



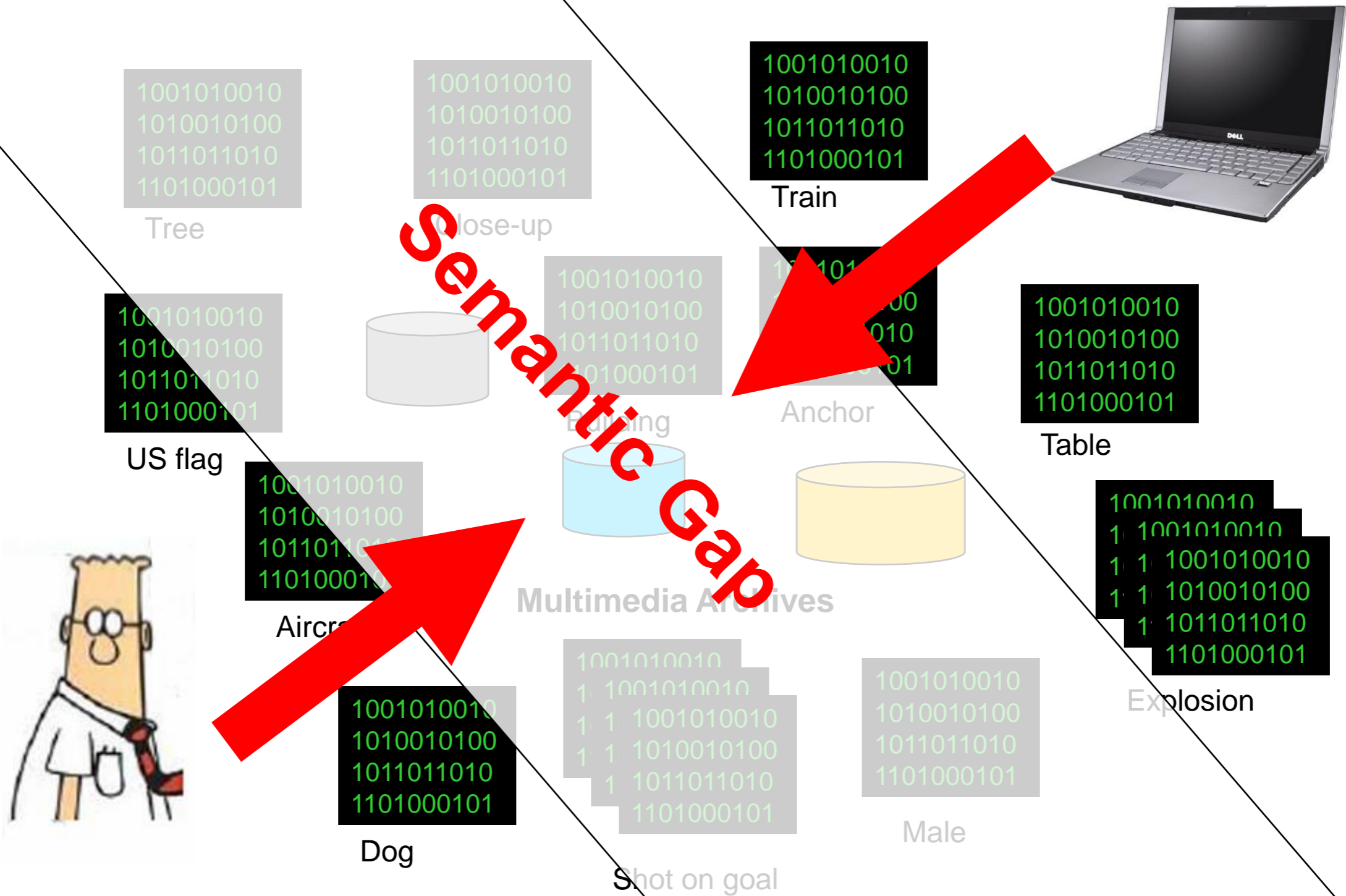
Learning rules for semantic video event annotation

Marco Bertini, Alberto Del Bimbo, **Giuseppe Serra**
University of Florence
{bertini, delbimbo, serra} @ dsi.unifi.it

International Conference on Visual Information Systems
September 11-12, 2008 – Salerno, Italy



