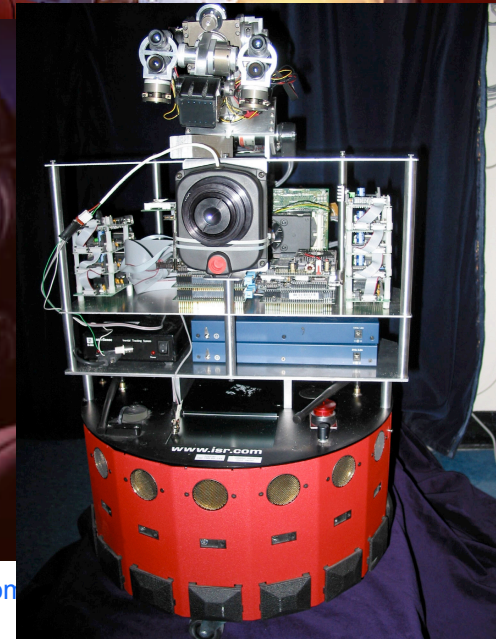
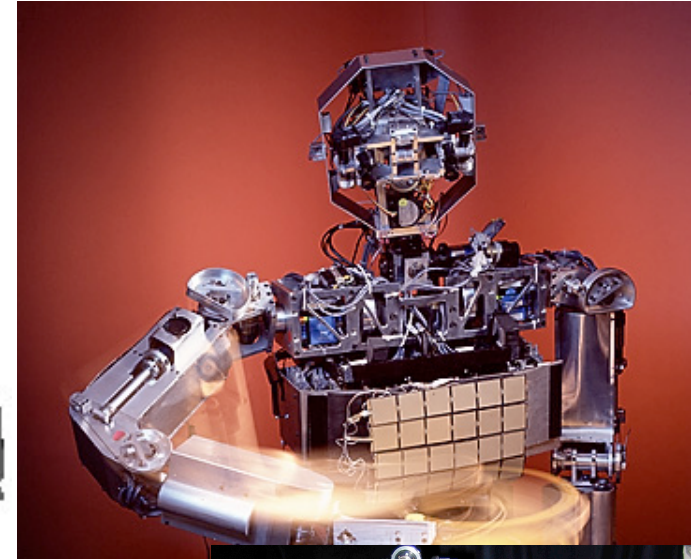
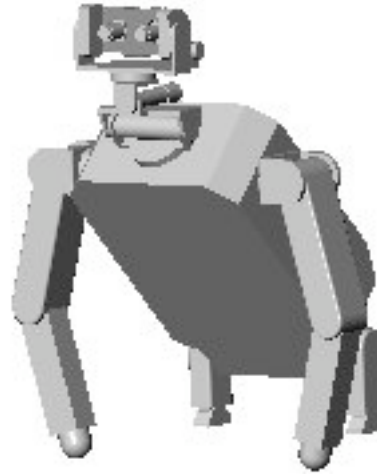




Natural Tasking of Robots Based on Human Cues -- MIT AI Lab

- **Challenge:** Make commanding robots as intuitive and natural as commanding soldiers
- We will build software systems (on 4 robot platforms) that
 - interact with people using intuitive social cues
 - are self-adapting through developmental learning
 - exploit their physical embodiment to simplify computation
 - use multi-modal sensory integration to maximize effectiveness





Research Issues: Learning by Imitating Humans

- **Knowing what to imitate**
 - what sensory signals are relevant to the current task?
- **Mapping between bodies**
 - how to convert sensory view of one agent to self-mapping?
- **Correcting failures and recognizing success**
 - how can the robot automatically develop success measures?
- **Chaining actions together**
 - how can the learned pieces be chained together into complete task sequences?
- **Generalizing to more complex tasks**
 - how to use invariants and common-sense knowledge to adapt the learned sequences to new situations
- **Intuitive interactions for the human**
 - how to use natural social signals to make it intuitive for humans and robots to communicate with each other



Our Approaches Lead to Specific Progress on Each Research Issue

Research Issues

Our Approaches

	Knowing what to imitate	Mapping between bodies	Correcting failures and recognizing success	Chaining actions together	Generalizing to more complex tasks	Intuitive Interactions
Social Interaction	Uses attentional cues to recognize task relevant events		Identifies and displays emotional states for reinforcement			Allows humans and robots to share similar social cues
Development	Limits search space by incremental refinement of perception	Simplifies mapping by incremental refinement of motor skills		Provides natural decomposition of complex tasks	Exploits incremental learning to build new skills	Provides methods for building social skills
Embodiment	Constrains movement which assists directed perception	Simplifies body mapping between human and robot	Enables simple but robust low-level behaviors	Places physical limits on successive actions		Allows the human to observe natural social feedback cues
Integration	Allows robot to recognize cues in multiple modalities		Increases robustness through redundancy		Assists in transfer to new modalities	Allows human to use natural modalities (voice, gesture, facial expressions)

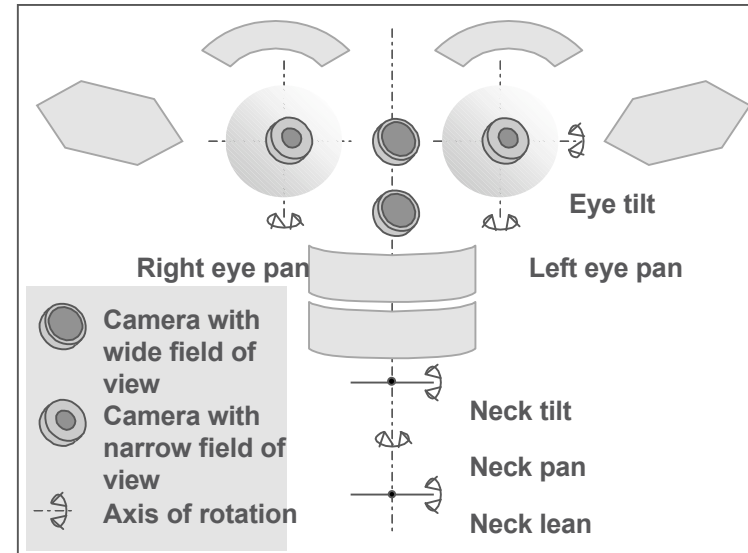
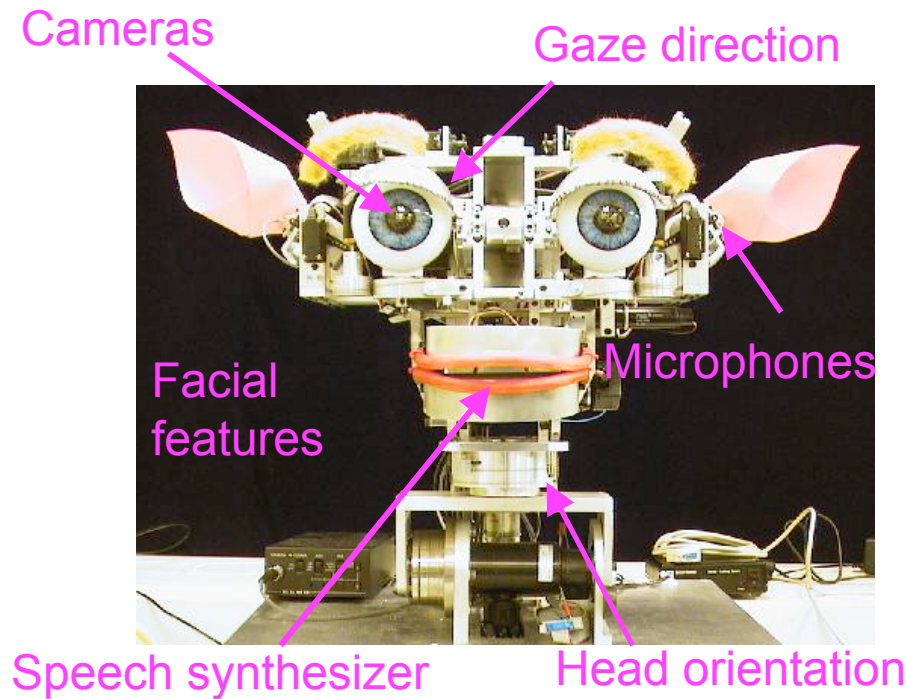


Contributors To This Talk

- **Artur Arsenio**
- **Lijin Aryananda**
- **Cynthia Breazeal**
- **Paul Fitzpatrick**
- **Matthew Marjanovic**
- **Giorgio Metta**
- **Una-May O'Reilly**
- **Brian Scassellati**
- **Paulina Varchavskaia**



Designing for the Human In the Loop





Social Cues and Interaction

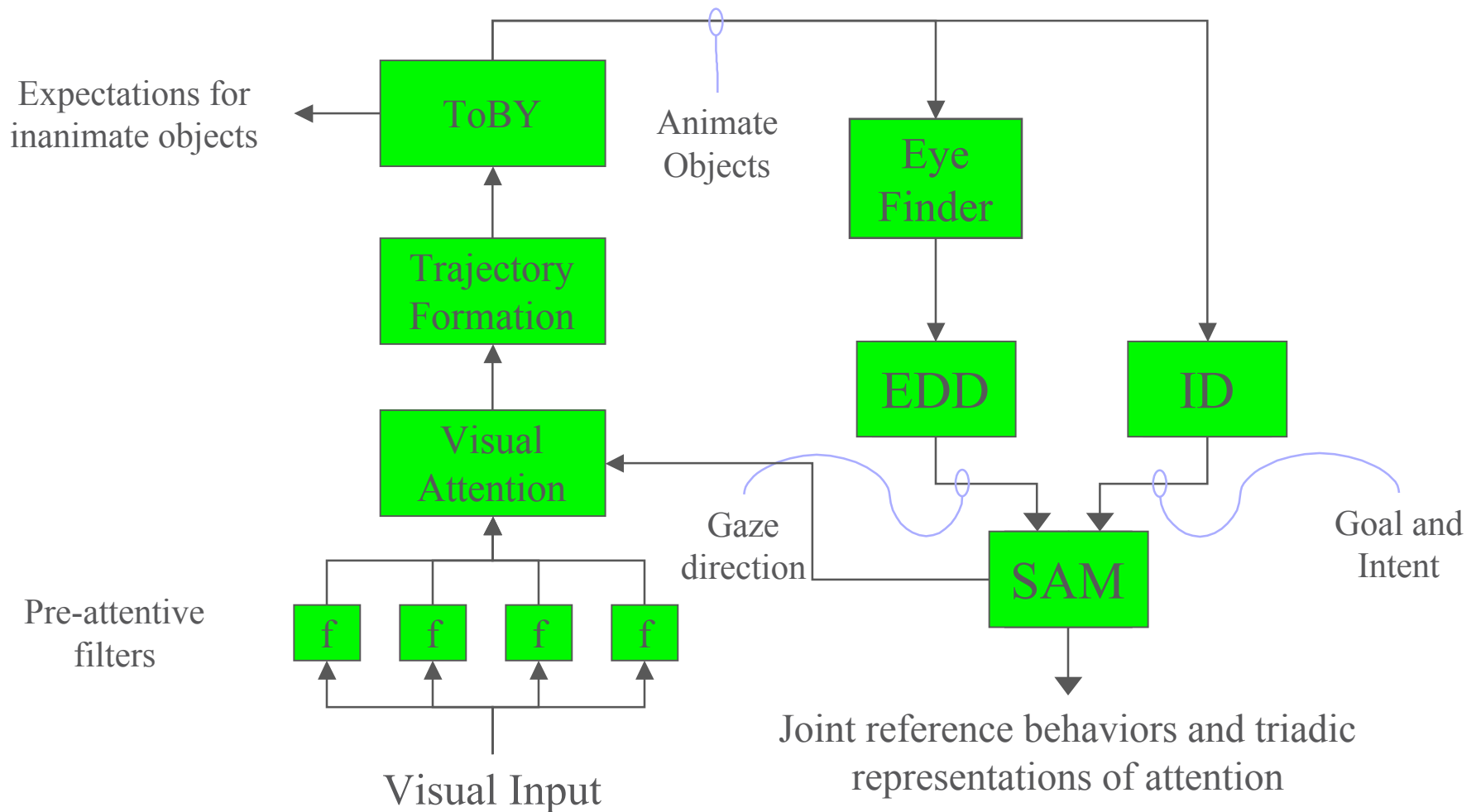


Theory of Mind

- For complex learning tasks the robot needs to
 - understand what it is the human knows
 - » about the world
 - » about the robot's current state
 - understand what the human is doing and why
 - understand what is the human

