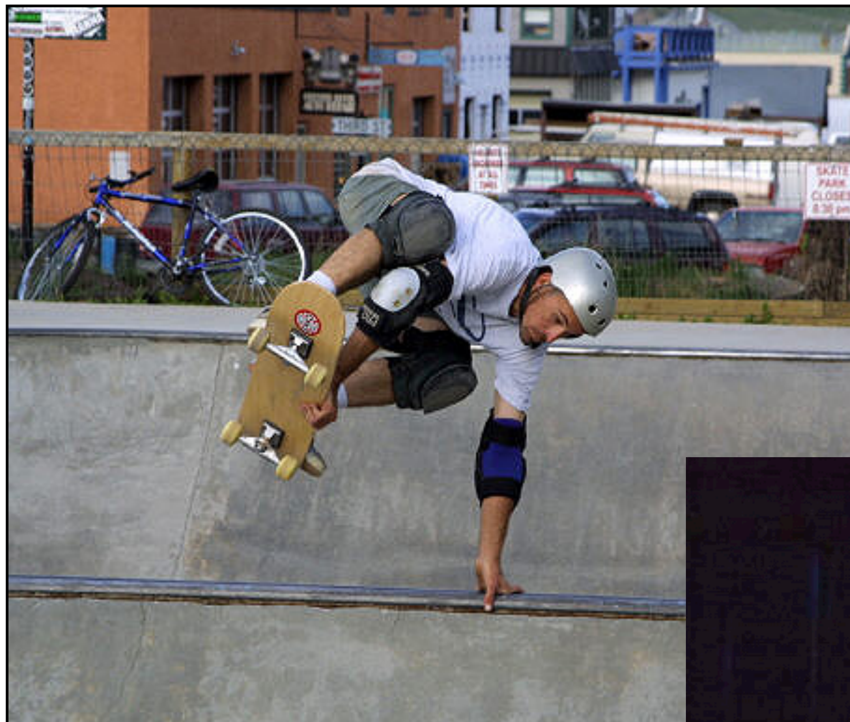
Human Motion Analysis

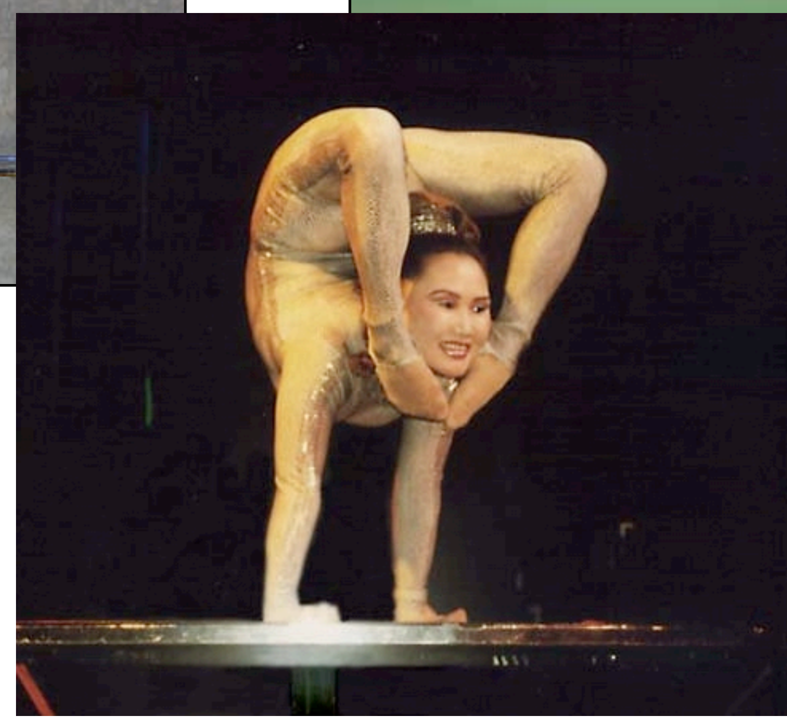
David Fleet
University of Toronto

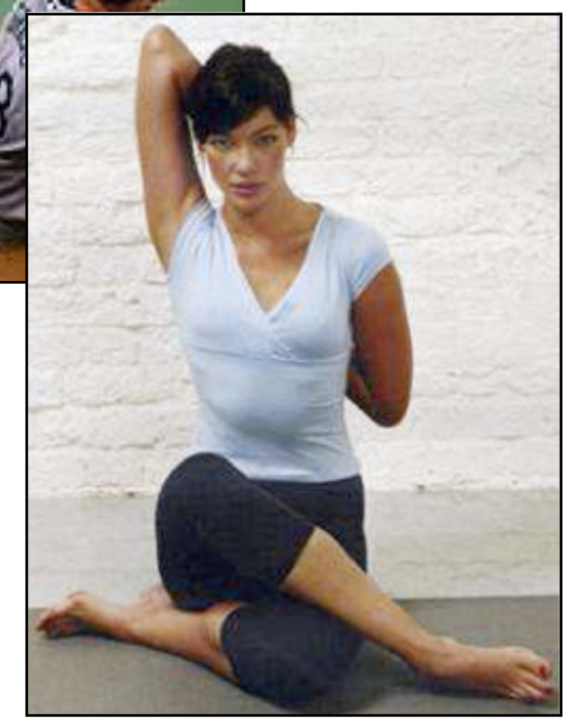
CIFAR Summer School, 2010

Looking at People

Challenges: Complex pose / motion







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

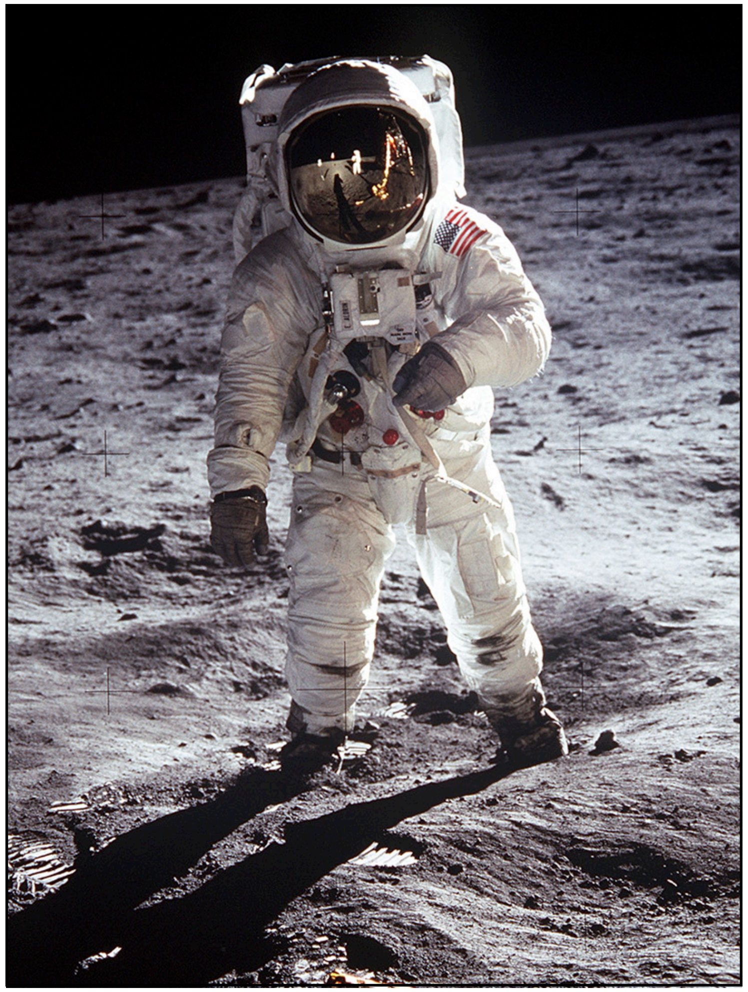
Challenges: Complex movements

People move in complex ways, often communicating with subtle gestures

Challenges: Complex movements & interactions

Interactions are fundamental

Challenges: Appearance, size and shape

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

Challenges: Appearance variability

Image appearance changes dramatically over time
due to non-rigidity of body and clothing and lighting.

Challenges: Appearance variability

Image appearance changes dramatically over time due to non-rigidity of body and clothing and lighting.

Challenges: Context dependence

Perceived scene context influences object recognition.

[Courtesy of Antonio Torralba]

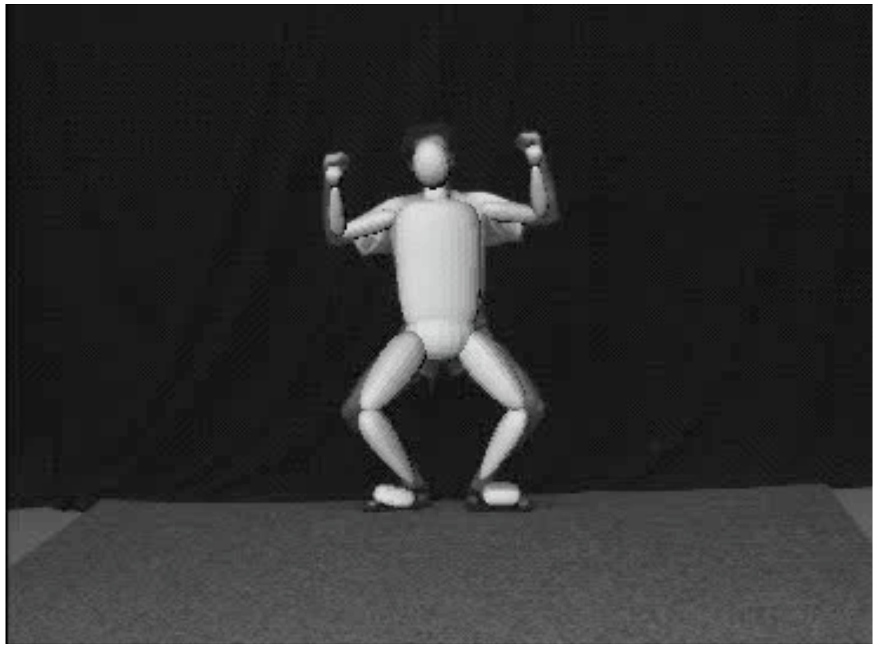
Challenges: Noisy and missing measurements

Ambiguities in pose are commonplace, due to

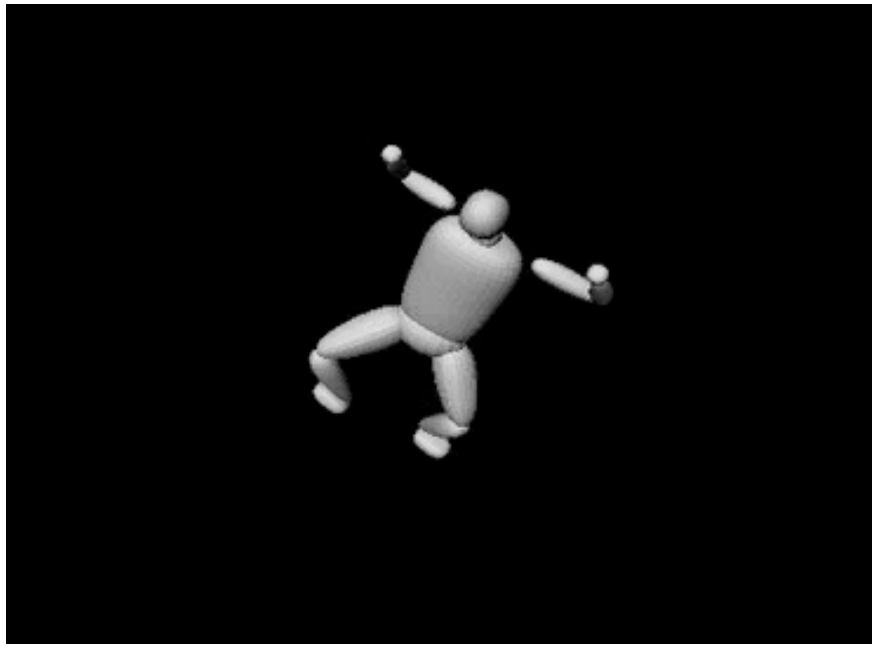
- background clutter
- apparent similarity of parts
- occlusions
- loose clothing
- ...

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

Estimation

Generative

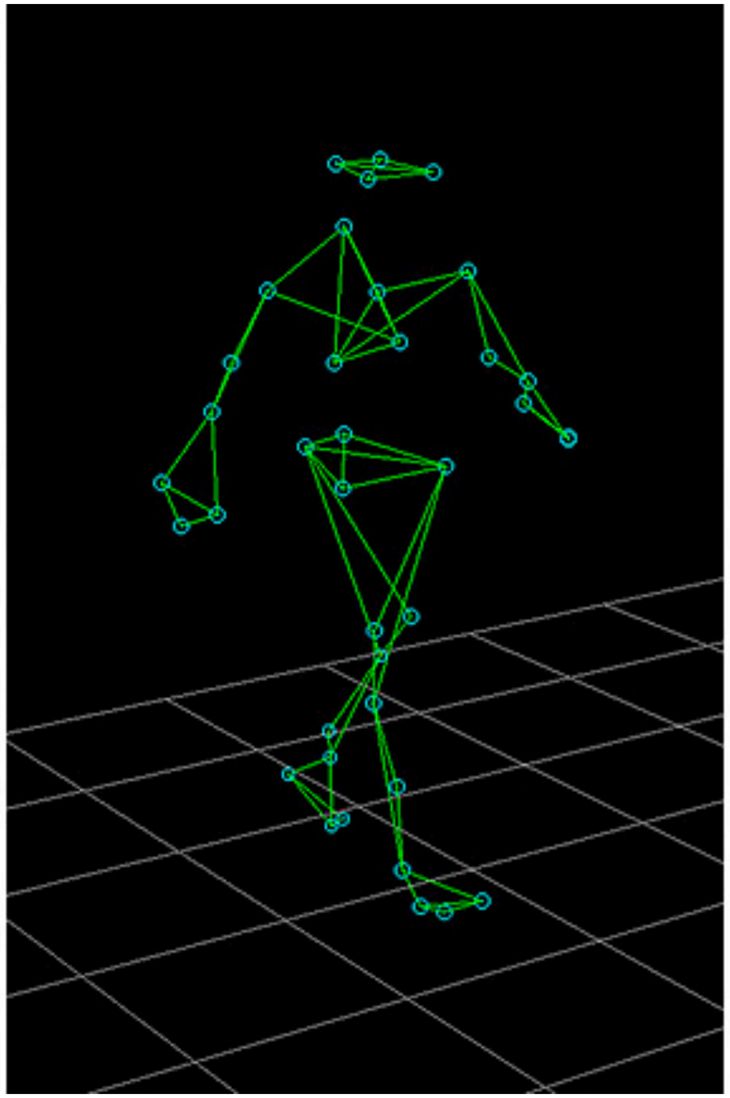
likelihood prior

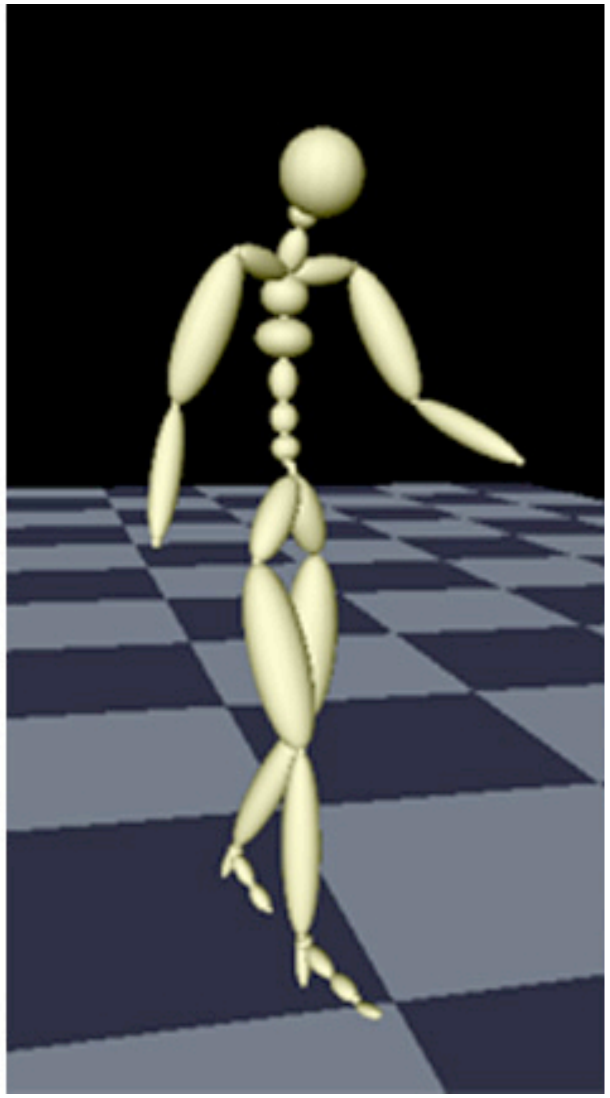
$$p(motion \,|\, video) \;=\; \frac{p(video \,|\, motion)\, p(motion)}{p(video)}$$

Discriminative

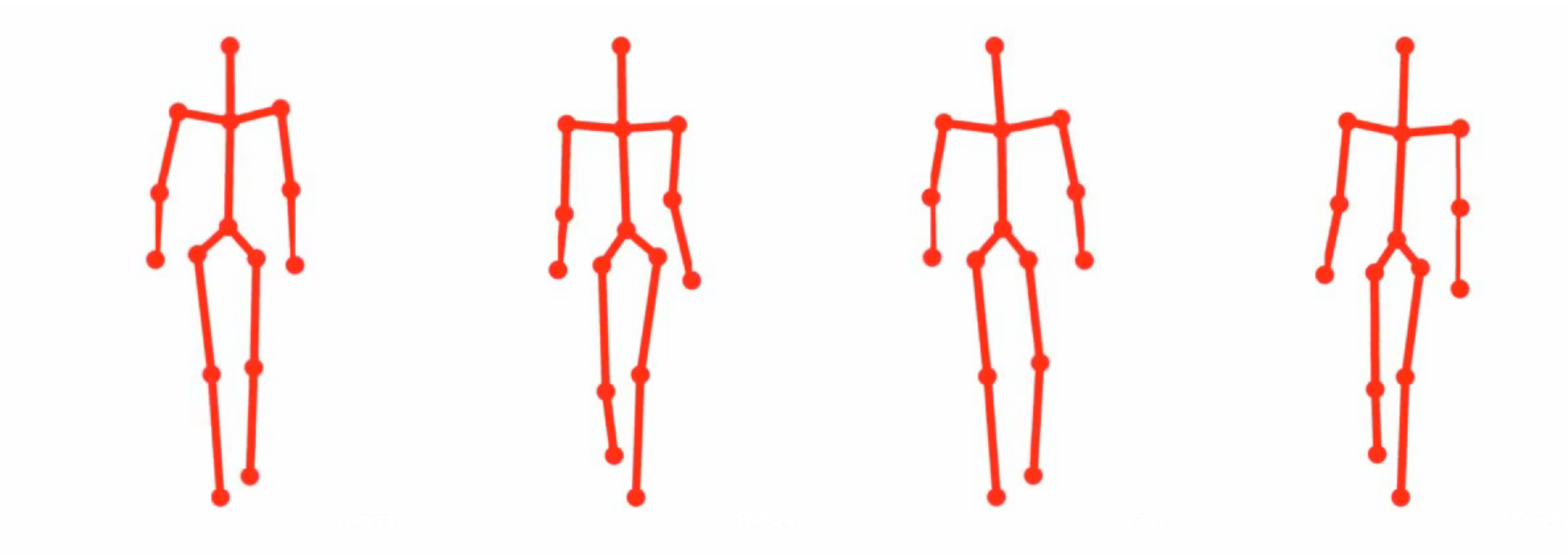
$$\begin{aligned} 3D\ pose \;&=\; E_{p(pose \,|\, image)}[\, f(pose)\,] \\ &\approx\; h(image\ measurements) \end{aligned}$$

Mocap training data





Mocap training data



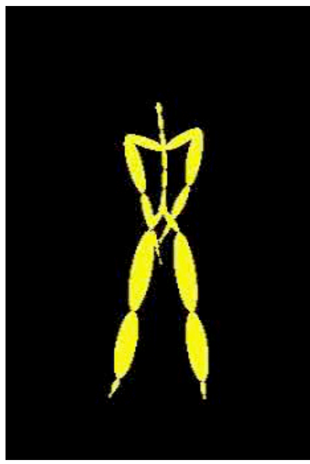
Outline

- Introduction
- Kinematic Models
- Discriminative Pose Estimation
- Physics-Based Models

Kinematic Motion Models

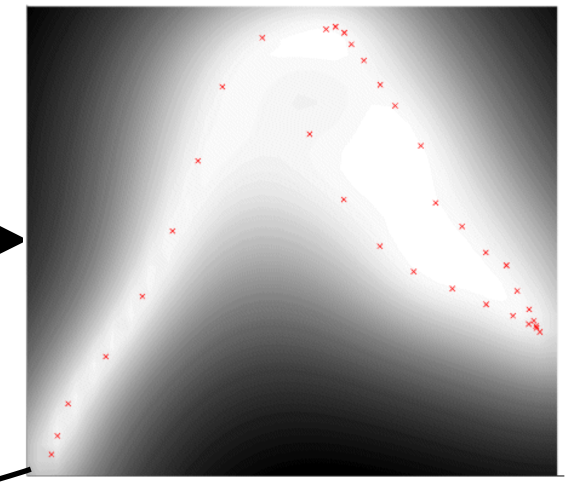
Kinematic density models

Off-line Learning



Mocap Data

Learning



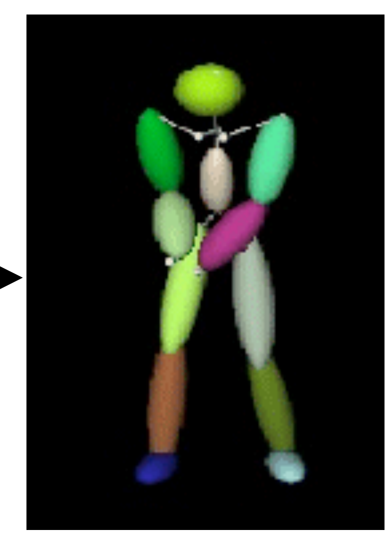
Motion/Pose Model

Prior

On-line Tracking

Video

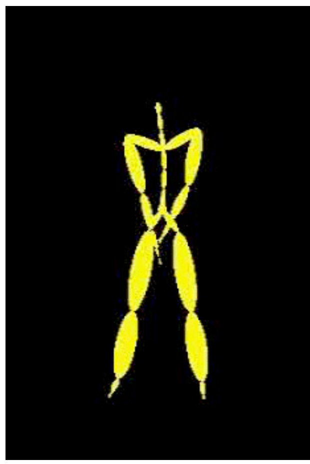
Tracking



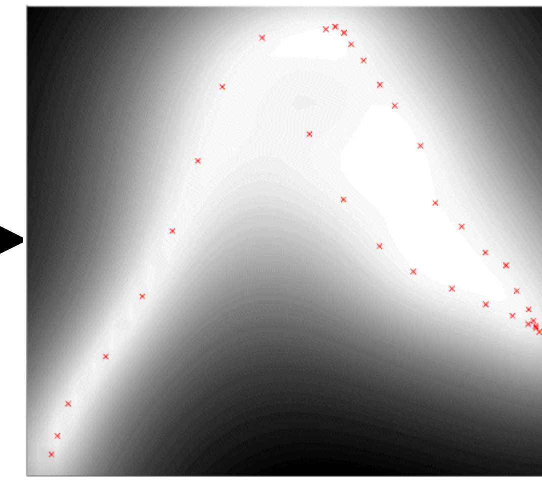
Pose

Model-based pose tracking

Off-line Learning



Learning

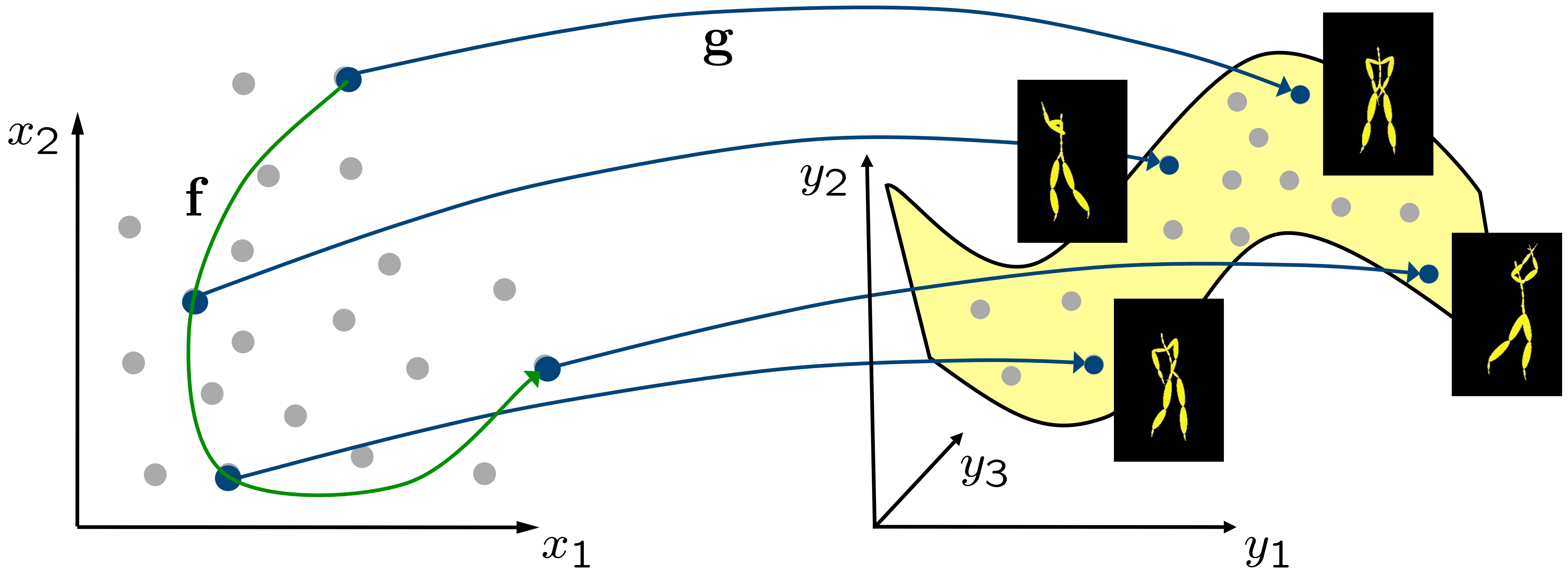


Mocap Data

Motion/Pose Model

Problem: Human pose data are high-dimensional, and difficult to obtain, so over-fitting and generalization are major issues.

Latent variable models



Low-dim. latent space $(\mathbf{x})$

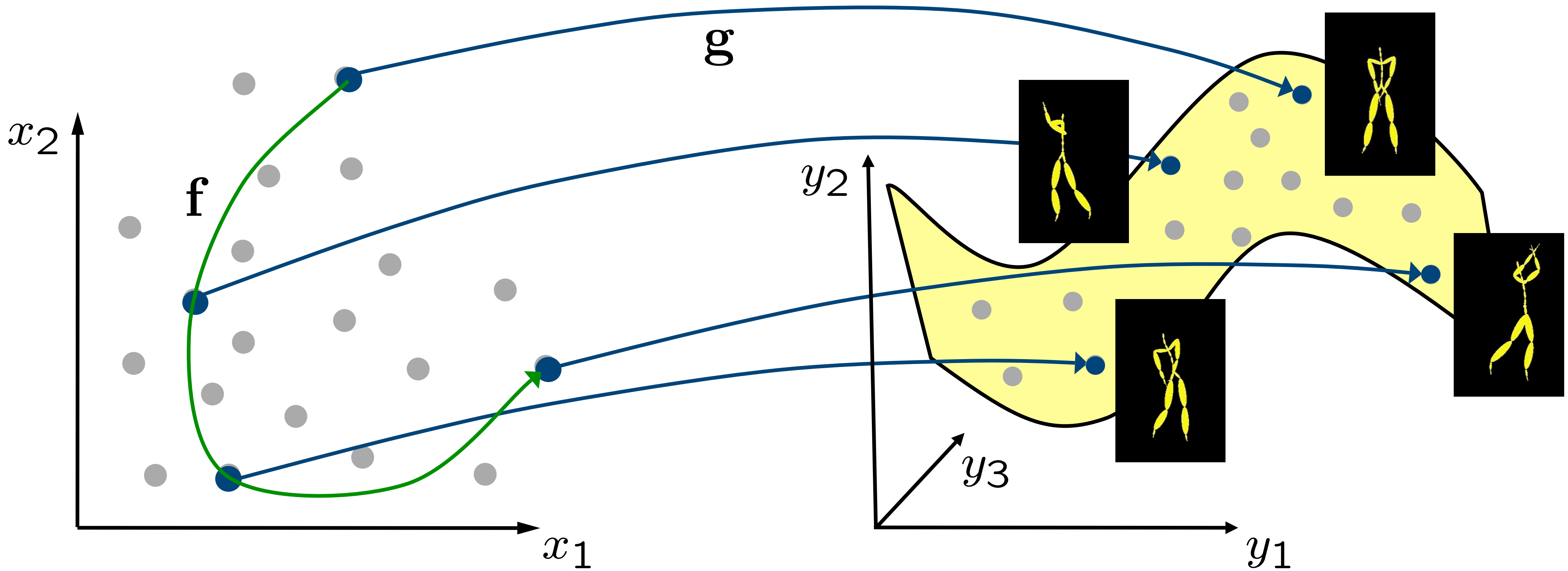
Joint angle pose space $(\mathbf{y})$

Mapping from latent positions to poses, $\mathbf{g}$

Latent dynamical model, $\mathbf{f}$

Density function over pose and motion (latent trajectories)

Latent variable models



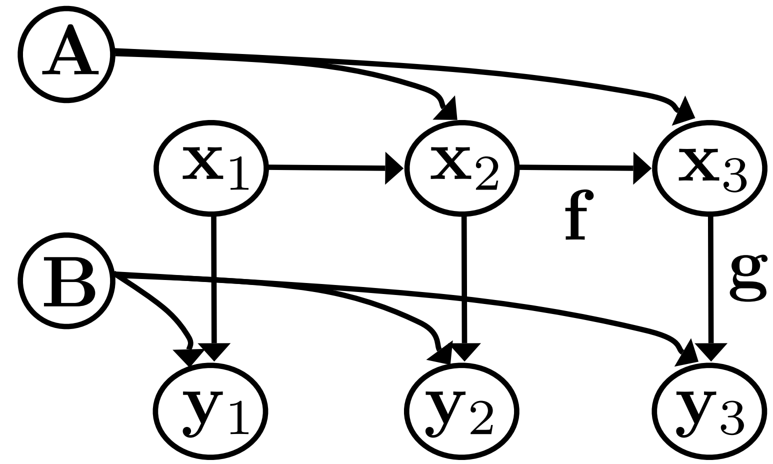
Low-dim. latent space $(\mathbf{x})$

Joint angle pose space $(\mathbf{y})$

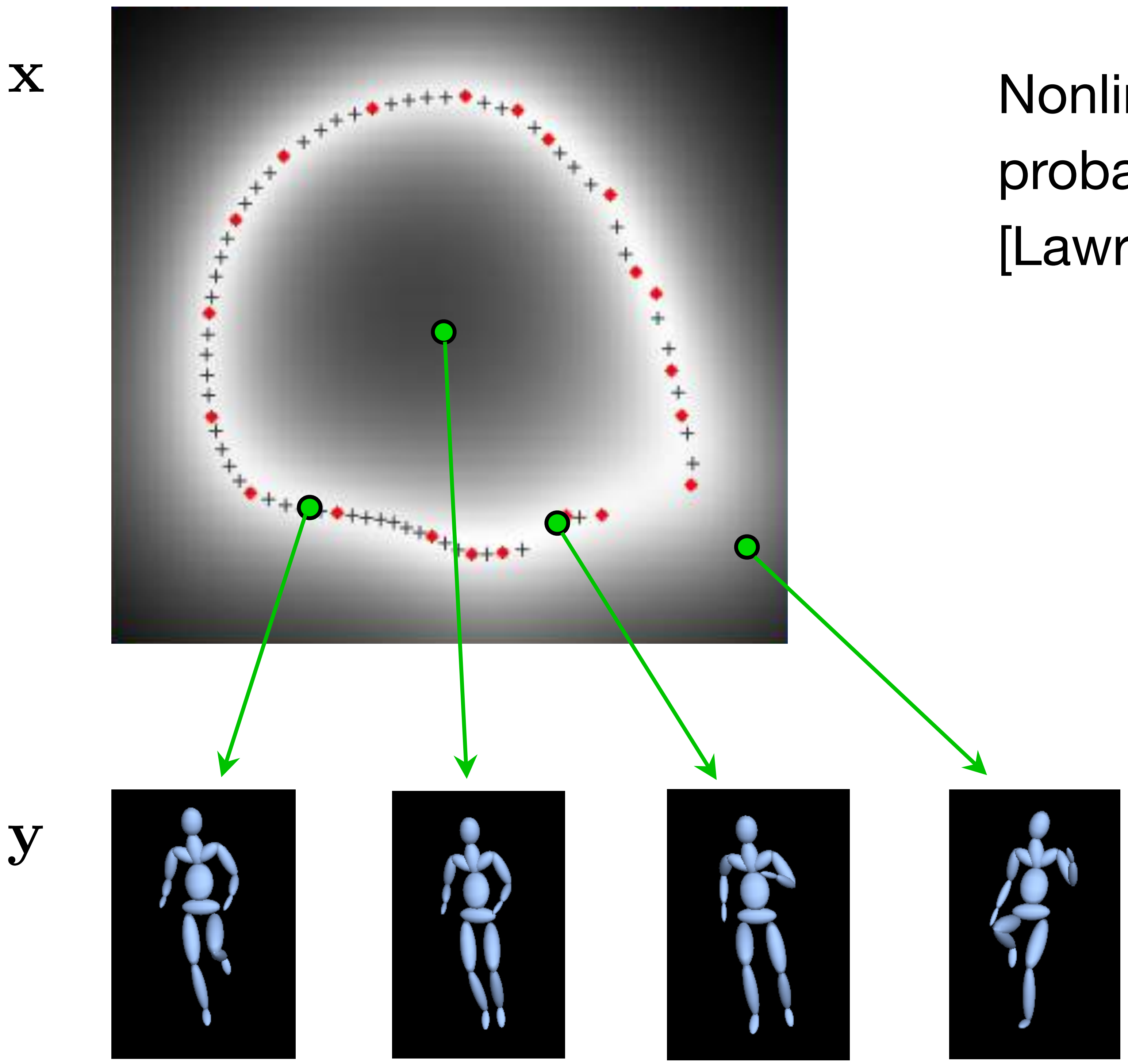
Linear dynamical systems:

$$\mathbf{x}_t = \mathbf{A}\,\mathbf{x}_{t-1} + \mathbf{n}_{x,t}$$

$$\mathbf{y}_t = \mathbf{B}\,\mathbf{x}_t + \mathbf{n}_{y,t}$$

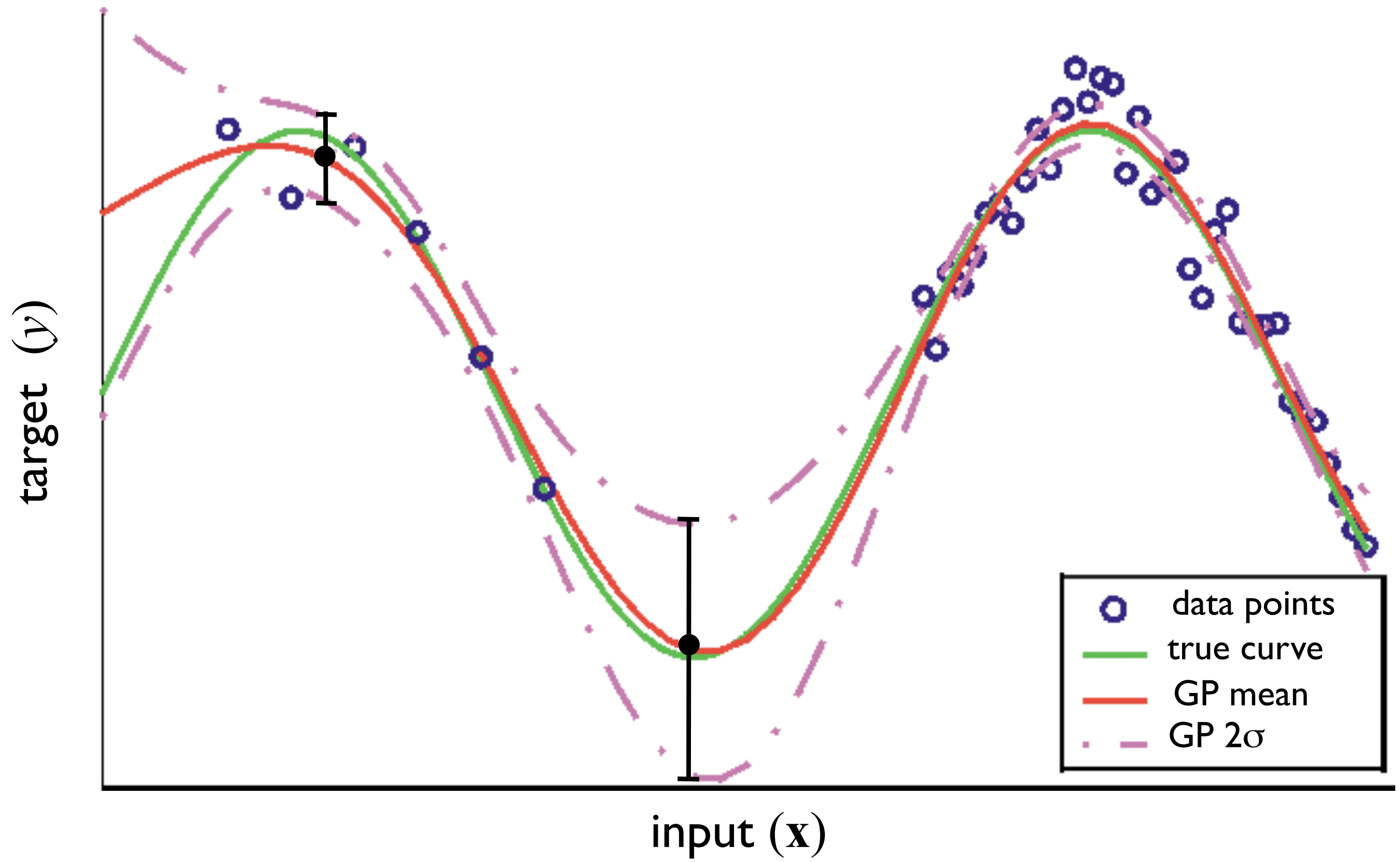


Gaussian Process Latent Variable Model



Nonlinear generalization of probabilistic PCA [Lawrence `05].

