School of Computing

UNIVERSITY OF LEEDS

Qualitative Spatial Representations for *Activity Recognition*

Tony Cohn

STRANDS Summer School, Lincoln, August 2015

Once upon a time …

UNIVERSITY OF LEEDS

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

Relational descriptions of object classes + supervised learning

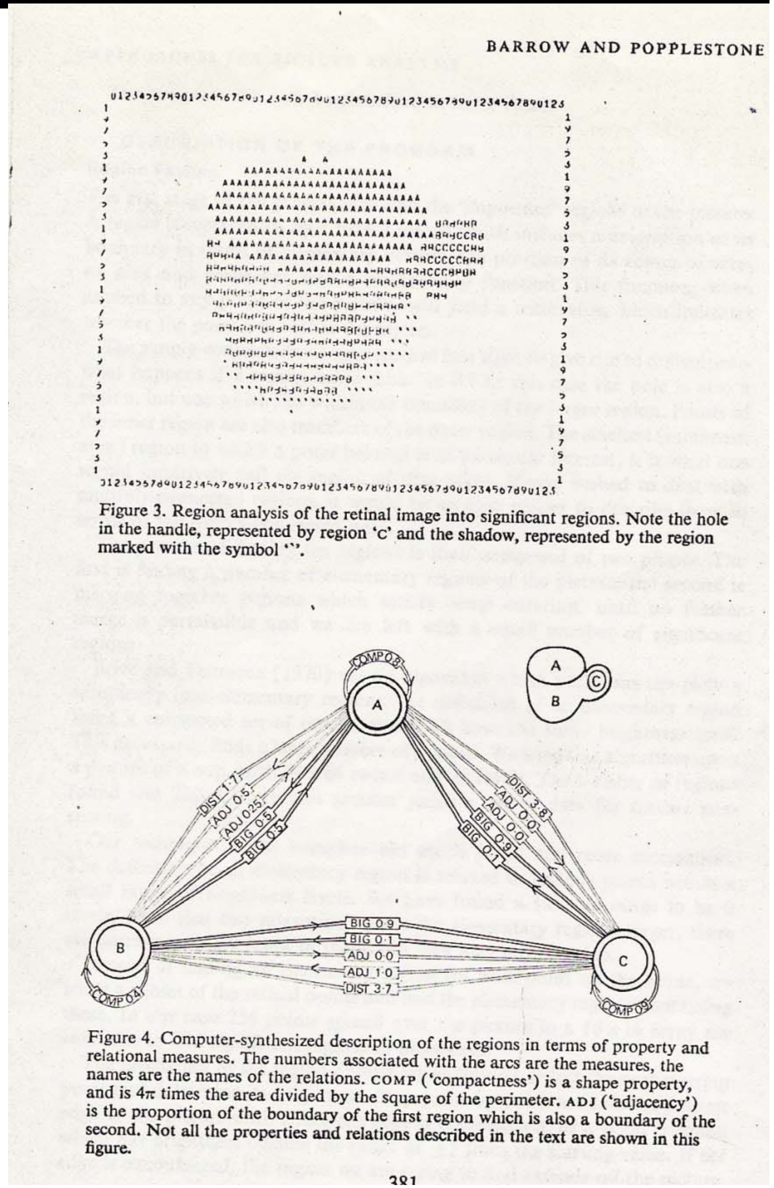

BARROW AND POPPLESTONE

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.

Over the decades

UNIVERSITY OF LEEDS

Artificial Intelligence

KR

Planning

ML

NLP

Computer Vision

. . .

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

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, **activity recognition**, 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

- relations may be taken as primitives, or defined in terms of other primitives

- in general consider disjunctions of basic relations too

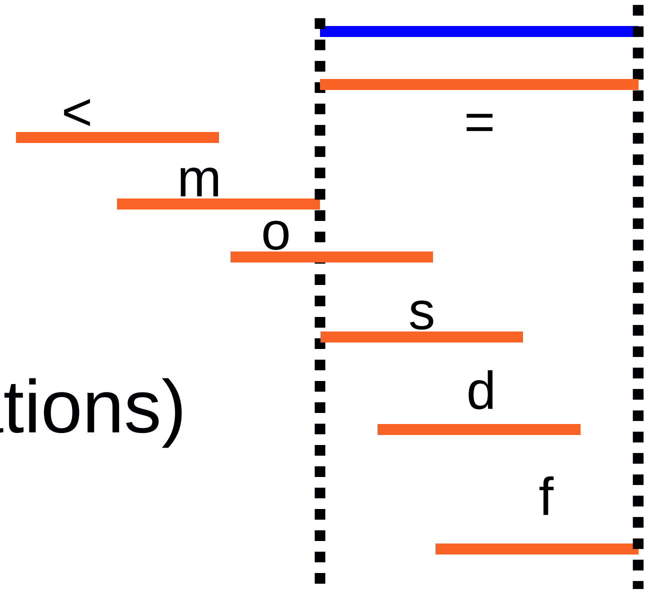
Qualitative temporal representations

UNIVERSITY OF LEEDS

- Vilain's & Kautz's point algebra -- 3 JEPD relations

  - Between temporal points (<,=,>)

- Allen’s interval calculus (IA) -- 13 JEPD relations



- INDU calculus (**in**tervals with **du**rations)

  – IA x PA = 25 JEPD relations

  <,m,o and inverses are split as to whether intervals are smaller (<), =, or larger (>)

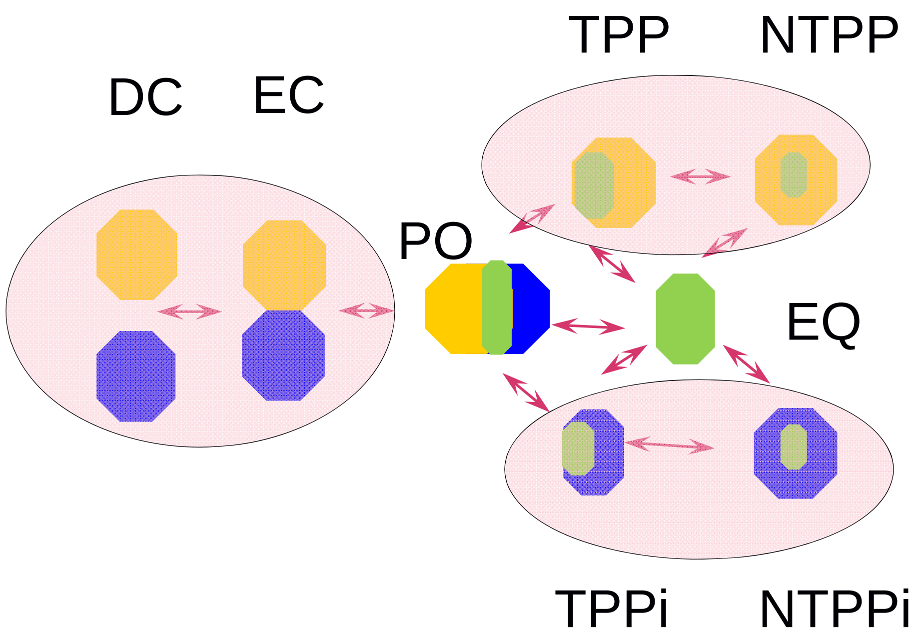
Qualitative spatial representations

UNIVERSITY OF LEEDS

Region Connection Calculus (RCC8)

- (mereo)topology
- definable from a primitive C(x,y)

Arrows indicate *conceptual neighbourhood:* continuous transitions



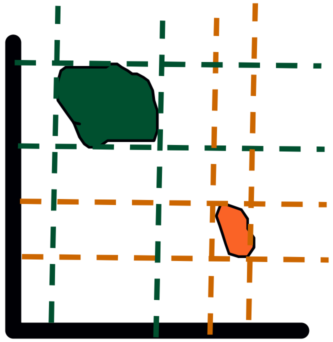
Simplification **RCC5** (tangential distinctions hard to make in practice in vision)

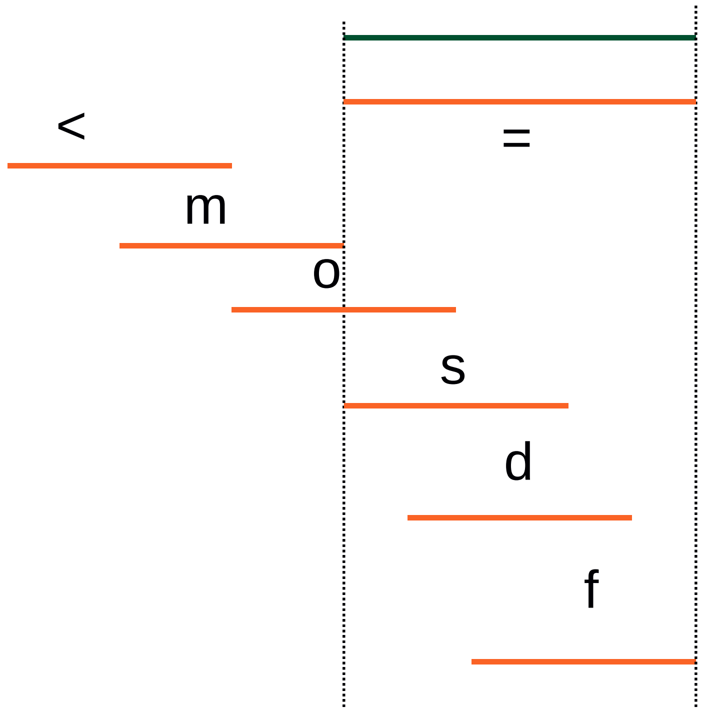
RCC doesn’t distinguish dimensionality

A 2D spatial calculus:
Rectangle Algebra:
combining topology and direction

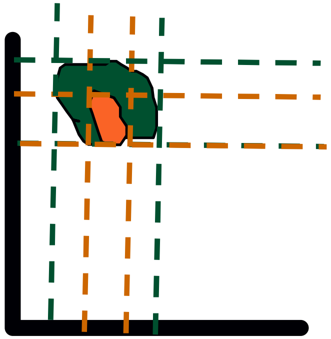
UNIVERSITY OF LEEDS

Apply Allen's interval calculus in 2D (*rectangle algebra: 13*13=169 relations*):





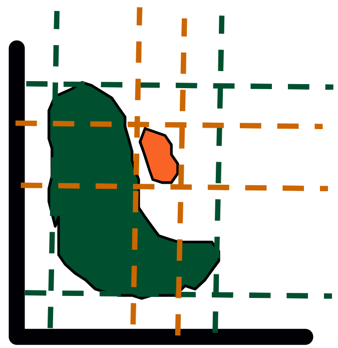
- E.g. Orange is SE of Green (>,<) above

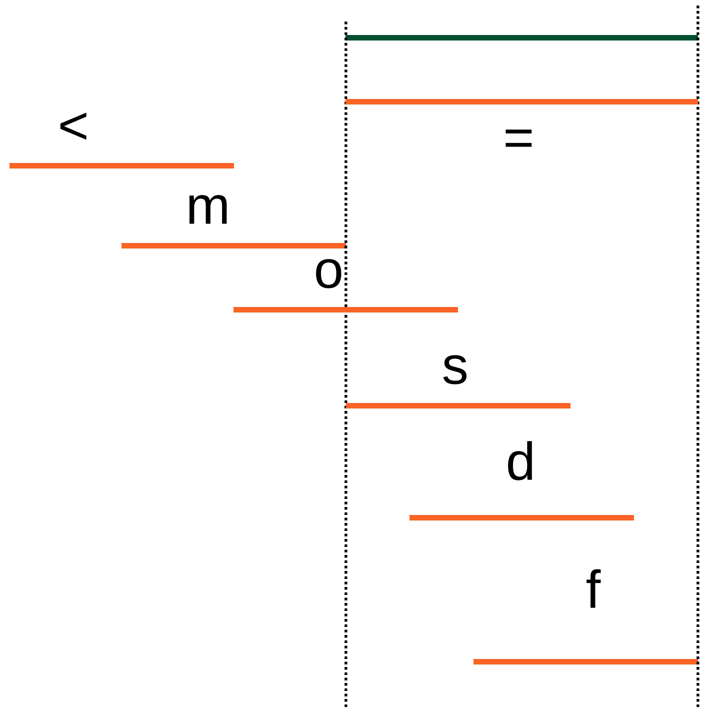


- E.g. Orange is part of Green and touches southern border (>,<) above

RA and non convex regions

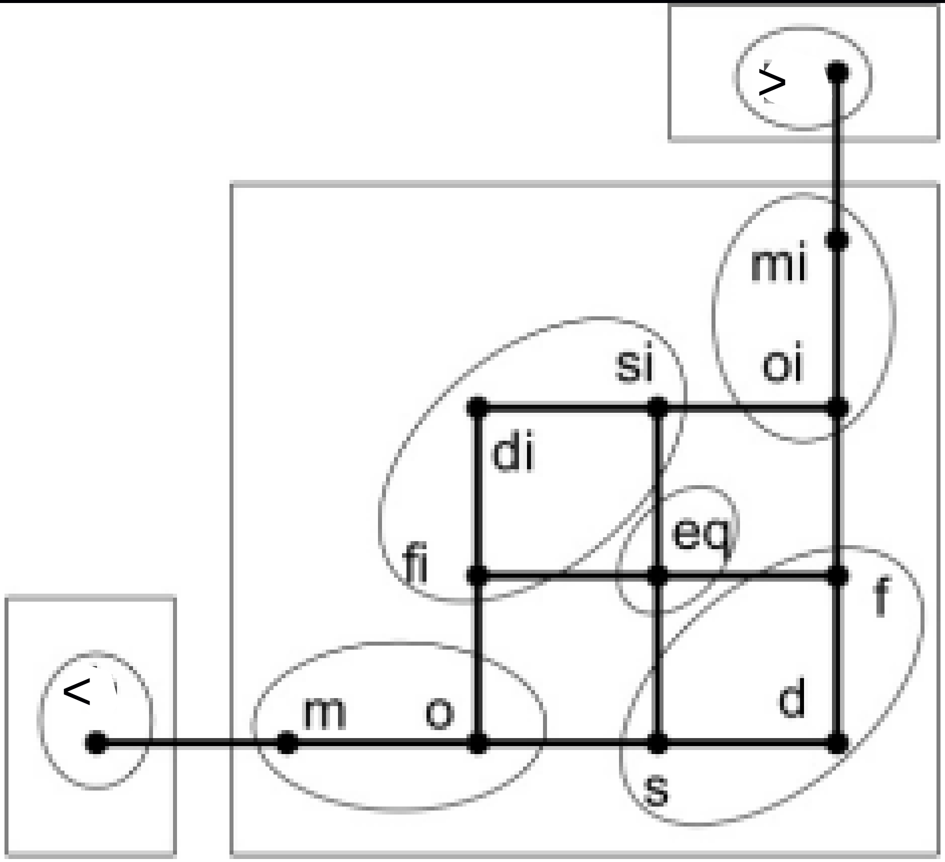
RA doesn't work so well for non convex regions:





13:35

Simplifications of the RA



DIR9 = IA3 x IA3

DIR49 = IA7 x IA7

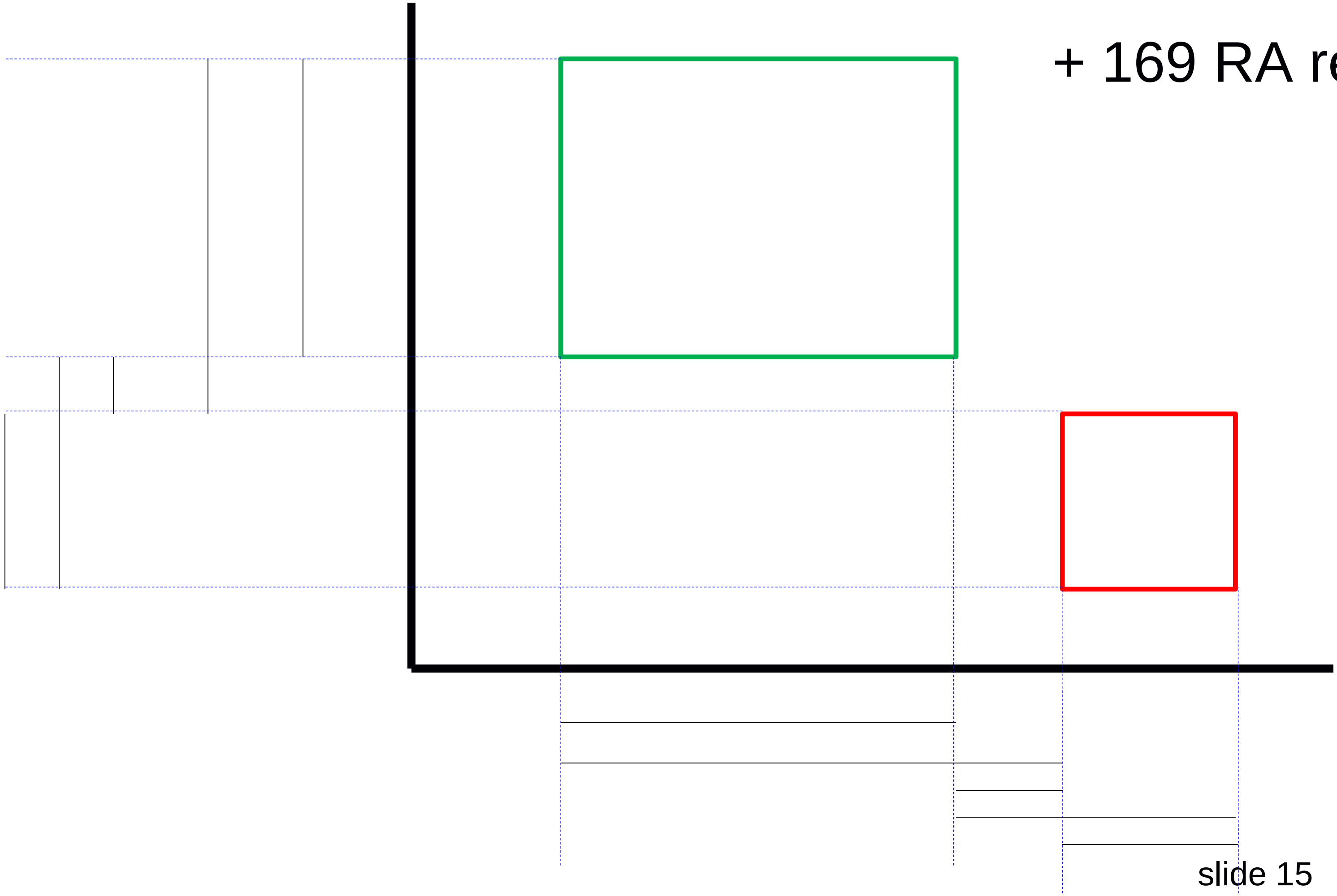
The conceptual neighbourhood graph of IA, where ellipses (boxes, resp.) represent basic relations in IA7 (IA3 , resp.).

CORE-9

2D version of INDU: up to 6 intervals on each axis

Can compare each of them pairwise – 66 possible relations

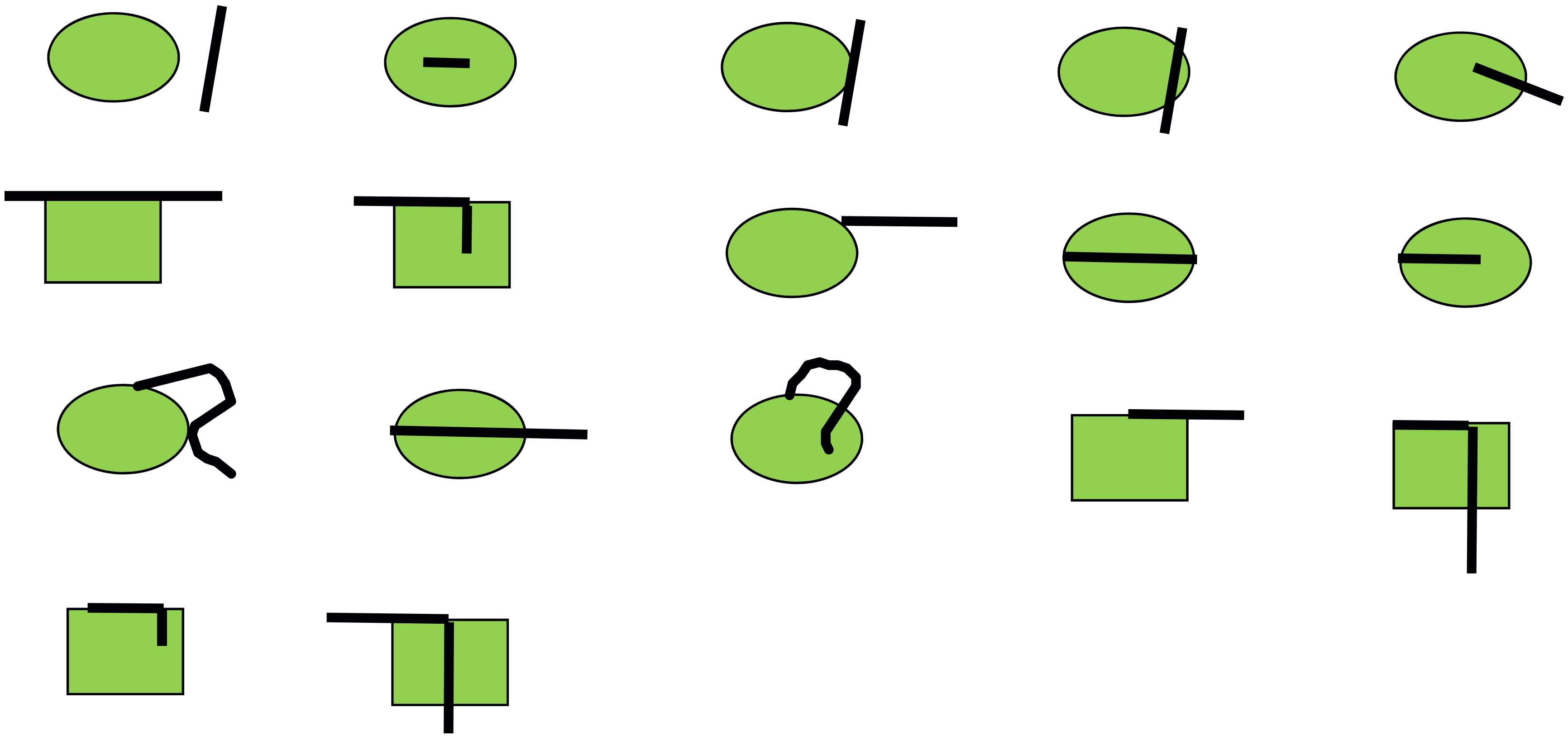
+ 169 RA relations



The 17 different L/A relations of the DEM (Dimension Extended Method)

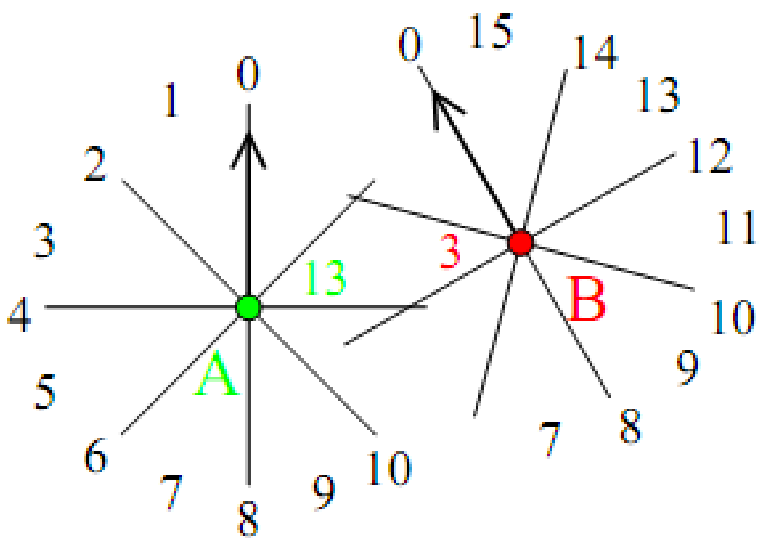
UNIVERSITY OF LEEDS

The 17 different L/A relations of the DEM



Direction calculi: Point based

E.g. Oriented Point Algebra (OPRA)



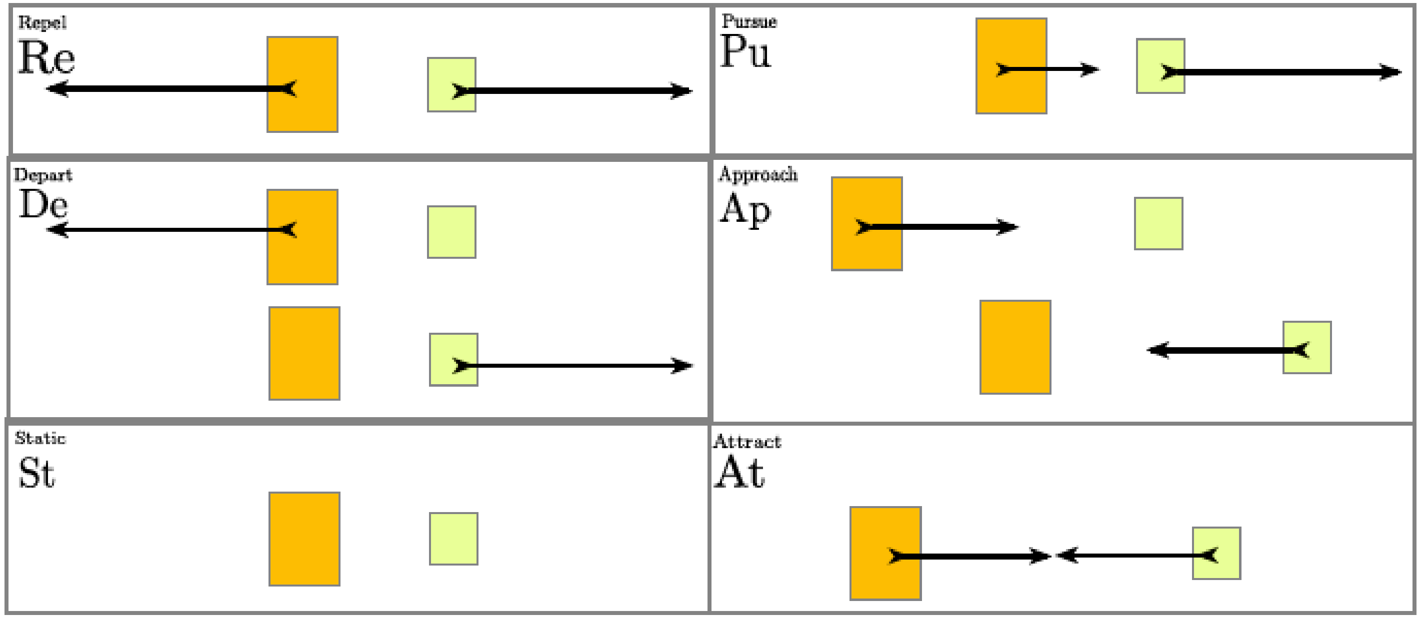
relation is: A (13,3) B

Qualitative Trajectory Calculus (QTC)

UNIVERSITY OF LEEDS

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

Relative Trajectories



- Can also record relative speed (faster +, slower -)
- Other QTC calculi distinguish 2D motions,…

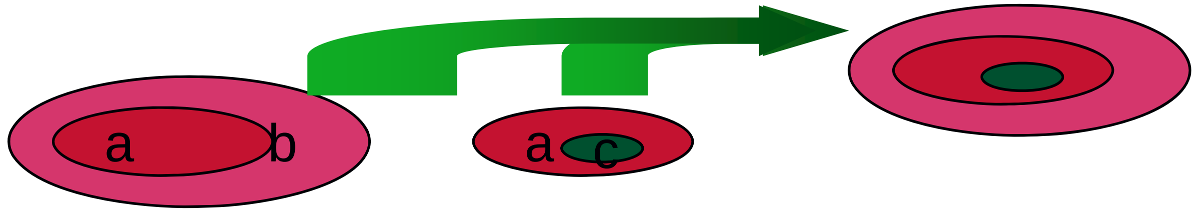
Reasoning

First order mereotopology is undecidable

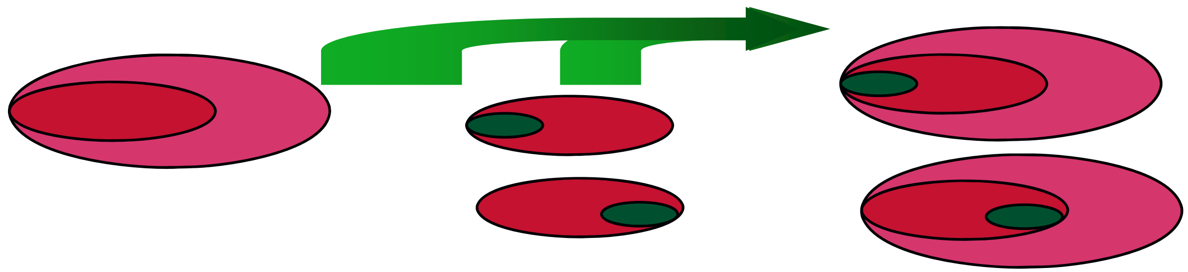
Decidable subtheories, e.g. constraint languages (RCC-8)

Composition based reasoning

$R1(a,b) \wedge R2(b,c) \quad \Rightarrow \quad R3(a,c)?$



In general R3 is a disjunction



Research has identified *tractable* subsets of constraint languages

QSTR and computer vision

UNIVERSITY OF LEEDS

Why might QSTR be useful in computer vision?

