KDD-09 PARIS • June 28th - July 1st 2009

The 15th ACM SIGKDD Conference On Knowledge Discovery and Data Mining

DynaMMo: Mining and Summarization of Coevolving Sequences with Missing Values

Christos Faloutsos

joint work with *Lei Li, James McCann, Nancy Pollard*

June 29, 2009

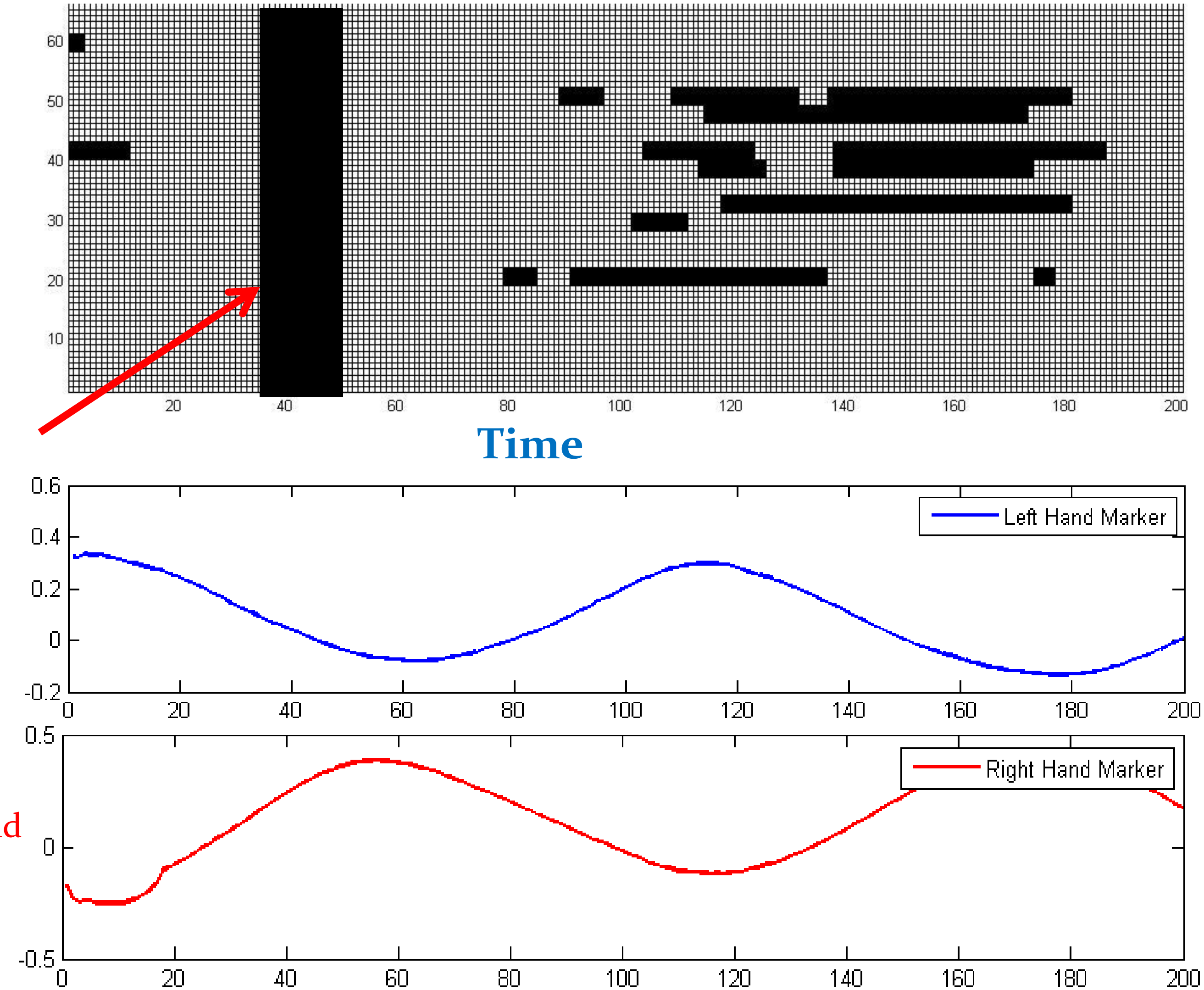
@Carnegie Mellon Databases

CHALLENGE

- Multidimensional coevolving time series:
  - Motion Capture sequence
  - Temperature/humidity monitoring
  - Daily Chlorine level measurement in drinking water system
- Big challenge:
  - Missing observations
  - Mining with missing values
    - Find hidden patterns
    - Use of hidden patterns
      - Forecasting
      - Compression
      - Segmentation
      - Clustering
      - and more …

