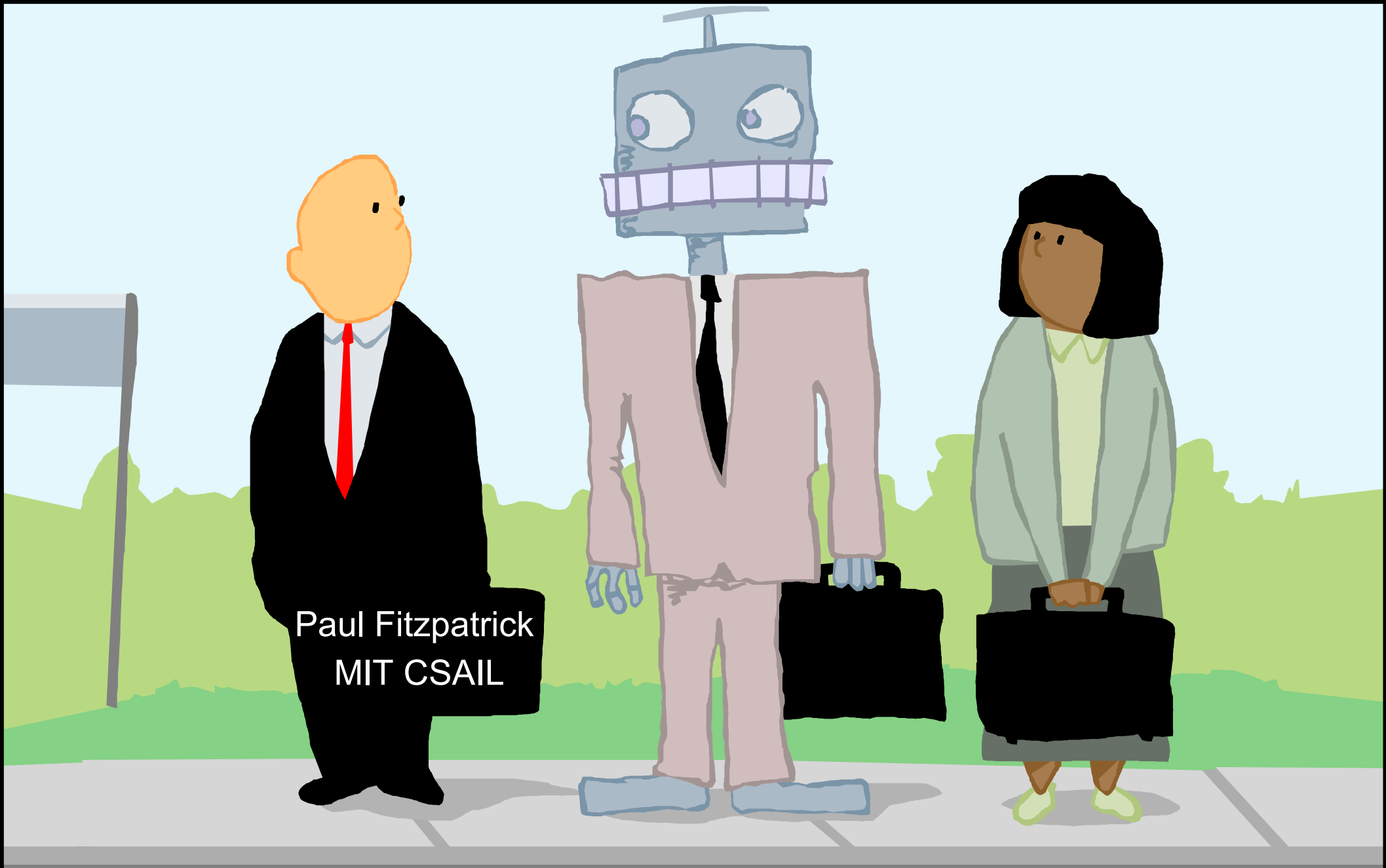


The DayOne Project

Building a Truly Flexible Humanoid

Day One

non-robotic robotics

- We need robots that are as flexible perceptually as they are becoming mechanically
- We're in luck! – mechanically flexible robots are uniquely well suited to flexible machine perception
- First step: create a class of robots that reverse the pejorative meaning of “robotic”
  - not dull, blinkered, scripted, endlessly-repeating, ...
  - instead opportunistic, meddlesome, persistent, ...
- DayOne: a step towards that first step

Motivation

Opportunism
Meddling
Acrobatics

Conclusions

Motivation

Opportunism
Meddling
Acrobatics

Conclusions

Day One

mechanical flexibility

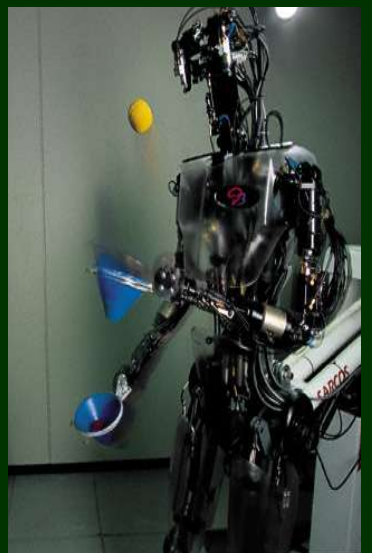
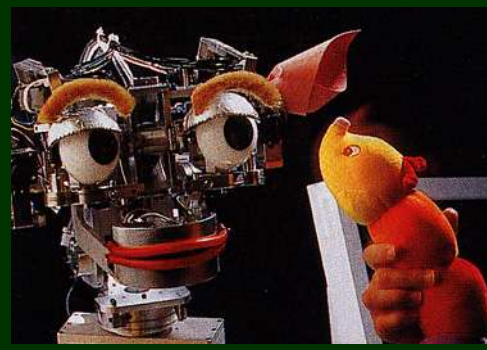
- Humanoid robots are improving mechanically by leaps and bounds
- Real progress, but some danger signs
  - Lots of synchronized dancing
  - More gesturing than grasping
  - More human interaction than object interaction

Day One

perceptual flexibility

- Machine perception for humanoid robots is often crude: bright objects, motion
- if not, it is generally imported from the computer vision and speech recognition communities
  - treats the robot's body as just an annoyingly noisy, unstable platform rather than an opportunity



Day One

the DayOne project

- **Ultimate goal**:
  - Robots with human-level perception
- **Intermediate goal**:
  - Maximize range of situations a robot can adapt to in one day
  - Inspired by ability of various “prey” species (e.g. ungulates) to rapidly adapt to their environment on their day of birth
  - A robot walking up a stairs is great mechanics; move/change the stairs to test perception

Day One

organizing principles

- **Be opportunistic**. Perception is sometimes easy. It is valuable to identify and exploit conditions that simplify perception, even if we can't rely on them entirely.
- **Be meddlesome**. Robots are not passive observers. They can shape their experience to their own advantage, and carry out experiments to resolve ambiguity.
- **Be acrobatic**. Information acquired opportunistically in one context can be used to learn and track properties across to other contexts, like a trapeze act.

Motivation

Opportunism

Meddling

Acrobatics

Conclusions

Day One

opportunistic perception

- Take advantage of occasional events or sporadic conditions
  - Complements “always-on” sensing
  - For example, depth perception using cast shadows versus stereo
- Why bother?
  - In rich, real environments, opportunities abound
  - No such thing as true “always-on” sensing anyway
  - Grist for learning, and robot can create the opportunities (next section)

Day One

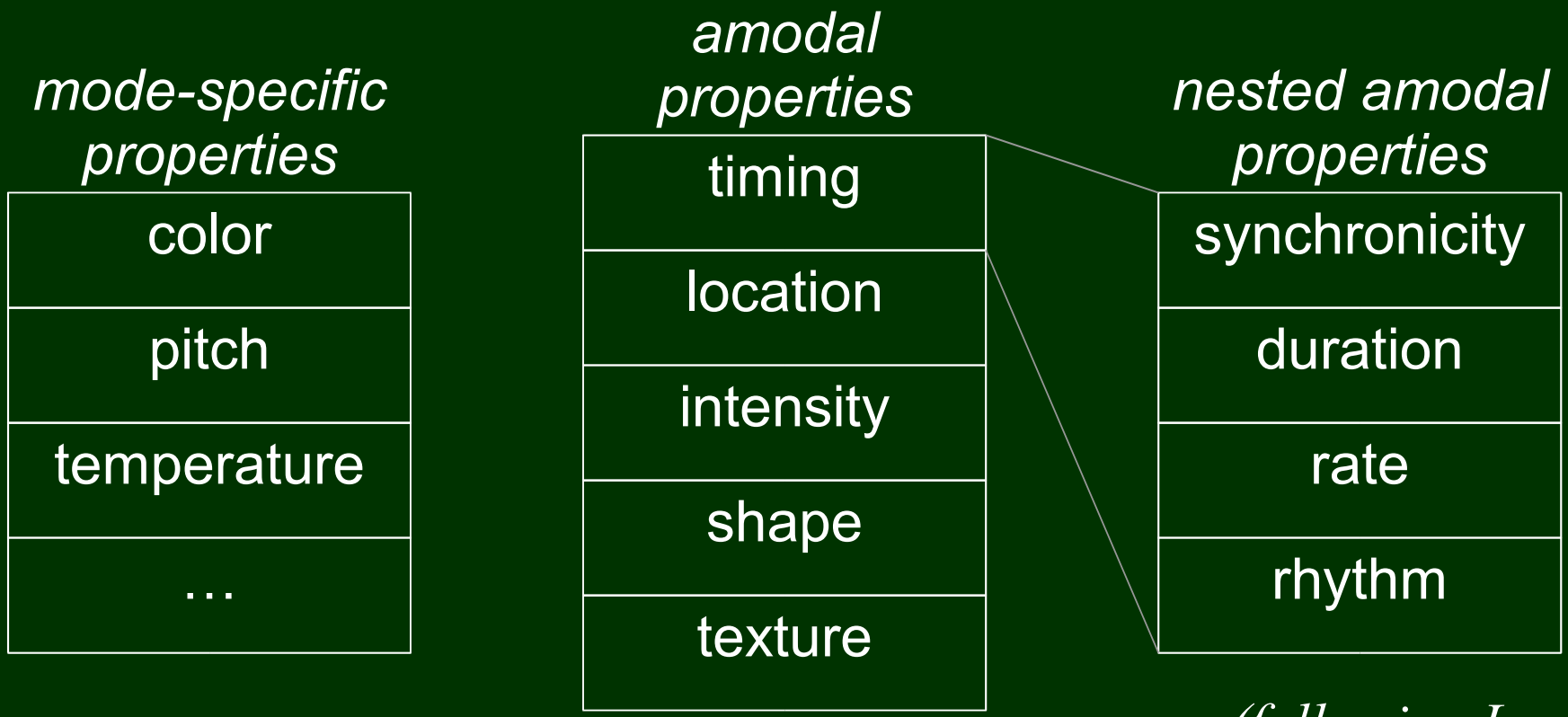
example 1: amodal cues

- An object/event is sensed in fragments
  - Different senses: vision, audition, touch, etc.
  - Different parts of the same sense: individual pixels, sound frequencies, locii of tactile stimulation, etc.
- Generally hard to pull this all together again
- But sometimes it is easy!
  - There are *amodal* cues that cross the senses, branding diverse signals as having a common origin
  - When present, they really simplify grouping

