Guest lecture: Visually Interpreting Human Actions for Intelligent Agent

Yezhou Yang
yz.yang@asu.edu

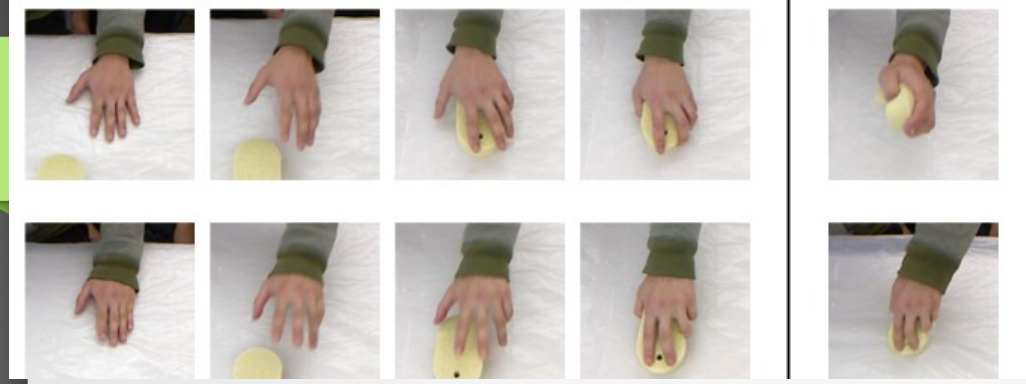
Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY

school of computing, informatics, & decision systems engineering

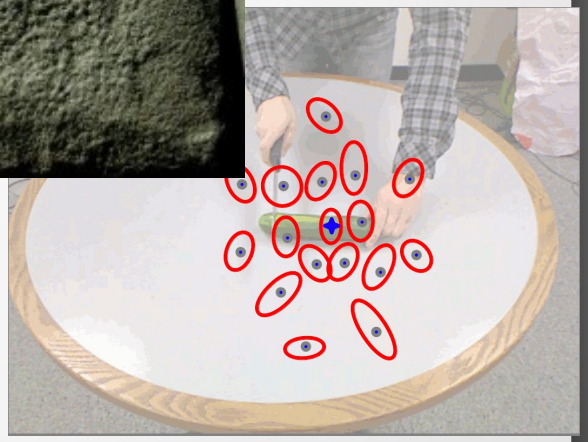
Prof. Kanade from CMU on Robotics System

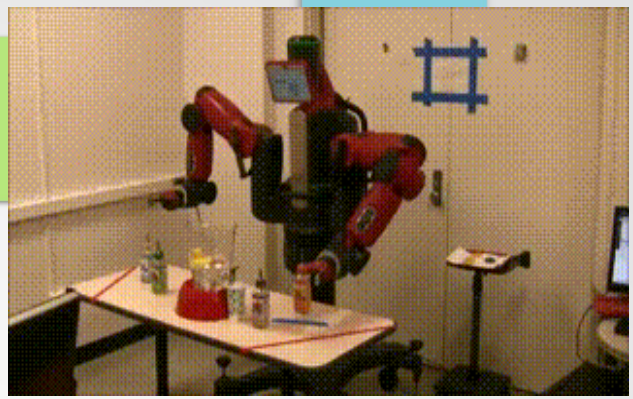
- Robotics/AI research is fun to do!
- You can always find something interesting to you (hopefully...) in Robotics.
- Roller Coaster
- Perception is (one of) the bottlenecks...
- … … ...