Gaussian Process



Model averaging (marginalization of the parameters) helps to avoid problems due to over-fitting and under-fitting with small data sets.

Gaussian Process

Output y is modeled as a function of input $\mathbf{x}$:

$$y \;=\; g(\mathbf{x}) \;=\; \sum_j w_j \, \phi_j(\mathbf{x}) \;=\; \mathbf{w}^T \boldsymbol{\Phi}(\mathbf{x})$$

If $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then $y \,|\, \mathbf{x}$ is zero-mean Gaussian with covariance

$$k(\mathbf{x}, \mathbf{x}') \;\equiv\; E[\, yy' \,] \;=\; \boldsymbol{\Phi}(\mathbf{x})^T \boldsymbol{\Phi}(\mathbf{x}')$$

A Gaussian process is fully specified by a mean function and a covariance function $k(\mathbf{x}, \mathbf{x}')$ and its hyper-parameters; E.g.,

Linear: $k(\mathbf{x}, \mathbf{x}') \;=\; \theta \, \mathbf{x}^T \mathbf{x}'$

RBF: $k(\mathbf{x}, \mathbf{x}') \;=\; \theta \, \exp(-\frac{\gamma}{2} \|\mathbf{x} - \mathbf{x}'\|^2)$

Gaussian Process Latent Variable Model (GPLVM)

Joint likelihood of vector-valued data $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_N]^T, \ \mathbf{y}_n \in \mathcal{R}^D$, given the latent positions $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N]^T$:

$$p(\mathbf{Y} \,|\, \mathbf{X}) \ = \ \prod_{d=1}^{D} \mathcal{N}(\mathbf{Y}_d; \, \mathbf{0}, \, \mathbf{K})$$

where $\mathbf{Y}_d$ denotes the d^{th} dimension of the training data, and the kernel matrix has elements $(\mathbf{K})_{ij} = k(\mathbf{x}_i, \, \mathbf{x}_j)$ and is shared by all data dimensions.

Learning: Maximize log likelihood of the data to find latent positions and kernel hyper-parameters, given an initial guess (e.g., use PCA).

Conditional (predictive) distribution

Given a model $\mathcal{M} = (\mathbf{Y}, \mathbf{X})$, the distribution over the data $\mathbf{y}_*$ conditioned on a latent position, $\mathbf{x}_*$, is Gaussian:

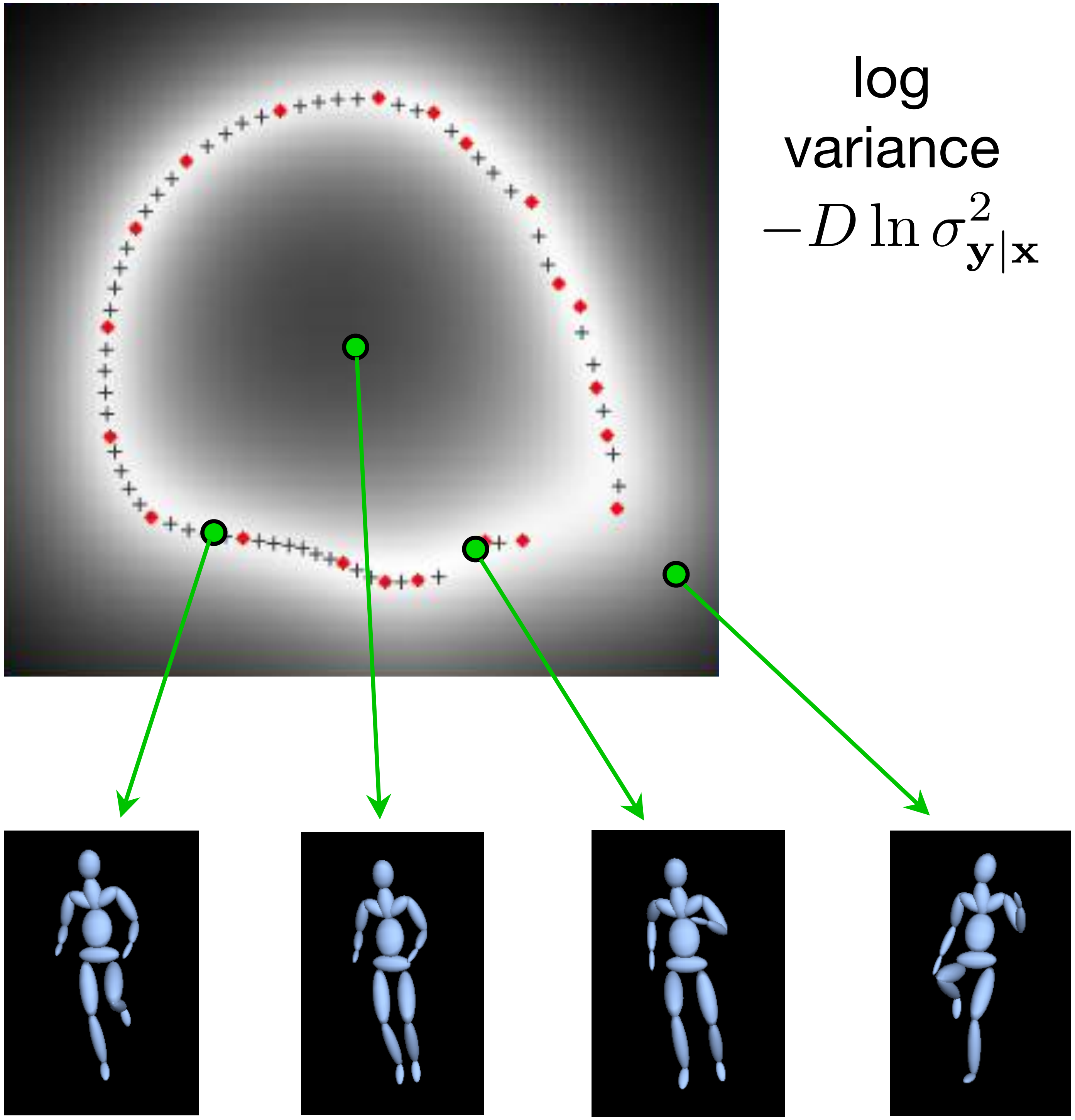
$$\boxed{\mathbf{y}_* \,|\, \mathbf{x}_*, \mathcal{M} \;\sim\; \mathcal{N}(\mathbf{m}(\mathbf{x}_*),\, \sigma^2(\mathbf{x}_*)\,\mathbf{I}_D\,)}$$

where

$$\begin{aligned}
\mathbf{m}(\mathbf{x}_*) &= \mathbf{Y}\,\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}_*) \\
\sigma^2(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)^T\,\mathbf{K}^{-1}\,\mathbf{k}(\mathbf{x}_*) \\
\mathbf{k}(\mathbf{x}_*) &= [k(\mathbf{x}_*, \mathbf{x}_1), ..., k(\mathbf{x}_*, \mathbf{x}_N)]^T
\end{aligned}$$

Gaussian Process Latent Variable Model

$\mathbf{x}$



log variance $-D \ln \sigma^2_{\mathbf{y}|\mathbf{x}}$

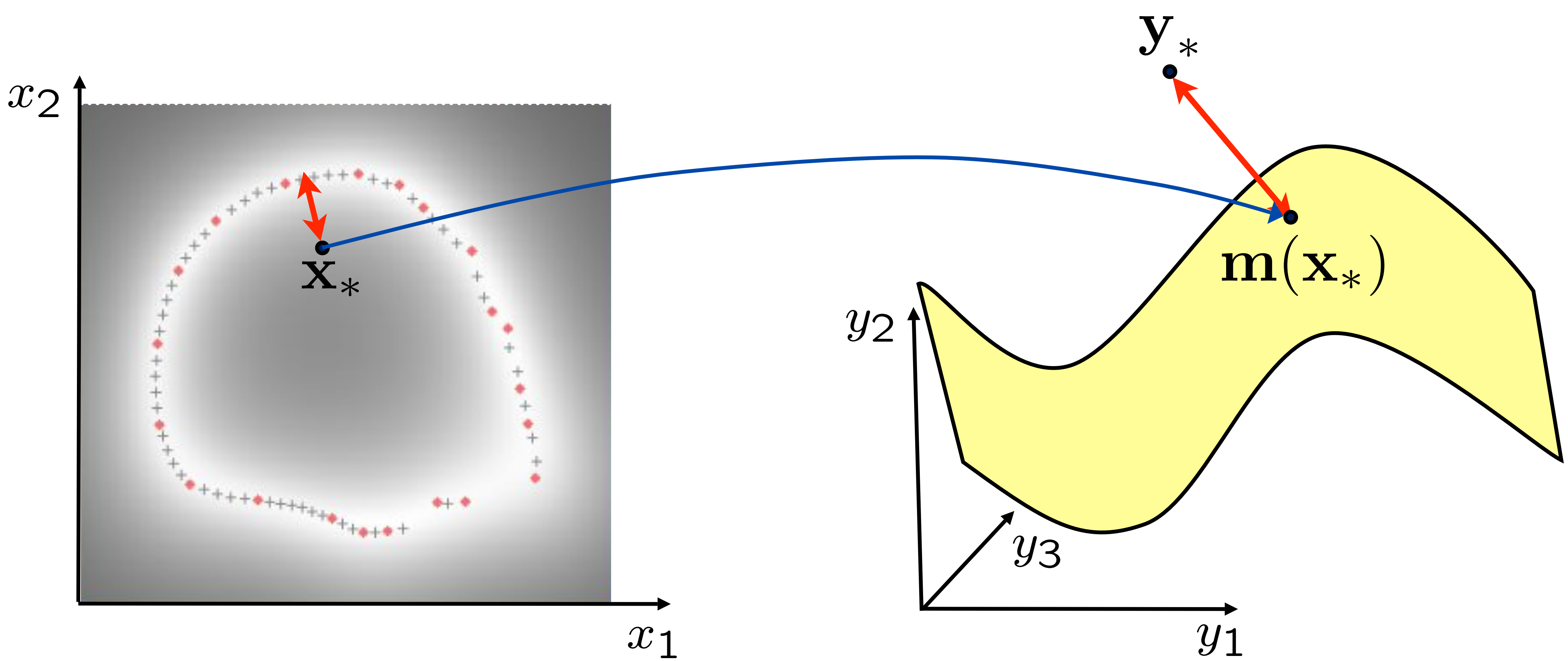
$\mathbf{y}$

mean pose $\mathbf{m}(\mathbf{x})$

Conditional (predictive) distribution

The negative log density for a new pose, given $\mathcal{M} \equiv (\mathbf{Y}, \mathbf{X})$, has a simple form:

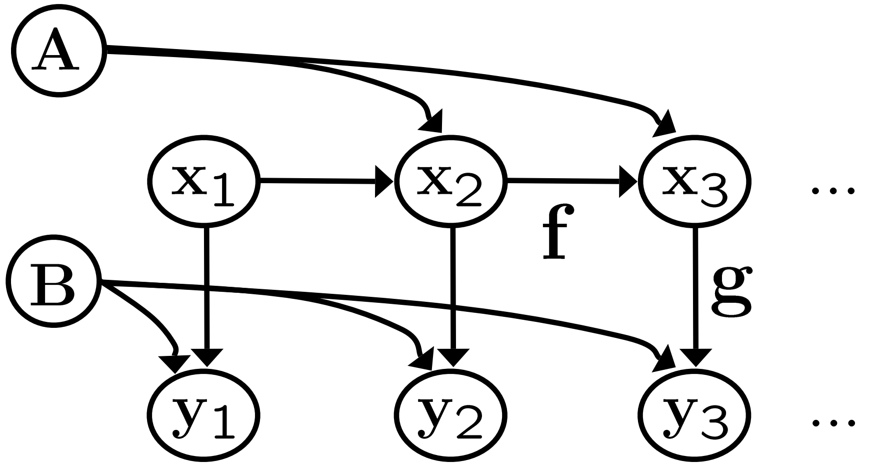
$$L(\mathbf{x}_*, \mathbf{y}_*; \mathcal{M}) = \frac{\|\mathbf{y}_* - \mathbf{m}(\mathbf{x}_*)\|^2}{2\sigma^2(\mathbf{x}_*)} + \frac{D}{2} \ln \sigma^2(\mathbf{x}_*)$$



Latent Space ($\mathbf{X}$) Pose Space ($\mathbf{Y}$)

Gaussian Process Dynamical Model (GPDM)

Latent dynamical model *[Wang et al 05]*:



$$\begin{aligned} \mathbf{x}_t &= \mathbf{f}(\mathbf{x}_{t-1}; \mathbf{A}) + \mathbf{n}_{x,t} \\ \mathbf{y}_t &= \mathbf{g}(\mathbf{x}_t; \mathbf{B}) + \mathbf{n}_{y,t} \end{aligned}$$

Assume IID Gaussian noise, and

$$\begin{aligned} \mathbf{f}(\mathbf{x}; \mathbf{A}) &= \sum_i \mathbf{a}_i \, \phi_i(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}; \mathbf{B}) &= \sum_j \mathbf{b}_j \, \psi_j(\mathbf{x}) \end{aligned}$$

with Gaussian priors on $\mathbf{A} \equiv \{\mathbf{a}_i\}$ and $\mathbf{B} \equiv \{\mathbf{b}_j\}$

Marginalize out $\{\mathbf{a}_i, \mathbf{b}_j\}$, and then optimize the latent positions, $\{\mathbf{x}, ..., \mathbf{x}_N\}$, to simultaneously minimize pose reconstruction error and prediction error on training sequence $\{\mathbf{y}, ..., \mathbf{y}_N\}$.

Reconstruction

The data likelihood for the reconstruction mapping, given centered inputs $\mathbf{Y} \equiv [\mathbf{y}, ..., \mathbf{y}_N]^T, \ \mathbf{y}_n \in \mathcal{R}^D$ has the form:

$$\boxed{p(\mathbf{Y} \,|\, \mathbf{X}, \vec{\beta}, \mathbf{W}) \;=\; \frac{|\mathbf{W}|^N}{\sqrt{(2\pi)^{ND}|\mathbf{K}_Y|^D}} \exp\left(-\frac{1}{2} tr(\mathbf{K}_Y^{-1}\mathbf{Y}\mathbf{W}^2\mathbf{Y}^T)\right)}$$

where

$\mathbf{K}_Y$ is a kernel matrix shared across pose outputs, with entries $(\mathbf{K}_Y)_{ij} = k_Y(\mathbf{x}_i, \mathbf{x}_j)$ for kernel function, e.g.,

$$k_Y(\mathbf{x}, \mathbf{x}') \;=\; \beta_1 \exp\left(-\frac{\beta_2}{2}||\mathbf{x} - \mathbf{x}'||^2\right) + \beta_3^{-1}\delta_{\mathbf{x},\mathbf{x}'}$$

with hyperparameters $\vec{\beta} = \{\beta_1, \beta_2, \beta_3\}$

$\mathbf{W} \equiv \mathrm{diag}(w_1, ..., w_D)$ scales the different pose parameters

Dynamical prior

The latent dynamical process on $\mathbf{X} \equiv [\mathbf{x}, ..., \mathbf{x}_N]^T, \ \mathbf{x}_n \in \mathcal{R}^d$ has a similar form:

$$\boxed{p(\mathbf{X} \,|\, \vec{\alpha}) \quad = \quad \frac{\mathcal{N}(\mathbf{x}_1; \mathbf{0}, \mathbf{I}_d)}{\sqrt{(2\pi)^{(N-1)\,d} \, |\mathbf{K}_X|^d}} \exp\left(-\frac{1}{2} tr(\mathbf{K}_X^{-1} \hat{\mathbf{X}} \hat{\mathbf{X}}^T)\right)}$$

where

$\hat{\mathbf{X}} = [\mathbf{x}_2, ..., \mathbf{x}_N]^T$

$\mathbf{K}_X$ is a kernel matrix defined by kernel function, e.g.,

$$k_X(\mathbf{x}, \mathbf{x}') \ = \ \alpha_1 \exp\left(-\frac{\alpha_2}{2} ||\mathbf{x} - \mathbf{x}'||^2\right) + \alpha_3 \mathbf{x}^T \mathbf{x}' + \alpha_4^{-1} \delta_{\mathbf{x}'}$$

with hyperparameters $\vec{\alpha}$

Learning

GPDM posterior:

$$p(\mathbf{Y}, \mathbf{X}, \underbrace{\bar{\alpha}, \bar{\beta}}_{\text{kernel hyperparameters}}, \mathbf{W}) \;=\; p(\mathbf{Y} \,|\, \mathbf{X}, \bar{\beta}, \mathbf{W}) \; p(\mathbf{X} \,|\, \bar{\alpha}) \; p(\bar{\alpha}) \; p(\bar{\beta})$$

reconstruction likelihood

dynamics likelihood

priors

training motions

latent trajectories

To estimate the latent coordinates & kernel parameters we minimize

$$\mathcal{L} \;=\; -\ln p(\mathbf{X}, \bar{\alpha}, \bar{\beta}, \mathbf{W} \,|\, \mathbf{Y})$$

with respect to $\mathbf{X}$, $\bar{\alpha}$, $\bar{\beta}$ and $\mathbf{W}$.

GPDM prior over new poses and motions

The model $\mathcal{M} \equiv (\mathbf{Y}, \mathbf{X}, \vec{\alpha}, \vec{\beta}, \mathbf{W})$ then provides a density function over new poses, with negative log likelihood:

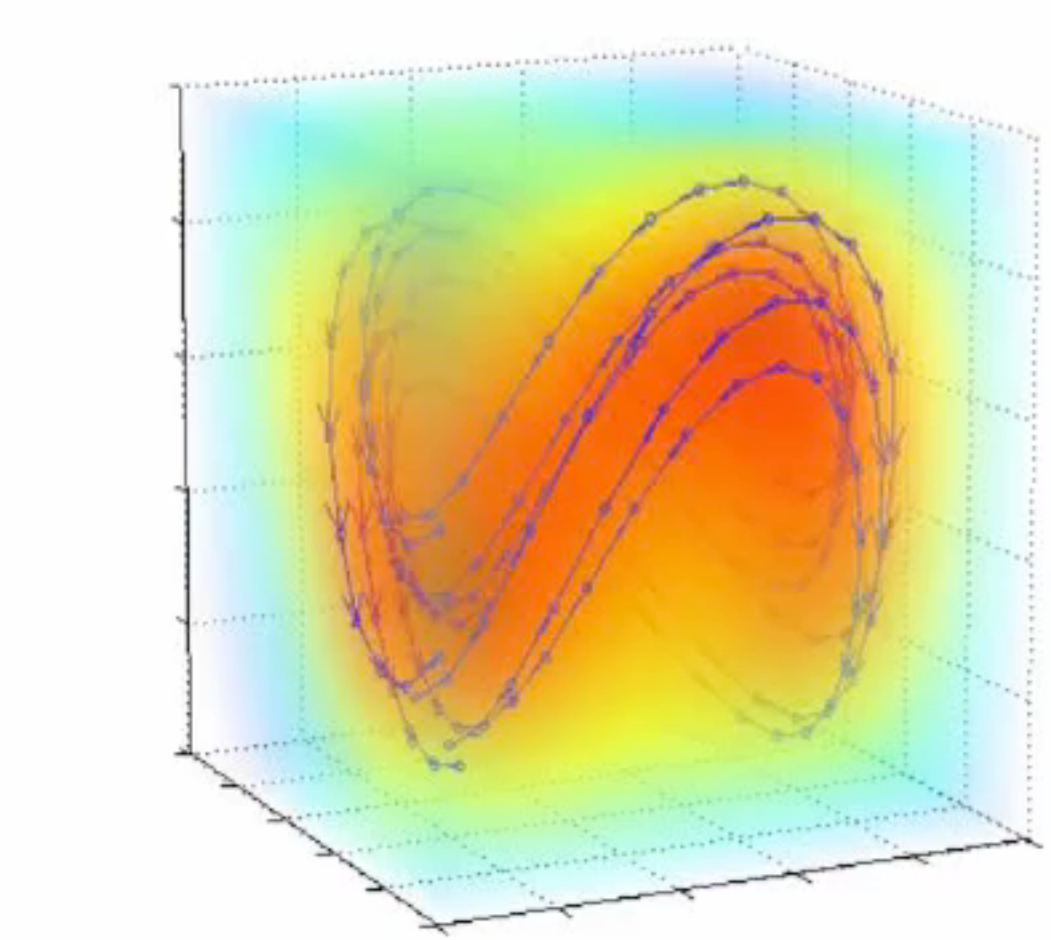
$$L(\mathbf{x}, \mathbf{y}; \mathcal{M}) \;=\; \frac{\|\mathbf{W}(\mathbf{y} - f(\mathbf{x}))\|^2}{2\sigma_Y^2(\mathbf{x})} + \frac{D}{2} \ln \sigma_Y^2(\mathbf{x})$$

and a density over latent trajectories, with negative log likelihood:

$$L_D(\bar{\mathbf{X}}; \bar{\mathbf{x}}_0, \mathcal{M}) \;=\; \frac{1}{2} tr\left(\bar{\mathbf{K}}_X^{-1} \bar{\mathbf{X}} \bar{\mathbf{X}}^T\right) + \frac{d}{2} \ln |\bar{\mathbf{K}}_X|$$

3D B-GPDM for walking

6 walking subjects,1 gait cycle each, on treadmill at same speed with a 20 DOF joint parameterization.



GPDM: log reconstruction variance $\ln \sigma^2_{\mathbf{y}} \,|\, \mathbf{x}, \mathbf{X}, \mathbf{Y}$

GPDM: mean trajectory

GPDM sample trajectories

[Urtasun et al, `06]

People tracking with GPDM

Image Observations: $\mathbf{I}_{1:t} \equiv (\mathbf{I}_1, ..., \mathbf{I}_t)$

State: $\phi_t = [\mathbf{G}_t, \ \mathbf{y}_t, \ \mathbf{x}_t]$

GPDM: $\mathcal{M}$

global pose joint angles latent coordinates

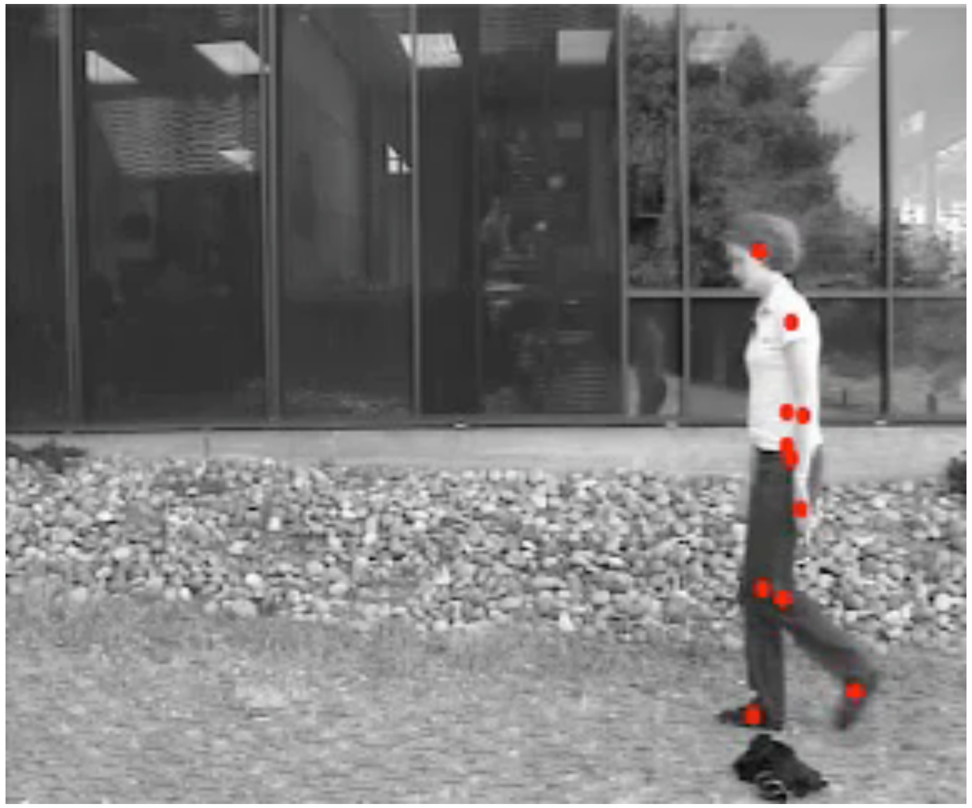
Inference: MAP estimation by gradient ascent on the posterior:

$$\underbrace{p(\phi_t \,|\, \mathbf{I}_{1:t}, \mathcal{M})}_{\text{posterior}} \;\propto\; \underbrace{p(\mathbf{I}_t \,|\, \phi_t)}_{\text{likelihood}} \; \underbrace{p(\phi_t \,|\, \mathbf{I}_{1:t-1}, \mathcal{M})}_{\text{prediction}}$$

Temporal predictions for the global DOFs based on a damped second-order Markov model.

[Urtasun et al, `06]

Measurement model

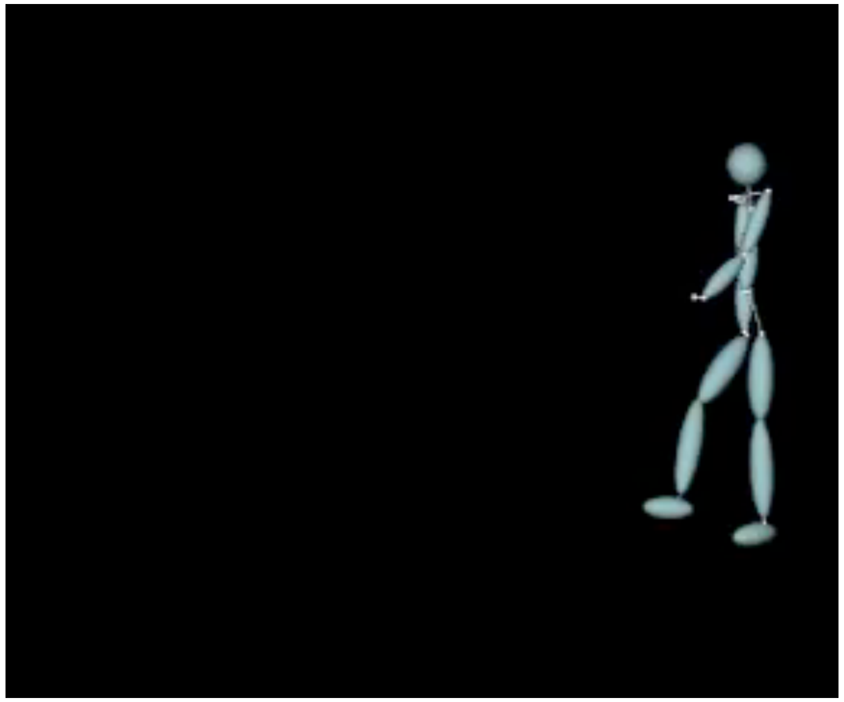


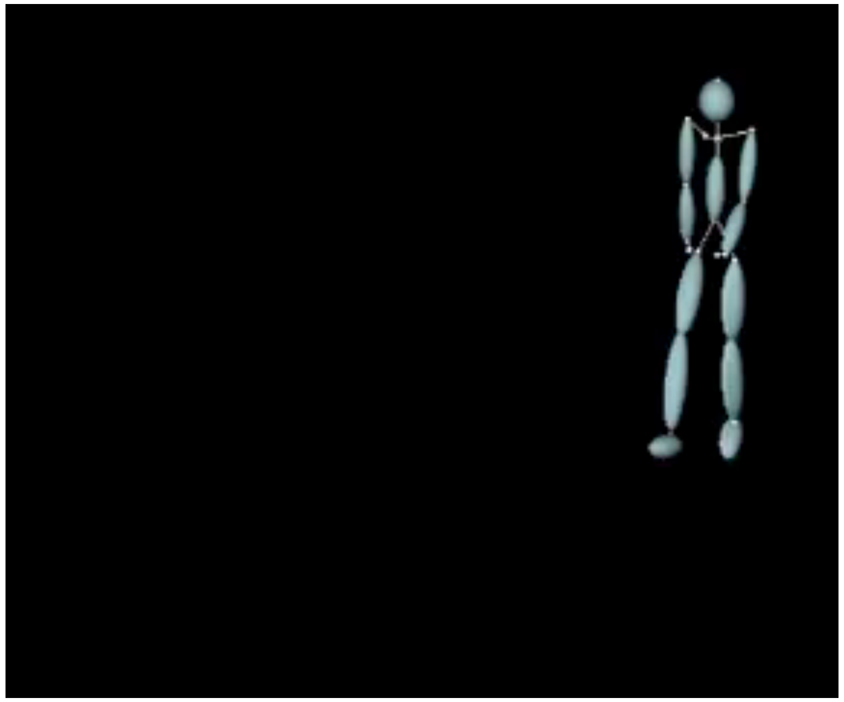
Measurements are the 2D image positions for several locations on the body, obtained with a 2D patch-based tracker *[Jepson et al 03]*.

Assume the measurements are corrupted with IID Gaussian noise.

