3D Human Motion Analysis and Manifolds

Kim Steenstrup Pedersen
DIKU Image group and E-Science center

Motivation

Goal: To give an overview of how manifolds and manifold learning are used in human motion analysis.

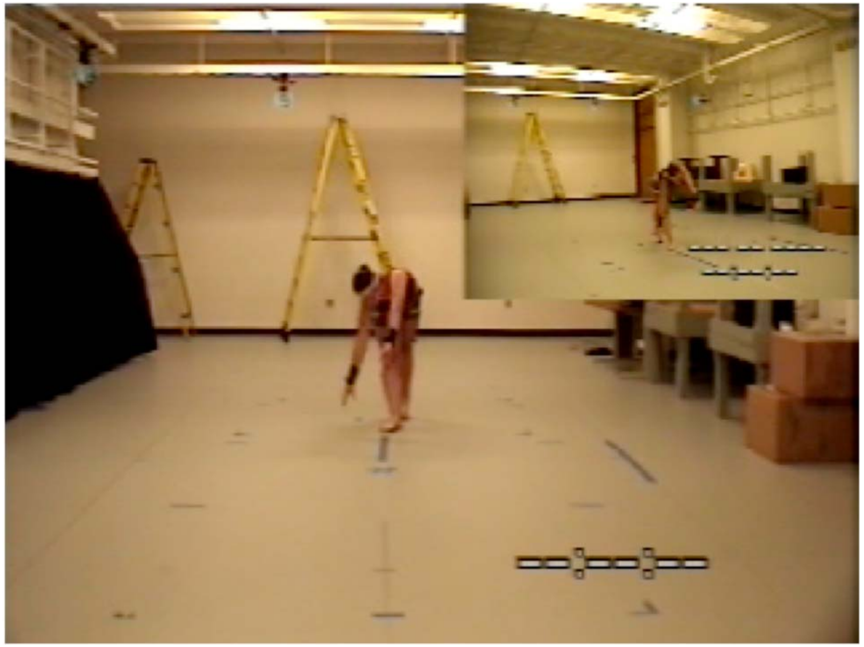
Outline of this lecture:

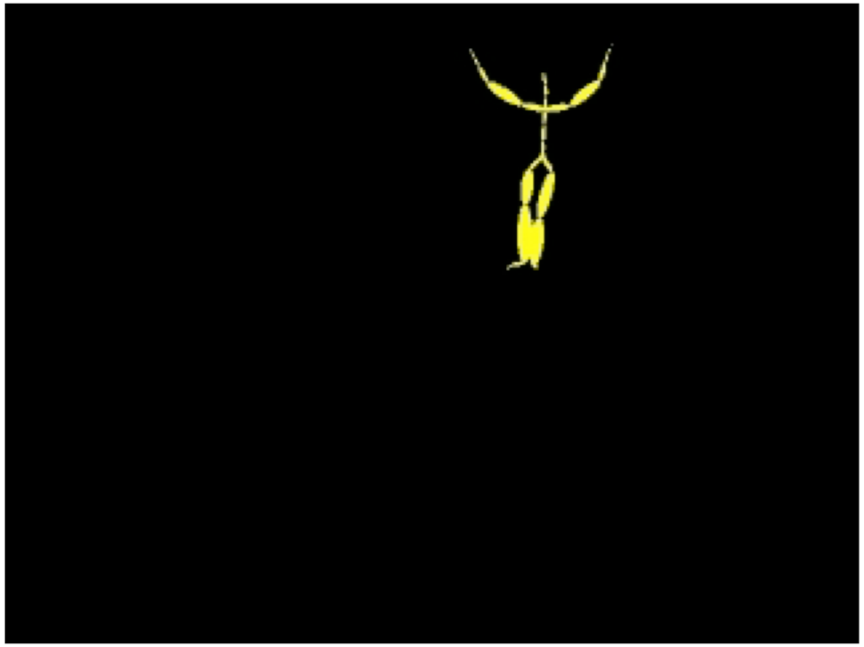
- 3D human motion analysis 101
- Manifolds in human motion analysis
- 2 - 3 concrete examples will be given

3D Human motion analysis

- Def.: Estimation of 3D pose and motion of an articulated model of the human body from visual data – a.k.a. motion capture.
- Marker-based motion capture (MoCap):
  - Outcome: Tracking markers on joints in 3D giving joint positions.
  - Markers: Acoustic, inertial, LED, magnetic, reflective, etc.
  - Cameras or active sensors.
- Marker-less motion capture (MoCap):
  - Outcome: 3D joint positions or triangulated surfaces and relation to video sequence.
  - Multi-view (several cameras / views)
  - Monocular (single camera / view)
  - Camera / view types: Optical camera, stereo pair, time-of-flight cameras, etc.

