High-resolution Traffic Sensing with Autonomous Vehicles

Sean Qian[1,2*], Shuguan Yang[1], Allison Plummer[3]

1. Department of Civil and Environmental Engineering, Carnegie Mellon University
2. Heinz College, Carnegie Mellon University
3. Uber Advanced Technologies Group**

*Corresponding author. Email: seanqian@cmu.edu
**Uber Advanced Technologies Group provided access to autonomous vehicle trip data in partnership with Carnegie Mellon University's Mobility Data Analytics Center. The views expressed in this article are the authors' own and do not represent a formal position by Uber.

Executive Summary

With the rapid development of autonomous driving technologies, including sensing, system control, communications and cybersecurity, the era of self-driving is arriving. Over the last two decades, a variety of advanced driver assistance systems (ADAS) have been developed and deployed on a wide range of vehicles. A number of L4-L5 (full autonomy) pilot projects have been conducted and tested on public roads. Over 100 cities around the world are either piloting or preparing for the arrival of automated vehicle technologies. [1]

Self-driving holds the potential to bring unprecedented changes to when, where and how people travel. This implies tremendous benefits and challenges to transportation systems, and our communities in general. Among many other challenges, how would transportation stakeholders design, plan and operate infrastructure in the era of automated vehicles? As transportation modality norms further diversify, what usage data can be leveraged in the maintenance of infrastructure?

Once they are deployed at scale, a fleet of automated vehicles (AVs) could serve as floating (or probe) sensors on a road network, acquiring and analyzing data collected from their own vicinities that can be used to make observations about traffic conditions. These observations, when sufficiently spatio-temporally dense, could be used to derive meaningful insights about how a transportation network is performing and changing. Transportation stakeholders could benefit from various types of traffic information, including, but are not limited to, travel speed, traffic density and traffic flow by vehicle classifications. The potential value of insights conceivably derivable from this data must be considered in the context of the significant cost and complexity associated with gathering, processing, modelling, transferring, storing, and securing this data, particularly at scale.

We further prove the concept of automated-vehicle-based traffic sensing by sensing traffic flow on surface streets in Pittsburgh through a fleet of automated vehicles from the Uber Advanced Technologies Group. In particular, we develop a general method for estimating traffic flow on roads, at the block level, and intersections using object detection/tracking data from automated vehicles. The proposed method is able to effectively extract various characteristics of traffic flow, including travel speed, traffic density, and traffic counts.

[1] Bloomberg/Aspen Institute: https://avsincities.bloomberg.org/global-atlas

Table of contents

Executive Summary

Background and motivation

Traffic sensing through AVs
- Overview
- Methodology
- Model training

Case studies

Conclusions and future work

Background and motivation

With the rapid development of autonomous driving technologies, including sensing, system control, communications and cybersecurity, the era of self-driving is arriving. Over the last two decades, a variety of advanced driver assistance systems (ADAS) have been developed and deployed on a wide range of vehicles. A number of L4-L5 (full autonomy) pilot projects have been conducted and tested on public roads. Over 100 cities around the world are either piloting or preparing for the arrival of automated vehicle technologies. [2]

Self-driving holds the potential to bring unprecedented changes to when, where and how people travel. This implies tremendous benefits and challenges to transportation systems, and our communities in general. Among many other challenges, how would transportation stakeholders design, plan and operate infrastructure in the era of automated vehicles? As transportation modality norms further diversify, what usage data can be leveraged in the maintenance of transportation infrastructure?

Once they are deployed at scale, a fleet of automated vehicles (AVs) could serve as floating (or probe) sensors on a road network, acquiring and analyzing data collected from their own vicinities that can be used to make observations about traffic conditions. These observations, when sufficiently spatio-temporally dense, could be used to derive meaningful insights about how a transportation network is performing and changing. Transportation stakeholders could benefit from various types of traffic information, including travel speed, traffic density and traffic flow by vehicle classifications.