- Abstract away from noise
- Abstract away from variation in event performance
- Descriptions of activities can be given in a “cognitive” way

And some challenges:

•Noise (inaccurate/missing detections)

•A small quantitative change might yield a different qualitative relation

- But one that is close in the conceptual neighbourhood

- Which QSTRs and at what granularity (e.g. RCC3 vs RCC5)?
- “Combined” calcluli (e.g. INDU, CORE-9,…) are representationally efficient but make it harder to do “feature selection” in learning

A “paradox”

Qualitative *Representations* seem to be more useful than Qualitative *Reasoning (Deduction)*

I.e. QSTRs are a useful abstraction

But since the video provides a *model* of the qualitative knowledge base it is “by definition” consistent

- Reasoning can be useful when there is partial knowledge

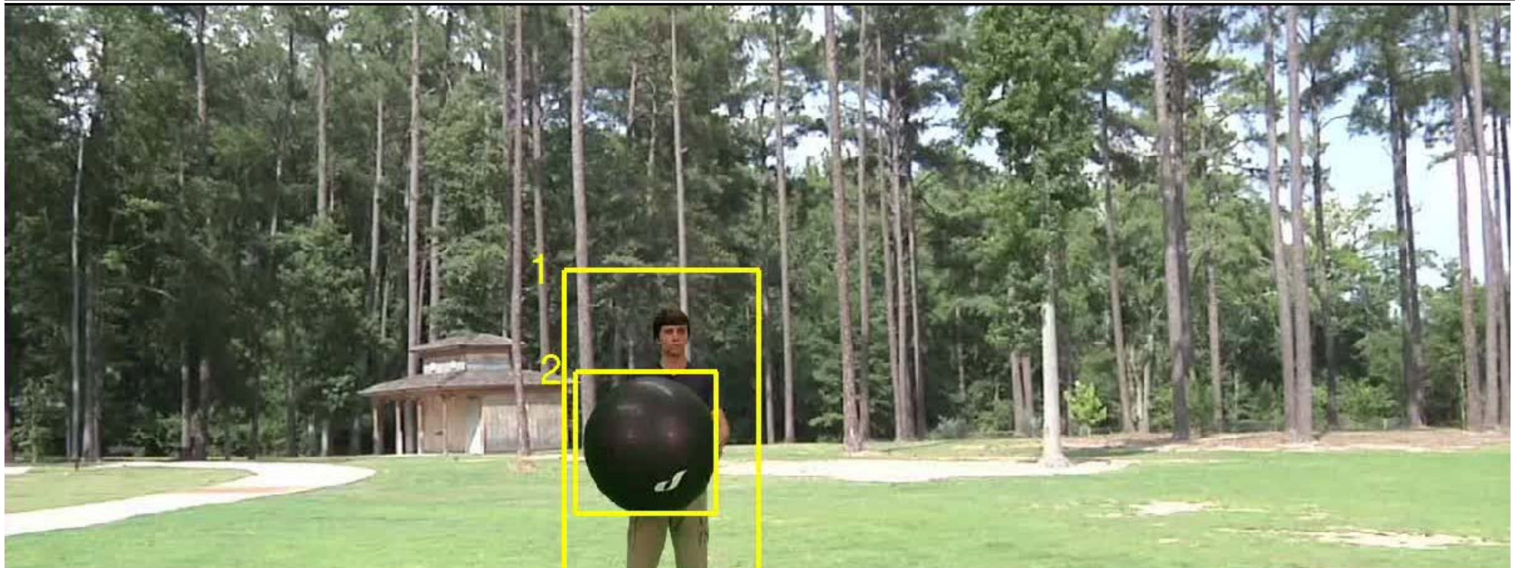
  (e.g. occlusions)

- Reasoning can be useful when there are multiple knowledge sources

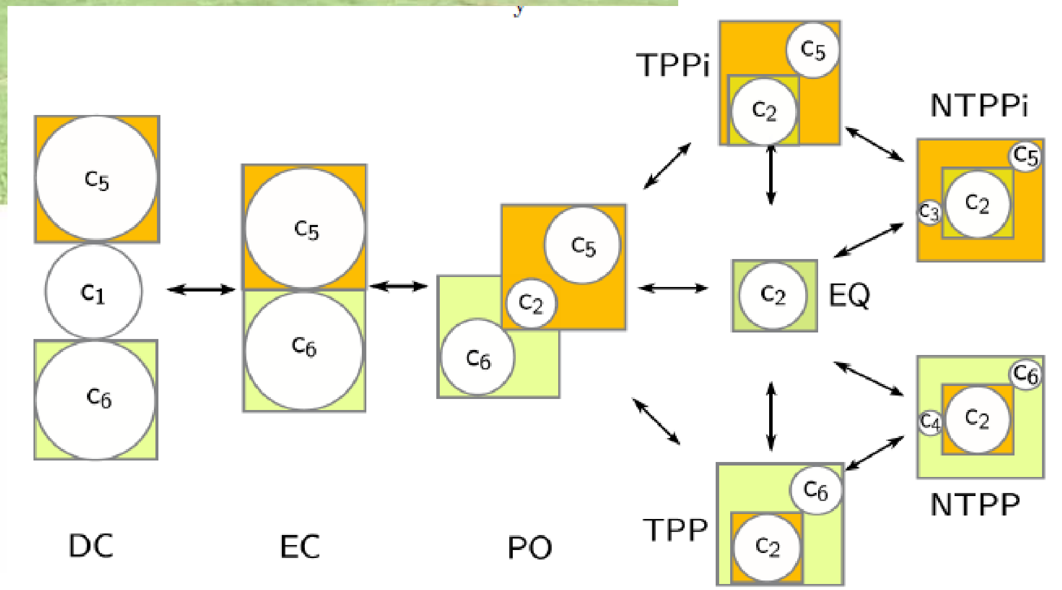
  - multiple cameras
  - video + language
  - not much investigated yet

- Induction (& abduction) more widely applied.

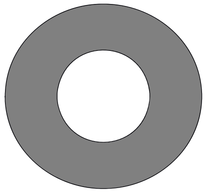
From video to QSR:
Using an HMM to ‘smooth’ relations



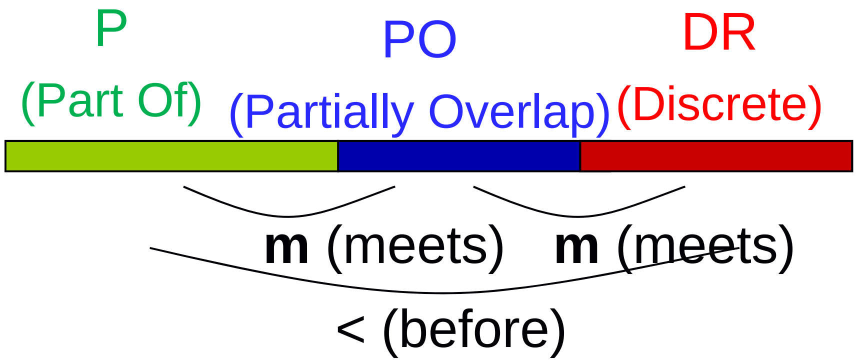
Sridhar et al.,
COSIT 2011
(best paper)

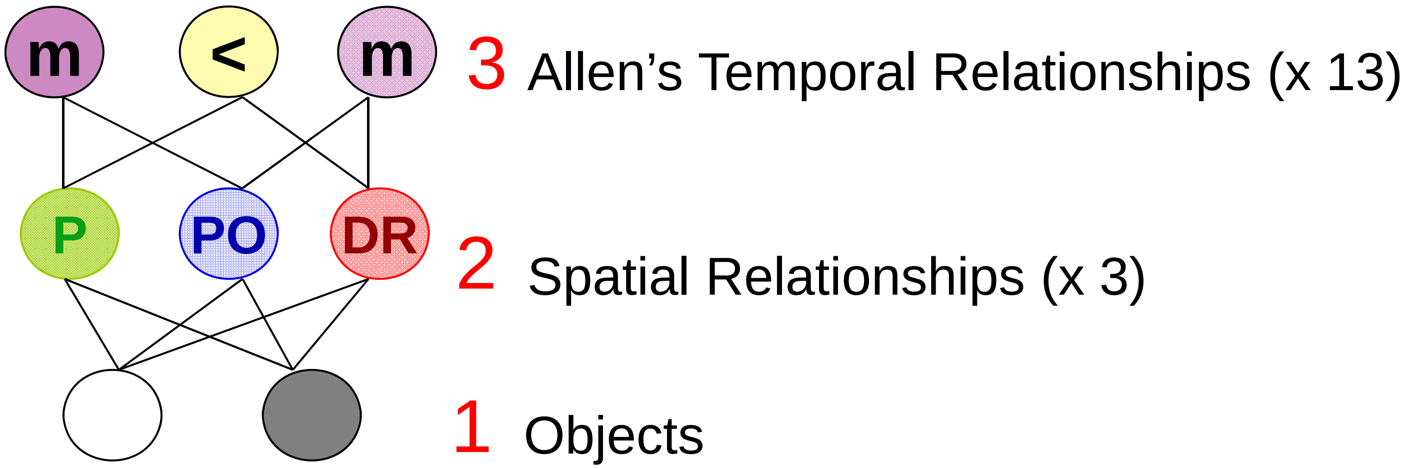


Representing interactions relationally



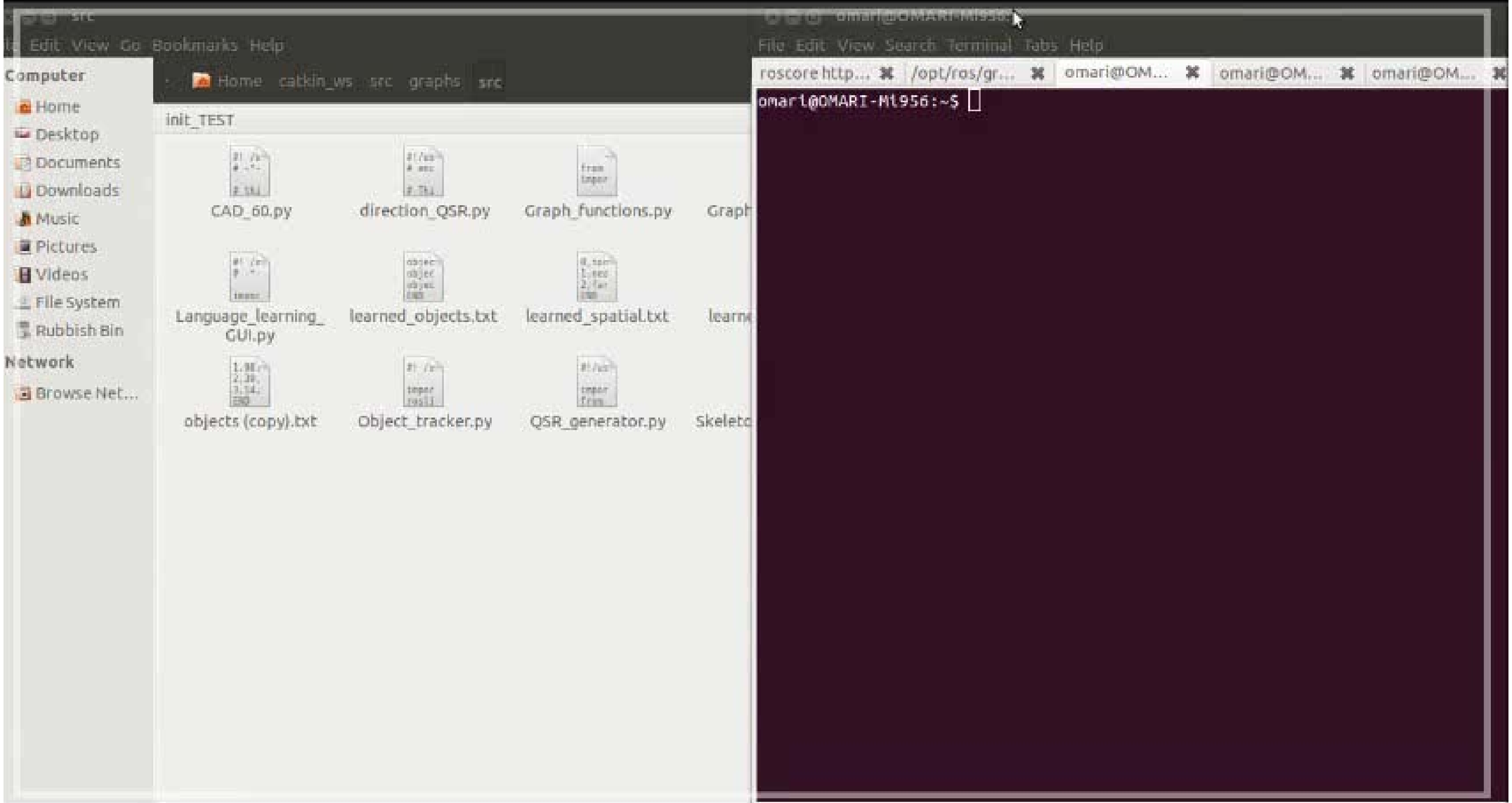
$$holds(X, Y, \mathsf{P}, I_1) \wedge holds(X, Y, \mathsf{PO}, I_2) \wedge holds(X, Y, \mathsf{DR}, I_3)$$
$$\wedge\ \mathsf{meets}(I_1, I_2) \wedge \mathsf{meets}(I_2, I_3) \wedge \mathsf{before}(I_1, I_3)$$





