EYE GAZING ENABLED DRIVING BEHAVIOR MONITORING AND PREDICTION

Xiaoyi Fan, Feng Wang†, Yuhe Lu*, Danyang Song*, Jiangchuan Liu**

*School of Computing Science, Simon Fraser University, Canada
†Department of Computer and Information Science, The University of Mississippi, USA
xiaoyif@sfu.ca, fwang@cs.olemiss.edu, {yuhel, arthur_song, jcliu}@sfu.ca

ABSTRACT

Automobiles have become one of the necessities of modern life, but also introduced numerous traffic accidents that threaten drivers and other road users. Most state-of-the-art safety systems are passively triggered, reacting to dangerous road conditions or driving behaviors only after they happen and are observed, which greatly limits the last chances for collision avoidances. Therefore, timely tracking and predicting the driving behaviors calls for a more direct interface beyond the traditional steering wheel/brake/gas pedal.

In this paper, we argue that a driver's eyes are *the interface*, as it is the first and the essential window that gathers external information during driving. Our experiments suggest that a driver's gaze patterns appear prior to and correlate with the driving behaviors for driving behavior prediction. We accordingly propose `GazMon`, an active driving behavior monitoring and prediction framework for driving assistance applications. GazMon extracts the gaze information through a front-camera and analyzes the facial features, including facial landmarks, head pose, and iris centers, through a carefully constructed deep learning architecture. Our on-road experiments demonstrate the superiority of our GazMon on predicting driving behaviors. It is also readily deployable using RGB cameras and allows reuse of existing smartphones towards more safely driving.

Index Terms— Gaze, Driving Assistant, Mobile Computing, Deep Learning

1. INTRODUCTION

Automobiles have become one of the necessities of modern life and deeply penetrated into our daily activities. They unfortunately also introduce numerous social problems, among which traffic accidents are most notoriously threatening automobile drivers and other road users. Besides well-developed passive safety equipments such as belt and air bag, active automobile safety systems are also under rapid development in recent years. They use positioning devices, built-in cameras, or laser beams to identify potentially dangerous events, so as to avoid imminent crashes. According to U.S. data [1], systems with automatic braking can reduce rear-end collisions by an average of 40%.

Despite being referred to as *active*, most of these systems remain passively triggered by a vehicle's surroundings and its driving interface (i.e., steering wheel, brake, and gas pedal) [2][3]. Such systems react to dangerous road conditions or driving behaviors only after they happen and are observed. Given the well-known *two-second rule*[1], such passive reaction can greatly limit the last chances for collision avoidances. For example, an alert from a *Blind Spot Warning* system occurs after the driver turns the steering wheel, which, on a highway, can be too late to avoid a collision if the speed is over 120 km/h. The *Adaptive Front-lighting* system, which has been developed to enhance night visibility, also follows the angle change of the steering wheel and accordingly changes the lighting pattern to compensate for the curvature of a road. The lag from steering wheel movement to light movement, however, is not negligible (being activated after 1/4 turn of the wheel and sometimes one or two full turns).

