Probability-based Dynamic Time Warping for Gesture Recognition on RGB-D data

Miguel Ángel Bautista[1,2]
Antonio Hernández-Vela[1,2]
Xavier Perez-Sala[2,3]
Victor Ponce Lopez[1,2]
Xavier Baro[2,4]
Oriol Pujol[1,2]
Cecilio Angulo[3]
Sergio Escalera[1,2]

[1]*Dept. Applied Mathematics and Analysis, Universitat de Barcelona*
[2]*Computer Vision Center, Universitat Autónoma de Barcelona*
[3]*CETpD, Universitat Politècnica de Catalunya*
[4]*EIMT, Universitat Oberta de Catalunya*

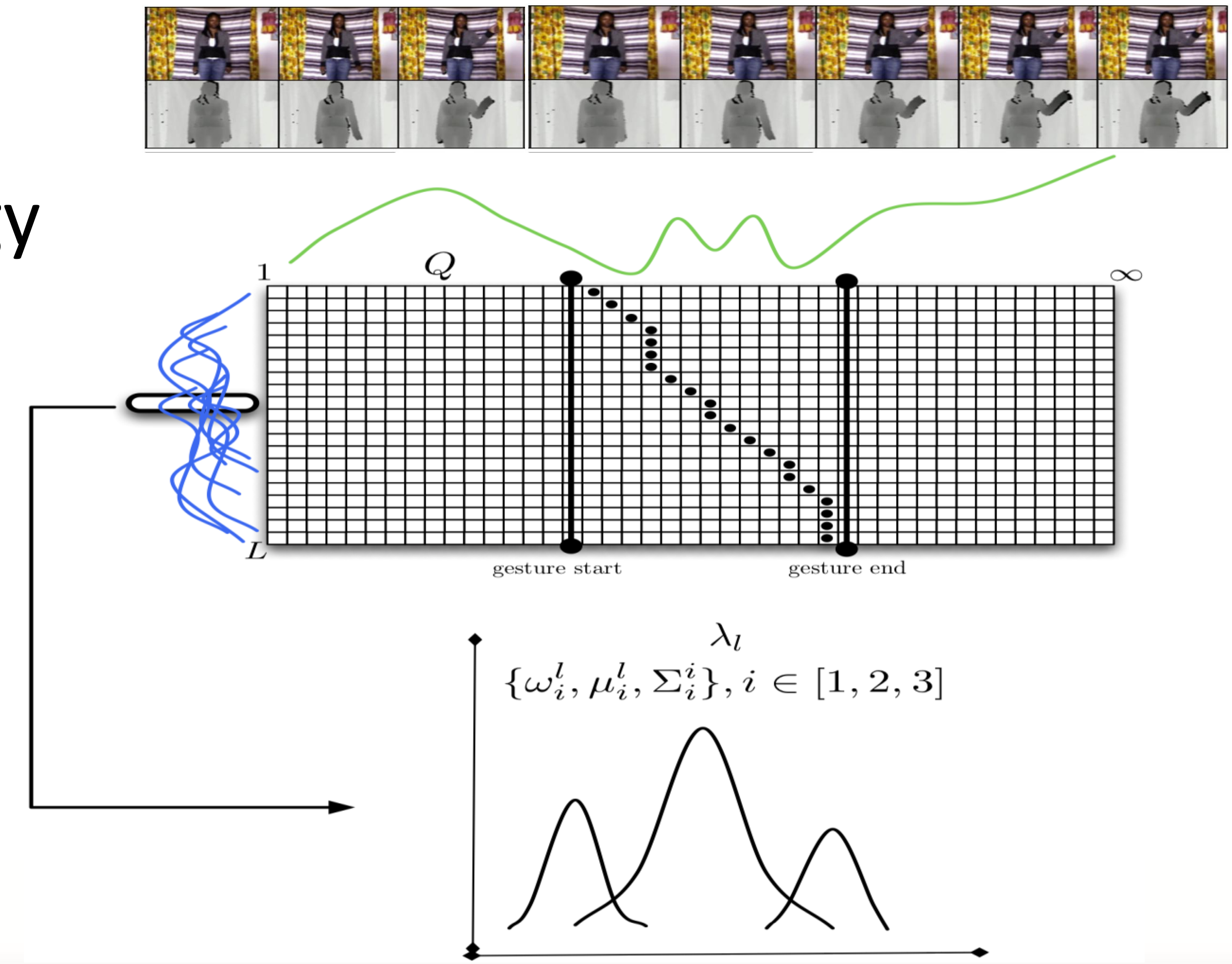
Computer Vision Center

Universitat Oberta de Catalunya

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Outline

1. Introduction
2. Methodology
3. Results
4. Conclusion

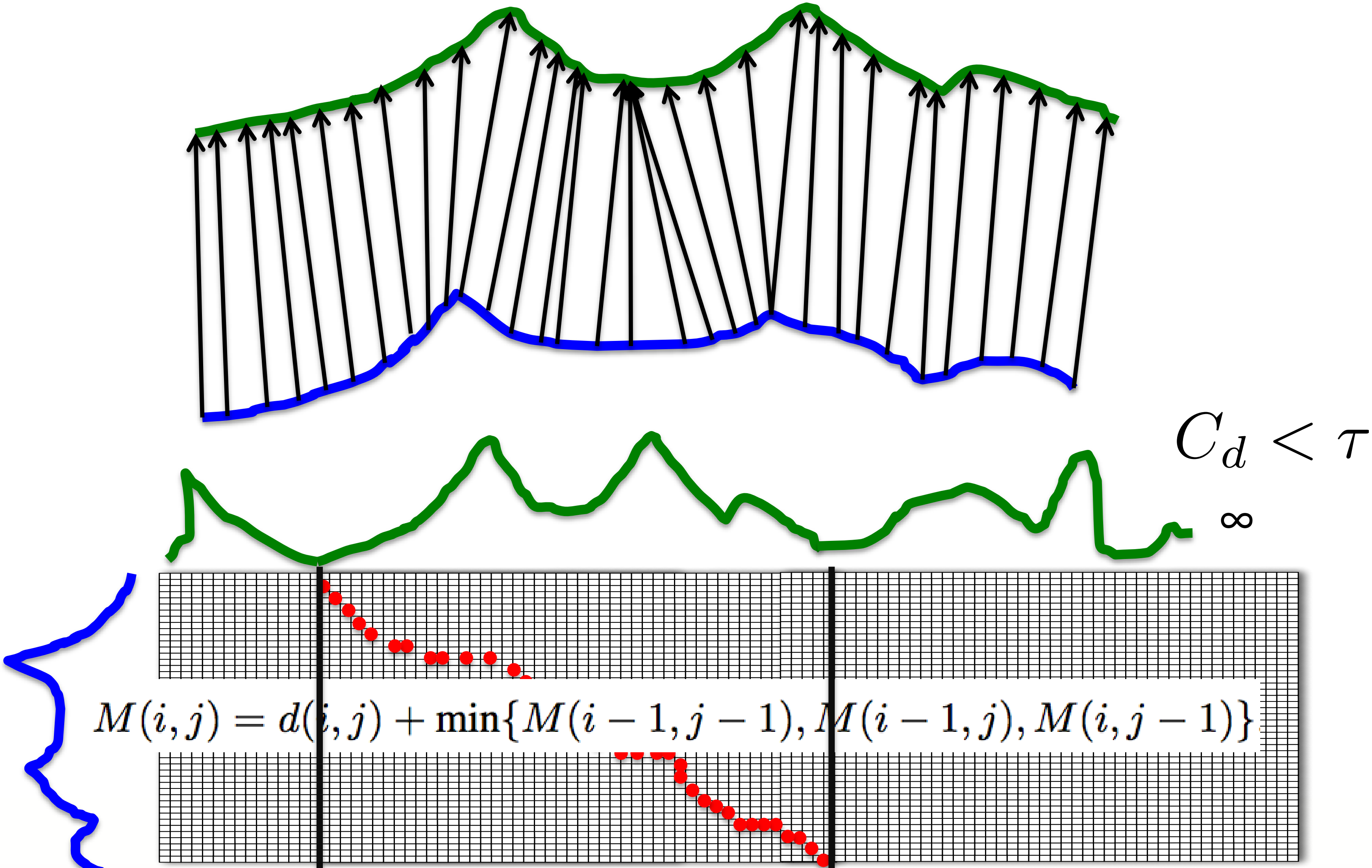


Multimodal Gesture Recognition on RGB-D data

- **Problem**:
  - **Continuous Gesture Recognition** in video sequences.
  - **Multimodal data**, RGB+D.

