Learning Manipulation Constraints from Human Demonstration for Humanoid Robots

Tamim Asfour
Robotics Institute Germany (RIG)
Karlsruhe Institute of Technology (KIT)

Robotics Everywhere

- Production
- Exploration
- Assistance
- Service

- Logistics
- Agriculture
- Construction
- Medicine

- Rehabilitation
- Laboratory Automation
- Rescue
- Decontamination

- Micro- /Nanorobotics
- Human Augmentation
- Education
- ....







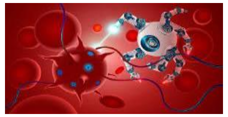






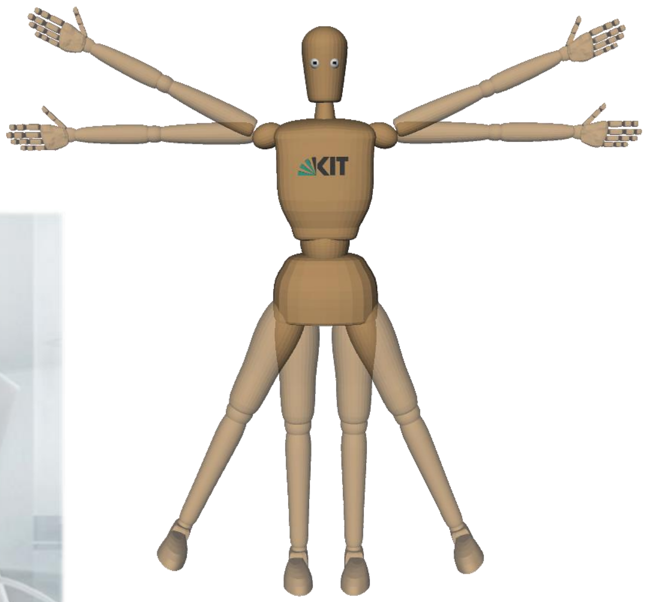




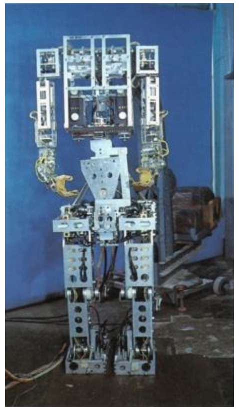


Humanoid Robotics

- Understanding human **versatility** and motion intelligence
- Creating multi-purpose machines with **physical intelligence**
- Engineering **human-centered technologies** to empower humans
- “Humanoid Robotics is a **Multi-Trillion Market**”



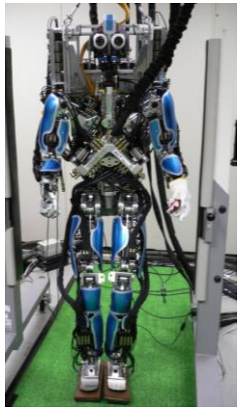
Humanoid Robotics has made Progress (1970 – 2023)


WABOT-I


ASIMO


Toyota


DB


CB


HRP-2


HRP-4C


Kenshiro


HUBO


iCub


ARMAR-III


Justin


Robonaut


REEM-C


Valkyrie


NAO


ALTAS

Humanoid Robotics has made Progress (2022 – 2025)

Apptronik
Apollo

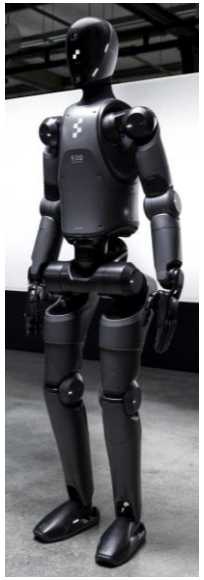
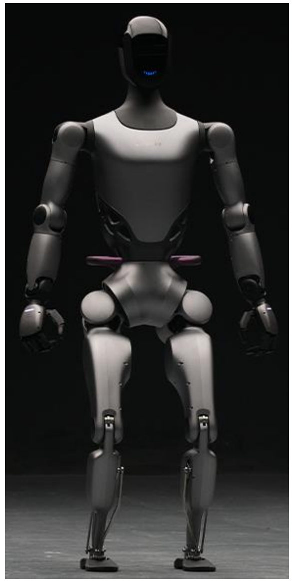
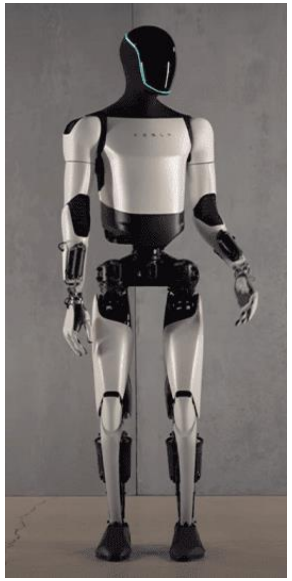


Figure AI
Figure 02

Unitree
H1 & G1



Fourier
GR-2



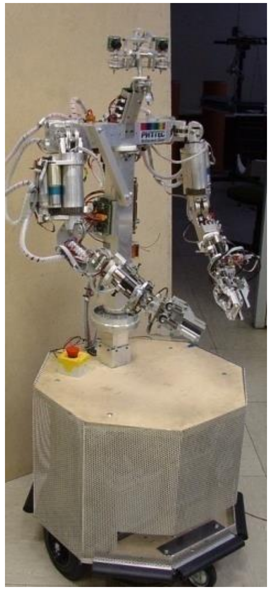
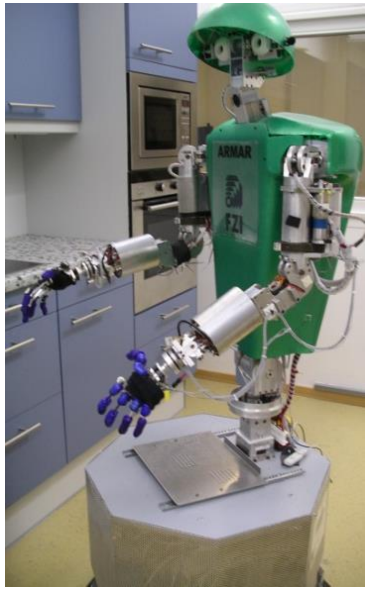
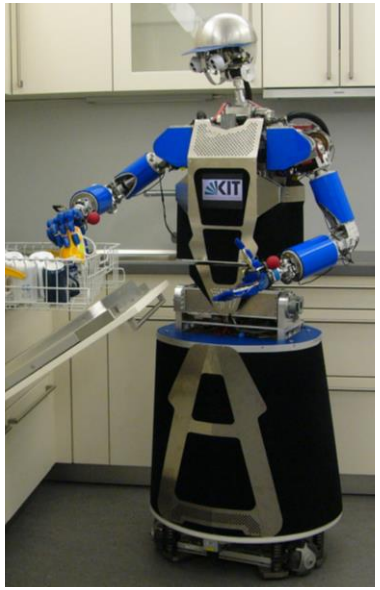
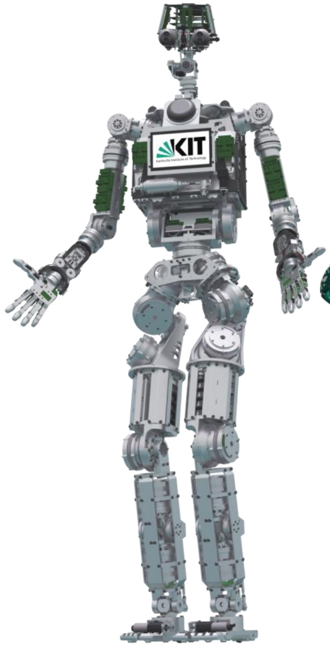
Tesla
Optimus

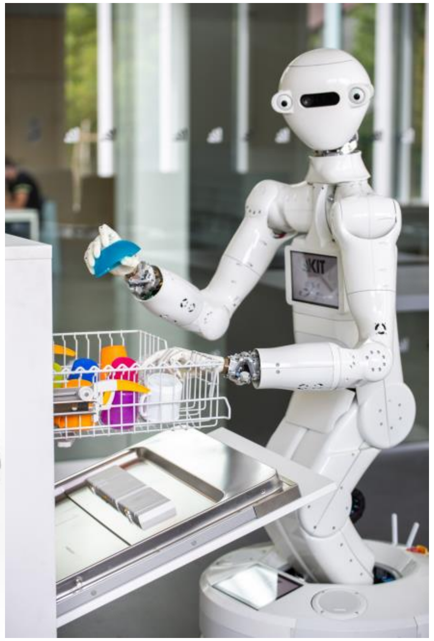
KIT
ARMAR-7

Today: Hundreds of new humanoids; large investments (in China and US)

The ARMAR Family | Robots

MADE IN GERMANY

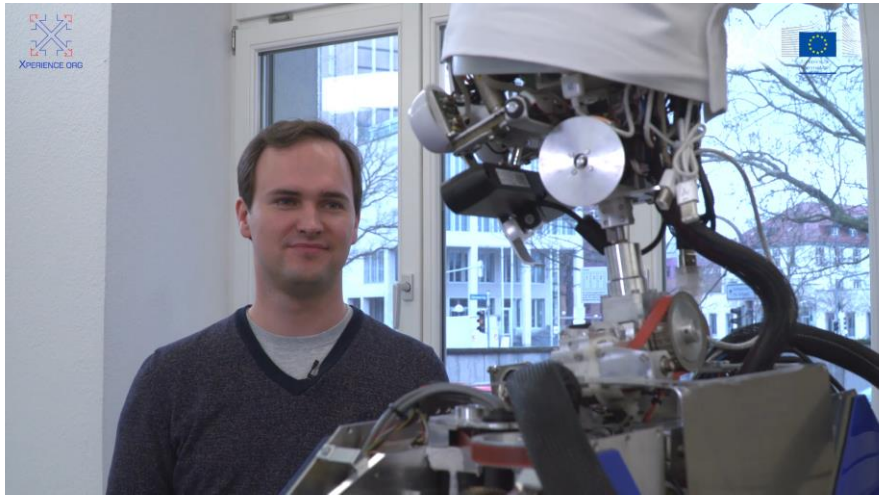
ARMAR-I	ARMAR-II	ARMAR-III	ARMAR-4	ARMAR-6	ARMAR-DE	ARMAR-7
2000	*2002*	*2006*	*2013*	*2017*	*2020*	*2024*

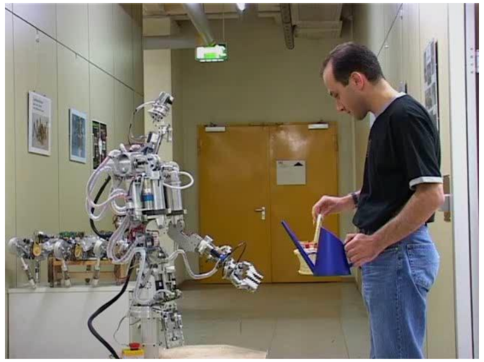
The ARMAR Family | Hands

TUAT/KIT Hand	ARMAR-III Hand	ARMAR-IV Hand	ARMAR-6 Hand V2	Prosthetic Hand V1	Prosthetic Hand V2.1	Sensorized Hand	Finger-Vision Hand	Prosthetic Hand V2.2	ARMAR-7 Hand
2000	*2006*	*2012*	*2018*	*2018*	*2020*	*2020*	*2020/21*	*2021*	*2024*

The ARMAR Humanoid Robots @ Home

- Versatile personal assistance for humans at homes, nursing and retirement homes, …
- Self-determined life of humans



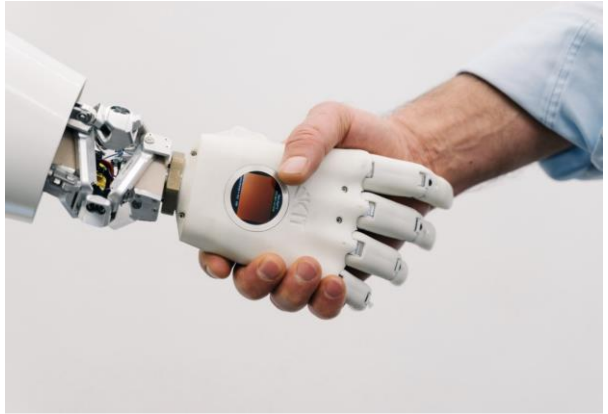


ARMAR-7
automatica 2025

ARMAR-7

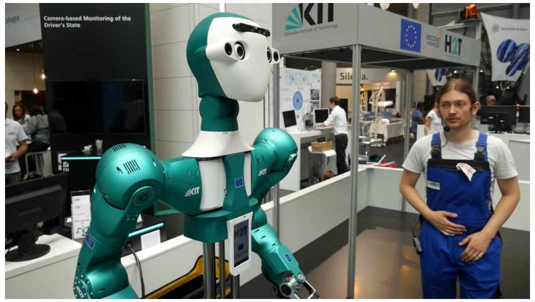
Humanoid Robots in Care Facilities

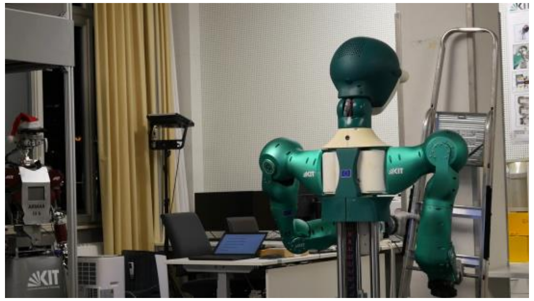
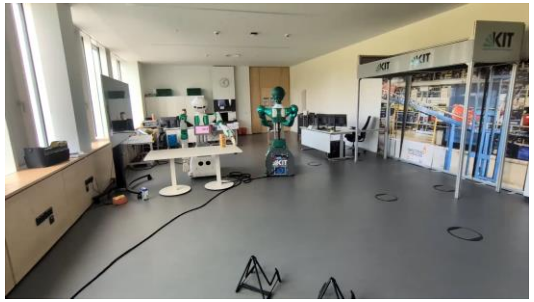
Humanoid robots – a solution for societal challenges

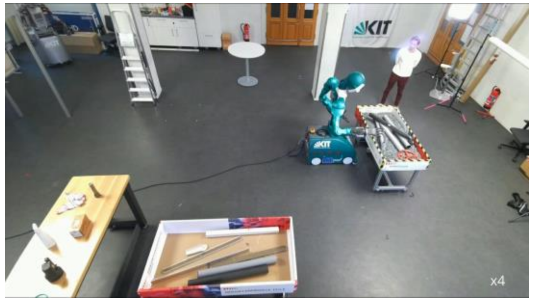


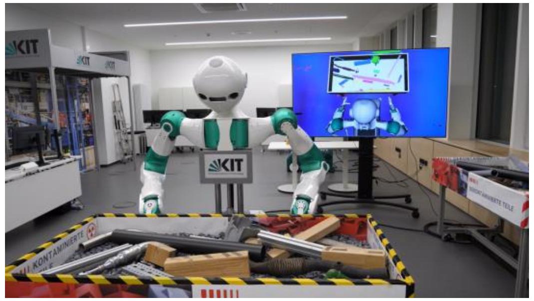
The ARMAR Humanoid Robots @ Work

- **Pro-active and adaptive assistance for workers in 3D jobs**
- **Meaningful work for humans**

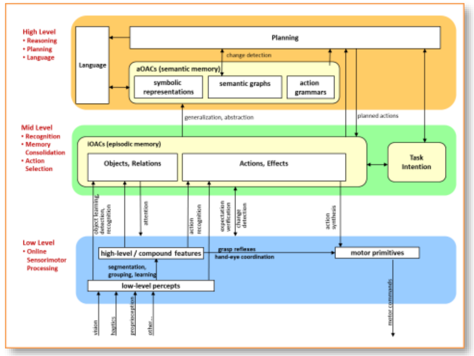


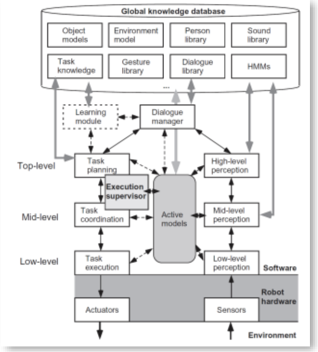





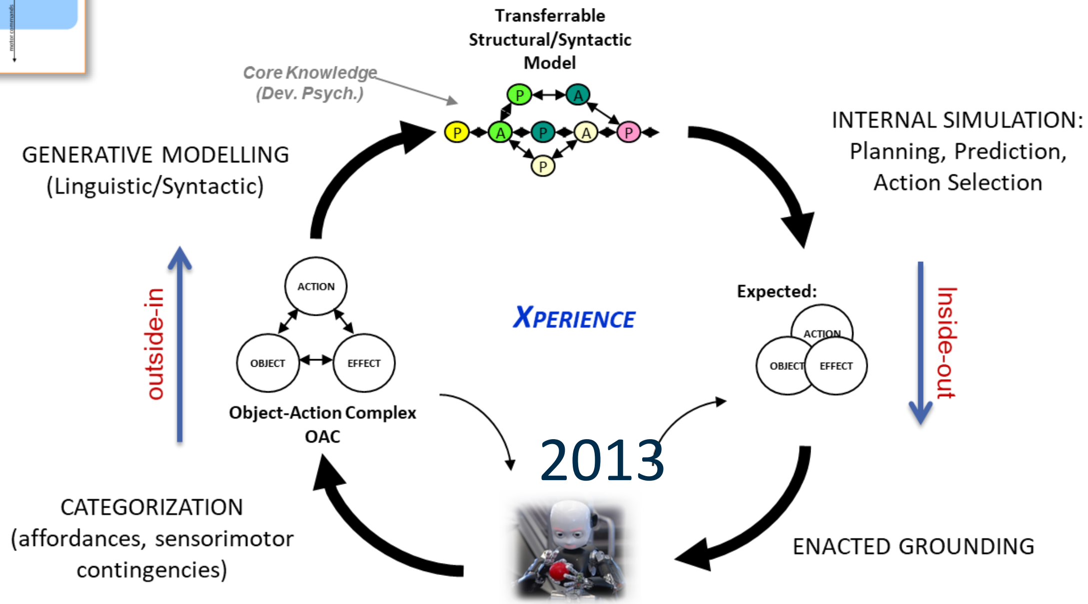

The ARMAR Robots Cognitive Architecture(s)

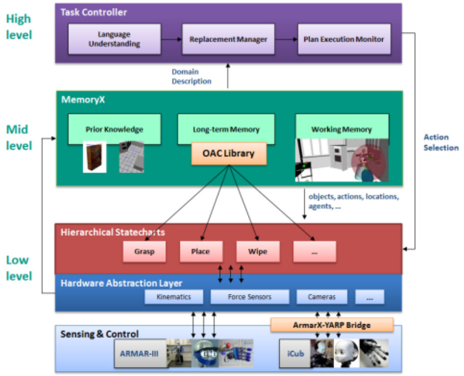


2009

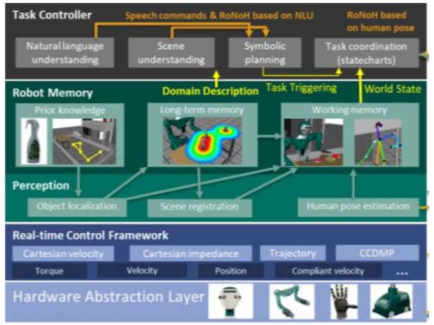


2005





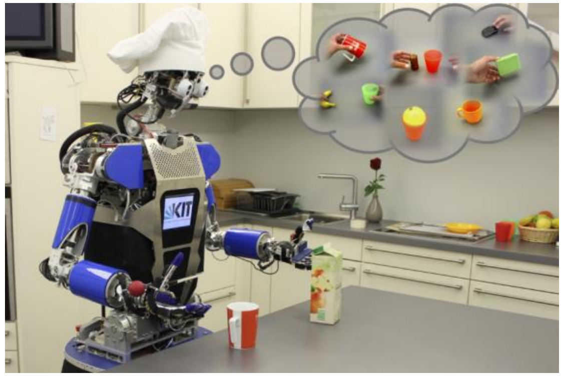
2016



2018

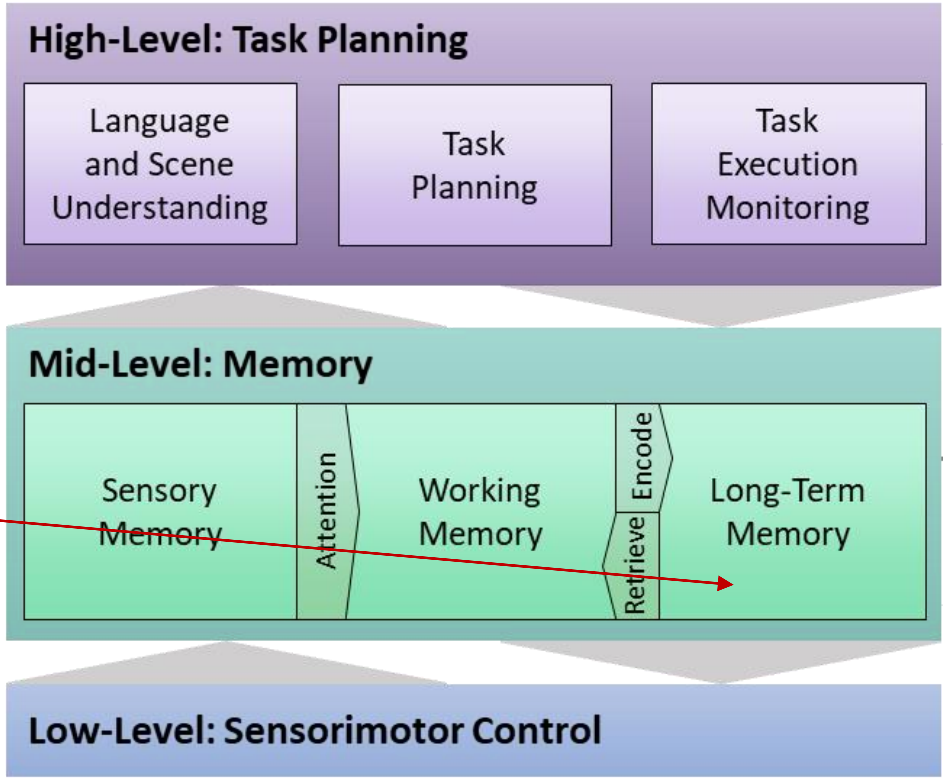
Hybrid, memory-based, transferable

Deep Episodic Memory



- How to **encode experiences?**
- How to **recall experience** and **predict actions?**
- How to **verbalize experience?**

