

From Intuitive Immersive Telepresence Systems to Conscious Service Robots

Sven Behnke

University of Bonn
Computer Science Institute VI –
Intelligent Systems and Robotics
Autonomous Intelligent Systems



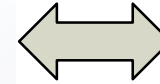
Telepresence Systems

- Enable a human operator to be present at a remote location
- Capture remote location with cameras, microphones, force & haptic sensors, etc.
- Display remote measurements to the operator
- Capture operator movements, speech, and expressions
- Transfer them to avatar robot

Operator Station



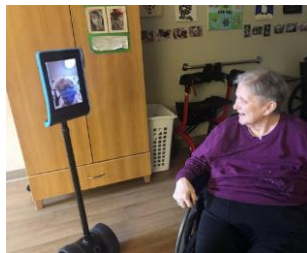
Avatar Robot



[TELESAR VI, Tachi et al. IJHR 2020]

Telepresence Applications

- Remote visits to family and friends
- Business trips
- Health care
- Personal assistance
- Remote work
- Disaster response
- Space
- Underwater
- Remote driving
- Many more



[Hung et al. 2023]



[OhmniLabs Ohmni]



[Intuitive Da Vinci]



[Pollen Reachy]



[Telexistence]



[KAIST DRC Hubo]



[NASA Robonaut]



[Stanford OceanOneK]



[Fetch]

Experience with Teleoperated Robots

- Multiple domains
- Often motivated by competitions and challenges



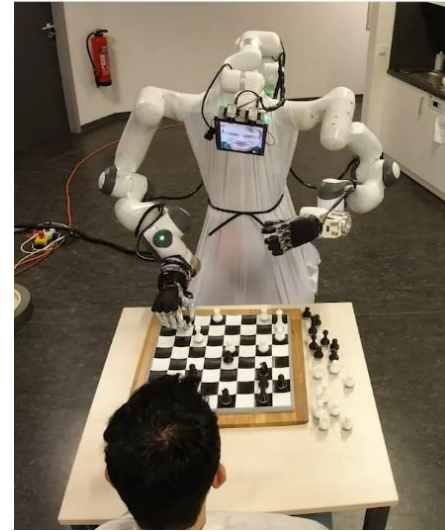
RoboCup@Home



DARPA Robotics Challenge
DLR SpaceBot Cup



CENTAURO



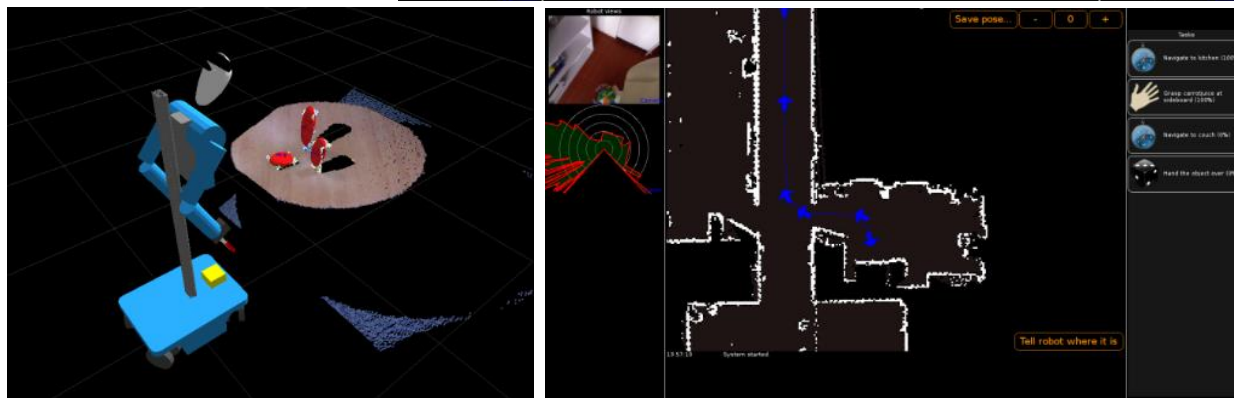
ANA Avatar XPRIZE

Cognitive Service Robot Cosero



Handheld Teleoperation Interface

- Three levels of autonomy/control:
 - Task level: Get me a beer!
 - Skill level: Grasp, place, navigate, ...
 - Direct control: Locomotion, manip.

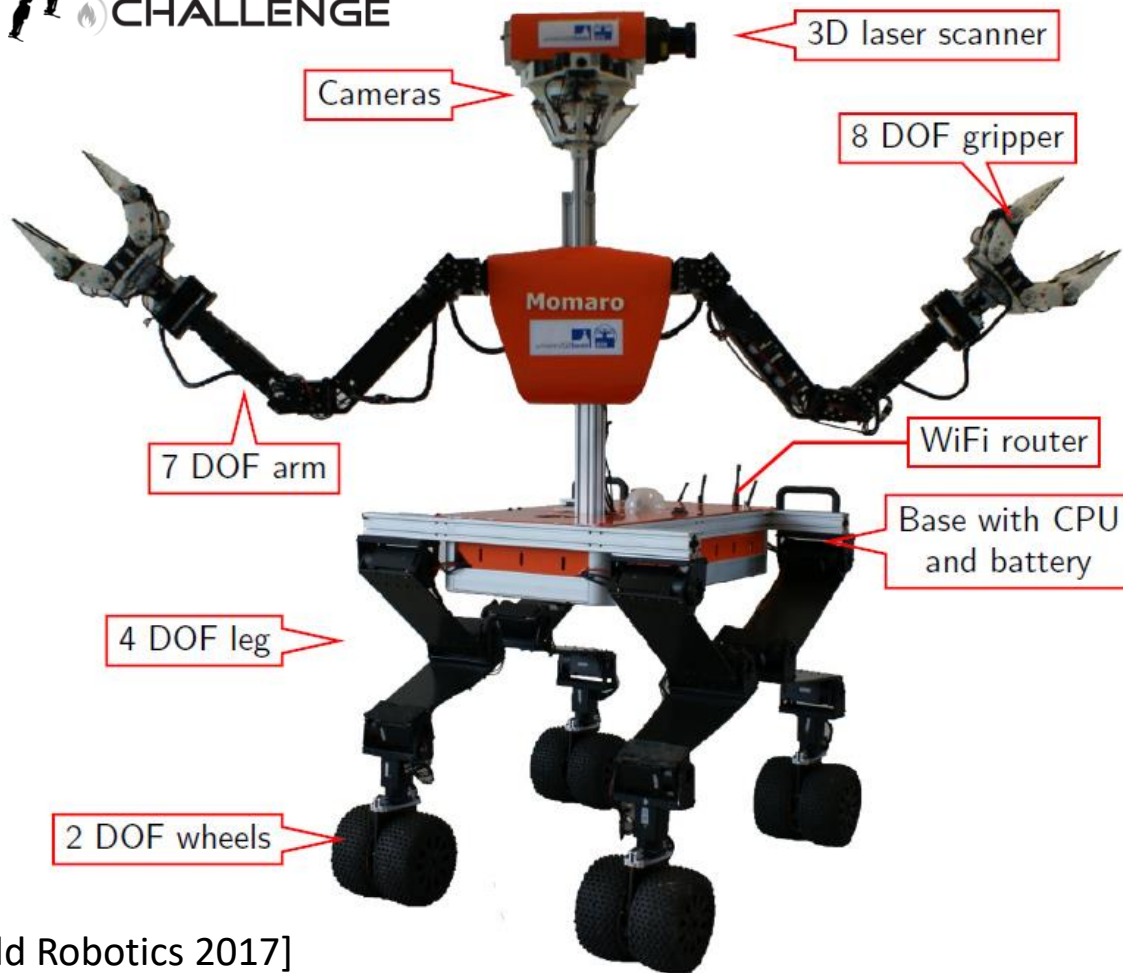


[Schwarz, Stückler, Behnke, HRI 2014]

Mobile Manipulation Robot Momaro

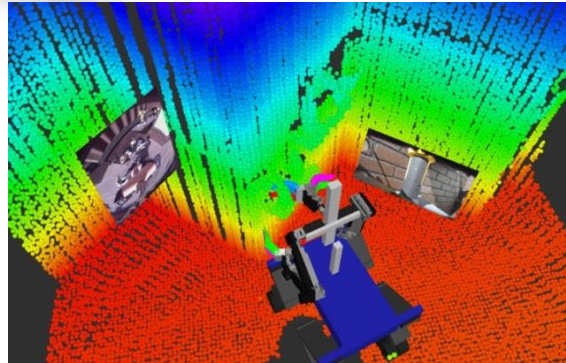


- Four compliant legs ending in pairs of steerable wheels
- Anthropomorphic upper body
- Sensor head
 - 3D laser scanner
 - IMU, cameras



Manipulation Operator Interface

- 3D head-mounted display
- 3D environment model + images
- 6D magnetic tracker



[Rodehutsors et al., Humanoids 2015]

DARPA Robotics Challenge



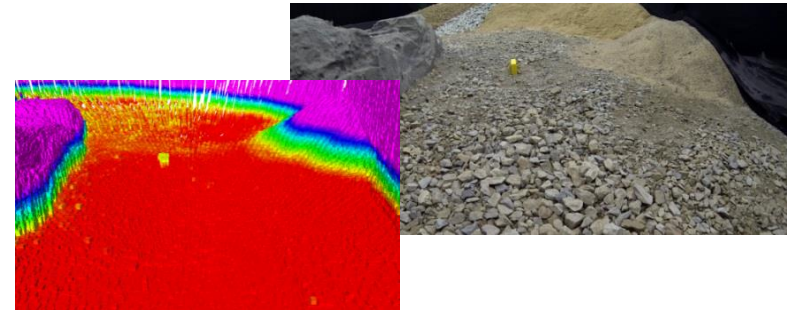
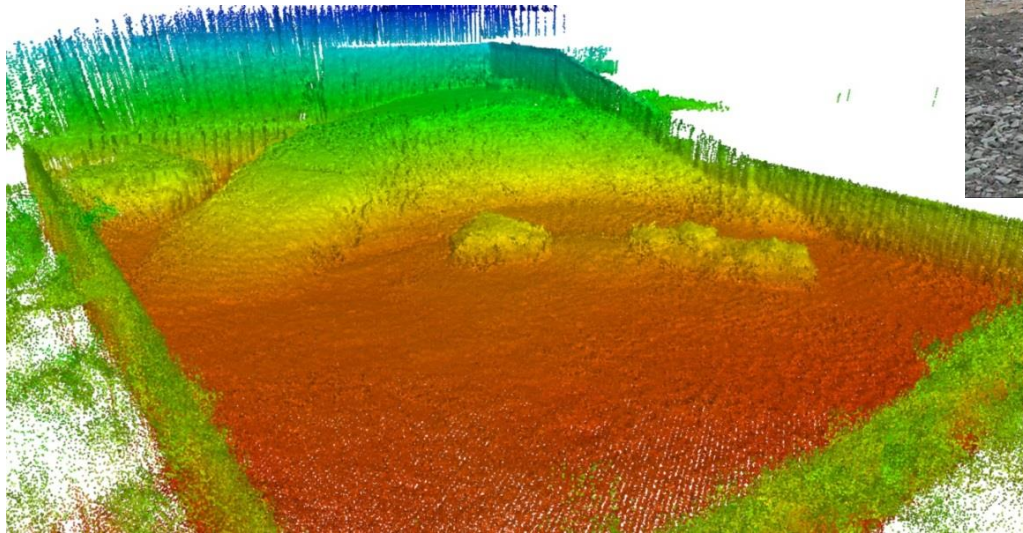
Team NimbRo Rescue



Best European Team (4th place overall),
solved seven of eight tasks in 34 minutes

DLR SpaceBot Cup 2015

- Mobile manipulation in rough terrain





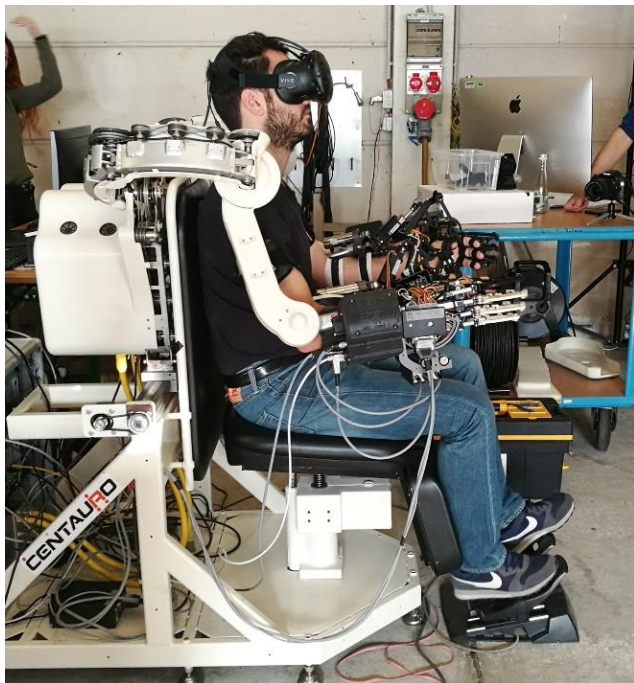
8X

[Schwarz et al., Frontiers on Robotics and AI 2016]

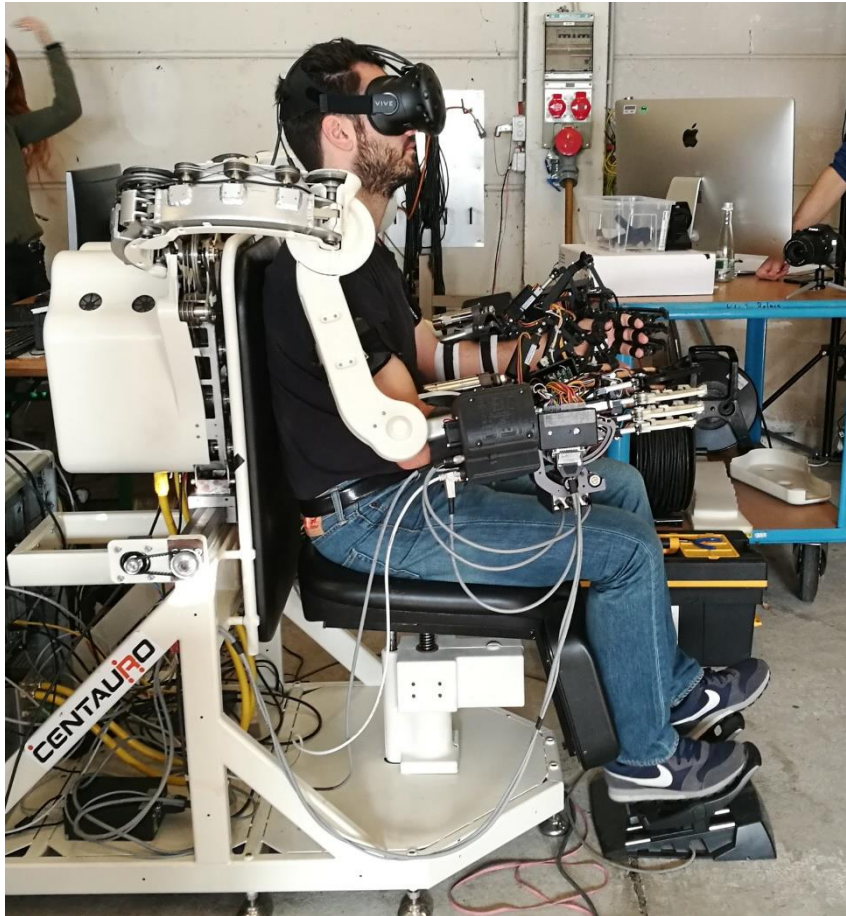
Robust Mobility and Dexterous Manipulation in Disaster Response by Fullbody Telepresence in a Centaur-like Robot

CENTAURRO

- Four-legged robot with steerable wheels and anthropomorphic upper body
- Immersive teleoperation through exoskeleton with HMD



Immersive Operator Interface



Stereo Visual and Audio Feedback

- Head-Mounted Display
- Audio Headset

Arm-Hand Movements

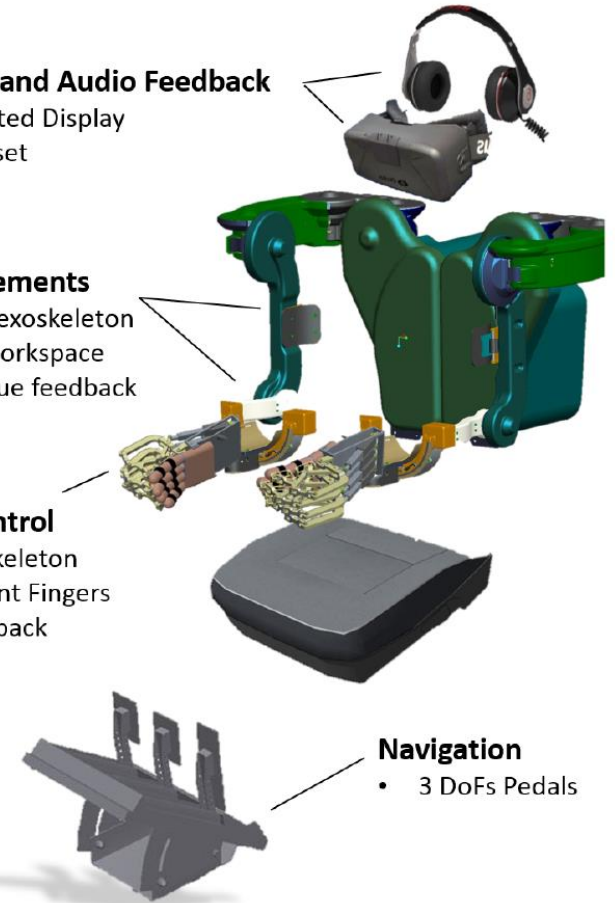
- Arm and wrist exoskeleton
- 7 DoFs, wide workspace
- Force and torque feedback

Grasping Control

- Hand exoskeleton
- Independent Fingers
- Force feedback

Navigation

- 3 DoFs Pedals



Teleoperation with Joystick and Spacemouse

3D VEROSIM
visualization

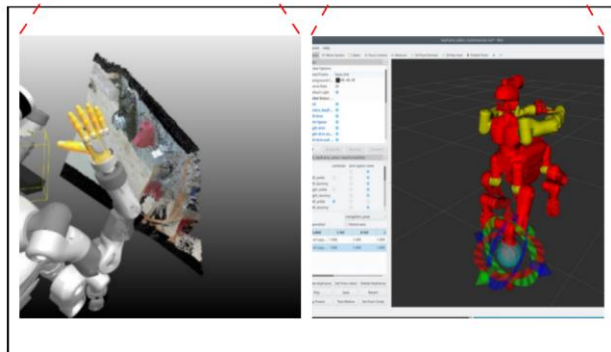
Robot state &
Keyframe editor

Foot
cameras

Panoramic view &
RGB Kinect image

Task specific
GUI

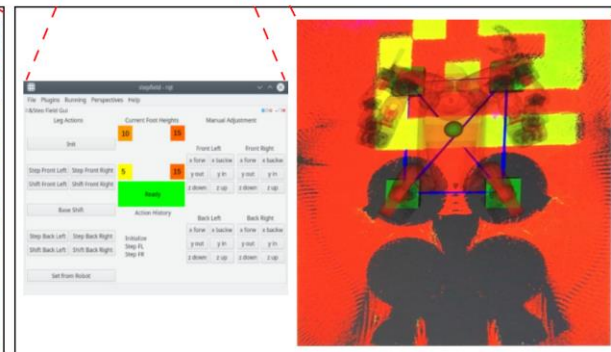
Pointcloud, ground
contact & COM markers



Monitor 1



Monitor 2



Monitor 3

- Flexible user interfaces for locomotion and manipulation tasks
- 3D situation awareness
- Motion editor



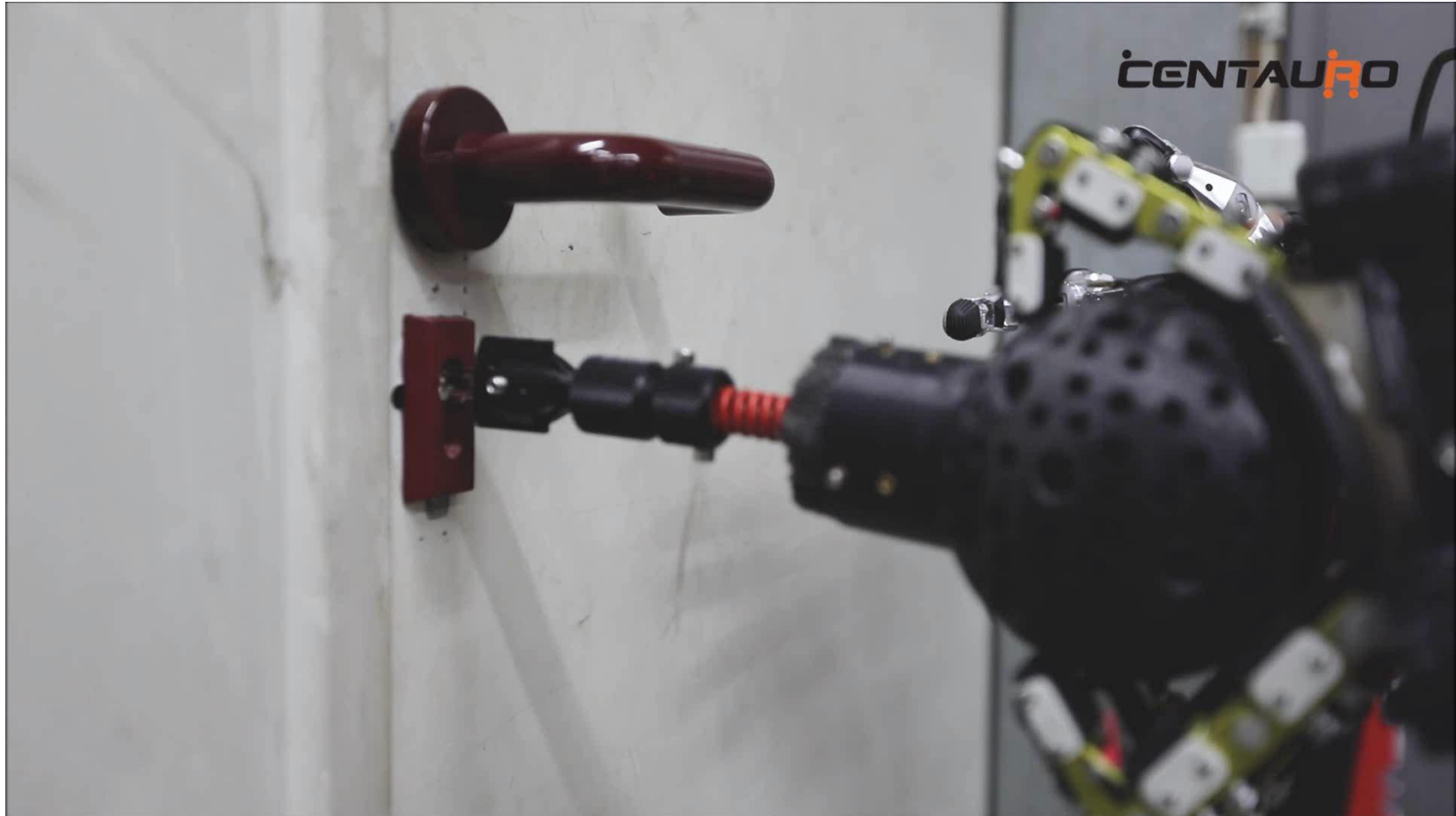
CENTAURO Evaluation @ KHG: Locomotion Tasks



