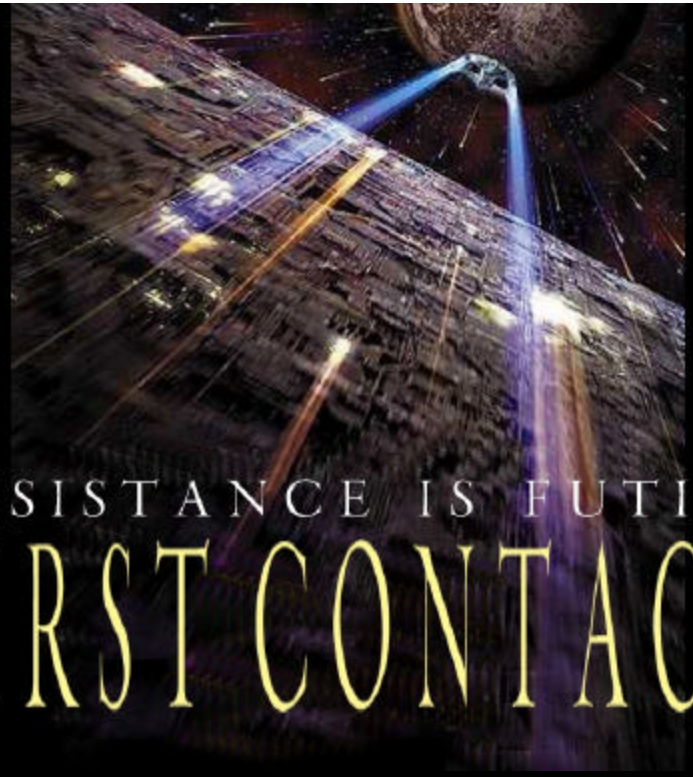


AN ACTIVE VISION APPROACH TO OBJECT SEGMENTATION

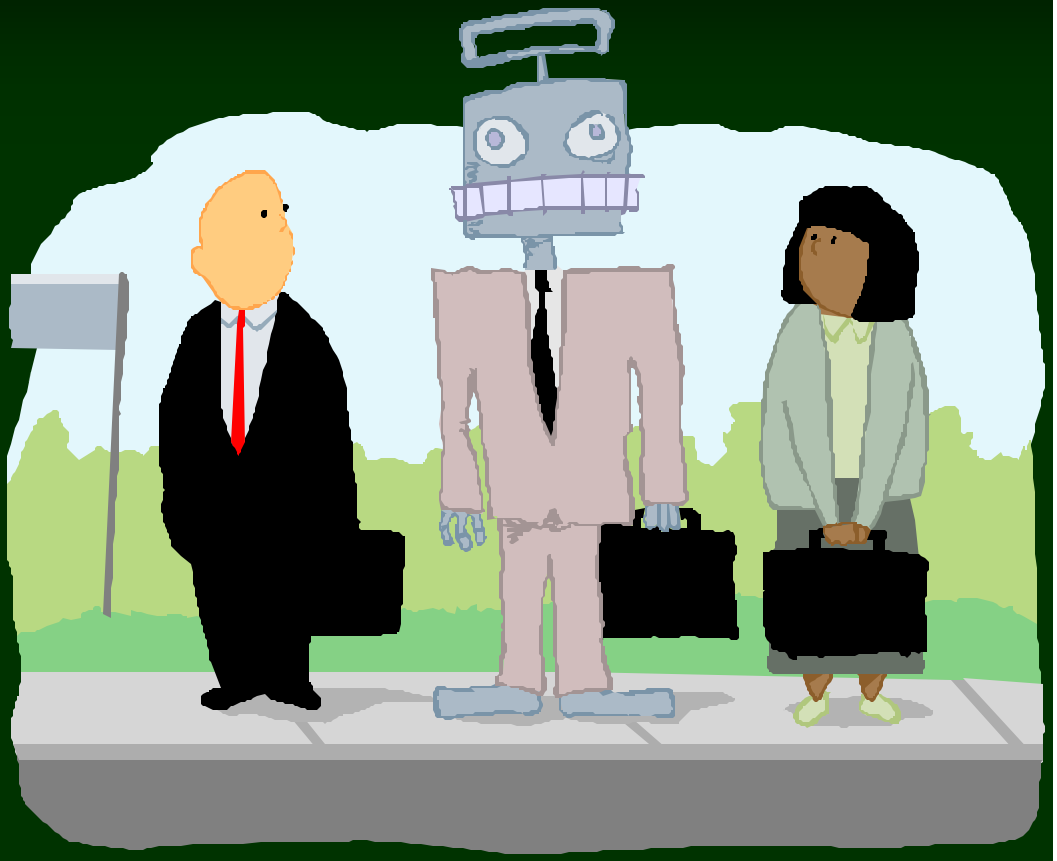
– Paul Fitzpatrick –

MIT CSAIL

flexible perception

Humanoid form is general-purpose, mechanically flexible

Robots that really live and work amongst us will need to be as general-purpose and adaptive perceptually as they are mechanically



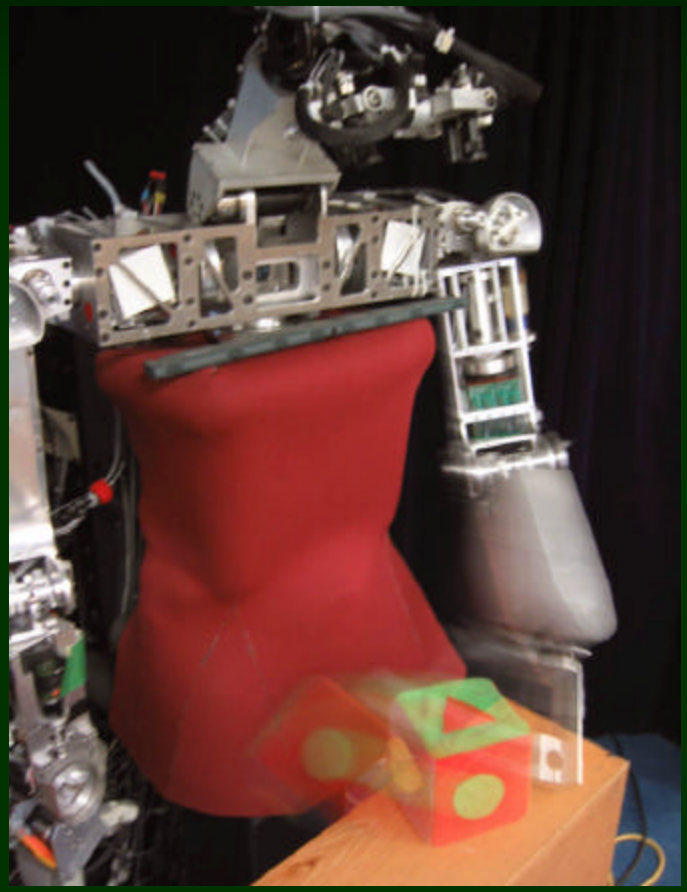
flexible visual perception

In robotics, vision is often used to guide manipulation

But manipulation can also guide vision

Important for…

- Correction – detecting and recovering from incorrect perception
- Experimentation – disambiguating inconclusive perception
- Development – creating or improving perceptual abilities through experience

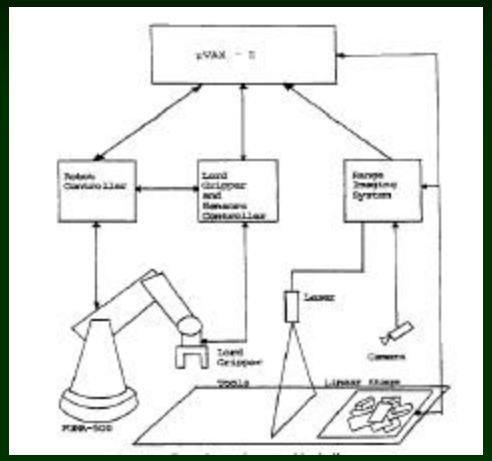


flexible visual object perception

Tsikos, Bajcsy, 1991

"Segmentation via manipulation"

Simplify cluttered scenes by moving overlapping objects

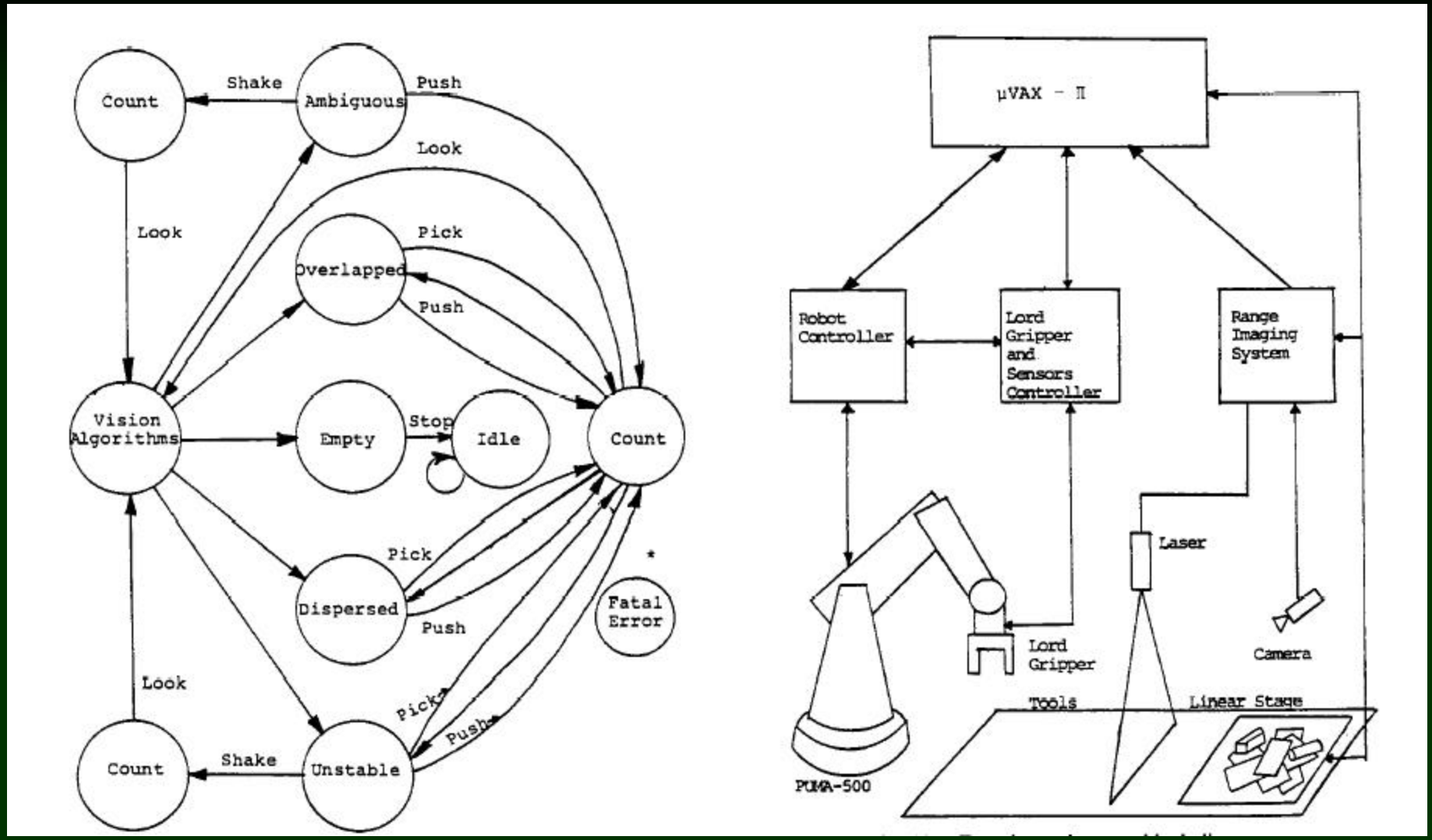


Sandini et al, 1993

"Vision during action"

