Object Detection in Context

David Hogg
University of Leeds

Introduction

Two challenges for research into object detection

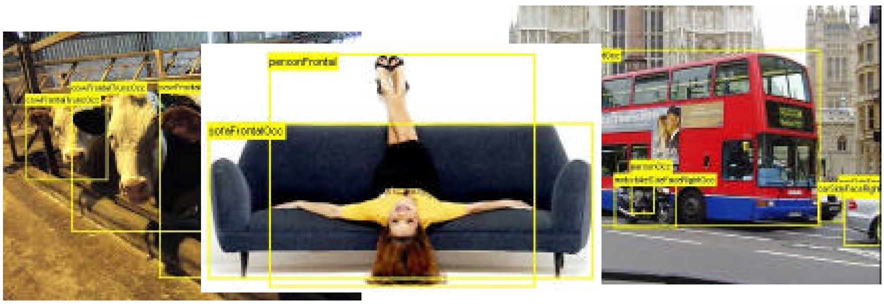
- Removing the need for supervision in learning
- Dealing with ambiguity and error

Explore object detection in the context of activity analysis

Learning object categories

Supervised learning is the dominant approach…

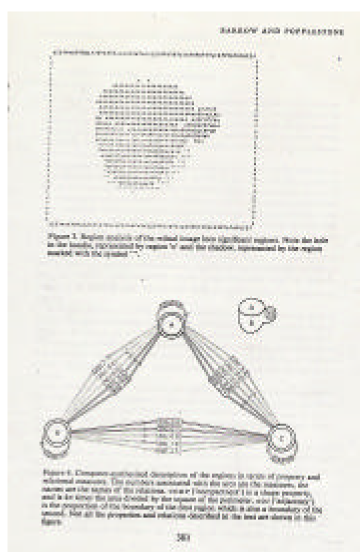
- PASCAL Visual Object Classes Challenge 2008
    - 20 classes (e.g. cow, bus, sofa, person)
    - predict absence/presence of each object
    - find bounding box for each object
    - ~4k training images depicting ~10k target objects



This approach has a long history…

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

Relational descriptions of object classes + supervised learning



…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 ......

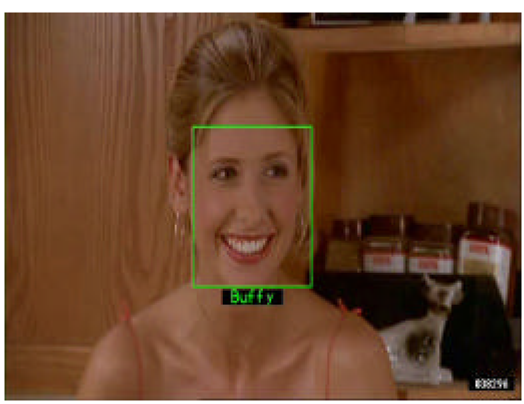
...... Such activities involve planning, inductive generalization, and, indeed, most of the capacities required by an intelligent machine. To develop a truly integrated visual system thus becomes almost co-extensive with the goal of producing an integrated cognitive system.'

Barrow and Popplestone, 1971.

A step in this direction…

Learning from video & text, for example:

- TV shows + subtitles + scripts (Everingham et al., BMVC 2006)

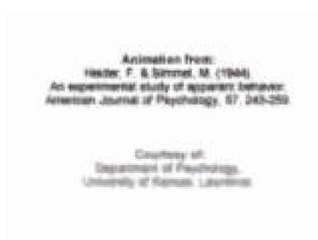


Going all the way - simulating evolution

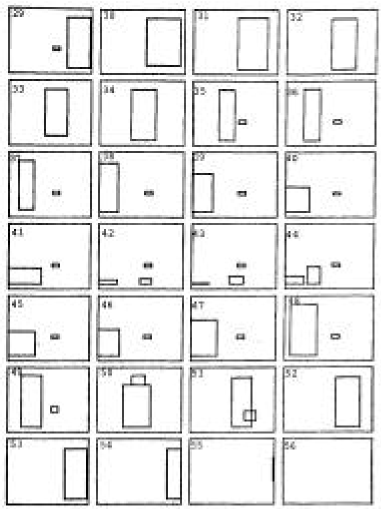
Karl Sims, Evolving Virtual Creatures, Siggraph 1994.

