Representation Learning and Activity Prediction from Video

Sven Behnke

Computer Science VI
Autonomous Intelligent Systems

Fadime Sener

Computer Science II
Visual Computing

UNIVERSITÄT BONN
AIS

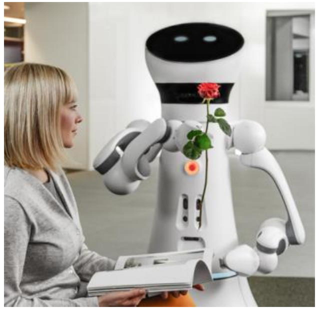
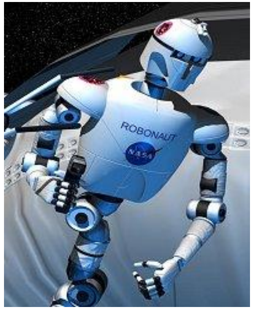
Many New Application Areas for Robots

- Self-driving cars
- Logistics
- Agriculture, mining
- Collaborative automation
- Personal assistance
- Space, search & rescue
- Healthcare
- Toys

Need more cognitive abilities!






Some of our Cognitive Robots

- Equipped with numerous sensors and actuators
- Complex demonstration scenarios

Soccer

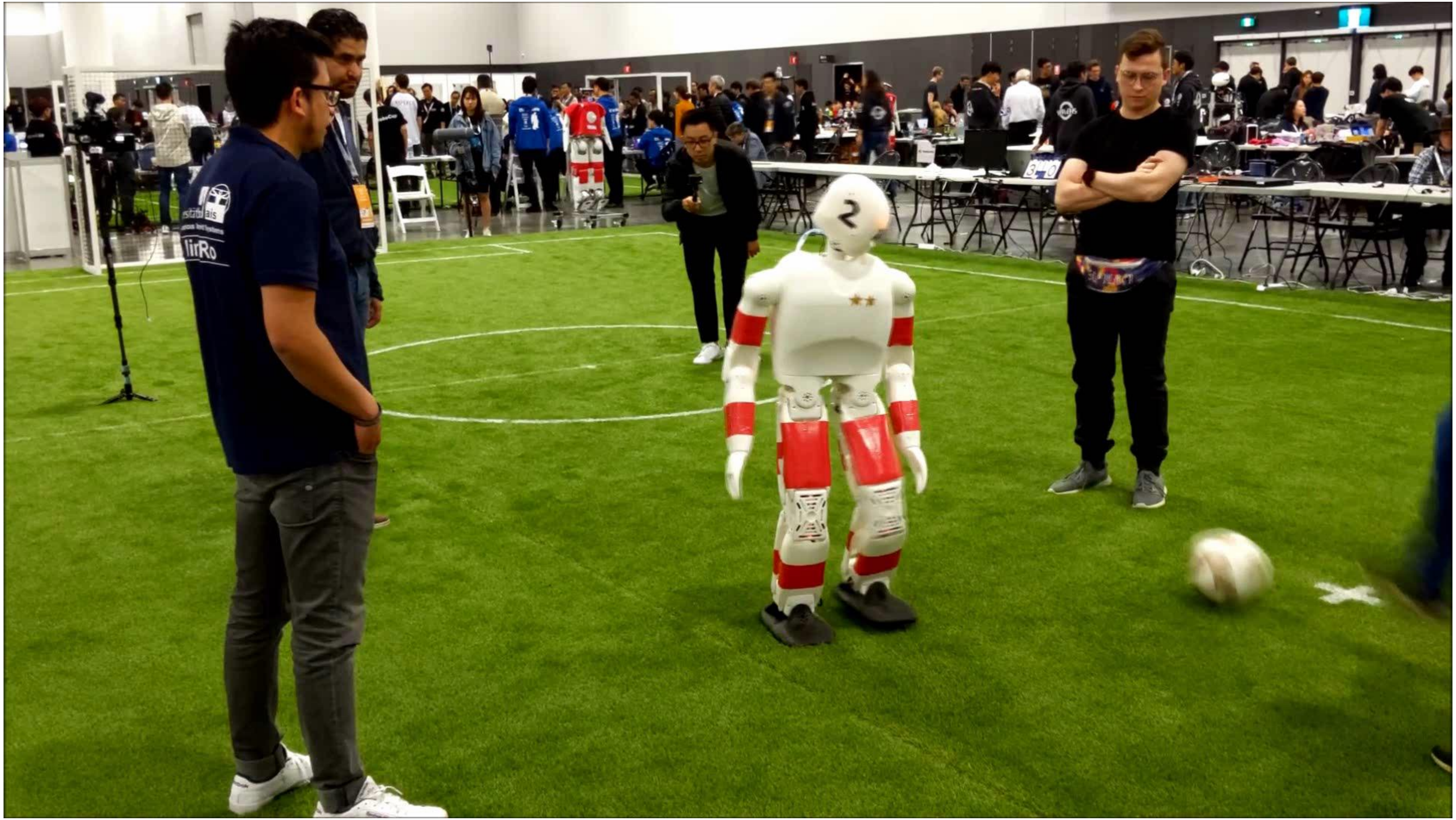
Domestic service

Mobile manipulation

Bin picking

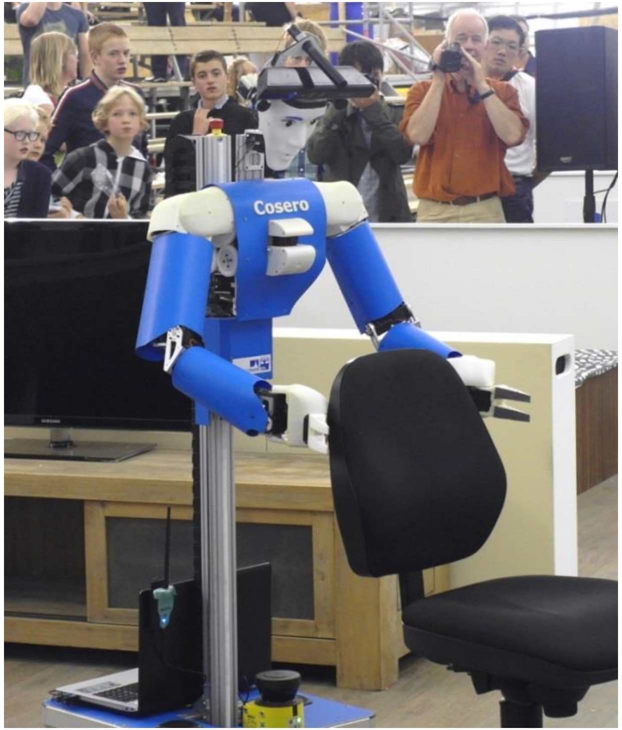
Aerial inspection

RoboCup 2019 in Sydney



Our Domestic Service Robots

Dynamaid



Cosero

- Size: 100-180 cm, weight: 30-35 kg
- 36 articulated joints
- PC, laser scanners, Kinect, microphone, …

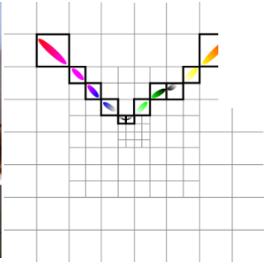
[Stückler et al.: Frontiers in Robotics and AI 2016]

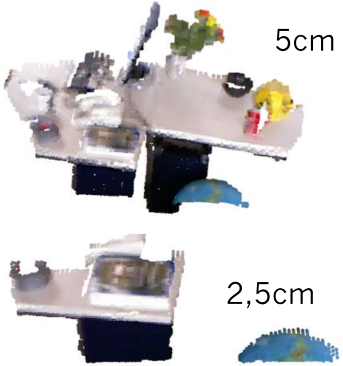
Cognitive Service Robot Cosero

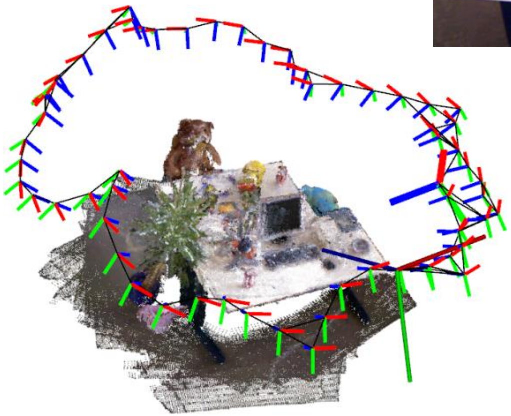
3D Mapping by RGB-D SLAM

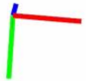
[Stückler, Behnke:
Journal of Visual Communication
and Image Representation 2013]

- Modelling of shape and color distributions in voxels
- Local multiresolution
- Efficient registration
  of views on CPU
- Global
  optimization

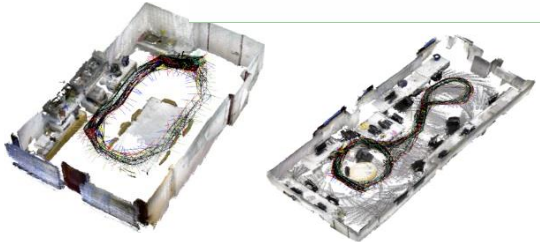








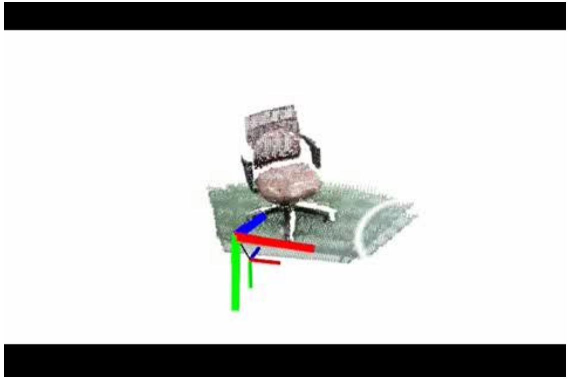
- Multi-camera SLAM



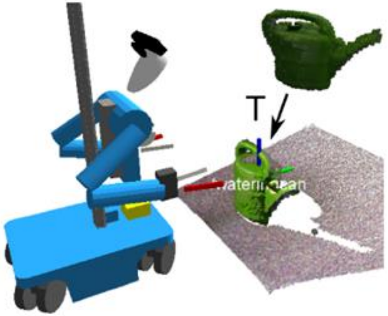
[Stoucken]

