School of Computing

UNIVERSITY OF LEEDS

Learning about Activities and Objects from Video

Tony Cohn

What does an agent need to know about the world?

- What kind of objects there are.
- What they do/can be used for.
- What kinds of events there are.
- Which objects participate in which events.

…

- *How can an agent acquire this knowledge?*
- *How should it represent it?*

Today’s talk

UNIVERSITY OF LEEDS

- Learning about
  - events: analyse activities in terms of event classes involving multiple objects
  - object categories via activity analysis
- Relational approach

This approach has a long history…

Barrow and Popplestone:
Relational descriptions in picture processing
Machine Intelligence 6, 1971

Relational descriptions of object classes + supervised learning

(re-)Connecting Logic and vision
(Kanade IJCAI’03)

From pixels to symbols to understanding

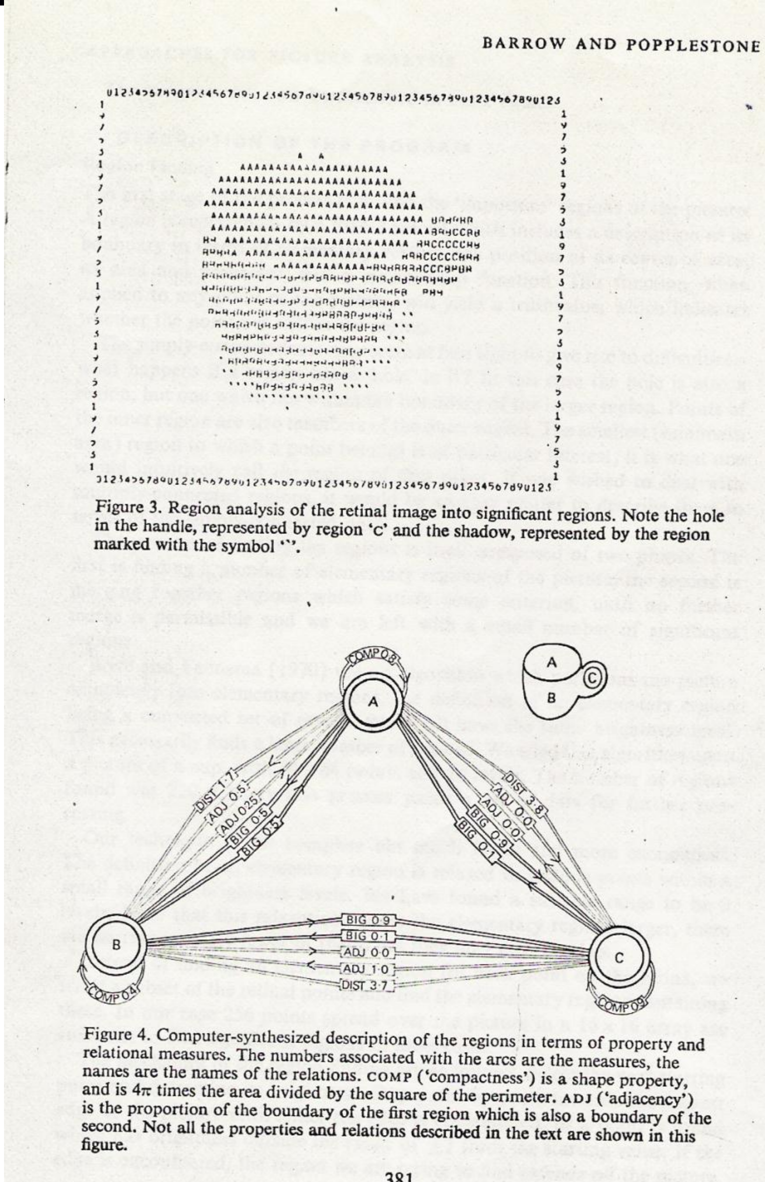


Figure 3. Region analysis of the retinal image into significant regions. Note the hole in the handle, represented by region ‘c’ and the shadow, represented by the region marked with the symbol ‘’’.

COMP 0·8
A
B
C
DIST 1·7
ADJ 0·5
ADJ 0·25
BIG 0·5
BIG 0·5
DIST 3·8
ADJ 0·0
ADJ 0·0
BIG 0·9
BIG 0·1
BIG 0·9
BIG 0·1
ADJ 0·0
ADJ 1·0
DIST 3·7
COMP 0·4
COMP 0·9

Figure 4. Computer-synthesized description of the regions in terms of property and relational measures. The numbers associated with the arcs are the measures, the names are the names of the relations. COMP (‘compactness’) is a shape property, and is 4π times the area divided by the square of the perimeter. ADJ (‘adjacency’) is the proportion of the boundary of the first region which is also a boundary of the second. Not all the properties and relations described in the text are shown in this figure.

…with an interesting conclusion

‘…let us consider the object recognition program in its proper perspective, as part of an integrated cognitive system. One of the simplest ways that such a system might interact with the environment is simply to shift its viewpoint, to walk round an object. In this way more information may be gathered and ambiguities resolved ......

...... Such activities involve ***planning, inductive generalization, and, indeed, most of the capacities required by an intelligent machine****. To develop a truly integrated visual system thus becomes almost co-extensive with the goal of producing an integrated cognitive system.’*

Barrow and Popplestone, 1971.

Movement can be at least as important as appearance in what we perceive:

Not just movement, but spatial relations between objects over time.

Heider & Simmel, 1944

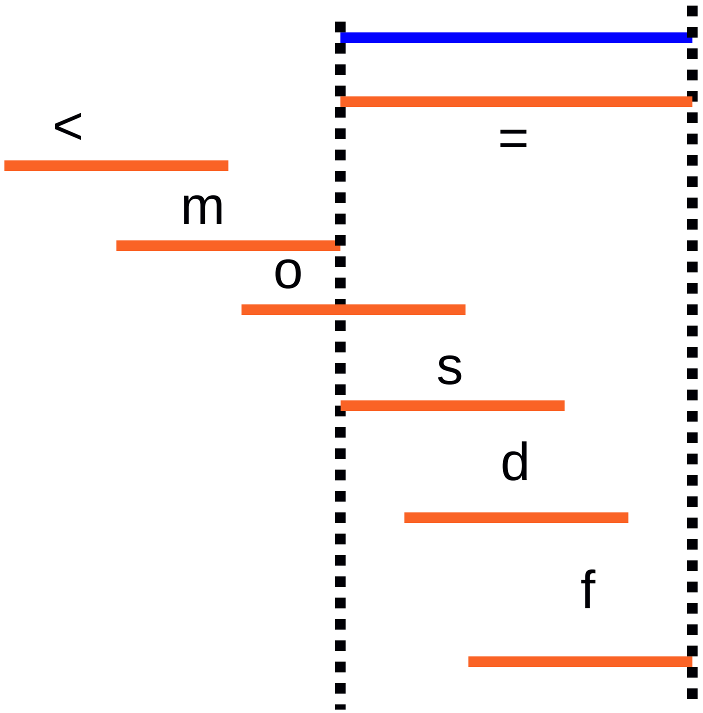
Qualitative spatial/spatio-temporal representations

- Complementary to metric representations
- Human descriptions tend to be qualitative
- Naturally provides abstraction
  - Machine learning
- Provide foundation for domain ontologies with spatially extended objects
- Applications in geography, **computer vision**, robotics, NL, biology…
- Well developed calculi, languages

Qualitative temporal representations

Allen’s interval calculus

13 jointly exhaustive and pairwise disjoint relations

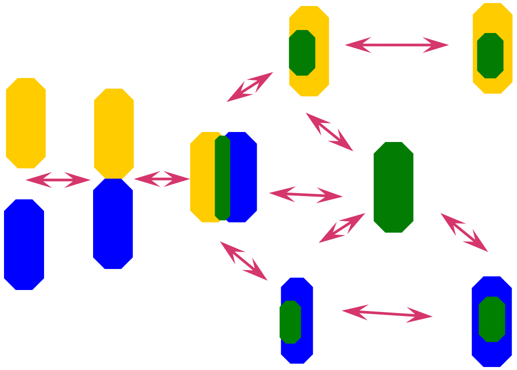


Qualitative spatial representations

UNIVERSITY OF LEEDS

Region Connection Calculus (RCC8)

