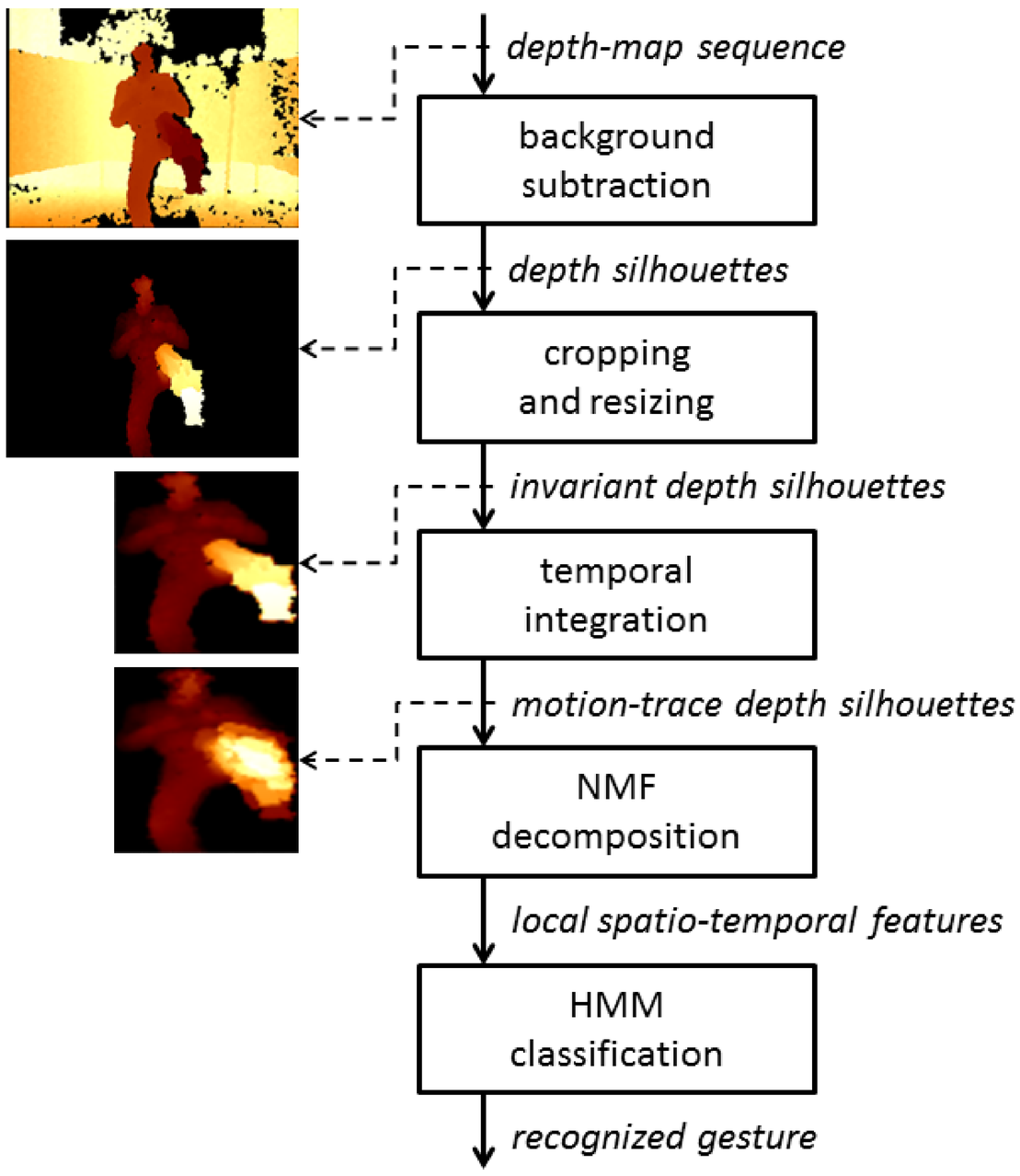
GESTURE RECOGNITION USING A NMF-BASED REPRESENTATION OF MOTION-TRACES EXTRACTED FROM DEPTH SILHOUETTES

