Doctoral thesis summary: Generating High-Level Knowledge to Enrich Maps for Autonomous Ground Vehicles (自律移動ロボット用の地図を豊かにする高レベルの知識生成に関する研究)

submitted by Thomas Westfechtel to the Graduate School of Information Sciences, Tohoku University

In the thesis, different methods to generate high-level knowledge are investigated with the purpose of enriching maps of the environment to increase the autonomy of ground vehicles.

Chapter 1: Introduction

In view of the aging population of Japan, robotic solutions will play an integral role to overcome the resulting problems of a shrinking workforce as well as the problem of an increasingly large portion of population in need of special care-taking services. However, currently it is still challenging for robots to easily interact with humans or even operate within an environment that is shared with humans due to the sheer complexity. The robot has to simultaneously localize itself in the environment, perceive and understand the surrounding, and finally plan its action according to its environment, nearby agents and its task. To alleviate this problem, a common way is to provide the robot with maps of the environment containing a-priori high-level knowledge about it. These maps can vastly help the robot in the perceiving and understanding task. This could be in form of static objects of high interest like the position of traffic lights in a street scene. Another form of high-level knowledge could be location specific rules about the environment like which lanes are used for turning at a crossroad. In this work, the problem of generating high-level knowledge to enrich maps for autonomous ground vehicles from different standpoints for three different tasks is investigated.

Chapter 2: Robust Stairway Detection and Localization Method for Mobile Robots using a Graph-Based Model and Competing Initializations

The first task has its background in a crawler-type robot that is supposed to patrol a manufacturing plant on a daily basis for inspections tasks to alleviate the burden of human workers. During the route, the robot has to climb several stairways. This study presents a method for accurately detecting, localizing, and estimating the characteristics of stairways using point cloud data.

The main challenge is the wide variety of different structures and shapes of stairways. This challenge is often aggravated by an unfavorable position of the sensor, which leaves large parts of the stairway occluded. This can be further aggravated by sparse point data.

These difficulties are overcome by introducing a three-dimensional graph-based stairway detection method combined with competing initializations. The stairway graph characterizes the general structural design of stairways in a generic way that can be used to describe a large variety of different stairways. By using multiple ways to initialize the graph, stairways can be robustly detected even if parts of the stairway are occluded. Furthermore, by letting the initializations compete against each other, the best initialization that accurately describes the measured stairway can be found. The detection algorithm utilizes a plane-based approach. Different planar segmentation algorithms were also investigated and

experimentally compared in an application-orientated manner.
The system accurately detects and estimates the stairway parameters with an average error of only 2.5mm for a variety of stairways including ascending, descending and spiral stairways. The method robustly works with different depth sensors for either small or large-scale environments and for dense and sparse point cloud data. Despite this generality, the system's accuracy is higher than most state-of-the-art stairway detection methods.

Chapter 3: Vehicle Detection and Localization on Bird's Eye View Elevation Images Using Convolutional Neural Network
The second task deals with self-driving vehicles. Self-driving vehicles will provide mobility for elderly people especially in rural regions with few public transport options. For autonomous vehicles, the ability to detect and localize surrounding vehicles is critical. It is one of the fundamental step for further processing like path planning. This study describes a convolutional neural network based vehicle detection and localization method using point cloud data acquired by a LIDAR sensor. Acquired point clouds are transformed into bird's eye view depth images, where each pixel represents a grid cell of the horizontal x-y plane. The images are processed by a two stage vehicle detector. As the pixels of the image represent real world positions, the bounding box of detected vehicles correspond directly to the real world position of the vehicles. The position of the vehicles is critical for path planning and collision avoidance. Furthermore, the proposed bird's eye view images utilize a three channel representation that allows us to utilize common RGB image based detection networks without modification. To evaluate the suitability of our method for further applications like path planning, the detection results are evaluated based on the real world localization error.

Chapter 4: Parking Spot Estimation and Mapping Method for Mobile Robots
The third task investigates further self-driving vehicles. To operate, the self-driving vehicles rely on detailed semantic maps of the environment. In this work, a method to autonomously generate a semantic map enriched with knowledge of parking spot locations is proposed. The method detects and uses parked vehicles in the surroundings to estimate parking lot topology and infer vacant parking spots via a graph-based approach. The method works for parking lot structures in different environments, such as structured parking lots, unstructured/unmarked parking lots, and typical suburban environments. Using the proposed graph-based approach to infer the parking lot structure, it is possible to extend the estimated parking spots by 57%, averaged over six different areas with ten trials each. The accuracy of the algorithm is increased when combining multiple trials over multiple days. With ten trials combined, the whole parking lot structure was estimated, and all parking spots were detected in four out of the six evaluated areas.

Chapter 5: Fusion of Camera and LIDAR Data for Large Scale Semantic Mapping
In the fourth task, the generation of semantic maps for self-driving vehicles was investigated. The main focus is to generate large scale semantic 3D maps of the environment where each point is classified depending on their semantic class (i.e. road, sidewalk, building, ...). In this work, the focus was on the fusion of RGB data

of a camera and depth data of a LIDAR.
With the recent rise of deep convolutional neural networks, the semantic segmentation of pictures has made a great boost in terms of performance. To build upon this performance boost and the color and semantic data gathered from a round view camera system with depth data gathered from a LIDAR sensor were fused. In the framework, each LIDAR scan was projected on the pictures to extract the color and semantic information while at the same time creating a large-scale 3D maps of the environment using a LIDAR-based SLAM algorithm. A random forest with combined geometrical features of the LIDAR and color and semantic features from the camera system was trained. The features complement each other and the combination of them yields an increase in IoU of around 20 percent points compared to only using geometrical features and around 13 percent points compared to using mainly the semantic information gathered from the pictures. The system achieves an average IoU-score of around 62% over 14 different semantic classes, including three different classes for ground types. Through this sensor fusion state-of-the-art results were achieved in comparison to other large scale outdoor semantic labeling methods, despite having more semantic classes and considerably less training data. Some examples on how the extracted high-level knowledge can be applied for further applications is shown, in particular the ability of the technique to retrieve trajectories of pedestrian and vehicles within the environment. In addition, using the acquired knowledge points of the specific classes can be removed from the overall point cloud, i.e. removing the traffic from the environment or removing the vegetation to get views on buildings from angles that are normally hidden by trees or other objects.

Chapter 6: General Conclusions
The last chapter concludes this study. In this study, different techniques for generating high-level knowledge to enrich maps for autonomous ground vehicles are shown. Firstly, object detection in cluttered environment techniques are demonstrated for the particular case of stairway detection (Chapter 2) as well as vehicle detection (Chapter 3). Secondly, a way to learn high-level knowledge through observation of other traffic participant is introduced for the specific case of parking spot detection (Chapter 4). Thirdly, a method for detailed semantic scene segmentation is shown for the case of urban street scenes (Chapter 5). These techniques can help overcome the current difficulties and limits of autonomous ground vehicles. By adding the acquired high-level knowledge to a map of the environment, the robot can rely on a broader a-priori environment knowledge. This does not only help to increase the robot's scene understanding but is an essential part for further action planning of the robot.