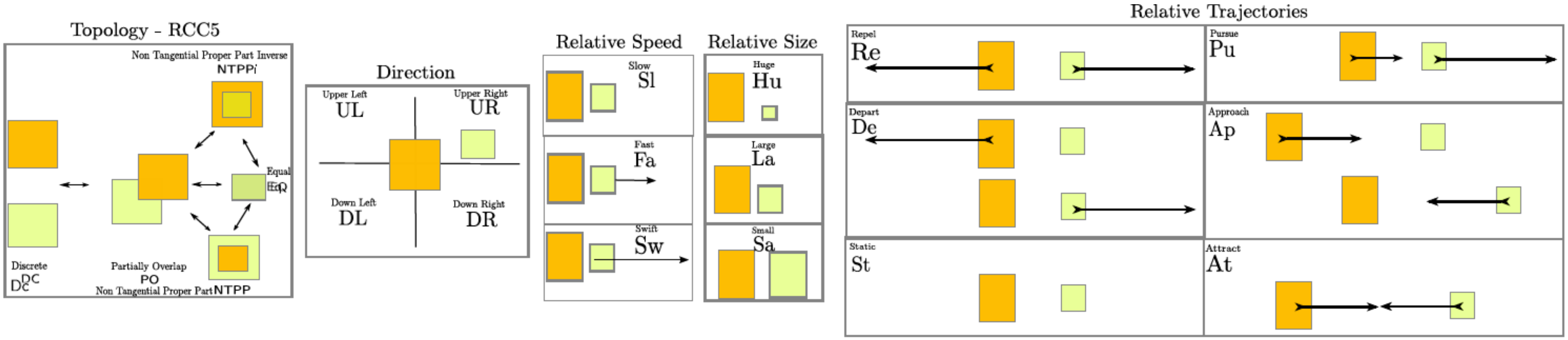
(mereo)topology



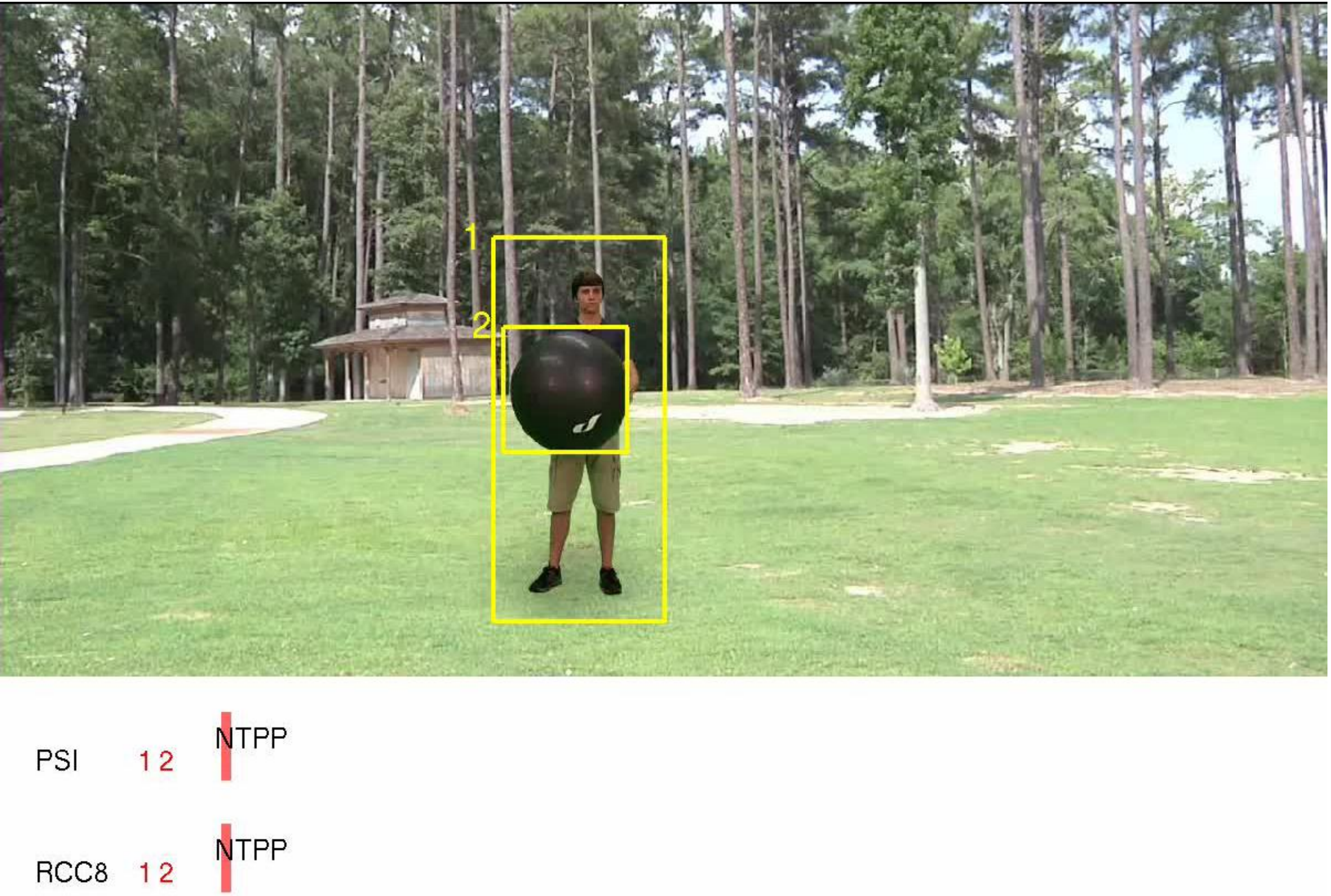
Qualitative spatial representations

Many other qualitative spatial calculi

- Other topological calculi
- Direction
- Size
- Distance
- Orientation

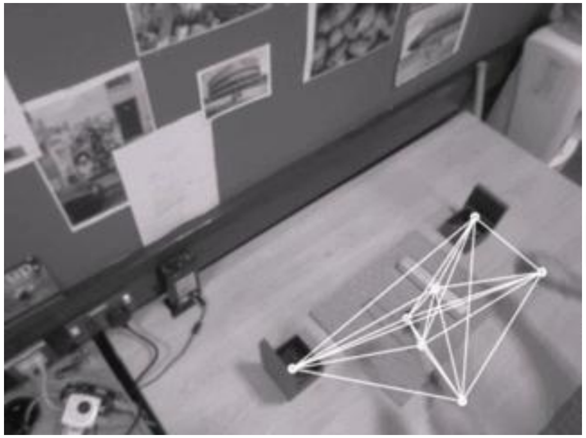


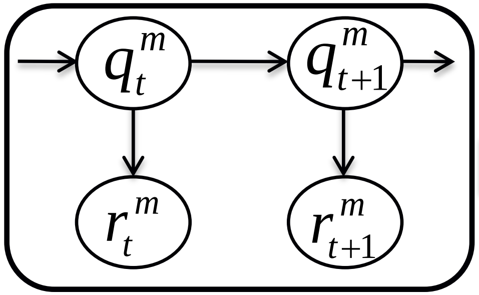
Using an HMM to ‘smooth’ relations

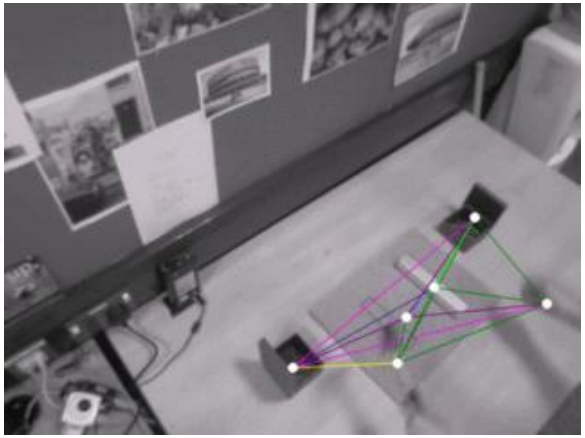


Sridhar et al., *COSIT 2011*

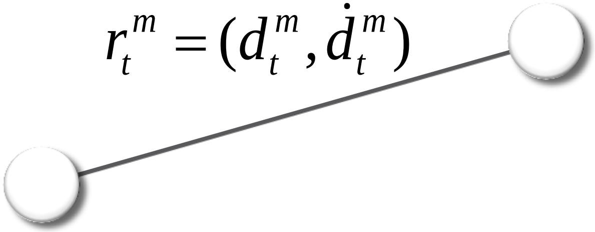
Learning relations

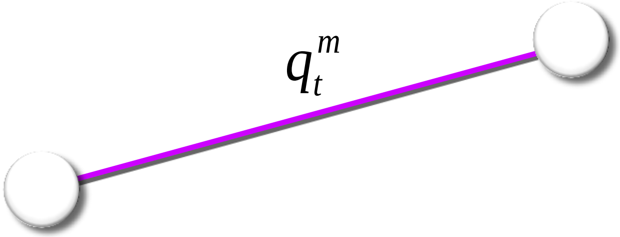






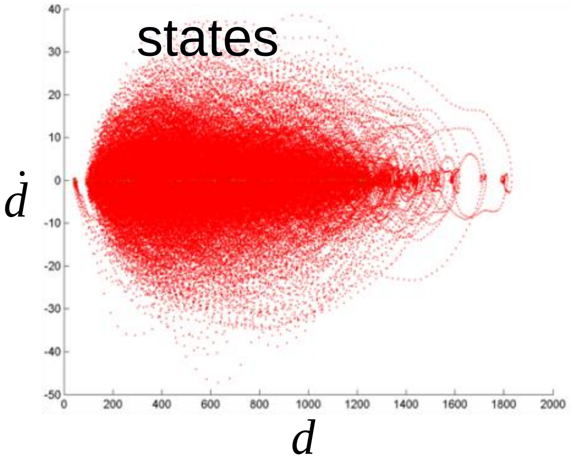
$$r_t^m = (d_t^m, \dot{d}_t^m)$$



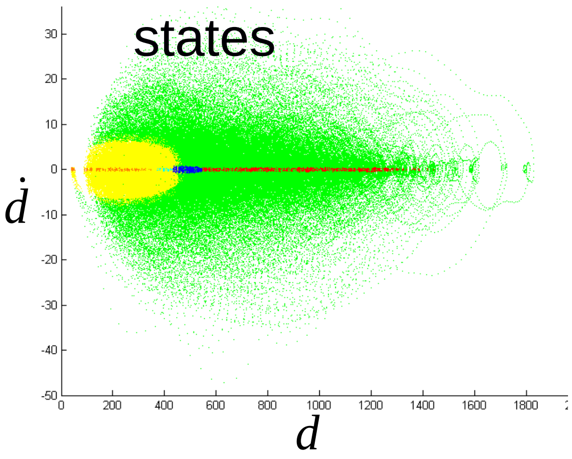


Quantisation of Relational Features

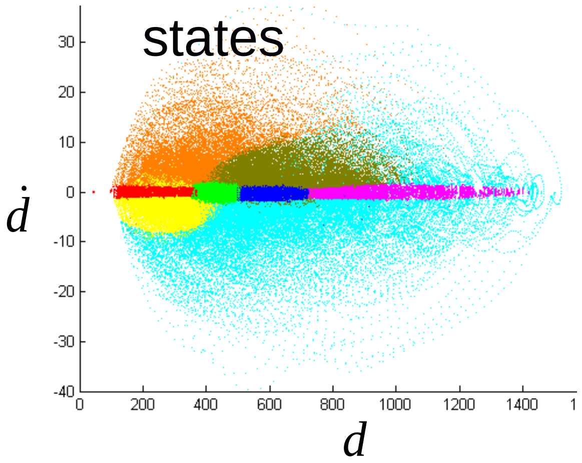
2 discrete states



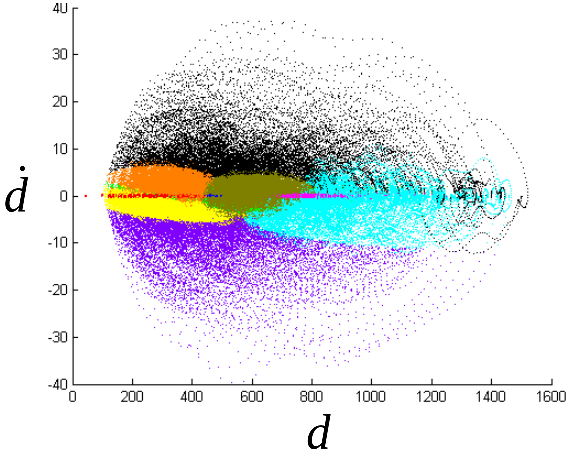
6 discrete states



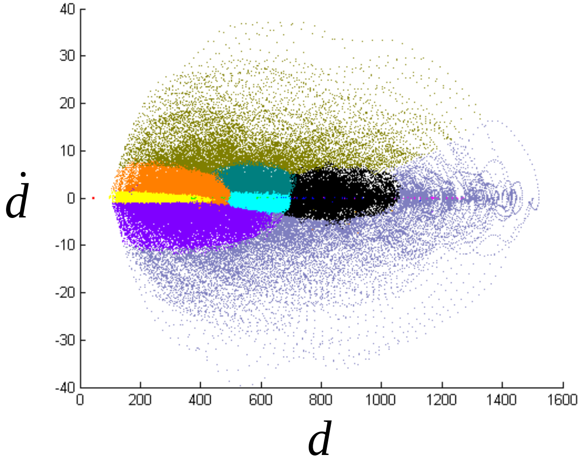
8 discrete states



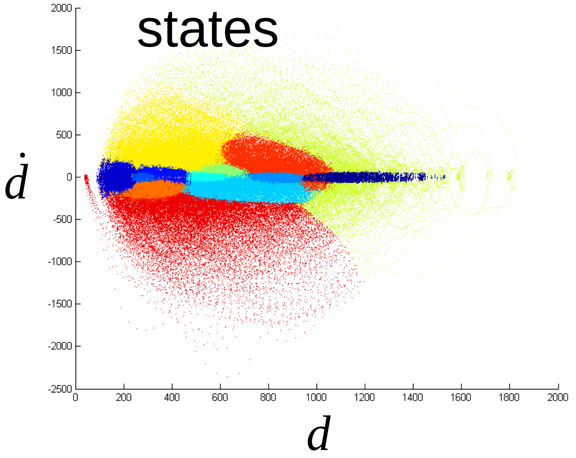
10 discrete states



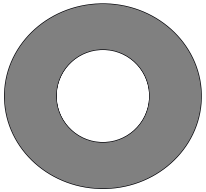
12 discrete states



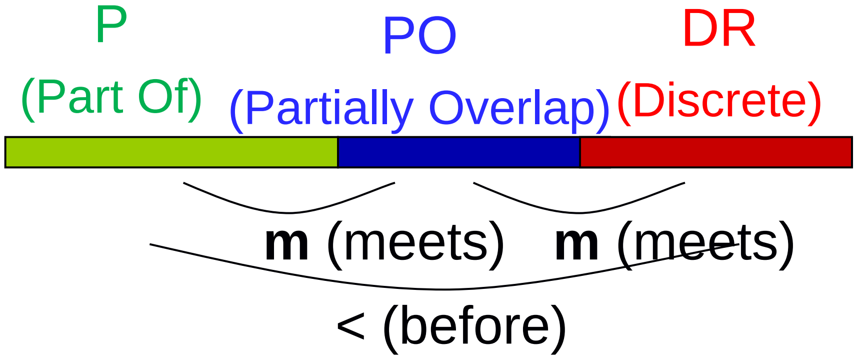
16 discrete states



Representing interactions relationally



$$holds(X,Y,\mathsf{P},I_1) \wedge holds(X,Y,\mathsf{PO},I_2) \wedge holds(X,Y,\mathsf{DR},I_3) \wedge \mathsf{meets}(I_1,I_2) \wedge \mathsf{meets}(I_2,I_3) \wedge \mathsf{before}(I_1,I_3)$$



P

(Part Of)

PO

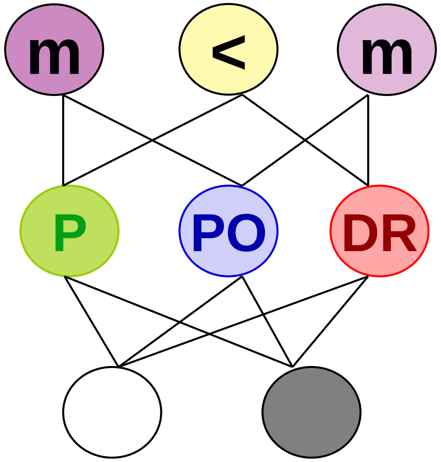
(Partially Overlap)

DR

(Discrete)

m (meets) **m** (meets)

< (before)



m < m

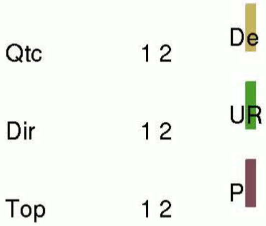
P PO DR

3 Allen’s Temporal Relationships (x 13)

2 Spatial Relationships (x 3)

1 Objects

Representing an entire video relationally



Steps towards Cognitive Vision (at Leeds)

System which learned traffic behaviours (ICCV'98, IVC'00)

- Qualitative spatio-temporal models

Learning of qualitative spatial relationships (ECAI'02)

- Allows domain specific distinctions to be learned

Reasoning about commonsense knowledge of continuity to improve tracking (ECAI'04)