- **Approaches:**
  - Probabilistic Graphical Models.
  - **Dynamic Time Warping.**

Dynamic Time Warping



$$M(i,j) = d(i,j) + \min\{M(i-1,j-1), M(i-1,j), M(i,j-1)\}$$

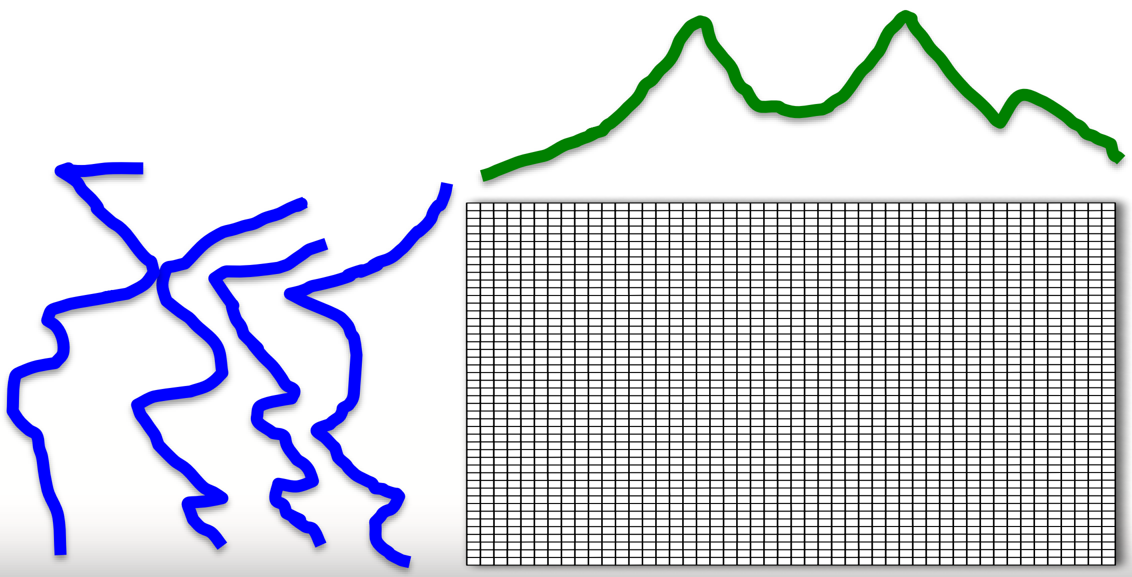
Probability-based DTW for gesture recognition on RGB-D data

Our **goal**:

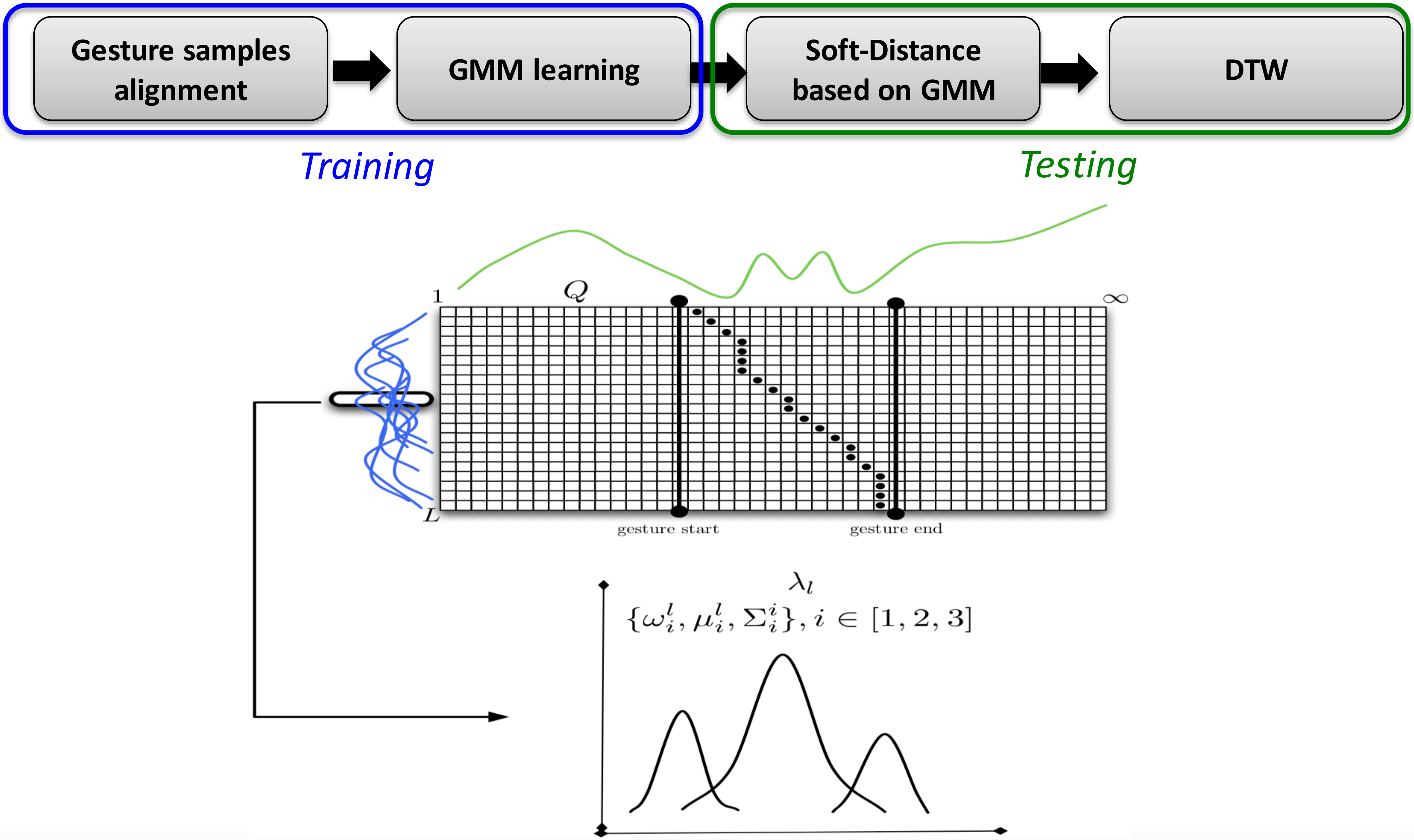
- **Improve the detection** by **encoding the variability** of a certain **gesture category** using RGB-D data.

Our **proposal**:

- Use DTW to **align gesture samples** in order to deal with temporal deformations.
- Use **Gaussian Mixture Models** to deal with pose deformations.
- Include a **soft-distance** based on **posterior probabilities** in the DTW algorithm.