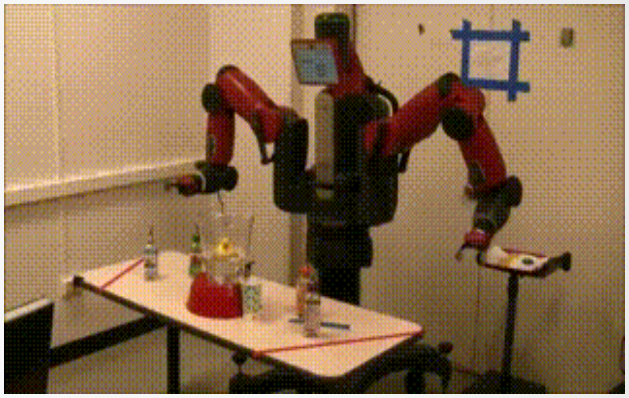
preparation

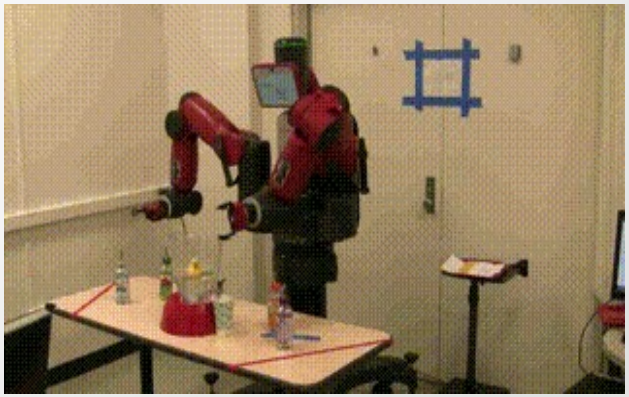


Power			Precision		
Cylindrical	Spherical	Hook	Pinch	Tripod	Lumbrical





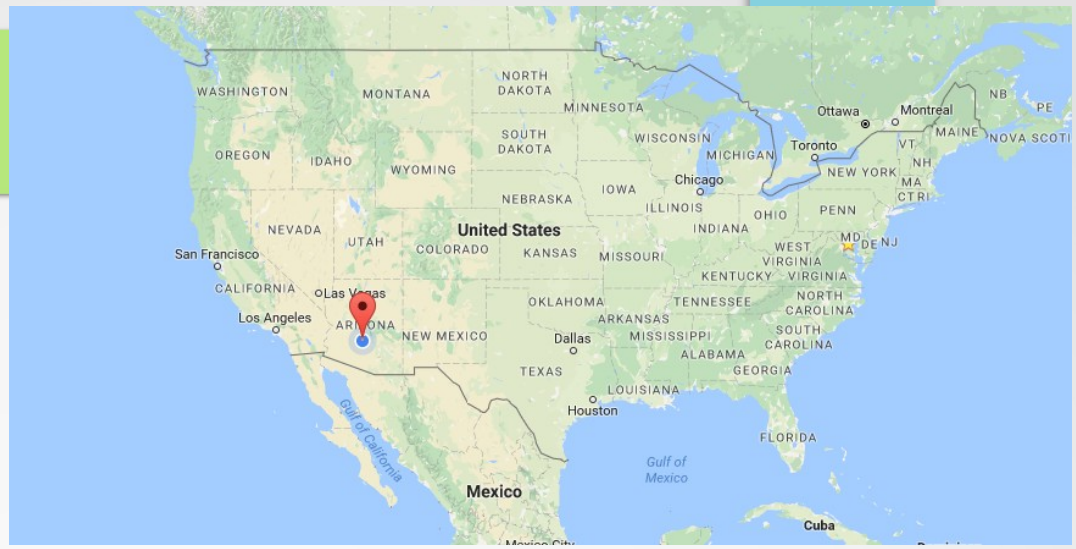


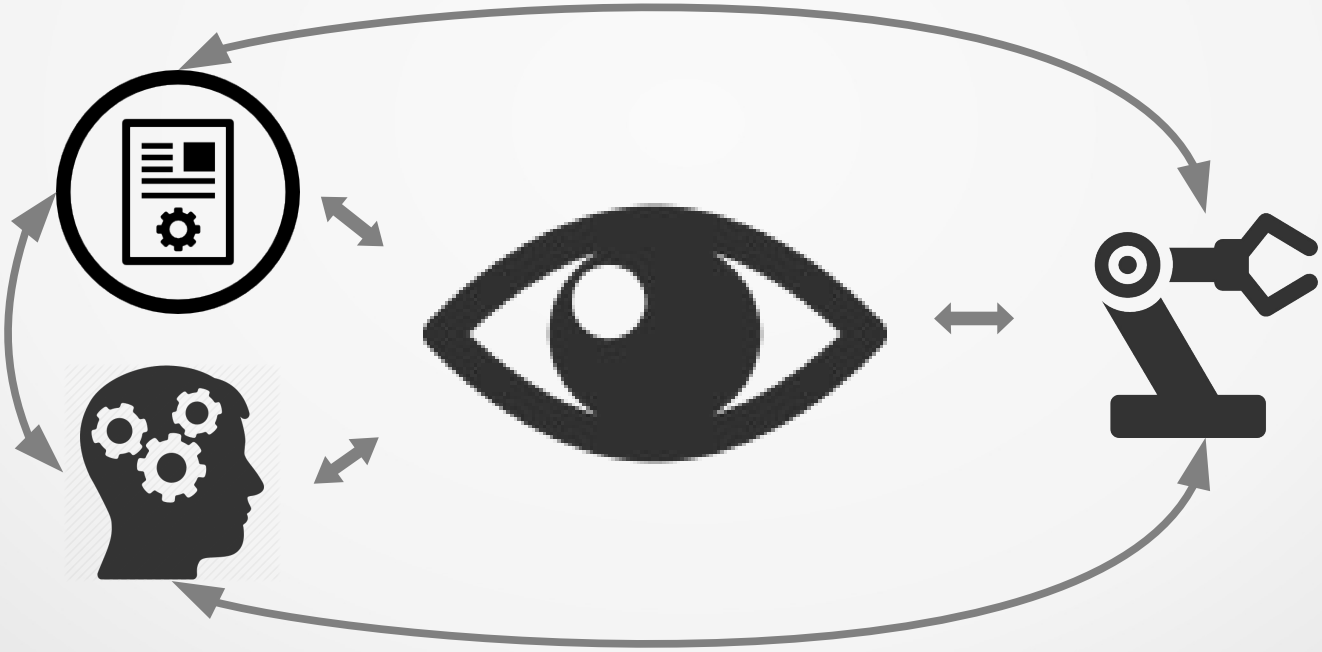


Manipulation actions

Ira A. Fulton Schools of Engineering
ARIZONA STATE UNIVERSITY

Assistant Prof.
2016-





Active Perception Group

Human Manipulation Action Understanding for Cognitive Robots

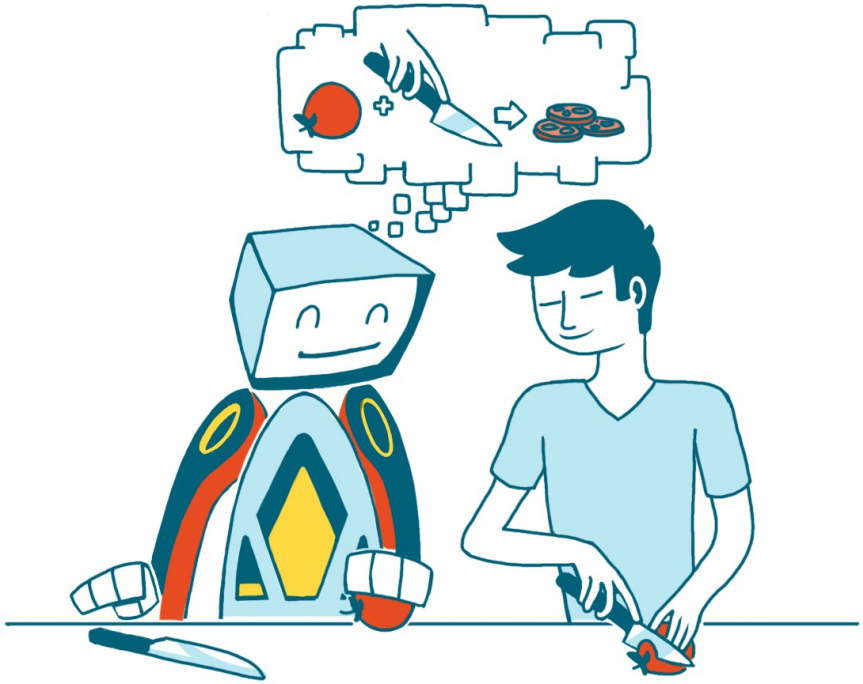
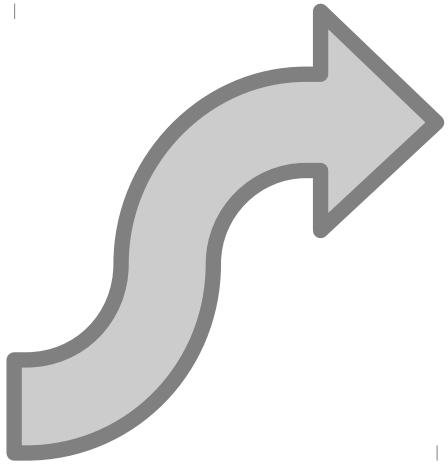
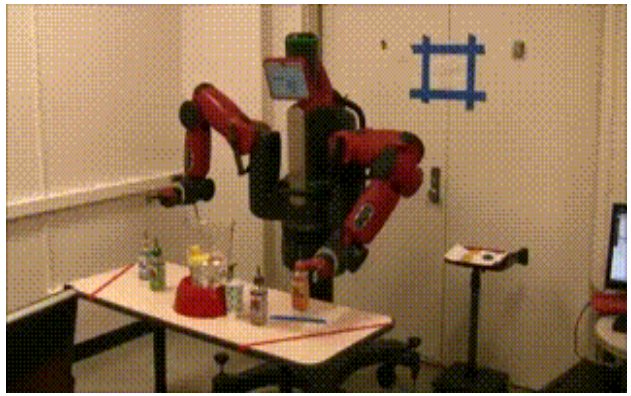
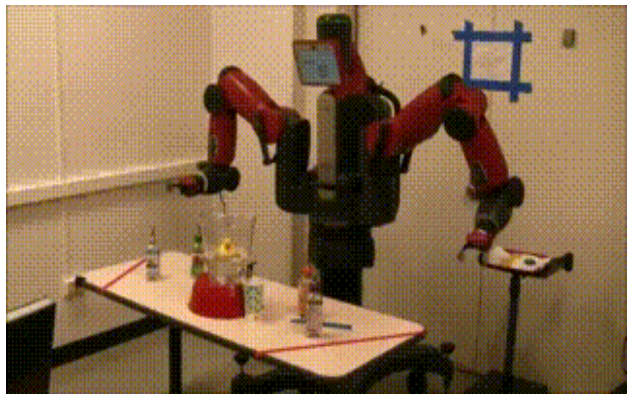
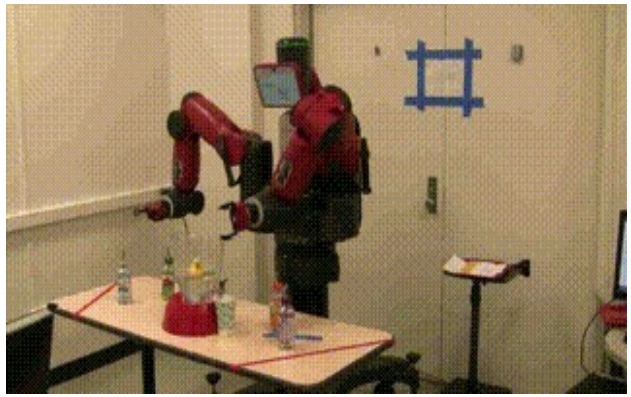


Illustration by Austin Myers









Manipulation actions

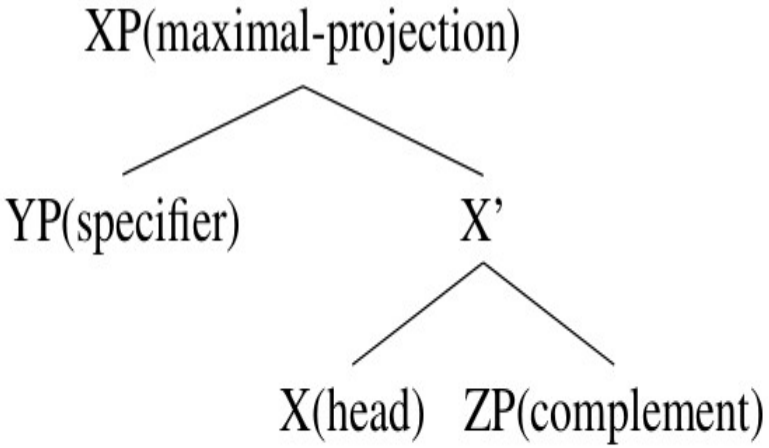
Theoretical Foundations

Cognitive Penetration

Vision shall be studied with cognitive context, and the top-down influence can even penetrate to the early vision stages.

Vetter, Petra, and Albert Newen. "Varieties of cognitive penetration in visual perception." Consciousness and cognition 27 (2014): 62-75.

Action Grammar

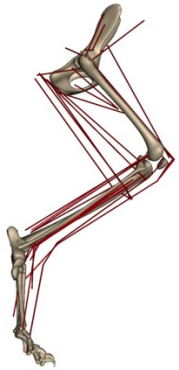


X-bar theory was first proposed by Noam Chomsky (1970), building on Zellig Harris's 1951 approach to categories and further developed by Ray Jackendoff (1977).

"The minimalist grammar of action." Pastra, Katerina, and Yiannis Aloimonos. Philosophical Transactions of the Royal Society of London B: Biological Sciences 367.1585 (2012)

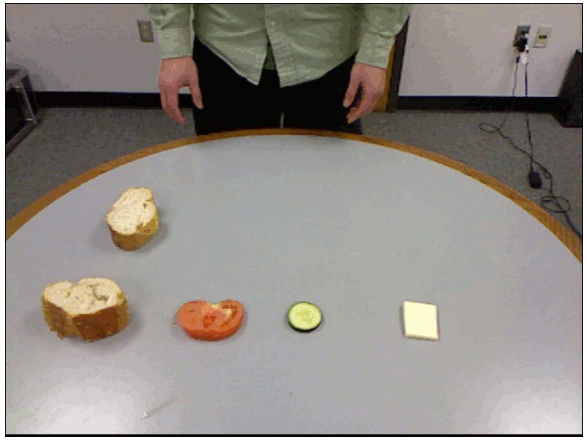
Disruption of Broca's Area Alters Higher-order Chunking Processing during Perceptual Sequence Learning
Luciano Fadiga et.al. From IIT, Journal of Cognitive Neuroscience 2016

Degrees of freedom problem (motor control)

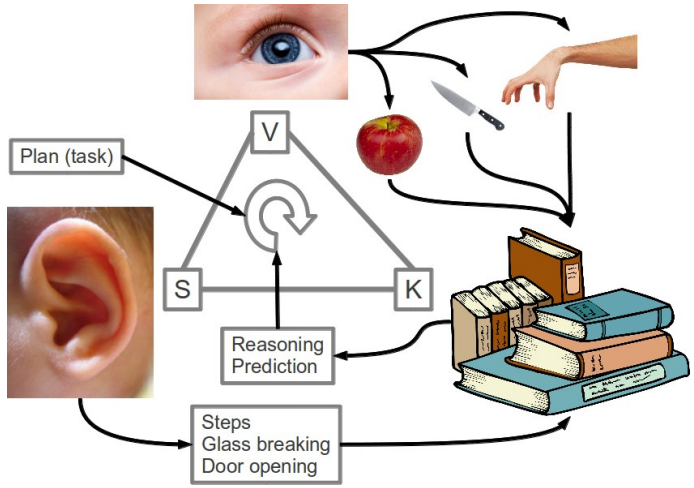


Perception-Action integration (Visual Grounding)

