Active visual learning on a humanoid robot

Aleš Ude

Humanoid and Cognitive Robotics Lab

Dept. of Automatics, Biocybernetics, and Robotics

Jožef Stefan Institute

Robot vision & manipulation

- Interaction of perception and action
  - perception provides data for motor control and planning
  - motor actions can facilitate perception,
  - all kinds of learning should be based on perception-action coupling
- Segmentation via manipulation
- Learning object representations
- Object singulation from a pile & grasping

Active exploration

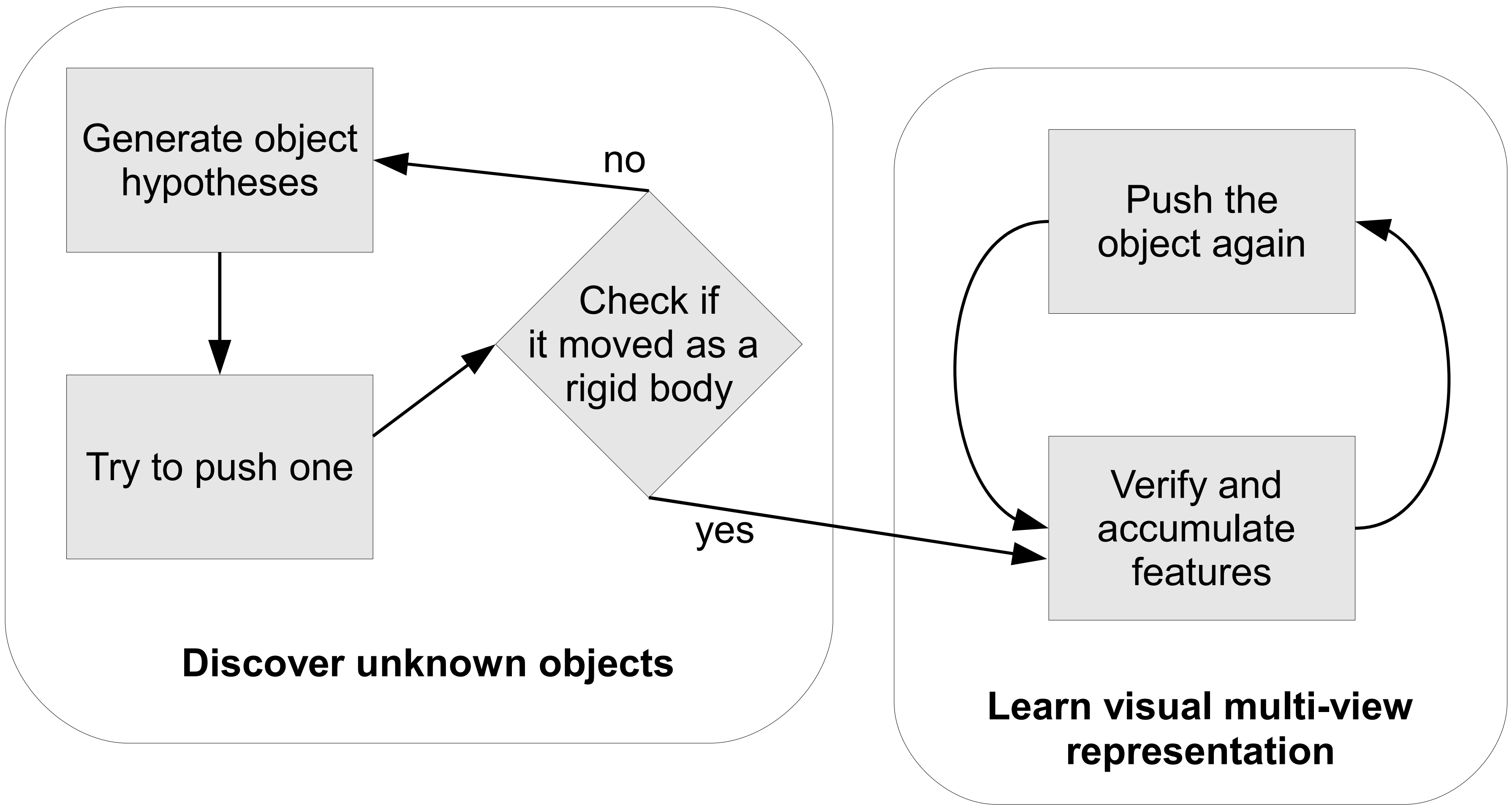
- Acquire visual experiences through experimental manipulation (Metta and Fitzpatrick, Adaptive Behavior, 2003; Fiztpatrick and Metta, Phil. Trans. Royal Society London A, 2003)
- It is much easier to define an object if the system is active.
  - Coherent motion is a very strong cue.
- Lately there has been lot of interest in interactive perception, especially to support manipulation tasks:
  - (Ude et al., IJHR 2008; Schiebener et al., Humanoids 2011; Ude et al., ICRA 2012; Kootstra et al., ICRA 2008; Kenney at al., ICRA 2009; Gupta et al., ICRA 2012; Krainin et al., IJRR 2011; Chang et al., 2012; ...)

