Driver's Perception Model in Driving Assist

Renzhi Tang
School of Information Science and Technologies
ShanghaiTech University
Shanghai, China
tangrzh@shanghaitech.edu.cn

Zhihao Jiang
School of Information Science and Technologies
ShanghaiTech University
Shanghai, China
jiangzhh@shanghaitech.edu.cn

***Abstract*—Vision is the primary way to perceive the environment during driving. However, due to its low spatial and temporal resolution, a driver may fail to perceive agents on the road, which may lead to collisions. Modern vehicles are equipped with sensors that can better perceive the driving environment, as well as ADAS to provide driving assist. However, ADAS does not consider the driver's perception, which may result in unnecessary warnings or actions against the driver's will. These false-positives may cause distractions and confusions in complex driving scenarios, which pose safety threat. In this project, we proposed a driving assist system which can reduce the number of unnecessary warnings by taking into account the driver's perception of the driving environment. The driver's perception model combines estimation of driving environment update and driver's observation. The driver's observation is obtained from gaze tracking and the driving environment update is estimated based on the last observation. In this paper, we formulated inference problem on the driver's perception, and developed a virtual driving simulator to evaluate the feasibility of the system.**

***Index Terms*—Gaze Tracking, Perception Model, State Estimation**

I. Introduction

Driving is a social activity that involves extensive interactions with other agents on the road. Human driver uses vision as the primary source of perception, which cannot fully capture and track all the agents in complex driving scenarios. Failing to correctly perceive other agents is a primary cause of traffic accidents, and can cause serious injuries or death due to collision [1], [2].

Modern vehicles are equipped with sensors like Lidar that have better spatial and temporal perception resolution than human vision. With the development of Advanced Driver-Assistance Systems (ADAS), vehicles perceive the driving environment based on these sensor data, identify imminent collision threats, warn the driver or even perform preemptive actions (i.e. braking) to avoid dangerous situations. i.e. the blind-spot monitoring system warns the driver when the driver change lane with another vehicle in the blind spot of the target lane. However, the ADAS and the driver are working independently in most driving scenarios. The ADAS may provide unnecessary warnings and/or perform actions that may "surprise" the driver in complex driving scenarios. i.e. the blind-spot monitoring system warns the driver when the driver decides to change lane, while the driver knows the lane change can be performed safely with the awareness of another vehicle in the blind spot. These false positives cause distractions and confusion to the driver, which result in mental fatigue and may raise additional safety concern [3], [4].

These false-positives exist because the ADAS does not know how the driver perceives the driving environment. If the ADAS can infer the driver's perception, the ADAS can provide more precise warning by comparing the driver's perception with its own perception. i.e. the ADAS can warn the driver when an agent poses imminent collision threat only if the agent is not perceived by the driver.

Cognitive models of the driver performing driving tasks have been extensively studied [4], [5]. However, these cognitive models are not integrated into the ADAS for real-time driving assistance. Gaze tracking has been used in various research studies to infer the driver's attention level during driving [6]. However, gaze tracking has not been used to infer driver's perception of the driving environment.

In this project, we infer driver's perception using gaze tracking, and propose a driving assist system based on driver's perception model. The driver's perception model represents how the driver perceive the current driving environment, which combines information from observations as well as predictions. The driver's observation of the driving environment is captured by tracking the driver's gaze on the road using a gaze tracking device. The system also mimics the driver's prediction of the unobservable agents in the driving environment based on historical observations. The system then compares the driver's perception model with the car's perception model, which has better observation with less uncertainty. The system only warns the driver if an agent on the road poses imminent collision threat and is not perceived by the driver. The system can effectively reduce the false-positives and unnecessary warnings generated by the ADAS, which can reduce the mental load of the driver in complex driving scenarios, and improves coordination between the driver and the vehicle. In this paper, we formulated inference problem on the driver's perception, and developed a virtual driving simulator to evaluate the feasibility of the system.

