Online Map Validation for Autonomous Driving

Extended Abstract

Andrea Fabris
Felix Drost
Luca Parolini
Andrea.Fabris@bmw.de
Felix.Drost@bmw.de
Luca.Parolini@bmw.de
BMW Group

Qing Rao
Andreas Rauch
Sebastian Schneider
Qing.Rao@bmw.de
Andreas.Rauch@bmw.de
Sebastian.SB.Schneider@bmw.de
BMW Group

Sebastian Wagner
Prof. Dr.-Ing. habil. Alois Knoll
sebastian.wagner@tum.de
knoll@mytum.de
Technische Universität München

ABSTRACT

Self-driving cars typically depend on high-definition (HD) maps for computing a driving strategy at areas inside and outside their field of view. Data in HD maps, however, can be outdated and erroneous. It is therefore of critical importance to validate this information before its use. We propose two complementary approaches for online map validation which promise sufficient performance for being effectively used on board in series production cars. The first approach builds a model-based framework. The second utilizes deep similarity learning.

ACM Reference Format:
Andrea Fabris, Felix Drost, Luca Parolini, Qing Rao, Andreas Rauch, Sebastian Schneider, Sebastian Wagner, and Prof. Dr.-Ing. habil. Alois Knoll. 2019. Online Map Validation for Autonomous Driving: Extended Abstract. In *Proceedings of CSCS19: Computer Science in Cars Symposium (CSCS19).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Autonomous vehicles promise to radically change and improve the way mobility affects society [7]. A recent study of *Continental AG* shows that safety is of great concern, e.g. 57% of the German population stated to be concerned about their reliability [9]. High Definition (HD) road maps contain both semantic, e.g. driving rules, and geometric features, e.g. lane structure, with accuracy in the range of approximately ten centimeters [5]. Map data can be considered as a sensor whose field of view is not affected by traffic conditions, but sends information with potentially unknown and extensive latency, e.g. maps can be days, or weeks old. In 2012, there have been 20.000 construction sites in Munich, Germany [3], which can lead to short or long term changes in the road structure. A discrepancy between the content of an HD map and the actual environment can lead to faulty system behavior which may cause violation of the safety constraints. In order to mitigate the risk of using data from invalid HD maps, it is of paramount importance to detect outdated or erroneous map data before this data is used for autonomous driving.

Previous work mainly focuses on algorithms for Simultaneous Localization and Mapping (SLAM) [6] and online road model generation [1], which both map the environment of a mobile robot or an autonomous vehicle. While these algorithms could be used for the purpose of map validation, their goal is to build a model of the vehicle environment. Our work aims to validate a HD map by directly comparing its content against sensor data, without having to build a model of the environment first. We expect that the proposed strategy will yield better performance than existing methods based on SLAM or online road model approaches with lower computational requirements.

CSCS19, October 08, 2019, Kaiserslautern, Germany

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 PROPOSED APPROACH

For the purpose of map validation, two approaches are developed. First, a framework models the validity of the map based on spatial and temporal correlation of sensor measurements. Probabilities for valid, invalid and unknown state of an HD map are represented in a grid structure. The second relies on a deep neural network for representation learning, in order to perform a binary classification between valid and invalid scenes. The two approaches complement each other and we expect strong benefits from a synergetic use of the two on board.

2.1 Model based approach

Spatial and temporal correlations are explicitly modeled and used for interpolating sensor measurements, which can only be gathered on limited areas and at irregular time intervals, e.g. due to sensor occlusion. This idea extends the concept of occupancy grids [4] for the characterization of the map validity: the environment is discretized and the value of each cell represents the probability, that the specific cell of the map is valid. Furthermore, in this framework, the hypothesis of independent cells is substituted by a notion of spatial correlation among points of the grid.

The validity of a map is modeled by a probability distribution defined over space and time. The correlation among measurements leads to a smooth variation of map validity that takes into consideration where and when sensor data has been collected. The probability distribution can take more than just two values, e.g. *valid* or *invalid*, by including values such as *unknown*.

By explicitly modeling the correlation of map validity, this approach can be exploited to estimate the validity of the map in areas where limited information is available. Although flexible, the proposed approach is subject to the assumption of known correlation values among measurements. While it is reasonable to assume correlation of spatial points close to each other, the exact definition of the correlation function is unknown and its approximation may largely affect the validation algorithm. Figure 1 shows an example of the probability distribution over space. In order to use the previous

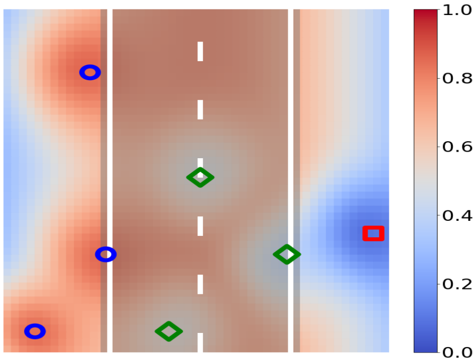

Figure 1: Spatial updating: circles represent valid, squares invalid and diamond unknown measurements. The color represents the validity probability of the map.

defined method, the knowledge of the surrounding environment is fundamental. For this reason, a fusion among different sensors is performed by a Bayesian Network [11]. The latter is able to represent a set of variables and their conditional dependencies via a directed acyclic graph combining the information in a probabilistic framework.

2.2 Deep Representation Learning

With this data driven approach, we focus on the validation of the geometrical aspects of HD maps. We expect it to be suited especially for elements, that are difficult to model such as complex geometric road shapes such as intersections.

