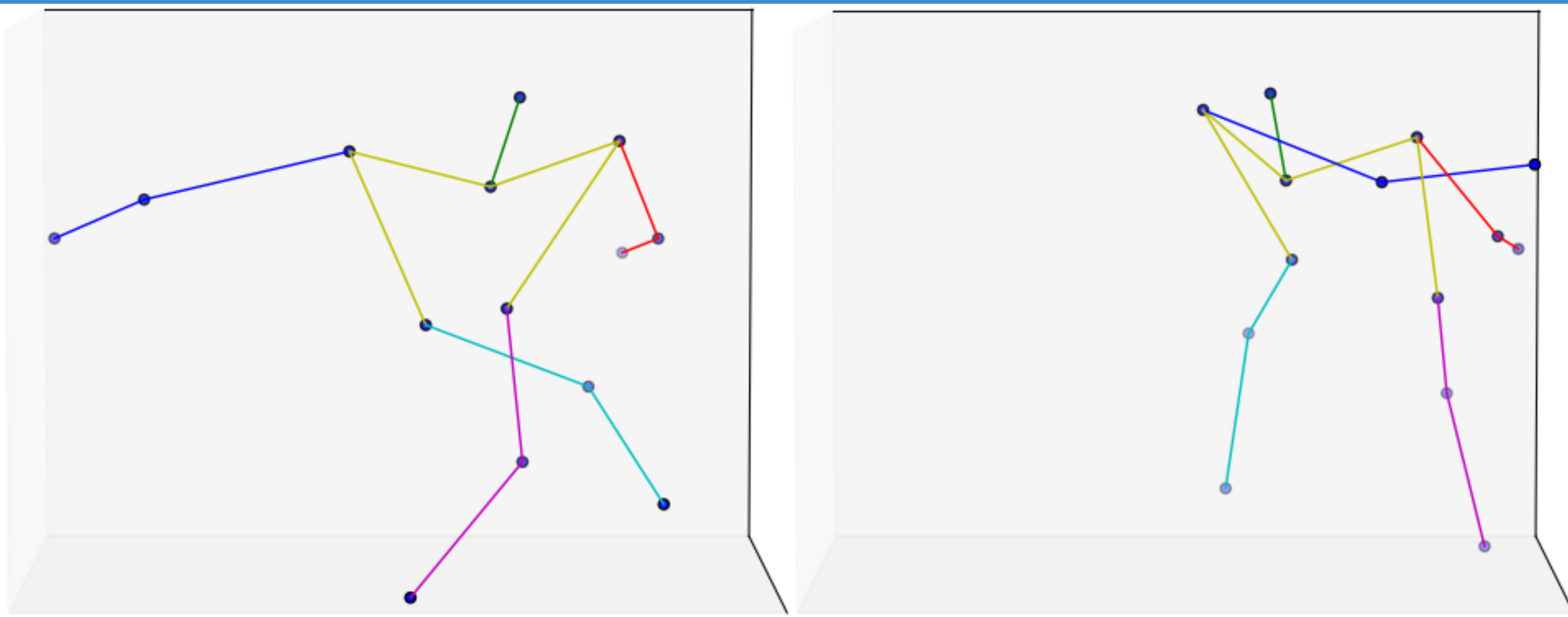
A Rotation Invariant Latent Factor Model for Moveme Discovery from Static Poses

Matteo Ruggero Ronchi, Joon Sik Kim, Yisong Yue

Wednesday, December 14th, ICDM 2016

ICDM 2016
IEEE International Conference on Data Mining
BARCELONA

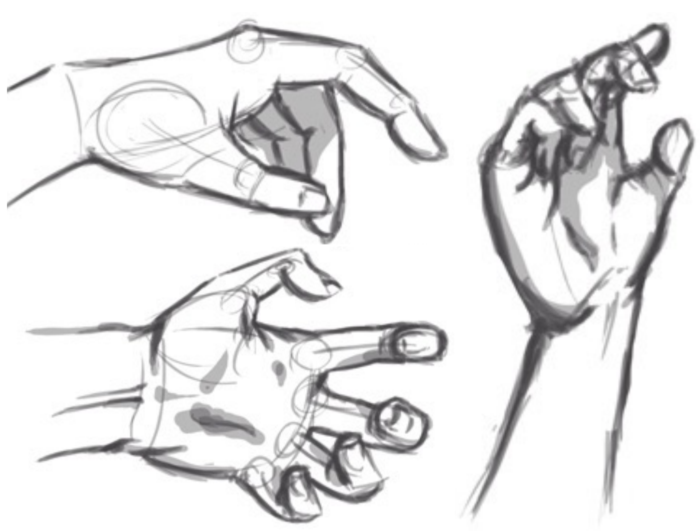


Human Movements

A priori unrestricted except for physically imposed joint angle limits.

Observable at different levels of detail:

- **Moveme:** short target-oriented trajectory ('reach','grasp','lift').
- **Action:** predefined ordered sequence of movemes ('drink from a cup','open a door').
- **Activity:** stochastic combination of actions over time ('dine','read','drive').



State of Human Movement Analysis

We know a lot from images*:

soccer | competition | football | athlete
match | man | stadium | adult | ball
uniform | outfit | soccer player
sports equipment | action | soccer field
web | game