II. Problem Formulation

A. Perception Basics

Ground truth

Let's denote the position of car as p which could be a 2D or 3D coordinate and the speed as v. And the state of car can be

978-1-7281-8915-4/20/$31.00 ©2020 IEEE
DOI 10.1109/QRS-C51114.2020.00047

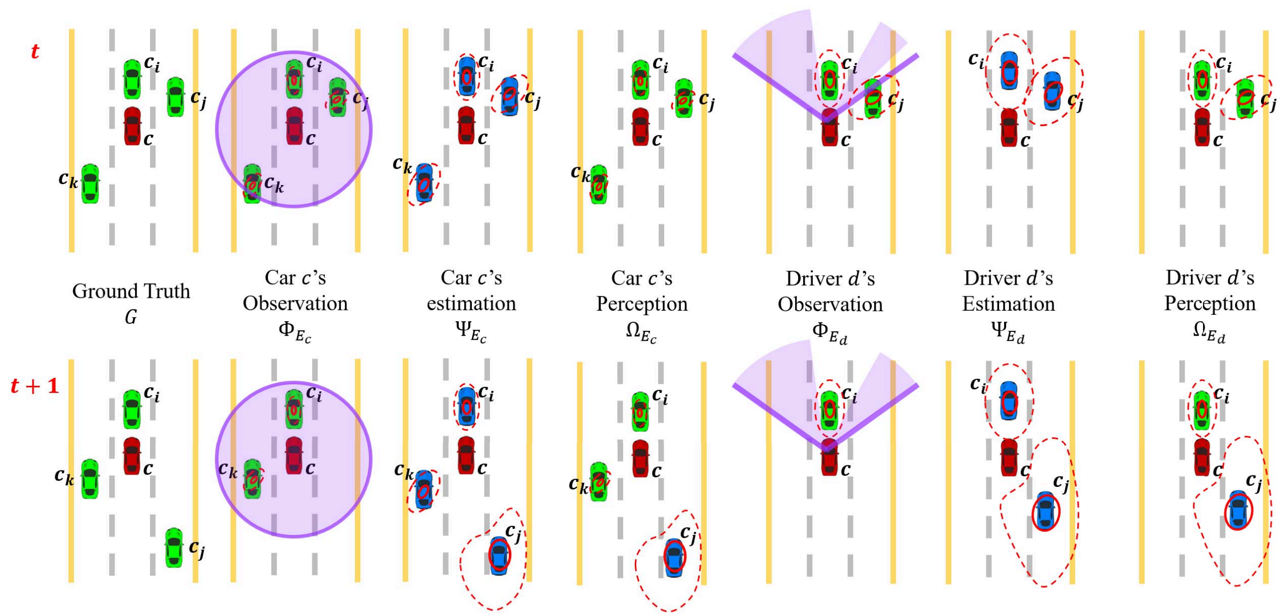

Fig. 1. How ground truth change affects observation (purple areas) and estimation uncertainty (red cicles) of the car and the driver.

represented as $c = \langle p, v \rangle$. The ground truth is defined as the combination of the cars around the ego vehicle c:

$$G_c = \{c, c_0, c_1, ..., c_n\}$$

Entity
An entity (i.e. a person or a vehicle) has the ability to making observation and estimation, which can be defined as:

$$E = \langle S, A \rangle$$

S is a set of sensors. A represents the available actions this entity can perform. The strategy means the probability of taking action a at a specific situation. The action can be in forms of $a \in \{\phi$, slide left, slide right, accelerate, slow down, maintain speed$\}$ in highway driving [7]. i.e. $E_c = \langle S, \phi \rangle$. This represents a traditional car with sensors S equipped on it.

Observation
An observation can be defined as a function which is related to the sensor that makes the observation. For example, sensor A and B may have different error while measuring the same object at the same condition. Here we use Φ to represent an observation andwe define an observation of entity E to measurable value x at time t as function

$$\Phi_{E,t}(x) = h(x_t) + v_t$$

Here v represents uncertainty.