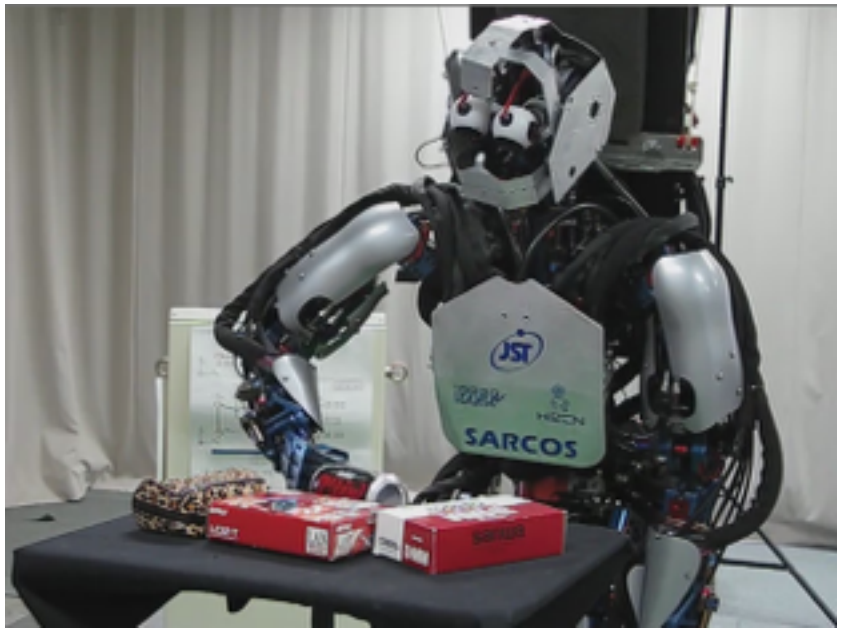
Active Exploration for Learning Object Representations

Bottom-up segmentation

- If action supports perception, crude bottom-up segmentation is fine
- Simple criteria for initial segmentation
- Refinement through action

System Overview





Stergraršek Kuzmič and Ude, Humanoids 2010
Schiebener et al., Adaptive Behavior, 2013
Schiebener et al., ICRA 2014

Generation of object hypotheses

- Calculate 3-D points from Harris interest points using stereo vision.
- Look for regular surfaces: find subsets of the points that lie on such surfaces.
- Feature proximity is another strong cue.
- Regular surface patches serve as initial object hypotheses.
- RANSAC.

Generation of object hypotheses

- Hypotheses are often incomplete and may sometimes be wrong
- Experiments (complex scenes):

