Spatial and Rotation Invariant 3D Gesture Recognition Based on Sparse Representation

Ferran Argelaguet[1], Mélanie Ducoffe[2], Anatole Lécuyer[1], Remi Gribonval[1]

[1] Inria, IRISA, France
[2] ENS Rennes, France

Introduction

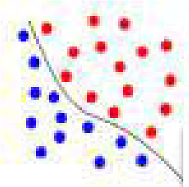
- Gesture interaction exploits the **affordances** and **experience** of users
  - User's knowledge of the real world


Tilt Brush

- A **gesture** can be considered as a meaningful and intentional movement
  - Encoded information based on the spatial, pathic, symbolic and affective characteristics [Mitra et al. 2007]

Context

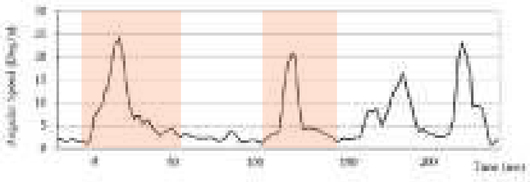
Pattern Analysis



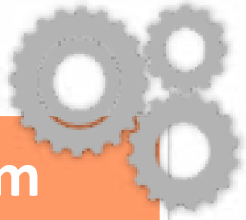
Classification

Recognition

Segmentation



User Interface



System Training

Usage

User Experience

Gesture Classification

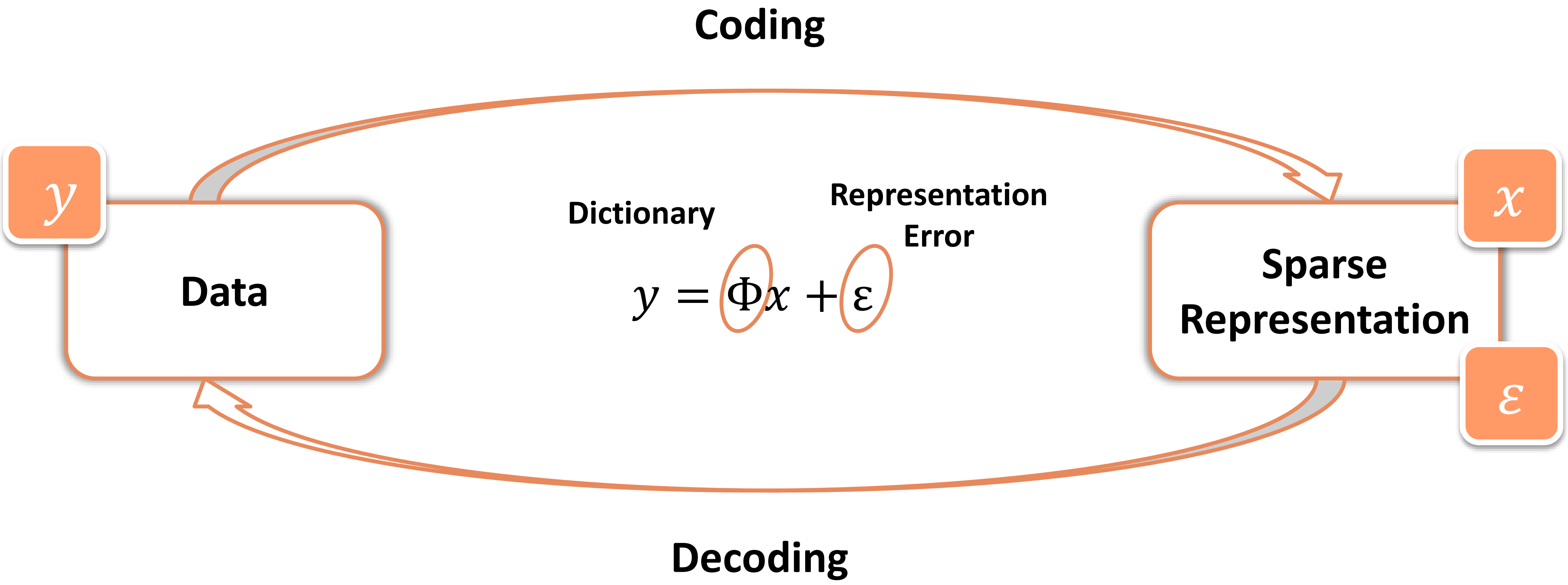
Feature-based

- Gesture characterization through a set of distinctive features
  - Mean speed, curvature... **[Chen et al. 2013]**
- A broad range of alternatives
  - Hidden Markov Models **[Chen et al. 2013]**
  - Nearest Neighbors **[Lai et al. 2012]**
  - Support Vector Machines **[Kela et al. 2006]**

Template-based

- Gesture characterization through a set of representative gestures
- Error minimization based on a distance function
  - Mean Square Error **[Woobroock et al. 2007]**
  - Angular Inverse Cosinus **[Li 2010]**
  - Dynamic Time Warping **[Liu et al. 2009]**

