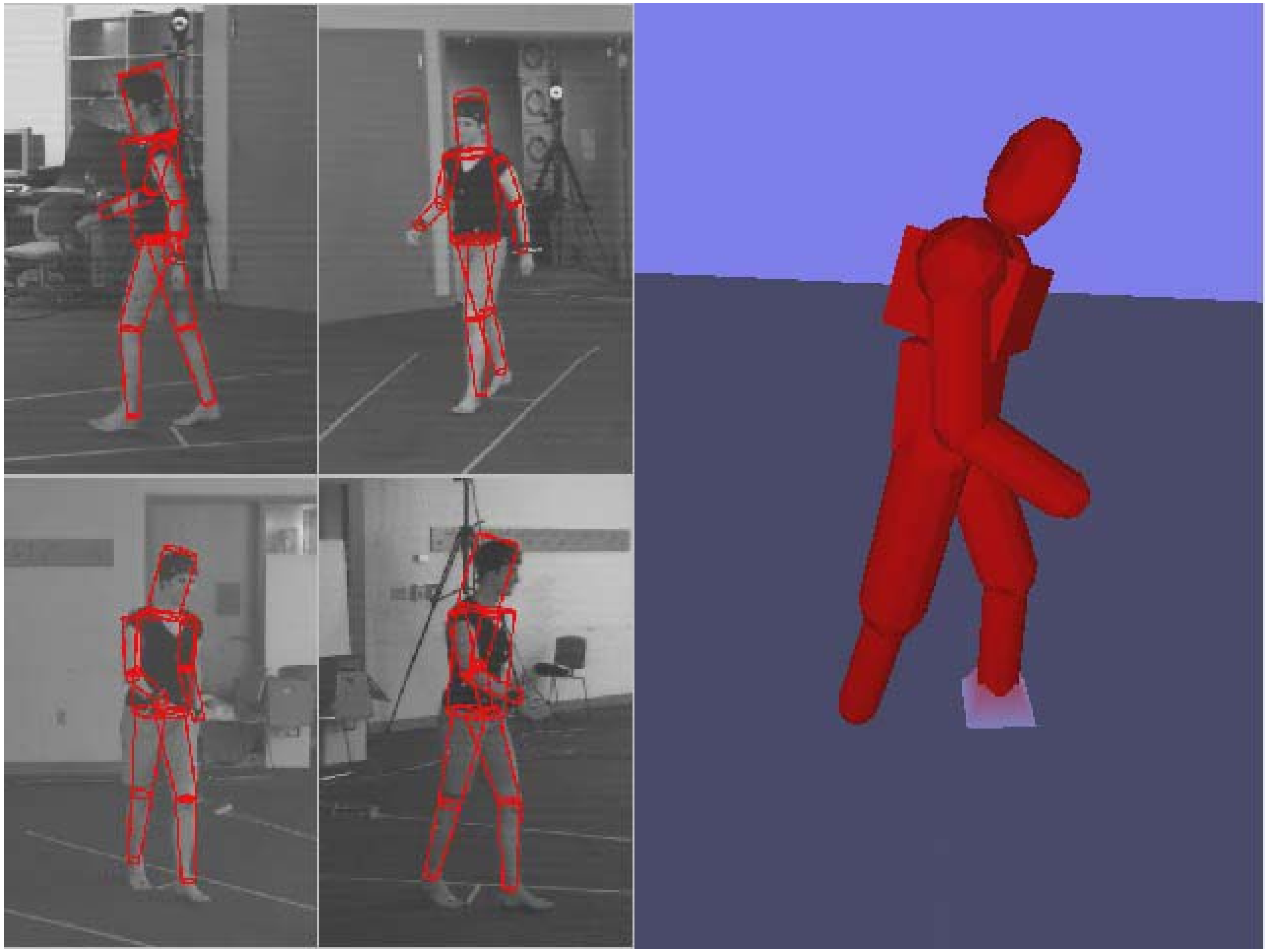
f+1



See:

[Brubaker et. al., '07]