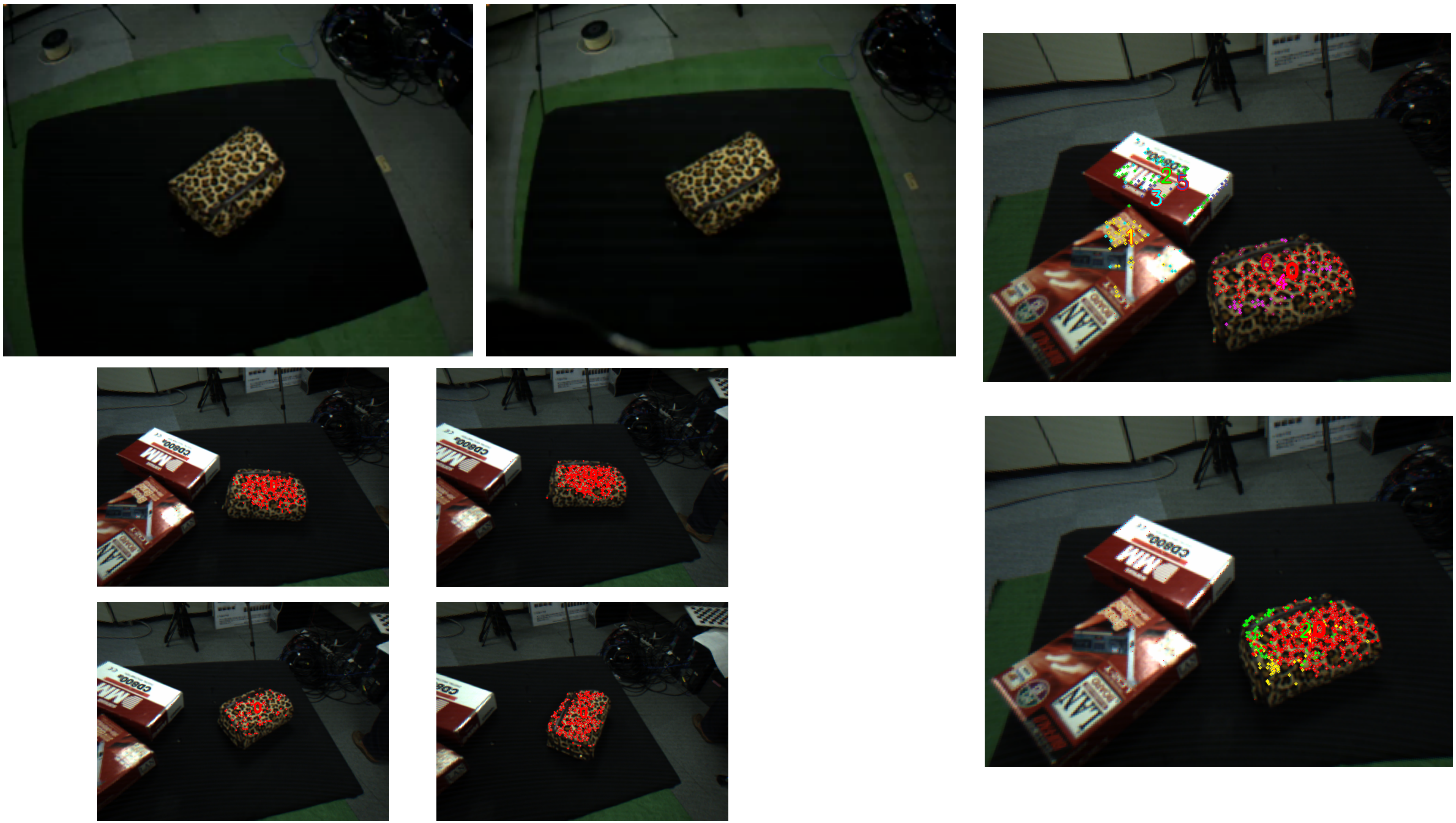
Good	part of object	bad
50 %	39 %	11 %

- Initial hypotheses are unreliable and incomplete but they are a good indication for possible objects and their location
- Inducing motion on an object allows its separation from the background

