from first contact to close encounters:

a developmentally deep perceptual system for a humanoid robot

– Paul Fitzpatrick –

thesis committee

Rodney Brooks, Trevor Darrell, Deb Roy

experimentation helps perception

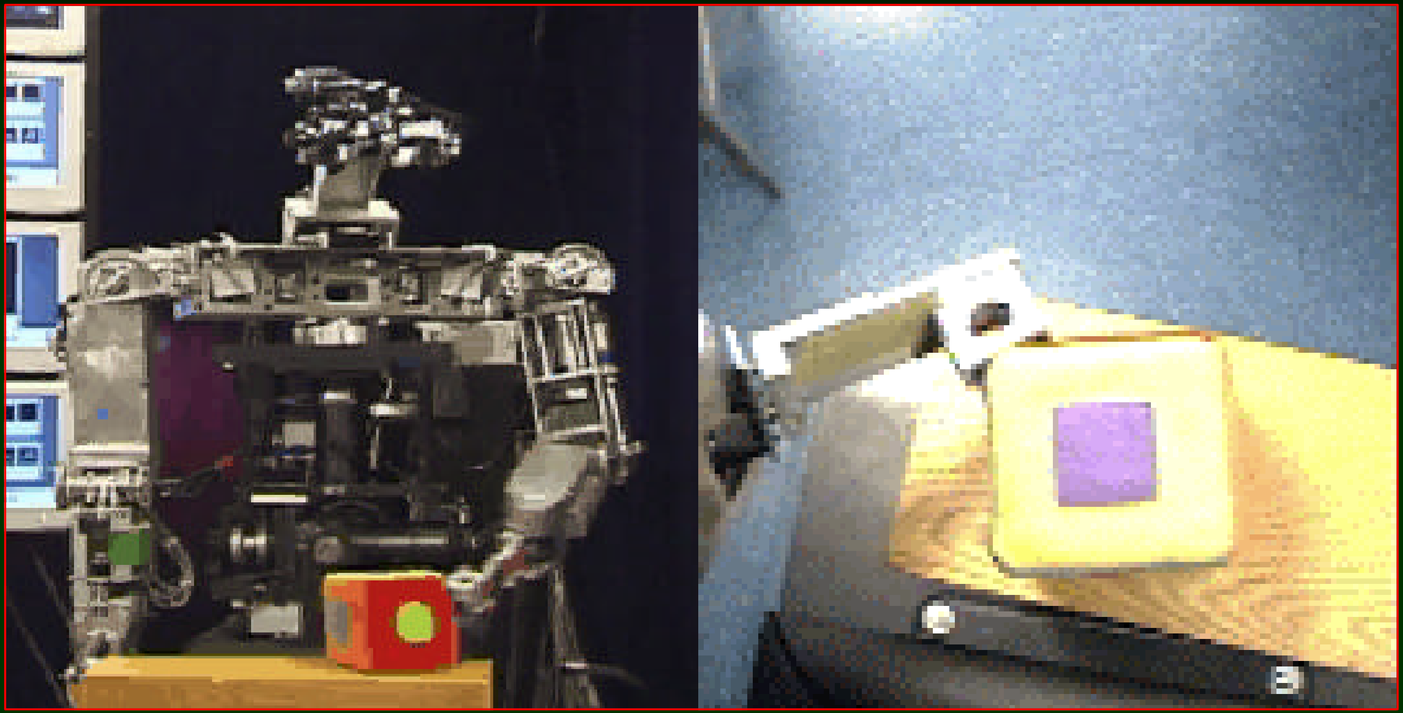
Rachel: We have got to find out if [ugly naked guy]'s alive.

Monica: How are we going to do that? There's no way.

Joey: Well there is one way. His window's open – I say, we poke him.

(brandishes the Giant Poking Device)

robots can experiment



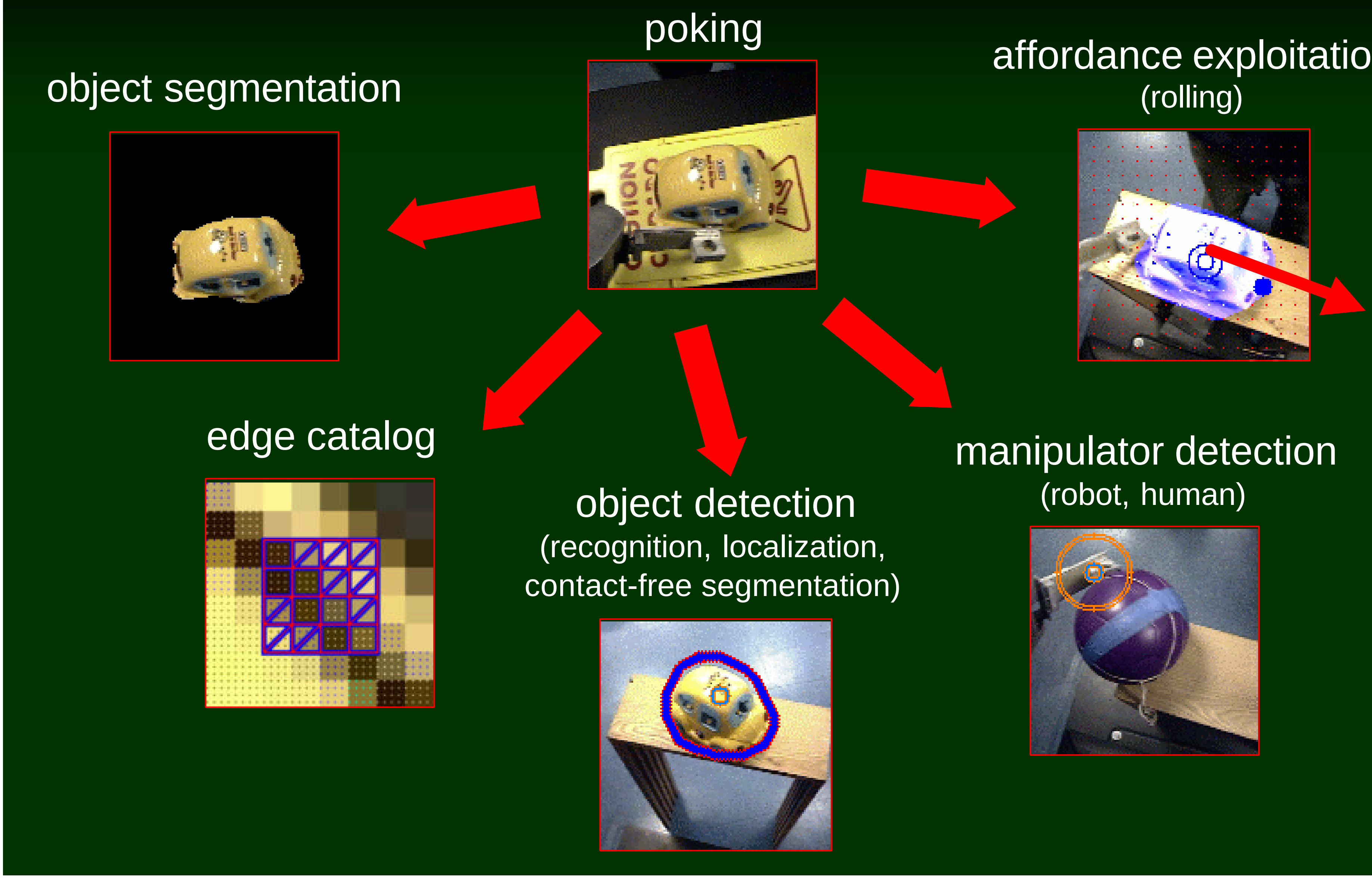
Robot: We have got to find out where this object’s boundary is.

Camera: How are we going to do that? There's no way.

Robot: Well there is one way. Looks reachable – I say, let’s poke it.

(brandishes the Giant Poking Limb)

the root of all vision



poking

object segmentation

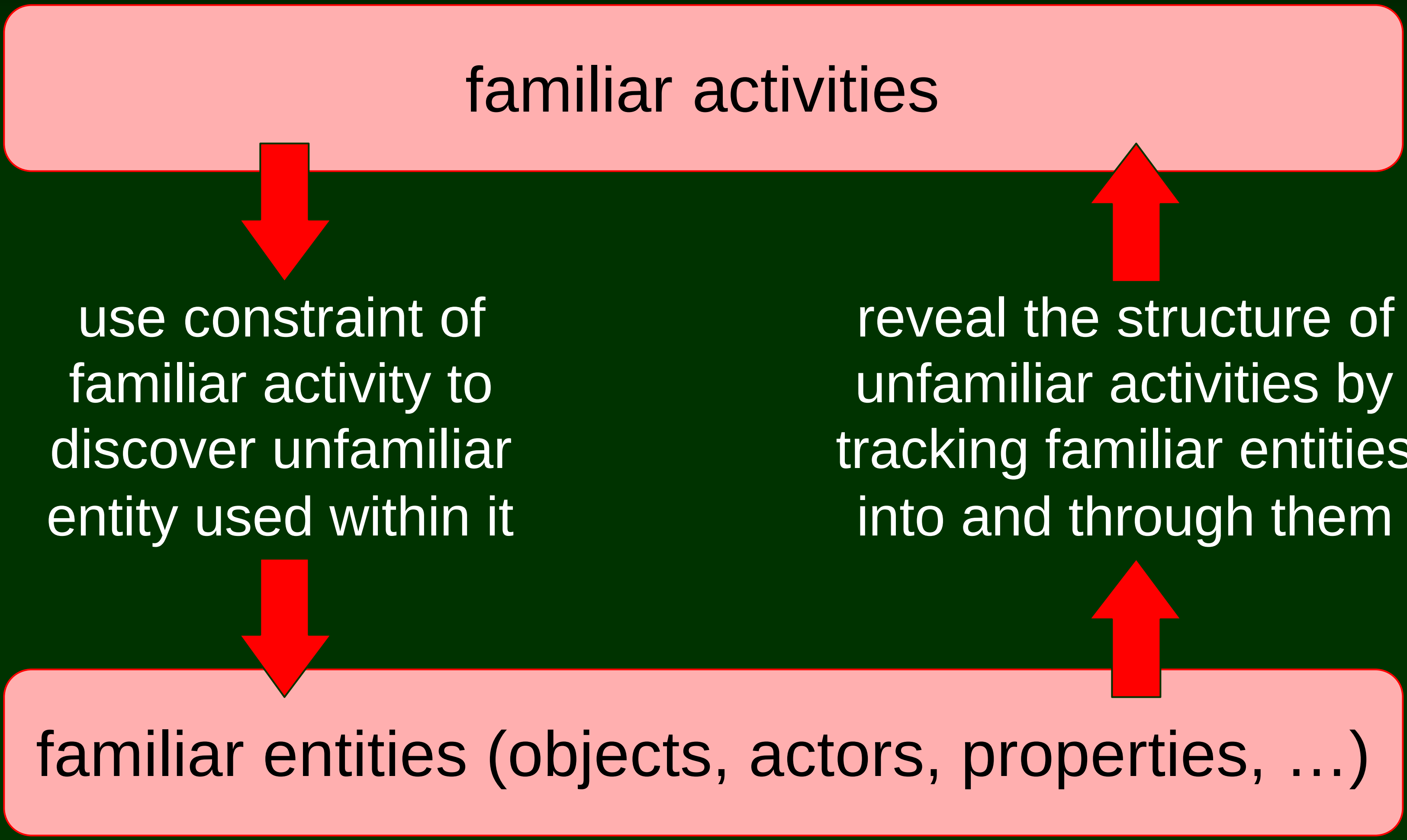
affordance exploitation
(rolling)

edge catalog

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

manipulator detection
(robot, human)

theoretical goal: a virtuous circle



familiar activities

use constraint of familiar activity to discover unfamiliar entity used within it

reveal the structure of unfamiliar activities by tracking familiar entities into and through them

familiar entities (objects, actors, properties, …)

practical goal: adaptive robots

Motivated by fallibility

- Complex action and perception will fail
- Need simpler fall-back methods that resolve ambiguity, learn from errors

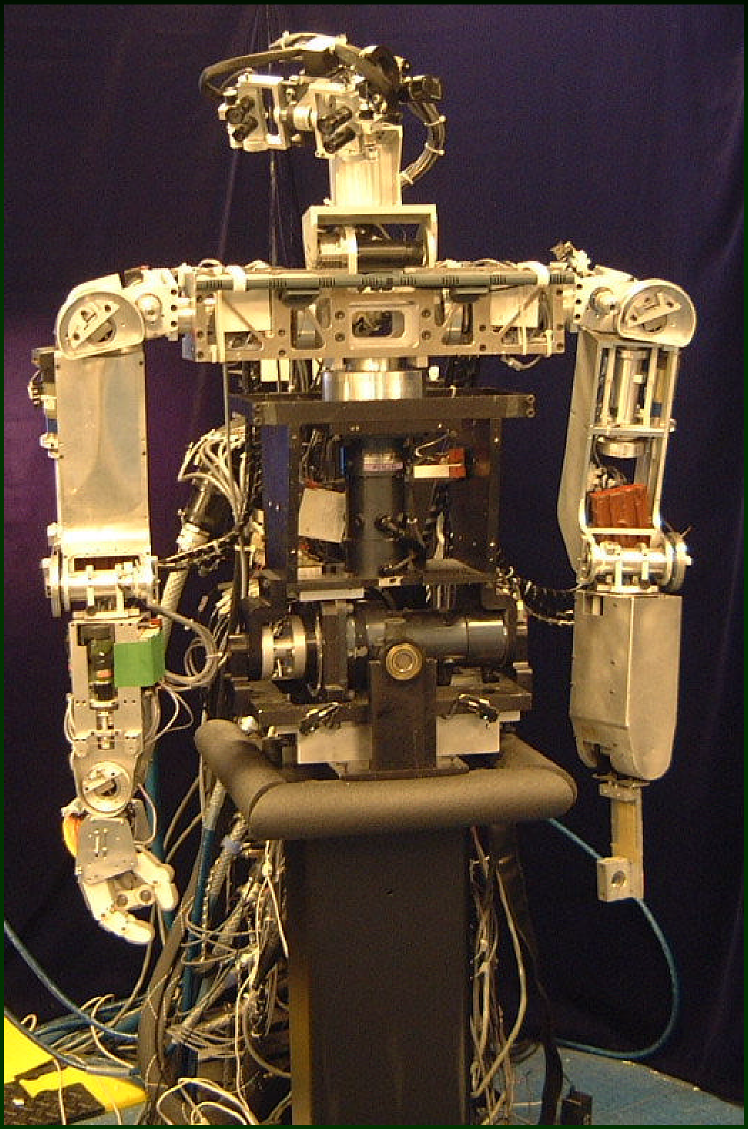
Motivated by transience

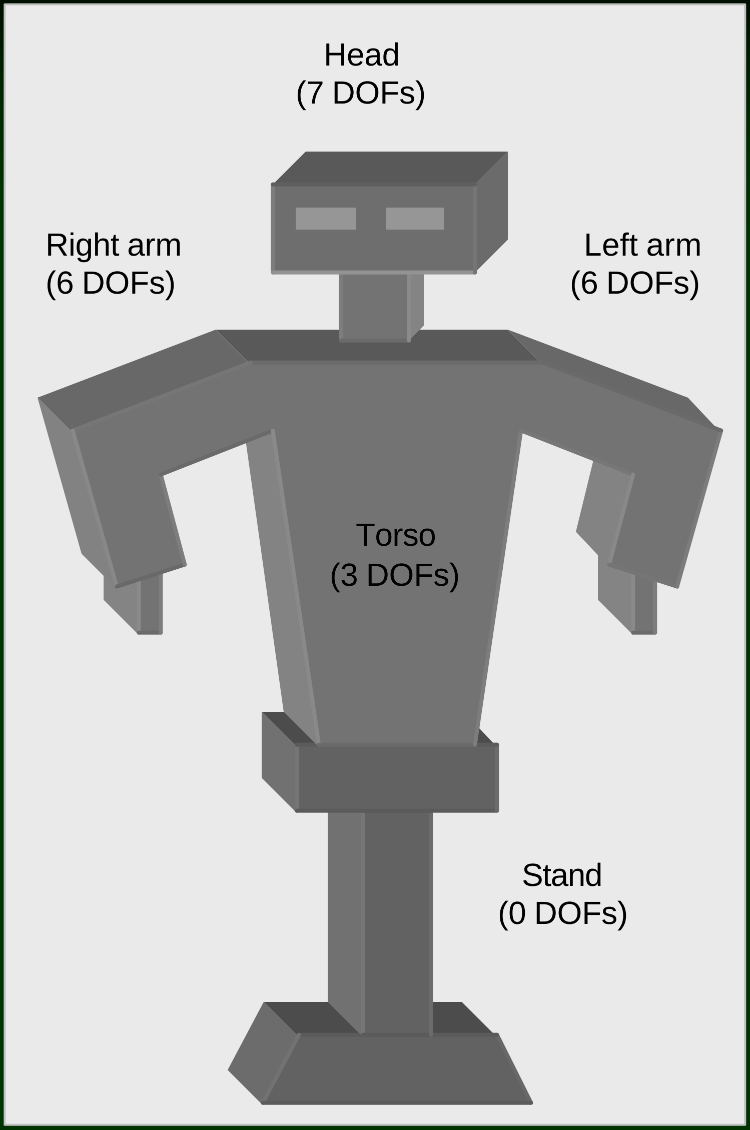
- Task for robot may change from day to day
- Ambient conditions change
- Best to build in adaptivity from very beginning

Motivated by infants

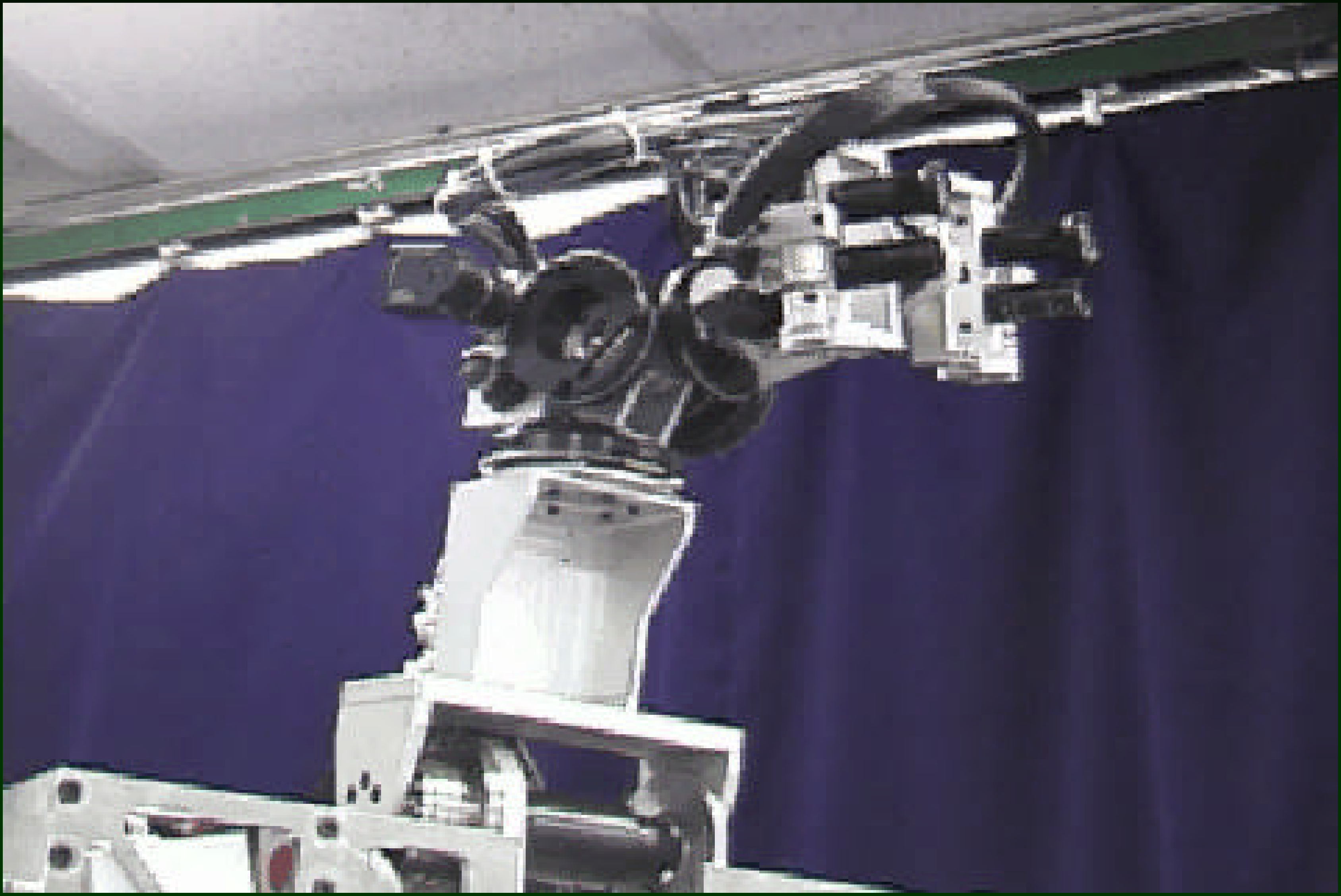
- Perceptual development outpaces motor
- Able to explore despite sloppy control

giant poking device: Cog





giant poking device: Cog



talk overview

Learning from an activity

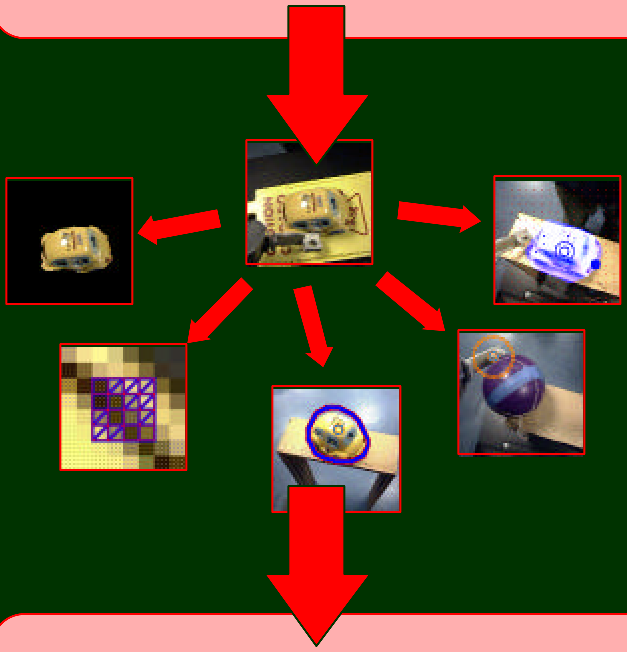
- Poking: to learn to recognize objects, manipulators, etc.
- Chatting: to learn the names of objects

Learning a new activity

- Searching for an object
- Then back to learning *from* the activity…

virtuous circle

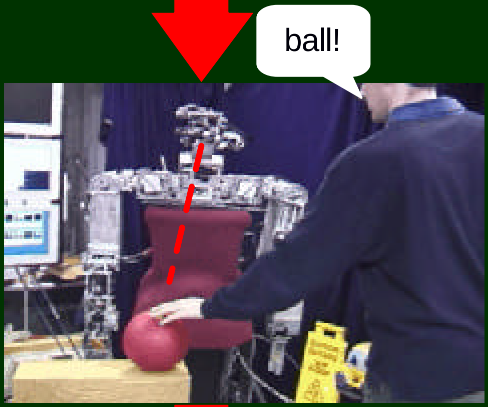
poking



objects, …

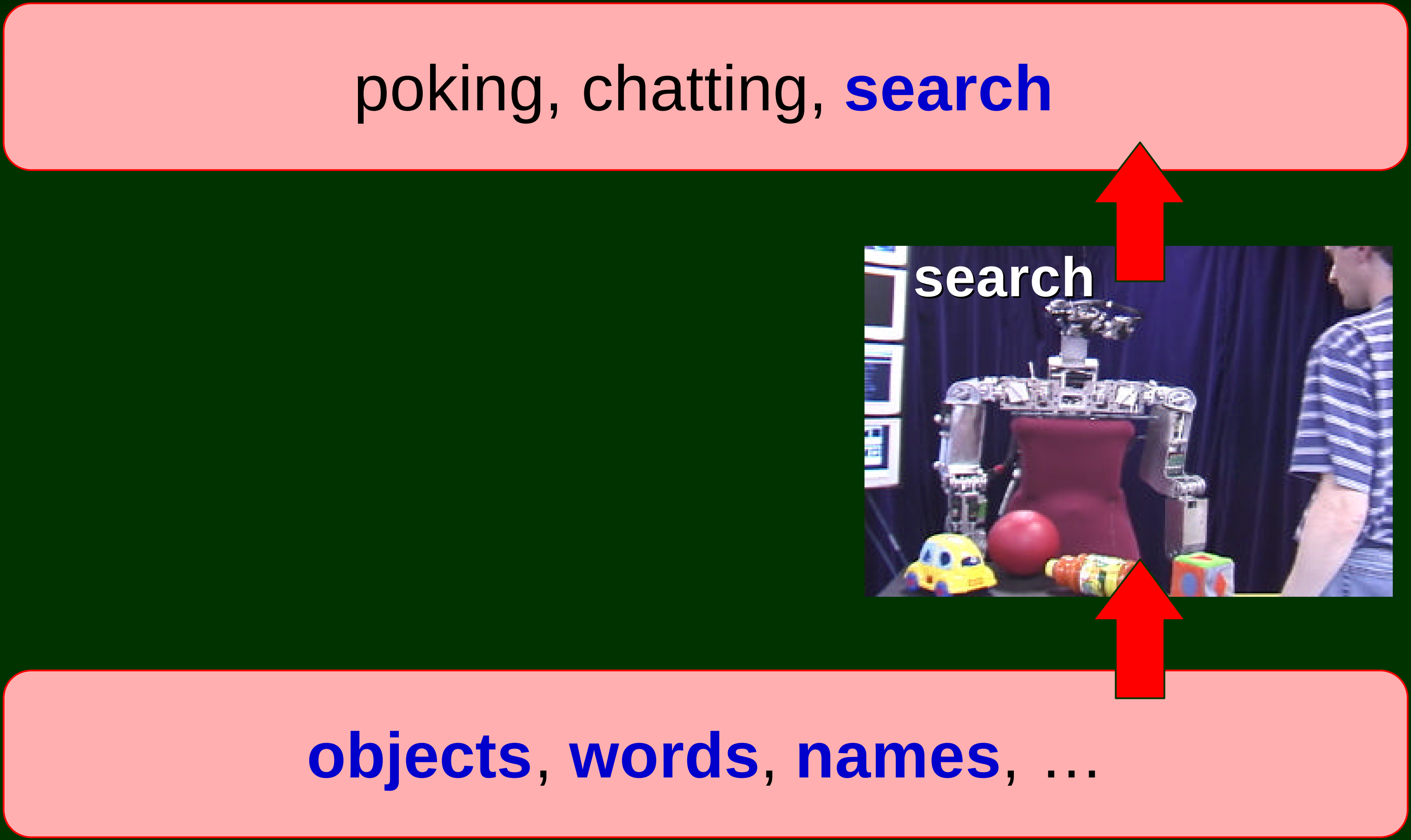
virtuous circle

poking, **chatting**



objects, **words**, **names**, …

virtuous circle

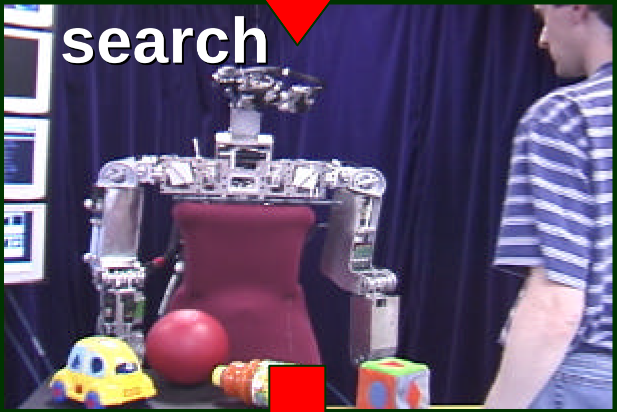


poking, chatting, **search**

objects, **words**, **names**, …

virtuous circle

poking, chatting, **search**



objects, words, **names**, …

talk overview

Learning from an activity

- Poking: to learn to recognize objects, manipulators, etc.
- Chatting: to learn the names of objects

Learning a new activity

- Searching for an object
- Then back to learning *from* the activity…

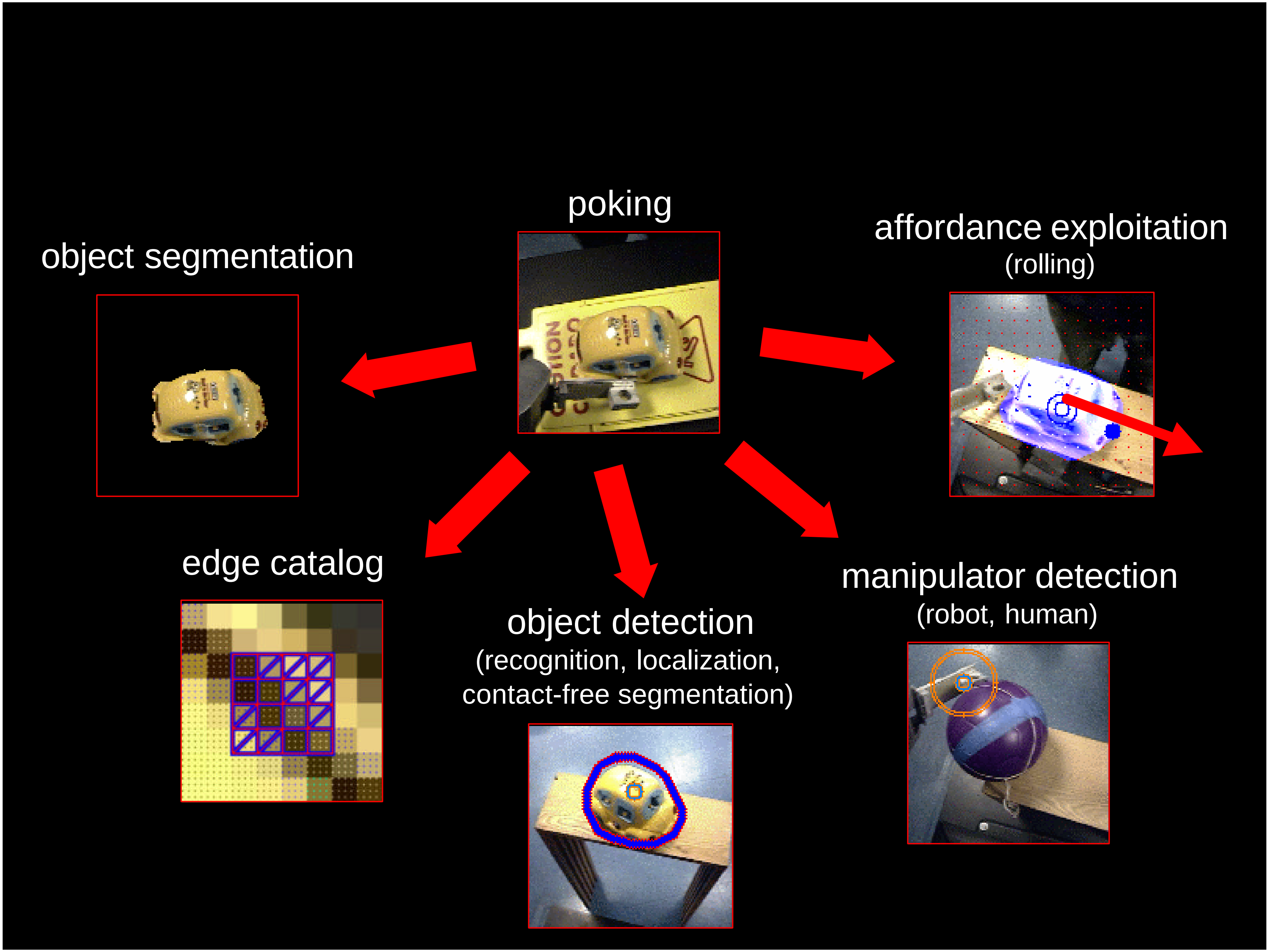
talk overview

Learning from an activity

- Poking: to learn to recognize objects, manipulators, etc.
- Chatting: to learn the names of objects

Learning a new activity

- Searching for an object
- Then back to learning *from* the activity…



poking

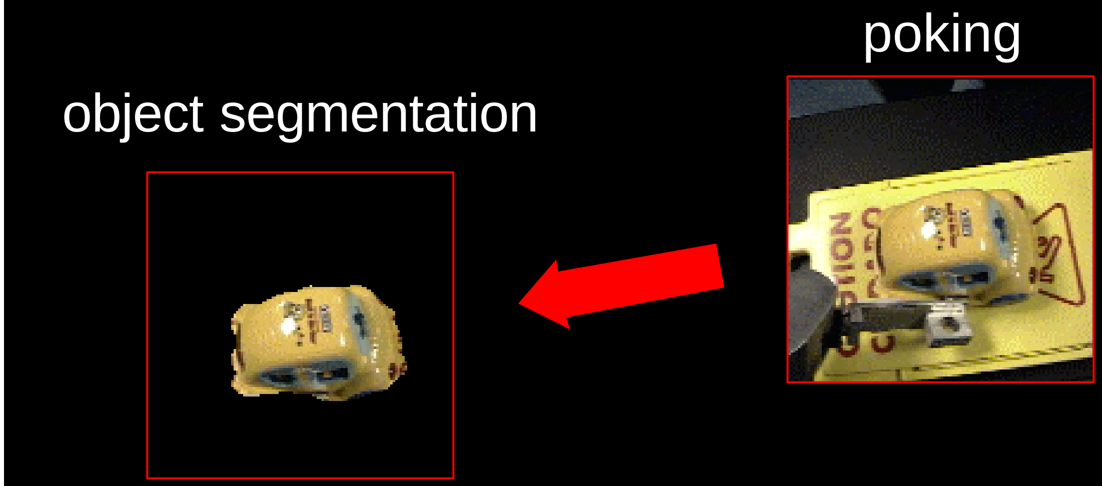
object segmentation

affordance exploitation
(rolling)

edge catalog

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

manipulator detection
(robot, human)



poking

object segmentation

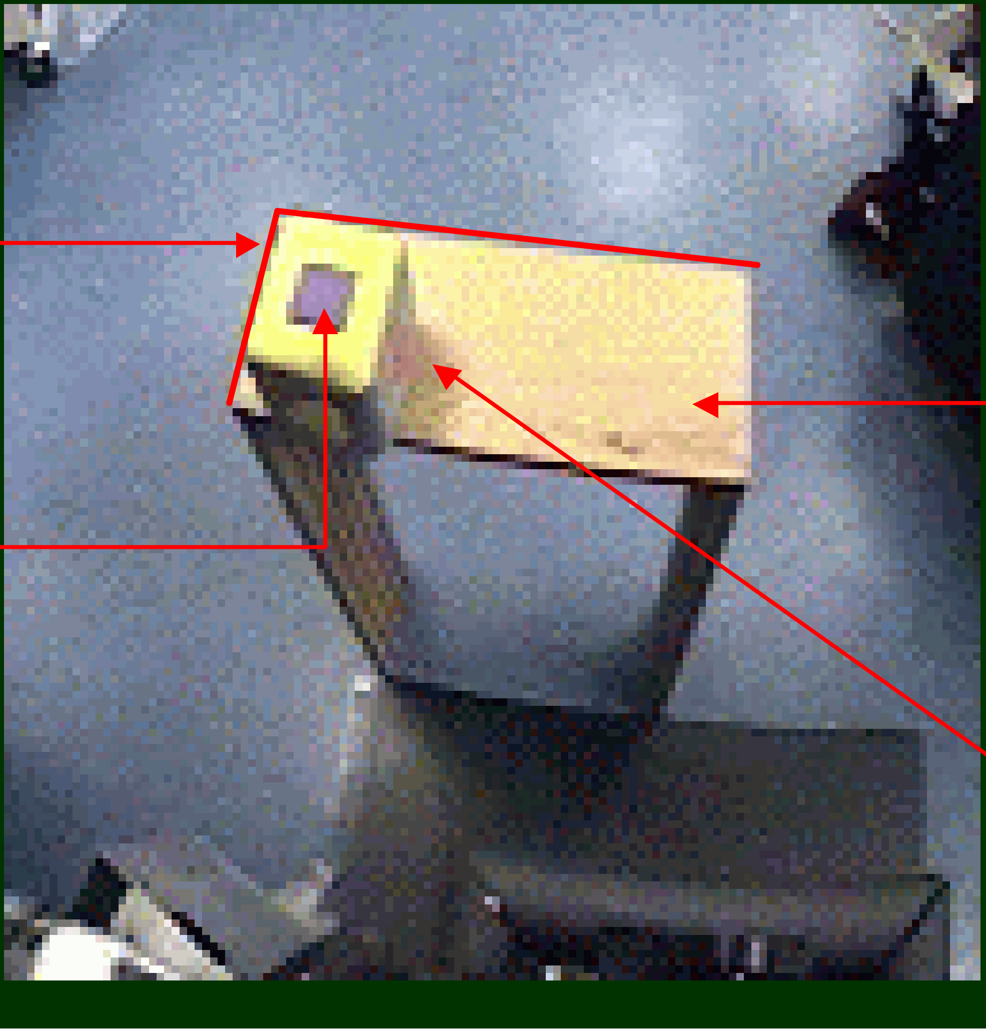
“Active Segmentation”

segmenting objects
through action

“Active Segmentation”

segmenting objects
by coming into contact with them

a simple scene?