Large-scale, multi-source data from sensors datasets made possible by autonomous vehicle sensor packs, equipped on automated vehicles, such as LiDAR, radar, and video cameras, etc. can be used to detect and track objects in the vicinity of those automated vehicles. As this data is collected by self-driving technology developers for guiding autonomous driving, there is an opportunity to leverage data already gathered for other purposes. The potential value of insights conceivably derivable from this data must be considered in the context of the significant cost and complexity associated with gathering, processing, modelling, transferring, storing, and securing this data, particularly at scale.

Share of streets given to automobiles is subject to change. As transportation stakeholders consider more creative and dynamic solutions for their transportation networks in the face of mounting congestion and the changing nature of transportation demand, the approach discussed in this paper could inform investments in more flexible road networks, flex parking and the optimization of networks of shared lanes, and understanding the impacts of lane reconfiguration projects. This notion is especially relevant when considering the shift towards greater modal diversity in urban environments and the unique potential to visualize the changing nature of transportation.

In the rest of this article, we use traffic sensing as a particular example to demonstrate the potential of rich spatio-temporally dense observations for measuring urban traffic patterns.

[2] Bloomberg/Aspen Institute: https://avsincities.bloomberg.org/global-atlas

Traffic sensing through AVs

Overview

Currently, the prevailing approaches to collect traffic information fall into one of two categories:

1. Fixed location sensing (Eulerian sensing). To monitor traffic at certain road segments or intersections in the roadway network, one popular solution is deploying fixed-location sensors at these target locations. Possible sensors include loop detectors, radar, Bluetooth, tubes and traffic cameras. Those sensors all have respective drawbacks. For instance, some cannot be used to distinguish pedestrians or cyclists; traffic speed measurement may not be accurate nor reliable. Traffic cameras can provide a direct measurement of multi-modal traffic. However, this solution is not adequately reliable, and its costly. Computer vision algorithms implemented on a single fixed-location camera generally do not perform well under adverse conditions (such as bumper-to-bumper congestion and hazardous weather conditions). In the US, a typical high-resolution traffic camera kit costs minimally $10K per unit, let alone high maintenance costs. This solution cannot scale to large-scale networks if intensive roads/intersection coverage is required.
2. Probe/floating sensing (Lagrangian sensing). In this method, locations and timestamps of a fleet of floating vehicles (or a group of mobile devices) are being tracked in real time. Given a road segment, average traffic speed and traffic flow of population vehicles can be inferred from the data of those probe/floating sensors. This approach is widely adopted by numerous traffic service providers, such as Google Maps, INRIX, HERE and TomTom. Although requiring less capital investment than fixed location sensors, probe sensing is based on a sample population of vehicles and people that oftentimes provide a bias inference upon data analytics and inference. Due to its sampling nature, traffic information inferred from probe sensing may be insensitive to detect non-recurrent traffic incidents. In addition, probe sensing is widely deployed to sense vehicles and people, but it is challenging to sense physical infrastructure using this approach.

In a nutshell, neither of the two approaches are perfect for reliable traffic sensing with high spatio-temporal coverage in a modally diverse transportation network.

Here we propose the concept of high-resolution traffic sensing in urban road networks using the data collected by a set of advanced sensors equipped on AVs. In particular, the AV-based approach is capable of measuring multi-modal traffic, such as speed, flow and density.

Very little research studied AV-based data for traffic sensing. The idea and methodology proposed in this article substantially differentiates from the past literature. In particular, a number of studies discussed the possibility of utilizing vehicular ad-hoc networks for traffic sensing (e.g., Eze et.al 2016). Most of them focus on the development of communication systems, whereas our study proposes a methodology for robust traffic sensing. More importantly, those methods using vehicular ad-hoc networks consider features of the probe vehicles only, whereas our approach incorporates features of all vehicles, as well as the features of probe vehicles. Chen et.al (2017) proposed a ‘nearby traffic flow modelling’ framework based on onboard cyber-physical-system sensors. However, how sensing data are integrated and used to estimate traffic flow is not elaborated. On the other hand, our study presents a complete approach for AV-based traffic sensing. Another study relevant to AV-based traffic sensing is MIDAS (Potluri and Mirchandani, 2018), short for Managing Interacting Demand and Supplies. MIDAS is a cyber-physical system framework aiming at the estimation of queue lengths and subsequent traffic signal timing optimization. The main data source of MIDAS is from video cameras installed in vehicles. Our approach uses processed data originally collected from AVs equipped with a variety of sensors, which enables accurate object detection in the vicinity of AVs.

