Corso di Dottorato

"Sistemi e Servizi Cognitivi per l'Intelligenza di Ambiente e le Telecomunicazioni"

Learning to classify the visual dynamics of a scene

Nicoletta Noceti

Dipartimento di Informatica e Scienze dell'Informazione

Università degli Studi di Genova

http://slipguru.disi.unige.it

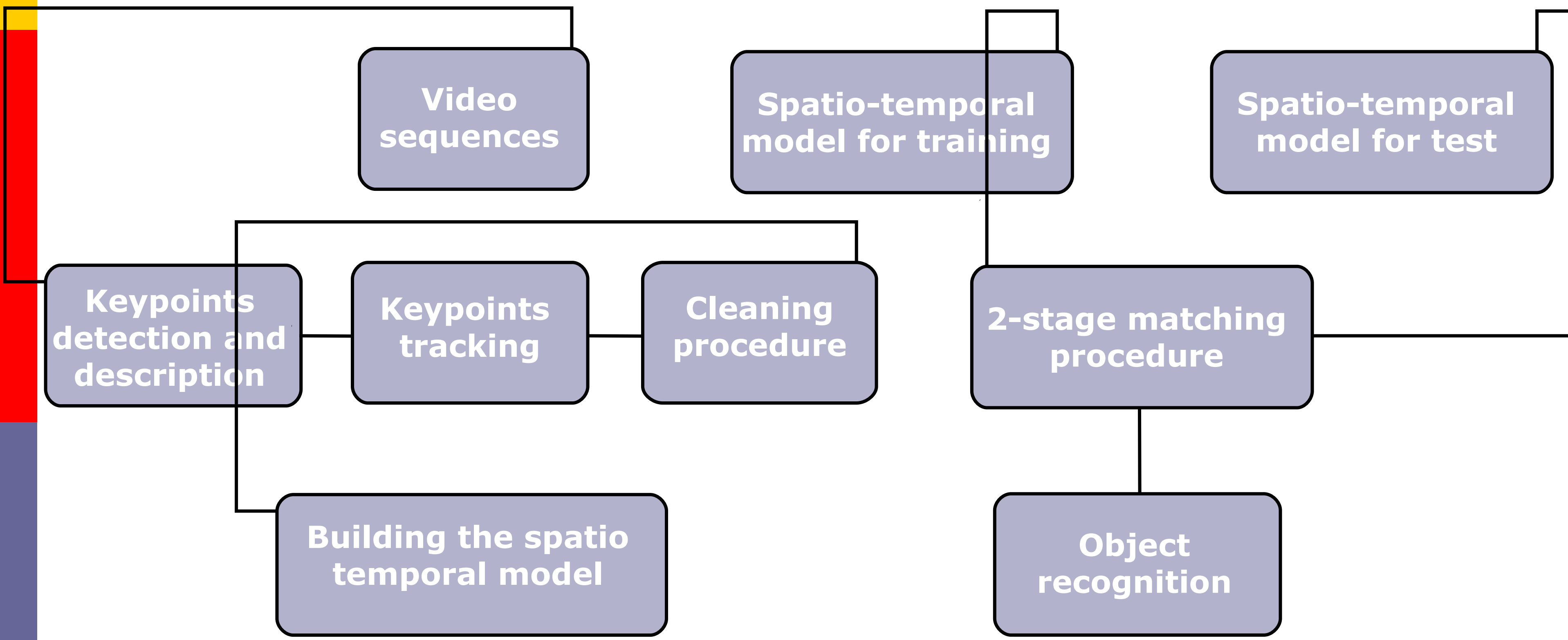
Outline of the presentation

- From past…
  - our 3D object recognition system
  - a demo
- …to future
  - Research proposal
  - Scenario and aims
  - Problem statement

