2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

Analyzing the Relationship Between Head Pose and Gaze to Model Driver Visual Attention

Sumit Jha and Carlos Busso

Abstract— **Monitoring driver behavior is crucial in the design of *advanced driver assistance systems* (ADAS) that can detect driver actions, providing necessary warnings when not attentive to driving tasks. The visual attention of a driver is an important aspect to consider, as most driving tasks require visual resources. Previous work has investigated algorithms to detect driver visual attention by tracking the head or eye movement. While tracking pupil can give an accurate direction of visual attention, estimating gaze on vehicle environment is a challenging problem due to changes in illumination, head rotations, and occlusions (e.g. hand, glasses). Instead, this paper investigates the use of the head pose as a coarse estimate of the driver visual attention. The key challenge is the non-trivial relation between head and eye movements while glancing to a target object, which depends on the driver, the underlying cognitive and visual demand, and the environment. First, we evaluate the performance of a state-of-the-art head pose detection algorithm over natural driving recordings, which are compared with ground truth estimations derived from AprilTags attached to a headband. Then, the study proposes regression models to estimate the drivers' gaze based on the head position and orientation, which are built with data from natural driving recordings. The proposed system achieves high accuracy over the horizontal direction, but moderate/low performance over the vertical direction. We compare results while our participants were driving, and when the vehicle was parked.**

I. INTRODUCTION

An effective in-vehicle active safety system has to consider the driver, the vehicle, and the road environment. This is a challenging task due to the unpredictable behavior of drivers who maneuver the vehicle to reach their goals. An important aspect to understand driver behavior is to estimate the direction of his/her visual attention, which can signal cognitive state [1], situational awareness [2] and attention level [3]. It can also play an important role in in-vehicle situated dialog systems [4]. In a controlled environment, gaze can be tracked with high accuracy, especially with affordable devices recently introduced in the market. However, robustness in naturalistic driving environment is not yet achieved (varying lighting conditions, uncontrolled head poses and obstructions due to hand and glasses).

For many applications such as detecting mirror-checking actions [2], [5], or lane change intentions [6], a coarse estimate of the glance direction is enough to infer his/her visual attention. For these cases, it is often posited that the driver's head pose can provide a coarse estimation of the driver's gaze [7]. The relationship between head pose and gaze, however, depends on multiple factors (e.g., when driving versus when the car is stopped, location of visual attention, cognitive and visual demands of the underlying driving task). Acknowledging that head pose does not provide the exact gaze [8], we hypothesize that head pose can provide cues to predict visual attention of the drivers.

The authors are with the Multimodal Signal Processing Lab at the University of Texas at Dallas, Richardson, TX-75080, USA sxj146830@utdallas.edu, busso@utdallas.edu

This paper explores models relying on the driver's head pose to infer his/her visual attention. How reliable are current systems to obtain accurate head pose estimates? How reliably can the head pose be used to estimate coarse gaze location? What kind of relationship do we observe between head pose and gaze? We address these questions by analyzing driver's natural head poses while looking at pre-defined markers on the windshield, mirrors, side windows, speedometer panel, radio, and gear. We compare the position of the target markers, assumed as the intended gaze direction, with the actual head pose estimated with an AprilTag-based headband. We propose linear regression models that are effective to predict gaze location, especially along the horizontal direction. The paper also studies the bias introduced by eye movement during glance actions concluding that the bias increases when the participants are driving the car. The analysis in this paper opens interesting research directions to infer driver visual attention which can serve as a building block for applications in safety, navigation, and infotainment systems.

II. RELATED WORK

It is particularly important to consider visual attention while studying driver behavior. Liang and Lee [9] studied the effect of cognitive and visual distractions, and their combination, concluding that visual distractions are the most detrimental type of distraction affecting the primary driving task. Given the role of visual distraction, several studies have proposed ADAS systems relying on the driver's gaze as a cue for driver behavior [10]–[12]. However, it is challenging to implement gaze estimation in a real-world scenario, due to changes in illumination, occlusions, and non-frontal faces.