In fall 2018, the Mobility Data Analytics Center at CMU and Uber ATG worked together to test the concept of traffic sensing through Uber ATG AVs in two urban environments in Pittsburgh. We use the object detection and tracking data extracted from Uber ATG AVs to estimate multi-modal traffic speed, flow rate and density at several locations where substantial activities of Uber ATG AV testing were recorded. Under a certain AV penetration rate, this method is expected to have a higher spatio-temporal resolution and accuracy than fixed-location sensing and probe sensing. This is because AV-based traffic sensing tracks all actors in the proximity of AVs, with a set of onboard sensors, including Lidar, radar and cameras.

In the proposed method, traffic-related information collected by a fleet of Uber ATG AVs is integrated and fed into an estimation model, which outputs all types of traffic conditions in a pre-defined spatial temporal resolution, such as in 5-min intervals on each road segment of 200-500 meters. The model is able to estimate multiple characteristics of traffic conditions, including speed, flow, and density. In particular, the estimations are provided for different vehicle classifications for all directions of travel lanes.

Methodology

In the proposed method, the following information is continuously collected from individual AVs and used as input for the estimation model:
(1) All detected objects of interests on the road, in this case, all cars, trucks, bikes and pedestrians, are called 'actors'. We incorporate multiple attributes of an actor in our method, include relative location to the AV, speed, headway, lane, the type of actor, whether they are currently occluded by other objects, etc. In a simplified manner, one can simply take the average of the estimated speed of all detected actors from AVs as the estimated traffic speed of all population vehicles, and count the total number of actors on the road as the estimated traffic density for this road segment. While the additional information including relative location, speed and whether occluded give us may further reveal how 'trustworthy' the data are, thus the traffic characteristics estimation can be improved with the machine learning techniques.
(2) The current phrase of all nearby traffic signals. In urban road network, the instantaneous traffic speed/flow/density is heavily impacted by the rolling phrases of traffic signals.
(3) Location and direction of the AV which provides sensing data. This provides the road segment to be estimated.

To estimate the multi-modal traffic over a time interval (e.g. 5 mins), we propose a two-step method as explained in Figure 1 and below:

1. Observations: As individual AVs moving along road segments, we first estimate the instantaneous traffic flow/density/speed. This step can also be seen as taking 'snapshots' of the temporary traffic conditions given the time and location. We denote individual snapshot as one observation, $\boldsymbol{x}_{ti}$, where *t* is the timestamp and *i* is the location index of the AV. Due to the impacts of traffic signal phrases, it is unreasonable to assume a constant rate of traffic over a period of time. Our method incorporates the detected traffic signal information by multiplying the original $\boldsymbol{x}_{ti}$ by a corresponding factor, to make it an estimation of the average traffic flow/density/speed over a complete traffic signal cycle based on this 'snapshot'. Each observation contains a set of values including the estimated means and variances of all types of traffic, denoted as expectation $E(x_{tim})$, and variance $\sigma(x_{tim})$, where *m* indexes the type of traffic specified by a mode and/or a movement direction.

2. Interpolations: Given the time interval and road segment for estimation, we aggregate all relevant observations from nearby AVs and estimate the average traffic over the time interval via kriging, which is a Gaussian process based interpolation method. By doing so, we are essentially considering multiple ‘votes’ from a fleet of AVs, which gives us a more comprehensive estimation of the traffic conditions on average. Thus, the estimation accuracy of the proposed traffic sensing method will improve as the AV penetration rate increases in the near future.

