ISTITUTO ITALIANO
DI TECNOLOGIA

Visual learning of objects and tools on the iCub robot

Lorenzo Natale

iCub Facility

Istituto Italiano di Tecnologia, Genova, Italy

Workshop on Towards Intelligent Social Robots – Current Advances in Cognitive Robotics
November 3rd, 2015, Seoul, South Korea

Italy


WWW.FREEWORLDMAPS.NET
ICELAND
FAROE ISLANDS
NORWAY
SWEDEN
FINLAND
RUSSIA
ESTONIA
LATVIA
LITHUANIA
BELARUS
IRELAND
UNITED KINGDOM
NETHERLANDS
BELGIUM
LUX.
GERMANY
POLAND
CZECH REPUBLIC
SLOVAKIA
UKRAINE
MOLDOVA
FRANCE
SWITZERLAND
AUSTRIA
HUNGARY
ROMANIA
SLOVENIA
CROATIA
BOSNIA
SERBIA
BULGARIA
MON.
KOS.
MAC.
ALBANIA
PORTUGAL
SPAIN
ITALY
GREECE
TURKEY
CYPRUS

Genova

Italy

Italian Institute of Technology

Genova

Italy

ISTITUTO ITALIANO DI TECNOLOGIA

Genova

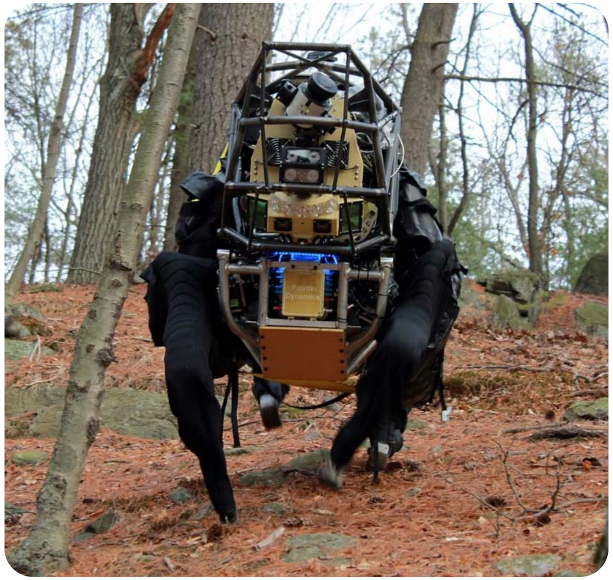
Italian Institute of Technology

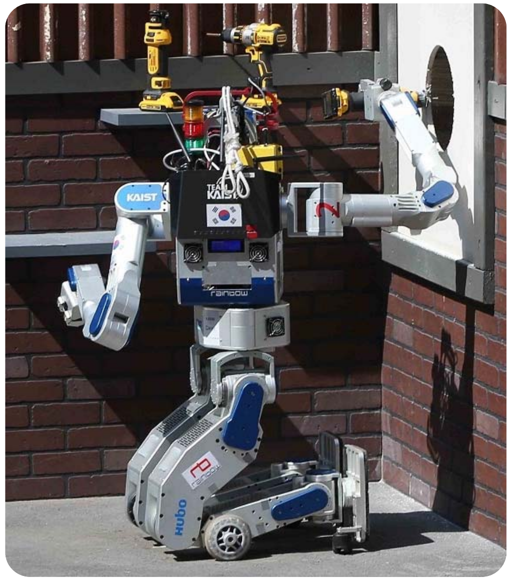
Italy

Robotics @ IIT

iCub Facility

Motivations

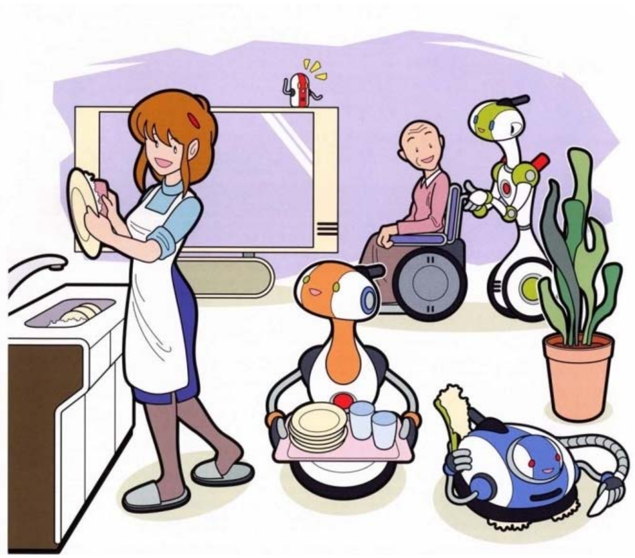


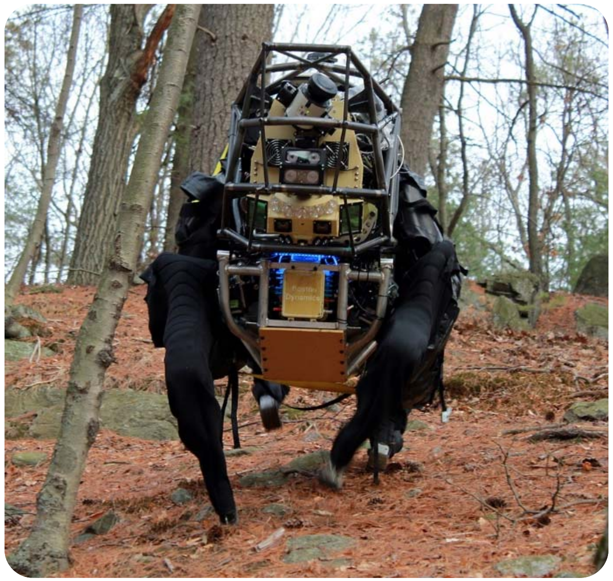


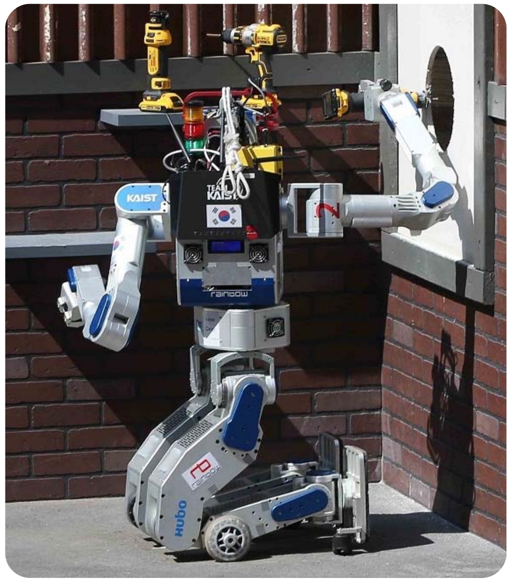
Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

Autonomous



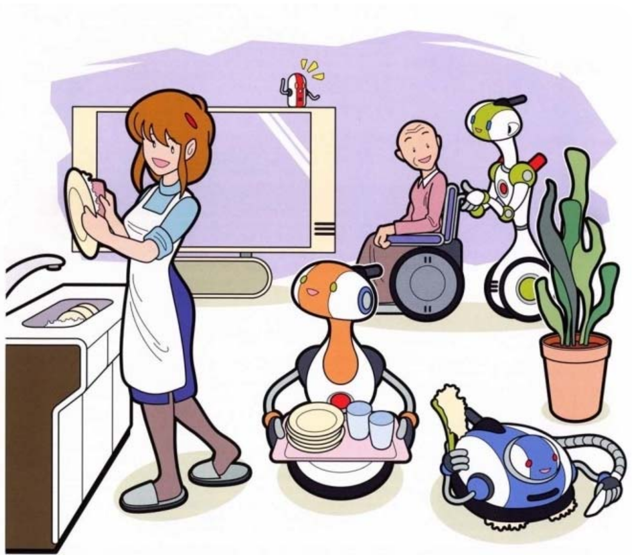


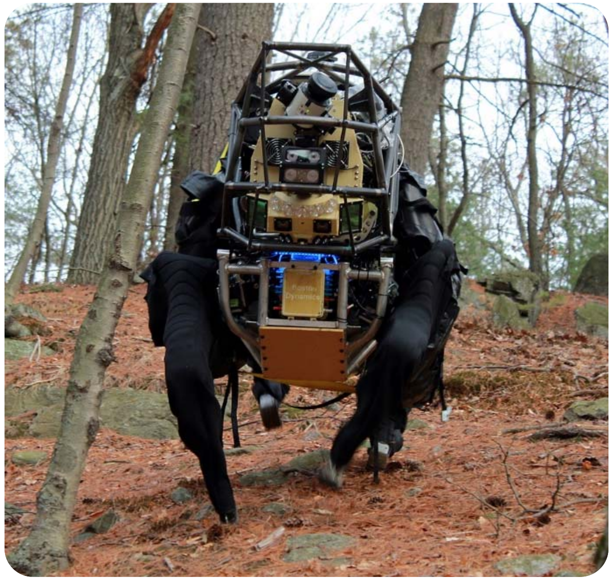


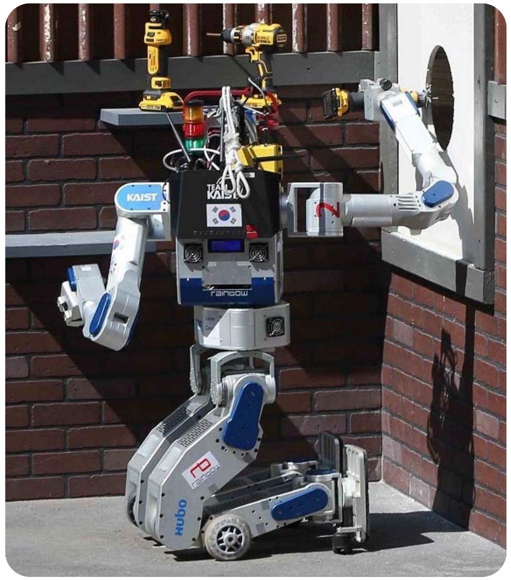
Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

Autonomous
Friendly (humans)



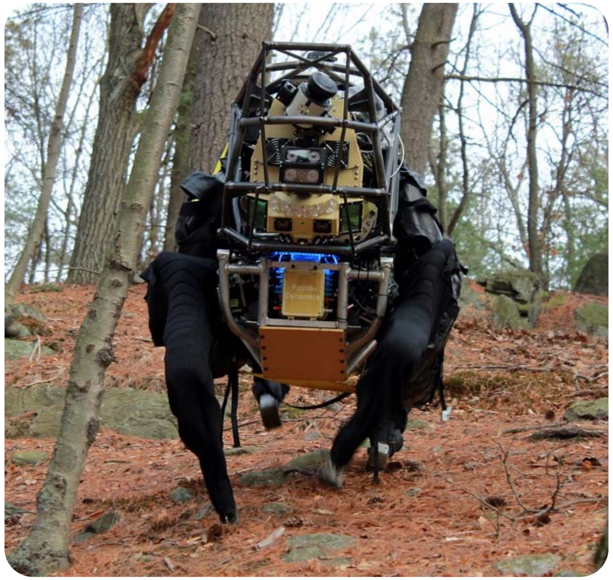


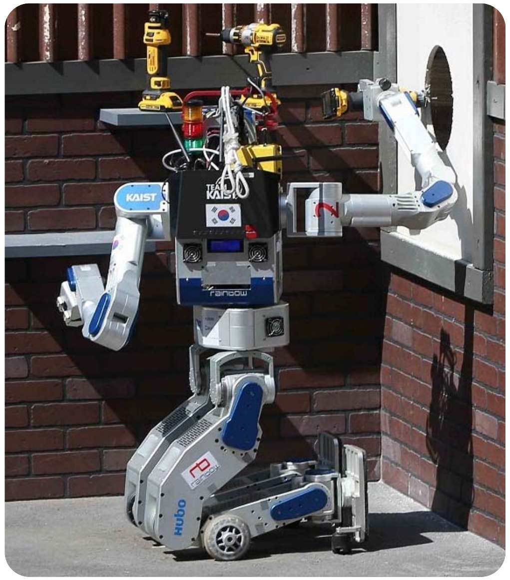


Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

Autonomous
Friendly (humans)
Perception & control

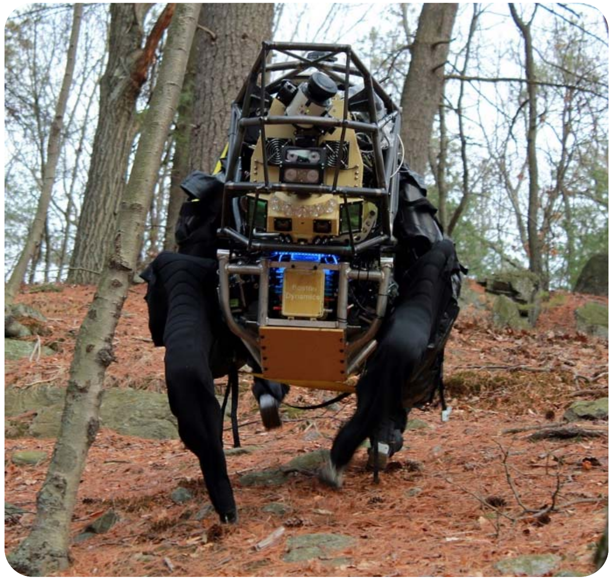


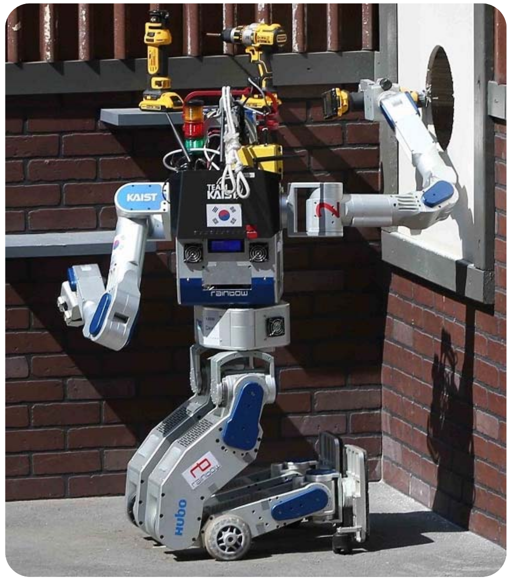


Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

Autonomous
Friendly (humans)
Perception & control
Size/Weight/Power

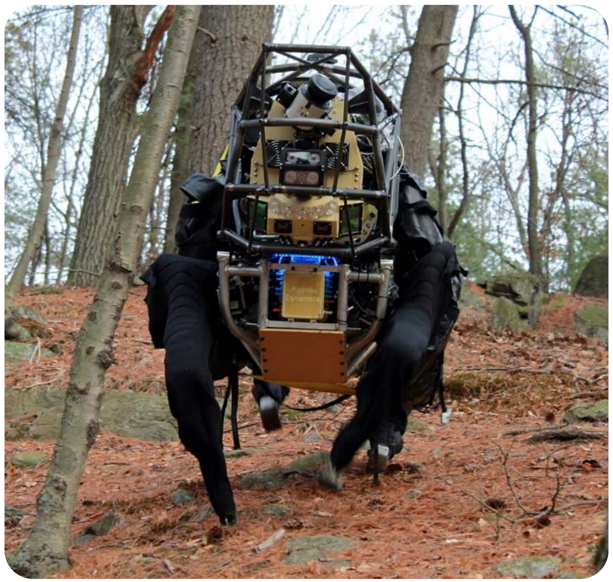




Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

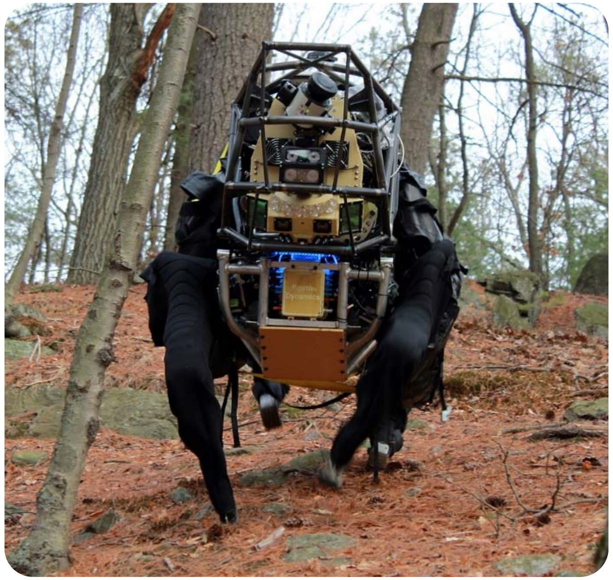
- Autonomous
- Friendly (humans)
- Perception & control
- Size/Weight/Power
- Safety

