Manifold Alignment of High-Dimensional Datasets

PIs: Sridhar Mahadevan, Rui Wang

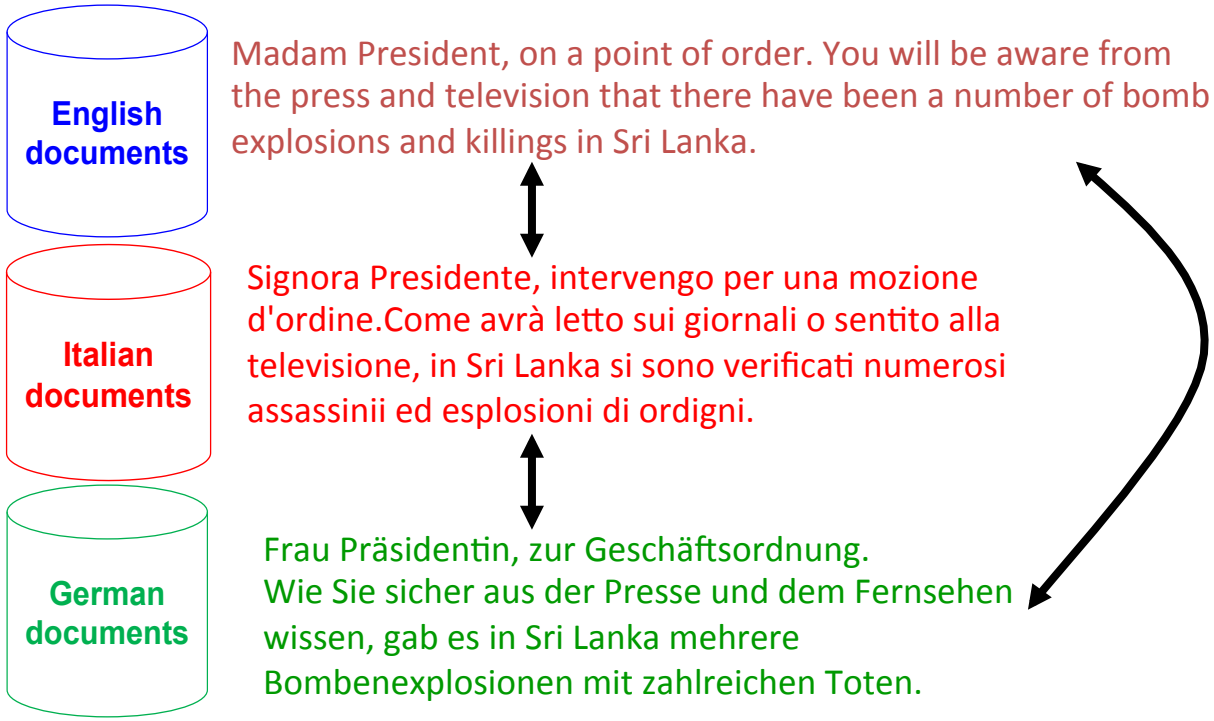
Blake Foster, Armita Kaboli, Peter Krafft,
Chang Wang (IBM Research)

Department of Computer Science
University of Massachusetts, Amherst

Outline

- Problem: transfer learning across domains
  - How to transfer knowledge across tasks?
- Solution:
  - Find manifold-based projections of original data
  - Align projected data in lower-dimensional space
- Applications:
  - Cross lingual IR, activity recognition, reinforcement learning
- Acceleration using GPU