Semantic Gap





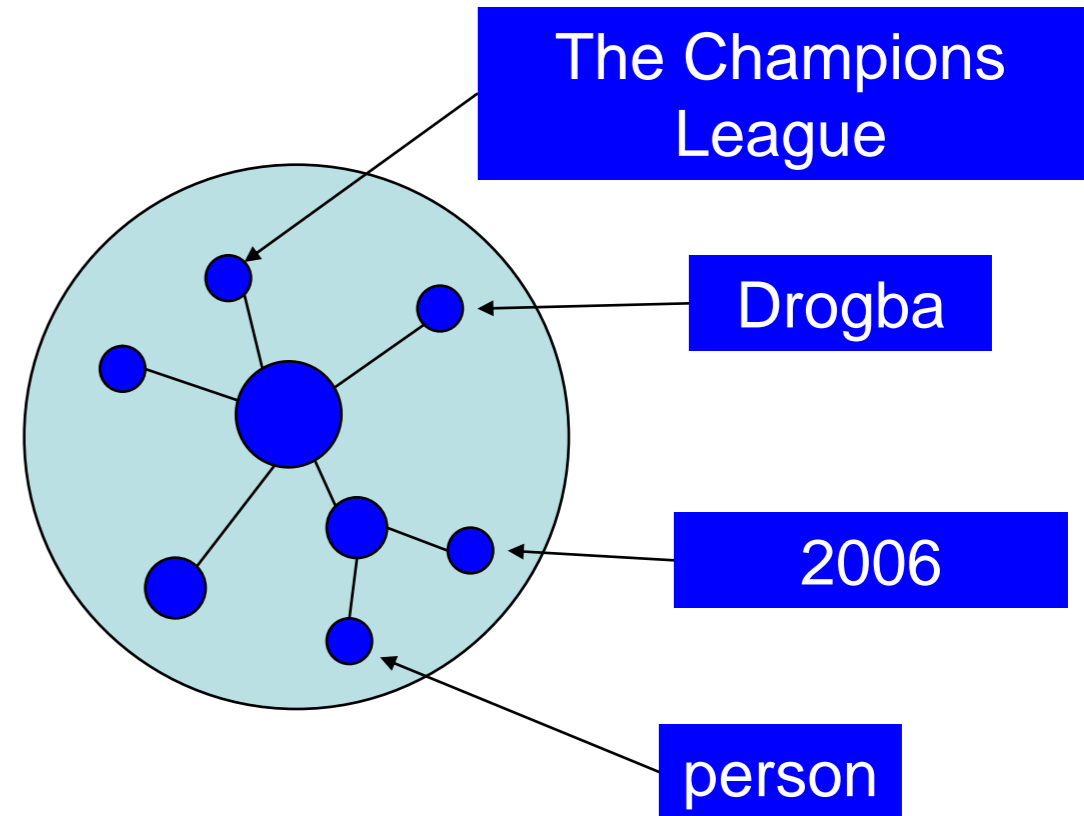
Ontology

Ontologies are formal, explicit specifications of a domain knowledge: they consist of concepts, concept properties and relationships between concepts.

Ontologies have been developed with the aim of proving a common vocabulary that encodes levels of semantics to overcome semantic heterogeneity for information.

For example:

- Content semantics (general knowledge)
playfield, players, crowd
- Content semantics (private knowledge)
Drogba
- Retrieval semantics:
“good” shot on goal
- Situation semantics:
2006
- Social semantics:
The Champions League





