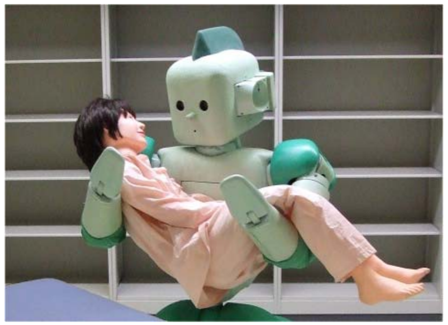
On-line Human Motion Prediction with Multiple Gaussian Process Dynamical Models

Takamitsu Matsubara[1,2]
Sang-Ho Hyon[2,3]
Jun Morimoto[2,3]

Graduate School of Information Science, NAIST[1]
Department of Brain Robot Interface, ATR-CNS[2]
ICORP-JST[3]

Human-robot interaction

- Importance of human-robot interaction in
  - Social interactions (Kanda et al. 2004)
  - Care of humans (Onishi et al. 2007)
  - Robot suits (Kawamoto et al. 2003)
  - Imitation learning (Nakaoka et al. 2005)


Our Goal

- Our goal
  - Assemble human-motion predictor (modeling)
    - Enables prediction in on-line manner (even learning is off-line)
    - Can also be used as motor primitive for imitation learning

- Difficulties in creating human behavior models
  - Nature of human motion such as:
    - Wide variety
    - High-dimensionality
    - Nonlinearity

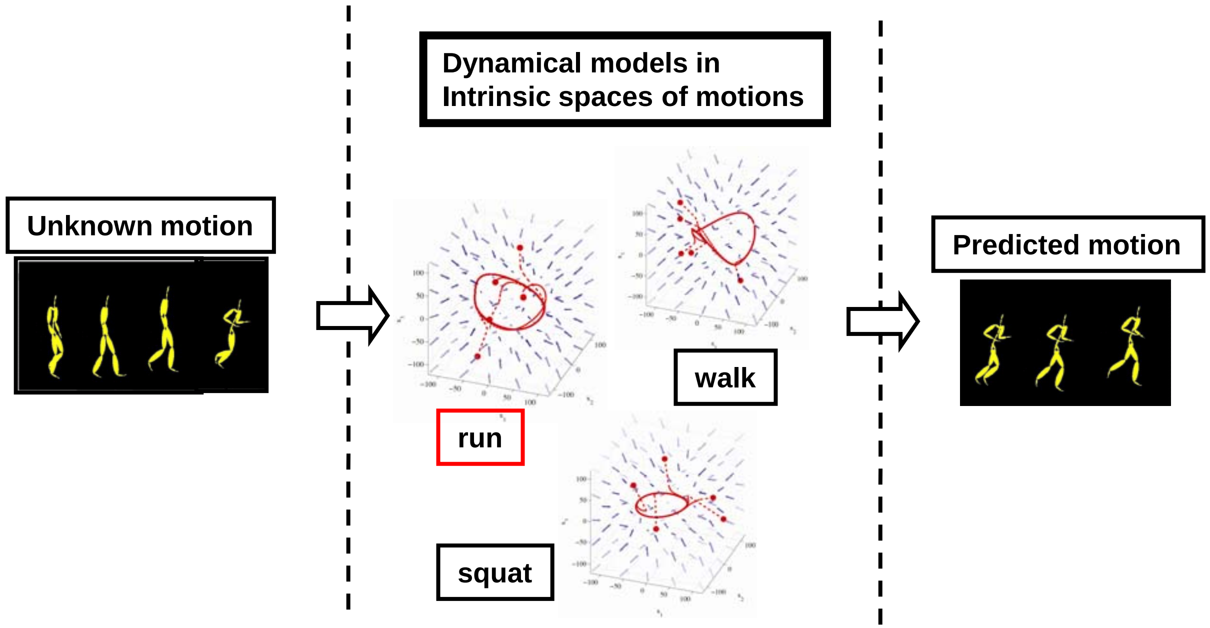
Related works

- Hidden Markov Models (HMMs)
  - Whole-body imitation between human and humanoid robot (Inamura et al. 2001)
    - Demonstrated learning and generation of whole-body motions
    - Requires discretization of poses (e.g. key pose)
  - Switching linear dynamical models (Pavlovic et al. 2000)
    - Potentially more powerful than simple HMMs due to its piece-wise continuous dynamics
    - Requires learning large number of parameters by an approximated manner

Learning for such complex parametric models with limited number of samples often suffers from over-fitting problems