Edges of table and cube overlap

Cube has misleading surface pattern

Color of cube and table are poorly separated

Maybe some cruel grad-student faked the cube with paper, or glued it to the table

active segmentation

active segmentation

where to poke?

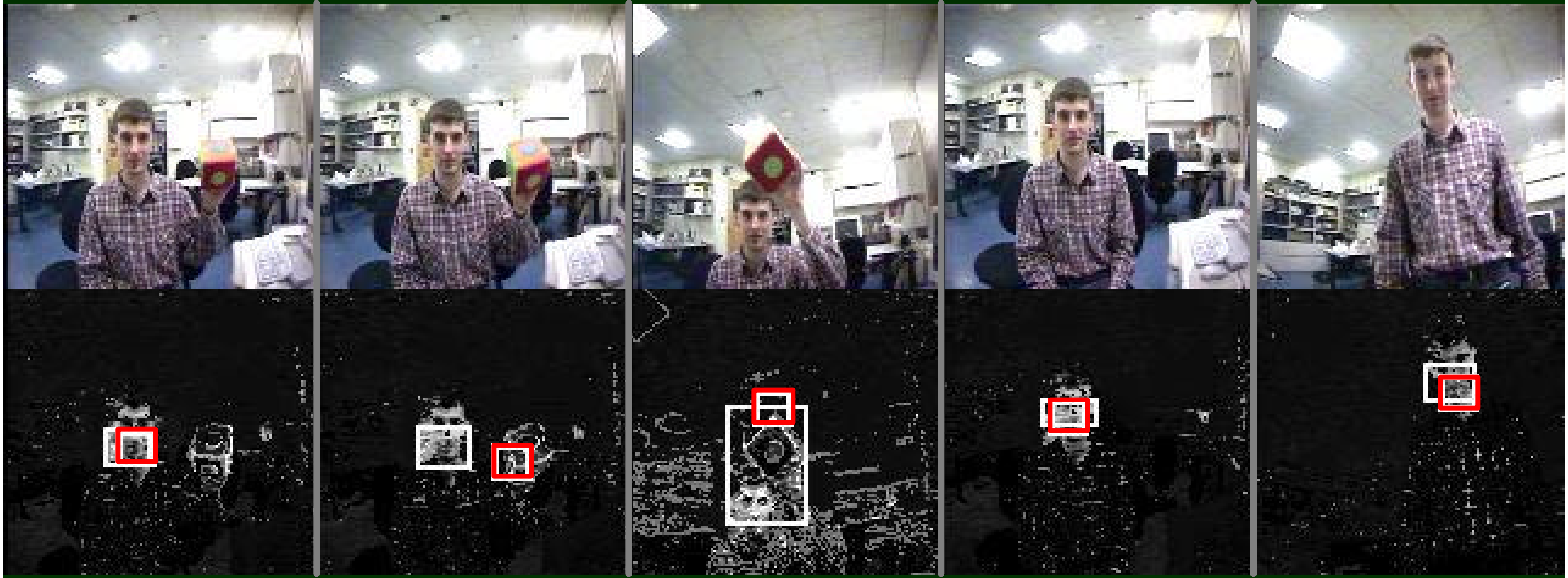
Visual attention system

- Robot selects a region to fixate based on salience (bright colors, motion, etc.)
- Region won’t generally correspond to extent of object

Poking activation

- Region is stationary
- Region reachable (right distance, not too high up)
- Distance measured through binocular disparity

visual attention system



person approaches	shakes object	moves object	hides object	stands up
attracted to skin color	attracted to bright color, movement	smooth pursuit	attracted to skin color	smooth pursuit

(Collaboration with Brian Scassellati, Giorgio Metta, Cynthia Breazeal)

tracking

poking activation

evidence for segmentation

Areas where motion is observed upon contact

- classify as ‘foreground’

Areas where motion is observed immediately before contact

- classify as ‘background’

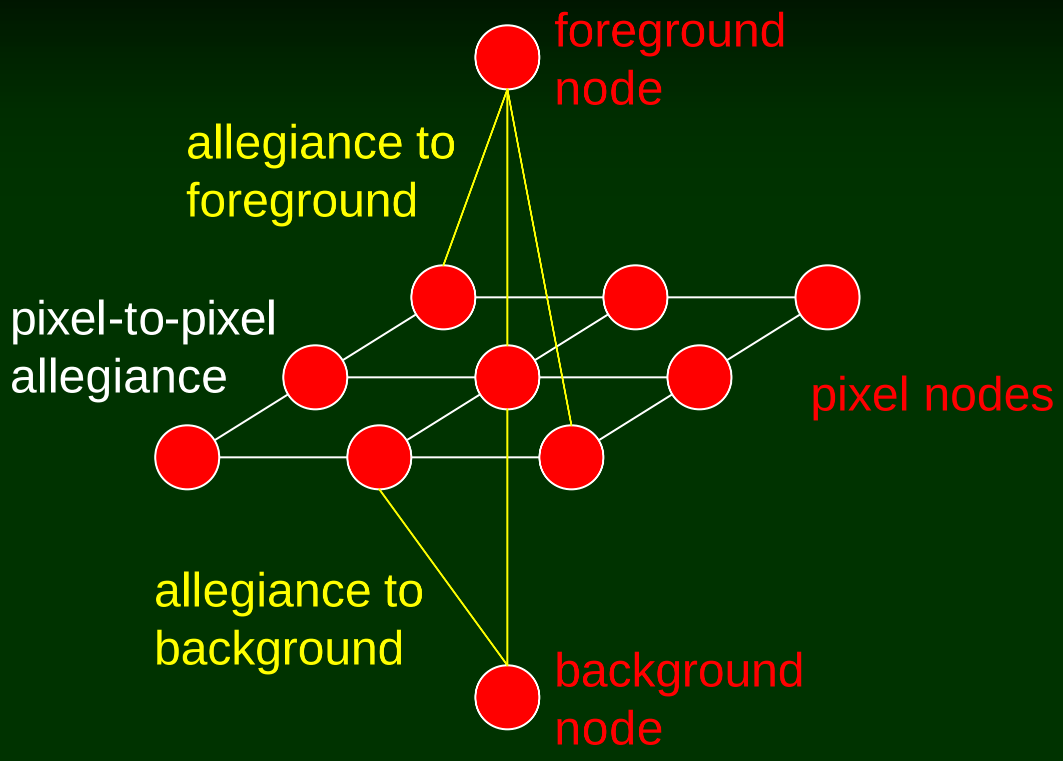
Textured areas where no motion was observed

- classify as ‘background’

Textureless areas where no motion was observed

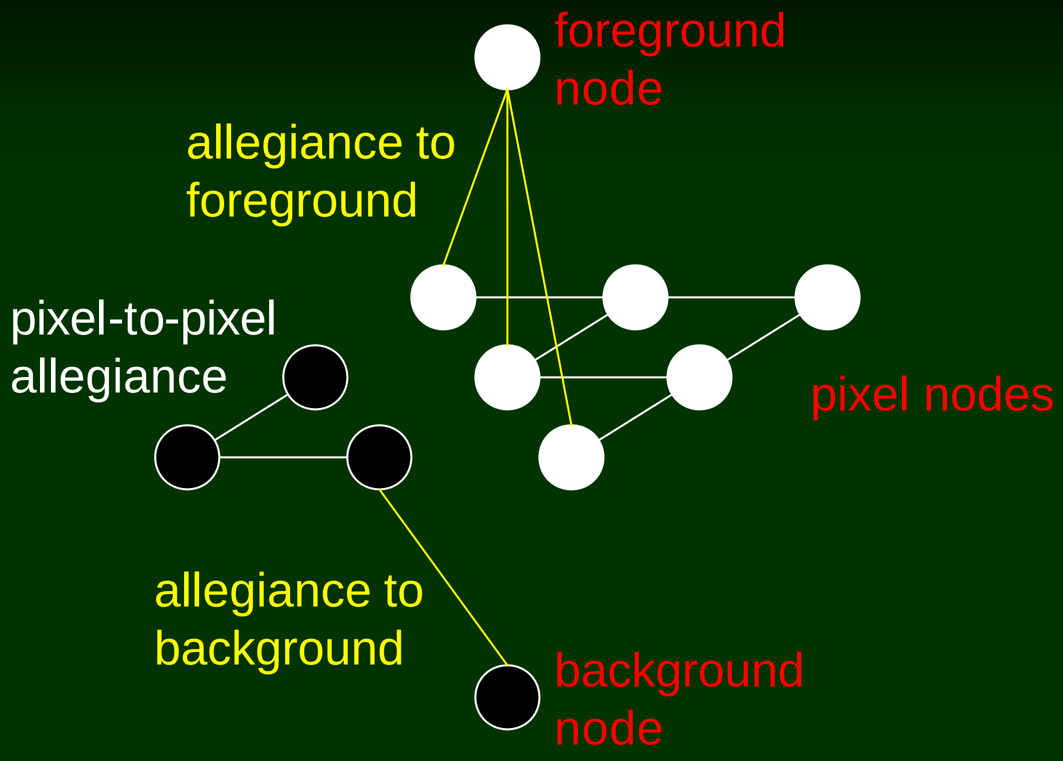
- no information

minimum cut



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

minimum cut



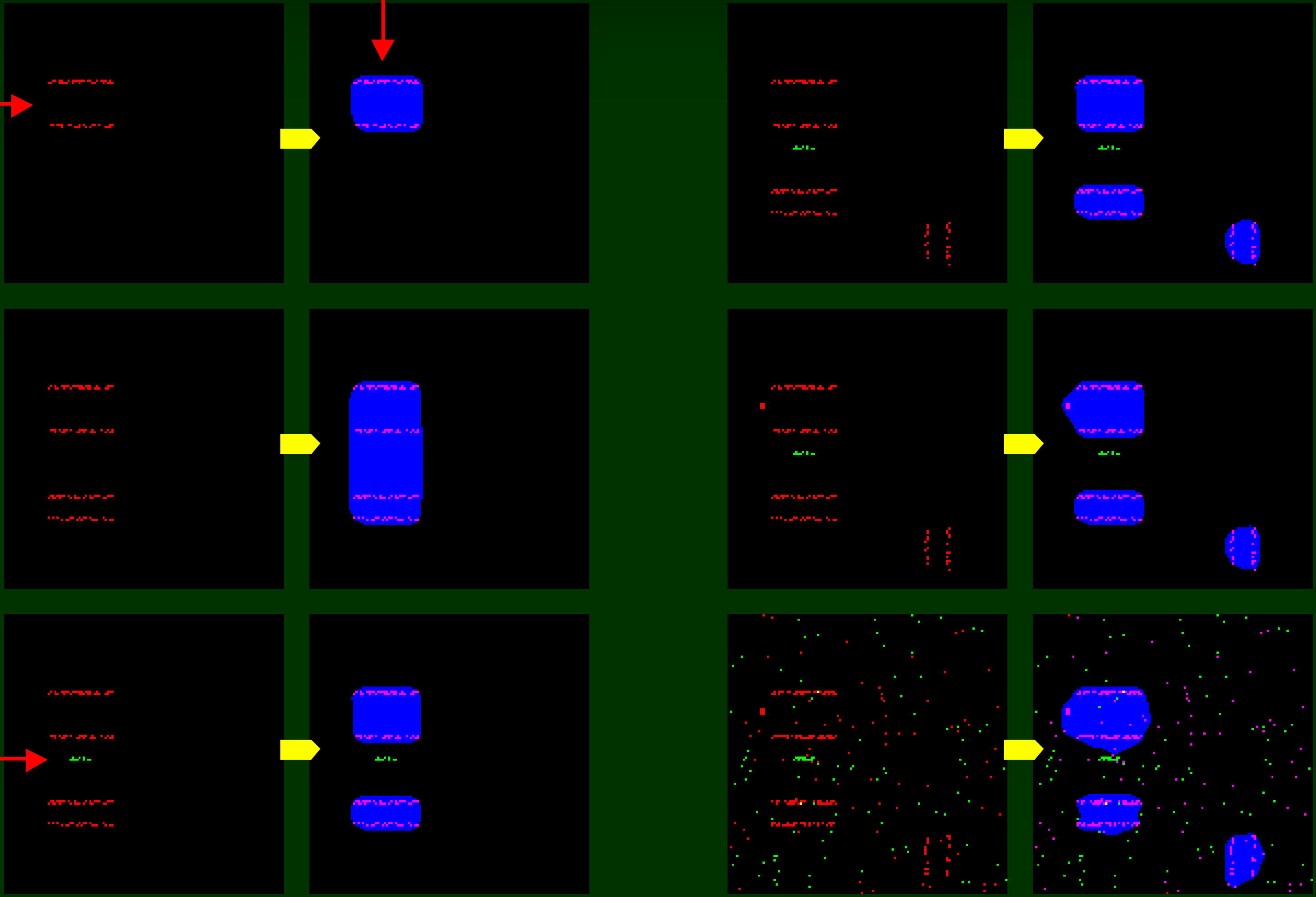
“allegiance” = cost of assigning two nodes to different layers (foreground versus background)

grouping (on synthetic data)

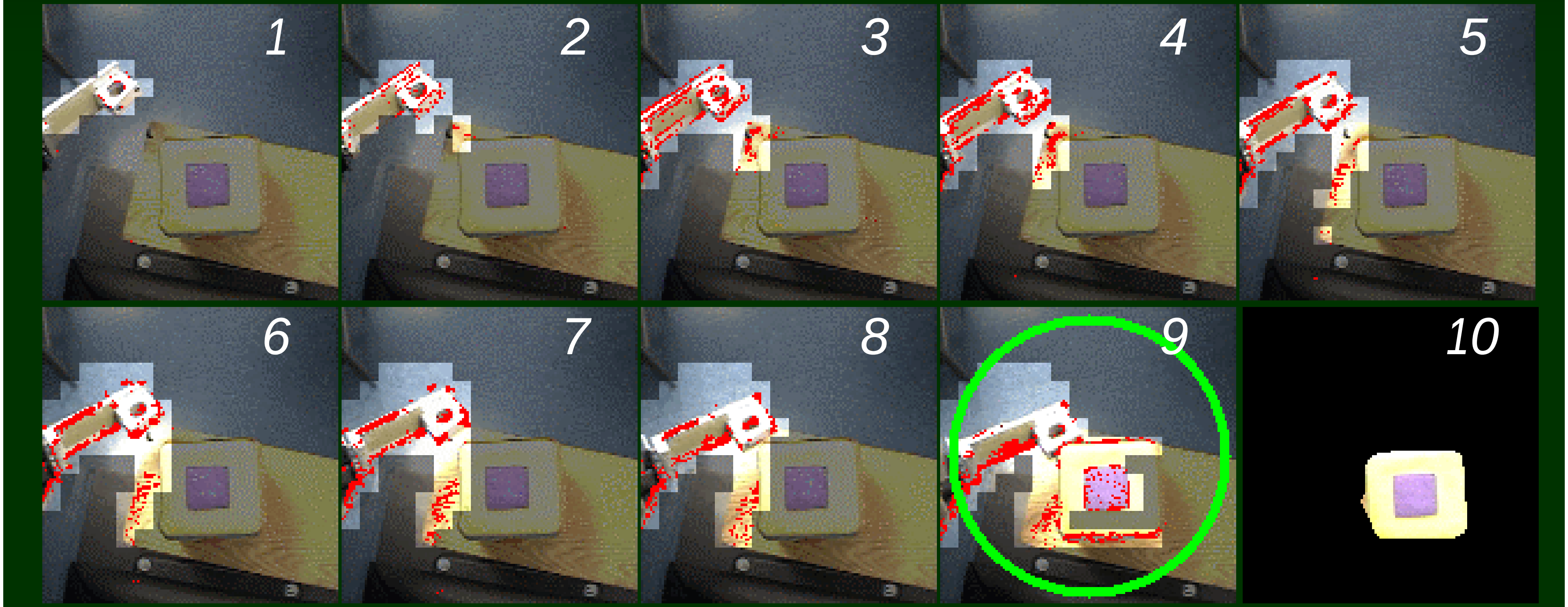
proposed segmentation

“figure” points (known motion)

“ground” points (stationary, or gripper)



point of contact

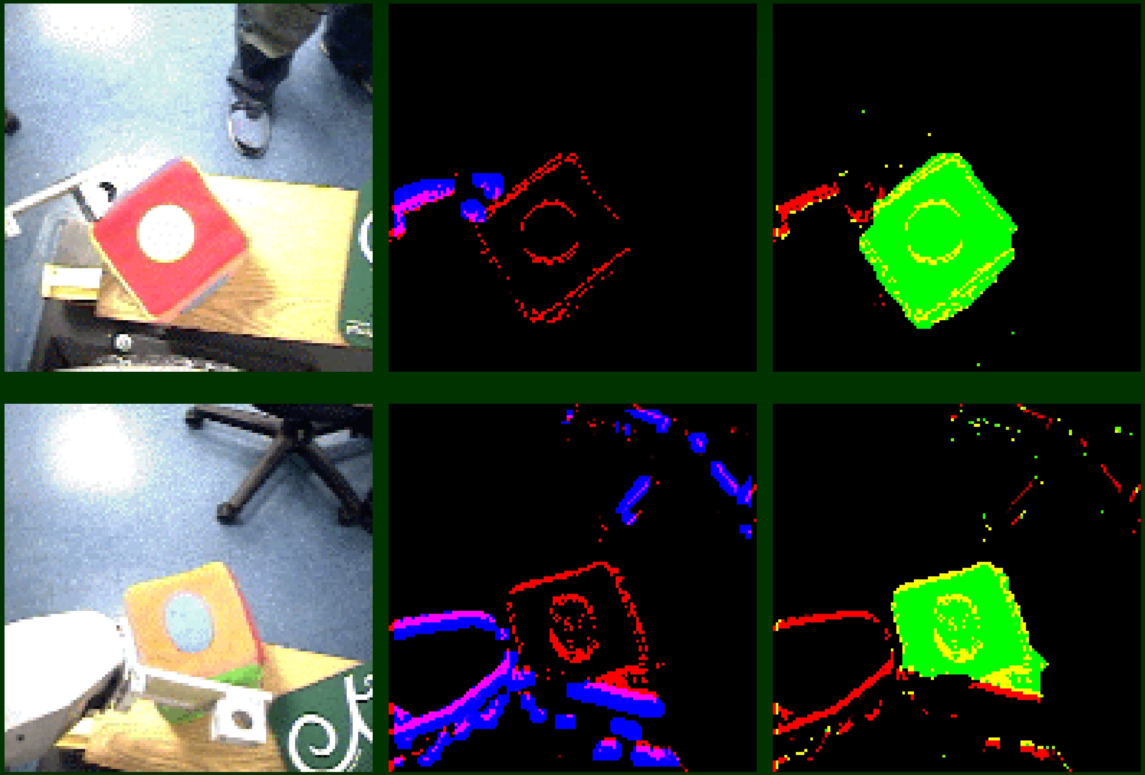


Motion spreads continuously
(arm or its shadow)

Motion spreads suddenly, faster than the arm itself → contact

point of contact

segmentation examples



Side tap

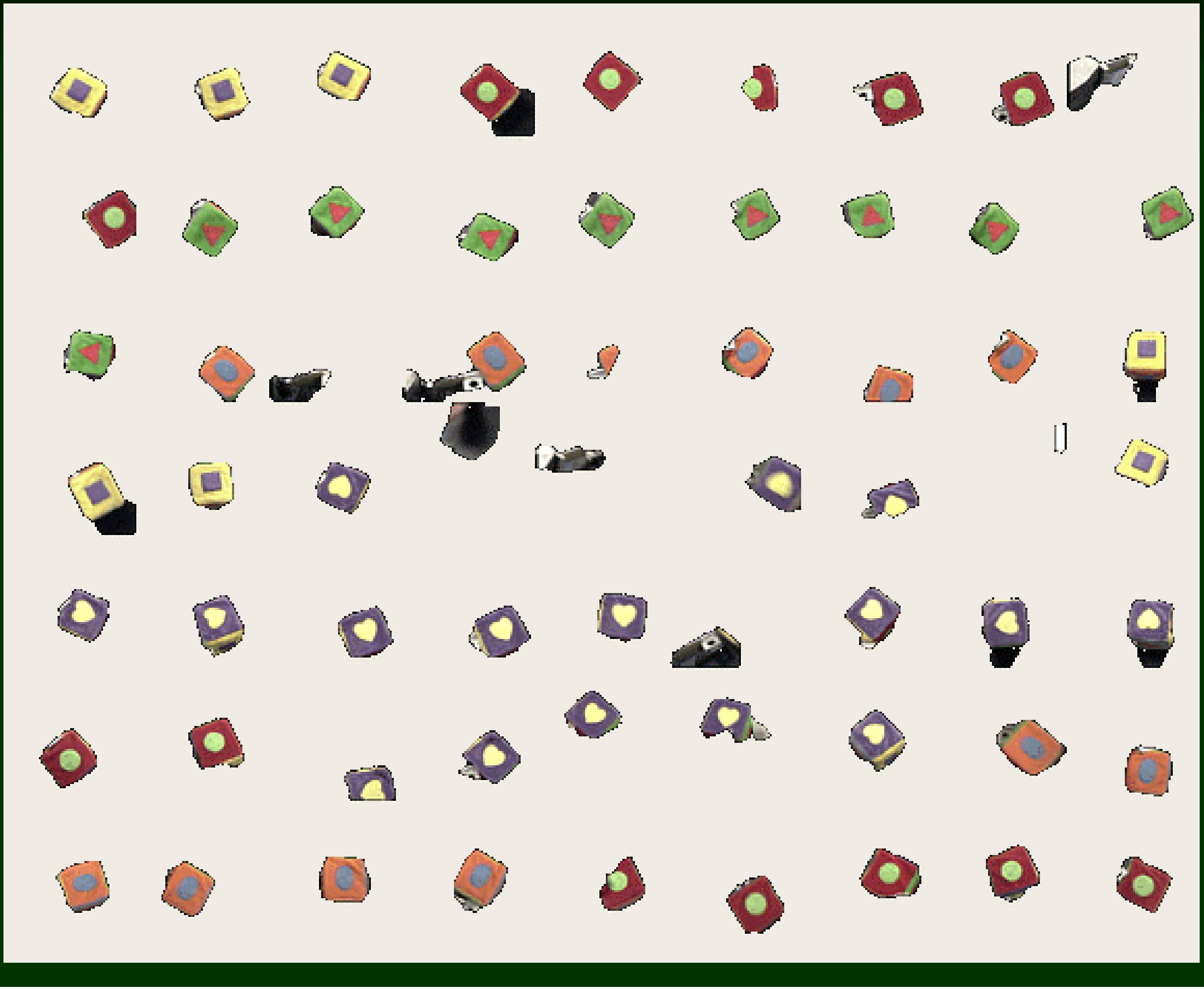
Back slap

Impact event

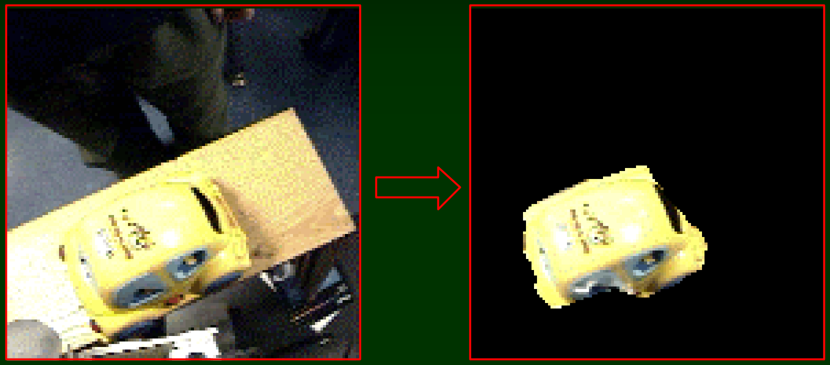
Motion caused
(red = novel,
Purple/blue = discounted)

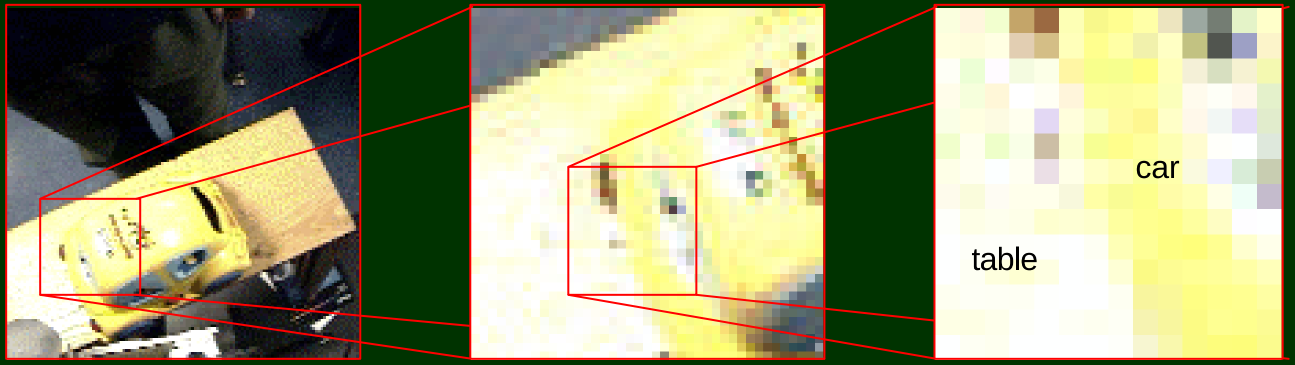
Segmentation
(green/yellow)