Head pose is easier to estimate than gaze, however, their relationship depends on the driving task, visual and cognitive demands, and the driver. Glances are characterized by combination of pupil and head movements. While acknowledging that eye gaze is a better indicator, Zhang *et al.* [7] claimed that head pose alone can provide good information about driver's intentions. Several researchers have used head pose as a coarse indicator of the driver gaze, avoiding calculating the exact gaze from the pupils location. Doshi and Trivedi [6] analyzed the driver's intent to change lanes using either the drivers exact gaze or with his/her head pose. The study found that head pose alone can be more helpful in predicting change

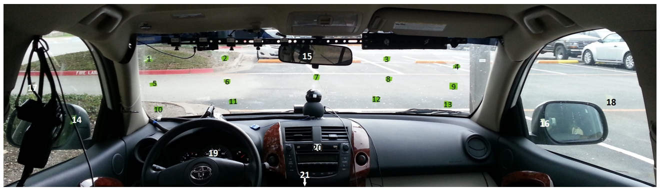


Fig. 1. Numbered markers placed at the windshield, mirrors, side windows, speedometer panel, radio, and gear. By asking the drivers to look at the markers, we establish reference labels for their gaze.

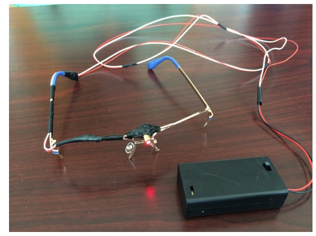


Fig. 2. Laser pointer mounted on a glasses' frame used in phase 3 (controlled head pose).

of lanes. A common approach is to classify the driver's gaze into predefined zones, using the head pose of the driver as input [13]–[15]. Murphy and Trivedi [16] proposed to monitor the driver's attention using a head tracking system. Rezaei and Klette [17] attempted to correlate the driver's head pose to potential hazards on the road.

The dynamics between head pose and gaze has been of a subject of interest in human physiology and driving behavior studies. Zangemeister and Stark [18] analyzed the correlation between the eye gaze and head movement, defining four cases: synchronous head and eye movements which occurred 34% of the time, late head movement which occurred 4% of the time, early head movement which occurred 43% of the time and early head movement with a late independent eye saccade which occurred 19% of the time. Robinson *et al.* [8] studied the head position of the driver during lane change and intersections. They studied the relationship between the eye and head movements in laboratory settings, concluding that head movement data are sufficient when the re-fixation angle (i.e., angle between two consecutive fixated point) is more than 20°. Doshi and Trivedi [19] studied the correlation between head pose and gaze for stimulus-driven (i.e., when the gaze shift is due to some external stimuli) or pre-meditated gaze shift (i.e., when the gaze shift is due to a planned task). At the fixation point, these two types of glances produces similar head poses.

This paper analyzes the information about the driver's gaze direction that can be obtained from head poses. First, we evaluate the effectiveness of head pose estimation algorithms in natural driving conditions. We benchmark the results by using a headband with multiple AprilTags. Then, we use regression models where the input is the position and orientation of the drivers' head, and the output is the predicted gaze location.

III. DATA COLLECTION

There are clear benefits in studying visual attention of the driver using data from naturalistic driving conditions, instead of simulators. Recordings in real cars bring challenges that computer vision algorithms have to solve before these systems can be deployed into ADAS. This study relies on recordings using the UTDrive platform, a vehicle preinstalled with multiple sensors (cameras, microphone array, data from the *controller area network* (CAN bus)). Our goal is to design a solution that can be easily implemented in regular vehicles. Therefore, we use a commercially available two-channel dash camera (Blackvue DR650GW-2CH). The front camera is positioned to capture the road, and is referred to as road camera (30fps, 1920×1080 resolution). The rear camera is positioned to capture the driver's face, and is referred to as face camera (30fps, 1280×720 resolution). While our analysis relies on offline processing, the dash camera has WiFi connection which facilitates real-time implementation.

