Semantic Perception for Cognitive Robots

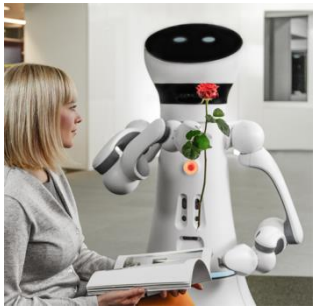
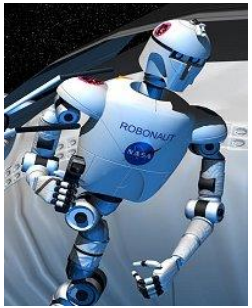
Sven Behnke
University of Bonn, Germany
Computer Science Institute VI
Autonomous Intelligent Systems

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!**

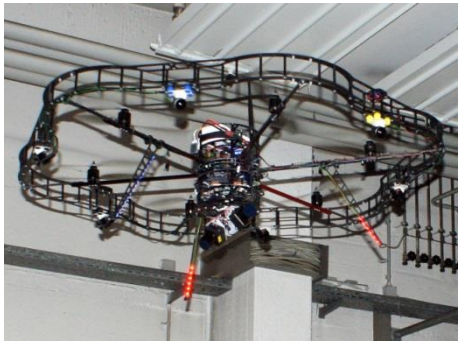





Some of Our Cognitive Robots

- Equipped with many sensors and DoFs
- Demonstration in complex scenarios



MAV

Soccer robot

Service robot

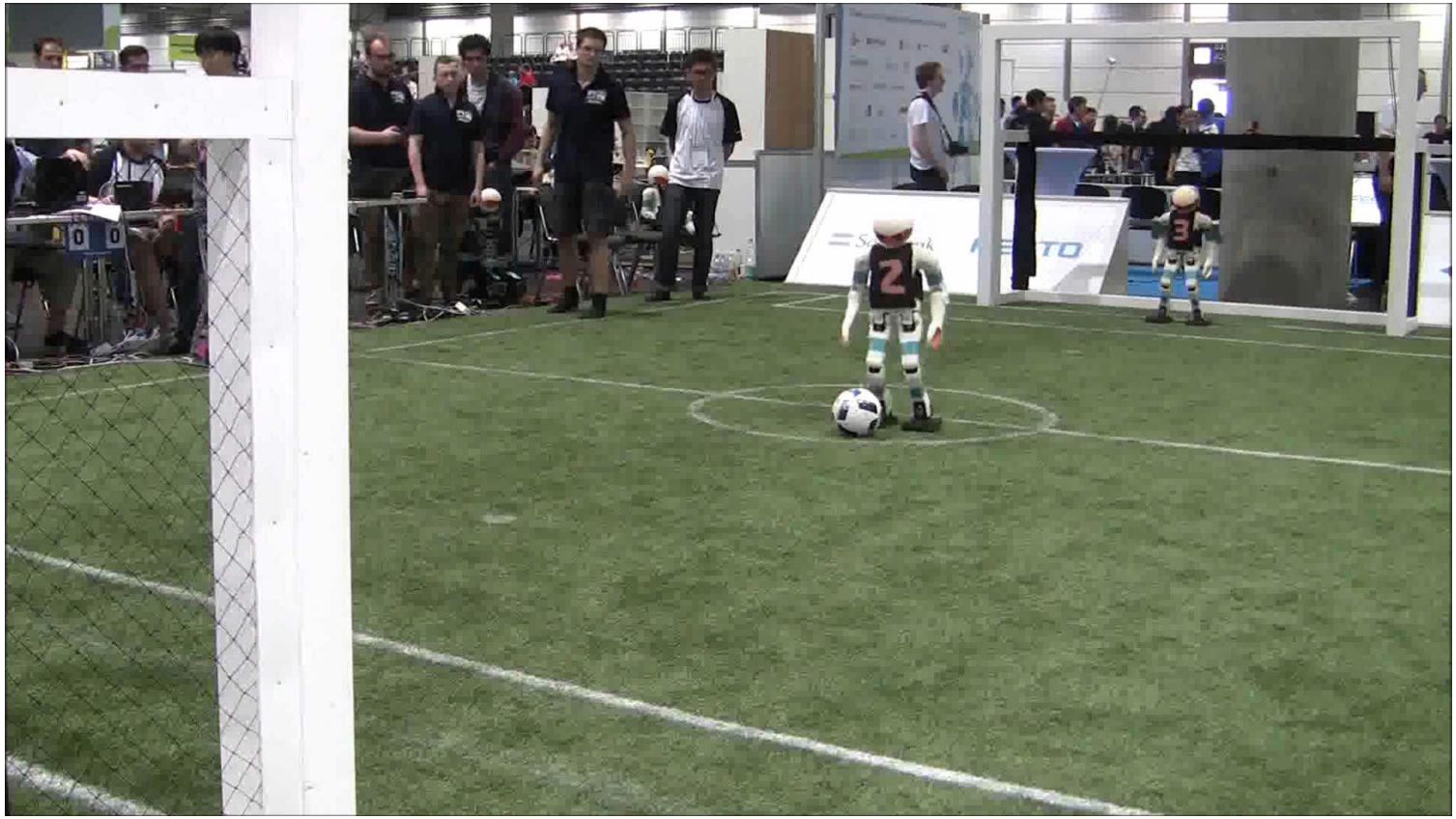
Exploration robot

Picking robot

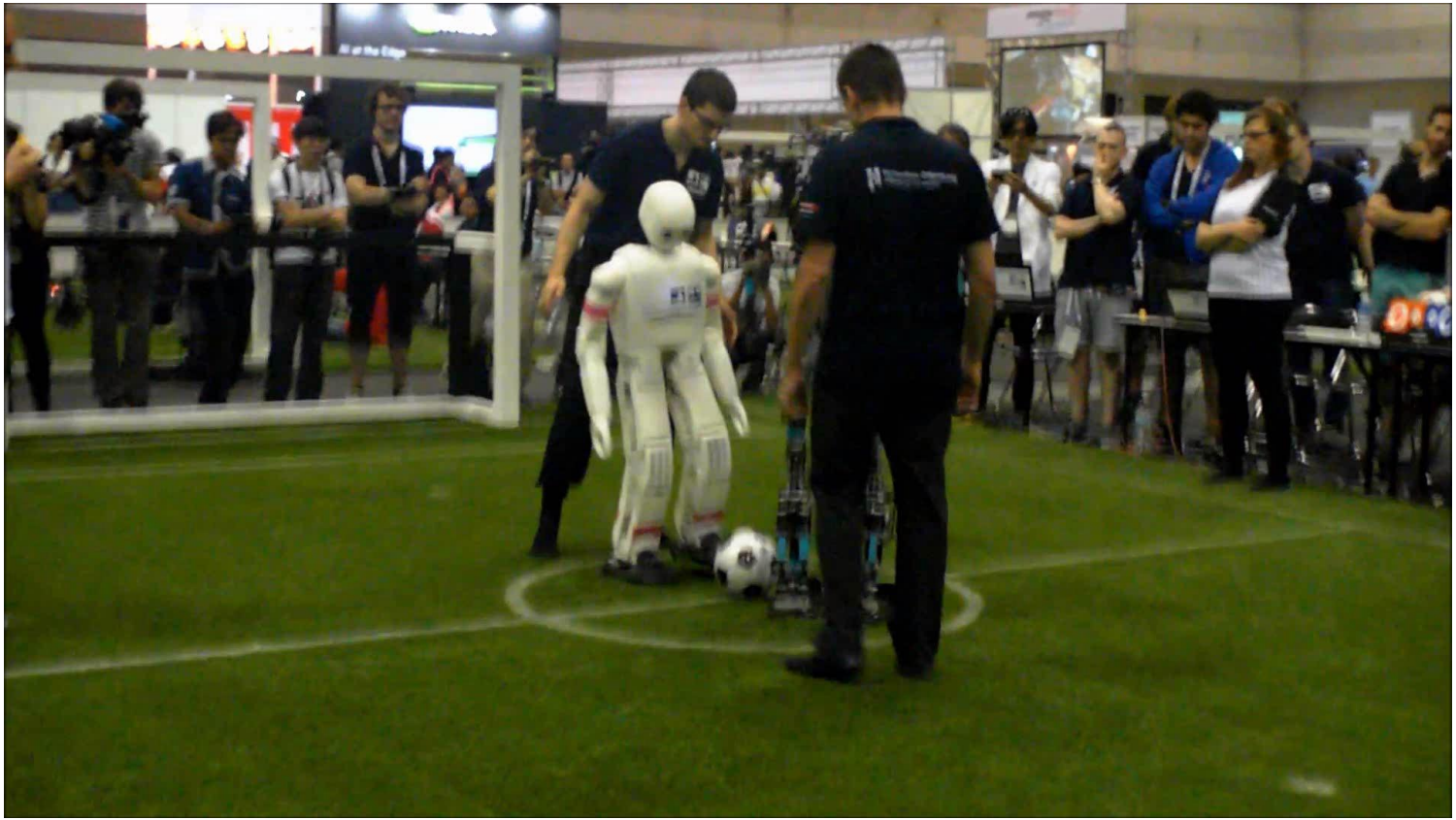
Visual Perception of Soccer Scene

[Farazi & Behnke, RoboCup 2016]

RoboCup 2016 TeenSize Final



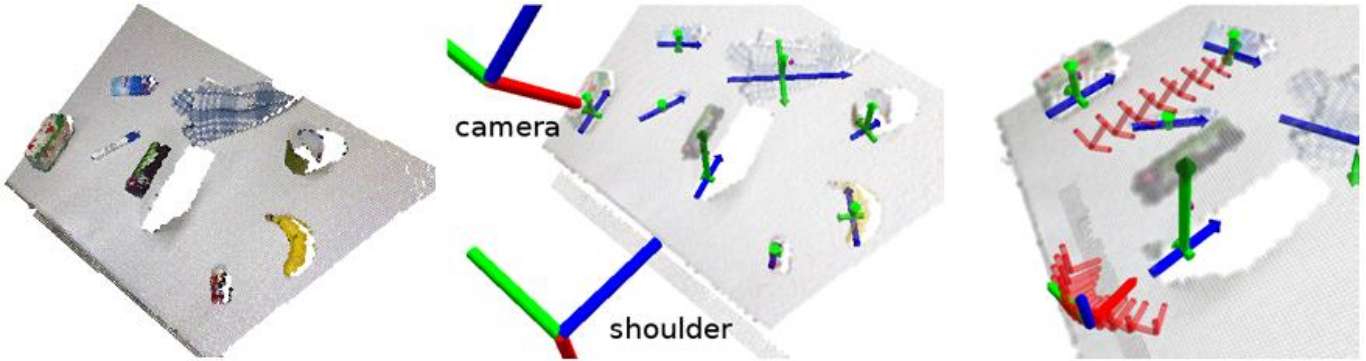
RoboCup 2017 AdultSize Final



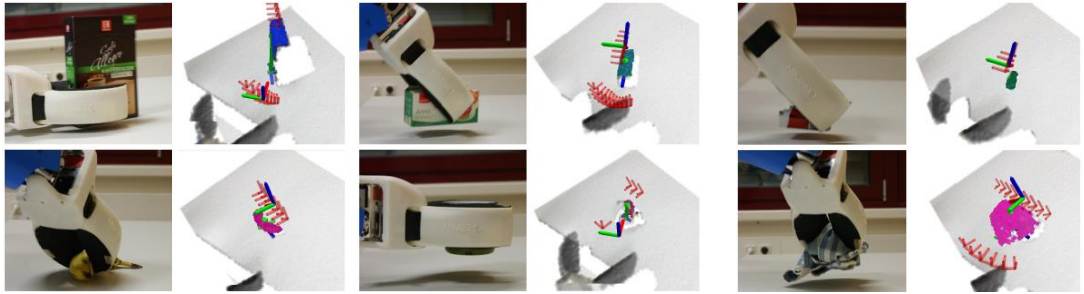
Cognitive Service Robot Cosero

Table-top Analysis and Grasp Planning

- Detection of clusters above horizontal plane
- Two grasps (top, side)



- Flexible grasping of many unknown objects

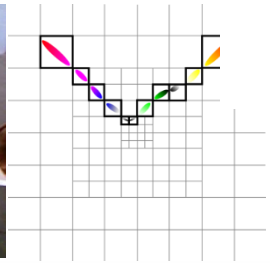


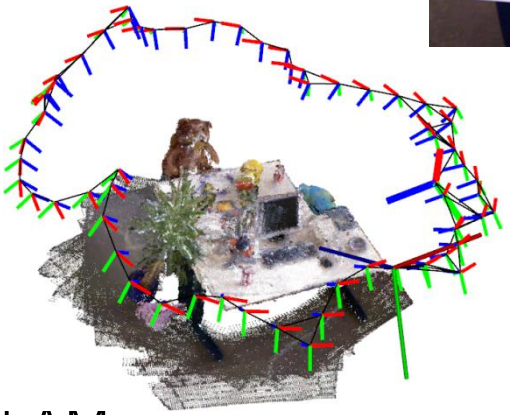
[Stückler et al, Robotics and Autonomous Systems, 2013]

3D Mapping by RGB-D SLAM

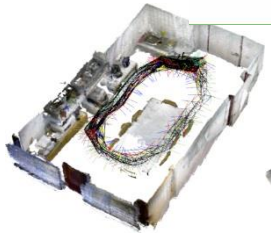
[Stückler, Behnke:
Journal of Visual Communication
and Image Representation 2013]

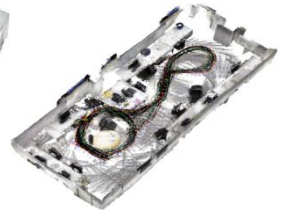
- Modelling of shape and color distributions in voxels
- Local multiresolution
- Efficient registration of views on CPU
- Global optimization





- Multi-camera SLAM



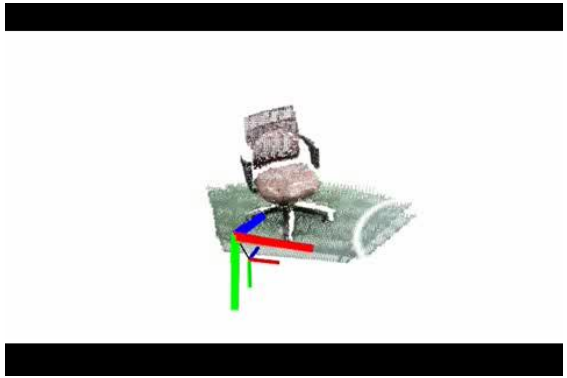


[Stoucken]