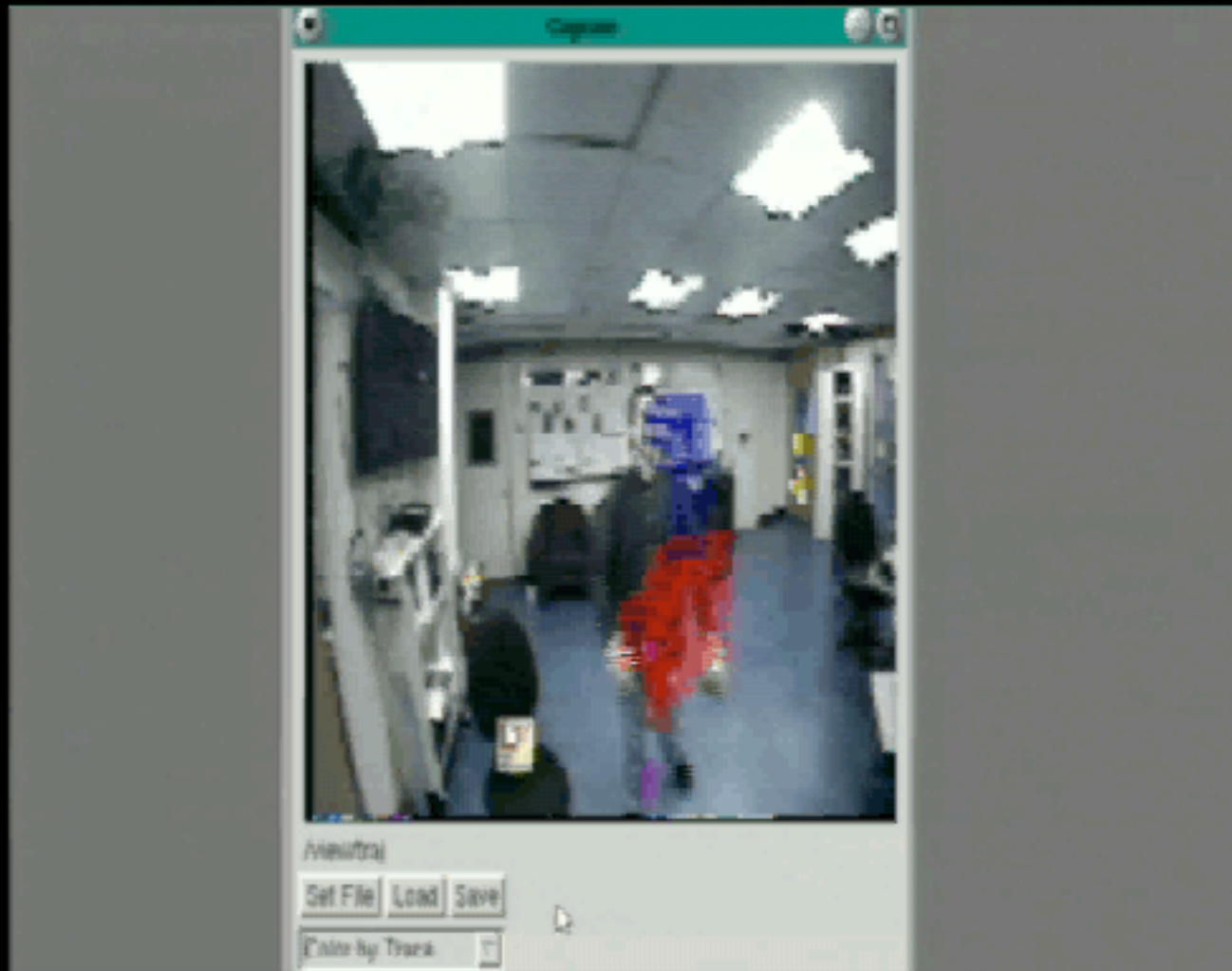


Scassellati's Theory of Mind





Trajectory Example





Animacy Results



Animate



Inanimate

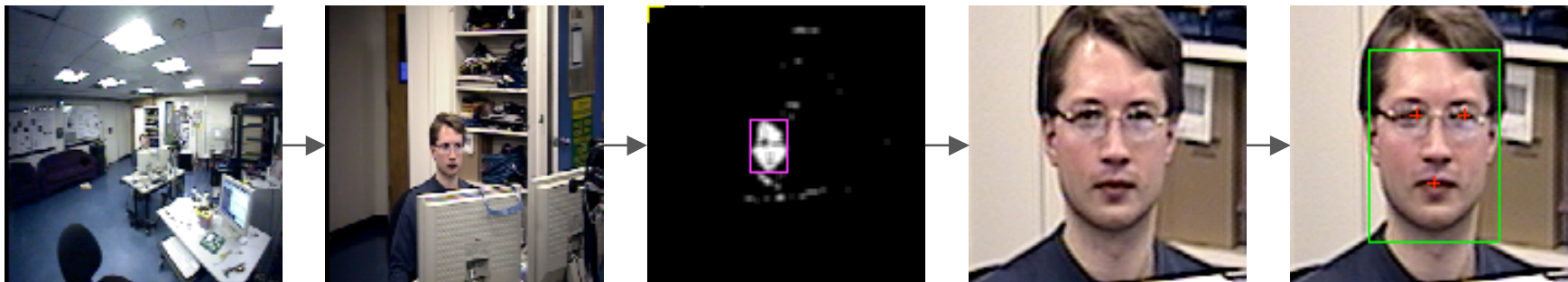
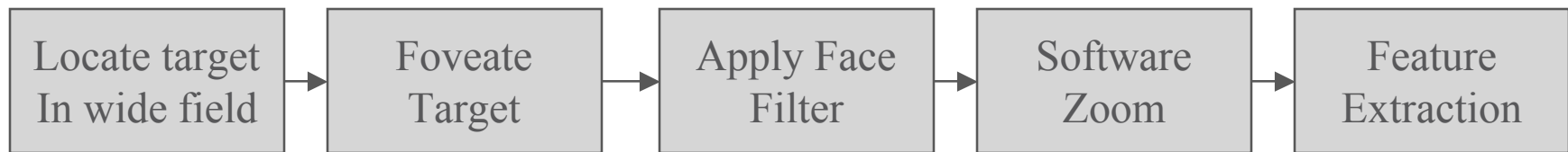


Animate

- Arbitration methods
 - Weighted sum
 - Winner-take-all



Post-Attentive Visual Processing: Finding Faces and Eyes





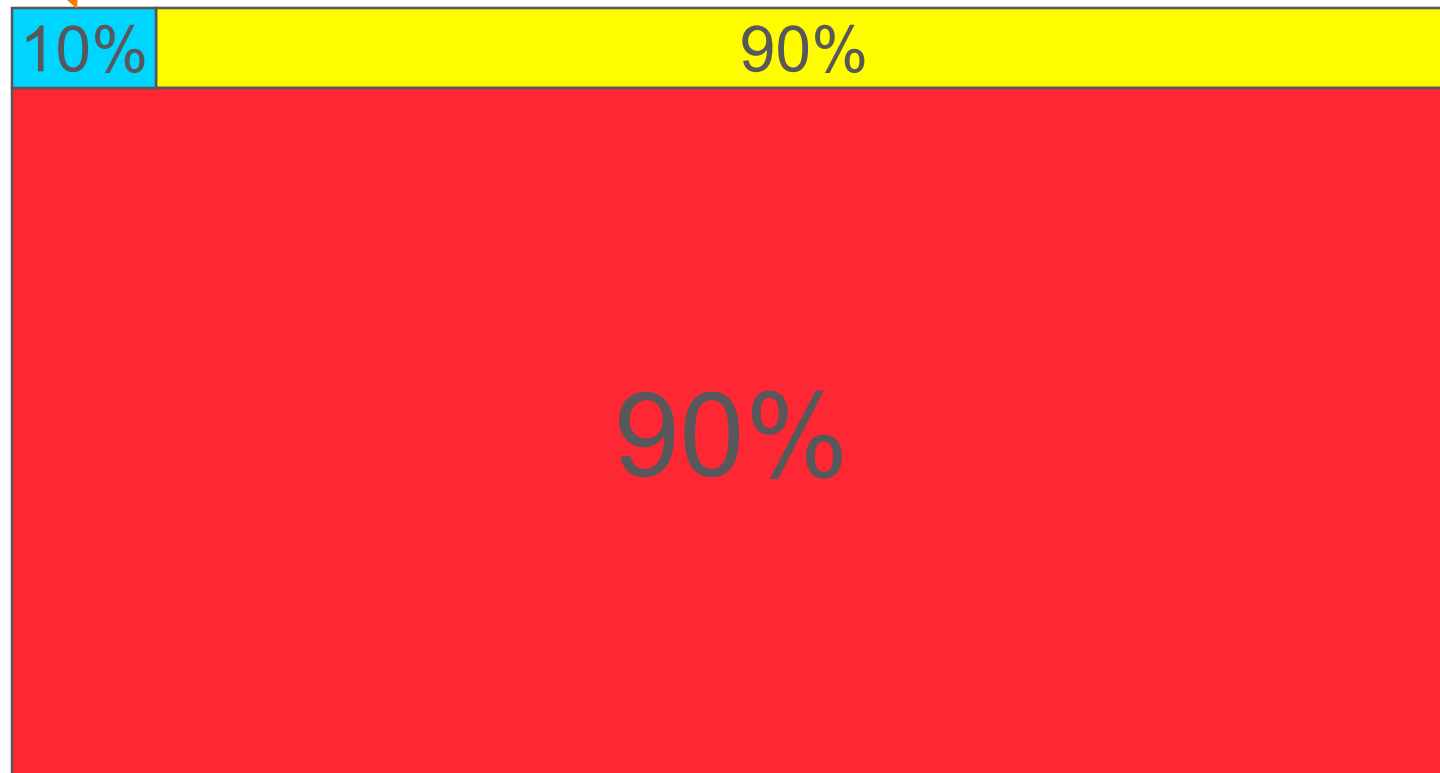
Shared Attention (New)



Getting Real Robots Into Service

Research ideas
and demonstrations

Developing to the point
of military demonstrations





Current Work Areas

Research Issues

Our Approaches

	Knowing what to imitate	Mapping between bodies	Correcting failures and recognizing success	Chaining actions together	Generalizing to more complex tasks	Intuitive Interactions
Social Interaction	Uses attentional cues to recognize task relevant events		Identifies and displays emotional states for reinforcement			Allows humans and robots to share similar social cues
Development	Limits search space by incremental refinement of perception	Simplifies mapping by incremental refinement of motor skills		Provides natural decomposition of complex tasks	Exploits incremental learning to build new skills	Provides methods for building social skills
Embodiment	Constrains movement which assists directed perception	Simplifies body mapping between human and robot	Enables simple but robust low-level behaviors	Places physical limits on successive actions		Allows the human to observe natural social feedback cues
Integration	Allows robot to recognize cues in multiple modalities		Increases robustness through redundancy		Assists in transfer to new modalities	Allows human to use natural modalities (voice, gesture, facial expressions)

