Probabilistic Estimation of the Driver's Gaze from Head Orientation and Position

Sumit Jha and Carlos Busso
Multimodal Signal Processing (MSP) Lab, Eric Jonsson School of Engineering and Computer Science
The University of Texas at Dallas, Richardson, Texas 75080, USA
Email: sumit.jha@utdallas.edu, busso@utdallas.edu

***Abstract*—Visual attention is an important factor when studying driver behavior. While the location of the pupil can provide accurate information about gaze, the vehicle environment poses challenges that prevent the use of off-the-shelf gaze detection algorithms in the car. Head pose can be used to approximate the driver's visual attention, providing a coarse estimate which can be good enough for many applications. However, the relation between head pose and gaze is not one-to-one, depending on the driver, cognitive load, and visual task. Instead of detecting a precise gaze direction, this paper proposes a novel approach which creates a probabilistic map describing visual attention. The approach relies on *Gaussian process regression* (GPR), which takes the position and orientation of the drivers' head to estimate the probability that the driver is looking at a given point. The approach creates confidence regions describing the most likely gaze directions. We evaluate the proposed approach with naturalistic recordings in real roads, where we estimate the position and orientation of the driver's head using a headband with fiducial markers. The experimental evaluation demonstrates that 89.4% of drivers' gaze are included in the 95% confidence region predicted by our model. The proposed system can provide valuable information for navigation, infotainment, safety and communication systems.**

I. Introduction

The safety of a vehicle is a major concern for the automobile industry, where lives can be saved by developing new tools for *advanced driver assistance systems* (ADAS). At the core of safety is the driver. According to the survey conducted by the *National Highway Traffic Safety Administration* (NHTSA) [24], the reason responsible for 94% of the total crashes in The United States was human error. Therefore, it is important to track the behaviors of the driver, detecting lack of attention and lack of situational awareness. This safety capability is even more relevant as we transition to autonomous vehicles, especially during *Level 2* and *Level 3* of autonomous vehicles [19], where the driver and vehicle share the control of the car.

It is important to consider the visual attention of the drivers to monitor their awareness. Driving primary relies on visual resources to maneuver the vehicle. Therefore, it is important to assess where the visual attention of the driver is directed. The knowledge of the driver's visual attention can help the system to infer situational awareness [17], the driver's cognitive state [16], and visual distractions [14], [18]. While there are various gaze detection solutions for controlled settings [3], [15], detecting gaze in the car is a challenging problem due to illumination changes, occlusions and extreme head rotations. A simplified solution is to rely on head pose to predict the driver's visual attention. Detecting head pose in driving environment is still a challenge [10], however, it is a more feasible computer vision task. A coarse solution for gaze detection based on head pose estimation is enough in most in-vehicle applications (e.g., mirror checking [17] and lane changing [6]). Most relevant studies in the field have used gaze zone classification to predict pre-defined regions of gaze in the driving environment [5], [13], [25]. However, the mapping between head pose and gaze is many-to-many, depending on the driver, his/her cognitive load, and the underlying visual task [9]. It is important to develop better solutions to estimate the visual attention of the driver from his/her head poses.

This paper proposes a novel approach to predict the gaze region. Instead of attempting to predict the exact position that the driver fixates his/her attention, we redefine the problem as a probabilistic mapping, where the task is to create a saliency map to characterize visual attention. The approach relies on *Gaussian process regression* (GPR), which takes as input the position and rotation of the driver's head, predicting the probability that the driver is looking at any given point. The probabilistic estimation of the driven's gaze provides confidence regions describing visual attention, which can be valuable information for navigation, infotainment, safety and communication systems.

We evaluate the approach with recordings collected with the UTDrive platform [1], where drivers were asked to look at predefined markers in the windshield. We obtain precise estimations of the position and rotation of the driver's head with a headband containing fiducial markers. The experimental evaluation shows that the driver's gaze lie 89.4% within the 95% confidence region predicted by our model.