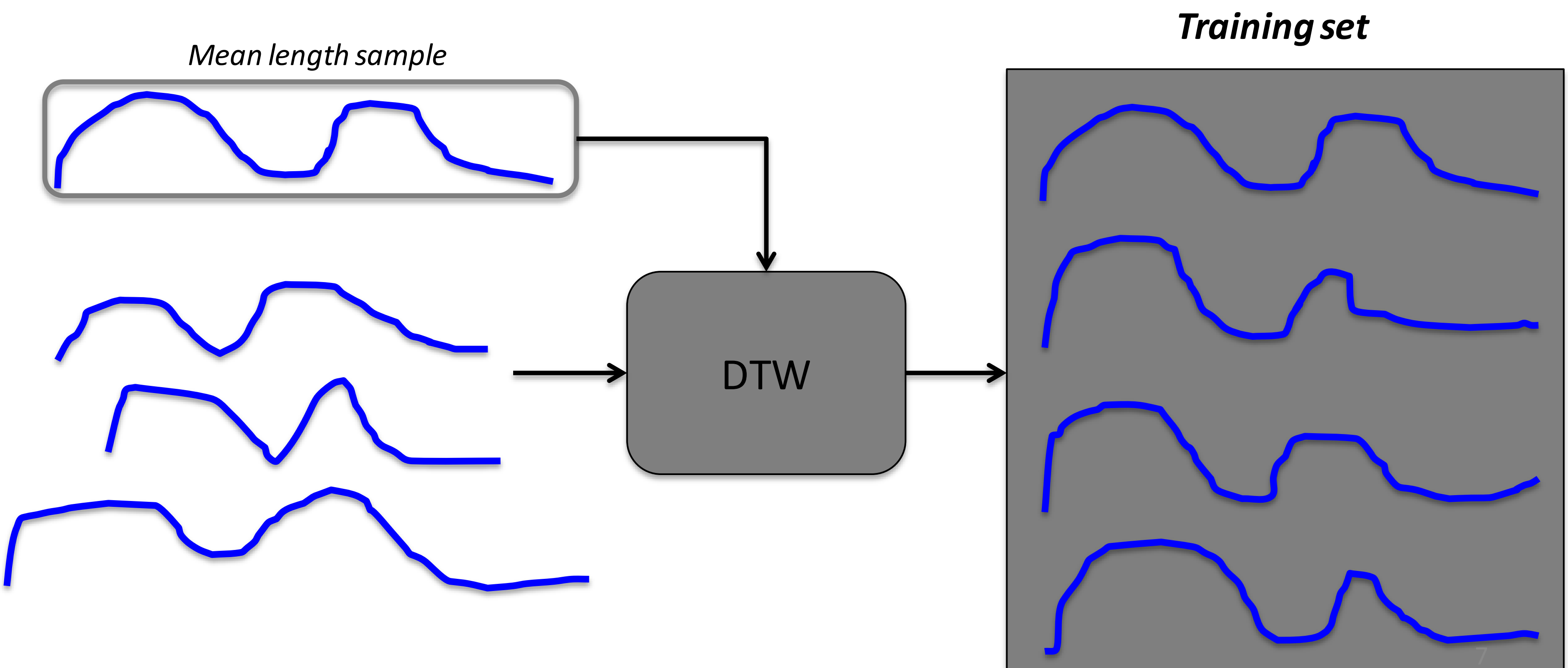


BoVDW for Gesture Recognition



Gesture samples alignment

- **Different samples** are used to model the **pattern gesture**.
- To deal with **temporal deformations** all **samples are aligned** with the mean length sample **using classic DTW**.