segmentation examples

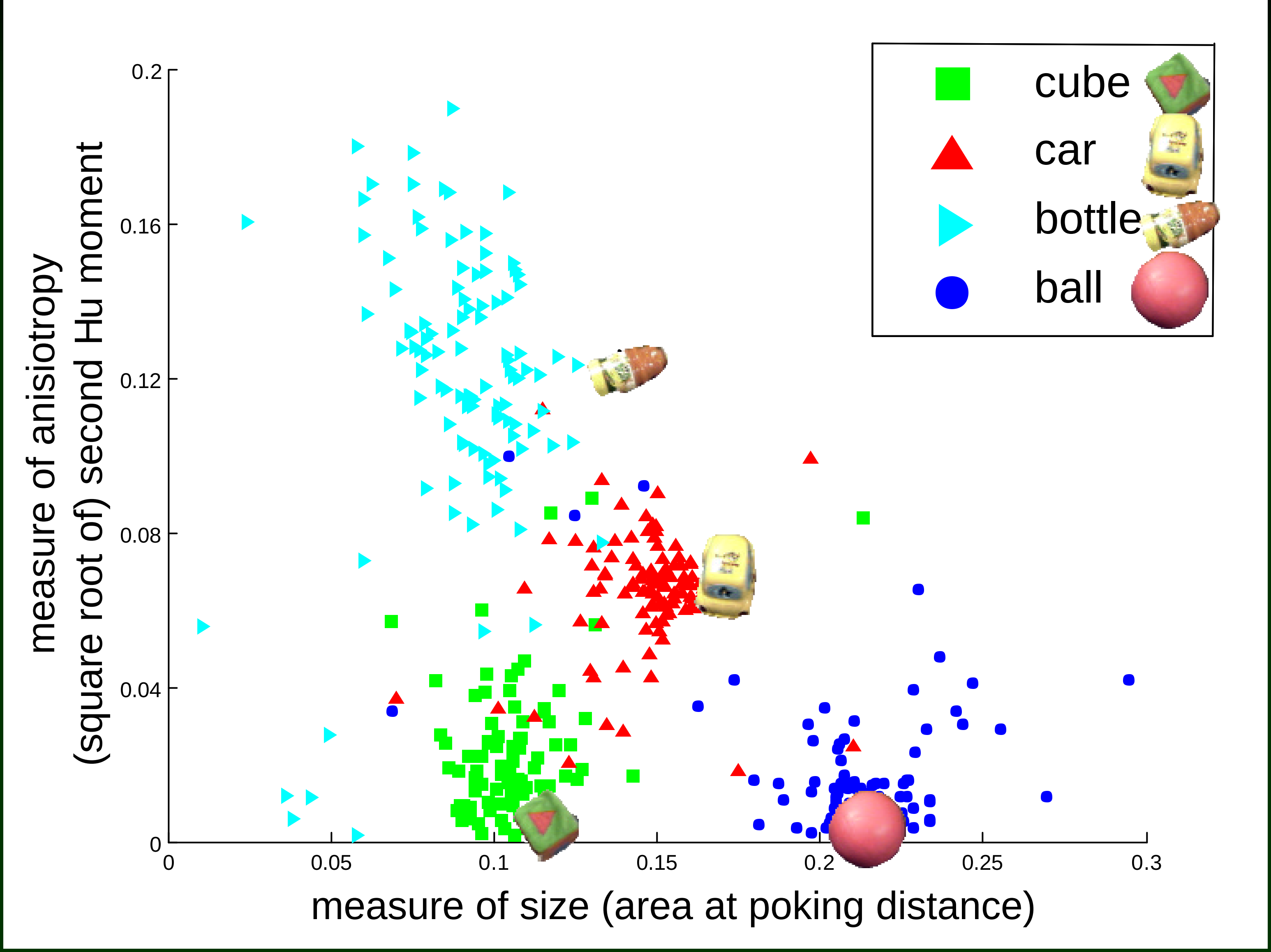


segmentation examples

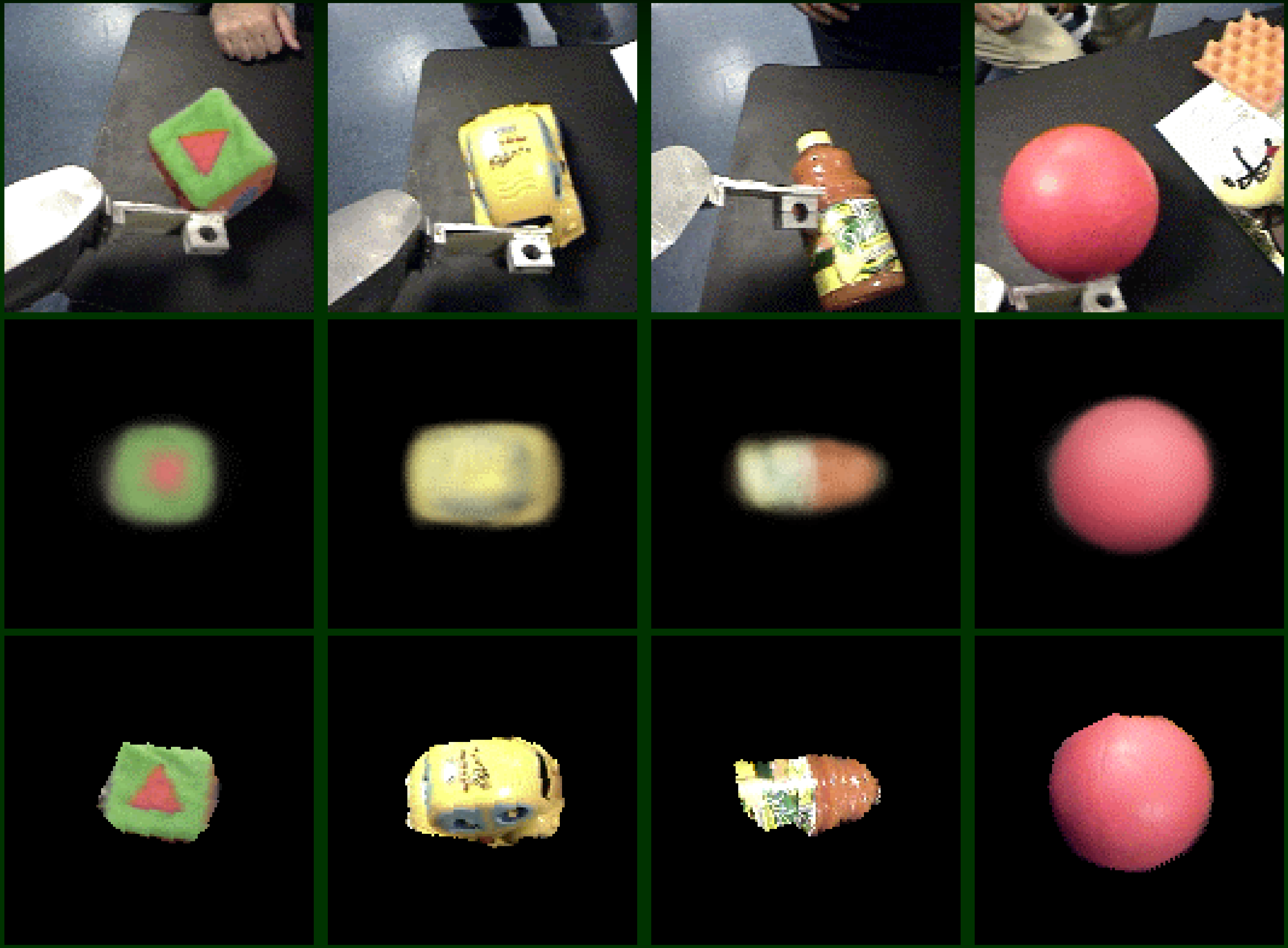


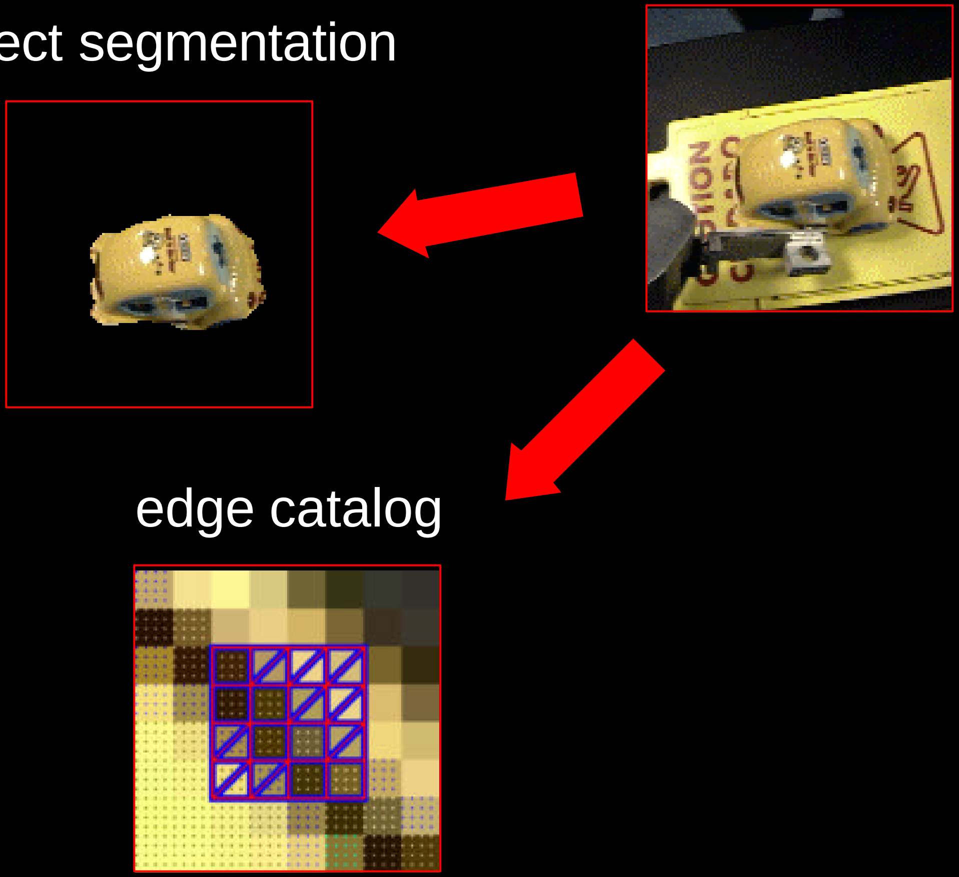


boundary fidelity



signal to noise





poking

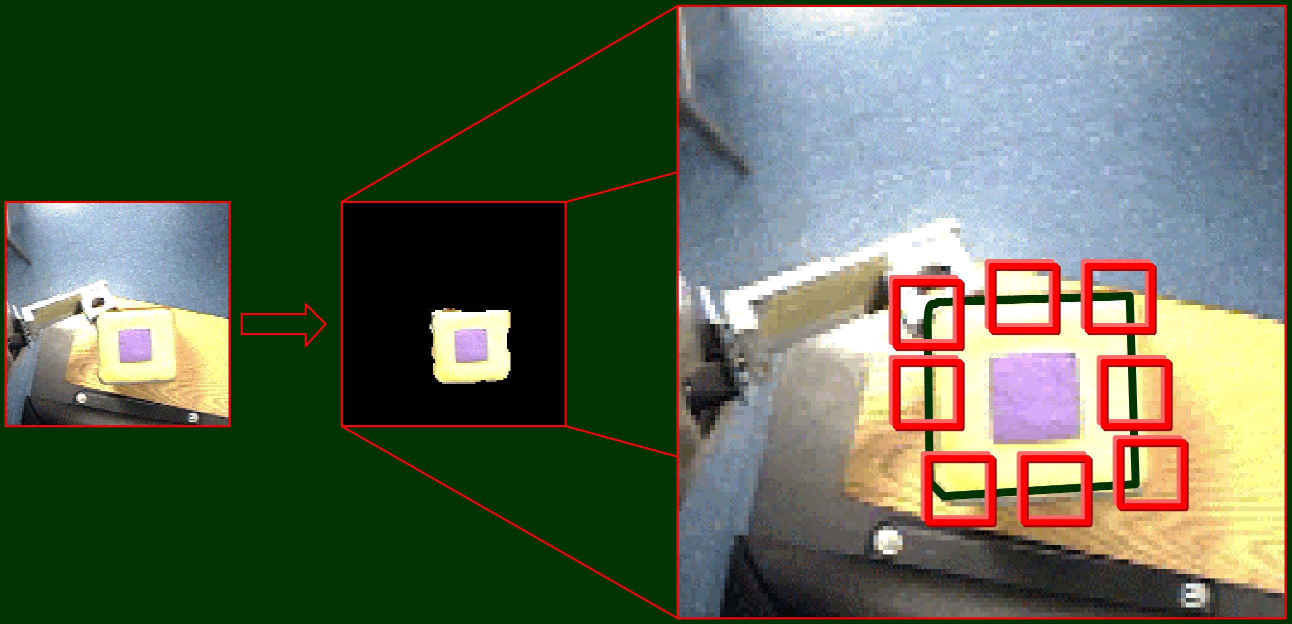
object segmentation

edge catalog

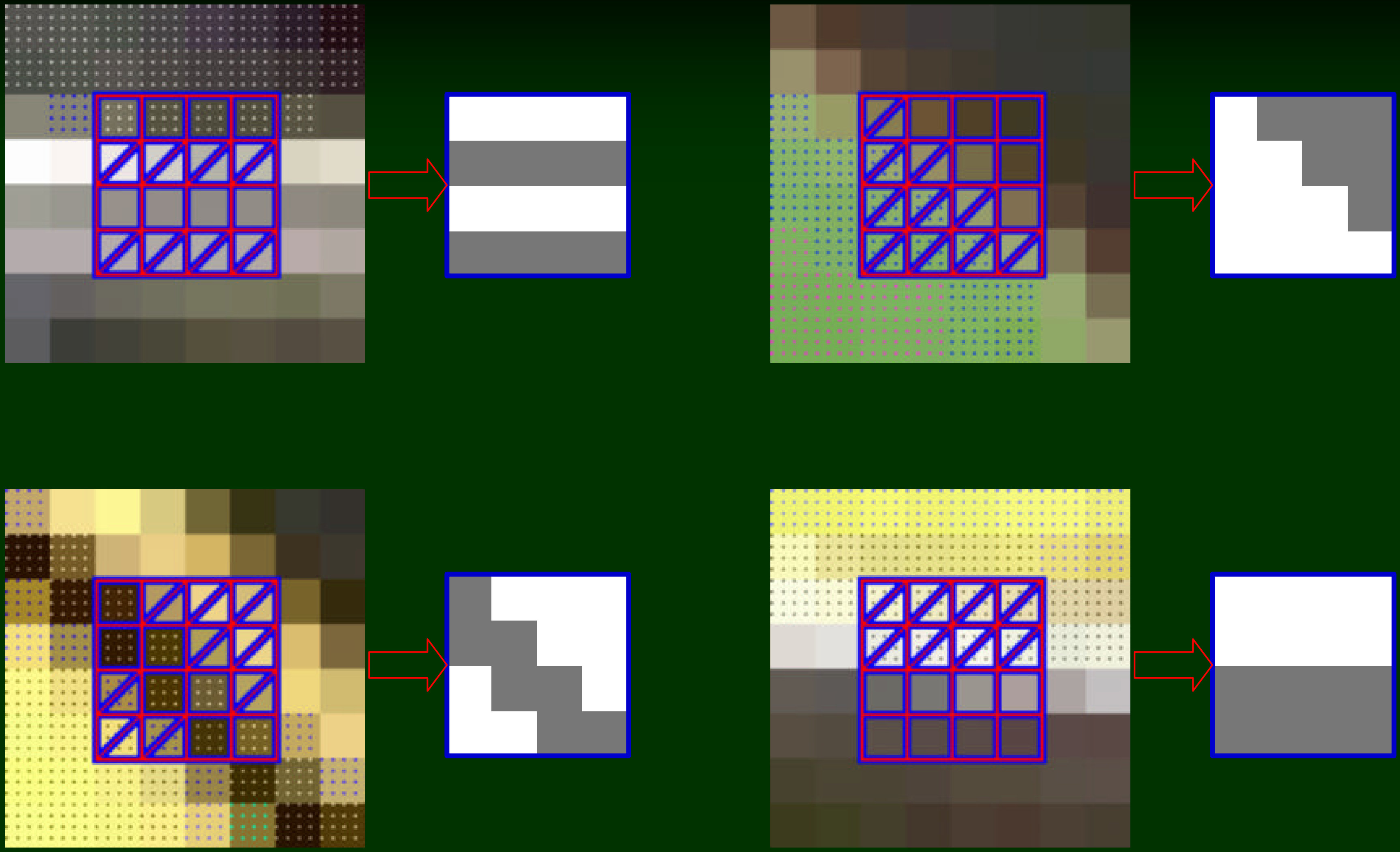
“Appearance Catalog”

exhaustively characterizing
the appearance of a low-level feature

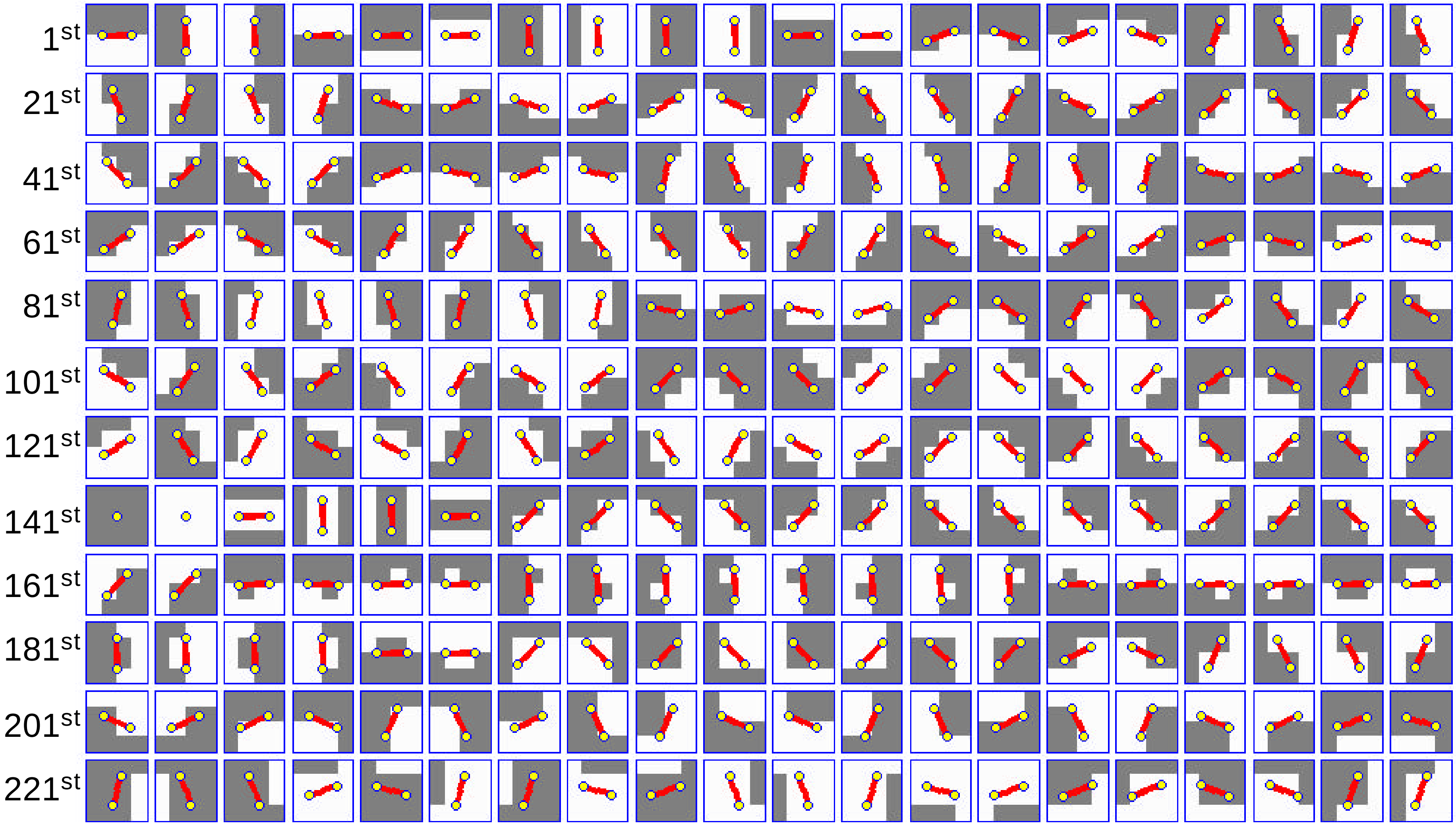
sampling oriented regions



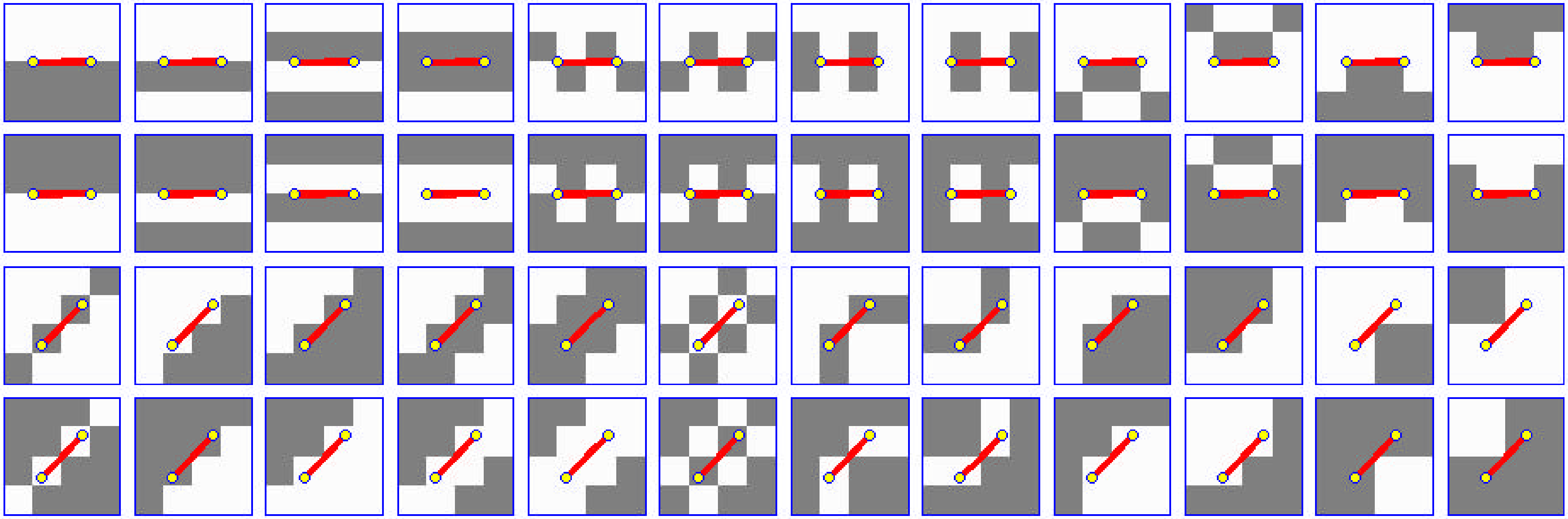
sample samples



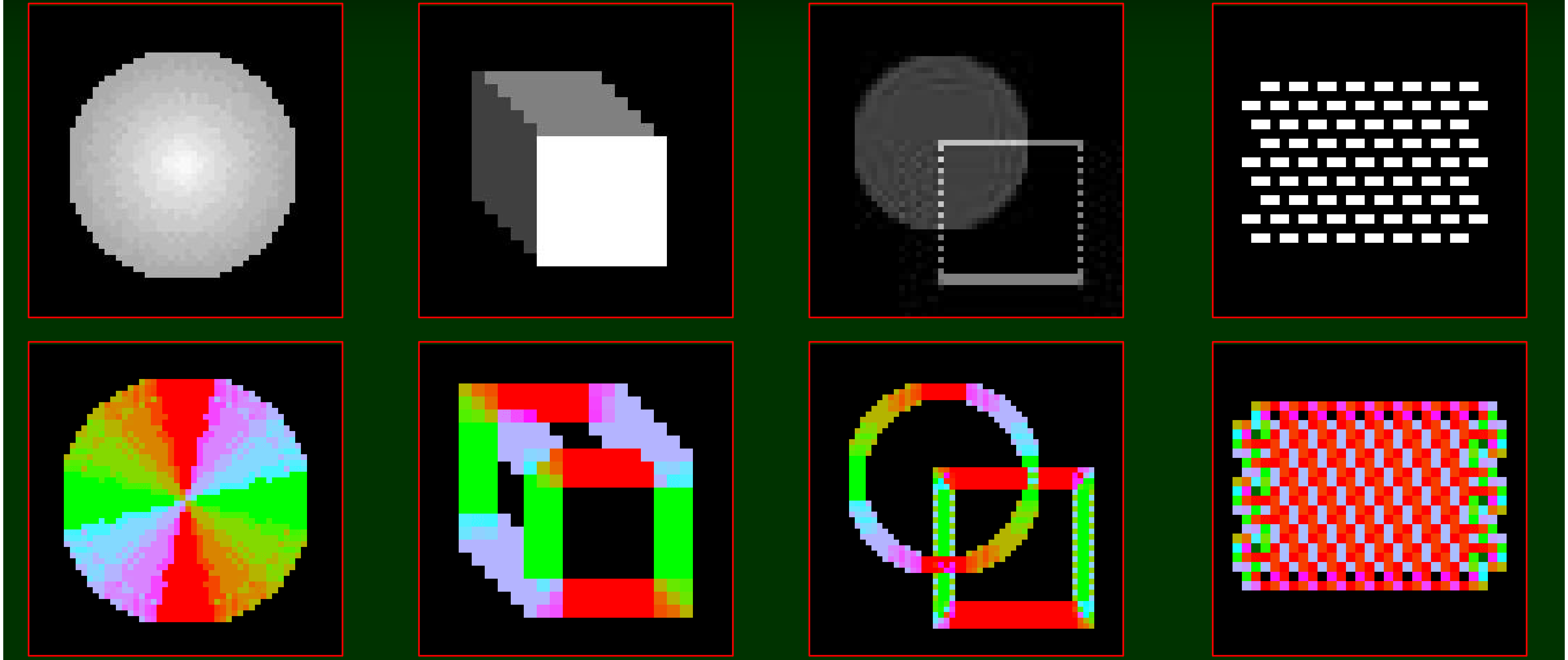
most frequent samples



selected samples

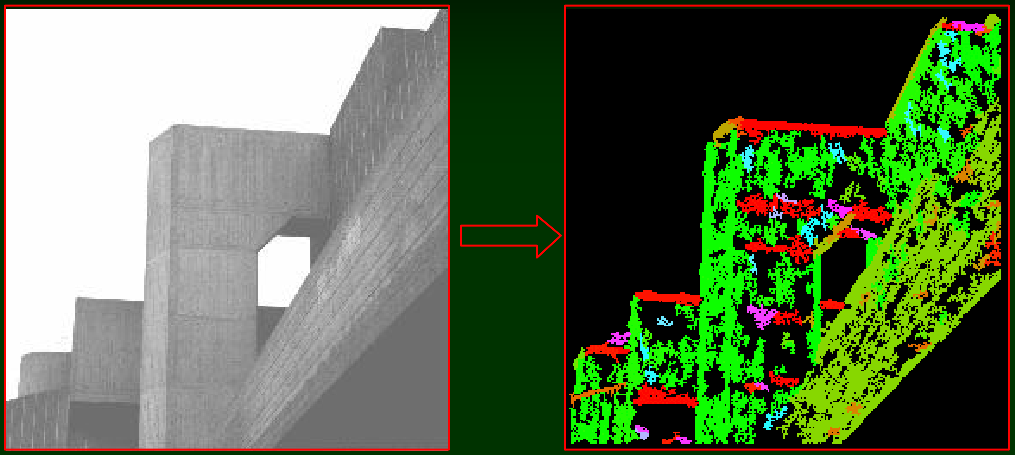


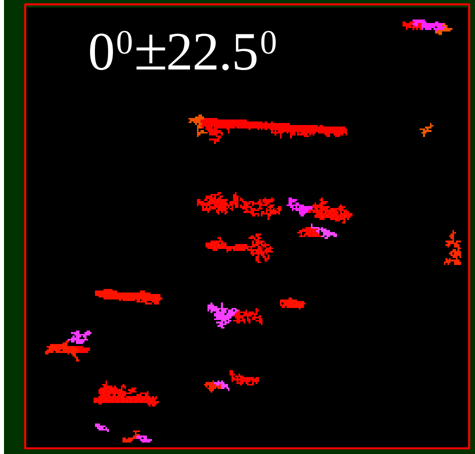
some tests

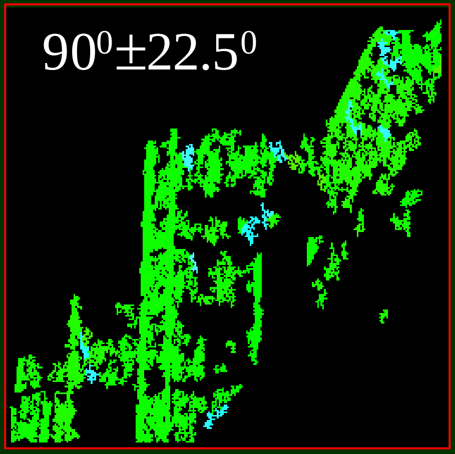


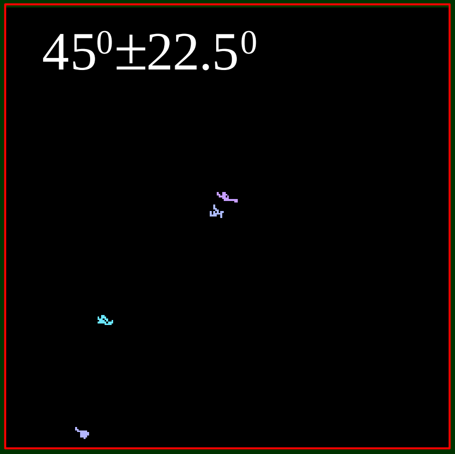
Red = horizontal   Green = vertical

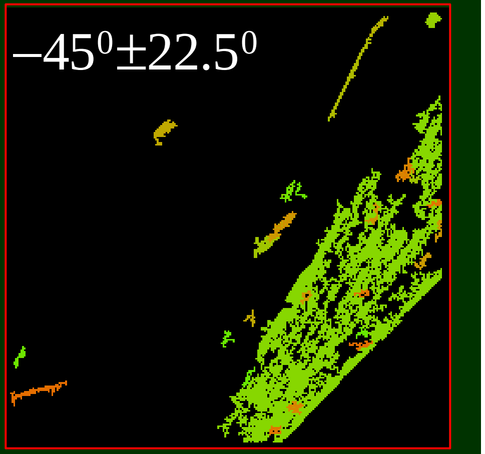
natural images

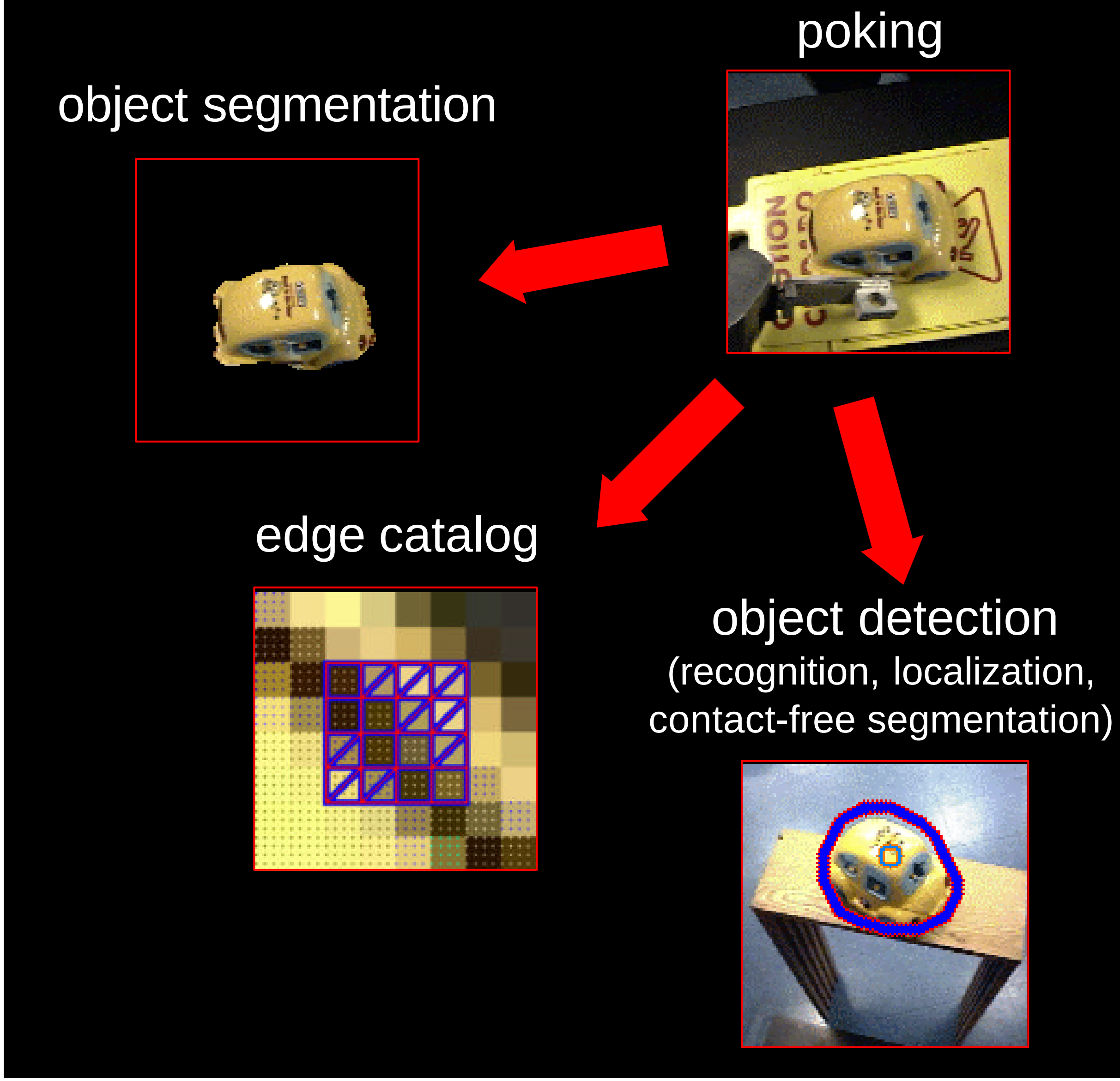












poking

object segmentation

edge catalog

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

“Open Object Recognition”

detecting and recognizing
familiar objects,
enrolling unfamiliar objects

object recognition

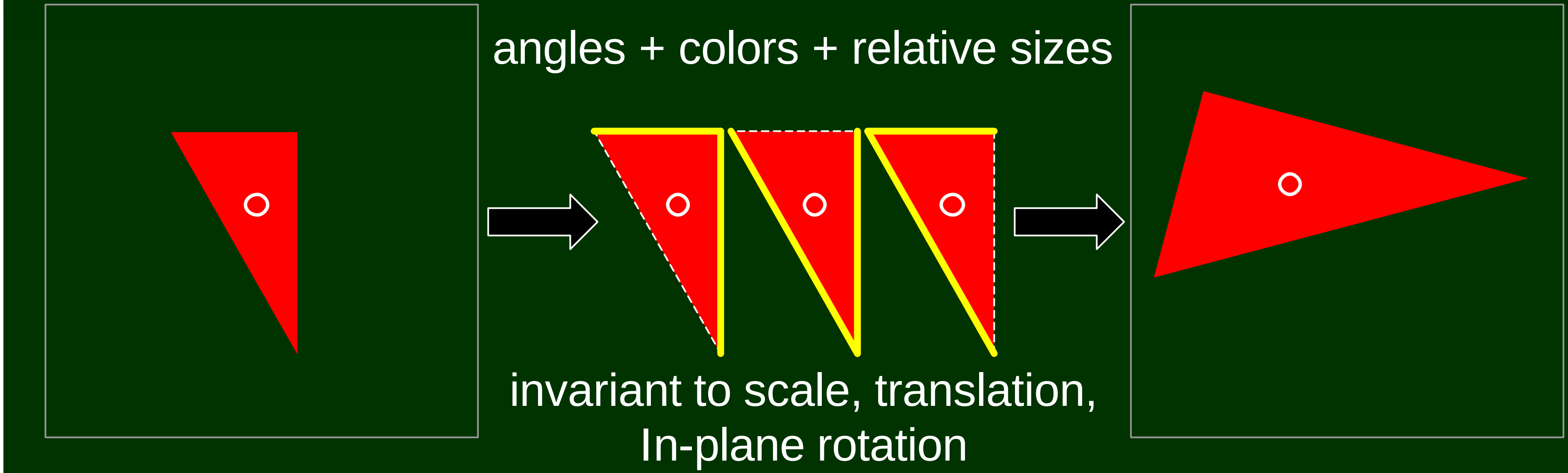
Geometry-based

- 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


angles + colors + relative sizes
invariant to scale, translation,
In-plane rotation