Multi-view Tracking

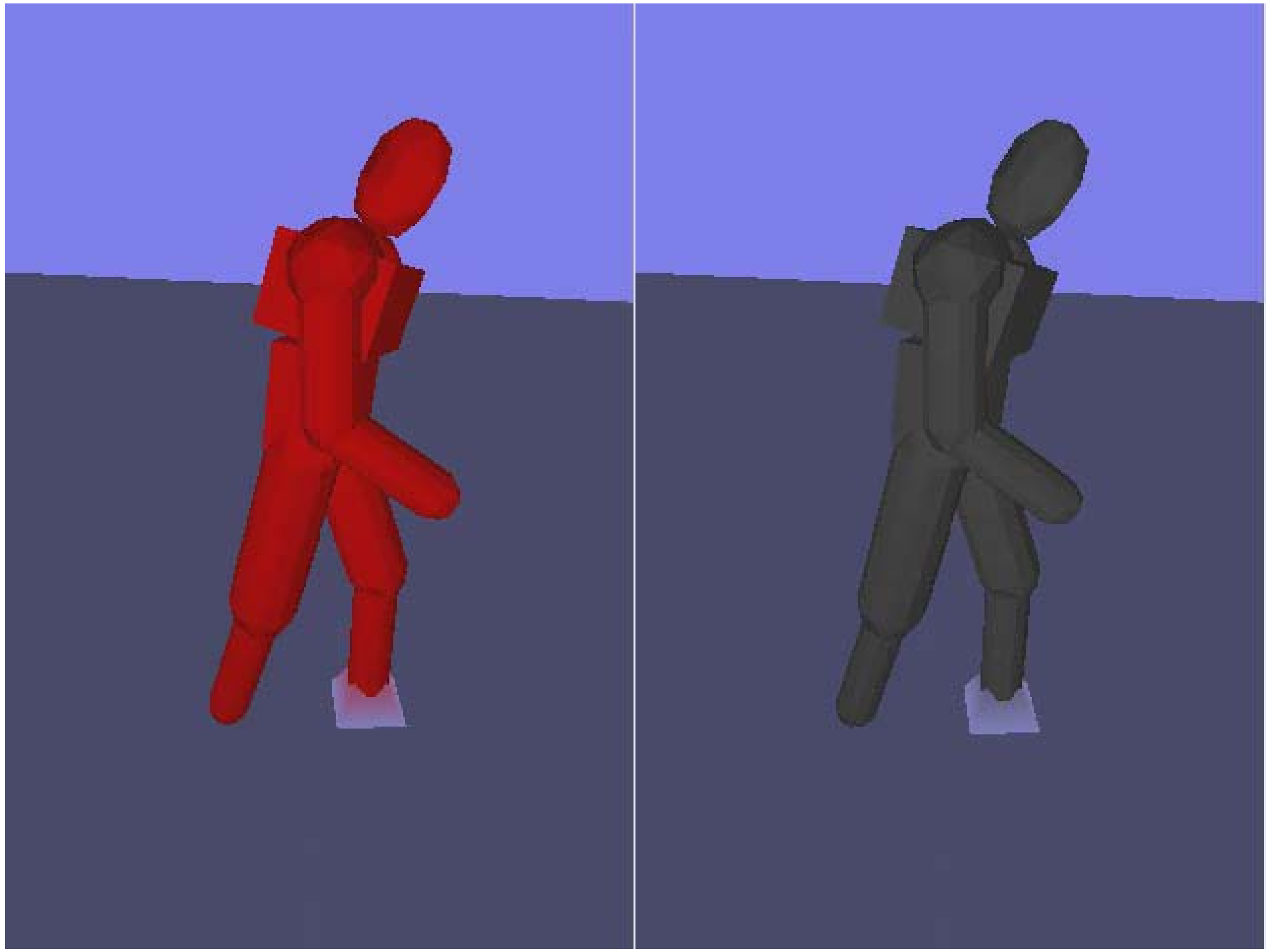


Purple square: Ground contact

Inference: Particle Filtering
Motion Prior: Physics-based

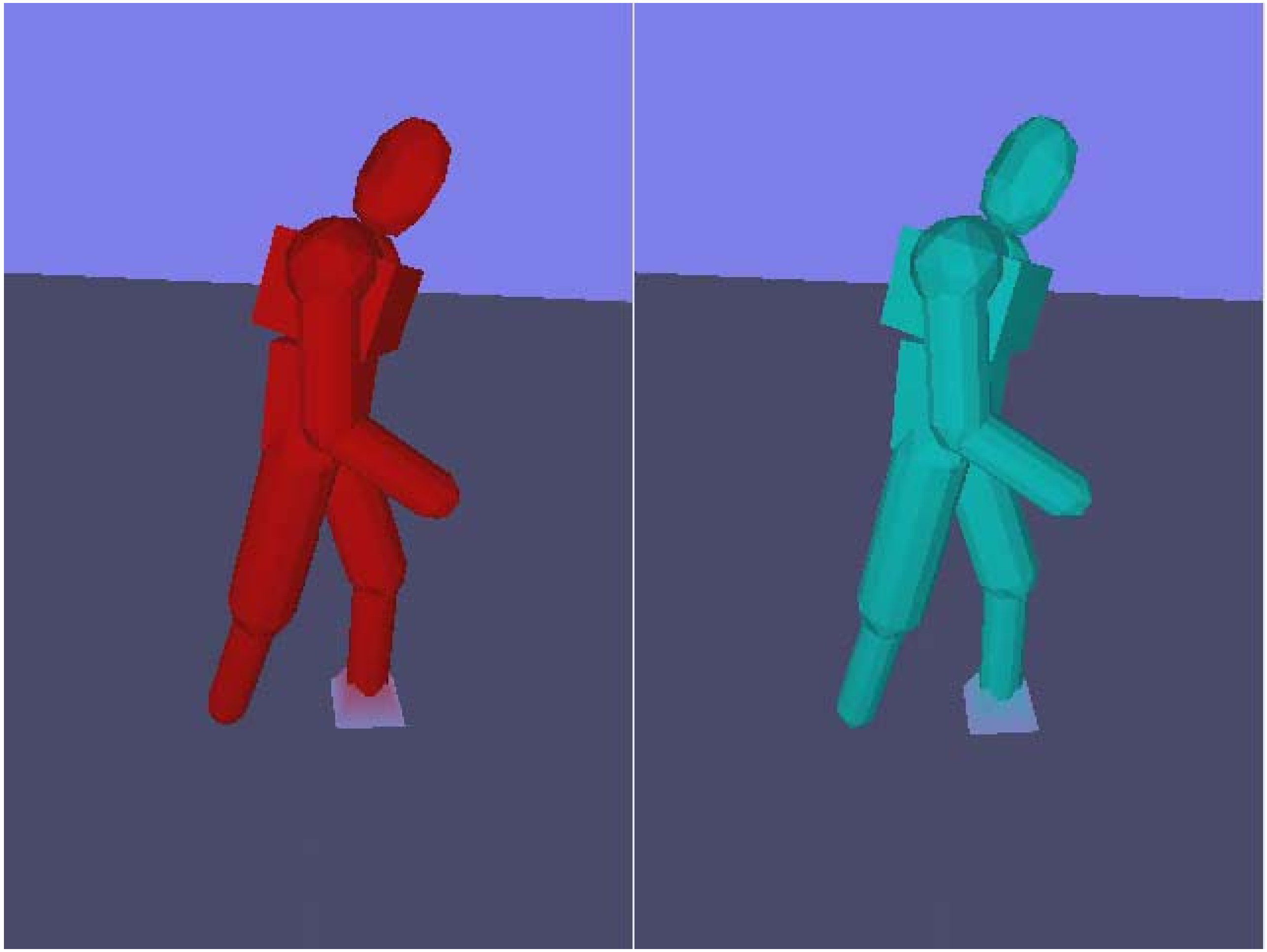
Multi-view Tracking: Comparison

with equal number of particles



Inference:	**Particle Filtering**	**Particle Filtering**
Motion Prior:	**Physics-based**	**Smooth prior**