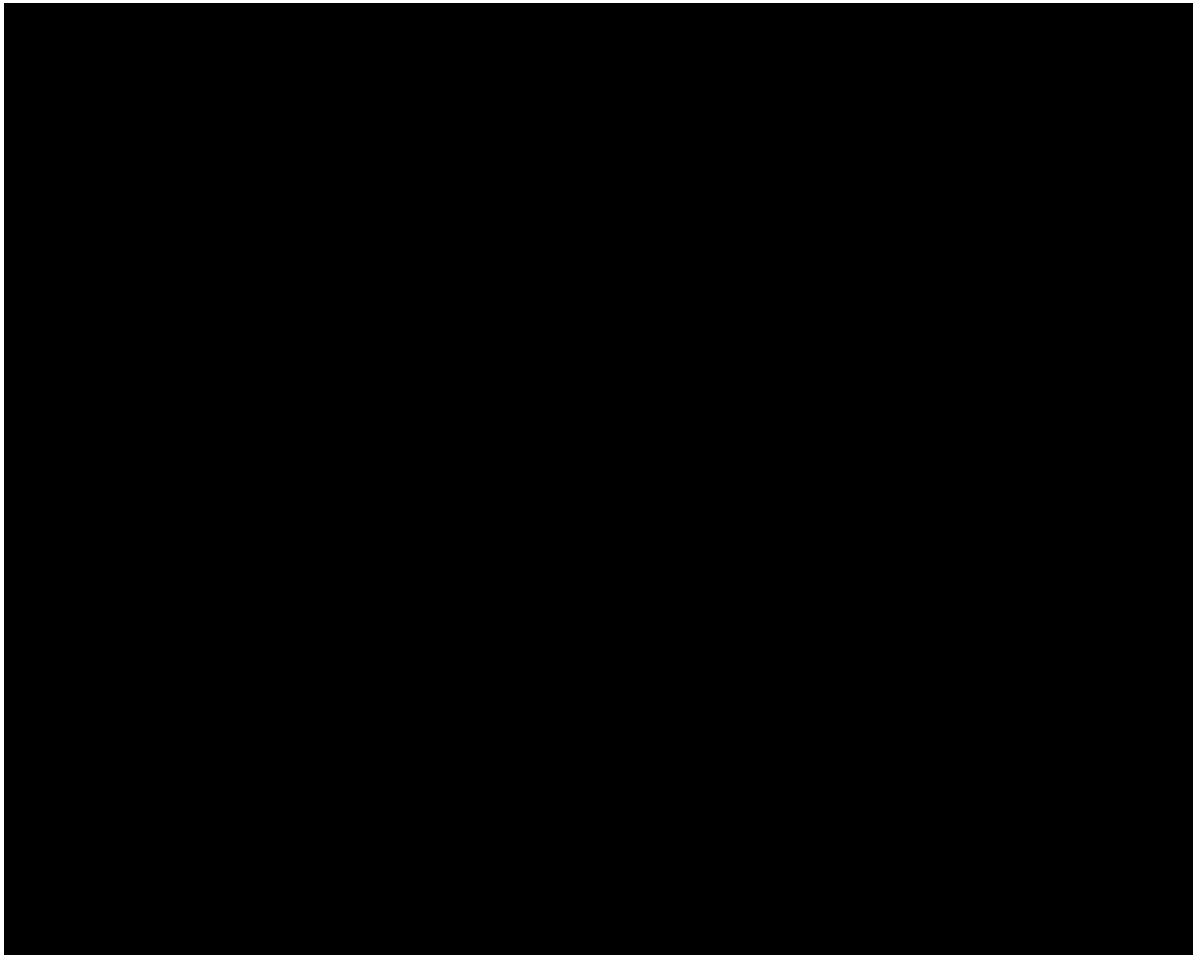
3D Object Recognition

- We observe an object from slightly different viewpoints and exploit local features distinctive in space and stable in time to perform recognition
  - Obtain a 3D object recognition method based on a **compact description of image sequences**
  - Exploit temporal continuity and spatial information **both on training and test**