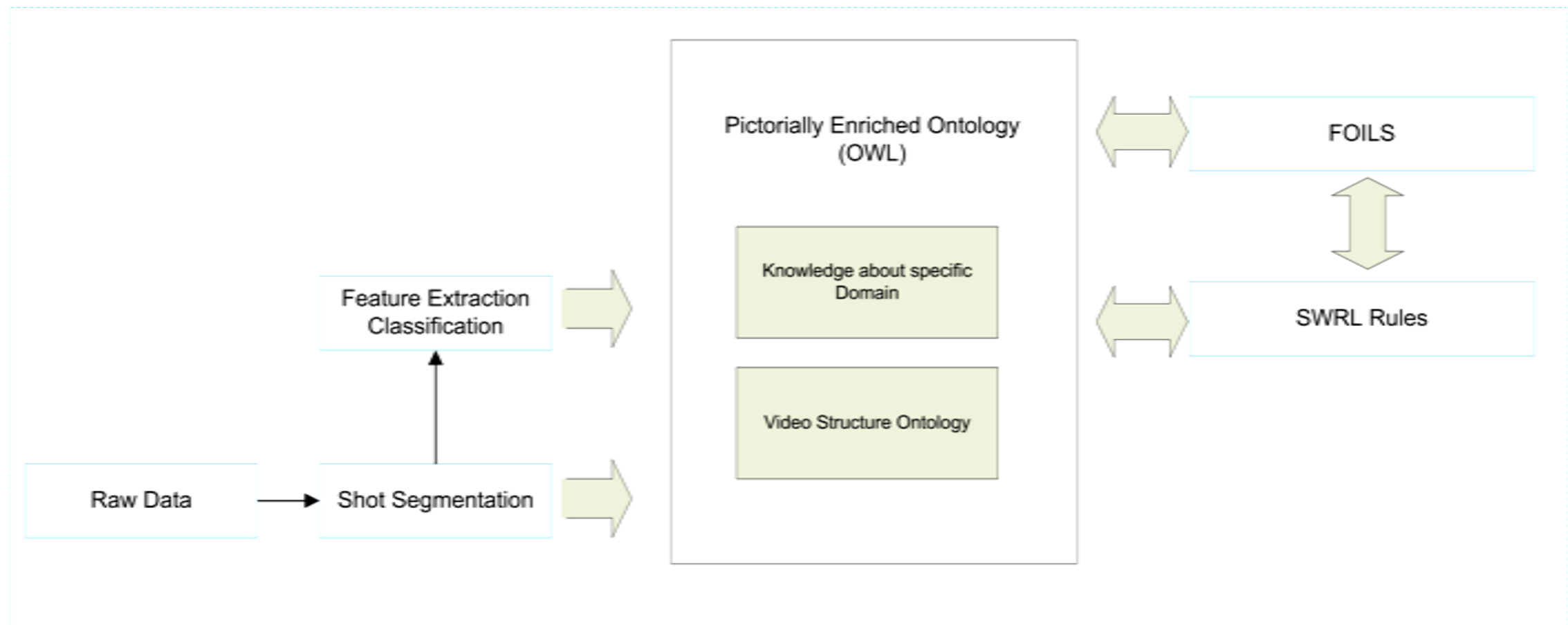
Ontologies

- Several researcher projects have addressed the problem of using ontologies for semantic annotation and retrieval by content from audio-visual digital libraries.
- Several approaches have been proposed:
 - System where the ontology provides the conceptual view of the domain at the schema level and appropriate classified play the role of entities detectors.
 - System that includes in the ontology an explicit representation of the visual knowledge to perform reasoning not only at the schema level but also at the data level.
- Recently for event recognition several authors have exploited the ontology schema using temporal reasoning based on rules over object and events.
 - These methods defined rules, used to describe events, that were created by human experts; thus, these approaches are not practical for the definition of a large set of actions.
- To overcome this problem some researcher have studied techniques to learn automatically a set of rules. None of these is based on ontologies.



System architecture

- We propose a framework for video event annotation that exploits the Pictorially Enriched Ontology model (OWL), that includes concepts and their visual descriptors, and a method to learn set of first-order logic rules that describe events defined in the ontology
- The learned rules, defined using the SWRL, are applicable directly to an ontology defined using the OWL
- The proposed learning method is First Order Inductive Learner for Semantic Web (FOILS)
 - Based on the FOIL technique defined by Quinlan





FOILS Algorithm

- At the beginning the algorithm **starts** with the head that we want to find in the rule and an initial body.
- The algorithm **iterates** searching the new literals that have to be added to the body of the rule.
- The algorithm **finishes** when negative are all excluded no more negative examples are excluded for a certain number of loops.
- Two important issues have to be addressed: the **generation of hypothesis candidates** and the choice of the **most promising candidate**.