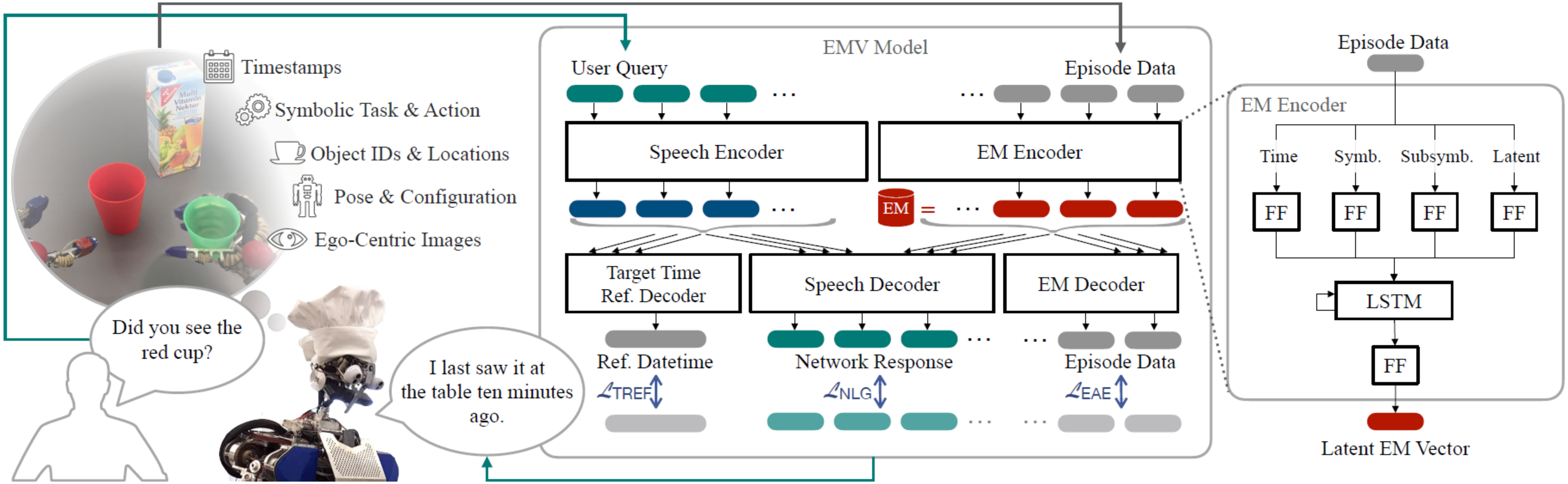
Episodic Memory



Deep Episodic Memory

Verbalization of Robot Experience

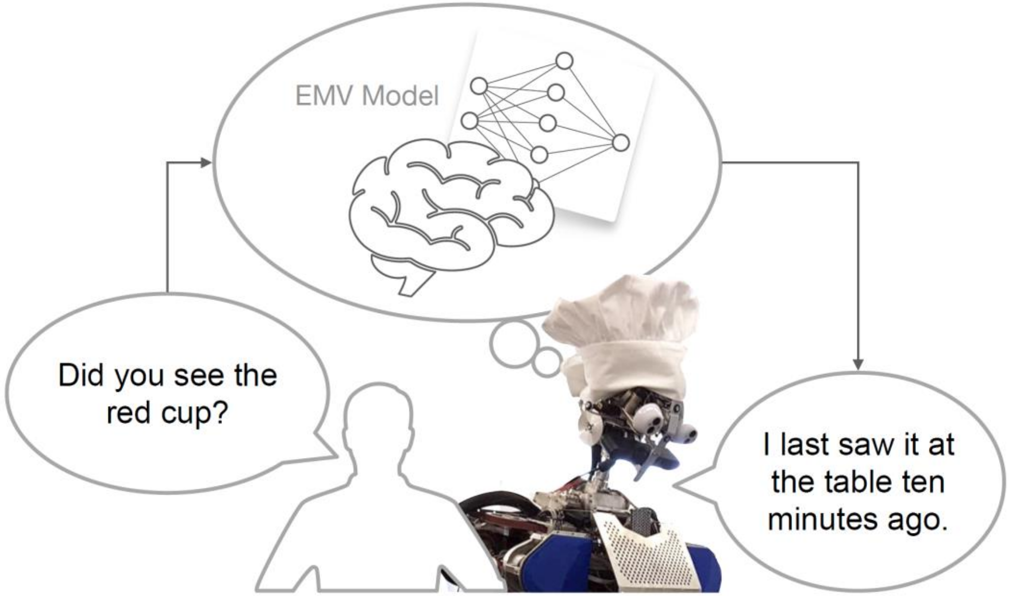
Deep neural end-to-end architecture



Deep Episodic Memory

Verbalization of Robot Experience

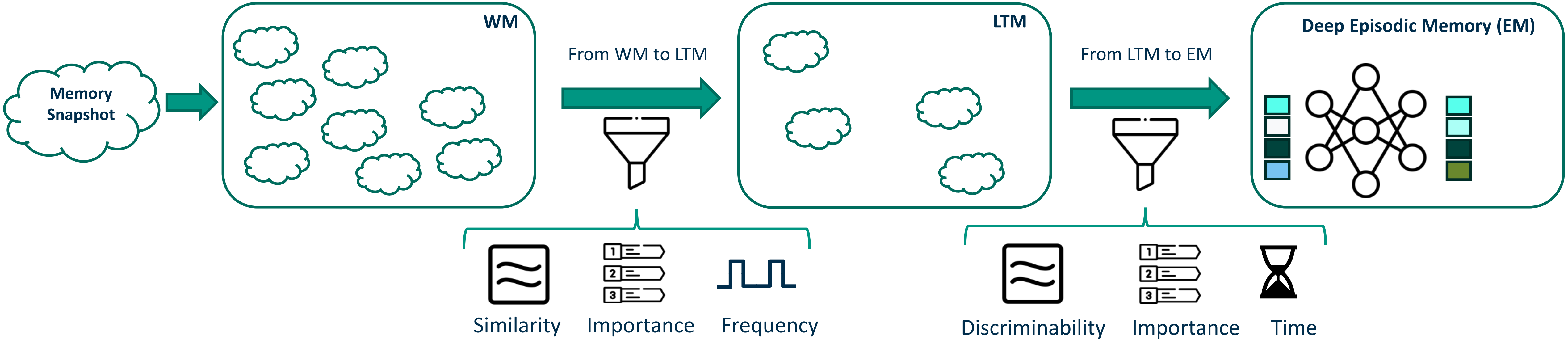
Deep neural end-to-end architecture



Deep Episodic Memory

Forgetting

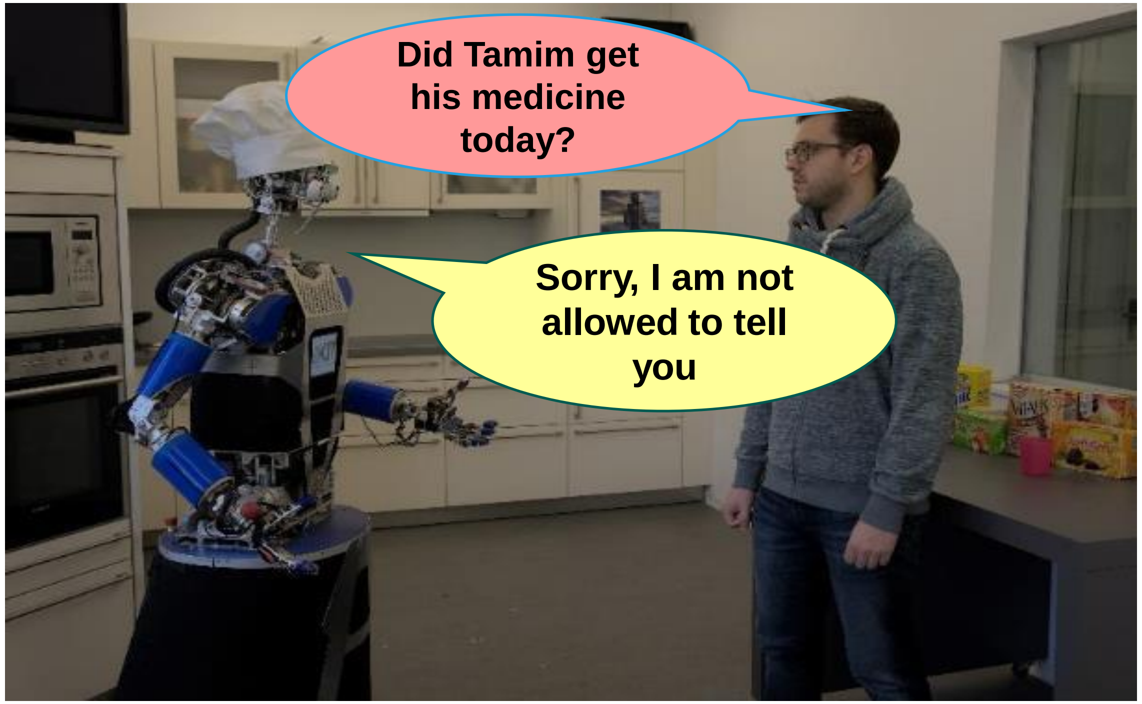
- Robots generate overwhelming amount of data (ARMAR-6 350 GB/h = 8.4 TB/day = 3.1 PB/year), significant even in latent space; much of the data contains redundant or similar information
- **Human-inspired forgetting as a solution**



Freedman, S. and Adams, J., *Filtering Data Based on Human-Inspired Forgetting*, IEEE Transactions on Systems, Man, and Cybernetics, 2011

Episodic Memory
Personalization and User Privacy

- How to adapt skills to user **preferences, habits and abilities**
- How to ensure user's **privacy**?
- What is a robot **allowed** to do?



Manipulation

Manipulation

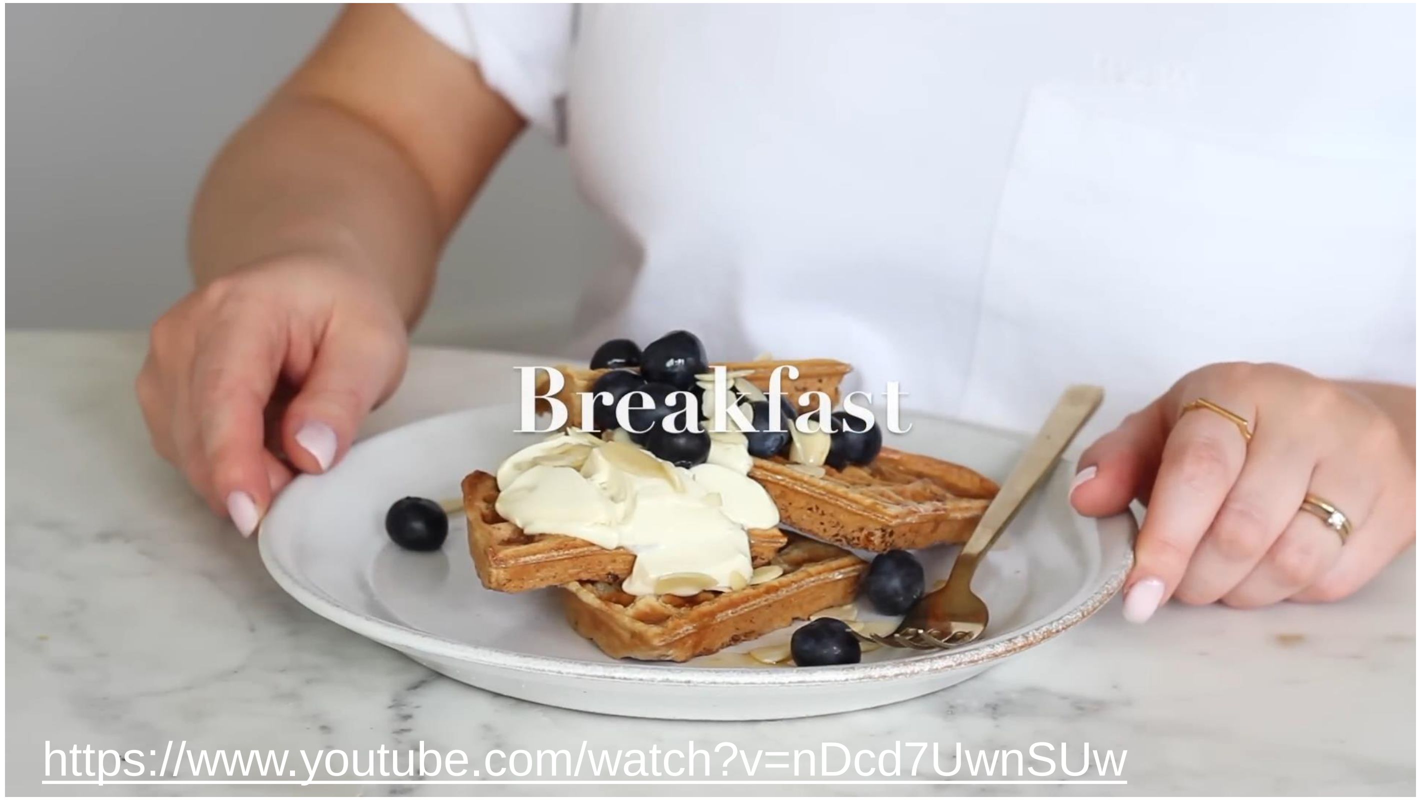
- Robotic manipulation stands as one of the grand challenges in robotics.
- **Manipulation** is not only a technical hurdle but a fundamental **gateway to intelligence**.
- Manipulation – the ability to change the physical world – is intrinsically intertwined with intelligence – the ability to detect change and adapt to it.

2008 – now

2018 – now

Manipulation – Understanding Constraints

- Manipulation is about understanding and leveraging **constraints**
- Constraints provide a language that bridges the gap between perception, planning, and control in manipulation.



Manipulation – Understanding Constraints

- How to detect constraints?
- How to formulate constraints?
- How to learn constraints from human demonstrations and through environmental interaction?
- How to leverage constraints for learning and control?

Taxonomies

- **Taxonomies as structured frameworks for understanding the space of possible constraints**

- Taxonomies for
  - Grasping tasks (Keller, Cutkosky, Kamakura, GRASP, Bullock and Dollar, Pollard, …)
  - Bimanual manipulation tasks (Kelso, Guiard, Kantak, …)
  - Loco-manipulation tasks (Borras Sol, …)

Bimanual Manipulation Taxonomy

bimanual manipulation

uncoordinated | ***coordinated***

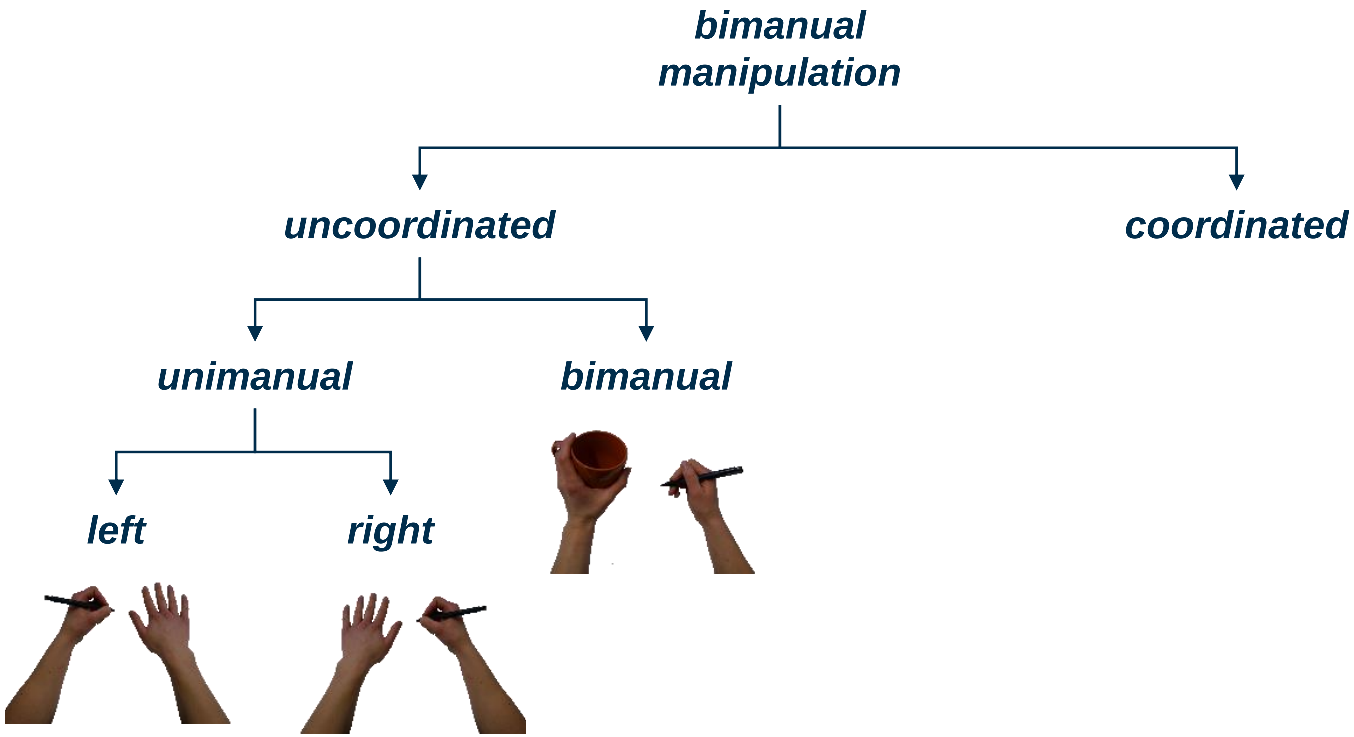
unimanual

left | ***right***

coordination

https://www.youtube.com/watch?v=nDcd7UwnSUw

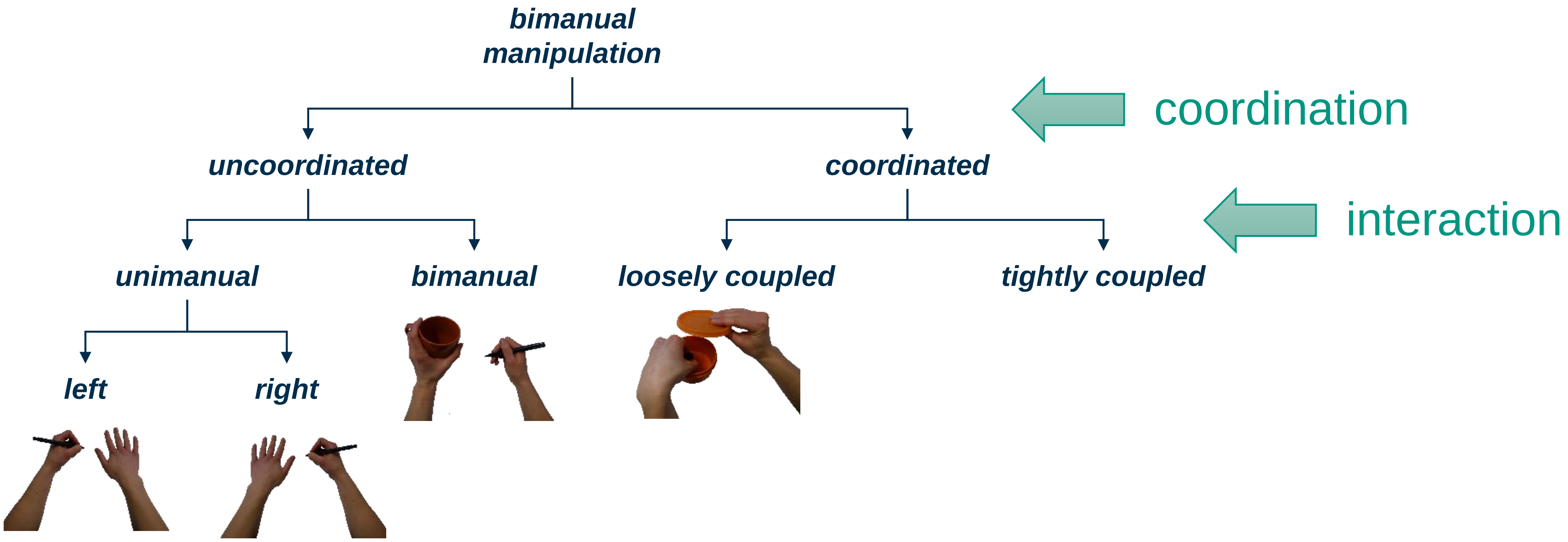
Bimanual Manipulation Taxonomy



coordination

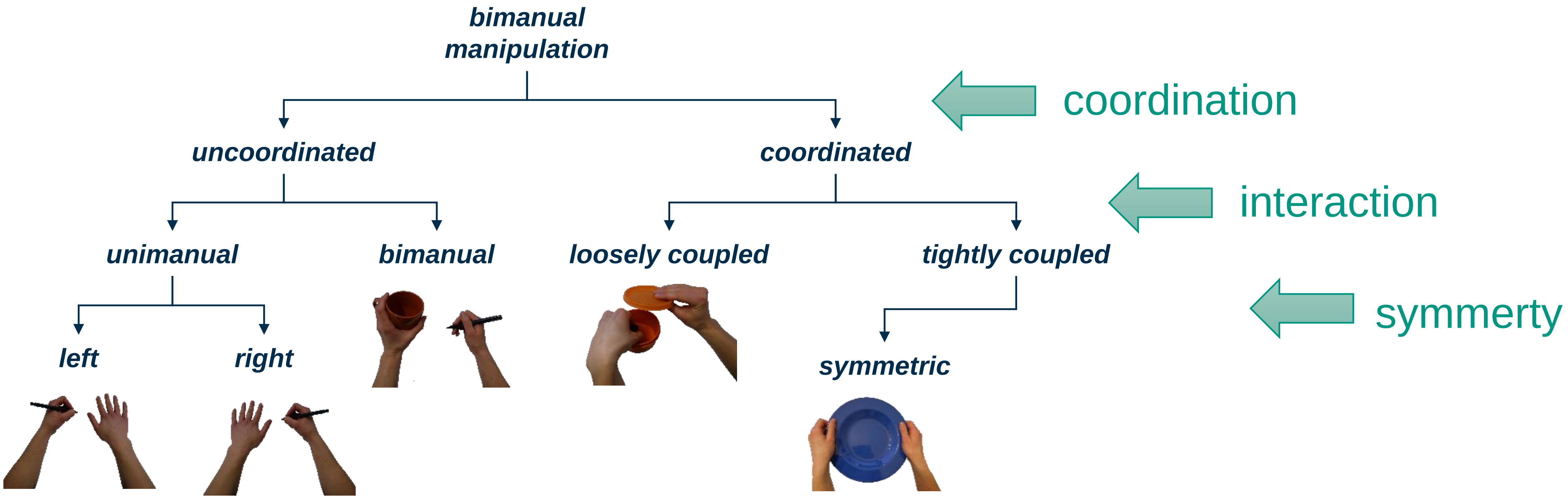
https://www.youtube.com/watch?v=nDcd7UwnSUw

