Human Motion Analysis

David Fleet
University of Toronto

CIFAR Summer School, 2010
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

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

\[ p(\text{motion} \mid \text{video}) = \frac{p(\text{video} \mid \text{motion}) p(\text{motion})}{p(\text{video})} \]

Discriminative

3D \ pose \ = \ E_{p(\text{pose} \mid \text{image})}[f(\text{pose})] \\ \approx \ h(\text{image measurements})
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

On-line Tracking

Video → Tracking → Pose
Model-based pose tracking

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 \((x)\)

Joint angle pose space \((y)\)

Mapping from latent positions to poses, \(g\)

Latent dynamical model, \(f\)

Density function over pose and motion (latent trajectories)
Latent variable models

Linear dynamical systems:

\[ \mathbf{x}_t = A \mathbf{x}_{t-1} + \mathbf{n}_{x,t} \]
\[ \mathbf{y}_t = B \mathbf{x}_t + \mathbf{n}_{y,t} \]
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.
Output $y$ is modeled as a function of input $x$:

$$y = g(x) = \sum_j w_j \phi_j(x) = w^T \Phi(x)$$

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

$$k(x, x') \equiv E[yy'] = \Phi(x)^T \Phi(x')$$

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

Linear: $k(x, x') = \theta x^T x'$

RBF: $k(x, x') = \theta \exp(-\frac{\gamma}{2} \|x - x'\|^2)$
Gaussian Process Latent Variable Model (GPLVM)

Joint likelihood of vector-valued data $Y = [y_1, \ldots, y_N]^T$, $y_n \in \mathcal{R}^D$, given the latent positions $X = [x_1, \ldots, x_N]^T$:

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

where $Y_d$ denotes the $d^{th}$ dimension of the training data, and the kernel matrix has elements $(K)_{ij} = k(x_i, x_j)$ 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 $\mathcal{M} = (Y, X)$, the distribution over the data $y_*$ conditioned on a latent position, $x_*$, is Gaussian:

$$y_* \mid x_*, \mathcal{M} \sim \mathcal{N}(m(x_*), \sigma^2(x_*) I_D)$$

where

$$m(x_*) = Y K^{-1} k(x_*)$$
$$\sigma^2(x_*) = k(x_*, x_*) - k(x_*)^T K^{-1} k(x_*)$$
$$k(x_*) = [k(x_*, x_1), \ldots, k(x_*, x_N)]^T$$
Gaussian Process Latent Variable Model

\[ \log \text{variance} - D \ln \sigma_y^2 | x \]

mean pose \( m(x) \)
**Conditional (predictive) distribution**

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

$$L(x_*, y_*; \mathcal{M}) = \frac{\|y_* - m(x_*)\|^2}{2\sigma^2(x_*)} + \frac{D}{2} \ln \sigma^2(x_*)$$
Gaussian Process Dynamical Model (GPDM)

Latent dynamical model [Wang et al 05]:

\[
\begin{align*}
x_t &= f(x_{t-1}; A) + n_{x,t} \\
y_t &= g(x_t; B) + n_{y,t}
\end{align*}
\]

Assume IID Gaussian noise, and

\[
\begin{align*}
f(x; A) &= \sum_i a_i \phi_i(x) \\
g(x; B) &= \sum_j b_j \psi_j(x)
\end{align*}
\]

with Gaussian priors on \( A \equiv \{a_i\} \) and \( B \equiv \{b_j\} \)

Marginalize out \( \{a_i, b_j\} \), and then optimize the latent positions, \( \{x, \ldots, x_N\} \), to simultaneously minimize pose reconstruction error and prediction error on training sequence \( \{y, \ldots, y_N\} \).
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:

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

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

- $\mathbf{W} \equiv \text{diag}(w_1, ..., w_D)$ scales the different pose parameters
Dynamical prior

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

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

where

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

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

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

with hyperparameters $\tilde{\alpha}$
Learning

GPDM posterior:

\[
p(Y, X, \bar{\alpha}, \bar{\beta}, W) = p(Y | X, \bar{\beta}, W) p(X | \bar{\alpha}) p(\bar{\alpha}) p(\bar{\beta})
\]