II. Related Work

A driver obtains the majority of the information to perform the driving task through visual cues. Hence, it is important to know where the driver is looking at to evaluate the performance of the driver, predict what is he/she going to do next, and understand ambiguous commands. Many studies have used the visual pattern of the driver to understand their behaviors. Robinson et al. [23] studied the visual search pattern of drivers in different sections of the road. Their study demonstrated that glance patterns strongly depend on the driving tasks, showing longer searching times at stop signs, and shorter searching times for maneuvers such as changing lanes. Underwood and

978-1-5386-1526-3/17/$31.00 ©2017 IEEE

(a) UTDrive platform (b) Layout of Markers

Fig. 1. (a) Vehicle used for data collection, and (b) markers placed at the windshield (1-13), mirrors (14-16), side windows (17-18), speedometer panel (19), radio (20), and gear (21). The subjects were asked to look at these markers.

Crundall [26] compared the visual search pattern in novice and experienced driver, showing that experienced drivers display different visual search patterns depending on the type of road. In contrast, novice drivers presented similar pattern across types of road. These studies showed that visual attention depends on the driving task and the driver's expertise.

Previous studies have proposed different ADAS that incorporates the driver's gaze as a cue to study inattentive drivers. Fletcher and Zelinsky [8] predicted inattention by studying the relationship between road signs and driver's gaze, which was estimated with the commercial software faceLAB™. This study considered recording in real driving conditions. Chien et al. [4] proposed an AdaBoost classifier to track the driver's eyes and estimate the gaze. The face orientation was obtained from the location of the eyes and nose and was adjusted using the pupil location. Using the orientation, the horizontal gaze angle was detected between $-\pi/4$ and $+\pi/4$ with $+\pi/16$ resolution. Ji and Yang [11] proposed the use of IR sensors to predict the driver's gaze, but the evaluation considered simulations. It is not clear how the algorithms will work in real-world driving scenarios, where occlusions (e.g., use of glasses, hands), non-frontal faces, and variant lighting conditions challenge the robust estimation of gaze.

A coarse estimation of the driver's visual attention is often sufficient for various practical scenarios. The driver's head pose can be helpful for this purpose, which is significantly easier to estimate. Zhang et al. [27] suggested that head pose alone can provide good information about driver's intentions as compared to the eye movement. Several studies have used different head pose estimation systems to classify the driver's gaze into several predetermined areas [5], [13], [25]. Rezaei and Klette [22] used the head pose of the driver as a feature to estimate the driver's attention. Doshi and Trivedi [6] studied the driver's head pose and eye movement behavior before changing lanes concluding that head pose is a good feature to predict the driver's intention to change lanes. They followed up this study by considering the causes that triggered gaze shifts, focusing on planned actions (e.g., glances before and after conducting driving tasks), and stimulus-driven actions (external objects such as other vehicles and pedestrians) [7]. For gaze fixation, head pose and gaze were highly correlated for both scenarios.

Instead of estimating the precise location of the drivers' gaze

(a) Dash camera

(b) Driver's camera (c) Car's camera

Fig. 2. Data collection. (a) Dash camera (Blackvue) used for data collection, (b) frame from the driver's camera, (c) frame from the road's camera.

using head pose, which depends on the driver, the task, and the cognitive load of the drivers, we propose to create a probabilistic visual map defining regions of confidence characterizing where the driver may be looking at. This concept is intuitive, and novel, offering contextual information that is relevant for a number of in-vehicle applications for security, infotainment and navigation. To the best knowledge of the authors, this is the first study that models the relationship between gaze and head motion using a probabilistic visual map.

III. Data Collection

The study relies on the database discussed in Jha and Busso [9]. This data was collected using the UT Drive platform (Fig. 1(a)), where the task for the drivers was to look at 21 predefined locations marked on the windshield, mirrors, side windows, speedometer panel, radio, and gear (Fig. 1(b)). These numbered markers are called at random, and the driver is asked to fixate their gaze on the target point. The road and the drivers are simultaneously recorded using a commercially available two-channel dash camera (Blackvue DR-650GW-2ch - Fig. 2(a)). The data collection has three phases. This study uses the first two phases, which are described below.

In **Phase 1**, the evaluation is conducted while the car is parked. The subject is asked to take the driver's seat and look at each marker five times in random order. This phase helps

(a) AprilTag (b) Headband with AprilTags

Fig. 3. (a) Sample of AprilTags, (b) Headband with AprilTags for robust head pose estimation.

the driver to get familiar with the task in a safe environment. The driver looked at the markers by moving their eyes and head in a natural manner.

