Autonomous Assistance Functions for Mobile Manipulation Robots and Micro Aerial Vehicles

Sven Behnke
University of Bonn
Computer Science Institute VI
Autonomous Intelligent Systems
Direct Control vs. Autonomous Assistance

- Direct teleoperation offers a high degree of flexibility
- Requires special operator interfaces, good data connection, extensive operator training, and induces high cognitive load on the operator
- Not all DoFs can be mapped directly
- => Use autonomous assistance functions on all levels of control!

[Klamt et al., Journal of Field Robotics 2020]
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]
At the DARPA Robotics Challenge, Momaro demonstrated driving a car.
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]
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 $(x, y, \theta)$ cost map

[Klamt and Behnke, IROS 2017]
3D Driving Planning \((x, y, \theta)\): A*

- 16 driving directions
- Orientation changes

=> Obstacle between wheels

[Klamt and Behnke, IROS 2017]
Making Steps

- If non-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 a Challenging Scenario

[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

- Planning in the here and now
- Far-away details are abstracted away
## Hybrid Driving-Stepping Locomotion Planning: Abstraction

<table>
<thead>
<tr>
<th>Level</th>
<th>Map Resolution</th>
<th>Map Features</th>
<th>Robot Representation</th>
<th>Action Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.5 cm, 64 orient.</td>
<td>Height</td>
<td><img src="image1" alt="Image" /></td>
<td>Individual Foot Actions</td>
</tr>
<tr>
<td>2</td>
<td>5.0 cm, 32 orient.</td>
<td>Height, Height Difference</td>
<td><img src="image2" alt="Image" /></td>
<td>Foot Pair Actions</td>
</tr>
<tr>
<td>3</td>
<td>10 cm, 16 orient.</td>
<td>Height, Height Difference, Terrain Class</td>
<td><img src="image3" alt="Image" /></td>
<td>Whole Robot Actions</td>
</tr>
</tbody>
</table>

[Klamt and Behnke, IROS 2017, ICRA 2018]
Learning Cost Functions of Abstract Representations

[Planning problem]

[Klamt and Behnke, ICRA 2019]
Abstraction CNN

- Predict feasibility and costs of local detailed planning

Training data
- generated with random obstacles, walls, staircases
- costs and feasibility from detailed A*-planner
- ~250,000 tasks

[Klamt and Behnke, ICRA 2019]
Learned Cost Function: Abstraction Quality

- CNN predicts feasibility and costs better than manually tuned geometric heuristics

![Graph showing comparison between CNN and manually tuned geometric heuristics]

<table>
<thead>
<tr>
<th></th>
<th>random</th>
<th>simulated</th>
<th>real</th>
</tr>
</thead>
<tbody>
<tr>
<td>feasibility correct, man.tuned</td>
<td>79.27%</td>
<td>65.35%</td>
<td>69.77%</td>
</tr>
<tr>
<td>Error($C_a$,man.tuned)</td>
<td>0.057</td>
<td>0.021</td>
<td>0.103</td>
</tr>
<tr>
<td>feasibility correct, CNN</td>
<td>95.04%</td>
<td>96.69%</td>
<td>92.62%</td>
</tr>
<tr>
<td>Error($C_a$,CNN)</td>
<td>0.027</td>
<td>0.013</td>
<td>0.081</td>
</tr>
</tbody>
</table>

[Klamt and Behnke, ICRA 2019]
Experiments – Planning Performance

- Learned heuristics accelerates planning, without increasing path costs much

Heuristic preprocessing: 239 sec

[Klamt and Behnke, ICRA 2019]
CENTAURÔ Evaluation @ KHG: Locomotion Tasks
Object Detection

- Adapted DenseCap approach for image-based object detection

[Johnson et al. CVPR 2016]
CENTAURO Tools Data Set

129 frames, 6 object classes

https://www.centauro-project.eu/data_multimedia/tools_data
Detection Examples
Semantic Segmentation

- Adapted RefineNet approach [Lin et al. CVPR 2017]
- Synthesis of training images by capturing object views on turn table and inserting them into complex scenes
6D Object Pose Estimation

- Cutting out individual object segments
- Pose estimation by neural network

[Periyasamy et al. IROS 2018]
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
Transfer of Manipulation Skills

Knowledge Transfer
Learning a Latent Shape Space

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

\[ T_1 = C + GW_1 \]
\[ T_2 = C + GW_2 \]
\[ T_i = C + GW_i \]
\[ T_T = C + GW_T \]

\[ Y = \begin{bmatrix} w_1' \ \\
                     w_2' \\
                     \vdots \\
                     w_T' \end{bmatrix} \]
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
CENTAURO: Complex Manipulation Tasks

[Klamt et al. RAM 2019]
Regrasping for Functional Grasp

- 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 Experiments

[ Pavlichenko et al. Humanoids 2019]
Part-based Non-rigid Object Registration

■ Objects consist of parts
■ Learn shape spaces of parts individually
■ Captures object shapes better
■ Robust against outliers, noise and initial pose misalignment

[Rodriguez et al. VISAPP 2022]
Dense Convolutional 6D Object Pose Estimation

- Extension of PoseCNN [Xiang et al. RSS 2018]
- Dense prediction of object center and orientation, without cutting out

[Capellen et al., VISAPP 2020]
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.