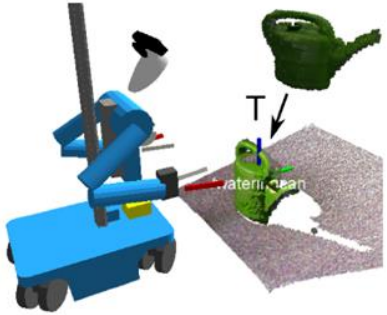
universitätbonn ais

Learning and Tracking Object Models

- Modeling of objects by RGB-D-SLAM

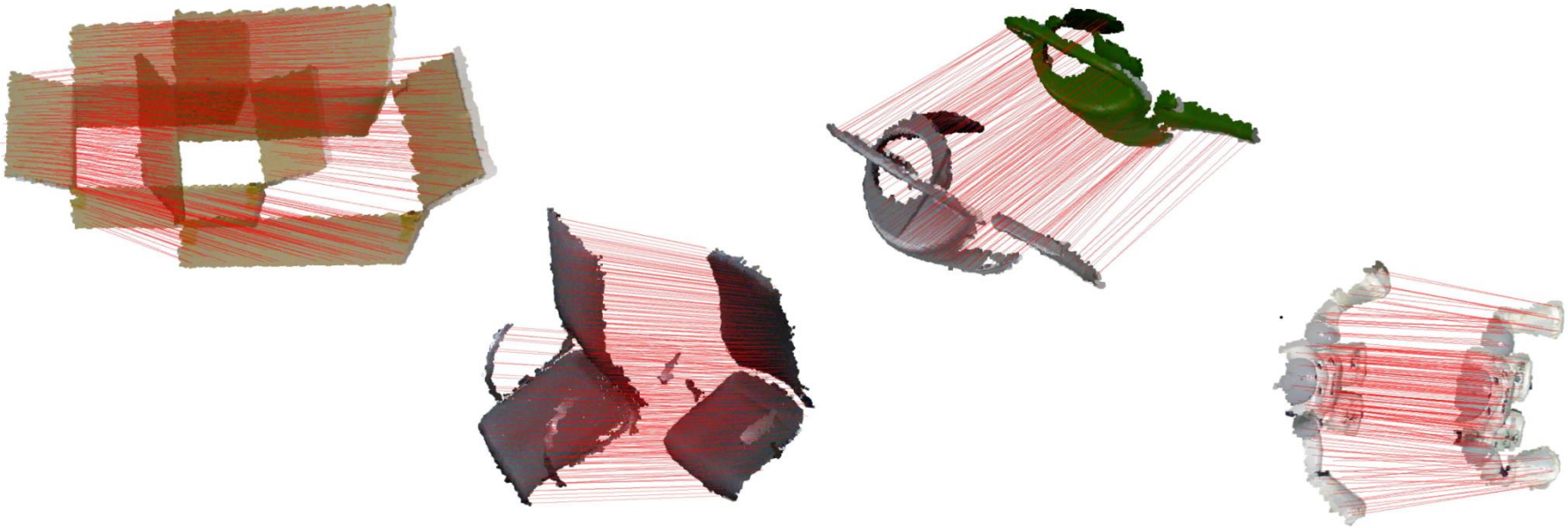


- Real-time registration with current RGB-D frame



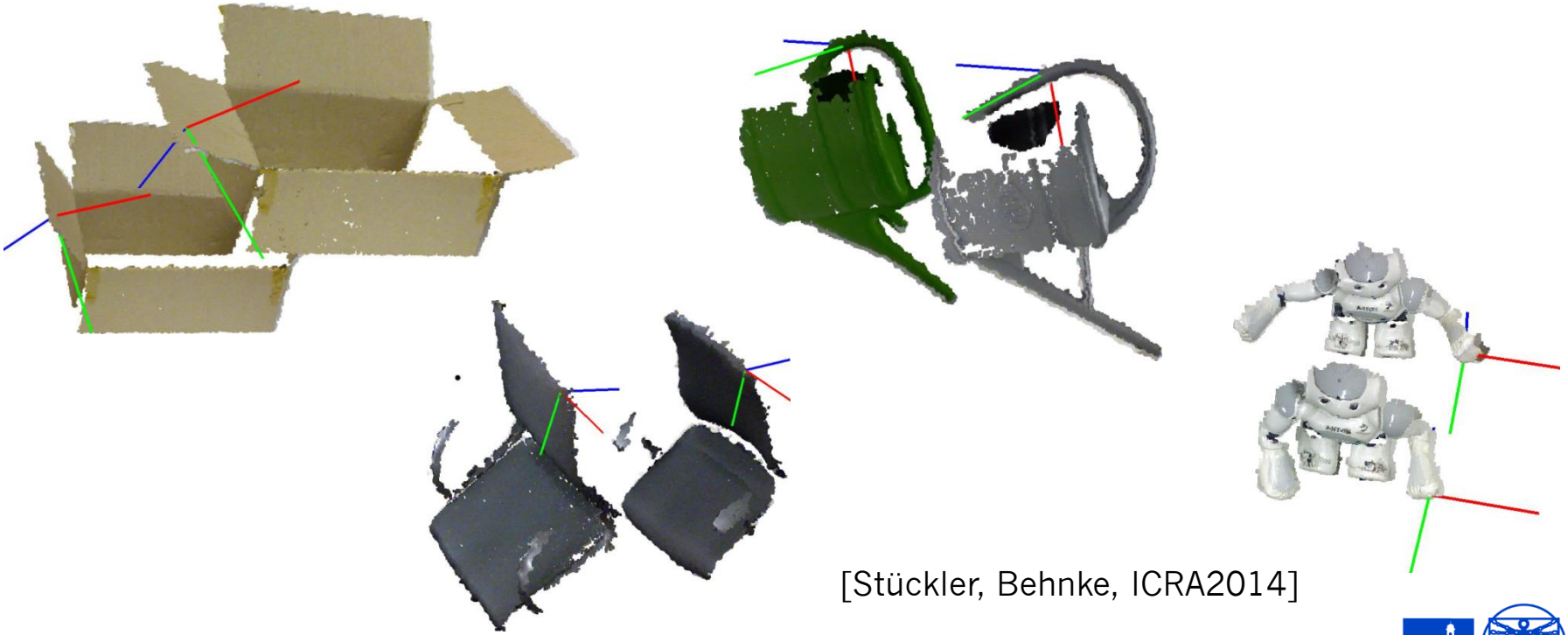
Deformable RGB-D-Registration

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



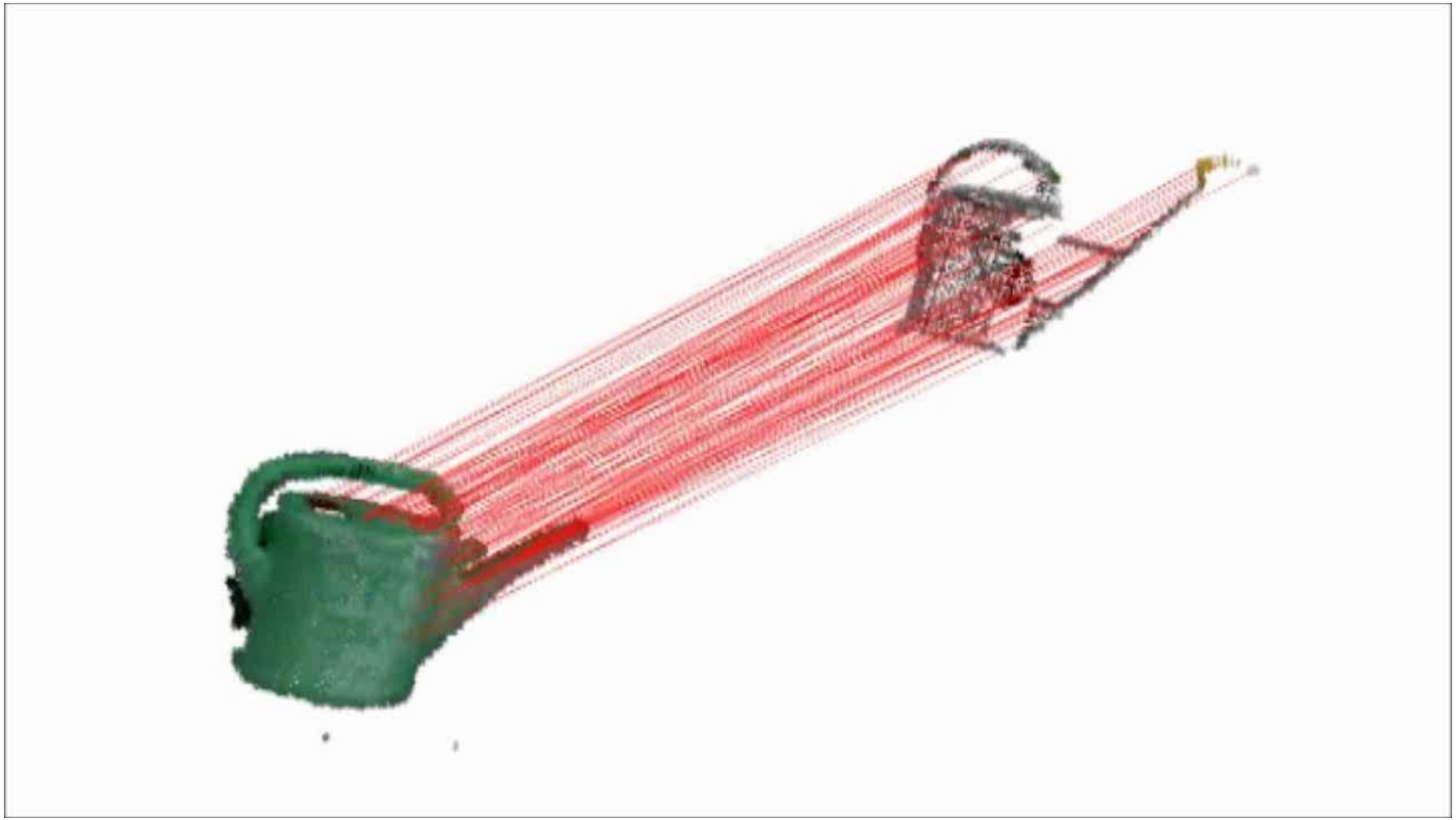
Transformation of Poses on Object

- Derived from the deformation field



[Stückler, Behnke, ICRA2014]

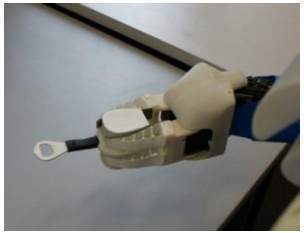
Grasp & Motion Skill Transfer

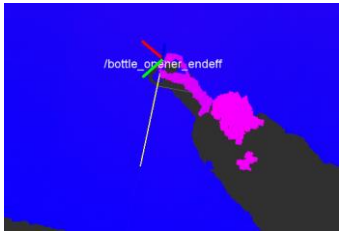


[Stückler, Behnke, ICRA2014]

Tool use: Bottle Opener

- Tool tip perception



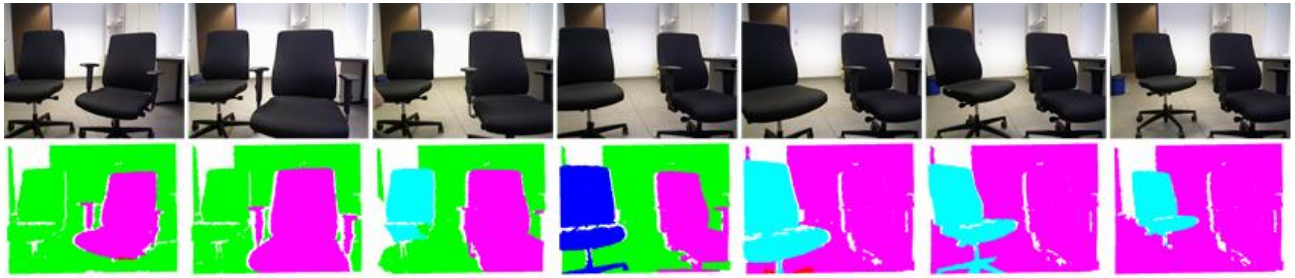


- Extension of arm kinematics
- Perception of crown cap
- Motion adaptation

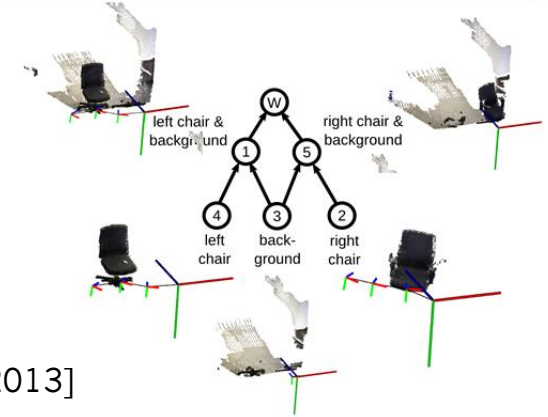
[Stückler, Behnke, Humanoids 2014]

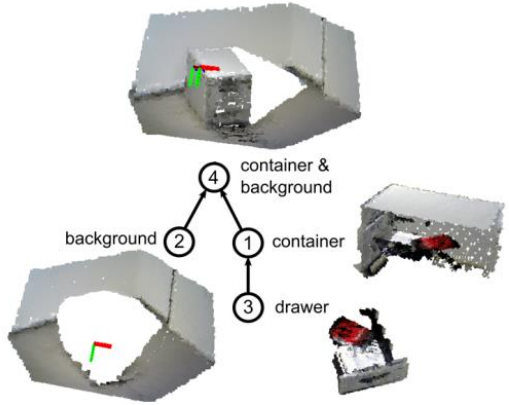
Hierarchical Object Discovery trough Motion Segmentation

- Simultaneous object modeling and motion segmentation



- Inference of a segment hierarchy

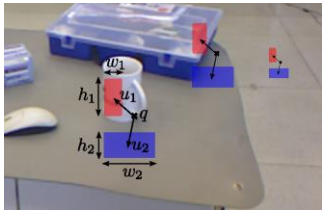


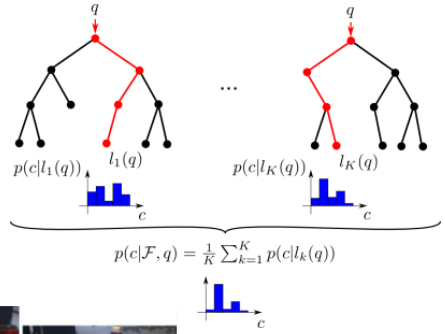


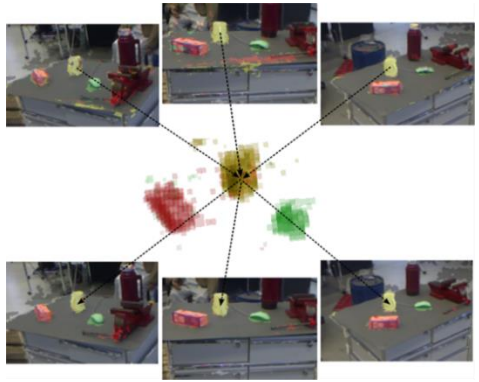
[Stückler, Behnke: IJCAI 2013]

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





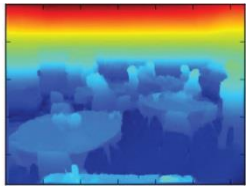


[Stückler, Biresev, Behnke: IROS 2012]

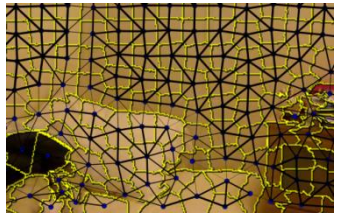
Accuracy in %	Ø Classes	Ø Pixels
Silberman et al. 2012	59,6	58,6
Couprie et al. 2013	63,5	64,5
Random forest	65,0	68,1
3D-Fusion	**66,8**	

Learning Depth-sensitive CRFs

- SLIC+depth super pixels
- Unary features: random forest
- Height feature

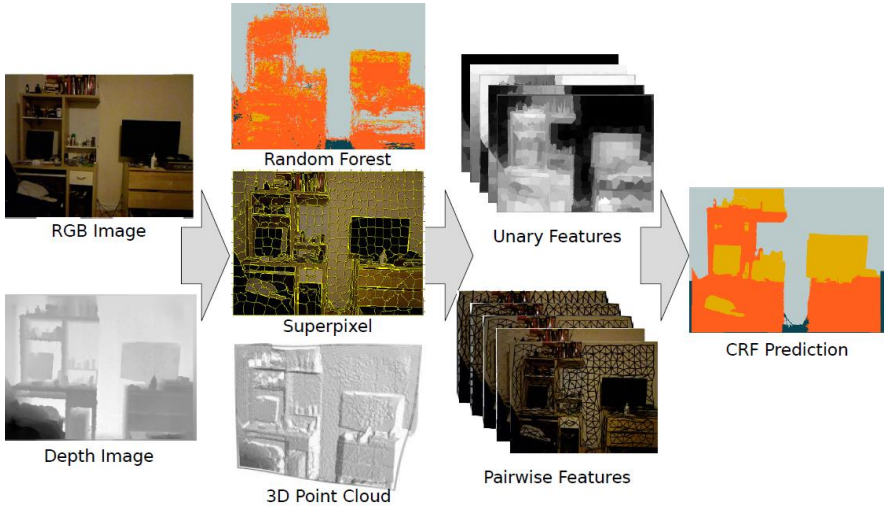


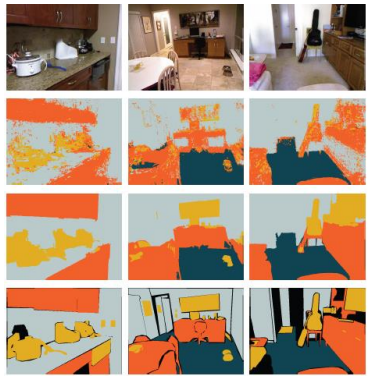
- Pairwise features
  - Color contrast
  - Vertical alignment
  - Depth difference
  - Normal differences



- Results:

	class average	pixel average
RF	65.0	68.3
RF + SP	65.7	70.1
RF + SP + SVM	70.4	70.3
RF + SP + CRF	**71.9**	**72.3**
Silberman *et al.*	59.6	58.6
Couprie *et al.*	63.5	64.5