Sparse Representation



Sparse Representation

Encoding

y — **Data**

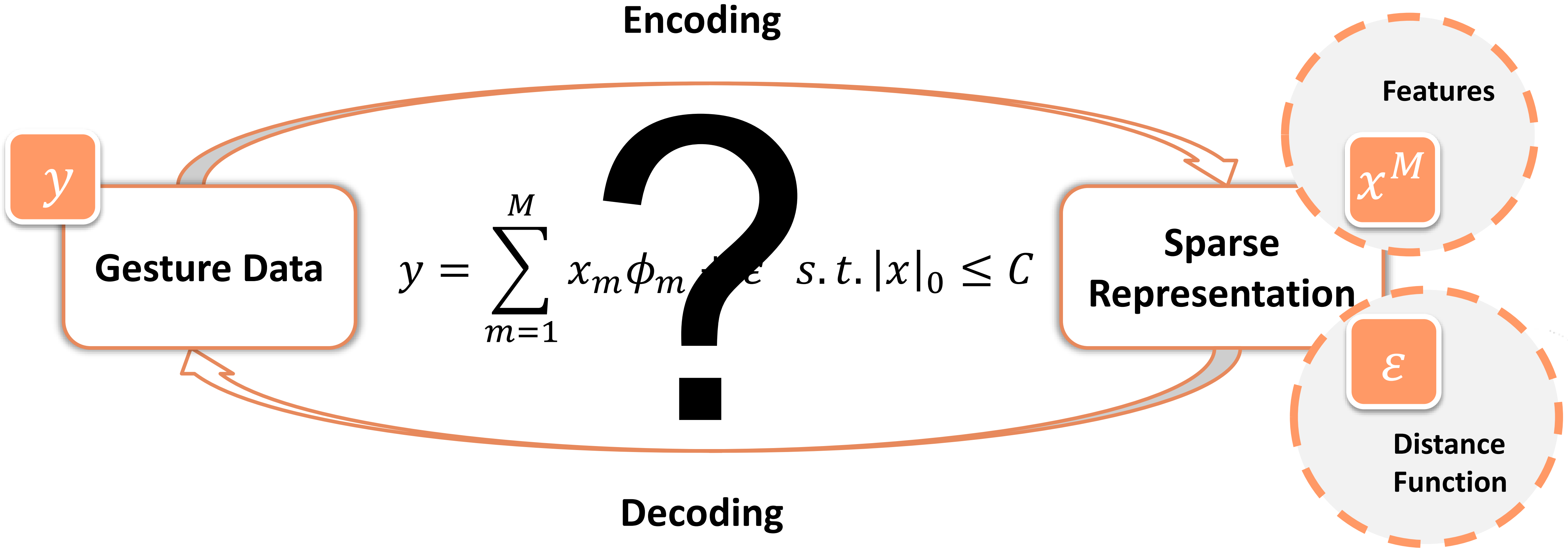
Sparsity Contraint

$$y = \sum_{m=1}^{M} x_m \phi_m + \varepsilon \quad s.t. |x|_0 \leq C$$

Sparse Representation — x^M, ε

Decoding

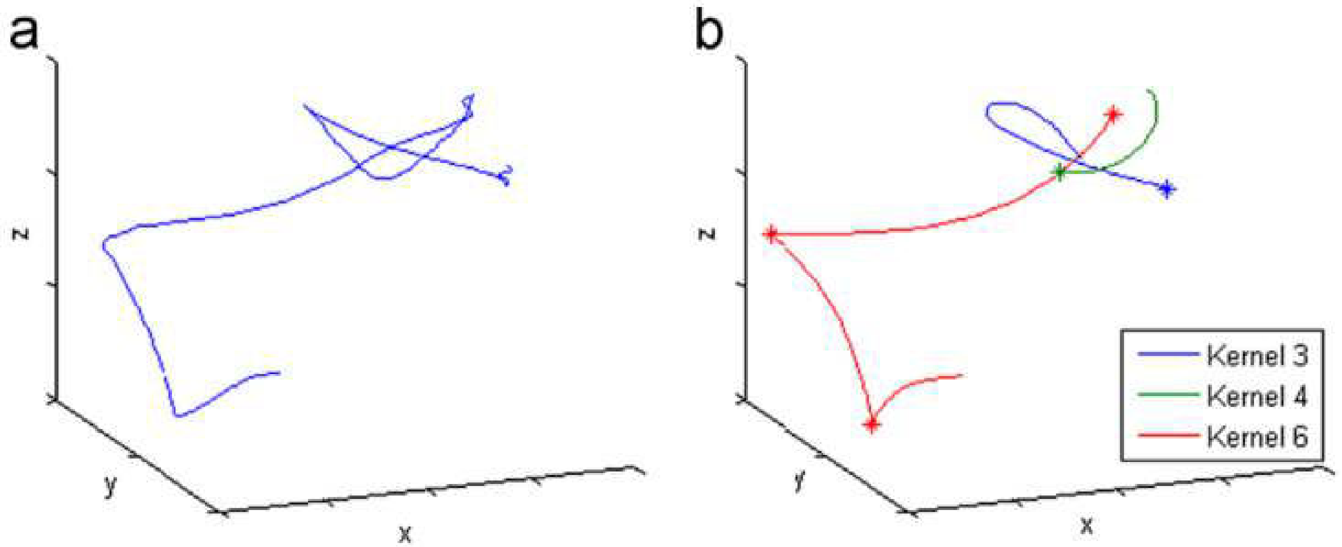
Sparse Representation-based Gesture Classification



Representation of Human Motion

- 3D rotation invariant sparse representation **[Barthelemy et al. 2014]**
- The dictionary is invariant to:
  - Scale
  - Rotation
  - Time

$$y = \Phi x + \varepsilon = \sum_{m=1}^{M} \underline{x_m}\,\underline{R_m}\,\phi_{m,\underline{\tau_m}} + \varepsilon \quad s.t. |x|_0 \leq C$$