Algorithm 1 FOILS algorithm schematization

```
Pos ← Positive examples
Neg ← Negative examples
Rule ← Initial rule
repeat
  Candidate_literals ← Generating hypothesis candidates
  Best_literal ← arg maxL Rule_Gain(L,Rule)
  Add Best_literal to Rule preconditions
  Pos ← subset of Positive examples that satisfy Rule
  Neg ← subset of Negative examples that does not satisfy Rule
until Neg is empty or no more Neg examples are excluded
for l loops
```



FOILS Algorithm

- **Generation of hypothesis candidates**

Suppose that the current rule being considered is

$$(L_1 \wedge L_2 \dots \wedge L_n) \rightarrow P(x_1, x_2, \dots, x_k)$$

The generation candidates follow the next rules:

- $Q(v_1, \dots, v_r)$ where Q is any class and properties of the ontology and where the v_i are either a new variable or already present in the rule. At least one of the v_i in the created literal must already exist as a variable in the rule.
- $Equal(x_j, x_k)$ where x_j and x_k are variables already present in the rule.
- **Examples:** $Airplane(?p) \rightarrow IsTakingOff(?p)$
 - candidates $IsOnGround(?p, ?g1)$ or $IsOnSky(?p, ?g1)$
- **Most promising candidate**

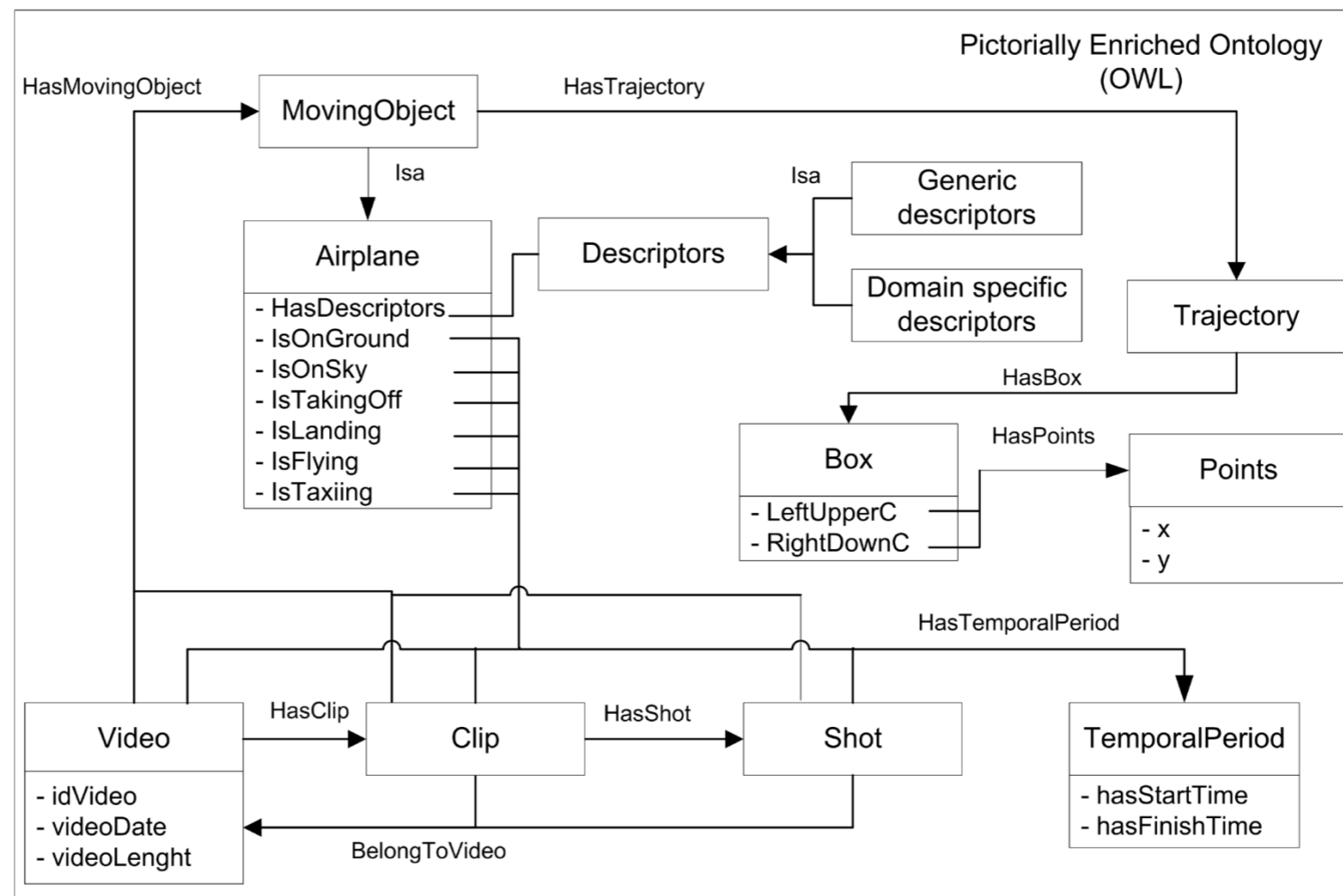
The evaluation function used to estimate the utility of adding a new literal is based