Bimanual Manipulation Taxonomy



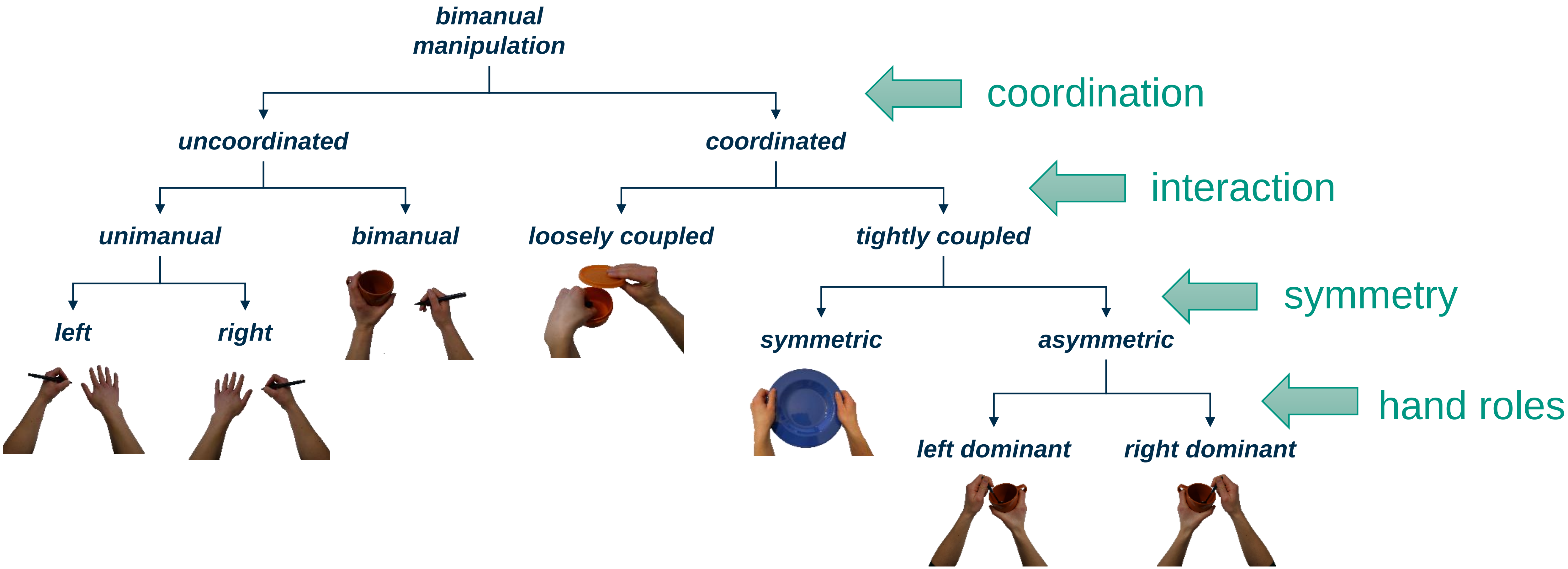
https://www.youtube.com/watch?v=nDcd7UwnSUw

Bimanual Manipulation Taxonomy

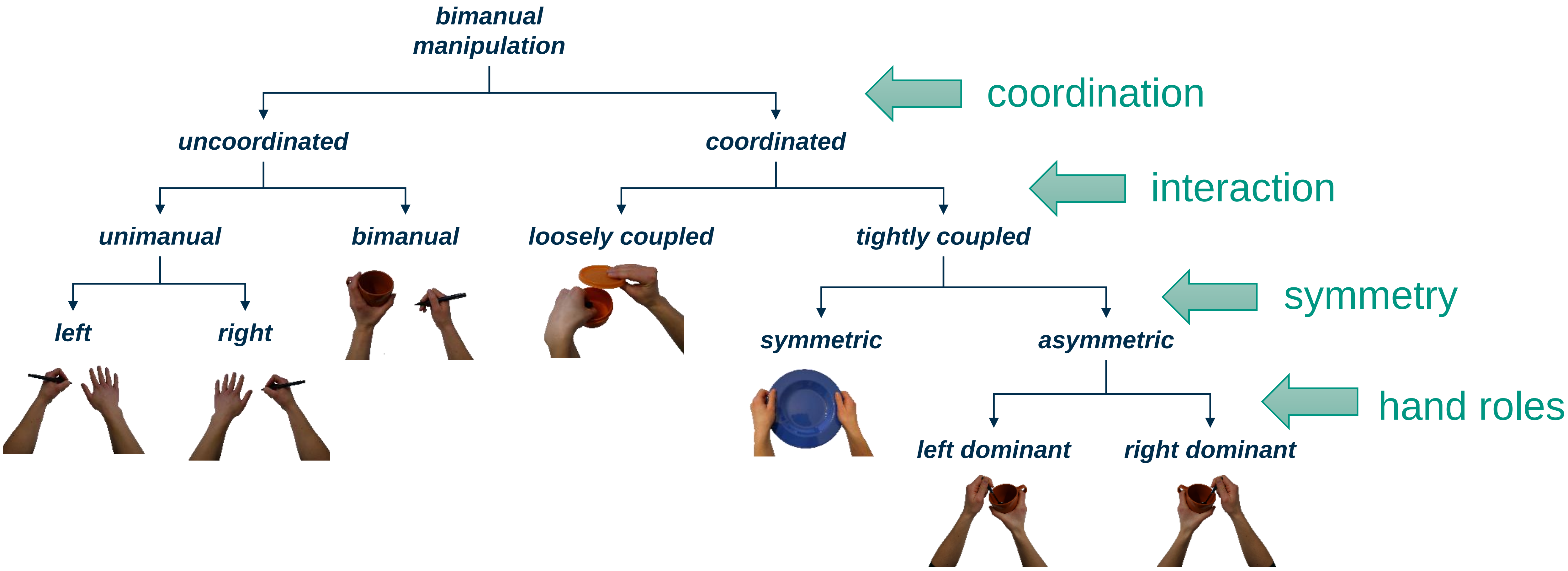


https://www.youtube.com/watch?v=nDcd7UwnSUw

Bimanual Manipulation Taxonomy

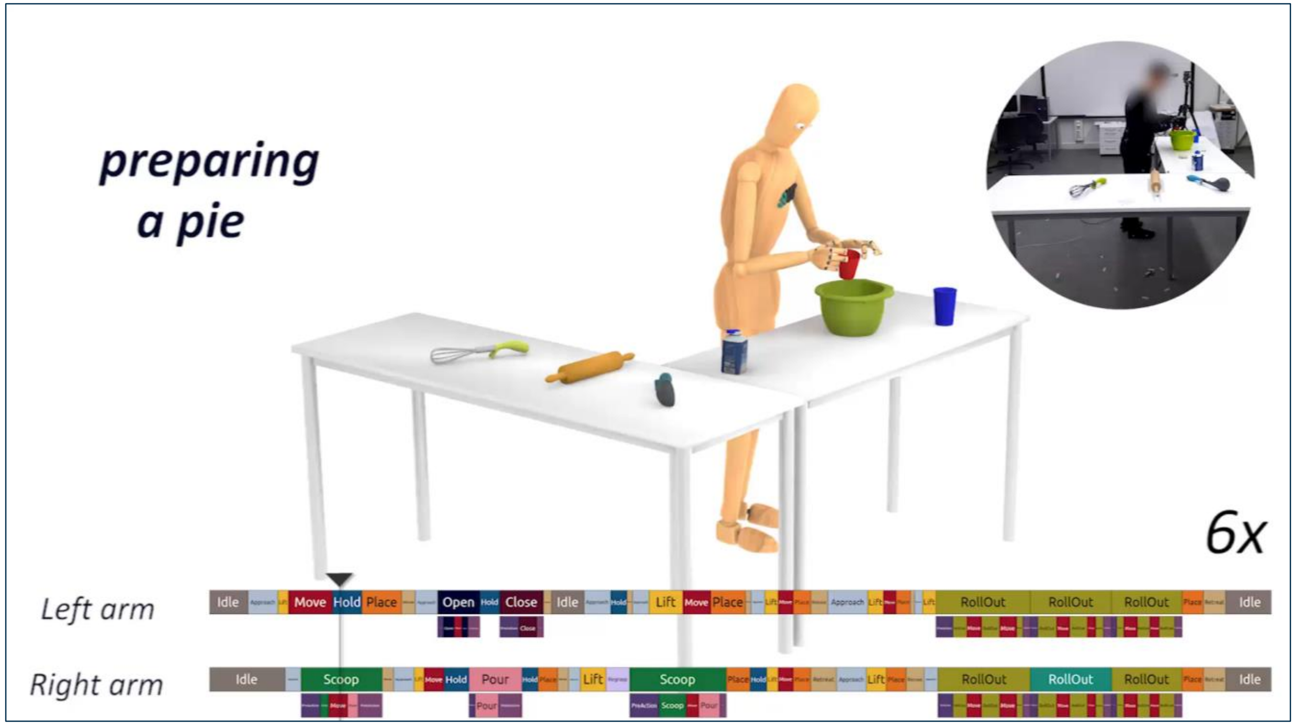


Bimanual Manipulation Taxonomy



Franziska Krebs and Tamim Asfour, *A Bimanual Manipulation Taxonomy*, IEEE Robotics and Automation Letters (RA-L), 2022

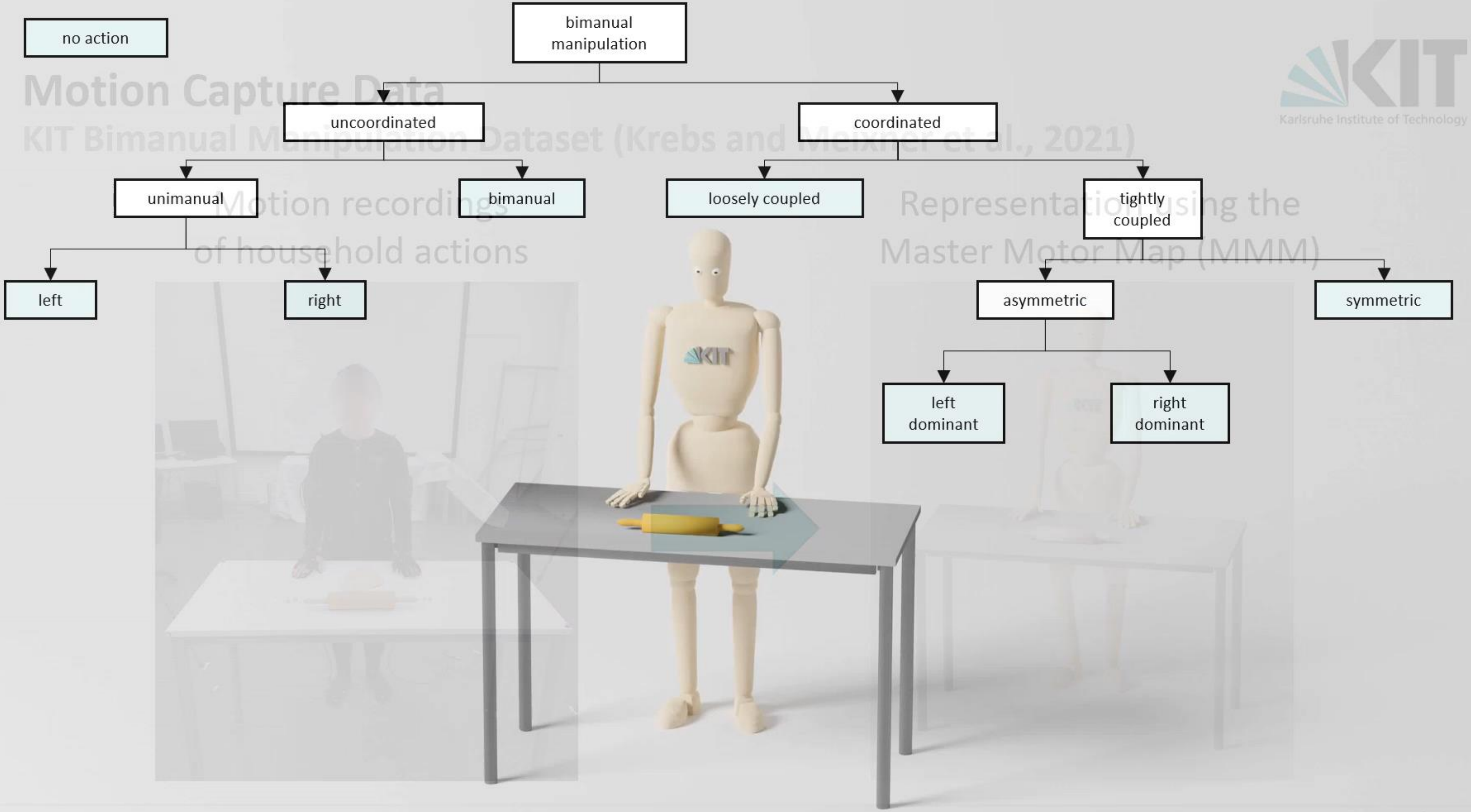
KIT Whole-Body Human Motion Database



42 hours of manually labeled human motion data (including object information); **9375 motions; 234 (112/41)** subjects and **158** objects.

motion-database.humanoids.kit.edu
https://gitlab.com/mastermotormp

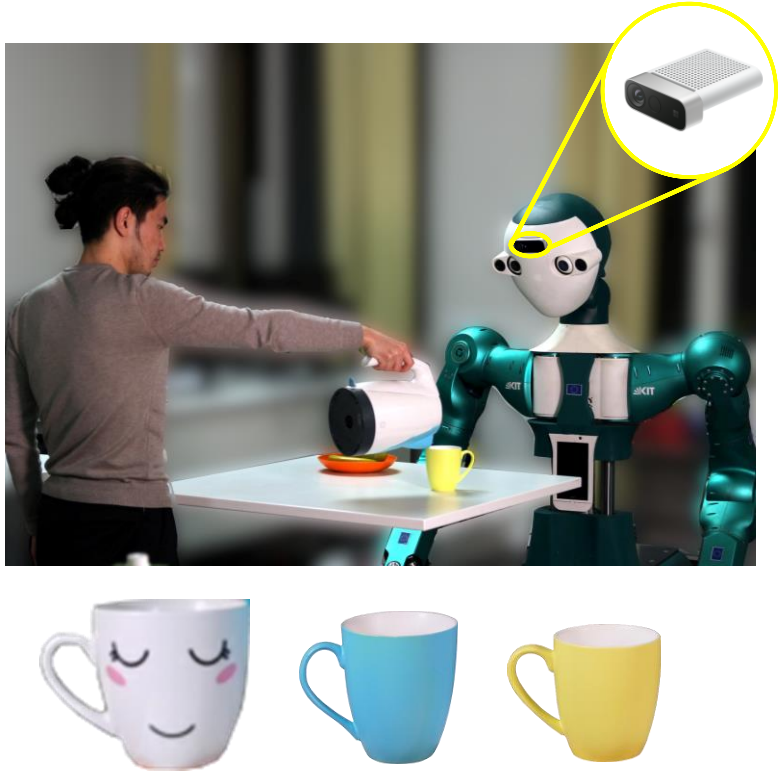
Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N. and Asfour, T., *Unifying Representations and Large-Scale Whole-Body Motion Databases for Studying Human Motion*, IEEE Transactions on Robotics (T-RO), 2016



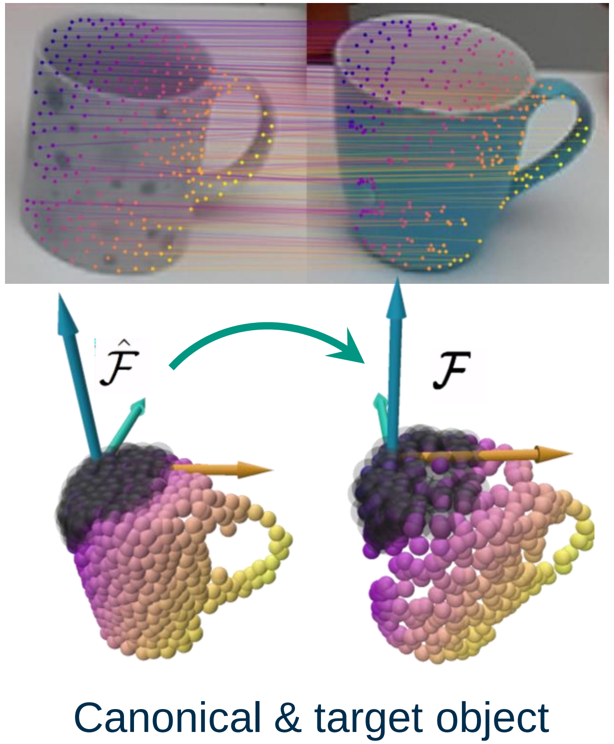
no action
bimanual manipulation
uncoordinated
coordinated
unimanual
bimanual
loosely coupled
tightly coupled
left
right
asymmetric
symmetric
left dominant
right dominant
Motion Capture Data
KIT Bimanual Manipulation Dataset (Krebs and Meixner et al., 2021)
Motion recordings of household actions
Representation using the Master Motor Map (MMM)
Karlsruhe Institute of Technology