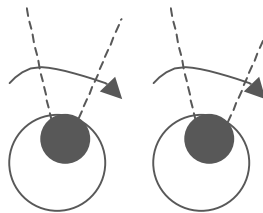
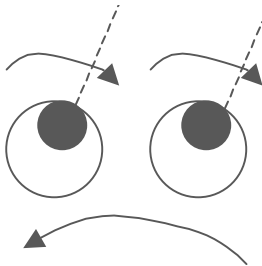
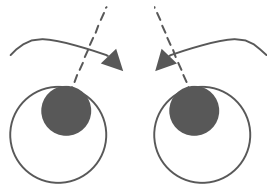
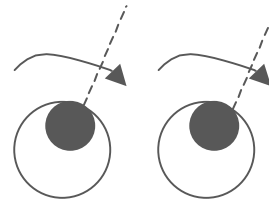


Particular Topics

- **Platform-portable vision system**
- **Understanding of self**
- **Understanding of coworkers and others**
- **Learning to communicate with coworkers**
- **Objects, navigation and interacting with others**



Integrating Full Animate Vision



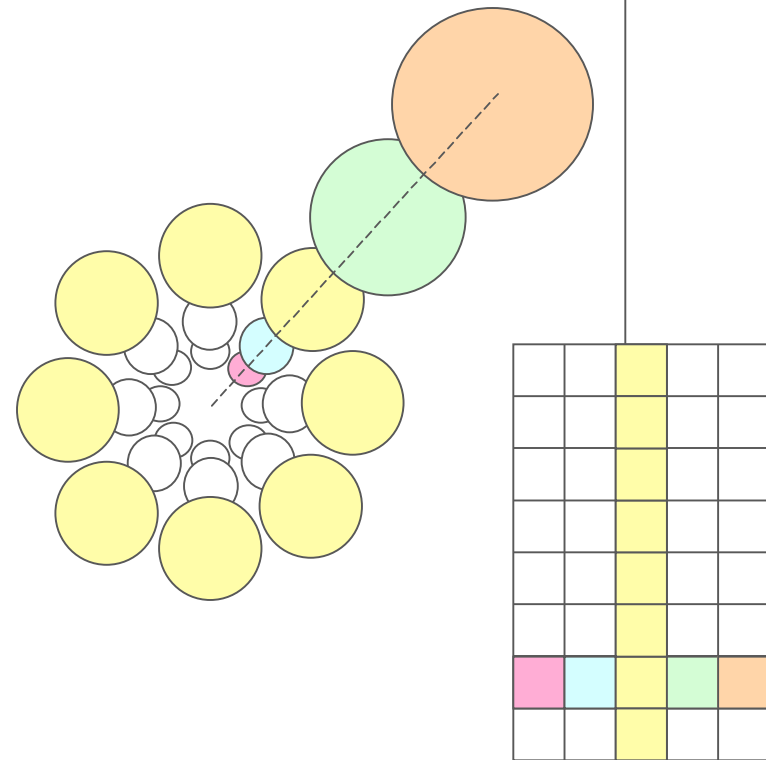
- Pursuit
- Vergence
- VOR
- Head/eyes coordination
- Saccades

We are experimenting with log-polar representations for this

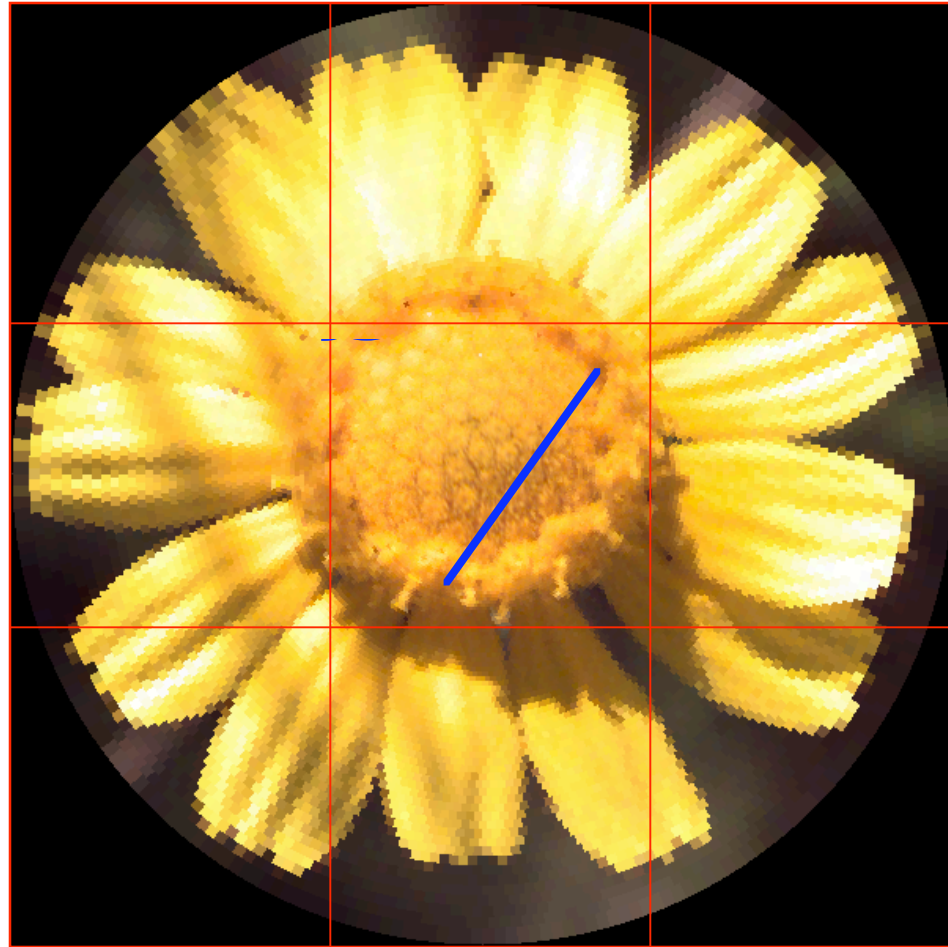
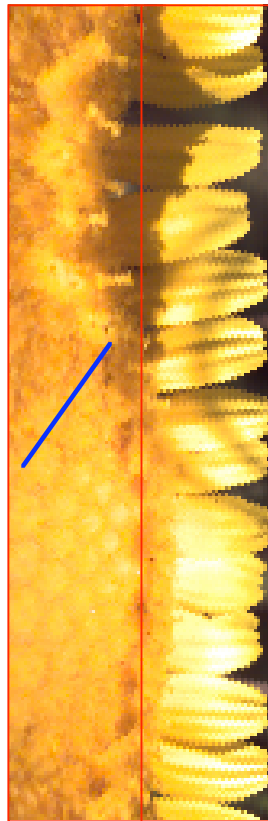


Log-Polar Transformation

- Same field of view of the rectangular image
- Reduced number of pixels
- Max resolution maintained in the fovea
- Invariant to rotation and scale... but non-invariant to translations