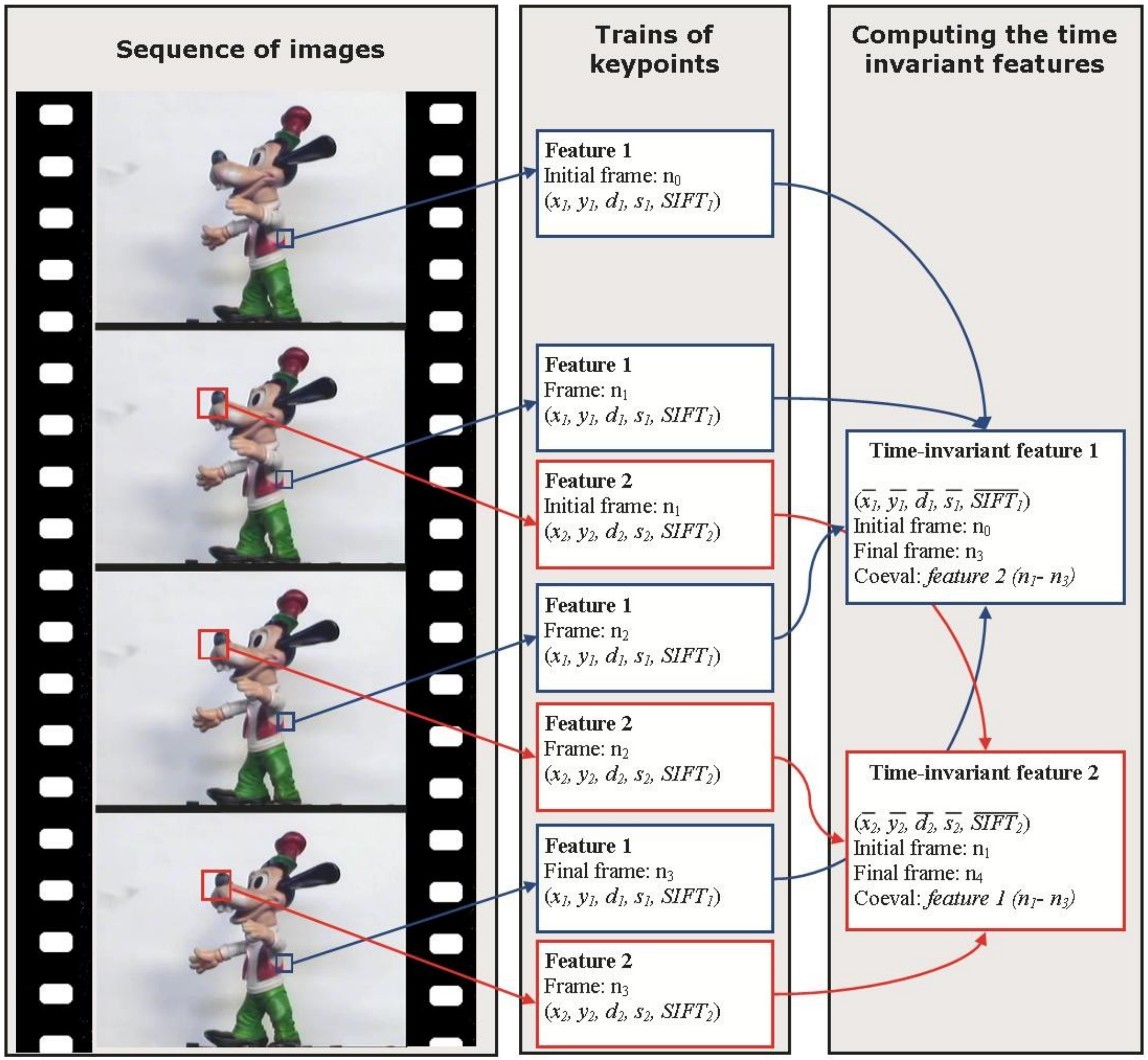
Recognizing objects with ST models



From sequence to spatio-temporal model



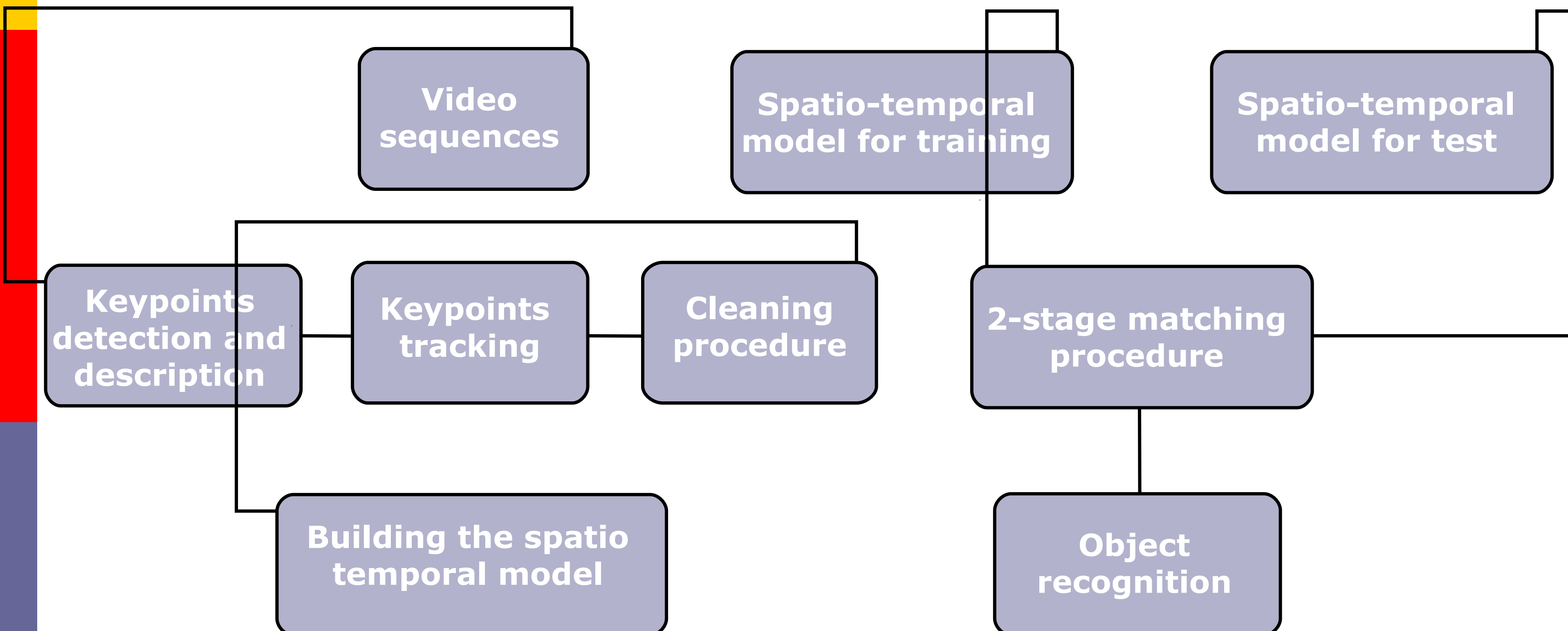
From sequence to spatio-temporal model



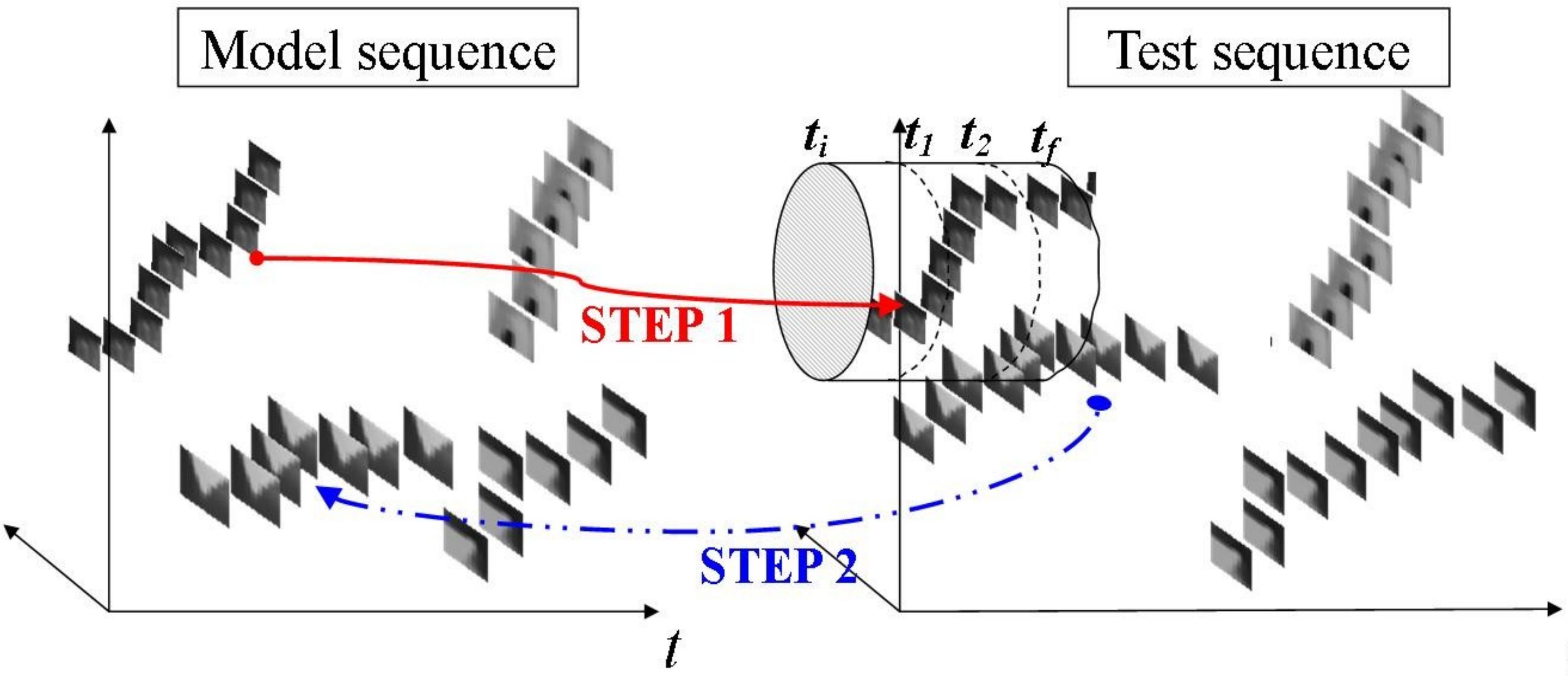
Time invariant feature

- We obtain a set of *time-invariant features*:
  - a spatial appearance descriptor, that is the average of all SIFT vectors of its trajectory
  - a temporal descriptor, that contains information on when the feature first appeared in the sequence and on when it was last observed

Recognizing objects with ST models



Matching of sequence models



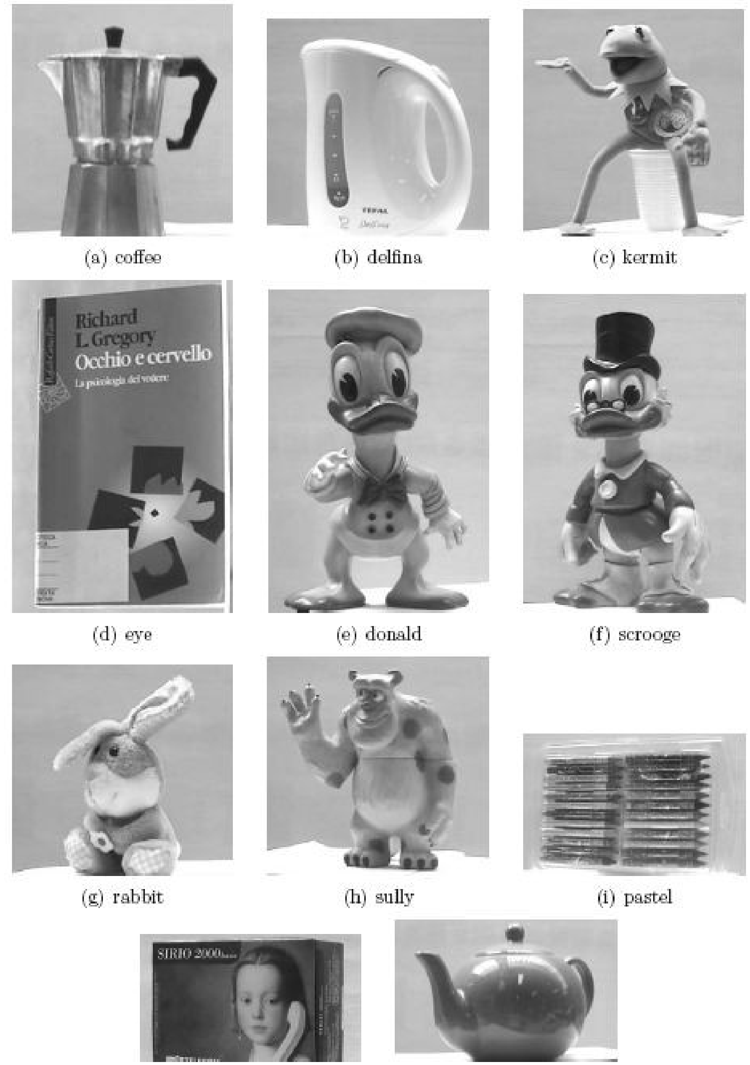
Experiments and results

- Matching assessment
  - Illumination, scale and background changes
  - Changes in motion
  - Increasing the number of objects
- Object recognition on a 20 objects dataset
- Recognition on a video streaming

E. Delponte, N. Noceti, F. Odone and A. Verri
Spatio temporal constraints for matching view-based descriptions of 3D objects
In WIAMIS 2007