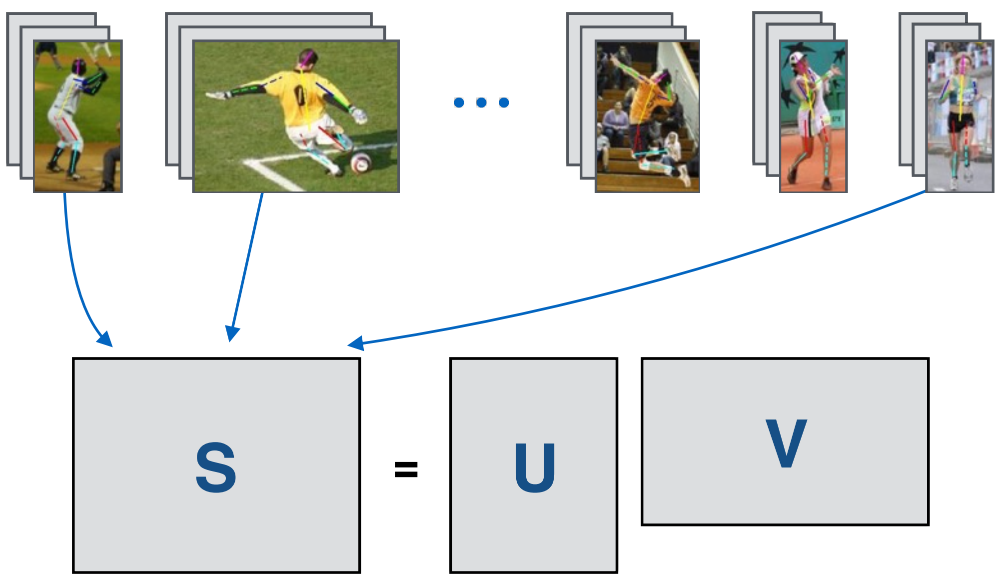
*clarifai.com

However, human behaviour is difficult to analyze at a *fine scale of dynamics*:

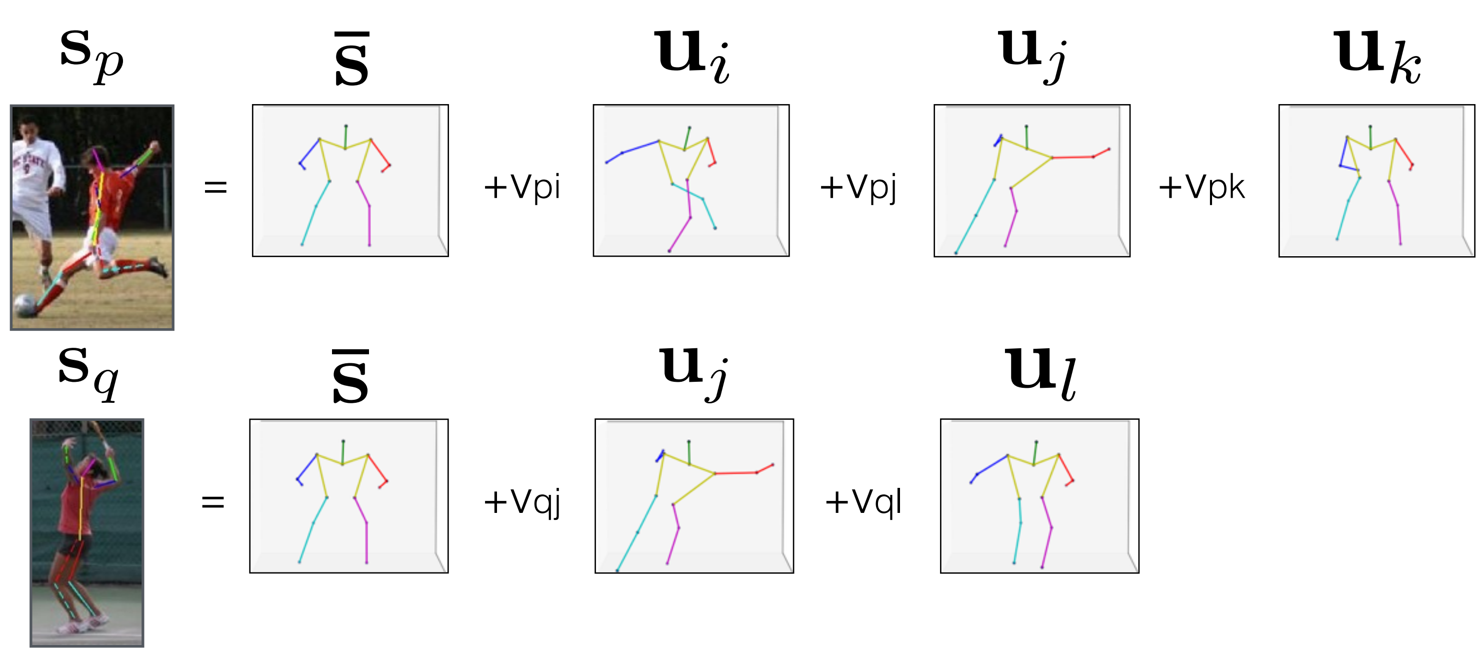
- What are the typical ranges of motion for the human body?
- What type of leg movements correlate with specific shoulder positions?
- How can we expect the arms to move given the current body pose?

Automatic Moveme Discovery

Our primary goal is to learn a global basis to smoothly capture movemes from a collection of 2-d joints.

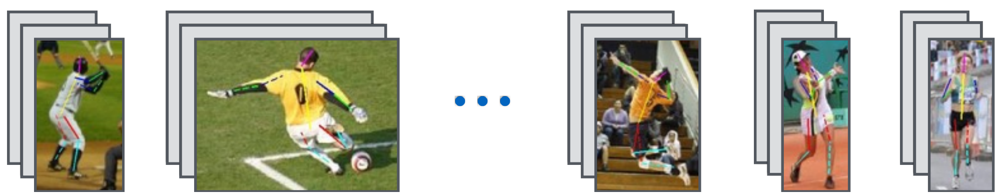


- Rotation Invariant
- Sparse
- Good Generalization
- Complementary Representation



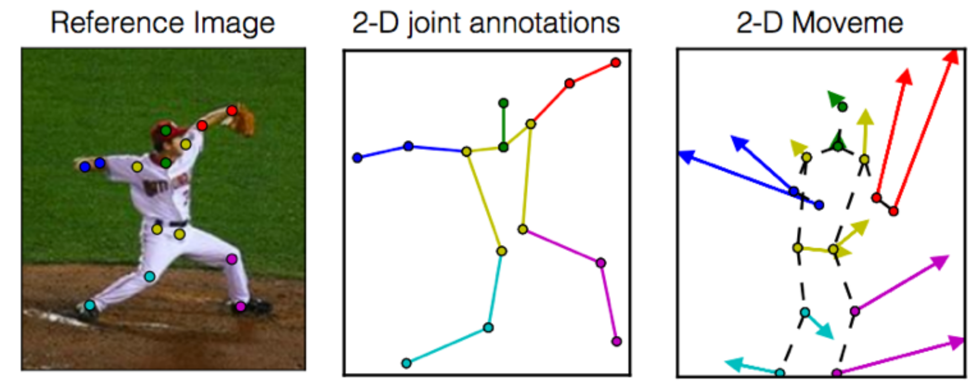
Framework and Model

LSP data: 2000 images 8 activities



$\mathbf{S} = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{j=1}^{n} \qquad \mathbf{S} \in \Re^{2d \times n}$

