Title: A High-fidelity Multimodal Synthetic Dataset Generation Framework for Off-road Unstructured Terrain Navigation Training of Autonomous Robots

Author Names: Liyana Wijayathunga, Dulitha Dabare, Alexander Rassau, Douglas Chai & Syed Mohammed Shamsul Islam

Abstract: The success of deep learning methods in a wide range of application areas has inspired many recent developments in the urban and off-road autonomous navigation domain. In particular, techniques for semantic scene understanding, a key aspect of the navigation pipeline, have been researched extensively, resulting in many real-world and synthetic datasets. However, in comparison to urban semantic segmentation datasets, the availability of datasets for off-road environments remains sparse. In this paper, we aim to overcome this challenge by introducing a methodology capable of efficiently generating photorealistic synthetic datasets for off-road environments with support for multiple sensor modalities. The developed approach is evaluated by presenting a synthetic multimodal dataset of four Western Australian outdoor scenarios suitable for ground robot navigation testing and training. This dataset contains monocular and stereo RGB image and LiDAR point cloud data, with respective semantic annotations covering 16 classes, along with IMU sensor data. The simulation of the four environments was carried out using the NVIDIA Isaac Sim robotics simulator platform. Experiments were conducted using state-of-the-art off-road and universal semantic segmentation models trained on the generated dataset to highlight the Sim2Real capabilities of photorealistic synthetic datasets in real-world off-road robot navigation applications and the limitations of existing datasets for generalised use.

https://link.springer.com/article/10.1007/s10846-025-02340-2