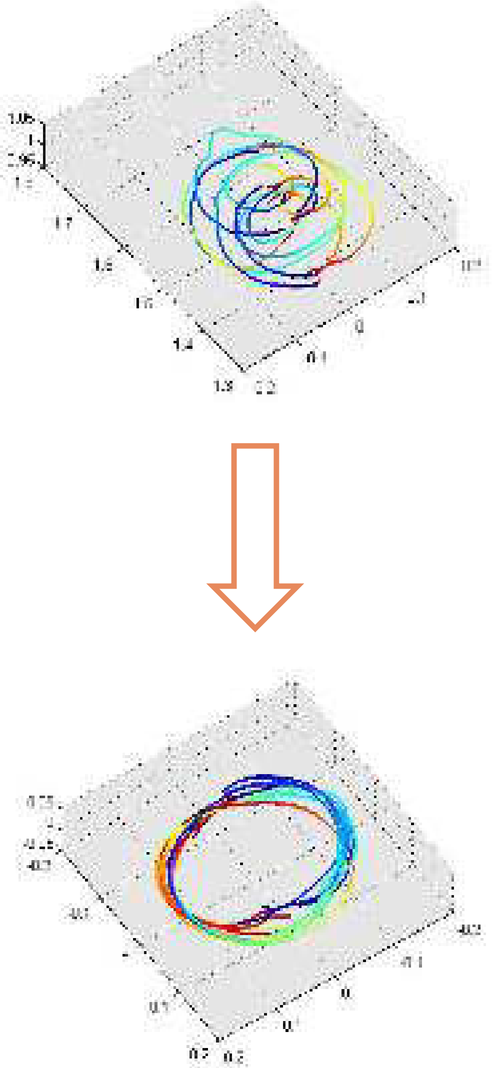
Contributions

- ➢Gesture Normalization and Resampling
- ➢Classification Strategies
- ➢Dictionary Learning Strategies
- ➢Evaluation
  - Off-line and On-line

Gesture Normalization and Resampling

$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon$$

- Decrease gesture variability in terms of **speed** and **position**
- Transform the gesture motion into a speed invariant domain
  - Curvi-linear abscissa
- The centroid of the gesture is removed

