Institut de Robòtica i Informàtica Industrial

ICT - 287617

SEVENTH FRAMEWORK PROGRAMME

CSIC

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Multimodal Object Recognition using Random Clustering Trees

M. Villamizar mvillami@iri.upc.edu

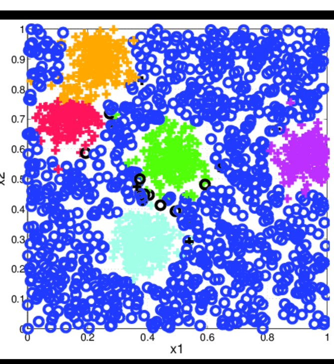
A. Garrell agarrell@iri.upc.edu

A. Sanfeliu sanfeliu@iri.upc.edu

F. Moreno-Noguer fmoreno@iri.upc.edu

Goal

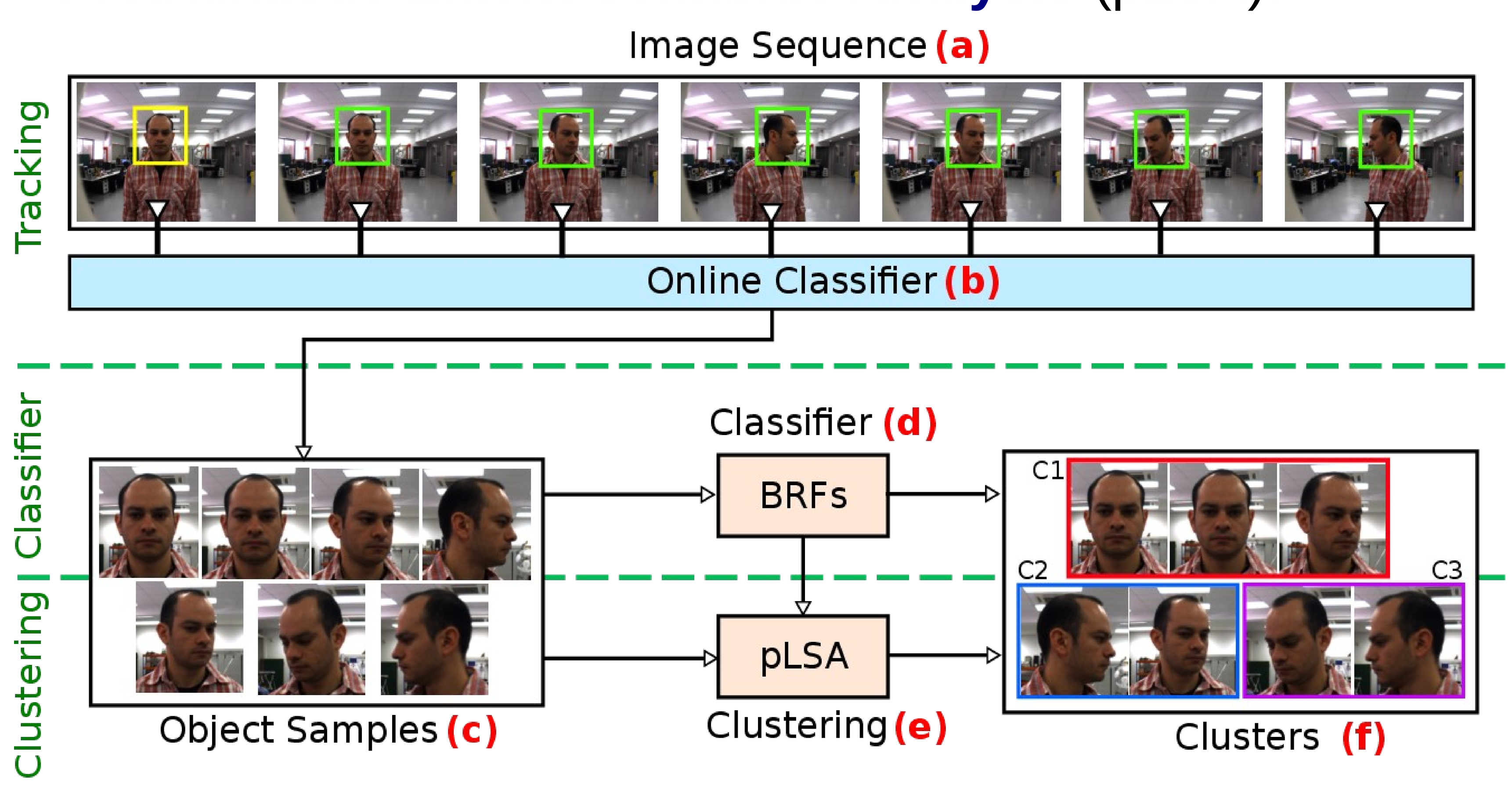
Efficient and **discriminative** approach to recognize objects and to discover automatically multiple **intra-class modalities** (object's views).