Interpret motion during manipulation to deduce object boundaries

segmentation via manipulation



Tsikos, Bajcsy, 1991

vision during action

Sandini et al, 1993

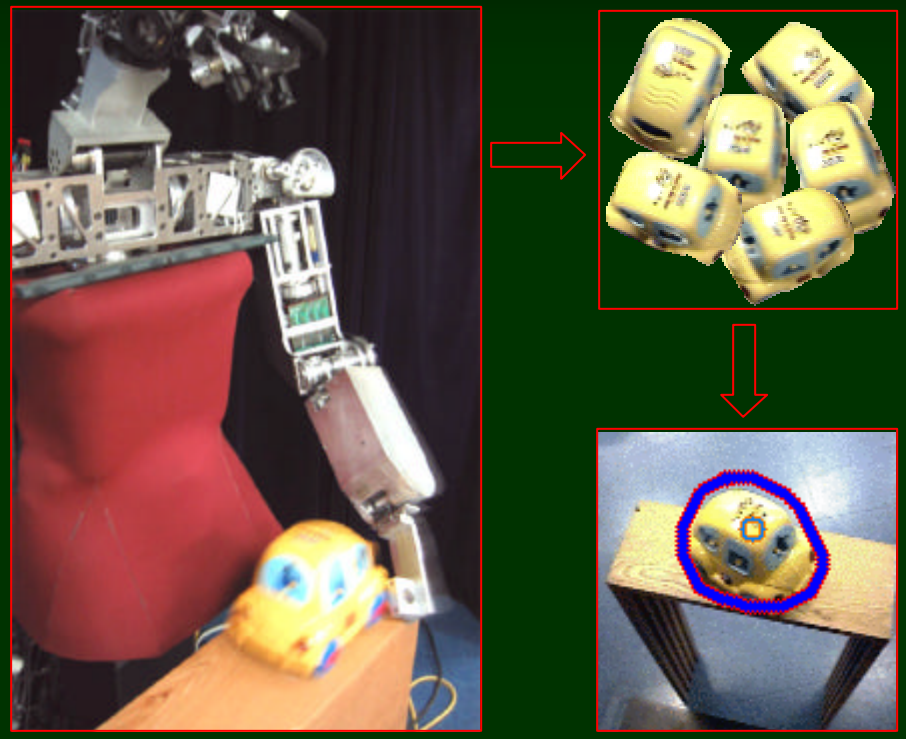
active segmentation

Object boundaries are not always easy to detect visually

Solution: Cog sweeps arm through ambiguous area

Any resulting object motion helps segmentation

Robot can learn to recognize and segment object without further contact



active segmentation

evidence for segmentation

Areas where motion is observed upon contact

- classify as ‘foreground’

Areas where motion is observed immediately before contact

- classify as ‘background’

Textured areas where no motion was observed

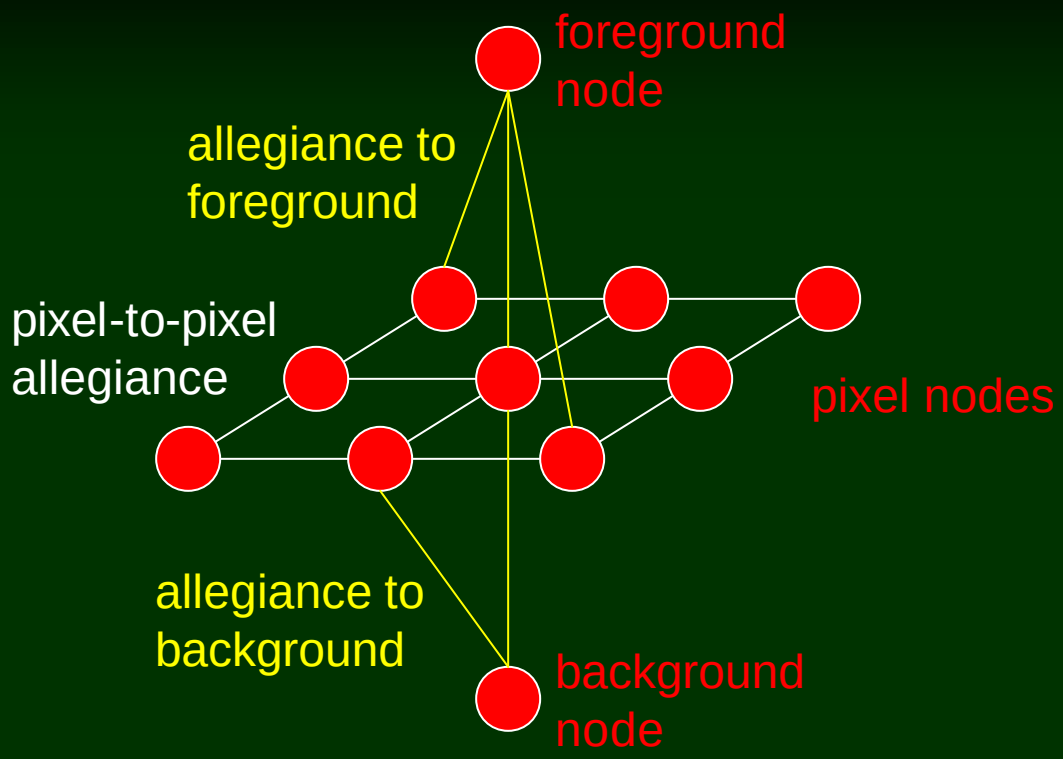
- classify as ‘background’

Textureless areas where no motion was observed

- no information

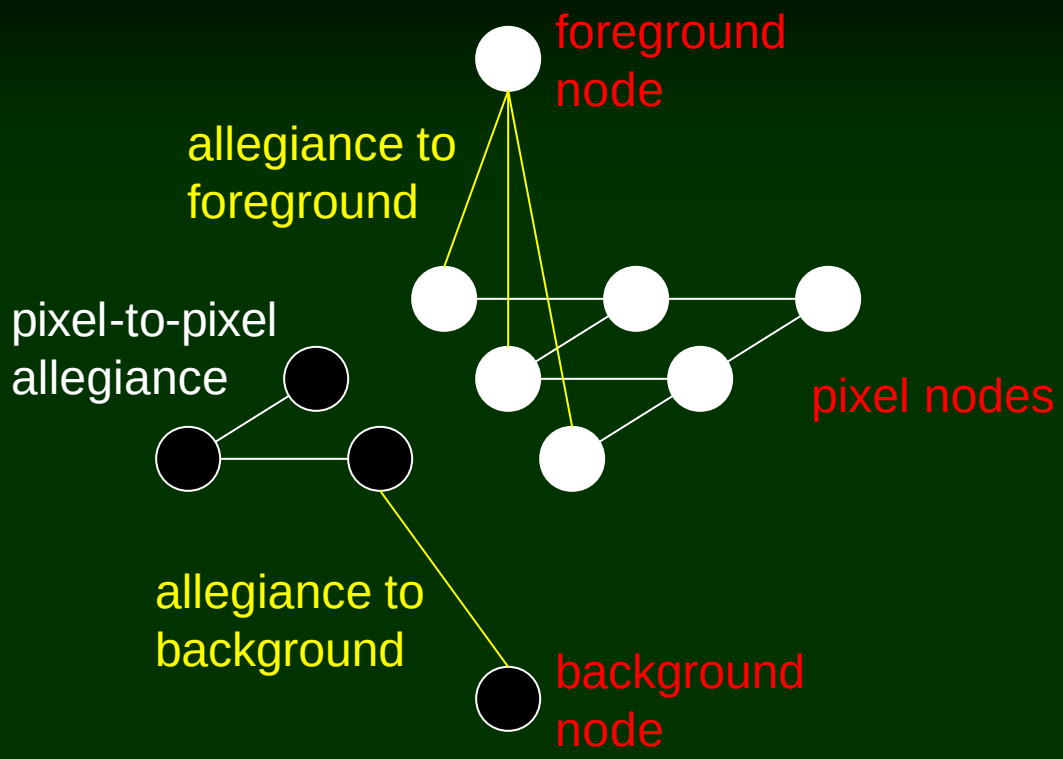
No need to model the background!

minimum cut



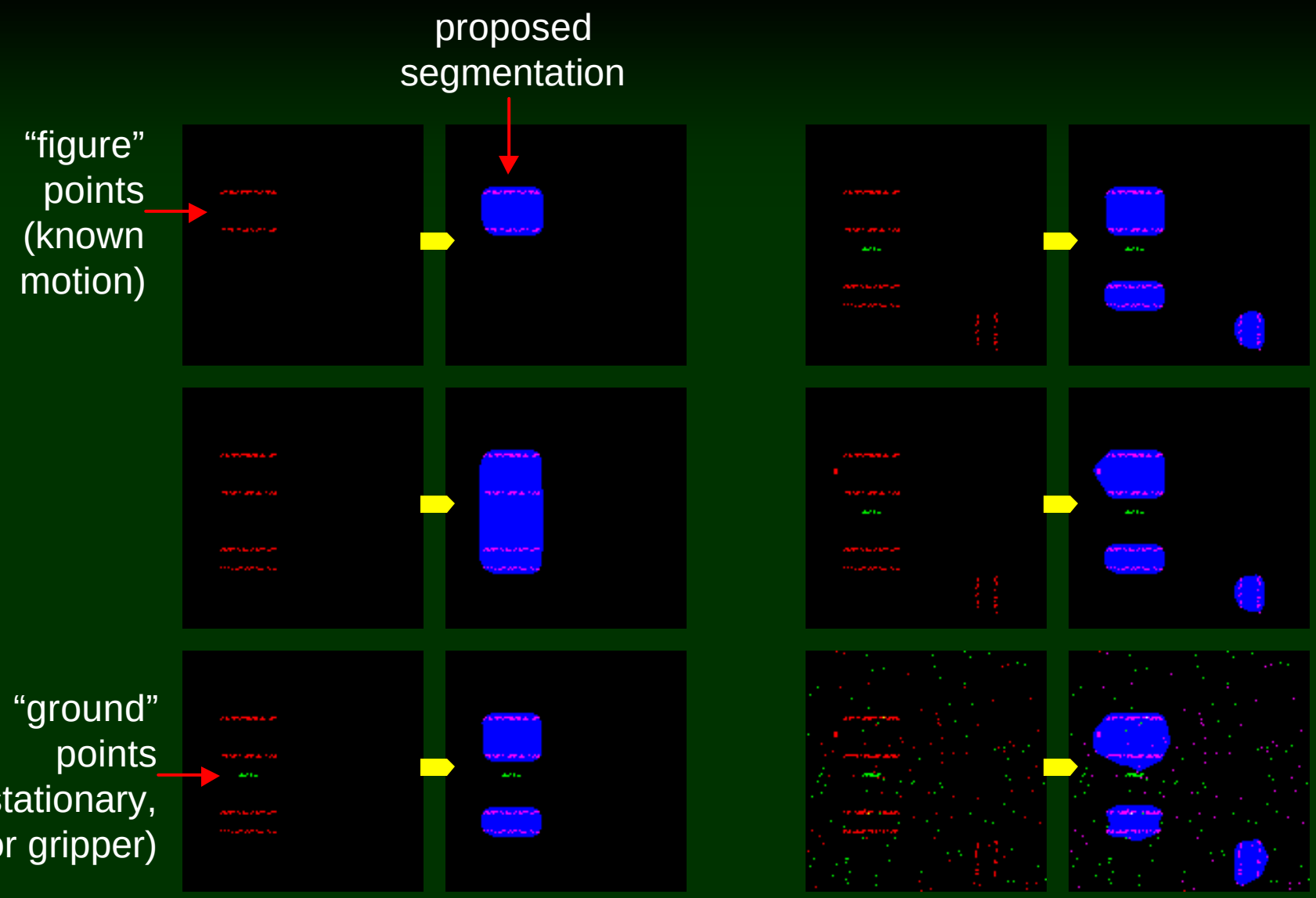
“allegiance” = cost of assigning two nodes to different layers (foreground versus background)