Our approach

- Human-motion prediction:
  - Non-parametric modeling with Gaussian Processes
    - low-dimensional dynamics in latent space
    - Inverse-inference map (from observation to latent space)
    - Multiple models with on-line gating
  - Less suffering from over-fitting problem



Gaussian Process for regression 1/3

- Gaussian Process
  - Definition: *A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution* (Rasmussen and Williams 2006)
  - For regression…

Input: $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$

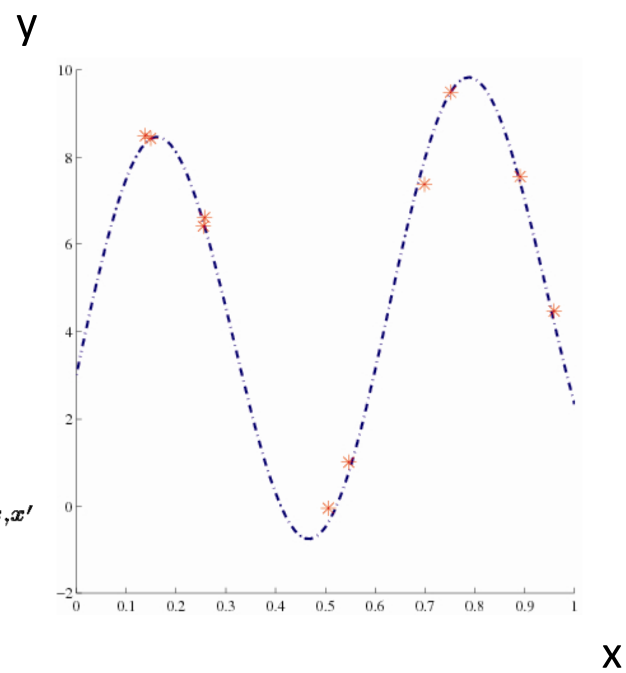
Output: $\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$

modeling

$$\mathbf{y} \sim \mathcal{N}\left(\mathbf{0}, K(\mathbf{x}, \mathbf{x})\right)$$

$$K(\mathbf{x},\mathbf{x}) = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_N) \\ & \ddots & \\ k(x_N,x_1) & & k(x_N,x_N) \end{bmatrix} \quad k(x,x') = \gamma_1 \exp\left(-\frac{\gamma_2}{2}||x-x'||^2\right) + \gamma_3^{-1}\delta_{x,x'}$$

$$\bar{\gamma} = (\gamma_1, \gamma_2, \gamma_3)^T$$



k(.,.): kernel function, $\bar{\gamma}$:hyper-parameters

(defines a correlation between two outputs based on attached inputs)

Gaussian Process for regression 2/3

- Prediction on query point (test data) x_*
  - Consider joint distribution

$$[\mathbf{y}, y_*]^T \sim \mathcal{N}\left(0, \left[\begin{array}{cc} K(\mathbf{x},\mathbf{x}) & k(\mathbf{x}, x_*) \\ k(x_*, \mathbf{x}) & k(x_*, x_*) \end{array}\right]\right)$$

Bayes thorem:

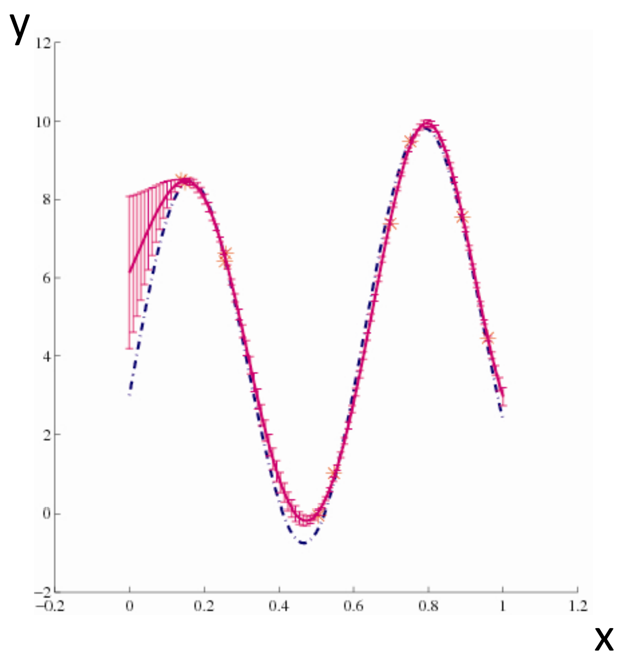
$$p(y_*|\mathbf{y}, \mathbf{x}, x_*) = \frac{p(\mathbf{y}, y_*|\mathbf{x}, x_*)}{p(y_*|x_*)}$$

