Lab Cognitive Robotics
Seminar Cognitive Robotics
Projektgruppe Kognitive Robotik

Raphael Memmesheimer and Sven Behnke

12.07.2024

Rheinische Friedrich-Wilhelms-Universität Bonn
Computer Science Institute VI: Autonomous Intelligent Systems
memmesheimer@ais.uni-bonn.de

BEFORE WE START

There is no need to take notes, you can download the slides here:

http://ais.uni-bonn.de/SS24/Lab_Cognitive_Robotics/lab.pdf

Cognitive robotics:
Active research area at the border between AI and robotics.

Investigation and implementation of mental functions:

- Perception of the environment,
- action planning, and
- learning.

Cognitive Robotics Lab (or Projektgruppe for Bachelor students):

- Implementation of state-of-the-art algorithms
- Work in small groups
- Good preparation for a Bachelor/Master thesis!

DATES & REGISTRATION

Registration (limit: 12 students PG/lab, 6 students seminar):

1. Register with us here: PG/Lab: `https://forms.gle/AGBqViD79q55kr67A`, Seminar: `https://forms.gle/k1TqmguBhu5oxG279`
2. On 18.07.2024we will close the registration. If there are too many applications, we will choose 12 students randomly.
3. You will get an e-mail with your registration result on 19.07.2024.
4. If you have a place, you can register in BASIS for your exam (bachelor/master). Uninvited BASIS registrations will be unregistered again!

Work period for Lab / PG: **26.08.2024 - 13.09.2024**

Presentations and Seminar: **20.09.2024**

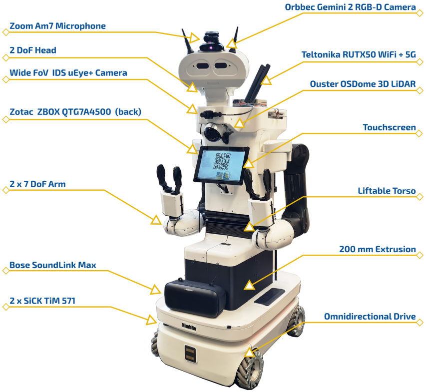
REQUIREMENTS

These are your contributions during the lab:

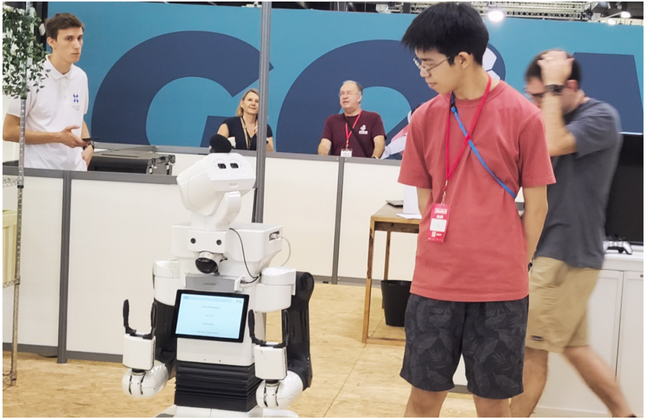
1. Full-time work during the lab period
2. Presentations: 10+5min for lab course, 30+15min for seminar
3. Report: Maximum eight pages, recommended six pages. Template will be provided.

All three will be graded.