Movemes: displacement from the mean pose



Rotation Invariant 3D Latent Factor Model:

- **Factorization into a pose basis matrix U and a coefficient matrix V**

S = **U** **V**

$\mathbf{U} \in \Re^{3d \times k}$

$\mathbf{V} \in \Re^{k \times n}$

- global and 3d

$$arg\,min_{\mathbf{U},\mathbf{V},\theta} \mathcal{E}(\mathbf{U},\mathbf{V},\theta) + \lambda_U \|\mathbf{U}\|_F^2 + \lambda_V \|\mathbf{V}\|_1$$

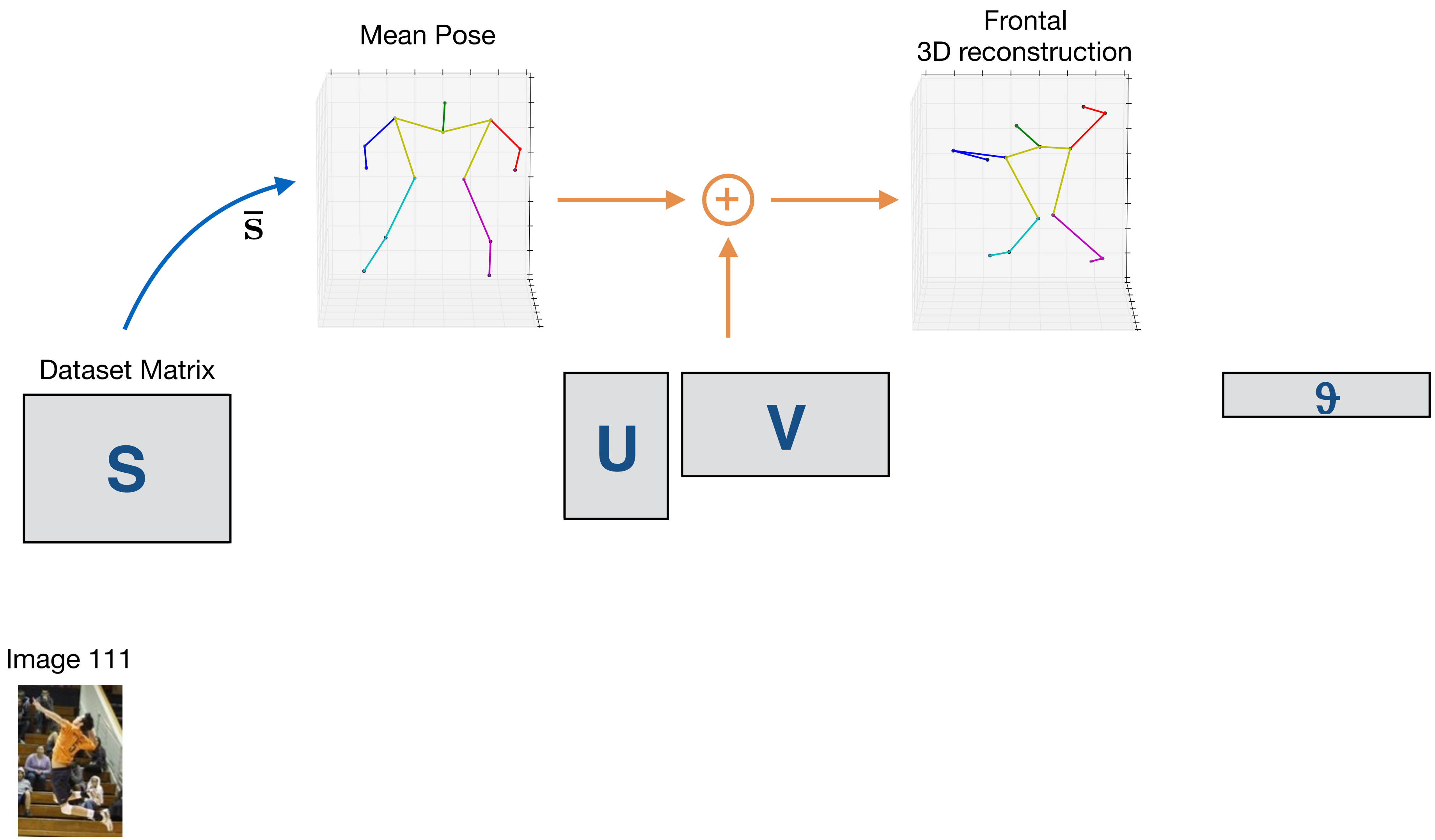
- sparse

$$\mathcal{E}(\mathbf{U},\mathbf{V},\theta) = \sum_j (\mathbf{s}_j - \underline{f(\bar{\mathbf{s}} + \mathbf{U} \cdot \mathbf{v}_j, \theta_j)})^2$$