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

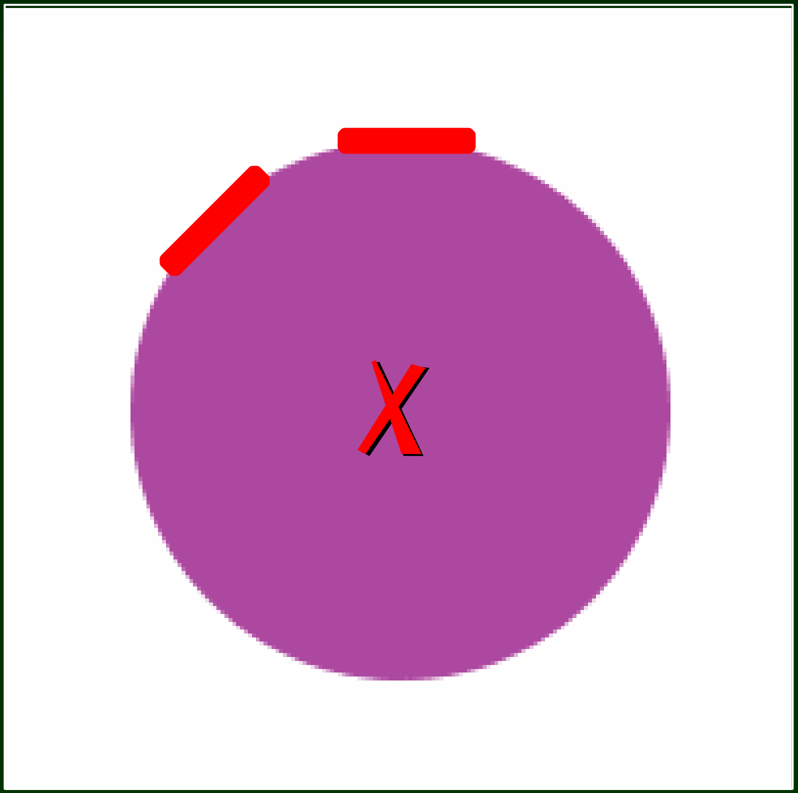
Output of feature match:

- Predicts approximate center and scale of object if match exists

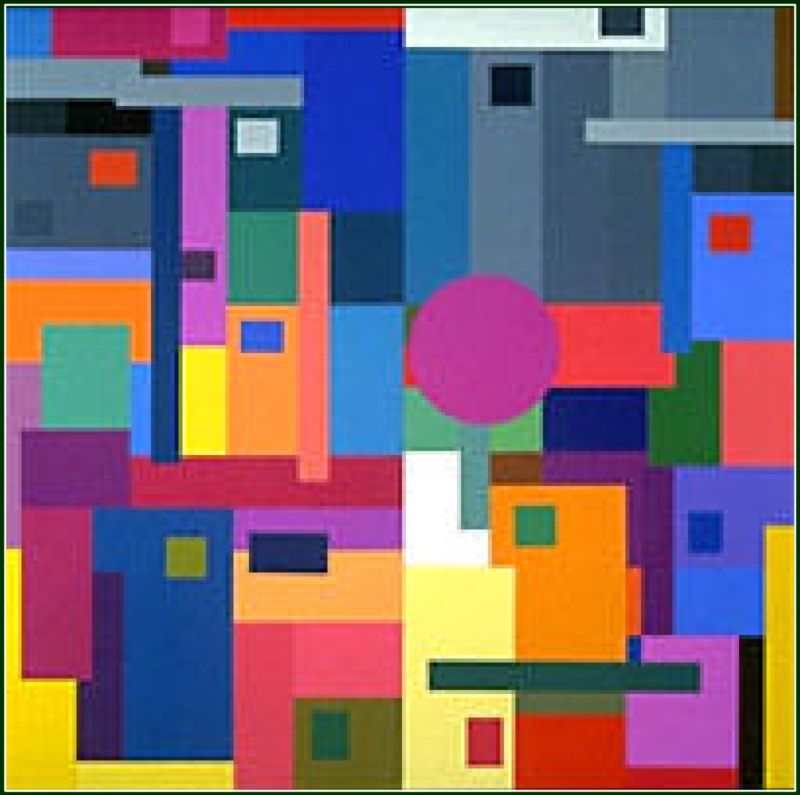
Weighting for combining features:

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

localization example

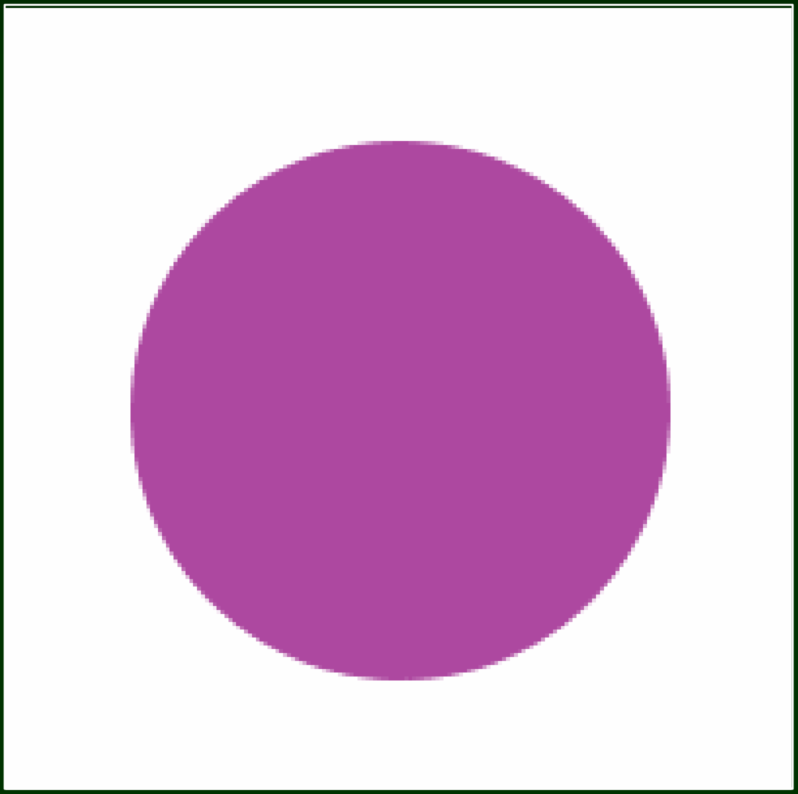


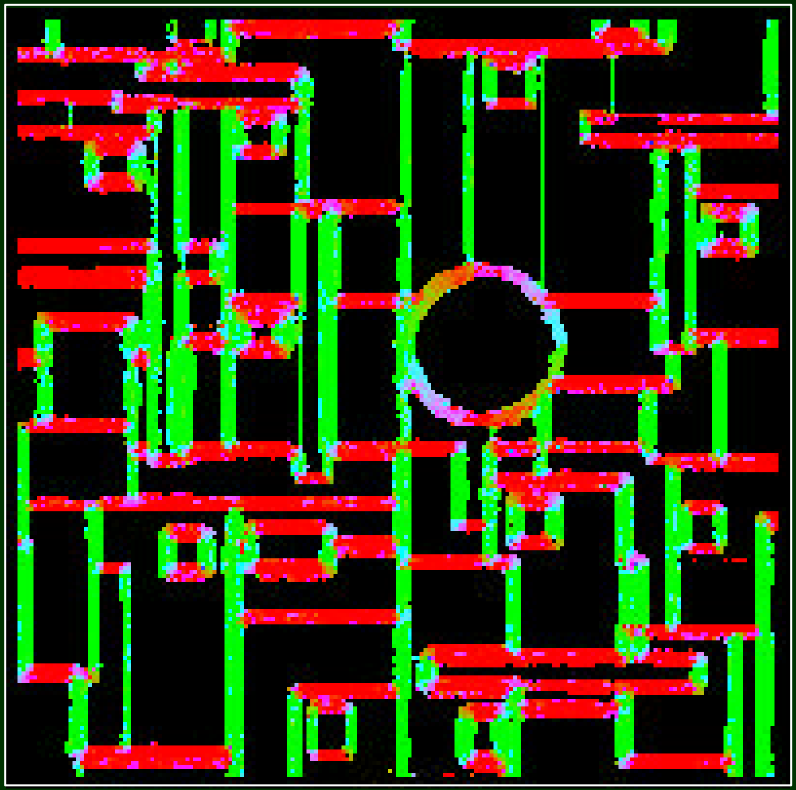
look for this…



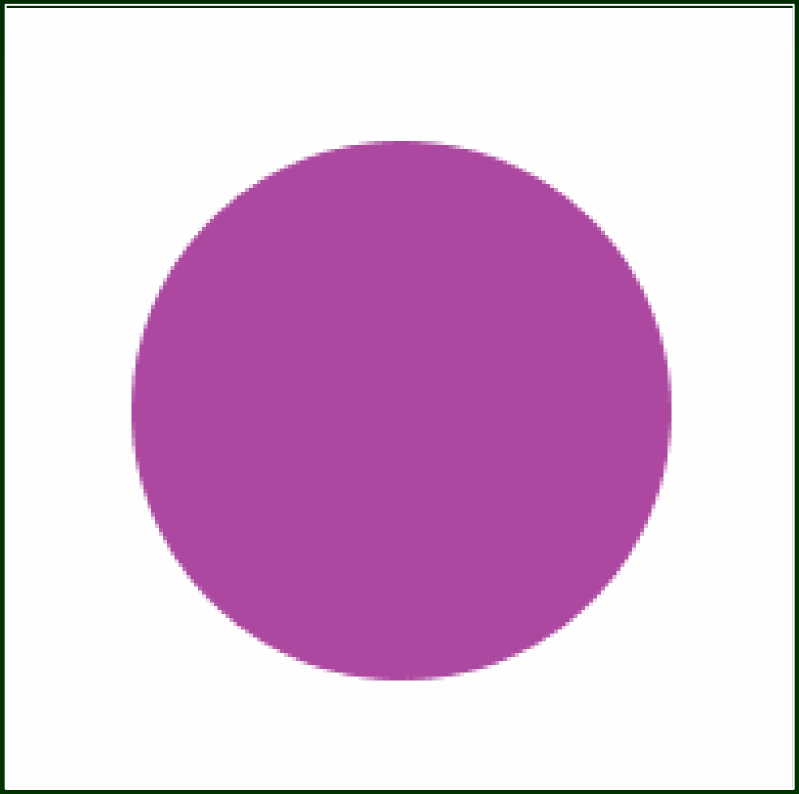
…in this

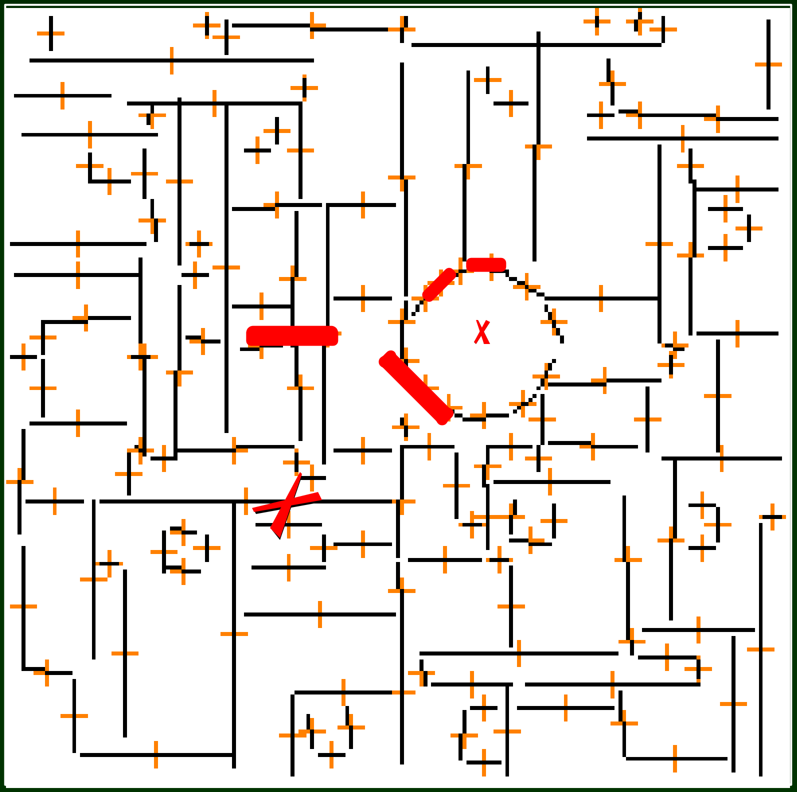
localization example





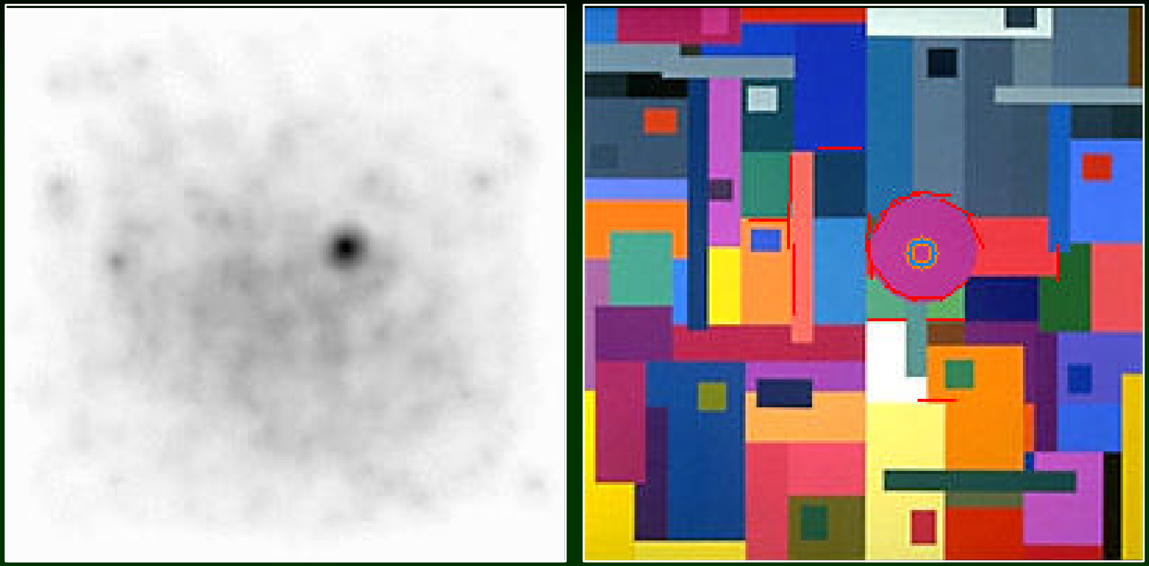
localization example



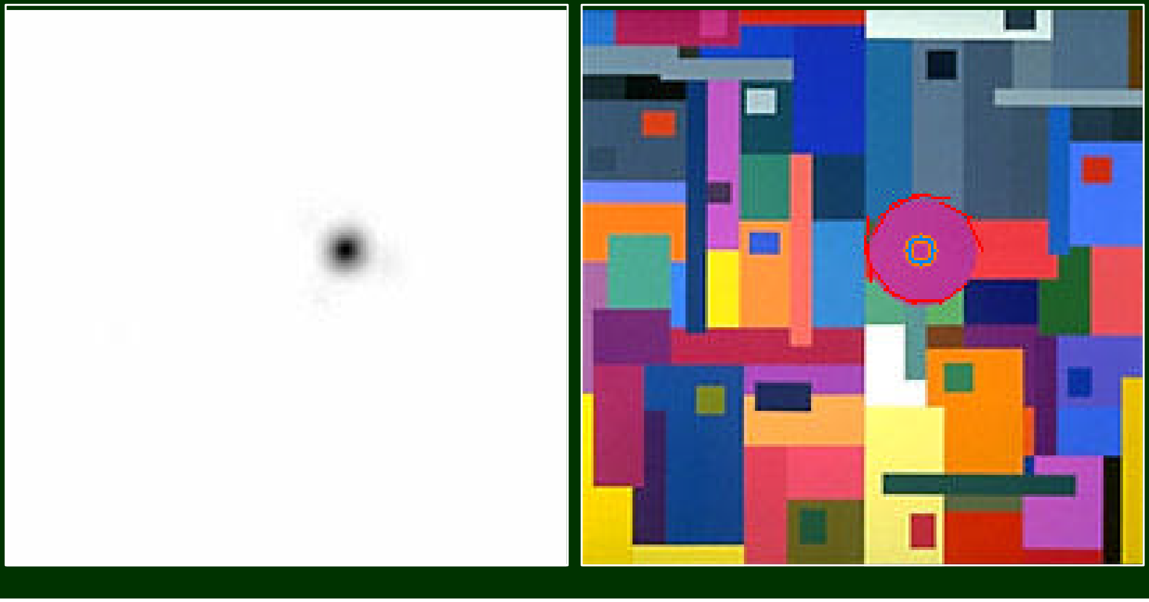


localization example

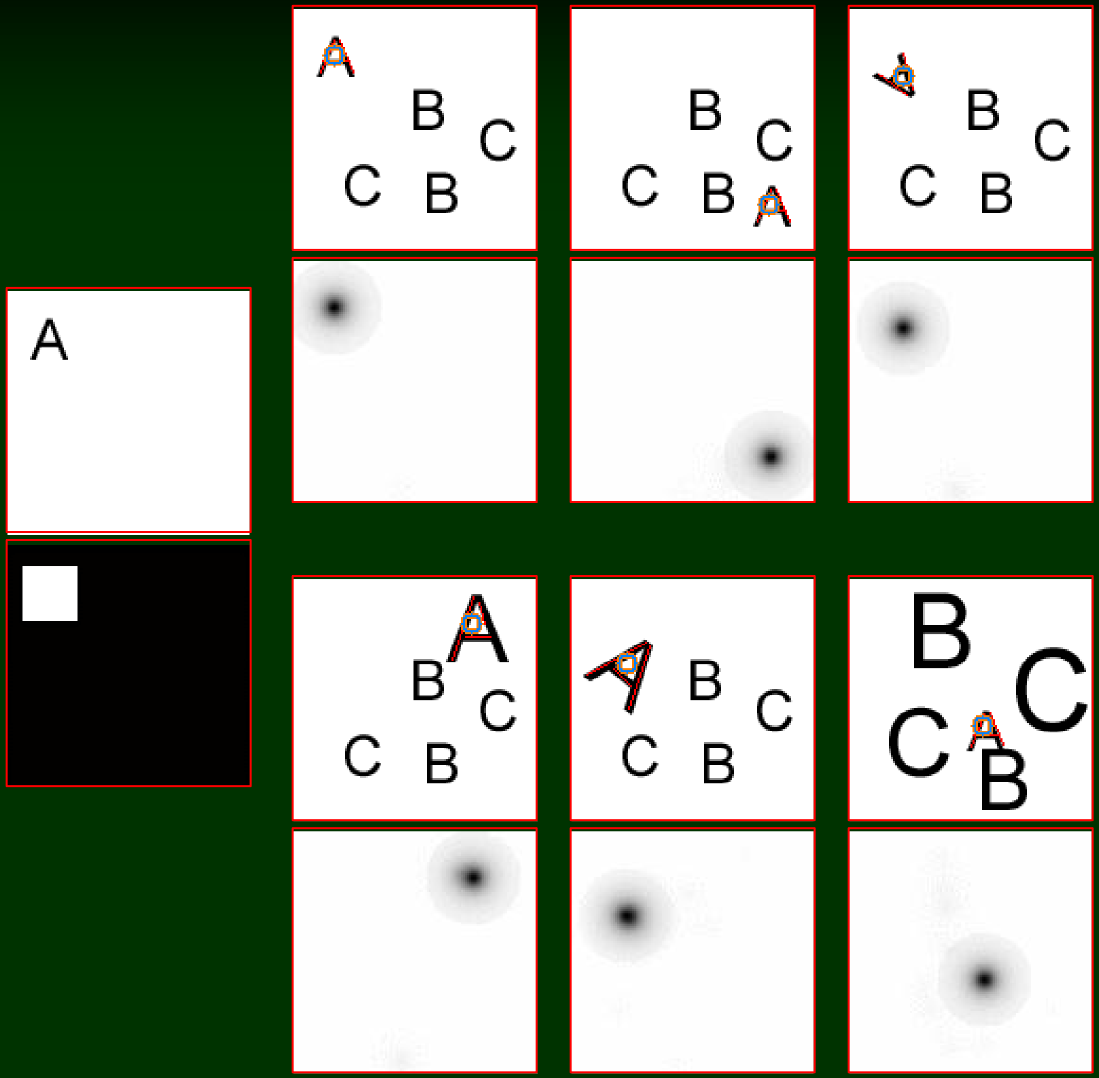
just using geometry



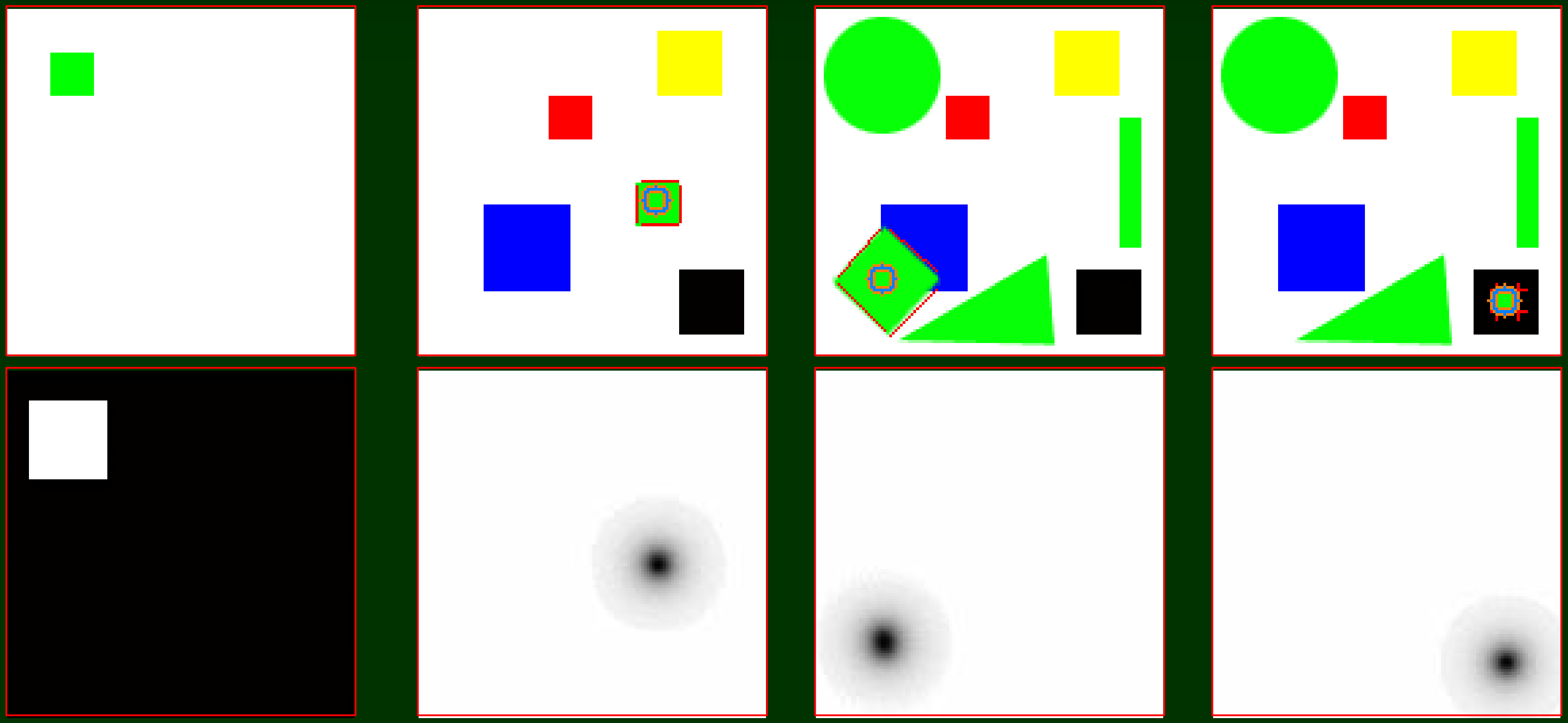
geometry + appearance



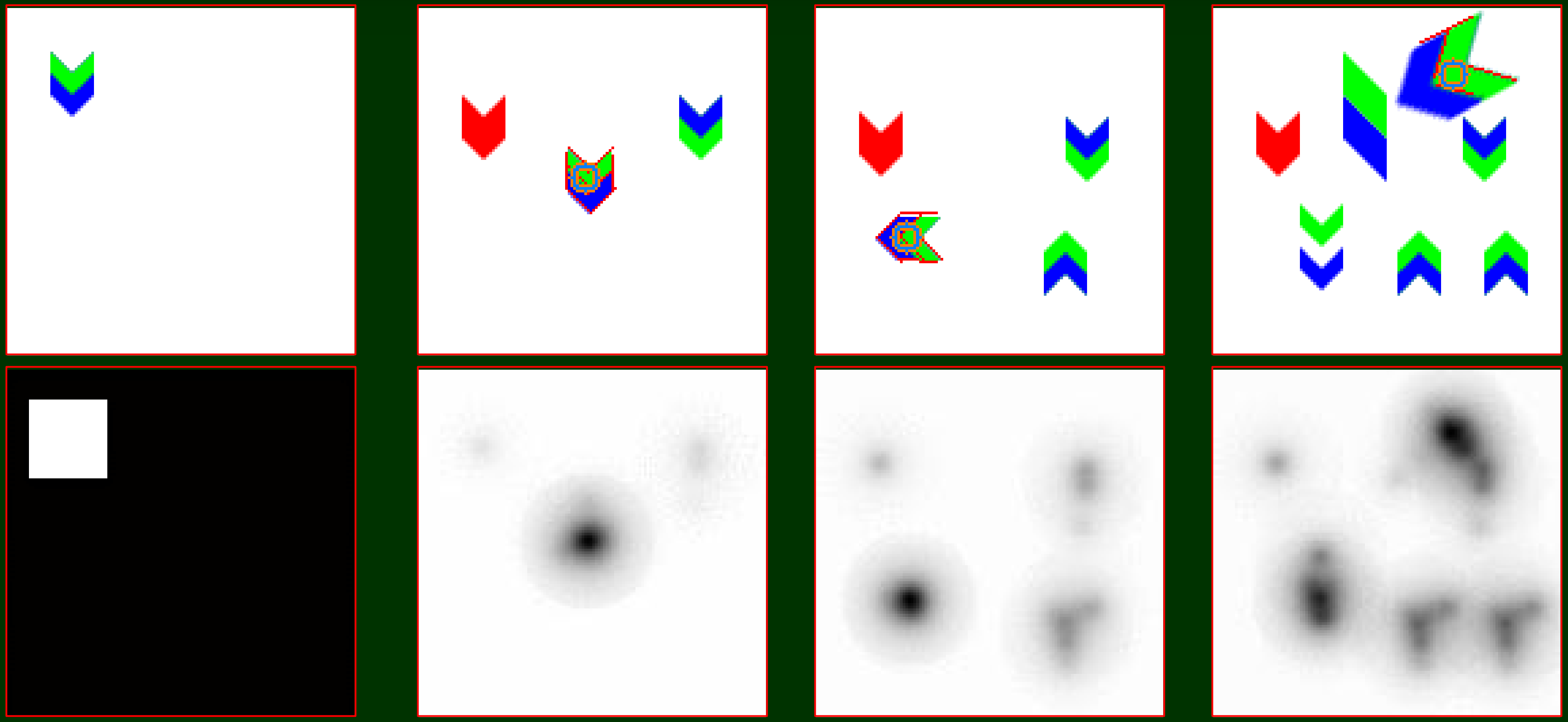
other examples



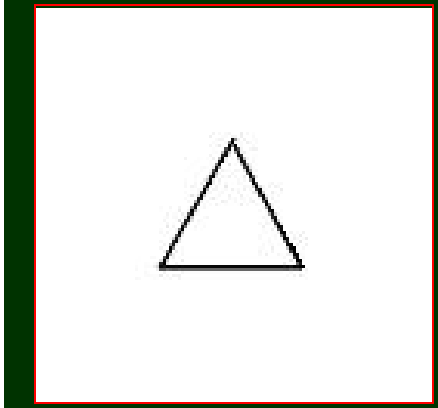
other examples



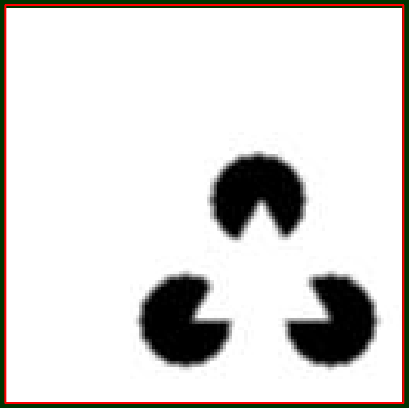
other examples



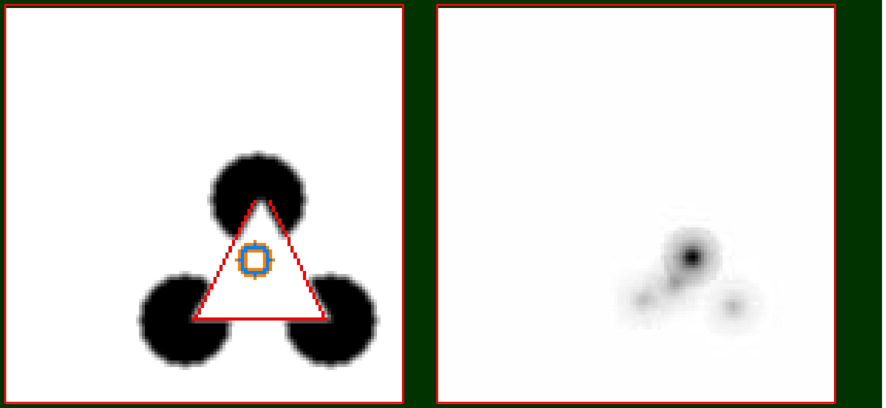
just for fun



look for this…

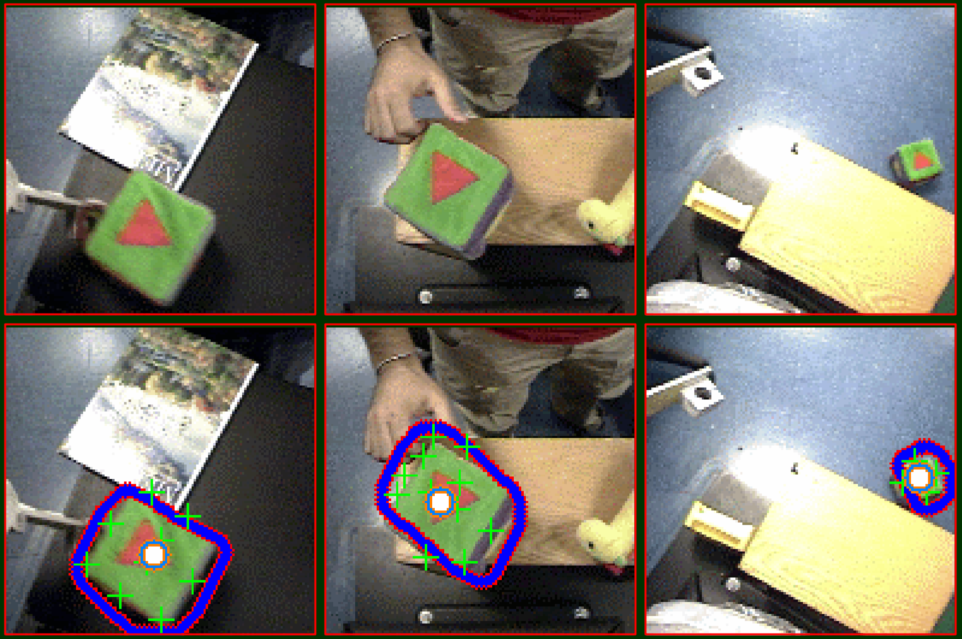


…in this



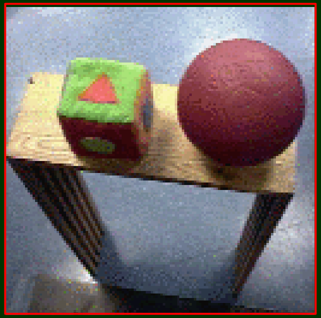
result

real object in real images

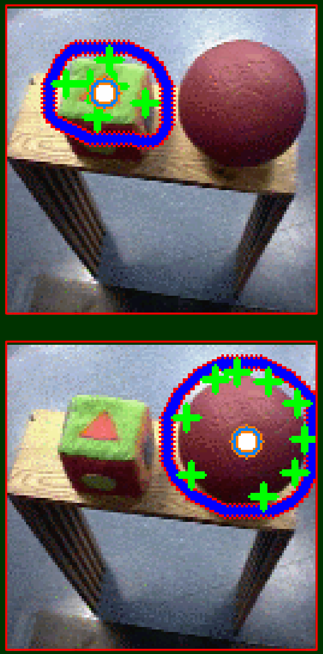


yellow on yellow

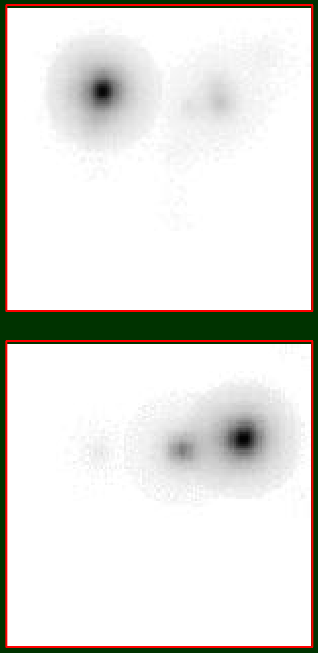
multiple objects



camera image

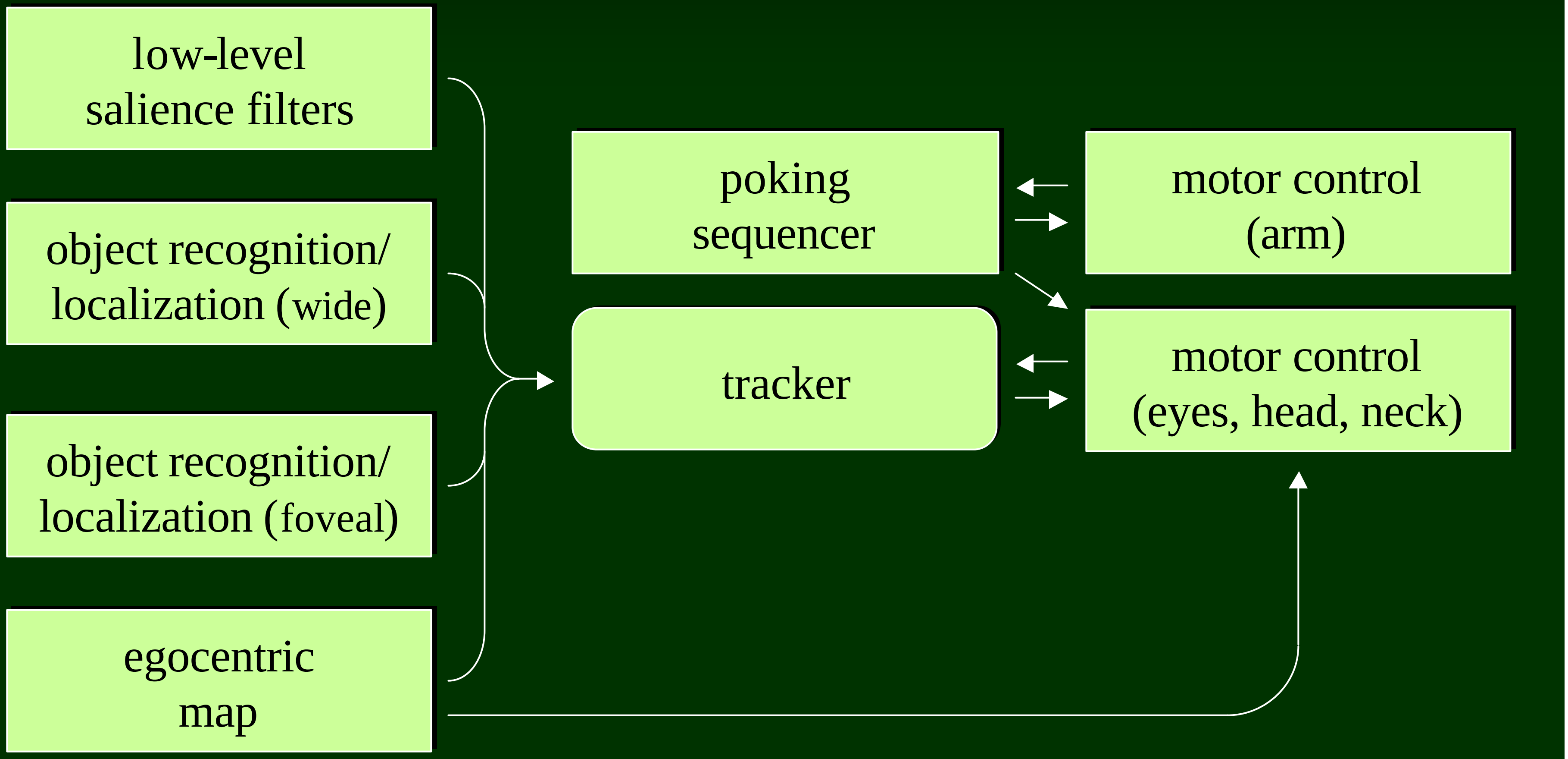


implicated edges
found and grouped



response for
each object

extending the attention system



low-level salience filters

object recognition/ localization (wide)