Example: Cross-Lingual Retrieval

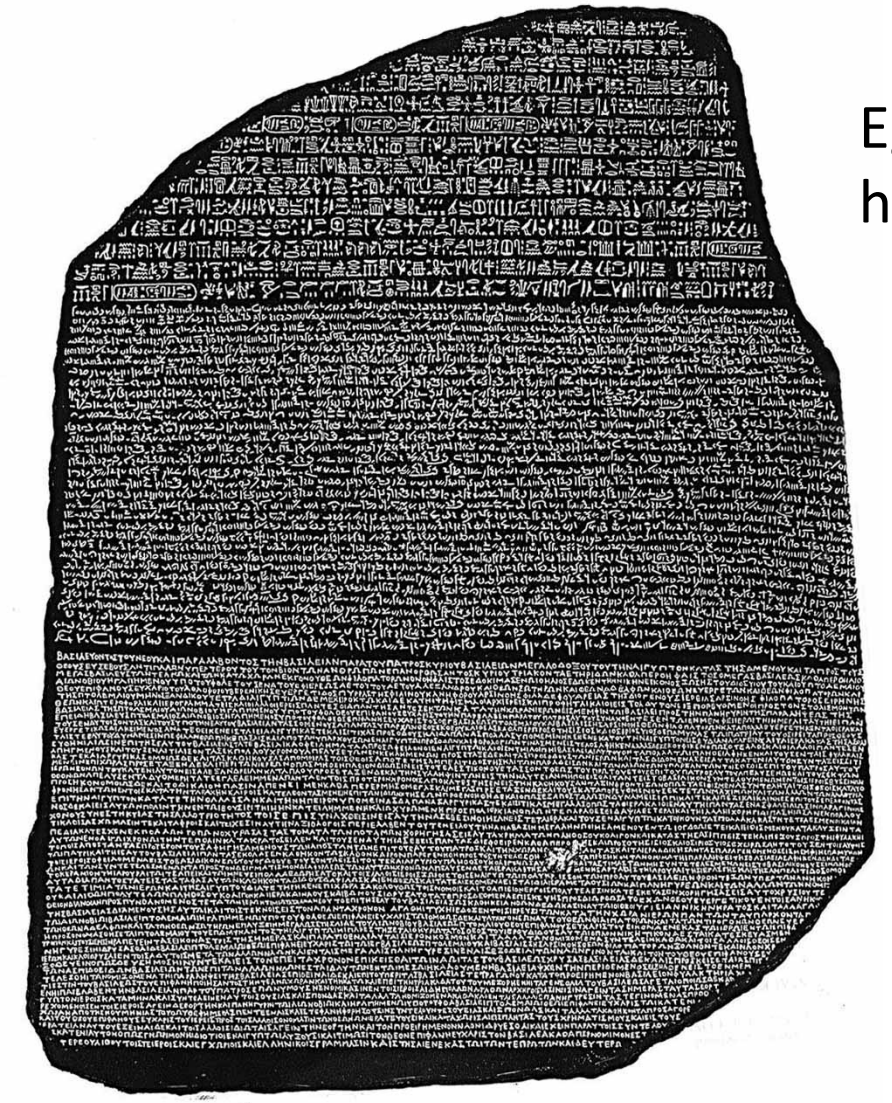


Rosetta Stone

Carved in 196 BC

Discovered in 1799 AD

Paved the way to decipherment of hieroglyphics

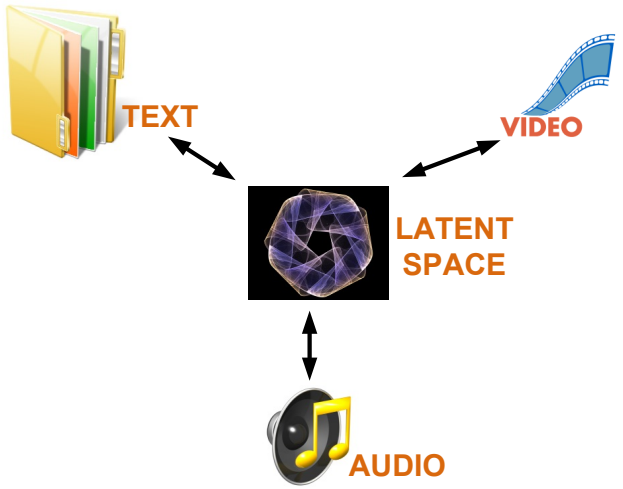


Egyptian hieroglyphics

Demotic

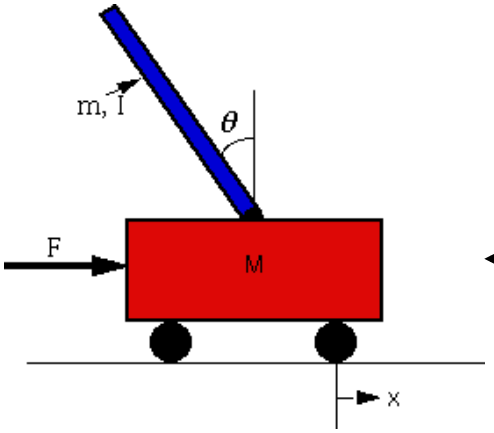
Greek

Knowledge Transfer



flowers, grass, tiger, water

Knowledge Transfer in Reinforcement Learning



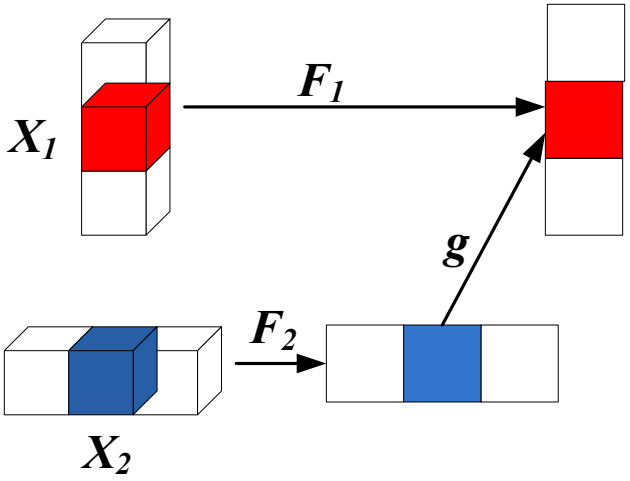
Why is Transfer Difficult

- Data is usually high-dimensional (~100,000 dimensions or more)
- The source and target domains may have distinct features
- Correspondences in the original space are difficult to find due to the high dimensionality

Classical Solutions to Alignment

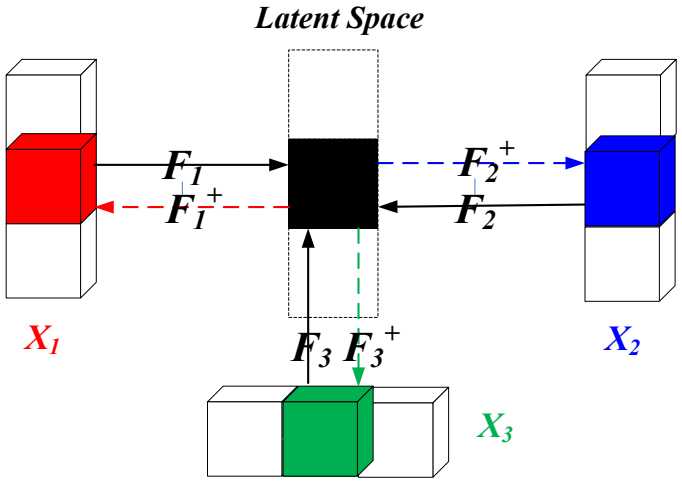
- Canonical Correlational Analysis (Hotelling)
  - Analogous to PCA across two data sets
  - Find a linear projection of each domain that maximizes correlation across domains
- Dynamic Time Warping (Juang)
  - Used in aligning time series data
  - Based on dynamic programming
- Our approach: find alignments by explicitly modeling the underlying data manifold

Manifold Alignment



Two-step alignment

Example: Procrustes alignment
(Wang and Mahadevan, ICML 2008)