Grasping an Unknown Power Drill and Fastening Screws



CENTAURO: Complex Manipulation Tasks



ANA Avatar XPRIZE Competition

- Organized by XPRIZE Foundation
- Sponsored by All Nippon Airways (ANA)
- **Objective:** Create a robotic avatar system that can transport human senses, actions, and presence to a remote location in real time
 - Expanding human connection
 - Transferring skills
 - Exploring dangerous or inaccessible places
- Panel of 22 expert judges
- Launched 03/2018
- **Prize purse of \$10M**
- 99 teams registered by 09/2019

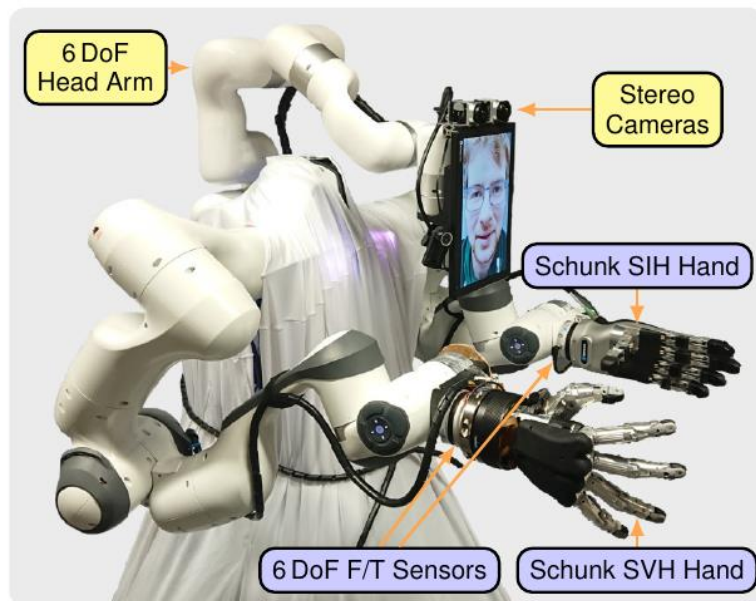
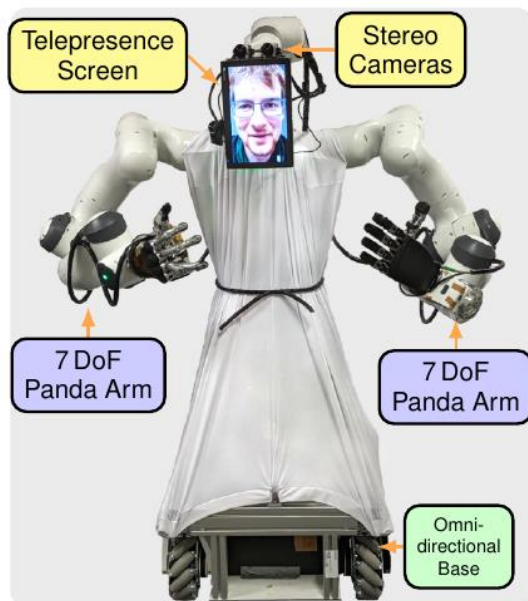


[XPRIZE]

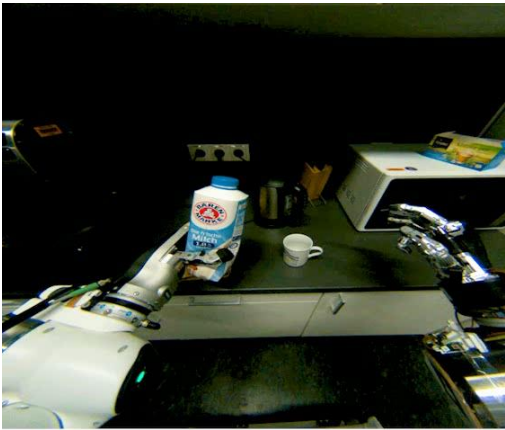
- Required mobility, manipulation, human-human interaction
- Focused on the **immersion** in the remote environment and the **presence** of the remote operator



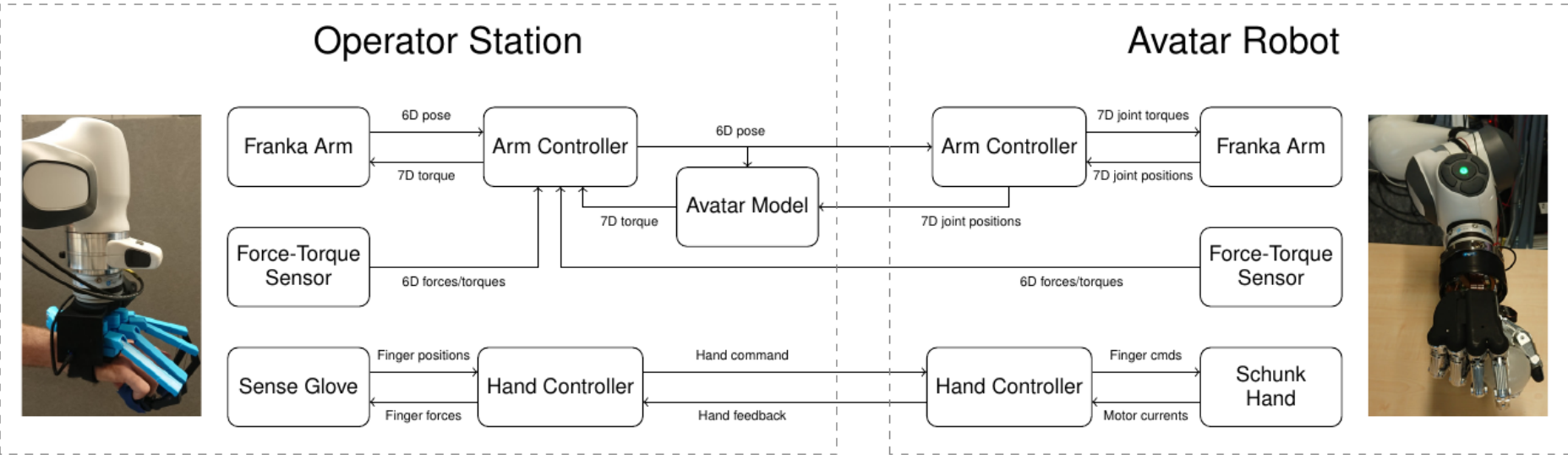
- Two-armed avatar robot designed for teleoperation with immersive visualization & force feedback
- Operator station with HMD, exoskeleton and locomotion interface



Team NimbRo Semifinal Submission



Manipulation with Force and Haptic Feedback



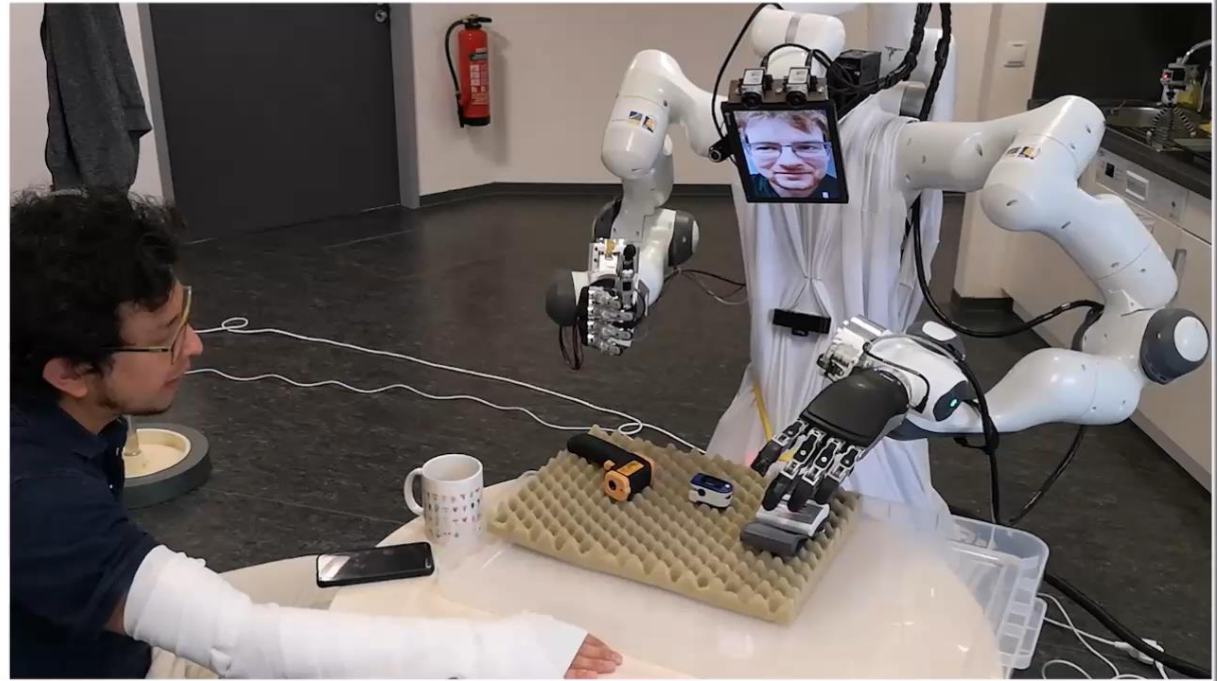
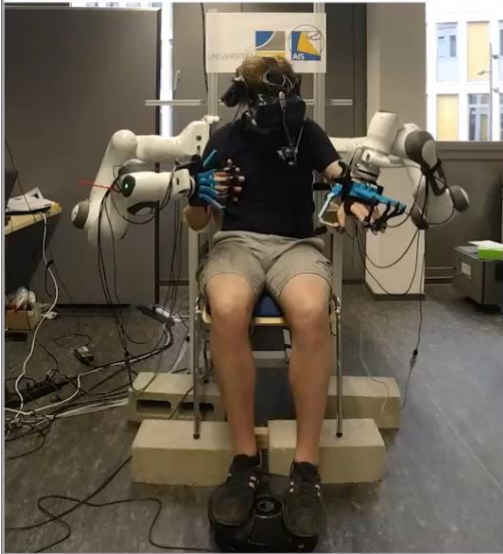
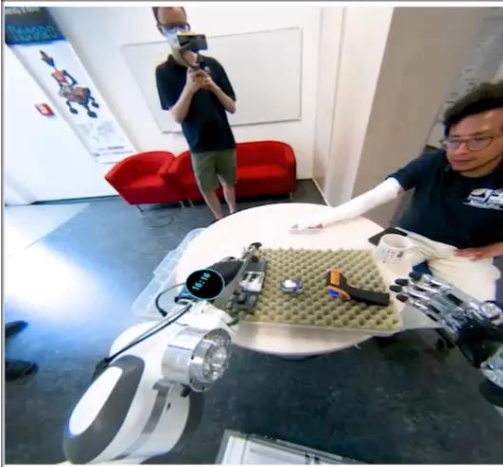
- Arm exoskeleton (Franka Emika Panda), F/T sensor (Nordbo + OnRobot HEX), hand exoskeleton (SenseGlove)
- Avatar side: Arm + F/T sensor + Schunk SVH / SIH hand
- Provides force feedback for wrist and haptic feedback for fingers
- Avatar limit avoidance using predictive model to reduce latencies

Team NimbRo

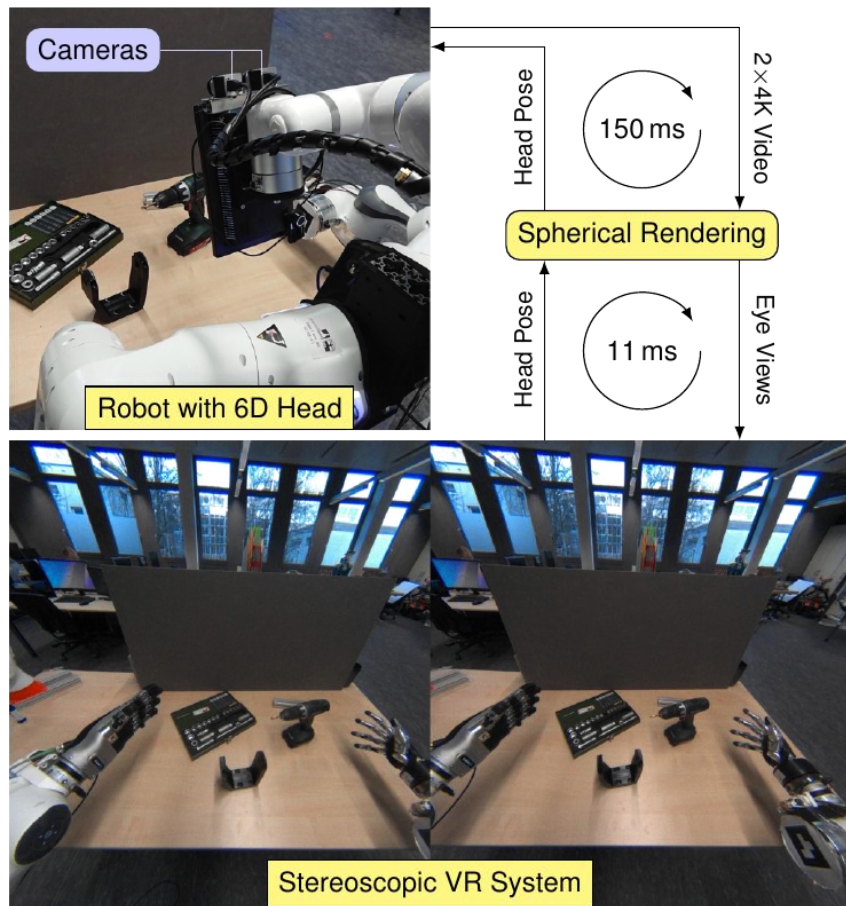
Semifinal Team Video

Tasks

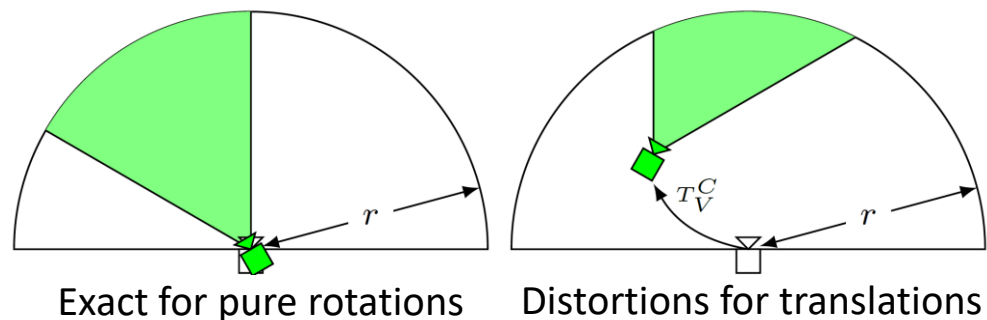
1. Make a coffee
2. Greet the recipient
3. Measure temperature
4. Measure blood pressure
5. Measure oxygen saturation
6. Help recipient with jacket



NimbRo Avatar: Immersive Visualization

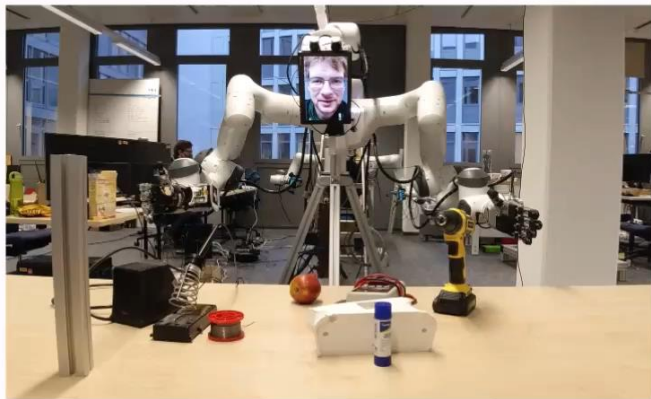


- 4K wide-angle stereo video stream
- 6D neck allows full head movement
 - Very immersive
 - Good hand-eye coordination
- Spherical rendering technique hides movement latencies
 - Assumes constant depth

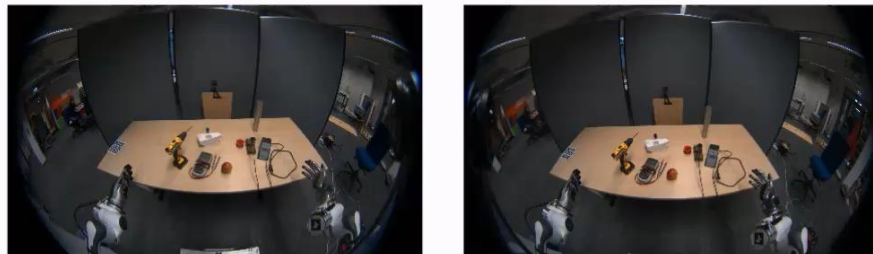


NimbRo Avatar: Immersive Visualization

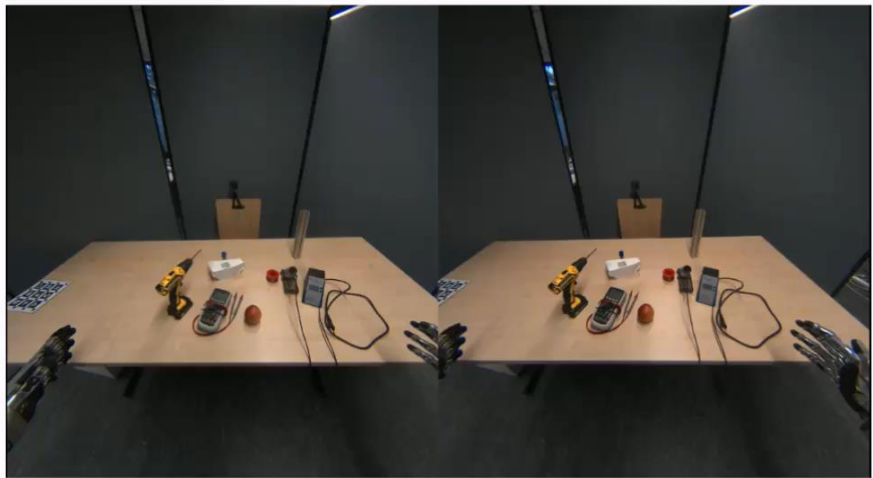
Avatar Robot



Wide-Angle Stereo



HMD View



Operator



NimRo Avatar: Operator Face Animation

- Operator images without HMD
- Capture mouth and eyes
- Estimate gaze direction and facial keypoints
- Generate animated operator face using a warping neural network



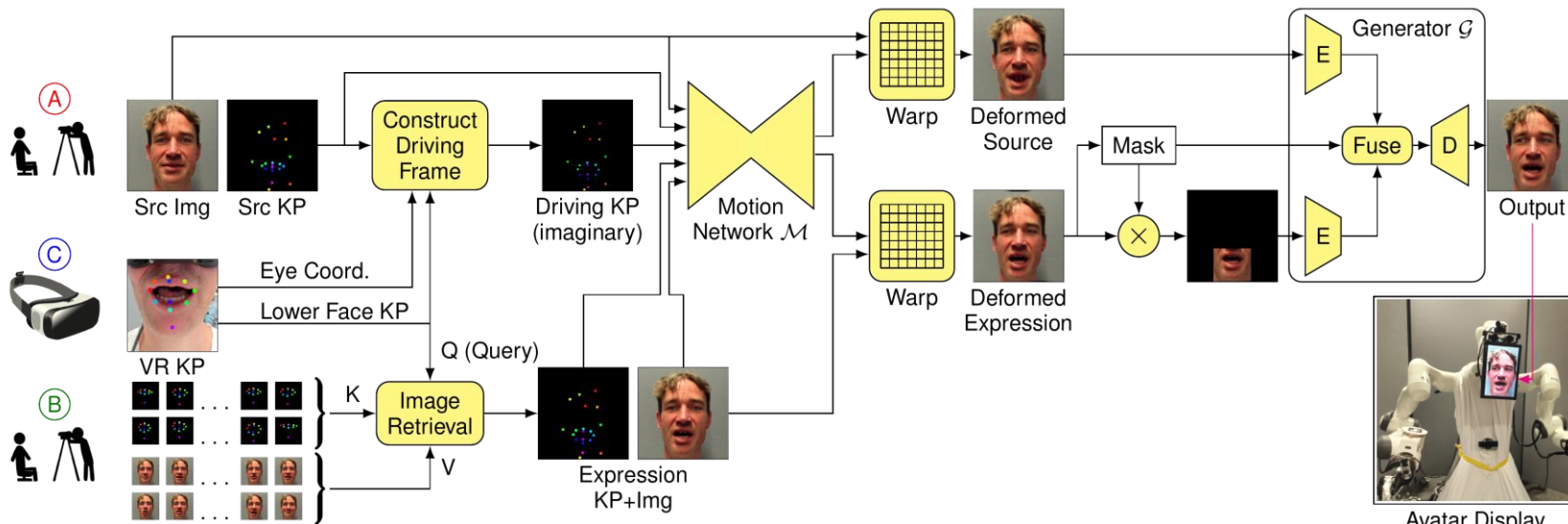
Left Eye



Mouth



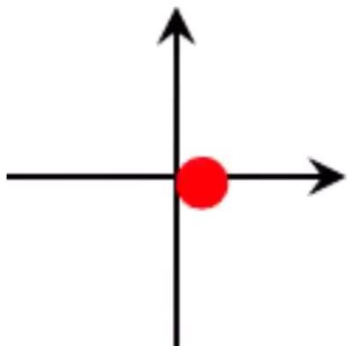
Right Eye



[Rochow et al. IROS 2022]