Learning Keypoint-based Geometric Constraints

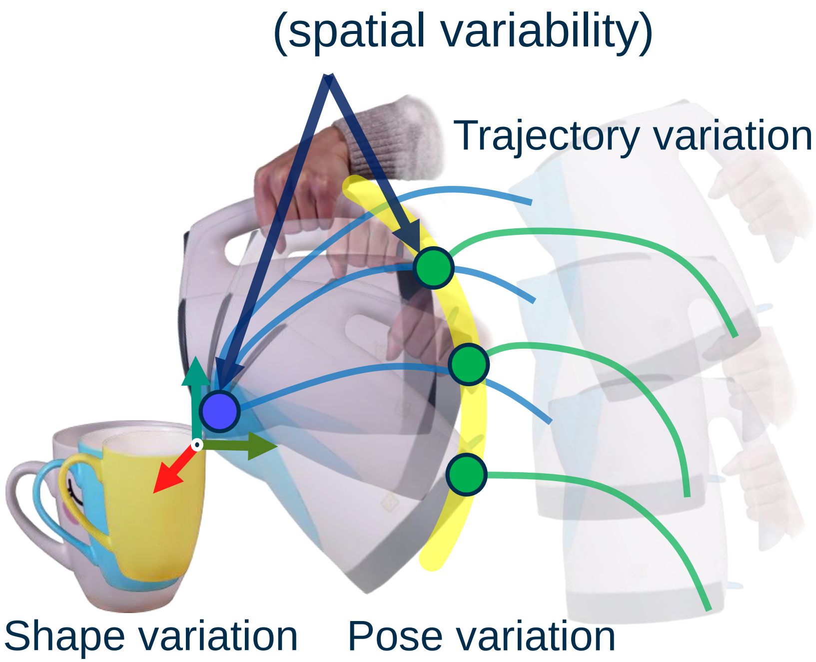
Observational Learning



Generalization



Statistical Evidence



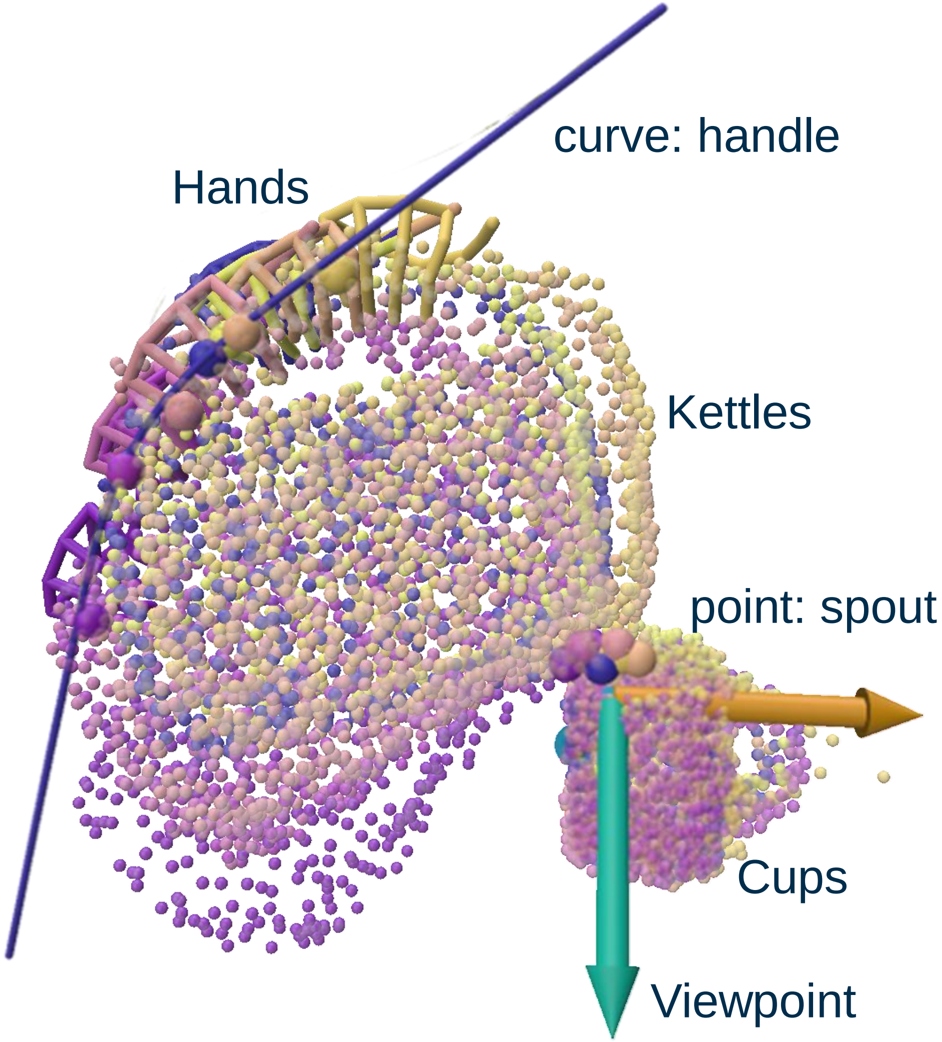
Jianfeng Gao, Zhi Tao, Noemie Jaquier, and Tamim Asfour, K-VIL: Keypoints-based visual imitation learning, Transactions on Robotics (T-RO), 2023

Learning Keypoint-based Geometric Constraints

Human Demonstrations (≤ 10)

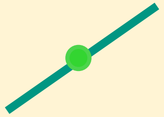
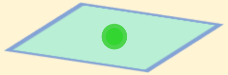
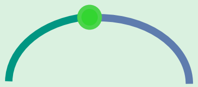
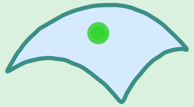
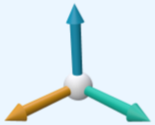
Statistical Evidence

Invariant Task Feature



Joint Extraction of **Sparse** Keypoints, Constraints and Viewpoints

Keypoint-based Geometric Constraints

(Principal) Manifold		Constraints	Statistical Method
	Point	point-to-point (p2p)	
	Line	point-to-line (p2l)	Principal Component Analysis (PCA) Pearson, 1901
	Plane	point-to-plane (p2P)	
	Curve	point-to-curve (p2c)	Principal Manifold Estimation (PME) Meng and Eloyan, 2021
	Surface	point-to-surface (p2S)	
	$\mathbb{R}^3 \times S^3$	6D Pose	Gaussian Mixture Model (GMM) Laird et al., 1977

Evaluation

- Learn **six daily tasks** from different number of demonstrations

Press Button

Fetch Tissue

Hang Hat

Insert Stick

Clean Table

Pour Water

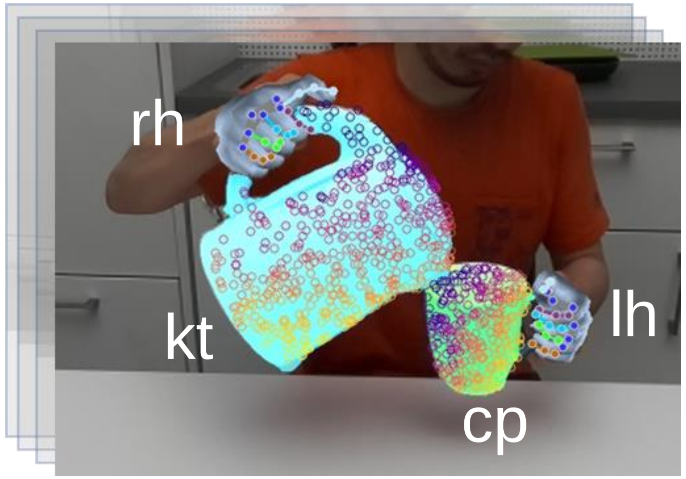
Evaluation

- Extracted constraints converge as more demonstrations are provided ( )

# Demo.	1	3	4	5	8	11
Press Button	3 x p2p	p2p, p2l	p2p	p2p	-	p2p
Fetch Tissue	3 x p2p	p2p, p2l	p2p	p2p	-	p2p
Hang Hat	3 x p2p	p2p	p2p	p2p	-	p2p
Insert Stick	3 x p2p	p2p, p2l	p2p, p2l	p2p, p2l	-	p2p, p2l
Clean Table	3 x p2p	p2l, p2l	p2l, p2l	p2l, p2l	-	p2l, p2l
Pour Water	3 x p2p	p2p, p2l	p2p, p2P	p2p, p2P	p2p, p2c	p2p, p2c

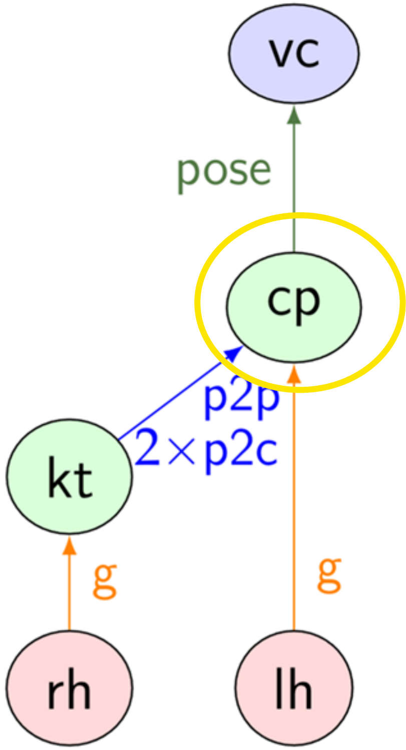
Keypoints-based Visual Imitation Learning

Bimanual Tasks

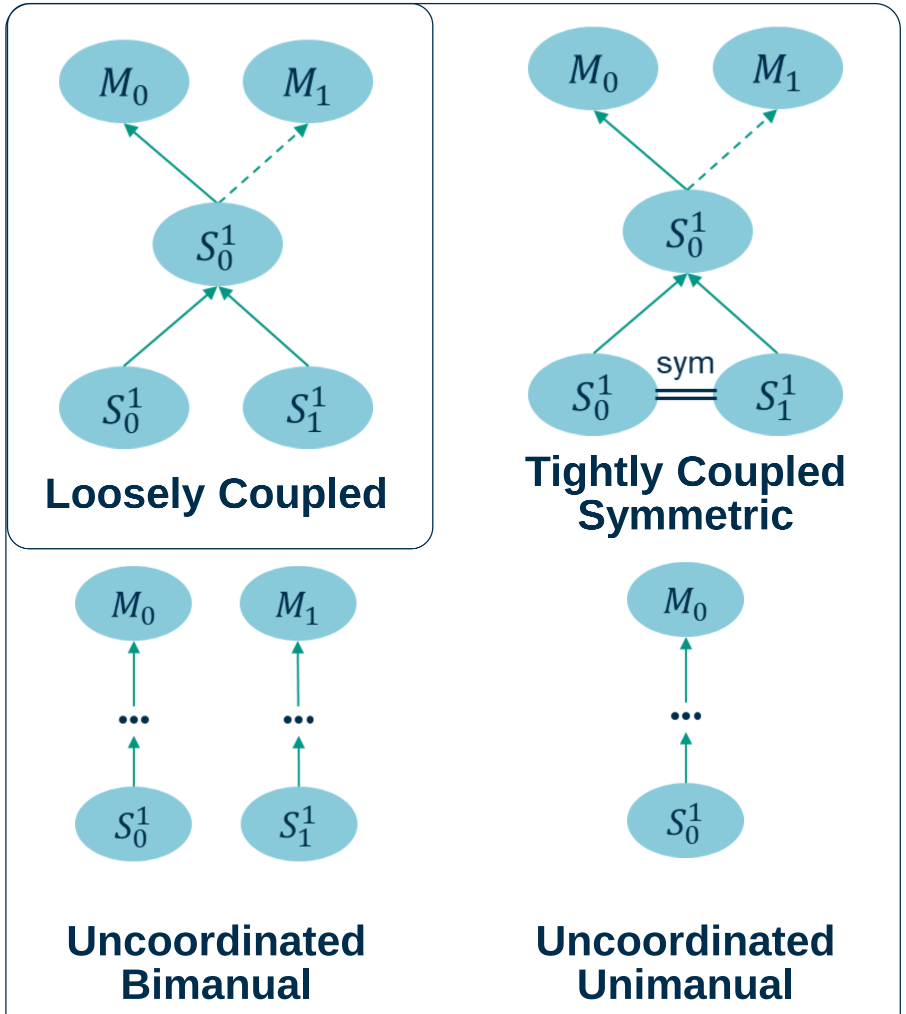


vc: virtual cup

Eight Demonstrations



Object Relation Graph



Coordination Strategy

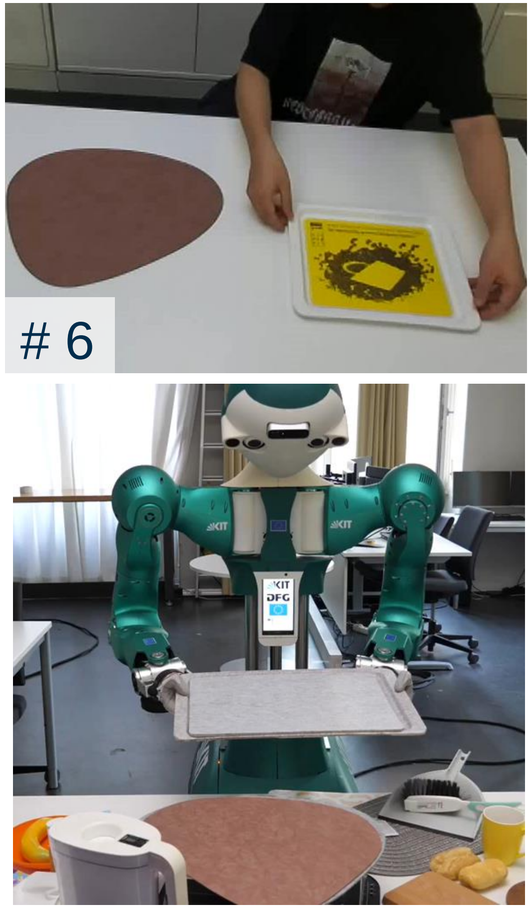
Jianfeng Gao, Xiaoshu Jin, Franziska Krebs, Noemie Jaquier, and Tamim Asfour,
Bi-KVIL: Keypoints-based Visual Imitation Learning of Bimanual Manipulation Tasks, ICRA 2024

Keypoints-based Visual Imitation Learning

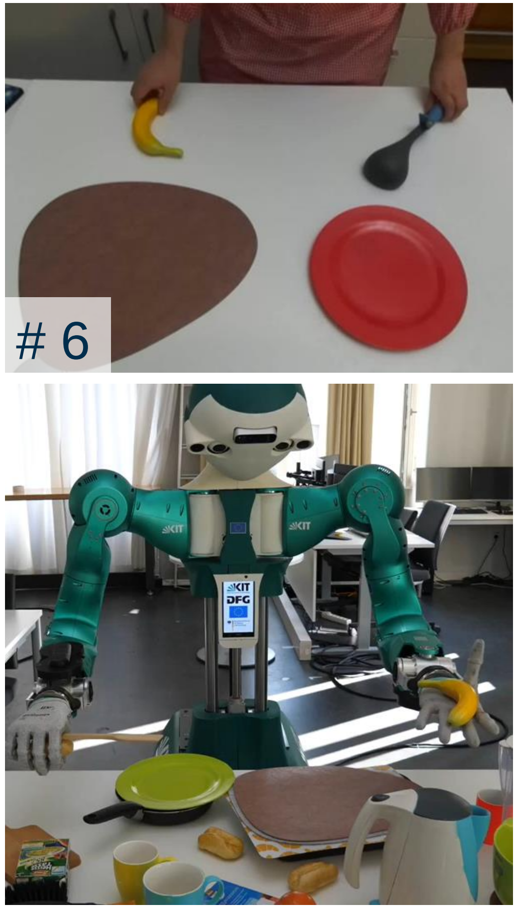
Bimanual Tasks

Coordinated
Loosely Coupled

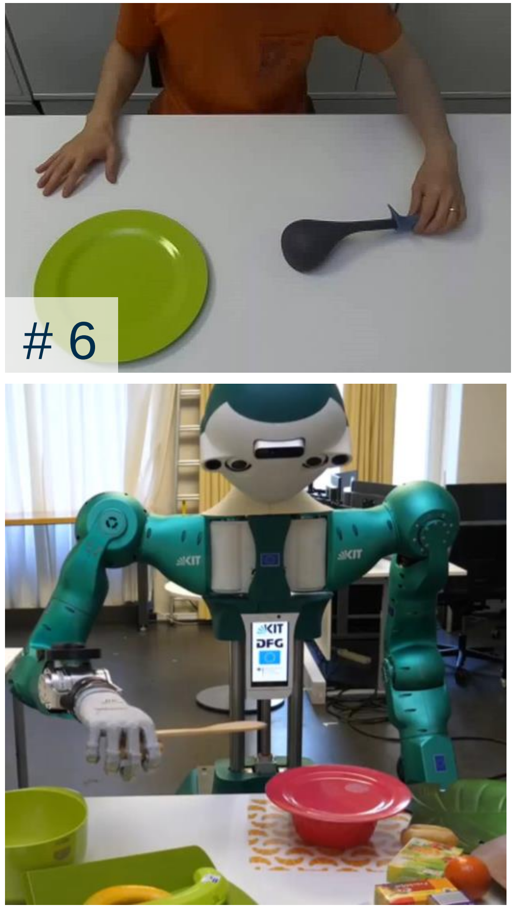
Coordinated
Tightly Coupled



Uncoordinated
Bimanual

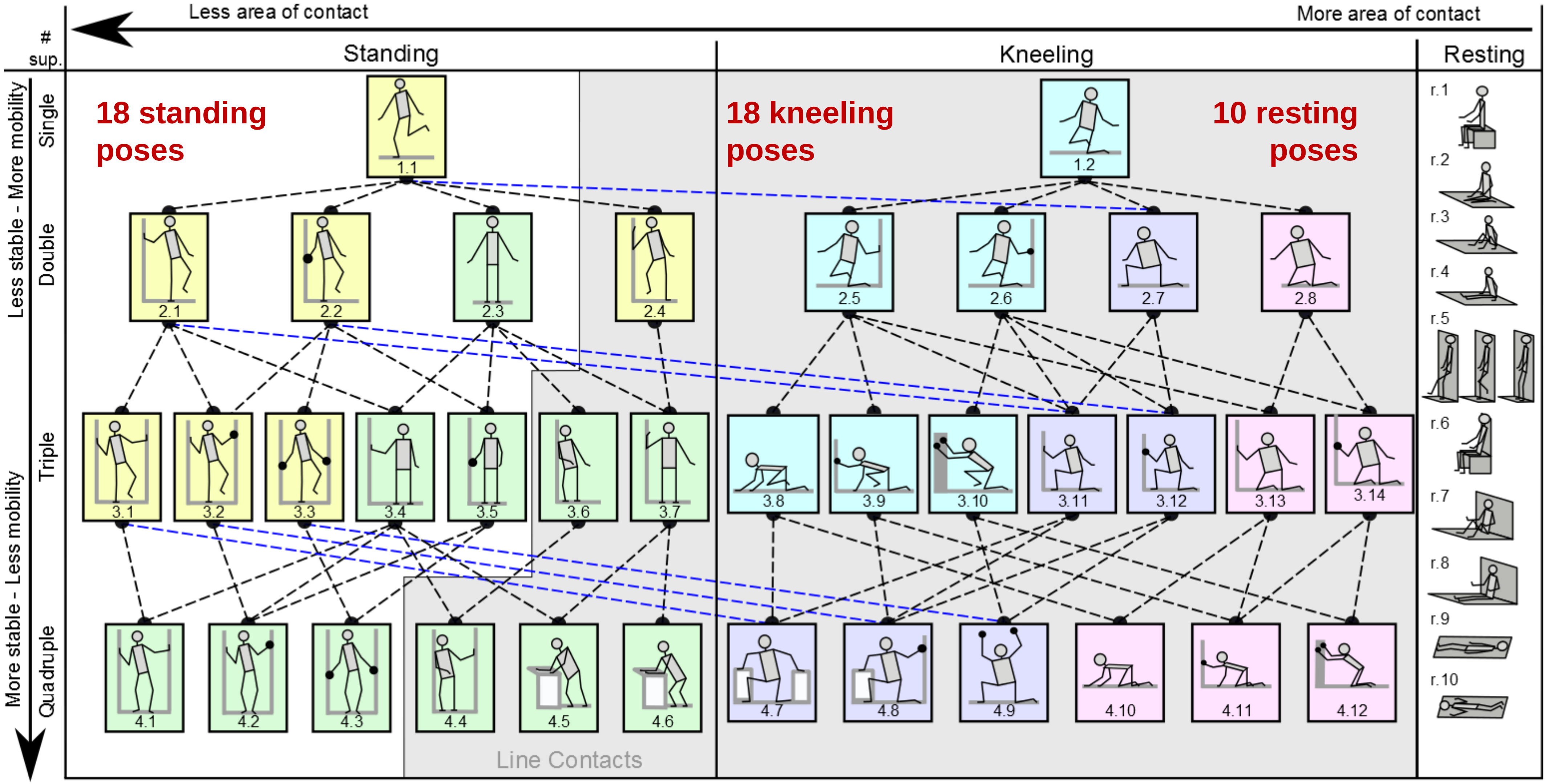


Uncoordinated
Unimanual



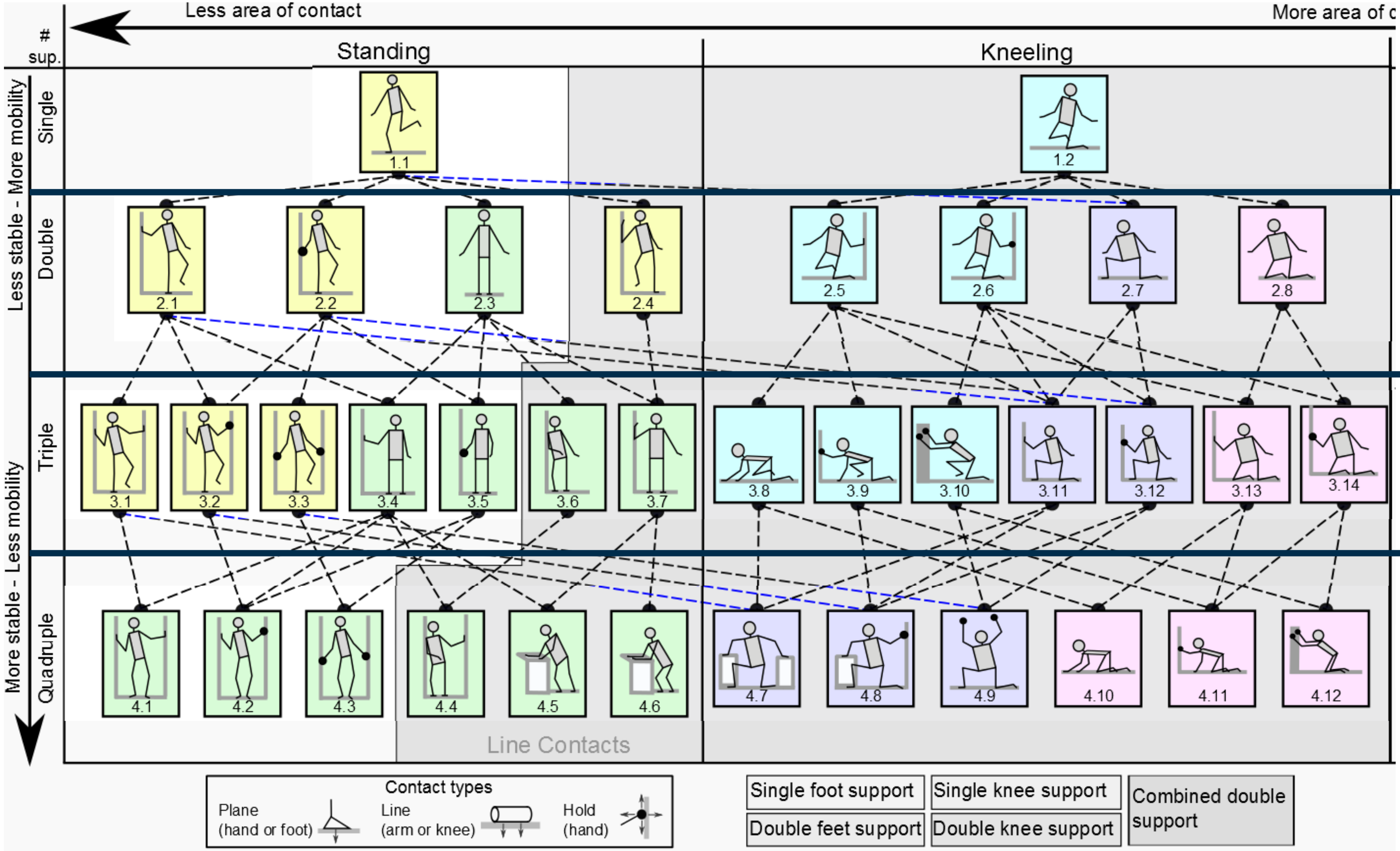
Whole-Body Poses Taxonomy

Total: 46 classes



Julia Borràs Sole, Christian Mandery and Tamim Asfour, A Whole-Body Support Pose Taxonomy for Multi-Contact Humanoid Robot Motions, Science Robotics, 2017