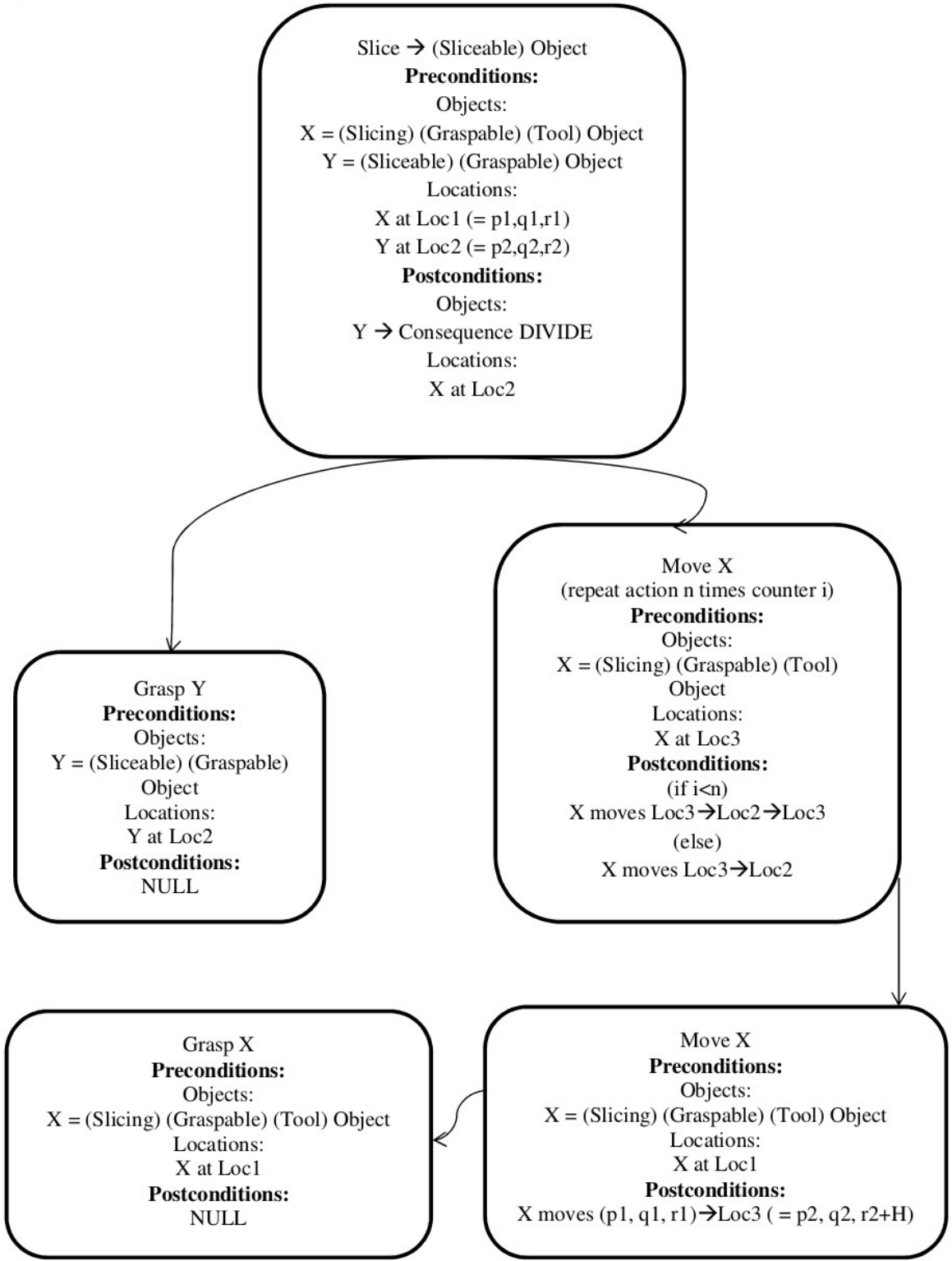
Action Sequencing



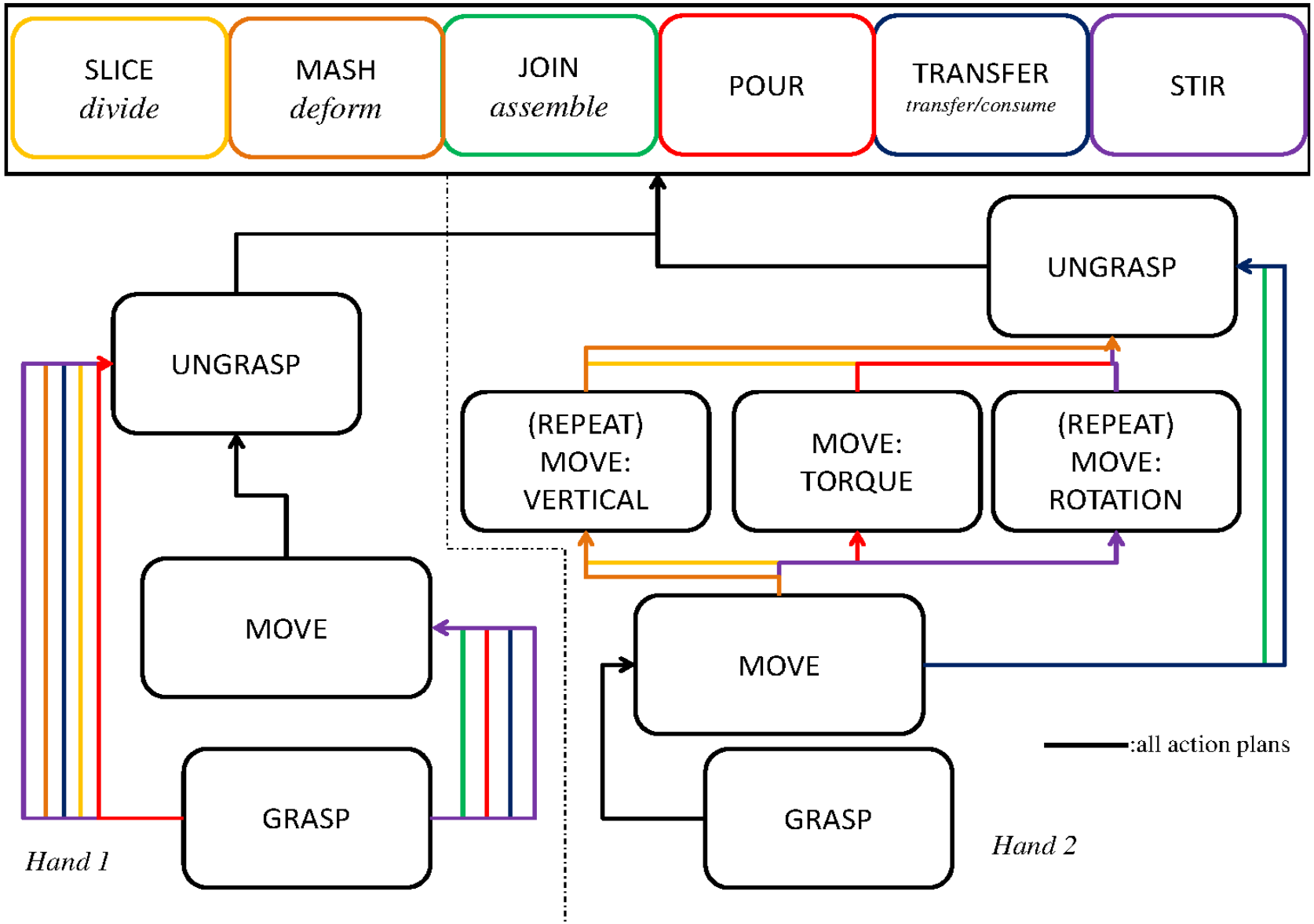
Learning



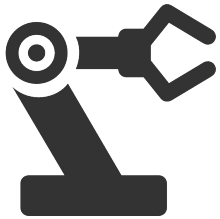
1. SLICE


Slice → (Sliceable) Object
Preconditions:
Objects:
X = (Slicing) (Graspable) (Tool) Object
Y = (Sliceable) (Graspable) Object
Locations:
X at Loc1 (= p1,q1,r1)
Y at Loc2 (= p2,q2,r2)
Postconditions:
Objects:
Y → Consequence DIVIDE
Locations:
X at Loc2
Grasp Y
Preconditions:
Objects:
Y = (Sliceable) (Graspable) Object
Locations:
Y at Loc2
Postconditions:
NULL
Move X
(repeat action n times counter i)
Preconditions:
Objects:
X = (Slicing) (Graspable) (Tool) Object
Locations:
X at Loc3
Postconditions:
(if i<n)
X moves Loc3→Loc2→Loc3
(else)
X moves Loc3→Loc2
Grasp X
Preconditions:
Objects:
X = (Slicing) (Graspable) (Tool) Object
Locations:
X at Loc1
Postconditions:
NULL
Move X
Preconditions:
Objects:
X = (Slicing) (Graspable) (Tool) Object
Locations:
X at Loc1
Postconditions:
X moves (p1, q1, r1)→Loc3 (= p2, q2, r2+H)