Object detection in the context of activity analysis

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



Heider & Simmel, 1944



An object has appeared at left of scene, call it object A.
Object A has begun moving.
Object A looks like a person, call it Fred.
An object has appeared in middle of scene, call it object B.
Object B came from Fred.
Fred placed object B in the scene.
Object B looks like an inanimate object.
Fred has disappeared at right of scene.
An object has appeared at right of scene, call it object C.
Object C has begun moving.
Object C looks like a person, call it Mary.
Mary has completely crossed the scene.
Mary has stopped moving.
Mary has laid down.
Mary has stood up.
Mary has begun moving.
Mary has completely crossed the scene.
Object B has disappeared in middle of scene.
Object B was engulfed by Mary.
Mary picked up object B.
Mary has stopped moving.
Mary has disappeared at right of scene.

Learning functional object categories from activity analysis

Krishna Murali, Cohn, and Hogg, ECAI-08.

Overview

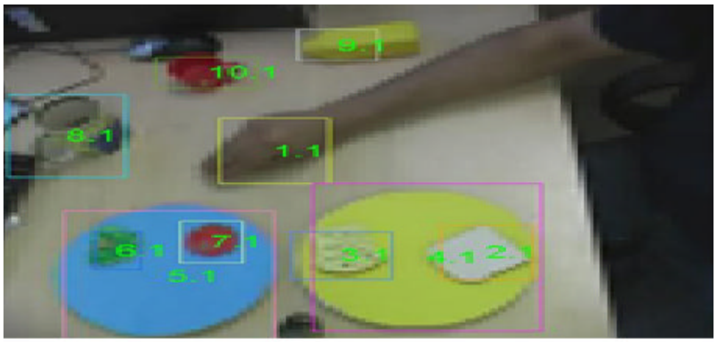
- Learn event classes from patterns of qualitative spatio-temporal relations
- Cluster objects by their role in these activities

Focus on food preparation

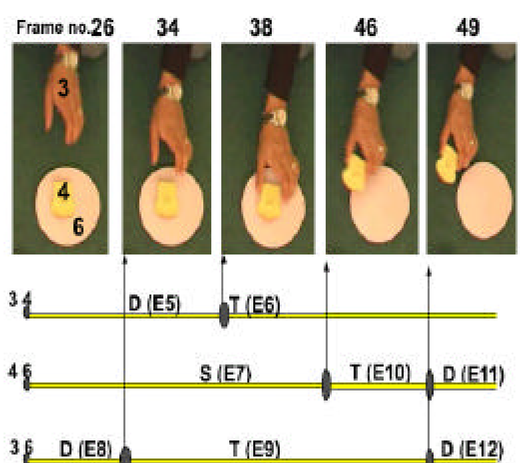
- Large set of objects with a rich taxonomy
- Repeated patterns of events involving these objects

Object discovery

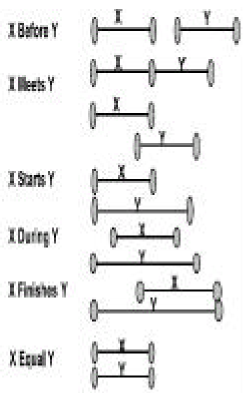
Colour-based blob detection and tracking