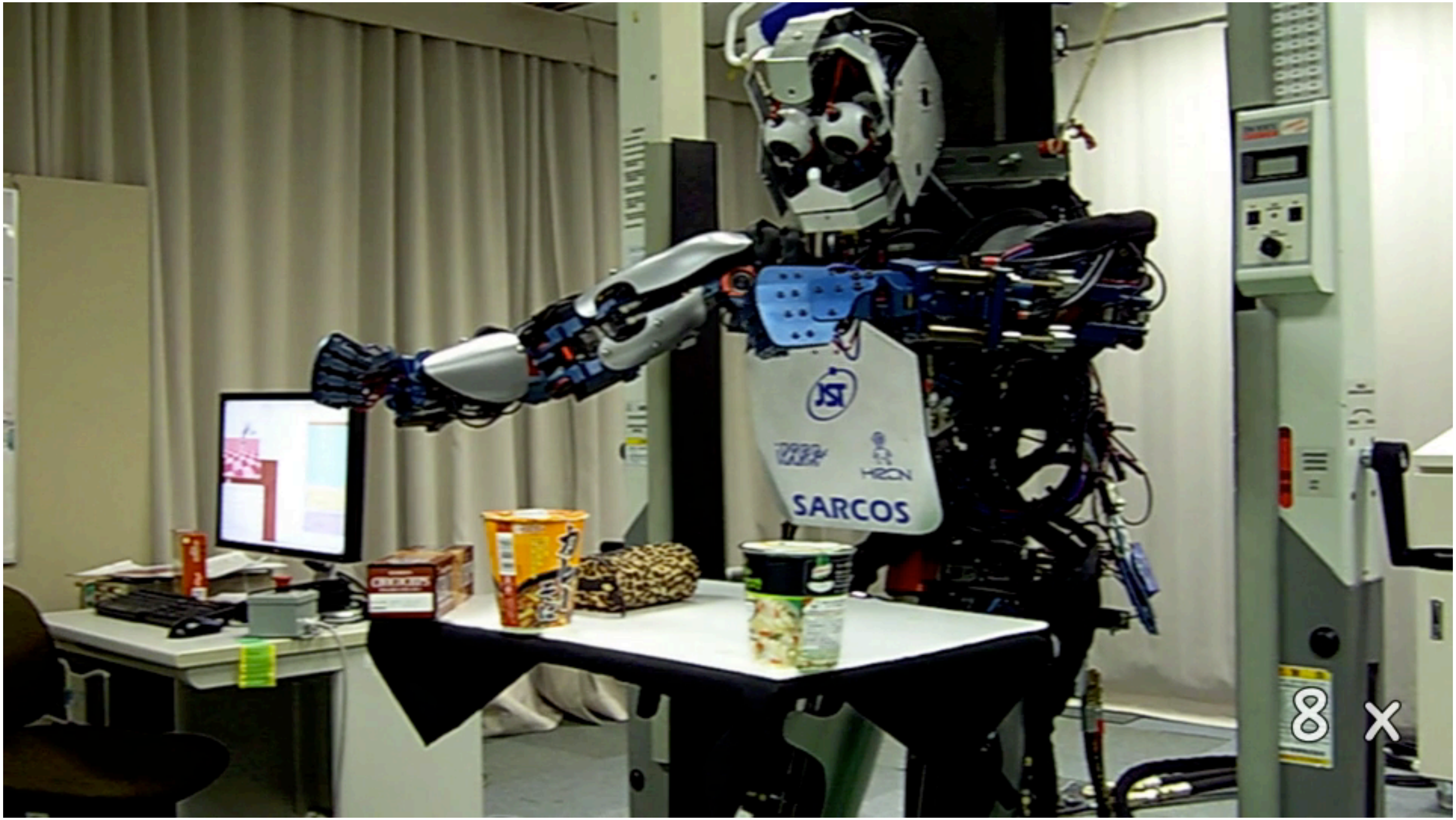
Verification by Pushing

- Randomly generated pushing actions
  - Planning?
- Hypothesis verification by RANSAC: check if the hypothetical object moved as a rigid body.

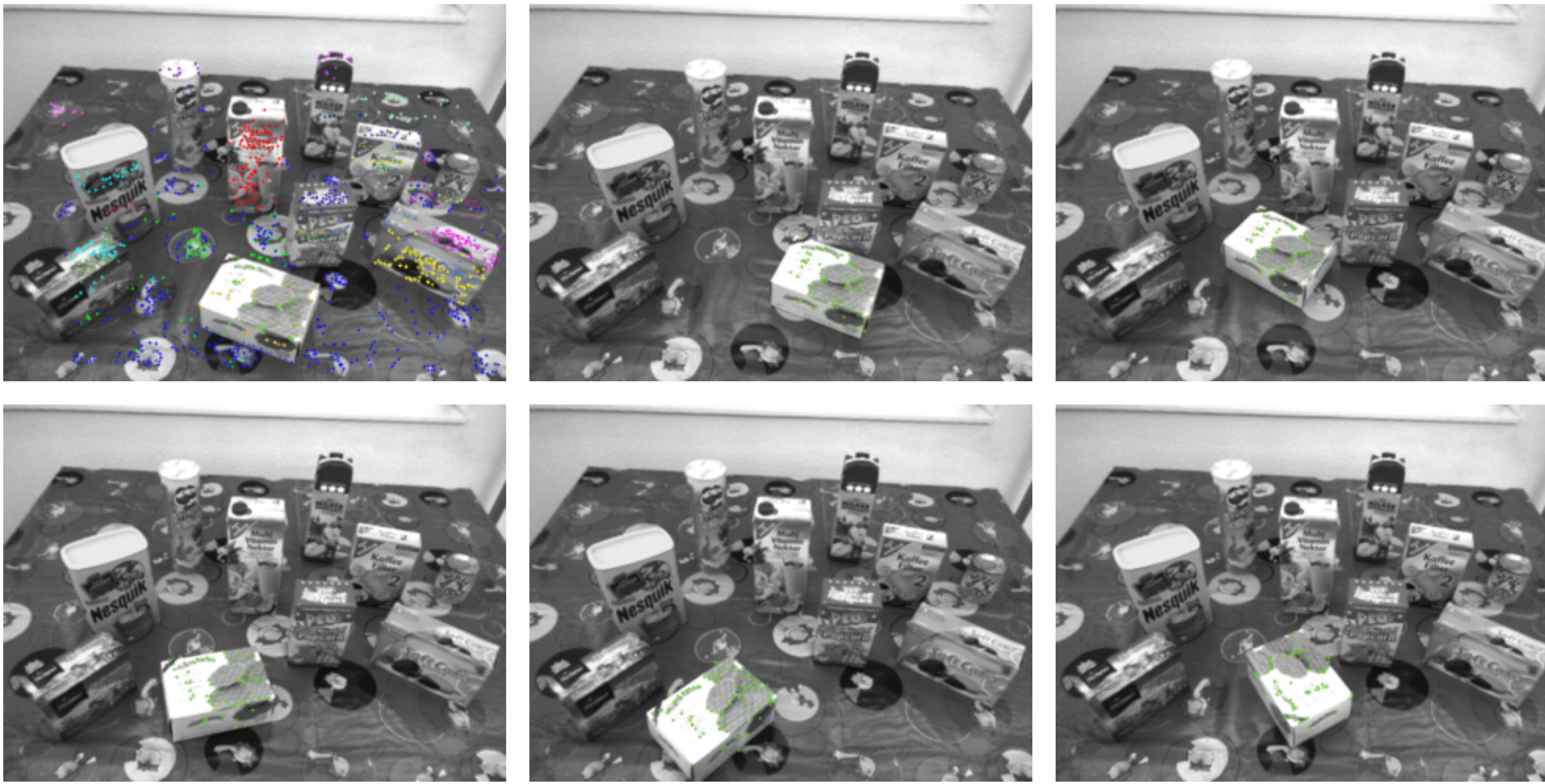
Feature Accumulation



Object Learning by Bimanual Pushing



Acquired Sequences



Enhancements: Force feedback

Schiebener et al., Humanoids 2012

Object Learning for Recognition

- "Bag of Features" model with SIFT descriptors at Harris interest points and MSER feature descriptors
  - Segmentation provided by manipulation
  - Doesn't require reliable long-term feature tracking
- Global hue histogram in the area spanned by interest points
- Recognition: BoF model and hue histograms
- Segmentation through pushing