Whole-Body Poses Taxonomy



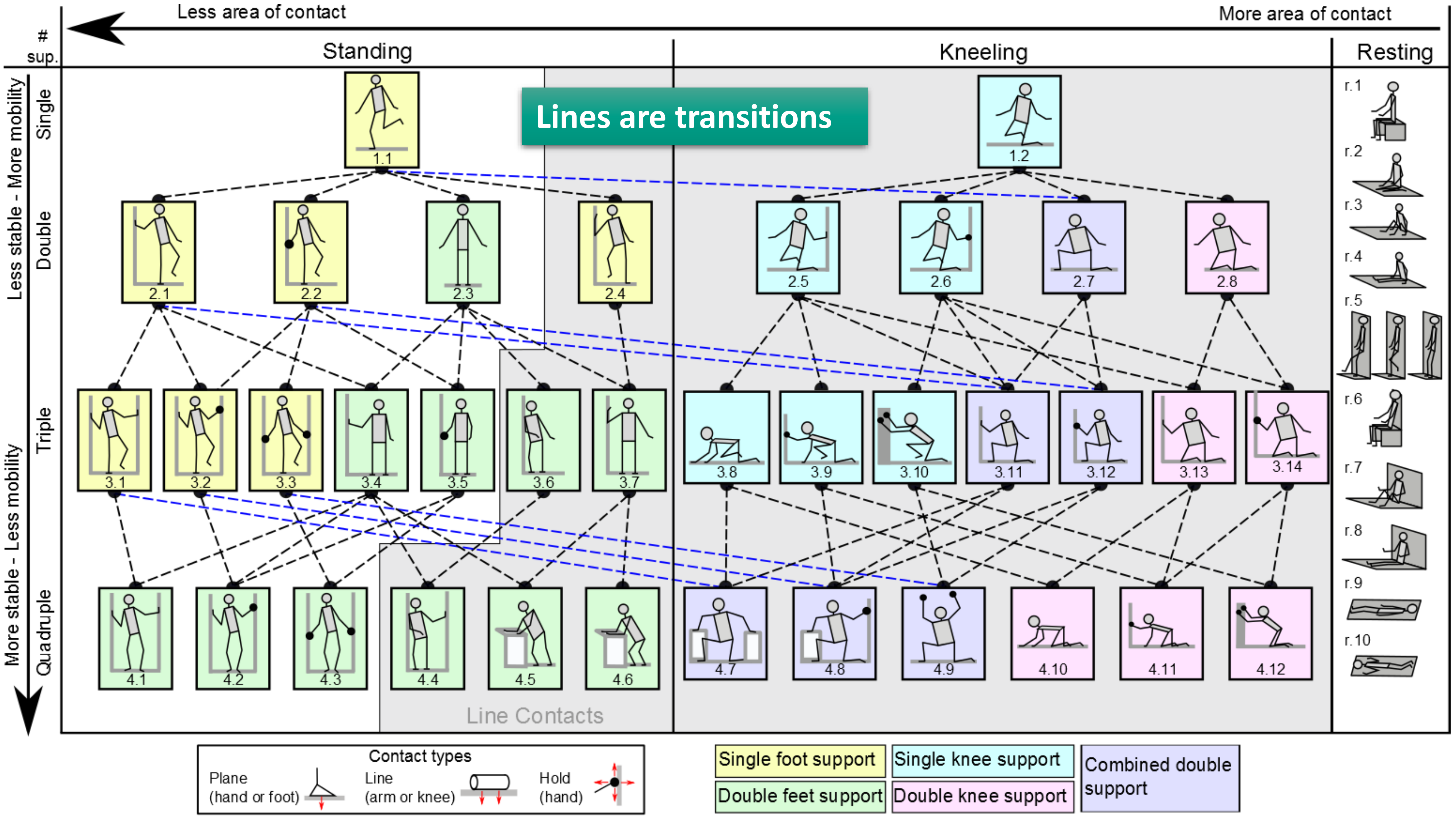
1 contact

2 contacts

3 contacts

4 contacts

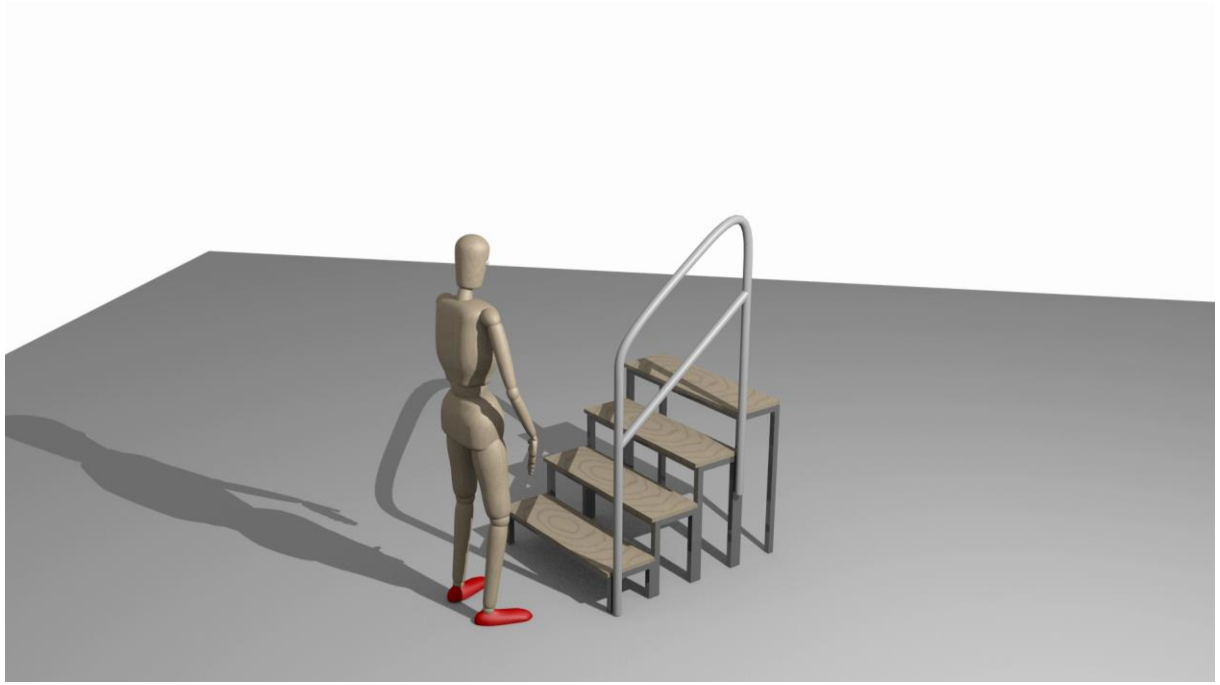
Whole-Body Poses Taxonomy



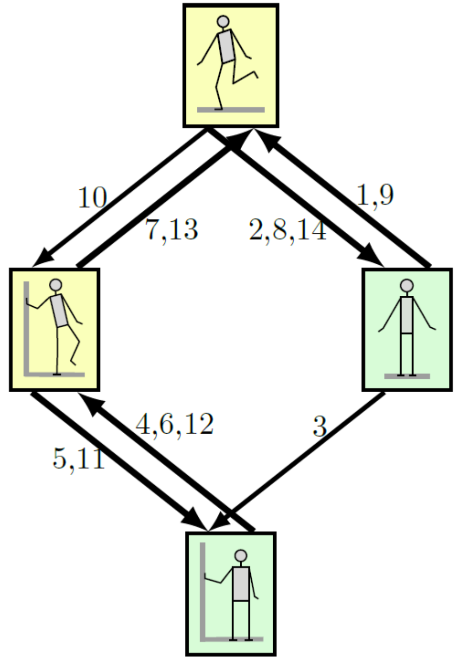
Whole-Body Poses Taxonomy

Taxonomy-Driven Human Motion Analysis

Going upstairs with a handle: Detection of **support contacts** highlighted in **red**

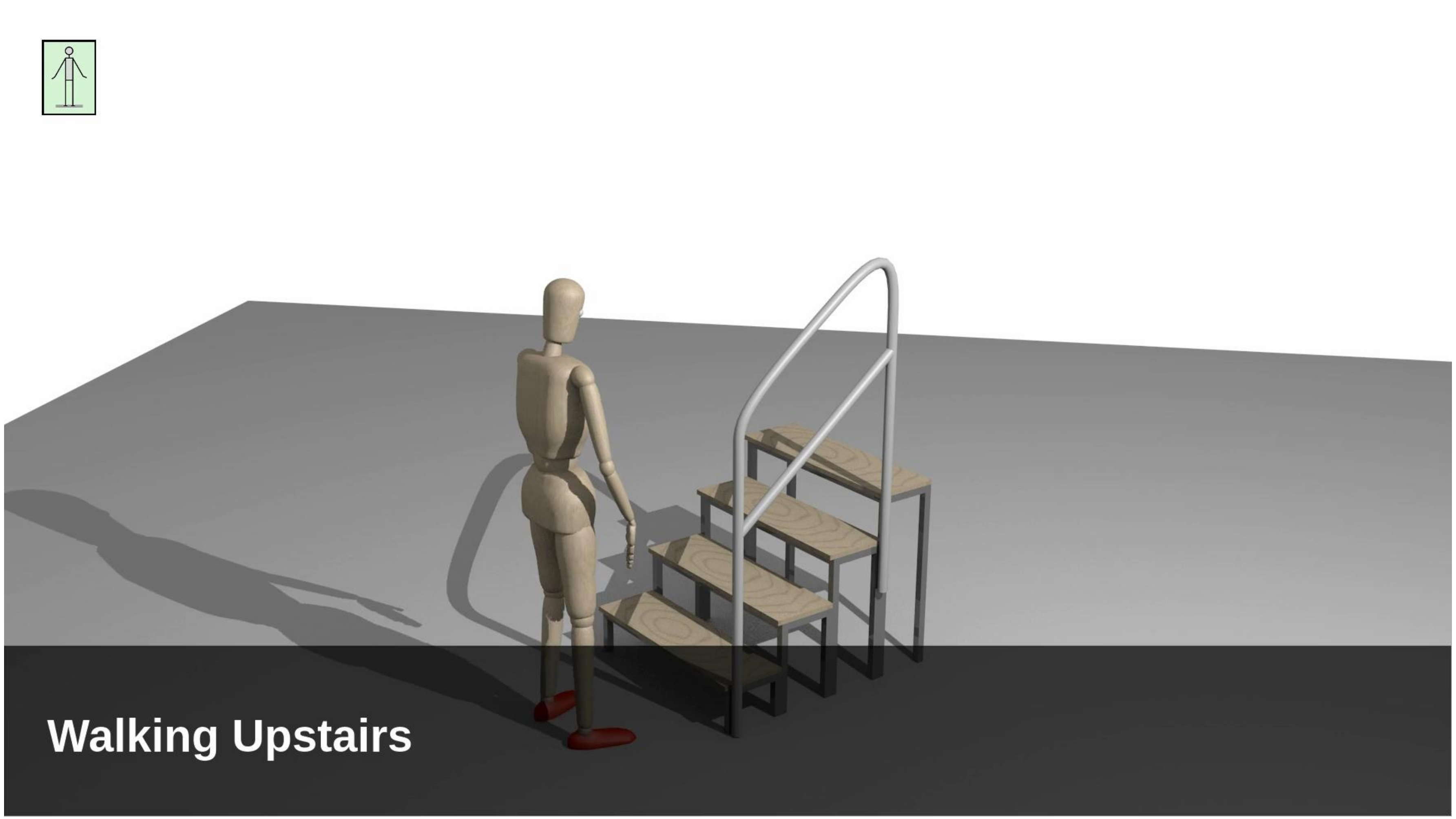


Subject swings left leg with a *right foot – right hand* support pose



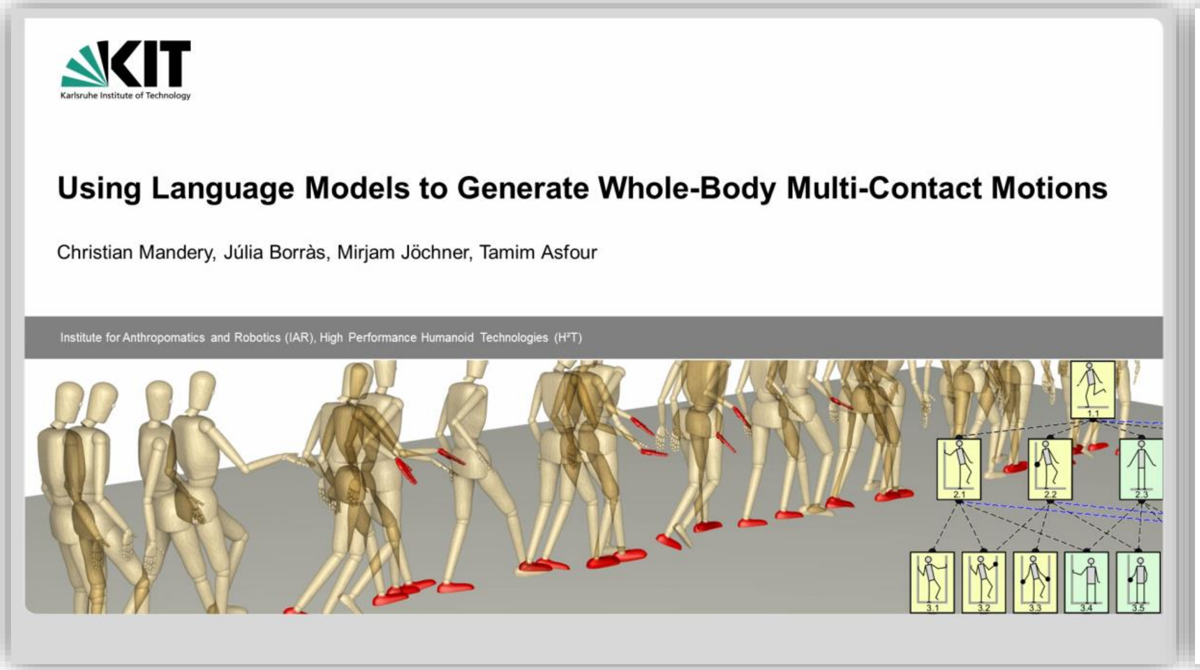
Generated graph of transitions

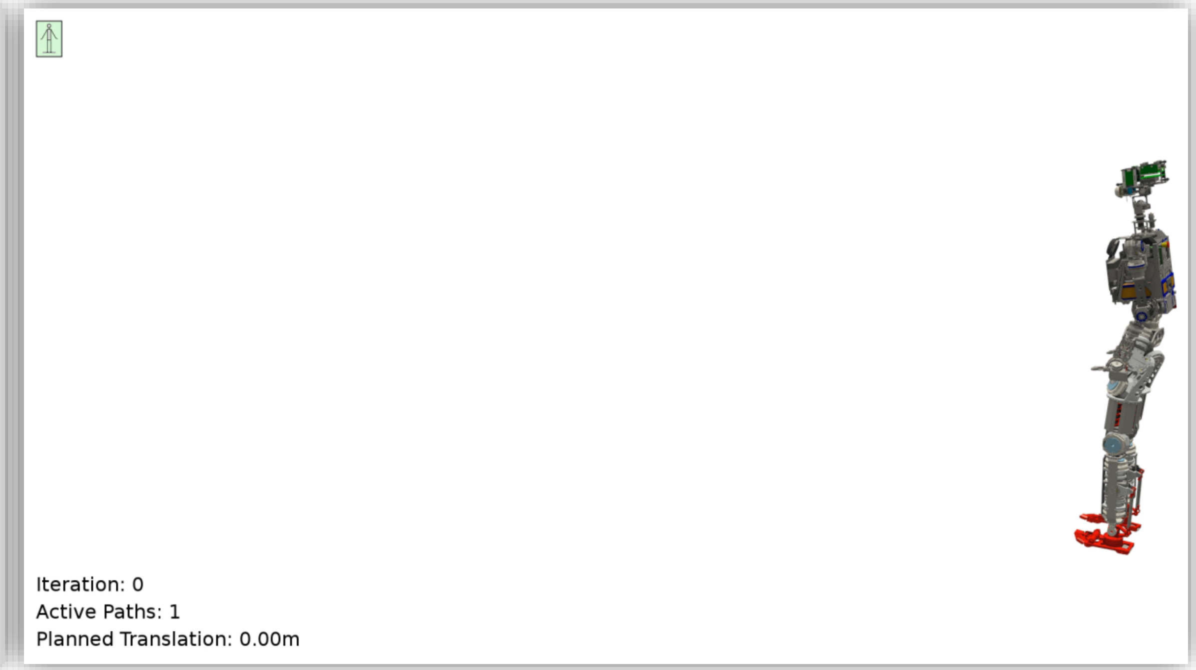
Whole-Body Poses Taxonomy
Semantics of Human Motion



Motion (Sentence) as Sequences of Whole-Body Poses (Words)

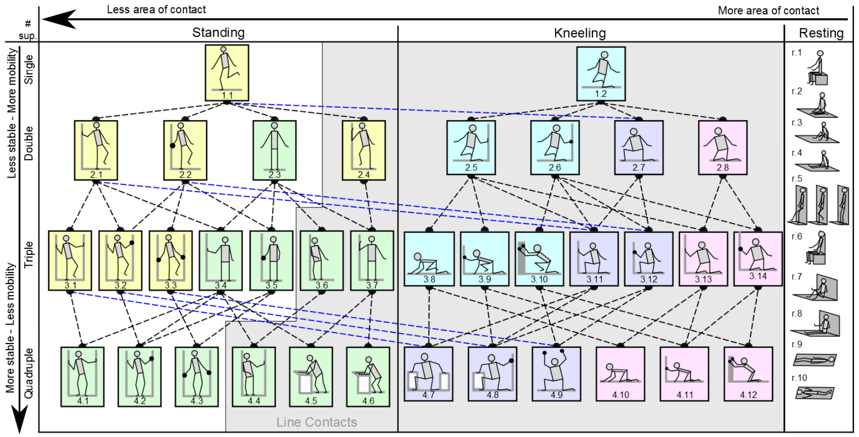
Statistical learning of conditional transition probabilities between whole-body poses (*n-gram language model* learned from segmented human motion)



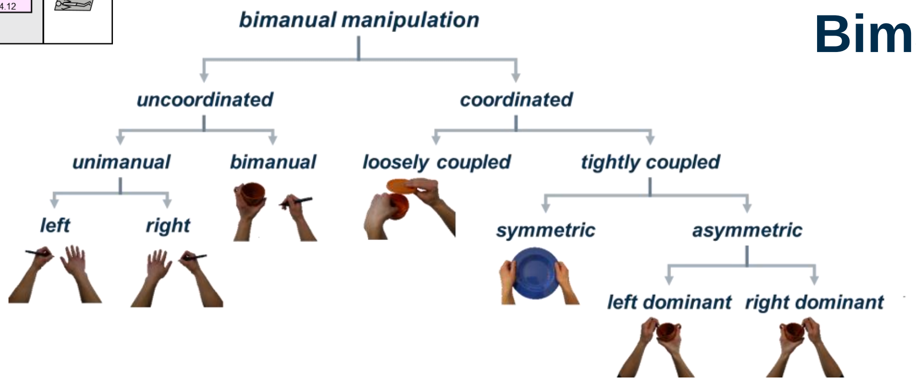


Mandery, C., Borràs, J., Jöchner, M. and Asfour, T., *Using Language Models to Generate Whole-Body Multi-Contact Motions*, IROS 2016

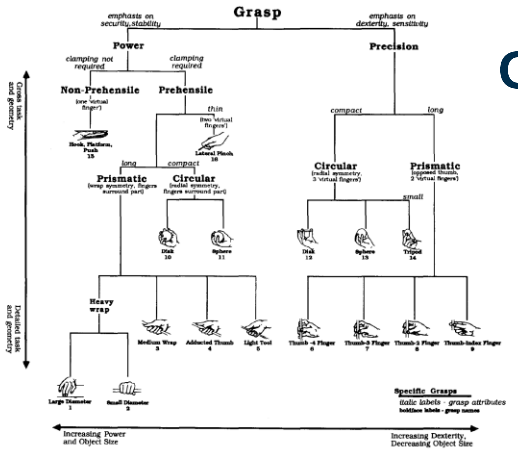
Concatenation of Taxonomies ?