Demo of relational graph generation from video (running in ROS)

UNIVERSITY OF LEEDS



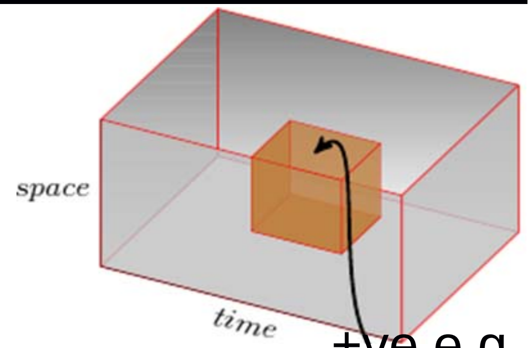
touch near far

Supervised event learning using ILP

UNIVERSITY OF LEEDS

Look what's *happening* over *there*

- "Deictic supervision"



+ve e.g.

- Just specify a rough s-t region for +v examples

- No need to specify *exactly* which objects are involved
- We have developed a *transactional*, *typed* Inductive Logic Programming (ILP) system to induce rules.

REMIND (**R**elational **E**vent **M**odel **IND**uction)

What is Inductive logic programming?

UNIVERSITY OF LEEDS

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

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

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

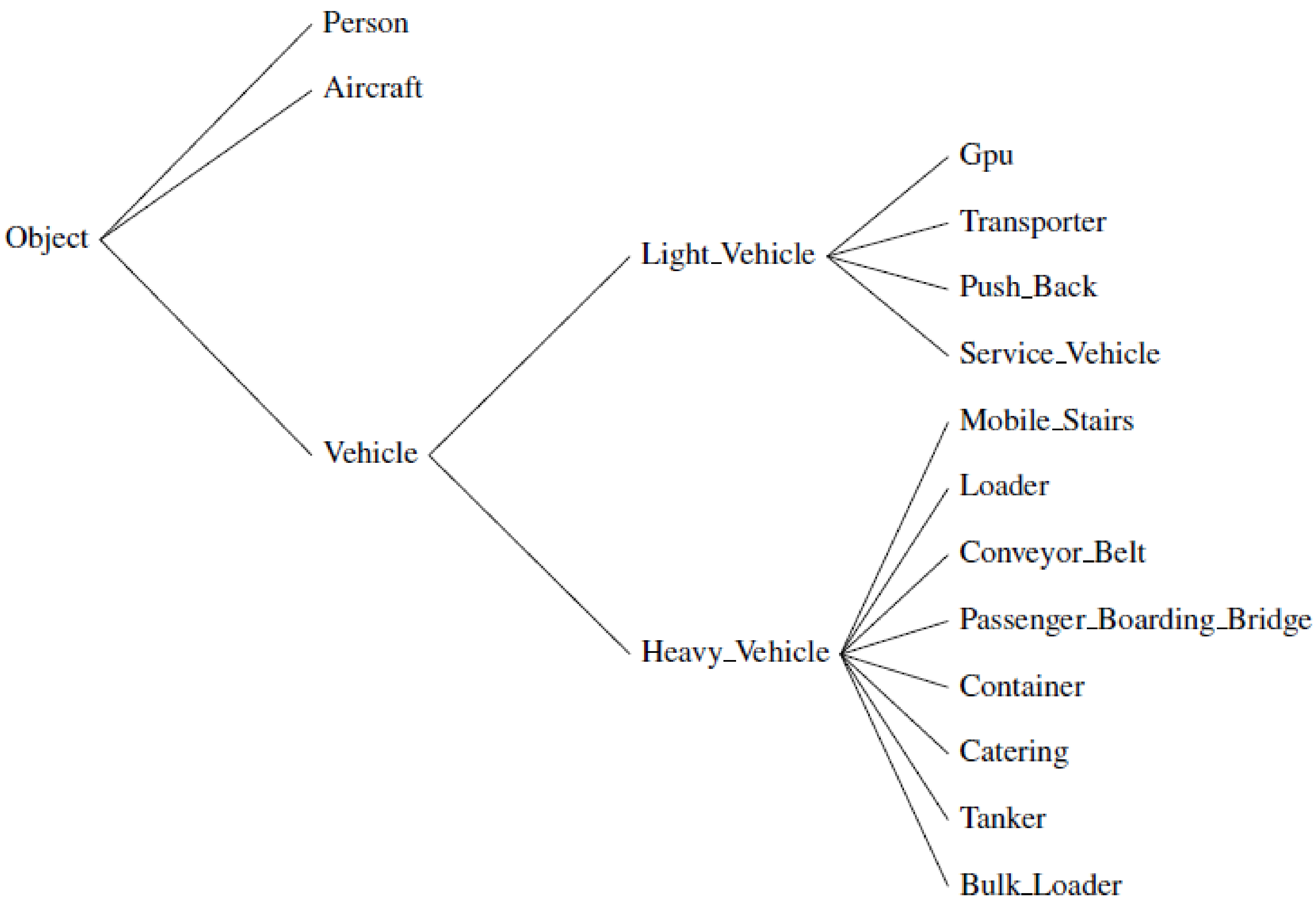
  - reduce overgeneralisation from noisy examples

  - improve efficiency during ILP hypothesis verification

Type hierarchy for aircraft turnarounds

UNIVERSITY OF LEEDS

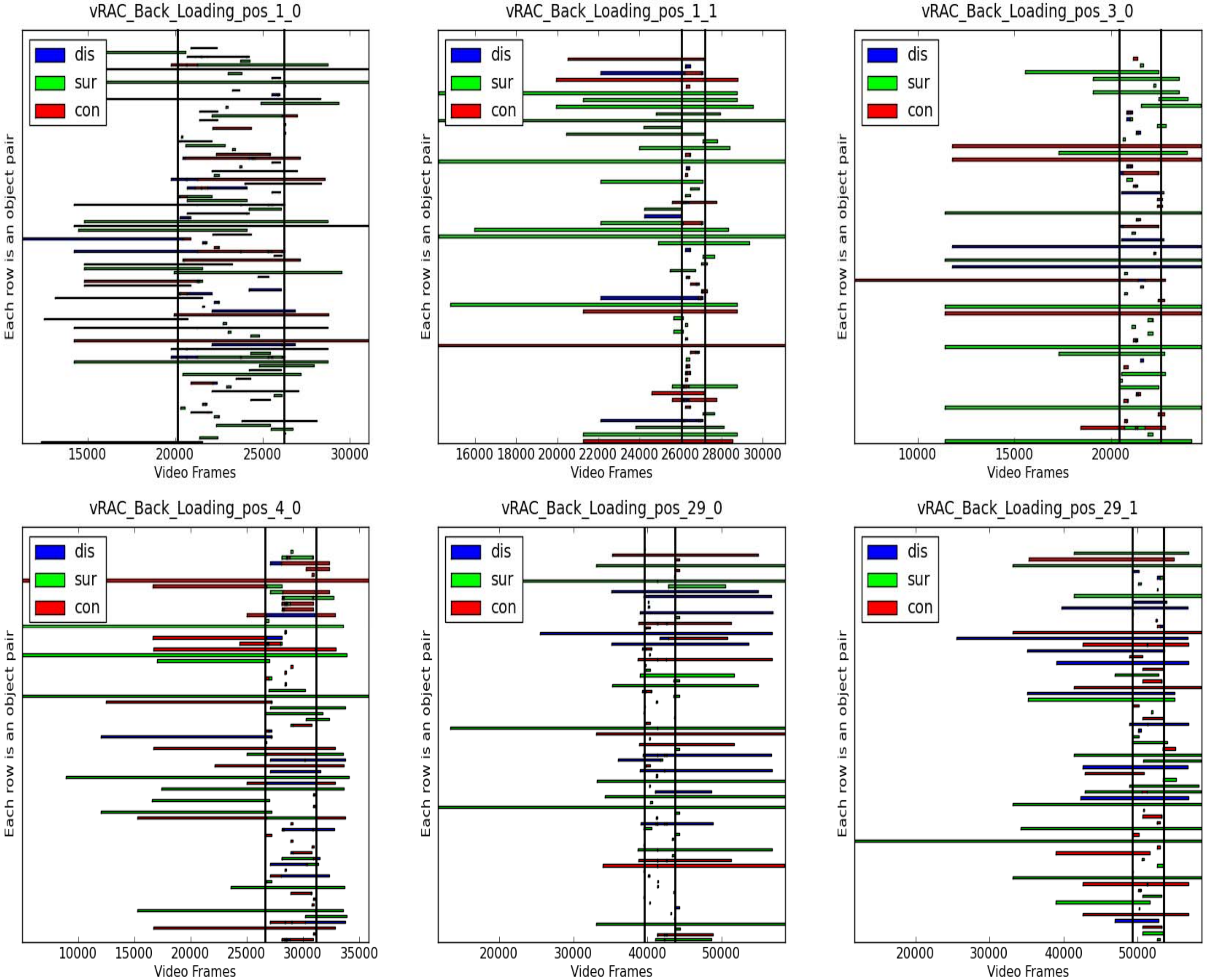
Hand built hierarchy, organised by perceptual similarity



“Learning from Interpretations” setting

UNIVERSITY OF LEEDS

Each positive example is represented as a separate Database



Search Strategy

Search the hypothesis lattice for a model that maximizes

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

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

Search moves

Rule specialisation:

- Initially RHS of rule is empty
- Add conditions to specialise rule to avoid negative examples
- Ordering on conditions to avoid duplicate generation

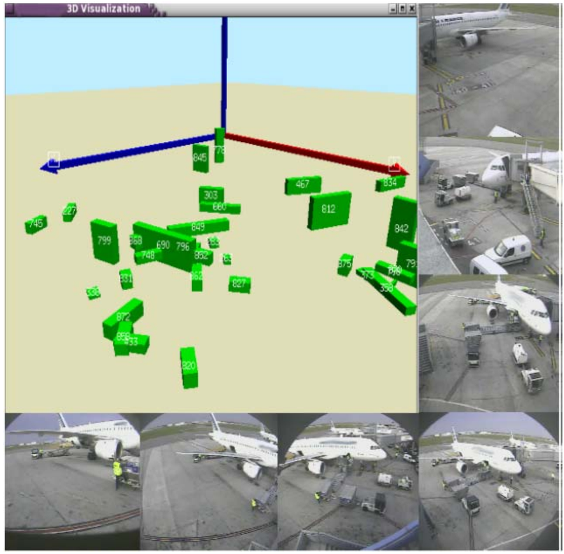
Type generalisation:

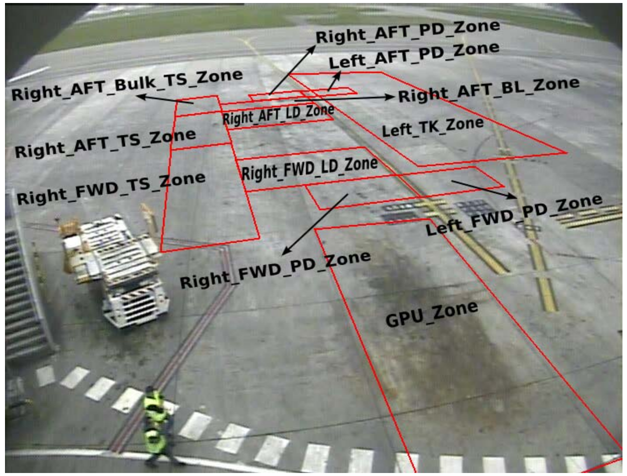
- Replace a type for some term with the next type up in the hierarchy.

