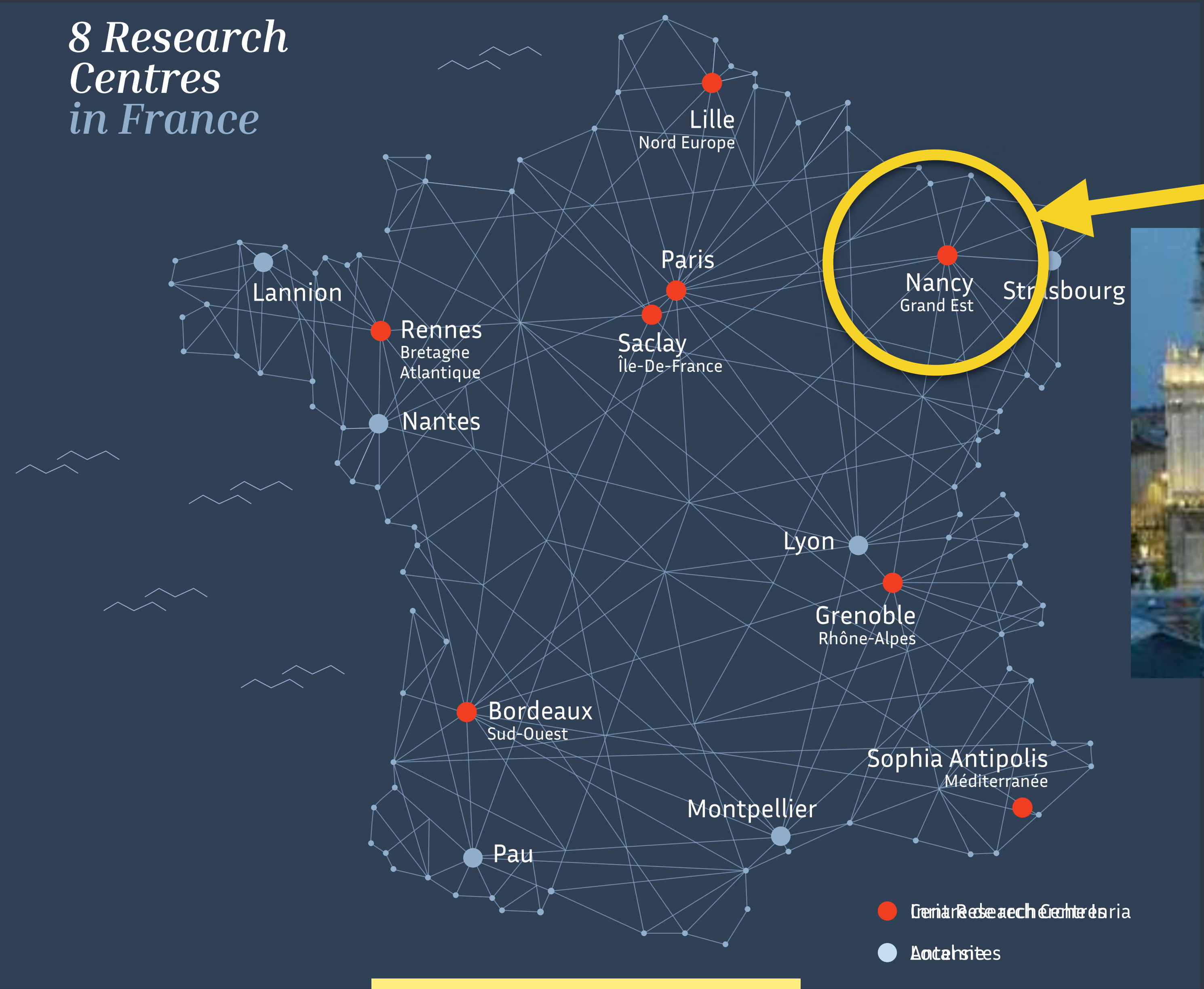
Improving object grasping and manipulation by exploiting uncertainty and human interaction

Serena Ivaldi

Team LARSEN, INRIA
serena.ivaldi@inria.fr

Workshop “Task informed grasping” @ RSS - 22/06/2019

Inria Nancy / Loria
Team Larsen



European Commission

Horizon 2020
European Union funding
for Research & Innovation

erc
European Research Council

Outline

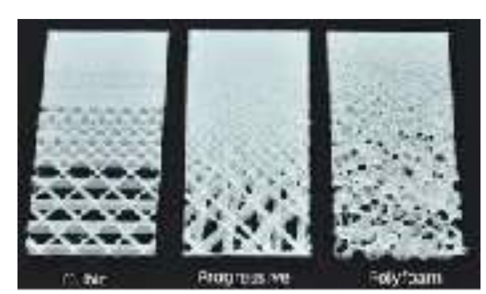
Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Multimodal learning of the visual appearance of objects (w/ Kinect)

Demonstrating assembly with kinesthetic teaching

Demonstrating whole-body grasping with tele-operation

HEAP project



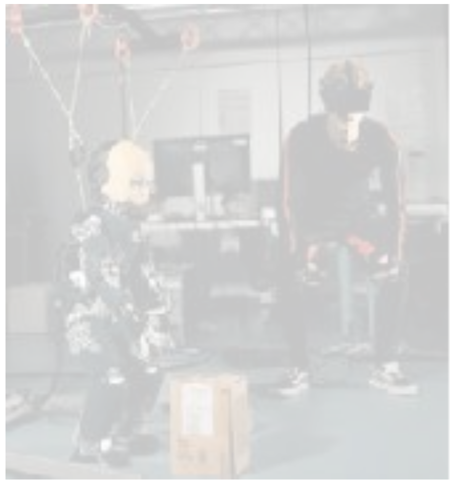
HEAP object dataset

Outline

Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

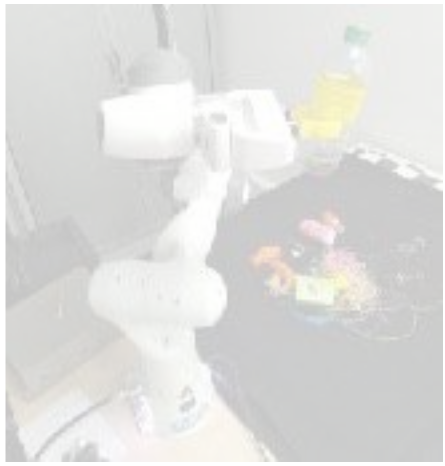
Multimodal learning of the visual appearance of objects (w/ Kinect)

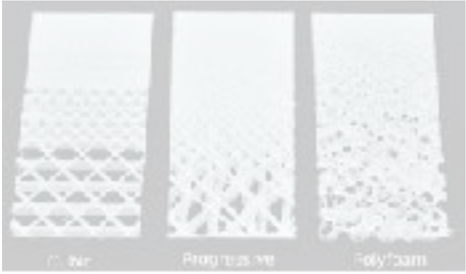
Demonstrating assembly with kinesthetic teaching



Demonstrating whole-body grasping with tele-operation

HEAP project

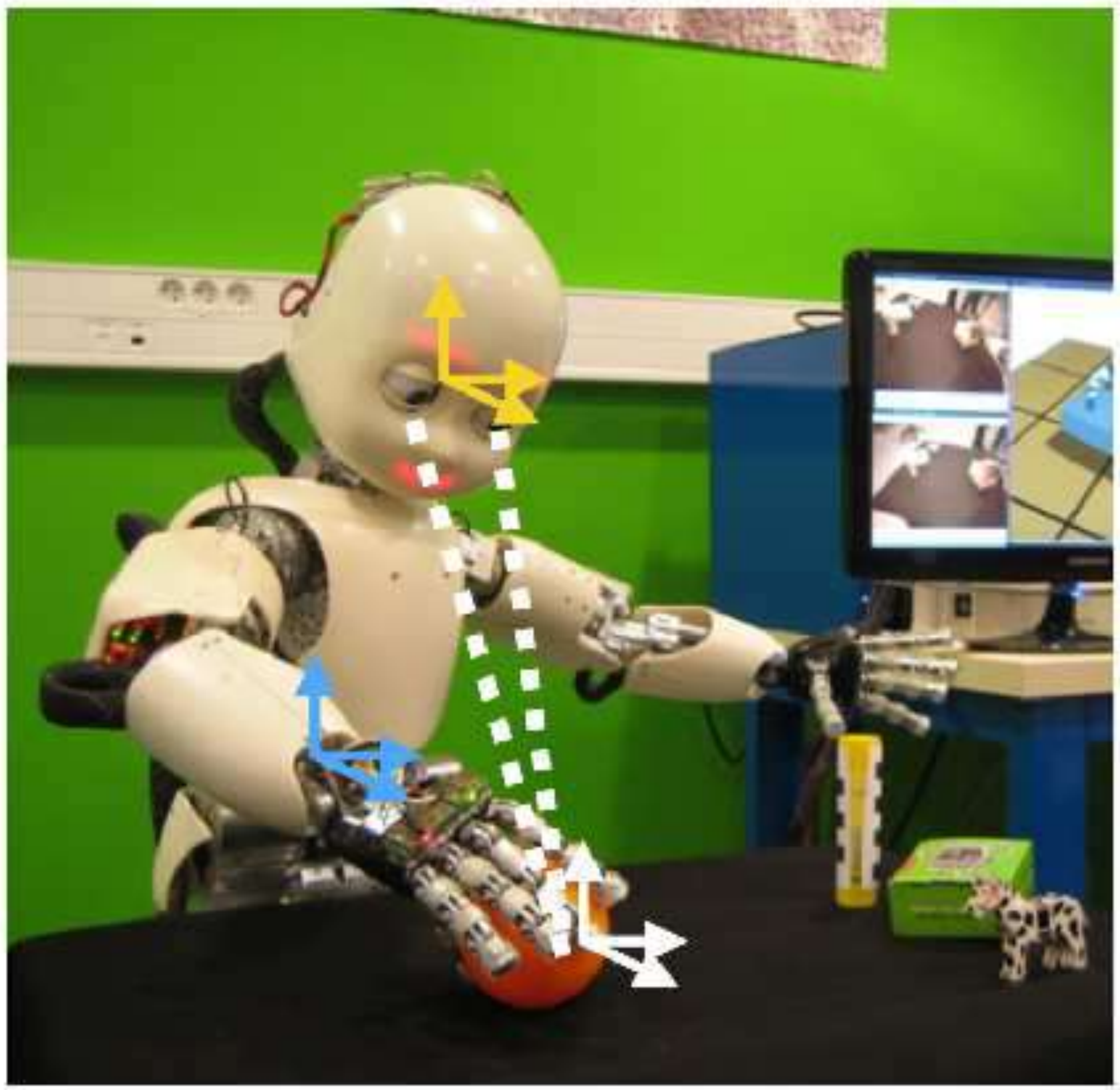


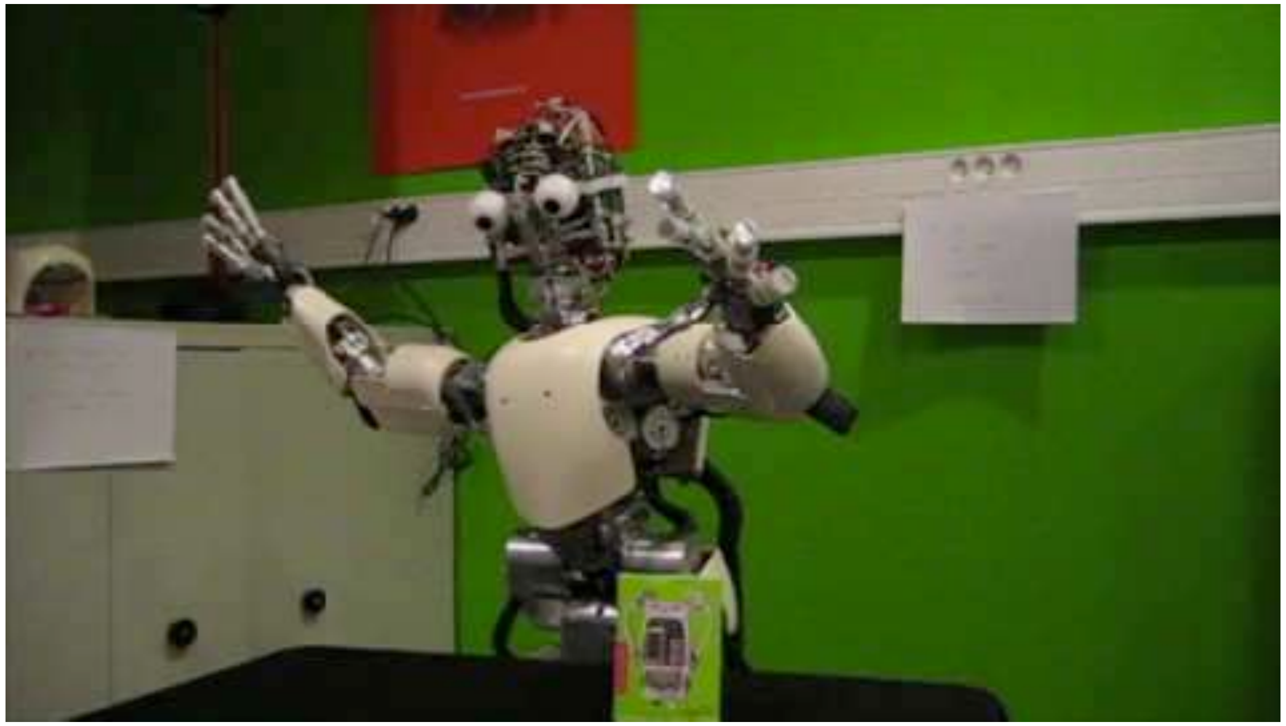


HEAP object dataset

Can we do the same with the eyes' cameras?

- Ideal grasping of perfectly localised objects using the eyes' cameras: eye-hand calibration + object pose (vision) + object size/shape (vision) + correct grasp = success!



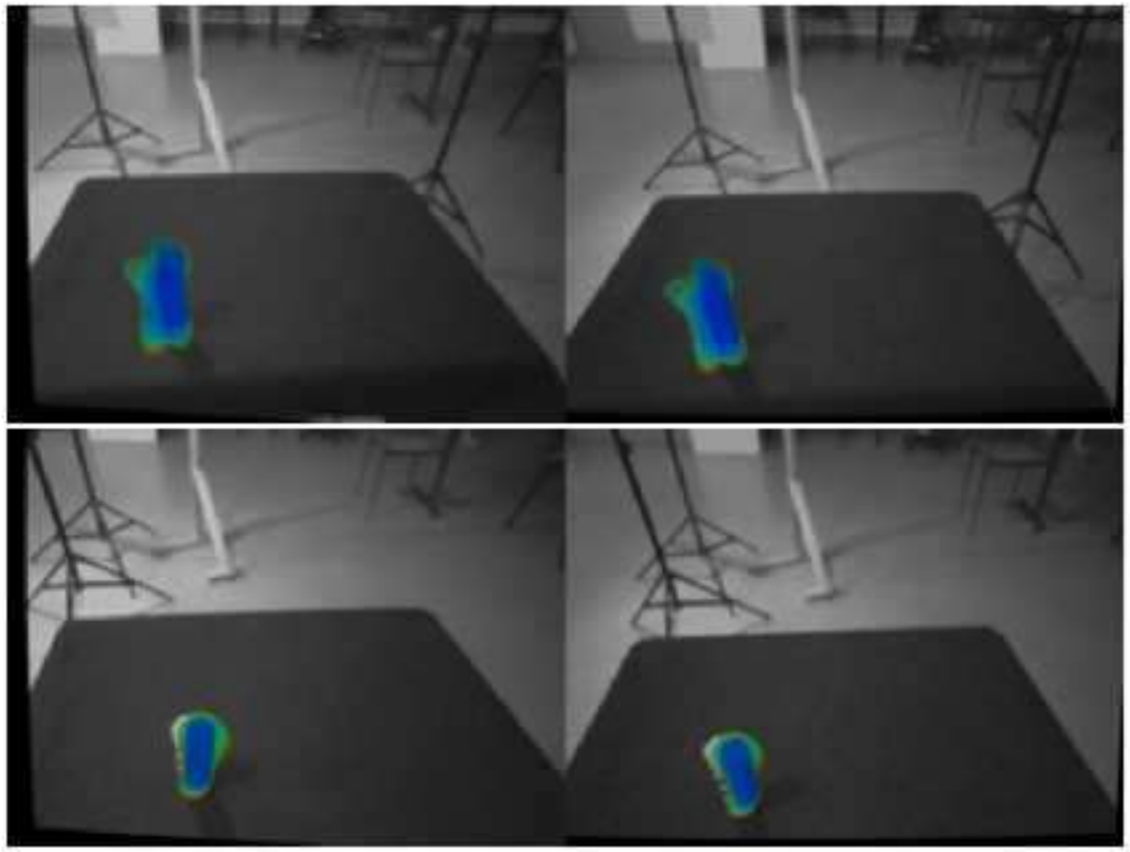


success!

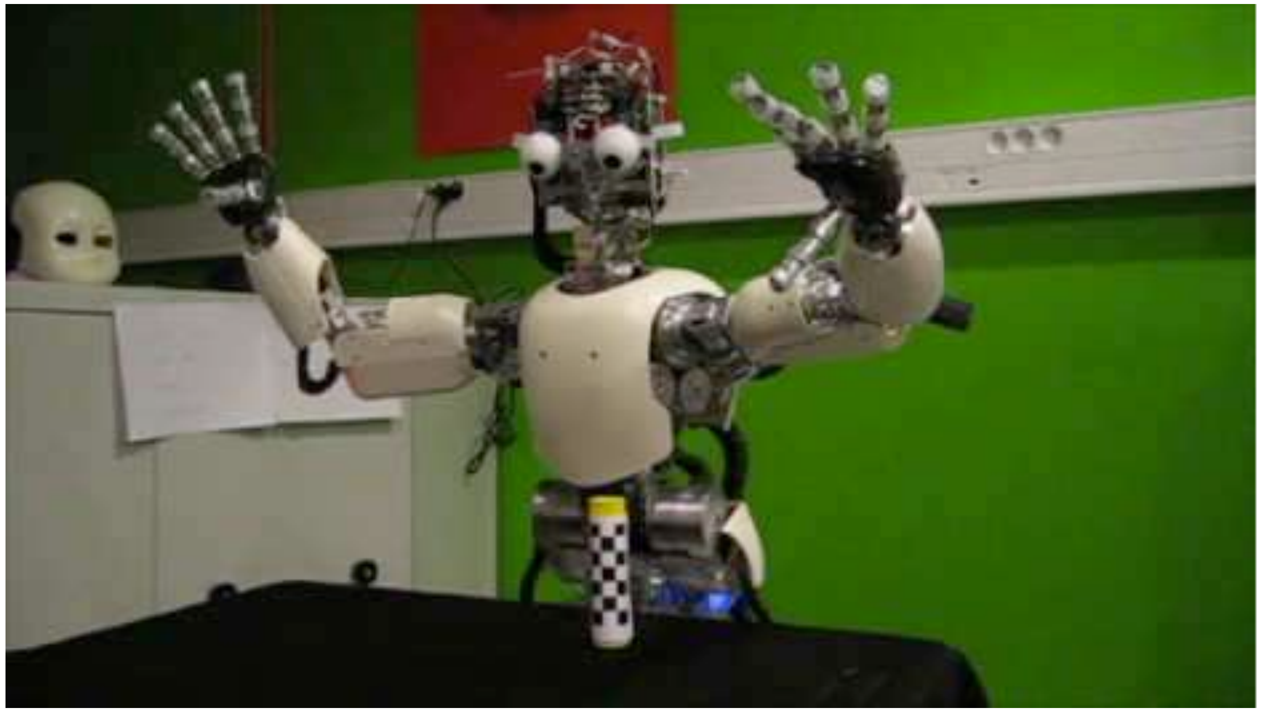
Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Unfortunately, the cameras bring limitations…

- Error in object pose estimation is inevitable, particularly when the object pose is estimated through low-resolution cameras
- Grasping is very sensitive to the accuracy of the object pose estimation
  ➡ failure!



inaccurate object pose estimation



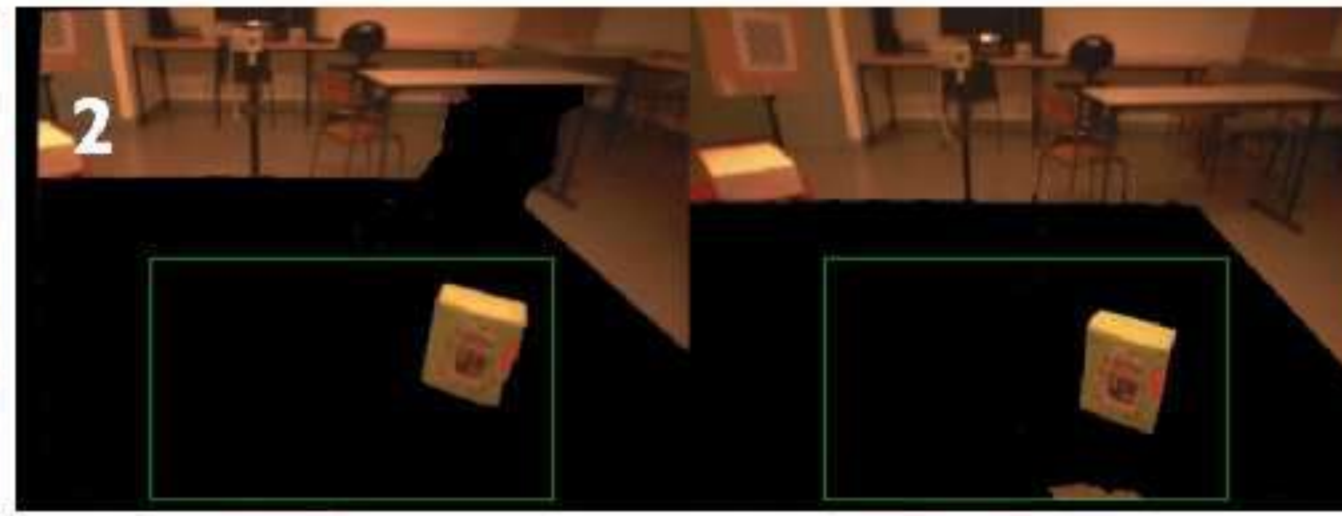
failure!