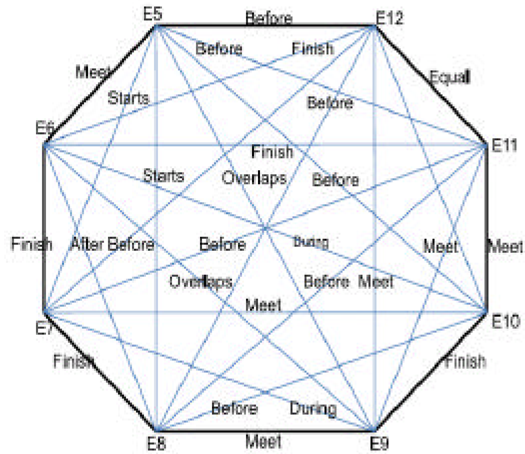
Spatial relations



Temporal relations and the '*activity graph*'



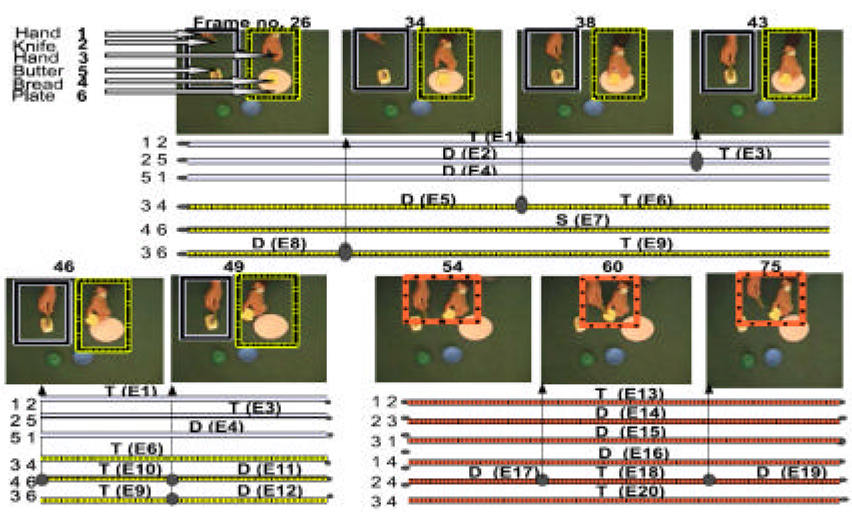
Allen's relations



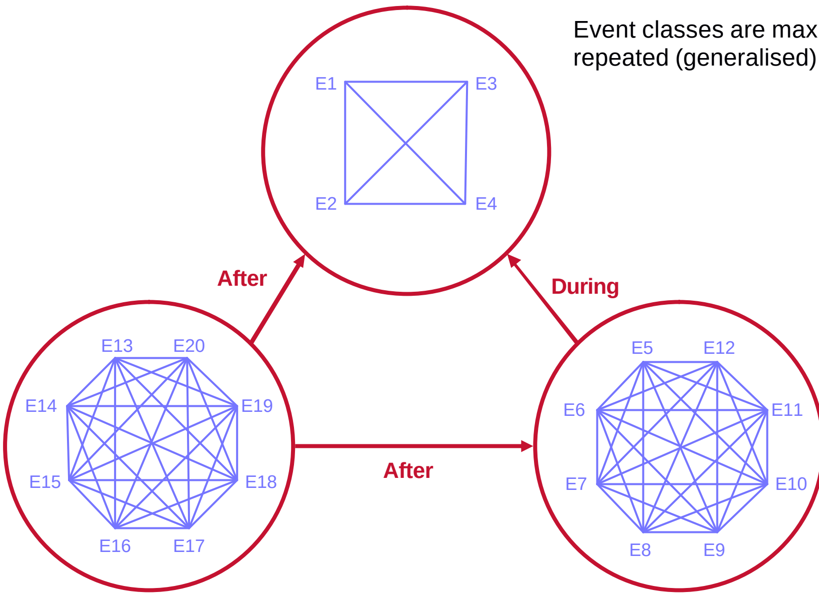
Using Allen's relations facilitates dealing with gaps, partial ordering and parallel activities