Occlusion

3D
model
overlaid
on video

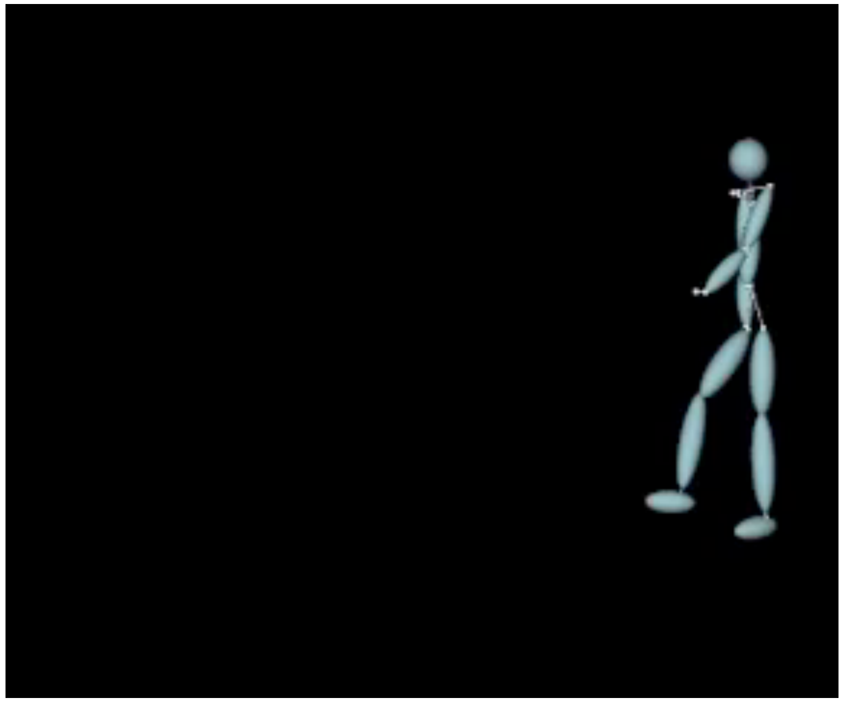


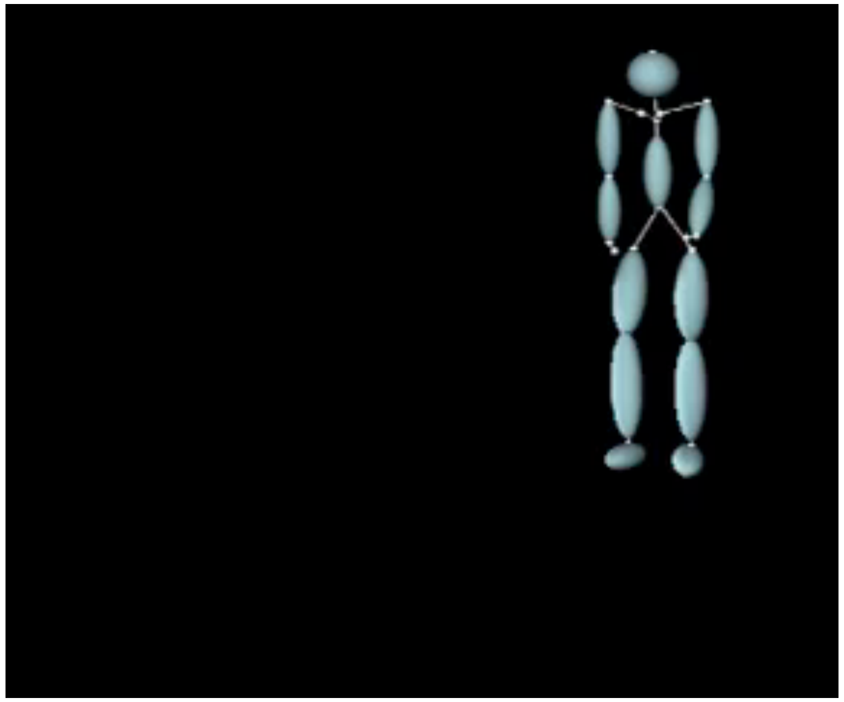


3D animated characters

Occlusion

3D
model
overlaid
on video

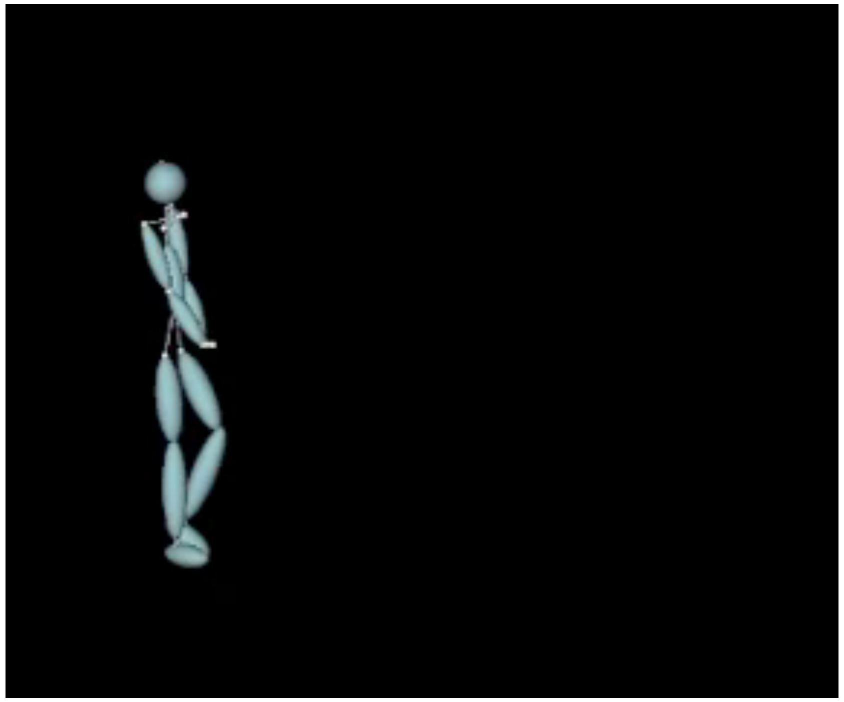


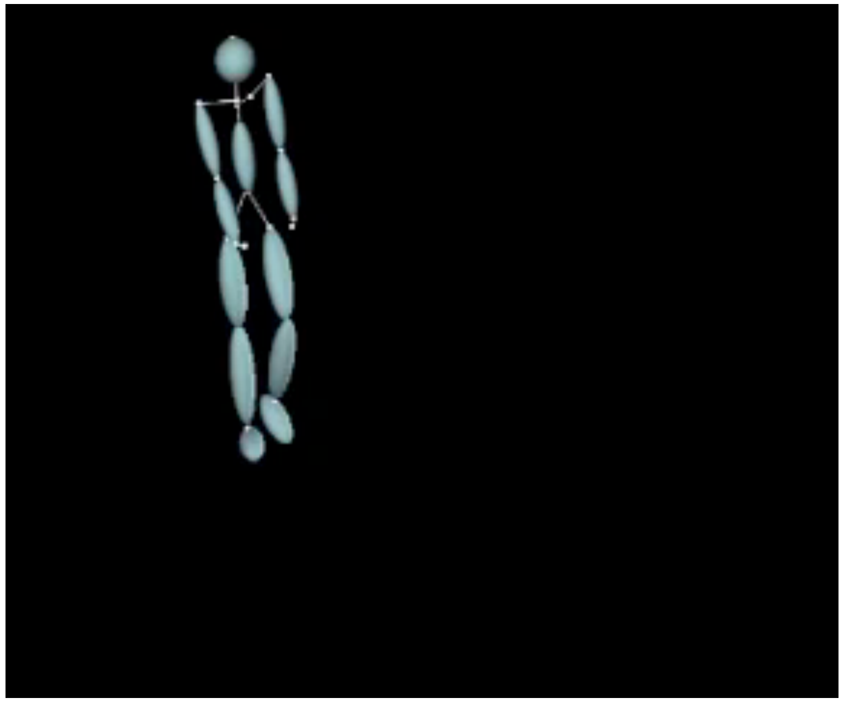


3D animated characters

Exaggerated gait

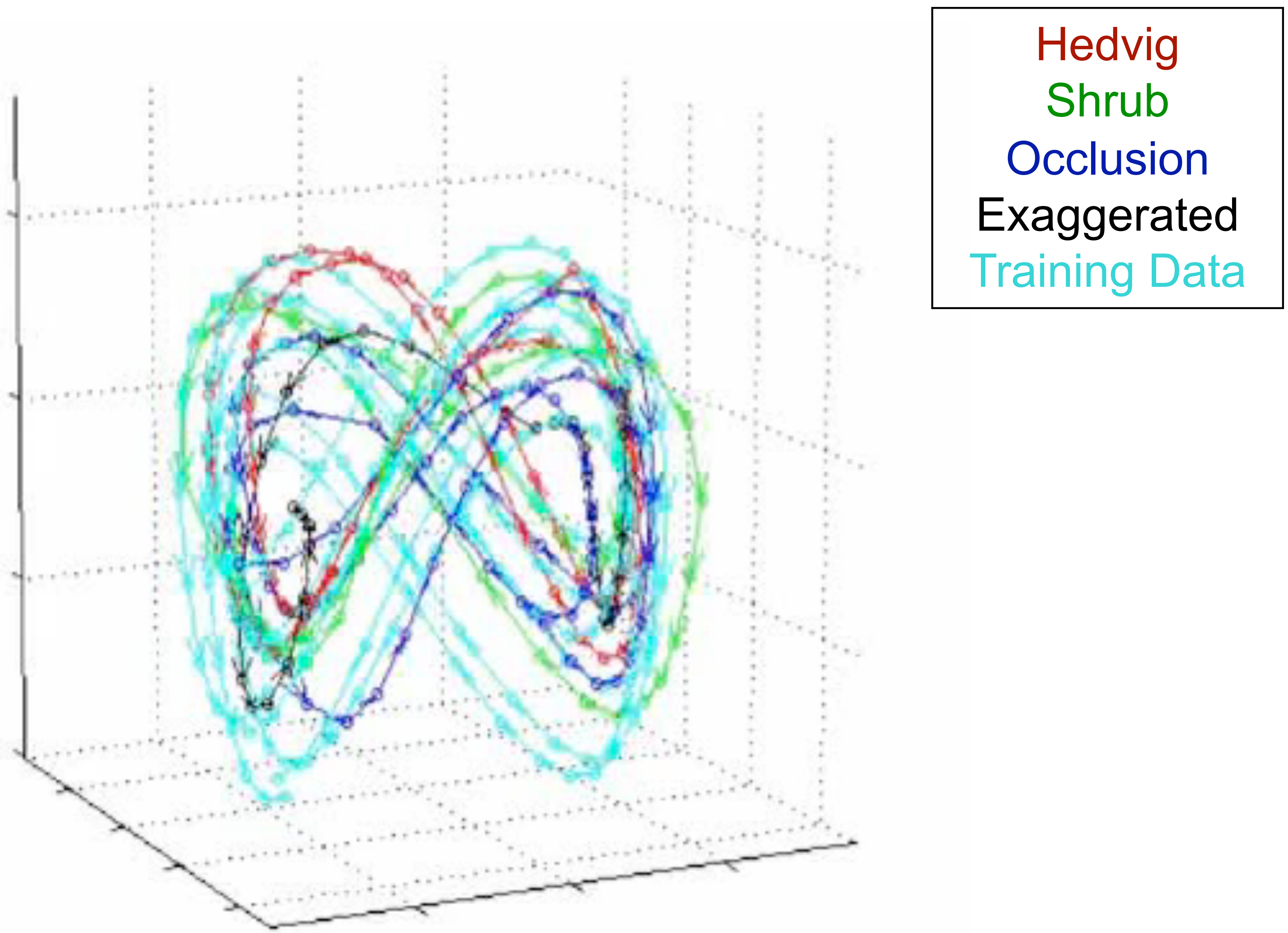
3D
model
overlaid
on video





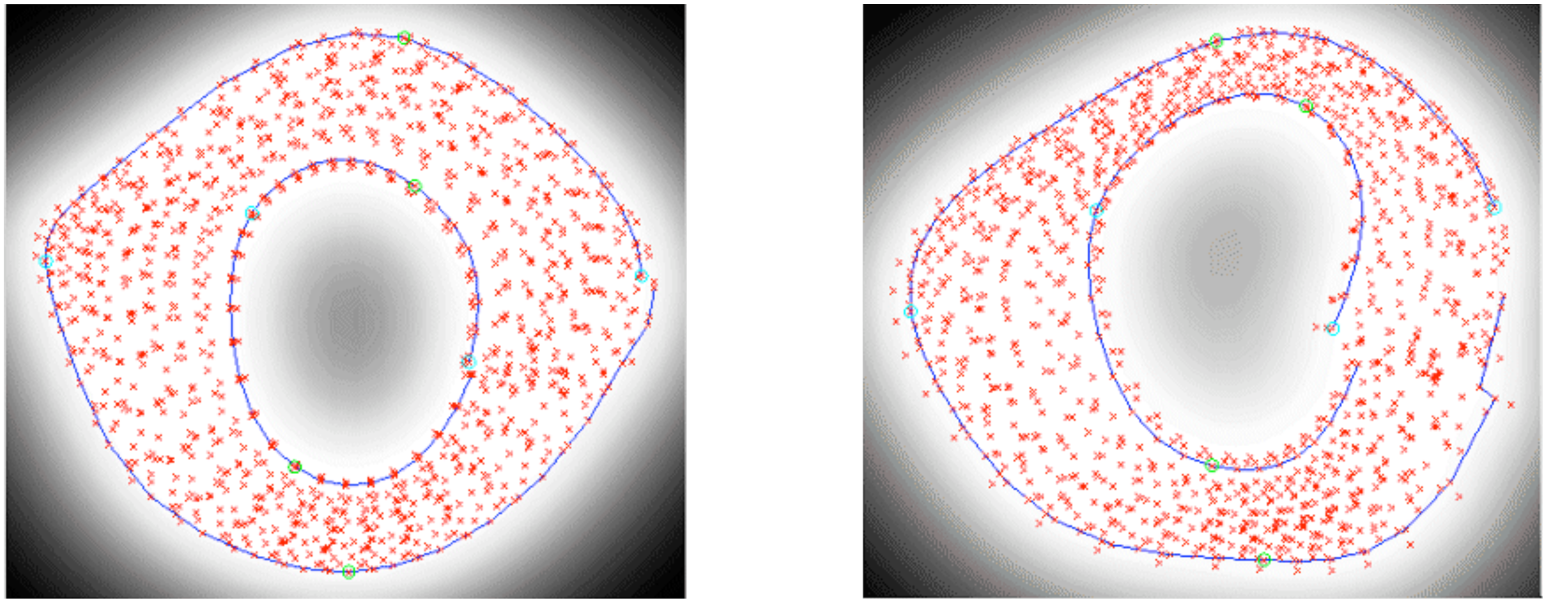
3D animated characters

Latent trajectories

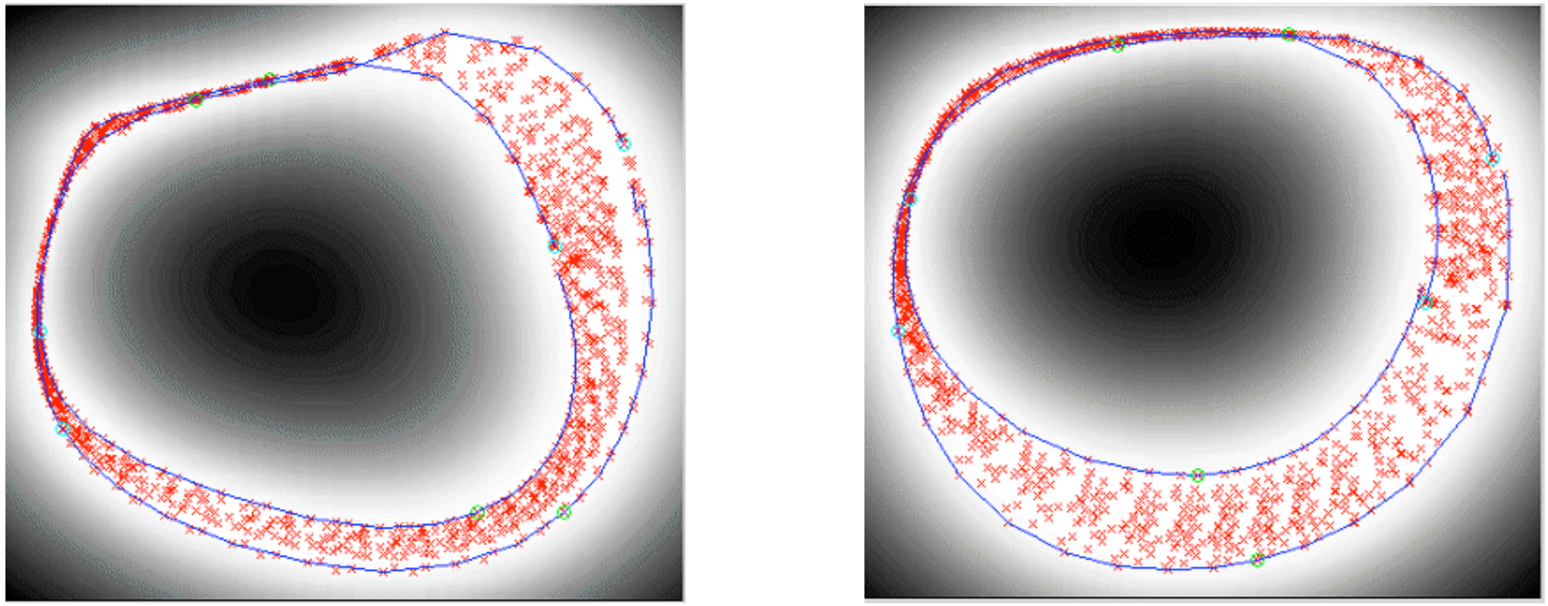


Multiple speeds and visualization of pathologies

Two subjects, four walk gait cycles at speeds 3-7 km/hr



Two subjects with a knee pathology



GPLVM / GPDM Extensions

- Multifactor GPLVM (stylistic diversity)
  [Wang et al, ICML 2008]

- Back constraints (smooth inverse mappings)
  [Lawrence and Quinonero-Candela, ICML 2006]

- Topologically-constrained GPLVM (structured latent manifolds)
  [Urtasun et al, ICML 2009]

- Hierarchical GPLVM (compositional models)
  [Moore and Lawrence, ICML 2008]

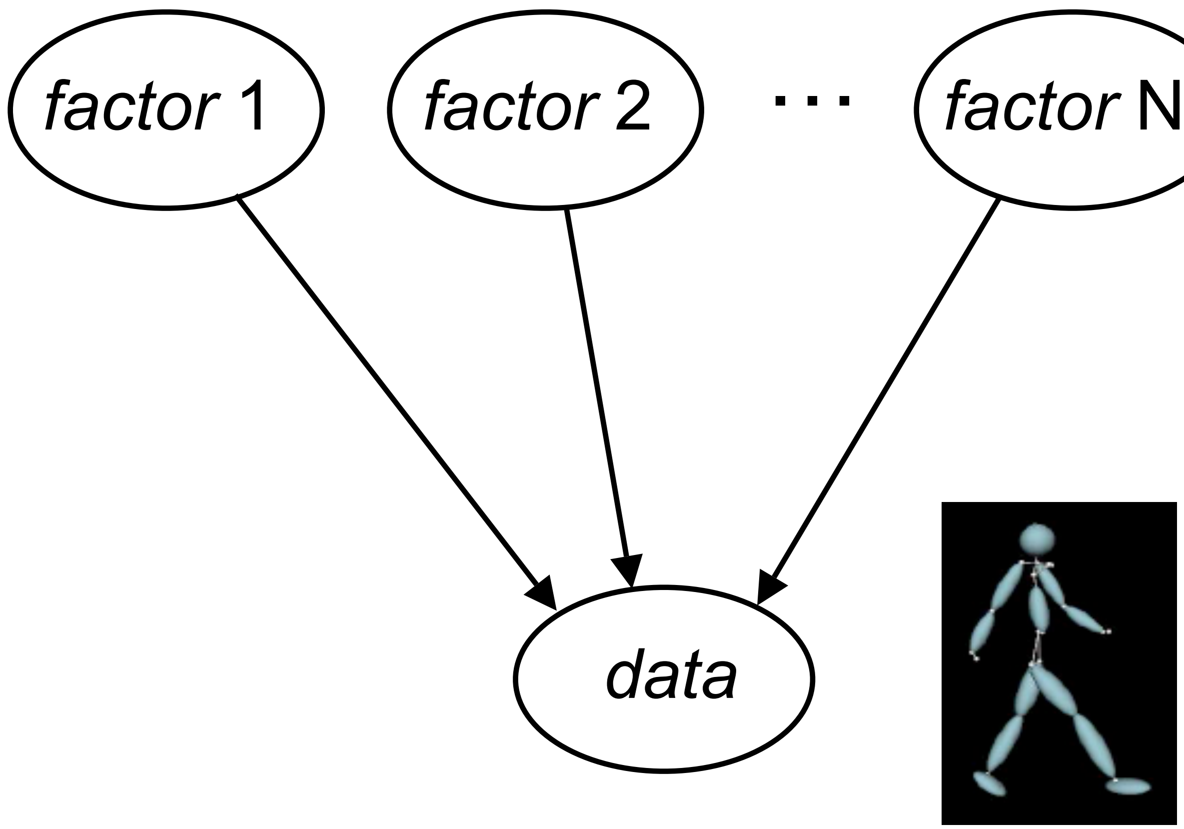
To appear (if we ever finish it): "GPs for modeling human motion" Lawrence, Fleet, Hertzmann, and Urtasun

Open problems

Modeling arbitrary motions, spanning a wide range of activities, with:

- atomic motion primitives, with suitable transitions
- part-based compositionality
- good generalization to styles and environments
- context and interactions

Multifactor LVMs



gait,
phase,
identity,
gender,
...

pose

$$y = \sum_{i,j,k,..} w_{ijk\ldots}\, a_i b_j c_k \cdots + \epsilon$$

Multilinear style-content models
[Tenenbaum and Freeman ’00;
Vasilescu and Terzopoulos ‘02]

$$y = \sum_{i,j} w_{ij}\, a_i \phi_j(\mathbf{b}) + \epsilon$$

Nonlinear basis functions
[Elgammal and Lee ‘04]

Multifactor GPLVM

Suppose y depends linearly on latent style parameters $s_1, s_2, \ldots,$ and nonlinearly on $\mathbf{x}$:

$$y = \sum_i s_i g_i(\mathbf{x}) + \epsilon = \sum_i s_i \mathbf{w}_i^T \mathbf{\Phi}(\mathbf{x}) + \epsilon$$

where $\mathbf{\Phi}(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_{N_x}(\mathbf{x})]^T$

If $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\epsilon \sim \mathcal{N}(\mathbf{0}, \beta^{-1})$, then $y \,|\, \mathbf{x}$ is zero-mean Gaussian, with covariance

$$E[yy'] = \mathbf{s}^T \mathbf{s}' \, \mathbf{\Phi}(\mathbf{x})^T \mathbf{\Phi}(\mathbf{x}') + \beta^{-1}\delta$$

where $\mathbf{s} = [s_1, \ldots, s_{N_s}]^T$

$k_s(\mathbf{s}, \mathbf{s}')$ $k_x(\mathbf{x}, \mathbf{x}')$

[Wang et al. ICML '07]

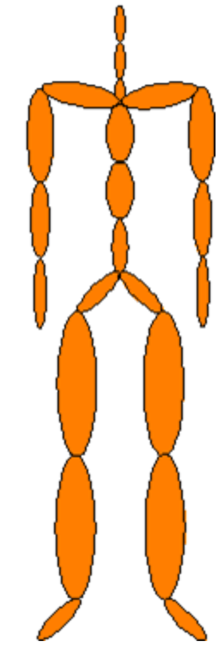
Multifactor locomotion model

Three-factor latent model with $\mathcal{X} = \{\mathbf{s}, \mathbf{g}, \mathbf{x}\}$:

$\mathbf{s}$: identity of the subject performing the motion

$\mathbf{g}$: gait of the motion (walk, run, stride)

$\mathbf{x}$: current state of motion (evolves w.r.t. time)



Covariance function:

$$k_d(\mathcal{X}, \mathcal{X}') = \theta_d \, \mathbf{s}^T \mathbf{s}' \; \mathbf{g}^T \mathbf{g}' \; e^{-\frac{\gamma}{2}||\mathbf{x}-\mathbf{x}'||^2} + \beta^{-1}\delta$$

scale of variance for each dimension

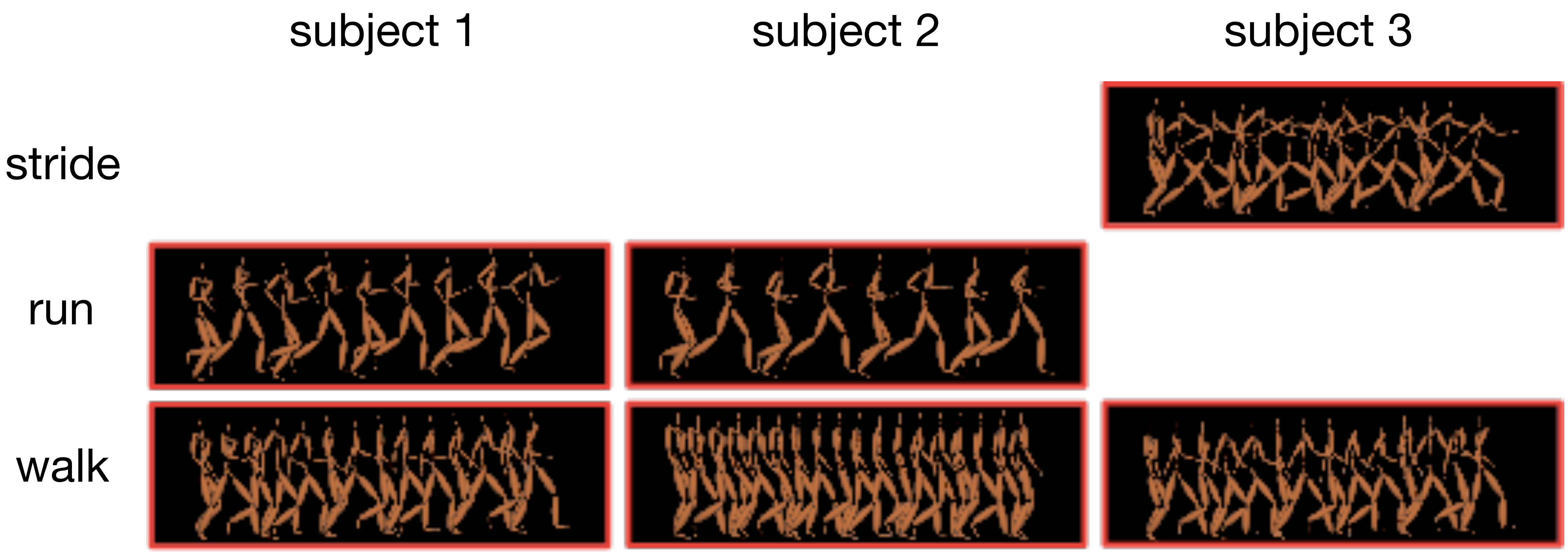
linear kernels for identity and gait (style)

RBF kernel for state (content)

additive white process noise

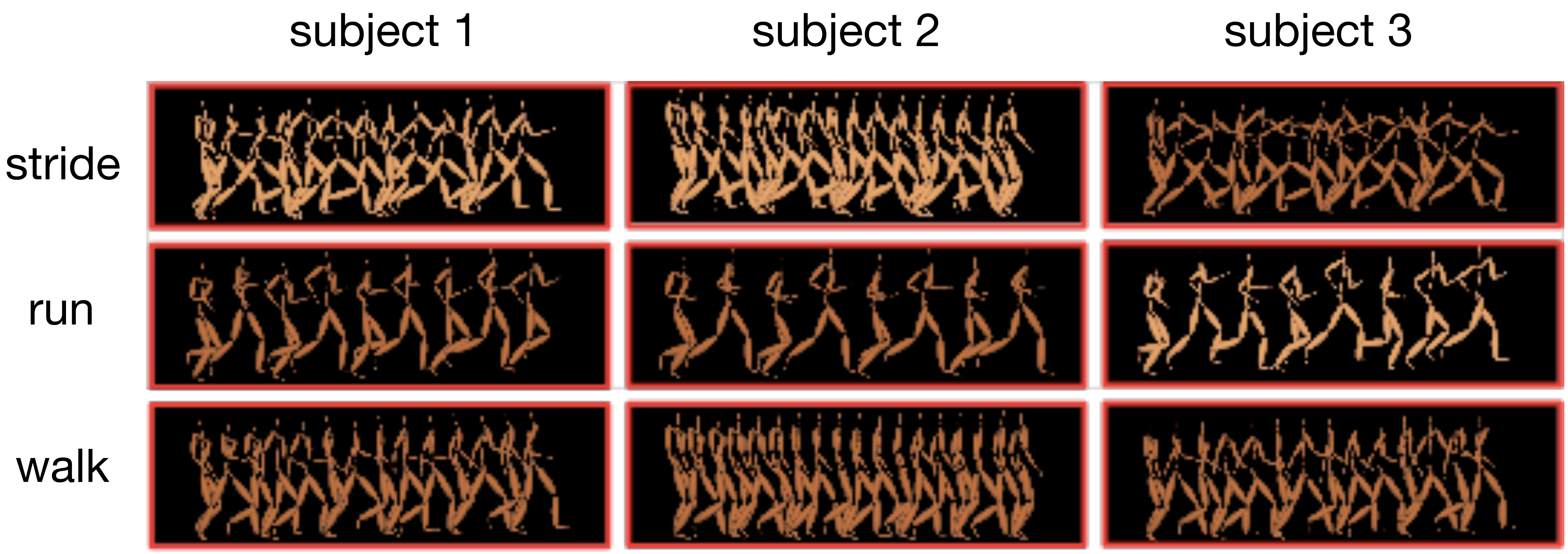
[Wang et al. ICML '07]

Training data



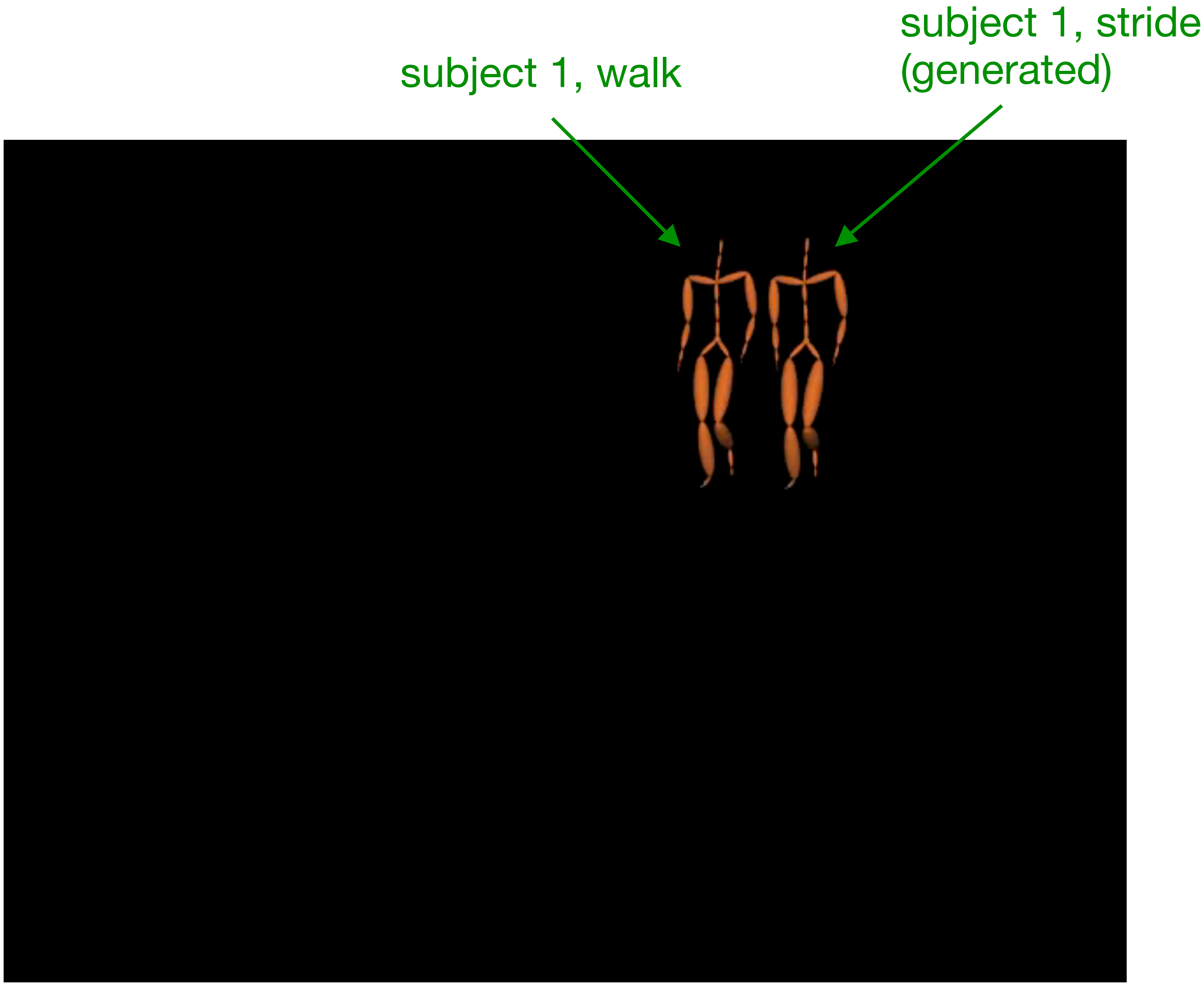
Each training motion is a sequence of poses, sharing the same combination of subject ($\mathbf{s}$) and gait ($\mathbf{g}$).

Generating new motions



The GP model provides a Gaussian prediction for new motions.
We use the mean to generate motions with different styles.

Generating new motions

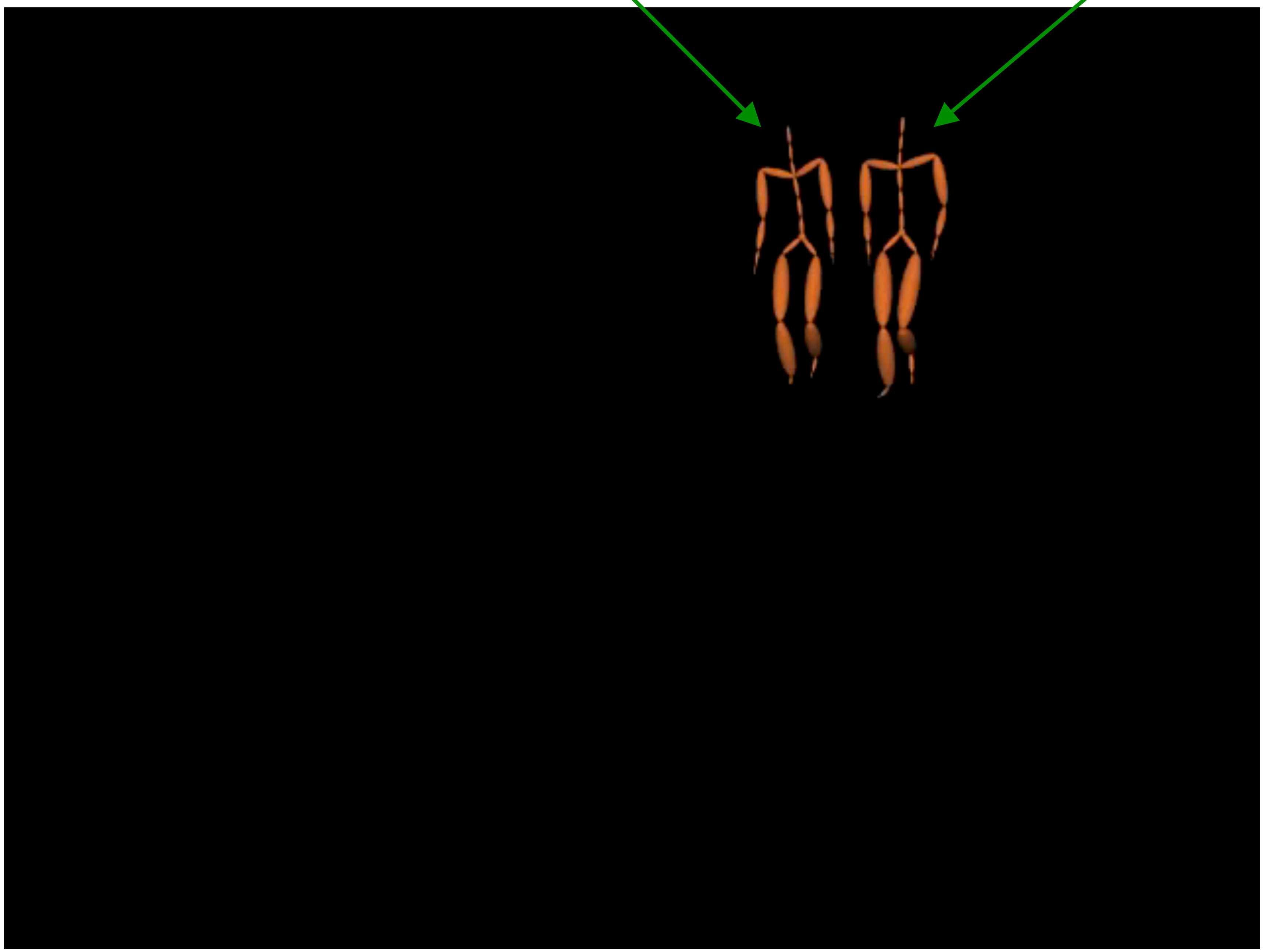


[Wang et al. ICML '07]

Generating new motions

subject 3, stride

subject 1, stride (generated)

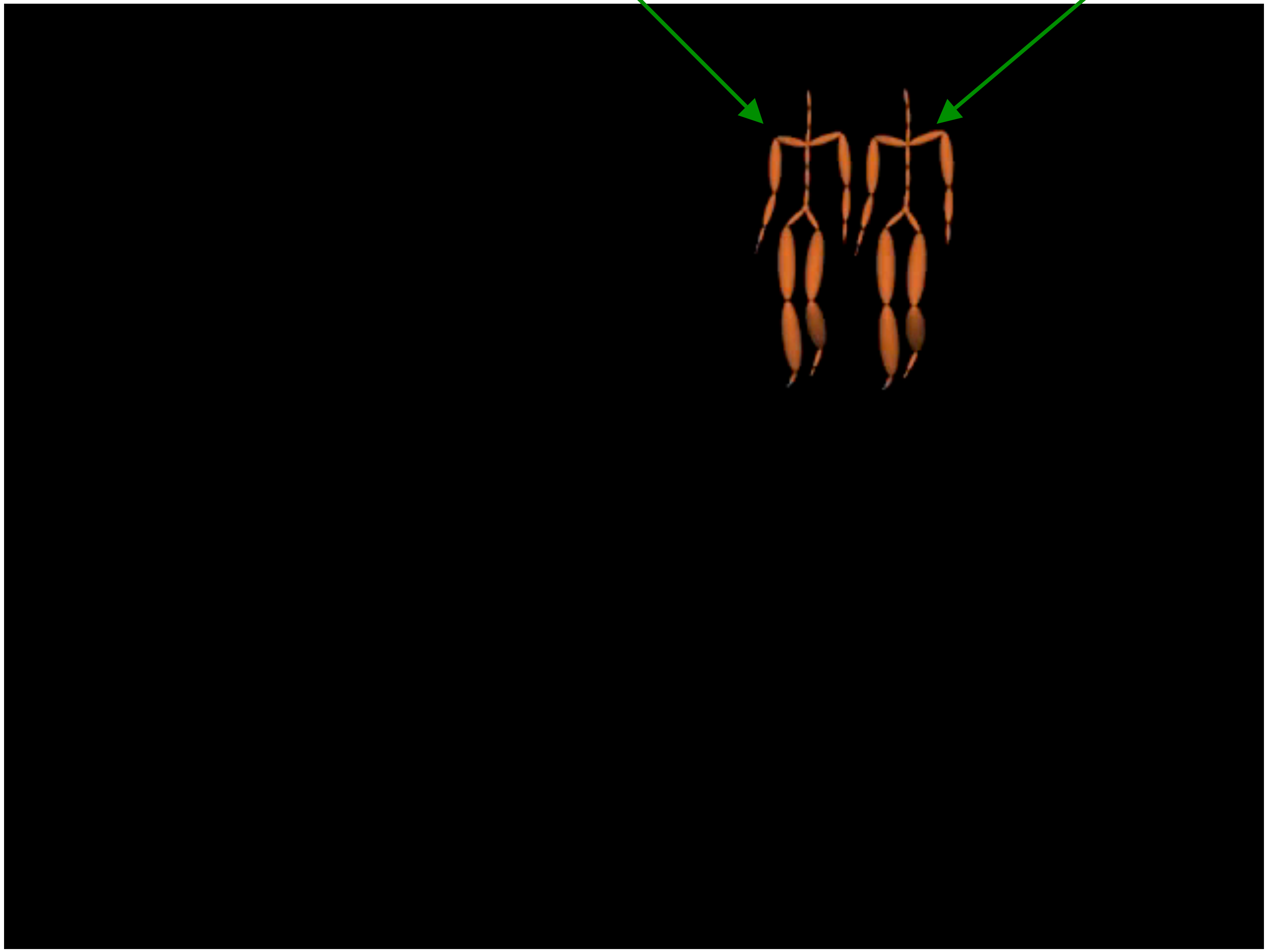


[Wang et al. ICML '07]

Generating new motions

subject 2, walk

subject 2, stride (generated)

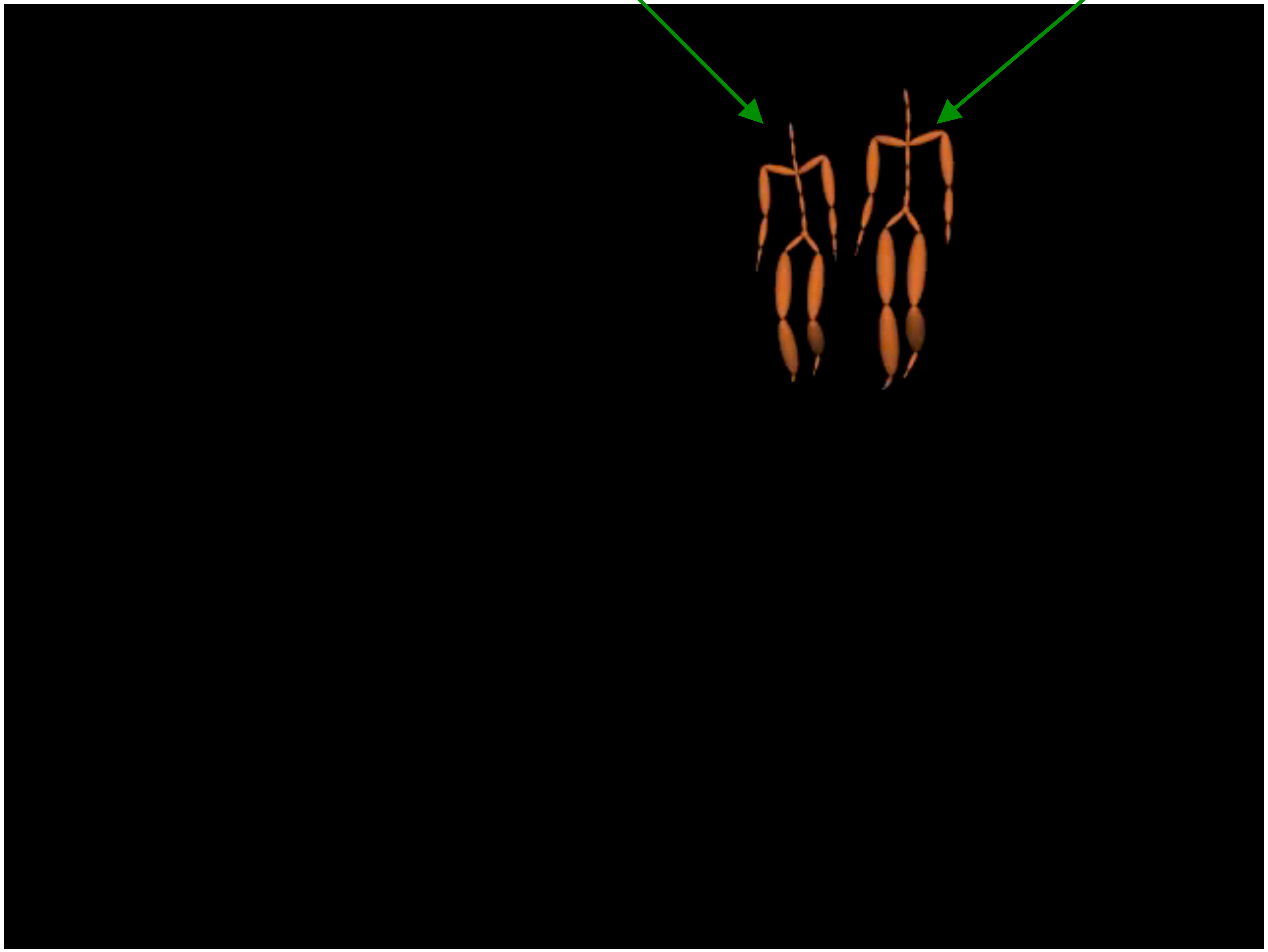


[Wang et al. ICML '07]