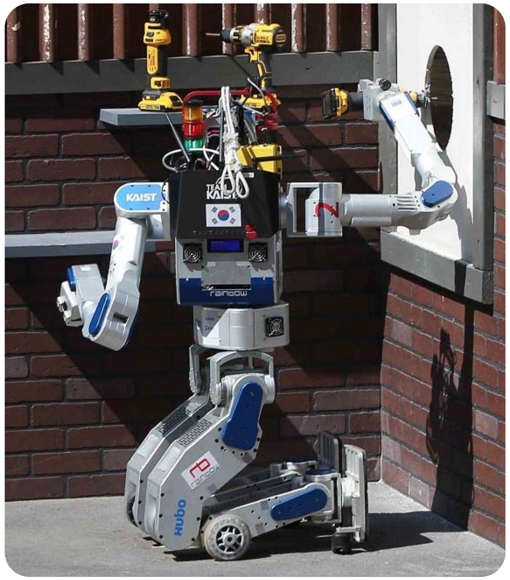


Credits: Boston dynamics, Italian Institue of Technology, Kaist

Motivations

Autonomous
Friendly (humans)
Perception & control
Size/Weight/Power
Safety





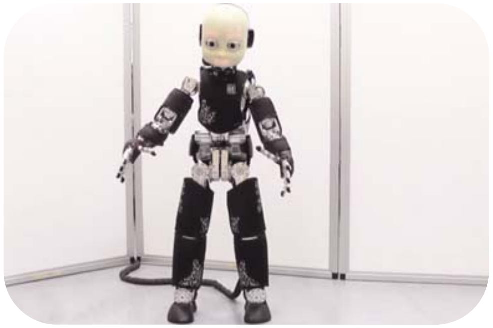
Credits: Boston dynamics, Italian Institue of Technology, Kaist

Research at the iCub Facility

- Engineering
- Research/science

Research at the iCub Facility

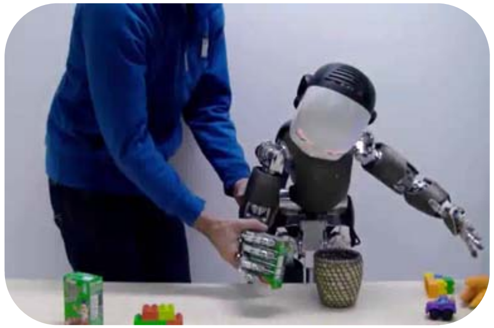
platform



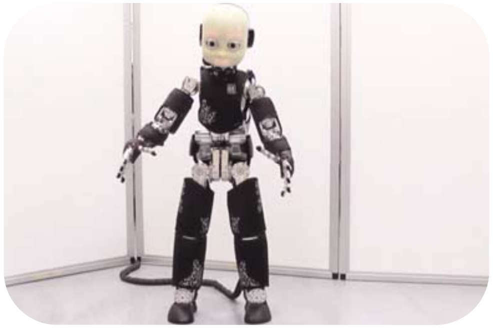
- Engineering
- Research/science

Research at the iCub Facility

interaction



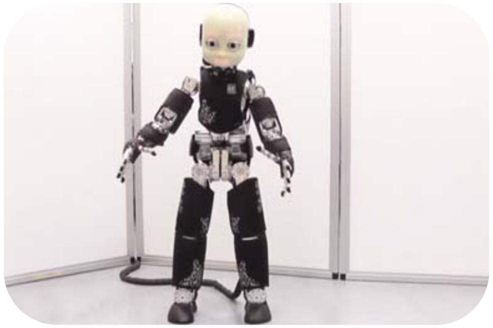
platform



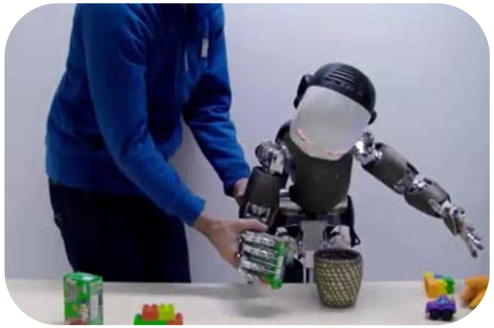
- Engineering
- Research/science

Research at the iCub Facility

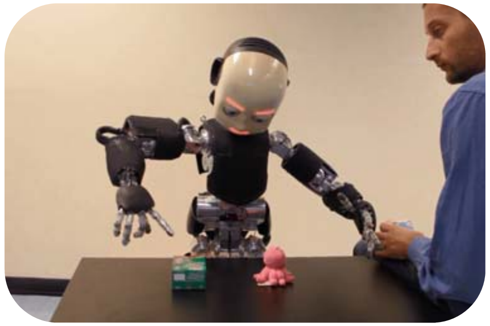
platform



interaction



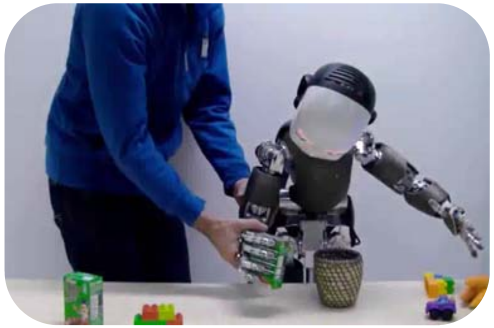
objects



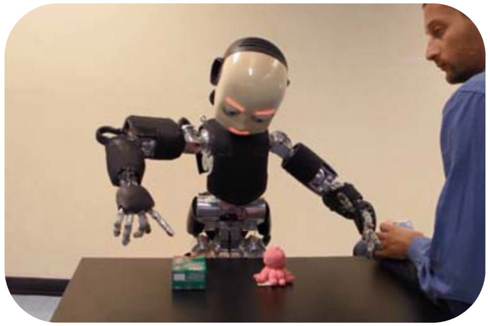
- Engineering
- Research/science

Research at the iCub Facility

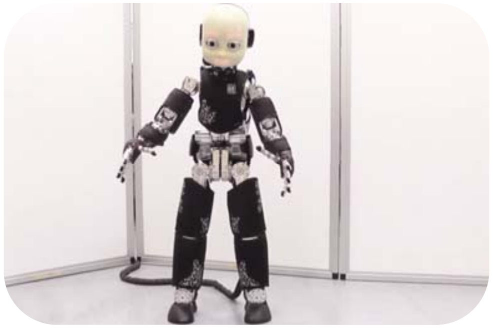
interaction



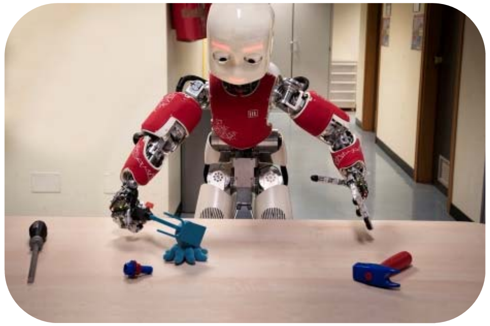
objects



platform



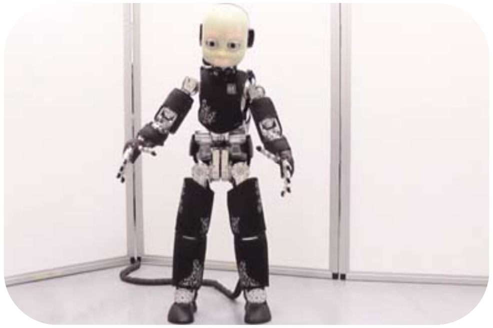
tools



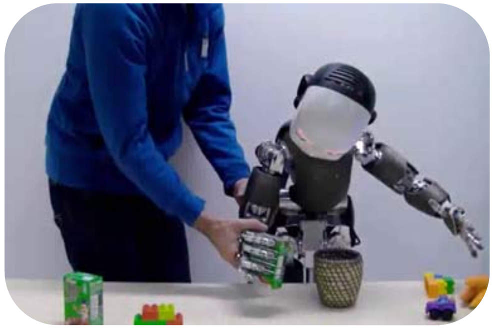
- Engineering
- Research/science

Research at the iCub Facility

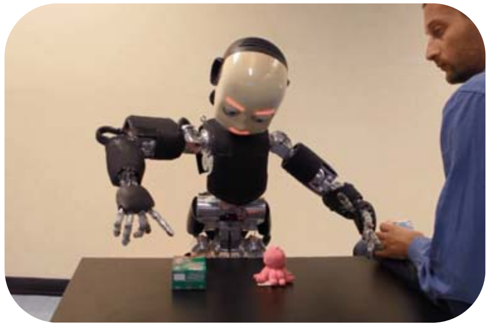
platform



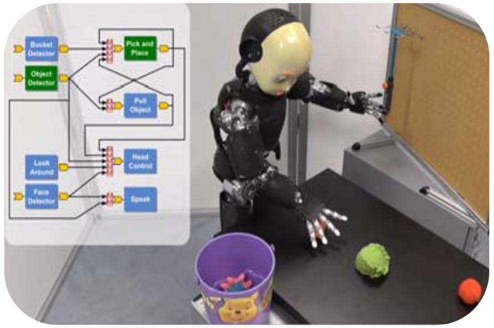
interaction



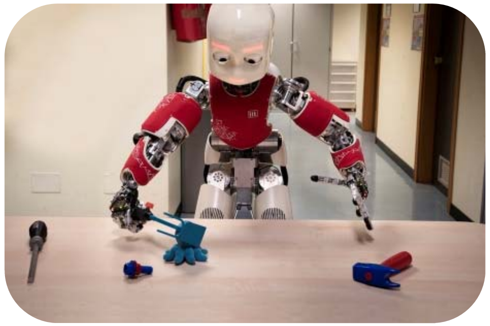
objects



system integration



tools



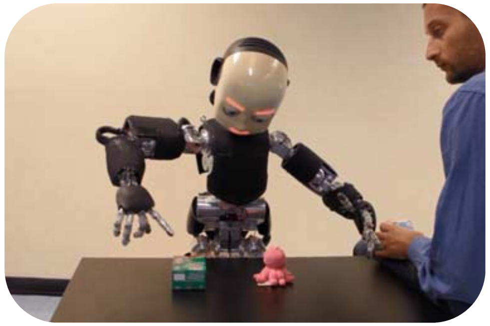
- Engineering
- Research/science

Research at the iCub Facility

platform

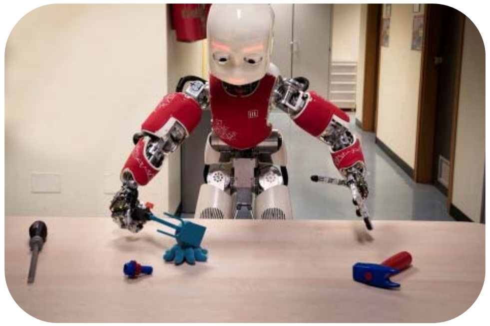
interaction

objects



system integration

tools

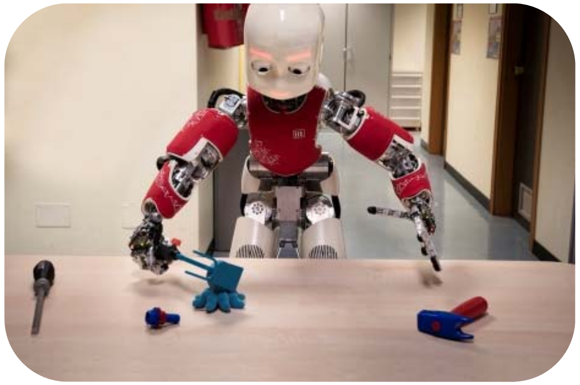


- Engineering
- Research/science

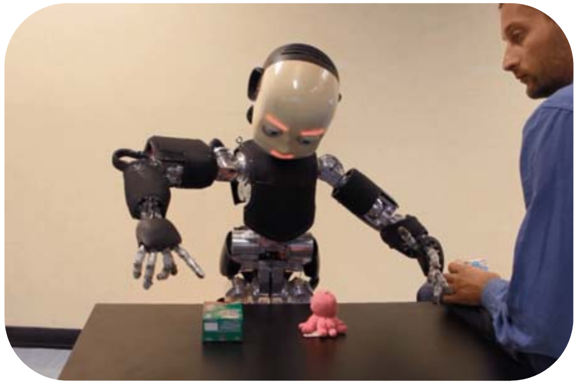
Learning

- Autonomous
- Continuous, online, incremental
- Multimodal, exploit interation with the environment