Attention

Focus on *atomic events*: maximal sub-graphs involving a constant set of connected (S,T) objects, at least one of which must move



Replace activity graph by a graph over atomic events



Event classes are maximal repeated (generalised) subgraphs

Learn functional object categories from event roles

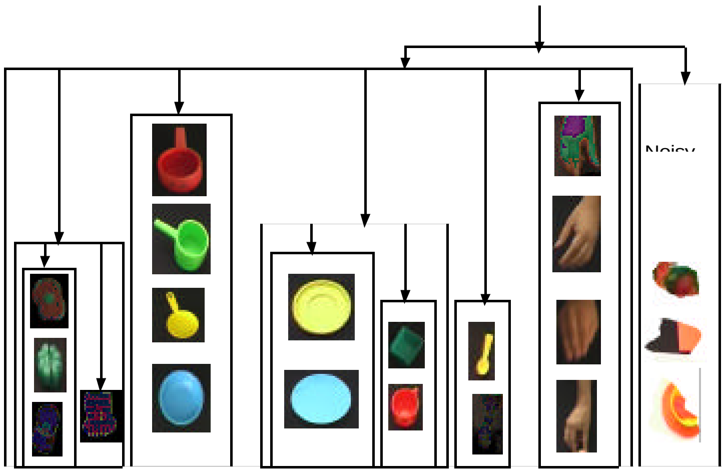
Objects within each event class (+ partially generalised classes)

	X1	X2….Xn		Y1… ..Ym	Z1…. …Zl	A1… … An
O1	1	0	0	1……………..		
O2	0	1	0	…….1……….		
O3	0	1	0	…….1……….		
O4						
.						
.						
.						
.						
.						

Create matrix of roles played by each object
Reduce using PCA
Obtain object taxonomy from hierarchical clustering of rows

Experimental Results

Video of 5 minutes: preparing breakfast with tea, and a simple vegetable curry
50 objects, ~ 3000 roles. PCA reduces to 50 x 20

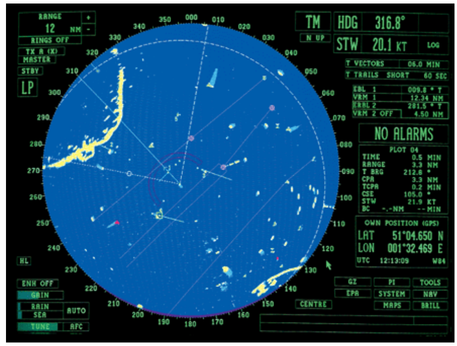


Dealing with detection errors and ambiguity

Radar tracking

Dealing with

- missed detections
- spurious detections



Long history from radar literature and elsewhere:

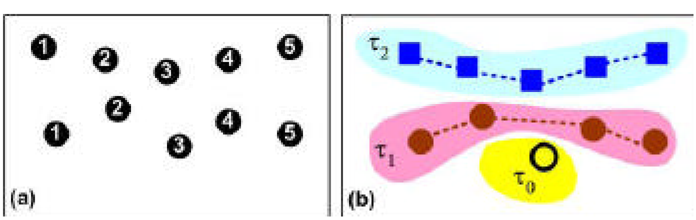
Ingemar Cox, A Review of Statistical Data Association Techniques for Motion Correspondence, International Journal of Computer Vision, vol. 10, pp. 53-66, 1993.

Standard approach

Find the optimal global explanation:

Given a set of noisy observations Y over a period of time.