Learning symbolic descriptions of intentional behaviours

- Use ILP to induce rules for simple games (AIJ 2004, ECAI'04,…)

Learning Qualitative Spatio-temporal event class descriptions

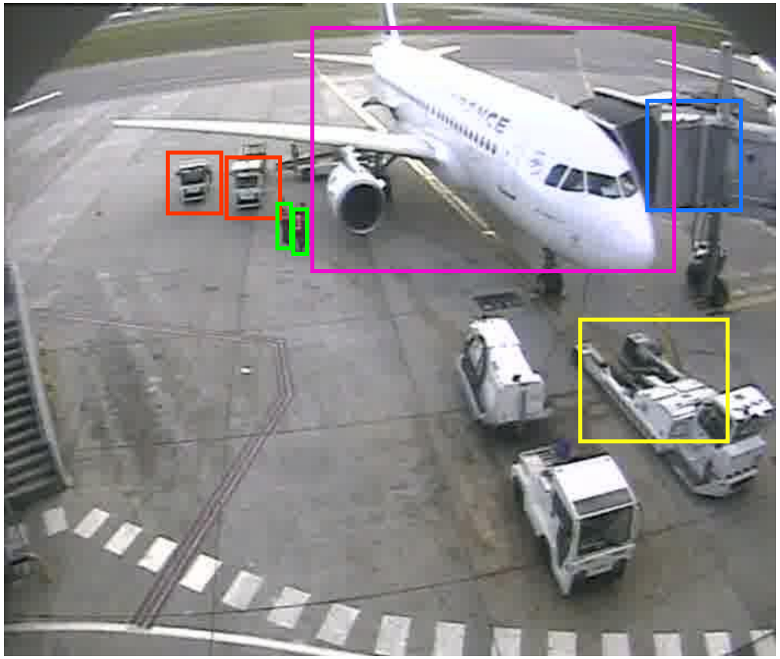
- Supervised (ECAI'10a, ILP-11)
- Unsupervised (ECAI'08, AAAI'10, ECAI'10, STAIRS'10, COSIT-11)

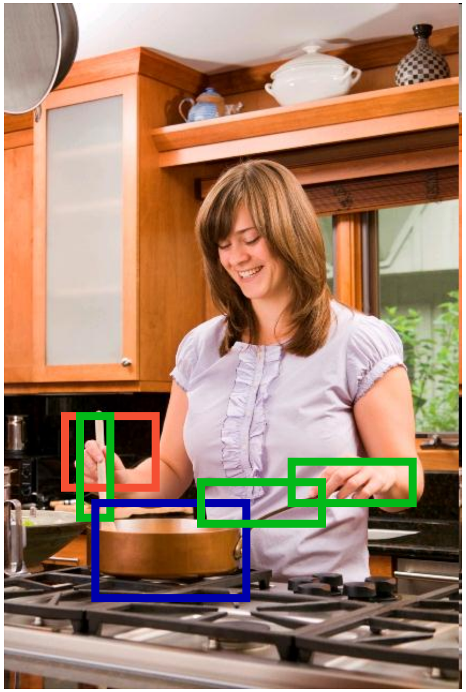
Functional Object Categories from a Relational Spatio-Temporal Representation (ECAI'08)

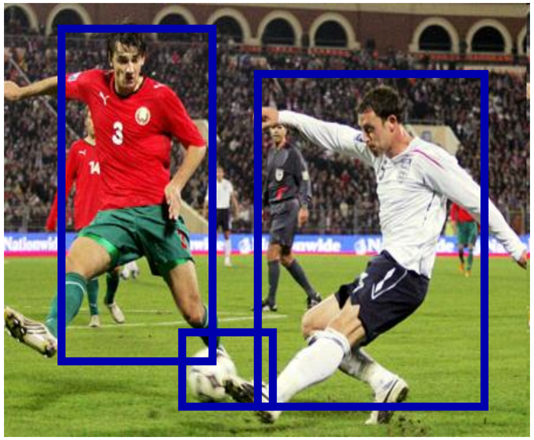
Workflow Activity Monitoring using the Dynamics of Pair-wise Qualitative Spatial Relations (MMM-12) …

Problem: Understanding Activities

- Point a camera at a scene with **complex activities** where objects are interacting.
- We start our analysis after obtaining **object tracks**.







sli

Problem: Understanding Activities

- Activities consist of **events.**

- Events are **goal directed interactions** between a subset of objects.

- **Events are patterns –** instances of event classes
  - **but may be hidden by noise/coincidental interactions.**

- Can we **learn events** from complex activities in an **unsupervised** way despite the **presence of noise/coincidences**?

Events and Event Classes

What is an event?

- a set of spatio-temporal histories
  - some set of objects *interacting* at a particular time
  - each event is *unique*

What is an event class?

- at some level of abstraction, events will have *similar* descriptions
  - *qualitative* spatio-temporal change
  - *frequent* occurrences of similar events

Mining event classes

- What do we mean by *interacting*?
  - *How many objects* involved?
- What do we mean by *similar*?
- How *frequent*?
- *Complete* object histories, or *partial*?
  - How to split?
- Distinguishing *simple from complex* events
- Distinguishing *between contemporaneous events*
- How to find event classes efficiently?
  - How to search?