In **Phase 2**, the subject is asked to repeat the evaluation while driving the car. The subject is asked again to look at each marker five times in random order. We selected a straight road, which minimized the risk in the evaluation. The driver was asked to look at the markers only when it was safe to conduct this task.

We recruited 16 subjects (10 males, 6 females) for the data collection. Figures 2(b) and 2(c) show examples of frames for the cameras facing the driver and road.

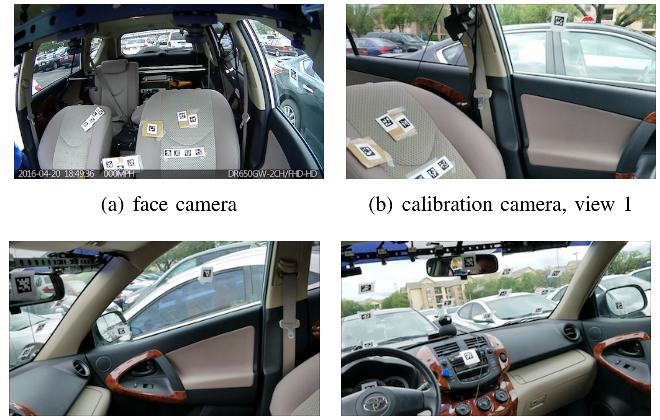
IV. Head Pose Estimation Using AprilTags

The proposed approach uses the position and rotation of the driver's head as input. Ideally, this information can be automatically derived from the camera using computer vision solutions. We have studied head pose estimation algorithms in driving environment [10]. Our analysis demonstrated the complexity of the task due to illumination, extreme head poses and occlusions. State-of-the-art algorithms cannot detect the driver's face in over 20% of the frames in our corpus. While we expect that better head pose estimation algorithms will improve the robustness for in-vehicle applications, the current study requires more accurate estimates. Therefore, we rely on a headband with AprilTags [20] to estimate the head position and orientation. AprilTags (Fig. 3(a)) are fiducial markers that can be robustly detected in an image. It is possible to estimate their position and orientation from 2D images, since their patterns and sizes are known. AprilTags are used in camera calibration, augmented reality and applications with robots.

We designed a headband with 17 different AprilTags as illustrated in the structure shown in Figure 3(b). The subject wore the headband during the data collection. The design of the headband allows the camera to capture AprilTags even for extreme head motion, from which we estimate the position and orientation of the headband. We translate the angles estimated across AprilTags to the central AprilTag, defining a reference for the head position and orientation. We select the median value of the angles provided by AprilTags as the estimate of the head orientation.

V. Methodology

This paper aims to create a probabilistic map describing the driver's visual attention. The approach requires to calibrate the 3D location of the markers relative to the reference coordinate



(a) face camera (b) calibration camera, view 1

(c) calibration camera, view 2 (d) calibration camera, view 3

Fig. 4. Coordinate calibration: Calibration for marker location with the face camera

system. Section V-A discusses the calibration approach used to obtain the 3D location of target markers. The system also requires to normalize the angle to generalize the models across drivers. This normalization seeks to create a driver independent system that maintains performance when evaluated on a new driver, whose recordings were not included during training. Section V-B discusses the speaker-dependent normalization of the the recordings. Section V-C presents the proposed technique to predict the probabilistic map based on *Gaussian Process Regression* (GPR).

A. Calibration of the Target Markers

The method requires estimating the 3D location of the target markers in the windshield, mirrors, side windows, speedometer panel, radio, and gear. This calibration process is done only one time, since the location of the cameras and the markers are fixed during the data collection. We conduct this calibration using AprilTags, by placing a unique AprilTag on each marker (see Fig. 4(d)). A challenge in this process is that most of these AprilTags are visible by neither of the dash-cameras (cameras facing the road and driver). We addressed this problem by adding extra AprilTags at different locations in the visual range of the camera facing the drivers (Fig. 4(a)). The goal is to find the homogeneous transformations between AprilTags, which will allow us to place all the angles in a single coordinate system. After placing extra AprilTags, we took pictures from multiple angles using a portable camera Figures 4(b), 4(c) and 4(d) are examples of these pictures. We also take frames using the camera facing the driver (Fig. 4(a)). Then, we detect the location and angular direction of each AprilTag that is visible in each figure. We determine the relative orientation between two images using the set of AprilTags visible on both images using the Kabsch algorithm [12]. The result of this calibration is the 3D locations of all the markers in the reference coordinate system of the camera facing the driver.

