Reinforcement Learning for Autonomous Exploration of Cluttered Environments using Collision-resilient Aerial Robots

Overview: This thesis aims to develop a reinforcement learning approach applicable to continuous state and action spaces and partially observed Markov decision processes, wherein the reward for the agent is contingent upon the agent's efficient exploration of the environment. The method should be able to utilize a sliding window of LiDAR (and/or camera) observations in order to identify reference paths that a) maximize the anticipated exploration reward, while b) minimizing the likelihood of a collision and necessarily ensuring that if a collision is bound to happen then the kinetic energy at contact-time is such that the collision-tolerant robot can remain safe. To enable robust performance in very long-term and large-scale deployments we seek to develop a method that does not assume a consistent online reconstructed 3D map of the environment but rather aim to encode the ability to "develop map memory" through recurrent networks. The reinforcement learning policy will work in combination with a) a multi-modal localization pipeline allowing for the robust pose estimation of the robot, as well as b) an onboard Model Predictive Control strategy either up to the velocity or up to the position control level (design choice to be made during the project). The method is to be developed with the combination of simulation and real-life data and it is aimed to be tested with the "Gagarin" collision-resilient aerial robot (depicted in the Figure on the right) developed in the framework of the DARPA Subterranean Challenge.

Tasks and Sub-objectives

- Literature review - understanding of Q-Learning/Q-Learning in Continuous State/Action Spaces, study of the Deep Deterministic Policy Gradients method.
- Modeling of the Exploration Planning Problem - understanding of the application of Q-Learning for Partially Observable Markov Decision Processes - understanding of the Recurrent Deterministic Policy Gradients method.
- Modeling of the Gagarin collision-tolerant robot in simulation (prior model in Gazebo available).
- Research to propose Policy, Objective function approximation and Network design for Reinforcement Learning-based Resilient Exploration based on Collision-tolerant Aerial Robots.
- Derivation of Combined Simulation- and Experiments-based Dataset for the Autonomous Exploration of Subterranean Settings (e.g., caverns/mines) and other constrained environments.
- Implementation on the Gagarin Aerial Robot and Evaluation in real-life Field Experiments.

Starting Literature

[1] [1] Russell Reinhart, Tung Dang, Emily Hand, Christos Papachristos, and Kostas Alexis, "Learning-based Path Planning for Autonomous Exploration of Subterranean Environments", IEEE International Conference on Robotics and Automation (ICRA) 2020, May 31 - June 4 2020, Paris, France.

[2] [2] Mina Kamel, Thomas Stastny, Kostas Alexis, Roland Siegwart, "Model Predictive Control for Trajectory Tracking of Unmanned Aerial Vehicles Using ROS", Springer Book on Robot Operating System (ROS) – The Complete Reference (Volume 2)

[3] [3] Mihir Rahul Dharmadhikari, Tung Dang, Lukas Solanka, Johannes Brakker Loje, Dinh Huan Nguyen, Nikhil Vijay Khedekar, and Kostas Alexis, "Motion Primitives-based Path Planning for Fast and Agile Exploration using Aerial Robots", IEEE International Conference on Robotics and Automation (ICRA) 2020, May 31 - June 4 2020, Paris, France.

[4] [4] Wierstra, D., Foerster, A., Peters, J. and Schmidhuber, J., 2007, September. Solving deep memory POMDPs with recurrent policy gradients. In International Conference on Artificial Neural Networks (pp. 697-706). Springer, Berlin, Heidelberg.

[5] [5] Sutton, R.S. and Barto, A.G., 2018. Reinforcement learning: An introduction. MIT press.

[6] [6] Bertsekas, D.P., 2019. Reinforcement learning and optimal control. Athena Scientific.