Random forest

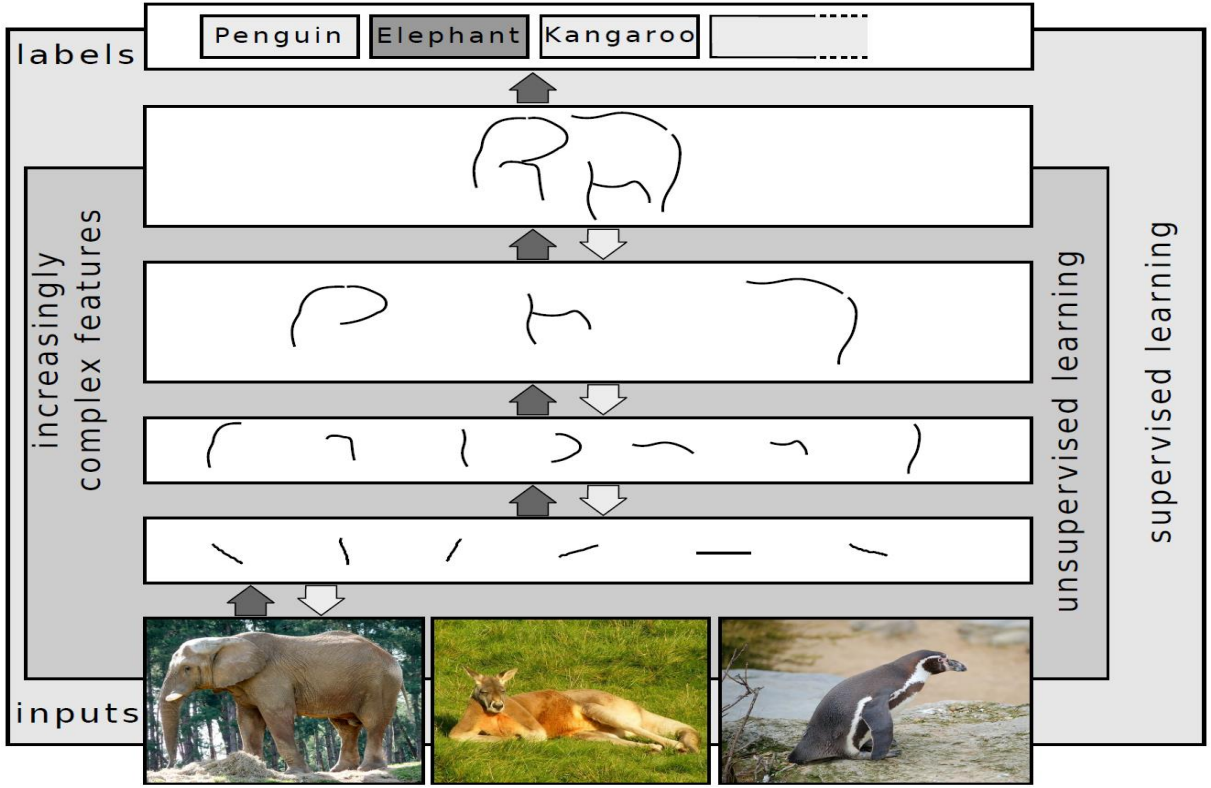
CRF prediction

Ground truth

universitätbonn ais

Deep Learning

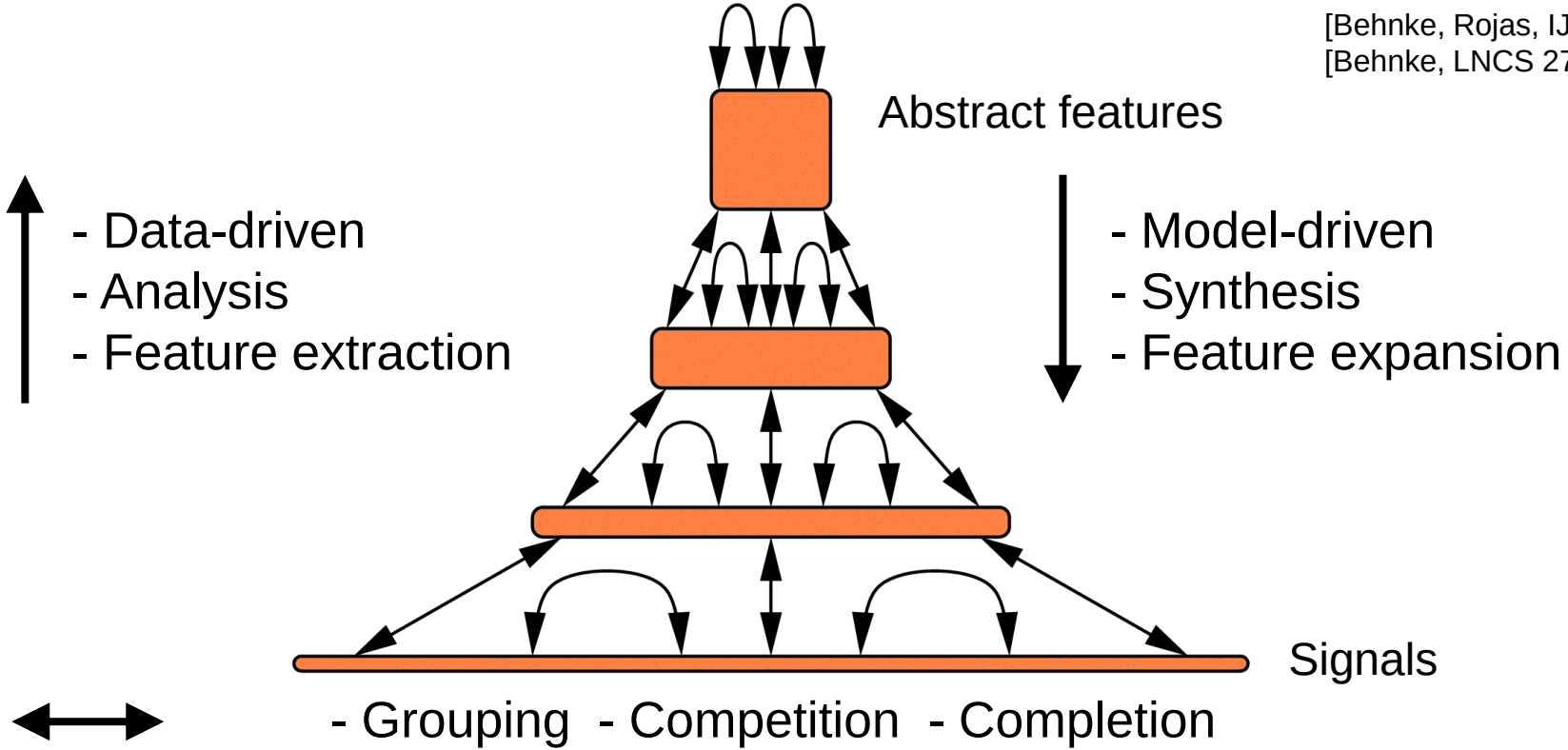
- Learning layered representations



[Schulz; Behnke, KI 2012]

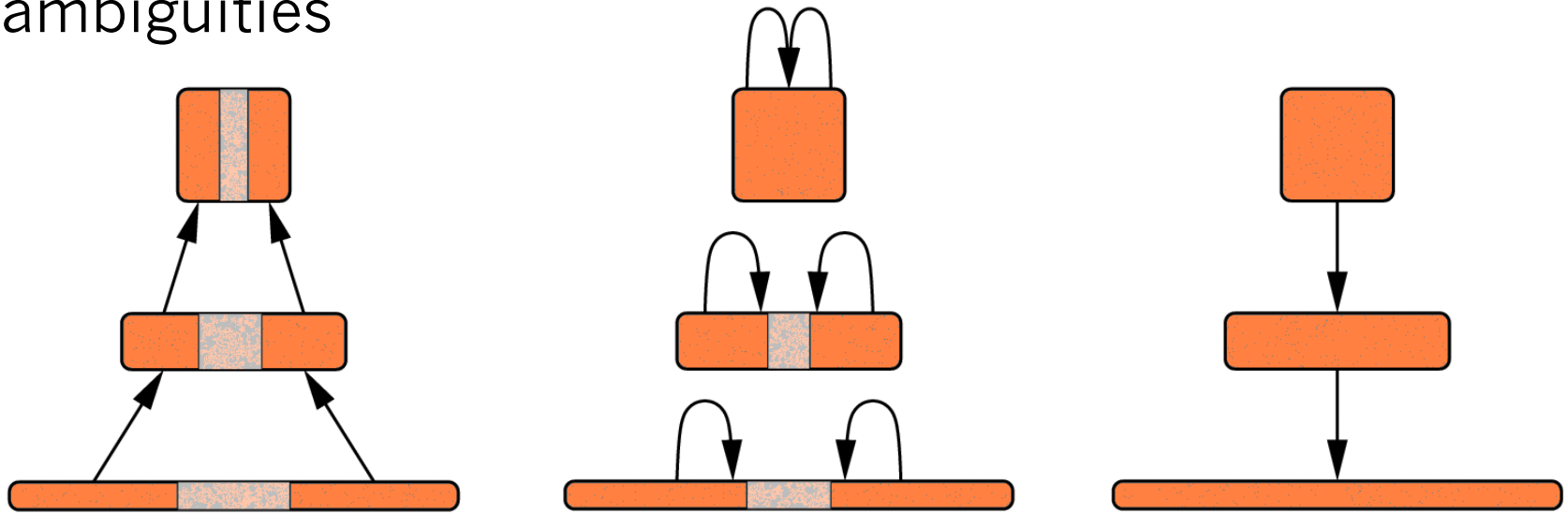
Neural Abstraction Pyramid

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



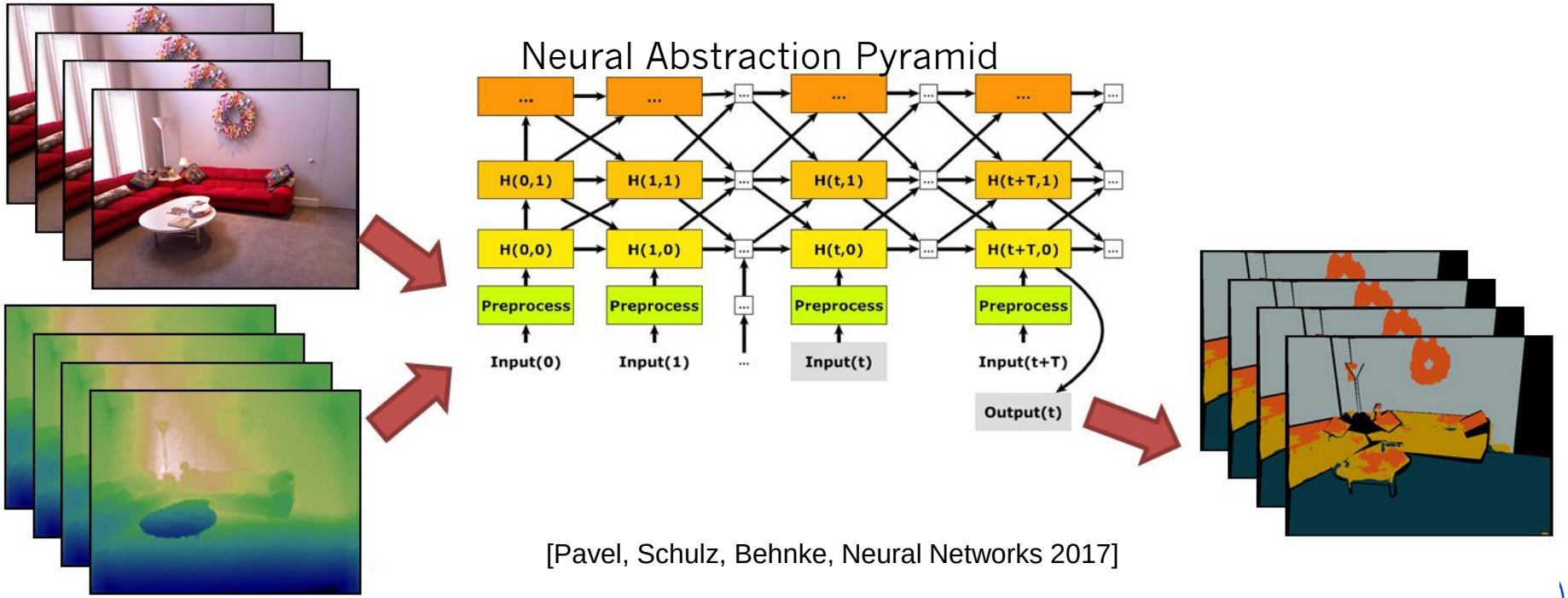
Iterative Image Interpretation

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



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

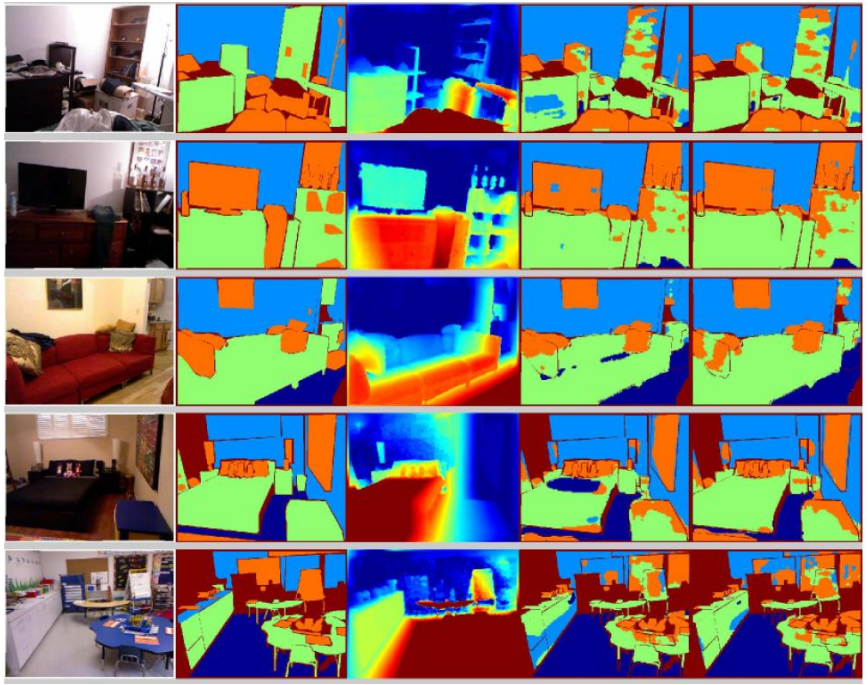
- Recursive computation is efficient for temporal integration



[Pavel, Schulz, Behnke, Neural Networks 2017]

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

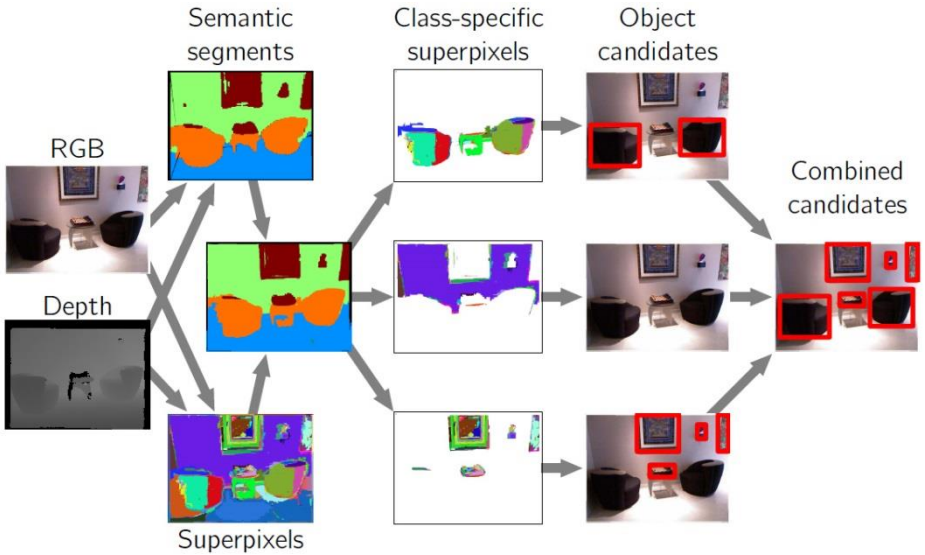
- New **geometric** feature: distance from wall
- **Semantic** features pretrained from ImageNet
- Both help significantly



[Husain et al. RA-L 2017]

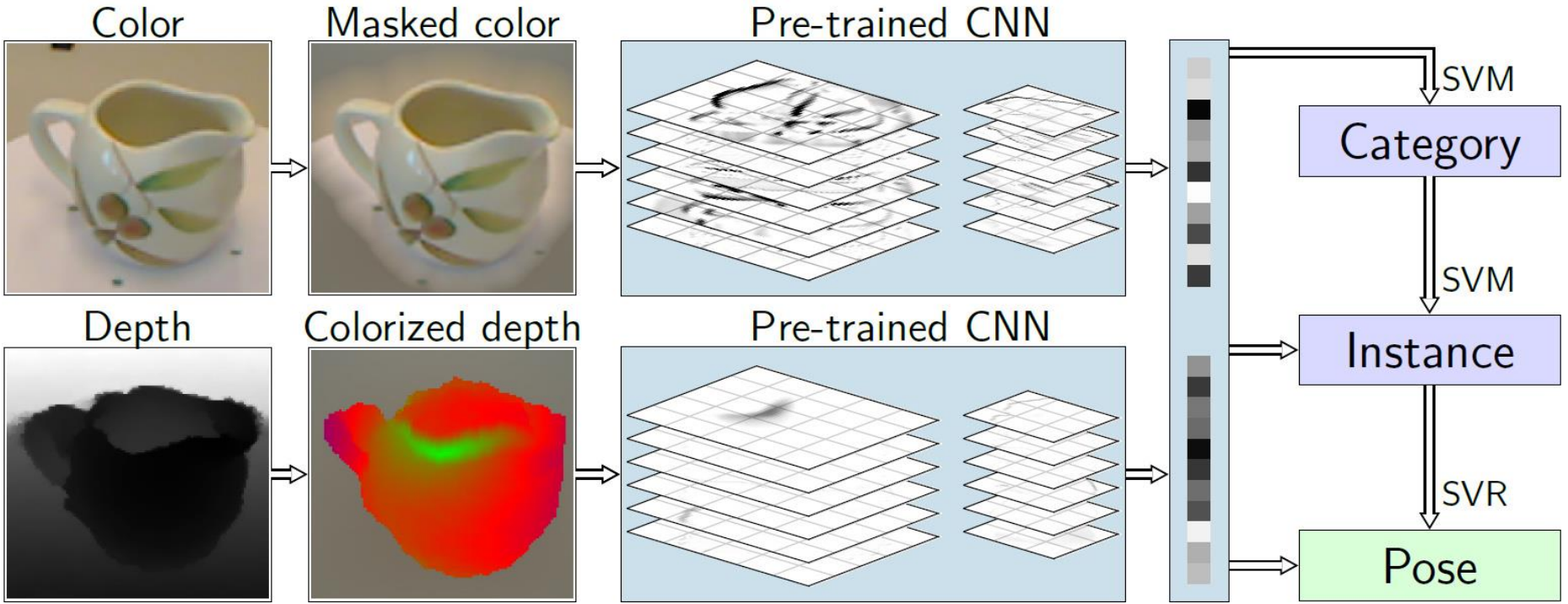
Semantic Segmentation Priors for Object Discovery

