Perception and Planning for Cognitive Robots

Sven Behnke

University of Bonn
Computer Science Institute VI
Autonomous Intelligent Systems

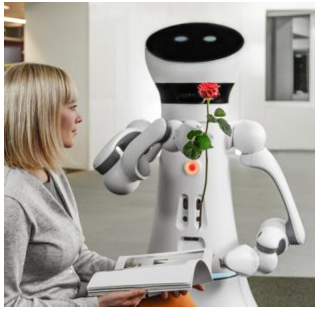
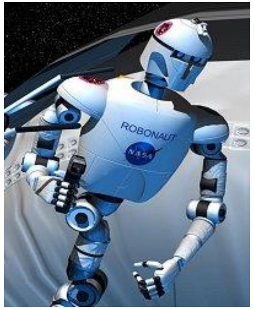
Many New Application Areas for Robots

- 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

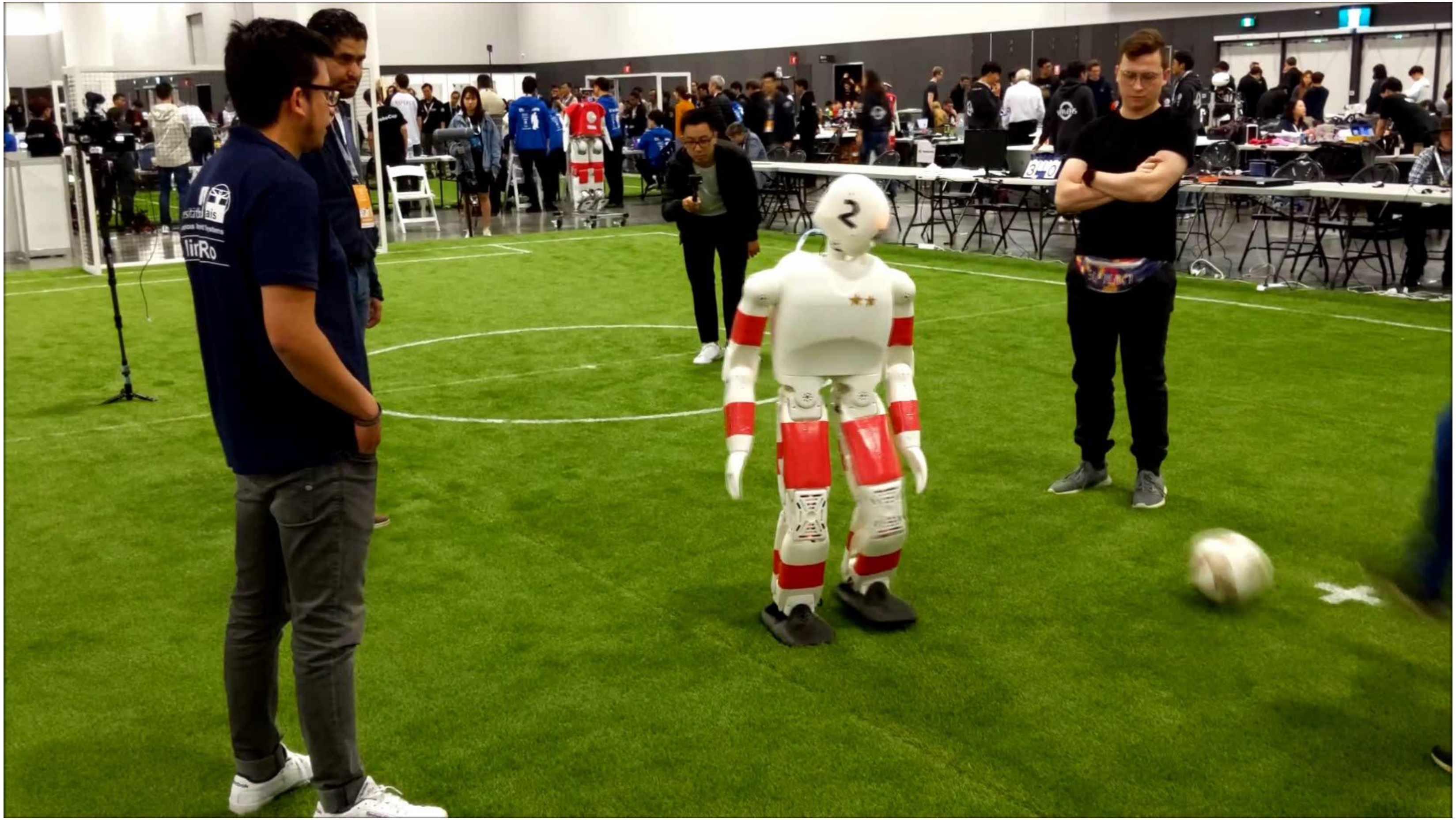
Domestic service

Mobile manipulation

Bin picking

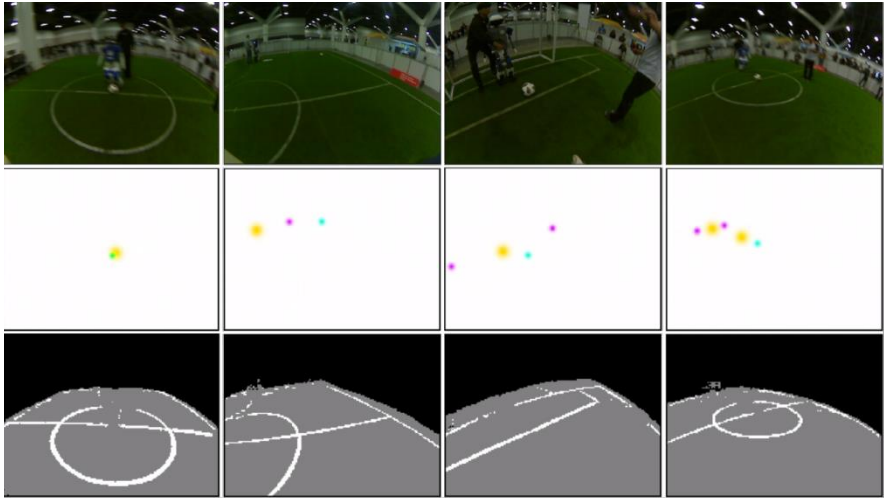
Aerial inspection

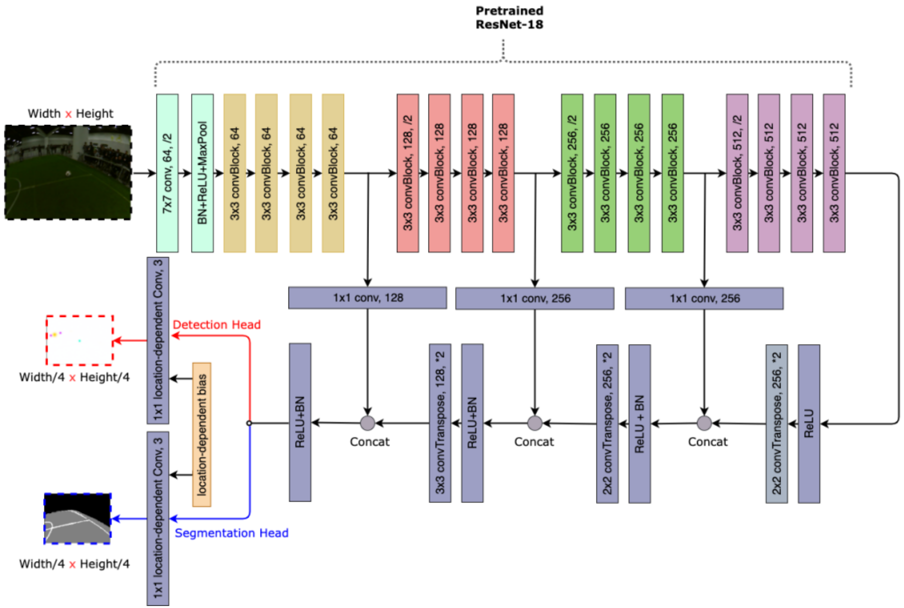
RoboCup 2019 in Sydney



Visual Perception

- Encoder-decoder network
- Two outputs
  - Object detection
  - Semantic segmentation
- Location-dependent bias



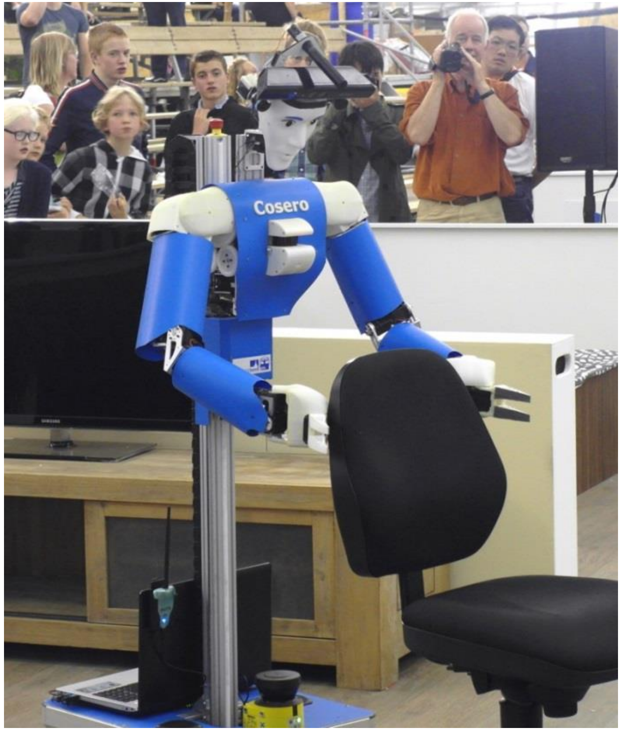


- Detects objects that are hard to recognize for humans
- Robust to lighting changes

[Rodriguez et al. , 2019]

Our Domestic Service Robots

Dynamaid



Cosero

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

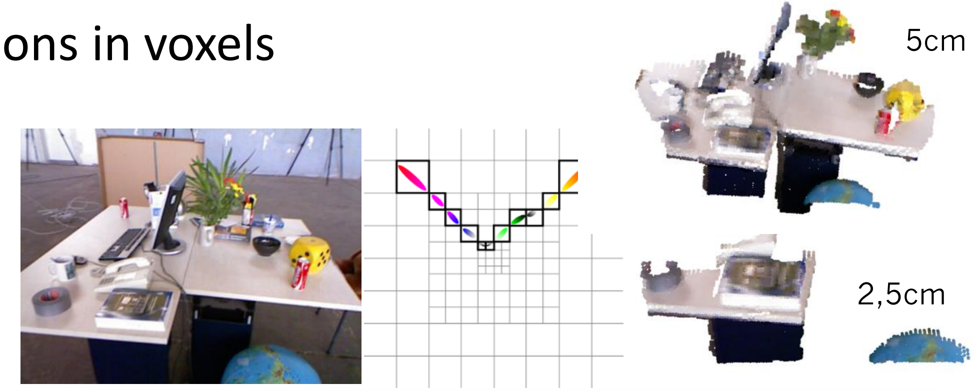
[Stückler et al.: Frontiers in Robotics and AI 2016]

Cognitive Service Robot Cosero

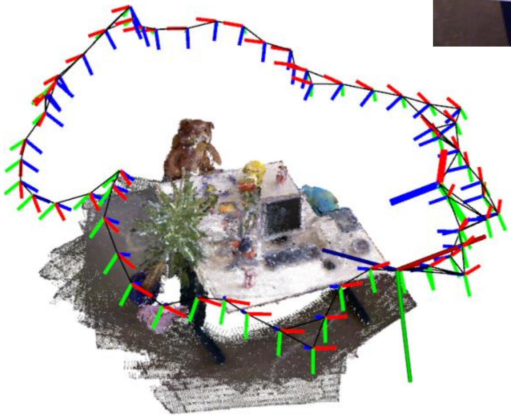
3D Mapping by RGB-D SLAM

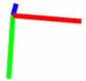
[Stückler, Behnke:
Journal of Visual Communication
and Image Representation 2013]

- Modelling of shape and color distributions in voxels

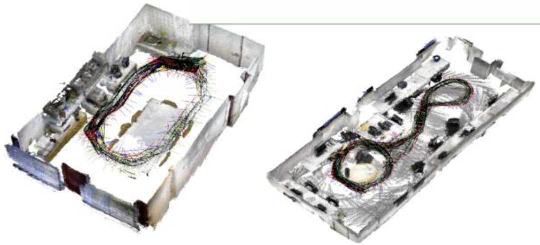


- Local multiresolution
- Efficient registration of views on CPU
- Global optimization





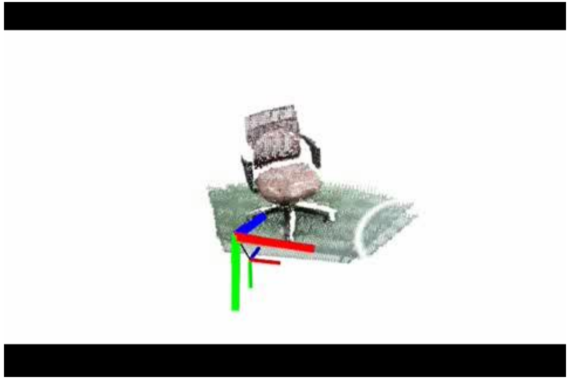
- Multi-camera SLAM



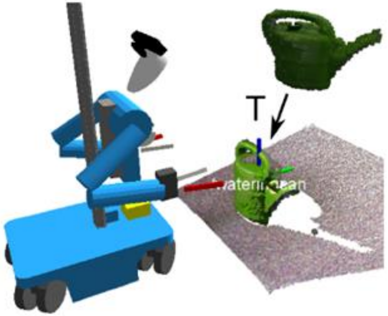
[Stoucken]

Learning and Tracking Object Models

- Modeling of objects by RGB-D-SLAM

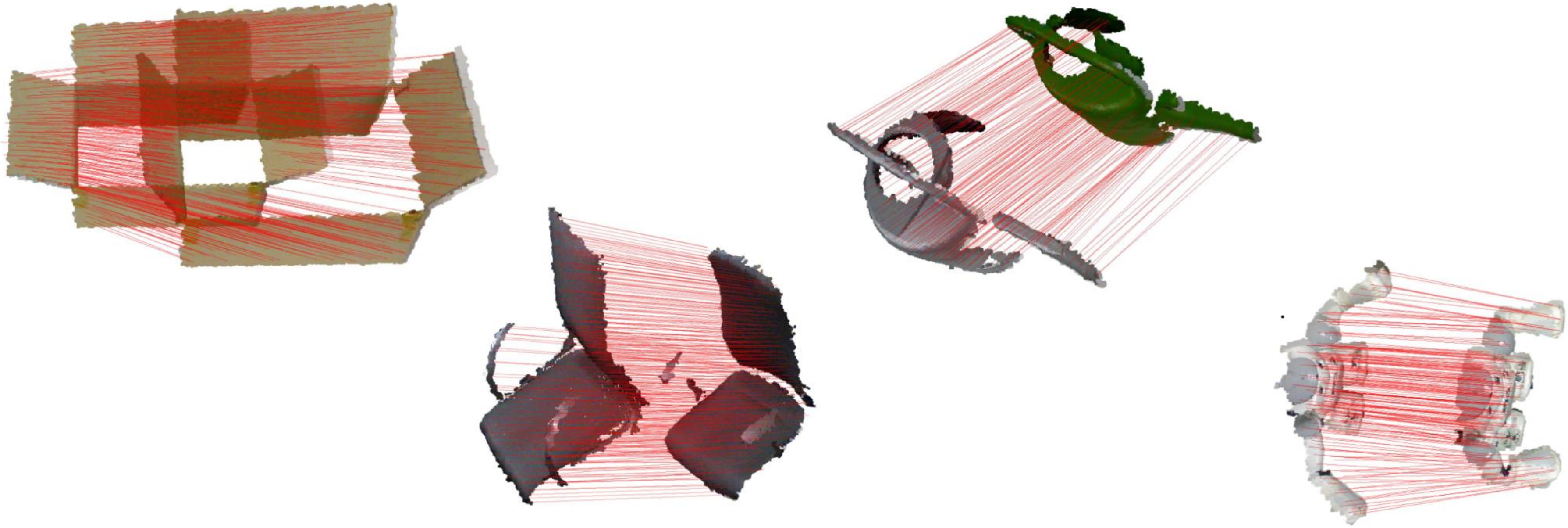


- Real-time registration with current RGB-D frame



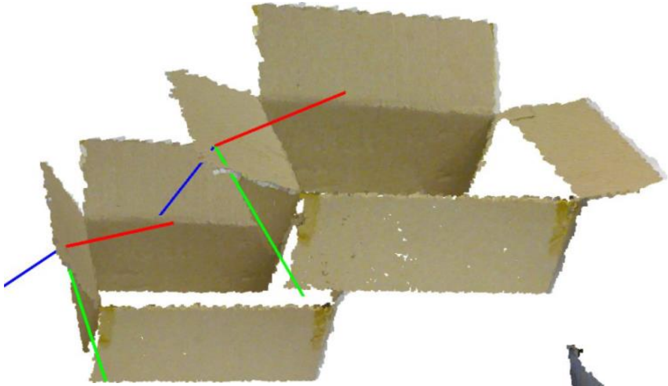
Deformable RGB-D-Registration

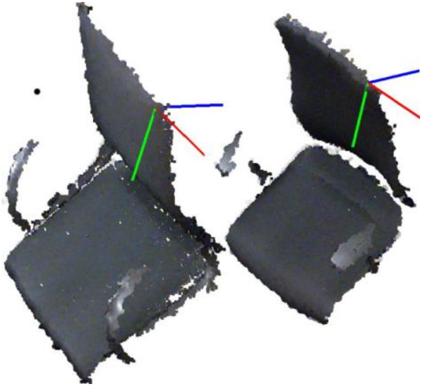
- Based on Coherent Point Drift method [Myronenko & Song, PAMI 2010]
- Multiresolution Surfel Map allows real-time registration