$$Foil_Gain(L, R) \equiv t \left(\log_2 \frac{p_1}{p_1 + n_1} - \log_2 \frac{p_0}{p_0 + n_0} \right)$$



Used Case

- We have applied the automatic video annotation ontology to detection of events related to airplanes, selecting them from the revised list of LSCOM. Four events analyzed are:
 - Airplane flying; Airplane takeoff; Airplane landing; Airplane taxiing
- An ontology defined for these events has been developed





Airplane and sky/ground Detector

- The **airplane detector** has been created using the Viola&Jones object detector.
- Positive and negative examples used to train the detector have been selected from standard image datasets such as Caltech, VOC2005 and VOC2006.
- We have trained five different detectors, using five configurations, with different numbers of positive and negative examples.



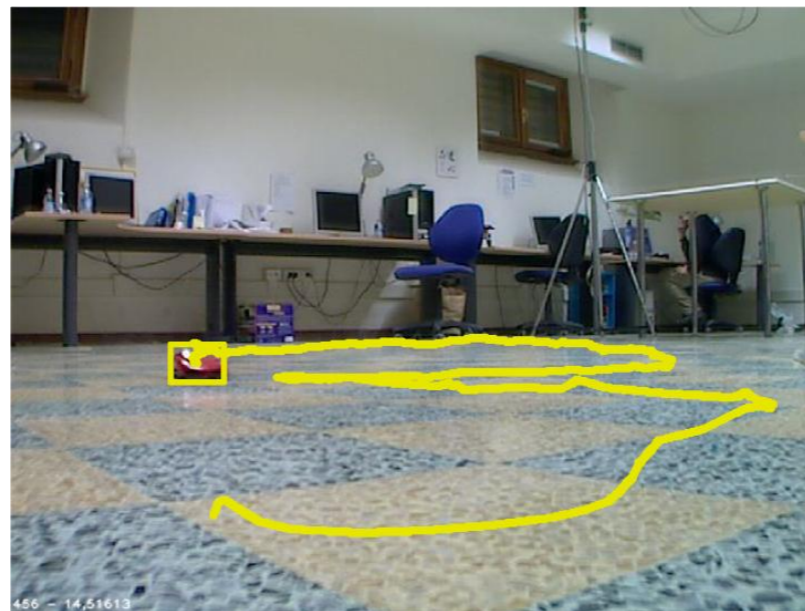
N. detector	N. steps	Neg. examples	Pos. examples	Window size	Precision	Recall
1	17	3000	800	50×30	0.20	0.74
2	18	1500	800	50×30	0.19	0.83
3	20	1500	800	50×30	0.32	0.65
4	20	1500	800	25×10	0.75	0.55
5	22	1500	1040	50×30	0.41	0.66

- The **sky/ground detector** evaluates statistical parameters of the luminance of the blobs around the detected airplane.