Tools

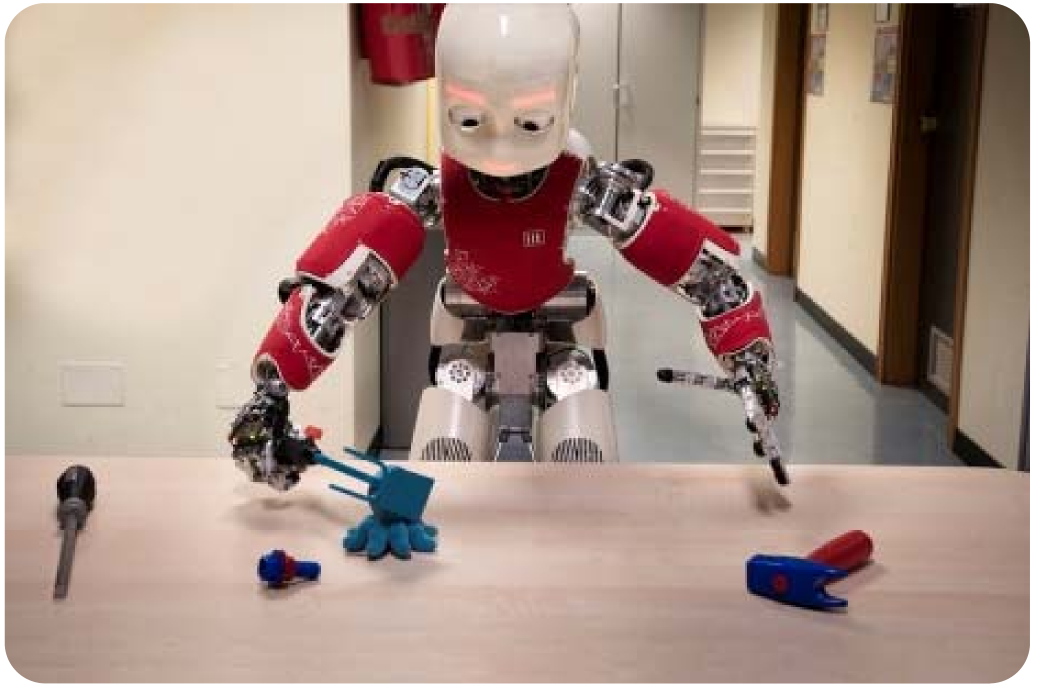


Objects



ISTITUTO ITALIANO
DI TECNOLOGIA

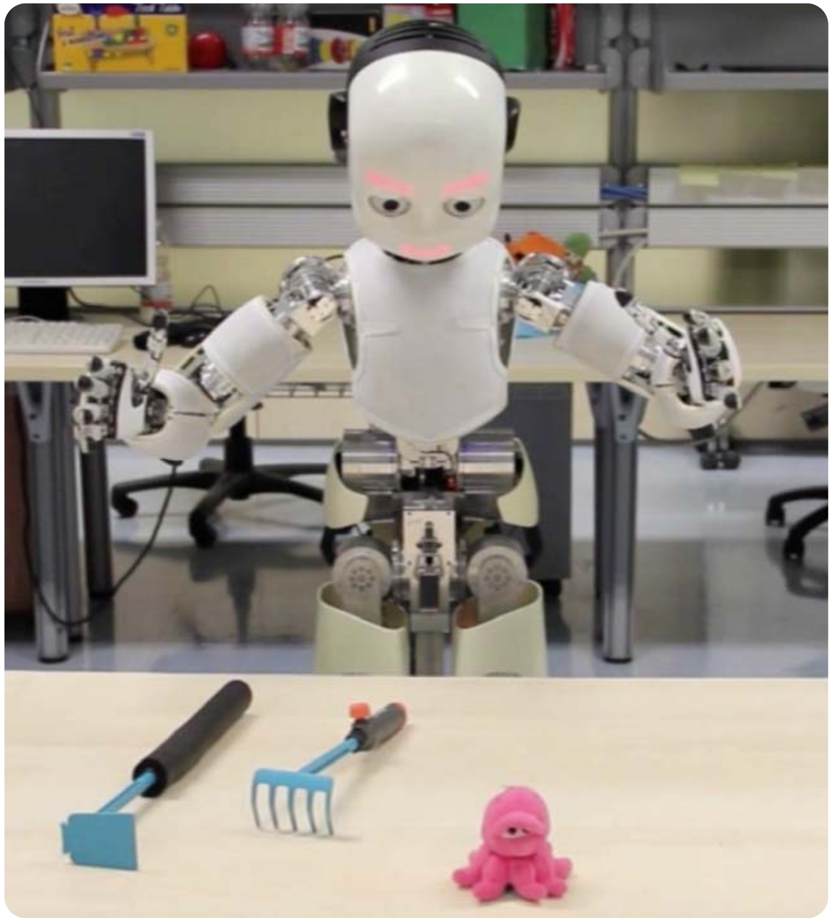
Tools



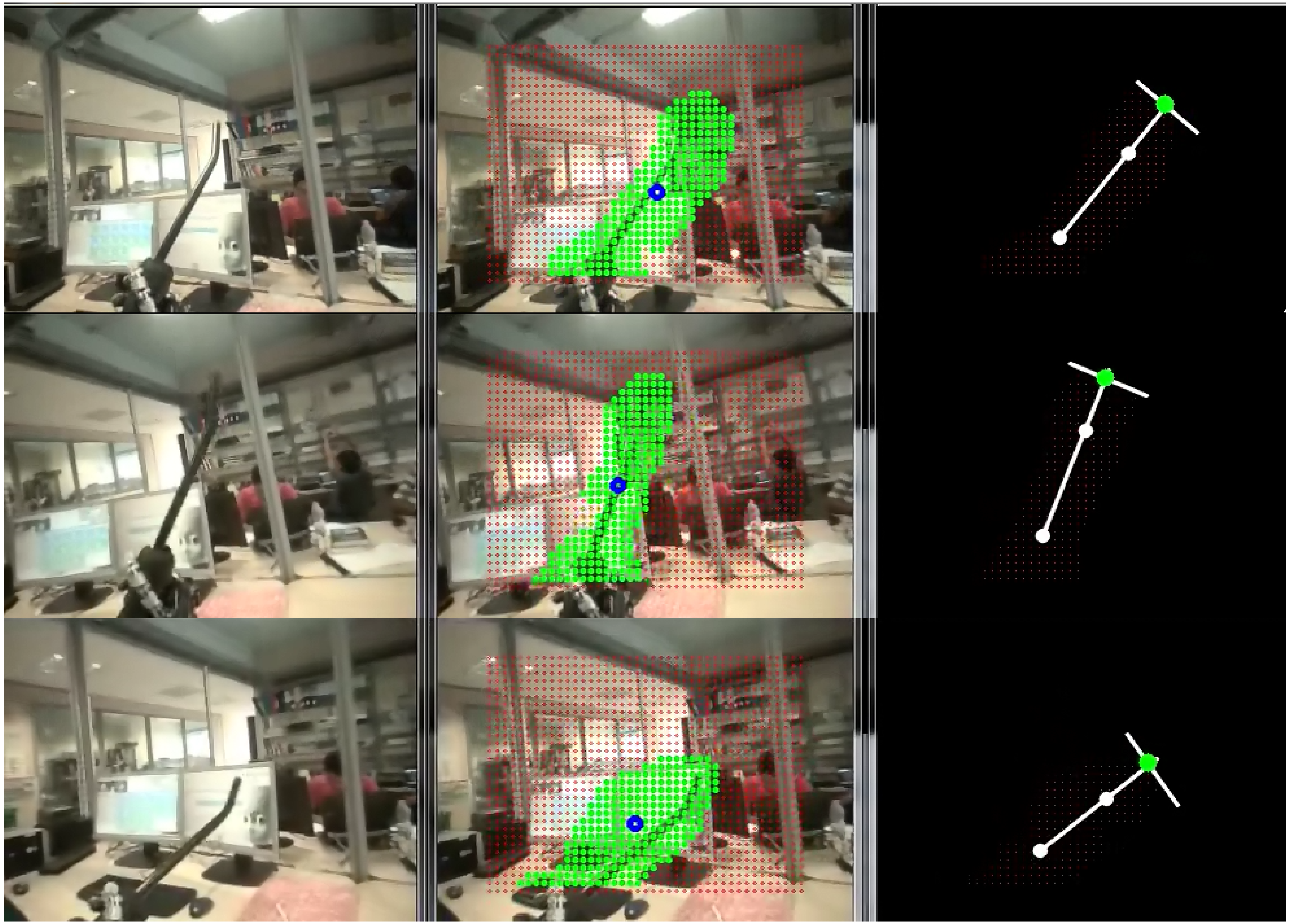
ISTITUTO ITALIANO
DI TECNOLOGIA

Exploring Affordances

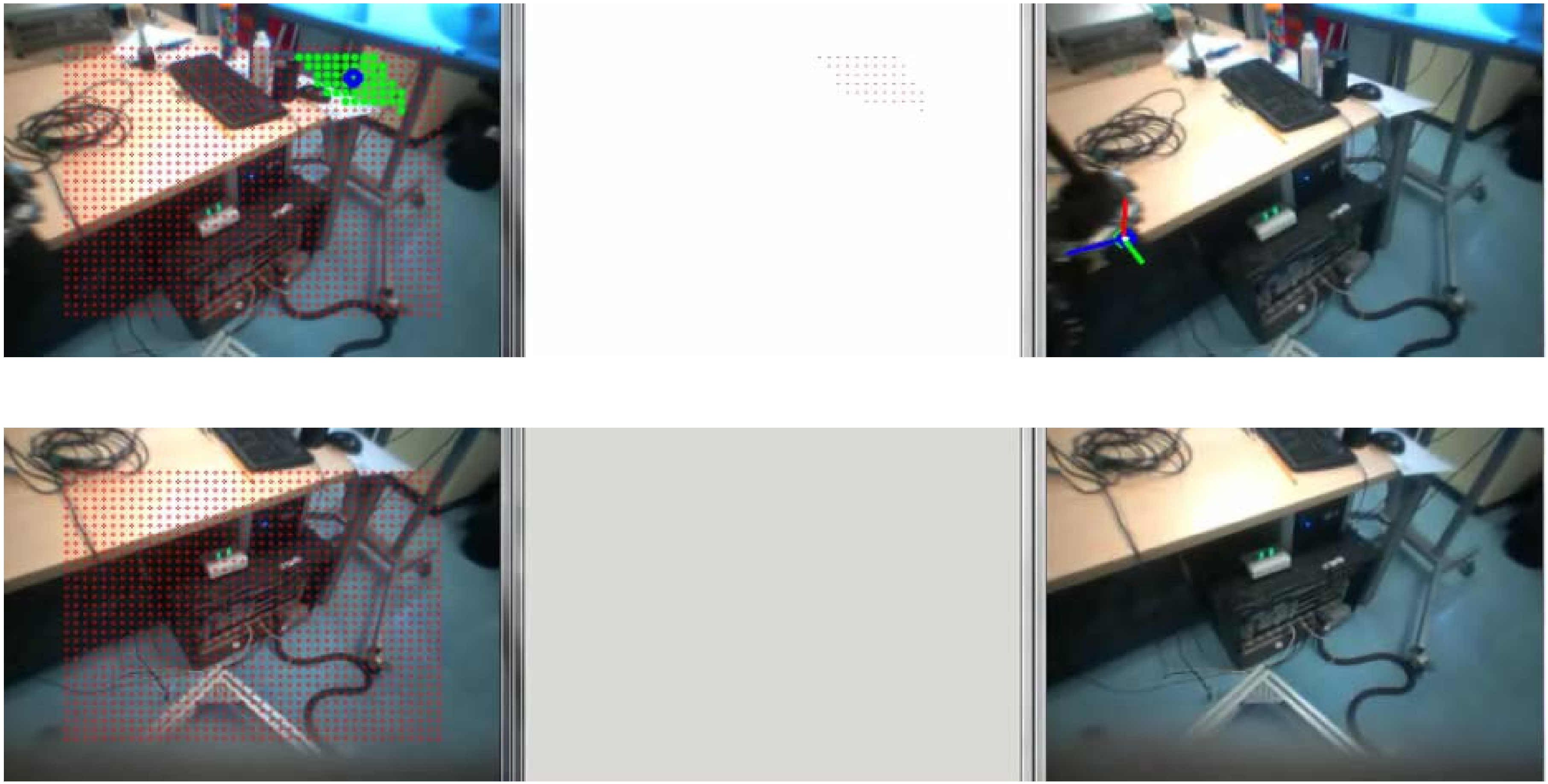
- Self-supervised learning of pulling actions
- Exploring tool size
- Exploring tool affordances

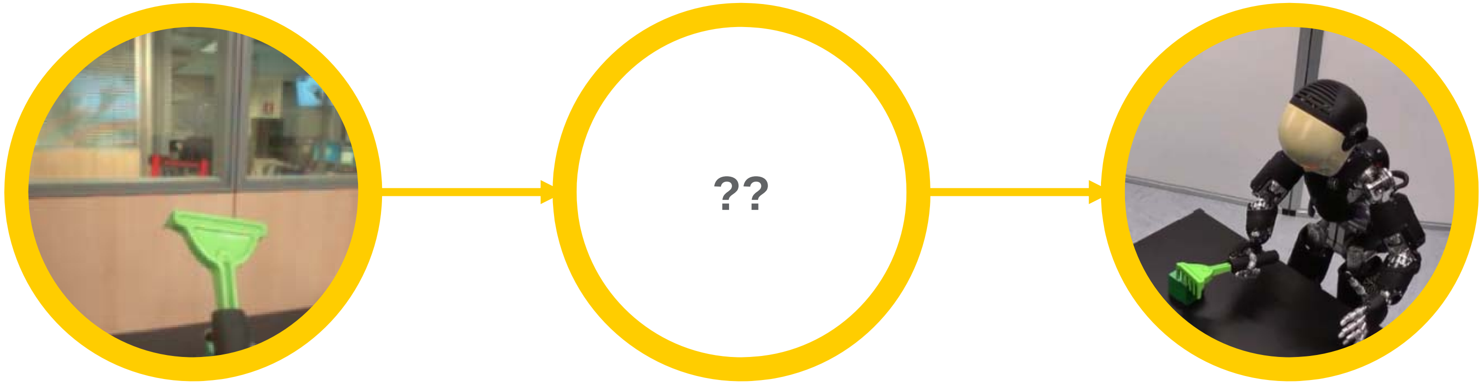


Exploring tool size



Exploring tool size





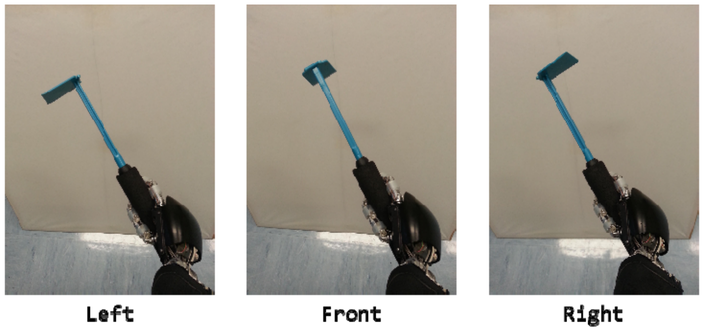
Geometrical features

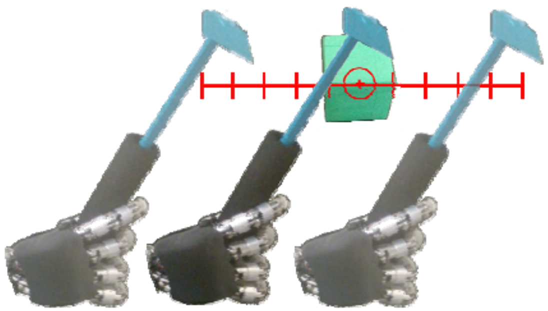
Discover categories

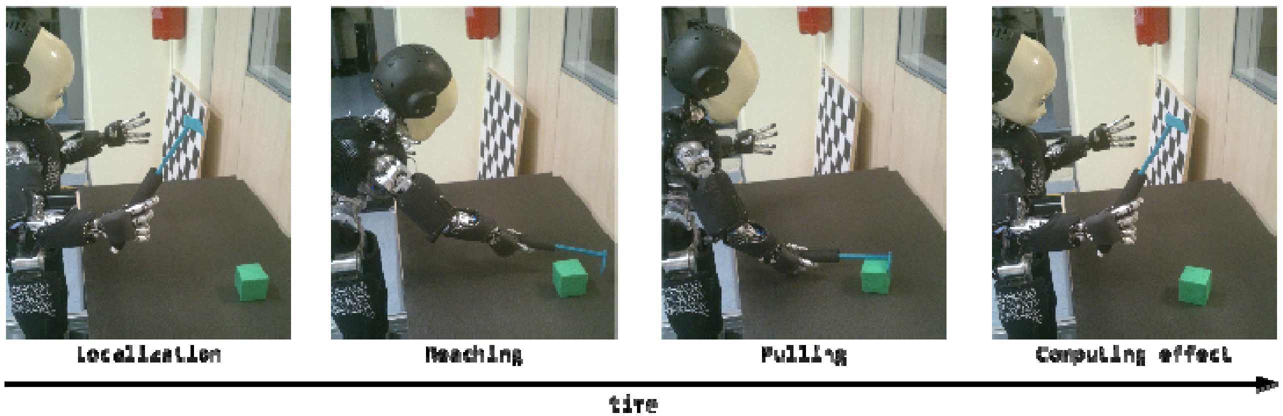
Best action

Exploring Affordances

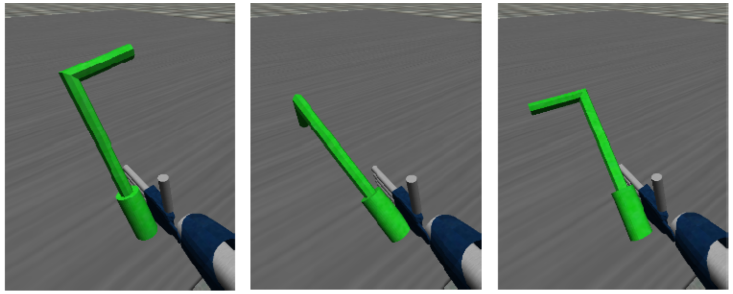
- Learn effect of pulling actions
- Depends on tool and tool pose



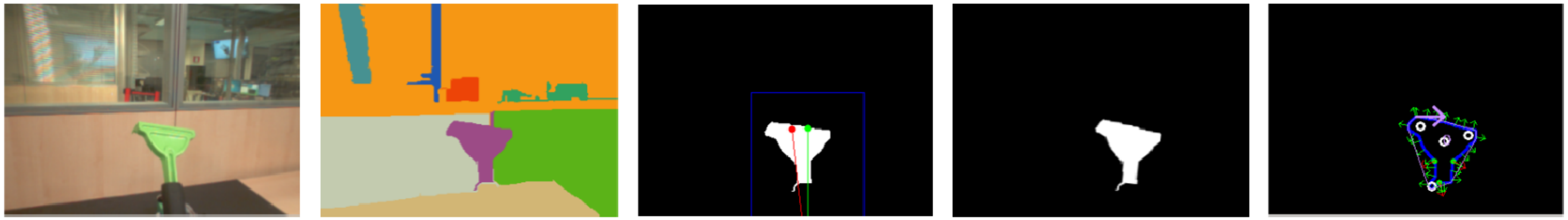




Characterizing the tool



Left Front Right

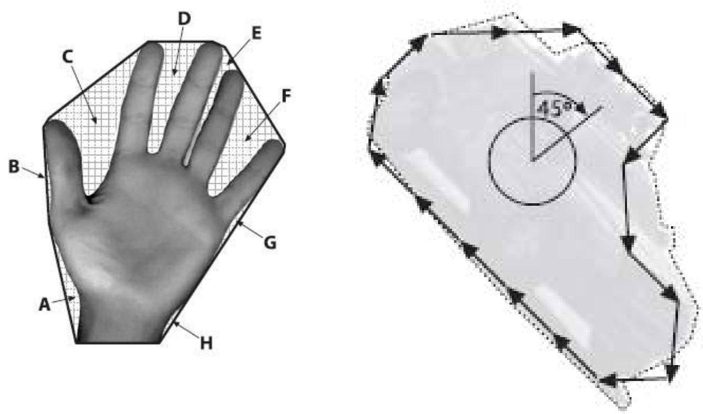


Camera Image | Graph Based Segmentation | Perspective Normalization | Crop Region of Interest | Feature Extraction

Processing Stages

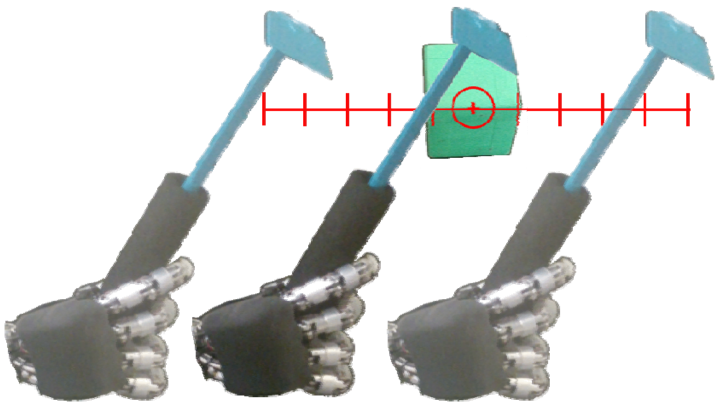
Details: considered features

- **Based on convex hull**
  - Depth of the 5 larger convexity defects
  - Histogram of bisector angles at convexity defects
  - Area of the convex hull
  - Solidity
- **Based on thinning**
  - Number of skeleton bifurcations to the left, right, under and above
  - Number of skeleton endings to the left, right, under and above the blob's center of mass
- **Based on moments**
  - Normalized central moments
- **Shape descriptors**
  - Area, perimeter, compactness
  - Major principal axis (length), Minor principal axis (width)
  - Aspect ratio, Extension, Elongation, Rectangularity
- **From the angle signature**
  - Bending energy (sum of squares of the angle variation along the contour, divided by the number of points in their contour)
  - Angle signature histogram
- **Signature of distance contour to centroid**
  - Fourier coefficients
  - Wavelet coefficient