Multi-view Tracking: Comparison

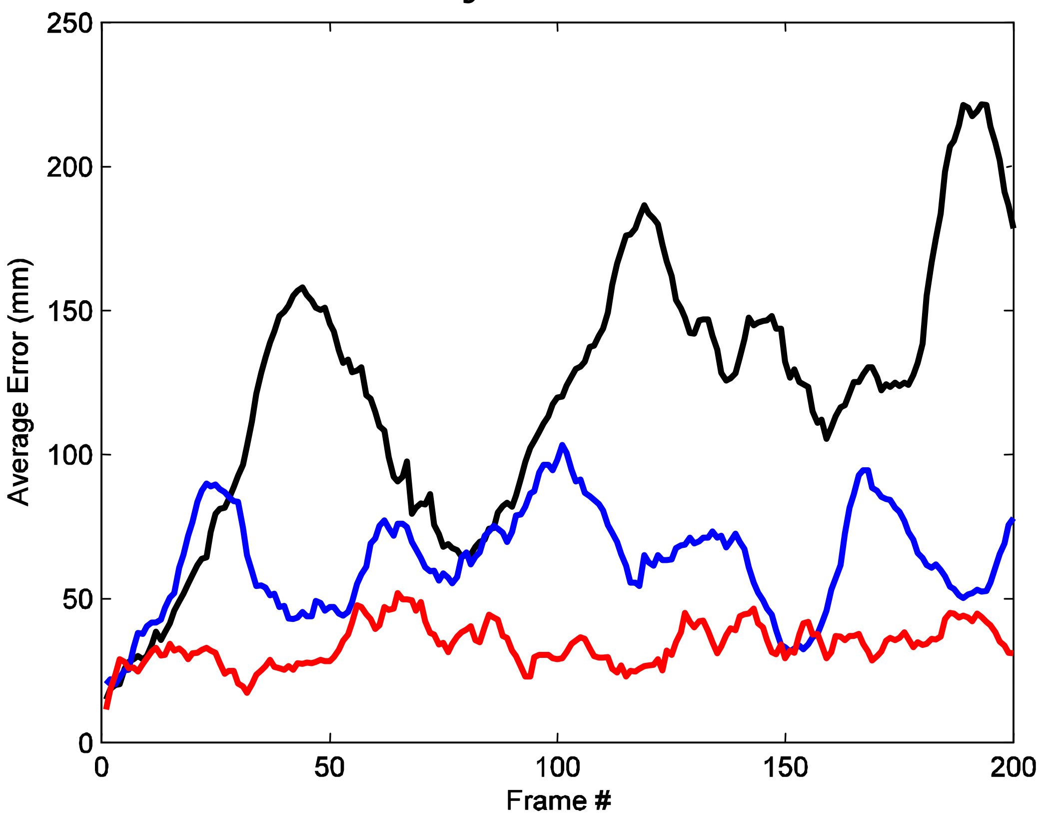


Inference:	**Particle Filtering**	**Annealed Particle Filtering**
Motion Prior:	**Physics-based**	**Smooth prior**

Quantitative Comparison: Multi-view

PF	APF 5	PF Physics
120.7 ± 46.9	**63.5** ± 17.9	**33.9** ± 7.2

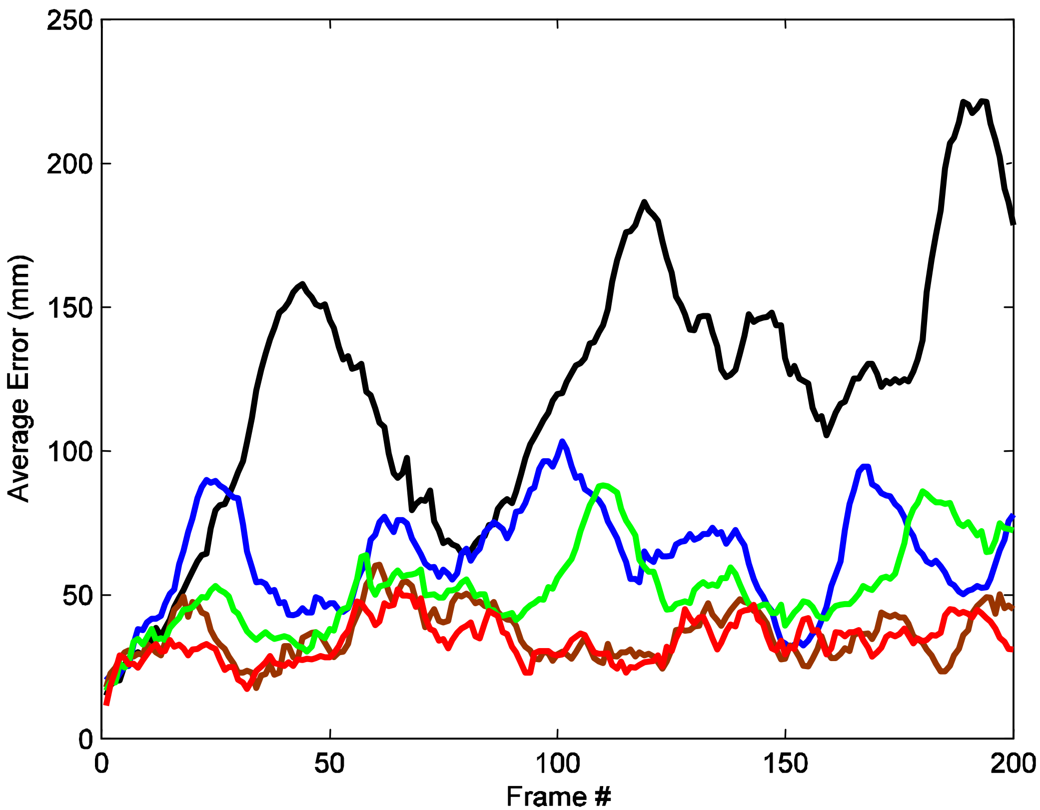
- HumanEva Dataset **[Sigal et. al., '06]**
  - contains synchronized motion capture data and multi-view video



