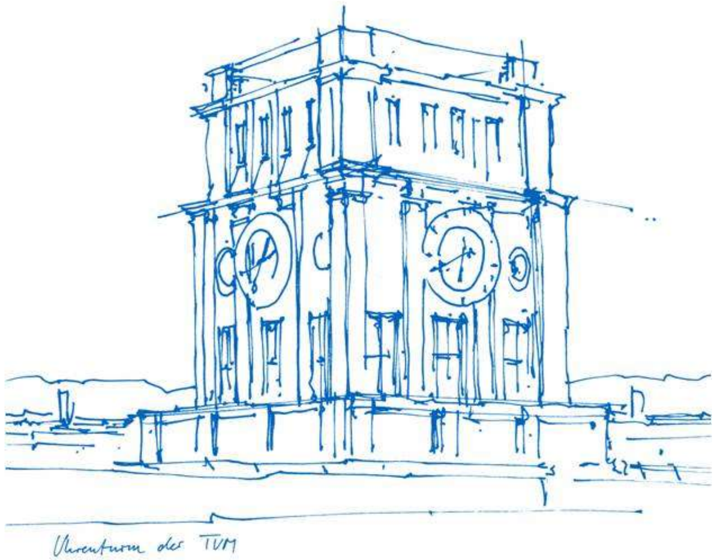
Efficient Motor Skill Learning in Robotics

Dongheui Lee

Department of Electrical and Computer Engineering
Technical University of Munich (TUM)

Institute of Robotics and Mechatronics
German Aerospace Center (DLR)



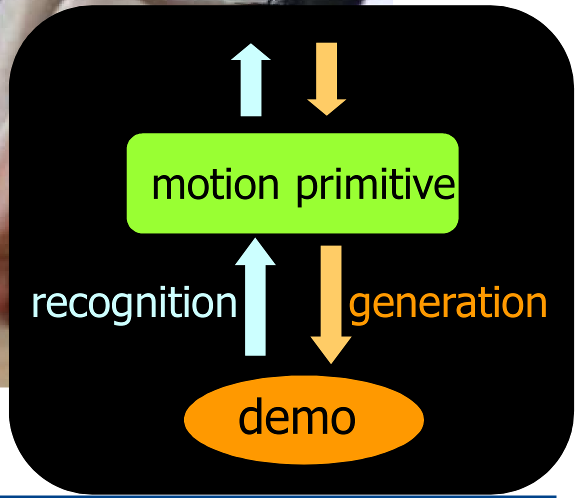
Reinforcement Learning School, April 8, 2021

Overview

- Learning from Demonstrations
- Reinforcement Learning in Robotics
- Human Robot Interaction Learning
- Complex Manipulation Task Learning

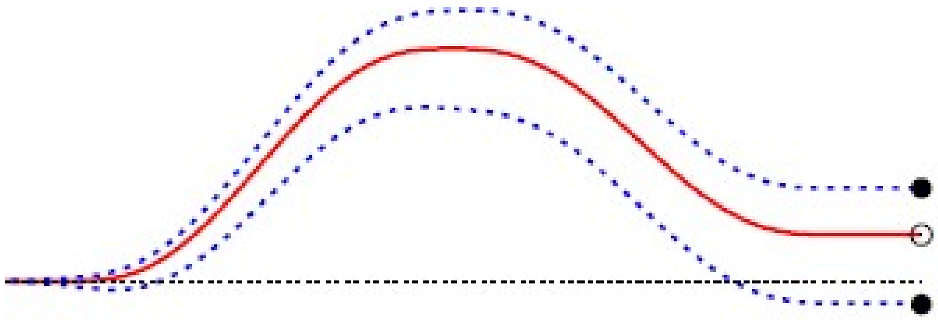
Imitation Learning

- Developmental Learning
- Neuroscience
- Optimal Control
- Psychology



Imitation Learning in Robotics: Generation vs. Generalization

Reaching to a different goal

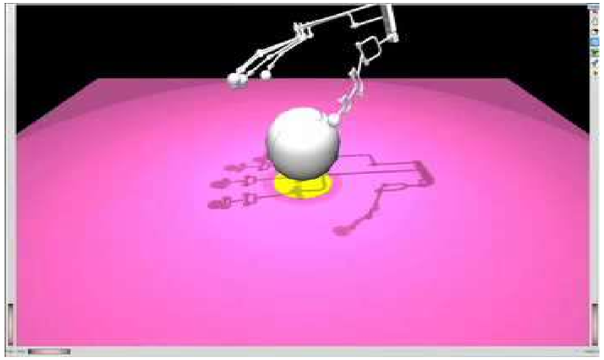


[Schaal et al]

A different intermediate goal

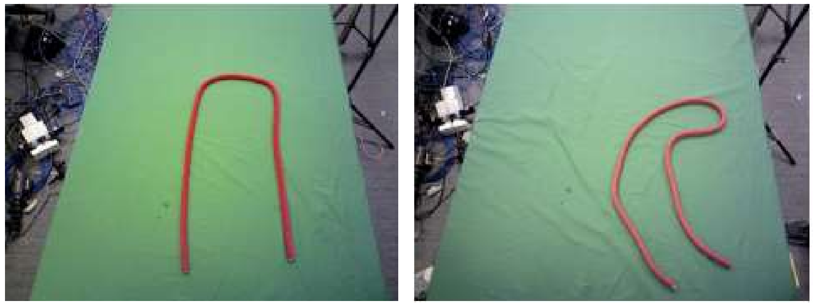
[Pervez, Lee, 2017]

Grasping a different size ball



[Schmidts, Peer & Lee]

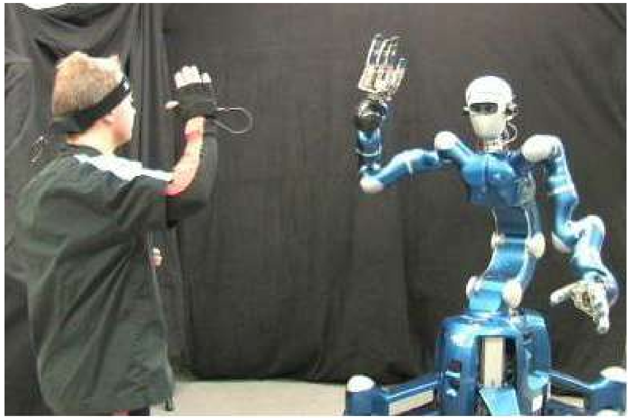
Knot Tying



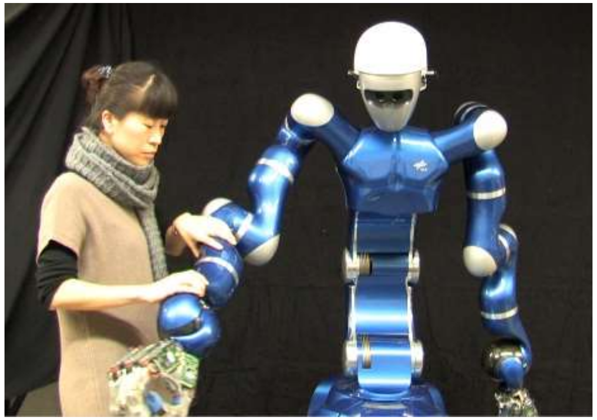
[Abbeel et al]

Learning from Demonstrations: Teaching modalities

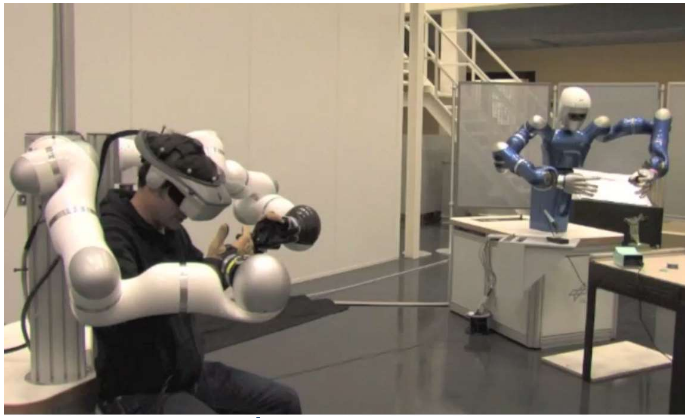
Motion Imitation



Kinesthetic teaching



Teleoperation



Intuitive

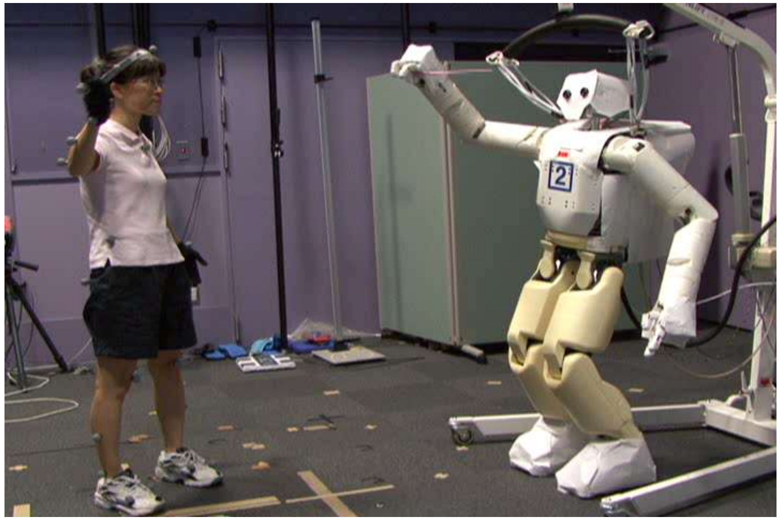
Exteroceptive

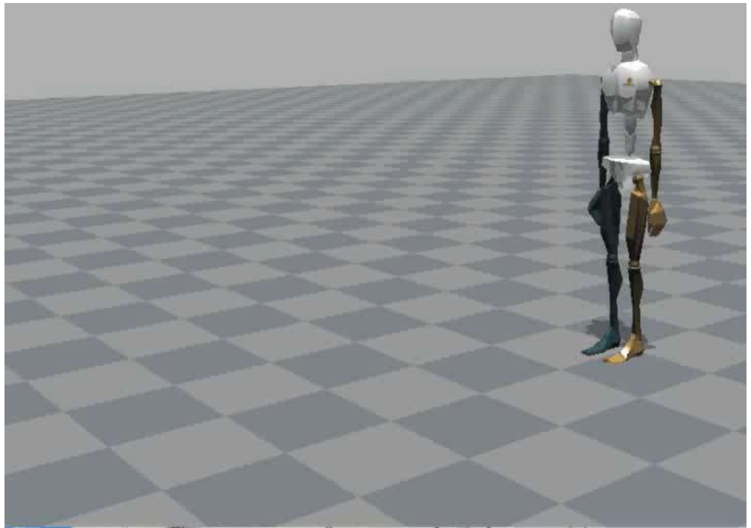
High burden

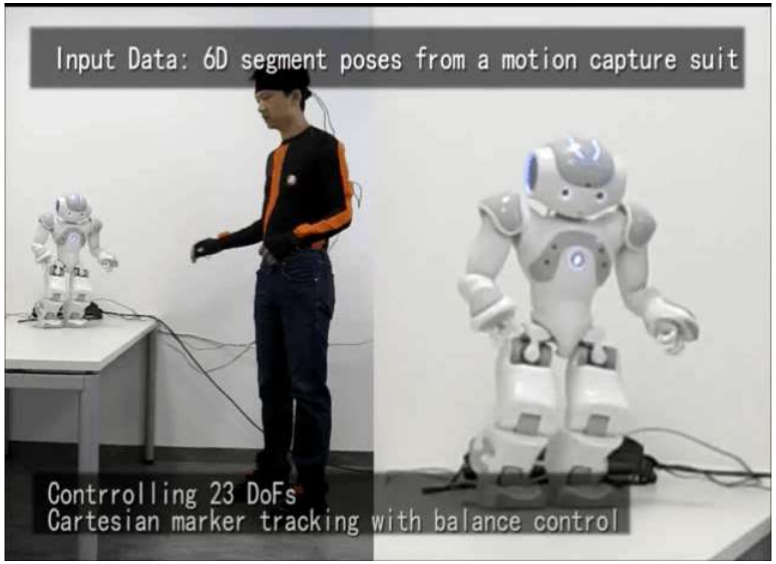
Proprioceptive

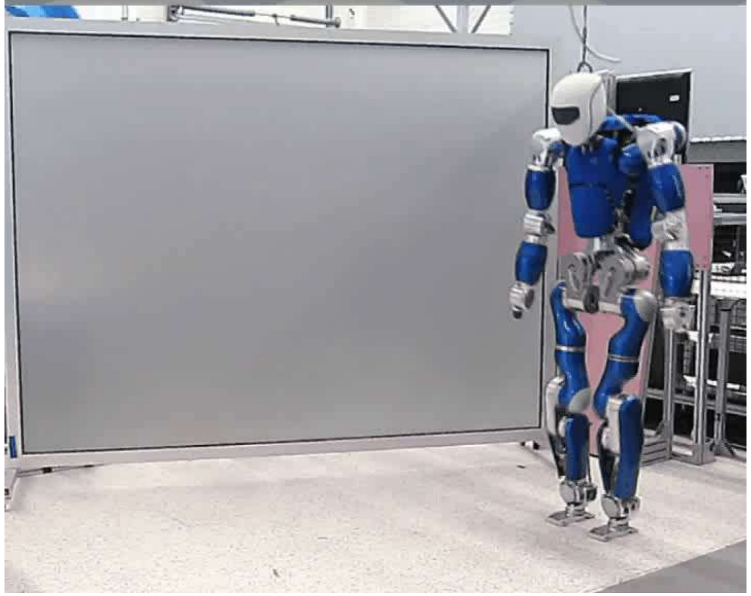
[Calinon, Lee, 2017]