  - Predictive distribution

$$y_* \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

$$\mu = \mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{y}$$

$$\sigma^2 = k(x_*, x_*) - \mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{k}_*$$

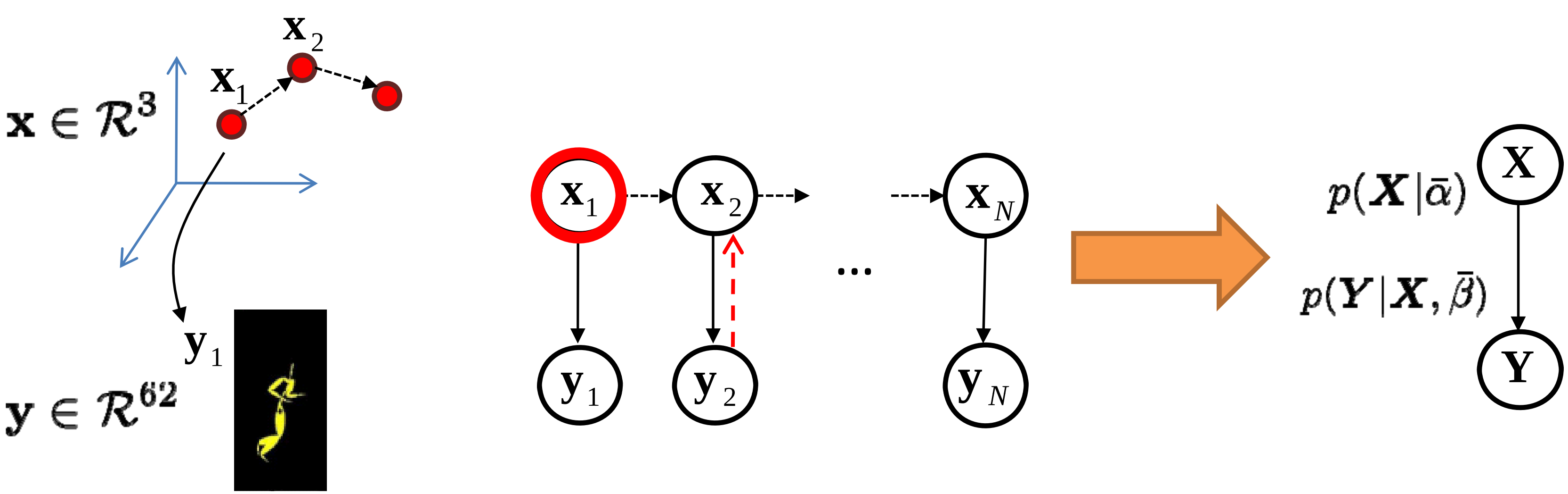


Gaussian Process for regression 3/3

- Why GP?
  - GP does not have "parameters" to be estimated
    - Interpreted as having been marginalized out
      → less suffering from over-fitting problem
    - No need to select appropriate complexity unlike parametric models
  - GP still has hyper-parameters, but they can be estimated by optimization of marginal likelihood (e.g., Mackay 1999)

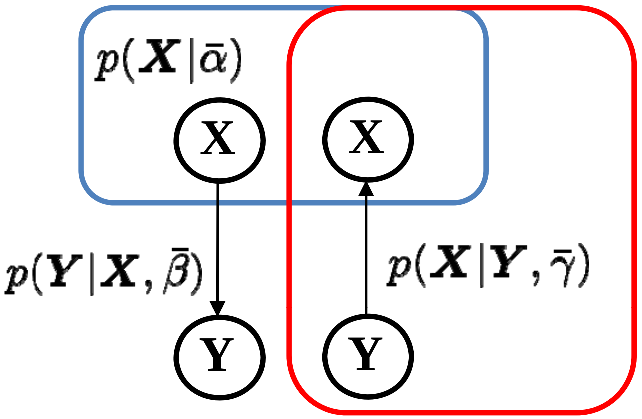
GPs for times series

- Guassian Process Dynamical Models (Wang et al. 2005)
  - Latent variable model for time series
  - Learning is to find MAP estimate of X, α and β
  - Predictions in both GPs can be made, but should be started from latent space…



Our approach

- Includes inverse inference map p(X|Y) as GP in advance and consider the three GPs simultaneously for learning X
  - Allows prediction in observation space through latent space dynamics
  - Idea (inverse-map, back-constraints) has been seen in GP-LVM (Shon et al. 2005, Lawrence et al. 2006)



