Physics-Based Models for People Tracking: Introduction and Motivation

Marcus Brubaker[1] Leonid Sigal[1,2] David J Fleet[1]

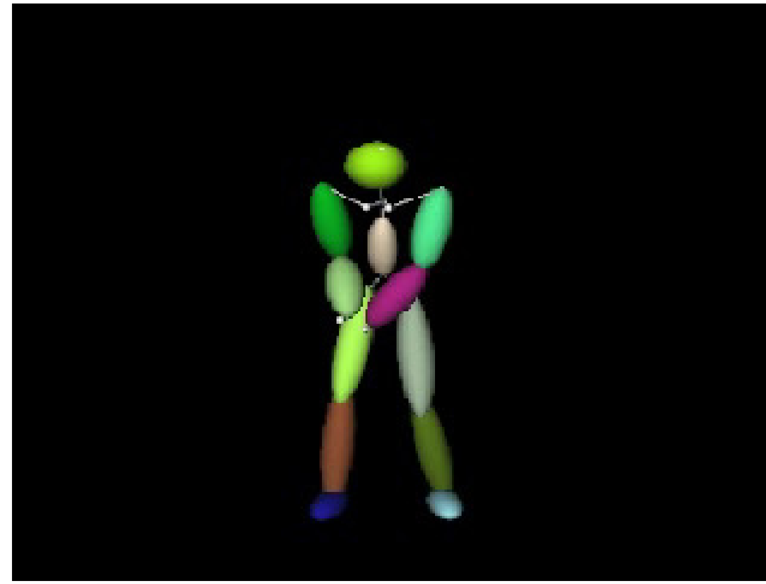
[1]University of Toronto

[2]Disney Research, Pittsburgh

Looking at people

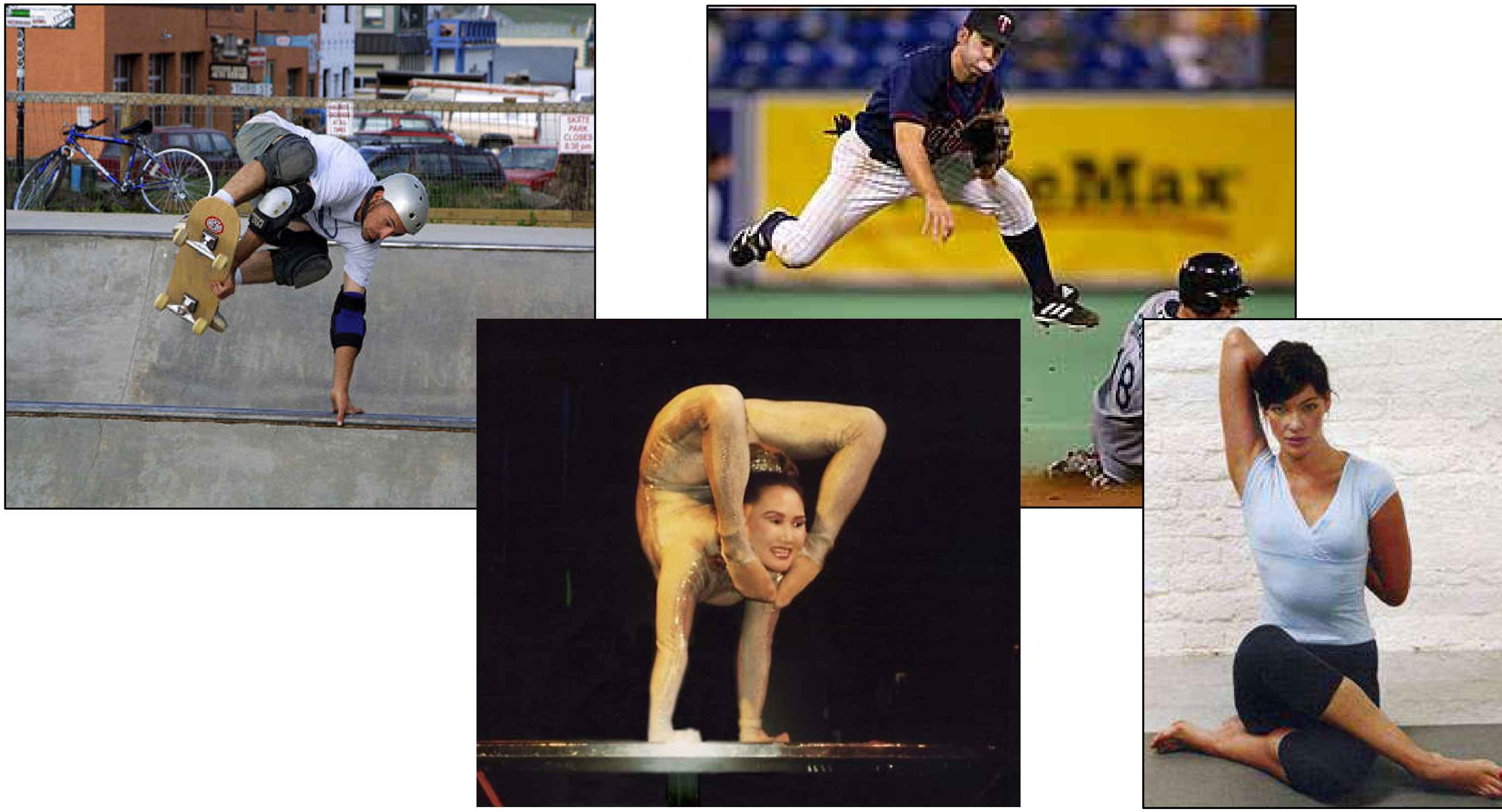
3D pose tracking

Video input



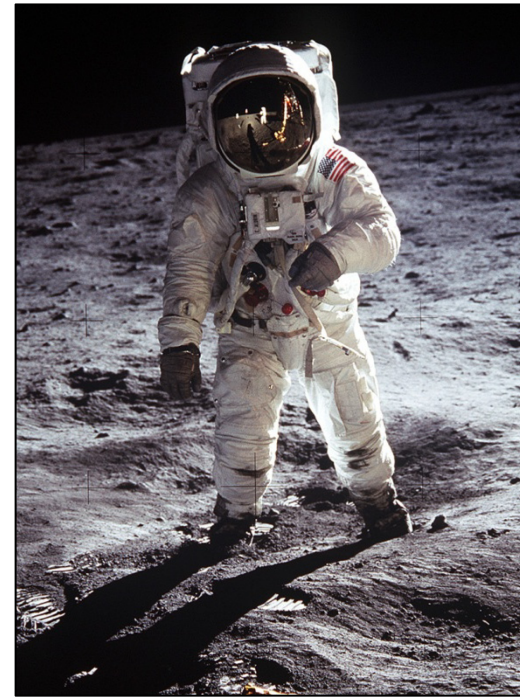
3D articulated model

Challenges: Complex pose / motions



People have many degrees of freedom, comprising an articulated skeleton overlaid with soft tissue and deformable clothing.