- Combine bottom-up object discovery and semantic priors
- Semantic segmentation used to classify color and depth superpixels
- Higher recall, more precise object borders



[Garcia et al. ICPR 2016]

RGB-D Object Recognition and Pose Estimation

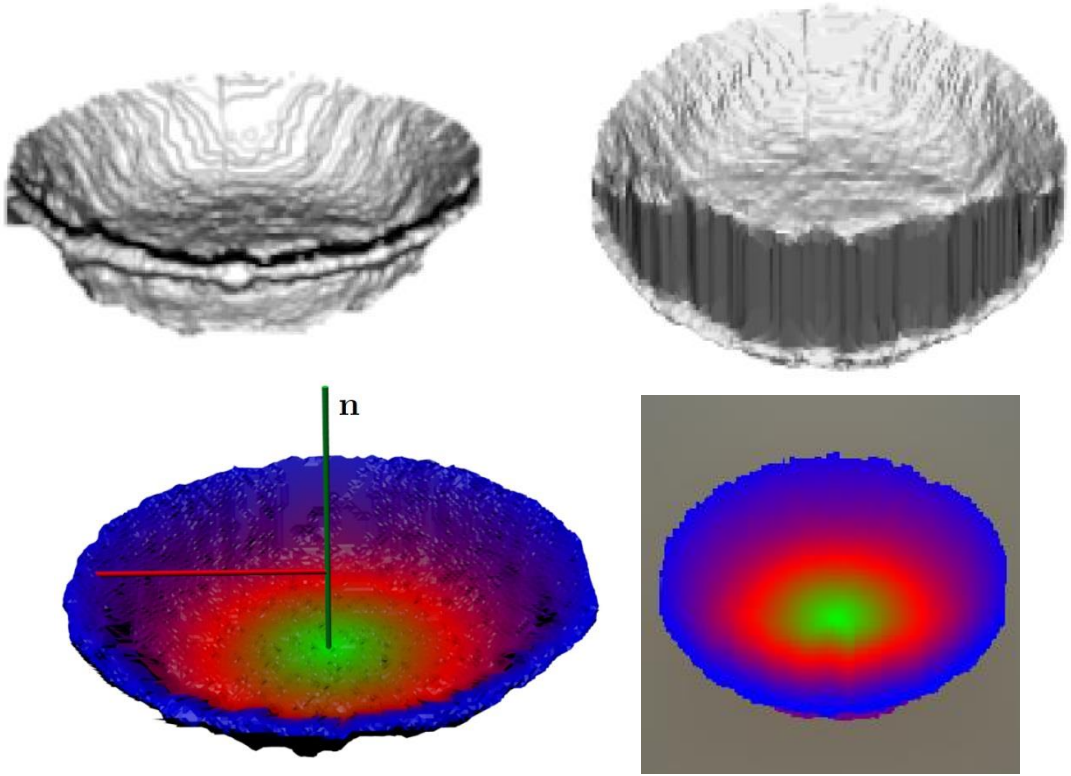


[Schwarz, Schulz, Behnke, ICRA2015]

Canonical View, Colorization

- Objects viewed from different elevation
- Render canonical view

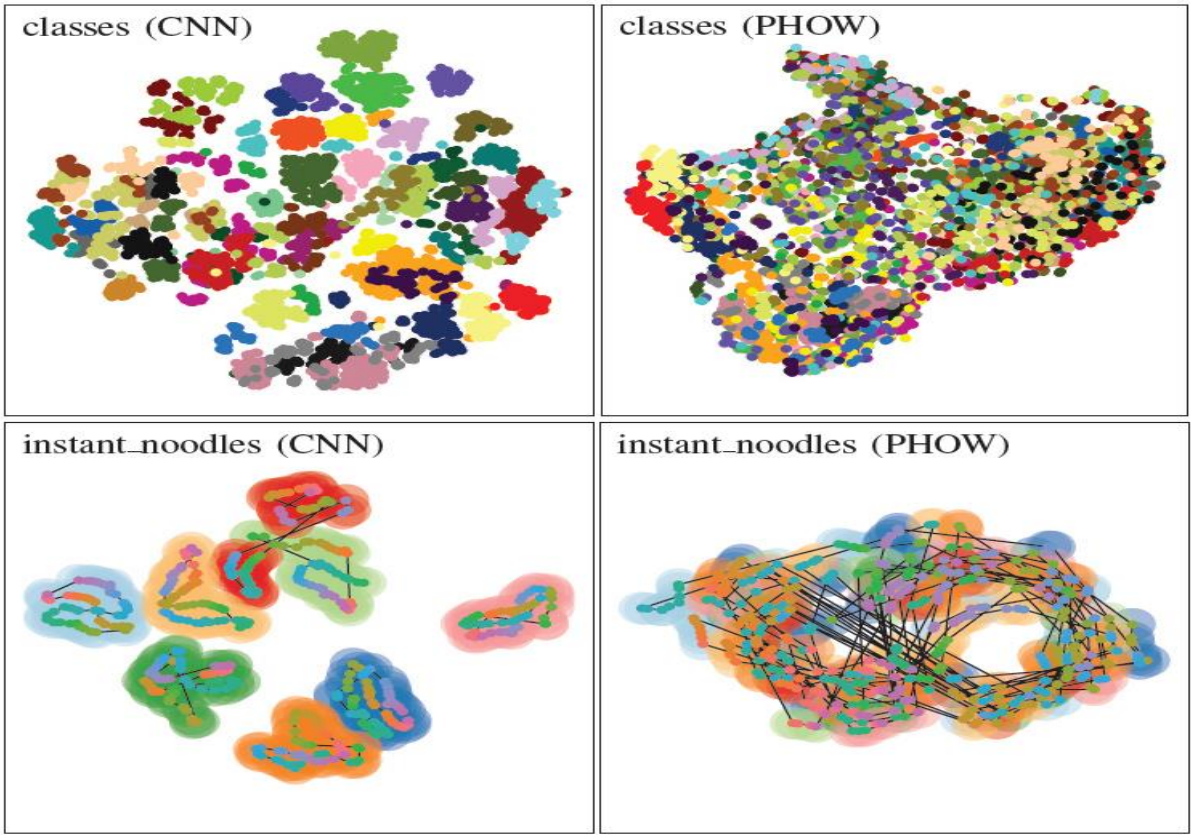
- Colorization based on distance from center vertical



[Schwarz, Schulz, Behnke, ICRA2015]

Pretrained Features Disentangle Data

- t-SNE embedding



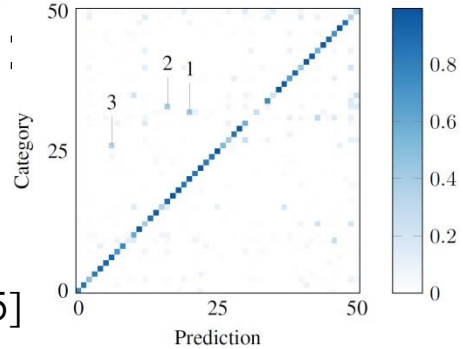
[Schwarz, Schulz, Behnke ICRA2015]

Recognition Accuracy

- Improved both category and instance recognition

Method	Category Accuracy (%)		Instance Accuracy (%)	
	RGB	RGB-D	RGB	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:



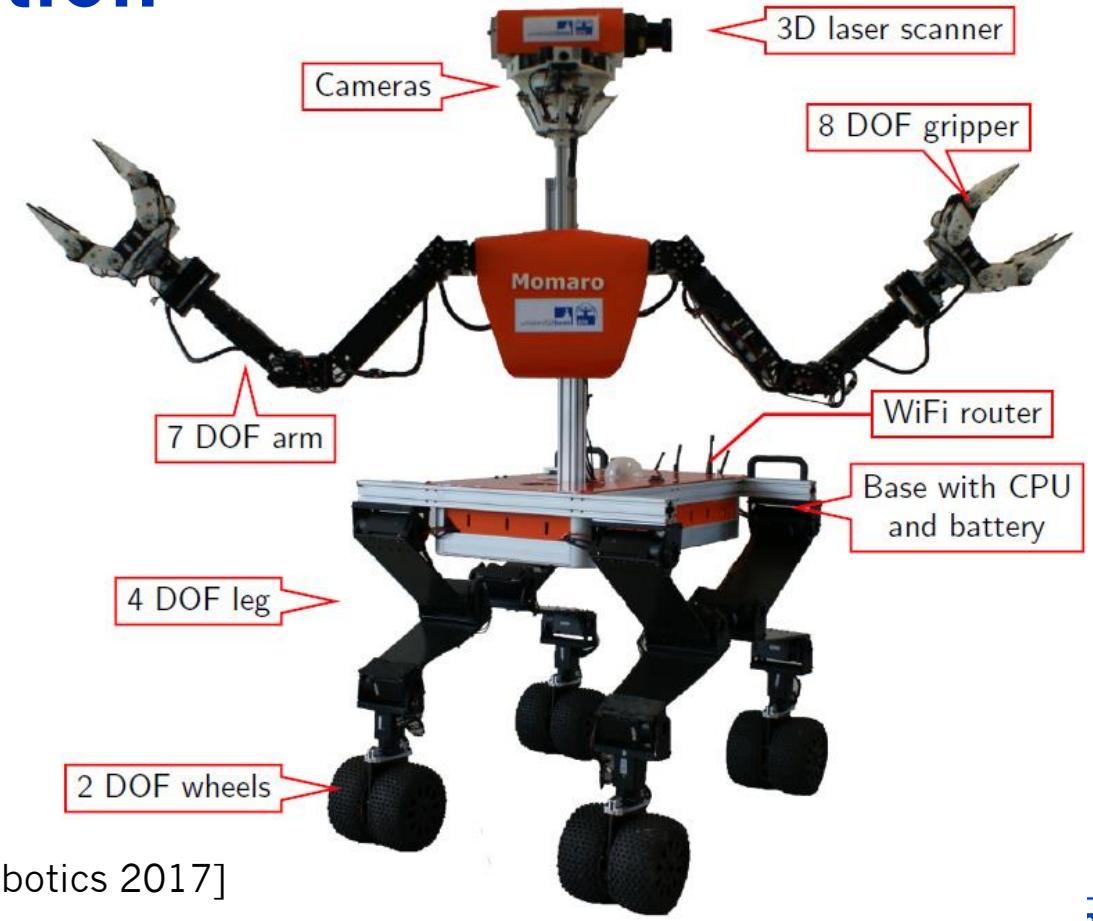
[Schwarz, Schulz, Behnke ICRA2015]

1: pitcher / coffe mug

2: peach / sponge

Mobile Manipulation Robot Momaro

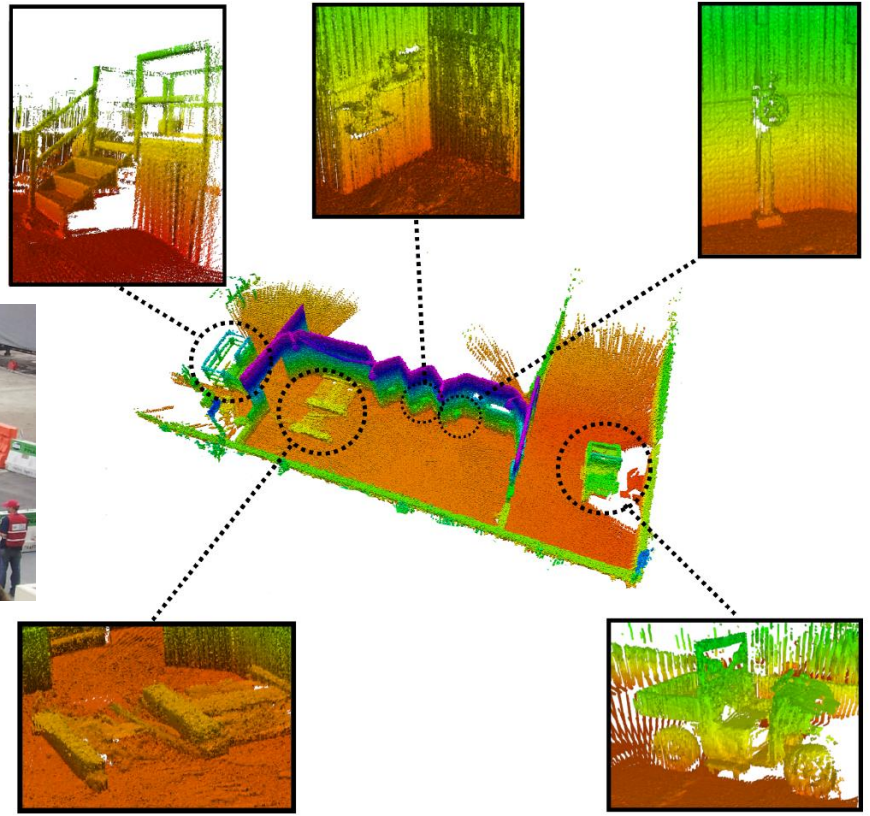
- Four compliant legs ending in pairs of steerable wheels
- Anthropomorphic upper body
- Sensor head
  - 3D laser scanner
  - IMU, cameras



[Schwarz et al. Journal of Field Robotics 2017]

Laser-based 3D Mapping

- Registration of egocentric maps by graph optimization



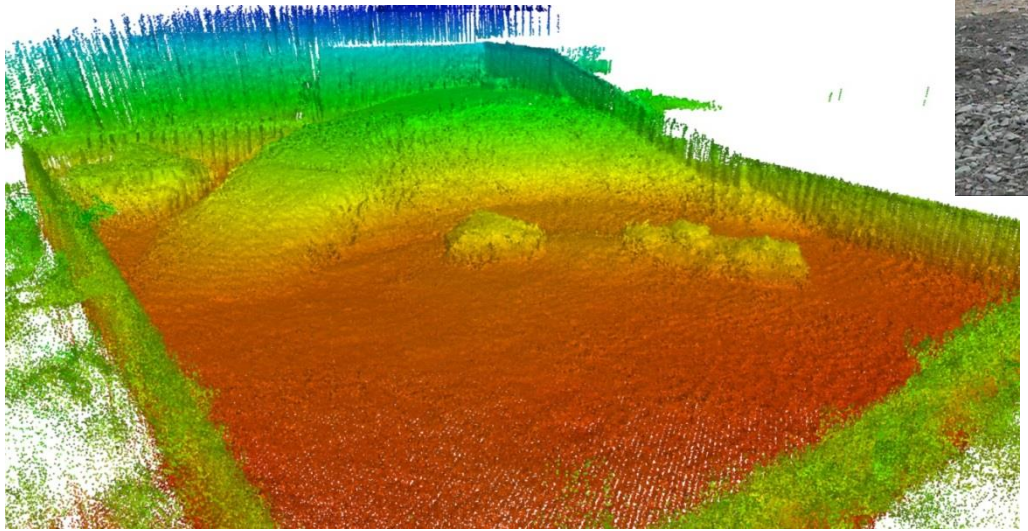
[Droeschel et al., Robotics and Autonomous Systems 2017]

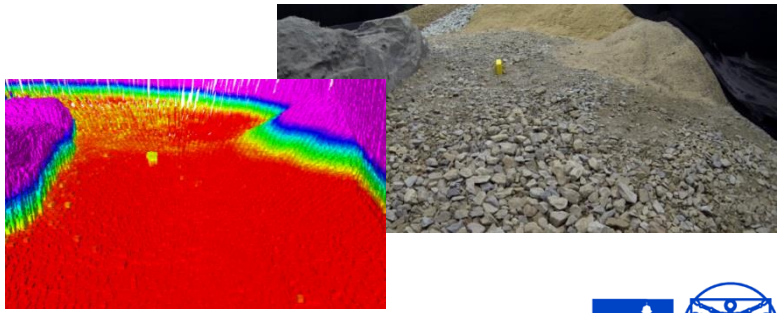

DARPA
CHALLENGE
2015
23:36:46 05/06/2015 UTC
4x

DLR SpaceBot Cup 2015

- Mobile manipulation in rough terrain

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

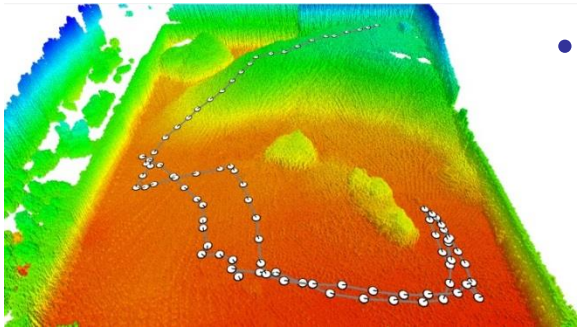




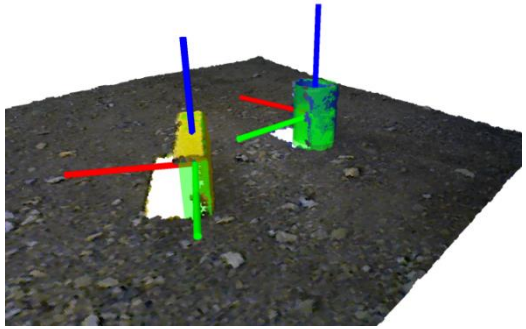

8X

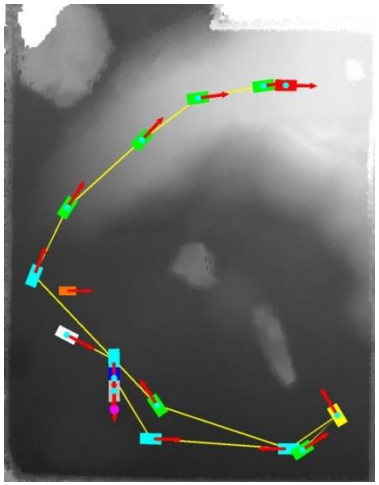
Autonomous Mission Execution

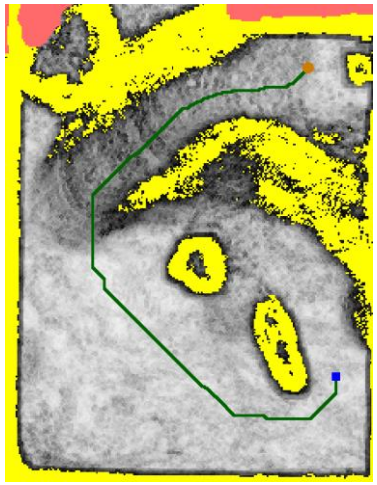
- 3D mapping, localization, mission and navigation planning