3D objects

(a) bambi (b) box (c) dewey (d) biscuit
(e) bookGeo (f) bookSvm (g) dino
(h) teddy (i) pino (j) telephone
(k) goofy (l) tommy (m) winnie



(a) coffee (b) delfina (c) kermit
(d) eye (e) donald (f) scrooge
(g) rabbit (h) sully (i) pastel
(j) easyBox (k) teapot

Recognizing 20 objects

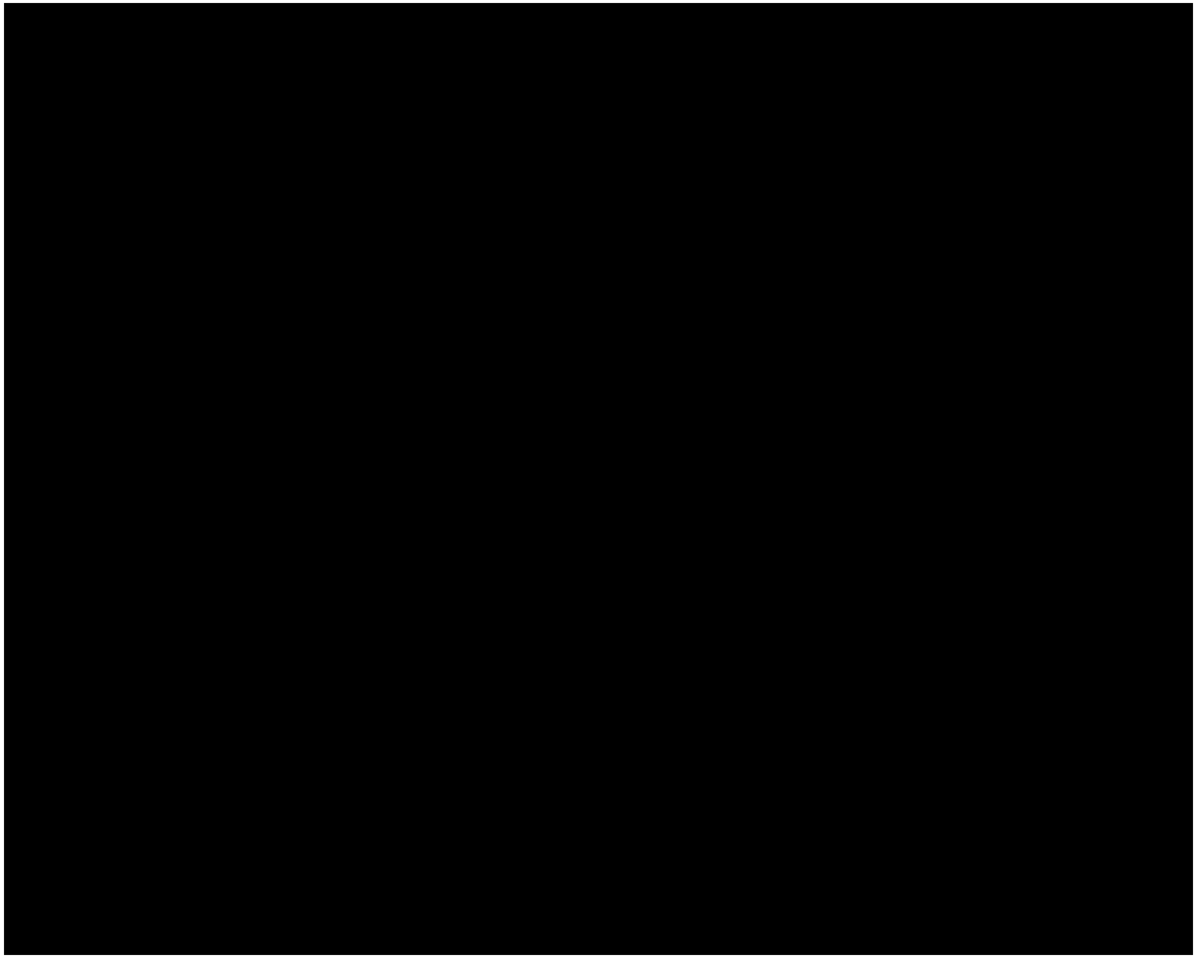
Number of experiments: 840

TP=51	FN=13
FP=11	TN=765

$$\text{RECALL} = \frac{TP}{TP + FN} = 80\%$$

$$\text{PRECISION} = \frac{TP}{TP + FP} = 82\%$$

Recognition on a video stream



But my research proposal is…

"Learning to classify the visual dynamics of a scene"

