Gaussian Process Dynamical Models (GPDM) for Motion Analysis

Zhenbo LI

Zhenbo.li@IT-SudParis.eu

28/04/08

Outline

- Motion Models
- Linear Motion Models (PCA)
- Gaussian Process Dynamical Models
- Motion Recognition
- Conclusion

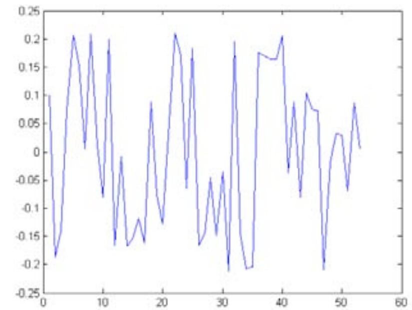
Motion Models

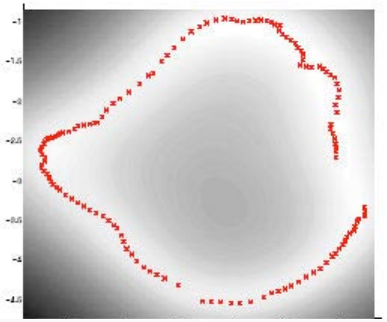
- **What motions are likely?**
- **Applications:**
  - Computer animation
  - Computer vision

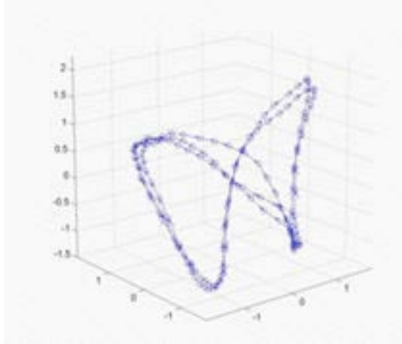
NBA Live 2008

TELECOM & Management SudParis

Motion Models



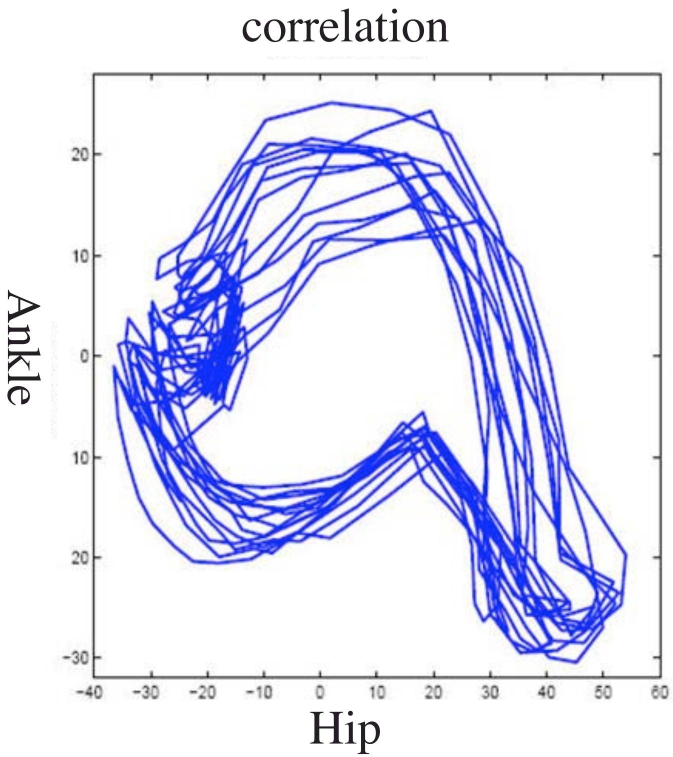




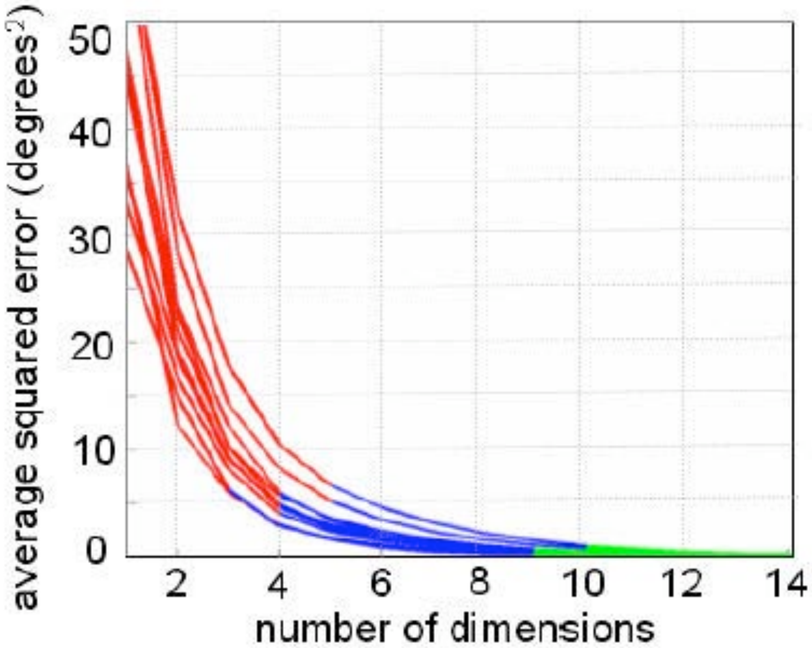
- $M(t) = \{p(t), Q(t)\}$
- Little work on motion models [Urtasun 2006]
- Linear motion models
- Non-Linear motion models