Interactivity: *an attention mechanism for isolating subsets of objects over intervals*

All objects interact evenly over interval

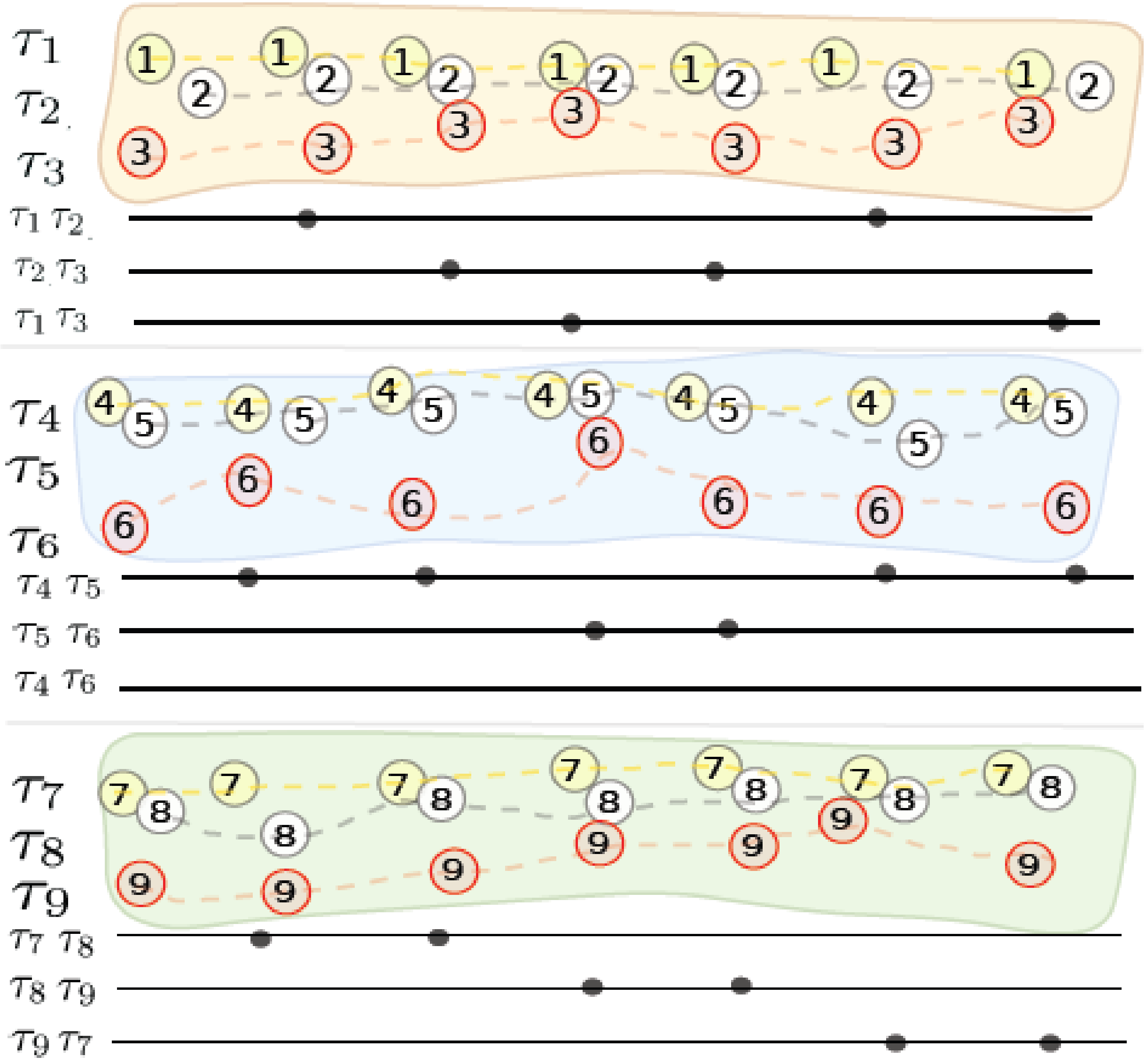
✓

τ_6 doesn't interact much

X

Interactions temporally extended

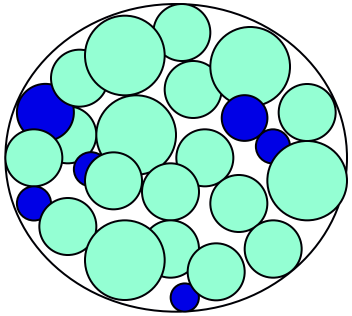
X



Definitions

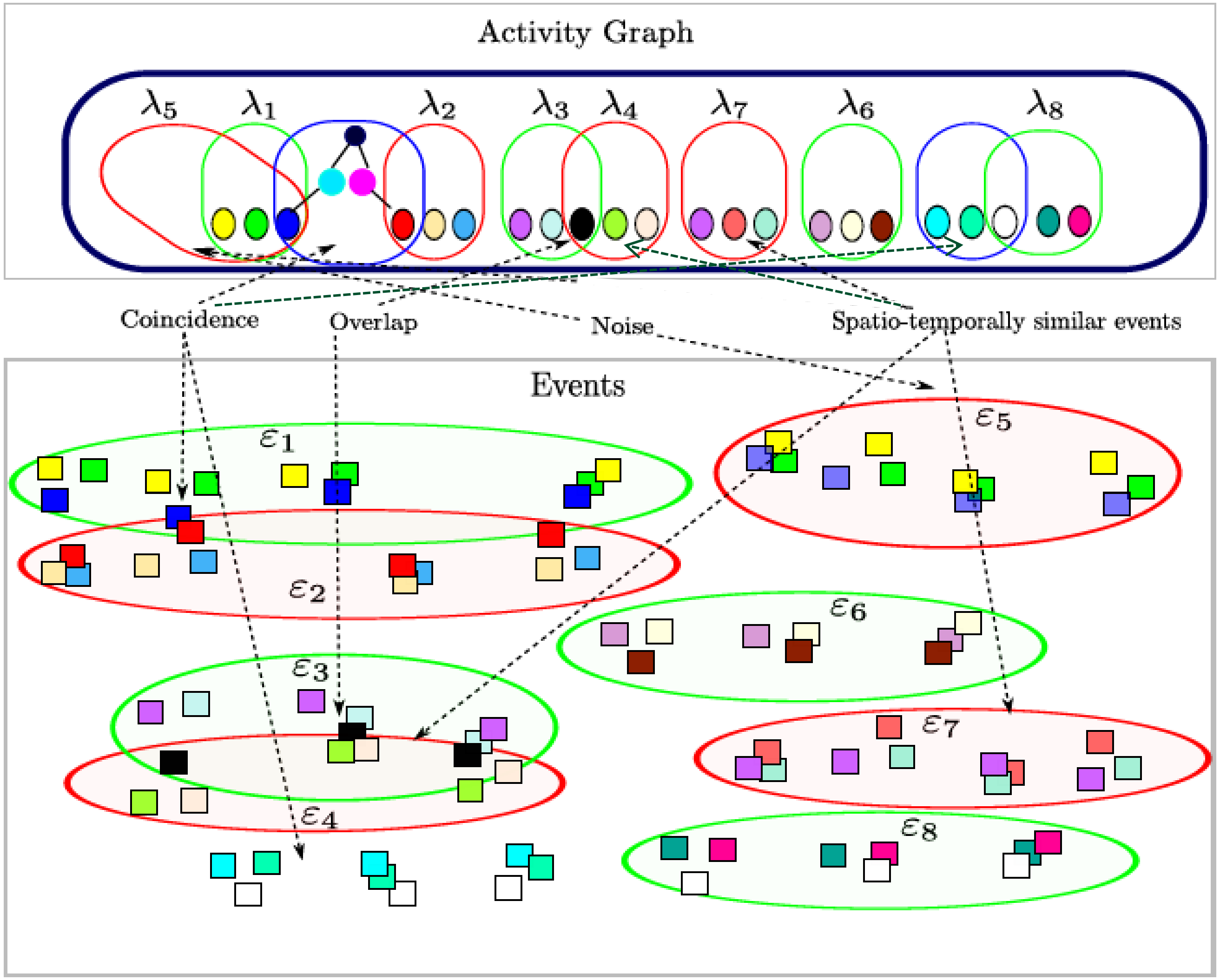
An *interaction graph* is a sub-graph of the activity graph that contains all spatial/temporal relations between its objects over some time interval.

A *cover* Λ for an activity graph A is a subset of the interaction graphs that jointly cover the (relevant) nodes of the activity graph. Furthermore, each interaction graph is labelled as **event** or **coincidence**.

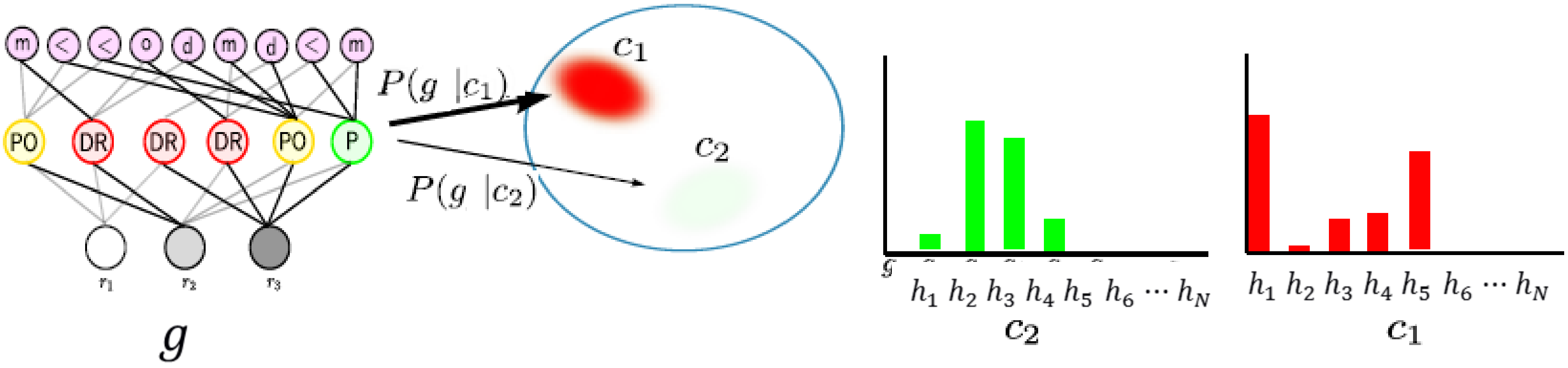


A *model* Θ defines a probability function over interaction graphs.