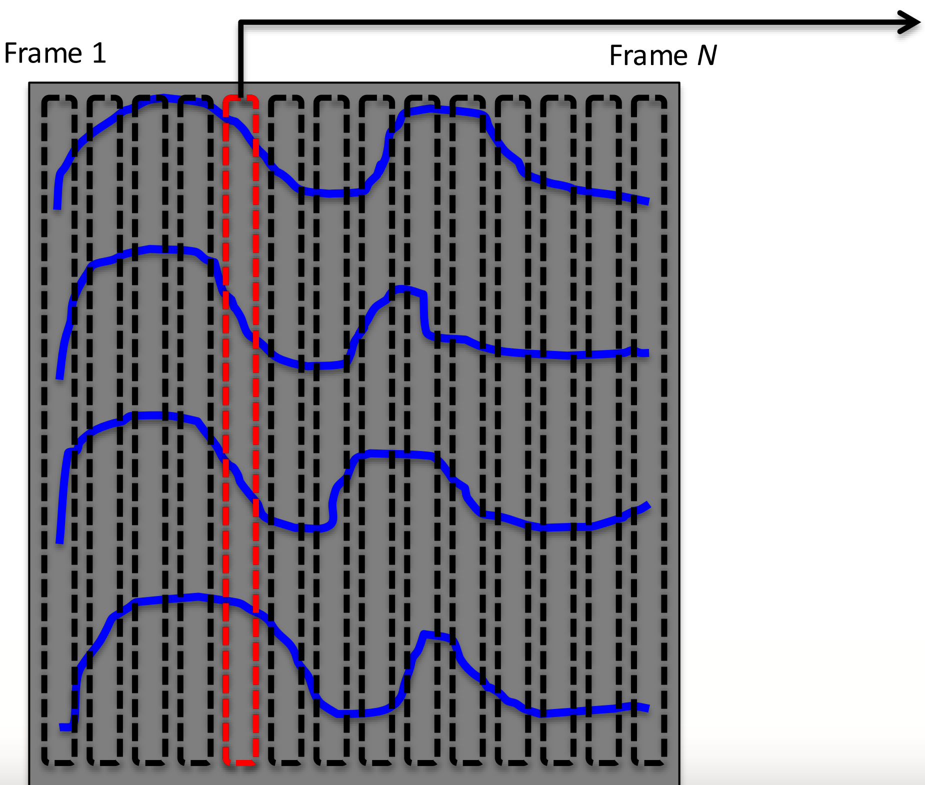


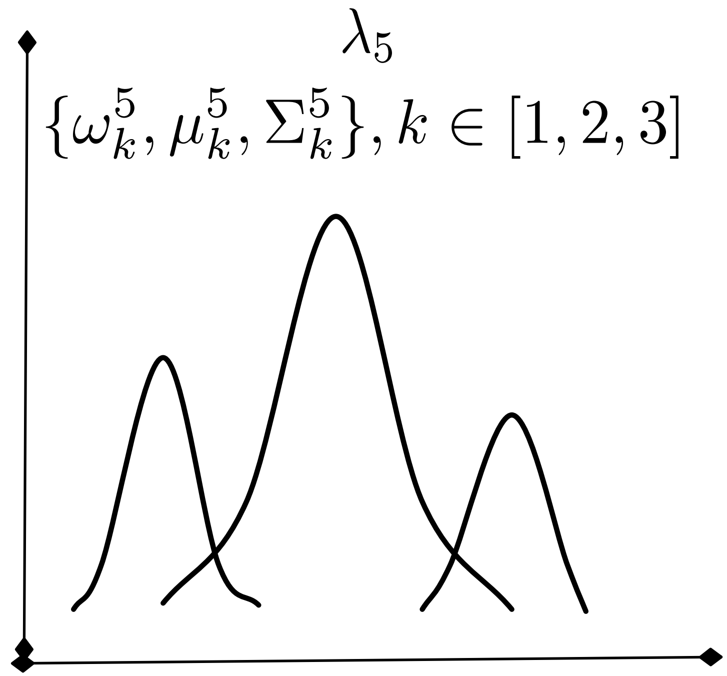
GMM learning

Gesture samples alignment → **GMM learning** → Soft-Distance based on GMM → DTW

- When the gesture samples are aligned we use a **Gaussian Mixture Model** to learn each set of elements overall sequences.