Proposed Approach

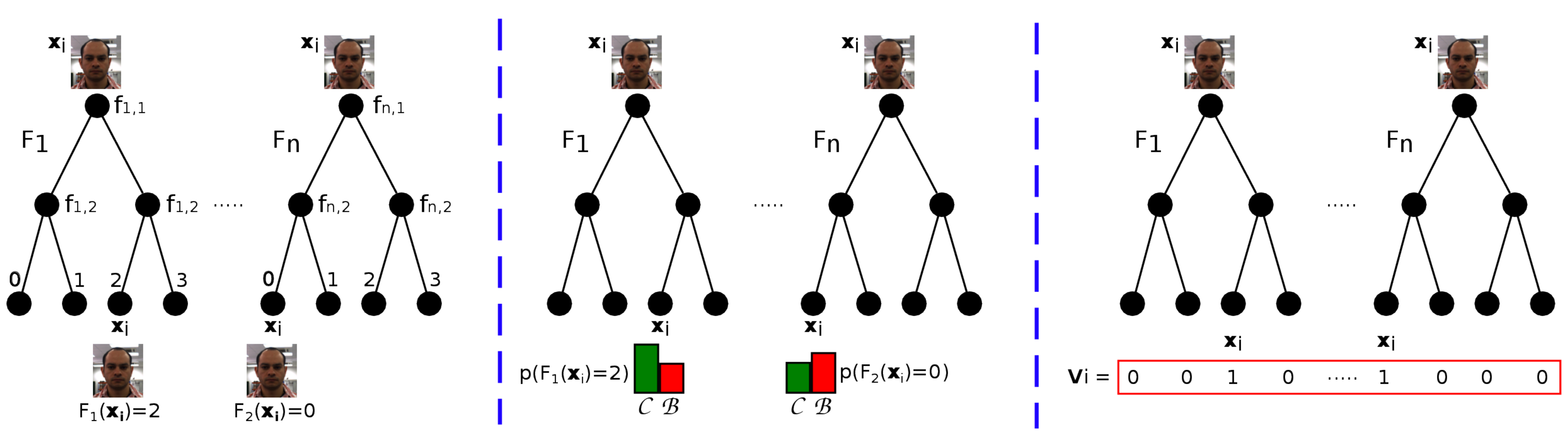
Combined approach to learn objects with multiple modalities:

- **Tracking step: online** classifier to **track** the object during a video sequence and to acquire the **training data** (object and background samples).
- **Classifier step**: computation of a discriminative and robust tree-structured classifier, **Boosted Random Ferns** (BRFs), using the acquired training data in the tracking step.
- **Clustering step**: **clustering** of the classifier output using **Probabilistic Latent Semantic Analysis** (pLSA).



Object Tracking

- Object **tracking** to extract **automatically** a set of training samples which are used later to compute the object classifier.
- Extremely randomized trees (**random ferns**) are used to track **online** the object during the video sequence.
- Online random ferns [1] are computed/updated **incrementally** using its own detection hypotheses in images (**self-learning**).



Object Classifier

- **Efficient** and **discriminative** object classifier based on the **boosted** combination of **random ferns** [2].
- Each fern is a set of **binary features** (pixel comparisons) computed at specific image location.
- The most discriminative **ferns** are chosen via **AdaBoost**.

$$H(\mathbf{x}) = \sum_{t=1}^{T} h_t(\mathbf{x}) > \beta \qquad h_t(\mathbf{x}) = \frac{1}{2}\log\frac{p(F_t(\mathbf{x}) = r|\mathcal{C}) + \epsilon}{p(F_t(\mathbf{x}) = r|\mathcal{B}) + \epsilon}$$

References

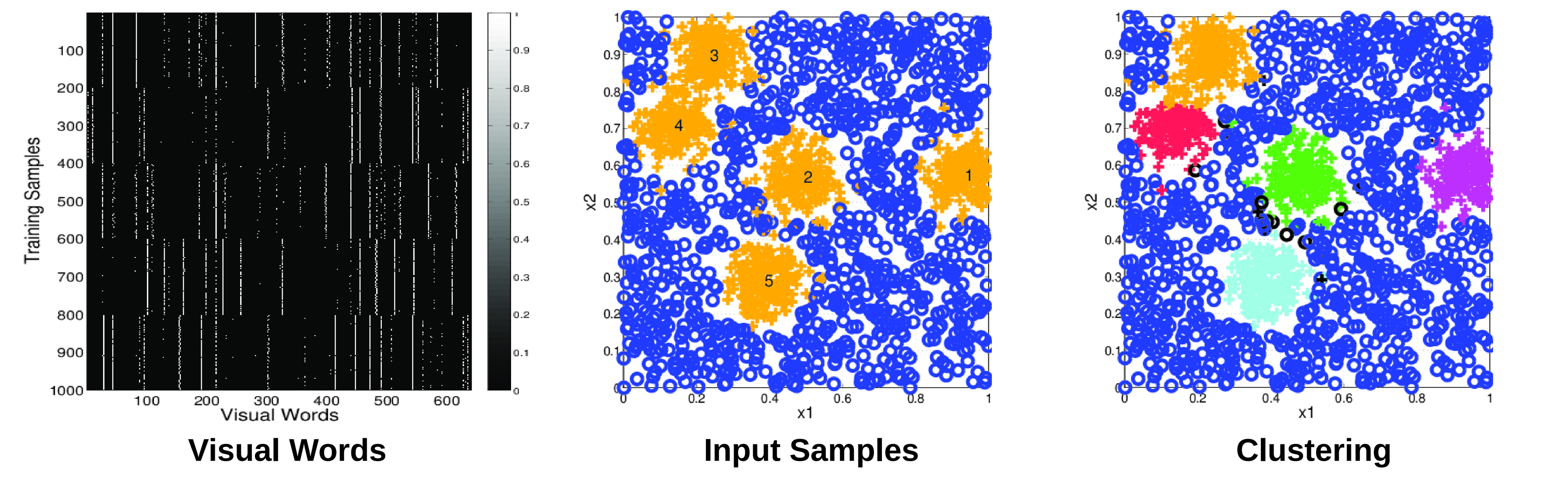
[1] M. Villamizar, A. Garrell, A. Sanfeliu, and F. Moreno-Noguer. Online human-assisted learning using random ferns. In ICPR, 2012.