Estimation
An estimation is a partial or a complete prediction to the ground truth. It's based on some previous observation. We use Ψ to represent an estimation. And it can be written as the following form with noise w

$$\Psi_{E,t}(x) = f(x_0, \Phi_{E,x,1:t}) + w_t$$

Perception
For a specific ground truth, different perceptions will lead to different views. For brief, it's a combination of observations and estimations. And we use Ω to represent a perception.

$$\Omega_{E,t}(x) = g(\Psi_{E,t}(x), \Phi_{E,t}(x))$$

The distribution $P(x_t|x_0, \Phi_{E,1:t}(x))$ can be calculated according to Bayes' theorem.

$$P(x_t|x_0, \Phi_{E,1:t}(x)) \propto P(\Phi_{E,t}(x)|x_t)P(x_t|x_0, \Phi_{E,1:t-1}(x))$$

It's similar with Kalman filter. And we can use the similar way to get the uncertainty.

Differences between perceptions
The differences between perceptions in our paper is defined as a function contains the distribution of position and velocity at time t. We use $|\Omega_{E_0,t}(x) \ominus \Omega_{E_1,t}(x)|$ or $|\Omega_{E_0,t}(x)|$ to represent it. By comparing two different perception $\Omega_0(G)$ and $\Omega_1(G)$ at time t, there are 2 situations.

1) $|\Omega_0(c_j) \ominus \Omega_1(c_j)|$
This means there's a difference at c_j in perception Ω_0 and perception Ω_1. For example, c_i in car c's perception $\Omega_{E_c,t+1}$ and driver d's perception $\Omega_{E_d,t+1}$ in figure 1.
2) Do not exist $\Omega_0(c_j)$ or $\Omega_1(c_j)$
We call this situation a loss of c_j in Ω_0 or Ω_1. In figure 1, there's a loss of c_k in driver d's perception $\Omega_{E_d,t}$ compared with car c's perception $\Omega_{E_c,t}$. And for this situation, the difference between 2 perceptions comes from $\Omega_0(c_j)$ or $\Omega_1(c_j)$ only. It can be represented as $|\Omega_0(c_j)|$ or $|\Omega_1(c_j)|$.

In figure 1, we shows the driver d and the car c are the two entities defined as $E_d = \langle S_d, A_d \rangle$ and $E_c = \langle S_c, \phi \rangle$ for

S_d = {eyes, ears, hands, ...}, S_c = {Lidar, cameras, IMU, ...}. Ground truth can be defined as

$$G = G_c = \{c, c_i, c_j, c_k\}$$

Let us take car c at time t as an example. Car c's observation is

$$\Phi_{E_c,t}(G) = \{\Phi_{E_c,t}(c), \Phi_{E_c,t}(c_i), \Phi_{E_c,t}(c_j), \Phi_{E_c,t}(c_k)\}$$

And its estimation is represented as

$$\Psi_{E_c,t}(G) = \{\Psi_{E_c,t}(c), \Psi_{E_c,t}(c_i), \Psi_{E_c,t}(c_j), \Psi_{E_c,t}(c_k)\}$$

Thus,car c's perception is

$$\begin{aligned}\Omega_{E_c,t}(G) &= g(\Psi_{E,t}(G), \Phi_{E,t}(G)) \\ &= \{\Phi_{E_c,t}(c), \Phi_{E_c,t}(c_i), \Phi_{E_c,t}(c_j), \Phi_{E_c,t}(c_k)\}\end{aligned}$$

Here, we use observation Φ_E to update the estimation Ψ_E based on previous observation $\Phi_{E,1:t-1}$. The update mostly is not complete as we only observe a part of ground truth with noise. For the agents we haven't observed, we use the estimation in perception. And the rest observed part, we use the estimation updated by observation in perception. As figure 1 shows at time $t+1$, car c has no observation on car c_j. So in perception $\Omega_{E_c,t+1}$,we have estimated c_j and updated c_i and c_k.