NimbRo Avatar: Operator Face Animation

Gaze
Direction



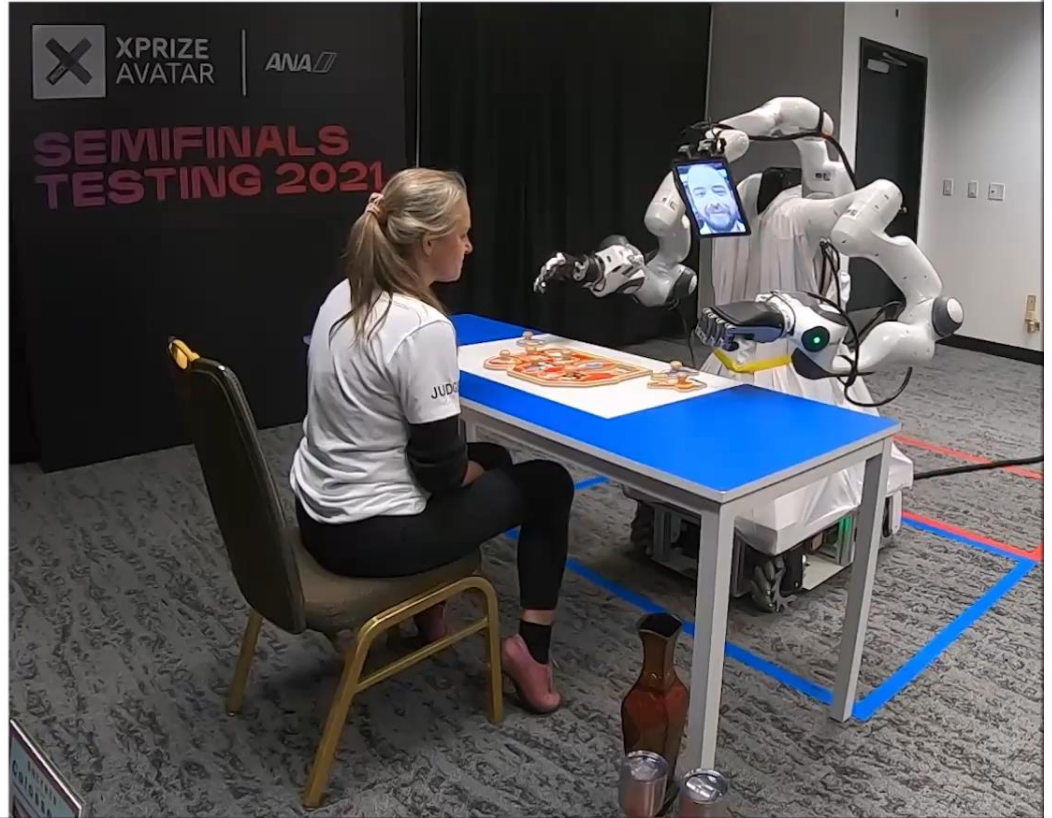
Output

Mouth Cam



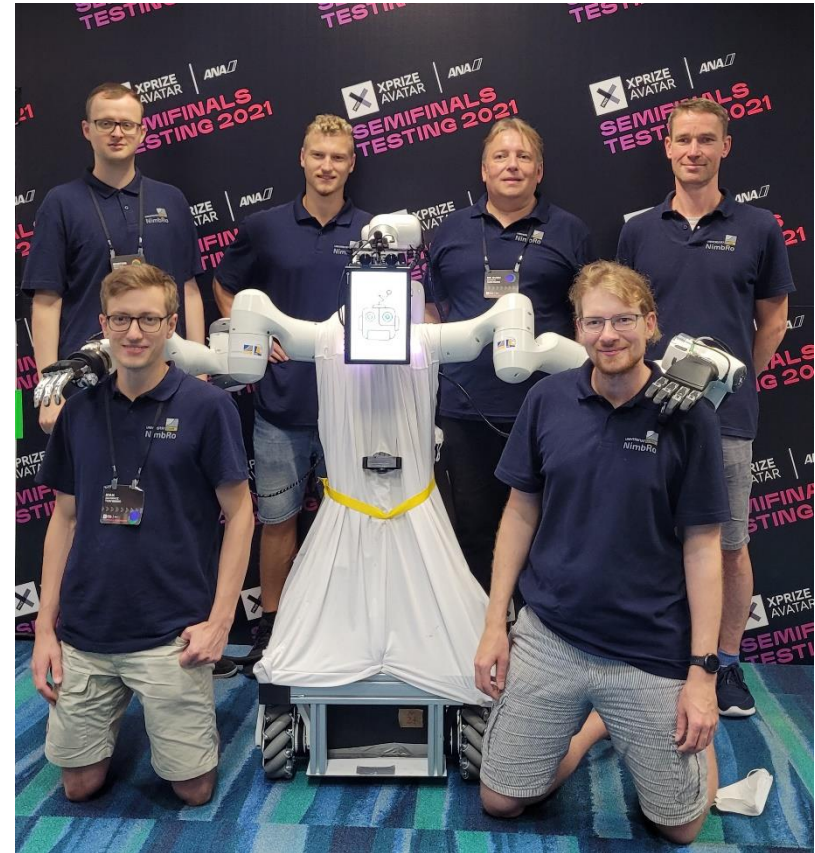
NimbRo Avatar

Avatar XPRIZE Semifinals



Semifinals Conclusions

- Designed an Avatar system for intuitive immersive telepresence
- Very good immersive visualization
- Operator-Recipient interaction with facial animation
- Bimanual human-like manipulation with force and haptic feedback
- Omnidirectional drive with birds-eye navigation view
- Scored 99/100 points, ranked 1st in the Semifinals
- Judges seemed to enjoy our system



Semifinals Results

Rank	Team Name	Country	Tested in	Score
1	NimbRo	Germany	Miami	99
2	iCub	Italy	own lab	95.25
3	i-Botics	Netherlands	own lab	93.75
4	Team Northeastern	Unites States	Miami	93
5	Dragon Tree Labs	Singapore	Miami	93
6	AVATRINA	United States	Miami	92.75
7	Avatar Hubo	United States	Miami	92
8	Tangible	United States	Miami	92
9	AlterEgo	Italy	own lab	91.75
10	Cyberselves	Un. Kingdom	Miami	90.75
11	Team SNU	South Korea	Miami	89.5
12	Pollen Robotics	France	Miami	89,5
13	Last Mile	Japan	Miami	88.5
14	Enzo	Colombia	own lab	87.25
15	Team UNIST	South Korea	Miami	86
16	Inbiodroid	Mexico	Miami	84.5
17	Rezillient	United States	Miami	84
18	Touchlab	Un. Kingdom	Miami	82.5
19	AvaDynamics	United States	Miami	80.5
20	Janus	France/Japan	own lab	80

[XPRIZE]

New Finals Requirements

- Untethered avatar robot, more mobility
 - Movable operator station
 - Mission on a distant planet
 - 10 tasks must be solved in given sequence
 - 11/2022: Qualification day, two testing days with daily down-selection of teams
- => System reliability extremely important



Long Beach, CA, USA



Finals Testing Arena

Finals Teams

- 17 teams from 10 countries
- Top research groups and companies



Inbioidroid Avatar-Hubo AvaDynamics SNU UNIST AlterEgo iCub i-Botics Cyberselves Tangible NimRo Northeastern AVATRINA Last Mile Pollen Janus Dragon Tree Labs

Finals Tasks

- Three domains:
 - Connectivity
 - Exploration
 - Skill transfer
- Incl. judging object weight and remote feeling of texture
- One point per task
- Tasks fulfillment had highest importance in scoring
- Trial time to break ties



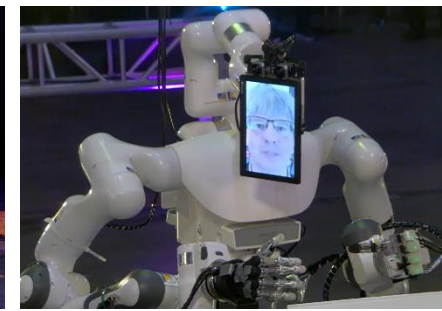
Start



1. Move



2. Introduce



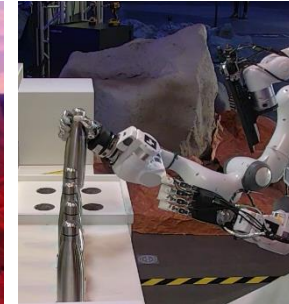
3. Confirm mission



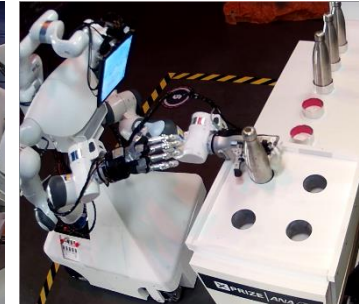
4. Activate switch



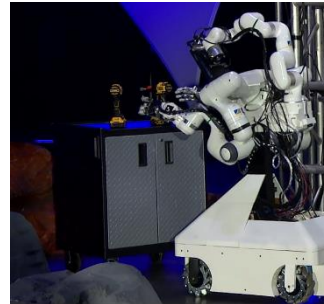
5: Travel planet



6. Identify full canister



7: Place it



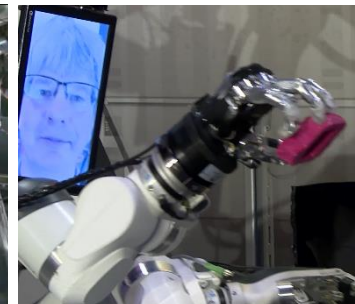
8. Narrow pathway



9: Use drill



10. Feel texture



Finish

[XPRIZE]

Finals Judged Scoring

■ Operator Experience (3 points)

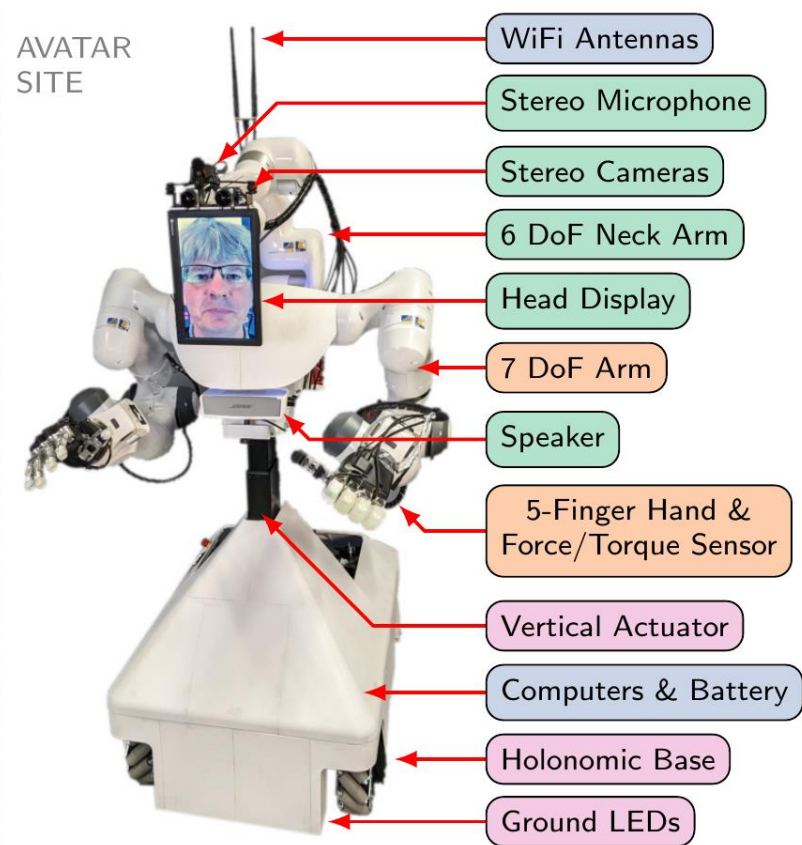
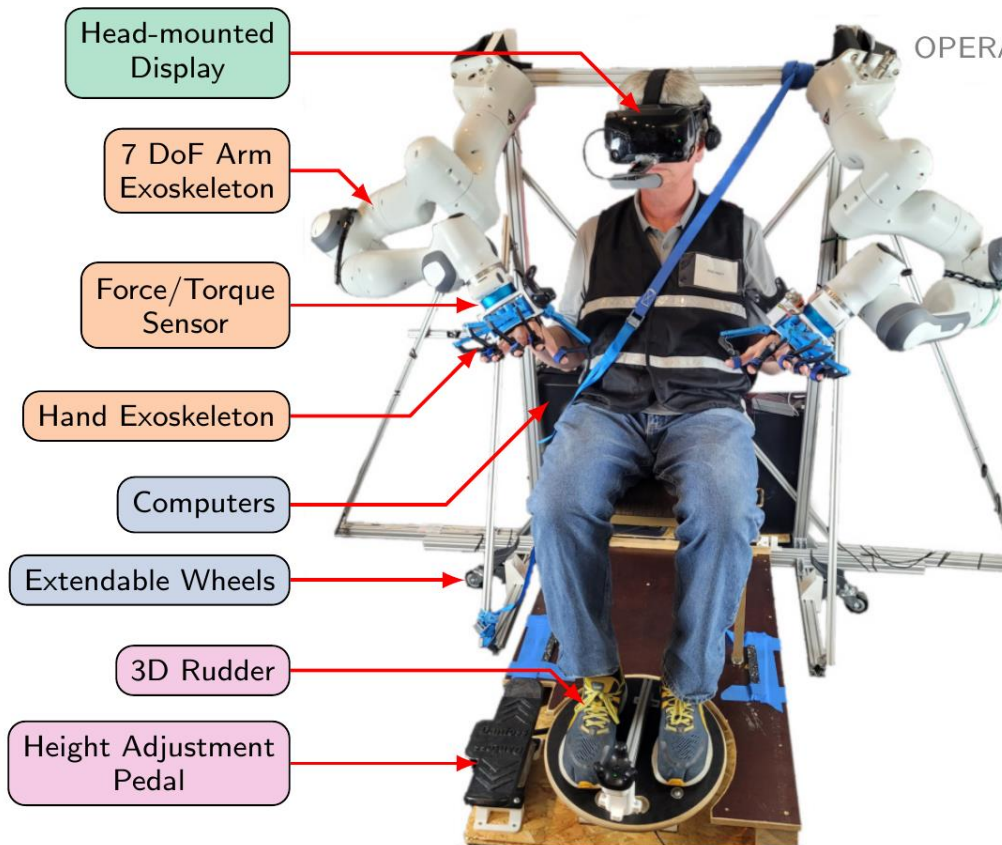
- The avatar system enabled the operator judge to **feel present** in the remote space and conveyed appropriate sensory information.
- The avatar system enabled the operator judge to **clearly understand** (both see and hear) the recipient.
- The avatar system was **easy and comfortable** to use.

■ Recipient Experience (2 points)

- The avatar robot enabled the recipient judge to feel as though the **remote operator was present** in the space.
- The avatar robot enabled the recipient judge to **clearly understand** (both see and hear) the operator.



NimbRo Avatar Finals System

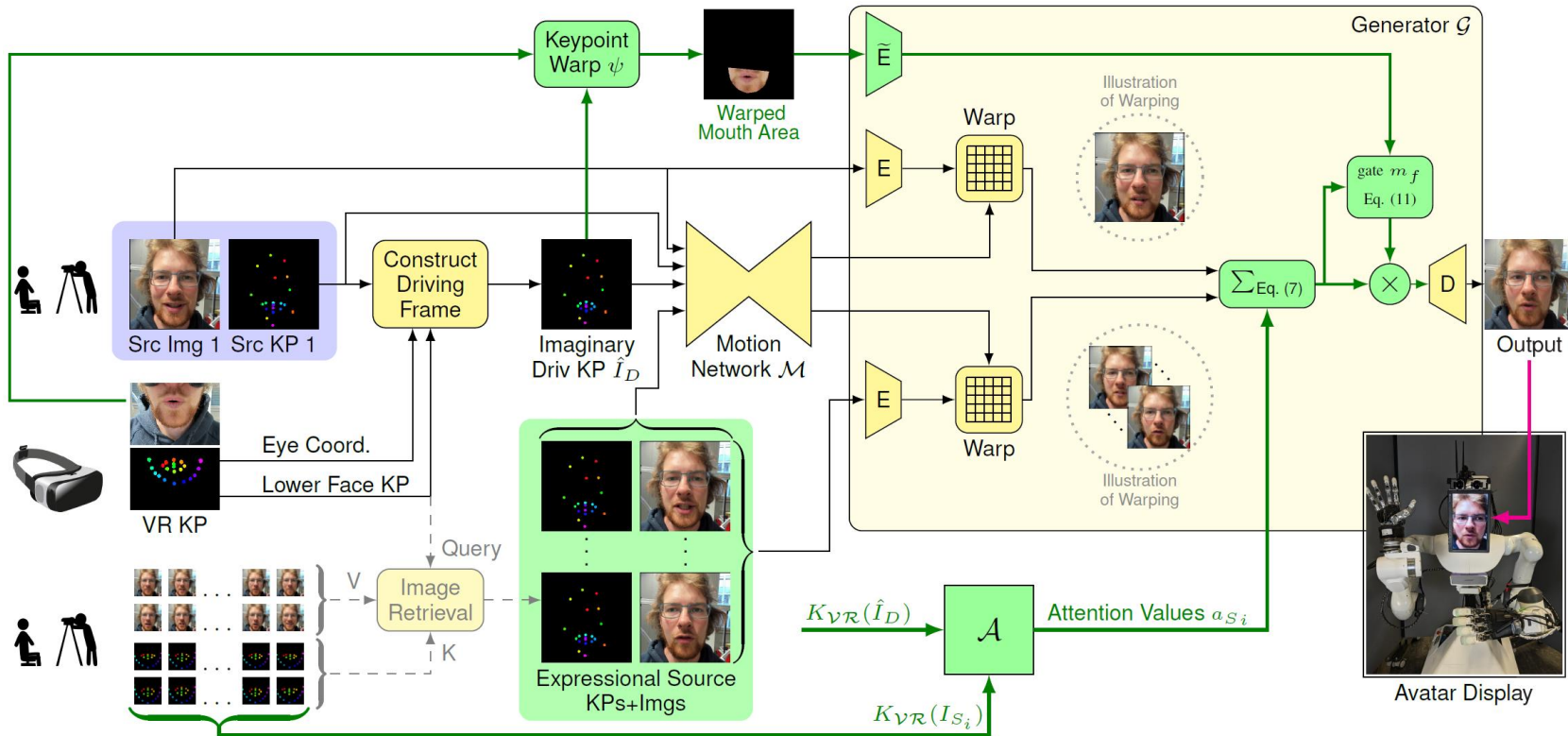


Finals Test Run Day 1



Improved Operator Face Animation

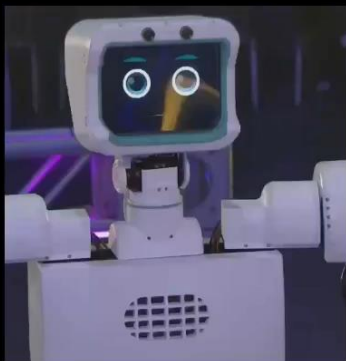
- Better temporal continuity
- Direct incorporation of mouth video



[Rochow et al. IROS 2023]

Face Animation @ Finals

Team UNIST



Ours (NimbRo)



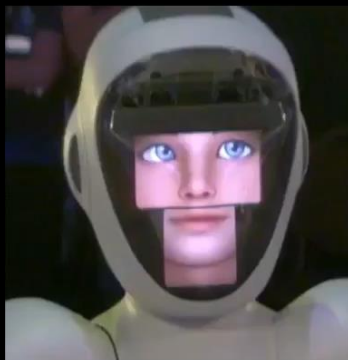
Team AVATRINA [13]



Source: Official XPRIZE Avatar live stream



Northeastern [12]



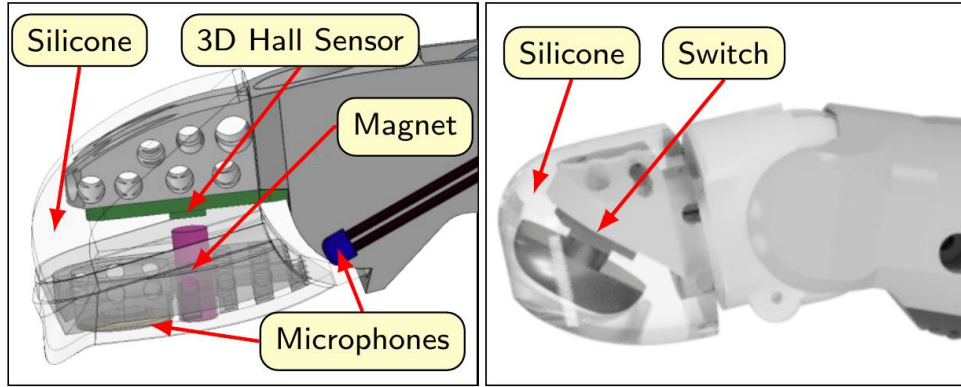
i-BOTICS



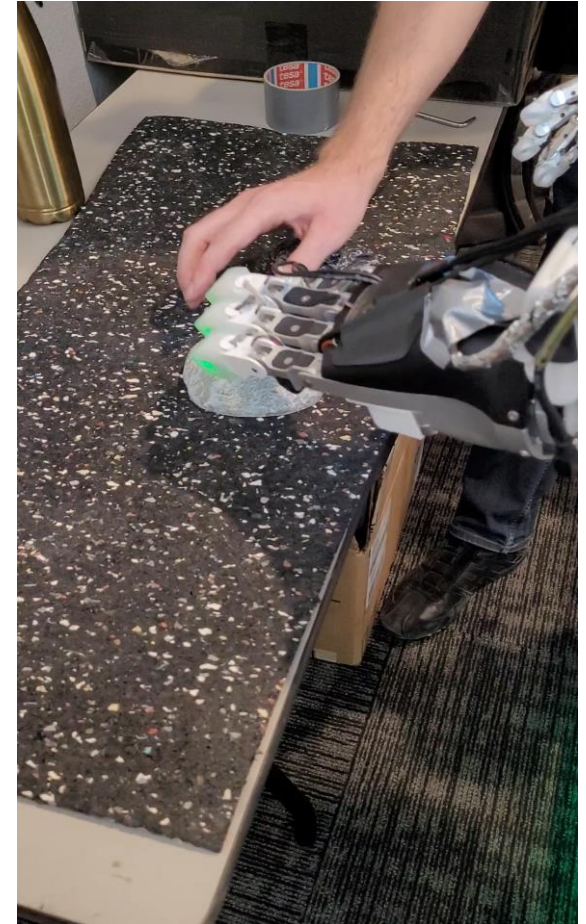
Pollen Robotics

Haptic Perception

■ Sensors in the finger tips

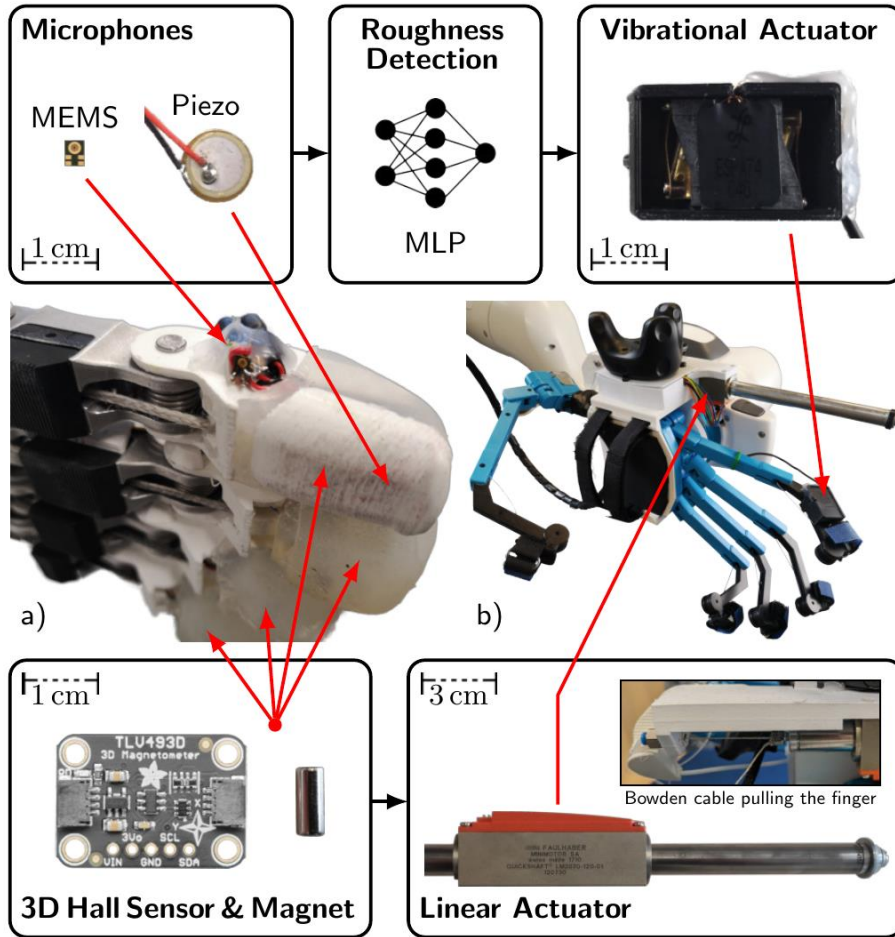


■ Actuators on the hand exoskeleton



[Pätzold et al. SMC 2023]