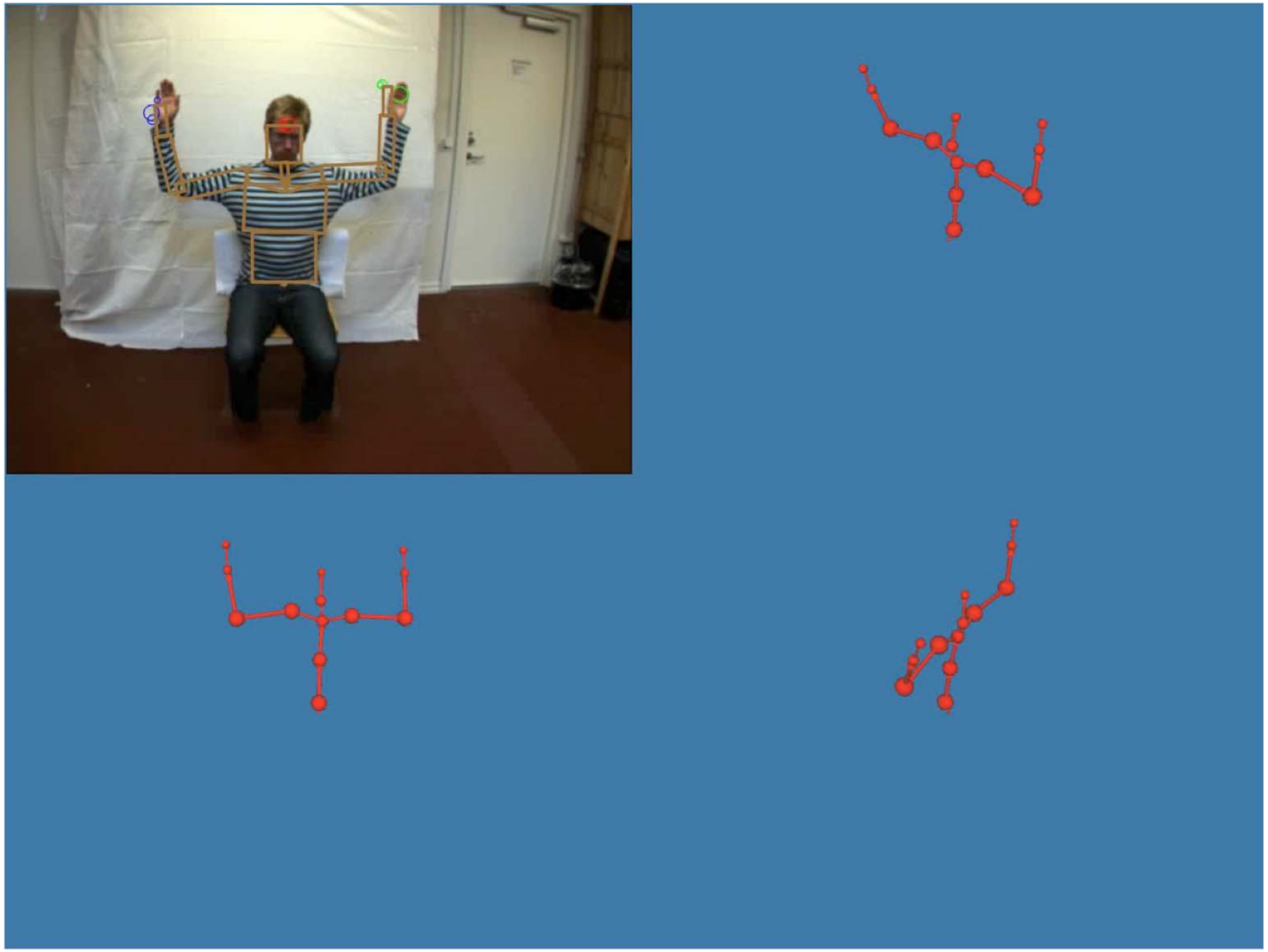
3D Marker based motion capture





[http://mocap.cs.cmu.edu/]

3D Marker-less motion capture (Upper body)



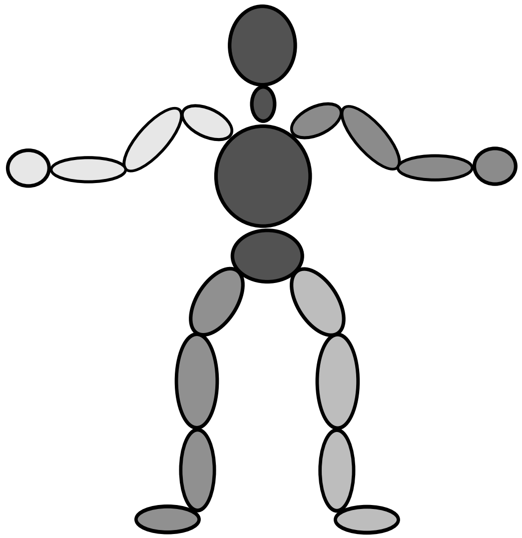
[Hauberg et al, 2009]

Why do we want to do human motion analysis?

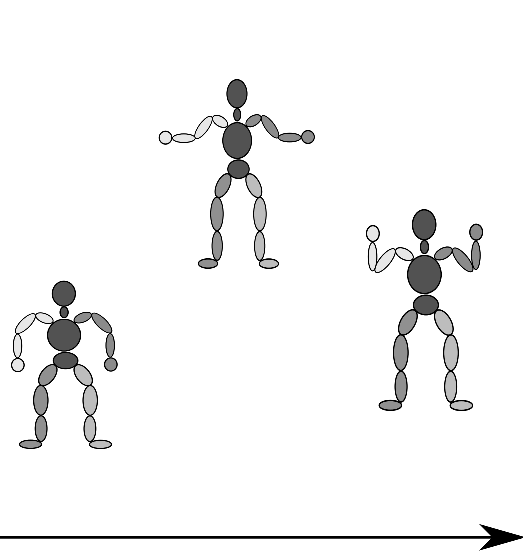
- Human computer interaction: Non-invasive interface technology
- Computer animation: Entertainment (movies and games), education, visualization
- Surveillance: Suspicious behavior recognition, movement patterns
- Physiotherapeutic analysis: Sports performance enhancement, patient treatment enhancement
- Biomechanical modeling

Human body model

- The human body is commonly modeled as an articulated collection of rigid limbs connected with joints.
- **Common representation:**
  - Vector $y = \left[\theta_1, \ldots, \theta_D\right]^T$ of joint angles together with some representation of global position and orientation.
  - Geometric shapes for modeling limb extend (boxes, ellipsoids).



[Hauberg et al, 2009]



- Other representations:
  - Joint positions
  - End-effector positions
  - Surface models
  - …

Human body model constraints

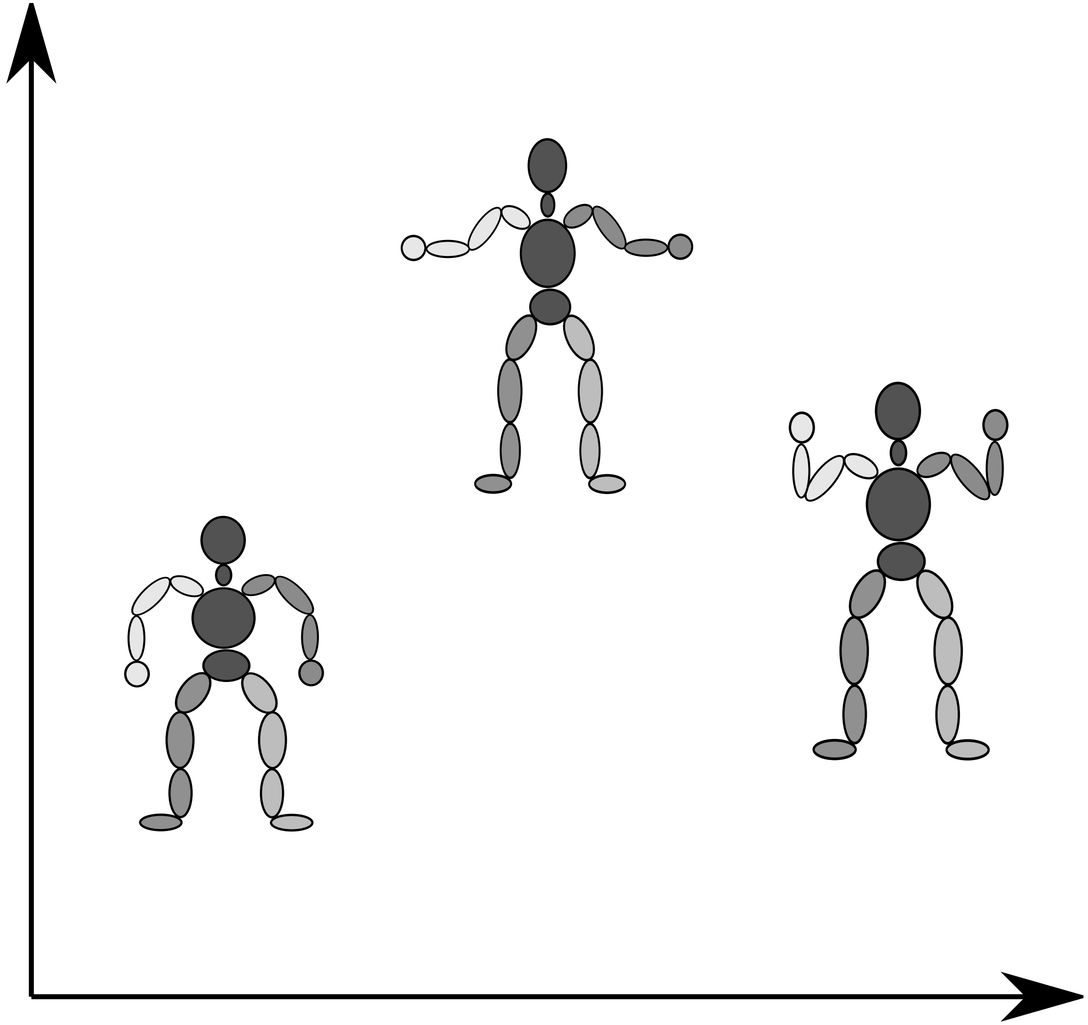
- Natural physical constraints:
  - Body limitations, e.g. joint angle limits, limb dimensions (volume, length, etc.), …
  - Non-penetrability of limbs
  - Angular velocity and acceleration limits