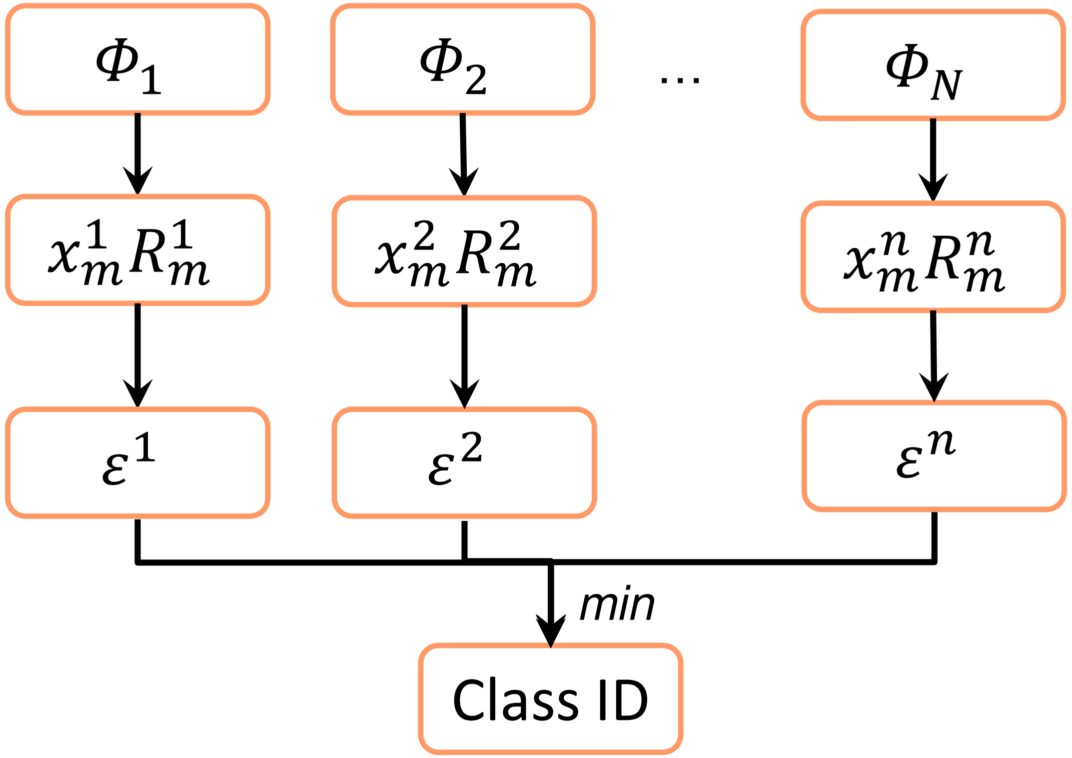


Gesture Classification Strategies

$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon$$

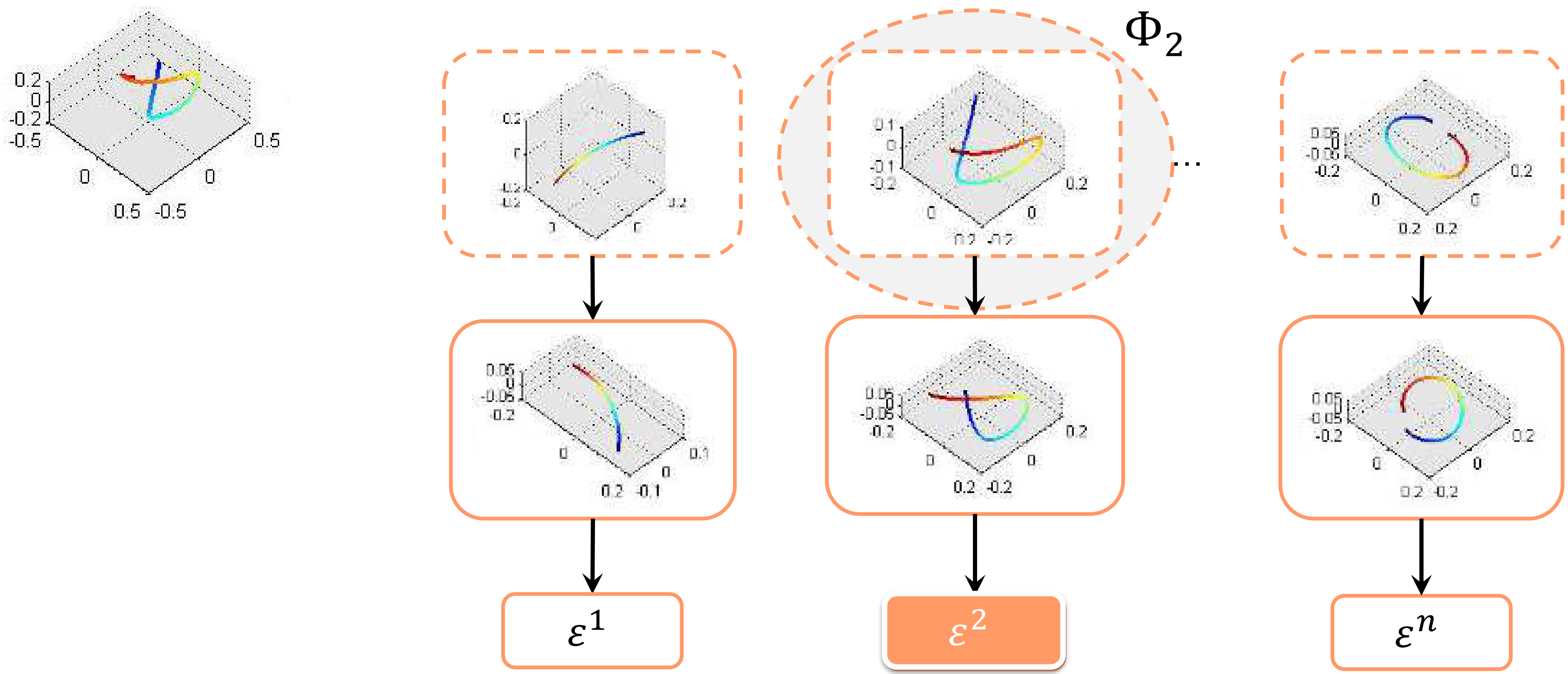
Feature-based x^M

Template-based ε

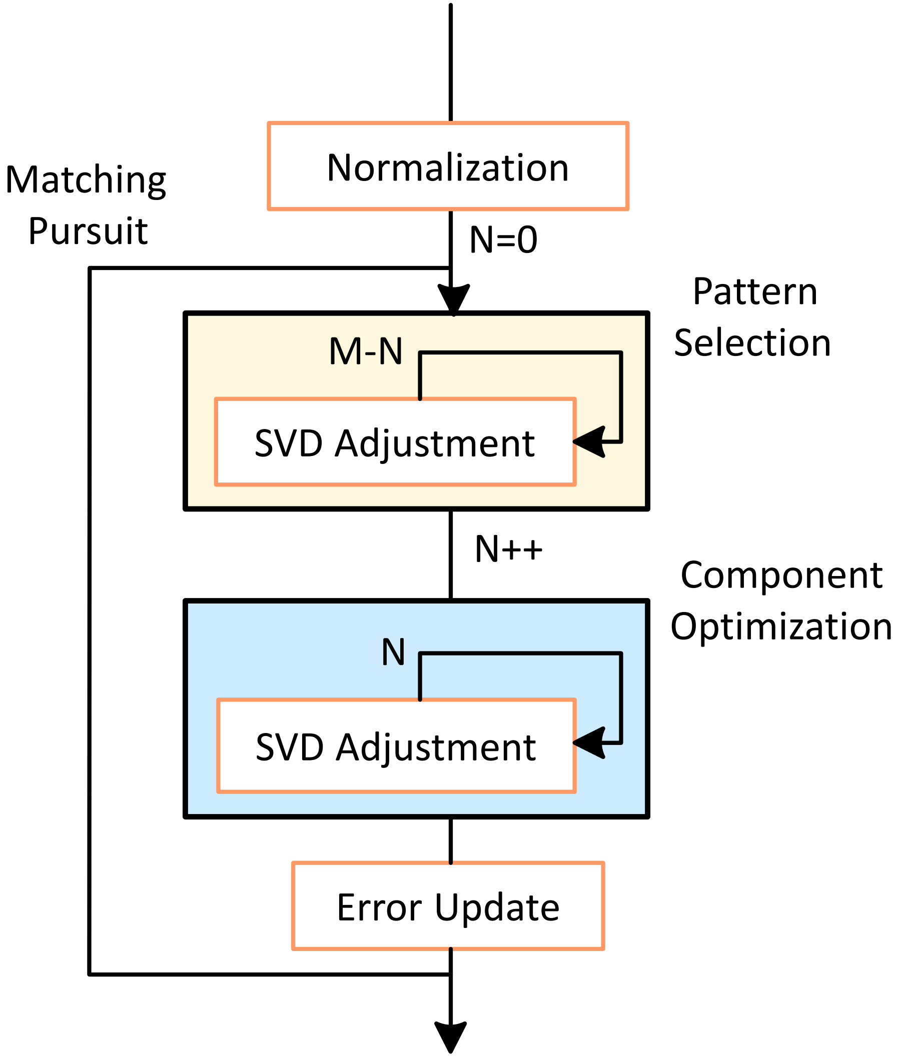


Gesture Classification Strategies