Day One

time

- Time is a basic property that gets encoded in all senses but is unique to none of them

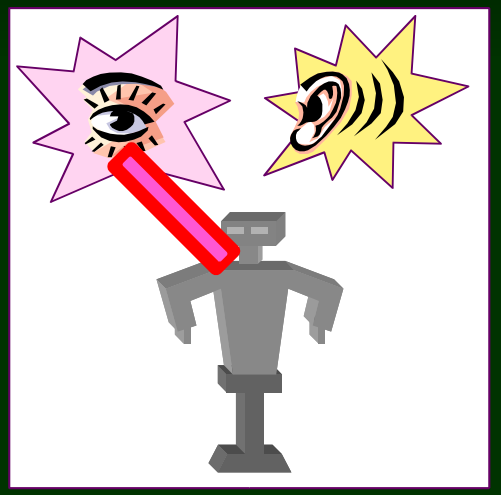


(following Lewkowicz)

Day One

rate matching

- Group compatible repeating signals
  - Check for equal rates, or multiples
  - Repetition gives redundancy, phase information



- Real-time implementation
  - Applied to sound, vision, and prioprioception
  - Repetitive events involving any combination of these three senses are detected automatically
  - Used to train recognizers that work *without* repetition

Day One

how accurate can this be?

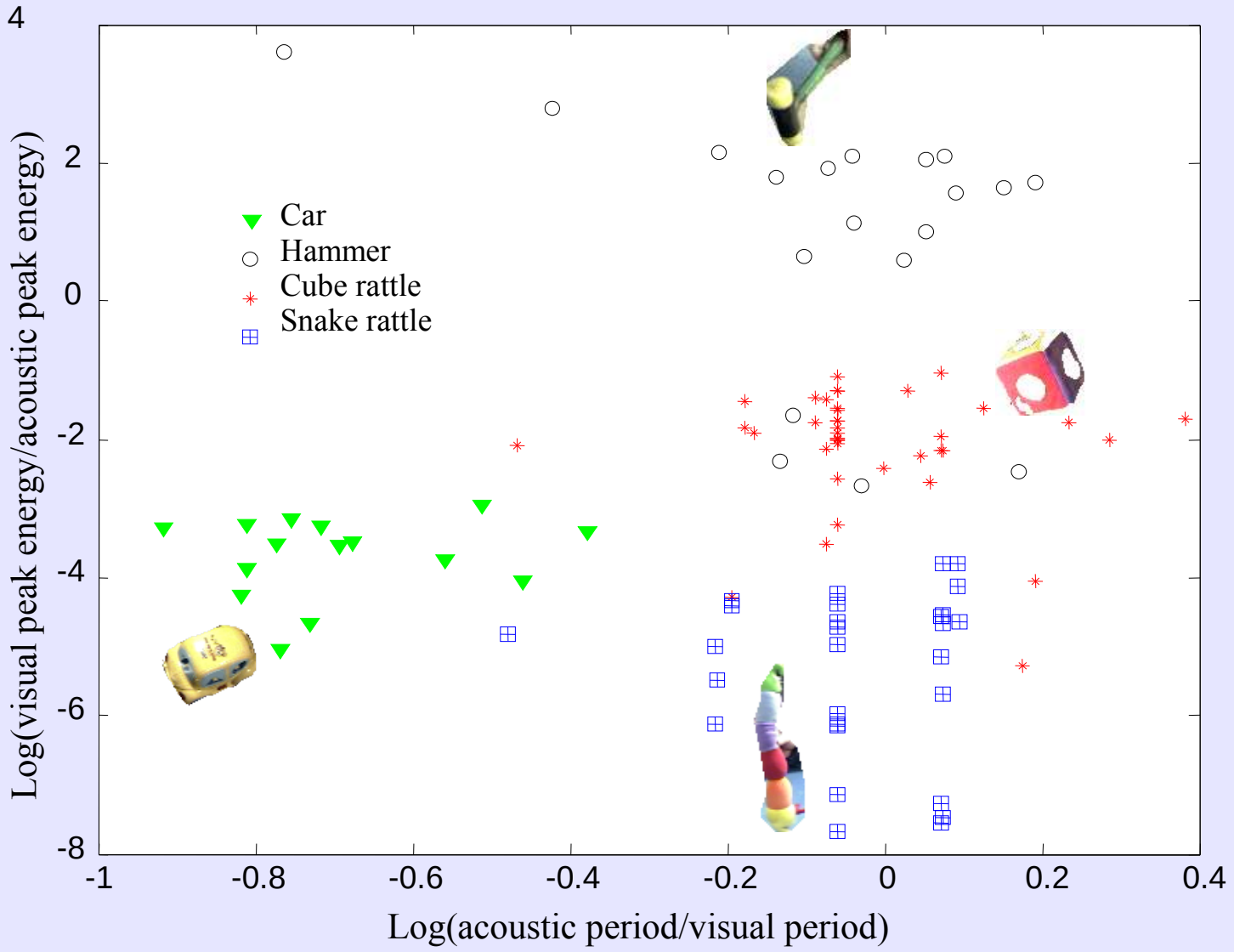
- Two moving objects
  - One noisy object (a plane)
  - One silent object (a mouse)
- How to link sound to right object?
  - Easy if different rates, but here they are almost equal (up to a factor of 2)
  - Could try to physically interpret sound and relate to vision
  - Drift reveals all – sound stays nailed to visual trajectory of plane, drifts slowly for mouse

Day One

recognition

- Robot learns to recognize what it detects this way
  - Appearance-based model for visual recognition
  - Eigensound approach for auditory recognition
- (Visual) recognition doesn't need further repetition
- When repetition is present:
  - Extra cues are available from *cross*-modal relations
  - E.g. plane makes noise at velocity extremes (two per visual period), hammer bangs at one extreme of position, bell tends to clang at both extremes of position

Day One recognition without sound or vision



Cross-modal recognition rate: 82%

Day One

what's the point?

- Amodal cues are low-hanging fruit
- Opportunistically establish links between the senses
- Kick-start modal perception
- Good match with human showing behavior
- *Not* exploiting these cues will be unforgivable in future robots

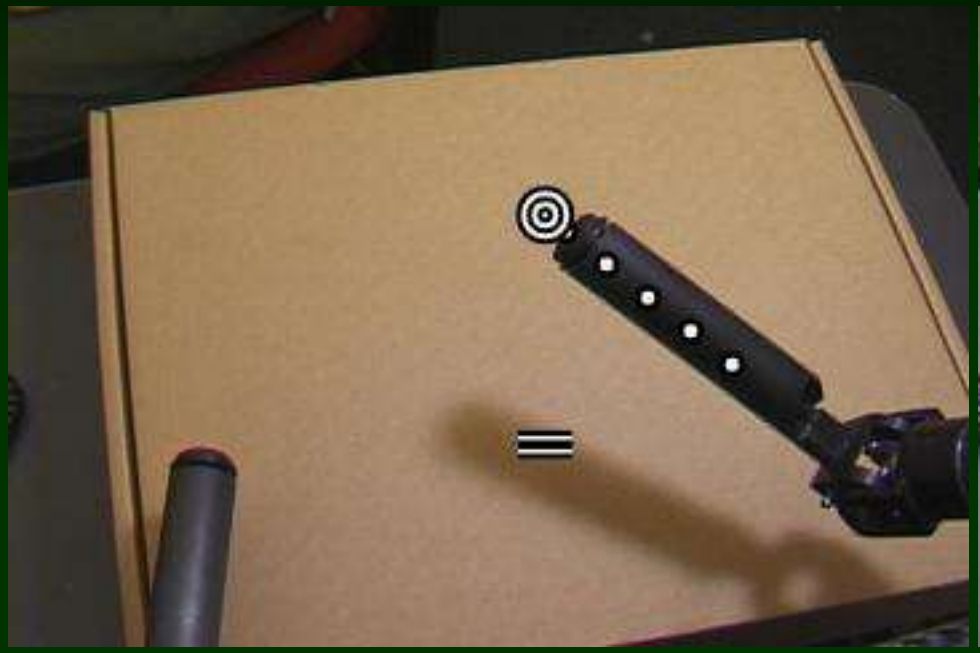
Day One

example 2: shadows

- How can a robot predict imminent contact between its hand and a surface?
- Full 3D scene recovery is one approach
  - The arm gets in the way though ...
- A complementary, opportunistic approach
  - Hand, its shadow(s) and (inter)reflections converge at impact, both in space and time
  - Shadows are complicated, but this is a *moving* shadow of an object (the hand) we control, so we've got some cross-modal knowledge