One of the important challenges while studying the gaze of drivers is the absence of ground truth labels describing the true direction in which the driver is looking. We address this challenge by placing 21 numbered markers at predetermined locations (Figure 1). Markers 1 to 13 are placed on the windshield. Markers 14 to 16 are placed on the mirrors. Markers 17 and 18 are placed on the side windows. The last three markers are in the speedometer panel (19), radio (20), and gear (21). We ask the drivers to look at these predetermined locations following the protocol described in Section III-A. In total, 16 subjects participated in the recordings (10 males, 6 females).

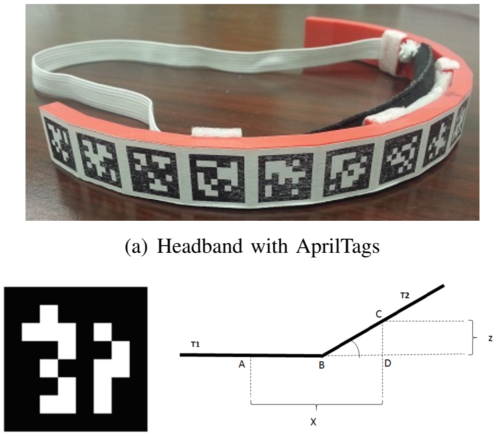
A. Protocol

The protocol for data collection has the following three phases recorded in order.

In **phase 1**, the vehicle is parked and the subject is seating in the driving seat. We ask the subject to look at the markers. A number is randomly generated and read aloud by the researcher conducting the data collection. We asked the subjects to look at each marker five times to collect natural glance behaviors. The aim of this phase is to compare the glance behavior when the driver does not have any other competing task requiring cognitive or visual resources. Furthermore, the subjects get familiar with our core task in a safe environment.

In **phase 2**, the driver repeats the procedure while driving the car. We select a route without many curves or intersections following the protocol approved by our *institutional review board* (IRB). The driver is asked to look at the markers only when it is safe to take the eye-off-the-road for short durations. The researcher points the target marker such that he/she does not have to spend time searching for the required location. The subject is asked to fixate his/her glance at the marker and then look back at the road immediately. The safety of the driver was our first priority. This phase provide natural glance behaviors while driving. We collect five repetitions per marker.

In **phase 3**, the car is parked again and the test is repeated. This time, the driver is asked to turn his/her head completely toward the marker, so that the head pose and gaze are aligned toward the target location. To ensure this requirement, the



(a) Headband with AprilTags

(b) AprilTag (c) Design of headband

Fig. 3. AprilTag- based headband. (a) headband with 17 AprilTags used during the data collection, (b) a sample of AprilTag and (c) the design of the headband.

subjects wore a glasses' frame with a laser pointer mounted at the center of the frame (Fig. 2). Their task was to direct the laser toward the marker. Each marker is requested three times to collect multiple samples for this controlled recording. The aim of this phase is to collect reference of ideal head pose that would point to the marker without any bias due to pupil movement.

B. Use of AprilTags for Head Pose Estimation

One of the goal in the recordings of this corpus is to assess the performance of current computer vision algorithms to track the orientation and position of the head. As analyzed in Section IV, current algorithms are not able to track the driver's face for some frames and, hence, they cannot provide head pose estimations, especially for cases with high rotations. The performance is also significantly affected by variations in illumination, which occurs frequently in a naturalistic driving scenario. Furthermore, estimating the position of the face in 3D using a single RGB camera is challenging. To be able to analyze the head pose regardless of the head rotation of the driver, we rely on additional reference markers with the use of AprilTags.

