School of Computing

UNIVERSITY OF LEEDS

Building Qualitative Models of Spatio-Temporal Behaviour

Tony Cohn

What does an agent need to know about the world?

UNIVERSITY OF LEEDS

- What kind of objects there are.
- What they do/can be used for.
- What kinds of actions and events there are.
- Which objects participate in which actions/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
  - Qualitative spatio-temporal relations

This approach has a long history…

UNIVERSITY OF LEEDS

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

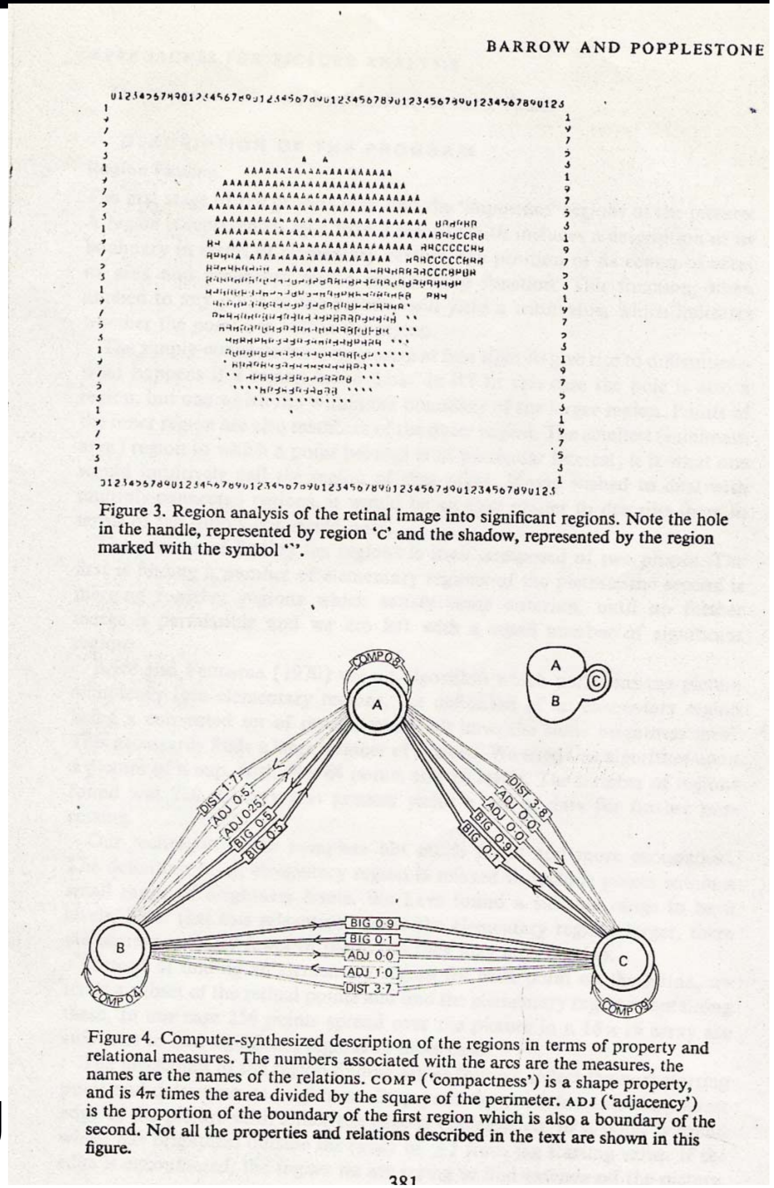


Figure 3. Region analysis of the retinal image into significant regions. Note the hole in the handle, represented by region 'C' and the shadow, represented by the region marked with the symbol '''.

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.

381

…with an interesting conclusion

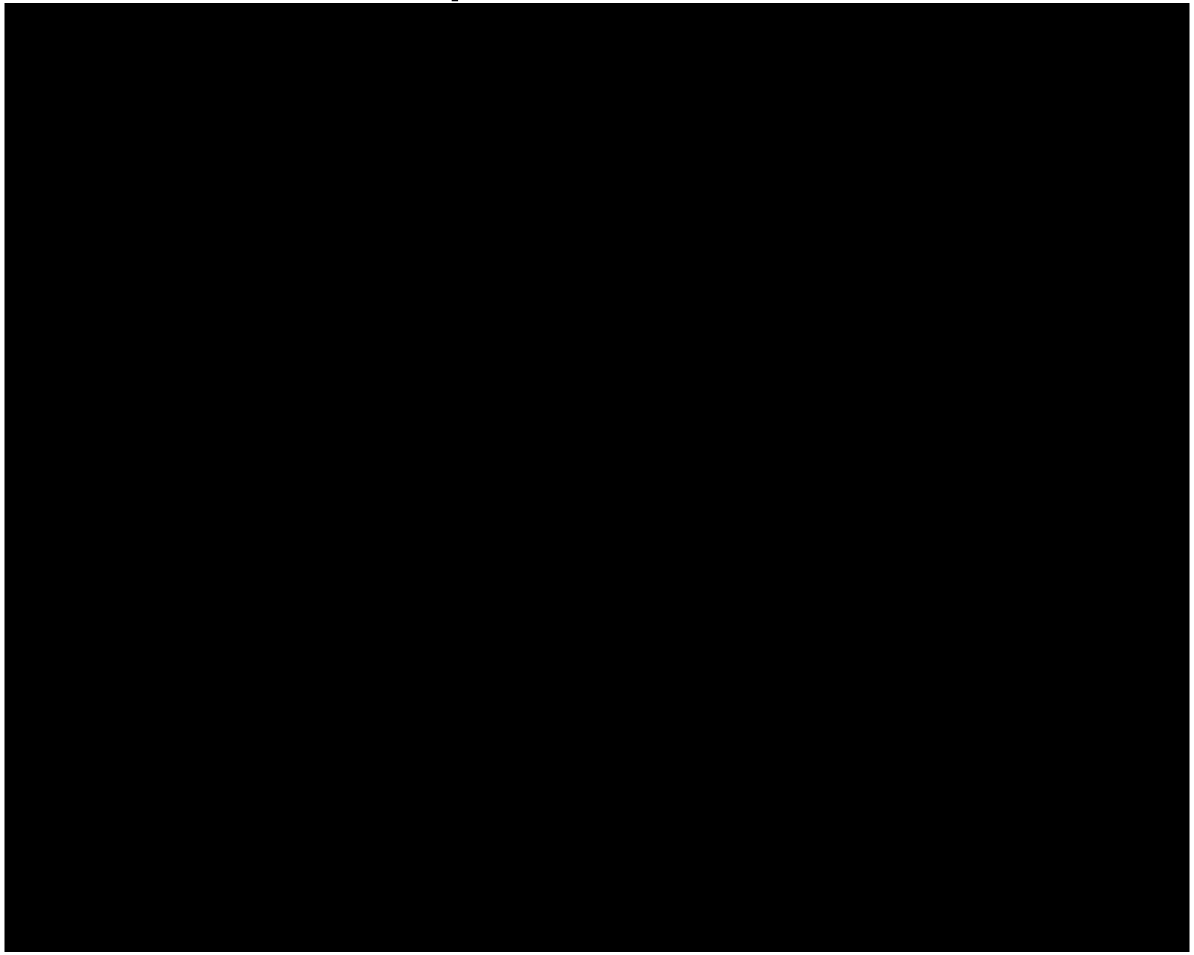
UNIVERSITY OF LEEDS

'…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 ***<u>planning</u>*****, 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

UNIVERSITY OF LEEDS

- 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

A ***brief*** tour of qualitative s-t languages/reasoning

UNIVERSITY OF LEEDS

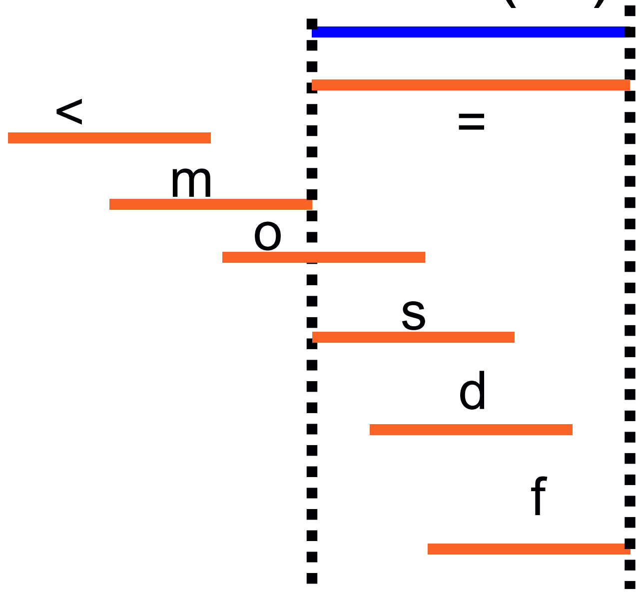
Sets of *Jointly Exhaustive and Pairwise Disjoint* (JEPD) relations

- Temporal – ~3 calculi
- Spatial – 100's of calculi
- Spatio-temporal – some calculi

Qualitative temporal representations

UNIVERSITY OF LEEDS

- Vilain's & Kautz's point algebra -- 3 JEPD relations
- Allen's interval calculus (IA) -- 13 JEPD relations