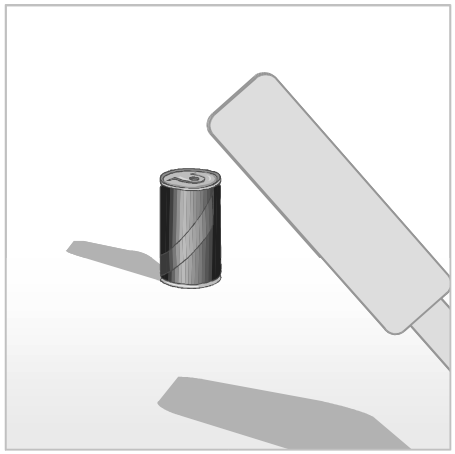
Day One

cast shadows/reflections

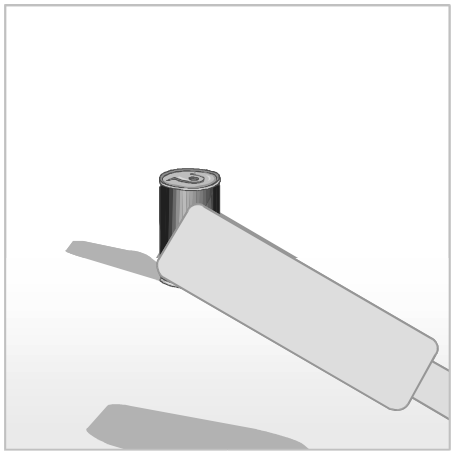


Day One

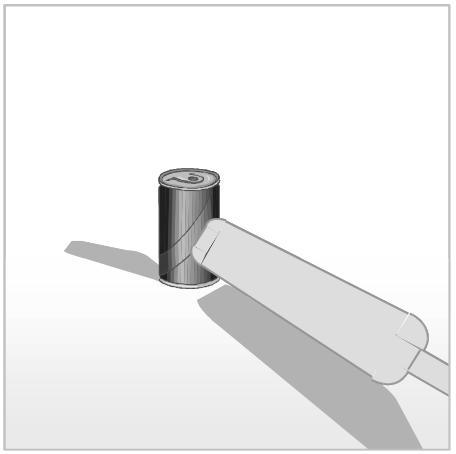
shadows for control



Robot sees target, arm, and arm's shadow



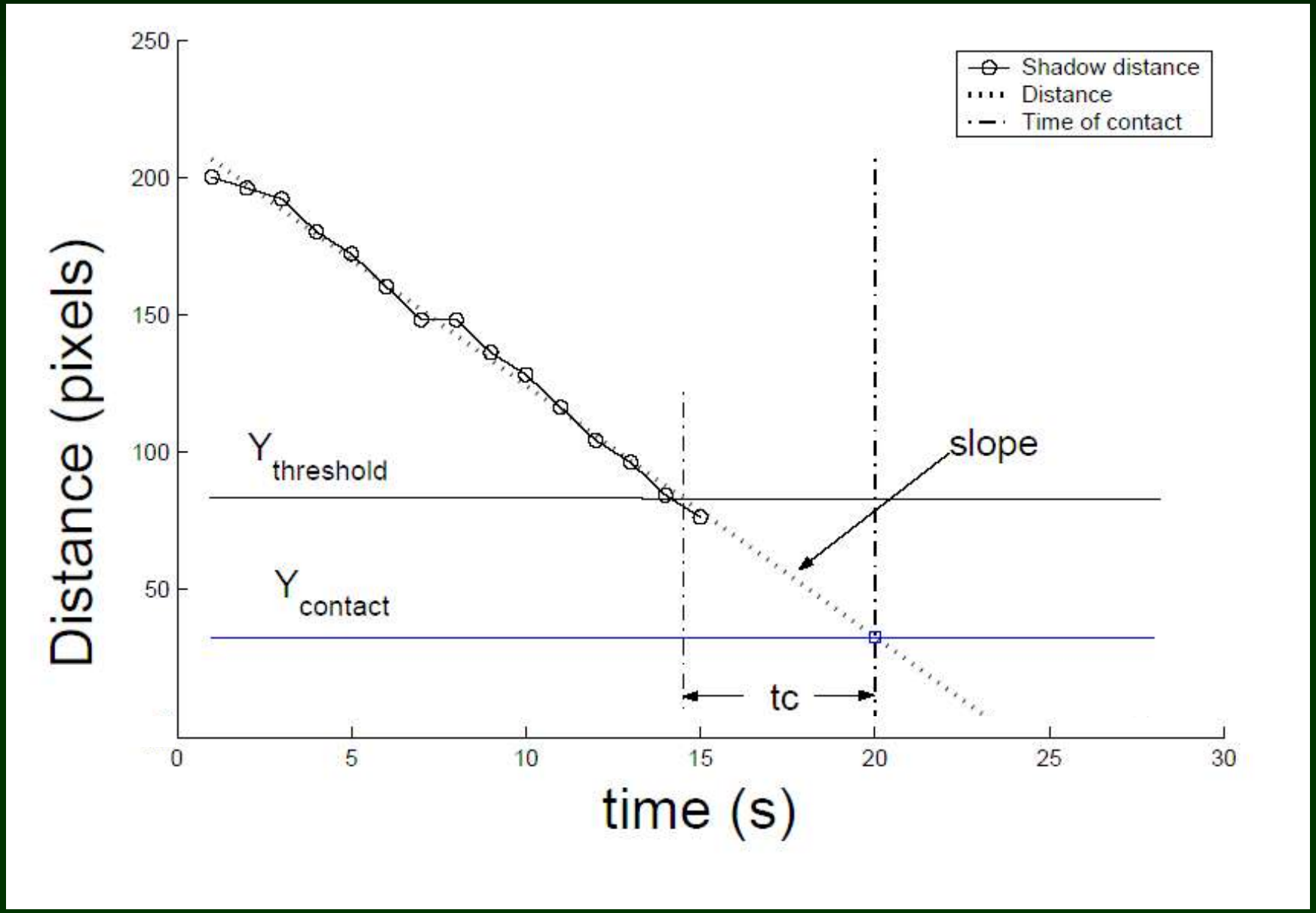
Robot moves to reduce visual error between arm and target