Roughness Perception



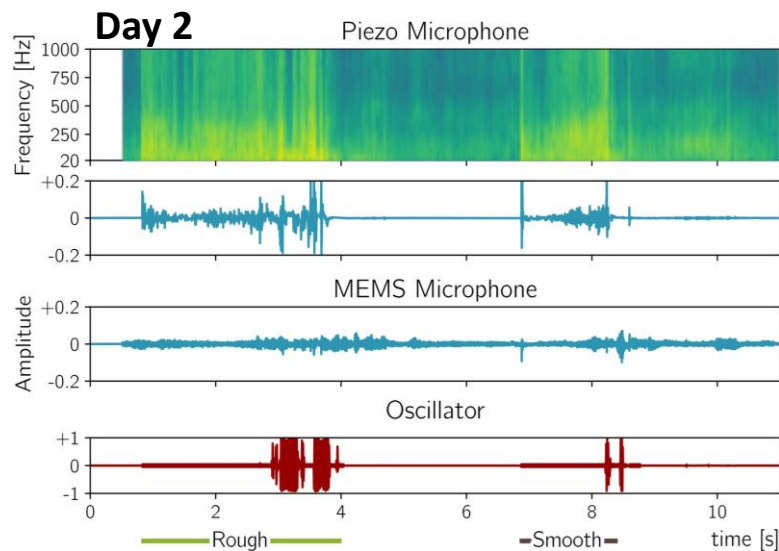
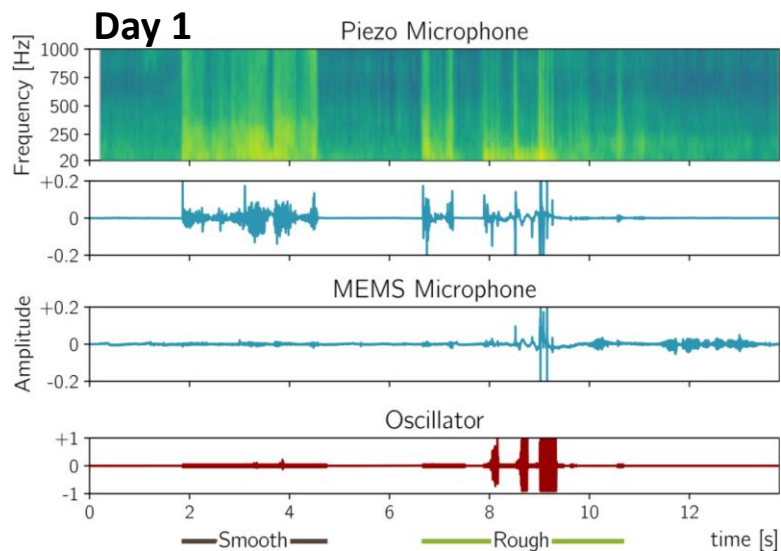
Dataset of Rough and Smooth Objects



[Pätzold et al. SMC 2023]

Finals Task 10: Retrieve a Rough Stone

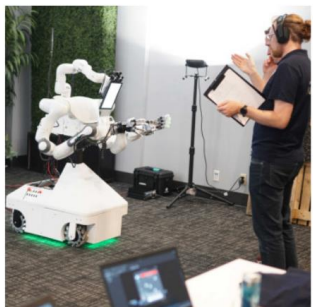
- Vision partially blocked by a curtain
- 5 stones (3 smooth + 2 rough)



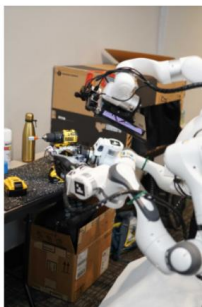
Operator Training



(a) Introduction



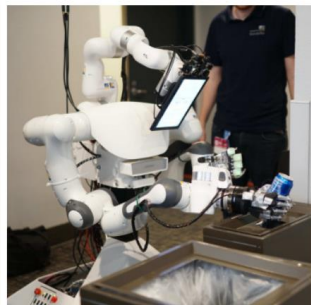
(b) Locomotion



(c) Grasping



(d) Monitoring crew



(e) Free experiments

Training	Time [min]
System overview	3
Face animation video w/o HMD	2
Put on HMD	1
Face animation video with HMD	2
Strap in hands	4
Enable arm and hand control	3
Locomotion training (T1, T5, T8)	4
Training switch and canister (T4, T6, T7)	5
Training power drill (T9)	5
Training stones (T10)	10
Enjoy the system	3
System recovery & recap	3
Total training	45

- Dedicated roles: Communication with operator, Software control, Face animation, Hardware support
- Trade-off between learning by own exploration vs. explicit instruction

Moving into the Arena



- Seamless roaming / disconnection handling with UDP data streaming
- No calibration/initialization/button press
- Essential: Operator room crew in the voice loop during setup
- Gamepad control

Operator Crew GUI

Anna

control_box/Clock: **400:37**

Anna/symon/state

Battery	Power supply 100%
CPU	Usage 15.11%
Temperature	CPU: 88° PCH: 67° SSD: 44°
HDD	Usage 32% (596G free)
USB	All 11 devices checked
Ping	All 6 connections checked
Network	All 3 connections checked
Basler Left	46.3 Hz (delay 0.09s)
Basler Right	45.8 Hz (delay 0.07s)
Brio Front	19.7 Hz (delay 0.13s)
Brio Rear	15.1 Hz (delay 0.15s)
Hand Cam	15.0 Hz (delay 0.11s)
Hand Left	1: 46°, 2: 48°, 3: 46°, 4: 44°
Hand Right	48.9 Hz (delay 0.04s)
Magnet	3 sensors
SVH Contact	193.2 Hz (delay 0.04s)
Head	Delay: 0.02s
Arm Left	Delay: 0.02s
Arm Right	Delay: 0.02s
FT left	480.2 Hz (delay 0.04s)
FT right	479.9 Hz (delay 0.04s)
Wheels	Delay: 0.05s
Spine	0.50m (57%)
Audio	Running
Face display	Human
E-Stop	OK
Bagfile	Paused

▼ Anna Core

▼ /rosmon_anna_core/state

Node	CPU:
/anna/audio/carla	0.00
/anna/audio/haptics	0.03
/anna/audio/interface	0.05
/anna/audio/jack	0.00
/anna/audio/player	0.01
/anna/audio/thru_comm	0.26
/anna/audio/thru_comm02	0.00
/anna/audio/thru_comm05	0.00
/anna/audio/thru_speak	0.01
/anna/audio/thru_haptic	0.01
/anna/audio/thru_speak_02	0.00
/anna/audio/thru_speak_03	0.00
/anna/monitor	0.01
/anna/network_control	0.00
/anna/operator_repub	0.02
/anna/service_receiver	0.00
/anna/sylog	0.00
/anna/symon	0.10
/anna_tf_static_agg	0.00
/anna_tf_transceiver	0.08
/anna/transceiver	0.04
/atlas_receiver	0.02
/atlas_sender	0.00
/config_server	0.00
/ping_node	0.00

▼ Otto

rosmon arms

rosmon_otto_arms/state

Node	CPU:
/arduno	0.00
/otto/raulhaver_comm	0.02
/otto/left/driver	0.00
/otto/left	0.04
/otto/right/driver	0.28
/otto/right	0.07
/otto/rudder_3d	0.01
/otto/state_pub	0.05

▼ /rosmon_network

rosmon_otto_network/state

No message

▼ sense_glove/GestureGUI

Rudder	0.08 rad	Thumb rot
Pedal	0.690	Max
Eye Tracking	-0.09 rad	Thumb flt

VR Calibration: Trackers/Arms not working

Audio	Running	1.31s	Min
Jamulus Otto	Registered on server	0.10 rad	Max other
Jamulus	Paused	0.280	Max

Recording: Recording

Record	FT left	Off
Run	Y Offset	On

▼ /anna/left/commander

Status: SS2	Reason: NOMINAL	Lock	Power off
-------------	-----------------	------	-----------

▼ /anna/right/commander

Status: SS2	Reason: NOMINAL	Lock	Power off
-------------	-----------------	------	-----------

▼ Otto/left/commander

Status: SS2	Reason: NOMINAL	Lock	Power off
-------------	-----------------	------	-----------

▼ Otto/right/commander

Status: SS2	Reason: NOMINAL	Lock	Power off
-------------	-----------------	------	-----------

▼ VR

VR Calibration: Up Calib, 90° Calib

network_display/network_display

Freq: 5.76 GHz
Associated since: Signal: -64 dBm
RX: 390 MB/s MCS 8 20M 14000000

5 GHz

TX: 390 MB/s MCS 8 20M 14000000

5.88 MB/s

XPRIZE

Router

22.30 MB/s

Ping RTT 0.0ms

28.22 MB/s

5.56 MB/s

2.4 GHz

TX: 26 MB/s MCS 8 20M 14000000

5.56 MB/s

▼ Otto config

▼ Anna Config

▼ Otto context

▼ Log

```
Filter:
Time Node Message
15:41:34 /otto/monitor Right tracking pose is not valid (tracker turned off?)
15:41:53 /avatar_vr /anna/basler/right/lnage/h264: waiting for transform: Query anna_basler_right_optical_frame <- anna_nominal_head_link: Would require extrapolation
/anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:41:15 /avatar_vr Can not get Senseglove data. Please check US connection.
15:41:40 /sense_glove Opening bag file: /home/avатар/eye_bags/bag_2022-11-05-23-41-34.bag
15:41:59 /otto/eye_recorder Recording stopped.
15:41:63 /avatar_vr /anna/basler/right/lnage/h264: waiting for transform: Query anna_basler_right_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:41:25 /otto/eye_recorder Recording stopped.
15:42:07 /anna/right/driver Otto right arm command is too old (81.248448935s)
15:42:08 /anna/left/driver Otto left arm command is too old (81.255314612s)
15:42:46 /otto/monitor Left tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:42:22 /avatar_vr /anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:42:29 /otto/monitor Right tracking pose is not valid (tracker turned off?)
15:42:07 /otto/monitor Left tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:43:19 /avatar_vr long delay in decoder
15:43:07 /anna/right/driver Otto right arm command is too old (141.248074364s)
15:43:08 /anna/left/driver Otto left arm command is too old (141.256047258s)
15:43:15 /avatar_vr /anna/birds_eye/out/compressed: Dropping old frames
15:43:98 /sense_glove Can not get Senseglove data. Please check US connection.
15:43:19 /otto/left/driver E-Stop released (mode 1), back to control
15:43:19 /otto/left/driver Franka:ControlException: L2ofranka: Move command rejected: command not possible in the current mode
15:43:58 /rosmon_otto_arms rosmon: /otto/left/driver died from signal 6
15:43:51 /rosmon_otto_arms rosmon: starting '/otto/left/driver'
15:43:23 /otto/left/driver Robot is locked, I'm going to unlock it...
15:43:49 /otto/left/driver Setting brakes to 0
15:43:08 /otto/left/commander Getting it: eResourcePending'
15:43:08 /otto/left/driver Could not lock/unlock brakes: state ABORTED/Got error from Franka: eResourcePending
15:43:71 /otto/left/driver Checking if resource is present...
15:43:88 /otto/left/driver Operator is present, not disabling.
15:43:26 /rosmon_otto_arms rosmon: /otto/left/driver died from signal 6
15:43:27 /rosmon_otto_arms rosmon: starting '/otto/left/driver'
15:43:24 /otto/left/driver Waiting for E-Stop release...
15:43:69 /otto/monitor Could not get kinematic tracker pose: Lookup would require extrapolation 0.09378322s into the past. Requested time 1667680190.18899593 but the earliest data is at time 1667680190.282778025, when looking up transform from frame [otto_arm_left_tracker_link] to frame [vr_link]
15:43:24 /otto/left/driver Waiting for E-Stop release...
15:43:71 /avatar_vr /anna/basler/right/lnage/h264: waiting for transform: Query anna_basler_right_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:43:24 /otto/left/driver Waiting for E-Stop release...
15:43:09 /otto/monitor Could not get kinematic tracker pose: Lookup would require extrapolation 9.993624202s into the past. Requested time 1667680190.18899593 but the earliest data is at time 1667680200.182815552, when looking up transform from frame [otto_arm_left_tracker_link] to frame [vr_link]
15:43:72 /avatar_vr /anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation
```

Basler

Freq: 5.76 GHz
Associated since: Signal: -64 dBm
RX: 390 MB/s MCS 8 20M 14000000

5 GHz

TX: 390 MB/s MCS 8 20M 14000000

5.88 MB/s

XPRIZE

Router

22.30 MB/s

Ping RTT 0.0ms

28.22 MB/s

5.56 MB/s

2.4 GHz

TX: 26 MB/s MCS 8 20M 14000000

5.56 MB/s

▼ Otto config

▼ Anna Config

▼ Otto context

▼ Log

Left Eye

Eye calibration: Waiting Start Stop

Bag file: /home/avатар/eye_bags/bag_2022-11-05-23-41-34.bag

#images: L:356, R:356, D:1639

Train error: 3.891891 deg

Train error: 3.891891 deg

▼ Bird's Eye

▼ Hand

Force

Force

Torque

Torque

▼ Roughness

roughness_detector_client/confidence

▼ /otto/haptics_sender/stats

Packet rate

Packet Loss Concealment

Operator Crew GUI

Anna

control_box/Clock: **400:37**

Anna Core

Anna/symon/state	On	Off
Movement	On	Off
Send Cmds	Usage	Off
Head Control	On	Off
Right Hand	On	Off
Left Hand	On	Off
Force / Torque	On	Off
Otto	On	Off
Anna Feedback	On	Off
Anna Limits	On	Off
Atlas	On	Off
Drive	On	Off
Spine	On	Off
Recording	On	Off
Record	On	Off
Run	On	Off
Y Offset	On	Off

Anna Network

System	0.39 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Feedback	5.32 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
TF	4.16 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Cam Left	7.16 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Cam Right	7.33 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Aux Image	4.25 MB/s	5GHz	0 p/s	2.4GHz	0 p/s

Anna Network Control Status

System	0.00 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Control	0.17 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
TF	1.35 MB/s	5GHz	0 p/s	2.4GHz	0 p/s
Aux Image	1.92 MB/s	5GHz	0 p/s	2.4GHz	0 p/s

Anna Left/Right Commander Status

Status: EXEC	Reason: NOMINAL	Action: inactive	Lock	Power off
Status: SS2	Reason: NOMINAL	Action: inactive	Lock	Power off

VR Calibration

VR Calibration	Tracked/Arms not working
Audio	Running
Jamulus Otto	Registered on server
Jamulus	Paused
Recording	Running
HDMI1	58.2 Hz (delay 0.06s)
Bagfile	Paused

network_display/network_display

Freq: 5.76 GHz
Associated since: Signal: -44 dBm
RX: 390 MB/s MCS 8 20M 40M 80M

5 GHz

TX: 390 MB/s MCS 8 20M 40M 80M

Robot

5.88 MB/s

XPRIZE

Ping RTT 0.0ms

28.22 MB/s

22.30 MB/s

Ping RTT 0.1ms

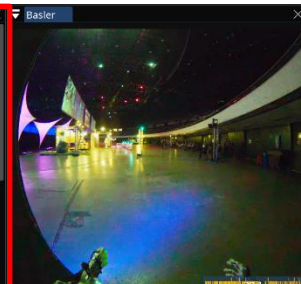
Freq: 2.412 GHz
Associated since: Signal: -53 dBm
RX: 58 MB/s MCS 0 20M 40M 80M

2.4 GHz

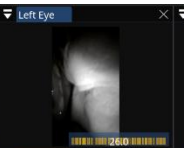
TX: 26 MB/s MCS 0 20M 40M 80M

536 MB/s


Basler




Left Eye




Right Eye



Mouth



Reconstruction



Eye calibration

Waiting Start Stop

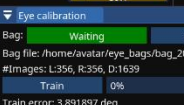
Bag file: /home/avатар/eye_bags/bag_2022-11-05-23-41-34.bag

#images: L:356, R:356, D:1639

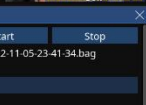
Train 0%

Train error: 3.891891 deg

Bird's Eye



Hand



Force

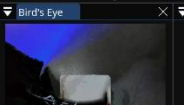
10 X
0 Y
-10 Z

Torque


4 X
0 Y
-4 Z

Roughness


roughness_detector_client/confidence



Packet rate



Packet Loss Concealment



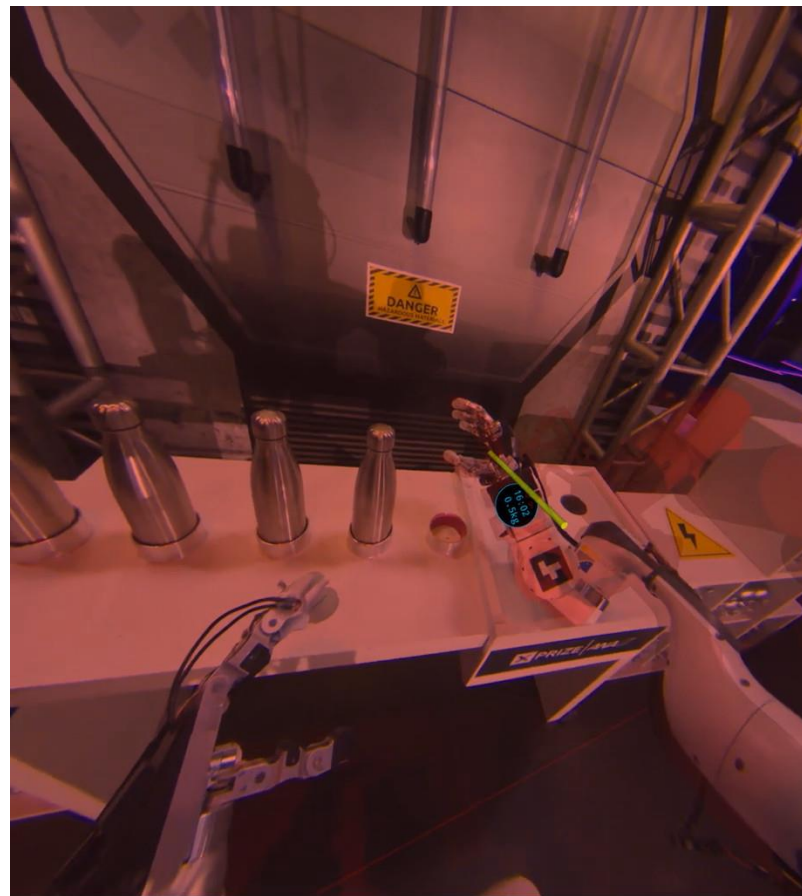
Otto config

Filter

Time	Node	Message
15:41:34	/otto/monitor	Right tracking pose is not valid (tracker turned off?)
15:41:53	/avатар_vr	/anna/basler/right/lnage/h264: waiting for transform: Query anna_basler_right_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:41:53	/avатар_vr	/anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:41:15	/avатар_vr	Could not get Senseglove data. Please check US connection.
15:41:40	/sense_glove	Opening bag file: /home/avатар/eye_bags/bag_2022-11-05-23-41-34.bag
15:41:59	/otto/eye_recorder	Recording stopped.
15:41:63	/avатар_vr	Recording stopped.
15:41:25	/otto/eye_recorder	Otto right arm command is too old (81.240440935s)
15:42:07	/anna/right/driver	Otto left arm command is too old (81.255314612s)
15:42:08	/anna/left/driver	Left tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:42:46	/otto/monitor	Left tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:42:22	/avатар_vr	/anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:42:29	/otto/monitor	Right tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:42:87	/otto/monitor	Left tracking pose is not valid (tracker turned off?) (connected=true, valid=true, result=101)
15:43:19	/avатар_vr	long delay in decoder
15:43:07	/anna/right/driver	Otto right arm command is too old (141.240074364s)
15:43:08	/anna/left/driver	Otto left arm command is too old (141.256047258s)
15:43:15	/avатар_vr	/anna/birds_eye/out/compressed: Dropping old frames
15:43:98	/sense_glove	Could not get Senseglove data. Please check US connection.
15:43:19	/otto/left/driver	E-Stop released (mode 1), back to control
15:43:19	/otto/left/driver	Franka:ControlException: L2ofranka: Move command rejected: command not possible in the current mode
15:43:50	/rosmon_otto_arms	rosmon: /otto/left/driver died from signal 6
15:43:51	/rosmon_otto_arms	rosmon: starting '/otto/left/driver'
15:43:23	/otto/left/driver	Robot is locked, I'm going to unlock it...
15:43:49	/otto/left/driver	Setting brakes to 0
15:43:08	/otto/left/commander	Checking if eResource is Pending
15:43:08	/otto/left/driver	Could not lock/unlock brakes: state ABORTED/got error from Franka: eResourcePending
15:43:71	/otto/left/driver	Checking if present to transform...
15:43:88	/otto/left/driver	Operator is present, not disabling.
15:43:26	/rosmon_otto_arms	rosmon: /otto/left/driver died from signal 6
15:43:27	/rosmon_otto_arms	rosmon: starting '/otto/left/driver'
15:43:24	/otto/left/driver	Waiting for E-Stop release...
15:43:69	/otto/monitor	Could not get kinematic tracker pose: Lookup would require extrapolation 0.09378322s into the past. Requested time 1667680190.18899593 but the earliest data is at time 1667680190.282770025, when looking up transform from frame [otto_arm_left_tracker_link] to frame [vr_link]
15:43:24	/otto/left/driver	Waiting for E-Stop release...
15:43:71	/otto/left/driver	/anna/basler/right/lnage/h264: waiting for transform: Query anna_basler_right_optical_frame <- anna_nominal_head_link: Would require extrapolation
15:43:24	/otto/left/driver	Waiting for E-Stop release...
15:43:09	/otto/monitor	Could not get kinematic tracker pose: Lookup would require extrapolation 9.993024202s into the past. Requested time 1667680190.18899593 but the earliest data is at time 1667680200.182815552, when looking up transform from frame [otto_arm_left_tracker_link] to frame [vr_link]
15:43:72	/avатар_vr	/anna/basler/left/lnage/h264: waiting for transform: Query anna_basler_left_optical_frame <- anna_nominal_head_link: Would require extrapolation

Reliability Features

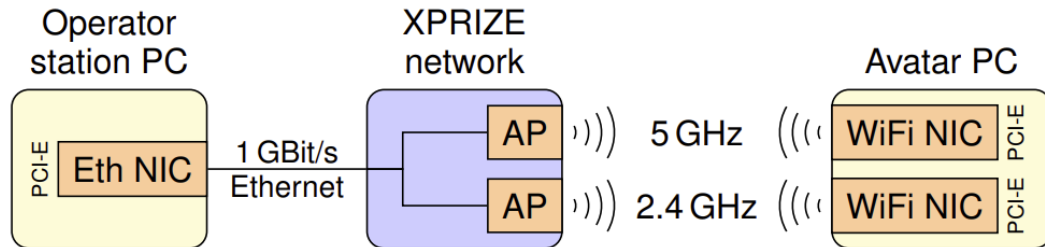
1. Operator crew awareness
2. Automatic arm resets
3. ROS node respawn
4. State- and connectionless network system (pure UDP)
5. Redundant WiFi connections
6. PC watchdog