One-step alignment

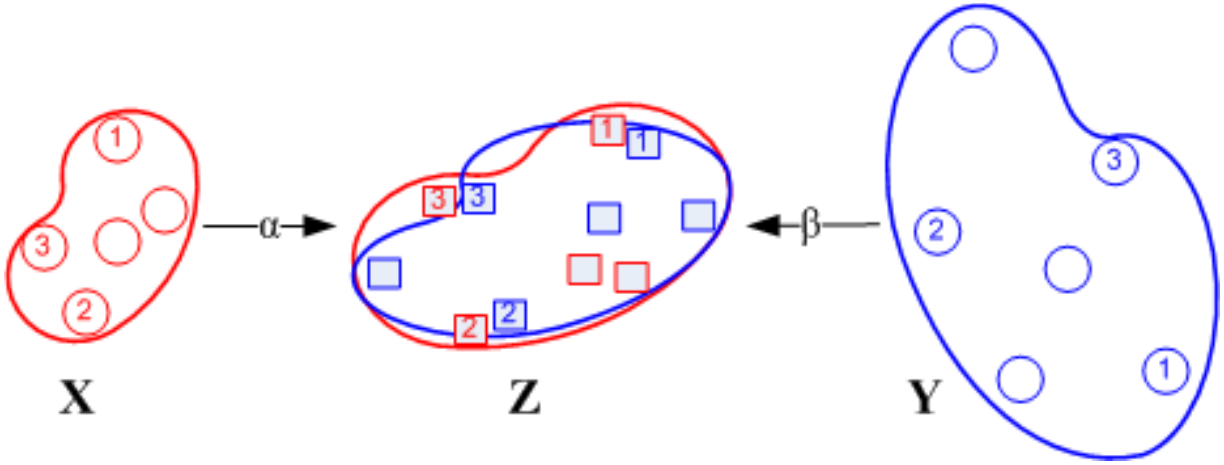
Example: Manifold Projections
(Wang and Mahadevan, IJCAI 2009)

A Spectrum of Manifold Alignment Approaches

	Given correspondences	***Given labels***	***Unsupervised alignment***
Preserve Local geometry			
Preserve Global geometry			
One-step alignment			
Two-step alignment			
Feature-level			
Instance-level			

Procrustes alignment ***Manifold Projections (MP)*** ***Extensions of MP***

Feature-level Manifold Projection



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

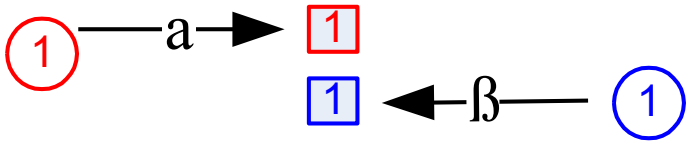
$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

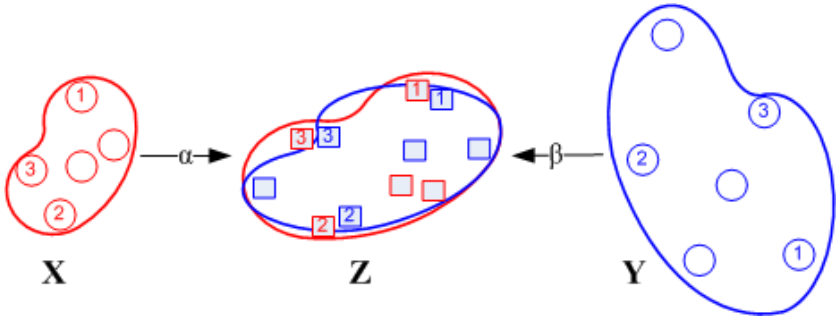
Projection Results:

$$x_i \rightarrow \alpha^T x_i$$

$$y_j \rightarrow \beta^T y_j$$



Manifold Projection and Alignment

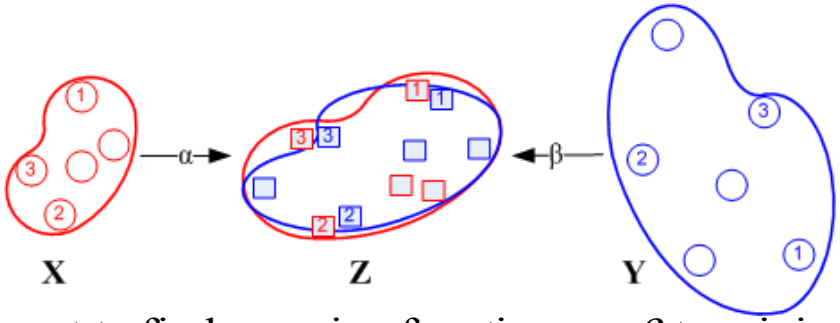


$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

Manifold Projection and Alignment



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

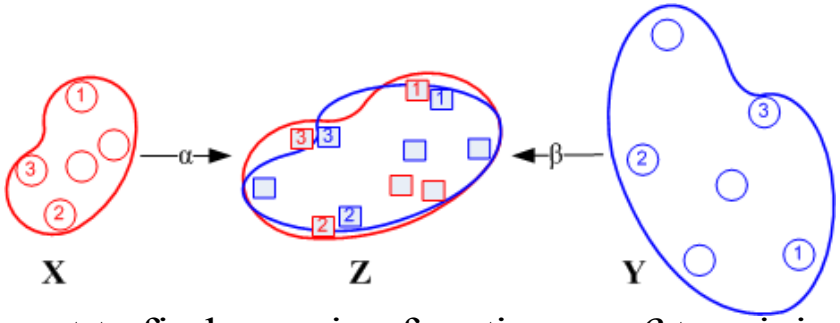
$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

We want to find mapping functions α, β to minimize the cost function $C(\alpha, \beta)$, where

$$C(\alpha,\beta) = \mu \sum_i \sum_j (\alpha^T x_i - \beta^T y_j)^2 W^{i,j} + 0.5 \sum_{i,j} (\alpha^T x_i - \alpha^T x_j)^2 W_x^{i,j} + 0.5 \sum_{i,j} (\beta^T y_i - \beta^T y_j)^2 W_y^{i,j}$$

Manifold Projection and Alignment



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

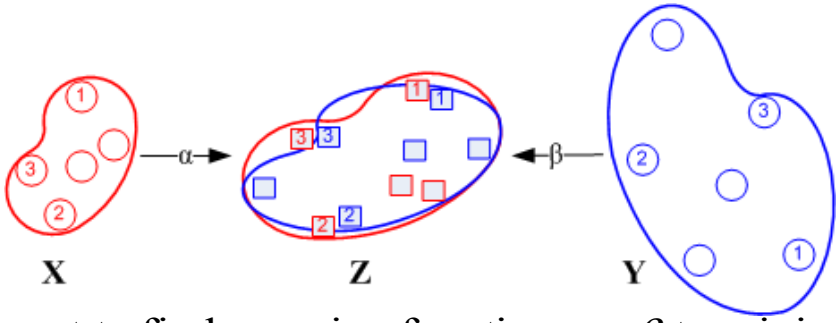
We want to find mapping functions α, β to minimize the cost function $C(\alpha, \beta)$, where

$$C(\alpha,\beta) = \mu \sum_i \sum_j (\alpha^T x_i - \beta^T y_j)^2 W^{i,j} + 0.5 \sum_{i,j} (\alpha^T x_i - \alpha^T x_j)^2 W_x^{i,j} + 0.5 \sum_{i,j} (\beta^T y_i - \beta^T y_j)^2 W_y^{i,j}$$

The ***first*** term encourages the corresponding instances from different domains to be projected to similar locations.