Robot moves to reduce visual error between arm's shadow and target

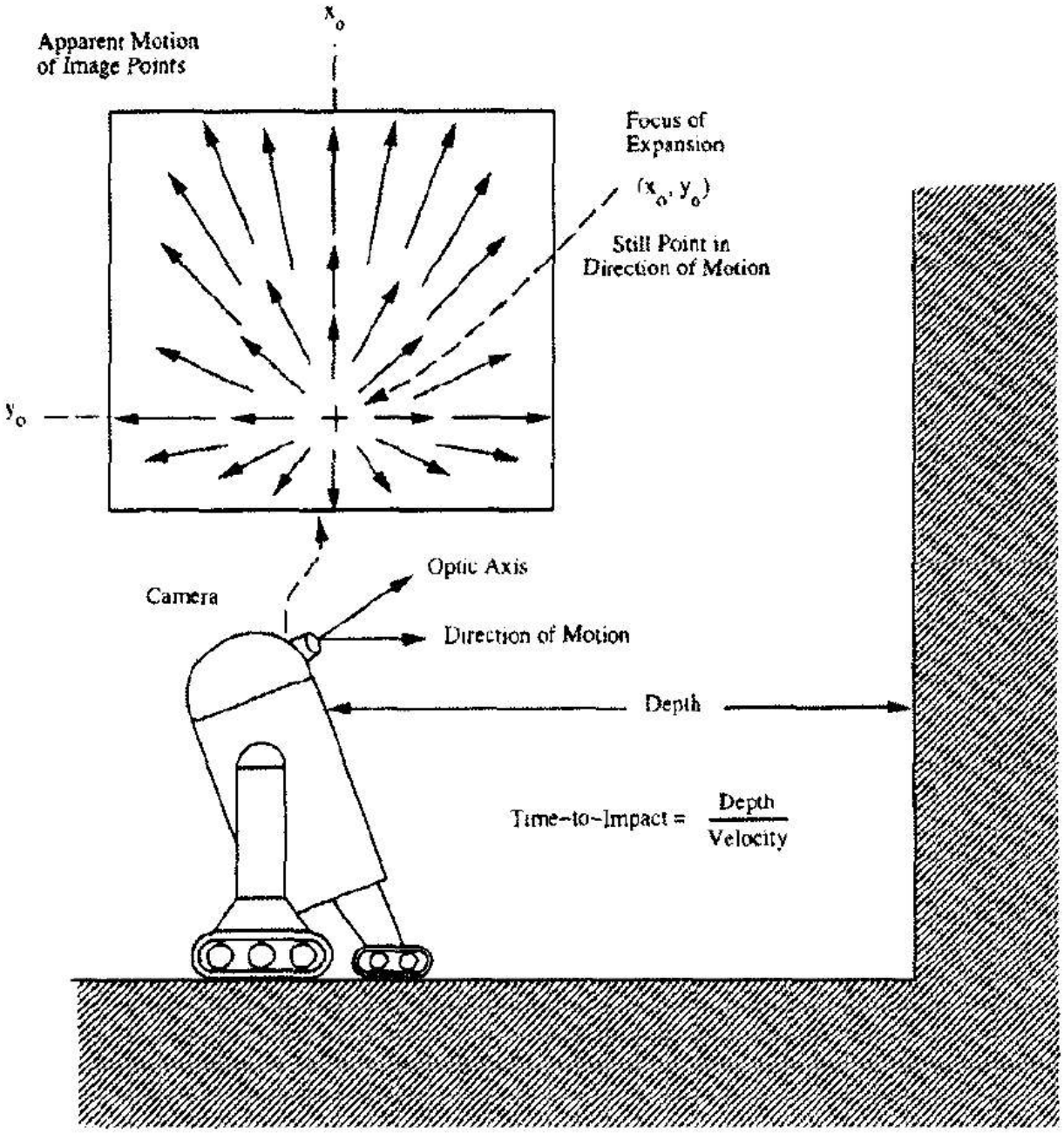
Day One

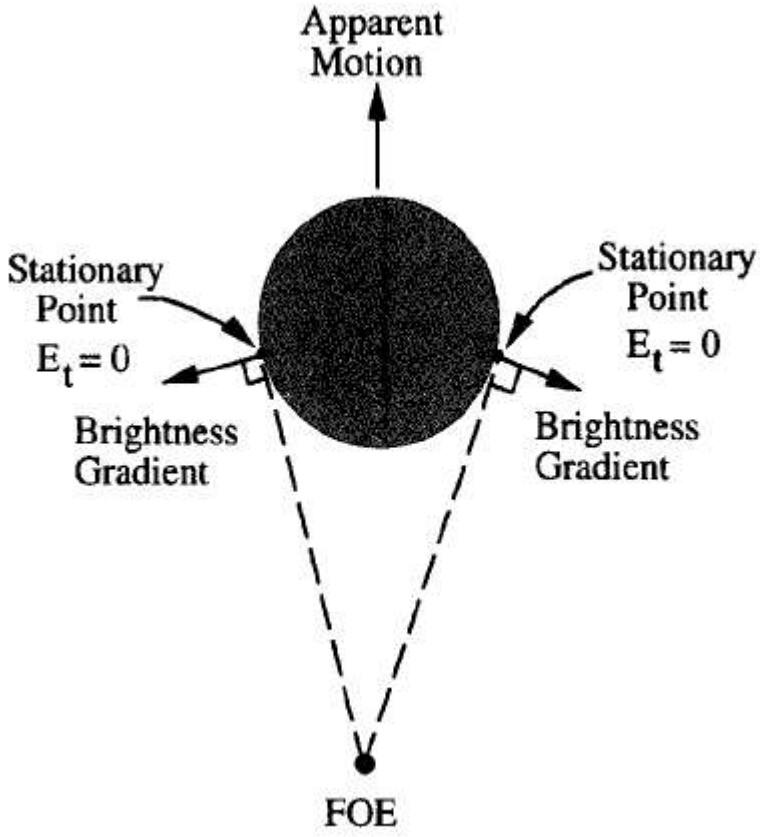
time to contact estimation



Day One

McQuirk, Horn et al



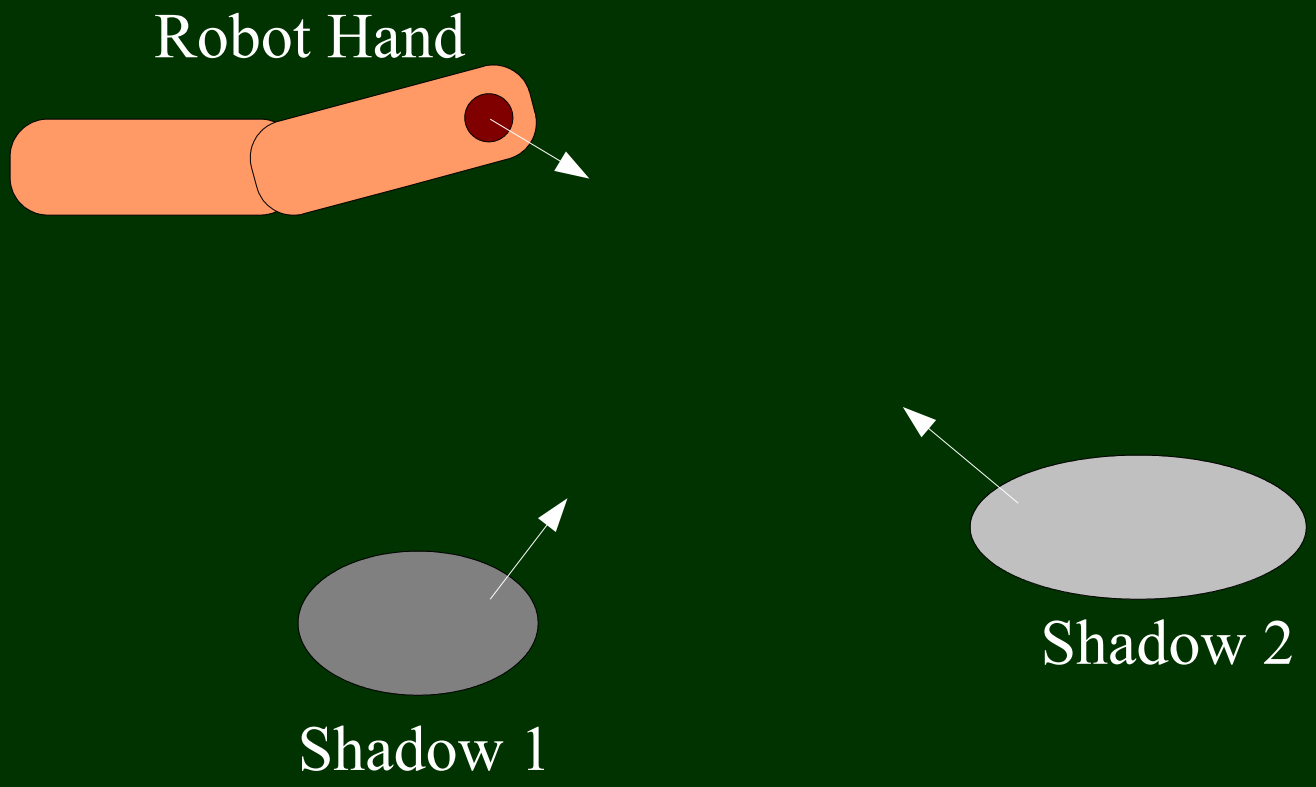


For a moving camera, find:
FOE (Focus Of Expansion)
TTC (Time To Collision)

Day One

shadows/hand convergence

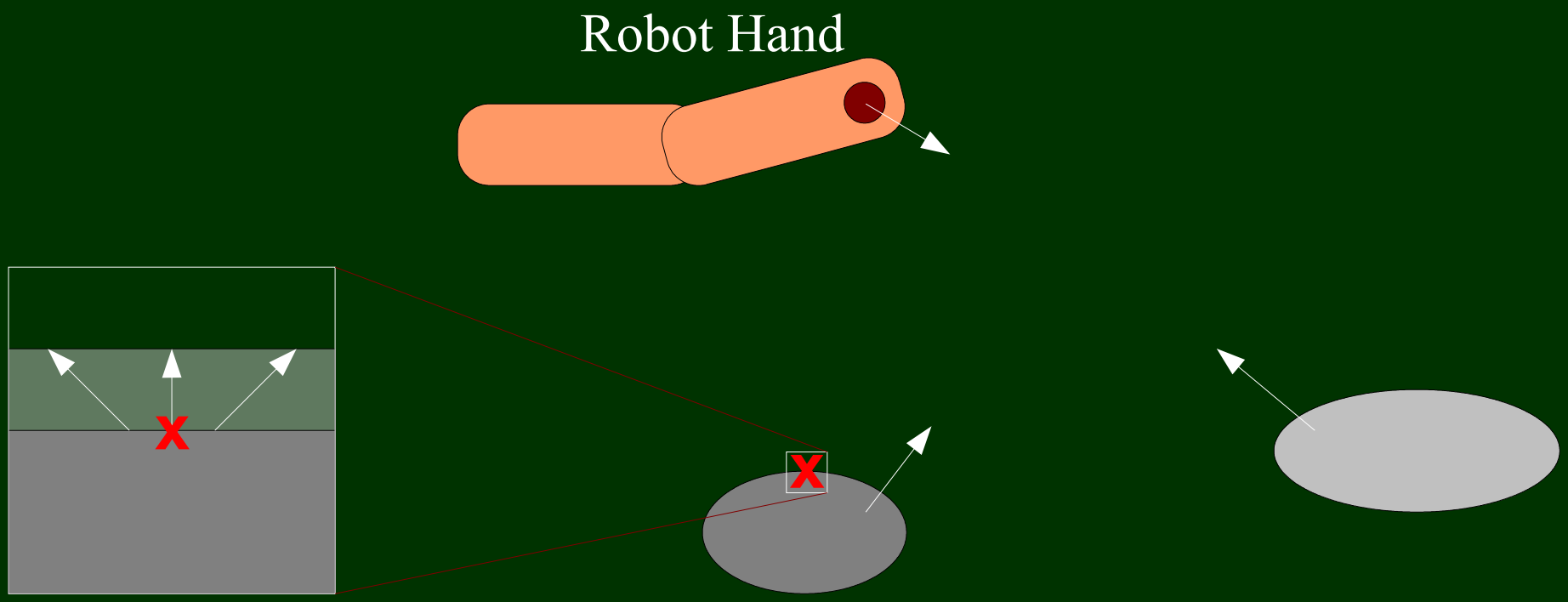
- Can we predict convergence without explicit shadow tracking?



Day One

shadows/hand convergence

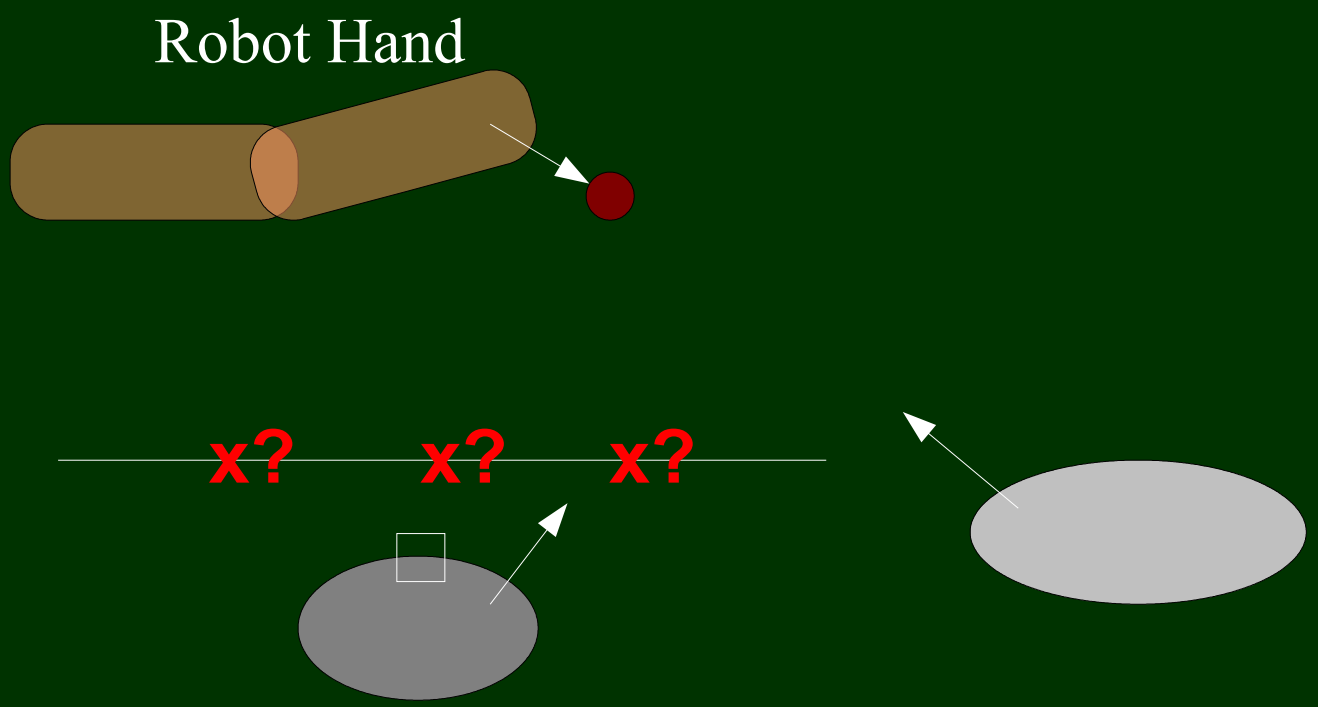
- Can we predict convergence without explicit shadow tracking?



aperture problem –
can only determine component of flow
that lies along the intensity gradient