minimum cut

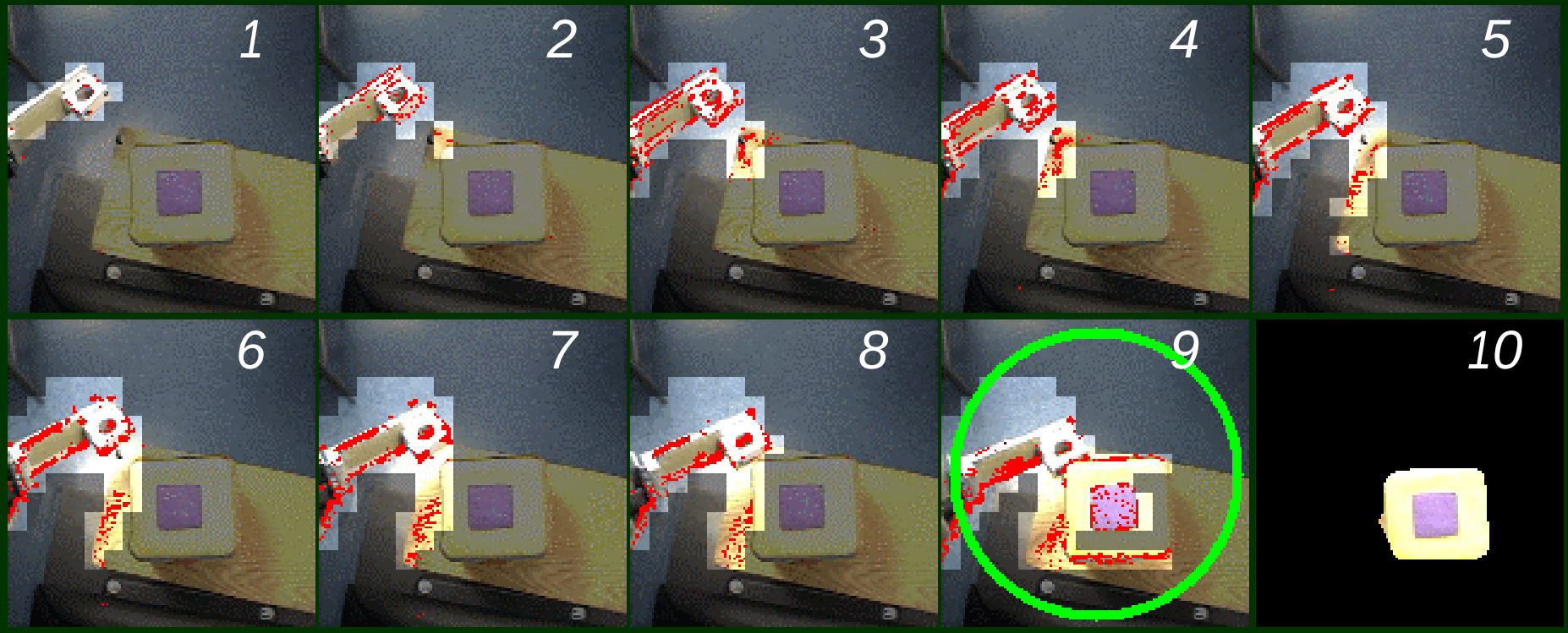


“allegiance” = cost of assigning two nodes to different layers (foreground versus background)

grouping (on synthetic data)



point of contact



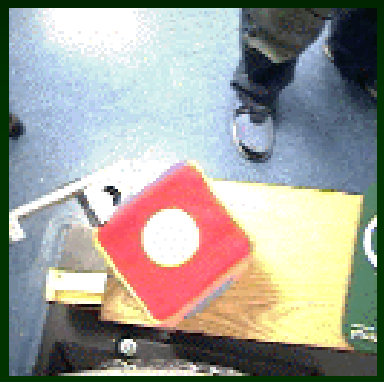
Motion spreads continuously
(arm or its shadow)

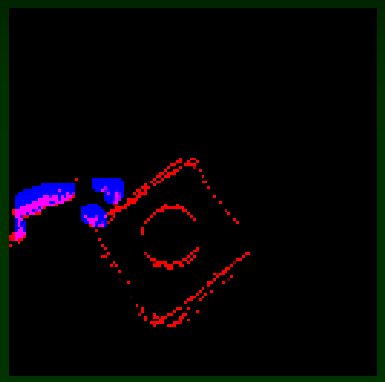
Motion spreads suddenly, faster than the arm itself → contact

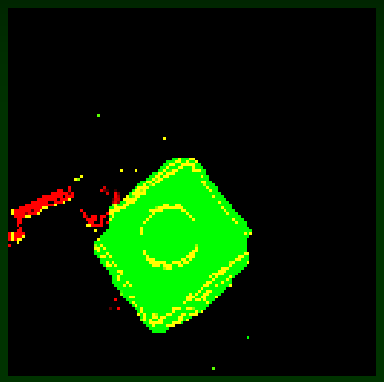
point of contact

segmentation examples

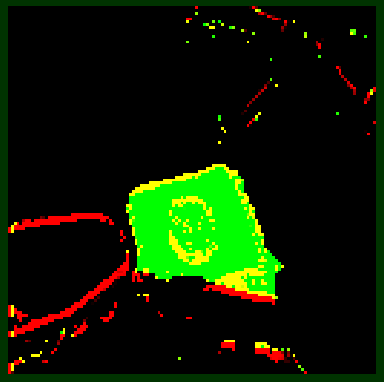
Side tap







Back slap

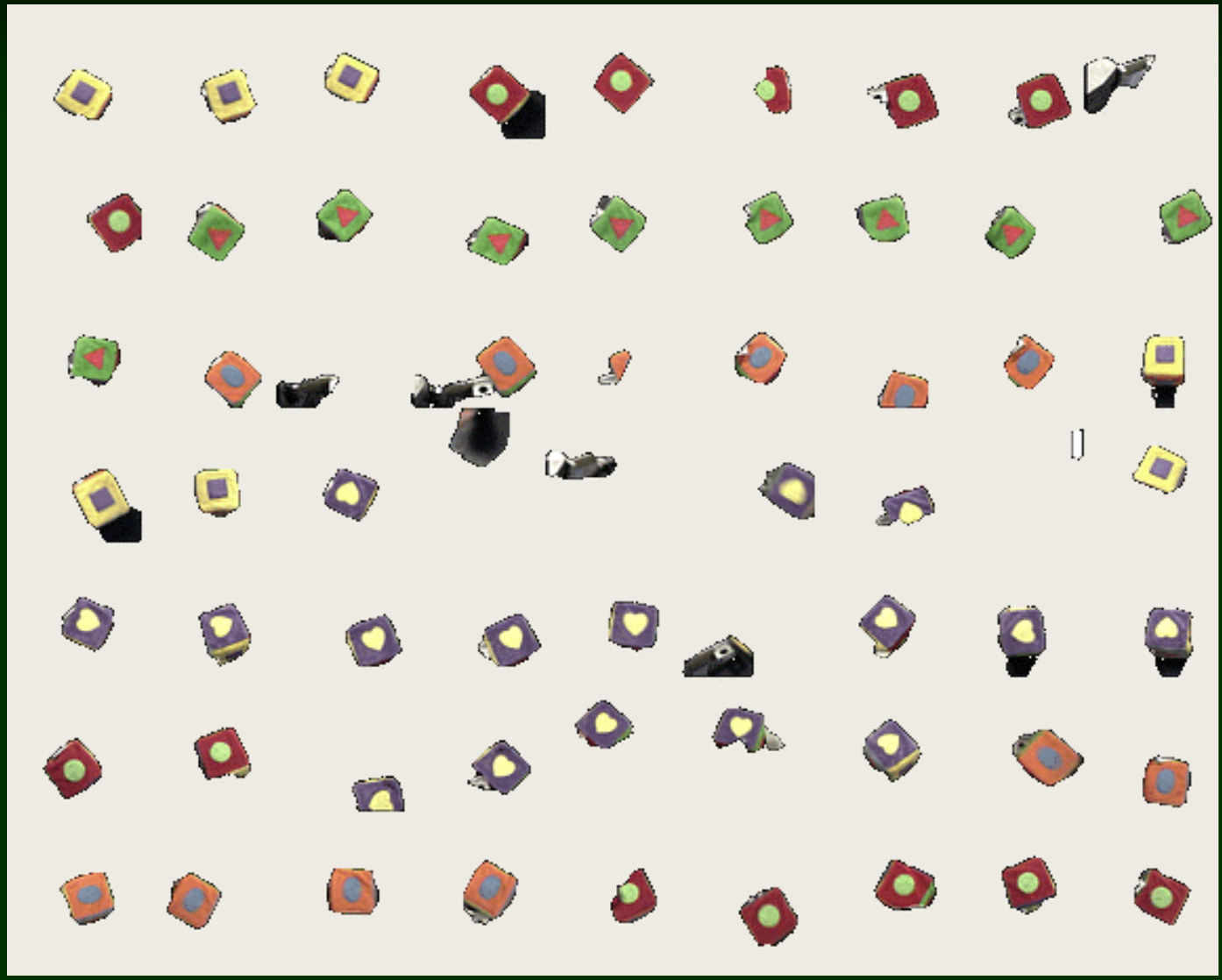


Impact event

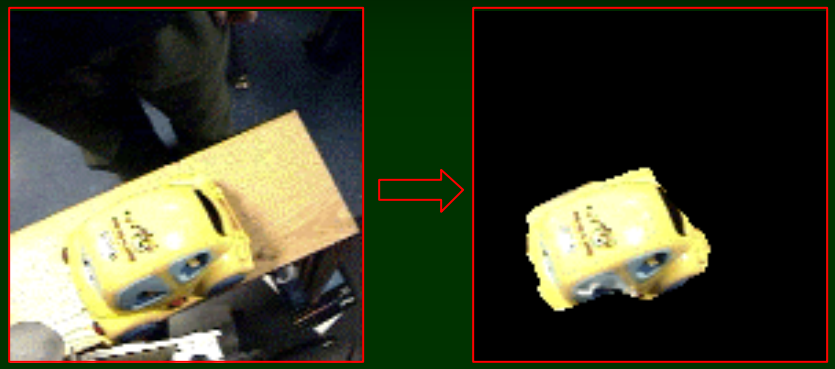
Motion caused
(red = novel,
Purple/blue = discounted)

Segmentation
(green/yellow)