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Unfortunately, the cameras bring limitations…

- Extracting the point cloud from the stereo cameras of iCub

camera images after distortion correction



extracted objects (w/ Grabcut algorithm)

2D features (SURF)

matched features

Grasping objects localised from noisy point clouds

- **Problem:** the point cloud from the stereo cameras has too few points to run classical algorithms, such as the Iterative Closest Point (ICP, in PCL)

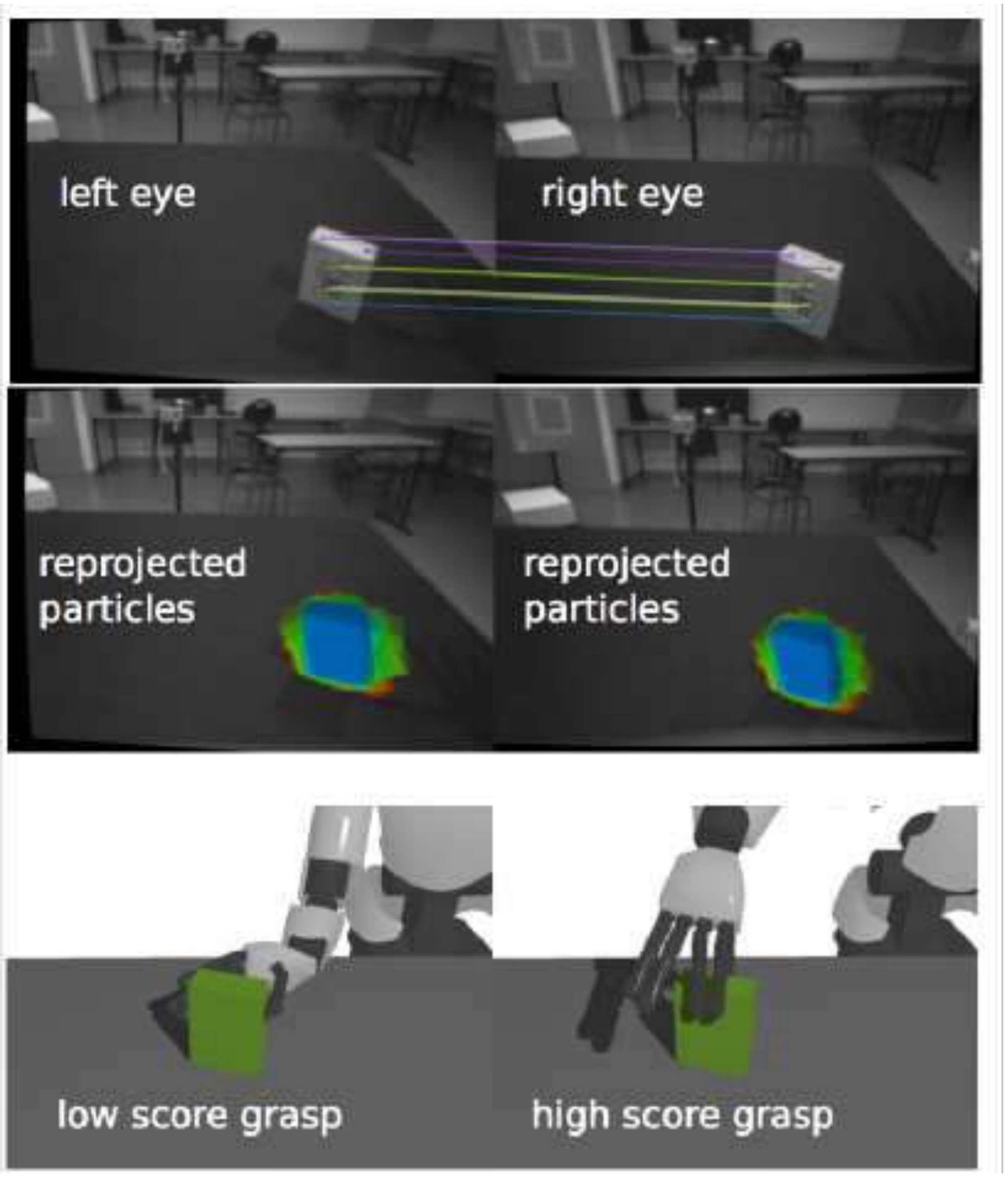
20 points | 8 points | 9 points

- Small errors in the estimated pose may cause the planned grasp to fail
- Difficult to validate a grasp when tactile or force sensing is missing

➡ **find grasps that are less sensitive to the pose uncertainty**

Grasping objects localised from noisy point clouds

➡ **Proposed method:**
grasp planning method that explicitly considers the pose uncertainty to compute the best grasp configuration

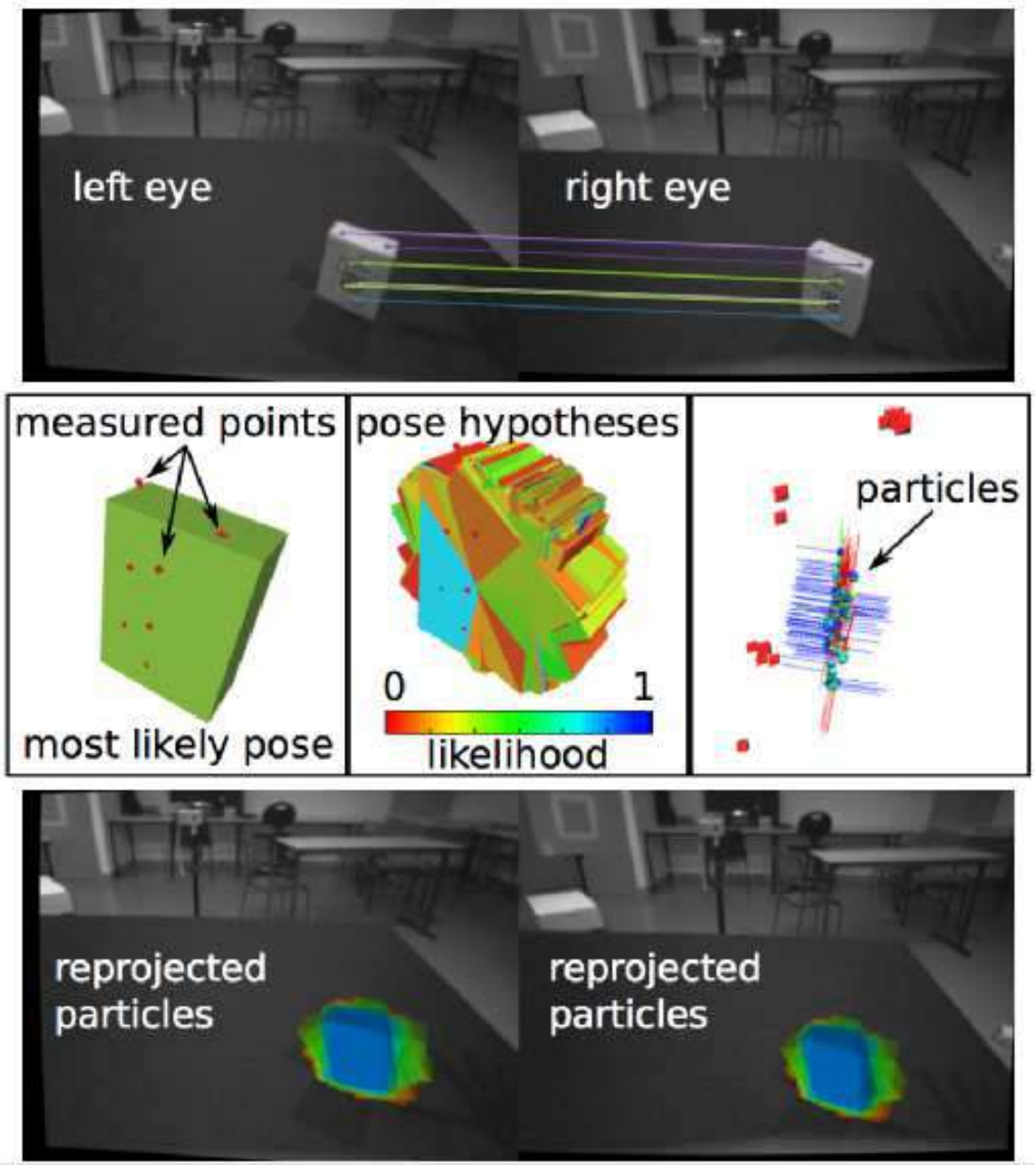


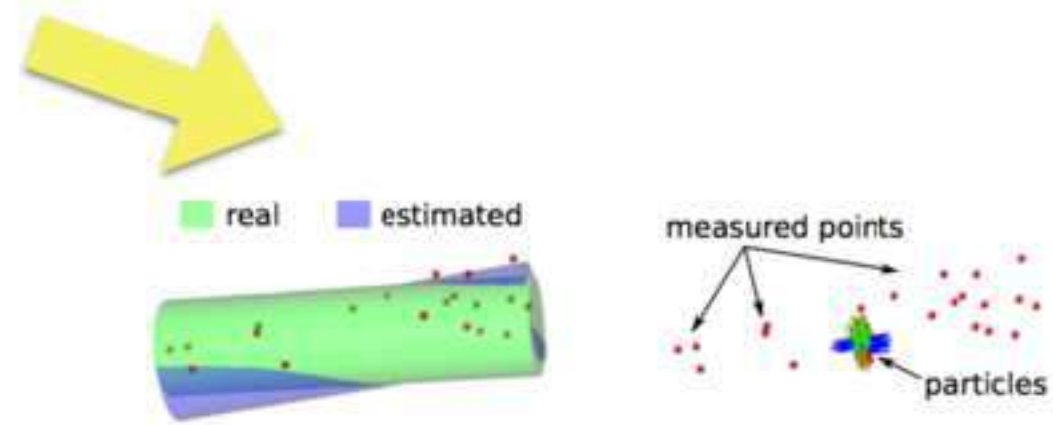
- Inputs: point cloud, object model (primitive or 3D mesh)
- Step 1: Estimate the probability distribution of the object pose with a set of particles/hypotheses
- Step 2: Build a set of stable grasps & compute scores

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. RAS

Grasping objects localised from noisy point clouds

- Step 1: Estimate the probability distribution of the object pose and a set of particles/hypotheses





- a particle X is an hypothesis on the object pose
- we have m measured points (from the point cloud)
- we can compute the probability of a candidate object pose X given m measured points d

$$p(X|d_1, \ldots, d_m)$$

Grasping objects localised from noisy point clouds

- Step 2: Build a set of stable grasps & compute scores

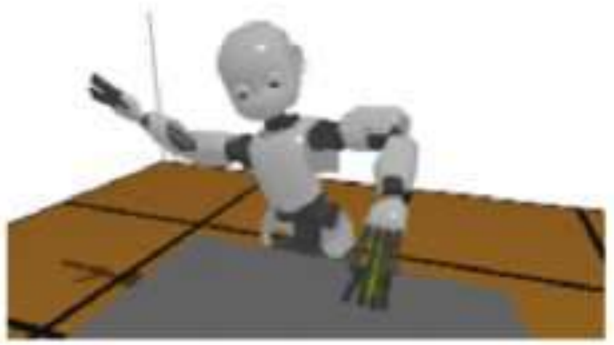
n=144 evaluations of grasps

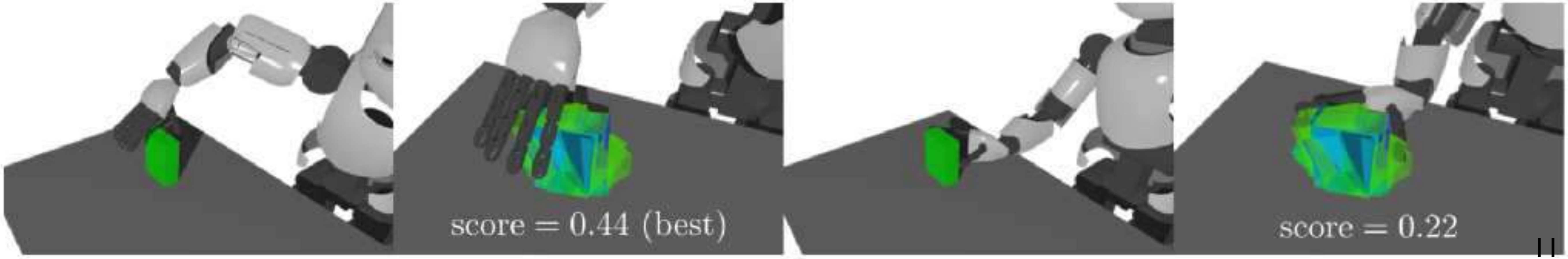
compute probability of success S of the candidate grasp T:

$$p(S|T_{grasp}) = \frac{1}{n}\sum_{i=1}^{n} p(X_i|d_1,\ldots,d_m)\,p(S|T_{grasp},X_i)$$

probability of the object pose X given the observations d, computed at step 1

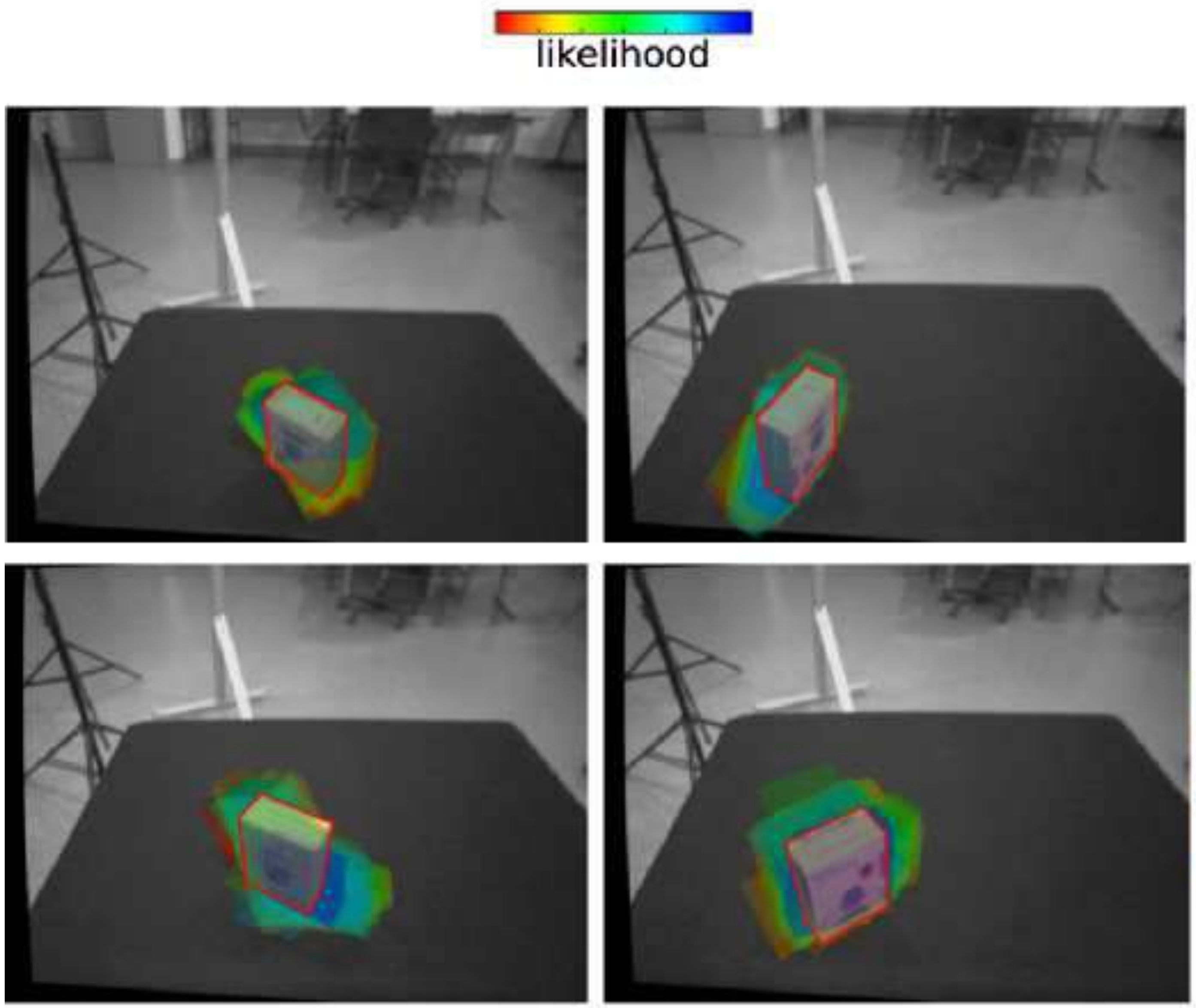
= {0=invalid, 1=valid}, evaluated in simulation



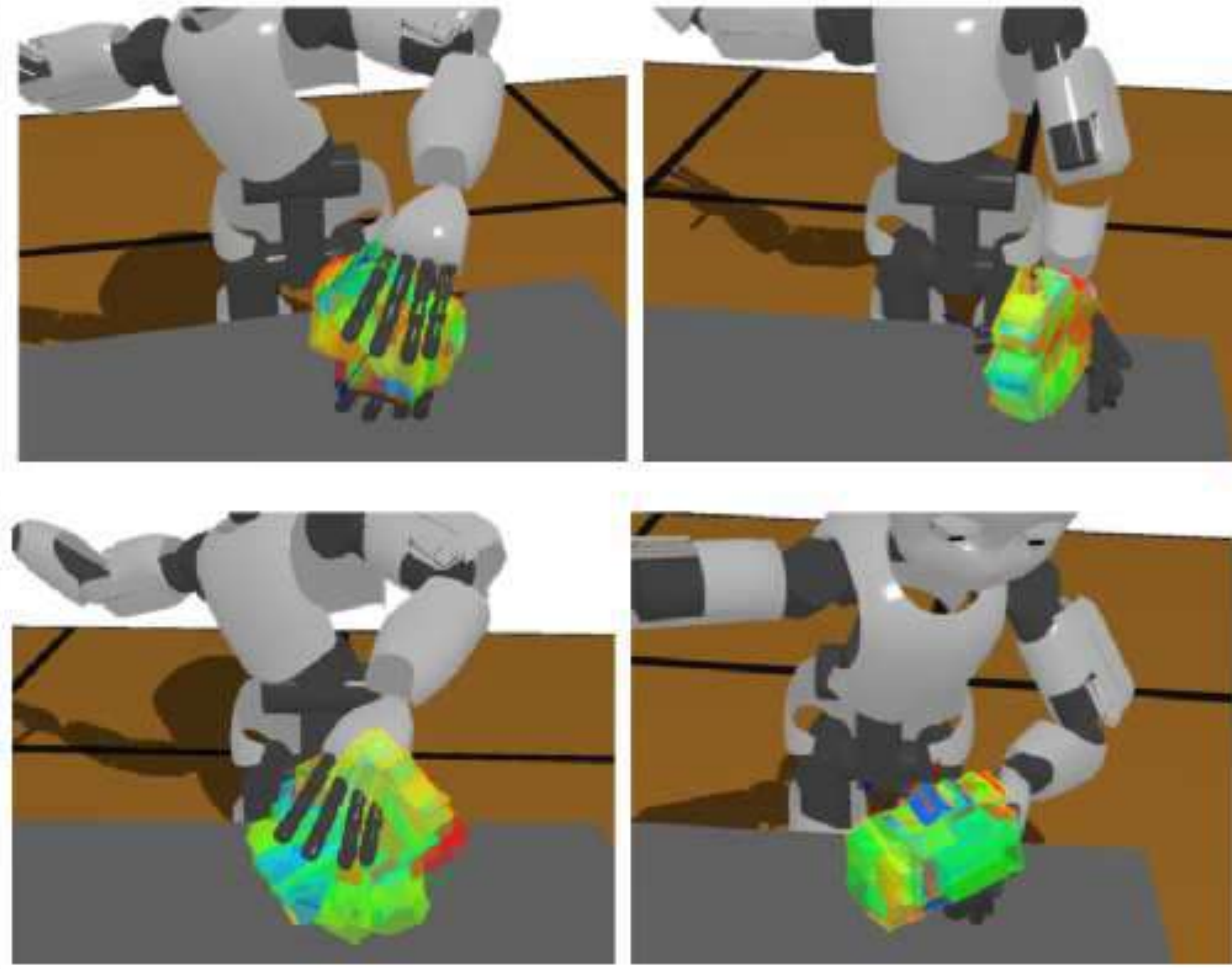


Grasping objects localised from noisy point clouds

- Each different pose & orientation of the object yield different grasps



Reprojection of the particle set on the left image



Grasp that was ranked first in the likelihood to succeed

Grasping objects localised from noisy point clouds

- We use the probability distribution of the object pose to help selecting the grasp that is more likely to succeed considering the possible poses

- **Pro**:
  - Successful grasp with (noisy) stereo cameras, compensating the absence of tactile sensing in the fingers

- **Cons**:
  - Need a prior object model
  - Computational time required to find the most suitable grasp (~seconds, less than ICP in any case)

- **Future work:**
  - Integrating human feedback in the grasp scoring system

Saut, Ivaldi, Sahbani, Bidaud (2014) Grasping objects localised from uncertain point cloud data. Robotics and Autonomous Systems, 62(12): 1742-1754.

Outline

Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Multimodal learning of the visual appearance of objects (w/ Kinect)

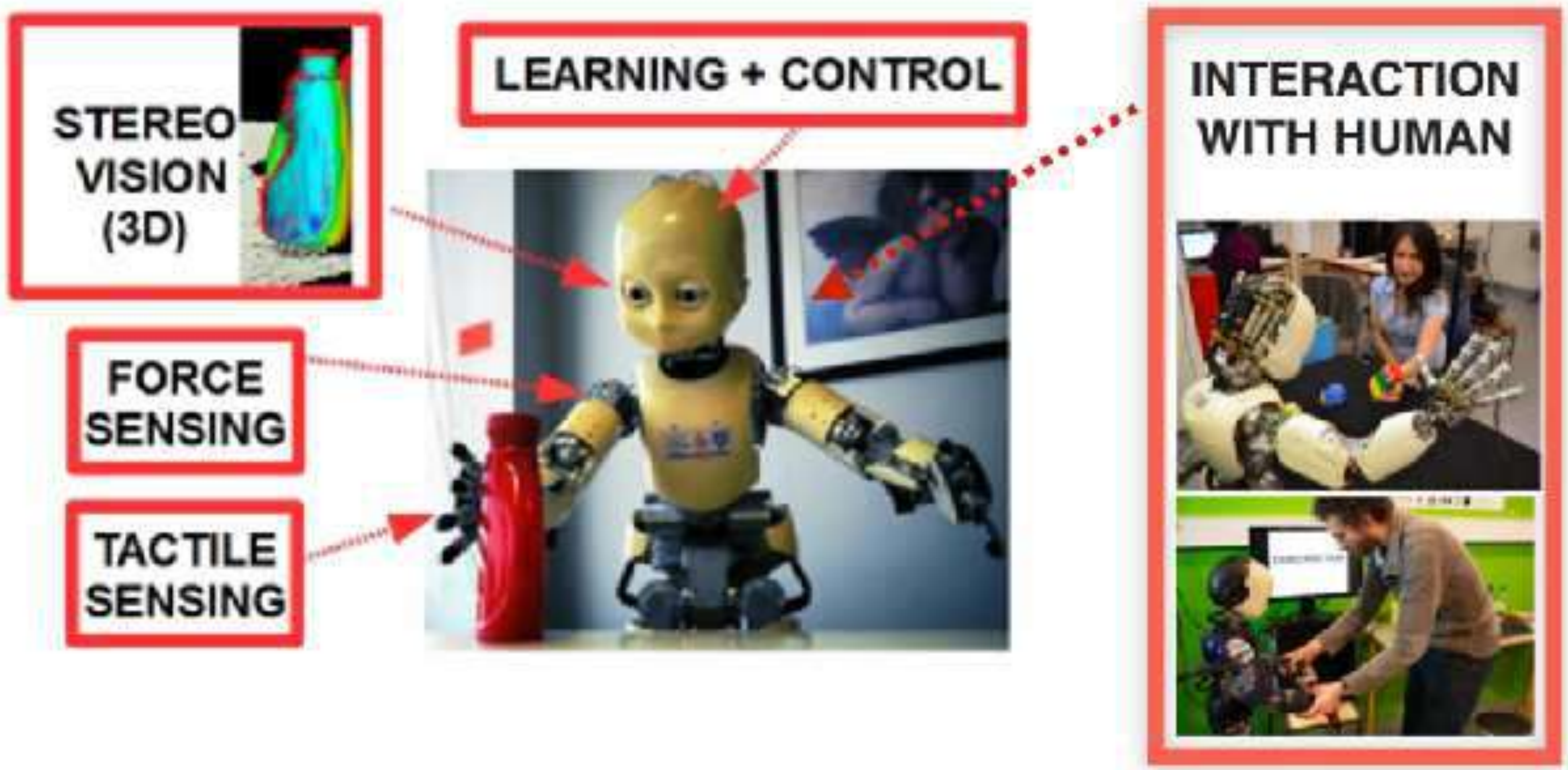
Demonstrating assembly with kinesthetic teaching