Loco-manipulation



Bimanual



Grasping

Manipulation – Constraints & Taxonomies

- Taxonomies
  - structured frameworks for understanding the **space of possible constraints**
  - powerful **inductive bias for learning**
  - enable transfer learning and generalization through cross-task analogy

- However, taxonomies
  - are rigid and may not capture the continuous and force-based nature of robotic manipulation tasks
  - cannot easily handle edge-cases (e.g. **manipulation of deformable objects**)
  - risk oversimplify complex problems
  - may ignore context dependencies
  - can over-constrain learning

Manipulation – Constraints & Taxonomies

- Taxonomies are powerful tools but not the complete solution

- Provide initial structure to make learning tractable while remaining flexible enough to capture the complexity of real-world manipulation tasks.

- (Many) open questions
  - How can constraint manifolds guide learning algorithms toward efficient constraint representations?
  - How can manipulation tasks be described as sequences of coordination patterns defined by taxonomies?
  - How to integrate dynamics into taxonomies, especially force?

Manipulation of Deformables

KIT in ROMANDIC

High Performance Humanoid Technologies

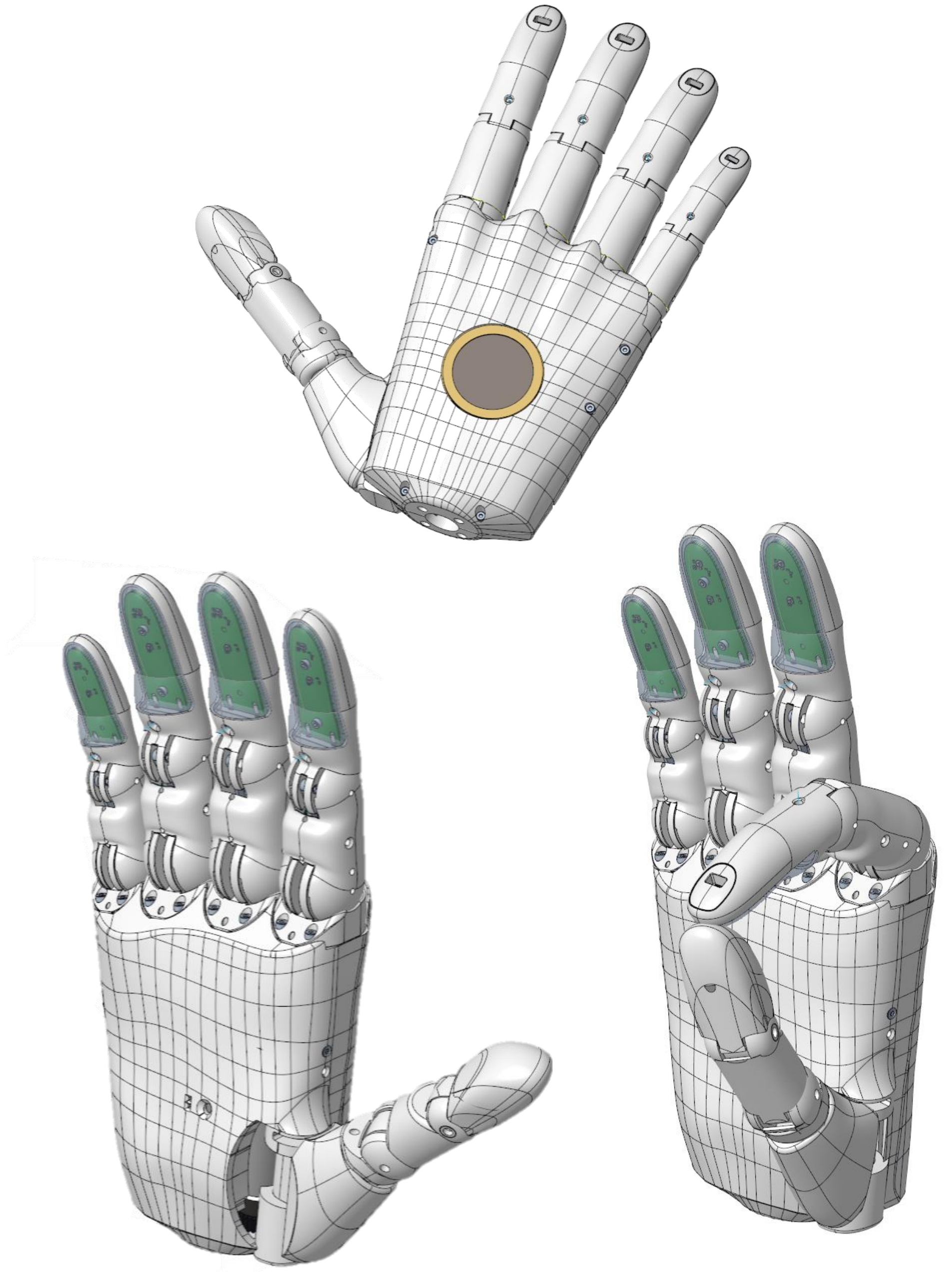
Research Topics

Bimanual manipulation of deformable objects with humanoid robots

- Dexterous hands/grippers for manipulation of deformable objects
- Learning skills for manipulation from demonstration

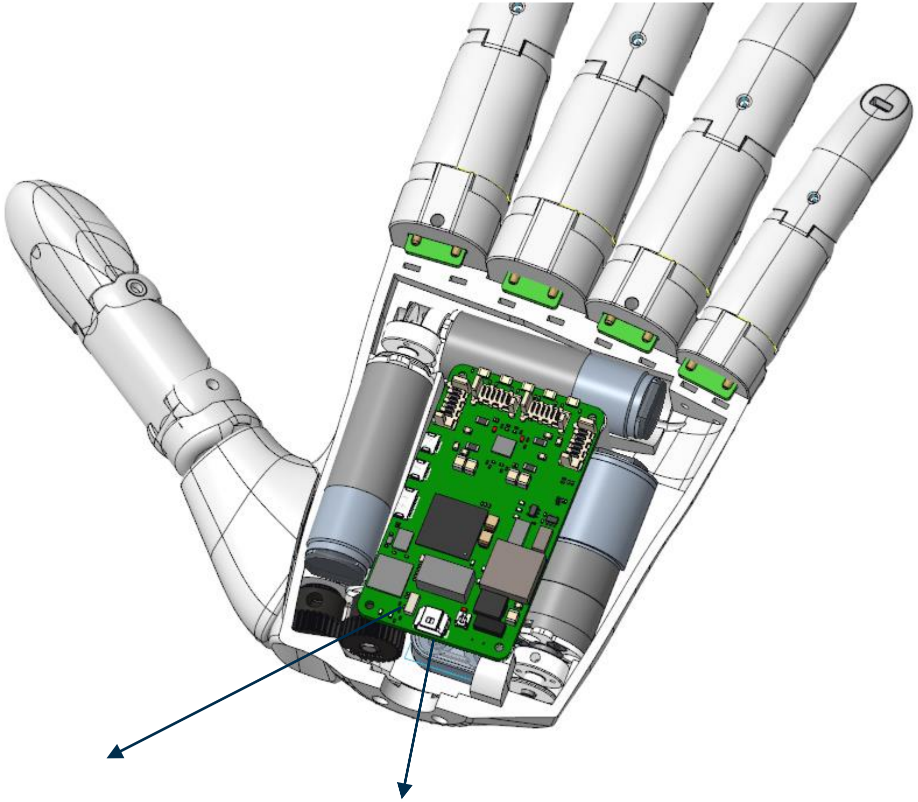
ROMANDIC Hand

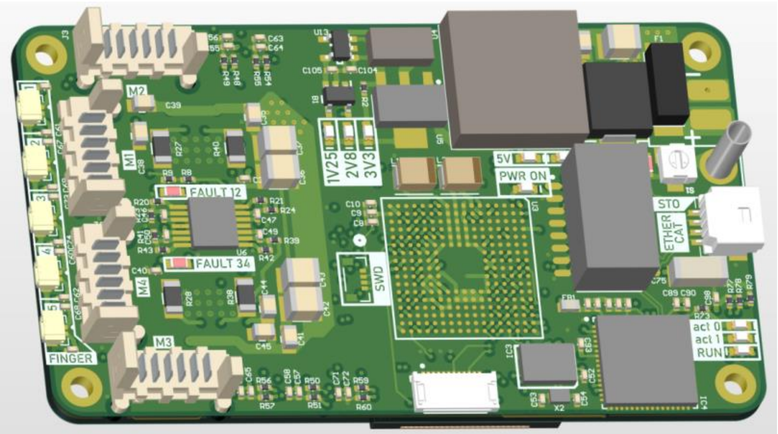
- 5 Finger Humanoid Hand
- 4 Degree of Freedom
  - 2 DoF Thumb (Flexion + Cicumduction)
  - 1 DoF Index
  - 1 DoF for Middle, Ring, Little fingers
- Back-driveable
- Sensorized Fingers



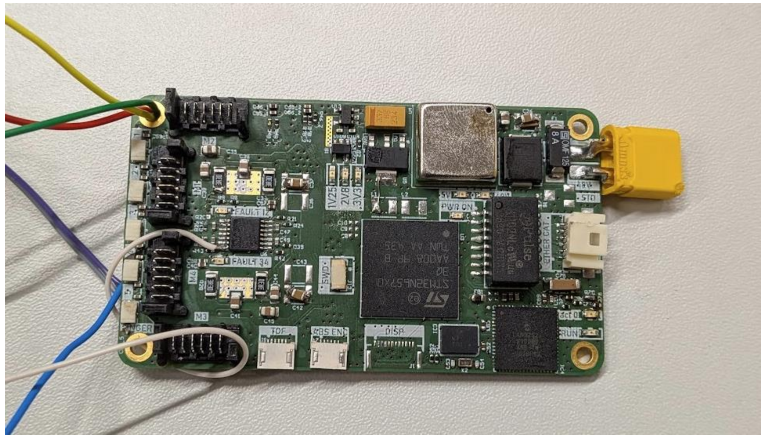
ROMANDIC Hand | Electronics

- Hand-internal control electronics
  - Finished first prototype, only partially working
  - Currently working on 2nd version
- Features
  - 48V input, EtherCAT interface
  - 4 Motor drivers (48V @2,5A)
  - 5 Finger Interfaces
  - Display
  - **Camera** for In-Hand Vision
  - **ToF Sensor**
  - Bluetooth
  - Processor: Cortex-M55 @800MHz, with 600 GOPS neural accelerator for image processing





PCB Layout

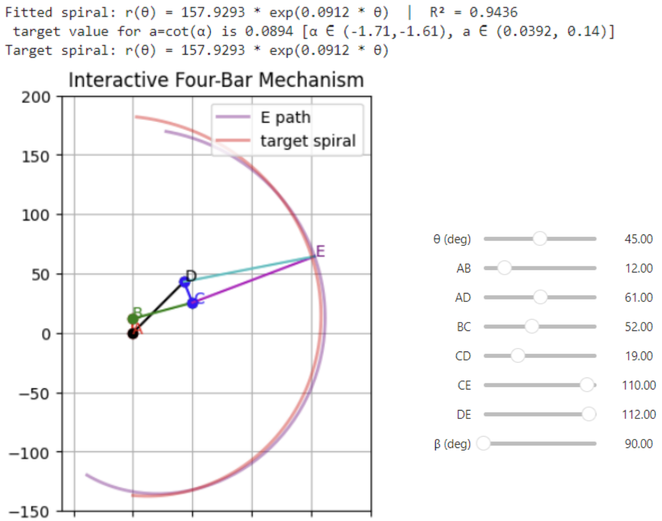


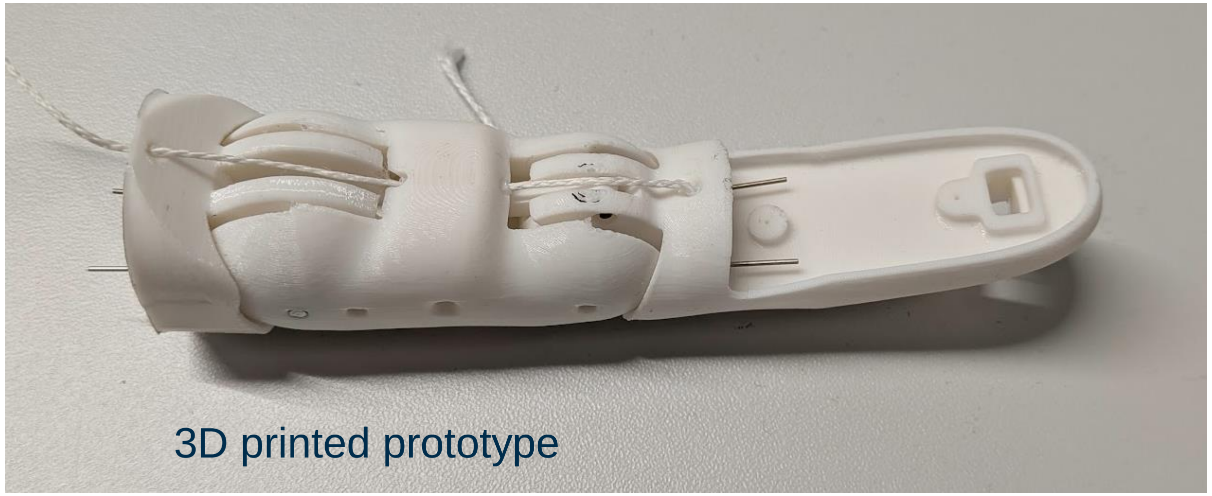
Prototype

ROMANDIC Hand | Fingers

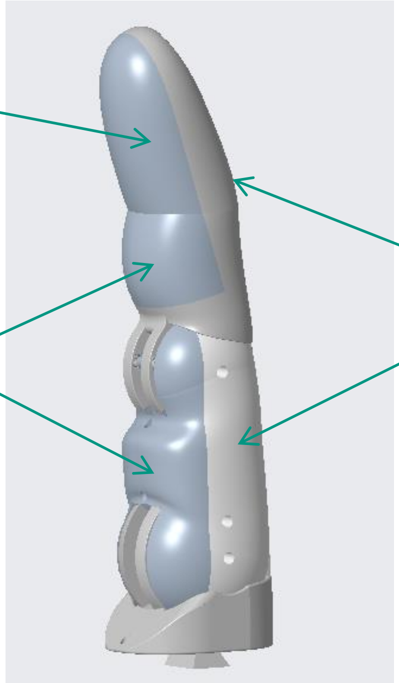
- **Trajectory optimzation**
  - Finger Trajectories of 4-bar mechanism, follows human fingertip trajectory

- **CAD design**
  - First prototype built





3D printed prototype

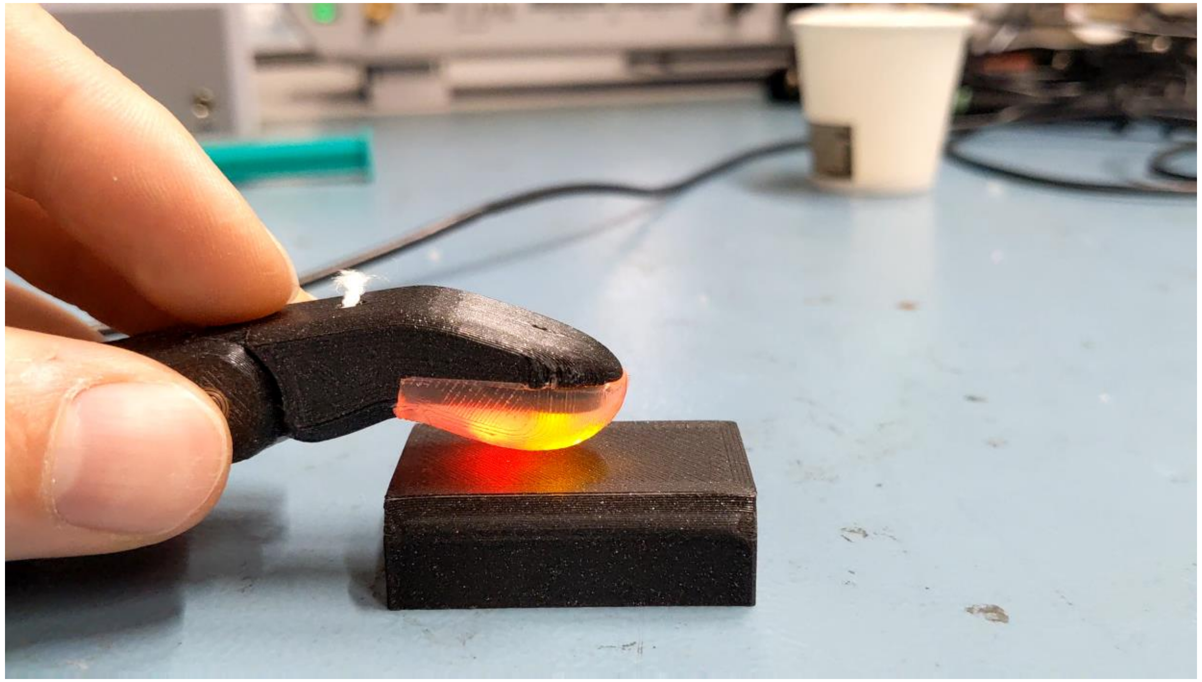


ROMANDIC Hand | Tactile Sensors

Tactile Sensors: Evaluation of different sensor modalities:

- Slip
- Pressure (2 Taxel)

Slip detection

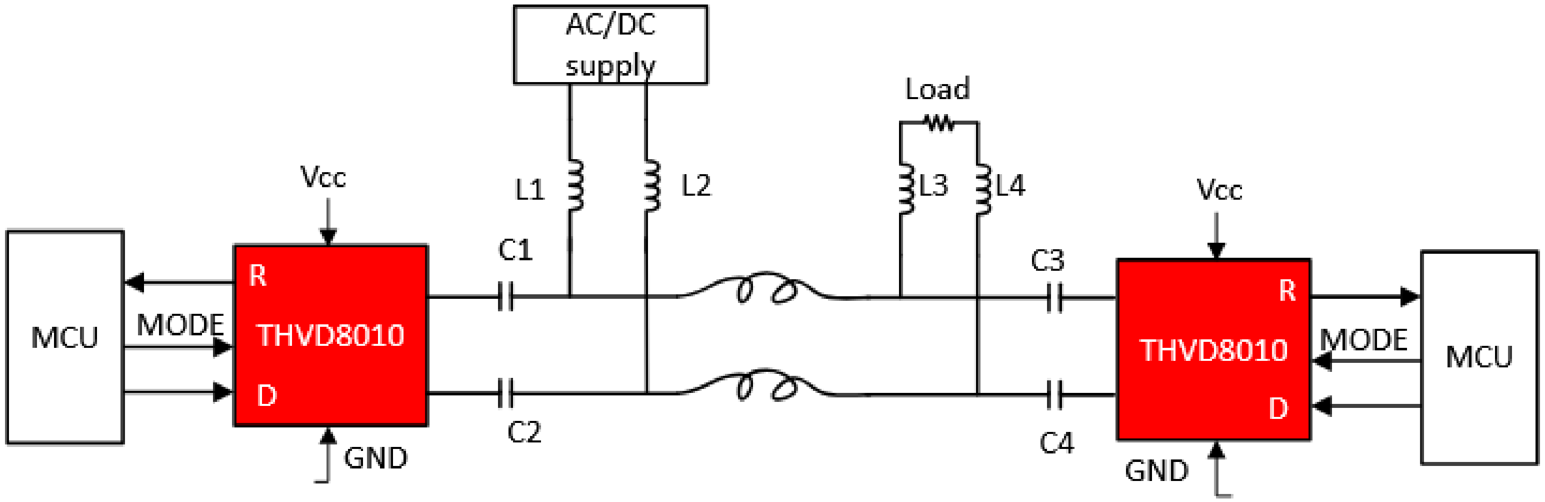


Pressure detection

ROMANDIC Hand | Finger Interfaces

- **Use mechanical torsion springs to transmit power & data**
  → No cable inside finger required, increased robustness and repairability

- Powerline-transceiver for signal/power coupling:



2 pin finger sensor interface

Soft Hands at H²T

Sensorized Soft Hand

Finger-Vision Soft Hand

Underactuated humanoid hands with **three** degrees of freedom

Soft material fingers

Sensors:

- **KIT Finger-Vision Soft Hand:** camera in each fingertip
- **KIT Sensorized Soft Hand:** tactile sensors in each fingertip

