Mobility21

A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Carnegie Mellon University

Penn
University of Pennsylvania

The Ohio State University

CCAC

A Scenario-based Database for Connected and Autonomous Driving in A Smart City

Ding Zhao,
https://orcid.org/0000-0002-9400-8446

FINAL RESEARCH REPORT

Disclaimer:

A Scenario-based Database for Connected and Autonomous Driving in A Smart City

Ding Zhao (dingzhao@cmu.edu)

Introduction:

A high-quality driving dataset is a key infrastructure to thrive in the autonomous vehicle industry in Pittsburgh and build a smart city for the residence. In this project, we aim to build the world's first scenario-based driving database that is dedicated to connected and autonomous vehicles. We plan to record and model the dynamic traffic information in Pittsburgh from heterogeneous driving data such as lidar point cloud, vision information, GPS, etc. Then, scalable machine learning approaches, including unsupervised learning, will be applied to automatically facilitate the extraction of typical driving scenarios.

Our data collection platform is equipped with multiple advanced sensors including Lidar, high-resolution camera, radar, GPS, IMU units, and vehicle information such as steering wheels and braking pedals. The platform is able to capture the complex and informative real-world driving scenarios and categorize them as high-dimensional and heterogeneous time series data. After that, we first propose several representative vehicle behavior categories and traffic scenarios in order to extract semantic segmentation from traffic data. An unsupervised learning approach based on nonparametric Bayesian will also be applied to learn and recognize driving scenarios. A user-friendly web application will be developed to provide the dataset to the public from a scenario perspective.

Data Collection System:

Our data collection platform is equipped with multiple advanced sensors including Lidar, high-resolution camera, radar, GPS, IMU units, and vehicle information such as steering wheels and braking pedals. The platform is able to capture the complex and informative real-world driving scenarios and categorize them as high-dimensional and heterogeneous time series data.

Figure 1: Data collection platform

After collection, the data is upload to a server. The raw will be processed into a specific format for relational retrieval and organization. Due to the large heterogeneity, a unified framework is applied to organize or visualize the data. The driving data will be segmented into meaningful pieces that have distinct semantics.

Vehicle Behavior & Traffic Scenario:

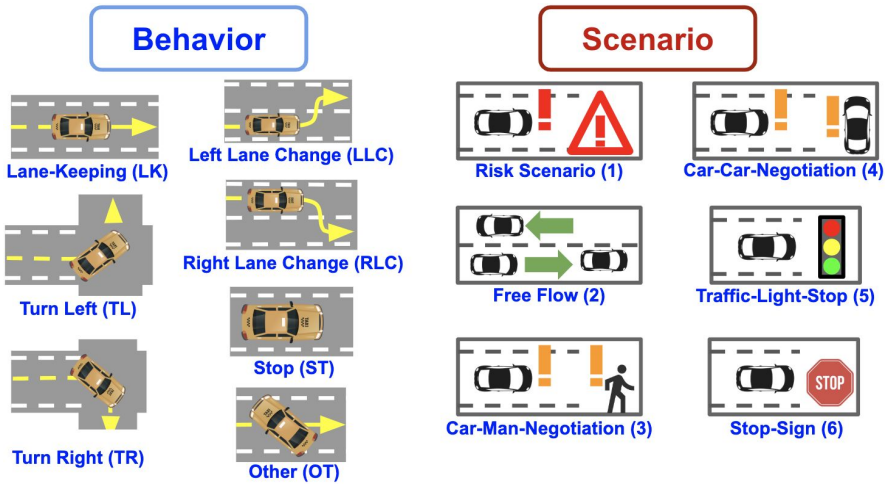
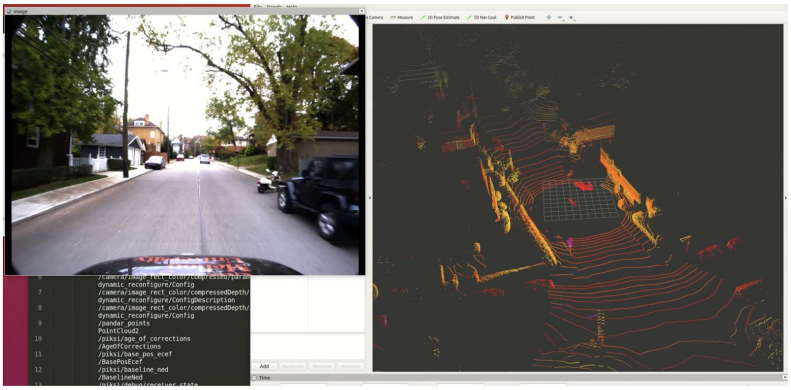