[2] M. Villamizar, J. Andrade-Cetto, A. Sanfeliu, and F. Moreno-Noguer. Bootstrapping boosted random ferns for discriminative and eficient object classication. Pattern Recognition, 2012.

[3] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 2001.

Clustering

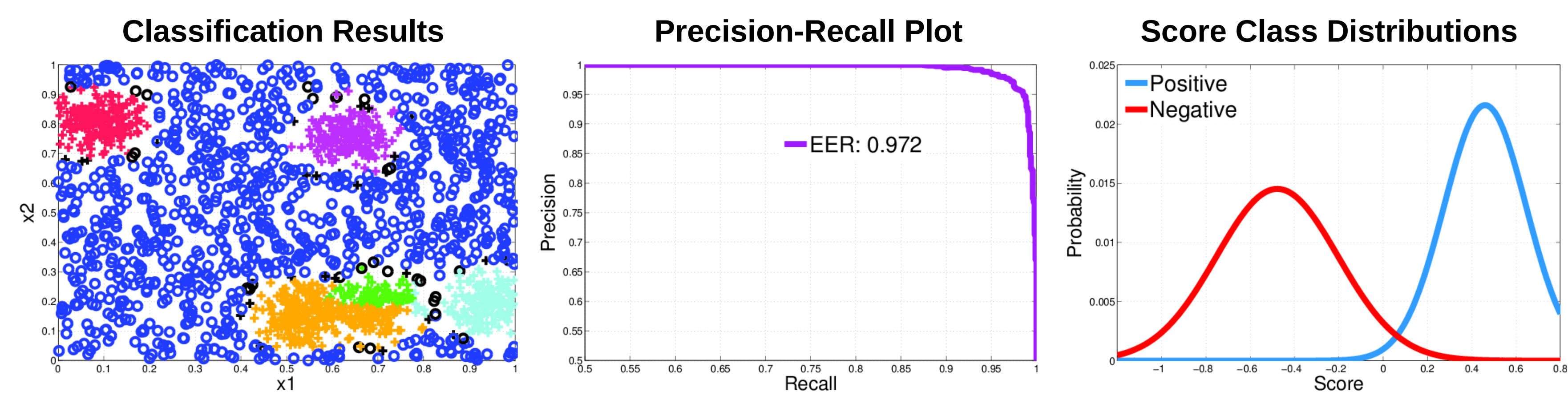
- **Probabilistic Latent Semantic Analysis** [3] are used to discover **automatically** intra-class object modalities (**latent** variable). The overall object **appearance** is discretized in multiple sample **clusters** with strong feature similarity.
- Tree-structured **visual words**, obtained from the **output** of the **BRFs** classifier, are used to perform the **clustering**.



Synthetic Experiments

2D classification problem involving two **complex** classes.

- Classification results of BRFs for multimodal distributions:



- Average classification results of BRFs and RFs classifiers:

Classification Performance														
	RFs			BRFs										
	# Clusters (K)			# Clusters (K)			# Ferns (R)				# Features (S)			
	3	5	10	3	5	10	5	10	20	50	1	3	5	7
EER(%)	90.6	92.4	84.4	96.9	97.4	96.0	95.4	96.6	96.9	97.2	83.1	96.1	96.9	97.2
Distance(%)	59.0	68.8	46.5	83.1	90.6	73.2	77.2	80.4	83.1	88.8	41.2	73.0	83.1	88.4

- Average confusion values in the clustering labels:

Clustering Results												
	BRFs+pLSA				K-means (Euclidean)				BRFs+K-means			
D	2	5	10	20	2	5	10	20	2	5	10	20
K=3	0.097	0.001	0.033	0.000	0.147	0.100	0.133	0.067	0.177	0.036	0.086	0.113
K=5	0.240	0.022	0.019	0.020	0.180	0.139	0.163	0.201	0.304	0.127	0.114	0.173
K=10	0.514	0.144	0.092	0.096	0.367	0.143	0.159	0.116	0.548	0.251	0.207	0.102

Real Experiments

Detection and pose estimation of faces and 3D objects.