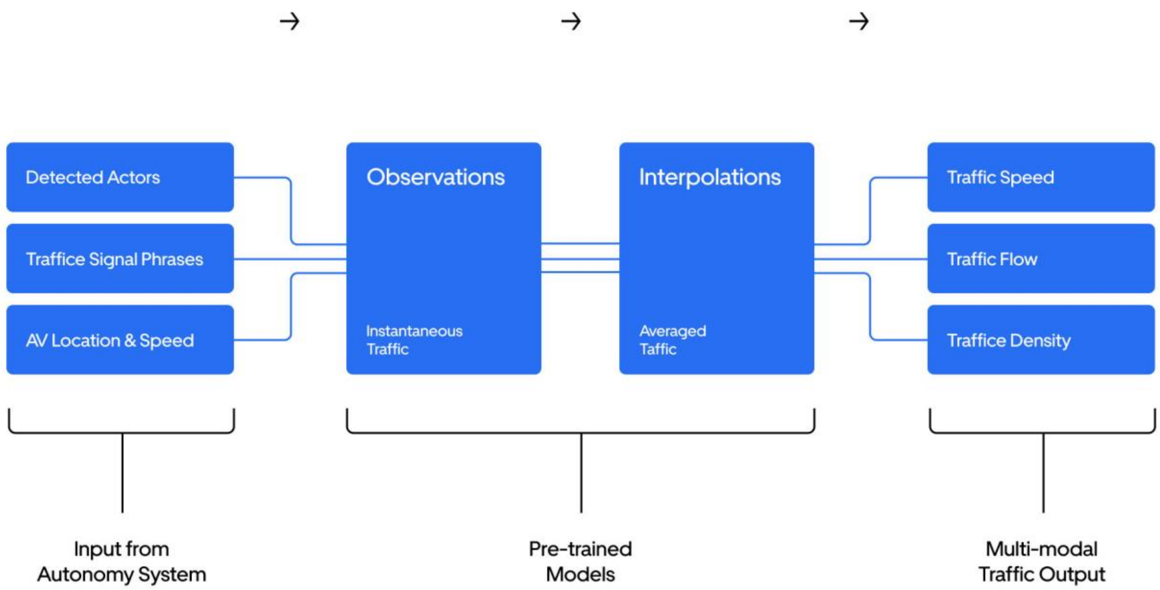

Figure 1 Method overview: Multi-modal traffic estimation

Due to the nature of visual/LiDAR perceptions, the accuracy of detection, classification and tracking declines with respect to the distance of the objects from the AV. Also, the efficiency of traffic sensing is impacted by lane configurations and the direction of travel. For example, it is easier to count the number of vehicles on the opposite direction of the AV than on the moving direction, as the view of the AV is more likely to be blocked by large objects next to the AV along its moving direction. To tackle these problems, we adopt a ‘zone-based’ method in the observation step, the surrounding area of an AV is separated into several zones. Different machine learning models are trained and applied to each of the zone when estimating instantaneous traffic. A sample configuration of zones for a three-lane road with bike lanes is illustrated in Figure 2.