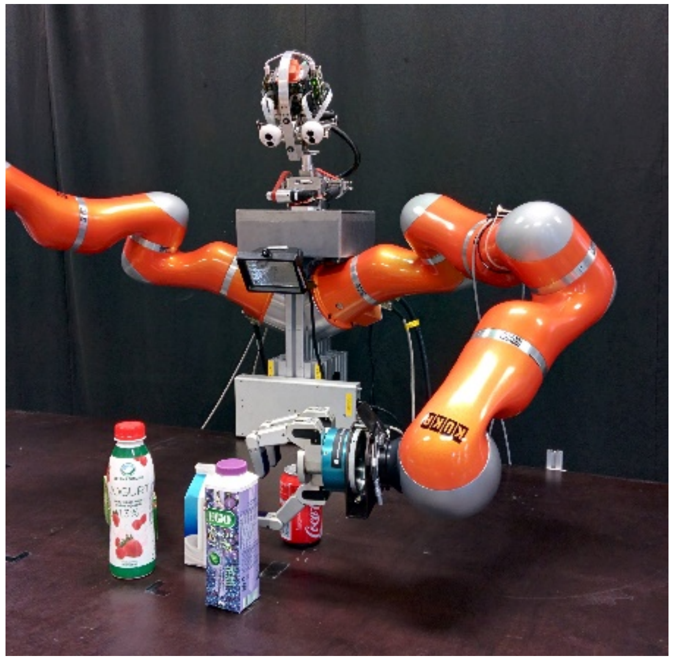
Recognition rates (15 objects learned with our approach + 25 from images)

init. hyp.	1 push	2 pushes	3 pushes
77 %	86 %	96 %	98 %

Interactive Object Learning

Enhancements: Foveal Vision

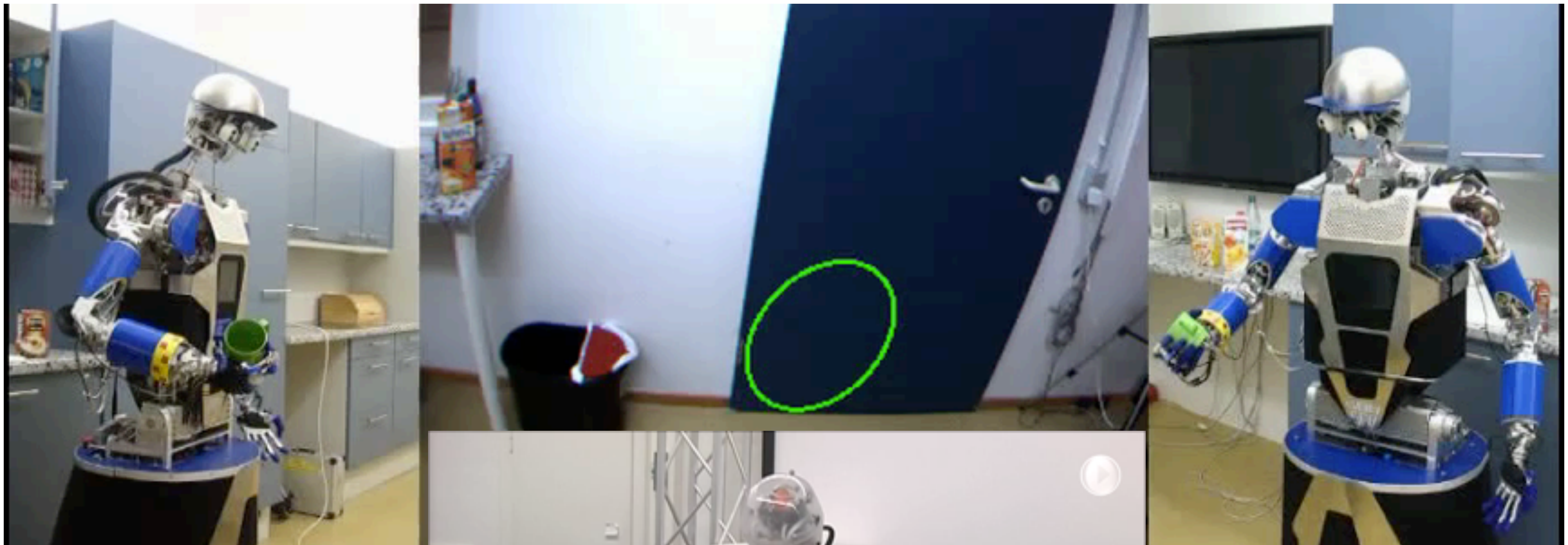
- Integration of foveal vision and robot manipulation for learning object representations and for recognition.
- Similar visual processing.
- Improved models and recognition rates.