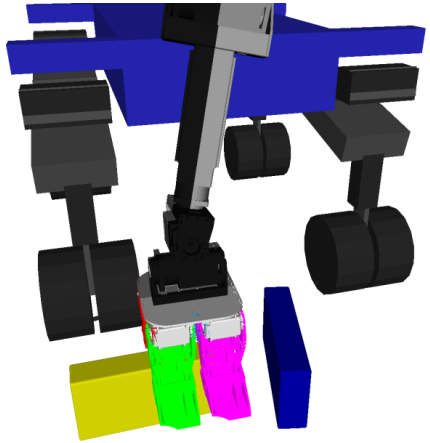


- 3D object perception and grasping







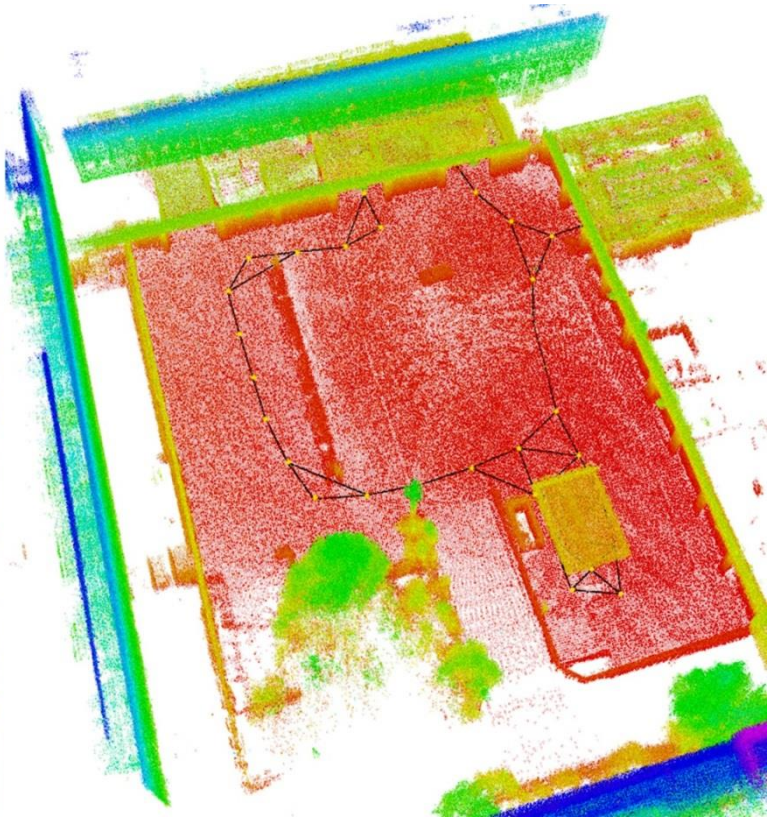


[Schwarz et al. Frontiers 2016]

New Sensor Head

- Continuously rotating Velodyne Puck VLP-16
  - 300,000 3D points/s
  - 100 m range
  - Spherical field of view
- Three wide-angle color cameras (total FoV 210×103°)
- Kinect V2 RGB-D camera on pan-tilt unit

3D Map of Indoor+Outdoor Scene

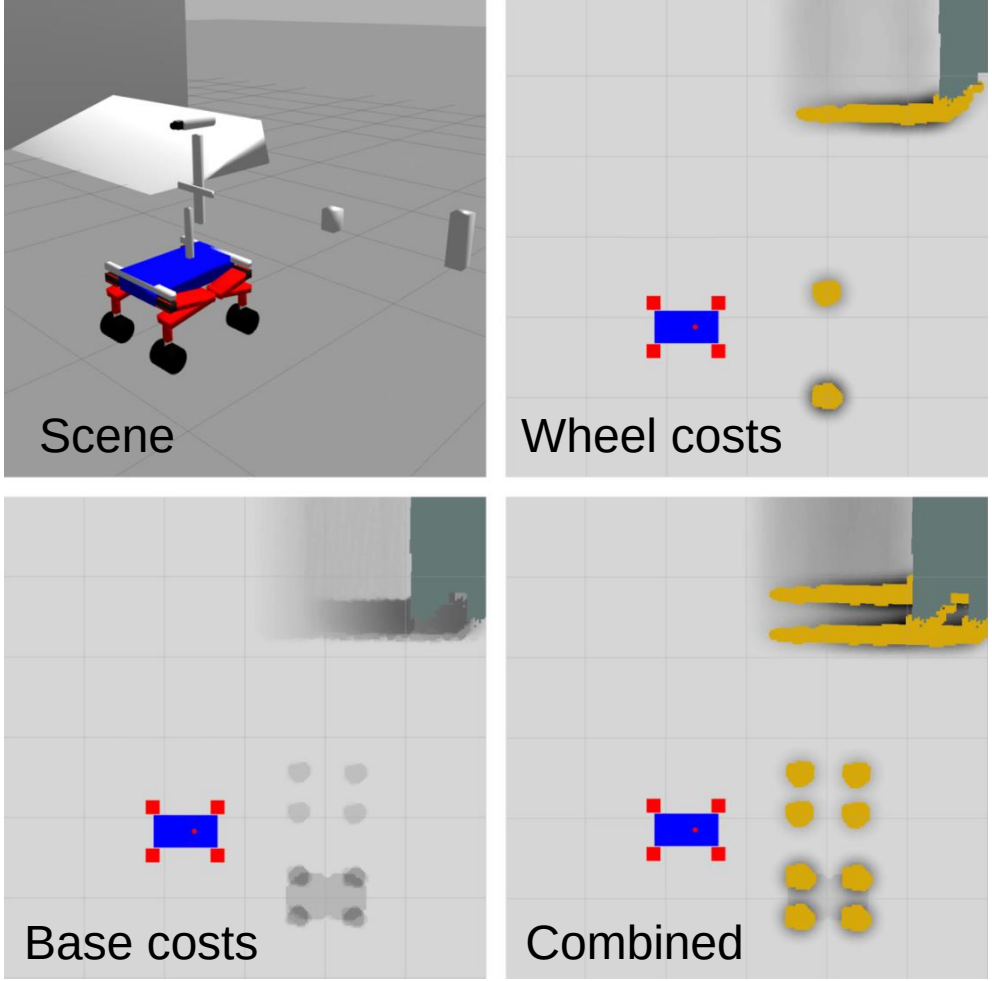


[Droeschel et al., Robotics and Autonomous Systems 2017]

Considering Robot Footprint

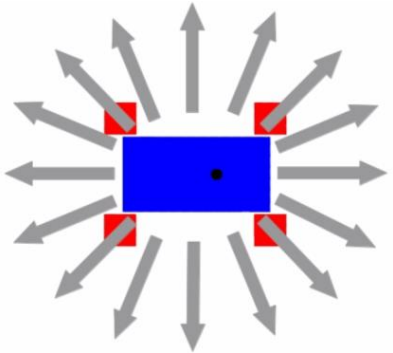
- Costs for individual wheel pairs from height differences
- Base costs
- Non-linear combination yields 3D (x, y, θ) cost map

[Klamt and Behnke, IROS 2017]

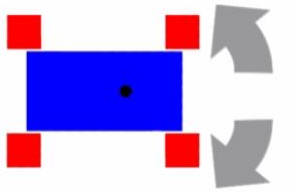


3D Driving Planning (x, y, θ): A*

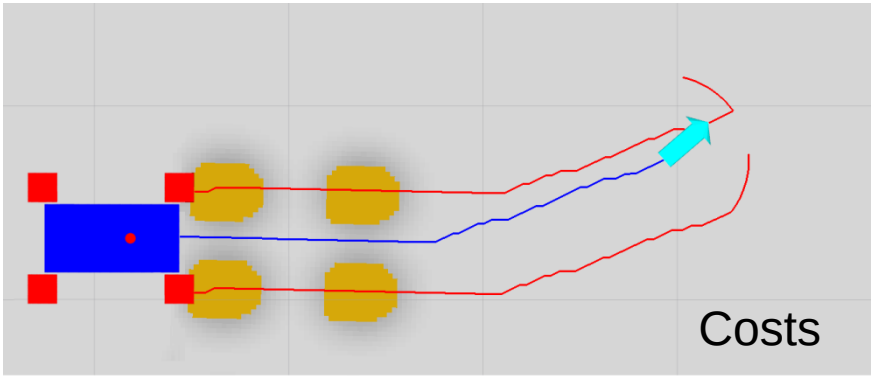
- 16 driving directions

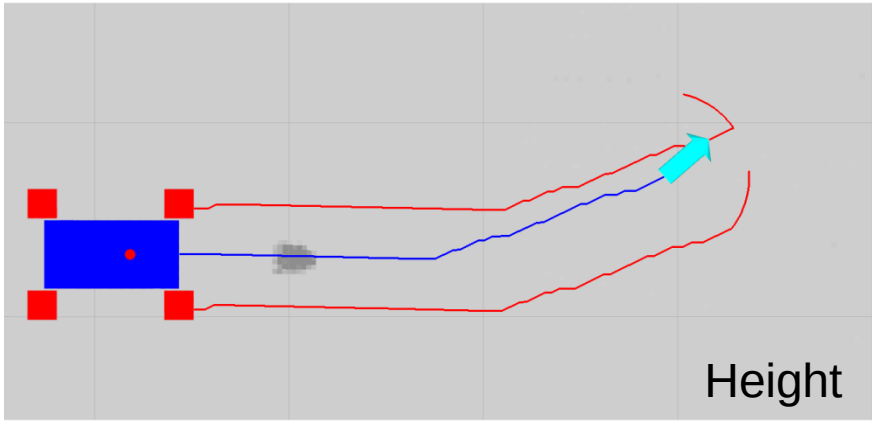


- Orientation changes



=> Obstacle between wheels

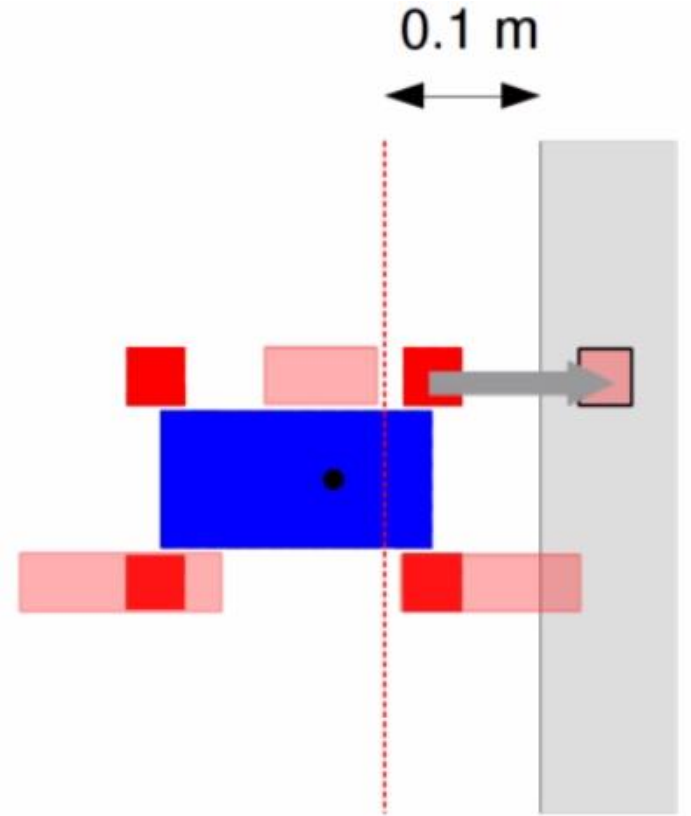




[Klamt and Behnke, IROS 2017]

Making Steps

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

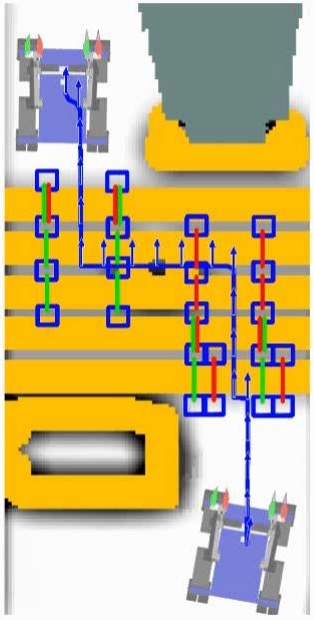


[Klamt and Behnke: IROS 2017]

Expanding Abstract Steps to Detailed Motion Sequences

[Klamt and Behnke: IROS 2017]

Planning for Challenging Scenarios



[Klamt and Behnke: IROS 2017]

New Centauro Robot

- Serial elastic actuators
- 42 main DoFs

Joint	DoF
Ankle	2
Knee	1
Hip	2
Waist	1
Shoulder	3
Elbow	1
Wrist	3
Neck	2

- Schunk hand
- 3D laser
- RGB-D camera
- Color cameras
- Two GPU PCs

[Tsagarakis et al., IIT 2017]

Climbing over a Gap

Walking over a Step Field

Transfer of Manipulation Skills

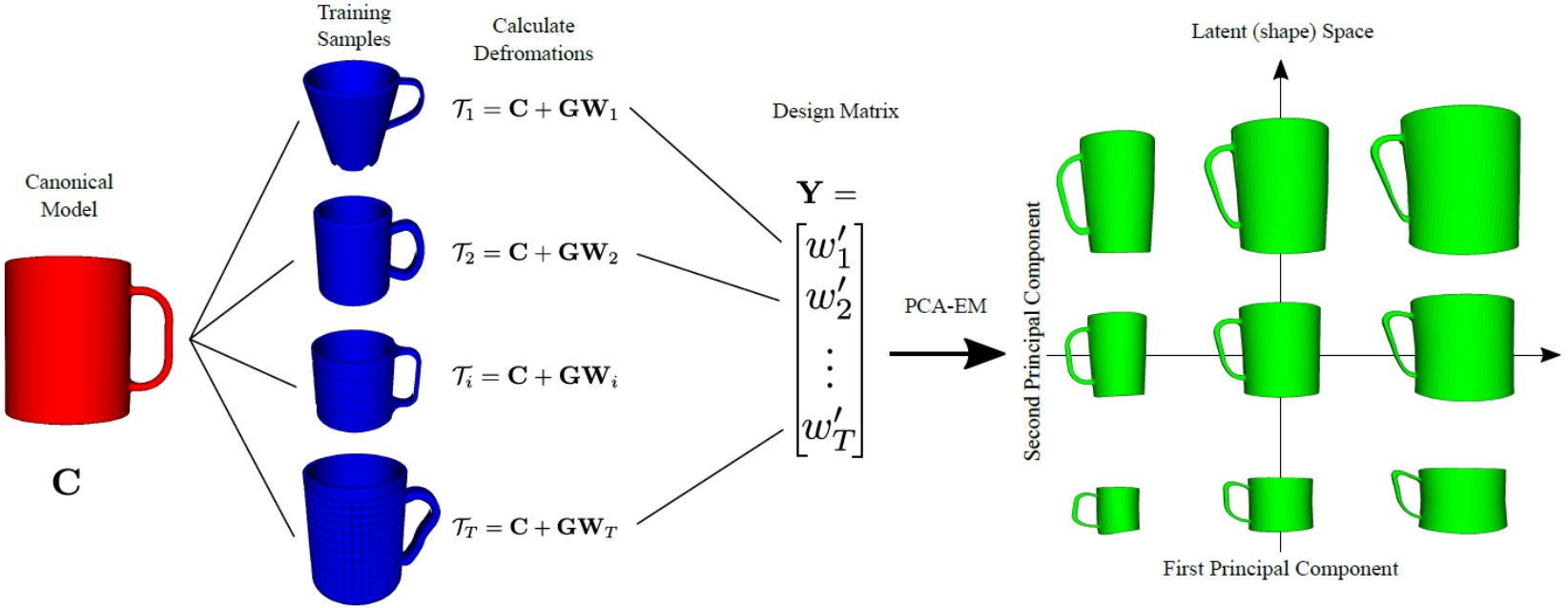
- Objects belonging to the same **category** can be handled in a very similar manner.