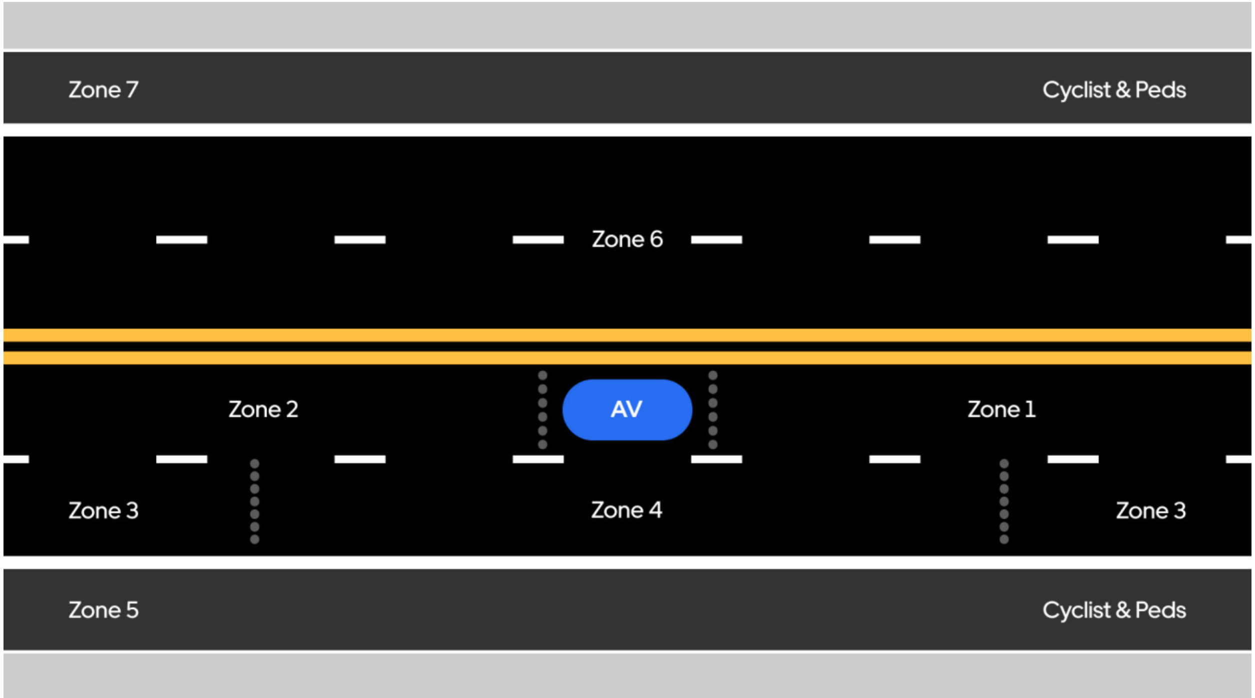

Figure 2 'Zone-based' observation: Sample zone configuration

In addition, we use a probabilistic approach that estimates the means and variances of the instantaneous flow/density/speed. The variances represent how confident we are for the estimation. For instance, given two observations with different variances, we would lean more on the one with a lower variance by taking the weighted average as the output. On the other hand, unlike vehicle control that requires close-to-zero false positive rate, reasonable variances in traffic sensing can be acceptable for the purpose of traffic management. Then, the expectation and variance of the estimated multi-modal traffic are obtained via the following equations, where t, i, m, z stands for the indices of time, road segment, vehicle classification and zone respectively. $\boldsymbol{\theta}_E$ and $\boldsymbol{\theta}_\sigma$ are the parameters for zone-based expectation and variance, which are optimized during training using actual observations.

$$E(x_{tim}) = \sum_z (\theta_{Ez} \cdot E(x_{timz}))$$

$$\sigma(x_{tim}) = \sum_z (\theta_{\sigma z} \cdot E(x_{timz}) \cdot \sigma(x_{timz}))$$

In the interpolations step, the model takes the estimated $E(x_{tim})$, $\sigma(x_{tim})$ from a series of observations, and use Kriging to fit a curve representing the spatial and temporal pattern of the traffic. Kriging is a Gaussian process based regression model for spatial and temporal interpolation. In this case, Kriging uses discrete observations from AVs on different road segments at various time points, to estimate the traffic flow/speed/density of any time and location, Kriging outputs a weighted average of all observation. In Kriging or Gaussian processing, it is assumed that the joint probability of any two observations x_{tim} and $x_{t'i'm'}$ follows a multivariate Gaussian distribution, as:

$$(x_{tim}\ x_{t'i'm'}) \sim N((\mu_{tim}\ \mu_{t'i'm'}), (K\ K'\ K'^T\ K''))$$

Where μ_{tim} is the mean of the observations and K is the covariance matrix or a kernel function measuring the correlation between observations tim and $t'i'm'$. Some of the popular choices of kernel functions includes:

- constant kernel $K(x, x') = C$

- linear kernel $K(x, x') = |x - x'| + c$
- RBF kernel: $K(x, x') = exp\left(-\frac{|x-x'|^2}{2\,l^2}\right)$.