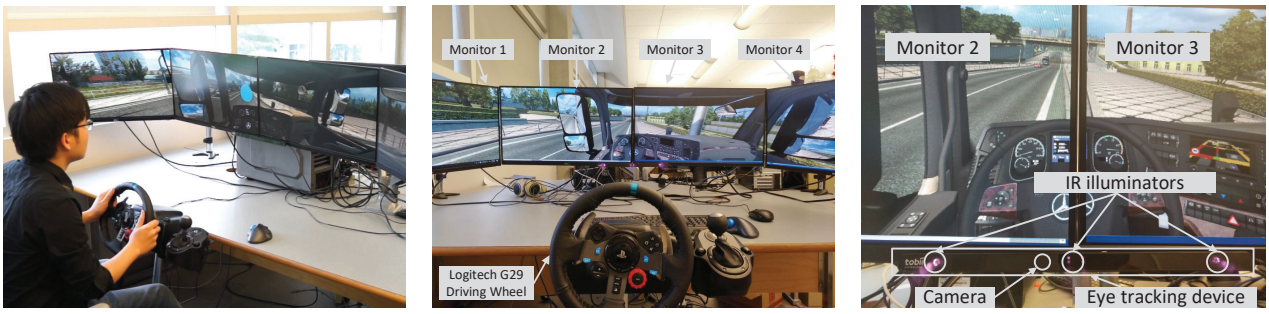
In short, timely tracking and predicting the driving behaviors is essential and important towards improving driving safety, and we need a new and more direct interface beyond the traditional steering wheel/brake/gas pedal. We argue that a driver's eyes are *the interface*, as this is the first and the essential window that gathers external information. Our crowdsourcing measurements reveal strong correlations between the eye-gazing patterns and the driving behaviors, which are further confirmed by our on-road experiments to be discussed later. In particular, gaze patterns occur prior to the corresponding driving behaviors, which offers a great chance to overcome the two-second rule.

To this end, we develop `GazMon`, an active driving behavior monitoring and prediction framework for driving assistance applications. GazMon extracts the gaze information from a front-camera and predicts driving behaviors based on the gaze patterns. The patterns are analyzed through a supervised deep learning architecture. In particular, we incorporate a joint Convolutional Neural Network (CNN) and Long Short

This work is supported by an NSERC Discovery Grant and an NSERC E.W.R. Steacie Memorial Fellowship. This research is partly supported by an NSF I/UCRC Grant (1539990).

[1] A driver usually needs about two seconds to react to avoid accident.

978-1-5386-1737-3/18/$31.00



(a) Driving Simulator (b) Four Monitors and Driving Wheel (c) Eye Tracking Device

Fig. 1: GazMon immersive emulating environment

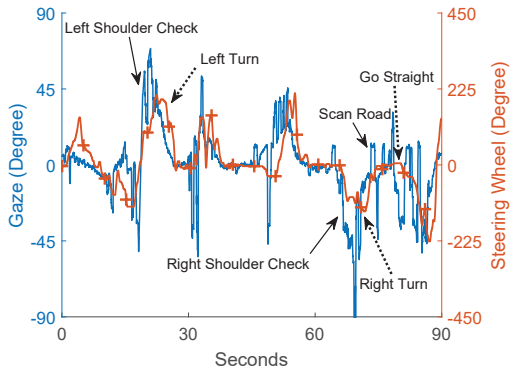


Fig. 2: Gaze patterns appear prior to driving behaviors

Term Memory (LSTM) network, which first identifies low-level activities, and then scales up to predict complex high-level driving behaviors.

Our GazMon does not rely on very advanced and high-cost eye tracking devices (e.g., Tobii EyeX[2]). It is readily deployable using RGB cameras and can be easily integrated to intelligent in-vehicle systems, e.g., CarPlay and Android Auto, minimizing/reducing the reliance on extra hardware. It also allows the reuse of existing smart phones for driving behavior prediction. Our GazMon demonstrates that a careful design can turn a smartphone from an accident contributor into a crash preventer. With GazMon, driving applications can warn and return feedbacks to drivers without distracting them, e.g., through voice instructions, to improve the safety. We have deployed the trained deep learning models of GazMon with Mobile TensorFlow on Android smart phones, e.g., Google Pixel and Vivo X9 Plus. We conduct extensive on-road experiments for driving behavior prediction, which also provide additional feedbacks to GazMon to fine-tune the deep learning model. The evaluation results report significant prediction accuracy improvements over different state-of-art solutions.