Particle filter-based visual tracker

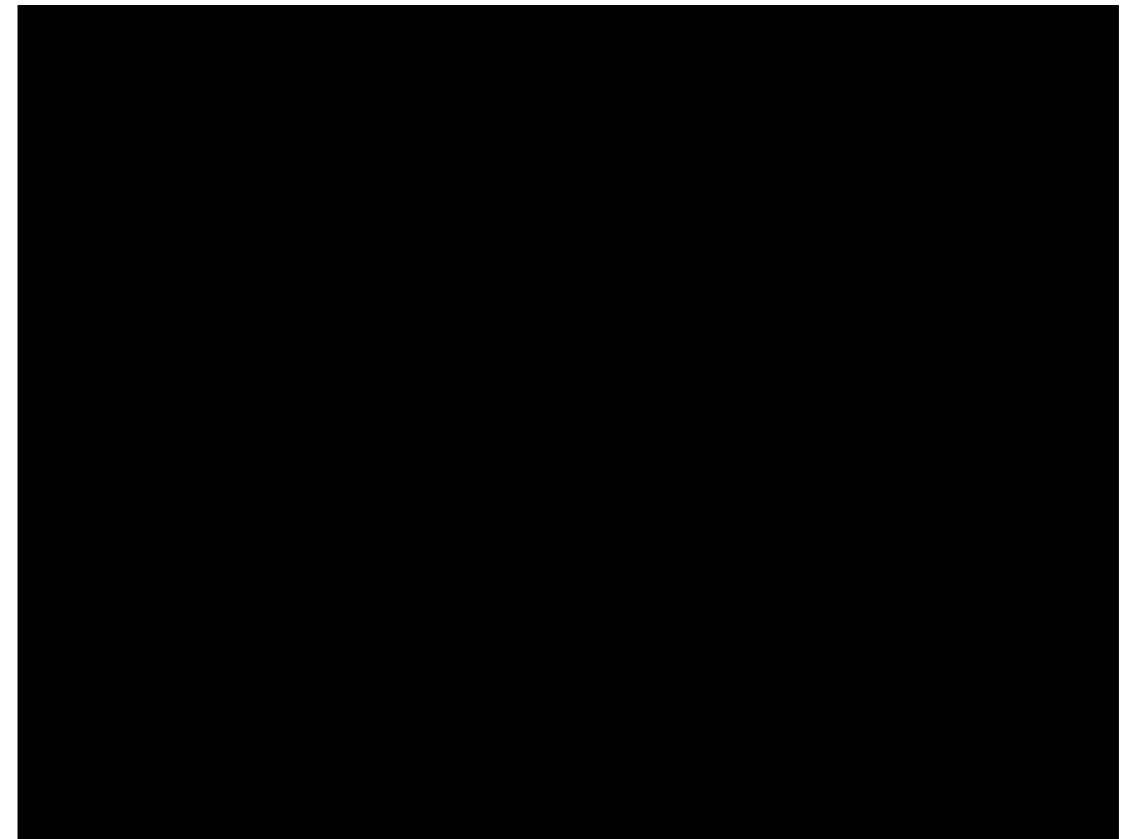
- We used a particle filter-based visual tracker such that:
 - uses a 1st-order dynamic model (tracks position, size, motion direction and velocity)
 - tracks targets in a 8-dimensional state space (400-800 particles)
 - uses a color-based appearance model (in HSV color space)
 - implements a novel method to estimate and adapt the uncertainty in the dynamic model
 - shows improved robustness to occlusions, erratic target motion and appearance changes
 - runs in real-time