Human Motion Imitation by Humanoids



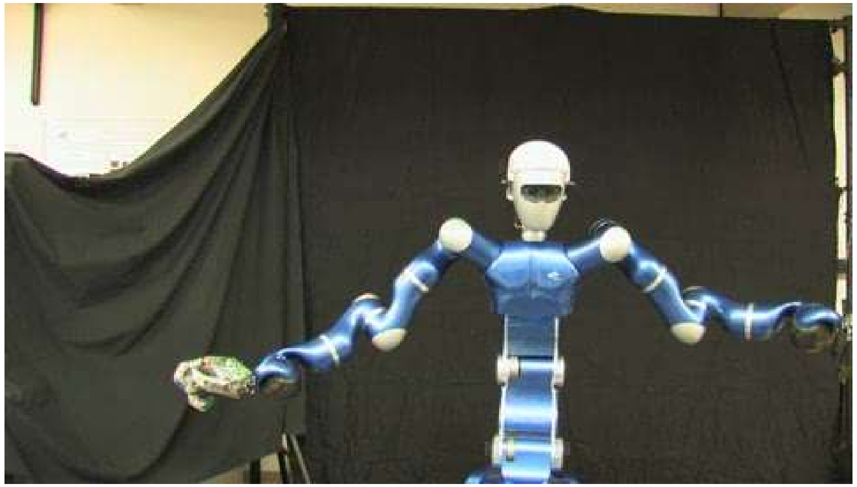






[Humanoids 2008, SYROCO2012, AT 2012, ICRA2014]

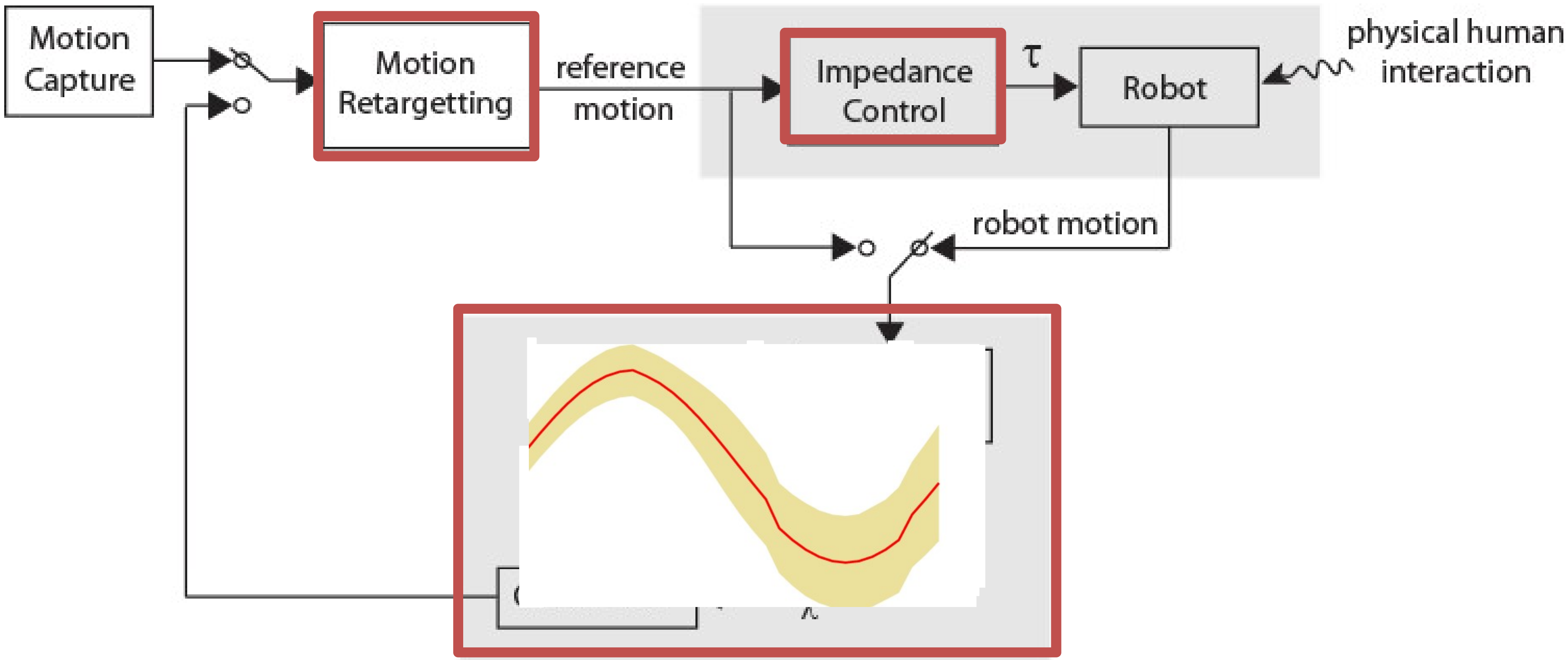
Teaching Pulp Fiction Dance



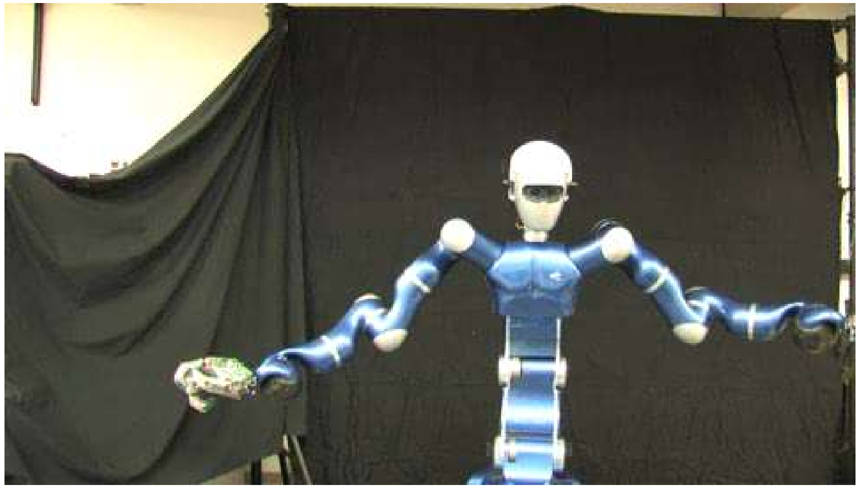
Learning from human motion retargeting

Refine a skill by kinesthetic teaching

$$\tau = g(q) + M(q)\ddot{q}_d + C(q,\dot{q})\dot{q}_d - D\dot{\tilde{q}} - s(\tilde{q})$$

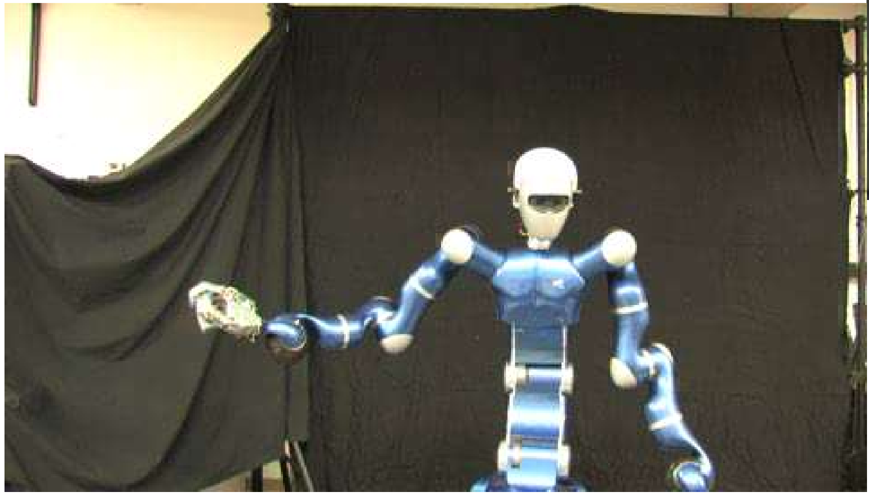


Incremental Learning Steps



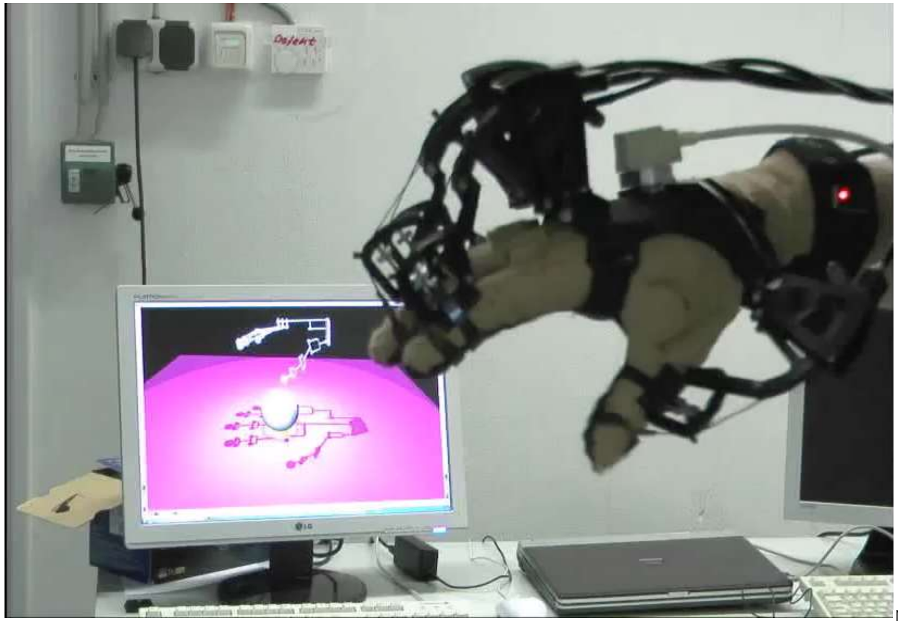
Imitation learning

Kinesthetic Coaching

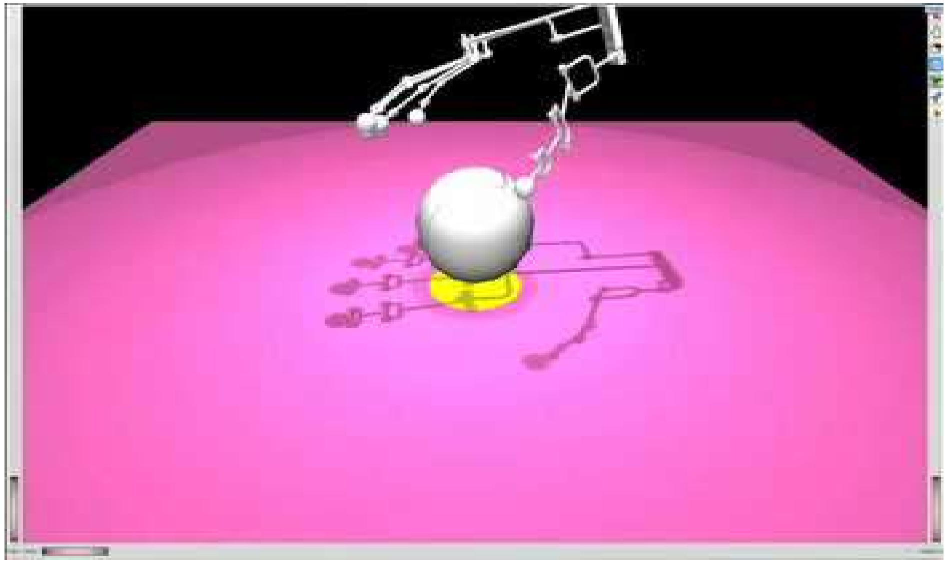


[Autonomous Robots 2011, IROS 2010]

Grasping Skill Learning from Motion & Force Data



Teleoperation using Cyberglove,
Flock of Birds, & Cybergrasp
(Haptic Feedback)



r[cm]	max(f^{in})[N]		$\overline{f}^{in}$[N]		ΔT[ms]	
3.6	3.21	-*	3.20	-*	28	-*
4.0	**3.21**	**5.41**	**3.20**	**5.10**	**11**	**209**
4.8	3.21	7.12	3.20	7.04	39	371
5.6	3.21	12.92	3.20	12.84	88	531
6.0	3.21	-*	3.20	-*	106	-*
Force control	ON	OFF	ON	OFF	ON	OFF

* unsuccessful grasping attempt

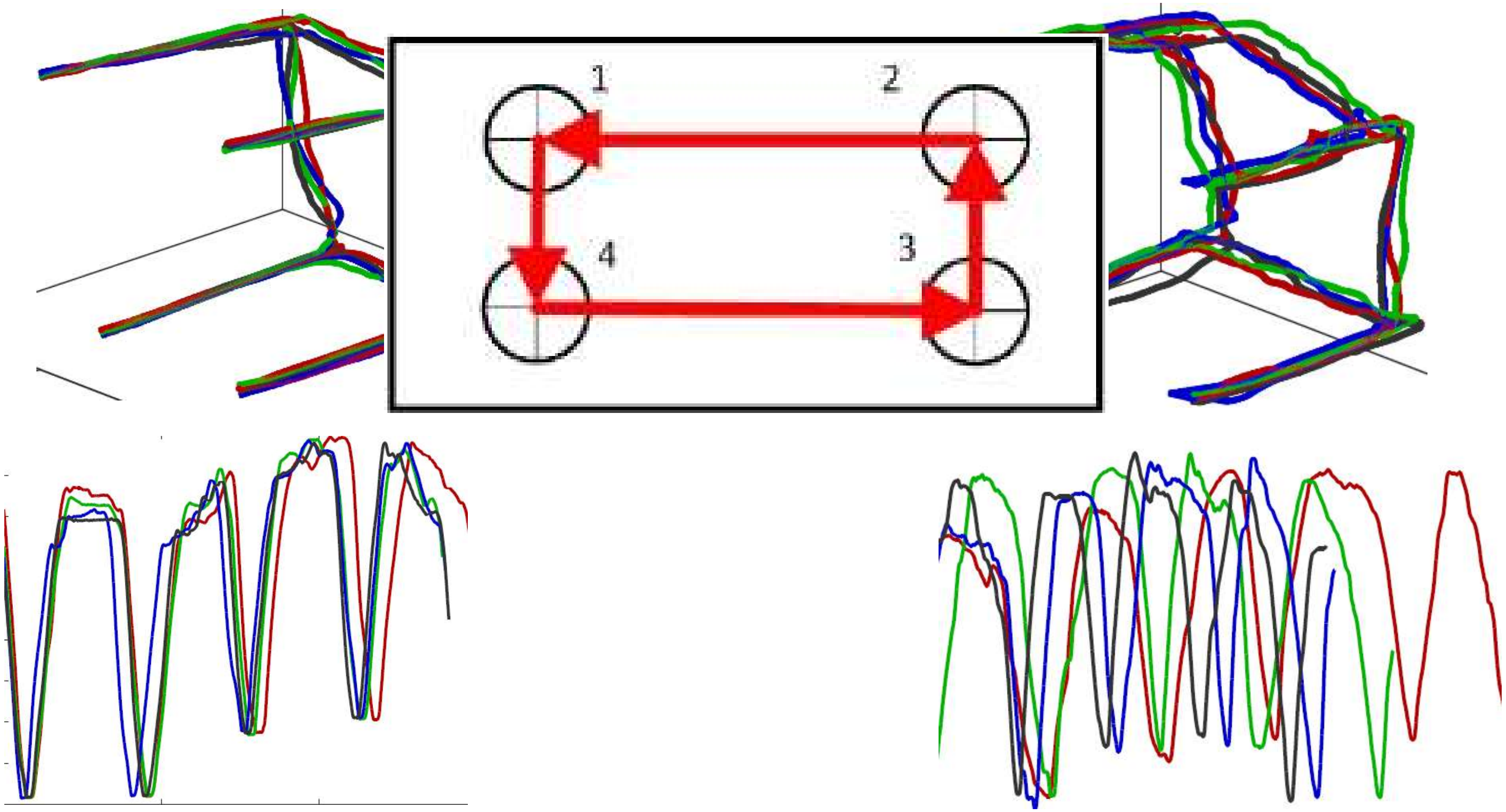
[Schmidts, Lee and Peer, IROS 2011]

What are Challenges in Teaching by Teleoperation?