Problem

- High-dimensional
- Time related (dynamic)
- Stylistic diversity and variation

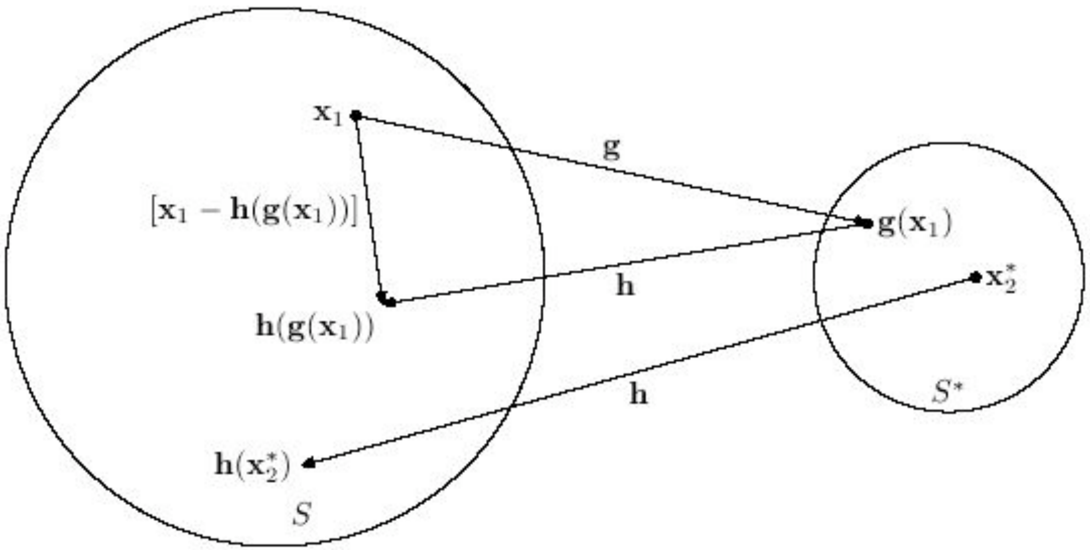


[Pullen&Bregler'2002]



[Safonova et. al'2004]

Motion Models



- **S: space of human poses**

  ☹ high number of DOF
- **S*: lower dimensional subspace**

  ☞ **Mapping function between S and S***

Linear Motion Models (PCA)

Subspace from PCA

- x is the vector of joint angles
- F is the projection of x in eigenspace

$$F_1 = u_{11}x_1 + u_{21}x_2 + \cdots + u_{p1}x_p$$

$$\cdots$$

$$F_m = u_{1m}x_1 + u_{2m}x_2 + \cdots + u_{pm}x_p$$

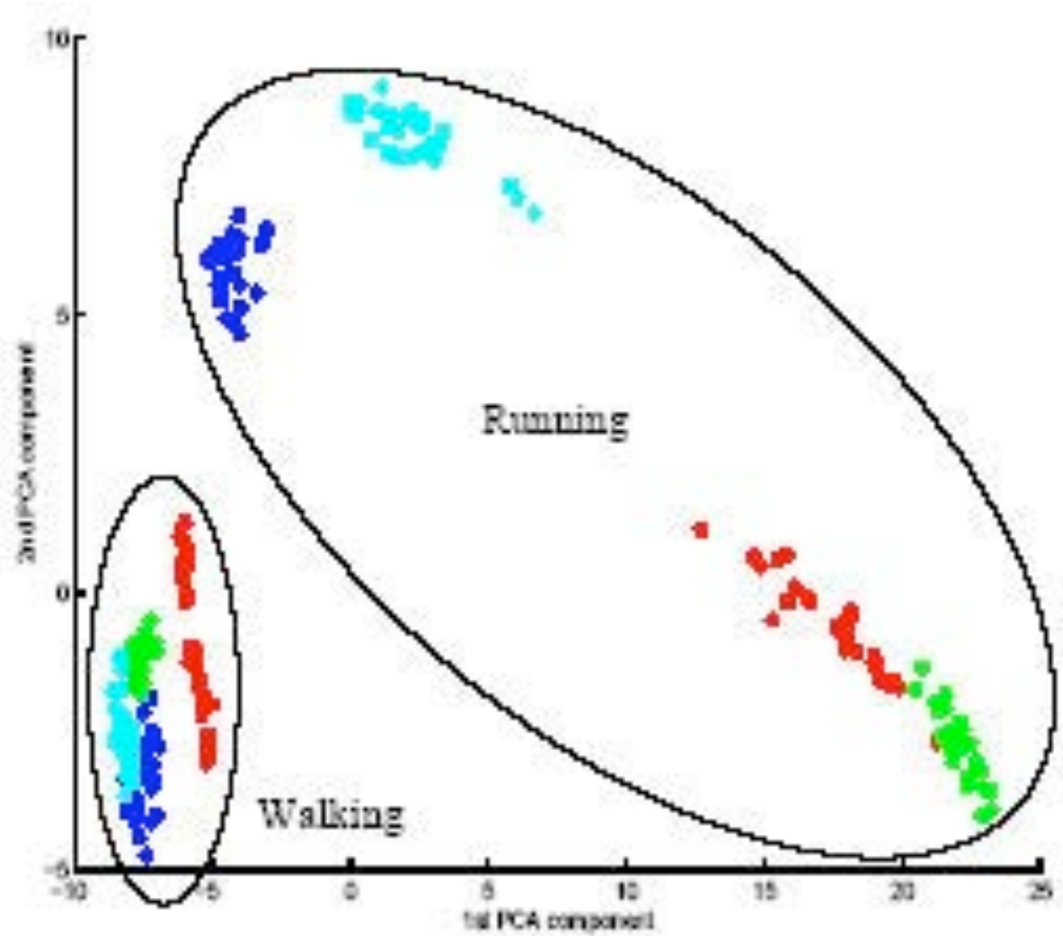
$$E_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Contributive Factor

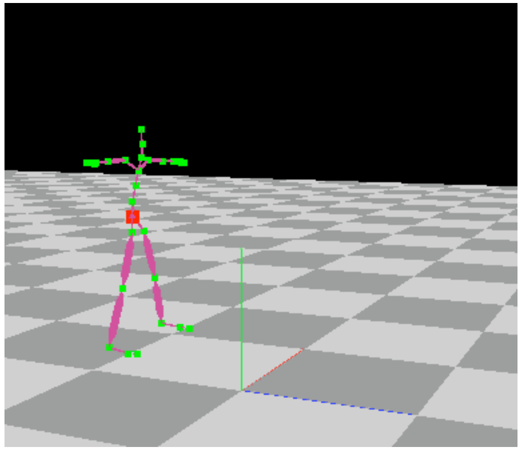
- ☞ **Low dimensional description of motions**