Demonstrating whole-body grasping with tele-operation

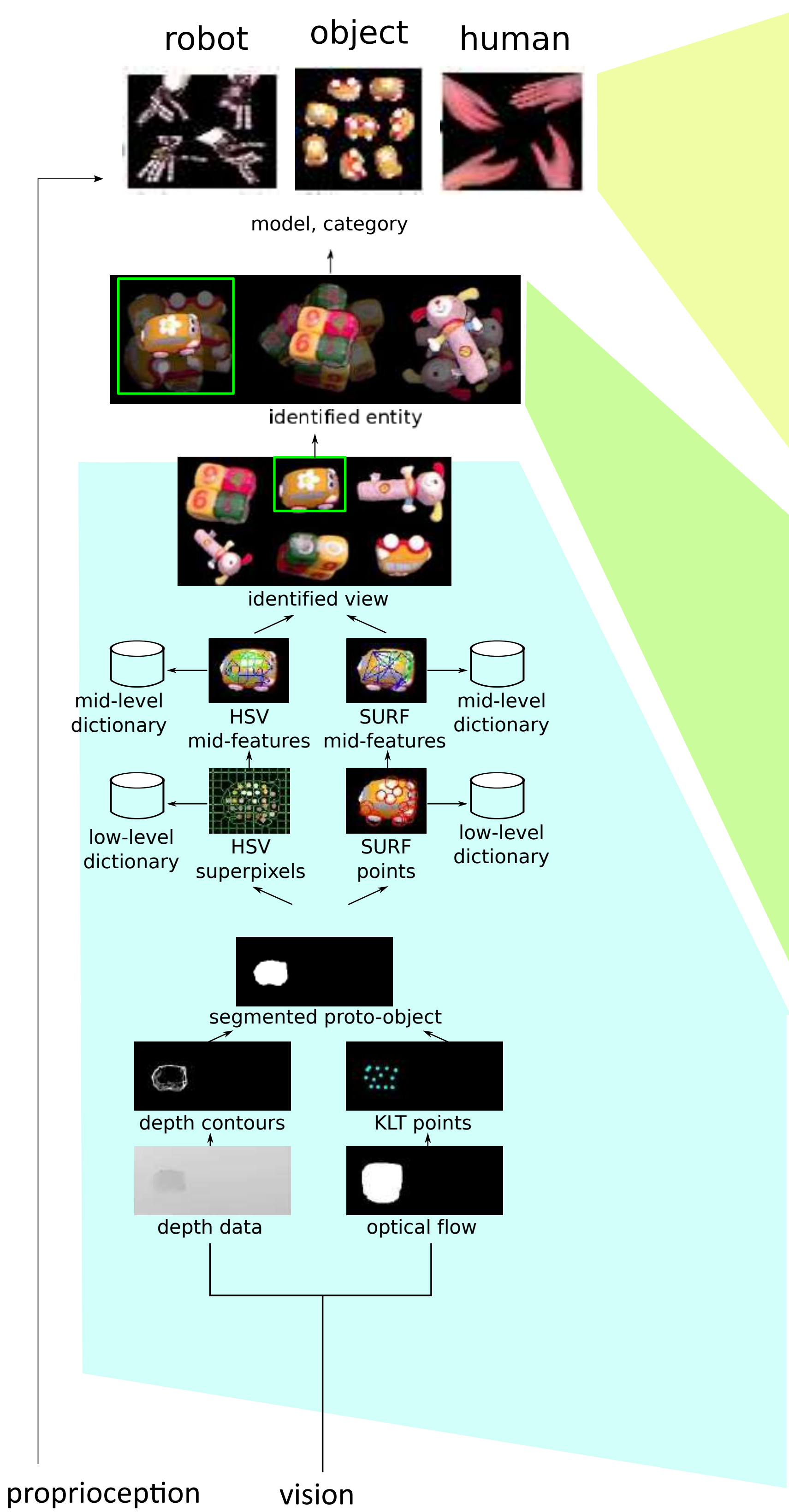
HEAP project

HEAP object dataset

Multimodal learning of objects



Multimodality for object learning



exploration and interaction
(better models with categories)

↑

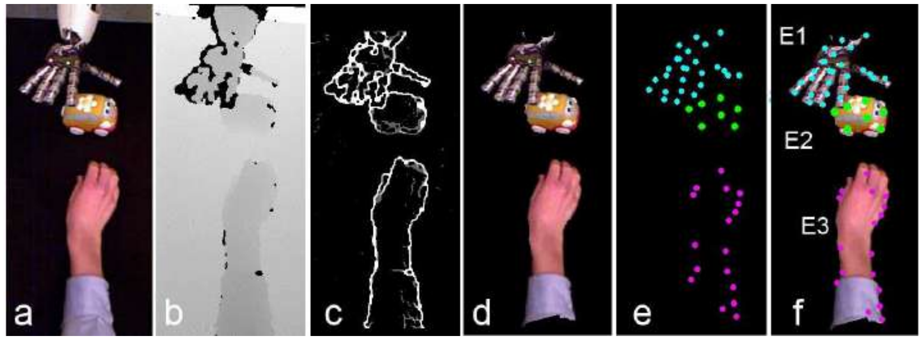
active exploration
(manipulation, better models)

↑

observation
(pure vision: models and entities)

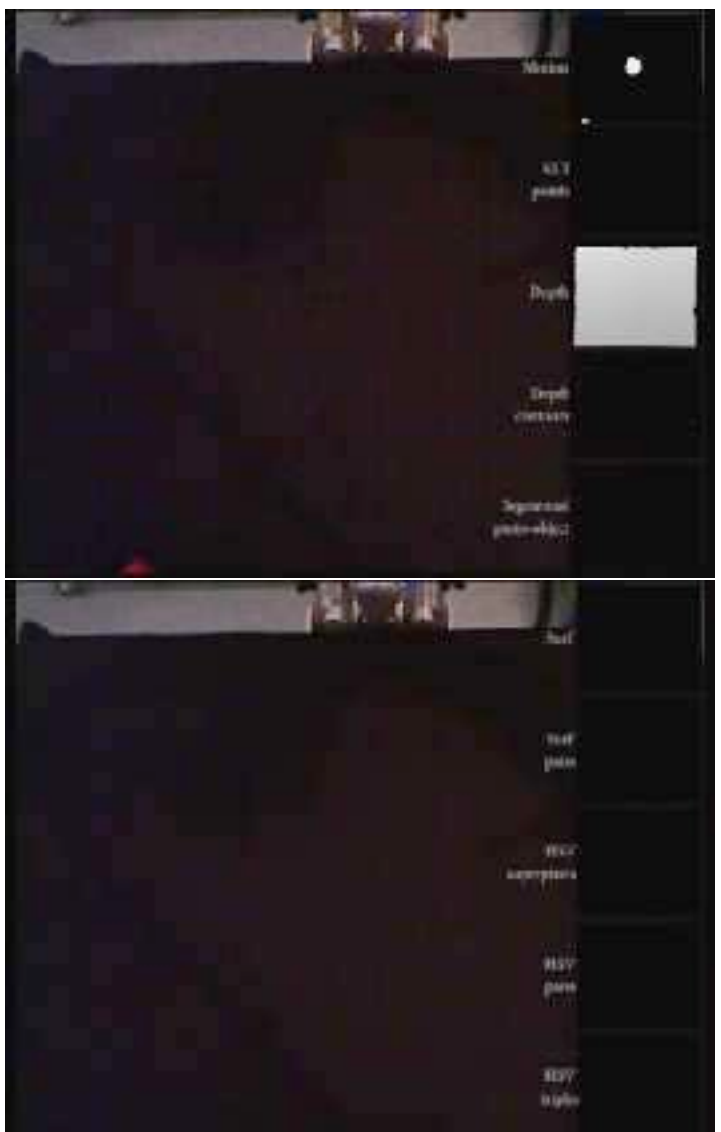
Observation alone is not enough

vision



entities + collected views

model



The robot learns the objects demonstrated by the human.

The robot has not yet learnt to identify its body, hence all entities are labeled by an "unknown" category.

Pushing objects

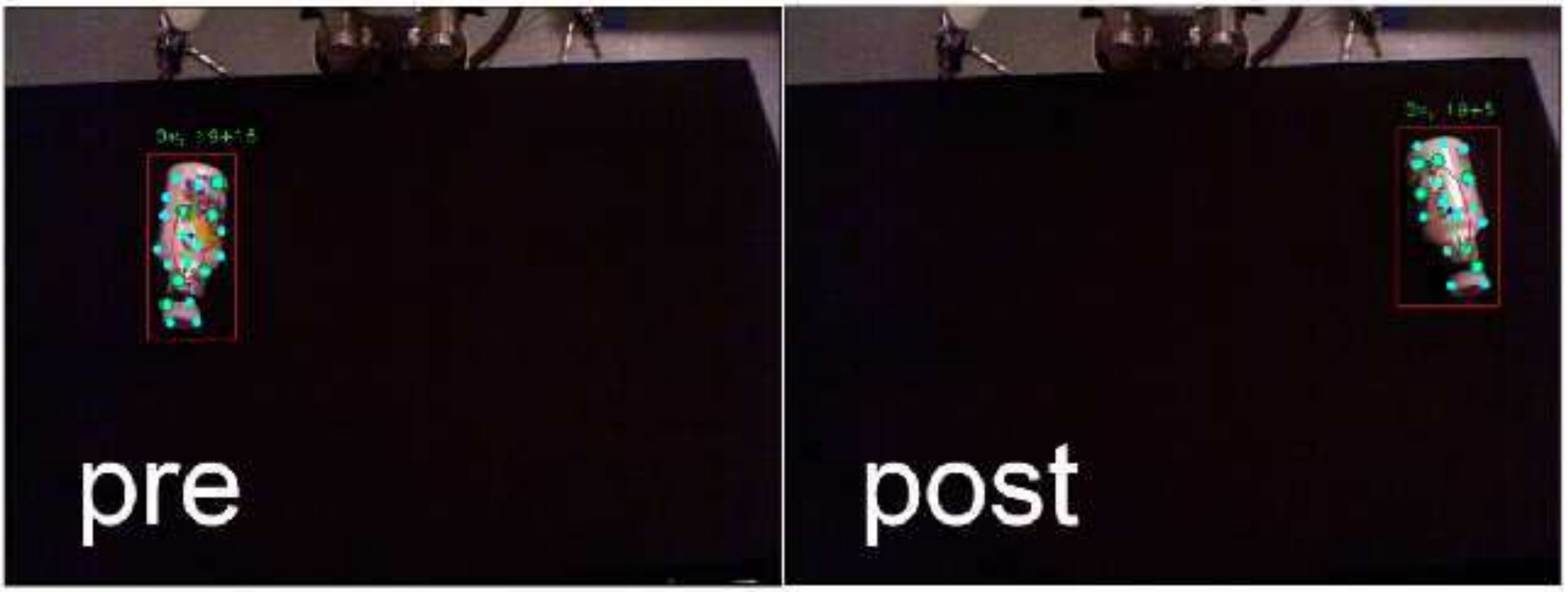
grasp
lift
throw

grasp
lift
rotate
put

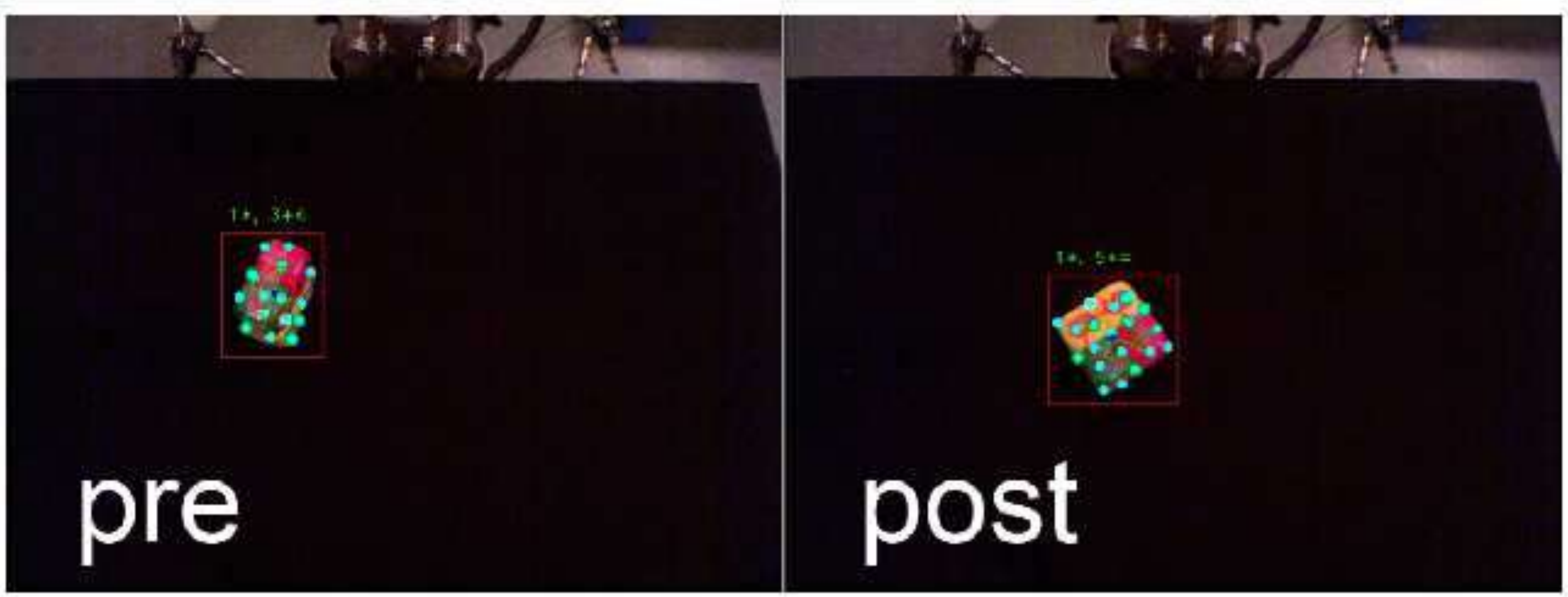
Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

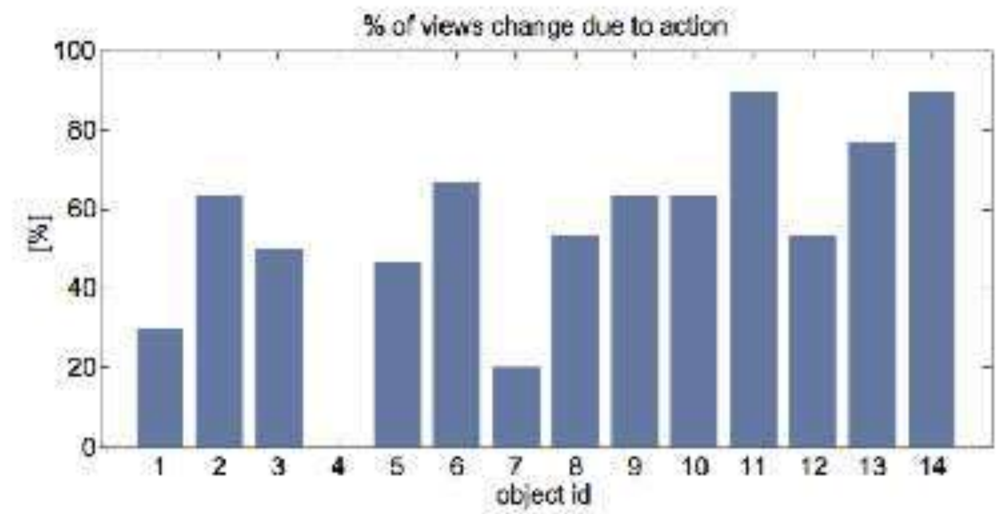
Active exploration of objects

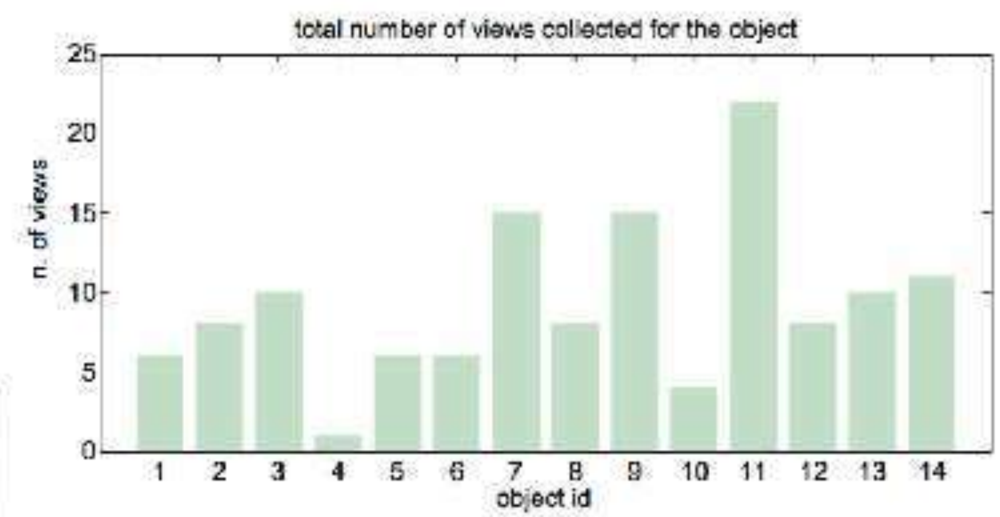
action does not change the view



action provokes a new view

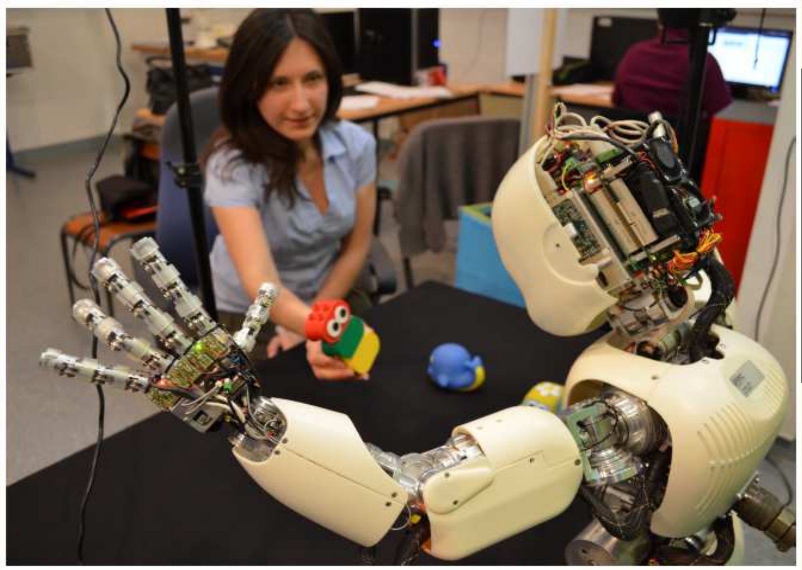






Active exploration & social guidance

social exploration

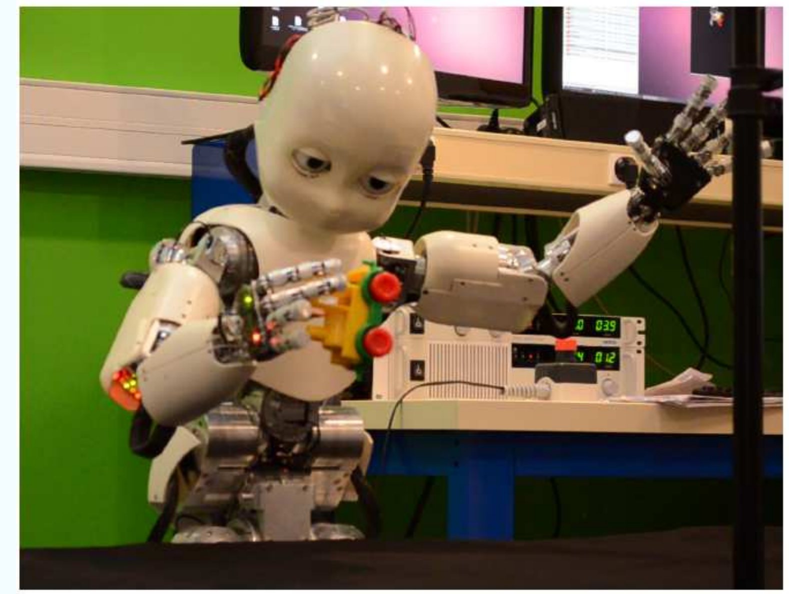


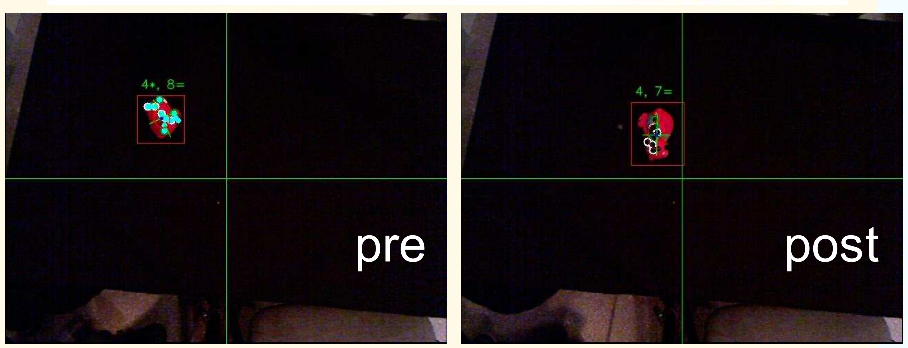
intrinsic motivation
SGIM-ACTS

exploration strategy

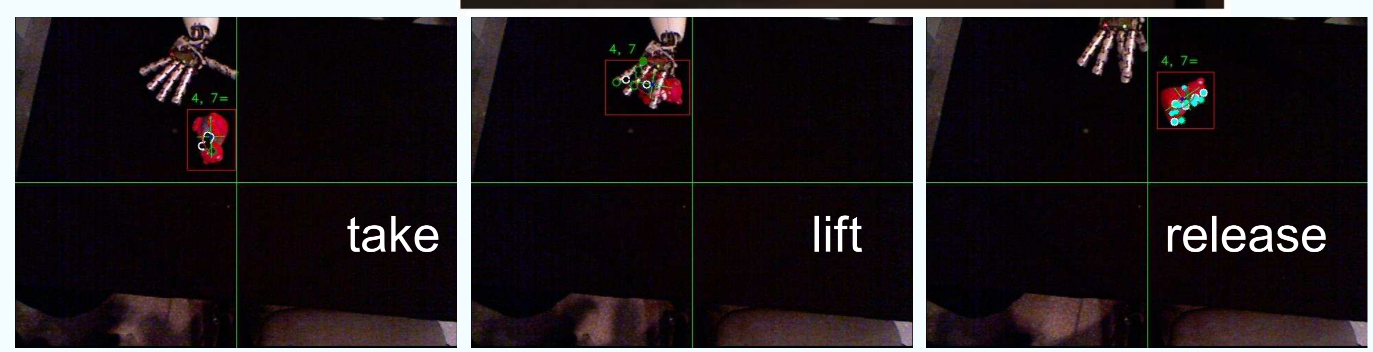
actor action object

autonomous exploration

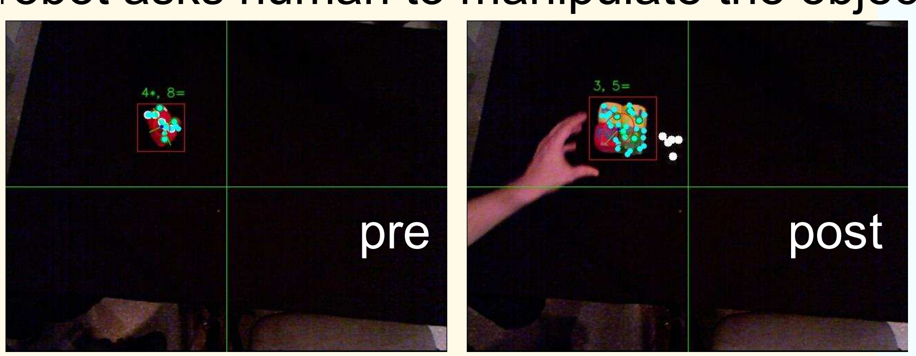




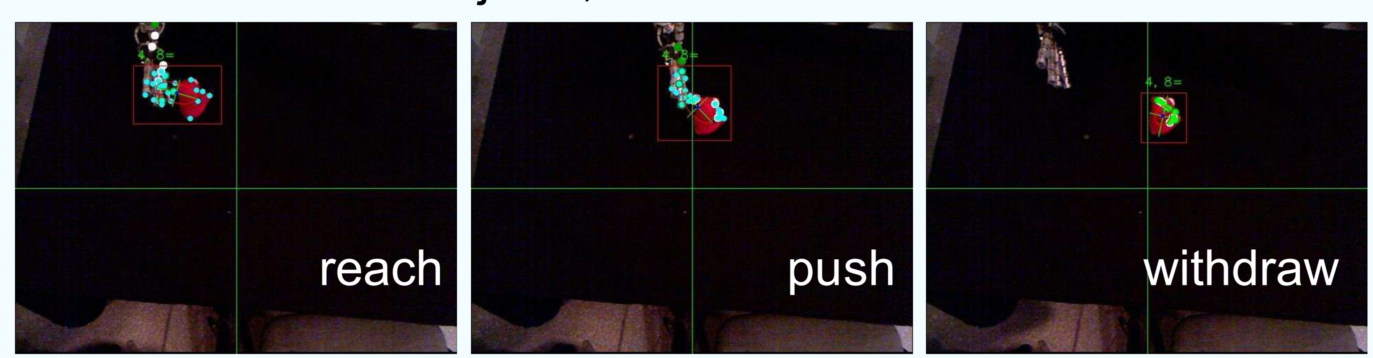
robot asks human to manipulate the object