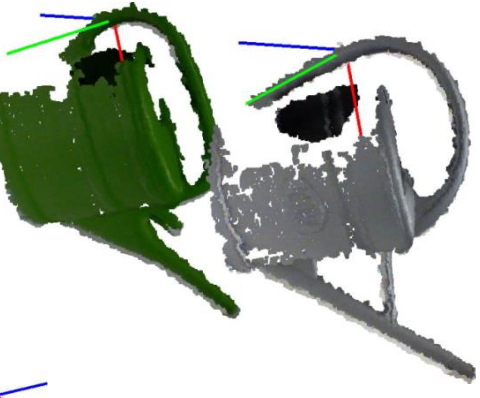


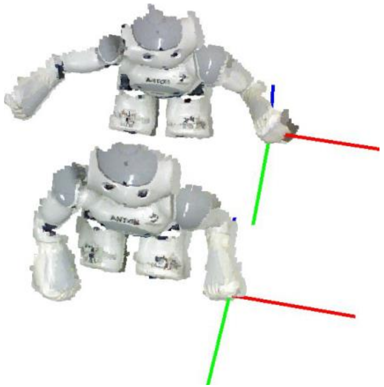
Transformation of Poses on Object

- Derived from the deformation field



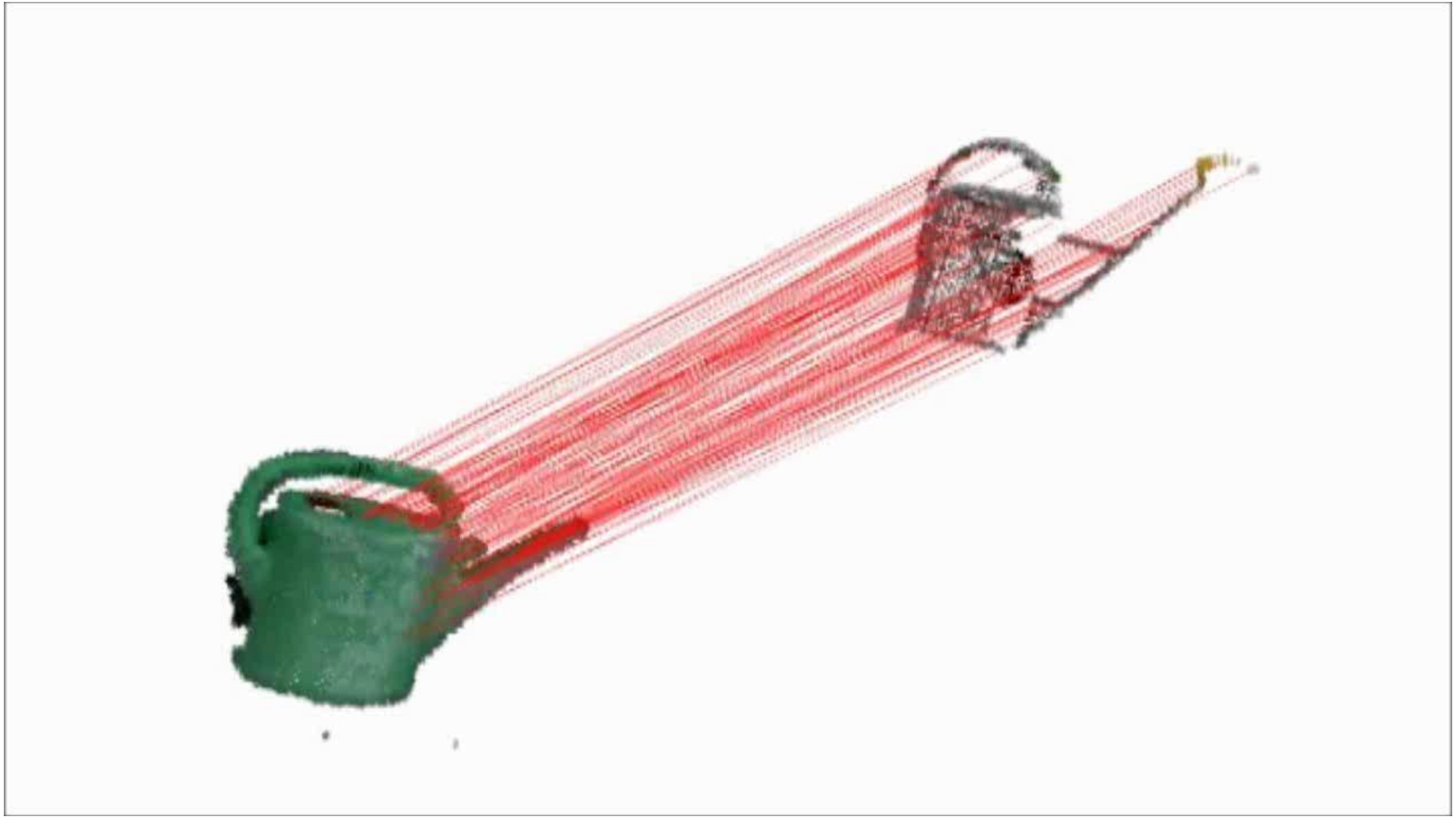






[Stückler, Behnke, ICRA2014]

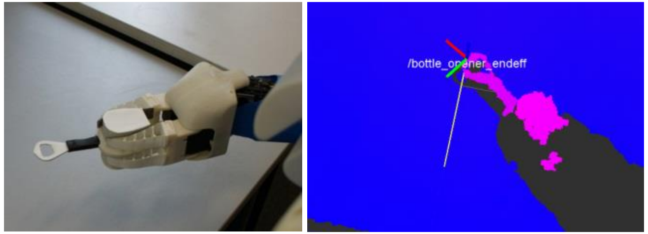
Grasp & Motion Skill Transfer



[Stückler, Behnke, ICRA2014]

Tool use: Bottle Opener

- Tool tip perception

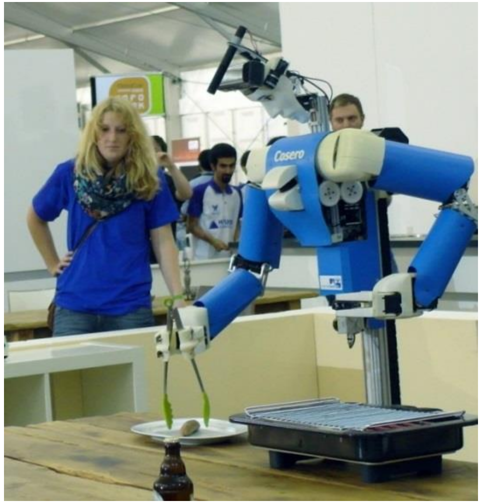


- Extension of arm kinematics
- Perception of crown cap
- Motion adaptation

[Stückler, Behnke, Humanoids 2014]

Picking Sausage, Bimanual Transport

- Perception of tool tip and sausage
- Alignment with main axis of sausage

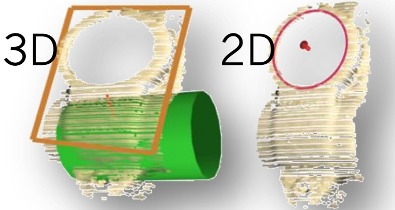


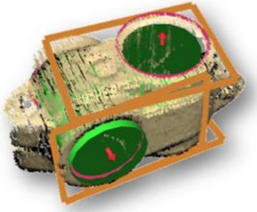
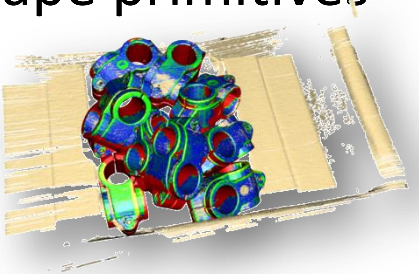
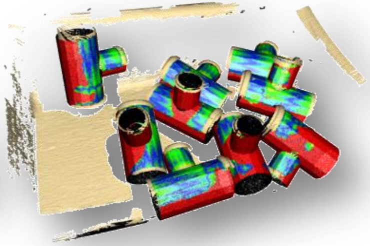
- Our team NimbRo won the RoboCup@Home League in three consecutive years

Bin Picking

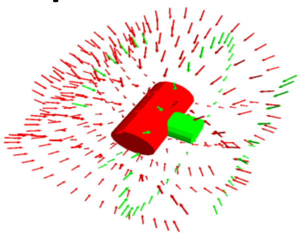
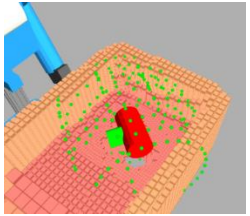
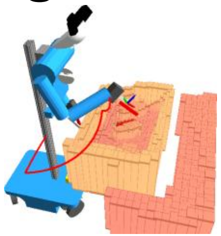
- Known objects in transport box


- Matching of graphs of 2D and 3D shape primitives



- Grasp and motion planning

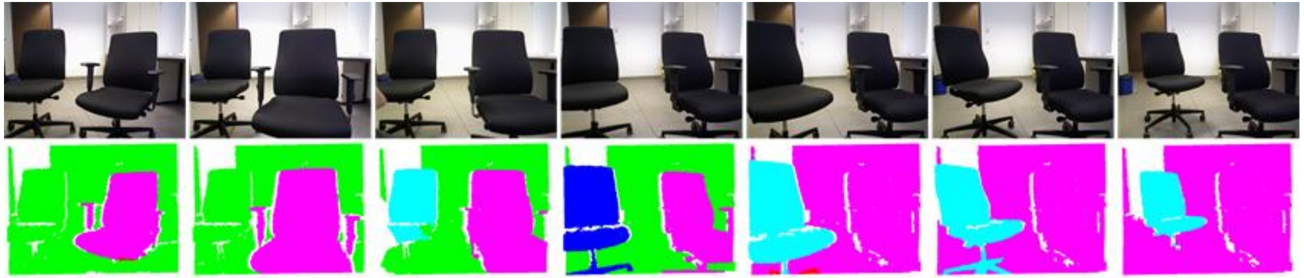




Offline Online

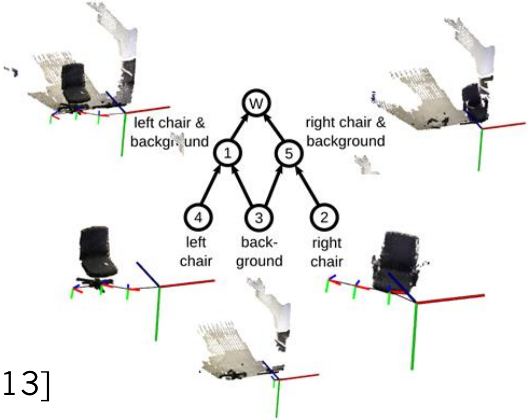
[Nieuwenhuisen et al.: ICRA 2013]

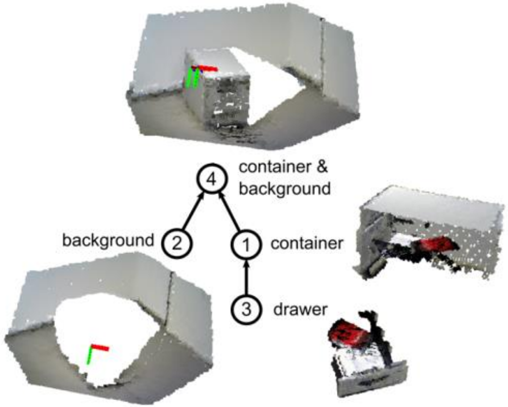
Hierarchical Object Discovery trough Motion Segmentation

- Simultaneous object modeling and motion segmentation



- Inference of a segment hierarchy

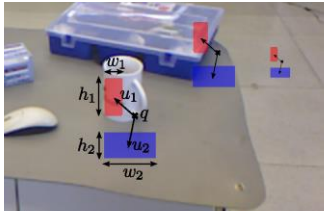


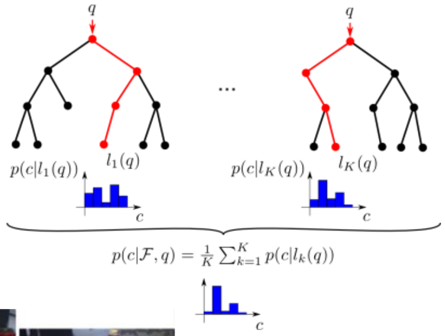


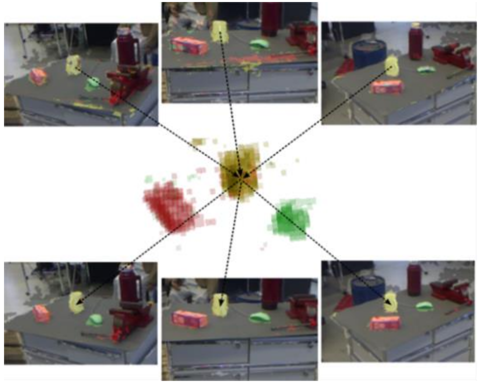
[Stückler, Behnke: IJCAI 2013]

Semantic Mapping

- Pixel-wise classification of RGB-D images by random forests
- Compare color / depth of regions
- Size normalization
- 3D fusion through RGB-D SLAM
- Evaluation on NYU depth v2







[Stückler, Biresev, Behnke: IROS 2012]

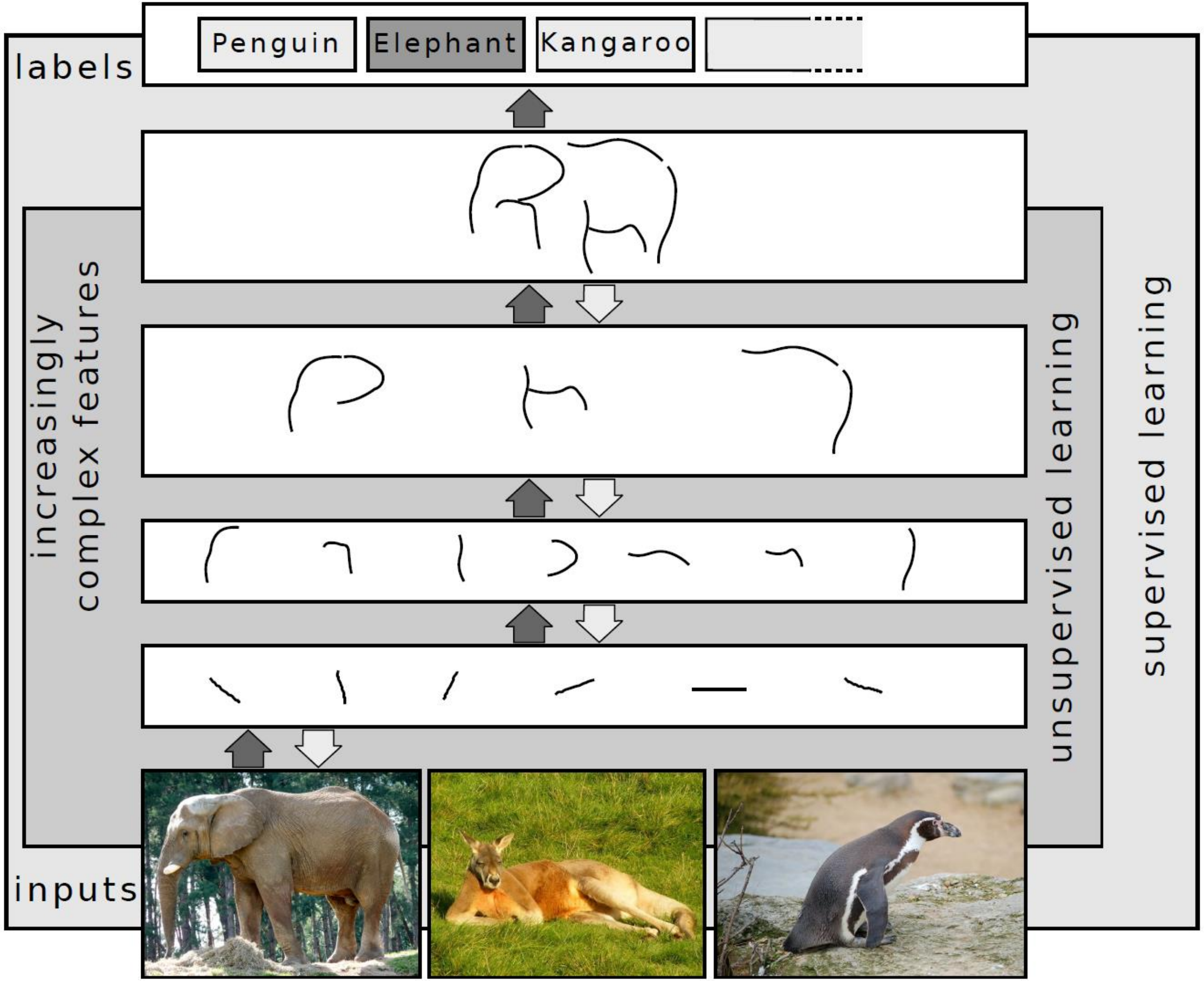
Ground truth

Segmentation

Accuracy in %	Ø Classes	Ø Pixels
Silberman et al. 2012	59,6	58,6
Couprie et al. 2013	63,5	64,5
Random forest	65,0	68,1
3D-Fusion	**66,8**	

Deep Learning

- Learning layered represen-tations

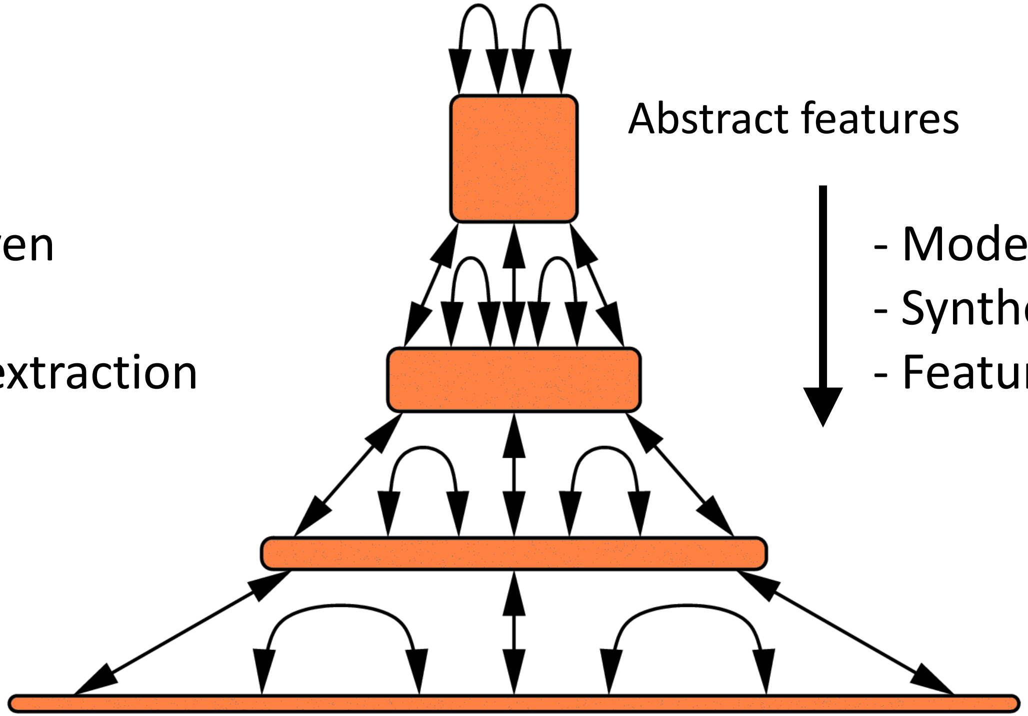


[Schulz; Behnke, KI 2012]

UNIVERSITÄT BONN AIS

Neural Abstraction Pyramid

[Behnke, Rojas, IJCNN 1998]
[Behnke, LNCS 2766, 2003]



Abstract features

- Data-driven
- Analysis
- Feature extraction

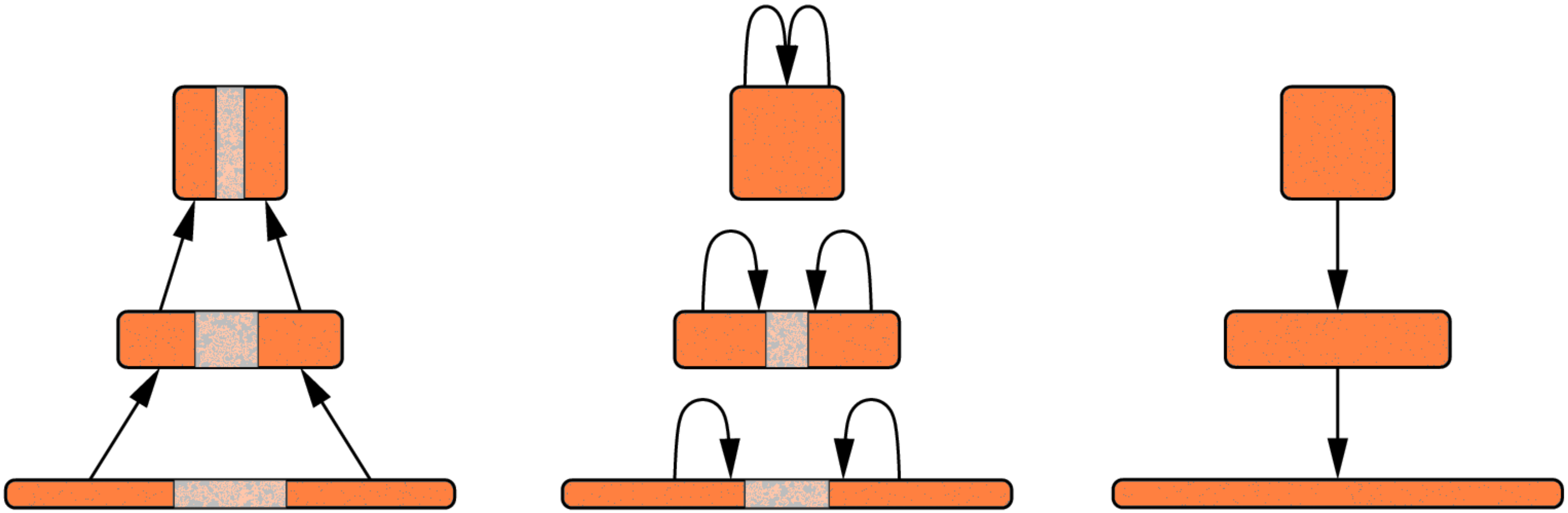
- Model-driven
- Synthesis
- Feature expansion