Challenges: Appearance, size and shape



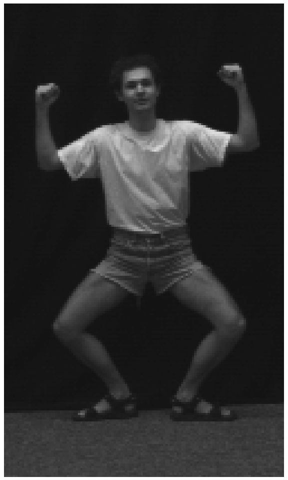
People come in all shapes and sizes, with highly variable appearance.

Challenges: Noisy and missing data

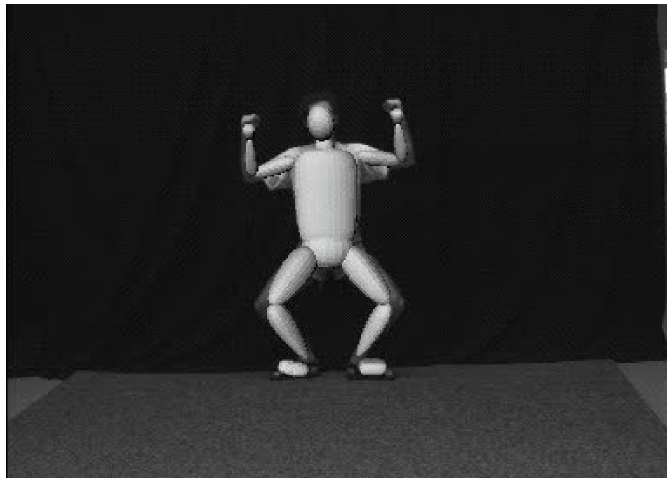
Ambiguities in pose are commonplace, e.g., due to

- background clutter
- apparent similarity of parts
- occlusions
- loose clothing
- …

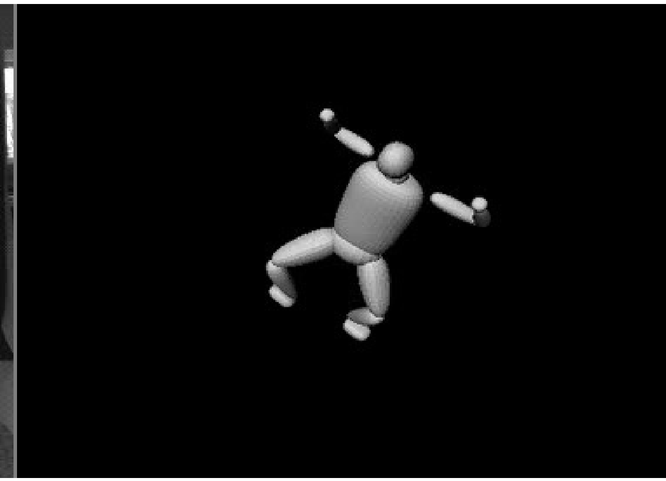
Challenges: Depth and reflection ambiguities



image



3D model
(camera view)



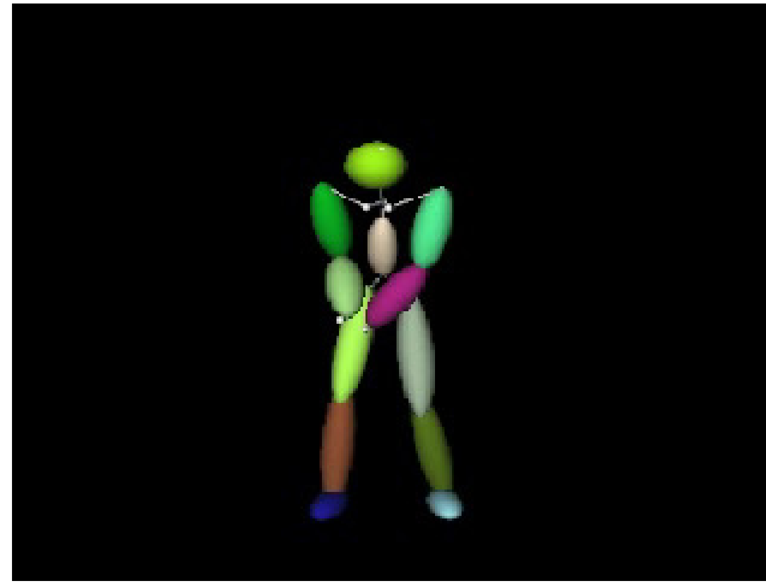
3D model
(top view)

Multiple 3D poses may be consistent with a given image.

[*courtesy of Cristian Sminchisescu*]

Model-based pose tracking

Video input



3D articulated model

Pose tracking as Bayesian filtering

Generative Models

$$p(motion \,|\, video) = \frac{p(video \,|\, motion)\; p(motion)}{p(video)}$$

Pose tracking as Bayesian filtering

Generative Models

$$p(motion \mid video) = \frac{\overbrace{p(video \mid motion)}^{\textit{likelihood}}\ p(motion)}{p(video)}$$

The likelihood models the consistency of image measurements with the projection of the body into the image (e.g., using silhouettes, edges, optical flow, foreground color, etc).

Pose tracking as Bayesian filtering

Generative Models

$$p(motion \mid video) = \frac{p(video \mid motion)\, \overset{\textit{prior}}{p(motion)}}{p(video)}$$

The prior models how people typically move (e.g., it can be manually specified, data-driven, or learned).

Pose tracking as Bayesian filtering

Generative Models

$$p(motion \,|\, video) \;=\; \frac{p(video \,|\, motion)\; p(motion)}{p(video)}$$

Inference (estimation) involves Bayesian filtering, often with Monte Carlo approximation, or gradient-based optimization.