Action Primitives



ICRA 2012, IROS 2013

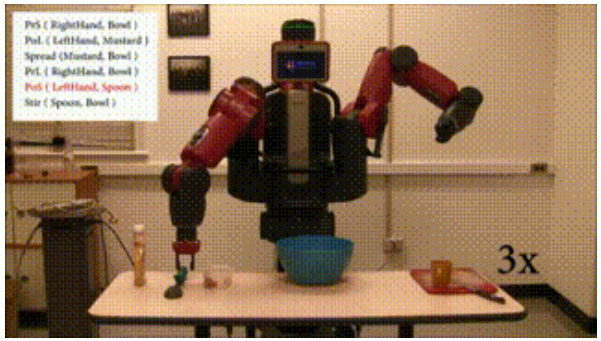


Robotic Execution

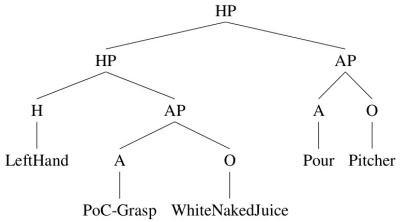
Action Sequencing

Visual Grounding

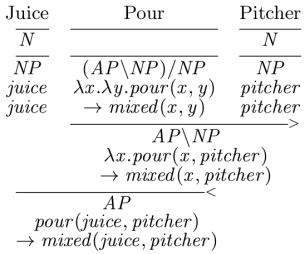
Sensory Input



ongoing



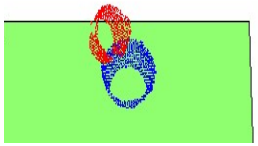
CogSys 14, AAAI 15



ACL 15, ongoing

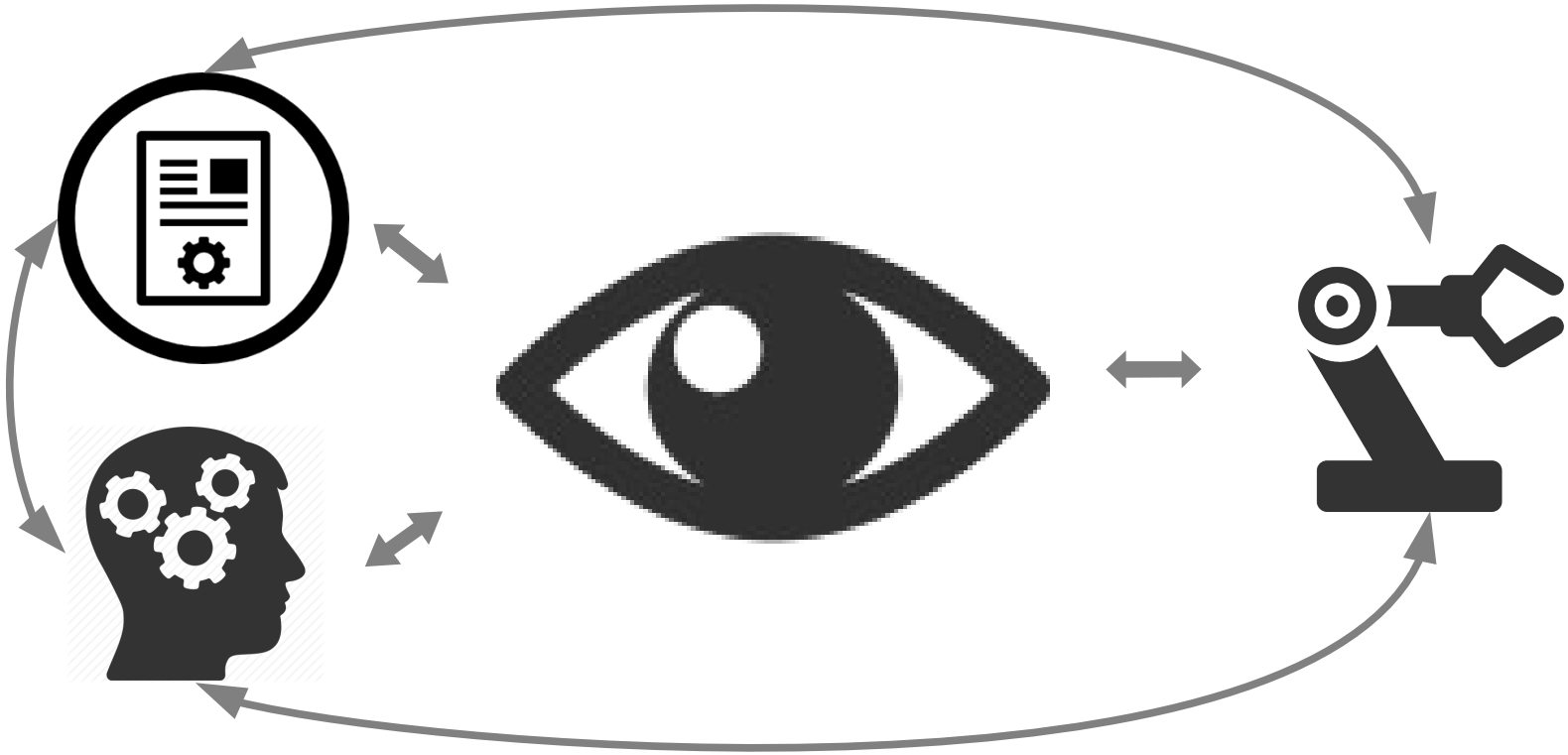
CVPR 15 , ongoing

CVPR 13



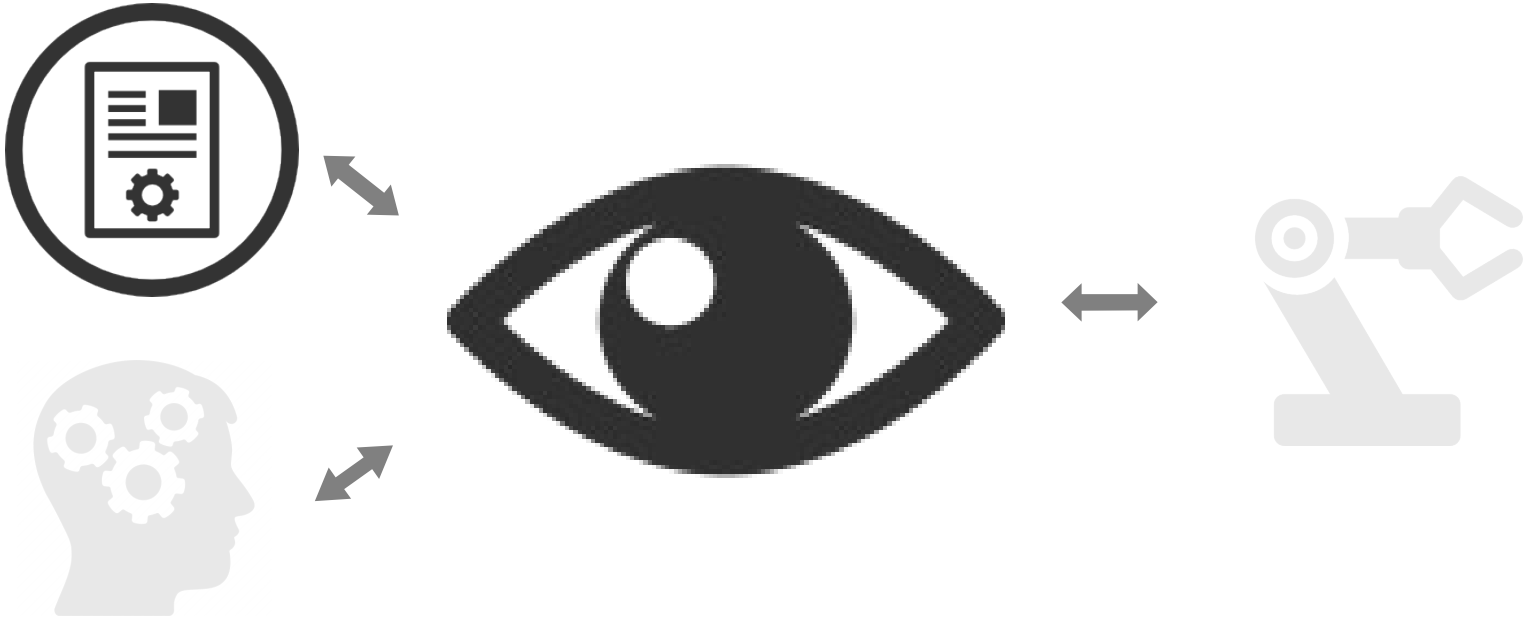
ICRA 15

Active Vision: A Modern Perspective

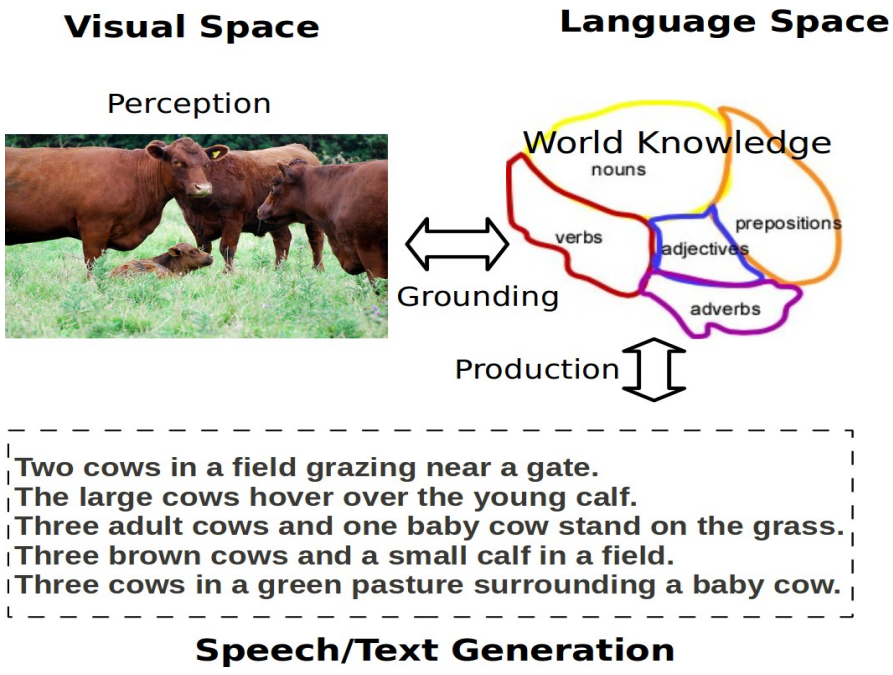


"...An agent is an active perceiver if it knows why it wishes to sense, and then chooses what to perceive, and determines how, when and where to achieve that perception."

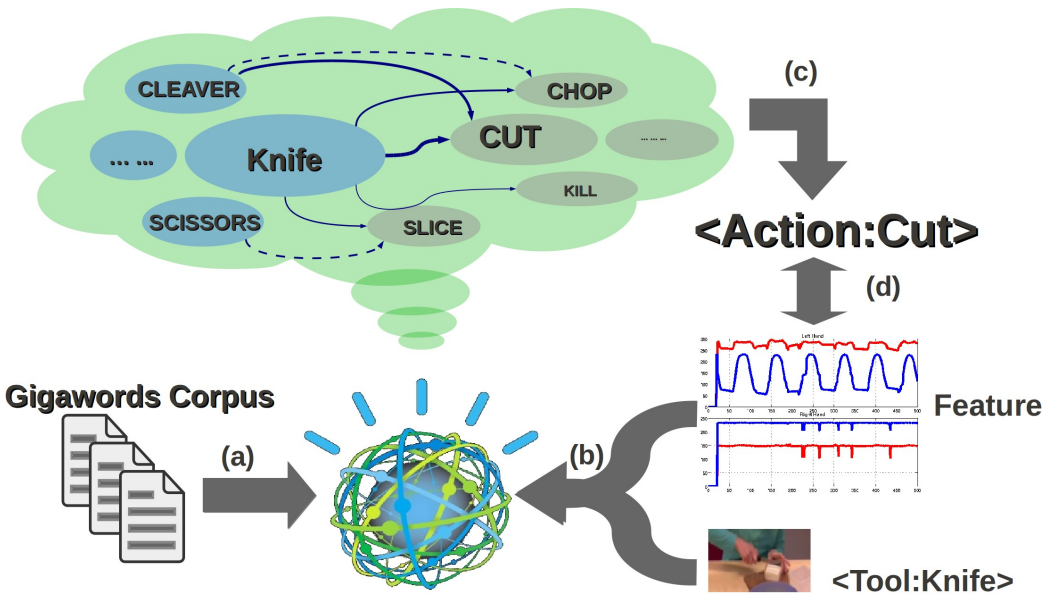
Revisiting Active Perception, R Bajcsy, Y Aloimonos, JK Tsotsos - arXiv preprint arXiv:1603.02729, 2016



