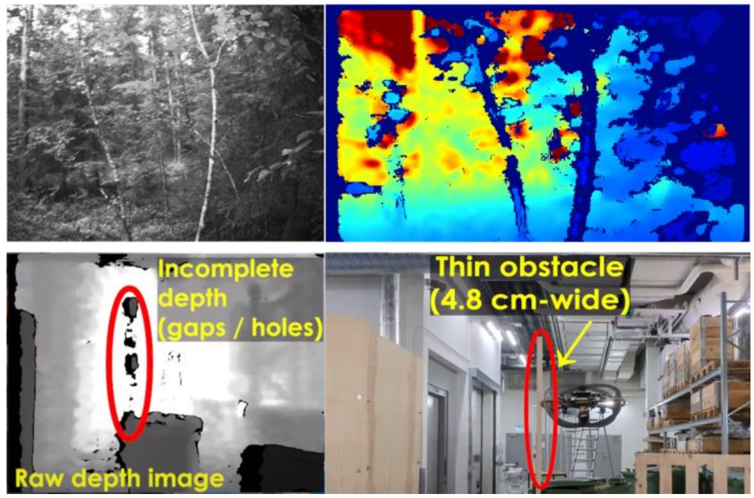
# Multi-modal learning-based navigation


**Abstract**:




Agile navigation has been demonstrated using only a lightweight sensor suite including RGB-D cameras and IMU sensor, harnessing the power of deep learning to directly map noisy sensor observations to low-level command output in a low-latency manner. Nevertheless, existing works often utilize either visual or depth data only. While visual data can offer rich information about the environments, navigation policies trained with visual image input only can suffer from generalizability issue when deployed on the real system, thus requiring extensive domain adaptation/randomization or appropriate intermediate abstract representation [1]. Depth images can be used as an abstract input representation for navigation task [2,3], allowing successful sim-to-real transfer in multiple scenarios. However, depth cameras can miss important details in the environment, such as thin obstacles or shiny surfaces (see above pictures), hence using depth camera alone is not sufficient in many real-world situations. This project and thesis aim to address this challenge by utilizing both visual and depth image inputs to train a collision-free navigation policy for aerial robots.


**Tasks:**

- Study and understand the basic deep learning/aerial robot's dynamics knowledge.
- Setup the simulation environment with simulated (noisy) depth and RGB image inputs.
- Familiarize yourself with the current learning-based navigation methods of the lab [5,6].
- Collect real/simulated visual/depth data and train two VAEs [4] to compress both visual and depth data
- Train and test a navigation policy using the compressed latent space in simulation
- Deploy on a real quadrotor platform (with NVIDIA Jetson GPU boards).

**Literature (indicative):**


[1] B. Zhou, P. Krähenbühl, and V. Koltun, "Does computer vision matter for action?" Sci. Robotics, vol. 4, no. 30, 2019.

[2] A. Loquercio, E. Kaufmann, R. Ranftl, M. Muller, V. Koltun, and D. Scaramuzza, "Learning high-speed flight in the wild," in Science Robotics, October 2021.

[3] D. Hoeller, L. Wellhausen, F. Farshidian, and M. Hutter, "Learning a state representation and navigation in cluttered and dynamic environments," IEEE Robotics and Automation Letters, vol. 6, no. 3, pp. 5081–5088, 2021.

[4] C. Doersch, "Tutorial on variational autoencoders," arXiv preprint 2016.

[5] H. Nguyen, S. H. Fyhn, P. De Petris, and K. Alexis, "Motion primitives based navigation planning using deep collision prediction," in 2022 (ICRA). IEEE, 2022, pp. 9660–9667.


**Relevant Project information**


- **Digiforest & SENTIENT projects,zFunding Agencies:** European Commission, Research Council of Norway


**Main supervisor:** Kostas Alexis, Professor, NTNU | **Co-supervisor:** Mihir Kulkarni, PhD Candidate, NTNU