Learning to Recognize Parallel Combinations of Human Motion Primitives with Linguistic Descriptions using Non-Negative Matrix Factorization

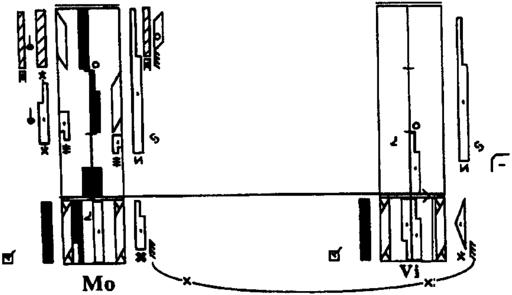
Olivier MANGIN, Pierre-Yves OUDEYER

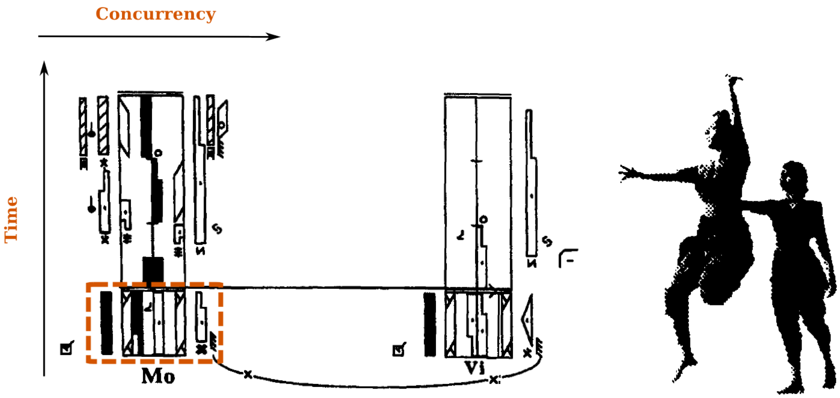
INRIA Flowers, Université Bordeaux 1, France

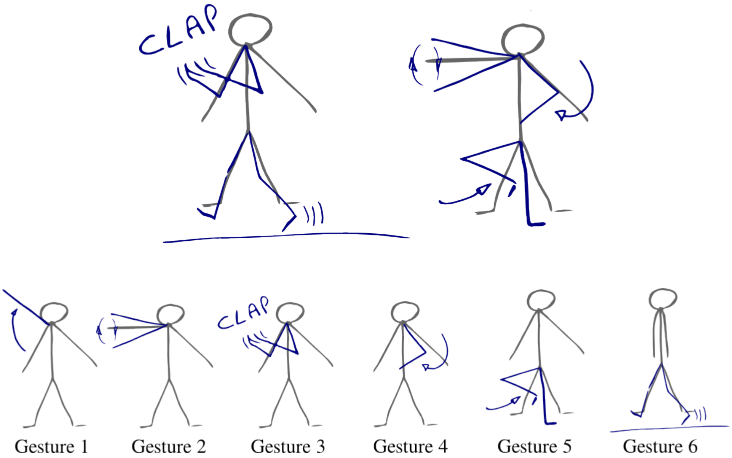
October 09, 2012

Motions and behaviors are complex

Picture: `http://www.hongkiat.com/`

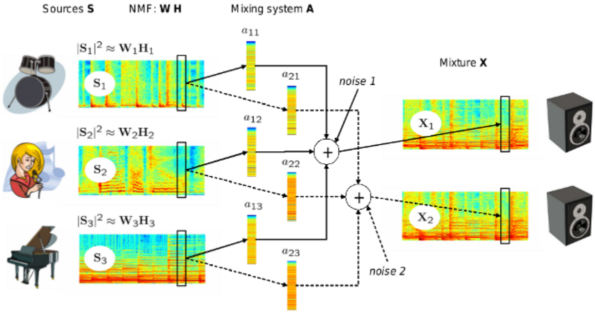

Mo
Vi


Concurrency
Time
Mo
Vi

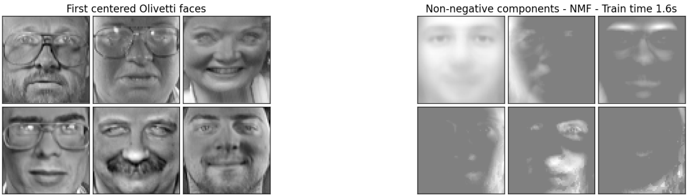


Learn a dictionary of **primitive motion** to explain/represent complex behaviors.