Training motions \rightarrow latent trajectories \rightarrow kernel hyperparameters

reconstruction likelihood \quad dynamics likelihood \quad priors

To estimate the latent coordinates & kernel parameters we minimize

\[
\mathcal{L} = -\ln p(X, \bar{\alpha}, \bar{\beta}, W | Y)
\]

with respect to \( X, \bar{\alpha}, \bar{\beta} \) and \( W \).
The model $\mathcal{M} \equiv (Y, X, \tilde{\alpha}, \tilde{\beta}, W)$ then provides a density function over new poses, with negative log likelihood:

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

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

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

GPDM same sample trajectories

[Urtasun et al, `06]
People tracking with GPDM

Image Observations: $I_{1:t} \equiv (I_1, ... , I_t)$

State: $\phi_t = [G_t, y_t, x_t]$

GPDM: $\mathcal{M}$

Inference: MAP estimation by gradient ascent on the posterior:

$$p(\phi_t | I_{1:t}, \mathcal{M}) \propto p(I_t | \phi_t) p(\phi_t | I_{1:t-1}, \mathcal{M})$$

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

[Urtasun et al, `06]
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

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

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 $x$:

$$y = \sum_i s_i g_i(x) + \epsilon = \sum_i s_i w_i^T \Phi(x) + \epsilon$$

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

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

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

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

$[Wang et al. ICML '07]$
Multifactor locomotion model

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

- **s**: identity of the subject performing the motion
- **g**: gait of the motion (walk, run, stride)
- **x**: current state of motion (evolves w.r.t. time)

Covariance function:

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

[Wang et al. ICML '07]
Training data

Each training motion is a sequence of poses, sharing the same combination of subject (s) and gait (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

subject 1, walk

subject 1, stride (generated)

[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)


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


Discriminative Pose Estimation
Discriminative pose estimation

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

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 level features
- 2D joint positions or full 2D pose

Best to learn the features, but this can be hard or expensive
We want to find a “mapping” from features to 3D poses; i.e., a conditional distribution \( p(3D \, pose \mid features) \)

**Problem:**

- Approximate with locally linear mappings.

**Solution:**

- Approximate with locally linear mappings.
Multi-valued regression: Mixtures of experts

\[ p(y | x) \propto \sum_{k=1}^{M} p_{g,k}(k | x, \theta_{g,k}) p_{e,k}(y | x, \theta_{e,k}) \]

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

\[ \mathbf{x} \in \mathbb{R}^n \]  

query image features

\[ \mathbf{y} \in \mathbb{R}^m \]  

3D pose space

\[ p_{g,k}(k | \mathbf{x}) \]  

gating network

\[ p_{e,k}(\mathbf{y} | \mathbf{x}) \]  

mapping

feature space

feature space

feature space
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

- **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

[Poon and Fleet, 01]
Implausible motions

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

[Urtasun et al. ICCV `05]
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**
- point-mass at hip, massless legs with prismatic joints, and impulsive toe-off force
- inverted pendular motion

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

*References*
- Blickhan & Full 1993; Srinivasan & Ruina 2000
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 $\tau_{to}$, $\tau_h$, $\tau_{k1}$, $\tau_{k2}$ (for torso, hip, & knees)
- impulse applied at toe-off (with magnitude $\iota$)
- gravitational acceleration (w.r.t. ground slope $\gamma$)

[Brubaker and Fleet `08]
The *Kneed Walker*

Joint torques are parameterized as damped linear springs.

For hip torque

\[ \tau_h = \kappa_h \left( \phi_{t_2} + \phi_{t_1} - \phi_h \right) - d_h \left( \dot{\phi}_{t_2} + \dot{\phi}_{t_1} \right) \]

with stiffness and damping coefficients, \( \kappa_h \) and \( d_h \), and resting length \( \phi_h \).
The Kneed Walker

Equations of motion

\[ \mathcal{M}(\ddot{q}) = f_s(\vec{k}, \vec{d}, \vec{\phi}) + f_g + f_c \]

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 \( \vec{x} = (\vec{q}, \vec{\dot{q}}) \) that produce cyclic locomotion at speed \( s \), step length \( \ell \), and slope \( \gamma \), with minimal “energy”.

