LATVIJAS UNIVERSITĀTE
DATORIKAS
FAKULTĀTE

ELEKTRONIKAS UN
DATORZINĀTŅU
INSTITŪTS

INSTITUTE OF
ELECTRONICS AND
COMPUTER SCIENCE

Development of Robot Cognition Through Applications of Machine Learning

Thesis progress report

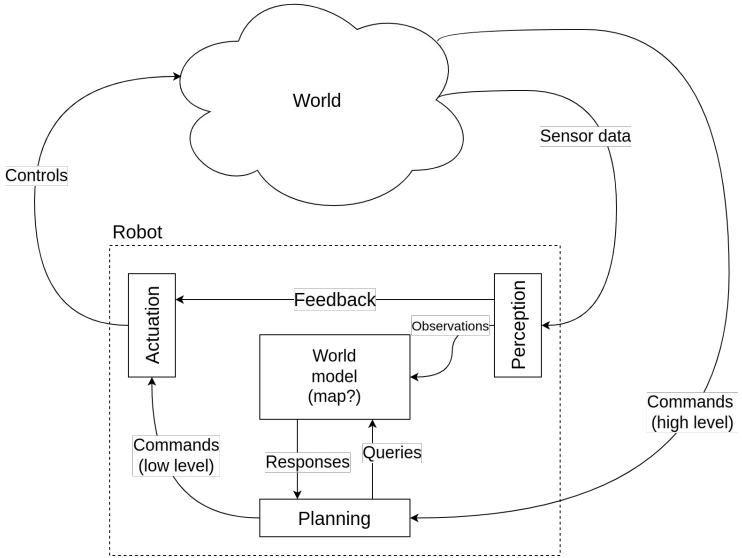
Author: Peteris Racinskis
Supervisor: Dr.sc.comp. Modris Greitans

Contents

- Topic description and motivation
- Personal background
- Stated plan for the first year
- Revised plan for the first year
- Work on prior / short term projects
  - Validation and refinement of a 6DOF object pose estimation system
  - Acoustic drone detection and direction finding system
- Current work – RoLISe / Edge AI task 4.1
  - System spanning two projects – concept overview
  - Review paper – Semantic Mapping
  - RoLISe 4.1 technical leadership

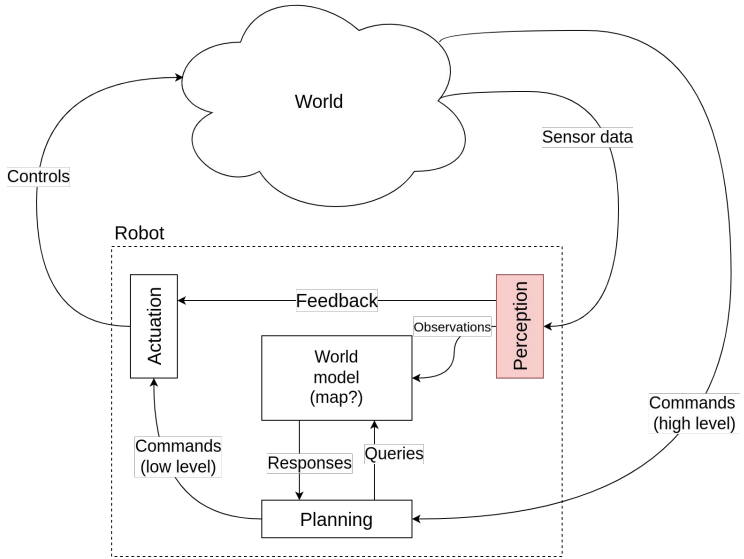
Topic

- Development of Robot Cognition through the Application of Machine Learning
  - “*Mašīnmācīšānās metožu pielietojums robotu kognitīvo spēju attīstīšanā*”
- Three main directions of study
  - Perception
  - Environment mapping / modelling
  - Planning
  - All exist on a continuum!



Topic – perception

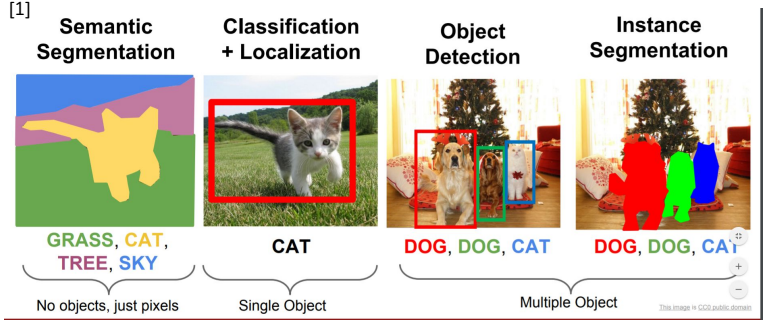
- Given raw sensor measurements, produce meaningful observations
- Visual inputs
  - Classification
  - Object detection
  - Segmentation
  - ...
- Other modalities
  - Radar
  - LiDAR
  - Audio
  - ...