- rotation invariant

- ϑj: angle of view estimate
- f: rotation + projection

3D Latent Factor Model Optimization



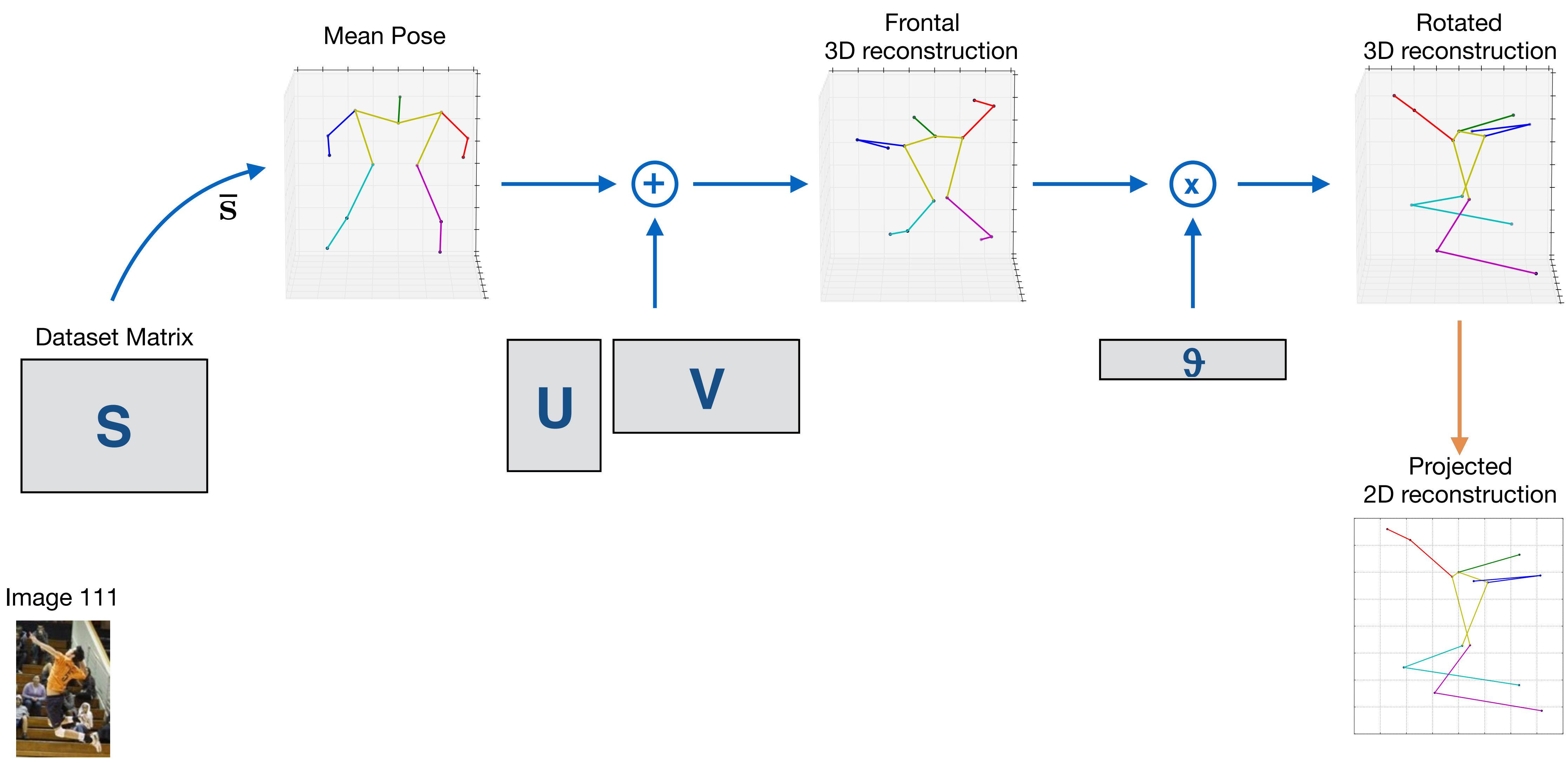
Initialization:

- **U:** random between -1,1
- **V:** random between 0,1
- **ϑ:** heuristic

Frontal Reconstruction:

$$\widetilde{\mathbf{s}}_{111} = \bar{\mathbf{s}} + \mathbf{U} \cdot \mathbf{v}_{111}$$

3D Latent Factor Model Optimization



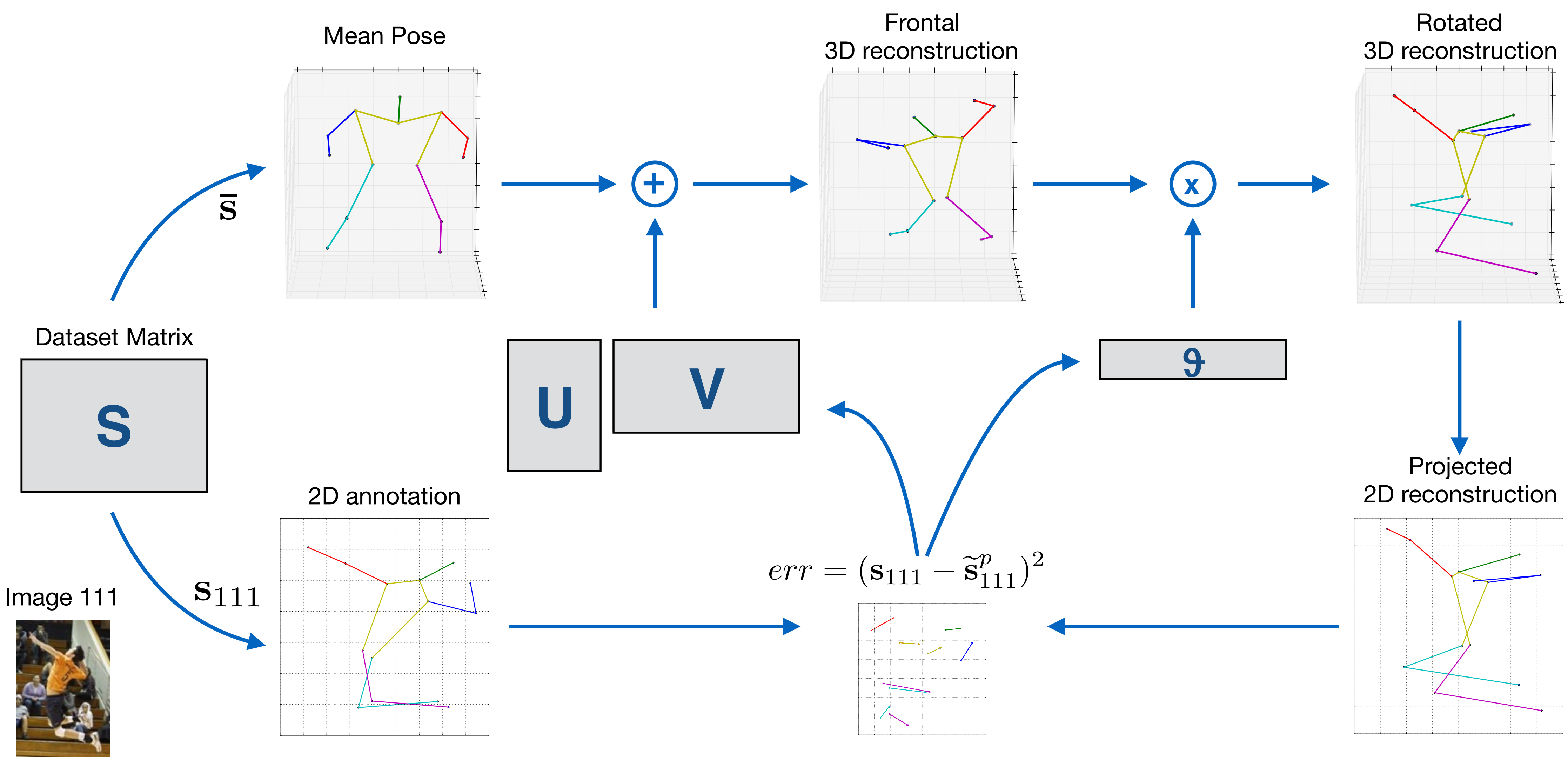
Rotation:

$$\widetilde{\mathbf{s}}^{\mathbf{r}}_{111} = \mathbf{Q}(\theta_{111}) \cdot \widetilde{\mathbf{s}}_{111}$$

Projection:

$$\widetilde{\mathbf{s}}^{p}_{111} = \left[\widetilde{\mathbf{s}}^{\mathbf{r},(\mathbf{x})}_{111}, \widetilde{\mathbf{s}}^{\mathbf{r},(\mathbf{y})}_{111}\right]$$

3D Latent Factor Model Optimization



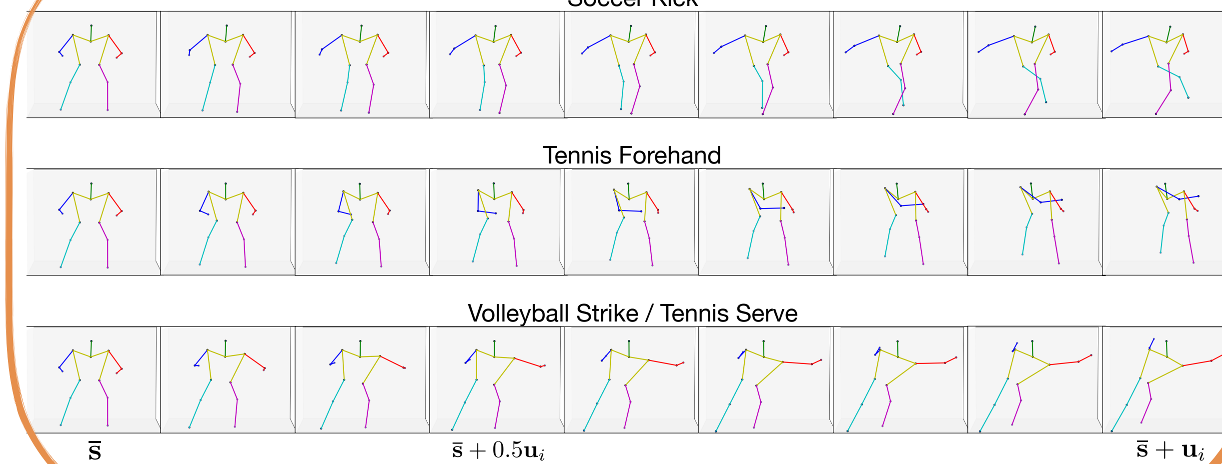
Alternate Gradient Update:

- **Representation:** $\dfrac{\delta err}{\delta \mathbf{U}}$ $\dfrac{\delta err}{\delta \mathbf{V}}$
- **Angle of view:** $\dfrac{\delta err}{\delta \theta}$