Signals

- Grouping - Competition - Completion

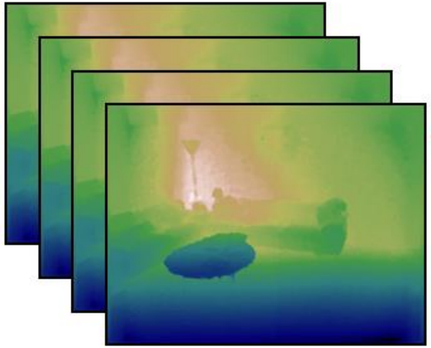
Iterative Image Interpretation

- Interpret most obvious parts first
- Use partial interpretation as context to resolve local ambiguities

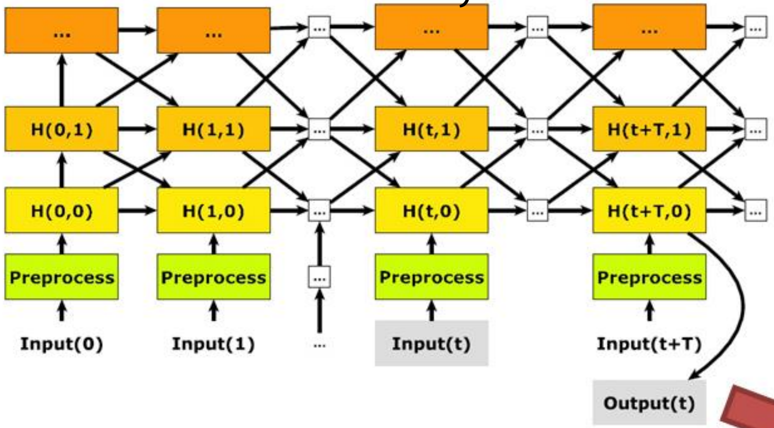


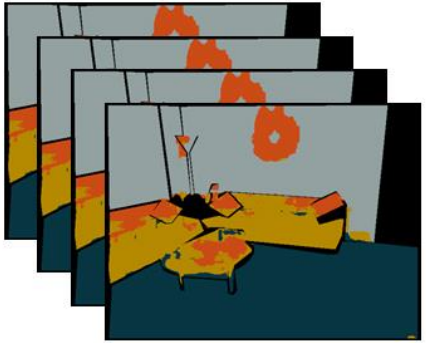
Neural Abstraction Pyramid for RGB-D Video Object-class Segmentation

- Recursive computation is efficient for temporal integration



Neural Abstraction Pyramid





[Pavel, Schulz, Behnke, Neural Networks 2017]

The Data Problem

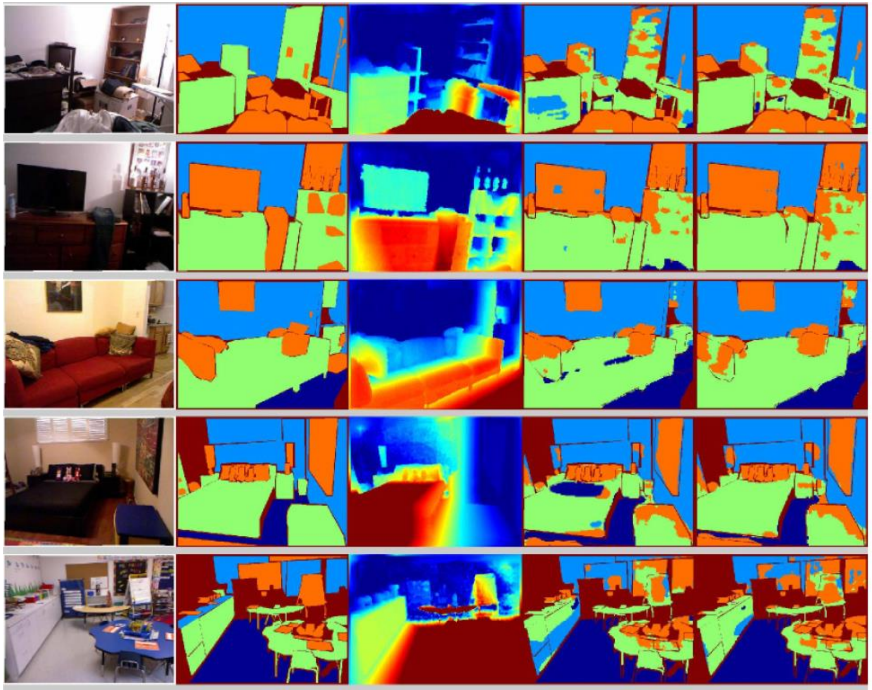
- Deep Learning in robotics (still) suffers from shortage of available examples
- We address this problem in two ways:

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

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

Geometric and Semantic Features for RGB-D Object-class Segmentation

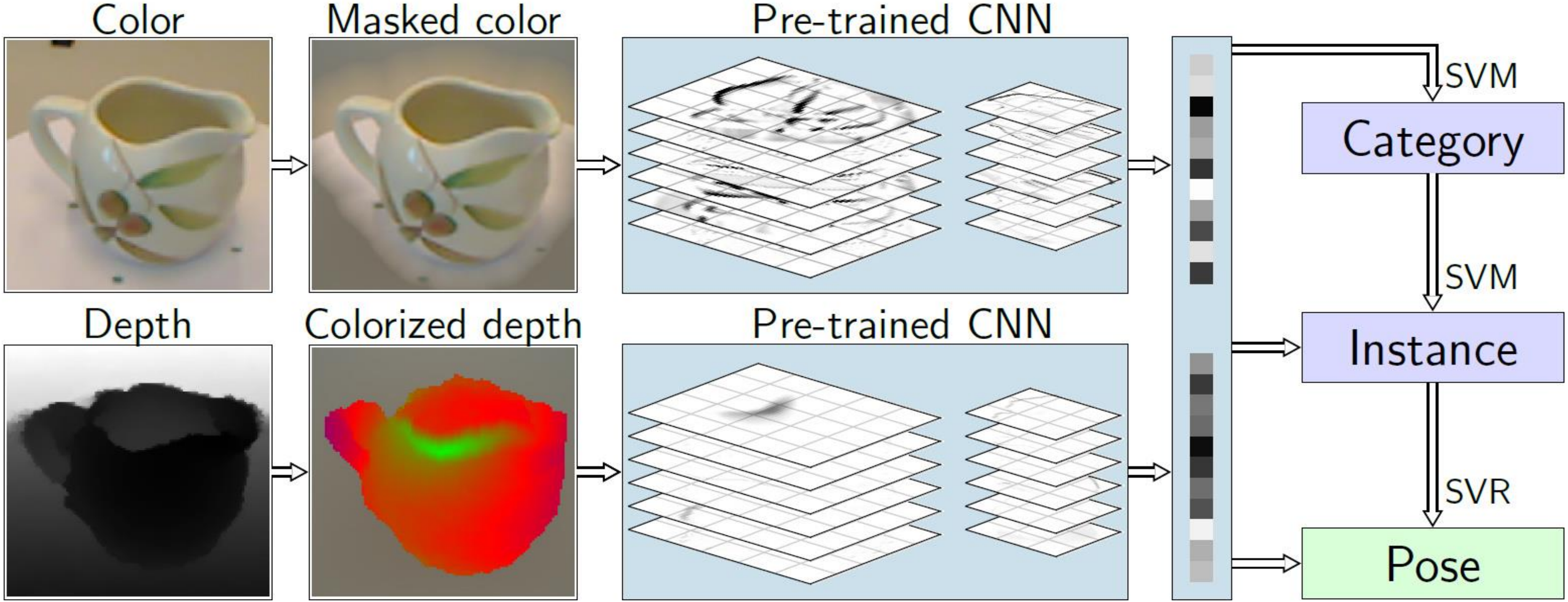
- New **geometric** feature: distance from wall
- **Semantic** features pretrained from ImageNet
- Both help significantly



RGB Truth DistWall OutWO OutWithDistWall

[Husain et al. RA-L 2017]

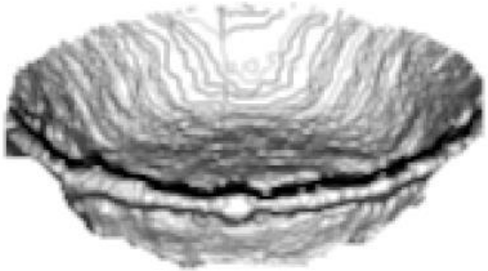
RGB-D Object Recognition and Pose Estimation

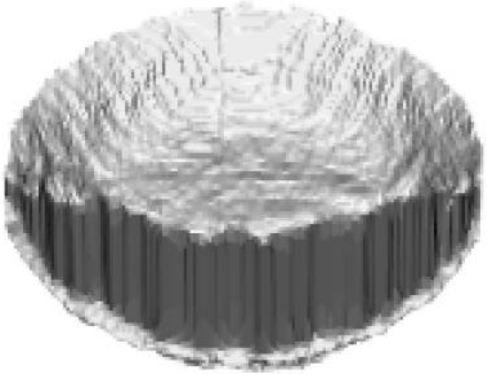


[Schwarz, Schulz, Behnke, ICRA2015]

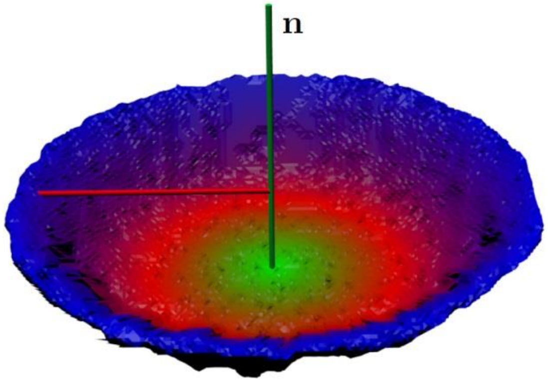
Canonical View, Colorization

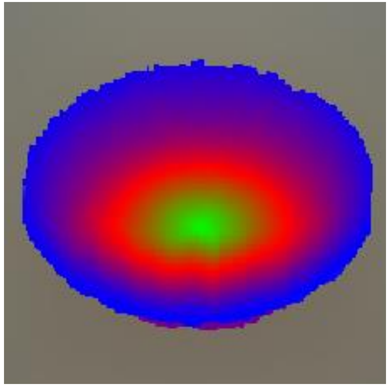
- Objects viewed from different elevation
- Render canonical view





- Colorization based on distance from center vertical

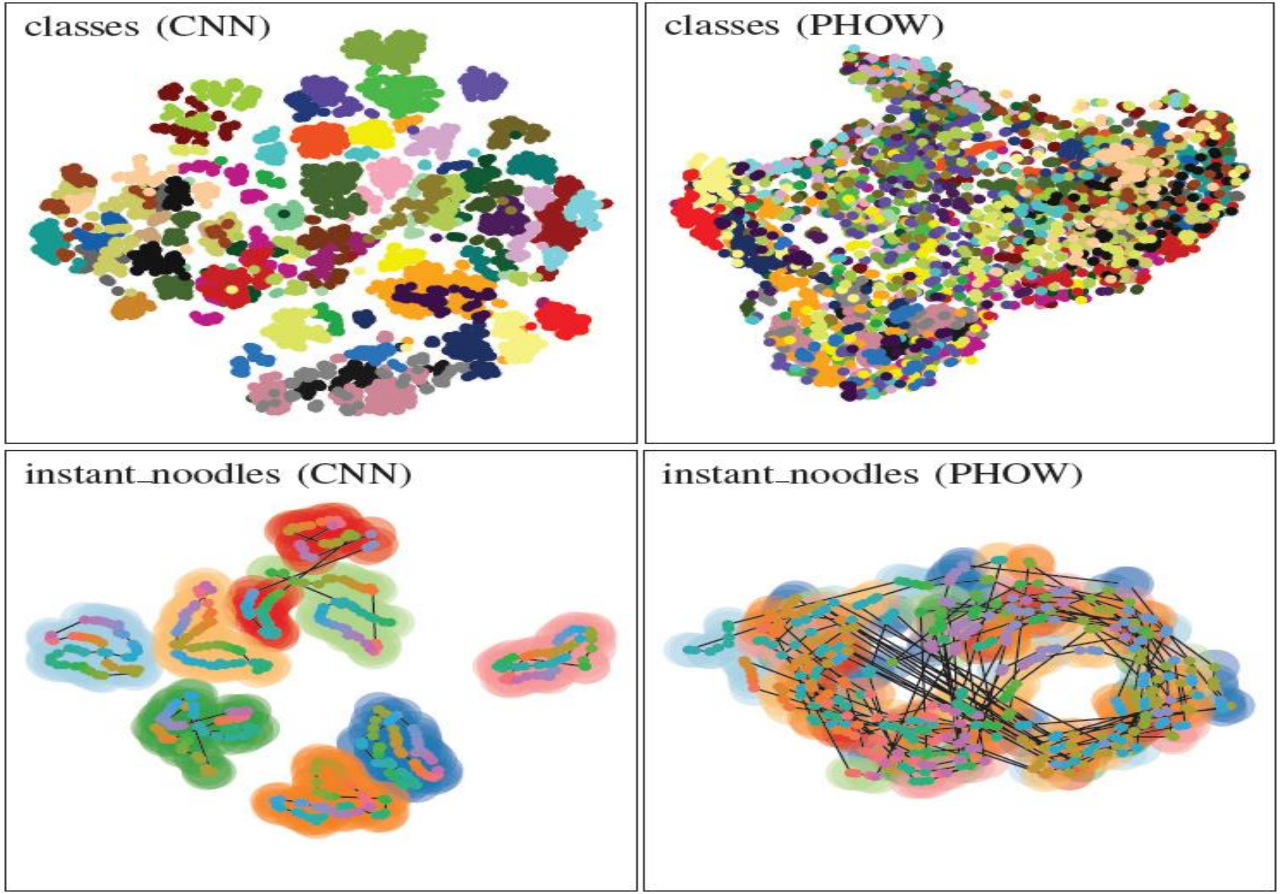




[Schwarz, Schulz, Behnke, ICRA2015]

Pretrained Features Disentangle Data

- t-SNE embedding



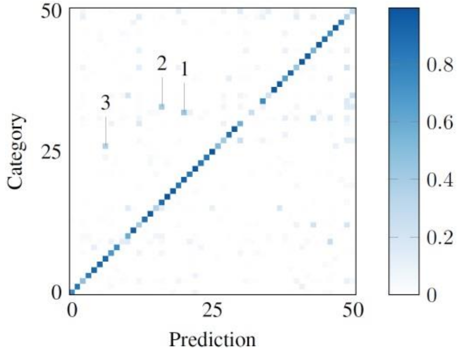
[Schwarz, Schulz, Behnke ICRA2015]

Recognition Accuracy

- Improved both category and instance recognition

Method	Category Accuracy (%) RGB	Category Accuracy (%) RGB-D	Instance Accuracy (%) RGB	Instance Accuracy (%) RGB-D
Lai *et al.* [1]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo *et al.* [2]	82.4 ± 3.1	87.5 ± 2.9	**92.1**	92.8
PHOW[3]	80.2 ± 1.8	—	62.8	—
Ours	$\mathbf{83.1 \pm 2.0}$	88.3 ± 1.5	92.0	**94.1**
Ours	$\mathbf{83.1 \pm 2.0}$	$\mathbf{89.4 \pm 1.3}$	92.0	**94.1**

- Confusion:



1: pitcher / coffe mug


2: peach / sponge


[Schwarz, Schulz, Behnke, ICRA2015]

Object Capture and Scene Rendering

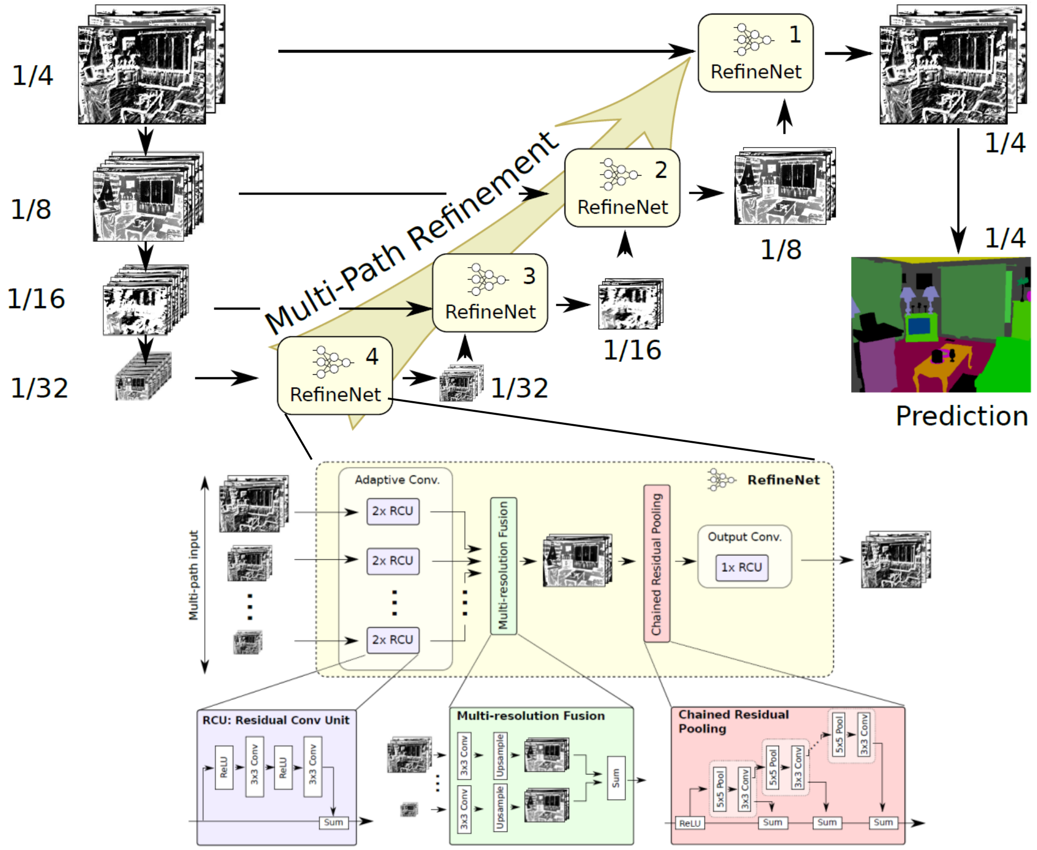
- Turntable + DLSR camera

- Rendered scenes

[Schwarz et al. ICRA 2018]

RefineNet for Semantic Segmentation

- Scene represented as feature hierarchy
- Corse-to-fine semantic segmentation
- Combine higher-level features with missing details

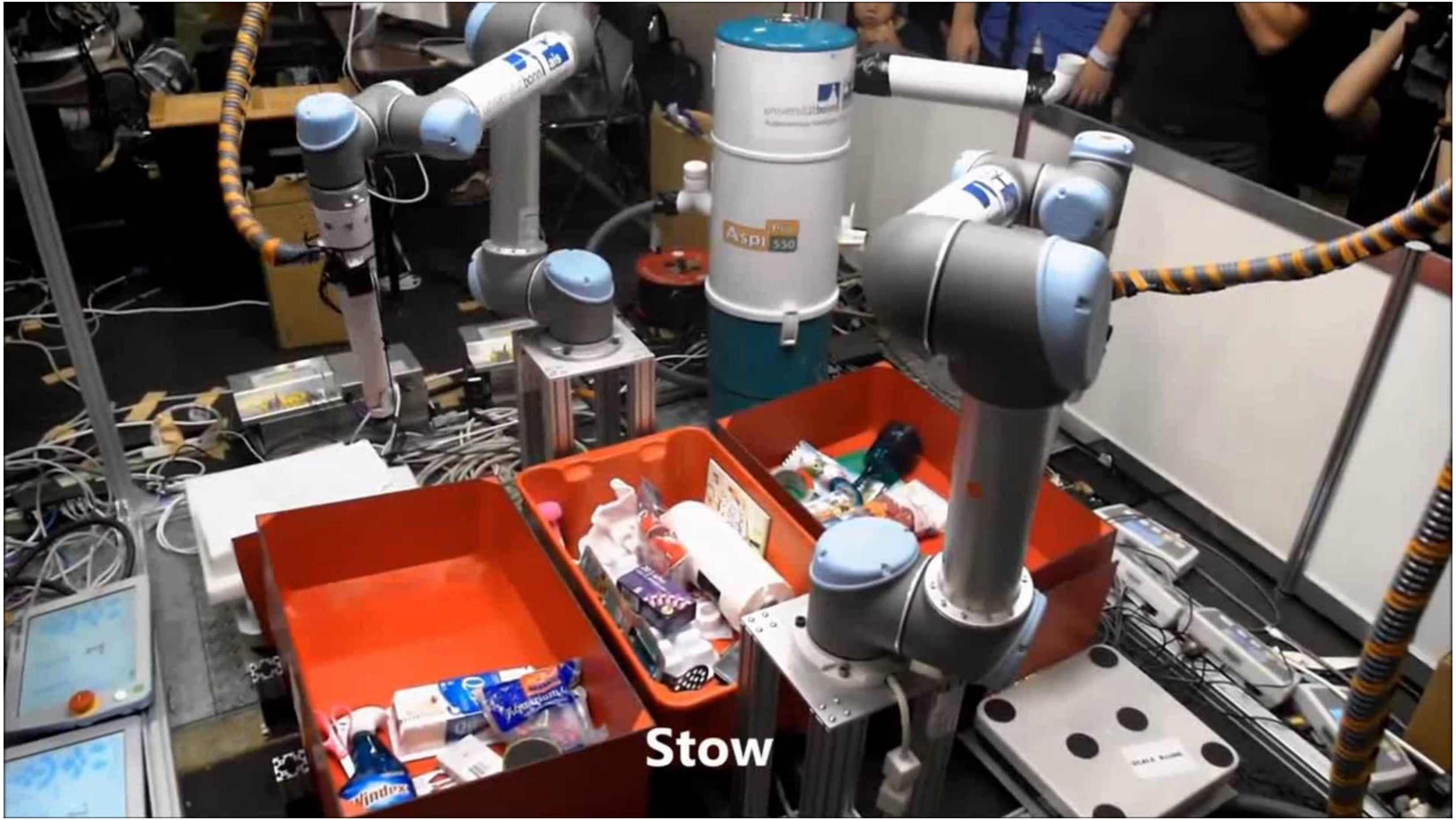


[Lin et al. CVPR 2017]