Dictionary learning in other fields



(Source http://perso.telecom-paristech.fr/˜fevotte/)



(Source Scikit-learn documentation)

An ambiguous problem

- invariances in the decomposition,
- levels of decomposition,
- real ambiguity in the behavior (e.g. cultural, contextual),
- ...

An ambiguous problem

- invariances in the decomposition,
- levels of decomposition,
- real ambiguity in the behavior (e.g. cultural, contextual),
- ...

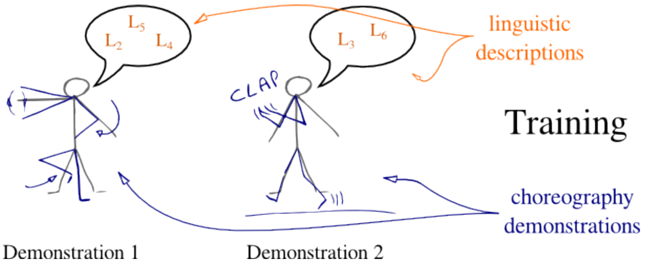
Heuristics to guide the dictionary learning:

- **structural**

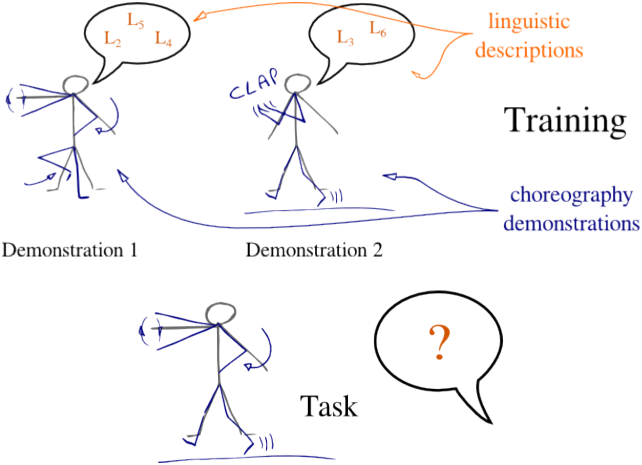
  non-negativity
- **social and multimodal**

  decomposition shaped by linguistic modality (also models **language grounding**)

Learning setup



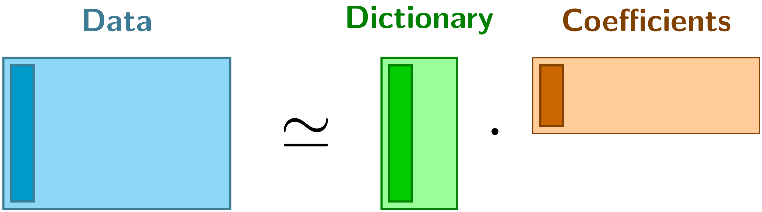
Learning setup



This setup is symetric to [Driesen et al., 2009].

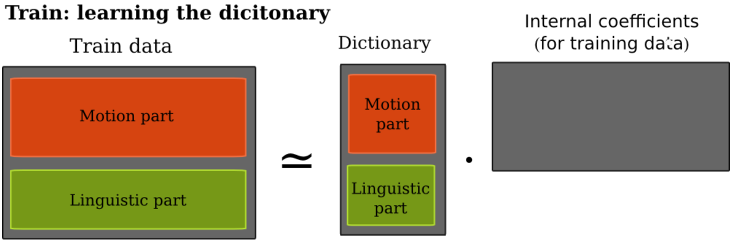
Non-negative matrix factorization (NMF)

The big picture:

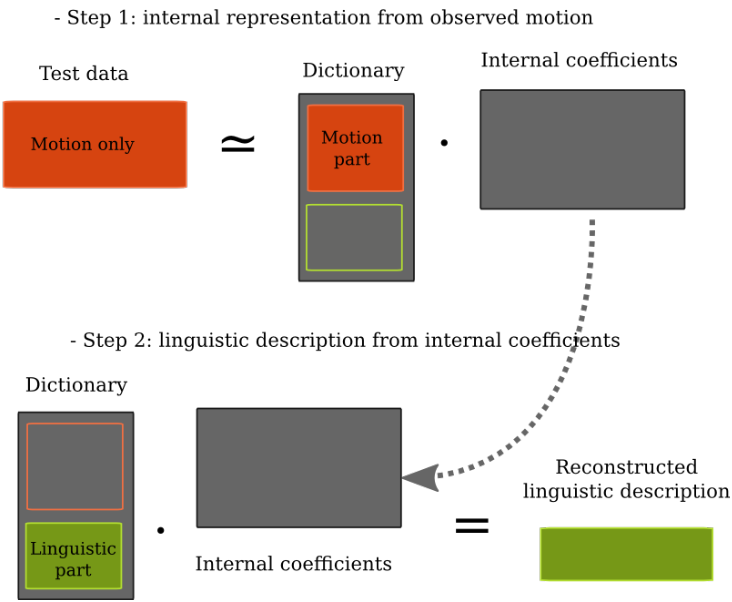


NMF for multimodal data

[Driesen et al., 2009, Mangin & Oudeyer, 2012b]