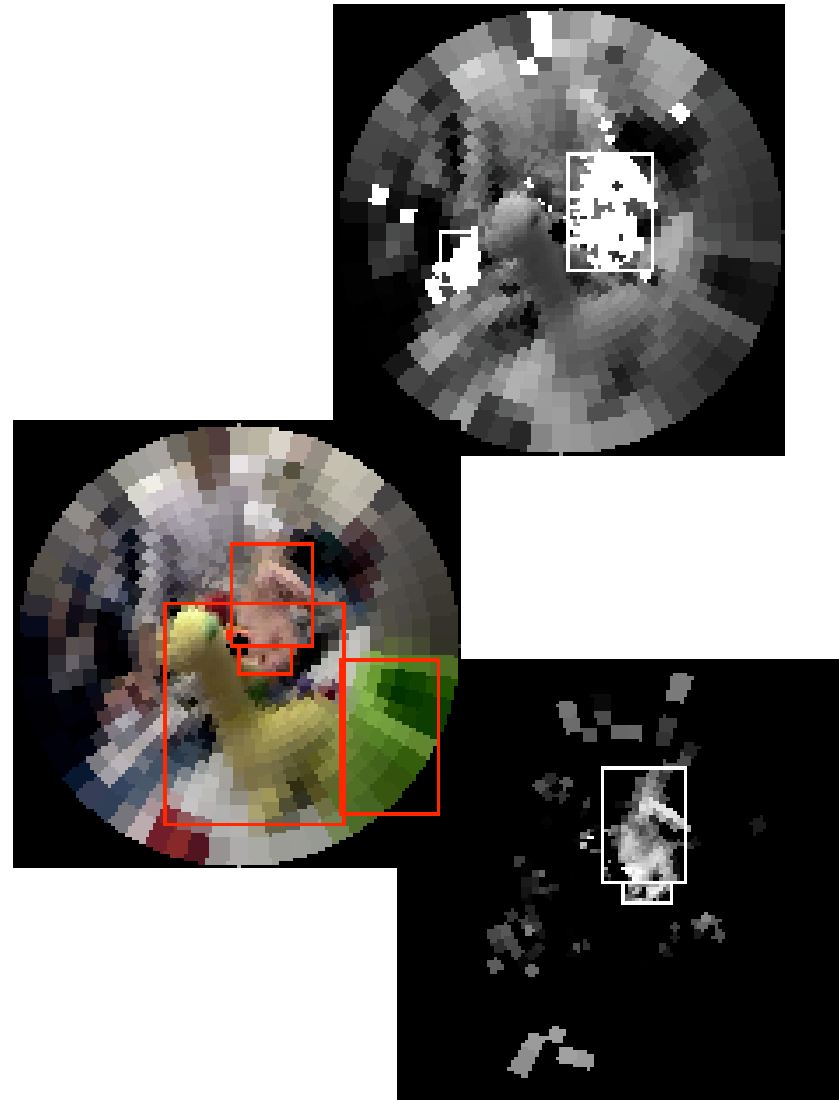
Log-polar Vision





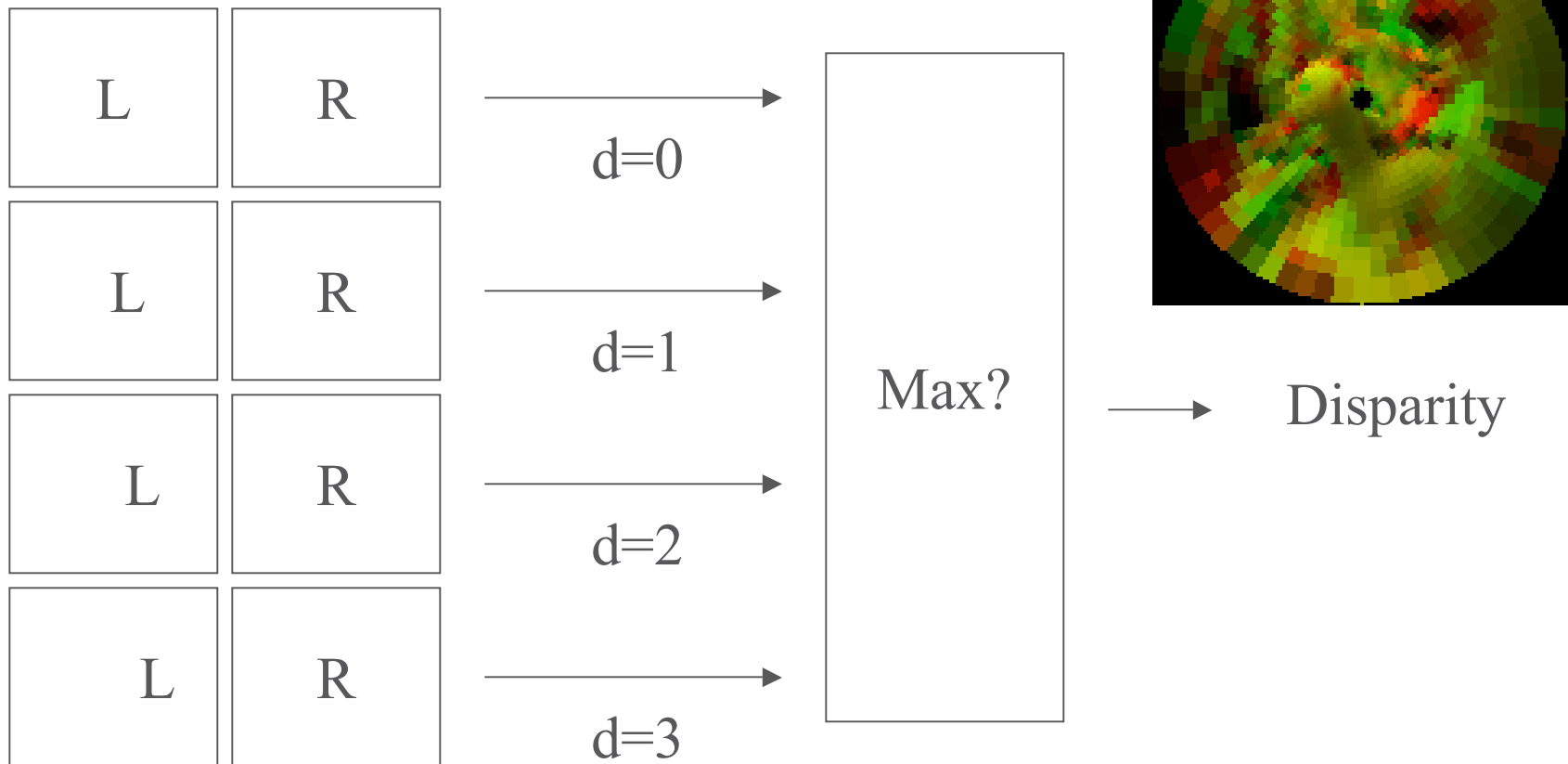
Color Processing

- Histogram based segmentation. *HSV* color space.
- Color blob detection region growing in *HS* plane.
- Skin color detector



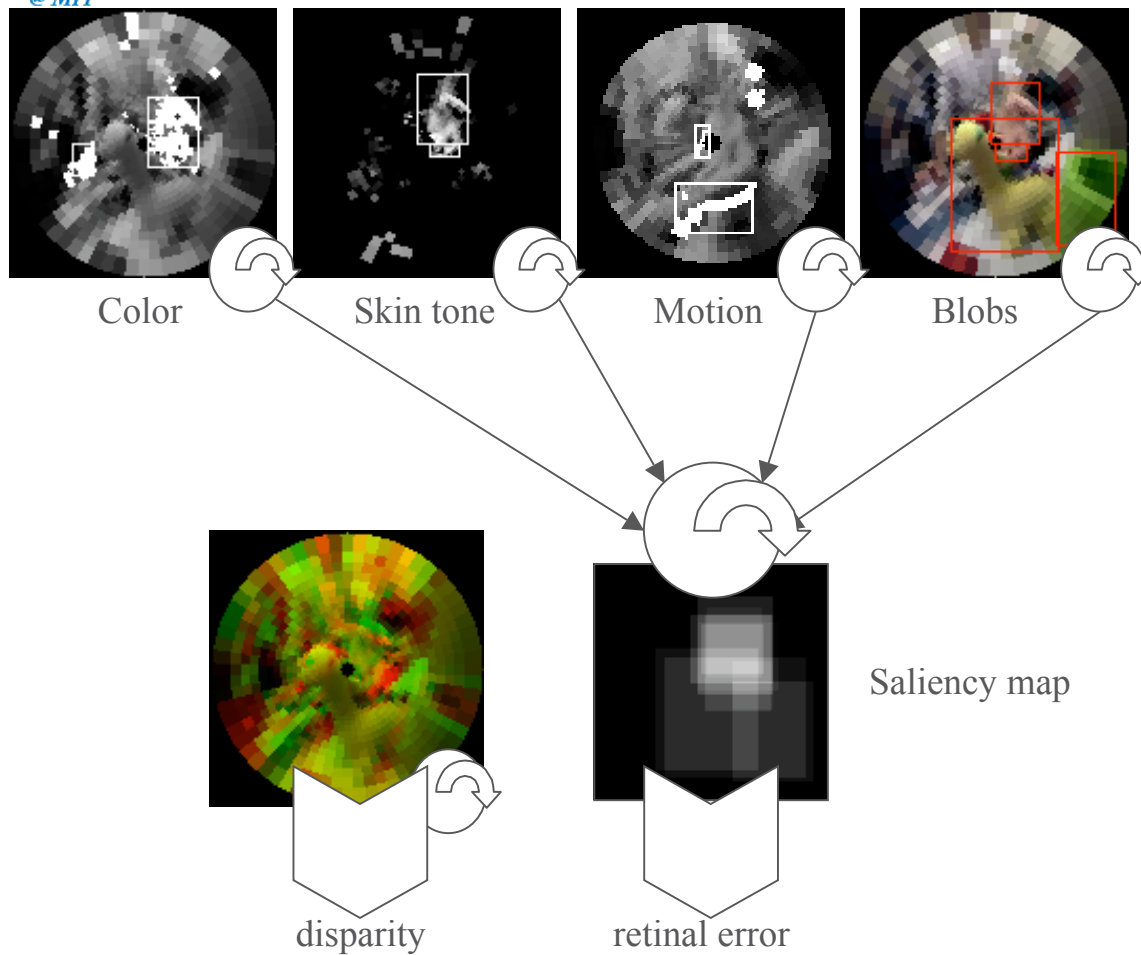


Disparity Computation





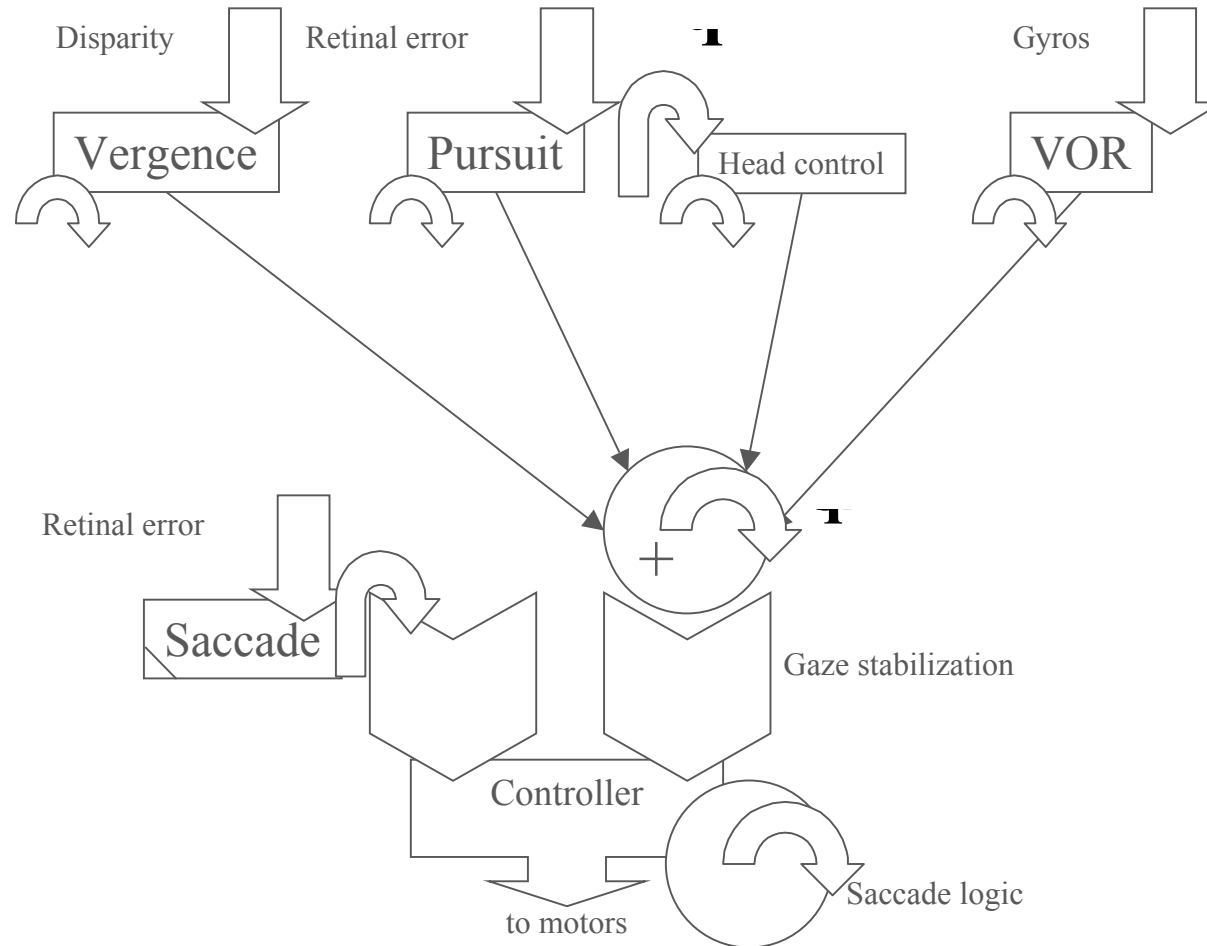
Same Sort of Attention Mechanism



- Pursuit/tracking signal (retinal error).
- Disparity signal.



Integrated Control System





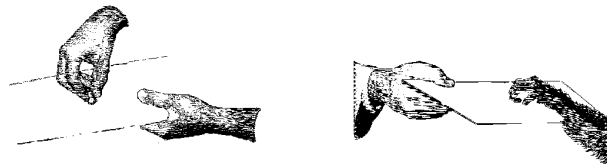
The Integrated System Operating



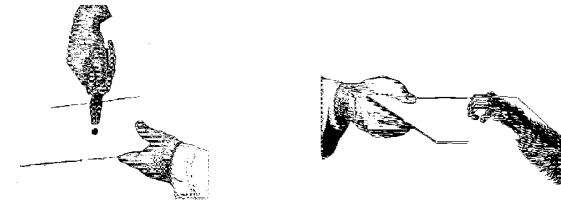
Some Neuroscience Results

Mirror neurons

A



B

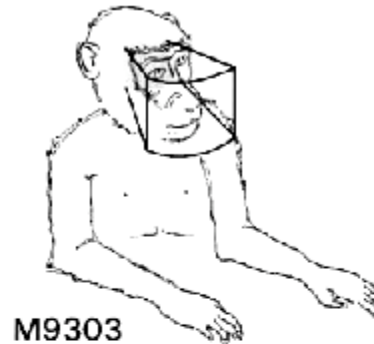


From: Fadiga, L., L. Fogassi, V. Gallese, and G. Rizzolatti, *Visuomotor neurons: ambiguity of the discharge or “motor” perception?* International Journal of Psychophysiology, 2000. **35**: p. 165-177.



F4 Neurons

- Visual, tactile, and motor responsive neurons.
- Representation of peripersonal space carved out of visual information because of motor activity.
- Receptive field linked to body part.



From: Fogassi L., Gallese V., Fadiga L., Luppino G., Matelli M., Rizzolatti G. *Coding of peripersonal space in inferior premotor cortex (area F4)*. Journal of Neurophysiology **76** (1) 1996.



The Great Leap Forward?