B. Head Position and Orientation Normalization

Drivers have different heights and they have different seating preferences. There are also differences in the placements of the headband across drivers. Therefore, it is important to normalize the 6D head estimates so the models generalize to new drivers. To compensate for these variabilities, we normalize for each driver the angles and positions provided by the AprilTags. For phase 1, we estimate the mean and standard deviation of the head pose estimation while the subject was conducting task associated with this phase. For phase 2, we calculate the mean and standard deviation of the head pose estimation during the time that a subject was driving. This normalization is conducted for each of the 6D parameters using Equation 1.

$$x_{norm} = \frac{x_{obs} - \mu}{\sigma} \tag{1}$$

This approach provides an implicit calibration technique to normalize the data and generalize the models to new drivers for which the system is not trained. In Section V-A, we obtained the 3D location of the markers with respect to the reference system of the camera facing the driver. With the normalization performed on the head pose, the reference axes also change. Therefore, we apply the same mapping operation to the 3D location of the target markers to maintain consistency.

C. Estimation of Probabilistic Visual Map

The relation between head pose and gaze while driving is not one-to-one [9]. The drivers move their eye and head to direct his gaze to the target location. The visual and cognitive load at that moments will dictate the interplay between head and eye movement. As a result, we propose to estimate a visual map that indicate the probability that a driver is looking at a given point. This approach is different from traditional methods that aim to detect the precise gaze of the drivers. We argue that having a probability map for the gaze is more flexible and robust.

The framework for the proposed model is *Gaussian Process Regression* (GPR) [21]. GPR is a probabilistic method that aims to predict the output, which is considered a Gaussian process. Any finite subset in the output space forms a joint Gaussian distribution. The key difference of this framework, compared to other regression models, is that the goal is not to predict a single value, but to predict the probability of the output. To design our model, we obtain a unit vector in the direction between the head and the location of the target gaze. The vector is represented by the horizontal angle θ and the vertical angle ϕ. We train two separate GPR models to predict these angles

The model predicts a Gaussian distribution for the horizontal (θ) and vertical (ϕ) angles describing visual attention of the driver. The distributions are parametrized by the mean and standard deviation of these angles. The mean is obtained from a linear basis function using the 6D head parameters ($\vec{x} = [x, y, z, \alpha, \beta, \gamma]$). The covariance function between two vectors is obtained using the squared exponential kernel function:

$$K(\vec{x_1}, \vec{x_2}) = \sigma_f^2 \exp\left(\frac{-|\vec{x_1} - \vec{x_2}|^2}{2l^2}\right) \tag{2}$$

where, l and σ_f^2 are parameters determined during training process. The probabilistic region provide a coarse estimate of the head pose, where the coarseness depends on the confidence that we have for a given head pose. For example, if we are more confident in the estimation of the gaze for a given head pose, the model will provide a small confidence region for the gaze.

VI. Experimental Evaluation

The evaluation relies on driver dependent partitions for training and testing using *leave-one-out cross-validation* (LOOCV). In each fold, we use 15 subjects to train the system, evaluating the results on data collected from the remaining driver. Therefore, the testing set has recordings from a driver whose data was not included in the training set, which help us to study the generalization of the proposed model. The results correspond to the average values observed across the 16 folds.

We evaluate the approach with the instances where the drivers were asked to look at the target markers, for which we have ground truths for the intended gaze direction. We separately analyze the results for phase 1 (i.e., data collection when the car was parked) and phase 2 (i.e., data collection when the car was moving) of the corpus, training and testing the models in matched conditions.

A. Predicting Location of Target Markers

The first evaluation consists of predicting the location of target markers. For each head pose, we estimate the confidence regions using the probabilistic model for θ and ϕ. We evaluate the 50%, 75% and 95% confidence intervals. Notice that the 50% confidence interval is a subset of the 95% confidence interval, which covers a larger region. The size of these regions depend on the certainty of the model which is a function of the position and rotation of the head (see Fig. 6). Then, we estimate the proportion of target gazes included in a given confidence interval.

