Robot Exploration in Unknown Environments

rijksuniversiteit groningen

Author: Ivo Gorgias Meijers, S3810313

1st supervisor: prof. dr. ir. M. (Ming) Cao
Daily supervisor (PhD): B. Yu
2nd supervisor: dr. A.J. (Albert) Bosch

Table of Contents

1. Introduction

Robot exploration (RE) of unknown indoor environments (UIE) has become increasingly popular due to the emergence of indoor applications of mobile robotics. (Cadena et al., 2016) RE methods are used in different (indoor) applications, such as search and rescue applications (Niroui, 2019), human assistance in crowded areas (Mammolo, 2019) and fire searching (Marjovi et al., 2021). Moreover, RE methods have developed significantly over the last decade as they are able to operate in more complex environments for a longer period of time. (Kunze et al., 2018). Thereby, many new applications, new sensors and new computational looks have been developed (Cadena et al., 2016)

In the field of RE, multiple different exploration strategies exist, i.e., different paradigms, methods, and software. Though, all strategies aim for the same goal: construct a model of the UIE efficiently and/or reach a certain goal or perform a task (Cadena et al., 2016). Maps are useful for two reasons; (i) for path planning and to provide an intuitive visualisation for a human operator and (ii) to allow limiting the error committed in estimating the state of the robot (Cadena et al, 2016).

Moreover, the popularity of implementing reinforcement learning (RL), a subpart of machine learning (ML), has increased (Niroui et al., 2019). RL is characterised by computer learning in sequential decision-making problems in which there is limited feedback (Otterlo & Wiering, 2012). RL approaches have two limitations in RE of UIE, namely that state features need to be handcrafted and that it suffers from the curse of dimensionality (Arulkumaran, 2017). A more advanced proposed approach of RL is deep reinforcement learning (DRL). Contrary to RL, DRL learns state features automatically and dimensionality can be reduced through iterative interactions with the environment (Arulkumaran, 2017). Implementing DRL in RE of UIE, is still in a developing phase. (Niroui et al., 2019)

According to Cadena et al. (2016), RE is entering a new era, called *robust-perception age*, in which challenges await. These challenges are among others *robustness* and *time-efficiency*, which will be the two main focuses of this project. *Robustness (*or *robust performance)* entails that the RE method can operate in noisy environments (Cadena et al., 2016). Time-efficiency entails that the robot can explore as much as possible of the UIE in as little time as possible. (Cadena et al., 2016)

Due to the above-mentioned challenges and based on the state-of-the-art RE policies, a chance to design a RE policy based on DRL arises. Therefore, the main objective in this research is to design a RE policy based on DRL. Thereby, the challenges of robustness and time-efficiency will be analysed by putting the to-be-designed RE policy in comparison with the traditional RE policies.

2. Stakeholder analysis

In RE two stakeholders could be distinguished, namely:

- Discrete Technology and Production Automation (DTPA); the DTPA is the main stakeholder in this project, as it is the problem owner, and therefore has high interest in this project. The DTPA has also relatively high power in this project, as they can intervene if they desire.
- Overall users of RE; this stakeholder entails all people that make use of RE. They could be interested in this project since a new policy of RE will be designed. Other users of RE have very low power in this project, as they are not involved in the design of the new RE policy. However, these users do have interest in this project, as it contributes to the development of RE.

No competitors are considered in this project, as most of the RE libraries are open source. (Cadena et al., 2016)

3. System description

The robot will examine the environment by 2D-LIght Detection And Ranging (2D-LiDAR) sensors, which are able to analyse contours of an environment (Vasilii et al., 2018). Next to 2D-LiDAR, the *odometry* of the robot will be used as an input. Odometry entails the use of motion sensors to assess the robot its movement in space. (Vasilii et al., 2018) The combination of the input of the 2D-LiDAR sensors and the input of odometry sensors could be considered as the observation of the robot. Based on the current observation, Simultaneous Localization and Mapping (SLAM) is used to construct the environment map and to calculate the current location of the robot on the map. (Chaplot et al., 2020)

Based on the given map so far and the location of the robot, the robot follows a certain *exploration policy* to determine its next movement to the most informative destination. The design of this exploration policy is the main objective of this project. For the determination of a robot its next movement, there will be made use of a Markov Decision Process (MDP) as described by Otterlo & Wiering (2012). This decision will lead to a movement of the robot within the environment. Then the robot will receive new observations from the environment by its sensors and therefore a feedback loop will be created. An overview of this RE system is given in *figure 1*:

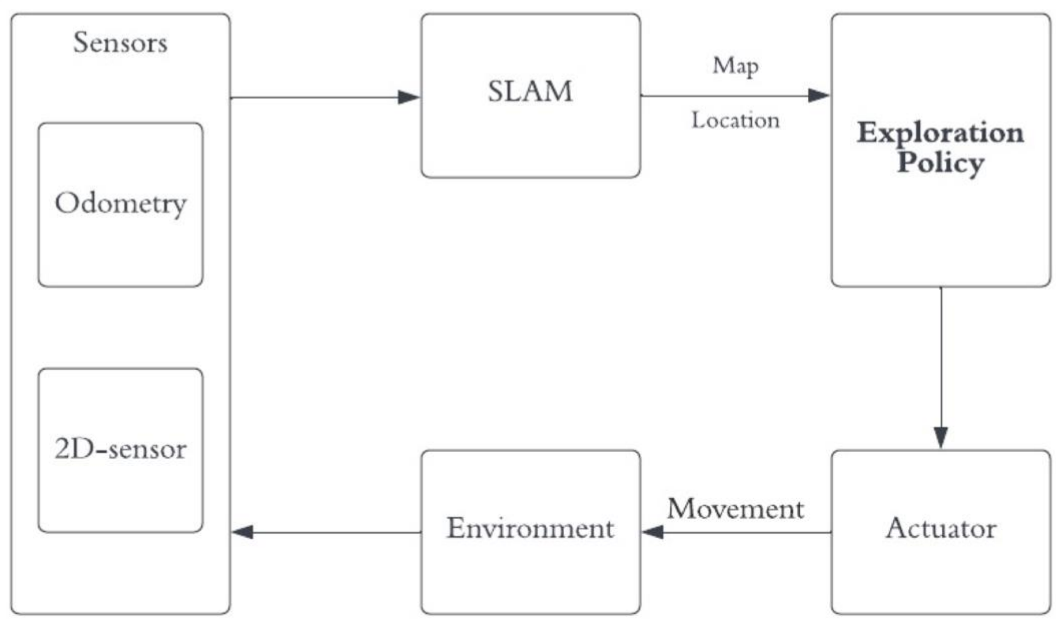

Figure 1 - System overview of the RE method

This project will merely focus on UIE, which means that no prior knowledge is known in any way. This excludes both manually built maps of artificial beacons in the environment and GPS systems (the GPS satellites can be considered as moving beacons at known locations) (Cadena et al., 2016). Thereby, only indoor environments will be considered since this is a demarcated area and GPS has little function within indoor areas (Cadena et al., 2016).

4. Research problem

Robot exploration in *unknown indoor environments* using *reinforcement learning* is still in a development phase. Thereby, challenges arise when entering the *robust-perception age* regarding robustness and time-efficiency.

5. Goal statement

The main goal of this design science project is to design an RE policy based on (deep) reinforcement learning. Thereafter, the designed RE policy will be assessed on the robustness and time-efficiency, which will be compared with the traditional RE policies. Ultimately, the designed RE policy will be implemented in the real world.

These objectives result in the following goal statement:

This research aims to design and implement a robot exploration policy based on (deep) reinforcement learning which can explore unknown indoor environments while taking robustness and time-efficiency into consideration, within a time span of 17 weeks.

This goal is considered as SMART, as it specifically aims at designing a new RE policy, measurable due to the degree of time-efficiency and robustness, attainable due to the broad variance of existing literature, relevant due to the contribution to existing RE methods and policies, and time-bound due to the time-restriction of 17 weeks. The creation of SMART goals is based on Williams (2012).

6. Main research question

How to design and implement a *robot exploration policy* based on *reinforcement learning* which explores *unknown indoor environment*?

6.1 Research sub-questions

I. What causes the challenge of robustness?
   a. How to analyse and measure the challenge of robustness?

II. What causes the challenge of time-efficiency?
   a. How to analyse and measure the challenge of time-efficiency?

The first two sub-questions contribute to the main research question by better understanding the challenges which come with the *robust-perception age*. First, the roots of these challenges are examined for better understanding the actual problem, followed by methods on how to analyse these challenges.

III. How to design a RE policy based on RL?

The third sub-question contributes to the main research question by investigating how a RE policy based on RL could be designed.

IV. How does the to-be-designed RE policy compare to traditional RE policies?

The fourth sub-question contributes to the main research question by putting the to-be-designed RE policy in perspective with other traditional RE policies. This will firstly help improving the to-be-designed RE policy, as useful elements from traditional RE policies could be used. Moreover, this clarifies on which elements the to-be-designed RE policy differentiates from traditional RE policies.

V. How to implement the to-be-designed RE policy in the real world?

The fifth sub-question contributes to the main research question as it examines how the to-be-designed RE policy could be implemented in the real world.

7. Research strategy

Firstly, this research project is in-depth research, as its aim is to yield knowledge on a specific niche subject, namely RE (based on RL). However, the project will begin broader by examining challenges of robustness and time-efficiency.

Secondly, this research project includes both qualitative as quantitative parts. On the one hand, qualitative research is done by means of literature research about traditional RE methods and robustness and time-efficiency. Moreover, the system-thinking approach of programming the to-be-designed RE policy could be considered as qualitative. On the other hand, this project will be somewhat quantitative, as the programming will include data analysis and data science.

Thirdly, the main focus and goal of this project is considered as empirical, as a RE policy should be designed. The implementation of this policy in the real world is even more empirical. However, in order to arrive at this stage; in the first stage desk research should be done in order to analyse robustness and time-efficiency. Moreover, desk research about traditional RE policies should be done, in order to be compared with the to-be-designed RE policy.