Pose tracking as Bayesian filtering

Generative Models

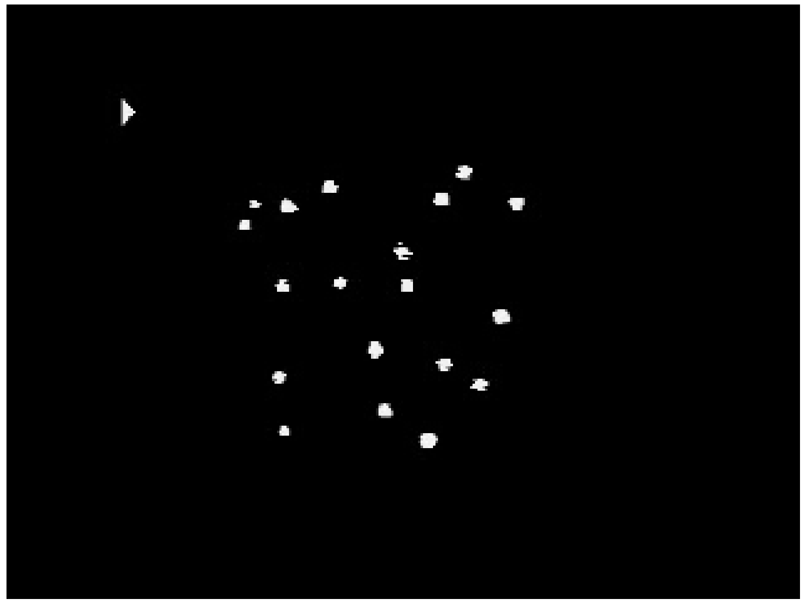
$$p(motion \,|\, video) = \frac{p(video \,|\, motion)\; p(motion)}{p(video)}$$

Discriminative Models

$$3D\ pose \;=\; f(\,image\ measurements\,)$$

Regression models are learned from training data (e.g., using k-NN, RVM regression, mixtures of experts, latent variable regression, etc).

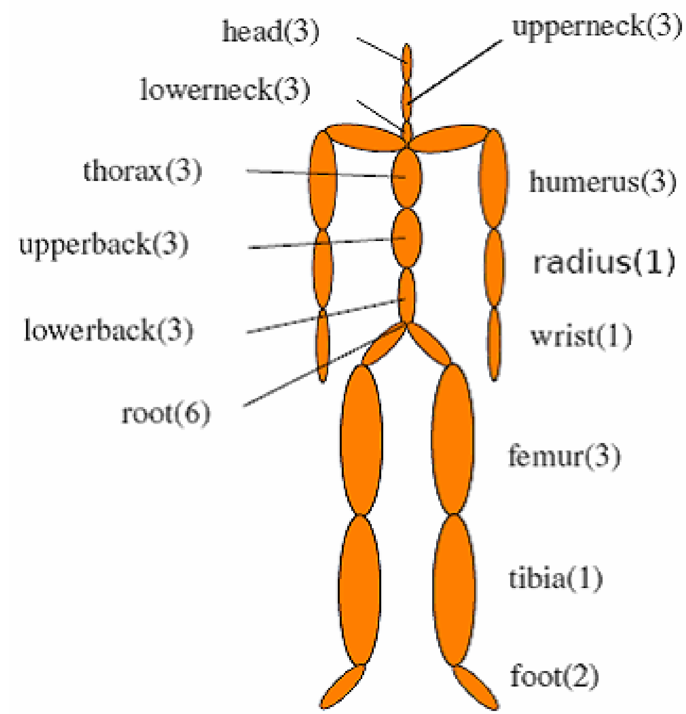
Motion capture data



[*Johansson, 1973*]