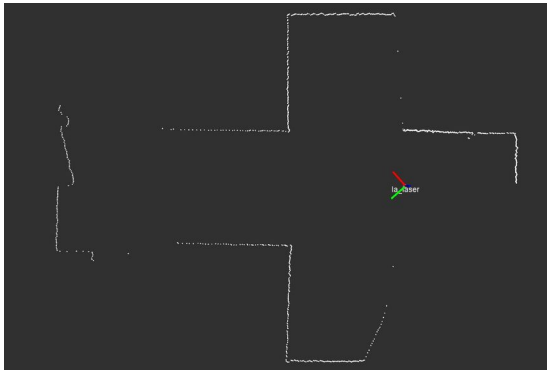
Topic – perception

- Given raw sensor measurements, produce meaningful observations
- Visual inputs
  - Classification
  - Object detection
  - Semantic segmentation
  - Instance segmentation
  - ...
- Other modalities
  - Radar
  - LiDAR
  - Audio
  - ...

[1]



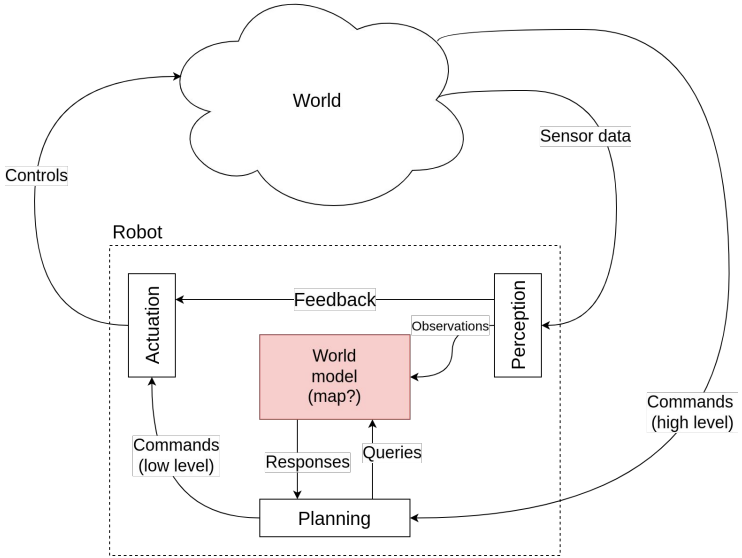
[2]



[1] https://data-flair.training/blogs/wp-content/uploads/sites/2/2020/09/segmentation-types.png
[2] https://www.researchgate.net/figure/2D-laser-scan-left-taken-with-a-Sick-LMS-100-Photo-credit-IST-TU-Graz_fig15_304987927

Topic – environment modelling

- Given meaningful observations, construct an *actionable* model of the environment
- What is the state of the world at a given place, time? Needed for planning
- Most common form - maps
  - Metric
  - Topological
  - Semantic
  - Implicit
- Many AI-based planning approaches still skip this step!



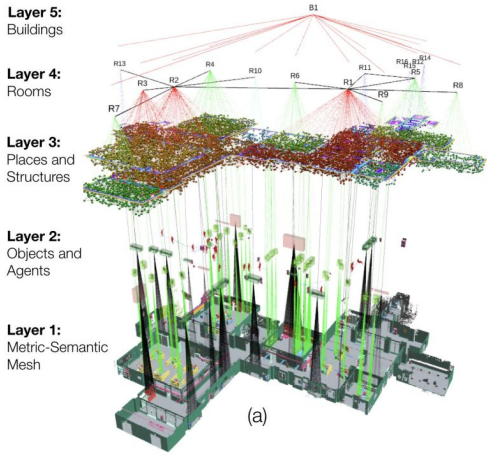
Topic – environment modelling

Forest LiDAR demo

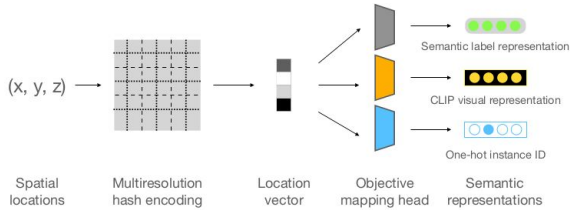
Topic – environment modelling

- Given meaningful observations, construct an *actionable* model of the environment
- What is the state of the world at a given place, time? Needed for planning
- Most common form - maps
  - Metric
  - Topological
  - Semantic
  - Implicit
- Many AI-based planning approaches still skip this step!