More accurate performance with lower variance

Quantitative Comparison: Multi-view

PF	APF 5	PF Physics	PF Physics	PF Physics
120.7 ± 46.9	**63.5** ± 17.9	**33.9** ± 7.2	**36.3** ± 9.0	**52.5** ± 15.0
	Training Subject	**L1**	**L1, S1, S2, S3**	**S1, S2, S3**
	Test Subject	**L1**	**L1**	**L1**

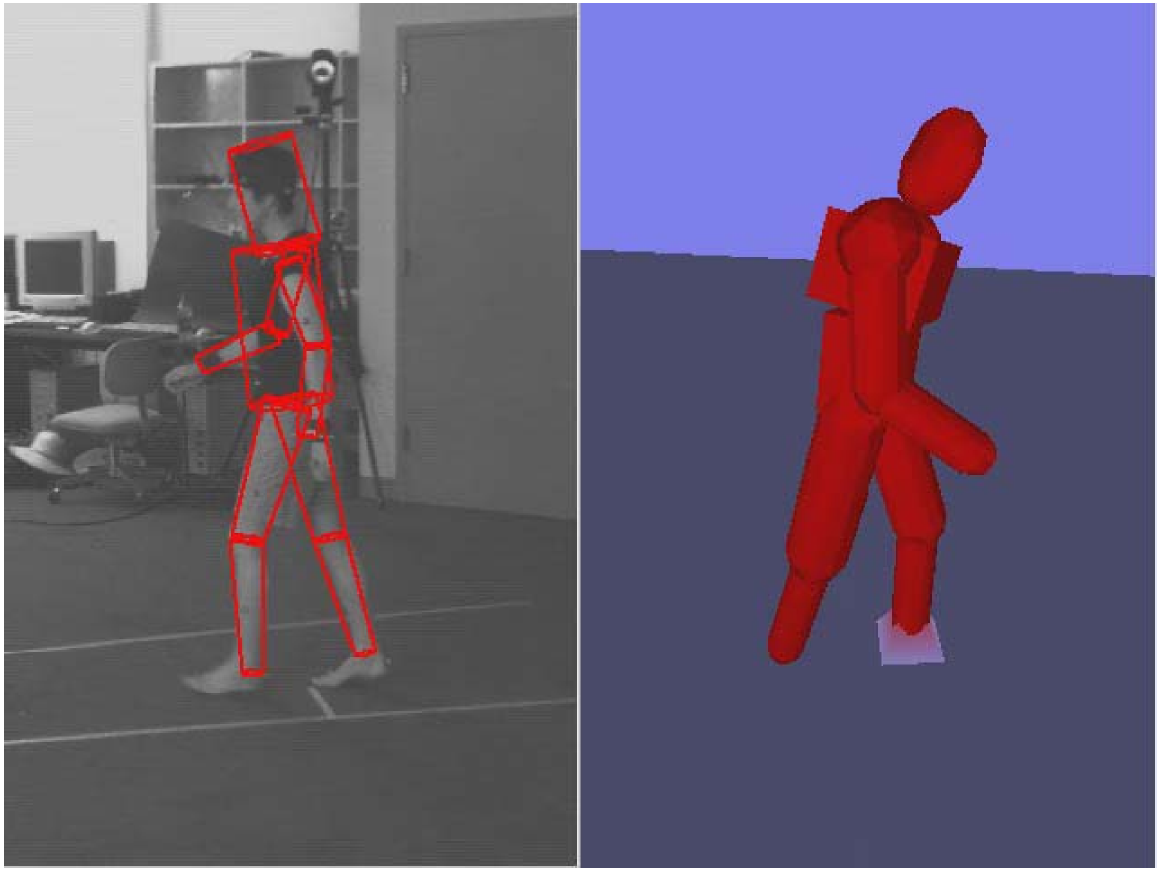


More accurate performance with lower variance

Performance does not degrade with larger training sets

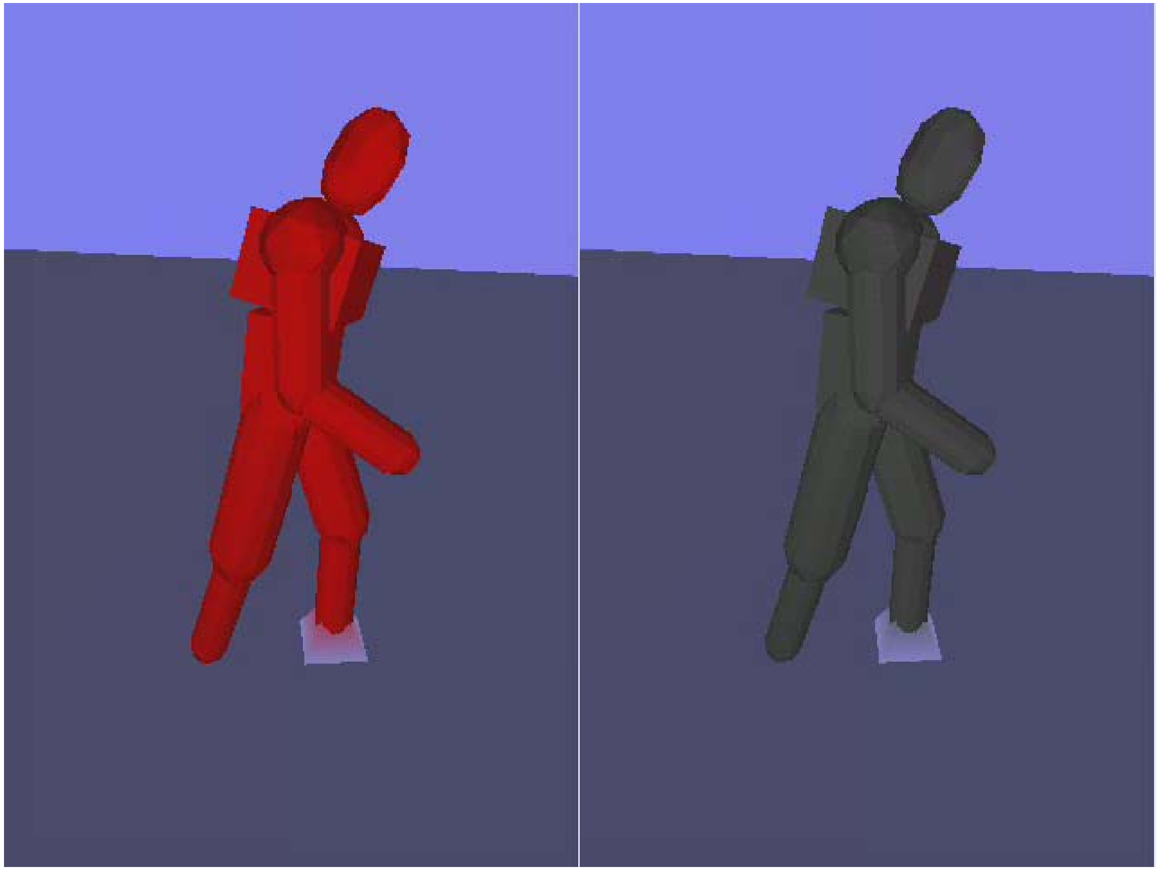
Can generalize to new people (at least for simple motions)