Learning by Foveated Vision

Bevec & Ude, Humanoids 2013

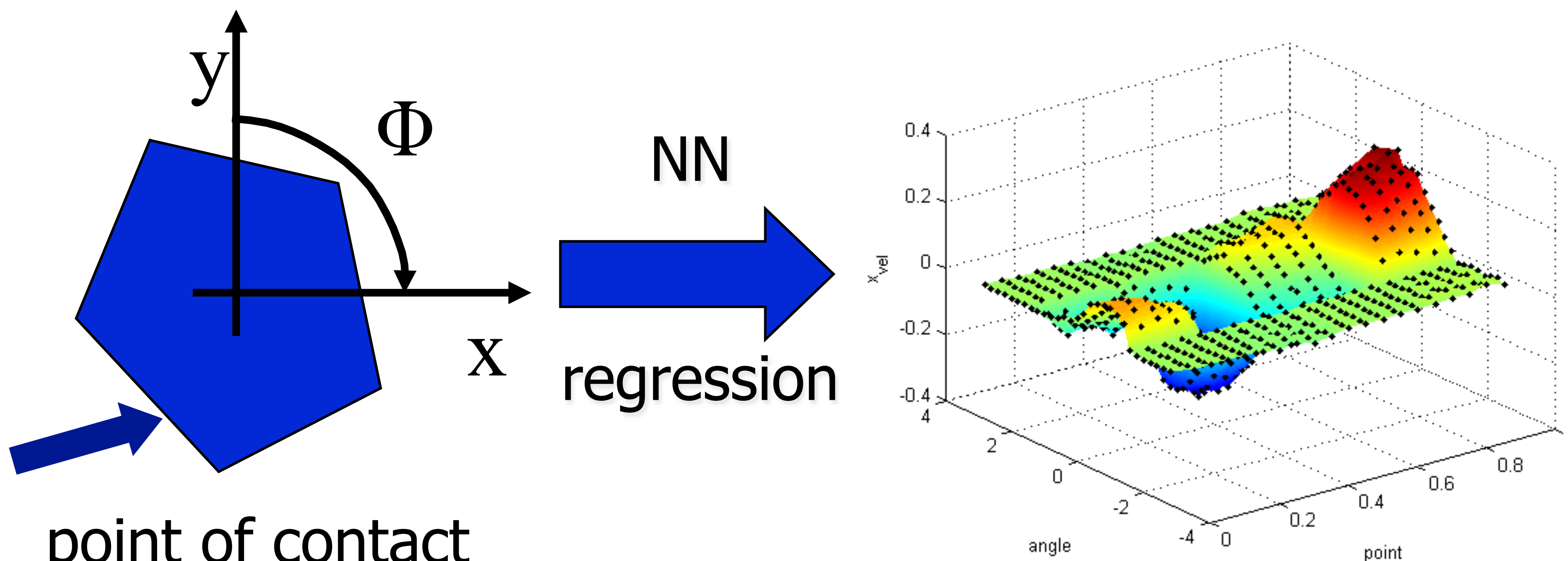
Enhancements: Reactive Grasping & In-Hand Manipulation