[1]



[2]

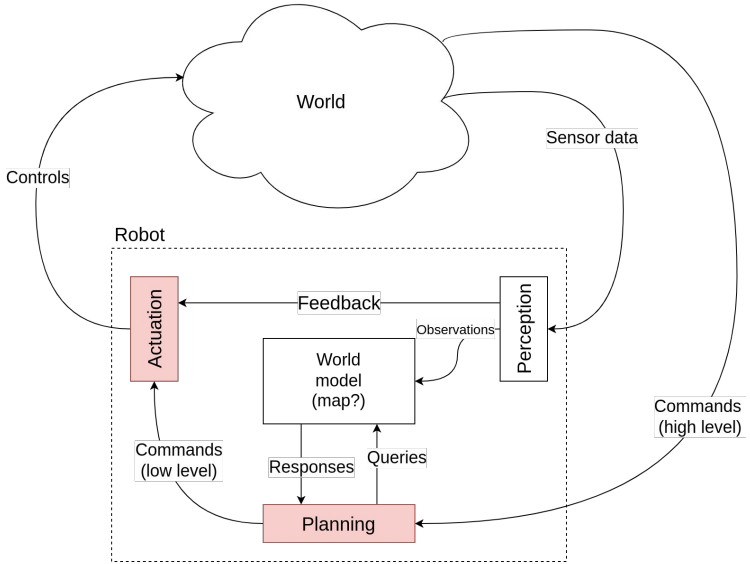


[1] Rosinol, Antoni, Arjun Gupta, Marcus Abate, J. Shi and Luca Carlone. "3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans." ArXiv abs/2002.06289 (2020): n. pag.

[2] Shafiullah, Nur Muhammad (Mahi), Chris Paxton, Lerrel Pinto, Soumith Chintala and Arthur D. Szlam. "CLIP-Fields: Weakly Supervised Semantic Fields for Robotic Memory." *ArXiv* abs/2210.05663 (2022): n. pag.

Topic - planning, actuation

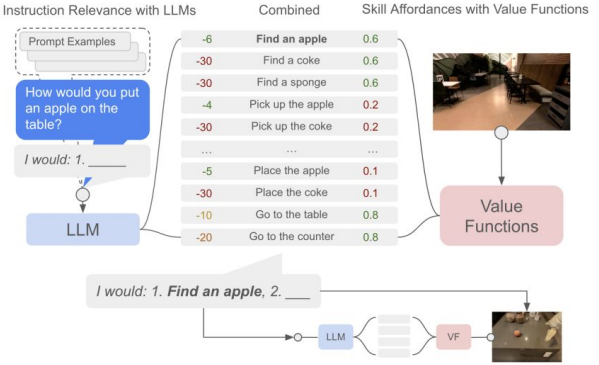
- Given commands and observations of the environment (possibly augmented by an internal world-model), produce a control output
- Classic approaches – find feasible / optimal trajectories in configuration space
- ML-based approaches
  - End-to-end learned policies
  - Motion primitives
- Planning vs control
  - Planner – construct sequence of commands or setpoints
  - Controller – physically drive the execution of commands, track setpoints through feedback



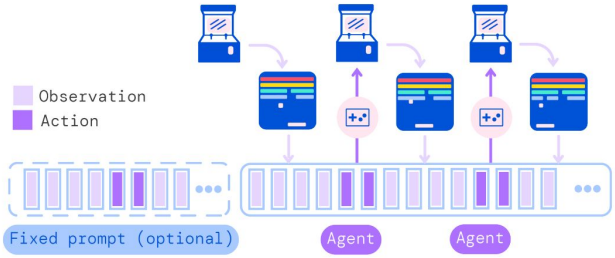
Topic - planning, actuation

- Given commands and observations of the environment (possibly augmented by an internal world-model), produce a control output
- Classic approaches – find feasible / optimal trajectories in configuration space
- ML-based approaches
  - End-to-end learned policies
  - Motion primitives
- Planning vs control
  - Planner – construct sequence of commands or setpoints
  - Controller – physically drive the execution of commands, track setpoints through feedback

[1]



[2]



[1] Ahn, Michael, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alexander Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil Jayant Joshi, Ryan C. Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego M Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, F. Xia, Ted Xiao, Peng Xu, Sichun Xu and Mengyuan Yan. "Do As I Can, Not As I Say: Grounding Language in Robotic Affordances." Conference on Robot Learning (2022).

[2] Reed, Scott, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley D. Edwards, Nicolas Manfred Otto Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar and Nando de Freitas. "A Generalist Agent." *ArXiv* abs/2205.06175 (2022): n. pag.