Figure 2: Predefined vehicle behavior and traffic scenario categories

In order to find the semantical information in traffic data. We used a hierarchical structure to describe the traffics. The ego-vehicle behaviors are categorized as Lane-Keeping, Turn-Left, Turn-Right, Left Lane Change, Right Lane Change, Stop, and Other. While the traffic scenarios are categorized as Risk scenario, Free Flow, Car Man Negotiation, Car Car Negotiation, Traffic Light, and Stop sign. As a result, we can represent the ego vehicle behavior, multi-vehicle interaction, as well as other road user behaviors comprehensively. We plan to record and model the dynamic traffic information in Pittsburgh from heterogeneous driving data such as lidar point cloud, vision information, GPS, etc. Dynamic unsupervised learning then will be applied to automatically extract typical driving scenarios automatically.



Behavior: TR[0,7], LK[7,12], ST[12,22], LK[22,48], ST[48,54]
S: 2[0,12], 6[12,22], 1[22,32], 4[32, 42], 2[42,48], 6[48,54]

Figure 3: Vehicle behavior and traffic scenario annotation of raw data

Scenario-Based Dataset for Autonomous Vehicles

http://3.91.193.80/index.html

A web application will be built to provide the dataset to the public from a scenario

perspective, from which users can query the driving scenario data from a dynamic interaction perspective. The web application is hosted on Amazon Web Service. The MySQL database is used to store the annotation and index information. A LAMP (Linux Apache MySQL PHP) archetypal model us used to support the web service. Meanwhile, the collected data can be visualized using open sources application.

Figure 4:The 3D visualize toolkit applied in this project

In order to let users query data according to specific traffic scenarios. An innovative grid-based user interface is implemented. As a result, traffic data can be accessed more easily and efficiently.

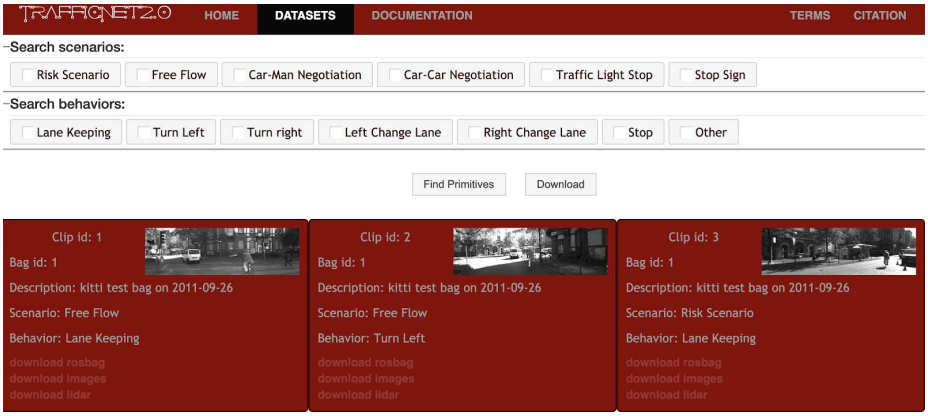


Figure 5:The query page of the scenario-based dataset.

Particularly, we plan to work closely with the department of mobility of the city and integrate the DSRC and smart cities information (e.g. traffic light, grid, event, weather, etc) into our analysis. The confidence of success stems from the lab's accumulated efforts in developing automated vehicle platforms and unsupervised machine learning theories supported by Toyota, Uber, Ford, Mcity, etc.