Weiner, P., Hundhausen, F., Grimm, R. and Asfour, T., "Detecting Grasp Phases and Adaption of Object-Hand Interaction Forces of a Soft Humanoid Hand Based on Tactile Feedback", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021

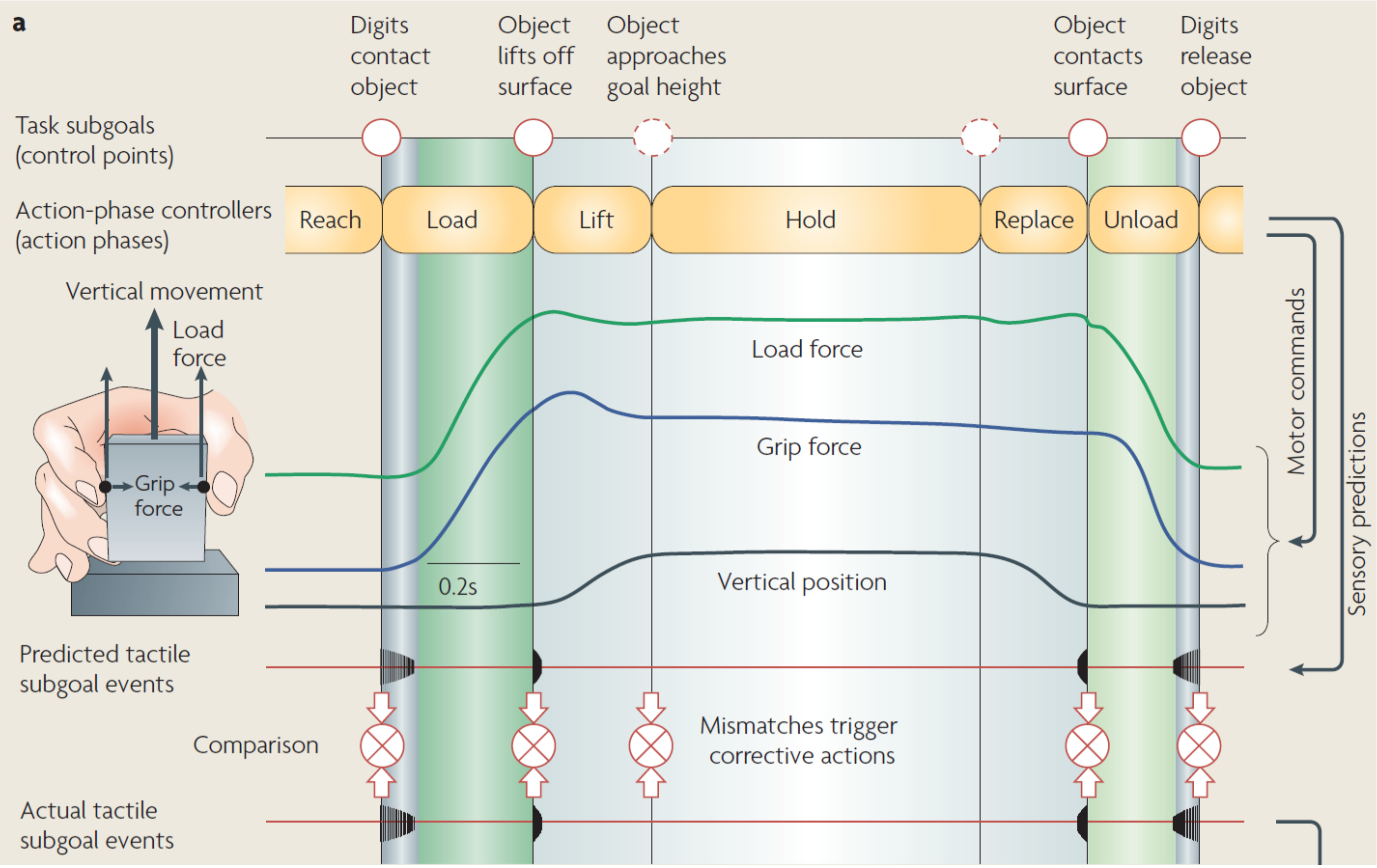
Hundhausen, F., Grimm, R., Stieber, L. and Asfour, T., Fast Reactive Grasping with In-Finger Vision and In-Hand FPGA-accelerated CNNs, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021

Grasp Phases in Humans

Humans divide the grasping process into distinct action phases: reach, load, lift, hold, replace and unload

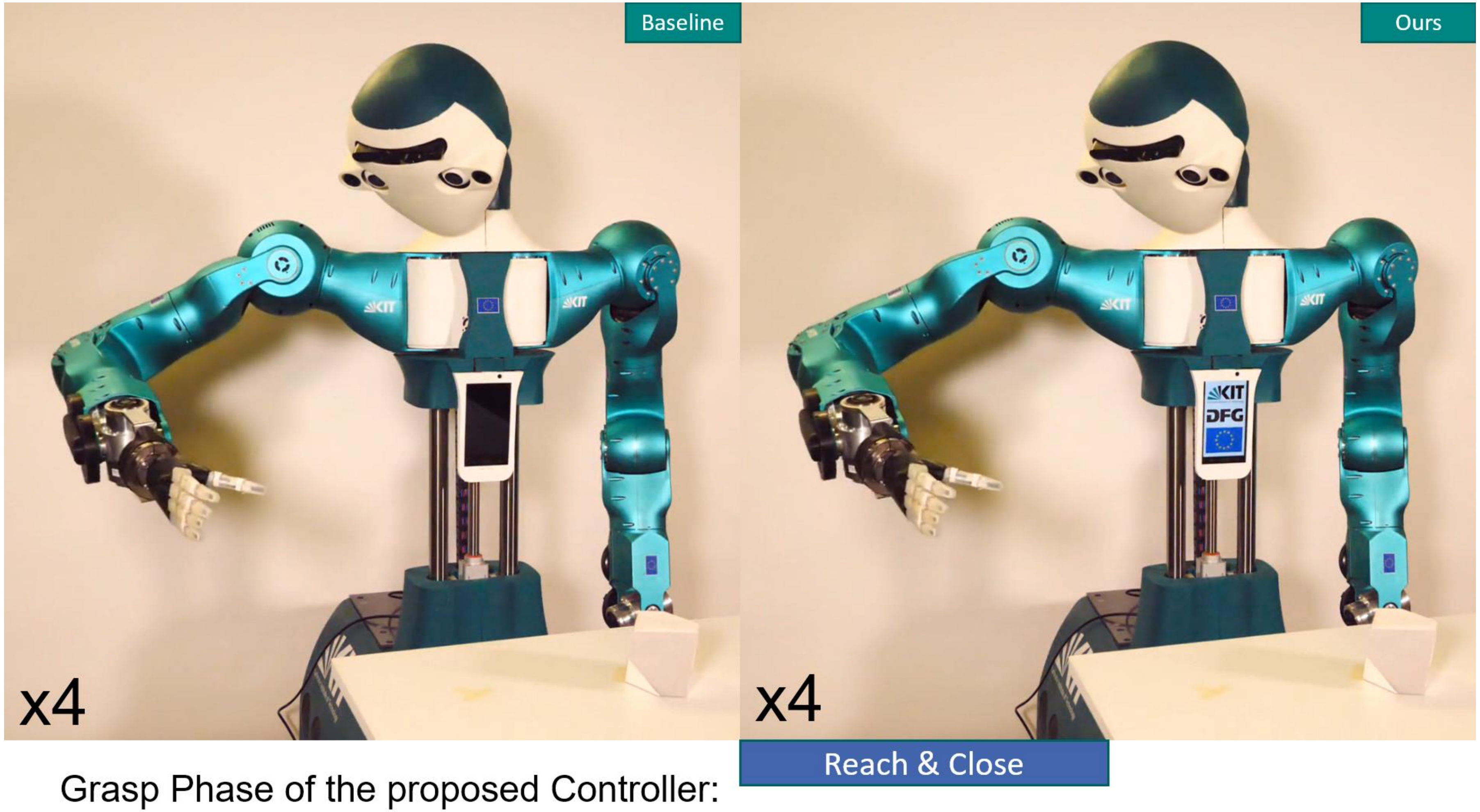
Each phase is triggered by sensory events dependent on the phase

Each phase is associated with specific control goals



Johansson, R., Flanagan, J. Coding and use of tactile signals from the fingertips in object manipulation tasks. Nat Rev Neurosci 10, 345–359 (2009)

Grasping with Soft Hands and Tactile Feedback



Weiner, P., Hundhausen, F., Grimm, R. and Asfour, T., *Detecting Grasp Phases and Adaption of Object-Hand Interaction Forces of a Soft Humanoid Hand Based on Tactile Feedback*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021

Grasp Phases Controller – Evaluation

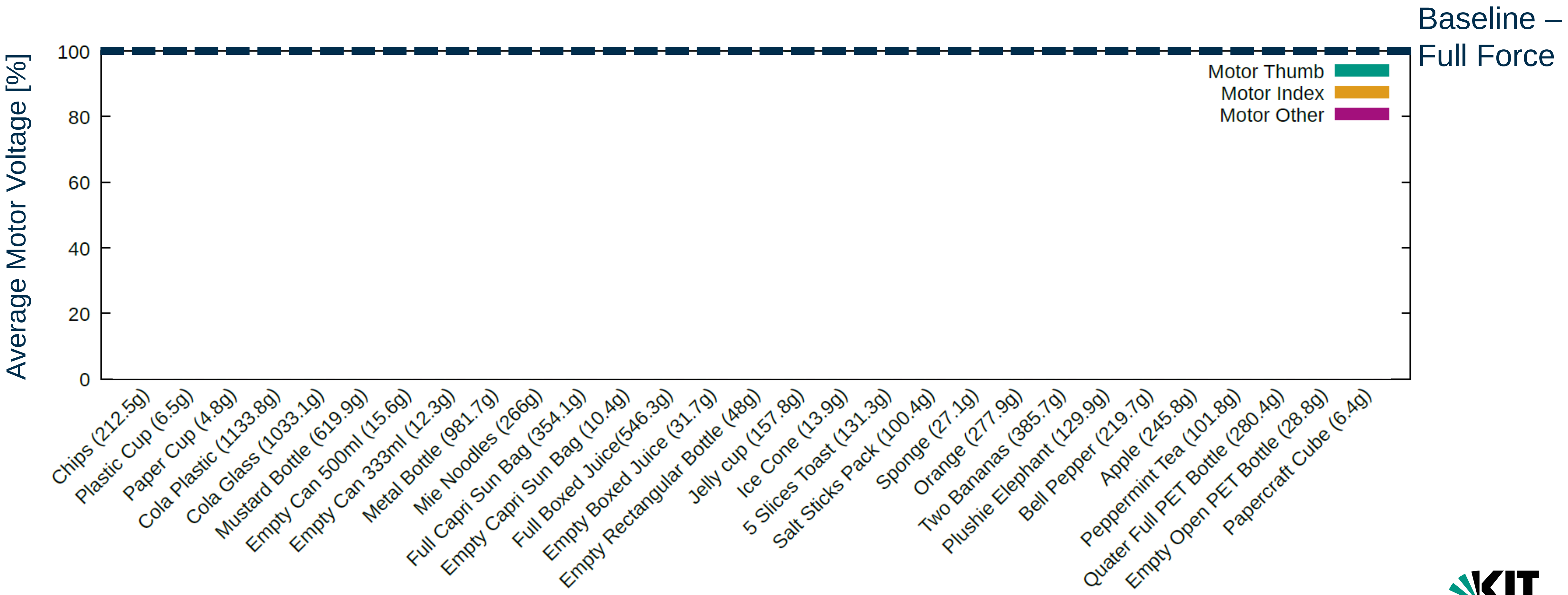
Evaluation with 31 household and food items

Baseline: grasping with full force without tactile feedback

Criteria

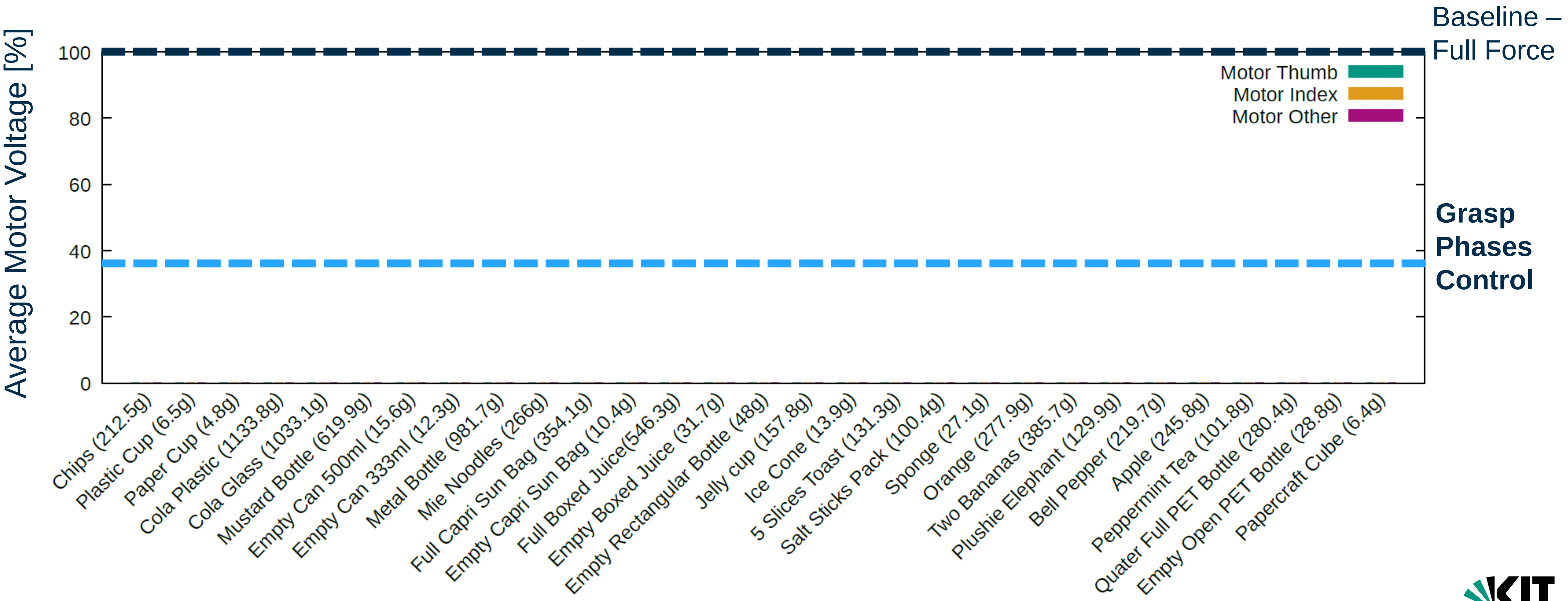
- Grasp force
- Object damage
- Grasp success

Experimental Evaluation – Grasp Force



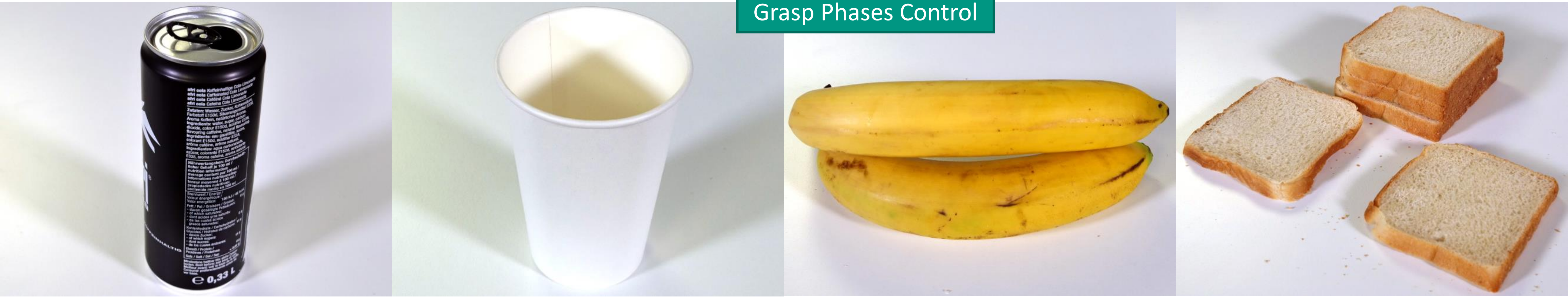
Experimental Evaluation – Grasp Force

Average motor voltage of 35.64% over all trials and all three motors

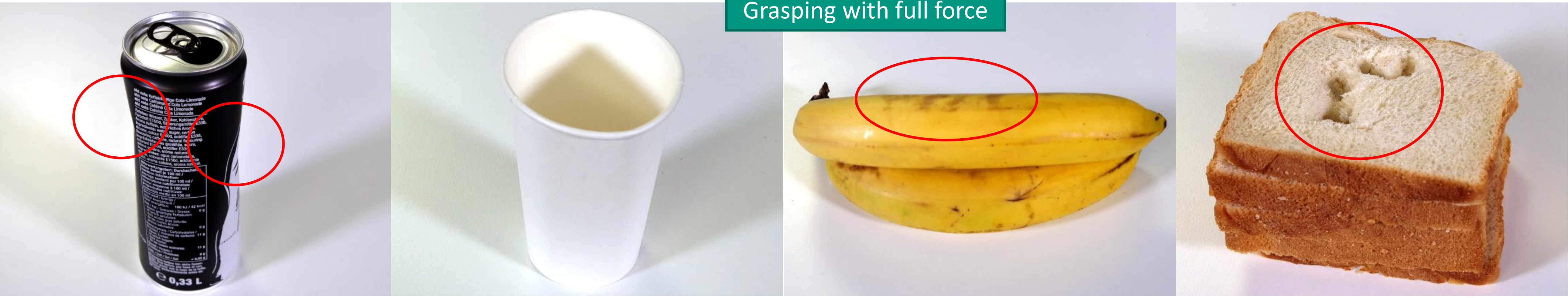


Experimental Evaluation – Object Damage Examples

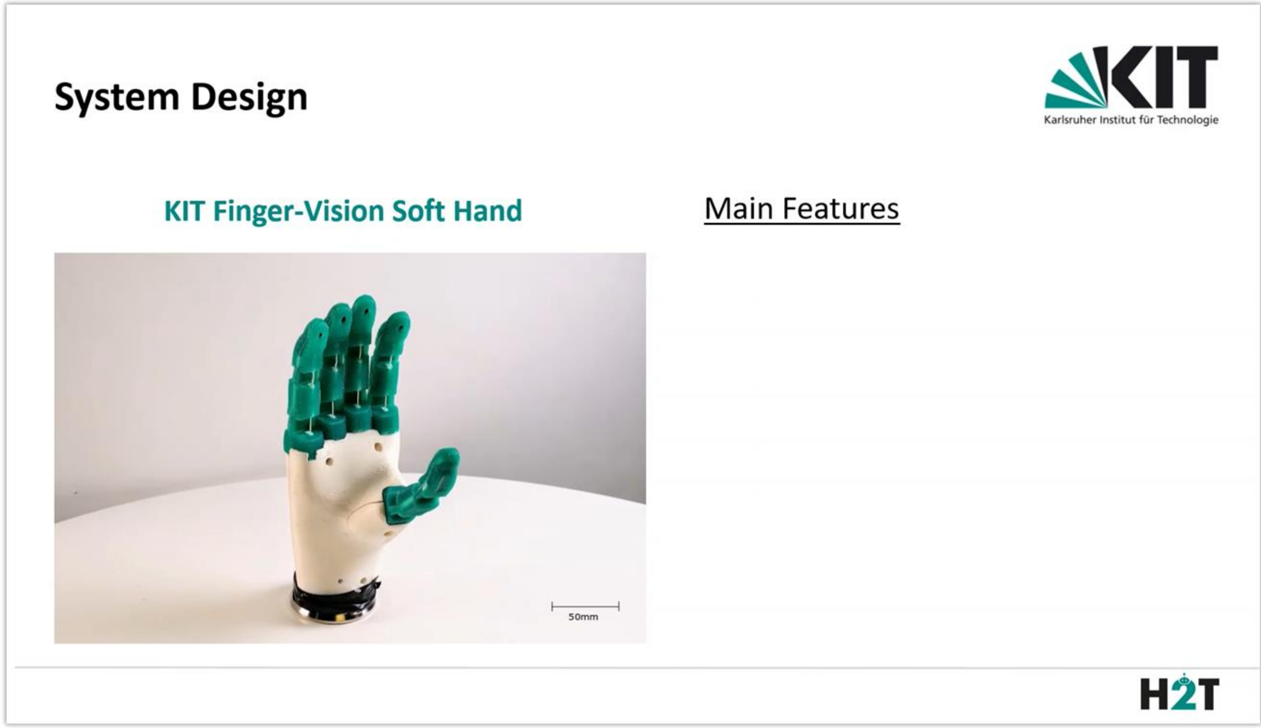
Grasp Phases Control



Grasping with full force



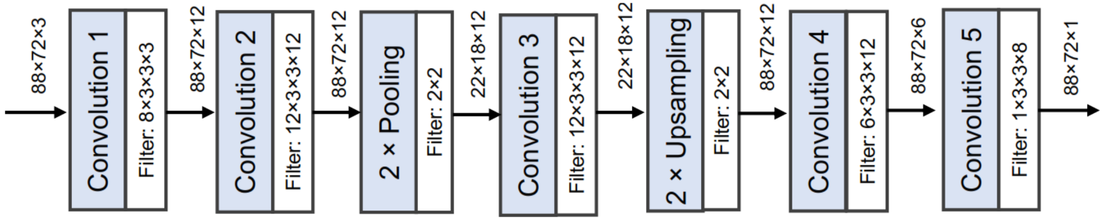
In-Finger Vision Soft Hand



In-Finger-Vision Soft humanoid hand with in-hand accelerated CNN

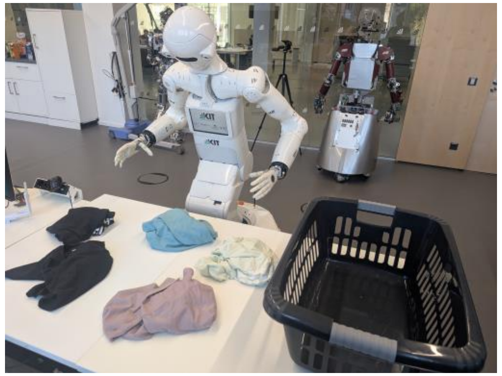
Miniature Camera in each finger-tip

Encoder-Decoder CNN for object-segmentation processed on in-hand FPGA

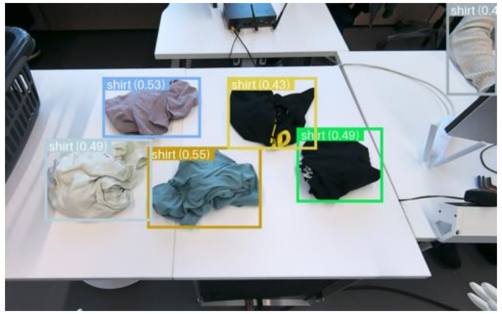


Hundhausen, F., Grimm, R., Stieber, L. and Asfour, T., *Fast Reactive Grasping with In-Finger Vision and In-Hand FPGA-accelerated CNNs*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021