Learning GPDM

- MAP estimation (Lawrence 2004, Wang et al. 2005)
  - Finds latent variables X and hyper-parameters to minimize the following objective function:

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

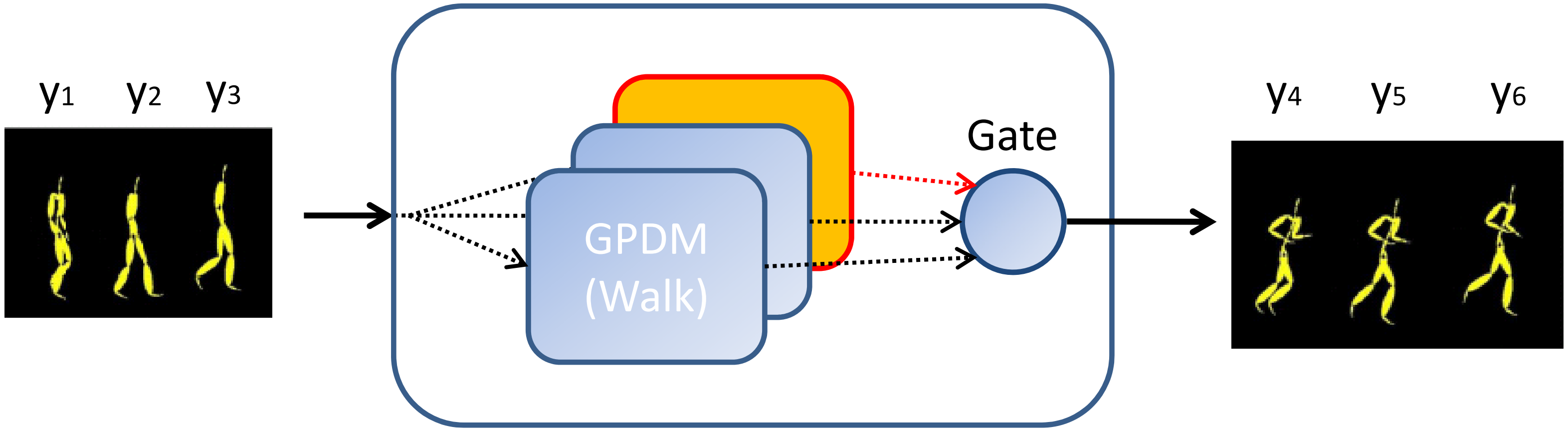
where, $p(\mathbf{Y}|\mathbf{X},\bar{\beta}) = \frac{1}{\sqrt{(2\pi)^{ND}|\mathbf{K}_y|^D}}\exp\left(-\frac{1}{2}\mathrm{Tr}(\mathbf{K}_y^{-1}\mathbf{Y}\mathbf{Y}^T)\right)$

$$k_y(\mathbf{x},\mathbf{x}') = \beta_1\exp\left(-\frac{\beta_2}{2}|\mathbf{x}-\mathbf{x}'|^2\right) + \beta_3^{-1}\delta_{\mathbf{x},\mathbf{x}'} \quad \bar{\beta} = (\beta_1,\beta_2,\beta_3)^T$$

  - Uses a conjugate gradient algorithm to find MAP estimate of X and hyper-parameters as in (Lawrence 2004)

For variety of human motions

- Simple modular structure
  - Prepare a variety of human behavior models as GPDM
  - On-line gating from a few observations
  - Best suited model predicts its future state

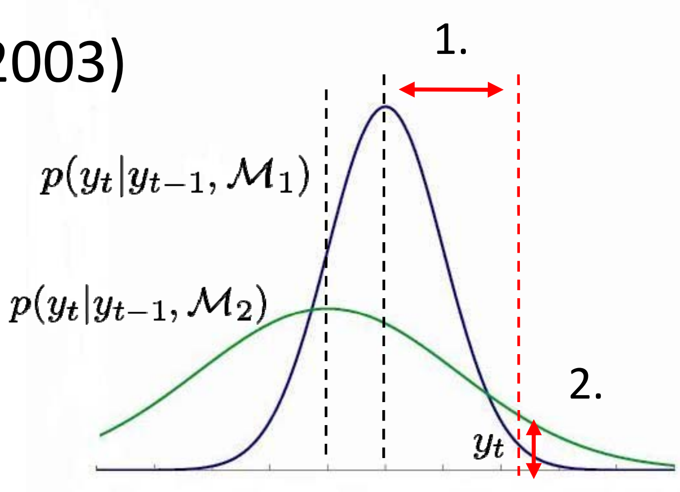


Criteria for on-line gating

1. Prediction error (squared error)
   - Easily achieved by using mean predictions of each GP
   - But, not effectively used variance information of predictive distributions
2. Approximated marginal likelihood
   - Criterion according to Bayesian Model selection

$$\hat{\mathcal{M}} = \arg\max_i p(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathcal{M}_i)$$