AprilTags [20] are 2D barcodes that are useful in augmented reality, robotic applications, and camera calibration. These tags contain unique black and white patterns that are very easy to detect from an image. Figure 3(b) gives an example. The relative position and orientation of each tag can be estimated from the size and shape of its projection into the 2D image plane. We designed a solid headband structure with 17 flat areas of 2x2 cm each, which are separated by an angle of 12°(Fig. 3(c)). We placed a different AprilTag to each of the 17 sides, where the tag size is 1.6cm x 1.6cm. Figure 3(a) shows the headband, which was used by the subjects during the data collection. Even when one of the tags is visible, we can derive the orientation of the headband, and, therefore, the head pose of the person wearing the headband.

C. Orientation of the Central Tag

For each frame, we detect the orientation of the central tag using the orientation of all the detected tags. The headband contains 17 tags. Given the structure of the headband, a subset of the tags will be visible by the face camera. Since the relative orientation of the tags with respect to each other is fixed, we can determine the orientation of the central tag with respect to each of the detected tags. The process involves simple homogenous transformation matrices. First, we calculate the homogeneous transformation between two tags adjacent to each other. Let $T1$ and $T2$ be 2 adjacent tags as shown in Figure 3(c). To obtain T2 from T1, the rotation about the Y axis is 12°and around the X and Z axes is 0°. The translation from the center of T1 to the center of T2 is:

$$\begin{aligned} x &= AB + BD = AB + BC\cos 12^\circ = 1.9781cm \\ z &= CD = BC\sin 12^\circ = -0.2079cm \end{aligned}$$

Therefore, the overall homogeneous transformation matrix from $T1$ to $T2$ is given by (see Fig. 3(c)).

$$^{T1}H_{T2} = \begin{bmatrix} \cos 12^\circ & 0 & \sin 12^\circ & 1.9781 \\ 0 & 1 & 0 & 0 \\ -\sin 12^\circ & 0 & \cos 12^\circ & -0.2079 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Once the relation between adjacent tags is known, the process can be repeated to calculate the orientation of each tag from the orientation of another tag. Using this method, the orientation of the central tag (tag #8) is calculated from all the detected tags. We observed that few outlier estimations could skew the result away from the true estimate. To minimize this problem, we take the median of the calculated values as the final orientation estimate.

We also calibrate the system to derive the 3D position of the markers using AprilTags. A key challenge is that the markers do not appear in the road or face cameras. Therefore, we place additional AprilTags on the markers and on certain locations in the vehicle (e.g., driver's seat). Then, we use a portable camera to take pictures including subsets of AprilTags from multiple perspective, from which we learn their homogeneous transformations. This approach gives the 3D coordinate of each of the markers, which are used for driver normalization (Sec. III-D) and performance evaluation (Sec. V-A).

D. Calibration of Tag Orientation to Head Pose

Each subject may have different variations between their true head orientation and the orientation estimated from the their headband (e.g., different placements of the headband). We compensate for these differences as follow. First, we define the homogeneous transformation between the head and headband as $^{Head}H_{tag}$. Let T_{tag} be the orientation of the headband and T_{head} be the orientation of the head (see Fig. 4). From T_{head}, we can obtain the equation of the line normal to the head's plane. In the data collected during phase 3, the driver's head pose is oriented toward the markers, so the normal vector from the head should cross the location

Fig. 4. Transformation from headband orientation to head orientation.

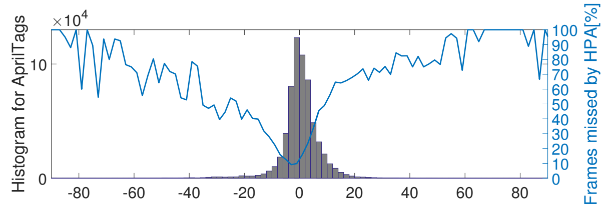
of the marker. From the 3D location of the markers (Sec. III-C) and T_{tag} obtained from the AprilTag-based headband, we solve for ${}^{Head}H_{tag}$ such that the mean square distance between the line and the marker location is minimized.