UNIVERSITY OF LEEDS

Evaluation in aircraft turnaround domain

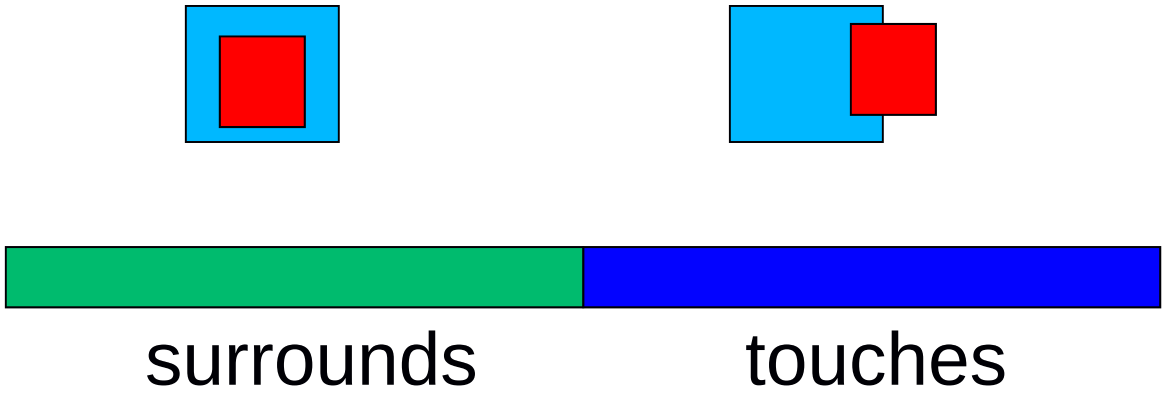
- 15 aircraft turnarounds
- 50,000 frames each turnaround
- 7 camera views
- Obtain tracks on 2D ground-plane
- ~350 spatial facts/video +temporal
- 10 event classes, 3-15 examples for each
- Many errors:
  - false/missing/displaced objects
  - broken/switched tracks
- Generate spatial relations between objects/IATA-zones
- Prolog rules determining temporal relations are in Background
- Leave-one-out (from turnarounds) testing





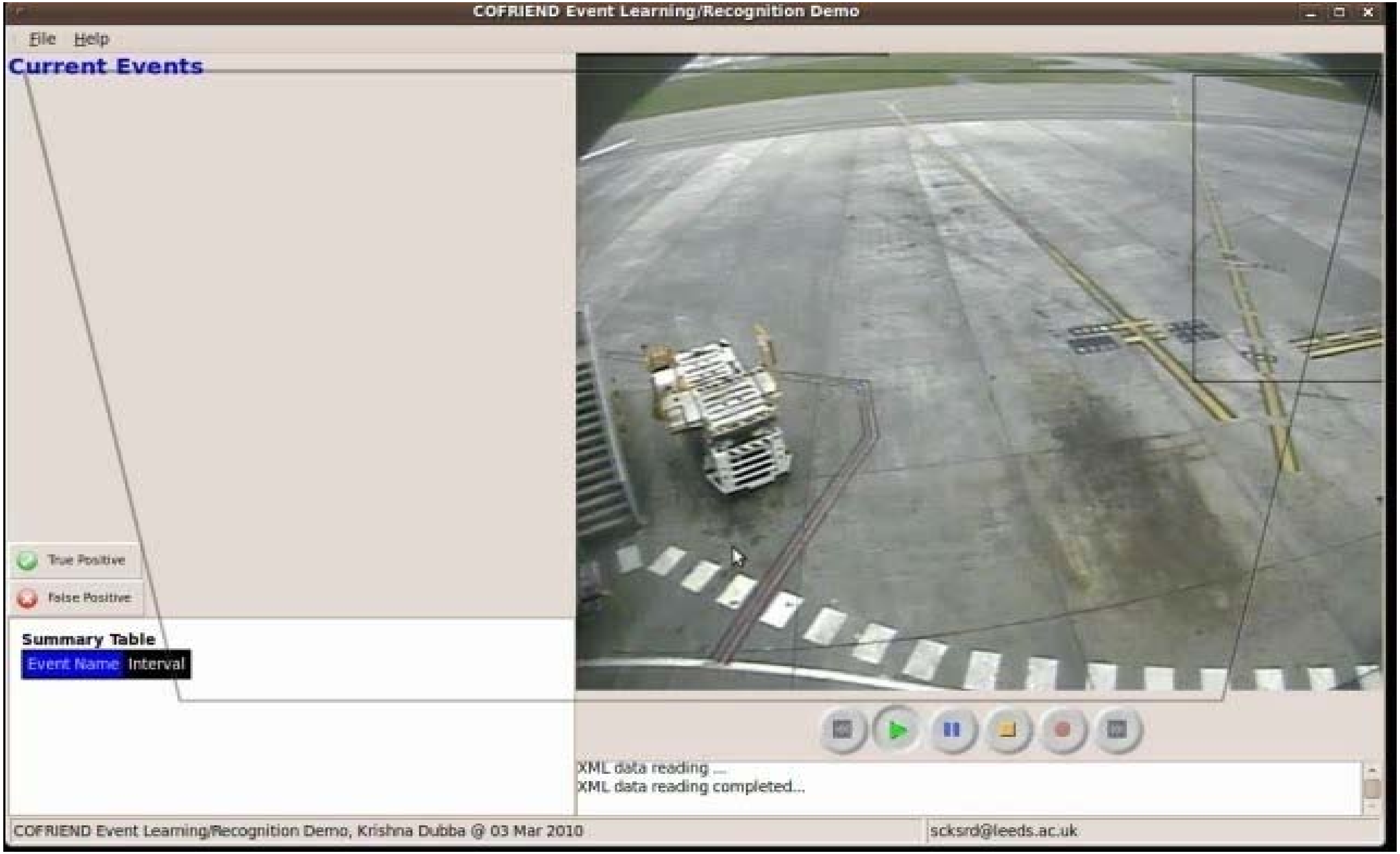
A Learned Event Model:

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

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

Interleaving induction and abduction (IIA)

Problem: noisy data tends to produce too many rules and overfit the data; more data can help but what if it's not available?

Idea: explain away noisy instances using abduction so that rules are not explicitly generated to cover these (Dubba et al 2012)
- Assume that noise in examples is random

Domain independent spatial theory:
- Basic calculus properties (e.g. JEPD relations, symmetry…)
- Conceptual neighbourhood axioms
- Composition Table
- Axioms linking different calculi (e.g. topology + size)

Abductive Explanations

Given a theory T and observations (example) G, find an explanation Δ s.t. (Kakas et al 92):

- $T \bigcup \Delta \models G$
- $T \bigcup \Delta$ is consistent

Reduce # explanations:

- Basic (not explain another explanation
- Minimal (not subsume another explanation)
- Satisfy (spatial) theory
- Look for *low cost* explanations

Explanation cost

Lowest cost:

extending the interval when a spatial relation holds

Medium cost:

change of spatial relation (to a conceptual neighbour)

Highest cost:

introduction of a hypothetical object
(to cover case where vision system fails to detect object)

Interleaving abduction and induction: results

UNIVERSITY OF LEEDS

Airport Events	pos_ex	●	□	◇	RoI	PoI	RIA	PIA
FWD_CN_LoadUnload	5	2	1	2	0.8	0.3	0.8	0.4
GPU Positioning	15	5	3	4	1	0.2	1	0.4
Aircraft Arrival	15	5	2	5	0.38	0.26	0.33	0.32
Aircraft Departure	15	5	2	7	0.8	0.15	0.71	0.26
AFT_Bulk_LoadUnload	12	5	2	4	0.63	0.43	0.63	0.65
Left Refuelling	6	2	1	2	0.66	0.5	0.66	0.55
PB Positioning	15	4	3	2	0.33	0.34	0.33	0.42
AFT_CN_LoadUnload	7	3	1	3	0.57	0.4	0.57	0.51
PBB Positioning	15	4	3	2	1	0.57	1	0.62
PBB_Removing	15	5	2	5	0.54	0.23	0.54	0.31
FWD_Bulk_LoadUnload	3	2	1	1	1	1	1	1

● Num of rules with only Induction □ Num of rules with $\mathcal{IIA}$ ◇ avg num of examples covered by abd

IIA in a "verbs" domain

Verb Events	#pos	●	□	◇	RoI	PoI	RIA	PIA
Approach	584	12	5	45	0.73	0.12	0.74	0.12
Arrive	8	2	1	2	0.50	0.05	0.50	0.05
Attach	48	6	3	12	1.00	0.14	1.00	0.17
Bounce	22	2	2	0	0.95	0.06	0.95	0.08
Catch	201	7	4	31	0.59	0.11	0.56	0.11
Chase	108	11	7	19	0.59	0.08	0.57	0.08
Collide	101	6	4	14	0.98	0.16	0.98	0.18
Dig	140	10	7	21	0.96	0.38	0.96	0.39
Drop	44	2	2	0	1.00	0.16	1.00	0.16
Exchange	18	6	3	4	0.40	0.03	0.40	0.03
Fall	134	8	5	18	0.92	0.35	0.90	0.35
Give	552	27	20	54	0.94	0.56	0.94	0.60
Jump	150	6	4	14	0.98	0.13	0.98	0.13
Kick	48	4	3	6	1.00	0.15	1.00	0.15
Leave	116	10	4	34	0.67	0.20	0.67	0.22
Lift	78	8	5	17	0.67	0.24	0.67	0.24
Pass	76	8	4	13	0.87	0.10	0.87	0.12
Pickup	40	6	4	8	0.81	0.13	0.81	0.16
Run	76	7	5	7	0.57	0.12	0.57	0.12
Throw	26	3	2	5	0.67	0.11	0.67	0.11

● Num of rules with only Induction □ Num of rules with $\mathcal{IIA}$ ◇ avg num of examples covered by abd

An alternative way of handling noise

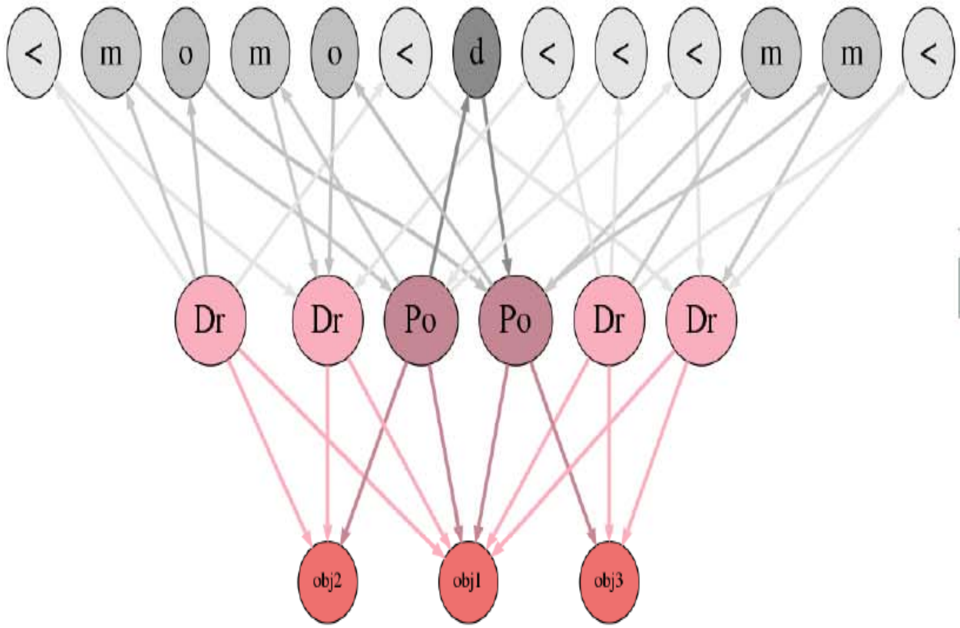
UNIVERSITY OF LEEDS

- Represent video portions as histogram of relational features
- Use metric learner (SVM, KNN…) to model event classes

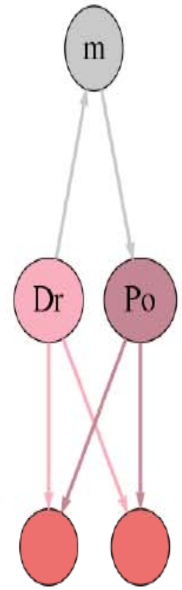
Graph Formulation

UNIVERSITY OF LEEDS

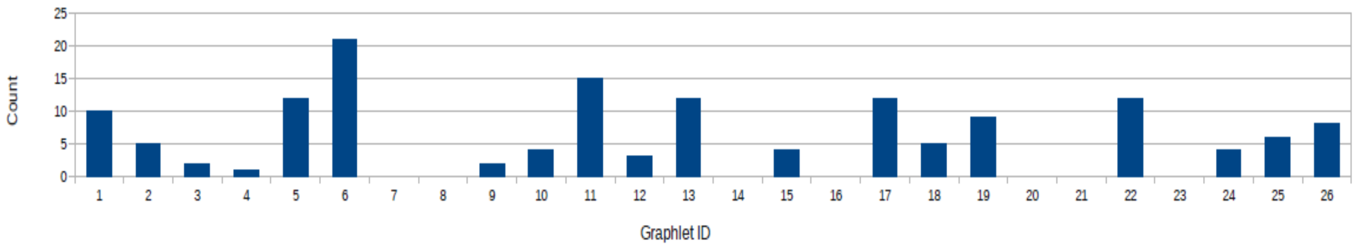
Activity Graph:



Graphlet:



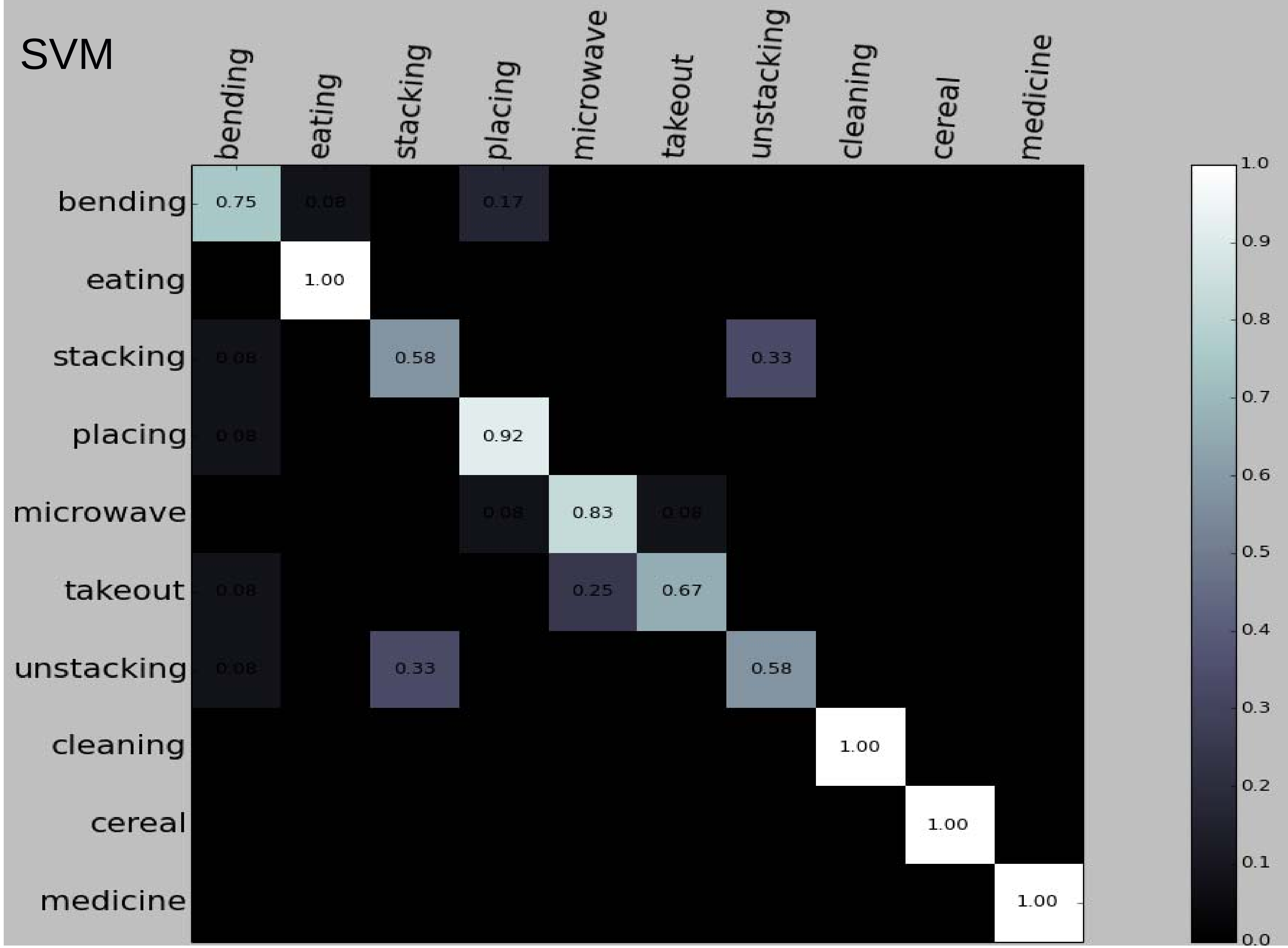
Histogram of Graphlet Count:



CAD120: 85% Precision & 85% Recall

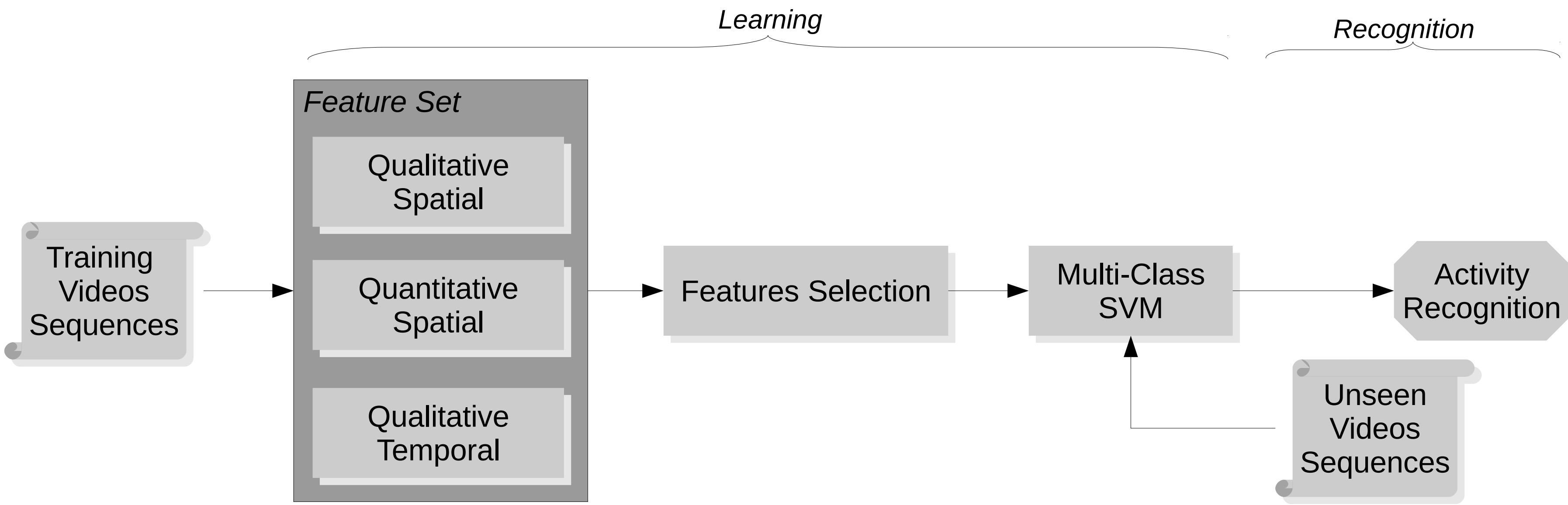
Leave-one-subject-out Cross Validation

UNIVERSITY OF LEEDS



Activity recognition with feature selection

Need more feature expressivity, but which ones?



Feature Set

UNIVERSITY OF LEEDS

$$F = \langle F_1, F_2, F_3 \rangle$$

F_1 Qualitative Spatial Relationships

Count R_i in RCC-3

< R_1>

< R_1 R_2>

< R_1 R_2 R_3>

< R_1 R_2 R_3 R_4 >

F_2 Qualitative Temporal Relationships

For each pair of Consecutive relations, *Compute relative length*

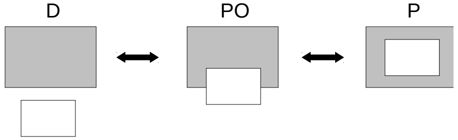
$r = | R_2 | / |R_1|$

Use k-means to bin r into = , long, short

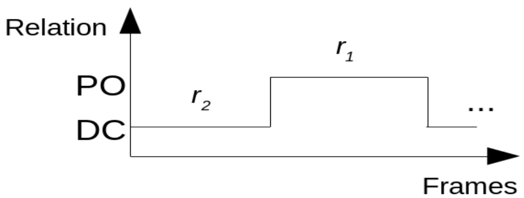
F_3 Quantitative Spatial Relationships

Compute descriptive statistics of distances and direction of motion between joints of skeleton and objects across all frames:

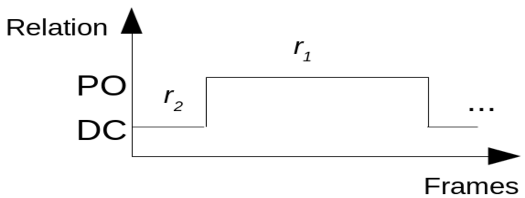
- *Mean*
- *Standard deviation*
- *Skewness*
- *Kurtosis*



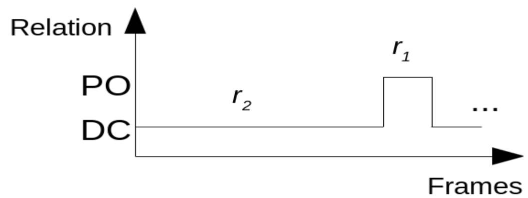
Equal (E)



Long (L)



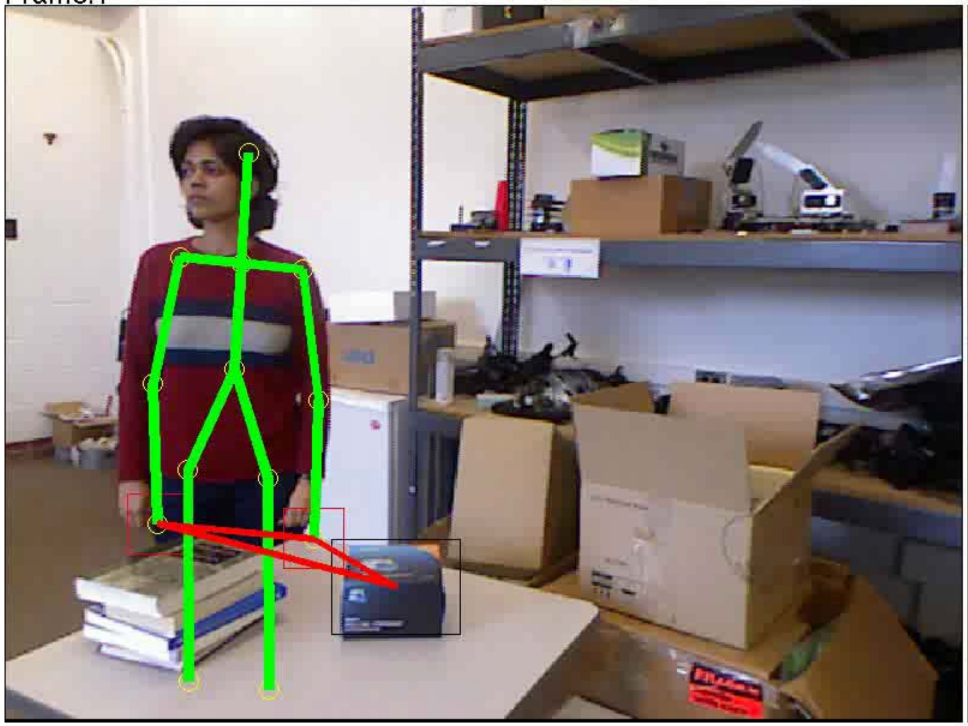
Short (S)



Feature generation

QUALITATIVE FEATURES

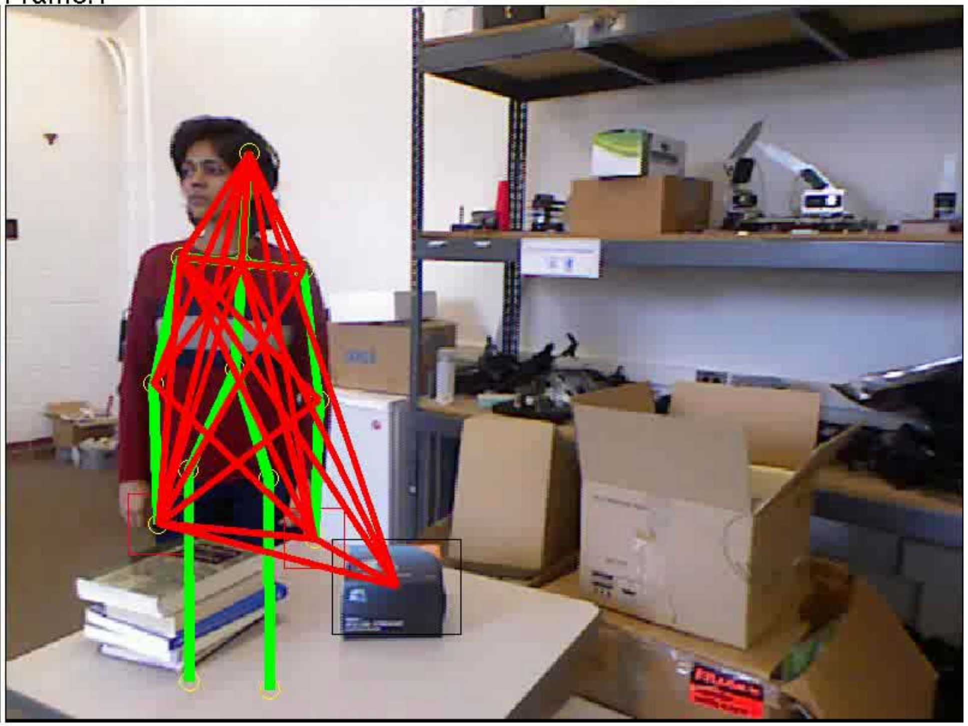
Frame:1



ELEMENT PAIR	RCC RELATION	TEMPORAL RELATION
HandR - HandL	Disconnect	-
HandR - Object	Disconnect	-
HandL - Object	Parttially Overlap	-