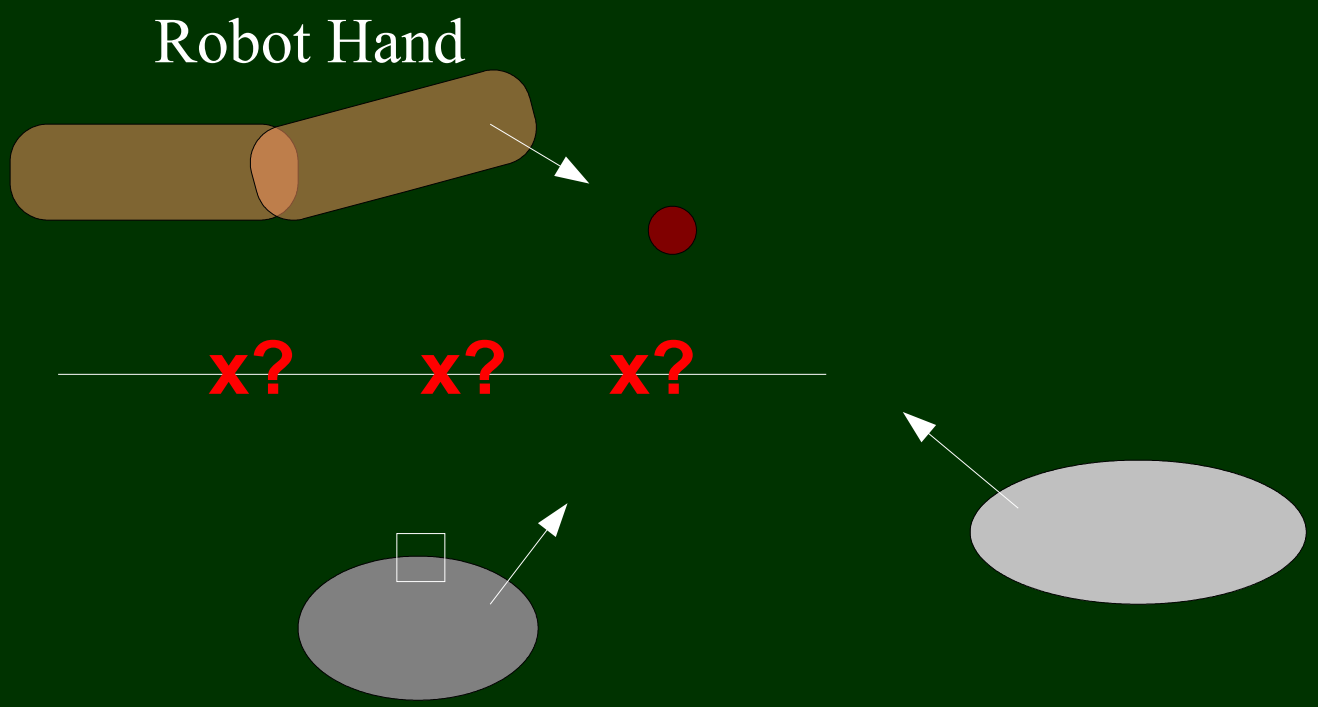
Day One

shadows/hand convergence



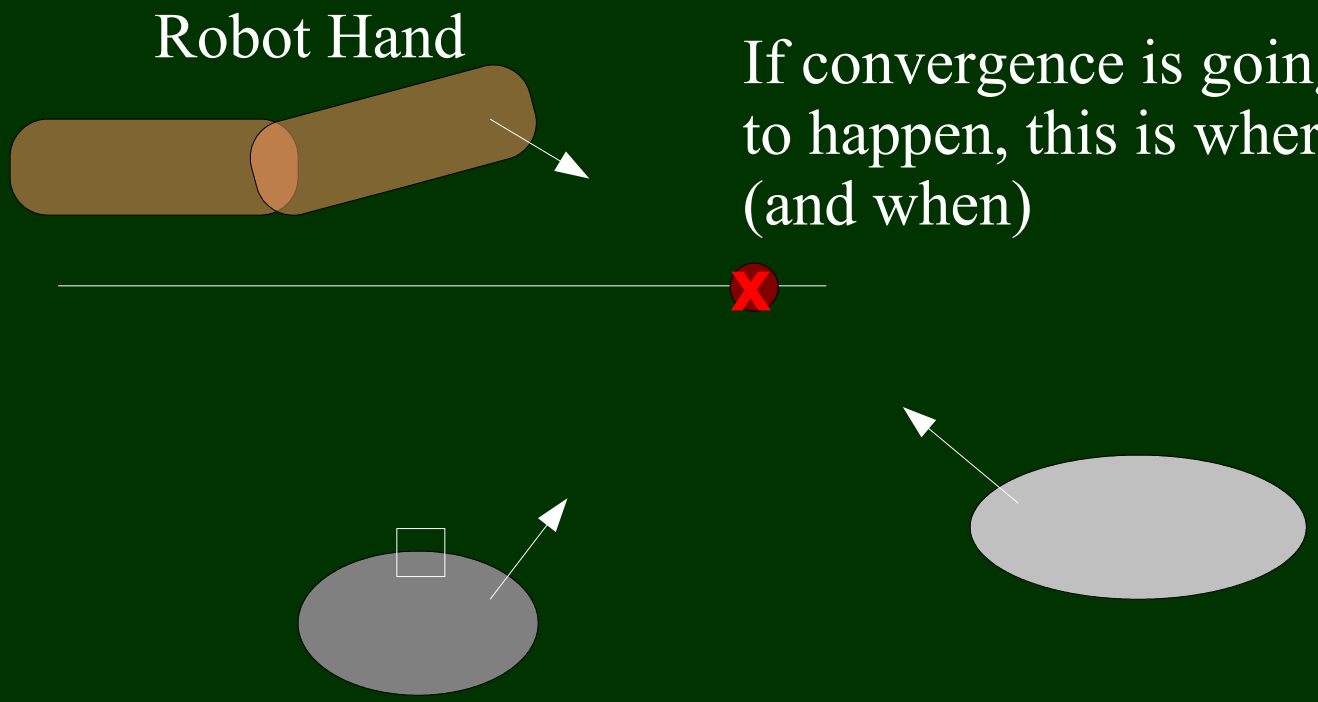
Day One

shadows/hand convergence



Day One

shadows/hand convergence



Day One

what's the point?

- Shadows/reflections/interreflections are mostly irritants in computer vision, but could be great for robotics – like have a second body
- Unlikely to be as “always-on” as stereo, but a good opportunistic complement
  - Works well on textureless surfaces, unlike stereo
- Could this approach kick-start surface perception?

Motivation

Opportunism

Meddling

Acrobatics

Conclusions

Day One

active perception

- Robots don't need to wait for opportunity to knock
  - They have a huge and growing freedom of action
  - Action can help perception (*Bajcsy, Aloimonos, etc*)
  - “Active perception” isn't just moving cameras anymore – we've got robust hands and arms, and can get into real mischief!
  - Meddling approach: if you leave anything near a robot, it should be all over it, touching and tinkering
- Manipulation *demands* active perception
  - Decomposition of action and sensing is impractical

Day One

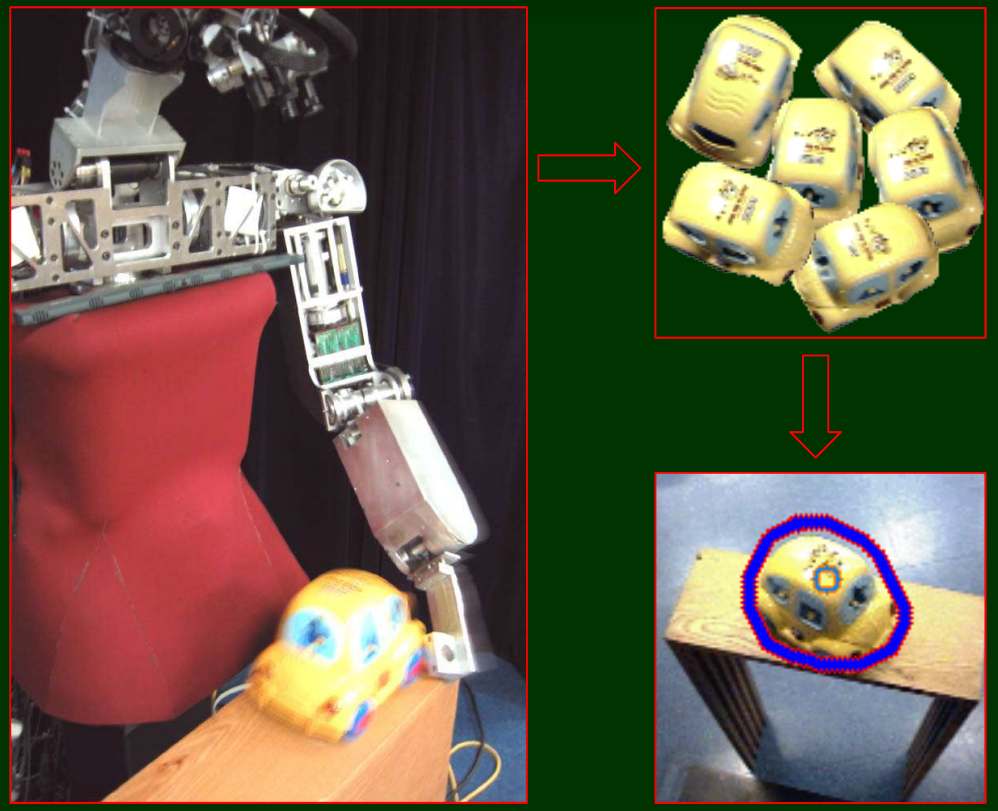
role reversal

- In robotics, vision is often used to guide manipulation
- But manipulation can also guide vision
  - Correction – detecting and recovering from incorrect perception
  - Experimentation – disambiguating inconclusive perception
  - Development – creating or improving perceptual abilities through experience

Day One

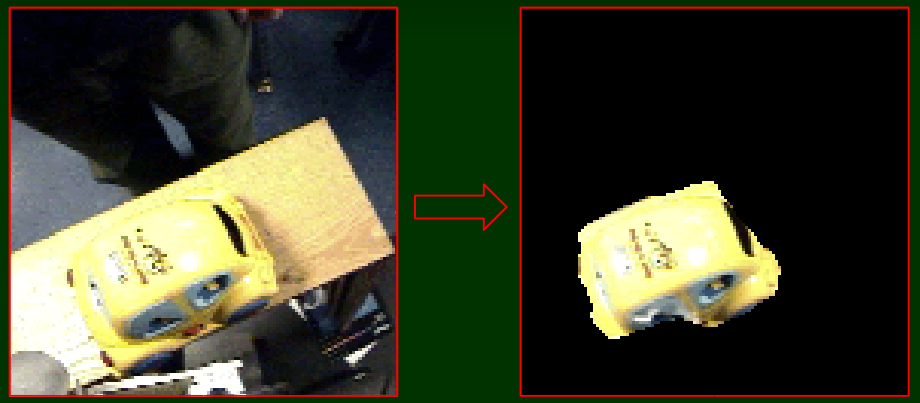
example 1: poking

- Object boundaries are not always easy to detect visually
- Solution: Robot sweeps arm through ambiguous area
- Any resulting object motion helps segmentation
- Robot can learn to recognize and segment object without further contact



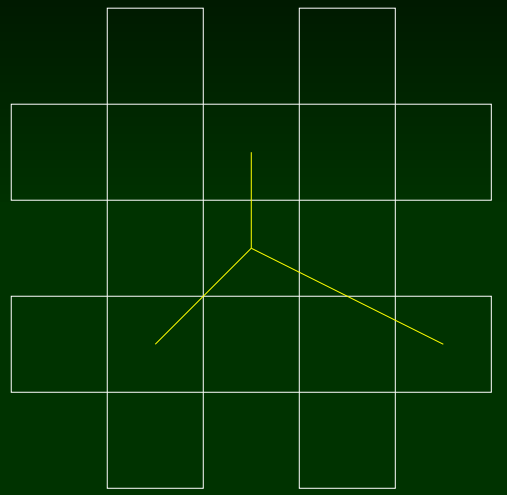
Day One

segmentation example



Day One

algorithm



- Minimum-cut segmentation into foreground and background
- 8-connected plus Knight-moves
- Each pixel classified as known-background, known-foreground, or unknown

$$A(x,y) = \begin{cases} -1, & I(x,y) \text{ is background} \\ 0, & I(x,y) \text{ is unassigned} \\ 1, & I(x,y) \text{ is foreground} \end{cases}$$

- Weights between connected pixels:

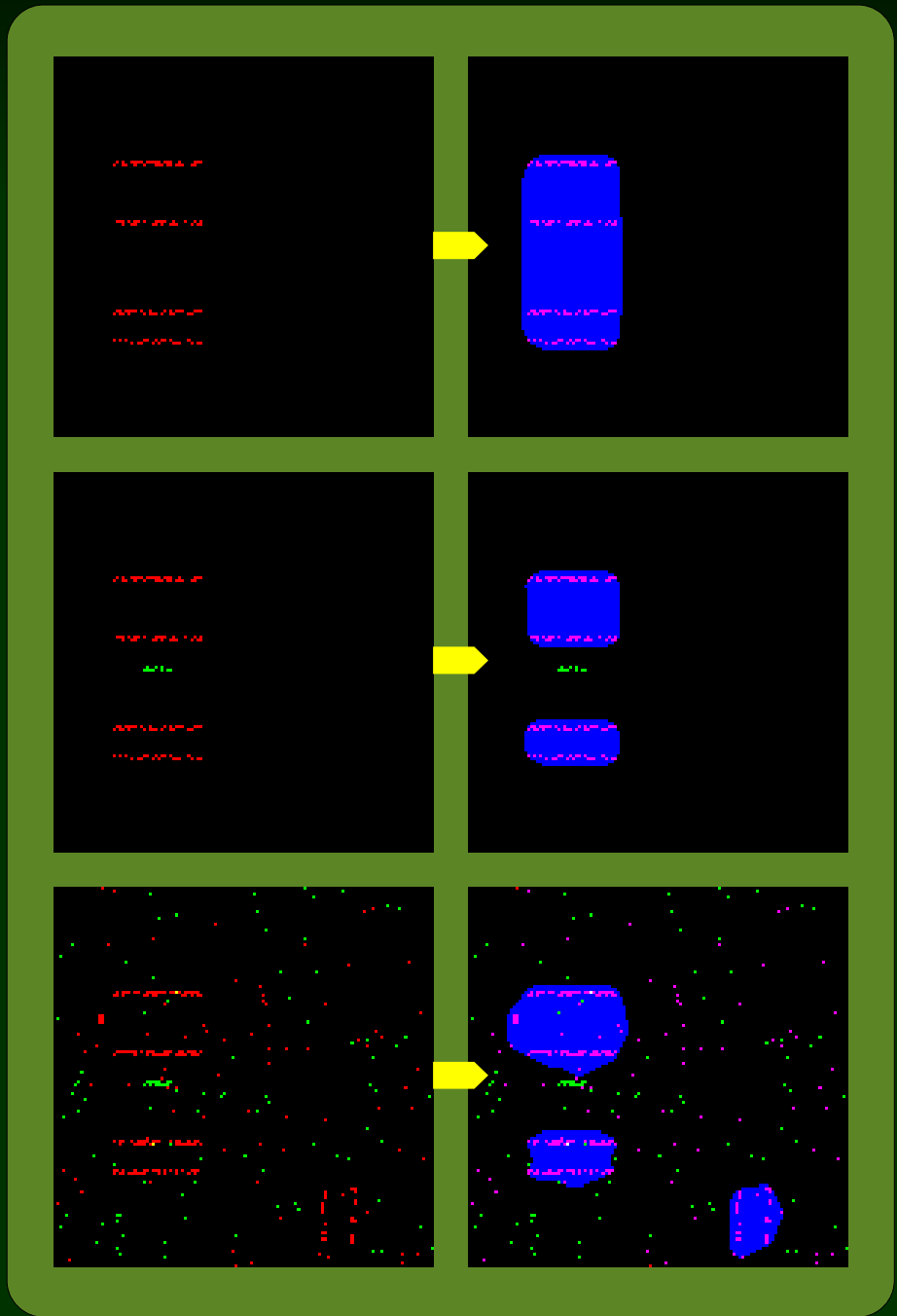
$$\mathscr{C}(N(x_0,y_0),N(x_1,y_1)) = \begin{cases} D, & A(x_0,y_0)=0, \\ & A(x_1,y_1)=0 \\ (1+k)D, & otherwise \end{cases}$$

$$\text{where } D = \sqrt{(x_0-x_1)^2+(y_0-y_1)^2}$$

Day One

operation

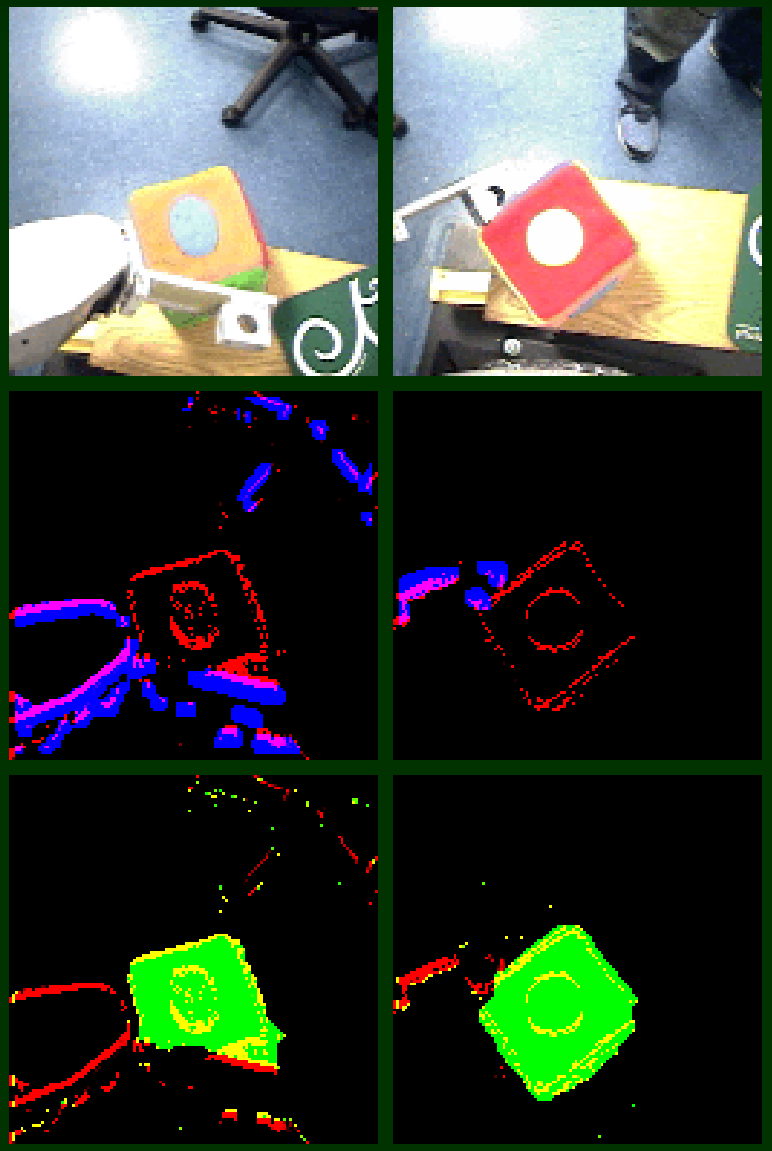
- Legend -
  - Red: known foreground
  - Green: known background
  - Blue: final segmentation
- There is a weak assignment to the background at image border
- Cost of cut is essentially the segmentation perimeter length, plus penalties for overriding assignments



Day One

advantages

- Deals well with sparseness of optic flow information
  - Mostly present just in edges perpendicular to direction of motion
- Allows us to naturally discount arm and any other objects whose movement doesn't start at the moment of impact



Day One

what's the point?

- Robot has a way to learn about unfamiliar objects
- Robot *doesn't* have to always poke something before it can see it properly

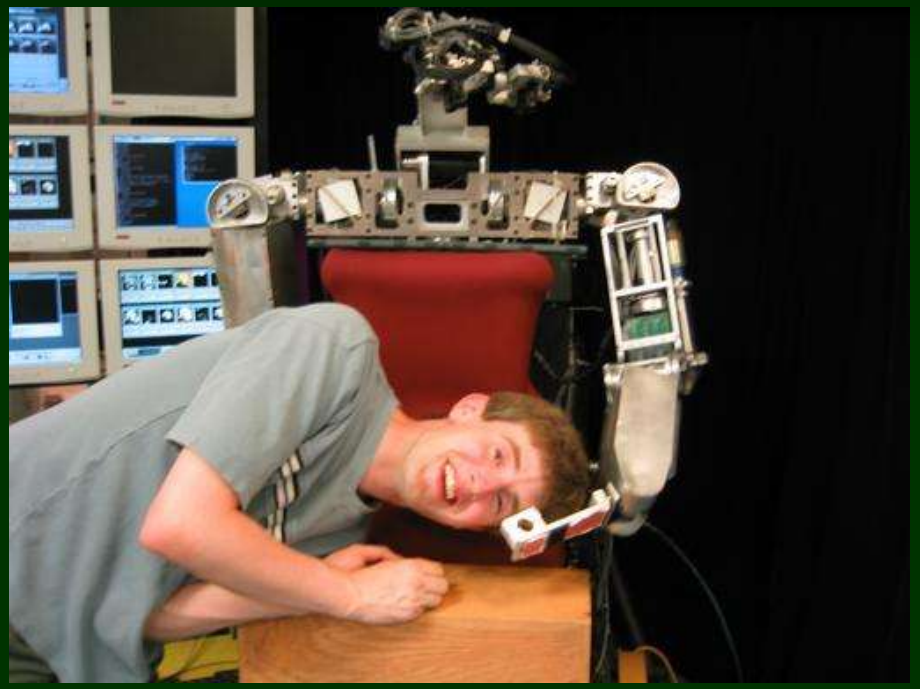
  - Learns fast – nice clean segmentations are ideal for training an object recognition system online
  - Familiar objects are detected without further contact

- Now, higher level behaviors can be layered onto a robust, adaptive foundation

- Leads naturally to exploration and exploitation of object affordances

Day One

what's the point?