robot lifts the objects, then makes it fall on the table



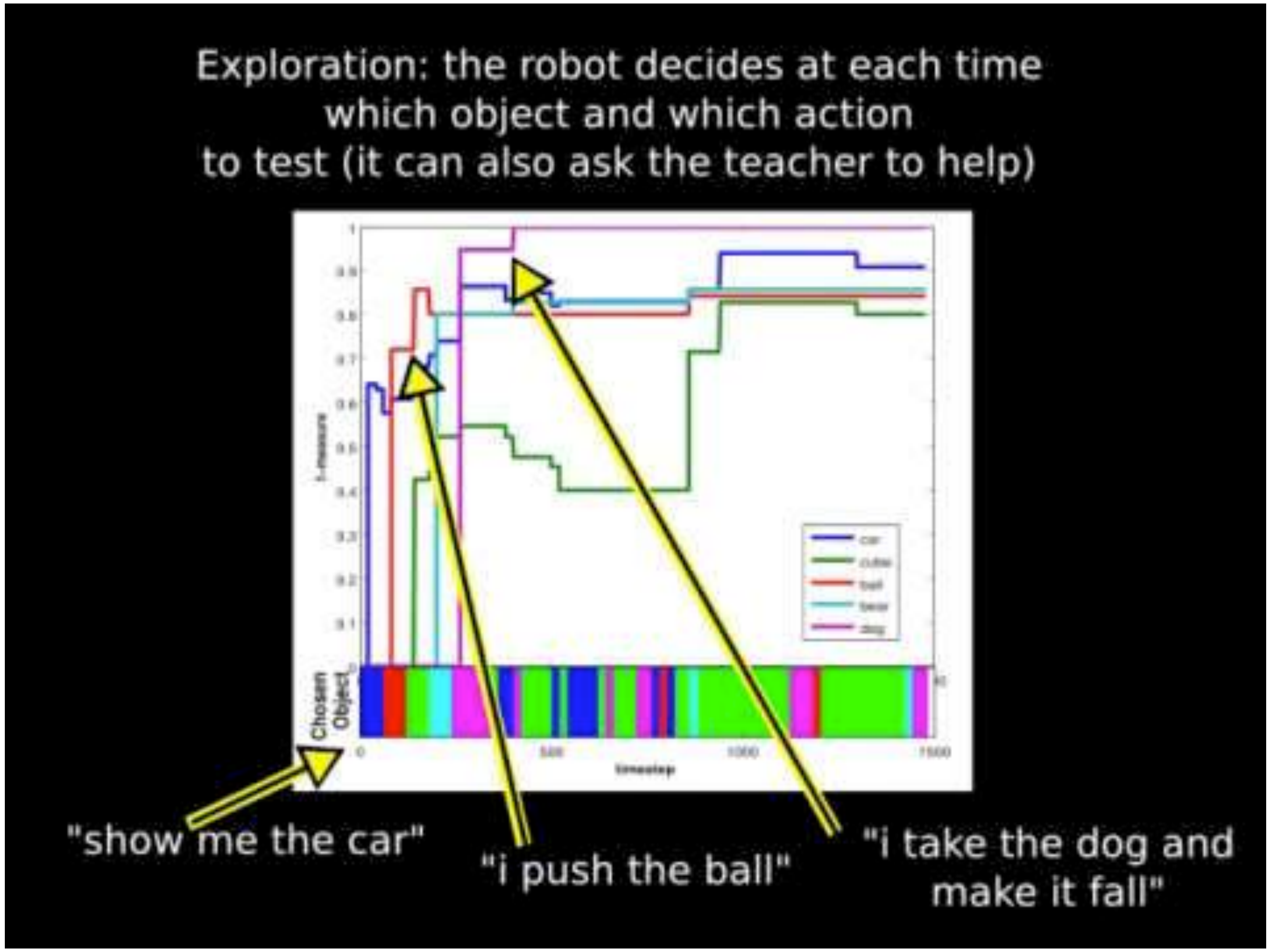
robot asks human to show a new object



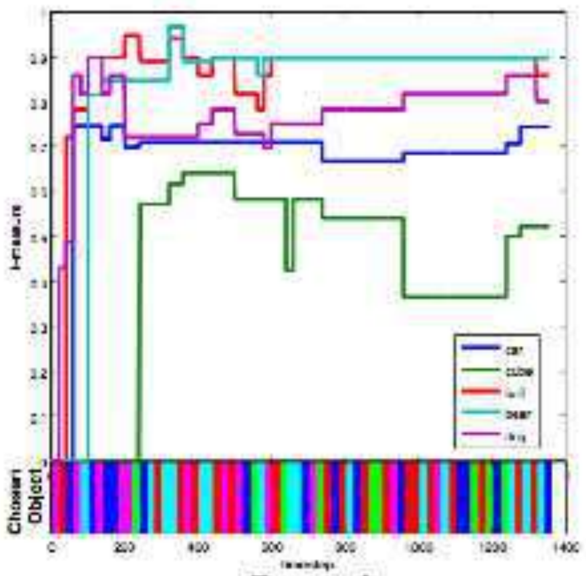
robot pushes the object

Curiosity-driven exploration of objects

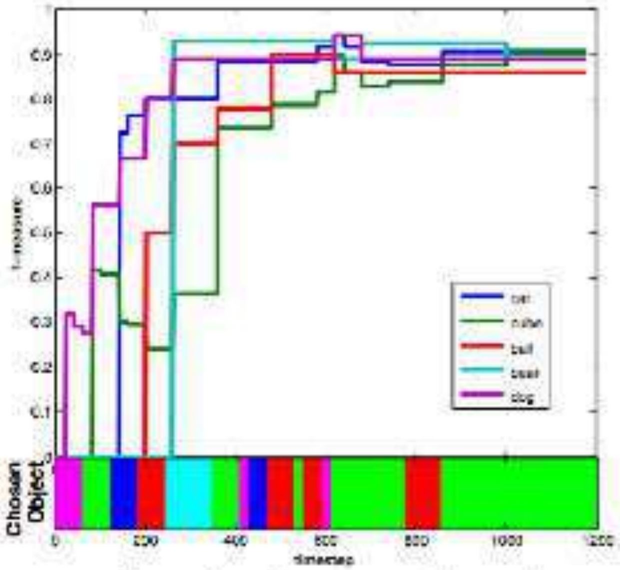
- Focusing on the objects that are not yet learned
- choosing the appropriate action for each object



ball → yellow car → red bear → …



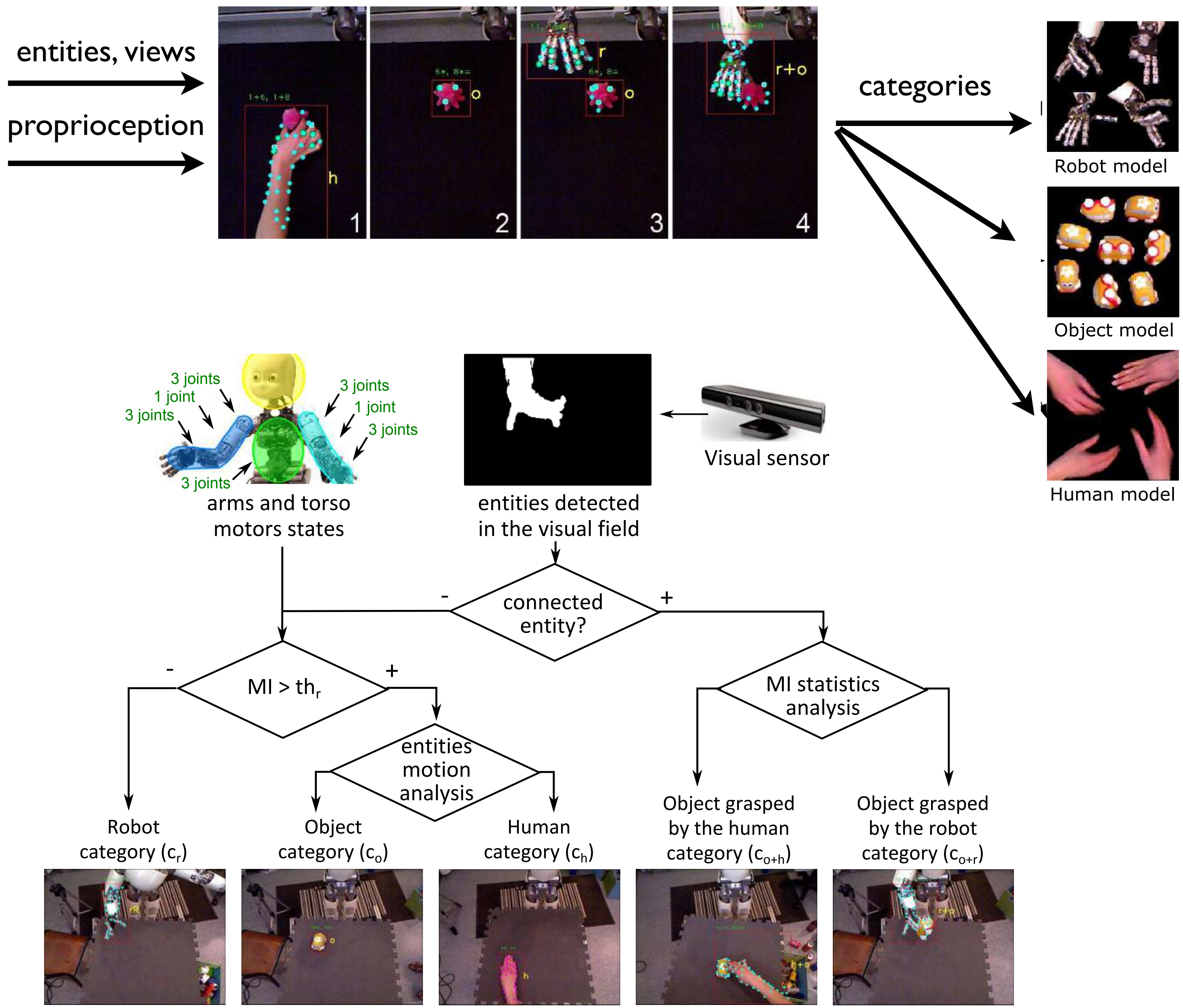
Random



Intrinsic motivation

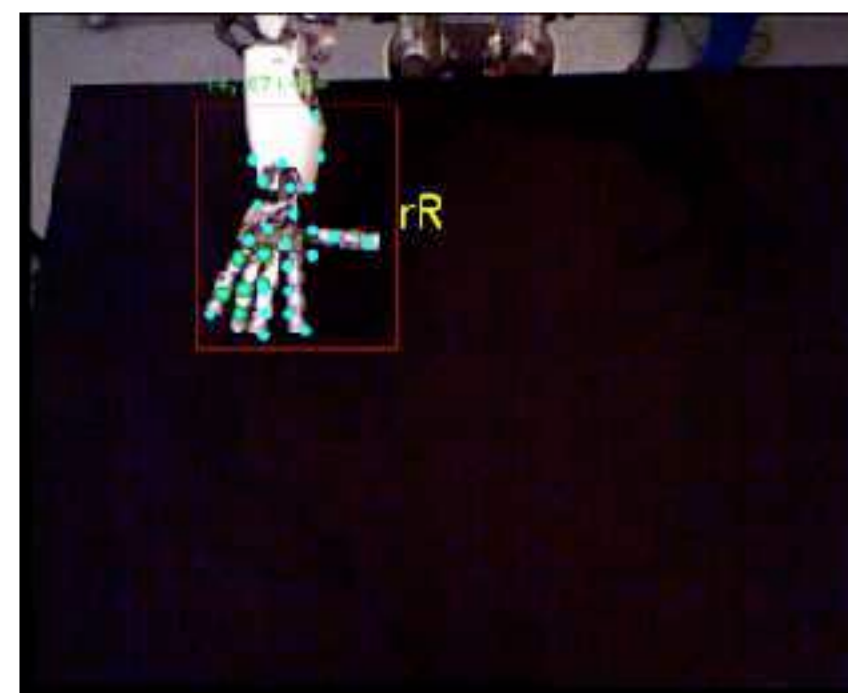
Ivaldi, Nguyen, Lyubova, Droniou, Padois, Filliat, Oudeyer, Sigaud (2014) Object learning through active exploration. IEEE Transactions on Autonomous Mental Development.

Better learning with action and interaction



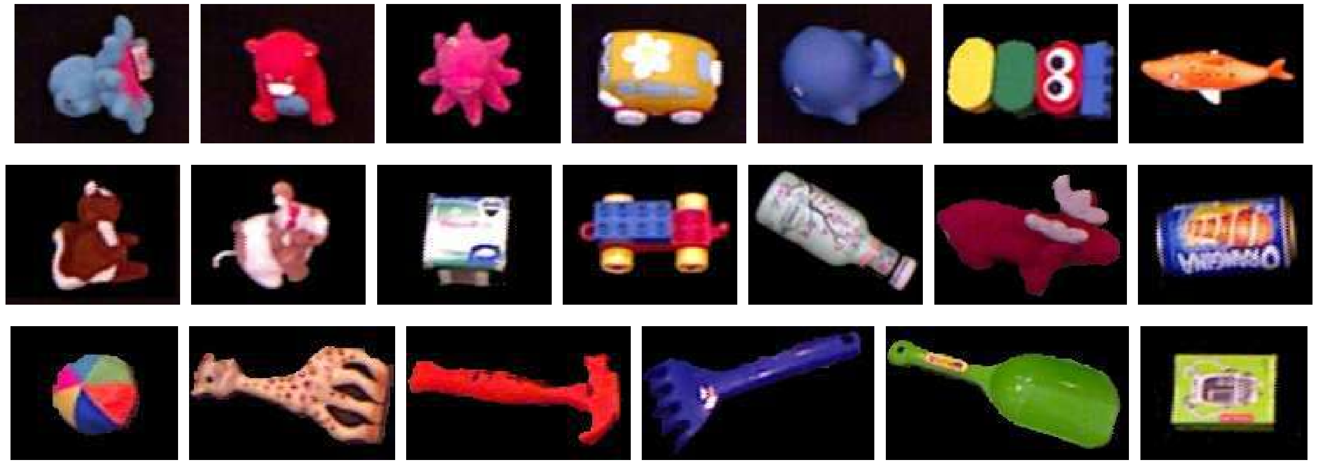
The robot learns the objects through manipulation.

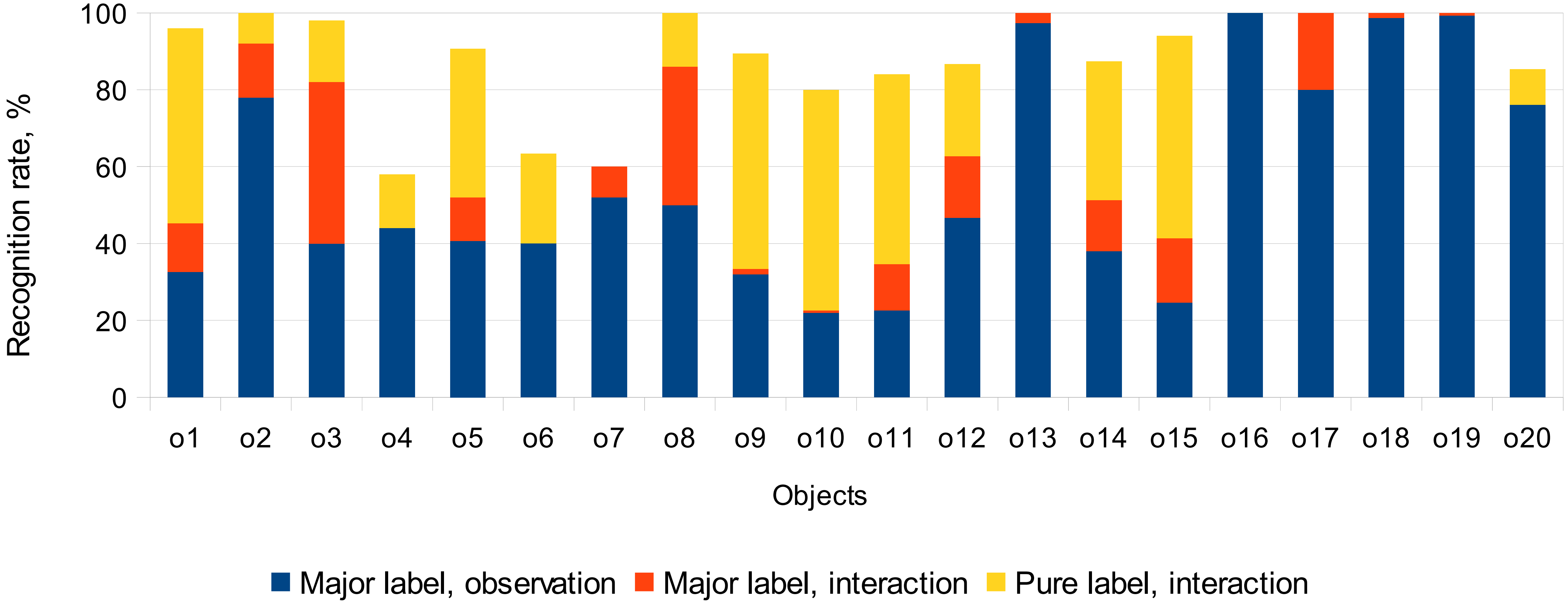
The robot learns to identify its body, hence entities can be categorized as "robot hand", "human hand" and "object".



Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

Better object recognition





Lyubova, Ivaldi, Filliat (2016) From passive to interactive object learning and recognition through self-identification on a humanoid robot. Autonomous Robots, 40(1):33-57.

Outline

Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Multimodal learning of the visual appearance of objects (w/ Kinect)

Demonstrating assembly with kinesthetic teaching

Demonstrating whole-body grasping with tele-operation

HEAP project

HEAP object dataset

Ordinary people teach iCub how to assembly an object

1. How do people behave (gaze, touch, posture, ...) during physical interaction?
2. How much force do they apply on the robot?
3. Do these measures change depending on their expertise with robots, their personality and attitudes?
4. Could we use the demonstrated trajectories to teach autonomous assembly to the robot?

Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Ordinary people teach iCub how to assemble an object

- 56 subjects
- age : 36,95±14,32 (min 19, max 65)
- sex : 19 male, 37 females

naive participants (no experience in robotics)

Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Ordinary people teach iCub how to assembly an object

1. How do people behave (gaze, touch, posture, ...) during physical interaction?

- Extroverts talk more to the robot
- People with higher NARS (negative attitude towards robot) look more at the hands and less at the robot face
- Tendency to be far from the robot

2. How much force do they apply on the robot?

- Variable, men tend to apply less contact force than women
- Extroverts and old people apply less force
- People with high NRS apply bigger forces

3. Do these measures change depending on their expertise with robots, their personality and attitudes?

- Yes!

4. **Could we use the demonstrated trajectories to teach autonomous assembly to the robot?**

- No, too much noise, too much variability. Primitives not sufficient.

Ivaldi, S.; Lefort, S.; Peters, J.; Chetouani, M.; Provasi, J.; Zibetti, E. (2016) Towards engagement models that consider individual factors in HRI: on the relation of extroversion and negative attitude towards robots to gaze and speech during a human-robot assembly task. Int. Journal Social Robotics

Outline

Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)

Multimodal learning of the visual appearance of objects (w/ Kinect)

Demonstrating assembly with kinesthetic teaching

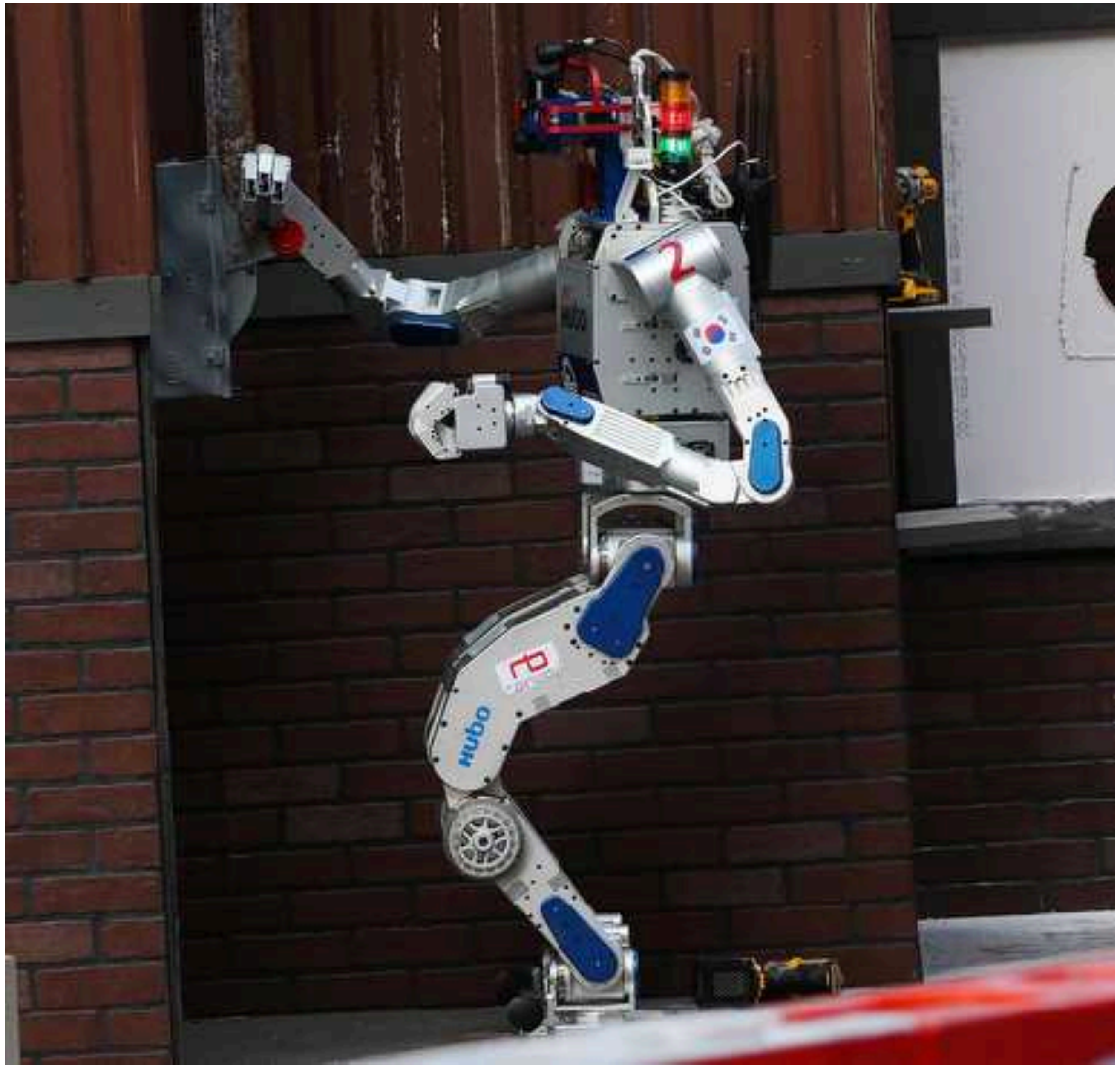
Demonstrating whole-body grasping with tele-operation