object recognition/ localization (foveal)

egocentric map

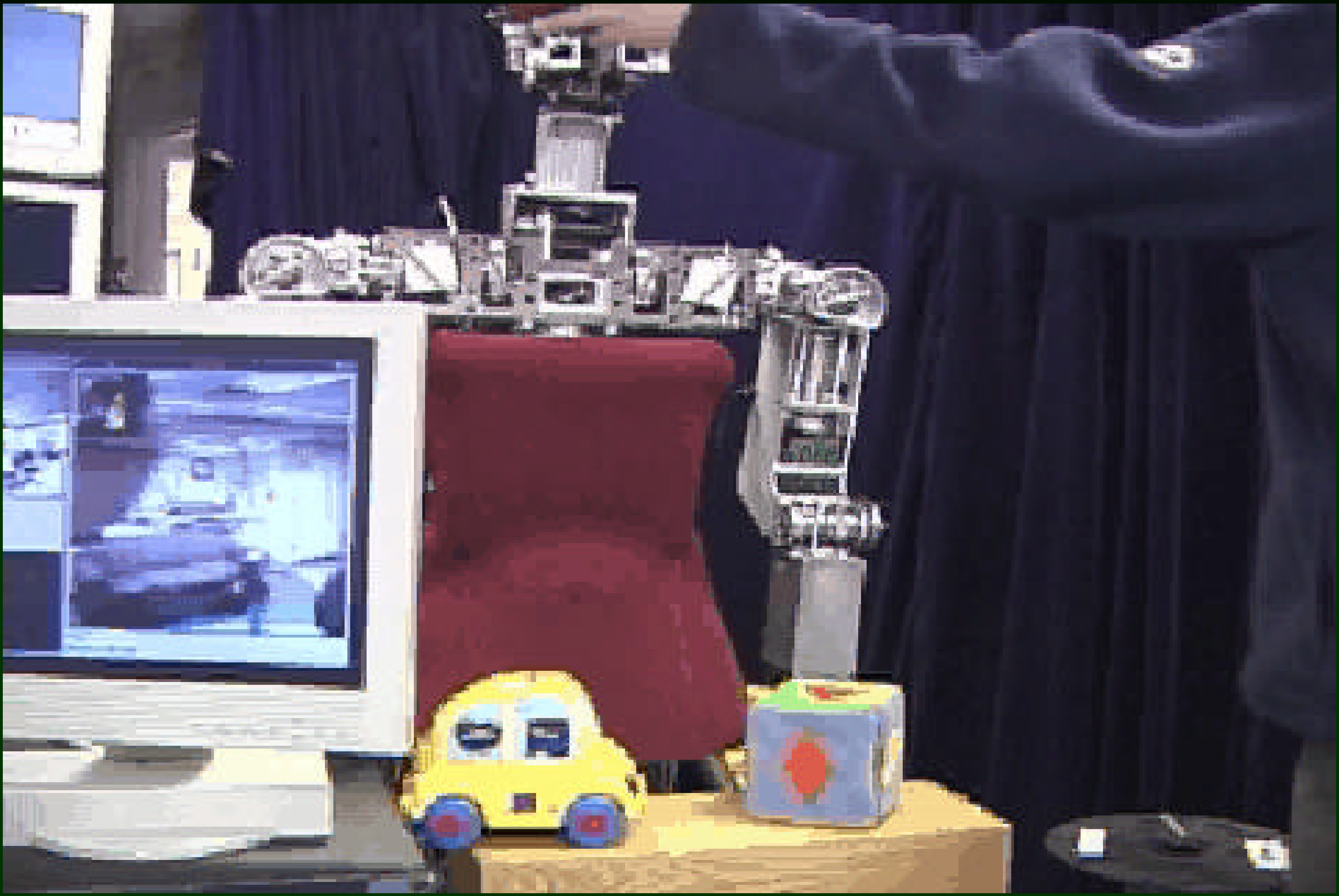
poking sequencer

tracker

motor control (arm)

motor control (eyes, head, neck)

attention



working with one example

Method gives best-guess location

Location evaluated to determine if object is really there

Thresholds determined by variation in match strengths seen over all examples

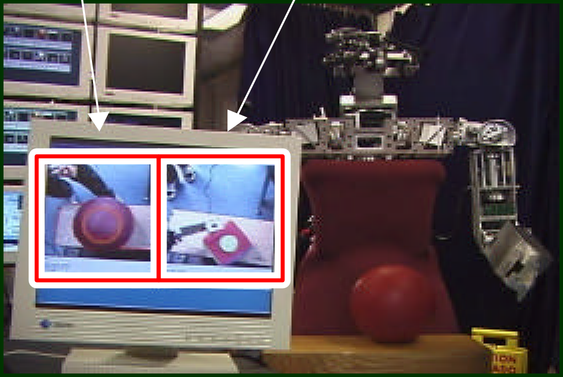
Hard to set sensibly with only one example

Solution: be tolerant, and allow for online correction

open object recognition

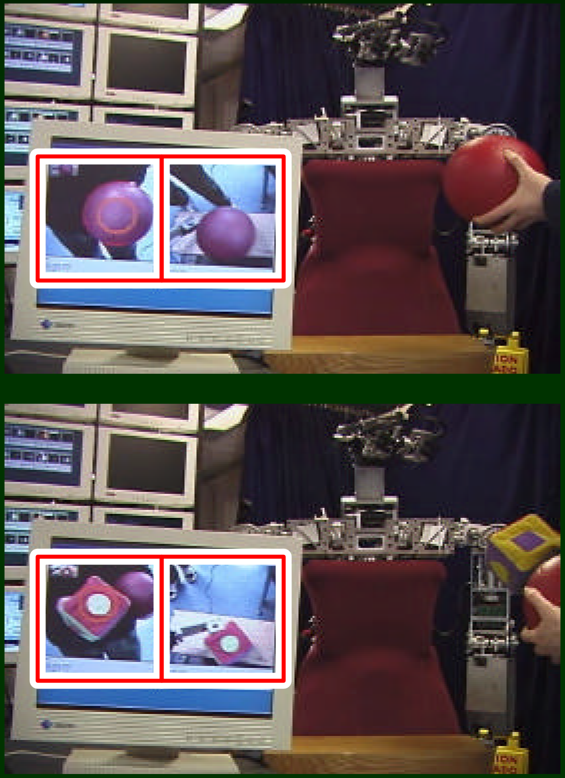
robot's current view

recognized object (as seen during poking)



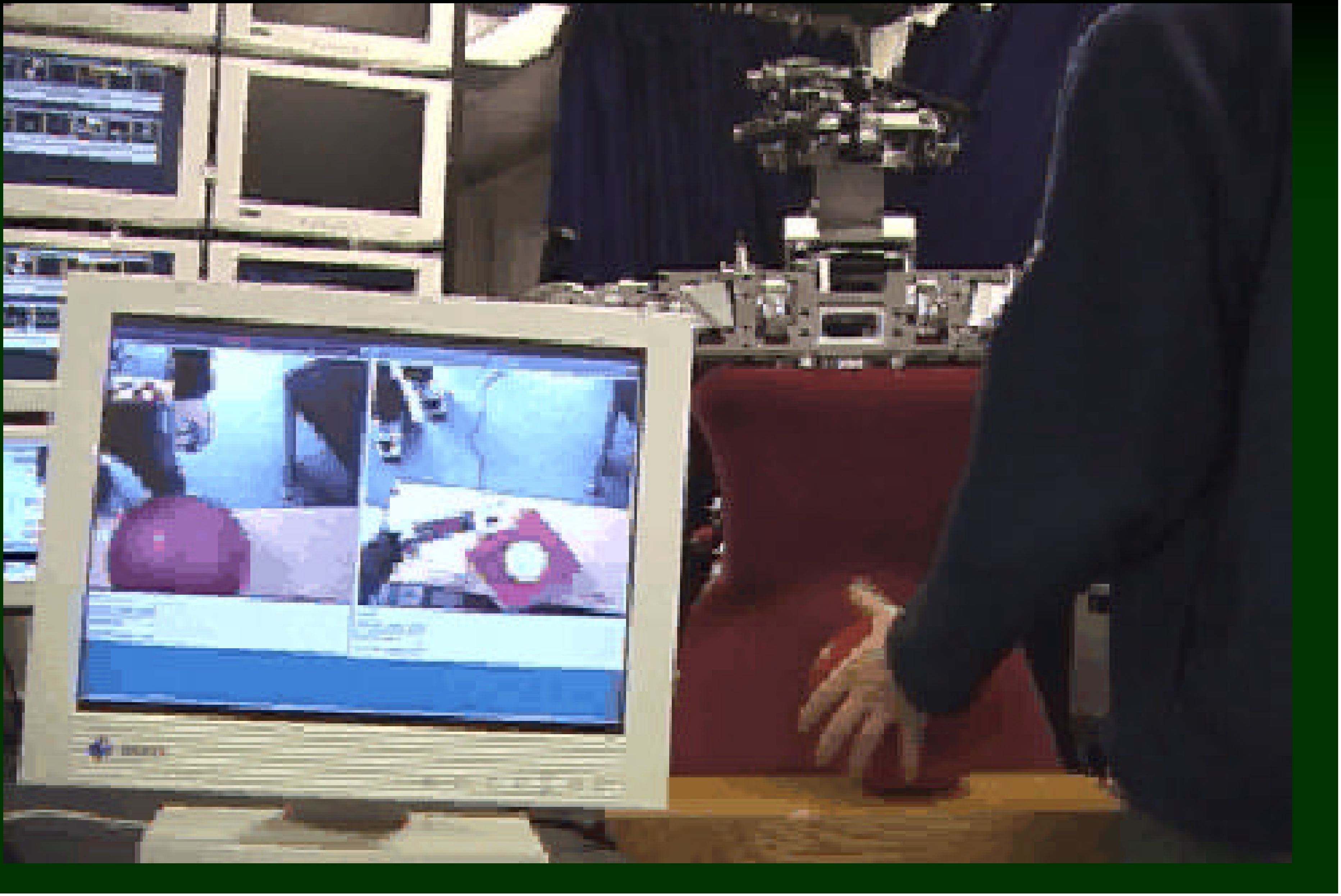
sees ball,
"thinks" it is cube

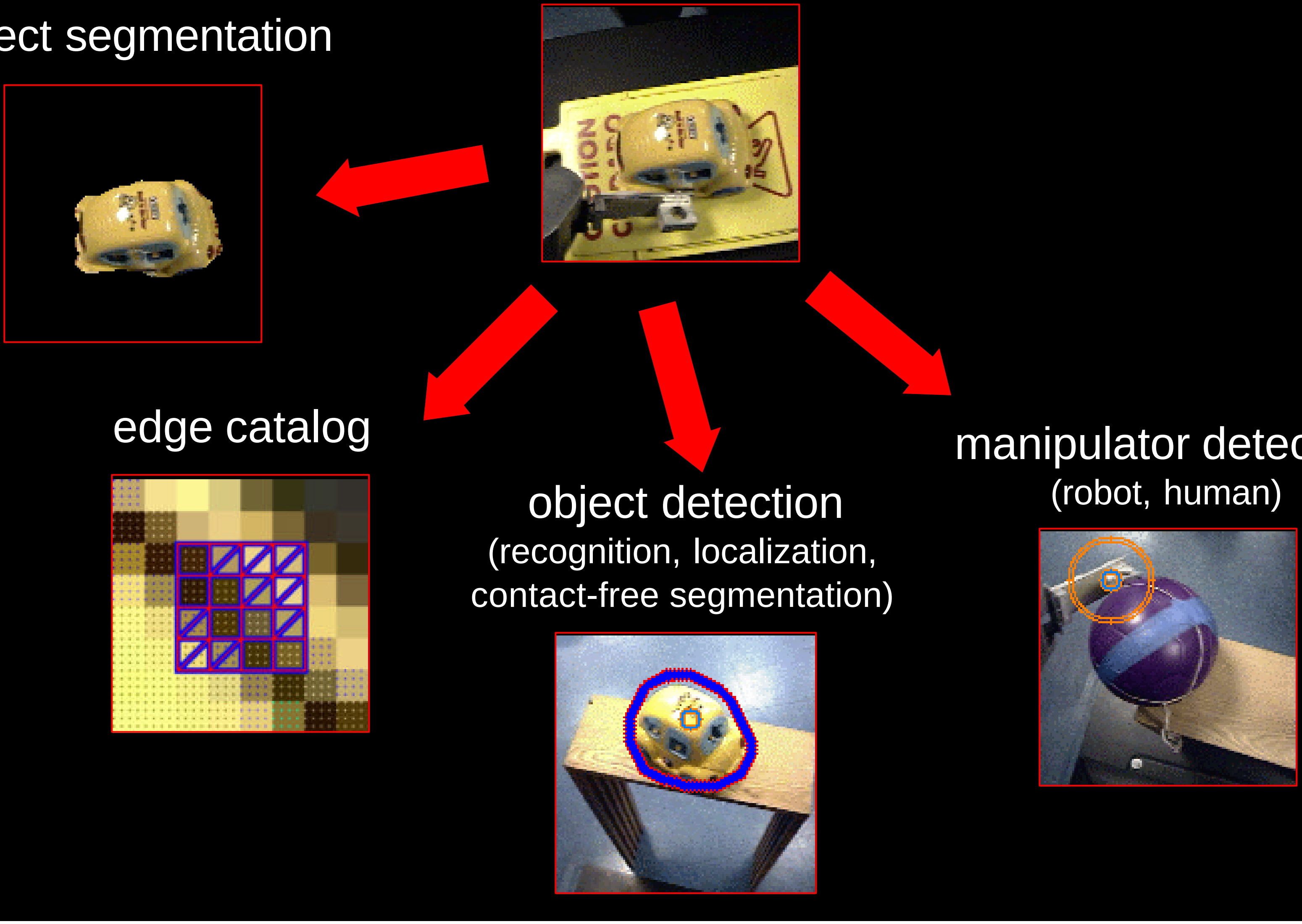
pokes,
segments
ball



correctly differentiates
ball and cube

open object recognition





poking

object segmentation

edge catalog

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

manipulator detection
(robot, human)

finding manipulators

Analogous to finding objects

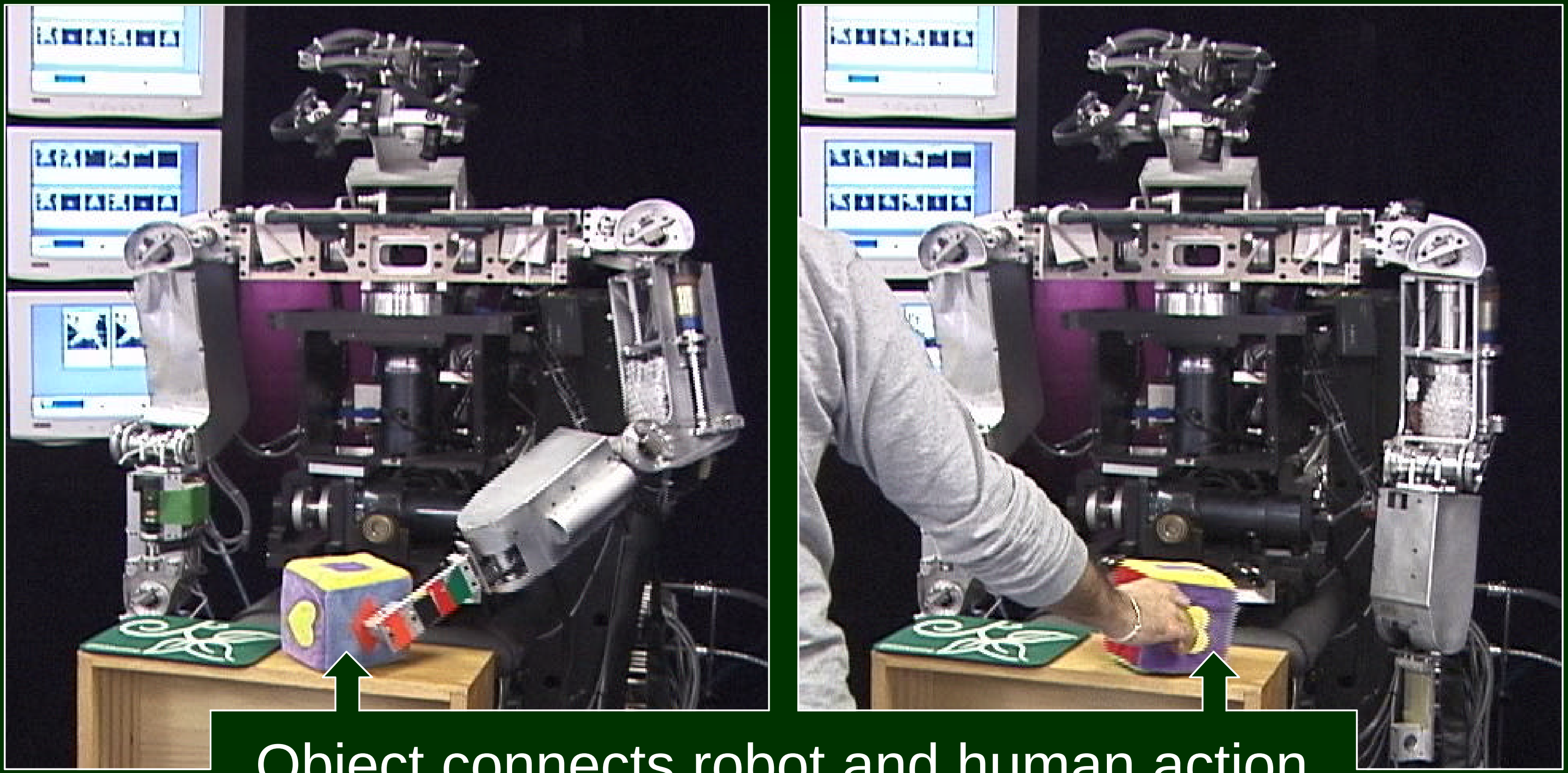
Object

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

Actor

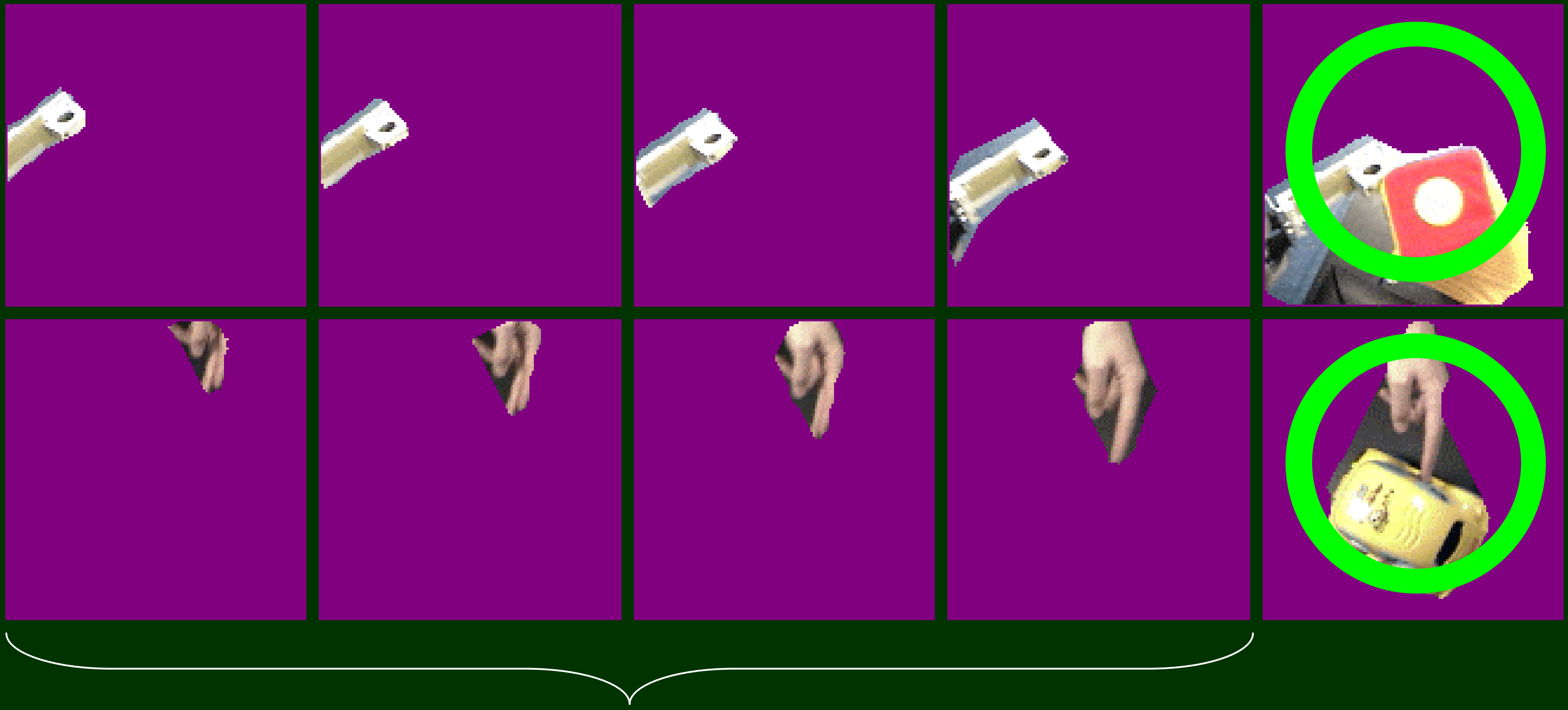
- *Definition*: something that acts on objects
- *How to find one*: see what pokes objects

similar human and robot actions



Object connects robot and human action

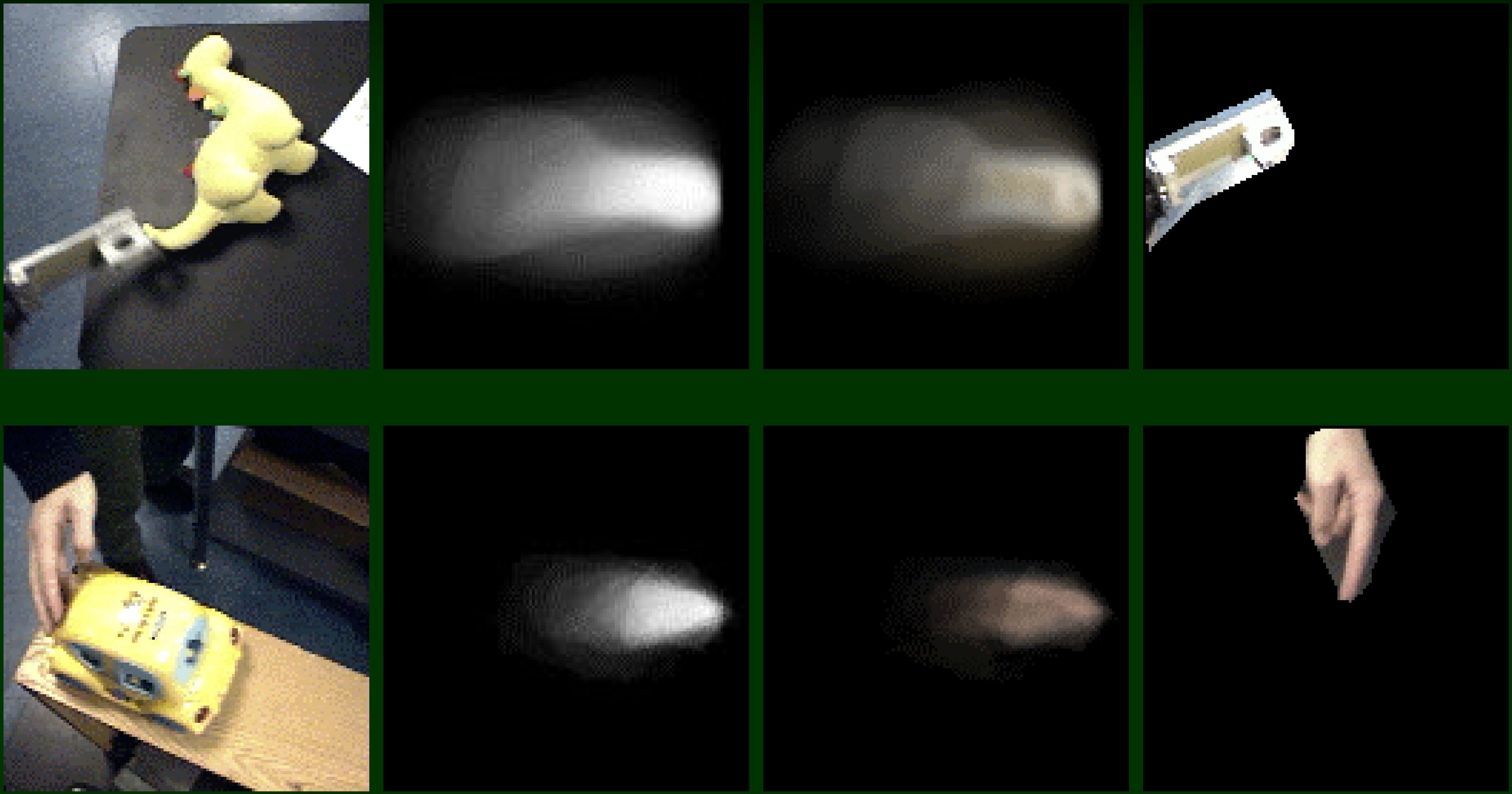
catching manipulators in the act



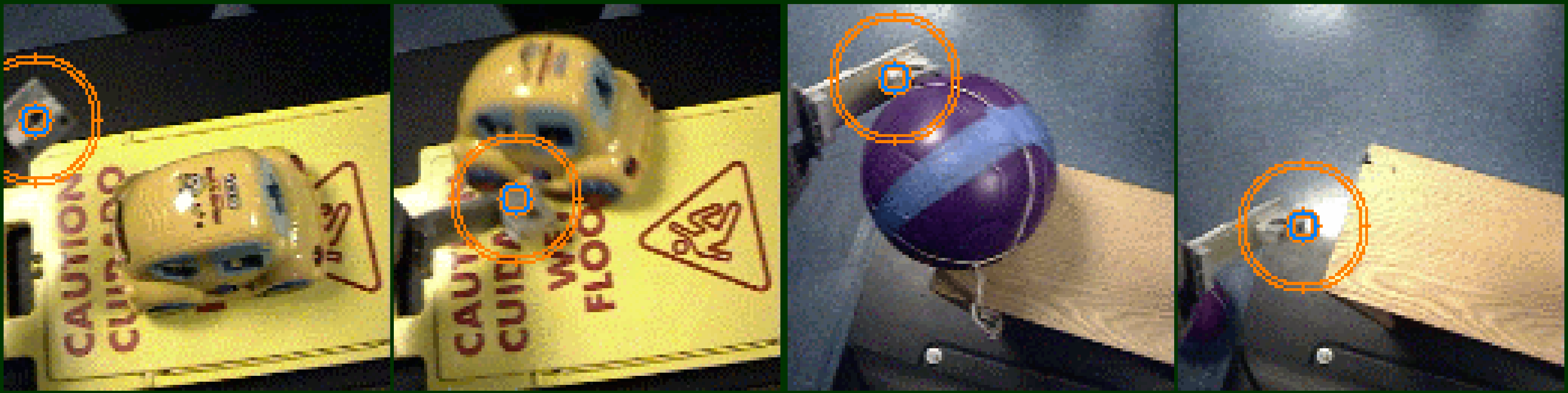
manipulator approaches object

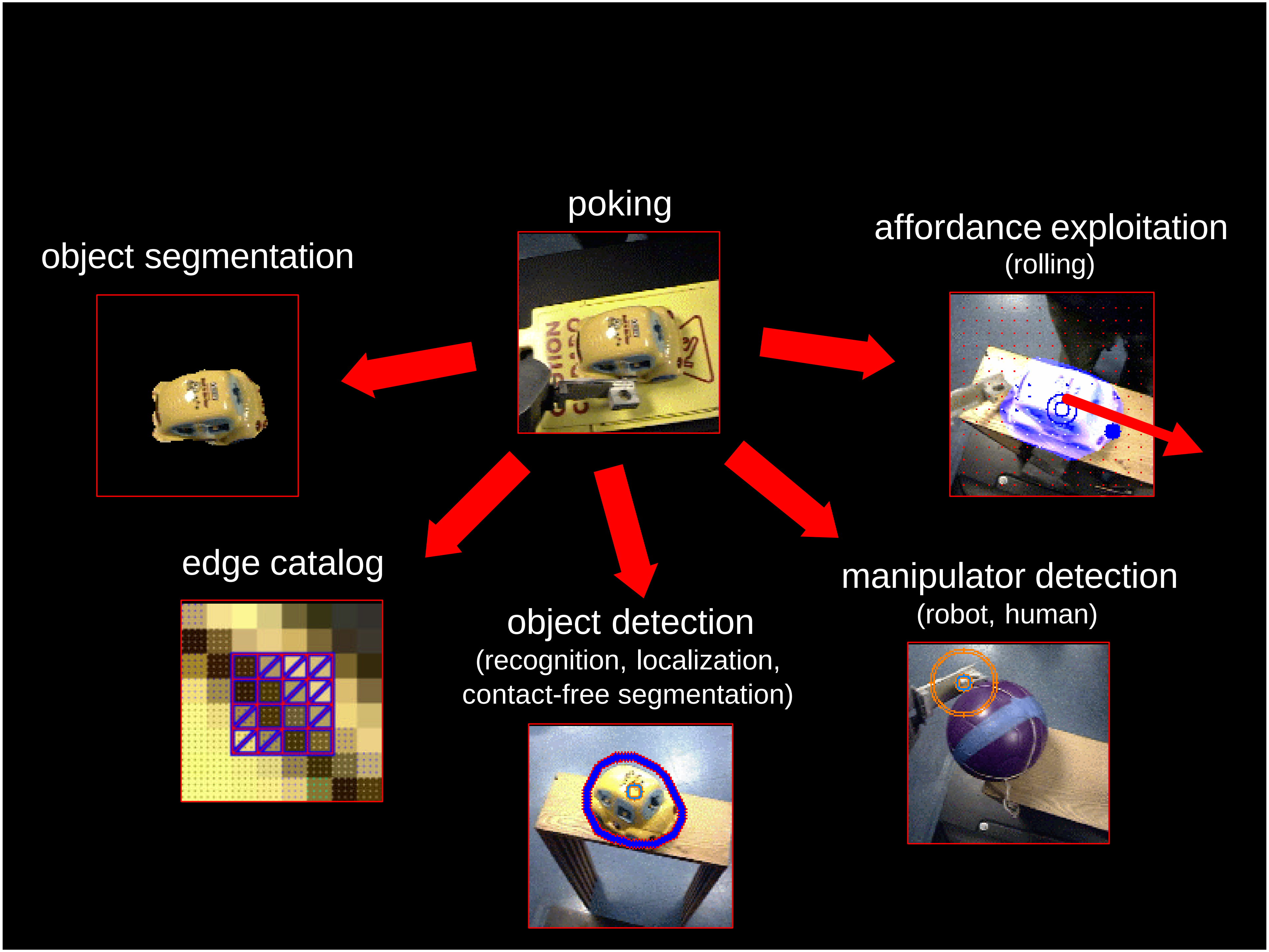
contact!