Ude et al., International Journal of Humanoid Robotics, 2008

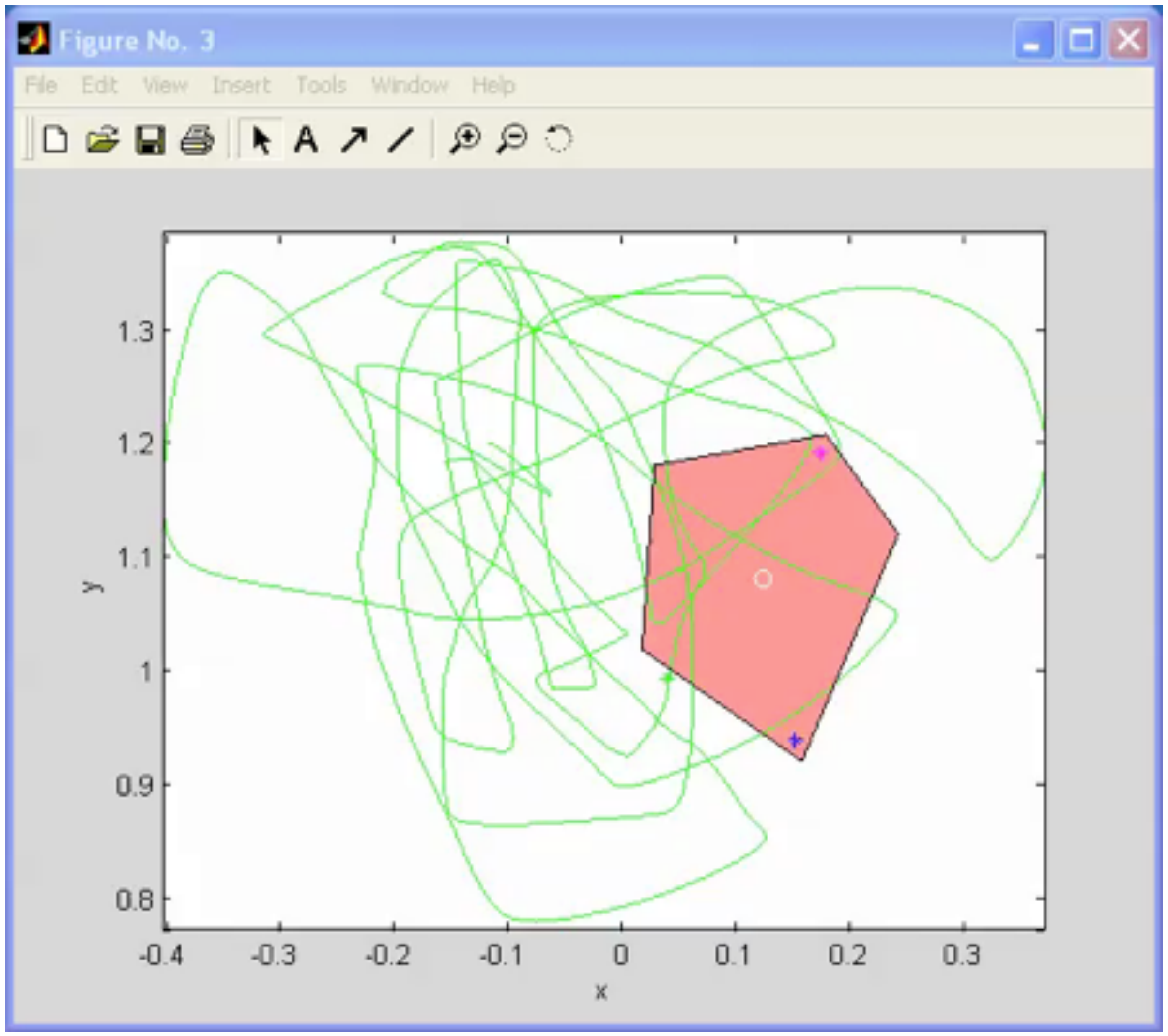
Schiebener et al., Humanoids 2012

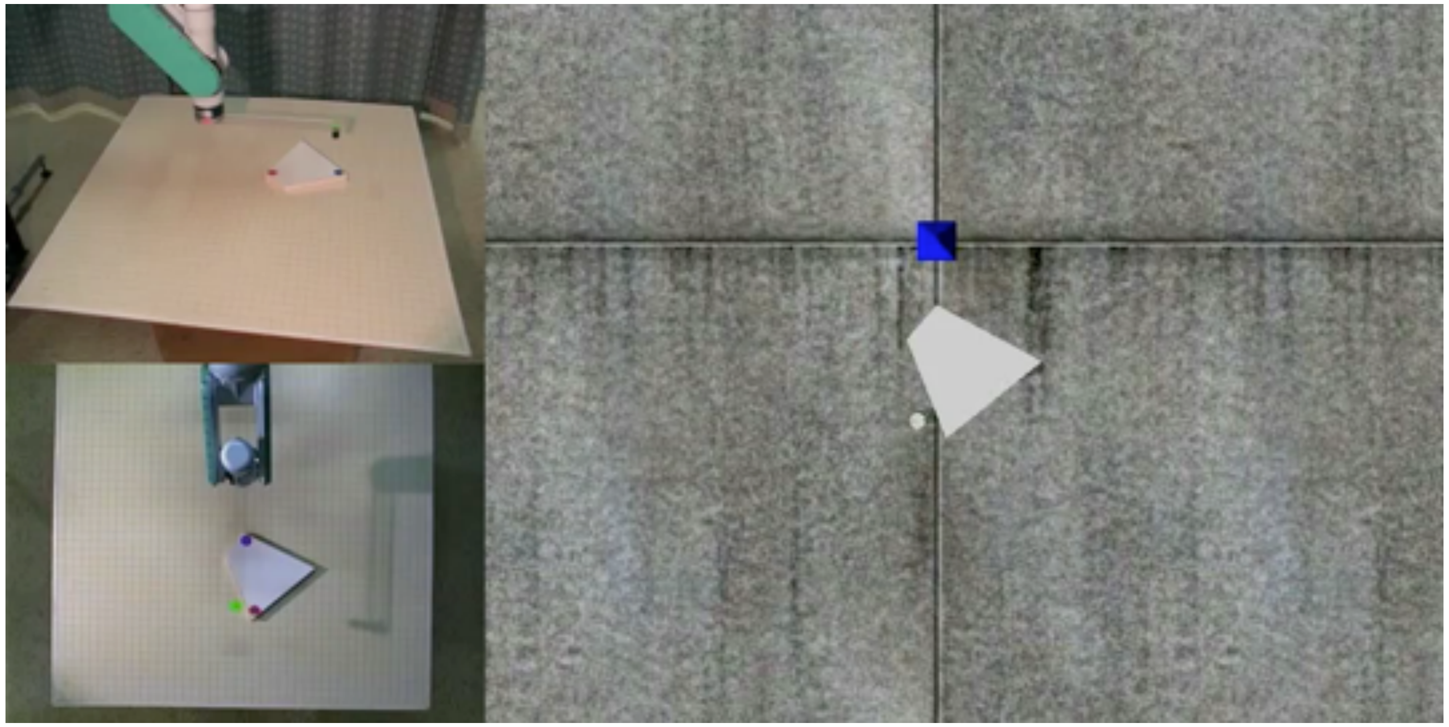
Learning pushing actions



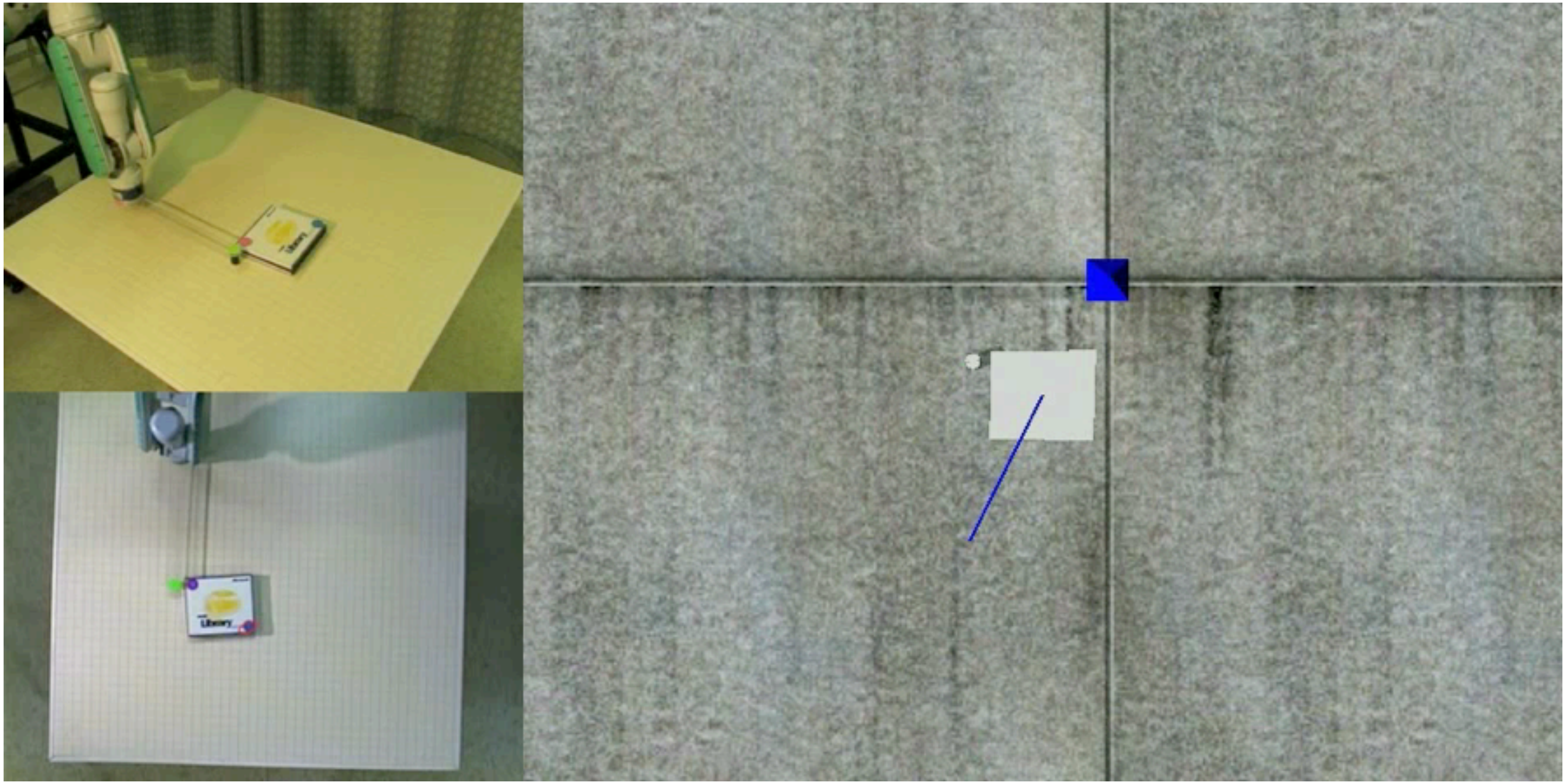
point of contact
angle of contact

Data acquisition





Pushing control



Omrčen et al., Humanoids 2008

Summary

- More integration between control, vision and planning is necessary.
- Better integration with other modalities, especially tactile sensing.

- Contributions by: David Schiebener, Damir Omrčen, Robert Bevec, Jun Morimoto, Tamim Asfour
- Funding: PACO+, Xperience, NICT (Japan Trust)