- Not always practical!
- No good for objects the robot can view but not touch
- No good for very big or very small objects

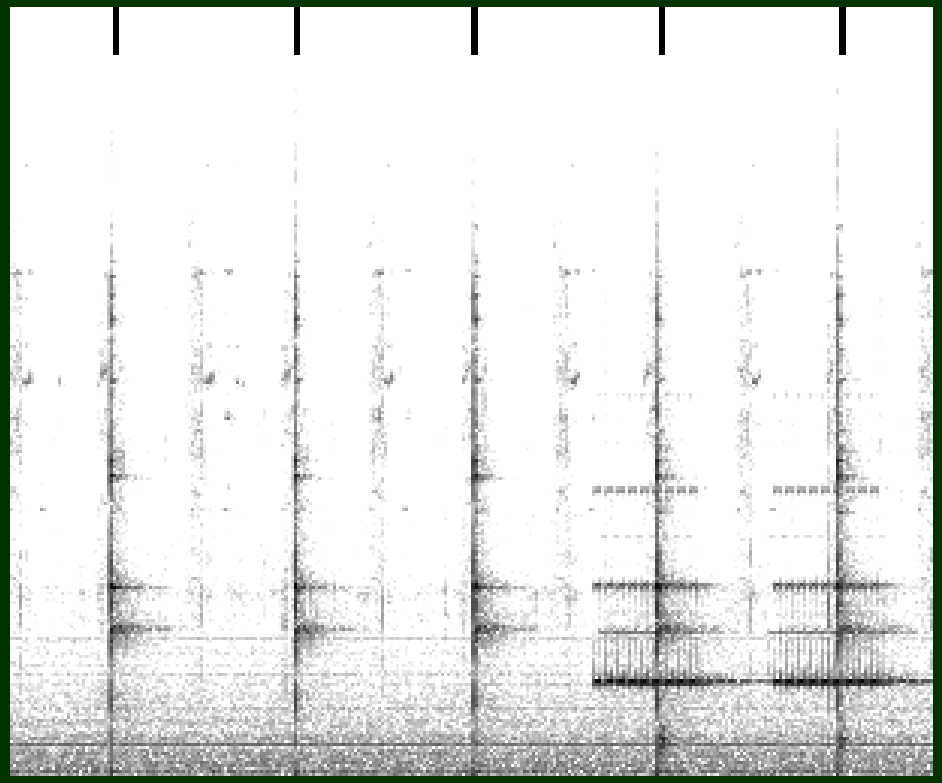
- Don't segment people this way!

- Key point: ideal for objects the robot is expected to manipulate

Day One

example 2: tapping

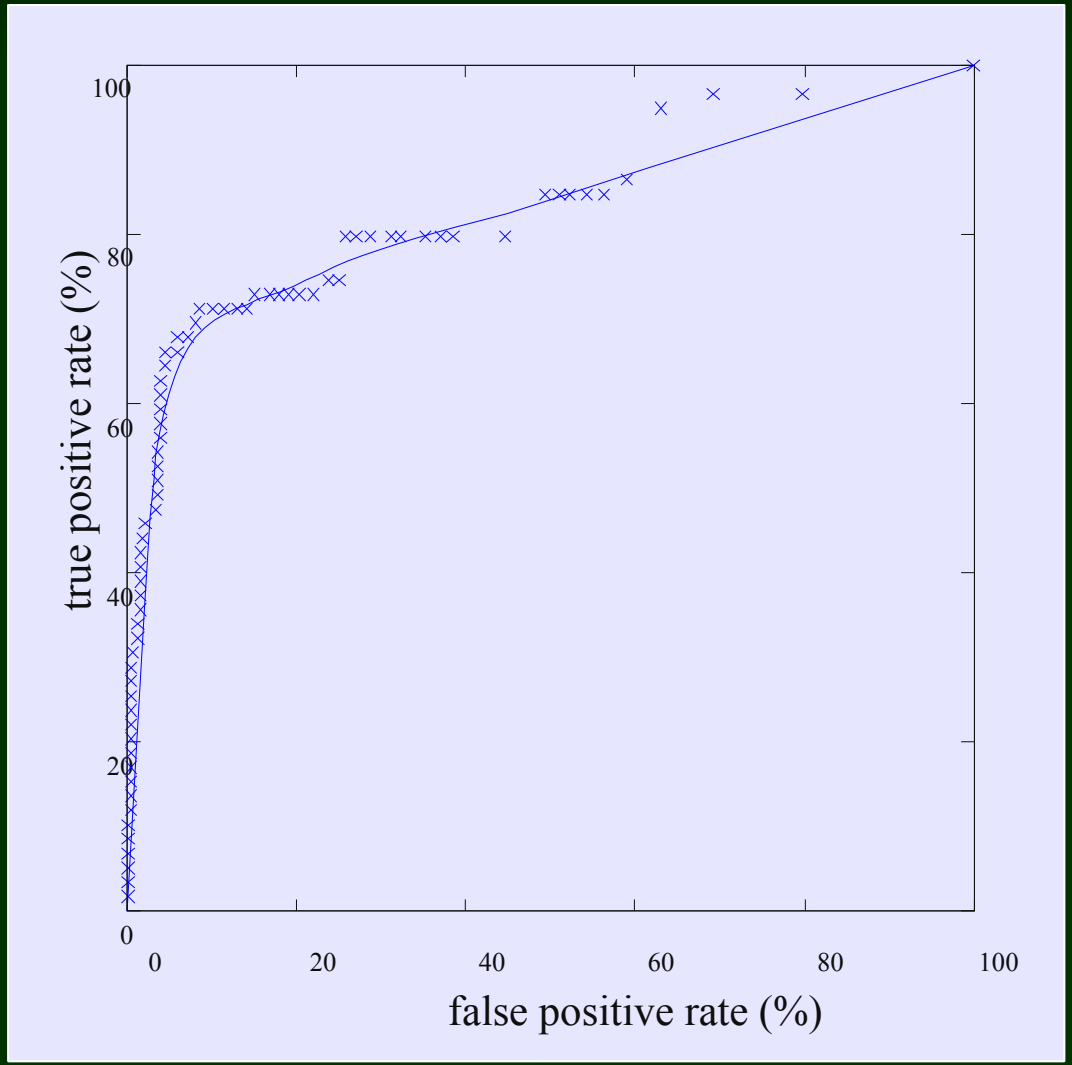
- Hybrid of amodal work and contact work – robot taps objects to learn what they sound like



Day One

performance

- Use very naïve comparison of spectral histograms
- We can match a tapping episode with 50% of previous instances involving the same object, if we accept 5% false matches



Motivation

Opportunism
Meddling
Acrobatics

Conclusions

Day One

the big picture

- Opportunism lets the robot perceive a little bit beyond what it normally can
  - Useful for its own sake
  - But crucial for learning – these increments can be aggregated, generalized, and built upon
- How far can this go?
  - You can't learn anything you don't almost already know (Patrick Winston and others)
  - Opportunities are very specific to a particular context, and so presumably will run out of steam fast

Day One

acrobatic perception

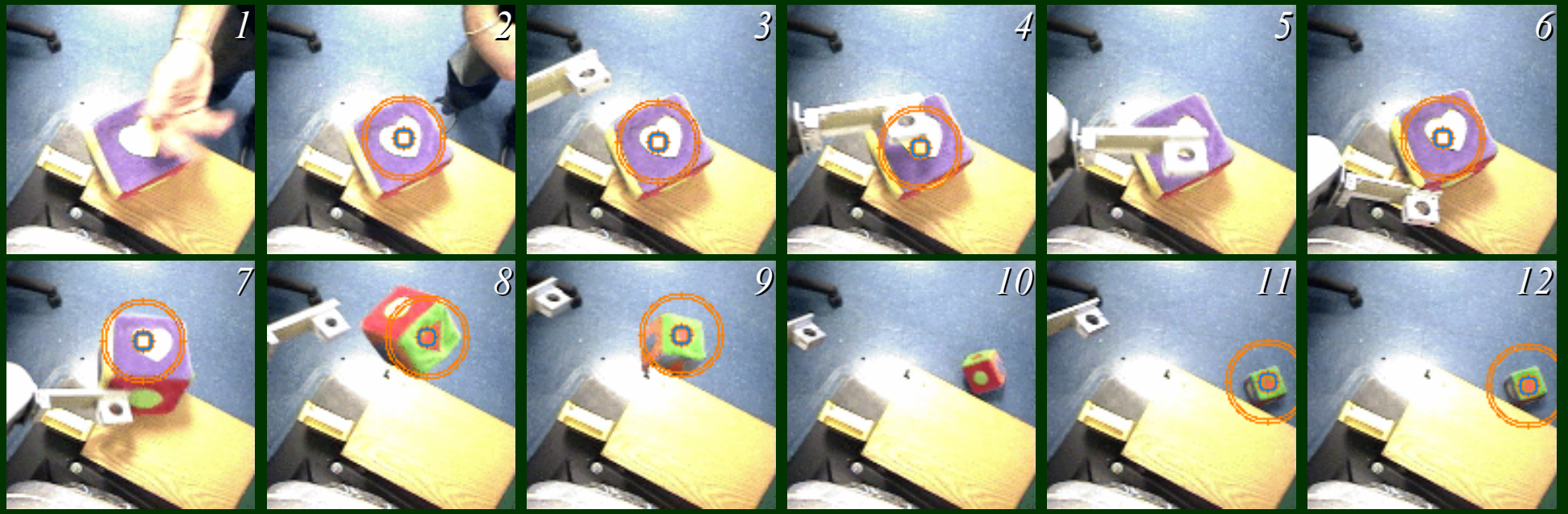
Opportunities are limited in scope ...

... But can interlock in happy ways

Day One

example 1: poking+tracking

- Poking reveals a 2D view of an object – robot may not recognize different sides as being views of the same object
- Tracking can link these opportunistically – e.g. when cube falls below, three side views are linked



Day One

example 2: a new opportunity

- In the following videos, the robot is observing a “search” activity that follows a regular script
- First, it sees searches for familiar objects, allowing it to learn the structure of the activity
- Then, it sees a search for an unfamiliar object, and uses the activity structure to make a novel inference
- (based on Tomasello '97 – word learning in infants non-ostentive contexts)