8. Methods and Tools

In this project, data will be both gathered and generated. On the one hand, data about different RE methods will be gathered. This includes not only information about how these methods are built up and work, but also the actual coding itself. Thereby, information about robustness and time-efficiency is gathered. On the other hand, data will be generated by means of designing a RE policy. This data generation will be based on the data gathered beforehand.

Extensive literature, modelling programmes and programming language are available and thereby literature is provided by the project its supervisor preliminarily and intermediately.

The design of the to-be-designed policy will be programmed in Python. The simulation of this policy will be done in Gazebo. Finally, the software which will be used to implement the to-be-designed policy in the real world is Robot Operating System (ROS).

9. Deliverable

The deliverable of this research project is an RE policy based on reinforcement learning and a comparison on robustness and time-efficiency of the to-be-designed RE policy with traditional RE policies.

9.1 Validation

Internal validation will be achieved by gathering information about traditional RE methods and its corresponding coding and robustness and time-efficiency. External validation will be reached by designing an RE policy based and first simulating this method and subsequently implementing this policy in a real-world environment.

10. Risk analysis and feasibility

The main risk of this research project is the time limit. Firstly, this project is rather challenging as it implies both extensive and in-depth research, which will preoccupy much time. Moreover, this project is empirical and practice oriented. Designing the RE policy will probably take much time.

The feasibility of reaching all the research objectives in this project is uncertain, as the time limit constraint might cause the final object, implementation in a real-world environment, to not be achieved.

11. Planning

A preliminary planning is made, which is built up in four different stages:

Stage 1: Orientation stage which will be used to investigate traditional RE methods. Moreover, this stage will be used to get familiar with modelling software and programming languages. This stage will take approximately 3 to 4 weeks.

Stage 2: Deployment stage which will be used to get familiar with the modelling and programming of traditional RE methods in the simulation environment. This stage will take approximately 1 to 2 weeks.

Stage 3: Learning-based exploration stage in which learning based methods and policies will be examined and designed. This design will be compared to traditional policies in this stage as well. This stage will take approximately 5 to 8 weeks.

Stage 4: Implementation stage in which the designed policy will be implemented in a real-world environment. This stage will take approximately 3 to 5 weeks.

Below, the project's planning is depicted in a Gantt chart:

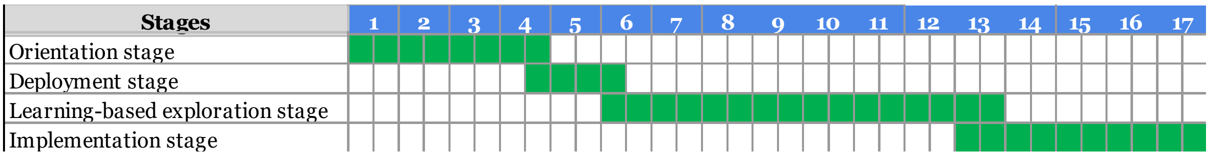

Figure 2 - Gantt chart of research planning

12. References

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26-38.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., ... & Leonard, J. J. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, *32*(6), 1309-1332.

Chaplot, D. S., Gandhi, D., Gupta, S., Gupta, A., & Salakhutdinov, R. (2020). Learning to explore using active neural slam. arXiv preprint arXiv:2004.05155.

Keidar, M., & Kaminka, G. A. (2014). Efficient frontier detection for robot exploration. The International Journal of Robotics Research, 33(2), 215-236.

Kunze, L., Hawes, N., Duckett, T., Hanheide, M., & Krajník, T. (2018). Artificial intelligence for long-term robot autonomy: A survey. IEEE Robotics and Automation Letters, 3(4), 4023-4030.

Mammolo, D. (2019). Active slam in crowded environments (Master's thesis, Autonomous Systems Lab, ETH Zurich).

Marjovi, A., Nunes, J. G., Marques, L., & De Almeida, A. (2009, October). Multi-robot exploration and fire searching. In 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 1929-1934). IEEE.

Niroui, F., Zhang, K., Kashino, Z., & Nejat, G. (2019). Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments. IEEE Robotics and Automation Letters, 4(2), 610-617.

Otterlo, M. V., & Wiering, M. (2012). Reinforcement learning and markov decision processes. In Reinforcement learning (pp. 3-42). Springer, Berlin, Heidelberg.

Ramakrishnan, S. K., Jayaraman, D., & Grauman, K. (2021). An exploration of embodied visual exploration. *International Journal of Computer Vision*, *129*(5), 1616-1649.

Vasilii, K., Vladimir, A., Andrey, P., & Vera, K. (2018). The lidar odometry in the slam. Proceedings of the Xxth Conference of Open Innovations Association Fruct, 23, 180–185. Retrieved March 24, 2022, from INSERT-MISSING-URL.

Verschuren, P., Doorewaard, H., & Mellion, M. (2010). *Designing a research project* (Vol. 2). The Hague: Eleven International Publishing.

Williams, C. (2012). MGMT (5th ed.). USA: South- Western College Publishing.