Personal background

- BEng Mechatronics Engineering – RTU, 2020
- MSc Computer Science – LU, 2022
- Research Assistant at EDI – 2022-2023
  - Robotics and Machine Perception Laboratory, Robotics group
- Researcher at EDI – 2023-
  - Robotics and Machine Perception Laboratory, Robotics group
- Areas of relative strength
  - Electrical and mechanical engineering
  - Software development (primarily in Python, but not constrained by languages)
  - Practical skills in machine learning
- Weaknesses, things to work on
  - Lack of mathematical background (working to fix this now)

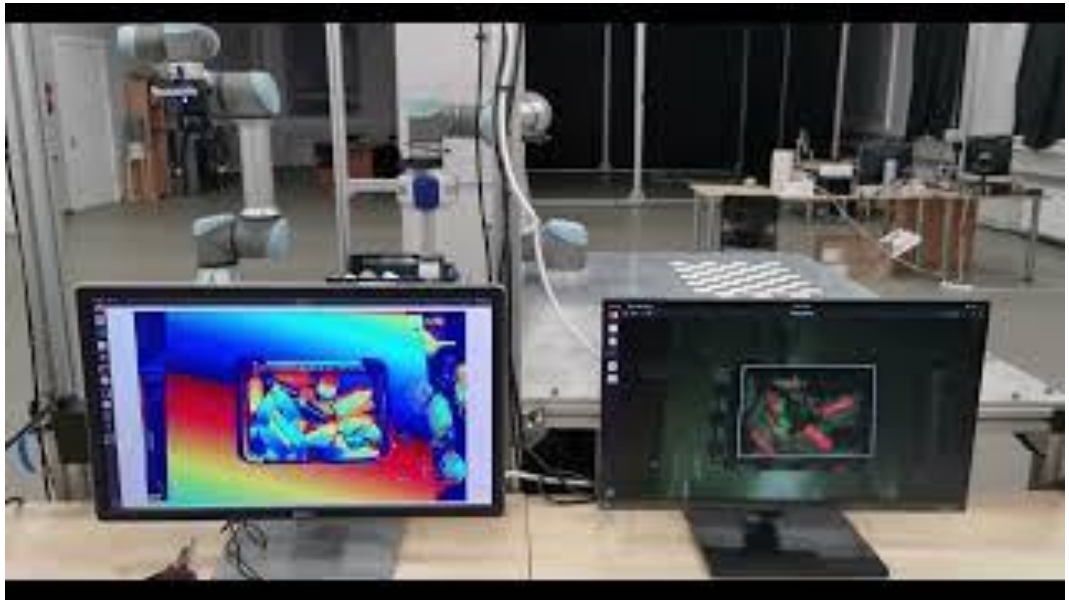
Plan - year 1 (revised)

- Mandatory courses – 8 credits
- Theoretical courses – pushed back to year 2 (took too long to negotiate)
- Research activities – 40+ credits
  - 6DOF pose estimation from RGBD data – validation, porting to FPGA, switch to direct inference of planar projection poses
    - paper at EDI conference
  - Semantic mapping overview article, technical specification – RoLISe + EdgeAI
  - Start working on technical implementation – RoLISe
  - Conference poster (drone detector)
  - Start drafting theoretical part of PhD thesis
- Educational activities – 2-4 credits (?)
  - Guest lectures – robotics seminar, image processing, deep learning(?)
  - Internship supervision
    - 1 x BSc mechanical engineering (done)
    - 1 x BSc CS (upcoming this summer)
  - Advising MSc theses in Intellectual Robotics (RTU)
    - NLP-conditioned control
    - Motion primitives for mobile manipulators

Plan - year 2 (revised)

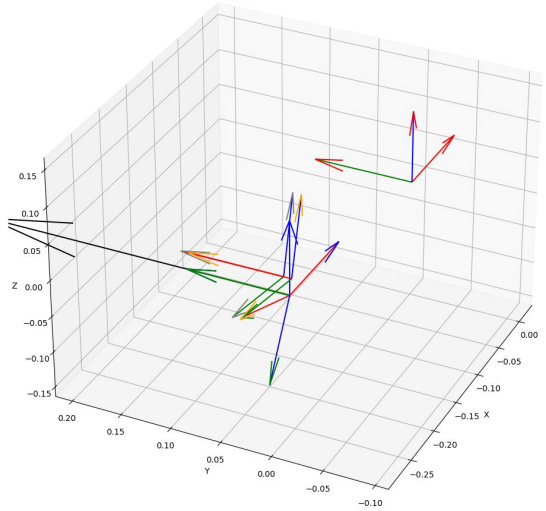
- Theoretical courses – mathematics
  - Year 2 - fundamentals (analysis, algebra)
  - Goal for Years 3, 4 - reach abstract algebra, topology
- Research activities
  - Technical specification and implementation – RoLISe + EdgeAI
  - Conference article (data set annotation), presented at ICARA 2024
  - Conference article planned (map construction)
  - Journal article planned (mapping system integration, overview of work done in RoLISe)
  - Journal article (publicly available data set)
  - Start drafting theoretical part of PhD thesis (some time)
- Educational activities – 2-4 credits (?)
  - Guest lectures – robotics seminar
  - Internship supervision
    - 1 x Programming technician (Gustavs Krasņikovs)
  - Qualification project supervision (Ārija Kalniņa)
  - Advising MSc thesis in GNSS + VIO integration (Andris Lapiņš)
  - Advising MSc theses in Intellectual Robotics (RTU)
    - NLP-conditioned control (T.E. Zinars)
    - Motion primitives for mobile manipulators (Oskars Vismanis)