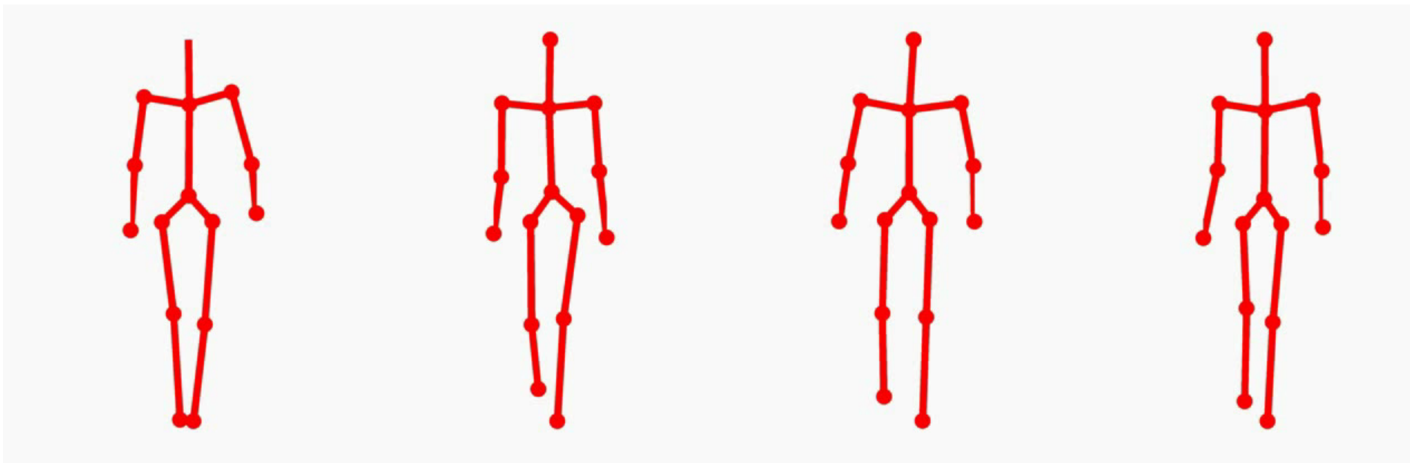
Motion capture data

motion capture

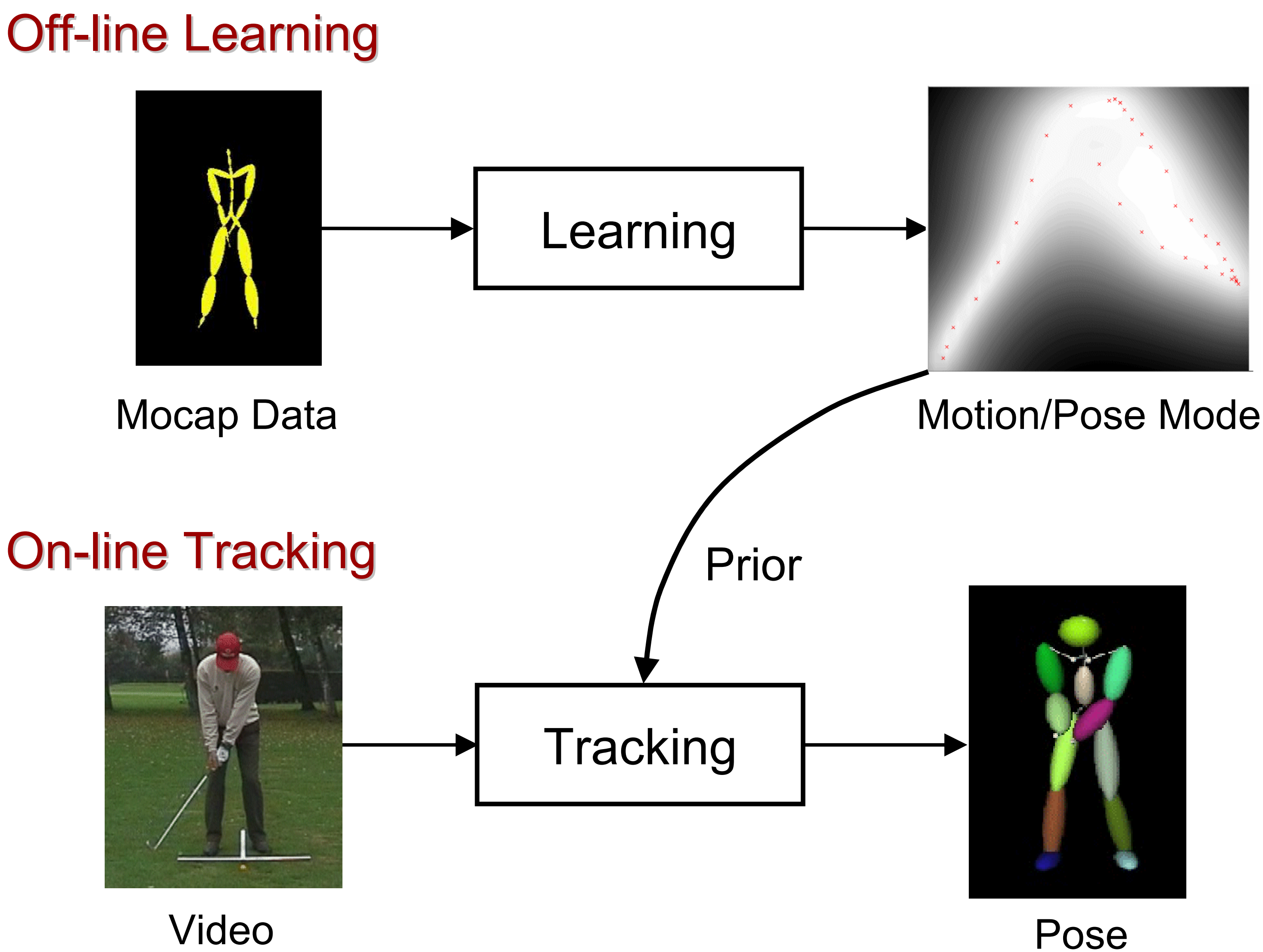


3D articulated model

Motion capture data

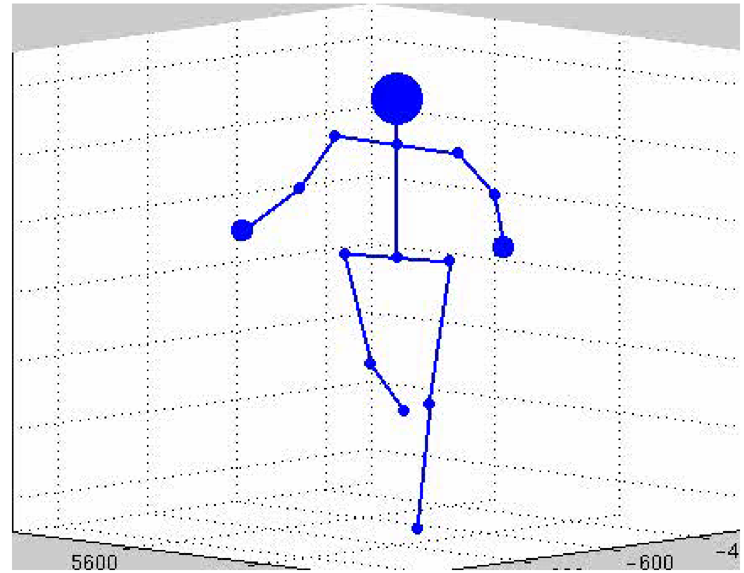


Model-based pose tracking



How well will kinematic models generalize?

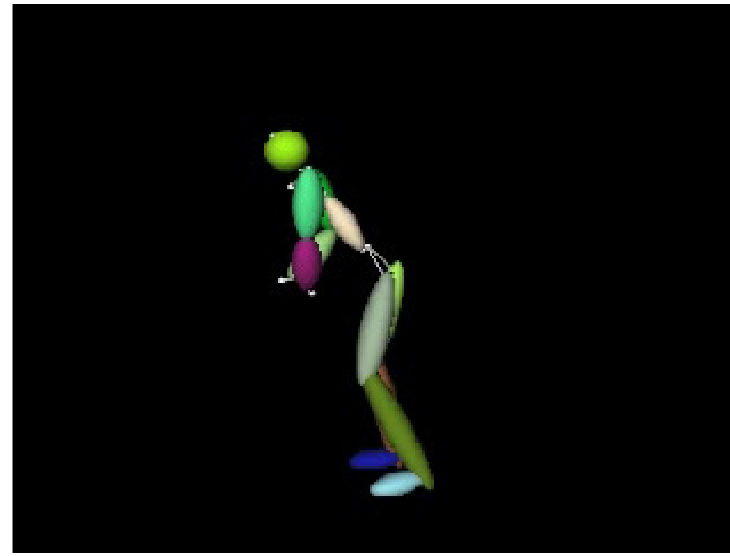
Implausible motions



[*Poon and Fleet, '01*]

- Kinematic Model: damped 2nd-order Markov model with Beta process noise and joint angle limits
- Observations: steerable pyramid coefficients (image edges)
- Inference: Hybrid Monte Carlo particle filter

Implausible motions



[*Urtasun et. al., ‘05*]

- Kinematic Model: GPLVM for pose, with 2nd-order dynamical model
- Observations: tracked 2D patches on body (WSL tracker)
- Inference: MAP estimation (hill climbing)

Discriminative Results

[*Kanaujia et al., '07*]

- Inference: mixture of experts (with semi-supervised learning)
- Features: hierarchical image encoding

Will learning scale?

Problem: Learning kinematic pose and motion models from motion capture data, with dependence on the environment and interactions with other bodies, may be untenable …

Physics-based models

Physics specifies the motions of bodies and their interactions in terms of inertial descriptions and forces, and generalize naturally to account for:

- balance and body lean (e.g., on hills)
- sudden accelerations (e.g., collisions)
- static contact (e.g., avoiding footskate)
- variations in style due to speed and mass distribution (e.g., carrying an object)
- …

Why physics-based models?