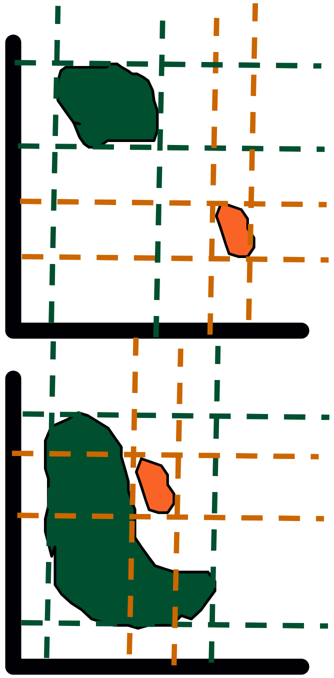


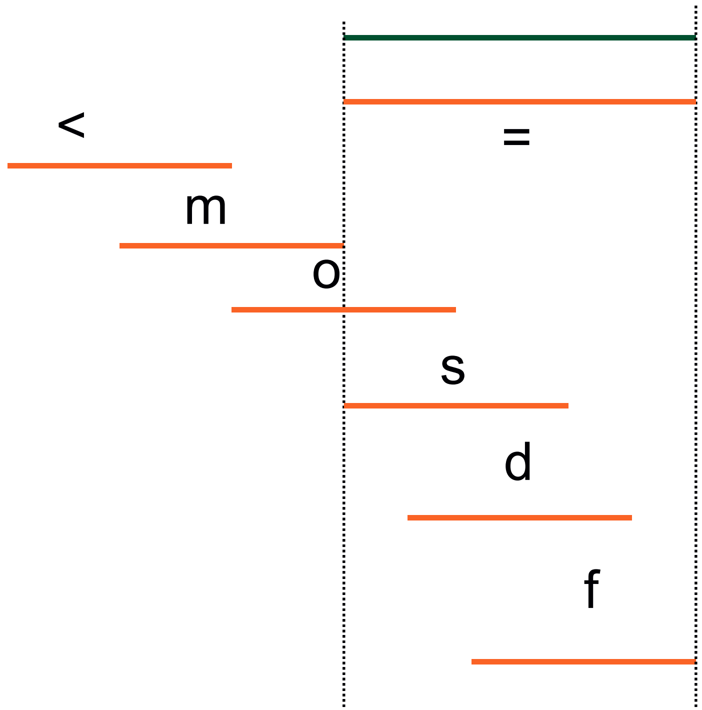
- INDU calculus (intervals with durations)
  - IA x PA = 25 JEPD relations

Richness of QSR derives from multi-dimensionality

UNIVERSITY OF LEEDS

- Consider trying to apply Allen's temporal interval calculus in 2D (*rectangle algebra*):





- Can work well for particular domains -- e.g. envelope/address recognition (Walischewski 97)
- Extension to INDU + ranking of intervals – CORE-9 (KR-12)

Aspects of QSR

- ontology, topology, orientation, distance, shape...
- spatial change
- Vagueness and uncertainty
- reasoning mechanisms
- pure space v. domain dependent

Ontology of Space

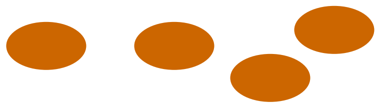
extended entities (regions)?

points, lines, boundaries?

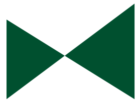
mixed dimension entities?

Open/closed/regular/non regular regions?

Multi-piece (disconnected)?



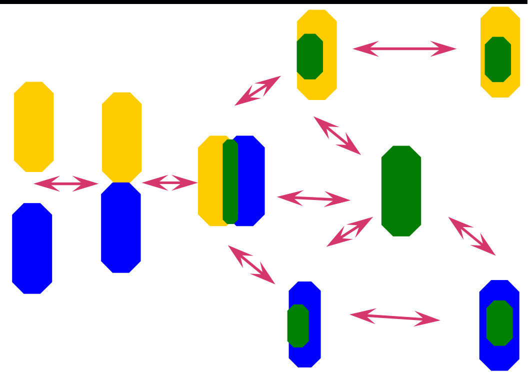
Interior connected?



What is the embedding space?

- connected? discrete? dense? dimension? Euclidean?...

What entities and relations do we take as primitive, and what are defined from these primitives?



Region Connection Calculus (RCC8)

- Defined from primitive: C(*x*,*y*)

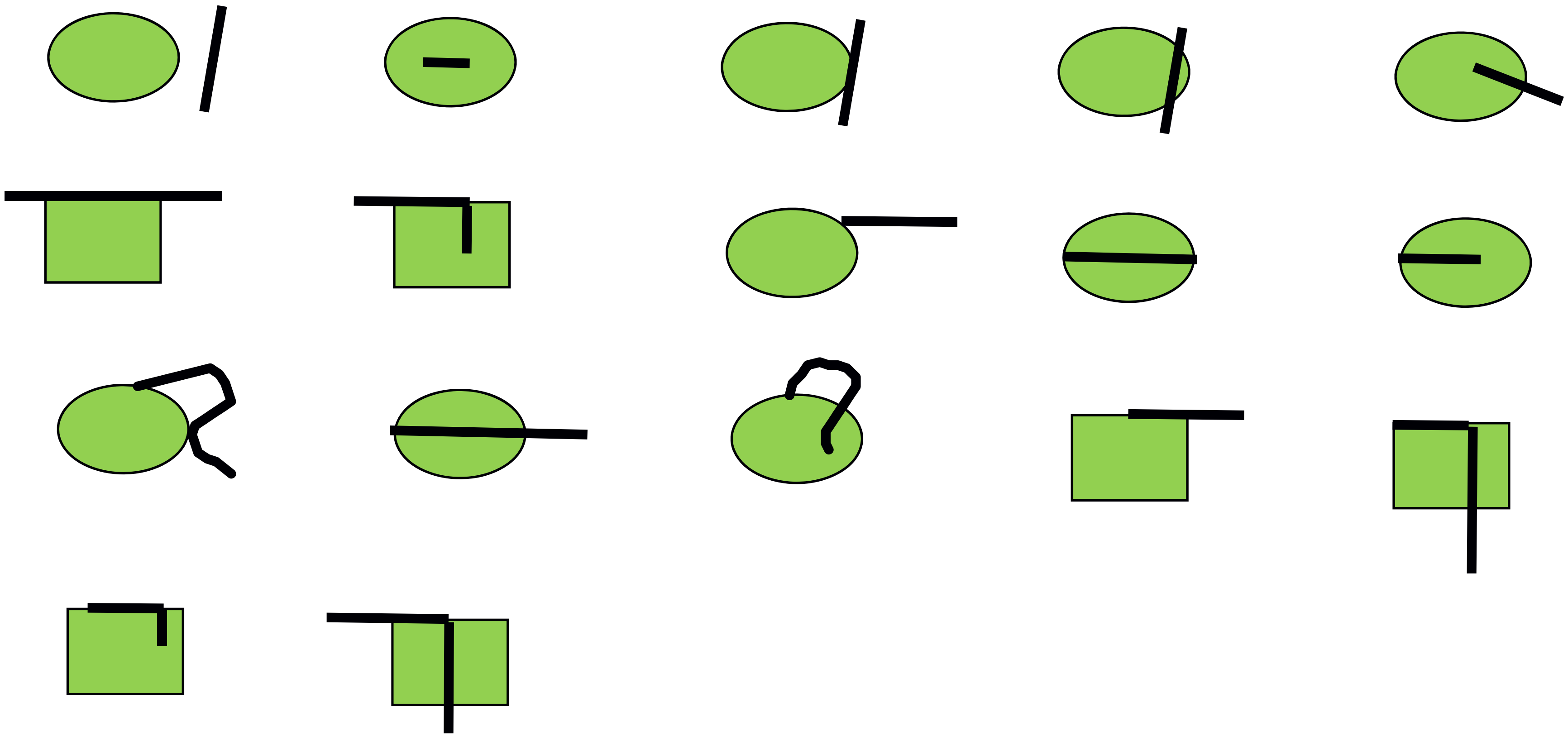
Alternatively: 9-intersection model

$\cap$	boundary(*y*)	interior(*y*)	exterior(*y*)
boundary(*x*)	$\neg$	$\emptyset$	$\neg$
interior(*x*)	$\emptyset$	$\emptyset$	$\emptyset$
exterior(*x*)	$\neg$	$\emptyset$	$\neg$

- $2^9 = 512$ combinations
  - 8 relations assuming planar regular point sets
- potentially more expressive

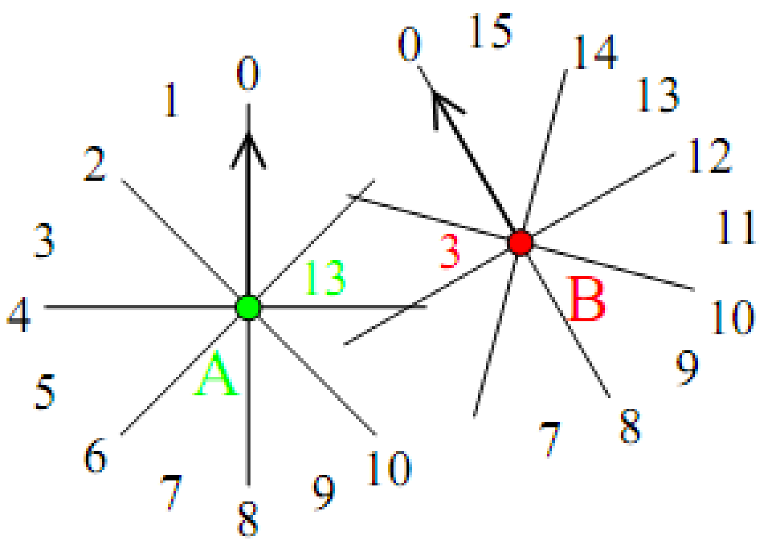
Extension of 9IM: replace ¬ with dimension of the intersection