$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon$$



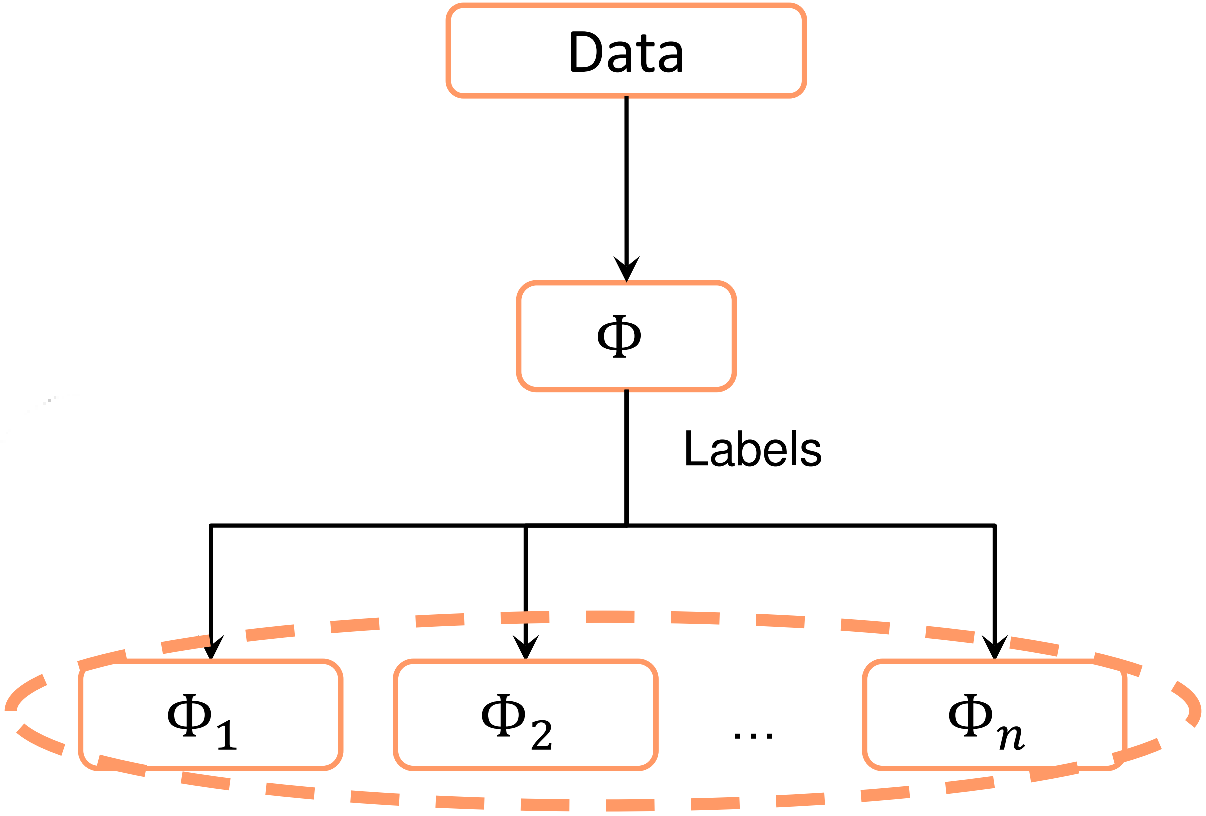
Sparse Coding Algorithm



Dictionary (Φ) Learning Strategies

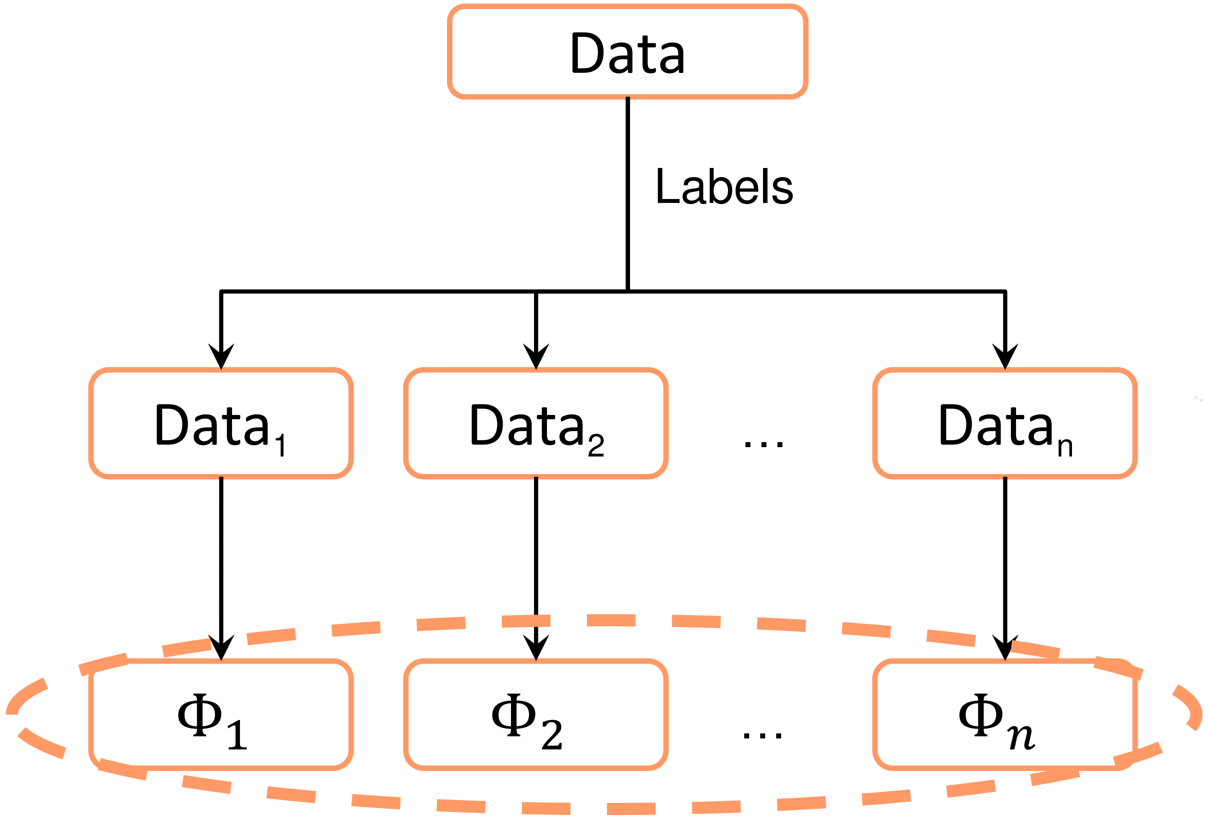
$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon$$