Activity graph and cover



Mixture distribution over interaction graphs



$h_1\ h_2\ h_3\ h_4\ h_5\ h_6 \cdots h_N$

$$p(g|c_i) = \sum_{k=1}^{N} q_k^i \mathcal{K}_d(g, h_k) \qquad \text{where} \sum_{k=1}^{N} q_k^i \ = 1$$

A similarity measure between graphs

In summary:

• Each event class is specified non-parametrically by a set of event graphs

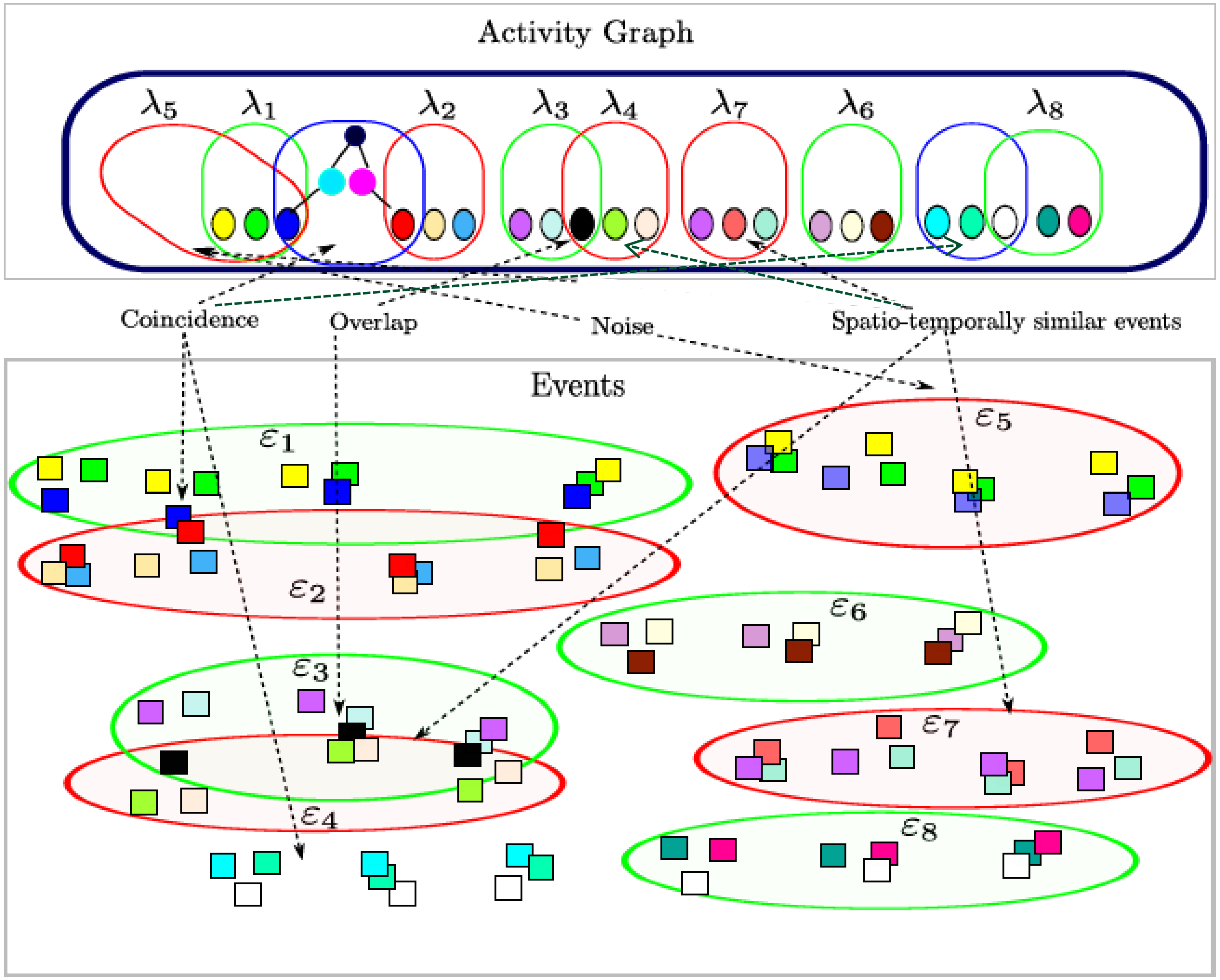
• Try to explain data by finding a set of event classes such that the tracks can be divided into sets of tracklets each of which obeys the spatio-temporal constraints of some event class

• A good explanation:

- *explains as much as possible*
- minimizes number of event classes

  (classes will tend to have many instances, and will be `large’)
- has event classes all of whose graphs are similar
- has events which have high degree of tracklet interaction and *low* object sharing with other events

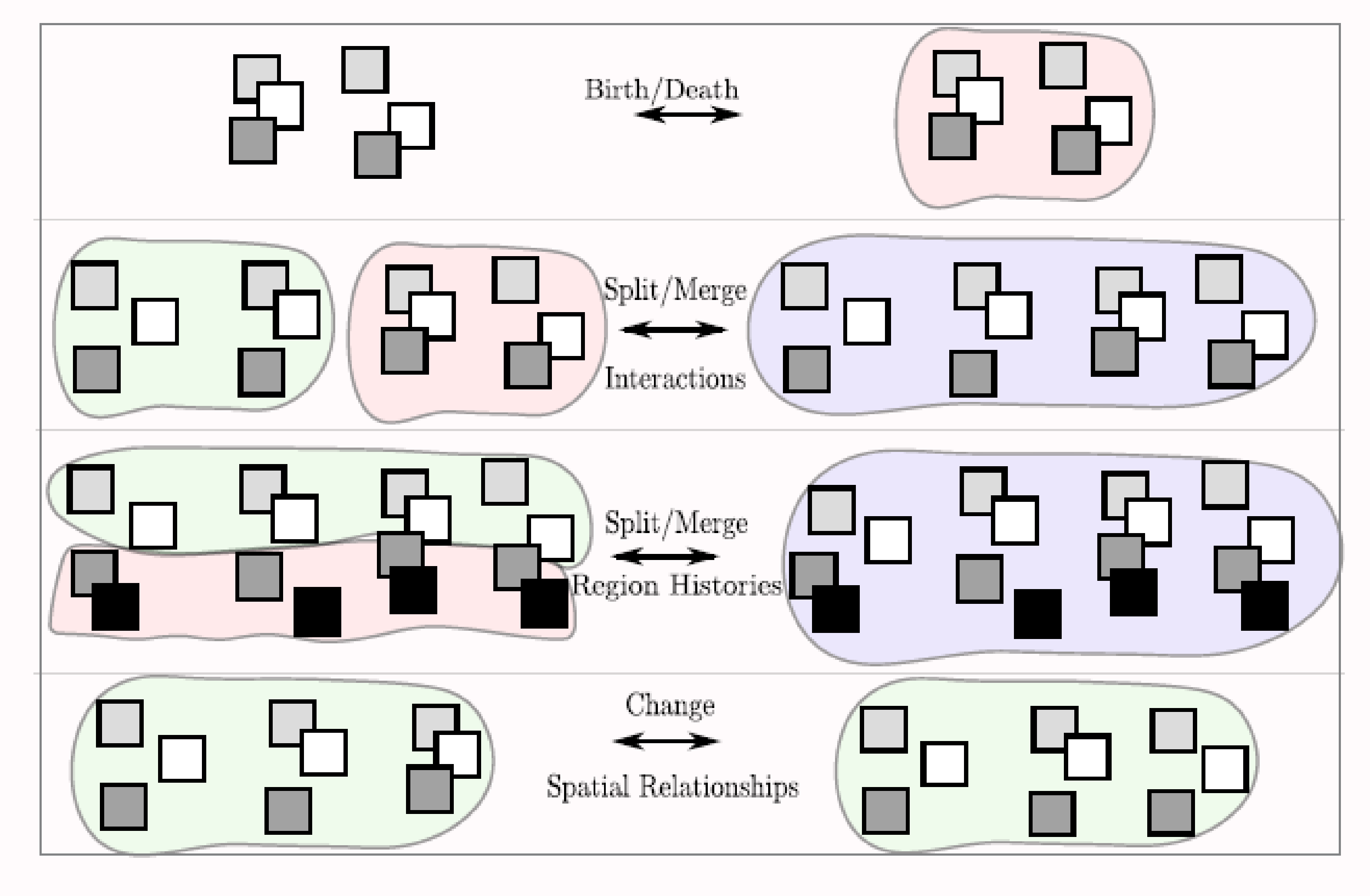
Activity graph and cover



Find MAP interpretation using MCMC

UNIVERSITY OF LEEDS

MCMC moves



Evaluation in aircraft domain

24 aircraft turnarounds – 37 hours

Single viewpoint

Semi-automated tracking of the *plane, trolley, carriage, loader, bridge, plane-puller*

Discard class labels

Obtain RCC3 relations in image plane

Results

Discovered two classes

	Semantic category	True positives	False positives	False negatives	Precision	Recall
Class 1	(un)loading	14	4	6	78	70
Class 2	bridge-on-off & plane-puller-on	16	7	6	70	80

Learning object classes from behaviour (not appearance)

Most computer vision work on learning object classes recognises objects from their appearance

Can we categorise objects by what they do, not what they look like?

Inducing a functional object taxonomy (ECAI-2008)

UNIVERSITY OF LEEDS

Form Boolean matrix of the role played by objects in each event class (+ partially generalised classes)