IV. Evaluating Head Pose Estimation

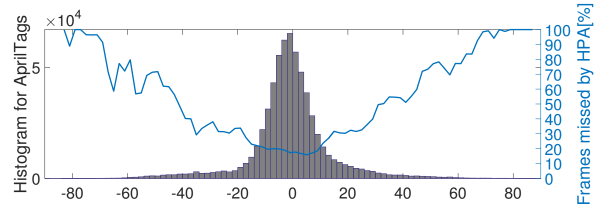
Ideally, an in-vehicle safety system has to rely on non-intrusive sensors to track the behaviors of the drivers. Therefore, it is important to have reliable computer vision algorithms to estimate the position and orientation of the head. This section compares a state-of-the-art *head pose estimation algorithm* (HPA) with the estimation derived from the headband. We use the head pose estimations from the headband as baseline, acknowledging that their estimates are accurate but not perfect. We selected the publicly available head pose estimation algorithm in Baltrusaitis *et al.* [21]. We evaluate each frame in the videos, extracting estimations from the headband and the automatic head orientation algorithm. As illustrated in Figure 4, the orientations of the head and headband are not exactly the same. For each subject, we estimate rotation matrices by estimating rotation angles between them for all the frames, which are then averaged.

In 73.2% of the frames, we obtain estimates from both AprilTag-based headband and head pose estimation algorithm. In 5.3% of the frames, the AprilTag-based headband was not able to provide an estimate. In 24.5% of the frames, the HPA was not able to provide an estimate. In 21.51% of the frames, only the AprilTag-based headband gave an estimate, while in 2.3% of the frames, only the HPA provided an estimate. Few frames were not detected by the AprilTag-based headband because of blurred images due to motion, while the face could be tracked. For some frames (3%), we do not have either of the estimates, because of very high or low exposure to light or the driver's face was completely out of the frame (e.g., looking back).

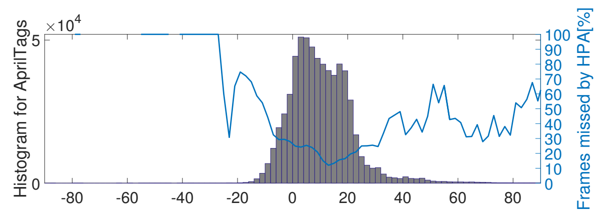
We consider the data for which AprilTags were detected but the HPA did not detect the face. The major reason for missed those frames is when the driver turned away from the frontal position. Figure 5 shows the percentage of frames missed by the HPA at different angles of rotation for α (roll), β (yaw) and γ (pitch) directions (solid line). The histograms show the number of frames detected by AprilTag in each bin, for different angles (gray bars). Figure 5 shows that most



(a) angle α (Roll)



(b) angle β (Yaw)



(c) angle γ (Pitch)

Fig. 5. Percentage of the frames missed by the *head pose estimation algorithm* (HPA) as a function of the angles (solid line). The figure also shows the histogram of the head pose angles estimated with the AprilTag-based headband.

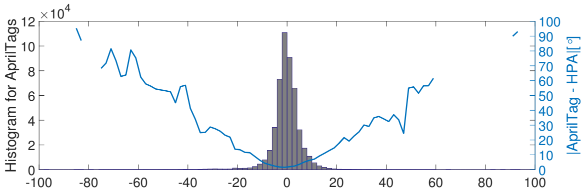
head orientations are between -10°and 10°, for roll and yaw. For pitch, the distribution is skewed to the right (-10°and 20°). Less frames are missed by the HPA when the angles are close to zero – i.e. when the face is frontal. When the magnitude of the head orientation is larger than 20°, more frames are missed by the HPA. The encouraging result is that there are few frames under this condition (see Fig. 5).