Generating new motions

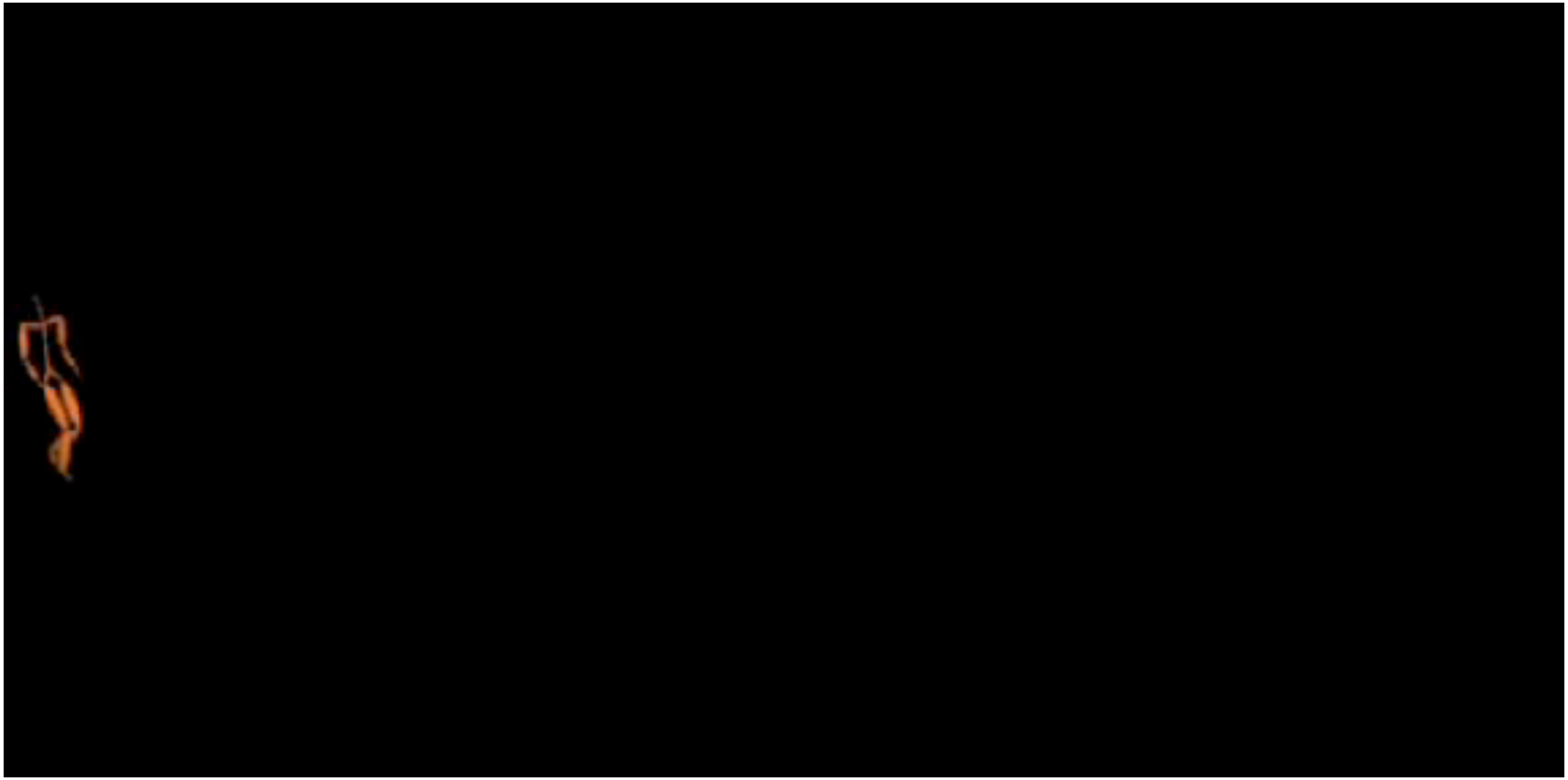
subject 3, stride

subject 2, stride (generated)



[Wang et al. ICML ’07]

Generating new motions



Transitions

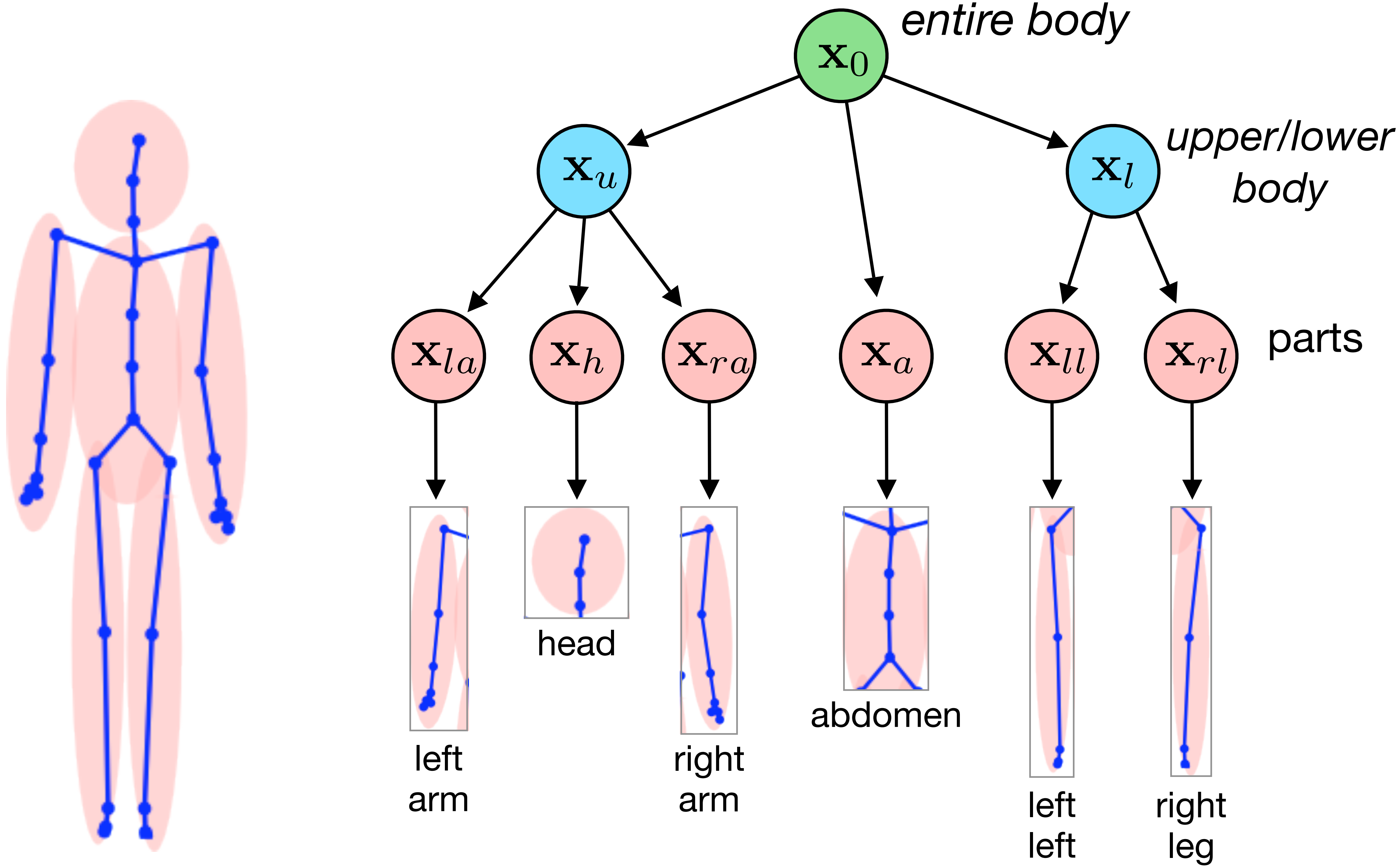
[Wang et al. ICML '07]

Generating new motions

Random motions

[Wang et al. ICML '07]

Hierarchical GPLVM



Hierarchical GPLVM *[Lawrence and Moore ICML '07]*

Selected references for GP models

Lawrence N, Probabilistic nonlinear principal components analysis with Gaussian Process latent variable models. *JMLR* 6, 2005 (also see NIPS 2004)

Moore A and Lawrence N, Hierarchical Gaussian process latent variable models. *Proc ICML*, 2007

Quinonero-Candela & Rasmussen, A unifying view of sparse approximate Gaussian Process regression. *JMLR* 6, 2006

Urtasun R et al., People tracking with the Gaussian process dynamical model. *Proc IEEE CVPR,* 2006

Urtasun R et al., Topologically constrained latent variable models. *Proc ICML* 2008

Wang J et al ., Multifactor Gaussian process models for style-content separation. *Proc ICML,* 2007.

Wang J et al, Gaussian Process dynamical models for human motion. *IEEE Trans PAMI* 30(2), 2008 (also see NIPS 2005)

Discriminative Pose Estimation

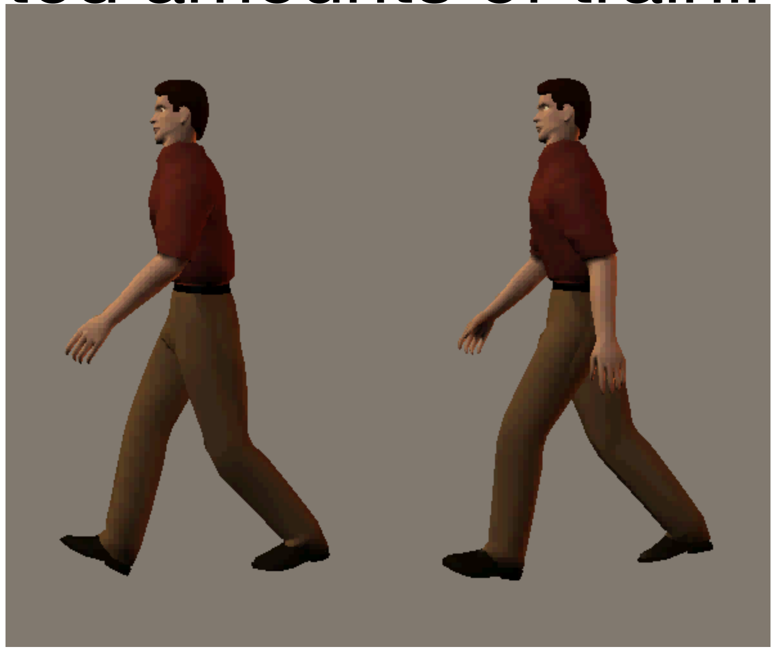
Discriminative pose estimation

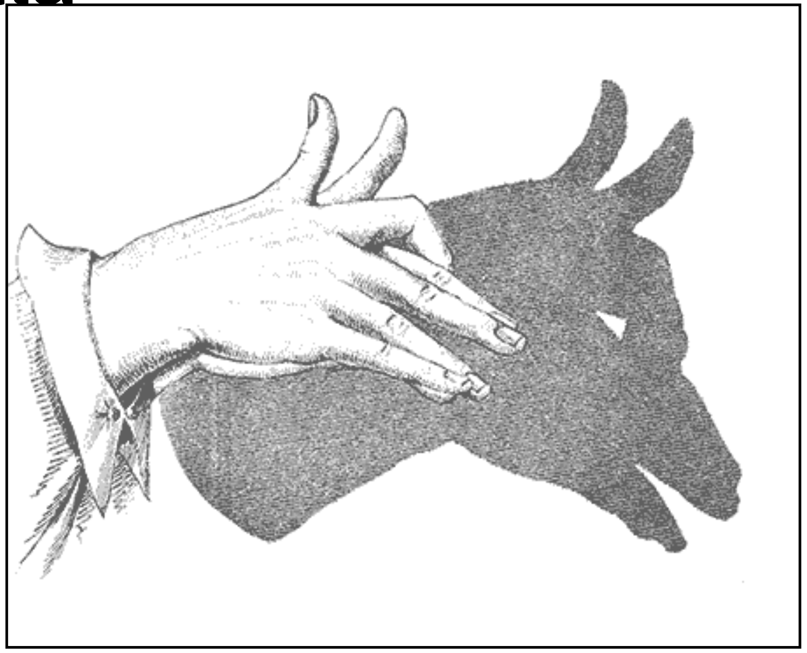
Parameterized model for the conditional density $p(\mathbf{y} \mid \mathbf{x})$

3D pose   image features

Challenges:

- high-dimensional features / high dimensional pose
- ambiguities imply a multi-modal regression problem
- limited amounts of training data



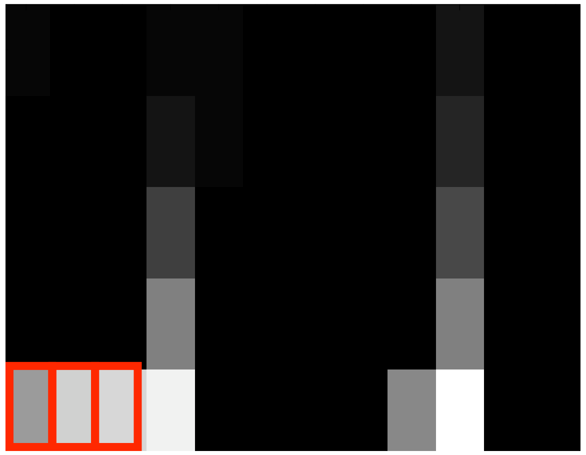


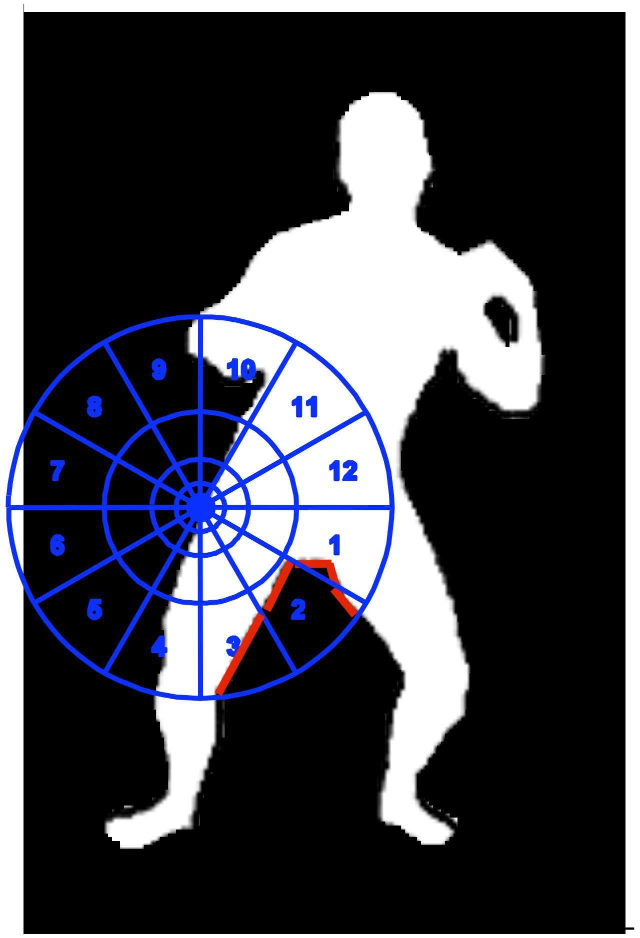
Features

Image descriptor

- HOG (or SIFT)
- Shape Context
- Hierarchical Descriptors (HMAX, Spatial Pyramid, Vocabulary Tree, Multilevel Spatial Blocks, ...)

Shape Context: log-polar histogram of edge points





Features

Image descriptor

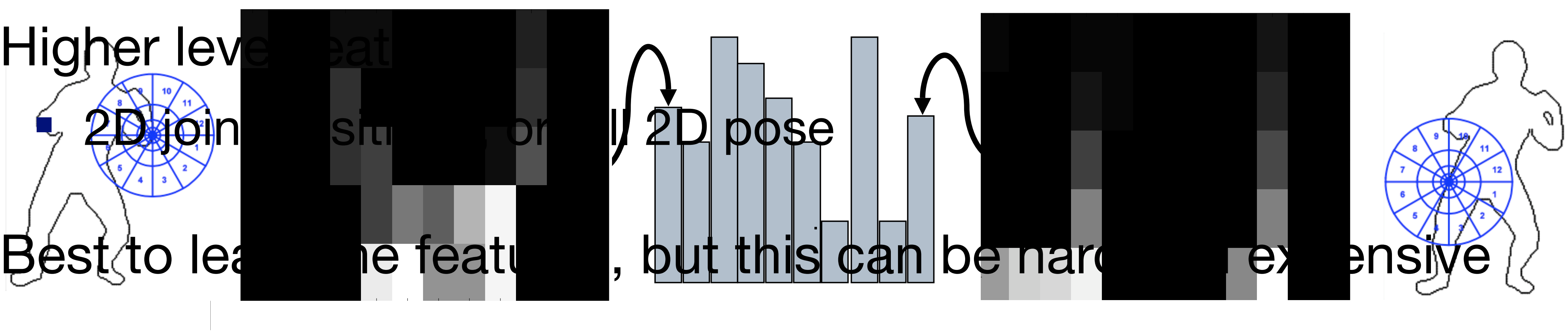
- HOG (or SIFT)
- Shape Context
- Hierarchical Descriptors (HMAX, Spatial Pyramid, Vocabulary Tree, Multilevel Spatial Blocks, ...)

Vector quantization (reduce descriptor dimensionality)

Higher leve[illegible]

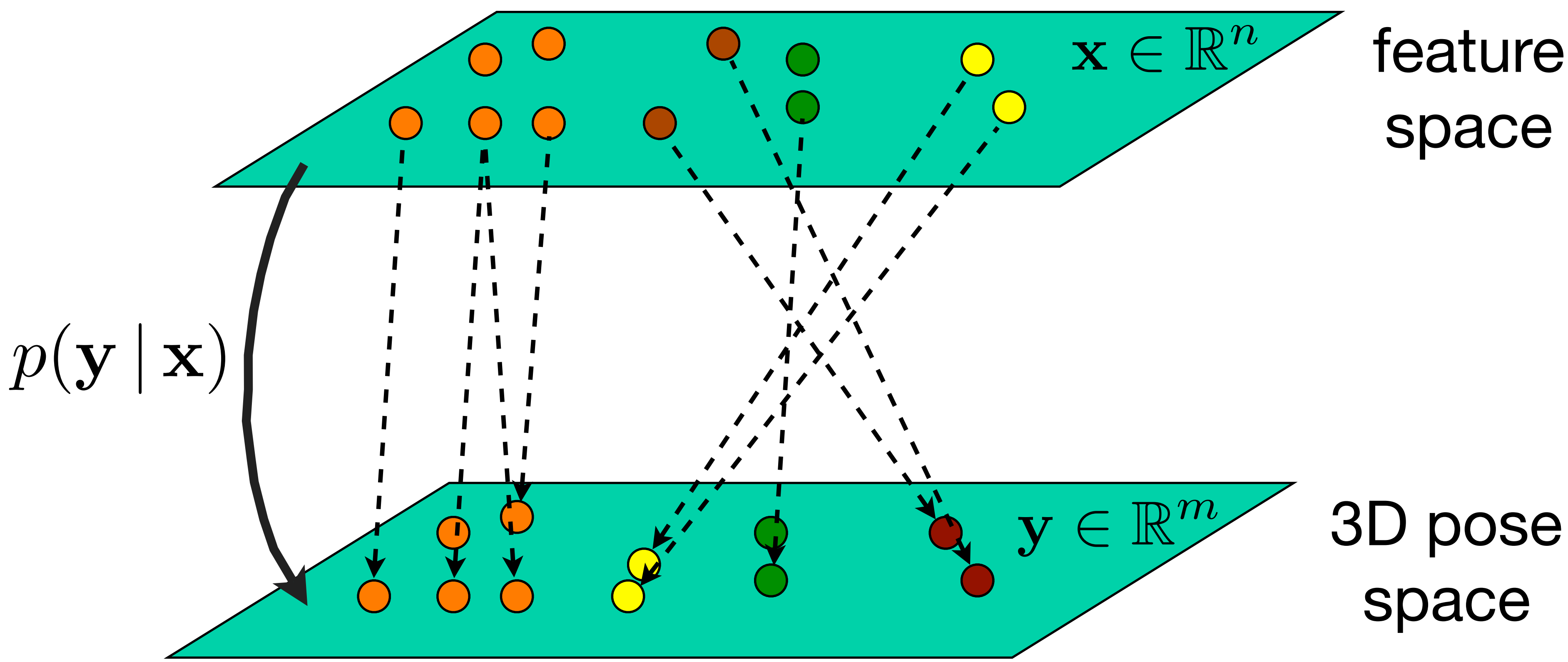
- 2D join[illegible] or [illegible] 2D pose

Best to lea[illegible] featu[illegible], but this can be [illegible] ex[illegible]ensive



Multi-valued Regression: Mixtures of experts

We want to find a "mapping" from features to 3D poses;
i.e., a conditional distribution $p(3D\ pose \,|\, features)$



Problem: $p(\mathbf{y} \,|\, \mathbf{x})$ is non-linear, and not one-to-one.

Solution: Approximate $p(\mathbf{y} \,|\, \mathbf{x})$ with locally linear mappings.

Multi-valued regression: Mixtures of experts

pose features gating functions experts (mappings)

$$p(\mathbf{y} \mid \mathbf{x}) \;\propto\; \underbrace{\sum_{k=1}^{M}}_{\text{over experts}} p_{g,k}(k \mid \mathbf{x}, \theta_{g,k}) \; p_{e,k}(\mathbf{y} \mid \mathbf{x}, \theta_{e,k})$$

Experts – ridge regression with constant offset

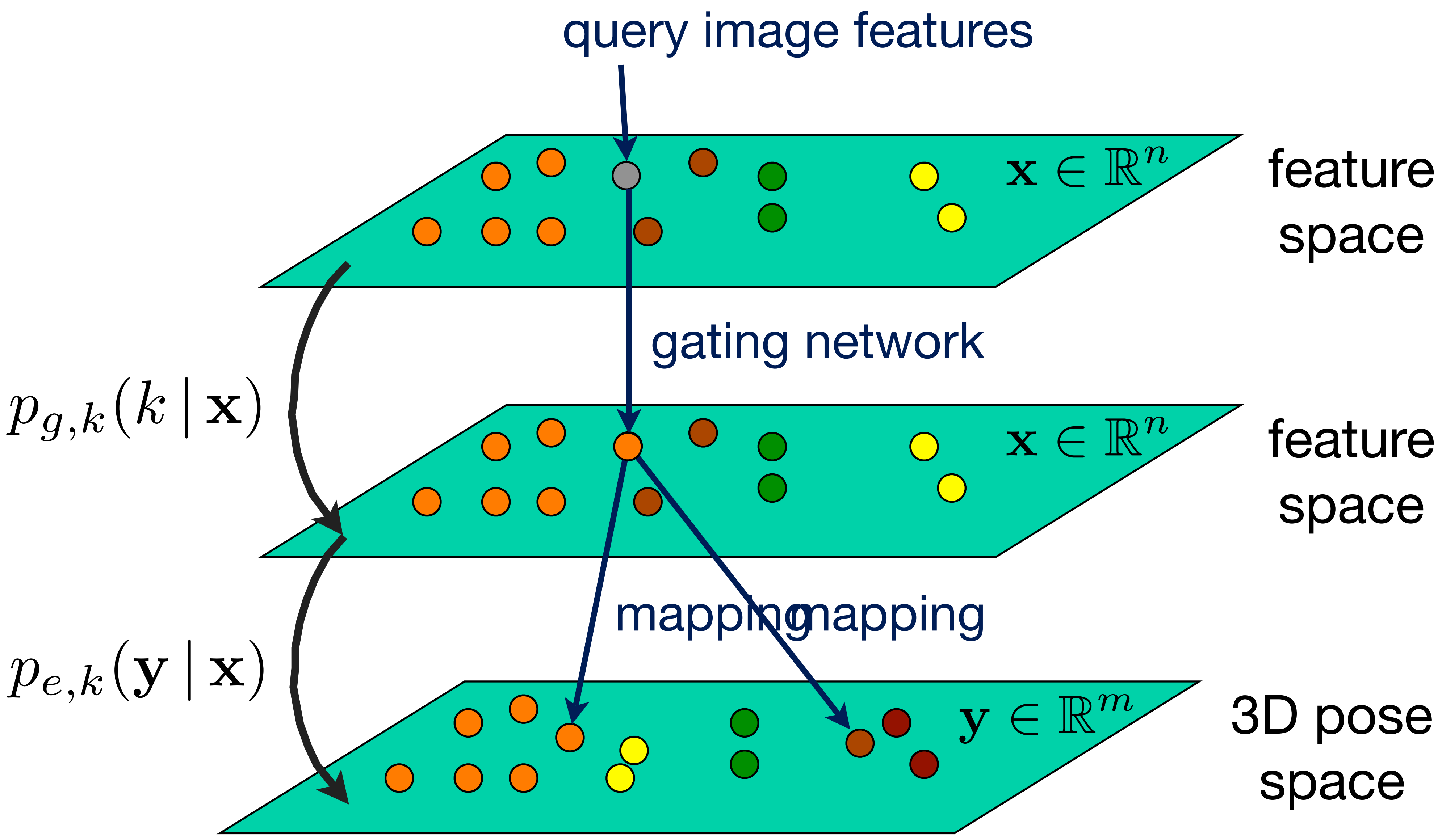
Gating functions – Gaussian

Training – similar to EM for Gaussian mixture models

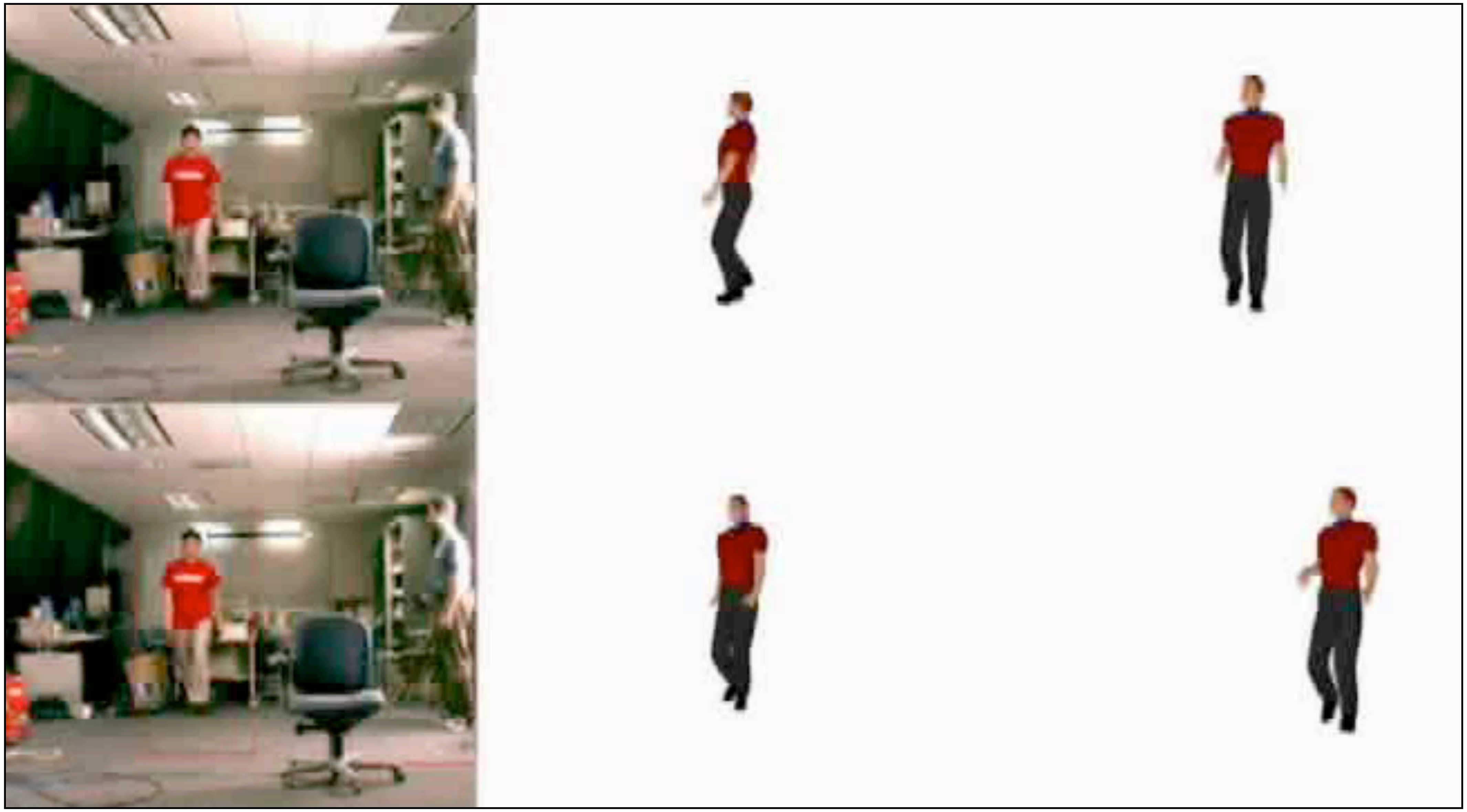
[Jordan and Jacobs, 94]

[Waterhouse et al, 96]

Multi-valued regression: Mixtures of experts



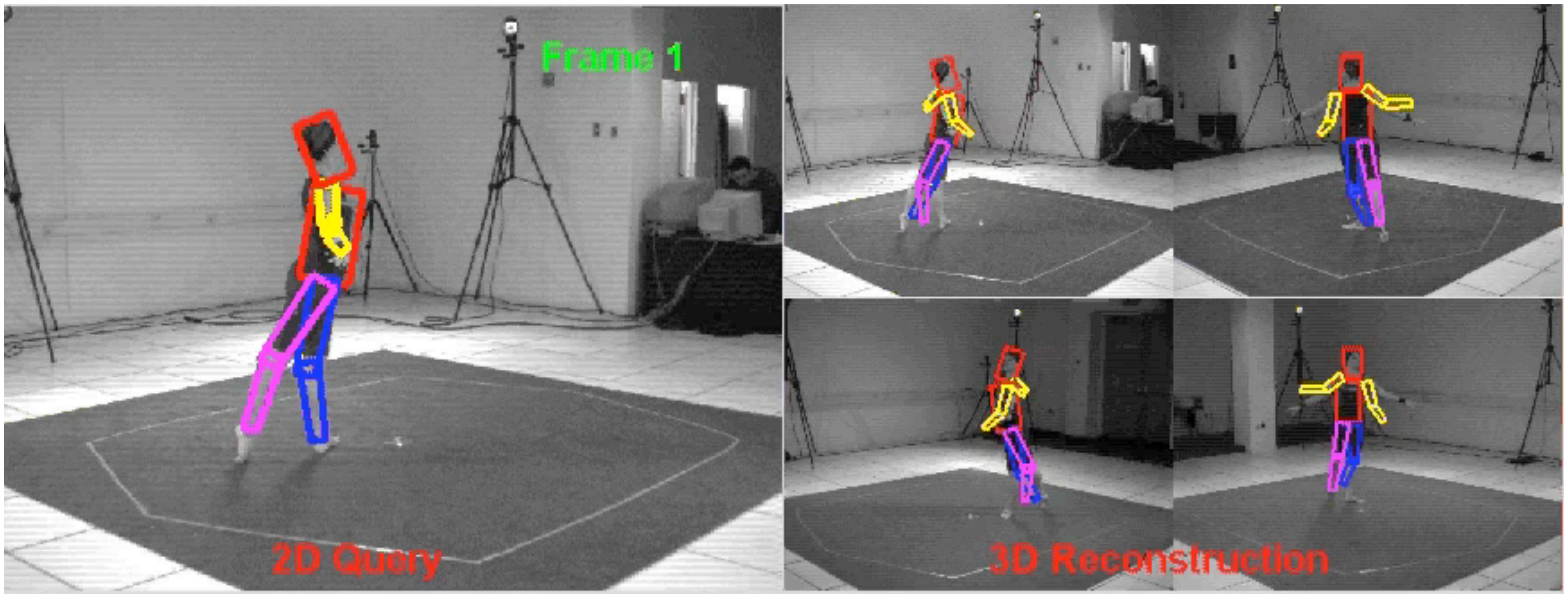
Mixtures of experts: Results



[Sminchisescu et al, CVPR'06]

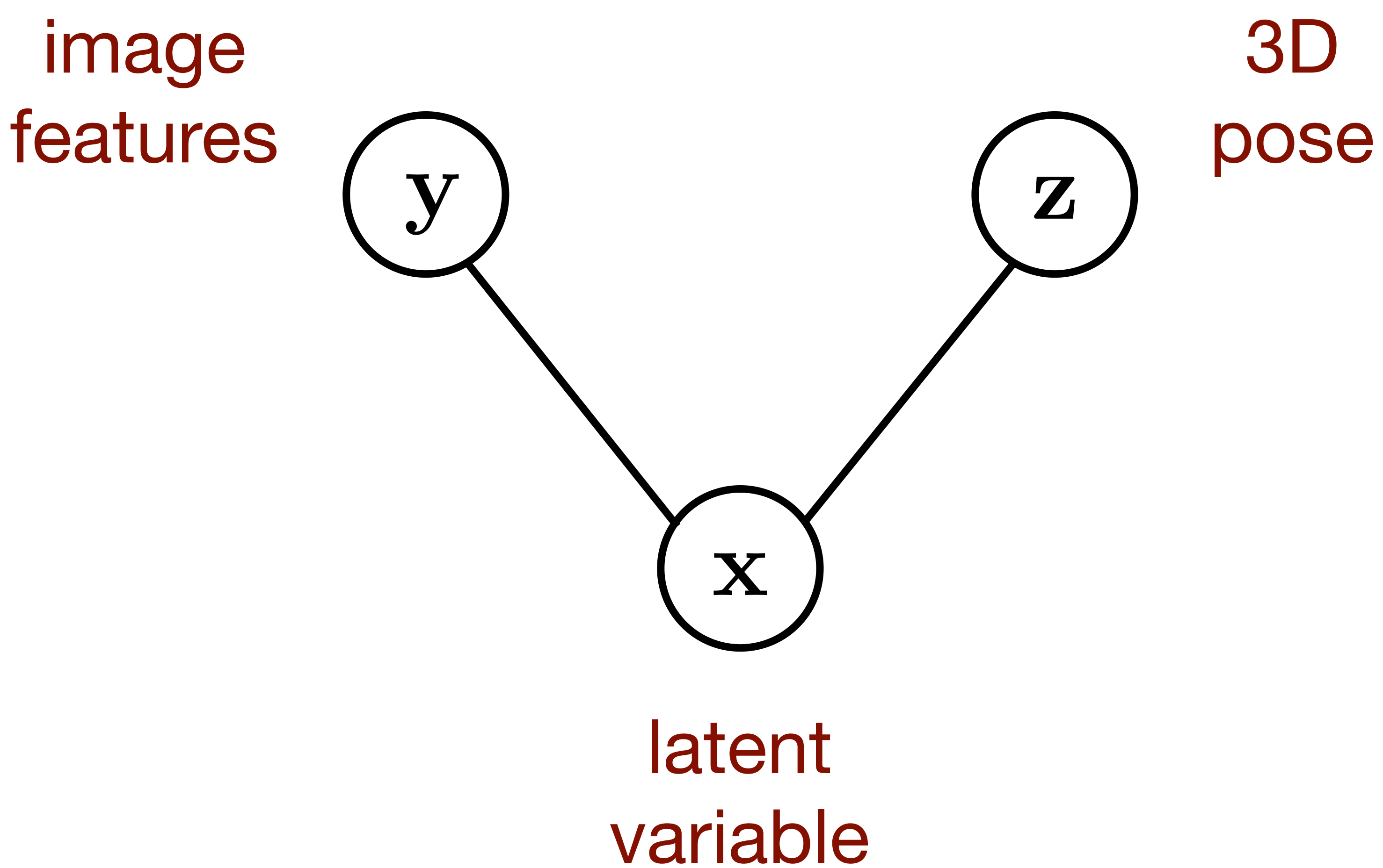
Mixtures of experts: Results

Estimated 2D pose is the input feature:



[Sigal and Black, AMDO'06]

Shared latent variable models



E.g.: sGPLVM *[Navaratnum et al 2007]*, sKIE *[Sigal et al 2008]*, ...