What are Challenges in Teaching by Teleoperation?

Kinesthetic

Teleoperation



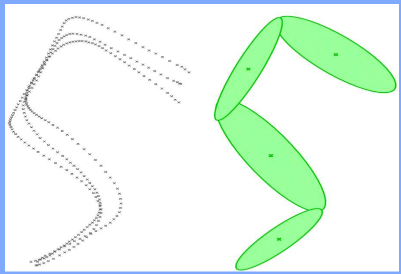
- High level of spatial-temporal variations.
- High cost for demonstration

Learning Repetitive Teleoperation Tasks with DMP/GMM

Canonical System $\dot{s} = \tau\omega$

DMP $\dot{v} = \tau\alpha_x(\beta_x(g-x)-v)+\tau a\mathcal{F}(s)$

GMM Encoding



$$\begin{bmatrix} \mathcal{F}_1(s_0) & x_{1,0} & \boxed{\begin{matrix} s_0 \\ \vdots \\ s_n \end{matrix}} \\ \vdots & \vdots & \\ \mathcal{F}_1(s_n) & x_{1,n} & \end{bmatrix}^\top$$

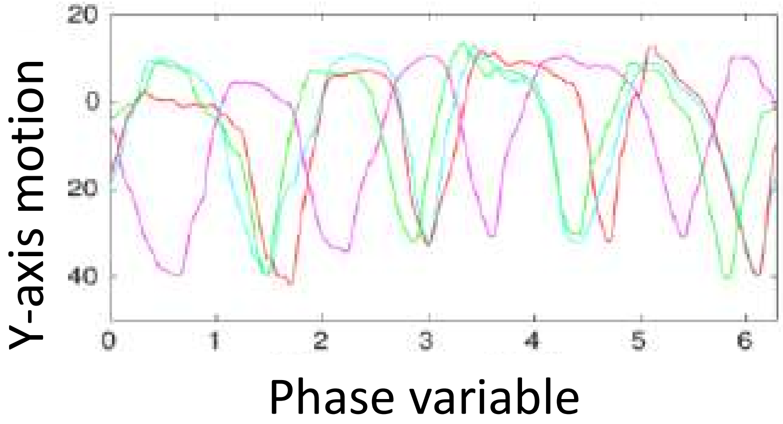
unknown

EM
GMM Update

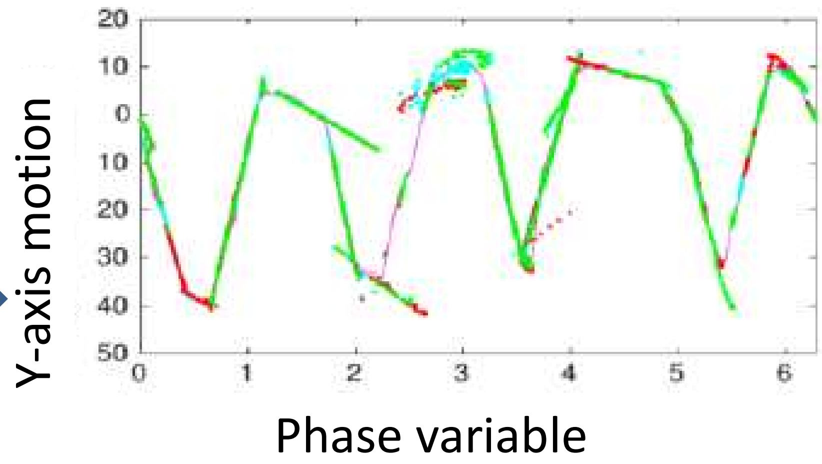
s Update

[WHC 2017]

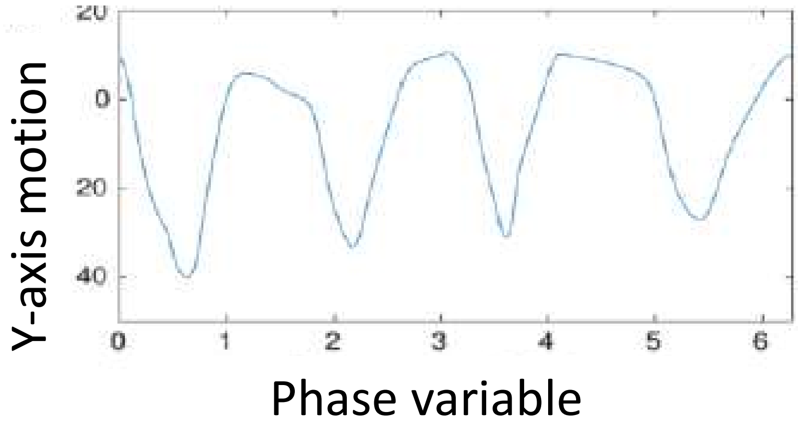
Results



Asynchronous trajectories



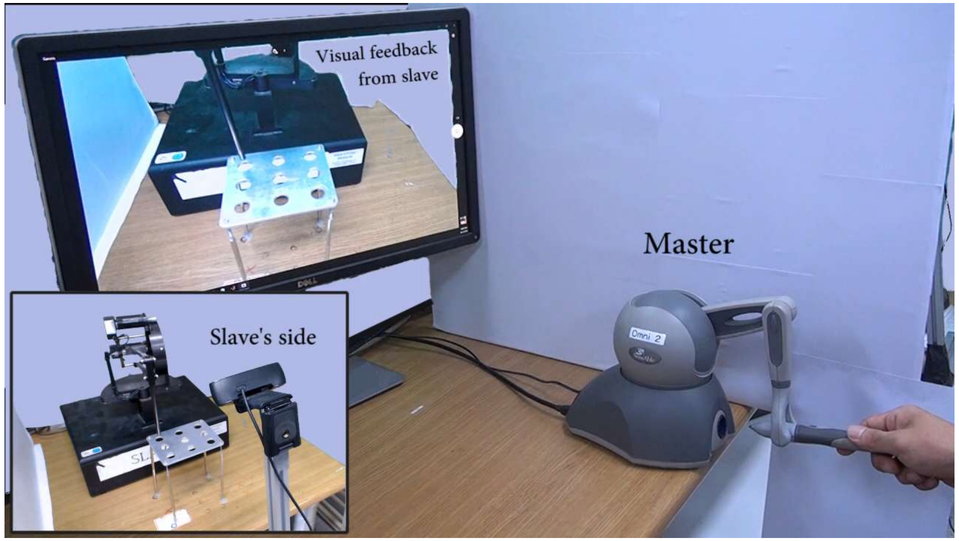
Synchronized data

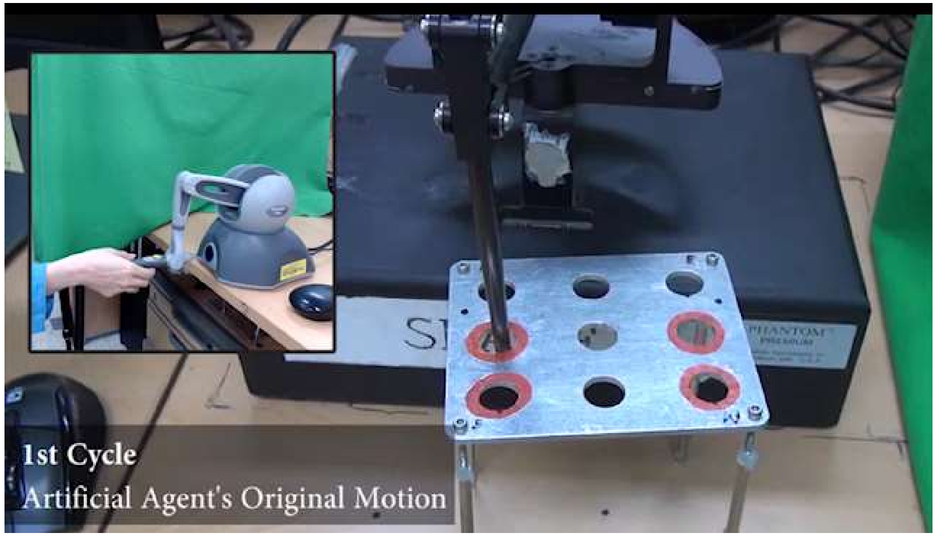


Reproduced trajectory

[WHC 2017]

Supervisor Teleoperation with Kinesthetic Coupling





Shared Control
- agent: horizontal motion
- human: vertical motion

Re-train the learned skill on the fly by dynamic authority and kinesthetic coupling

[AURO2019, ICRA 2020]

Overview

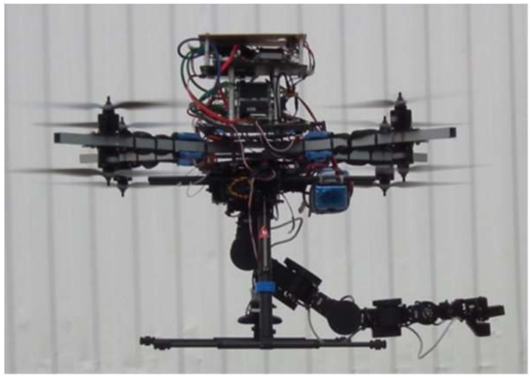
- Learning from Demonstrations
- Reinforcement Learning in Robotics
- Human Robot Interaction Learning
- Complex Manipulation Task Learning

Reinforcement Learning in Robotics

- Robots can learn how to execute a task by trial-and- error.
- Can learn complex and highly dynamic tasks
- Limited or no knowledge of robot/environment dynamics needed

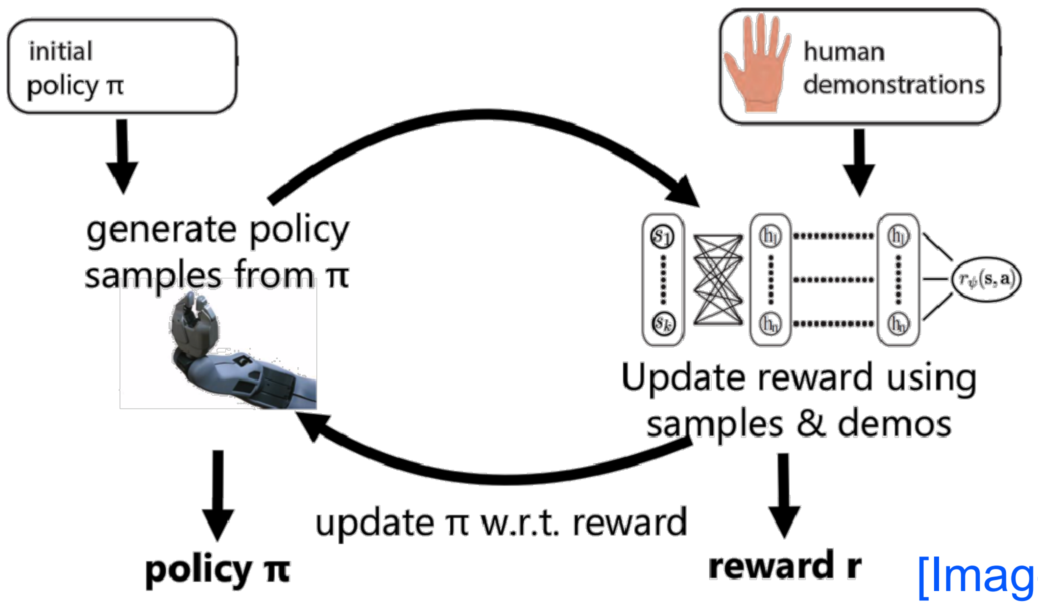
- Typical problems of RL in robotic domain:
  - Continuous and high dimensional state and action space
  - Many rollouts in real world → Time consuming, noisy measurement
  - Exploration with real robot: robot damages

[Kormushev+ 2010]

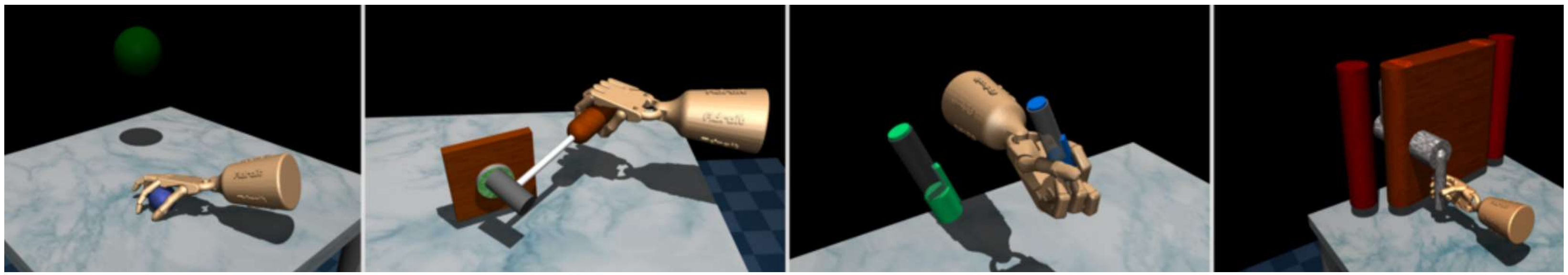


Imitation Learning combined with RL

Inverse Reinforcement Learning



[Image source: Finn]