- Constraints can be modeled as either hard or soft constraints.

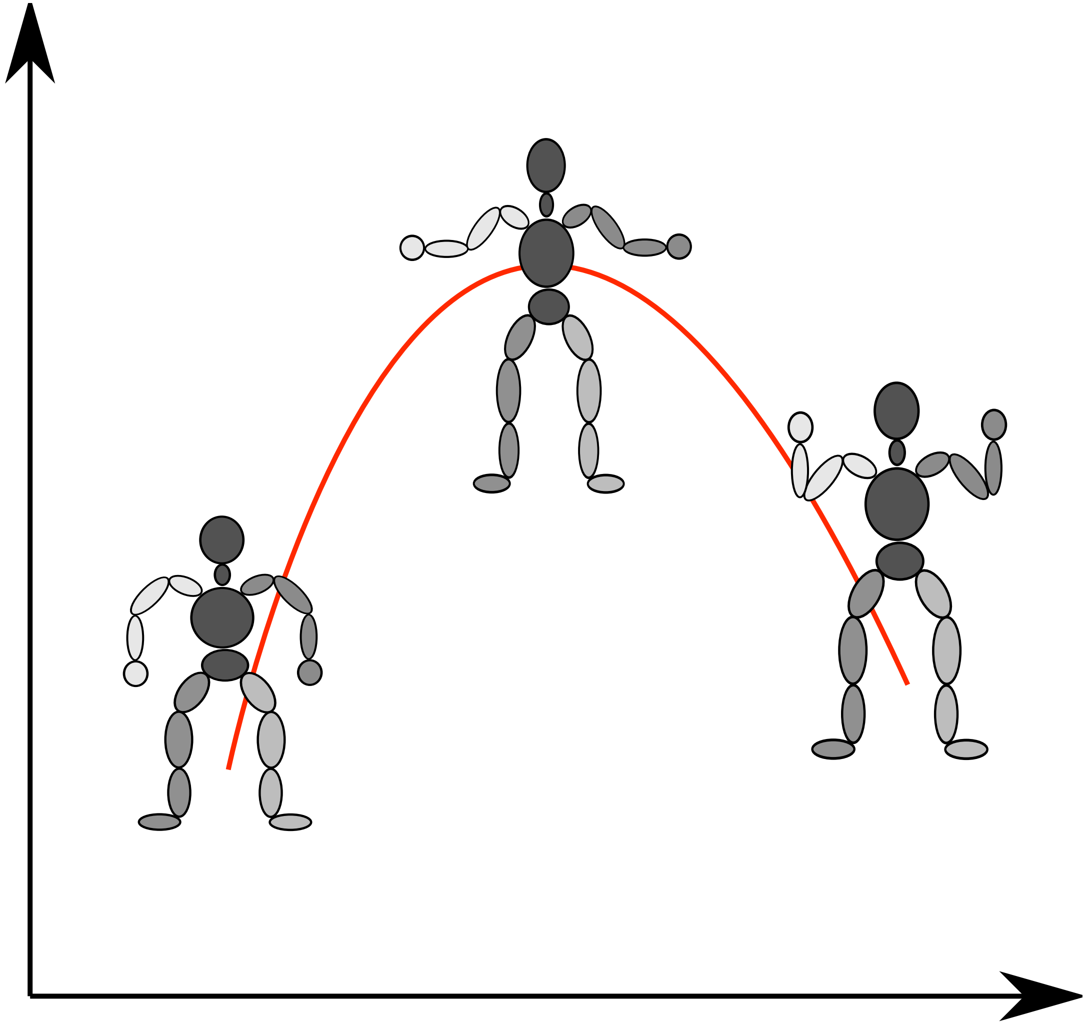
Manifolds in human motion analysis

- The manifold representation is a natural choice because:
  - Human motion is sparsely distributed in pose space with low intrinsic dimensionality. This is especially true for activity specific motion, such as walking.
  - Human motion is generally continuous and smooth – joint angles does not change instantaneously in large jumps (governed by Newton laws). Hence we would like dimensionality reduction which respect this (locality preservation).
  - Constraints leads to boundaries and maybe to holes in manifolds.
- Added benefits: Dimensionality reduction
  - Necessary to make robust estimates of model parameters from small data sets.
  - Will make most tracking algorithms more feasible.