Learning and Tracking Object Models

- Modeling of objects by RGB-D-SLAM

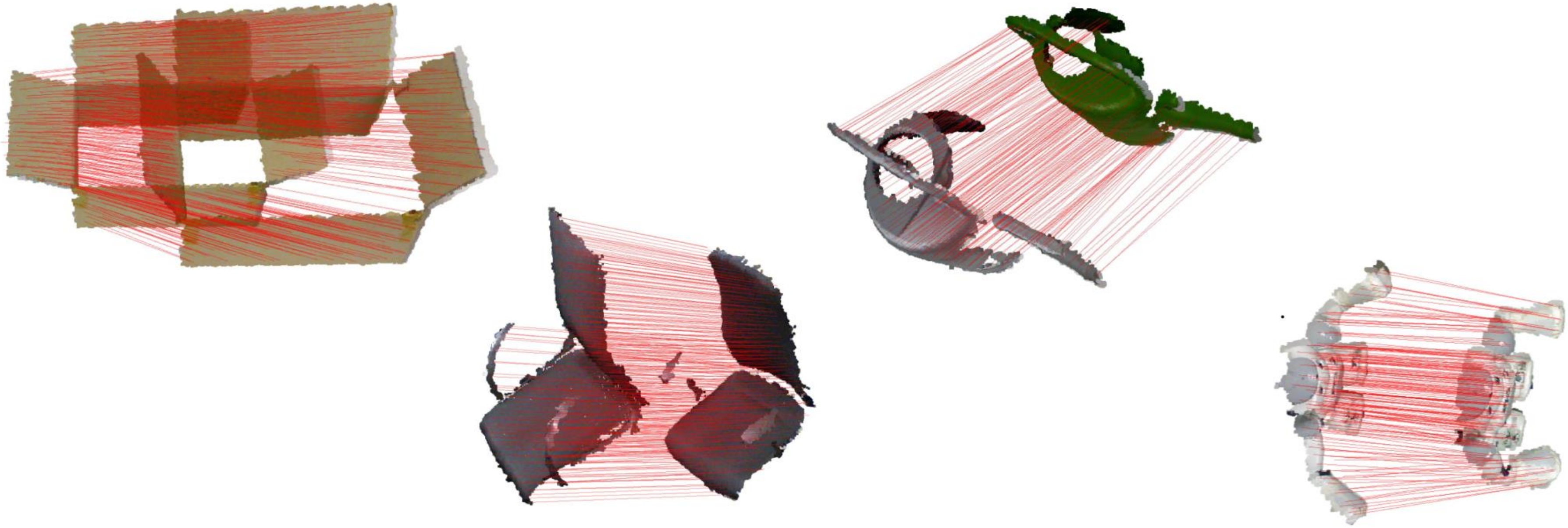



- Real-time registration with current RGB-D frame

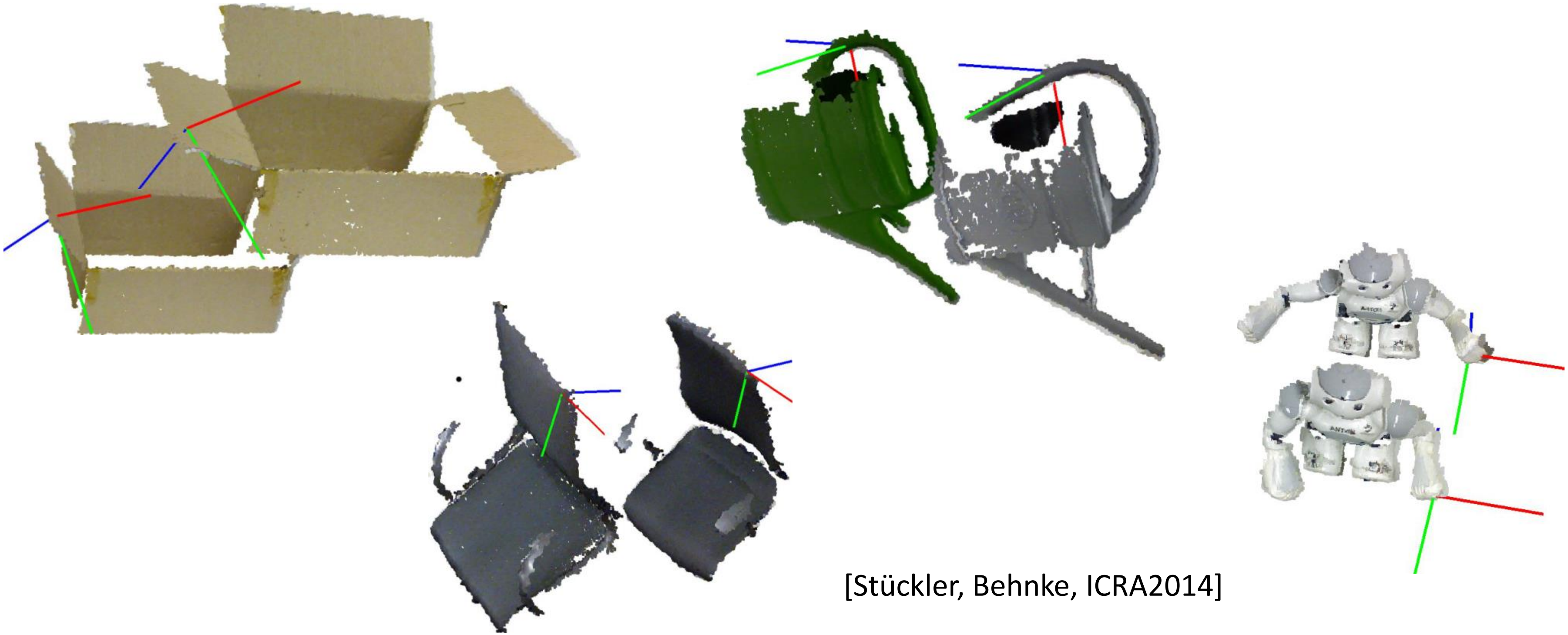
Deformable RGB-D-Registration

- Based on Coherent Point Drift method [Myronenko & Song, PAMI 2010]
- Multiresolution Surfel Map allows real-time registration



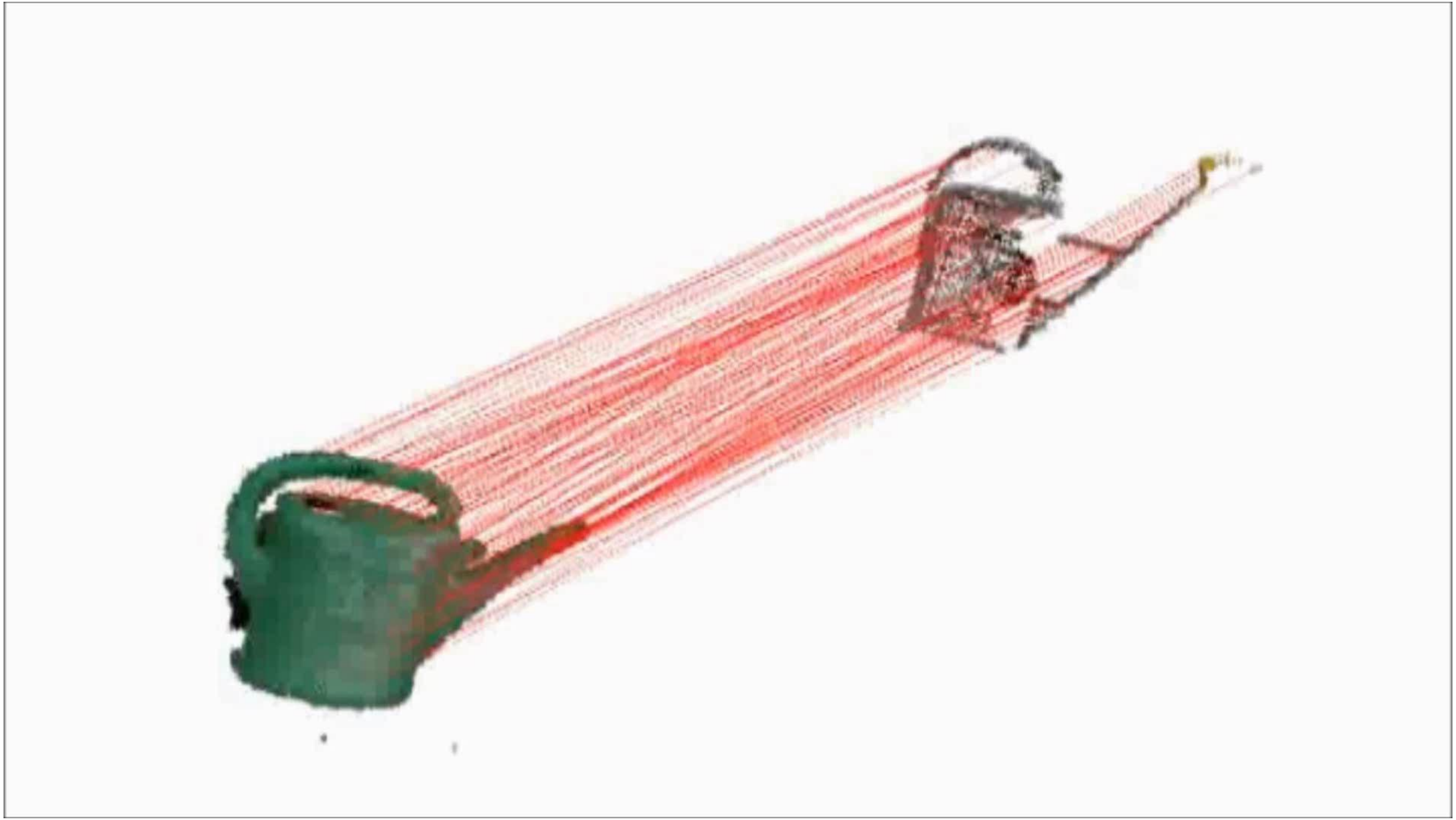
Transformation of Poses on Object

- Derived from the deformation field



[Stückler, Behnke, ICRA2014]

Demonstration of Complex Manipulation Tasks



The Data Problem

- Learning in robotics suffers from shortage of available examples
- We address this problem in two ways:

1. **Generating data**:
   Automatic data capture,
   online mesh databases,
   scene synthesis

2. **Improving generalization**:
   Object-centered models,
   deformable registration,
   transfer learning,
   semi-supervised learning

Object Capture and Scene Rendering

- Turntable + DLSR camera

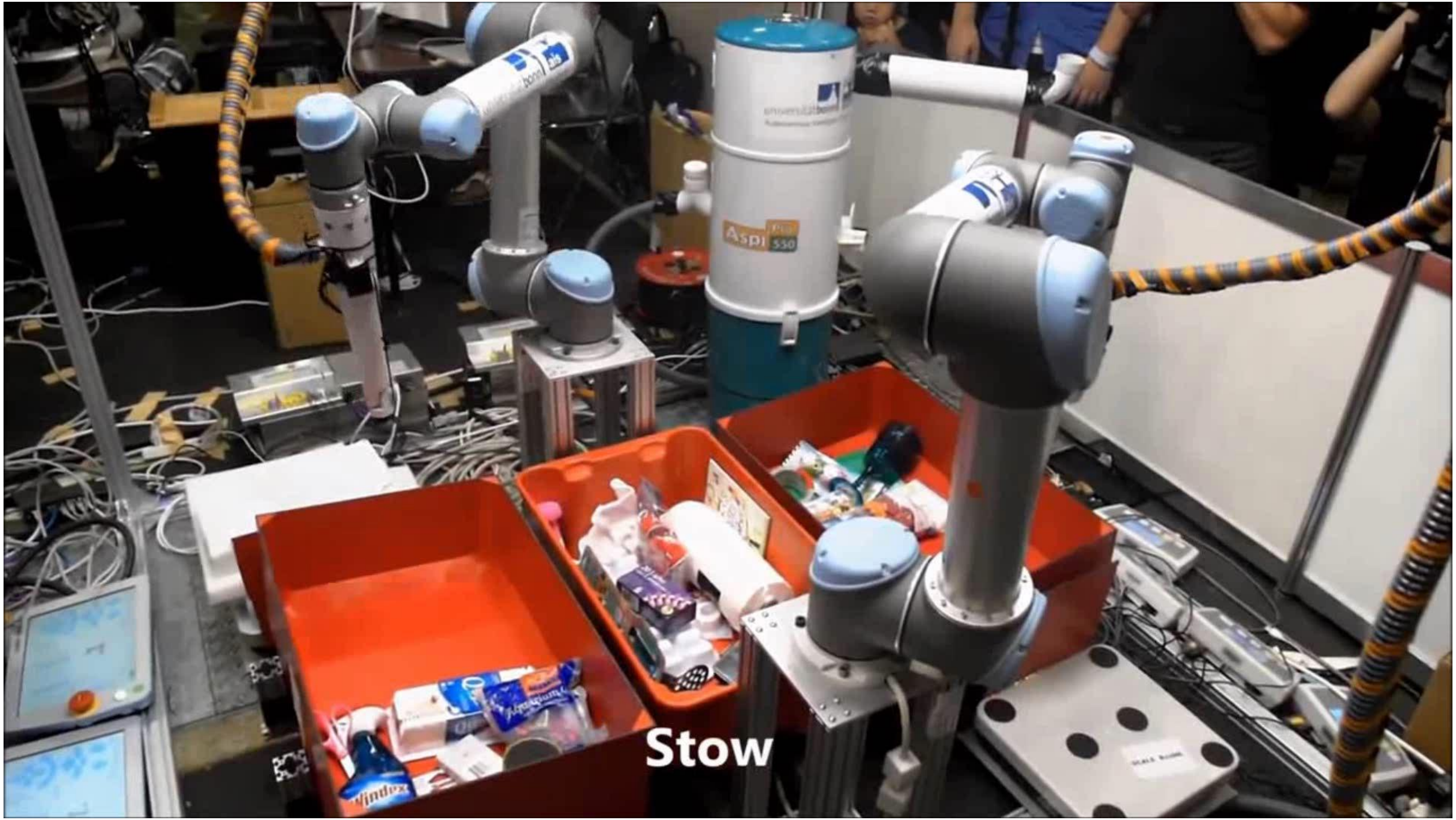
[Schwarz et al. ICRA 2018]