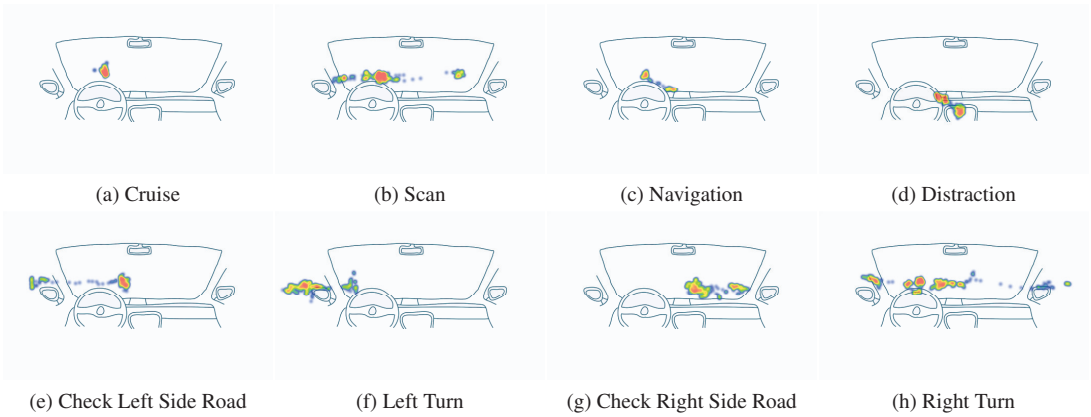
[2]https://tobiigaming.com/

2. WHY WE INCORPORATE GAZE PATTERNS INTO DRIVING PREDICTION?

In this paper, we explore the opportunities to predict driving behaviors through analyzing drivers' gaze patterns. We seek to first answer the following question: *Do a driver's gaze patterns appear prior to the driving behaviors?* To investigate the correlations between them, we capture the driver's gaze patterns and the steering wheel through our testbed. For safety concerns, our testbed runs in a virtual reality environment as shown in Fig. 1(a). The driving simulator platform runs on a customized PC, which is connected to four 27-inch monitors as shown in Fig. 1(b), where the NVIDIA Surround Technology enables to combine displays to create the most immersive emulating environment. As illustrated in Fig. 1(c), we choose Tobii eyeX 4C[3] as the eye-tracking device to collect the users' gazing data due to its affordable price for our testbed, suitable sampling rate, and reasonable accuracy. The eye-tracking device consists of three illuminators and one camera, where the illuminators create the pattern of near-infrared light on viewer's eyes, and the camera captures high-resolution images of the driver's eyes and the patterns. In this simulation platform, volunteers play a driving simulation game, namely *Euro Truck Simulator 2*, which makes people feel as driving a vehicle in real life. We record a driver's behaviors with the gaming wheel and pedals set and capture the driver's gazing patterns with the eye-tracking device.

We perform experiments over 50 experienced drivers on the gaze patterns to explore their potential relationships with the driving behaviors. The results reveal that the driver's gaze patterns appear prior to the drivers' behaviors, thus opening new opportunities to explore. Fig. 2 shows a typical example of the gaze patterns collected from a volunteer and the steering wheel turning behaviors, which is the most important feature in driving a vehicle. We plot the driver's gaze patterns at the horizontal direction, where a positive degree means that the driver is looking on the left and a negative one means looking on the right. And the steering wheel turning behavior is plotted in a similar way. It is clear to see that the gaze

[3]https://tobiigaming.com/eye-tracker-4c/



(a) Cruise (b) Scan (c) Navigation (d) Distraction

(e) Check Left Side Road (f) Left Turn (g) Check Right Side Road (h) Right Turn

Fig. 3: Gaze Patterns from Typical Driving Behaviors

patterns highly correlate with the driving behaviors but come ahead of some time advance, e.g. shoulder check comes prior to left turn for about 10-15 seconds. We count the time gap that gaze patterns appear prior to driving behaviors, which is approximately 5.09 seconds on average and large enough for the two-second rule to apply.