Manifolds in human motion analysis



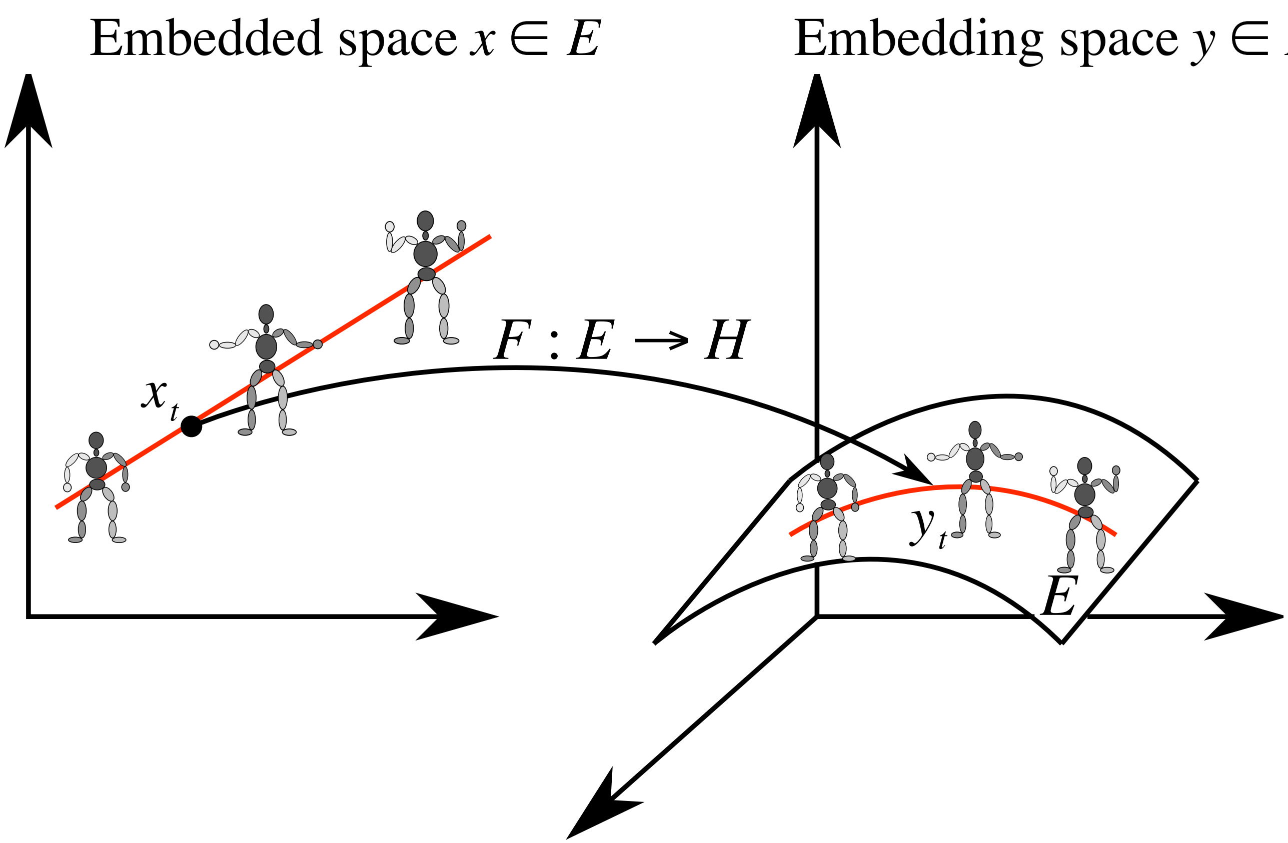
Manifolds in human motion analysis



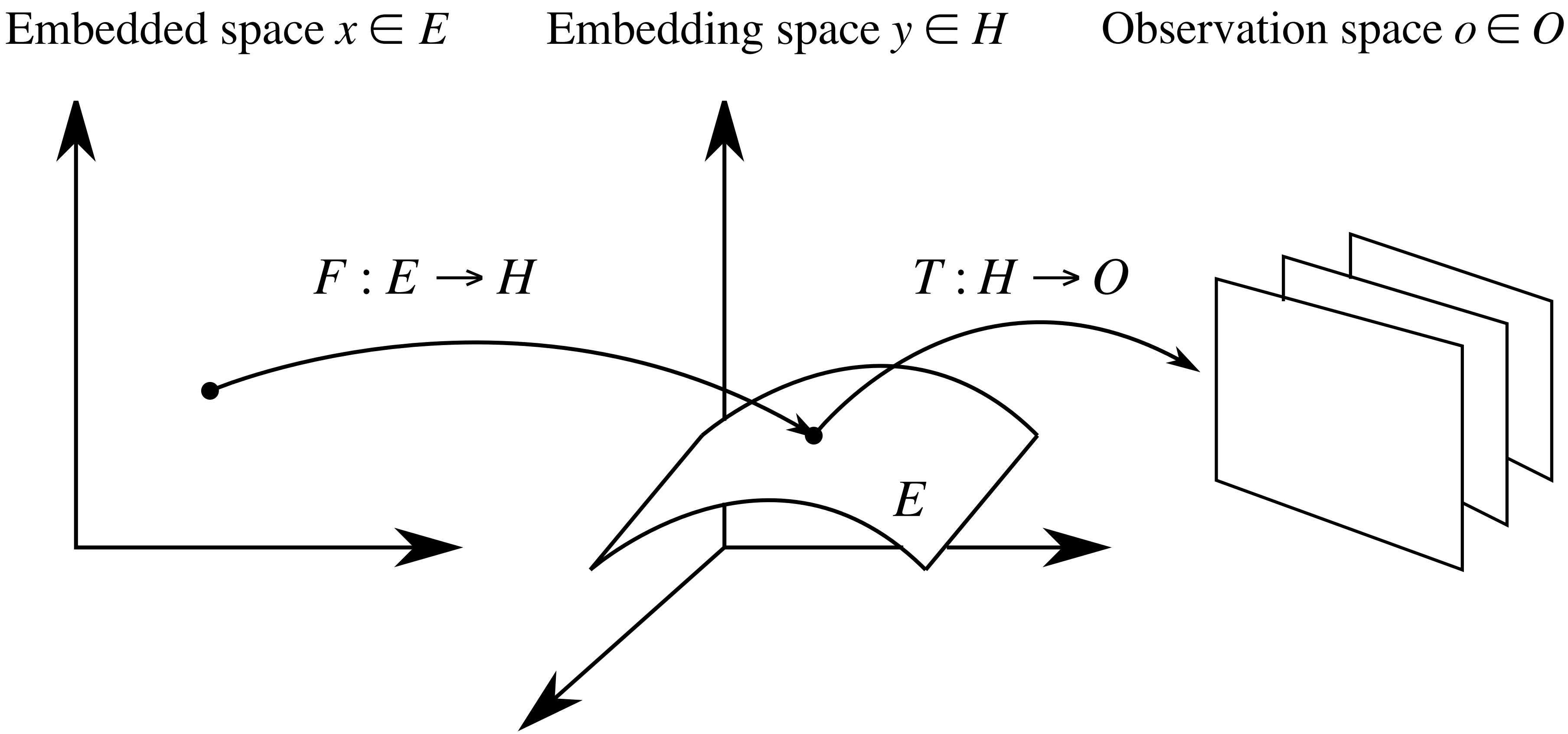
Motion in pose space

- Motion is modeled as temporal curves in pose space

$$y_t = [\theta_1(t), \ldots, \theta_D(t)]^T \quad , \quad x_t = [x_1(t), \ldots, x_d(t)]^T$$



Manifolds in human motion analysis



Goal: Estimate poses and motion from observations. Unkowns: *y, x*
In general we need to learn parameters of the mappings *F* and *T*.

Tracking of human motion (Bayesian framework)

Apply tracking algorithms to sequentially estimate the pose.