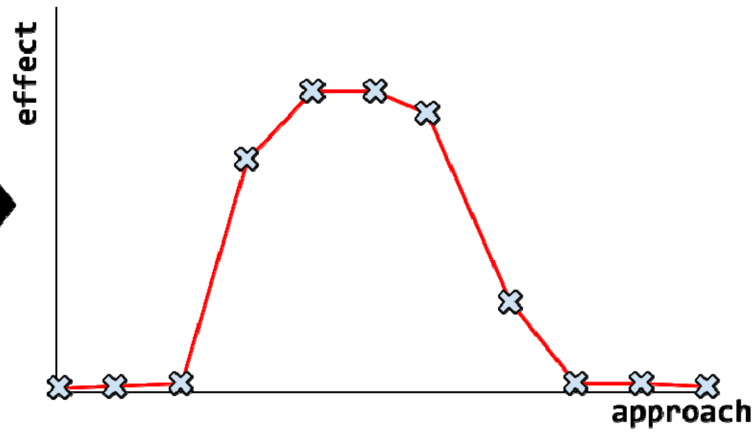


Characterizing Effect

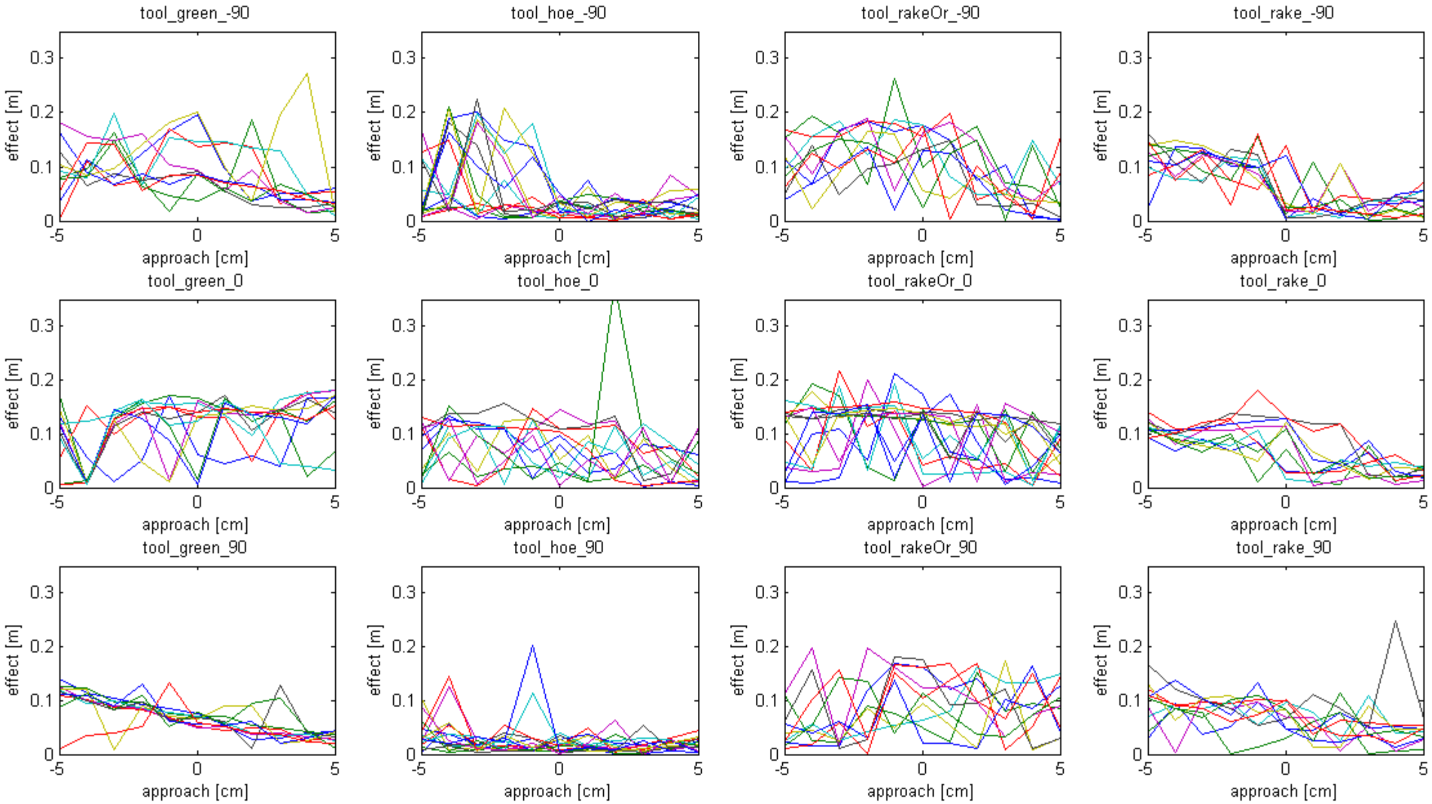
- How close is the object to the robot after the action, given tool position w.r.t object →affordance vector

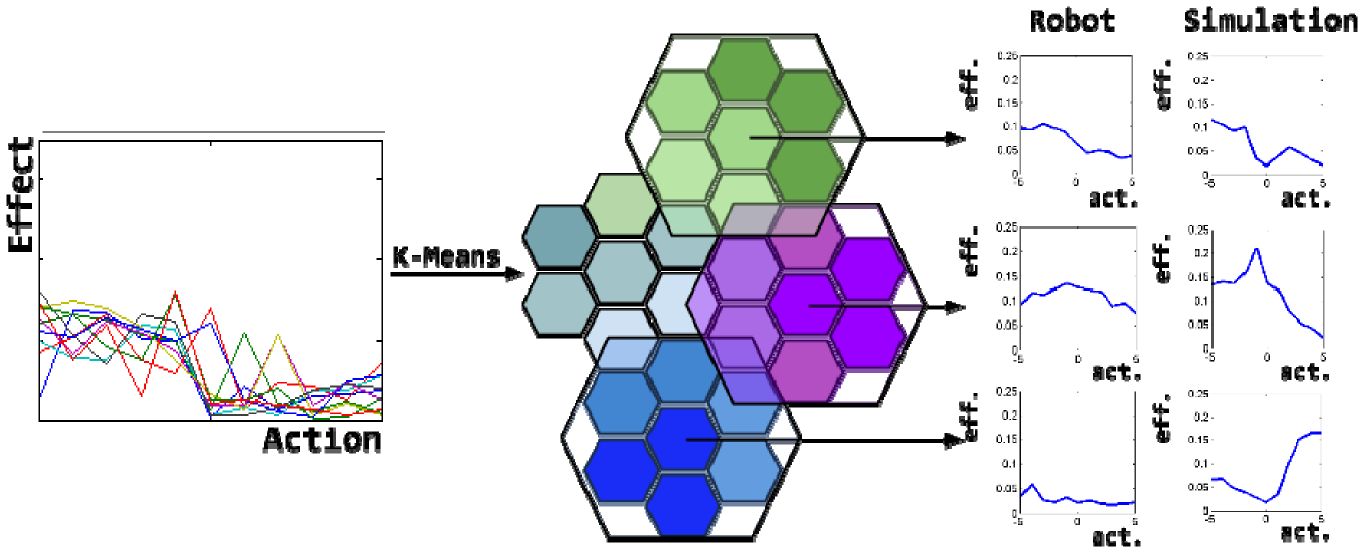


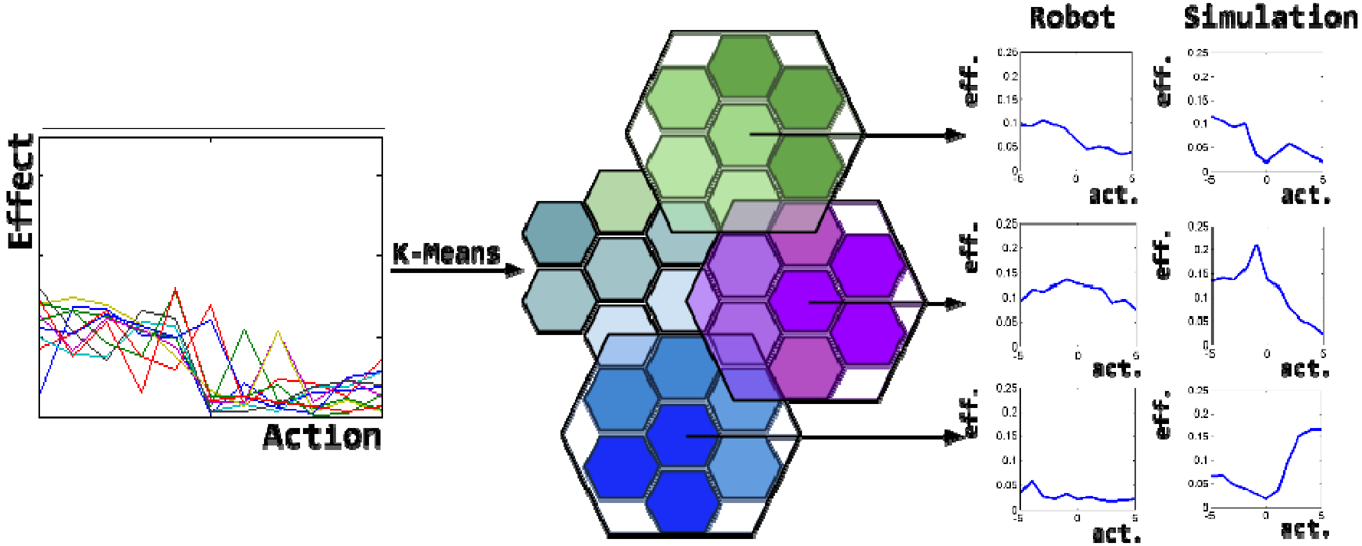
For each orientation & position…

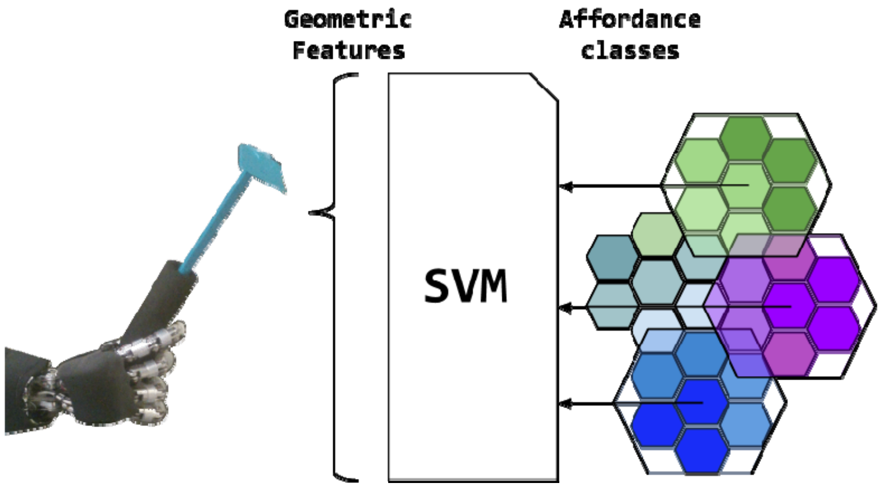


…extract an "affordance vector"

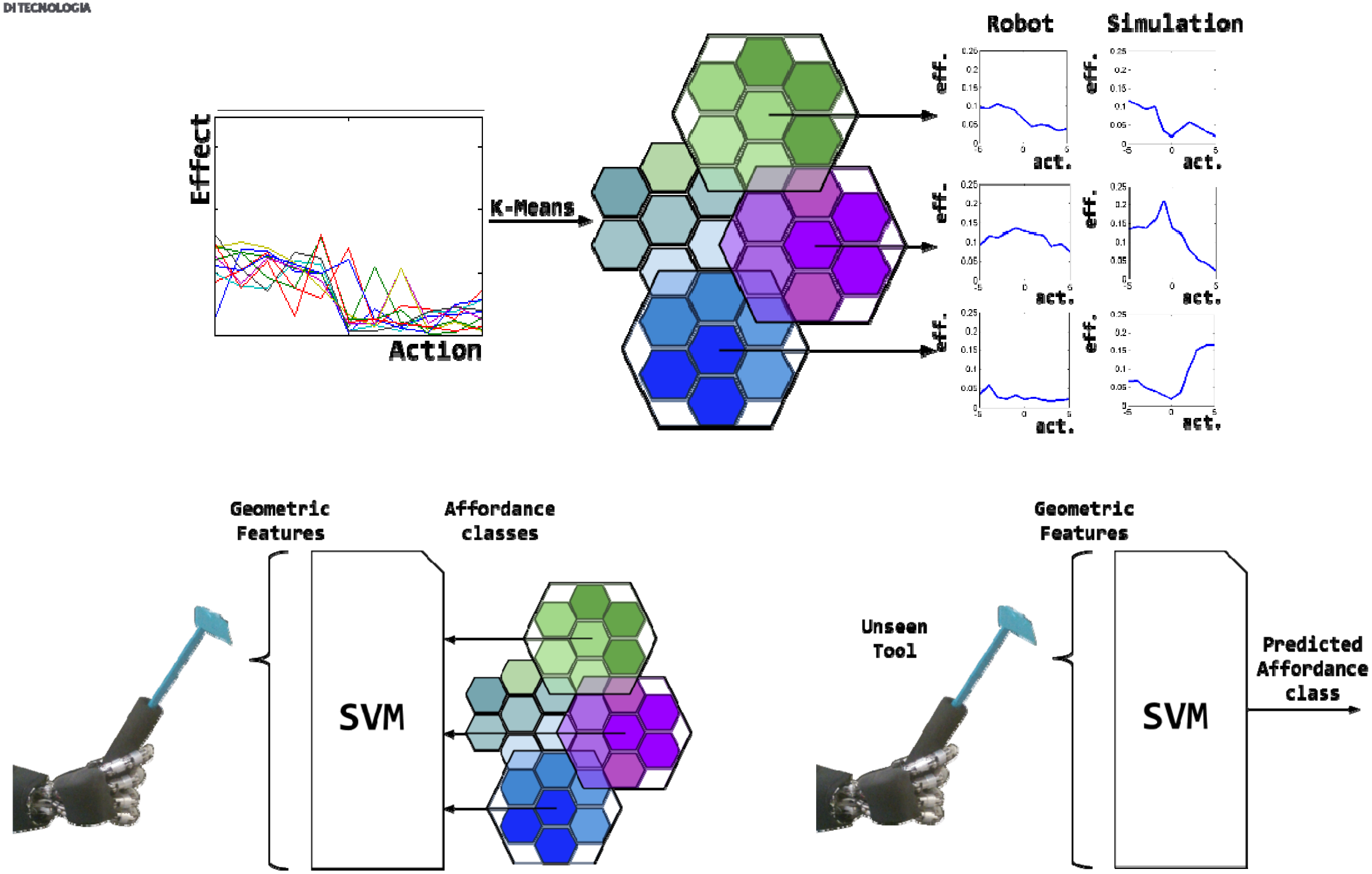

tool_green_-90
tool_hoe_-90
tool_rakeOr_-90
tool_rake_-90
tool_green_0
tool_hoe_0
tool_rakeOr_0
tool_rake_0
tool_green_90
tool_hoe_90
tool_rakeOr_90
tool_rake_90
effect [m]
approach [cm]


Effect
Action
K-Means
Robot
Simulation
eff.
act.


Effect
Action
K-Means
Robot
Simulation
eff.
act.