   - Gaussian approximation (Girard et al. 2003)



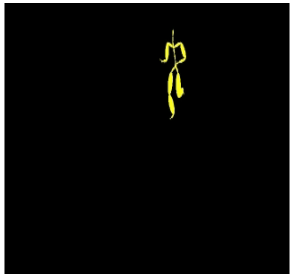
Modeling several human behaviors

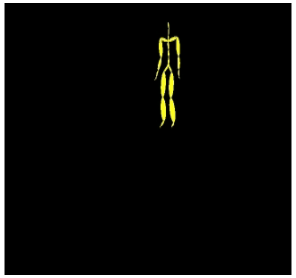
- Modeling by GPDMs
  - Data: CMU Graphics lab motion capture data base
    - 62 dim (56 joints and 6 for root position and orientation)
    - Segmented and labeled by hand
    - Use running, jumping and soccer-kicking
  - Latent space
    - Assume a first-order Markov dynamics in 3dim latent space
    - Square Exponential kernel function for all GP mappings
    - Initialize latent variables X by PCA
    - Set hyper-parameters by trial and error

Learned GPDMs

- Learned latent trajectories
  - Smooth and compact trajectories are obtained

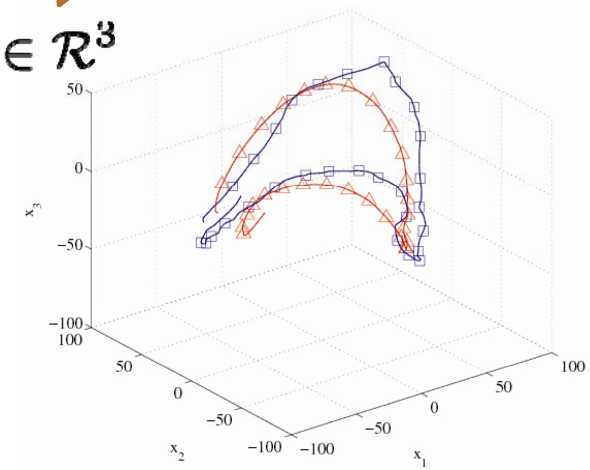
$\mathbf{y} \in \mathcal{R}^{62}$



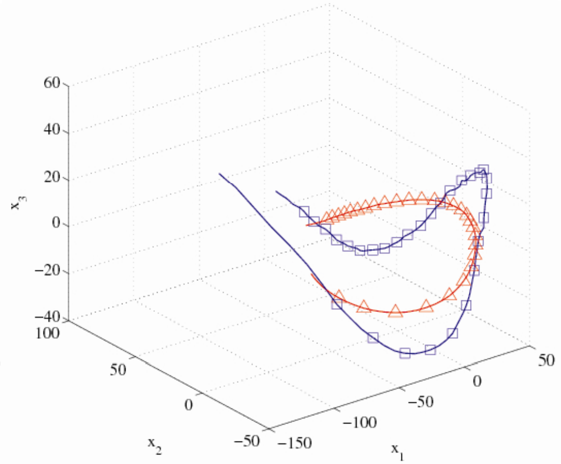


(From CMU Graphics Lab Motion Capture Database)

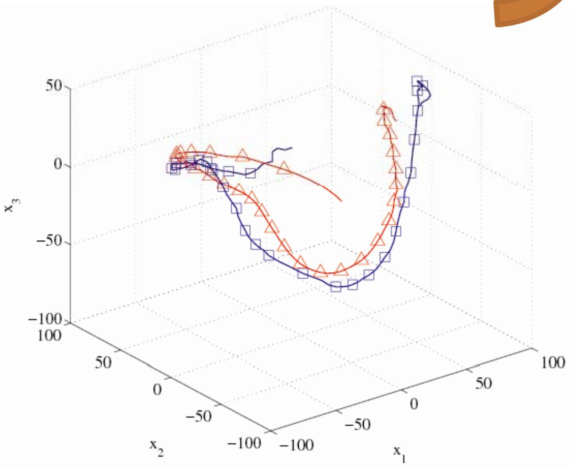
$\mathbf{x} \in \mathcal{R}^{3}$



Running (09_09.amc)