- Key ingredients of a sequential Bayesian framework:
  - Observation model: $p_O(o_t \mid y_t)$
  - Prior on poses: $p_H(y_t)$
  - Prior on embedded space: $p_E(x_t)$
  - Dynamical model: $p_H(y_t \mid y_{1:t-1})$ or $p_E(x_t \mid x_{1:t-1})$

- Estimation:
  - Sequential stochastic filtering are commonly used – e.g. Kalman and particle filtering. Sometimes deterministic optimization is also possible.
  - Example: 1^{st} order Markov chain example of filtering on manifold:

$$p(x_t \mid o_{1:t}) \propto \int p(o_t \mid F(x_t))\, p(x_t \mid x_{t-1})\, p(x_{t-1} \mid o_{1:t-1})\, dx_{t-1}$$

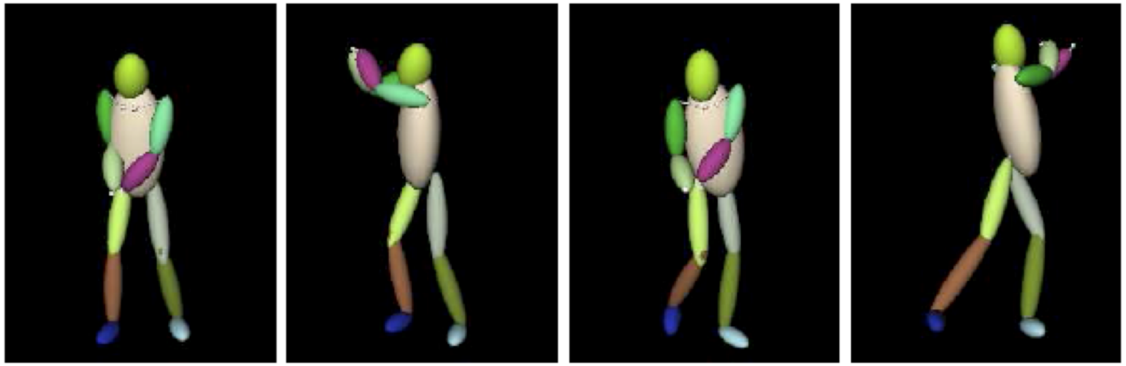
Pose and motion prior models

- Priors on pose: Which poses are probable?
  - Activity specific pose models: Walking, running, golfing, jumping, etc. Examples: [Urtasun et al, 2005b; Sminchisescu et al, 2004].
  - Constraints: Joint angle limits, non-penetrability of limbs, etc.
- Priors on motion: What types of motion are probable?
  - Activity specific motion models: Walking, running, golfing, jumping, etc. Examples: [Urtasun et al, 2005a, 2006].
  - Markov chain models (e.g. 1st and 2nd order models, HMM, etc.)
  - General stochastic processes
  - Constraints: Angular velocity and acceleration limits.
- Priors on plausible human poses and motion are especially important for monocular 3D tracking in order to handle occlusion, depth ambiguity, and noisy observations.

Motion and pose prior: PCA [Urtasun et al, 2005a]

- Prior model for golf swings:
  Learn a joint model on motion and poses from motion capture data using PCA. Use the prior to track golf swings in 3D.

- Training set:
  - 10 motion capture golf swing samples (from CMU data set).
  - Time warp samples to meet 4 key postures and sample with N=200 time steps. Use normalized time in [0,1].



[Urtasun et al, 2005a]

Motion and pose prior: PCA [Urtasun et al, 2005a]

- Model:
  - D=72 angles (+ global 3D position and 3D orientation).
  - Angular motion vector, N*D=14400 dim.: $y = [\psi_{\mu_1}, \ldots, \psi_{\mu_N}]^T$

  ψ_{μ_i} row vector of joint angles at normalized time μ_i

  - Motion model: $y \approx \Theta_0 + \sum_{i=0}^{d} \alpha_i \Theta_i$

  d=4 principle components Θ_i of the training set.

  Θ_0 denotes the mean of the training set.

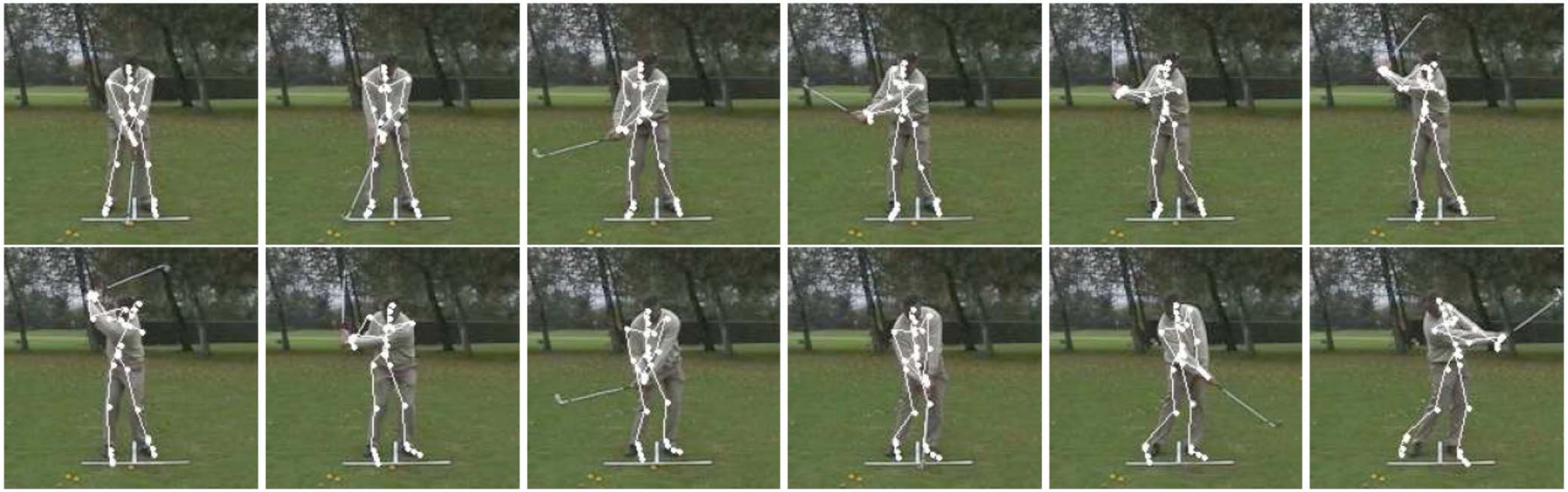
  Embedded coordinates $x = \left[\alpha_1, \ldots, \alpha_d\right]^T$

Motion and pose prior: PCA [Urtasun et al, 2005a]

- Estimation of motion:
  - Sequential least squares minimization of PCA coefficients, global position and orientation, and normalized times over a sliding window of *n* frames.
  - Objective function include observation model and global motion smoothing terms.
  - Linear global motion model.