Classification of activities in eigenspace

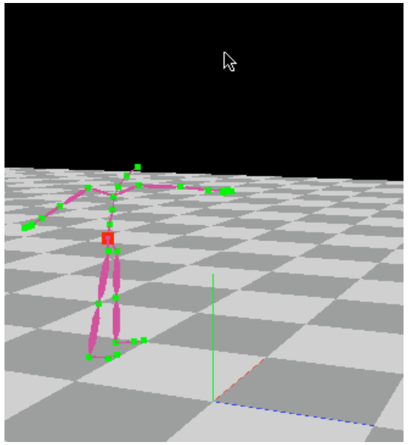
- [Urtasun 2006]



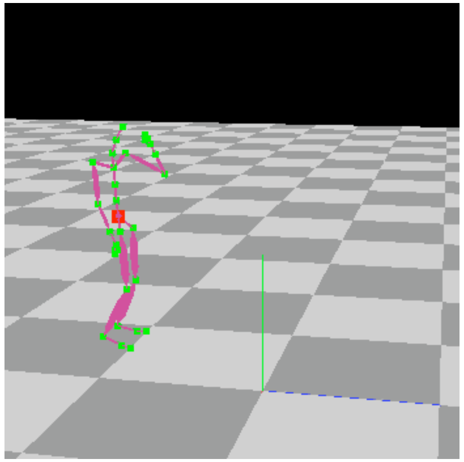
Motion generation



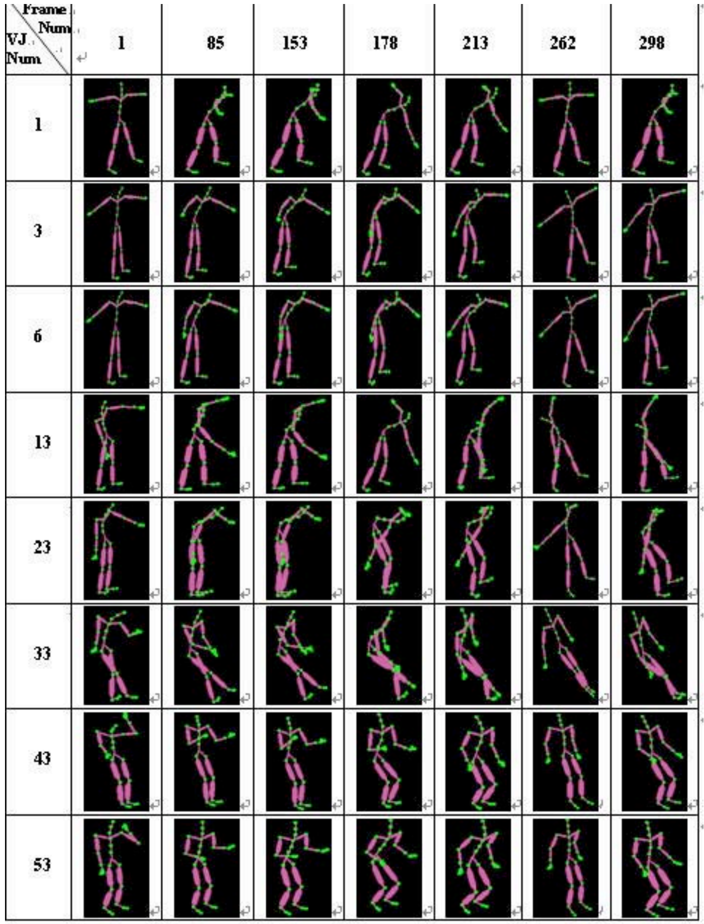
1



6

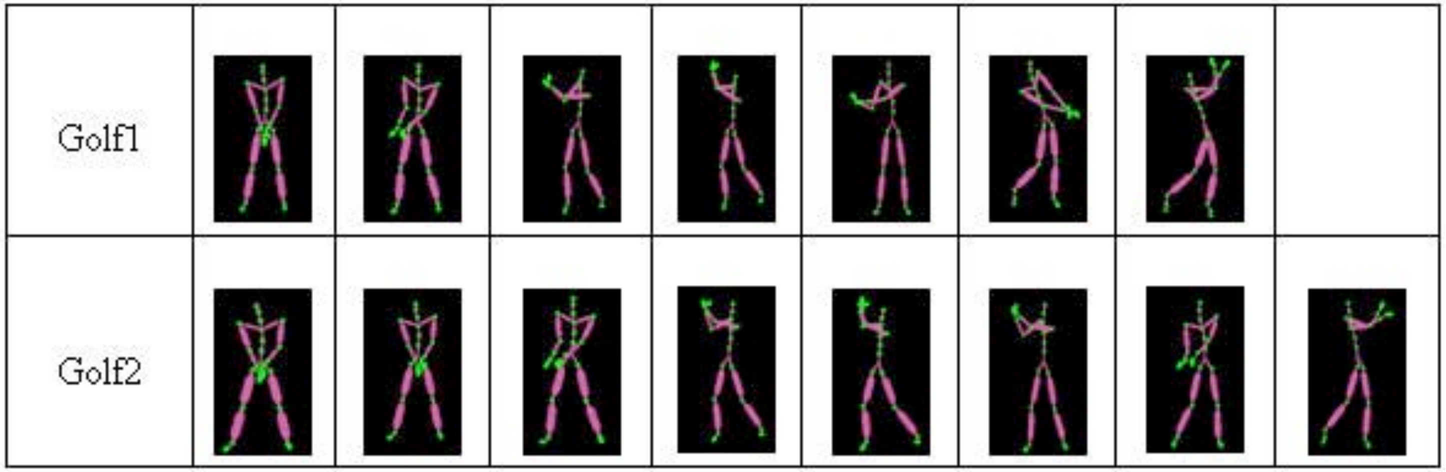


53

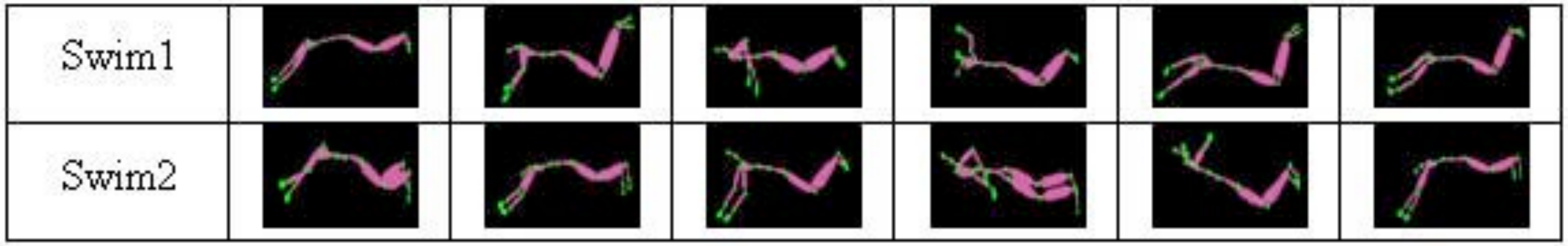


Motion similarity

$$P_{MN} = \frac{\sum_{i=1}^{Dim} M(i) * N(i)}{\sqrt{\sum_{i=1}^{Dim} (M(i))^2 * \sum_{i=1}^{Dim} (N(i))^2}}$$