Autonomous Intelligent Systems

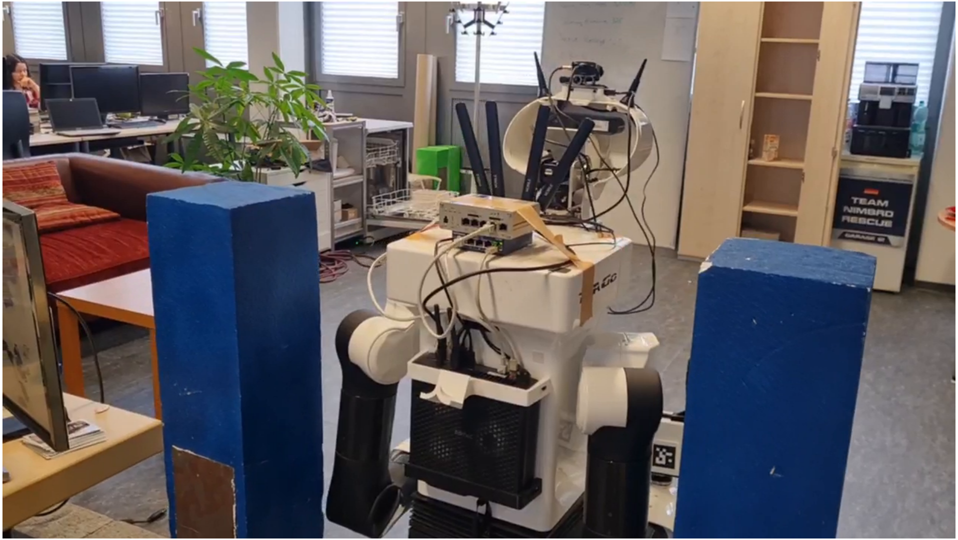


TIAGo

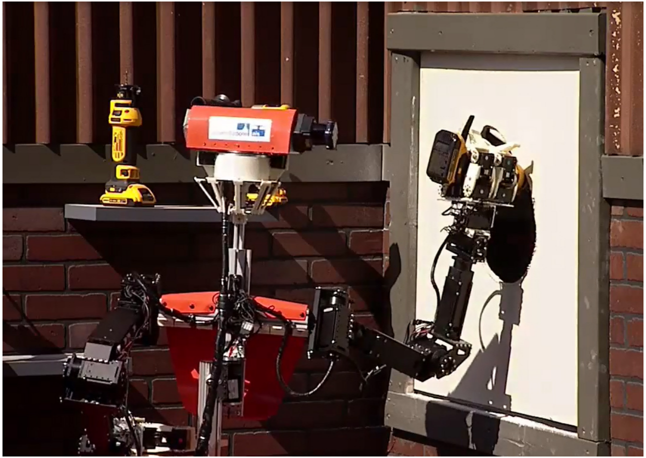


Receptionist

ROBOCUP@HOME: CLEAN THE TABLE TASK



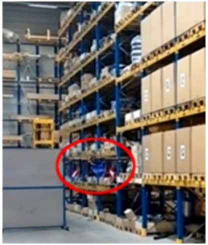
SEARCH & RESCUE ROBOTS

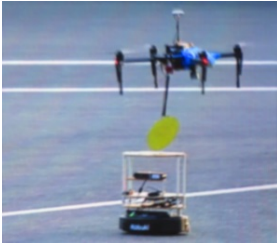


[Schwarz et al.: Frontiers in Robotics and AI 2016, JFR 2017]

[Allgeuer et al.: Humanoids 2015, 2016]

MICRO AERIAL VEHICLES

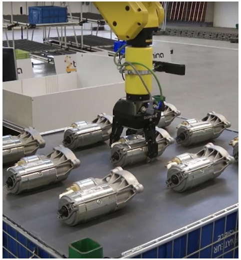




[Nieuwenhuisen et al.: JINT 2015, Droeschel et al.: JFR 2016]