$W^{i,j}$=1, when x_i and y_j *are in correspondence; 0, otherwise.*

Manifold Projection and Alignment



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

We want to find mapping functions α, β to minimize the cost function $C(\alpha, \beta)$, where

$$C(\alpha,\beta) = \mu \sum_i \sum_j (\alpha^T x_i - \beta^T y_j)^2 W^{i,j} + 0.5 \sum_{i,j} (\alpha^T x_i - \alpha^T x_j)^2 W_x^{i,j} + 0.5 \sum_{i,j} (\beta^T y_i - \beta^T y_j)^2 W_y^{i,j}$$

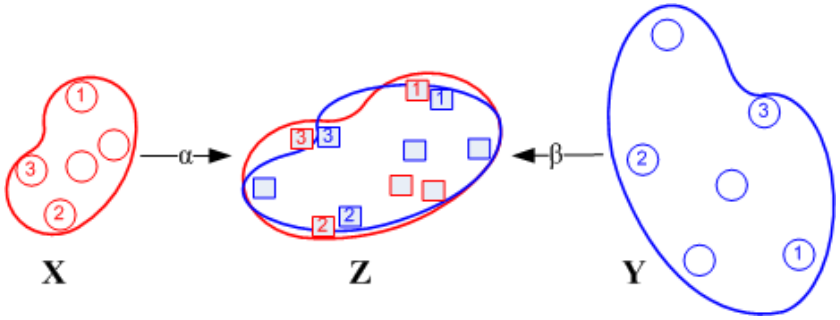
The ***first*** term encourages the corresponding instances from different domains to be projected to similar locations.

$W^{i,j}$=1, when x_i and y_j *are in correspondence; 0, otherwise.*

$$W = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & 0 & \\ & & & & & & \ddots \\ & & & & & & & 0 \end{bmatrix}$$

- ***When 1:1 correspondence is given*** *($x_i \leftrightarrow y_i$ for $i <= l$):*
- ***When many:many correspondence is given****, set corresponding entries to 1.*
- ***When nothing is given****, we can use local geometry information to fill in this matrix. (IJCAI 2009)*

Manifold Projection and Alignment



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$
$$Y = [y_1, \ldots, y_n], y_j \in R^q$$
$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

We want to find mapping functions α, β to minimize the cost function $C(\alpha, \beta)$, where

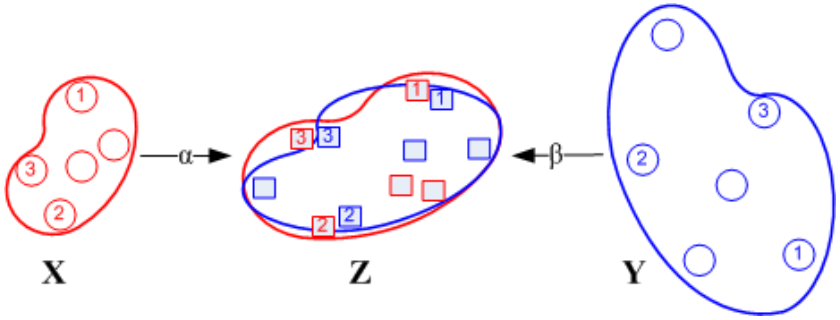
$$C(\alpha,\beta) = \mu \sum_i \sum_j (\alpha^T x_i - \beta^T y_j)^2 W^{i,j} + 0.5 \sum_{i,j} (\alpha^T x_i - \alpha^T x_j)^2 W_x^{i,j} + 0.5 \sum_{i,j} (\beta^T y_i - \beta^T y_j)^2 W_y^{i,j}$$

The ***second*** and ***third*** terms encourage the neighborhood relationship within X and Y to be preserved in the mappings.

$W_x^{i,j}$: similarity of x_i and x_j in the original space.

$W_y^{i,j}$: similarity of y_i and y_j in the original space.

Comparison with Canonical Correlation Analysis (CCA)



$$X = [x_1, \ldots, x_m], x_i \in R^p.$$

$$Y = [y_1, \ldots, y_n], y_j \in R^q$$

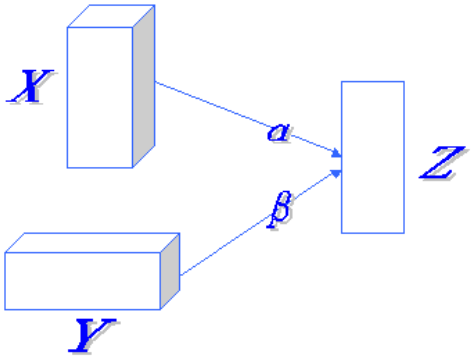
$$x_i \leftrightarrow y_i \text{ for } i \in [1, l]$$

We want to find mapping functions α, β to minimize the cost function $C(\alpha, \beta)$, where

$$C(\alpha,\beta) = \mu \sum_i \sum_j (\alpha^T x_i - \beta^T y_j)^2 W^{i,j} + 0.5 \sum_{i,j} (\alpha^T x_i - \alpha^T x_j)^2 W_x^{i,j} + 0.5 \sum_{i,j} (\beta^T y_i - \beta^T y_j)^2 W_y^{i,j}$$

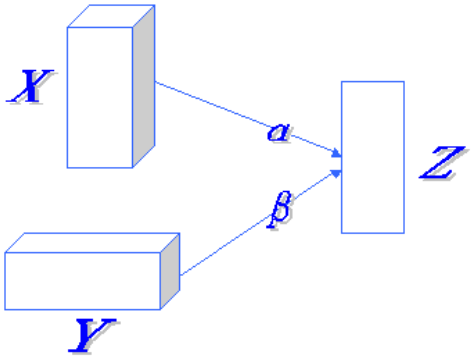
How to compute projections?