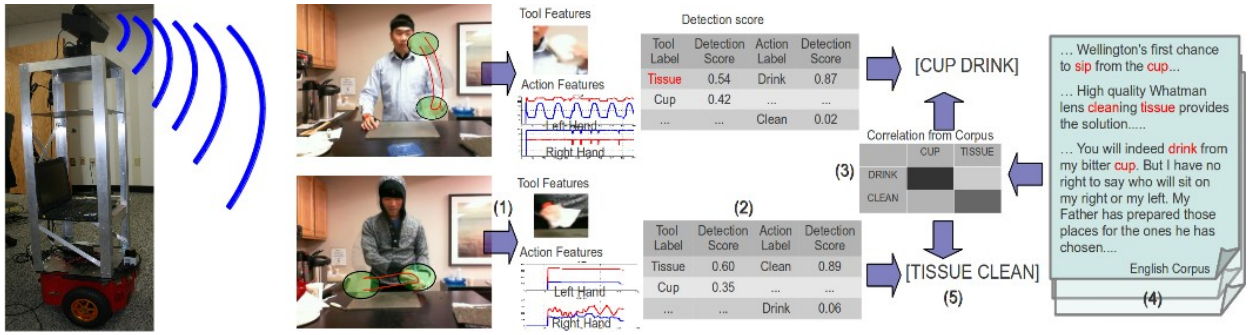
Vision + Language



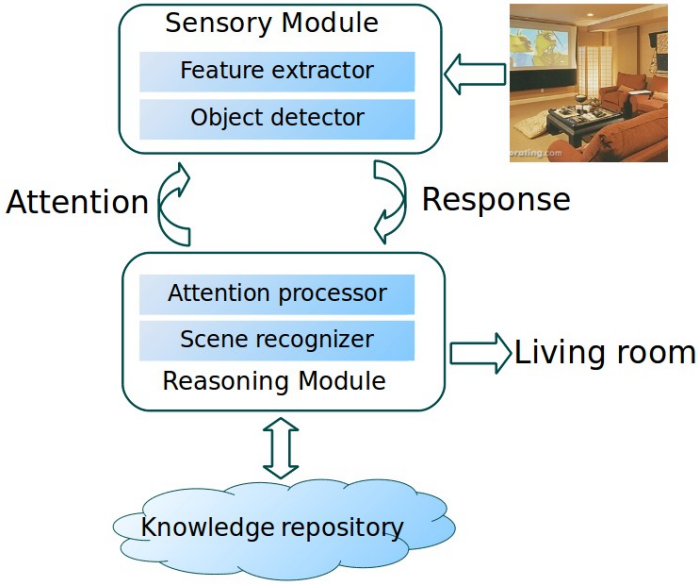
EMNLP 11



ICRA 12



ICRA 13



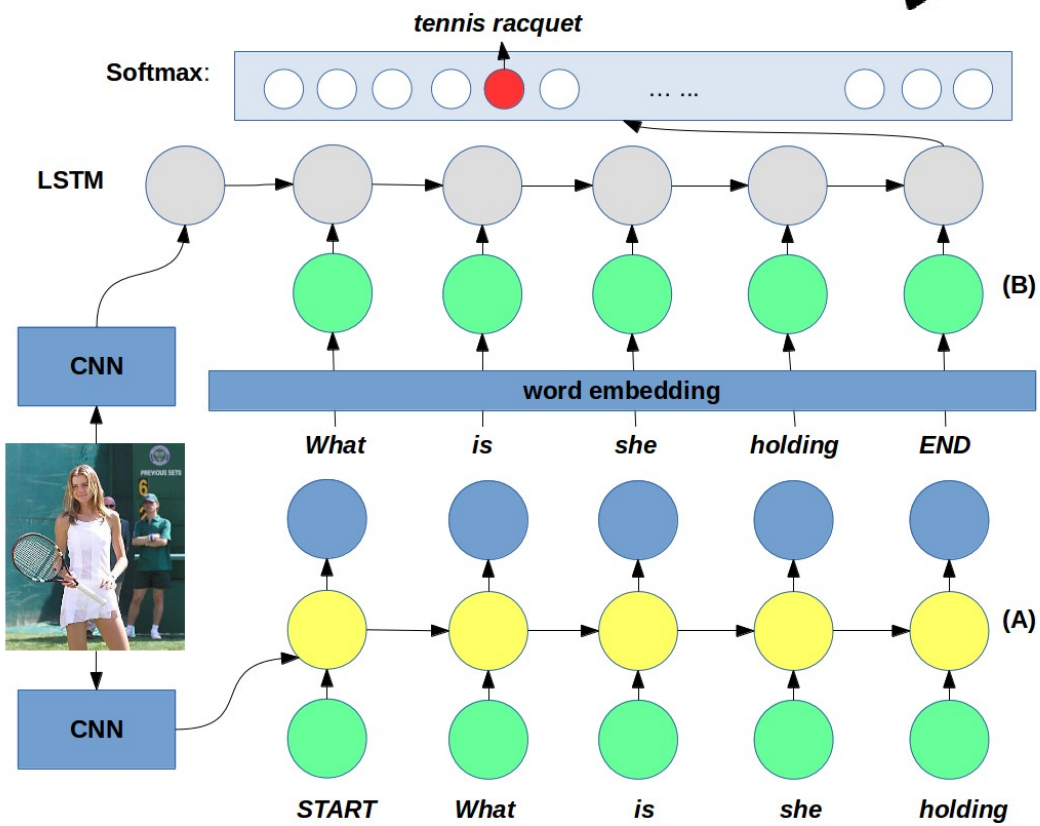
ICCV 11

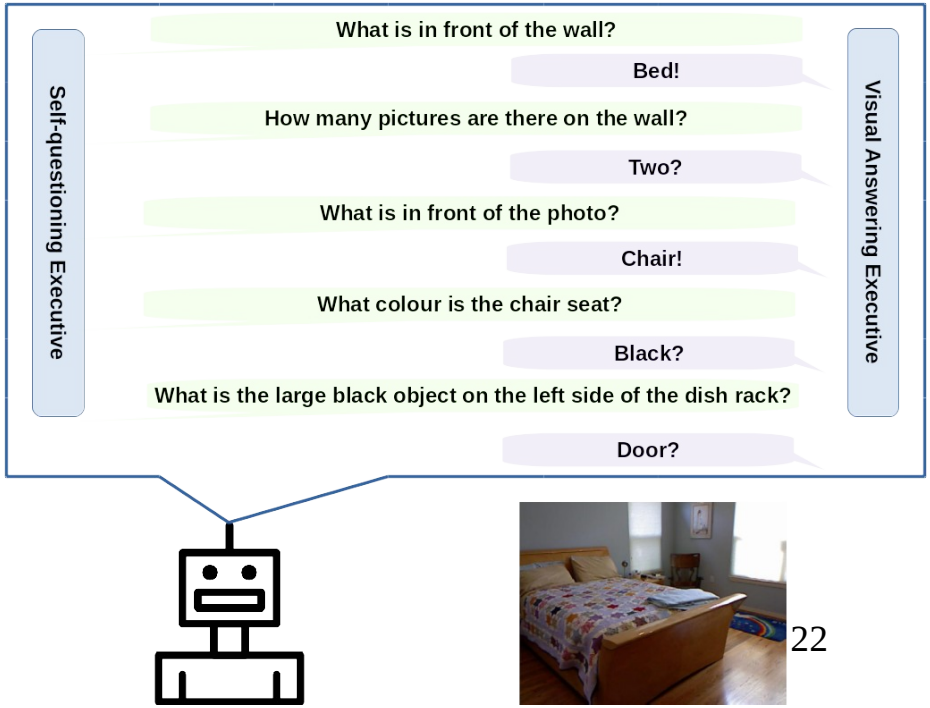
Image Captioning
Image Question Answering
Cognitive Dialogue

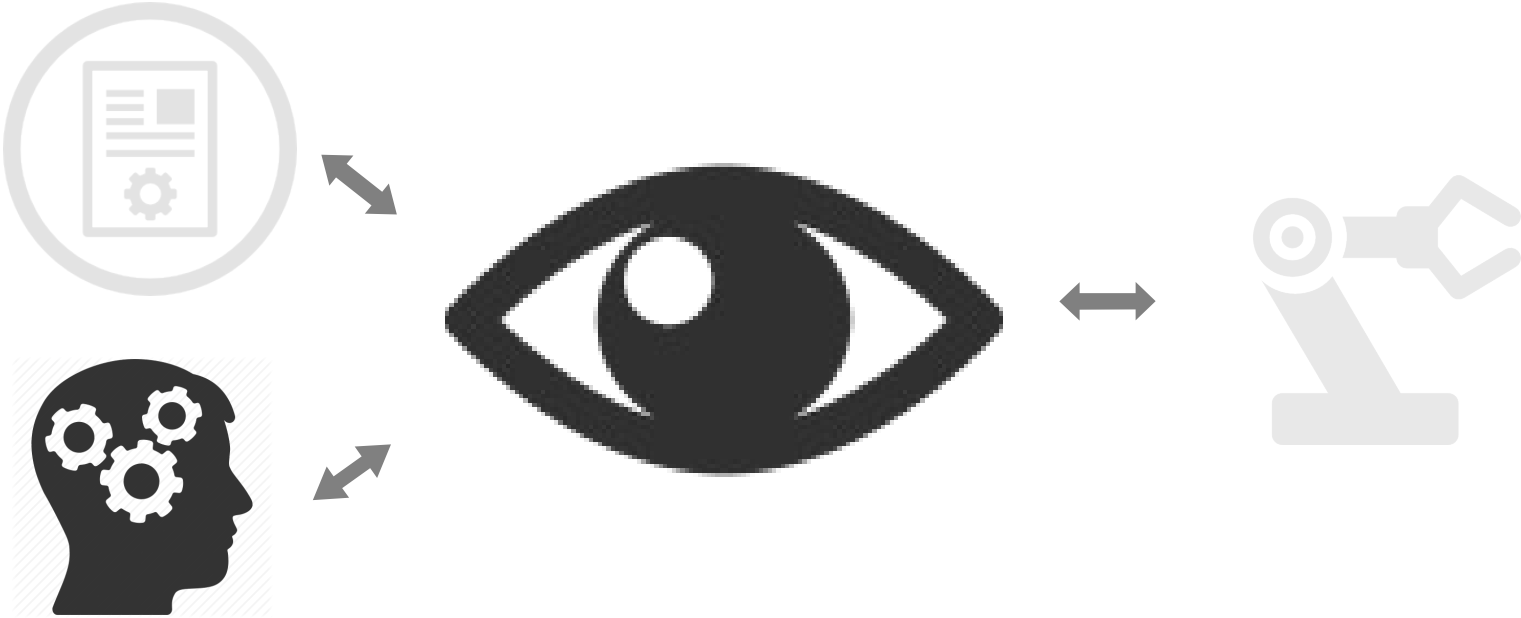
Where is the scene?.....Indoor room
Are they partying?.....Yes
So, birthday party?.....Seems so
How many candles are there?.....Six?
So, how old is she?.....Turning six
… … …

VQA Visual Question Answering

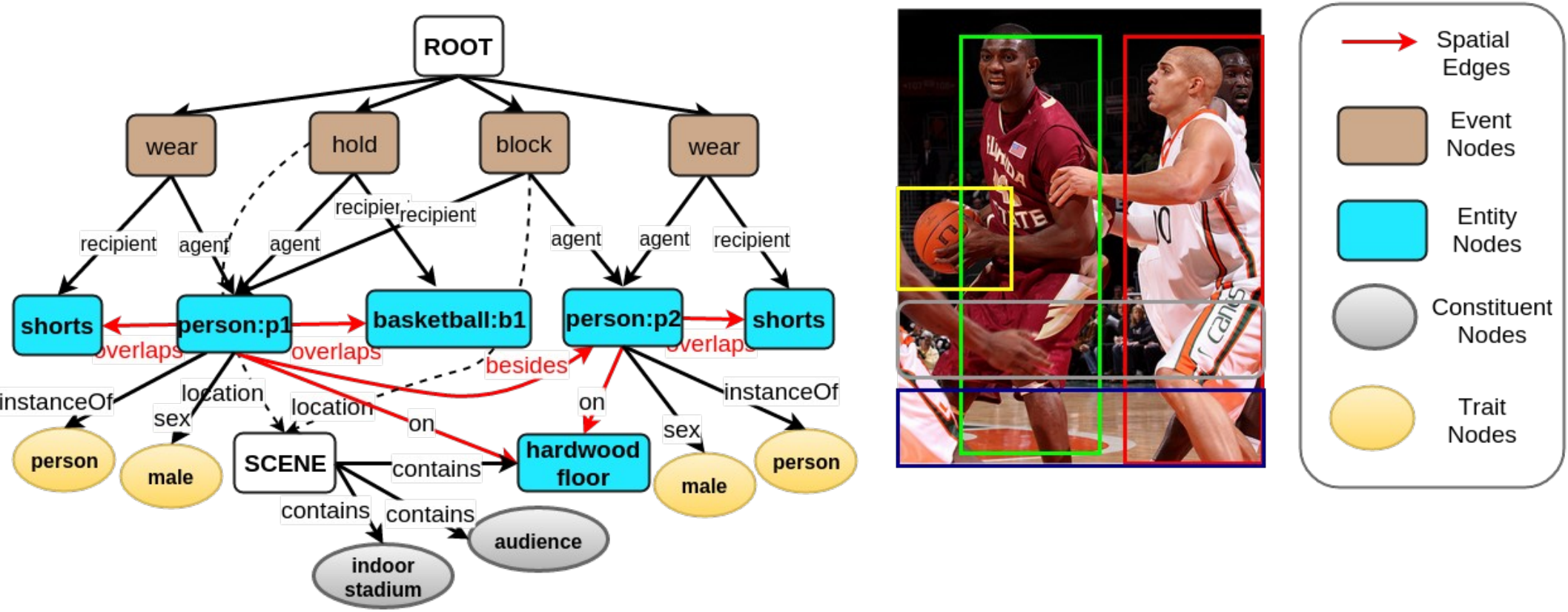
[Parikh et. al.]