- Rendered scenes

Semantic Segmentation Example

[Schwarz et al. ICRA 2018]

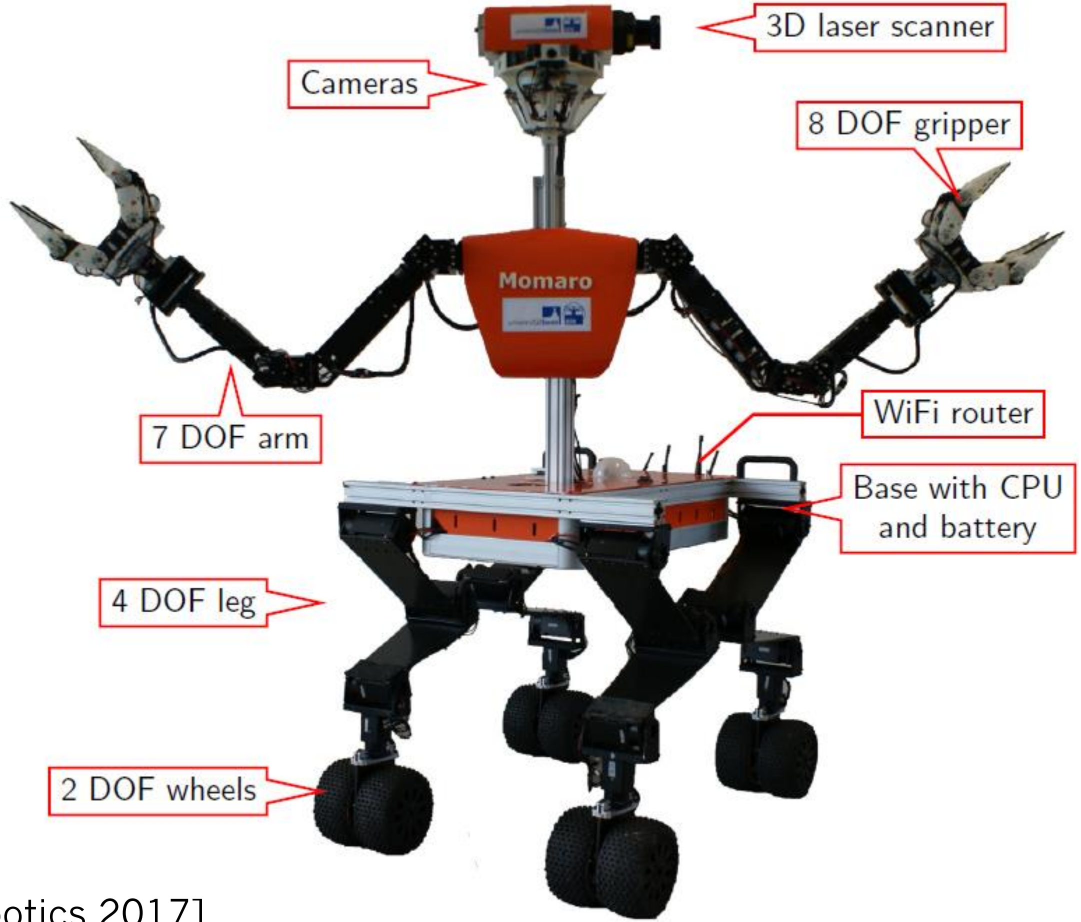
Amazon Robotics Challenge 2017



UNIVERSITÄT BONN AIS

Mobile Manipulation Robot Momaro

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



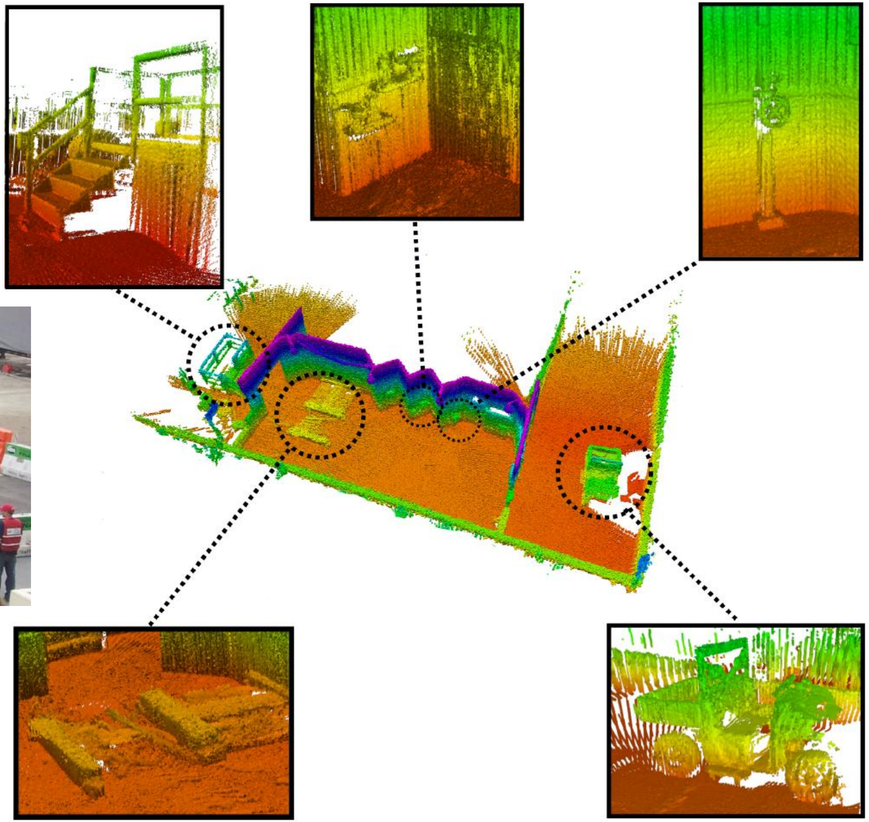
[Schwarz et al. Journal of Field Robotics 2017]

DARPA Robotics Challenge

UNIVERSITÄT BONN AIS

Allocentric 3D Mapping

- Registration of egocentric maps by graph optimization

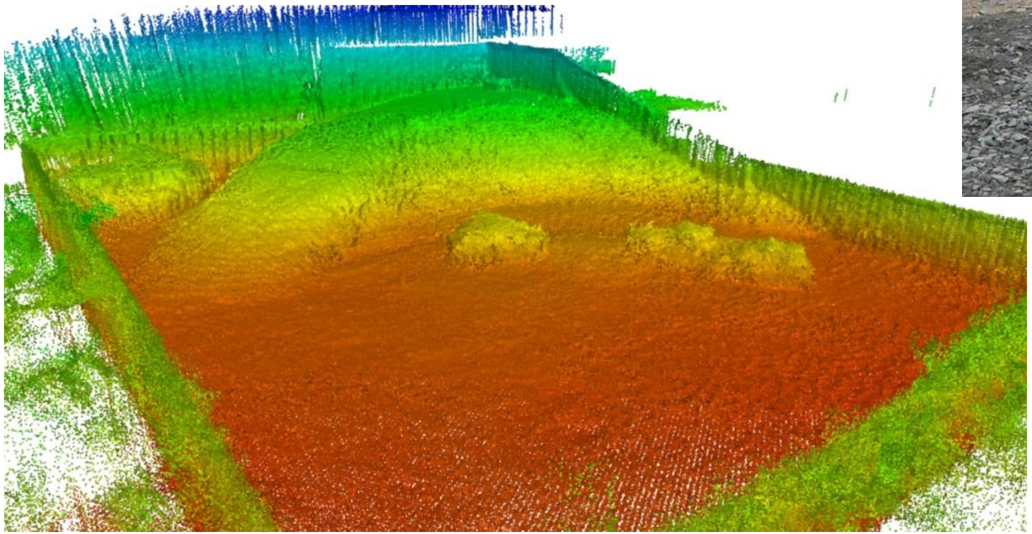


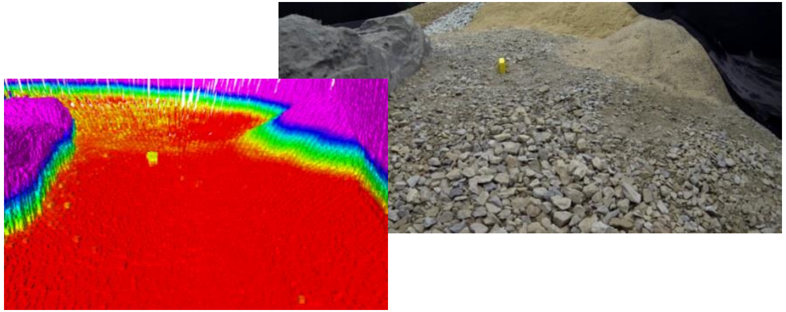
[Droeschel et al., Robotics and Autonomous Systems 2017]