QUANTITATIVE FEATURES

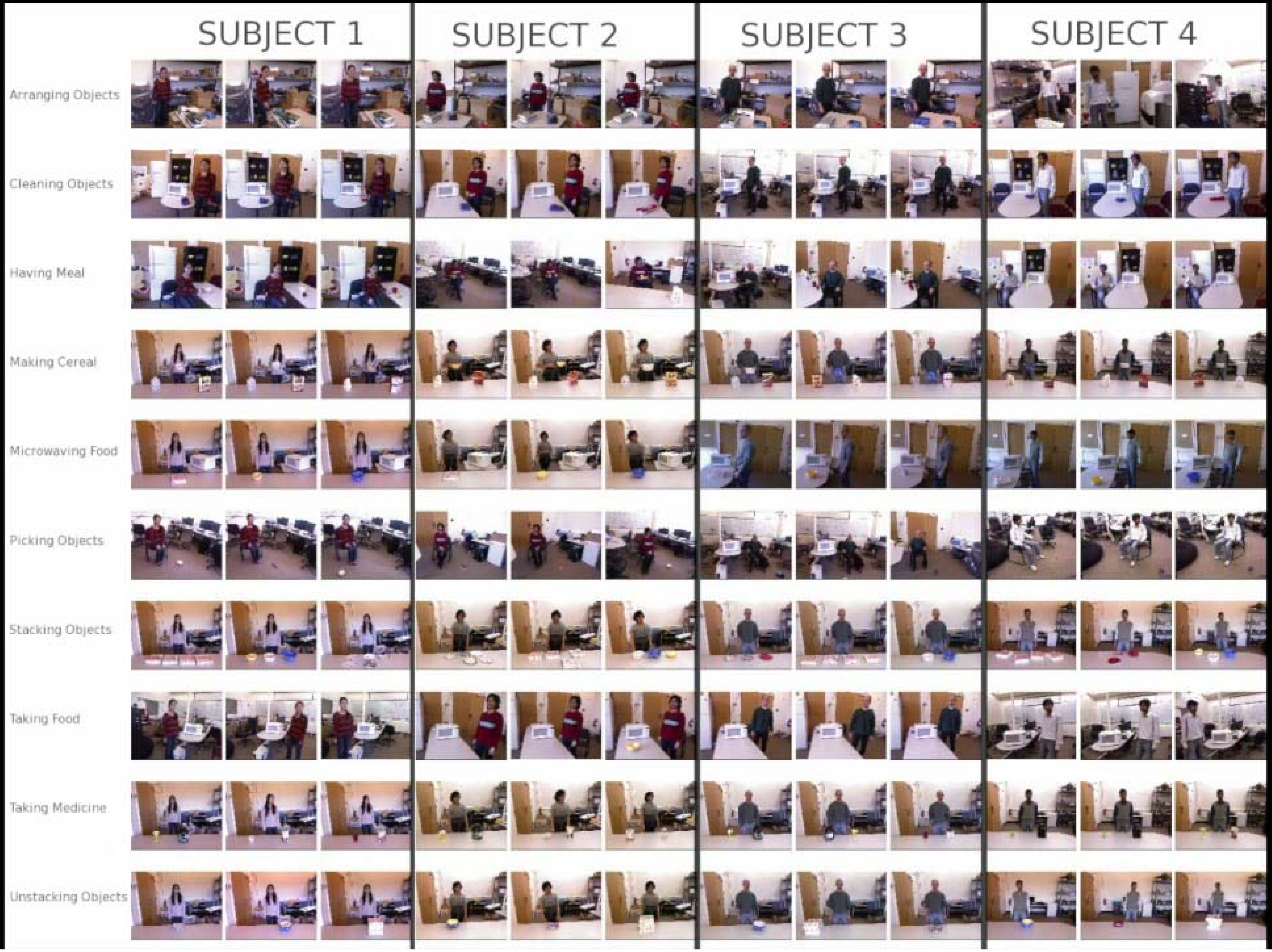
Frame:1



ELEMENT PAIR	EUCLIDEAN DISTANCES
HandR - HandL	104.7706
HandR - Object	63.6027
HandL - Object	164.6822
HandL - ShoulderR	195.1514
HandL - Head	255.0876
...	

Results of 4 fold cross evaluation

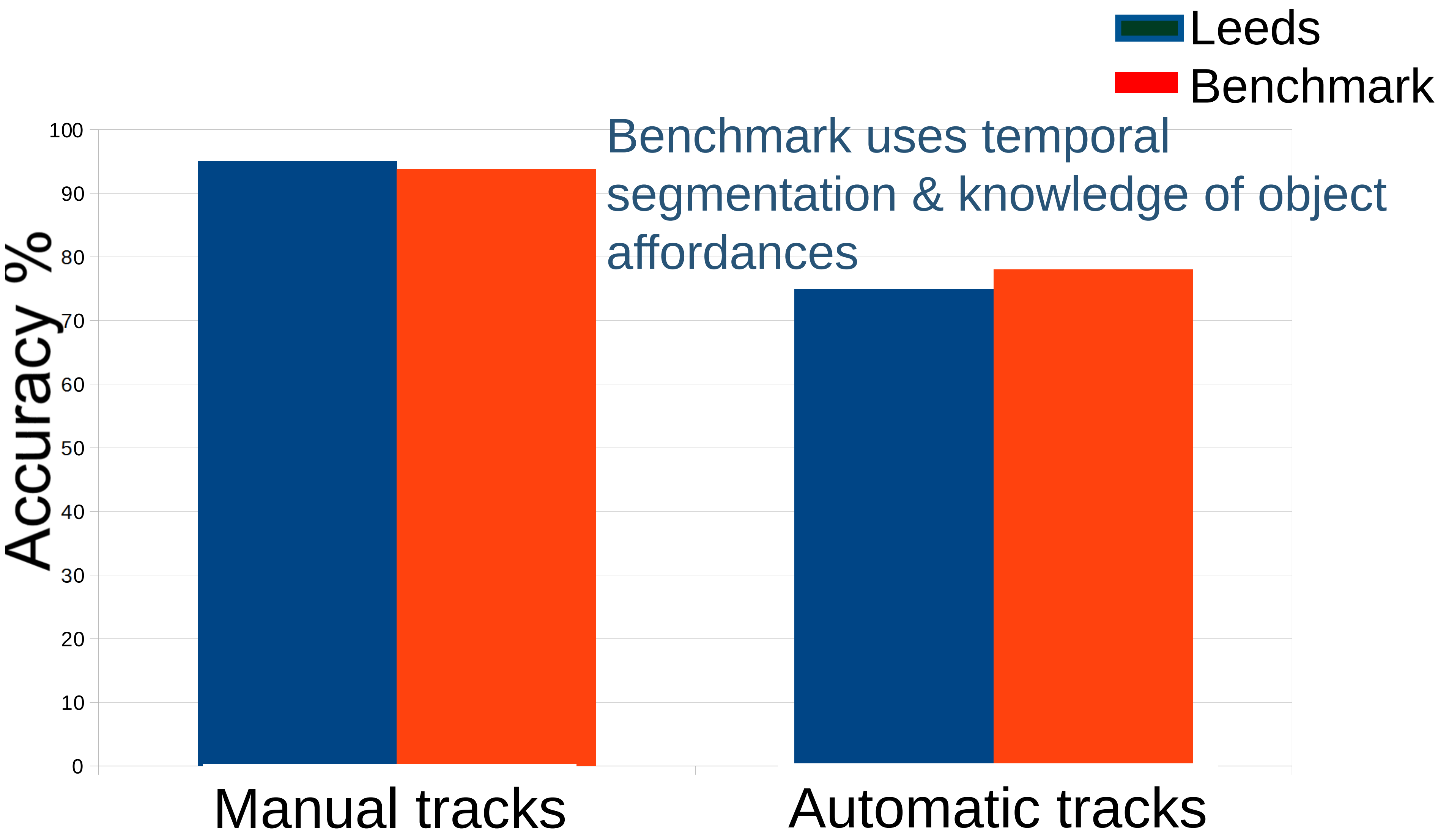
UNIVERSITY OF LEEDS



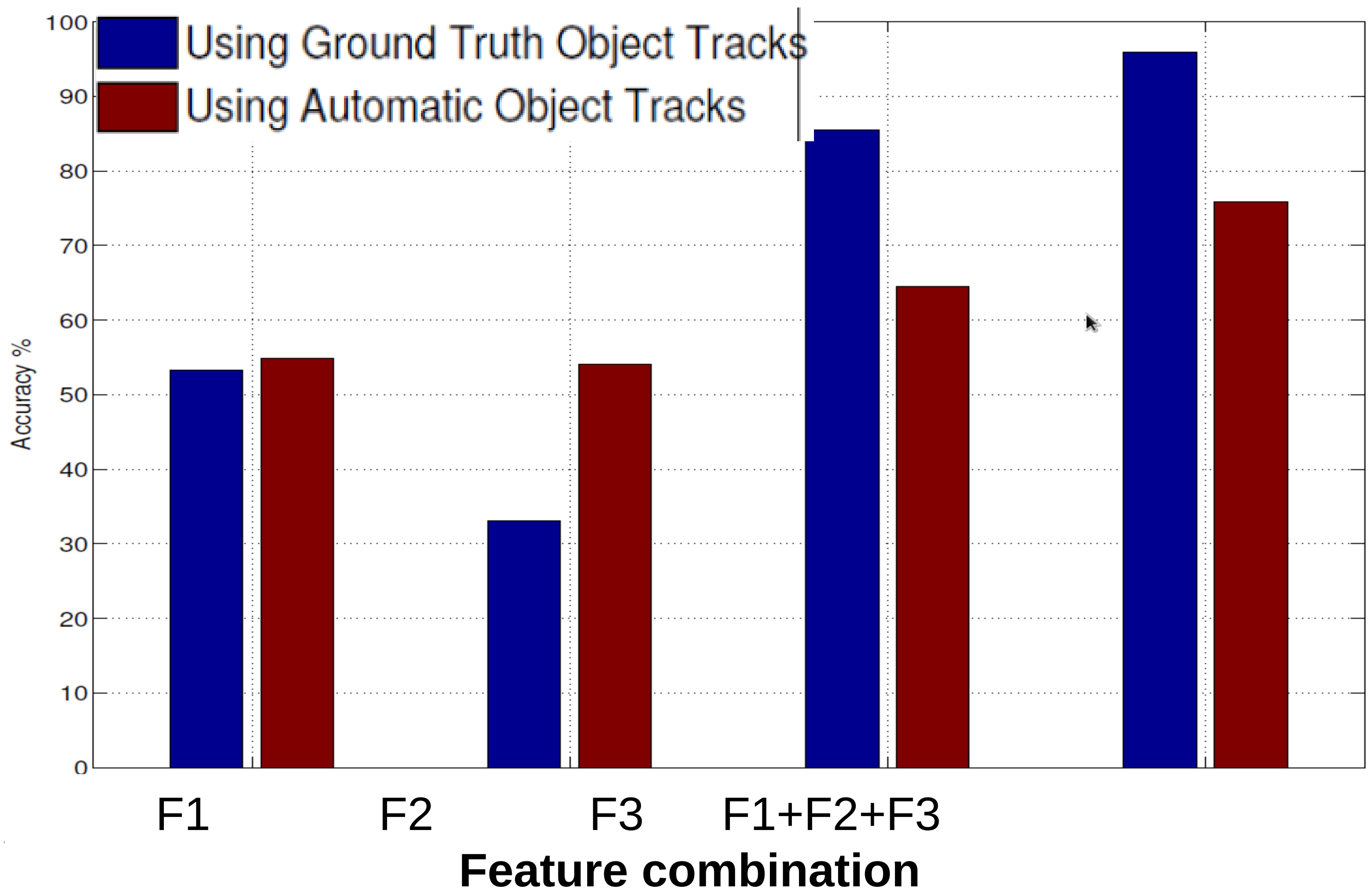
Each video will turn red/green on classification after completion.

Experiments: CAD120

UNIVERSITY OF LEEDS



Comparison of features

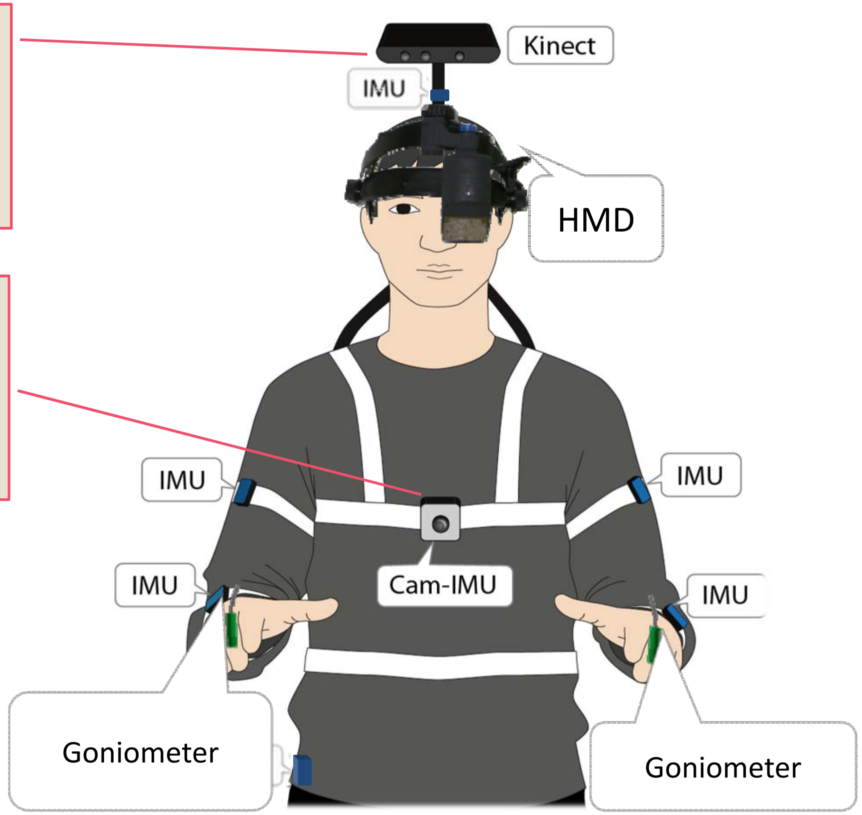


Cognito project: Learning workflows

UNIVERSITY OF LEEDS

Object recognition

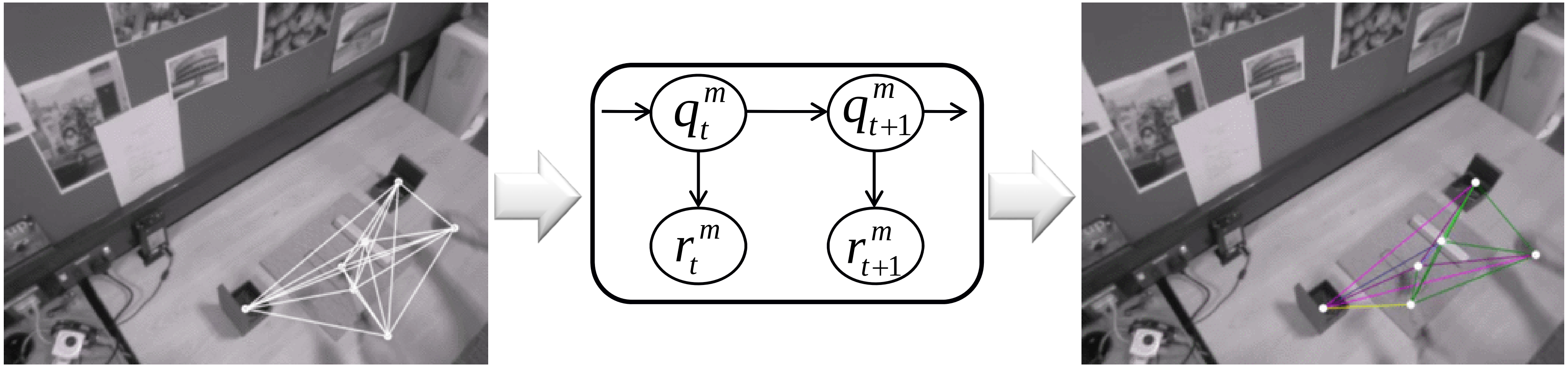
Wrist recognition



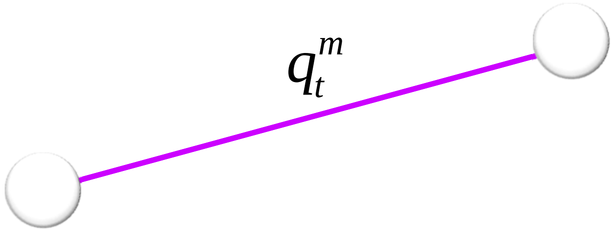
Intended application: learn workflow from few experts, then guide novices; e.g. for maintenance tasks, construction tasks…