B. Inference of the Driver's Perception from Gaze Tracking

In our example, driver d's observation is $\Phi_{E_d}(G)$. And at time t,

$$\Phi_{E_d,t}(G) = \{\Phi_{E_d,t}(c), \Phi_{E_d,t}(c_i), \Phi_{E_d,t}(c_j)\}$$

When reaching time $t+1$, it turns to

$$\Phi_{E_d,t+1}(G) = \{\Phi_{E_d,t+1}(c), \Phi_{E_d,t+1}(c_i)\}$$

Car c_j has moved out from driver d's sight. He could only see c_i while looking forward. Therefore, driver d's perception at time $t+1$ has updated $\Psi_{E_d,t+1}(c_i)$ with $\Phi_{E_d,t+1}(c_i)$ and kept estimation on c_j which is $\Psi_{E_d,t+1}(c_j)$. As a result

$$\Omega_{E_d,t+1}(G) = \{\Phi_{E_d,t}(c), \Phi_{E_d,t}(c_i), \Psi_{E_d,t}(c_j)\}$$

C. Smart Collision Warning Using the Driver's Perception

First of all, a collision will happen in T is defined as ξ_T. And $e_{E_d,a}$ represents the event that driver d takes action $a \in$ {slide left, slide right, accelerate, slow down, maintain speed}. e_{G,Ω_0,Ω_1} represents the event that in the driving scenario with ground truth G and perceptions Ω_0, Ω_1. We will give a warning on action a in the situation below

$$e_{E_d,a} \wedge e_{G,\Omega_0,\Omega_1} \rightarrow \xi_T$$

A simple way to judge the left side of the logic expression is to calculate the probability of the event that if driver d takes action a leads to getting into a larger $|\Omega_0 \ominus \Omega_1| > \epsilon$ area. Here ϵ is a threshold. In figure 1, this may means car c takes action $a = slide\ left$. And the larger $|\Omega_{E_c,t+1} \ominus \Omega_{E_d,t+1}| > \epsilon$ area is occupied by car c_k which has never been seen by driver d.

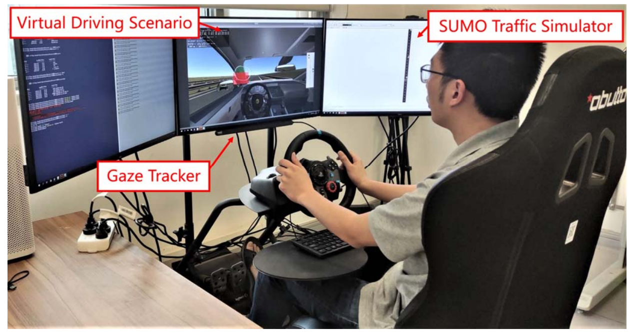

Fig. 2. The virtual cockpit for system evaluation

III. Research Roadmap

In this project, we adapt to the following principles:

- From Virtual to Real
  - **Virtual driving environment to real driving environment:** The system is first evaluated in virtual environment. The results obtained from virtual environment are reproducible and more interpretable.For safety-critical systems, the system can be evaluated in more extreme conditions without causing real world damage.
  - **Virtual driver to real driver**: The ground truth of a driver's perception during the evaluation cannot be determined. In this project, we construct a virtual driver who uses a camera as perception input, and make action decisions based on the perception model. After demonstrating the effectiveness with the virtual driver, the system can then be evaluated with real drivers.
- From Simple to Complex: In this project, we first prove the feasibility of our system in simple driving conditions like highway driving, and gradually increase the complexity to reflect real world driving conditions. The complexity of the virtual driver will also be gradually increased to mimic the behaviors of a real driver.

IV. Experimental Setup

In this section, we introduce our early effort to evaluate the feasibility of the driver assist system.

A. Virtual Driving Simulator

Driving simulators have been widely used to evaluate driving safety [8]. As shown in Fig. 2, a driving cockpit has been set up to mimic real driving scenarios. The virtual driving environment is constructed using the Unity game engine. Drivers can operate the ego vehicle in the driving environment using Logitech G29 gaming controller. Dynamic traffic is simulated in SUMO driving simulator [9] with identical road map as in the Unity environment. Locations of the agents in SUMO simulator are mapped to the Unity environment and the location of the ego vehicle is mapped to the SUMO simulator to achieve closed-loop interactions. A gaze tracker is