Semantic Segmentation Example

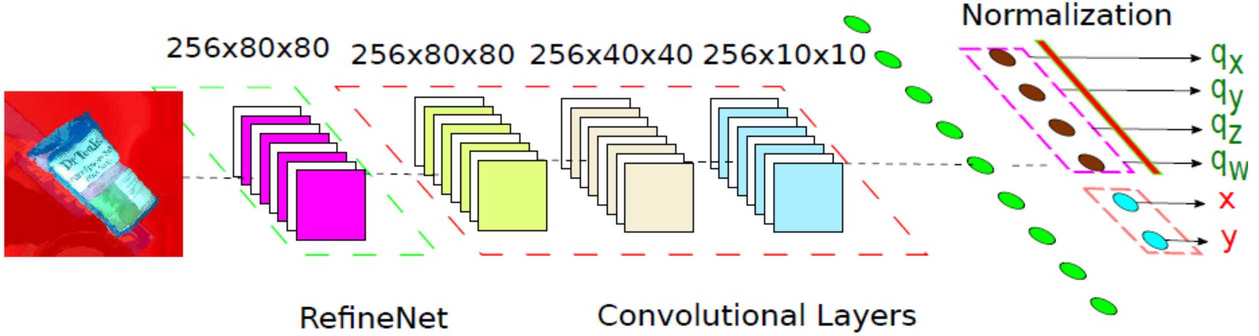
[Schwarz et al. ICRA 2018]

Amazon Robotics Challenge 2017



Object Pose Estimation

- Cut out individual segments
- Use upper layer of RefineNet as input
- Predict pose coordinates



Input

Predicted pose

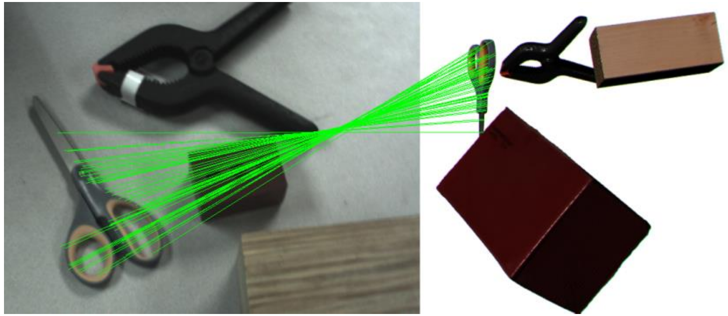
[Schwarz et al. ICRA 2018, Periyasamy et al. IROS 2018]

From Turntable Captures to Textured Meshes

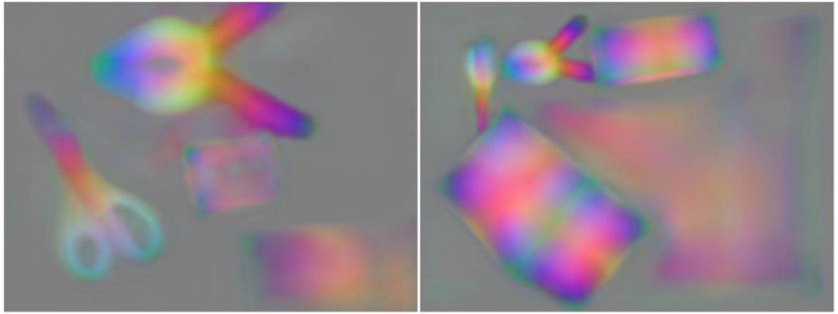
Fused & textured result

Self-Supervised Surface Descriptor Learning

- Feature descriptor should be constant under different transformations, viewing angles, and environmental effects such as lighting changes
- Descriptor should be unique to facilitate matching across different frames or representations
- Learn dense features using a contrastive loss



Known correspondences



Learned features

[Periyasamy, Schwarz, Behnke Humanoids 2019]

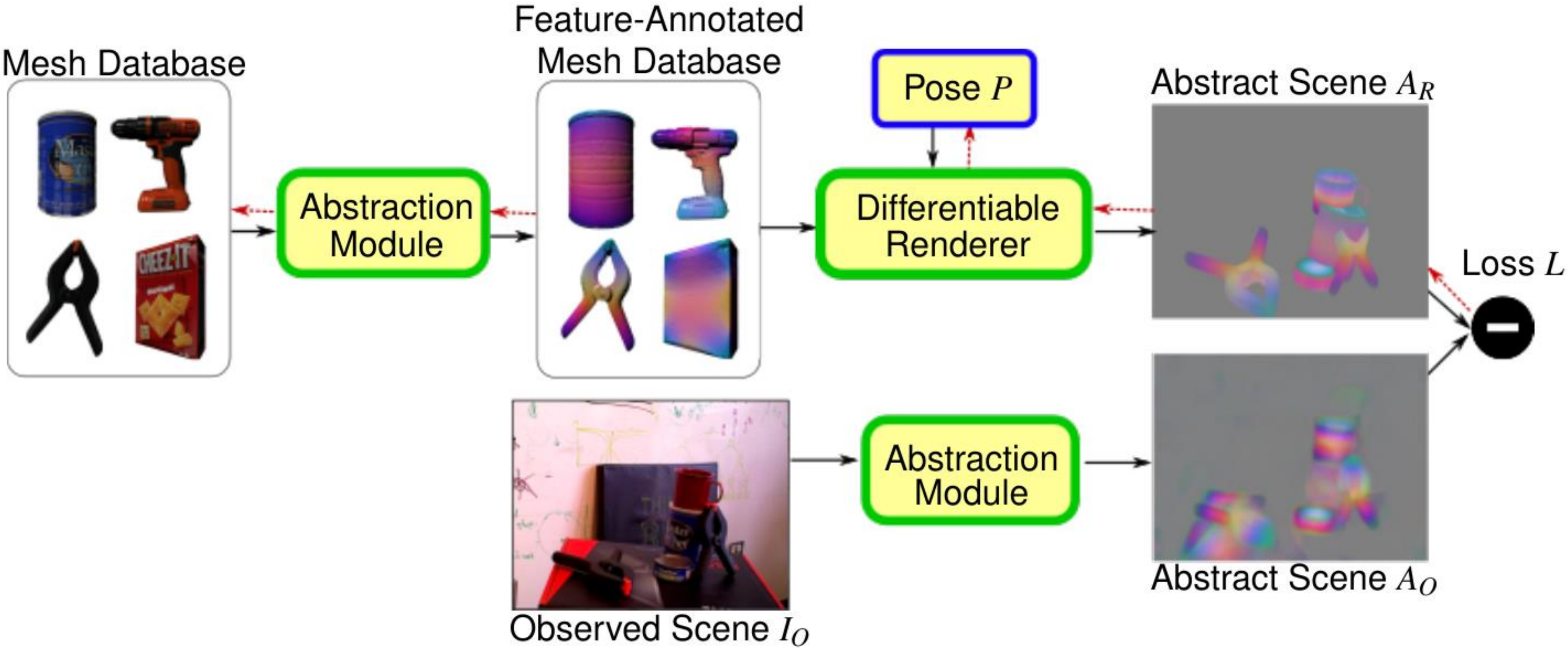
Descriptors as Texture on Object Surfaces

- Learned feature channels used as textures for 3D object models
- Used for 6D object pose estimation

UNIVERSITÄT BONN AIS

Abstract Object Registration

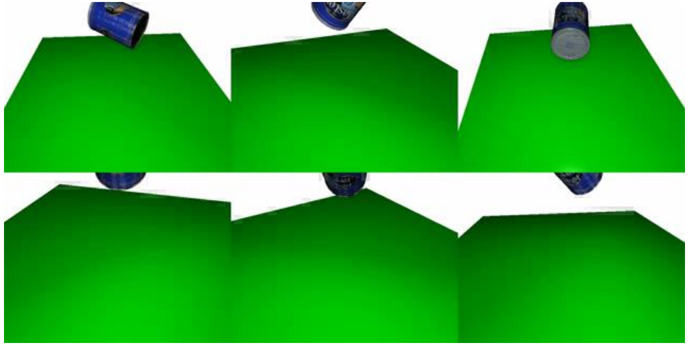
- Compare rendered and actual scene in feature space
- Adapt model pose by gradient descent

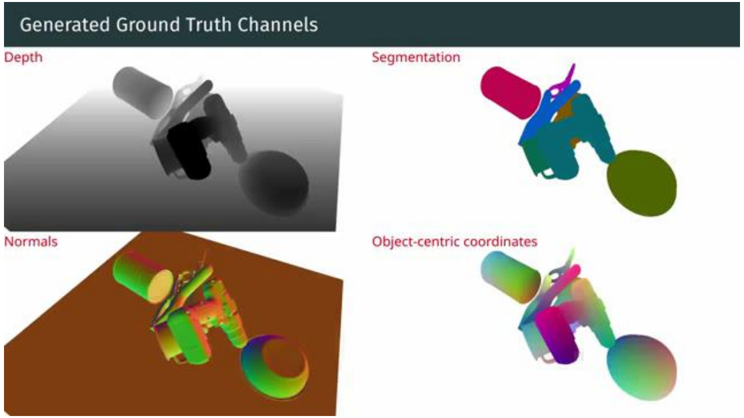


Registration Examples

Learning from Synthetic Scenes

- Cluttered arrangements from 3D meshes
- Photorealistic scenes with randomized material and lighting including ground truth
- For online learning & render-and-compare
- Semantic segmentation on YCB Video Dataset
  - Close to real-data accuracy
  - Improves segmentation of real data

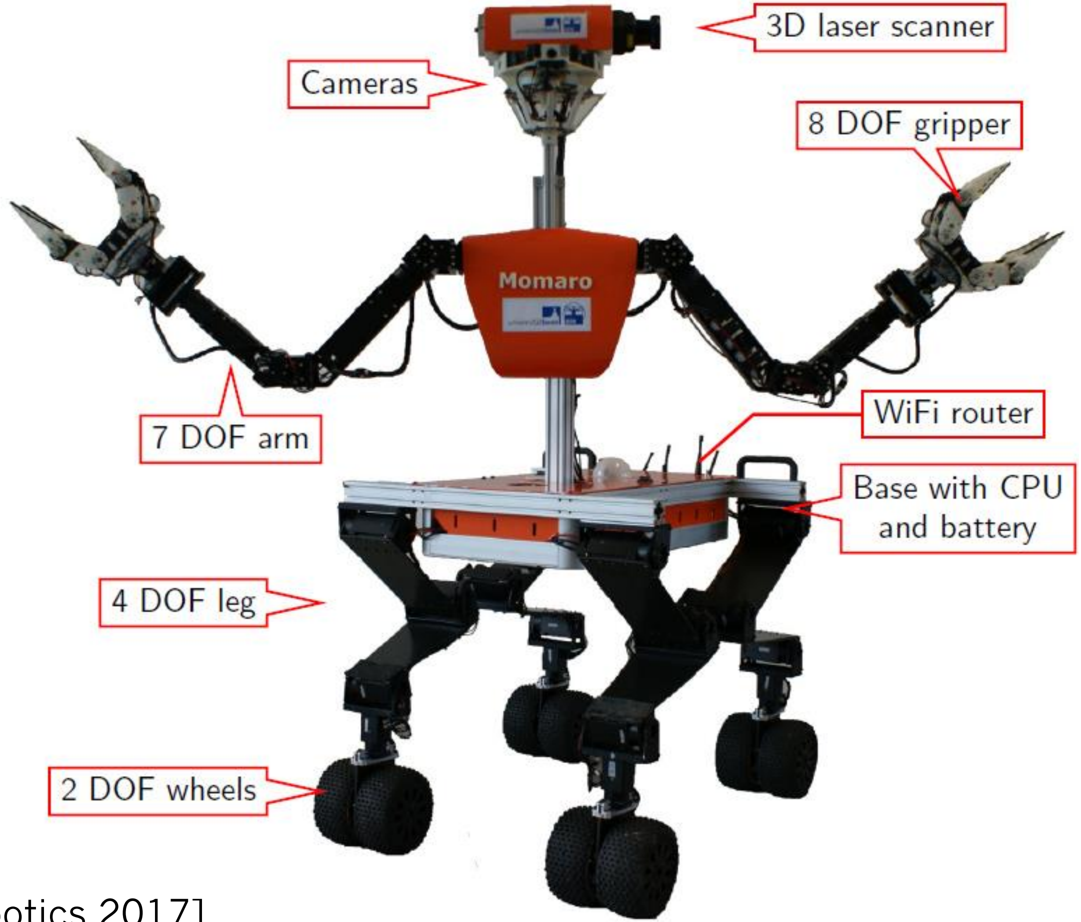




[Schwarz et al. 2020 (submitted)]

Mobile Manipulation Robot Momaro

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

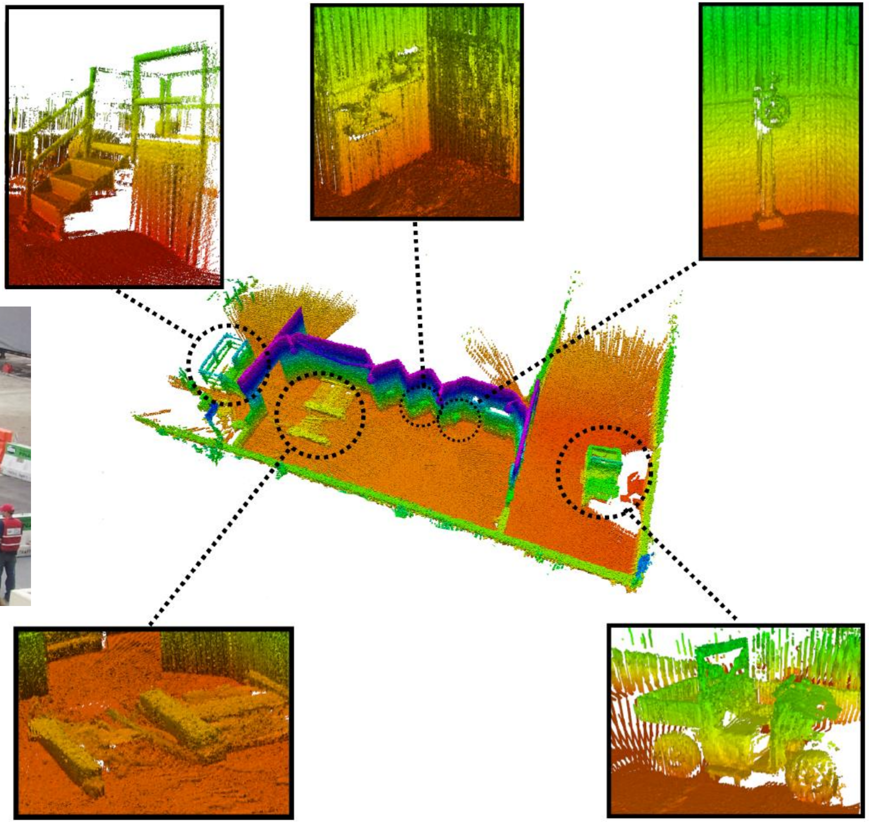


[Schwarz et al. Journal of Field Robotics 2017]

DARPA Robotics Challenge

Allocentric 3D Mapping

- Registration of egocentric maps by graph optimization

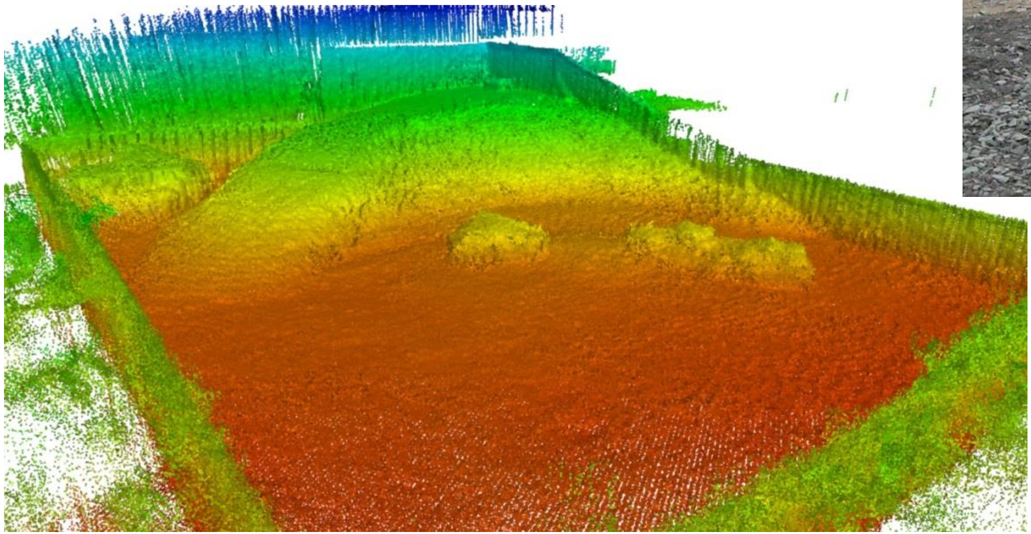


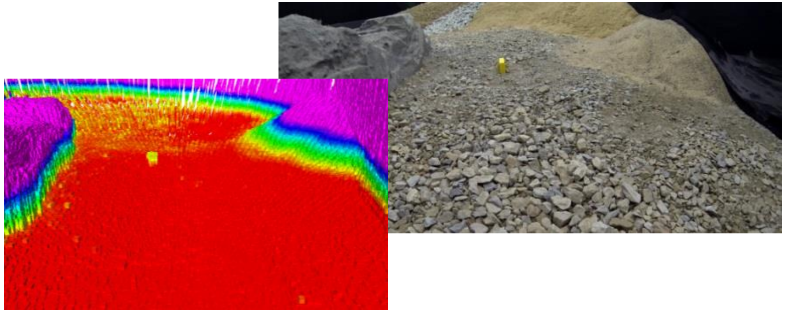
[Droeschel et al., Robotics and Autonomous Systems 2017]