Motion and pose prior: PCA Results

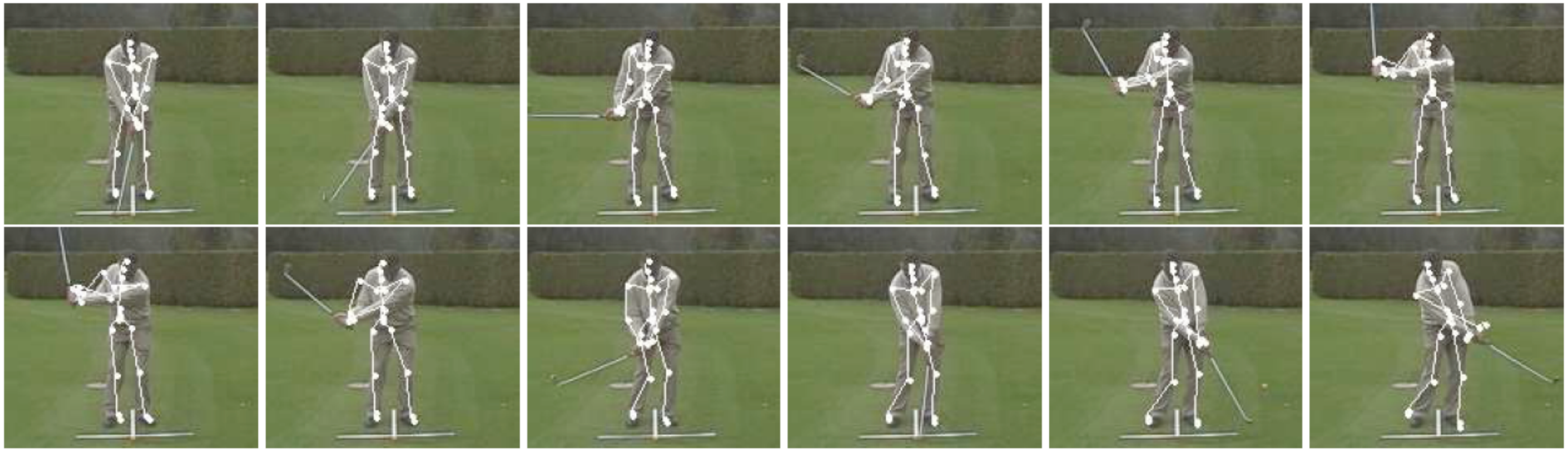
Full swing



[Urtasun et al, 2005a]

Motion and pose prior: PCA Results

Short swing



[Urtasun et al, 2005a]

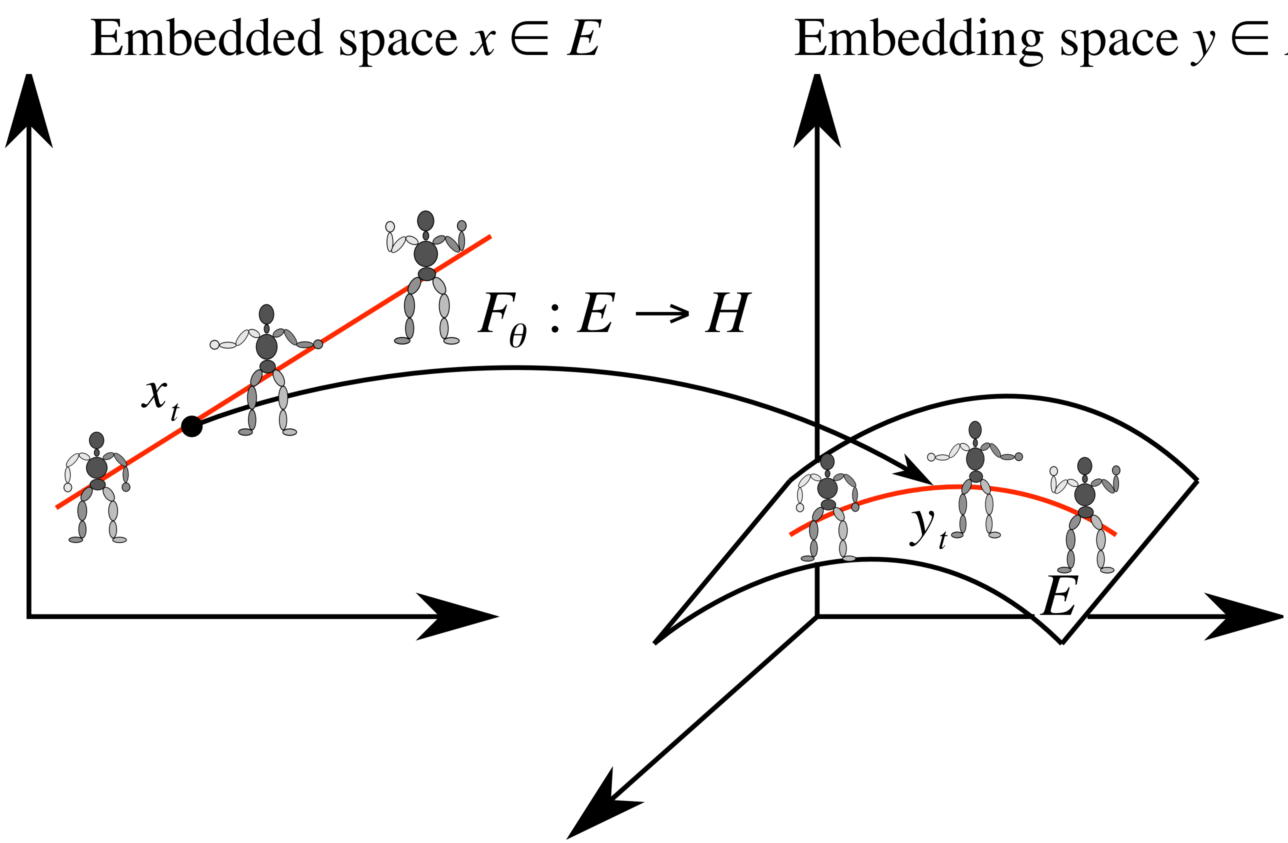
Priors on poses: Laplacian eigenmaps [Sminchisescu et al, 2004]

- Priors for poses using Laplacian eigenmaps:
  - Activity specific, but combinations of activities are possible as we shall see.

- Outline:
  - Embedded manifold *E* is learnt from MoCap training data (CMU database) using Laplacian eigenmaps.
  - Intrinsic dimensionality can be estimated by the Hausdorff dimension.
  - Use a first order Markov chain dynamical model in embedded space *E*.
  - Tracking is performed by standard sequential Bayesian estimation using Covariance scaled sampling.

Mapping to manifold

– Learn global smooth mapping F_θ from embedded space *E* to embedding space *H* (angle representation) by kernel regression using the training set.