Missing values in Mocap

Marker/Sensor



blankout

Time

Left Hand position

Right Hand position

GOAL

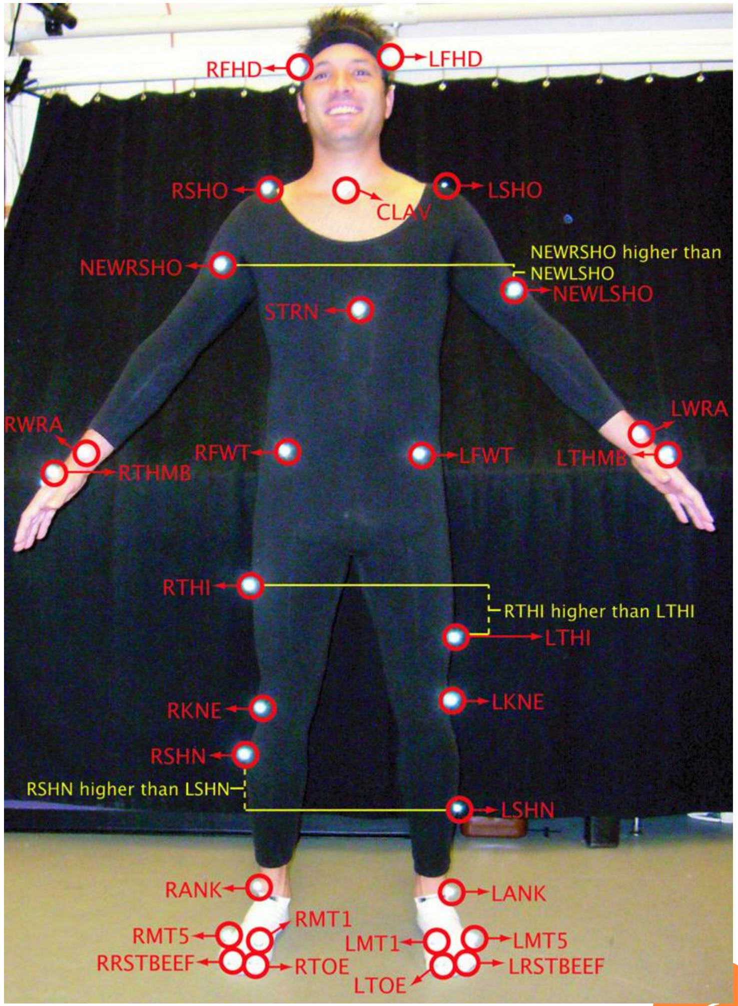
- We want recovering, mining and summarization algorithms to be:
  1. Effective: low reconstruction error, agreeing with human intuition (e.g. natural reconstructed motion for mocap)
  2. Scalable: to time-duration T of the sequences.
  3. Black-outs: It should be able to handle “black-outs”, when all markers dissappear (eg., a person running behind a wall, for a moment).
  4. Automatic: The method should require no parameters to be set by the user.

OUTLINE

- Scenario and Motivation
- Proposed Methods – DynaMMo
  - Recovering missing values
  - Compression and summarization
  - Forecasting
  - Segmentation
- Experimental Results
- Conclusion

Scenario: Motion Capture

- Motion Capture:
  - Markers on human actors
  - Cameras used to track the 3D positions
  - Duration: 100-500
  - 93 dimensional body-local coordinates after preprocessing (31-bones)
- Challenge:
  - Occlusions
- Other general scenario:
  - Missing value in Sensor data: Out of battery, transmission error, etc
  - Unable to observe, e.g. historical/future observation



From mocap.cs.cmu.edu

Observation and Motivation

- Dynamics: temporal moving pattern
- Correlation among multiple markers

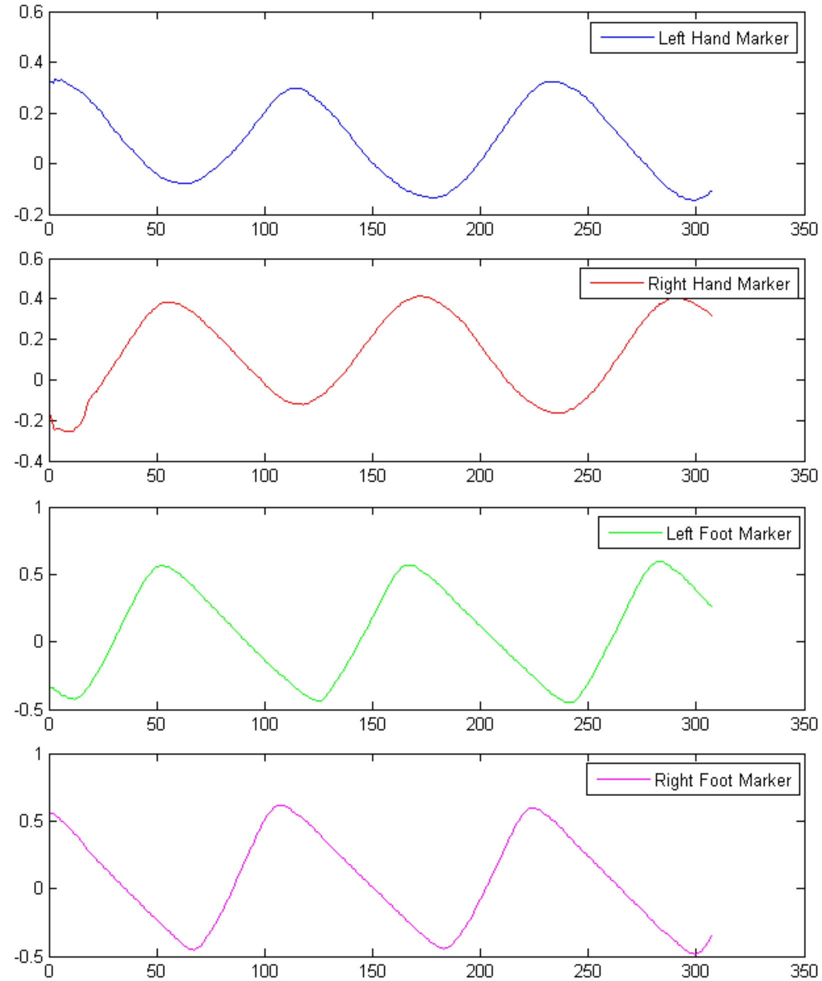
Use both dynamics and correlation to solve occlusion.