Geometric Features
Affordance classes
SVM

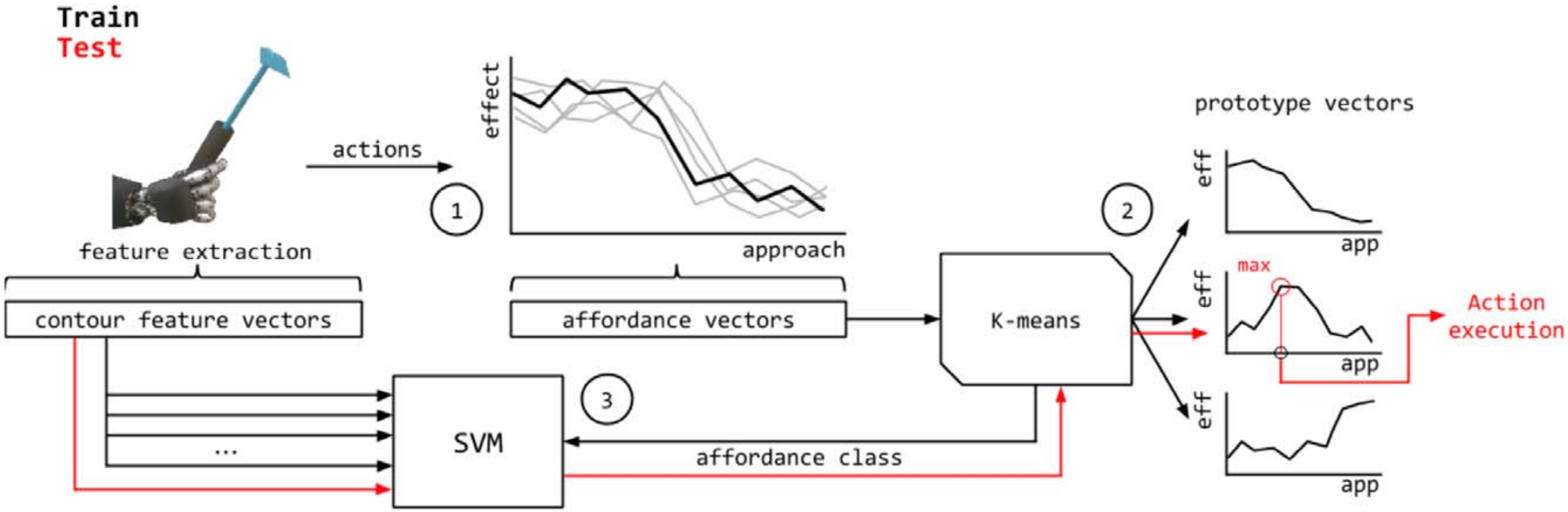
Training



Training

Prediction

Putting it all together



Details of the experiments

- Each trial consists of:
  - 11 pull actions (various approaches from -5 to 5 cm to either side of the object)
  - The 11 pairs action-effect represent an affordance vector which describe how well a particular tool-pose affords pulling as a function of the approach position w.r.t the object
    - Between 20 and 25 of such affordance vectors have been recorded in simulation
    - And 10 vectors for each of the tool-poses on the real robot

total of 567 vectors (6237 pulls) on simulation and

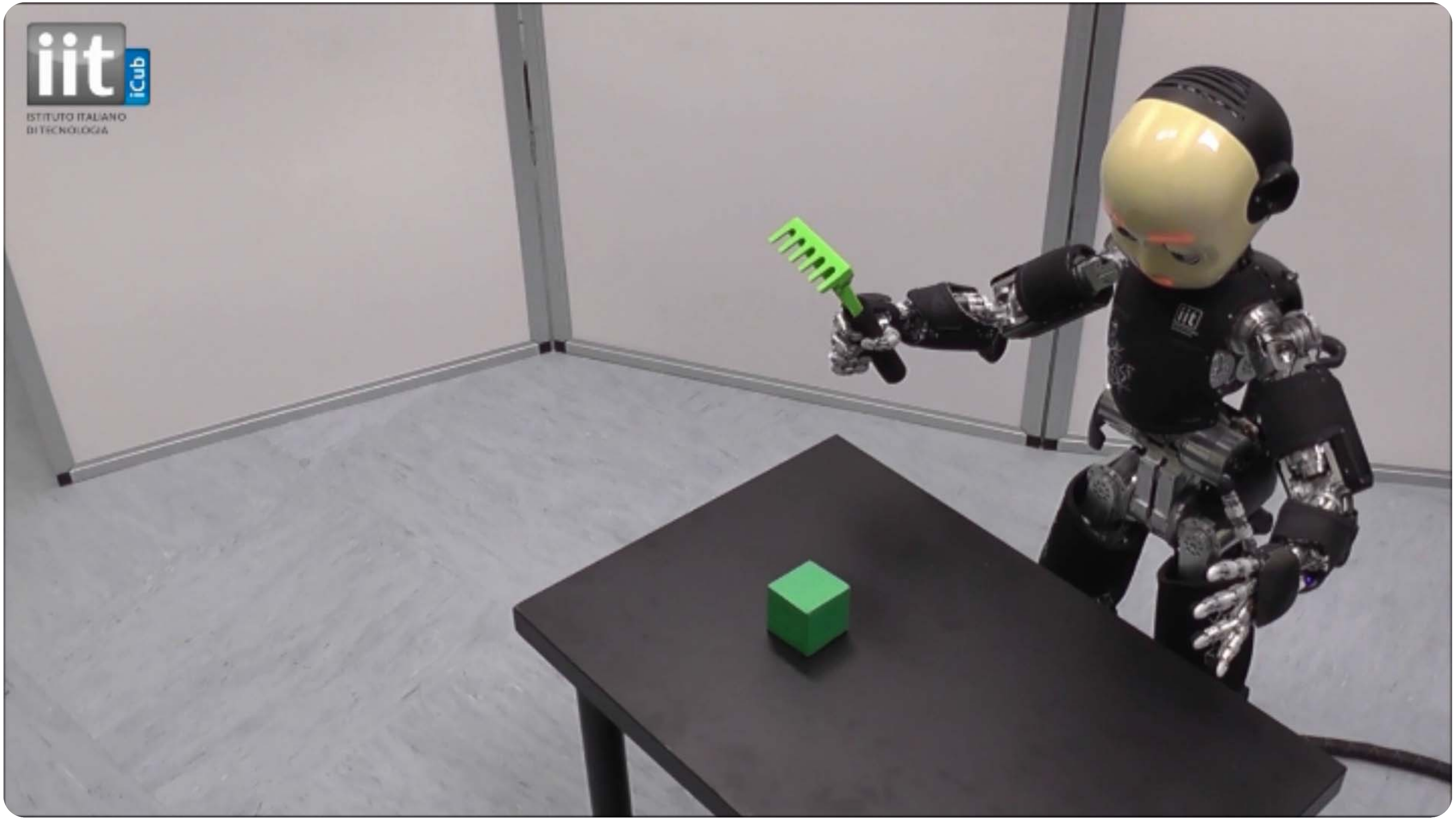
138 vectors (1518 pulls) on the real robot

Results

Test	Env.	Class. Acc. (%)	rMSE [m]
Prediction	Sim.	81.9 %	0.064
Prediction	Robot	64.1 %	0.051

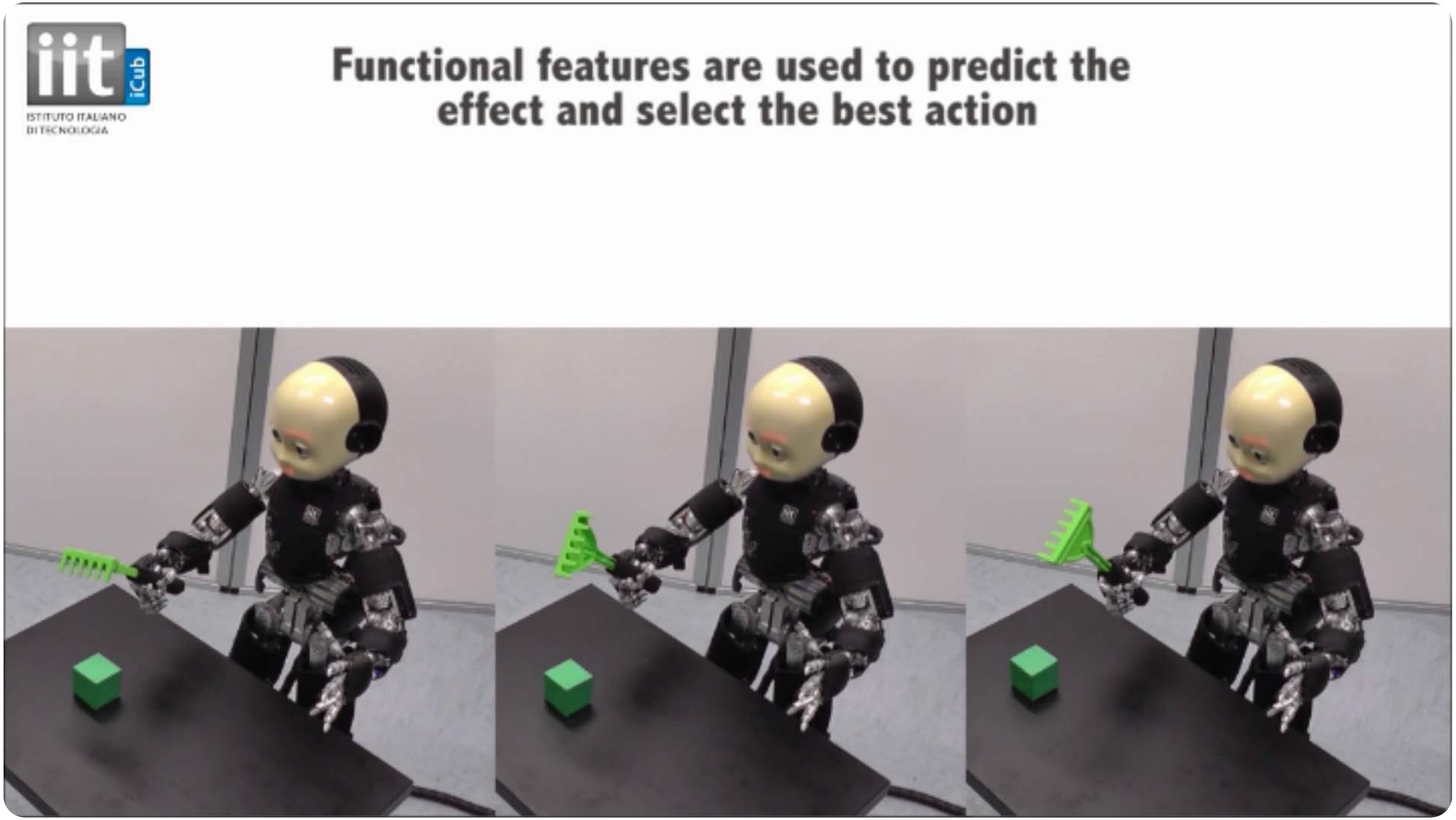
Environment	Goal Acc. (%)	Avg. Diff [m]
Simulation	86.51 %	0.064
Robot	86.11 %	0.056