The 17 different L/A relations of the DEM



Direction calculi: Point based

E.g. Oriented Point Algebra (OPRA)



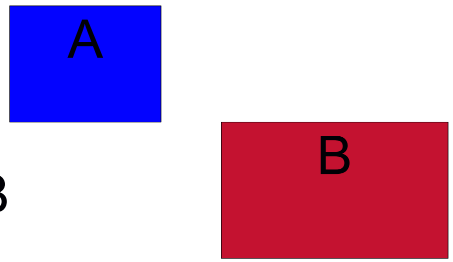
relation is: A (13,3) B

Direction calculi: region based

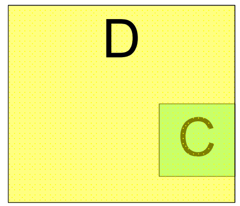
UNIVERSITY OF LEEDS

E.g. use Rectangle Algebra (IA x IA)

- Take MBRs of regions



A (>,m) B



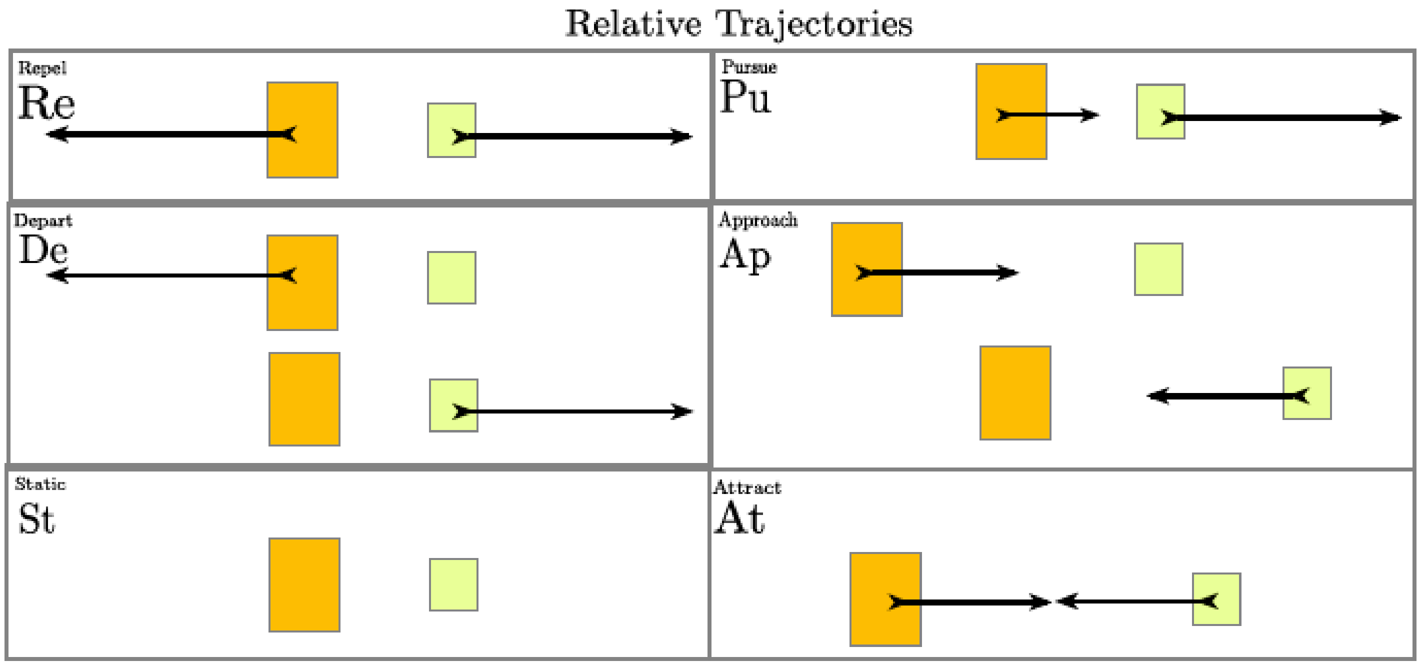
C (f,d) D

- Note possibility of “internal directions”: London is in SE of UK
- Extension: use INDU not RA
    - E.g. Ireland is to the West of England and smaller

Qualitative Trajectory Calculus

UNIVERSITY OF LEEDS

Record whether two objects moving towards –) or away (+) from each other:



Can also record relative speed

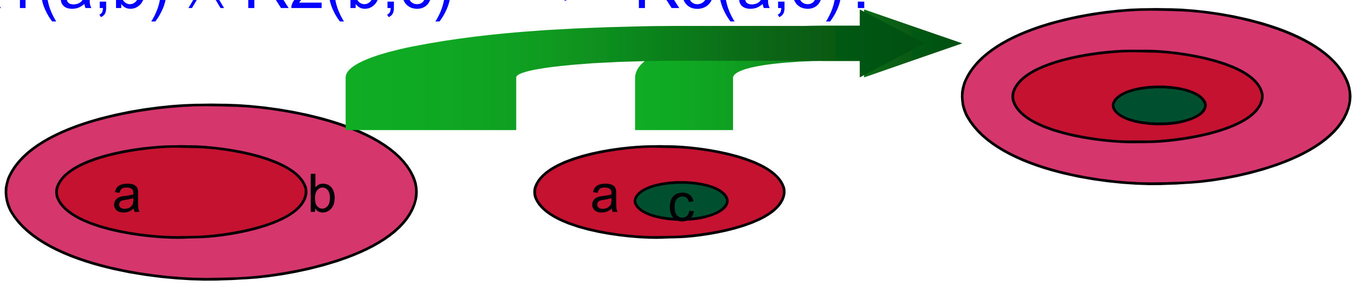
Reasoning

First order mereotopology is undecidable

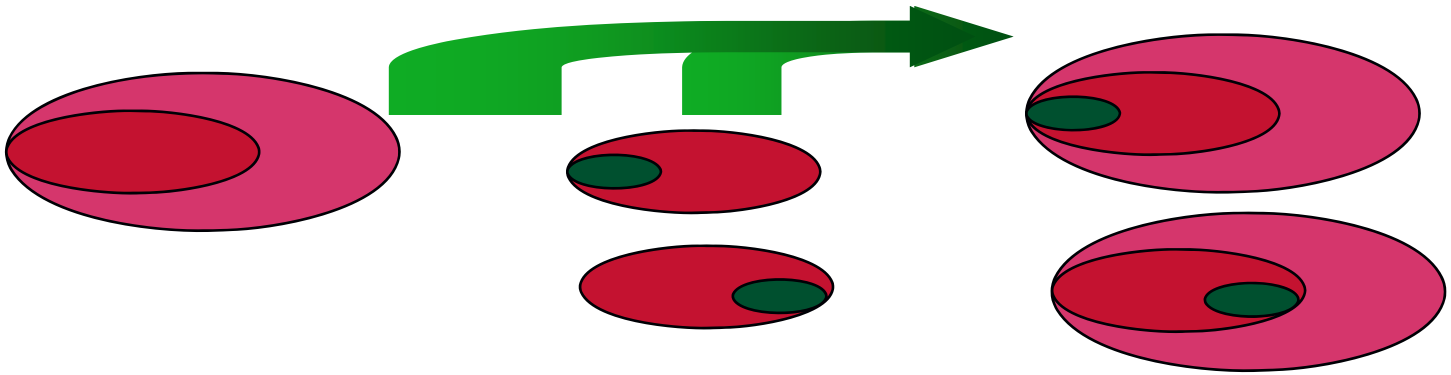
Decidable subtheories, e.g. constraint languages (RCC-8)

Composition based reasoning

R1(a,b) ∧ R2(b,c) => R3(a,c)?