DLR SpaceBot Cup 2015

- Mobile manipulation in rough terrain

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

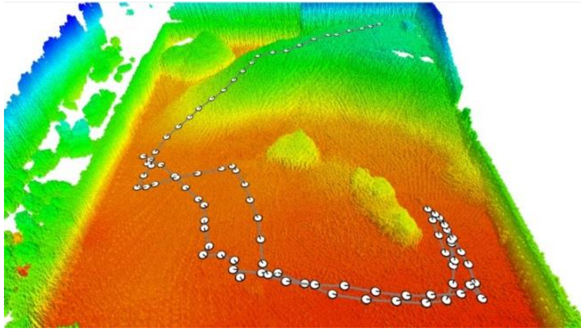


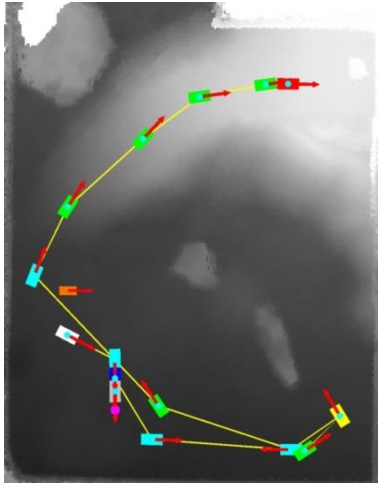


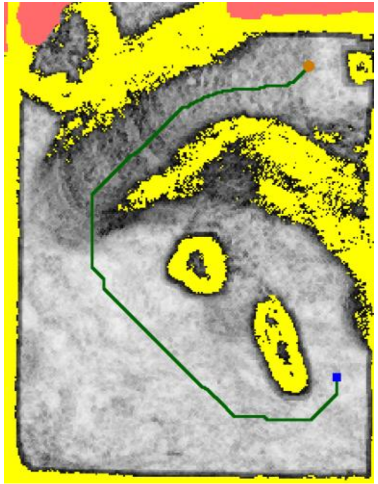

8X

Autonomous Mission Execution

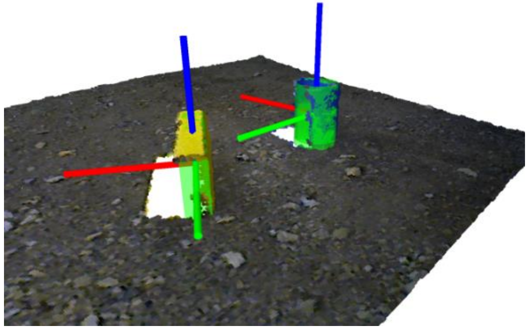
- 3D mapping, localization, mission and navigation planning

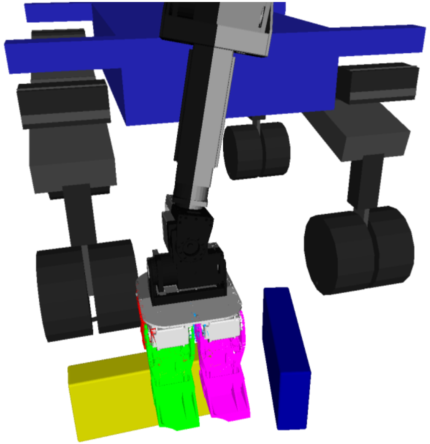






- 3D object perception and grasping

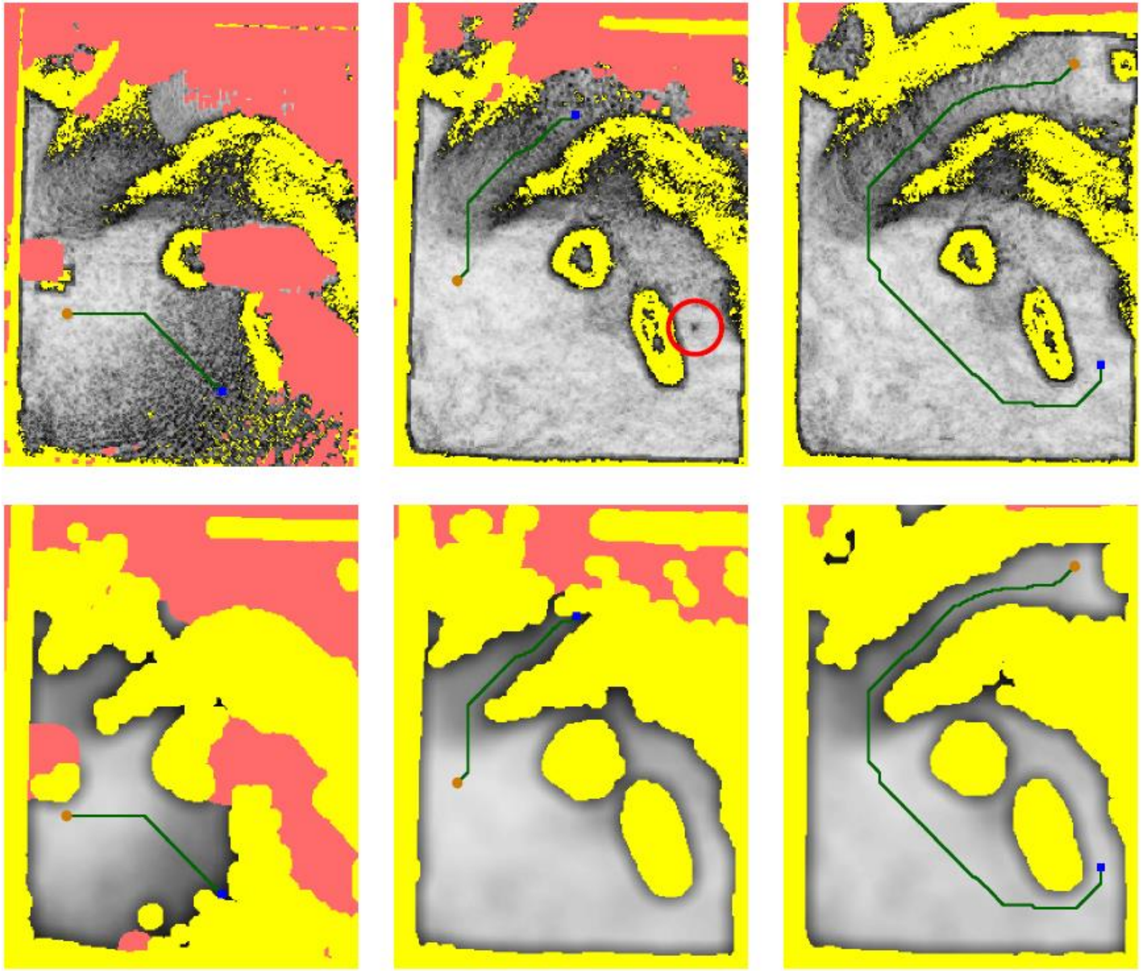




[Schwarz et al. Frontiers 2016]

Navigation Planning

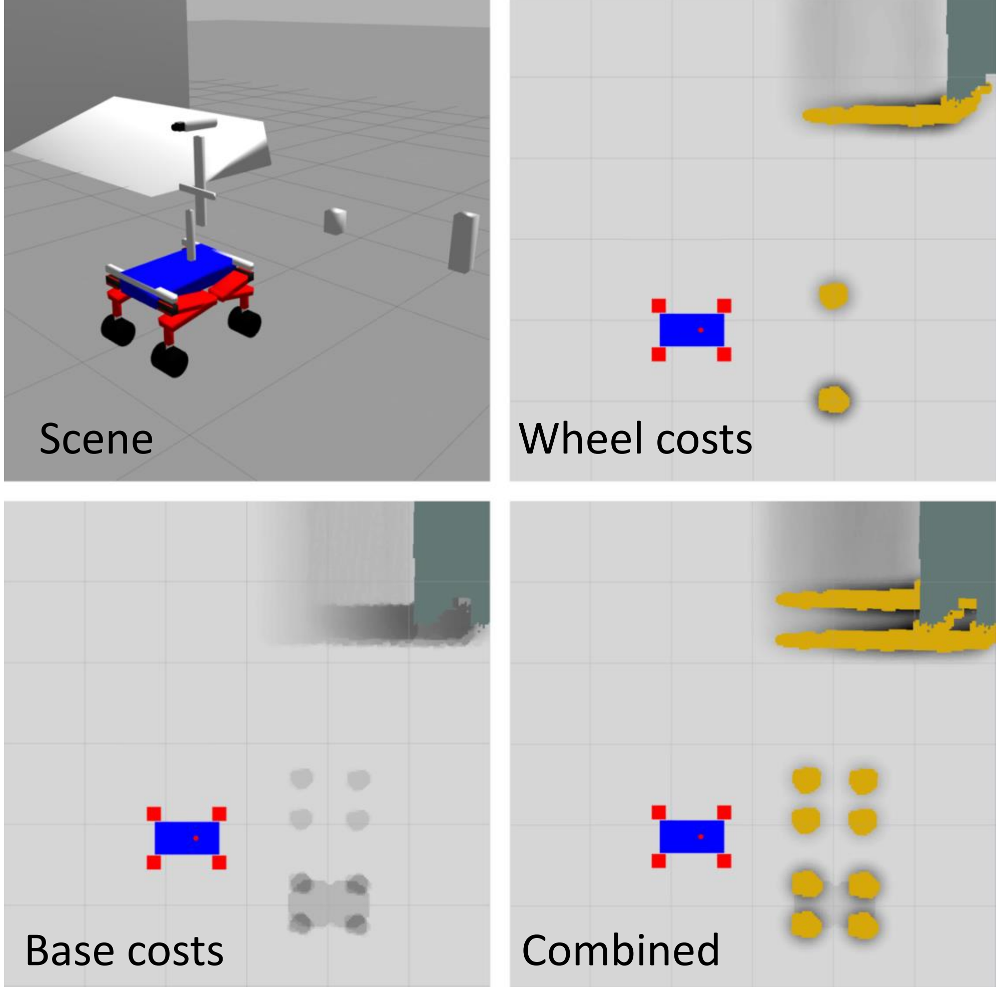
- Costs from local height differences
- A* path planning



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

Considering Robot Footprint

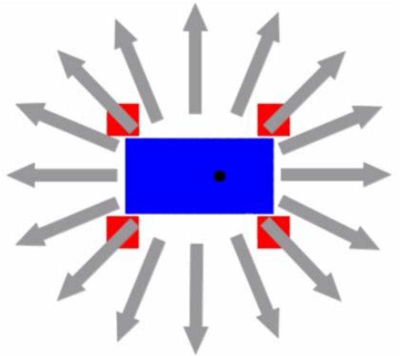
- Costs for individual wheel pairs from height differences
- Base costs
- Non-linear combination yields 3D (x, y, θ) cost map



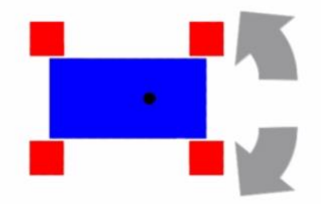
[Klamt and Behnke, IROS 2017]

3D Driving Planning (x, y, θ): A*

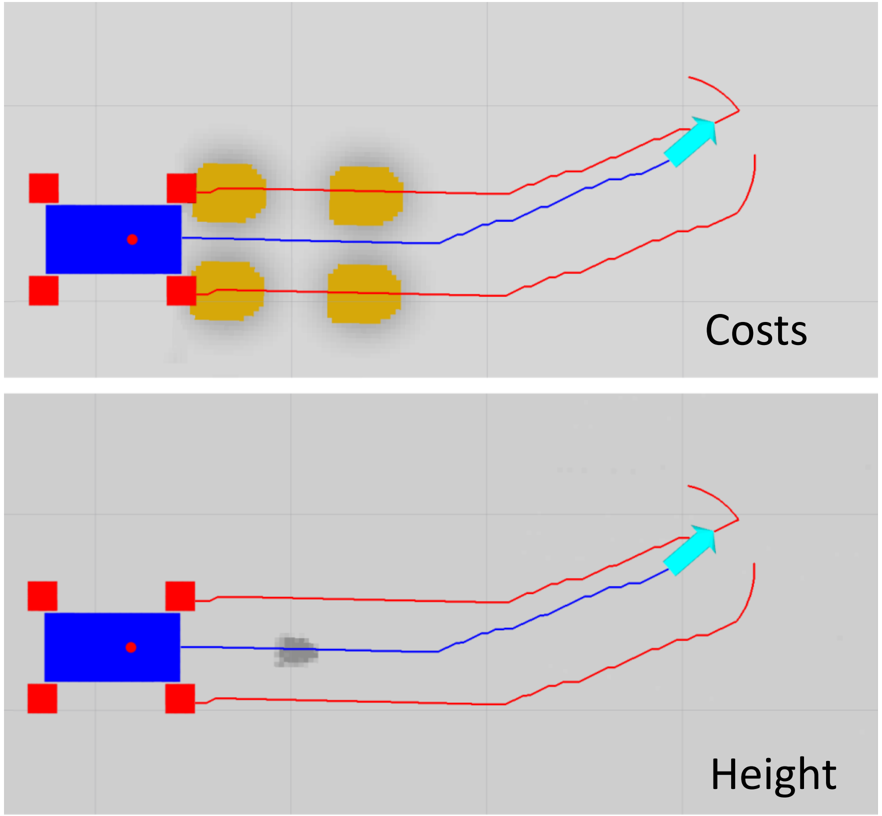
- 16 driving directions



- Orientation changes



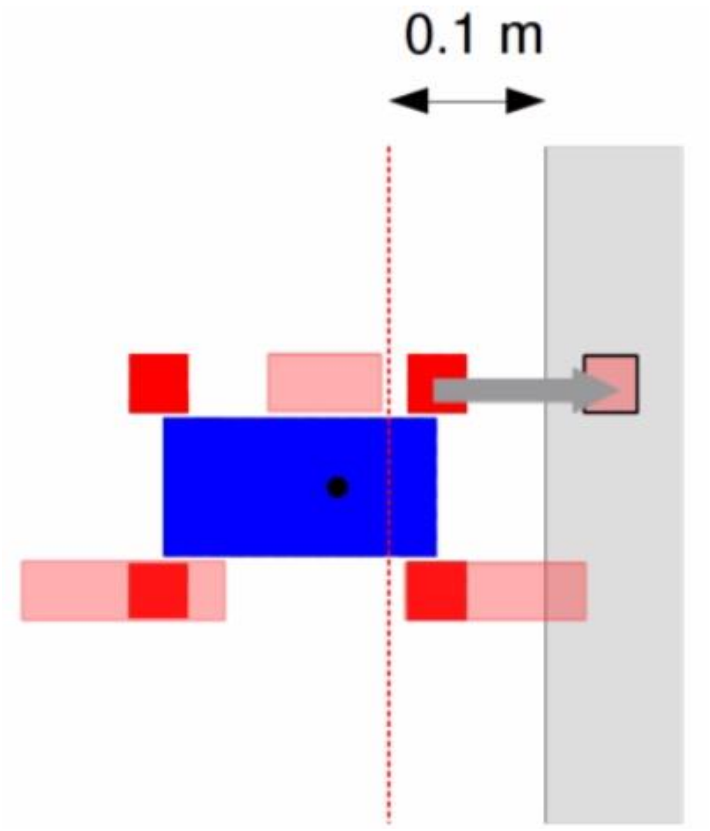
=> Obstacle between wheels



[Klamt and Behnke, IROS 2017]

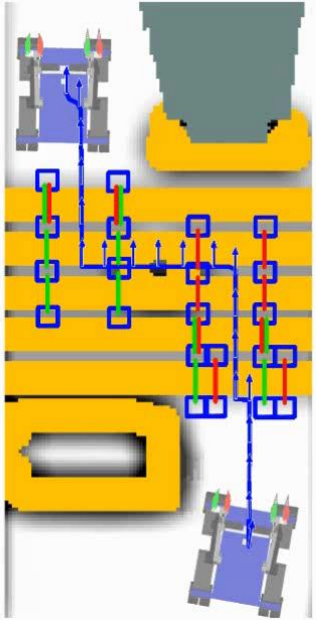
Making Steps

- If not drivable obstacle in front of a wheel
- Step landing must be drivable
- Support leg positions must be drivable



[Klamt and Behnke: IROS 2017]

Planning for Challenging Scenarios



[Klamt and Behnke: IROS 2017]

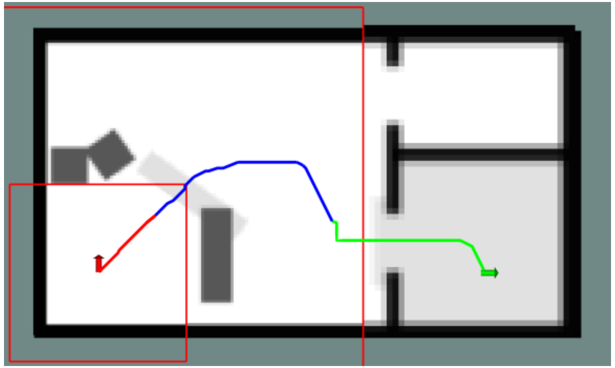
Centauro Robot

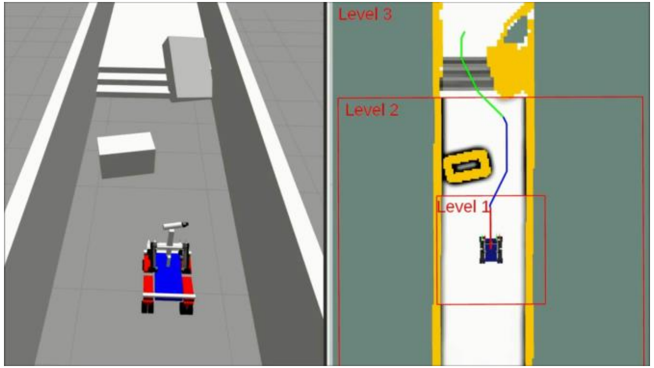
- Serial elastic actuators
- 42 main DoFs
- Schunk hand
- 3D laser
- RGB-D camera
- Color cameras
- Two GPU PCs

[Tsagarakis et al., IIT 2017]

Hybrid Driving-Stepping Locomotion Planning: Abstraction

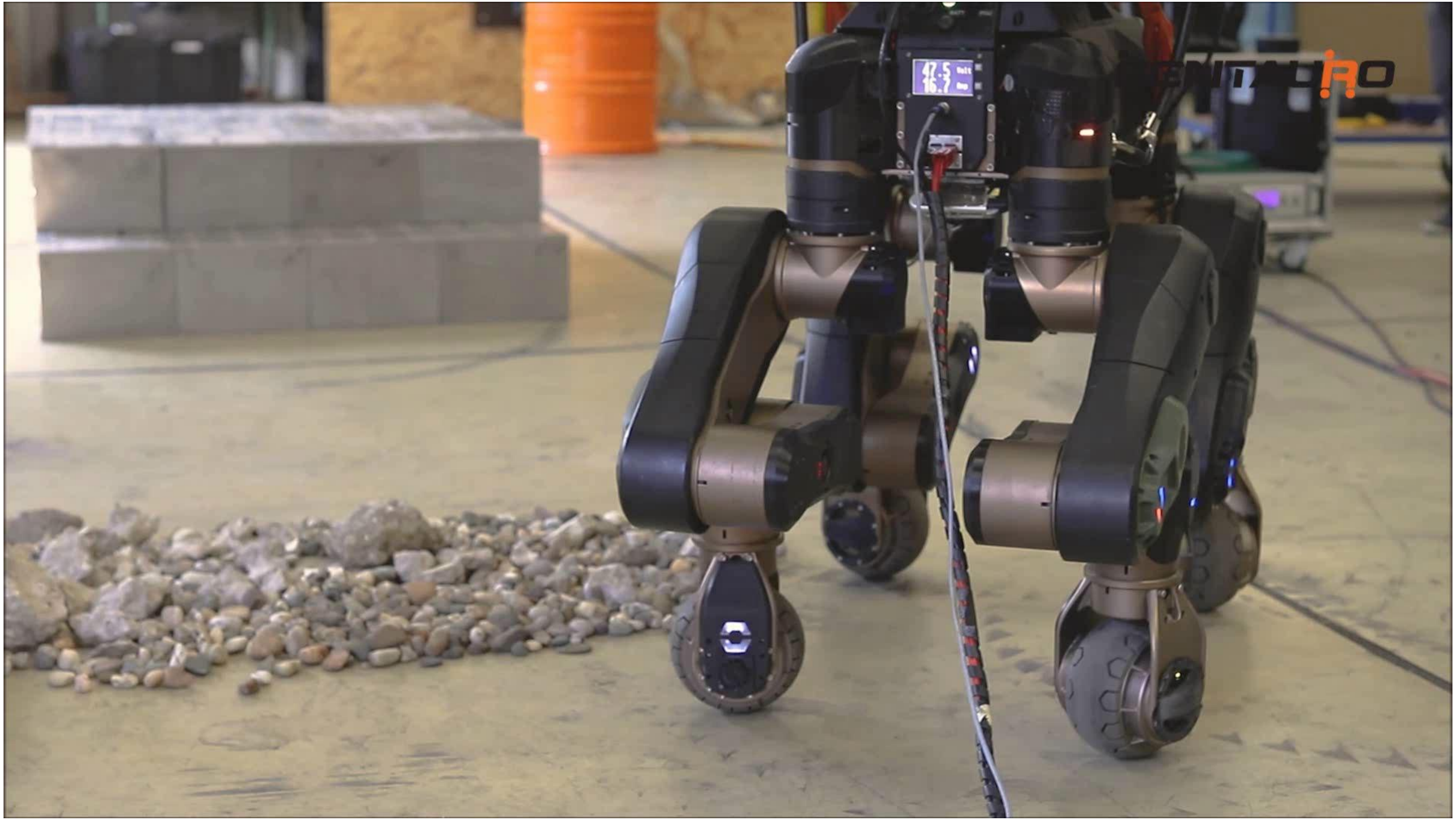
Level	Map Resolution	Map Features	Robot Representation	Action Semantics
1	• 2.5 cm • 64 orient.	• Height		• Individual Foot Actions
2	• 5.0 cm • 32 orient.	• Height • Height Difference		• Foot Pair Actions
3	• 10 cm • 16 orient.	• Height • Height Difference • Terrain Class		• Whole Robot Actions