An explanation is a partition of these observations $w = \{t_0, t_1, \ldots t_K\}$ where each part defines a track and t_0 contains all spurious observations (false alarms)



Seek $\underset{w \in \Omega}{\operatorname{argmax}}(p(w \mid Y))$

Formulation from Oh, Russell and Sastry, CDC-04

Defining $p(w \mid Y)$

Assumptions:

(1) each track behaves as a stochastic linear system:

$$x_{t_{i+1}} = Ax_{t_i} + \mathbf{h}$$
$$y_{t_i} = Cx_{t_i} + \mathbf{u}$$

(note that matrix A and noise term scaled according to the width of interval

(2) new objects and false alarms occur as Poisson processes

(3) objects disappear and are undetected with fixed probability at each time-step

---

For a given w at time-step t, assume:

e_t objects persist from t-1

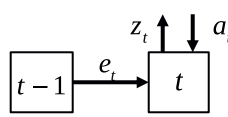
a_t new objects appear

z_t objects disappear

d_t objects detected

f_t false alarms

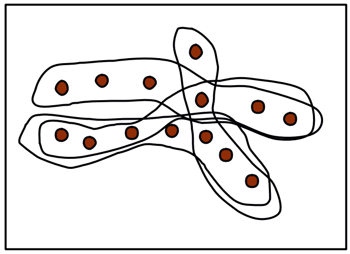
$u_t = e_t - z_t + a_t - d_t$ objects undetected



$$P(w \mid Y) = \frac{1}{Z} \prod_{t=1}^{T} \underbrace{p_Z^{z_t} (1-p_z)^{e_t - z_t}}_{\text{track terminations}} \underbrace{p_d^{d_t} (1-p_d)^{u_t}}_{\text{missing observations}} \underbrace{\frac{l_b^{a_t}}{a_t!} \frac{l_f^{f_t}}{f_t!}}_{\text{new objects and false alarms}} \prod_{t \in w \setminus \{t_0\}} \prod_{i=1}^{|t|-1} \underbrace{?\,(t(t_{i+1}) \mid C\overline{x}_{t_{i+1}}(t), B_{t_{i+1}}(t))}_{\text{stochastic linear system}}$$

Integer Programming
Morefield, IEEE-TAC 1977

- Create a large set of feasible tracks F (a covering), many of which will be inconsistent with one another.



- Seek the optimal partition from a subset of these tracks + false alarms

$$\operatorname*{argmax}_{\substack{w \subset F \\ w \in \Omega}} (p(w \mid Y))$$

Example
from *Leibe, Schindler, and Van Gool, ICCV 2007*

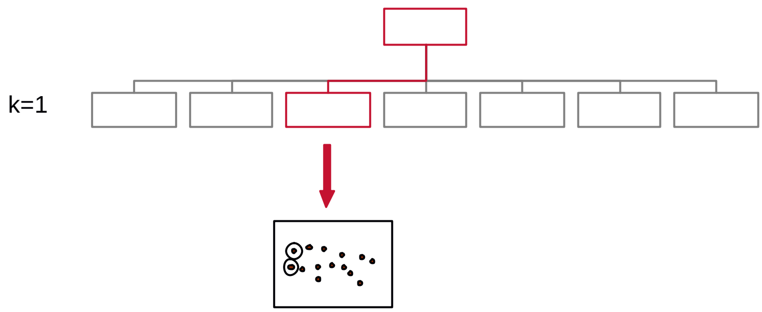
Uses a trained pedestrian detector operating on each frame

from http://www.vision.ee.ethz.ch/~bleibe/index.html

Multiple-Hypothesis Tree (MHT)

Reid, IEEE-TAC 1979

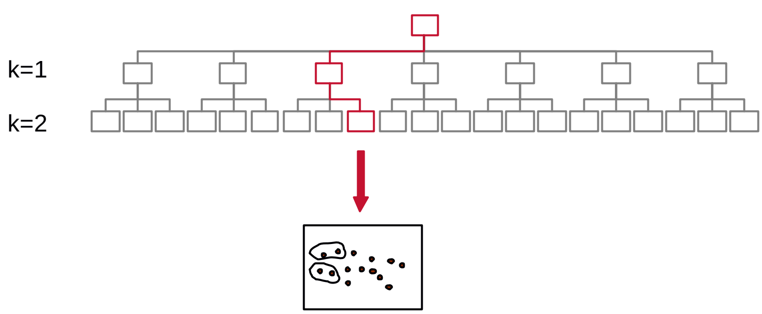
- Iteratively extend partial tracks at each time-step
- Pursue multiple hypotheses where there is ambiguity
- Prune unlikely hypotheses to keep search tractable