Video: exploration



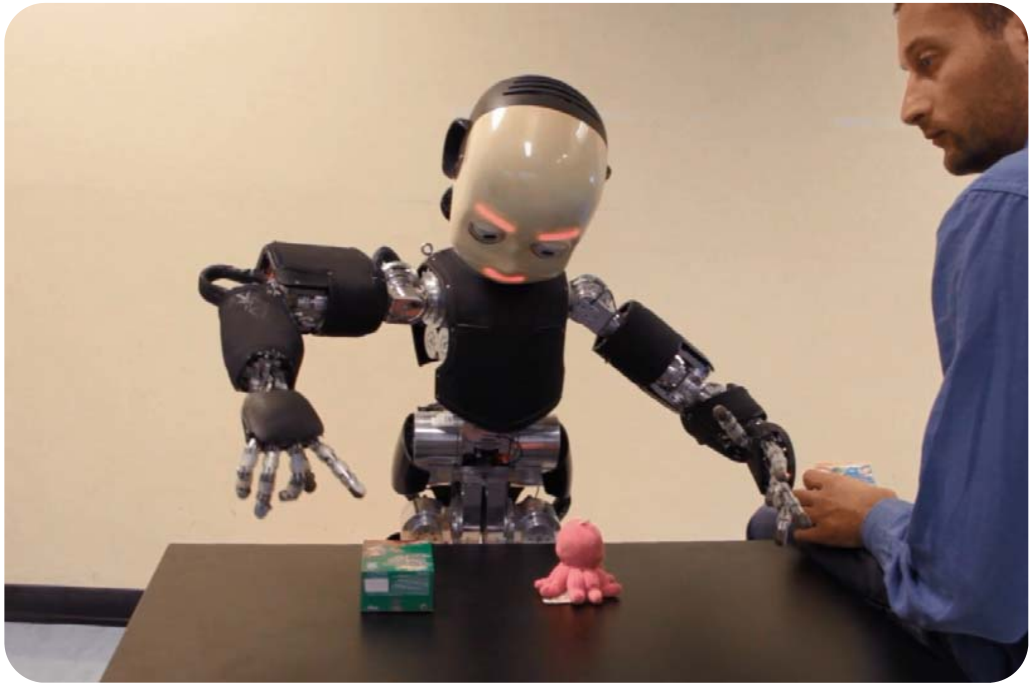
Mar et al. IROS 2015, Humanoids 2015

Video: prediction

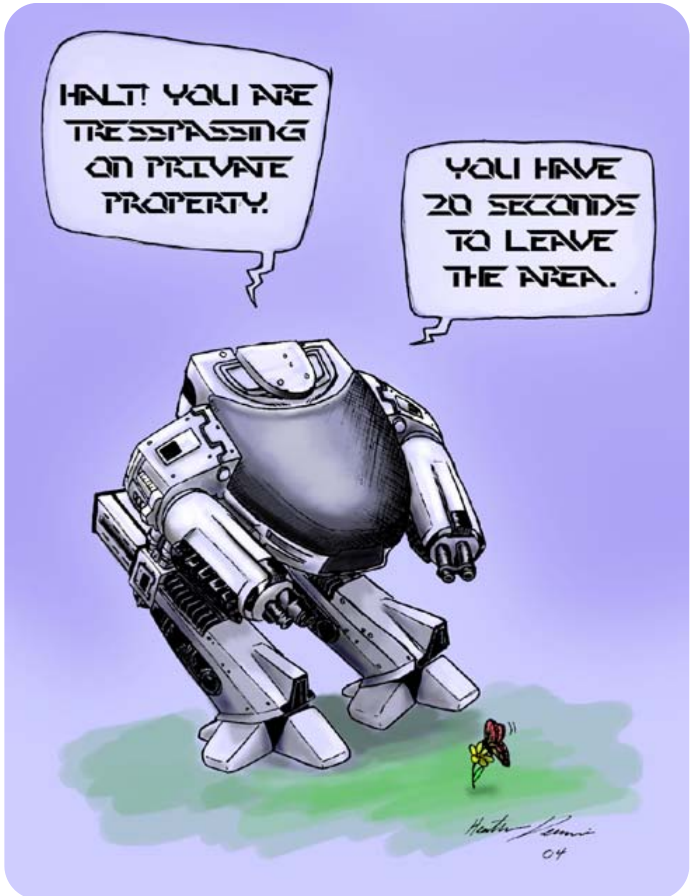


Mar et al. IROS 2015, Humanoids 2015

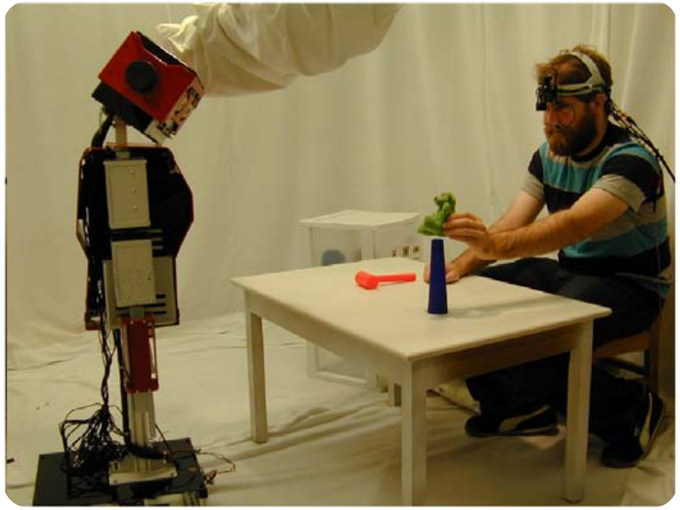
Objects

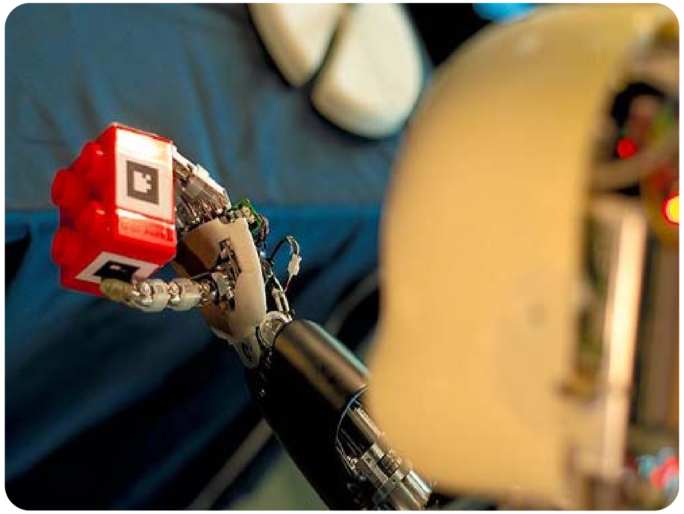


ISTITUTO ITALIANO
DI TECNOLOGIA

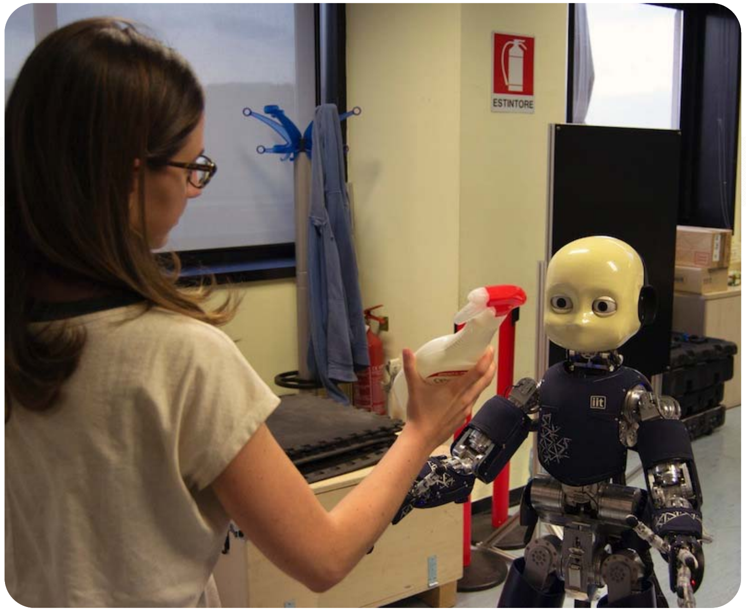


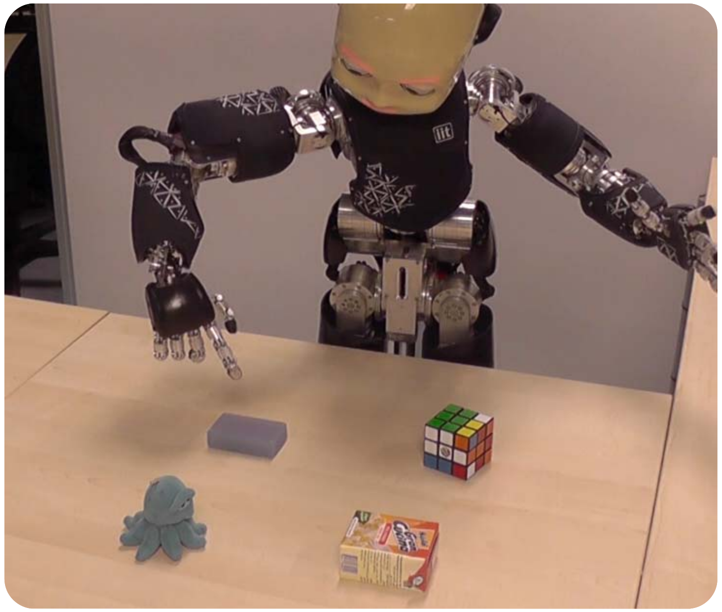
Vision in robotics





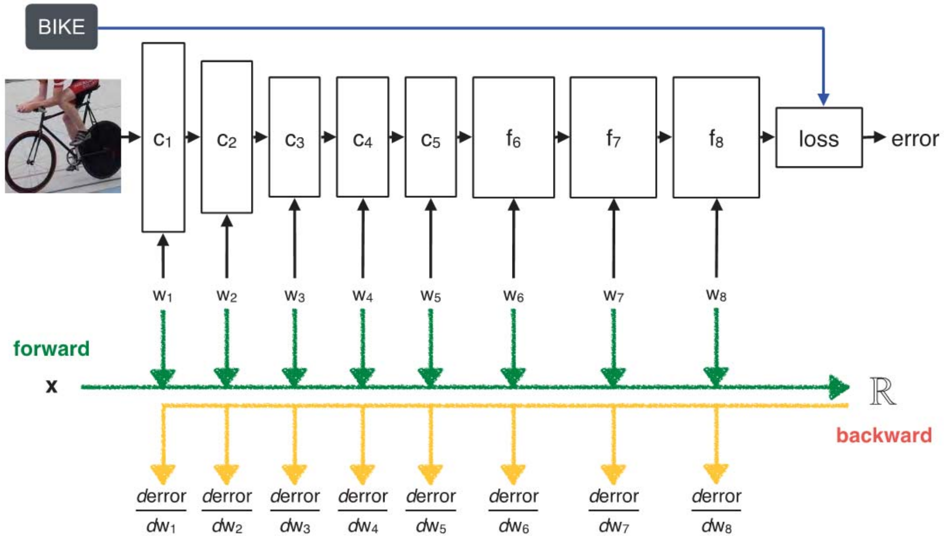
Autonomous learning





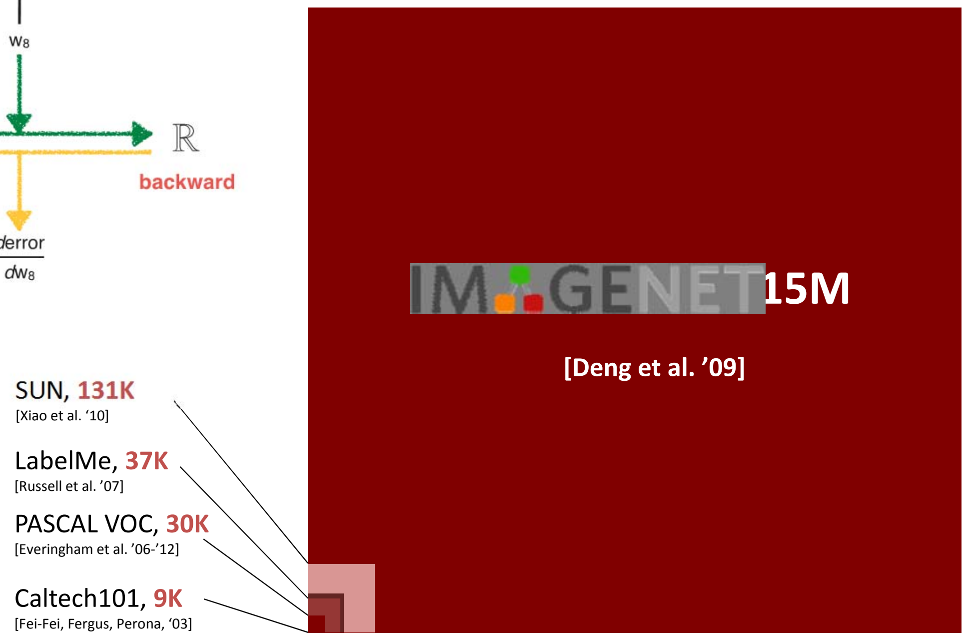
Breakthrough in Computer Vision

DEEP NETWORKS (GPU)



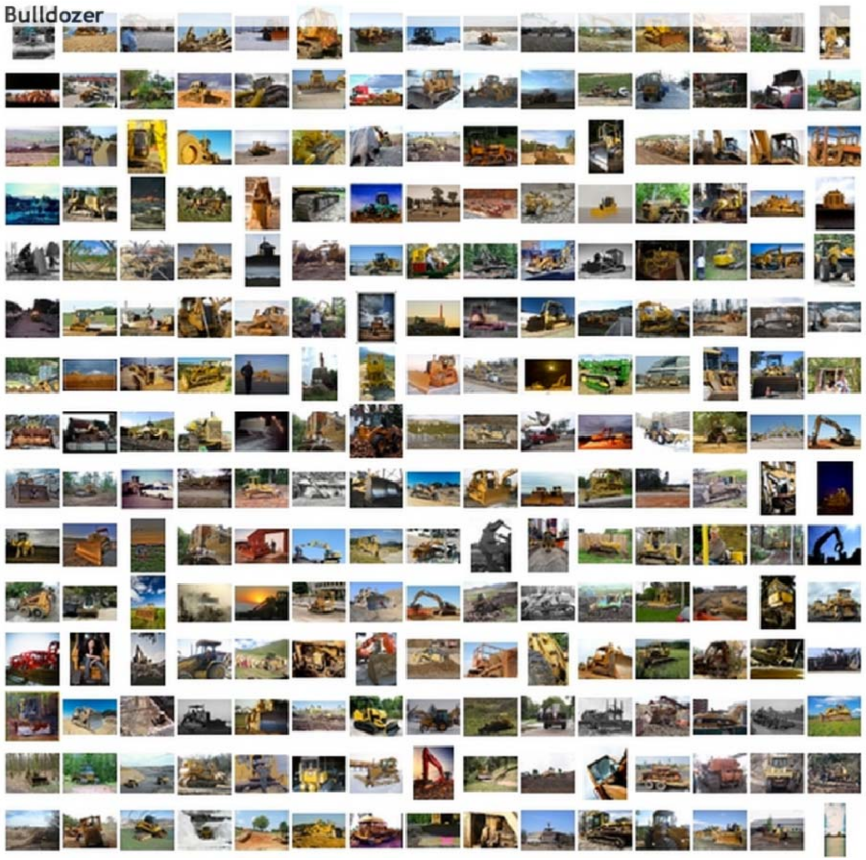
Credits: A. Vedaldi

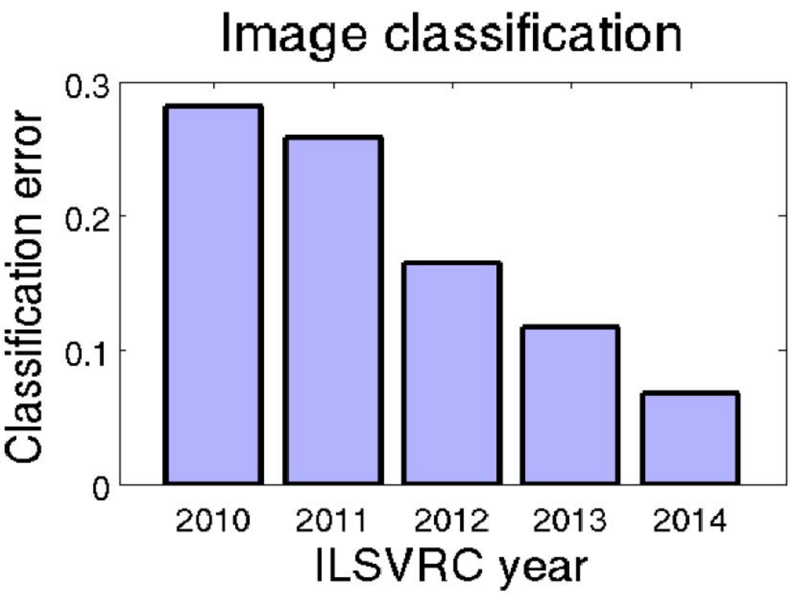
Large Datasets



Credits: Fei-Fei Li

Computer Vision

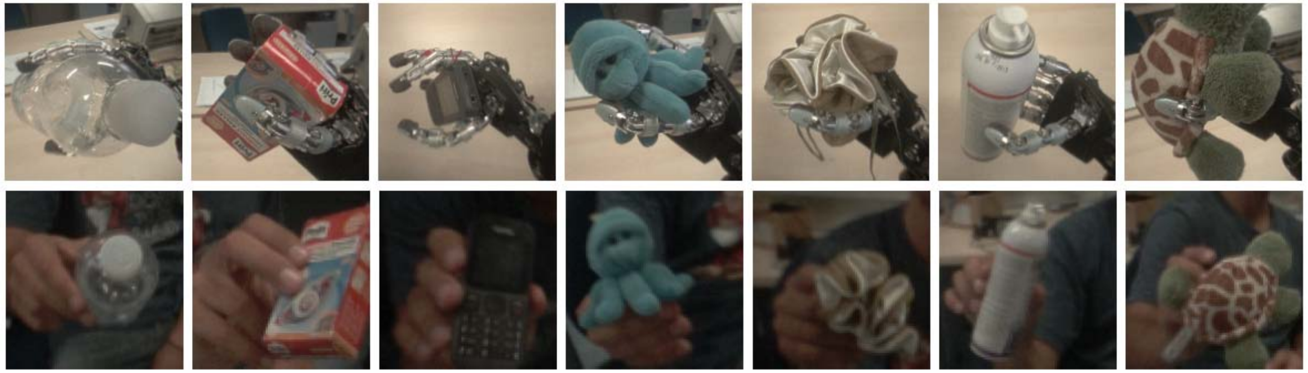




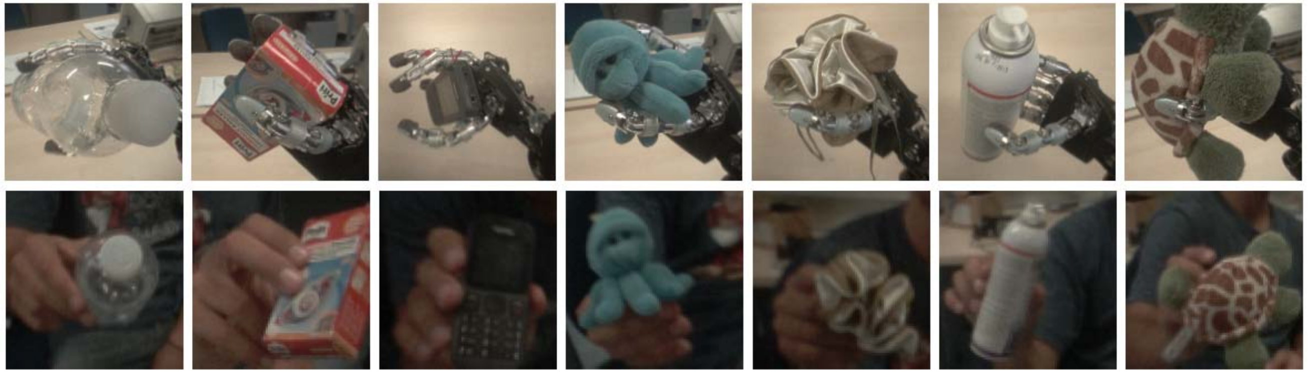
Approaching human performance
on the same dataset!

Russakovsky et al. 2015

The visual experience of a robot

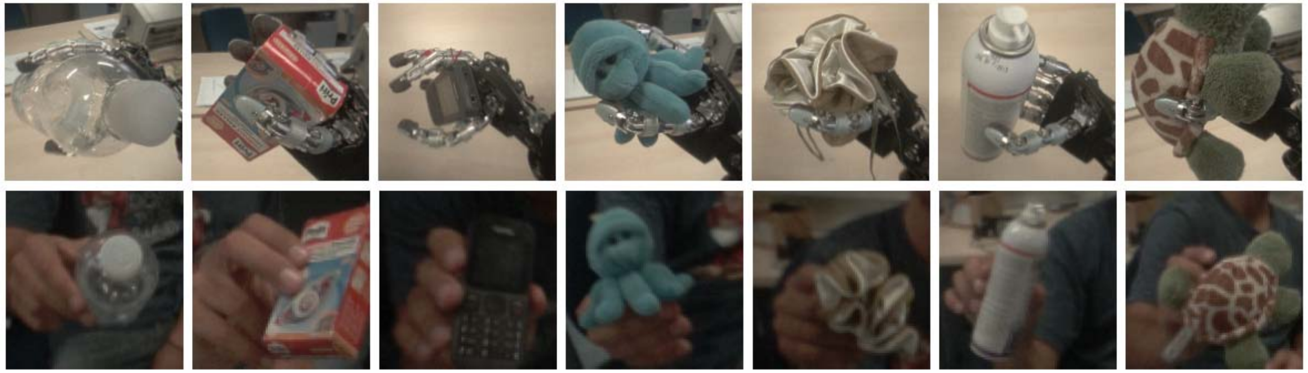


The visual experience of a robot



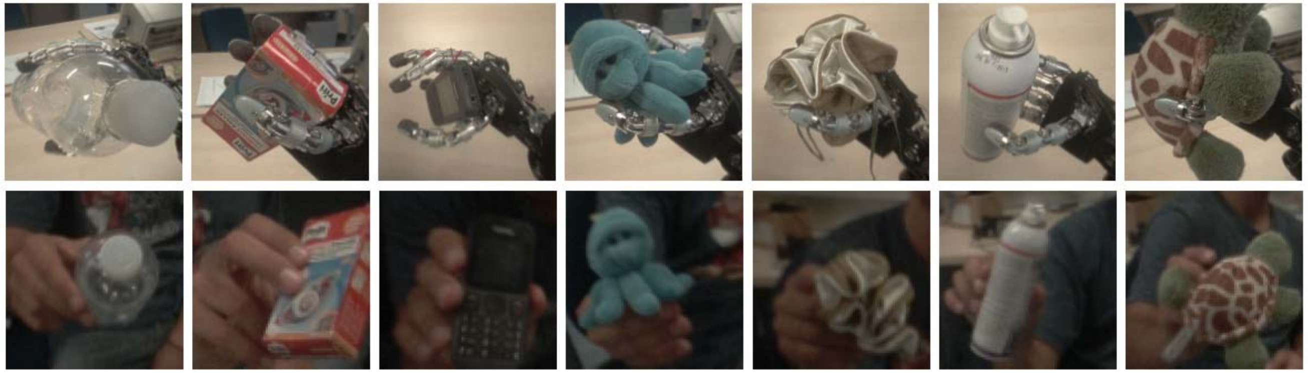
- Supervision is expensive and inaccurate

The visual experience of a robot



- Supervision is expensive and inaccurate
- Need for online learning

The visual experience of a robot

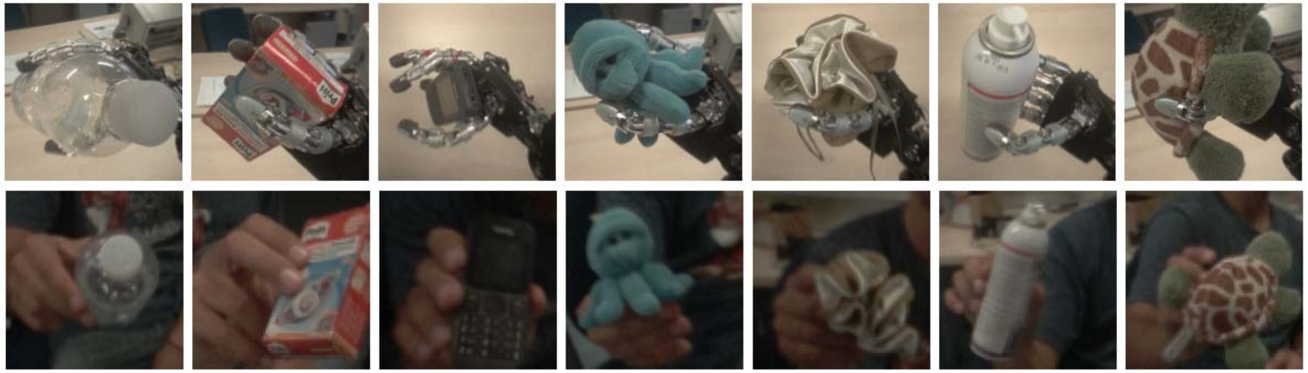


- Supervision is expensive and inaccurate
- Need for online learning
- Objects: large variability (scale, viewpoint)

The visual experience of a robot

- Supervision is expensive and inaccurate
- Need for online learning
- Objects: large variability (scale, viewpoint)
- Background: little variability