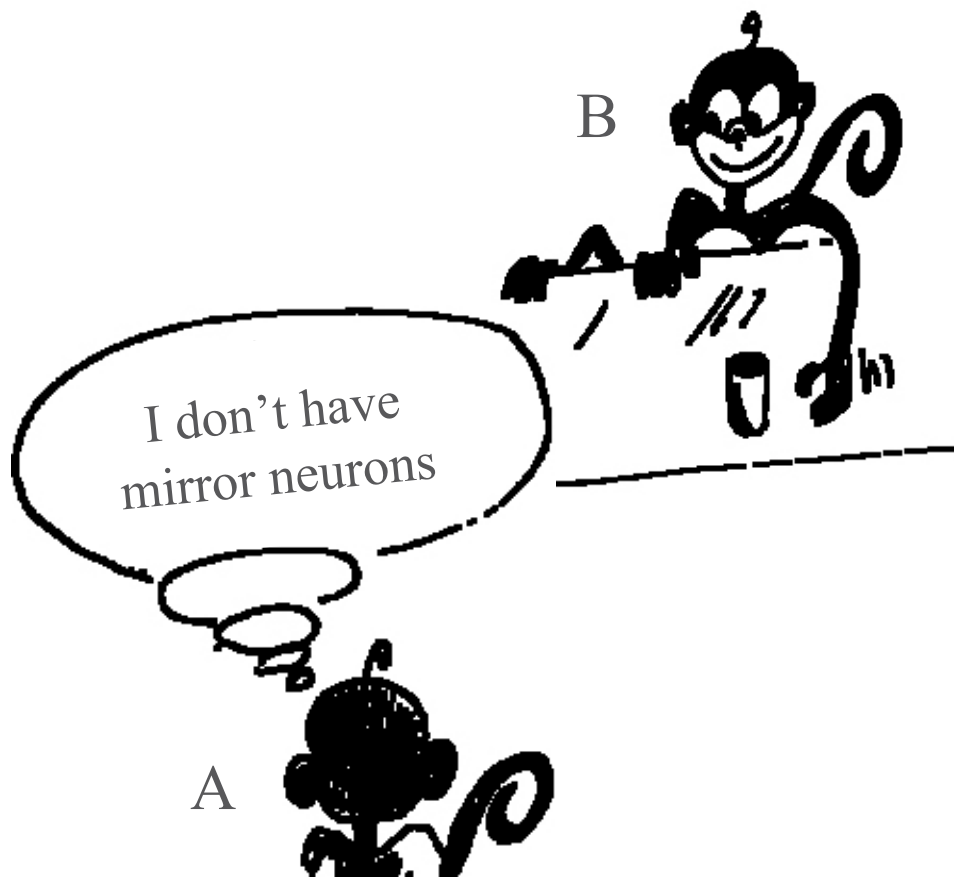
Why the sudden explosion (often called the "great leap") in technological sophistication, widespread cave art, clothes, stereotyped dwellings, etc. around 40 thousand years ago, even though the brain had achieved its present "modern" size almost a million years earlier?

A mechanism for Lamarckian learning based on mirror neurons and imitation learning... liberates us from pure gene based evolution (cultural inheritance).

From: V.S. Ramachandran, *MIRROR NEURONS and imitation learning as the driving force behind "the great leap forward" in human evolution*. Talk delivered at the Silver Jubilee meeting of the Society for Neuroscience.

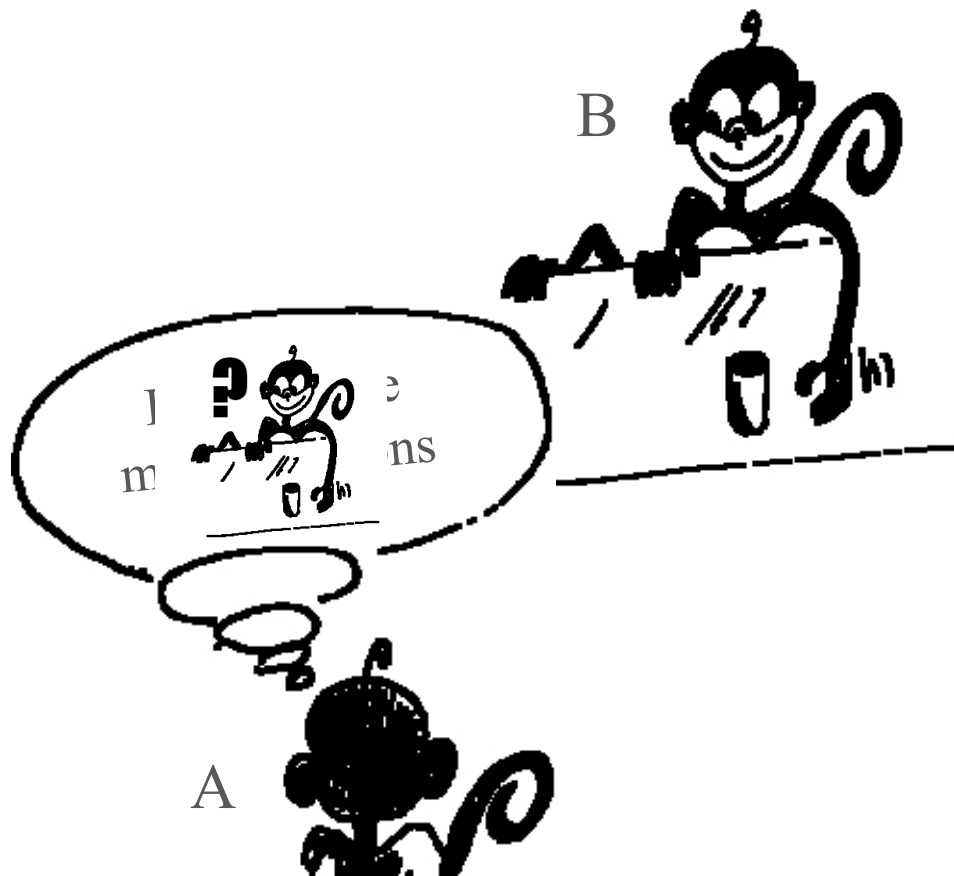


Can We Learn A “Mirror” Representation?



- Monkey A watches monkey B while grasping.
- A “incorrectly” maps the visual cues of the observed action into its own motor repertoire and tries to imitate B.
- Of course A fails (unless particularly lucky). What is then the error signal to drive the improvement of the mapping? Which are the motor primitives to map into?

Indeed We Can!



- Same situation as before.
- Monkey A notes that the action is goal-directed towards the glass.
- Monkey A remembers a case of performing a similar action towards a similar object (same goal).
- Monkey A maps the observed action into its own motor repertoire.
- The link between “A grasping” and “A watching” can be reinforced.



Main idea

Artificial

Reaching is a lot simpler if we directly map gaze direction into the motor commands required to reach the fixation point.

From: Metta, G., G. Sandini, and J. Konczak. *A Developmental Approach to Sensori-motor Coordination in Artificial Systems*. in *IEEE Conference on System, Man and Cybernetics*. 1998. San Diego (USA).

Biological

“This suggests that head-eye-hand coordination plays an important role in the organization of these movements and leads to the hypothesis that a representation of current gaze direction may serve as a reference signal for arm motor control”.

From: Flanders, M., L. Daghestani, and A. Berthoz, *Reaching beyond reach*. *Experimental Brain Research*, 1999. **126**(1): p. 19-30.

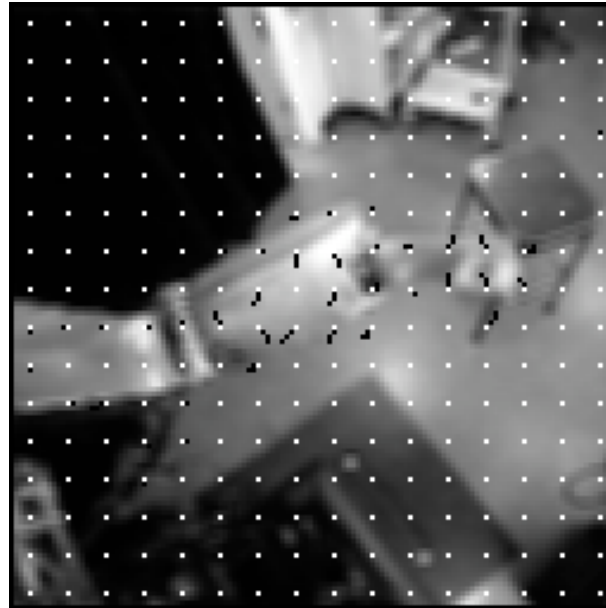


Arm Localization

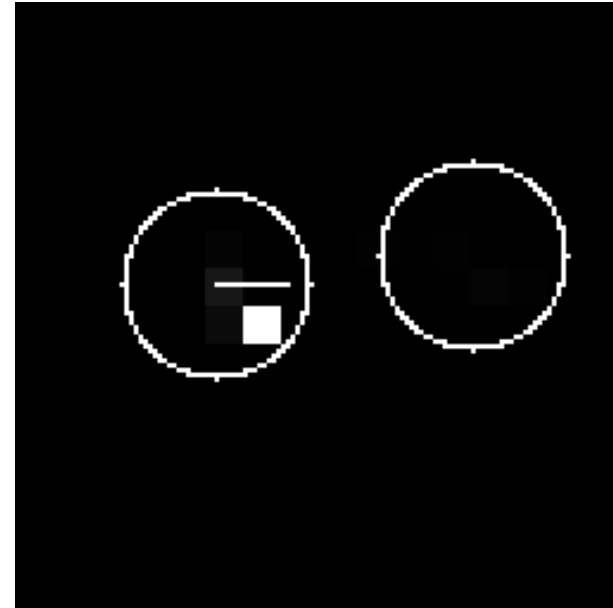
- **Correlation between vision and motor information:**
 - Visual: optical flow – Motor: velocity of arm joints.
- **Why it works:**
 - The important aspect is the temporal synchronicity.
 - Visuo-proprioceptive cues are combined.
 - The robot builds an expectation of what the optic flow would be given the known generated movement (in terms of the sign only).



A Typical Example...



Optical flow

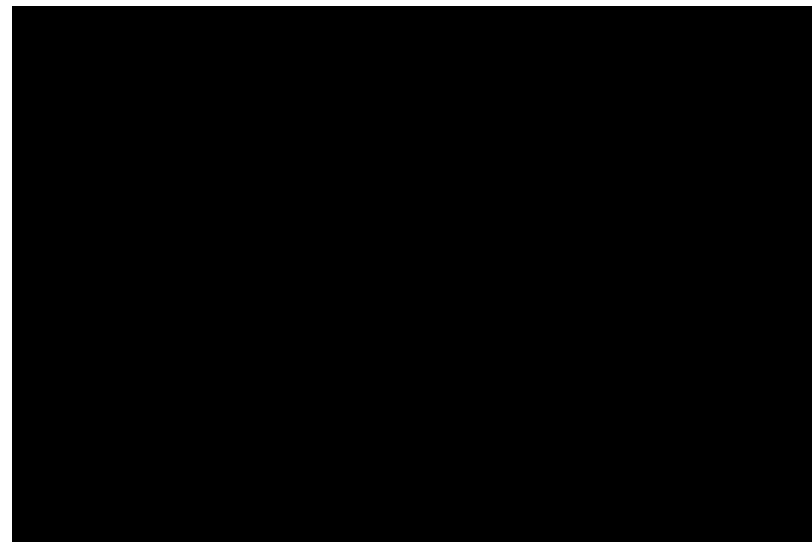
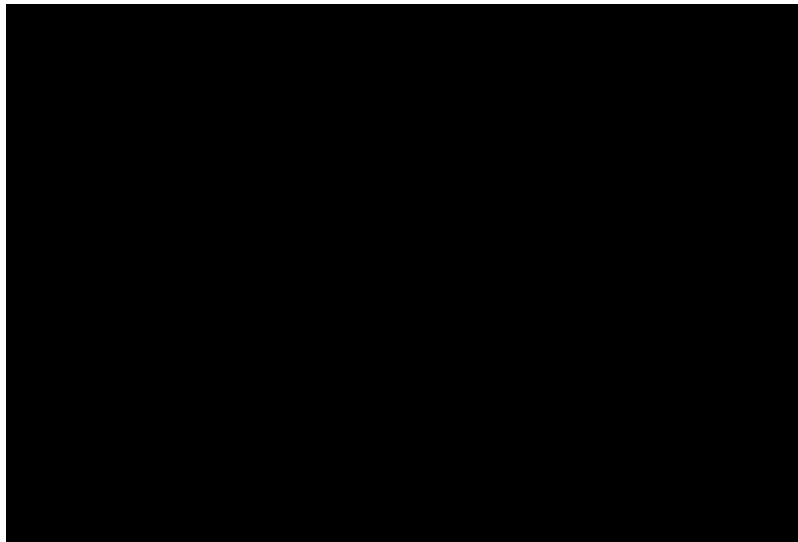