- In general R3 is a disjunction



Tractability of consistency

- Under certain circumstances using composition repeatedly can decide consistency
- Precompute compostion tables
- Non-tractable in general (disjunctive constraints)
- Isolation of maximal tractable subsets
  - automatic procedure (Renz)

Decidable Spatiotemporal modal logics (Wolter & Zakharyashev)

UNIVERSITY OF LEEDS

Combine point based temporal logic with RCC8Combine RCC, Boolean operations on regions and temporal modal logic

- temporal operators: **S**ince**, U**ntil

  - Next (O), Always in the future □+, Sometime in the future ◊+

- Decidable: satisfiability is PSPACE complete

E.g. **¬** □+P(Kosovo,Yugoslavia)

Kosovo will not always be part of Yugoslavia

E.g. EQ(Germany, □+Eurozone)

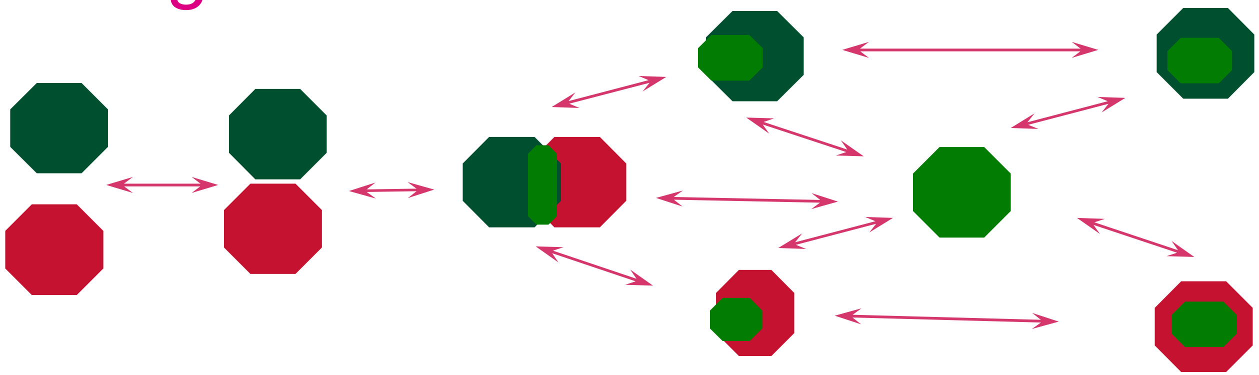
Only Germany will use the Euro for ever

E.g. O EQ(UK,GB+N.Ireland)

Tomorrow the UK will (still) be composed of GB + N.Ireland

Continuity Networks/ Conceptual Neighbourhoods

- What are next qualitative relations if entities transform/translate continuously?
  - E.g. RCC-8



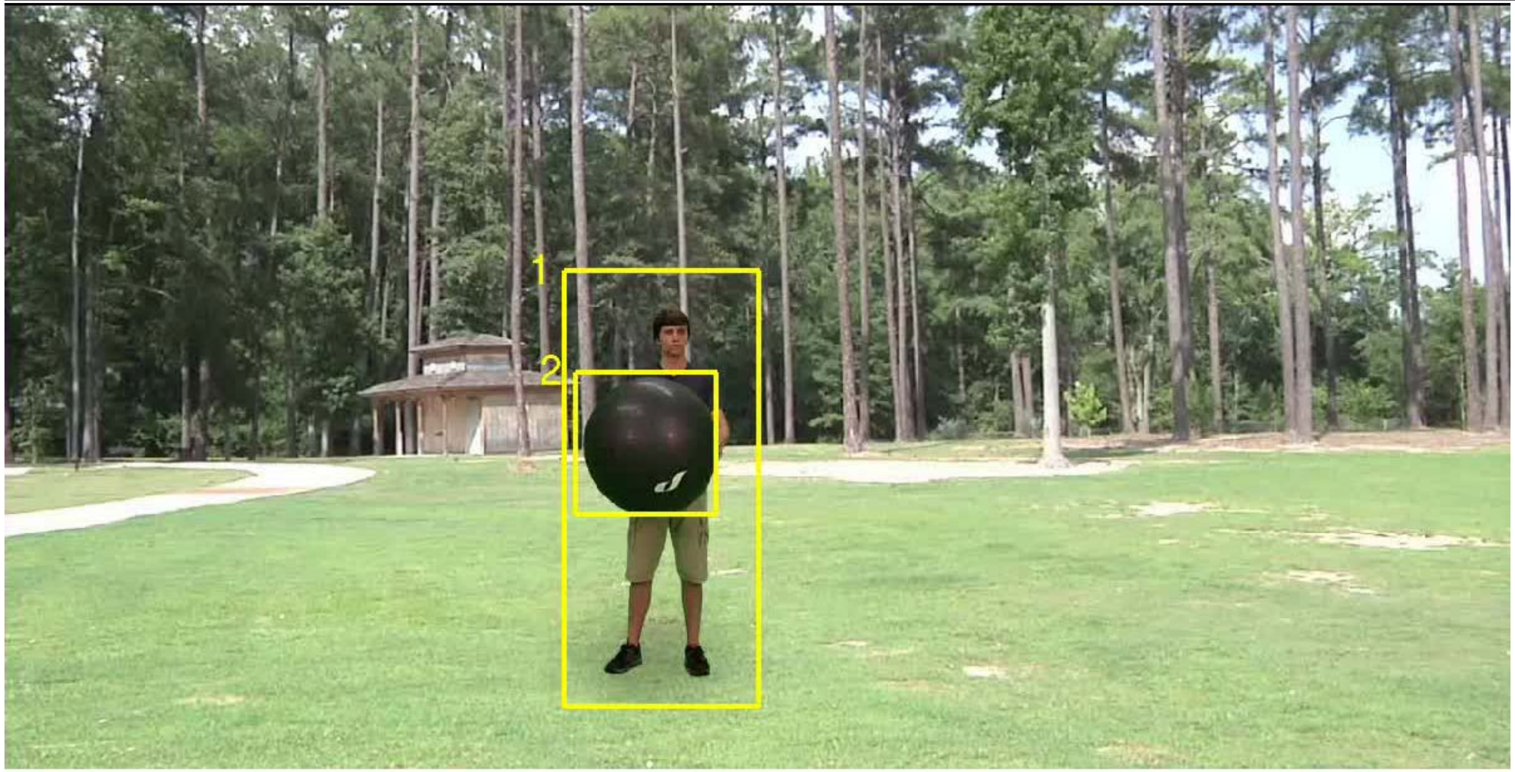
- If uncertain about the relation what are the next most likely possibilities?

  Uncertainty of precise relation will result in connected subgraph (Freksa 91)

- Can be used for qualitative simulation algorithm

From video to QSR:
Using an HMM to 'smooth' relations

UNIVERSITY OF LEEDS

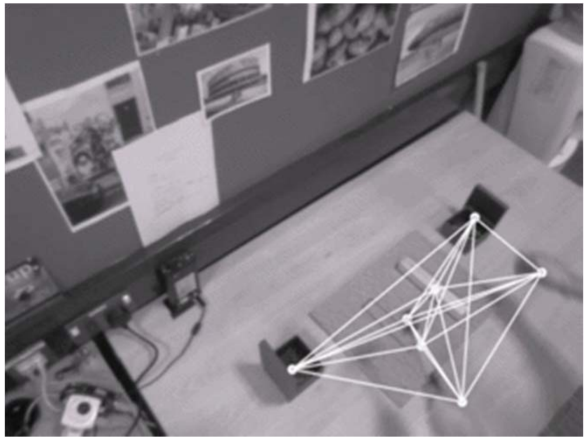


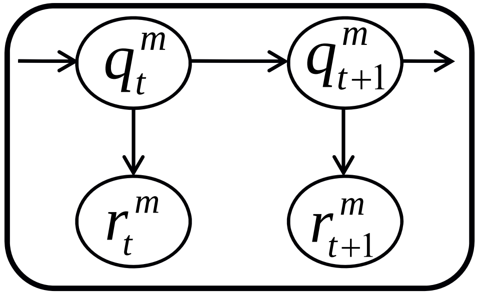
PSI 2 1 NTPP

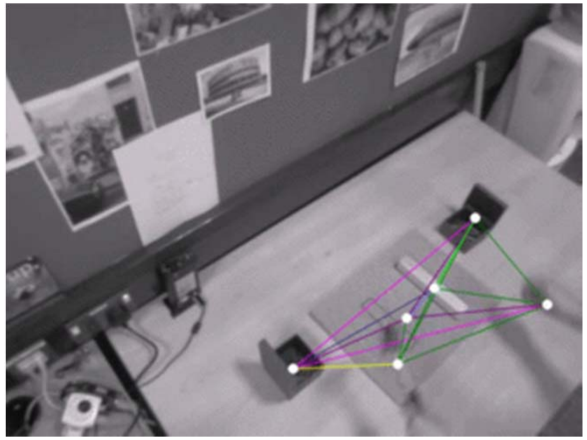
RCC8 2 1 NTPP

Sridhar et al.,
COSIT 2011
(best paper)