Multiple-Hypothesis Tree (MHT)

Reid, IEEE-TAC 1979

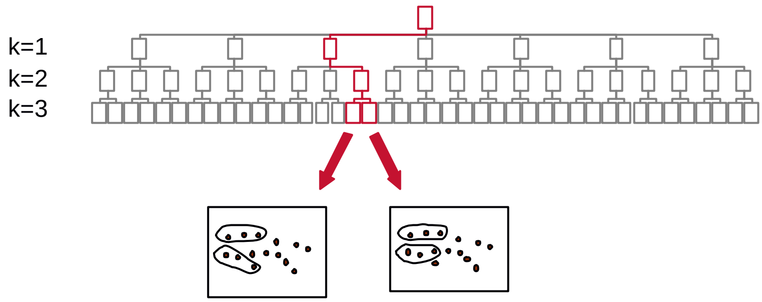
- Iteratively extend partial tracks at each time-step
- Pursue multiple hypotheses where there is ambiguity
- Prune unlikely hypotheses to keep search tractable



Multiple-Hypothesis Tree (MHT)

Reid, IEEE-TAC 1979

- Iteratively extend partial tracks at each time-step
- Pursue multiple hypotheses where there is ambiguity
- Prune unlikely hypotheses to keep search tractable



Markov Chain Monte Carlo Data Association

Oh, Russell, and Sastry, CDC-04, 2004

- Draw samples from posterior $p(w \mid Y)$ and select the maximum. Use Markov Chain Monte Carlo (MCMC) to do this efficiently.

initialise w
repeat many times
  Sample w’ from proposal distribution $q(w, w')$
  Replace w by w’ with (acceptance) probability:

$$A(w,w') = \min\left(1, \frac{p(w' \mid Y)q(w',w)}{p(w \mid Y)q(w,w')}\right)$$

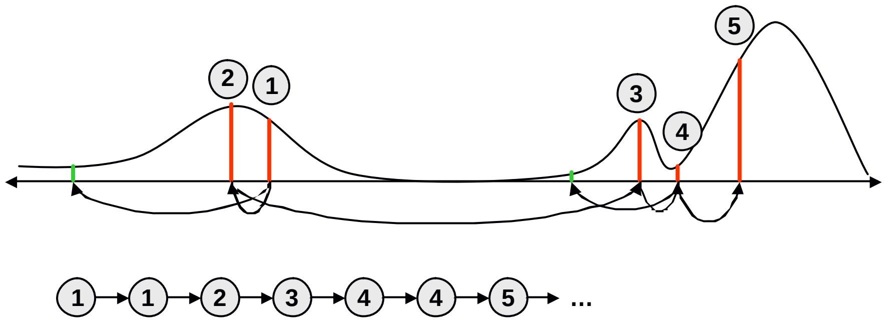
end

Introduction to MCMC

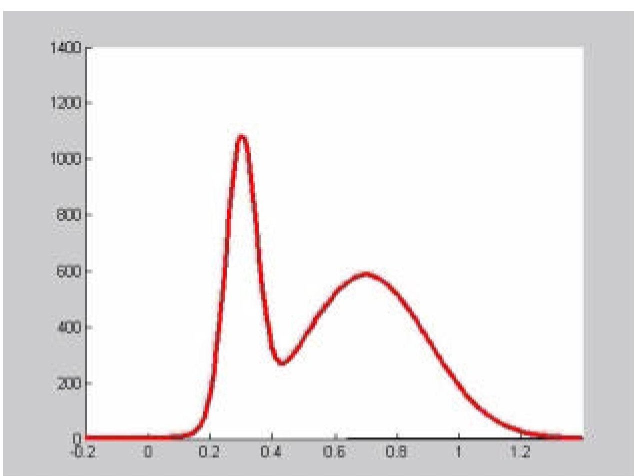
MCMC – Markov Chain Monte Carlo

When to use?