[ICDL 2021, submitted]

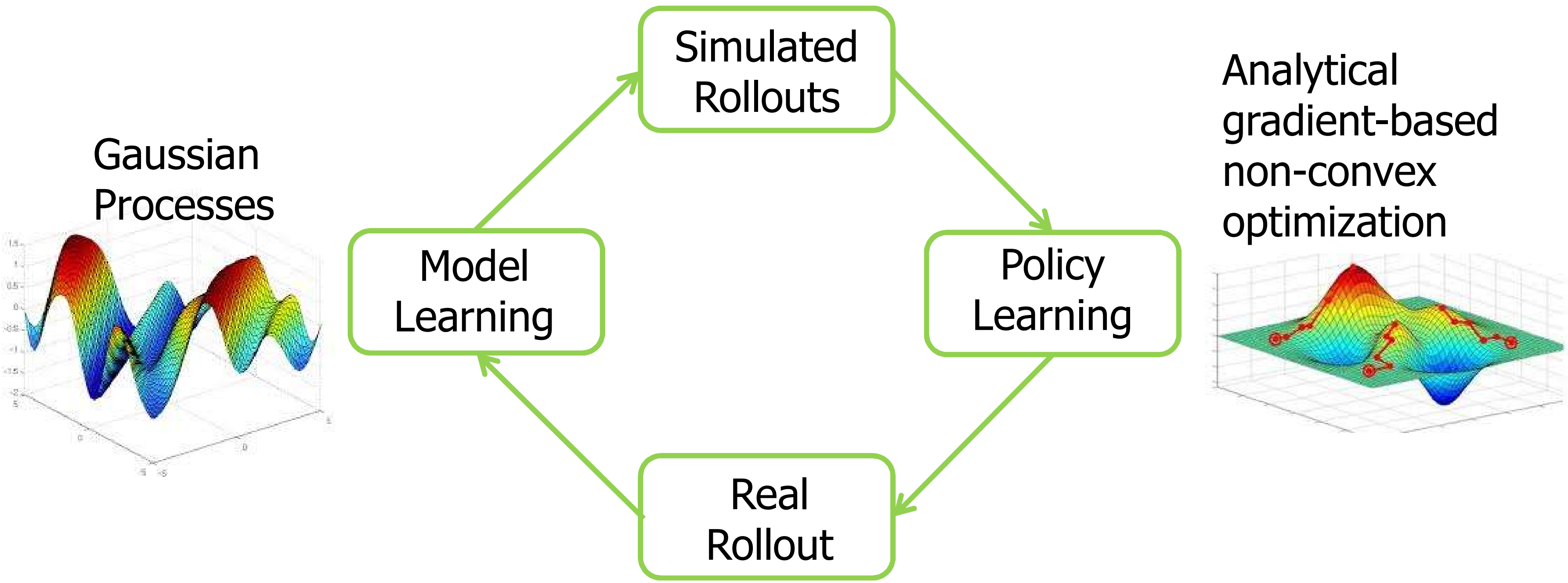
Imitation Learning combined with RL

PoWER (Policy Learning by Weighting Exploration with the Returns) [Kober+ 2009]

- Simple and computationally efficient update rule
- Learn with minimal prior knowledge
- **Policy initialized** with human demonstration

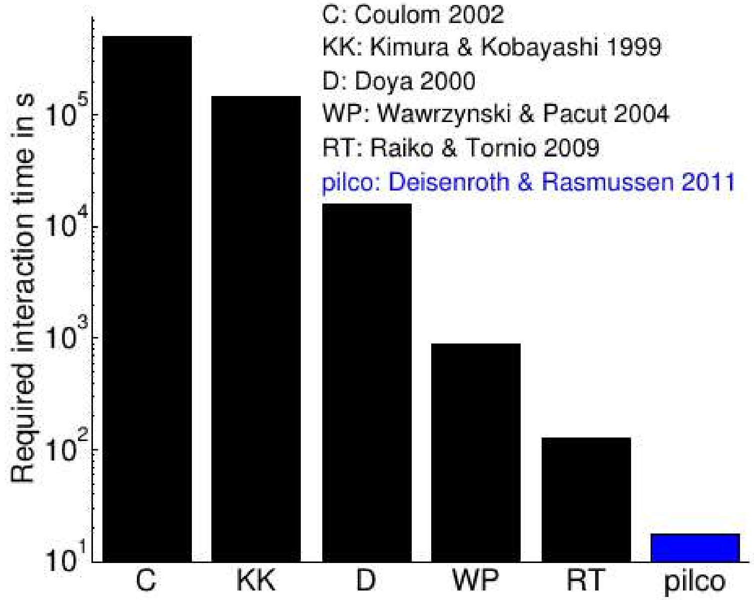
Probabilistic Inference for Learning COntrol (PILCO)

- Model-based policy search approach: Use data collected during the rollout to learn a model of the robot in a data-efficient way
- Find optimal policy on the learned model using simulation
  - Probabilistic long-term prediction to reduce model bias learning problem



[Deisenroth+ 2015]

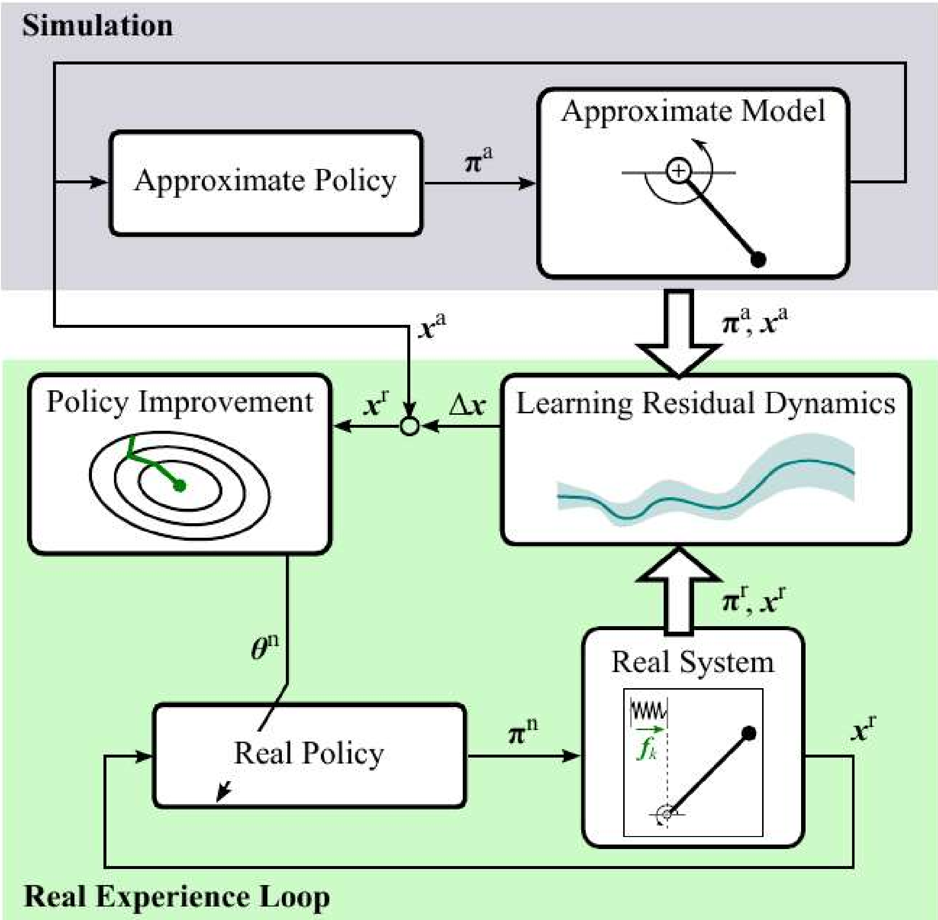
Benchmark: Cart-Pole Swing-up



- No knowledge about nonlinear dynamics
- Cost function $c(x) = -\exp(-\|x - x_{target}\|^2)$
- Fast learning speed compared to state of the art
- Learned dynamics models are only confident in areas of the state space previously observed

[Deisenroth+, PILCO, 2015]

Policy Improvement with REsidual Model learning (PI-REM)



Approximate Model
Known a priori and deterministic

Additive Gaussian noise
E.g., sensor noise

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}^a(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{f}^u(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{w}$$

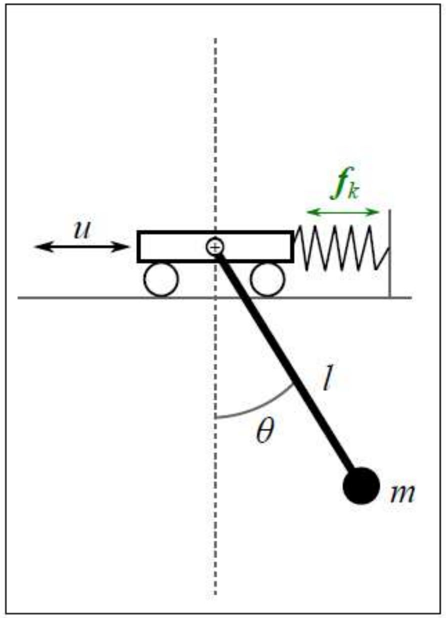
Unknown *residual* dynamics
E.g. joint friction, contact forces

$$\boldsymbol{f}^u(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{w} \sim \text{Gaussian Process}$$

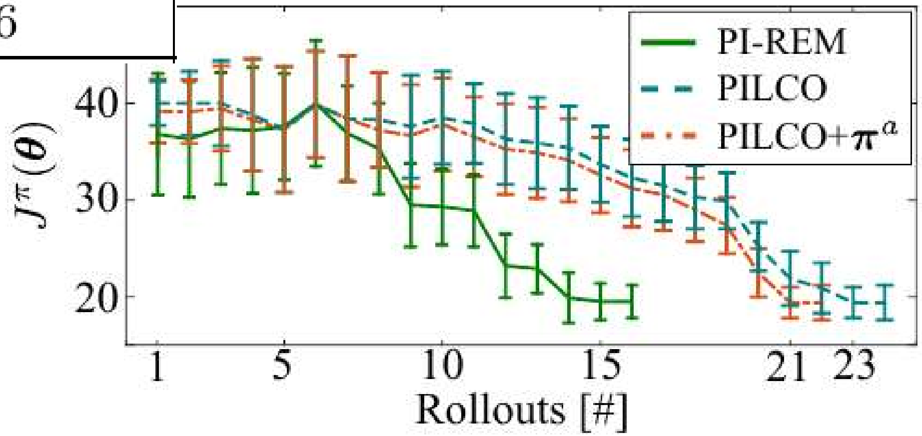
[Saveriano+ 2017] [Mouret+]

Cart-Pole Swing-up

- Approximate model : Cart-Pole without f_k
- State $\boldsymbol{x} = [p, \dot{p}, \theta, \dot{\theta}]^{\mathrm{T}}$
- Goal $\boldsymbol{x}_g = [0, 0, \pi, 0]^{\mathrm{T}}$



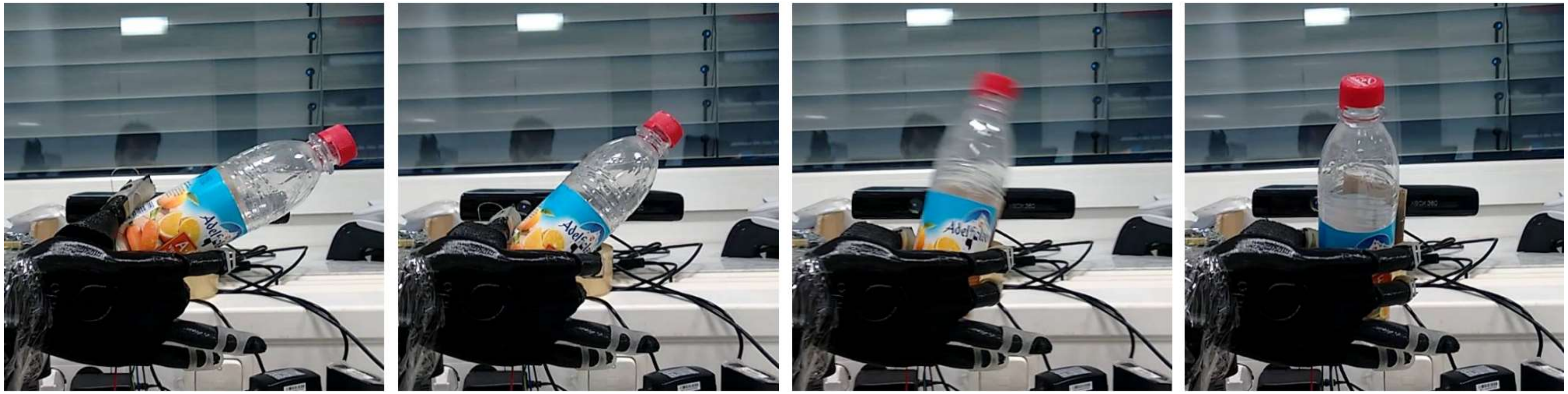
	Stiffness [N/m]	Real rollouts [#]	Real experience [s]
PI-REM	25	**2**	**8**
PILCO	25	5	20
PI-REM	50	**3**	**12**
PILCO	50	6	24
PI-REM	120	**15**	**30**
PILCO	120	23	46



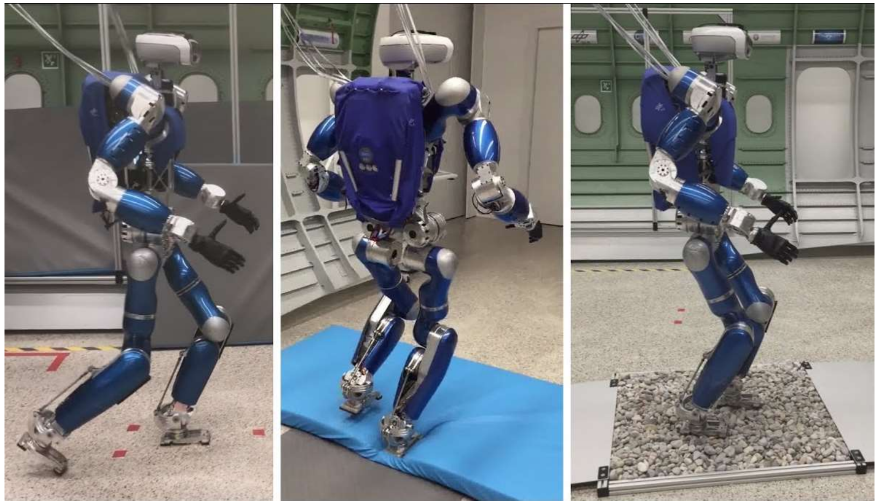
[Saveriano+ 2017]