[Klamt and Behnke, IROS 2017, ICRA 2018]

Evaluation @ KHG: Locomotion Tasks



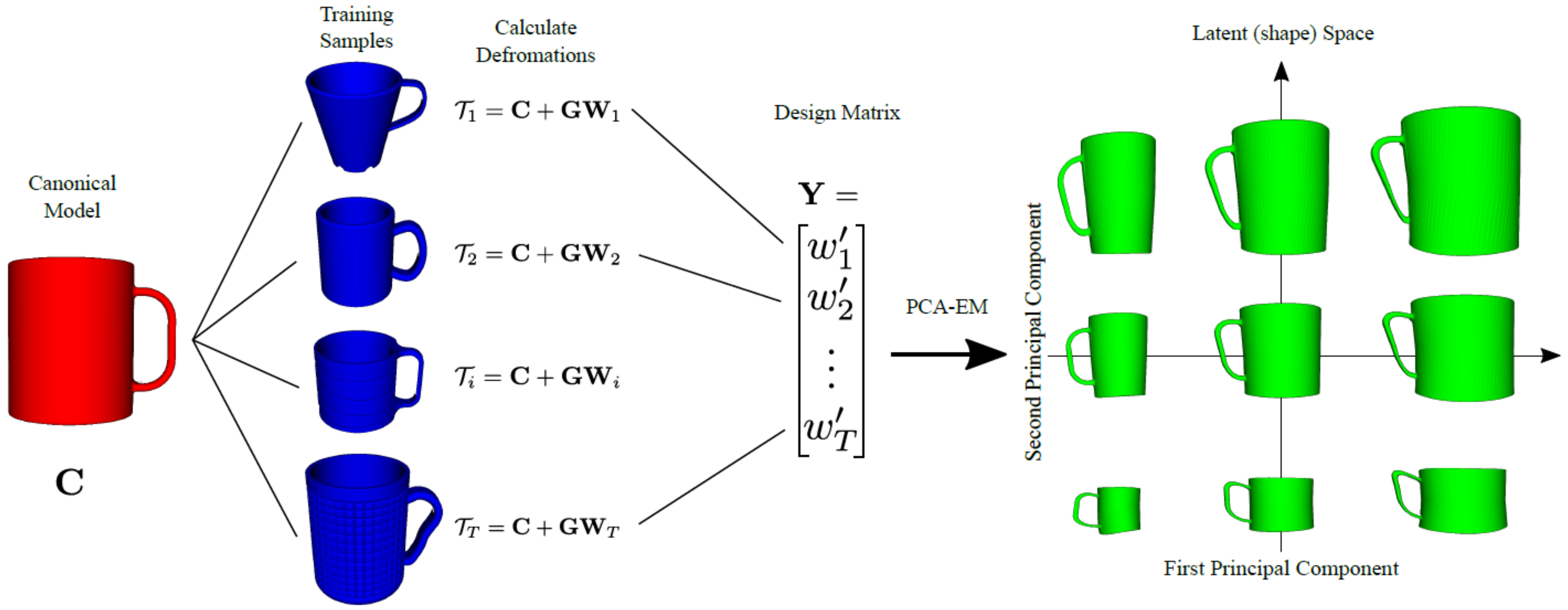
[Klamt et al. RAM 2019]

Transfer of Manipulation Skills

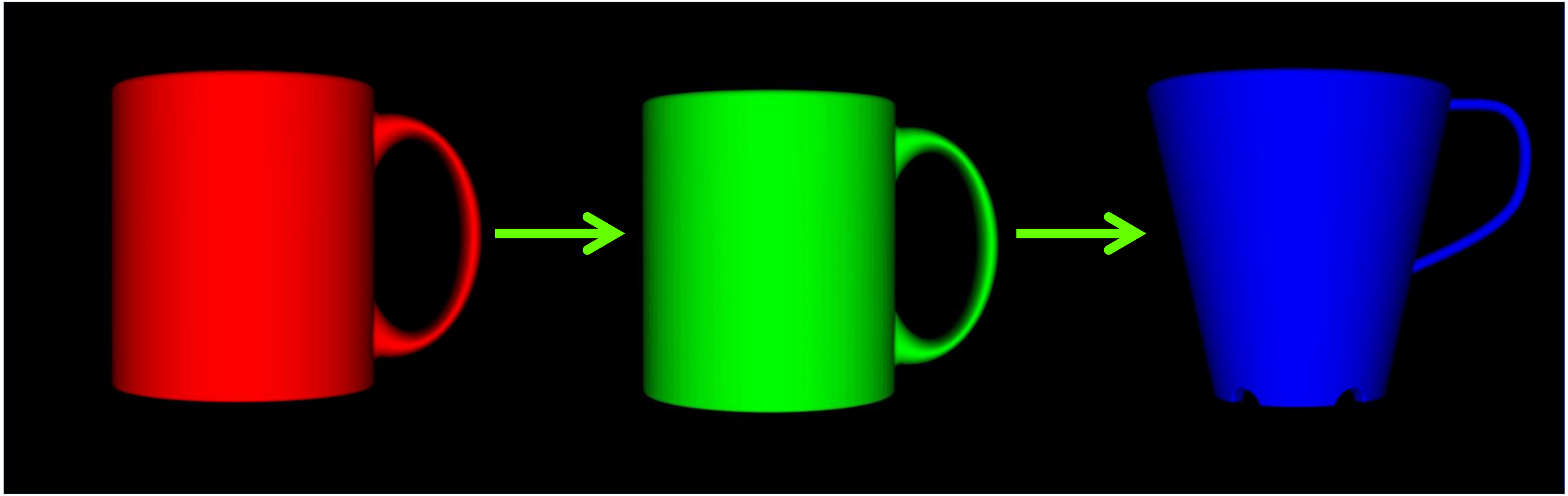
Knowledge
Transfer

Learning a Latent Shape Space

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



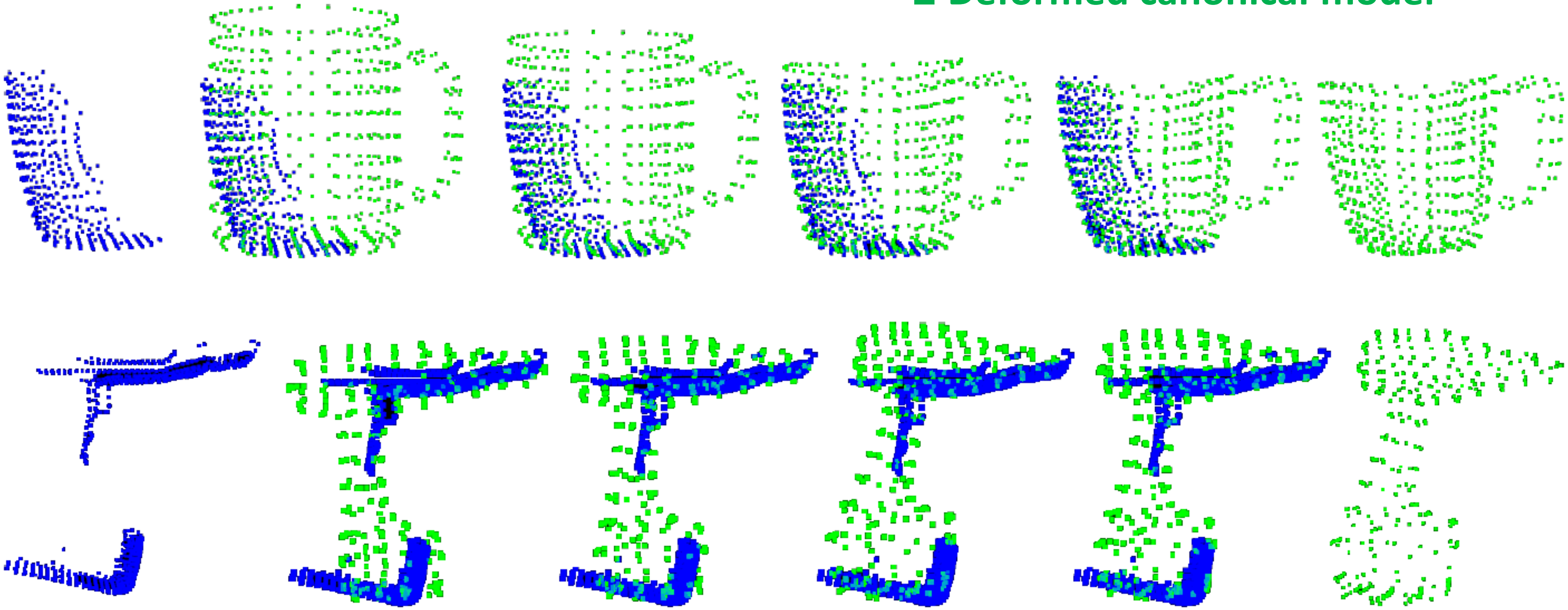
Interpolation in Shape Space



[Rodriguez and Behnke ICRA 2018]

Shape-aware Non-rigid Registration

- Partial view of novel instance
- Deformed canonical model

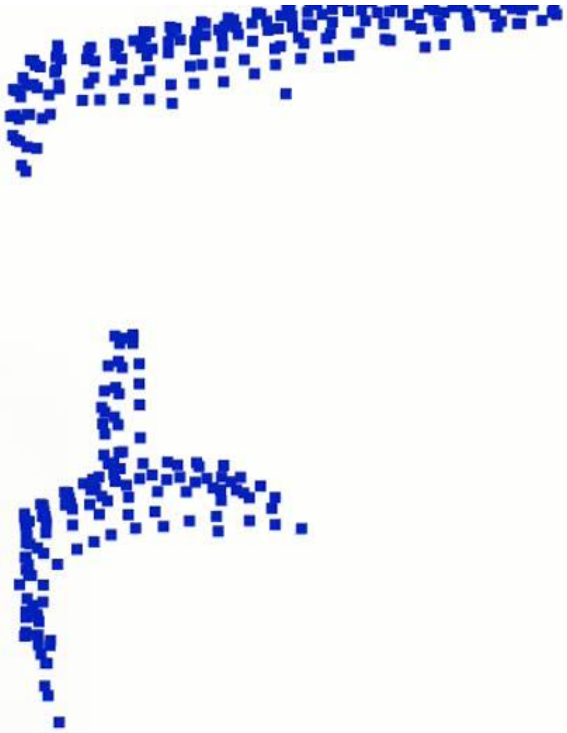


[Rodriguez and Behnke ICRA 2018]

Shape-aware Registration for Grasp Transfer

- Full point cloud

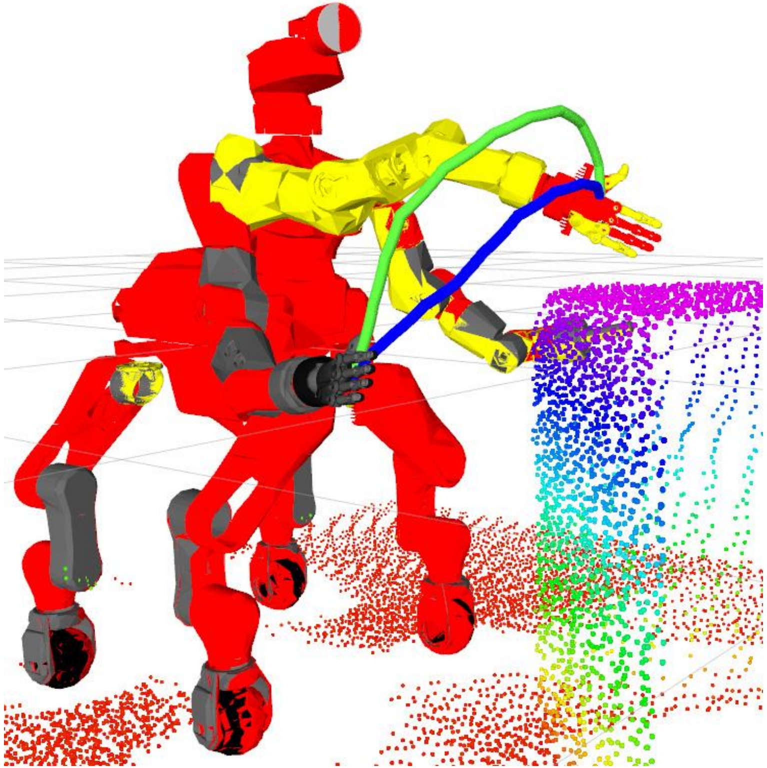
- Partial view



Collision-aware Motion Generation

Constrained Trajectory Optimization:

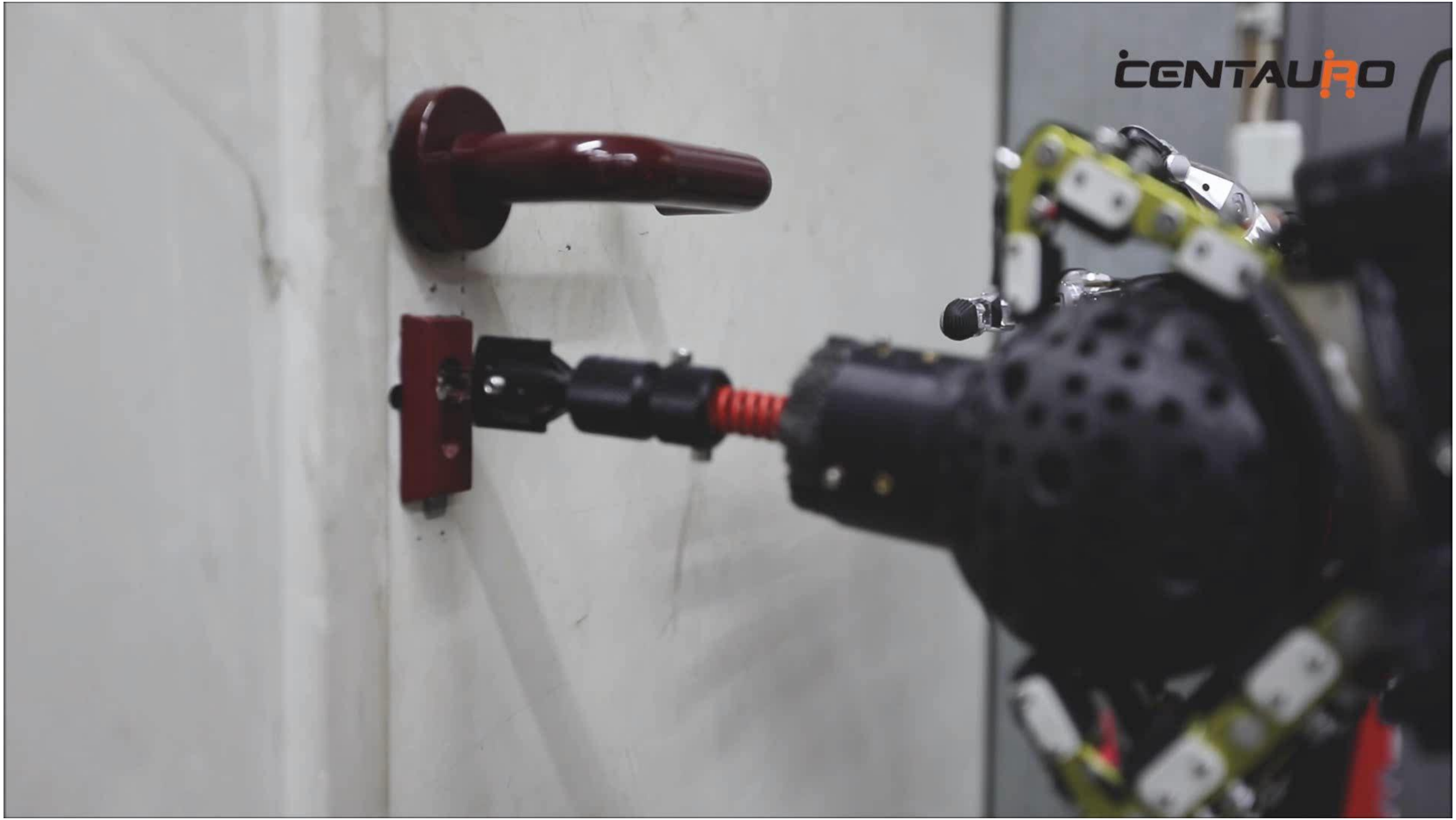
- Collision avoidance
- Joint limits
- Time minimization
- Torque optimization



[Pavlichenko et al., IROS 2017]

Grasping an Unknown Power Drill and Fastening Screws

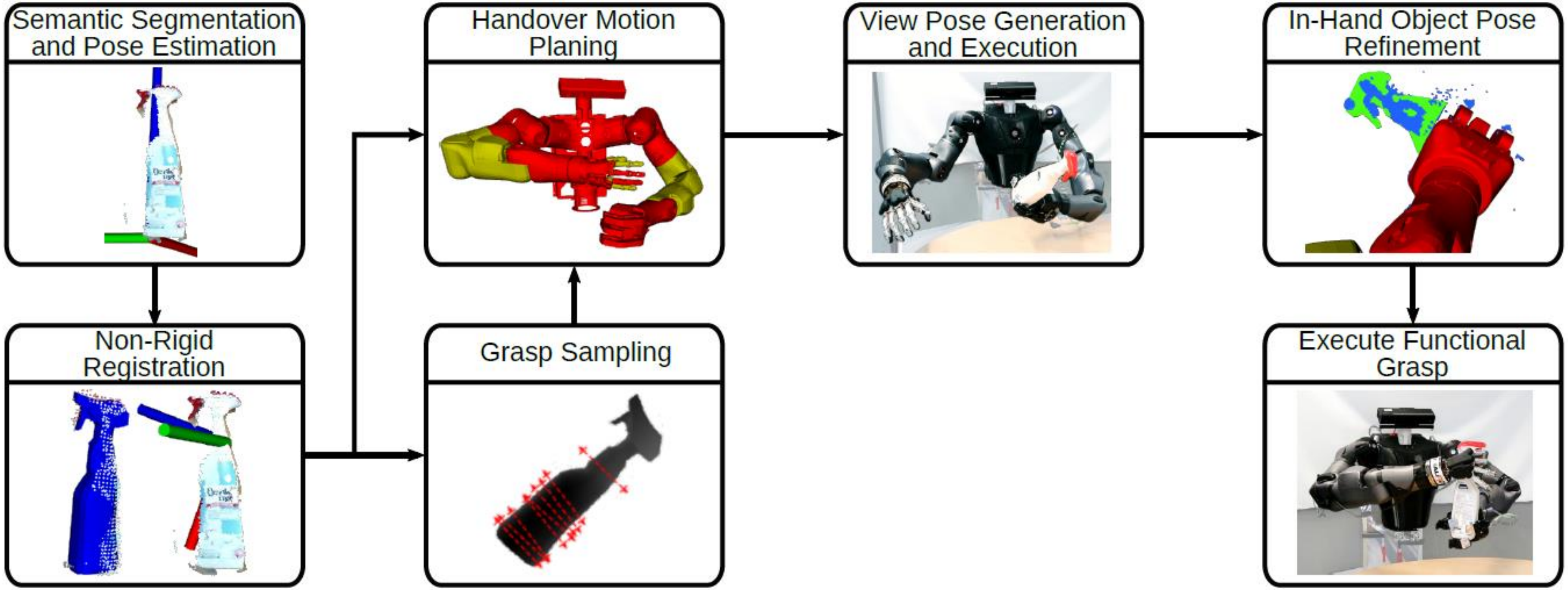
Complex Manipulation Tasks



[Klamt et al. RAM 2019]