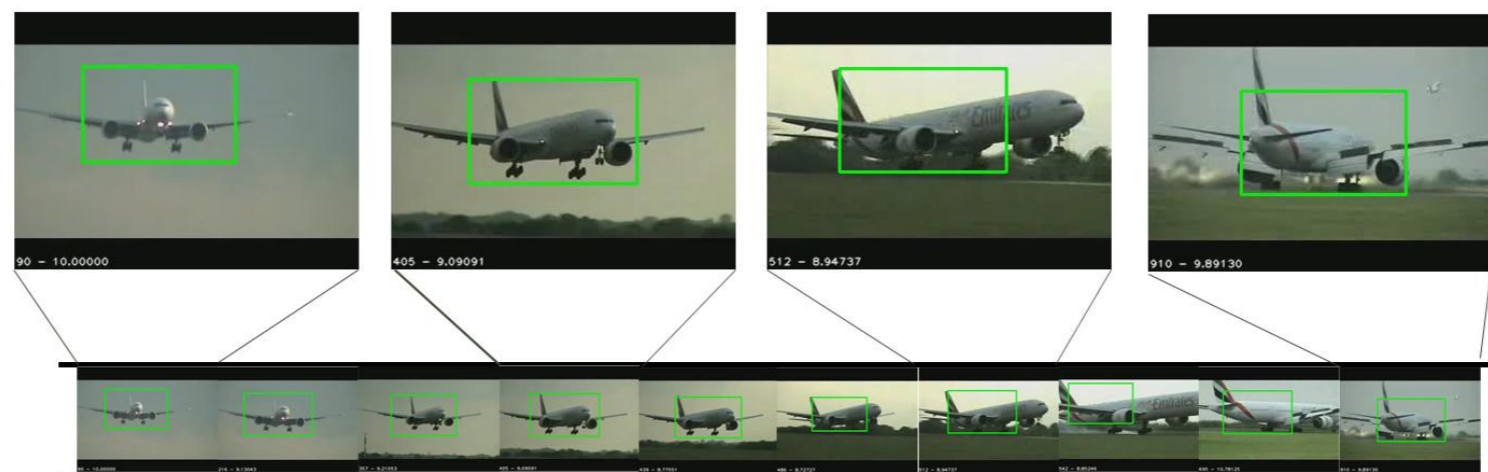
Airplane detector and tracking

- Examples





Event temporal evolution



Airplane is on Ground



Airplane is on Sky



Airplane is on Ground



Airplane is on Sky





Learned rule Experiments

Rule: Airplane TakingOff

Initial rule:

$$Airplane(?p) \wedge Clip(?c) \rightarrow IsTakingOff(?p, ?c)$$

Result rule:

$$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge IsOnGround(?p, ?g2) \wedge$$

$$Temporal : after(?g1, ?g2) \wedge HasTemporalPeriod(?c, ?g3) \wedge Temporal : contains(?g3, ?g1) \wedge$$

$$Temporal : contains(?g3, ?g2) \wedge MovingObject(?p) \rightarrow IsTakingOff(?p, ?c)$$

Rule: Airplane Landing

Initial rule:

$$Airplane(?p) \wedge Clip(?c) \rightarrow IsLanding(?p, ?c)$$

Result rule:

$$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge IsOnGround(?p, ?g2) \wedge$$

$$Temporal : notafter(?g1, ?g2) \wedge HasTemporalPeriod(?c, ?g3) \wedge Temporal : contains(?g3, ?g1) \wedge$$

$$Temporal : contains(?g3, ?g2) \wedge MovingObject(?p) \rightarrow IsLanding(?p, ?c)$$

Rule: Airplane Flying

Initial rule:

$$Airplane(?p) \wedge Clip(?c) \rightarrow AirplaneFlying(?p, ?c)$$

Result rule:

$$Airplane(?p) \wedge Clip(?c) \wedge IsOnSky(?p, ?g1) \wedge$$

$$HasTemporalPeriod(?c, ?g2) \wedge Temporal : contains(?g2, ?g1) \rightarrow IsFlying(?p, ?c)$$

Rule: Airplane Taxiing

Initial rule:

$$Airplane(?p) \wedge Clip(?c) \rightarrow IsTaxiing(?p, ?c)$$

Result rule:

$$Airplane(?p) \wedge Clip(?c) \wedge IsOnGround(?p, ?g1) \wedge$$

$$HasTemporalPeriod(?c, ?g2) \wedge Temporal : contains(?g2, ?g1) \rightarrow IsTaxiing(?p, ?c)$$



Event recognition Experiments

- Precision and recall of Airplane flying, Airplane takeoff, Airplane landing and Airplane taxiing for different datasets.

Data Set	Airplane Action	Precision	Recall
Web Dataset	Airplane flying	0.96	0.94
Web Dataset	Airplane takeoff	0.76	0.80
Web Dataset	Airplane landing	0.84	0.96
Web Dataset	Airplane taxiing	1	0.76
TRECVID 2005	Airplane flying	0.94	0.5
TRECVID 2005	Airplane takeoff	0.3	0.42
TRECVID 2005	Airplane landing	0.66	0.66
TRECVID 2005	Airplane taxiing	1	0.76
Web Dataset + TRECVID 2005	Airplane flying	0.96	0.71
Web Dataset + TRECVID 2005	Airplane takeoff	0.61	0.70
Web Dataset + TRECVID 2005	Airplane landing	0.90	0.90
Web Dataset + TRECVID 2005	Airplane taxiing	0.94	0.84



Conclusions and Future work

- Pictorially Enriched Ontology has been defined to perform automatic semantic annotation of video.
- A set of Rule used to describe events have been learned from positive and negative examples using FOILS technique
- Next steps are to answer to these questions.
- How can we insert qualitative temporal information?
- How to deal with noisy data?



Thank you