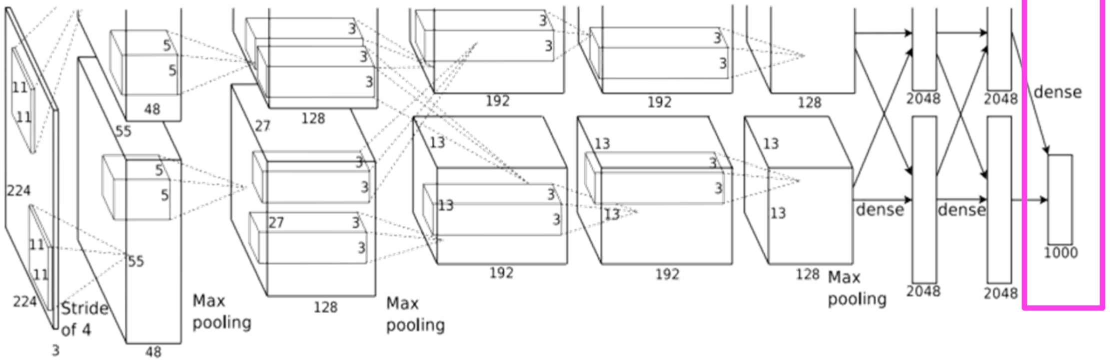
The visual experience of a robot



- Supervision is expensive and inaccurate
- Need for online learning
- Objects: large variability (scale, viewpoint)
- Background: little variability
- Limited resolution

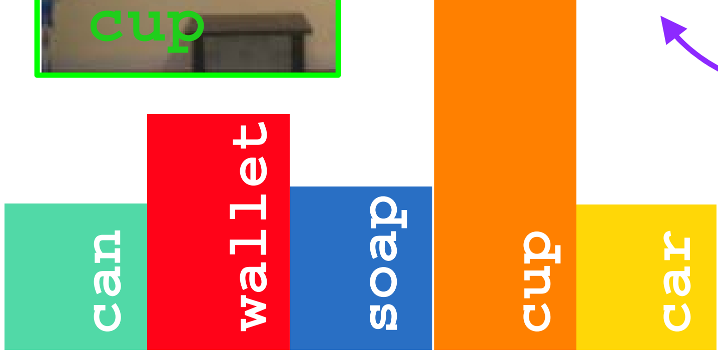
Methods

Deep Convolutional Network



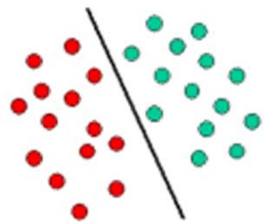
Krizhevsky «Ale» network [Caffe BVLC Reference *CaffeNet*]

code



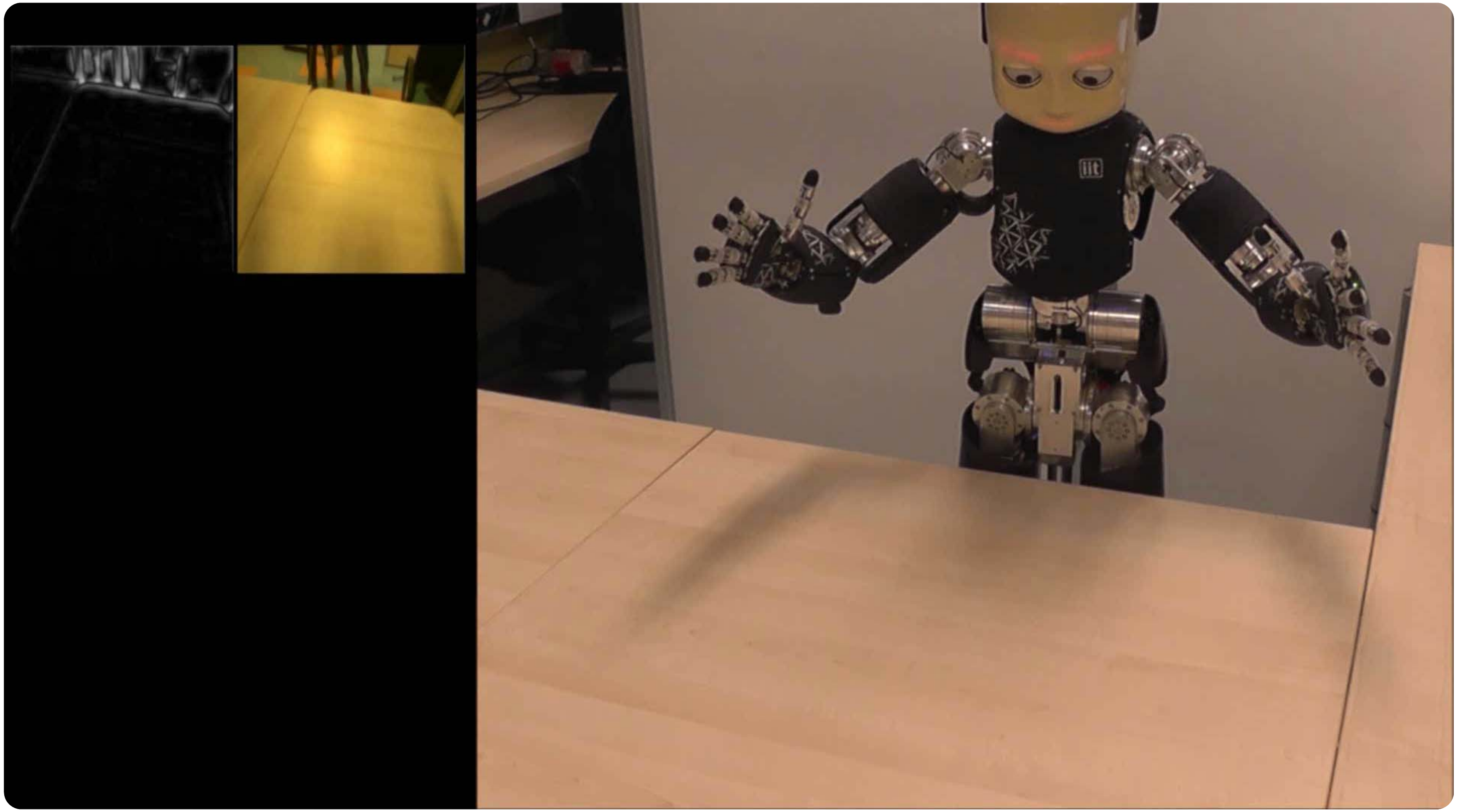
scores

Linear Classifier



RLS [GURLS]

An initial evaluation



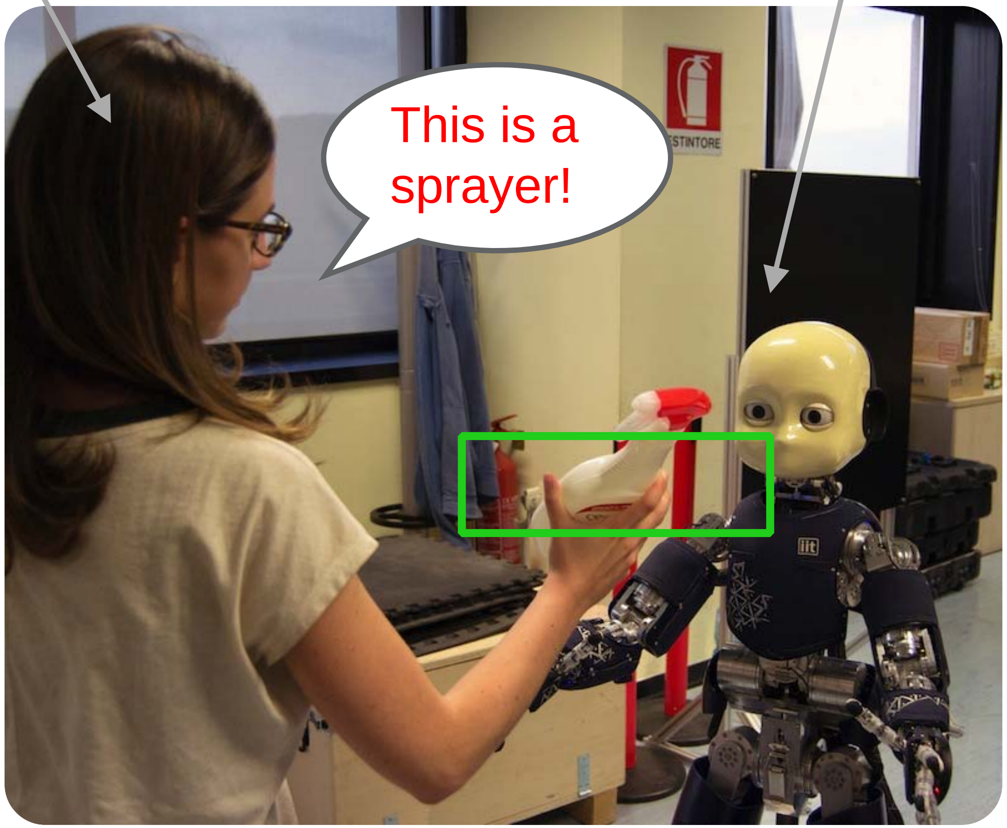
Some questions

- To what extend does clutter affect performance?
- Scalability. How do iCub recognition capability decrease as we add more objects to distinguish?
- Can we use assumptions on physical continuity to make recognition more stable?
- Incremental Learning. How does learning during multiple sessions affect the system recognition skills?
- Generalization. How well does the system recognize objects “seen” under different settings?

On the fly recognition

Verbal instructions
of a “teacher”

Robot’s attention (motion/disparity)

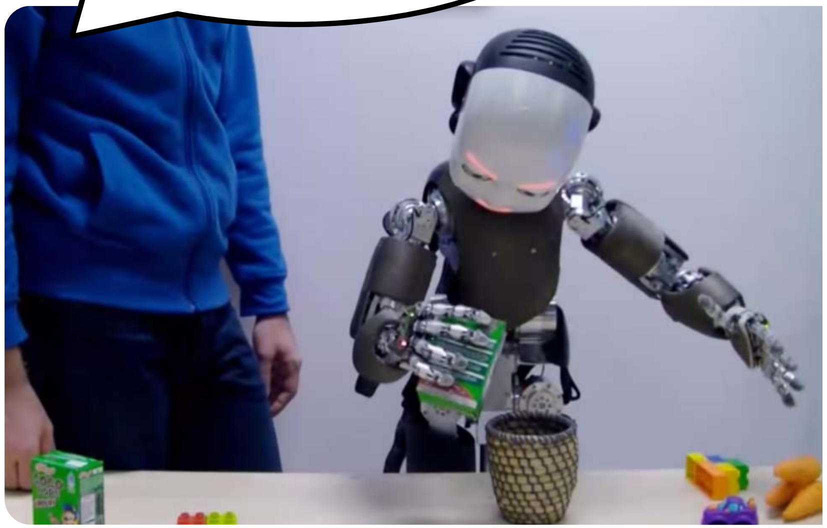


Benchmarking the iCub visual system
iCubWorld Dataset

***Enabling Depth-driven Visual Attention on the iCub robot: Instructions for Use and New Perspectives (online:** arxiv)*

We start by focusing on instance recognition

Pour from the **green box** into the **brown cup**, please.



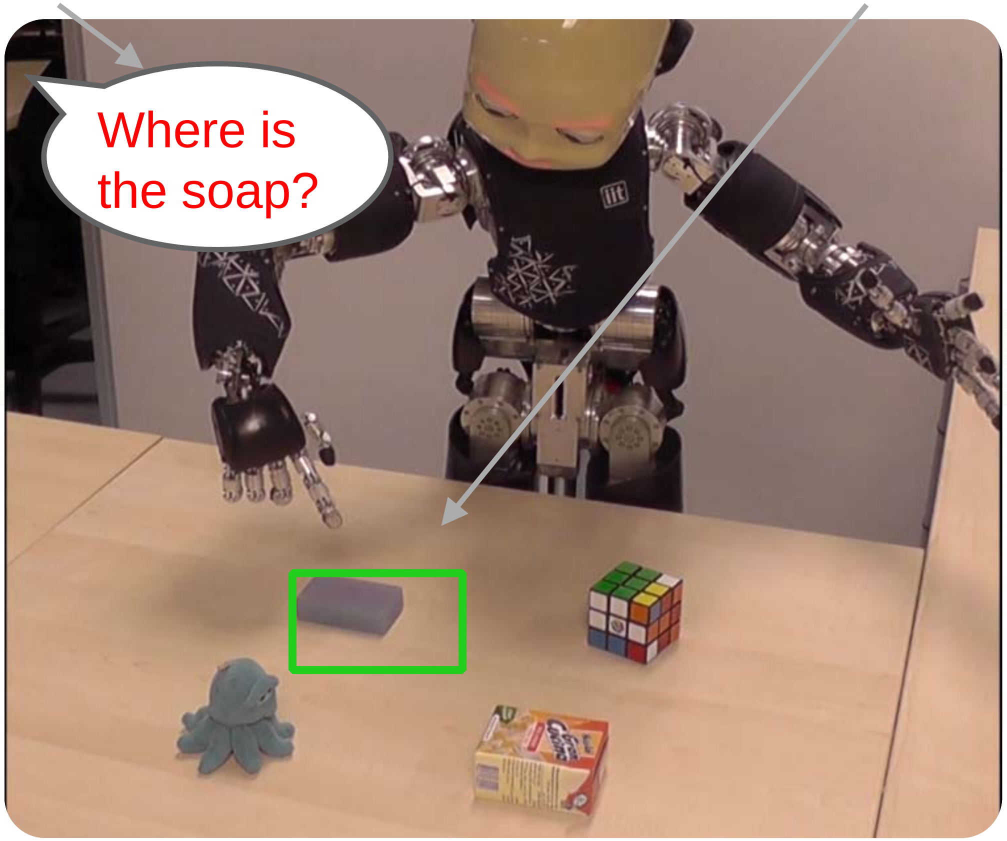
Instance Recognition

ISTITUTO ITALIANO DI TECNOLOGIA

Interactive Object Learning

Verbal instructions of a “teacher”

Robot’s attention (motion, color-based segmentation, disparity)



Benchmarking the iCub visual system

- Growing dataset collecting images from a real robotic setting
- Provide the community with a tool for benchmarking visual recognition systems in robotics
- 28 Objects, 7 categories, 4 sessions of acquisition (four different days)
- 11Hz acquisition frequency
- ~50K Images

http://www.iit.it/en/projects/data-sets.html

ISTITUTO ITALIANO DI TECNOLOGIA

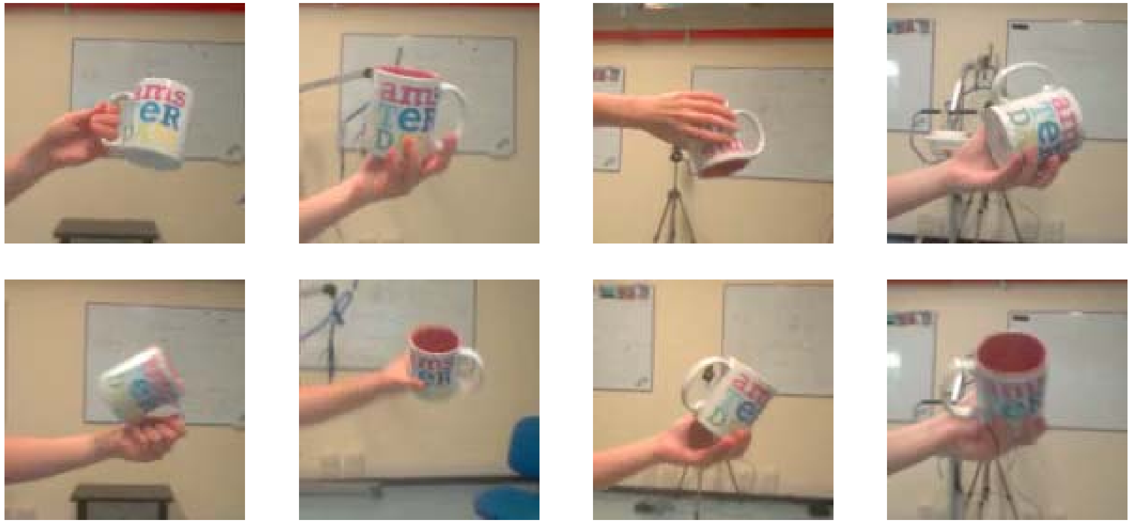
iCubWorld28 Dataset
Examples of Acquired Videos