Then we need to answer the following question: *How a driver's gaze patterns correlate with the driver's behaviors?* As we know, the single gaze point is ineffective to predict driving behaviors. Our experiments reveal that if we stack the gaze points across a small time interval into a vector, then this vector can be a good indicator of different driving behaviors. Fig. 3 shows the gaze patterns from eight typical driving behaviors, i.e., cruising, scanning, looking at navigator, distracting, checking left side road, left turn, checking right side road and right turn. This example shows that gaze patterns are distinct with different driving intentions, and we can predict the driving behaviors through analyzing gaze patterns.

3. SYSTEM IMPLEMENTATIONS

Our GazeMon framework does not rely on a particular eye-tracking hardware. In the long run, advanced eye tracking solutions could be seamlessly integrated into the vehicles' on-board systems with affordable cost, and our GazeMon will benefit from it. On the other hand, we also note that nowadays mobile phones are ubiquitous and widespreadly used, where more than a third of the world's population is estimated to have smartphones by 2019. Given that people carry their phones everyday everywhere, the phones have great potentials to serve as eye gazing tools in vehicular environments, since mobile phones can directly capture images from the front RGB camera and require no modifications to the existing on-vehicle systems. Another benefit is that the high adoption rate of technology upgrades on mobile phones can lead to rapid development and deployment of new camera technology and allow the use of computationally expensive methods.

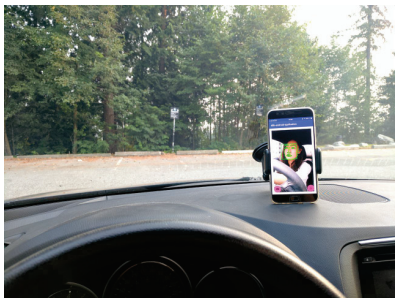


Fig. 4: Smartphone in the real-world experiment

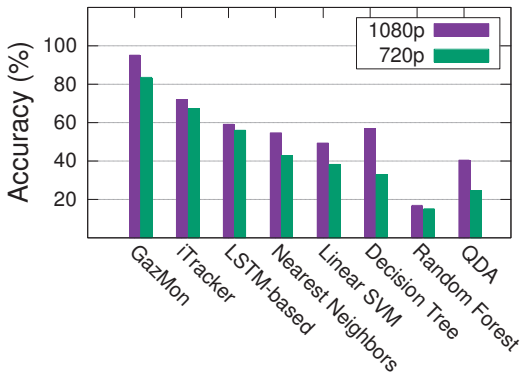


Fig. 5: Overall Performance of GazMon System

Our GazMon is the first attempt towards this direction, which can achieve high prediction accuracy in a timely manner as later demonstrated by our on-road experiments.

The mobile phone part of GazMon is implemented as an app on Android OS 5.1.1 as shown in Fig. 4. On startup, the GazMon app launches an Android activity (CameraActivity.java) which basically accesses the camera by using the Android Camera2 package. Then GazMon uses the supported JNI (Java Native Interface) procedures to interact with dlib-

Table 1: The Accuracy of Driving Behavior Prediction versus Prediction Gap

	Cruise			Left Turn			Right Turn			Left Line			Right Line		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
1	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00
2	1.00	1.00	1.00	0.98	1.00	0.99	1.00	0.98	0.99	0.96	1.00	0.98	1.00	0.97	0.99
3	0.97	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.99	0.99	0.97	0.98
4	0.96	1.00	0.98	1.00	1.00	1.00	1.00	0.95	0.98	0.98	0.98	0.98	0.99	0.99	0.99
5	0.96	0.87	0.91	0.94	0.92	0.93	0.88	0.94	0.91	0.88	0.98	0.93	0.95	0.90	0.92
6	0.96	0.98	0.97	0.84	0.96	0.90	0.89	0.91	0.90	0.84	0.91	0.88	0.97	0.78	0.86
7	0.94	0.93	0.94	0.93	0.61	0.74	0.98	0.94	0.96	0.94	0.98	0.96	0.75	0.96	0.84
8	0.98	0.98	0.98	0.85	0.57	0.68	0.97	0.90	0.93	0.83	0.96	0.89	0.72	0.87	0.79

android engine and the recent proposed dlib library to extract a sequence of eye gazing features including facial landmarks, head pose, and iris centers from the incoming image stream.