Policy Learning Robust to Irreversible Events

- In-hand manipulation [RA-L 2018]

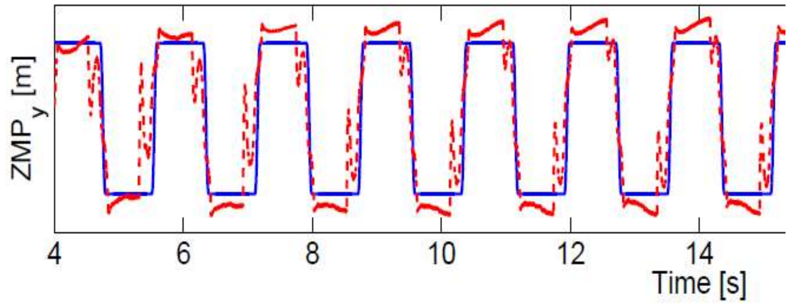


- Bipedal locomotion

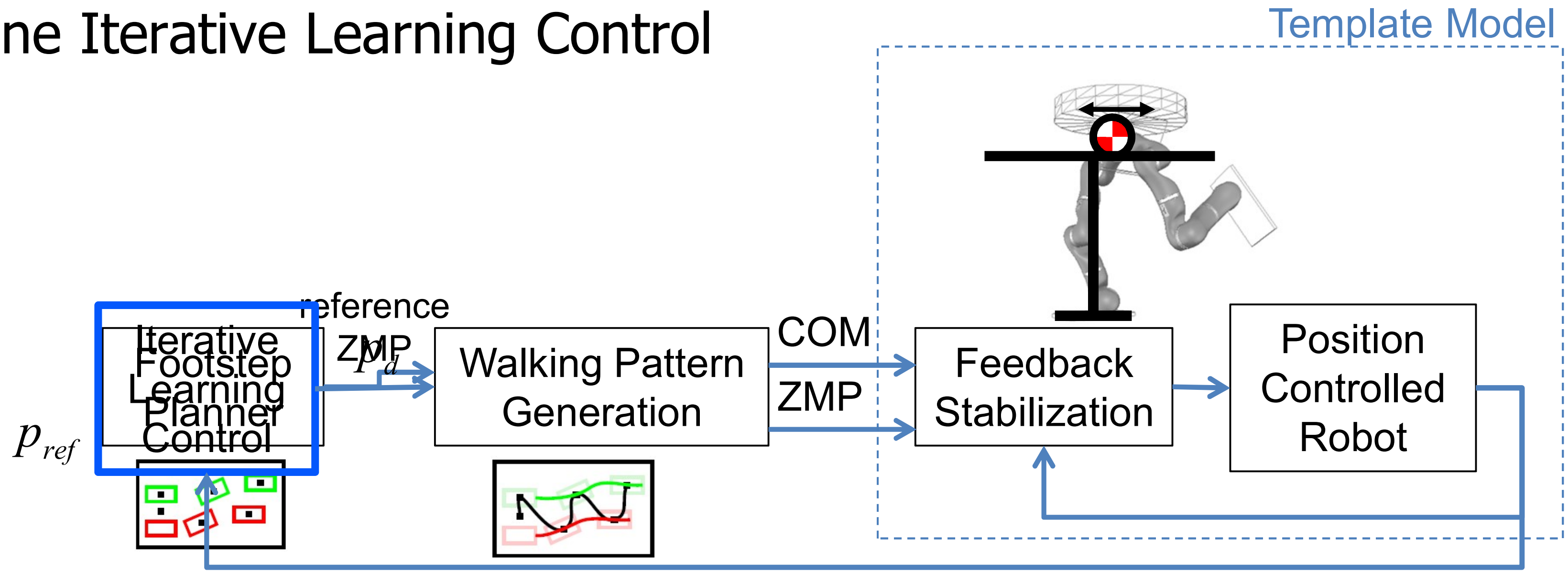


Bipedal Walking

- Conventional ZMP Based Walking
  - Feedback stabilization → tracking template model behavior
  - Dynamically consistent walking pattern generation
  - Existing ZMP tracking error