6DOF pose estimation from RGBD, trained on synthetic data



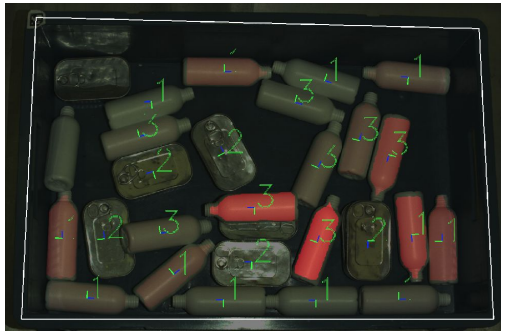
AI4DI – validation

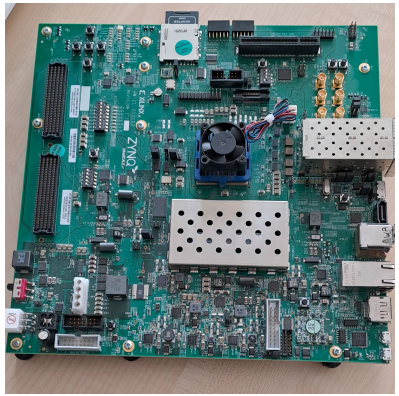
- Models trained on synthetic data
- Use synthetic ground truth to quantify model performance
- Validation pipeline
  - Generate evaluation data sets
  - Compute “optimal” grasps
    - not trivial when rotational symmetries involved!
  - Compare with system predictions
  - Devise and compute metrics of interest



AI4DI – refactoring, embedded execution

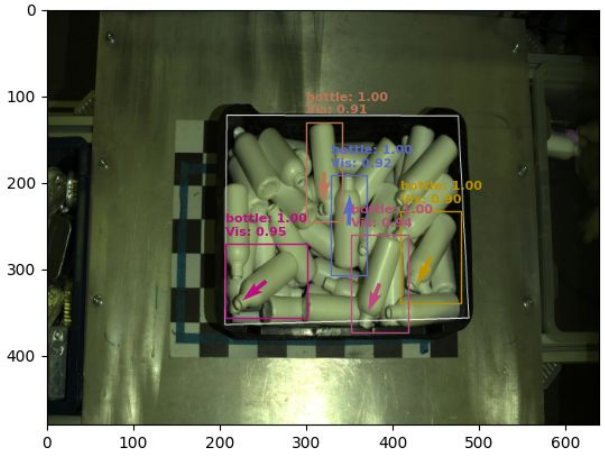
- Rewrite entire inference pipeline
  - Fragile filesystem-based “protocol” -> REST API
  - ~300x speedup of direction estimation algorithm through reimplementation
  - Allow pose estimation for multiple objects
  - Bring system to workable state by eliminating a long list of implementation flaws and overlooked edge cases
- Switch out inference back-end for embedded execution
  - MaskRCNN – infeasible on FPGA accelerator
  - Train SOLO model, implement inference server in C++





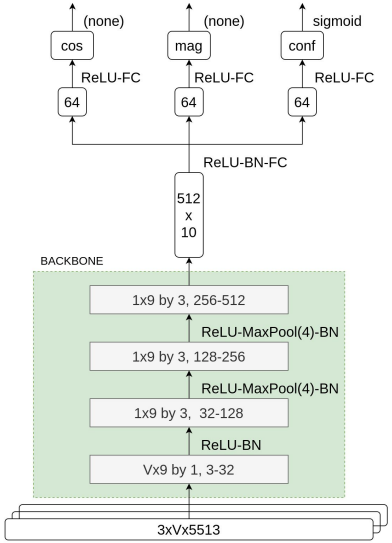
AI4DI – switch to direct prediction of planar poses

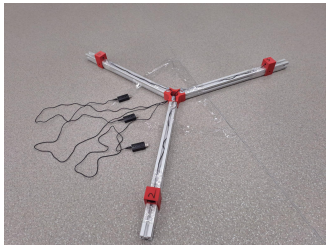
- Handcrafted direction finding algorithm – unreliable, reliant on features found in bottles
- After project end, system slated for use in Digital Innovation Hub
- Retrofit inference pipeline to directly predict object directions using a neural network
  - Base model – DETR
  - Add inference heads – direction, visibility
  - Sort outputs by visibility, backproject directly inferred directions and positions derived from bounding boxes
- Upcoming conference publication – June, EDI conference
- As far as I'm aware, this is a novel application of a neural network
- Idea, technical leadership - me; implementation – Andris Lapiņš