Correlation measure

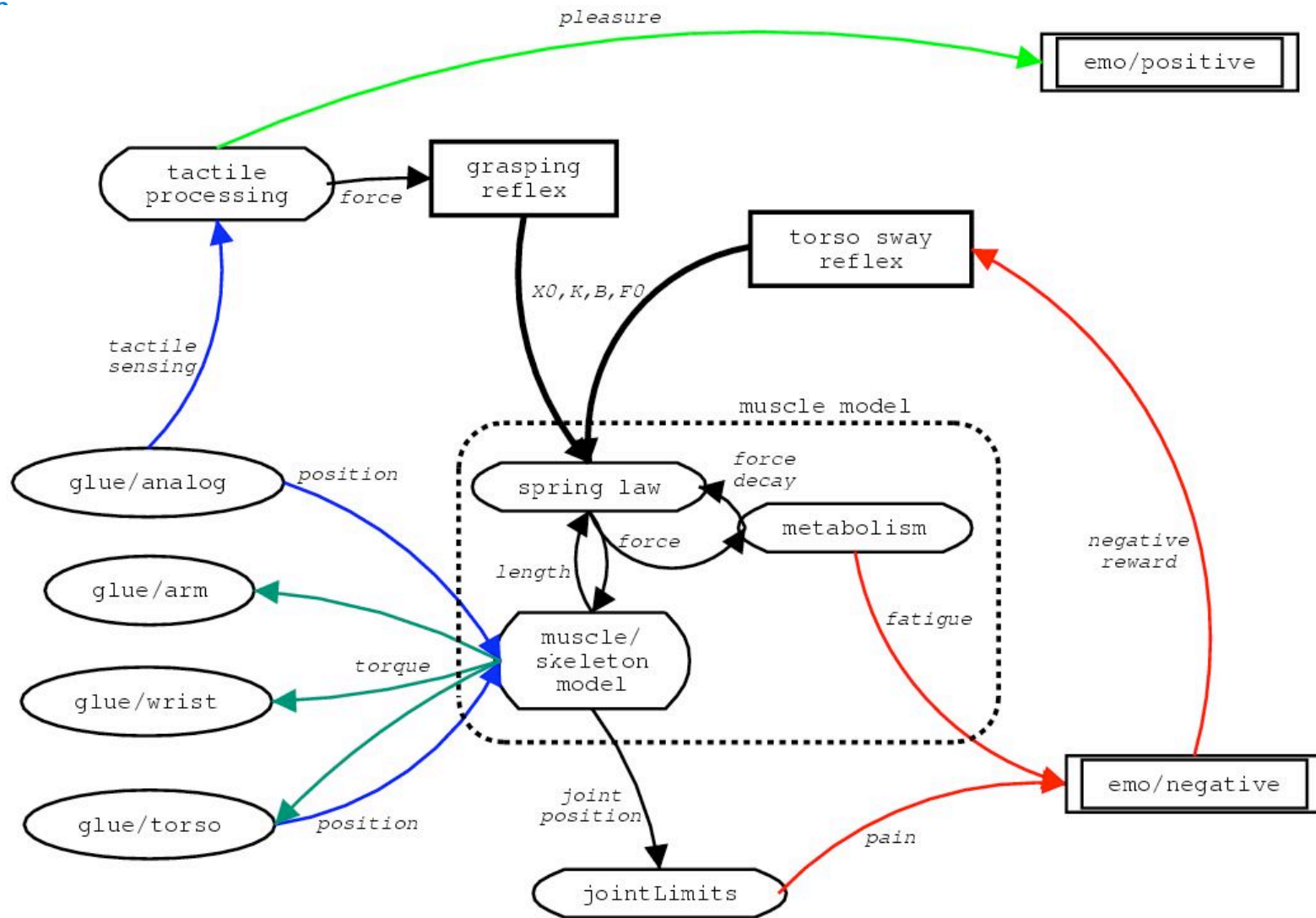
The white dot is the maximum of the correlation.



Correlation with Proprioception

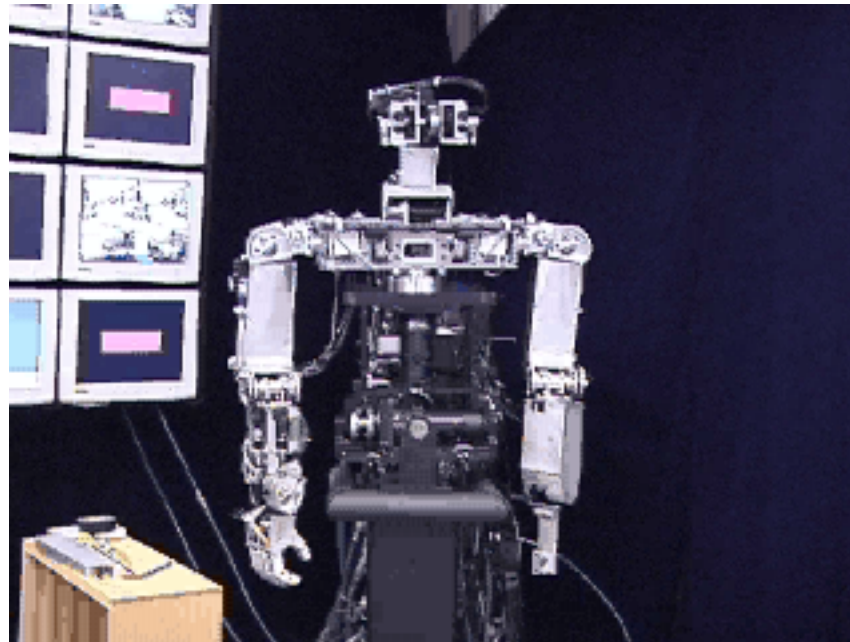


Self Awareness of Body





Learning To Sit Up Straight



No kinematic model (and therefore adaptive)
Metabolism based on energy use
Learns how to minimize energy consumption



Recognizing and Remembering Individuals



- Individual recognition is crucial for complex interactions among social animals: parental care, group breeding, cooperation, etc.
- Advantages for social robot?
 - The ability to interact with people as individuals allows complex social behavior: dislike, loyalty, cooperation, etc.
 - Many social competencies require the understanding of people as distinct individuals: theory of mind and social referencing



Individual Recognition in Nature

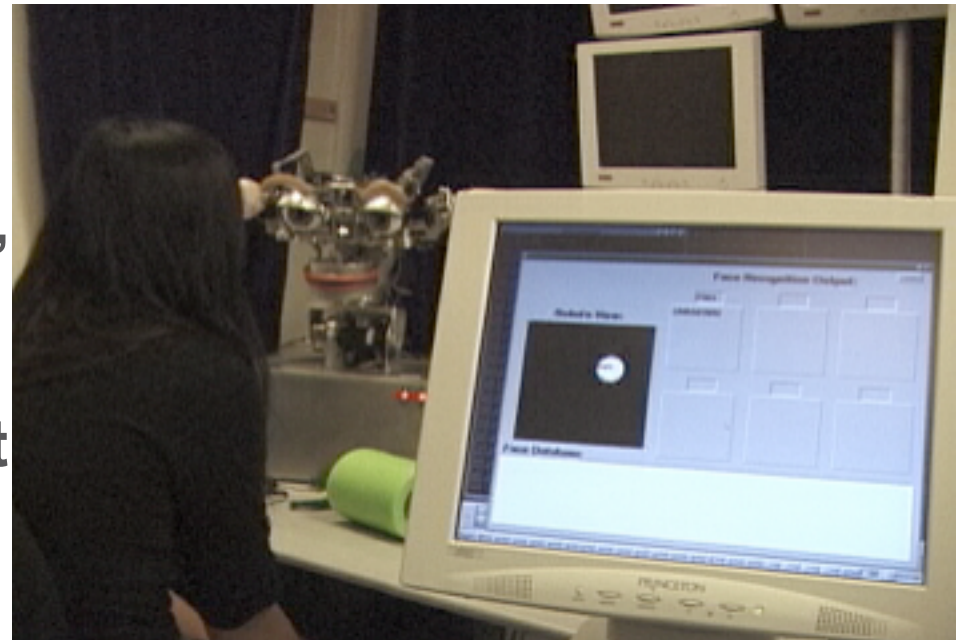
- **Mantis Shrimps**
 - avoid empty cavities with the odor of those that have defeated them in the past, but they enter cavities with the odor of those they have beaten.
 - “are able both to **recognize** other individuals and to **remember** their ‘reputation’ as fighters (Caldwell, 1986).
- **Dolphins**
 - form coalitions where each group possesses a fertile female. Often coalitions of coalitions are formed, but each sub-coalition mates only with its own female.
 - demonstrate the ability to • **recognize** friends • **keep track** of friends' past behavior • use this knowledge to predict future actions.





Individual Recognition in Humanoid Robots: Translation

- **Recognize:** person identification
 - Multi-modal Recognition: face, body, gesture, voice, etc.
- **Remember:** acquisition of contextual information about individuals
 - Appearance, past behavior, habit, belief, etc.



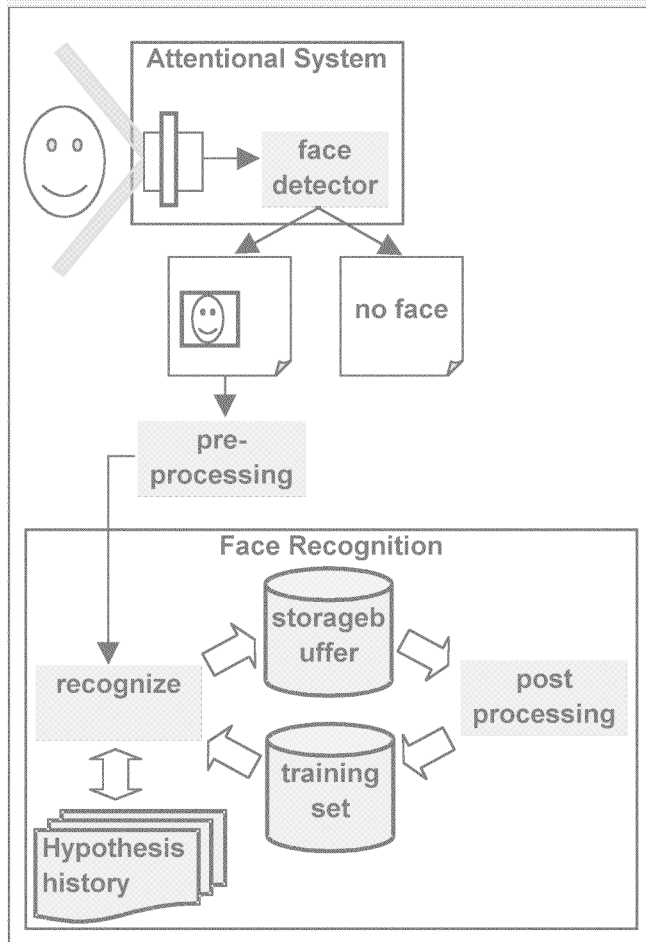


Online Unsupervised Face Recognition

- In existing face recognition technology, training images are collected, processed, and labeled manually.
- Contextual information to be learned about individuals (i.e. past behavior, habit, belief, etc) is very complex.
- Learning by manually encoding information into a database is limiting.
- The robot must learn about individuals through embodied social interaction (online and unsupervised), thus directly experiencing the richness of environment.