Optimal Solution:



How to compute projections?

Optimal Solution:



(1) Construct Z, L, D using X, Y and W (the correspondences).

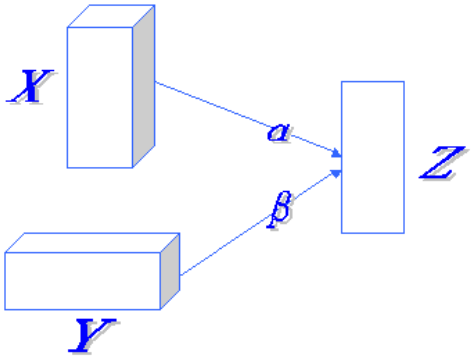
D_x is a diagonal matrix: $D_x^{ii} = \sum_j W_x^{ij}$.
$L_x = D_x - W_x$.
D_y is a diagonal matrix: $D_y^{ii} = \sum_j W_y^{ij}$.
$L_y = D_y - W_y$.
Ω_1 is an $m \times m$ diagonal matrix, and $\Omega_1^{ii} = \sum_j W^{i,j}$.
Ω_2 is an $m \times n$ matrix, and $\Omega_2^{i,j} = W^{i,j}$.
Ω_3 is an $n \times m$ matrix, and $\Omega_3^{i,j} = W^{j,i}$.
Ω_4 is an $n \times n$ diagonal matrix, and $\Omega_4^{ii} = \sum_j W^{j,i}$.

$$Z = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}.$$
$$D = \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}.$$
$$L = \begin{pmatrix} L_x + \mu\Omega_1 & -\mu\Omega_2 \\ -\mu\Omega_3 & L_y + \mu\Omega_4 \end{pmatrix}.$$

Create a joint domain.
(use correspondences if available)

How to compute projections

Optimal Solution:



(1) Construct Z, L, D using X, Y and W (the correspondences).

D_x is a diagonal matrix: $D_x^{ii} = \sum_j W_x^{ij}$.
$L_x = D_x - W_x$.
D_y is a diagonal matrix: $D_y^{ii} = \sum_j W_y^{ij}$.
$L_y = D_y - W_y$.
Ω_1 is an $m \times m$ diagonal matrix, and $\Omega_1^{ii} = \sum_j W^{i,j}$.
Ω_2 is an $m \times n$ matrix, and $\Omega_2^{i,j} = W^{i,j}$.
Ω_3 is an $n \times m$ matrix, and $\Omega_3^{i,j} = W^{j,i}$.
Ω_4 is an $n \times n$ diagonal matrix, and $\Omega_4^{ii} = \sum_j W^{j,i}$.

$$Z = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}.$$
$$D = \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}.$$
$$L = \begin{pmatrix} L_x + \mu\Omega_1 & -\mu\Omega_2 \\ -\mu\Omega_3 & L_y + \mu\Omega_4 \end{pmatrix}.$$

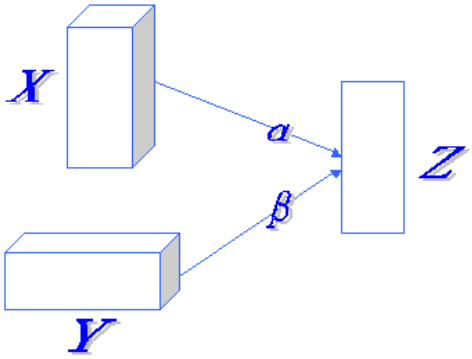
Create a joint domain.
(use correspondences if available)

Project the joint domain to a lower dimensional space.

(2) Theorem 1 : α, β to minimize $C(\alpha, \beta)$ are given by the eigenvectors corresponding to the smallest eigenvalues of $ZLZ^T\gamma = \lambda ZDZ^T\gamma$.

How to compute projections?

Optimal Solution:



(1) Construct Z, L, D using X, Y and W (the correspondences).

D_x is a diagonal matrix: $D_x^{ii} = \sum_j W_x^{ij}$.
$L_x = D_x - W_x$.
D_y is a diagonal matrix: $D_y^{ii} = \sum_j W_y^{ij}$.
$L_y = D_y - W_y$.
Ω_1 is an $m \times m$ diagonal matrix, and $\Omega_1^{ii} = \sum_j W^{i,j}$.
Ω_2 is an $m \times n$ matrix, and $\Omega_2^{i,j} = W^{i,j}$.
Ω_3 is an $n \times m$ matrix, and $\Omega_3^{i,j} = W^{j,i}$.
Ω_4 is an $n \times n$ diagonal matrix, and $\Omega_4^{ii} = \sum_j W^{j,i}$.

$$Z = \begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}.$$
$$D = \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}.$$
$$L = \begin{pmatrix} L_x + \mu\Omega_1 & -\mu\Omega_2 \\ -\mu\Omega_3 & L_y + \mu\Omega_4 \end{pmatrix}.$$