Network Details

- Separate ROS cores operator station and avatar
- Pure UDP, no re-connect / initialization
- Main camera stream (stereo 2472×2178 @46 fps) is HEVC-encoded & decoded on GPU (NVENC).
Total Bandwidth: ~14 MBit/s
- Control data is sent redundantly
- Monitor packet loss
- The core software is already open source, more to come:

https://github.com/AIS-Bonn/nimbro_network



WiFi Bandwidth Requirements

Downlink from avatar				Uplink to avatar			
Channel	MBit/s	5 GHz	2.4 GHz	Channel	MBit/s	5 GHz	2.4 GHz
Arm feedback	8.5	✓	×	Arm control	4.9	✓	✓
Transformations	4.1	✓	×	Transformations	1.4	✓	×
Main cameras	14.7	✓	×	Operator face	5.7	×	✓
Hand camera	5.5	×	✓	Audio	0.4	✓	✓
Diagnostics	0.4	✓	✓				
Audio	0.4	✓	✓				
Total [MBit/s]		28.1	6.3	Total [MBit/s]		6.7	11.0

Audio Details

- Low-latency solution utilizing the *JACK Audio Connection Kit*
- Redundant UDP transmission via the *OPUS audio codec*
- *NVIDIA MAXINE* for GPU-accelerated *acoustic echo cancelation*
- *Jamulus* for team communication with operator and recipients



Finals Day 2 Testing



Rank	Team name	Time	Task score	Judged score	Total
1	NimbRo (DE)	5:50	10	5	15
2	Pollen Robotics (FR)	10:50	10	5	15
3	Team Northeastern (US)	21:09	10	4.5	14.5
4	AVATRINA (US)	24:47	10	4.5	14.5
5	i-Botics (NL)	25:00	9	5	14
6	Team UNIST (KR)	25:00	9	4.5	13.5
7	Inbiodroid (MX)	25:00	8	5	13
8	Team SNU (KR)	25:00	8	4.5	12.5
9	AlterEgo (IT)	25:00	8	4.5	12.5
10	Dragon Tree Labs (SG)	25:00	7	4	11
11	Avatar Hubo (US)	25:00	6	3.5	9.5
12	Last Mile (JP)	25:00	5	4	9

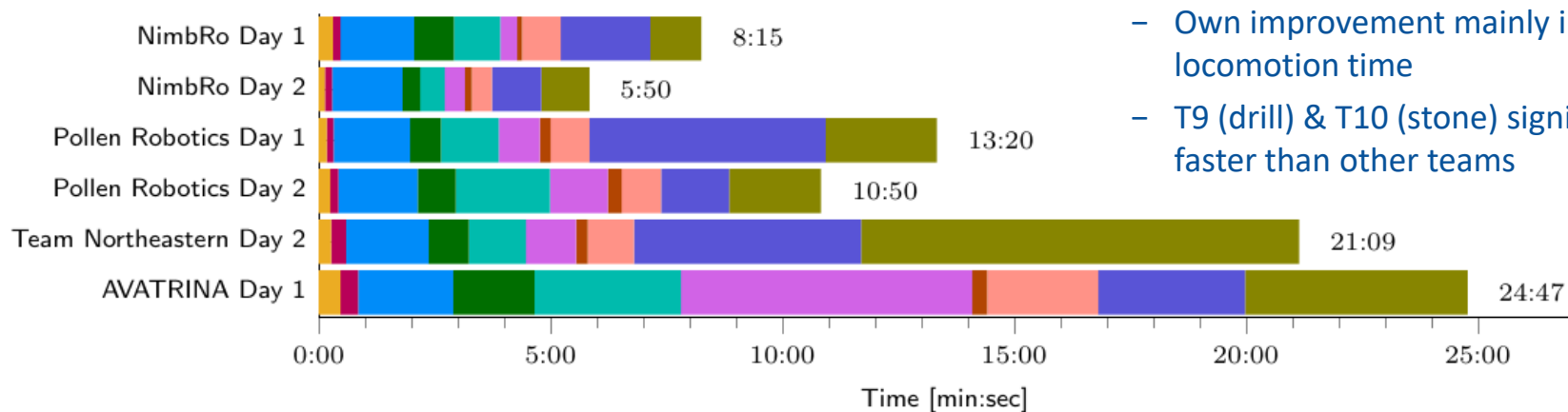
[XPRIZE]

Team NimbRo



Finals Timings

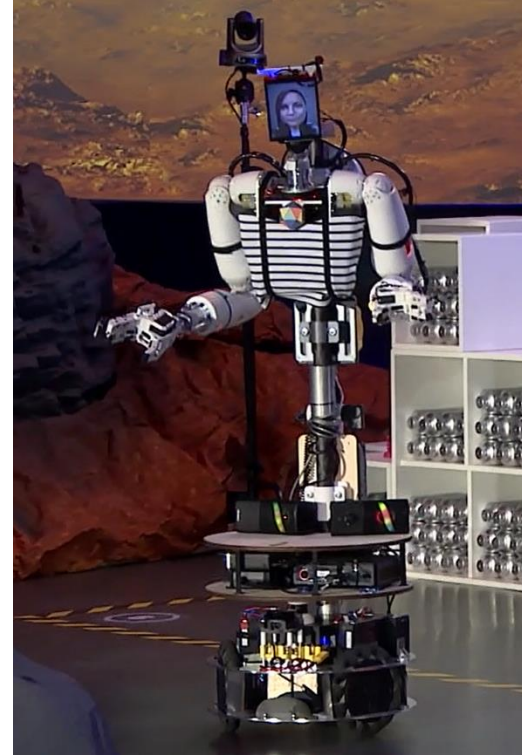
Team	Day	Time ¹ [mm:ss]											Total
		Start ²	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	
NimbRo	1	00:00	00:18	00:10	01:35	00:52	01:00	00:22	00:06	00:50	01:56	01:06	08:15
	2	00:00	00:08	00:09	01:31	00:23	00:32	00:26	00:09	00:26	01:04	01:02	05:50
Pollen Robotics	1	00:00	00:10	00:09	01:39	00:40	01:15	00:53	00:14	00:50	05:06	02:24	13:20
	2	00:00	00:15	00:09	01:43	00:49	02:02	01:15	00:18	00:51	01:28	01:59	10:50
Team Northeastern [25]	1	00:00	00:33	00:24	02:08	01:43	04:03	01:27	00:36	01:56			12:50
	2	00:00	00:16	00:19	01:47	00:52	01:14	01:05	00:15	01:00	04:54	09:27	21:09
AVATRINA [26]	1	00:00	00:28	00:23	02:03	01:45	03:10	06:17	00:19	02:24	03:10	04:48	24:47
	2	00:00	00:24	00:12	01:39	01:05	02:50	00:48	00:11	01:30	02:43		11:22
i-Botics [51]	1	00:00	00:13	00:26	01:23	01:53	01:57	01:52	02:07	02:57	09:47		22:35
	2	00:00	00:19	00:12	01:36	03:25							05:32



- Own improvement mainly in locomotion time
- T9 (drill) & T10 (stone) significantly faster than other teams

2nd Place: Pollen Robotics

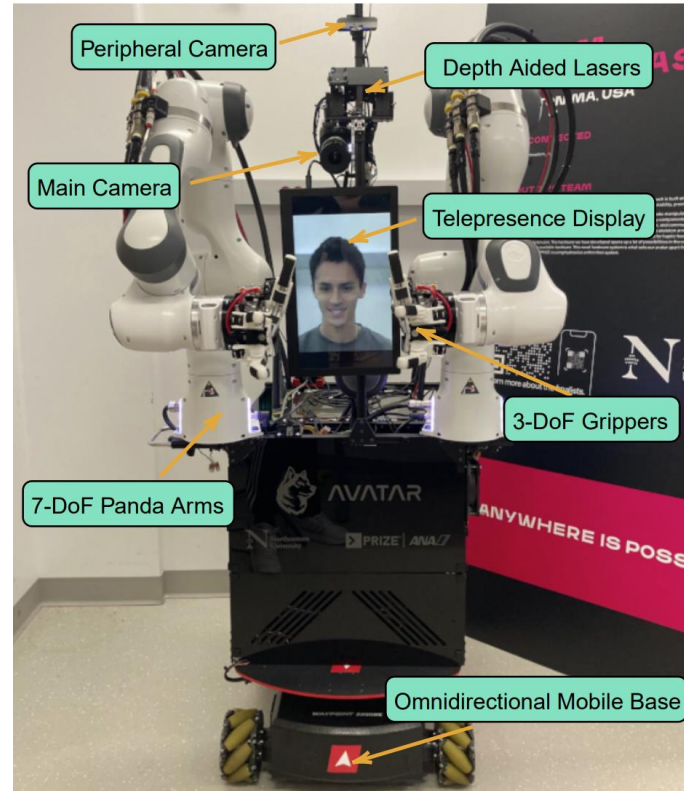
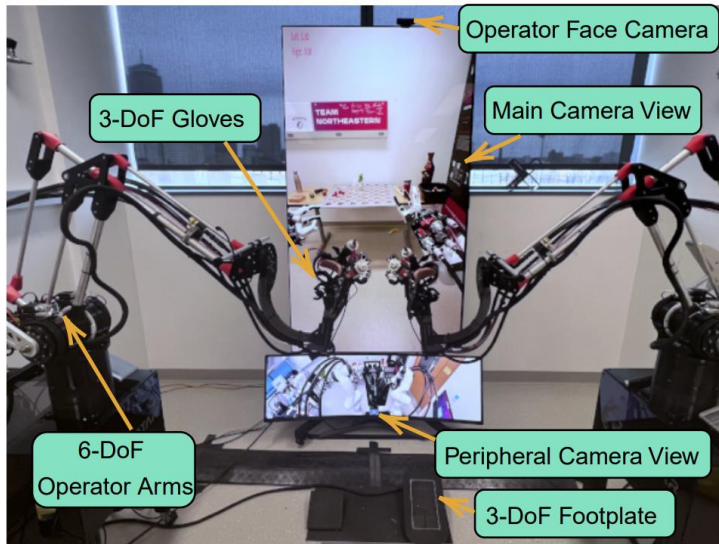
- Cost-effective design of robot and operator station
- Human-like upper body with Orbita 3 DoF actuators in shoulder, wrist, and neck



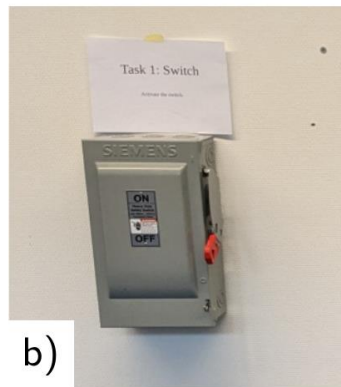
[Pollen Robotics]

3rd Place: Team Northeastern

- Hydraulically actuated glove-gripper pair for haptic force feedback
- Non-immersive visualization with two monitors
- Projected laser lines aid 3D perception



After the Competition: User Study in our Lab

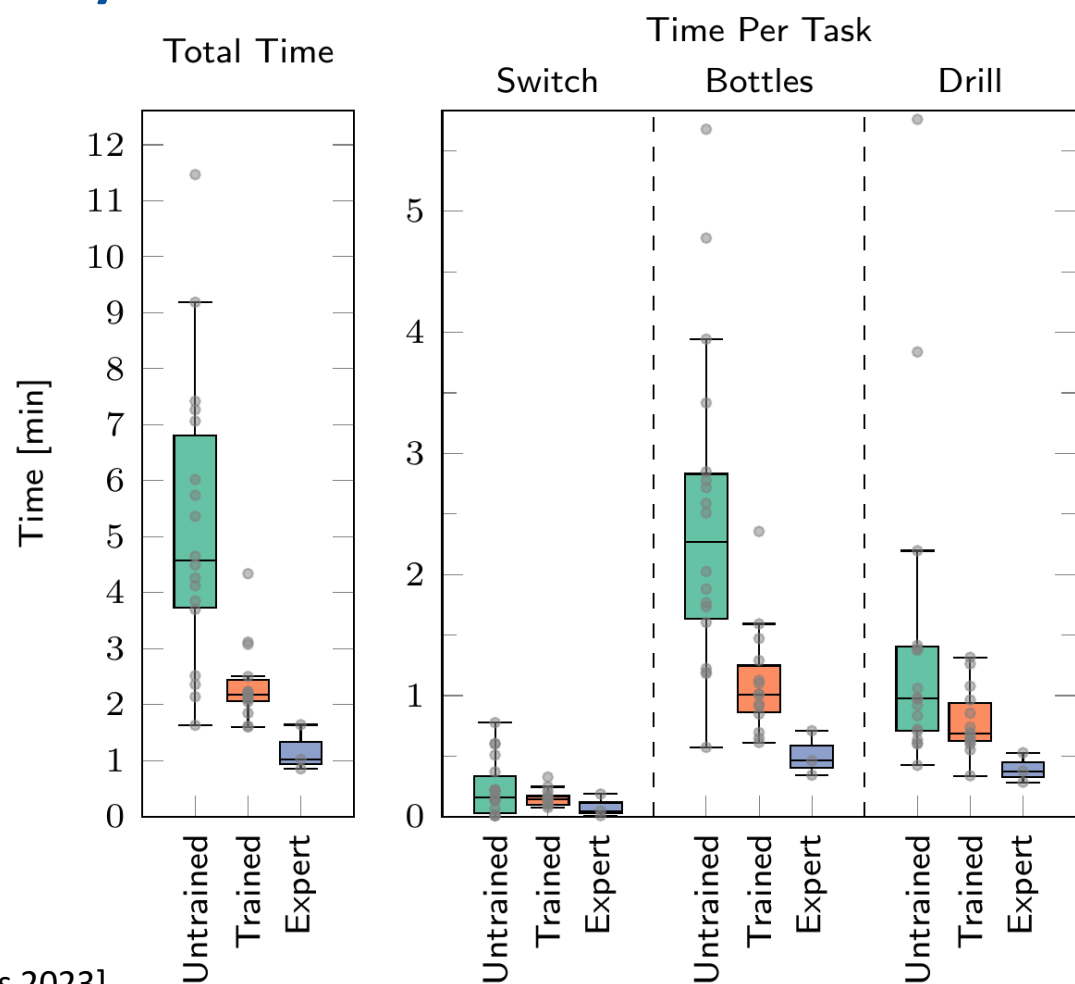


- Three tasks, similar to finals: a),b) Locomotion+Switch, c) Bottles, d) Drill
- 35 participants, 32 with no prior experience of the system
- All participants: 2 min intro video explaining the system (task agnostic)
- Three groups:
 - No training
 - 10 min task training, similar to Finals
 - Expert team members

After the Competition: User Study in our Lab

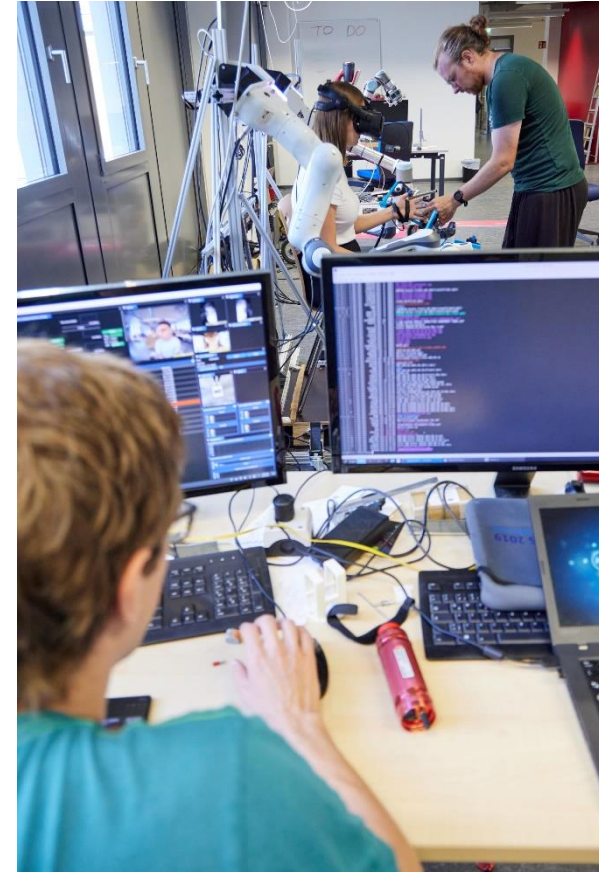
- Unsurprising: Clear advantage of training (2× over untrained)
- Unsurprising: Expert operators are very fast (2× trained)
- Untrained operators could still solve all tasks in reasonable time
- All participants were able to solve the tasks

=> System is very intuitive, but short instruction on tasks improves completion time.



Lessons Learned

- Robustness is key
- Latency is the enemy of direct teleoperation
- Frequent testing under competition conditions is essential: System & team!
- 1:1 correspondence is best
- 6D head motion simplifies manipulation control
- Sparse immersive control overlays – don't break immersion!
- Facial animation and gestures: Head & gaze direction enables shared awareness
- Had to modify components



[Photographer: Volker Lannert]

What is Next?

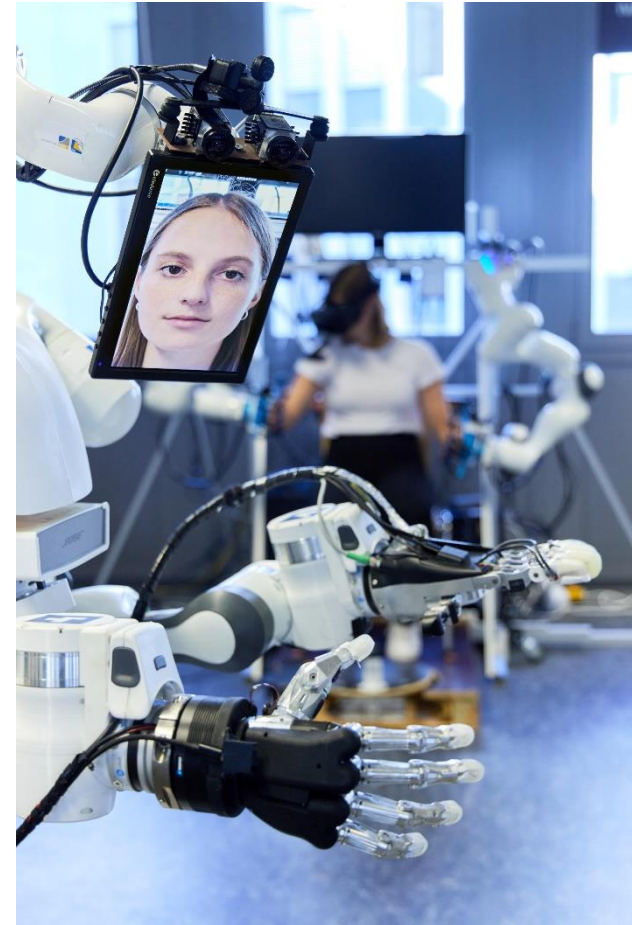
- Transfer to real applications
 - Complex avatar systems could be further developed e.g. for
 - Dangerous or hard-to-reach domains,
 - Disaster relief,
 - Medical assistance in isolation wards
 - Everyday virtual travel requires simpler and more affordable systems
- Research questions include
 - How much human-likeness avatars should assume?
 - How to address latencies and bandwidth limitations?
 - How to balance and interface direct control and autonomy?



[Photographer: Volker Lannert]

Motivation for Autonomy

- Longer latencies require less direct control
 - Use autonomous skills, such as grasping an object or navigating to a waypoint
 - Shared autonomy where the operator controls high-level behavior and autonomy fills-in the low-level details (horse metaphor, Flemisch 2003)
- Operator might not always be available
 - 1:1 control often too costly
=> one operator must supervise many robots
 - Issues of privacy and of being in operator's dept
- AI: Understanding intelligence by creating intelligent artefacts

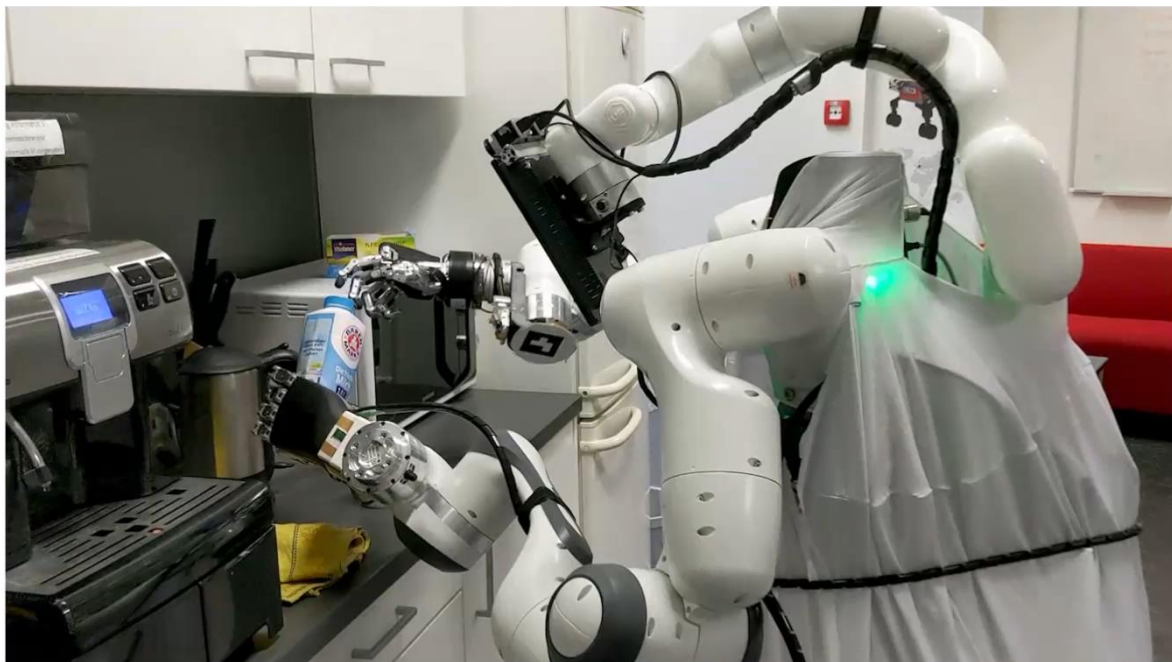


[Photographer: Volker Lannert]

Unmatched Human Operators



- Humans can solve many tasks by teleoperation
 - Can cope with novel situations, quickly learn new tasks
 - Recognize and mitigate errors
- Far beyond the capabilities of autonomous robots