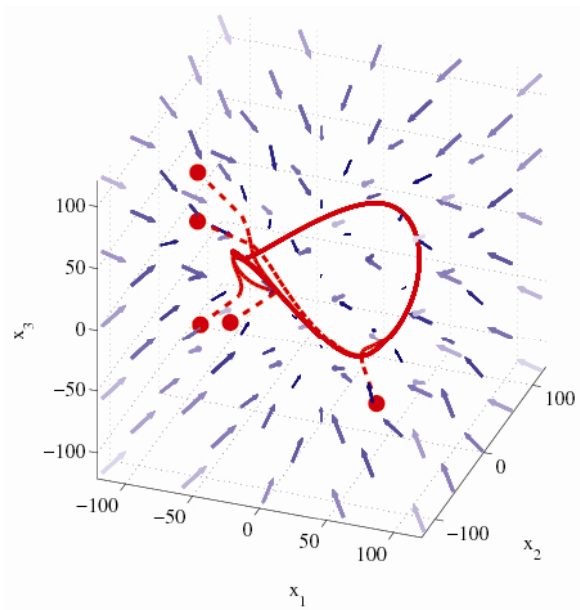
Jumping (16_09.amc)



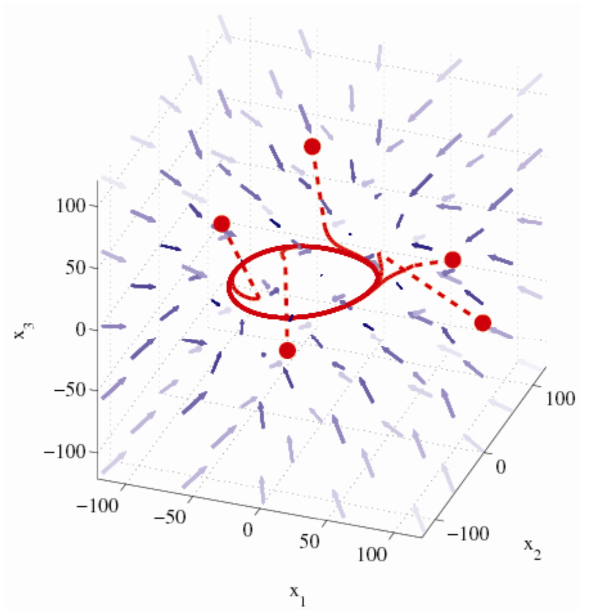
Soccer (10_03.amc)

Analyze latent dynamics

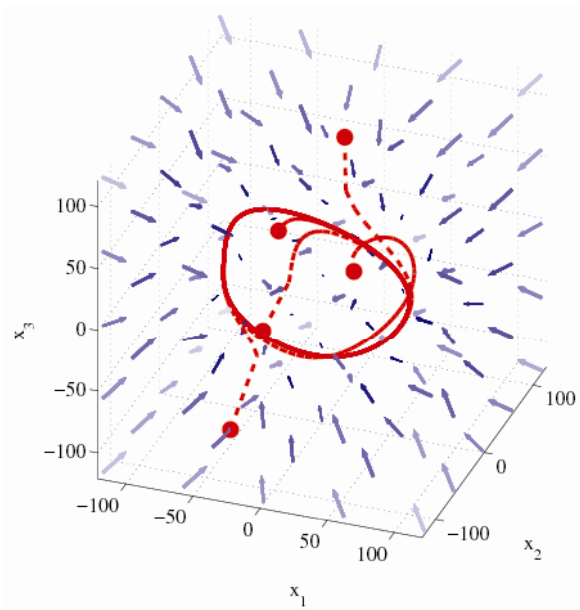
- Acquired latent dynamics
  - Smooth attractor dynamics over wide range of latent space
  - Plot latent dynamical GP prediction at several query points
    - Blue arrow (size: mean, shade: variance)
    - Red-dashed line (obtained by long term predictions)



Walking (09_09.amc)