Event classes

Objects

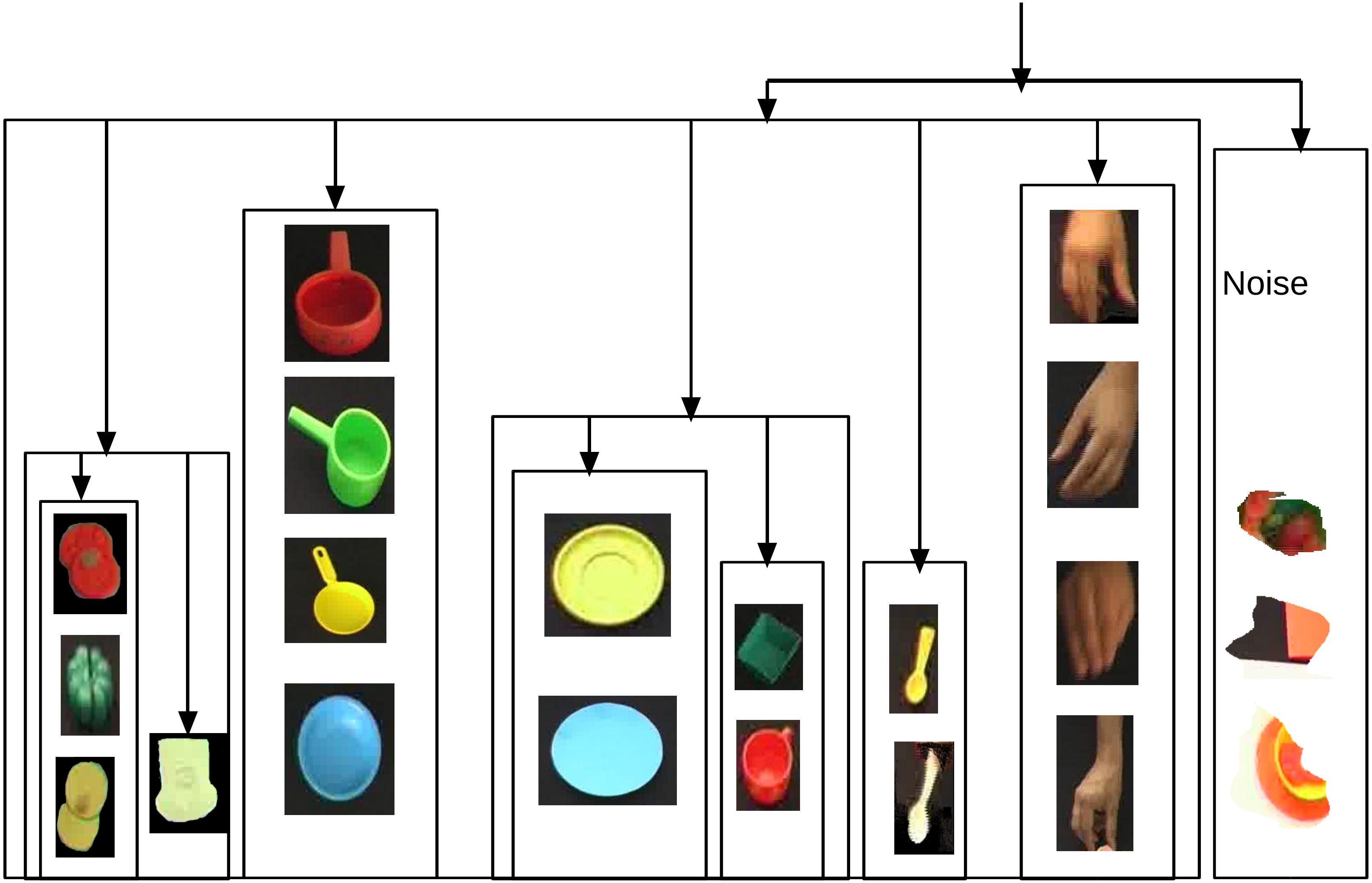
$$\begin{array}{c|cccccccccc}
 & & E_1 & & E_2 & & & & E_m & & \\
 & (\bullet & \bullet & \bullet) & (\bullet & \bullet) & \cdots & (\bullet & \bullet & \bullet & \bullet) \\
\hline
o_1 & 0 & 1 & 0 & 0 & 0 & & 0 & 0 & 1 & 0 \\
o_2 & 1 & 0 & 0 & 0 & 1 & & 0 & 0 & 0 & 0 \\
\vdots & & & & & & & & & & \\
o_n & 1 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0
\end{array}$$

Compress the rows (pattern for each object) using PCA

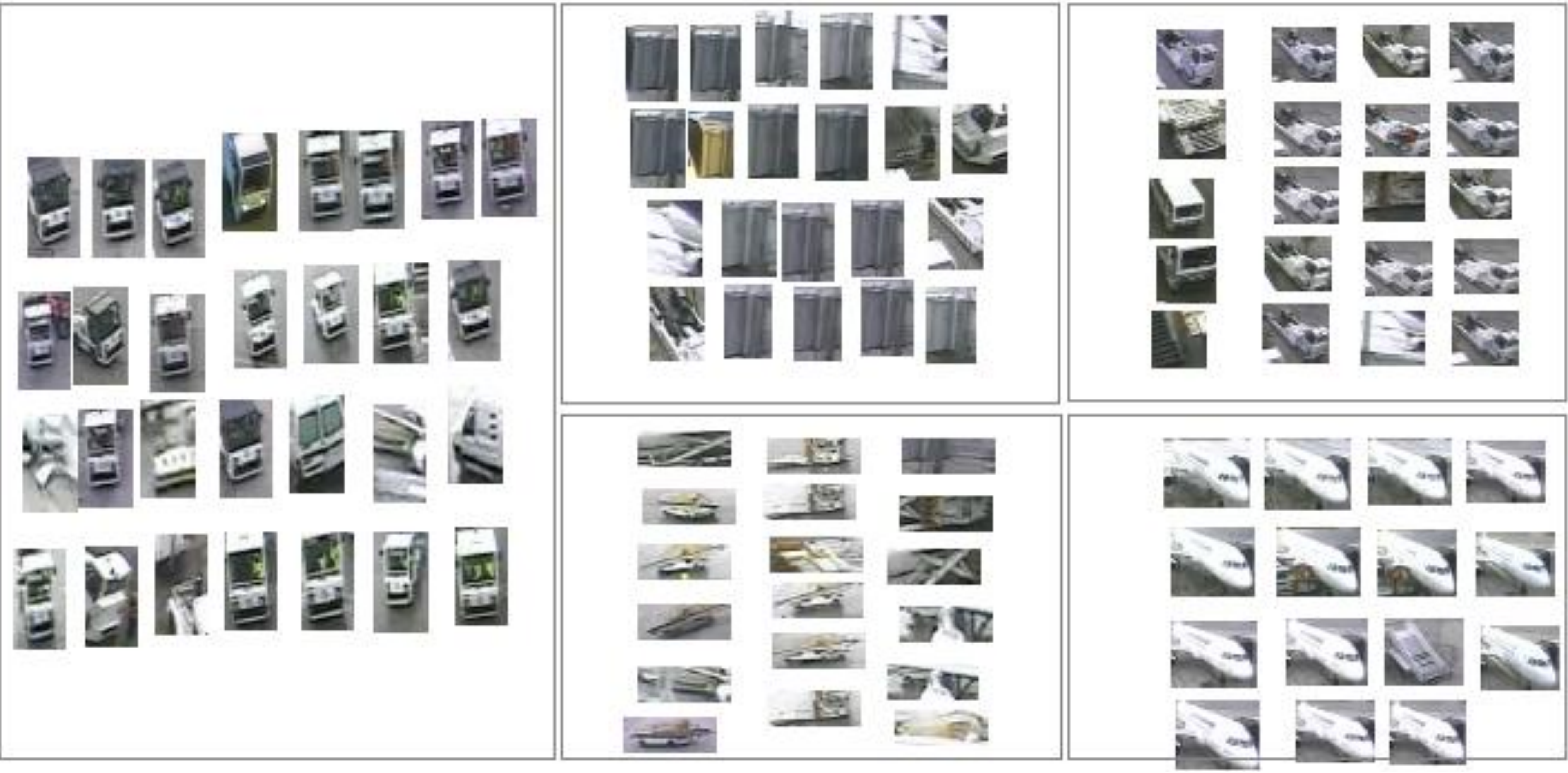
Obtain object taxonomy by hierarchical-clustering of the compressed rows

Emergent object classes: *Toy kitchen hierarchy*

UNIVERSITY OF LEEDS



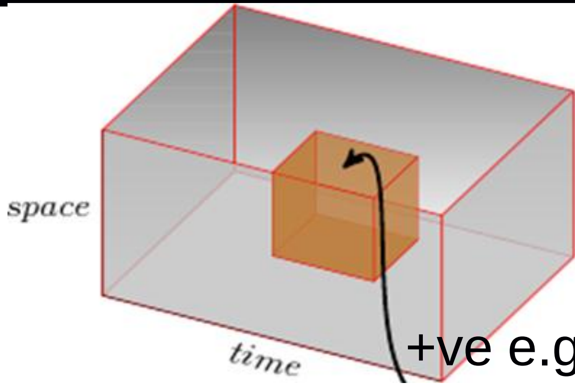
Emergent object classes: *Aircraft domain*



Semi-supervised event learning

Look what's *happening* over *there*

- "Deictic supervision"



- Just specify a rough spatio-temporal region for positive examples

  - No need to specify *exactly* which objects are involved in the event.

- We have developed a *transactional*, *typed* Inductive Logic Programming (ILP) system to induce rules.

What is Inductive logic programming?

- Machine learning, where the hypothesis space is the set of all logic programs
- Logic programs are a subset of First Order Logic
- A set of rules of the form:

$$\text{Event}(\ldots) \leftarrow \text{Condition}_1(\ldots) \wedge \ldots \wedge \text{Condition}_n(\ldots)$$

- Very expressive
- Learning consists of finding a set of rules such that all (most?) of the examples are correctly labelled by these rules.
- We use types to:

  - improve efficiency

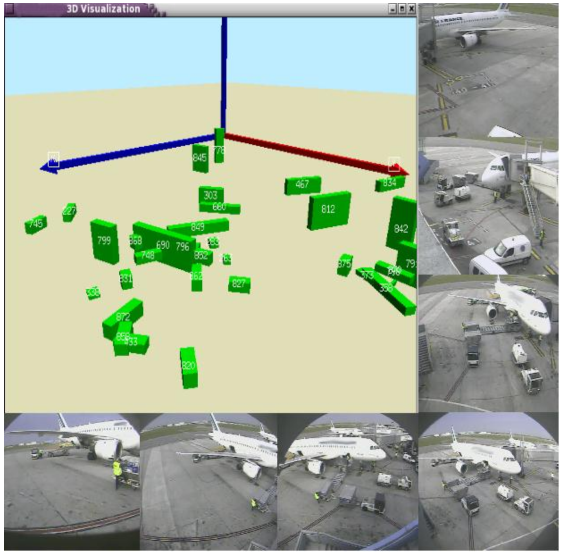
  - reduce overgeneralisation from noisy examples