We also compare the differences in the estimation as a function of the angle. For this analysis we consider the frames where both methods provided an estimate (73.2% of the frames). Figure 6 plots the mean absolute differences between the estimates (solid line). The figure also plots the histogram with the distribution of different angles estimated by the AprilTag-based headband. The estimates of both approaches are similar for frontal or semi frontal faces (i.e., yaw between -40°and 40°, pitch between -20°and 20°). Yaw is one of the most important angle for in-vehicle applications (horizontal glance patterns). For yaw, the differences are less than 5°when the angles are between -20°and 30°. For angles with higher magnitude, there are fewer frames provided by the HPA so the average results are less reliable.

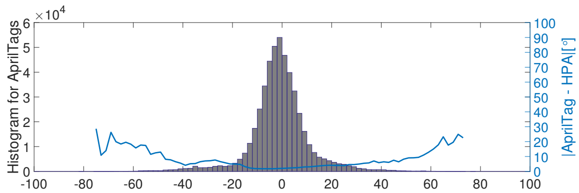
V. Prediction of Driver's Gaze

A. Regression Model to Estimate Gaze

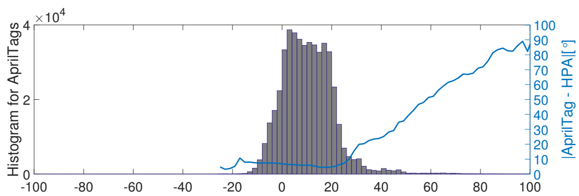
We explore linear regression to study the dependency between the orientation and position of the head and gaze.



(a) Angle α (Roll)



(b) Angle β (Yaw)



(c) Angle γ (Pitch)

Fig. 6. Absolute differences between the estimates from the *head pose estimation algorithm* (HPA) and the AprilTag-based headband (solid line). The figure shows the histogram of the head pose angles estimated with headband (bars).

Given the challenges in detecting head pose using automatic algorithms (Sec. IV), this analysis relies on the estimates provided by the AprilTag-based headband. The independent variables correspond to the head position (x,y,z) and head orientation (α – roll, β – yaw, γ – pitch). The dependent variable are the gaze location characterized by the 3D position of the target marker (x_0,y_0,z_0), using the method described in Section III-C. Equation 1 gives the proposed model for x_0. Similar models are used for y_0 and z_0.

$$x_0 = a_0 + a_1 x + a_2 y + a_3 z + a_4 \alpha + a_5 \beta + a_6 \gamma \qquad (1)$$

The regression models are separately trained and tested with data from the three phases. The evaluation considers driver independent partitions for building and evaluating the models. The training set includes data from ten drivers, and the testing set includes data from the remaining six drivers. The drivers were randomly selected in these partitions.

Table I lists the R^2 values observed on the train and test partitions for the three phases. The R^2 represents the amount of variability in the dependent variable that the model account for. The table gives the R^2 values on the train and test partitions, where their small difference indicates that the regression models have good generalization. Across phases, we observe that the predictability in the Y (vertical axis) and Z (depth axis) directions are significantly lower than the predictability in the X (horizontal axis) direction. Most in-vehicle applications require good resolution on the horizontal direction (e.g., lane change, turns, mirror checking actions). Therefore, it is encouraging that the R^2 over the test partition is 0.65 for natural glance behaviors while driving (phase 2). The best performance of the regression models is for phase 3 (controlled recordings where head pose is directed toward the target marker). The R^2 values drop for phase 1 and phase 2, where the drivers complete the task with natural glance behaviors. The worst performance of the regression models is in phase 2.

TABLE I

REGRESSION MODELS TO ESTIMATE GAZE FROM HEAD ORIENTATION AND POSITION. THE RESULTS ARE PROVIDED FOR PHASE 1 (NATURAL HEAD POSE WHEN CAR IS PARKED), PHASE 2 (NATURAL HEAD POSE WHILE DRIVING), AND PHASE 3 (CONTROLLED HEAD POSE).