Solve

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

where \( E(\vec{\theta}, \vec{x}; s, \ell, \gamma) \) measures the “cost” of the motion, and \( C(\vec{\theta}, \vec{x}; s, \ell, \gamma) \) 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°
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: $z_{1:t} \equiv (z_1, \ldots, z_t)$

State: $s_t = [d_t, k_t]$

Posterior distribution:
$$p(s_{1:t} | z_{1:t}) \propto p(z_t | s_t) p(s_t | s_{1:t-1}) p(s_{1:t-1} | z_{1:t-1})$$

Sequential Monte Carlo inference:

- particle set $S = \{s_{1:t}^{(j)}, w_t^{(j)}\}_{j=1}^N$ approximates $p(s_{1:t} | z_{1:t})$
- step 1. sample next state: $s_t^{(j)} \sim p(s_t | s_{t-1}^{(j)})$
- step 2. update weight: $w_t^{(j)} = w_{t-1}^{(j)} p(z_t | s_t^{(j)})$
- resample when the effective number of samples becomes small
Bayesian people tracking

Proposals for re-sampling are given by Monte Carlo approximation,
\[ Q_t = \{ s_t^{(j)}, \hat{w}_t^{(j)} \}_{j=1}^N, \]
to the windowed smoothing distribution

\[ p(s_t \mid z_{1:t+\tau}) \propto \int_{s_{t+1:t+\tau}} p(z_{t:t+\tau} \mid s_{t:t+\tau}) p(s_{t:t+\tau} \mid z_{1:t-1}) \]

Re-sample \( S_t \) when the effective sample size \( \left[ \sum_j (\hat{w}_t^{(j)})^2 \right]^{-1} \) 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:

\[ s_t^{(k)} \leftarrow s_t^{(i)} \quad w_t^{(k)} \leftarrow w_t^{(i)} / \hat{w}_t^{(i)} \]
Image observations

Foreground model
Gaussian mixture model for colors in each part

Background model
mean color (RGB) and luminance gradient
\[ E[\bar{I}(x, y), \nabla L(x, y)] \]
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 (~10°)

MAP Pose Trajectory (half speed)
Sloped surface (~10°)

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

Controller optimized and tested for random 100 N pushes

[Wang et al, SIGGRAPH 2010]
Walking on narrow beam with disturbances

[Wang et al, SIGGRAPH 2010]
Walking on ice with motor noise

Baseline tested without motor noise

Trained and tested with motor noise (β = 100)

[Wang et al, SIGGRAPH 2010]
Interactive control

Looking for coffee before the SIGGRAPH deadline

Interactive demo: User-controlled heading direction and controller switching

[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

Hard Table

Mouse Pad
Bouncing ball

Hard Table

Mouse Pad
Bouncing ball

Hard Table

Mouse Pad

Path of ball

Contact force

Gravity

Contact force

Model

24 N/m  surface  15 N/m
stiffness
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:

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

\[ = \mathcal{A}_{int}(\mathbf{q}) + \mathcal{A}_{ext}(\mathbf{q}, \dot{\mathbf{q}}) + \mathcal{A}(\mathbf{q}, \dot{\mathbf{q}}) \]

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

\[ \tau_{ext}(\mathbf{q}, \dot{\mathbf{q}}) = \mathcal{F}(\mathbf{q}) \left[ f_g + f_r(\mathbf{q}, \dot{\mathbf{q}}) \right] \]

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 *fictitious root forces* are necessary:

$$\min_{\theta, f_{\text{root}}, \tau_{\text{int}}} \sum_t ||f_{\text{root}}(t, \theta)||^2$$

s.t. \[ M \ddot{q} = \mathcal{F}(\tau_{\text{int}}, \theta, f_{\text{root}}) + A \]

- 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

ankle

knee

hip

shoulder

walking

jogging
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

Net vertical ground force for left and right feet

Time (sec)

Force (N)

- mocap
- video
- video (smoothed)
Comparison of video & mocap

Joint torques for left and right knees

<table>
<thead>
<tr>
<th>Force (N)</th>
<th>Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mocap</td>
<td>video</td>
</tr>
<tr>
<td>video</td>
<td>video (smoothed)</td>
</tr>
</tbody>
</table>

mocap
video
video (smoothed)
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


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 ...