DLR SpaceBot Cup 2015

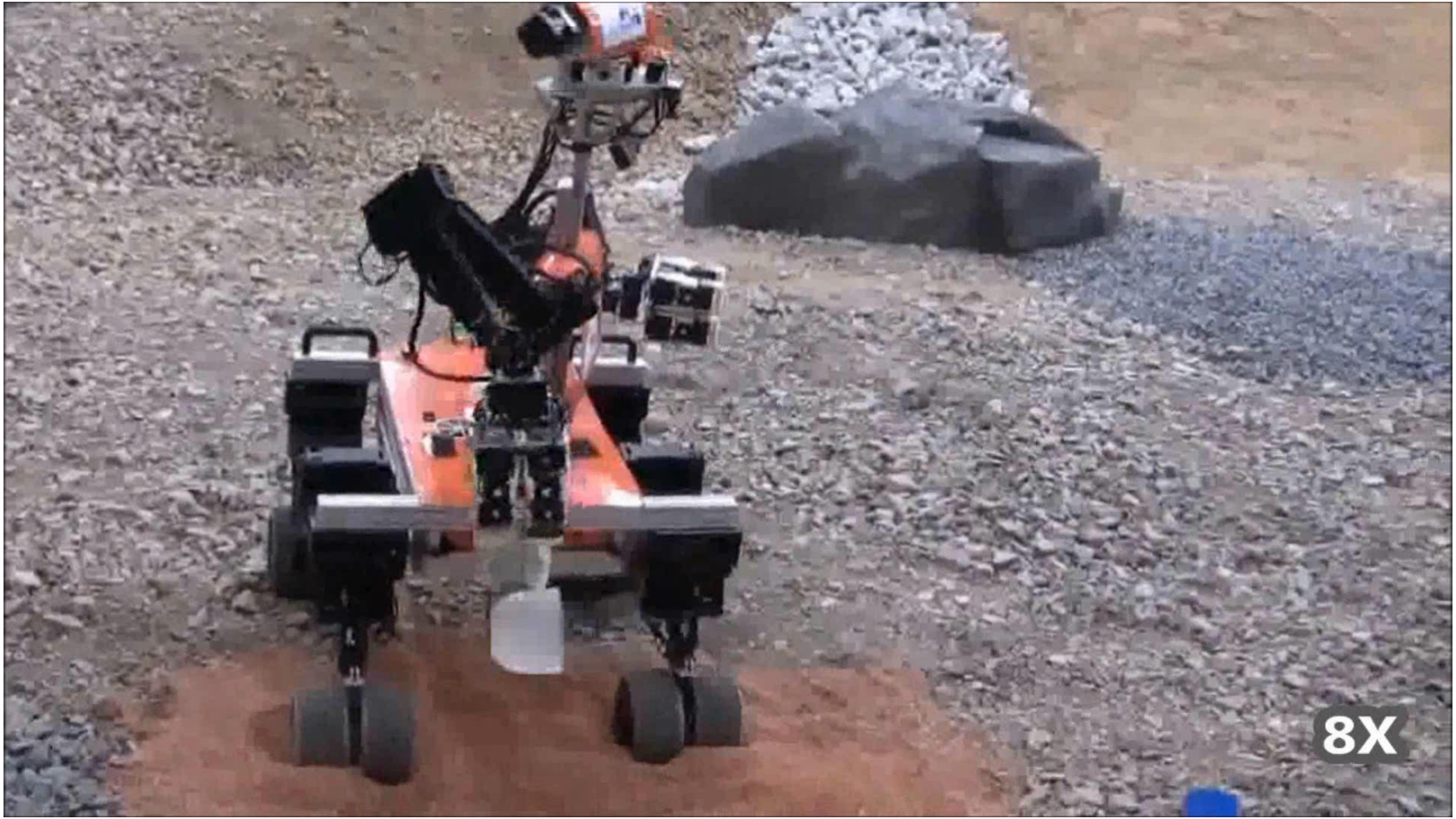
- Mobile manipulation in rough terrain

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



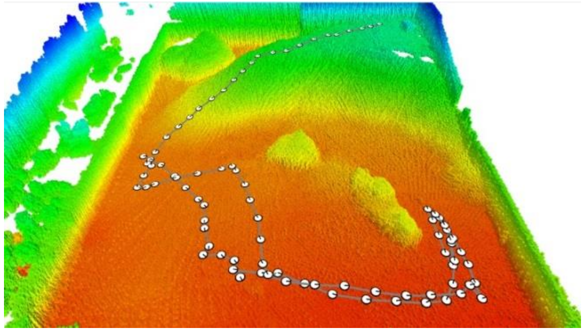


DLR SpaceBot Cup 2015

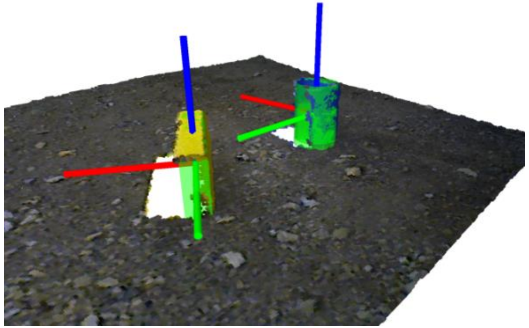


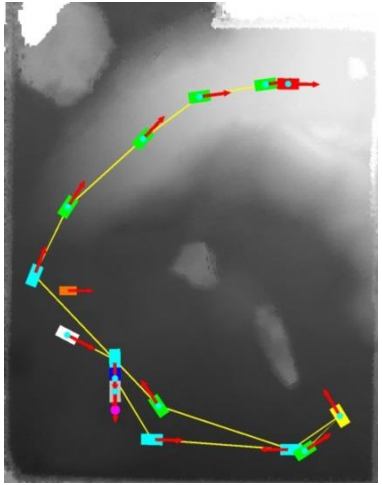
Autonomous Mission Execution

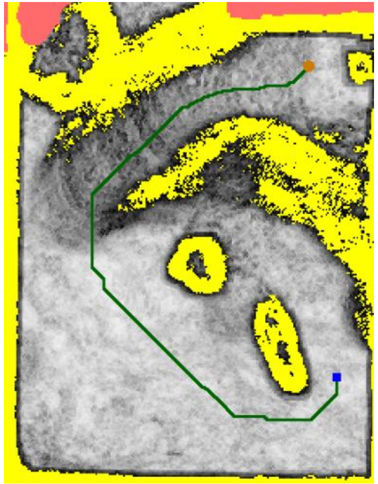
- 3D mapping, localization, mission and navigation planning

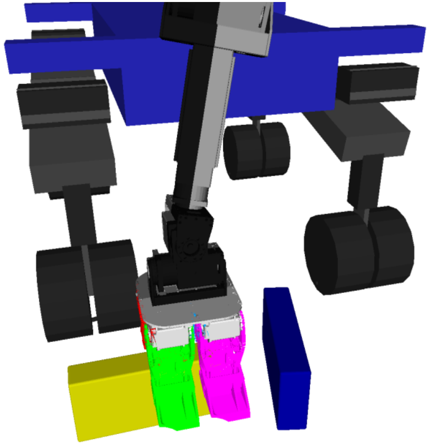


- 3D object perception and grasping









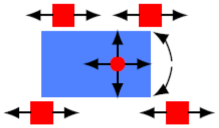
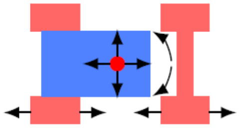
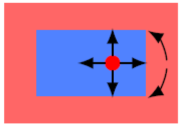
[Schwarz et al. Frontiers 2016]

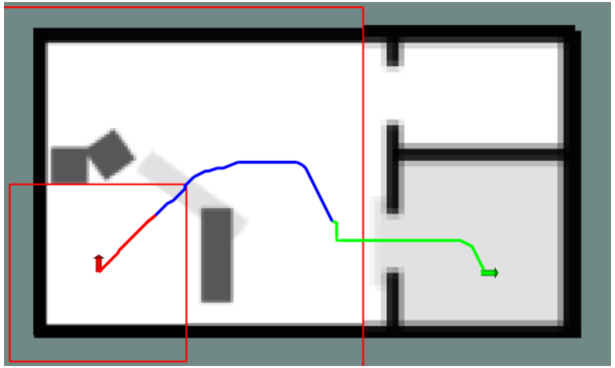
Centauro Robot

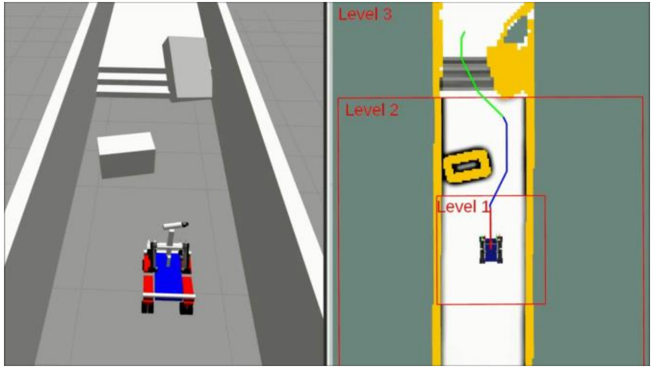
- Serial elastic actuators
- 42 main DoFs
- Schunk hand
- 3D laser
- RGB-D camera
- Color cameras
- Two GPU PCs

[Tsagarakis et al., IIT 2017]

Hybrid Driving-Stepping Locomotion Planning: Abstraction

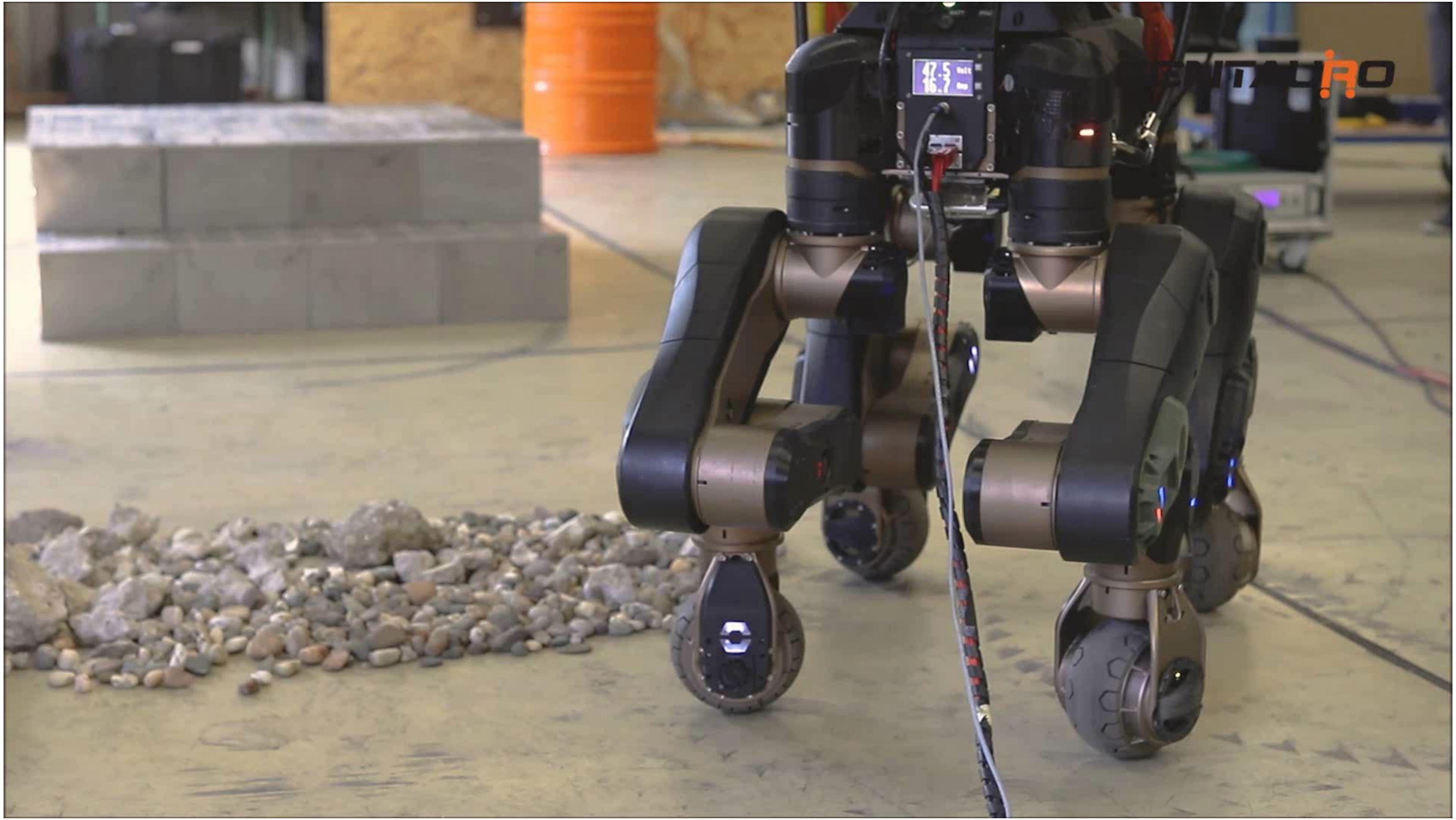
Level	Map Resolution	Map Features	Robot Representation	Action Semantics
1	• 2.5 cm • 64 orient.	• Height		• Individual Foot Actions
2	• 5.0 cm • 32 orient.	• Height • Height Difference		• Foot Pair Actions
3	• 10 cm • 16 orient.	• Height • Height Difference • Terrain Class		• Whole Robot Actions