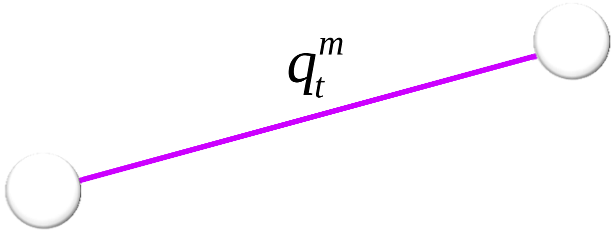
Learning relations







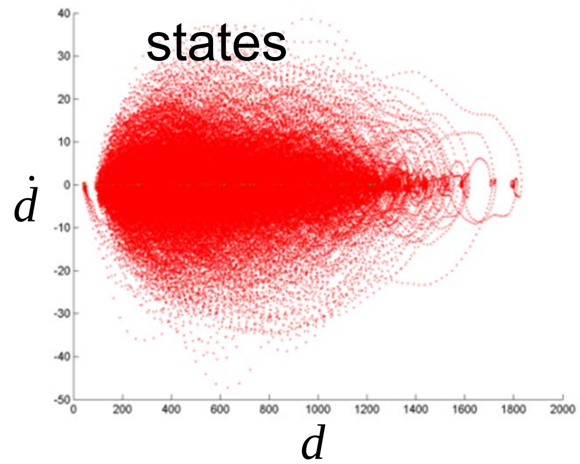
$$r_t^m = (d_t^m, \dot{d}_t^m)$$



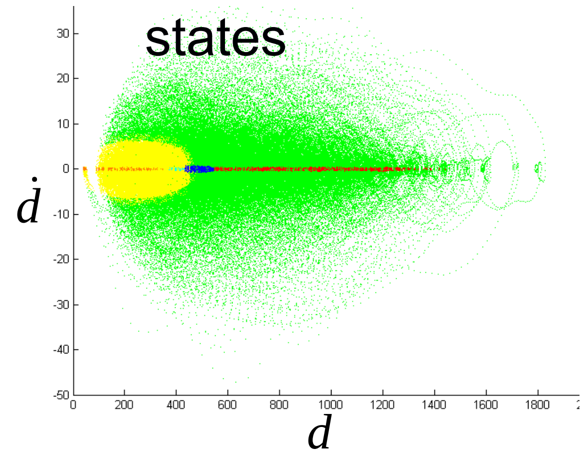
Quantisation of Relational Features

UNIVERSITY OF LEEDS

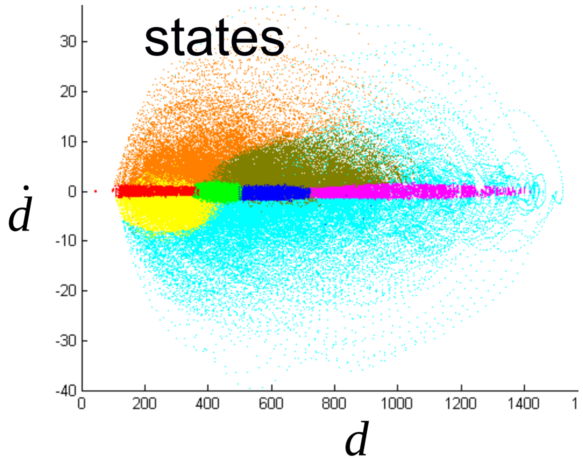
2 discrete states



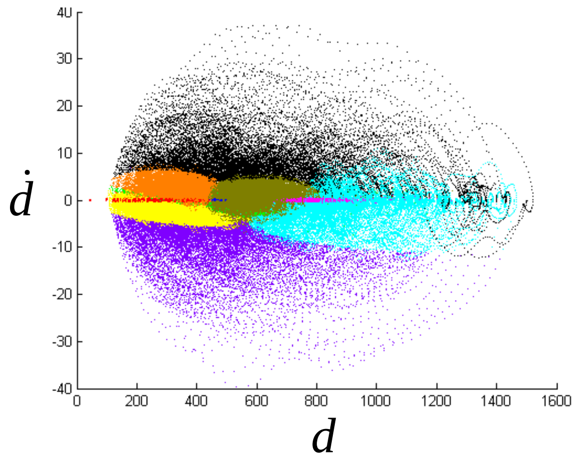
6 discrete states



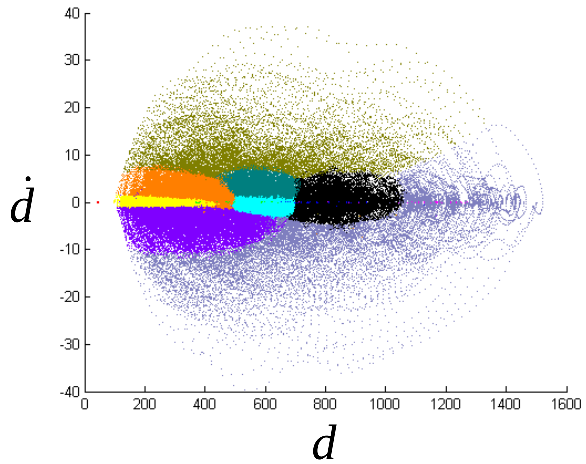
8 discrete states



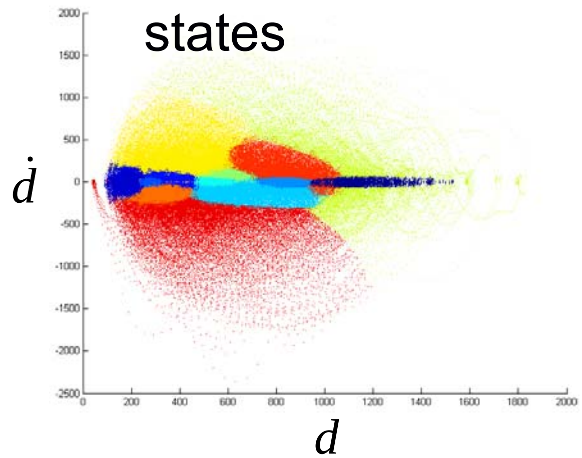
10 discrete states



12 discrete states



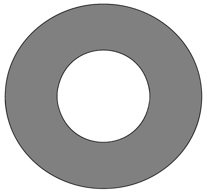
16 discrete states



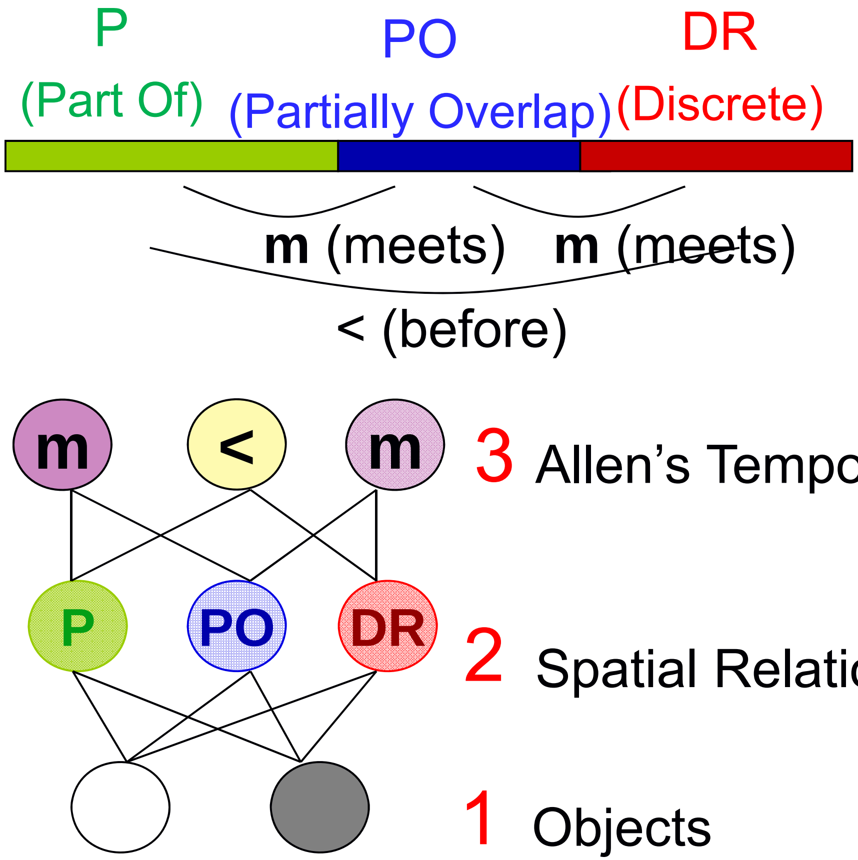
Use a Bayesian Information Criterion to optimize number of states/relations

Representing interactions relationally

UNIVERSITY OF LEEDS



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



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)

UNIVERSITY OF 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)**

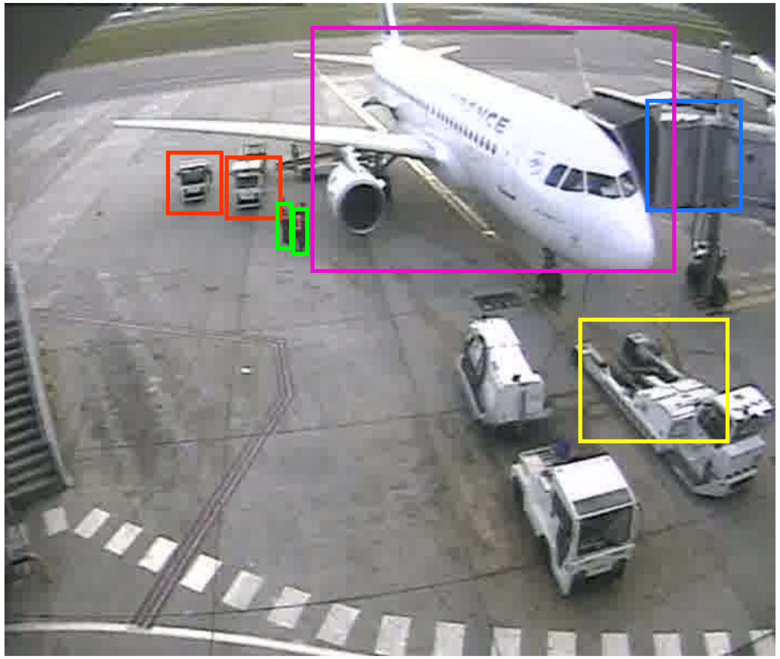
Functional Object Categories from a Relational Spatio-Temporal Representation (ECAI'08)