We use the linear kernel to measure the temporal correlations and the RBF kernel to measure the spatial correlations. By using RBF kernel, we have the flexibility to tune the spatial range within which we have a prior belief on how strongly traffic are correlated. The final Kernel functions is shown below:

$$K(x_{tim}, x_{t'i'm'}) = (|t - t'| + c) \cdot exp\left(-\frac{|d(i,i')|^2}{2\,l^2}\right) \cdot K_m(m, m')$$

Where $d(i, i')$ is the driving distance between *i* and *i'*, $K_m(m, m')$ is the pre-defined correlation between two modes of traffic flow. Finally, given any set of t', i', m', z', denote $x^* = x_{t',i',m',z'}$, the estimated mean and variance of x^* are:

$$E(X) = K(x^*, X) \cdot (K(X, X) + \sigma_x^* \cdot I)^{-1} \cdot E(X) + \mu_{x*}$$
$$\Sigma(X) = K(x^*, x^*) - K(x^*, X) \cdot (K(X, X) + \sigma_x^* \cdot I)^{-1} \cdot K(X, x^*)$$

Where X is the set of observations collected from relevant AVs.

Note that not all x_{tim} need to be collected on road segment *i*, as the AV onboard system is capable of sensing from a range or predicting future movement of objects from their past trajectories. Some can also come from adjacent segments on which other AVs observe the traffic heading towards to segment *i* based on the predicted trajectories. In other words, our method is capable of sensing traffic in a spatial-temporal manner. An illustration of using Kriging to interpolate the temporal pattern of traffic flow is shown in Figure 3. The total traffic flow over a 10-min time interval is estimated by calculating the area under the solid blue curve.

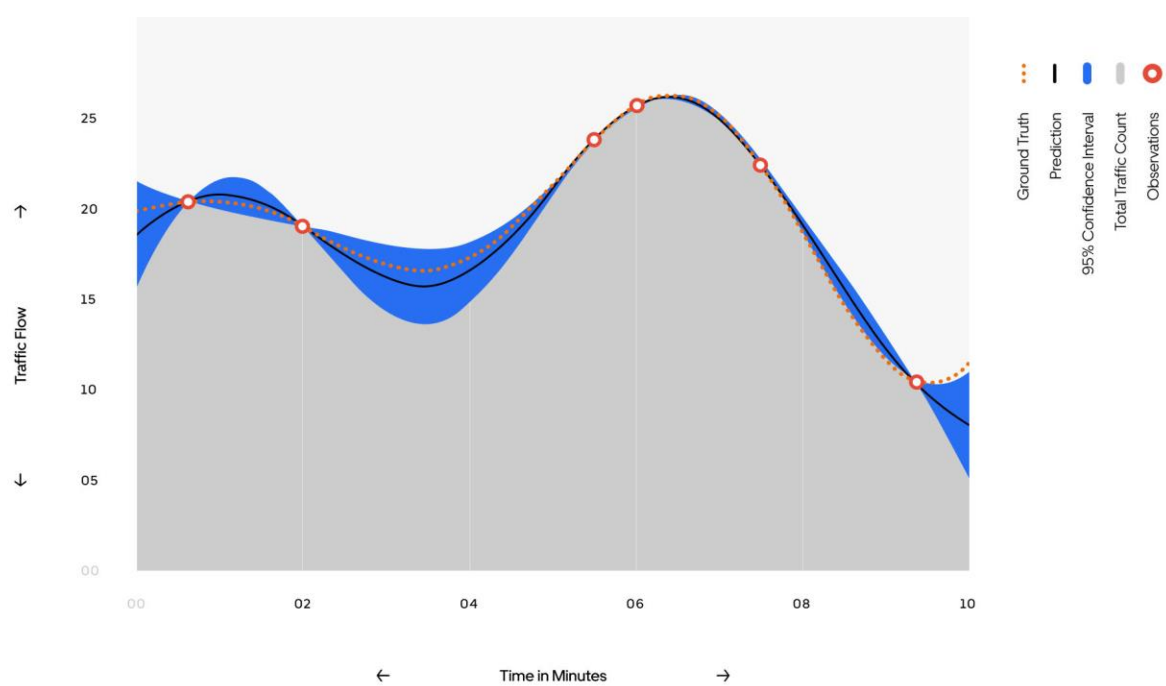

Figure 3 Kriging for traffic flow estimation

Model training

The proposed method contains multiple machine learning models which need to be trained and tested before being deployed. Model training for the observations part can be done using historical AV trips and ground truth traffic conditions that are human labeled. The resultant observation models can then be used to generate inputs for the interpolation training. The ground truth of the interpolation training is the average traffic flow, density and speed, which can be obtained from fixed-location traffic cameras or human-annotated ground truth.