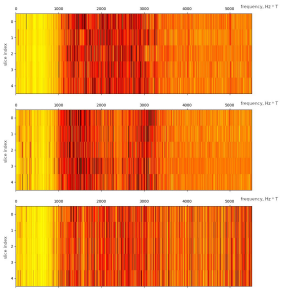


Audio drone detector

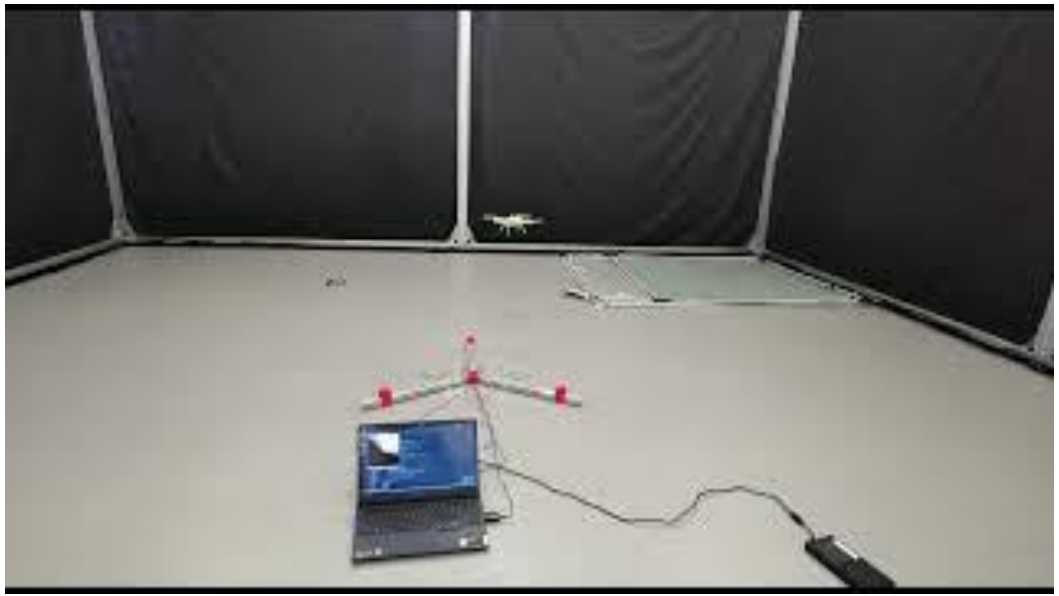
- Quick side project (~4 weeks of work)
- Original idea – at RTU defense makeathon 2022
- Sensors – array of 3 microphones
- Model – 2d-to-1d CNN
- Position ground truth data – motion tracking equipment
- Presented at DCOSS-IoT 2023 in Pafos, Cyprus

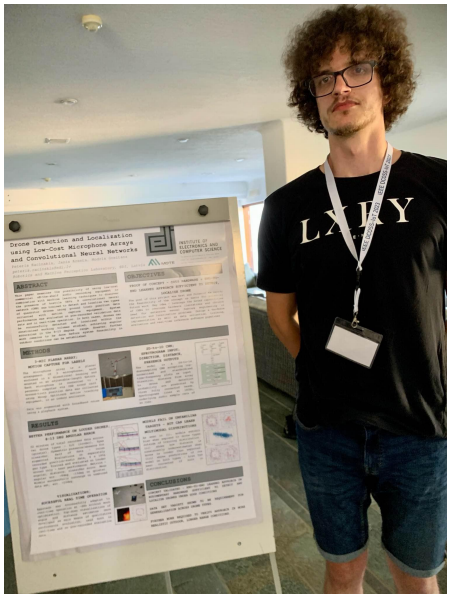




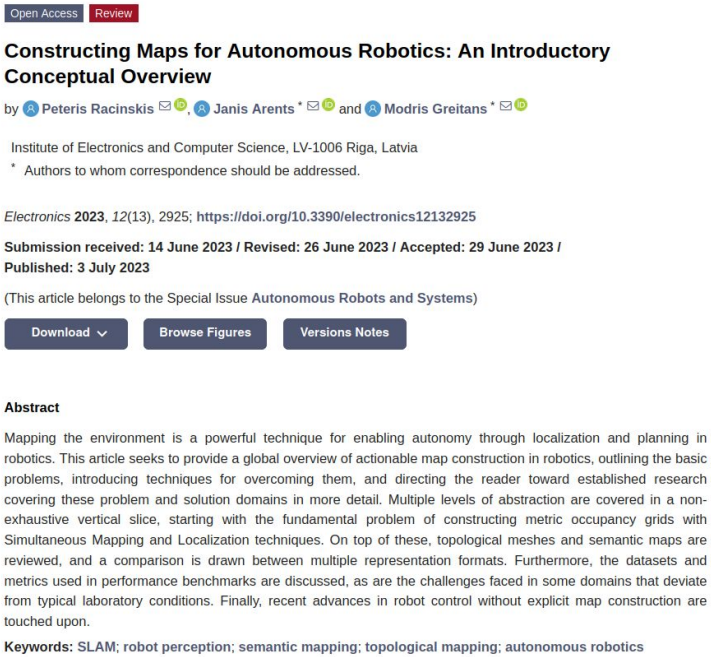


Drone detector





Survey article (promised last year)