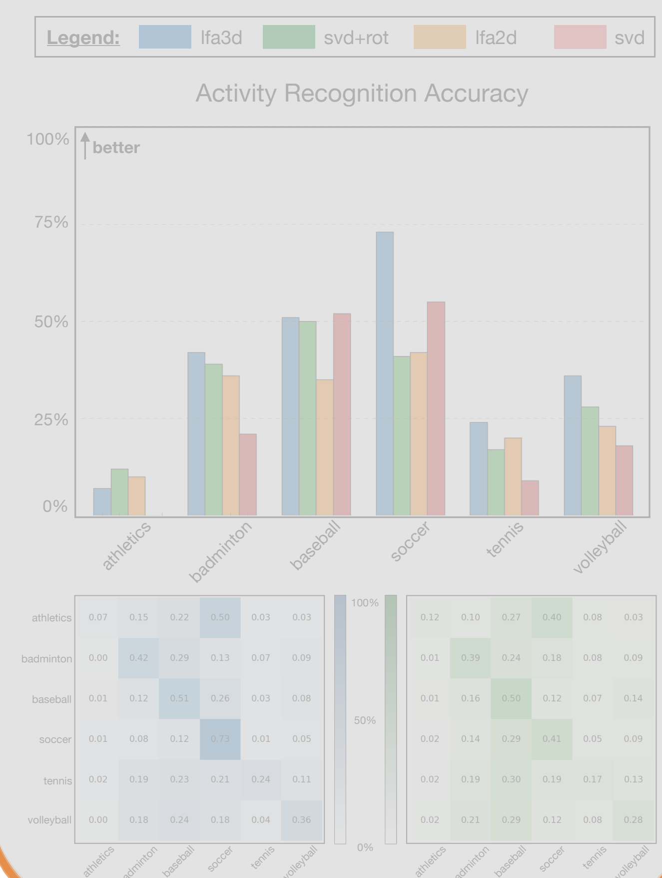
Experiments

OPTIMIZATION DETAILS

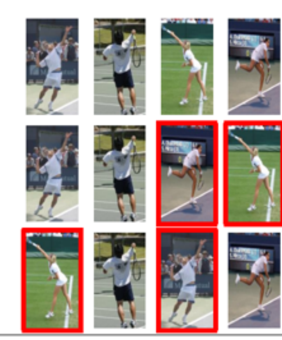
MOVEME VISUALIZATION



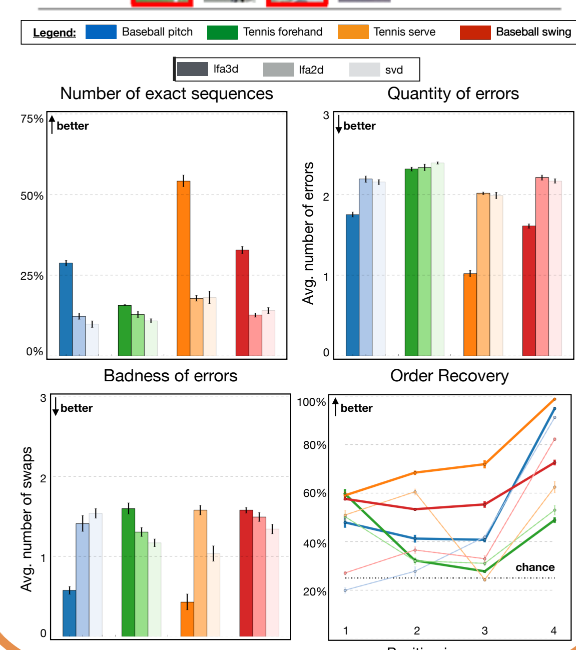
ACTIVITY RECOGNITION



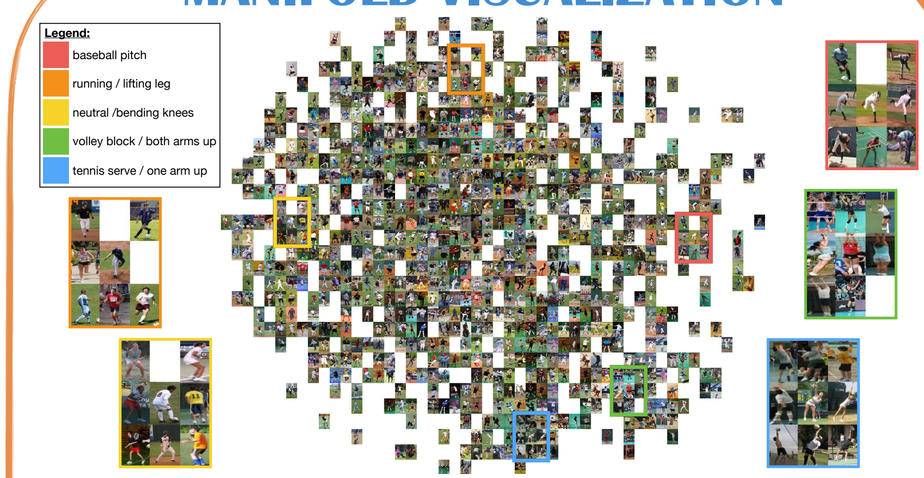
DYNAMICS INFERENCE

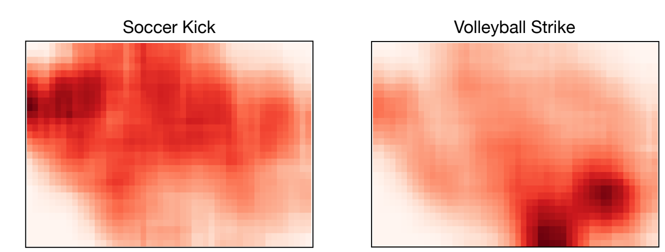


Method	Reordered Sequence	Errors	Swaps
lfa3d		0	0
lfa2d		2	1
svd		2	3



MANIFOLD VISUALIZATION