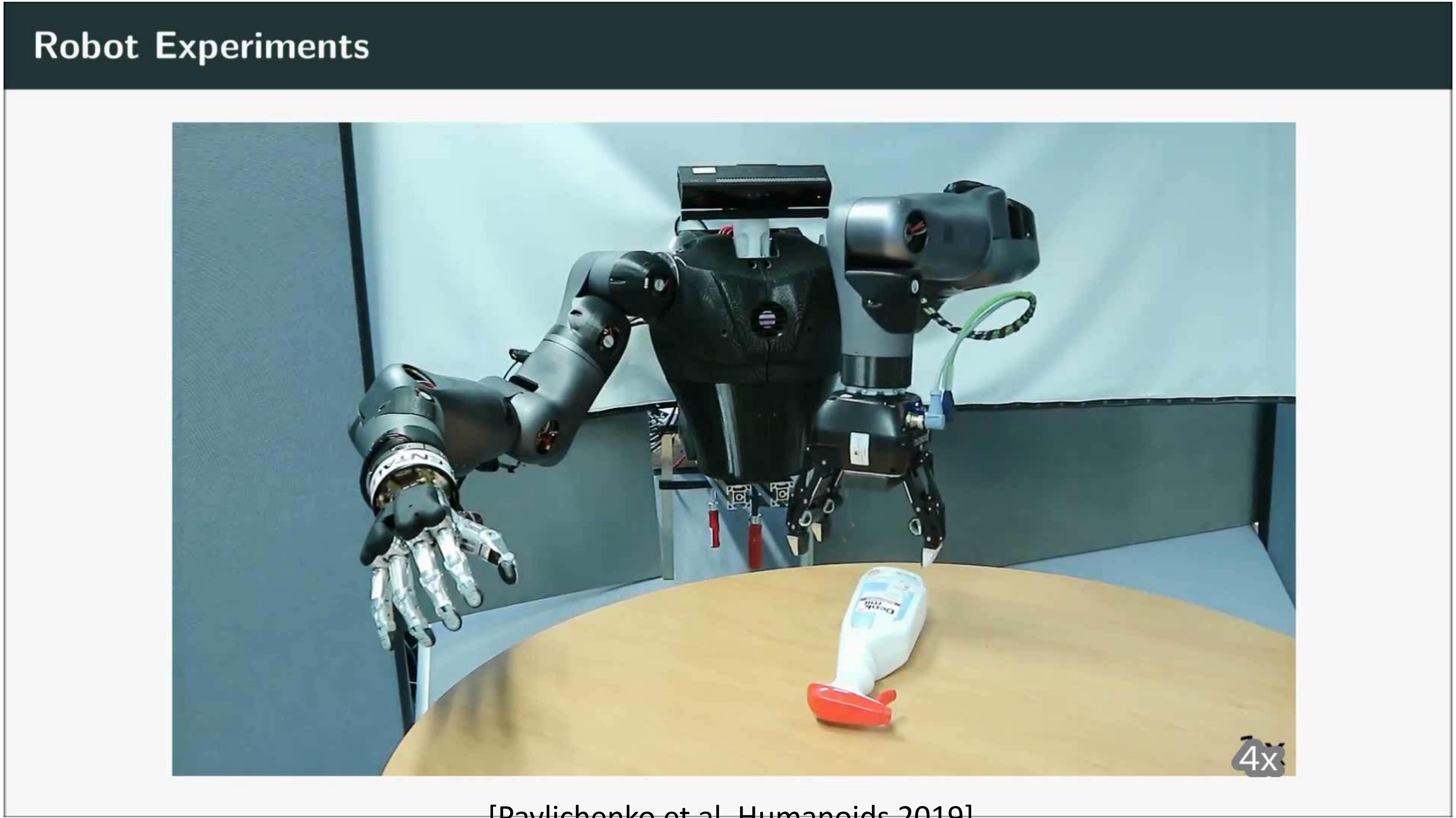
Regrasping

- Direct functional grasps not always feasible
- Pick up object with support hand, such that it can be grasped in a functional way



[Pavlichenko et al. Humanoids 2019]

Regrasping

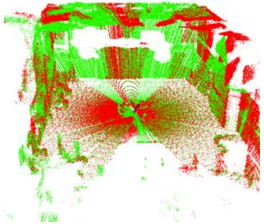


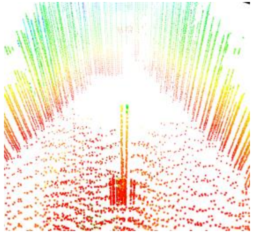
[Pavlichenko et al. Humanoids 2019]

Autonomous Flight Near Obstacles

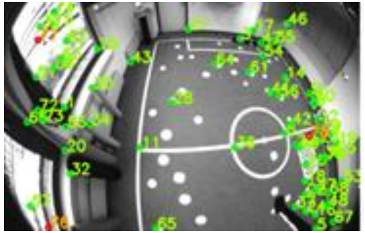
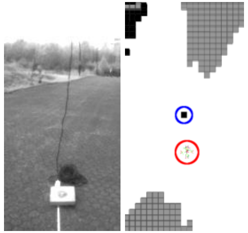
Multimodal obstacle detection

- 3D laser scanner

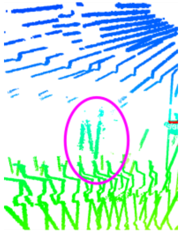
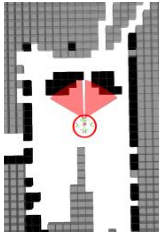


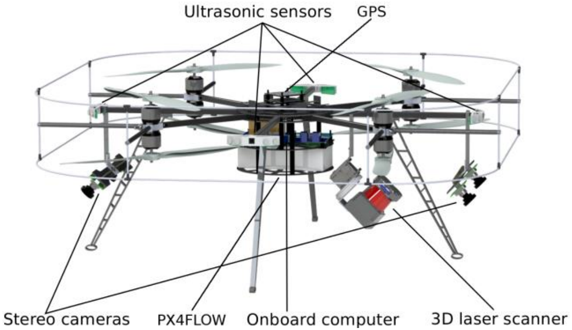
- Stereo cameras

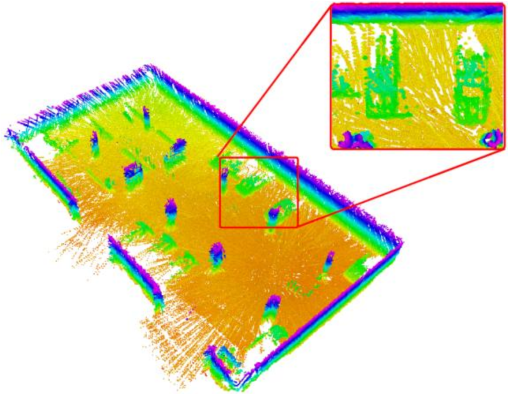




- Ultrasound



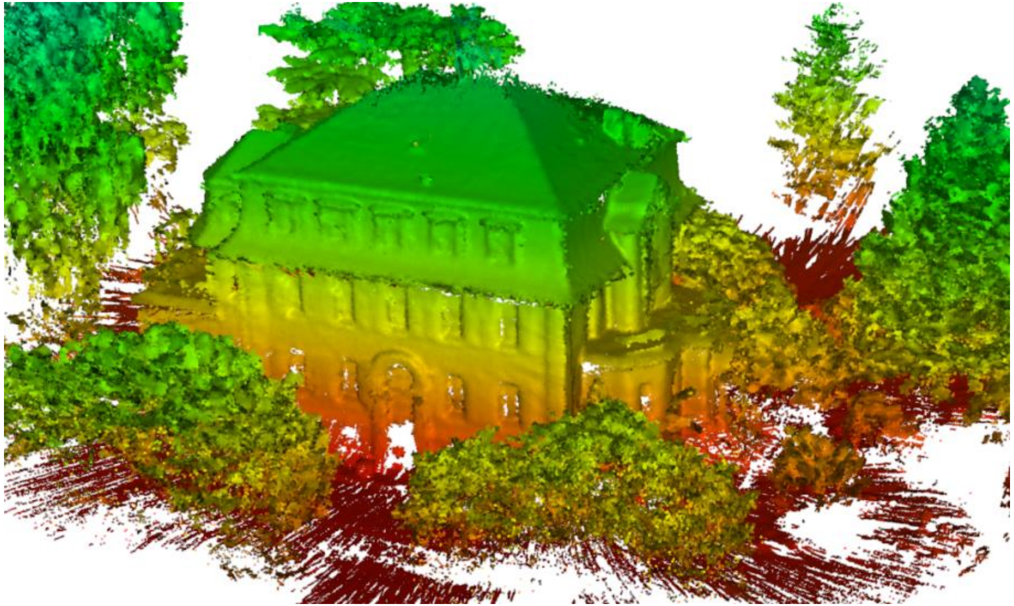




[Droeschel et al.: Journal of Field Robotics, 2015]

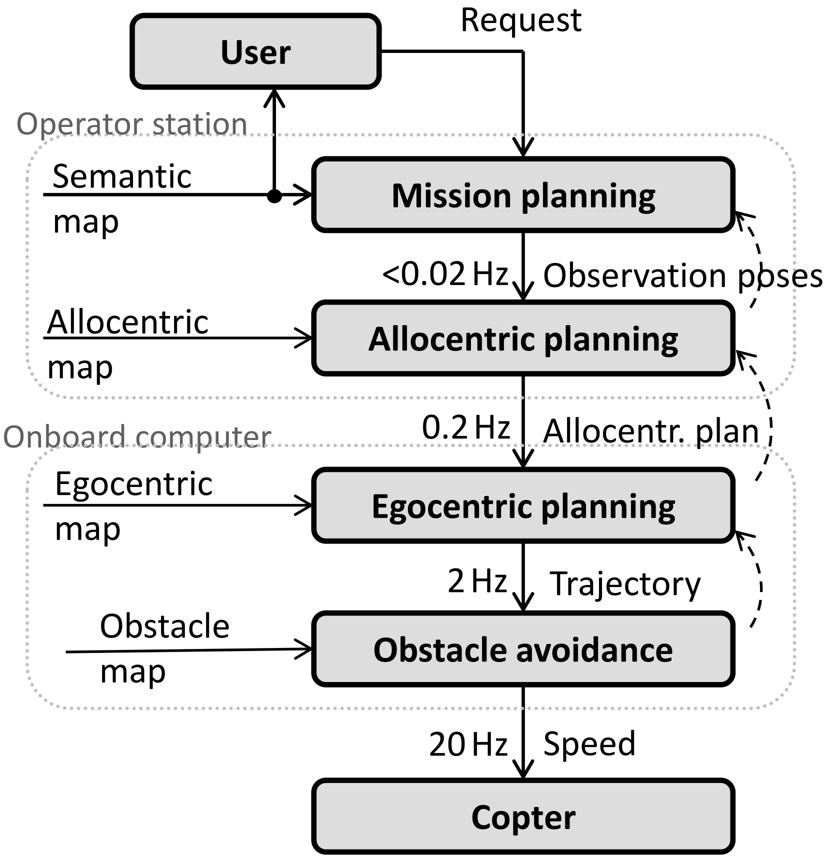
Allocentric 3D Map

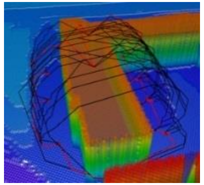
- Registration of egocentric maps
- Global optimization of registration error by GraphSLAM



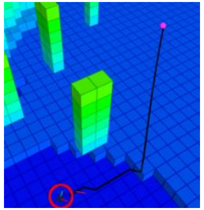
[Droeschel et al. JFR 2016]

Hierarchical Navigation

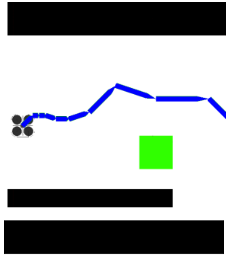




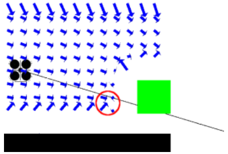
Mission plan



Allocentric planning



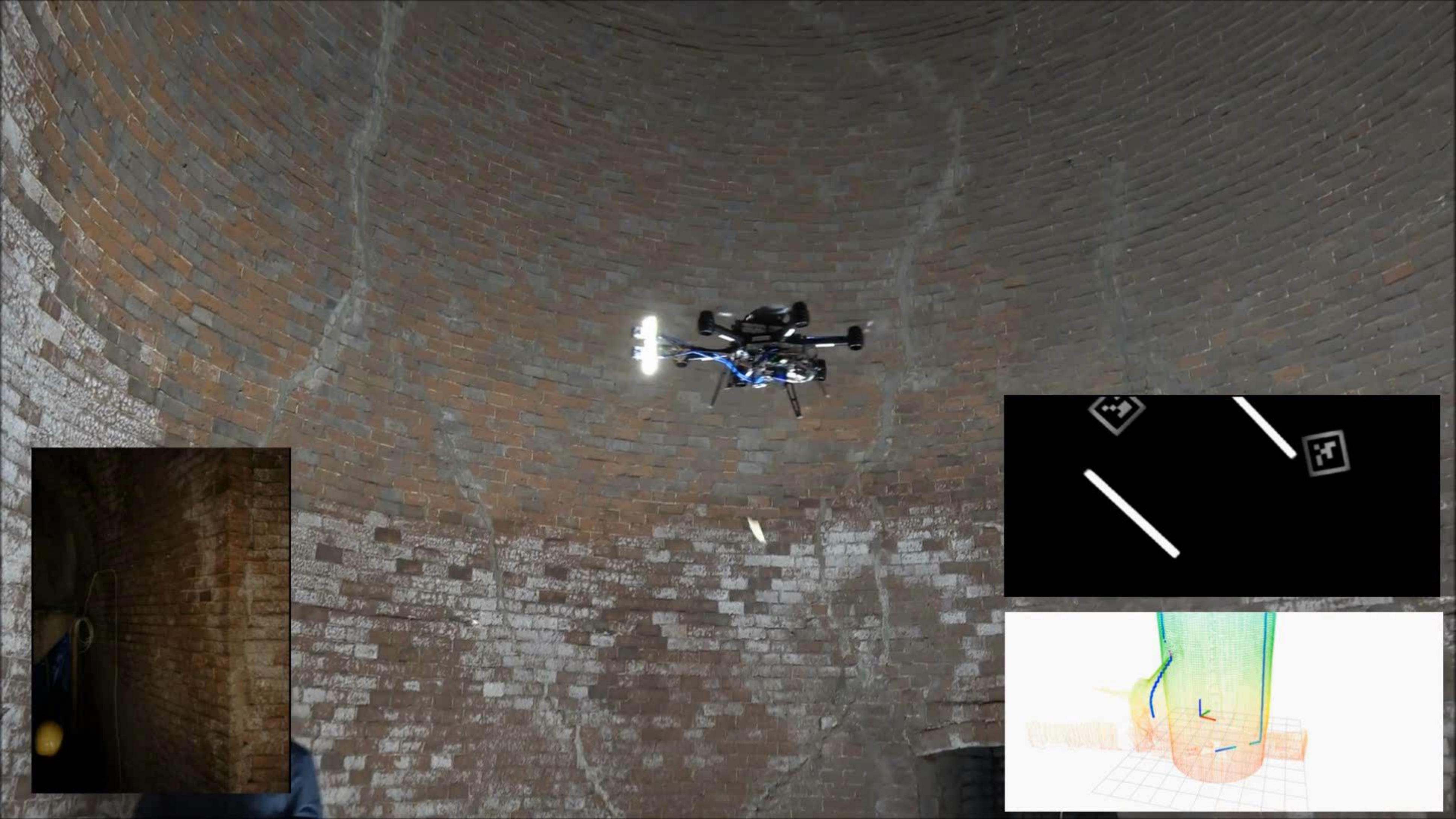
Egocentric planning



Obstacle avoidance

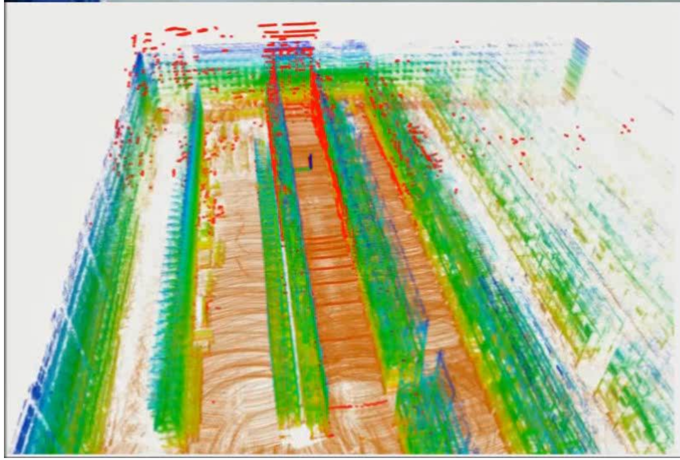
Mapping on Demand

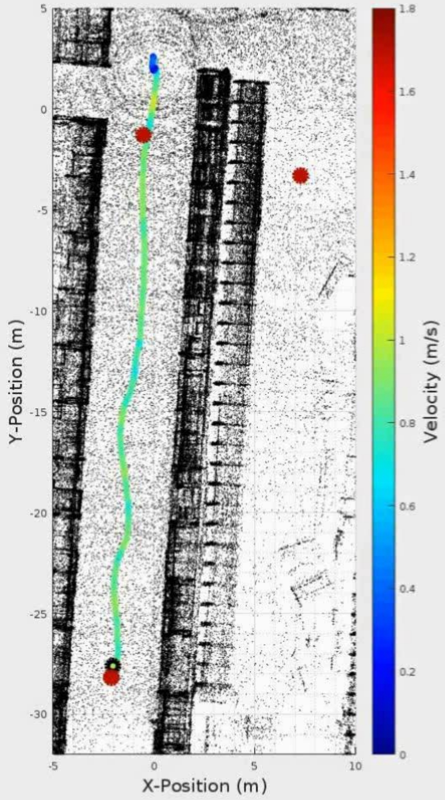
Autonomous Flight to Planned View Poses



DJI Matrice 600 with Velodyne Puck & Cameras

InventAIRy: Autonomous Navigation in a Warehouse

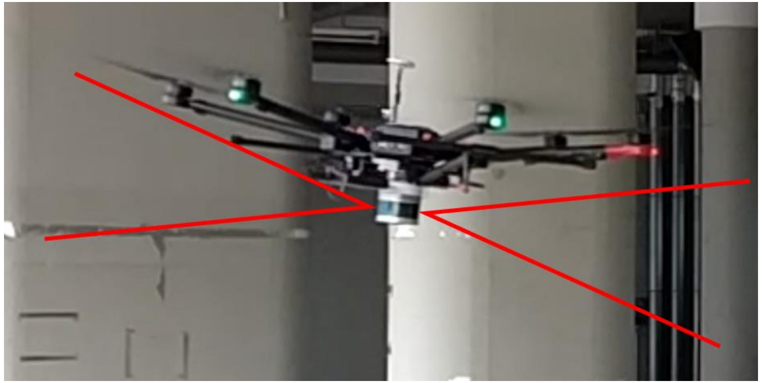




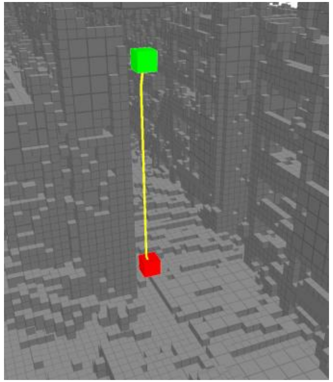
InventAIRy: Detected Tags in Shelf

Navigation Planning with Visibility Constraints

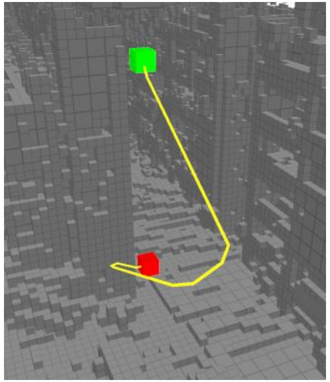
- Velodyne Puck has limited vertical field-of-view (30°)
- Must be considered in navigation planning
- Only fly in directions that can be measured