Transfer of Manipulation Skills

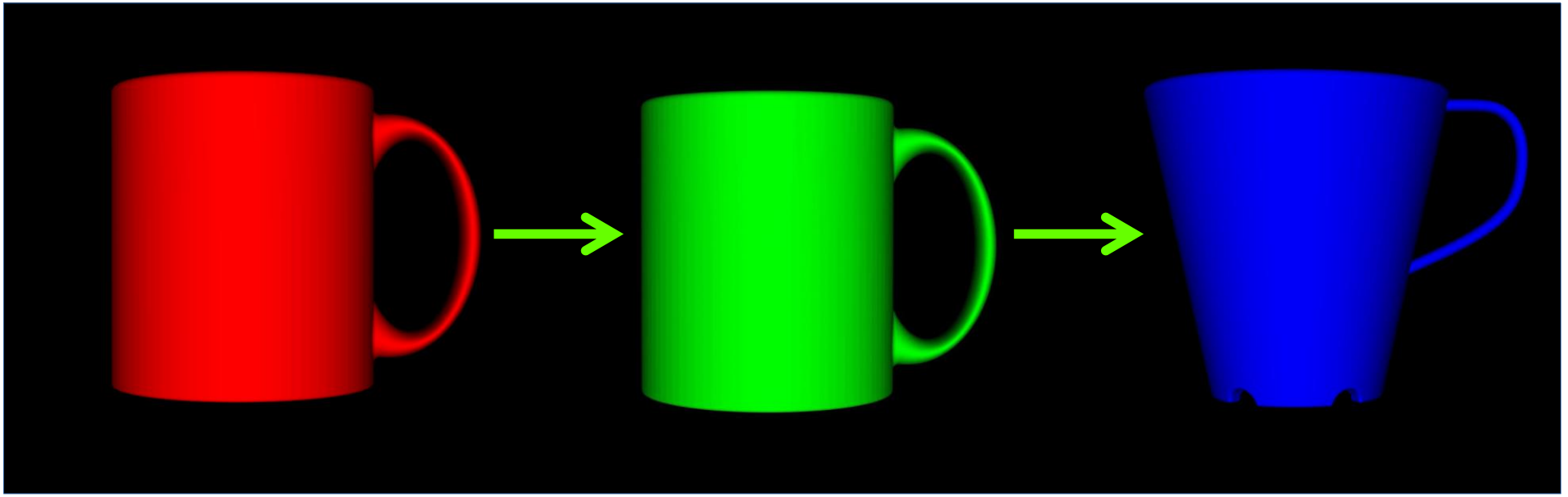
Knowledge Transfer

Learning a Latent Shape Space

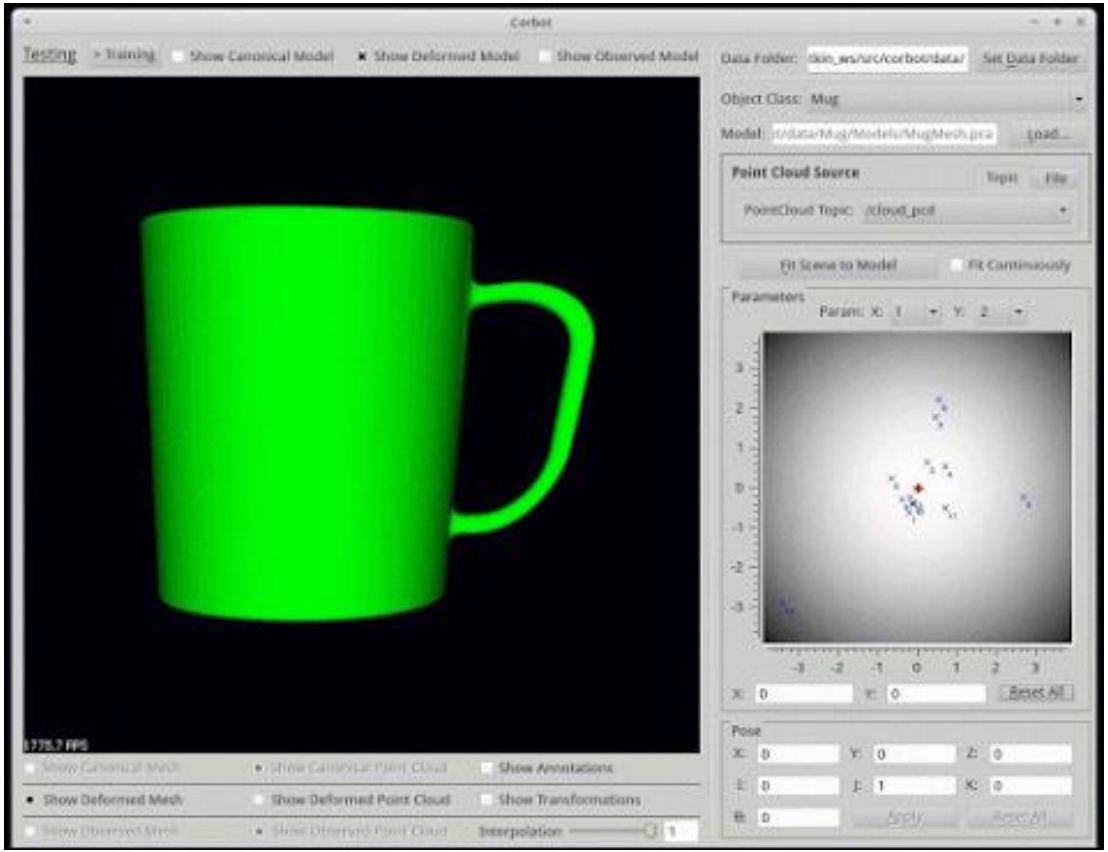
- Non-rigid registration of instances and canonical model
- Principal component analysis of deformations



Interpolation in Shape Space

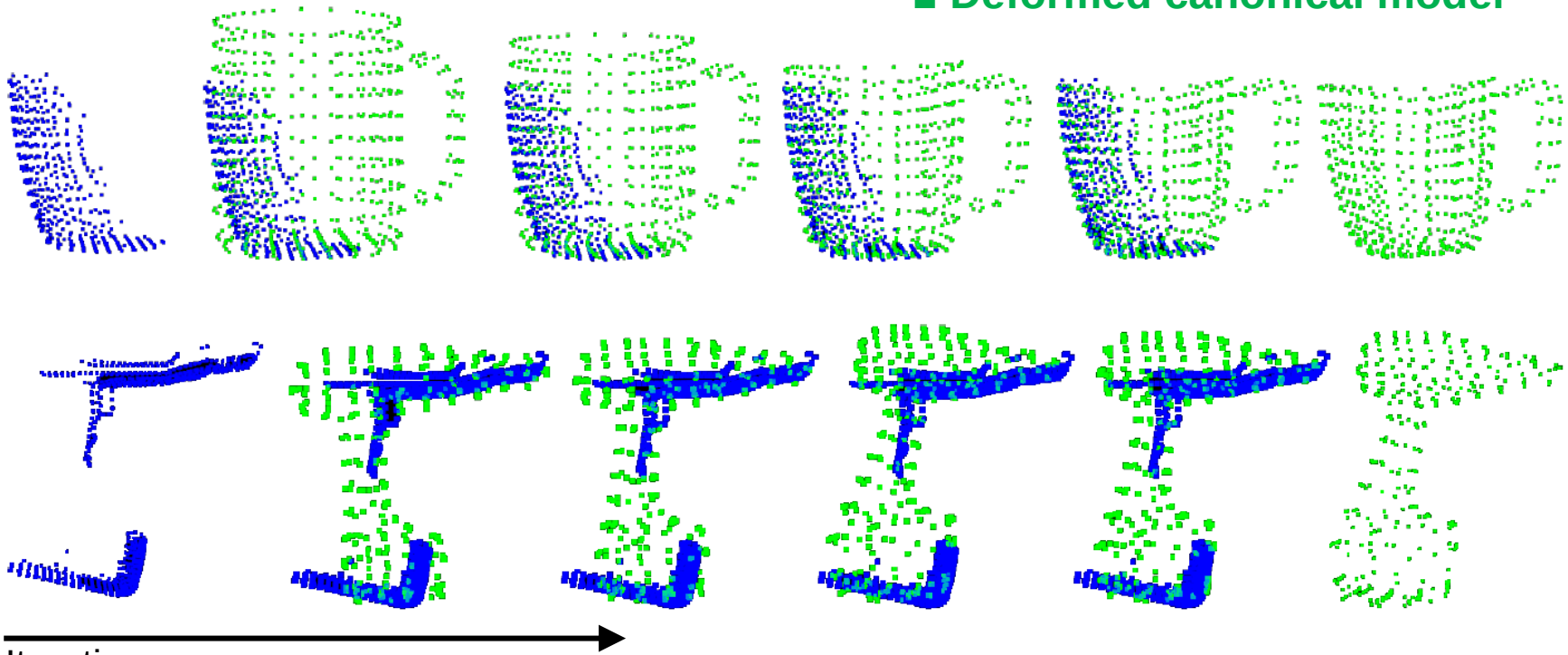


Interpolation in Shape Space



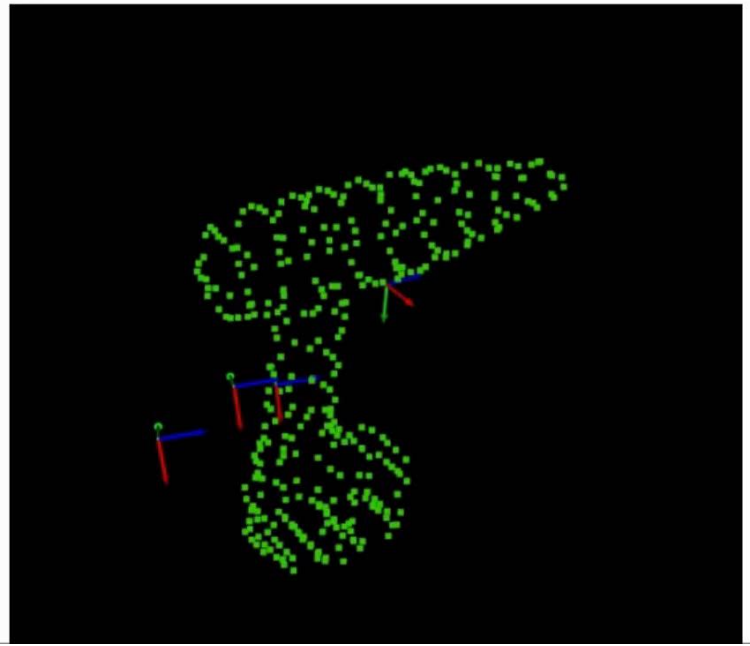
Shape-aware Non-rigid Registration

- Partial view of novel instance
- Deformed canonical model



Transference of Grasping Skills

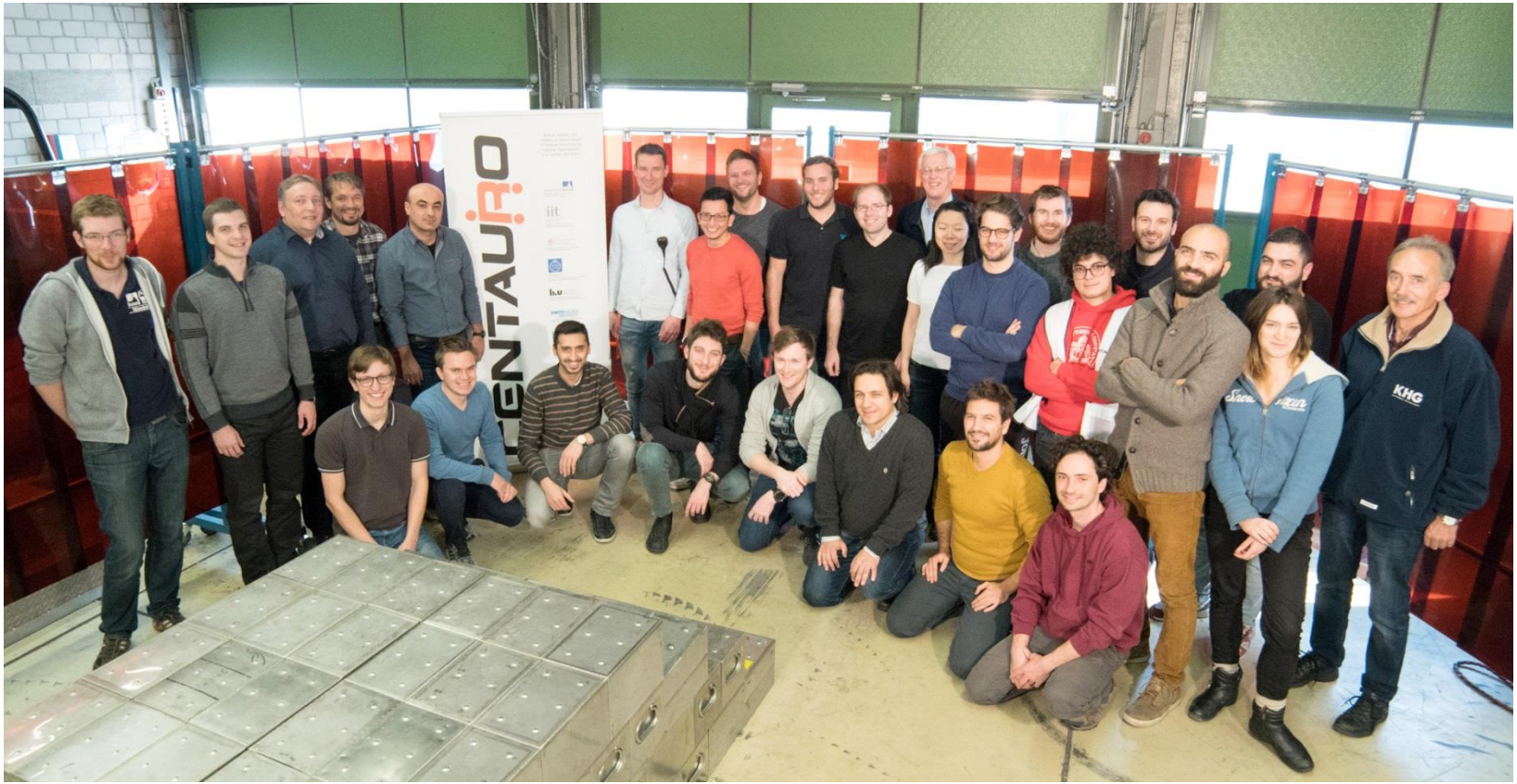
Warp grasping information



Grasping an Unknown Power Drill

Fastening a Screw

CENTAURO Team



Amazon Picking Challenge

- Large variety of objects
- Unordered in shelf or tote
- Picking and stowing tasks

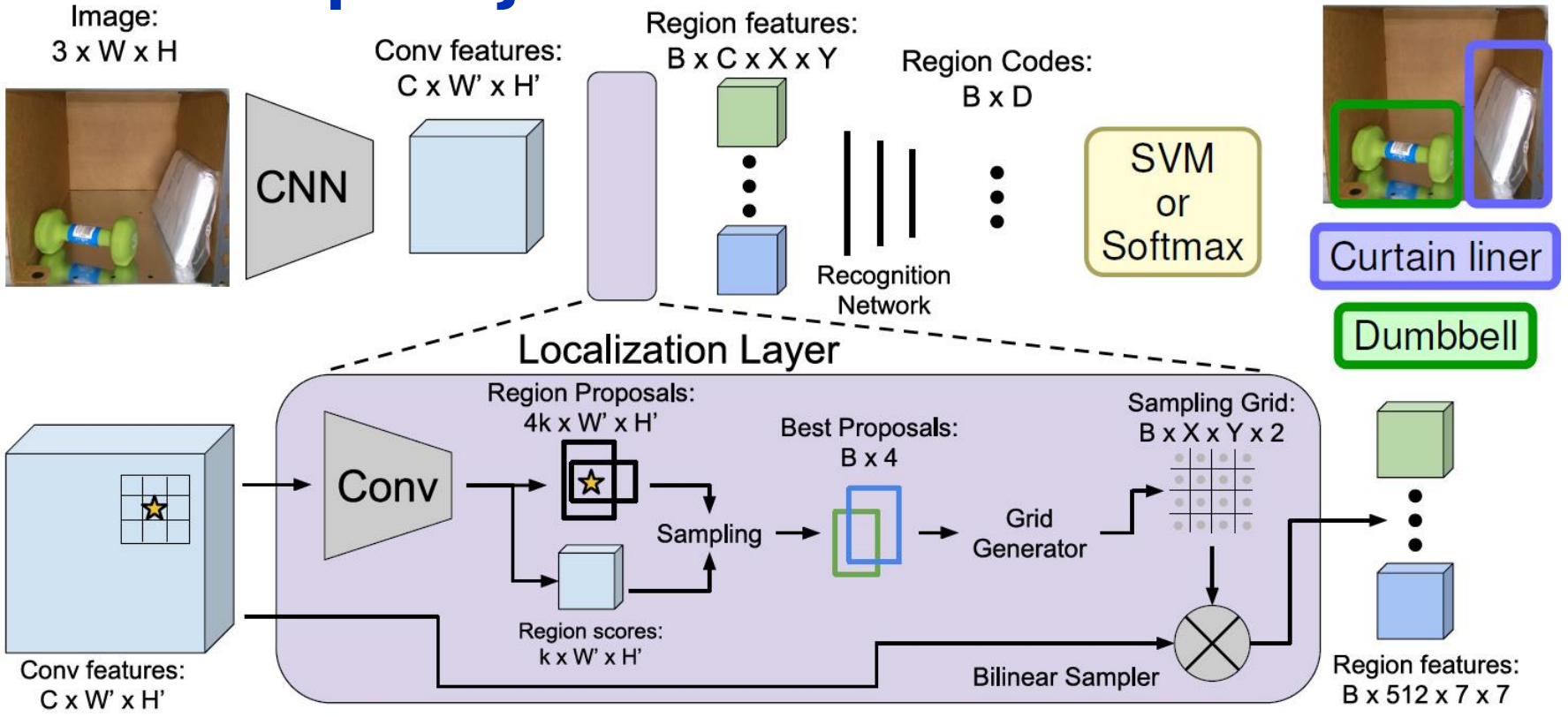
[Schwarz et al. ICRA 2017]