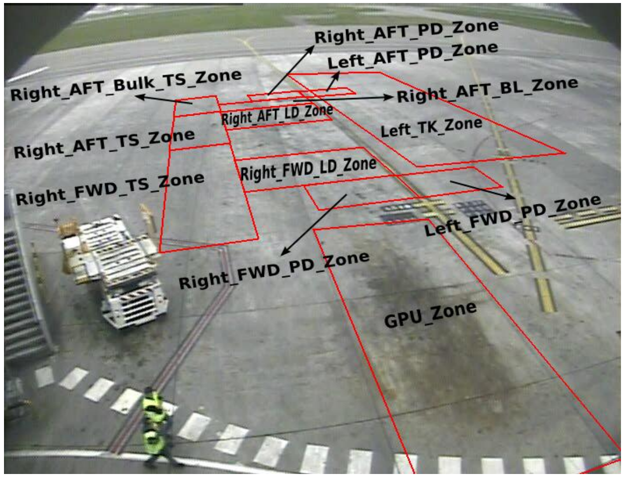
Evaluation

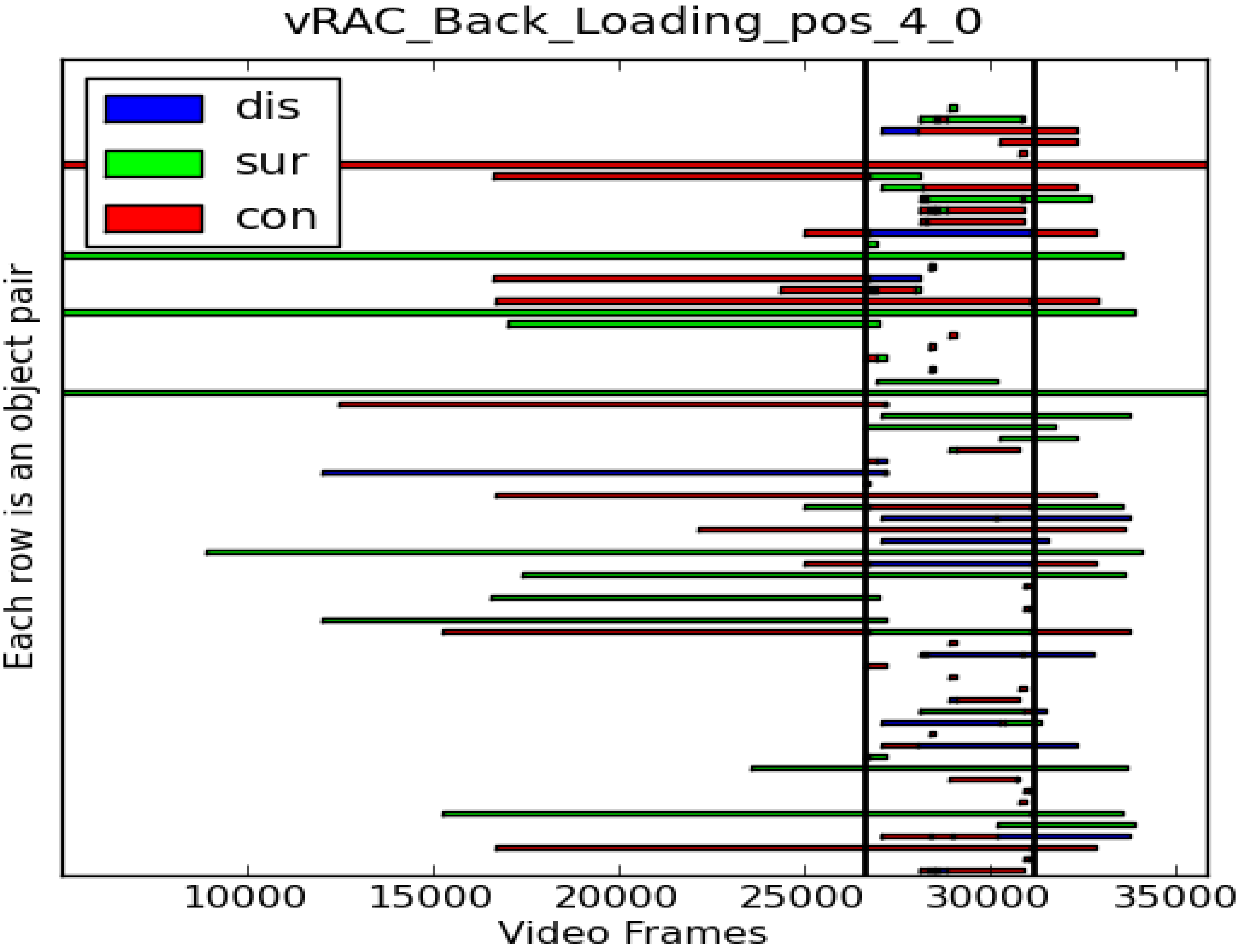
- 15 aircraft turnarounds
- 50,000 frames each turnaround
- 6 camera views
- Obtain tracks on 2D ground-plane
- ~350 spatial facts/video +temporal
- 10 event classes, 3-15 examples for each
- Many errors:

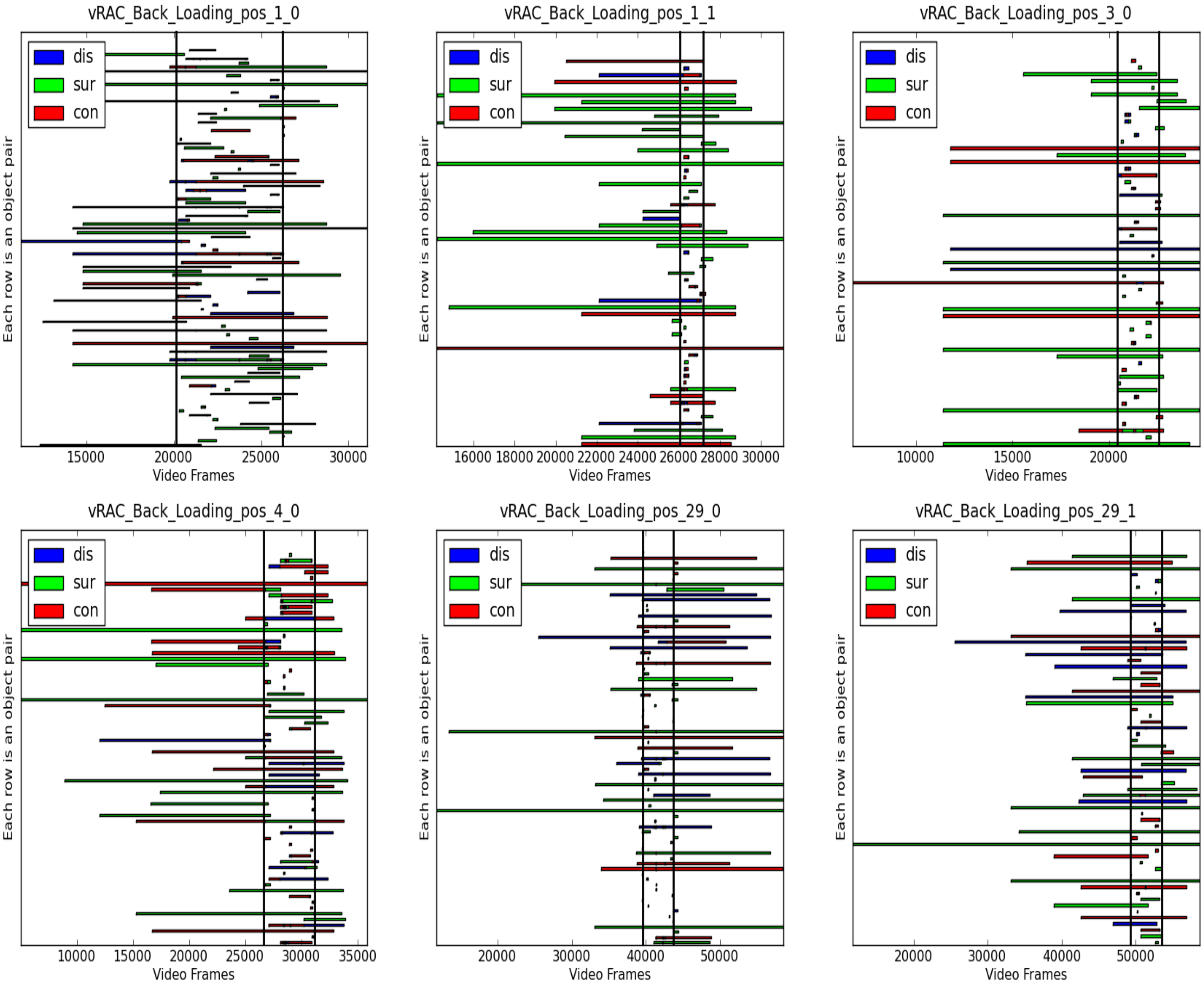
  - false/missing/displaced objects

  - broken/switched tracks
- Generate spatial relations between objects/IATA-zones
- Prolog rules determining temporal relations are in Background
- Leave-one-out (from turnarounds) testing





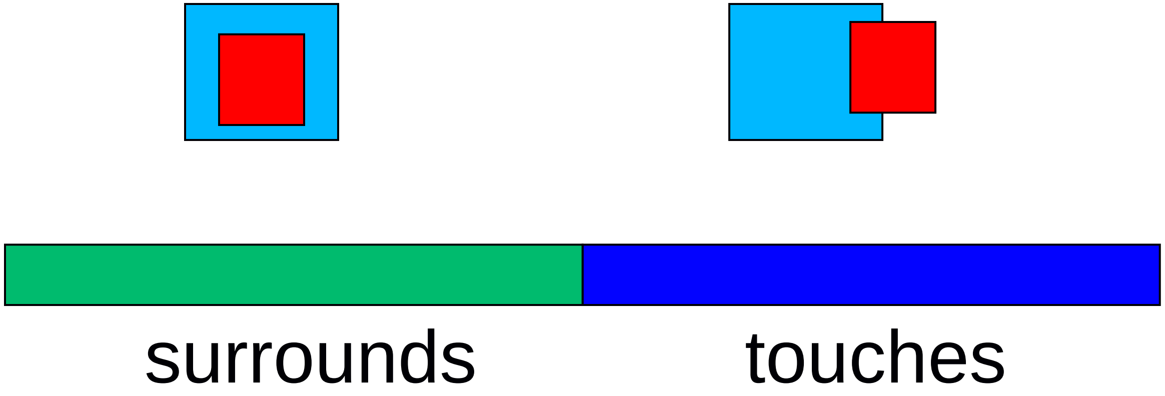

vRAC_Back_Loading_pos_4_0
dis
sur
con
Each row is an object pair
10000
15000
20000
25000
30000
35000
Video Frames


vRAC_Back_Loading_pos_1_0
dis
sur
con
Each row is an object pair
15000
20000
25000
30000
Video Frames
vRAC_Back_Loading_pos_1_1
dis
sur
con
Each row is an object pair
16000
18000
20000
22000
24000
26000
28000
30000
Video Frames
vRAC_Back_Loading_pos_3_0
dis
sur
con
Each row is an object pair
10000
15000
20000
Video Frames
vRAC_Back_Loading_pos_4_0
dis
sur
con
Each row is an object pair
10000
15000
20000
25000
30000
35000
Video Frames
vRAC_Back_Loading_pos_29_0
dis
sur
con
Each row is an object pair
20000
30000
40000
50000
Video Frames
vRAC_Back_Loading_pos_29_1
dis
sur
con
Each row is an object pair
20000
30000
40000
50000
Video Frames

A Learned Event Model:

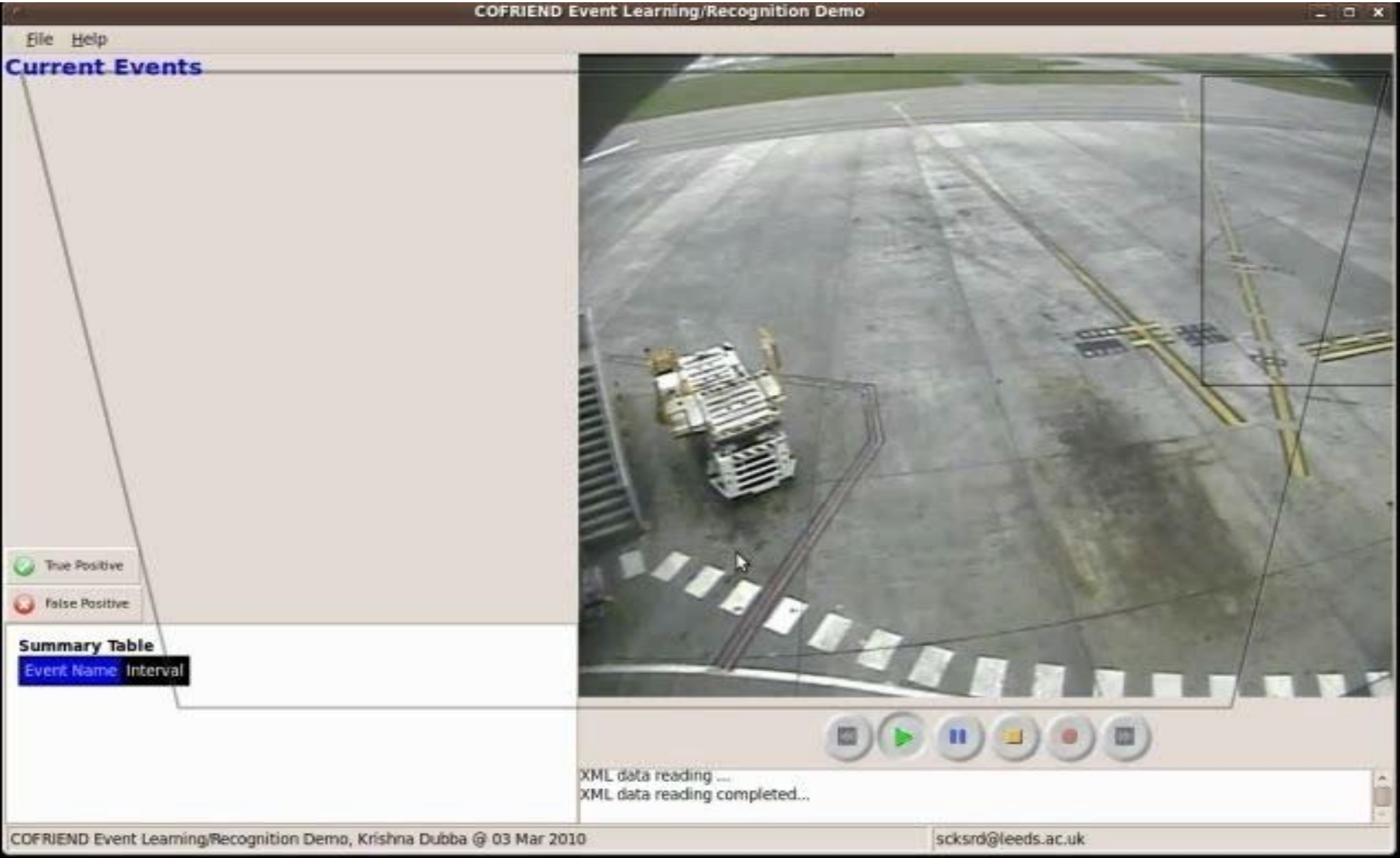
UNIVERSITY OF LEEDS

- aircraft_arrival([intv(T1,T2),intv(T3,T4)]) ←
  surrounds(obj(aircraft(V)), right_AFT_Bulk_TS_Zone, intv(T1,T2)),
  touches(obj(aircraft(V)), right_AFT_Bulk_TS_Zone, intv(T3,T4)),
  meets(intv(T1,T2),intv(T3,T4)).



Applying the learned rules:

UNIVERSITY OF LEEDS



COFRIEND Event Learning/Recognition Demo

Summary Table

Event Name	Interval

XML data reading ....
XML data reading completed...

COFRIEND Event Learning/Recognition Demo, Krishna Dubba @ 03 Mar 2010

scksrd@leeds.ac.uk

Results

Event	# examples	Learned rules		Hand-crafted rules	
		precision	recall	precision	recall
FWD_CN_LoadingUnloading_Operation	5	**0.71**	0.3	0.04	**0.6**
GPU_Positioning	4	**1**	0.2	0.02	**0.5**
Aircraft_Arrival	15	**0.15**	0.06	0.04	0.06
AFT_Bulk_LoadingUnloading_Operation	12	**0.83**	**0.11**	0.04	0.03
Left_Refuelling	6	**0.38**	**0.5**	0	0
PB_Positioning	15	**0.25**	**0.5**	0.09	0.2
Aircraft_Departure	10	**0.33**	**0.14**	0	0
AFT_CN_LoadingUnloading_Operation	7	**0.54**	**0.4**	0.05	0.27
PBB_Positioning	15	**0.92**	0.05	0.07	**0.37**
FWD_Bulk_LoadingUnloading_Operation	3	1	**1**	1	0.02

UNIVERSITY OF LEEDS

Summary/novelty

- From pixels to symbolic, relational, qualitative behaviour/event descriptions
- Minimal supervision
- Multiple objects, shared objects, multiple simultaneous events,
- Robust computation of qualitative relations via HMM
- Functional object categorisation through event analysis

See papers for related work discussion

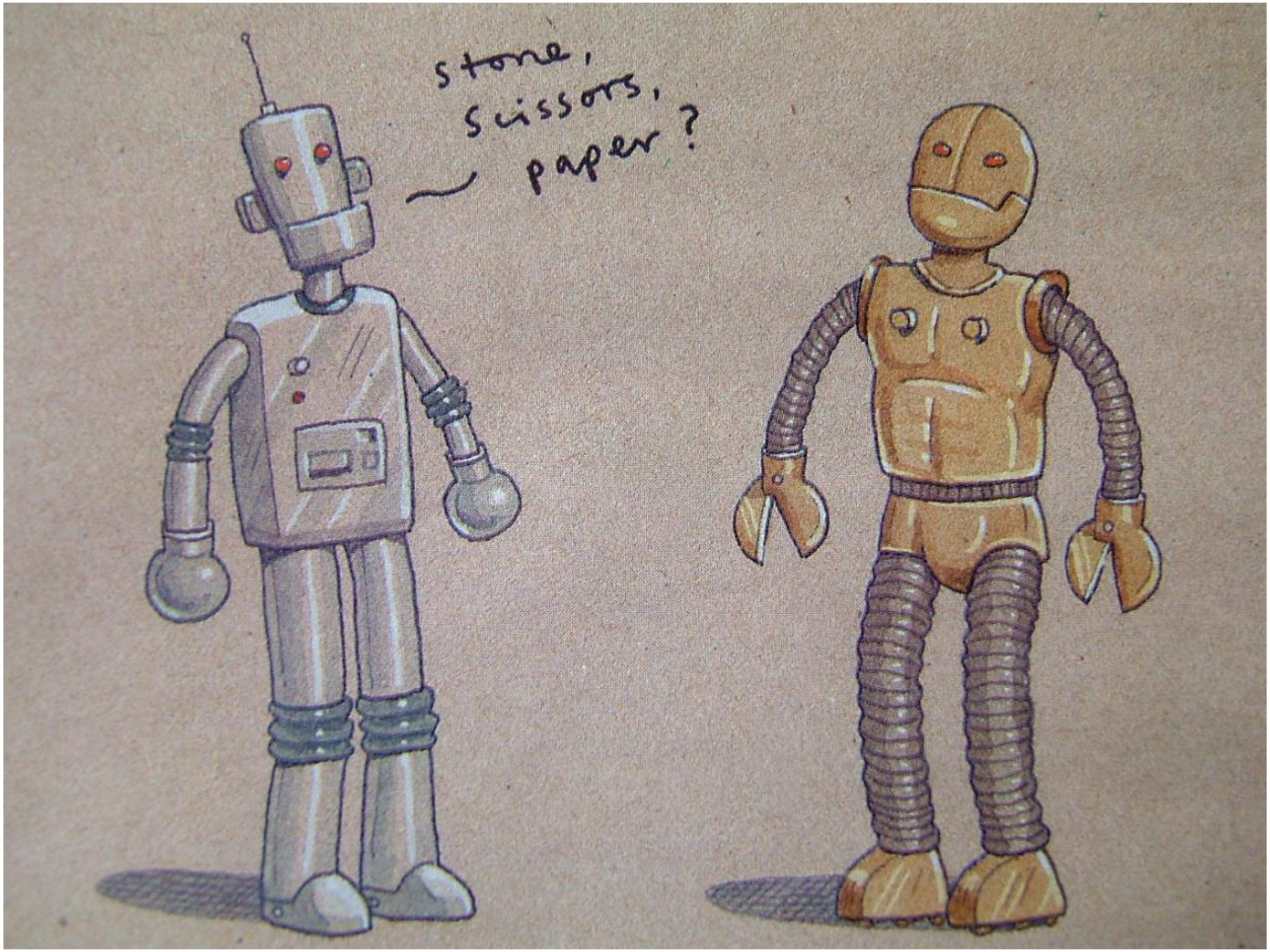
www.comp.leeds.ac.uk/qsr/publications.html

Research challenges/ongoing work

- New domains, longer scenes
  - Cognito project: learning workflows
  - Mind’s Eye project: spatio-temporal semantics of verbs
- Further experimentation with different sets of spatial relations
- Use induced functional categories to supervise appearance learning
- Learning probabilistic weights for rules (MLN)
- Interleaving induction/abduction to mitigate noise
- Cognitive evaluation of event classes and functional categories
- Learning a global model
  - temporal sequencing of event classes
- Online learning and Ontology alignment
- Language (+ vision)
- …

UNIVERSITY OF LEEDS

Any Questions?



Thanks to:

EPSRC, EU (CoFriend, Cognito), DARPA (Mindseye/Vigil)

David Hogg,

Krishna Sridhar,

Sandeep Dubba,

QSR and CV groups at Leeds