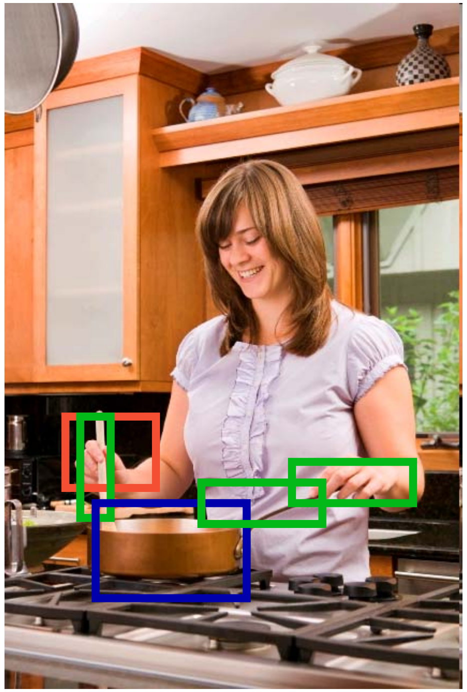
Workflow Activity Monitoring using the Dynamics of Pair-wise Qualitative Spatial Relations (MMM-12) …

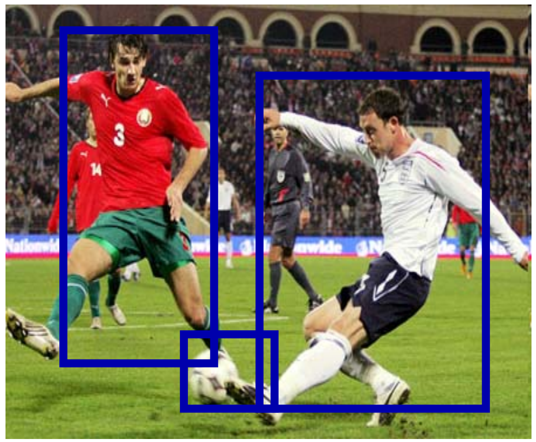
Problem: Understanding Activities

UNIVERSITY OF LEEDS

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







IC

Problem: Understanding Activities

UNIVERSITY OF LEEDS

- 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

UNIVERSITY OF LEEDS

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

UNIVERSITY OF LEEDS

- 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*

UNIVERSITY OF LEEDS

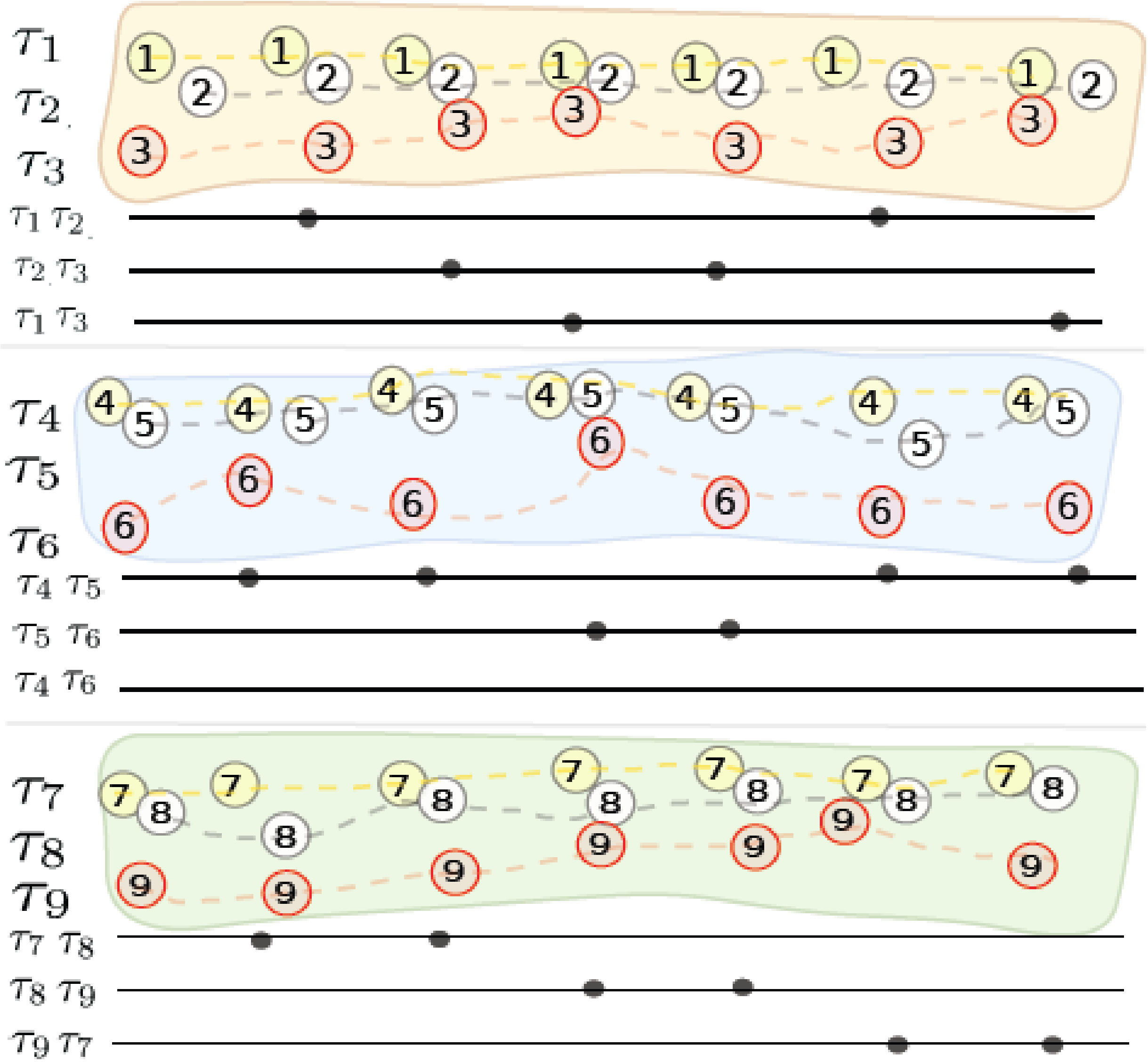
All objects interact evenly over interval

~9 doesn't interact much

X

Interactions temporally extended

X



Unsupervised Learning Summary:

UNIVERSITY OF LEEDS

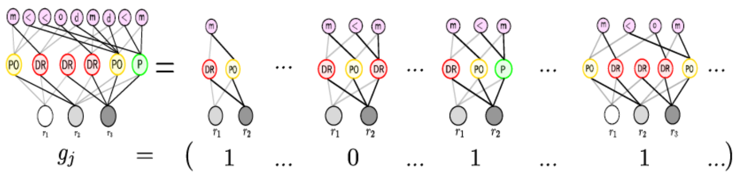
• Event classes specified non-parametrically by set of event graphs

• 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 (but *coincidences still possible)*
- minimizes number of event classes
- has event classes all of whose graphs are similar

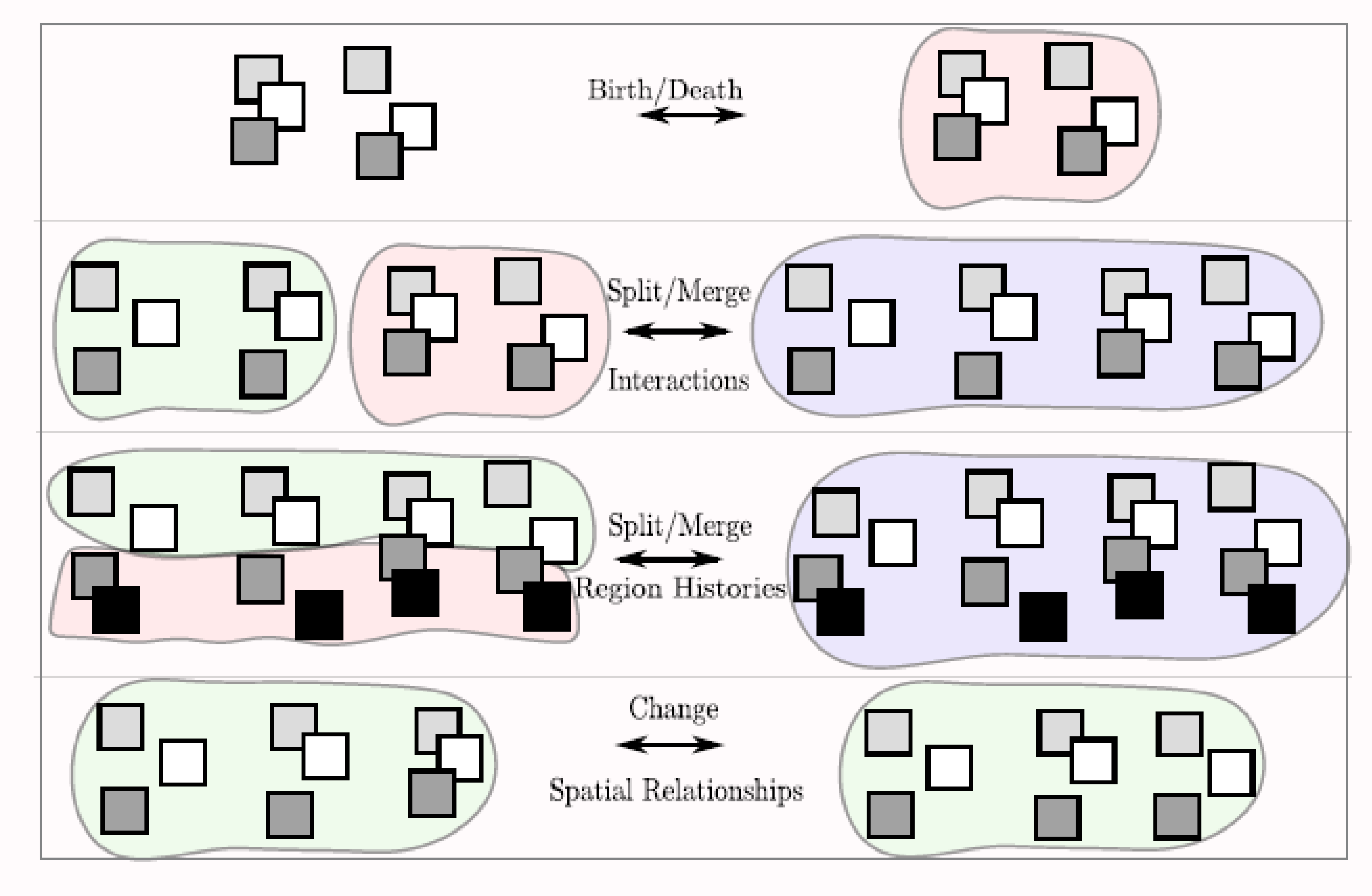
*(similarity computed with **bag-of-graphlets** representation)*