[Klamt and Behnke, IROS 2017, ICRA 2018]

Evaluation @ KHG: Locomotion Tasks

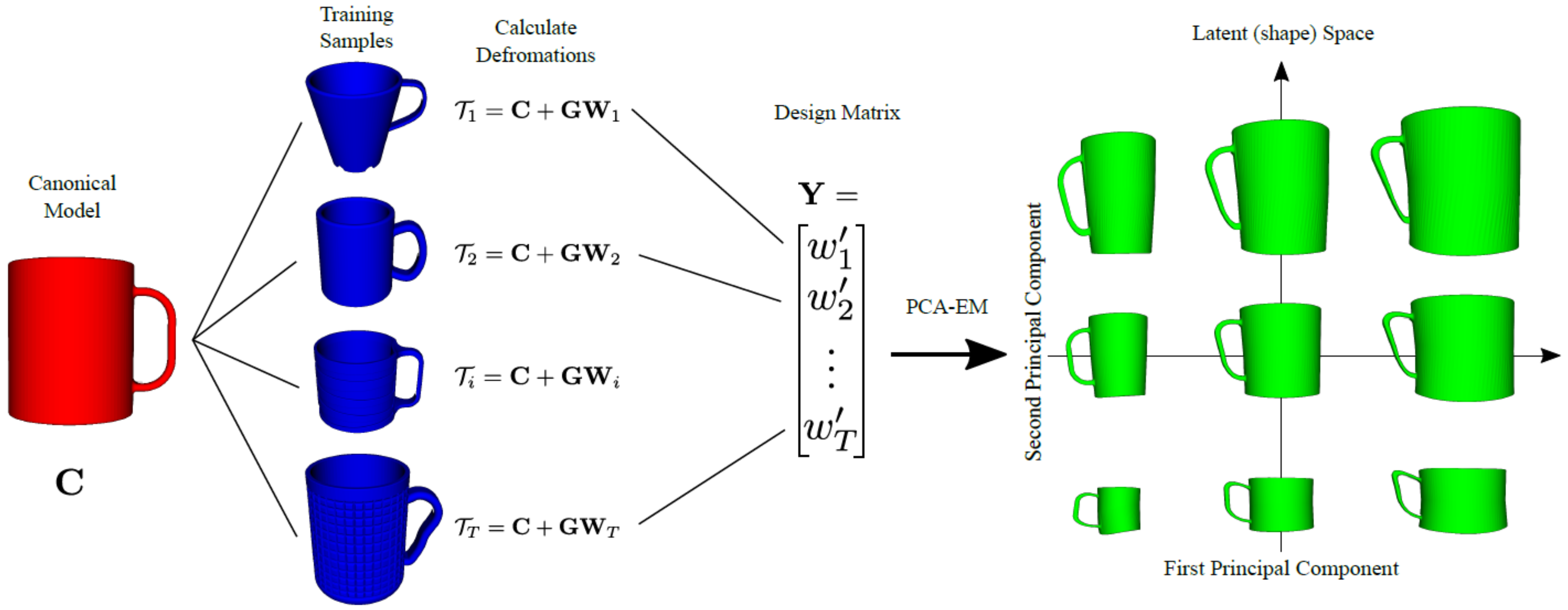


Transfer of Manipulation Skills

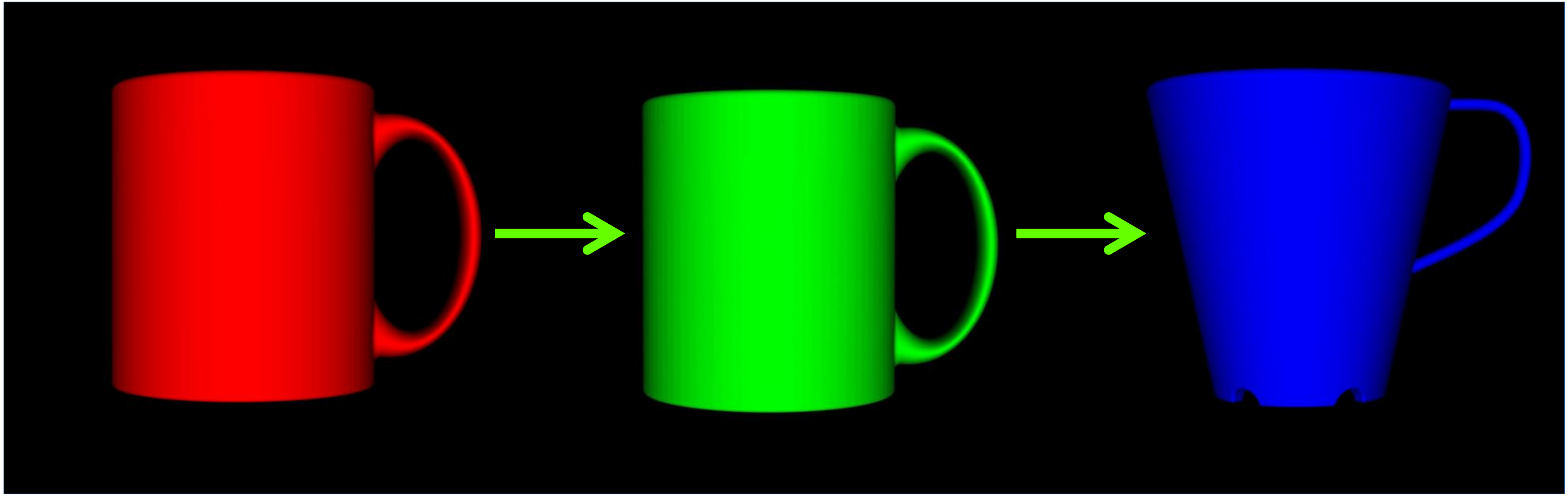
Knowledge Transfer

Learning a Latent Shape Space

- Non-rigid registration of instances and canonical model
- Principal component analysis of deformations

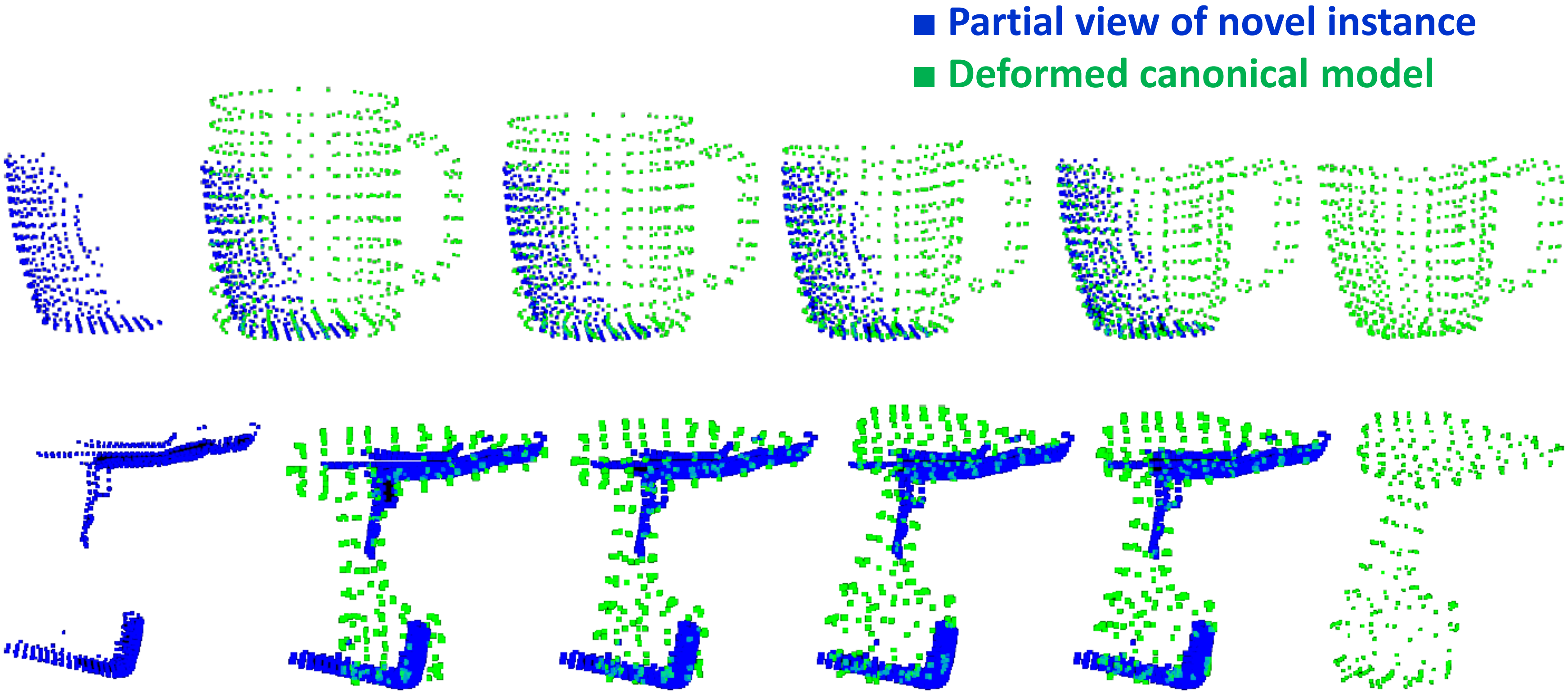


Interpolation in Shape Space



[Rodriguez and Behnke ICRA 2018]

Shape-aware Non-rigid Registration

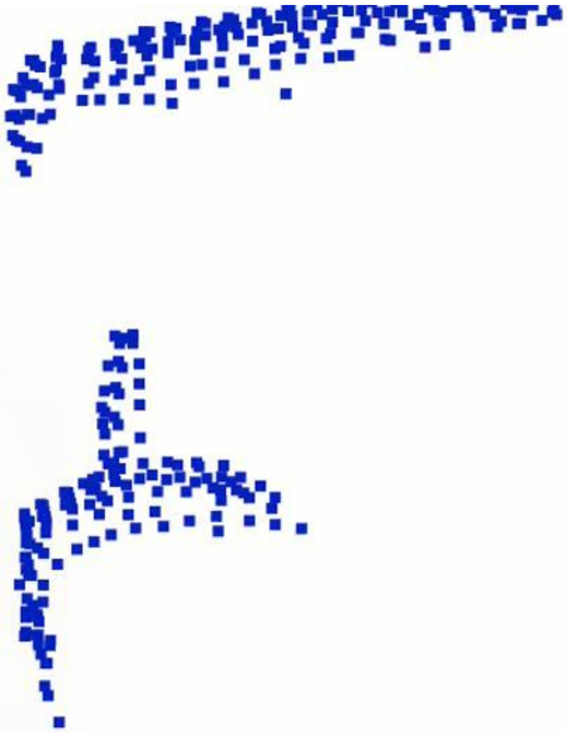


[Rodriguez and Behnke ICRA 2018]