Open Access Review

Constructing Maps for Autonomous Robotics: An Introductory Conceptual Overview

by Peteris Racinskis, Janis Arents * and Modris Greitans *

Institute of Electronics and Computer Science, LV-1006 Riga, Latvia

* Authors to whom correspondence should be addressed.

Electronics **2023**, *12*(13), 2925; **https://doi.org/10.3390/electronics12132925**

Submission received: 14 June 2023 / Revised: 26 June 2023 / Accepted: 29 June 2023 / Published: 3 July 2023

(This article belongs to the Special Issue **Autonomous Robots and Systems**)

Download Browse Figures Versions Notes

Abstract

Mapping the environment is a powerful technique for enabling autonomy through localization and planning in robotics. This article seeks to provide a global overview of actionable map construction in robotics, outlining the basic problems, introducing techniques for overcoming them, and directing the reader toward established research covering these problem and solution domains in more detail. Multiple levels of abstraction are covered in a non-exhaustive vertical slice, starting with the fundamental problem of constructing metric occupancy grids with Simultaneous Mapping and Localization techniques. On top of these, topological meshes and semantic maps are reviewed, and a comparison is drawn between multiple representation formats. Furthermore, the datasets and metrics used in performance benchmarks are discussed, as are the challenges faced in some domains that deviate from typical laboratory conditions. Finally, recent advances in robot control without explicit map construction are touched upon.

Keywords: **SLAM**; **robot perception**; **semantic mapping**; **topological mapping**; **autonomous robotics**

RoLISe / Edge AI 4.1

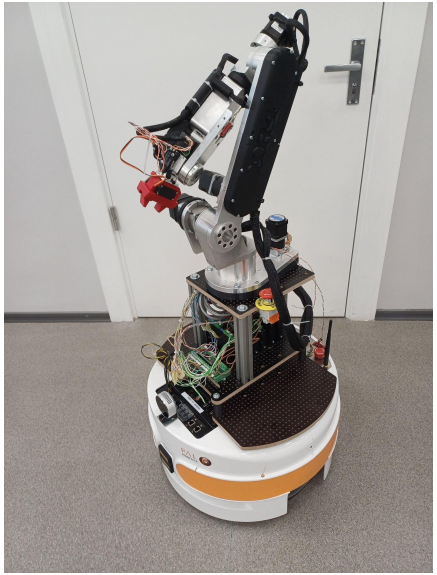
- Autonomous mobile robots, mobile manipulation
- RoLISe – part of VPP Fotonika / MOTE
  - Robotics, Internet of Things, Sensors
  - term – 2 years (end in Nov/Dec 2024)
  - focus of task 4.1 – perception, mapping framework
- Edge AI – EU-funded project
  - term – 3 years (end in Dec 2025)
  - focus of task 4.1 – NLP, planning, partner perception blocks, actuation
- Demonstrators
  - RoLISe – sensor suite and mapping framework, outdoor focus
  - Edge AI – autonomous agent using the above, potentially integrated with drone “scout”
- Role
  - RoLISe 4.1 - tech lead
  - Edge AI 4.1 - leadership of certain aspects

[1]

[1] https://robotnik.eu/wp-content/uploads/2021/02/home-gallery-mobile-manipulator-100121.jpg