Phase	Gaze	R^2		constant	position			rotation		
		Train	Test	a_0	a_1	a_2	a_3	a_4	a_5	a_6
Phase 1	x_0	0.78	0.77	-0.11	**-1.04**	**-1.14**	**-0.51**	-0.05	**-1.09**	**0.12**
	y_0	0.36	0.12	0.01	**2.15**	**1.88**	**0.59**	0.00	**0.12**	**-0.10**
	z_0	0.25	0.10	**0.33**	**-3.32**	**-2.25**	0.09	**-0.19**	**-0.57**	**-0.46**
Phase 2	x_0	0.69	0.73	**-0.30**	**-1.07**	**-2.26**	**-1.17**	**-0.20**	**-1.24**	0.01
	y_0	0.36	0.16	**-0.24**	**3.16**	**2.27**	**0.35**	**0.19**	**0.22**	0.01
	z_0	0.24	0.12	**0.37**	**-4.50**	**-1.09**	**0.28**	**-0.47**	**-0.78**	**-0.30**
Phase 3	x_0	0.91	0.87	**0.16**	**0.45**	**-1.65**	0.16	-0.01	**-0.70**	-0.01
	y_0	0.66	0.31	**-0.19**	**1.35**	**2.08**	**0.20**	-0.01	0.02	**-0.08**
	z_0	0.31	0.25	**0.65**	**-2.00**	**1.29**	**1.68**	-0.05	**-0.22**	-0.05

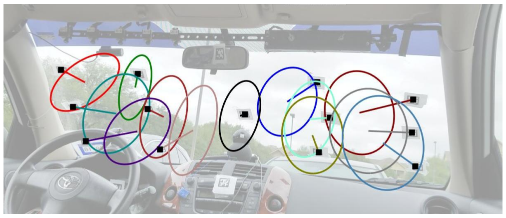
Coefficients in bold are statistically significant at p-value$<$0.05.

Table I also lists the coefficients of the regression models. If we assert significance at p-value$<$0.05, most of the independent variable are useful to estimate the gaze location. Table I highlights in bold these coefficients. We notice that a_5 (β, yaw) is important for estimating x_0. The independent variables for the head's position (i.e., a_1-a_3) are important to estimate the target gaze location.

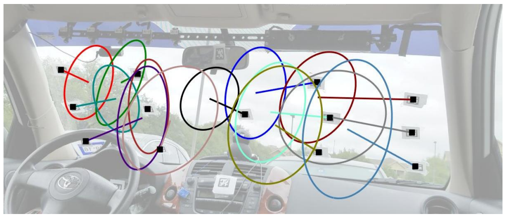
B. Bias Introduced by Eye Movement

The final analysis explores the bias introduced by eye movement while glancing at the target marker. We project the head direction vectors into the windshield by using the homogeneous transformation between the head and headband (${}^{Head}H_{tag}$) estimated in Section III-D. For each marker, we estimate an ellipse centered at the mean of the head pose projections, with radius given by the standard deviation. This analysis identifies the head pose direction used by the drivers to look at the marker. The difference between the head pose direction and the marker position is due to the eye movement, which this analysis refers to as bias. This bias is illustrated by a solid line from the center of an ellipse and the corresponding marker position.

Figure 7 shows that the bias is low for markers in front of the drivers, and increases for markers further from the drivers. When the angle increases, the variability of the head pose for each gaze location also increases (e.g., bigger ellipses), as the driver depends more on eye movement rather than head rotation. The figure shows interesting differences between phase 1 (when the car is parked, Fig 7(a)) and 2 (when the subject is driving, Fig. 7(b)). First, the size of the ellipses are bigger when the subject is driving indicating higher variability. We hypothesize that different driving conditions affect the manner in which the driver glances at the target marker. Another interesting observation is that



(a) Natural head pose distribution when car is parked



(b) Natural head pose distribution when driving

Fig. 7. Distribution of head pose directions projected into the windshield. Each ellipse characterizes a target marker, where its center corresponds to the mean of the head pose direction and it size represents its standard deviation. The solid line connecting the marker to the center of its ellipse is the mean bias introduced by eye movement during eye glance actions.

the average bias (solid lines) is higher, indicating that the center of the ellipses on the right side of the windshield are shifted to the left. The primary driving task increases the cognitive/visual demand, so drivers complete the glance action by moving the eye limiting the head rotation.