HEAP project

HEAP object dataset

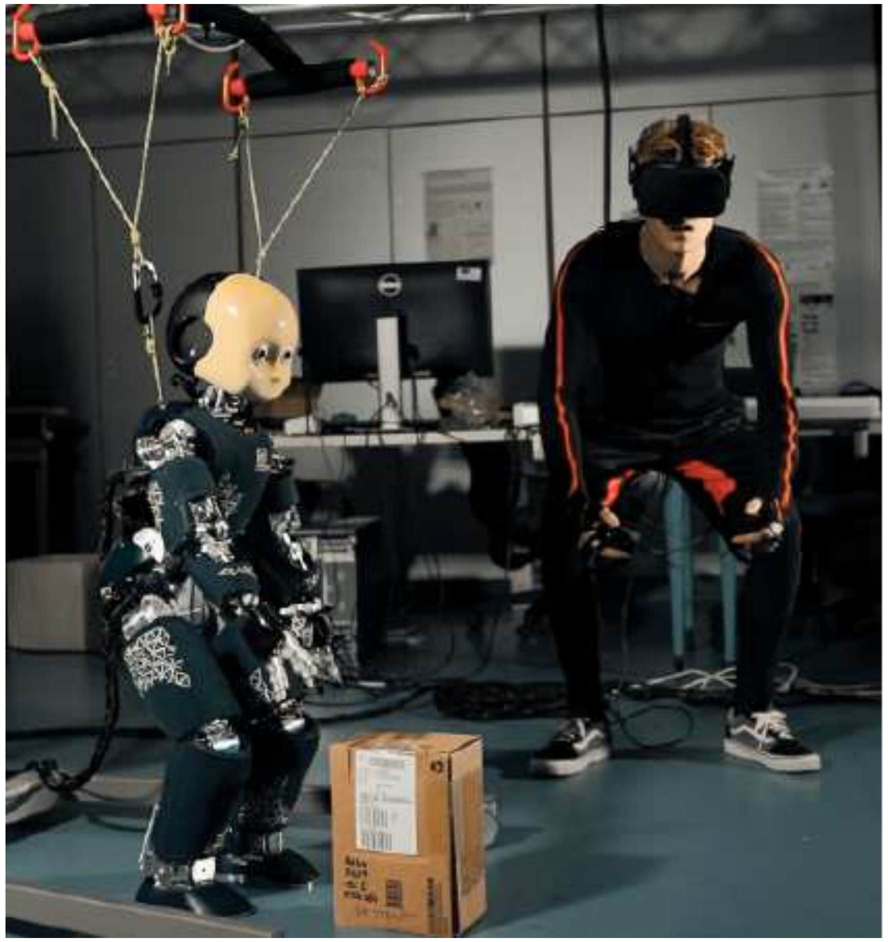
Towards whole-body grasping & manipulation

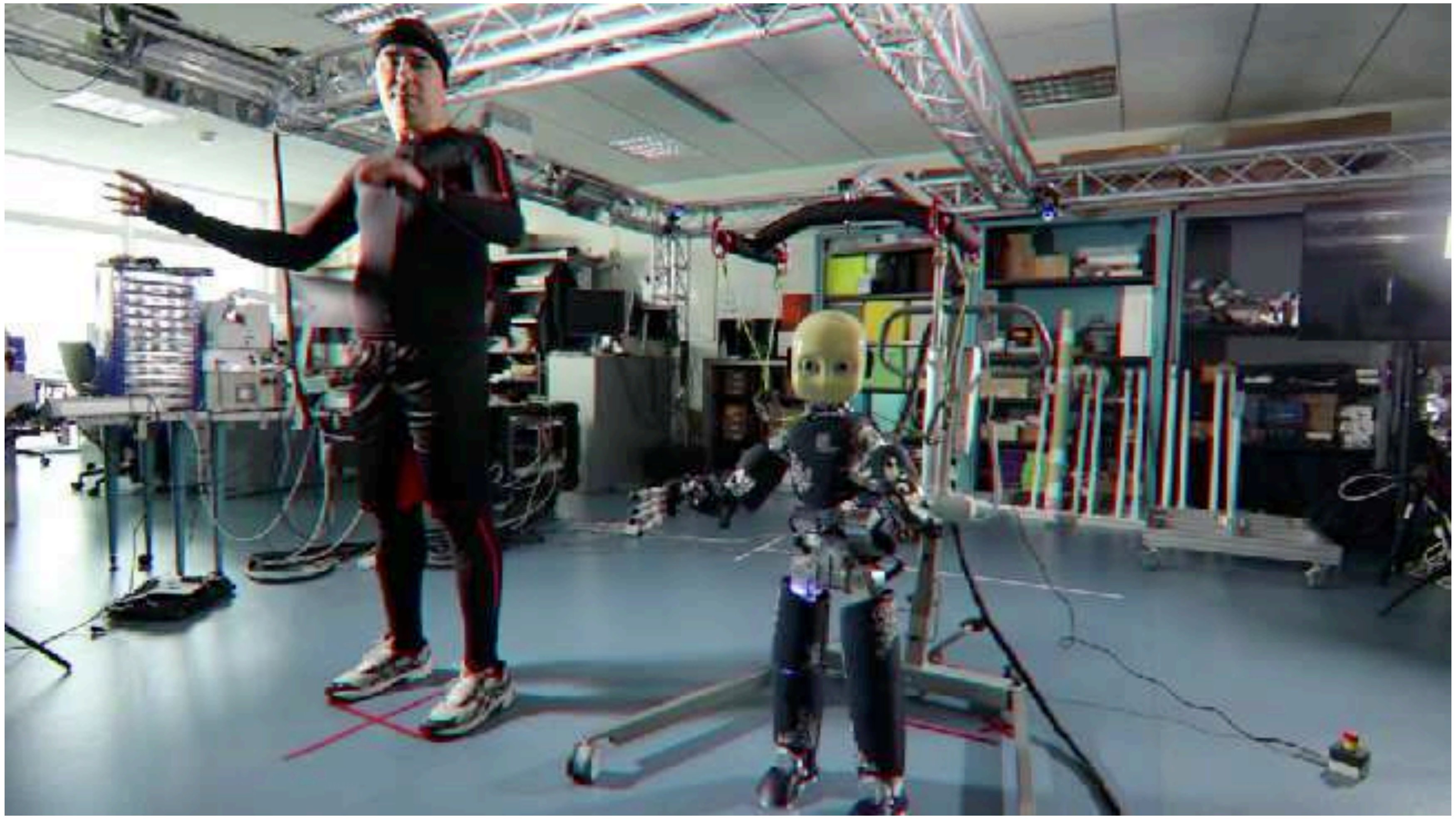
Grasping an object is a particular task adding to the balancing and locomotion tasks of the whole-body robot controller.



Towards whole-body grasping & manipulation

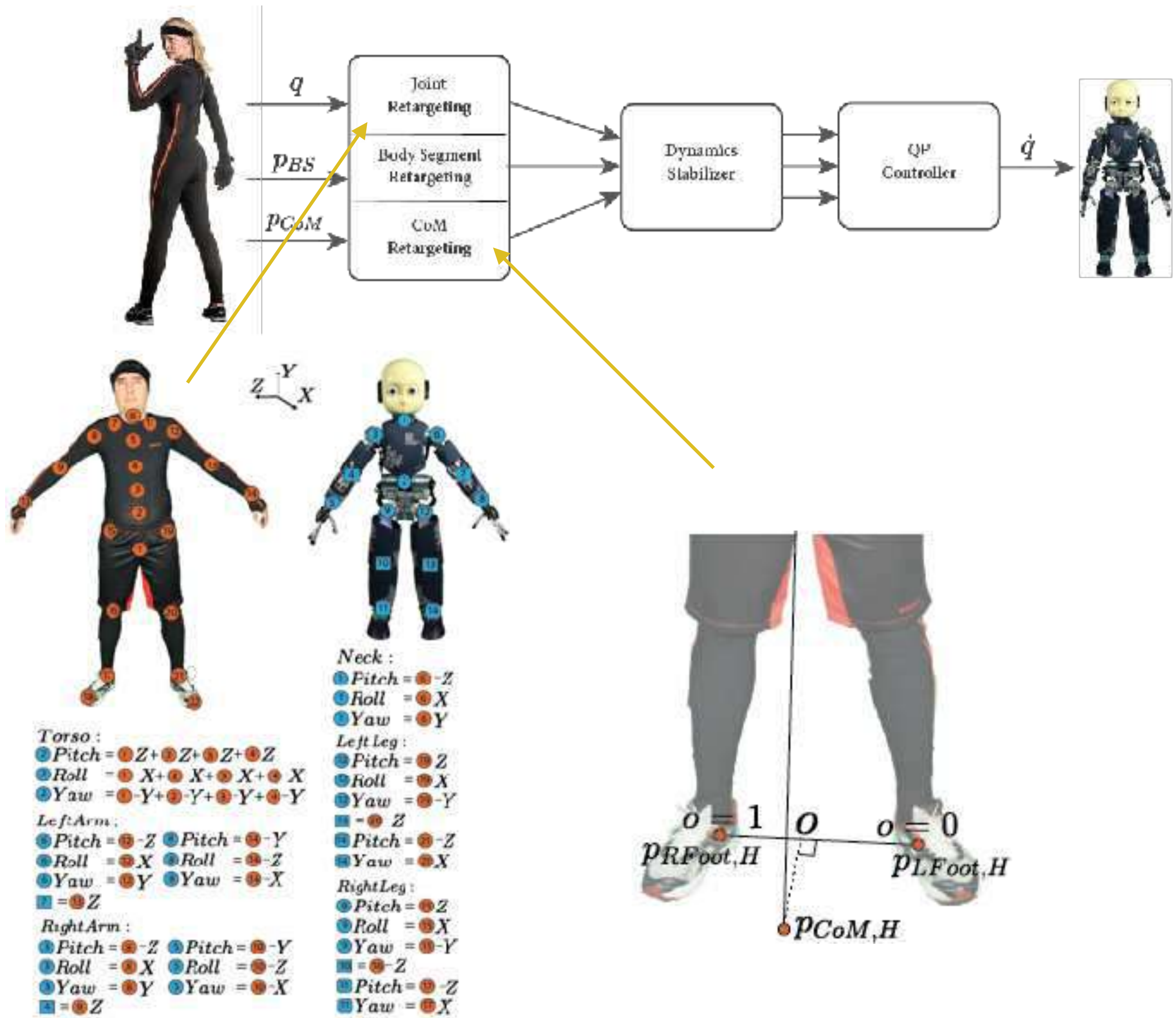
Tele-operation/retargeting is the whole-body kinesthetic teaching





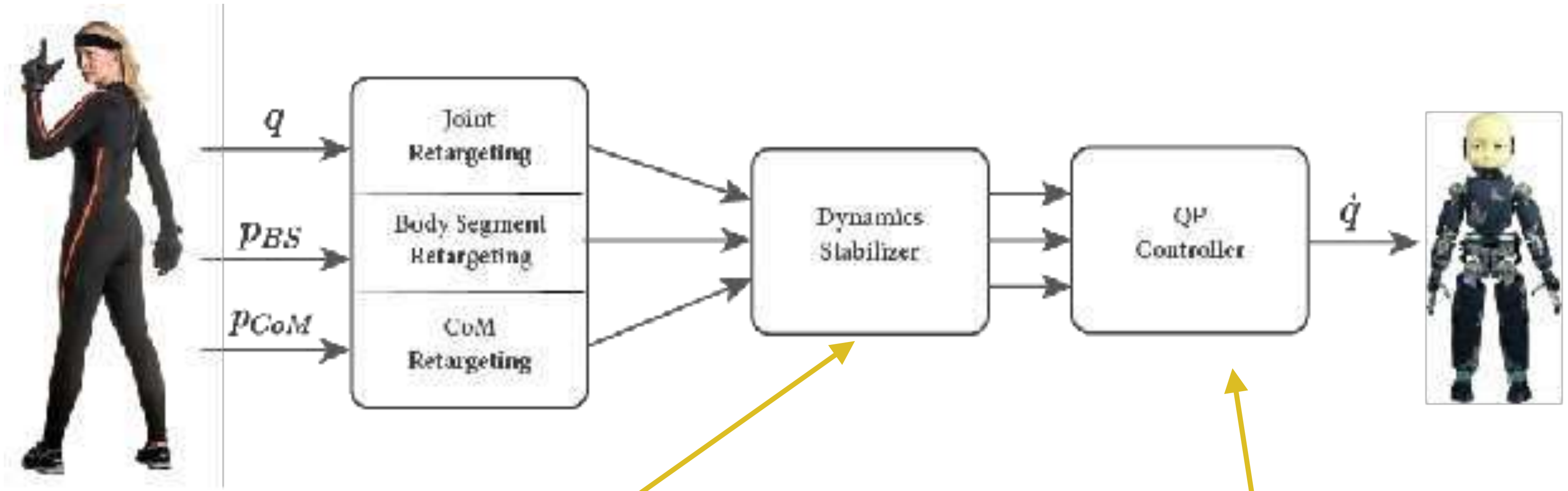
Penco et al (2018) **Robust real-time whole-body motion retargeting from human to humanoid.**
Proc. IEEE/RAS International Conf. on Humanoid Robots (HUMANOIDS).

Towards whole-body grasping & manipulation



Penco et al (2018) **Robust real-time whole-body motion retargeting from human to humanoid.** Proc. IEEE/RAS International Conf. on Humanoid Robots (HUMANOIDS).

Towards whole-body grasping & manipulation



$$\begin{aligned} \min_{p_{ZMP}} \quad & (\dot{p}_{CoM}^{des} - \dot{p}_{CoM})^T R\, (\dot{p}_{CoM}^{des} - \dot{p}_{CoM}) \\ \text{s.t.} \quad & \dot{p}_{CoM} = \dot{p}_{CoM}^{t-1} + \frac{Tg}{h^{t-1}}(p_{CoM} - p_{ZMP}) \\ & lb_{SP} < p_{ZMP} < ub_{SP} \end{aligned}$$

ZMP correction

$$\min_{x} \;|| \; Ax - b \; ||_W^2 \qquad s.t. \quad l \leq Cx \leq u$$

$$stack = (lFoot + rFoot + head_{sub}) / \\ (com_{sub} + base_{sub} + torso_{sub} + lArm_{sub} + rArm_{sub});$$

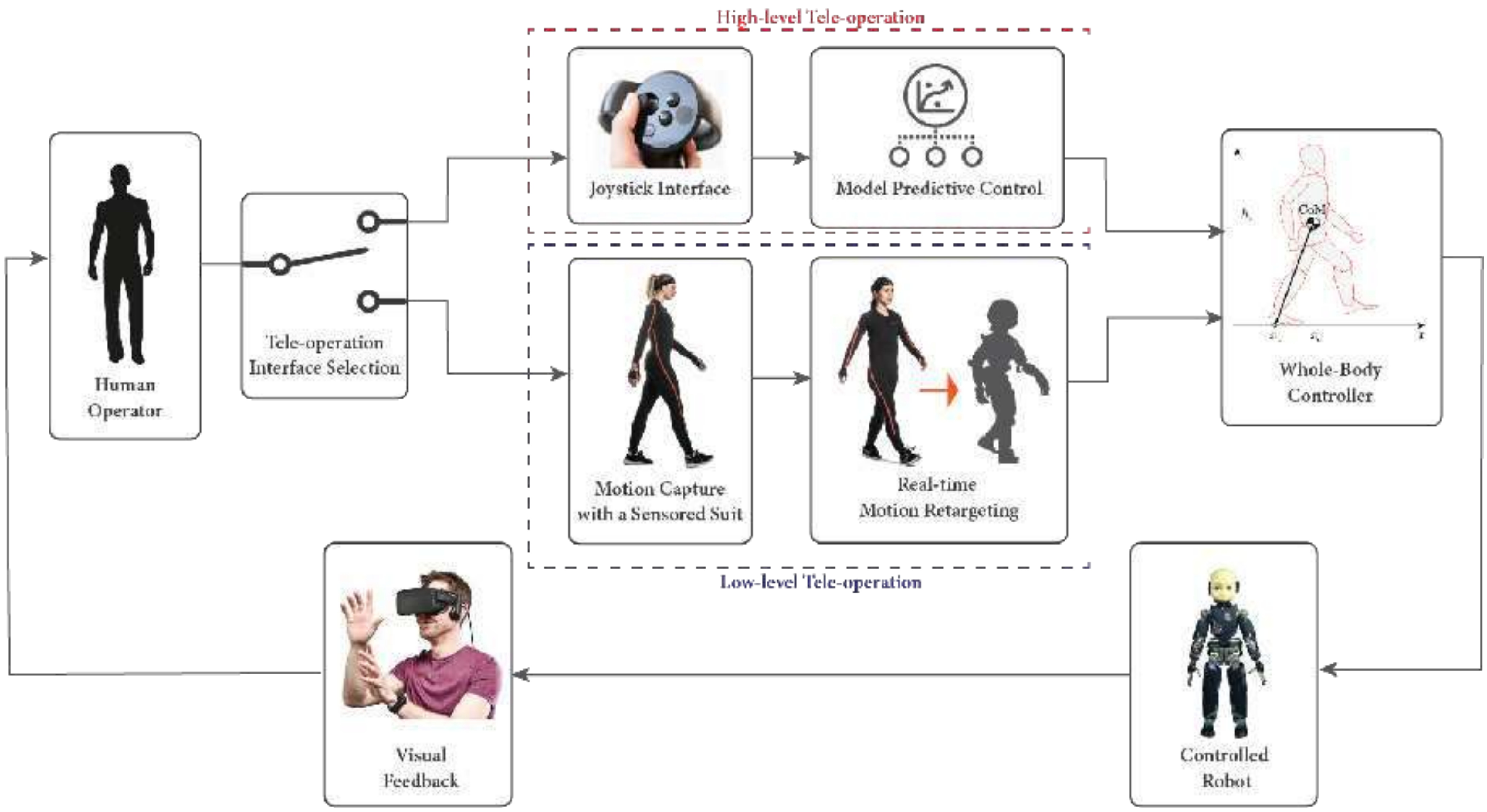
Multi-task QP controller with strict priorities (OpenSoT)

Grasp-related tasks must be put here!

Penco et al (2018) **Robust real-time whole-body motion retargeting from human to humanoid.** Proc. IEEE/RAS International Conf. on Humanoid Robots (HUMANOIDS).

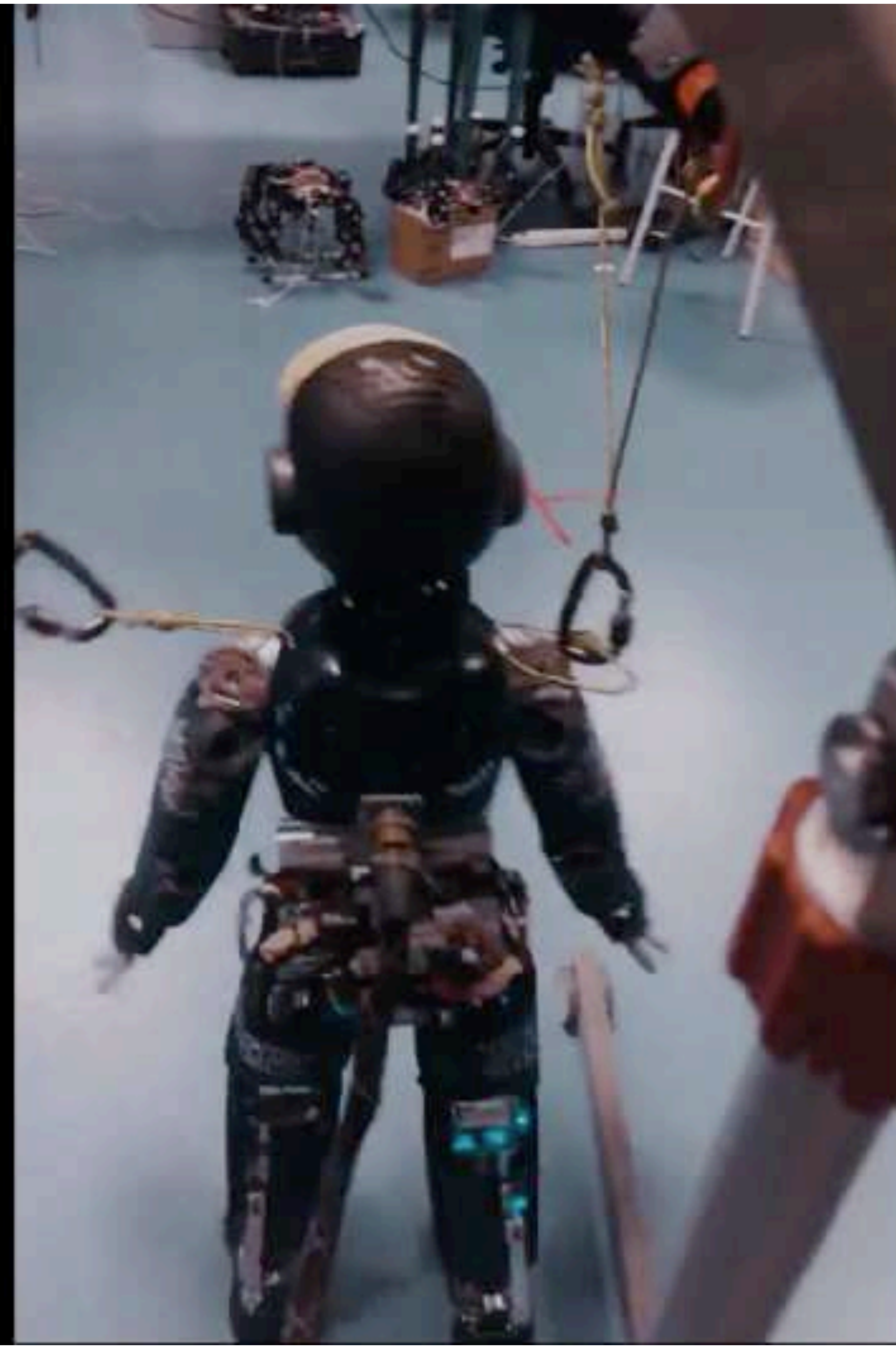
Towards whole-body grasping & manipulation

Tele-operation/retargeting is the whole-body kinesthetic teaching



Penco*, L.; Scianca*, N.; Modugno*,V.; Lanari, L.; Oriolo, G.; Ivaldi, S. **A Multi-Mode Teleoperation Framework for Humanoid Loco-Manipulation.**

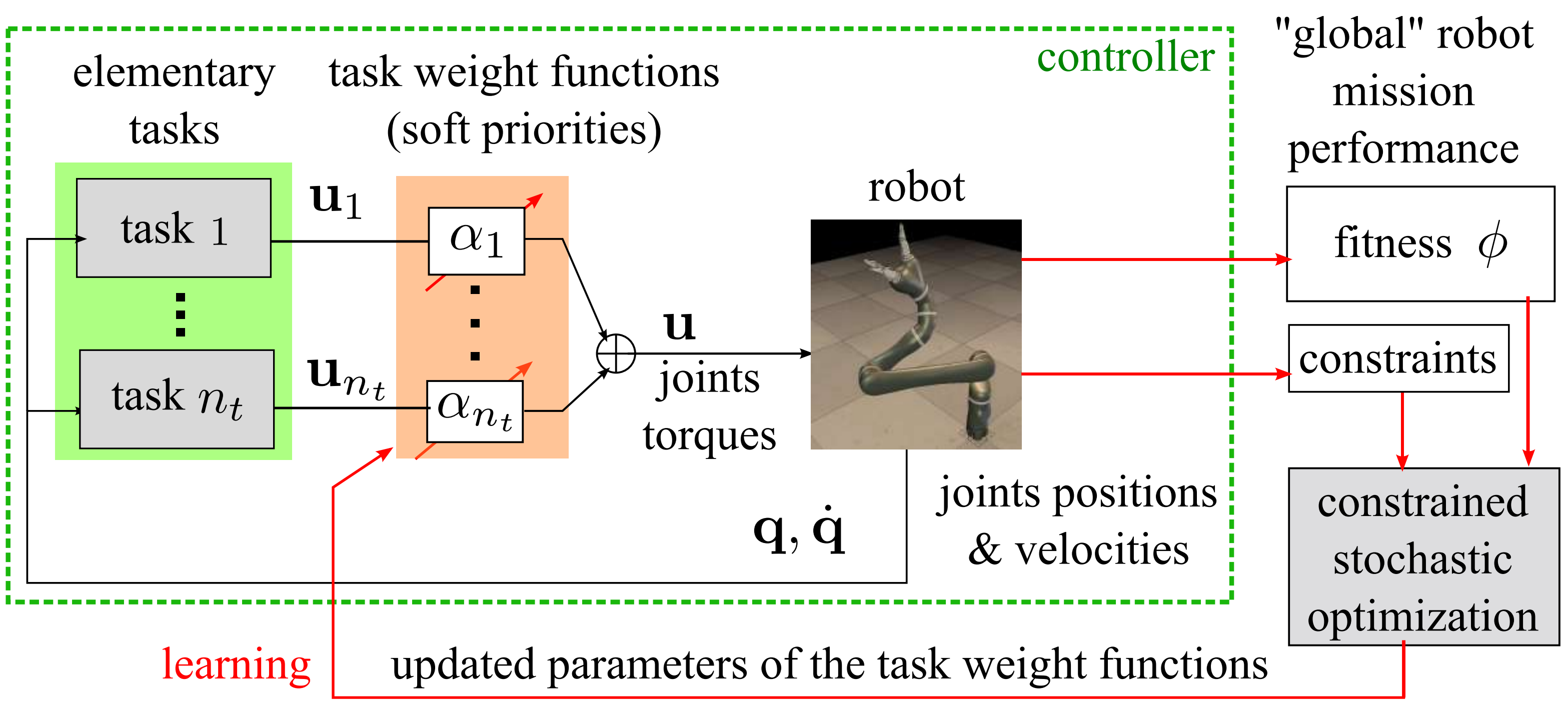
Towards whole-body grasping & manipulation



Penco et al (2018) **Robust real-time whole-body motion retargeting from human to humanoid.**
Proc. IEEE/RAS International Conf. on Humanoid Robots (HUMANOIDS).