2x

Human Cognitive System

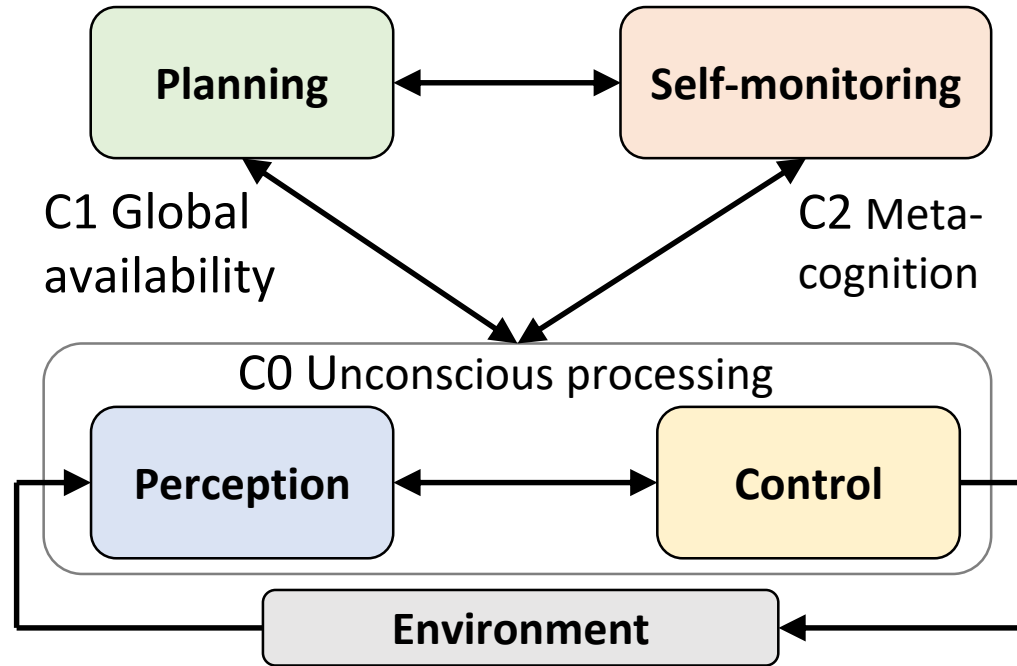
- Cognitive architecture of the human mind has evolved to continuously interact with changing environments and self-monitor

System 2

- slow, serial
- flexible
- **conscious**

System 1

- fast, parallel
- habitual
- unconscious



Cognitive functions according to Kahneman (2011) and Dehaene (2017)

My Objective

- Develop methods for learning perception and planning for service robots, which go beyond unconscious routine tasks by incorporating **conscious processing** to cope with novel situations and self-monitor



Overall Approach

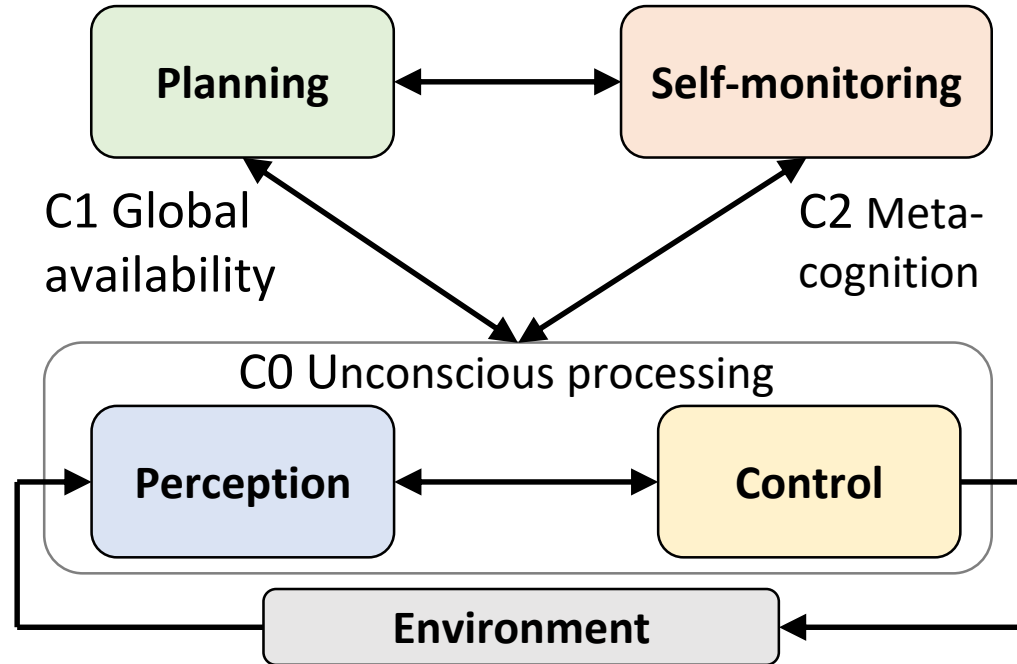
- Equip service robots with key elements of human cognitive architecture
- **Bottom-up** approach ensures **grounding** of conscious processing

System 2

- slow, serial
- flexible
- conscious

System 1

- fast, parallel
- habitual
- unconscious



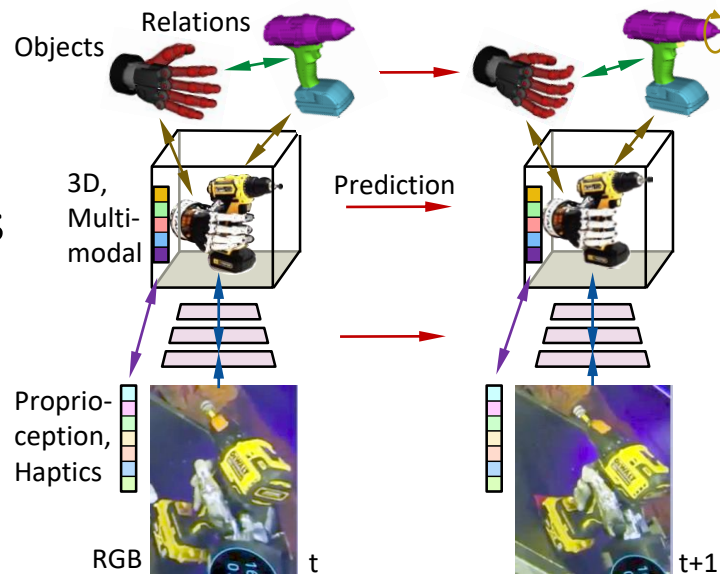
Cognitive functions according to Kahneman (2011) and Dehaene (2017)

Unconscious Perception & Tracking

1. Learning hierarchical representations
2. Learning 3D multimodal scene models
3. Learning object models & relations
4. Learning prediction and tracking

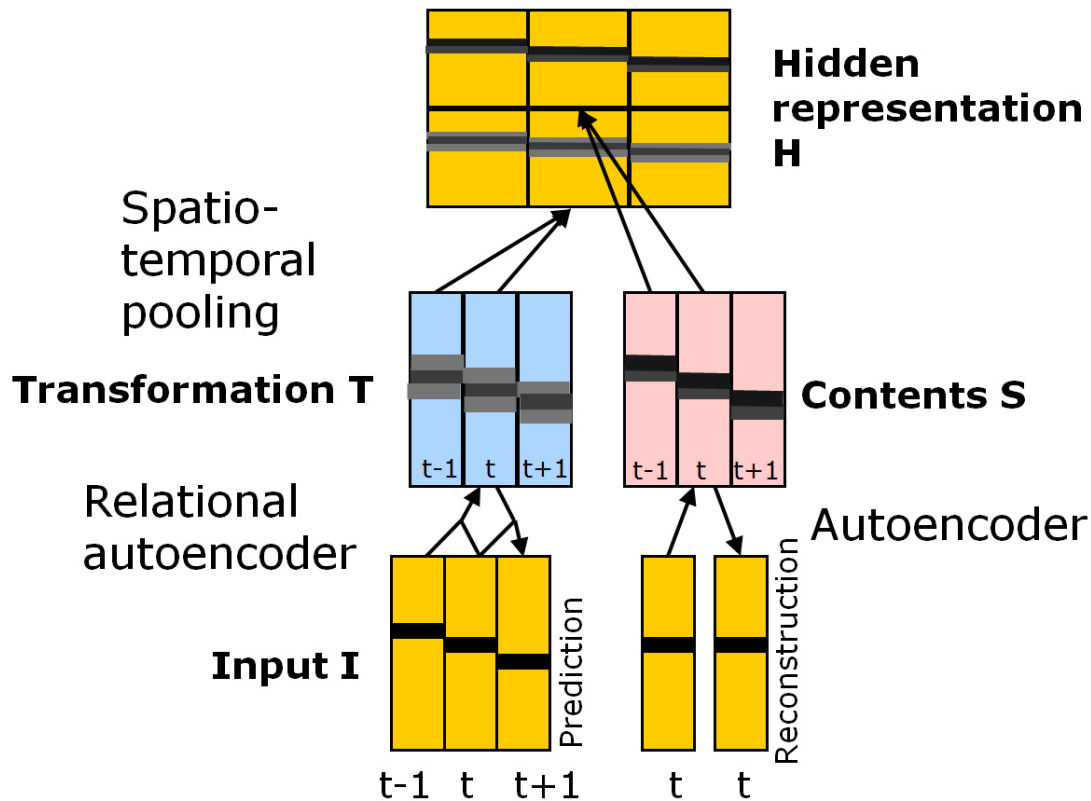
■ Scene compositionality

- Objects and scenes described by their constituent parts and their relations
 - Infinite variants from a finite set of building blocks
- Exploit inductive biases like canonical frames, 3D projective geometry, camera motion, object relations, compositional structure, hierarchical categorization, ...



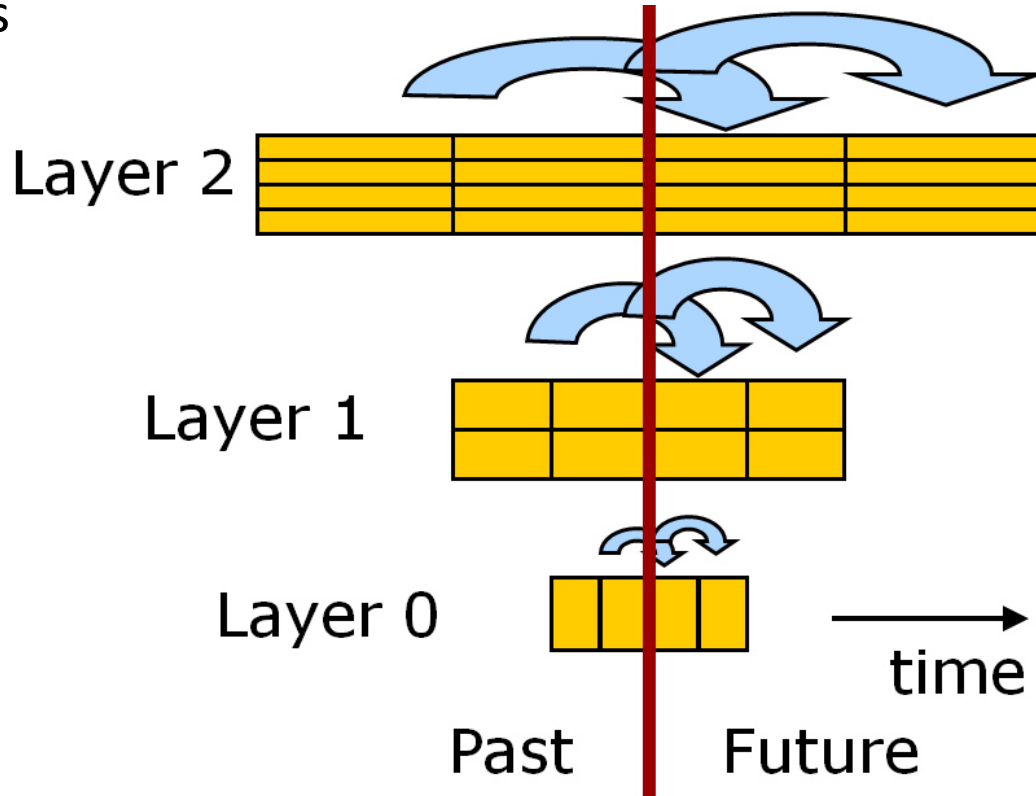
Learning of Hierarchical Representations for Prediction

Local learning module



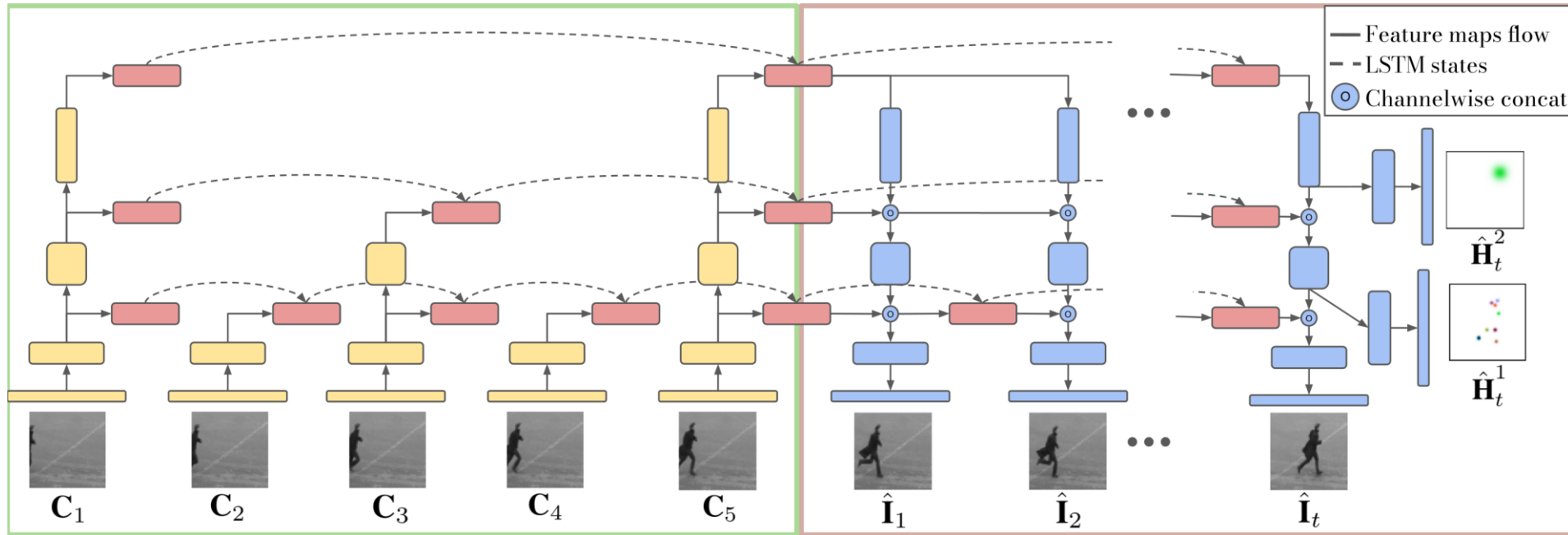
Learning of Hierarchical Representations for Prediction

- Coarser, more abstract predictions for longer time horizons in higher layers



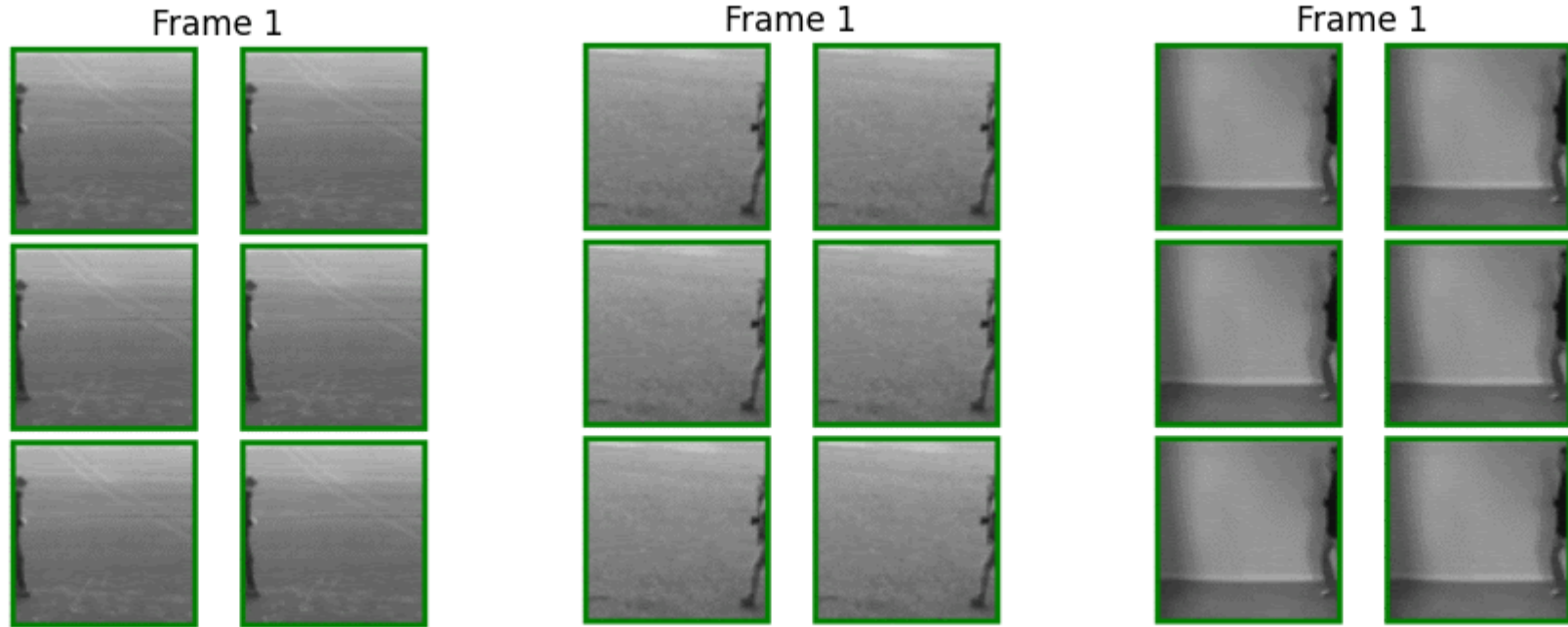
MSPred: Video Prediction at Multiple Spatio-Temporal Scales

- Coarser, more abstract predictions for longer time horizons in higher layers
- Predict image itself, human pose joint keypoints, and human body position

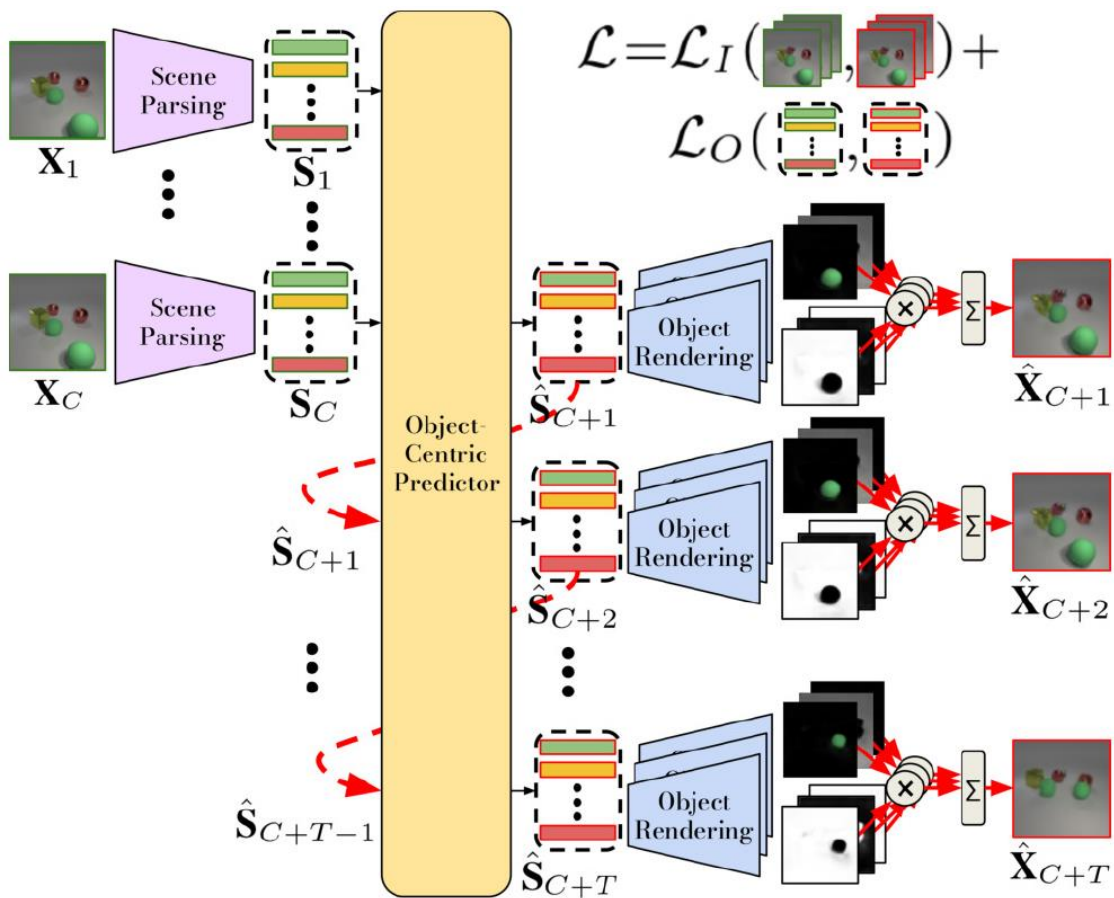


MSPred: Video Prediction at Multiple Spatio-Temporal Scales

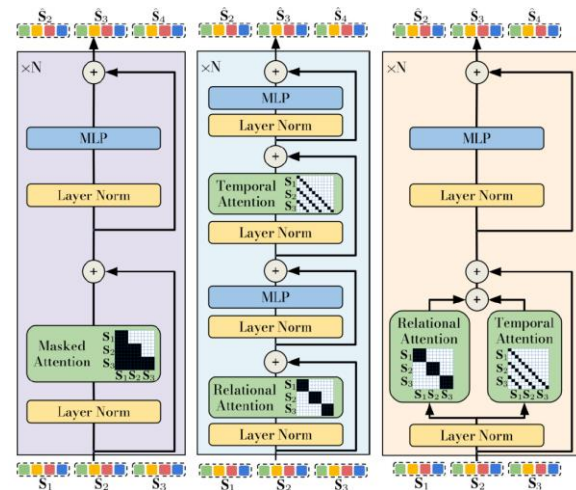
- Coarser, more abstract predictions for longer time horizons in higher layers
- Predict image itself, human pose joint keypoints, and human body position



Object-centric Video Prediction Decoupling Dynamics and Interaction



- Scene parsing into object slots
- Video synthesis from objects and masks
- Predictor decouples temporal and relational attention



Object-centric Video Prediction Data Sets

Obj3D

- Synthetic 3D objects
- Ball colliding with static objects
- Given 5 frames, predict next 5



MOVi-A

- Synthetic 3D objects
- Complex dynamics and occlusions
- Given 6 frames, predict next 8