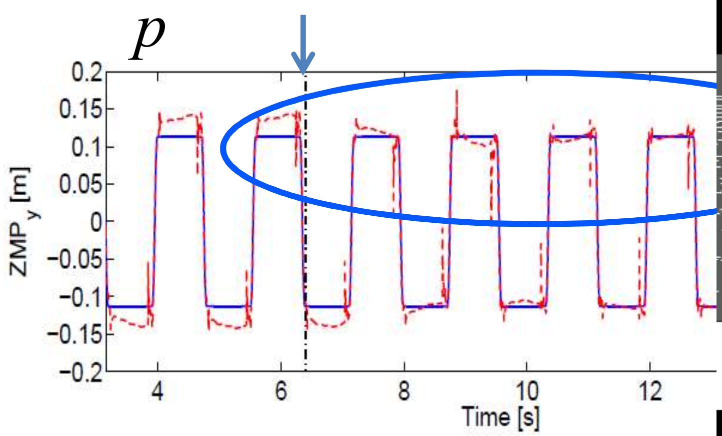


- Online Iterative Learning Control

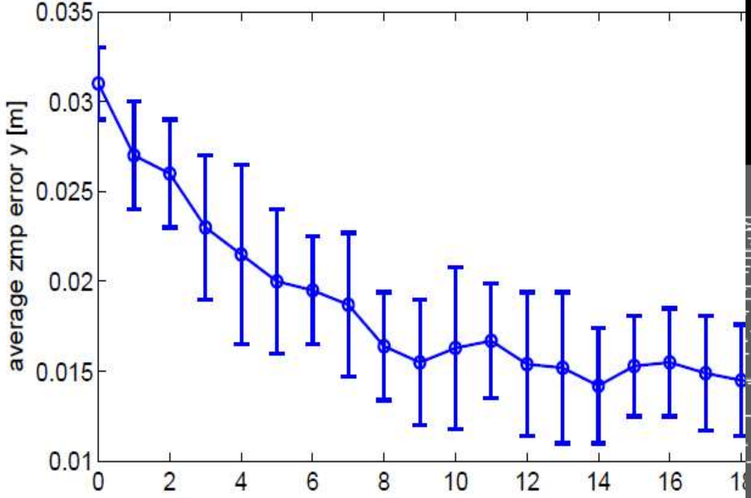


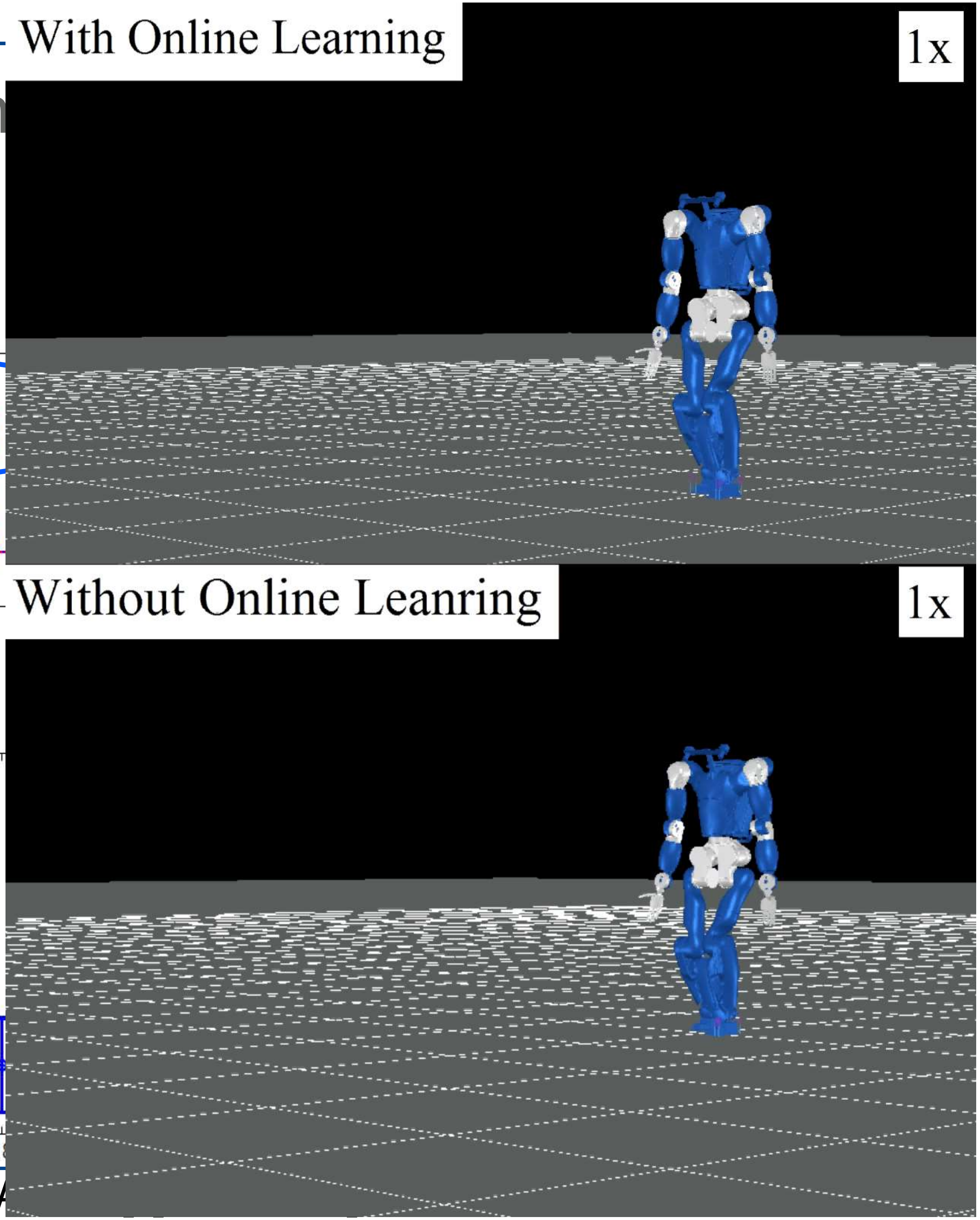
Simulation & Experim

Simulations



Experiments

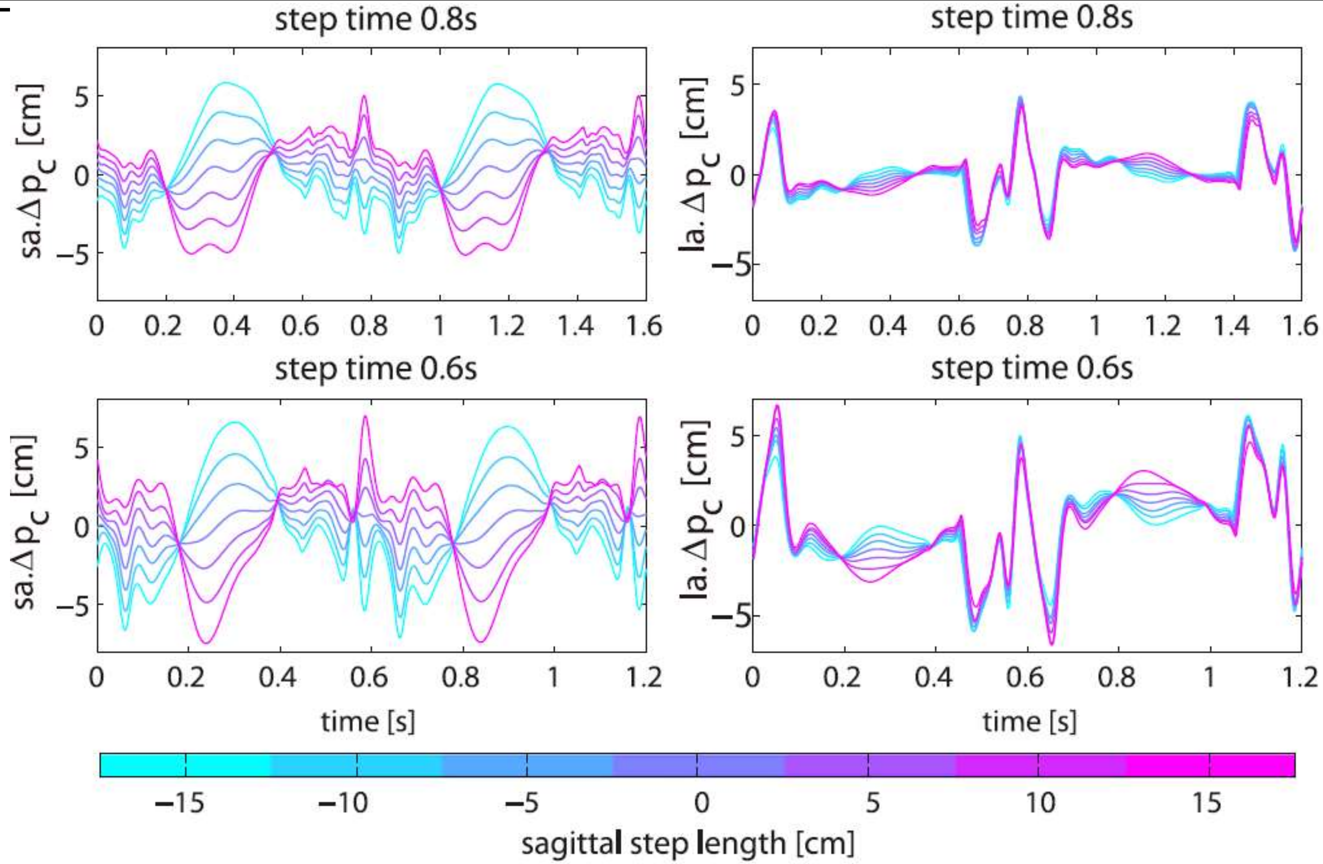




[ICRA

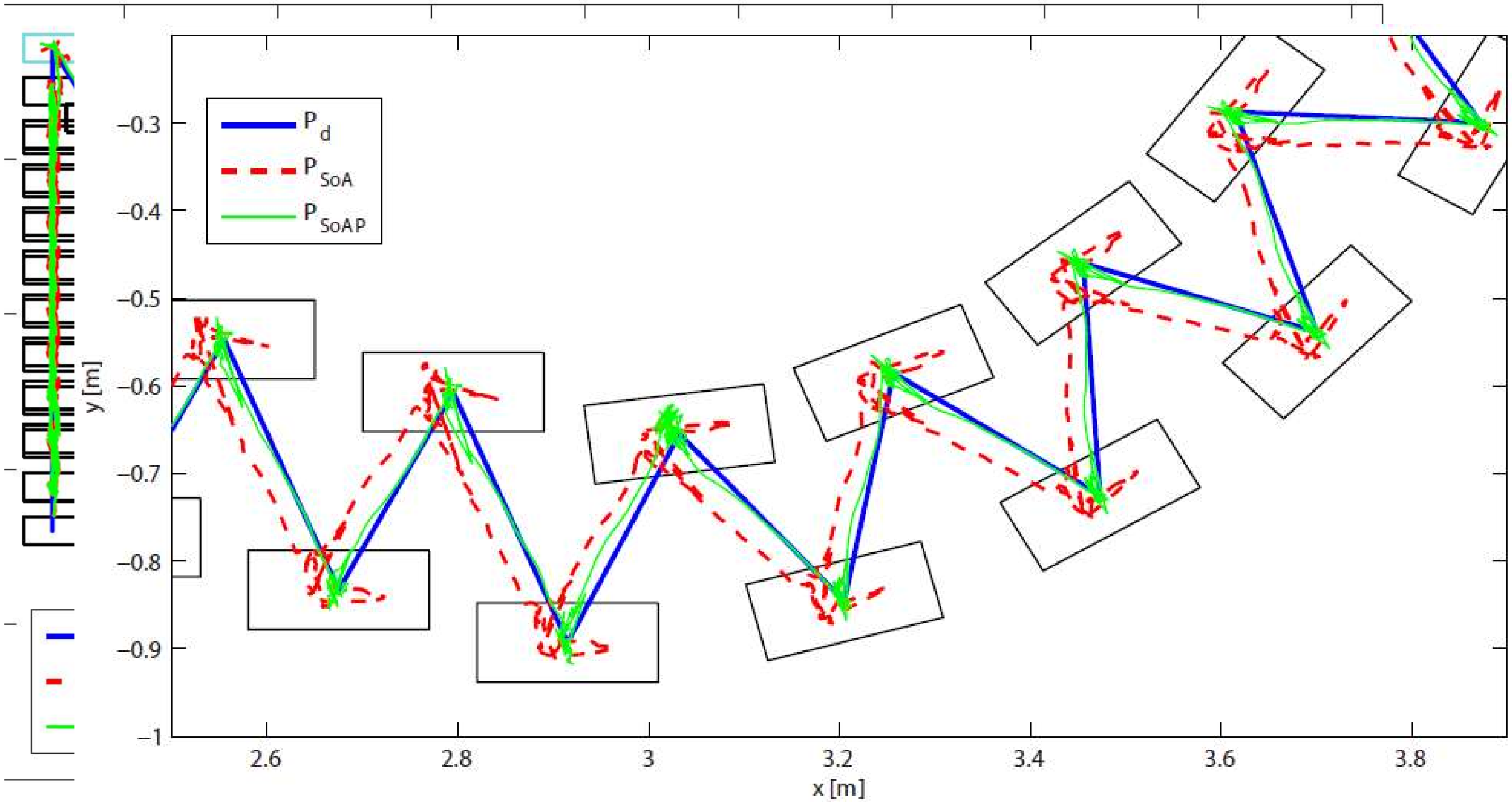
Learning Dataset of Compensative ZMP Term

Sagittal Straight Walking (SSW)	Lateral Straight Walking (LSW)	Circle Walking
$d_{sa} = \{-15, -10, -5, 0, 5, 10, 15\}$ cm	$d_{sa} = 0$ cm	$r = \{0.5, 0.75, 1\}$ m
$d_{la} = 0$ cm	$d_{la} = \{-7.5, -5, -2.5, 0, 2.5, 5, 7.5\}$ cm	$\Delta\alpha = \{10, 20\}^\circ$
$T_s = \{0.6, 0.8, 1.0\}$ s	$T_s = \{0.6, 0.8, 1.0\}$ s	$T_s = \{0.6, 0.8, 1.0\}$ s



Learning Dataset of Compensative ZMP Term

Sagittal Straight Walking (SSW)	Lateral Straight Walking (LSW)	Circle Walking
$d_{sa} = \{-15, -10, -5, 0, 5, 10, 15\}$ cm	$d_{sa} = 0$ cm	$r = \{0.5, 0.75, 1\}$ m
$d_{la} = 0$ cm	$d_{la} = \{-7.5, -5, -2.5, 0, 2.5, 5, 7.5\}$ cm	$\Delta\alpha = \{10, 20\}^{\circ}$
$T_s = \{0.6, 0.8, 1.0\}$ s	$T_s = \{0.6, 0.8, 1.0\}$ s	$T_s = \{0.6, 0.8, 1.0\}$ s

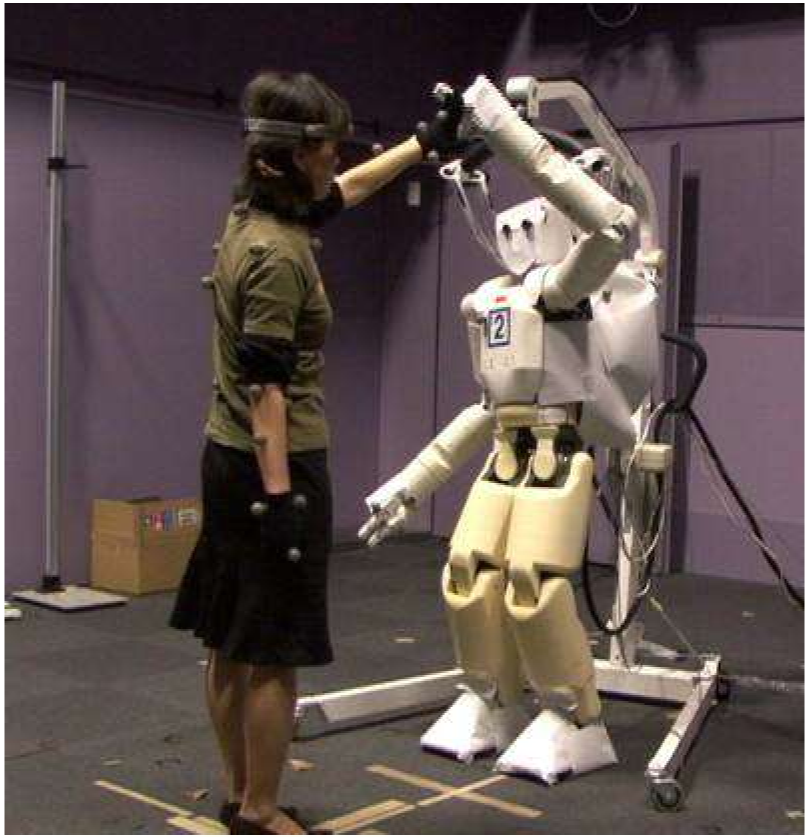


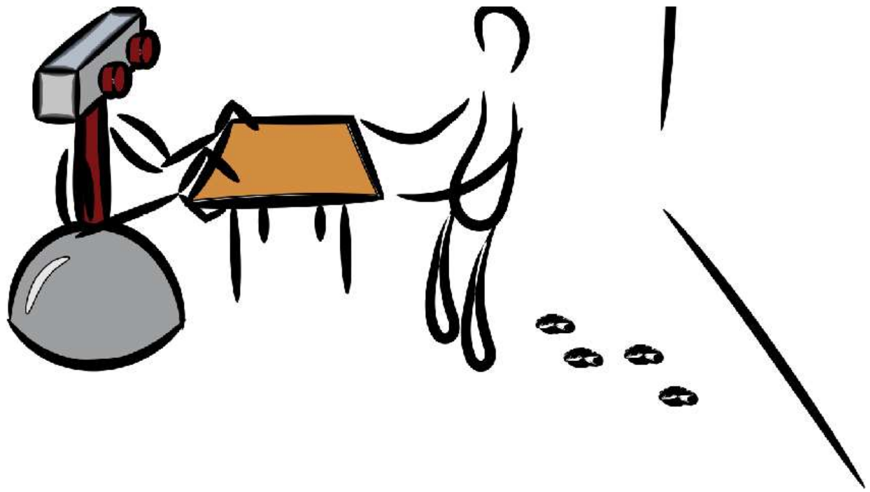
[ICRA 2015] [TRO 2016]

Overview

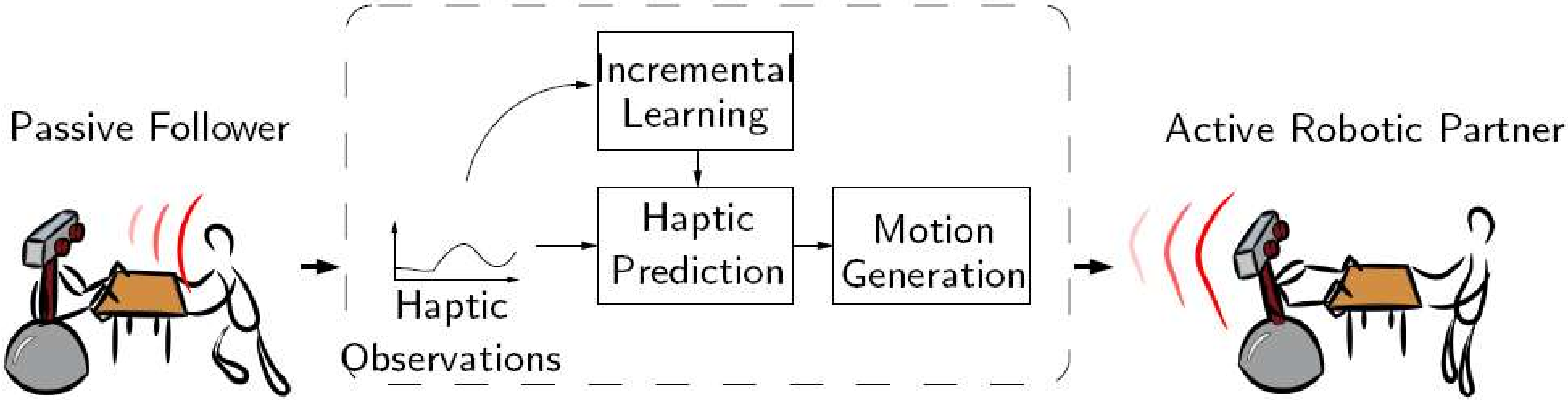
- Learning from Demonstrations
- Reinforcement Learning in Robotics
- Human Robot Interaction Learning
- Complex Manipulation Task Learning

Human is not a passive entity, but active, and full of uncertainty.



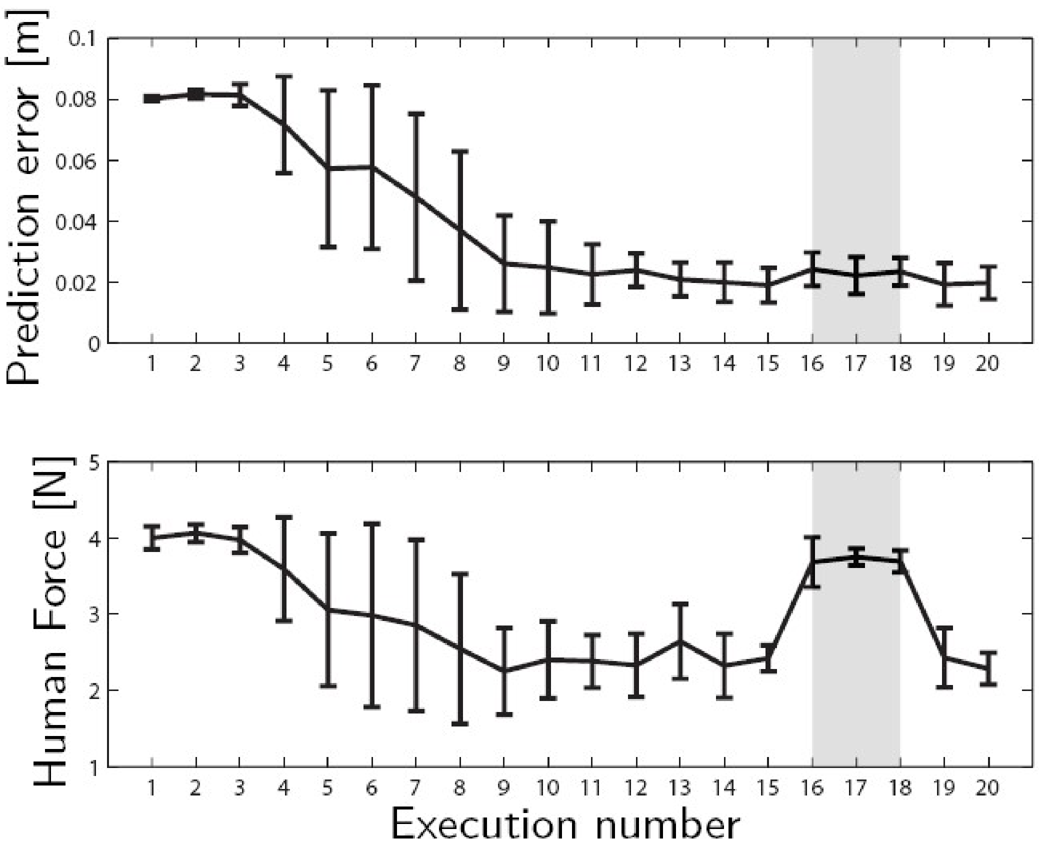


Experience-Driven Robotic Assistant



Experiment in 2D Virtual Scenario

- Robot learning, predicting and assisting during execution
- Repetitions 16 to 18 without assistance