Aymeric Masurelle, Slim Essid, Gaël Richard
Institut Mines-Télécom/Télécom ParisTech, CNRS-LTCI, Paris, France
aymeric.masurelle@telecom-paristech.fr

Introduction

- **Task :**
To recognize full-body gestures from depth-map sequences captured by a monocular depth sensor placed in front of the performer as in Human-Computer Interaction applications.
- **Model :**
Gesture = a concatenation of atomic sequences of motions and poses of different body parts.
- **Proposal :** A novel approach that classifies full-body human gestures using original spatio-temporal features obtained by applying non-negative matrix factorisation (NMF) to an extended depth silhouette representation.

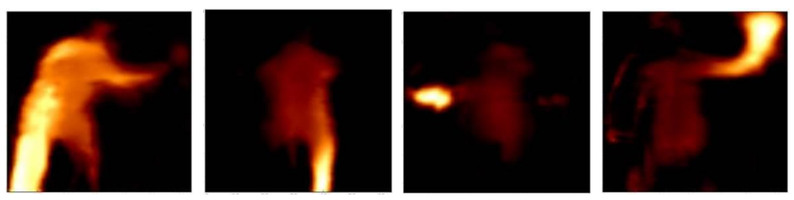
Overview of our system



NMF representation process

To achieve the proposed **NMF representation process**, we have to perform two consecutive steps:

- learning step : to create the **local spatio-temporal feature dictionary**, **W** (randomly initialized), using the entire training set.
- representation step : to project each incoming motion-trace on the dictionary components to obtain the **activation vectors** which are used as motion features.



NMF component examples from the created local spatio-temporal feature dictionary.

Experimental evaluation

- **Dataset :**
A subset of the *Huawei/3DLife 3D human reconstruction and action recognition Grand Challenge* database (`http://mmv.eecs.qmul.ac.uk/mmgc2013/`) :

# of performers	8
# of different gestures	8
# of gesture repetitions per participant	5

- **Evaluation procedure :**
  - *leave-one-participant-out* cross-validation (8 folds),
  - adjustment of motion feature dimension (PCA, NMF),
  - model hyper-parameter tuning (HMM),
  - metrics : F-measure uniformly averaged over all gestures and over all folds.
- **Reference systems :**
  - Angles deduced from main body joints [1]
  - PCA-representation of main body joints trajectories [2]
  - PCA-representation of motion trace, adapted from [3]

Classification results

Best classification results in F-measure on 8 gesture classes with HMM classifiers

Approaches			
[1]	[2]	[3]	**proposed**
78% (Q=6)	89% (w=6, d=30, Q=2)	89% (w=5, d=40, Q=2)	**91%** (w=6, k=128, Q=7)

(w: window size in samples, d and k feature dimension (resp. PCA and NMF), Q: number of HMM states)

- The method [1] only based on global pose features is outperformed by the other methods. that incorporate motion dynamics and a decomposition process through their feature representation.
- Moreover the use of a NMF decomposition technique is more suitable to represent spatio-temporal motion features compared to PCA as the performance of our system is superior to the other reference systems.

Conclusion

- Our system obtains better performance than reference systems based on static features, PCA decomposition and/or 3D body joint positions.
- This highlights the importance of considering a gesture as a concatenation of atomic sequences of motions and poses of different body parts.

[1] G.Th. Papadopoulos, A. Axenopoulos and P. Daras, Real-time Skeleton-tracking-based Human Action Recognition Using Kinect Data, in *Proc. of the 20th International Conference on MultiMedia Modeling*, 2014.

[2] A. Masurelle, S. Essid and G. Richard, Multimodal Classification of Dance Movements using Body Joint Trajectories and Step Sounds, in *International Workshop on Image and Audio Analysis for Multimedia Interactive Services WIAMIS*, 2013.

[3] R. Muñoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas and A. Carmona-Poyato, Depth silhouettes for gesture recognition, in *Pattern Recognition Letters*, vol.29, no.3, pp.319-329, 2008.