Shape-aware Registration for Grasp Transfer

- Full point cloud

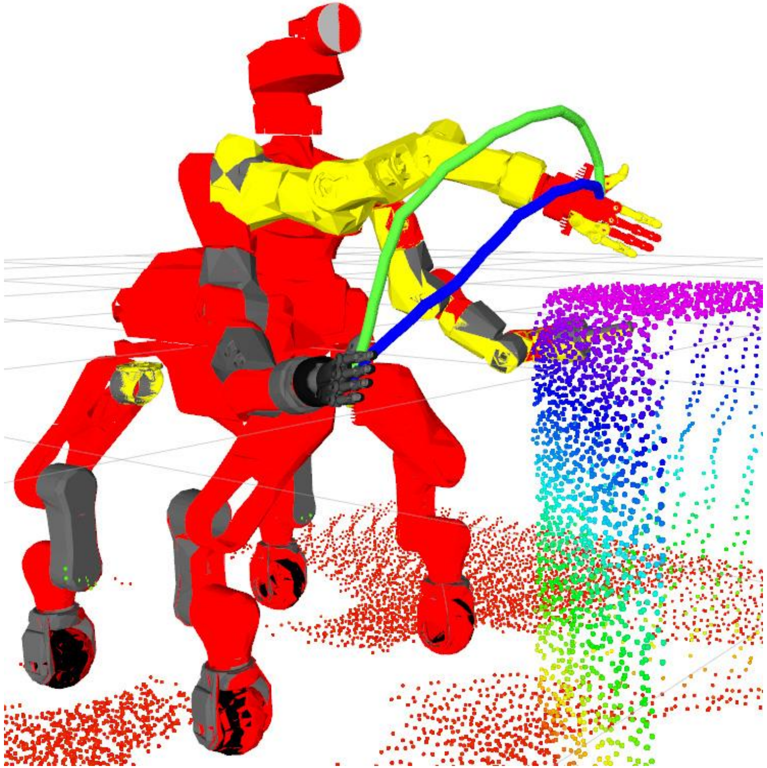
- Partial view



Collision-aware Motion Generation

Constrained Trajectory Optimization:

- Collision avoidance
- Joint limits
- Time minimization
- Torque optimization

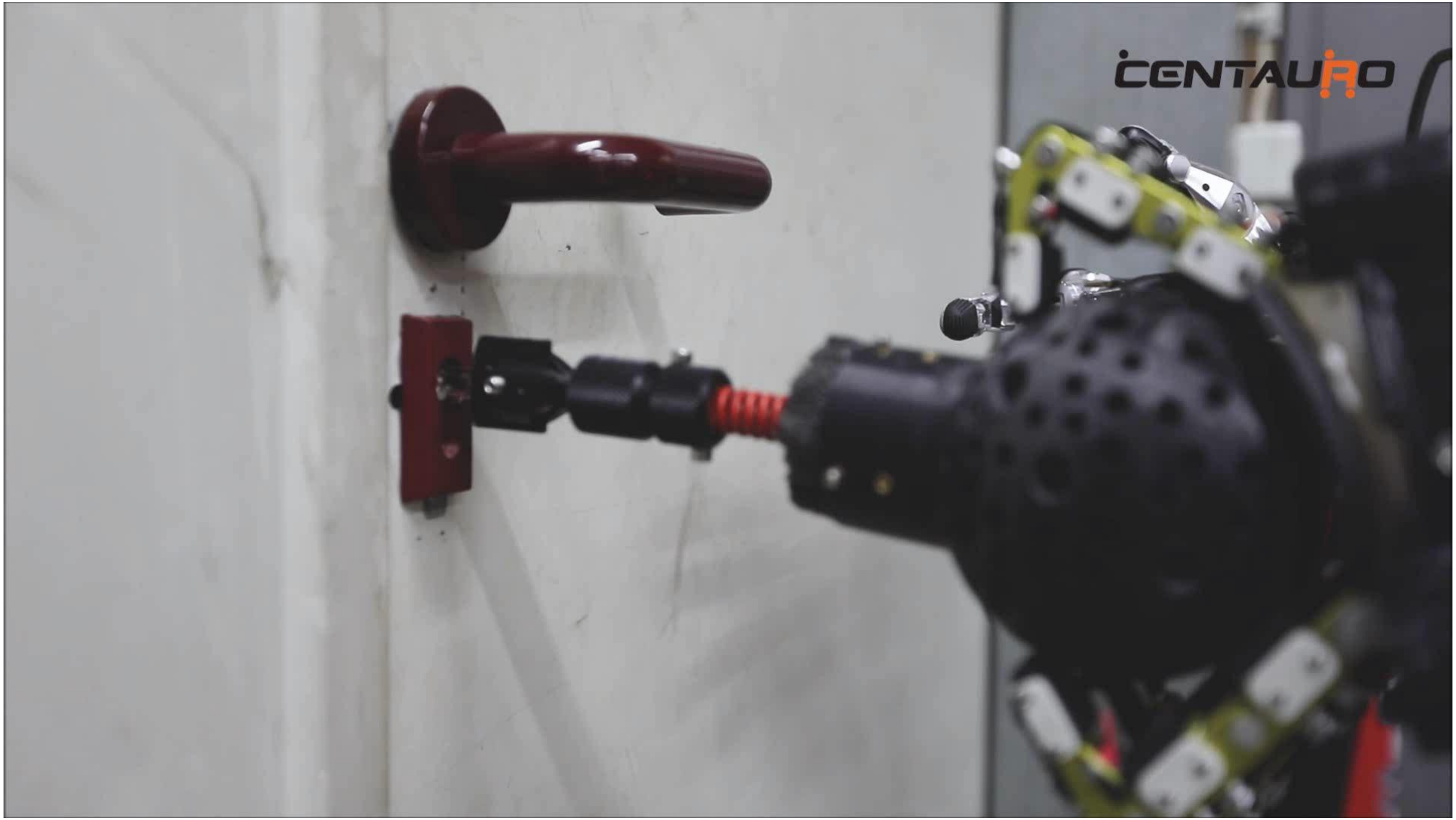


[Pavlichenko et al., IROS 2017]

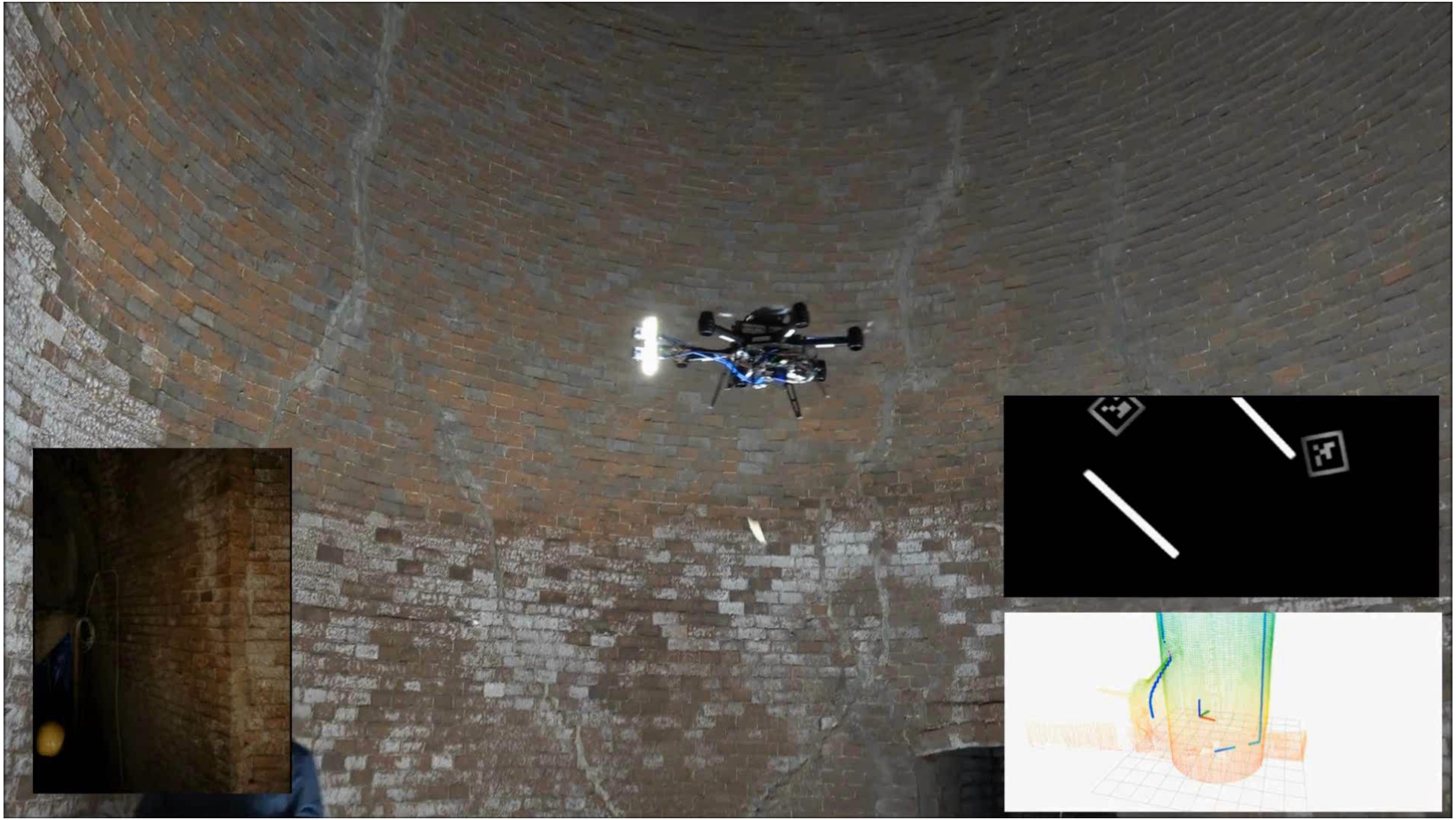
Grasping an Unknown Power Drill and Fastening Screws