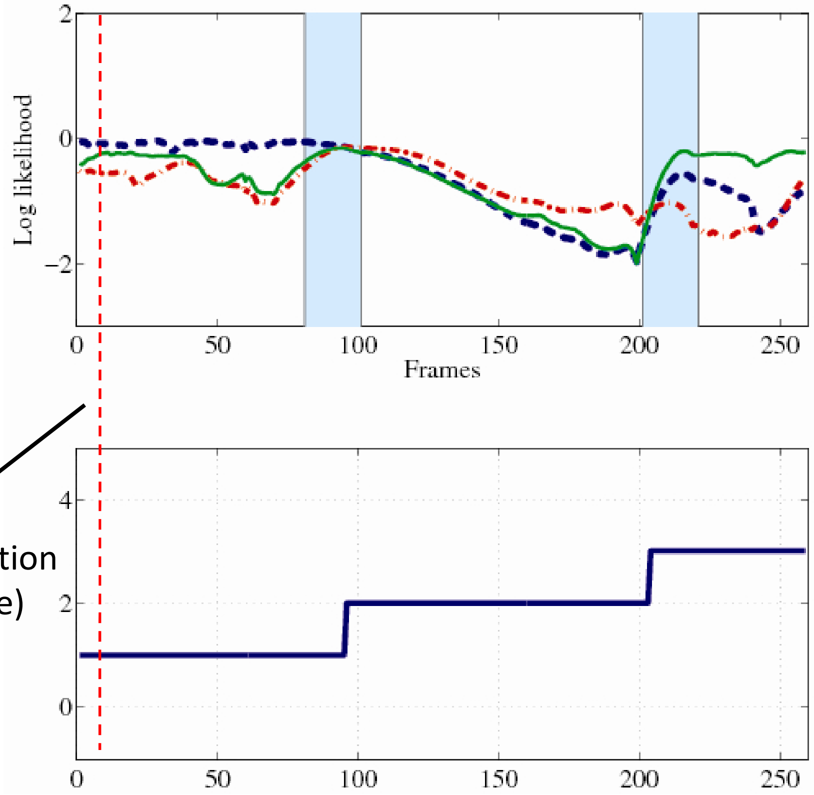
Jumping (16_09.amc)



Soccer (10_03.amc)

Human motion recognition by learned multiple GPDMs

- Recognitions by predictive error
  - Smoothly synthesized human motion data
    - composed of three behaviors (test trial data of same subjects)
      - Running (09_08.amc), Jumping (13_13.amc), Soccer (10_03.amc)

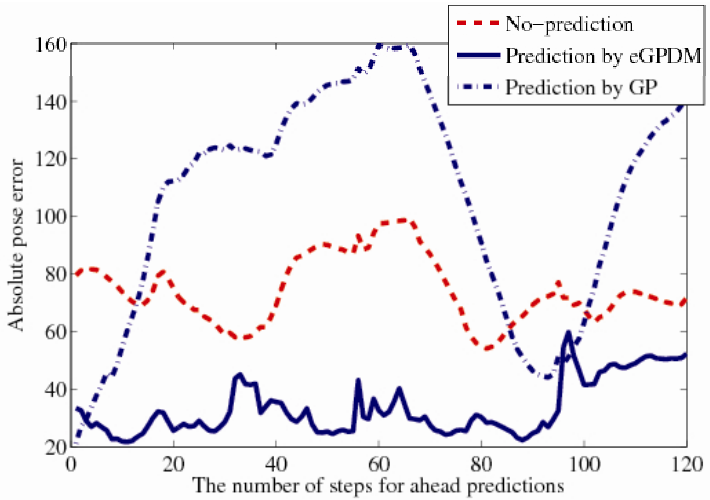


• Starting point for long term prediction (Results are presented in next slide)

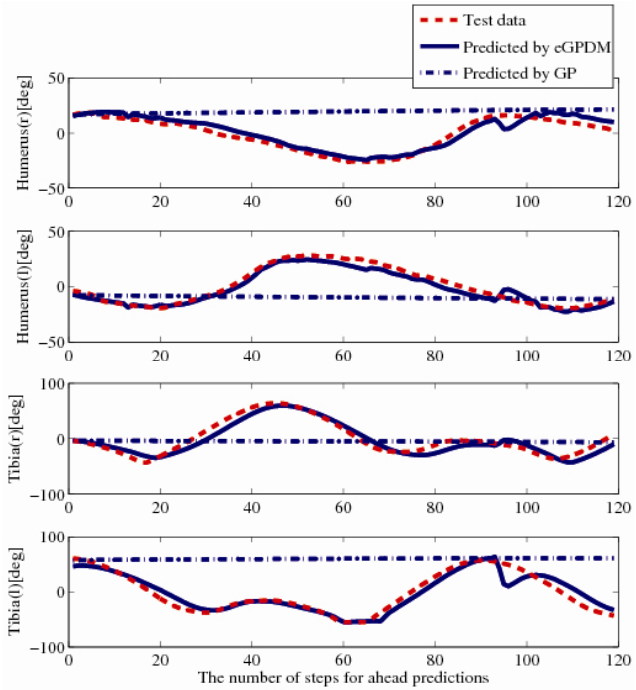
Simple predictive error effectively works for recognition