modeling manipulators



manipulator recognition





poking

object segmentation

affordance exploitation
(rolling)

edge catalog

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

manipulator detection
(robot, human)

“Affordance Recognition”

switching from
object-centric perception
to recognizing action opportunities

(collaboration with Giorgio Metta)

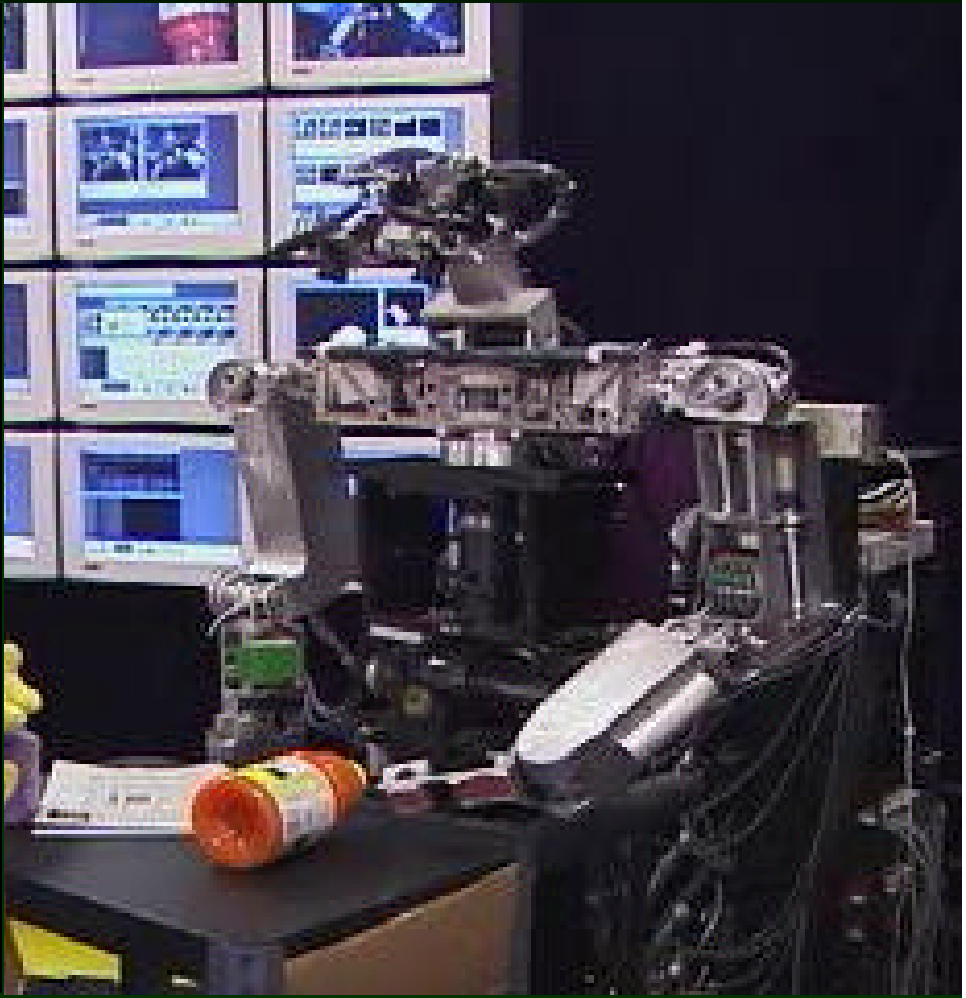
what is an affordance?

A leaf affords

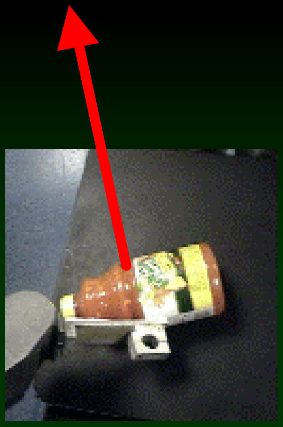
rest/walking to an ant …

… but not to an elephant

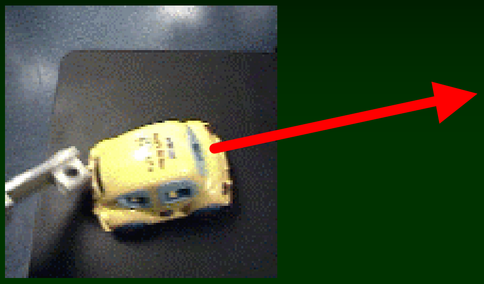
exploring affordances



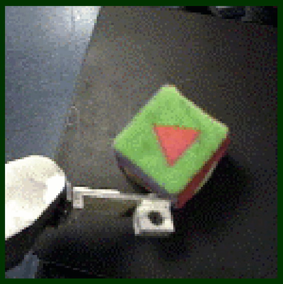
objects roll in different ways



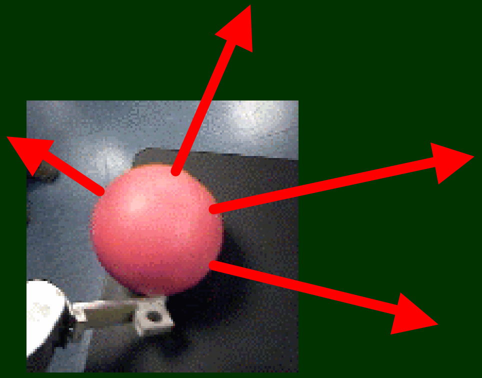
a bottle
it rolls along its side



a toy car
it rolls forward

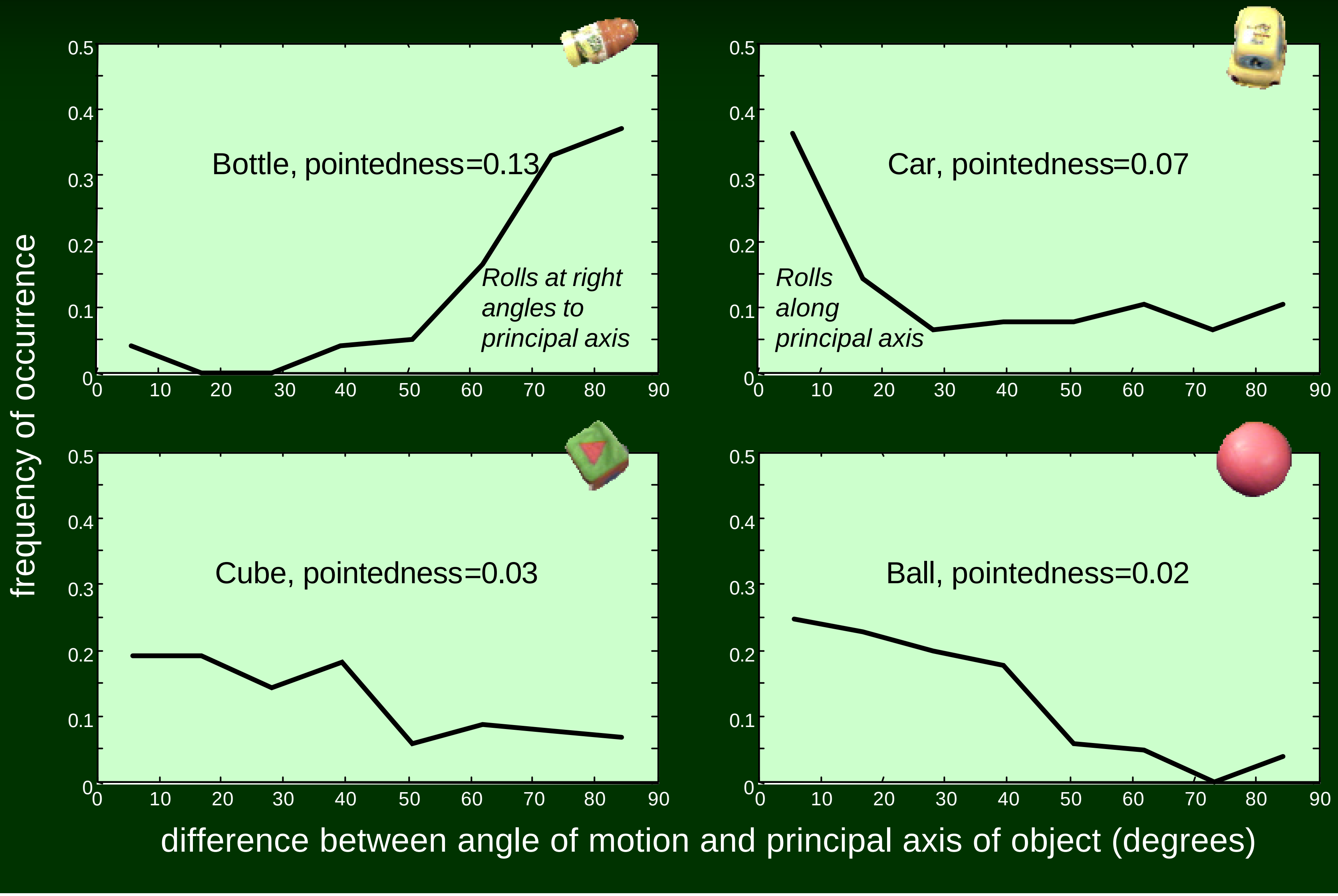


a toy cube
it doesn’t roll easily



a ball
it rolls in
any direction

preferred direction of motion



Bottle, pointedness=0.13

Rolls at right angles to principal axis

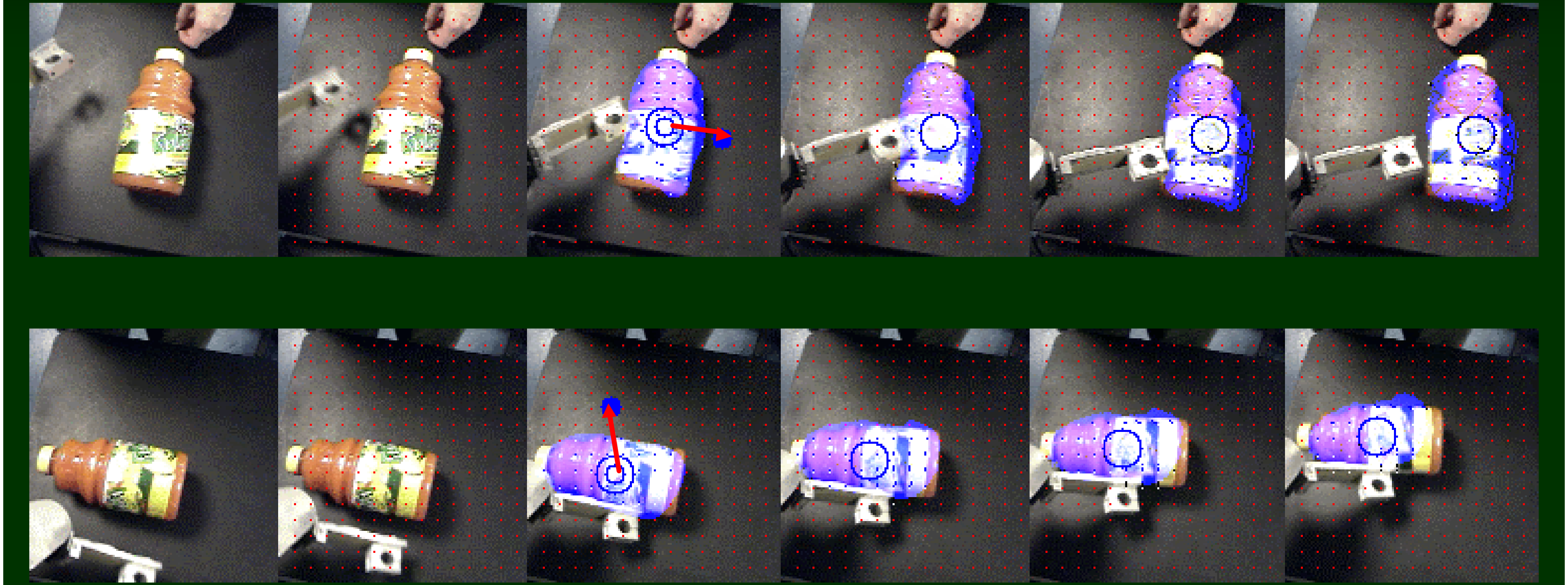
Car, pointedness=0.07

Rolls along principal axis

Cube, pointedness=0.03

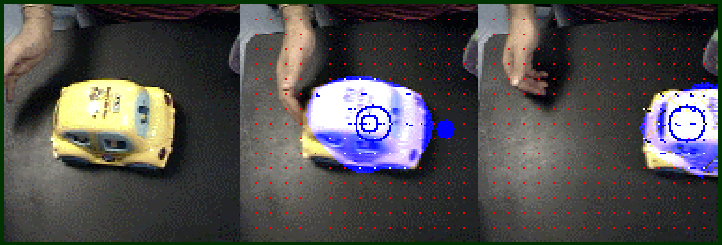
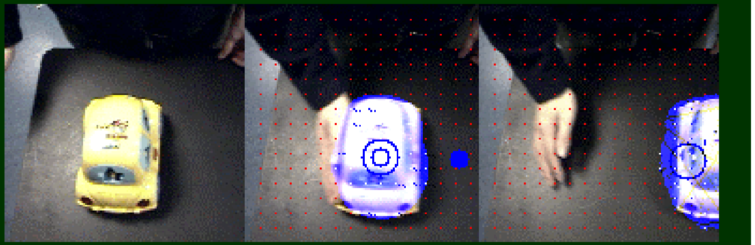
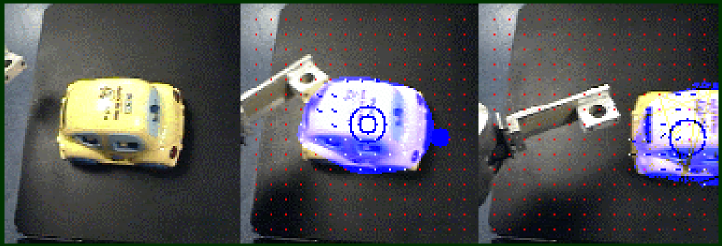
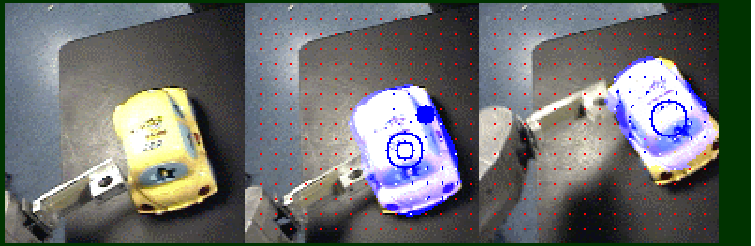
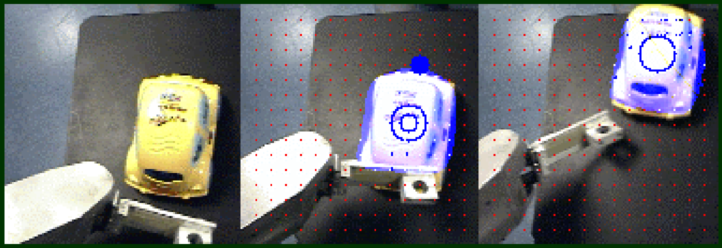
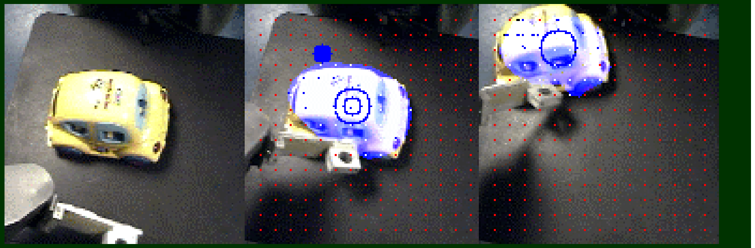
Ball, pointedness=0.02

affordance exploitation

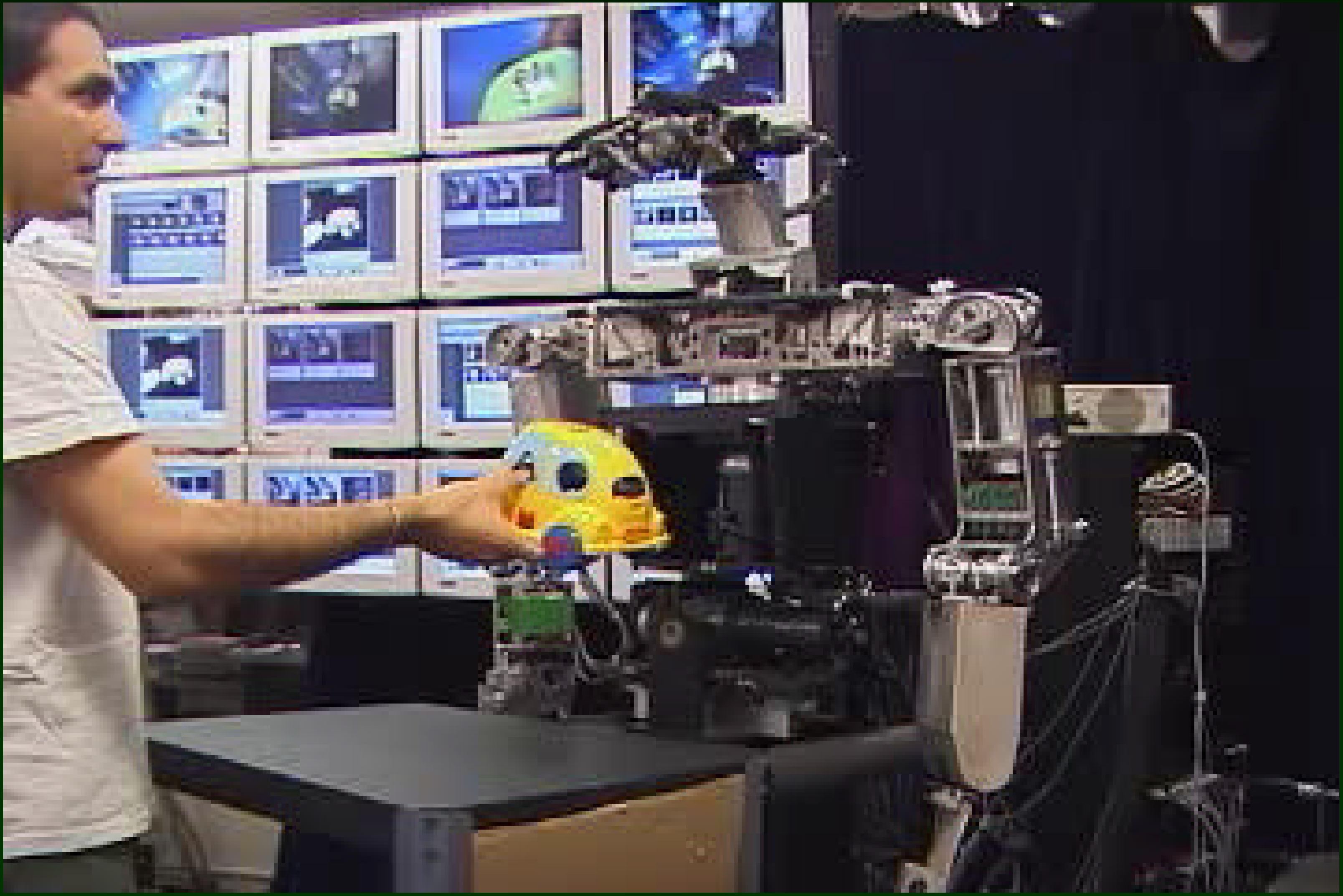


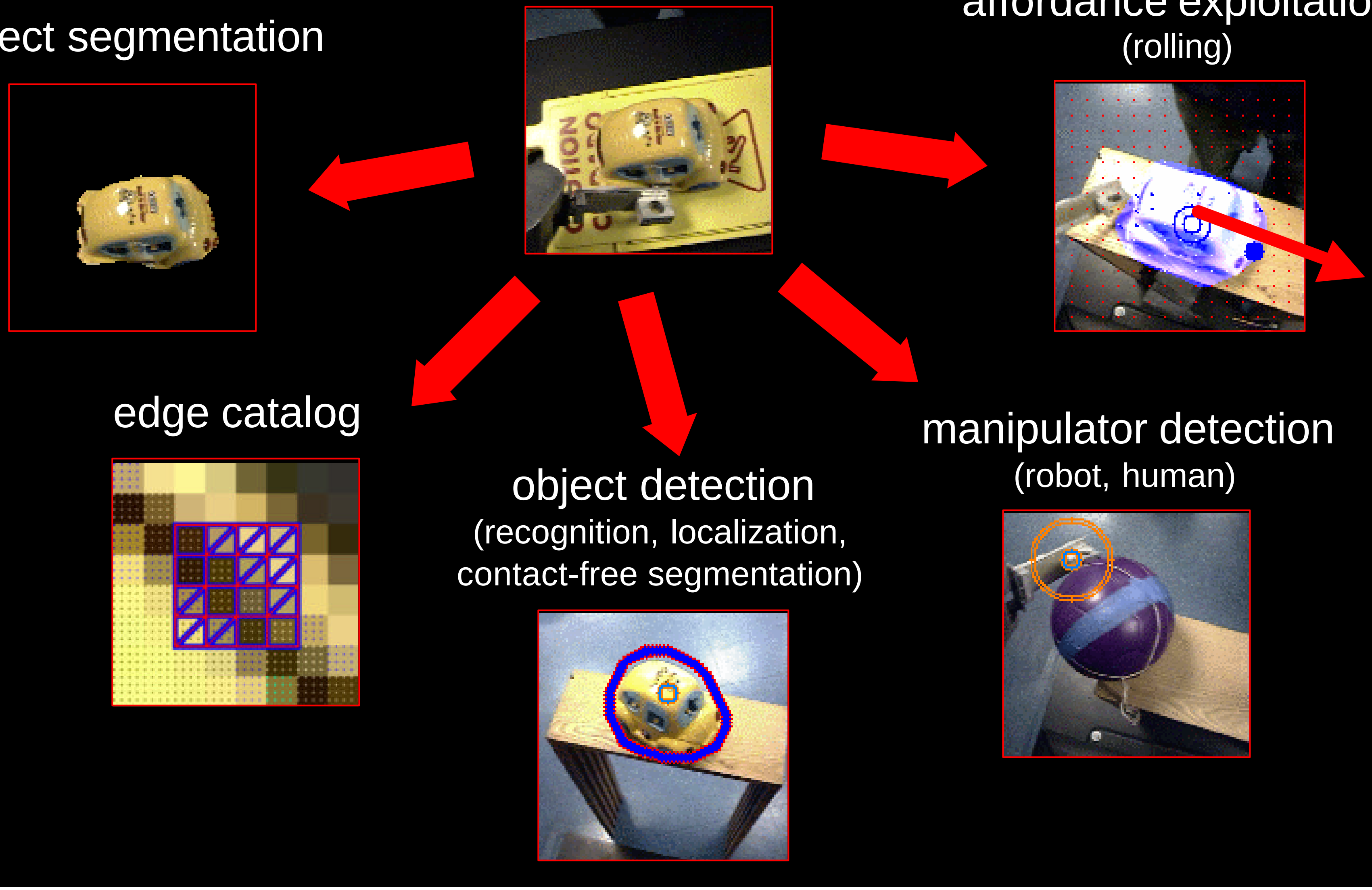
Caveat: this work uses an early version of object detection (not the one presented today)

mimicry test

	Invoking the object's natural rolling affordance	Going against the object's natural rolling affordance
Demonstration by human		
Mimicry in similar situation		
Mimicry when object is rotated		

mimicry test





poking

object segmentation

affordance exploitation
(rolling)

edge catalog

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

manipulator detection
(robot, human)

talk overview

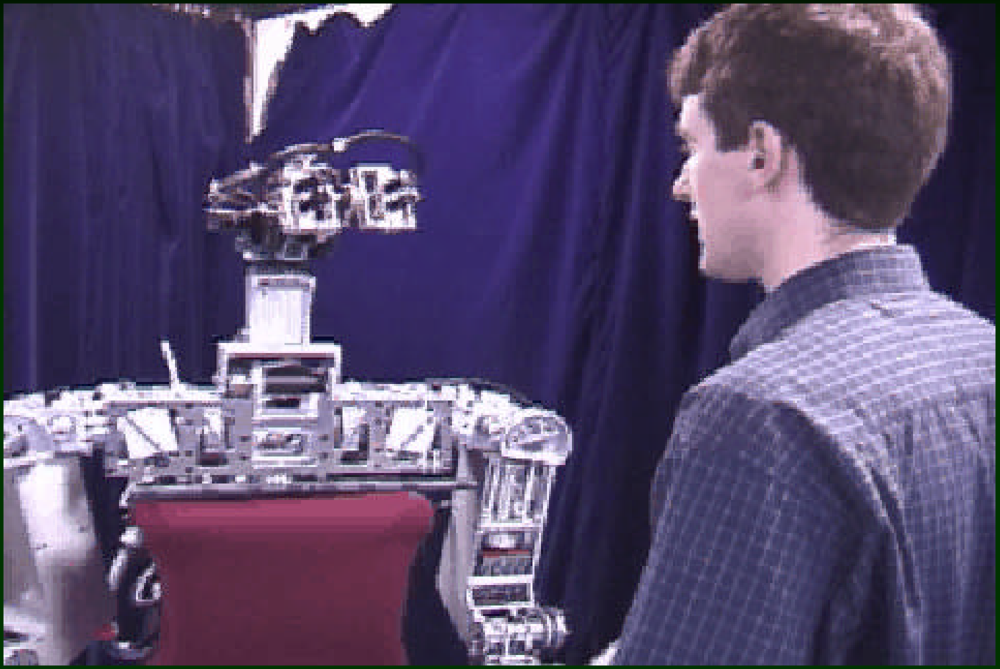
Learning from an activity

- Poking: to learn to recognize objects, manipulators, etc.
- Chatting: to learn the names of objects

