

Semantic SLAM with Quality-adaptive Properties

Jianhao Zheng

Professor : Alcherio Martinoli

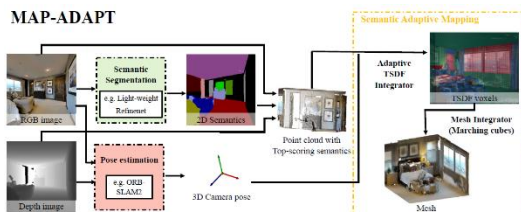
Assistant(s) : Lucas Wälti

Creating 3D semantic reconstructions of the world around us is paramount for many applications, including those on autonomous agents. However, we do not always need everything in anything. The ability to reconstruct the world in a quality-adaptive manner can reduce computation, storage, and overall cost. Although prior work has explored the creation of geometry-guided quality-adaptive maps, very few examine how to use semantic information. I propose MAP-ADAPT, a real-time approach for quality-adaptive semantic 3D reconstruction using RGBD frames. MAP-ADAPT comprises the first adaptive semantic 3D mapping algorithm that generates directly a single map with regions of different quality based on the semantic information. I include semantic SLAM capability and evaluate end-to-end the adaptive semantic reconstruction of indoor spaces on two state-of-the-art 3D semantic datasets.

I evaluate MAP-ADAPT on creating accurate and complete geometric and semantic 3D maps with adaptive resolution, and compare with the fixed voxel size Voxblox at different resolution levels, as well as with Panoptic Multi-TSDFs, the only other method that creates individual object instance maps of different resolutions leveraging semantic information. One of the evaluation datasets is Matterport3D.



Left: reconstructed map on the MP3D dataset. Right: Colors in voxels denote the three quality levels, where red is high, green is middle, and blue is coarse.

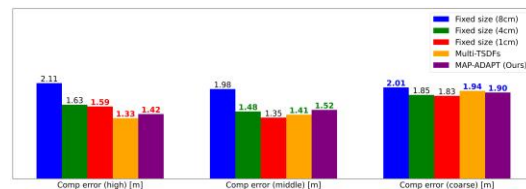


System overview of MAP-ADAPT

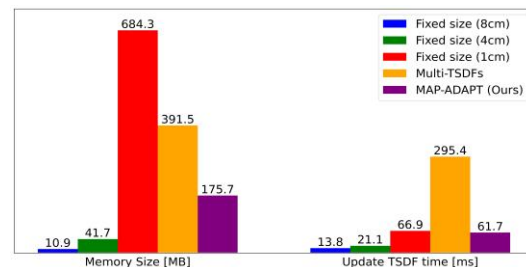
Given a set of RGB-D frames, MAP-ADAPT uses the RGB images and depth images to estimate the camera pose and predict the semantic segmentation map with RGB images only. It employs this information to create a quality-adaptive map in an online manner. The main contribution of this project is the component of semantic adaptive mapping, which is built on Voxblox, a real-time TSDF-based mapping system which can run on a CPU-only platform. In addition to the original Voxblox, I mainly implemented the following two additional features:

- 1) Semantic SLAM capability incorporation
- 2) Reconstruct regions in different quality depending on their semantics.

MAP-ADAPT, when compared to baselines, provides a lightweight semantic 3D map that is comparable in geometric accuracy to using a fixed-sized map for certain regions. As regards to Multi-TSDFs, our method performs similar in the geometric metrics while taking far less computational time and memory.



Completion error per method on the MP3D dataset



Running time and memory on the MP3D dataset