Priors on poses: Laplacian eigenmaps

- Priors in embedded space and embedding space:
  - Physical constraints (joint limits, angular velocity limits, non-penetrability of limbs, etc.) naturally defined in the original representation (embedding space) *H.*
  - Prior in embedded space *E* given by learning a mixture of Gaussian from training data:

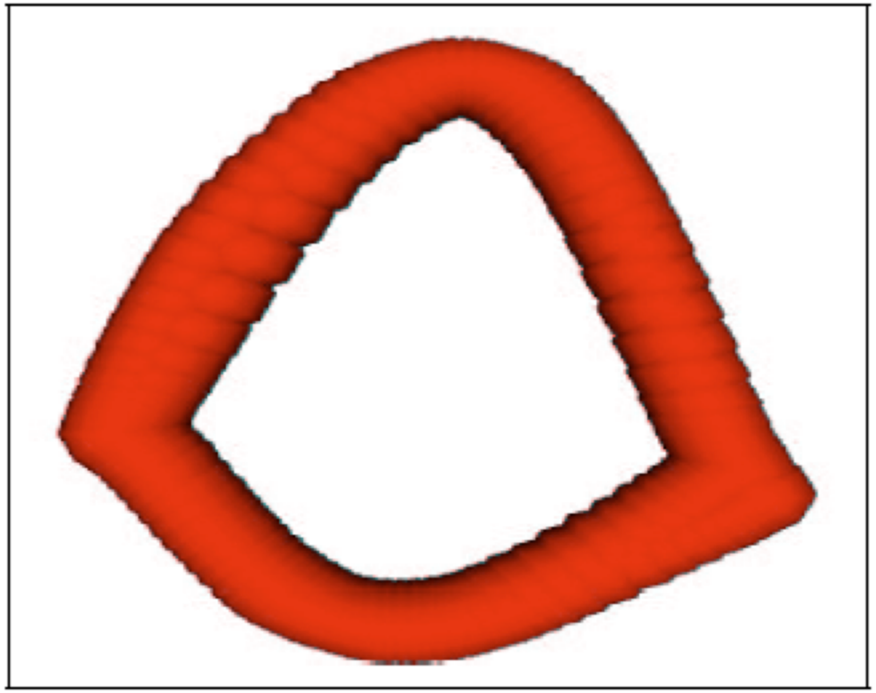
$$p_E(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x, \mu_k, \Sigma_k)$$

  - **Solution - Embedded flattened prior**:
    - Prior in original space *H* (physical constraints) is used to produce flattened prior in embedded space *E*:

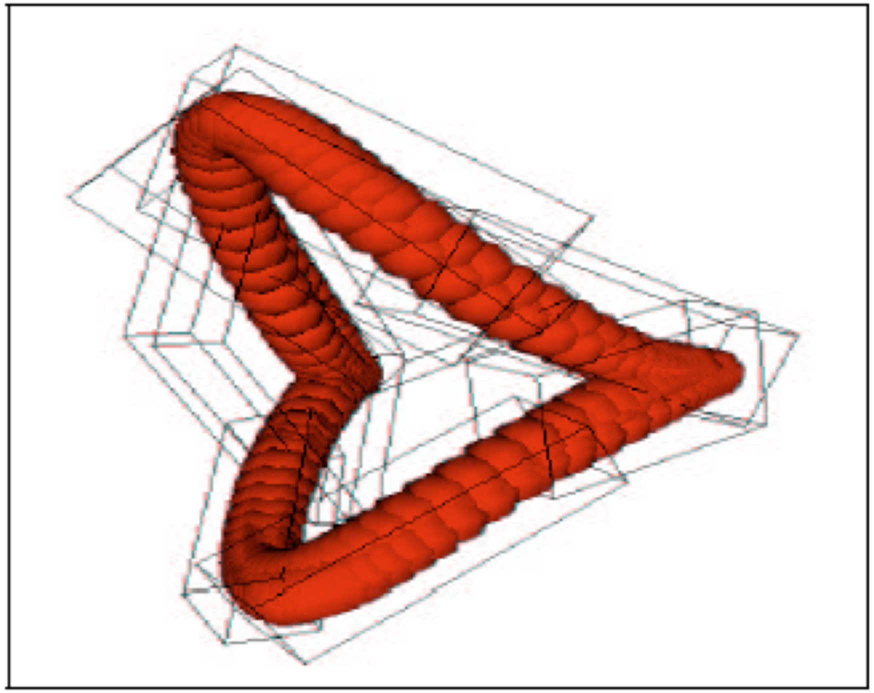
$$p(x) \propto p_E(x) \cdot p_H(F_\theta(x)) \cdot \left| J_{F_\theta}(x)^T J_{F_\theta}(x) \right|^{1/2}$$

Priors on poses: Walking prior $p_E(x)$

2D embedding



3D embedding



[Sminchisescu et al, 2004]

Priors on poses: Interaction [Sminchisescu et al, 2004]

Priors on poses: Effect of embedding prior

[Sminchisescu et al, 2004]

Priors on poses: GP's and latent variables

- Priors for pose derived from a small training set using a scaled Gaussian processes (GP) latent variable model [Urtasun et al, 2005b]:
  - Activity specific model learnt from motion capture training data.
  - Can learn and generalize from a single training motion example.
  - Learn the mapping $y = F(x)$ from *E* to *H* and optimize the latent variable positions at the same time.
  - Learn a joint distribution *p*(*x*,*y*) on embedded *E* and embedding spaces *H*. Assign high probability to new *x* near training data.

Priors on poses: GP's and latent variables

- Training:
  - Mean zero training data: $Y = \left[y_1, \ldots, y_N\right]^T \;, y_i \in R^D$
  - Unknown model parameters: $M = \left\{ \left\{x_i\right\}_{i=1}^{N}, \alpha, \beta, \gamma, \left\{w_j\right\}_{j=1}^{D} \right\}$
  - GP require that:

$$p(Y \mid M) = \frac{\left|W\right|^N}{\sqrt{(2\pi)^{ND}\left|K\right|^D}} \exp\left(-\frac{1}{2}\mathrm{tr}(K^{-1}YW^2Y^T)\right)$$

$W = \mathrm{diag}(w_1, \ldots, w_D)$ and $K_{ij} = k(x_i, x_j)$ is a RBF with parameters α, β, γ

  - Model parameters are learned by finding the MAP solution, using a simple prior on hyperparameters and an isotropic i.i.d. Gaussian prior for latent positions *x*.