Learning a new activity

- Searching for an object
- Then back to learning *from* the activity…

open speech recognition

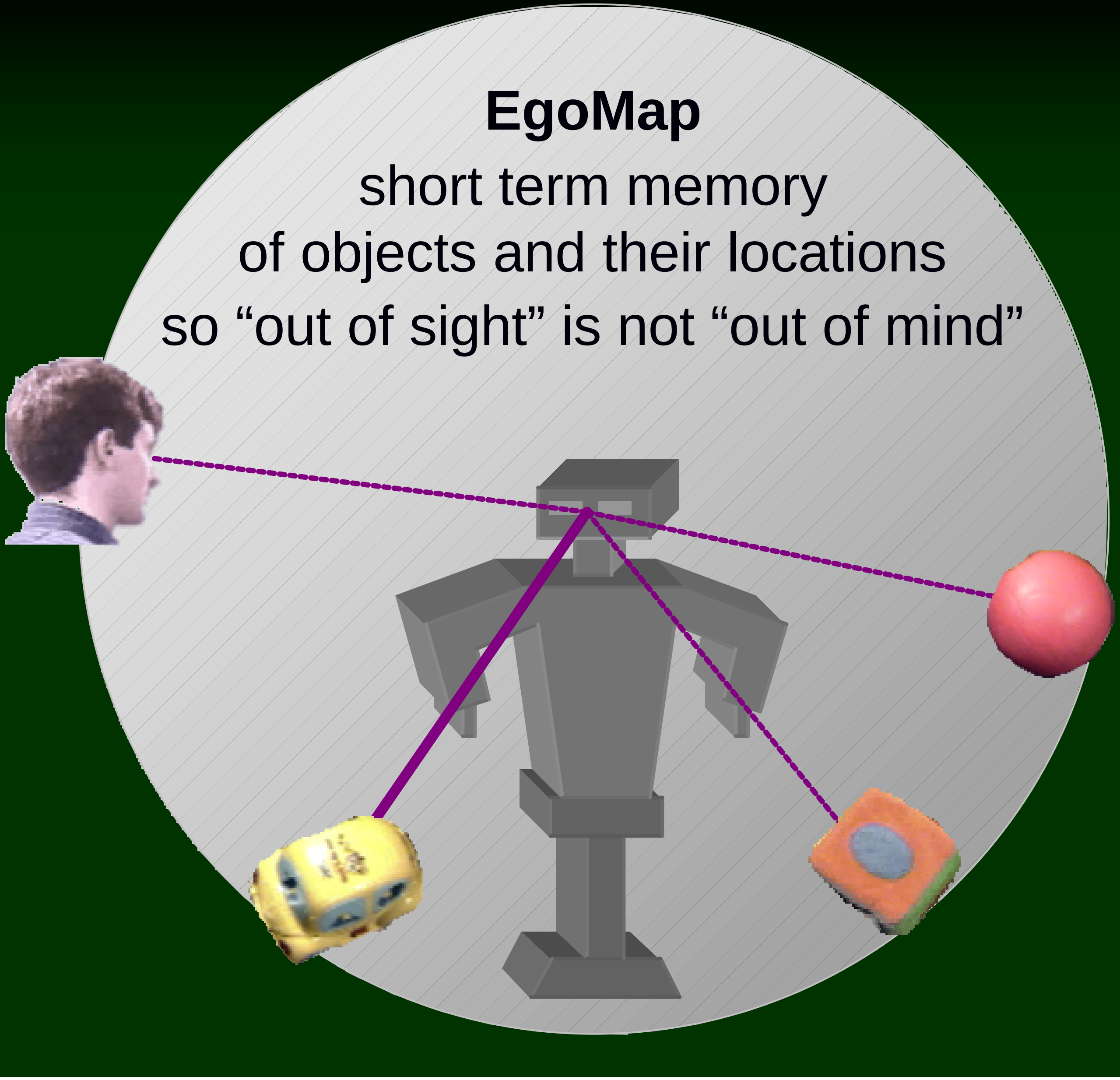


Vocabulary can be extended at any time

Assumes active vocabulary is small

Isolated words only

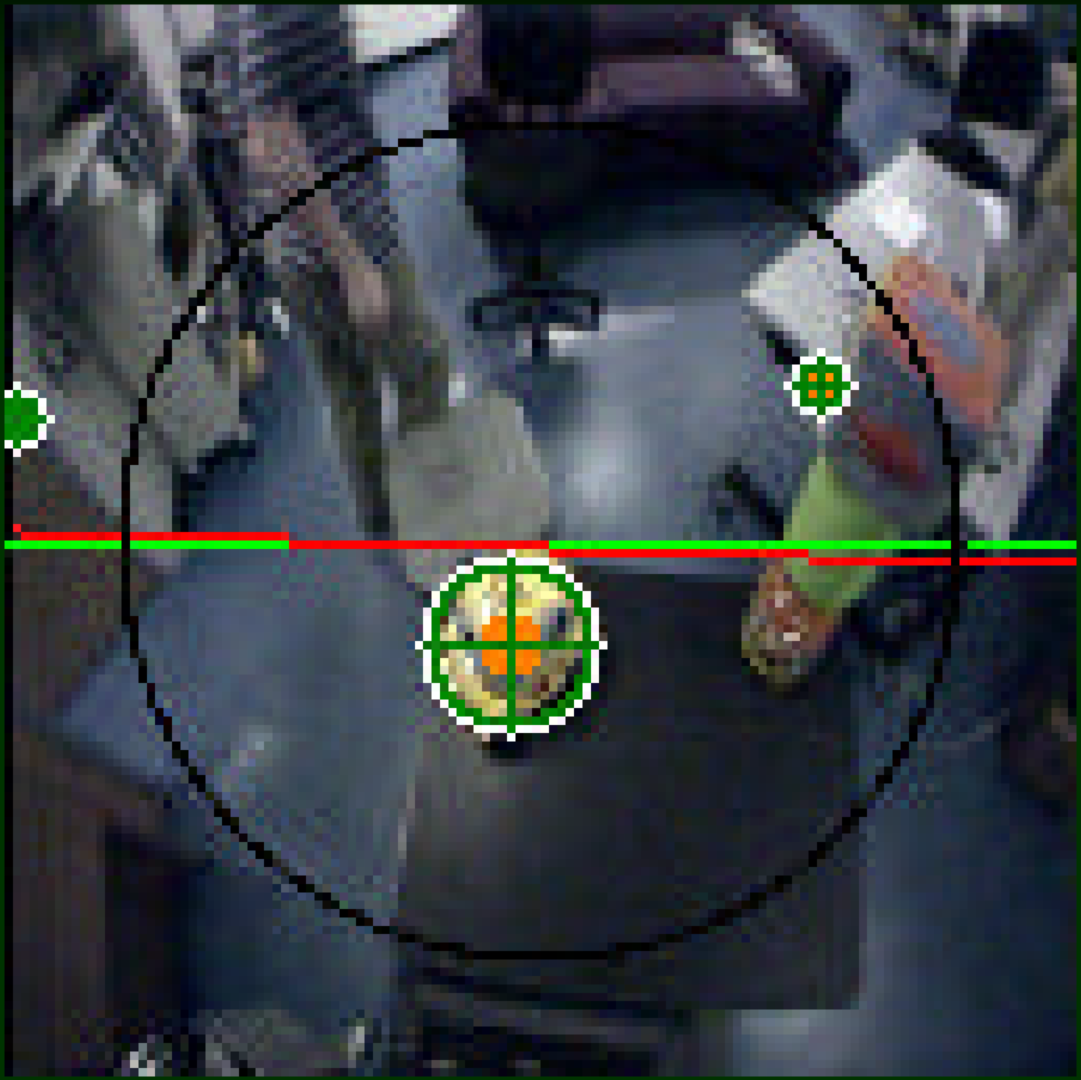
keeping track of objects



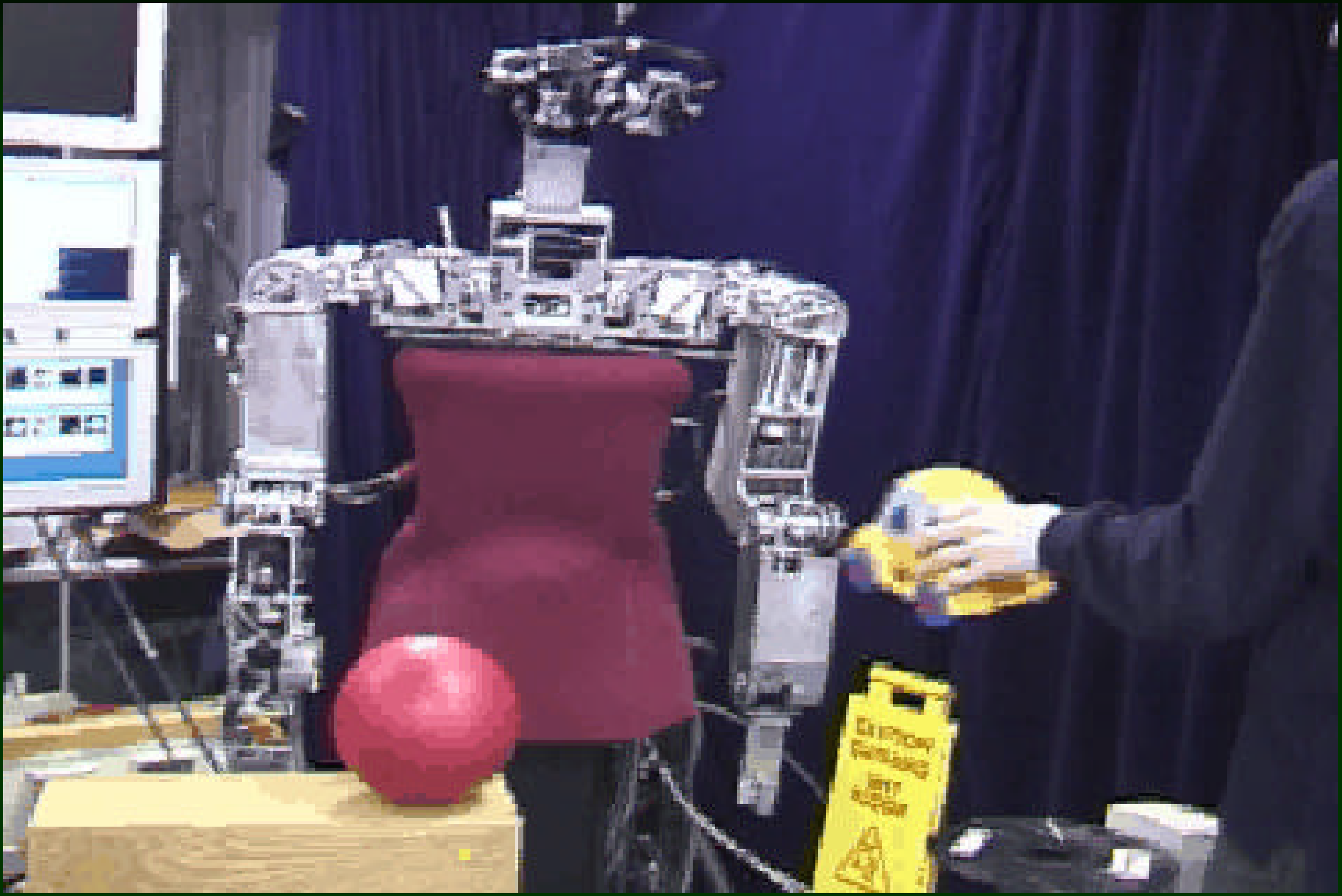
EgoMap

short term memory
of objects and their locations
so “out of sight” is not “out of mind”

keeping track of objects



speech and space: chatting



talk overview

Learning from an activity

- Poking: to learn to recognize objects, manipulators, etc.
- Chatting: to learn the names of objects

Learning a new activity

- Searching for an object
- Then back to learning *from* the activity…

Tomasello’s experiments

Designed experiments to challenge constraint-based theory of language acquisition in infants

Wants to show infants learn words through real understanding of activity (‘flow of interaction’), not hacks

Great test cases! Get beyond direct association

(But where does knowledge of activity come from?)

“let’s go find the toma!”

Infant plays with set of objects

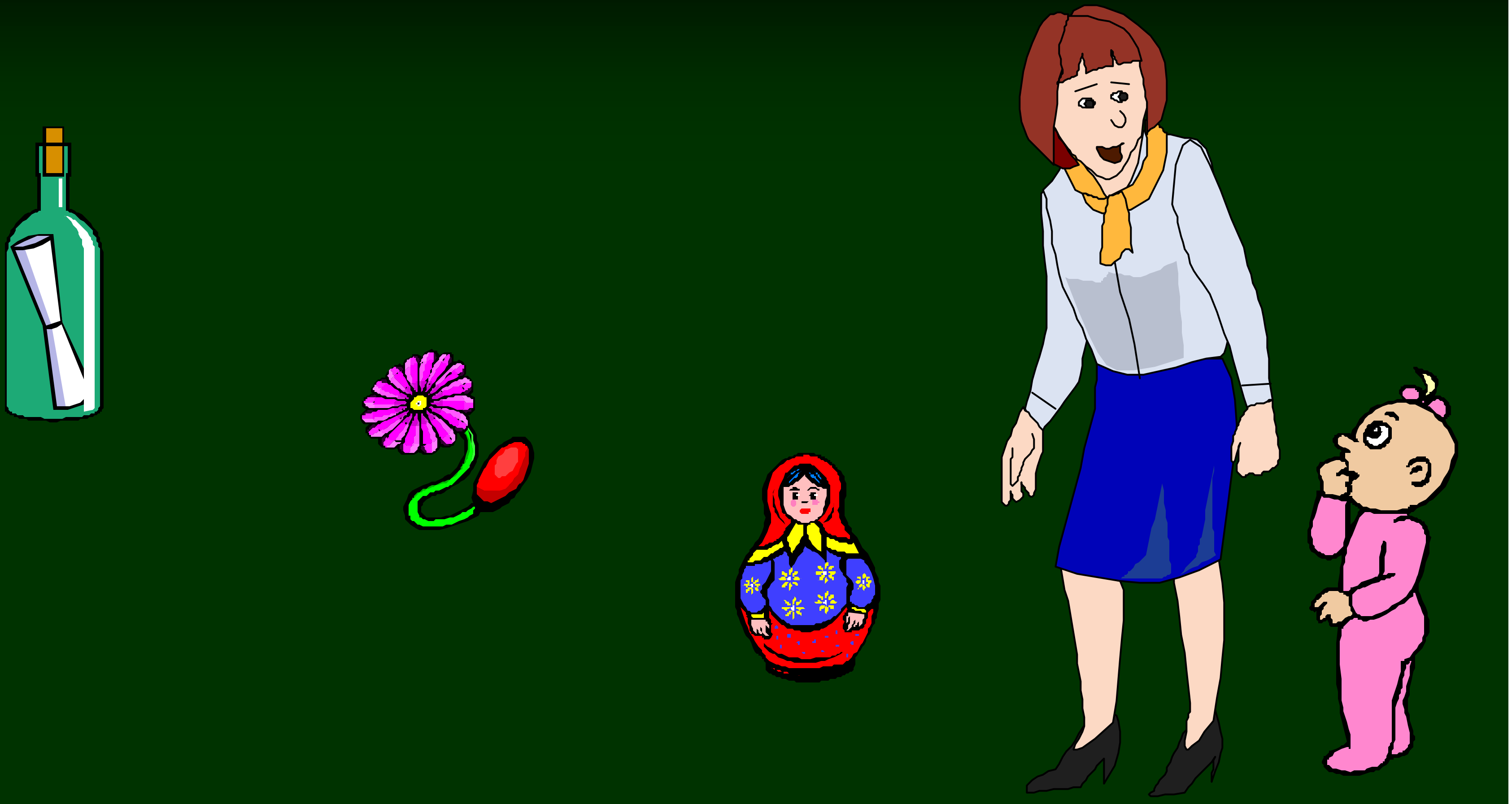
Then adult says “let’s go find the toma!” (nonce word)

Acts out a search, going to several objects first before finally finding the ‘toma’

Later, infant tested to see which object it thinks is the ‘toma’

Several variants (e.g. ‘toma’ placed in inaccessible location with the infant watching – adult is upset when trying to get it)

“let’s go find the toma!”

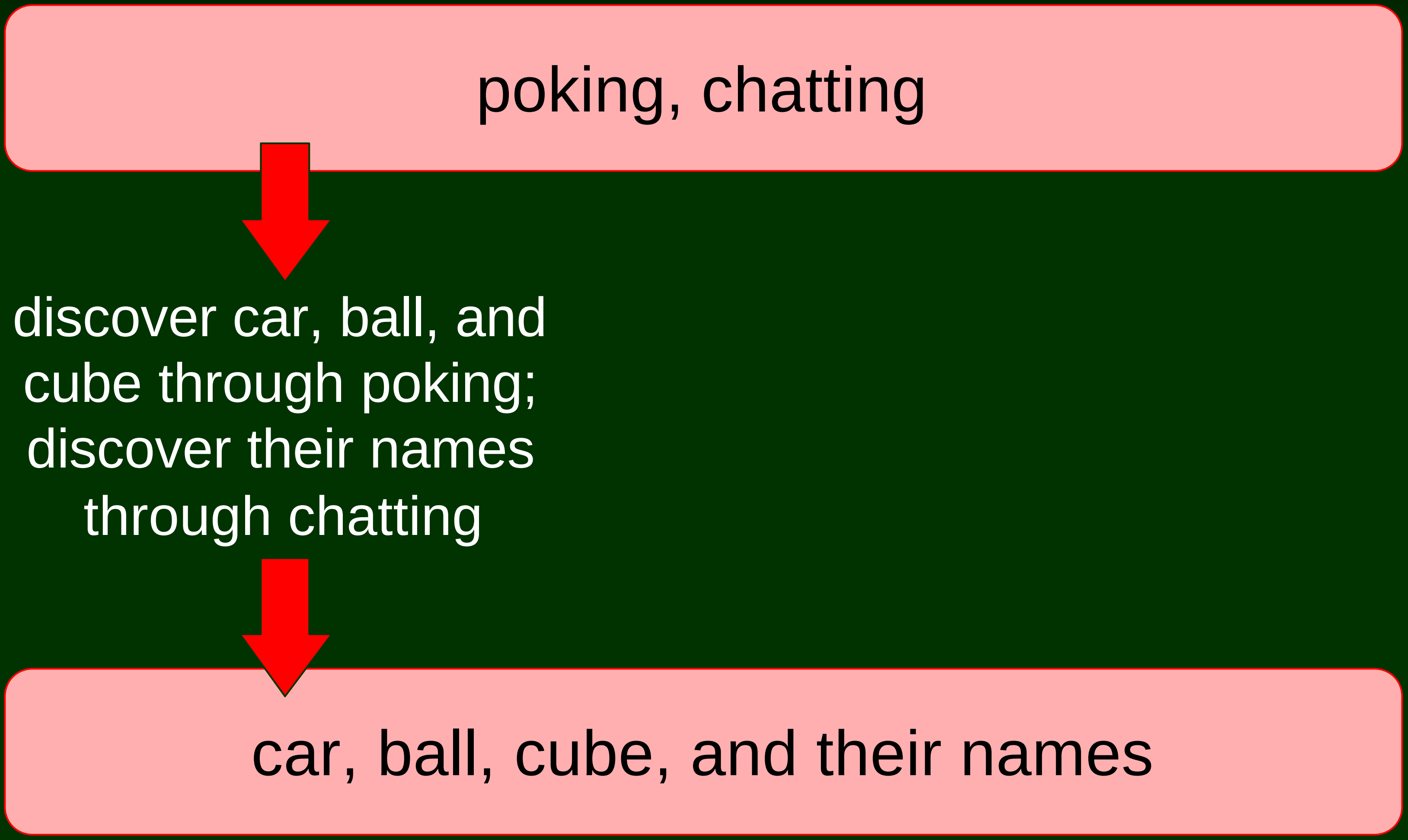


goal

Have robot learn about search activity from examples of looking for known objects

Then apply that to a “find the toma”-like scenario

virtuous circle

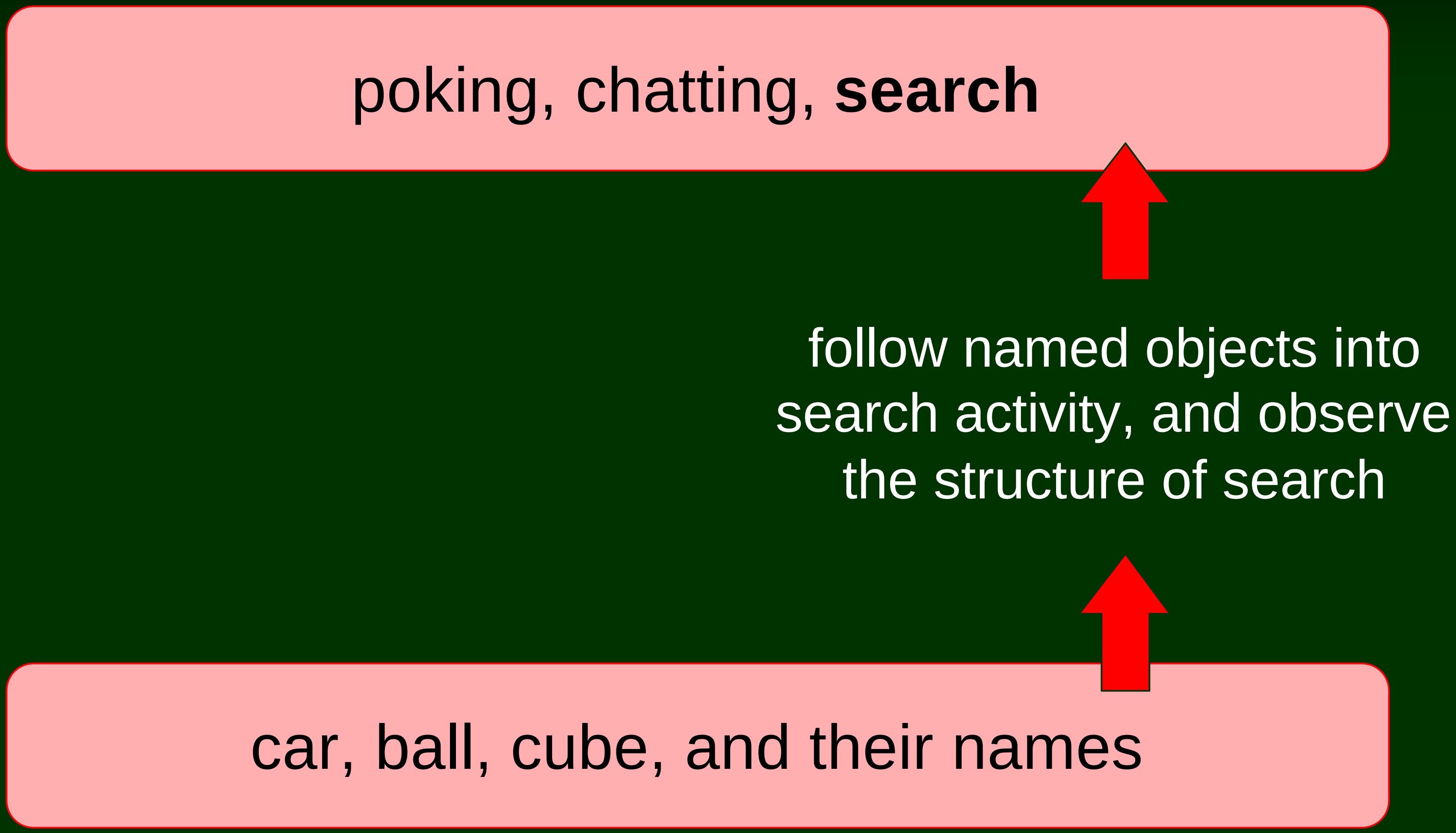


poking, chatting

discover car, ball, and cube through poking; discover their names through chatting

car, ball, cube, and their names

virtuous circle

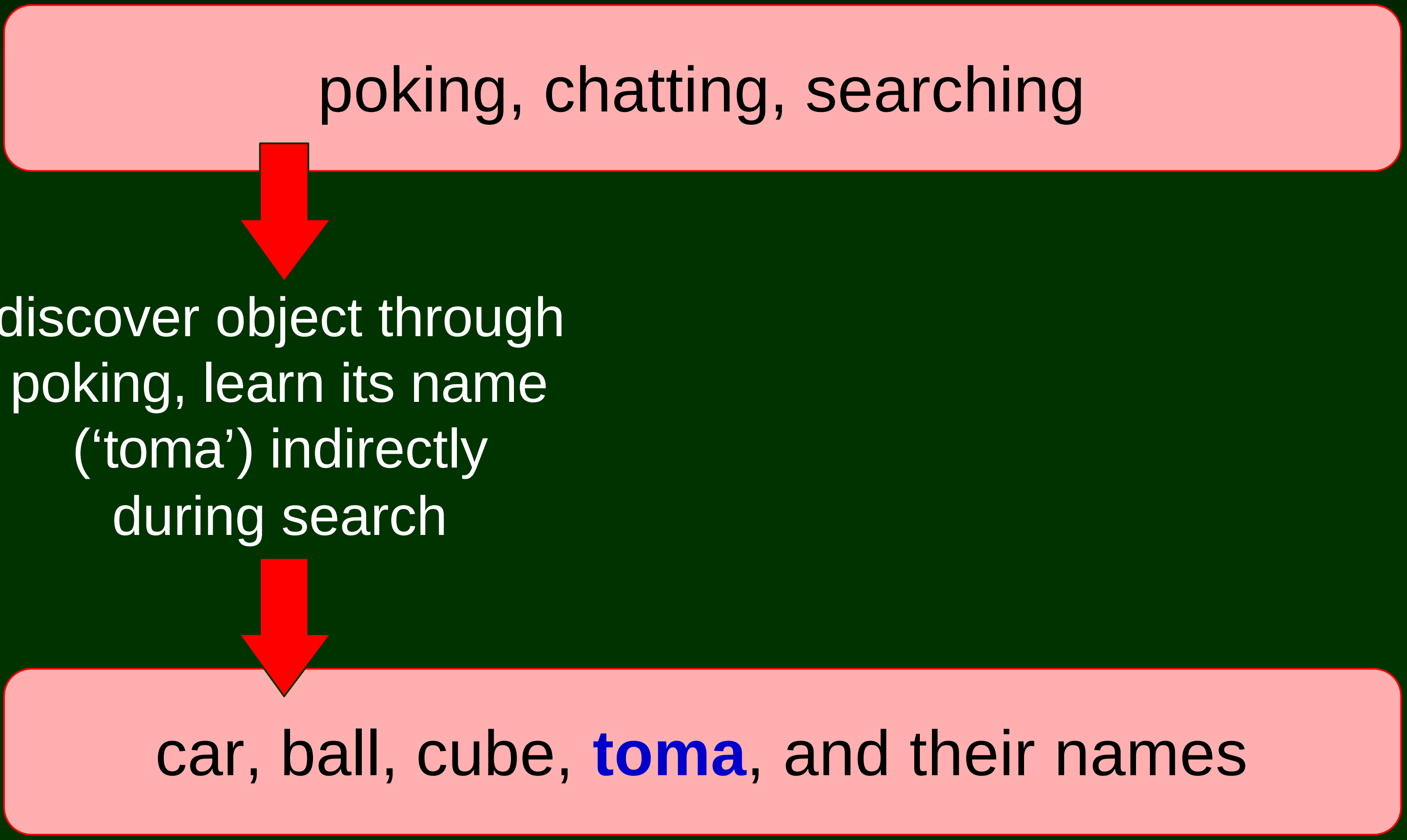


poking, chatting, **search**

follow named objects into
search activity, and observe
the structure of search

car, ball, cube, and their names

virtuous circle

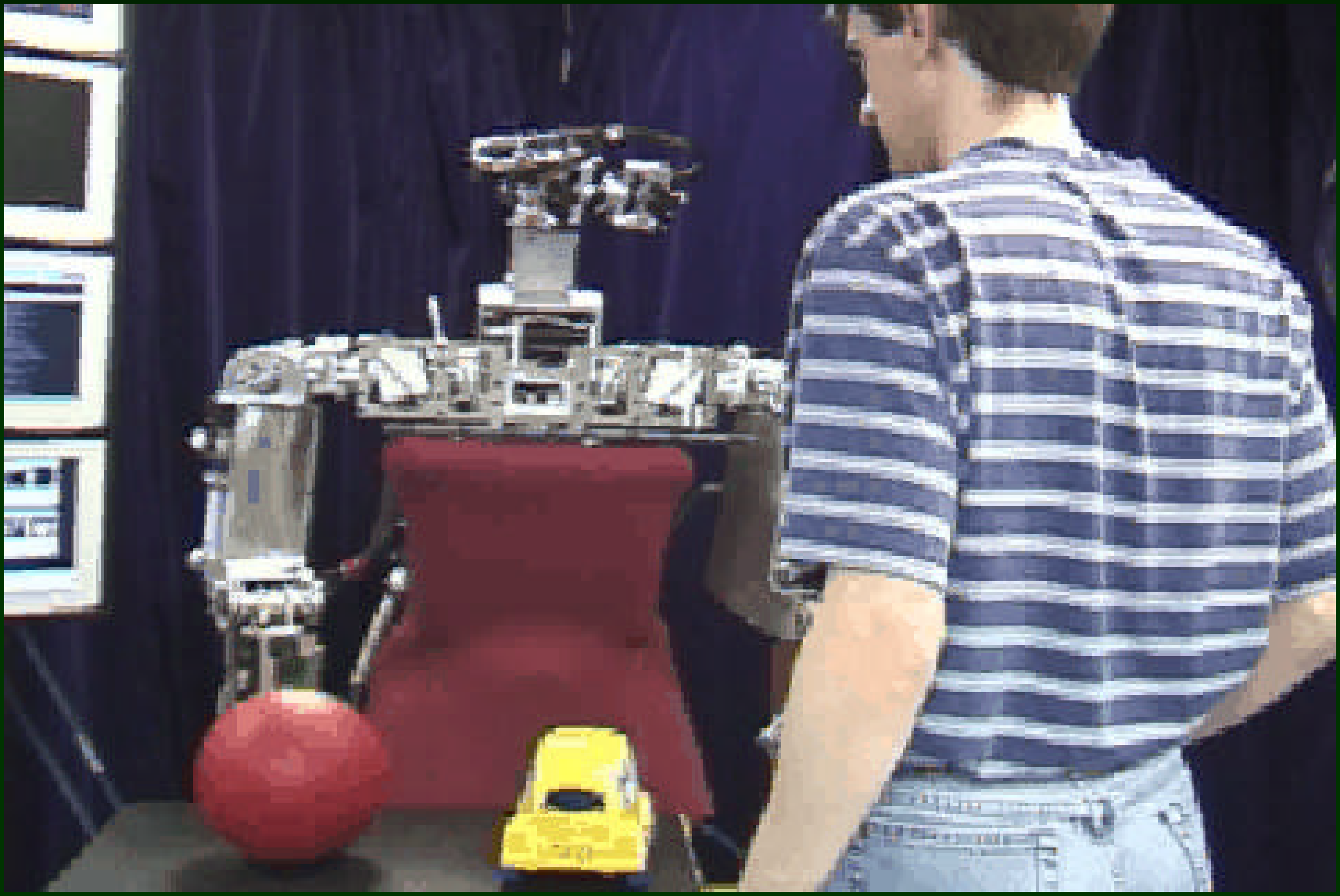


poking, chatting, searching

discover object through poking, learn its name (‘toma’) indirectly during search

car, ball, cube, **toma**, and their names

learning about search



what the robot learns

‘Find’ is followed by mention of an absent object

‘Yes’ is said when a previously absent object is in view

how it learns this

Look for reliable event/state combinations, sequences

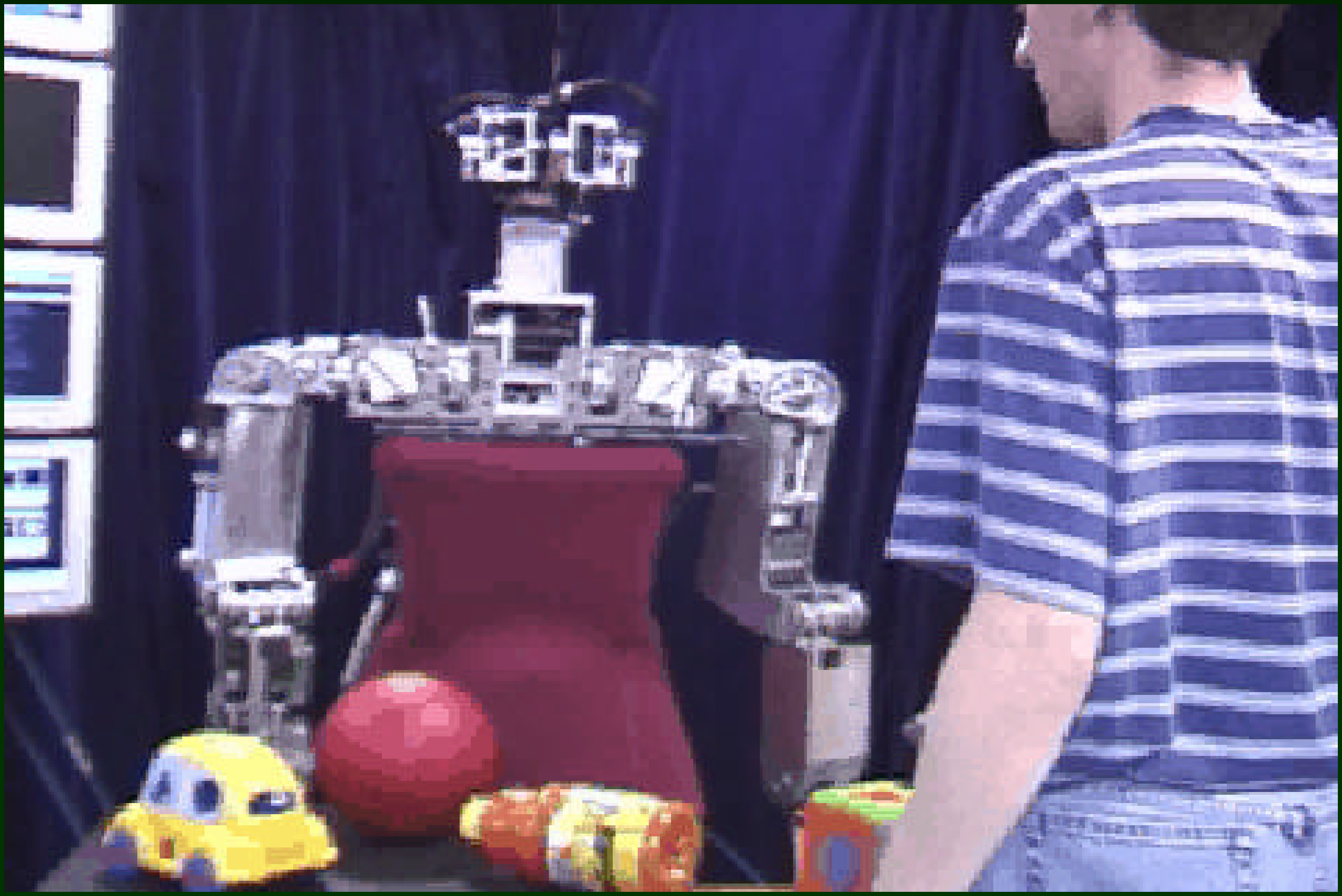
Events are:

- hearing a word
- seeing an object

States are:

- recent events
- situation evaluations (object corresponding to word not present, mismatch between word and object, etc.)

finding the toma



caveats

Much much less sophisticated than infants!

Cues the robot is sensitive to are very impoverished

Slightly different from Tomasello’s experiment

Saved state between stages – wasn’t one complete continuous run

conclusions: why do this?