Golf1(447 frames), Golf2(492 frames), The comparability is: $P_{MN} = 0.9691$



Swim1(1536 frames), Swim2(1780 frames), The comparability is: $P_{MN} = 0.9236$

Motion similarity

P_{MN}	Liu_walk1	Liu_walk2	Liu_walk3	Liu_walk4
Liu_walk1	1	0.9653	0.9383	0.8405

P_{MN}	Rory_run1	Steve_walk1	Rory_walk1	Takeo_walk1
Liu_walk1	0.5366	0.6996	0.7080	0.7759

P_{MN}	Golf1	Swim1	Kickball
Liu_walk1	0.3481	0.1046	0.5889

Gaussian Process Dynamical Model (GPDM)

- Proposed by Wang, Fleet, Hertzmann in 2005.
- Based on the work of Gaussian Process Latent Variable Model proposed by Lawrence in 2004.

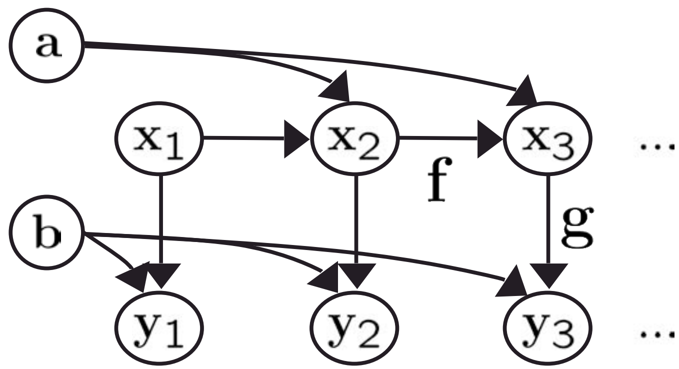
Gaussian Process Dynamical Model (GPDM)

Latent dynamical model:

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t; \mathbf{b}) + \mathbf{n}_{y,t}$$ pose reconstruction

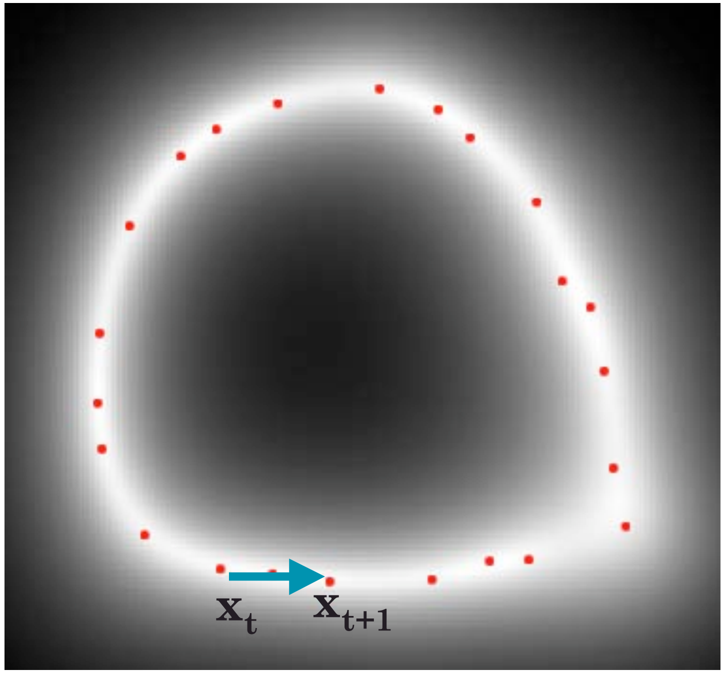
$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1}; \mathbf{a}) + \mathbf{n}_{x,t}$$ latent dynamics

Here, $x_t \in \mathbb{R}^l$ denotes the l dimensional latent coordinates at time t, f and g mappings parameterized by a and b



$n_{x,t}$ *and* $n_{y,t}$ are zero-mean, isotropic, white Gaussian noise

Dynamical models



$$\mathrm{x}_t = \mathrm{f}(\mathrm{x}_{t-1}; \mathrm{a}) + \mathrm{n}_{x,t}$$

Reconstruction

The data likelihood for the reconstruction mapping, given centered inputs $\mathbf{Y} \equiv [\mathbf{y}_1, \cdots, \mathbf{y}_N]^T, \ \mathbf{y}_n \in \mathcal{R}^d$ has the form:

$$p(\mathbf{Y} \,|\, \mathbf{X}, \bar{\beta}, \mathbf{W}) = \prod_{j=1}^{d} \mathcal{N}(\mathbf{Y}_j;\, \mathbf{0},\, \mathbf{K}_Y / w_j^2)$$

$$= \frac{|\mathbf{W}|^N}{\sqrt{(2\pi)^{Nd} |\mathbf{K}_Y|^d}} \exp\left(-\frac{1}{2} \mathrm{tr}\left(\mathbf{K}_Y^{-1} \mathbf{Y} \mathbf{W}^2 \mathbf{Y}^T\right)\right)$$

where

$\mathbf{Y}_j$ contains the j th-dimension of each training pose

$\mathbf{K}_Y$ is a kernel matrix with entries $(\mathbf{K}_Y)_{ij} = k_Y(\mathbf{x}_i, \mathbf{x}_j)$ for kernel function (with hyperparameters $\bar{\beta} = \{\beta_1, \beta_2, \beta_3\}$)

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

$\mathbf{W} \equiv \mathrm{diag}(w_1, ..., w_d)$ scales different pose dimensions

Dynamics

The latent dynamic process on $\mathbf{X} \equiv [\mathbf{x}_1, \cdots, \mathbf{x}_N]^T$, $\mathbf{x}_n \in \mathcal{R}^l$ has a similar form:

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

where

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

$\mathbf{K}_X$ is a kernel matrix defined by kernel function

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

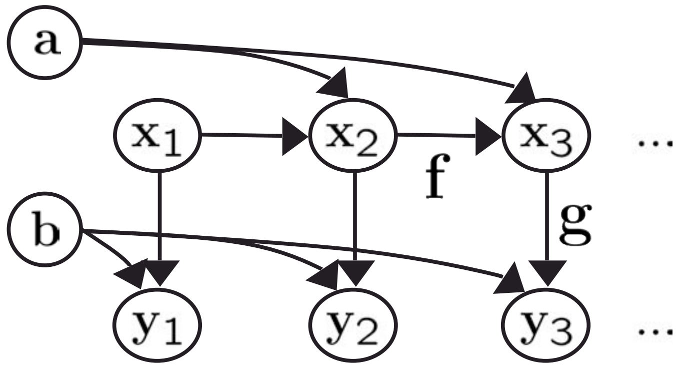
with hyperparameters $\bar{\alpha}$

Markov Property

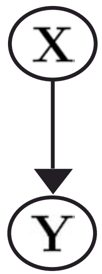
Subspace dynamical model:

$$\mathbf{x}_t = \mathbf{f}(\mathbf{x}_{t-1};\mathbf{a}) + \mathbf{n}_{x,t}$$

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t;\mathbf{b}) + \mathbf{n}_{y,t}$$



Remark: Conditioned on $\{\mathbf{a}_i, \mathbf{b}_j\}$, the dynamical model is 1st-order Markov, but the marginalization introduces longer temporal dependence.



Dynamical models

- **Hidden Markov Model (HMM)**
- **Linear Dynamical Systems (LDS)**
  [van Overschee et al ‘94; Doretto et al ‘01]
- **Switching LDS**
  [Ghahramani and Hinton ’98; Pavlovic et al ‘00; Li et al ‘02]
- **Nonlinear Dynamical Systems**
  [e.g., Ghahramani and Roweis ‘00]

Learning

Training poses: $\mathbf{Y} \equiv [\mathbf{y}_1, \cdots, \mathbf{y}_N]^T, \ \mathbf{y}_n \in \mathcal{R}^D$

Model parameters:

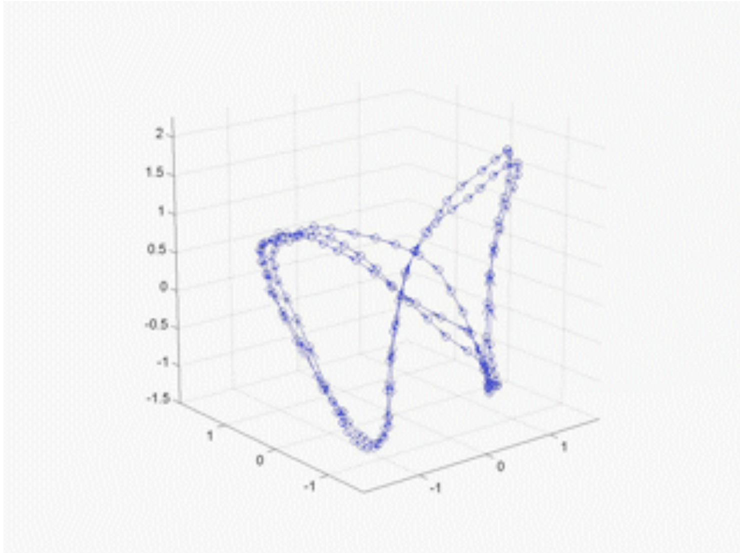
- 3D latent coordinates: $\mathbf{X} \equiv \{\mathbf{x}_n\}_{n=1}^N$
- RBF kernel hyperparameters: $\bar{\beta} = \{\beta_j\} \quad \bar{\alpha} = \{\alpha_j\}$
- weights on output dimensions: $\mathbf{W} = \mathrm{diag}(w_1, ..., w_D)$

Learning: estimate parameters by maximizing

$$\underbrace{p(\mathbf{Y}}_{\text{training motions}}, \underbrace{\mathbf{X}}_{\text{latent trajectories}}, \underbrace{\bar{\alpha}, \bar{\beta}, \mathbf{W}}_{\text{hyperparameters}}) = \underbrace{p(\mathbf{Y} \,|\, \mathbf{X}, \bar{\beta}, \mathbf{W})}_{\text{reconstruction likelihood}} \ \underbrace{p(\mathbf{X} \,|\, \bar{\alpha})}_{\text{dynamics likelihood}} \ \underbrace{p(\bar{\alpha}) \ p(\bar{\beta})}_{\text{priors}}$$

TELECOM & Management SudParis

~2.5 gait cycles (158 frames)

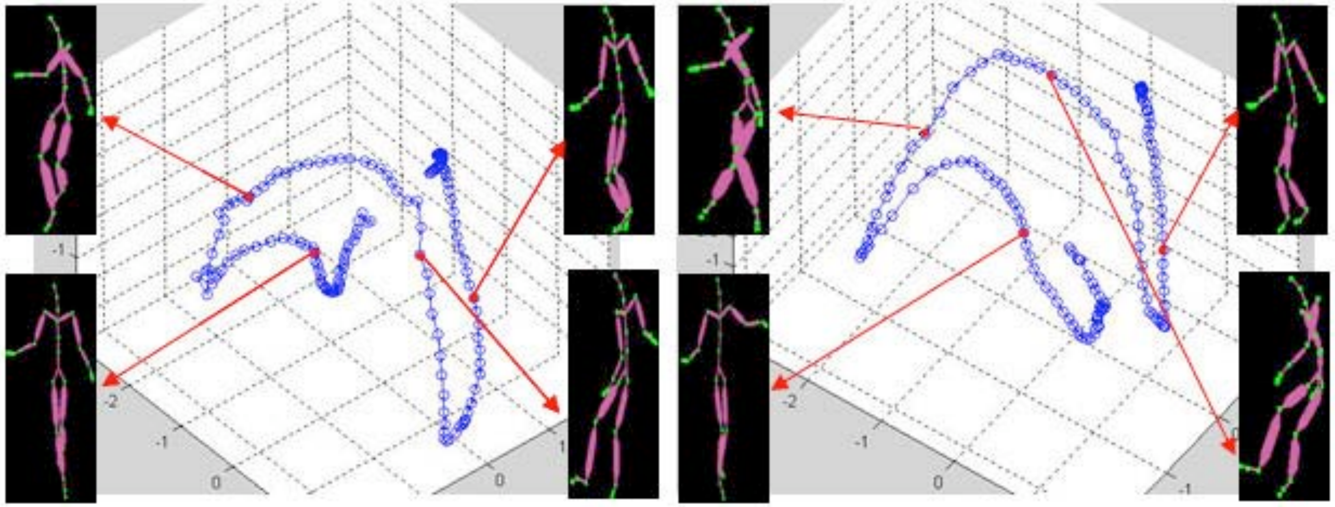


Learned latent coordinates
(1st-order prediction, RBF kernel)