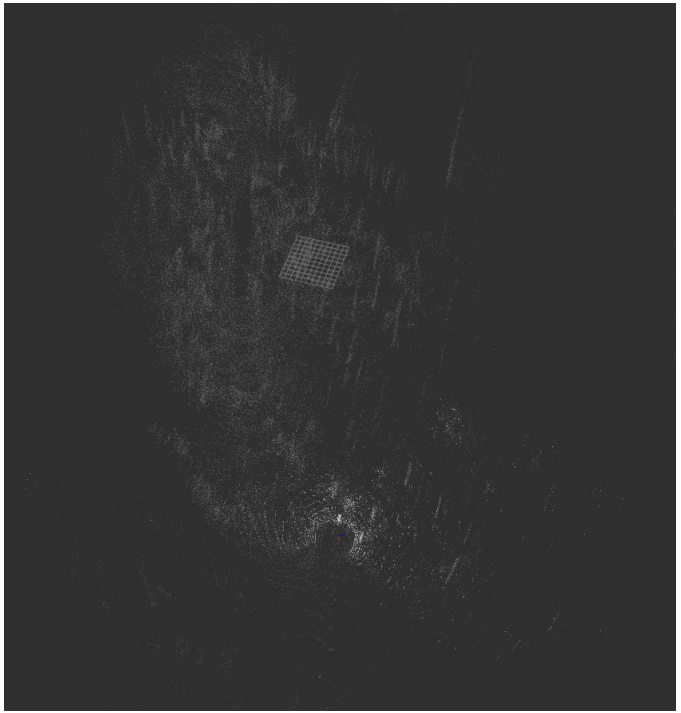
RoLISe / Edge AI 4.1

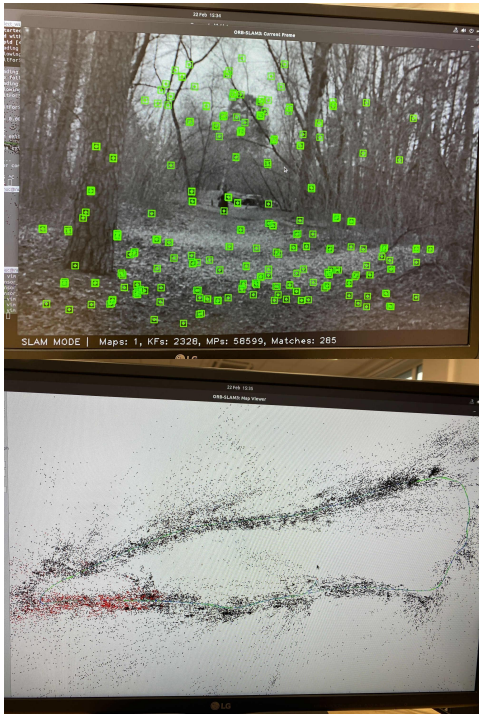
- Large task - have to delegate and cooperate
  - Map construction, software integration, data set collection, publication planning and writing - me
  - Localization - Andris Lapiņš
  - NLP queries - Toms Eduards Zinars
  - Planning primitives - Oskars Vismanis
  - Multi-agent exploration - Ārija Kalniņa
  - Mobile manipulator software integration - Rodions Saltanovs
  - Terrain segmentation - Guntis Bārzdiņš
- Assistance on other tasks
  - Vehicle re-identification in video data (Tomass Zutis, Gustavs Krasņikovs)
  - Synthetic data generation (Gustavs Krasņikovs)



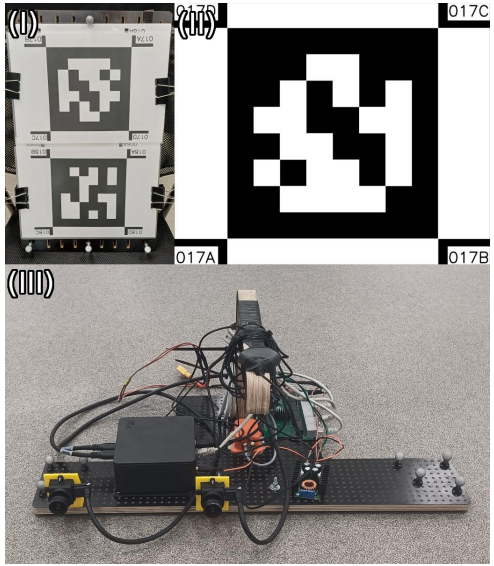
[1] https://robotnik.eu/wp-content/uploads/2021/02/home-gallery-mobile-manipulator-100121.jpg

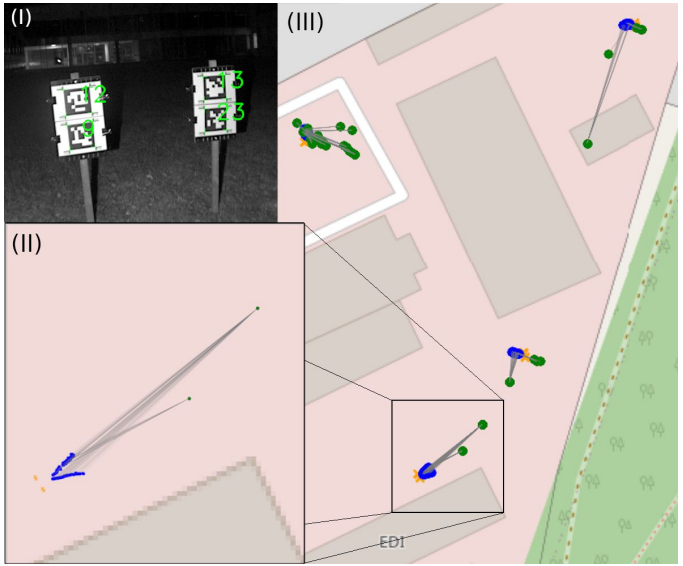
Implementation progress (RoLISe, EdgeAI)





Conference publication (data set annotation)





Thank you for your attention!