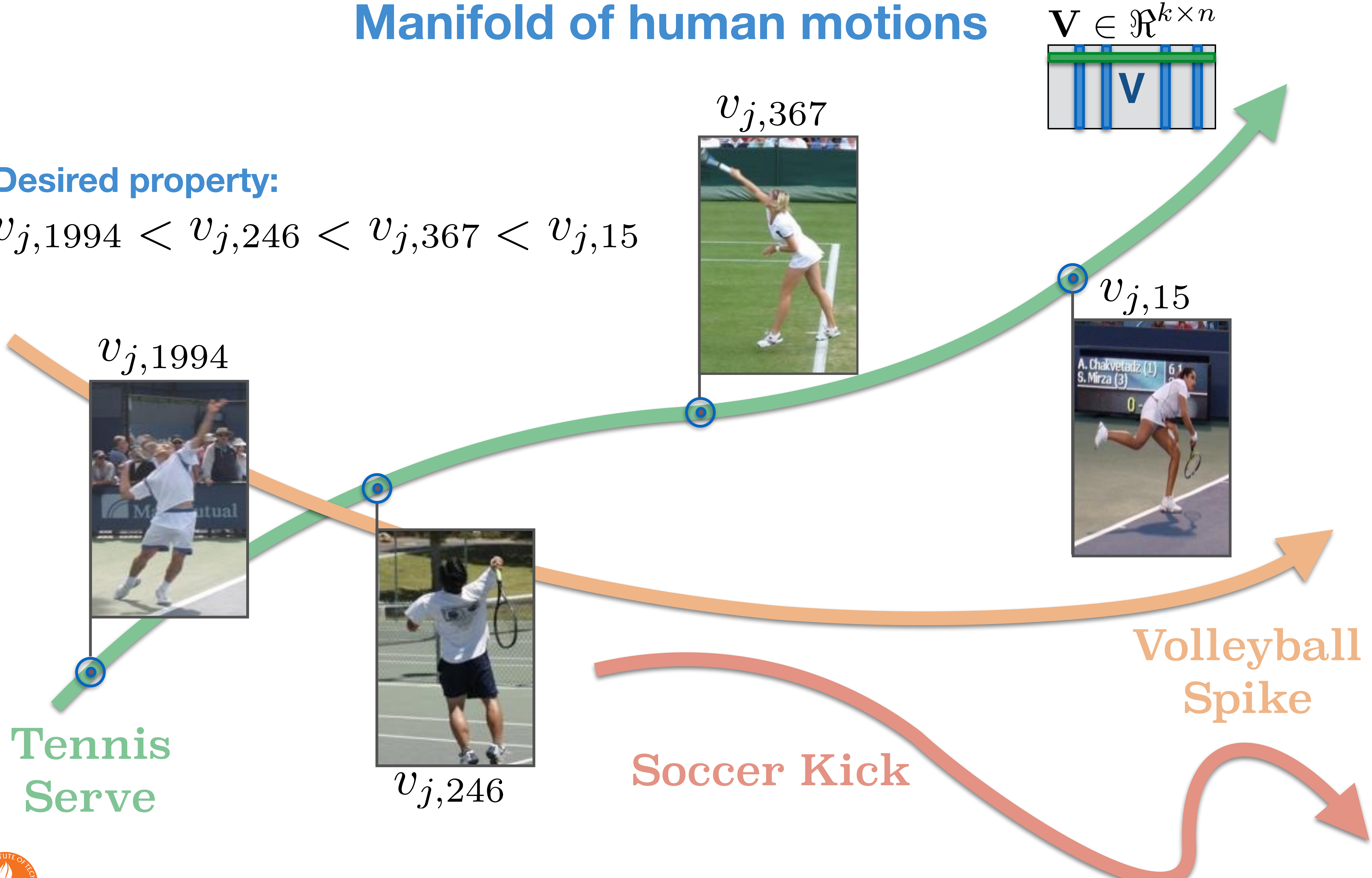


Action Dynamics Inference

Manifold of human motions

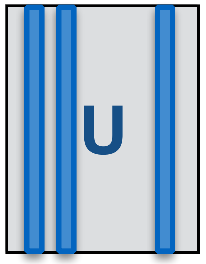
Desired property:

$$v_{j,1994} < v_{j,246} < v_{j,367} < v_{j,15}$$



Moveme Visualization

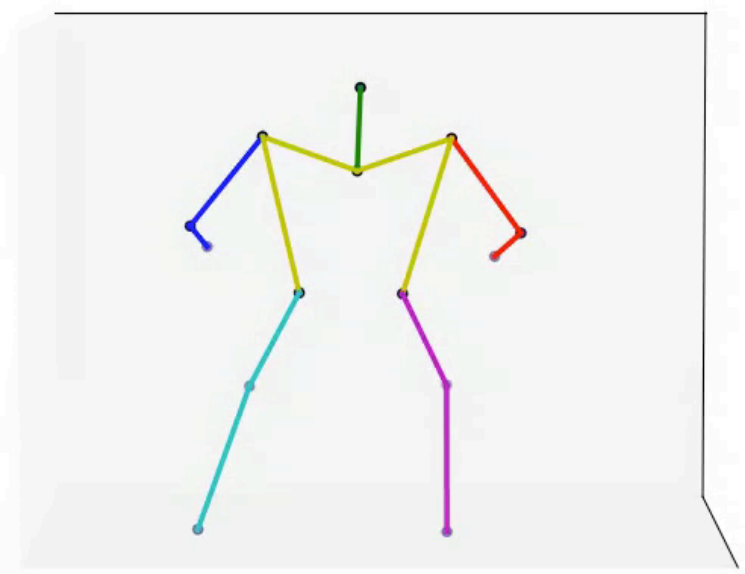
$\mathbf{U} \in \Re^{3d \times k}$



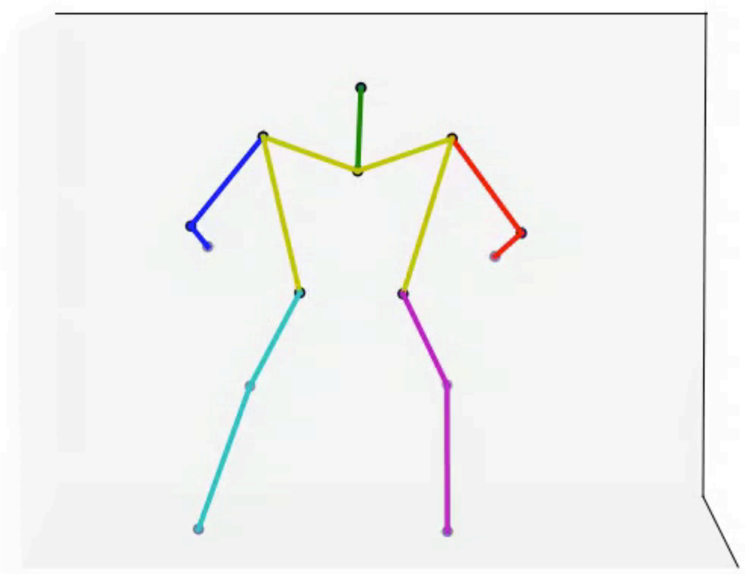
We can synthesize realistic human motions from static joint locations in images with the matrix U:

- **By adding an increasing percentage of a basis to the mean pose.**

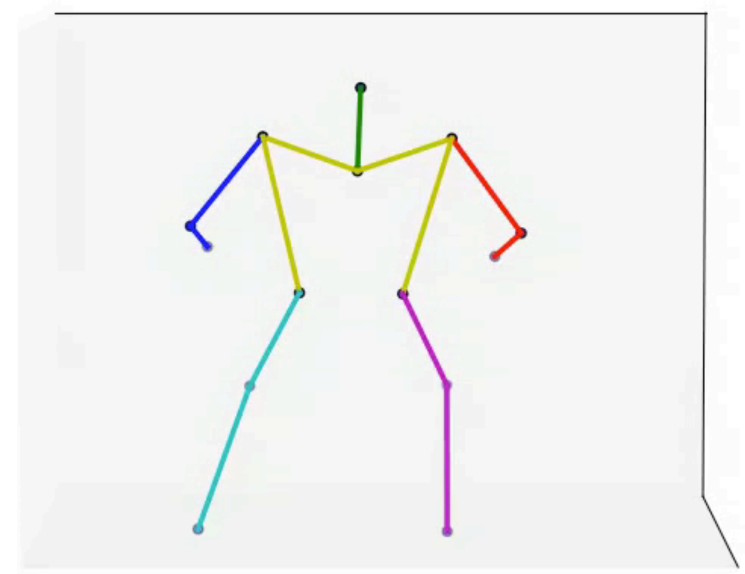
Soccer Kick



Volleyball Spike



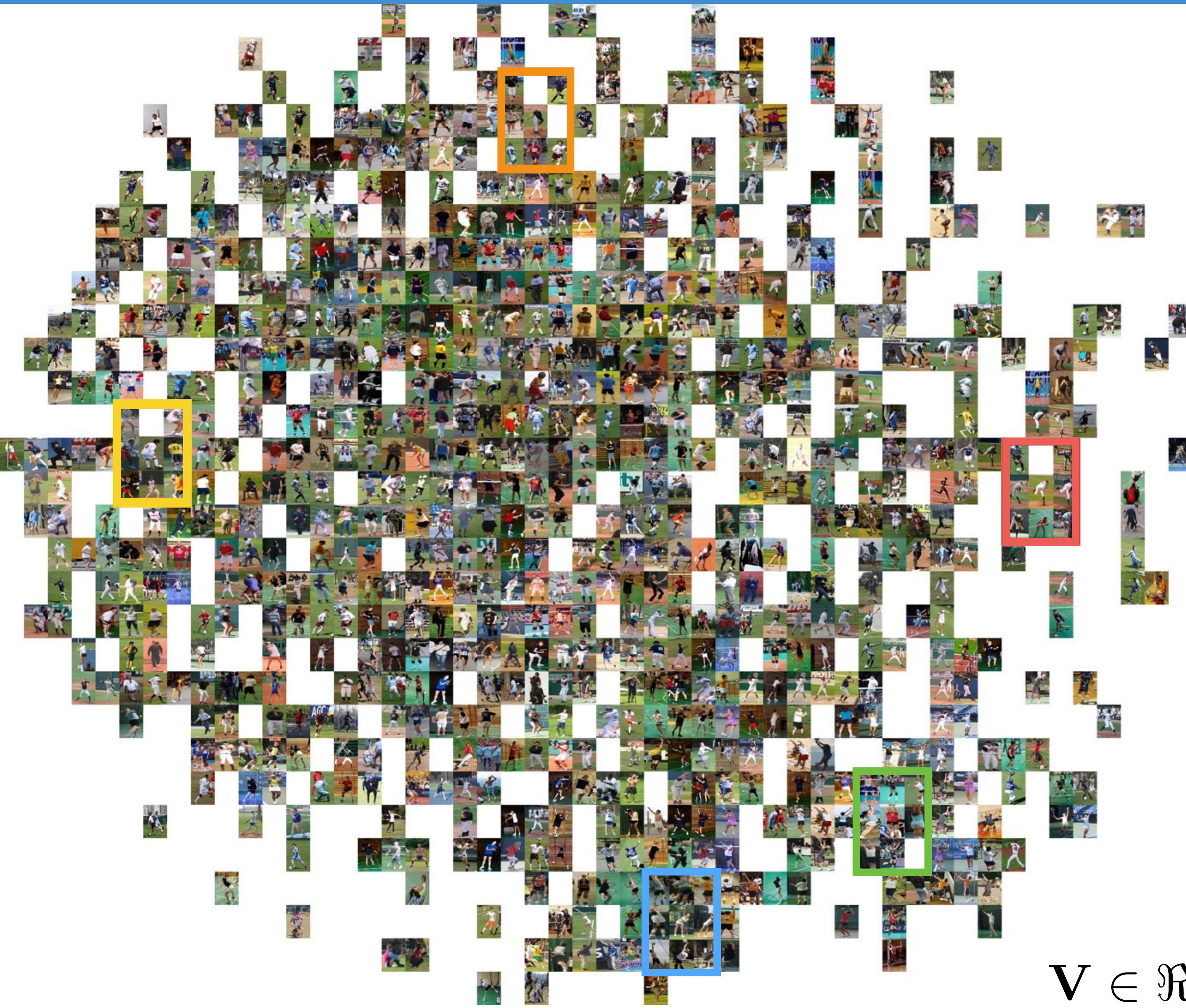
Tennis Forehand



Two parameters mainly affect visualization quality:

- **Number of latent factors:** set as the number of movemes in the data.
- **Coefficients of the V matrix:** constrained between 0 and 1.

Manifold Visualization

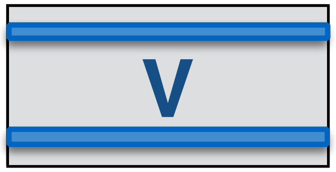


Legend:

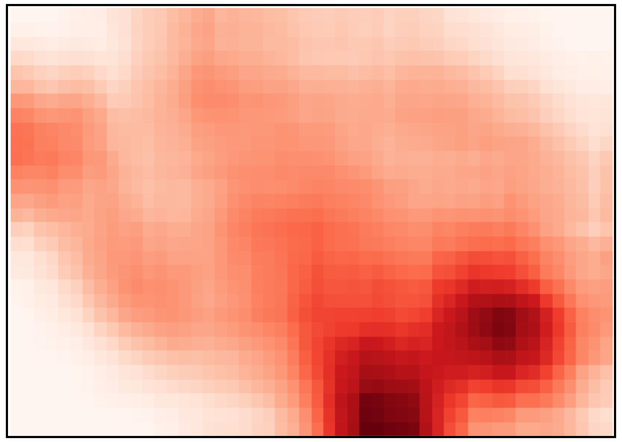
- baseball pitch
- running / lifting leg
- neutral /bending knees
- volley block / both arms up
- tennis serve / one arm up

bending knees lifting leg
baseball pitch
volleyball block tennis serve

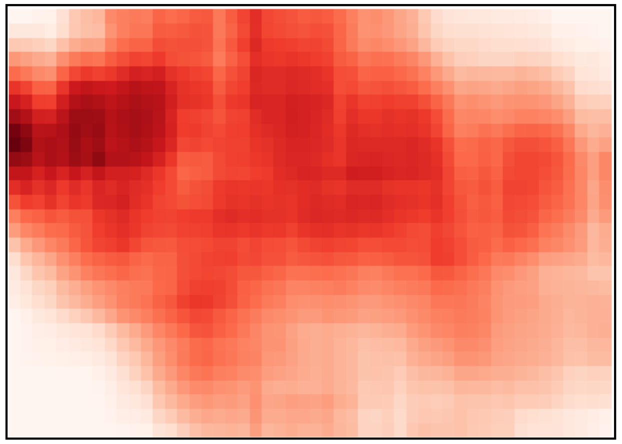
$$\mathbf{V} \in \Re^{k \times n}$$



Volleyball Strike



Soccer Kick



Summary

We developed an algorithm that can learn a 3-d basis for human movemes with the following properties:

- Global
- Rotation Invariant
- Sparse
- Complementary Representation

Limitations and possible improvements:

- Fixed number of latent factors
- Needs a more complex initialization
- End-to-end pose denoising and moveme learning
- Incorporate symmetry into movemes
- Apply to larger and more general/higher dimensional datasets

Questions?

Fork me on GitHub

github.com/matteorr/rotation_invariant_movemes