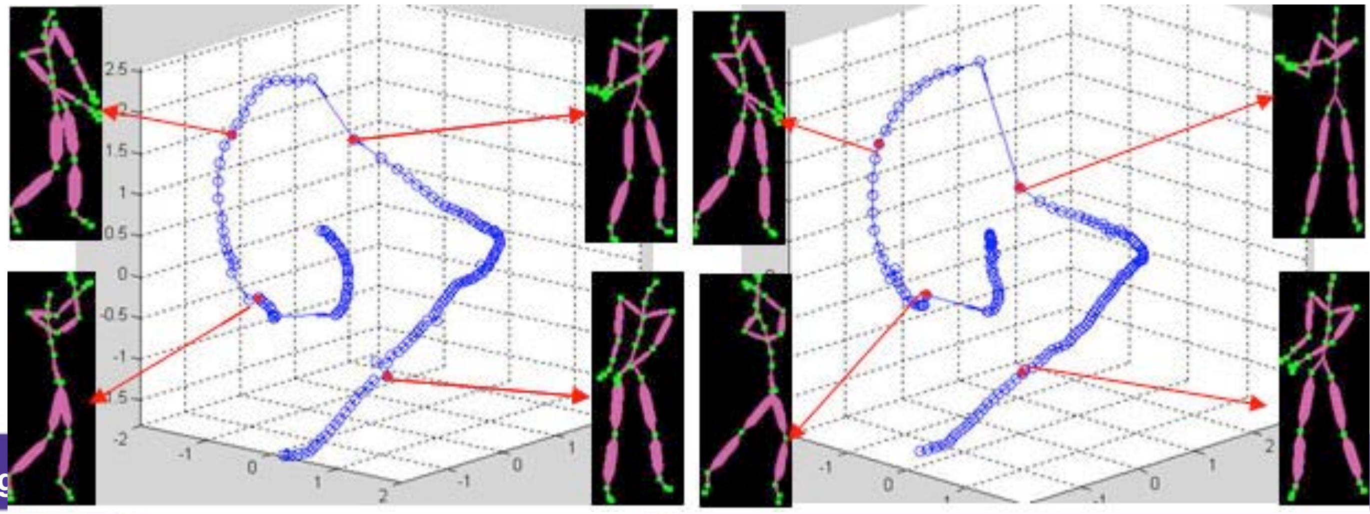
56 joint angles + 3 global translational velocity + 3 global orientation
from CMU motion capture database