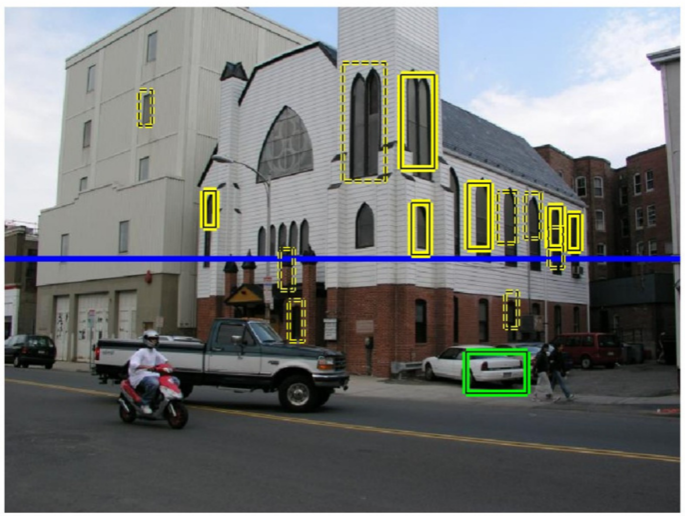
Three things:

- ensure physically plausible tracking results
- reduce dependence of models on mocap (generalization)
- incorporate interactions into tracking formulation

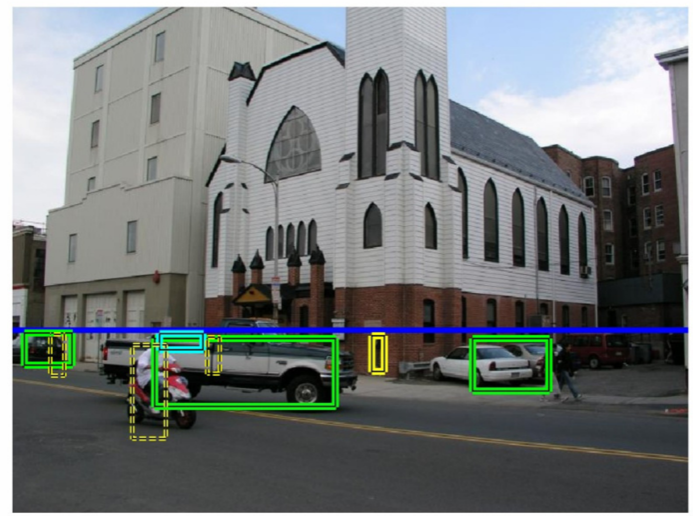
Importance of context

Knowledge of the ground plane places strong constraints on the expected locations and motions of objects.

no context



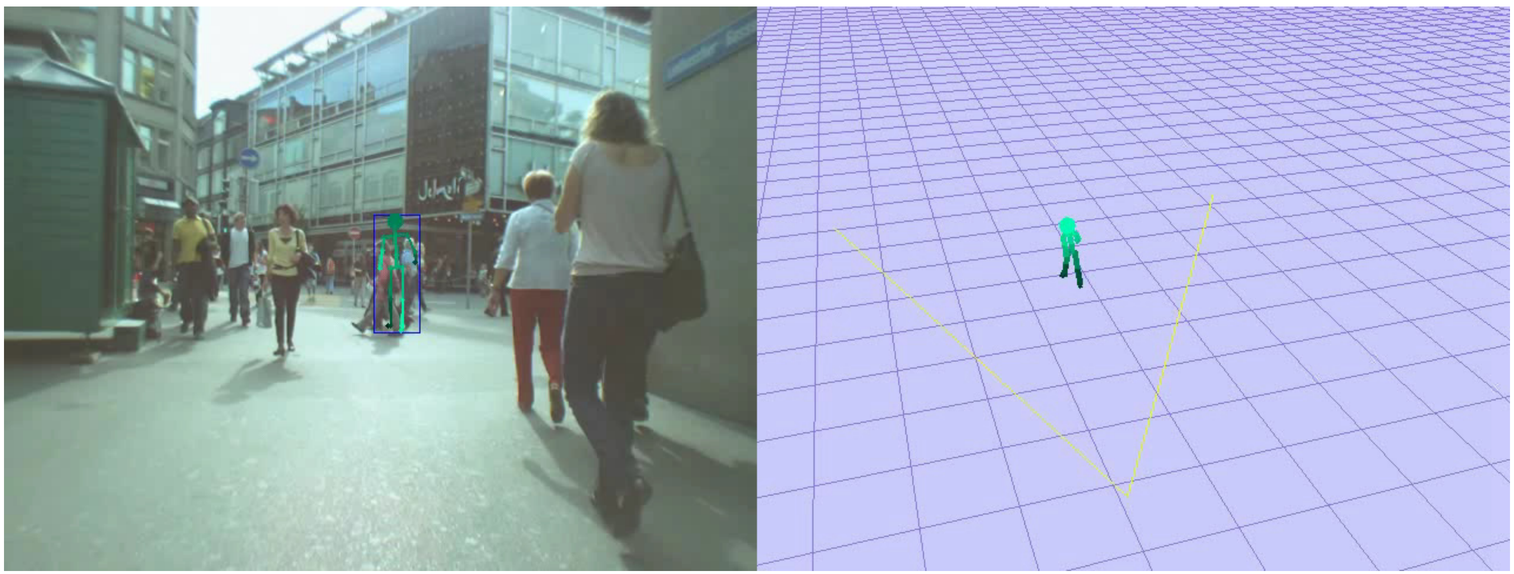
with context



[*Hoiem et al., ‘09*]

Importance of context

Knowledge of the ground plane places strong constraints on the expected locations and motions of objects.



[*Gammeter et al., ‘08*]

Why physics-based models?

Three things:

- ensure physically plausible tracking results
- reduce dependence of models on mocap (generalization)
- incorporate interactions into tracking formulation

Why physics-based models?

Physics-based models in several related fields provide ideas for models suitable for human pose tracking and analysis in vision.

Robotics

Control of machines that interact with unpredictable environments.

Humanoid robotics

Machines that move like people and perform everyday tasks.

[*Honda’s Asimo Robot*]