VI. CONCLUSIONS

The paper analyzed the relationship between head pose and gaze, in naturalistic driving environment. It presented a carefully designed corpus where glance behaviors are analyzed while driving and when the car is parked. The paper evaluated the performance of a state-of-the-art head pose estimation algorithm showing that 24.5% of the frames are not correctly processed, suggesting open challenges in creating computer vision algorithms that are robust in vehicle environment. The paper also demonstrated that linear regression models are effective to estimate the gaze location, having reasonable performance along the horizontal direction. Finally, the paper analyzed the bias introduced by eye movement during glance actions, which increases when the participants are driving. The strong correlation between head pose and the gaze direction can be utilized to estimate the visual attention of the driver with varying level of confidence depending on the situation.

Most of the analysis relied on estimates derived from the AprilTag-based headband. It is important to develop robust solution based on non-intrusive sensors to estimate the orientation and position of the driver's head. Various techniques can be investigated, including using multiple cameras [13], infrared sensors, or time-of-flight cameras.

ACKNOWLEDGMENT

We thank Dr. John Hansen for his support with UTDrive.

REFERENCES

[1] N. Li and C. Busso, "Predicting perceived visual and cognitive distractions of drivers with multimodal features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 51–65, February 2015.

[2] 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*, vol. 17, no. 4, pp. 980–992, April 2016.

[3] N. Li, J.J. Jain, and C. Busso, "Modeling of driver behavior in real world scenarios using multiple noninvasive sensors," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1213–1225, August 2013.

[4] T. Misu, "Visual saliency and crowdsourcing-based priors for an in-car situated dialog system," in *International conference on Multimodal interaction (ICMI 2015)*, Seattle, WA, USA, November 2015, pp. 75–82.

[5] N. Li and C. Busso, "Driver mirror-checking action detection using multi-modal signals," in *The 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems*, Seoul, Korea, September-October 2013, pp. 101–108.

[6] A. Doshi and M.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*, vol. 3, no. 10, pp. 453–462, September 2009.

[7] 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.

[8] 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*, vol. 14, no. 4, pp. 315–323, August 1972.

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

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

[11] L. Fletcher and A. Zelinsky, "Driver inattention detection based on eye gaze - road event correlation," *The international journal of robotics research*, vol. 28, no. 6, pp. 774–801, June June.

[12] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, no. 5, pp. 357–377, October 2002.

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

[14] 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*, vol. 12, no. 1, pp. 254–267, March 2011.

[15] 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)*, Columbus, OH, USA, June 2014, pp. 165–170.

[16] E. Murphy-Chutorian and M.M. Trivedi, "HyHOPE: Hybrid head orientation and position estimation for vision-based driver head tracking," in *IEEE Intelligent Vehicles Symposium (IV 2008)*, Eindhoven, The Netherlands, June 2008, pp. 512–517.

[17] 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)*, Columbus, OH, June 2014, pp. 129–136.

[18] W.H. Zangemeister and L. Stark, "Types of gaze movement: variable interactions of eye and head movements," *Experimental Neurology*, vol. 77, no. 3, pp. 563–577, September 1982.

[19] A. Doshi and M.M. Trivedi, "Head and eye gaze dynamics during visual attention shifts in complex environments," *Journal of vision*, vol. 2, no. 12, pp. 1–16, February 2012.

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

[21] 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)*, Sydney, Australia, December 2013, pp. 354–361.