Selected readings for discriminative methods

Local Models

- Nearest-neighbor *[Mori and Malik, ECCV 02]*
- Locally weighted regression *[Shakhnarovich et al, ICCV 03]*
- Gaussian processes regression *[Urtasun and Darrell, CVPR 08]*

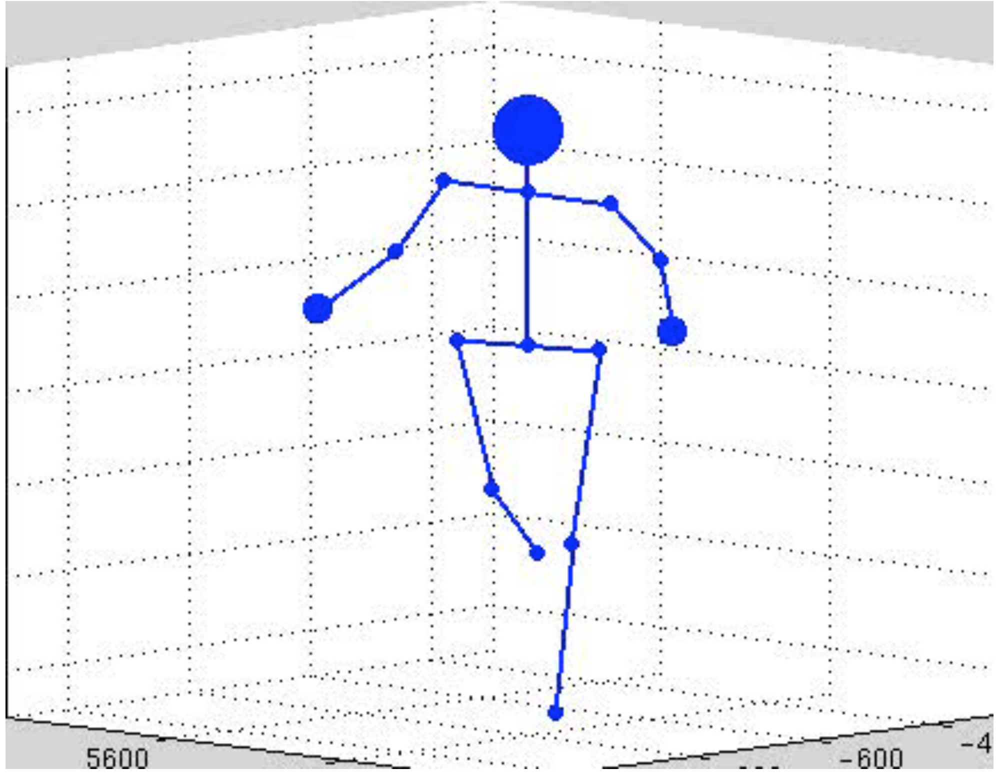
Global Models

- Linear regression, RVM regression, mixtures of regressors *[Agarwal & Triggs, ICML 04, CVPR 04/05]*
- Mixtures of experts *[Sminchisescu et al, CVPR 05/06]*
- Gaussian Process LVMs *[Navaratnam et al, ICCV 07]*
- Spectral LVMs *[Kanaujia et al, ICCV 07]*
- Kernel information embeddings *[Sigal et al, CVPR 09]*

Physics-Based Models

Physics-based models

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. ICCV `05]

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

Will learning scale?

Problem: Learning kinematic pose and motion models from mocap data, with the environment and interactions, 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:

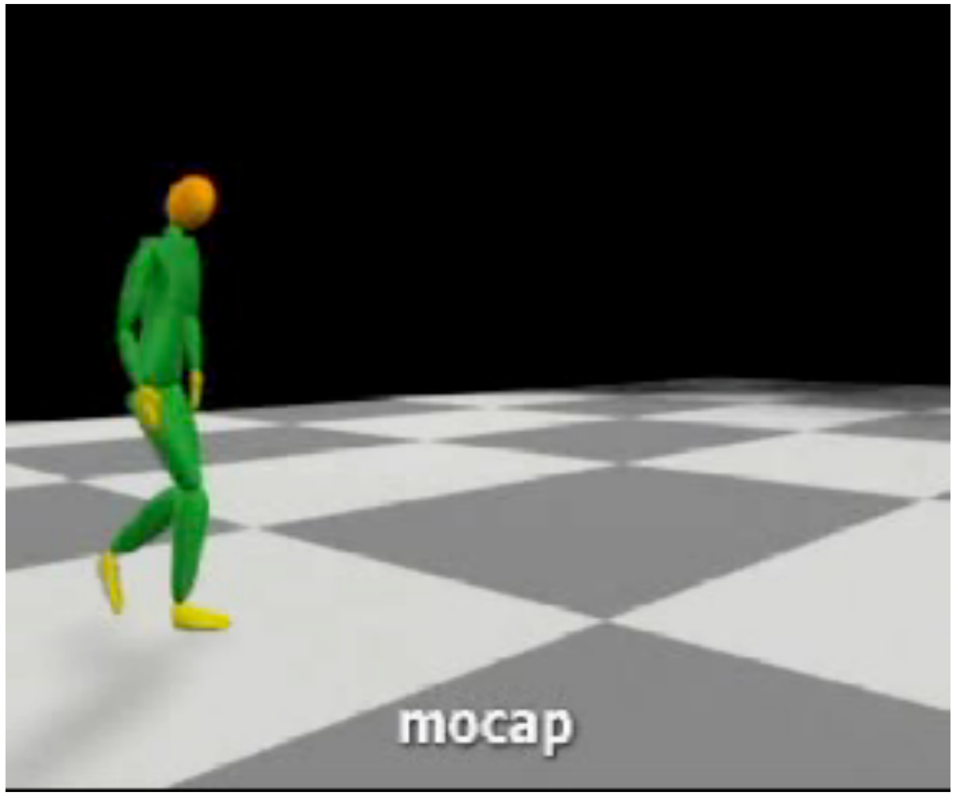
- balance and body lean (e.g., on hills)
- sudden accelerations (e.g., collisions)
- static contact (e.g., avoiding footskate)
- variations in style due to speed and mass distribution (e.g., carrying an object)
- ...

Physics-based models for pose tracking

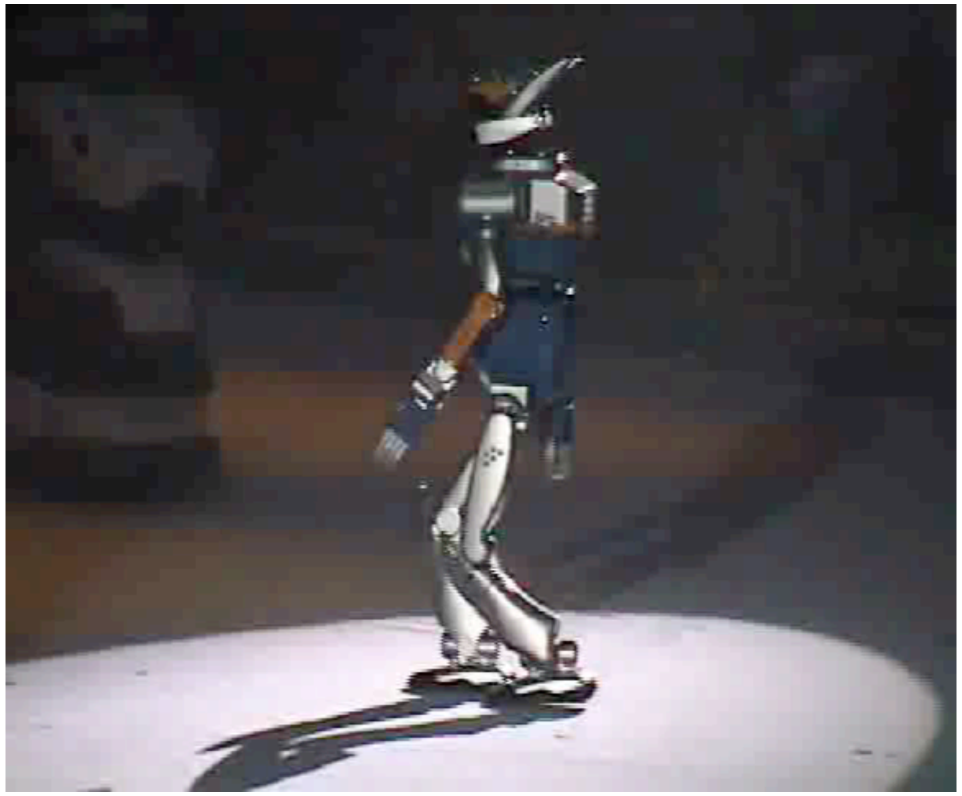
Incorporate basic principles of physics into models of biological motion:

- ensure physically plausible pose estimates
- reduce reliance on mocap data
- model interactions

Modeling full-body dynamics is difficult



[Liu et al. `06]



[Kawada Industries HRP-2. `03]

Passive dynamics

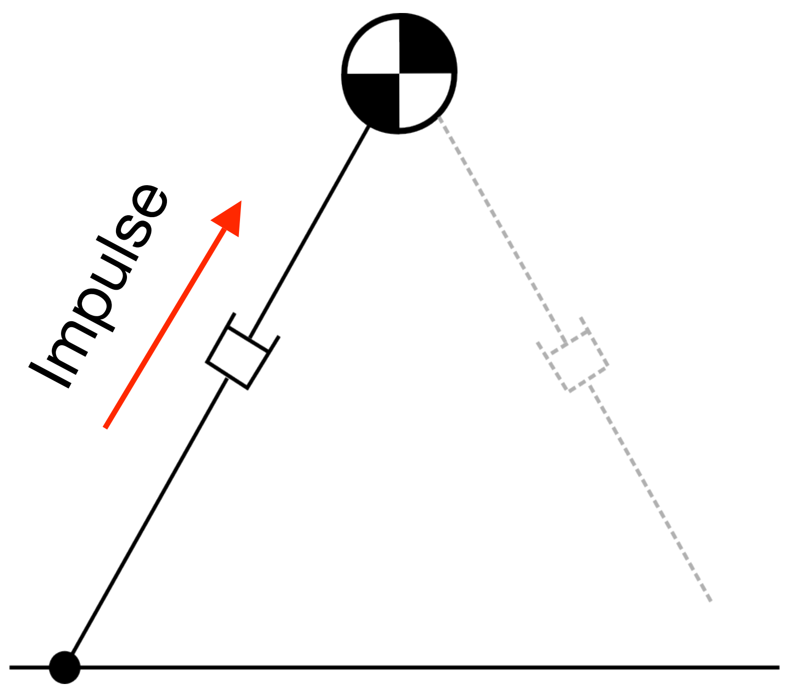
But much of walking is essentially passive.

[McGeer 1990]

[Collins & Ruina 2005]

Simplified planar biomechanical models

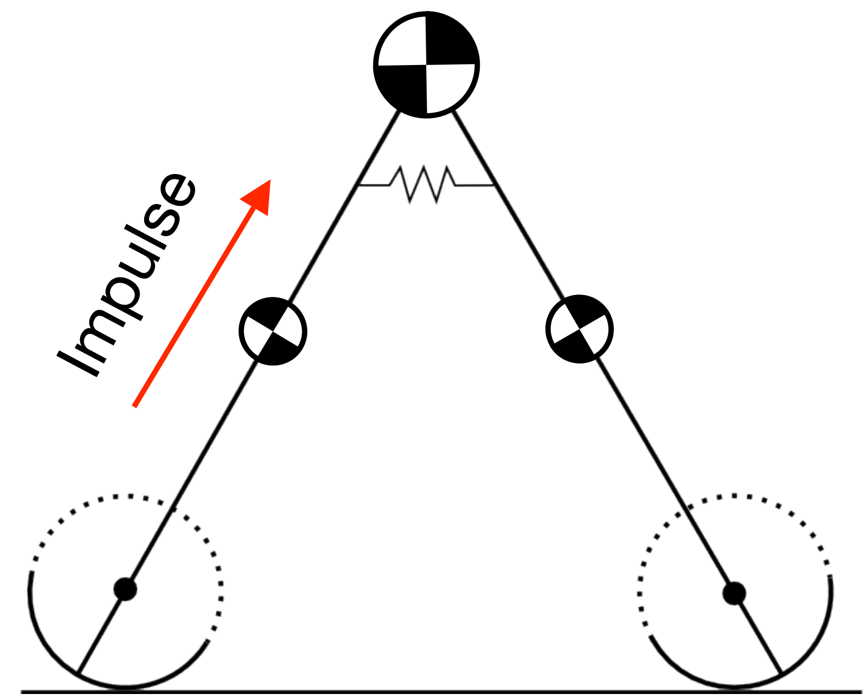
Monopode



[Blickhan & Full 1993; *Srinivasan & Ruina 2000]*

- point-mass at hip, massless legs with prismatic joints, and impulsive toe-off force
- inverted pendular motion

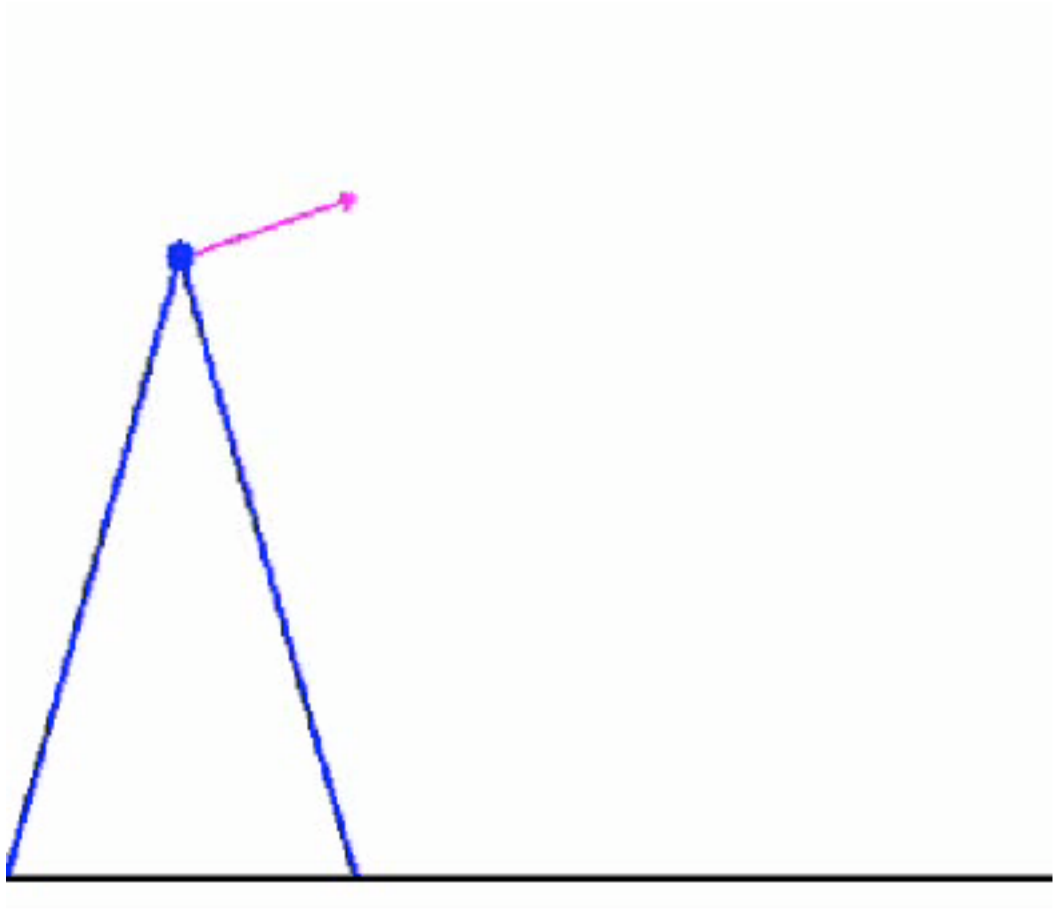
Anthropomorphic Walker



[McGeer 1990; Kuo 2001,2002]

- rigid bodies for torso and legs
- forces due to torsional spring between legs and an impulsive toe-off

Anthropomorphic walker gait



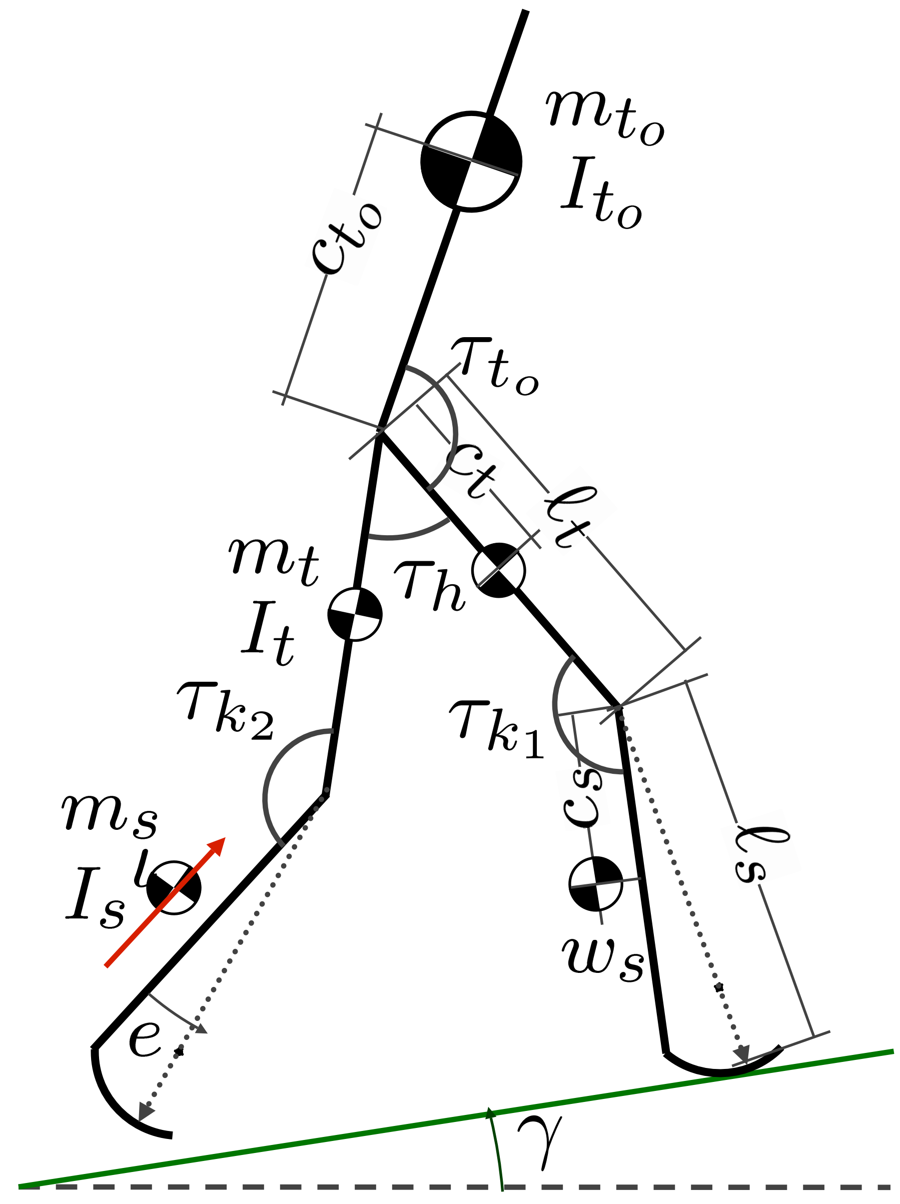
The *Kneed Walker*

Kneed planar walker comprises

- torso, legs with knees & feet
- inertial parameters from biomechanical data

Dynamics due to:

- joint torques $\tau_{t_o}, \tau_h, \tau_{k_1}, \tau_{k_2}$ (for torso, hip, & knees)
- impulse applied at toe-off (with magnitude ι)
- gravitational acceleration (w.r.t. ground slope γ)



[Brubaker and Fleet `08]

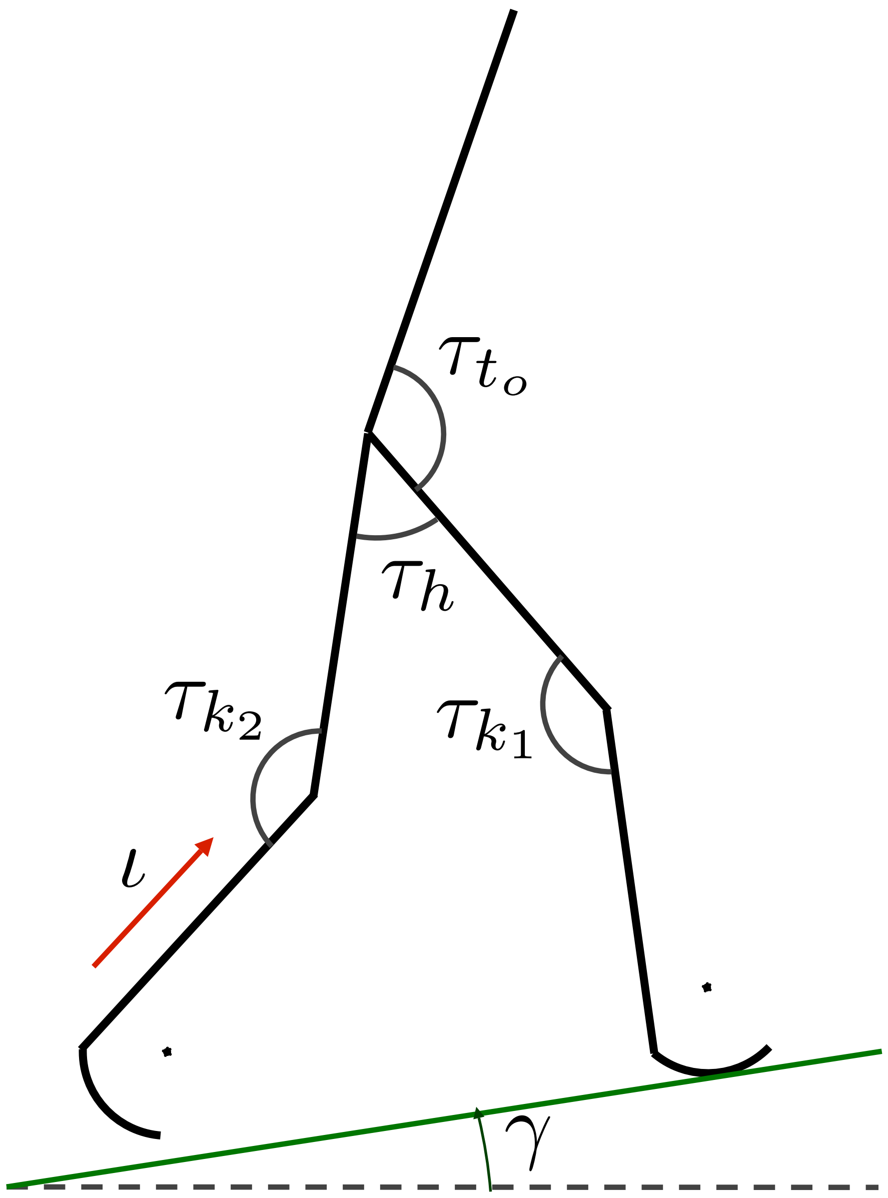
The *Kneed Walker*

Joint torques are parameterized as damped linear springs.

For hip torque

$$\begin{aligned}\tau_h &= \kappa_h\,(\phi_{t_2}+\phi_{t_1}-\phi_h) \\ &\quad - d_h(\dot\phi_{t_2}+\dot\phi_{t_1})\end{aligned}$$

with stiffness and damping coefficients, κ_h and d_h, and resting length ϕ_h



The *Kneed Walker*

Equations of motion

$$\mathcal{M}\ddot{\mathbf{q}} = f_s(\vec{\kappa}, \vec{d}, \vec{\phi}) + f_g + f_c$$

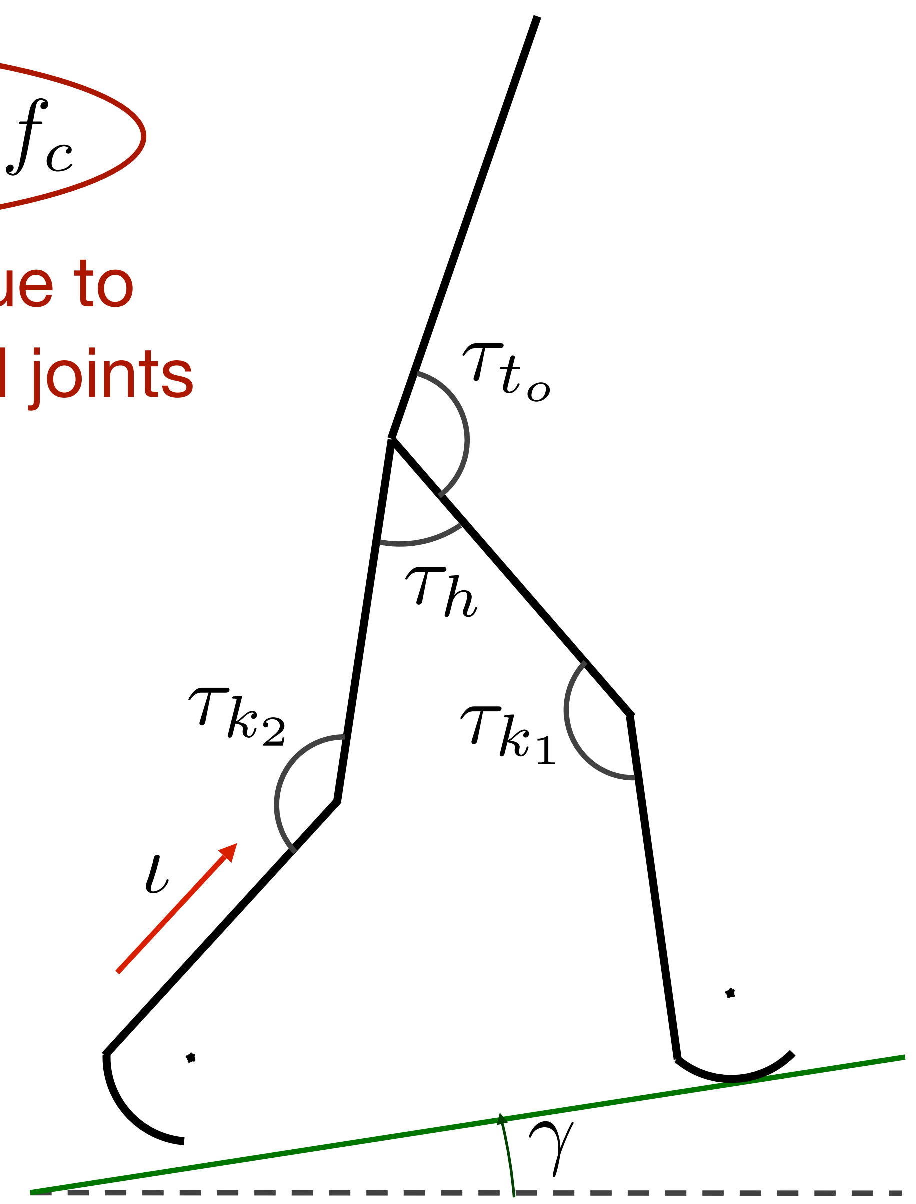
generalized mass matrix

acceleration

spring torques

forces due to gravity and joints

... plus ground collisions and joint limits (esp. knee)



Prior for the Kneed Walker

How do we design a prior density over dynamics for walking?

Assumption: Human walking motions are characterized by efficient, stable, cyclic gaits.

Approach:

- Find control parameters that produce optimal cyclic gaits over a range of speeds & step lengths, for various surface slopes, with minimal energy.
- Assume additive process noise in the control parameters to capture variations in style.

Efficient, cyclic gaits

Search for dynamics parameters $\vec{\theta} = (\vec{\kappa}, \vec{d}, \vec{\phi}, \iota)$ and initial state $\mathbf{x} = (\mathbf{q}, \dot{\mathbf{q}})$ that produce cyclic locomotion at speed s, step length ℓ, and slope γ, with minimal "energy".

Solve

$$\min_{\vec{\theta}, \mathbf{x}} \; E(\vec{\theta}, \mathbf{x};\, s, \ell, \gamma) \quad \text{s.t.} \;\; C(\vec{\theta}, \mathbf{x};\, s, \ell, \gamma) < \epsilon$$

where $E(\vec{\theta}, \mathbf{x};\, s, \ell, \gamma)$ measures the "cost" of the motion, and $C(\vec{\theta}, \mathbf{x};\, s, \ell, \gamma)$ measures the deviation from periodic motion with the target speed and step-length.

Efficient, cyclic gaits

Speed: 5.8 km/hr; Step length: 0.6 m; Slope: 0^o

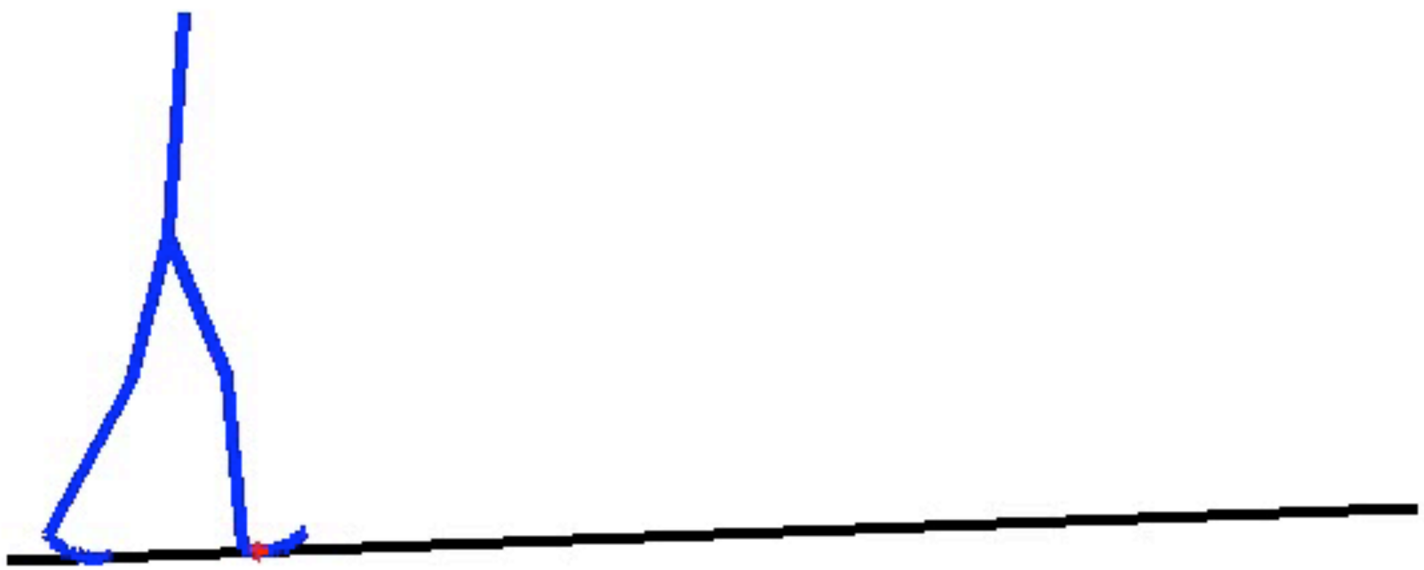
Efficient, cyclic gaits

Speed: 6.5 km/hr; Step length: 0.6 m; Slope: 4.3°

Efficient, cyclic gaits

Speed: 3.6 km/hr; Step length: 0.4 m; Slope: 4.3°

Efficient, cyclic gaits



Speed: 5.0 km/hr; Step length: 0.6 m; Slope: 2.1°

Efficient, cyclic gaits

Speed: 4.3 km/hr; Step length: 0.8 m; Slope: -2.1°

Efficient, cyclic gaits

Speed: 5.8 km/hr; Step length: 1.0 m; Slope: -4.3°

Stochastic dynamics

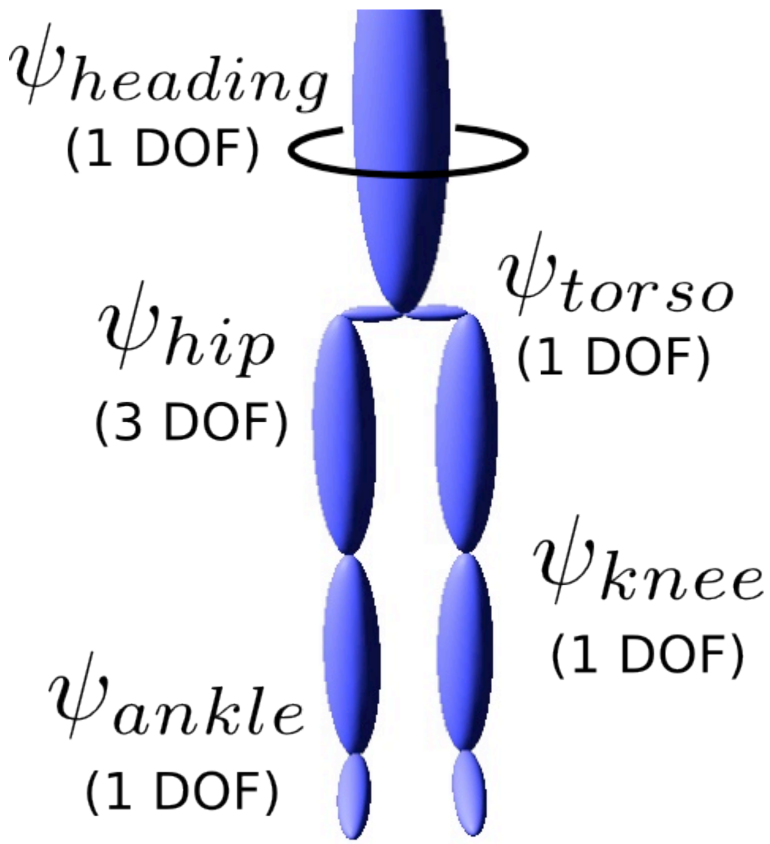
Our prior over human walking motions is derived from the manifold of optimal cyclic gaits, plus

- additive noise on the control parameters (i.e., spring stiffness, resting lengths, and impulse magnitude).
- additive noise on the resulting torques.

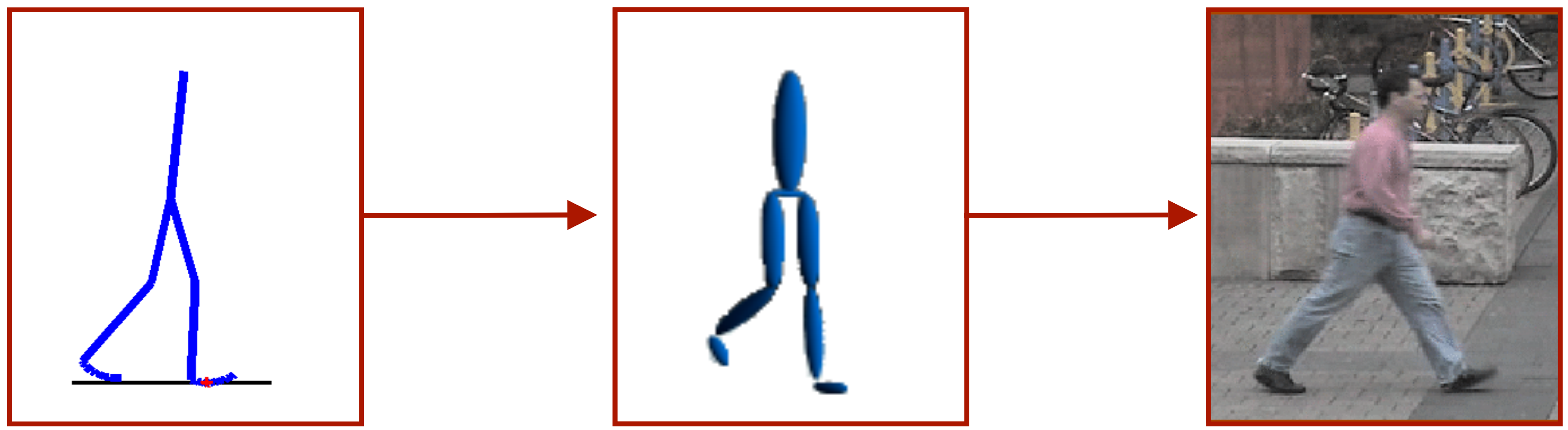
3D kinematic model

Kinematic parameters (15D) include global torso position and orientation, plus hips, knees and ankles.

- dynamics constrains contact of stance foot, hip angles (in sagittal plane), and knee/ankle angles
- other parameters modeled as smooth, second-order Markov processes.