Semi-supervised Learning



Variable number of patterns

Supervised Learning

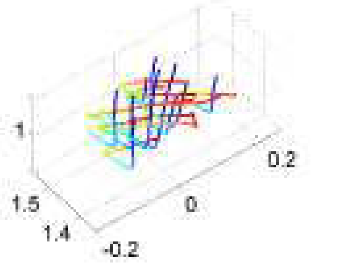


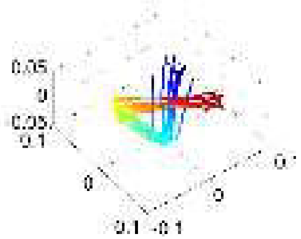
Fixed number of patterns

Overview

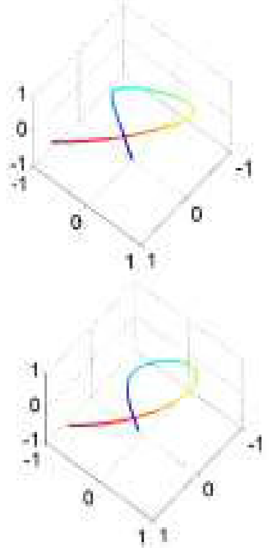
$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon$$

Training

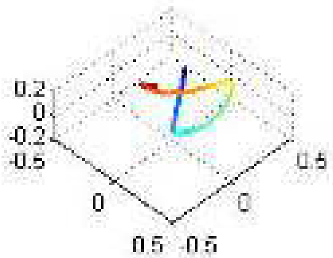


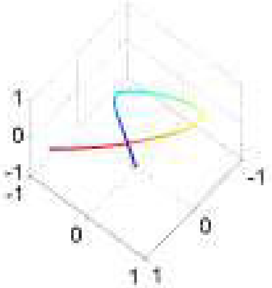


$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon \;\; s.t \;\; |\, x\,|_0 \leq C$$



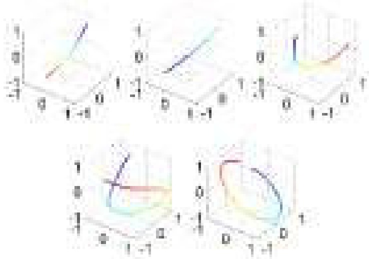
Classification





$$y = \sum_{m=1}^{M} x_m R_m \phi_{m,\tau_m} + \varepsilon \;\; s.t \;\; |\, x\,|_0 \leq C$$

$\Phi=$



$x^M, R^M, \tau^M, \varepsilon$

Evaluation

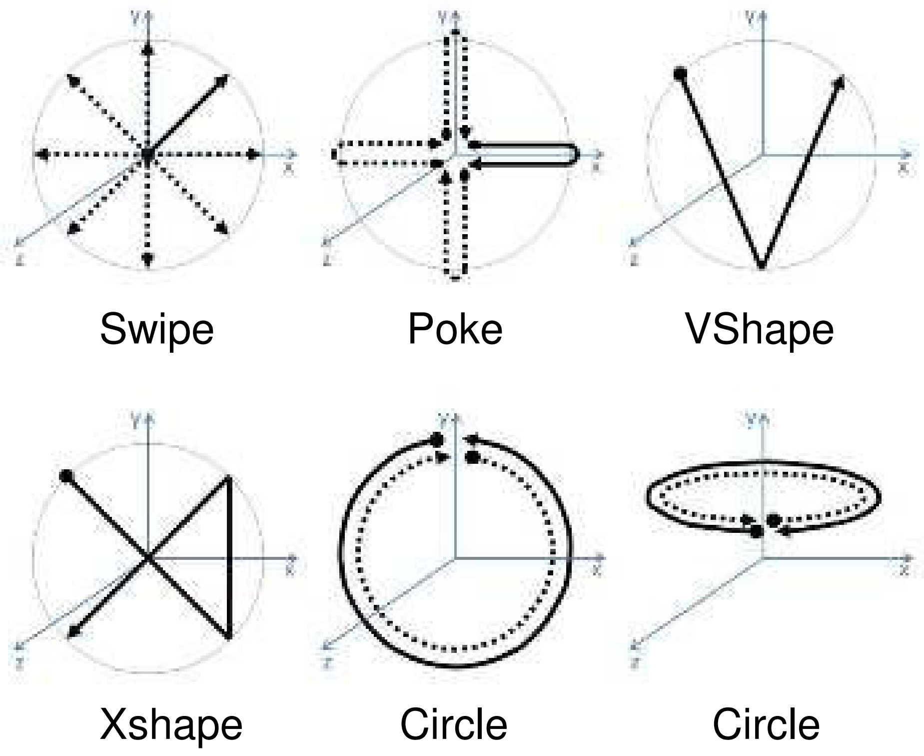
Off-Line Evaluation

- 6DMG Gesture Database **[Chen et al. 2013]**
  - 5600 samples from 28 users (7 left-handed)
  - 20 different gestures
  - 6 gesture classes (only five are considered)

- Dictionary Parametrization
  - Sparse components – 1 Component
  - Gesture sub-sampling - 60 Samples
  - Dictionary Patterns – 5 to 10 Patterns

- Conditions
  - User Dependent vs User Independent
  - Supervised DLA vs Semi-Supervised DLA
  - Class Recognition vs Total Recognition

6DMG Database



Off-Line Evaluation Results

- User Dependent vs User Independent
  - Higher recognition rate for the user dependent condition
  - No impact on users' dominant hand

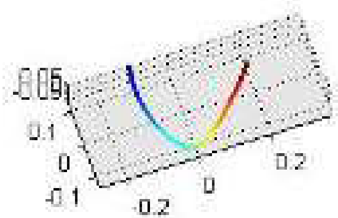
- Supervised vs Semi-Supervised
  - No significant differences

- Class vs Total Classification
  - Lower recognition rate for the total condition

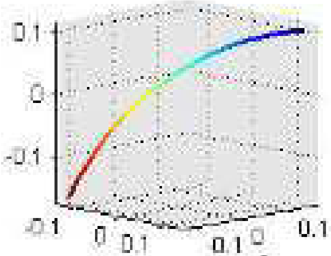
	3D Sparse Representation			
	Supervised		Semi-Supervised	
	Class	Total	Class	Total
UD	99.74%	**99.55%**	**99.75%**	98.22%
UI	99.18%	97.73%	**99.33%**	**97.82%**

Off-Line Evaluation Results (II)