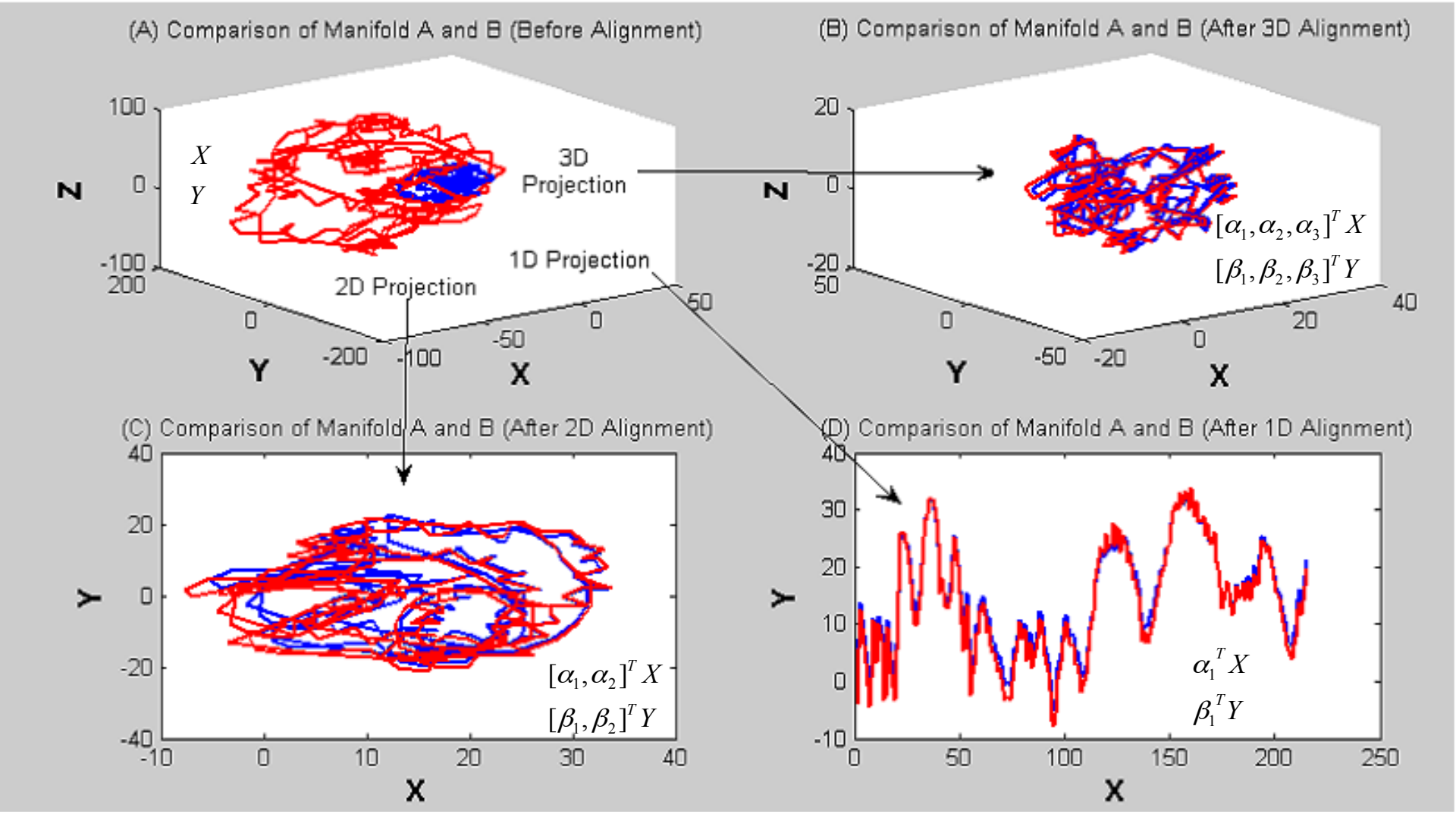
Create a joint domain.
(use correspondences if available)

Project the joint domain to a lower dimensional space.

(2) Theorem 1 : α, β to minimize $C(\alpha, \beta)$ are given by the eigenvectors corresponding to the smallest eigenvalues of $ZLZ^T\gamma = \lambda ZDZ^T\gamma$.

(3) $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = [\gamma_1, \ldots, \gamma_d]$, where γ_i is the i^{th} minimum eigenvector.

Protein Alignment



General Framework for *Manifold Projection*

- Instance-level

$$C(\mathcal{Y}_1,\cdots,\mathcal{Y}_c)=0.5\mu_1\sum_{a=1}^{c}\sum_{b=1}^{c}\sum_{i=1}^{m_a}\sum_{j=1}^{m_b}\|\mathcal{Y}_a^i-\mathcal{Y}_b^j\|^2W_{a,b}^{i,j}+0.5\mu_2\sum_{k=1}^{c}\sum_{i=1}^{m_k}\sum_{j=1}^{m_k}\|\mathcal{Y}_k^i-\mathcal{Y}_k^j\|^2W_k^{i,j}$$

1. *c=1*: Laplacian eigenmaps (Belkin, Niyogi, 2003)
2. *c=2, μ_2=1, $W_{a,b}$* is an identity matrix: Semi-supervised alignment (Ham, et. al. 2005)
3. *c=2, μ_2=0, $W_{a,b}$* is an identity matrix: Non-linear Canonical Correlation Analysis
4. ……

- Feature-level

$$C(\mathcal{F}_1,\cdots,\mathcal{F}_c)=0.5\mu_1\sum_{a=1}^{c}\sum_{b=1}^{c}\sum_{i=1}^{m_a}\sum_{j=1}^{m_b}\|\mathcal{F}_a^Tx_a^i-\mathcal{F}_b^Tx_b^j\|^2W_{a,b}^{i,j}+0.5\mu_2\sum_{k=1}^{c}\sum_{i=1}^{m_k}\sum_{j=1}^{m_k}\|\mathcal{F}_k^Tx_k^i-\mathcal{F}_k^Tx_k^j\|^2W_k^{i,j}$$

2. *c=2, μ_2=0, $W_{a,b}$* is an identity matrix: Canonical Correlation Analysis (Hotelling, 1936)
3. *Unsupervised manifold alignment* (Wang, Mahadevan, IJCAI 09)
4. *c=2, μ_2=1, $W_{a,b}$* is a symmetric matrix
5. *c>2*: Multiple manifold alignment
6. ……