Deep Learning Semantic Segmentation

- Adapted from our segmentation of indoor scenes [Husain et al. RA-L 2016]

[Schwarz et al. ICRA 2017]

DenseCap Object Detection

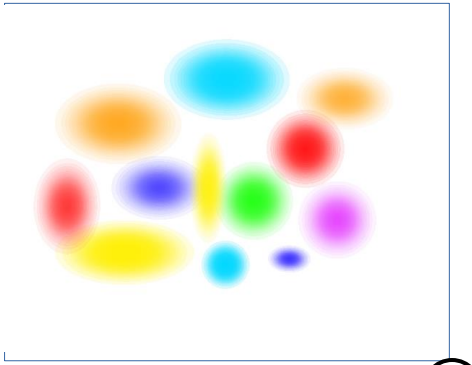


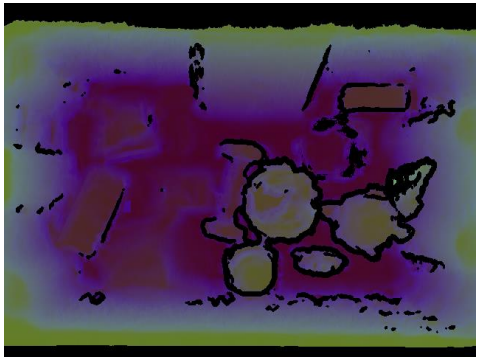
[Schwarz et al. ICRA 2017]

[Johnson et a CVPR 2016]

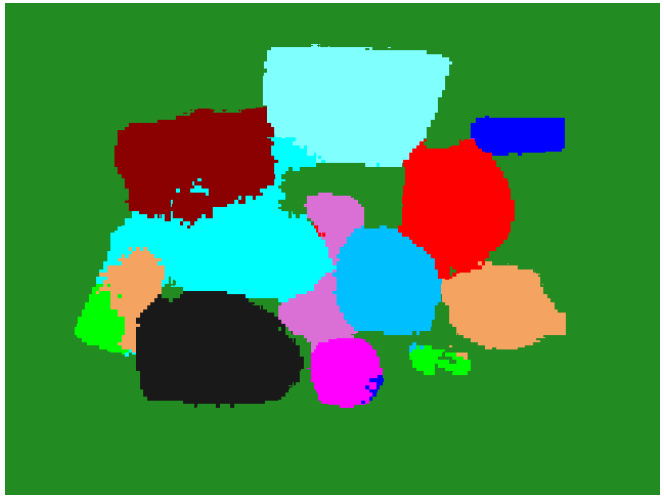
Combined Detection and Segmentation

Detection



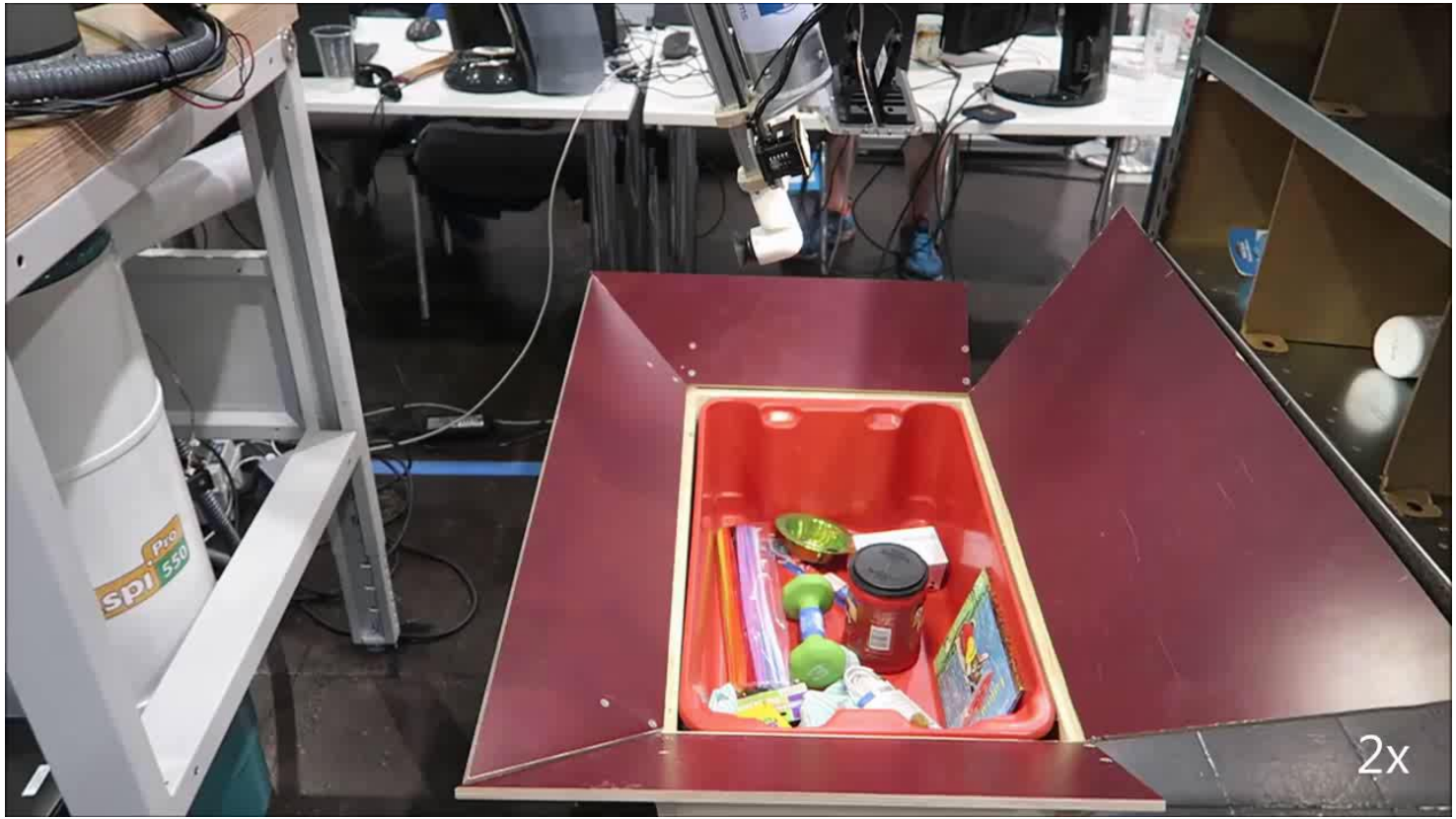


Segmentation



[Schwarz et al. IJRR 2017]

Stowing



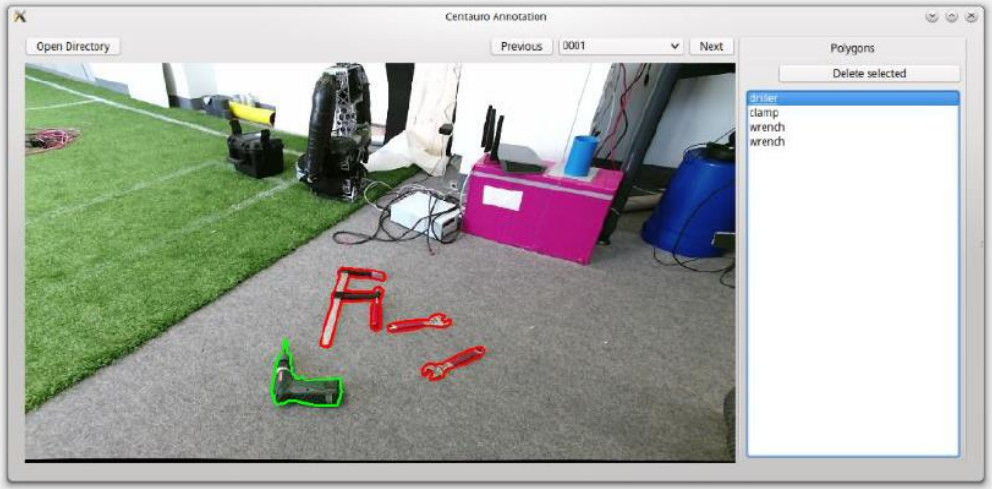
Picking

NimbRo Picking APC 2016 Results

- 2nd Place Stowing (186 points)
- 3rd Place Picking (97 points)

[Schwarz et al. IJRR 2017]

CENTAURO Workspace Perception Data Set

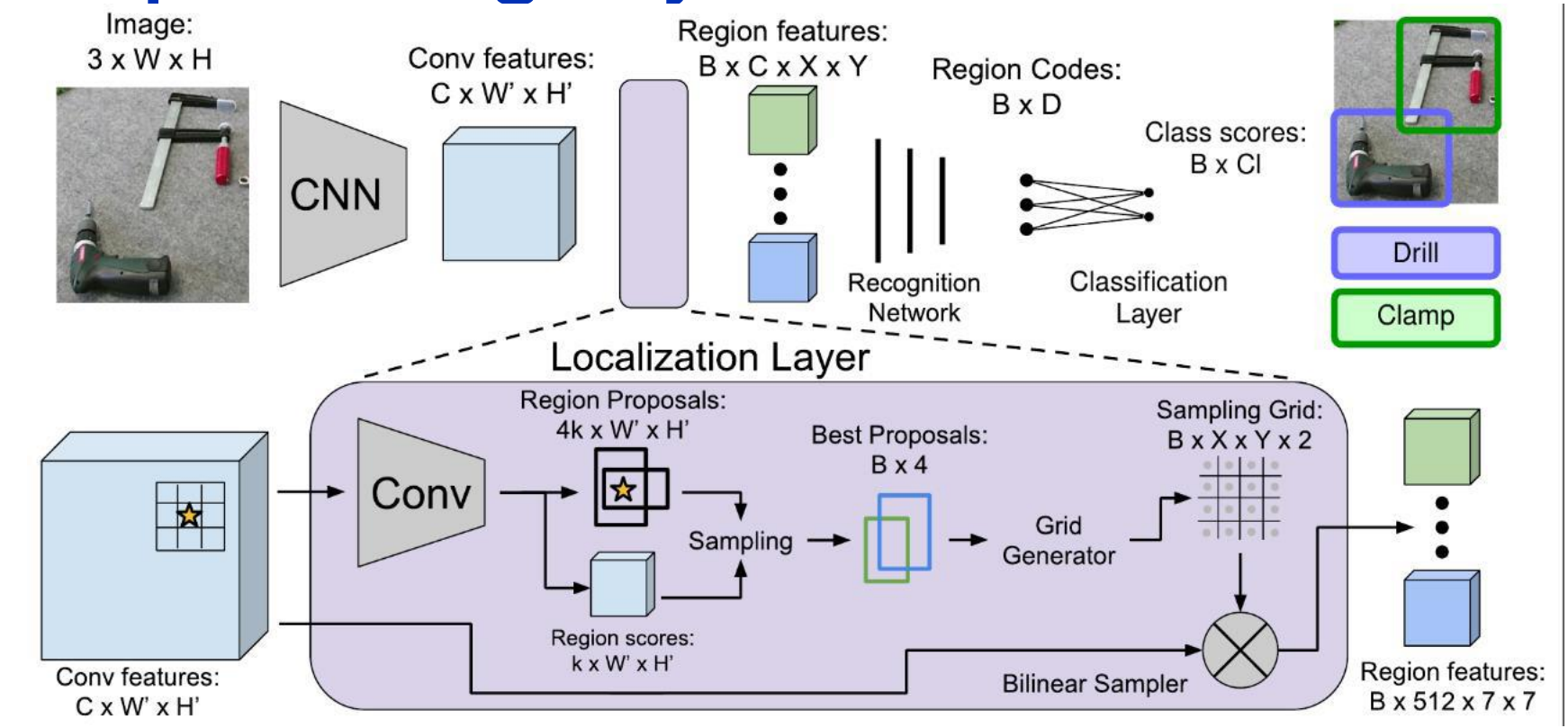


129 frames, 6 object classes

https://www.centauro-project.eu/data_multimedia/tools_data

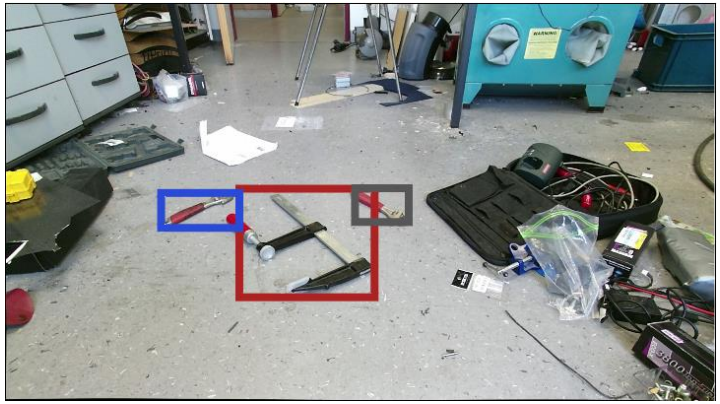
universitätbonn ais

Deep Learning Object Detection



[Johnson et al. 2015]

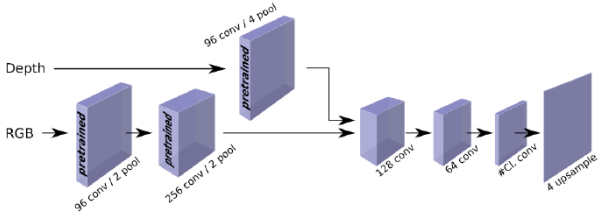
Detection of Tools



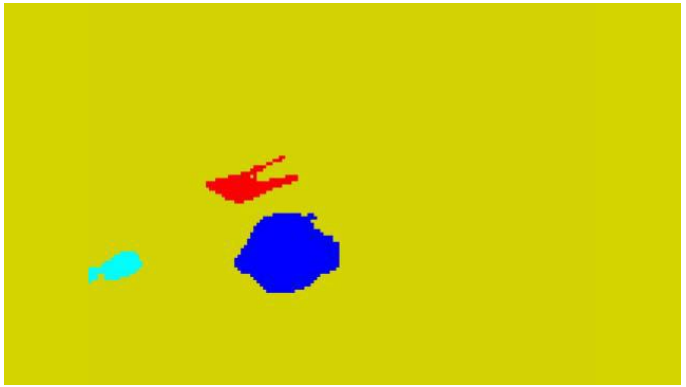
[Schwarz et al. IJRR 2017]

Semantic Segmentation

- Deep CNN



[Husain et al. RA-L 2016]

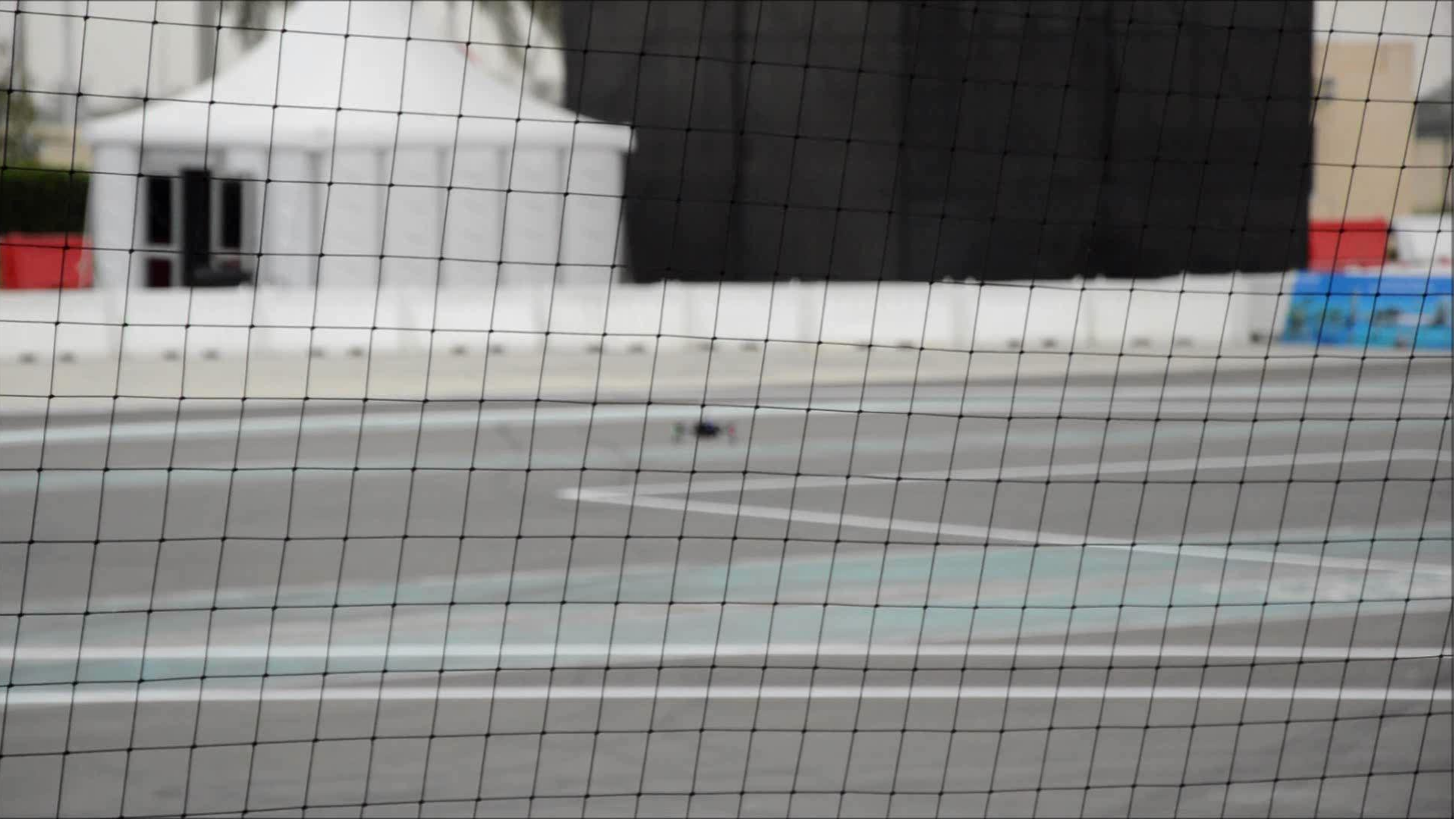


Pixel-wise accuracy:

Clamp	Door handle	Driller	Extension	Stapler	Wrench	Background	Mean
0.727	0.751	0.769	0.889	0.775	0.734	0.992	0.805

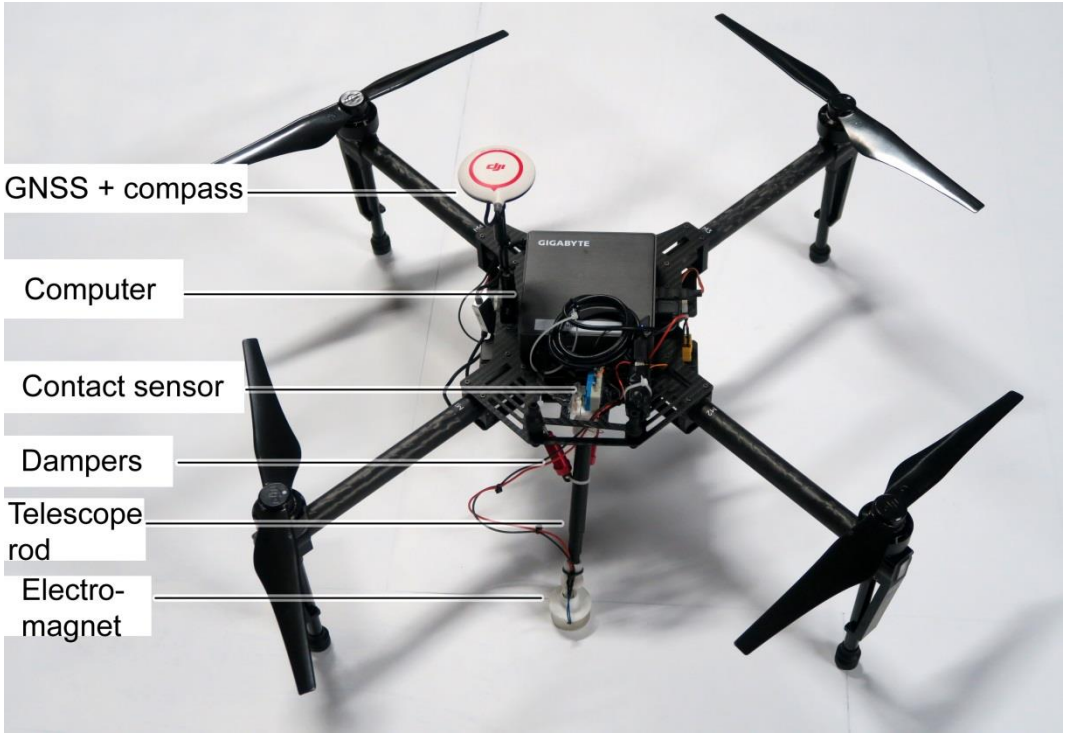
MBZIRC Challenge 2

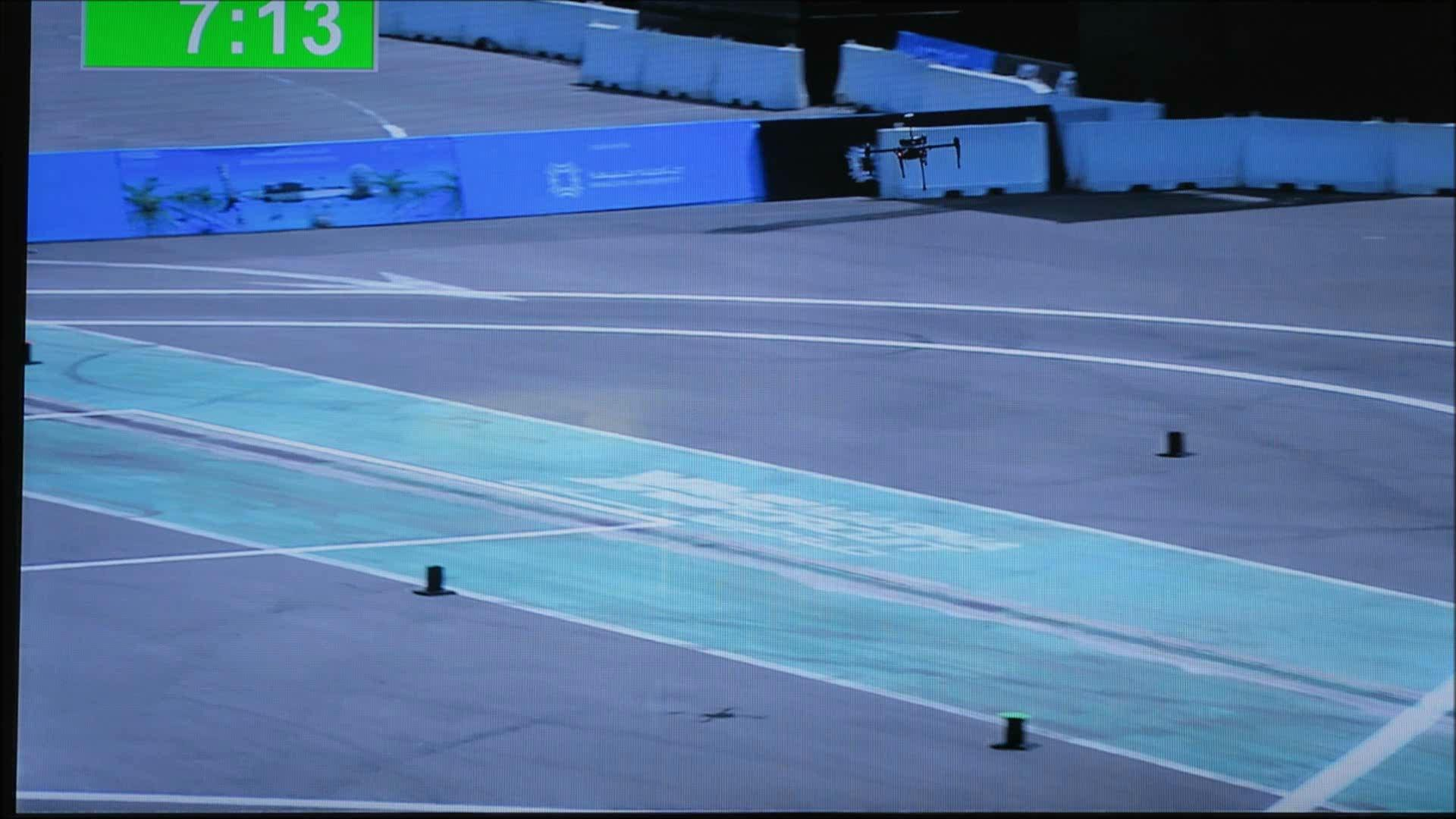
Wrench Selection: Detection of Tool Ends



Picking Copter DJI Matrice 100

- Wide-angle downward looking color camera
- Electromagnetic gripper
- Laser-distance sensor to ground
- Dual-core PC




7:13

MBZIRC Team NimbRo

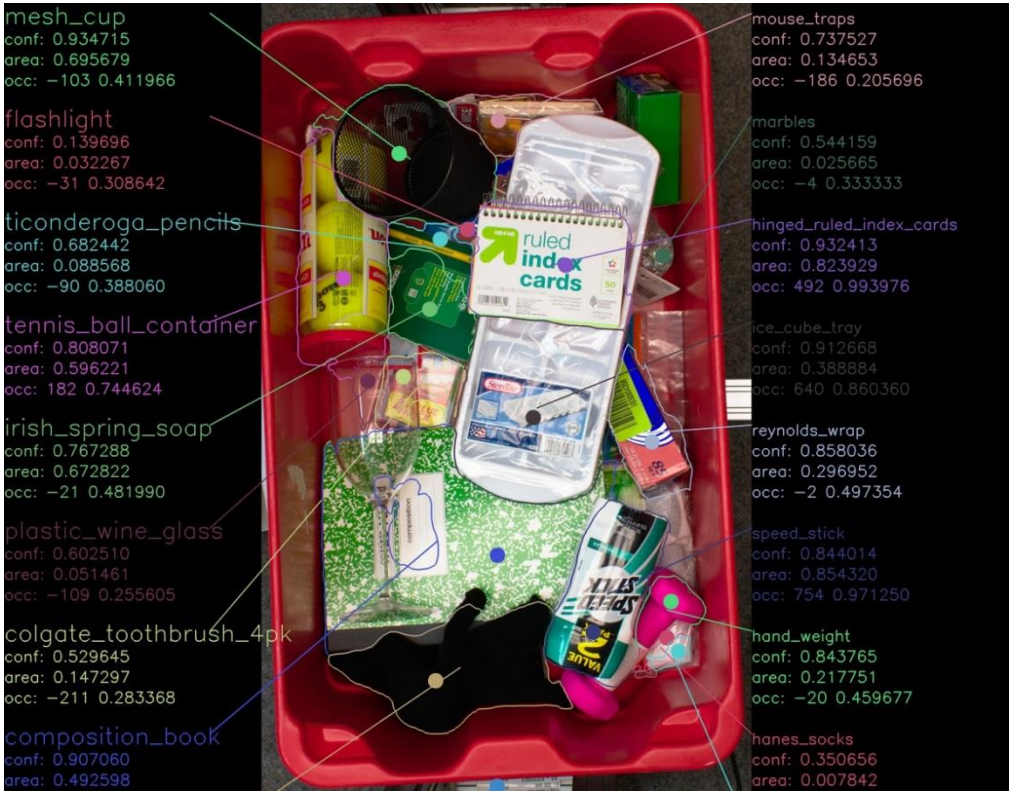
DJI Matrice 600 with Velodyne Puck

Autonomous Indoor Navigation

Fully Autonomous indoor flight without external tracking.

Amazon Robotics Challenge 2017

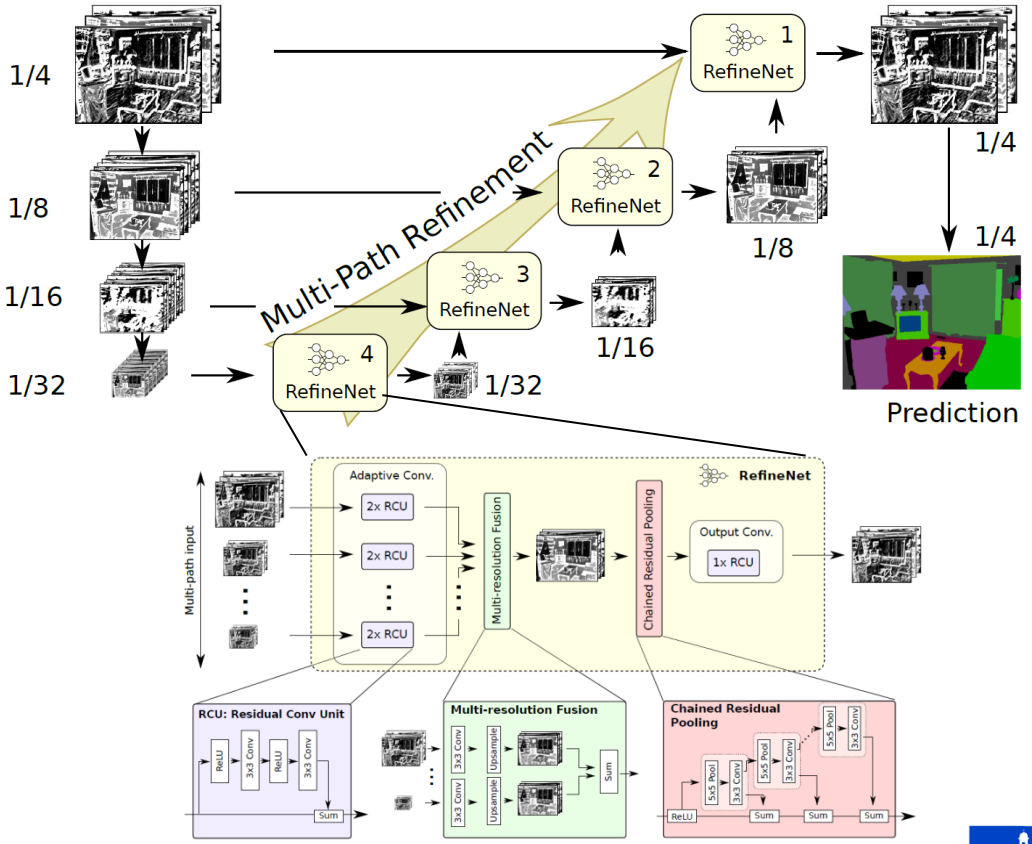
- Quick learning of novel objects
- Training with rendered scenes



RefineNet

[Lin et al. CVPR 2017]

- Increase reso-lution by using features from the higher resolution
- Corse-to-fine semantic segmentation



Object Capture and Scene Rendering

- Turn table + DLSR

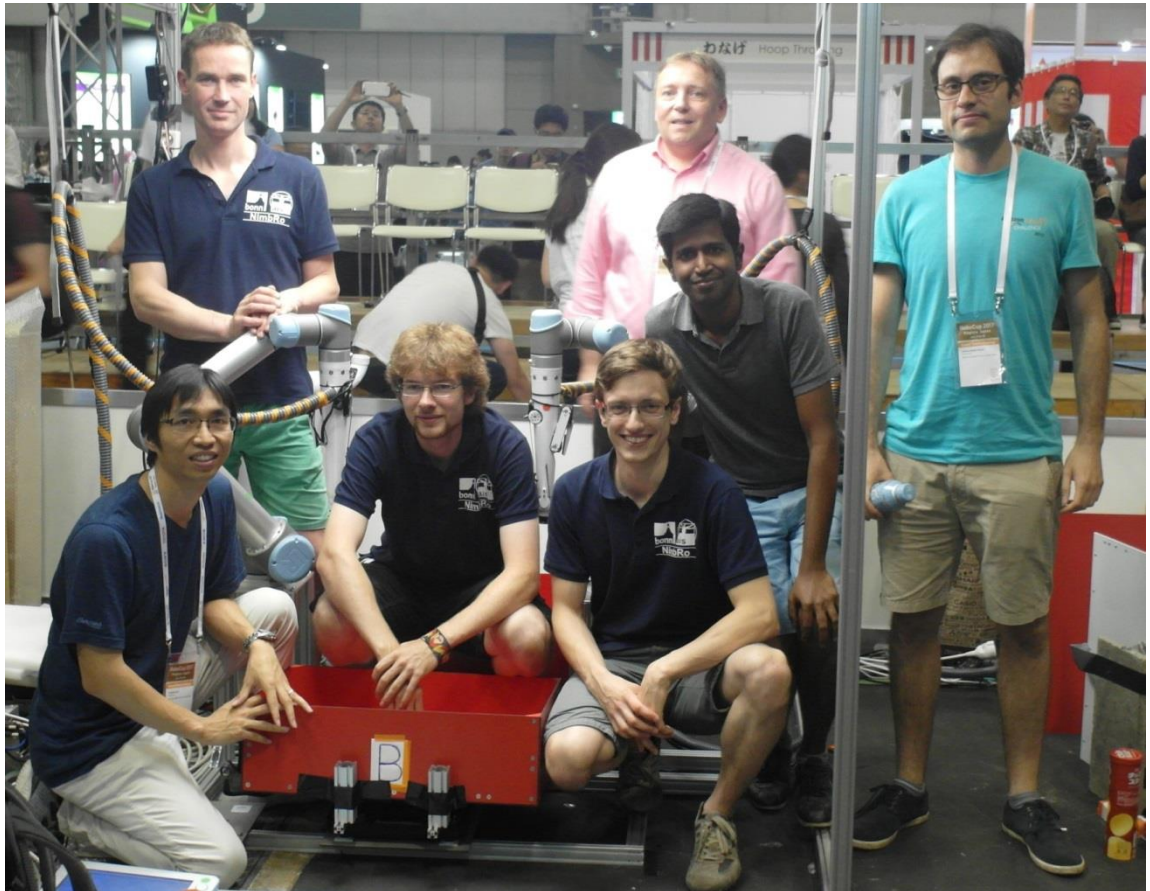
Rendered scenes

ARC 2017 Perception Example

Amazon Robotics Challenge 2017 Final

NimbRo Picking 2017 Team

- 2nd place Pick
- 2nd place Stow-and-Pick Final



Conclusions

- Semantic perception is challenging
- Simple methods rely on strong assumptions
- Depth helps with segmentation, allows for size normalization, geometric features, shape descriptors
- Deep learning methods work well
- Transfer of features from large data sets
- Synthetic training data generation
- Many open problems, e.g. total scene understanding, incorporating physics, ...

We are Hiring!

- PhD students and postdocs
- **ais.uni-bonn.de/jobs.html**

universitätbonn ais