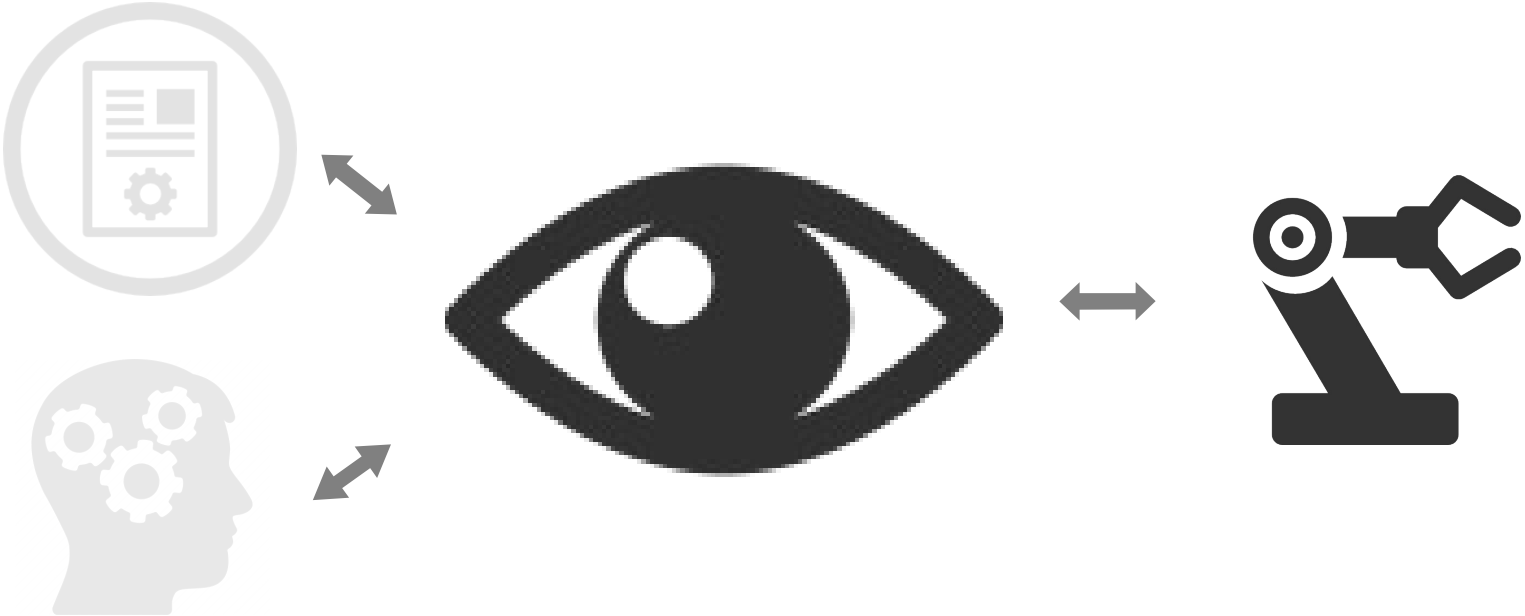


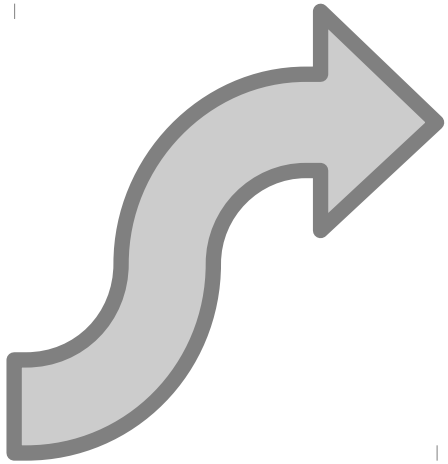


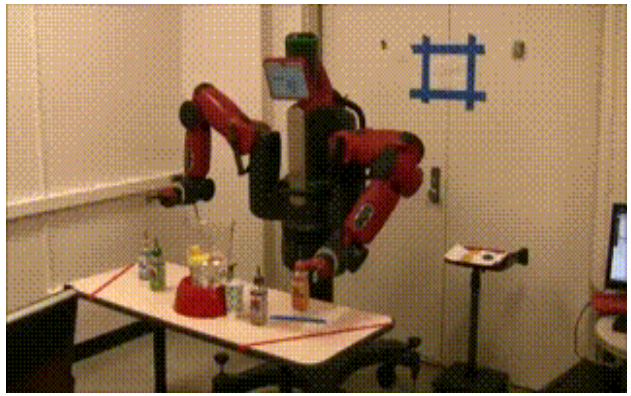
DeepIU: Deep Vision + Commonsense reasoning

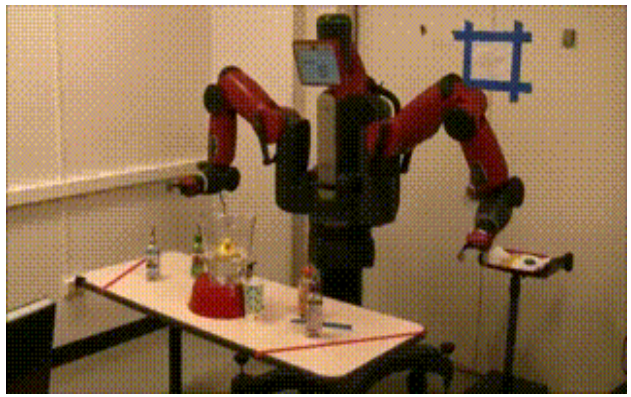


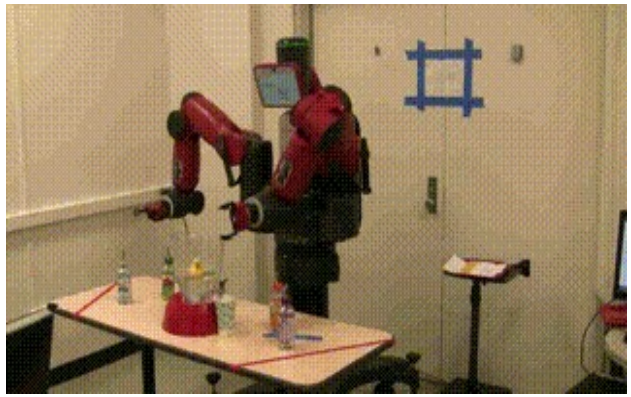
Somak Aditya, Chitta Baral, Yezhou Yang etc.
To appear, Advances in Cognitive Systems, 2016