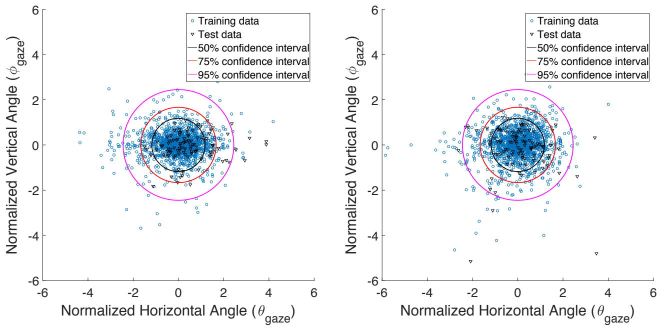
Table I gives the average percentage of samples included in the 50%, 75% and 95% confidence regions of the Gaussian distribution. We include the results from the training set, and from the testing set to study the generalization of the models. We observe that when the subject is driving (phase 2), 89.43% of the true gazes from the test samples are included in the 95% region of the distribution predicted by the model. Even for more restrictive confidence intervals, the approach is able to include the majority of the intended gaze directions from the test set. The values for samples in the training set are higher than the one for the testing set, which suggest that the models can be improved. We expect that increasing the number of drivers in the training set will produce more robust models.

We also study the detection performance for markers in different regions. For this purpose, we divided the markers into three groups. The *left windshield* group includes all the point near the frontal gaze region (i.e., markers 1, 2, 5, 6, 10, and 11 in Fig. 1(b)), the *right windshield* group includes all the markers in the windshield to the right of the driver (i.e., markers 3, 4, 7, 8, 9, 12, and 13 in Fig. 1(b)), and

TABLE I

GAZE ESTIMATION RESULTS FOR PHASE 1 AND 2 OVER THE TRAIN AND TEST SAMPLES. THE RESULTS ARE LISTED FOR EACH *confidence interval* (CI).

C	All Markers		Left Windshield		Right Windshield		Outside Windshield	
	Training Data	Test Data	Training Data	Test Data	Training Data	Test Data	Training Data	Test Data
Phase 1 (while parked)								
50% CI	**77.77%**	**61.34%**	**92.20%**	**83.80%**	**66.98%**	**51.32%**	**76.38%**	**53.22%**
75% CI	**89.45%**	**78.44%**	**97.78%**	**94.81%**	**84.18%**	**73.96%**	**87.80%**	**70.03%**
95% CI	**96.51%**	**90.35%**	**99.17%**	**98.34%**	**95.29%**	**91.25%**	**95.59%**	**83.51%**
Phase 2 (while driving)								
50% CI	**74.52%**	**56.32%**	**80.87%**	**69.44%**	**74.64%**	**56.19%**	**69.59%**	**46.52%**
75% CI	**88.54%**	**76.55%**	**94.82%**	**89.93%**	**89.32%**	**79.83%**	**83.09%**	**63.46%**
95% CI	**96.88%**	**89.43%**	**99.75%**	**97.89%**	**98.09%**	**91.96%**	**93.62%**	**80.67%**



(a) when stopped - phase 1 (b) when driving - phase 2

Fig. 5. Normalized distance between the intended gaze angles and the mean value estimated by the GPR model. The circles correspond to different confidence regions.

the *outside windshield* group includes all the markers in mirrors, side windows, speedometer panel, radio, and gear (i.e., markers 14 – 21 in Fig. 1(b)). We evaluate the performance of the model in these regions, listing the results in Table I. The model has the best performance for markers in the *left windshield* group. The worst performance in terms of accuracy and generalization is for markers outside the windshield. The system is very accurate for regions in front of the drivers, but its performance drops for markers further from the drivers where the relationship between gaze and head pose is more ambiguous.