Monocular Tracking



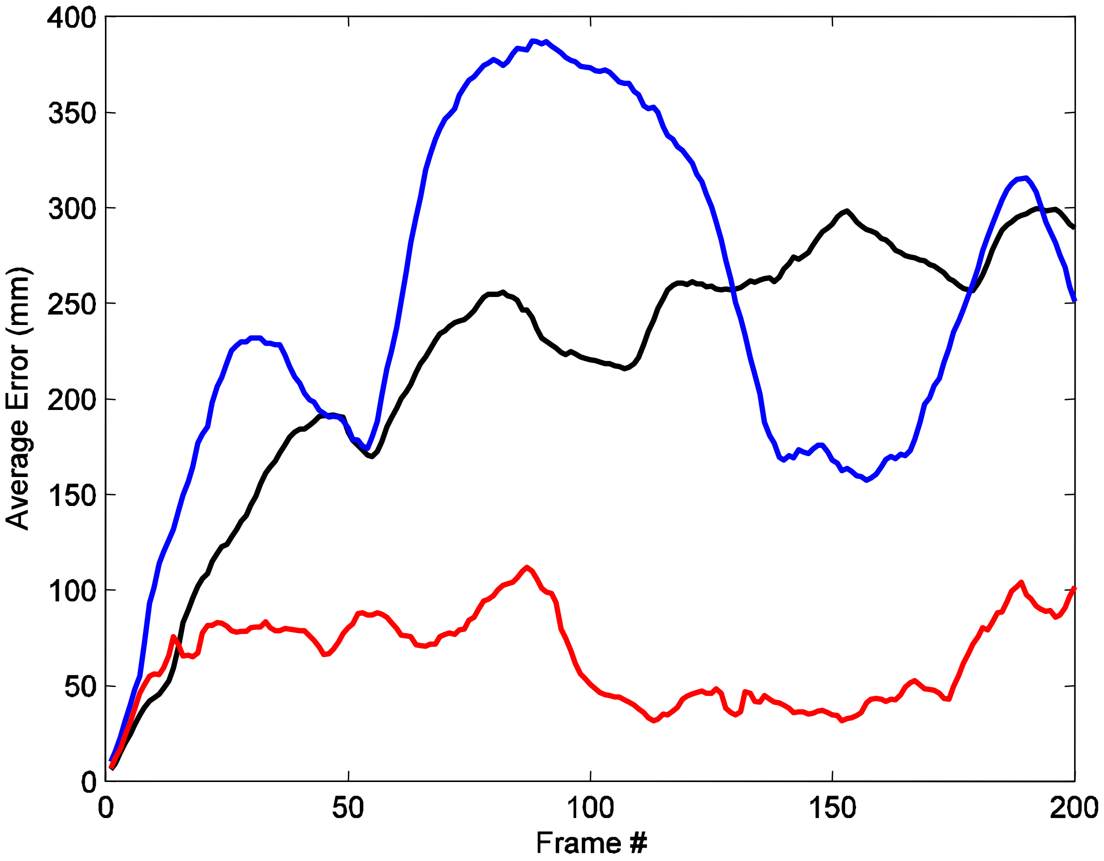
Inference: **Particle Filtering**
Motion Prior: **Physics-based**

Monocular Tracking: Comparison



Inference:	**Particle Filtering**	**Particle Filtering**
Motion Prior:	**Physics-based**	**Smooth prior**