Case studies

To evaluate the efficiency of traffic sensing through AVs, we conducted case studies on two road segments in the strip district of Pittsburgh downtown: Smallman Street and Liberty Ave.
In both cases, the AVs are used to estimate westbound traffic density while going eastbound. Traffic density is measured by the total number of actors per meter. In the case studies, densities of multiple modals of traffic, including cars, trucks, bicycles and pedestrians are estimated using the proposed method. The scope of the study is illustrated in Figure 4.

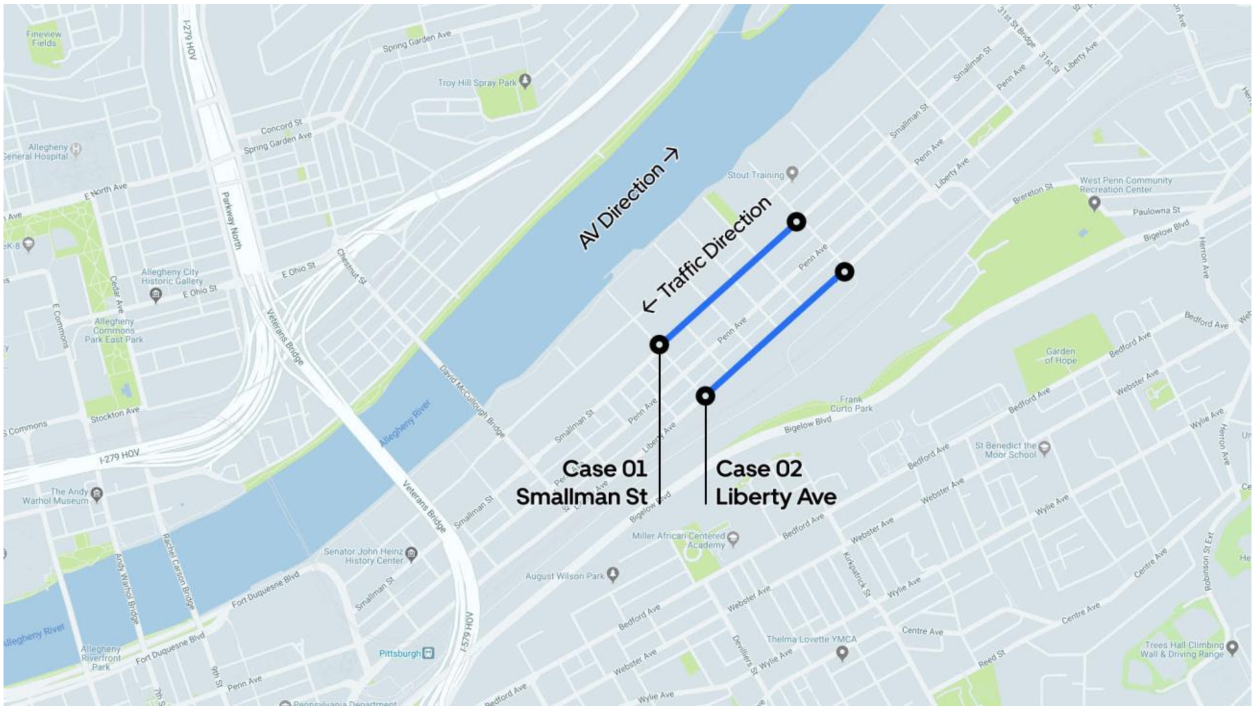

Figure 4 Scope of the study: Strip district in Pittsburgh

To conduct the case studies, we first collected historical Uber AV trips data taken between March 2016 and August 2017 on the two road segments of study.[3] Among the selected trips, we then find all scenarios where at least three AVs have passed through the segment within any 5-min intervals. For each scenario, the first, third and subsequent AV trips are used as the input of the two-step method, which outputs the estimated traffic density at the time the second AV passing