Towards whole-body grasping & manipulation

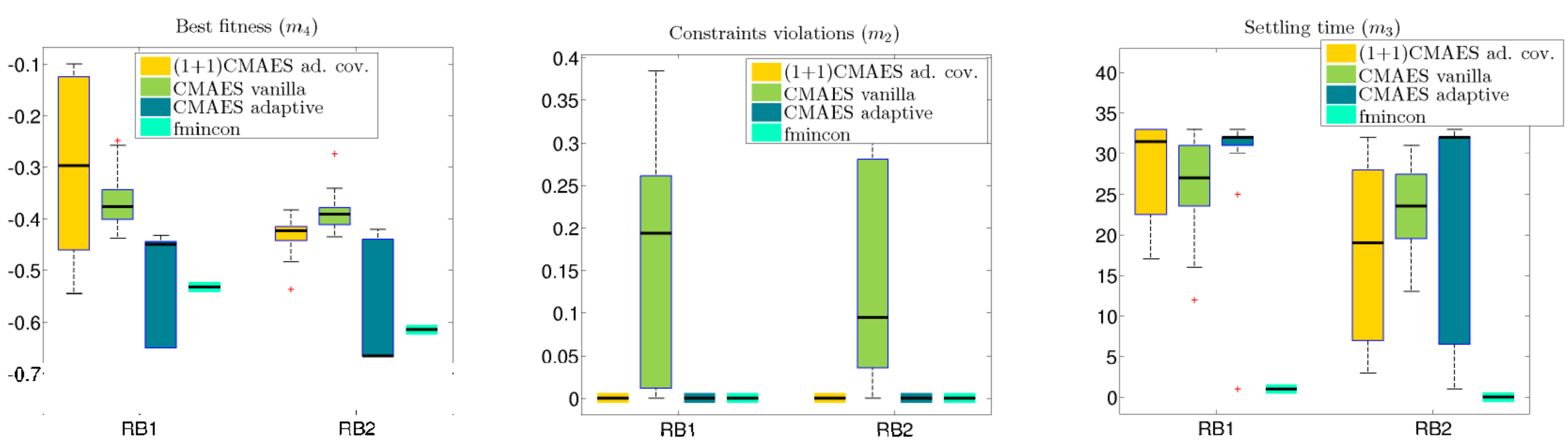
How to optimize the task priorities and/or the desired task trajectories to make sure the robot fulfill its mission while never violating the problem's constraints?



$$\boldsymbol{w}^{\circ} = \arg\max_{\boldsymbol{w}} \phi(\boldsymbol{w})$$

subject to

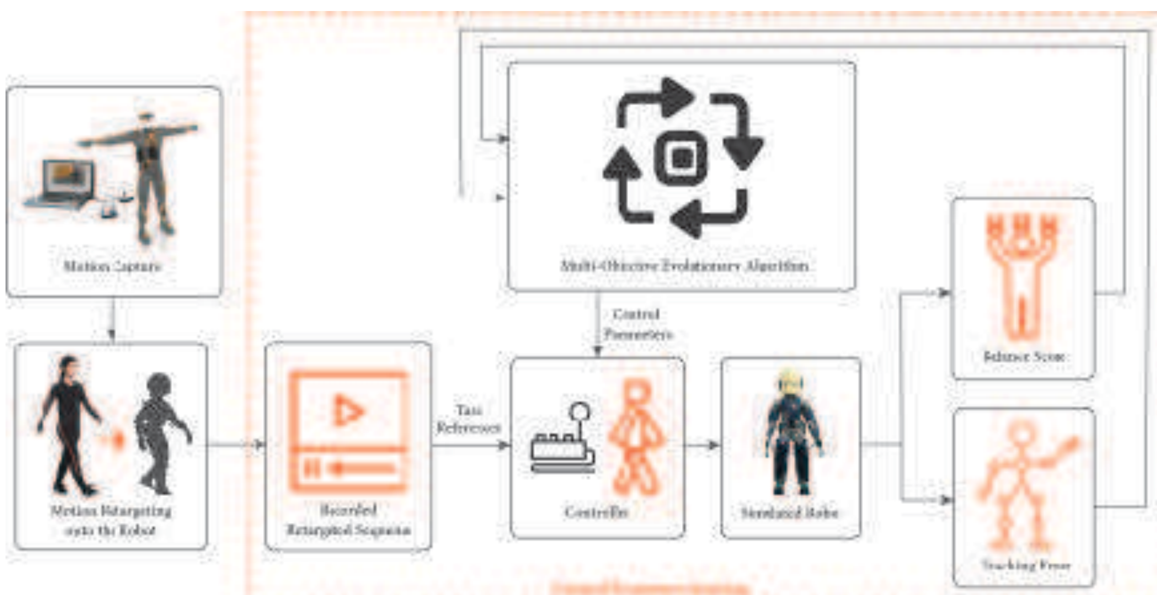
$$g_i(\boldsymbol{w}) \leq 0 \quad i = 1, \ldots, n_{IC}$$
$$h_i(\boldsymbol{w}) = 0 \quad i = 1, \ldots, n_{EC}$$



Modugno et al. (2016, 2017) Humanoids, Charbonneau et al. (2018) Humanoids

Towards whole-body grasping & manipulation

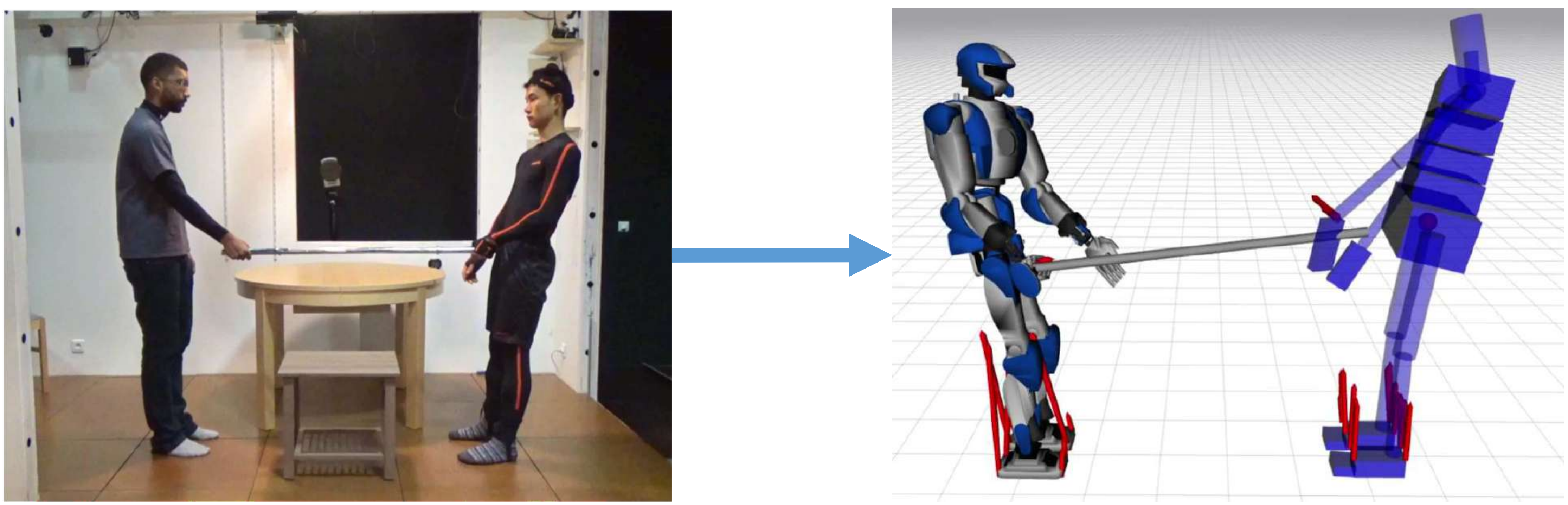
Learning the control structure and the parameters that enable the robot to perform a variety of motions, including whole-body manipulation



Penco et al. - in submission for Humanoids 2019

Whole-body co-manipulation with a human

- Take into account the entire human dynamics in a multi-task QP controller for collaborative manipulations
- Joint level controller for the robot, but capable of reacting to the human



Human collaborator (to be replaced by robot)
Recorded human
Robot collaborator
Simulated human

K. Otani, K. Bouyarmane, S. Ivaldi (2018) ICRA

Towards whole-body grasping & manipulation

Proposed approach:

- Model the human as a robot → multi-robot QP controller
- Reason in terms of balance of the couple human+robot, not robot only

$$\underset{\ddot{q},\tau,f}{\text{minimize}} \quad \sum_k w_k ||\ddot{g}_k - \ddot{g}_k^{des}||^2$$

Individual or combined tasks (e.g. combined CoM for balance)

$$\text{subject to} \quad M\ddot{q} + N = J_0^T F^0 + (J_1 - \Psi^T J_2)F^- + S\tau$$

$$J_0\dot{q} = 0$$

Equal and opposite contact forces between robots

$$(J_1 - \Psi^T J_2)\dot{q} = 0$$

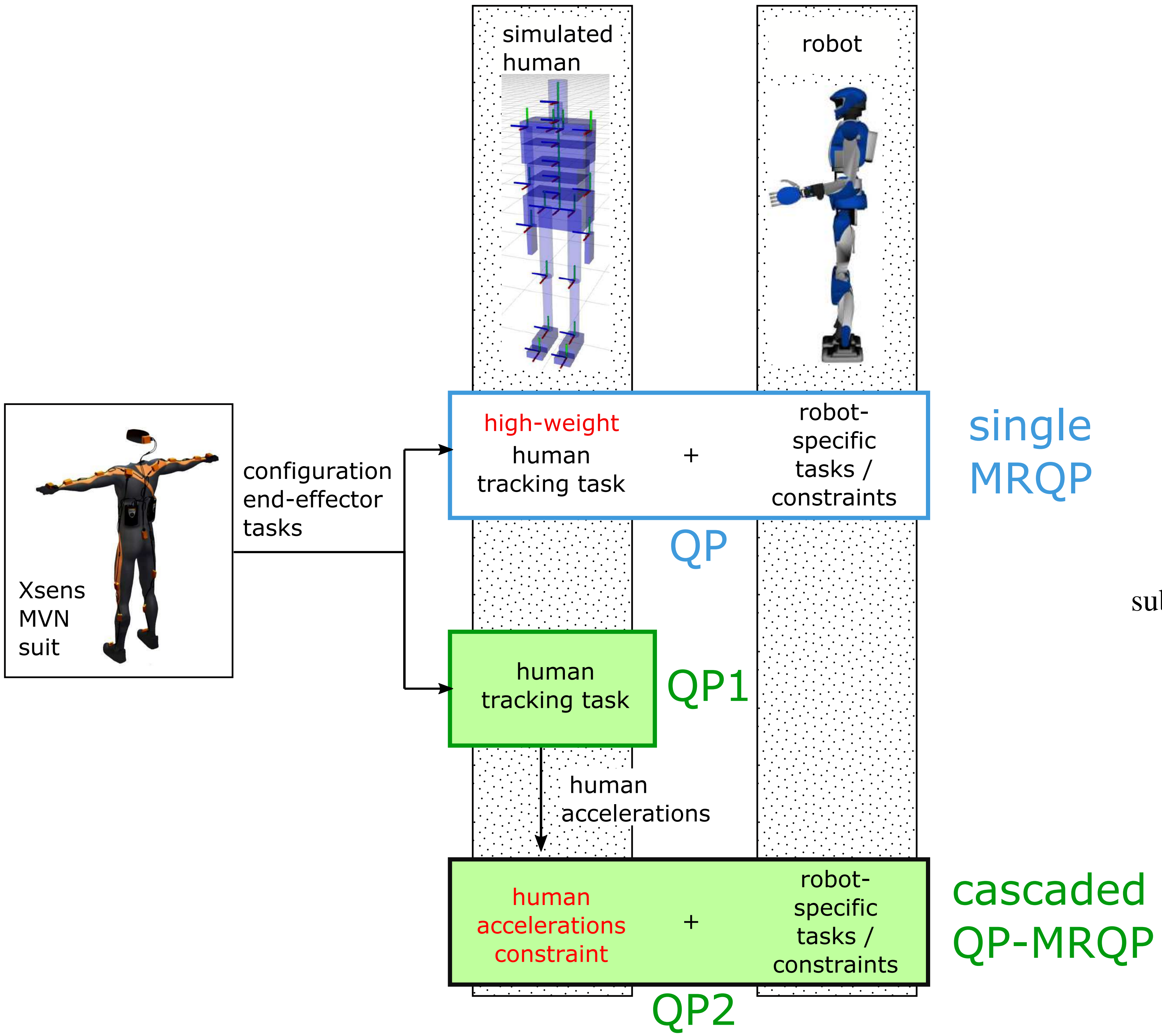
Non-slipping contacts between robots

$$f \in \mathcal{C}$$

torque limits, joint limits, collision avoidance

K. Otani, K. Bouyarmane, S. Ivaldi (2018) ICRA

Towards whole-body grasping & manipulation



$$\min_{\ddot{q},\tau,\lambda^0,\lambda^1} \sum_k w_k ||\ddot{y}_k - \ddot{y}_k^d||^2 ,$$

$$\text{subject to: } \begin{cases} \text{combined dynamics equation} \\ \text{contacts no-slip (with environment)} \\ \text{forces within friction cones} \\ \text{joint position, velocity, torque limits} \\ \text{collision avoidance} \\ \text{contacts no-slip (between robots)} \end{cases}$$

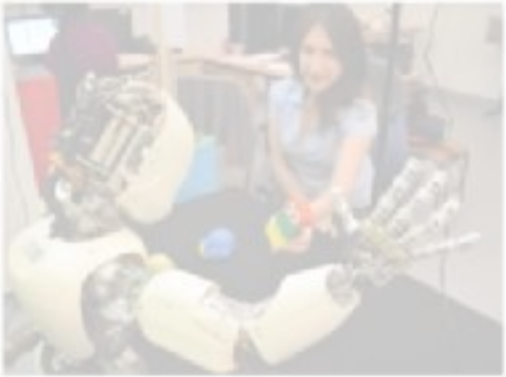
K. Otani, K. Bouyarmane, S. Ivaldi (2018) ICRA

Towards whole-body grasping & manipulation

K. Otani, K. Bouyarmane, S. Ivaldi (2018) ICRA

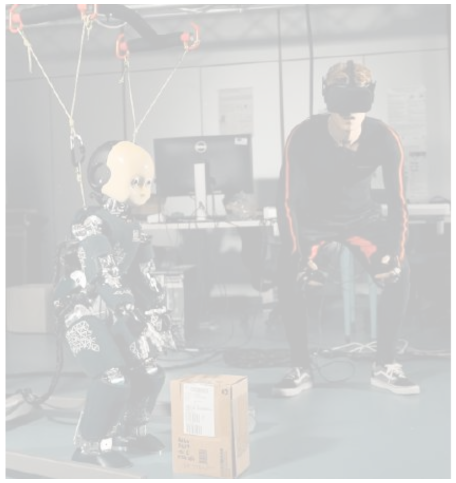
Outline

Grasping objects localised by noisy point clouds, acquired by stereo cameras (w/ eyes)



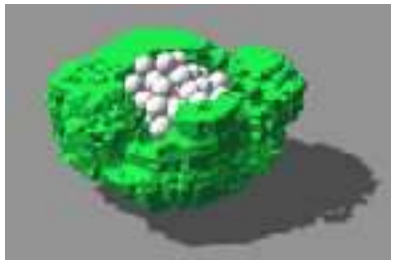
Multimodal learning of the visual appearance of objects (w/ Kinect)

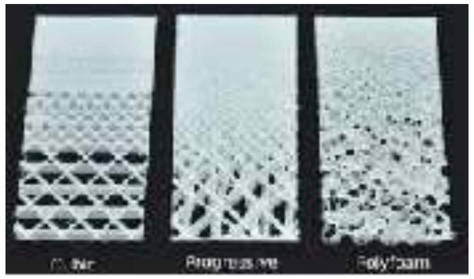
Demonstrating assembly with kinesthetic teaching



Demonstrating whole-body grasping with tele-operation

HEAP project





HEAP object dataset

EU Project HEAP

EU Project HEAP

Human guided learning of robotic heap sorting

- Chist-Era Call:
  Object recognition and manipulation by robots: Data sharing and experiment reproducibility (2017)

- <u>Kick-off: May 2019</u>

- Consortium:
  - Univ. of Lincoln (Gerard Neumann)
  - Inria (Serena Ivaldi)
  - IIT (Lorenzo Natale)
  - TU Wien (Markus Vincze)
  - IDIAP (Sylvain Calinon)

EU Project HEAP

chist-era

Inspired by problems in nuclear waste sorting:

- irregular objects, articulated, often broken
  - 3D models are not always available
- items may break during grasping and transportation
- human operators

Nuclear waste

Nuclear waste

Nuclear waste

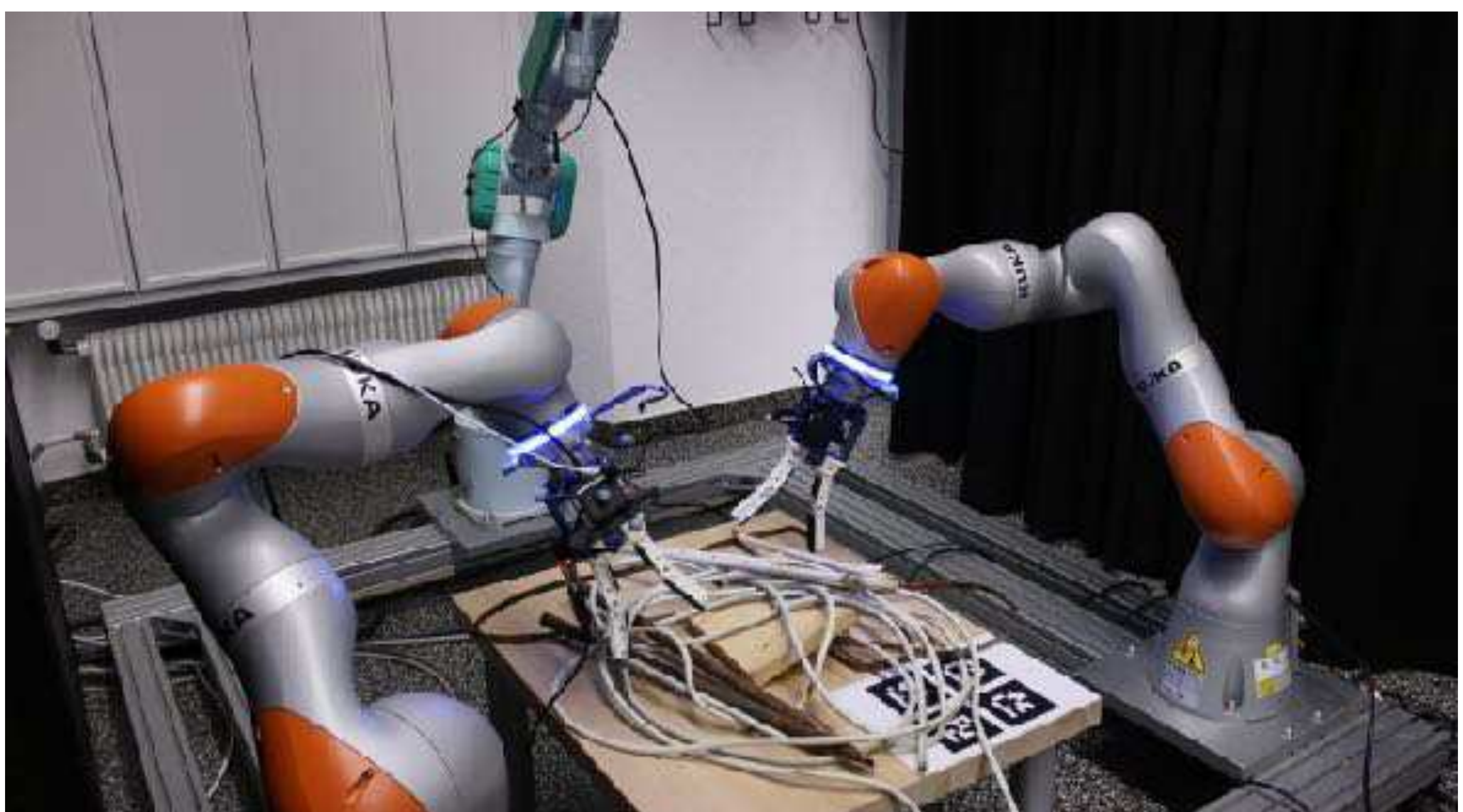
e- waste

EU Project HEAP

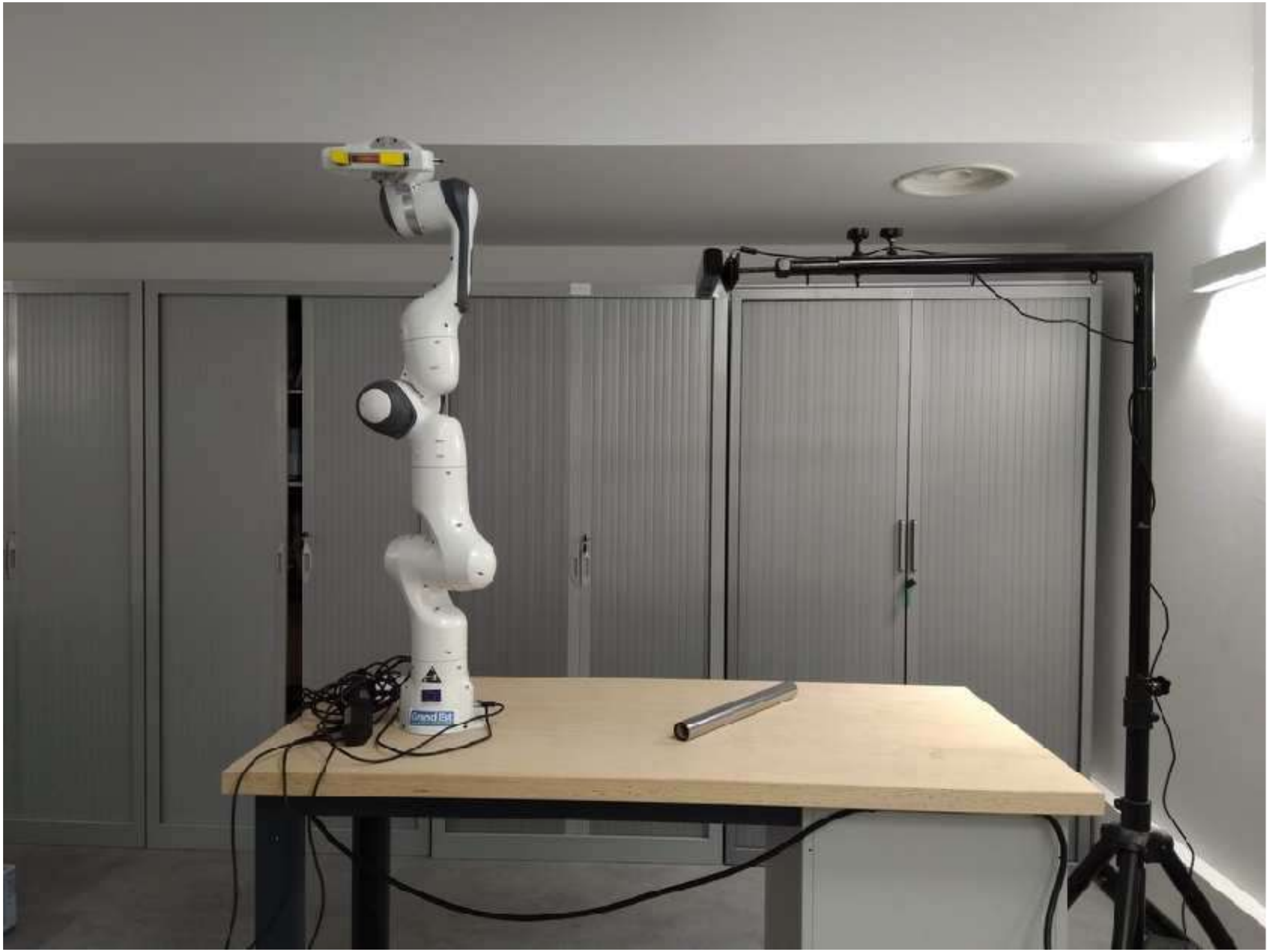
chist-era

Inspired by problems in nuclear waste sorting:

- irregular objects, articulated, often broken
  - 3D models are not always available
- items may break during grasping and transportation
- human operators



HEAP benchmarking setup



EU Project HEAP

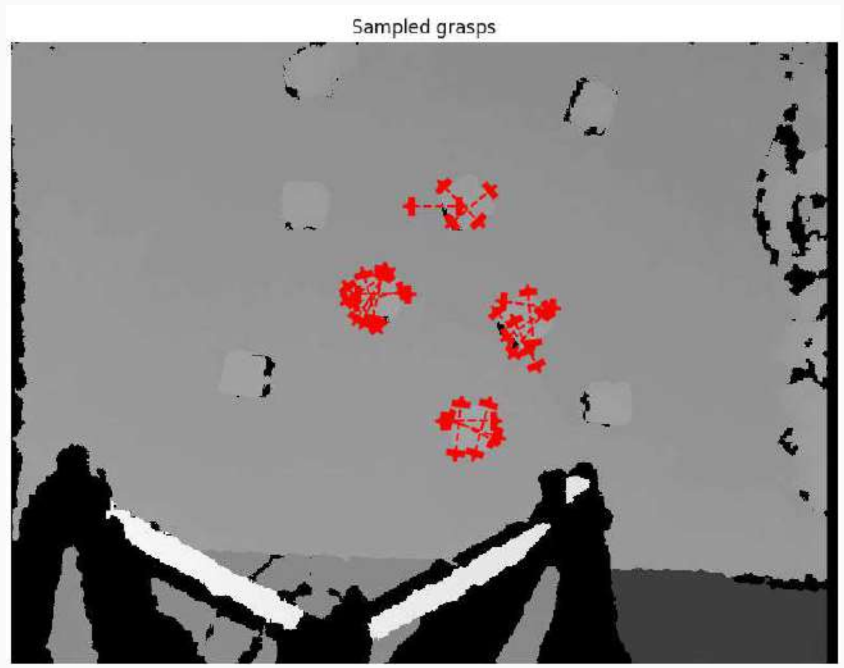
Human guided learning of robotic heap sorting

- Scientific challenges:
  - Improved object recognition, segmentation and grasp-pose estimation
  - Interactive perception
  - Human-in-the-loop and share control
  - Human guided learning
  - Reproducible Setups
  - Benchmarking

Why do we need human guidance?

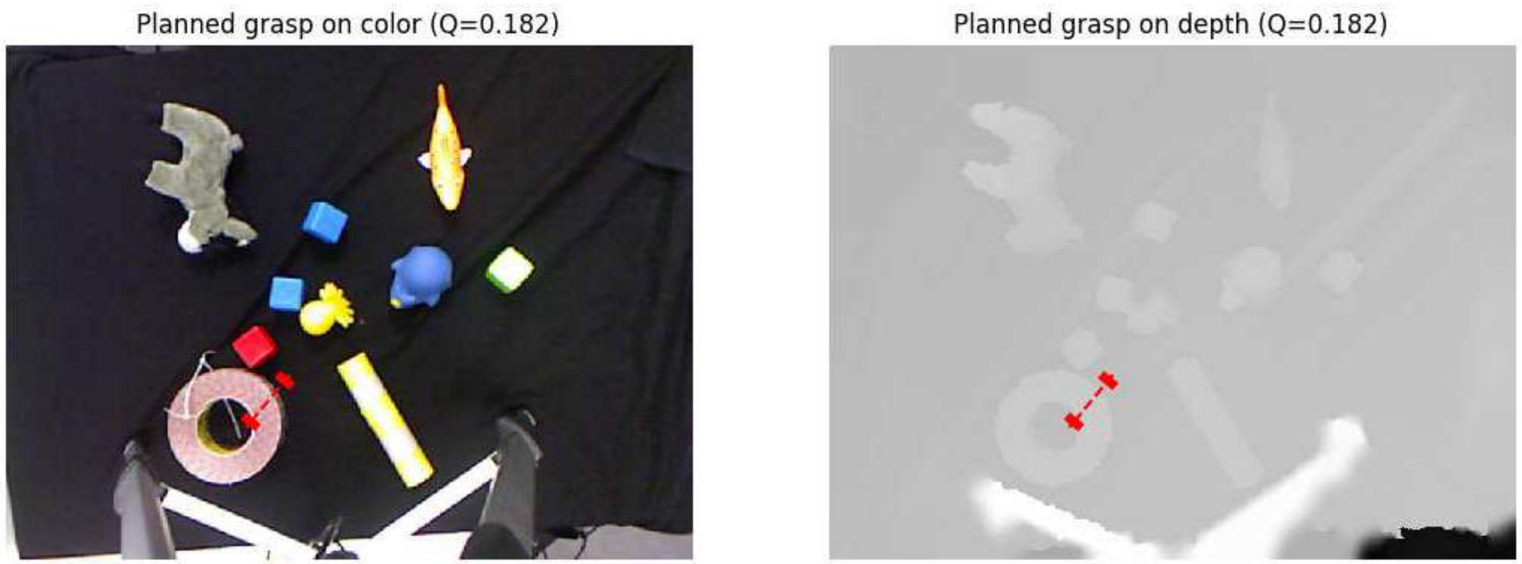
Why do we need human guidance

We already have plenty of grasping algorithms that we can use to find the best candidate grasp for the objects in the scene…



=> Dexnet 2.0
Malher et al., RSS 2017

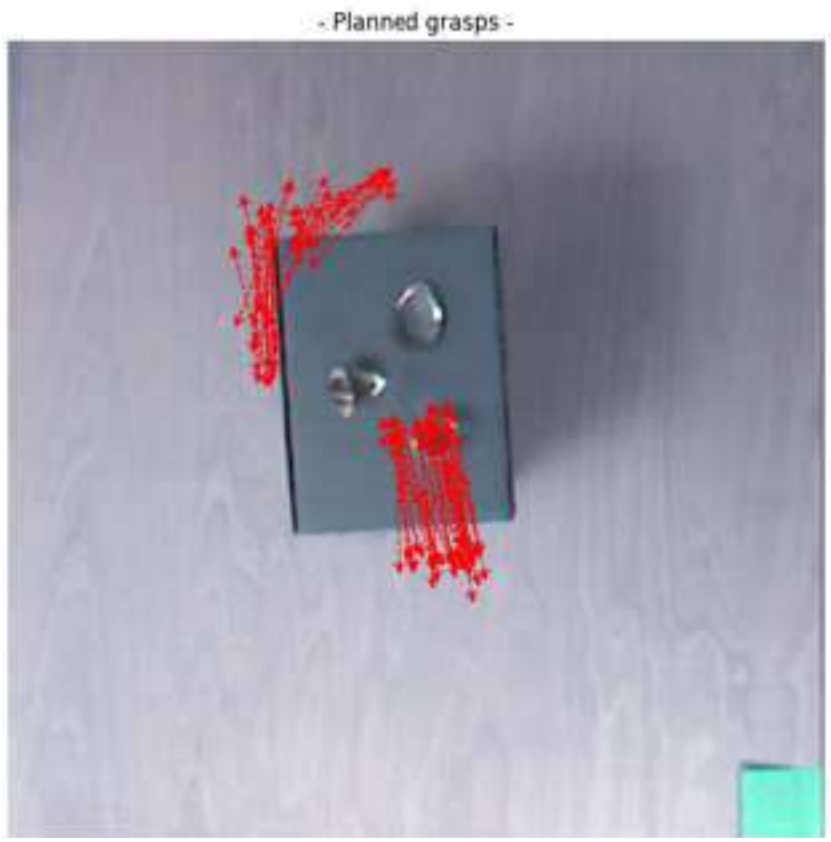
https://github.com/BerkeleyAutomation/gqcnn