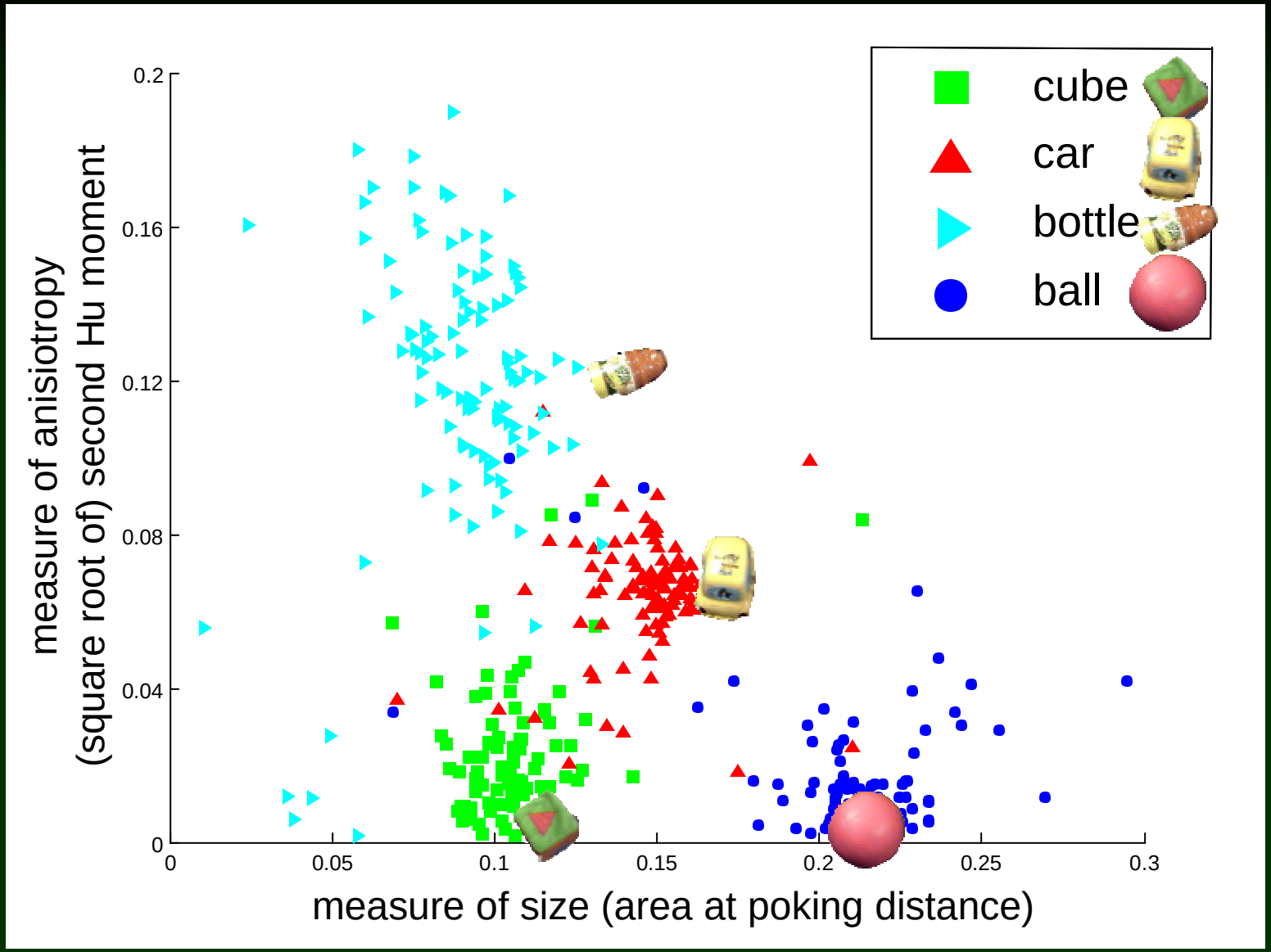
segmentation examples



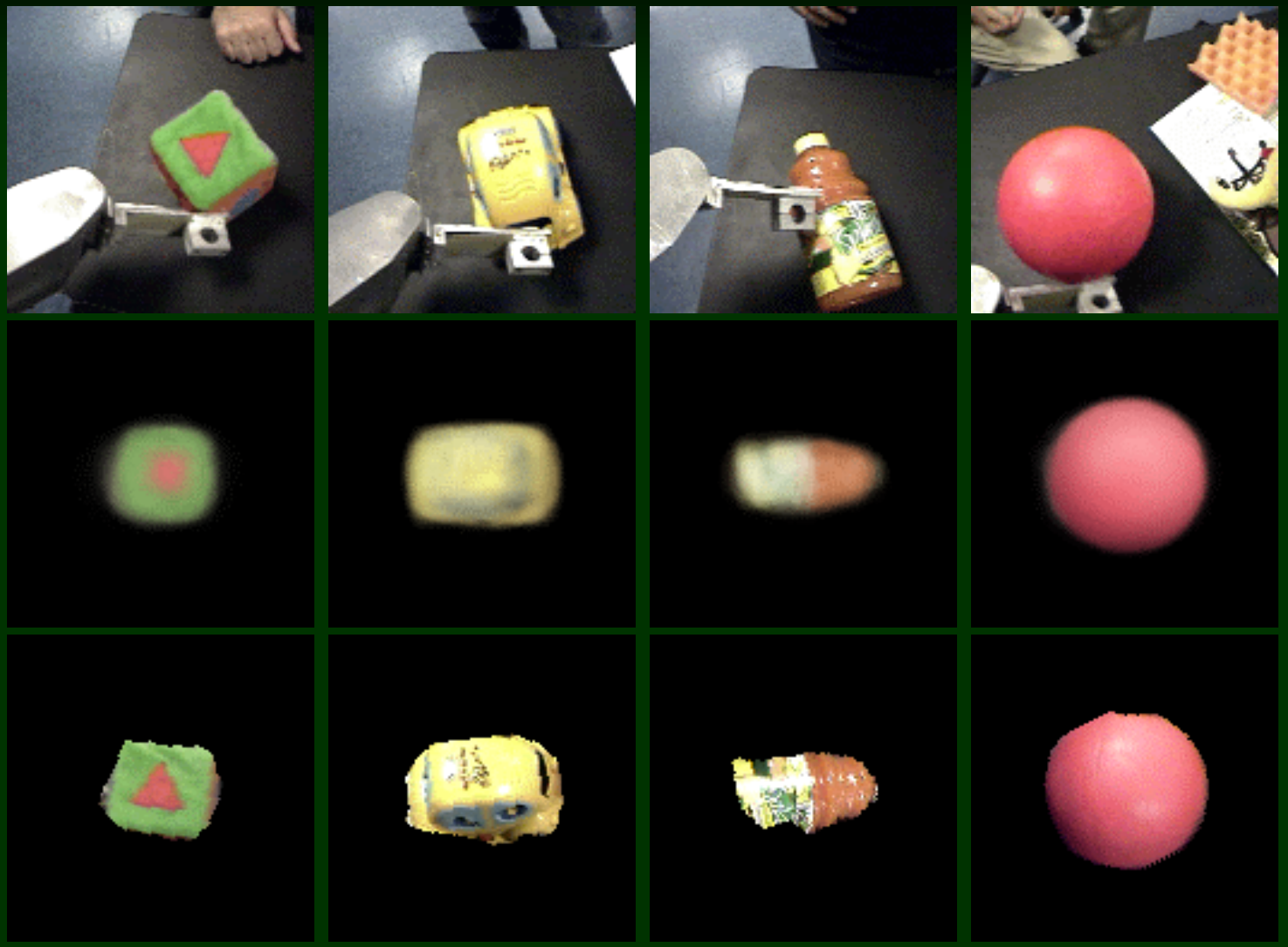
segmentation examples



boundary fidelity



signal to noise



active segmentation

Not always practical!

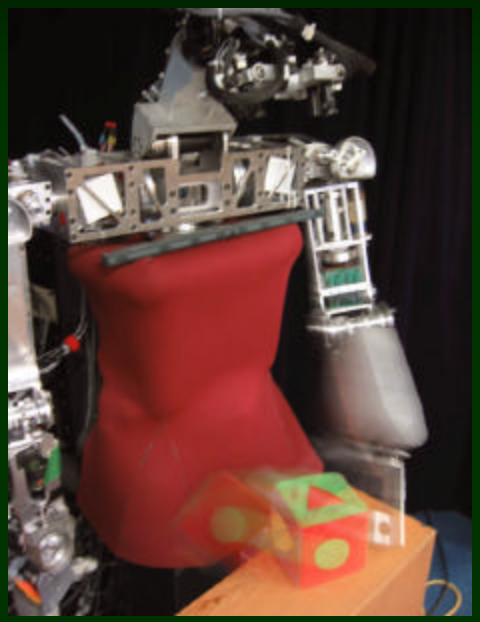
No good for objects the robot can view but not touch

No good for very big or very small objects

But fine for objects the robot is expected to manipulate

Head segmentation the hard way!

other approaches



robot manipulation,
first person perspective

(Paul Fitzpatrick,
Giorgio Metta)

human manipulation,
external perspective

(Artur Arsenio)

human manipulation,
first person perspective

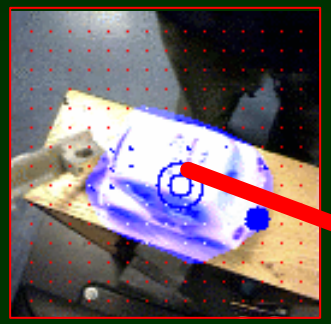
(Charlie Kemp)

from first contact to close encounters

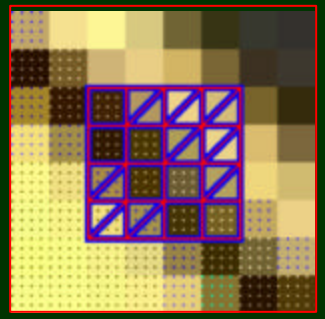
active segmentation

segmentation catalog

affordance exploitation
(rolling)



edge catalog



object detection
(recognition, localization,
contact-free segmentation)

manipulator detection
(robot, human)

object recognition

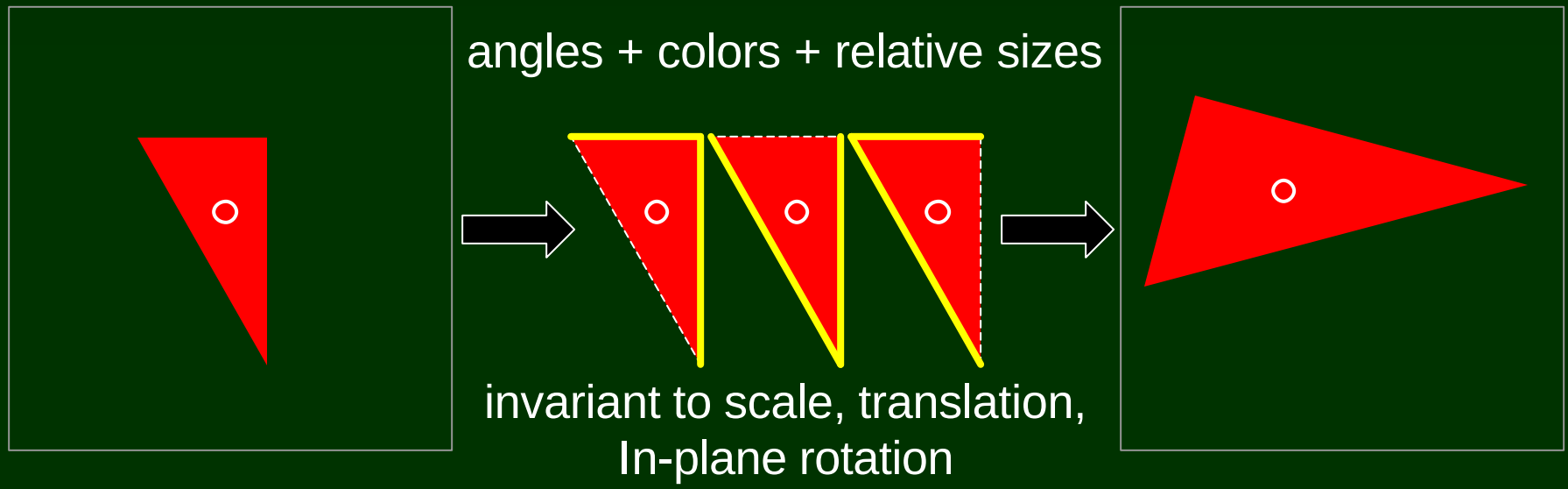
Geometry-based

- Objects and images modeled as set of point/surface/volume elements
- Example real-time method: store geometric relationships in hash table