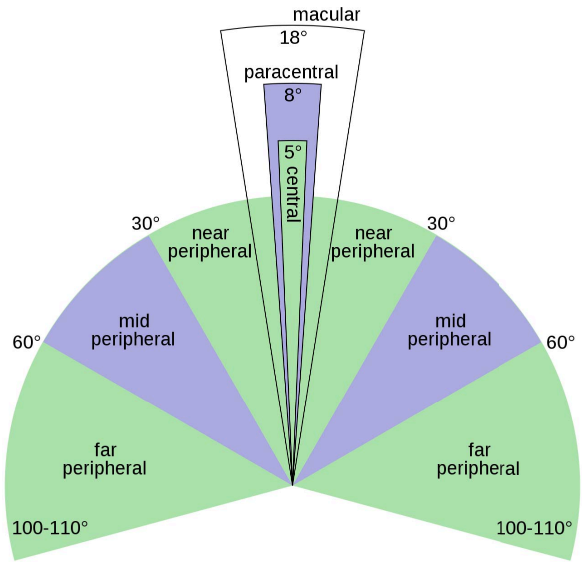

Fig. 3. Human vision regions

used to monitor the driver's observation in the virtual driving environment. The driving simulator can be used to collect data on the driver's observation of the driving environment as well as the driver's actions under different driving context.

B. *The Virtual Driver*

In the early stage of system development, a virtual driver is developed to evaluate our system. A camera inside the ego vehicle in the Unity environment is used to represent the eye of the virtual driver. Agents captured within the vision of the camera are regarded as observations of the driver. An estimation model is developed to mimic how human driver estimates the existence and location of unobserved agents on the road. i.e. when an agent move out of sight, the driver still estimates the movement of the agent using the state at the last observation. The virtual driver's perception combines the estimation and the observation, and will be used to determine the next action by a decision-tree based controller.

V. Summary and Future Work

In this project, we proposed a driving assist system which can reduce the number of unnecessary warnings by taking into account the driver's perception of the driving environment. In this paper, we formulated inference problem on the driver's perception, and developed a virtual driving simulator to evaluate the feasibility of the system.

Peripheral vision is known to have less accuracy compared to macular vision [10]. As the next step, we will assign different uncertainties on the peripheral vision and the macular vision to achieve more accurate representation.

In [7], Chen et al. proposed to use driver's behavior model to predict driver's actions in different driving context. However, the driving context was assumed to be the ground truth. With our driver's perception model as driving context, the driver's behavior model can be more accurate.

References

[1] X. S. Zheng and G. W. McConkie, "Two visual systems in monitoring of dynamic traffic: Effects of visual disruption," *Accident Analysis and Prevention*, vol. 42, no. 3, pp. 921 – 928, 2010.

[2] R. MA and N. LM, "Effects of verbal and spatial-imagery tasks on eye fixations while driving." *J Exp Psychol Appl*, vol. 6, no. 2, pp. 31–43, 2000.

[3] L. Qin, Z. R. Li, Z. Chen, M. A. Bill], and D. A. Noyce, "Understanding driver distractions in fatal crashes: An exploratory empirical analysis," *Journal of Safety Research*, vol. 69, pp. 23 – 31, 2019.

[4] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: Less than the sum of its parts," *Accident Analysis and Prevention*, vol. 42, no. 3, pp. 881 – 890, 2010.

[5] J. He, J. S. McCarley, and A. F. Kramer, "Lane keeping under cognitive load: Performance changes and mechanisms," *Human Factors*, vol. 56, no. 2, pp. 414–426, 2014.

[6] P. Konstantopoulos, P. Chapman, and D. Crundall, "Driver's visual attention as a function of driving experience and visibility. using a driving simulator to explore drivers' eye movements in day, night and rain driving," *Accident Analysis and Prevention*, vol. 42, no. 3, pp. 827 – 834, 2010.

[7] X. Chen, E. Kang, S. Shiraishi, V. M. Preciado, and Z. Jiang, "Digital behavioral twins for safe connected cars," in *Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, ser. MODELS '18, 2018, p. 144–153.

[8] L. N. Boyle and J. D. Lee, "Using driving simulators to assess driving safety," *Accident Analysis and Prevention*, vol. 42, no. 3, pp. 785 – 787, 2010.

[9] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. [Online]. Available: https://elib.dlr.de/124092/

[10] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: A review," *Journal of Vision*, vol. 11, no. 5, pp. 13–13, 12 2011.