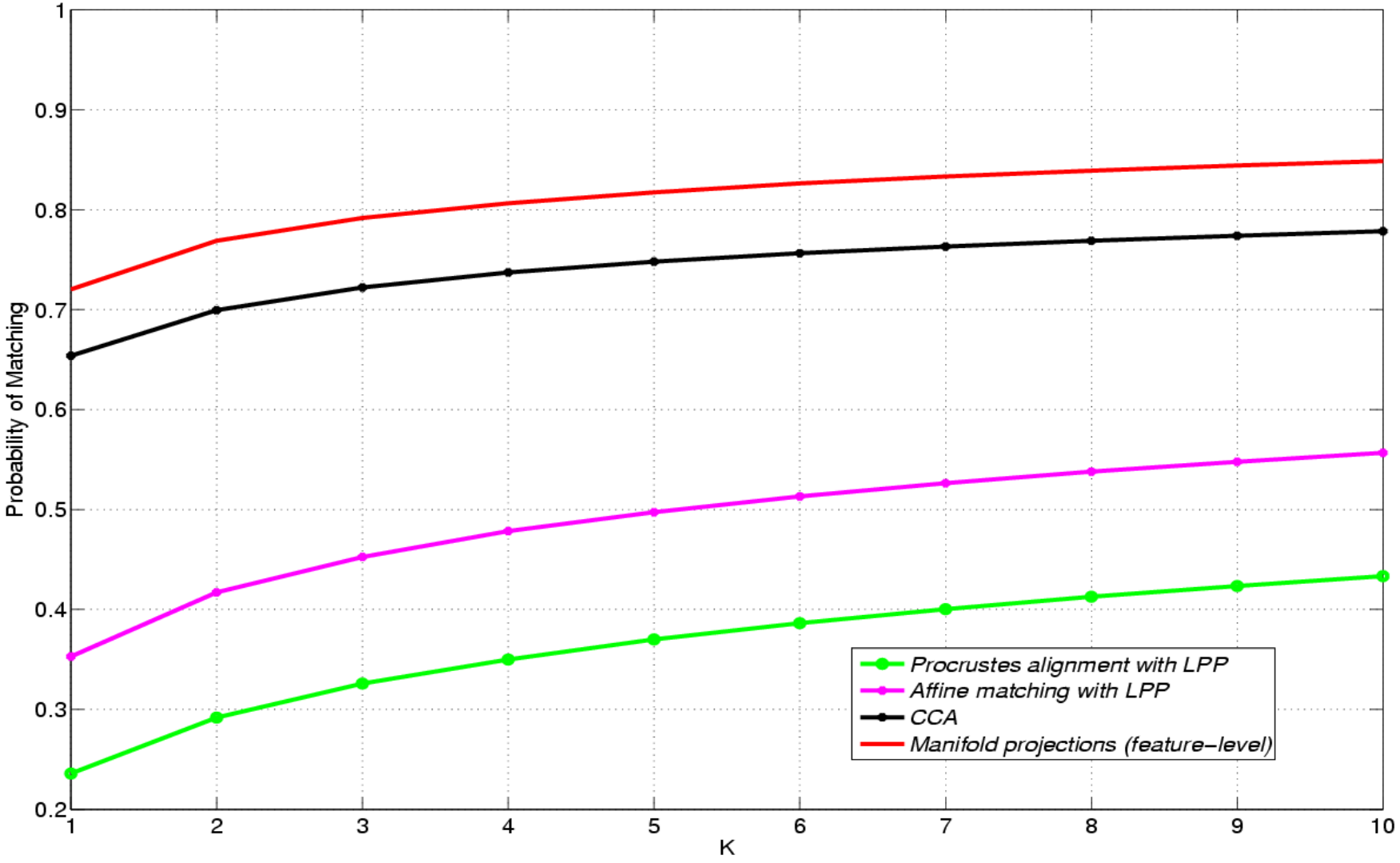
[Wang and Mahadevan, AAAI Fall Symposium 2009; IJCAI 2009]

European Parliament Parallel Corpus

- **Data:**
  - **70,458** English-Italian-German document triples.
  - Those documents have > **170,000,000** words.

- **Features:**
  - **X_1:** English documents are represented by the most commonly used **2,500 English words**.
  - **X_2:** Italian documents are represented by the most commonly used **2,500 Italian words**.
  - **X_3:** German documents are represented by the most commonly used **2,500 German words**.

EU Parallel Corpus: Results



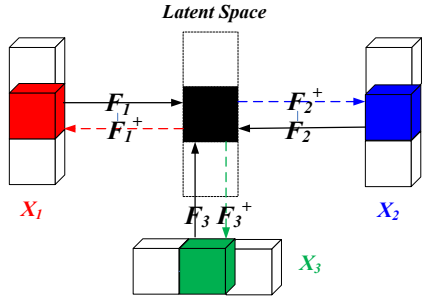
EU Parallel Corpus

- Interpretation of mapping functions:

  F1=[α_1 α_2 …α_{200}]: 2500*200 matrix

  F2=[β_1 β_2 …β_{200}]: 2500*200 matrix

  F3=[γ_1 γ_2 …γ_{200}]: 2500*200 matrix



	Top Terms
α_1	fisheries fishing agency fishermen negotiated applause protocol nos sustainability ports
α_2	consumers internet consumer strategies b5 bulgaria behaviour b4 discharge november
α_3	strategies swedish courage denmark telecommunications nato credibility wine regional brings
α_4	interinstitutional parliaments repeated guarantees century rail finland british choose conciliation
α_5	unemployment thursday heads portuguese economies declarations balkans widespread islands india

	Top Terms
β_1	pesca agenzia a5 applausi protocollo ripartizione pescatori bilaterali sostenibilita tonnellate
β_2	consumatori reca internet consumatore strategie discarico bulgaria novembre allargamento chiusa
β_3	strategie svedese interrogazioni occidentali danimarca regionale kyoto coraggio credibilita segretario
β_4	parlamenti sentenza interistituzionale aprile ferroviario britannica tecnici essenzialmente unanimita indipendenza
β_5	disoccupazione giovedi scientifica portoghese balcani aeree firmato turco maggio piccoli

	Top Terms
γ_1	fischerei fischereipolitik agentur protokolls fischer protokoll ablehnen a5 tatsächlichen arten
γ_2	verbrauchern verbraucher strategien internet bulgarien entlastung anfragen todesstrafe b4 zukünftigen
γ_3	schwedischen strategien regionalpolitik frist anfragen westlichen mut nato regionalen minuten
γ_4	parlamente interinstitutionelle b4 parlamenten urteil spezielle folgt anmerkungen nächster beschluß
γ_5	portugiesischen personal arbeitslosigkeit offenen wissenschaftlichen polizei verordnungen donnerstag inseln äußerungen