Appearance-based

- Objects and images modeled as set of features closer to raw image
- Example real-time method: use histograms of simple features (e.g. color)

geometry+appearance



Advantages: more selective; fast

Disadvantages: edges can be occluded; 2D method

Property: no need for offline training

details of features

Distinguishing elements:

- Angle between regions (edges)
- Position of regions relative to their projected intersection point (normalized for scale, orientation)
- Color at three sample points along line between region centroids

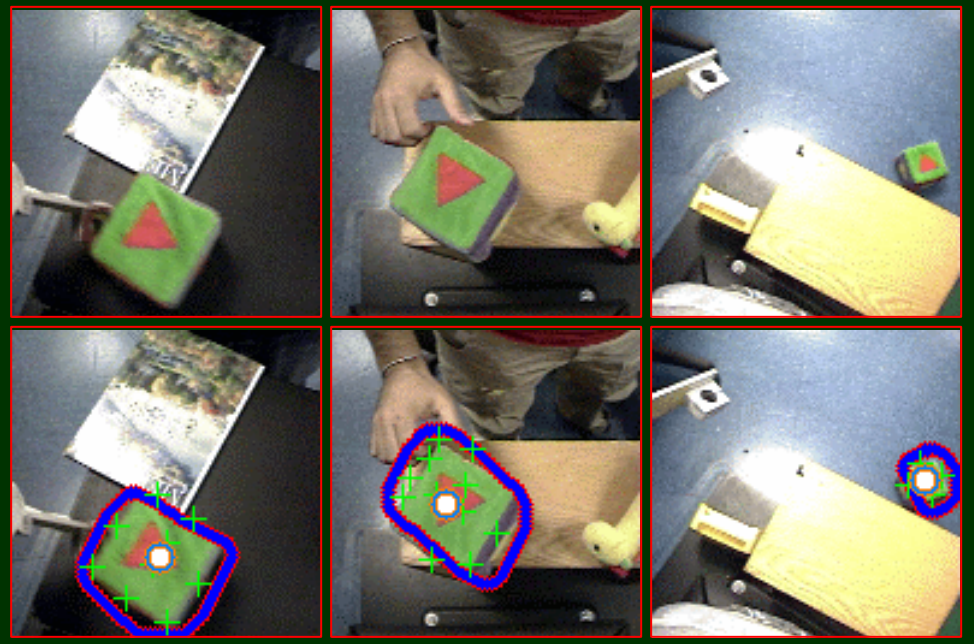
Output of feature match:

- Predicts approximate center and scale of object if match exists

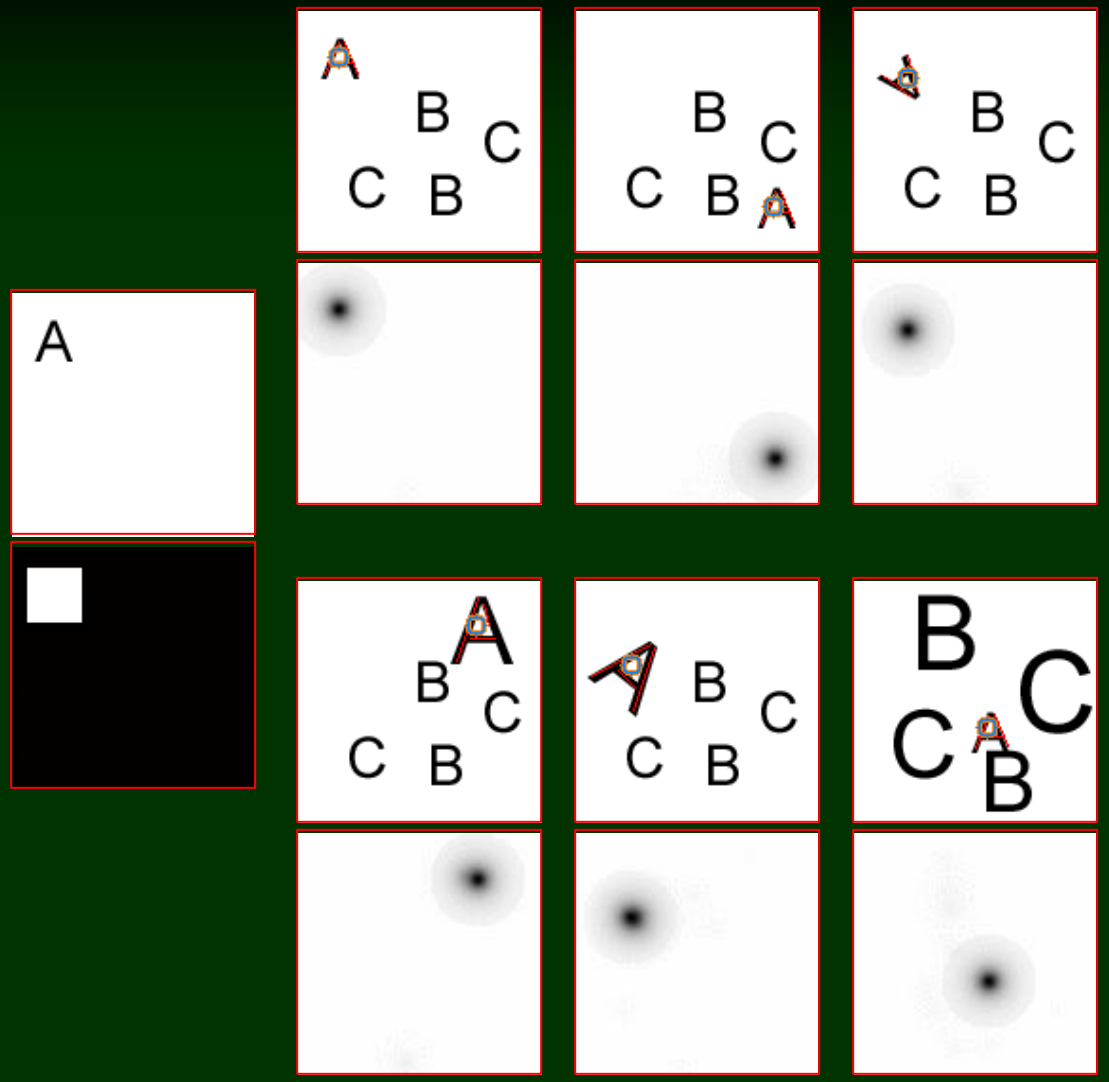
Weighting for combining features:

- Summed at each possible position of center; consistency check for scale
- Weighted by frequency of occurrence of feature in object examples, and edge length