Left Hand

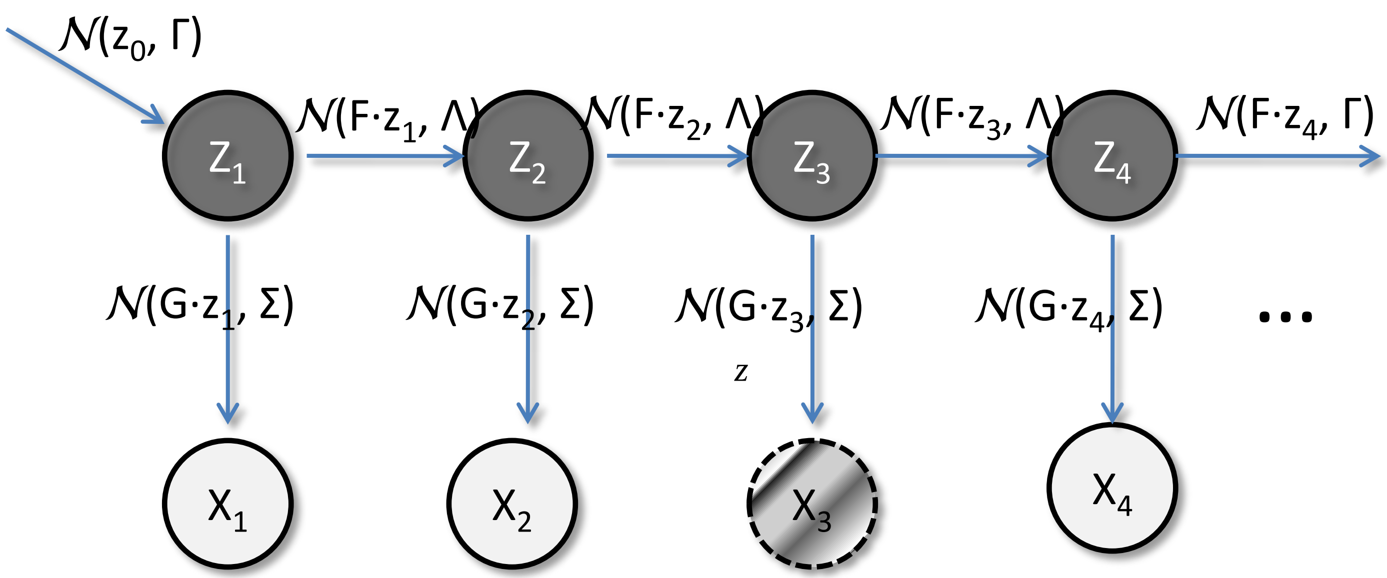
Right Hand

Left Foot

Right Foot



The Underlying Time Series Model
Linear Dynamical Systems



Model parameters:

$\theta=\{z_0, \Gamma, F, \Lambda, G, \Sigma\}$

$$z_1 = z_0 + \omega_0$$
$$z_{n+1} = F \cdot z_n + \omega_n$$
$$x_n = G \cdot z_n + \varepsilon_n$$

DynaMMo Recovering Algorithm

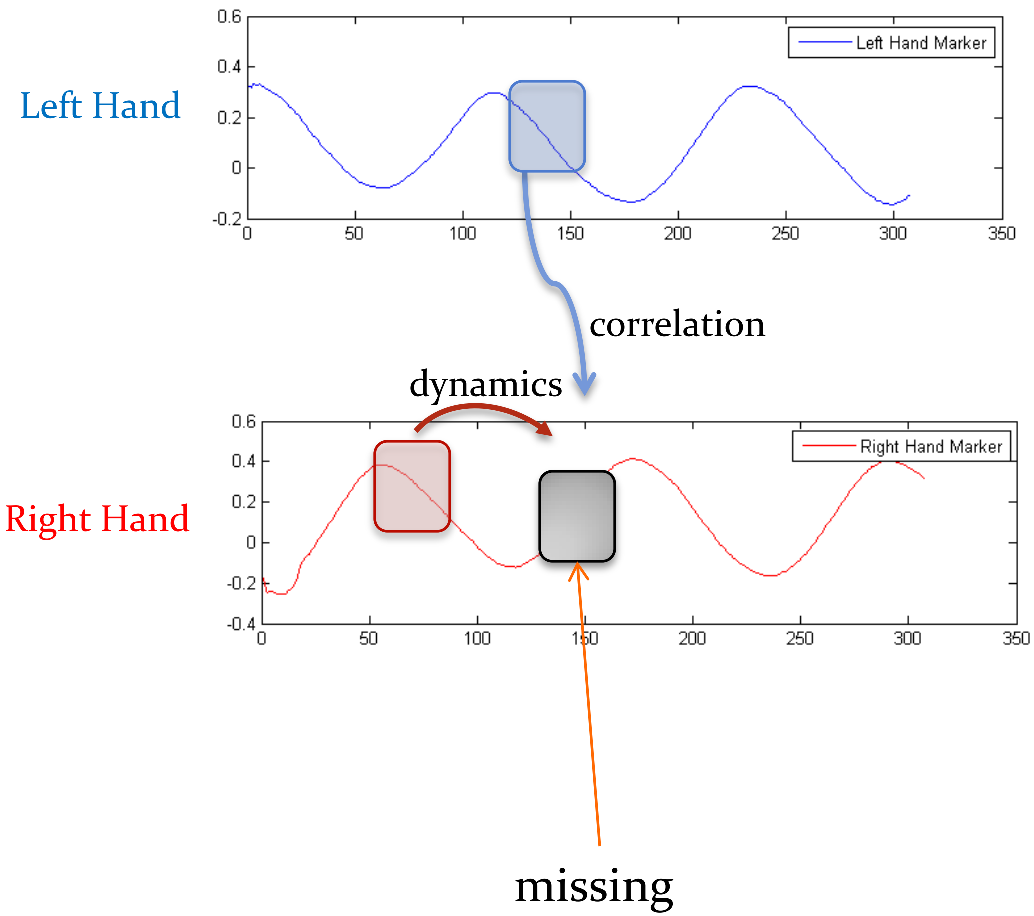
- Expectation Maximization
- Intuition:

  Finding the best model parameters (θ) and missing values for X to minimize the expected loglikelihood:

$$
\begin{aligned}
Q(\theta) \quad = \quad & \mathbb{E}_{\mathcal{X}_m,\mathcal{Z}|\mathcal{X}_g,\mathcal{W}}[-D(\mathbf{z}_1, z_0, \Gamma) \\
& - \sum_{t=2}^{\mathrm{T}} D(\mathbf{z}_t, \mathbf{F}\mathbf{z}_{t-1}, \Gamma) \\
& - \sum_{t=1}^{\mathrm{T}} D(\mathbf{x}_t, \mathbf{G}\mathbf{z}_t, \Sigma) \\
& - \frac{1}{2}\log|\Gamma| - \frac{\mathrm{T}-1}{2}\log|\Lambda| - \frac{T}{2}\log|\Sigma|]
\end{aligned}
$$

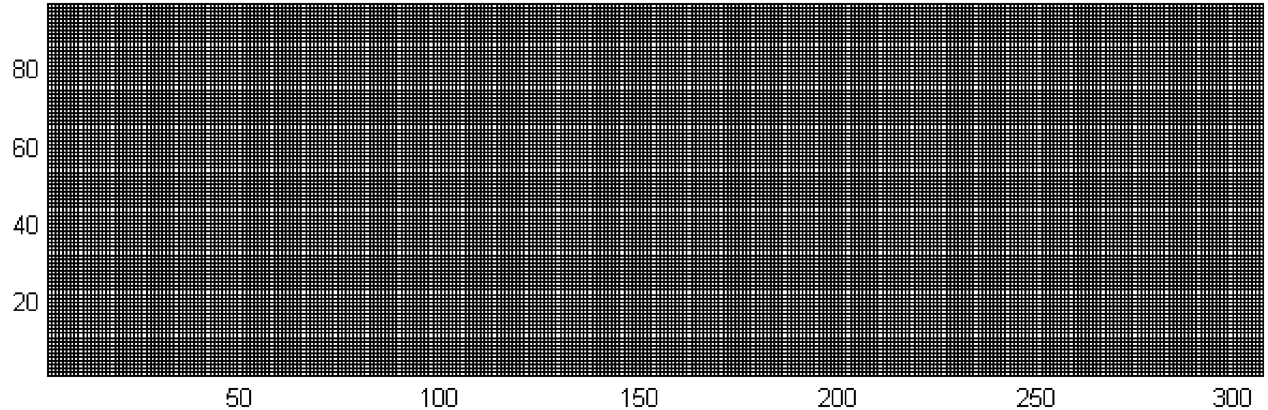
See details in paper

DynaMMo Intuition:

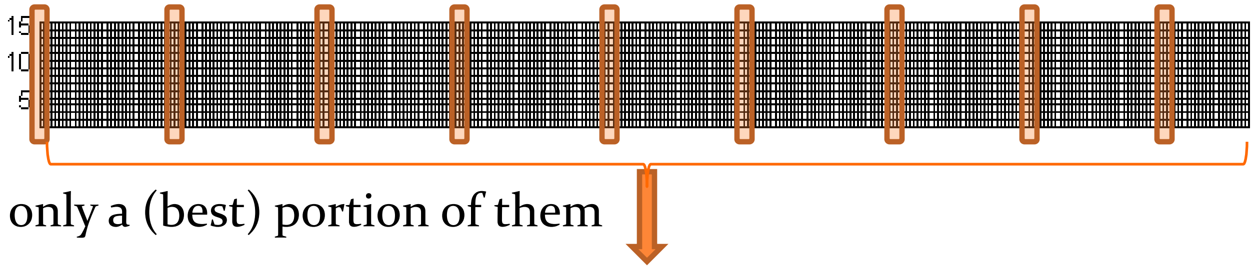


DynaMMo Compression: Intuition

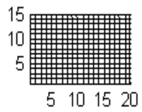
observations w/ missing values



get hidden variables and model parameters



keep only a (best) portion of them

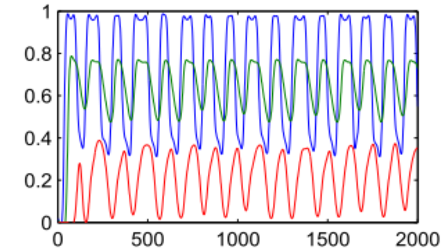


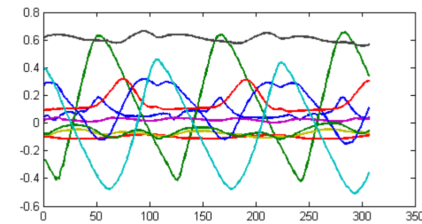
and model parameters

Same idea could be used in segmentation and forecasting

Experiment

- Dataset:
  - Chlorine: Chlorine level in drinking water system
    - Duration 4310 time ticks
    - 166 sequences
  - Mocap: full body human motion capture dataset
    - 58 motions
    - each with duration 100-500, 93 dimensions
    - marker positions in body local coordinates
- Occlusion: random mask out
- Baseline:
  - linear interpolation and spline
  - MSVD:
    - Missing value SVD algorithm
    - EM flavored version of SVD.