Training set

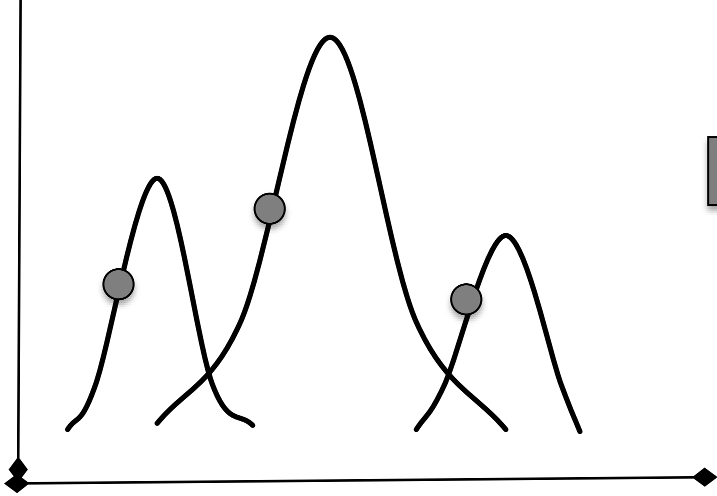




Soft-Distance based on GMM posterior probability

$$\lambda_l$$

$$\{\omega_k^l, \mu_k^l, \Sigma_k^l\}, k \in [1,2,3]$$



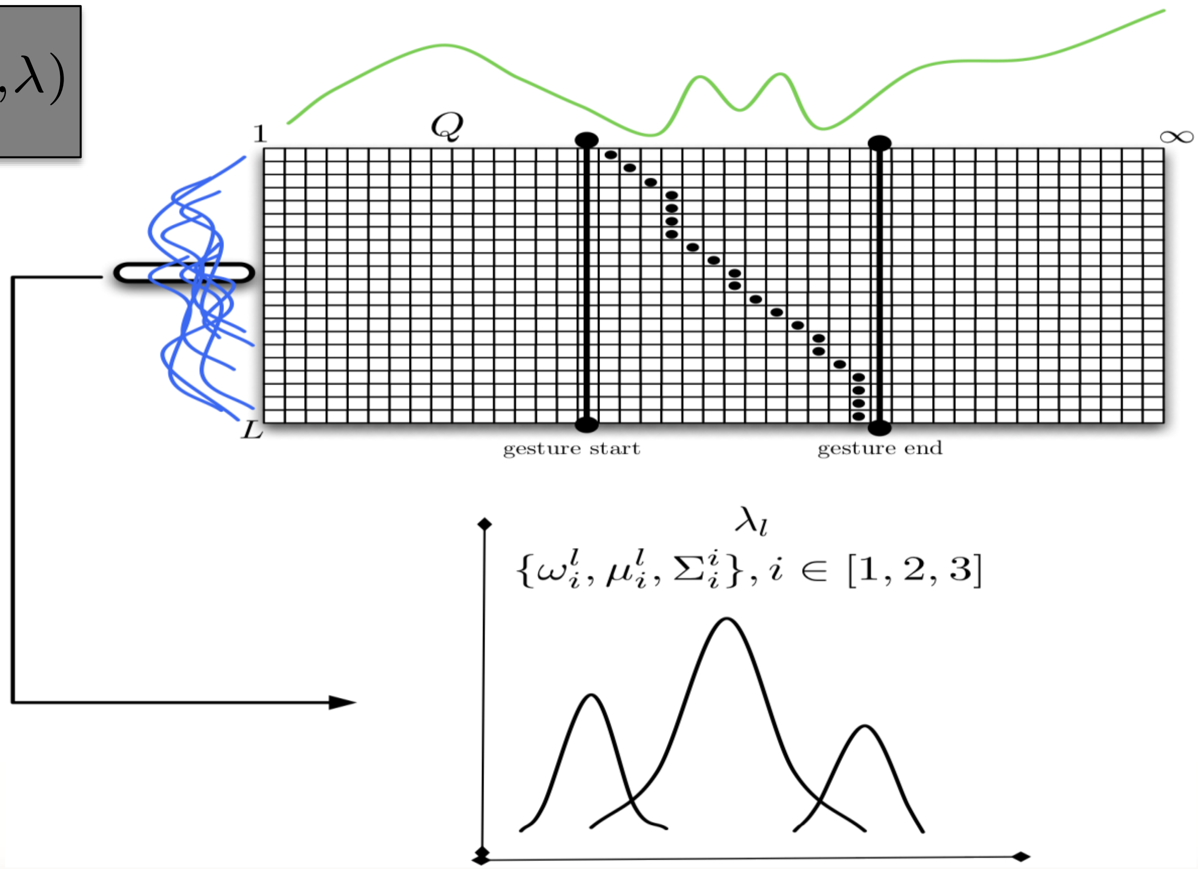
$$P(x,\lambda) = \sum_{k=1}^{M} \omega_k \cdot P_k(x)$$