Humanoid robotics

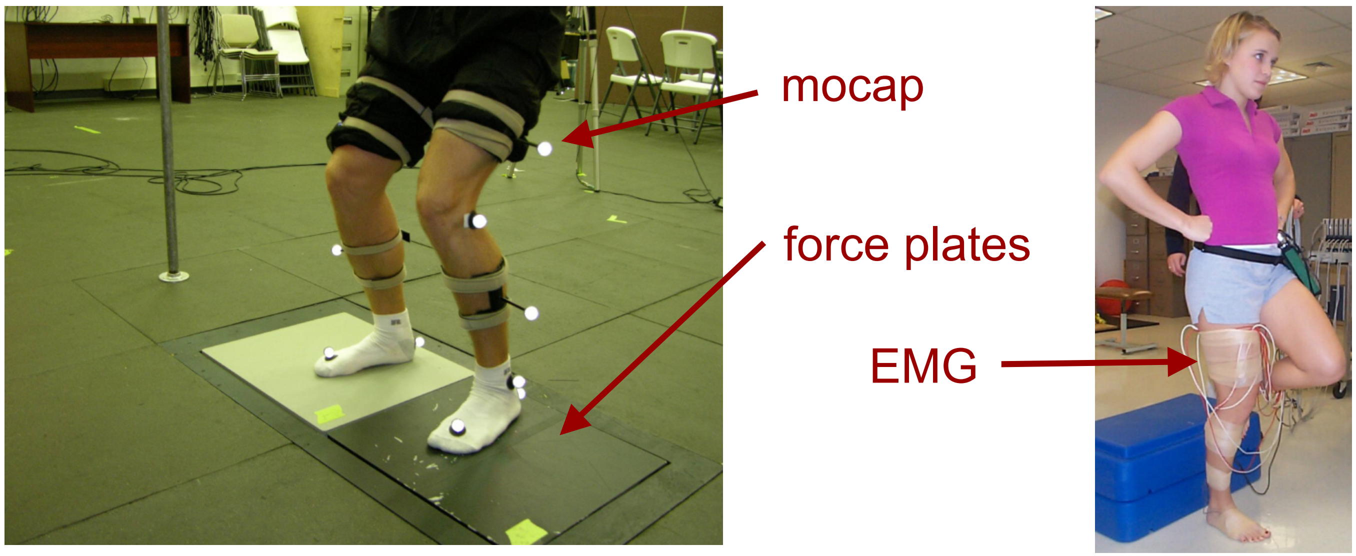
HRP2 performing a Japanese folk dance

[*Nakaoka et al 2004*]

Robots typically have different mass and inertial parameters, torque limits, stability criteria, etc, but approaches to control may be useful.

Biomechanics

Study of human morphology, kinematics and kinetics.



Knowledge of musculature, kinematics and kinetics of locomotion, balance recovery etc. But the neural basis for motor control and the principles underlying human motion less well understood.

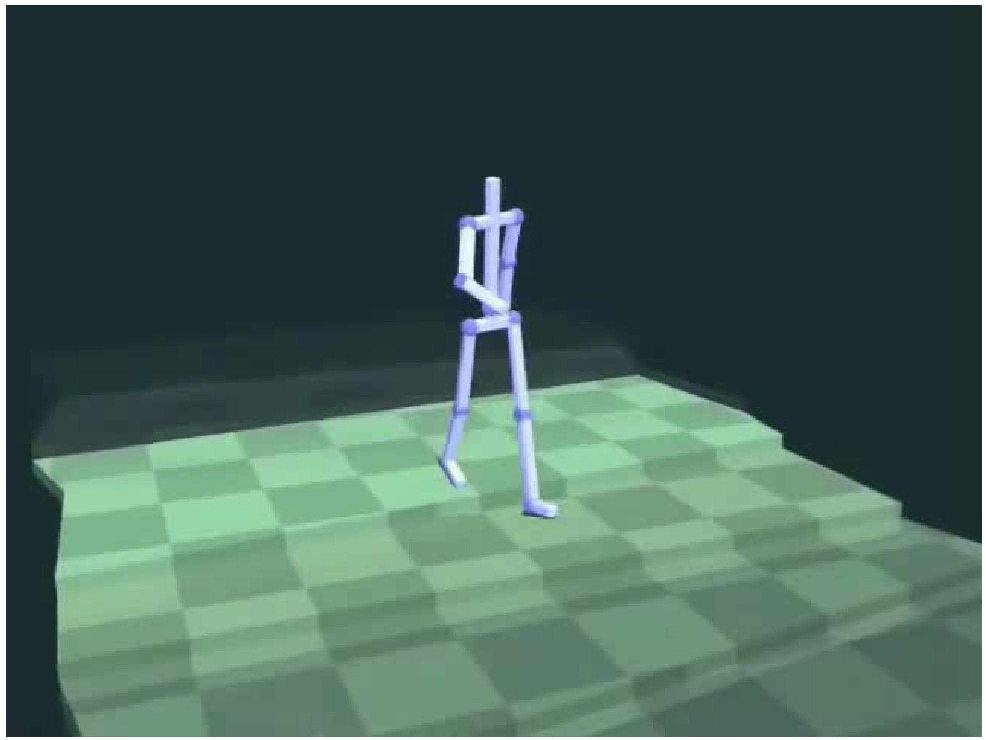
Computer graphics: Space-time optimization

Search for physically plausible motions satisfying user-specified constraints (e.g., foot placements) often while minimizing energy.

[*Liu et al., ’06*]

Computer graphics: Controllers

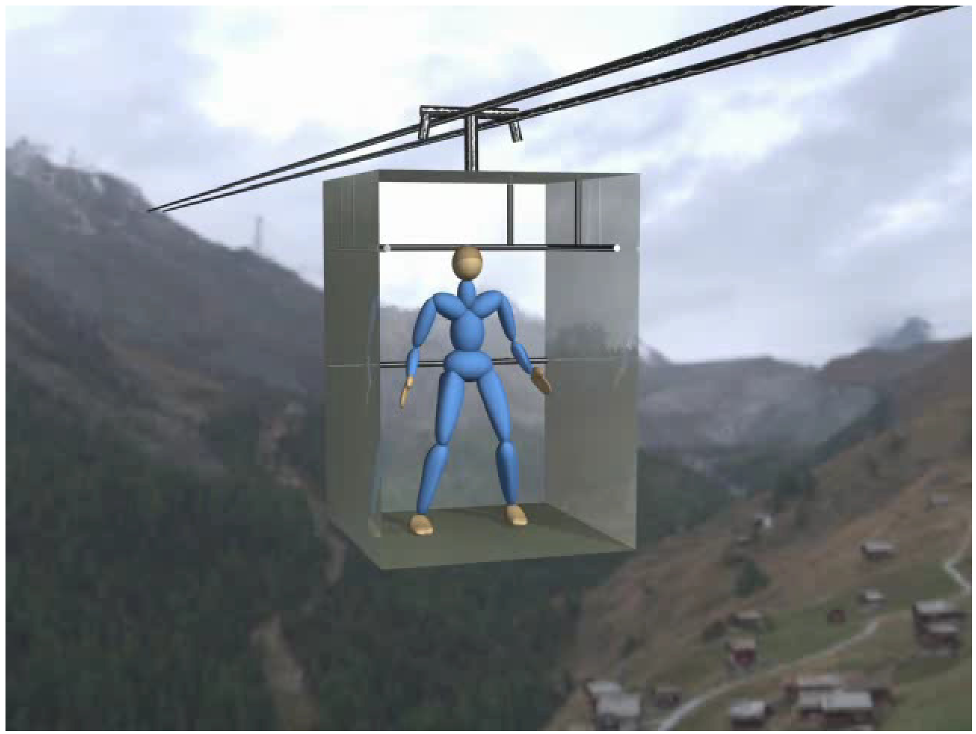
Compact representations for repetitive motions that are reactive and robust to environmental perturbation