2014: “Household”

day1 day2 day3 day4

TRAIN

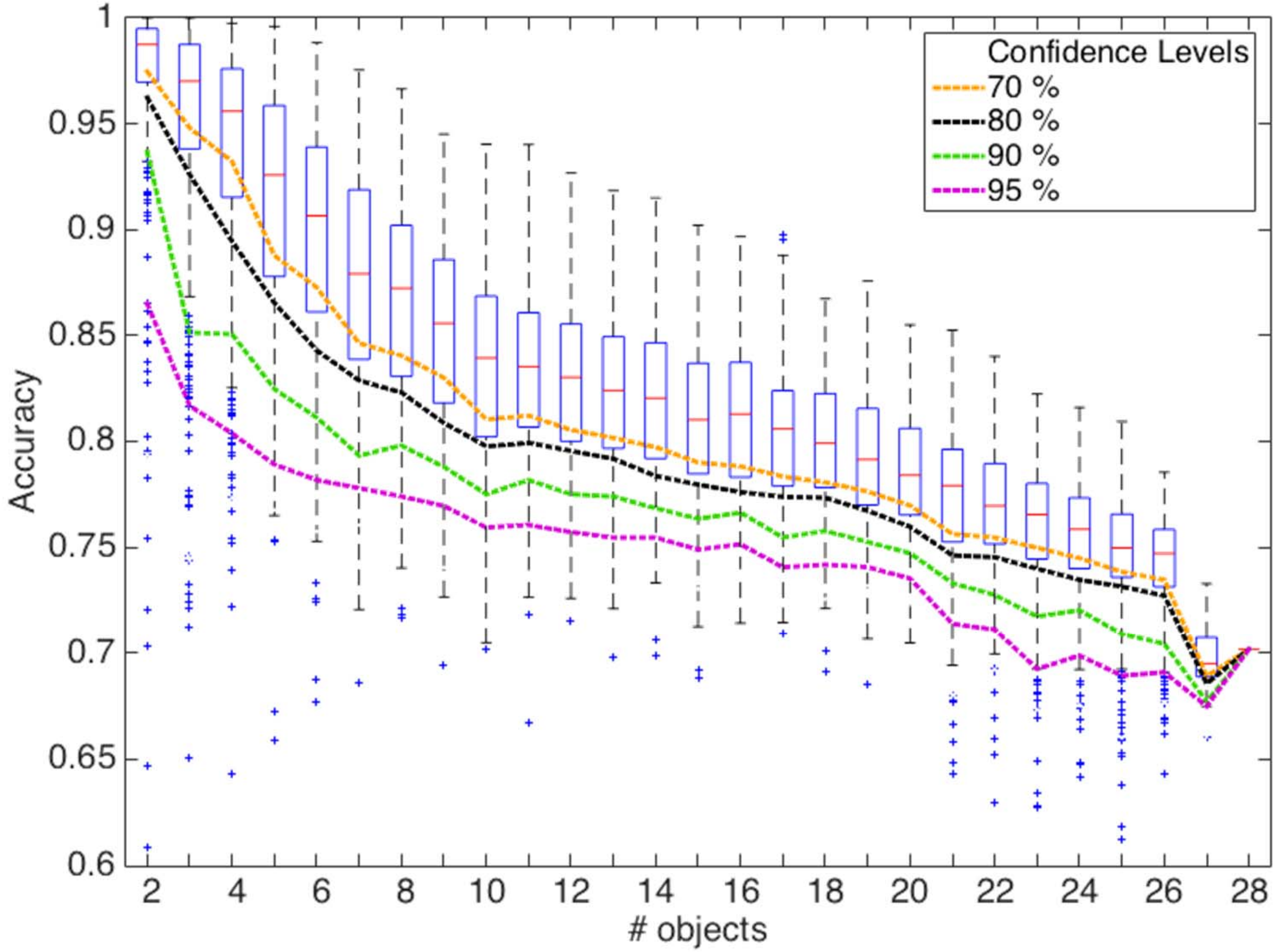
TEST



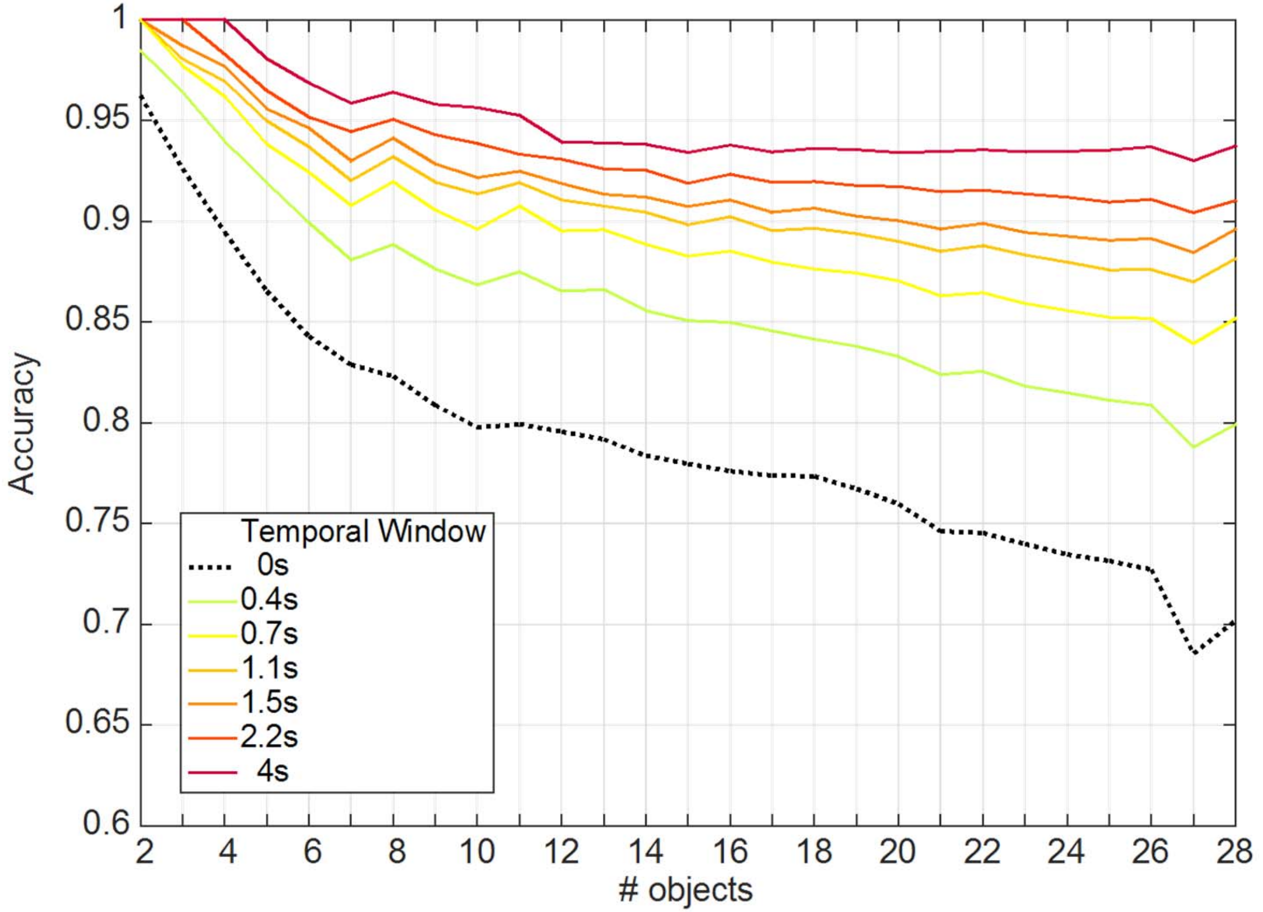
Benchmarking deep Conv Nets for Real-world Object Recognition: How many Objects can iCub Learn?

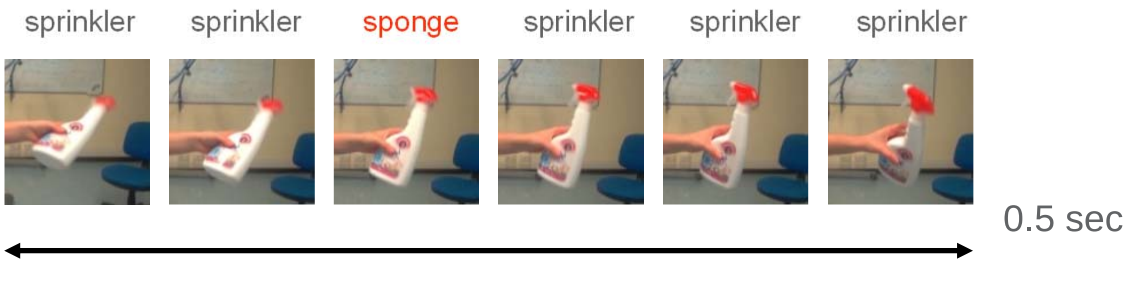
arXiv: 1504.03154, ***http://www.iit.it/it/projects/data-sets.html***

Recognition datasheet



Exploiting time continuity



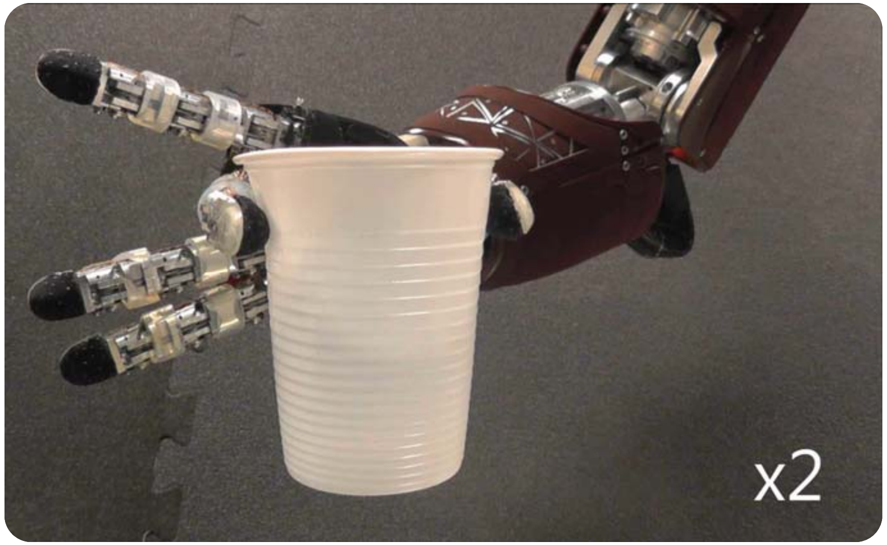


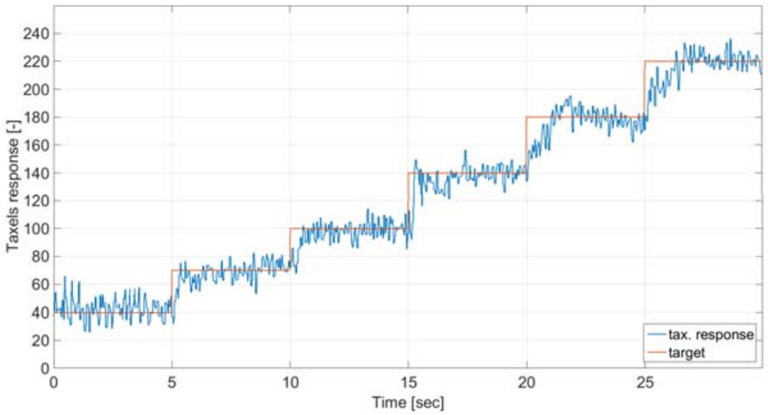
More objects, more variability

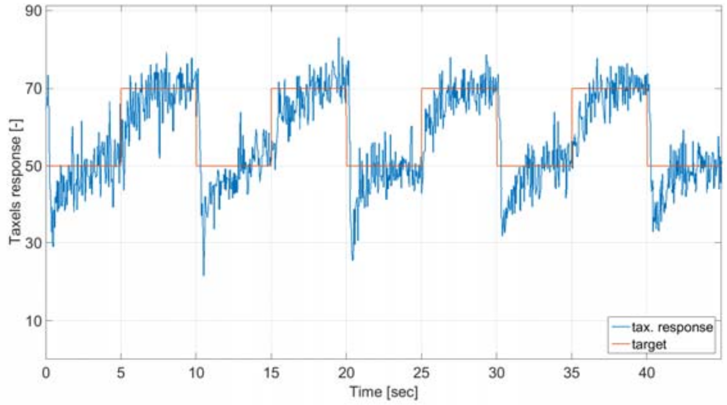
Toys	Kitchen & Food	Bathroom	Office	Daily Use	Work Tools
TEDDY BEAR	SODA BOTTLE	PERFUME	MOUSE	WALLET	SHOVEL
BARBIE DOLL	MUG	BODY LOTION	RING BINDER	CELL PHONE	RAKE
CAR	GLASS	SOAP DISPENSER	PENCIL CASE	SUNGLASSES	PAINT BRUSH
…	OVEN GLOVE	SPRAYER	REMOTE	HAIR CLIP	SCISSORS
	SQUEEZER	HAIRBRUSH	DAISY FLOWER	BOOK	ADHESIVE TAPE
	FOOD CAN	SOAP	CALCULATOR	…	…
	CANDY PACK		STAPLER		

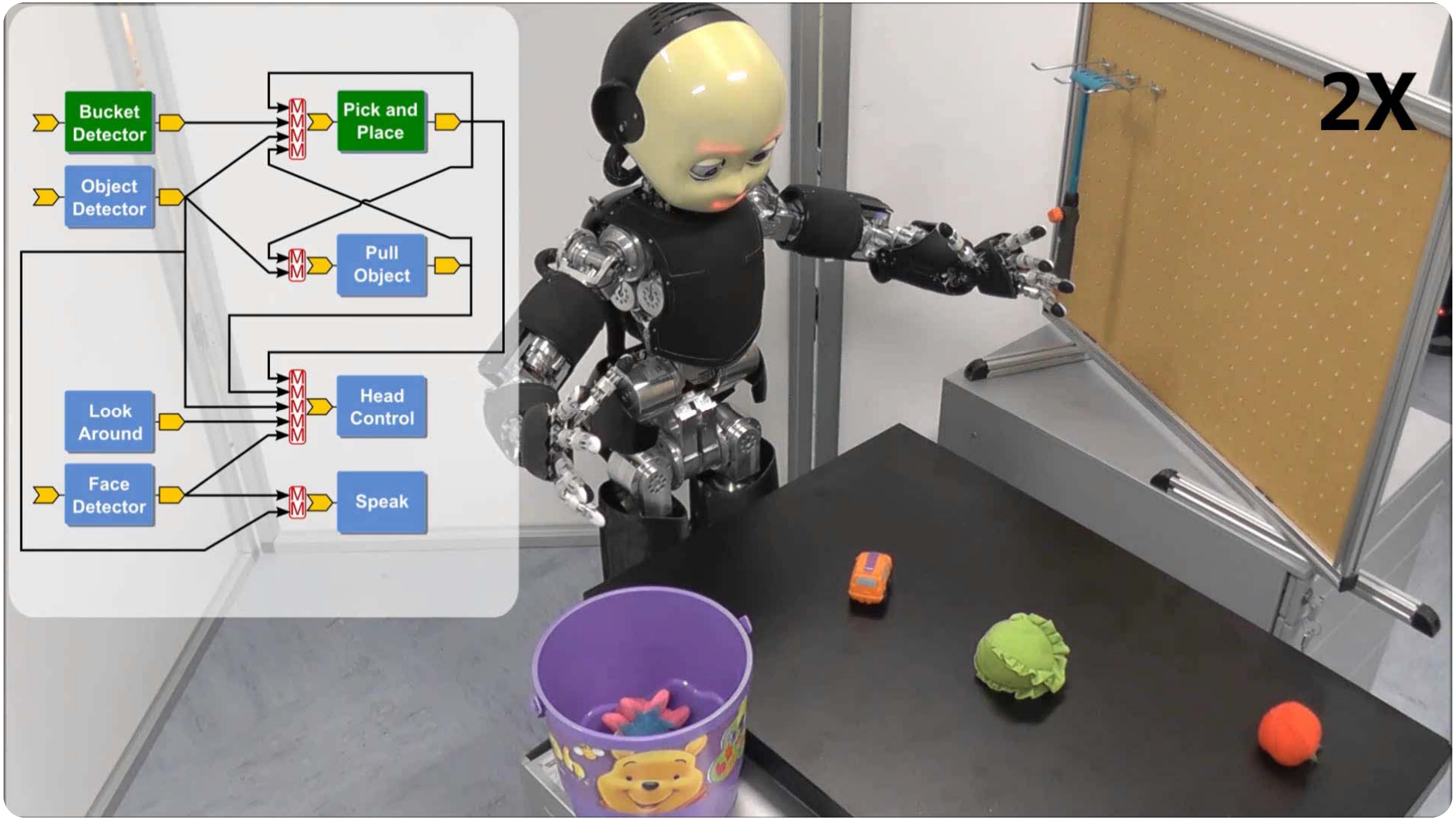
- 20 categories x 10 samples: 200 objects
- 5 different days, 600K images
- 12 hours of acquisition
- Soon to be released: http://www.iit.it/en/projects/data-sets.html
- Continuously expanding dataset, will involve other labs:
  public code for data acquisition & automatic processing

Future directions, touch









Paikan et al. IROS 2014, JOSER 2015

Wrap up

- **Tool use**:
  - A framework for self-supervised learning of pulling affordances, linking effect of actions with visual appearance of the tool
  - Improve actions and generalize to different actions
  - 3D features (see Tanis Mar presentation here at Humanoids 2015)
- **Object learning**:
  - Hierarchical methods with pre-learned representation
  - Methodology for acquiring large data set, iCubWorld
  - State-of-the-art much better, but still need improvement
    - Time/spatial continuity?
    - Incremental learning?

Acknowledgements

Giulia Pasquale
Tanis Mar
Ali Paikan
Massimo Regoli
Nawid Jamali
Carlo Ciliberto

Vadim Tikhanoff
Ugo Pattacini
Lorenzo Rosasco
Giorgio Metta