TELECOM & Management SudParis

Different motions

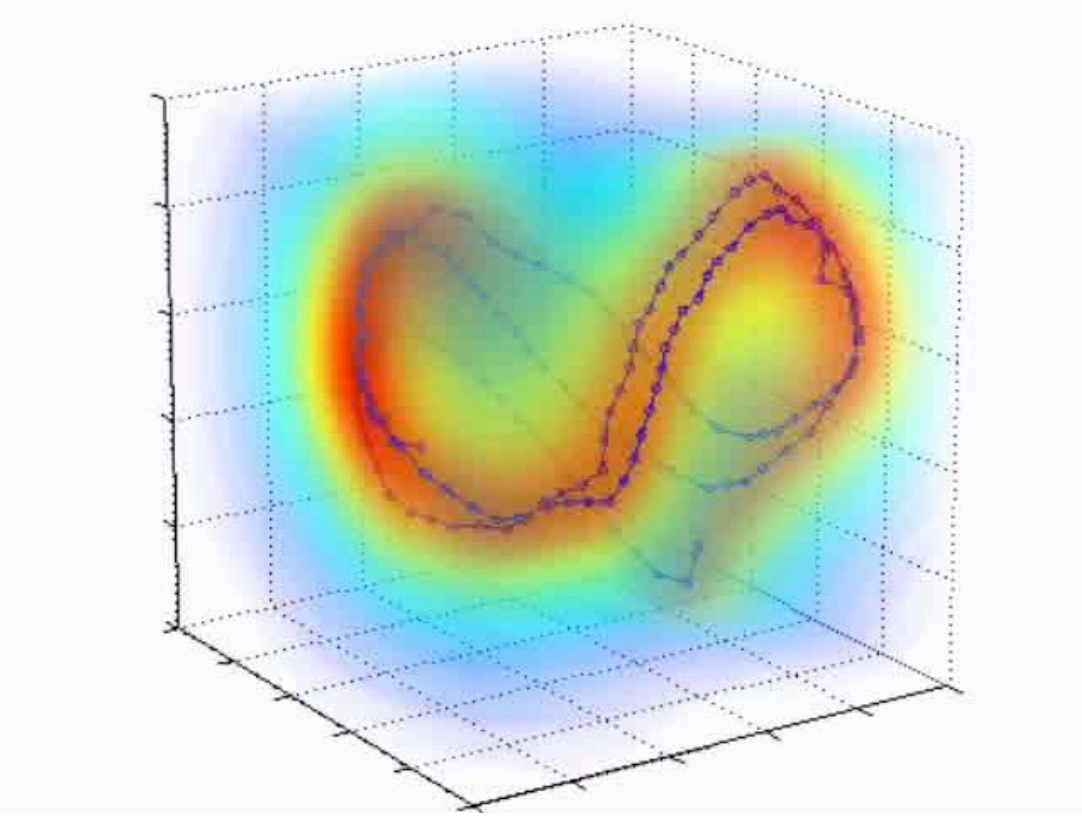


Football



Golf

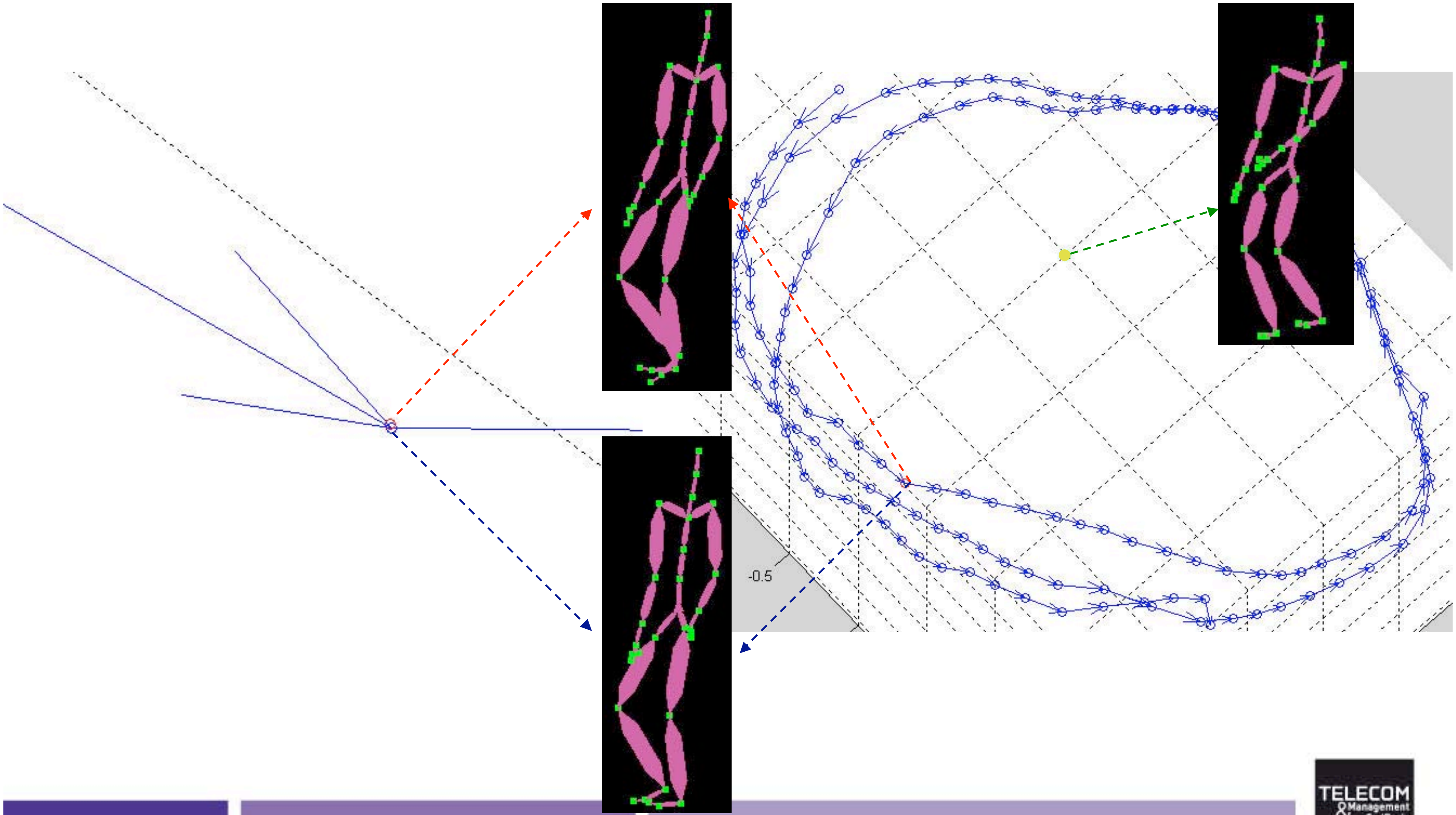
Reconstruction Variance



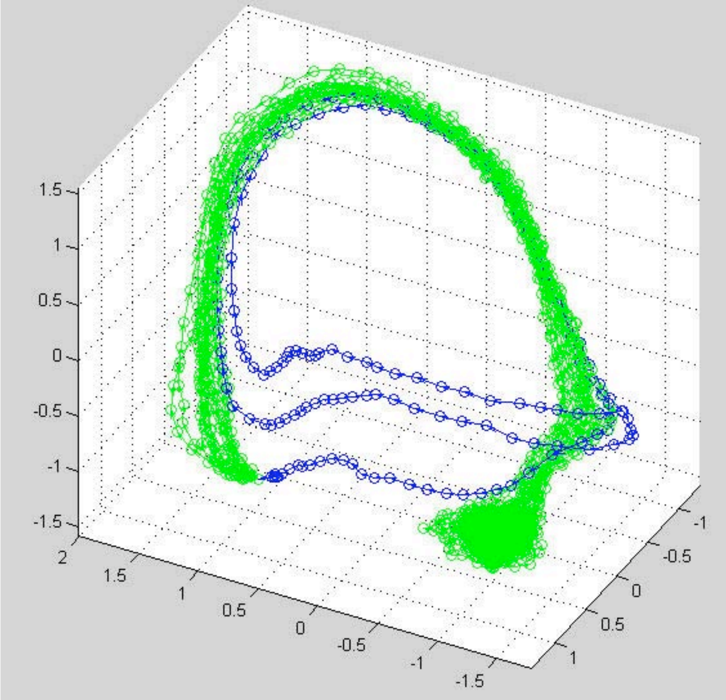
Volume visualization of $\ln \sigma^2_{\mathbf{y} \mid \mathbf{x}, \mathbf{X}, \mathbf{Y}}$

[Wang 05]