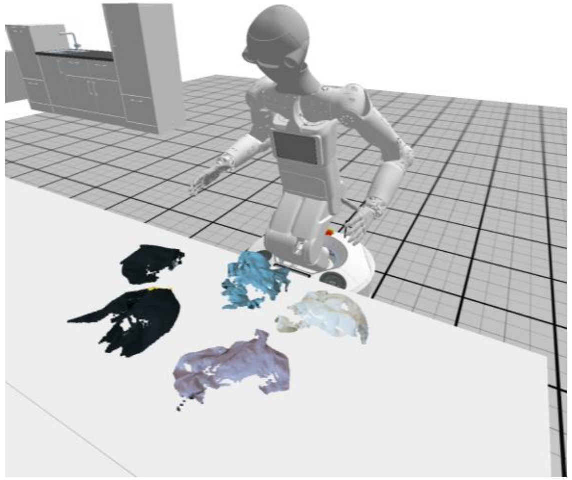
Manipulation of Garments



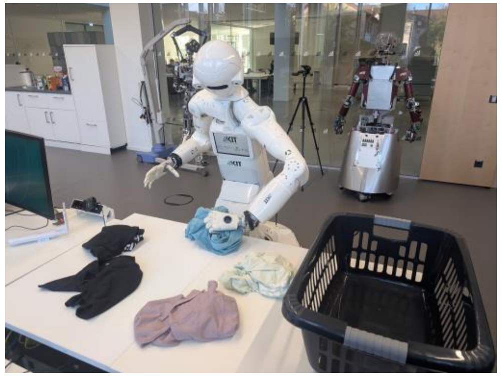
High-fidelity point cloud
via FoundationStereo

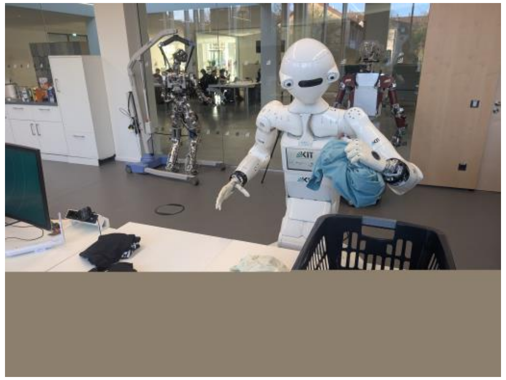


Open-vocabulary
segmentation
via Grounded SAM 2



Crop from
point cloud



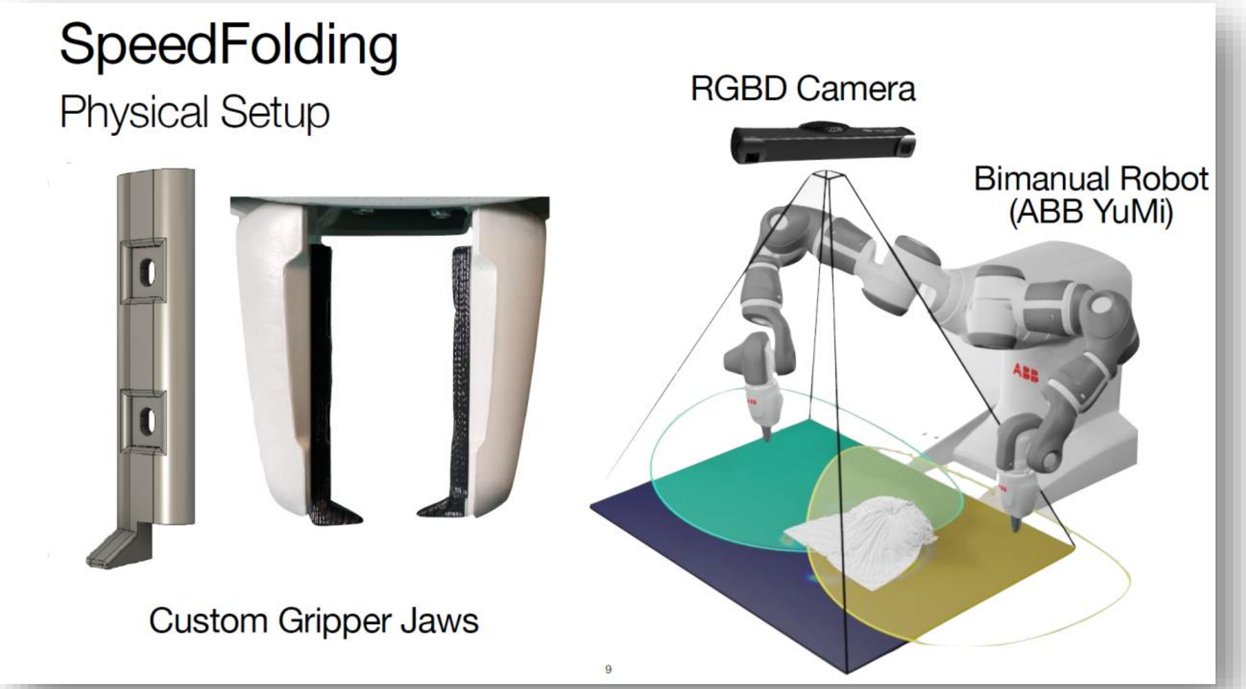


Open-loop grasping and
dropping into basket

Manipulation of Garments

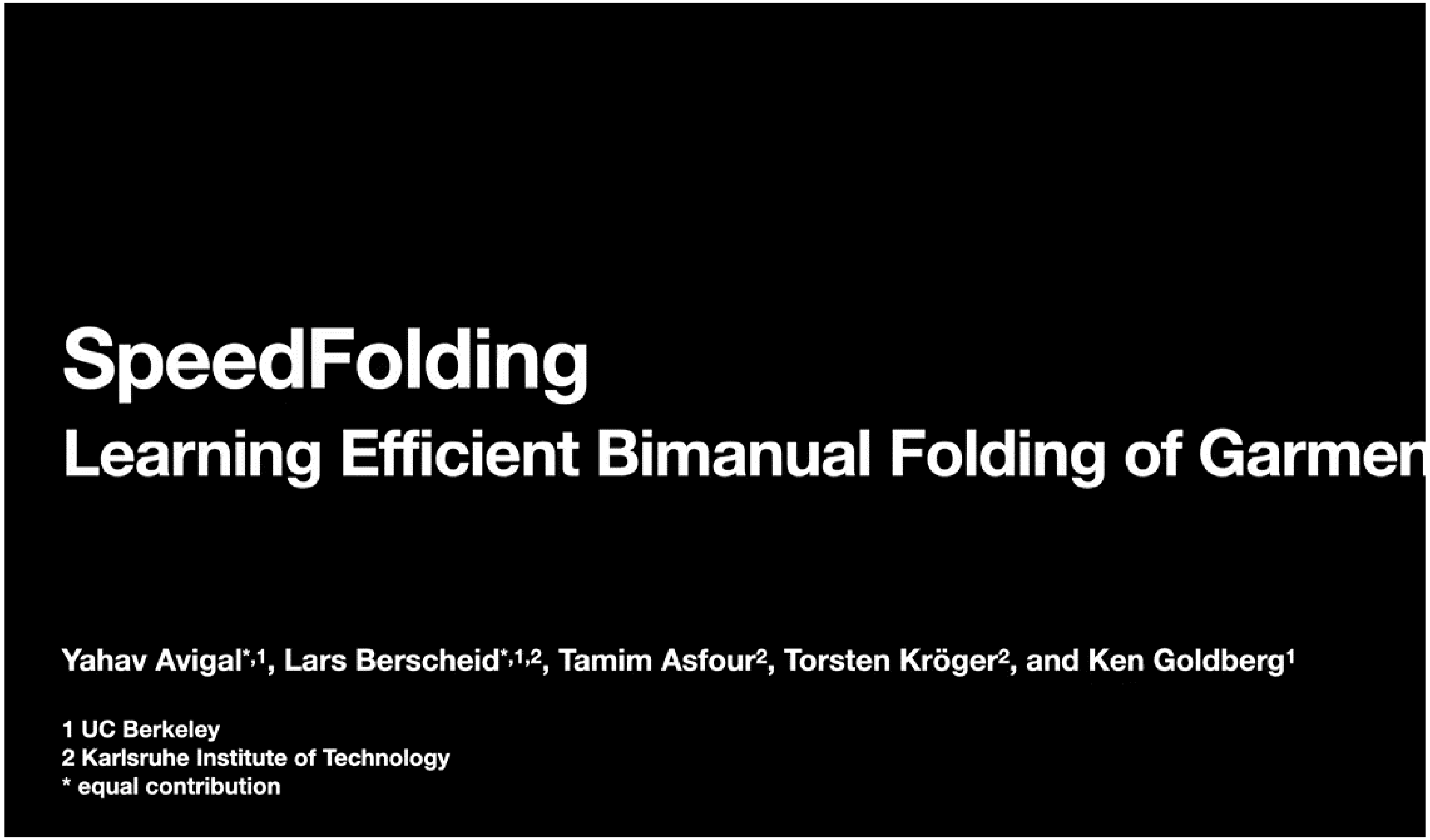
SpeedFolding

- BiMaMa-Net architecture for bimanual manipulation
- End-to-end robotic system for efficient smoothing and folding
- **Smoothing → Folding**
- **Bimanual Motion Primitives**
  - Fling
  - Drag
  - Pick and Place



Avigal, Y., Berscheid, L., Asfour, T., Kroeger, T. and Goldberg, K., *SpeedFolding: Learning Efficient Bimanual Folding of Garments*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1-8, 2022 (**Best Paper Award**)

SpeedFolding



30-40 fold per hour (FPH) compared to 3-6 FPH in previous work

Avigal, Y., Berscheid, L., Asfour, T., Kroeger, T. and Goldberg, K., *SpeedFolding: Learning Efficient Bimanual Folding of Garments*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1-8, 2022 (**Best Paper Award**)

What are the most important challenges in robotic manipulation of deformable objects?

Key Research Questions

- Manipulation of Deformables/taxtiles requires **Bimanulality**
- **Capable Hands!**
- Efficient representations for manipulation of Deformables
  - See talks of Júlia Borràs Sol and Noémie Jaquier
- Learning manipulation strategies from a few Examples
- Perception for manipulation of Deformables

Trivial for (some) humans, challenging for (all) robots

H²T Team

Humanoids@KIT

Thanks to...

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Hector Stiftungen

European Commission

KIT Karlsruhe Institute of Technology

Carl Zeiss Stiftung

Federal Ministry of Research, Technology and Space

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Thank you!

Tamim Asfour
Robotics Institute Germany (RIG)
Karlsruhe Institute of Technology (KIT)

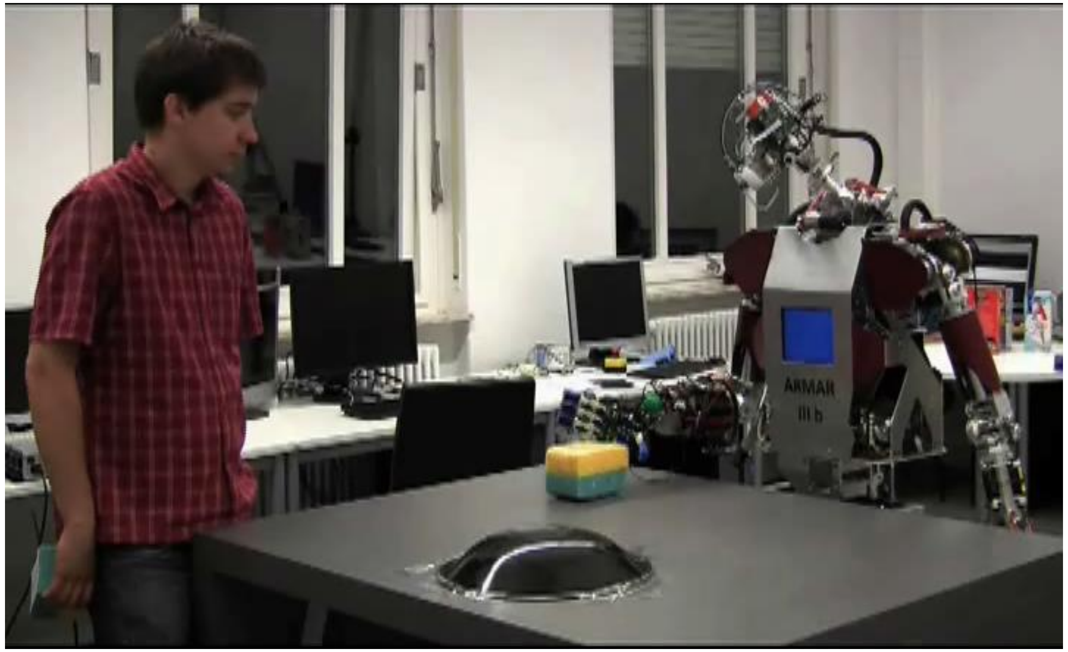
Robotics Institute Germany

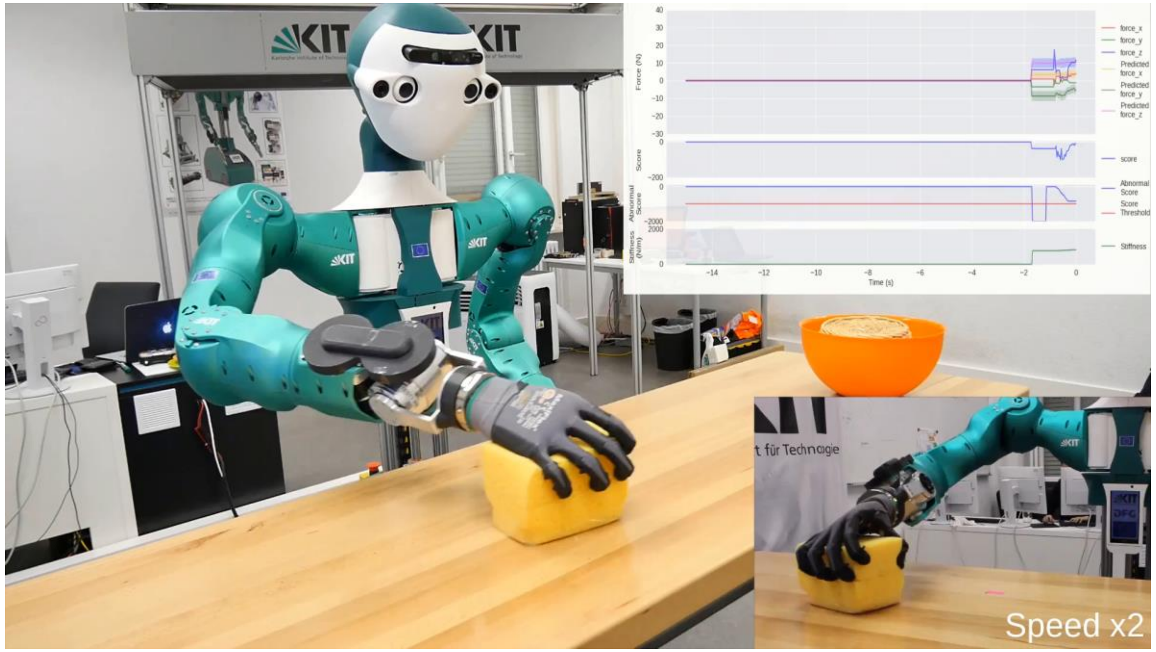
KIT

H2T
High Performance Humanoid Technologies

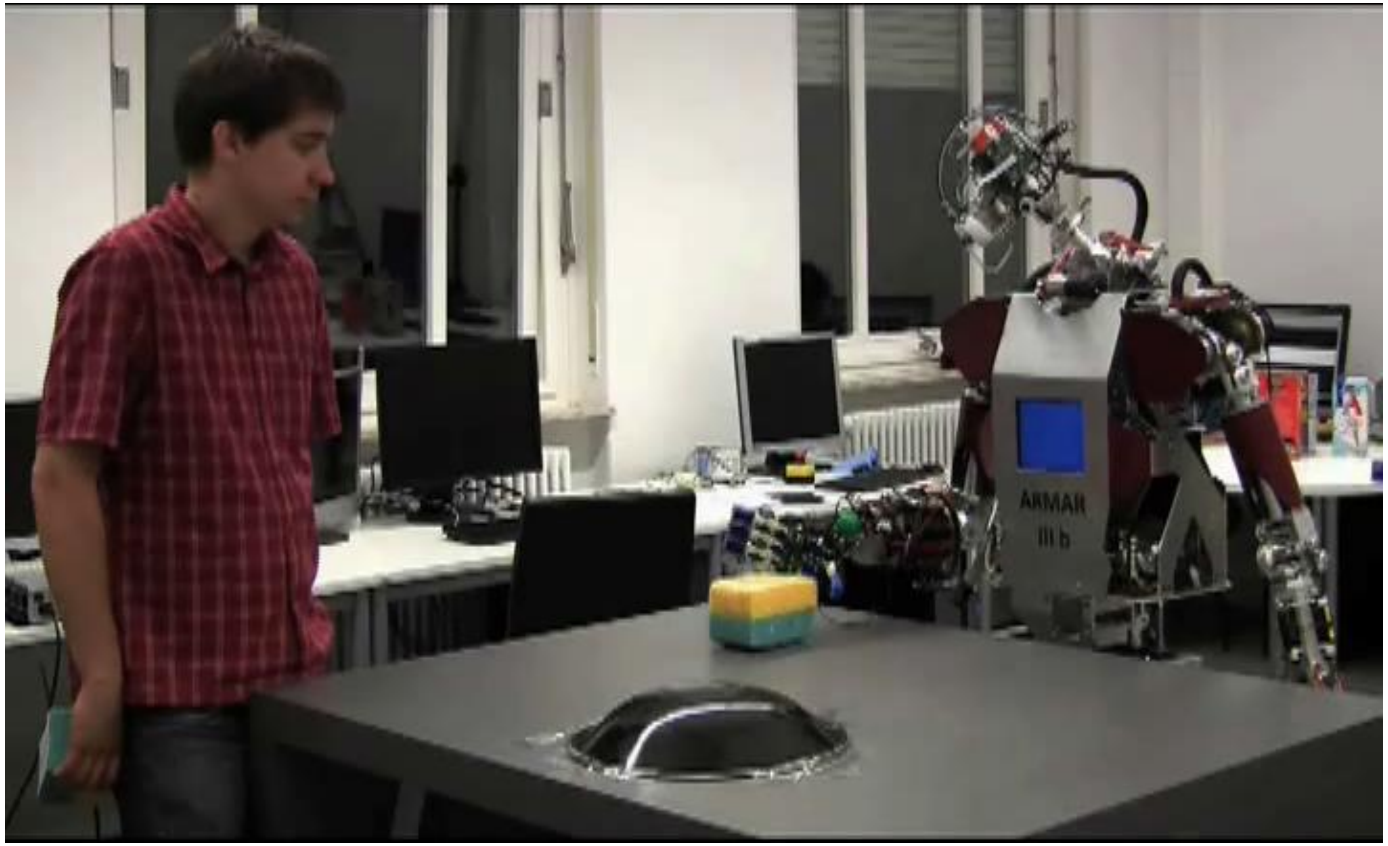
Learning from Human
Imitation Learning

- Library of motion primitives (motion alphabet)
- Tasks as sequences of motion primitives





Learning from Andrej (Gams)!



Gams, A., Do, M., Ude, A., Asfour, T. and Dillmann, R., *On-Line Periodic Movement and Force-Profile Learning for Adaptation to New Surfaces*, IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 560-565, December, 2010

Thank you!

Tamim Asfour

Robotics Institute Germany (RIG)

Karlsruhe Institute of Technology (KIT)

Robotics Institute Germany

KIT

H2T

High Performance Humanoid Technologies