- Idea: to combine classical computer vision techniques and learning approaches to understand and classify dynamic events
  - **Modeling of common behaviours**
  - **Anomaly detection**

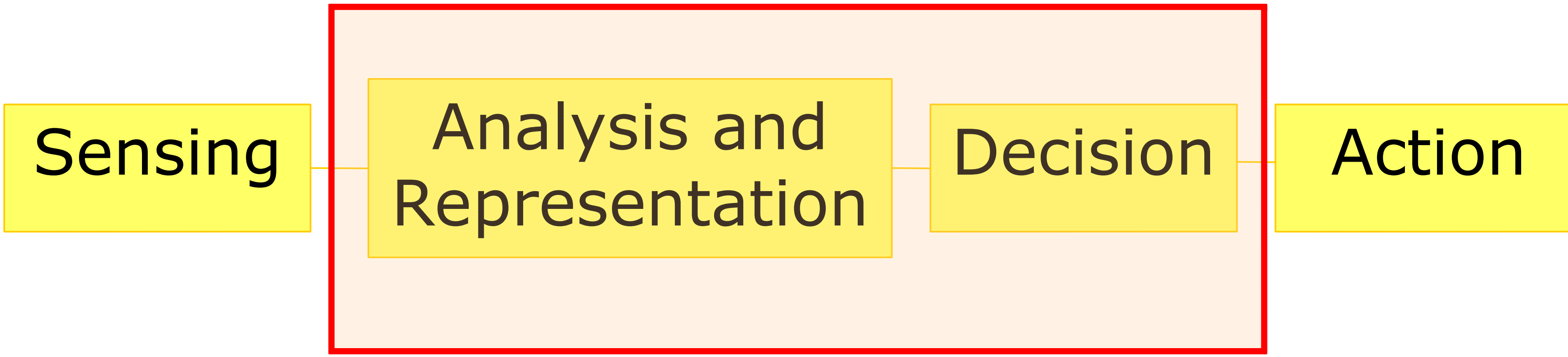
State of the art

- In the video surveillance framework there is a growing need for *adaptive systems*, able to learn behaviour models by long time observations

- In the last decades it has been accepted that many computer vision application are better dealt with a learning from example approach

- Focusing on video description, there are some promising works but the research has still many open issues

The cognitive cycle



- Description of video content
- Event classification to decide what is an anomaly event and how it is described
  - Now decision made by humans → automation

Analysis and representation

- The focus of the first part of our work will be on **video processing**
  - to study robust spatio-temporal features obtaining a reliable video content description
  - Low-level blob description exploring more features (shape, color, texture) than the ones usually used (position, area, perimeter)
  - To look for a balance between computational complexity (real time needed…) and efficency

Analysis and representation

- Why do we need a robust blob description?
  - Blobs will be tracked but there are some problems to deal with:
    - Illumination changes
    - Velocity variations
    - Occlusion
    - Trajectories intersection
    - Features local nature
- A reliable blob description allows to obtain a robust tracker

From representation to decision

- Blobs trajectories built by tracking are the starting point of the classification step

- Idea: to integrate motion analysis with statistical learning techniques to exploit the knowledge coming from *previously seen* scenario

  - **Unsupervised learning**
  - **Manifold learning**

Learning techniques

- Unsupervised learning
  - Method of machine learning where a models is fit to observations. It is disinguished from supervised learning from the fact that there is no *a priori* output

- Manifold learning
  - High dimensional data can be difficult to interpret. One approach to semplifications is to assume that the data of interest lies on an embedded non-linear manifold within the higher dimensional space

From representation to decision

- …but our representation is not suitable for a learning framework…
  - At this computation point, an event is related to one (or more) blob trajectory
- Two possible solutions:
  - Appropriate handling of the description
  - Design of appropriate similarity functions

Case studies

- Today: medium distance video of indoor scenes

- Long term objective: wide area monitoring
  - Analysis of complex crowded scenes (train stations, airports)
  - From blob tracking to the study of the whole scene motion (optical flow based)

Collaborations

- **Imavis**: IMAge and VISion
  - Development and software consulting company with headquarters in Bologna and a reserach and development office in Genova

    www.imavis.com

- **SINTESIS project**: Sistema INTegrato per la Sicurezza ad Intelligenza diStribuita
  - DIBE, DIST, DISI, XXX altro?

Thanks for your attention!