Graphical model



2D dynamics

3D kinematics

image observations

Bayesian people tracking

Image observations: $\mathbf{z}_{1:t} \equiv (\mathbf{z}_1, \ldots, \mathbf{z}_t)$

State: $\mathbf{s}_t = [d_t, k_t]$

dynamics pose

Posterior distribution:

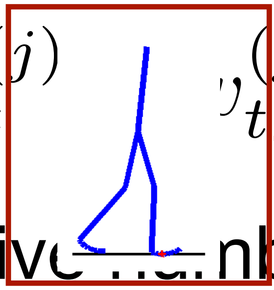
$$p(\mathbf{s}_{1:t} \mid \mathbf{z}_{1:t}) \;\propto\; p(\mathbf{z}_t \mid \mathbf{s}_t)\, p(\mathbf{s}_t \mid \mathbf{s}_{1:t-1})\, p(\mathbf{s}_{1:t-1} \mid \mathbf{z}_{1:t-1})$$

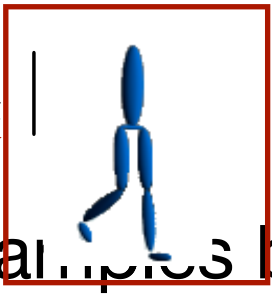
likelihood transition posterior

Sequential Monte Carlo inference:

- particle set $\mathcal{S} = \{\mathbf{s}_{1:t}^{(j)}, w_t^{(j)}\}_{j=1}^{N}$ approximates $p(\mathbf{s}_{1:t} \mid \mathbf{z}_{1:t})$
- step 1. sample next state: $\mathbf{s}_t^{(j)} \sim p(\mathbf{s}_t \mid \mathbf{s}_{t-1}^{(j)})$
- step 2. up... ight: $w_t^{(j)}$ ... $w_{t-1}^{(j)}\, p(\mathbf{z}_t \mid$
- resample when the effective number of samples becomes small

$(\iota, \kappa, \ldots)$



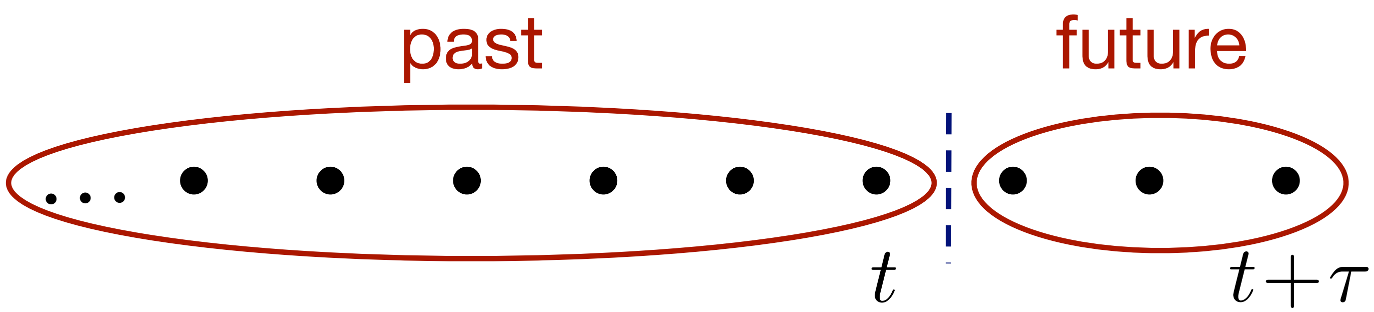


sample control parameters

simulate dynamics

sample kinematics

Bayesian people tracking



Proposals for re-sampling are given by Monte Carlo approximation, $\mathcal{Q}_t = \{\mathbf{s}_t^{(j)}, \hat{w}_t^{(j)}\}_{j=1}^N$, to the windowed smoothing distribution

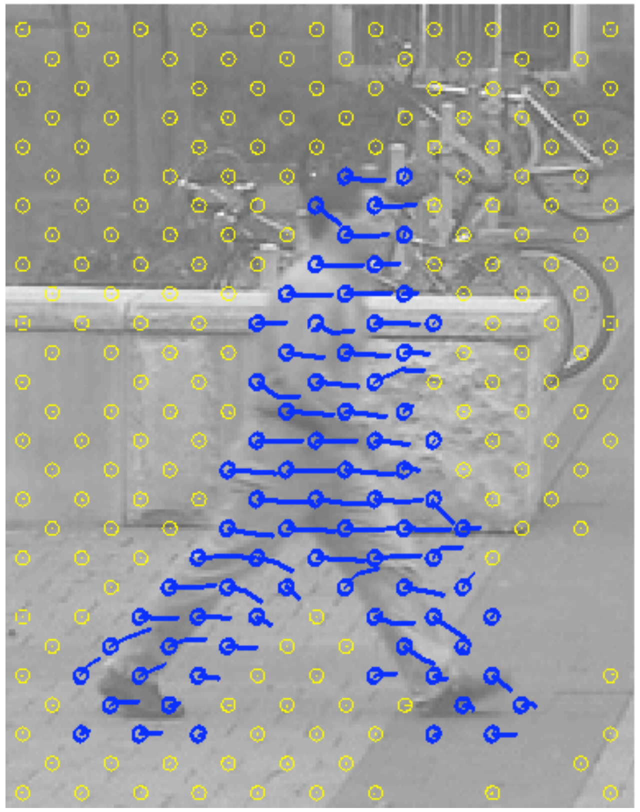
$$p(\mathbf{s}_t \,|\, \mathbf{z}_{1:t+\tau}) \propto \int_{\mathbf{s}_{t+1:t+\tau}} p(\mathbf{z}_{t:t+\tau} \,|\, \mathbf{s}_{t:t+\tau})\, p(\mathbf{s}_{t:t+\tau} \,|\, \mathbf{z}_{1:t-1})$$

Re-sample $\mathcal{S}_t$ when the effective sample size $[\sum_j (\hat{w}_t^{(j)})^2]^{-1}$ drops below threshold. Then,

- draw sample index $k(i) \sim multinomial\{\hat{w}_t^{(j)}\}_{j=1}^N$
- assign samples and perform importance re-weighting:

$$\mathbf{s}_t^{(k)} \leftarrow \mathbf{s}_t^{(i)} \qquad w_t^{(k)} \leftarrow w_t^{(i)} / \hat{w}_t^{(i)}$$

Image observations



Foreground model

Gaussian mixture model for colors in each part

Background model

mean color (RGB) and luminance gradient $E[\vec{I}(x,y),\ \nabla L(x,y)]$ with covariance matrix

Optical flow

robust regression for translation in local neighborhoods

Calibration and Initialization

- Camera calibration and ground plane are known
- Body position, pose & dynamics coarsely set manually

Speed change

input video sequence

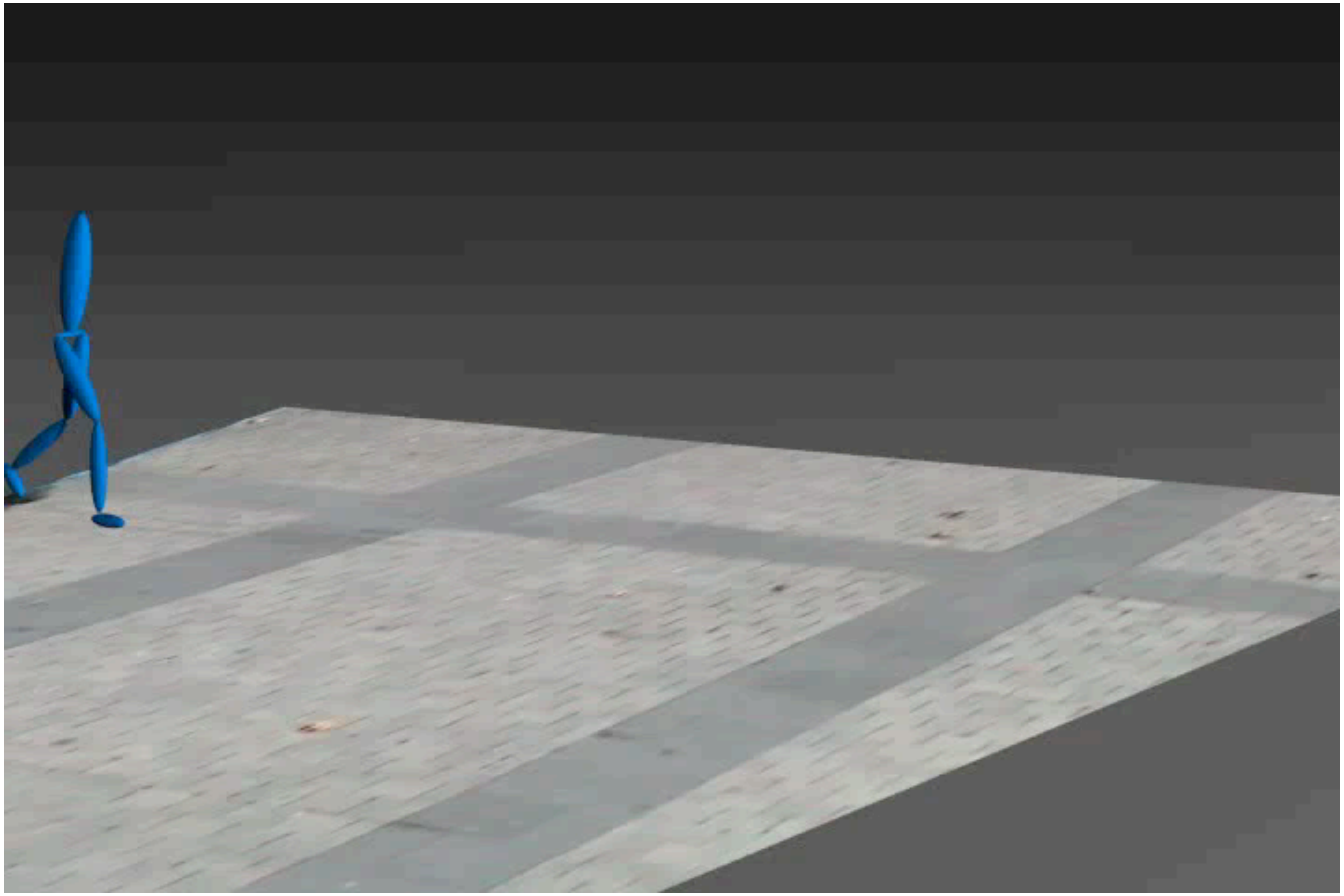
Image observations

negative log background likelihood

Speed change

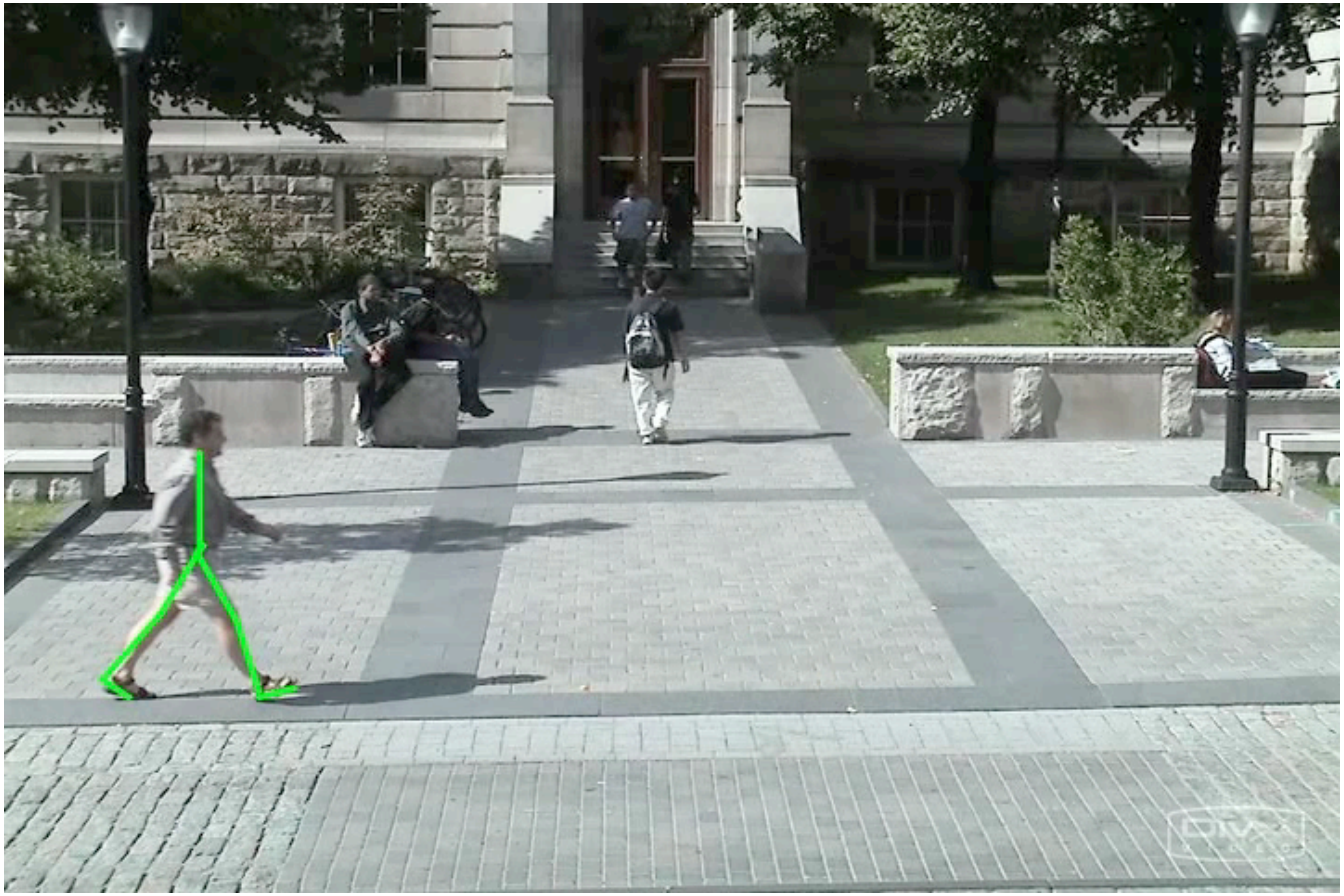
MAP Pose Trajectory (half speed)

Speed change



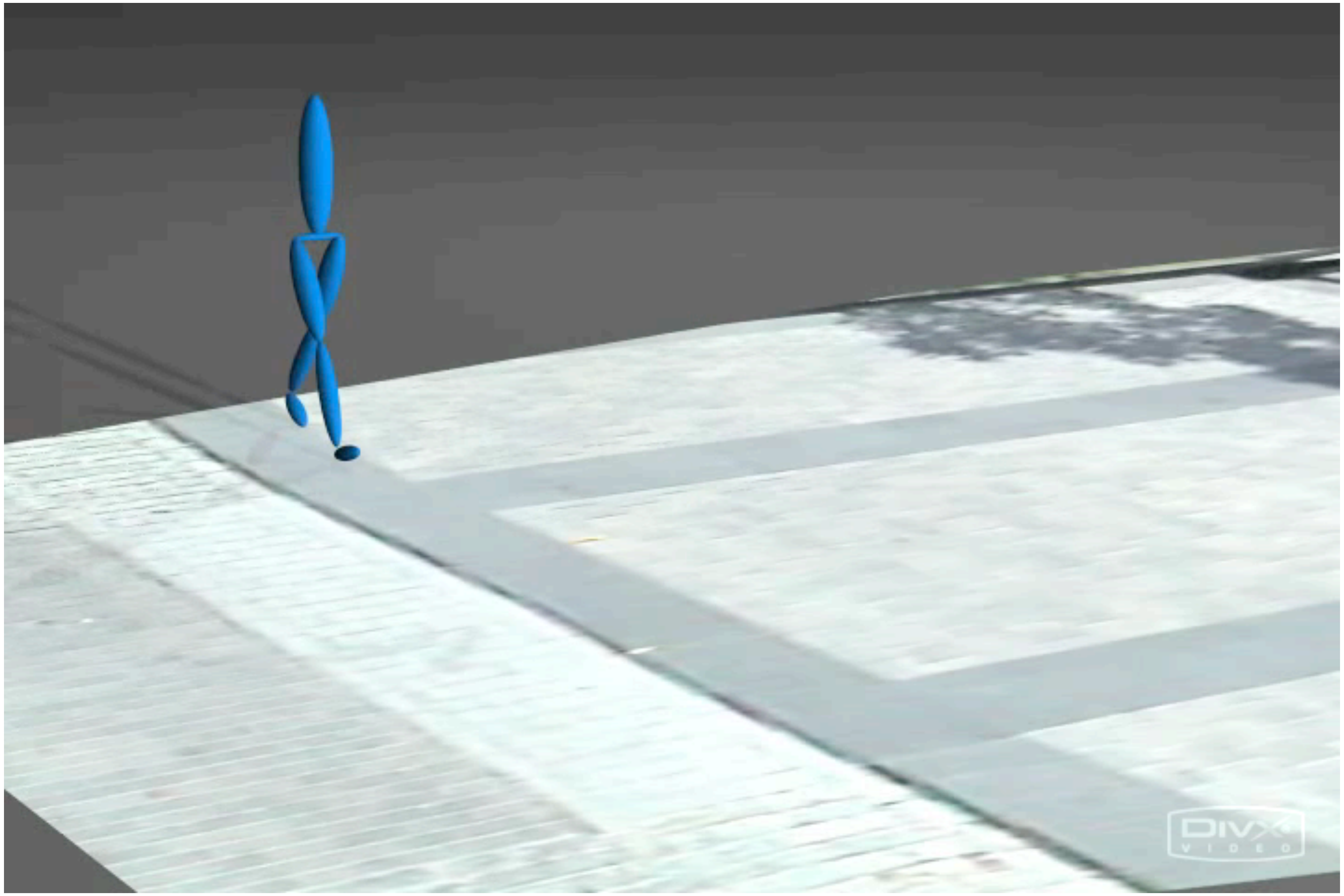
Synthetic rendering of MAP Pose Trajectory (half speed)

Occlusion



MAP Pose Trajectory (half speed)

Occlusion

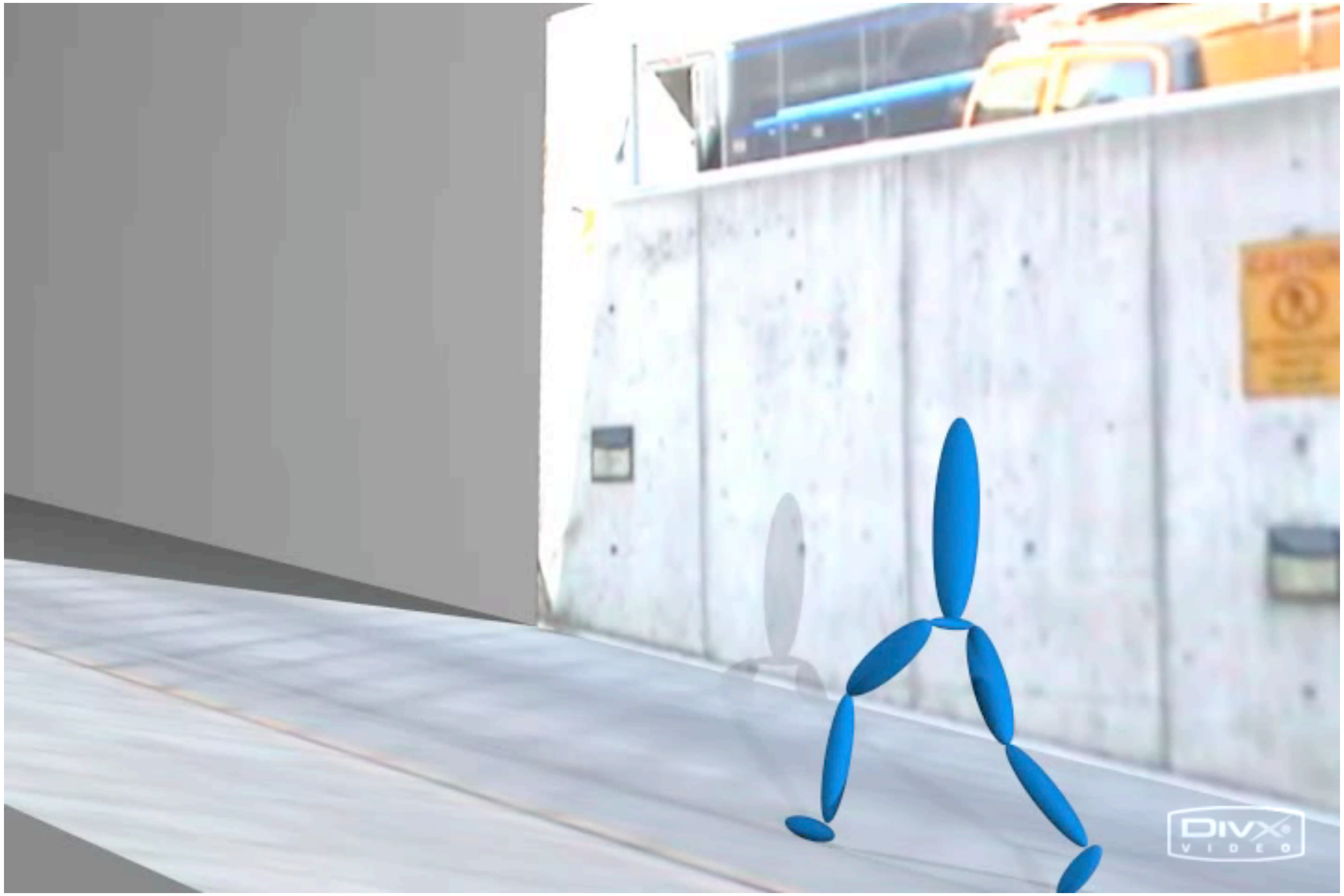


Synthetic rendering of MAP Pose Trajectory (half speed)

Sloped surface (~10º)

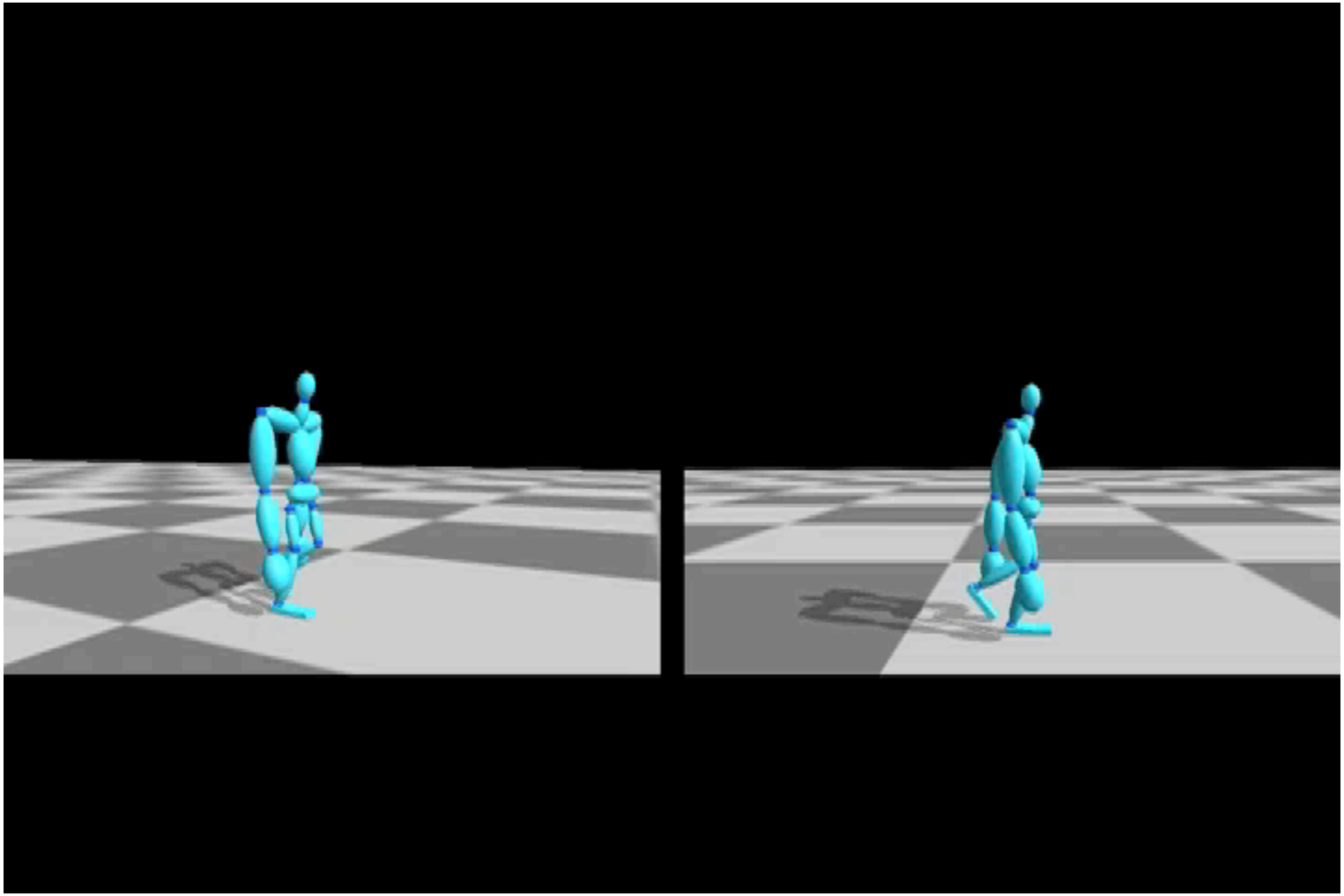
MAP Pose Trajectory (half speed)

Sloped surface (~10º)



Synthetic rendering of MAP Pose Trajectory (half speed)

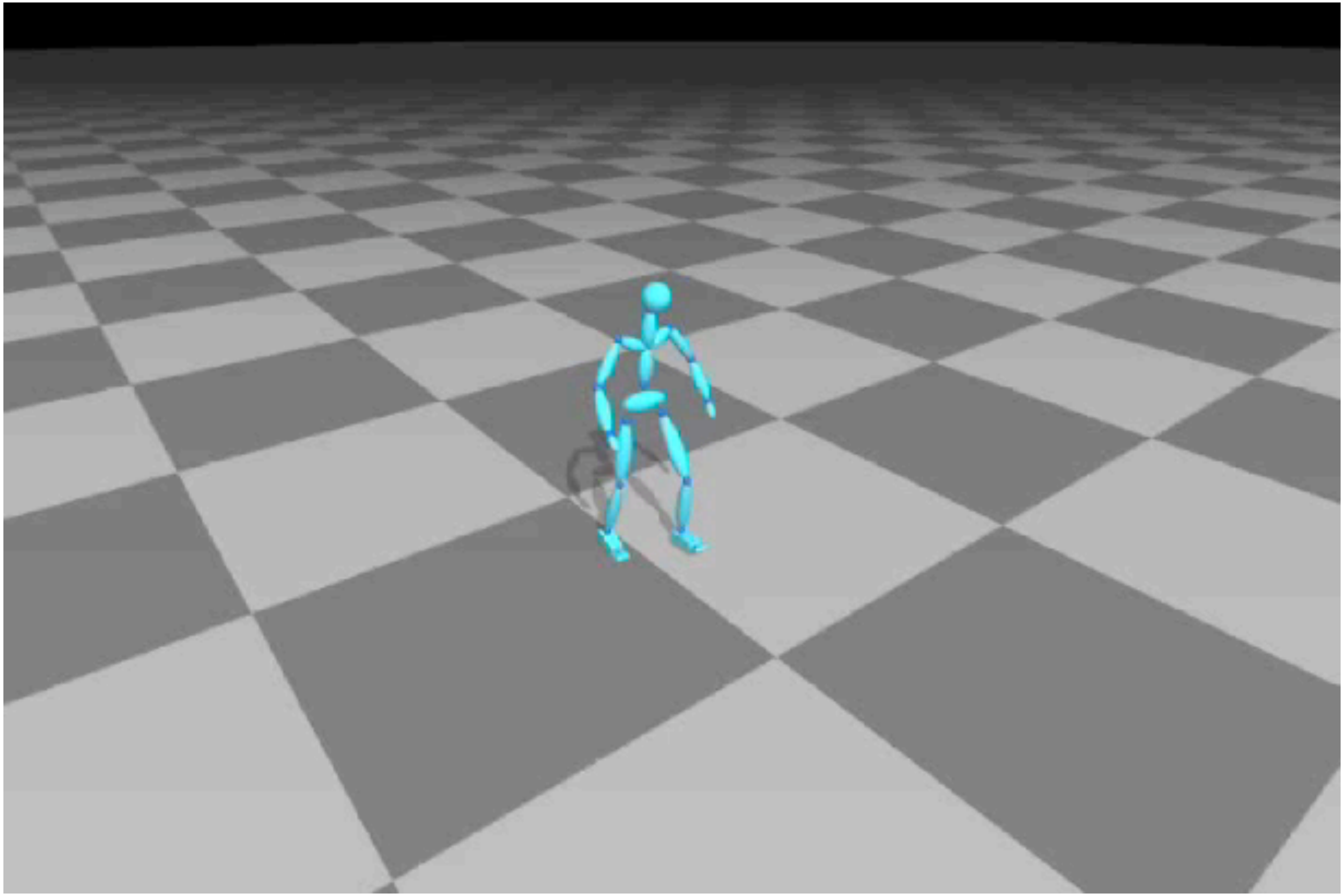
Control of 3D full-body dynamics



[Wang et al, SIGGRAPH Asia 2009]

Control under uncertainty

Optimization under deterministic conditions



Control under uncertainty

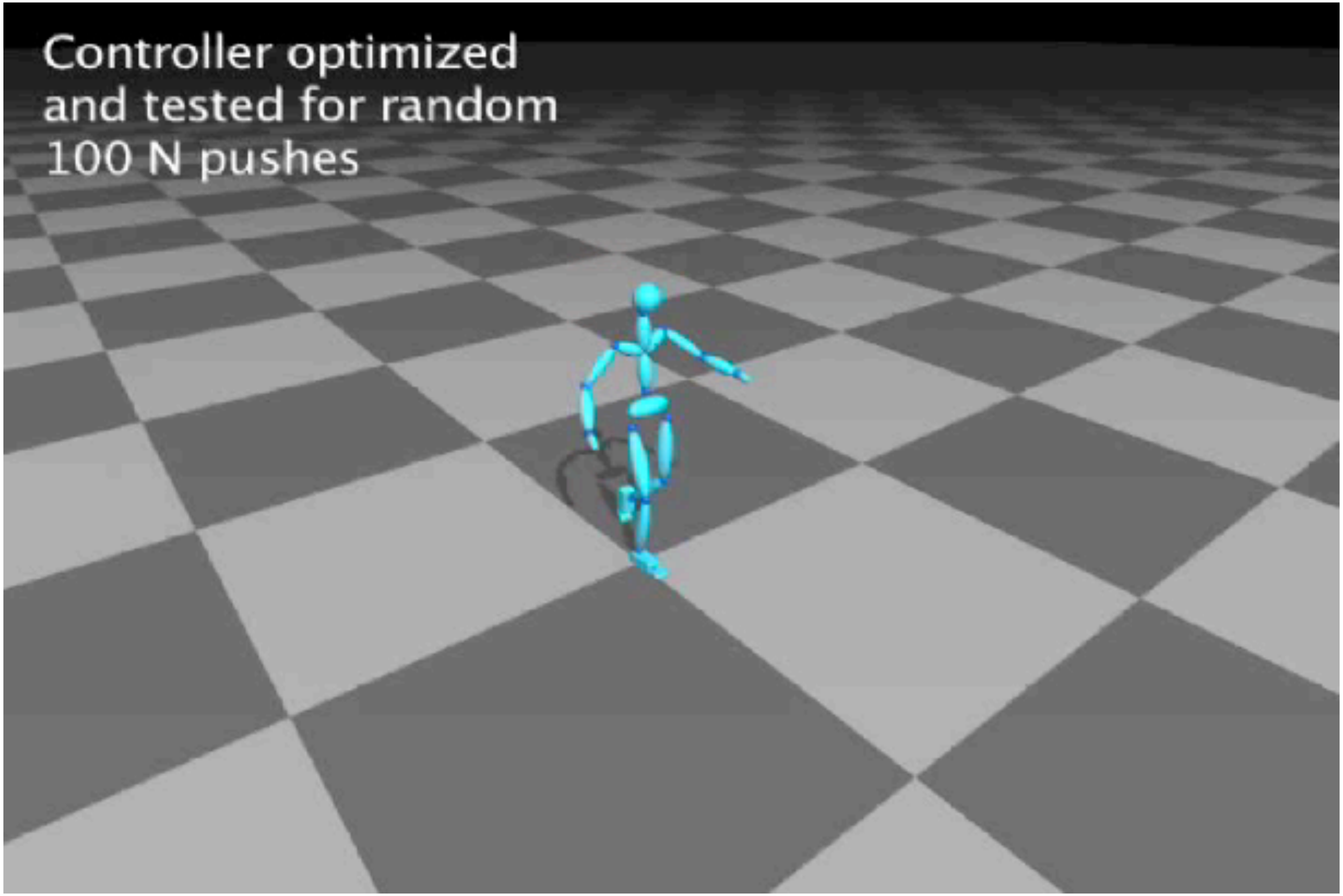
Controllers must account for uncertainty:

- signal dependent neural motor noise
- perceptual and proprioceptive uncertainty
- external disturbances
- user inputs in interactive animation

Controller design (optimization) under uncertainty:

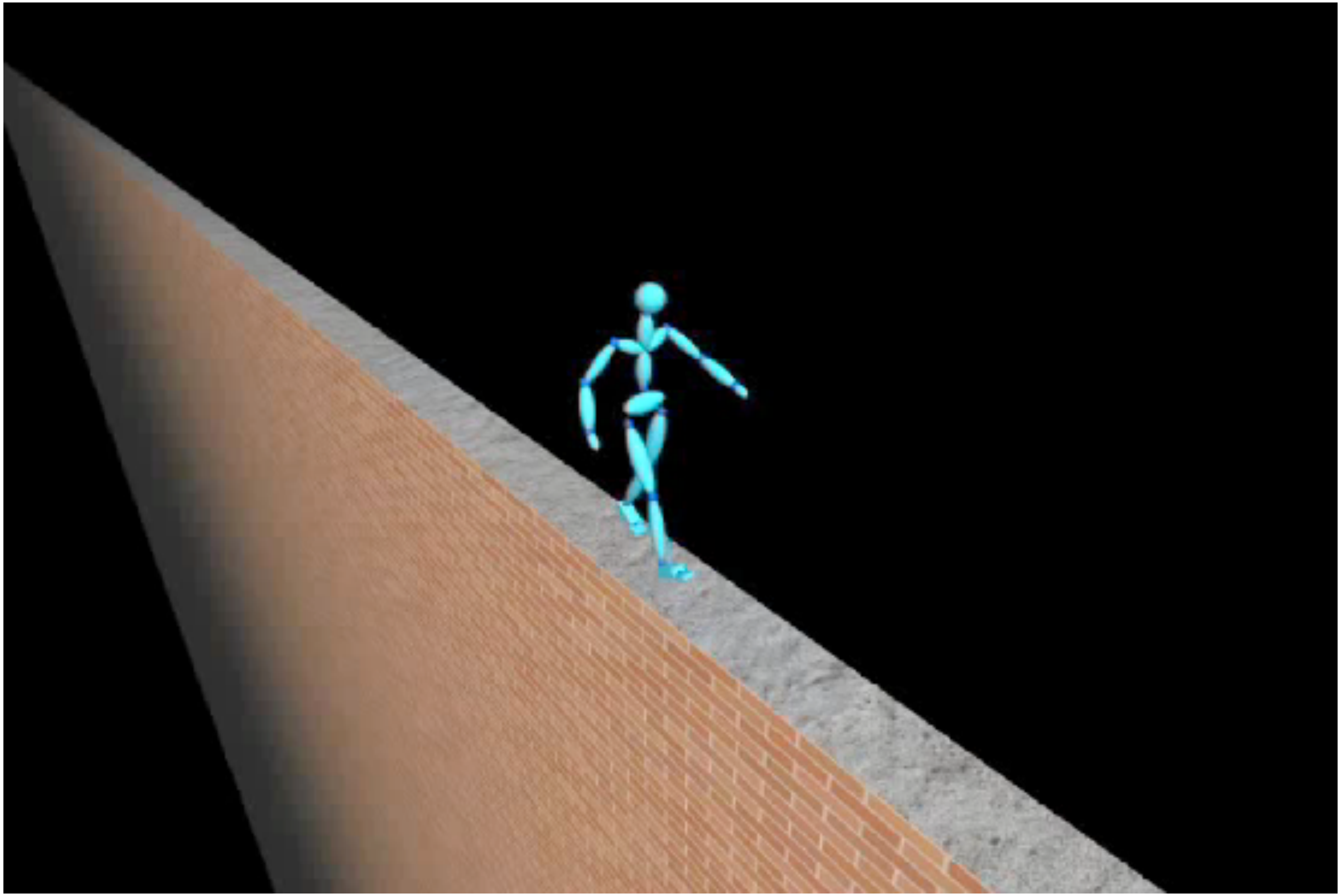
- robustness
- style adaptation to environmental constraints and noise
- ease in controller composition

Robustness to external disturbances



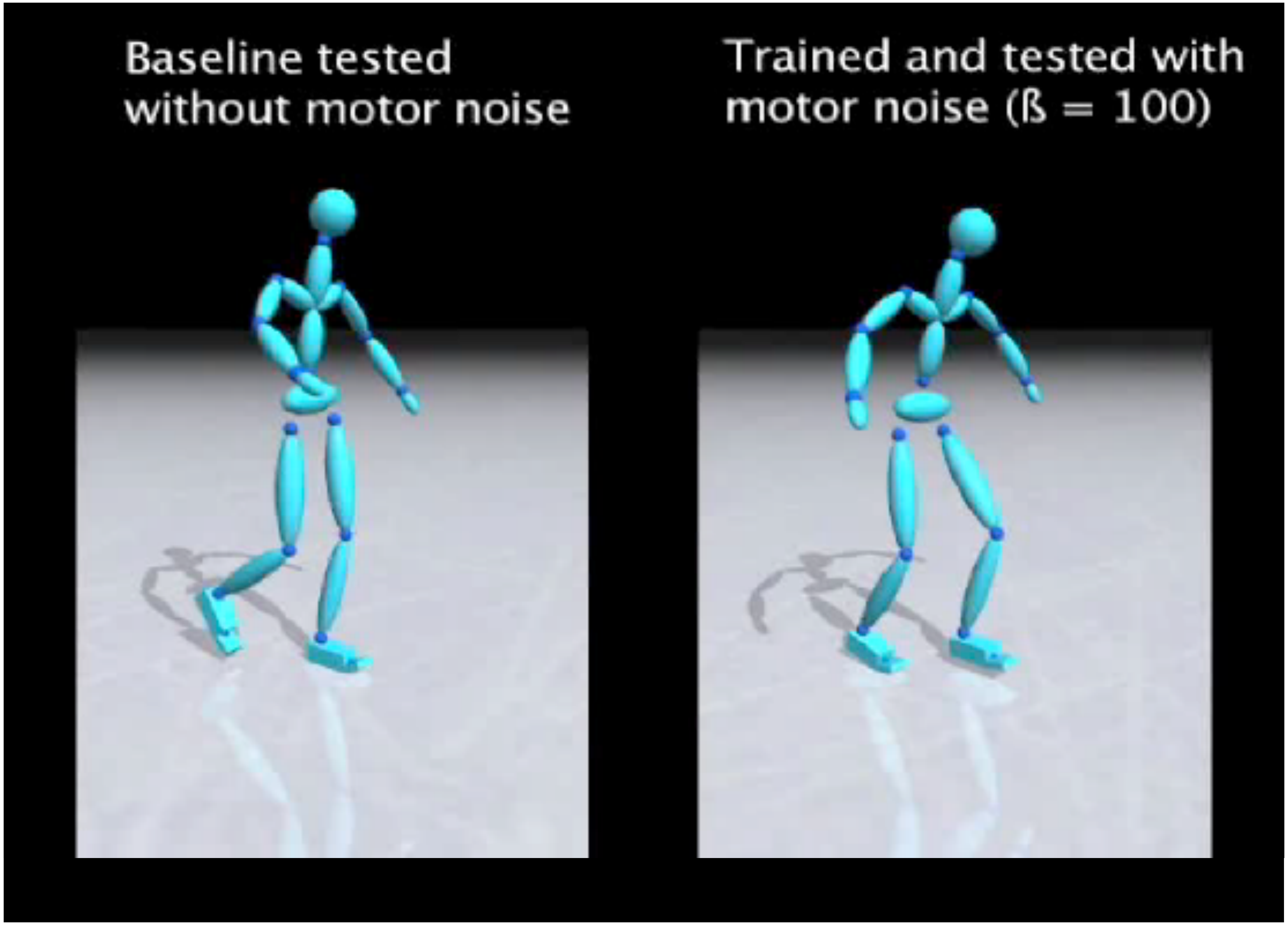
[Wang et al, SIGGRAPH 2010]