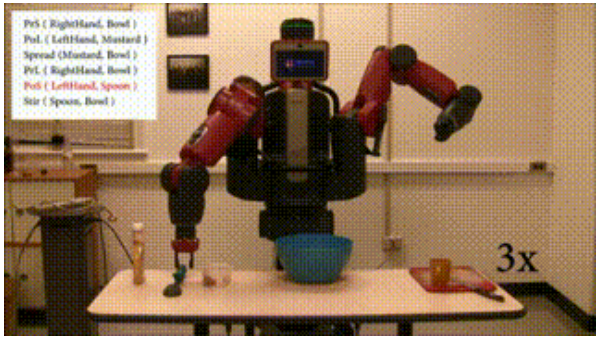






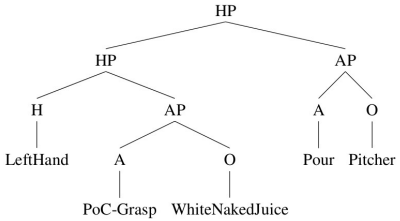
Manipulation actions

Robotic Execution

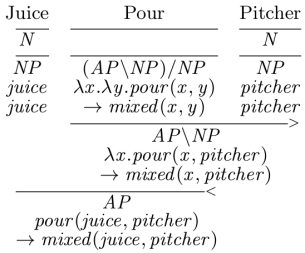


ongoing

Action Sequencing



CogSys 14, AAAI 15

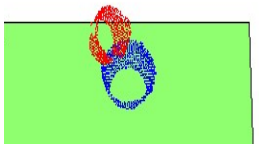


ACL 15, ongoing

Visual Grounding

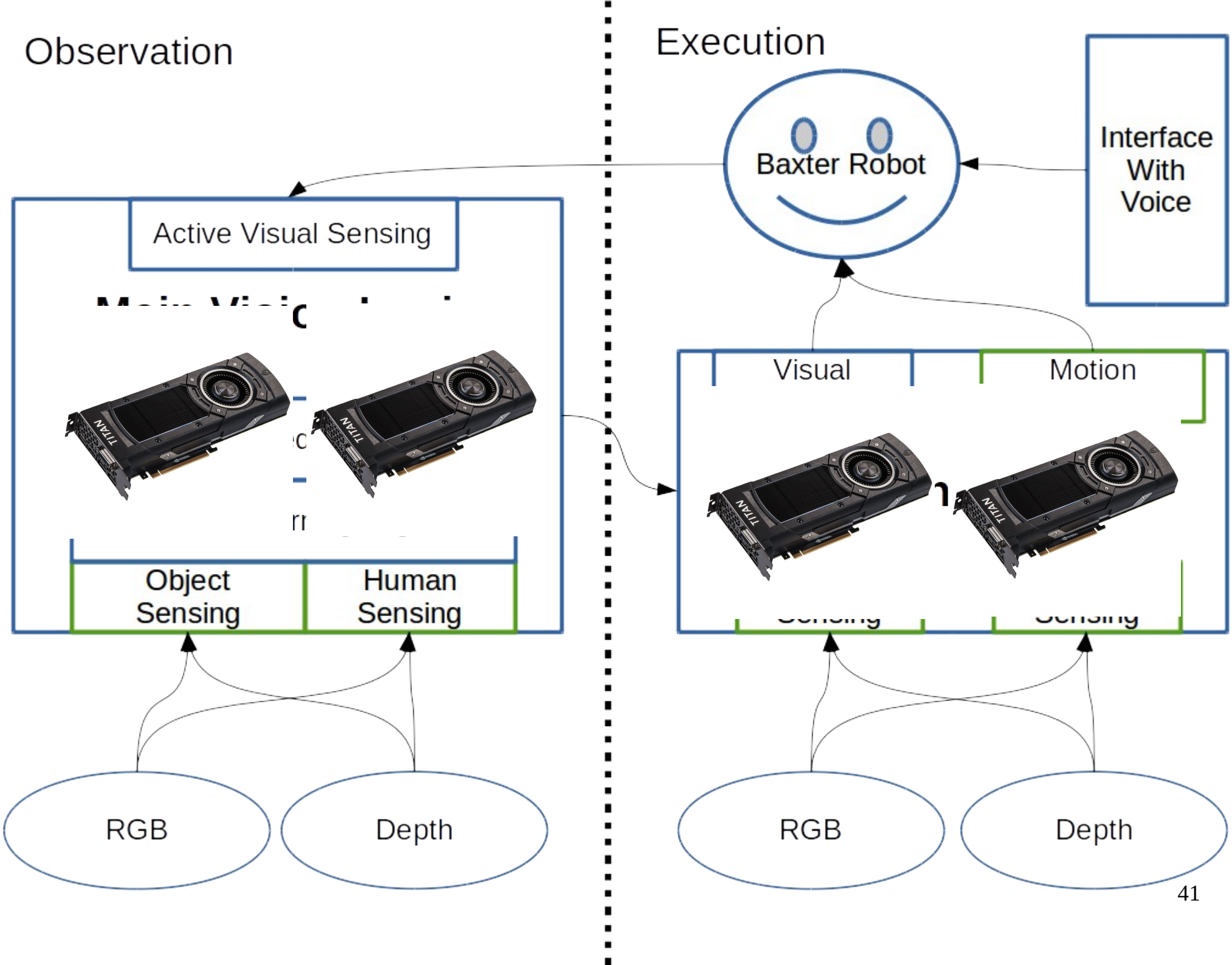
CVPR 15 , ongoing

CVPR 13



ICRA 15

Sensory Input

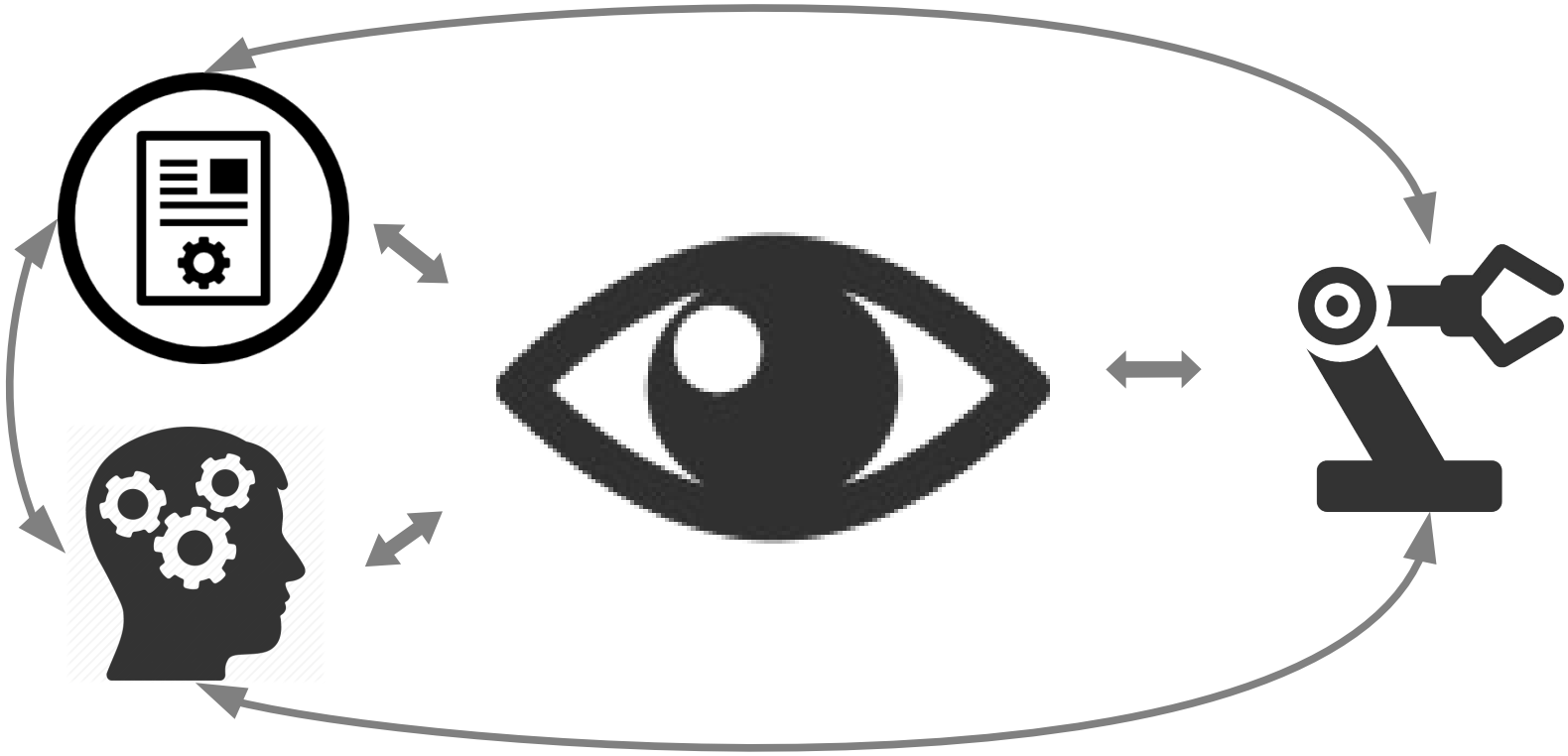

Observation
Execution
Baxter Robot
Interface With Voice
Active Visual Sensing
Visual
Motion
Object Sensing
Human Sensing
RGB
Depth
RGB
Depth

Robot Procedural Learning by Observation

Voice by Stratis Aloimonos


The Robot Visual Learner
by
UMD's
Robot Training Academy

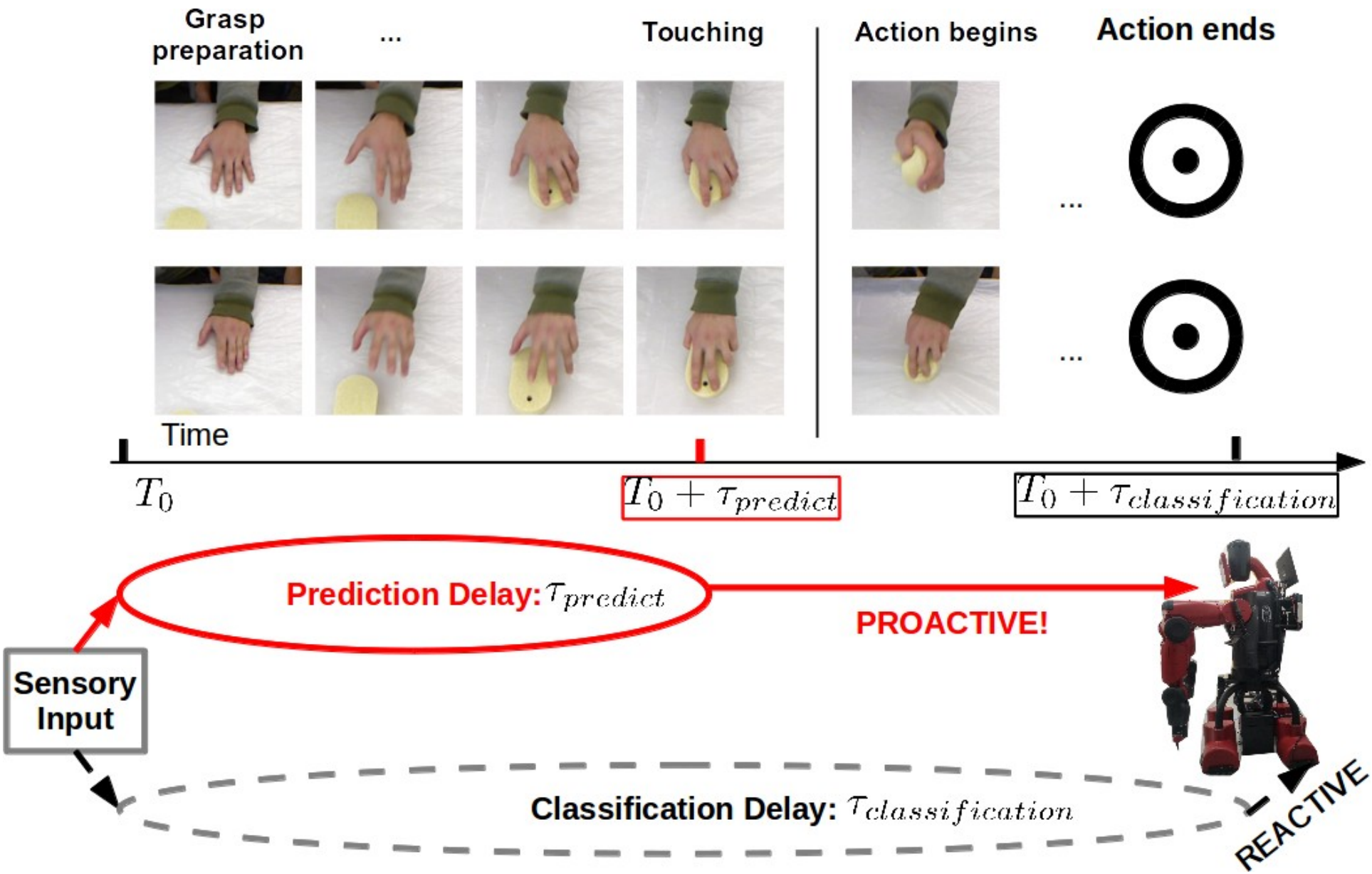
Active Vision/Perception



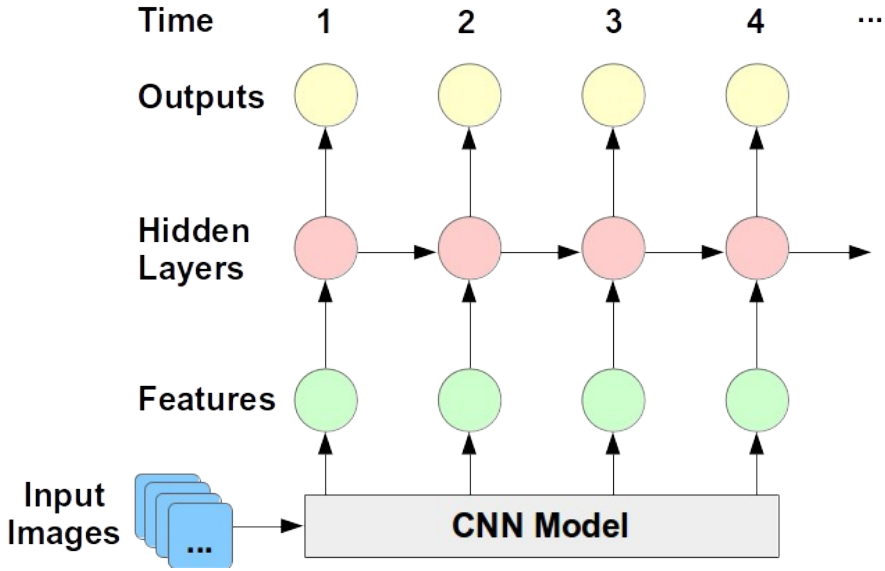
… … …

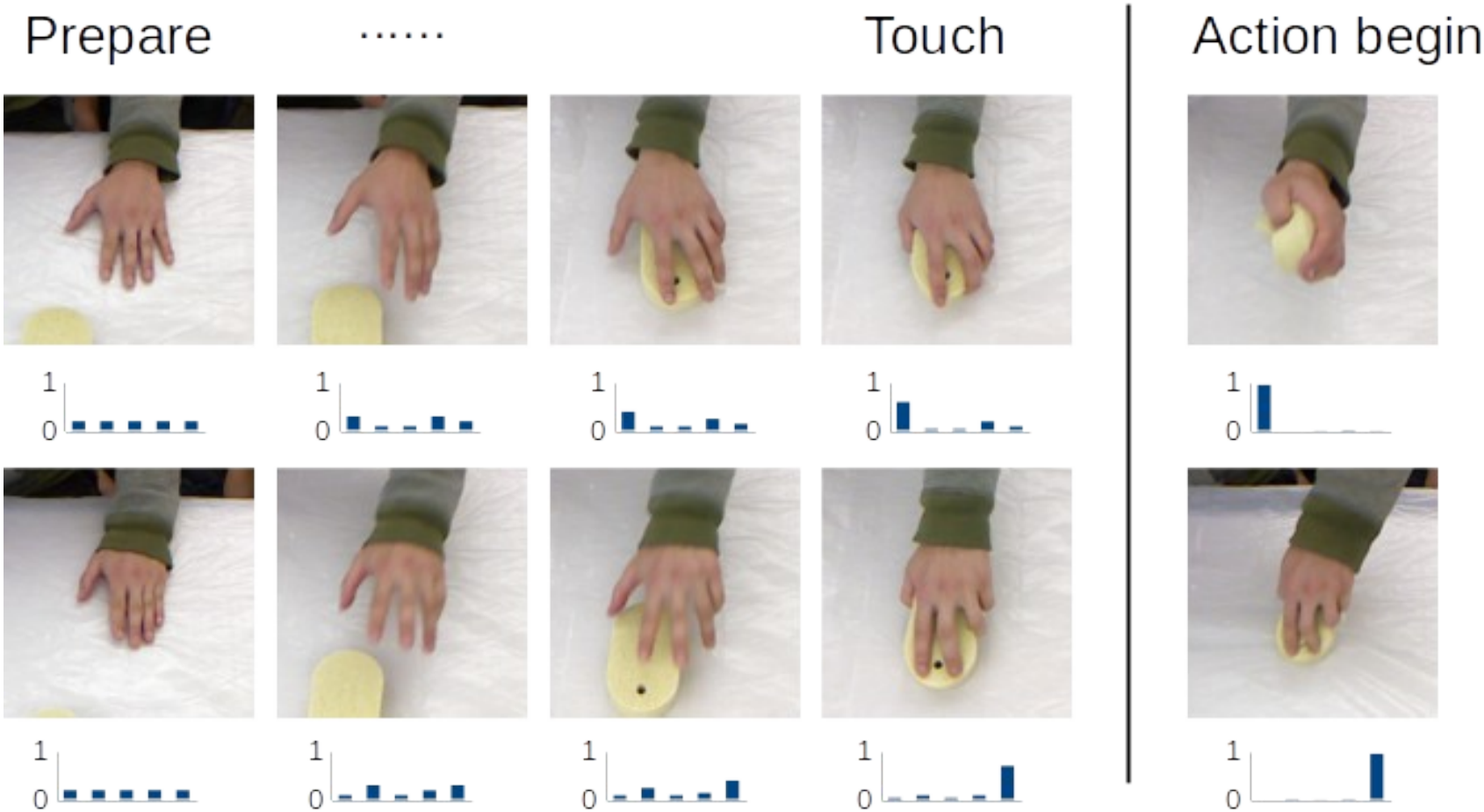
Recognition or Prediction?