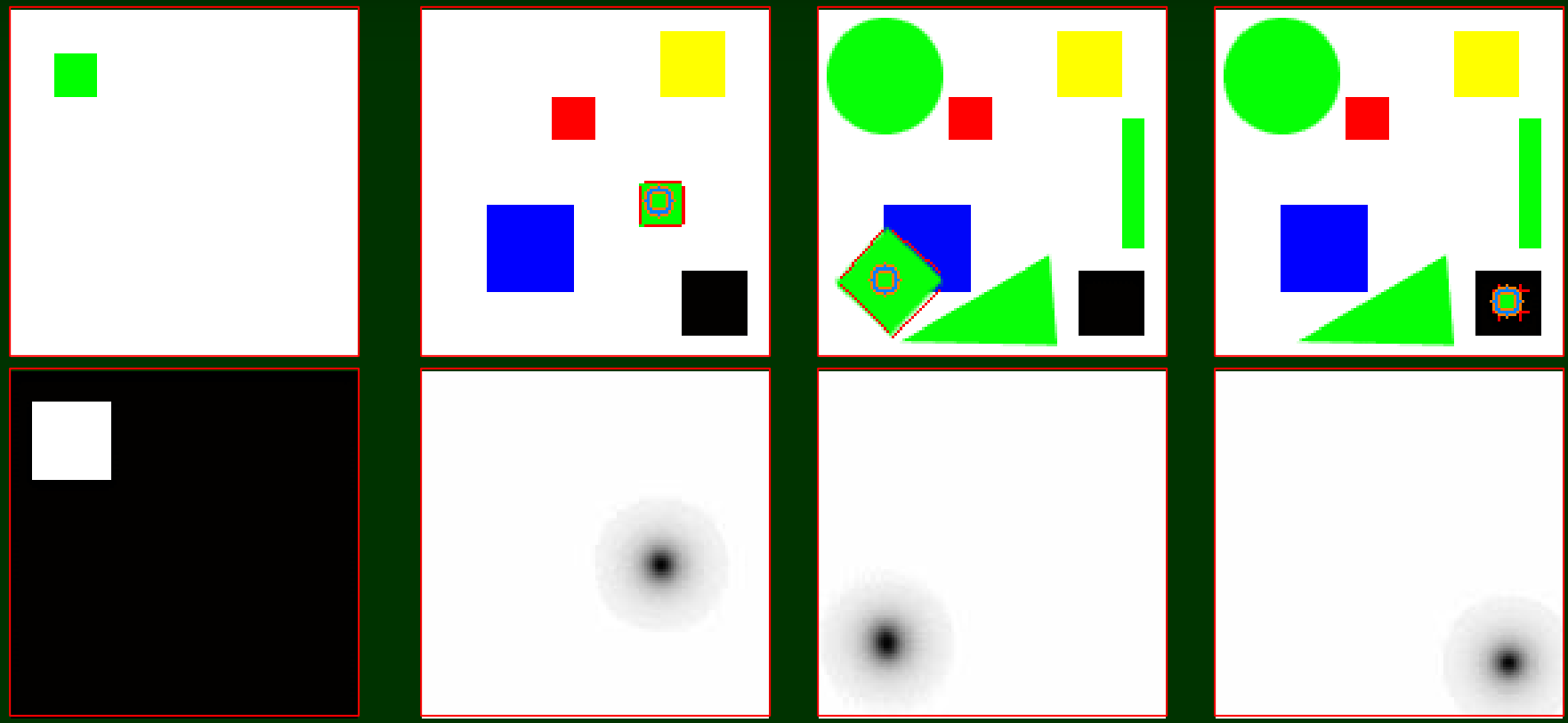
real object in real images



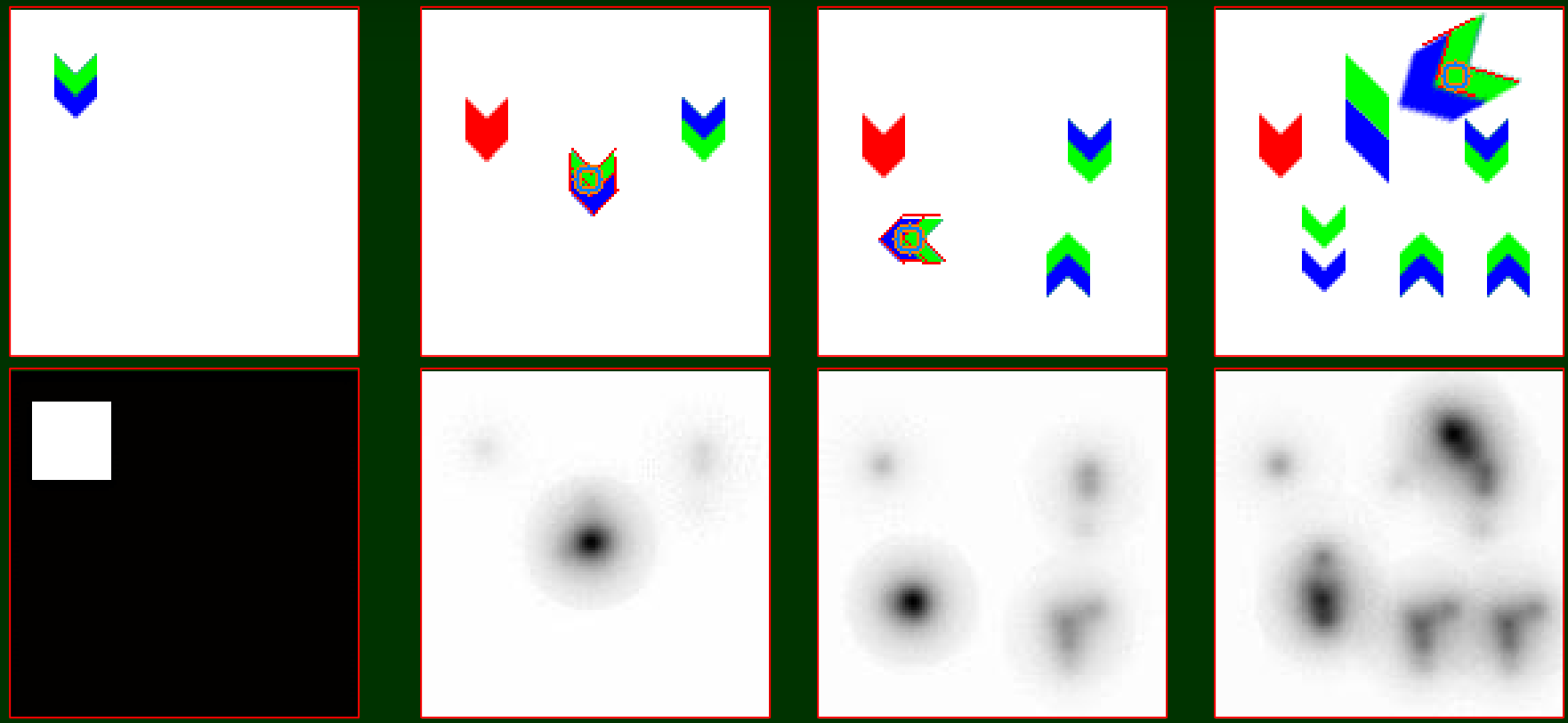
other examples



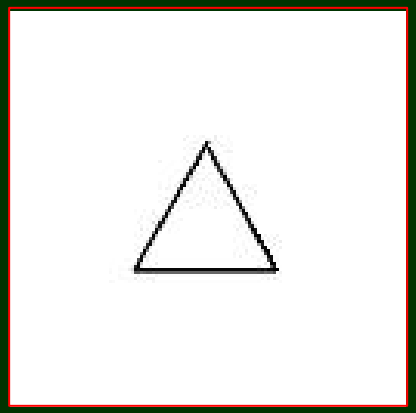
other examples



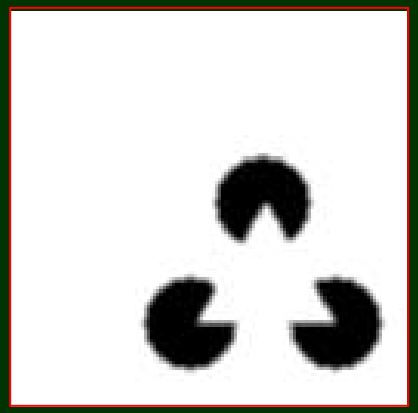
other examples



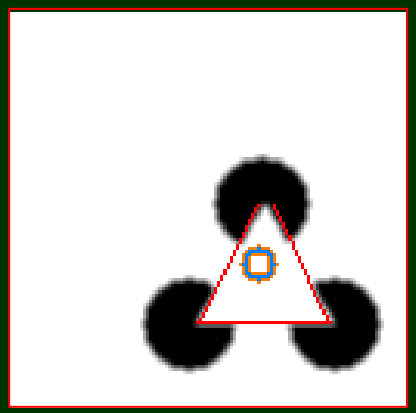
just for fun

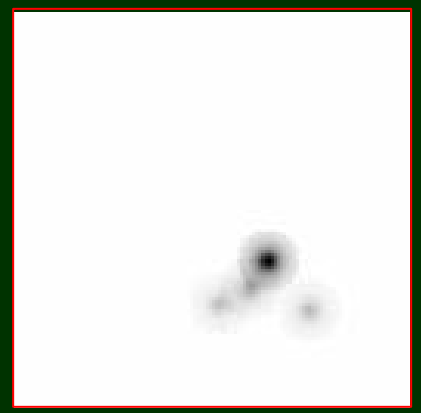


look for this…



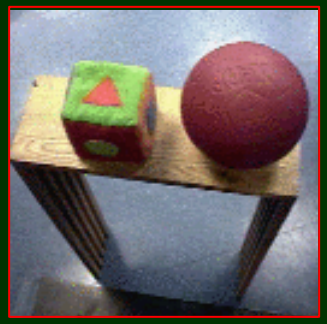
…in this



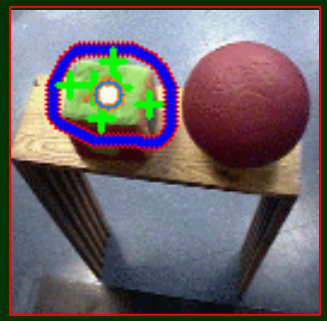


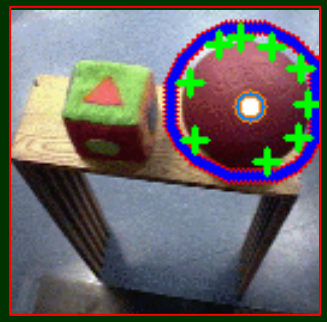
result

multiple objects

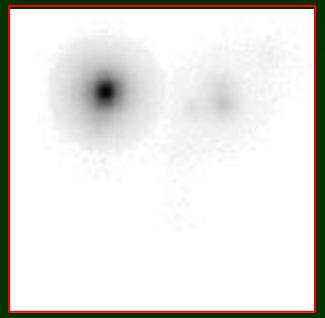


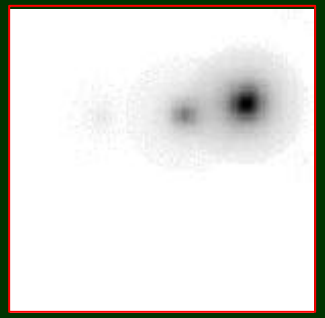
camera image





implicated edges
found and grouped





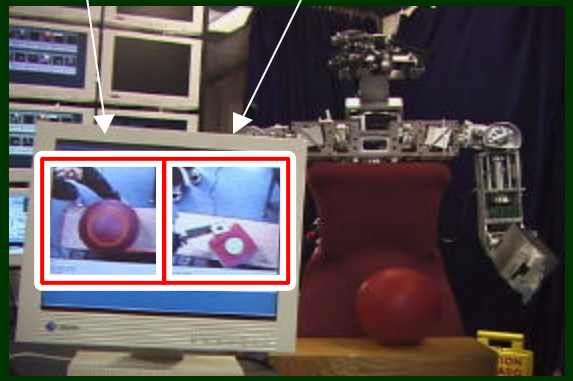
response for
each object

yellow on yellow

first time seeing a ball

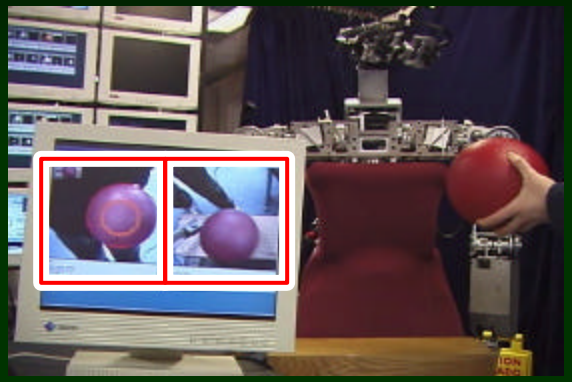
robot's current view

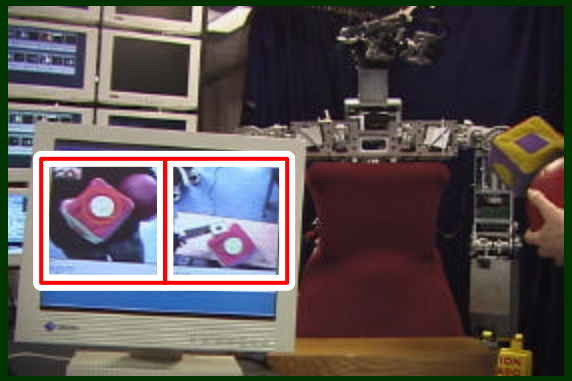
recognized object (as seen during poking)