Uses all the ‘alternative essences of intelligence’

- Development
- Social interaction
- Embodiment
- Integration

Points the way to really flexible robots

- today the robot should sort widgets from wombats (neither of which it has seen before)
- who knows what it will have to do tomorrow

conclusions: contributions

active segmentation	through contact
appearance catalog	for oriented features
open object recognition	for correction, enrollment
affordance recognition	for rolling
open speech recognition	for isolated words
virtuous circle of development	learning about and through activity

conclusions: the future

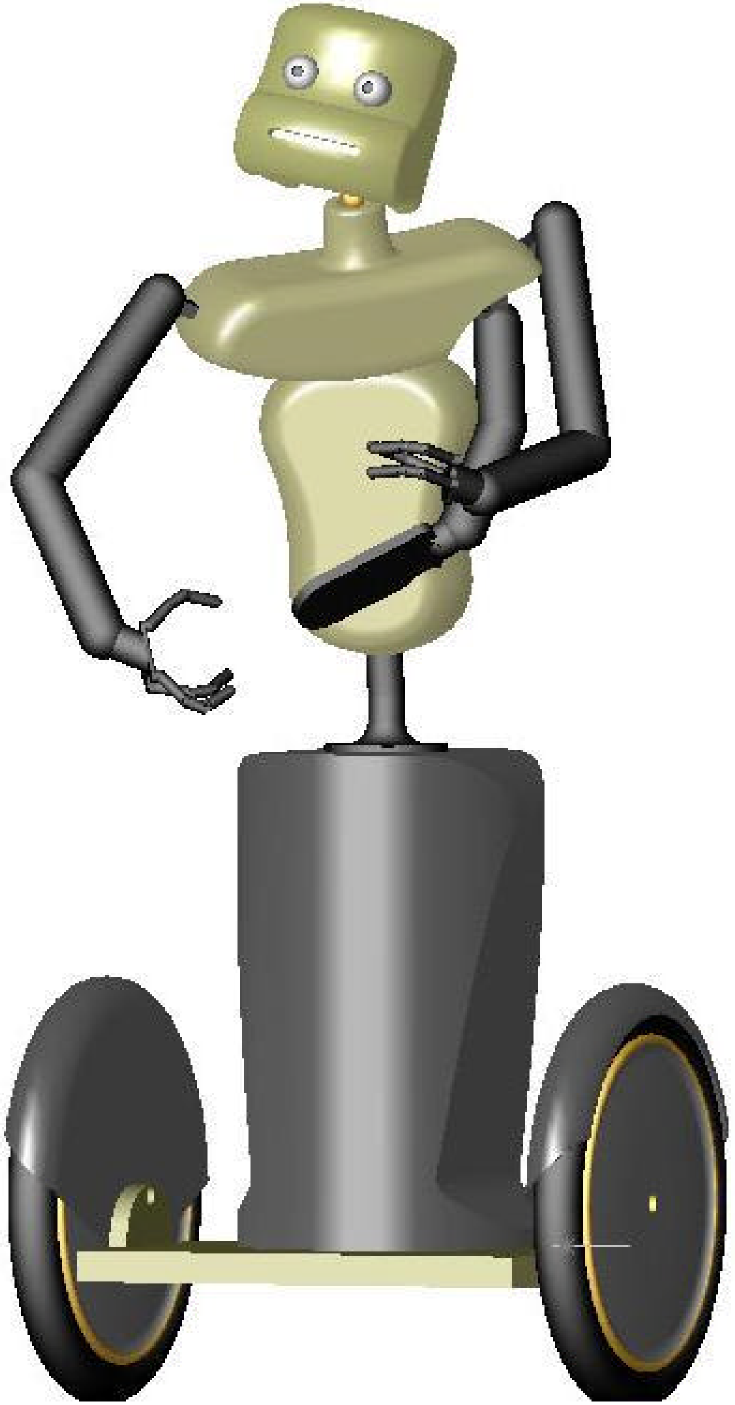
Dexterous manipulation

Object perception (visual, tactile, acoustic)

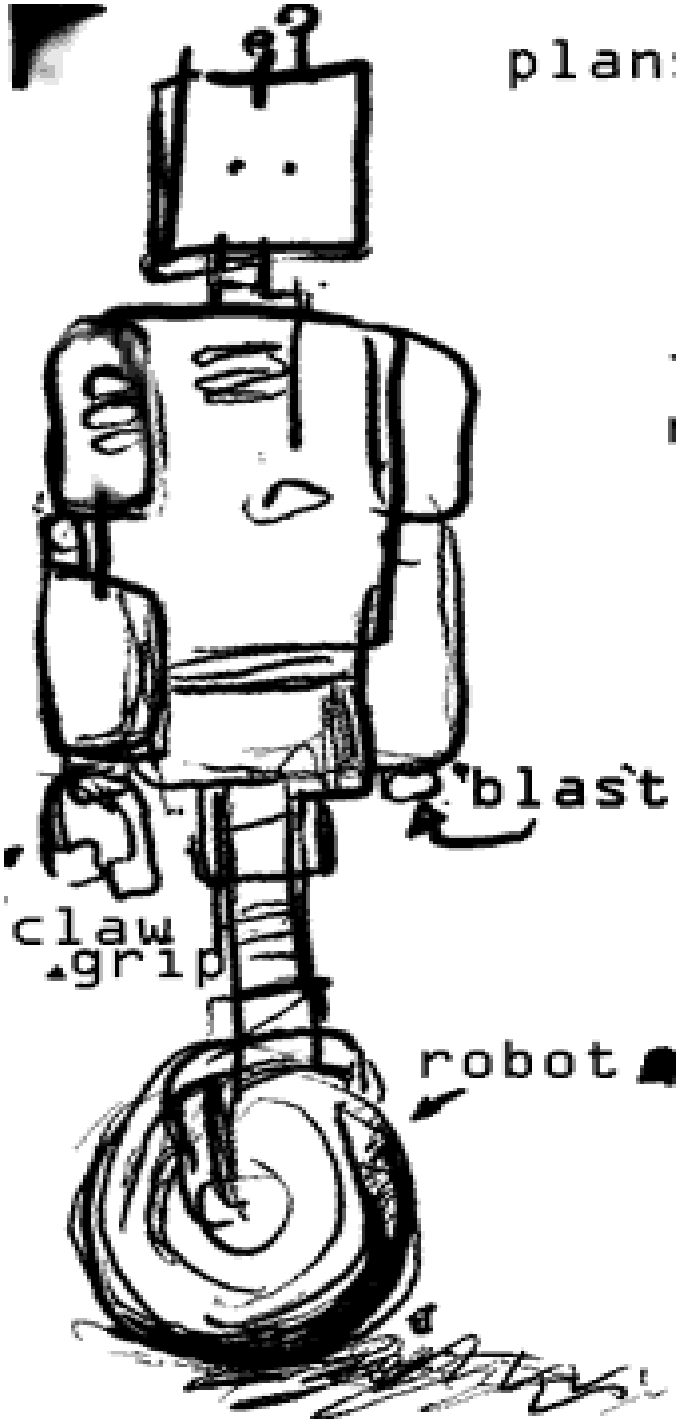
- During dextrous manipulation
- During failed manipulation

Integration with useful platform

- Socially enabled
- Mobile



plans for building ~~ka~~ killer robot



first you will need ~~tok~~ the robot parts[1]

robot brain
blastocannon v150
germanium diodes (3)
robot eyes (2)
robot shell incl.
wheel, claw grip

connect the parts as shown in the diagram

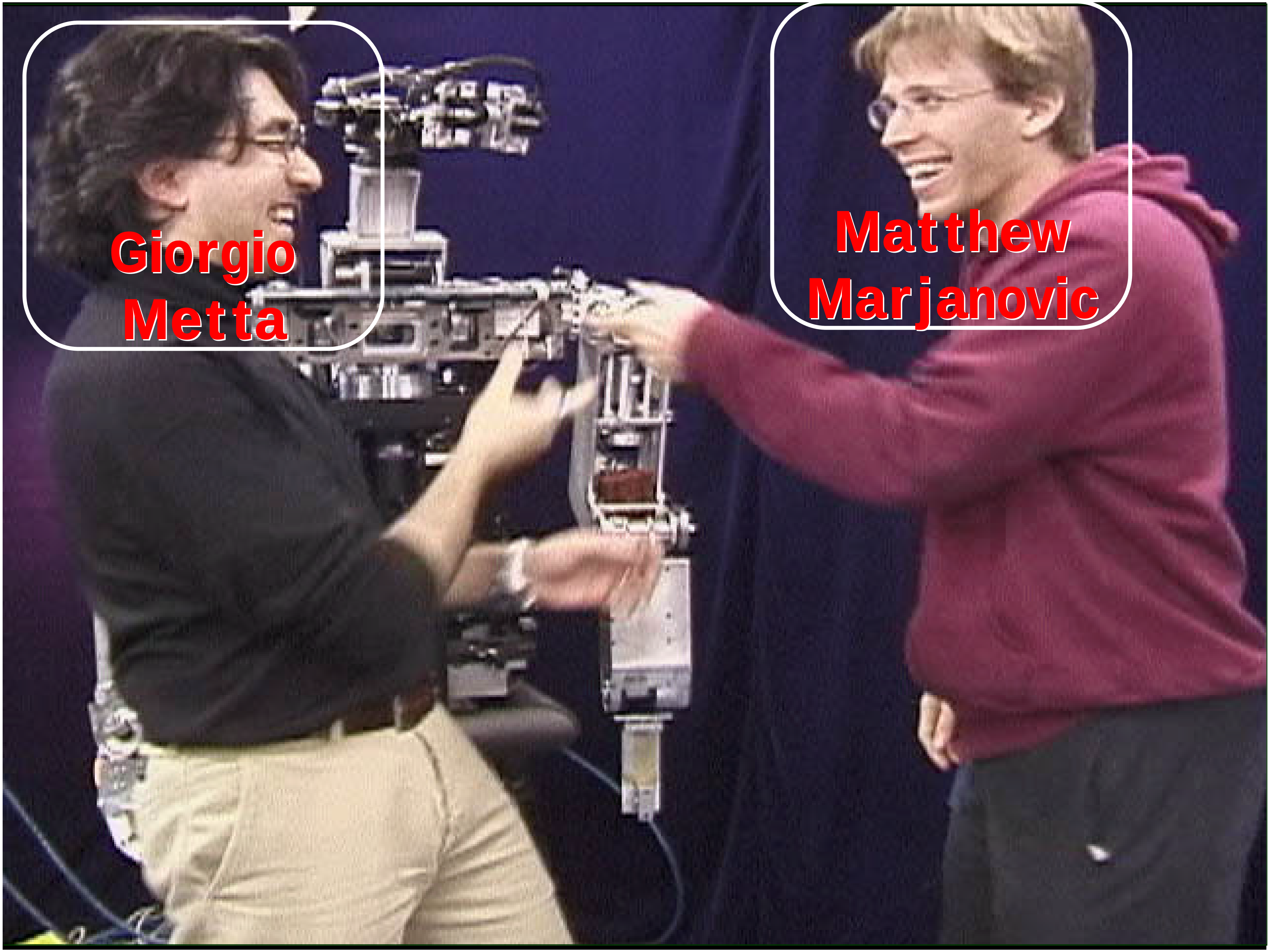
give voice commands

GO KILL STOP

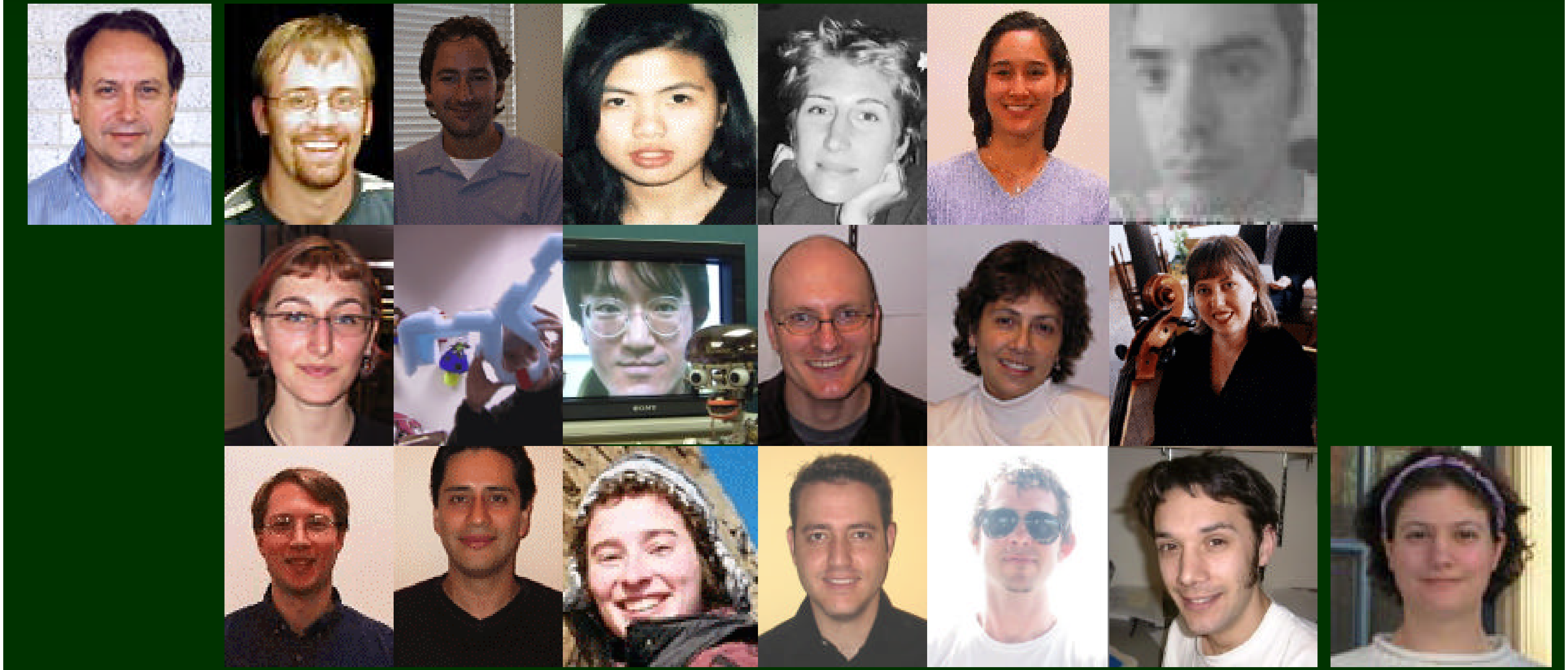
to control your robot

[1] available
~~available~~ at robot shack

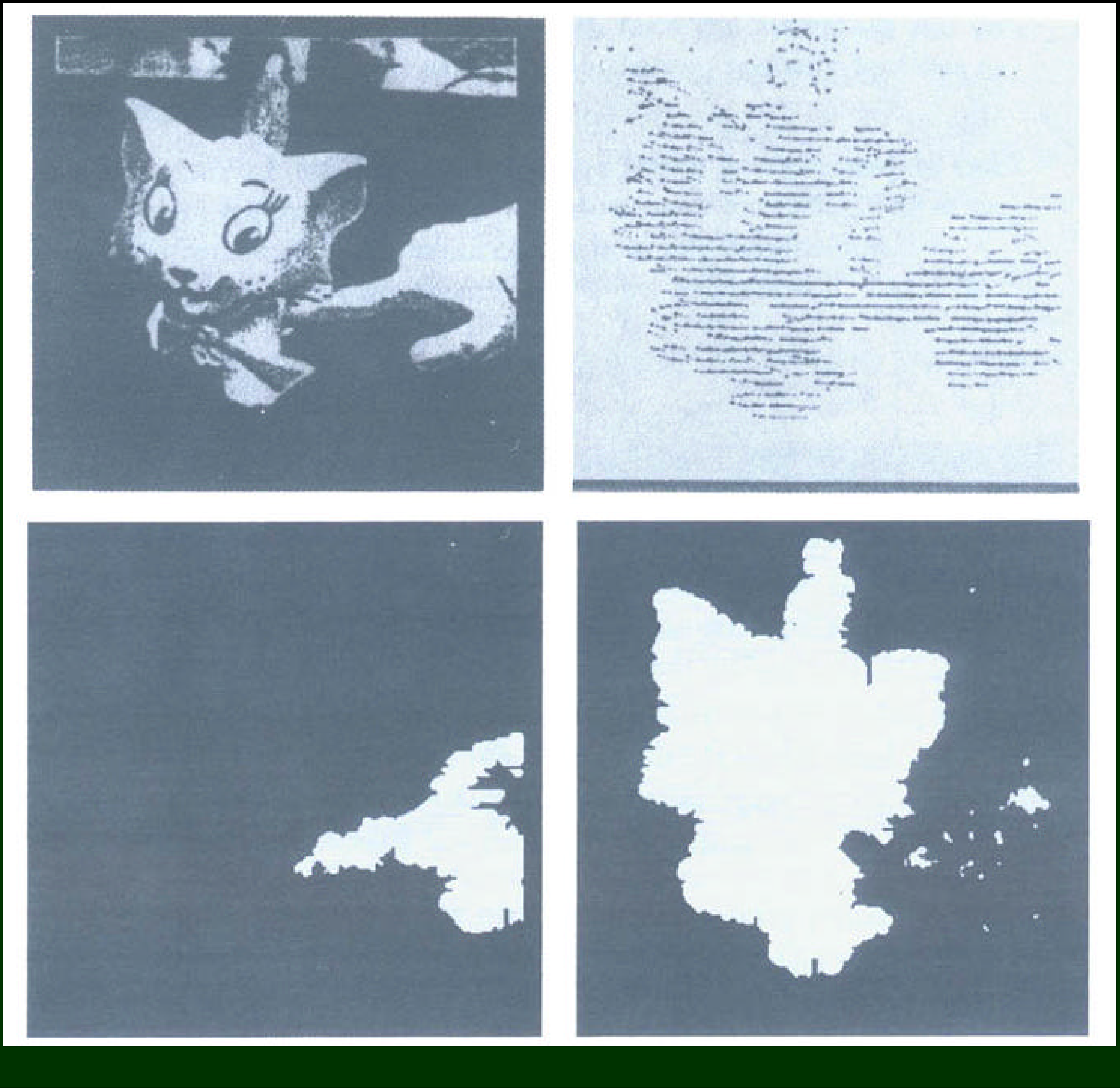
Tom Murphy (www.cs.cmu.edu/~tom7)


Giorgio Metta
Matthew Marjanovic

thanks!



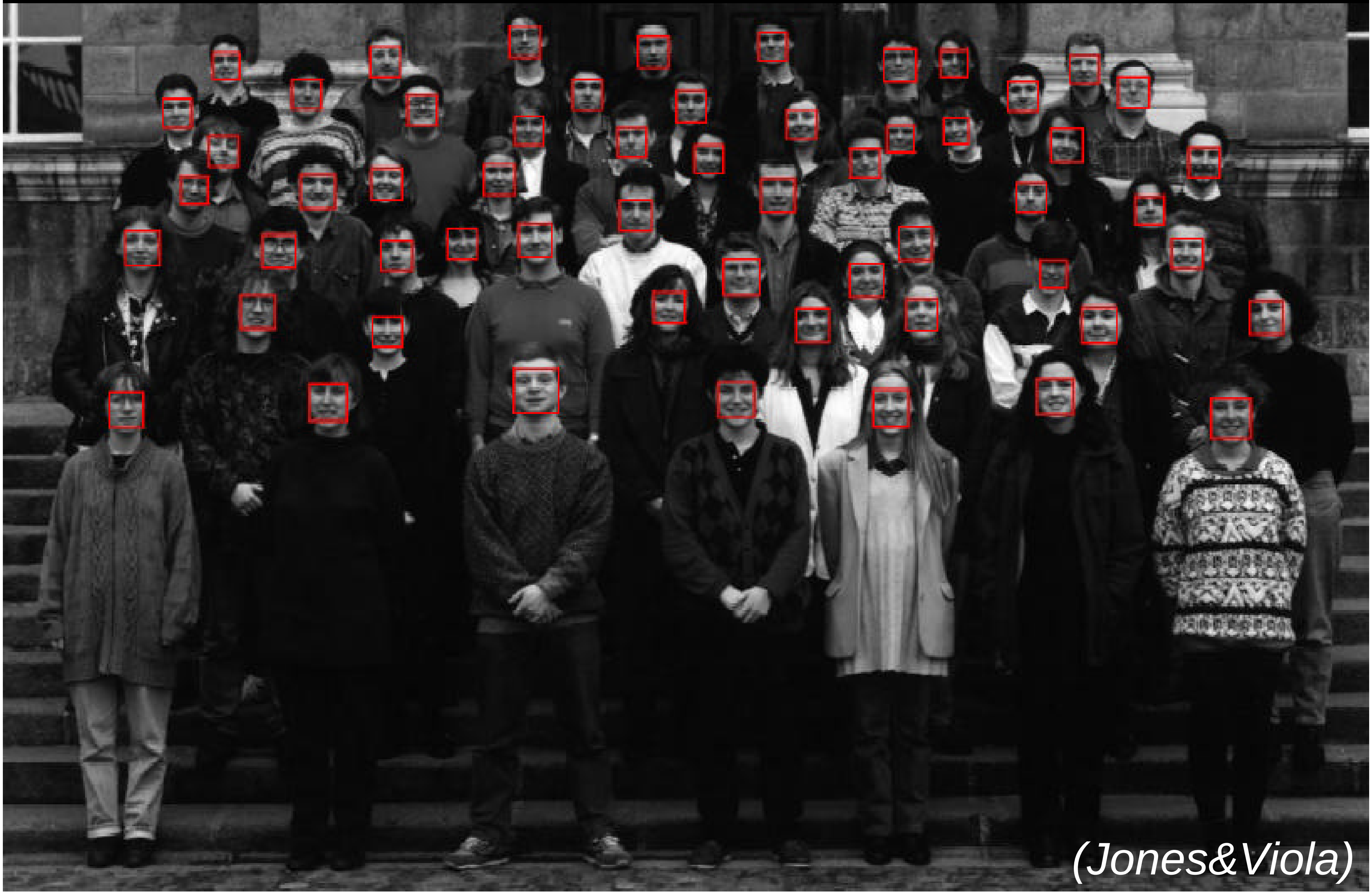
Sandini et al, 1993



machine perception

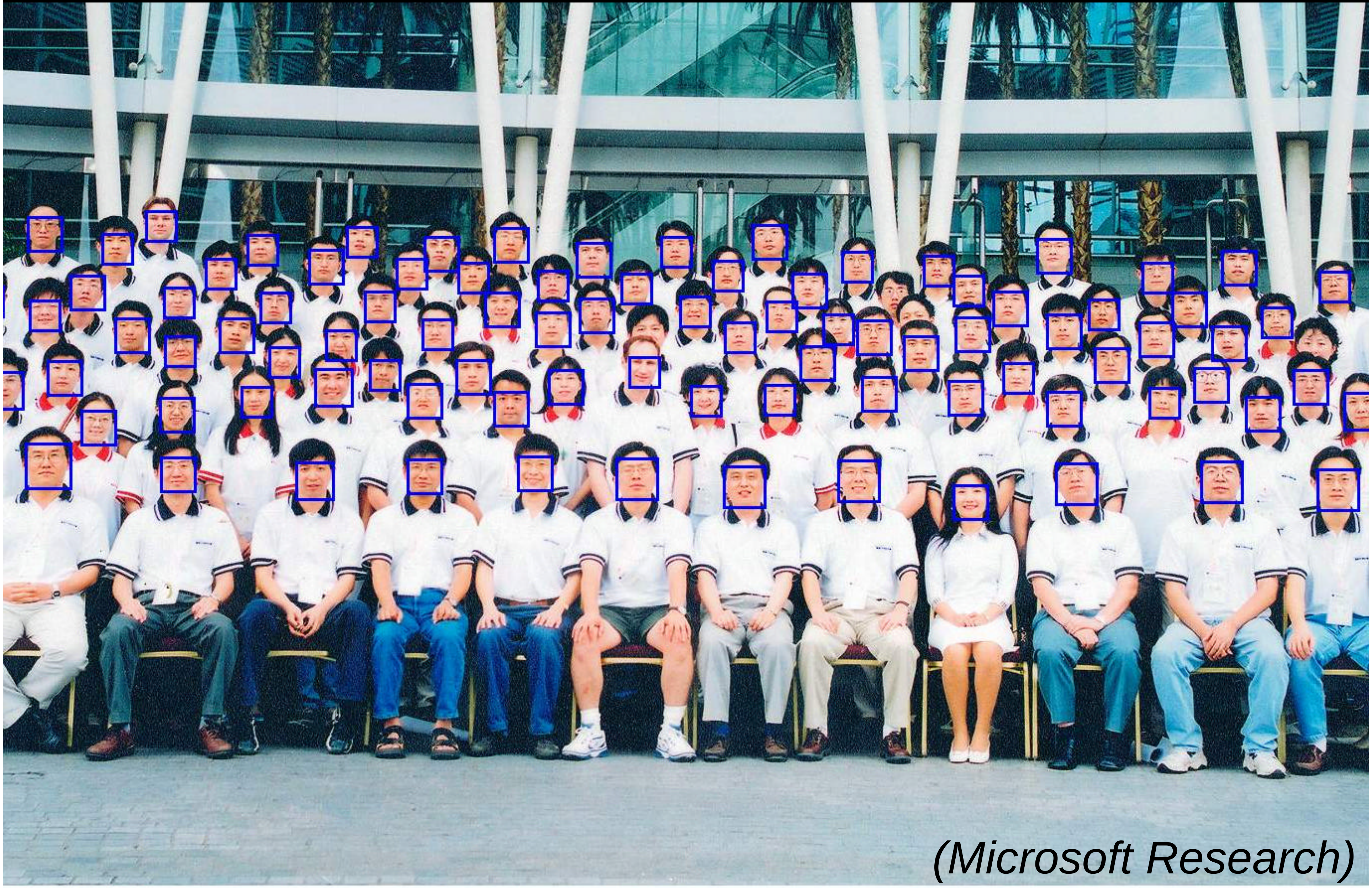
algorithms + **training data** = robust perception

face detection



(Jones&Viola)

face detection

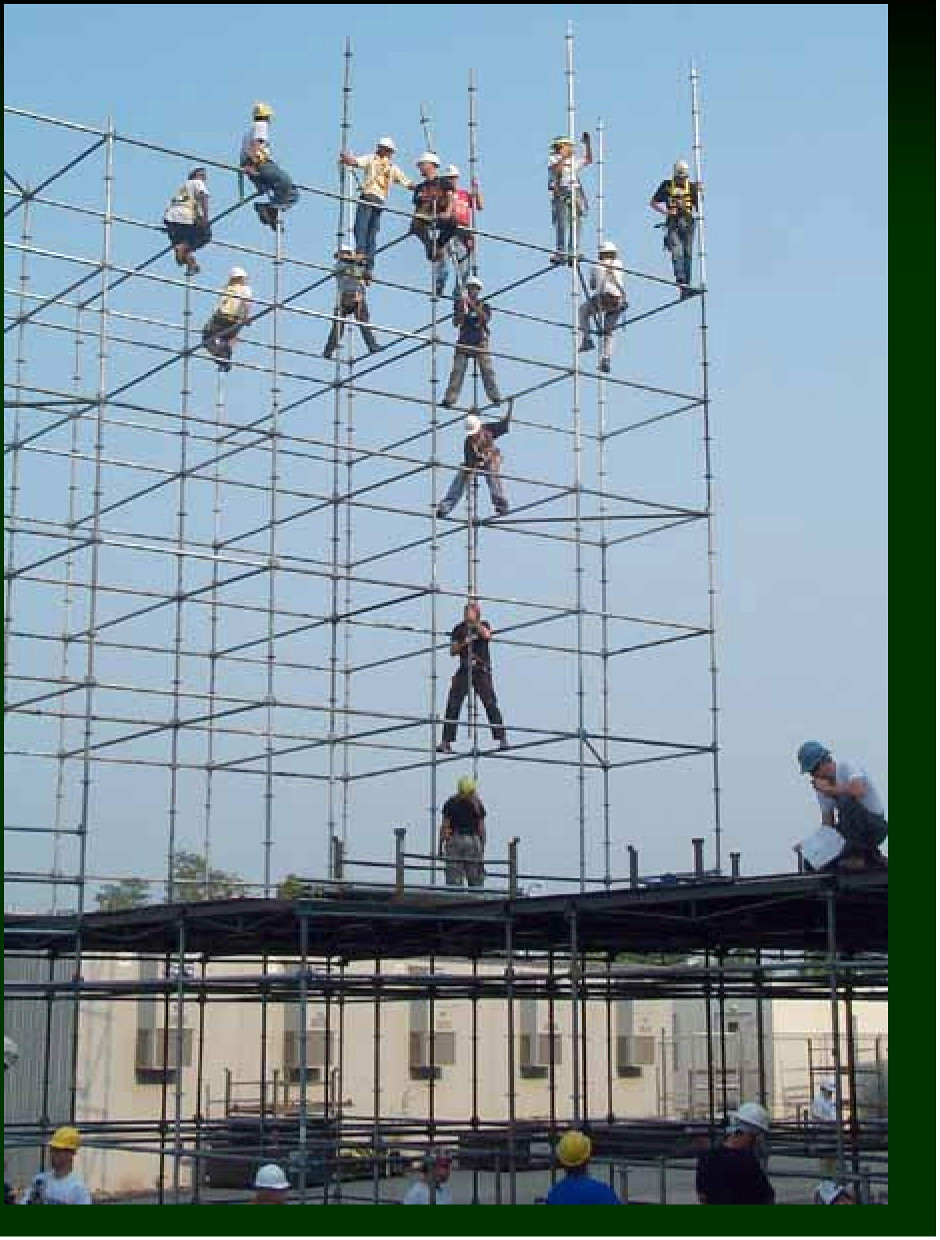


(Microsoft Research)

stable problem

Build once, use many times

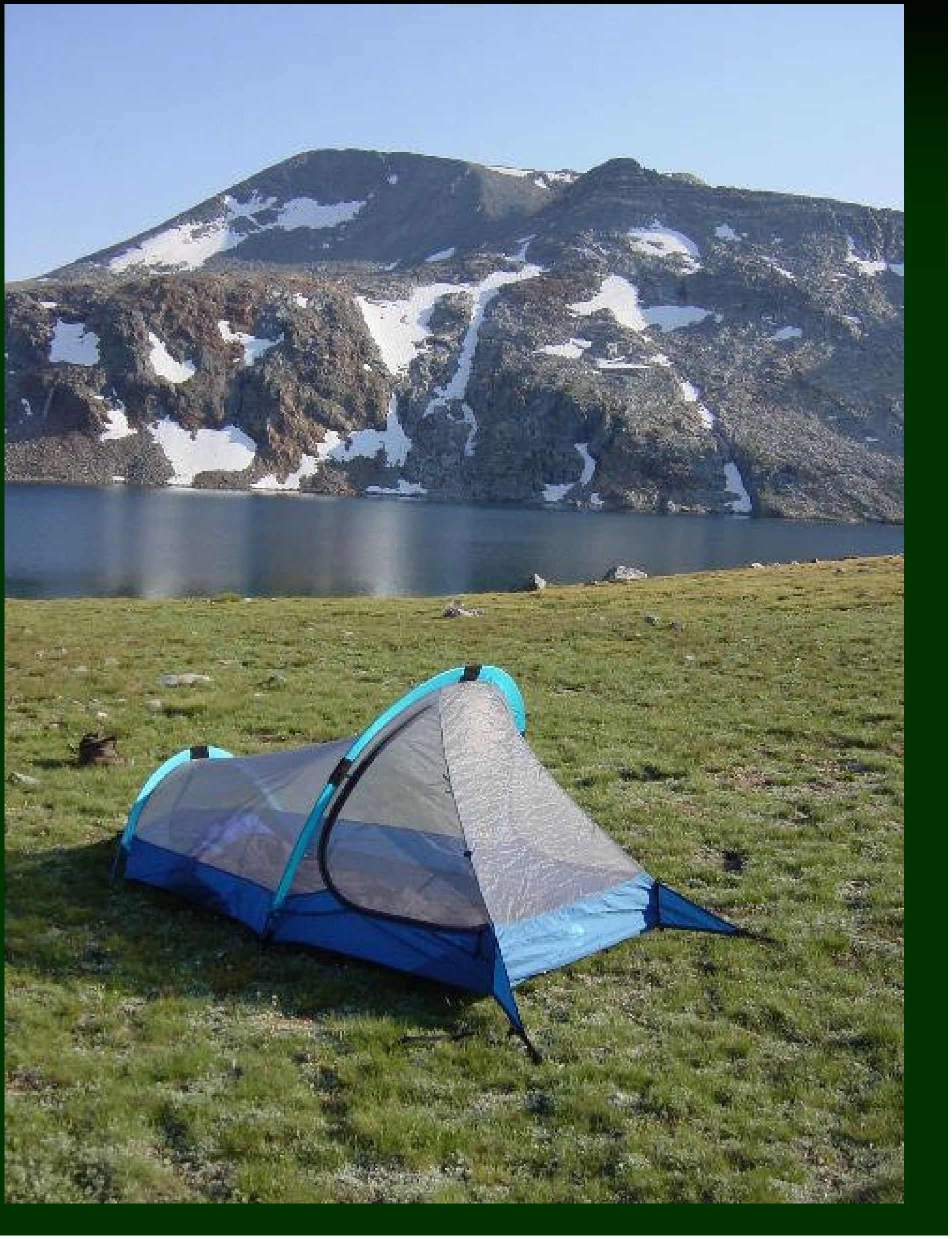
Can use and then remove scaffolding



unstable problem

Frequent changes

Better if scaffolding remains part of the structure



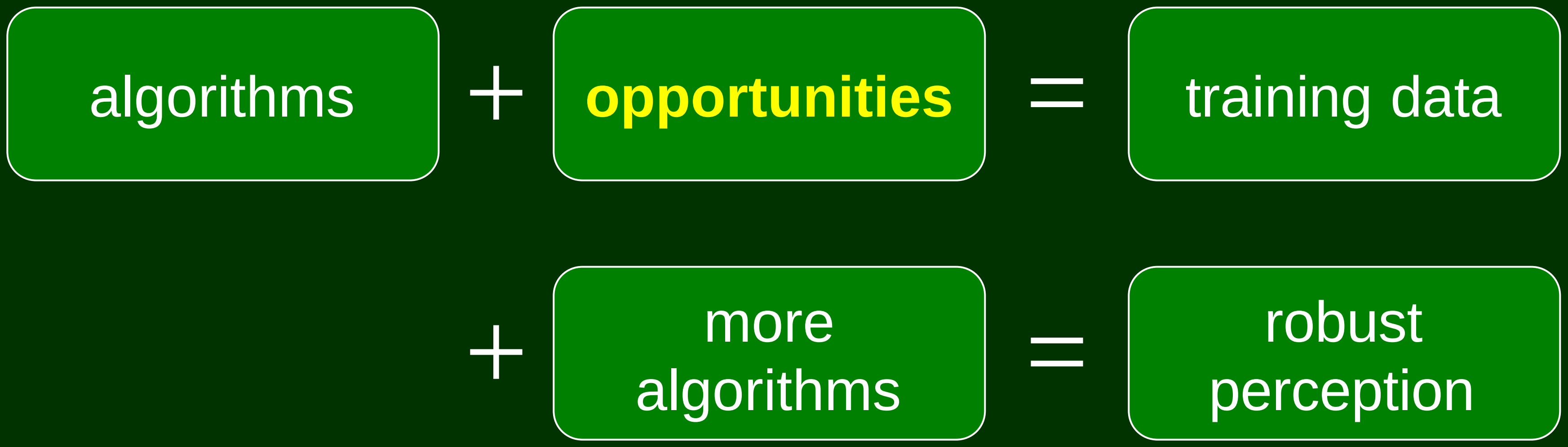
object recognition

(COIL 100)

machine perception

algorithms + **training data** = robust perception

self-training perceptual system



algorithms + **opportunities** = training data

+ more algorithms = robust perception