Quantitative Comparison: Monocular

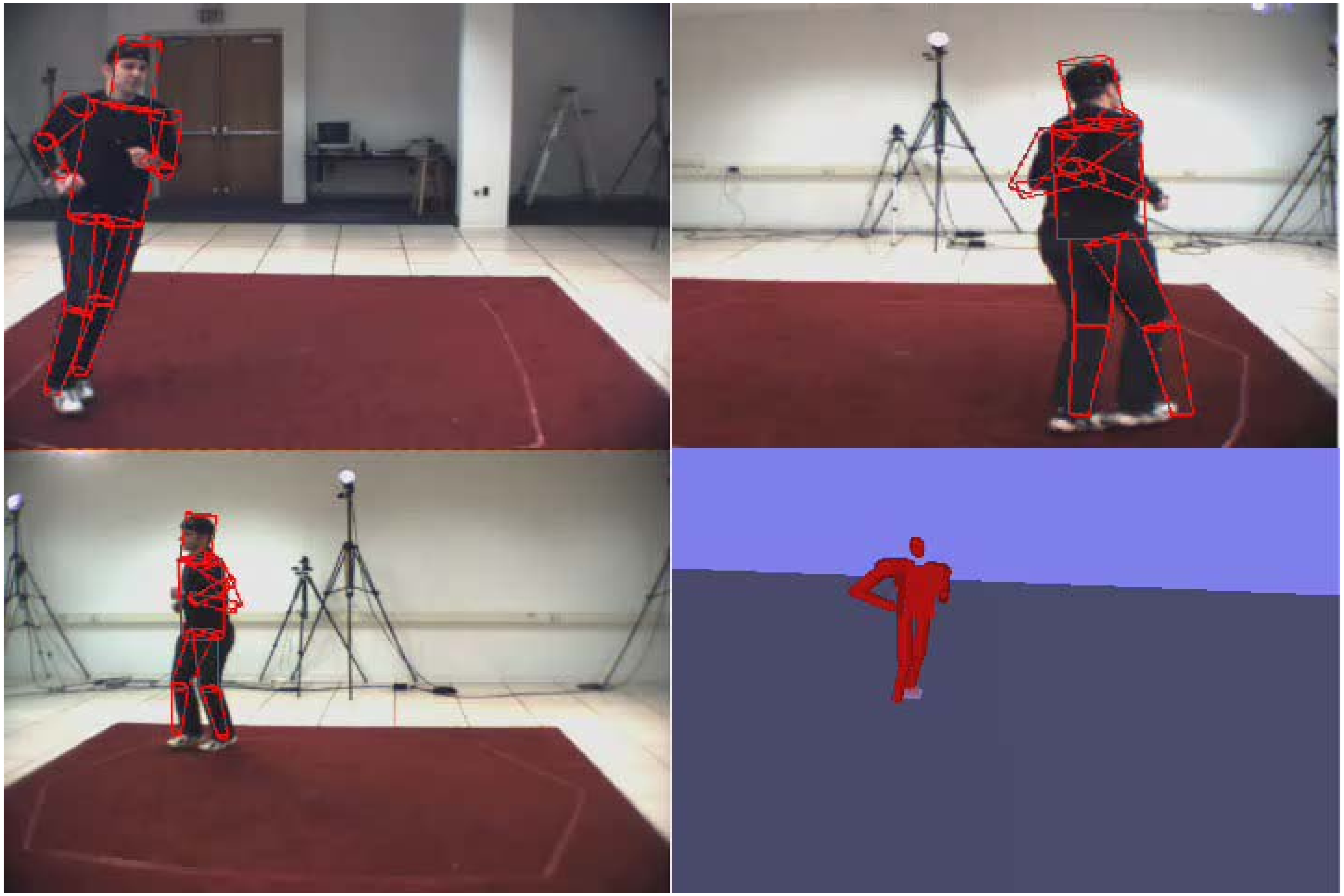


PF	APF 5	PF Physics
219.2 ± 72.1	**249.6** ± 91.0	**64.2** ± 23.4

More accurate performance with lower variance

What about other motions?

e.g., Jogging



Conclusions

- **Framework for physically-plausible tracking**
  - Incorporates the full-body physics-based simulation as a temporal prior within the context of Bayesian filtering
    - Non-linear non-stationary dynamics of the human body
    - Interactions with the environment
  - We also introduce novel hybrid constraint-based controller
  - We show both qualitatively and qualitatively that the resulting framework is more accurate and physically plausible than results obtained using standard priors in Bayesian filtering methods

Future Work

- **More effective inference methods**
  - Better proposal functions for particle filtering
  - Richer observation models
- **More realistic models of human motion**
  - Better models of ground contact with multi-segment feet
  - Active balancing
- **Weaker reliance on motion capture data**
  - Action-specific controllers (or combinations of such)
  - Learning of priors over motor forces for specific classes of motion (hard problem even for biomechanics)

Acknowledgments

- **Funding**
  - ONR Young Investigator Award (C. Jenkins)
  - PECASE (C. Jenkins)
- **Personal thanks**
  - Michael J. Black, Brown University
  - Alexandru Balan, Brown University
  - David Fleet, University of Toronto
  - Matt Loper, Brown University
  - Morgan McGuire, Williams College
  - German Gonzalez, EPFL
  - Jaroslav Semancik, Charles University in Prague