Lidar field-of-view



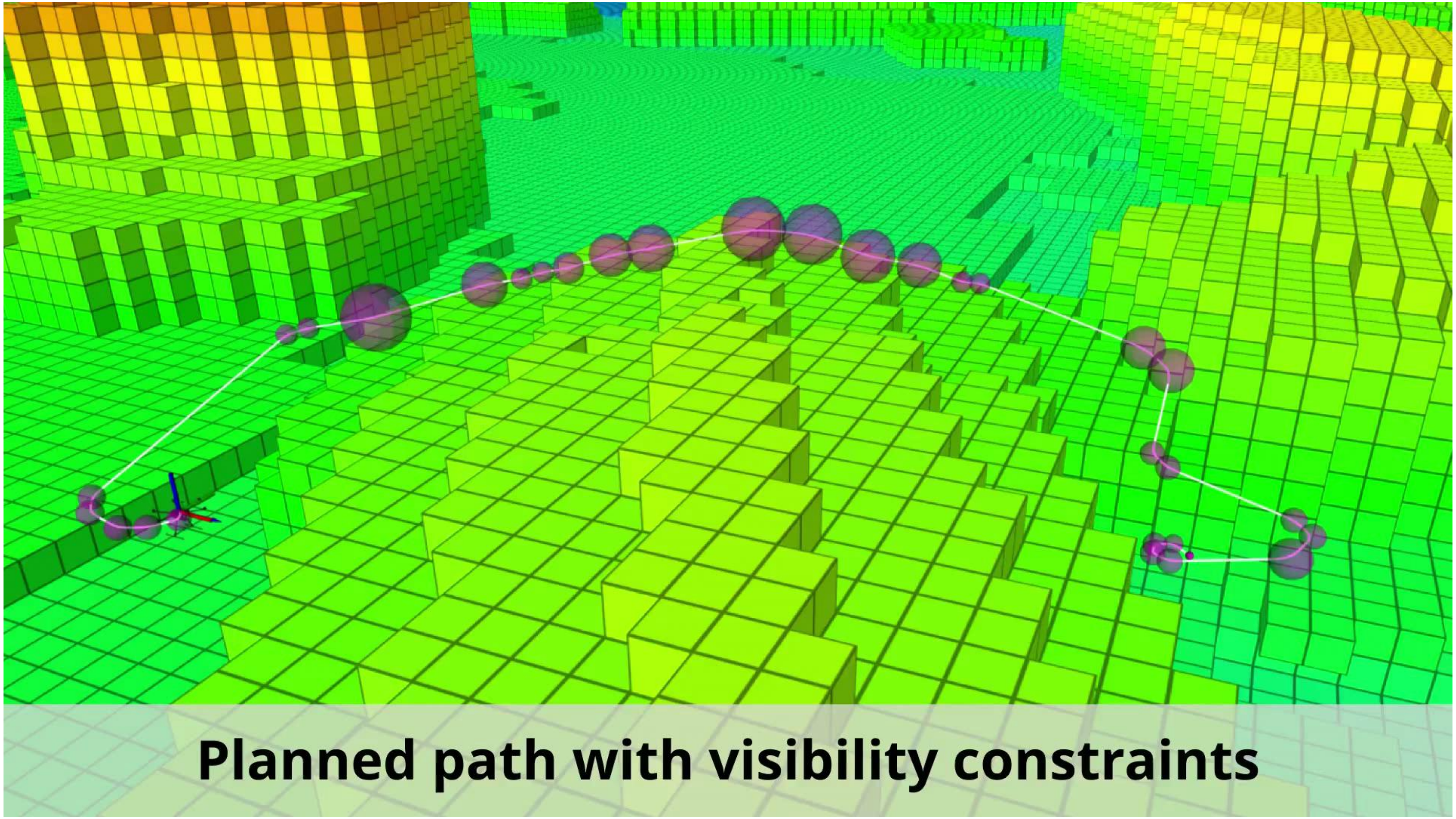
Fastest path



Safe path

[Nieuwenhuisen and Behnke, ICRA 2019]

Navigation Planning with Visibility Constraints

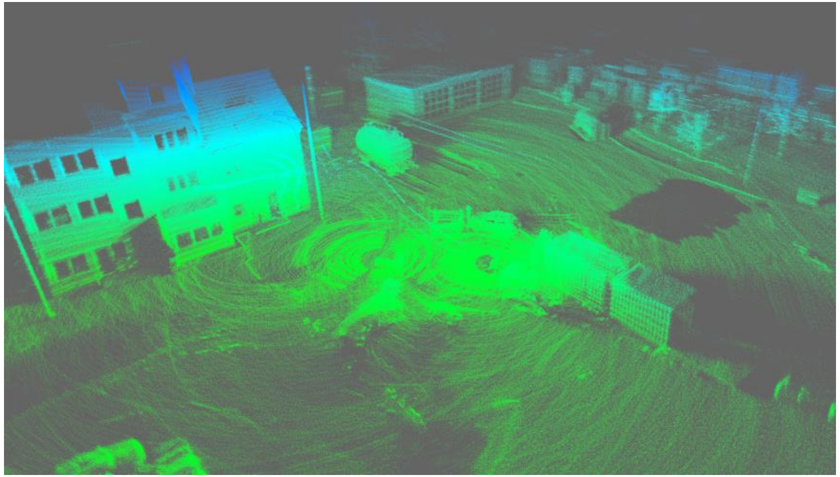


Lidar-based SLAM from MAV

[Droeschel & Behnke, ICRA 2018]

Supporting Fire Fighters (A-DRZ)

- Added thermal camera
- Flight at Brandhaus Dortmund



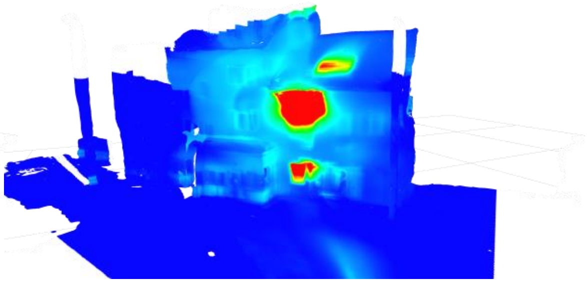
[Rosu et al. SSRR 2019]

Mesh-based 3D Modeling + Textures

- Model 3D geometry with mesh
- Appearance and temperature as high-resolution texture


Mesh geometry


RGB texture

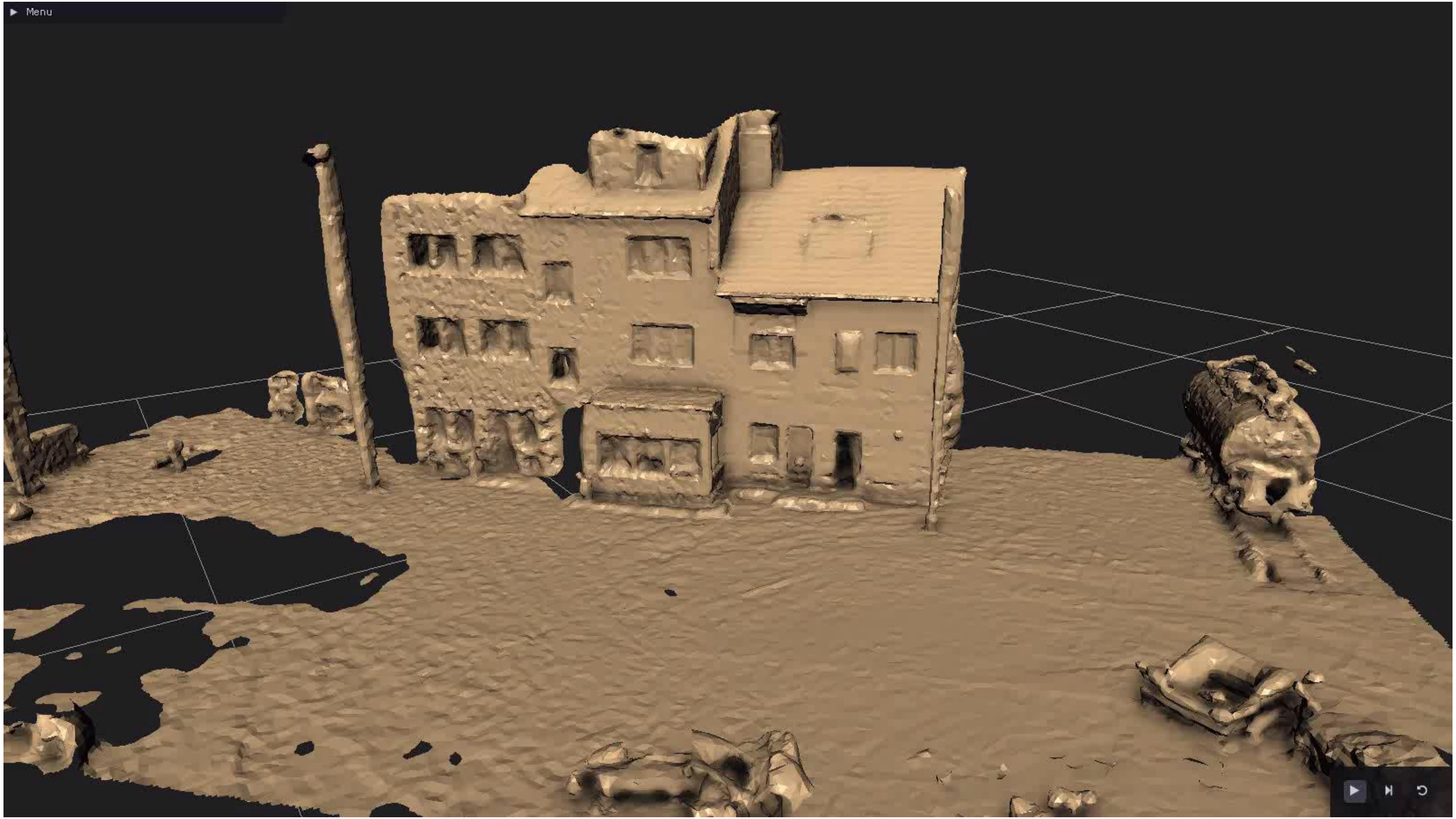

Thermal texture

- Mapping from 3D mesh to 2D texture


Texture mapping

[Rosu et al. SSRR 2019]

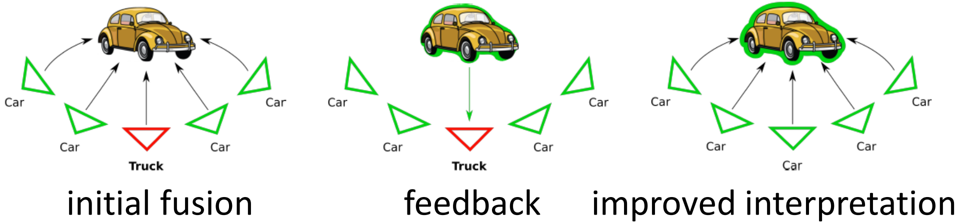
Modeling the Brandhaus Dortmund

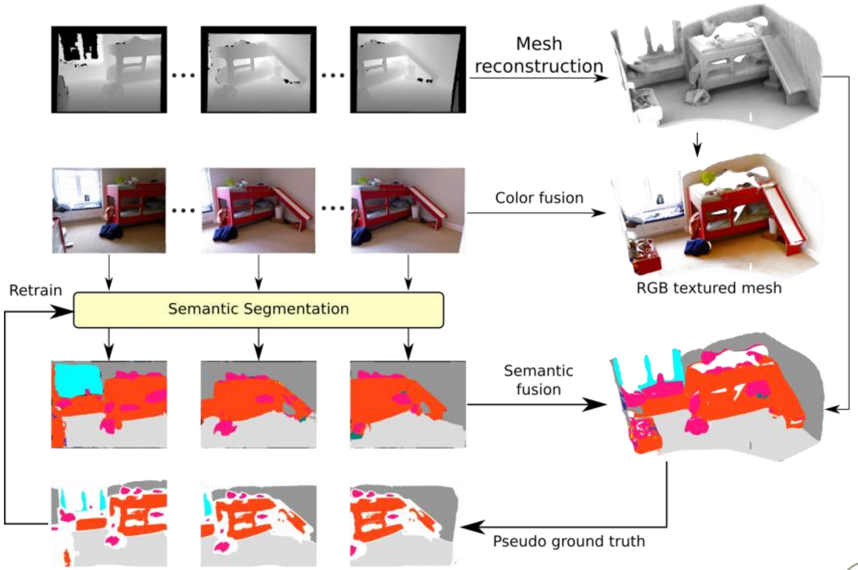


[Rosu et al. SSRR 2019]

3D Semantic Mapping

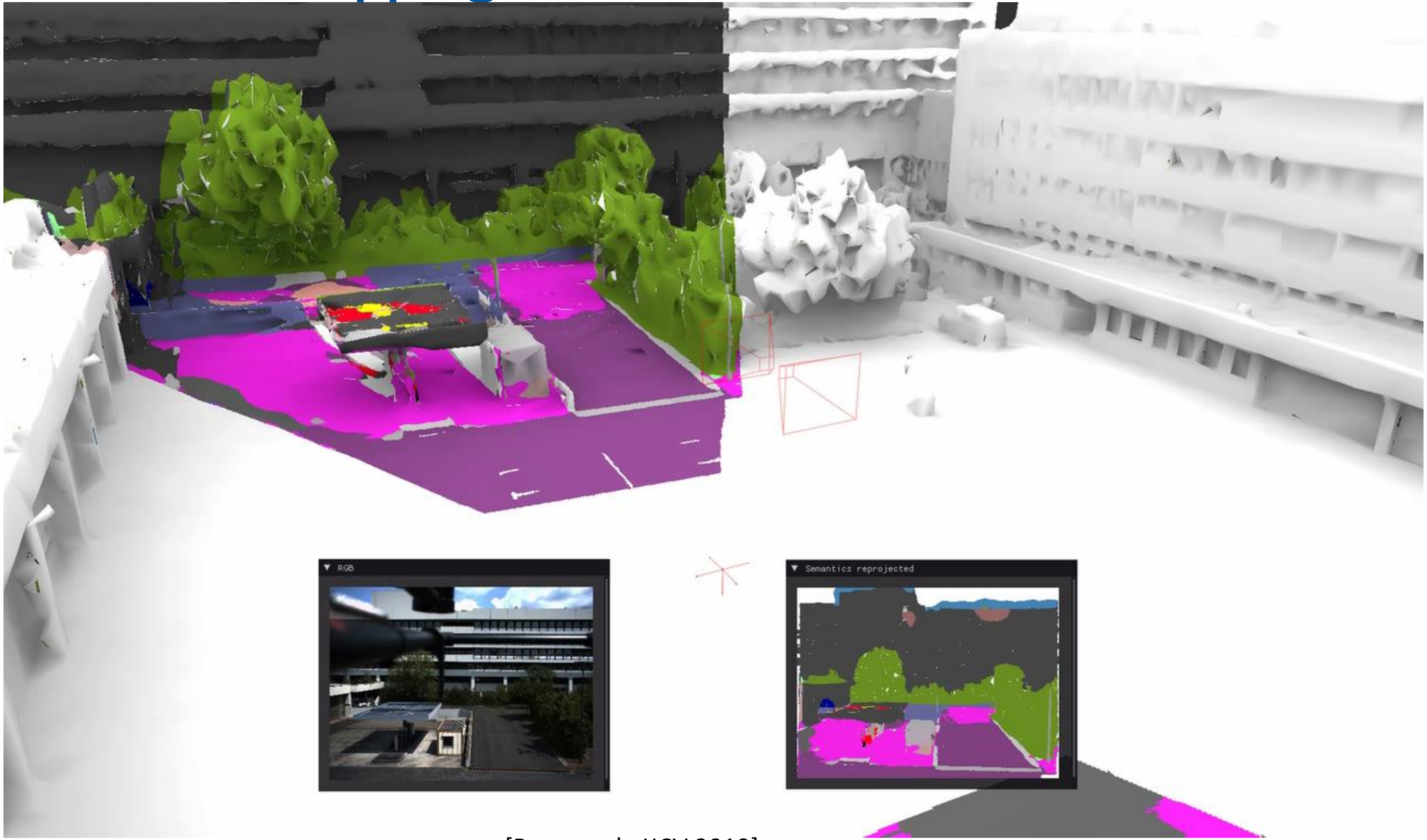
- Image-based semantic categorization, trained with Mapillary data set
- 3D fusion in semantic texture
- Backprojection of labels to other views





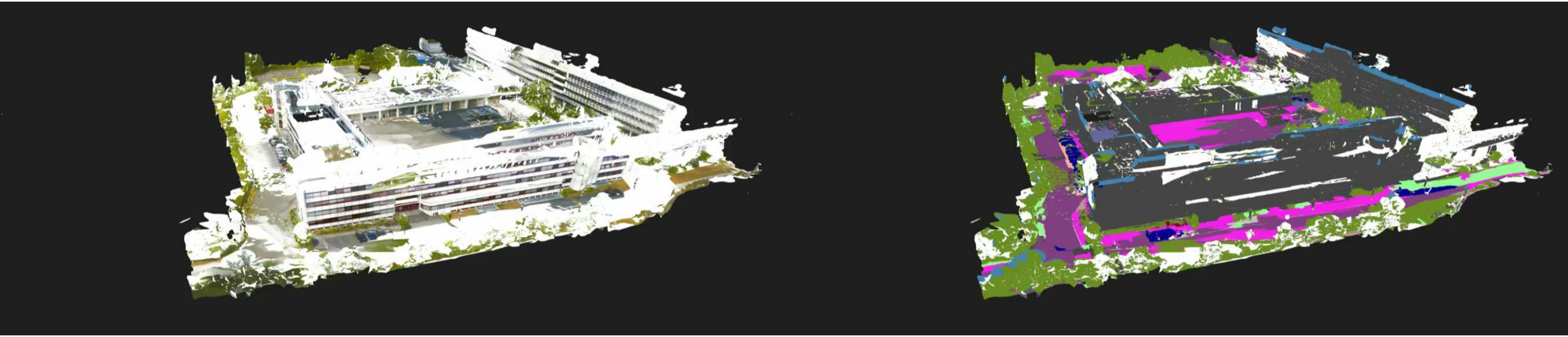
[Rosu et al., IJCV 2019]

3D Semantic Mapping



[Rosu et al., IJCV 2019]

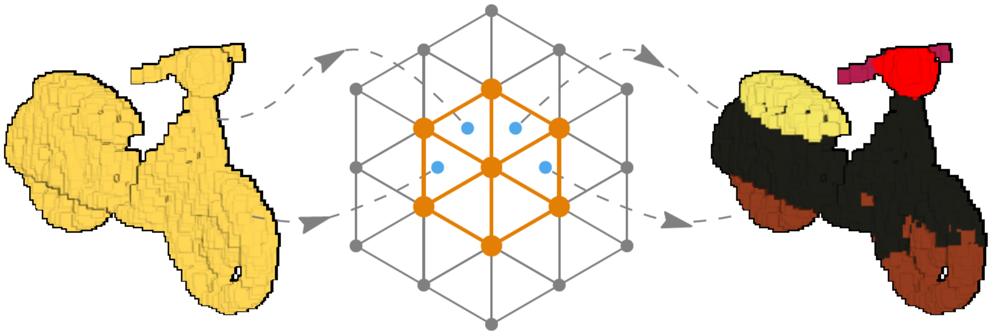
3D Semantic Map



[Rosu et al., under review]

Fast Point Cloud Segmentation Using Permutohedral Lattices

- Point cloud embedded into sparse permutohedral lattice
- Low memory footprint
- Fast 3D convolutions
- U-net semantic segmentation
- Good results on three data sets



ShapeNet

ScanNet

SemanticKITTI

[Rosu et al. 2020 (*submitted*)]

Conclusions

- Developed capable robotic systems for challenging scenarios
  - Humanoid soccer
  - Domestic service
  - Bin picking
  - Disaster response
  - Aerial inspection
- Challenges include
  - Capable and affordable robot platforms
  - 4D semantic perception
  - High-dimensional motion planning
- Promising approaches
  - Shared autonomy
  - Instrumented environments