Online and Unsupervised Face Recognition

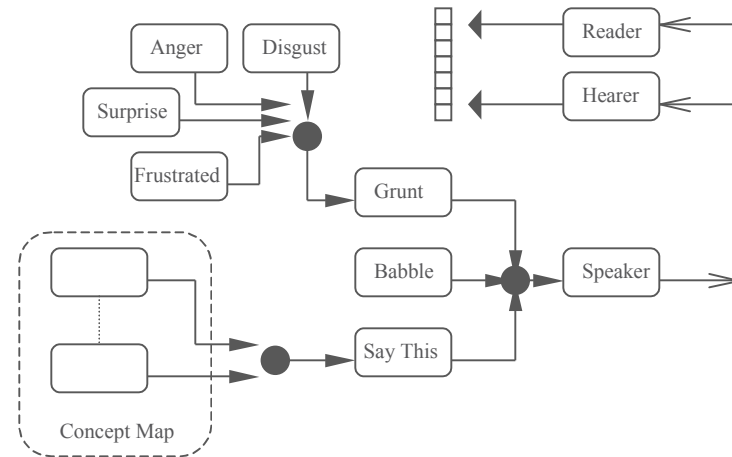


- **Receives input directly from robot's camera**
- **Uses a robust face detector developed by (Viola and Jones, 2001)**
- **Opportunistically collects, labels, and learns the faces of people interacting with the robot, starting with an empty database**
- **Uses eigenface technique (Turk and Pentland, 1991)**

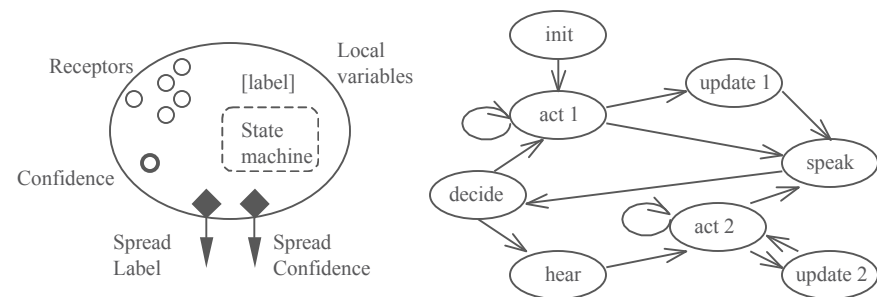


Protoverbal Behaviors on Kismet

- **Pragmatic language/concept acquisition**
 - not just hypothesis refinement in external sensor space
 - rather, includes the internal states that the external objects and behaviors engender in the robot



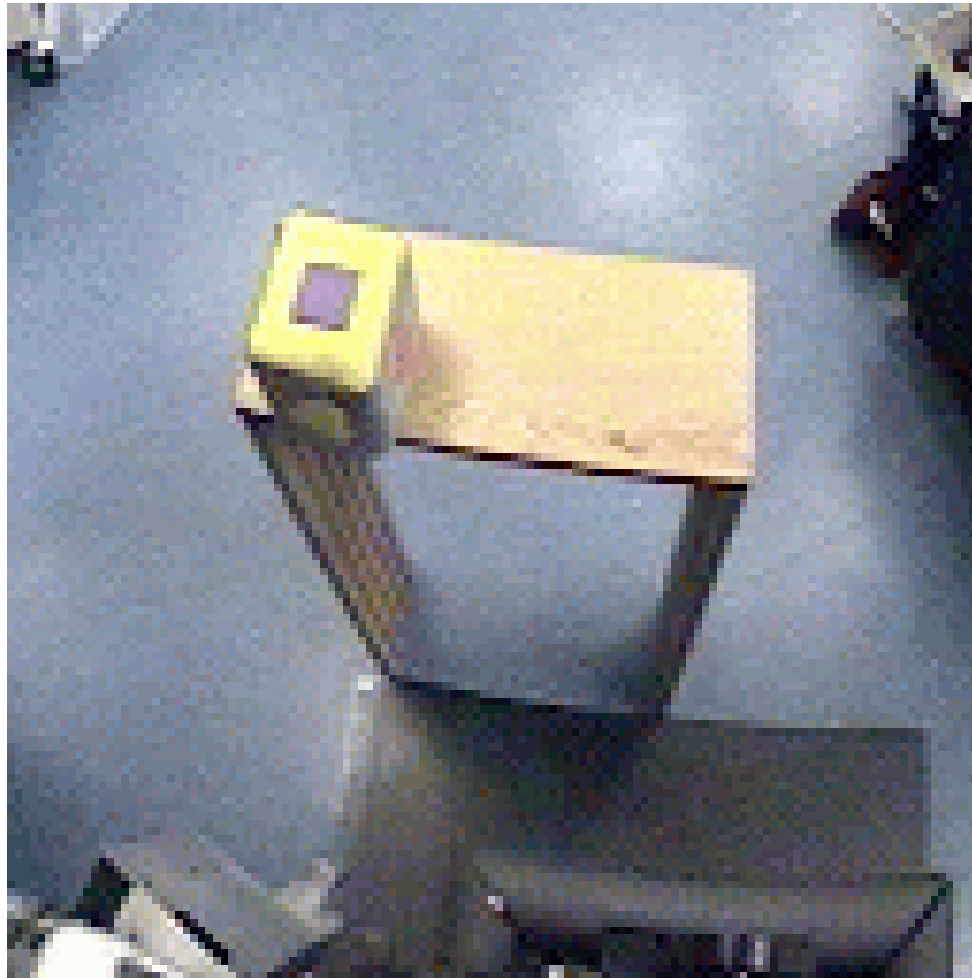
Architecture of protoverbal behaviors



Design of a single protoverbal behavior



A Simple Scene?



February 26, 2002

DARPA Mobile Autonomous Robot Software BAA99-09

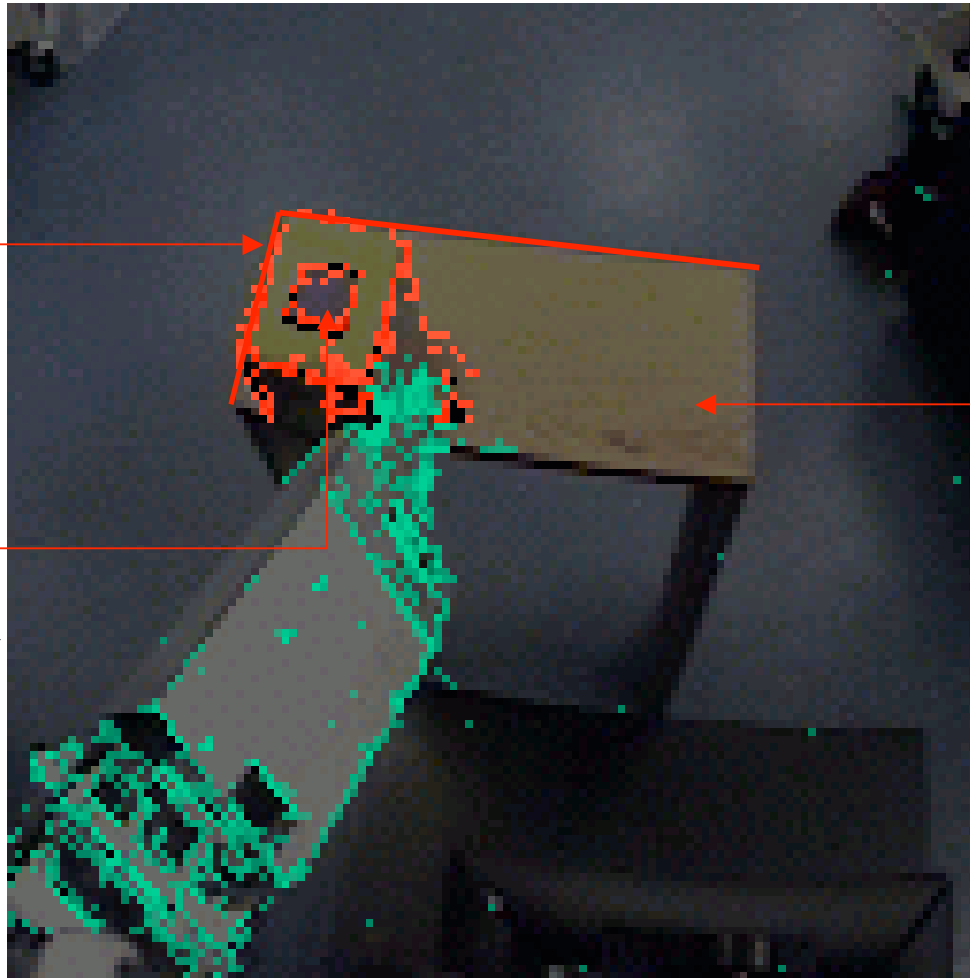


A Simple Scene?