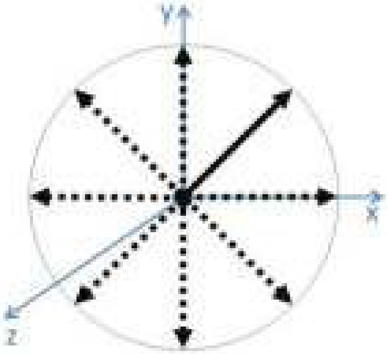
- Confusions between the Swipe and VSphape for the user independent condition

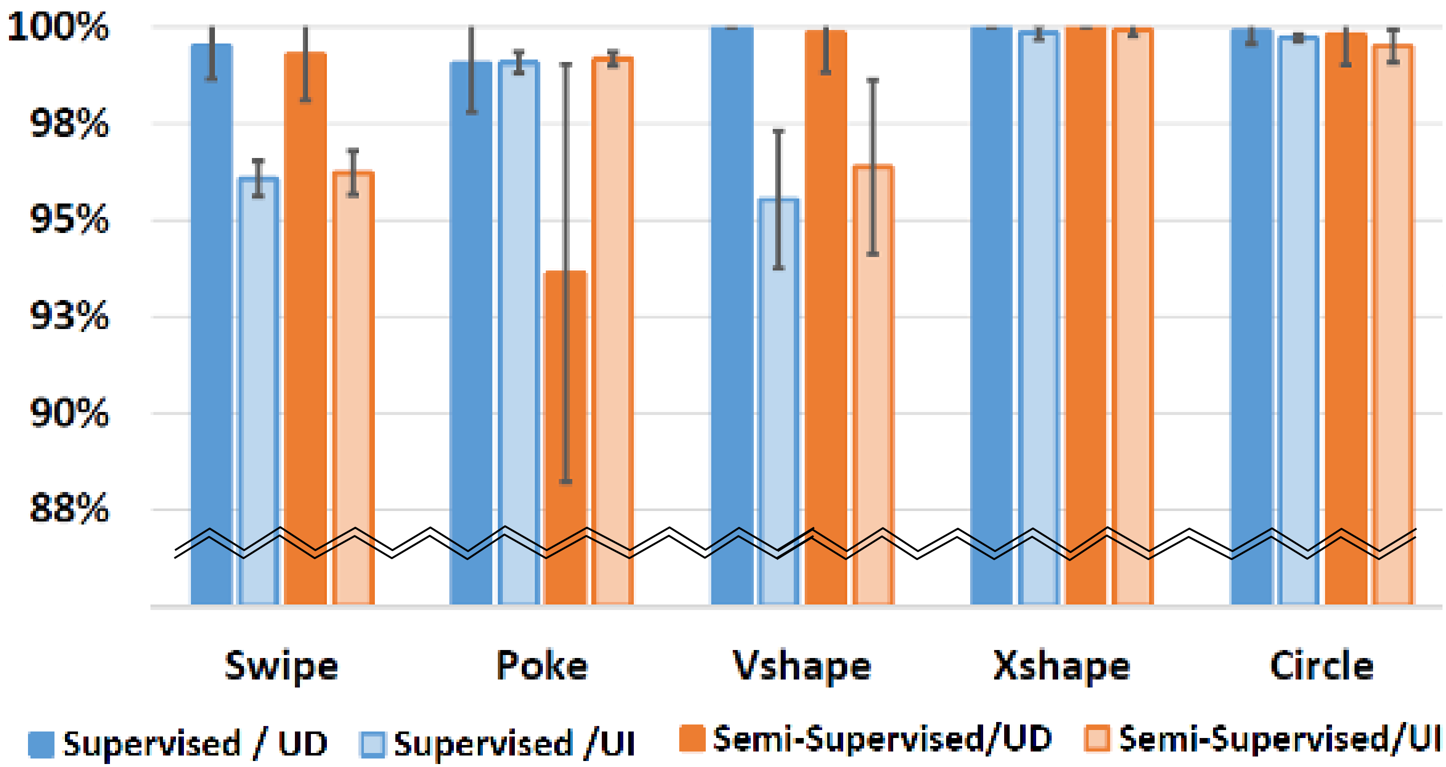


VSphape



Swipe

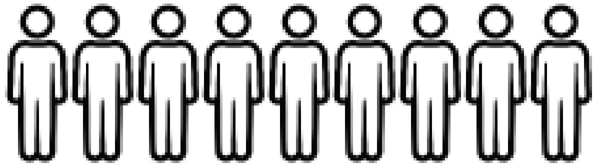




Real-Time User Evaluation

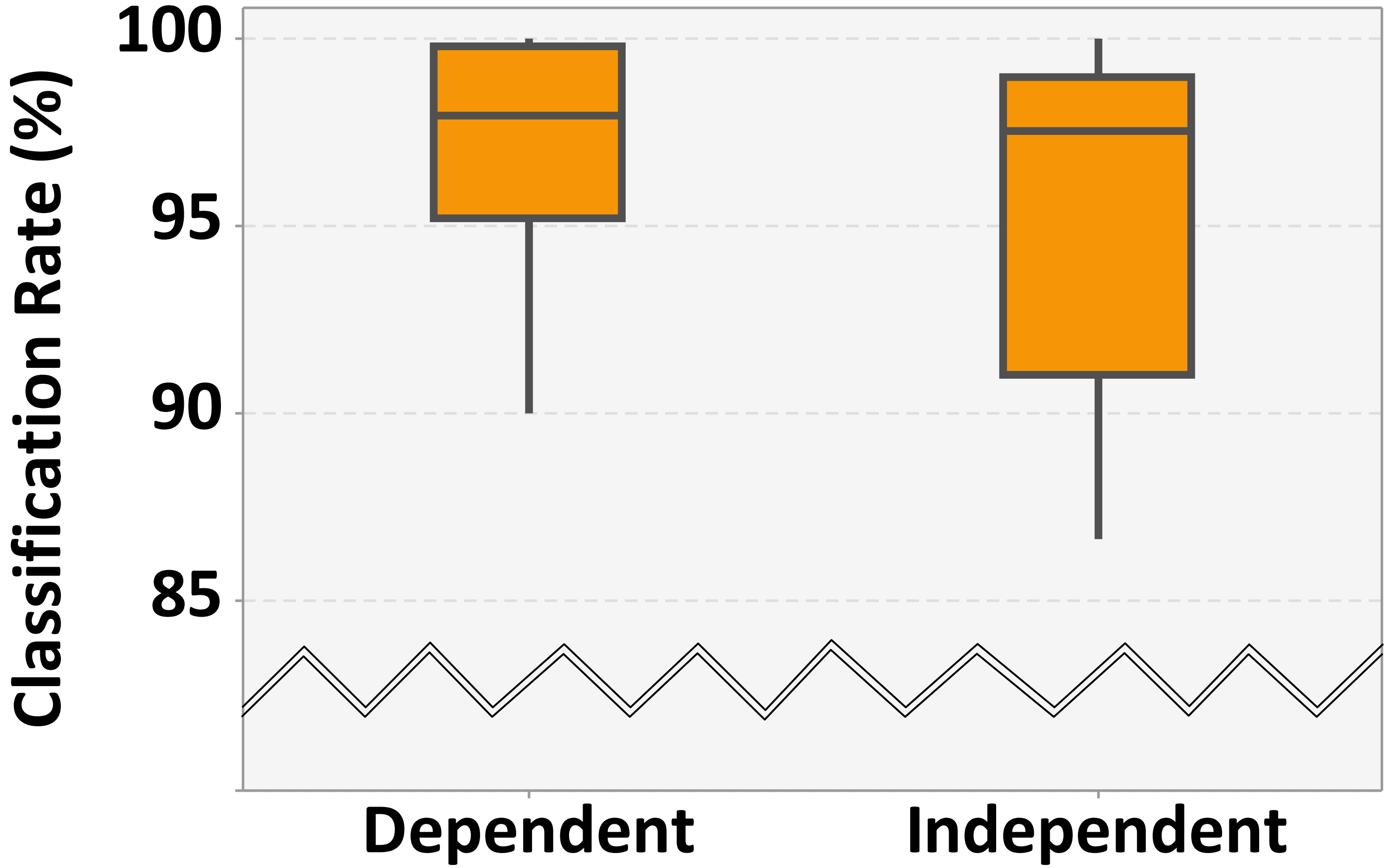
- Evaluate the gesture recognition system in real time conditions
  - Button-based segmentation

- User Independent Scenario
  - Precomputed dictionary based on the 6DMG database

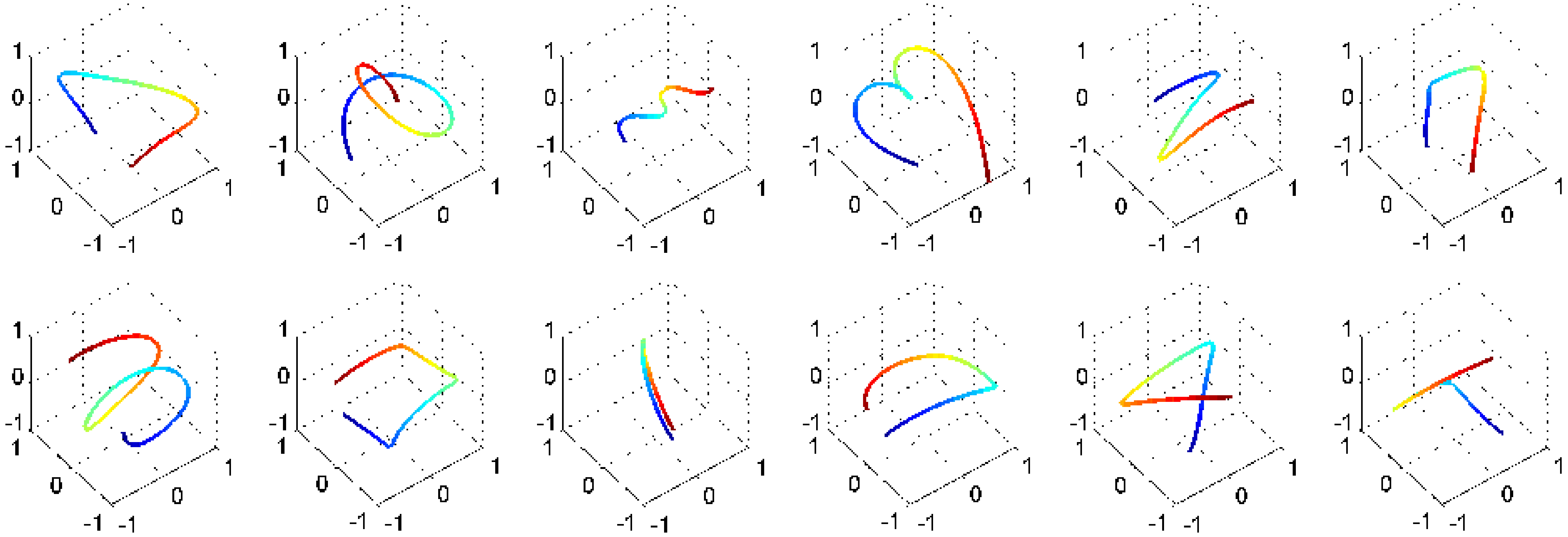


- User Dependent Scenario
  - Design of six gestures
  - The dictionary is trained by the user

Real-Time User Evaluation Results



User Defined Gestures



Conclusion

Conclusion

- Need of robust and efficient gesture recognition classifiers
  - Error-prone motion-based interfaces

- Sparse-based gesture classification algorithm
  - Speed, scale and rotation invariances
  - Reduced number of gesture samples to train the classifier
  - Off-line and real-time experimental validation

- Future perspectives
  - Additional evaluations with other classification algorithms and gesture databases
  - Support segmentation and additional degrees of freedom