sees ball, "thinks" it is cube

pokes, segments ball

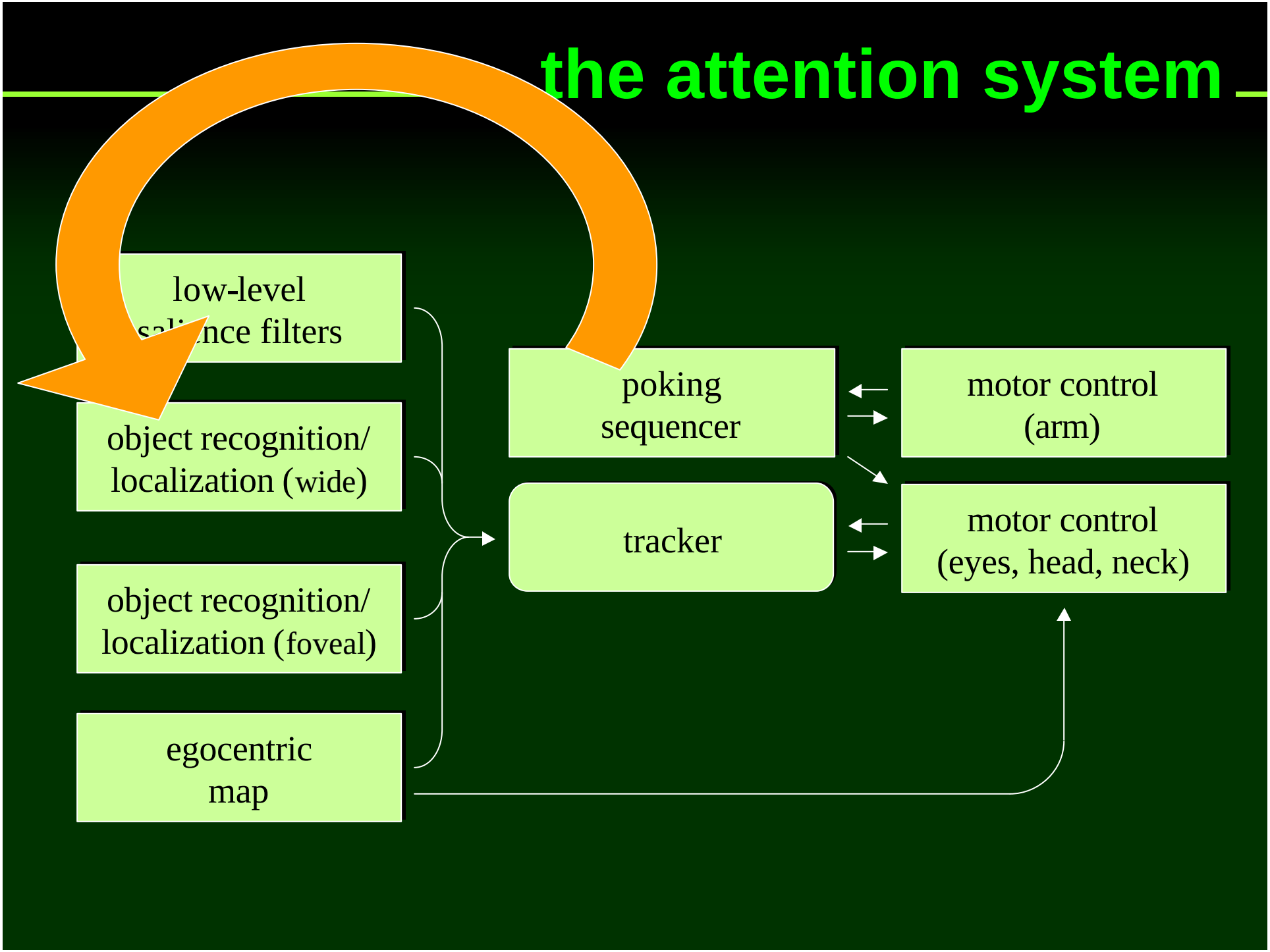




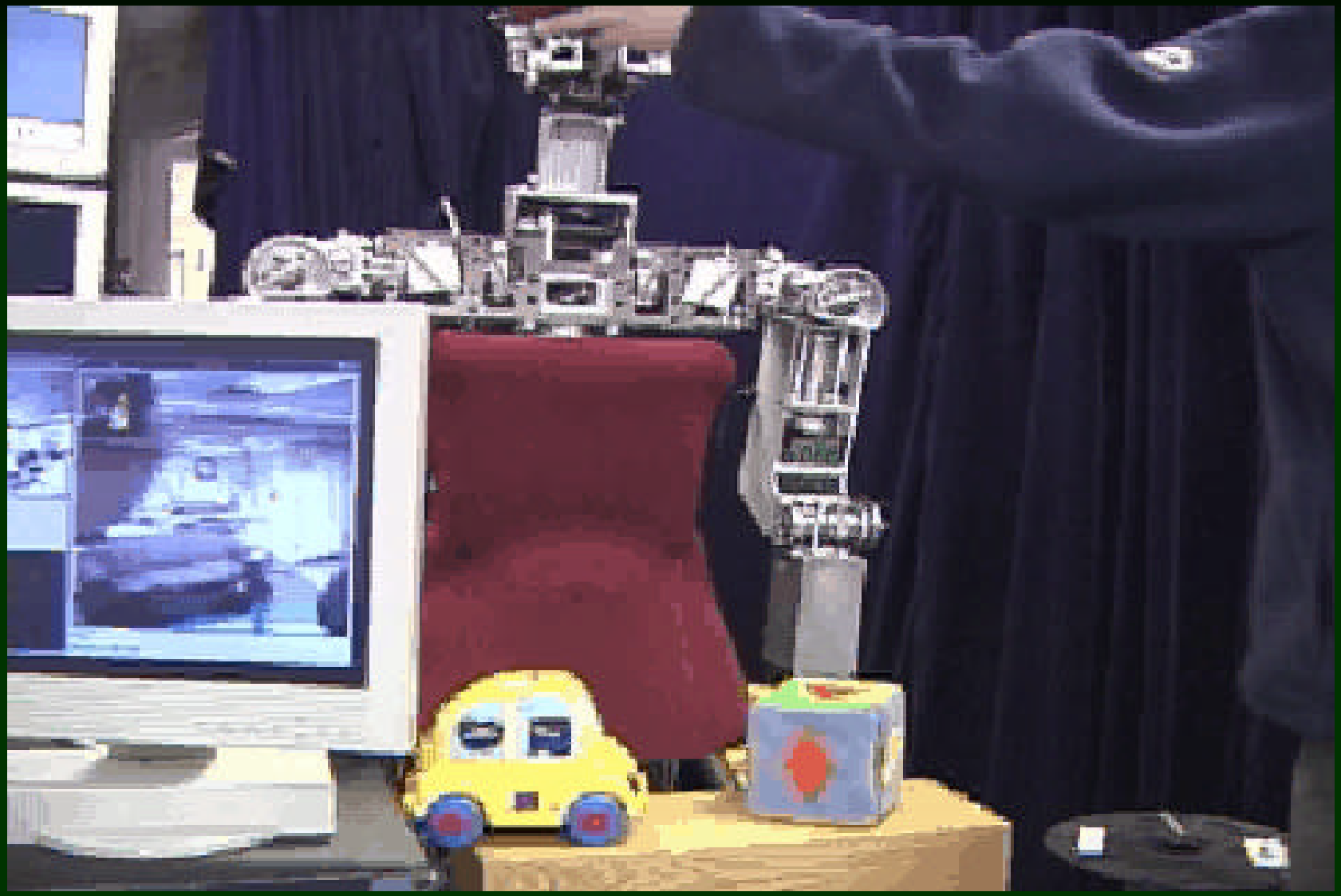
correctly differentiates ball and cube

open object recognition

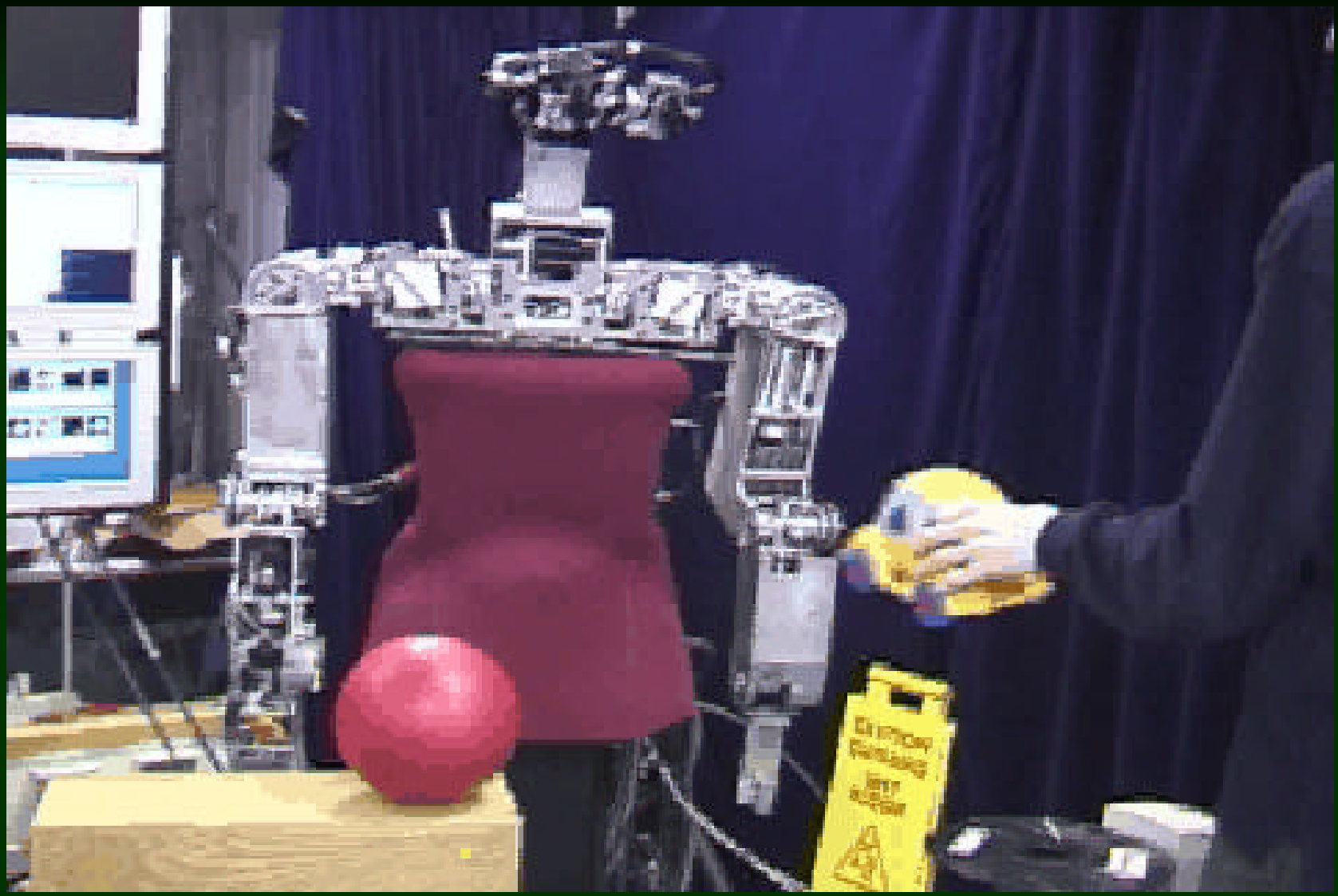
the attention system



attention



and so on…



finding manipulators

Analogous to finding objects

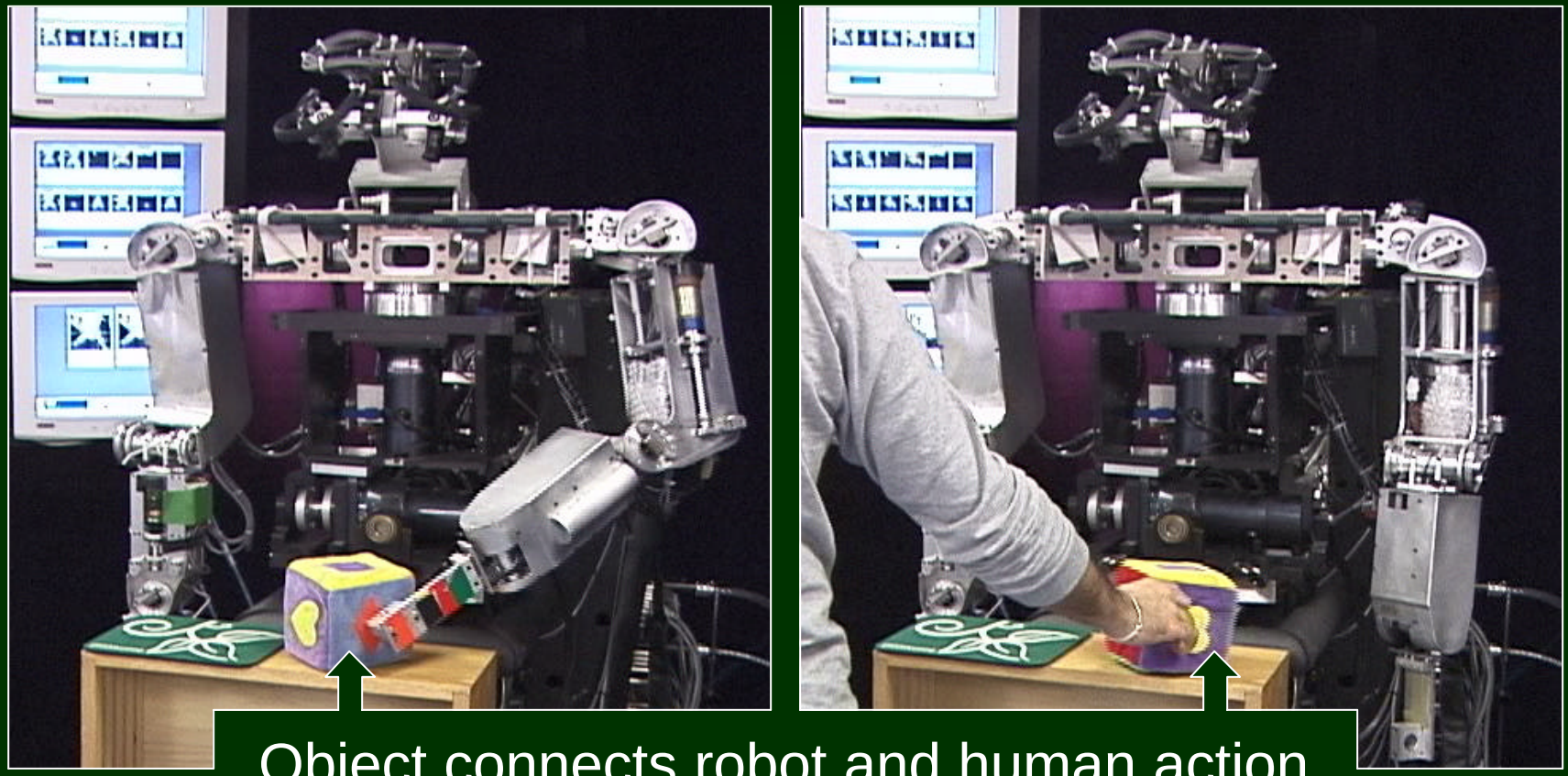
Object

- *Definition*: physically coherent structure
- *How to find one*: poke around and see what moves together