Priors on poses: GP's and latent variables

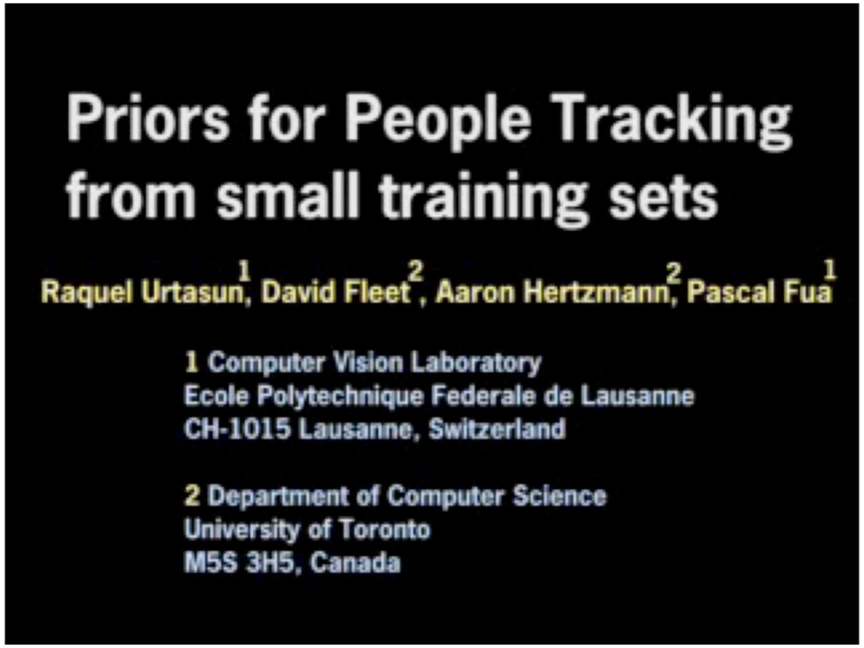
- Pose prior:
  - Joint probability on new latent positions *x* and poses *y*:

$$p(x,y \mid M,Y) = \exp\left(-\frac{x^T x}{2}\right) \frac{\left|W\right|^{N+1}}{\sqrt{(2\pi)^{(N+1)D}\left|\hat{K}\right|^{D}}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\hat{K}^{-1}\hat{Y}W^{2}\hat{Y}^{T}\right)\right)$$

$$\hat{Y} = \left[y_1,\ldots,y_N,y\right]^T \ , \ \hat{K} = \begin{pmatrix} K & \mathbf{k}(x) \\ \mathbf{k}(x)^T & k(x,x) \end{pmatrix} \ , \ \mathbf{k}(x) = \left[k(x_1,x),\ldots,k(x_N,x)\right]^T$$

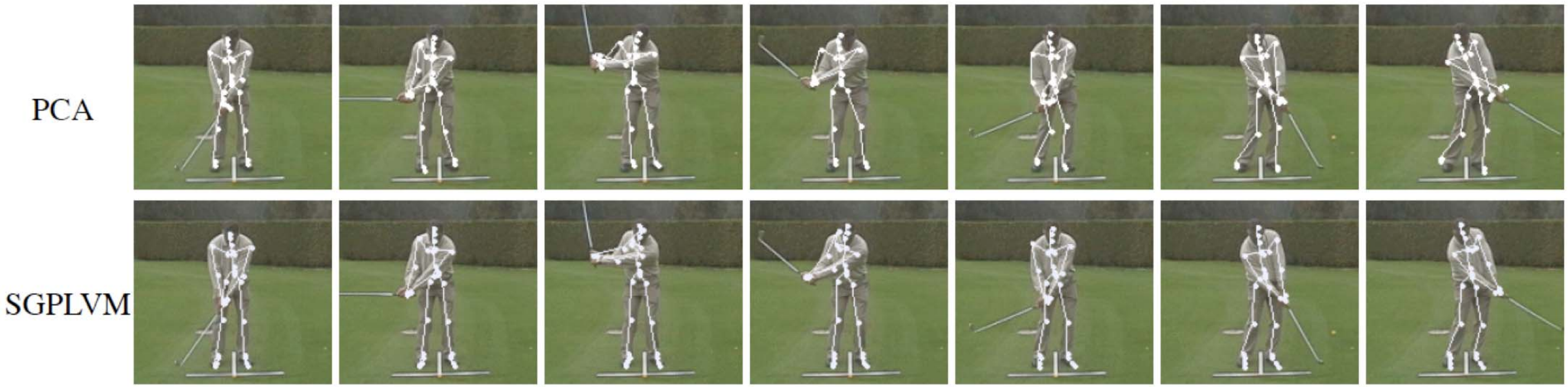
  - Learned mean mapping: $F(x) = \mu + Y^T K^{-1}\mathbf{k}(x)$
  - Learned variance: $\sigma^2(x) = k(x,x) - \mathbf{k}(x)^T K^{-1}\mathbf{k}(x)$
- Tracking:
  - Sequential MAP estimation of *x,y* based on model and observations with 2nd order Markov dynamics.
  - Solved by deterministic optimization.

Priors on Poses: GP’s and Latent variables



[Urtasun et al, 2005b]

Priors on Poses: GP’s and Latent variables



[Urtasun et al, 2005b]

Summary

- We have seen how pose and motion manifolds appear in human motion analysis:
  - Human motion have low intrinsic dimensionality, especially activity specific motion
  - Human motion is smooth
  - Physical limitations – joint limitations, non-penetration, etc.
- Strong prior models are especially needed in monocular 3D tracking.
- I have given a couple of concrete examples:
  - PCA prior model of pose and motion
  - Laplacian eigenmaps for learning pose prior
  - Gaussian processes latent variable model for pose prior

Suggested reading material

- R. Poppe: Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108 (1-2): 4-18, 2007.
- R. Urtasun et al.: Monocular 3-D tracking of the golf swing. Proceeding of CVPR'05, 2005.
- C. Sminchisescu and A. Jepson: Generative modeling for continuous non-linearly embedded visual inference. ICML’04, pp. 759--766, 2004.
- R. Urtasun et al.: Priors for people tracking from small training sets. Proceeding of ICCV'05, pp. 403 - 410, 2005.
- CMU Graphics lab motion capture database: http://mocap.cs.cmu.edu/

Additional references

- S. Hauberg, J. Lapuyade, M. Engell-Nørregård, K. Erleben and K. S. Pedersen. Three Dimensional Monocular Human Motion Analysis in End-Effector Space. In Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), pp. 235-248, 2009.
- M. Engell-Nørregård, S. Hauberg, J. Lapuyade, K. Erleben, and K. S. Pedersen. Interactive Inverse Kinematics for Monocular Motion Estimation. VRIPHYS'09, submitted, 2009.
- R. Urtasun, D. J. Fleet, and P. Fua: Temporal motion models for monocular and multiview 3D human body tracking. Computer Vision and Image Understanding, 104: 157-177, 2006.
- Z. Lu et al.: People Tracking with the Laplacian Eigenmaps Latent Variable Model. NIPS'07, 2007.
- A. Elgammal and C.-S. Lee: Tracking people on a torus. IEEE T-PAMI, 31(3): 520-538, 2009.
- N. D. Lawrence: Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data. Advances in Neural Information Processing Systems, pp. 329-336, 2004.