During the recent years supervised, deep learning models have shown great performance in computer vision tasks on image data [10]. In multi-modal representation learning, the input of different sources is converted to a joined latent space [8]. This approach is used for online map validation in order to compare the on-board sensor measurements to the map data. For generating training data, map information and different sensor modalities of the vehicle are recorded during test drives. Subsequently, they are converted into an image representation as shown in Figure 2. Data pairs of the same recording time are labeled as valid samples. Due to the sparsity of erroneous map data on the test routes, invalid samples are created by matching map and sensor images with different time stamps. The trained models are evaluated against an unseen test set consisting of construction sites, which have not been modelled in the HD maps.

The binary decision of the map's validity is based on comparing map and sensor data in a compressed representation. In previous work *Siamese Neural Networks* are used to determine, whether two images show the same content e.g. for face recognition [2]. A body consisting of two convolutional neural networks (CNN) of the same architecture takes map and sensor data as input. Each network is processing one source of input, respectively. These CNNs are trained to extract high level structures of their input to feature vectors in the same latent space. Based on these feature representation of map and sensor data, binary classification is performed by a decision head. The feature extractor and the classification network are trained simultaneously in an end-to-end fashion to reduce the cost over the assigned labels.

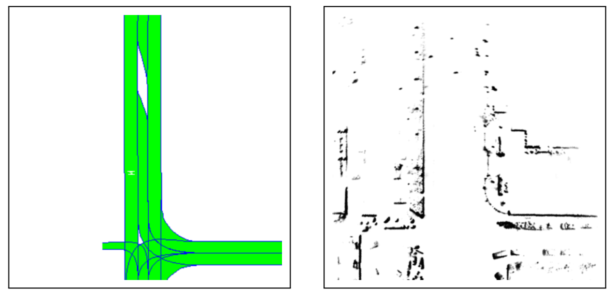

Figure 2: Sample input of a valid data pair. Left: HD map data. Right: occupancy grid representation of the static environment.

A challenge for training a deep learning model for map validation is the limited amount of invalid samples. In this approach, training and testing data have different structure, which is usually critical in deep learning algorithms. In the training data, map and sensor images from different time stamps are shown and therefore differ largely in the shown content. The test set features construction sites which change only parts of the road structure (e.g. block a lane). Training samples which match the test set better can lead to better results, but are hard to obtain.

3 CONCLUDING REMARKS

In the model-based approach, the developed framework can be easily extended for different sources of information. It is suitable when the modelling process is known or easy to approximate and permits to define the relation between multiple sources of information directly in a way which is understandable for a human observer. One drawback is the high number of parameters which have to be defined manually.

For a deep learning approach, the model is not described explicitly since it is directly recovered from the data. Complex comparisons between different data sources can be made if enough training data is available. Especially the sparsity of invalid samples is increasing the difficulty in this approach. Furthermore, deep learning models add high computational costs on the vehicle's processing unit.

Since the occurrence of map invalidity will cause reduced driving functionality or human intervention, further work has to be conducted on localizing the concrete invalid part of the map. If the false information does not affect the driving strategy, or missing information can be compensated by the vehicle's sensors, driving experience will benefit greatly.

REFERENCES

[1] E. Casapietra, T. H. Weisswange, C. Goerick, F. Kummert, and J. Fritsch. 2015. Building a probabilistic grid-based road representation from direct and indirect visual cues. In *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, Seoul, South Korea, 273–279. https://doi.org/10.1109/IVS.2015.7225698

[2] Sumit Chopra, Raia Hadsell, and Yann Lecun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *In Proc. of Computer Vision and Pattern Recognition Conference*. IEEE Press, San Diego, CA, USA, 539–546.

[3] Baureferat der Landeshauptstadt München. 2013. Jahresbericht 2012 2013. (2013). https://www.muenchen.de/rathaus/dam/jcr:021e3d71-5065-48b5-ab59-d6a92b63007f/Baureferat_Jahresbericht_2012-13.pdf

[4] A. Elfes. 1989. Using occupancy grids for mobile robot perception and navigation. *Computer* 22, 6 (June 1989), 46–57. https://doi.org/10.1109/2.30720

[5] Kichun Jo, Chansoo Kim, and Myoungho Sunwoo. 2018. Simultaneous Localization and Map Change Update for the High Definition Map-Based Autonomous Driving Car. *Sensors* 18 (09 2018), 3145. https://doi.org/10.3390/s18093145

[6] Jesse Levinson, Michael Montemerlo, and Sebastian Thrun. 2007. Map-Based Precision Vehicle Localization in Urban Environments.. In *Robotics: Science and Systems*, Wolfram Burgard, Oliver Brock, and Cyrill Stachniss (Eds.). The MIT Press, Georgia, USA, 352. http://dblp.uni-trier.de/db/conf/rss/rss2007.html#LevinsonMT07

[7] T. Litman. 2013. *Autonomous Vehicle Implementation Predictions: Implications for Transport Planning*. Victoria Transport Policy Institute, Victoria, British Columbia. https://books.google.de/books?id=G3FKnwEACAAJ

[8] Brian McFee and Gert Lanckriet. 2011. Learning Multi-modal Similarity. *Journal of Machine Learning Research* 12 (Feb. 2011), 491–523. http://dl.acm.org/citation.cfm?id=1953048.1953063

[9] Continental Corporate Media Relations. 2018. Where are we heading? Paths to mobility of tomorrow. (December 2018).

[10] Neha Sharma, Vibhor Jain, and Anju Mishra. 2018. An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science* 132 (2018), 377 – 384. https://doi.org/10.1016/j.procs.2018.05.198 International Conference on Computational Intelligence and Data Science.

[11] Aakriti Singhal and Christopher Robert Brown. 1997. Dynamic Bayes net approach to multimodal sensor fusion. In *Sensor Fusion and Decentralized Control in Autonomous Robotic Systems*, Vol. 3209. Pittsburgh, PA, United States.