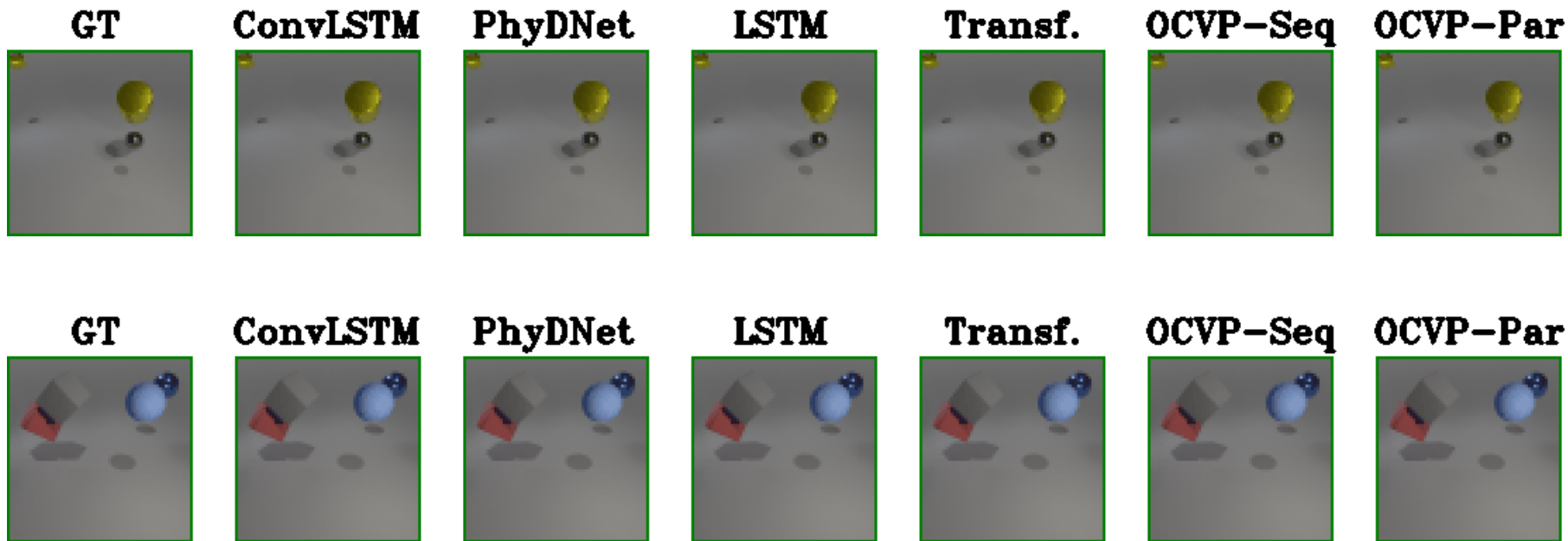


Object-centric Video Prediction: Obj3D



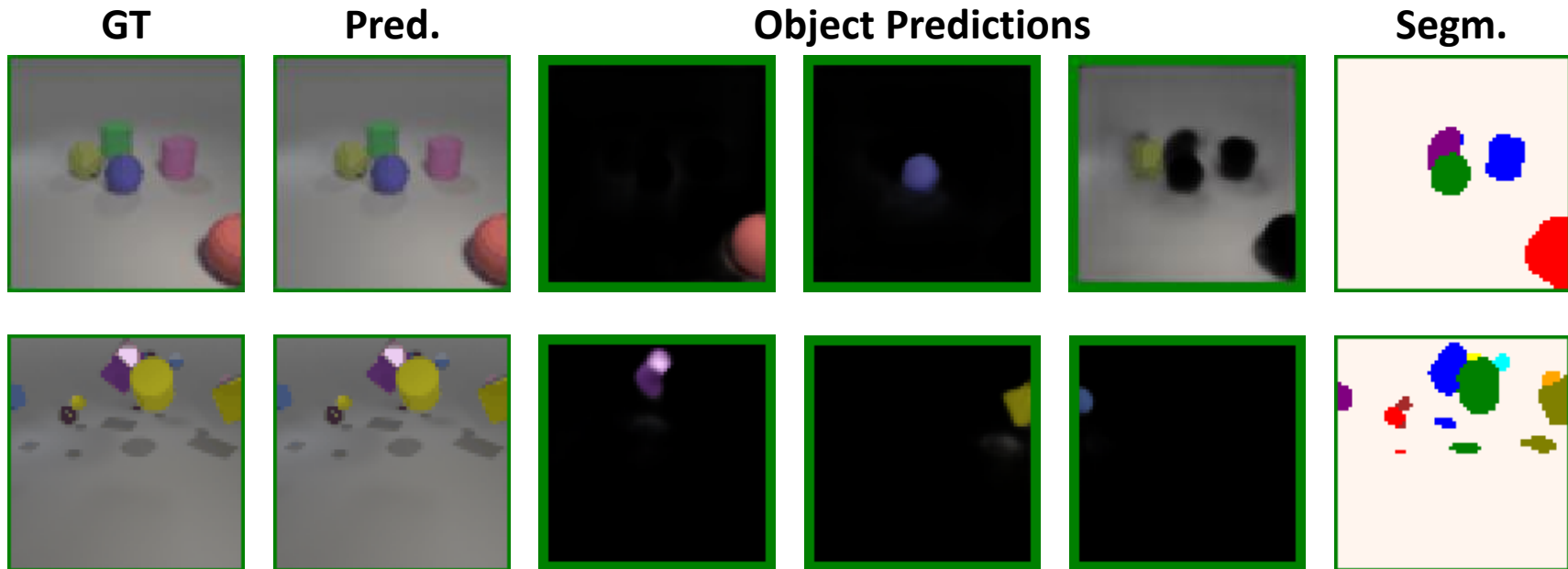
[Villar-Corrales et al. ICIP 2023]

Object-centric Video Prediction: MOVi-A



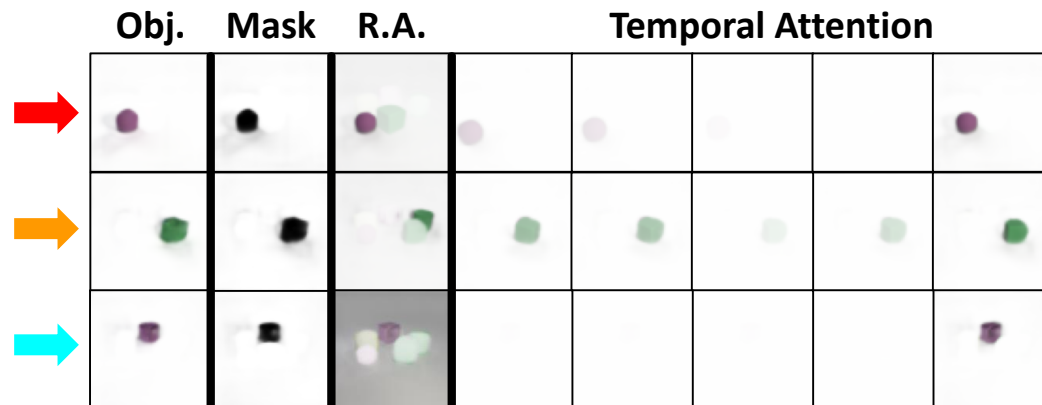
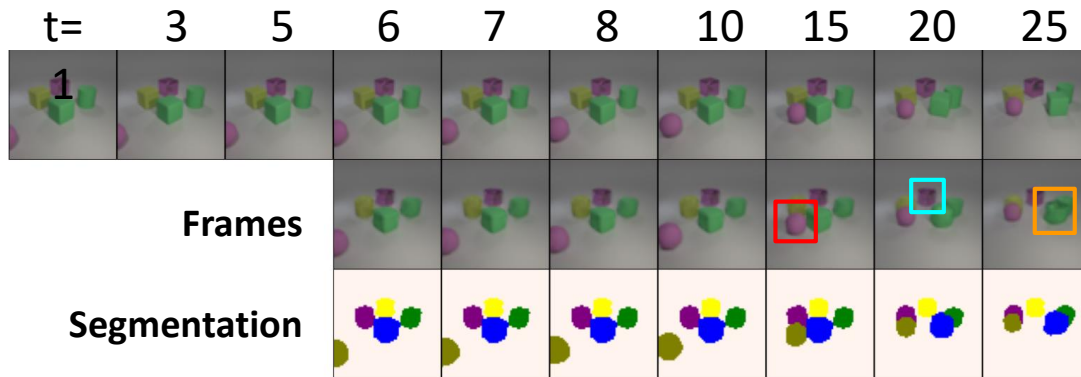
[Villar-Corrales et al. ICIP 2023]

Object-centric Video Prediction: Object Predictions



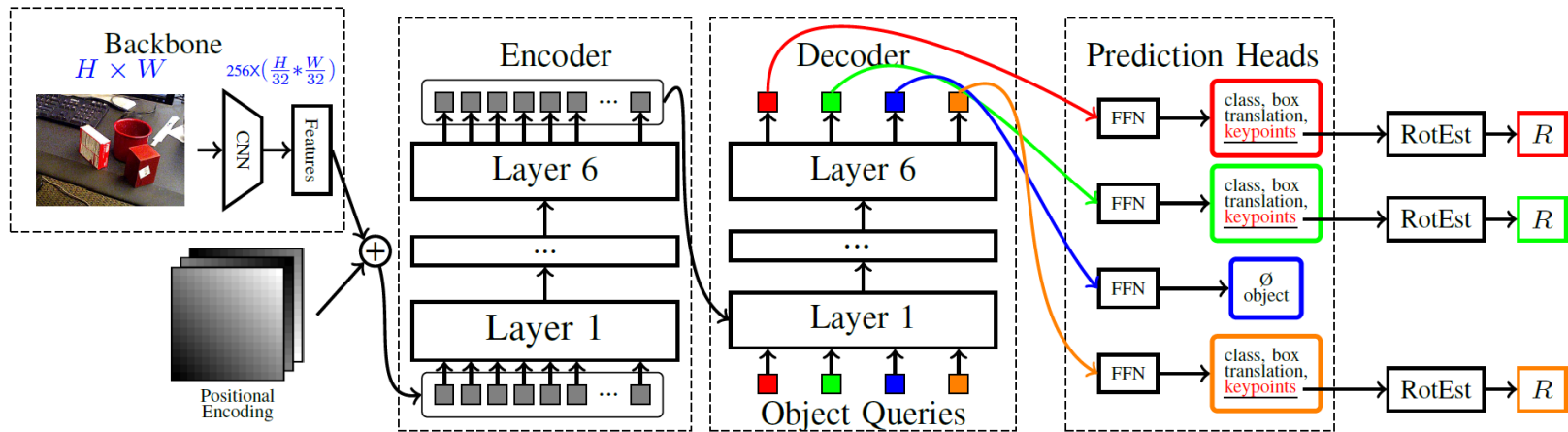
[Villar-Corrales et al. ICIP 2023]

Object-centric Video Prediction: Model Interpretability



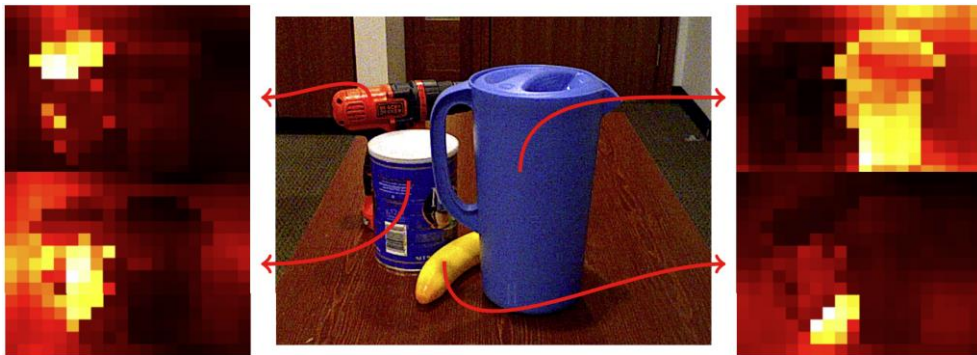
[Villar-Corrales et al. ICIP 2023]

Multi-Object 6D Pose Estimation using Keypoint Regression

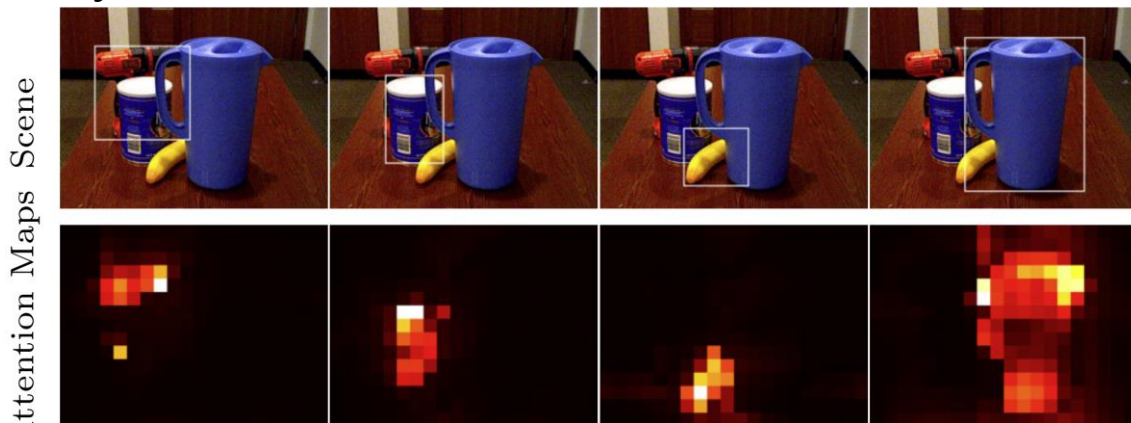


YOLOPose: Multi-Object 6D Pose Estimation using Keypoint Regression

Encoder self-attention

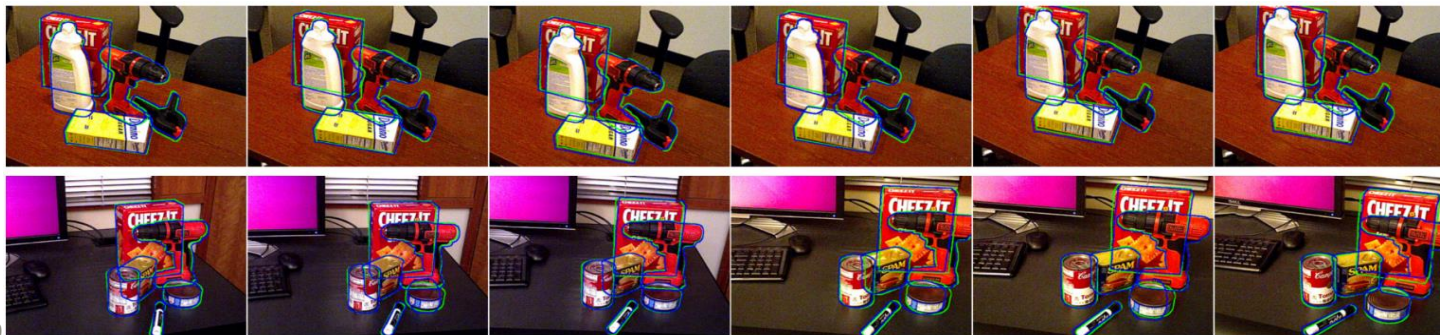
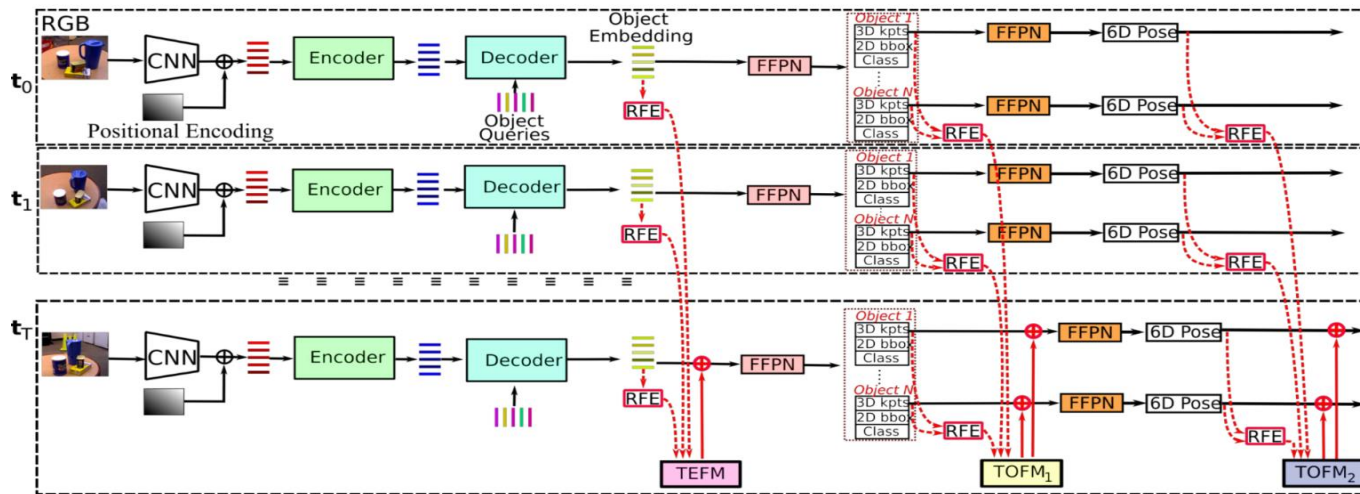


Object detections and decoder cross-attention



MOTPose: Attention-based Temporal Fusion for Multi-object 6D Pose Estimation

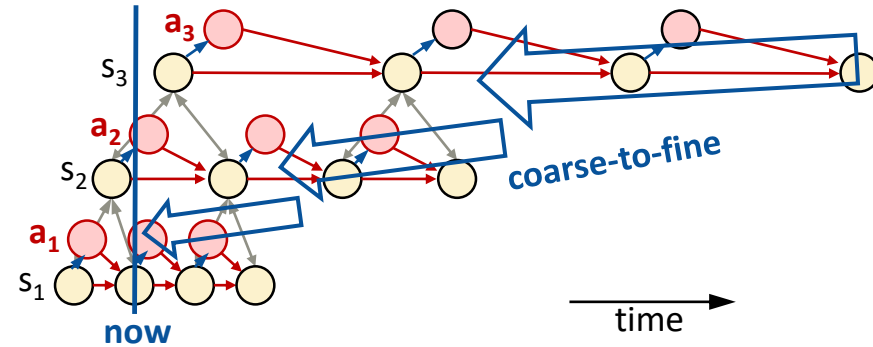
- Propagating object embeddings, object descriptors, and poses



[Periyasamy, 2023]

Unconscious Prediction and Control

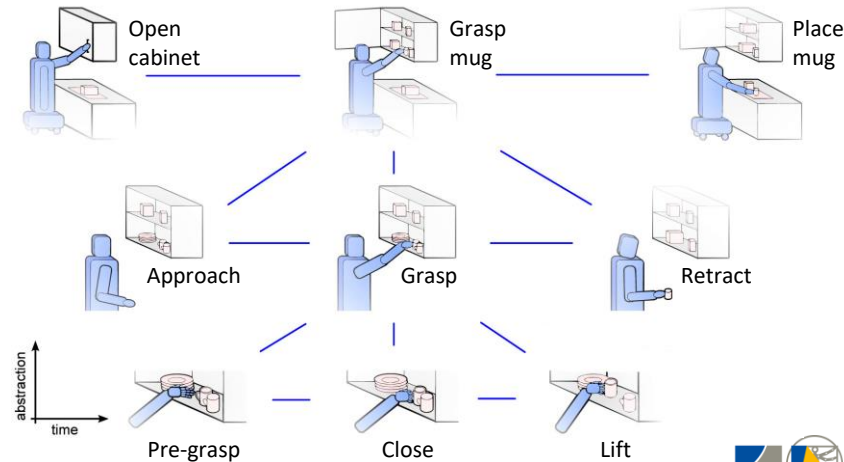
1. Learning action-conditioned prediction
2. Learning to control in the now
3. Learning reusable skills
4. Learning from imitation and real-robot experience



■ Action compositionality

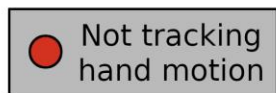
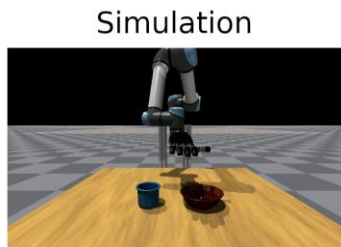
- Activities consists of sequence of actions, which can be decomposed into movement primitives

- Exploiting inductive biases like hierarchical structure, object binding, planning in the now, ...

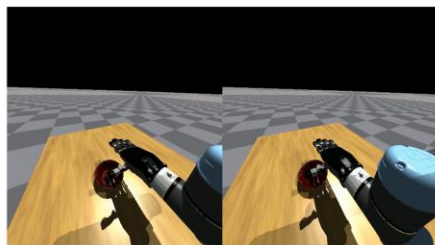


Immersive Manipulation Demonstration

- Immersive VR visualization
- Hand and finger tracking
- Haptic feedback



HMD View



Operator



Operator

Headset



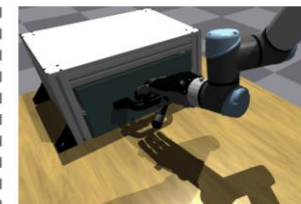
Hand



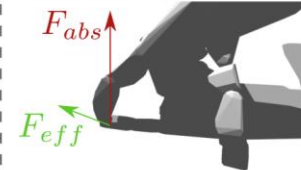
Markov decision process $\rightarrow \frac{90}{c}$ Hz

Physics simulation $\rightarrow 90$ Hz

VR camera



Contact forces



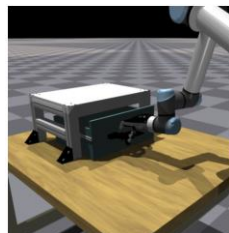
update pose \rightarrow

\leftarrow update view

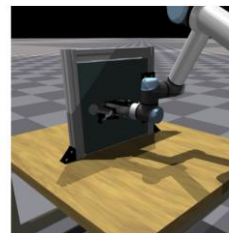
$a_t \rightarrow$

\leftarrow haptic feedback

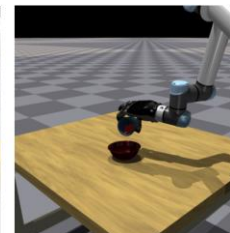
■ Four tasks



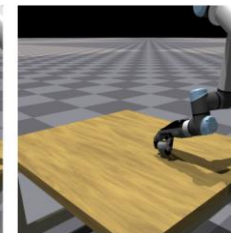
OpenDrawer



OpenDoor

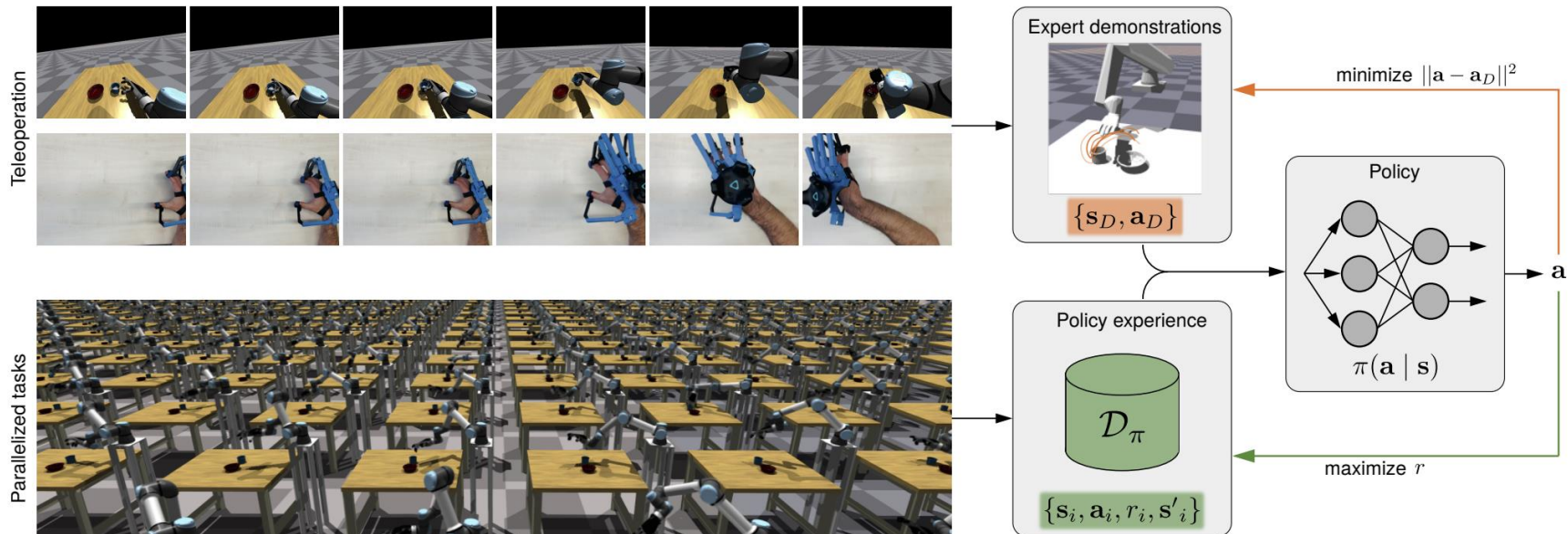


PourCup



LiftObject

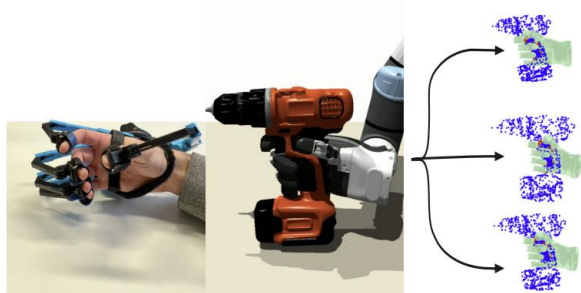
Accelerating Interactive Human-like Manipulation Learning with GPU-based Simulation and High-quality Demonstrations