Walking on narrow beam with disturbances



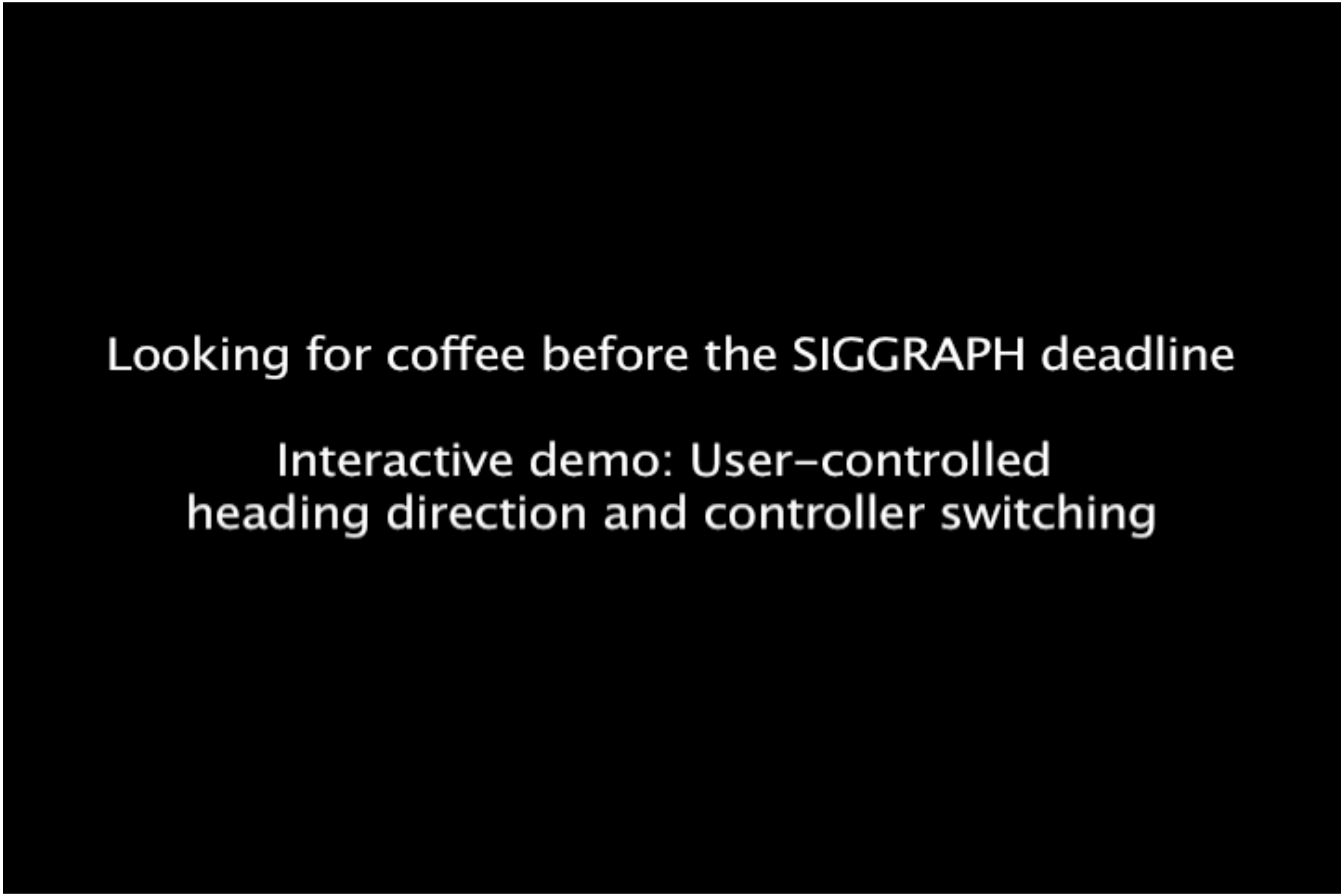
[Wang et al, SIGGRAPH 2010]

Walking on ice with motor noise



[Wang et al, SIGGRAPH 2010]

Interactive control



[Wang et al, SIGGRAPH 2010]

Estimating Contact Dynamics

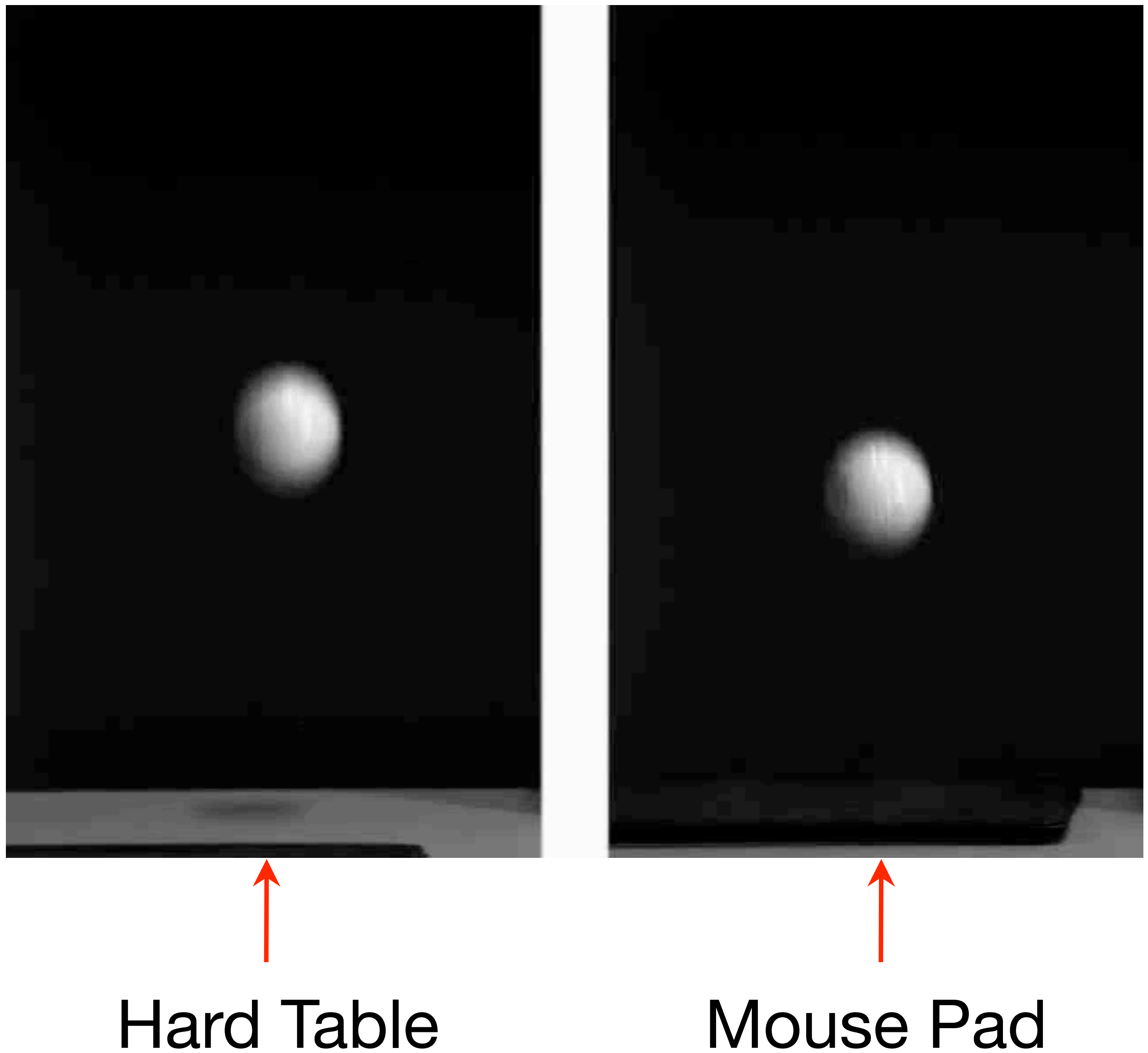
How can we infer the forces acting on a body from motion?

Contact dynamics

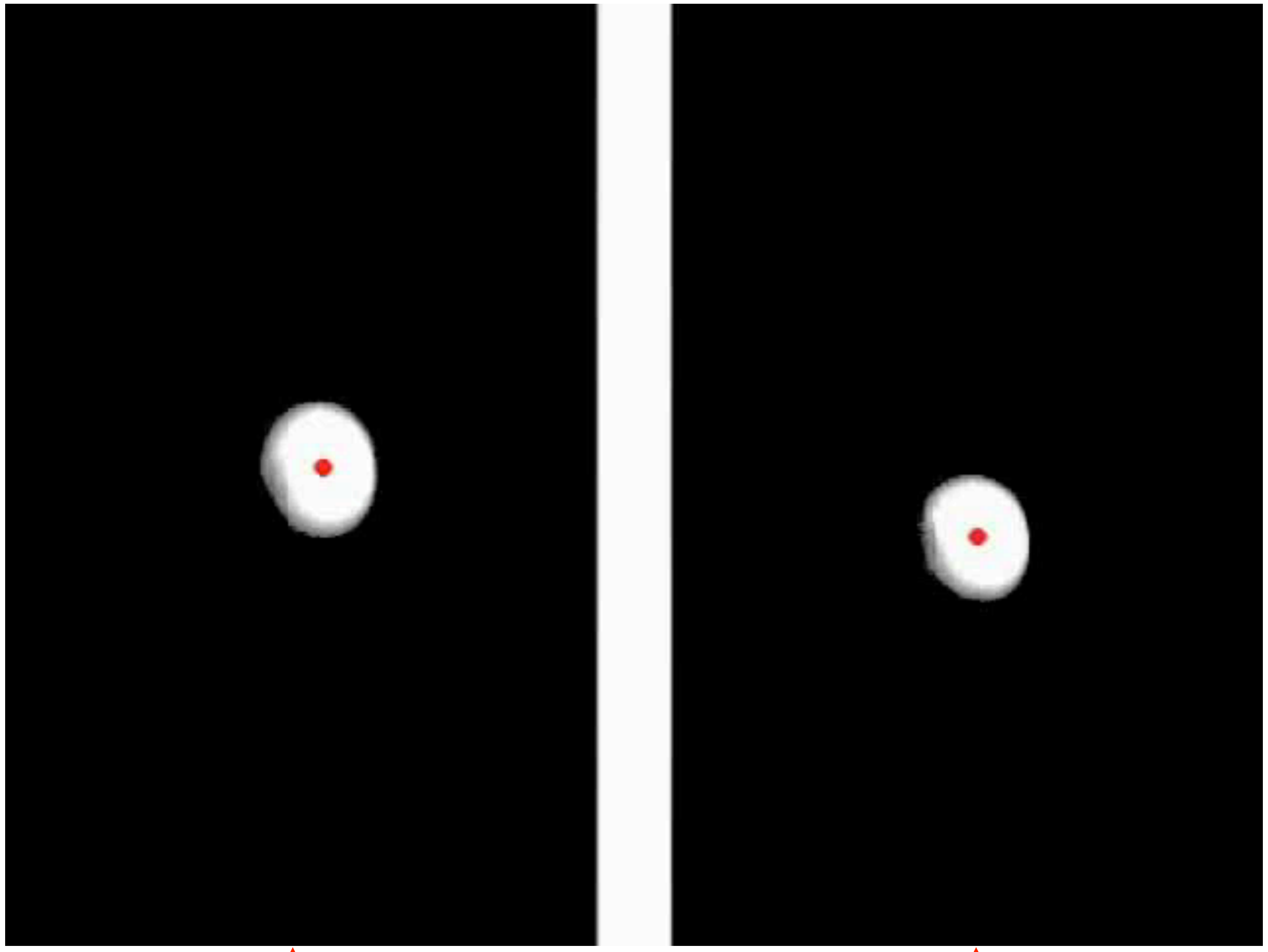
How can we infer, from motion, the internal and external forces acting on the body, in terms of

- the geometry and timing of surface contact?
- the dynamics of contact?
- the internal joint torques of the body?

Bouncing ball



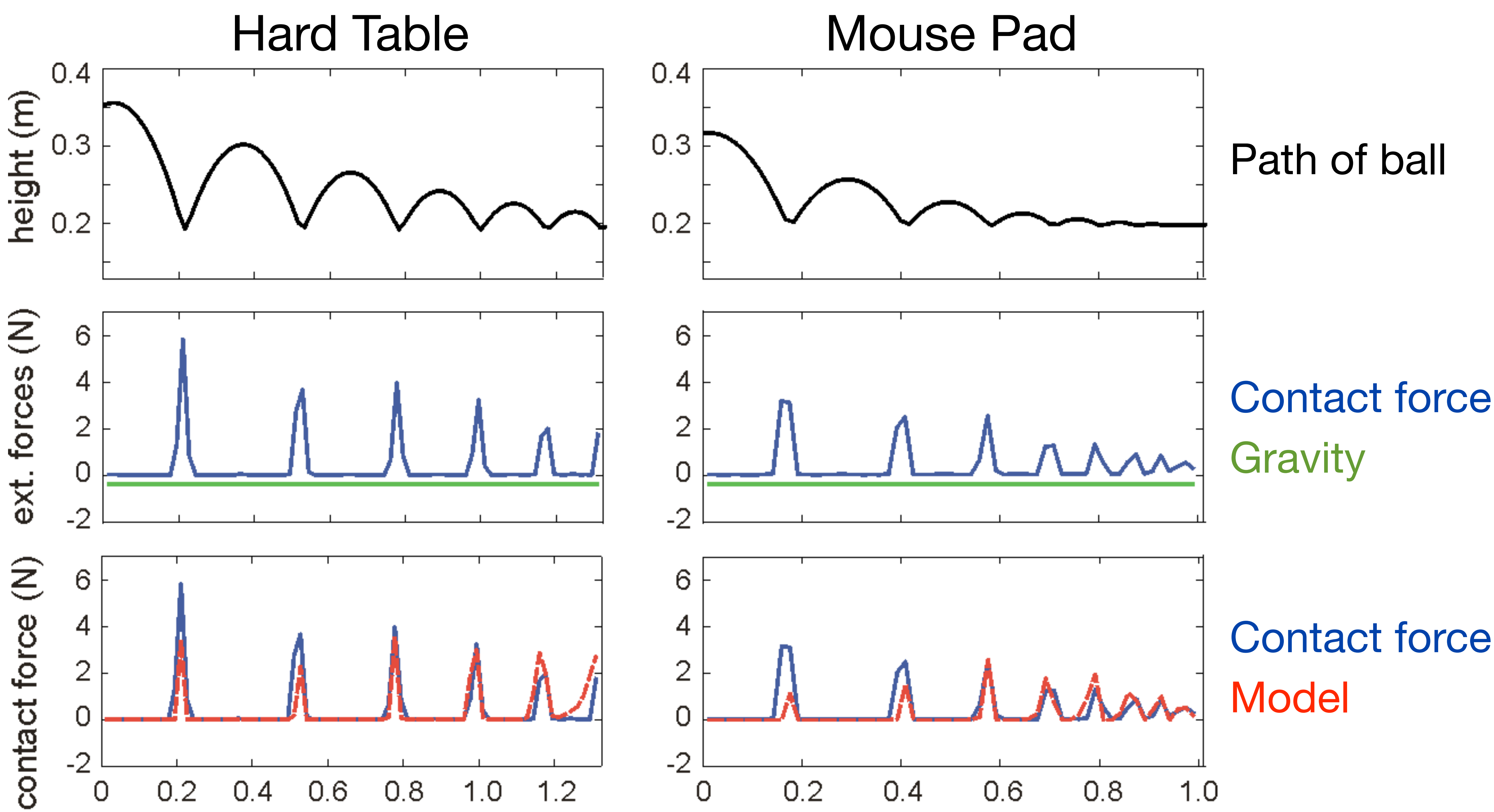
Bouncing ball



Hard Table

Mouse Pad

Bouncing ball



24 N/m surface stiffness 15 N/m

People and surface contact

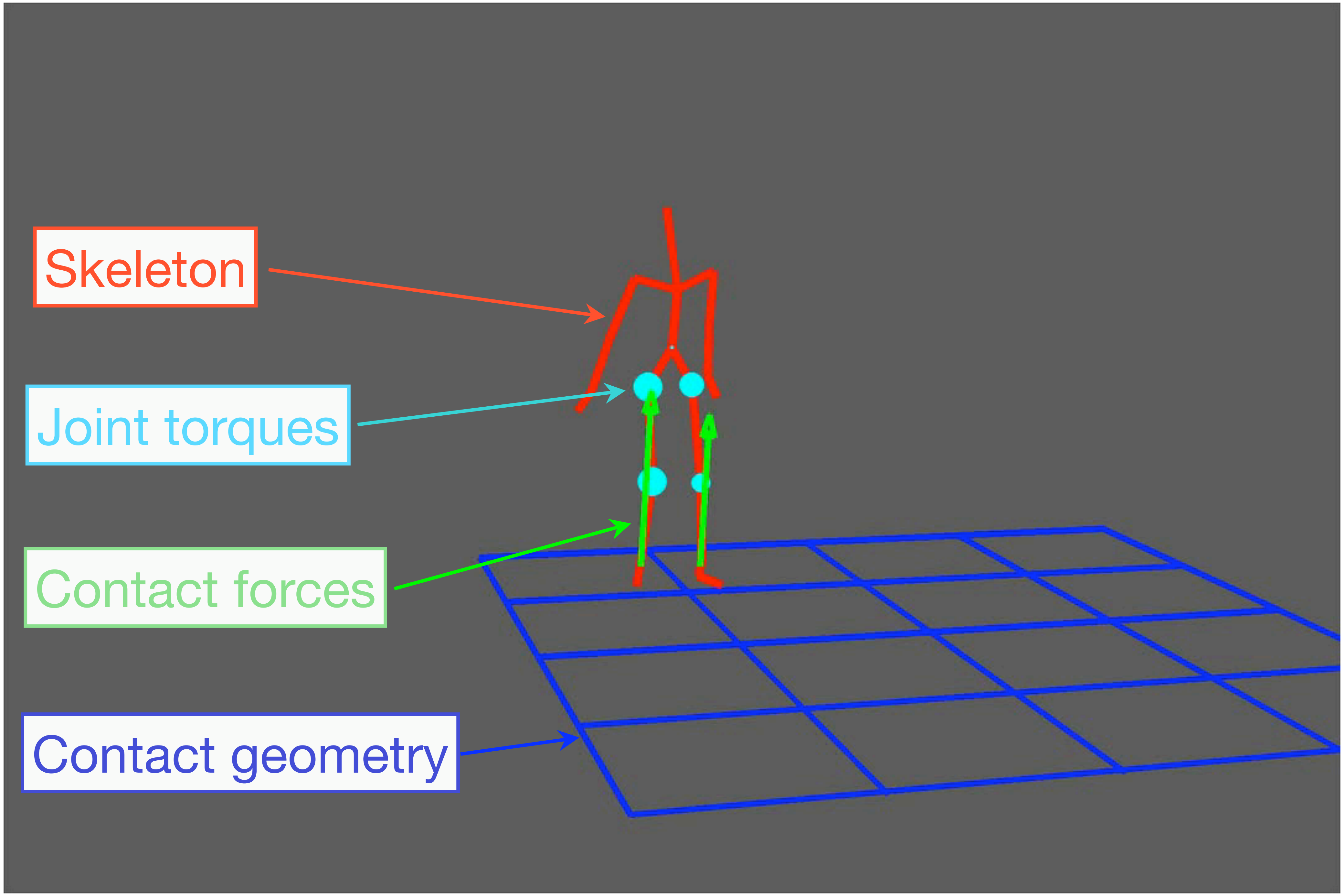
People are more complicated than the ball …

- high-dimensional articulated system
- internal and external forces
- multiple points of contacts
- ...

But the principle is essentially the same

- laws of physics are used to relate forces to state (i.e., articulated pose) and its time derivatives
- forces acting on the body are explained in terms of joint torques, gravity, and a surface contact model

Estimating contact dynamics



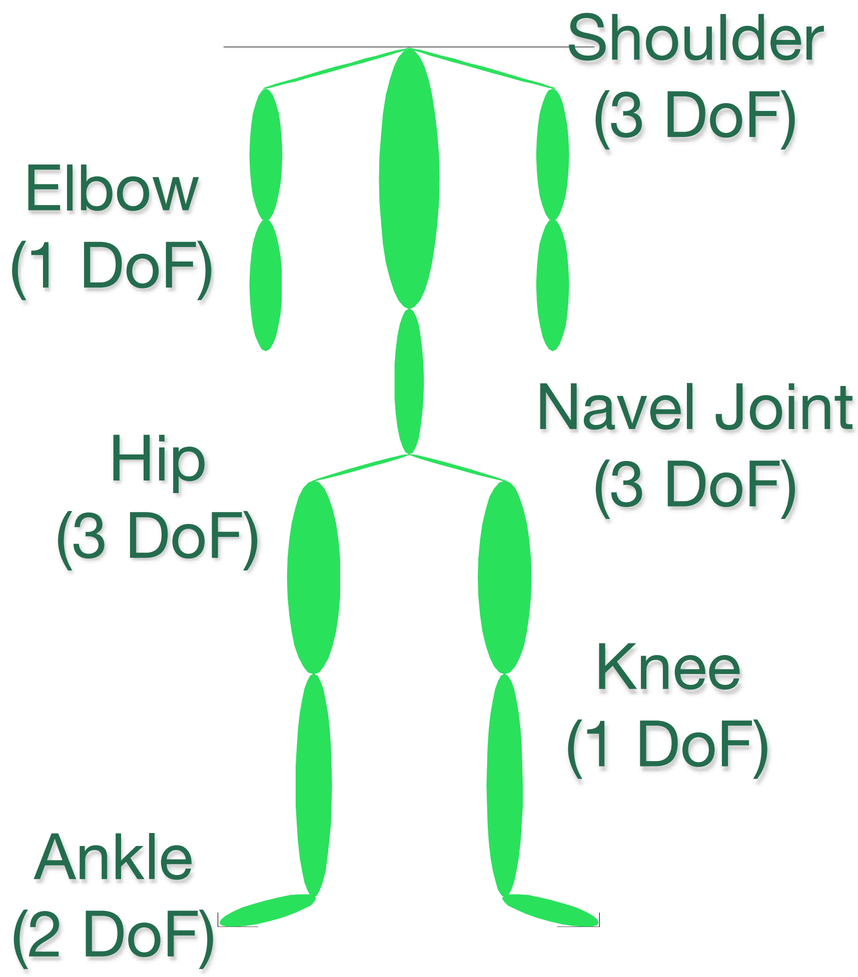
[Brubaker, Sigal and Fleet `09]

Articulated model

Model comprises 12 rigid parts and with 11 joints:

- 23 joint angle DoFs
- 6 DoFs for root position and orientation

Pose specified by generalized coordinates, $\mathbf{q} \in \mathbb{R}^{29}$.



Decomposition of generalized forces

Equations of motion:

$$\mathcal{M}(\mathbf{q})\,\ddot{\mathbf{q}} = \mathcal{F}(\mathbf{q},\dot{\mathbf{q}}) + \mathcal{A}(\mathbf{q},\dot{\mathbf{q}})$$

$$= A_{int}\,\tau_{int} + \tau_{ext}(\mathbf{q},\dot{\mathbf{q}}) + \mathcal{A}(\mathbf{q},\dot{\mathbf{q}})$$

generalized mass matrix

generalized accelerations (29D)

generalized forces

convective acceleration and other terms

internal joint torques (23D)

net external forces

External generalized forces modeled naturally in terms of forces/torques on individual parts of the articulated body

$$\tau_{ext}(\mathbf{q},\dot{\mathbf{q}}) = F(\mathbf{q})\left[\,\mathbf{f}_g + \mathbf{f}_?(\mathbf{q},\dot{\mathbf{q}})\,\right]$$

mapping to generalized forces

forces/torques on parts due to gravity and other external factors

How should we model the unexplained external forces?

Optimization

Explain as much of the observed accelerations as possible with internal joint torques and a fixed number of scene surfaces.

That is, minimize the residual accelerations for which *ficticious root forces* are necessary:

$$\min_{\theta, \mathbf{f}_{root}, \tau_{int}} \sum_t || \, \mathbf{f}_{root}(t, \theta) \, ||^2$$

$$\text{s.t.} \quad \mathcal{M} \, \ddot{\mathbf{q}} \; = \; \mathcal{F}(\tau_{int}, \theta, \mathbf{f}_{root}) + \mathcal{A}$$

internal joint torques

surface params to determine contact forces

root forces

Simple contact model

Contact locations:

- contact points on the articulated model located at ends of each body segment.
- environment is a single planar surface.

Contact dynamics:

- interface is modeled with a modulated, damped spring.

Parameters:

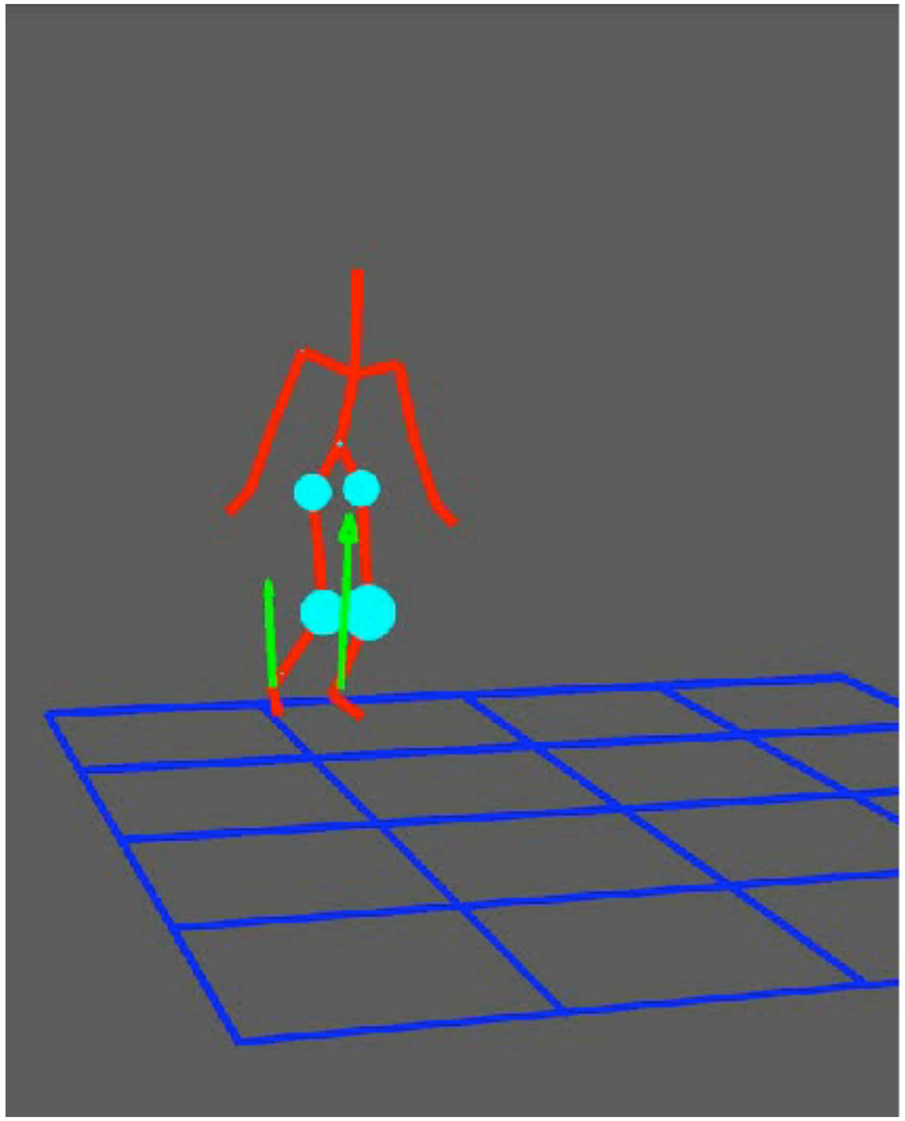
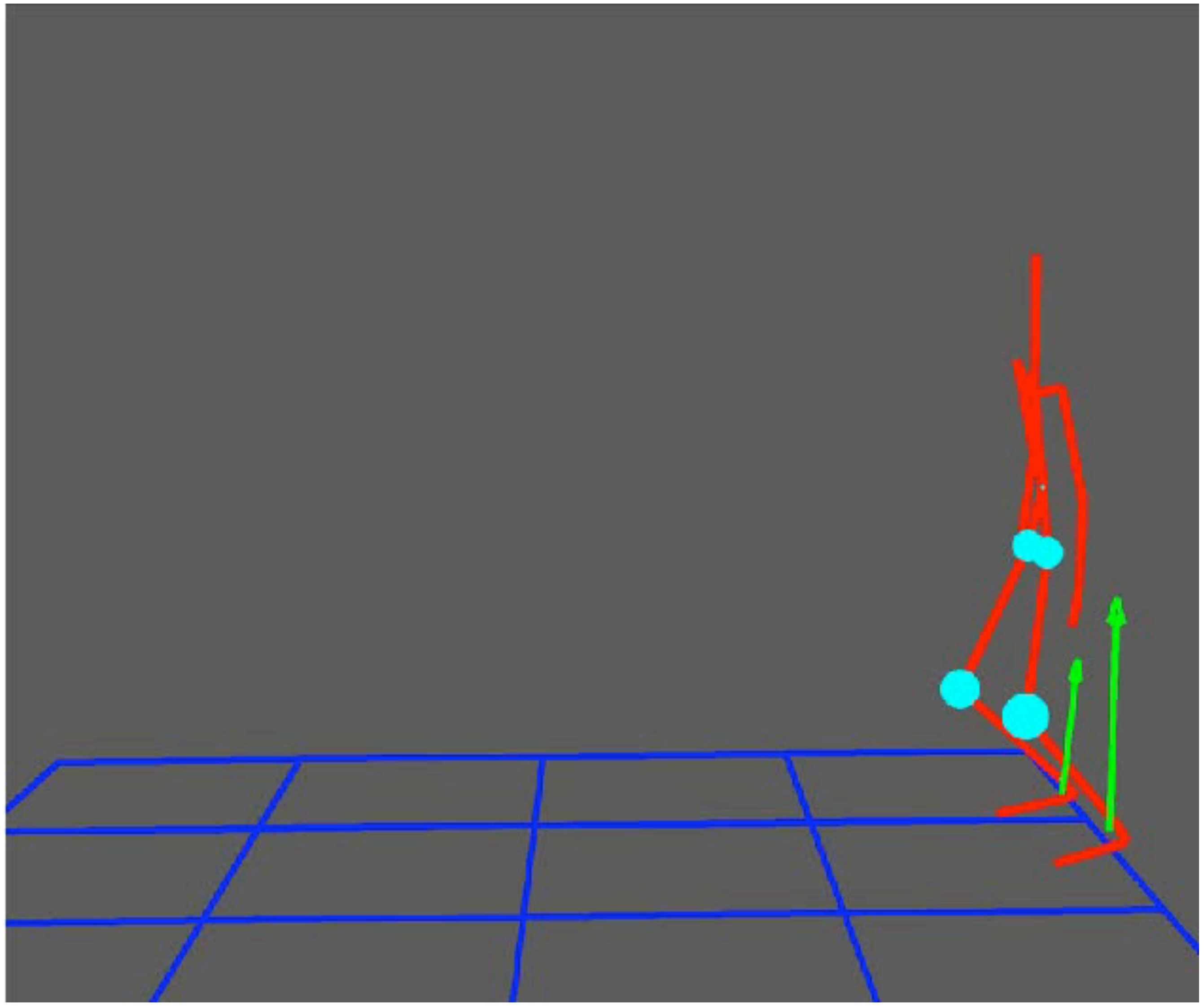
- plane orientation and position
- spring stiffness and damping coefficient normal to surface, plus tangential damping coefficient

Demos

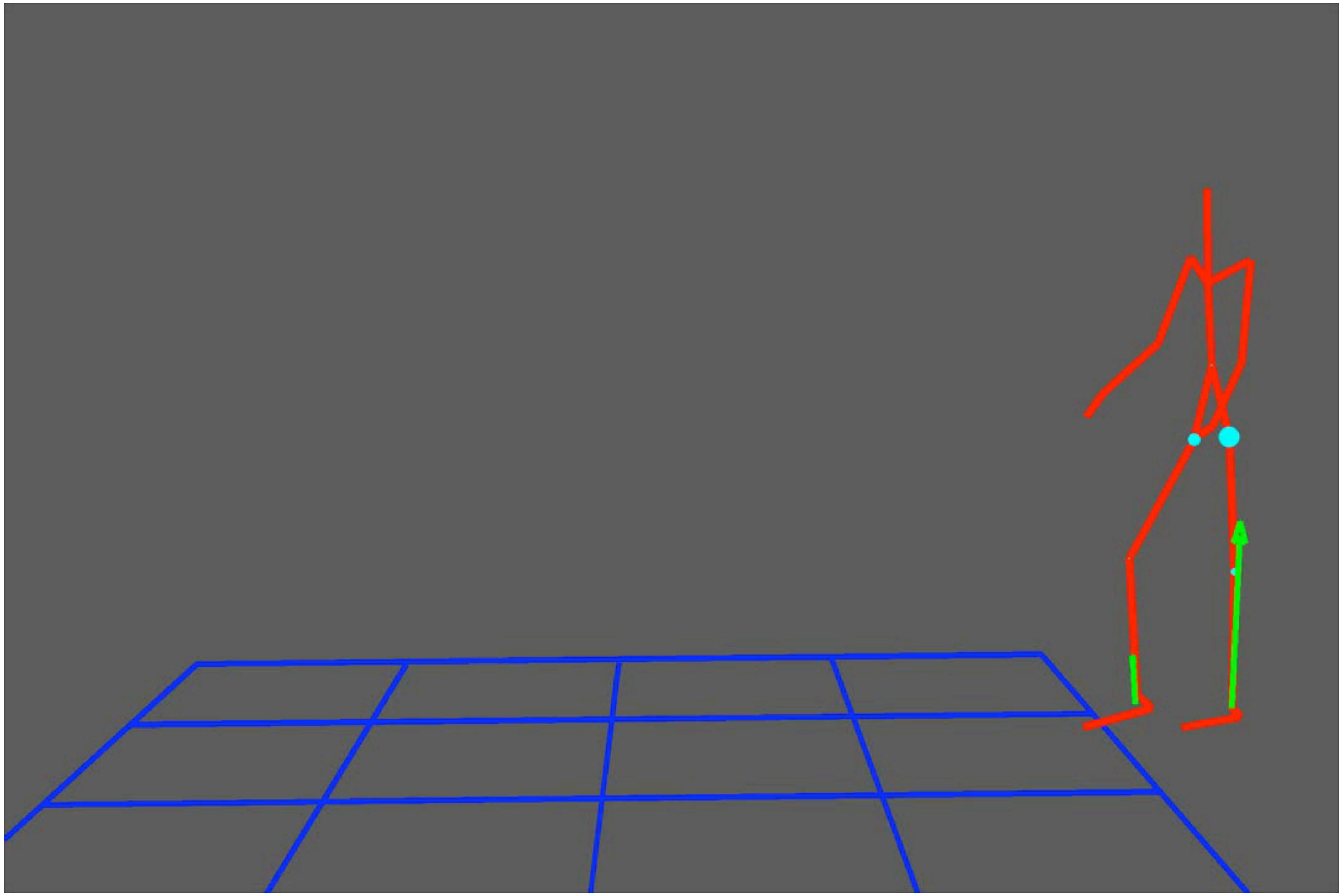
Input data

- 115 subjects, each with 2-4 samples of walking and jogging (~520 motions)
- 5 subjects with jumping, hopscotch, cartwheels, walking and jogging, all with synchronized MoCap and video (two views)