BIN PICKING ROBOTS

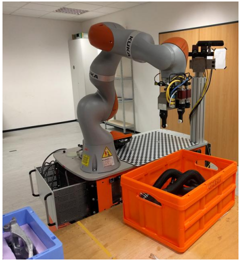
ActReMa



STAMINA

APC

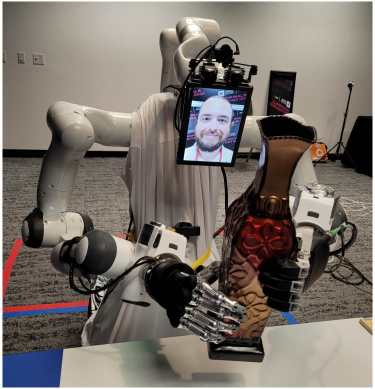
EuRoC C1



EuRoC C2

ARC

ANA AVATAR XPRIZE CHALLENGE

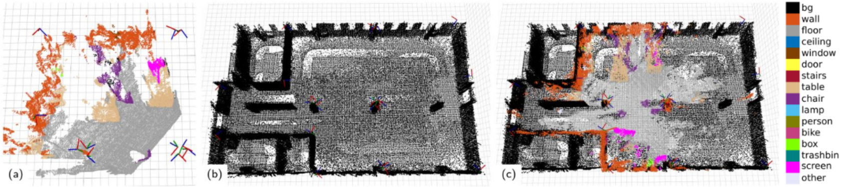


https://www.youtube.com/watch?v=8AwgGSpcAe8

Lab Topics

1) POINT CLOUD COMPRESSION

Beneficial: C++, ROS | Students: Bachelor/Master | Supervisor: Simon Bultmann



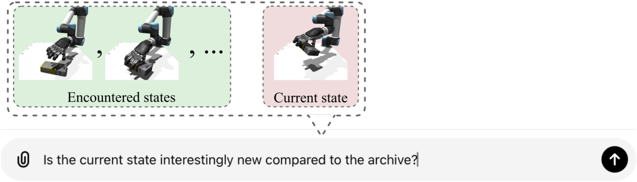
3D semantic mapping: (a) semantic point cloud of a single sensor, (b) prior map, (c) fused semantic map.

- Utilizing a network of smart edge sensors with RGB-D cameras for collaborative 3D semantic scene perception[1]
- High network bandwidth required for transmission of point cloud data, temporal redundancy can be further exploited for compression[2]
- Aim: Develop real-time temporal point cloud compression approach for stationary RGB-D cameras, transmitting only measurements that deviate from the static scene and updating variance parameters when scene geometry changes

[1]: Bultmann and Behnke, 3D Semantic Scene Perception using Distributed Smart Edge Sensors, IAS 2022
[2] Kammerl et al., Real-time compression of point cloud streams, ICRA 2012

2) FOUNDATION MODELS TO UNDERSTAND STATE-NOVELTY FOR EFFECTIVE EXPLORATION

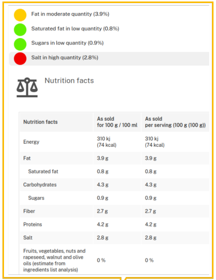
Beneficial: PyTorch, Python | Students: Master | Supervisor: Malte Mosbach



- Exploration and scene understanding are key issues in reinforcement learning for robotics.
- Foundation models (FMs) embed world knowledge, useful for addressing this challenge [1]
- Given an image, the goal is use FMs to understand the relevance and novelty of the encountered state
- For this, you will need to design prompts for the FM to analyze the image content and to compare efficiently (for example via RAG [2]))
- Assess the results in terms of how states are grouped and which interesting behaviors were discovered.

[1]: Lu, Cong, Shengran Hu, and Jeff Clune. Intelligent Go-Explore: Standing on the Shoulders of Giant Foundation Models. arXiv preprint arXiv:2405.15143 (2024).
[2]: Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." Advances in Neural Information Processing Systems 33 (2020): 9459-9474

Beneficial: Python / NumPy / PyTorch | Students: Bachelor/Master | Supervisor: Jan Nogga



Fat in moderate quantity (3.9%)
Saturated fat in low quantity (0.8%)
Sugars in low quantity (0.9%)
Salt in high quantity (2.8%)

Nutrition facts

Nutrition facts	As sold for 100 g / 100 ml	As sold per serving (100 g (100 g))
Energy	310 kj (74 kcal)	310 kj (74 kcal)
Fat	3.9 g	3.9 g
Saturated fat	0.8 g	0.8 g
Carbohydrates	4.3 g	4.3 g
Sugars	0.9 g	0.9 g
Fiber	2.7 g	2.7 g
Proteins	4.2 g	4.2 g
Salt	2.8 g	2.8 g
Fruits, vegetables, nuts and rapeseed, walnut and olive oils (estimate from ingredients list analysis)	0 %	0 %

Implemention of a language-image retrieval system for our domestic service robots which is capable of specifying the nutritional values of food given natural language or visual input.

- The OpenFoodFacts database [1] is augmented by computing OpenCLIP [2] text and image embeddings for each entry.
- These form an index for identification of database entries similar to given query embeddings.
- Finally the retrieval system is integrated on one of our domestic service robots to demonstrate that it can provide helpful information when assisting humans in everyday situations.

[1]: `https://world.openfoodfacts.org/`,
[2]: `https://arxiv.org/pdf/2212.07143.pdf` Cherti et al.

4) SMART-EDGE NEURAL FIELDS

Beneficial: Python, PyTorch, ROS | Students: Master | Supervisor: Evgenii Kruzhkov

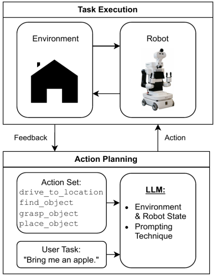
Creation of Neural Fields-based maps of our lab using multiple fixed sensors and analyse implicit and explicit representations.