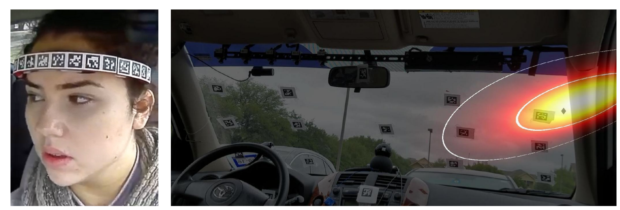
The performance for phase 1 (while parked) is in general better than the one for phase 2 (while driving). The 95% confidence region contained 90.35% of the intended gazes from the test data. Notice that detecting visual attention while driving (phase 2) may be more important than when the car is stopped (phase 1), so our future effort will focus on increasing the performance for phase 2.

To visualize the results, we calculate the distance between the mean of the predicted distribution and the intended target angle for each sample using Equation 3. We can visualize samples for all the markers in a single figure after subtracting the mean and dividing by the standard deviation. Figure 5 shows the confidence intervals and the samples from the training and testing sets for one of the 16 folds. The figure shows that most of the points are included in the confidence intervals.

$$dist(y, model) = \frac{y_{true} - \mu_{model}}{\sigma_{model}} \tag{3}$$

(a) Face camera (b) Mapping on windshield, phase 2

(c) face camera (d) Mapping on windshield, phase 2



(e) face camera (f) Mapping on windshield, phase 2

Fig. 6. Example of mapping the gaze region on the windshield (white ellipses give 50% and 95% CIs). The target marker is highlighted with a black circle.

B. Mapping the Gaze Angles to the Windshield

The probability map can be easily mapped into the windshield. The proposed models estimate the probabilistic map for the angles θ and ϕ. There is not depth information, so it is not possible to predict the exact point along the line where driver's gaze is directed. However, we have the 3D position of the markers, obtained after the calibration (Sec. V-A). The distance between the target marker and the head is used to map the probabilistic map of the gaze into the windshield. First, we map the gaze region in the surface of a sphere. Then, we project this spherical region onto a camera image of the windshield. Figure 6 provides three examples of the probabilistic map for recordings during phase 2. The black mark is the true gaze location and the heat map shows the confidence region, where bright pixels indicate higher probabilities. These figures visualize the confidence region predicted by the model. Notice that the size of the ellipses are not equal, which vary according to the uncertainty in the relationship between head pose and gaze given the orientation and position of the head. The white lines represent the 50% and 95% CIs. A smaller ellipse implies higher confidence in the data, while a large ellipse implies more ambiguity. Interestingly, the information is directly learned from the data. The proposed ideas can be extended to project the probabilistic map for the gaze onto the road camera image.

VII. Conclusions

The paper presented a novel probabilistic approach to detect gaze regions given the position and orientation of the driver's head. Instead of directly estimating the actual direction of gaze, which depends on the driver, cognitive load and the visual task, the approach estimate a visual attention map with the probability that the driver is looking to a given direction. We project confidence regions into the windshield, which convey the uncertainty associated with the relationship between head pose and gaze. This approach provides important information about visual attention of the driver, opening research opportunities for applications in navigation, infotainment, safety and communication systems.

Our future work includes mapping the confidence regions from the windshield to the road. This step will allow us to identify objects and events outside the vehicle that attract the attention of the drivers, opening new opportunities for ADAS (e.g., lack of awareness of pedestrians in the road). To achieve this goal, we need a transformation that projects the probability distribution onto a 2D map on the road image. Another area of further improvement is to train more powerful models to capture the relationship between gaze and head pose, which will result in smaller and more localized confident regions for the probabilistic models. An straightforward extension is to model the temporal dynamic of head pose, as opposed to evaluating isolated frames. Another alternative is to rely on deep learning architectures that are appropriate for this task. A limitation of the approach is the need of the headband to estimate the position and orientation of the driver's head. We will extend the work by using head pose estimations obtained with computer vision solutions such as OpenFace [2]. We are also planning to increase the size of the corpus to build more robust models.

References

[1] P. Angkititrakul, M. Petracca, A. Sathyanarayana, and J. Hansen. UT-Drive: Driver behavior and speech interactive systems for in-vehicle environments. In *IEEE Intelligent Vehicles Symposium*, pages 566–569, Istanbul, Turkey, June 2007.

[2] T. Baltrusaitis, P. Robinson, and L. P. Morency. Constrained local neural fields for robust facial landmark detection in the wild. In *IEEE International Conference on Computer Vision Workshops (ICCVW 2013)*, pages 354–361, Sydney, Australia, December 2013.