Proactive Vision



C. Fermuller, F. Wang, Y. Yang et. al. Under Review

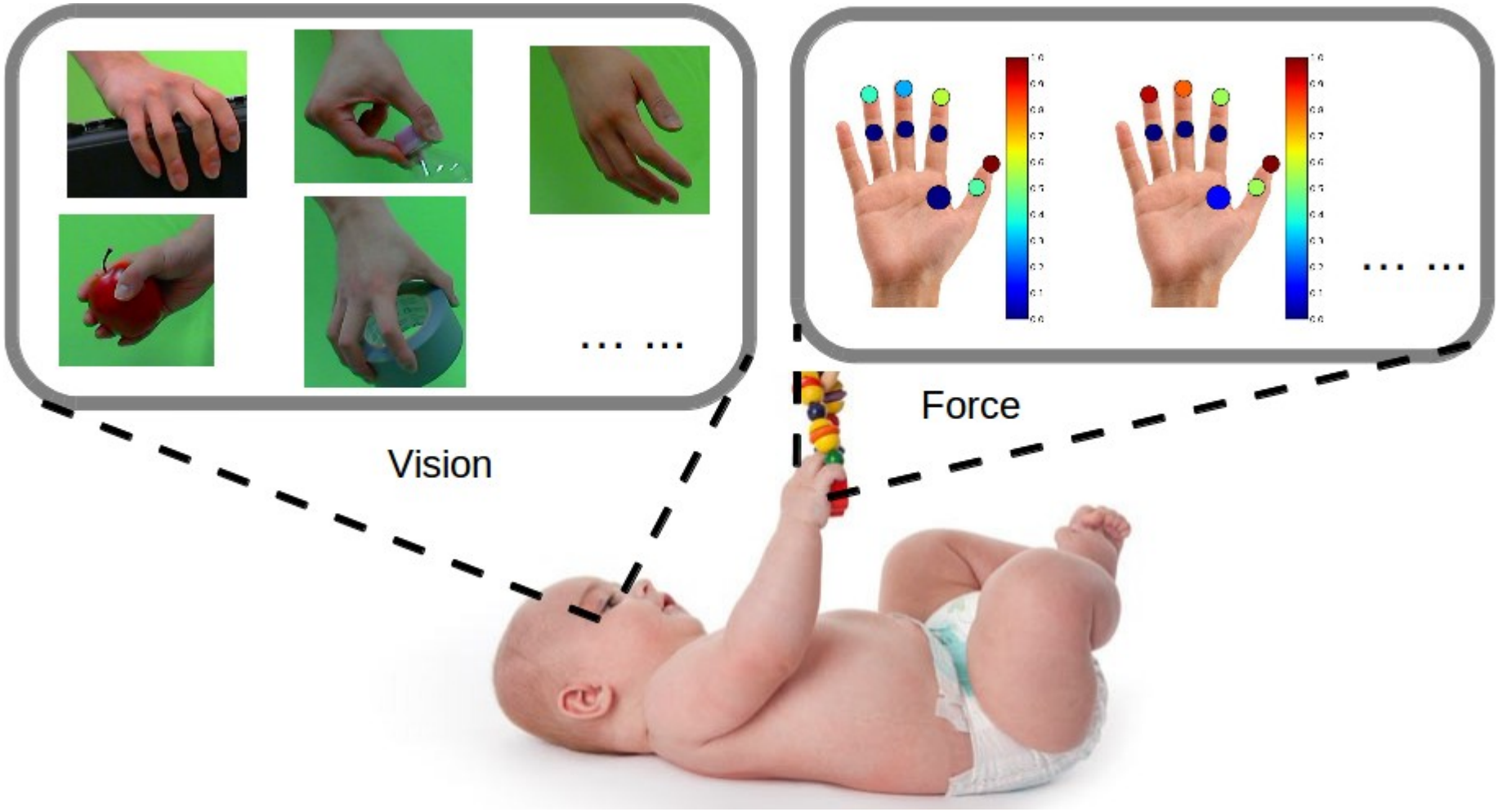

Time
1
2
3
4
...
Outputs
Hidden Layers
Features
Input Images
...
CNN Model

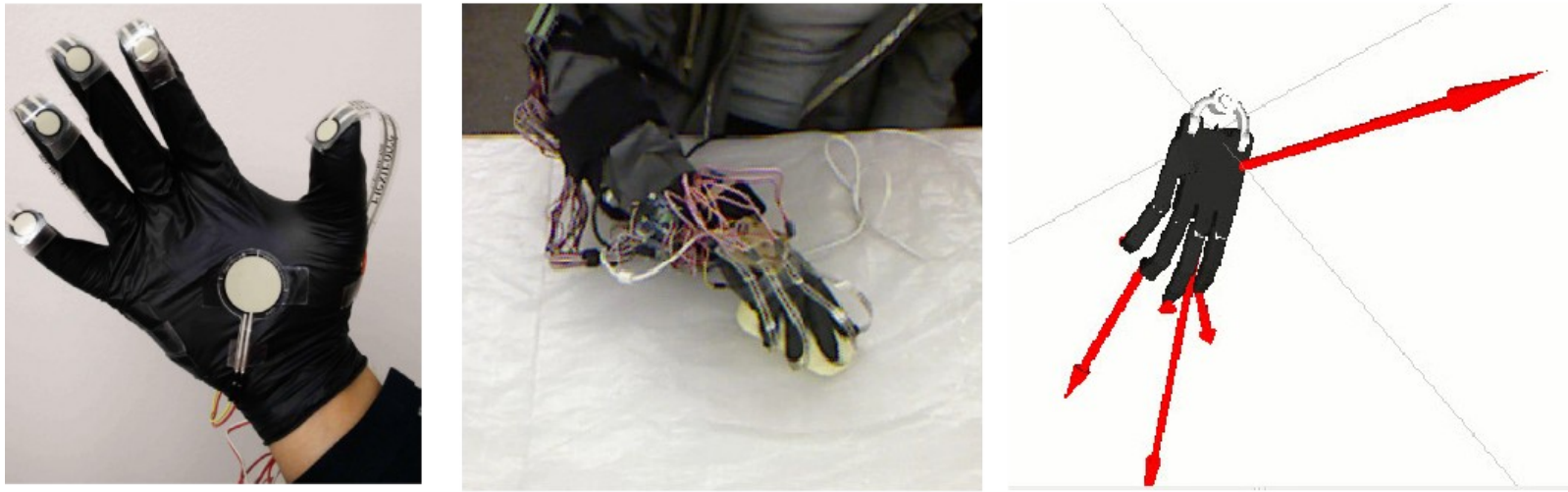

Prepare
.....
Touch
Action begin
1
0

Manipulation Action Prediction

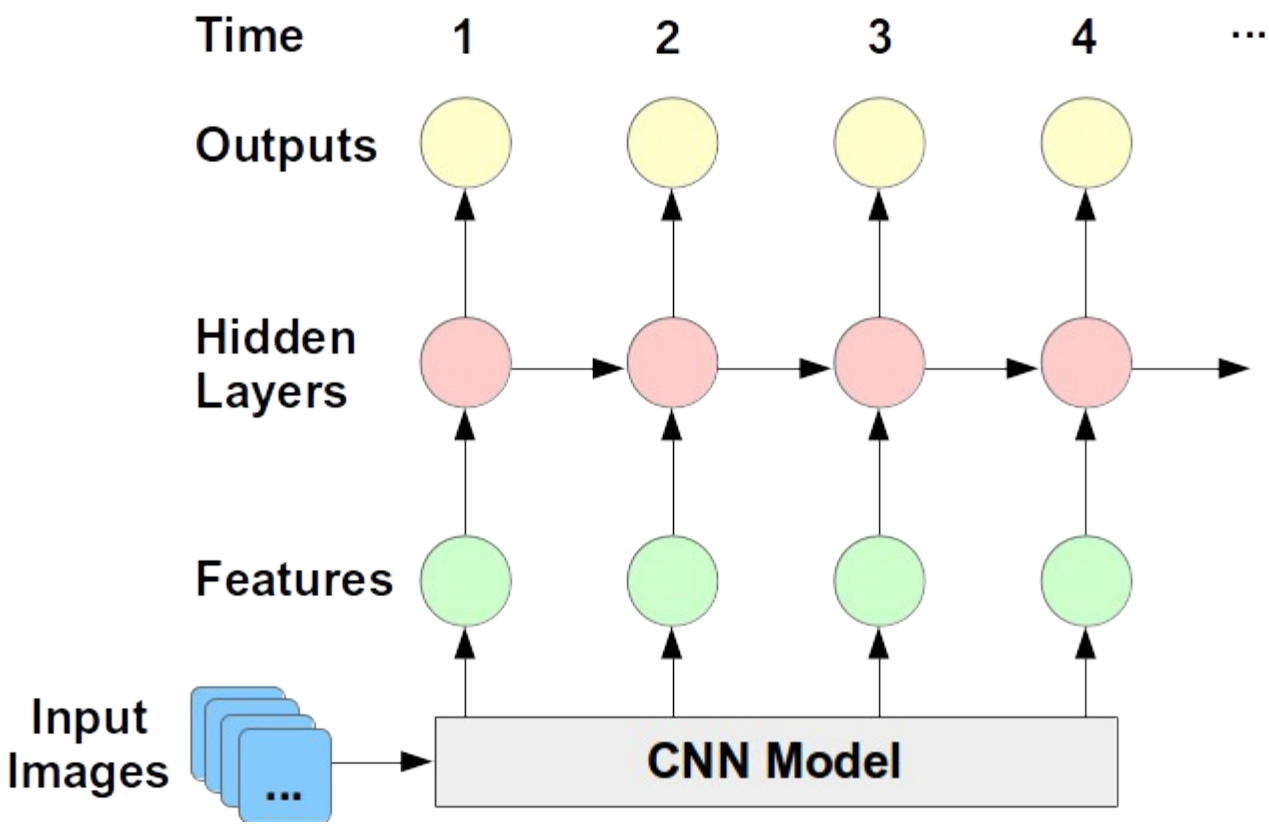
The following video shows the results of the RNN for continuous classification (over 25 object-action pairs)

Recognition or Sensation?


Vision
Force



(a) (b) (c)


Time
1
2
3
4
...
Outputs
Hidden Layers
Features
Input Images
...
CNN Model

Hand Force Estimation

The following video shows the results of the RNN as regressor (of the six forces on the palm and finger tips)

yz.yang@asu.edu