- High quality map of the environment.
- Modern neural fields -based environment representation in robotics applications.
- Distributed real-time neural mapping
- Test and compare modern neural fields-based mapping techniques

[1]: https://www.matthewtancik.com/nerf,
[2]: https://dazinovic.github.io/neural-rgbd-surface-reconstruction/,

5) LLM-BASED COMMUNICATION AGENT FOR HUMAN-ROBOT INTERACTION

Beneficial: ROS2 / Python | Students: Bachelor/Master | Supervisor: Bastian Pätzold



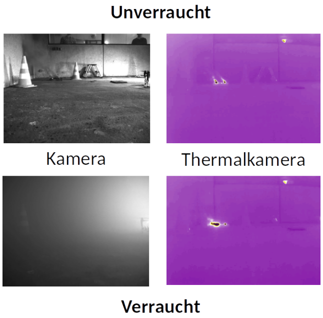
Develop an LLM-based agent to achieve complex human-robot communication goals.

- Enhance our existing approach through improved turn taking, sanity checking and adaptable style.
- Embed it into an existing framework based on the function calling API of state-of-the-art LLMs (OpenAI or Mistral).
- Integrate it using existing speech recognition and synthesis pipelines on a PAL Robotics TiaGo platform.

[1]: Kim et al., Understanding large-language model (LLM)-powered human-robot interaction, HRI 2024.

[2]: Pages et al., TIAGo: The modular robot that adapts to different research needs, IROS 2016.

Beneficial: C++ / ROS | Students: Bachelor/Master | Supervisor: Jan Quenzel



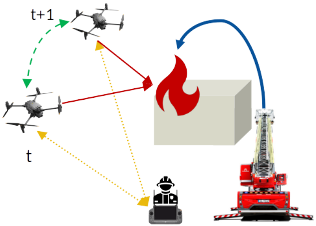
Adapt existing Visual-Inertial Odometry (VIO) for thermal cameras to track camera pose in low-visibility conditions, such as darkness or smoke-filled environments. Compare against other state-of-the-art methods.

- Research on adapting VIO for thermal cameras
- Implementation and testing of the adapted VIO
- Comparison against other state-of-the-art methods

[1]: Geneva et al., ”OpenVINS: A Research Platform for Visual-Inertial Estimation”, ICRA 2020
[2]: Das et al., ”Online Photometric Calibration of Automatic Gain Thermal Infrared Cameras”, R-AL 2021

7) ROBUST MULTI-VIEW FIRE LOCALIZATION

Beneficial: ROS / Python / PyTorch | Students: Bachelor/Master | Supervisor: Daniel Schleich



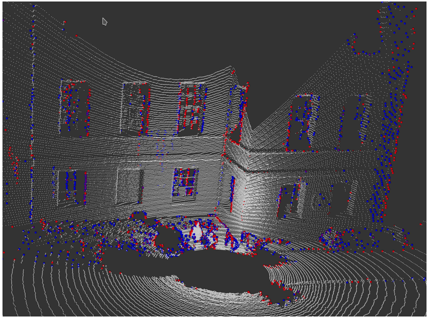
Develop a robust temporal stereo matching algorithm for accurate fire localization from multiple time-varying views using Color/Thermal data. Data generation with Fire/Smoke Simulation in Unity and validation on real-world data.

- Research on robust temporal stereo matching for dynamic fires
- Implementation and testing of the algorithm using Color/Thermal data
- Data generation using Fire/Smoke Simulation in Unity
- Validation of the algorithm on real-world data

[1]: Rosu et al., "Reconstruction of Textured Meshes for Fire and Heat Source Detection", SSRR 2019
[2]: Quenzel et al., "Beyond Photometric Consistency: Gradient-based Dissimilarity for Improving Visual Odometry and Stereo Matching", ICRA 2020

8) EDGE EXTRACTION FOR LIDAR ODOMETRY

Beneficial: ROS, C++, Python | Students: Bachelor/Master | Supervisor: Jan Quenzel



- Man-made environments contain many flat surfaces joined at edges.
- Edges provide motion constrains in two directions instead of only one for surfaces.
- LOAM's edge extraction [1] has been widely adopted, but classifies only a small subset (blue/red points).
- The aim is to robustly extract edges from LiDAR scans and to replace LOAM's extraction within LIO-SAM[2] to improve the accuracy of localization and mapping.