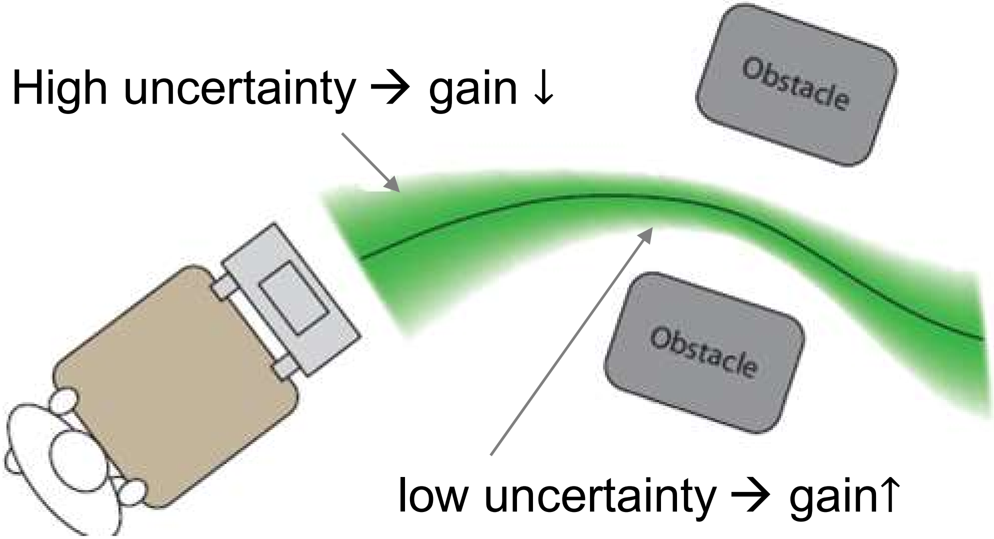


Overall results

- Robot assistance ↗
- Human force contribution ↘
- Execution time ↘
- Prediction error ↘

[IJRR 2011, IROS2011, ICRA 2012, IROS 2012]

Risk-sensitive Optimal Feedback Control



- Risk-averse
  gain ↑ → dominant
- Risk-seeking
  gain ↓ → passive

- Assistive behavior considering both human model *uncertainties*

$$u = K(\xi - \hat{\xi})$$

- Probabilistic human model for desired trajectory and exerted force

$$\hat{\xi} = \{\hat{\mu}_\xi, \hat{\Sigma}_\xi\}, \quad \hat{u} = \{\hat{\mu}_u, \hat{\Sigma}_u\}, \text{ with } \xi = (x \; \dot{x})^{\mathrm{T}}$$

- Risk sensitive stochastic optimal control

$$J = \sum_{k=1}^{T} \left((\xi_k - \hat{\mu}_\xi)^{\mathrm{T}} \hat{\Sigma}_{\xi,k}^{-\frac{1}{2}} Q \hat{\Sigma}_{\xi,k}^{-\frac{1}{2}} (\xi_k - \hat{\mu}_\xi) + u_{rk}^{\mathrm{T}} R u_{rk} \right)$$

[ICRA2012, IROS2012]

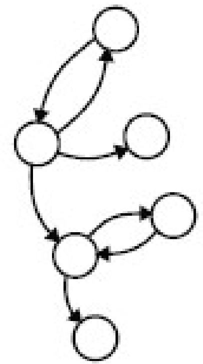
Overview

- Learning from Demonstrations
- Reinforcement Learning in Robotics
- Human Robot Interaction Learning
- Complex Manipulation Task Learning

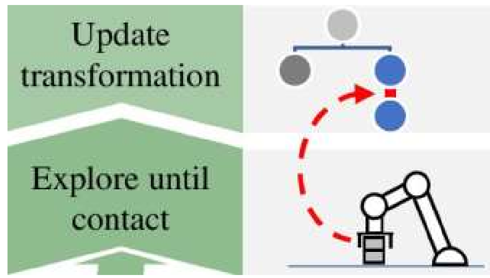
Learning complex tasks

Knowledge Representation: To find embedded structure of a task from demonstrations

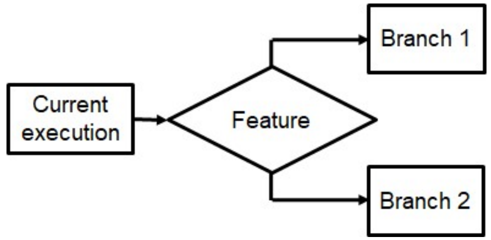
- Temporal structure
  - Clustering skills and learning sequencing order (transition probabilty)



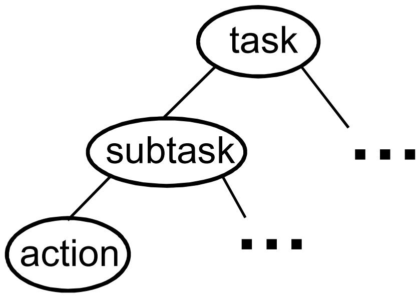
- Spatial structure
  - task parameters (e.g. coordinate system) of a skill
  - Spatial relation between skills



- Conditional Tasks
  - Decision making based on conditional reasoning

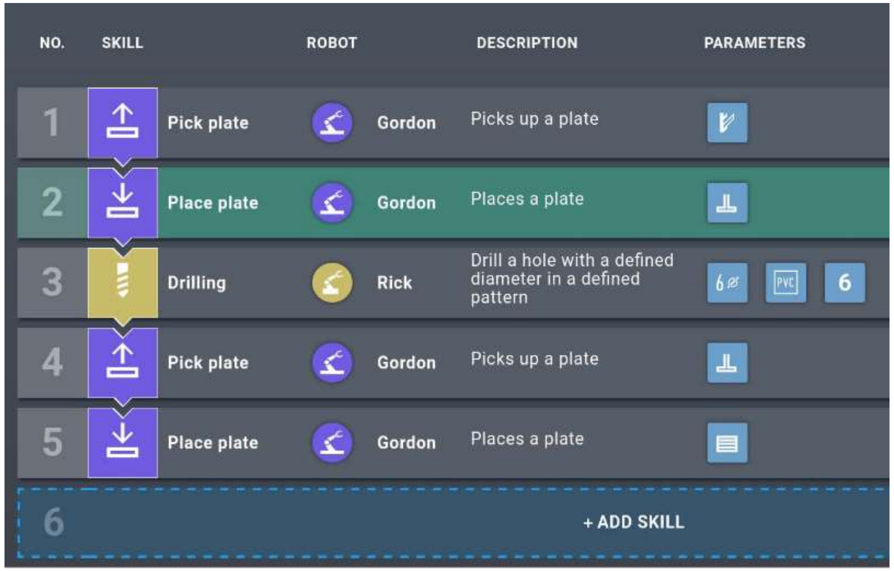


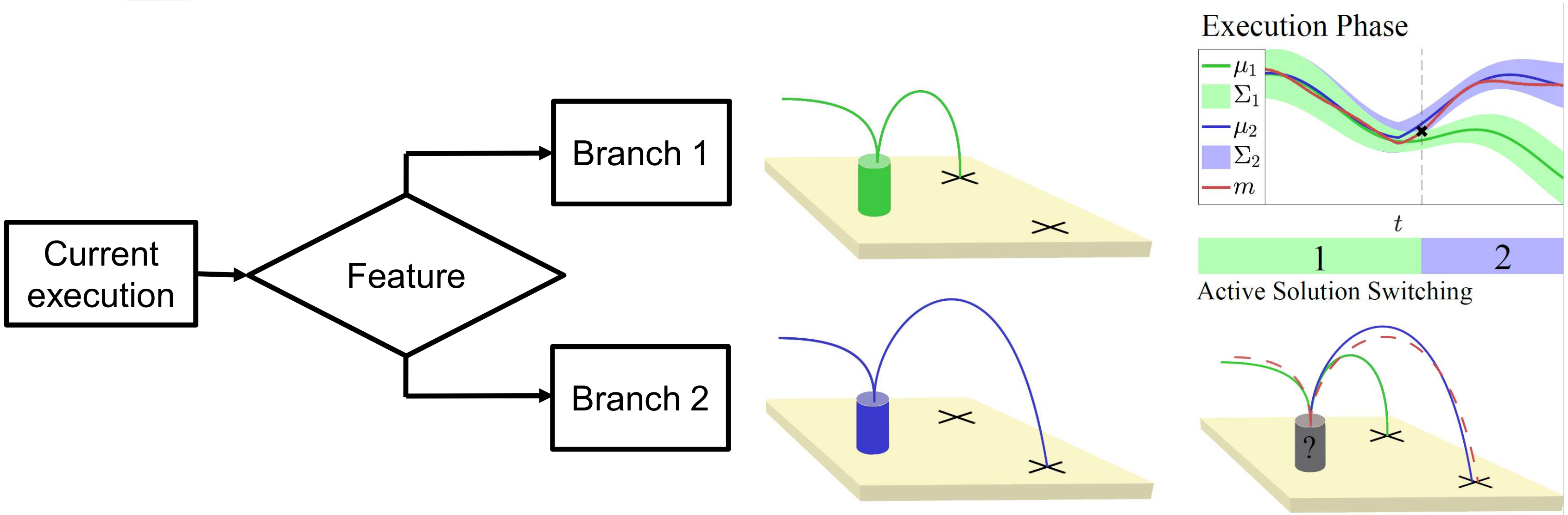
- Hierarchy in symbolic abstraction level
  - Task (e.g. make a coffee) – subtasks (e.g. add water) – skills (e.g. move A to B)



[IJRR11, Int. Service Robotics17, Humanoids17, ICRA19, AuRo18, RAL20]

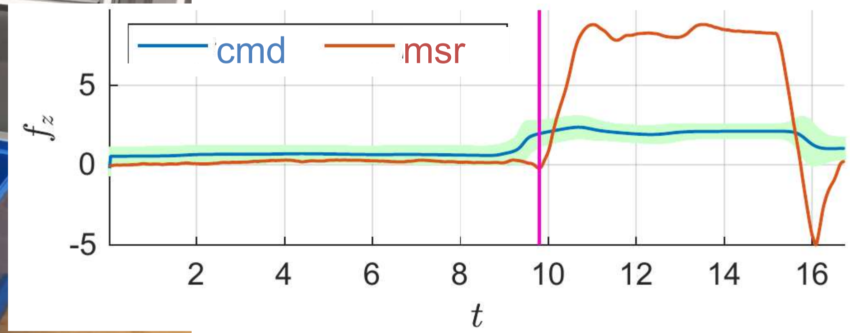
Fixed Sequencing → Conditional Sequencing





Experiment: Milk Carton Sorting

Demonstrations **w/o labeling**

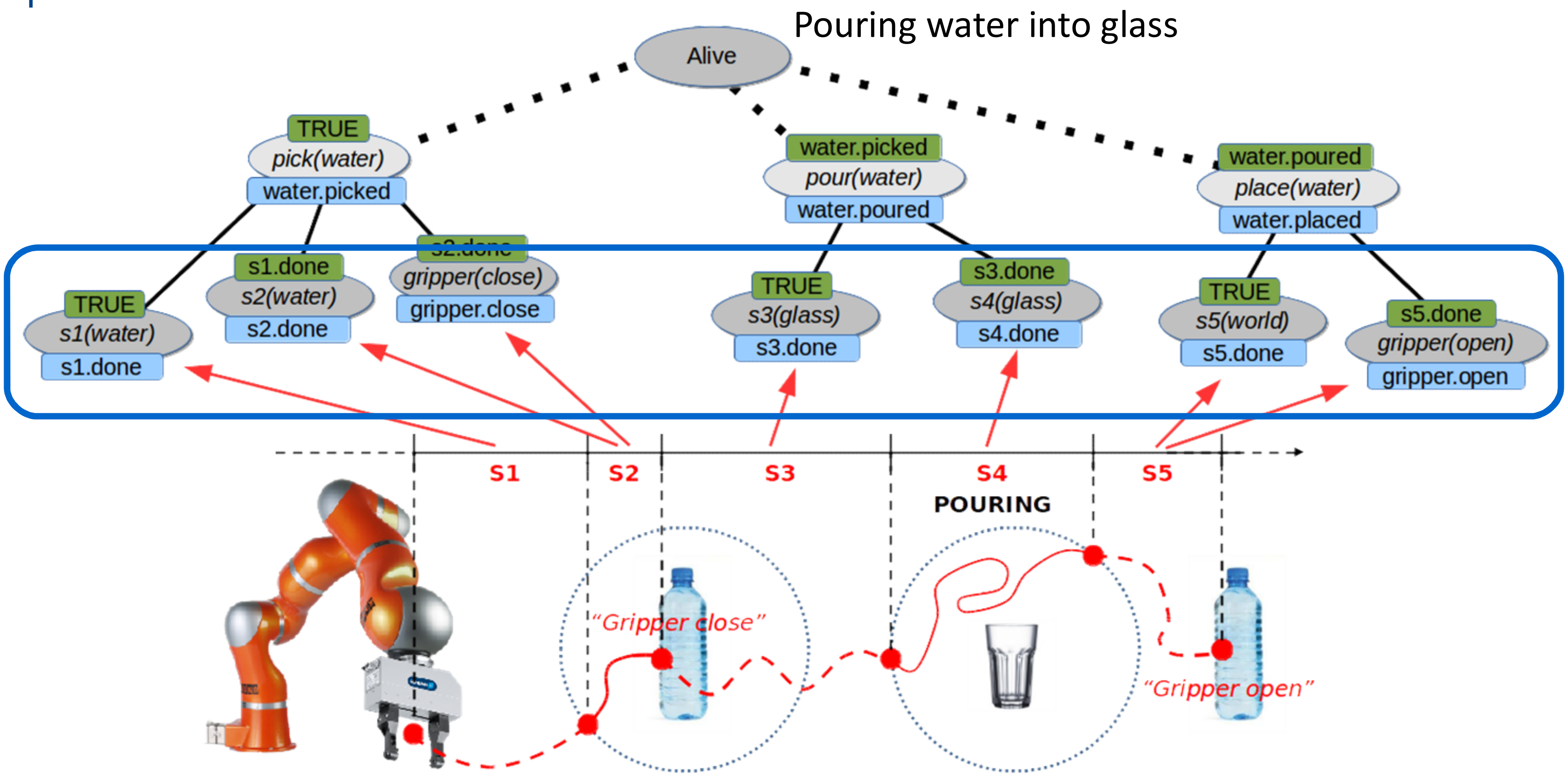


Anomaly detection

Execution

[Eiband et al, Learning Conditional Tasks by Demo. of Multi Solutions, RAL, 2019]