[*Muico et al., ’09*]

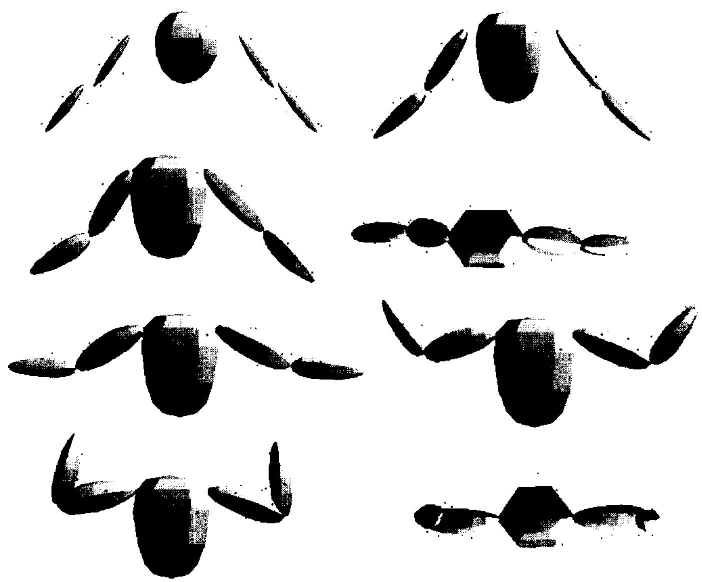
Computer graphics: Interactions

Constrained space-time optimization techniques for handling dynamic interactions with the environment (batch or on-line)

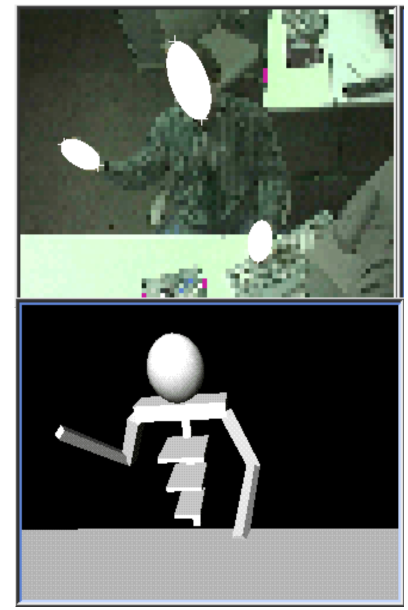


[*Jain et al., ’09*]

Physics-based models in computer vision



[*Metaxas and Terzopoulos,'93*]

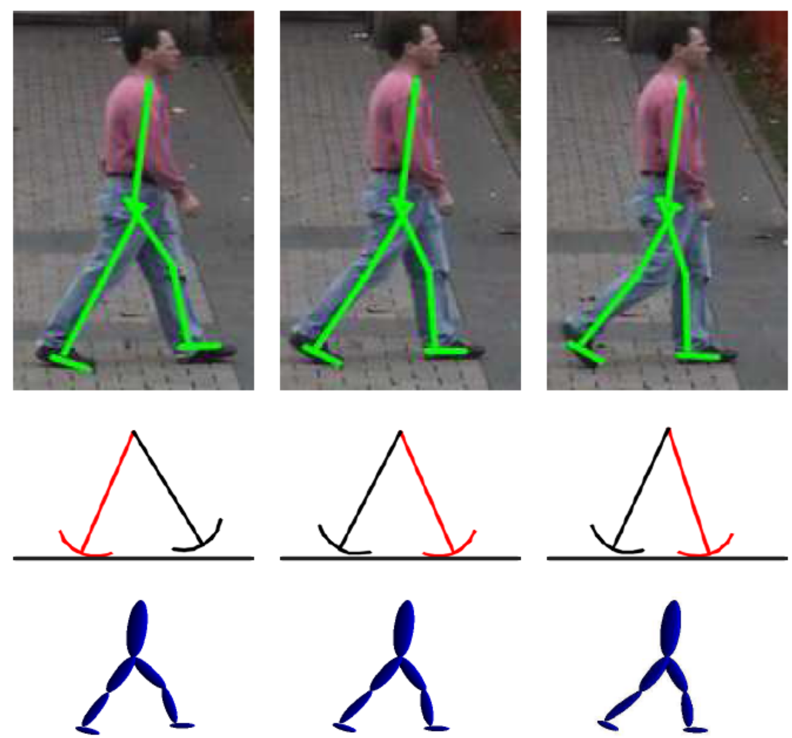


[*Wren and Pentland, '98*]

- Dynamics: upper body only, no contact
- Observations: 3D markers, stereo cameras
- Inference: Kalman filter

Physics-based models for pose tracking

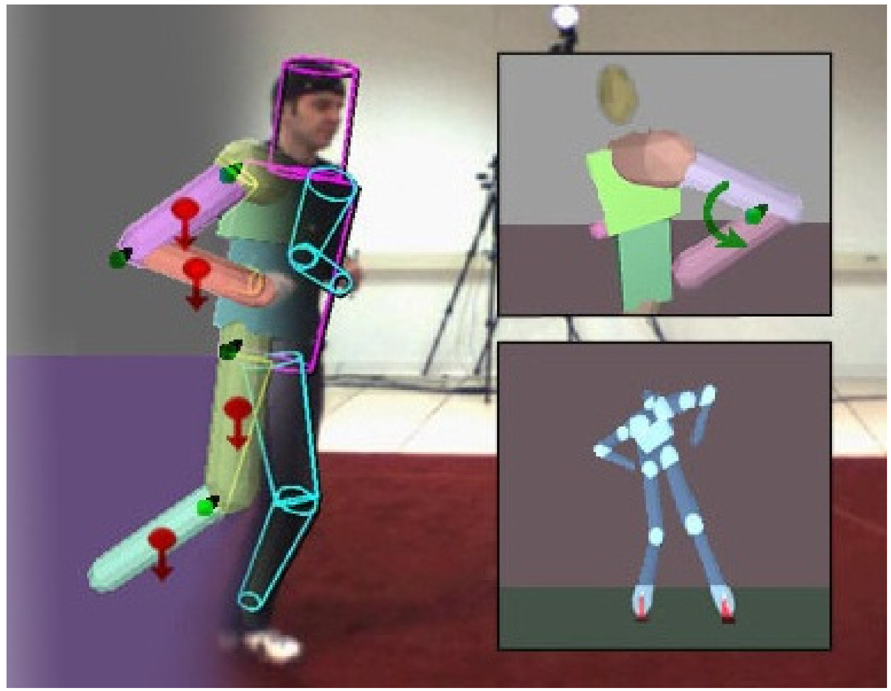
Low-dimensional, biomechanically inspired models of locomotion.