Actor

- *Definition*: something that acts on objects
- *How to find one*: see what pokes objects

similar human and robot actions



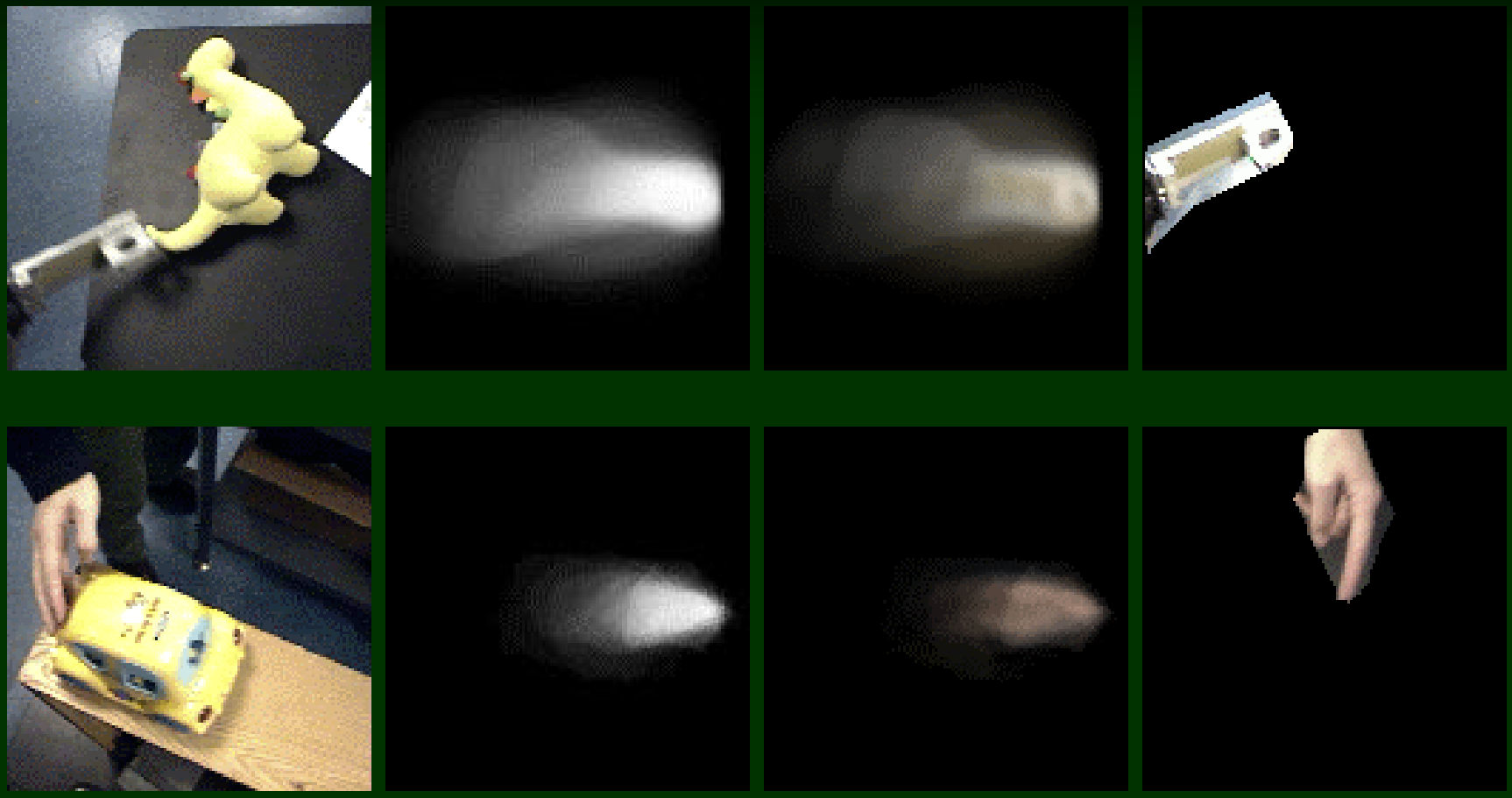
Object connects robot and human action

catching manipulators in the act

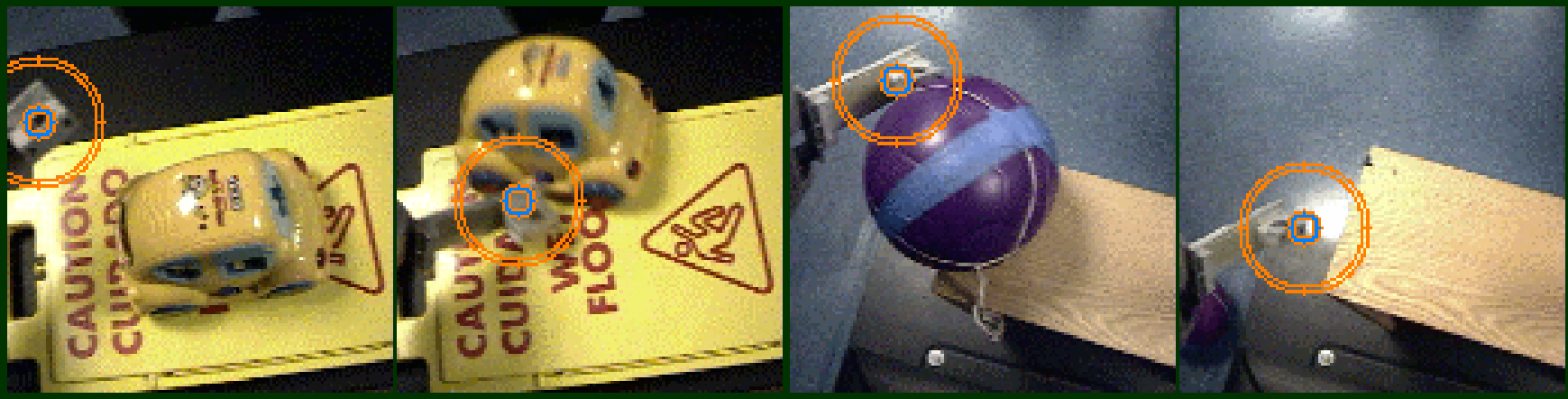
manipulator approaches object

contact!

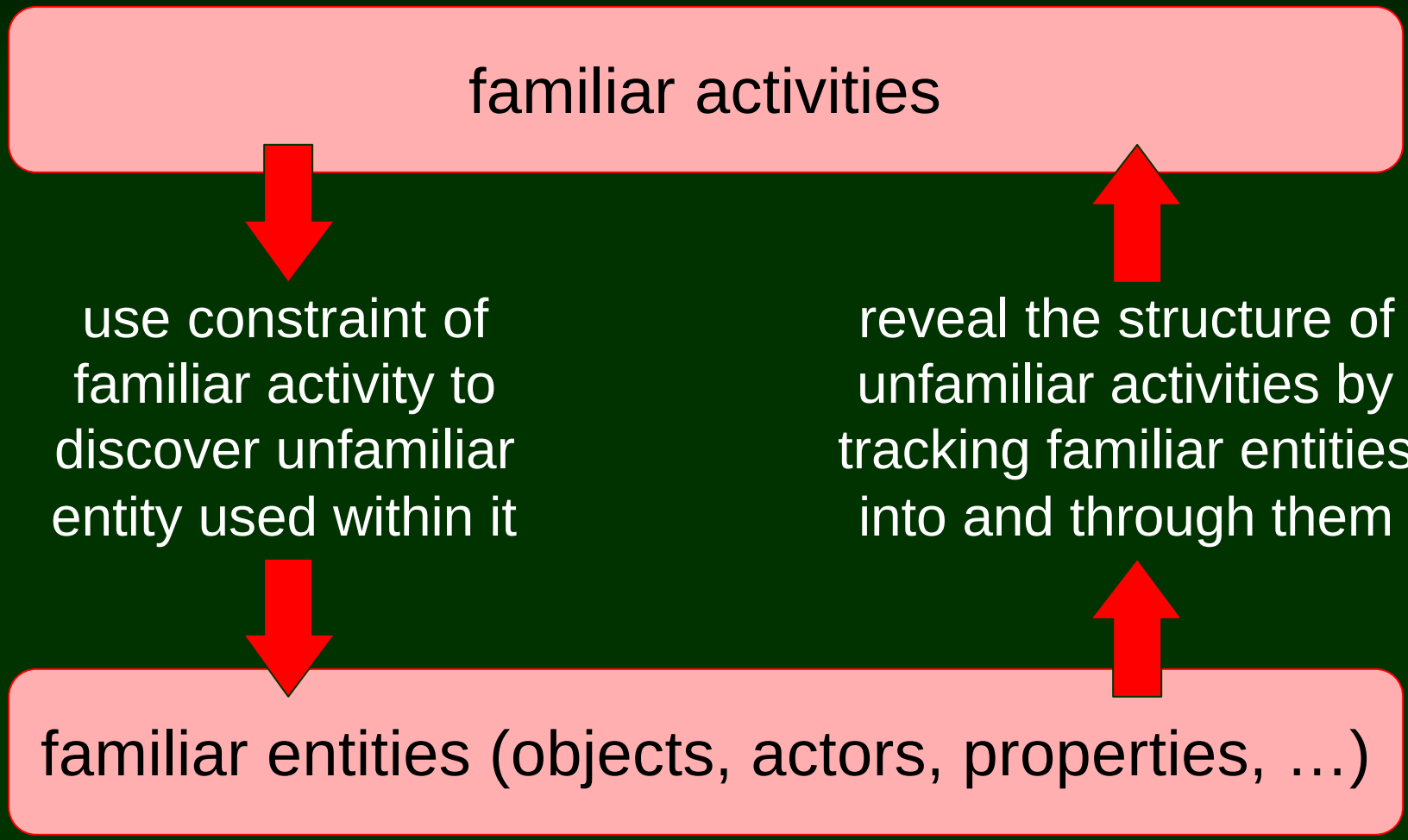
modeling manipulators



manipulator recognition



theoretical goal: a virtuous circle



conclusions

Active segmentation can make any segmentation strategy better

Ideal for learning about manipulable objects – an important class of object for humanoid robots

Doesn't require a database of objects to be built up by hand

(at least, not by human hand)