[3] All data is anonymized and aggregated to ensure no personally identifiable information or user behavior can be surfaced.

through the road segment. By doing so, we do not estimate the averaged density over the 5-min interval. Rather, we evaluate the performance of this method in interpolating a random time point within the time intervals of a set of observations. The proposed method reaching a satisfactory accuracy would demonstrate its abilities in estimating the averaged traffic at an arbitrary time. In the case studies, each scenario is converted into a data point. All those data points are then used to train and test the models. On the Smallman St, we find in all 616 such data points, while on Liberty Ave, 142 data points are found.

The performance of the method is evaluated by the testing mean absolute percentage error (MAPE) of the estimated multi-modal traffic density. For the case of the Smallman St, an overall MAPE of 0.24 is reached, while for the Liberty Ave, the MAPE is 0.69. Among all the data points, a significant portion was taken during night time when congestion was particularly low. If we focus on scenarios with sufficiently dense traffic, e.g. during peak hours, the performance of the model is improved. For the case of the Smallman St, we select a subset of scenarios with traffic density higher than the 60 percentiles of all the 616 data points. The experiment reaches an overall testing MAPE of 0.22. As for the case of the Liberty Ave, we select scenarios with density higher than the 80 percentiles, then achieve a better MAPE as 0.45.

Conclusions and future work

Large-scale, multi-source data from sensors datasets made possible by autonomous vehicle sensor packs, equipped on automated vehicles, such as LiDAR, radar, and video cameras, etc. can be used to detect and track objects in the vicinity of those automated vehicles. As this data is collected by self-driving technology developers for guiding autonomous driving, there is an opportunity to leverage data already gathered for other purposes. These observations, when sufficiently spatio-temporally dense, could be used to derive meaningful insights about how a transportation network is performing and changing. The potential value of insights conceivably derivable from this data must be considered in the context of the significant cost and complexity associated with gathering, processing, modelling, transferring, storing, and securing this data, particularly at scale.

We further prove the concept of automated-vehicle-based traffic sensing by sensing traffic flow on surface streets in Pittsburgh through a fleet of automated vehicles from Uber ATG. In particular, we develop a general method for estimating traffic flow at the level of road segments and intersections using object detection/tracking data from automated vehicles. The proposed method is able to effectively extract various characteristics of traffic flow, including travel speed, traffic density, and traffic counts. The estimation results are reasonably accurate and satisfactory.

In the near future, we will conduct more studies evaluating the accuracy of this traffic sensing approach over larger network areas to better understand potential sampling biases vis a vis fixed sensor and probe vehicle approaches. We will also look into improving the current machine learning models by optimizing each zone for the actor usage profiles and various road configurations. One possible improvement is to apply spatial temporal kriging with different types of kernels. In this case, not only AVs on the target road segment, but also those on adjacent road segments can be incorporated to the estimation models, which lead to the increase in the average number of observations and thus, smoother and more confident interpolations of traffic characteristics over larger, more representative urban areas.

References

Chen, B., Yang, Z., Huang, S., Du, X., Cui, Z., Bhimani, J., Xin, X. & Mi, N. (2017). Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles. In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)* (pp. 1-6). IEEE.

Eze, E. C., Zhang, S. J., Liu, E. J., & Eze, J. C. (2016). Advances in vehicular ad-hoc networks (VANETs): Challenges and road-map for future development. *International Journal of Automation and Computing*, *13*(1), 1-18.

Potluri, V. & Mirchandani, P. (2018). Developing and testing proactive signal control strategies for arterial-freeway diamond interchanges. Technical report for SOLARIS Consortium, Center for Advanced Transportation Education and Research. Accessed Feb 2019. https://www.unr.edu/Documents/engineering/solaris/Mirchandani_Task%20G%20Final%20Report.pdf