Why egocentric?: movement between workspaces; no need for fixed cameras; reduces chance of occlusion

Learning relations



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



Continuous relations

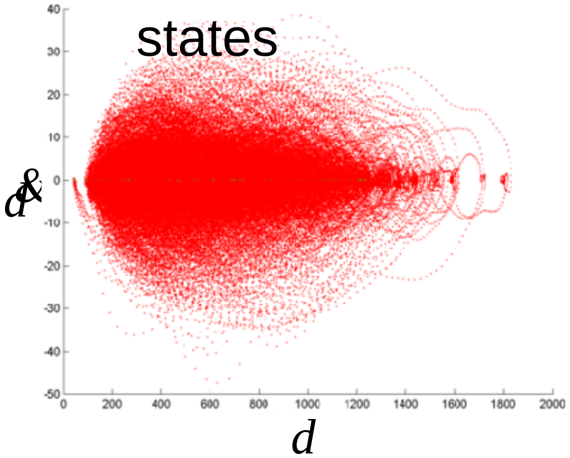
Finite discrete relations

Global, or for each pair of object types

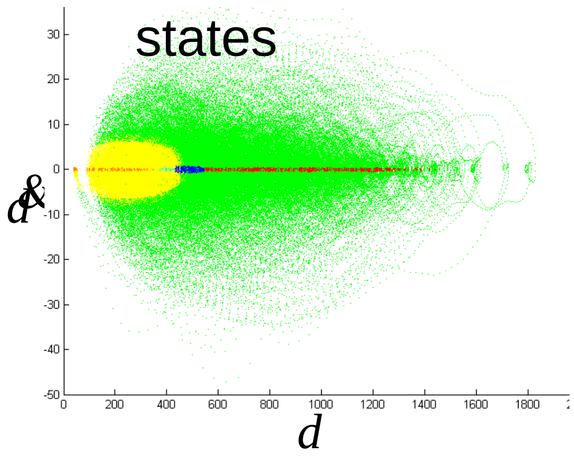
Quantisation of Relational Features

UNIVERSITY OF LEEDS

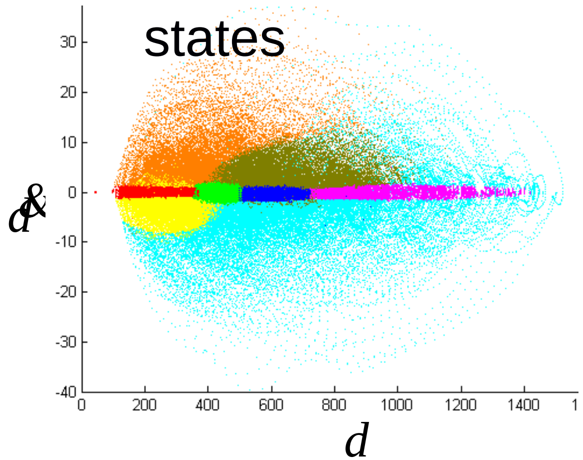
2 discrete states



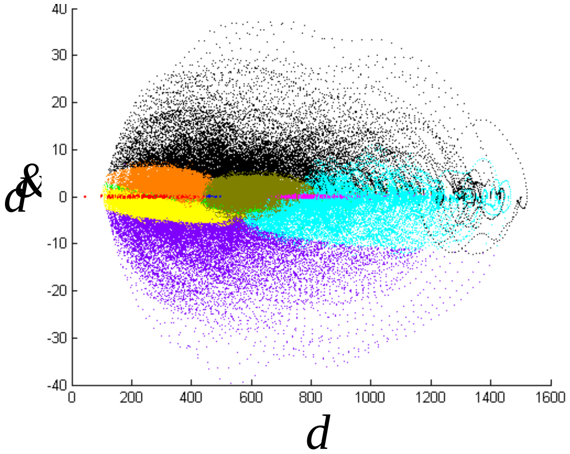
6 discrete states



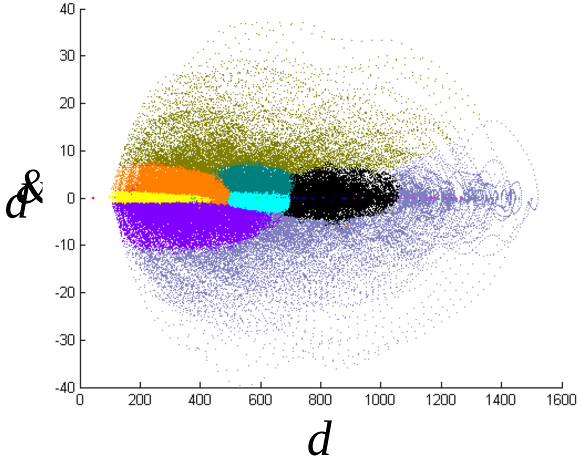
8 discrete states



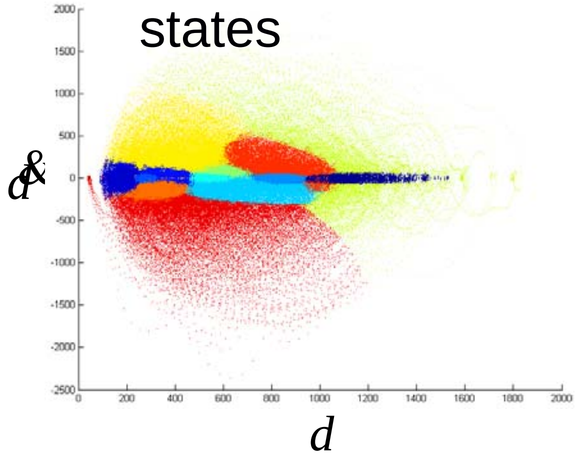
10 discrete states



12 discrete states



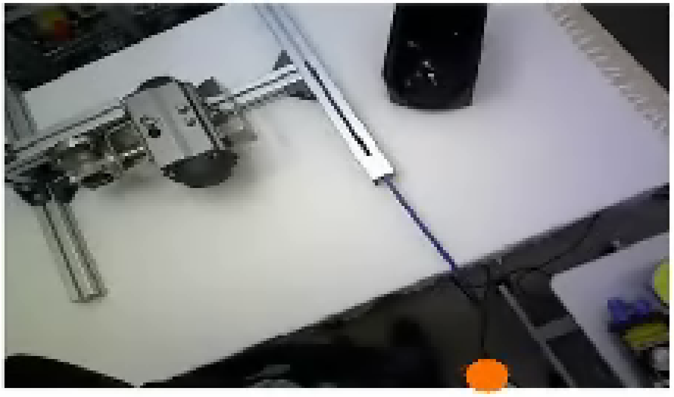
16 discrete states

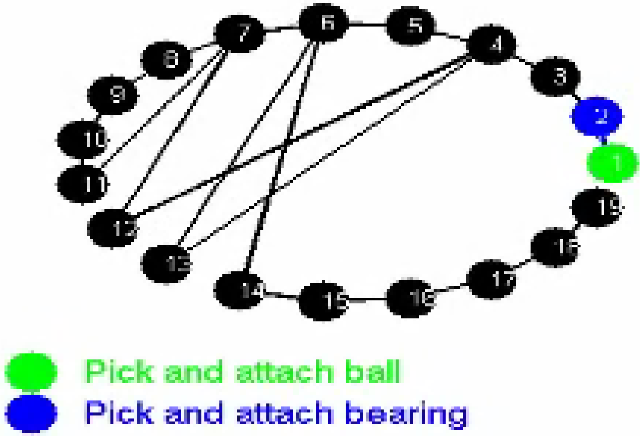


Use a Bayesian Information Criterion to optimize number of states/relations

Ball valve example

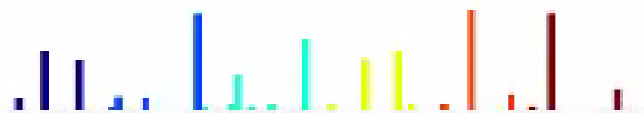
Relational Graph





Bag-of-Relations : Object-Object

Bag-of-Relations : Upper-Body Model



Instructions given to user via a Head Mounted Display

UNIVERSITY OF LEEDS

Summary/novelty

- Many QSR calculi available
- From pixels to symbolic, relational, qualitative behaviour/event descriptions
- Supervised and unsupervised
- 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

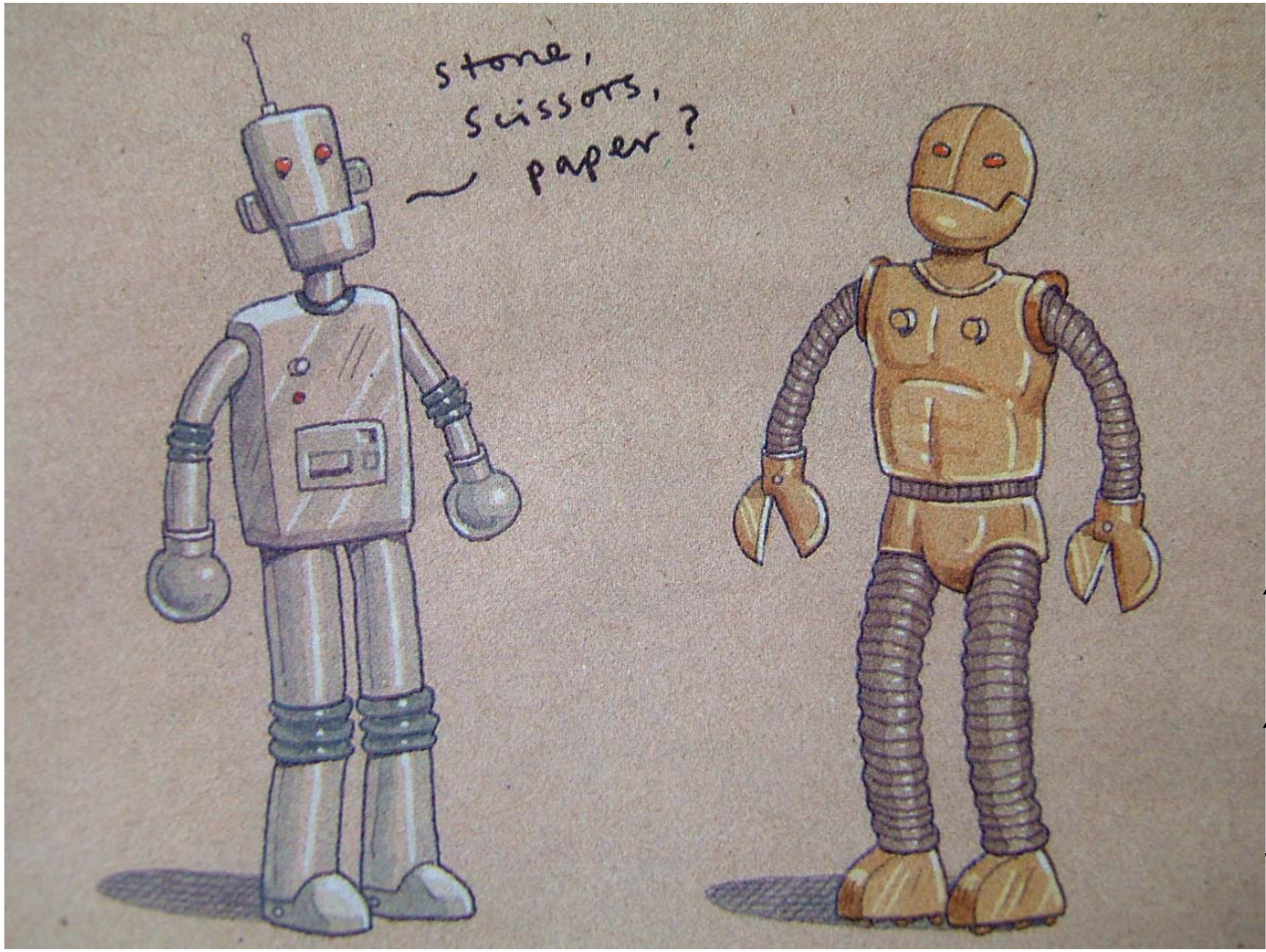
www.comp.leeds.ac.uk/qsr/publications.html

Research challenges/ongoing work

- New domains, longer time frames, larger environments
  - STRANDS project: aiming for 4 months continuous
  - Learning a global model – temporal sequencing
  - Daily, weekly, monthly routines
  - Activities and subactivities
- Further experimentation with different sets of spatial relations
- Use induced functional categories to supervise appearance learning
- Learning probabilistic weights for rules (MLN)
- Cognitive evaluation of event classes and functional categories
- Online learning and Ontology alignment
- Language (+ vision)
- …

UNIVERSITY OF LEEDS

Any Questions?



Thanks to:

EPSRC,

EU *(CoFriend, Cognito, RACE, STRANDS),*

DARPA *(Mindseye/Vigil)*

David Hogg,

Krishna Sridhar,

Sandeep Dubba,

Ardhendu Behera,

Paul Duckworth,

Aryana Tavanai,

Muhannad al Omari,

Jawad Tayyub,

Eris Chinellato,

Yiannis Gatsoulis