Reconstruction

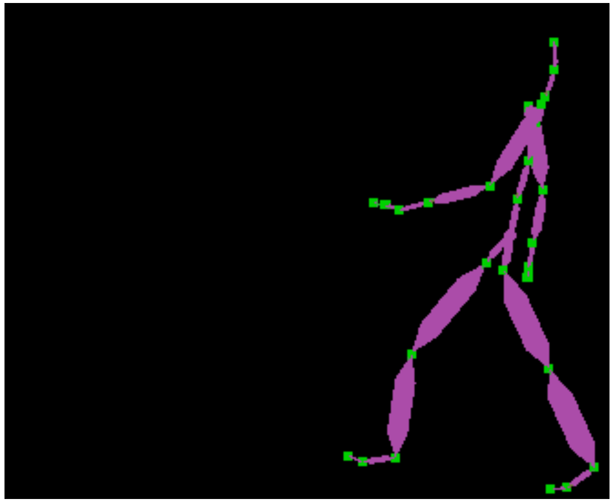


Motion Simulation

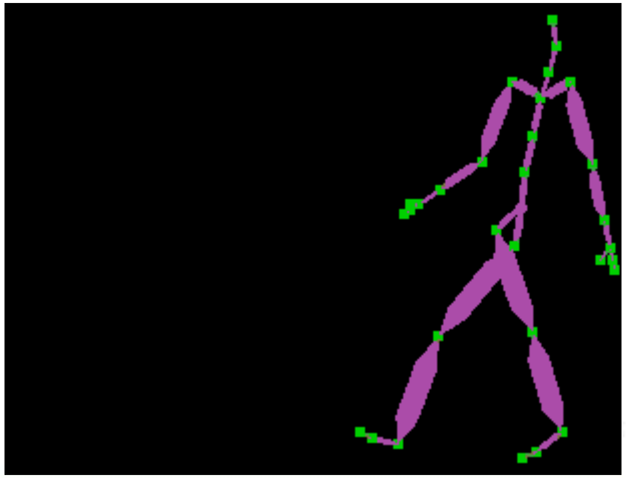


Random trajectories

(~1 gait cycle, 60 steps)

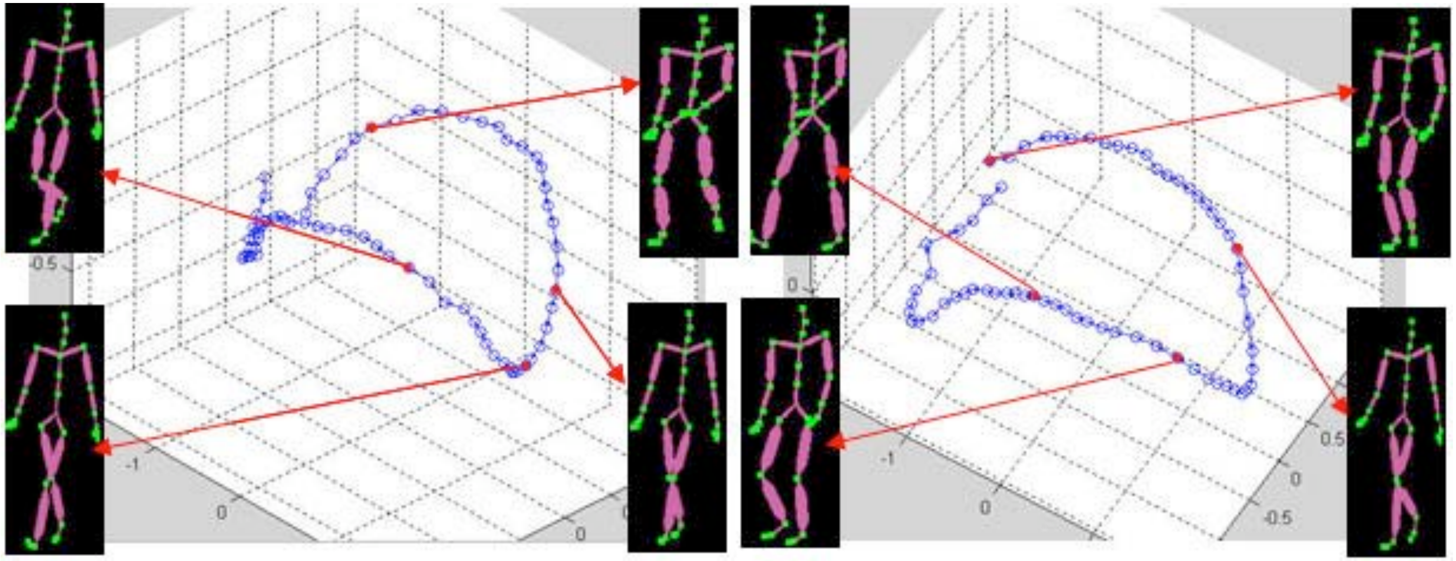


Original



Generated

Motion	Walking1	Walking2	Football1	Football2	Golf1	Golf2
Length	135	156	369	302	313	292



GPDM based motion analysis

- 3D curve moment invariants

GPDM based motion analysis

Motions	m_0	m_1	m_2	m_3
Walking 1	10.317	19.832	109.500	136.198
Walking 2	10.728	22.619	135.738	137.800
Walking 3	9.693	21.052	120.148	108.560
Football 1	19.507	60.025	1138.477	6738.425
Football 2	19.527	56.614	1023.861	5954.281
Golf 1	15.262	47.914	741.376	3705.910
Golf 2	16.295	51.833	854.212	4486.878
Golf 3	16.718	43.669	603.590	2650.057

3D curve moment invariants of different motions

GPDM based motion analysis

D	Walking 1	Walking 2	Walking 3	Football 1	Football 2	Golf 1	Golf 2	Golf 3
Walking 1	0	0.056	0.039	1.564	1.410	0.934	1.106	0.770
Walking 2	0.056	0	0.061	1.525	1.370	0.892	1.064	0.725
Walking 3	0.039	0.061	0	1.563	1.408	0.930	1.102	0.771
Football 1	1.564	1.525	1.563	0	0.164	0.642	0.469	0.827
Football 2	1.410	1.370	1.410	0.164	0	0.492	0.322	0.666
Golf 1	0.934	0.892	0.930	0.642	0.492	0	0.174	0.223
Golf 2	1.106	1.064	1.102	0.469	0.322	0.174	0	0.377
Golf 3	0.770	0.725	0.771	0.827	0.666	0.223	0.377	0

The Euclidean distance of different motions

Conclusion

- **Linear motion models provide a low-dimensional parameterization for simple motions.**

- **GPDM provide motion models from small training data sets.**

- **Motion models be used to constrain inference to plausible poses and motions.**

- **Perspectives:**
  - learn models with many different activities.
  - as prior models to tracking complex motions.

TELECOM & Management SudParis

Thank you.

28/04/08