Test: producing linguistic data from demonstrated motions



Data representation

Motion features

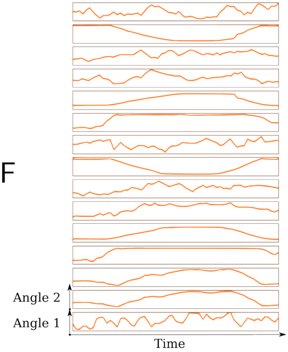
moving skeleton (from Kinect)

Data representation

Motion features

moving skeleton (from Kinect)

$\longrightarrow$ angles and angle velocities for each DOF



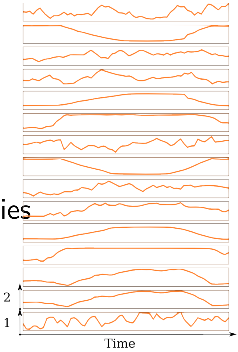
Data representation

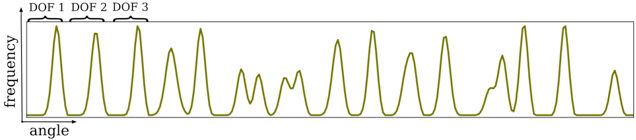
Motion features

moving skeleton (from Kinect)

$\longrightarrow$ angles and angle velocities for each DOF

$\longrightarrow$ histograms of positions or position-velocities





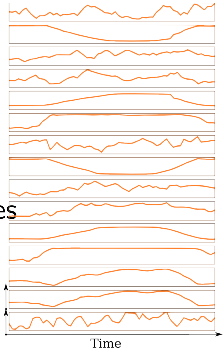
Data representation

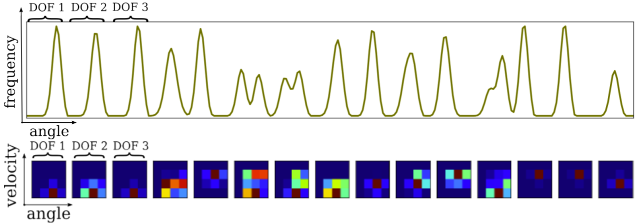
Motion features

moving skeleton (from Kinect)

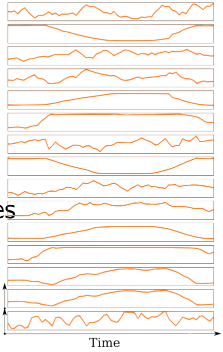
⟶ angles and angle velocities for each DOF

⟶ histograms of positions or position-velocities





Data representation



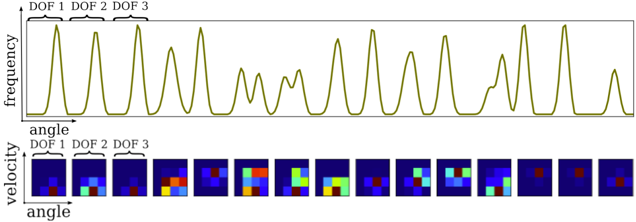
Motion features

moving skeleton (from Kinect)

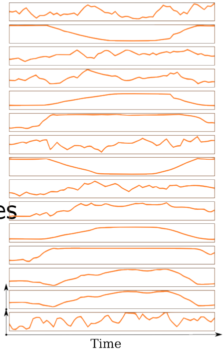
⟶ angles and angle velocities for each DOF

⟶ histograms of positions or position-velocities

⟶ flattened in a non-negative vector



Data representation



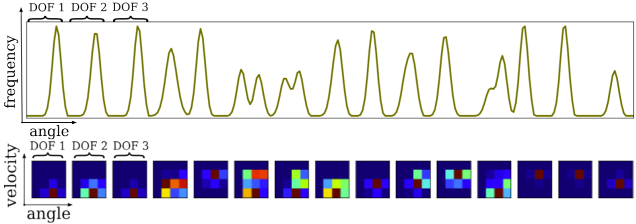
Motion features

moving skeleton (from Kinect)