[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

[Periyasamy, Schwarz, Behnke Humanoids 2019]
Abstract Object Registration

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

[Periyasamy, Schwarz, Behnke Humanoids 2019]
Registration Examples

[Periyasamy, Schwarz, Behnke Humanoids 2019]
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 and Behnke, ICRA 2020]
SynPick: A Dataset for Dynamic Bin Picking Scene Understanding

- Object arrangement and manipulation simulation using NVIDIA PhysX
- Untargeted and targeted picking actions, as well as random moving

[Periyasamy et al. CASE 2021]
T6D-Direct: Transformers for Multi-Object 6D Pose Direct Regression

- Extends DETR: End-to-end object detection with transformers [Carion et al. ECCV 2020]
- End-to-end differentiable pipeline for 6D object pose estimation

![Diagram showing the T6D-Direct method]

Encoder self-attention

Object detections and decoder attention

[Amini et al. GCPR 2021]
YoloPose: Multi-Object 6D Pose Estimation using Keypoint Regression

[Image of YoloPose architecture]

Amini et al. IAS 2022
Attention Maps

- Encoder self-attention

- Decoder cross-attention

[Amini et al. IAS 2022]
Micro Aerial Vehicles: Hierarchical Navigation

- User
  - Request
  - Operator station
    - Semantic map
      - Mission planning
        - <0.02 Hz Observation poses
          - Allocentric planning
            - 0.2 Hz Allocentr. plan
              - Egocentric planning
                - 2 Hz Trajectory
                  - Obstacle avoidance
                    - 20 Hz Speed
                      - Copter

- Mission plan
- Allocentric planning
- Egocentric planning
- Obstacle avoidance

[Droeschel et al. JFR 2016]
InventAIRy: Autonomous Navigation in a Warehouse

[Beul et al. RA-L 2018]
InventAIRy: Detected Tags in Shelf

[Beul et al. RA-L 2018]
Label Propagation for 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]
German Rescue Robotics Center

Initial demonstrator

- Basis: DJI Matrice 600 Pro
- Sensors: Velodyne VLP 16, FLIR Boson, 2x FLIR BlackFly S
- Tiltable sensor head

Current demonstrator

- Basis: DJI Matrice 210 v2
- Sensors: Ouster OS-0, FLIR AGX, 2x Intel RealSense D455
- IP43 water resistance
Modeling the Brandhaus Dortmund

[Rosu et al. SSRR 2019]
Real-time LiDAR Odometry with Continuous-time Trajectory Optimization

- Simultaneous registration of multiple multiresolution surfel maps using Gaussian mixture models and temporally continuous B-spline
- Accelerated by sparse permutohedral voxel grids and adaptive choice of resolution
- Real-time onboard processing 16-20 Hz
- Open-Source
  https://github.com/AIS-Bonn/lidar_mars_registration

[Quenzel and Behnke, IROS 2021]
3D LiDAR Mapping

DRZ Living Lab
Semantic Perception: LiDAR Segmentation

- LatticeNet segmentation of 3D point clouds based on sparse permutohedral grid
- Hierarchical information aggregation through U-Net architecture
- LatticeNet is real-time capable and achieves excellent results in benchmarks

[Rosu et al., RSS 2020]
Semantic Fusion: 3D LiDAR Mapping

Categories:
- Building
- Floor
- Persons
- Vehicles
- Fence
- Vegetation

Segmented point cloud

Minimax-Viking fire house

Semantic multiresolution surfel map
Semantic Fusion: Temporal LatticeNet

- Semantic segmentation of sequences of 3D point clouds
- Integration of recurrent connections
- Trained on three scans of SemanticKITTI
- Distinguishing moving from parking vehicles

[Rosu et al. Autonomous Robots 2021]
Onboard Multimodal Semantic Fusion

- Real-time semantic segmentation and object detection (≈9Hz) with EdgeTPU / iGPU
  - SalsaNext for LiDAR
  - DeepLabv3 for RGB images
  - SSD MobileDet for Thermal/RGB

- Late-fusion for
  - Point cloud
  - Image segmentation

[Bultmann et al. ECMR 2021]
LiDAR-based Obstacle Avoidance

- Fast analytical collision check with 3D point cloud
- Planning of alternative trajectories if original trajectory causes collision
- Selection and execution of a collision-free alternative trajectory

[Beul and Behnke, SSRR 2020]
Dynamic 3D Navigation Planning

- Positions and velocities in sparse local multiresolution grid
- Adaptation of movement primitives to grid
- Optimization of flight time and control costs
- 1 Hz replanning

[Schleich and Behnke, ICRA 2021]
Planning with Visibility Constraints

- Extra costs for flight through unmapped volumes

- Consideration of sensor frustum:
  - Coupling of vertical and horizontal motion
  - Preferred forward flight with limited rotational speed
Observation Pose Planning

- Planning of observation poses with line of sight to the target object despite occlusions
- Target objects are defined by position, line of sight and distance
- Optimization of observation poses with regard to visibility quality and accessibility
Autonomous Flight without GNSS

DRZ Dortmund
Exploration

- Definition of target area w.r.t. satellite images or streets
- Simple exploration patterns (spirals, meanders, ...)
- Collision check
- TSP to determine segment sequence
- Continuous replanning

Campus Poppelsdorf
Terrain Classification for Traversability

- Based on voxel-filtered aggregated point cloud
- Terrain classification based on local height differences in the robot ground robot footprints
- Categories: drivable, walkable, unpassable
- Reachability analysis

[Schleich et al., ICUAS 2021]
Conclusions

- Developed capable robotic systems for disaster-response
  - Centaur-like ground robots
  - Micro aerial vehicles

- Challenges include
  - 4D semantic perception
  - High-dimensional motion planning

- Promising approaches
  - Prior knowledge (inductive bias)
  - Data generation (rendering, simulation)
  - Shared experience (fleet learning)
  - Shared autonomy (human-robot)
Challenges are HUGE, see Flooding in Erftstadt, Germany July 2021