Geometric Structural Awareness From Monocular Visual SLAM for Low-Cost Robotic Navigation

Samuel, Tristan (School: Good Hope Country Day School)

Autonomous robots rely on Simultaneous Localization and Mapping (SLAM) to build maps and estimate their position within unknown environments. While many modern robotic systems achieve high reliability using additional sensors such as LiDAR or dedicated depth cameras, these components significantly increase system cost and limit accessibility. Monocular visual SLAM systems use only a single camera and can estimate device pose and generate sparse 3D maps, but they do not directly provide the proximity and structural awareness information required for safe navigation. This project investigates whether outputs from a monocular SLAM system can be post-processed to generate meaningful device-centric distance measurements to surrounding surfaces and obstacles. Code was written to improve the results of open-source monocular visual SLAM systems. Video data collected from a smartphone camera was processed using ORB-SLAM3 to obtain camera trajectories and sparse point clouds. Depth estimation was enhanced using an AI model, and RANSAC plane fitting was applied to identify dominant structural surfaces. Remaining map points were clustered to approximate obstacle regions. Perpendicular distances from the camera to detected planes and clusters were computed and compared to ground-truth measurements. Although not yet as precise as LiDAR-based systems, given that these additional systems were able to accurately detect and distance walls, provide analytics about obstacles and best paths in real time, this approach suggests that low-cost camera-only platforms can provide meaningful structural awareness for navigation in structured environments, potentially expanding accessibility of autonomous robotic systems.

Awards Won: