Automated Vehicles and Roadmanship: *How Safe Robot Cars Can Read the Road Like Humans Do*

HUEI PENG, PhD
Roger L. McCarthy Professor of Mechanical Engineering
University of Michigan

MINGHAN ZHU, MSE
Perception Research Engineer
University of Michigan

JOHN K. LENNEMAN, PhD
Senior Principal Research Scientist
Toyota Motor North America
Collaborative Safety Research Center

Contents

1 Introduction
2 Background
3 Research Question
4 Methodology
8 Findings
10 Next Steps
11 References

INTRODUCTION

In a 2018 report [1], the RAND Corp. defined the term roadmanship as "the ability to drive on the road safely without creating hazards and responding well—regardless of legality—to the hazards created by others." An automated vehicle (AV) constantly slamming on its brakes to avoid obstacles even when it is not really a safety risk, for example, is annoying and could create a safety hazard for other vehicles.

Recently, two research projects supported by Mcity at the University of Michigan and Toyota Motor North America's Collaborative Safety Research Center (CSRC) defined a process to quantify roadmanship using naturalistic driving data from unprotected left-turns and roundabouts. Naturalistic data is collected from real-world, every-day driving. The basic concept is that if enough naturalistic driving data is collected, and that data is used to guide the decision-making process of an AV so that its behavior is statistically undistinguishable from vehicles driven by human drivers under similar situations, the AV will literally be undistinguishable from other human-driven vehicles and seamlessly blend into the local driving culture. This is important today, when most vehicles are operated by human drivers. As AVs achieve high market penetration, the expectation is that they will become better than human drivers. During the transition, establishing a clear definition of, and maintaining desirable, robust roadmanship is essential.

BACKGROUND

What are some examples of desirable roadmanship traits? Consider unprotected left-turns. Many YouTube videos show vehicles operating in automated mode forfeiting what many typical drivers would see as a safe distance, or gap, separating it from an oncoming vehicle. At other times, the automated vehicles start to move but quickly slam on the brake to abort. Better "gap acceptance" and avoiding unnecessary harsh braking are two examples of desirable roadmanship traits. Other examples, among many, include:

- Selecting a proper driving speed
- Using smooth acceleration and braking levels
- Maintaining a proper and non-threatening distance from vulnerable road users, such as pedestrians or cyclists, walking or riding along on the shoulder

The concept of roadmanship is also useful for assessing vehicle performance beyond safety. Perhaps more importantly, it has great potential to guide the design of automated vehicles so they will be more readily accepted and trusted by the general public. Insisting on good roadmanship is essential to realizing the potential benefits of widespread adoption of automated vehicles, including safer, greener, more equitable, and more accessible mobility options.

In 2019, Waymo, Cruise, Zoox and other companies filed a combined total of 105 crash reports involving prototype AVs to the California Department of Motor Vehicles [2], representing total testing mileage of 2.85 million miles and averaging around 27,000 miles between crashes. The majority of the reported crashes were minor, and, in most cases, the AVs were rear-ended or side-swiped by other human- driven vehicles.

In comparison, police-reported crashes occurred every 500,000 miles for human drivers.

The prototype AVs represented Level 4 automation as defined by SAE International: Humans are not required to take over driving in any situation, but the vehicles can operate only in specifically defined areas, known as operating design domains, which may be limited by weather conditions, roadway surface type, speed, and other factors.

At first glance it appears that the AVs were not at fault in the 105 reported crashes. However, further consideration of the crashes suggests that some of the AVs were too cautious and could have deviated too much from driving norms. In that sense, even though these crashes were not caused by the AVs, they demonstrate that improvements in AV roadmanship could lead to a more natural driving environment for surrounding vehicles, and contribute to a safer and more enjoyable driving experience.

RESEARCH QUESTION

A good example of undesirable roadmanship involves decision-making at an unprotected intersection or a roundabout. What is an appropriate gap to keep between vehicles? In situations like these, traffic rules (which can vary from region to region) frequently say something vague, such as, "Yield to oncoming traffic, proceed when safe." These statements spell out the general rules but are not directly useful when designing path planning and control algorithms for AVs. We have a good idea of what unsafe gaps look like, but it is not at all clear what safety margins a robot driver should maintain.

The main goal of this paper is to focus on the gap acceptance decision problem using two example scenarios—an unprotected left turn at an intersection, and an unprotected right turn at a roundabout—to illustrate our approach to data gathering and analysis. We hope much of what we learned also translates to other driving scenarios.

It should be noted that while human decisions in terms of gap acceptance have been studied in the literature, most of that research did not take advantage of the vast progress in machine vision and artificial intelligence in the last several years and used less accurate sensors.

To start, visualize a "conflict zone" at each intersection and roundabout; the yielding vehicle should clear the conflict zone before the next oncoming vehicle arrives. Assuming that the safety margin is measured in seconds—the time gap—what is the minimum time gap a robot should allow? 2 seconds? 5 seconds? What about the upper limit? When would a robot vehicle be perceived as too timid? With a 10-second time gap?

Figure 1: Gap acceptance decisions for drivers at unprotected turns near intersections and roundabouts

Every human driver is different. What other drivers perceive as desirable roadmanship behavior may change from city to city, state to state, even country to country. In several cities where level-4 AVs are deployed in larger numbers, including Phoenix and San Francisco, there have been many complaints that these AVs take too long to make an unprotected left turn, such as when an AV has a green light but not a green arrow, leaving the decision about when to turn up to the AV. In addition, less desirable roadmanship behavior was evident whenever one of these AVs had a safe gap to proceed, but then aborted the maneuver and slammed on the brake.

METHODOLOGY

Data collection at an intersection

There are many traffic cameras installed all over the world, and some of them make live streaming openly available. We needed data collected only during the unprotected left turn phase at an intersection, when a traffic light flashes yellow in the United States, for example. Ideally this location would also have varying climate conditions to provide data during rain and snow.

Footage reviewed from a live-streaming camera in Jackson, Wyoming, showed a dedicated left-turn lane, a dedicated traffic signal phase (flashing yellow light), and a good view of the oncoming traffic. While a quick review of the video suggests the camera provides all the key information, several machine vision and signal processing steps were needed before the collected videos could be used for our roadmanship analysis.

Figure 2: Example image from a roadside camera useful for roadmanship analysis

The machine vision processes we needed to convert images that humans can see and understand into useful data for subsequent analysis included:

- Converting the images to bird’s-eye-view
- Extracting accurate vehicle position and orientation information for gap acceptance analysis
- Detecting the “unprotected left turn” phase of the traffic signal

Converting the images to bird’s-eye-view is a relatively well-defined and straight-forward process if the camera’s intrinsic and extrinsic parameters are known. Camera intrinsic parameters fully define a mapping between the camera coordinates and pixel coordinates in the image frame. These parameters are camera dependent. The extrinsic parameters define the location and orientation of the camera with respect to the world frame and depend on the location and pose of the camera as it was installed. Since the traffic camera was neither purchased nor installed by the research team, these parameters are not known. Therefore, a calibration process was necessary to reverse-engineer these parameters. Google Earth images and known factors such as lane width and key markers were used as the basis for calibration.

Extracting accurate vehicle position and orientation was the most challenging part of the machine vision technique that had to be developed. The challenge can be readily visualized from Figure 3. Bird’s-eye-view images are inevitably distorted, especially for the pixels from objects farther away from the camera and higher up from the ground. The distortion makes the objects look very fuzzy. Therefore, pinpointing their precise X-Y position and orientation is very hard for human eyes. The good news is that these distortions are no more (but also no less) difficult to process for artificial neural networks (ANN), which essentially simulate a human brain. The key challenge is to secure enough data with accurate ground-truth labels, meaning they confirm the real-world conditions, to train these ANNs.

Figure 3: (a) Original image, (b) Bird's-eye-view, and (c) Extracted vehicles with 2D bounding-boxes (the vehicle footprints on the ground) which indicate position and orientation of the vehicles in our left-turn data collection.

Because images recorded from real-world traffic do not come with ground-truth information, we had to generate training labels by using images enhanced by augmented reality techniques. Figure 4 shows the basic ideas of the process. First, we gathered vehicle images of different types and models from the library of ShapeNet, a large-scale dataset of 3D shapes. We then inserted the images on an "empty intersection" image, to represent left-turning and on-coming vehicles at various locations with different orientations. We found that occlusions are important, especially for oncoming vehicles. Therefore, it is important to insert vehicles in a way that causes realistic occlusions. Finally, these augmented-reality images were converted to the bird's-eye-view and used as the training dataset. The trained ANN achieved better than 90 percent accuracy, even for occluded vehicles.

Figure 4: Simulated vehicles with known locations and orientations under diverse weather conditions inserted to help to train our ANN to achieve high perception accuracy.

The final machine vision task involved identification of the traffic signal state and focused only on the segments for unprotected left-turns (flashing yellow light). In other phases of the signal, there was no decision for the left-turning vehicle to make. This task was relatively easy as the traffic signal was at a fixed location.

After the positions and orientations of the vehicles were identified, it was necessary to find out what gaps human drivers accept or reject. Figure 5 shows the concept of identifying the "key frame," which defines the critical reference point in time a decision is made. Once a left-turning vehicle was found to arrive at the reference line, we examined the existence of a straight-driving vehicle and observed whether the left-turning vehicle accepted the "gap" or rejected it. For both cases, we went back in time to find out what vehicle states are the key variables for such a decision. This involved the definition of multiple lines and reference points, and significant effort in data analytics. The details of the data processing work can be found in [3].

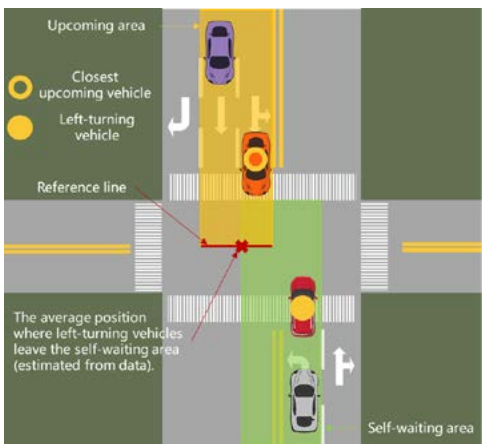

Figure 5: Reference points to compute accepted and rejected gaps at unprotected left-turns (shown in idealized bird's-eye-view).

In the end, we collected more than 5,000 accepted or rejected gaps, including 3,000 during sunny/fair weather and 2,000 on snowy days. We did not collect any data in the evening or at nights as the machine vision results were not found to be accurate enough.

Data collection at roundabouts

The basic concept and challenges in collecting roundabout data are similar to those for intersections. Figure 6 shows a streaming roundabout camera we found. The key challenges and machine vision tools were the same as for the intersection scenario, and the ANN for this camera was tuned and trained with another set of augmented reality images to achieve better perception accuracy. More than 22,000 data points were collected to indicate human drivers' acceptance/rejection decisions, including 9,000 in normal weather, 9,000 on rainy days and 4,000 on snowy days.

UNIVERSITY OF MICHIGAN

Figure 6: Original image (left) and extracted vehicles shown in 2D bounding boxes (the vehicle footprints on the ground) which indicate position and orientation for the roundabout data collection. Sometimes false positive vehicles, such as the image circled in red with ID 141 in the right photo, need to be filtered out by pixel locations.

FINDINGS

Sample roadmanship analysis

A common hypothesis related to human drivers' gap acceptance is that the most important factor is the "time gap," measured in seconds [4][5]. Since we had data from thousands of trips, we revisited this hypothesis, and analyzed the effect of many factors, including the distance between the two vehicles, speeds of both vehicles, weather (snow for this research), the presence of multiple oncoming vehicles, and how long the left-turning vehicle had been waiting during the flashing-yellow signal phase. The results showed, surprisingly, that the most important variable is the distance gap, not the time gap, as shown in Figure 7.

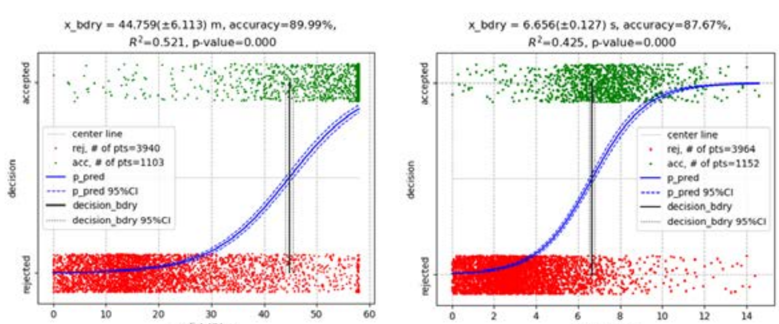

Figure 7: Logistic regression analysis of the gap acceptance/rejection data. Using the distance gap (left) results in a better separation of the data compared with using the time gap (right) as the variable.

Using the distance gap to separate accepted vs. rejected data points results in a single-variate model that has higher accuracy than using the time gap (89.8 percent vs. 83.3 percent). Even human eyes can see the better separation between the green data points (accepted gap) and the red data points (rejected gap) by using the distance gap. It is possible this observation, drawn from a single intersection, may not be as robust for other intersections.

The behavior of the left-turning vehicle arriving at the waiting line also showed statistical significance with the subsequent gap acceptance decision: a vehicle creeping forward (rather than fully stopped) tended to be more aggressive and accepted shorter gaps. Weather, snow for this study, the presence of multiple oncoming vehicles, and time waiting in the flashing yellow phase of the traffic signal did not impact the gap acceptance behavior noticeably.

For roundabouts, we used our own data, as well as data from two open datasets: the Roundabout Drone Dataset (RounD) [6] and the INTERACTION Dataset [7]. Both of these datasets were pre-processed, and vehicle locations and orientations were already available. In other words, no machine vision processing was necessary. Using these two open datasets made it possible to study the gap acceptance behavior at multiple roundabouts with varying size and shapes (see Figure 8).

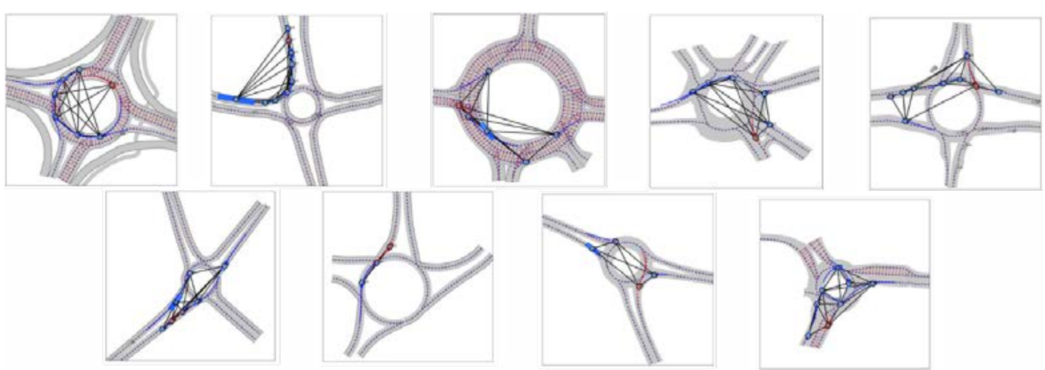

Figure 8: Roundabouts used in our gap acceptance analysis, including several from [6] and [7].

We found the most reliable variable to separate the accepted and rejected gaps to be arc_distance, defined as distance along the arc of the roundabout, and not the Cartesian distance between the two vehicles. The time gap is as good as the arc_distance gap as the independent variable, which is different from what we observed for the unprotected left-turn case. How do we explain why time gap is a good model variable for the roundabout scenario but not the left-turn scenario? Our conjecture is that human drivers can estimate the distance to the other vehicle reasonably accurately for both scenarios and have no problem estimating the traveling speed of the other vehicle in a roundabout scenario as the vehicle is traveling across the field of view. For the left-turn case, however, the oncoming vehicle simply "looms larger" and estimating its speed is harder for human drivers, especially at a distance.

The data and findings offer great insight in terms of how human drivers make decisions in "gap acceptance" at intersections and roundabouts. They can be the basis for judging whether a robot driver's decision exhibits desirable roadmanship behaviors.

NEXT STEPS

Collecting accurate data across multiple intersections and roundabouts proved to be a key challenge for this research. Data sharing, especially related to other aspects of human decisions and behaviors, may be a worthwhile opportunity for pre-competitive collaborations among vehicle manufacturers, AV technology developers, infrastructure engineers and others engaged in automated vehicle engineering and design.

About Mcity

Mcity at the University of Michigan is leading the transition to connected and automated vehicles. Home to world-renowned researchers, a one-of-a-kind test facility, and on-road deployments, Mcity brings together industry, government, and academia to improve transportation safety, sustainability, and accessibility for the benefit of society.

REFERENCES

1. Fraade-Blanar, Laura, Marjory S. Blumenthal, James M. Anderson, and Nidhi Kalra. “Measuring automated vehicle safety: Forging a Framework.” 2018. https://www.rand.org/pubs/research_reports/RR2662.html
2. Disengagement Reports, State of California Department of Motor Vehicles. https://www.dmv.ca.gov/portal/vehicle-industry-services/autonomous-vehicles/disengagement-reports/
3. [2103.15293] Monocular 3D Vehicle Detection Using Uncalibrated Traffic Cameras through Homography (arxiv.org).
4. Pollatschek, Moshe A., Abishai Polus, and Moshe Livneh. “A decision model for gap acceptance and capacity at intersections.” Transportation Research Part B: Methodological 36, no. 7 (2002): 649-663.
5. Li, Yingfeng, Hao, Haiyan, Gibbons, Ronald B., and Medina, Alejandra. “Understanding Gap Acceptance Behavior at Unsignalized Intersections using Naturalistic Driving Study Data.” Transportation Research Record (2021): 03611981211007140.
6. Krajewski, Robert and Moers, Tobias and Bock, Julian and Vater, Lennart and Eckstein, Lutz. “The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany.” Submitted. https://www.round-dataset.com/
7. Zhan, Wei and Sun, Liting and Wang, Di and Shi, Haojie and Clausse, Aubrey and Naumann, Maximilian and K\"ummerle, Julius and K\"onigshof, Hendrik and Stiller, Christoph and de La Fortelle, Arnaud and Tomizuka, Masayosh. “INTERACTION Dataset: An INTERnational, Adversarial and Cooperative moTION Dataset in Interactive Driving Scenarios with Semantic Maps, arXiv:1910.03088 [cs, eess], 2019. https://interaction-dataset.com/