$$P(x)_k = e^{-\frac{1}{2}(x-\mu_k)^T \cdot \Sigma_k^{-1} \cdot (x-\mu_k)}$$

$$\boxed{D(x,\lambda) = \exp^{-P(x,\lambda)}}$$

DTW

$$D(x, \lambda) = \exp^{-P(x,\lambda)}$$



Experimental Settings

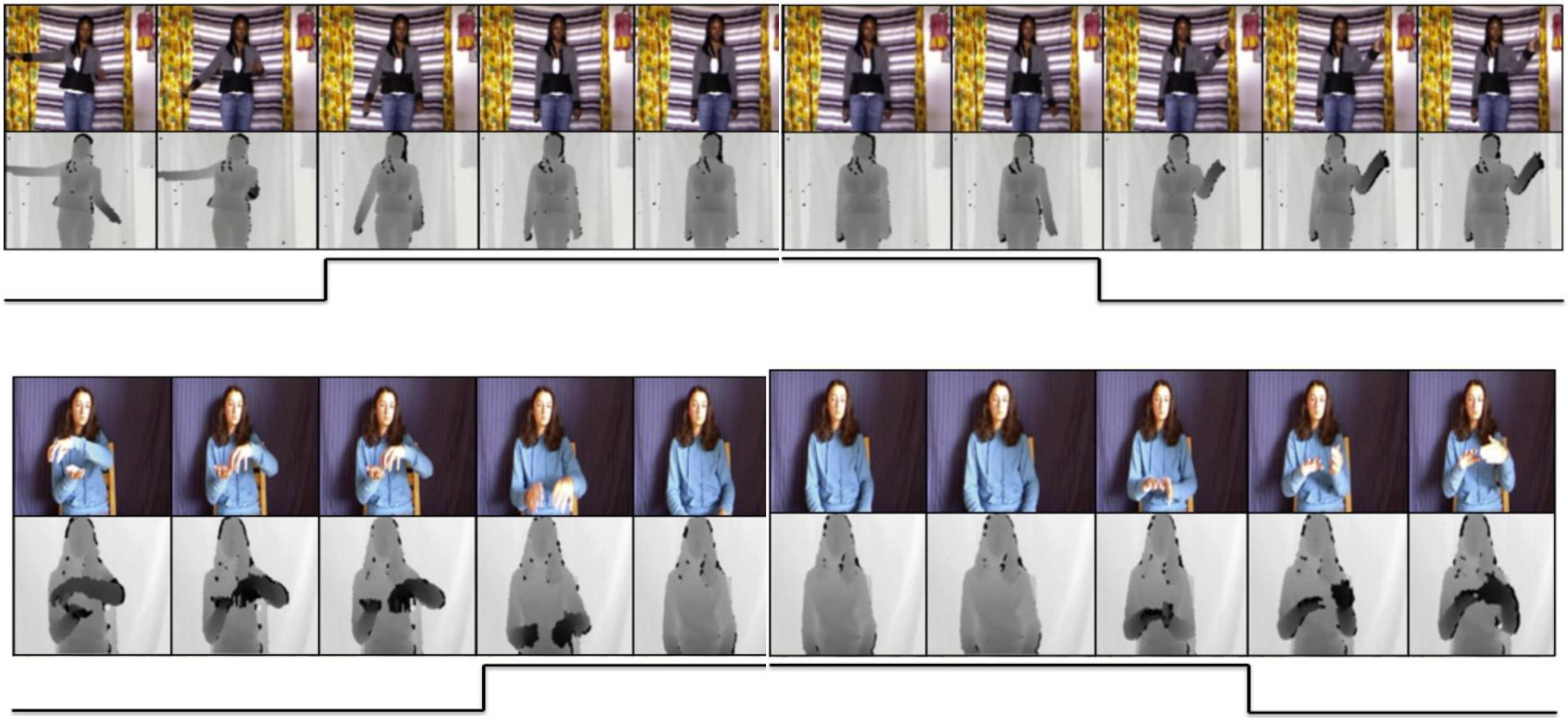
- **Data:**
  - **ChaLearn Dataset** (CVPR2012), in which, each video sequence shows an actor performing a set of gestures discriminated by an **Idle gesture** performed in between (more than 940 sequences).
  - Our goal is **to detect the Idle gesture** (more than 1000 samples available).
  - We defined a **10×10 grid approach** to extract **HOG+HOF** feature descriptors per cell.
  - We use **900 samples** of the gesture category in **a ten-fold validation procedure.**

- **Methods:**
  - Classic DTW with Euclidean distance.
  - Hidden Markov Model.
  - Probability-based DTW.

- **Evaluation:**
  - We obtain the **overlapping** metric (frame wise) and the **accuracy** metric of the number of gestures detected in each video sequences.

Qualitative results

- **Idle gesture detection** for two video sequences in the **ChaLearn Dataset**.



Quantitative results

- Results show how **the new approach outperforms classic DTW** and **HMM** by nearly **10% of overlapping**.
- When analyzing the accuracy, it can be seen that **the new approach easily detects more Idle gestures** than the classical approaches.
- Subtle differences found between **Euclidean DTW and HMM**.

	Overlap.	Acc.
Probability-based DTW	**39.08± 2.11**	**67.81±2.39**
Euclidean DTW	30.03±3.02	60.43± 3.21
HMM	28.51±4.32	53.28±5.19

Conclusions & Future work

- We proposed a **probability-based DTW** for **gesture recognition**.
- The **pattern model** is **learned from several samples** of the same gesture category using multimodal **RGBD data**.
- **Different sequences** were used to build a **Gaussian-based probabilistic model** of the gesture whose **possible deformations are implicitly encoded**.
- A **soft-distance based on the posterior probability** of the **GMM** was defined.
- The proposal is able to deal with **multiple deformations in data**, showing **performance improvements** compared to the **classical DTW** and **HMM** approaches.

Probability-based dynamic time warping for gesture recognition on RGB-D data

Thank you!

Miguel Ángel Bautista[1,2]
Antonio Hernández-Vela[1,2]
Xavier Perez-Sala[2,3]
Victor Ponce Lopez[1,2]
Xavier Baro[2,4]
Oriol Pujol[1,2]
Cecilio Angulo[3]
Sergio Escalera[1,2]

[1]*Dept. Applied Mathematics and Analysis, Universitat de Barcelona*
[2]*Computer Vision Center, Universitat Autónoma de Barcelona*
[3]*CETpD, Universitat Politècnica de Catalunya*
[4]*EIMT, Universitat Oberta de Catalunya*