Online Semantic Mapping for Autonomous Navigation and Scouting

Daniel Maturana*, Sankalp Arora*, Po-Wei Chou*, Dong-Ki Kim*, Masashi Uenoyama†, Sebastian Scherer*

*Robotics Institute, Carnegie Mellon University
{dmaturan, asankalp, poweic, dkkim, basti}@andrew.cmu.edu
†Yamaha Motor Corporation USA
mike_uenoyama@yamaha-motor.com

I. INTRODUCTION

The last decade has seen remarkable advances in 3D perception for robotics. Advances in range sensing and SLAM now allow robots to easily acquire detailed 3D maps of their environment in real-time. While purely geometric maps have been proven useful for robotic navigation, they fail to represent many aspects of the environment necessary for intelligent robotic decision making.

An improvement over purely geometric maps are so-called semantic maps, which annotate the 3D geometric features with task-oriented semantic labels. In other words, a map of what is where. For a survey on semantic mapping, see [4]; some relevant work includes [13, 2, 10, 5, 12, 11].

In order to be useful for robot autonomy, a semantic mapping must be able to accurately predict relevant semantic labels in real time. In this abstract, we describe a simple method for semantic mapping that meets this goal and its application to two tasks, autonomous off-road navigation with an All-Terrain Vehicle (ATV) and autonomous semantic exploration (or scouting) with a Micro-Aerial Vehicle (MAV). In both cases, we have conducted successful field tests closing the perception-planning loop using only onboard computation.

II. METHOD

In essence, our method, as implemented in the two applications described in this abstract, consists of two main stages: semantic segmentation and mapping. In each application, the semantic mapping creates a task-oriented semantic map, and concurrently the planning system uses the semantic map to accomplish the respective robot’s missions.

While specific details vary in each case to handle different goals and constraints, they remain conceptually similar and share a significant amount of implementation infrastructure. An overview of the system for the case of off-road driving is depicted in Figure 2. Below we summarize each stage in more detail, including the differences and commonalities in how we use them for our two applications.

A. Semantic Segmentation

First, we perform 2D semantic segmentation, in which one of K predefined labels is assigned to each pixel of an image. The choice of classes is task-dependent, and often an involved engineering decision. In our case, for off-road navigation we use Smooth Trail, Rough Trail, Low Vegetation, High Vegetation, Trees, and Obstacles; these classes were chosen for their different traversability properties, e.g. the vehicle should try to stay on Smooth Trail; it is acceptable to drive over Low Vegetation but not High Vegetation. For the scouting task, we simply use Vehicle/Not Vehicle, as the task is to find and gather high-resolution imagery of vehicles such as cars and trucks.

To perform semantic segmentation, we use custom Fully Convolutional Network (FCN) architectures [6]. In each case, given the lack of suitable publicly available datasets, we collected and labeled our own. We plan to release our implementations.

Fig. 1. Platforms used in our field tests.

Fig. 2. System flow for Off-Road Navigation.
these datasets shortly. **Figure 3** and **Figure 4** show some qualitative results of our networks on our two application datasets.

Fig. 3. Left: Validation set examples of semantic segmentations (smooth trail, low vegetation, sky, etc.) from our off-road driving dataset. Right: An accumulated semantic map.

Fig. 4. Left: Validation set examples of semantic segmentations (vehicle/not vehicle) from our scouting dataset. Right: An accumulated semantic map after a scouting mission, where red indicates a found car.

B. Mapping and Projection

In the second stage, the task is to project the 2D image labels into a persistent metric representation which can be used for planning. In both applications, this representation is a 2.5D grid map defined in a global frame. The grid map has a predefined, finite size, but can be moved as the vehicle moves around; this is done efficiently with a 2D circular buffer-type data structure. Each grid cell maintains the robot’s beliefs regarding the semantic class and height of the grid cell’s contents.

In both cases, the beliefs regarding semantic labels are accumulated additively using the log-odds of the probabilistic pixelwise predictions provided by the CNNs. This is analogous to Bayesian binary occupancy mapping algorithms [9], with $K$ classes instead of occupancy.

The method of projection differs in the two applications. The ATV used in off-road driving has a LiDAR sensor, which provides fast and reliable depth measurements. In this case, we use raycasting and the (known) LiDAR-camera extrinsics to build the semantic map online (Figure 5).

On the other hand, the MAV used in scouting has no LiDAR, and we wish to build maps that may be outside the range of binocular stereo. Thus, in this case, to build the semantic map we raycast onto publicly available Digital Elevation Maps (DEMs), relative to which we have a pose estimate, and employ heuristics regarding the elevation of semantic classes. The result is of lower quality than in our other application, but sufficient for our task.

C. Planning

The sole purpose of the semantic maps is to help the robot accomplish its goals. In particular, the aim is to provide the planning system with information it can use to make better decisions.

For the off-road driving application, the task is to drive towards a specified goal while staying on trails and avoiding obstacles. The trails are not necessarily known beforehand, and may be covered in grass, have puddles, rocks, etc. We have used different planning systems toward this goal. The simplest planner is a receding horizon planner which assigns a traversability cost to each of classes (e.g. low for Smooth Trail, high for High Vegetation) and chooses among a discretized set of constant-curvature trajectories based on their cost in the map.

In the MAV scouting application, the task is to gather high-quality data regarding objects belonging to some predefined class, a task which we call scouting. Currently, we use automotive vehicles (trucks, cars) as our target class. Thus, the semantic mapping system maintains an online prediction of where vehicles may be present, and the planning system, described in [1], continuously uses this map to build information-gathering plans. We have demonstrated this system autonomously finding and collecting high-quality imagery of two cars hundreds of meters apart (Figure 4).

III. CONCLUSIONS AND ONGOING WORK

In this abstract, we have described a simple yet robust system for semantic mapping and its use in two robotic applications. In ongoing work, we are particularly interested in improving segmentation by incorporating geometry using multiview stereo or depth sensing. For the latter, we have introduced a novel 2D/3D CNN integrating image and LiDAR data (Figure 6) for semantic segmentation. Finally, we are also working in using inverse reinforcement learning and self-supervised methods to alleviate the need for manually chosen labels.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of Qualcomm, the Office of Naval Research and the Yamaha Motor Corporation.
REFERENCES