Learning Hierarchical Structured Tasks

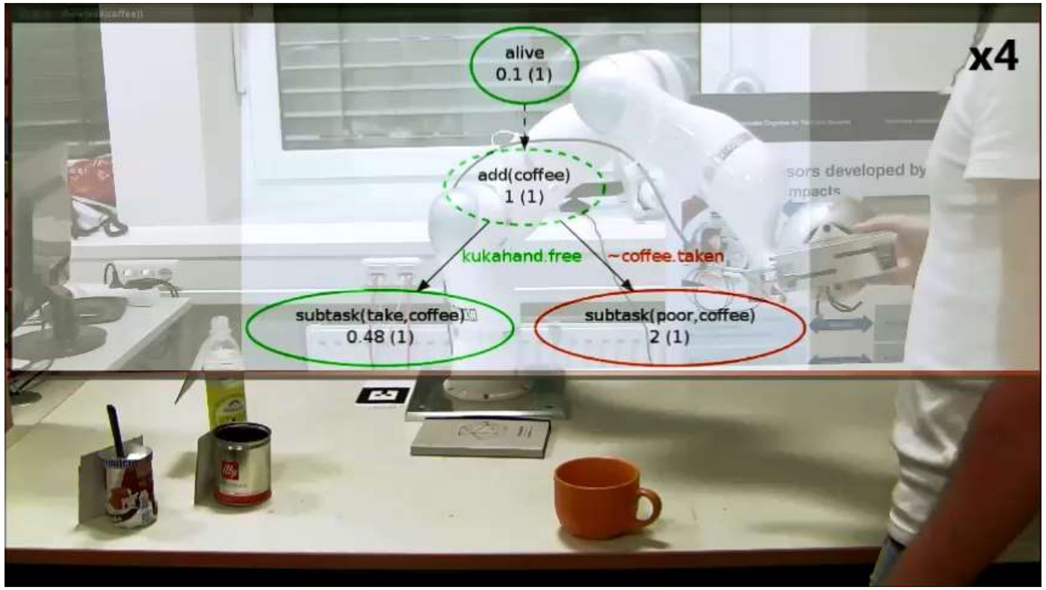


- Bridging low level MP learning and high level symbolic reasoning
- integrating imitation learning, attentional supervision, and cognitive control to learn and flexibly execute structured tasks

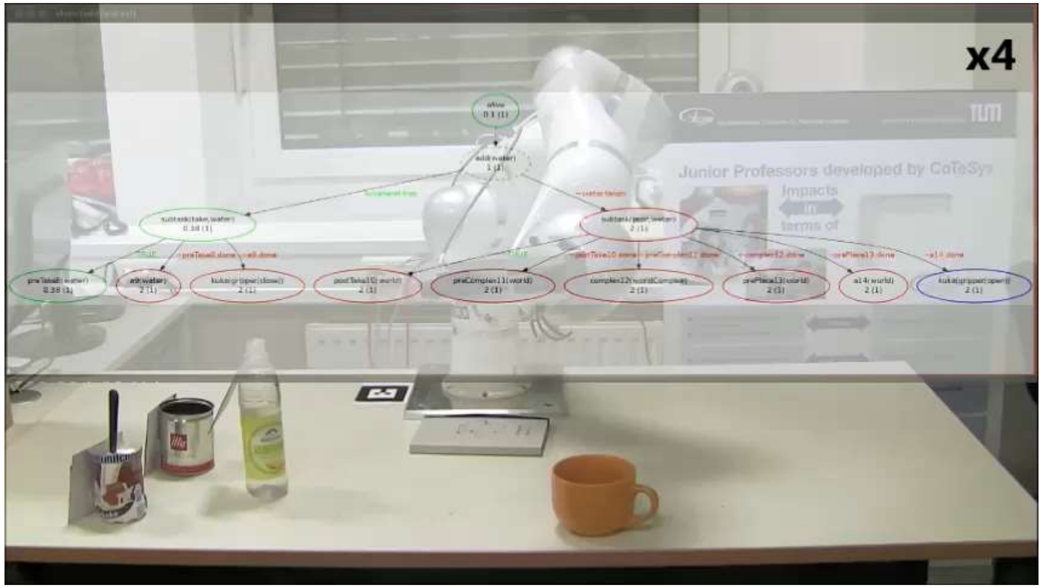
[Caccavale et al, Autonomous Robots 2018] [RA-L 2020]

Experiment: coffee making

Teaching



Execution

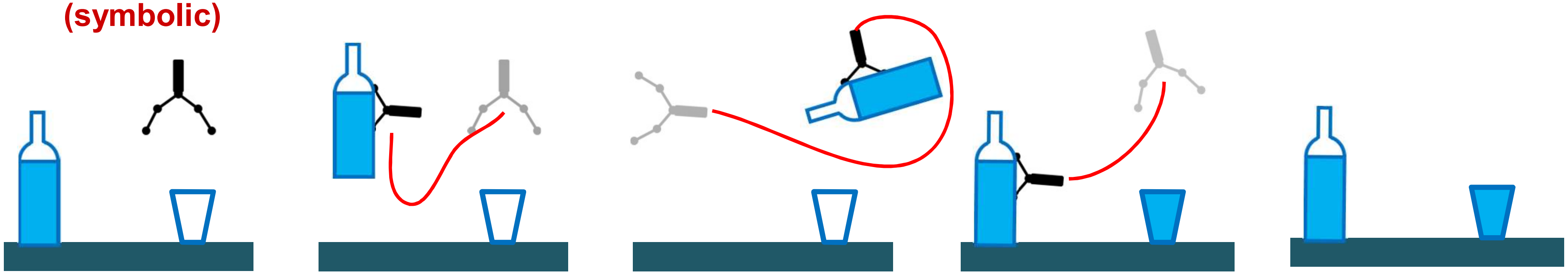


[AURO 2018]

Task Planning and Motion Planning

Initial state (symbolic)

Goal



on table bottle
on table glass
in bottle water
empty glass
empty hand

Pick bottle from table

```
(:action pick
:parameters
(bottle table)
:precondition (and
(on table bottle)
(empty hand))
:effect (and
(grasped bottle)
(not(on bottle table))
(not(empty hand))
```

Pour water in glass

```
(:action pour
:parameters
(water glass)
:precondition (and
(grasped bottle)
(empty glass))
:effect (and
(in glass water)
(not(empty glass))
```

Place bottle on table

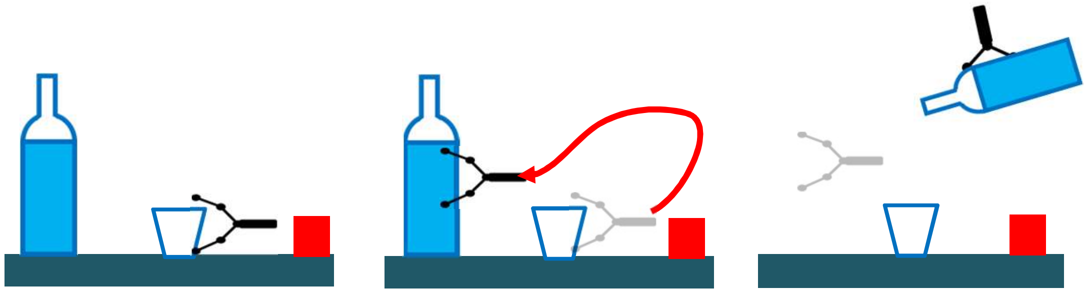
```
(:action place
:parameters
(bottle table)
:precondition (and
(grasped bottle)
:effect (and
(empty hand)
(on table bottle)
(not(grasped bottle))
```

in glass water
on table bottle

Planning Operators

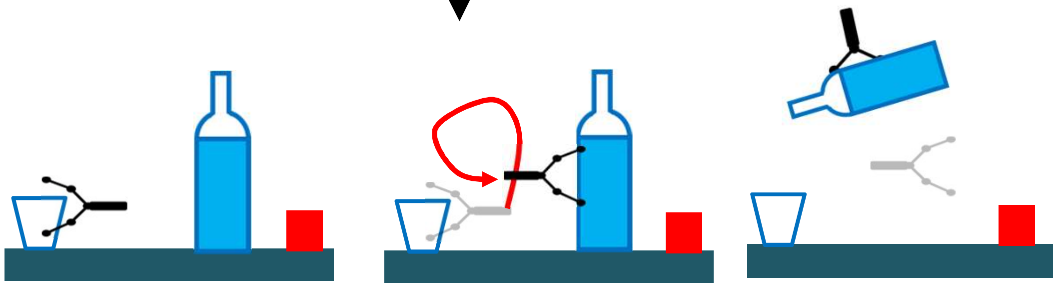
[Agostini et al, Manipulation Planning using Object-centered Predicates and Hierarchical Decomposition of Contextual Actions, RA-L 2020]

Action Context



Pick side bottle table

Same symbolic action but different motions!

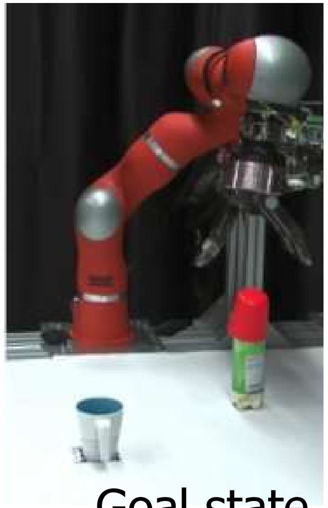


Pick side bottle table

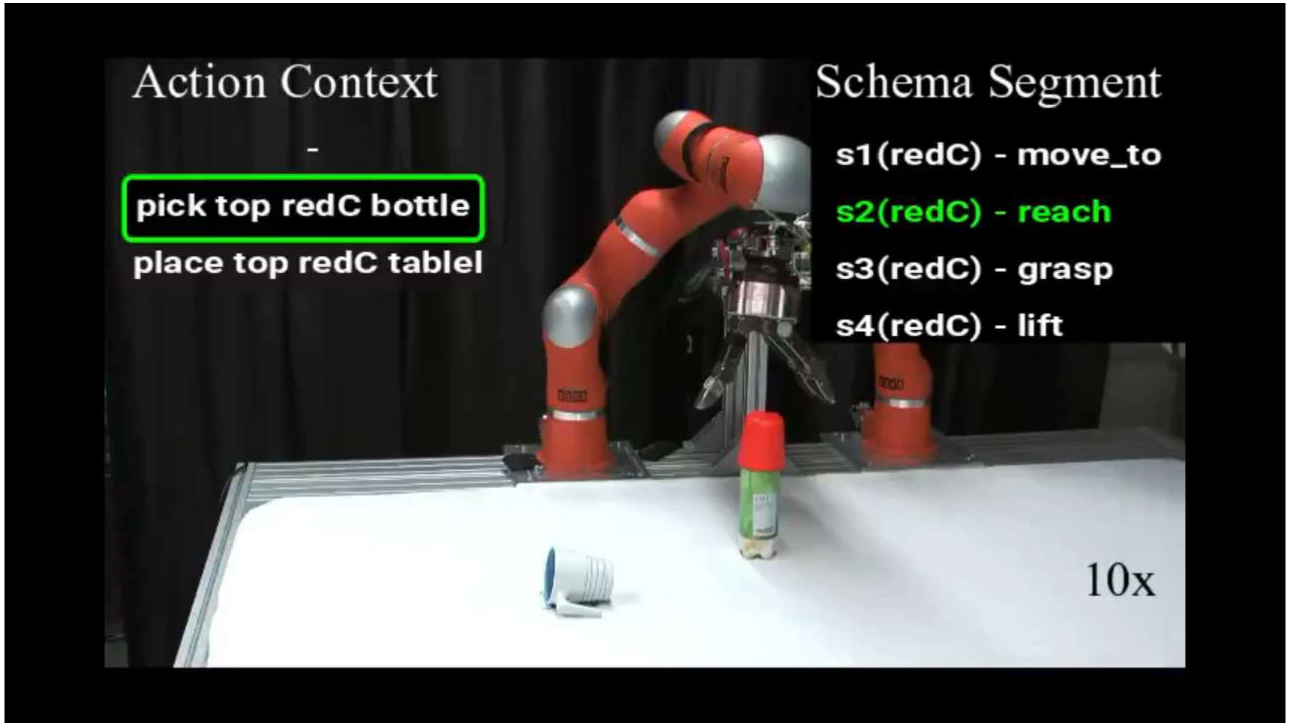
[Agostini et al, Manipulation Planning using Object-centered Predicates and Hierarchical Decomposition of Contextual Actions, RA-L 2020]

Experiments: Pouring Water

Initial state



Goal state



[Agostini et al, Manipulation Planning using Object-centered Predicates and Hierarchical Decomposition of Contextual Actions, RA-L 2020]

Summary: Challenges in Robot Learning

- *Skill transfer from Human to Robot* is a promising way towards intuitive programming and efficient motor skill learning.
- *Sample-efficient and Safe* Reinforcement Learning in *Physical World* can be acheived by leveraging imitation learning, approximate model knowledge, and learning in simulation.
- *Understanding human's behaviors and their uncertainties* leads to smooth and adaptive human robot interaction..
- In order to learn *complex robotic manipulation* tasks, it is essential to find the embedded structure of a task: sequencing, conditions, hierarchical abstraction.

What's next in robot learning?

- Robot learning requires an integrated architecture covering symbol grounding from sensing, symbolic reasoning, motion planning and adaptive control in physical world.

- Continual learning for Wide-Ranging Data
  - A robot can collect a large amount of information from a large variety of sensors, but rather low number of data. Simulator helps, but often do not reflect reality in a sufficient level of details.

- Social Interaction in Robot Learning Control
  - Account for the way in which data are collected.
  - Iterative interaction with the users can be exploited to influence the quality and nature of the collected data.
  - Linked with <u>active learning</u> with multimodal social interaction aspect.

Thank you for your attention

Thanks to Collaborators :

- Thomas Eiband
- Matteo Saveriano
- Affan Pervez
- Kai Hu
- Jose Medina
- Alejandro Agostini
- Pietro Falco
- Alberto Finzi
- Riccardo Caccavale
- Yoshihiko Nakamura
- Dana Kulic
- Jee-Hwan Ryu
- Christian Ott
- Angelika Peer
- Alexander Schmidts
- Sandra Hirche
- Hiba Latifee