Motivation

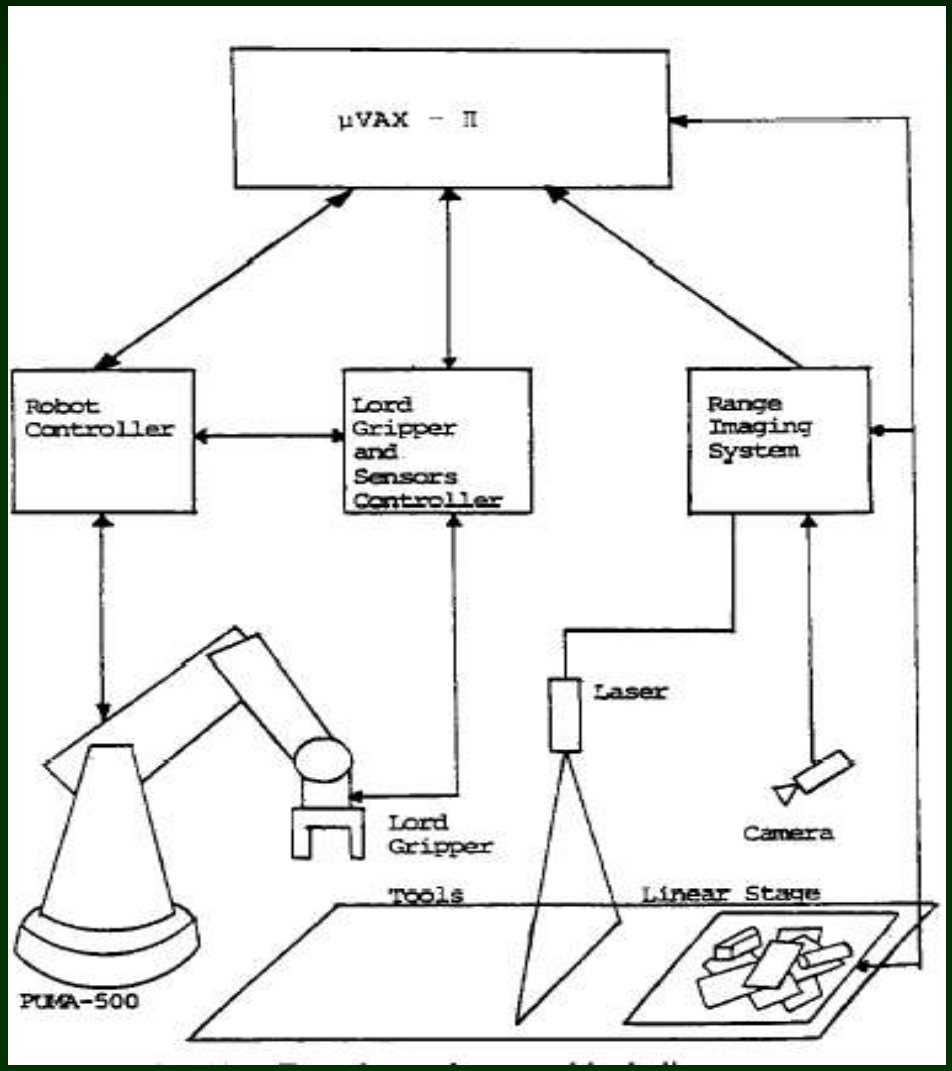
Opportunism
Meddling
Acrobatics

Conclusions

Day One

active segmentation, 1991

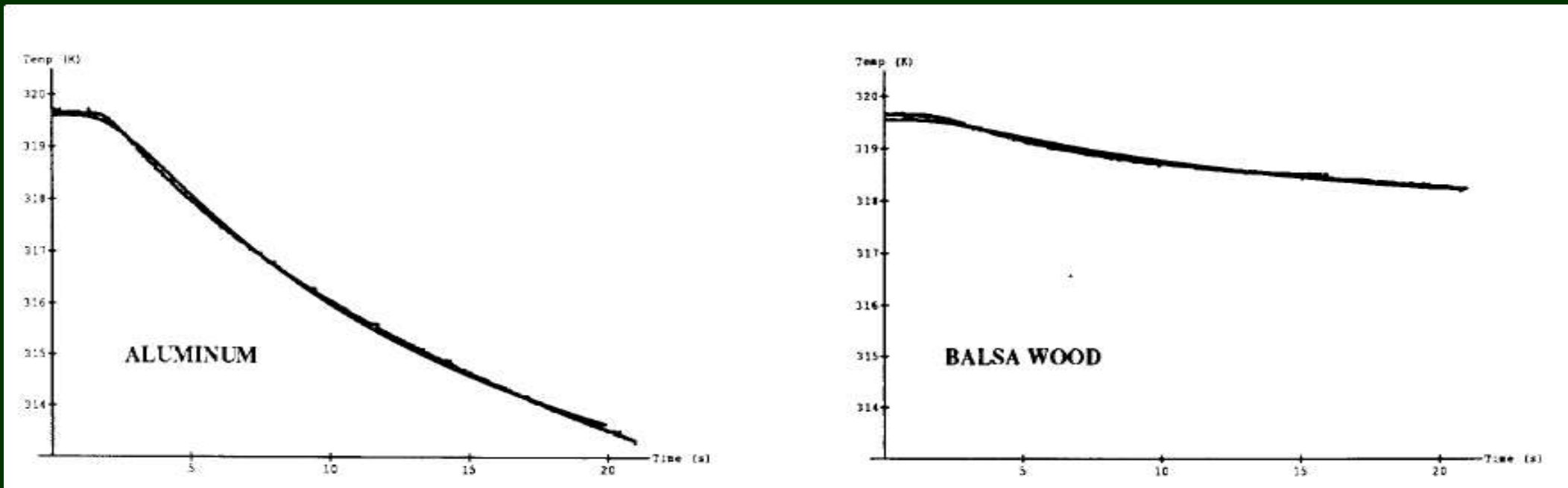
- Tsikos, Bajcsy, 1991
  - “Segmentation via manipulation”
  - Simplified understanding of cluttered scenes by physically moving overlapping objects



Day One

exploratory procedures, 1991

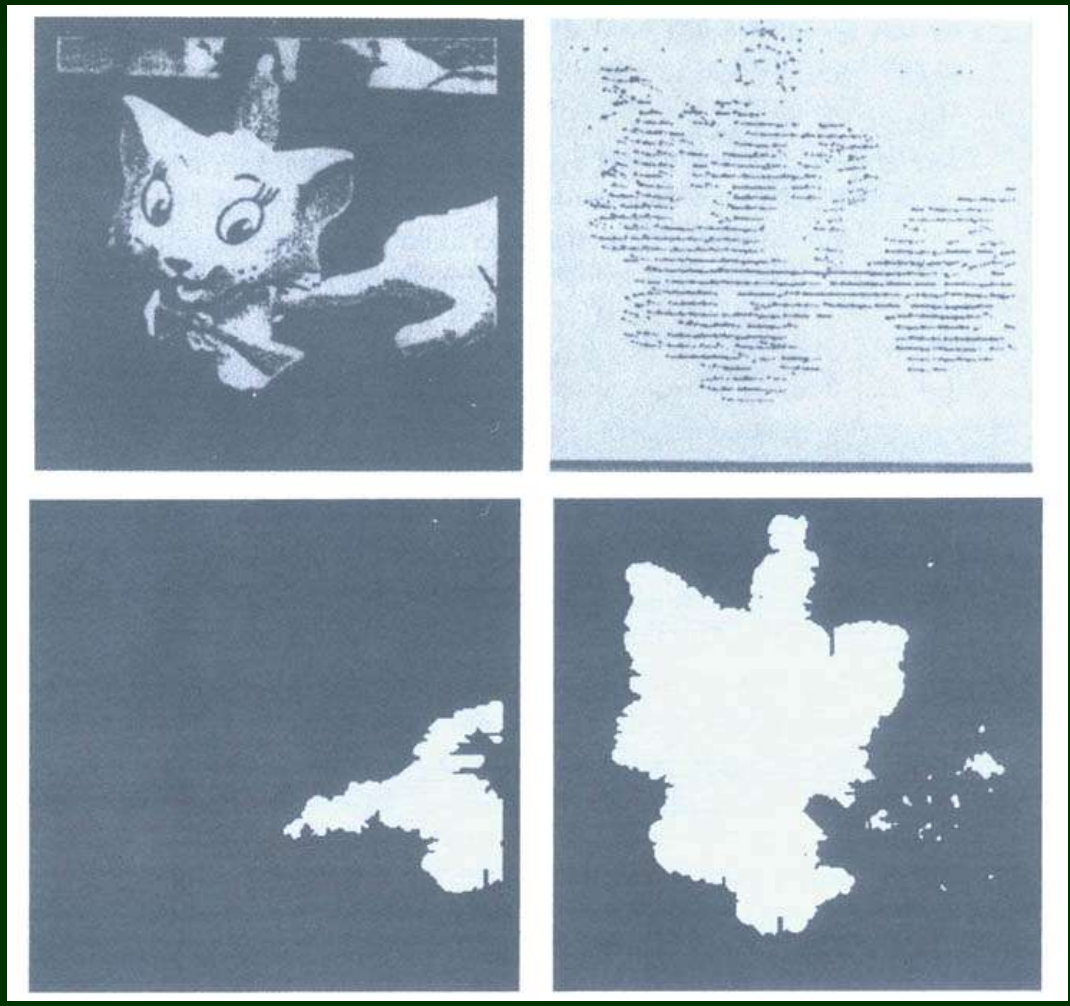
- Campos, Bajcsy, Kumar 1991 – observed that robots could use exploratory procedures identified in humans for haptic perception
- Implemented sensitivity to thermal diffusivity (distinguishes “cold” metal from “warm” wood)



Day One

poking, 1993

- Sandini et al, 1993
  - “Vision during action”
  - interpreted motion during manipulation to deduce object boundaries
- Same basic idea as poking
- Just didn't have processing power



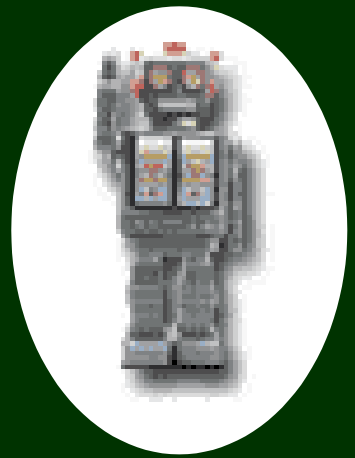
Day One

interesting times

- Robotics: a slashdot topic since March 04
  - “Chainsaw-wielding Robotic Submarine”

  ...
  - “Toyota to Employ Advanced Robots”
  - “First Peek at Robosapien V2”
  - “Humanoid Robot KHR-1 SDK Released”

  ...
  - “Camel-riding Robots”



Day One

interesting times

- Nov/Dec 04 white paper on “mobile manipulation”
  - “We are advancing this argument now because new developments regarding actuation and sensing promise to make robots more responsive to unexpected events in their immediate surroundings. This is a boon to mobility technology and is the “missing link” to producing integrated manipulation systems.”

Robert Ambrose	Andrew Fagg	Lawrence Leifer
Christopher Atkeson	Roderic Grupen	Maja Matarić
Oliver Brock	Jeffrey Hoffman	Randal Nelson
Rodney Brooks	Robert Howe	Alan Peters
Chris Brown	Manfred Huber	Kenneth Salisbury
Joel Burdick	Oussama Khatib	Shankar Sastry
Mark Cutkosky	Pradeep Khosla	Robert (Bob) Savely
	Vijay Kumar	Stefan Schaal

Day One

how to measure progress?

- Robotics is notoriously difficult to evaluate
  - Incomparable hardware, behavior, goals
- Mechanical progress effectively measured by video
  - Terrible, but not a complete and utter disaster
- What about progress in perception?
  - Much less visible – is action canned or responsive?
  - DayOne goal: consider *range* of behavior enabled
    - All the possible things the robot could do after (e.g.) 24 hours – not just the coolest one or two things (which could be canned)

Day One

mensa flexa in corpore flexo

- Perceptual ability is lagging mechanical ability in robotics, but that may soon change
  - Active perception is hugely more interesting with arms and hands (rather than just moving cameras)
- The behavior of general-purpose robots will be anything but “robotic”
- Now is the time

Day One

acknowledgments

- Illustrious leader:
  - Rod Brooks
- Cool collaborators:
  - Giorgio Metta (poking)
  - Artur Arsenio (amodal)
  - Eduardo Torres-Jara (tapping, shadows)
  - Lorenzo Natale (grasping, tapping)
- Body builders:
  - Matt Marjanovic, Matt Williamson, Brian Scassellati, Cynthia Breazeal, ...