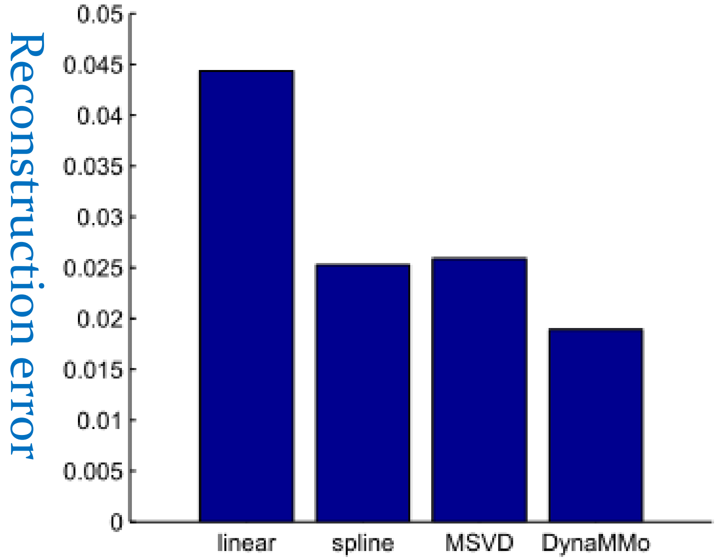


Results

- Reconstruction Error for random mask out
- Scalability: computation time to duration
- Forecasting case study
- Compression: error versus space
- Segmentation for synthetic and real data

DynaMMo Reconstruction Result
(average over 10 repeats)