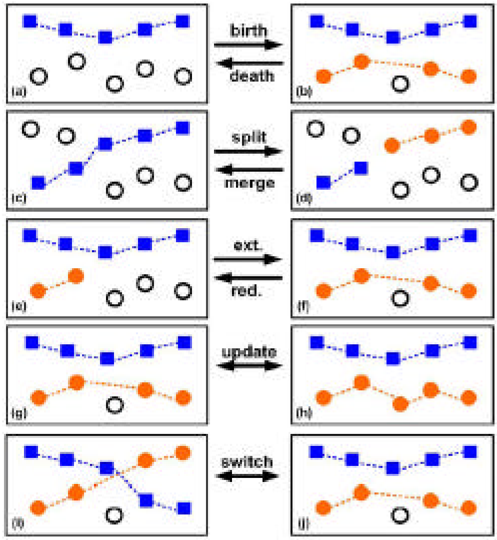
- You can't sample from the distribution itself
- Can evaluate it at any point



Introduction to MCMC



MCMC moves

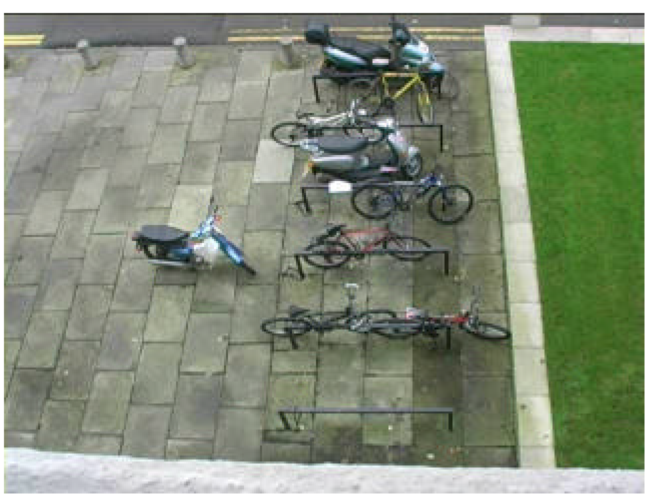


From Oh, Russell and Sastry, CDC-04, 2004

Detecting people parking and collecting bikes

Damen & Hogg, BMVC 2007

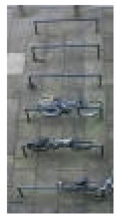
Task: linking people dropping-off and picking-up bikes

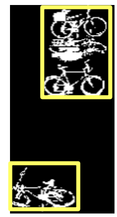


Method

- Track people (+/- bikes) entering and leaving the rack area
- Detect new clusters of dropped & picked bikes each time the rack area becomes empty
- List the possible new drop, pick and pass-through events, assuming people entering the rack, drop or pick no more than one bike
- Find optimal set of linked drop and pick events

$$\arg\max_{w}(p(w \mid Y))$$





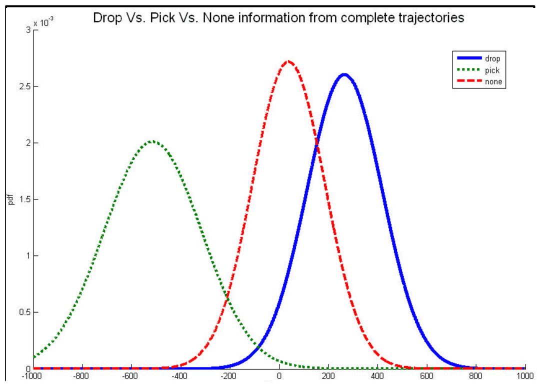
Defining $p(w \mid Y)$

Based on:

- Change in the area of person-blobs between entering and leaving rack
- Proximity of people to bike clusters
- Similarity of bike clusters between drop and pick
- Prior probabilities for the different events

$$p(w) = \left(\prod_{i=1}^{3} p_{e_i}{}^{n_{e_i}} \right) p_{D*}{}^{n_{D*}} \, p_{*P}{}^{n_{*P}} \, p_{DP}{}^{n_{DP}}$$

Likelihood of a person dropping, picking or passing through



Area difference of person-blob entering and leaving the rack

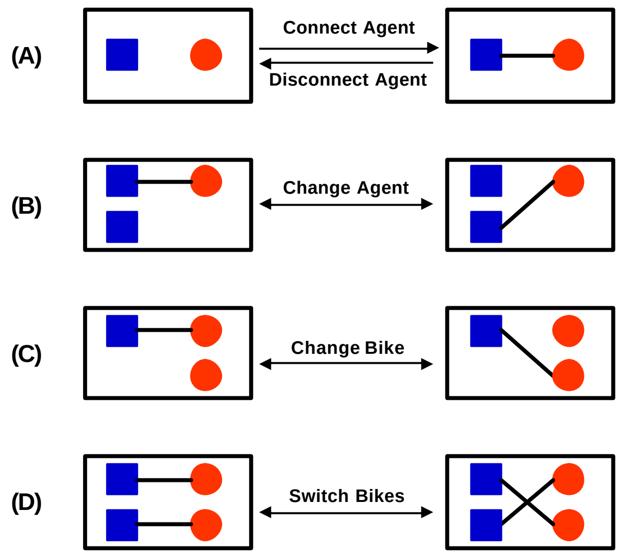
Likelihood of a drop/pick linkage

Drop

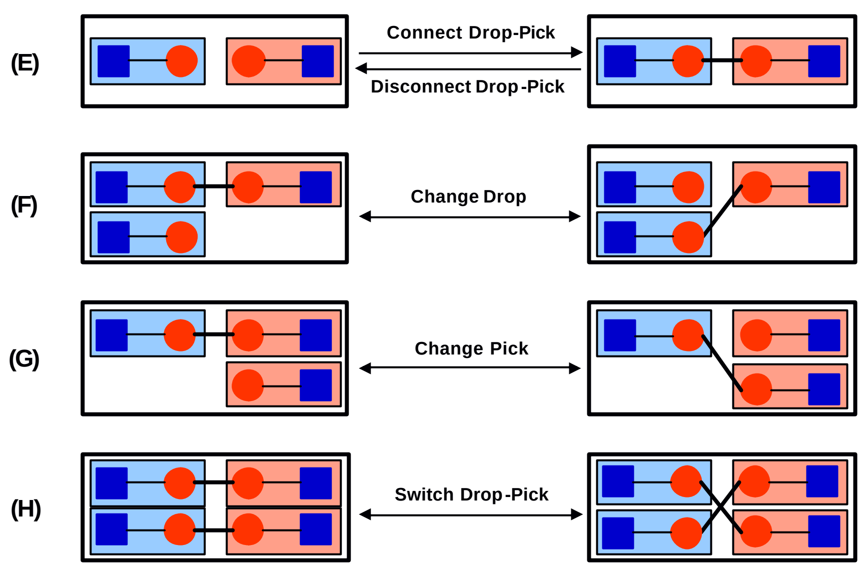
Pick

Pixel-wise comparison of bike-clusters

Possible moves - 1



Possible moves - 2



Results

Experiments	MHT (k=10)	RJMCMC (10 iterations)
1 hour (43 events)	93.10	93.10
9.5 hours (39 events)	93.75	94.53

% correct drop-pick connections

Summary

A wider scope of interest provides new ways of thinking about problems within a narrower focus.