Edges of table
and cube
overlap

Cube has
misleading
surface pattern

Color of cube
and table are
poorly separated





Active Segmentation



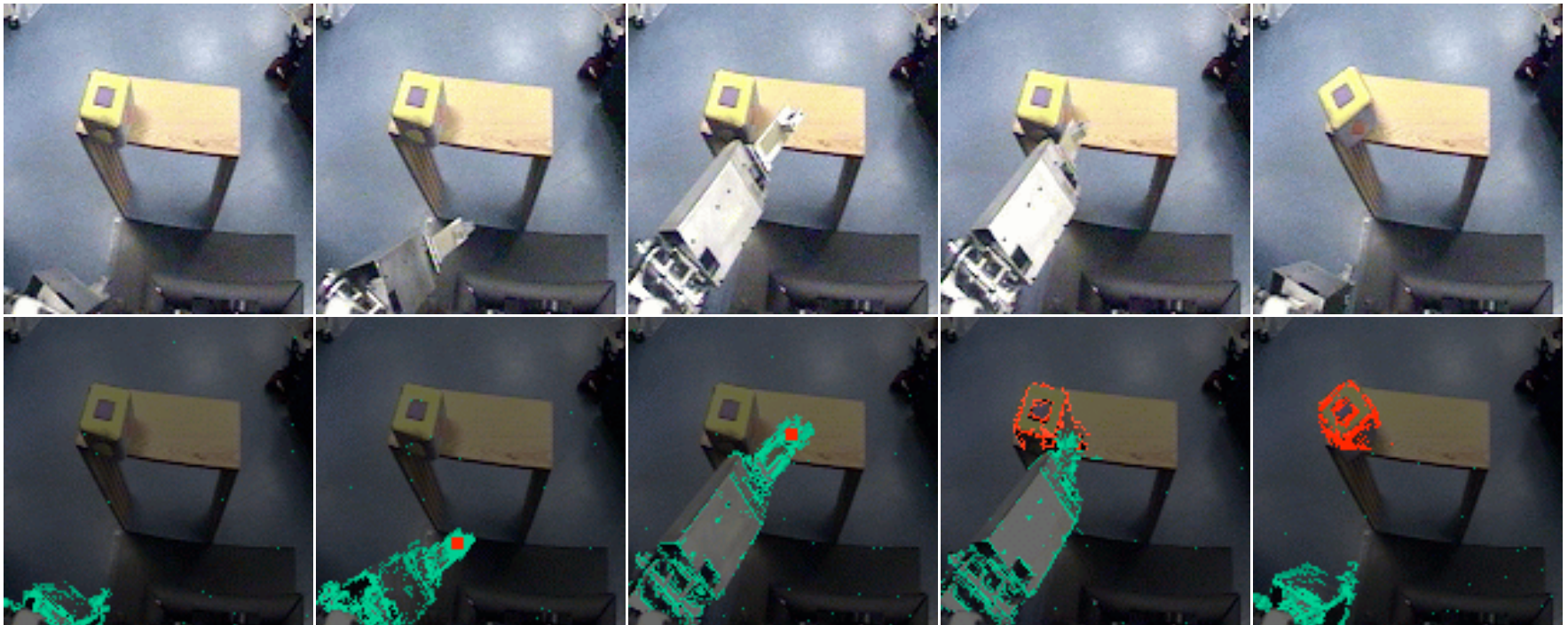
February 26, 2002



DARPA Mobile Autonomous Robot Software BAA99-09



Anatomy of a Poke



Begin

Find hand

Sweep

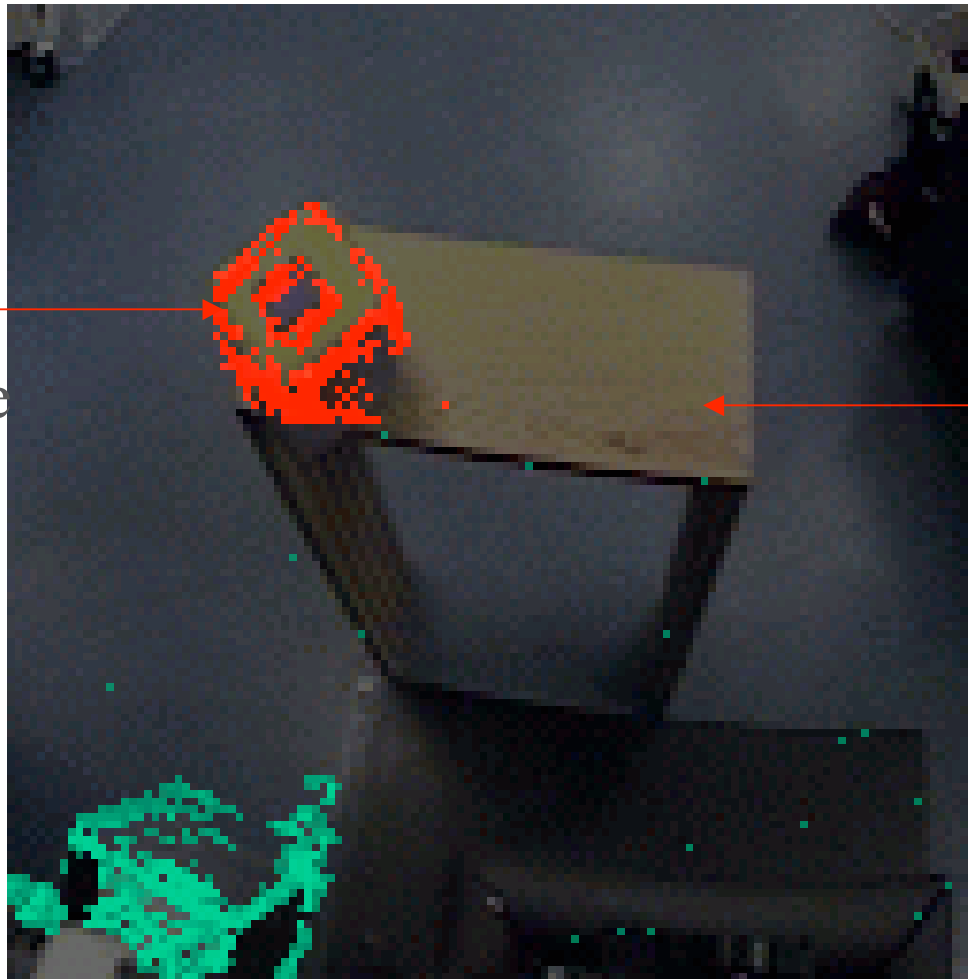
Contact!

Withdraw



Result

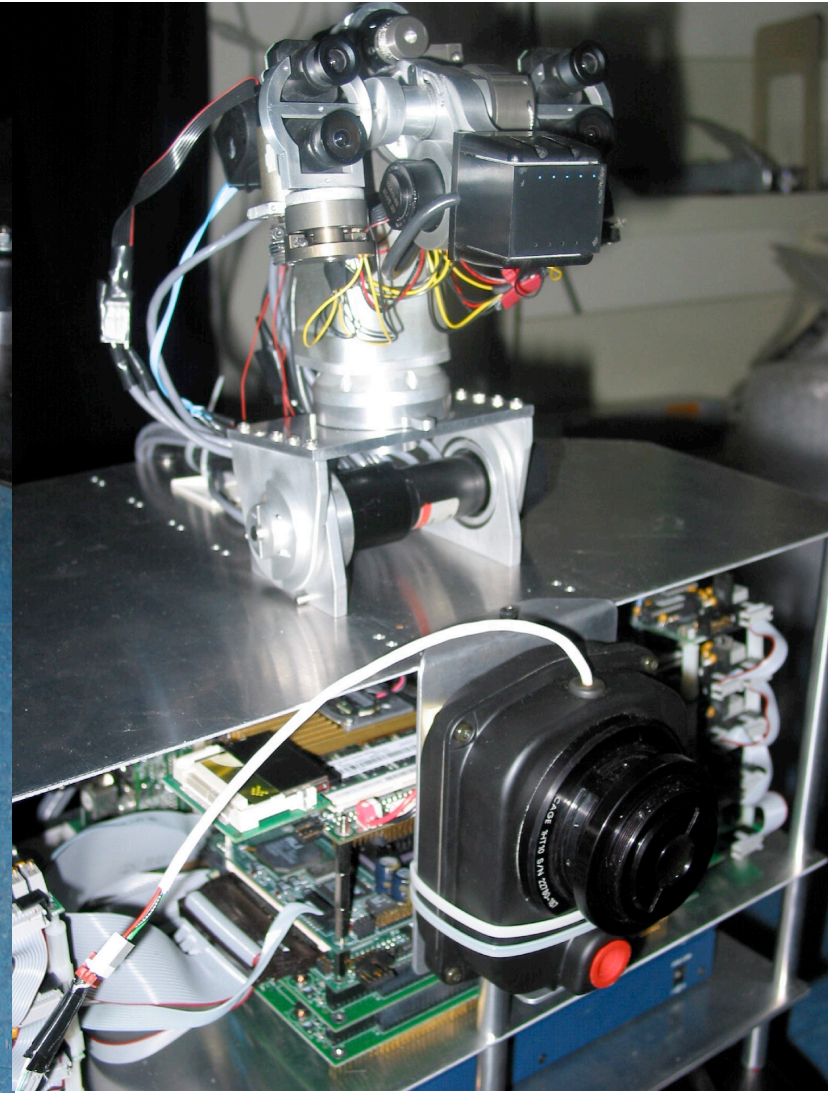
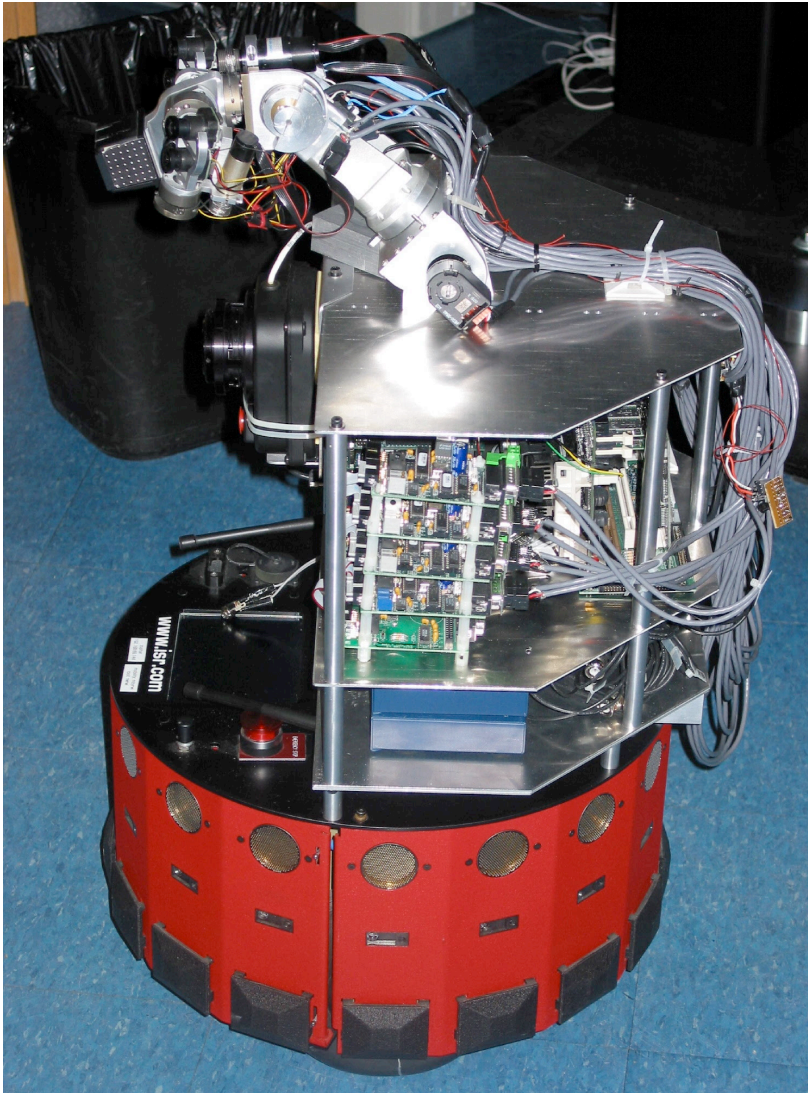
No confusion
between cube
and own texture



No confusion
between cube
and table



Combining Navigation

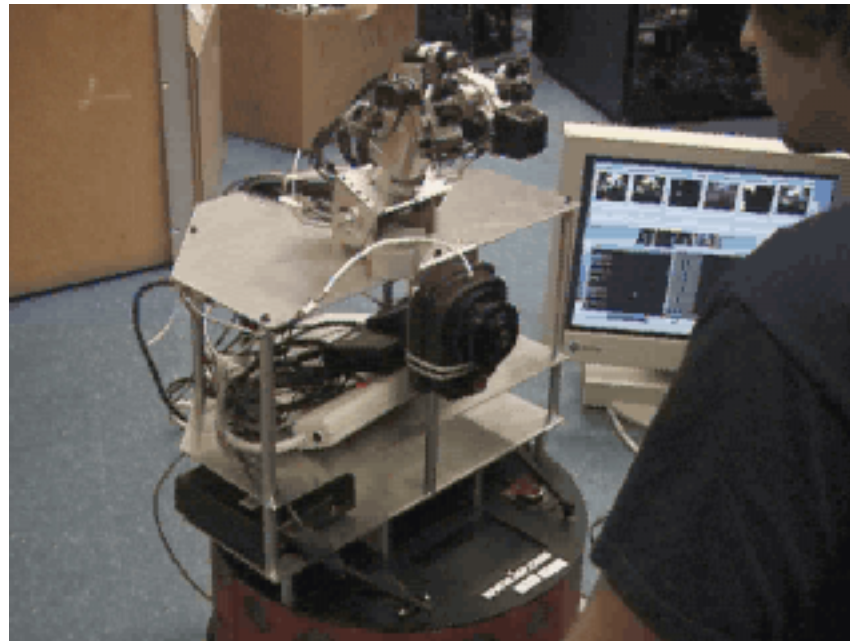


February 26, 2002

DARPA Mobile Autonomous Robot Software BAA99-09



Early Results



February 26, 2002

DARPA Mobile Autonomous Robot Software BAA99-09