$\longrightarrow$ angles and angle velocities for each DOF

$\longrightarrow$ histograms of positions or position-velocities

$\longrightarrow$ flattened in a non-negative vector



Language representation

$[0, 0, 1, 0, 0, 0, 0, 0, 1, 0] \quad \longrightarrow$ symbols 2 and 8

The choreography dataset

Available online:

http://flowers.inria.fr/choreography_database.html

3 separate sets of examples:

- **primitive:** only one primitive motion demonstrated in each example

  (**326** examples, **47** primitive motions)

- **mixed small:** complex choreographies

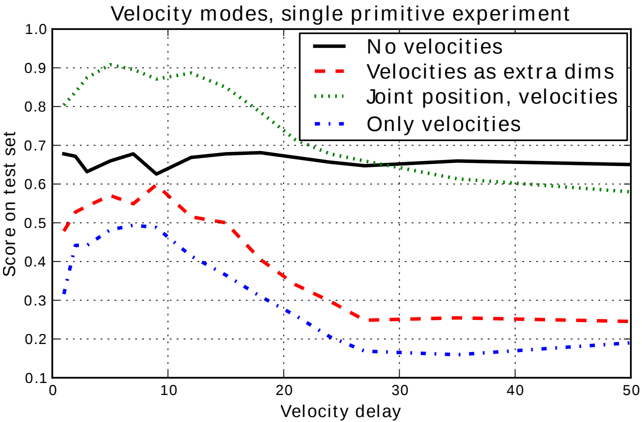
  (**137** examples, **16** primitive motions)

- **mixed full:** complex choreographies

  (**277** examples, **47** primitive motions)

Evaluation on classification

(only one primitive motion at a time)



Evaluation on reconstruction of multiple labels

(reconstructed description is thresholded and compared to human annotations)

	l_{full}
17 labels (SVM, linear)	0.818
17 labels (NMF, Frobenius)	0.854
17 labels (NMF, DKL)	0.789
47 labels (SVM, linear)	0.422
47 labels (NMF, Frobenius)	0.625
47 labels (NMF, DKL)	0.574

Evaluation on reconstruction of multiple labels

(reconstructed description is thresholded and compared to human annotations)

	l_{full}
17 labels (SVM, linear)	0.818
17 labels (NMF, Frobenius)	0.854
17 labels (NMF, DKL)	0.789
47 labels (SVM, linear)	0.422
47 labels (NMF, Frobenius)	0.625
47 labels (NMF, DKL)	0.574

Results for unobserved combinations

	l_{full}
17 labels (NMF, Frobenius)	0.568
17 labels (SVM, linear)	0.667
47 labels (NMF, Frobenius)	0.406
47 labels (SVM, linear)	0.206

Extensions

- **Other constraints** (sparsity, etc.)
- **Going to real language** (merge with [Driesen et al., 2009])
- **Learn position in time**
- **Other motion representations (towards imitation):** models of activities as combination of potential functions [Mangin & Oudeyer, 2012a]

Thank you!

Bibliography

Driesen, J., ten Bosch, L., & Van Hamme, H. (2009).
Adaptive Non-negative Matrix Factorization in a Computational Model of Language Acquisition.
In *Interspeech* (pp. 1–4).

Mangin, O. & Oudeyer, P.-Y. (2012a).
Learning the combinatorial structure of demonstrated behaviors with inverse feedback control.
In *to appear in third International Workshop on Human Behavior Understanding*, number 3 Vilamoura, Algarve (Portugal).

Mangin, O. & Oudeyer, P.-Y. (2012b).
Learning to recognize parallel combinations of human motion primitives with linguistic descriptions using non-negative matrix factorization.
In *to appear in International Conference on Intelligent Robots and Systems (IROS 2012)* Vilamoura, Algarve (Portugal): IEEE/RSJ.

Evaluation on reconstruction of multiple labels

(reconstructed description is thresholded and compared to human annotations)

	l_{full}	$l_{\text{given number}}$
17 labels (SVM, linear)	0.818	-
17 labels (NMF, Frobenius)	0.854	0.971
17 labels (NMF, DKL)	0.789	0.905
47 labels (SVM, linear)	0.422	-
47 labels (NMF, Frobenius)	0.625	0.755
47 labels (NMF, DKL)	0.574	0.679

Results for unobserved combinations

	l_{full}	$l_{\text{given number}}$
17 labels (NMF, Frobenius)	0.568	0.800
17 labels (SVM, linear)	0.667	-
47 labels (NMF, Frobenius)	0.406	0.653
47 labels (SVM, linear)	0.206	-