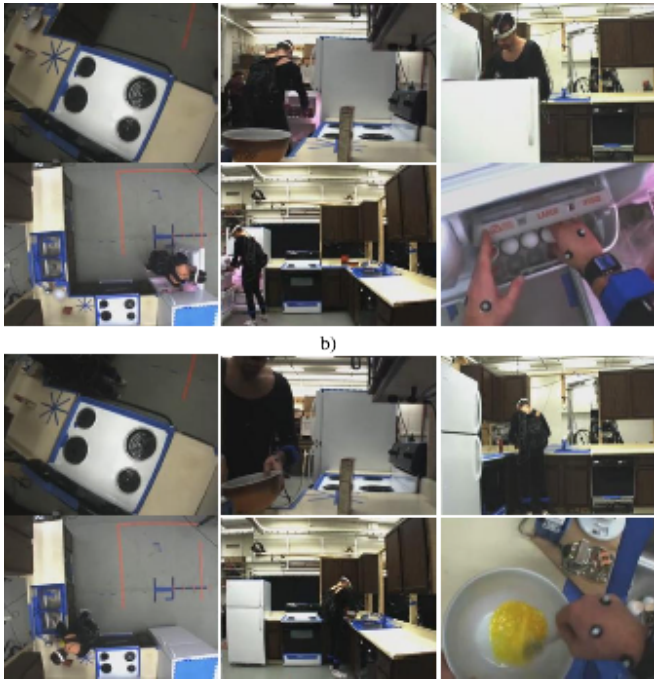
The Indus Script (2500 BC)

- **Data:**
  - **1548** "documents" – different seals
  - **7000** "words" – symbols on the seals (**377** unique)
  - Many competing inconsistent claims of decipherement
  - Is this a language (Rao et al., Science, PNAS)

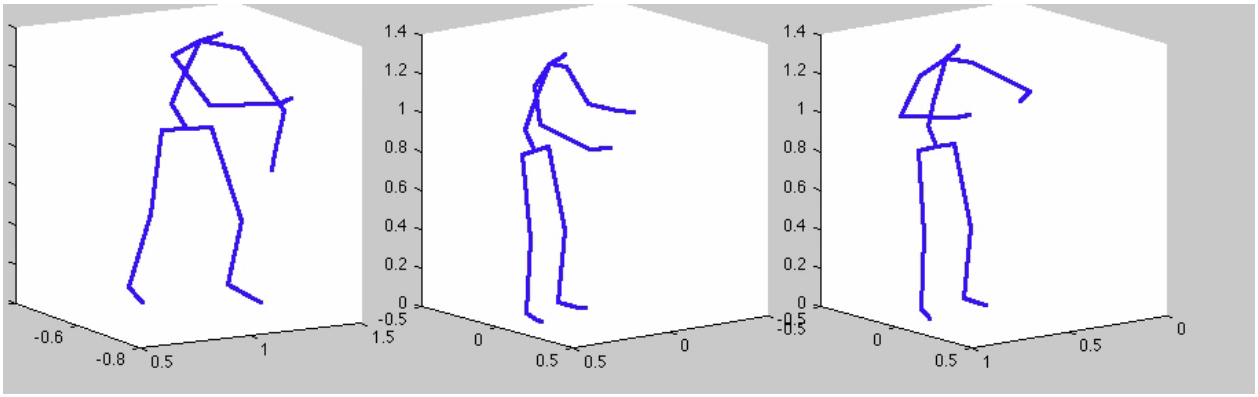
- **Challenges:**
  - We have no parallel corpus, no Rosetta Stone
  - The document-symbol matrix is sparse
  - Many of the documents are very short

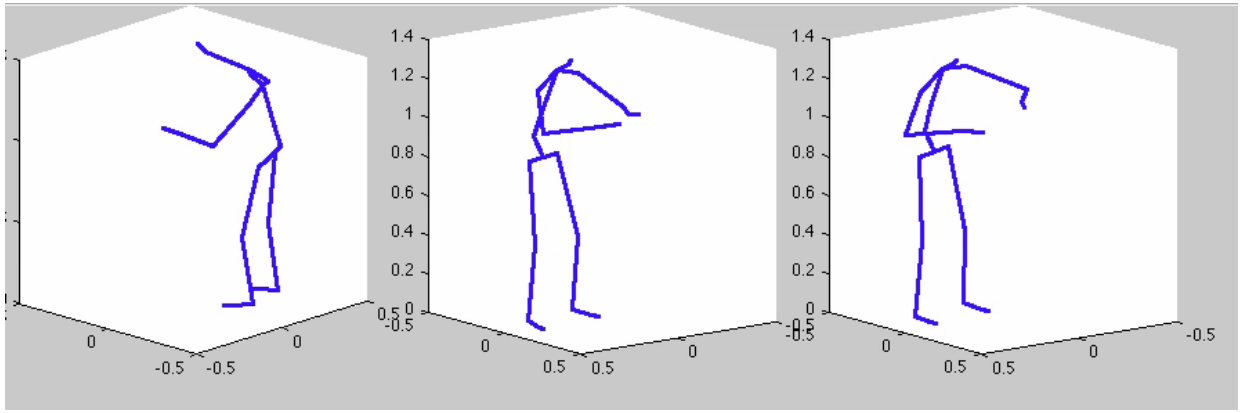
Manifold Alignment over time

- CMU Multimodal activity dataset
  - kitchen.cs.cmu.edu
- Measure human activity while cooking
  - 26 subjects
  - 5 different recipes
- Many sensors:
  - Cameras
  - RFID
  - Audio
  - Joint angles



Temporal Alignment of Activities





Scaling to Really Large Problems

- Data mining a million books!
  - NSF Large IIS Project: James Allan, PI, U.Mass
- Transfer learning in Reinforcement Learning
  - Backgammon (10^{20} states!), Go (10^{170} states!!)
- CMU Multi-modal activity dataset
  - Many sensor modalities, high frame rate