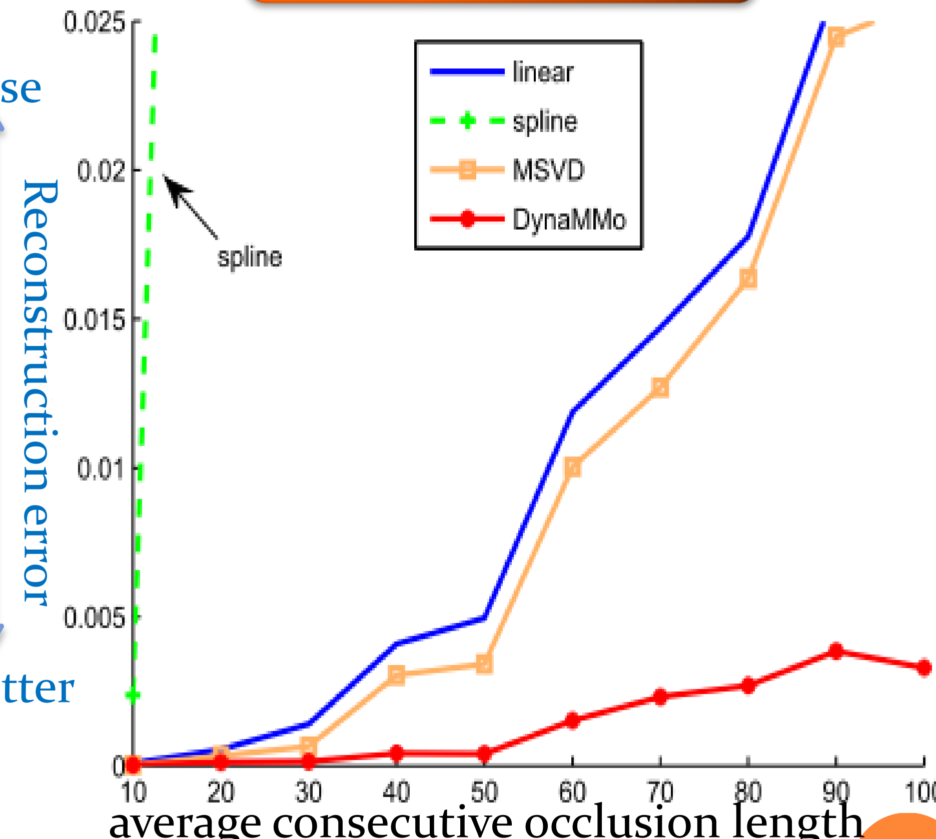
Chlorine dataset



worse

better

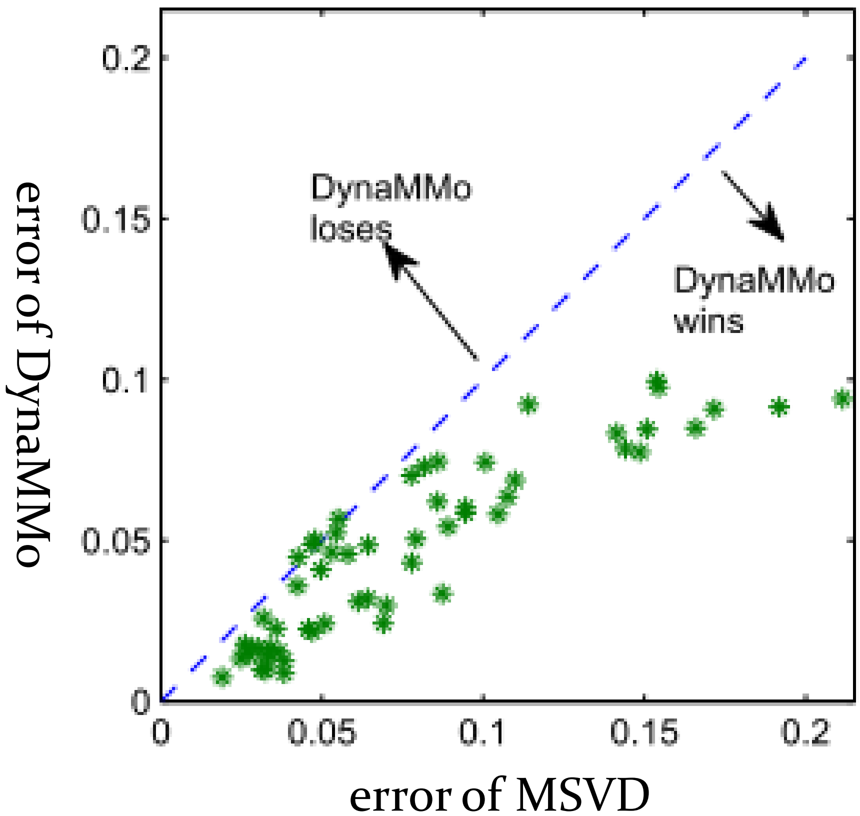
Mocap dataset



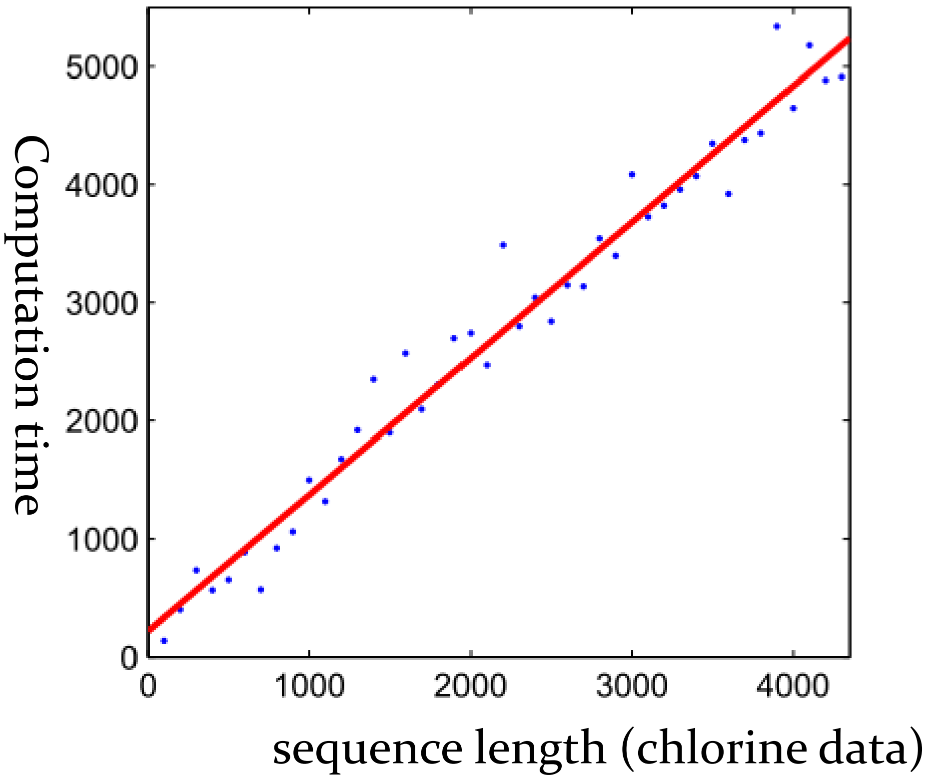
easy

difficult

SCATTER COMPARISON: DYNAMMO VS MSVD



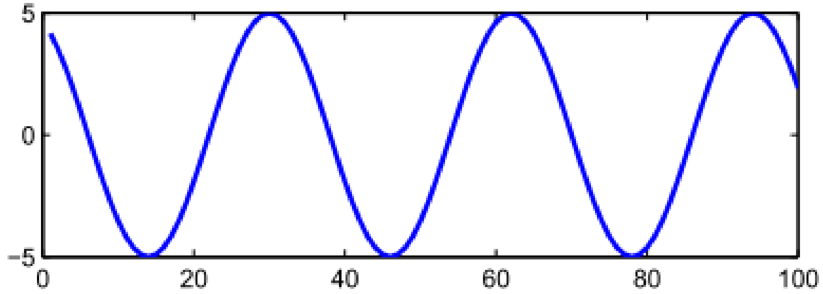
DynaMMo Scalability



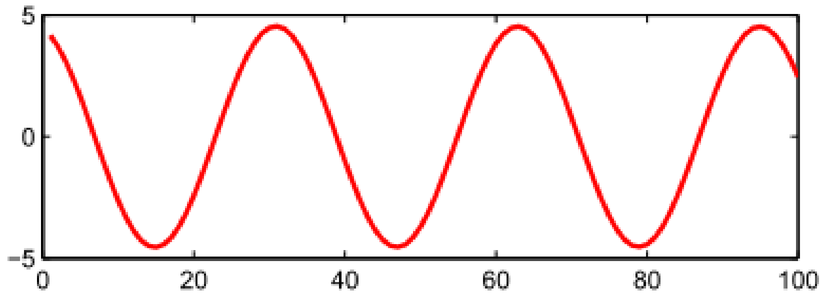
Li *et al.* **DynaMMo**: Mining and Summarization of Coevolving Sequences with Missing Values