- has events which have high degree of tracklet interaction

and *low* object sharing with other events

MCMC moves


Birth/Death
Split/Merge
Interactions
Split/Merge
Region Histories
Change
Spatial Relationships

Evaluation in aircraft domain

UNIVERSITY OF LEEDS

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

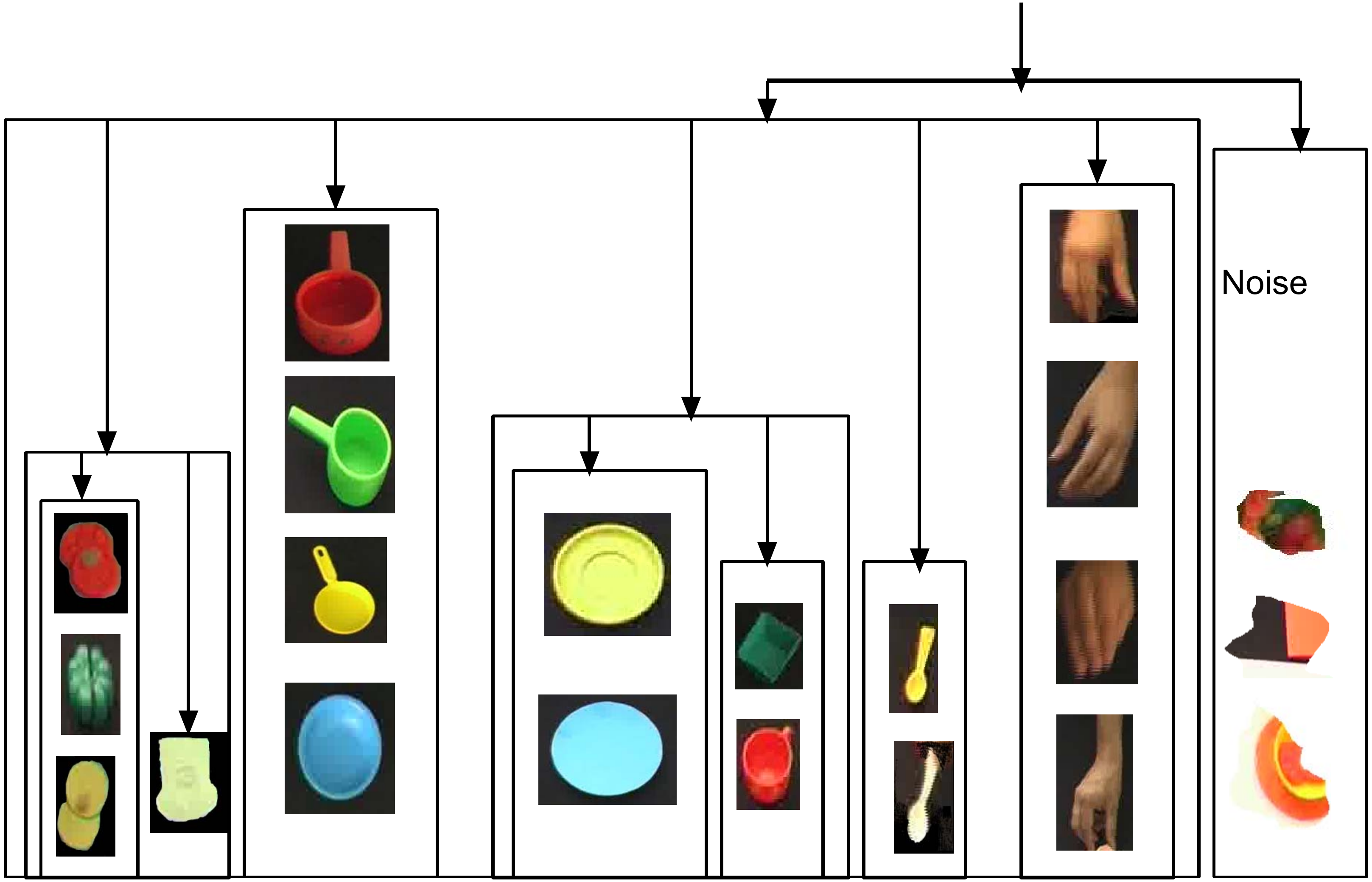
Objects

	E_1			E_2		$\cdots$	E_m			
	(•	•	•)	(•	•)		(•	•	•	•)
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

- Compress the rows (pattern for each object) using PCA
- Obtain object taxonomy by hierarchical-clustering of compressed rows

Emergent object classes: *Toy kitchen hierarchy*

UNIVERSITY OF LEEDS



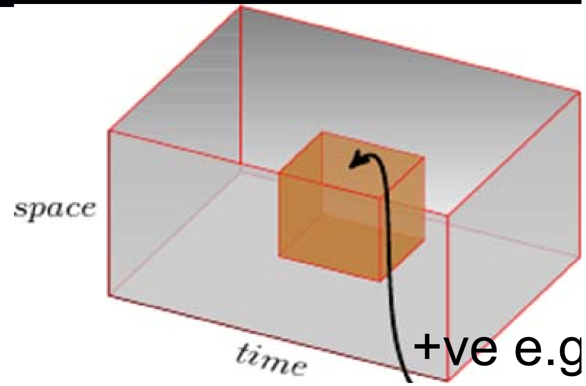
Emergent object classes: *Aircraft domain*

UNIVERSITY OF LEEDS

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?

UNIVERSITY OF LEEDS

- 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:

$$Event(\ldots) \leftarrow Condition_1(\ldots) \wedge \ldots \wedge Condition_n(\ldots)$$

- 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:

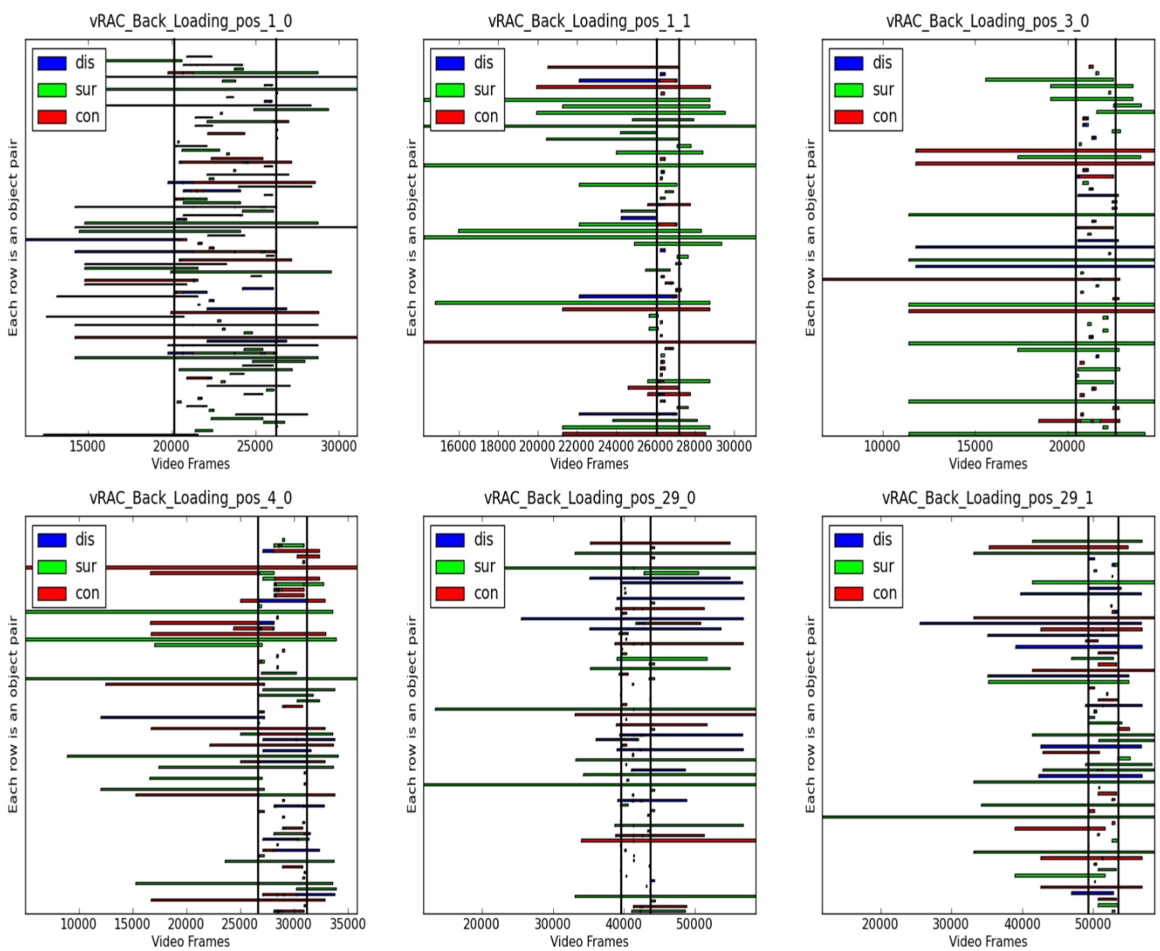
  - improve efficiency

  - reduce overgeneralisation from noisy examples

“Learning from Interpretations” setting

UNIVERSITY OF LEEDS

Each positive example is represented as a separate Database



Search Strategy

UNIVERSITY OF LEEDS

Search the hypothesis lattice for a model that maximizes

α*positives covered – β*negatives covered – #variables

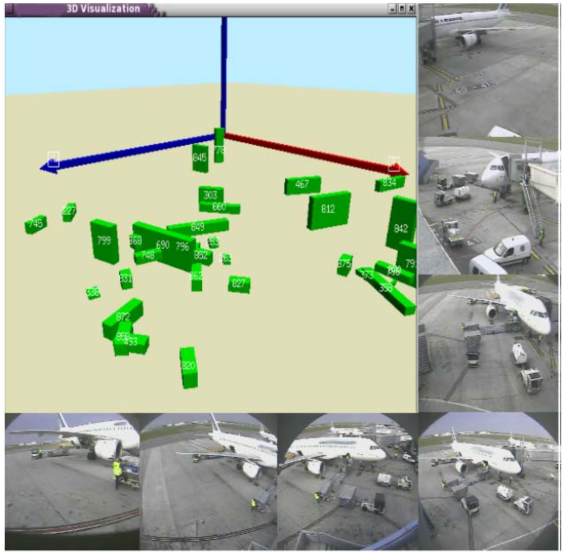
subject to generic s-t constraints, e.g.:

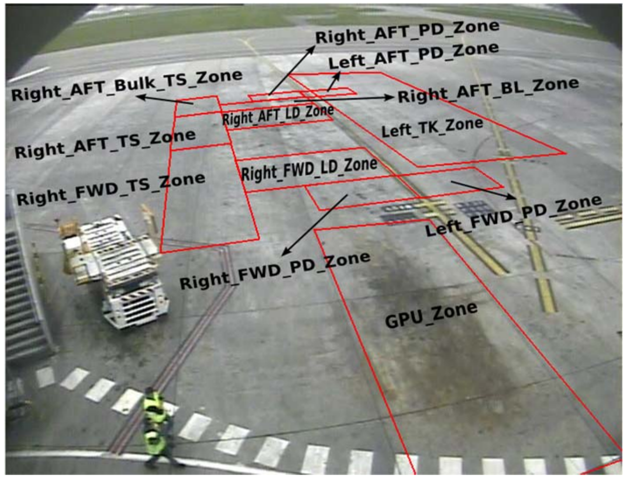
- Hypothesis should not have only temporal predicates.
- All intervals in temporal predicates should be present in some spatial predicate

UNIVERSITY OF LEEDS

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

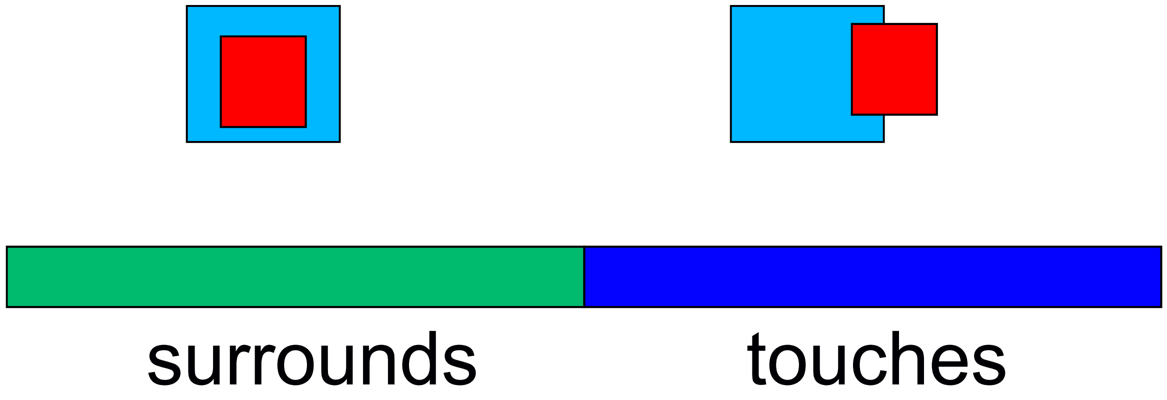




A Learned Event Model:

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

Results

UNIVERSITY OF LEEDS

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

Summary/novelty

- Many QSR calculi available
- 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

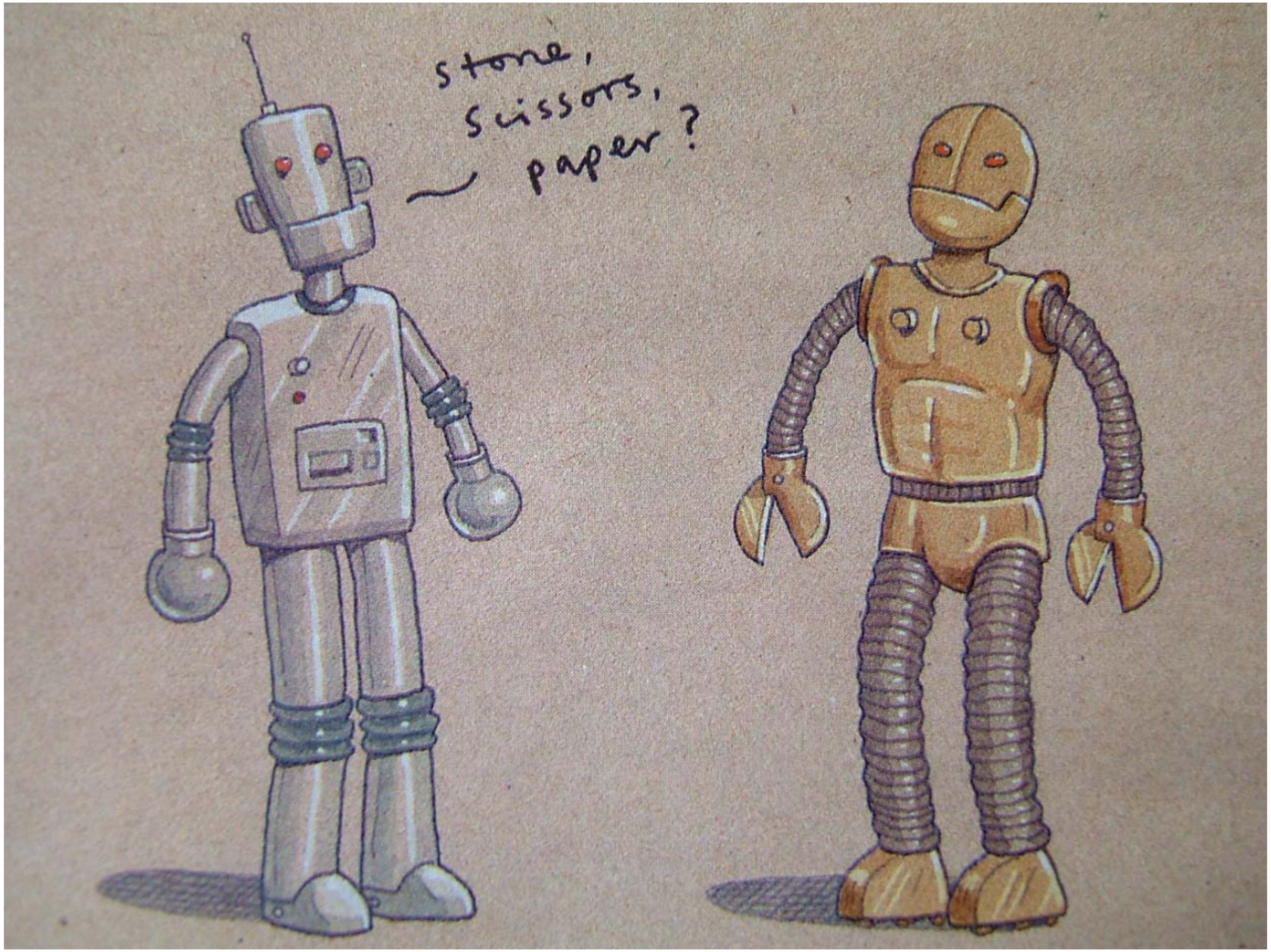
UNIVERSITY OF LEEDS

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)
- …

Any Questions?

UNIVERSITY OF LEEDS



Thanks to:

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

David Hogg,
Krishna Sridhar,
Sandeep Dubba,
QSR and CV groups at Leeds