UNIVERSITÄT BONN AIS

Complex Manipulation Tasks



Autonomous Micro Aerial Vehicles



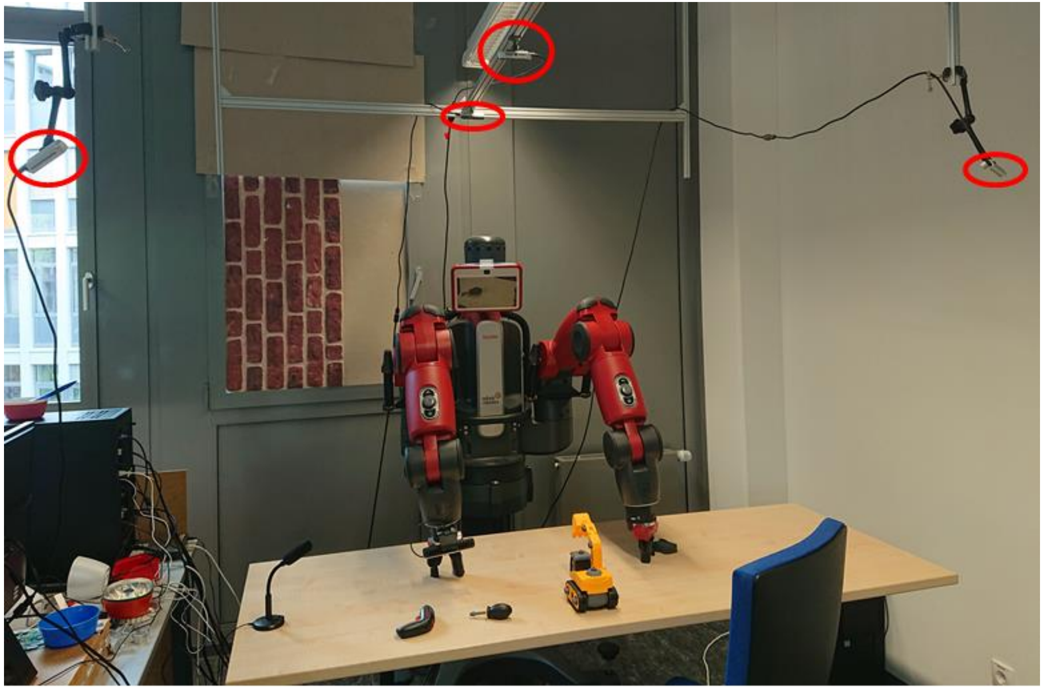
FOR2535 P7: Learning Hierarchical Representations for Anticipative Human-Robot Collaboration

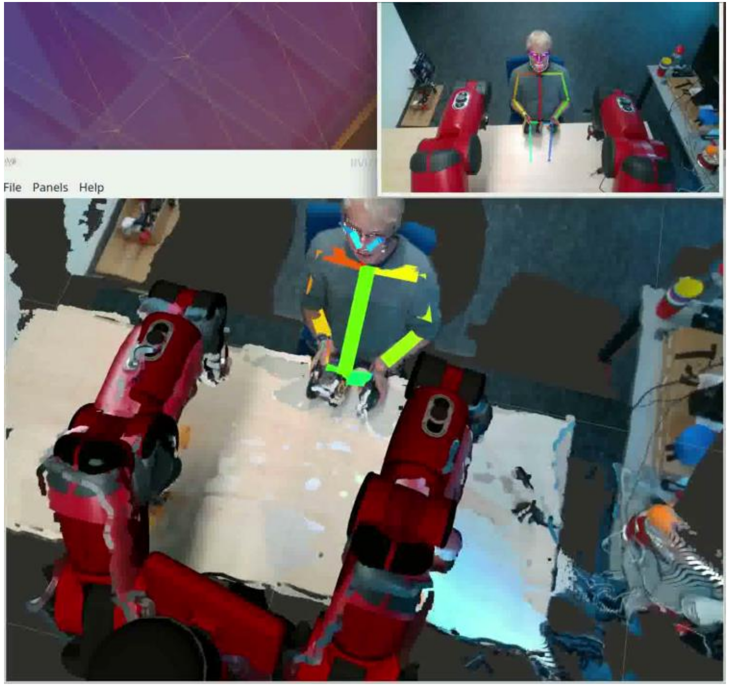
Objectives:

- Overall: Learn a sequence of representations of the shared human-robot workspace which are increasingly abstract and which allow for predictions for increasing time horizons such that future semantically meaningful states can be predicted and anticipatory robot behavior in human-robot collaboration can be realized
- O1: Unsupervised learning of hierarchical representations for prediction
- O2: Supervised prediction of semantic percepts
- O3: Anticipative human-robot collaboration

Collaborative Manipulation Setup

- Baxter robot + RGB-D cameras
- Human pose estimation (OpenPose), object detection and pose estimation
- Robot shall provide parts and tools needed for human assembly/cocking tasks

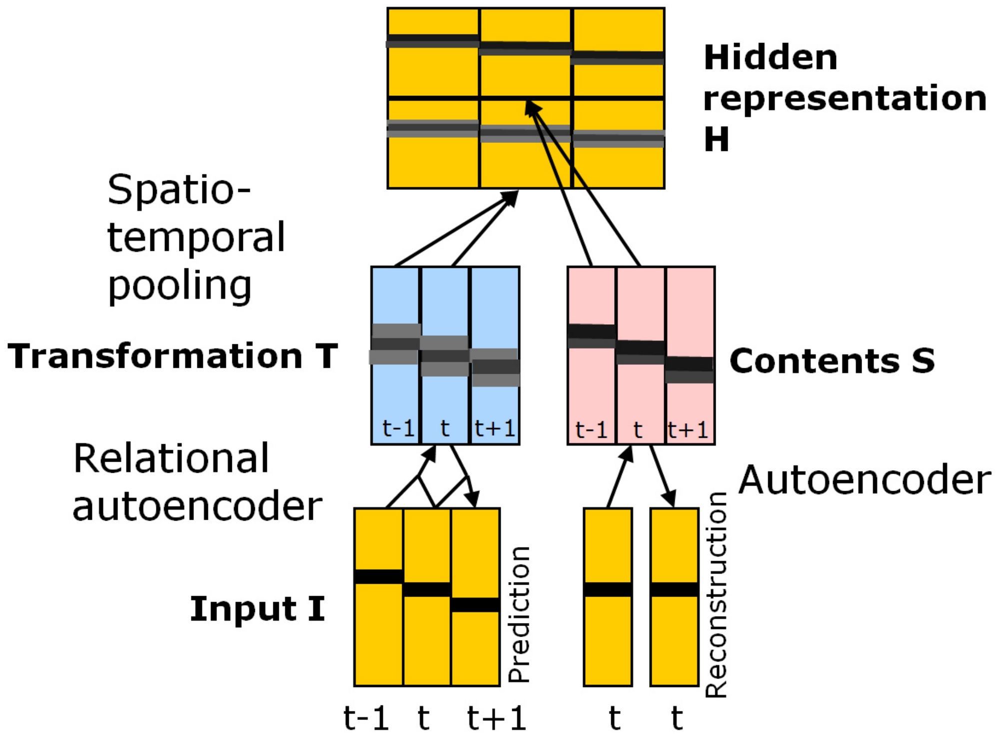




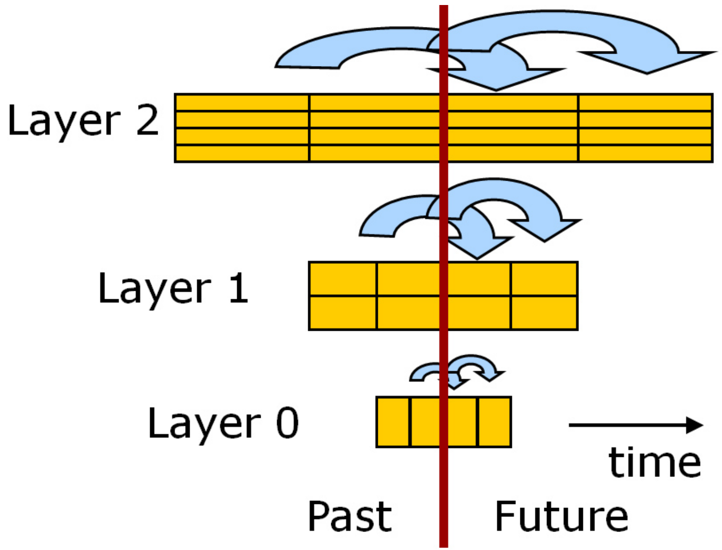
UNIVERSITÄT BONN AIS

Unsupervised Learning of Hierarchical Representations for Prediction

- Local learning module

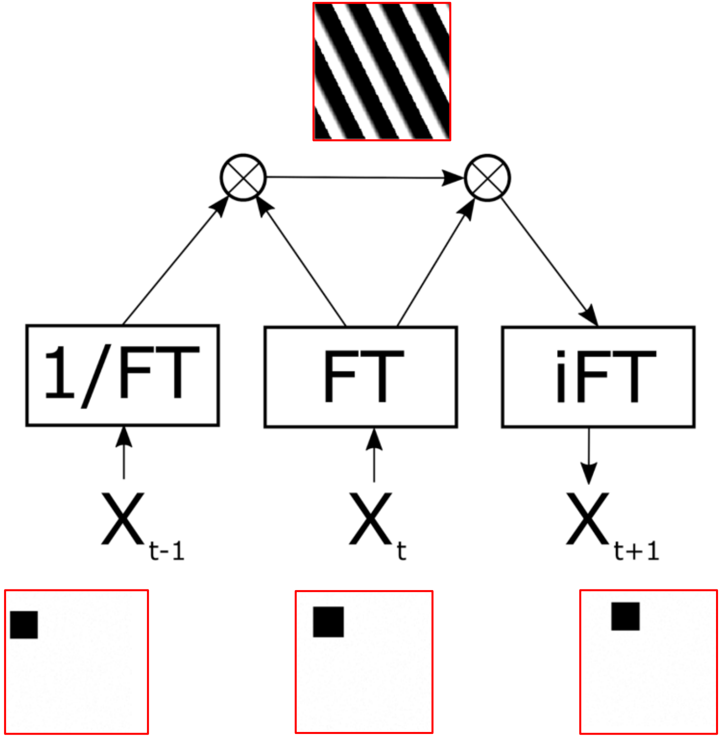


- More abstract predictions for longer time horizons



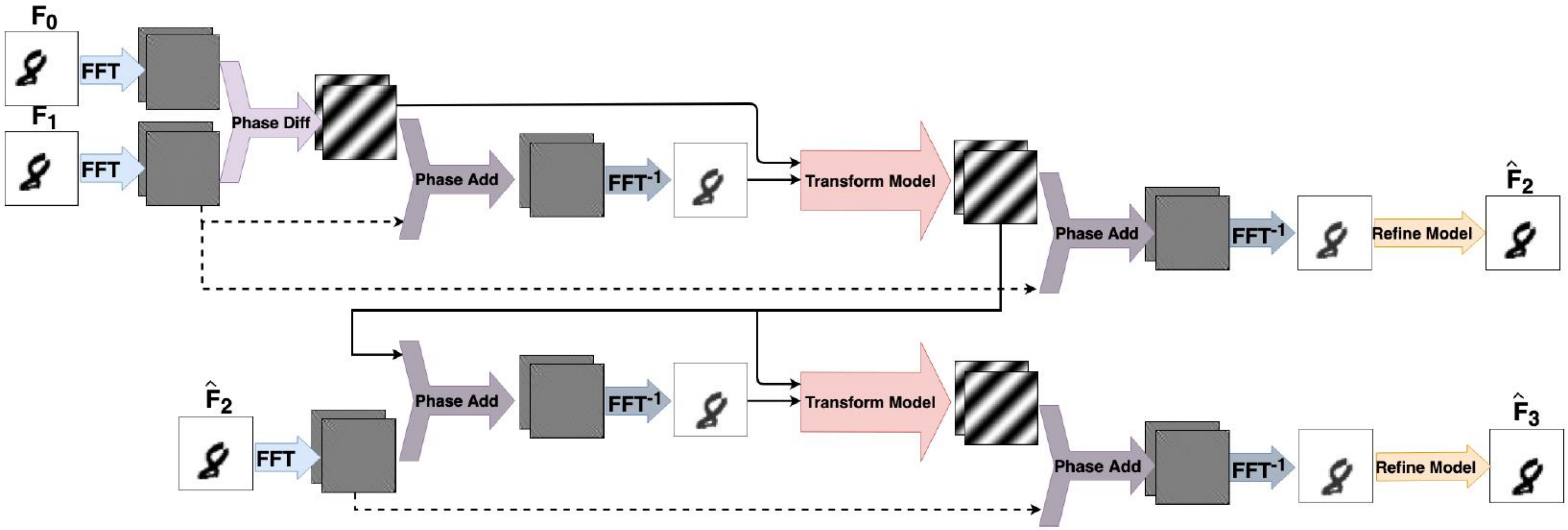
Motion Estimation in Frequency Space

- By element-wise division of Fourier coefficients
- Simple hard-wired computation graph