[1]: Zhang and Singh: "LOAM: Lidar Odometry and Mapping in Real-time", RSS 2014 [2]: Shan et al.: "LIO-SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping", IROS 2020

Own ideas are welcome!

Literature & Tutorials

Robotics: C++

See `http://www.cplusplus.com/doc/tutorial/` for a C++ tutorial.

Deep Learning: Python 3

See `https://docs.python.org/3/tutorial/` for a Python tutorial.

If you are familiar with another object oriented language, you will be fine.

For all robotics topics, we use the ROS framework: `http://www.ros.org/`.

We recommend the ROS tutorials for beginners:
`http://wiki.ros.org/ROS/Tutorials`

There will also be time during the lab to work through the tutorials.

Additional literature:

- Probabilistic Robotics - Dieter Fox, Sebastian Thrun, and Wolfram Burgard

For all deep learning topics, we use the PyTorch framework: `http://pytorch.org/`

See
`http://pytorch.org/tutorials/beginner/pytorch_with_examples.html`
for a tutorial. Notice that tutorial assumes familiarity with modern deep learning techniques.

There will also be time during the lab to work through the tutorials.

Seminar

Requirements for presentation and report:

- Fully understand the given paper, also look up relevant cited works.
- Present the proposed method in your own words.
- Discuss related work, i.e. why is this work better/different? Why is it not?
- Critically discuss the method. In your opinion, what are strengths, what are weaknesses? How could it be improved?

You will have at least one meeting with a supervisor during the lab period to discuss your paper and ask questions.

Output: 8 pages seminar report, seminar presentation (30 min)

UNIVERSITÄT BONN AIS

Selected works from the field of deep learning & robotics:

- Boyuan Chen et al. (2024). “Diffusion Forcing: Next-token Prediction Meets Full-Sequence Diffusion.” In: arXiv preprint arXiv:2407.01392
- Hidenobu Matsuki et al. (2024). “Gaussian splatting slam.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18039–18048
- Kan Chen et al. (2023). “Womd-lidar: Raw sensor dataset benchmark for motion forecasting.” In: arXiv preprint arXiv:2304.03834
- Fangqiang Ding et al. (2024). “RadarOcc: Robust 3D Occupancy Prediction with 4D Imaging Radar.” In: arXiv preprint arXiv:2405.14014
- Yinzhen Xu et al. (2023). “Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4737–4746
- Sammy Christen et al. (2023). “Learning human-to-robot handovers from point clouds.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9654–9664
- Ishika Singh et al. (2023). “Progprompt: Generating situated robot task plans using large language models.” In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 11523–11530
- An Dinh Vuong et al. (2024). “Language-driven Grasp Detection.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17902–17912
- Chenguang Huang et al. (2023). “Visual language maps for robot navigation.” In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 10608–10615

WHAT TO DO NOW?

If you intend to participate, please register in the Google form (see above).

For the **lab**: We will allocate you in groups of two. If you already have a lab mate, please submit your preferences together.

For the **seminar**: We will allocate your paper in two weeks (19.07.2024).

Thank you & see you soon!

http://ais.uni-bonn.de/SS24/Lab_Cognitive_
Robotics/lab.pdf

REFERENCES

- Chen, Boyuan et al. (2024). “Diffusion Forcing: Next-token Prediction Meets Full-Sequence Diffusion.” In: arXiv preprint arXiv:2407.01392.
- Chen, Kan et al. (2023). “Womd-lidar: Raw sensor dataset benchmark for motion forecasting.” In: arXiv preprint arXiv:2304.03834.
- Christen, Sammy et al. (2023). “Learning human-to-robot handovers from point clouds.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9654–9664.
- Ding, Fangqiang et al. (2024). “RadarOcc: Robust 3D Occupancy Prediction with 4D Imaging Radar.” In: arXiv preprint arXiv:2405.14014.
- Huang, Chenguang et al. (2023). “Visual language maps for robot navigation.” In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 10608–10615.

REFERENCES

Matsuki, Hidenobu et al. (2024). “Gaussian splatting slam.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18039–18048.

Singh, Ishika et al. (2023). “Progprompt: Generating situated robot task plans using large language models.” In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 11523–11530.

Vuong, An Dinh et al. (2024). “Language-driven Grasp Detection.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17902–17912.

Xu, Yinzhen et al. (2023). “Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4737–4746.