Long-term prediction for test data

- Long-term prediction from test data (running)
  - 1 sec prediction for whole body motion
  - Compare with a GP learns dynamics in observation space



Pose error at every frame



Principal joints in long-term prediction

The proposed method is effective for long term prediction

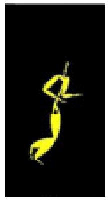
Effectiveness of low-dimensional latent space for smooth dynamics(1/2)

- What is advantage of considering latent space?
  - Comparison: learns dynamics directly in observation space by GP (running case), presents in 3dim space by PCA

The comparison:

$\mathbf{y} \in \mathcal{R}^{62}$

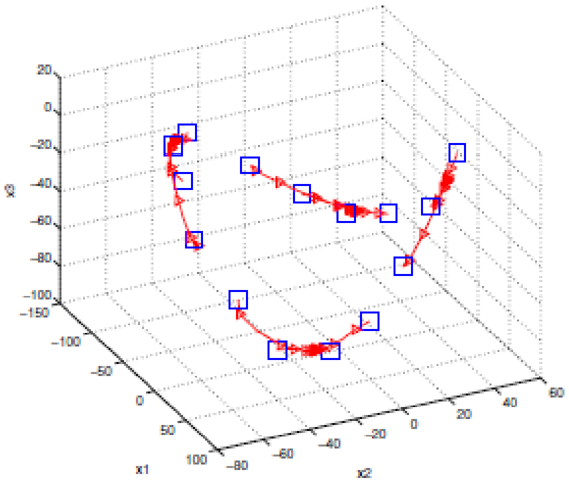
$$\mathbf{y}_1 \to \mathbf{y}_2 \to \cdots \to \mathbf{y}_N$$



GP modeling

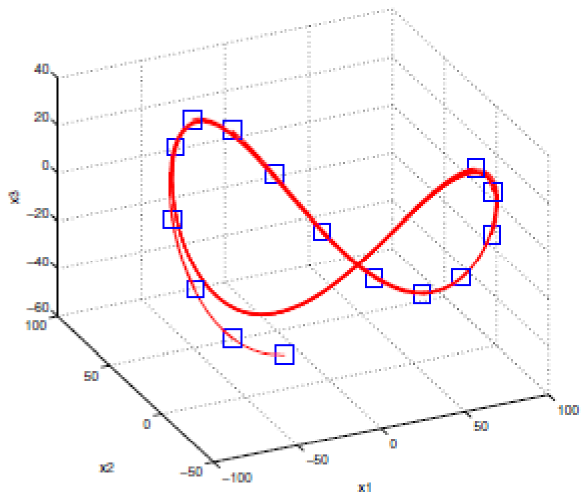
$$p(\mathbf{Y}_{2:N} | \mathbf{Y}_{1:N-1}, \bar{\beta})$$

Long-term Prediction in 3-dim space by PCA

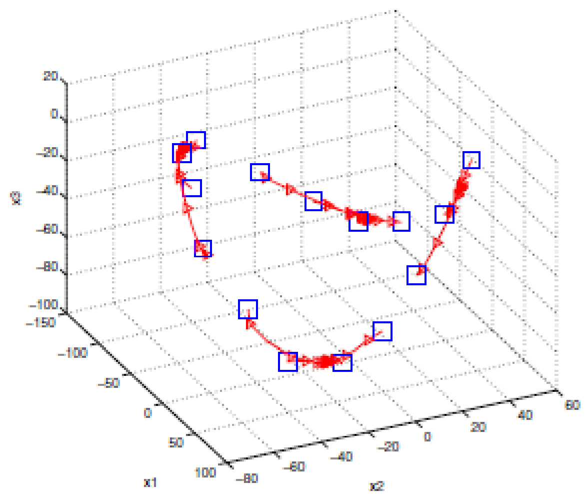


Effectiveness of low-dimensional latent space for smooth dynamics(2/2)

- What is advantage of considering a latent space?
  - Compare both in latent space (projected by PCA)



GPDM



GP (comparison)

Comparison could not make smooth dynamics
due to its high-dimensic

Conclusions and future work

- Explored multiple GPDMs and on-line gating criteria for on-line human motion prediction
- Demonstrated its basic effectiveness for simple experiments

- Auto segmentation of the human behavior data
- Auto selection of representative behavior models in daily lives