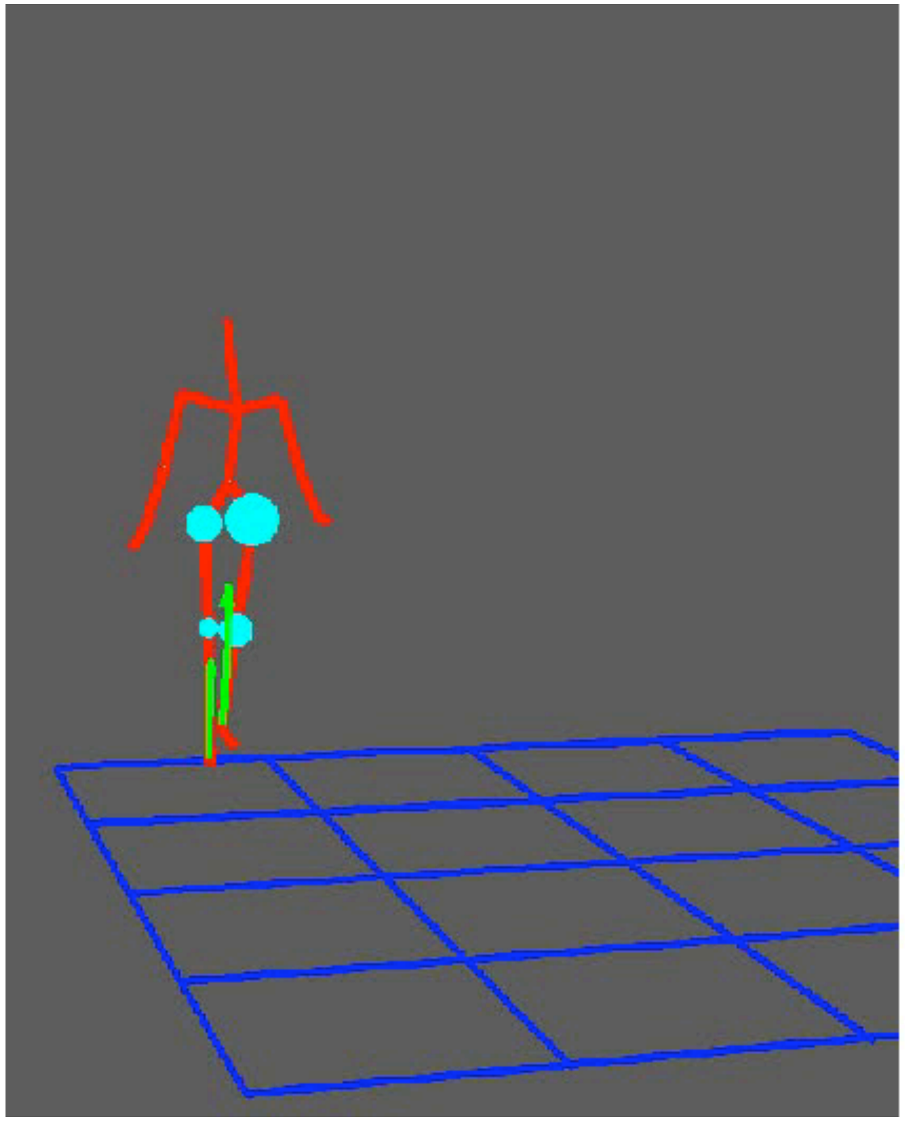
Results from mocap

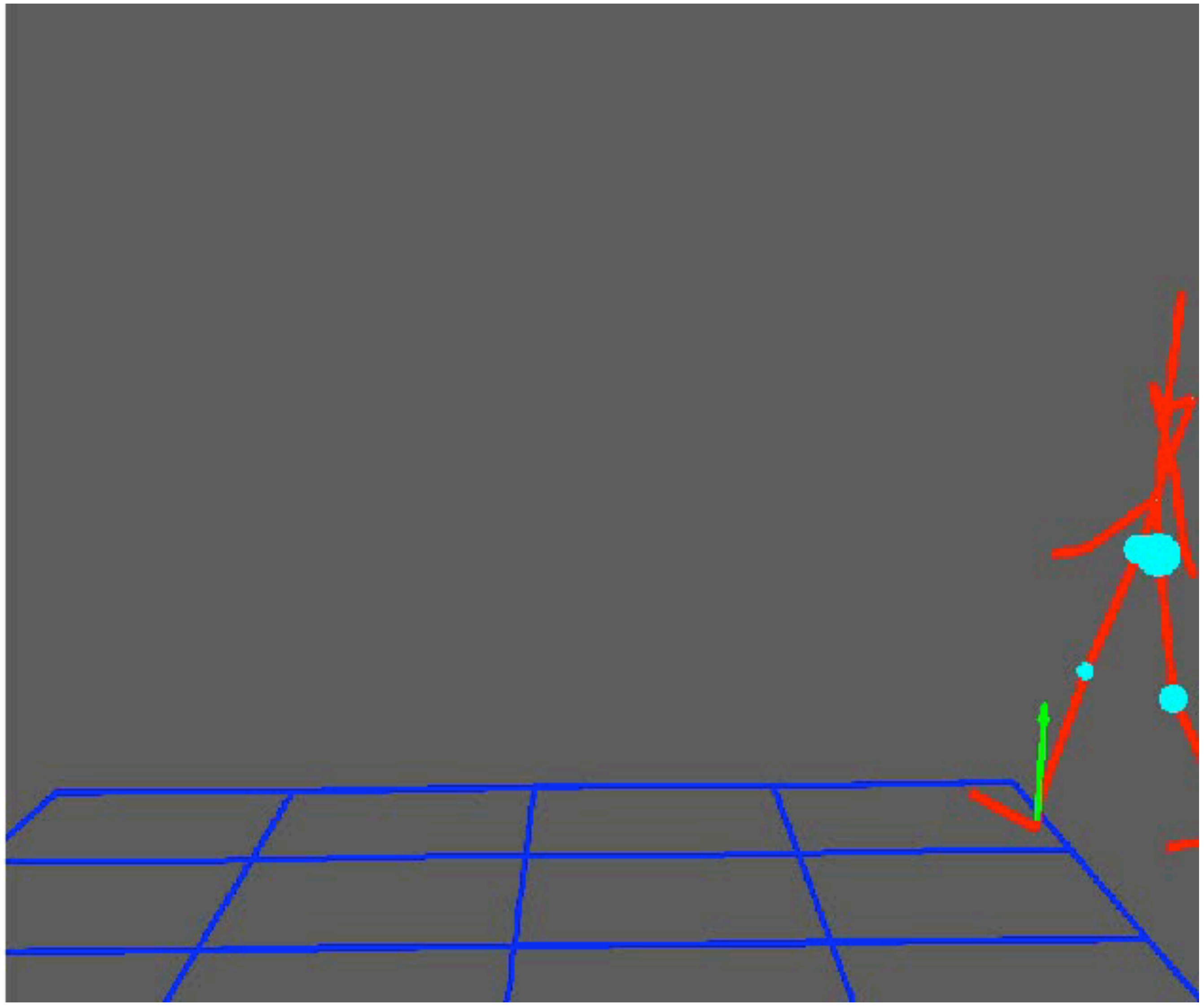



Results from mocap

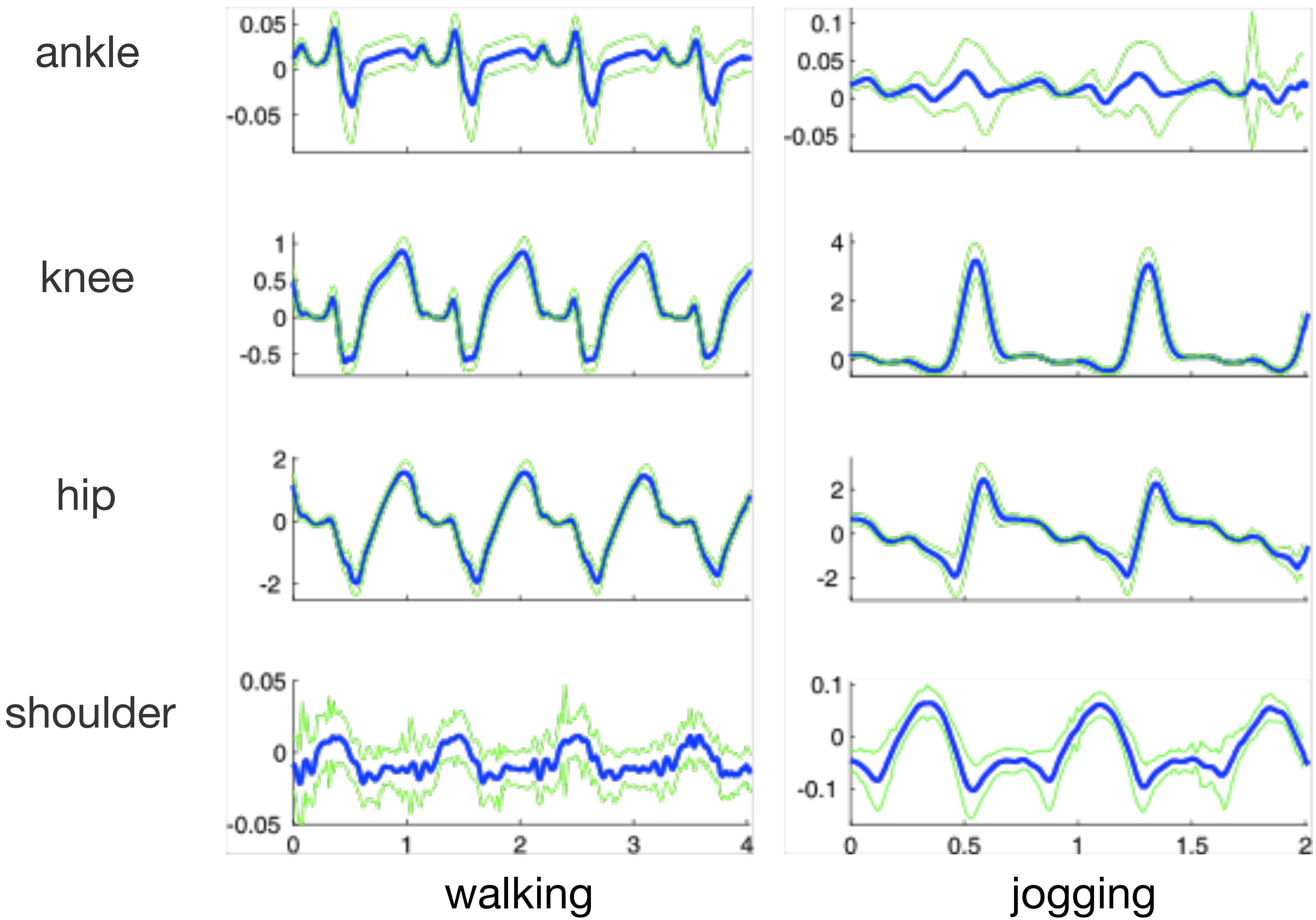


Results from mocap



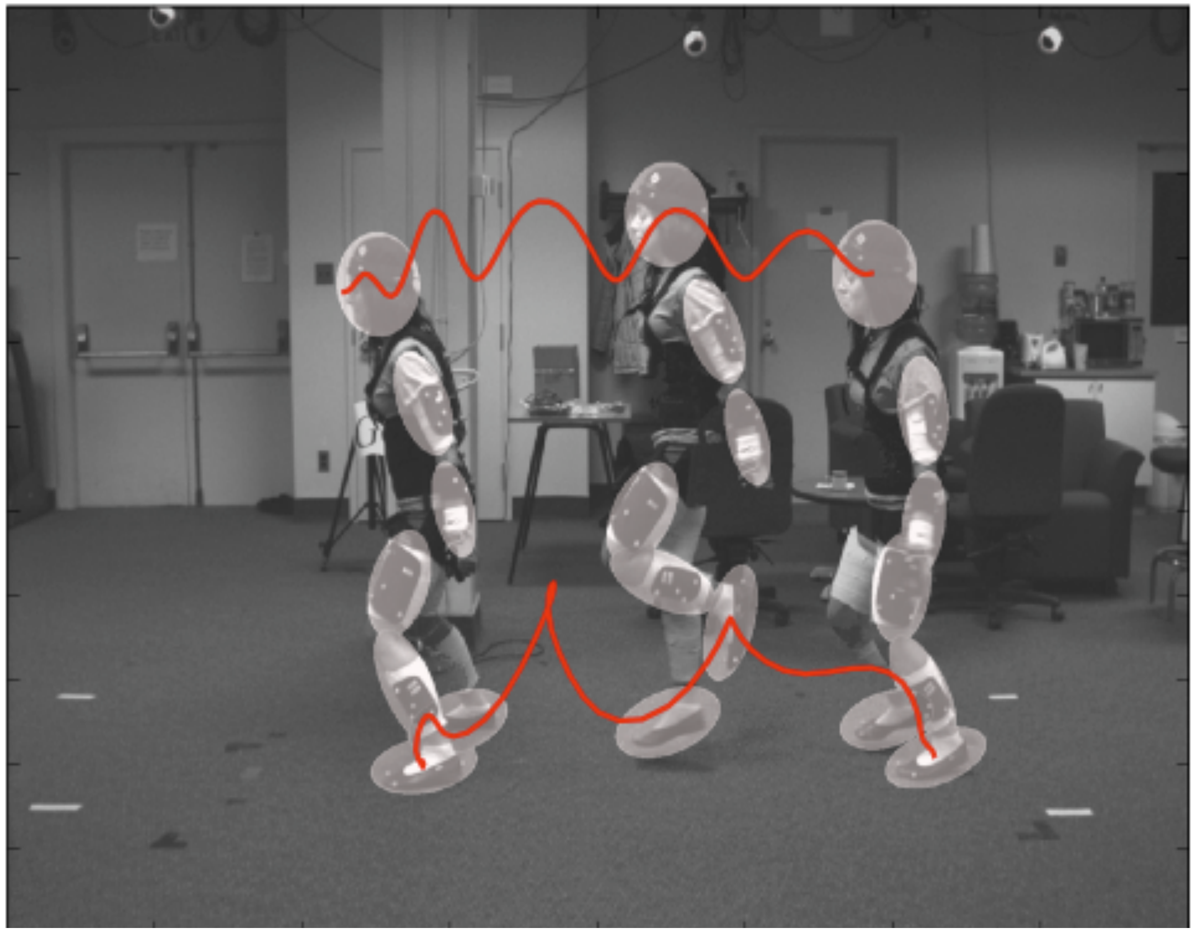


Dynamics for 115 people

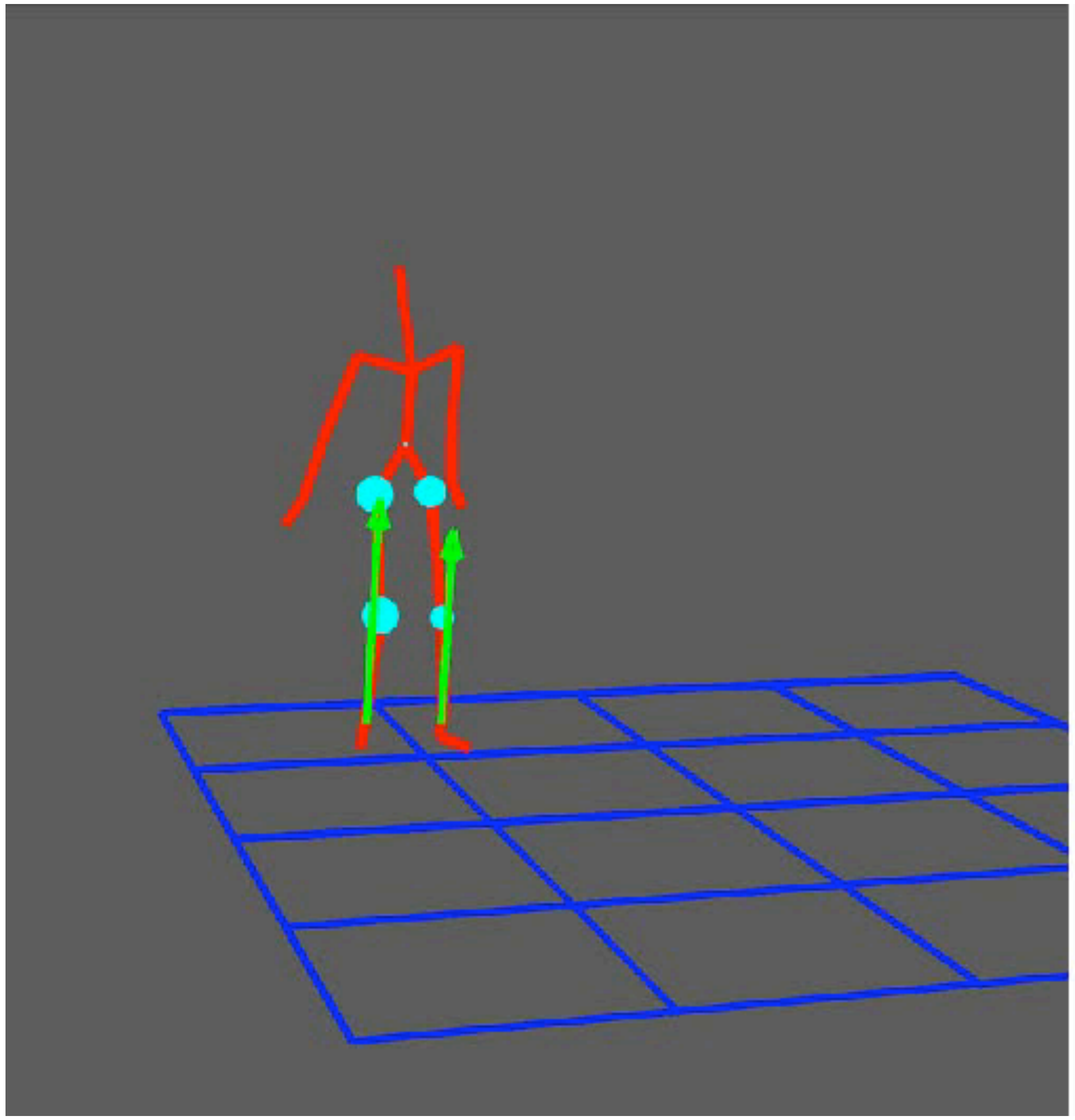


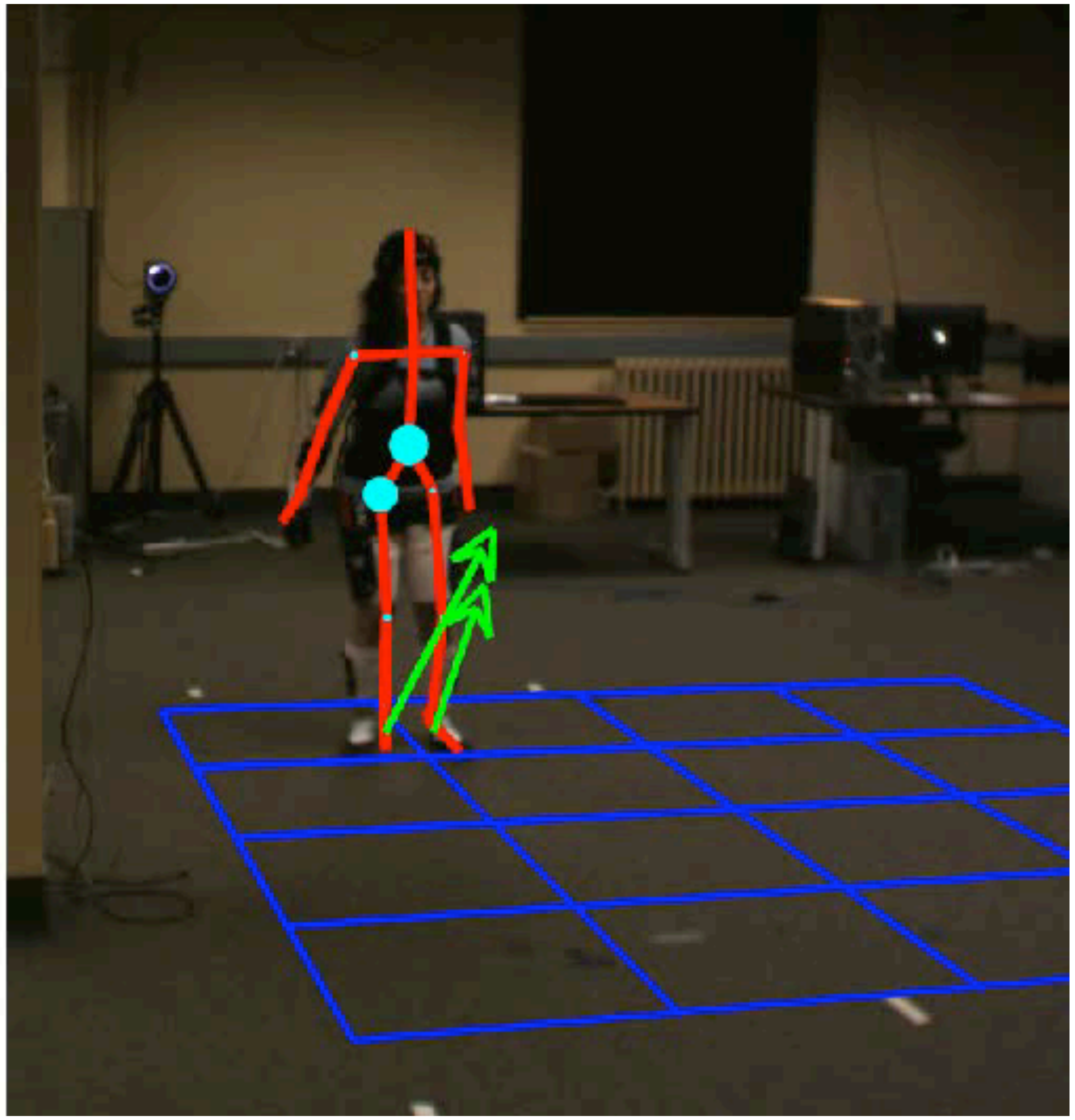
Video input

- Binocular tracking with an Annealed Particle Filter
- Prior: smoothness prior on joint accelerations
- Likelihood: background model and 2D WSL tracking



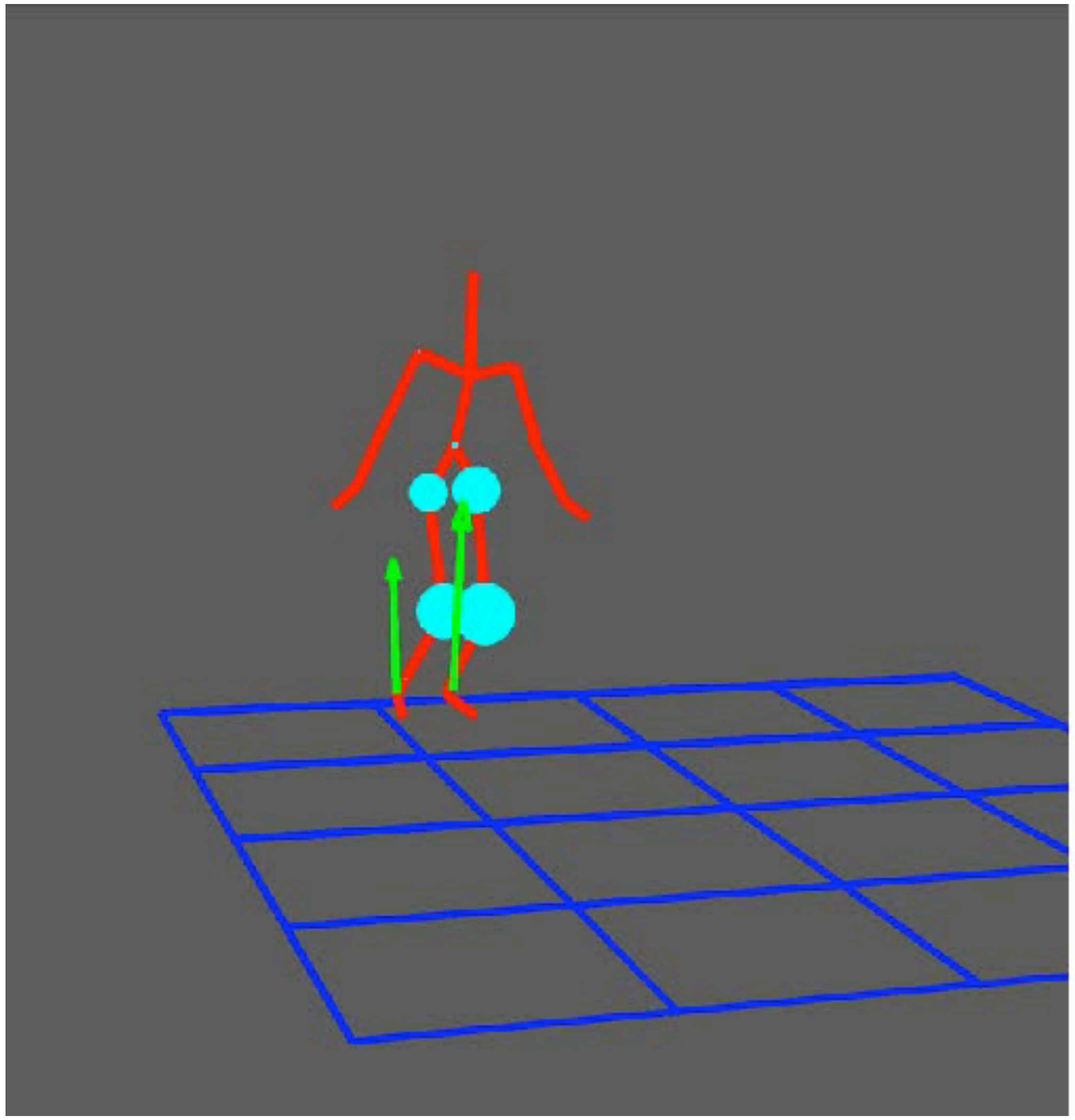
Video input

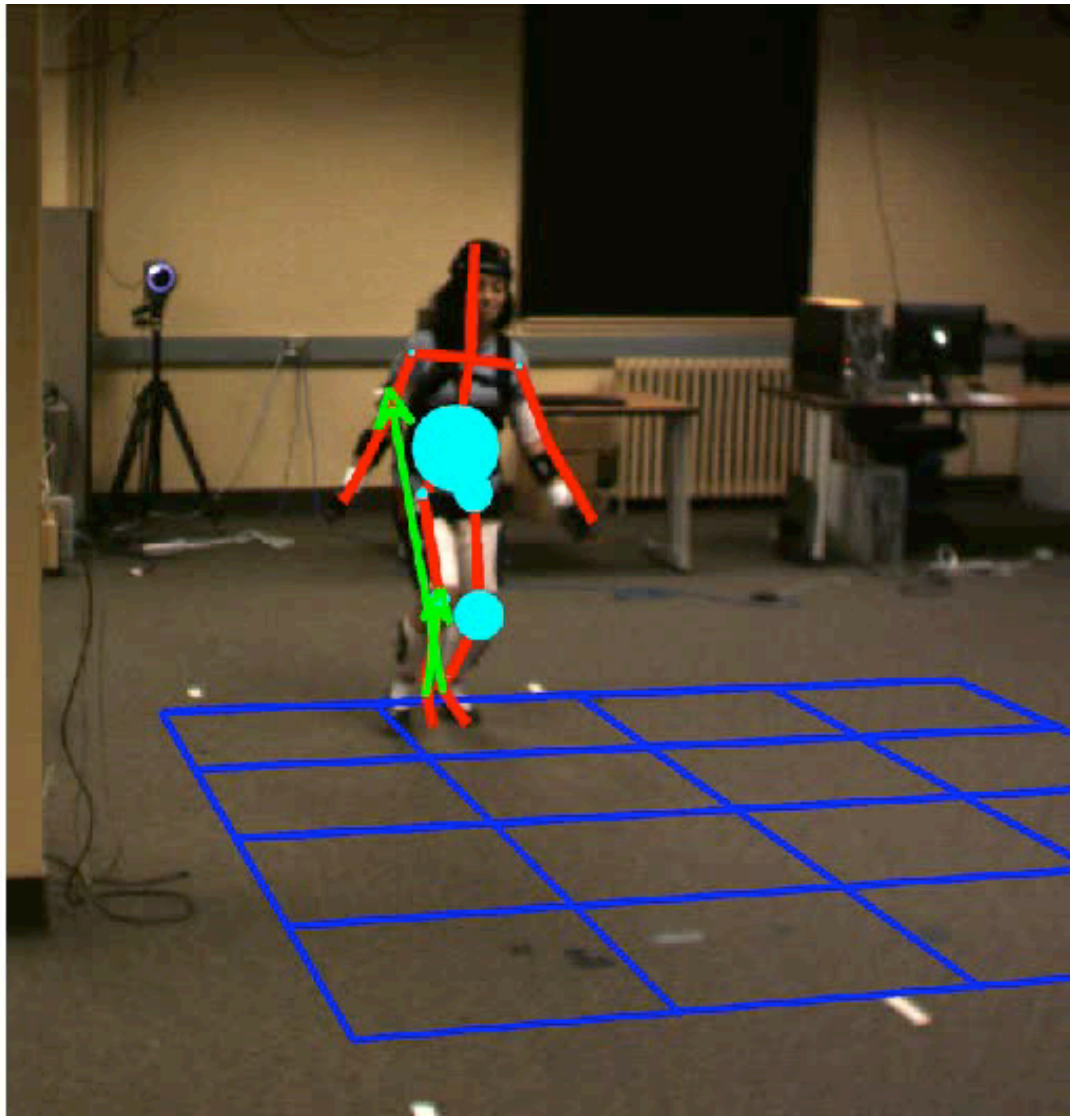




Estimates from mocap and video of same motion.

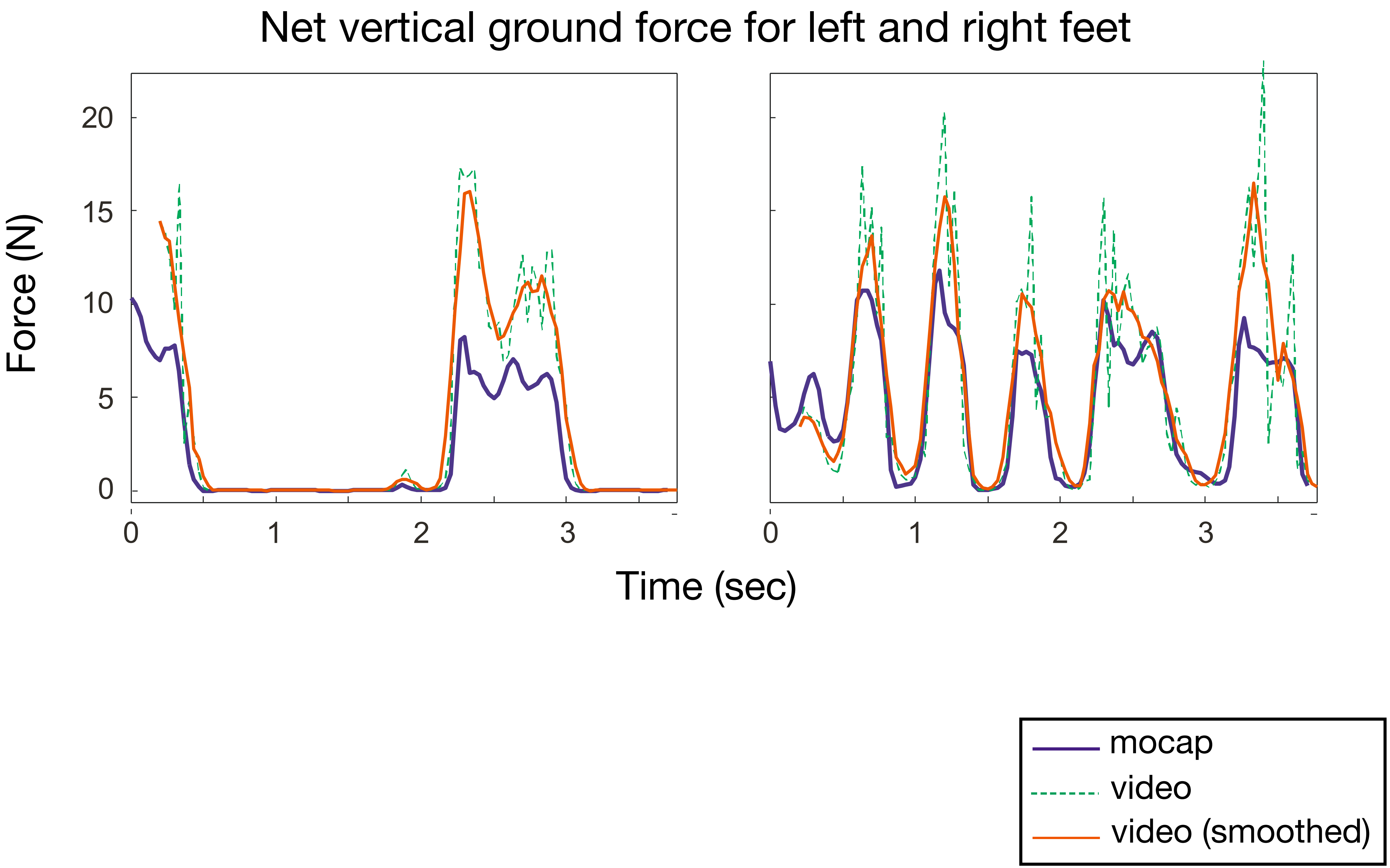
Video input



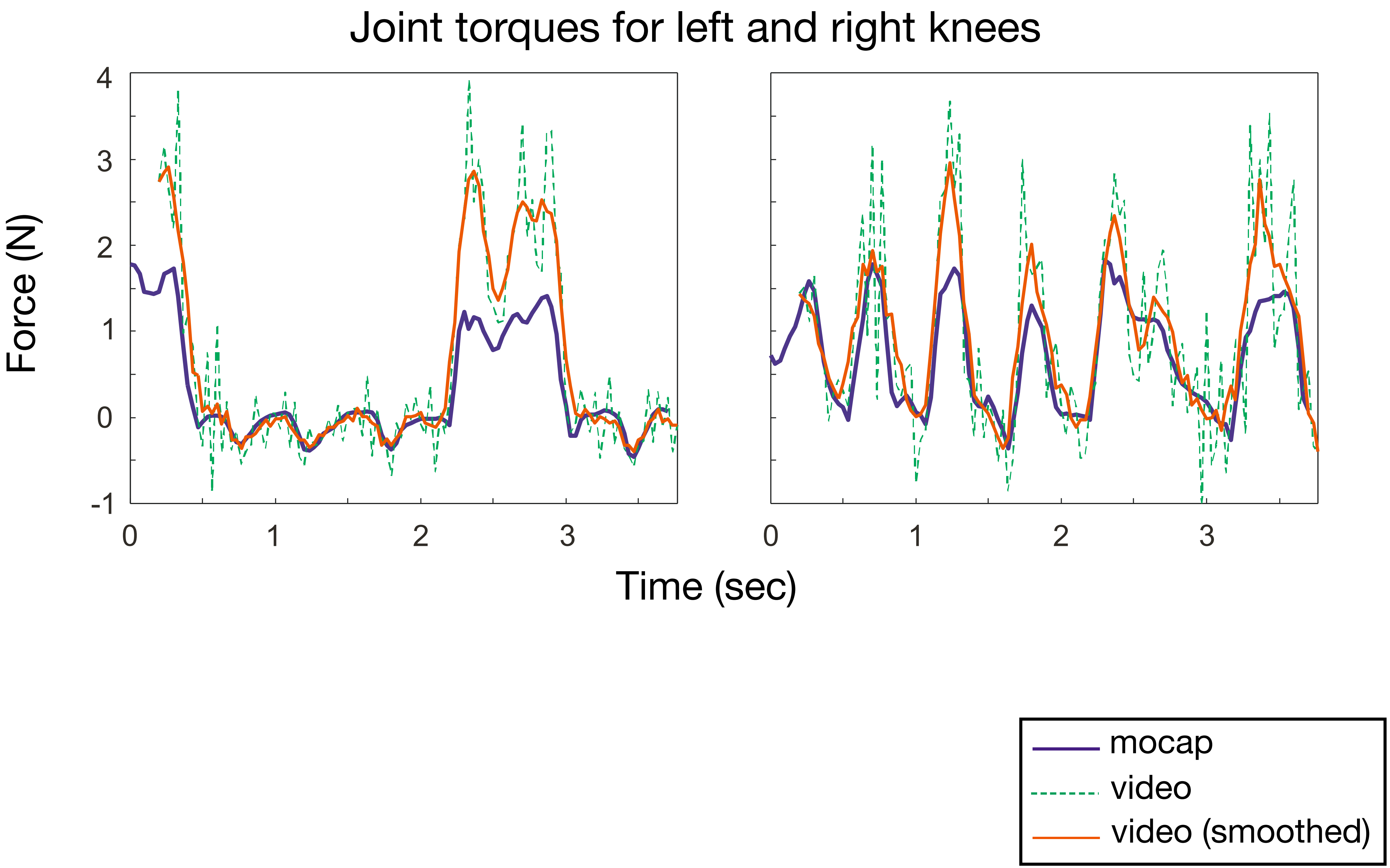


Estimates from mocap and video of same motion.

Comparison of video & mocap



Comparison of video & mocap



What’s a good model of human motion?

Selected readings for physics-based models

Brubaker M et al., Physics-based person tracking using the Anthropomorphic Walker. *IJCV*, 2010

Brubaker M and Fleet D. The Kneed Walker for human pose tracking. *Proc IEEE CVPR* 2008.

Brubaker M et al., Estimating contact dynamics. *Proc IEEE ICCV,* 2009

Collins S et al. Efficient bipedal robots based on passive-dynamic walkers. *Science* 207(5712), 2005

Kuo A. Energetics of actively powered locomotion using the simplest walking model. *J Biomech. Eng*. 124, 2002

McGeer T. Dynamics and control of bipedal locomotion. *J Theor. Biol.* 163, 1993

We need to get a lot of things right ...

Challenges:

- modeling pose and motion
- efficient search with effective proposals
- appearance
  - shape
  - reflectance
  - lighting
- understanding contact and interactions
- attribute inference
- activities ...