[*Brubaker et al., ’07 / ‘08*]

Physics-based models for pose tracking

Mocap-driven full-body control for pose tracking

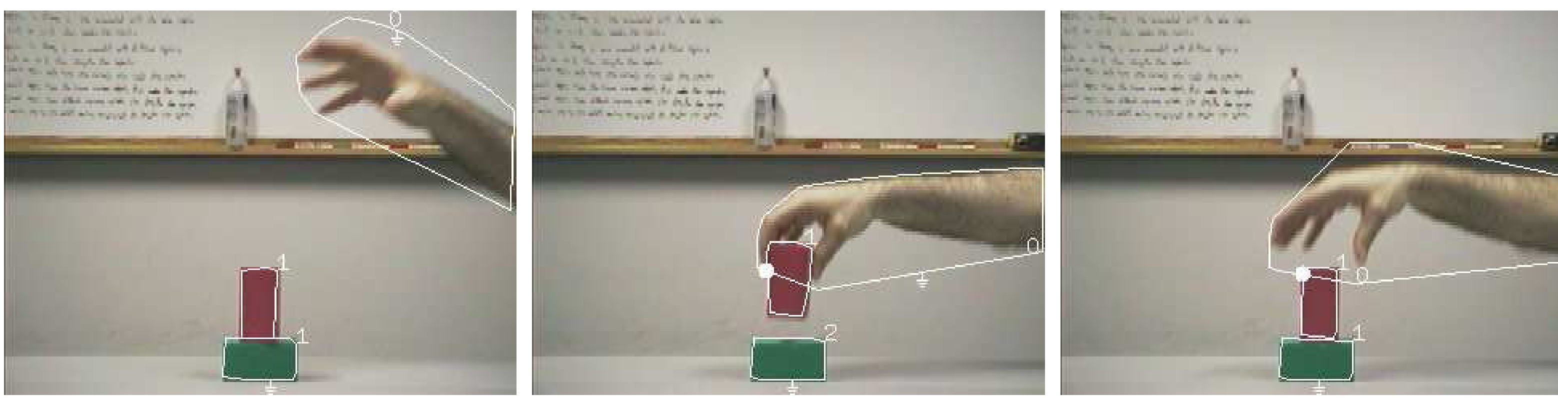


[*Vondrak et al., ‘08*]

Physics based models for scene dynamics

Interactions between objects in the scene can be reasoned about based on physical interactions and event logic

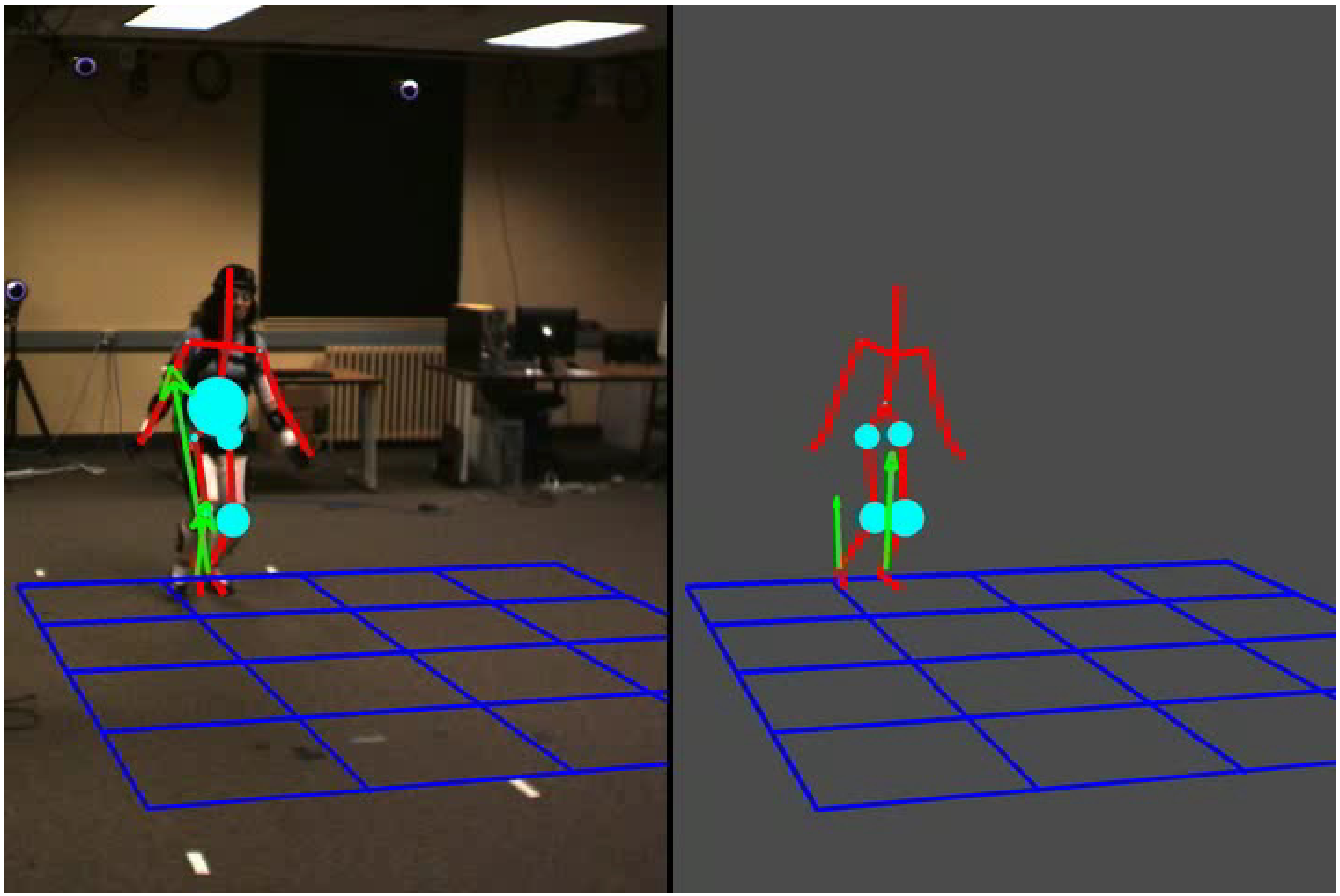
[*Mann and Jepson, 1998*]



[*Siskind, 2001*]

Estimating contact dynamics

Motion (through physical interaction) also carries information about the existence, location and compliance of the surfaces



[*Brubaker et al., ’09*]

Where do we go from here?

Code, slides and technical report available at

http://www.cs.toronto.edu/~ls/iccv2009tutorial