DynaMMo Forecasting

Actual Data



Predicted signal using learned model

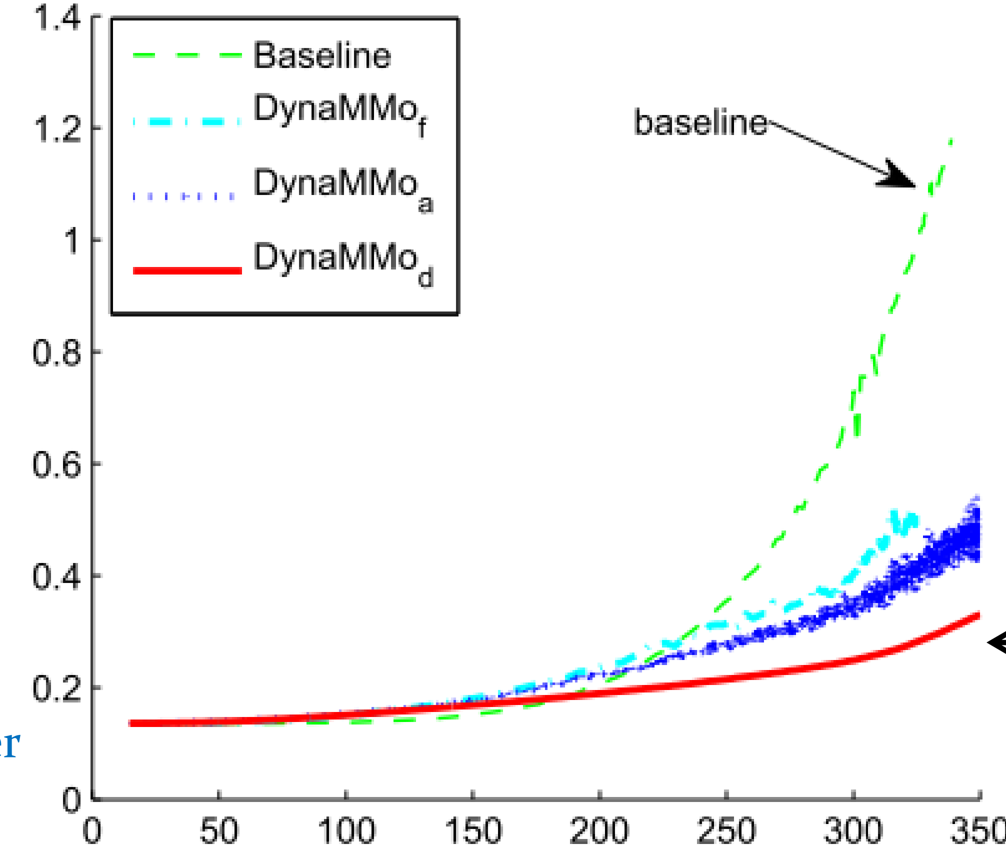


DynaMMo Compression

worse

error

better



Compression ratio

DynaMMo optimal compression

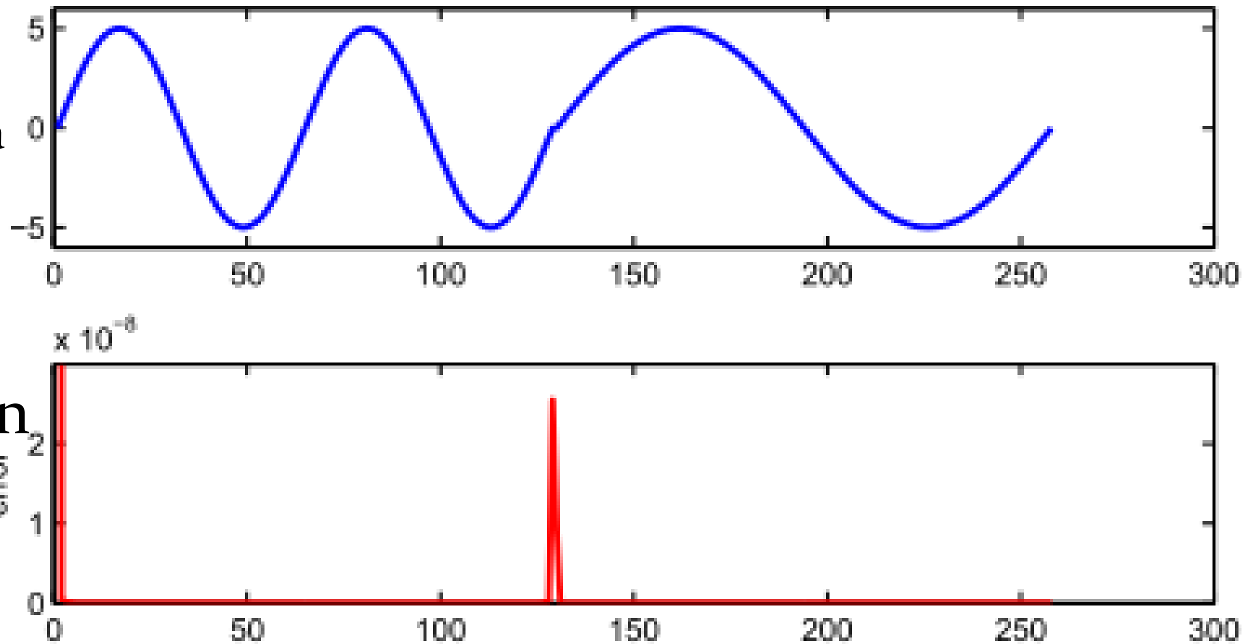
more space

less space

DynaMMo Segmentation

original data

reconstruction error

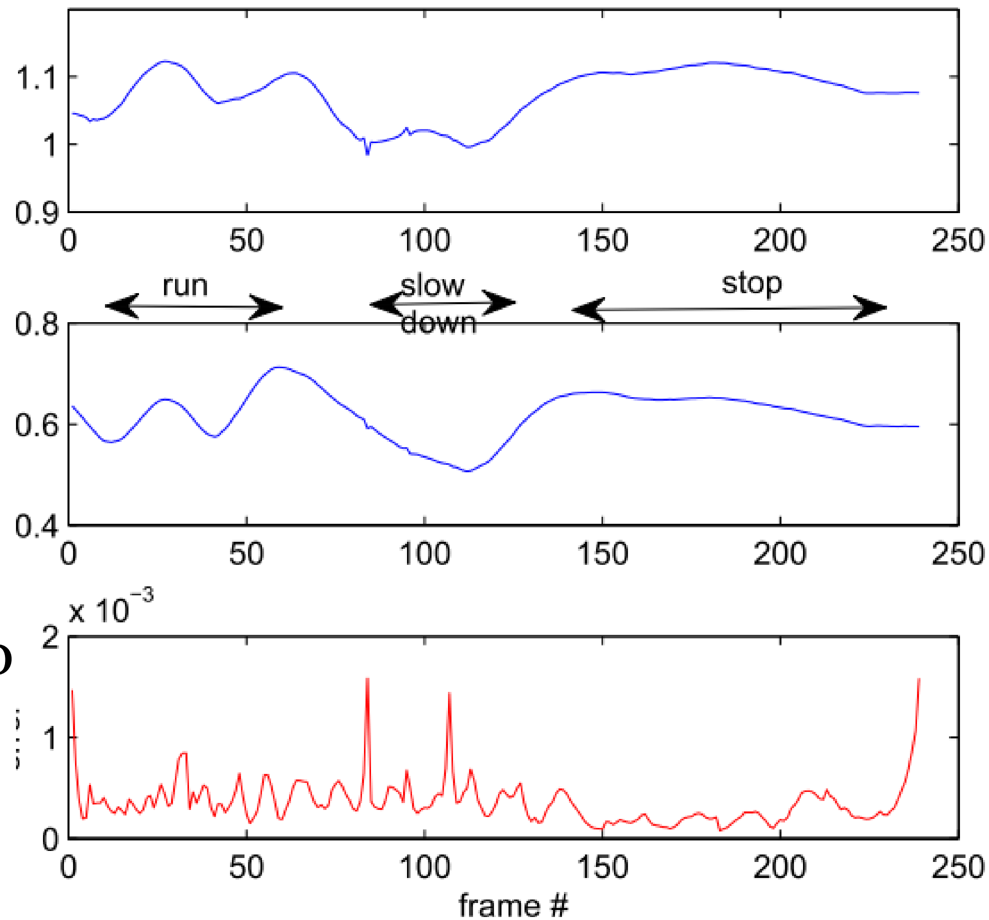


Mocap Segmentation
Running Transition motion (mocap#16.8)

left hip

left femur

reconstruction error



Li *et al.* **DynaMMo**: Mining and Summarization of Coevolving Sequences with Missing Values

RELATED WORK

- Time series representation and Indexing
  - using trajectory features (e.g. velocity), [Mehta, Parthasarathy, Machiraju, 06]
  - Symbolic representation (SAX)[Lin, Keogh, Lonardi, Chiu, 2003], iSAX[Shieh, Keogh, 2008],
  - uniform scaling indexing, [Keogh, Palpanas, Zordan, Gunopulos, Cardle, 2004]
- Time series classification
  - Skew distribution and concept shifts, [Gao, Ding, Fan, Han, Yu, 2008]
- Outlier detection
  - TARDO:sub-trajectory anomaly detection, [Lee, Han, Li, 2008]
- Missing value recovery
  - interpolation (e.g. spline) and autoregression models
  - PCA [Park, Hodgins, 2006]
  - Missing Value SVD [Srebro, Jaakkola, 2003]
  - mixture of local linear model [Liu, McMillan, 2006]
  - Gaussian process [Lawrence, Moore, 2007]
  - Human motion specific models, e.g skeleton based [Herda, Fua, PlÄankers, Boulic, Thalmann, 2000]

Contribution

- We propose algorithms DynaMMo for
  - Recovering missing values
  - Compression and summarization
  - Forecasting
  - Segmentation
- DynaMMo meets all goals:
  1. Effective: low reconstruction error, agreeing with human intuition (e.g. natural reconstructed motion for mocap) ✓
  2. Scalable: computation time linear to length/duration T of the sequences. ✓
  3. Black-outs: able to handle "black-outs", when all markers dissappear. ✓
  4. Automatic: The methods should require few parameter to be set by the user. ✓

QUESTION

- Thanks!
- Contact: leili@cs.cmu.edu

Christos Faloutsos

Lei Li

Jim McCann

Nancy Pollard