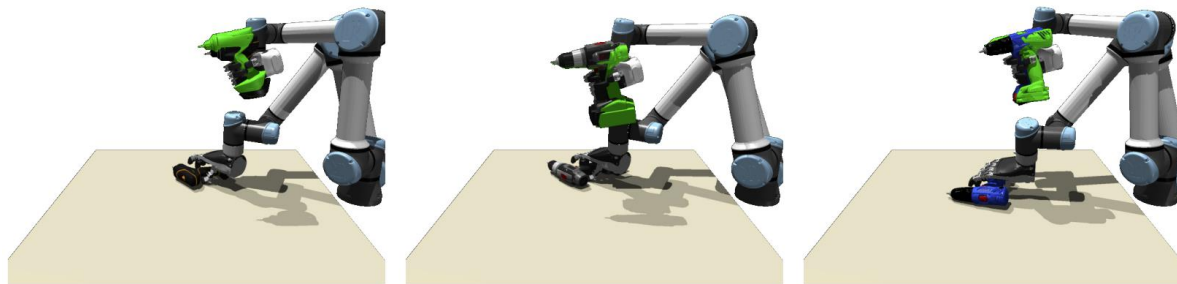
Method	OpenDrawer	OpenDoor	PourCup	LiftObject
BC	1.0 ± 0.0	0.96 ± 0.02	0.76 ± 0.23	0.27 ± 0.03
PPO-dense	1.0 ± 0.0	1.0 ± 0.0	0.98 ± 0.01	0.97 ± 0.08
PPO-sparse	1.0 ± 0.0	0.0 ± 0.0	0.48 ± 0.48	0.14 ± 0.33
DAPG-sparse	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0

Learning Interactive Functional Grasping

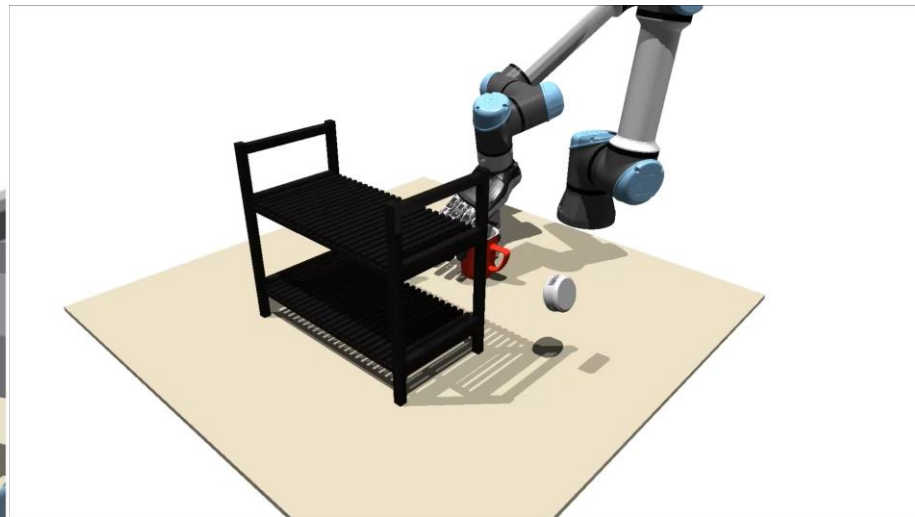
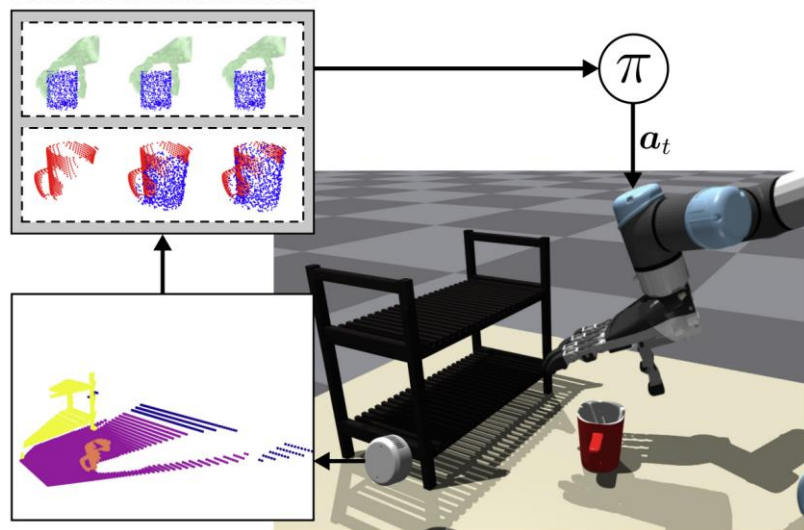
Generalization of a single demonstration



Interactive operation of unseen tools

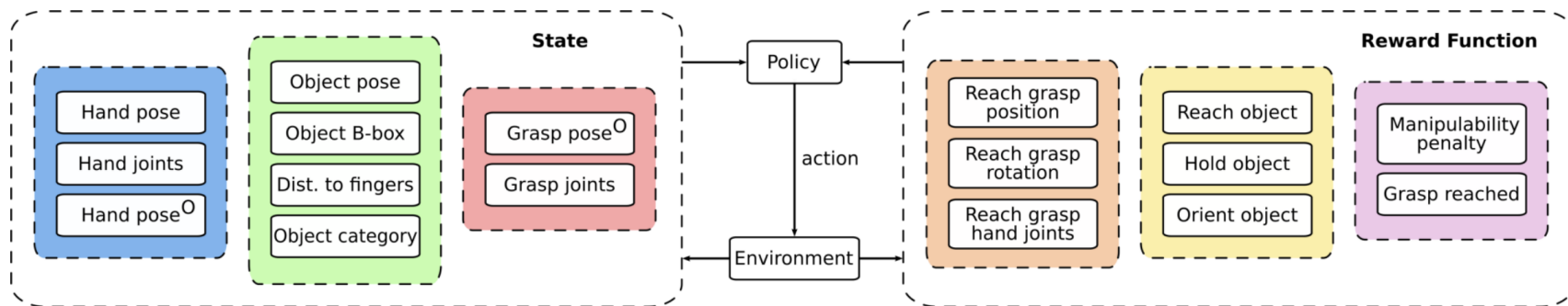


Generalized Demonstration

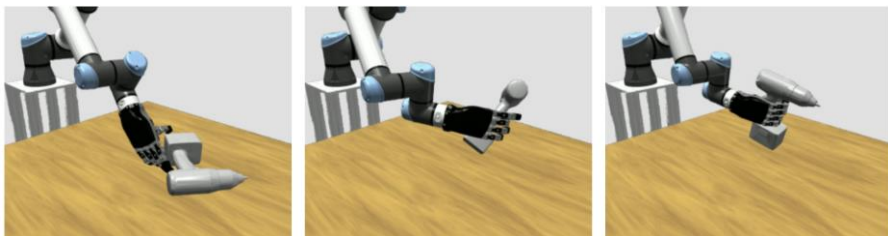


[Mosbach and Behnke CASE 2023, Best Paper Award]

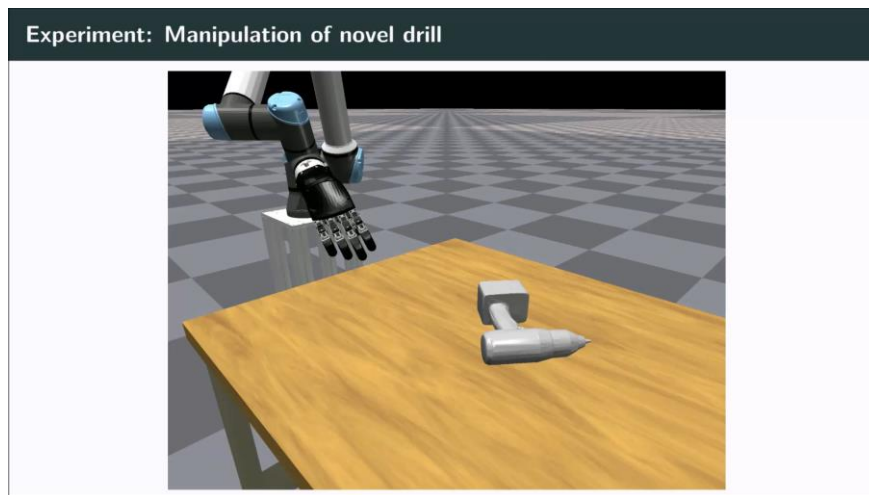
Learning Pre-grasp Manipulation for Human-like Functional Grasping



- Dense multi-component reward function encodes desired functional grasp

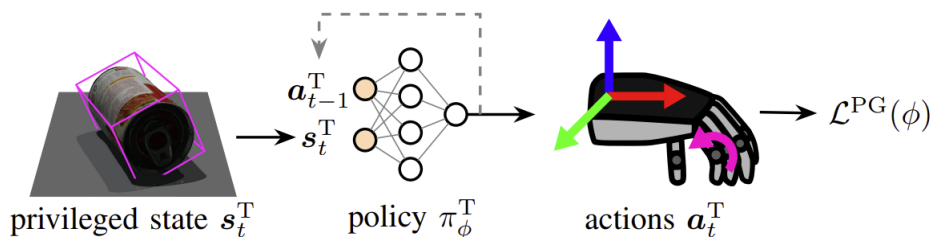


- Learns to reposition and reorient objects to achieve functional grasps

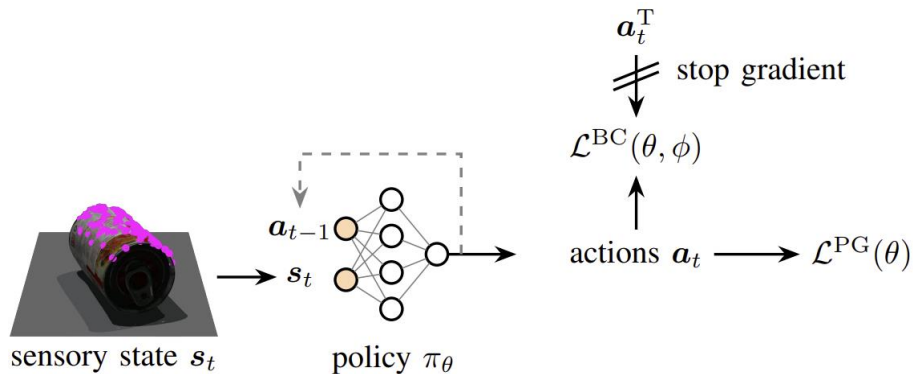


Grasp Anything: Augmenting Reinforcement Learning with Instance Segmentation to Grasp Arbitrary Objects

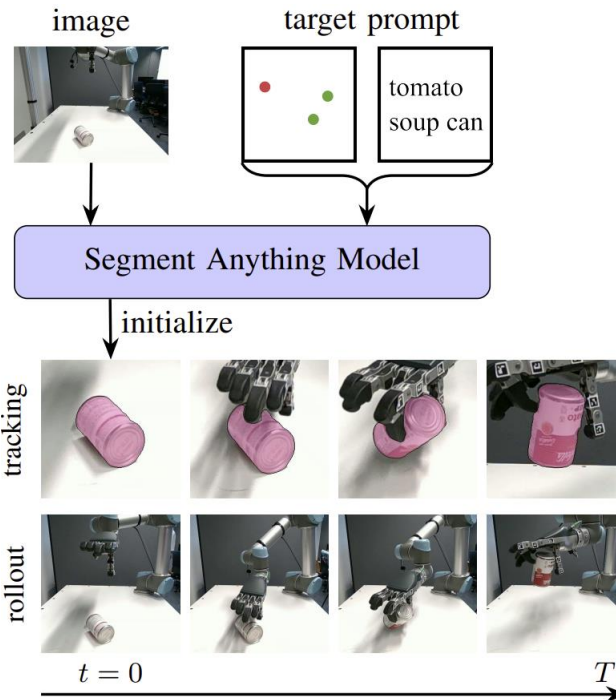
Teacher training



Teacher-guided sensorimotor learning

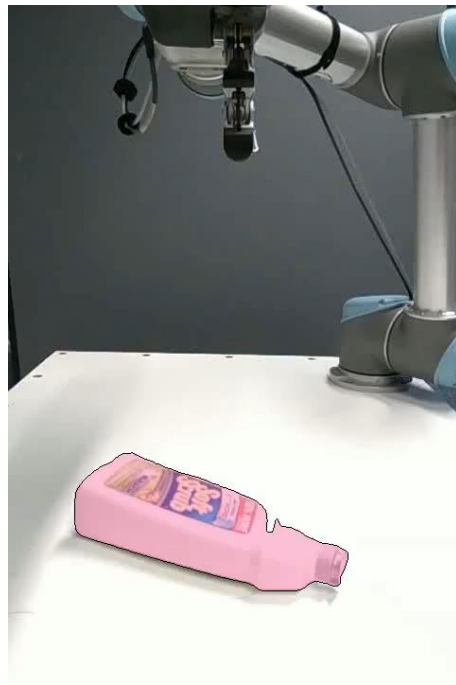
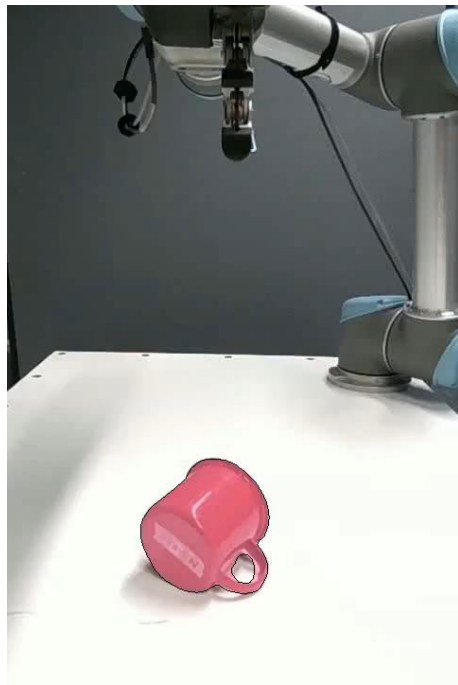
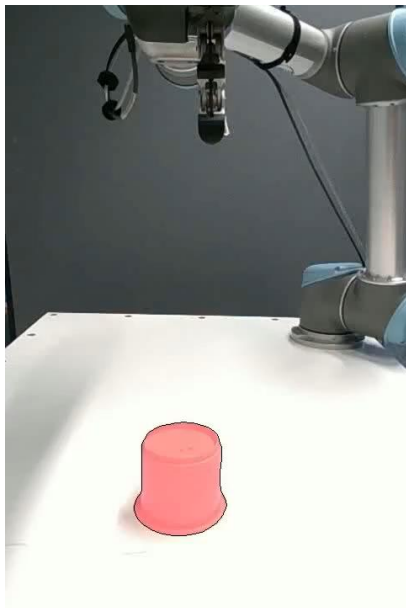


Real-world deployment of promptable grasping policy



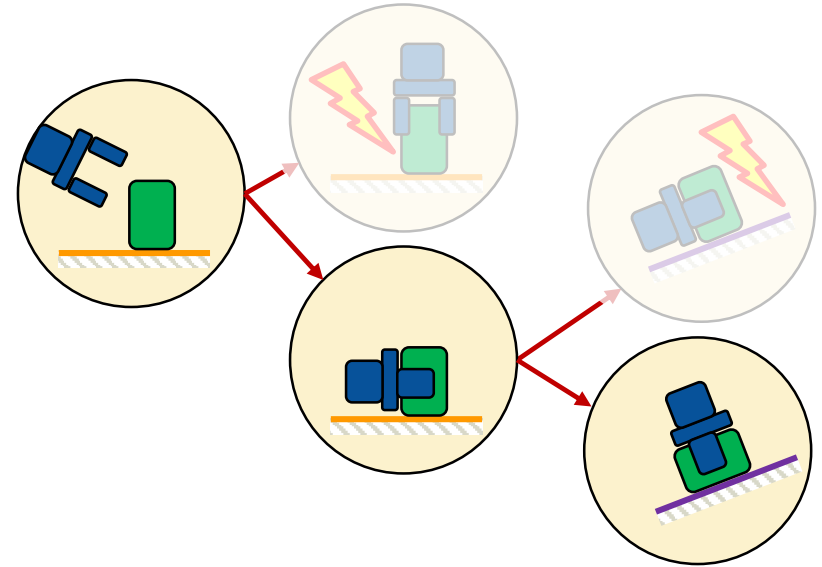
Grasp Anything: Augmenting Reinforcement Learning with Instance Segmentation to Grasp Arbitrary Objects

- Learned policy with improved object visibility is real-world deployable



Conscious Prediction and Planning

1. Learning a working memory
2. Learning working memory predictions
3. Learning conscious planning
4. Learning new conscious concepts



■ Systematic generalization

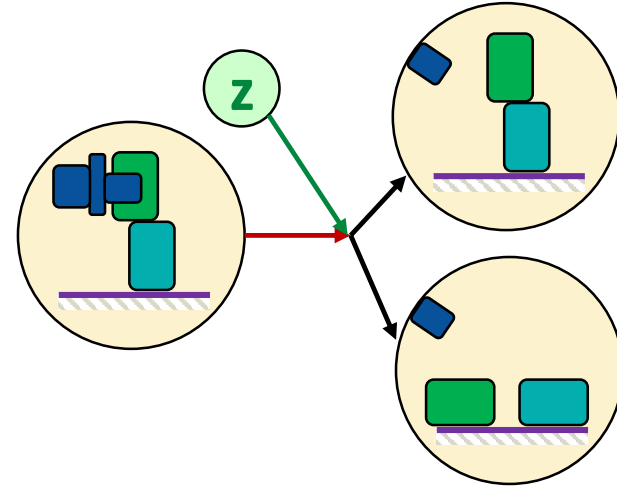
- Reuse task knowledge in infinitely many novel situations in which irrelevant items change

■ Working memory as communication bottleneck

- Focus on few items, ignore all others which are irrelevant for the task
- Must combine multiple lower-level items to larger, composite items

Conscious Self-monitoring

1. Representing uncertainty
2. Predicting multiple plausible futures
3. Error detection and mitigation
4. Interactive learning



■ Self-aware

- Being aware of own capabilities and limitations, dangers, etc.

■ Systematically model and use uncertainty

- Collect more information when needed
- Avoid dangerous situations
- Detect System 1 errors and mitigate them

Potential Impact

Consciousness is not a bug, but a feature!

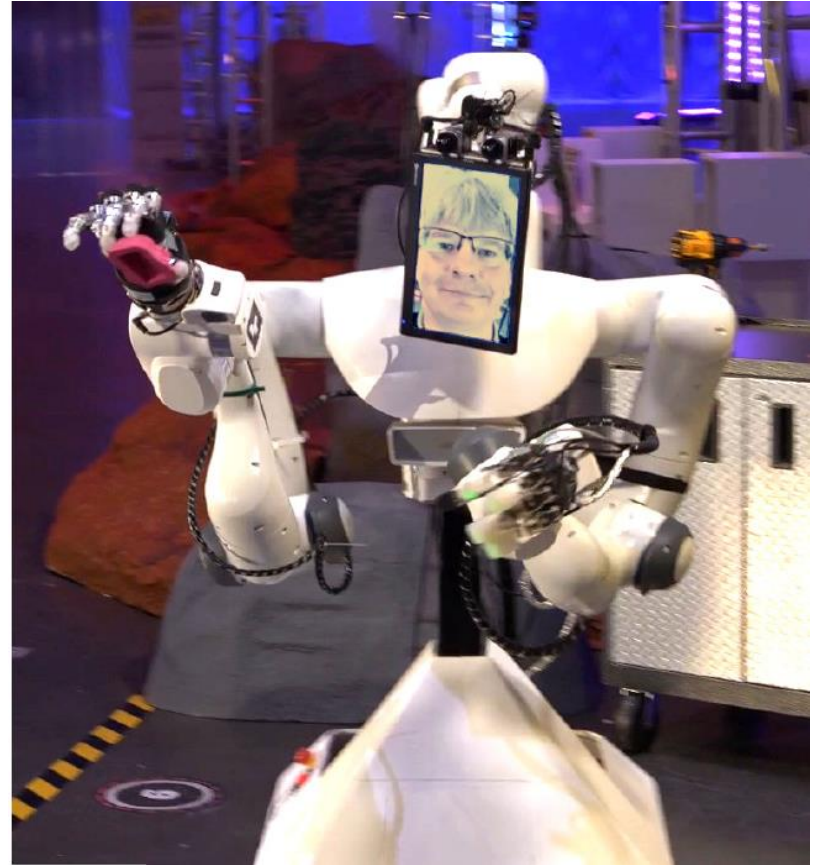
- Will bring service robots to the next level
 - **Systematically generalize** skills and cope with novel situations
 - **Self-monitor** perceptions and actions: obtain more information when needed, avoid risks, detect errors, and mitigate them
- Consciousness-inspired robots will have a high impact on economy and society since they will be **applicable to a large variety of open-ended domains**
- Will enable the creation of **personal service robots** which have the potential to change our society to the same degree personal and mobile computers changed it in the last decades



© DALL·E 2

Conclusions

- The ANA Avatar XPRIZE competition advanced immersive telepresence systems
- Potential follow-up could raise the bar
 - Bandwidth restrictions and latencies
 - Locomotion on more difficult terrain
 - More complex manipulation (e.g., bimanual tasks)
 - Additional interaction modalities (e.g., temperature or smell)
 - Deeper interactions between avatars and recipients including interpretation of subtle communication cues and direct physical contact
- More autonomy is needed
- Need to match human cognitive functions
- Demonstrations can guide RL
- Consciousness needed for systematic generalization and self-monitoring



[XPRIZE]

Questions?

Join the winning team!

We are hiring.