In training stage, GazMon uploads the drivers' videos with the preprocessed eye gazing features in a batch to the server, when the high-speed wireless connection is available. The preprocessed eye gazing features are used for training the deep learning architecture, where the ground truth of the driving behaviors is labeled based on the videos from the front cameras. The server part of the GazMon is deployed on our customized desktop, where our deep learning architecture incorporates a joint Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) network, so as to first identify low-level activities, and then scale up to predict complex high-level driving behaviors. The CNN and LSTM classifiers are implemented in Keras with cuDNN on Dual Nvidia GTX 1080Ti GPUs.

In prediction stage, the GazMon app running on a smartphone can timely process the images captured by the device's camera and predict the driving behaviors based on the deep learning architecture pre-trained by the aforementioned approach, so as to provide realtime services to users, where the preprocessed eye gazing features are fed into TensorFlow Mobile's core engine implemented by Google developers.

4. ON-ROAD EXPERIMENTS

Tab. 1 shows the details of prediction accuracy in precision (P), recall (R) and F-Score (F) of our *GazMon* approach, where each column denotes the driving activity performed and each row represents the prediction time gap (as shown at the beginning of each row). As shown in the table, the precision of driving behavior prediction is at least 0.96 in 4 seconds, which indicates that GazMon can allow 200% of the gap required by the two-second rule and still distinguish various driving behaviors with high accuracy. We thus use 4 seconds as the default predicted time gap for the remained experiments. We also observe that the left lane change (LL) has better prediction accuracy than the right lane change (RL) in longer predicted time gap, because the left lane change takes longer time as the vehicle needs accelerate to merge into the left lane. When the predicted time gap is larger than 5 seconds, the prediction accuracy decrease for the right lane change (RL) and left turn (LT). This is because the experienced drivers always have right shoulder check before both of those behaviors. If the predicted time gap is too large, it will cause that the prediction is mainly based on the right shoulder check and thus cannot well distinguish these two behaviors.

Fig. 5 shows the performance of our GazMon compared with different state-of-the-art approaches. To this end, we implement five commonly used classifiers (k-Nearest Neighbors, one-vs-all Linear SVM, Decision Tree, Random Forest, and Quadratic Discriminant Analysis) as well as the CNN based approach used in iTracker [4] and a LSTM based approach. The result clearly shows that GazMon can achieve 22% higher accuracy than iTracker that only uses CNN. This demonstrates the benefits of the LSTM architecture used in GazMon on learning dynamic temporal relationships from a sequential spectrum frames for driving behavior prediction. At the same time, GazMon also obtains 36% higher accuracy than the LSTM-based approach, which illustrates the necessity of the CNN architecture used in GazMon to efficiently extract the features for driving behavior prediction. Our GazMon also outperforms the other five commonly used classifiers, achieving 40% gain over the best approach (SVM) among them.

5. REFERENCES

[1] "Crashes avoided: Front crash prevention slashes police-reported rear-end crashes," *The Insurance Institute for Highway Safety (IIHS) Status Report*, vol. 51, no. 1, January 28, 2016.

[2] Cagdas Karatas, Luyang Liu, Hongyu Li, Jian Liu, Yan Wang, Sheng Tan, Jie Yang, Yingying Chen, Marco Gruteser, and Richard Martin, "Leveraging wearables for steering and driver tracking," in *Proceedings of IEEE INFOCOM 2016*.

[3] Dongyao Chen, Kyong-Tak Cho, Sihui Han, Zhizhuo Jin, and Kang G Shin, "Invisible sensing of vehicle steering with smartphones," in *Proceeding of ACM MobiSys 2015*.

[4] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba, "Eye tracking for everyone," in *Proceedings of IEEE CVPR 2016*.