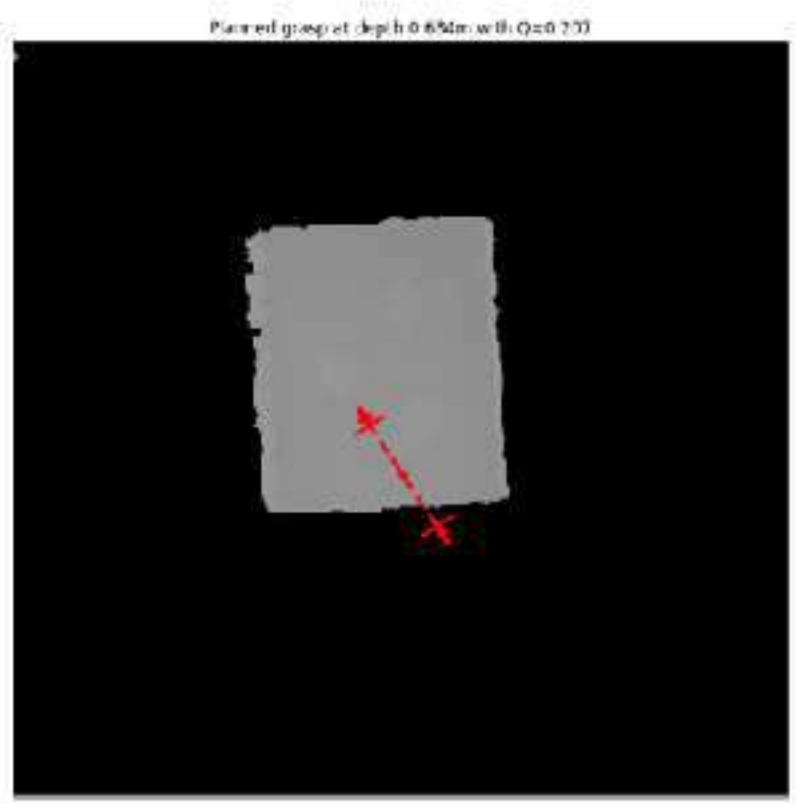


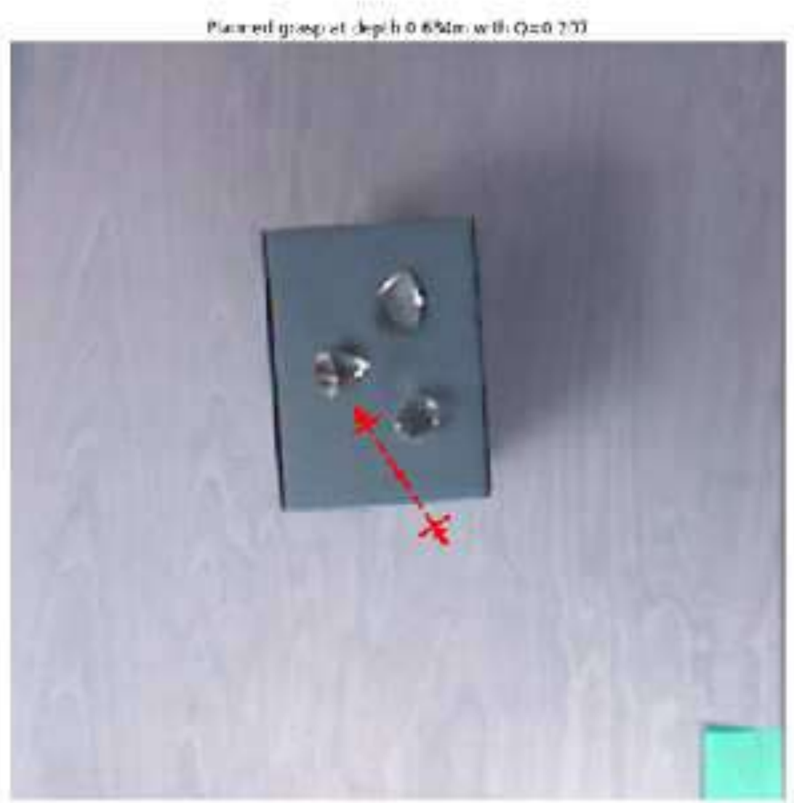
Why do we need human guidance

1) Because some objects challenge our cameras

➡ Crystals



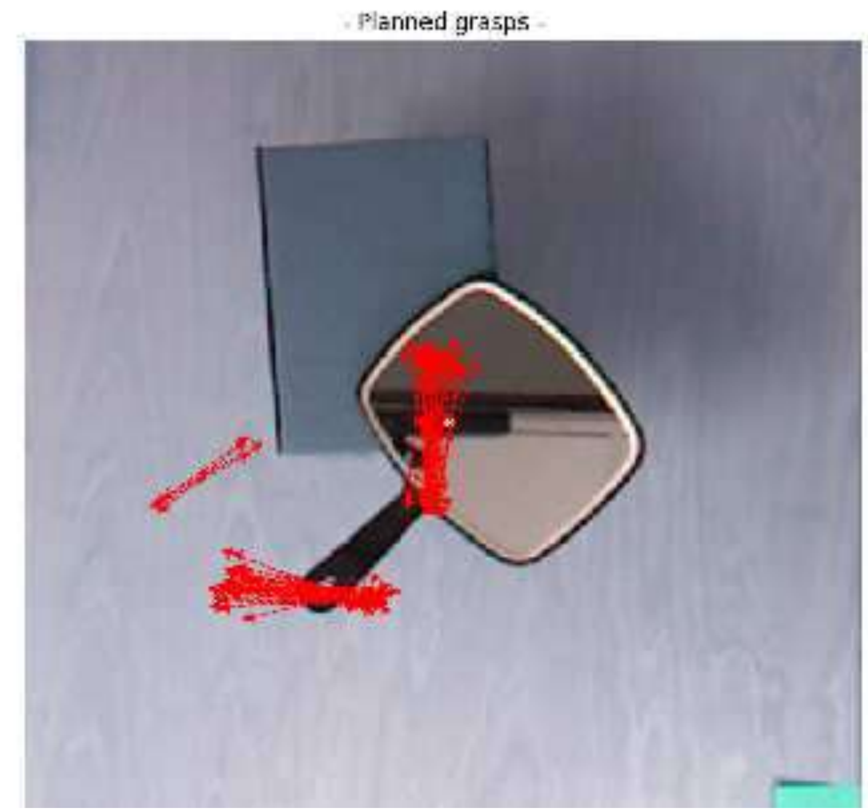


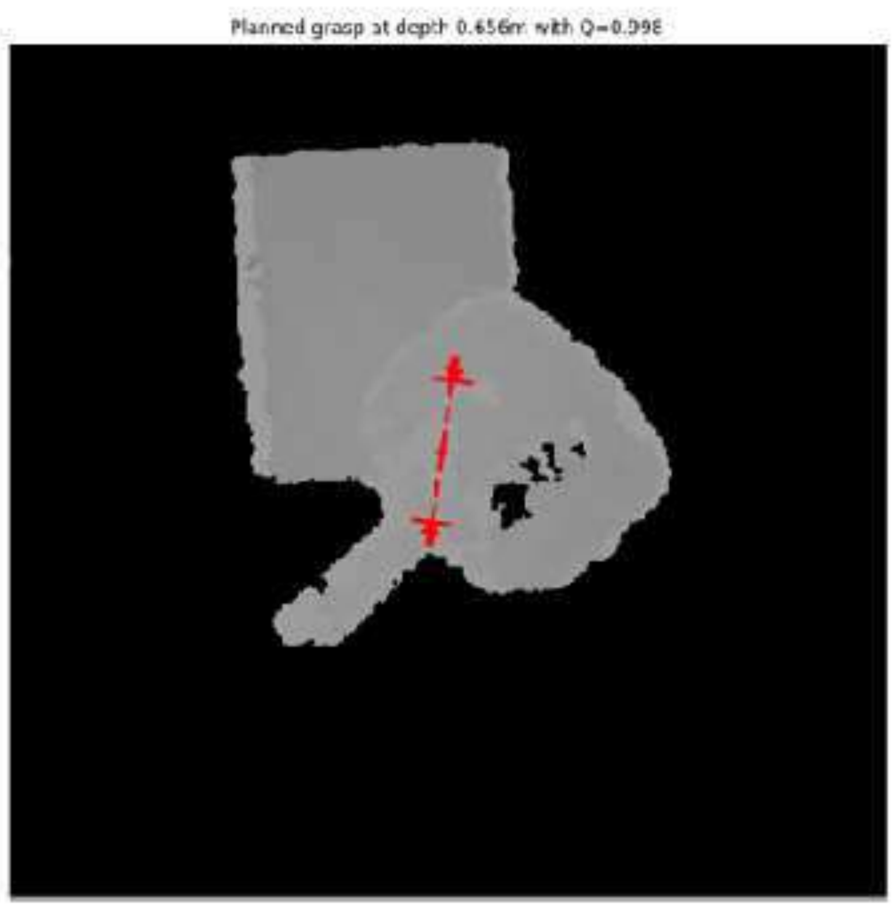


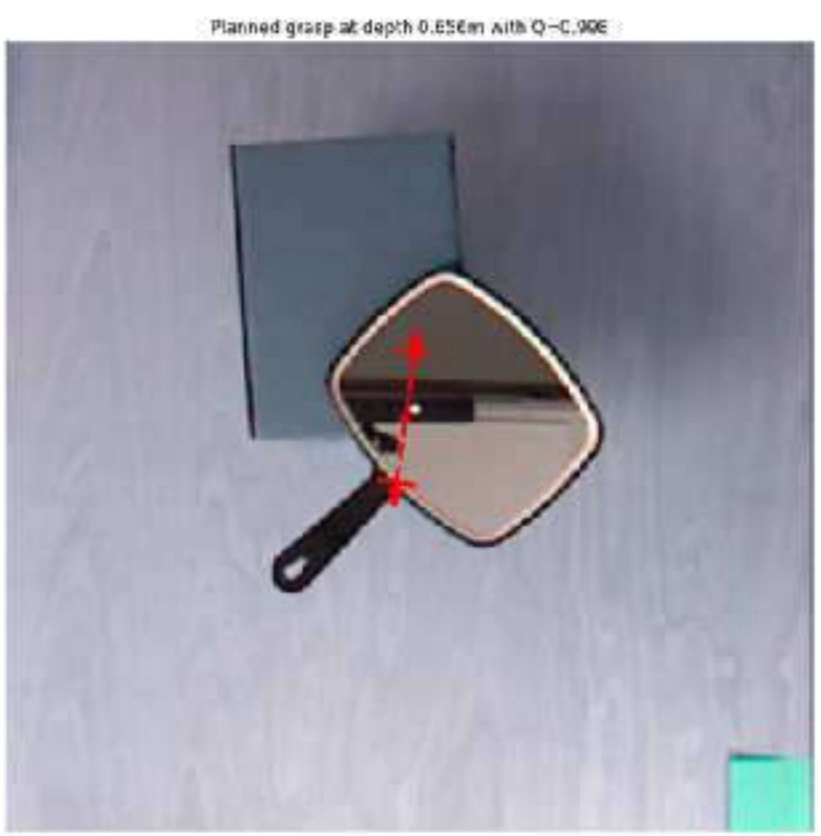
Why do we need human guidance

1) Because some objects challenge our cameras

➡Mirror

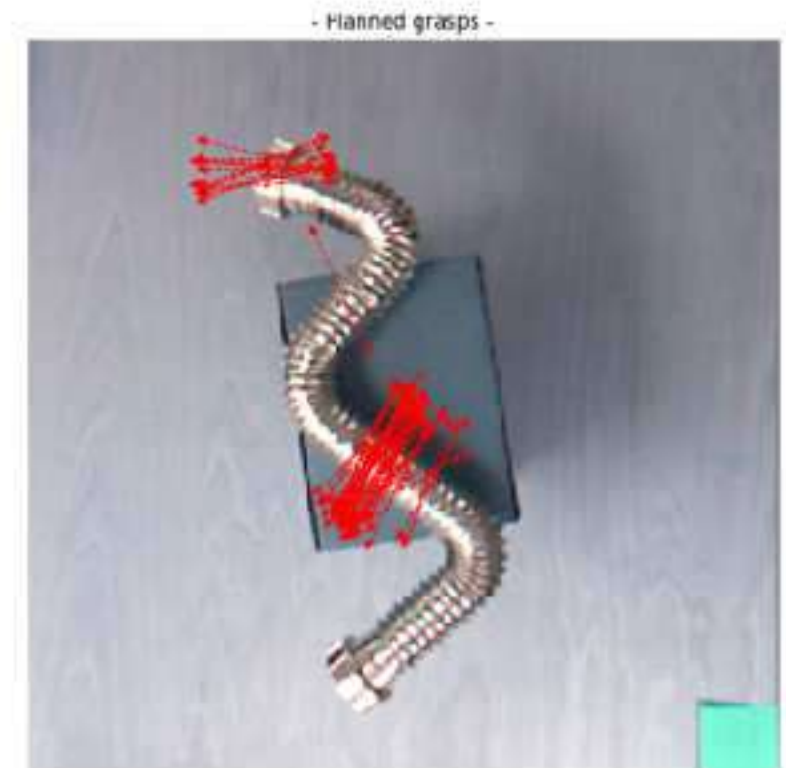




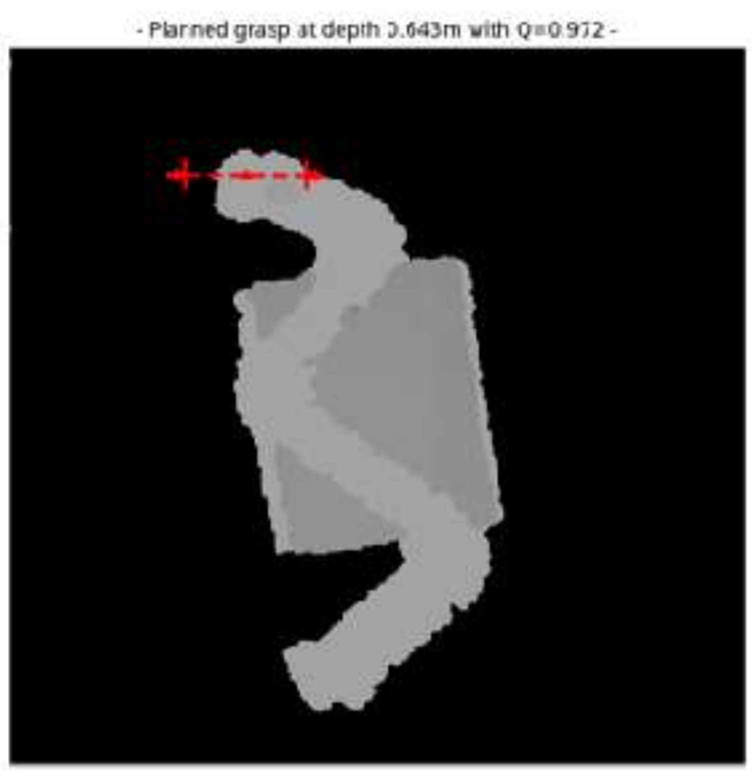


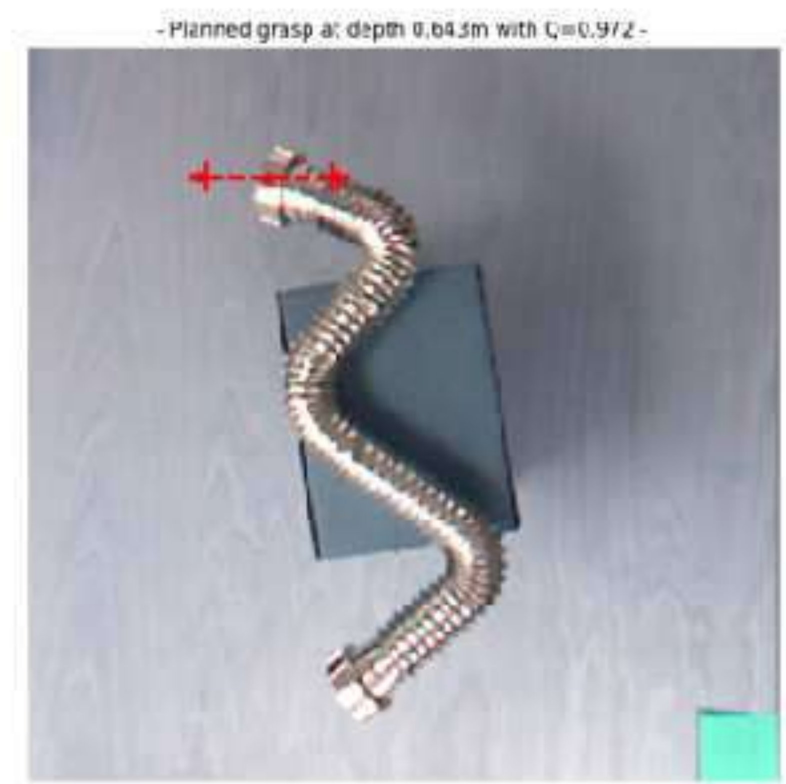
Why do we need human guidance

2) Because the best grasp candidate automatically computed by a grasping algorithm (here: Dexnet 2.0) is not the best choice according to the human, it is not what the human would do



➡Pipe

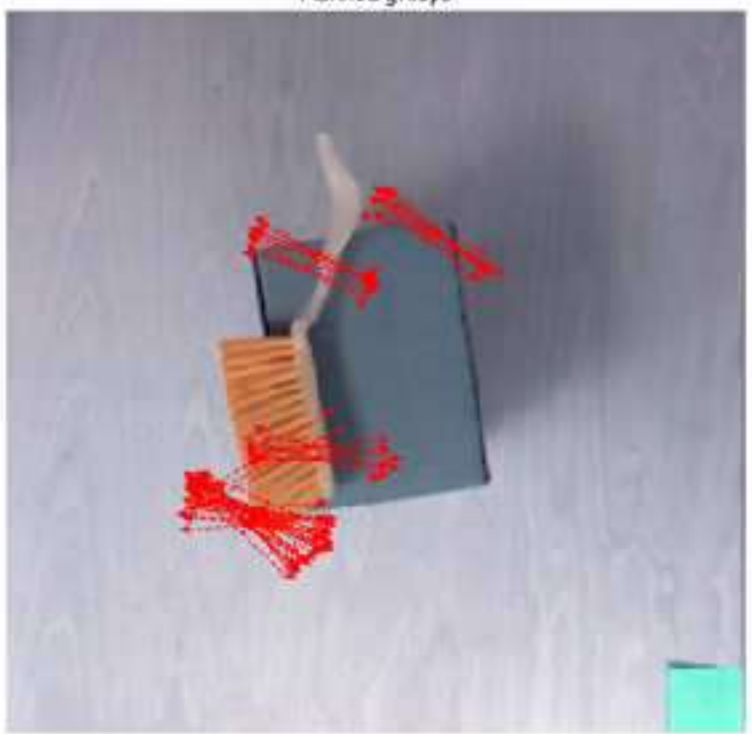




Why do we need human guidance

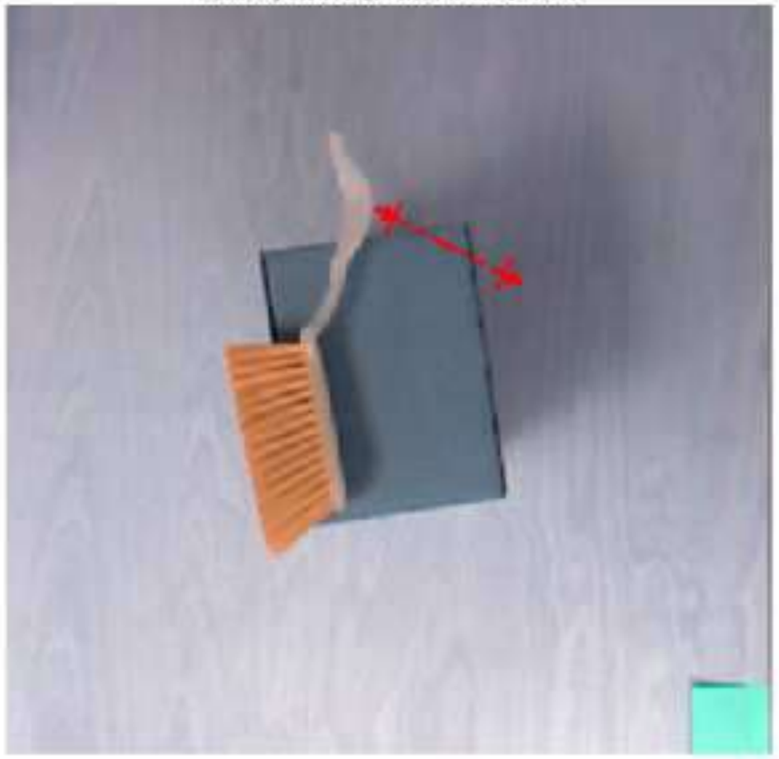
2) Because the best grasp candidate automatically computed by a grasping algorithm (here: Dexnet 2.0) is not the best choice according to the human, it is not what the human would do

- Planned grasps -

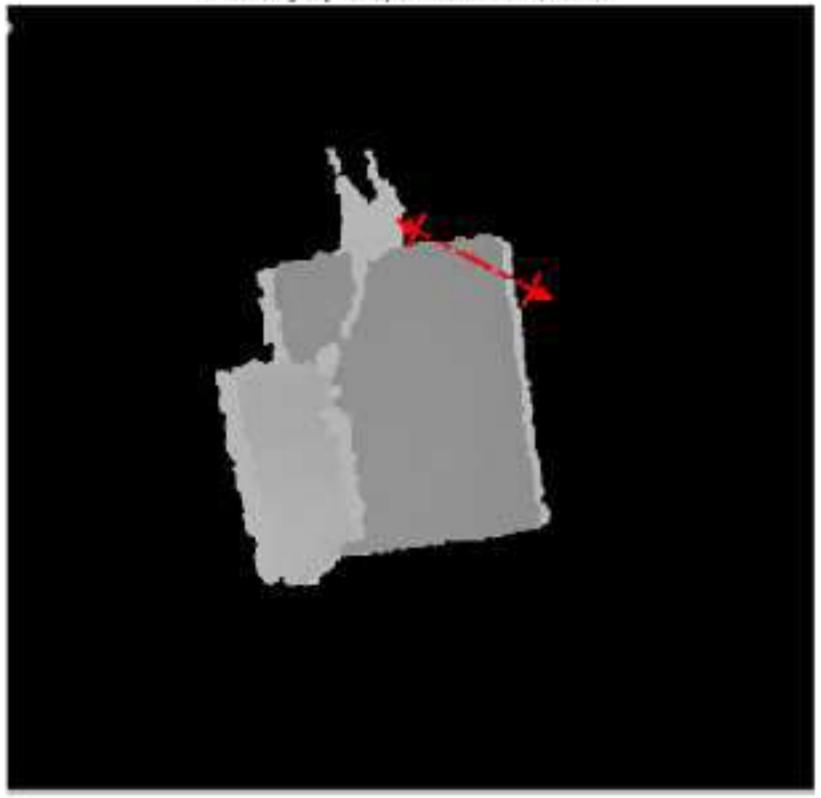


➡Broom

Planned grasp at depth 0.659m with Q=0.478



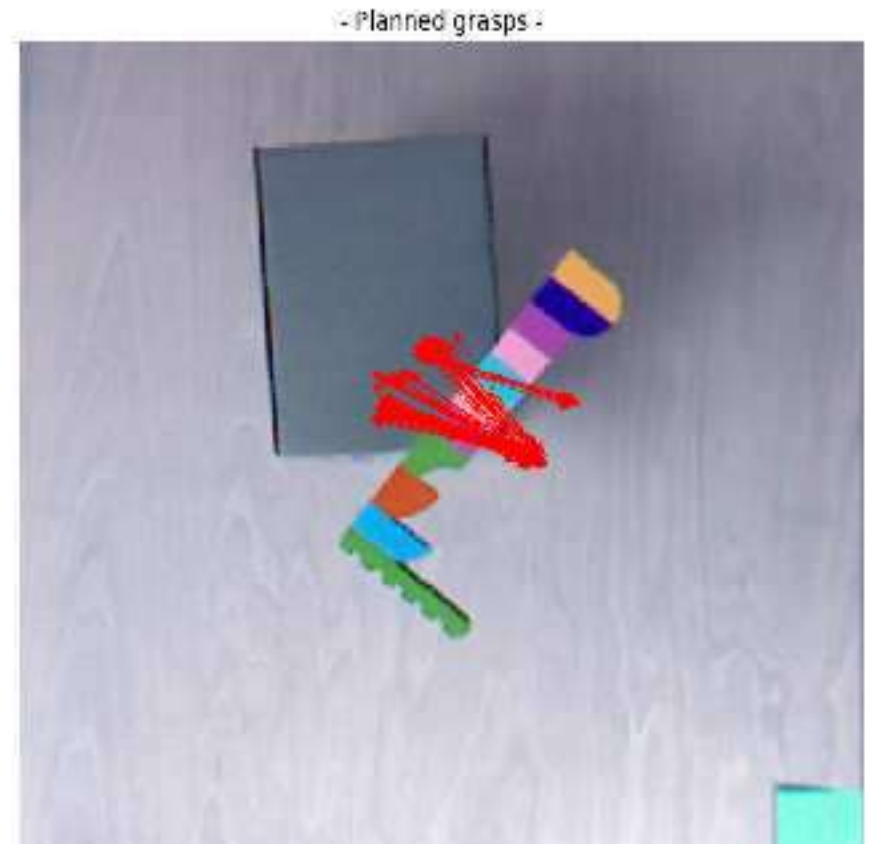
Planned grasp at depth 0.655m with Q=0.178

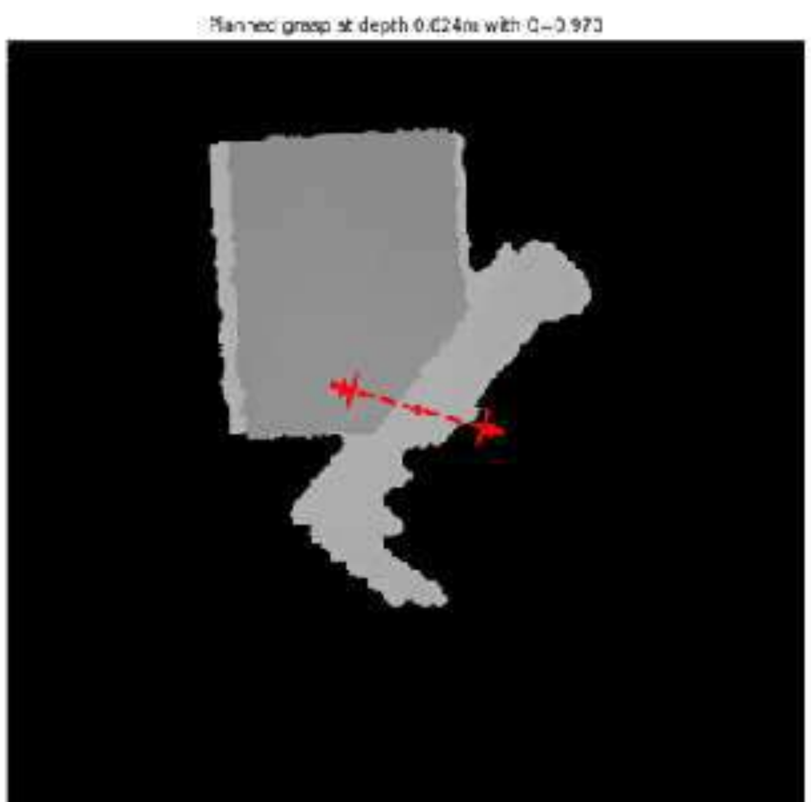


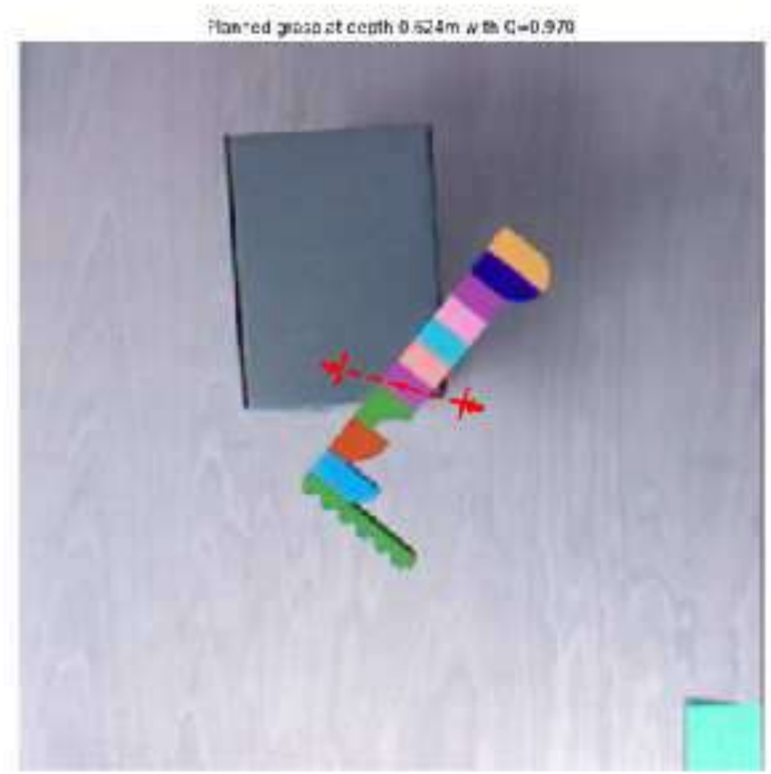
Why do we need human guidance

3) Because grasping algorithms do not reason about the objects fragility

➡ the Duplo tower will break during grasping and transportation







HEAP open dataset of objects

- Our first objective is to create (yet another?) <u>open</u> dataset of objects for benchmarking grasping algorithms
  - Inspired by waste sorting application

- Our dataset will contain a variety of different objects:
  - Irregular, deformable, articulated, different compliance
  - Challenging for vision and manipulation
  - Can break when grasped

  - Our dataset must be:
    - Reproducible
    - Easy to purchase / distribute
    - Cheap!

Andries et al. , HEAP object dataset

Existing object datasets for robotic manipulation

Object datasets: SIXD-challenge

[Hodan et. al ECCV18]

Existing object datasets for robotic manipulation

[Sun et. al Sensors18]

Birmingham nuclear waste dataset:
217 objects of 10 categories of objects

Existing object datasets for robotic manipulation

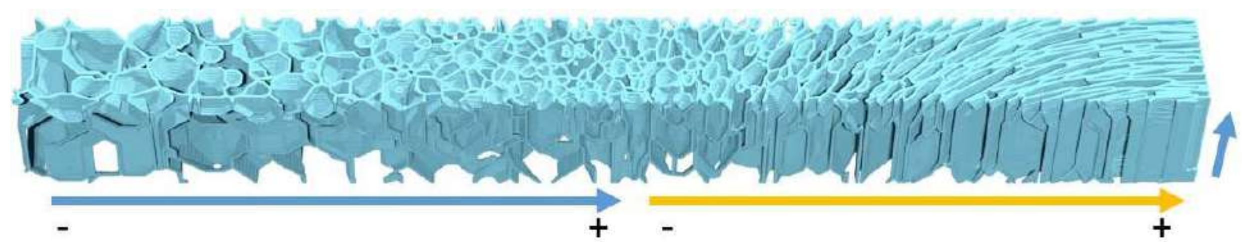
		Obj. variability	RGB-D scans	Simulation models	Physical models	Reproducible	Distributable
(Choi et al., 2009)	ALS dataset	✓	✗	✗	✓	✗	✗
(Lai et al., 2011)	Washington RGB-D dataset	✓	✓	✗	✗		
(Jiang et al., 2011)	Cornell grasping dataset	✓	✓	✗	✗		
(Kasper et al., 2012)	KIT object set	✗	✓	✓	✓	✗	✗
(Calli et al., 2015)	YCB dataset	✓	✓	✓	✓	✗	✓
(Correll et al., 2016)	Amazon picking challenge	✓	✗	✗	✓	✗	✓
(Sun et al., 2018)	Birmingham nuclear waste dataset	✗	✓	✗	✗		
(Matamoros et al., 2019)	Robocup @home	✓	✗	✗	✓	✗	✓

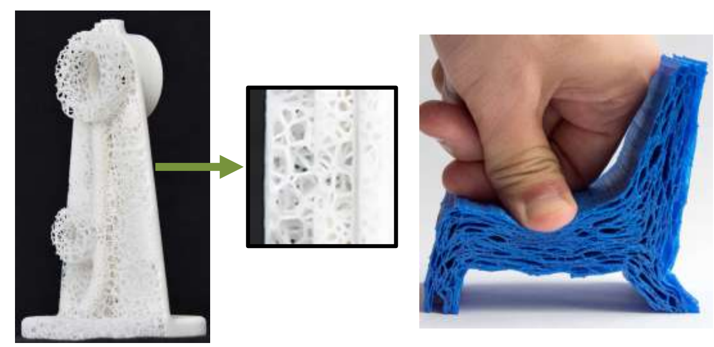
3D printable objects

Andries et al. , HEAP object dataset

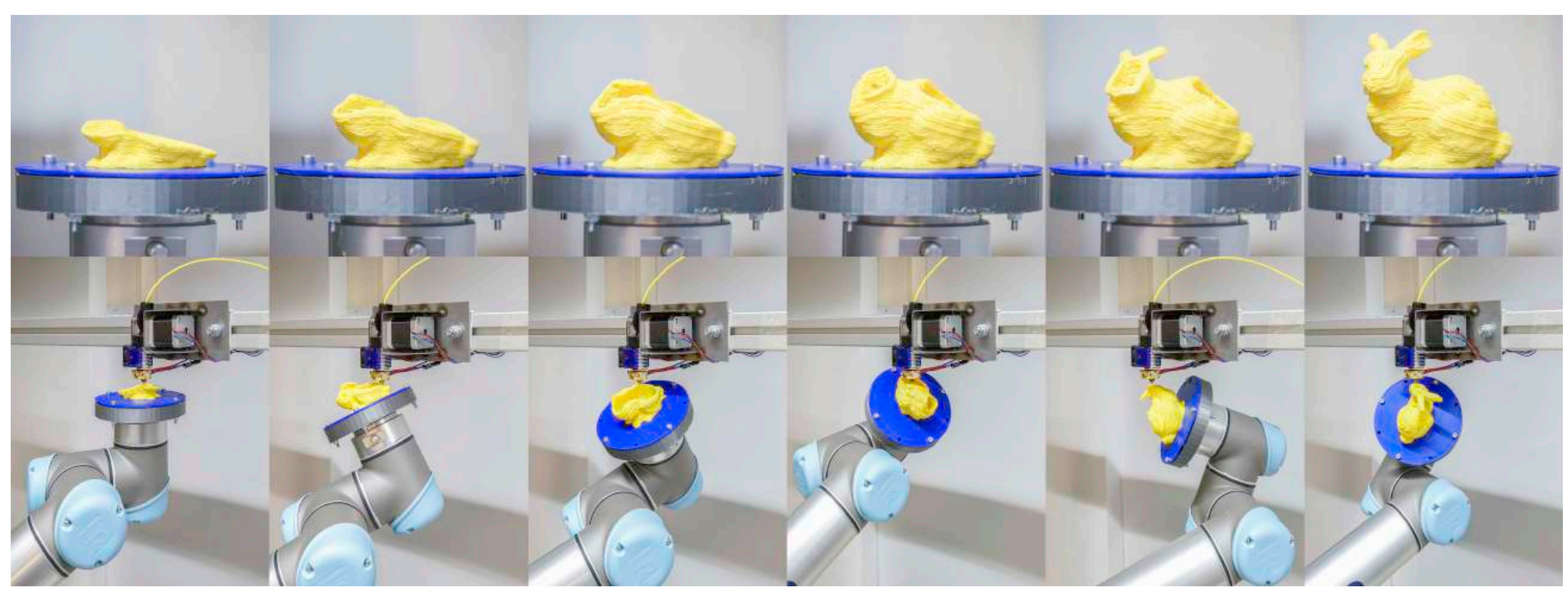
Generating 3D printable objects with specific properties

Polyedral Voronoi foams (ACM SIGGRAPH 2018)
- printable with low-cost filament printers
- closed-foam structure





Curved deposition (ACM SIGGRAPH 2018)



With S. Lefebvre

Generating 3D printable objects with specific properties

Do you have specific needs for your grasping and manipulation research?

We may include your specifications to create the objects in our dataset!

Write to:

serena.ivaldi@inria.fr
mihai.andries@inria.fr

Acknowledgments

J.-P. Saut, A. Sahbani

N.Lyubova, D. Filliat,
S.M. Nguyen, A. Droniou,
O. Sigaud, P.-Y. Oudeyer

MACSi

E. Zibetti, V. Padois,
M. Chetouani

CoDyCo

EDHHI

L. Penco, J.B. Mouret, V.
Modugno, E. Mingo

AnDy

HEAP

M. Andries, L.
Vianello, Y. Fleytoux

Thank you!   Questions ?

Thank you!   Questions ?

Do you have specific needs for your grasping and manipulation research?

We may include your specifications to create the objects in our dataset!

Write to:

serena.ivaldi@inria.fr
mihai.andries@inria.fr