[3] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical Report CMU-CS-94-102, Carnegie Mellon University, Pittsburgh, PA, USA, January 1994.

[4] J.-C. Chien, J.-D. Lee, and L.-C. Liu. A fuzzy rules-based driver assistance system. *Mathematical Problems in Engineering*, 2015:1–14, 2015.

[5] M. C. Chuang, R. Bala, E. A. Bernal, P. Paul, and A. Burry. Estimating gaze direction of vehicle drivers using a smartphone camera. In *EEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2014)*, pages 165–170, Columbus, OH, USA, June 2014.

[6] A. Doshi and M. Trivedi. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 3(10):453–462, September 2009.

[7] A. Doshi and M. Trivedi. Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of vision*, 2(12):1–16, February 2012.

[8] L. Fletcher and A. Zelinsky. Driver inattention detection based on eye gaze - road event correlation. *The international journal of robotics research*, 28(6):774–801, June June.

[9] S. Jha and C. Busso. Analyzing the relationship between head pose and gaze to model driver visual attention. In *International Conference on Intelligent Transportation Systems (ITSC 2016)*, pages 2157–2162, Rio de Janeiro, Brazil, November 2016.

[10] S. Jha and C. Busso. Analysis of head pose estimation of drivers during naturalistic recordings using existing tools. In *International Conference on Intelligent Transportation Systems (ITSC 2017)*, volume Submitted, Yokohama, Japan, October 2017.

[11] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, October 2002.

[12] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A.*, A34(Part 5):827–828, September 1978.

[13] S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim. Real-time gaze estimator based on driver's head orientation for forward collision warning system. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):254–267, March 2011.

[14] N. Li and C. Busso. Analysis of facial features of drivers under cognitive and visual distractions. In *IEEE International Conference on Multimedia and Expo (ICME 2013)*, pages 1–6, San Jose, CA, USA, July 2013.

[15] N. Li and C. Busso. Evaluating the robustness of an appearance-based gaze estimation method for multimodal interfaces. In *International conference on multimodal interaction (ICMI 2013)*, pages 91–98, Sydney, Australia, December 2013.

[16] N. Li and C. Busso. Predicting perceived visual and cognitive distractions of drivers with multimodal features. *IEEE Transactions on Intelligent Transportation Systems*, 16(1):51–65, February 2015.

[17] N. Li and C. Busso. Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):980–992, April 2016.

[18] N. Li, J. Jain, and C. Busso. Modeling of driver behavior in real world scenarios using multiple noninvasive sensors. *IEEE Transactions on Multimedia*, 15(5):1213–1225, August 2013.

[19] National Highway Traffic Safety Administration. Preliminary statement of policy concerning automated vehicles. https://www.nhtsa.gov, 2013.

[20] E. Olson. AprilTag: A robust and flexible visual fiducial system. In *IEEE International Conference on Robotics and Automation (ICRA 2011)*, pages 3400–3407, Shanghai, China, May 2011.

[21] C. E. Rasmussen. Gaussian processes in machine learning. In O. Bousquet, U. von Luxburg, and G. Rätsch, editors, *Advanced Lectures on Machine Learning*, Lecture Notes in Computer Science, pages 63–71. Springer Berlin Heidelberg, Berlin, German, October 2004.

[22] M. Rezaei and R. Klette. Look at the driver, look at the road: No distraction! no accident! In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, pages 129–136, Columbus, OH, June 2014.

[23] G. H. Robinson, D. J. Erickson, G. L. Thurston, and R. L. Clark. Visual search by automobile drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 14(4):315–323, August 1972.

[24] S. Singh. Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Technical report, National Highway Traffic Safety Administration, Washington, DC, 2015.

[25] A. Tawari and M. Trivedi. Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos. In *IEEE Intelligent Vehicles Symposium (IV 2014)*, pages 344–349, Dearborn, MI, June 2014.

[26] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 46(6):629–646, May 2003.

[27] H. Zhang, M. Smith, and R. Dufour. A final report of safety vehicles using adaptive interface technology: Visual distraction. Technical Report Phase II: Task 7C, Delphi Electronics and Safety, Kokomo, Indiana, February 2008.