Frequency Domain Transformer Networks (FDTN)

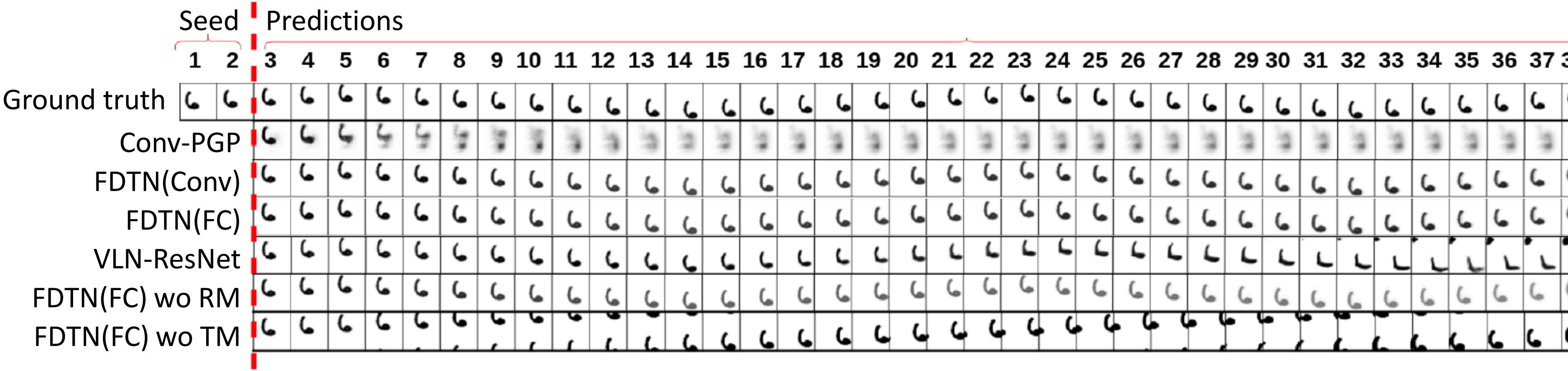
- Estimate motion by computing phase differences
- Model changes of motion, e.g. reflection at the image border (Transform model)
- Reconstruct signal (Refine model)



[Farazi & Behnke, ESANN 2019]

Frequency Domain Transformer Networks (FDTN)

- Moving MNIST: digit moving with constant random speed, bouncing at border

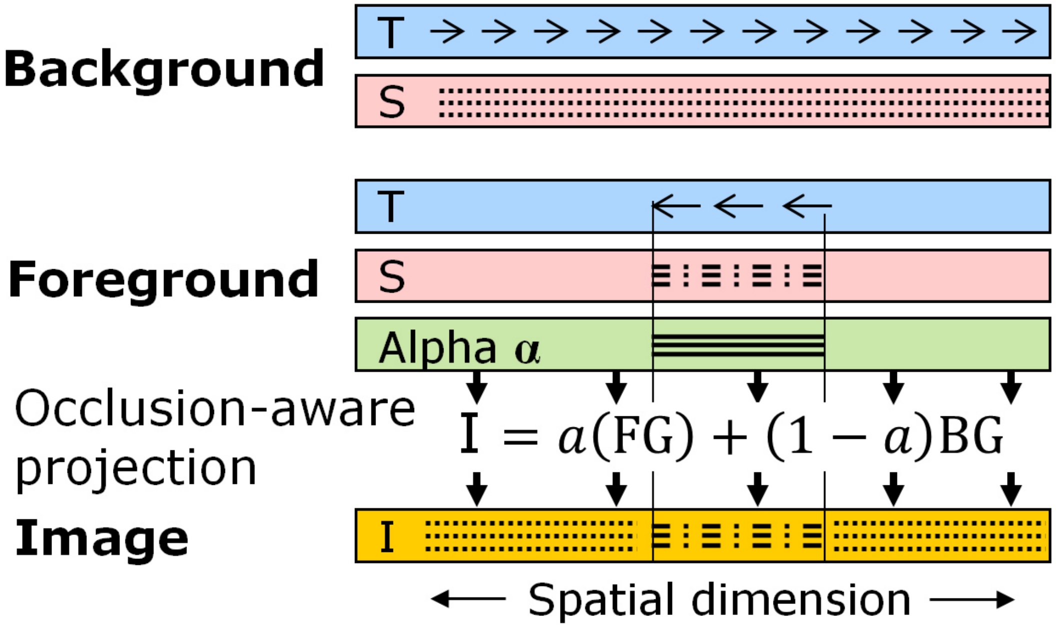


Model	MSE	Parameters
Conv-PGP	0.06963	32K
FDTN(Conv)	0.00316	**22K**
FDTN(FC)	**0.00285**	160K
VLN-ResNet	0.00544	1.3M

[Farazi & Behnke, ESANN 2019]

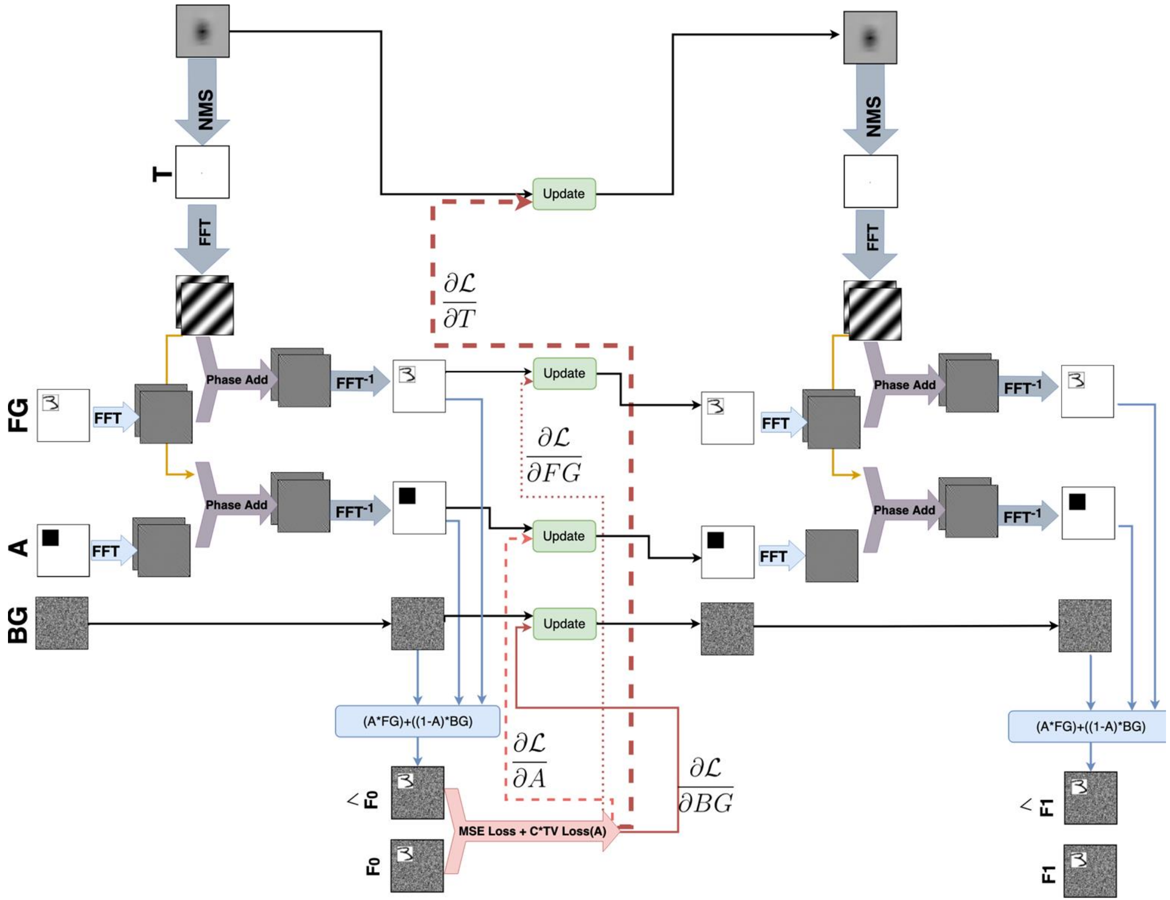
Depth-layered Models for Prediction

- Motion not uniform in image
- Use layers with separate models for motion (T) and content (S)
- Model occlusion of layers



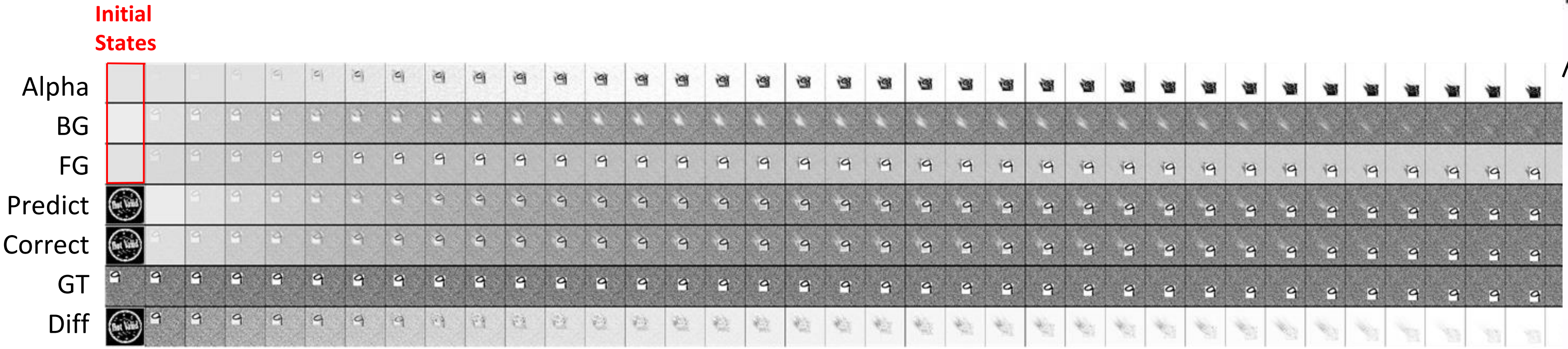
Frequency Domain Motion Segmentation

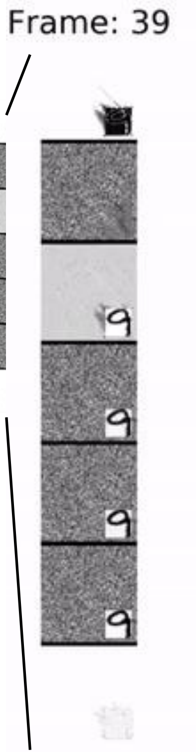
- Model stationary background and moving foreground
- Hard-wired network inspired by Kalman filter
- Prediction based on applying motion estimate
- Loss: MSE+TV(A)
- Correction by gradient descent



Frequency Domain Motion Segmentation

- Digit moving with constant random speed, stationary random background





Frequency Domain Motion Segmentation

- Digit moving with constant random speed, stationary random background